Australian Energy Market Commission

TECHNICAL REPORT: OPTIONAL FIRM ACCESS

Transmission Frameworks Review

AEMC Staff Paper

11 April 2013

REVIEW

Reference: EPR0019
Technical Report: Optional Firm Access
11 April 2013

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: EPR0019

Citation

AEMC 2013, *Transmission Frameworks Review*, Technical Report: Optional Firm Access, 11 April 2013, Sydney.

About the AEMC

The Council of Australian Governments (COAG), through its then Ministerial Council on Energy (MCE), established the Australian Energy Market Commission (AEMC) in July 2005. In June 2011, COAG established the Standing Council on Energy and Resources (SCER) to replace the MCE. The AEMC has two main functions. We make and amend the national electricity, gas and energy retail rules, and we conduct independent reviews of the energy markets for the SCER.

This report has been prepared by AEMC staff. It does not necessarily represent the views of the Commission or any individual Commissioner.

Contents

1 Introduction

1.1 Purpose of this Document

This document provides further detail of the proposed Optional Firm Access (OFA) model, building on the description presented in part 1 of the Transmission Frameworks Review: Final Report.[1]

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

- presenting a complete picture of how the OFA model would operate;
- providing confidence that the model contains no irresolvable difficulties or inconsistencies;
- facilitating qualitative and quantitative analysis of the possible impacts of the model on NEM efficiency;
- 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 further progressing of the model.

1.2 Changes from the Technical Report (August 2012)

This document is a revised version of the Technical Report (August 2012).[2]

We have made some modifications to the optional firm access model in response to stakeholders' suggestions, which for the most part seek to simplify the model without compromising its objectives. The main changes from the Technical Report (August 2012) are that:

- firm access would be limited by rated capacity, rather than by availability, and should be more attractive to intermittent generators as a result;
- the firm access standard would only apply during the set of normal operating conditions, resulting in a single-tier standard. The scaling factors that previously applied under lower tiers of the standard would no longer apply; and
- access pricing would use different forecasting models for the short, medium and long terms to avoid spurious accuracy and allay concerns about the objectivity of the forecasts.

1 AEMC, *Transmission Frameworks Review*, Final Report, 11 April 2013. All subsequent references to the Final Report mean this document.

2 AEMC, *Transmission Frameworks Review,* Technical Report: Optional Firm Access, 15 August 2012. All subsequent references to the Technical Report (August 2012) mean this document.

1.3 Acknowledgement

This Technical Report has been prepared by the staff of the Australian Energy Market Commission (AEMC). It does not necessarily represent the views of the Commission or any individual Commissioner.

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

1.4 Structure of this Document

This document is structured as follows:

- Chapter 2 describes the fundamental concepts of access, firmness and optionality that provide the foundations 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 10 consider each of the main model elements in turn.
- Chapter 11 considers potential changes to market behaviour that might be induced by the model.
- Chapter 12 provides technical detail on concepts, algorithms and processes used in the model.

Those chapters describing the model design (chapter 2 and chapters 4 to 10) are each subdivided 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 is being revisited because of the operational and commercial issues that it creates, relating to congestion management and access certainty. 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.

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* that have procured firm access.

The costs that this obligation creates for TNSPs are recovered from firm generators in *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.

2.2 Design Blueprint

2.2.1 Regional and Local Markets

NEM dispatch is conventionally thought of as a *regional* market clearing process operating as follows:[1]

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.[2]

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 common 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

   (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.

In summary, there is an inconsistency between:

---

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

2 Transmission losses are ignored in this discussion and in general in this document. They are not pertinent to the OFA model, which doesn't change the way they are calculated and applied. They are discussed in section 12.5 (Transmission Losses).

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

It is this fundamental inconsistency within the NEM design which lies at the root of problems such as disorderly bidding and access uncertainty. To address these issues, the inconsistency must be addressed.

2.2.2 Dispatch and Network Access

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[3] and be paid the *local price* for its output.
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:

$$\begin{aligned} \text{Pay\$} &= \text{Pay\$}_{\text{dispatch}} + \text{Pay\$}_{\text{network}} \\ &= \text{LMP} \times \text{G} + (\text{RRP} - \text{LMP}) \times \text{A} \end{aligned} \tag{2.1}$$

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) resolves to become the more familiar

$$\text{Pay\$} = \text{RRP} \times \text{G}$$

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. 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. Thus it is possible to change arrangements for the provision of network access without making any corresponding changes to the dispatch process.

3 Ie dispatched if the local price exceeds its offer.

2.2.3 Firm Access

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.[4] The level of network access is set independently of dispatch; instead it is dependent on the following factors:

- the amount of *firm access* agreed with the local TNSP;
- generator capacity and availability;
- transmission availability; and
- the firm access level, capacity and availability of nearby generators.

Firm access service is a new transmission service in the OFA model, provided by each TNSP to generators in its region.[5] Generators can choose the amount of firm access service that they wish to procure from their TNSP and are charged by the TNSP for this.

The maximum level of network access for each generator is equal to its capacity.[6] The reliability with which the network access is at, or close to, that maximum level is referred to as the *firmness* of access. A generator that procures a *firm access service* will receive a firmer level of network access than one who has not.

Because dispatch access is unchanged in the OFA model, it is only referred to in this section of this document. Thus, in other sections, *network access* will generally be referred to simply as *access*.

2.2.4 Settlement Balancing

Equation (2.1) above can be rewritten as follows

$$\text{Pay\$} = \text{RRP} \times \text{G} + (\text{RRP} - \text{LMP}) \times (\text{A} - \text{G}) \tag{2.2}$$

The first term is exactly the same as the generator settlement payment in the existing NEM design.[7] The second term is a *new* settlement payment introduced in the OFA model and is referred to as *access settlement*. A fundamental principle of the OFA model

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

5 It may be useful to clarify the OFA model terminology at this point, since it can be a little confusing. Access always means network access, being the MW volume on which RRP is paid in AEMO settlement. Firmness means the reliability with which anything (in this case access) is provided. Firm access is a service provided by TNSPs. The design of the OFA model ensures that a generator that has procured firm access service does actually obtain a firm level of network access. See Appendix A for the defined terms used in the OFA model.

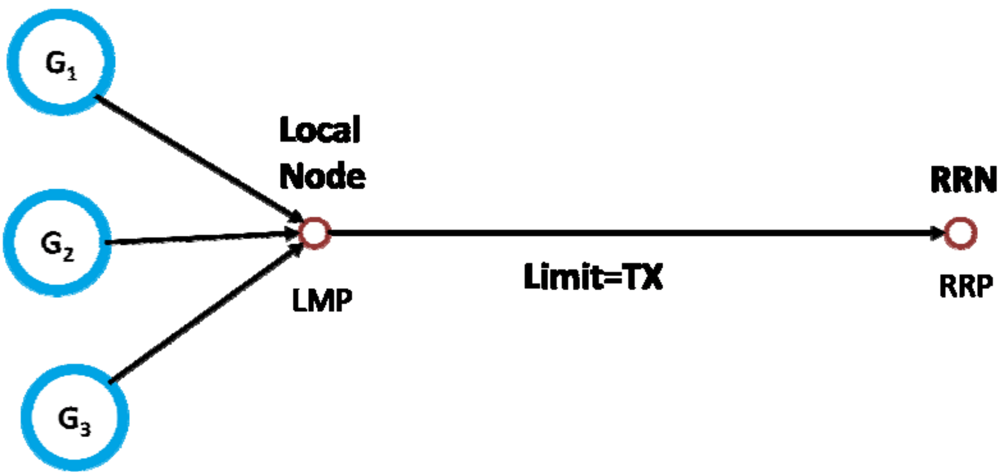
6 Ie its rated or nameplate generating capacity.

7 Again, ignoring transmission losses for simplicity.

is that *aggregate* settlement payments are unchanged. To achieve this, access settlement – when aggregated across all generators – *must* net out to zero.

The implications of settlement balancing are discussed below for a simple two-node network, shown in Figure 2.1. The more general situation of a meshed, multi-node network is discussed in section 4.2.6 (Flowgate Pricing and Settlement Balancing).

Figure 2.1 Two-node network example



In this example, all generators are connected to the same node, so LMP is the same for each generator.[8] 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:

$$\text{Pay\$}_i = (\text{RRP} - \text{LMP}) \times (A_i - G_i) \qquad (2.3)$$

The total payment across all generators is then:

$$\textstyle\sum_i \text{Pay\$}_i = (\text{RRP} - \text{LMP}) \times (\sum_i A_i - \sum_i G_i)$$

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

$$\textstyle\sum_i A_i = \sum_i G_i$$

But, since the line is congested we know that:

$$\textstyle\sum_i G_i = \text{TX}$$

8 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.

where:

TX = transmission capacity

Therefore, a sufficient condition for settlement balancing is:

$$\sum_i A_i = TX \quad (2.4)$$

Thus the setting of generator access levels involves sharing, or allocating, the available transmission capacity between various generators. A similar result holds in general on a meshed network, as explained below.

2.2.5 Flowgates, Usage and Entitlements

In the OFA model, the locations in the shared network where congestion *may* occur are referred to as flowgates. In the simple network shown in Figure 2.1 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.[9] Locations where congestion actually occurs are called *congested flowgates*.[10] In the simple network example, the flowgate is congested. In a real, meshed network, *several* flowgates may be congested at any point in time.[11]

Table 2.1 below lists the variables that are defined in the OFA model for each congested flowgate, and shows the values they take in the simple two-node example.

Table 2.1 Settlement variables and their equivalents in the two-node model

Variable	Acronym	Description	Value in 2-node model
flowgate price	FGP	the value of network access through the flowgate	RRP-LMP
flowgate usage	U	the amount of a generator's output that flows through the flowgate	G
flowgate entitlement	E	the amount of network access that a generator is allocated through the flowgate	A
flowgate capacity[12]	FGX	the maximum aggregate flowgate usage which the flowgate can accommodate	TX

9 In relation to stability constraints.

10 In the case of stability constraints, the congestion does not occur at a particular defined location but, nevertheless, a congested flowgate still exists *conceptually*.

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

12 Flowgate capacity is conceptually different to - albeit related to - transmission capacity, as discussed in section 12.2.3 (Transmission Capacity versus Flowgate Capacity).

Using the new terminology, we can rewrite equation (2.3) as:

$$Pay\$_i = FGP \times (E_i - U_i) \tag{2.5}$$

Equation (2.4) can also be rewritten, as:

$$\sum_i E_i = FGX \tag{2.6}$$

In the OFA model, equations (2.5) and (2.6) apply to *all* congested flowgates in all meshed networks. They are the basic building blocks of access settlement and are discussed further in chapter 4 (Access Settlement).

2.2.6 Access Allocation

When transmission capacity is allocated, *preferential* access is given to *firm generators*: those who have procured firm access service from the TNSP. Access is only provided to *non-firm generators* (those who have *not* procured firm access) if and when all firm generators have been provided with their target access level. Algorithms for allocating access between firm and non-firm generators are described in chapter 4 (Access Settlement).

2.2.7 Firm Access Standard

Equation (2.6) means that, *overall*, access can only be as firm as the transmission network: at the extreme, if there is no flowgate capacity, there can be no access. In the OFA model, a TNSP is required to ensure that a sufficient level of flowgate capacity is available to provide the necessary access firmness to all firm generators, individually and concurrently. This requirement is called the *firm access standard* (FAS). The FAS ensures that firm generators are provided with at least a specified level of access firmness. The FAS takes no account of non-firm generators, who will therefore receive an inferior level of access firmness.

Even for firm generators, access is required to be *firm* but not *fixed*. That is to say, the FAS allows firm generators' allocated access to be below target under specified circumstances: for example, when there are transmission outages. The FAS is discussed further in chapter 5 (Firm Access Standard).

2.2.8 Access Charges

Because a TNSP is required to expand and maintain its transmission network so as to comply with the FAS requirement, it incurs costs in providing firm access service to generators.[13] This cost is recovered from firm generators, through an *access charge*. The access charge is determined when new firm access is agreed. It is fixed[14] for the life of

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

14 Except for some defined indexation.

the *firm access agreement*, to ensure maximum financial certainty for firm generators. The charge is based on the forecast *incremental cost* associated with the new access.

A TNSP has no 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 (Access Pricing).

2.2.9 Flowgate Support

The discussion above considers situations where:

- there is no congestion and so *LMP=RRP*; or
- there is congestion which causes *LMP<RRP*.

A third possibility is that congestion causes *LMP>RRP*. This occurs where a generator's dispatch helps to *relieve* transmission congestion: the premium in the LMP reflects the value to the market of it doing this. In the OFA model, such generators are referred to as *flowgate support* generators. Irrespective of whether they are firm or non-firm, these generators are always paid the RRP. This is done by setting the entitlement level equal to the usage level, so that access settlement payments are zero (from equation 2.5). Usage - and hence entitlement - is based on dispatch so, for flowgate support generators, access and dispatch are *not* delinked. The rationale for this design decision is discussed in section 2.3.9 (Flowgate Support and Constrained-on Generators).

2.2.10 Summary

Setting access levels based on access agreements with TNSPs, rather than on dispatch, is a conceptually simple change, but it addresses many of the outstanding transmission issues in the transmission frameworks review, by simultaneously defining the level of access that generators are *entitled* to and the level of transmission capacity that TNSPs are *obliged to provide.* Clear rights and obligations are the foundation stone of an efficient market. The OFA model builds on this simple premise to construct a new access framework for the NEM.

2.3 Design Issues and Options

2.3.1 Disorderly Bidding

In the current NEM design, disorderly bidding occurs when there is congestion within a region. The driver for this behaviour is that network access is linked to dispatch level: a generator can only maintain its network access by maintaining its dispatch level and, during periods of congestion, it must reduce its offer price to ensure this.

Therefore, *any* NEM design that de-links network access from dispatch level will solve the problem of disorderly bidding.[15] Here, the delinking is just the first step in addressing several issues, of which disorderly bidding is only one.

2.3.2 Financial Certainty for Generators

A generator's *operating margin* is the difference between its revenue from AEMO settlement and its variable generating costs. If, for simplicity, we assume that a generators has a fixed *marginal cost* of generation, *C*, then its variable generating costs are $G \times C$. Using the revenue formula in equation (2.1) its operating margin in the OFA model is:

$$\begin{aligned} \text{Margin\$} &= \{\text{LMP} \times \text{G} + (\text{RRP} - \text{LMP}) \times \text{A}\} - \text{C} \times \text{G} \qquad (2.7) \\ &= (\text{LMP} - \text{C}) \times \text{G} + (\text{RRP} - \text{LMP}) \times \text{A} \\ &\equiv \text{dispatch margin} + \text{access margin} \end{aligned}$$

Note firstly that a generator is dispatched only if $LMP \geq$ *Offer Price*.[16] Therefore, as long as a generator bids *at cost (Offer Price = C)* or higher, the dispatch margin is never negative. This result reflects another fundamental principle of the OFA model: that a generator never *regrets* being dispatched. A positive dispatch margin, no matter how small is better than no dispatch margin at all.[17]

For a generator investor, the margin *certainty* is critical. Leaving aside the uncertainty of RRP itself (which can be hedged through forward contracts, as discussed in the next section), a major determinant of margin certainty is access firmness. In the current NEM design, if a generator is not dispatched, it has zero access. This means it loses its dispatch margin *and* its access margin. In the OFA model, access is independent of dispatch, so only the dispatch margin is lost, not the access margin. Consider a situation where transmission congestion causes – or coincides with – very high regional prices: RRP could be as high as $10,000/MWh, whereas LMP might be as low as $20/MWh.[18] In these circumstances, the loss of dispatch margin is immaterial; a loss of access margin is critical.

Since margin certainty now depends upon access firmness rather than dispatch firmness, firm generators will have more certainty than they do presently. However, non-firm generators will generally have *less* margin certainty, because their access

---

15 Assuming that it is narrowly characterised as bidding -$1,000 when there is intra-regional congestion. The bidding of flowgate support generators at the market price cap – which could also be characterised as disorderly bidding – is discussed further in section 2.3.9 (Flowgate Support and Constrained-on Generators).

16 As noted in section 2.2.2 (Dispatch and Network Access) this is a fundamental right associated with dispatch access, which is provided to all generators.

17 Clearly, if it is not dispatched, its dispatch margin is zero.

18 Based on the cost of the generators behind the constraint.

firmness in the OFA model is likely to be *worse* than their dispatch firmness currently.[19]

Another way of describing this feature is that firm access service provides a hedge against dispatch risk.

Recall equation (2.2) from above:

$$\text{Pay\$} = \text{RRP} \times \text{G} + (\text{RRP} - \text{LMP}) \times (\text{A} - \text{G}) \tag{2.8}$$

The first term is the current AEMO settlement and the second term is the new access settlement. If dispatch is reduced when regional prices are high, the amount in the first term falls but the amount in the second term rises to offset that fall. Thus, the access settlement payment *hedges* the generator against dispatch risk, similarly to how forward contract payments hedge RRP risk.

It can be seen from equation 2.8, that generators who are *access-long* ($A>G$) are paid *from* access settlement and those who are *access short* ($A<G$) must pay *into* access settlement. Since aggregate access settlement payments are zero, the total payments into access settlement must balance the total payments from access settlement. Informally, this situation can be thought of as access-short generators *compensating* access-long generators when the former's dispatch *above* their access levels causes the latter to be constrained-off *below* their access levels.

2.3.3 Forward Trading

The analysis above demonstrates that firm access reduces the margin risk associated with dispatch uncertainty. However, a major driver of margin risk remains: RRP volatility. Generators hedge this risk currently by selling forward contracts: swaps (or other derivative structures) against the RRP. A generator will receive payments under a forward swap equal to:

$$\text{Pay\$} = \text{F} \times (\text{FP} - \text{RRP})$$

Where:

F is the quantity of forward sales

FP is the forward price

When this payment is added to the equation (2.7) the adjusted operating margin becomes:

$$\begin{aligned}\text{Margin\$} &= (\text{LMP} - \text{C}) \times \text{G} + (\text{RRP} - \text{LMP}) \times \text{A} + \text{F} \times (\text{FP} - \text{RRP}) \\ &= \text{F} \times (\text{FP} - \text{C}) + (\text{A} - \text{F}) \times (\text{RRP} - \text{C}) + (\text{G} - \text{A}) \times (\text{LMP} - \text{C})\end{aligned}$$

19 Put another way, since the overall level of access firmness is dependent on transmission firmness, providing greater access firmness for firm generators inevitably means providing lower access firmness for non-firm generators.

= forward margin + spot regional margin + spot local margin (2.9)

Exposure to RRP is through the middle term: the spot regional margin.[20] The exposure depends upon the relative levels of access and forward sales and is independent of dispatch.

The middle term in equation (2.9) highlights the concerns that generators have under the existing NEM design and how these are addressed by the OFA model. In the current NEM, $A=G$, so if a generator is constrained off (ie *not* dispatched, because of congestion), its access is reduced and any high RRP will adversely affect its margin (because $A<F$). In the OFA model, a generator that procures sufficient firm access can be confident that A will be reliably higher than F under a range of transmission conditions and so risks from congestion coinciding with high RRP are substantially mitigated.

The spot local margin, $(G\text{-}A) \times (LMP\text{-}C)$, will always be non-negative for a generator that is fully available and bids at cost, because:

- If $(LMP\text{-}C)>0$ then the generator is fully-dispatched and so $(G\text{-}A)\geq 0$.[21]
- If $(LMP\text{-}C)<0$ then the generator is not dispatched and so $(G\text{-}A)\leq 0$.
- If $LMP=C$ then the margin equals zero.

The situation where a generator is unavailable is more complex. Currently, if a generator is unavailable, it cannot be dispatched and so cannot earn RRP, irrespective of whether there is any congestion. That means that the generator is *short* against RRP, unable to back its forward contracts and exposed to financial losses is RRP is high. Under the OFA model, when a generator is unavailable it will still receive payments from access settlement[22] equal to $A \times (RRP\text{-}LMP)$. If there is congestion, so that LMP is *low* relative to RRP, this can provide effective backing for the generator's forward contracts. However, if there is no congestion, RRP equals LMP, there is zero access payment, and the generator has the same exposure as currently.

2.3.4 LMP rather than Offer Price

The “no regrets” principle discussed above could have been achieved by paying the generator its offer price rather than LMP, so that equation (2.1) becomes instead:

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

20 A generator may also have exposure to RRP in the last term, when there is no congestion and so RRP = LMP. This situation is the same as in the current NEM design and the RRP risks are unaffected by the OFA model.

21 Recalling that access can never exceed generator capacity in the OFA model.

22 Note that this is a change to the OFA model that was proposed in the Second Interim Report, where an unavailable generator would receive no access.

LMP has been chosen in the design for two reasons: conceptual and pragmatic. Conceptually, LMP is the clearing price in the local market and this is what a generator with dispatch access, but no network access is entitled to. Pragmatically, if generators were paid at offer price, this would create a pay-as-bid market[23] at each generator node; when there was no congestion in a region, this pay-as-bid market would extend across the region. Pay-as-bid markets can be inefficient, because they encourage generators to rebid to earn close to the clearing price, potentially leading to disorderly bidding.[24]

Effective rebidding would in any case see generators bidding at, and earning, close to LMP. It is preferable, therefore, just to pay LMP in the first place and avoid the rebidding problem.

2.3.5 Firm Access, not Fixed Access

Access payments in the OFA model are similar in some ways to payments made under a regime of Fixed Transmission Rights (FTRs), as seen in some electricity markets in the US and elsewhere. A key difference, though, is that FTRs generally provide a *fixed* MW level of network access. Since transmission capacity still varies, but total access is fixed, transmission will at times exceed total access (creating an access settlement *surplus*) and at other times fall short of total access (creating an access settlement *deficit*). FTR markets absorb these surpluses and deficits by *smearing* settlement payments across settlement periods.

FTR markets typically have highly-meshed transmission systems, meaning that extreme settlement deficits or surpluses are unlikely. In the NEM's much less meshed network, a fixed access approach could give rise to very large deficits in some settlement periods which could not possibly be recovered from surpluses in other periods. Settlement would become untenable.[25] Therefore, a fundamental principle of the OFA model is that access settlement balances in each settlement period.[26] To achieve that principle, a *firm* - rather than *fixed* - access design has been chosen for the OFA model.

2.3.6 Optionality

Another fundamental principle of the OFA model is its *optionality*: generators are entitled to choose the level of firm access that they wish to pay for. An alternative approach would have been to provide firm access as a mandatory service to all generators. That approach would be similar to the Generator Reliability Standards option described in the first interim report.

---

23 A pay-as-bid market is a market design where each dispatched generator is paid its offer price rather than a common market clearing price.

24 Although not of the "bid -$1,000" variety. Rather, each generator would rebid constantly to chase the LMP as it varied up and down.

25 Unless an uplift charge were levied on customers to fund any unrecoverable deficits.

26 Ie each 30-minute trading interval.

Optionality is considered an important feature, because it allows generators to reveal and signal the level of access firmness that they require, rather than this being decided for them by a regulator. It means that, in planning a new generation investment, investors have two degrees of freedom: the location and the level of firm access.

2.3.7 Fixed Access Charge

In the OFA model, the access charge is fixed for the life of an access agreement, similar to connection charges currently. An alternative approach would have been for annual access charges to vary according to an annual pricing methodology, similar to demand-side transmission use of system (TUOS) charges. Fixed charges have been chosen for two reasons: to give certainty to generators; and to give certainty to TNSPs.

The certainty provided to generators is obvious, the certainty to TNSPs, less so. If charges were to vary annually, generators might respond by procuring *shorter term* access: generators would be reluctant to sign up to a long-term agreement where prices could be changed annually in an uncertain and unforeseeable way. At the expiry of the shorter term access agreements, generators would only renew their firm access if, in the near term, access prices were low or congestion risks expected to be high. The resulting variations in the level of firm access – and access revenue – would create major risks for TNSPs and also for demand-side users who, under proposed TNSP regulation, would be required to cover any access revenue shortfalls.[27]

Thus, having fixed (or, at least, reasonably stable and foreseeable) access pricing is the preferred design option when there is *optionality* in firm access procurement. Conversely, where transmission pricing varies annually, it is usually in the context of a mandatory (not optional) transmission service: for example, in TUOS pricing in the NEM.

2.3.8 Free Non-firm Access Service

It may appear that non-firm generators are getting something for nothing: network access (albeit at an uncertain level) for no access payment. This might be contrasted with non-firm gas shippers (say) who still pay a transmission charge, albeit lower than firm shippers.

This view is not entirely correct. Non-firm generators implicitly pay for the cost of transmission losses – whether or not they obtain network access – and losses represent *most* of the variable costs of electricity transmission. Apart from losses, variable transmission operating costs[28] are small and difficult to identify and measure. The OFA model implicitly approximates these costs as being zero: an approximation which is unlikely to have any material impact on generator behaviour or market efficiency.

---

27 Discussed in section 8.2.5 (Financial quality incentives).

28 That is to say, the incremental operating costs associated with a line being loaded compared to it not being loaded.

2.3.9 Flowgate Support and Constrained-on Generators

In the OFA model, a flowgate support generator is allocated an entitlement equal to its usage.[29] This ensures that a generator – whether firm or non-firm – is never paid a price (in settlement overall) higher than the RRP, even if its LMP is higher than RRP.

This approach might appear asymmetric and unfair: a generator without network access is only paid LMP when it is *lower* than RRP, but is not paid LMP when it is *higher* than RRP. Additionally, a generator that is *constrained on* (ie dispatched despite its offer price exceeding the RRP) will rebid unavailable if it is paid only RRP, but would willingly be dispatched if paid the LMP (which, as always, can be no lower than the offer prices of dispatched generators).

There are several reasons for the approach taken in the OFA model:

- Flowgate support generators are not currently paid LMP and it is not clear that it would be appropriate to do so.
- Where flowgate support generators assist a TNSP in maintaining the FAS, a TNSP could enter into network support agreements with them, just as they do currently in relation to demand-side reliability standards.
- Generators with pricing influence might, if paid LMP, be able to cause very high LMPs. These high LMPs would distort outcomes and would create high risks for access-short generators at associated flowgates.
- Constrained-on situations are dealt with currently through directions and direction compensation and it is not clear that modification to these arrangements is necessary.

In many ways flowgate support is a mirror image of flowgate access. It would be possible to create an *optional firm support* model as a mirror image of the OFA model, in which the *TNSP* procures and pay for flowgate support from generators and *firm support generators* are required to make payments *into* settlements if their dispatch is less than their *agreed support level*. Such a model could efficiently address constrained-on problems in the NEM, just as the OFA model addresses constrained-off issues. However, the model would be complex to design and the cost of its implementation would likely be disproportionate to the problems it aims to solve. It is therefore not being proposed at this stage. It could potentially be developed and introduced at a later date.

2.3.10 Interconnectors

The NEM has two different types of interconnectors: regulated interconnectors and market network service providers (or *MNSPs*). The OFA model provides optional firm access for both types.

---

29 Which will be negative, since its participation factor is negative.

MNSPs use TNSP shared networks in the two regions that they interconnect. In the region that they draw power from (the *exporting region*) they are a demand-side user and so beyond the scope of the OFA model. In the region they deliver power into (the *importing region*), they are similar to a generator, in the sense that they inject power into the shared network at a specific node. Therefore, access provision to an MNSP in the importing region is exactly the same as for a generator. An MNSP would decide how much firm access to procure, using similar criteria to a generator.[30]

A different form of firm access, referred to as *inter-regional access,* is provided to regulated interconnectors. This is described and discussed in chapter 9 (Inter-regional access).

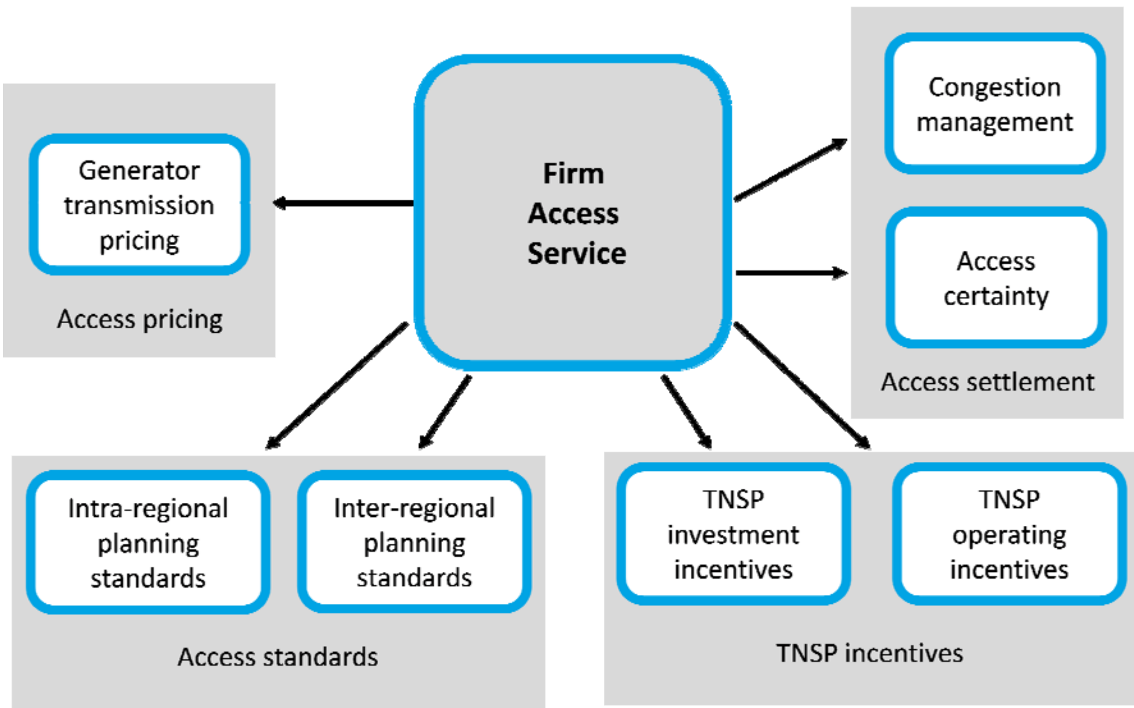
30 An MNSP can flow power in both directions and so there are potentially two importing regions in which it might seek firm access. However, the only MNSP in the NEM currently is Basslink, which connects to the RRN in Tasmania, and so would only require network access in Victoria.

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 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

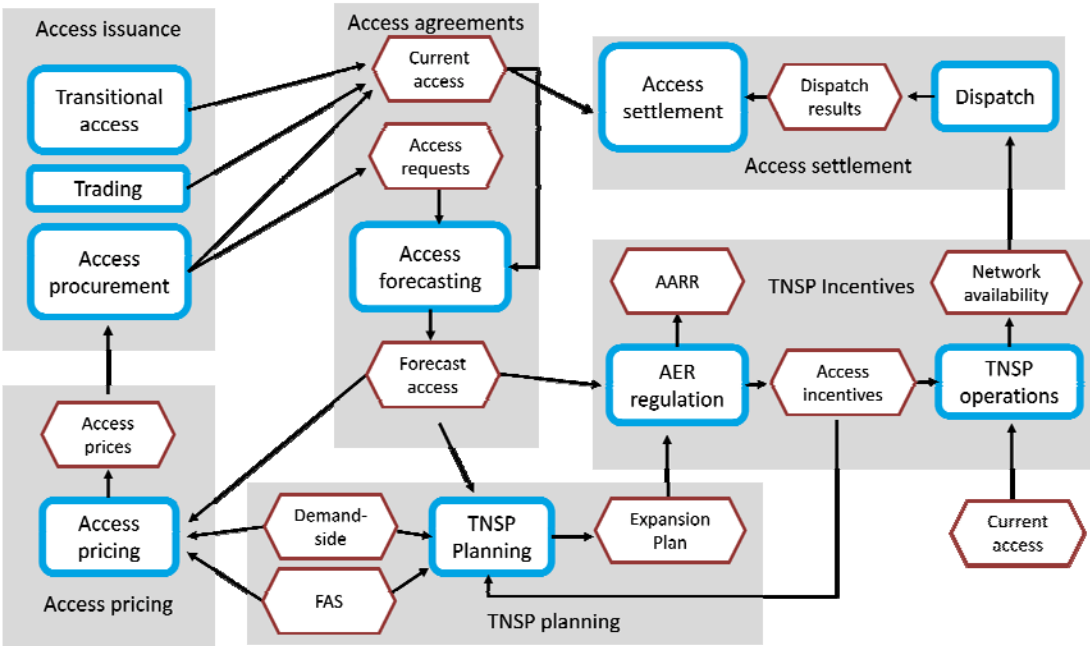


The OFA model does not address, and is not intended to address, transmission issues outside of this scope.

3.2 Model Architecture

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

Figure 3.2 Optional firm access 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 Section
Access Settlement	New	4
Firm Access Standard	New	5
Access Pricing	New	6
Access Procurement	New	7
TNSP Regulation	Augmented	8
Inter-regional Access	New	9
Transition	New	10

These processes are considered in turn in the sections below.

4 Access Settlement

4.1 Overview

Access settlement is the process which *effects* the de-linking of network access from dispatch, described in chapter 2 (Access). Network access is allocated to generators based on their agreed 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 network access cannot exceed flowgate capacities.

Existing AEMO settlement calculations and processes are unchanged.[1] 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 network 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 network access.

Access settlement occurs around congested *flowgates*: bottlenecks in the transmission network which are represented by binding transmission constraints in NEMDE. Typically, there are no more than a handful of congested flowgates in a region in any particular settlement period, so access settlement, whilst conceptually complex, should be straightforward for AEMO to implement.

A generator's *participation* in a flowgate is the proportion of its output that flows through the flowgate. Participation factors are currently calculated by AEMO for every generator and for every potentially congested flowgate, and appear as coefficients in the corresponding NEMDE constraint equation. A 1,000MW generator with a 10 per cent participation in a flowgate, say, would have just 100MW of its output flowing through the flowgate.[2] This 100MW is referred to as its flowgate *usage*.

Correspondingly, if that generator is to have 1,000MW of network access, it must have 100MW of access on that flowgate. In the OFA model, access on a flowgate is referred to as an entitlement. A *target entitlement* is the entitlement that would be required to provide a certain level of network access. However, to ensure settlement balances, total entitlements on a flowgate must equal the capacity of the flowgate and so not all target entitlements can be provided. An *entitlement scaling* process takes place in access settlement that gives priority allocation to firm generators so that their targets are met before any entitlements are allocated to non-firm generators.

For each generator at each congested flowgate, the access settlement payment is defined to be the difference between entitlement and usage, multiplied by the *flowgate*

1 With the exception of the calculation and allocation of the inter-regional settlements residue, discussed further in chapter 9 (Inter-regional access).

2 Notionally, of course. It is not possible to physically track generator output through a shared network.

price: the value of the corresponding NEMDE constraint. Generators whose entitlement exceeds their usage (typically constrained-off firm generators) will *receive payments* from access settlement. Generators whose usage exceeds their entitlement (typically dispatched non-firm generators) will *make payments into* access settlement.

4.2 Design Blueprint

4.2.1 Architecture

Recall from section 2.2.5 (Flowgates, Usage and Entitlements) the basic equations of flowgate settlement:

$$Pay\$_i = FGP \times (E_i - U_i) \tag{4.1}$$

$$\sum_i E_i = FGX \tag{4.2}$$

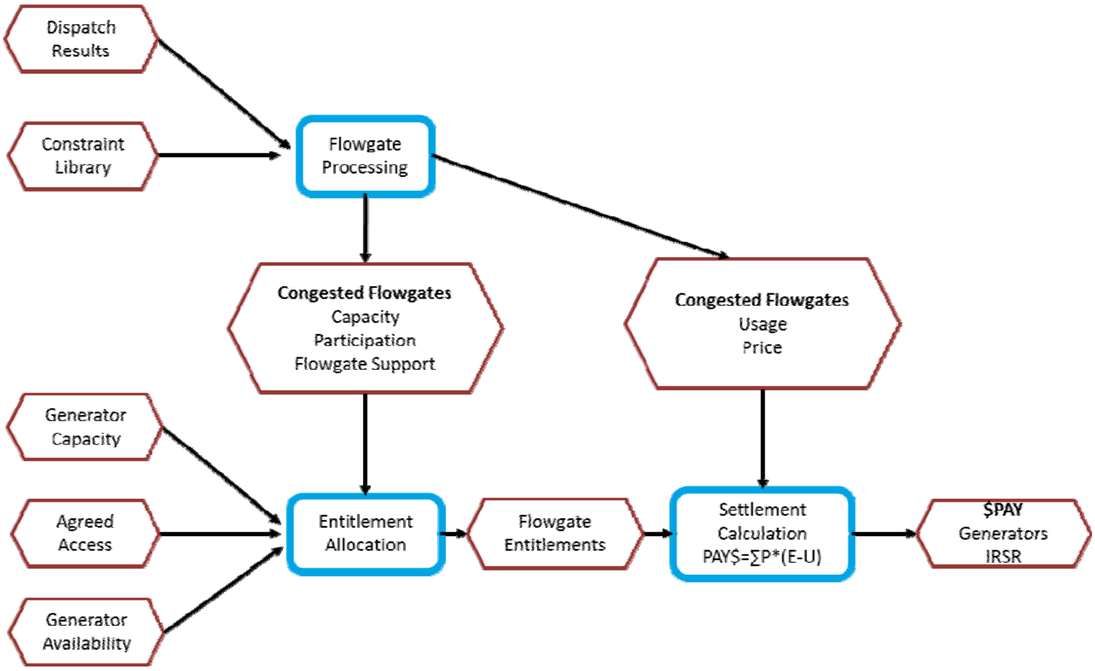
Access settlement calculates the amounts payable to or from each generator by applying these equations through three processes, presented in Table 4.1 below.

Table 4.1 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 4.1, below.

Figure 4.1 Access settlement processes



4.2.2 Flowgates

Although the term flowgate comes originally from gas transmission,[3] in the electricity context it means any bottleneck that potentially constrains dispatch and which is therefore represented by a transmission constraint in NEMDE.[4] The generic, linear form of transmission constraints[5] in NEMDE provide us with all of the parameters of a flowgate that we need for access settlement. Thus, a generic NEMDE constraint takes the form:[6]

$$\sum_i(\alpha_{ik} \times G_i) \leq RHS_k \qquad (4.3)$$

where:

i is the generator index and *k* is the flowgate index

3 In that context, it is where gas does literally flow through a gate: typically a point of connection between two gas pipelines at a point where gas flow is both commonly constricted and easily measured.

4 In the AEMC congestion management review, *constraint support price* and *constraint support contract* were terms used to describe flowgate prices and flowgate access, respectively.

5 A transmission constraint is informally defined as any constraint that arises as a result of limitations on TNSP networks and for which a constrained generator is not compensated under current arrangements. Transmission constraints may be associated with thermal and stability limits, even though the latter are not related to clearly-defined "bottlenecks" in a conventional sense. They may also, in some cases, include FCAS constraints. Flowgate types are discussed further in section 4.3.8 (What Constitutes a Flowgate?).

6 Although access settlement does not rely on the constraint being in that form; only that the constraint coefficients, α_i, are clearly defined. This is discussed further in section 4.3.8 (What Constitutes a Flowgate?).

Equation (4.3) provides the following quantities used in access settlement:

α_{ik} is the flowgate *participation factor* for generator *i* on flowgate *k*

$U_{ik} \equiv \alpha_{ik} \times G_i$ is the usage of flowgate *k* by generator *i*

RHS_k is the *flowgate capacity* for flowgate *k*

Thus, the generic constraint is re-framed in the OFA model as:

total usage ≤ flowgate capacity

4.2.3 Target Entitlements

Recall that the existing NEM design provides a generator with a level of network access equal to its dispatch level. If we wished to give generator the same level of access in the OFA model, we would need to ensure that its settlement payments are unchanged; that is to say, that its access settlement payments are zero. From equation (4.1) we can see that this is achieved by providing that generator with an entitlement on each flowgate:

$$E_{ik} = U_{ik}$$

Where:

E_{ik} = is the *entitlement* of generator *i* on flowgate *k*

Using the formula for usage, we have:

$$E_{ik} = \alpha_{ik} \times G_i$$

And, since the network access level A_i is then equal to G_i, we have:

$$E_{ik}=\alpha_{ik}\times A_i \qquad (4.4)$$

Thus, in general, to provide a generator with access level, A, it needs to be allocated entitlements on a flowgate equal to $\alpha \times A$ using the relevant participation factor, α.

Equation (4.4) is another fundamental building block for access settlements. It provides values for *target* entitlements: the entitlements that would need to be allocated to deliver a target level of access. These targets 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.

Access settlement calculates three access amounts for each generator, based on the agreed access level and the offered availability, as described in Table 4.2 below.

Table 4.2 Calculation of access amounts

Access Level	Formula
Firm access amount	Lower of agreed access and capacity
Non-firm access amount	Amount (if any) by which availability exceeds firm access amount
Super-firm access amount	Amount (if any) by which agreed access exceeds capacity

For each access amount, a corresponding target entitlement is determined for each flowgate, being the entitlement amounts that would need to be allocated to deliver that access amount, using equation (4.4). A numerical example demonstrating the calculation of target entitlements is presented in section 12.7.2 (Target Entitlements).

4.2.4 Entitlement Scaling

The aggregate of all target entitlements on a congested flowgate will always *exceed* the flowgate capacity[7]and so not *all* entitlement targets can be met. An entitlement scaling algorithm is used to determine *actual entitlements* from the scaling back of target entitlements, based on the principles that:

- total actual entitlements must equal flowgate capacity;
- a single *firm scaling factor* is applied to all firm and super-firm entitlements, and a single *non-firm scaling factor* is applied to all non-firm entitlements;
- firm entitlements are only scaled back when non-firm actual entitlements have already been scaled back to zero; and
- super-firm actual entitlements are only provided to the extent necessary to offset the scaling back of firm entitlements.

The detailed algebra for determining the scaling factors and entitlements, together with a numerical example, is presented in section 12.7.3 (Actual Entitlements).

Generators can informally be placed into one of four categories according to their relative levels of agreed access and availability, as presented in Table 4.3 below.[8]

7 The aggregate of the target entitlements is what the total flowgate usage would be if all of the generators were dispatched at full output. If flowgate capacity exceeded this level the flowgate could not possibly be congested.

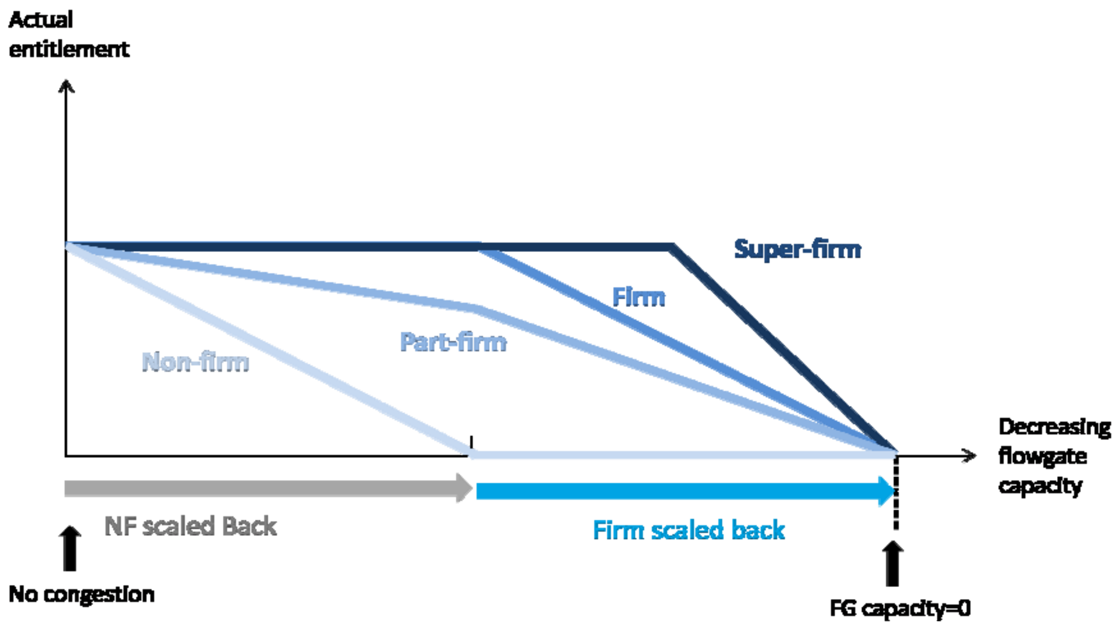
8 The categories are for illustration only and are not considered explicitly in the access settlement algebra.

Table 4.3 Generator access categories

Generator Type	Description
Super-firm generator	agreed access > capacity
Firm generator	agreed access = capacity
Part-firm generator	agreed access < capacity
Non-firm generator	agreed access = 0

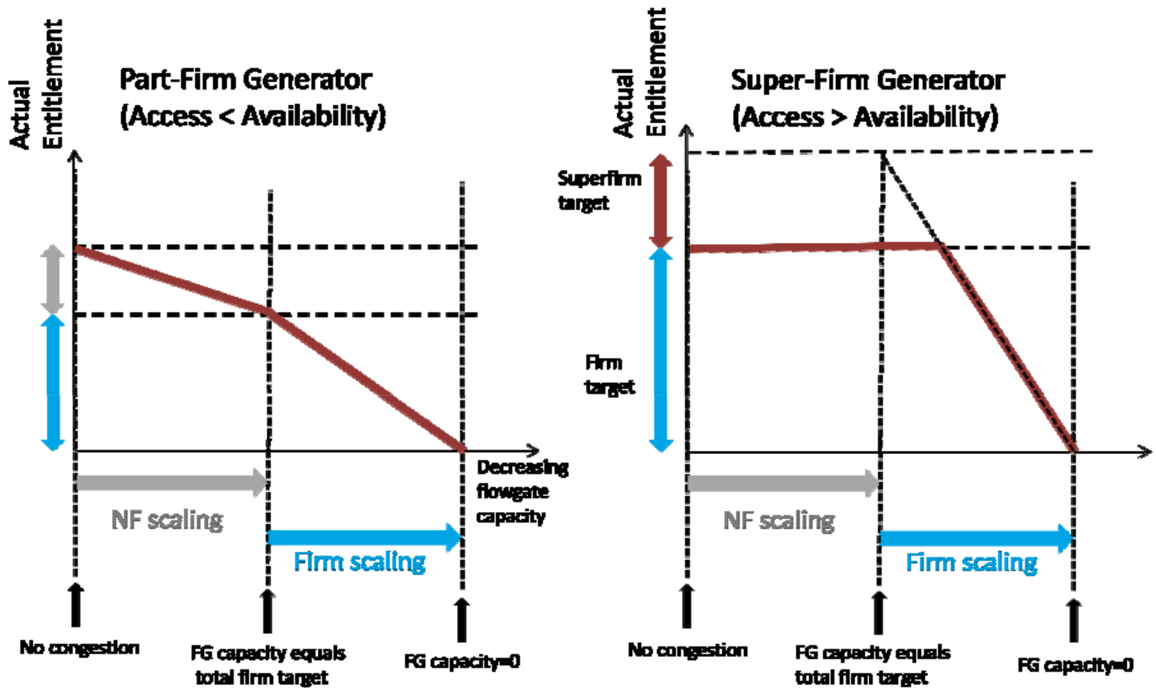
Figure 4.2, below, illustrates the level of entitlements that would be allocated to generators in these four access categories under decreasing levels of flowgate capacity. For simplicity, these generators are assumed to have identical capacities and participation factors and to be fully available.

Figure 4.2 Entitlement scaling for four access categories



Using similar assumptions, Figure 4.3 shows actual entitlements relative to target entitlements, for a super-firm and a part-firm generator. It will be seen that non-firm targets are only taken into account once firm targets have been fully met.

Figure 4.3 Actual entitlements compared to target entitlements



4.2.5 Flowgate Support Generation

In the description above, it has been implicitly assumed that all participation factors in a flowgate are *positive*. In fact, participation factors can be, and commonly are, *negative*. A generator with a negative participation factor is referred to in the OFA model as a *flowgate support* generator with respect to that particular flowgate. It will be noticed, from the formula for usage, that a dispatched flowgate support generator has *negative* usage, which means its dispatch *relieves* congestion on the flowgate (hence its name).

Flowgate support generators are always provided with (negative) actual entitlements equal to their (negative) flowgate usage.[9] The flowgate *support* amount (a positive number) is the total absolute level of entitlements allocated to flowgate support generators.

Since total actual entitlements must always equal flowgate capacity, flowgate support generators increase the *effective* level of flowgate capacity that is allocated through the entitlement scaling algorithm. For example, if flowgate capacity is 1,000MW and flowgate support is 100MW, the entitlement scaling algorithm allocates 1,100MW between the remaining generators.

9 This is to ensure that they have no exposure to access settlement and thus are simply paid, as now, the RRP on their dispatched output. Refer to section 2.3.9 (Flowgate Support and Constrained-on Generators) for an explanation of this design decision.

4.2.6 Flowgate Pricing and Settlement Balancing

Access settlement is based on the difference between entitlement and usage on a flowgate, multiplied by the *flowgate price*. The flowgate price is defined to be equal to the value (or shadow price) of the corresponding transmission constraint in the settlement period, as calculated by NEMDE.[10] Thus, for each generator and each congested flowgate:

Payment$ = Flowgate Price x (Actual Flowgate Entitlement – Flowgate Usage)

The sum of all settlement payments for a flowgate is therefore:

Total Payment$ = flowgate price x (Total Flowgate Entitlement – Total Flowgate Usage)

Since the flowgate is congested, *total flowgate usage equals flowgate capacity*. Furthermore, the entitlement scaling algorithm ensures that *total flowgate entitlement equals flowgate capacity*. Thus *total entitlement equals total usage* and so the total *payment$* is zero. Access settlement balances for each flowgate in each settlement period and, therefore, always balances in aggregate.

Section 12.2.10 (Generator Access Settlement) describes how access payments based on this settlement algebra ensures that generators receive the proper levels of network access and dispatch access, as defined in chapter 2 (Access).

4.2.7 Settlement Period

The settlement period is a trading interval (30 minute period), the same as for existing NEM settlement processes. However, many of the dispatch variables – such as flowgate prices, usages and capacities – are calculated by NEMDE each dispatch interval (5 minute period). These quantities are converted in access settlement to 30 minute equivalents, generally through simple arithmetic averaging. The approach taken is discussed in section 12.4 (Thirty-Minute Settlement).

4.3 Design Issues and Options

4.3.1 Access Grouping

The Technical Report (August 2012) described a mechanism by which generators could form an access group through which agreed access could be pooled. This would be advantageous to generators with relatively low availability (ie intermittent generators) whose availability is weakly correlated with fellow group members. This was in the context of the then design proposal that access would be limited to be no higher than availability. However, it is now proposed that access is limited only by generator

---

10 The economic meaning and relevance of flowgate prices is discussed in section 12.2.5 (Flowgate Prices).

capacity. In this context, grouping provides limited benefit to generators.[11] Therefore, the grouping mechanism no longer forms part of the proposed OFA design blueprint.

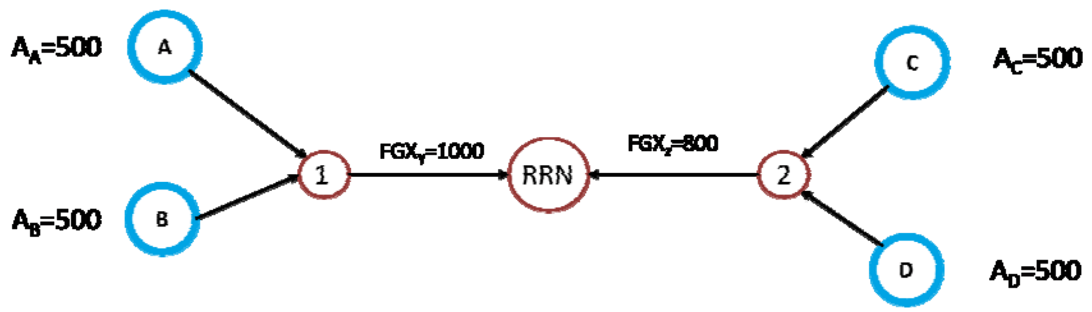
4.3.2 Use of Flowgate Prices rather than Nodal Prices

Chapter 2 (Access) describes access settlement in terms of the difference between RRP and LMP. However, LMPs are not mentioned at all in the settlement blueprint and, as discussed in section 12.2 (Flowgate Pricing and Local Pricing), it is complex to demonstrate that the use of FGPs and LMPs can be financially equivalent. Which raises the question: why not just use LMPs?

The reason for using flowgate pricing lies in the entitlement scaling process. The regional optional firm access model presented in the first interim report (ie package 4) used LMPs and then scaled back access settlements *pro rata* across a region to ensure settlement balance. The current OFA design is considered to be a substantial improvement on that previous design, for reasons discussed below.

Consider Figure 4.4 below. There are two congested flowgates: flowgate Y between node 1 and the RRN; and flowgate Z between node 2 and the RRN. The capacity of flowgate Y is sufficient to accommodate the firm generation connected to node 1 and using that flowgate.[12] However, the capacity of flowgate Z is insufficient to provide firm access levels to firm generators connected at node 2.

Figure 4.4 Example: Three-node radial network



Scaling of entitlements 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 scaling approach would have required effective access levels of *all* firm generators to be scaled back.

These characteristics make the flowgate approach not only fairer but also more efficient and transparent. It is more efficient because, in deciding on location and firm access levels, generators will only take account of flowgate capacity and firmness on

11 Because intermittency no longer affects access and so intermittent generators no longer have super-firm access that they can advantageously pool with other intermittent generators.

12 Nevertheless, some additional non-firm generators connected to node 1 (not shown in the figure) could cause flowgate X to be congested.

flowgates affecting their location. Access decisions should not be – and will not be – affected by congestion in other parts of the region.[13]

It is more transparent, because access settlement occurs only on congested flowgates. These will – in any particular settlement period – typically be few in number, and only a subset 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. It would much harder to monitor every single flowgate in the region.

4.3.3 Target Access not based on Preferred Output

A generator is constrained-off if its dispatched output is below its *preferred output*: the total MW amount that is offered at a price below RRP. Since the objective of access settlement is to compensate a firm generator that is constrained-off, it would be natural to set the target access level equal to the preferred output.

In the second interim report, it was proposed instead to base access on *availability*. It is now proposed to base firm access on generator *capacity*, but still to base non-firm access on availability. The reasons for rejecting the use of preferred output are described in this section. The reasons for deciding to use capacity and availability for firm and non-firm access, respectively are set out in the following two sections.

Preferred output is the level at which a generator *would* be dispatched (according to its dispatch offer) in the absence of congestion.[14] Preferred output represents the level at which a generator is seeking access. For example, when the RRP is low, a peaking firm generator is unlikely to wish to be dispatched (ie its preferred output is zero) and so is unlikely to require access.

The practical difficulty with preferred output is that it is dependent on the dispatch offer and so is easily manipulated by rebidding. If access were to be based on preferred output, the peaking firm generator mentioned could simply rebid to make it *appear* as though its preferred output was full output.[15] Therefore, basing access on preferred output would simply encourage rebidding and would give similar outcomes to basing access on availability. For these reasons, the use of preferred output has been rejected.

4.3.4 Target Firm Access Limited by Capacity

For the reasons discussed in the previous section, availability is the best *practical* proxy for preferred output. Thus, using availability keeps faith with the original concept of "firm access" that a generator would only be compensated for congestion if it was genuinely constrained off.

---

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

14 Assuming no change in the RRP.

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

However, there are some concerns with using availability rather than capacity to limit firm access. The major concern is that this may be seen as unfavourable for intermittent generators, who would pay the same access charge as conventional generators (for the same access amount), but would receive a much lower financial benefit, in terms of payments from access settlement. A subsidiary concern is that the use of availability makes the access service specific to a particular generator and so harder to trade between generators. For these reasons, capacity-based limiting is preferred.

This raises the question of whether access should be limited at all. If a 500MW generator chooses to be super-firm and purchase 800MW of access (say), should it not receive access settlement payments in relation to the full 800MW?

This would represent a significant change to the philosophy of the OFA model, from simply compensating a generator that is constrained off, to providing a purely financial transmission right that could potentially be held by anyone, even a non-generator. It might also exacerbate distortions to dispatch where a super-firm generator has local pricing influence.

On the other hand, limiting access creates a concern that – in the context of a single-tier FAS – super-firm generators might not actually obtain firmer access outside of normal operating conditions.[16] Nevertheless, the advantages of capacity-based limiting are considered to outweigh the disadvantages.

4.3.5 Unusual Constraint Formulations

Those familiar with NEMDE constraints will point out that these rarely have the simple form expressed in equation (4.3). For example, AEMO often uses feedback constraints, taking the form:.

> change in usage ≤ spare flowgate capacity in prior period + change in flowgate capacity

In a feedback constraint, and other non-standard constraint formulations, the RHS of the constraint does not represent flowgate capacity.

To ensure that access settlement correctly extracts the information it requires from NEMDE constraints, no matter what their form, the following approach is taken:

- participation factors are taken from the coefficients that are applied to generator dispatch variables on the LHS of the NEMDE constraint; and
- flowgate capacity is calculated from aggregate flowgate usage.

Coefficients must exist in all NEMDE constraint forms, because they are necessary for NEMDE to operate. Therefore, usages on a flowgate can always be calculated and aggregated. For a congested flowgate, flowgate capacity by definition equals total flowgate usage. It is not necessary to explicitly calculate the constraint RHS.

---

16 This issue is discussed further in section 5.3.2 (Single Tier FAS).

4.3.6 Target Non-firm Access based on Availability

It is proposed that target non-firm access is based on *availability*. This is the same approach as in the Technical Report (August 2012), but may appear somewhat inconsistent with the approach proposed for target firm access, which is now based on *capacity* rather than availability.

For a particular generator, target non-firm access will only be provided if target firm access is *less* than availability. In this situation; target firm access would *not* be limited anyway, either by an availability-based or a capacity-based approach. Thus the change to the target firm access approach has *no* consequential impact on the level of target non-firm access. The perceived inconsistency between the two approaches does not, therefore, create any issues for access settlement.

Nevertheless, it is worth considering whether the arguments, presented in the previous section, that prompted a change to the firm access approach would also apply in the context of non-firm access. In fact, they do not apply, because non-firm access is neither paid for, nor likely to be traded.

On the other hand, there are some good reasons for continuing to base non-firm access on availability:

- it is the best available proxy for preferred output, which is a fair basis for allocating non-firm access;
- it is similar to the *de facto* allocation of access under the current arrangements during disorderly bidding, assuming that generators have identical participation factors; and
- it is consistent with the Shared Access Congestion Pricing (SACP) model, which has generally been considered a fair way to allocate access in the absence of firm access rights.

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.

4.3.7 Firm Generator may be liable to pay into Access Settlement

In the original design of the OFA model, presented in the first interim report, a firm generator would *never* make payments *into* access settlement: it was either out-of-merit, dispatched or constrained-off. In these three cases it would receive nothing, RRP and compensation, respectively. That model implied that a firm generator would always receive agreed access if dispatched, even if other, constrained-off, firm generators had their access scaled back. Similarly, a non-firm generator could only receive a level of access that was at or below its dispatch level. In this design, then, there was some residual linkage between access and dispatch, creating the potential for some disorderly bidding to continue.

In the proposed OFA model, which in this respect remains unchanged from the second interim report, access and dispatch are *totally* de-linked. This creates the possibility of two counter-intuitive situations which could not have arisen in the original OFA model:

- a *firm* generator making payment into access settlement; and
- a *non-firm* generator receiving payment from access settlement.

The first situation could arise where a firm generator is fully dispatched and (coincidentally) has its entitlements scaled back. For example, suppose a 1,000MW generator procures 1,000MW of agreed access. In a settlement period, it is fully dispatched to 1,000MW but has its firm access scaled back to 800MW. It is access-short by 200MW and must pay *into* access settlement.

The second situation would arise where a non-firm generator is not fully dispatched, but is (coincidentally) allocated some (non-firm) entitlement. For example, a 500MW generator is dispatched to 100MW and receives 150MW of entitlement. It is access-long and so is paid *from* access settlement.

These situations will be relatively uncommon but can (and should) nevertheless occur from time to time. In short, a firm generator does not get fixed access; and a non-firm generator does not get *zero* access. Each gets different degrees of access firmness and these are unrelated to dispatch levels.

4.3.8 What Constitutes a Flowgate?

In the OFA model, every binding *transmission constraint* in NEMDE creates a corresponding flowgate in access settlement. But how exactly is a *transmission constraint* distinguished from other, non-transmission, constraints in NEMDE? An informal definition of transmission constraint is that it relates to any *constraint that arises as a result of limitations on TNSP networks and for which a constrained generator is not compensated*[17] *under current arrangements*.

AEMO lists three broad categories of constraint:[18]

- network;
- frequency standards; and
- other.

---

17 Where a generator is already compensated for being constrained-off, for example in relation to NCAS provision, the OFA model should avoid duplicating this compensation.

18 AEMO, *Constraint Formulation Guidelines*, 6 July 2010.

Network constraints cover three constraint types: thermal constraints, stability constraints[19] and network control schemes.[20] The former two types would certainly be treated as flowgates. The third type would probably not be so treated, as generators are currently paid for providing Network Control Ancillary Services (NCAS), although there may be some circumstances where they are not compensated for being constrained and these might give rise to the need for flowgates in the OFA model.

The term *flowgate* is borrowed from the gas industry and, as such, is a better description of thermal constraints than stability constraints, in that thermal constraints always apply to a *particular* transmission element (line or transformer) which must not be overloaded.[21] Therefore, the flowgate can easily be considered to be *physically located* on that element. Stability constraints are more nebulous.[22] They are typically modelled as limiting the aggregate power flow over a cutset[23]of the transmission network: usually, but not necessarily a regional boundary. They are not easily locatable: in particular, it is difficult to say which region the flowgate lies within. This does not matter for access settlement but is pertinent in relation to TNSP incentives, discussed in section 8.2.5 (Financial Quality Incentives).

Stability constraints are also complex in that the flowgate capacity is dependent upon the location, technology and connection equipment of synchronised generation. This makes it harder for TNSPs to predict and manage flowgate capacity. However, this is similarly not relevant to access settlement, but is pertinent to TNSP planning and operations.

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.[24] 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.

---

19 Covering voltage stability, transient stability and oscillatory stability.

20 Described by AEMO as the modelling of generator control schemes or reactive control devices on generator output.

21 Although, of course, gas pipelines do not overheat, but their transmission capacity is limited for other reasons.

22 They don't have any equivalent on gas pipelines.

23 A cutset is a set of transmission elements that, were they removed from service, would cause the transmission network to be separated into two parts. Informally, if one imagines cutting a map of the transmission network into two with a pair of scissors, the cutset contains all of the transmission elements that have been chopped in two.

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

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 takes the flow limit (and associated flowgate capacity) at face value[25] and applies the access settlement algebra accordingly.

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 flowgates: ie those relating to managing negative residues,[26] interconnector rate of change limits and discretionary limits;
- unit zero constraints might formally be considered flowgates, where they relate to network limitations; however, if flowgate capacity is zero then, by definition, there is zero flowgate usage, zero flowgate entitlements and therefore zero access settlement, except to the extent that there may be payment by a TNSP under an incentive scheme;[27]
- network support agreements generally impact only on flowgate support generators and so would not need to be treated as flowgates in access settlement;[28]
- constraints relating solely to generator limitations or non-conformance are not flowgates; and

25 Recalling that flowgate capacity on binding constraints is calculated based on aggregate usage, as discussed in section 4.3.5 (Unusual Constraint Formulations).

26 Although these are unlikely still to be required under an OFA regime.

27 Discussed in section 8.2.5 (Financial quality incentives).

28 Recall from section 2.2.9 (Flowgate Support) that support generators receive zero payments from access settlement, so settling flowgates in which only support generators participate is unnecessary.

- discretionary limits on generators may be treated as flowgate where they arise as a result of network limitations.

AEMO would be required to *tag* the NEMDE constraints that create flowgates and this tagging would be used to identify the binding flowgates required to be processed in access settlement.

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 ingredients are required to provide generators with confidence that service quality will be maintained: a *service standard* that specifies the minimum service quality that must be provided to each user; and a corresponding *network standard* that specifies the minimum level of transmission capacity that the TNSP must build and maintain to provide the minimum service quality to *all users*. The *Firm Access Standard* (FAS) performs both of these roles. A generator can obtain a service with a higher or lower effective firmness than this standard by procuring an *agreed access amount* that is higher or lower, respectively, than its generating capacity.

The agreed access amount specified in each access agreement is a nominal amount and is required to be provided under normal operating conditions (NOC) specified in the FAS during the period of the agreement. Under abnormal operating conditions (AOC), no minimum access level is specified by the FAS.

In planning and operating its network, a TNSP must ensure that it can provide the FAS-defined level of service to every firm generator concurrently: since it is possible that every generator will require access at the same time. Thus, the FAS – in combination with the set of all access agreements – defines a *network standard*: a minimum level of transmission capacity that must be provided under specified *normal operating conditions*. A TNSP must also ensure that it continues to maintain existing *demand-side reliability standards*, which still apply alongside the OFA model.

5.2 Design Blueprint

5.2.1 The Role of the FAS

The FAS provides the nexus between access agreements and other transmission processes such as network planning and operations, access pricing, and TNSP incentive regulation. A TNSP must ensure that, in real-time, it always has sufficient *available* transmission capacity to provide at least the minimum level of access that the FAS specifies. That obligation drives operational decisions and also, through the TNSP forecasting future access demand, drives planning decisions.

5.2.2 FAS Definition

The FAS defines a set of normal operating conditions (NOC) and requires that all firm generators are provided with their agreed access level on *congested* flowgates during these conditions. It is anticipated that NOC will include:

- *system normal:* all transmission elements are in service; and

- *planned outages:* some transmission elements are out of service due to planned maintenance

Note that, although target firm access is limited by generator capacity,[1] the FAS requirement is not: the flowgate capacity must be able to provide the full level of agreed access, which may exceed generator capacity. This is a design change from the Technical Report (August 2012) and is discussed further in section 4.3.4 (Target Firm Access Limited by Capacity).

Note also that FAS is an *operational* standard, which means that the required flowgate capacity must be provided under *secure* dispatch. AEMO's N-1 security standard means, then, that a system normal obligation is analogous to an N-1 planning standard.

Planned outages are included within NOC to ensure that TNSP outage schedules allow access levels to be maintained. This is likely to mean scheduling outages when congestion (of the depleted flowgate) is unlikely: ie during off-peak periods. However, it is recognised that defining and monitoring planned outages – and appropriately distinguishing them from *un*planned outages - may be difficult. The desirability of including them in NOC will be decided during OFA implementation.

In defining the NOC set it is important that:

- it is *clearly defined*, such that the NOC and AOC can be unambiguously distinguished within settlement timescales;[2]

- it does not encourage *perverse* TNSP behaviour: for example, deliberately taking a line out of service so that its FAS obligation is reduced; and

- it is *relevant* to generators: for example, if generators are most concerned about congestion during planned outages, these should be ideally be covered by NOC.

Section 12.8 (TNSP Planning and Operations under the Firm Access Standard) discusses in more detail how a TNSP might monitor and manage its FAS obligations.

5.2.3 FAS Implications for Flowgate Capacity

As discussed in section 4.2.3 (Target Entitlements), for a generator to receive some level of access, *A*, it must be provided with an entitlement on each flowgate equal to the product of the participation factor and that access level:

$$E_i = \alpha_i \times A_i$$

Where:

α_i is the participation factor of generator i

1 Chapter 4 (Access Settlement).

2 This is to allow TNSP incentive payments to be cleared through AEMO settlement, discussed further in section 8.2.5 (Financial quality incentives).

A TNSP must provide sufficient effective flowgate capacity to provide the FAS access obligation to all firm generators *concurrently*:

$$\text{Effective Flowgate Capacity} \geq \Sigma\, \alpha_i \times A_i \qquad (5.1)$$

The RHS side of this inequality is referred to as the *target flowgate capacity*. Thus

$$\text{Target Flowgate capacity} = \Sigma\, \alpha_i \times A_i \qquad (5.2)$$

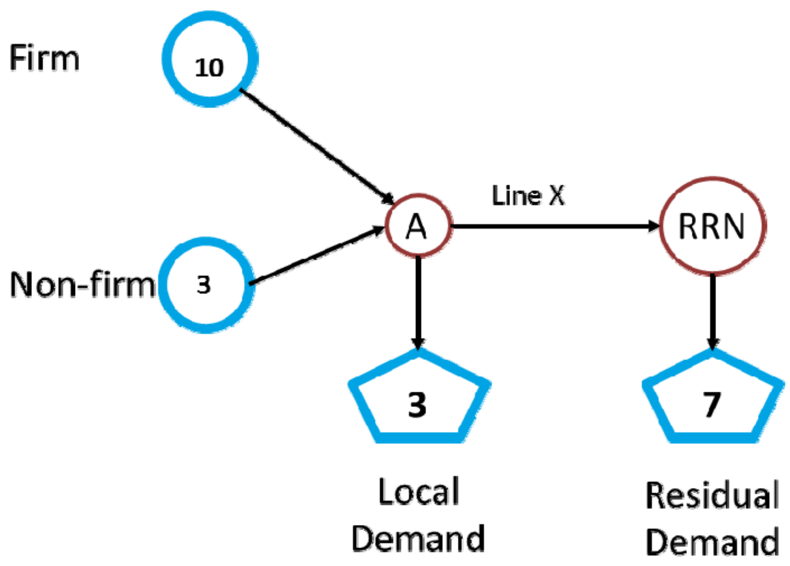
The RHS of equation 5.2 is the total usage of the flowgate by firm generators dispatched at their agreed access level. Thus, a sufficient condition for maintaining FAS is for a TNSP to ensure that, through a combination of transmission capacity and network support agreements with flowgate support generators, the effective flowgate capacity is sufficient to allow all firm generators to be dispatched simultaneously at their agreed access level.

Flowgate capacity takes account of *local demand* (ie demand not located at the RRN). *Local demand* close to a generator causes an increase in the capacity of flowgates used by that generator. Thus, being able to dispatch all firm generators simultaneously effectively means that the following load flow is feasible (does not breach transmission constraints) in which:

- all firm generators are dispatched at their agreed access levels;
- all local demand is supplied, based on expected demand levels for the particular study; and
- residual demand (ie total firm generation minus total local demand) is supplied at the RRN.

Figure 5.1 illustrates this load flow for a simple two node network.

Figure 5.1 Simple two-node network



If aggregate firm generation *exceeds* aggregate forecast demand (including demand at the RRN) then the residual demand at the RRN will exceed the forecast demand there. Thus, the target flow capacity required under the sufficient FAS condition described above might never be fully utilised. For example, to be feasible, the load flow in Figure 5.1, above, requires line X capacity to be at least 7GW.[3] However, if forecast demand at the RRN is only 6GW, then the line flow is only 6GW. It is only *necessary* for the TNSP to provide sufficient flowgate capacity so that the flowgate is never congested. In short:

> Flowgate Capacity Required = min(target flowgate capacity, maximum actual flowgate usage)

By definition, super-firm generators[4] cannot be dispatched at their agreed access level. Where super-firm generators are present, flowgate usage might therefore never reach target flowgate capacity, allowing TNSPs to provide a below-target level of flowgate capacity without breaching FAS. This is discussed further in section 5.3.3 (Super-firm Access under a Single-tier FAS).

Alternatively, if aggregate firm generation is less than aggregate forecast demand, then the residual demand placed at the RRN in the study is *negative*: ie it is notional *generation*. Thus, for the purposes of FAS, the TNSP is permitted to rely on a notional generator being dispatched at the RRN. Since this generator does not really exist, the amount of transmission capacity developed pursuant to FAS may be insufficient to ensure that demand can be reliably supplied. In the example above, if the forecast demand at the RRN were 8GW, line X capacity would need to be at least 8GW.[5] This issue is discussed further in section 5.3.7 (Demand-side Reliability Standards will continue).

Note also that, in the case of super-firm generators, the level of agreed access exceeds the amount of generation that can be available to supply demand. Therefore, it is not even sufficient, for demand-side reliability, that aggregate *agreed* access covers peak demand.

5.2.4 FAS is flowgate specific

It is natural to think of a transmission operating condition applying to the network as a whole. For example, the condition *system normal* is typically interpreted to mean that every transmission element in the network is in service; a *single* transmission outage then means that the network is no longer system normal. Under this interpretation, at any point in time, *every flowgate in a region* is either under NOC (and thus subject to target flowgate capacity) or under AOC.

However, because access settlement is flowgate specific, the NOC definition can and should be applied at the flowgate level. This can be done by requiring AEMO to tag

---

3 Giving the target flowgate capacity of 10GW, after taking account of local demand at node Z.

4 Those with agreed access higher than their generation capacity.

5 Permitting 1GW of non-firm generation to be dispatched to meet the demand shortfall.

each NEMDE transmission constraint as either an NOC or AOC flowgate, as discussed below.

Each NEMDE constraint prepared by AEMO relates to a particular transmission condition. For example, AEMO may prepare several constraints preventing contingent overloads on a line X: a constraint XN for system normal conditions, XY for when line Y is on a planned outage, XZ for when line Z has a forced outage and so on. In the above example, AEMO would label XN and XY as NOC and XZ as AOC.

AEMO arranges for prepared constraints to be *applied* in NEMDE only when the prevailing transmission conditions correspond to the conditions assumed when they were formulated. In a system normal, constraint XN would be applied; if line Y was on outage, constraint XY would be applied; and so on. If, during a system normal situation, a forced outage occurred in a zone *remote* from line X, AEMO would leave constraint XN in NEMDE since the remote outage would not affect contingent flows on line X. In other words, although the *network-as-a-whole* is no longer system normal, in the sense of *everything* being in service, the *system normal constraint XN* continues to apply.

If flowgate XN became congested in these conditions, the label on that would indicate that it was an NOC1 constraint and so the FAS target for flowgate capacity would still apply. The remote outage does not physically affect the capacity of flowgate XN and so it should not affect the FAS obligation, despite the fact that the *network-as-a-whole* is no longer, in the usual sense of the term, system normal.

In general, the FAS obligation applying to a flowgate is predicated on whether that *particular* flowgate is designed for an NOC or AOC condition and *not* to the condition of the transmission network overall.

5.2.5 Summary

The FAS defines the minimum level of firm access service quality that a firm generator is entitled to and, consequently, drives TNSP network planning and operation. The FAS obligation only applies under normal operating conditions. There is no minimum requirement under abnormal operating conditions.

5.3 Design Issues and Options

5.3.1 Firmness of FAS

Because it applies only under NOC, the FAS describes an access standard that is *firm* but not *fixed*. Why was this design choice made?

A fixed access FAS would mean a guaranteed agreed access level in *all* conditions. For settlement to balance, that means, in turn, a *fixed* target flowgate capacity. Achieving this is impractical: there is always the possibility of extreme conditions (multiple outages, extreme weather events etc) where a minimum level of transmission capacity cannot be maintained.

The use of the word "firm" comes from the gas industry, where *firm* transportation (as opposed to *non-firm* transportation) is provided. Firm is not *fixed* for gas transportation either. Events such as compressor failure will lead to reduced service.

Rather than over-engineer the transmission system to attempt to provide a level of firmness that it cannot inherently provide, it is preferable to define the FAS that can be economically be provided.

Of course, just because there is a zero FAS obligation under AOC, this does not mean that flowgate capacity will actually be zero. A TNSP will develop and maintain sufficient transmission capacity to meet FAS under NOC. A forced outage, or other AOC-triggering event, may well reduce flowgate capacity but – except under extreme circumstances – will not reduce flowgate capacity to zero.

Ideally, and notwithstanding that the FAS permits a lower level, access firmness should be maintained at an efficient level by TNSPs: ie to the point where the benefits of additional firmness no longer cover the costs of providing it. In particular, TNSPs should limit forced outage rates to efficient levels through appropriate design and maintenance. It is recognised that the FAS itself does not provide the incentives on TNSPs to do this and so these would need to be provided under separate performance incentive schemes, as discussed in section 8.2.5 (Financial quality incentives).

5.3.2 Single Tier FAS

In the second interim report, a multi-tier FAS design was proposed. This design requires definitions for multiple tiers of normal operating conditions (NOC1, NOC2, NOC3,…etc) and FAS scaling factors (between 0 and 1) for each tier. A single tier FAS is now proposed, as described in section 5.2.2 (FAS Definition). This is a significant and important change in the OFA design. The reasons for making this change are presented below.

The philosophy of the OFA model is that generators are able to choose their preferred level of access and, having made that choice, have substantially more certainty about the level of access that they will obtain than under the current arrangements. The access decisions of generators will then drive transmission development, to ensure that an efficient amount of transmission is provided, in terms of the preferences of generators.

The multi-tier FAS proposed in the second interim report was intended to align with this philosophy. The inclusion of lower tiers (NOC2, NOC3 etc) was intended to provide generators with greater certainty of access outside of NOC1. Ideally, the NOC tiers would cover the vast majority of operating conditions, with AOC just covering a residual, and fairly extreme, remnant.

A number of concerns about a multi-tier FAS were raised by stakeholders and have been considered carefully. These concerns can be summarised in the following questions:

- Is a multi-tier FAS practical?
- Does it really provide generators with more certainty?
- Does it really provide generators with more choice?

The practicality question covers concerns around designing, managing and monitoring a multi-tier FAS. The design difficulties would involve defining the various NOC tiers and deciding on appropriate FAS scaling factors. Simplicity and transparency considerations would suggest that definitions should be fairly generic (eg N-1, N-2) etc, but network characteristics mean that, in different situations, very different FAS scaling factors should then apply. For example, a single forced outage might reduce capacity by 20 per cent at one flowgate location but by 50 per cent at another location, depending upon the local topology. Setting a FAS scaling factor based on the "lowest common denominator" would make the FAS very weak at most locations, giving it a very limited role in creating TNSP incentives. On the other hand, setting a more aggressive scaling factor would make it very hard or expensive to maintain at some locations. This could be addressed by making NOC tier definitions location – or topology – specific, but this would add complexity.

For a TNSP to *manage* a multi-tier FAS, it would need to study and monitor all NOC tier conditions. This may become combinatorially complex for lower FAS tiers: eg there are a vast number of different possible network conditions under N-3. TNSP processes such as the RIT-T would become extremely complex and lose transparency as a result, to the detriment to the efficiency and appropriateness of the decisions made under these processes.

The FAS will only be effective if it can be properly *monitored* and enforced by the AER. That may be impractical for a multi-tier FAS.

So does a multi-tier FAS actually provide generators with more *certainty*? The main objection here is that FAS defines access obligations under each NOC tier, but says nothing about the timing or duration of each tier. A TNSP has no obligation to maximise the duration of the higher NOC tiers (eg tiers 1 and 2) and a clear incentive to do just the opposite, in order to reduce target flowgate capacities. Furthermore, most generators are unlikely to have the information or expertise needed to estimate NOC tier duration and timing when making location or access procurement decisions. The question of certainty is very much "in the eye of the beholder" and an uninformed generator cannot practically be provided with more certainty by a multi-tier FAS.

If not certainty, does a multi-tier FAS provide generators with more *choice*? The design of access procurement and access settlement means that the answer is, unfortunately, no. A generator could only choose its *nominal* level of access, which it would be provided with under NOC1. The levels of access provided under lower FAS tiers would then be defined by the FAS scaling factors. A generator could not obtain a higher level of access for just NOC2, say.[6]

Under a multi-tier FAS, this lack of choice could be detrimental to a generator. For example, a generator procuring 100MW of access would, assuming a FAS scaling factor of 80%, say, be provided with 80MW of access under NOC2. Provision of this level of NOC2 access may be quite expensive (as estimated by the access pricing methodology) and yet not highly valued by the generator. However, the generator can only avoid this cost by procuring a lower *nominal* level of access.

That is assuming that an access pricing methodology can actually be developed that is able to estimate the cost of providing access under the lower NOC tiers. This may be challenging; indeed, as discussed in section 6.2.2 (Medium-term and Long-term Forecasting), there is a need to further simplify access pricing, and this objective is incompatible with accurate costing of lower NOC tiers. In practice, only the NOC1 would likely be costed. In this case, the high cost of NOC2 in the example would not be borne by the generator, but by the TNSP and, ultimately, the demand-side user.

In the absence of effective choice and of costing of the lower NOC tiers, a multi-tier FAS does nothing to improve the efficiency of network provision and may do the opposite. By rigorously adhering to deterministic FAS obligations, a TNSP may be forced into network expenditure that is not required, valued or paid for by generators.

5.3.3 Super-firm Access under a Single-tier FAS

The Technical Report (August 2012) described[7] how, under a multi-tier FAS a generator could ensure sufficient access under lower FAS tiers by going *super-firm*. For example, a 100MW generator who required 100MW under NOC2 (with an 80% FAS scaling factor) could procure 125MW of access. A TNSP would then be required to ensure that the 100MW of flowgate capacity was provided under NOC2, although the access-limit meant that it would still only be required to provide 100MW during NOC1.

Under a single-tier FAS, there is no corresponding obligation on a TNSP. In recognition of this, the FAS definition has been changed somewhat since the Technical Report (August 2012). In order for super-firm access to be accommodated under a single-tier FAS, it would be necessary for the TNSP to provide flowgate capacity based on *agreed access* rather than (capacity-limited) *firm access*. In the above example, a TNSP would be required, under NOC, to provide flowgate capacity sufficient to provide 125MW of access to the super-firm generator, rather than 100MW as would have been the case under the previous FAS definition.

The implication of this means that, since FAS applies only to congested flowgates, the higher FAS requirement would only have any practical effect if the super-firm

---

6 The difficulties with customising in this way were discussed in section 5.3.6 of the Technical Report (August 2012).

7 Section 5.3.2 of the Technical Report (August 2012).

generator shared the flowgate with some non-firm generators, which would trigger congestion even if the flowgate capacity were built out to 125MW. Without non-firm generators, a flowgate with capacity above 100MW could never be congested (since the super-firm generator cannot operate above 100MW) and expanding it above this level would be unnecessary for the TNSP.

The super-firm generator will not benefit, during NOC, from a larger-than-100MW flowgate, since it would still only receive 100MW of access anyway. In fact, it is the non-firm generators who, in access settlement, would share between them the benefit of an extra 25MW of flowgate capacity. This might cause generators – who would not normally be inclined to assist their competitors – to prefer not to be super-firm.

On the other hand, under AOC conditions, benefits from the excess flowgate capacity *are* likely to flow to the super-firm generator. For example, if the 125MW of flowgate capacity falls by 20 per cent during a forced outage, the super-firm generator will still receive 100MW of access. This is similar to the outcome under a multi-tier FAS, the key difference being that there is now no specific *obligation* on the TNSP to provide 80 per cent of target flowgate capacity, this just happens to be the outcome as a result of the network topology chosen by the TNSP.[8] We note that this type of access may not be particularly useful, or well used by generators.

In summary, a single-tier FAS may be less effective than a multi-tier FAS in providing higher firmness to super-firm generators.

5.3.4 Establishing the FAS

If it is decided to implement the OFA model, the FAS will be developed as part of the implementation process. It is likely that the FAS would be developed in consultation with TNSPs and generators to ensure that it was both practical and useful. The OFA design places no limitations on FAS, except that NOC or AOC status can be identified for each flowgate (ie NEMDE transmission constraint): if not in advance of dispatch, then at least in time for access settlement.

5.3.5 Changing the FAS

Subsequent changes to the FAS after OFA implementation are somewhat problematic in that they affect the rights of generators with *existing* access agreements. It would be difficult – if not impossible – to grandfather existing access agreements and apply the new FAS only to subsequent agreements. Doing so might, in any case, largely defeat the purpose of the FAS change.

Alternatively, access charges payable on existing access agreements could be adjusted (up or down, as appropriate) to reflect a changed FAS. That may be rather more practical, but is conceptually problematic in that it forces changes on firm generators

8 Although mechanisms to encourage TNSPs to provide efficient levels of flowgate capacity to firm generators outside of NOC *are* included in the OFA model and are discussed in section 8.3.10.2 (Incentives outside the firm access standard).

who may not want them. Finally, it may be possible to adjust the agreed access amount on existing agreements to negate the impact of the FAS change.

In the light of these issues, the FAS would probably only be changed if necessary to address some unexpected – and unwanted – feature: for example, if definitions are ambiguous or able to be manipulated by TNSP decisions. The FAS could be embodied in the rules – or in schedules to the rules – which would mean that it would be changed through the usual rulemaking process.

5.3.6 Region-specific or NEM-wide FAS

The Technical Report (August 2012)[9] discussed this issue in the context of a multi-tier FAS. However, under a single-tier FAS, the issue is much simpler. Potentially, the only issue around a NEM-wide FAS would be harmonising differing definitions of planned outage and forced outage across different regions. This should be feasible. Thus, a NEM-wide FAS is proposed.

5.3.7 Demand-side Reliability Standards will continue

There is a large overlap between the FAS and the existing demand-side reliability standards. The FAS provides for minimum network access for firm generators under specified conditions, whereas demand-side reliability standards (except in Victoria) require sufficient network access under peak-demand conditions that enough generation can be dispatched to meet demand. By ensuring that firm generators can be dispatched, FAS helps in ensuring that demand can be supplied. However, neither standard dominates the other: that is, maintaining FAS is neither *sufficient* nor *necessary* for maintaining DSRS. The FAS does not replace the DSRS and so the latter must be retained if existing reliability of supply is to be maintained.

The key differences between FAS and DSRS are:

- FAS only applies to *firm* generation, whereas DSRS applies equally to firm and non-firm generation; and
- FAS only applies to *generation-side* flowgates, whereas DSRS applies to both generation-side and demand-side flowgates.

These two differences are explained below.

Implicitly, the DSRS can only be maintained if there is sufficient aggregate generation capacity to supply peak demand[10] and so this will be assumed in the discussion below. This requirement might be met by a combination of firm and non-firm generation, so it is possible that aggregate *firm* generation by itself is insufficient. In this case, in order to maintain DSRS, a TNSP shall be required to ensure that some

9 Section 5.3.7 of the Technical Report (August 2012).

10 The NEM generation market, together with the Reliability and Emergency Reserve Trader (RERT), ensure that this will be the case.

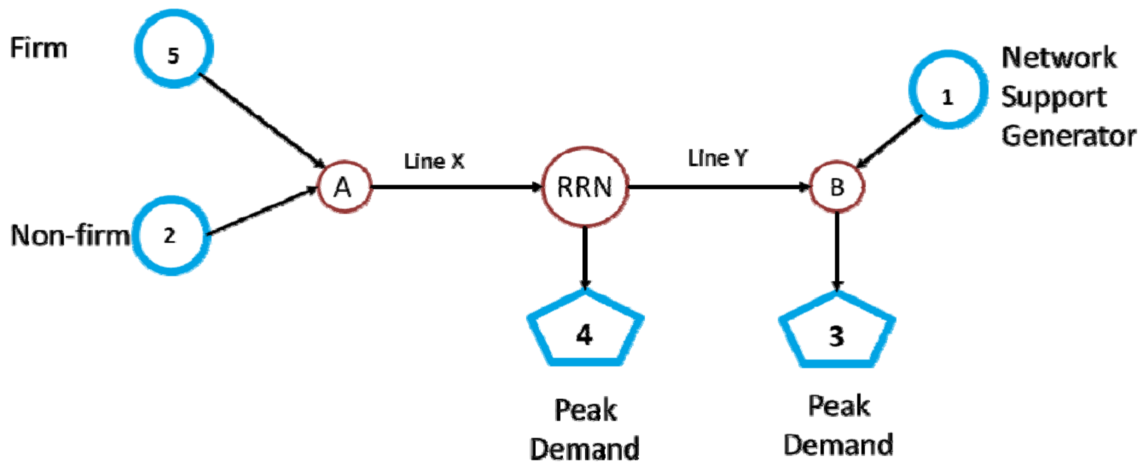
non-firm generation is able to be dispatched during the peak demand conditions referred to in the DSRS. Furthermore, this must be able to be dispatched without constraining-off firm generation – clearly, otherwise no extra generation in aggregate is dispatched – and therefore will receive RRP for its output: ie the non-firm generation receives some network access.

This access provided to non-firm generators is referred to in this document as *reliability access*.

In Figure 5.2, the firm generators connected at node A participate in the flowgate X (the flowgate associated with line X) but do not participate in line Y, which does not lie between the generators and the RRN.

FAS requires that the capacity of line X is at least 5GW. However, capacity of at least 6GW is necessary to maintain DSRS by ensuring that all forecast demand can be supplied.[11] The additional 1GW of capacity on line X that is required by DSRS provides a corresponding 1GW of reliability access, which is shared between the 2GW of non-firm generators at node A.

Figure 5.2 Simple three-node network



Reliability access is different from firm access in some important ways:

- it does not attract access charges (since the generator is non-firm);
- it is only provided if there is insufficient aggregate firm generation;
- it is only provided during DSRS peak periods; and
- it is not specific to a generator: in the example, 1GW of reliability access is shared between 2GW of non-firm generation

The provision and firmness of reliability access is uncertain and generators seeking certainty will prefer to procure firm access. Therefore, the continuation of DSRS does

11 Total demand is 7GW and 1GW of this can be supplied by the network support generator.

not negate the impacts and benefits associated with firm access. However, it may dilute them to some extent, if it means that more generators opt for non-firm access.

The second difference between FAS and DSRS is that FAS applies only to *generation-side* flowgates: ie flowgates in which at least one generator has positive participation. It places no obligation on TNSPs in relation to demand-side flowgates: ie flowgates without any positive-participation generators. In Figure 5.2 line X is a generation-side flowgate and line Y is a demand-side flowgate.

A TNSP must maintain sufficient demand-side flowgate capacity to meet DSRS, but this is not required to meet FAS, since demand-side congestion does not cause firm generators to be constrained-off. In the example, line Y has no FAS requirement, but must have capacity of at least 2GW to allow demand at node B to be supplied.

DSRS and reliability access have further implications for access pricing, which are discussed in chapter 6 (Access Pricing).

5.3.8 Reliability Access Safety Net

The need to preserve reliability standards whilst moving towards a regime of more market-driven expansion is a familiar concern for NEM design. Precisely this concern arose prior to NEM commencement, when a generation market was to be introduced and so generation expansion could no longer be centrally-planned to ensure sufficient generation capacity to reliably supply peak demand.

The solution to this difficulty was not to retain central planning of *all* generation development, but rather to rely on the market to provide the majority of new generation and to introduce a safety net process – known as the *Reliability and Emergency Reserve Trader (RERT)*[12] – to cover any shortfall. The RERT is operated by AEMO and is essentially a central planning process to ensure reliability.[13] When AEMO identifies an impending shortfall in generation capacity, it operates a tender process to purchase the additional generation capacity (ie capacity not already operating in the market) required, at lowest cost. The safety net has rarely been needed, but its presence provides confidence that reliability will be maintained, irrespective of market outcomes.[14]

An analogous safety net process could be designed for the OFA model. In this case, the process would be operated by TNSPs, who would identify where firm generation was insufficient to maintain DSRS. Because firm access is sold rather than purchased, the tender process would be a *reverse* auction: ie the TNSP would aim to *sell* the required additional firm access at lowest cost to the TNSP.

12 This previously was known as the "reserve trader".

13 We have recently published a final determination on a Rule Change which would result in the expiry of the RERT on 30 June 2016. See: AEMC, *Expiry of the Reliability and Emergency Reserve Trader*, Final Determination, 15 March 2012.

14 The RERT arrangements have been used twice to provide additional reserve capacity, in Victoria and South Australia. However, in both cases the additional capacity was not dispatched.

The cost in this case would be the aggregate *discount* to the standard access charge at which the firm access is sold in the tender. So non-firm generators choosing to participate in the tender would offer a $/kW discount price and the TNSP would choose the generators offering the low discount to sell the firm access to.

There was a concern at the time that the reserve trader safety net was designed that peaking generators would stay out of the market in order to get a better price from the reserve trader. Similarly, it is possible that the existence of the reliability access safety net could encourage generators to remain non-firm and then hope to purchase discounted firm access through the safety net process. The fears about the reserve trader were not realised. It is unclear what impact a safety net would have on the firm access market.

A reliability access safety net process is not part of the proposed OFA model, but it could be considered by the OFA implementation process as a potential *add-on* to the OFA design.

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 FAS, 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*. The purpose of *access pricing* is to estimate what these costs are.

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.[1]

LRIC is defined to be the difference between two costs: the *baseline cos t*, which is the NPV of the *baseline expansion plan* which is in place before the access request is received; and the higher *adjusted cost*, which is the NPV of the *adjusted expansion plan*: an amendment to the baseline expansion plan to accommodate the new access request:

LRIC = adjusted cost – baseline cost

The expansion 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.

1 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

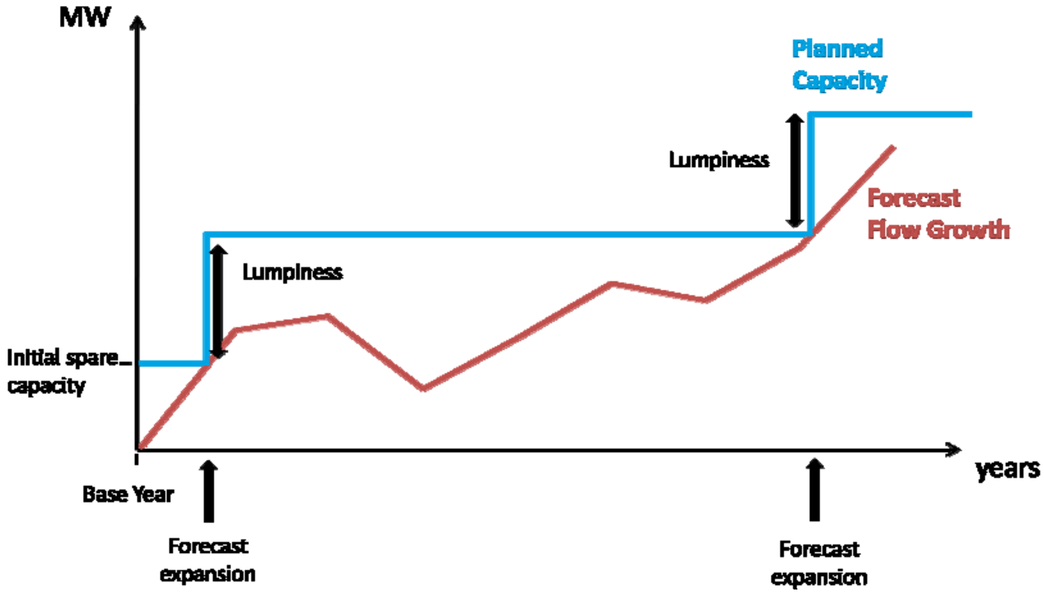
6.2.1 The Element-based Expansion Model

The access pricing methodology establishes a simplified model of transmission planning by assuming that separate, independent expansion plans are developed for each existing *branch element* (such as a transmission line or network transformer) of the shared transmission network. Each element's *baseline* expansion is based on three variables:

- *initial spare capacity*: the amount of spare capacity on the element in the base year;
- *annual flow growth*: the amount by which maximum flows on the element are forecast to increase each year; and
- *lumpiness*: reflecting the size of a practical and economic expansion of that element.

It is assumed that each element is expanded as soon as spare capacity is exhausted.[2] That expansion provides new spare capacity, which will be progressively eroded through subsequent flow growth until, eventually, a second expansion is required, and so on. This expansion model is illustrated in Figure 6.1 below.

Figure 6.1 Element baseline expansion model



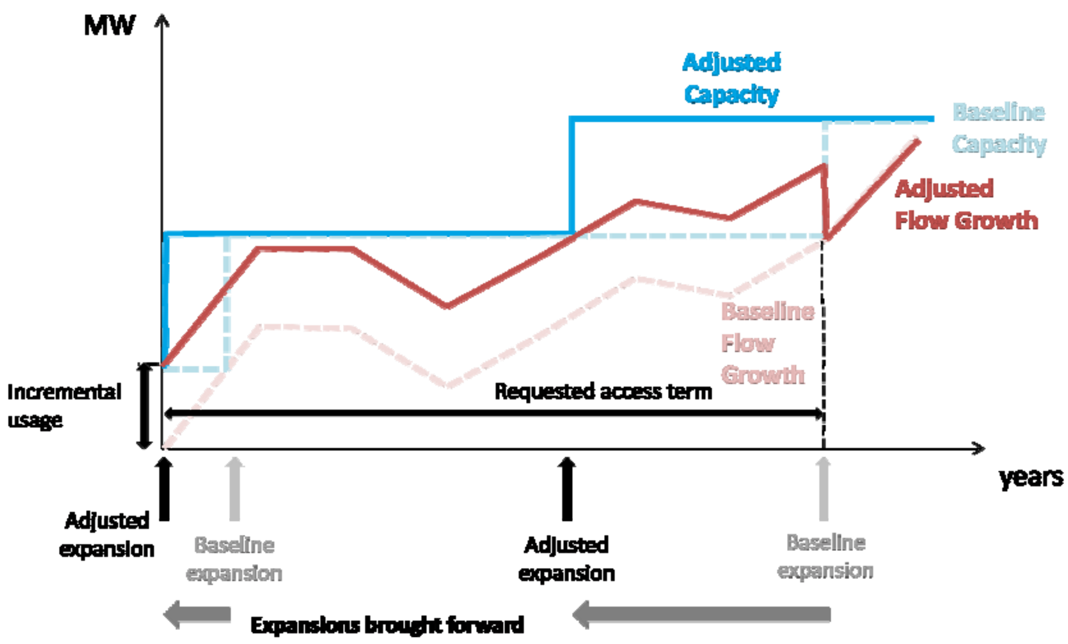
2 By assuming a piecewise linear demand growth each year, the time of expansion is estimated to a fraction of a year which, while unrealistic in practice, avoids the jerkiness associated with rounding to the nearest whole year.

To model the *adjusted* expansion, the impact of the new access request is included, which is represented by two further variables:

- *incremental usage*: the extra flow induced on the element by the access request; and
- *access term*: the period of the access request and so the period for which the extra flow occurs.

This incremental usage simply adds to the baseline flow growth and will cause the expansions in the baseline plan to be *brought forward* by varying amounts, as illustrated in Figure 6.2, below.

Figure 6.2 Element adjusted expansion model



The NPVs of baseline cost and adjusted cost are then calculated by applying an appropriate discount rate to the capital costs implied by the corresponding expansion plans. The access charge is the difference between these two NPVs, summed over all transmission elements in the network.[3]

The model is not a complete or realistic description of actual transmission planning and is not intended to be. In particular, new elements and changing connectivities are often introduced in expansion plans and these practices are not represented in the stylised model. On the other hand, it will be seen that the LRIC calculated by the model is smooth and proportionate: small changes in access request will generally give rise to small changes in LRIC. As discussed below, it also has characteristics which are similar to what one would expect of a true LRIC.

3 In practice, incremental usage will only be material on a subset of elements, generally those elements lying between the new access node and the RRN and so LRIC on only these elements needs to be calculated and summed.

6.2.2 Medium-term and Long-term Forecasting

The long-lived nature of transmission assets, together with the relatively low discount rate applicable to TNSP businesses, mean that the element-based model must cover many years into the future to avoid the problem of end-effects – eg a modelled expansion in the final year of the model – substantially impacting pricing outcomes. On the other hand, the forecasts upon which the model variables are predicated become increasingly uncertain into the future. Discounting also means that the influence of these longer-term forecasts is diminished. Thus, there comes a point at which inclusion of explicit, detailed forecasts simply generates spurious accuracy: ie the appearance but not the substance of improved accuracy.

To avoid this, and to simplify the access pricing model, forecasts beyond a certain point will be *stylised* rather than precise. For example, for further than 10 years (say) out, flow on an element could be assumed to grow at a fixed rate, rather than be calculated based on explicit demand and generation forecasts. The assumed rate could be standardised for different types of element: eg 3 per cent (of lumpiness) per year for *core* elements, 1 per cent per year for *local* elements.[4]

The point of delineation between the explicit *medium-term* forecasting and the stylised *longer-term* forecasting will be defined during the implementation process.

6.2.3 Estimating the Model Variables

Methods for determining the necessary model variables are described below.

Initial spare capacity

Various planning studies would be carried out, based on transmission, access and TUOS conditions expected under FAS normal operating conditions[5] in the *base year*: the first year of the access request. In each study, firm generators would be dispatched at their agreed access level. The *study spare capacity* on an element would be the difference between the study flow and the secure flow limit.[6] The *initial spare capacity* is the lowest value of all the study spare capacities.[7]

Annual flow growth

During the medium-term period, this will be based on end-user demand forecasts and firm generation forecasts for each node and each year of the pricing analysis, based on forecasts produced in the National Transmission Network Development Plan (NTNDP) or similar publication, as well as current access agreements and requests. A

---

4 It is acknowledged that this may not be straightforward given recent unpredictable and declining patterns of demand.

5 See chapter 5 (Firm Access Standard).

6 Ie maximum flow level consistent with secure dispatch.

7 Or, so that it is not too dependent upon a single study, it might be set at, say, the average of the lowest five study spare capacities.

standard load flow analysis would be used to convert these nodal forecasts into flow forecasts for each element in each study year.

For the longer-term period, a simple, stylised assumption for annual flow growth will be used.

Lumpiness

The lumpiness of an element is set equal to its expansion size *divided* by its *meshedness*, for reasons discussed briefly below and in more detail in section 12.9 (Meshedness in Access Pricing). The expansion size would be chosen, from the practical engineering alternatives available, based on minimising the NPV cost of current and future expansions.[8]

Meshedness

Meshedness is a measure of how many elements run in parallel to the studied element.[9] If four, identical lines operate in parallel, an expansion of one line by 1,000MW (say) is equivalent to a 250MW expansion across all four lines. This situation is best approximated in the model by assuming that each line can be expanded, independently, by 250MW at a time, not 1,000MW at a time, which gives the reason for the lumpiness formula, above.

Incremental usage

Incremental usage is equal to the amount by which flow on the element increases when the amount of the access request is dispatched across the transmission network to supply the incremental demand at the RRN. The access amount applied is the nominal amount and not the capacity-limited amount. That means that a super-firm generator pays the same for access as a firm generator (eg 80MW and 100MW generators would pay the same amount for 100MW of access).

Access term

The access term is based on the access request.

Expansion cost

The expansion cost is based on the capital cost of the expansion, divided by the meshedness.

8 This would be based on average flow growth and discount rate. Where flow growth was fast a relatively larger expansion size would be more efficient.

9 In practice, lines do not run exactly in parallel. Meshedness is defined more precisely by modelling a load flow between the two nodes at each end of an element and calculating the factor by which the total flow from node to node exceeds the flow on the element itself.

Discount rate

The discount rate would be based on an estimate of the TNSP's regulated cost of capital.

6.2.4 Special Situations

Some special situations which need to be managed in the LRIC pricing model are discussed below.

Counterflow incremental usage

An access request in a demand-rich location may create incremental usage which is in the *opposite* direction to baseline flows on many elements. This would have the effect of increasing spare capacity and potentially *deferring* future expansion and so creating a *negative* LRIC. However, spare capacity is only increased in practice if the associated power station is actually *available and dispatched* in the critical peak period. The TNSP is unable to rely on this on the basis of the access agreement alone.[10] Therefore, the element LRIC would be zero (rather than negative) in this situation.

Inter-regional effects

New access in one region may create incremental usage – and so LRIC – on some elements in a neighbouring *remote* region. Conversely, nodal demand or firm access changes in the remote region may cause changes to flows on elements in the local region. Thus, the access pricing model needs to include these remote elements and nodes to the extent there are material inter-regional impacts. Where the access request generates material LRICs on remote elements, corresponding payments should be made to the remote TNSP.

Reliability-driven expansion

Obligations on TNSPs to maintain jurisdictional *reliability standards* will continue in the OFA model. Possible future reliability expansion – expansion needed to maintain reliability standards – would need to be modelled and included in the baseline expansion plan. That would be done by including suitable *reliability generation* (ie any non-firm generation for which peak-period access must be provided in order to maintain reliability) in the relevant pricing studies.[11] If a reliability generator sought access, the *calculated* access price would then be zero, because the generator appears equally in the baseline studies and the adjusted studies. To avoid this anomaly, that particular generator would be removed from the baseline plan in that situation.[12]

10 However, the TNSP could enter into a separate network support agreement with the generator, under which the generator commits to be available and dispatched as needed and receives a payment from the TNSP in return. Similar agreements are made in the present NEM design.

11 The TNSP would need to decide which generators would be reliability generators, typically those non-firm generators to whom transmission access can be provided most cheaply.

12 In fact, similar issues arise in relation to all access requests, as discussed in section 6.3.7 (Including Pending Access Requests in Forecast).

6.2.5 Payment Profiling Algorithm

The access pricing methodology calculates a lump sum cost which would be recovered through annual payments[13] over the life of the access agreement. The payment profiling algorithm determines these annual payments.

There are a number of considerations relevant to payment profiling:

- the preference of the generator and the ability of the generator to negotiate variations from the standard payment profile;
- the cashflow and borrowing implications for a TNSP of timing mismatches between expansion costs incurred and access revenue received;[14] and
- the regulatory implications of revenues and costs varying within and between regulatory control periods.[15]

These considerations make payment profiling potentially complex. They are discussed further in section 12.10 (Annual Payment Profiling).However, it should be noted here that the profiling and payments are likely to depend on indices such as the consumer price index (CPI) and regulatory weighted average cost of capital (WACC). Therefore, annual access charges may vary as these indices vary, creating some modest uncertainty for the generator.

6.2.6 Summary

The access charge for an access request is based on the estimated long-run incremental cost incurred by the TNSP in expanding its network to accommodate the new access, in accordance with the FAS. That charge is calculated and agreed during the access procurement process and cannot be subsequently amended, apart from agreed future indexation based on CPI, regulated WACC or similar indices.

The access pricing methodology is based on a highly stylised model of transmission expansion which, nevertheless, is expected to broadly reflect the characteristics and levels of a true LRIC forecast. It is designed to provide smooth, transparent and robust prices which guide efficient generator behaviour whilst covering the cost to TNSPs of providing firm access services.

13 Perhaps broken down into twelve monthly payments.

14 Recognising that, in NPV terms, the revenues and costs should be similar.

15 These are discussed further in section 8.3.2 (TNSP risk from lumpy access-driven expansion costs).

6.3 Design Issues and Options

6.3.1 Why not use LRMC or deep connection

The OFA model uses an LRIC-based methodology as described above. However, there are two alternative approaches, referred to here as long run marginal cost (LRMC) and deep connection, which are used in other electricity markets.[16] These were considered for the OFA model, but were rejected for the reasons discussed below.

6.3.1.1 LRMC

An LRMC approach is similar to LRIC, with the essential differences that expansion *lumpiness* is *ignored*: it is assumed that, if an additional 233MW, say, of transmission capacity is required, exactly 233MW will be built. The access charge for 233MW will reflect that and will be set at 233 times the average $/MW cost of transmission expansion. In general, the LRMC on an element is the incremental usage on that element multiplied by the average expansion cost.

LRMC is a much simpler methodology than LRIC, because there is no need to take account of existing spare capacity or future planned expansions: capacity is expanded only as needed and so tracks the flow growth rather than occurring in steps. However, this simplification is also its flaw. Other things being equal, 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, despite the incremental cost of transmission being much higher at the latter location. Generators will choose locations that are best for them (in terms of land and fuel availability), rather than those where access can be provided more cheaply by the TNSP, due to existing spare capacity.

The materiality of this pricing inaccuracy is unclear and it may be that LRMC is actually quite a good proxy for LRIC. That will be revealed during more detailed examination of LRIC during the OFA implementation process and a decision whether to switch to an LRMC approach (perhaps driven by the relatively greater complexity of the LRIC methodology) could be made at that stage.

6.3.1.2 Deep Connection Charge

A deep connection charging approach levies only the *immediate* costs of transmission expansion through the access charge and takes no account of *future* costs as a result of future expansions being advanced. Rather than using a stylised methodology, a deep connection approach would rely on the TNSP (or other institution) determining *exactly* what needs to be built immediately to provide the new access and charging for the cost of that: analogous to what occurs currently for connection charging.

16 These terms can be used in various ways, but a specific meaning is applied here, as described below.

As with the LRMC approach, the fundamental problem with this approach is that the deep connection costs do not reflect the true incremental cost of access provision. Unlike with LRMC, however, it is clear in this case that the costing could be highly inaccurate. Consider a generator seeking access where there is limited spare capacity but an upgrade is planned in two years' time. The new generator would prompt immediate expansion and be charged the full cost of the expansion, even though it had simply caused the expansion to be brought forward by two years.

Another difficulty with deep connection is that the generator paying the cost is likely to demand the smallest possible (and hence cheapest in absolute terms) expansion, despite this being uneconomic in the longer term. There may be ways to help correct this inefficiency (eg by giving the generator some form of marketing rights on a larger expansion) but these introduce substantial additional complexity.

6.3.2 The value of spare capacity

Any new access will change the amount of spare capacity on an element. If the new access prompts immediate lumpy expansion, the amount of spare capacity is likely to increase, as the lumpy addition will typically exceed the new access requirement. Alternatively, if no immediate expansion is required, the amount of spare capacity must decrease, as some of it is now being used to provide access.

Although spare capacity is, by definition, *currently* unused, it is likely to have some value due to the possibility of it being used to provide some future access. Because of discounting, this (net 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 deep connection charging methods is that the former method charges (or pays) the new generator the value associated with any reduction (or increase) in spare capacity, whilst the latter method does not. Thus, where there is no immediate expansion, the LRIC reflects the value of the spare capacity now utilised by the generator. On the other hand, where there is immediate expansion, the LRIC reflects the cost of the expansion *minus* the value of the increase in spare capacity: the extent to which the lumpy expansion is not required by the new generator.

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.

As a special case, if spare capacity on an element is estimated to have zero value – because it is not anticipated to be used for future access – then LRIC and deep connection charge give identical outcomes. Thus, deep connection provides the correct price in this special case, but not in the general case where spare capacity has value.

6.3.3 Comparison of LRIC, LRMC and Deep Connection

Notwithstanding the disadvantages of the LRMC and deep connection methodologies, as discussed above, LRIC pricing may give similar outcomes to one or other of these models in certain circumstances. Such situations are discussed below.

6.3.3.1 Elements where LRIC will be similar to LRMC

As noted, LRMC assumes there is no lumpiness. In that case, expansion of an element notionally occurs annually, with each annual expansion exactly matching annual flow growth.

In the LRIC model, because there is lumpiness, expansion occurs in cycles rather than annually. For example, if lumpiness is 1,000MW and annual flow growth is 200MW per year, expansions will occur every 5 years. If the investment cycle is fairly short,[17] lumpiness is less material and LRIC will be broadly similar to LRMC.

These conditions are most likely to apply in the *core grid*: the main high voltage backbone of the transmission network. Here, large expansion lumps are offset by high meshedness,[18] and there is likely to be relatively high flow growth. Therefore, LRIC pricing of some elements in the core grid may be broadly similar to LRMC.

6.3.3.2 Elements where LRIC will be similar to Deep Connection

The opposite situation to the one described above is where the investment cycle is long: lumpiness is *high* relative to annual flow growth. At the extreme, where flow growth is zero, the investment cycle is infinite: ie there are no planned expansions of the element in the baseline expansion plan. In the adjusted expansion plan there will simply be:

- *no expansion*: if there is sufficient initial spare capacity to accommodate the incremental usage from the new access request; or
- *immediate lumpy expansion*: otherwise.

Recall that LRIC is based on the difference between the adjusted and baseline expansion cost, so in this case LRIC either charges *nothing* if no immediate investment is required, or the full expansion cost if it is. This is the same as the deep connection charge. More generally, if the investment cycle is long (relative to the discount rate), LRIC will be broadly similar to deep connection.

17 The cycle length is measured relative to the discount rate. At a 5 per cent discount rate, net present value is discounted by 21 per cent over 5 years, meaning broadly that LRIC might vary by 21 per cent depending upon the point in the investment cycle at which the new access commences: whether it is just prior to or just after a baseline expansion.

18 Recalling that in the LRIC model meshedness is defined as the expansion lump divided by the meshedness.

This analysis reaches – through a different route – the same conclusion as that in section 6.3.1: that deep connection charge and LRIC give the same outcomes when flow growth on an element is estimated to be zero.

These conditions are most likely to apply in the local grid: lower voltage lines either serving specific power stations or local load.[19]

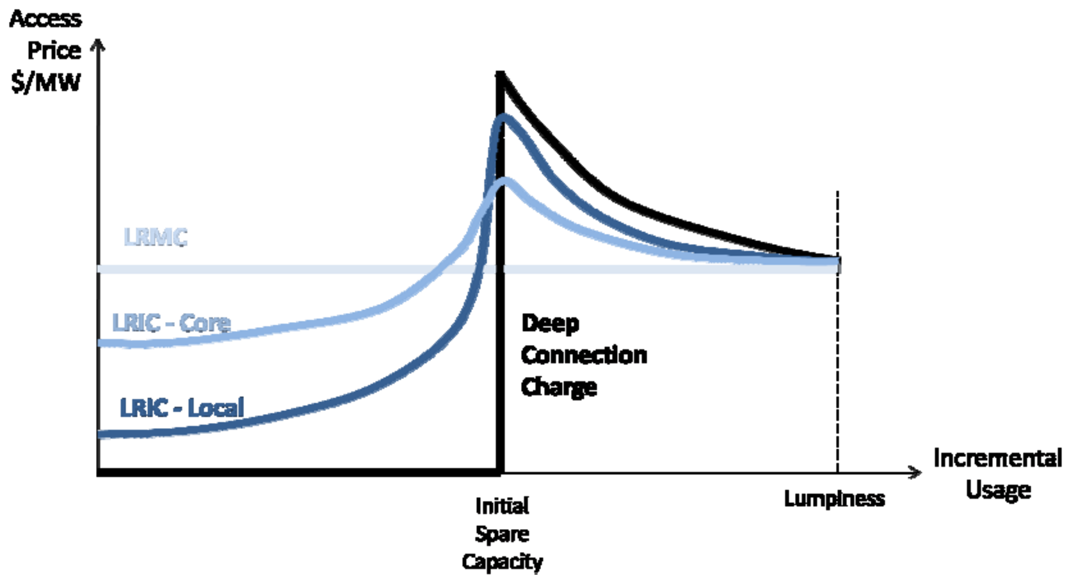
6.3.3.3 General Situation

Figure 6.3 illustrates how the incremental price (the incremental cost divided by the incremental usage) on an element varies by incremental usage, for the three methodologies. It will be seen from the figure that:

- if the incremental usage equals the expansion size, the three methodologies give identical outcomes, reflecting the cost of an immediate lumpy expansion;
- the LRMC price is constant, because the LRMC cost is proportional to the incremental usage;
- the Deep Connection charge is either zero or the full expansion cost: note that the incremental price is the cost divided by the incremental usage and so – given a fixed cost – decreases as usage increases;
- LRIC is closer to Deep Connection where there is a long investment cycle: low flow growth relative to lumpiness; and
- LRIC is closer to LRMC where there is a short investment cycle: high flow growth relative to lumpiness.

19 In the latter case, flow growth may even be negative, but since LRIC can never be negative, this is equivalent to the zero growth case.

Figure 6.3 Comparison of LRMC, LRIC and Deep Connection



The access charge is the aggregate of LRIC across all elements. The overall outcome, then, will depend upon the mix of long-investment-cycle and short-investment-cycle elements used to provide the requested access.

6.3.4 Central Pricing is not Central Planning

Ideally, access prices would be set by the market, like wholesale energy prices, rather than determined administratively. Of course, this is not possible, as TNSPs are monopolies and so there can be no competitive market for access provision. Thus, access pricing must be highly regulated, just as TUOS pricing is currently.

The fact of regulated prices, together with the way they are predicated on a central forecast of demand and generation, has led to concerns from some stakeholders that access pricing amounts to central (transmission) planning by stealth, contrary to the objective of the OFA model that transmission planning should, on the generation side, be more *market* driven.

It is acknowledged that prices will affect generator decisions and so centralised pricing necessarily establishes a central influence on generation. However, this does not make it *central planning*, which usually refers to a *command and control* approach to generation investment.

A more specific concern is that generation outcomes will reflect the forecasts embedded in the pricing method, making these forecasts a *self-fulfilling prophecy*: whatever scenario is used in the pricing model will eventually come about, because the prices guide generators to follow it. This concern reflects a misunderstanding of the characteristics of LRIC prices and the influence of forecasts on these.

For access forecasts to be self-fulfilling, higher levels of forecast firm generation at a location must lead to *lower* access prices, thus encouraging more generators to locate there. Conversely, lower levels of forecast generation must lead to higher prices. Higher forecast generation means higher flow growth on the elements connecting that location to the RRN. But does this actually lead to lower LRIC?

Recall from the discussion in section 6.3.3 (Comparison of LRIC, LRMC and Deep Connection), summarised in Figure 6.3, that the impact on LRIC of higher load flow growth on an element is to *flatten* the LRIC curve (ie make it less dependent upon the level of spare capacity and the size of the access request) and make it more similar to LRMC. This flattening means that, depending upon the particular access request and the level of spare capacity, higher load flow growth could lead to either *higher* prices or lower prices. In particular, on elements with high levels of spare capacity, higher flow growth will lead to higher prices. Thus, the forecasts in this situation become *self-denying* rather than self-fulfilling.

6.3.5 What if the Expansion Cost is much higher than Modelled

There is a risk that the access price could underestimate the true incremental cost of access provision, occasionally or systematically. Any pricing errors could cause inefficiency, to the extent they fail to signal the efficient location for new generation. However, any future inaccuracies must be compared to the status quo where there is no transmission price for generators to signal efficient locations.

Systematic errors should be able to be identified in the implementation process, using detailed analysis of access pricing under various generation scenarios and comparing those with actual expansion costs. The pricing model can then be recalibrated as necessary to ensure that costs are correct on average. To the extent under-pricing remains, the shortfall in costs will be recovered from demand-side users through higher TUOS charges: which, again, is likely to be no *worse* than what happens at present. Conversely, to the extent that there is over-pricing, demand-side users will benefit through lower TUOS charges.

6.3.6 How to Ensure Access Pricing is Objective and Transparent

The access prices calculated by the LRIC method are highly dependent on forecasts of flow growth on elements which, in turn, are dependent on forecasts of end-user demand and firm generation. For pricing and procurement efficiency, these forecasts must be accurate, objective and transparent.

To ensure this, it is proposed that the access pricing model and the background forecasts that drive it are managed and maintained by the NTP. Forecasts will be based on the NTNDP which is the product of an open and transparent process, or other similar information developed and published by AEMO.[20] Either TNSPs or a central

20 For example, we have also recommended as part of the Transmission Frameworks Review, that AEMO now be responsible for producing connection point demand forecasts for each region in the NEM.

pricing agency will use the model – or a faithful copy of the model – to calculate access prices.[21]

Under this approach, it may be possible to make the pricing model available for prospective firm generators to use independently, albeit informally, to help in deciding on their location and access level.

6.3.7 Including Pending Access Requests in Forecast

There is an interaction between access requests and access prices, since a TNSP is bound to take some notice of pending access requests in producing a baseline forecast. For example, a TNSP might make the assumption that 30 per cent of requests currently being processed will proceed through to completion. Therefore, when any request is priced, 30 per cent of the request will already be in the baseline and so the adjusted forecast would only add the remainder 70 per cent. Obviously, costing only the extra 70 per cent would significantly understate the true incremental cost.

Thus, there would be a need to create a special baseline scenario for each request, which would remove the request from the baseline, but leave all other pending requests.

6.3.8 Reliability Access

As discussed in section 5.3.7 (Demand-side Reliability Standards will continue), DSRS will continue under the OFA model. This means that, where aggregate firm generation is less than forecast peak demand, a TNSP will be required to provide some *reliability access* to non-firm generators.

A TNSP will incur some cost in providing reliability access. The more firm generation that there is, the *less* reliability access will need to be provided. This inverse relationship means that, whilst there is a *direct cost* to a TNSP in providing new firm access (as estimated by the LRIC method), there is also some *indirect saving* from the reduction in the amount of reliability access that must be provided. It is even possible that, had a generator not procured firm access, they would have been provided with reliability access instead, at exactly the same cost to the TNSP.

This indirect saving to the TNSP is not estimated in the LRIC pricing method. Indeed, care is taken in access pricing to prevent this happening, as discussed in section 6.2.4 (Special Situations). Some stakeholders have questioned whether this is appropriate and whether, as a result, firm access may be overpriced.

This is a difficult issue and there is no obvious correct approach. The root source of the difficulty is the DSRS and the result that non-firm generators may be provided access by the TNSP, in conflict with the fundamental approach of the theoretical OFA model ideal. It is, of course, not permissible to simply drop the DSRS on the generation-side

21 The decision as to whether TNSPs or a central agency will actually carry out pricing will be determined during OFA implementation.

and, if insufficient firm access is procured, potentially have a shortfall in meeting load requirements. The OFA model has to coexist with the DSRS. Therefore, the OFA ideal will inevitably be distorted in some way; the question is what approach to adopt that will minimise the impact of this distortion on generator and TNSP decision making.

Generators, understandably, would seem likely to prefer to receive an explicit discount associated with the reliability-access saving, but it is not clear that this will be the least distortionary. It is also not clear, practically, how the saving might satisfactorily be estimated in the access pricing methodology, especially given that the DSRS is (currently) different in each region whereas access pricing is NEM-wide.

On the other hand, there are two mechanisms already in the OFA blueprint that will provide some level of discounting of access prices.[22] The first mechanism is inherent in the LRIC method, since the presence of the DSRS – and the extra transmission capacity associated with it – lead to higher levels of spare capacity and so lower LRIC prices.

The second mechanism arises out of short-term access issuance. Reliability access will create spare capacity in the network which will facilitate additional short-term access issuance through the auction process described in section 7.2.4 (Short-term Firm Access Issuance). These auctions are likely to clear at prices less than the LRIC. So, the ST auction process is a way of converting reliability access into discounted firm access, for those generators prepared to pay the auction prices.

22 There is also a third possible mechanism, a reliability access safety net, described in section 5.3.8 (Reliability Access Safety Net).

7 Access Procurement

7.1 Overview

Access procurement covers the processes through which generators can procure new or additional *firm access service*, by entering into a *firm access agreement* with the TNSP in its region (the *local TNSP*) or by purchasing access from a generator who has previously procured its access from the TNSP.

Default access service terms and prices are regulated. These terms can be *customised* by mutual agreement, but only to the extent that this does not impact adversely on other transmission users. Primarily, though, the procurement process involves information exchange rather than commercial negotiation. Specifically, the generator seeks the combination of access level, location and term that best meets its needs.

Access pricing and procurement interact, since prices depend upon existing and prospective access agreements. Therefore each access request or agreement may affect the pricing of other, concurrent requests. The procurement process must be structured to manage these interactions so as to avoid placing undue risk and uncertainty on generators or TNSPs. A possible process is described below, but other processes, which could be developed by TNSPs in consultation with generators, are not ruled out.

The access agreement specifies the *access charge* and *service parameters*: the latter covering aspects such as term, amount and location. It may also include some standard terms such as prudential requirements, termination and assignment: given that, in procuring access, a generator is committing to pay a potentially large sum of money over a long period, effective prudential requirements may be paramount. However, most terms of service – such as service standard and liability – will lie outside the agreement, in rules and regulations.

Apart from the standard issuance process, TNSPs will be permitted to sell short-term access through an auction process. Generators are able to sell access into this auction, or alternatively sell directly to another generator, subject to TNSP approval. Prices in these other processes are not regulated or based on LRIC, but will instead depend upon the interaction of bids and offers.

There is no obligation on generators to procure firm access. Generators who do not do so will not be required to make any payment to the TNSP, but will receive a lower level of access firmness and so will receive fewer payments from – and make larger payments into – in access settlement.

7.2 Design Blueprint

7.2.1 Service Parameters

Through the procurement process, a generator decides upon a set of *service parameters* which best meet its access needs and for which it is prepared to pay the associated access charge.[1] The process will typically be iterative, with the generator submitting a request, the TNSP providing a price associated with that request and the generator then amending its request in response. The list of service parameters – and associated restrictions on them – is presented in Table 7.1 below.

Table 7.1 Firm Access Service Parameters

Parameter	Description	Restrictions
Amount (MW)	Nominal level of service	Not limited: eg by power station capacity.
Power Station(s)	Generating units to which the service applies	Must be connected to the shared network at a common point[2] (node).
Node	Transmission node from which access applies	Must be the point at which the power station(s) connects to the shared transmission network.
Term	Service commencement date and expiry date	Commencement may be delayed until transmission expansion can occur.
Profile	Variation of the nominal service level with time	Peak and/or off-peak, following forward energy contract convention.
Payments	Payment dates, amounts and indexation	Discussed in chapter 6 (Access Pricing).
Custom	Agreed variations from the default service terms	If these can be settled by AEMO, and do not adversely affect other users.

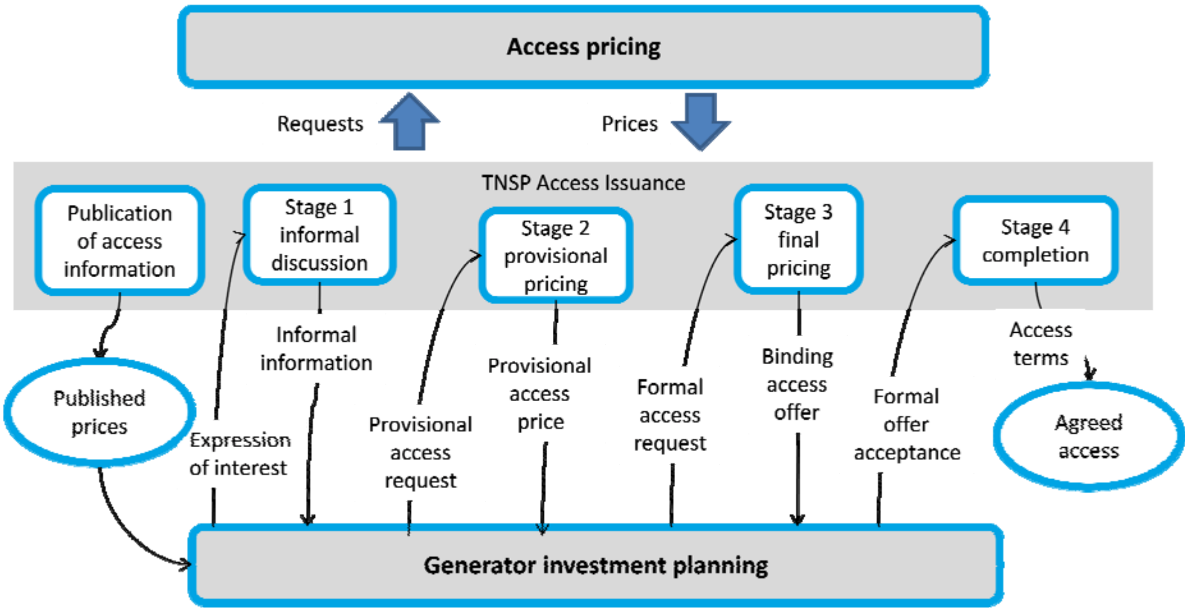
However, the role of the TNSP will be not simply to *provide a price* for the requests made, but also to advise the generator on the characteristics of the access pricing methodology and on possible service parameters that might best meet the generator's needs. An illustrative example of this is presented in section 7.3.1 (A Procurement Example).

1 See chapter 6 (Access Pricing).

2 For an embedded generator, this would be the point at which the relevant distribution network connects to the transmission network, discussed further in section 7.3.7 (Firm Access Procurement for Embedded Generation) for Embedded Generation.

7.2.2 Procurement Stages

Figure 7.1 Illustrative access procurement process



A possible staged procurement process is presented in Figure 7.1, above. This is intended to illustrate one way in which procurement objectives may be met, but other and better ways may be developed by, or in consultation with, TNSPs and generators. The stages are described in Table 7.2, below.

Table 7.2 Access procurement stages

Stage	Info provided by G to TNSP	Info provided by TNSP to G	Obligations[3]	Queuing
Publication of Access Information	None	Publication of indicative prices at all nodes for standard terms and amounts	TNSP to update information annually[4]	none
Stage 1: Informal Discussion	Indicative location(s) and access level	Indicative prices, key breakpoints in price for amounts, times, locations	TNSP to provide timely information in good faith	none
Stage 2: Provisional Pricing	Specific provisional access request	Provisional price for request	G to pay TNSP for costs incurred.	First-come-first-served queue

3 TNSP obligations will be set out in regulation, as discussed in chapter 8 (TNSP regulation).

4 And perhaps published in the annual NTNDP.

Stage	Info provided by G to TNSP	Info provided by TNSP to G	Obligations[3]	Queuing
Stage 3: Final Pricing	Formal access request	Binding, time-limited, price offer	G to provide refundable deposit.	One request at a time admitted in stage 2 queue order
Stage 4: Completion	Binding acceptance of stage 3 offer	Signed access agreement	G to provide non-refundable deposit (deducted from access charge)	none

A generator may withdraw from the procurement process at any stage, until the agreement is finalised in stage 4. Once completed, prudential arrangements would come into effect, but firm access would only be provided from the specified commencement date which – where a new power station or transmission expansion is being constructed – may be several years after procurement completion.

As discussed in section 6.3.6 (How to Ensure Access Pricing is Objective and Transparent) it may be feasible to make the pricing model available for generators to use informally. That might reduce reliance on TNSP input: substantially in stage 1, above, and partly in stage 2. However, the application process and the formal making and acceptance of an access offer would still rely on TNSP involvement.

7.2.3 Confidentiality

The details of any generation project are commercially sensitive in its early stages, but nevertheless its details are published in AEMO's Statement of Opportunities once it reaches a certain stage. Similarly, Stage 1 access requests would be confidential, but progress in later stages will be published to ensure transparency of the application and pricing processes. Once agreed, service parameters will be published. The question of whether details of access charges, payment arrangements and any customisations of service parameters should be published needs to be considered further.

7.2.4 Short-term Firm Access Issuance

FAS requires that effective flowgate capacity is no less than target flowgate capacity.[5] In practice, effective flowgate capacity might be or could be made to be significantly higher than the target in some parts of the network, due to:

- legacy transmission capacity: developed prior to commencement of the OFA regime;
- lumpy expansion;

5 Discussed in chapter 5 (Firm Access Standard).

- DSRS: providing reliability access to some non-firm generators; and
- TNSP undertaking operational actions to improve transmission capacity: eg weather-dependent line ratings.

It is proposed that this additional capacity may be used by a TNSP to back the issuance of short-term (ST) firm access, subject to:

- the TNSP remaining FAS compliant: ie there is sufficient flowgate capacity to accommodate both conventional and ST firm access;
- the horizon[6] of ST access being within a specified period: 3 months out is proposed, although an amount up to 12 months out could be considered;
- issuance taking place through an *open auction*: which will determine the amount and price of ST access sold;[7] and
- ST access not requiring or prompting any capital expenditure: in particular, revenues and liabilities associated with ST access must not be included within a RIT-T assessment.[8]

Within these restrictions, each TNSP would be free to develop the specifics of the auction process and parameters, although it would be advantageous if a common approach could be taken across the NEM and regional auctions take place simultaneously. This would also facilitate the issuance of ST inter-regional access.

Specifically, it would be necessary to decide:

- the structure and term of the ST access products: these would have to be able to be settled by AEMO and be useful to generators in backing their forward contracts: a quarterly term, with baseload and peaking structures, is suggested;
- the amount offered and any reserve price associated with it: although clearly the total amount issued, when added to existing firm access, would not be permitted to breach FAS limits;
- the flowgate constraints and capacities that would be applied in the auction to ensure that total the ST access issued would not cause any breach of FAS; and
- any credit requirements applying to auction bidders.

As discussed in section 8.2.5.2 (Short-term access incentive scheme), TNSP would have an incentive to maximise the revenue from ST access sales and so would ensure that the ST access product and auction would be attractive to generators.

6 The latest expiry date.

7 Access auction mechanisms are discussed in section 12.6 (Firm Access Allocation and Auctions).

8 A TNSP could be permitted to undertake unregulated capital expenditure but would not be allowed to recover the cost of this through its regulated revenue.

The use of a reserve price would allow a TNSP to sell ST access that had some operational cost associated with it: or, more exactly, some cost associated with increasing transmission capacity sufficiently to accommodate the new ST access with FAS. For example, it might cost a TNSP $10/MW/quarter to increase capacity on a flowgate. If this flowgate capacity is offered into the auction with a reserve price of $10 then:

- if the capacity is sold, it would earn at least $10/MW, sufficient to cover the cost of providing it; and
- if the capacity is unsold, the additional flowgate capacity is not required and so no operating costs are incurred.

The design and operation of the ST auction is discussed in more detail in section 12.6 (Firm Access Allocation and Auctions).

7.2.5 Secondary Trading - Bilateral

There may be occasions where a generator wishes to transfer all or part of its existing agreed access to another power station: whether one from within its portfolio or one belonging to another generating company. This is referred to as *secondary trading*. This may be done through two mechanisms:

- a bilateral agreement between the two generators, approved by the relevant TNSP; or
- the ST auction described in the previous section.

The bilateral process is described below. The auction process is described in the next section.

There are three potential changes, which may occur individually or in combination, arising out of a bilateral transfer of access:

- a change to the *power station* that the firm access applies to;
- a change to the *node* from which firm access applies, with possibly different participation in congested flowgates; and
- a change to the *generating company* who is the agreement counterparty and is obliged to make annual access payments.

TNSP approval criteria are summarised in Table 7.3.

Table 7.3 TNSP Approval Criteria for Bilateral Trading

Access Change	Approval Criteria
Change of Power Station	No approval required
Change of Node	Subject to no increase in current LRIC
Change of Company	Subject to agreed prudential arrangements

If the bilateral transfer is simply to a different *power station* connected at the same transmission node and owned by the same company, the impact on the TNSP is unlikely to be material (since the FAS obligations would remain the same) and granting permission should generally be straightforward.

If a bilateral transfer were to a different *node*, the TNSP would need to examine the FAS implications: target capacity might increase at some flowgates and decrease at others. One possible approach the TNSP could take would be to calculate the access price at the two nodes and approve the transfer only if the price at the new node was no higher than that at the old node.[9] If the price was higher, the amount of agreed access provided at the new node could be scaled back as needed to equate the two charges.

Irrespective of the calculated current access prices, the contractual access payments – reflecting the historically-calculated access prices - would remain unchanged. If the access transfer led to a requirement on the TNSP to change its transmission expansion plans, the transfer could be delayed by the TNSP to give reasonable time for this to occur.

If a bilateral transfer were to a different *generating company*, that company would acquire the obligation to make any *future* access payments specified in the access agreement; if only part of the access were transferred, payments would be shared pro rata between the two generating companies. The TNSP would need to establish prudential arrangements to ensure that these payments are made. Given the possibly different credit profile of the new counterparty, the new prudential arrangements may differ from those applying previously.

Irrespective of approval requirements, any trade would need to be notified to the relevant TNSP so that AEMO can then be formally notified. A TNSP would be entitled to levy fees on bilateral trades to recover reasonable costs incurred.

Any payments agreed between the two generating companies would be a purely bilateral matter and not relevant to the TNSP or subject to any rules or regulations.

9 Strictly speaking, the access price at the old node should be calculated using a Long Run Decremental Cost (LRDC) approach: ie the cost *saving* to the TNSP of no longer having to provide the access at that node.

7.2.6 Auction-based Transfers

Generators would be permitted to offer ST access into the auction arranged by the TNSP for selling *its* ST access, as described in section 7.2.4 (Short-term Firm Access Issuance) above. A generator would be able to offer any amount up to its existing agreed access level and at any reserve price.

If a generator sold some of its access in the auction, it would receive payment through the TNSP's auction settlement process. There would be some associated *purchase* of ST access by another generator, but the nature of the auction is that there is no explicit correspondence between buyers and sellers: ie a generator A cannot *buy* from generator B; rather generator A buys from the auction and generator B sells into the auction. Notwithstanding that a firm generator may have sold some of its firm access into the auction, it would still be required to make the same contractual payments to its TNSP.[10]

Generators selling into the auction will be competing with the TNSP for ST access sales. Other things being equal, the party with the lower reserve price has a higher chance of successfully selling its access.[11]

Given that the TNSP ensures - through the auction design - that auction trading does not have any adverse FAS or credit implications, no explicit TNSP approval of auction-based trading is required. A TNSP is entitled to levy auction fees to cover reasonable costs.

7.2.7 Summary

Deciding on the appropriate set of service parameters for firm access will be complex, especially for a new entrant generator who would be, in parallel, arranging to construct and fuel a new power station. Access pricing will be regulated and transparent, but its complexity nevertheless means there may be several iterations before a generator discovers its preferred service. The procurement process must ensure that a TNSP provides useful and timely information and advice to assist the generator with this discovery.

10 Note that this differs from the situation following a bilateral sale, where the purchaser acquires the payment obligation.

11 Although there can be no *exact* comparison between the TNSP and generator offers: because the TNSP is selling additional flowgate capacity, whereas the generator is selling (nodal) agreed access.

7.3 Design Issues and Options

7.3.1 A Procurement Example

An illustrative example of the procurement process is presented below.

An investor is planning to build a gas-fired power station. It has decided on the field from which it will buy the gas, and the pipeline on which it will be transported. However, it has a choice of pipeline off-take points and could potentially build a dedicated lateral to a site away from the pipeline.

Correspondingly, the investor has a choice of connection points to the transmission system, could commission a new connection point on an existing line, or could potentially build a dedicated transmission extension to a site remote from existing transmission lines.

Notwithstanding these options, the power station is likely to lie within a given *zone of interest*. The investor would review prices published in the TNSP access information report for relevant nodes and this might narrow down the choice somewhat. It would then, in stage 1, go to its TNSP to discuss access pricing as well as connection and extension practicalities and pricing. If the pricing model is available to generators for informal use informally, it is likely that the investor would make use of this facility.

The TNSP may highlight a number of relevant issues:

- *pending access requests* in the zone of interest which may affect access prices, depending upon whether or not they proceed;
- *transmission expansion plans*, which might also affect prices: since they affect the level of initial spare capacity;
- *elements with limited capacity in the zone of interest*: typically, it will be these local elements which will cause the biggest *variations* in access prices between different nodes in the zone of interest; and
- *breakpoints associated with this limited capacity*: eg up to 500MW could be accommodated without expansion; anything higher would prompt lumpy expansion which would increase access prices (obviously, this is most relevant where forecast flow growth on the element is low).

These issues – and the associated access price variations – will feed into the investor's decision-making, which will of course also be affected by site-specific costs and consents, gas supply costs and transmission connection and extension costs. Marginal loss factors may also be relevant.

Extensions to a node will generally be radial and not affect flows on the transmission network (compared to a direct connection to the node) and therefore not affect access pricing. On the other hand, a new connection point on an existing line would need to be explicitly modelled in the access pricing model at some point. However, the price difference from existing nearby connection points on the same line might be low and so not material: at least until stage 2 or 3.

Following the stage 1 discussions, the generator may settle on a proposed site for the power station and would then enter stage 2 to get a provisional access price(s) for the node(s) associated with that site, perhaps for different levels of agreed access.

Stage 3 would then probably not be entered until the site had been purchased, consents received or well advanced, and the project approaching financial close. That is because the firm offer made by the TNSP in stage 3 is time-limited and the generator would not be in a position to accept it and commit to it unless well advanced on all other stages of the project.

Stage 4 completion would occur at the same time as completion of other key components of the project (construction contract, gas supply contract etc). The TNSP would then commence undertaking the necessary expansion in parallel with the investor constructing the power station. Firm access would typically be scheduled to commence at the same time as power station commercial operation, although commencement might be delayed in some cases where there is a long lead time for the necessary transmission expansion.[12]

7.3.2 Form of Agreement

It is proposed that the basic terms of the agreement are as follows:

- the TNSP agrees to provide the agreed amount of firm access service; and
- the generator agrees to make the annual access payments.

The first term means in practice that the TNSP agrees that it will *notify* the agreed access amount (and the relevant generator node) to AEMO for each settlement period over the agreement term. AEMO will then use this notified amount in determining access settlement amounts in accordance with access settlement formulae which are described in the rules. In this respect, the role that AEMO plays is similar to its role in reallocation transactions[13] although more complex. It is not AEMO's role to question or verify the notified amounts.[14]

In making this notification, the TNSP discharges its explicit obligations to the generator. The consequential obligations to maintain service quality (through the FAS)

12 A generator could still access the network on a non-firm basis in the interim.

13 Where market participants notify the amount to be reallocated and AEMO adjusts settlements accordingly.

14 A generator could dispute the notified amount, through the normal settlement dispute process. In resolving the dispute, AEMO would need to refer to the firm access agreement to confirm the agreed access amount. To facilitate dispute resolution, agreed access amounts should be recorded in the agreement in standardised form which is lodged with AEMO in advance of access commencement.

are with the AER, not the generator and so the firm access agreement should not refer to FAS.[15]

The nature (as opposed to the volume) of the firm access service is predicated on the rules and on AER regulation. These may change from time to time and such changes are outside the control of the TNSP. Therefore, the agreement *cannot* contain any obligation on the TNSP to maintain service definitions or standards over the term of the agreement.

Prudential obligations of generators (firm and non-firm) in *access settlement* will be managed by AEMO, through a continuing AEMO obligation to manage prudential requirements across the whole of NEM settlements.[16] Therefore, the firm access agreement does not need to address this.[17]

Provisions of the firm access agreement associated with making annual access payments are likely to be similar to corresponding provisions in connection agreements where, similarly, annual payments must be made over the life of the agreement. Where there is a default on payment obligations, a TNSP would probably be relieved of its obligation to make continuing notifications of agreed access amounts to AEMO and of its corresponding FAS obligations.

7.3.3 Changes to standard forward contract peak/off-peak times

It is anticipated that the standard peak and off-peak periods which would apply to firm access agreements would be aligned with the convention used in forward contracts. These are agreed between market participants under the aegis of AFMA, which is beyond the jurisdiction of both TNSPs and the AEMC.[18] Although the rules, or firm access agreements, could easily reference this convention, this could create a governance issue in the future, where AFMA proposed to change the convention and this impacted adversely (through its link to firm access) on TNSPs or the NEM in general.

Therefore, it is proposed that peak period definitions are contained in the firm access agreement and then reflected in AEMO settlement procedures. Should the AFMA convention change and a generator wish to change the period definitions in existing access agreements, it would need to negotiate a change to the agreement with its TNSP. TNSPs may also decide to change (or may be required by the AER to change) the default period definitions for future agreements to align with the new AFMA standard.

---

15 To avoid a double jeopardy situation, where a breach of the FAS leads to a breach both of the regulations and the agreement.

16 Although access-short generators will be making payments into access settlement, taking into account other AEMO payments to generators, generators will continue to receive, net, monies from settlement.

17 Nor could it, because non-firm generators – whose prudential obligations must also be managed – do not have an access agreement.

18 Strictly speaking, AFMA only recommends non-binding standards but, in practice, all companies use the AFMA standards when agreeing forward contracts.

7.3.4 Agreed access not limited to registered capacity

The concept of super-firm access is discussed in chapters 4 (Access Settlement) and 5 (Firm Access Standard) and relies on a generator procuring an access amount greater than its generation capacity. The concept is a logical and straightforward way for a generator to obtain access that is firmer than the FAS standards. However, it does raise possible competition issues around hoarding capacity.

There is a potential scenario whereby a generator might deliberately buy up existing *spare* capacity, above its current needs, so as to pre-empt the use of that capacity by a future competitor. The future entrant would therefore be required to pay for transmission capacity to be expanded and this may create a barrier to entry.

However, this scenario is mitigated in a number of ways. Firstly, the LRIC methodology takes a long-term view of the level of spare capacity and its value. Spare capacity may not be substantially discounted in price compared to new capacity: unlike, say, under a deep connection charging approach, where the former has zero price attached.

Furthermore, and most importantly, whilst a generator can hoard *firm* access it cannot hoard access, *per se*. If the firm generator does not actually use the spare transmission capacity (and it obviously cannot generate above its generation capacity) the relevant flowgates will either remain uncongested or will have sufficient capacity for non-firm entitlements to be allocated. Therefore, a generator seeking to hoard firm access may simply be encouraging future entrants to free-ride by relying on non-firm access. Thus, hoarding is unlikely to be a tenable strategy, except perhaps for a short period: eg where a generator is planning to build a new power stations and wishes to "get in first".

In summary, our current view is that there seem unlikely to be any competition concerns in relation to super-firm access or access hoarding.

7.3.5 Customisation

Customisation of access terms would be permitted, subject to this customisation not impacting adversely on other users.[19] For example, the TNSP agreeing to a heavily-discounted access charge would imply correspondingly higher TUOS prices to recover the revenue shortfall and so would not be permitted.[20] On the other hand, a discounted access charge might be allowable in some situations, similar to discounting

19 Meaning all users of transmission services, not just firm generators. However, non-firm generators are not considered to be users as they do not have any agreement with the TNSP. And, of course, many new firm access agreements will – quite appropriately – impact adversely on some non-firm generators.

20 See chapter 8 (TNSP regulation).

of TUOS charges.[21] The regulation of customisation to ensure the TNSP complies with this principle is discussed in section 8.2.3 (Pricing regulation).

Since the agreement refers to only two matters (the TNSP notifying agreed access amounts to AEMO and the generator making annual access payments) only terms associated with these two matters can logically be customised.[22]

Customisation of agreed access amount might include variations such as:

- non-standard peak-period definitions;
- a call option structure: where access was only provided where RRP exceeded a threshold amount;
- scheduled outage windows: the agreed access amount could be set to zero for a period agreed each year between TNSP and generator; and
- future agreed access amounts being contingent on transmission expansion being completed.

Since some of these variations could make it more complex and expensive for AEMO to undertake access settlement, the TNSP would need to ensure that these variations did not adversely affect other users. Similarly, where a non-standard access price accompanied these variations, the TNSP would need to ensure that this reflected real changes to the cost of service provision, to ensure that it did not lead to costs being loaded onto other users.

Variations to annual access payments could include:

- discounts or premiums to the default access price to reflect a lower or higher service level;
- an upfront *capital contribution*;
- a change to the annual payment profile;
- a change to the indexation of the annual amounts; and
- different approaches to prudential management: eg parent company guarantees.

Again, the principle of *no adverse impact on other users* applies.

Customisation relating to anything apart from these two matters would not be permitted: for example (without limitation):

---

21 TUOS price discounting is currently permitted where this discounting provides benefits to users as a whole: eg by discouraging transmission bypass.

22 A TNSP would not be permitted to include terms not relating to these matters. For example, if a generator agreed to provide network support to the TNSP, this should be recorded in a separate agreement, not in the firm access agreement.

- options to renew or extend the agreement term;
- variations to the FAS;
- performance incentives or risk sharing; and
- bundling with other services (eg network support or network control ancillary services).

7.3.6 Recognising Transmission Expansion Lead Time

Transmission expansion will commonly be required prior to new firm access commencement to ensure no breach of the FAS. Expansion often has a long lead time and this needs to be reflected in the access procurement process: a generator whose access requirements necessitate expansion will need to provide the requisite time between procurement completion and access commencement.

On the other hand, a TNSP is not permitted to refuse or delay access because lumpy expansion may be underutilised in the medium-term: for example, where a 250MW access request prompts a 500MW expansion and the TNSP seeks to delay access until a second 250MW access request is received. This underutilisation is reflected in the access price and the TNSP can expect to recover the full cost of the expansion over time.

7.3.7 Firm Access Procurement for Embedded Generation

A generator procures access from its local transmission node: the connection point at which its power station connects to the shared transmission network. This raises the question of where an embedded generator (a generator connected to a distribution network) would buy access from.[23]

AEMO does not typically model distribution network constraints in NEMDE. If there were such constraints, they would be agreed between the generator and the distribution network service provider (DNSP) and reflected in the generator's dispatch offer.[24] If the generator required certainty on how it would be affected by distribution constraints it would need to enter into some arrangement with the DNSP, which is beyond the scope of the OFA model.

Where the output of an embedded generator materially affects transmission congestion, AEMO would include its output on the LHS of a transmission constraint and so it would have flowgate participation and usage, just like a transmission-connected generator. In formulating the constraint, AEMO would need to

23 Only scheduled embedded generators would need to consider procuring access, because non-scheduled generators are not dispatched by AEMO and so cannot be constrained off by transmission congestion.

24 For example, if the constraint meant that a 50MW generator could temporarily only generate at a maximum of 30MW, the generator would reduce its offered availability to 30MW.

know the node at which the embedded generator's output entered the transmission network. So long as this node did not change over time, an embedded generator could procure firm access at this node and it would receive the necessary network access to the RRN, just like a transmission-connected generator.

If the node did change – or if AEMO assumed that the entry of the embedded generator output was shared across multiple nodes – procuring firm access would be more complex. The embedded generator would need to enter into discussions with its DNSP, AEMO and the local TNSP to resolve this issue.

7.3.8 Mezzanine ST Access

Firm generators may be concerned that ST access issuance by TNSPs may degrade the firmness of their (conventional) access, particularly as a TNSP is incentivised to maximise the revenue from ST access sales and so may be tempted to oversell ST access: eg by being over-optimistic on flowgate capacity.

There are two possible approaches to addressing this concern:

- ensure that a TNSP fully compensates firm generators for any degradation in firmness caused by ST access sales: this is discussed in section 8.2.5.2 Short-term access incentive scheme; and/or
- make ST access *junior* to conventional access in the entitlement allocation process: explained below

As described in section 4.2.3 (Target Entitlements), non-firm target entitlements are *junior* to firm target entitlements in that they are scaled back first. Thus, there are two tiers of seniority in the entitlement scaling process, but there is no algebraic reason why there could not be more.

Specifically, it would be possible to give ST access a *mezzanine* firmness, meaning that it would be scaled back *before* conventional firm access entitlements but *after* non-firm entitlements. This would mean that, if a TNSP did oversell ST access, this would impact only on the firmness of ST access and not on conventional access. However, because the TNSP wishes to maximise revenue from ST access sales, and because generators will bid lower prices if they are concerned about the firmness of the ST access product, TNSPs have an incentive to ensure that ST access is not oversold.

The potential advantage of the mezzanine approach is that ST access sales may be higher, since generators, who would bear the risk of any non-firmness, will generally be less risk averse than TNSPs, who bear the risk in the former approach. On the other hand, the approach entails an increase in the complexity of the model, specifically in access settlement. It is not proposed at present, but could be considered as a future addition.

7.3.9 Grouped Procurement

Generators could be permitted to procure access in a *group* rather than individually. Access would be awarded to each group member in the same way as if they had applied individually. However, the TNSP would calculate a single access charge, based on the cost of the group's aggregate access. The group members would need to reach agreement on how to divide up this charge between themselves and then agree – in their access agreements – to make their respective payments to the TNSP.

The major benefit of grouping to a generator would be to share the cost of a lumpy expansion, where the LRIC calculation attributes most of this charge to new generator.[25] In that situation, the first generator may be the instigator of the grouping.

The benefit to a TNSP of grouping is that it avoids the difficulty of managing concurrent and inter-related access applications.

7.3.10 Information on Access Firmness

The multi-tier FAS proposed in the Technical Report (August 2012) gave a new generator an indication of the access firmness that a generator would be entitled to under different NOC tiers. Under the single-tier FAS now proposed, a generator is entitled to full access in NOC but receives no minimum guaranteed access under AOC.

Nevertheless, a TNSP should be able to give new generators an indication – at least for the medium term – of the level of firmness that might be expected under AOC. Given its location, the flowgates most likely to constrain a generators level of access under outage conditions are likely to be well-known and understood by the TNSP. Issuance regulations should place obligations on TNSPs to communicate this information to new generators in an accessible form.

25 Eg because additional use of the asset by future generation is not anticipated in the access pricing forecasts.

8 TNSP regulation

8.1 Overview

Firm access service is provided by the shared network, the operation of which is a natural and regulated monopoly. Since TNSPs are therefore monopoly providers of firm access service, the service is treated as a **prescribed service**. Regulation covers four areas: *issuance, pricing, revenue, and quality*. These are considered in turn.

Issuance regulation requires that, through the access procurement process, TNSPs offer the default service at a default price (determined by the access pricing methodology), provide timely and relevant information to allow a generator seeking firm access to choose its preferred service parameters, and negotiate and agree customised variations from the default in good faith, to the extent that these variations do not adversely affect other users. TNSPs would be able to specify the earliest date that the access term could commence (withim limits), to give time for necessary network expansion.

Pricing regulation requires that default prices for firm access should be calculated using the approved pricing methodology, consistent with LRIC pricing principles and requirements set out in the rules. The pricing methodology would be developed during implementation, should the OFA model proceed.

Revenue regulation will require that the *combined* revenue from TUOS services and firm access services are not forecast 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.

Quality regulation provides incentives for TNSPs to maintain access service quality at or above the minimum standard specified in the FAS. Incentives would be provided through financial penalties on the TNSP where breaches occur. Penalties will be based on – and not exceed - the cost to firm generators of shortfalls of flowgate capacity below the FAS standard. Through access settlement, payments by the TNSP will be allocated directly to the generators affected.

8.2 Design blueprint

8.2.1 Firm access service regulation

Firm access is a service provided by a TNSP using its shared network. Because the shared network is a natural and regulated monopoly, TNSPs are monopoly providers of firm access; the service cannot be provided on a competitive basis. For that reason, the service is treated as a *prescribed service* for the purposes of regulation, similar to TUOS service.

Regulation will apply to four areas:

- issuance;
- pricing;
- revenue; and
- quality.

As with TUOS service, regulatory principles and processes will be defined in the rules and these will be applied and operated by the AER.

These areas are discussed in turn below.

8.2.2 Issuance regulation

Regulation will require that TNSPs develop and operate an access procurement process similar to that described in chapter 7 (Access Procurement). TNSPs will be required to provide access seekers with timely and useful advice and information on access procurement.

As discussed in section 7.3.5 (Customisation), TNSPs are permitted to agree variations from default firm access terms and prices, so long as these are in the interests of all users and relate to agreed access amounts or to access payments.

This requirement will be overseen and enforced by the AER through:

- the AER specifying customisation principles, in accordance with high-level principles in the rules;
- TNSPs being required to develop an approved customisation policy and offering and agreeing variations in accordance with that policy; and
- the AER reviewing and approving the policy based on its compliance with the customisation principles.

The envisaged process is similar to that which applies to discounting of TUOS charges in relation to the TUOS service to demand-side users.

TNSPs would be required to operate a dispute resolution mechanism and those procuring access would refer disputes to this forum if they felt the TNSP was not meeting its obligations under its procurement procedure or customisation policy.[1]

1 This could be an existing mechanism or one especially created for access procurement.

8.2.3 Pricing regulation

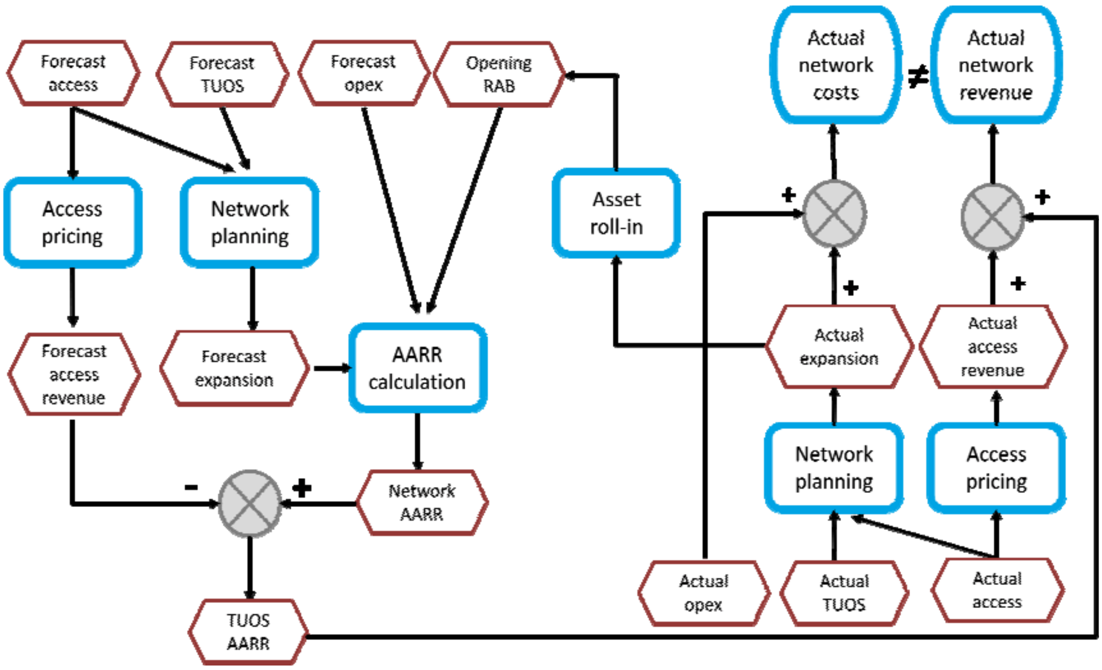
A TNSP will offer default prices for firm access which must be calculated using the access pricing methodology (see chapter 6 (Access Pricing)). As proposed in that section, there will be a single, common pricing methodology for all NEM regions.

In accordance with customisation principles, discounts could be permitted where the TNSP is able to demonstrate that these are in the interests of all users.

8.2.4 Revenue regulation

There will be a single regulatory framework for revenue from firm access service and TUOS service (which, collectively, will be referred to as shared network services). The regulatory framework for share network service revenue is presented in Figure 8.1 below.

Figure 8.1 Revenue regulation processes



8.2.4.1 Revenue allowances

The AER will determine, on each regulatory reset, an aggregate annual revenue requirement (AARR) for shared network services over a regulatory period. The AARR would be based on the efficient cost of building, owning and operating a shared network capable of providing shared network services to the relevant standards[2] based on current and forecast levels of these services.

This comprises the AER setting both capital and operational expenditure allowances. The capital and operating expenditure associated with investments to meet reliability

2 Ie for TUOS and firm access services, the reliability standard and the FAS, respectively.

standards would be set as it is currently and so reliability investment expenditure is not considered any further.

Ex ante revenue allowances provide a strong incentive for TNSPs to minimise their costs over the regulatory period since TNSPs are able to profit by spending less than their allowed revenue allowance. Ex ante revenue allowances also provide incentives for TNSPs to reduce their overall costs by making trade-offs across their network and prioritising projects.

The revenue allowance would take into account *committed* firm access. Typically, most access agreements, being long-term, will have been entered into prior to the regulatory period. In the next regulatory period the *actual* project cost of the investment would form part of the RAB, subject to an ex post efficiency review. If an operational decision is reached (or network support agreement entered into) then the economic regulation arrangements would ensure that the TNSP will have continuing access to a separately set operating expenditure allowance.

8.2.4.2 Recovery of revenue allowances

TNSPs recover their costs (as set in the revenue allowance) incurred in building and operating its transmission system from customers within its region. This occurs through recovering transmission use of system (TUOS) charges from customers. In order to calculate TUOS following the introduction of the OFA model, the amount of revenue expected to be received from providing firm access would be estimated and, by subtracting estimated access revenue from the allowed annual revenue requirement, a cap on the TUOS revenue to users of load services would be derived. Aggregate revenue from firm access sales would not be capped; instead, firm access *prices* would be regulated.

In other words, firm access revenue for the period would be estimated, based on current and agreed future access agreements (ie signed agreements), and *deducted* from the network AARR to determine a revenue cap for TUOS. TUOS prices will then be determined as they are at present, and subsequently adjusted as necessary to ensure that the TUOS *revenue cap* is not breached.

On the other hand, access revenue is not explicitly regulated through this mechanism: a TNSP may earn higher revenue than forecast if additional, unanticipated, access issuance occurs.[3] The additional access charges received through the sale of higher than forecast levels of firm access would provide TNSPs with a broadly appropriate amount of revenue to cover the additional costs.

If the actual volume of firm access sales within a regulatory control period was less than forecast at the time of the revenue determination, the TNSP would recover less revenue: the TUOS cap would prevent the TNSP from recovering the revenue shortfall from demand-side users. This would be appropriate since the TNSP's costs would be

3 Of course, it cannot increase prices on pre-existing agreed access, since prices are fixed prior to access commencement.

correspondingly lower. Alternatively, if access sales are higher than expected, access revenue will be higher and, since TUOS revenue is fixed, total revenue is higher. Again, this is appropriate, because the TNSP's costs are likely to be correspondingly higher.

Access pricing is designed to ensure that incremental access revenue and costs are *broadly* matched, but they will not *exactly* match and some risk will be borne by the TNSP.

There could be circumstances of revenue uncertainty to the TNSP during a regulatory control period where additional costs to the TNSP were either substantially higher or lower than the extra revenue: for example, where a particularly large expansion was required. There may also be a misalignment between when revenue is received and when costs would be incurred by the TNSP.

The AER may include regulatory mechanisms that mitigate this risk for a TNSP. This issue is discussed further in section 8.3.2 (TNSP risk from lumpy access-driven expansion costs).

In the absence of such a mechanism this uncertainty would be borne by the TNSP - increasing its risk. However, this risk is removed in the following regulatory period since the actual project would be rolled into the TNSP's RAB.

8.2.5 Financial quality incentives

The structure of the incentive schemes is an important component. In deciding whether or not to seek firm access, generators will consider the likelihood of the following occurring (when within the firm access standard):

- access curtailment; and
- receipt of compensation (both from other generators, and also the TNSP) if curtailment did occur.

We recognise that the development of incentive schemes is a complex, and sometimes lengthy, process. However, having effective incentives on TNSPs to maintain access service quality at or above the minimum standard specified in the firm access standard is an important feature of the model. Therefore, some incentives should be in place immediately or very soon after the introduction of the OFA model. We have proposed two low-powered incentive schemes that should apply, discussed below:

- operational incentive scheme; and
- short-term access incentive scheme.

8.2.5.1 Operational incentive scheme

As discussed in section 5.2.2 (FAS Definition), the FAS defines a level of target flowgate capacity that a TNSP must meet or exceed on every congested flowgate in every settlement period under normal operating conditions (NOC). The target flowgate capacity will be calculated by AEMO and any *capacity shortfalls* will be *identified* and *reported on*.

A capacity shortfall causes firm generators' entitlements to be scaled back and consequently their settlement payments will be reduced by an amount that – in aggregate – equals the product of the capacity shortfall and the flowgate price: the *shortfall value*. Through access settlement, payment by the TNSP associated with this would be allocated to generators.

The TNSP payment would be equal to some proportion of the costs to firm generators resulting from the breaches which would be achieved through the application of a sharing factor:

TNSP penalty = incentive sharing factor × shortfall value

The design and timing of any financial incentive scheme for FAS breaches would be decided by the AER. In the context of using the above formula, the AER would be required to set TNSP penalties.

The AER would set a sharing factor, ie "X" between zero and 100 per cent - based on an appropriate process as set out in the rules. "X" would likely start off low, and increase over time: sharpening incentives on TNSPs as they become more familiar with providing access. A series of principles to be followed in setting "X" would be developed. They would be contained in the rules in order to guide the AER when setting this sharing factor.[4]

This fixed sharing factor would apply until the aggregate penalties reached a predetermined limit (or "cap"), after which the sharing factor would be set to zero so that no further penalties would apply to the TNSP. The cap would also be set by the AER, and would be defined relative to a period of time – an annual basis might be appropriate.

A longer time period may result in too high an overall risk exposure, ie a TNSP reaching the cap early in the incentive period. This would mean it would not face incentives for the remainder of the period.

It may also be useful to define a cap by location or node to ensure that one breach does not exhaust the cap, eg where the cap was only based on a timescale, and a high-priced event occurred at one node resulting in the cap being reached. If there was no "location" cap, the TNSP in that region would not have to contribute to any other

4 The principles would be similar in nature to those principles specified in NER clause 6A.7.4(b) relating to the service target performance incentive scheme. This includes setting boundaries on the range of possible values that "X" could be.

shortfalls for the remainder of the period. If there was a "locational" cap, the TNSP would still be responsible for, and so incentivised, to provide firm access across the remainder of the region during that time.

A cap is necessary due to the high market price cap in the NEM. The market price cap is the maximum price at which generators can offer into the market, and is currently $12,900/MWh. It would only take relatively few shortfall periods at the market price cap for a TNSP to be facing, and responsible for, a large settlement shortfall.

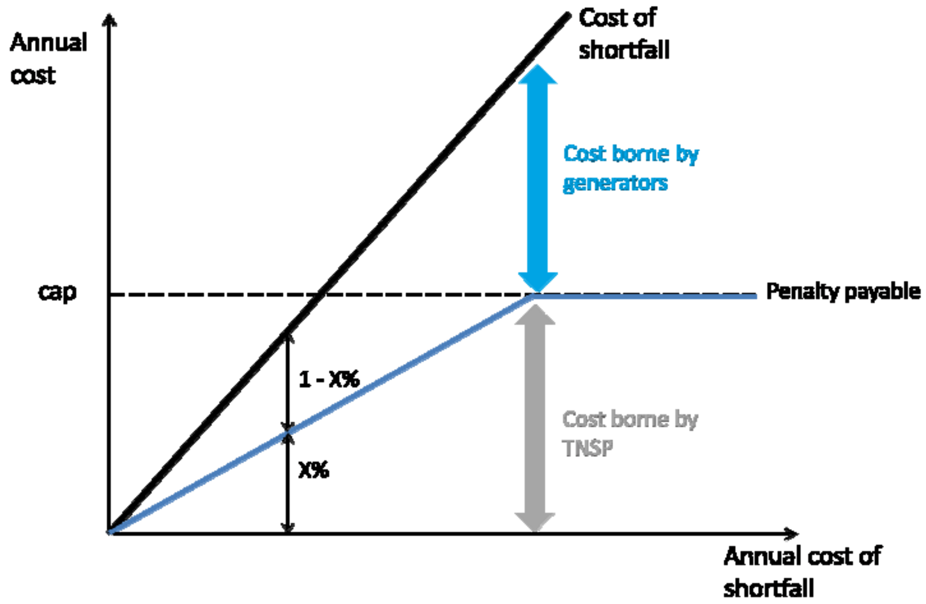
There are a number of principles that should be used by the AER in setting the cap. Specifically the cap should be set having regard to:

- the financial position of a benchmark-efficient TNSP;
- the impact of the risk on a benchmark-efficient TNSP, and its required rate of return to compensate for that risk; and
- the creation of sufficiently strong incentives for the TNSP to deliver firm access as efficiently as possible.

These principles would be included in the rules in order to guide the AER. We note that consideration would also need to be given to the interaction of caps set here, with the liability caps on TNSPs that exist through immunities in favour of NSPs under the NEL.

Figure 8.2 illustrates the operational incentive scheme. The cost of shortfall would be shared between TNSPs (X per cent) and generators through the scaling back of entitlements (1-X per cent) up until the cap was reached. Once the cap has been reached, the full cost of the shortfall is borne by generators (ie 100 per cent) – with the TNSP not liable for any further shortfall.

Figure 8.2 Operational incentive scheme



The penalty would be calculated and applied in access settlement whenever a capacity shortfall occurred, by calculating a TNSP support amount on the relevant flowgate based on the formula:

> TNSP support = incentive sharing factor × capacity shortfall

The TNSP would then be obliged to pay a penalty amount into access settlement based on the formula:

> TNSP penalty$ = flowgate price × TNSP support

The TNSP support *adds* to the *effective* flowgate capacity, meaning that the total effective flowgate capacity is:

> Effective Flowgate Capacity = Actual Flowgate Capacity + Flowgate Support + TNSP Support

As described in section 4.2.4 (Entitlement Scaling), the effective flowgate capacity is allocated between generators using the entitlement scaling algorithm. Thus, actual entitlements to firm generators are enhanced by the TNSP support and the TNSP penalty payment is automatically allocated between firm generators affected by the capacity shortfall.[5]

So long as the incentive sharing factor is less than 100 per cent, the FAS *obligations* on a TNSP to maintain flowgate capacity must remain. However, if at any time a 100 per cent sharing factor were applied,[6] firm generators would be indifferent to any capacity shortfalls, as they would be fully compensated. It could then be left to a TNSP to decide whether to maintain the FAS level of capacity or to pay the shortfall penalties.

8.2.5.2 Short-term access incentive scheme

As discussed in section 7.2.4 (Short-term Firm Access Issuance) a TNSP could release short-term access where it could create additional capacity on the network. The associated incentive scheme is discussed below.

The aim of offering short-term access is to encourage TNSPs to undertake operational actions to promote the most efficient use of the network. In order to incentivise TNSPs to take steps to maximise the available capacity, they would retain 100 per cent of the revenue associated with the sale of short-term access.

However, in order to balance the upside that exists through this scheme - TNSPs would be 100 per cent exposed to any shortfalls that result from not providing

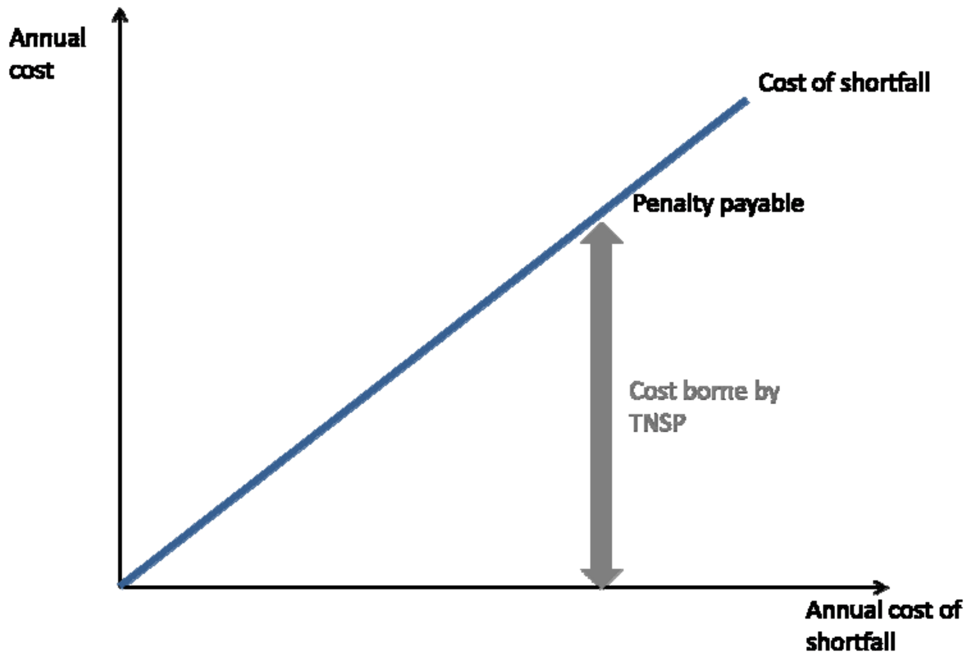
5 Since access settlement always balances.

6 At this point in time, this would appear likely to create unacceptable risk for a TNSP but, in time, TNSPs may find ways to manage such risk.

short-term access that has been released. This exposure would not be subject to any cap.[7]

The short-term access incentive scheme is illustrated in Figure 8.3.

Figure 8.3 Short-term access incentive scheme



8.2.6 Summary

Regulation will be designed to ensure that TNSPs, being monopoly providers of firm access service, will provide and price firm access in a way that promotes generator choice and market efficiency.

8.3 Design issues and options

8.3.1 Regulated or negotiated service

The firm access service has some similarities to negotiated services such as connection services, particularly in the procurement process: ie the service is agreed through bilateral discussions between TNSP and user, and agreed terms are enshrined in a long-term agreement between the two.

7 There is a possibility that this may be perceived as being inconsistent with the first revenue and pricing principle (TNSPs must be given reasonable opportunities to recover at least the efficient costs it occurs) as set out in the NEL; however, this is arguably addressed by the second revenue and pricing principle (a regulated network service provider should be provided with effective incentives in order to promote economic efficiency).

However, the firm access service, like the TUOS service, is provided by the shared network and these can only be provided by the TNSP.[8] It is impractical to differentiate between assets providing the firm access service and those providing the TUOS service. Therefore, the firm access should be treated as a prescribed service, as the TUOS service is.

8.3.2 TNSP risk from lumpy access-driven expansion costs

8.3.2.1 Capital costs and carrying costs

Errors in forecasting firm access levels, coupled with lumpy expansion requirements, could lead to a TNSP undertaking significantly more capital expenditure than anticipated in the regulatory reset process. Some of this will be recoverable through revenue from new access agreements; some of it may be borne by the TNSP, as discussed below.

It is an existing recognised feature of TNSP regulation (of TUOS revenue) that TNSPs bear the *capital* cost of transmission expansion and hold the network assets on their balance sheets, financed by equity and debt. The revenue framework then allows TNSPs to recover a *carrying cost:* a return on this asset base plus a recovery of the asset cost over time through a specified depreciation schedule and allowance. Over the life of an asset, its total recoverable carrying cost (in NPV terms) equals its capital cost, meaning the TNSP *gets its money* back in NPV terms.[9]

This feature would continue under the regulation of network services revenue, described above. Therefore, the risk to the TNSP from lumpy investment is not from the capital cost, *per se*, but rather from any mismatches between the carrying cost of the new investment and the incremental revenue from the new firm access provided.

8.3.2.2 Variances and impacts

A mismatch may potentially arise from two sources:

1. a difference between the *actual* capital cost of expansion triggered by the new access and the cost *estimated* by the access pricing methodology;[10] and

2. a *timing* difference in the recovery of the expansion cost, between the regulated carrying cost and the annual access payment profile.

---

8 In the sense of planning and operating the shared network assets collectively. A third party might construct individual assets, but that party does not provide the shared network.

9 One can think of the carrying costs as repayments on a loan. The NPV of loan repayments must always equal the loan value, using the loan interest rate as the discounting value for the NPV.

10 It should be recalled that the capital cost being talked about here is actually an incremental cost and may be associated with planned expansions being advanced, as well as new expansions.

The capital cost difference arises from a combination of two components: modelling errors or simplifications in the pricing methodology; and inefficiencies in the TNSP expansion process.

For example, if the pricing methodology estimates an expansion cost of $100m and the TNSP ends up spending $180m, the discrepancy could be due to:

- *pricing errors alone*: ie the true efficient expansion cost is $180m;
- *TNSP inefficiency alone*: ie the efficient expansion cost is $100m; or
- *a combination of both*: the efficient expansion cost is somewhere between the two.

It is not possible in practice to distinguish these components, which means that any regulatory mechanism to remove risks due to pricing errors would also remove incentives for capital efficiency, which is undesirable. Therefore it is not practically possible to mitigate pricing risk: except, of course, by carefully designing the pricing methodology so that such errors are minimised.

The *timing* discrepancy between actual costs and revenue recovery is similarly due to a combination of two factors:

- a timing discrepancy between actual and estimated costs; and
- a timing discrepancy between estimated costs and revenue recovery.

The first timing discrepancy is a function of the differences between *actual* expansion planning and the *stylised* expansion planning assumed in the access pricing methodology. The second discrepancy arises because it is not always possible to match the profile of access payments to the profile of estimated costs. This issue is discussed further in section 12.10 (Annual Payment Profiling).

A TNSP is exposed to these discrepancies only within the regulatory period in which the new access commences: the *first period*. In subsequent periods, the new physical assets from the access-driven expansion will be *rolled-in* to the TNSP's *regulatory asset base (RAB)* and remaining carrying costs will be recoverable under the network services AARR. To the extent there is a mismatch between remaining carrying costs and remaining access payments, this mismatch will then be borne by demand-side users, through TUOS charges, rather than the TNSP.

These risk drivers – and their impacts on TNSPs, firm generators and TUOS users – are summarised in Table 8.1, below: *positive* means reduced prices or improved profit; negative means increased prices or reduced profit; *none* means no effect.

Table 8.1 Transmission cost drivers and their impacts

Driver	TNSP impact (first period)	TUOS impact (subsequent periods)	Firm generator impact
TNSP expands inefficiently	negative	negative	none
TNSP expands super-efficiently[11]	positive	positive	none
Access price understates efficient expansion cost	negative	negative	positive
Access price overstates efficient expansion cost	positive	positive	negative
Expansion costs front-ended[12]	negative	positive	none[13]
Expansion costs back-ended[14]	positive	negative	none

The table reveals the following. Firstly, as one would expect, TNSP capital efficiency affects *total* costs which feed through to the TNSP and to TUOS users. It does not affect generators, whose charges are based on stylised, rather than actual, expansion costs. The other drivers do not affect *total* costs but rather how they are *shared*: so the impacts are *zero sum*.

Secondly, all of the drivers are *symmetrical*: they can operate in either direction and so associated impacts may be positive or negative, for each affected party. Ideally, the design of regulation (AARR forecasting and access pricing) should ensure these risks are unbiased, so that the positive and the negative average out over the longer term. Nevertheless, risk will remain and cannot be eliminated.

8.3.2.3 Numerical example

Table 8.2 below provides an example to illustrate these different risks. For simplicity, the amounts in this table are all NPV in the year prior to new access commencement, using the regulatory WACC as the discounting factor, meaning that total carrying cost equals capital cost.

11 Actual expansion costs are lower than would be expected from an efficient TNSP.

12 Expansion carrying costs exceed access payments in the first regulatory period.

13 Assuming that the generator is indifferent to the timing of access payments.

14 Access payments exceed expansion carrying costs in the first regulatory period.

Table 8.2 Example illustrating TNSP costs and revenues

($m)	Total revenue or cost	First period revenue or carrying cost	Subsequent periods revenue or carrying cost
actual new costs	60	12	48
access payments	50	5	45
Adverse Impact on TUOS users	+3	0	+3
Adverse Impact on TNSP	+7	+7	0

In this example, it is assumed that the new access that commences in the first period was not forecast *at all* during the previous regulatory reset and so had not been included in the AARR.

As discussed above, there are two drivers which affect prices or profits. Firstly, the access pricing methodology under-estimates the new expansion costs: $50m estimated versus $60m actual. That is due to a combination of pricing error and TNSP capital inefficiency.[15] As a result, there is a $10m shortfall between the incremental expansion costs and the incremental access revenue. The $10m burden is shared between the TNSP and TUOS users.

Secondly, there is a discrepancy between the timing of costs incurred and the timing of access payments. The table shows that only $12m of the total $60m (or 20 per cent) of carrying costs is borne in the first period, compared to recovery of just $5m (10 per cent), of estimated expansion costs. This leads to a $7m shortfall of revenue in that period, which is borne by the TNSP.[16]

After the first period, there are $48m of expansion costs remaining and just $45m of further access payments. The $3m shortfall will be recovered from TUOS users in subsequent periods.

In summary, the total $60m cost of the expansion caused by the new access is shared as follows:

- $50m borne by the new firm generator through access charges;
- $7m borne by the TNSP in the first regulatory period; and
- $3m borne by TUOS users in subsequent regulatory periods.

15 These components cannot be broken down in practice and so are not shown separately in Table 8.2 above.

16 Note that even if the pricing methodology had *correctly* estimated the expansion cost (of $60m) the same payment profiling (10 per cent in the first period) would mean only $6m of access revenue

These numbers are purely illustrative. Actual cost-sharing will depend upon the details of the access pricing methodology and the specific characteristics of the new access and associated expansion. As noted above, the risks are symmetric and (ideally) unbiased, so for every case like the example, where the generator is undercharged and the TNSP and/or TUOS users negatively impacted, there will be another case where the reverse applies.

8.3.3 Mechanisms to mitigate TNSP risk

The example above can also be used to illustrate four possible mechanisms for mitigating TNSP risk arising from discrepancies between estimated carrying cost and access revenue:

- *Access agreement front-ending:* the first-period access charges are increased (reduced) and the second-period access charges are correspondingly reduced (increased);
- *Contingent project reopener:* the TUOS revenue cap in the current period is reopened; it is increased to cover the cost discrepancy;
- *Shortfall roll-up*: part of the shortfall in the first period is capitalised, included in the RAB and so recoverable in future periods; and
- *Revenue drivers:* revenue drivers are a means of linking the revenue allowance to specific measurable events considered to influence costs, and are typically set on a dollar per unit of capacity basis.

It will be seen that all of these options transfer costs from the TNSP to demand-side users. It is not possible to transfer additional costs to firm generators since these costs are fixed by the access pricing methodology.

We favour the use of one of these mechanisms to mitigate TNSP risk - most likely the use of contingent projects or revenue drivers. These mechanisms have their advantages and disadvantages - therefore which specific mechanism should be used would need to be considered further in OFA implementation. However, we note the following:

- contingent projects are a well-established and understood mechanism in Australia, while revenue drivers may require more development in the Australian context;
- contingent projects require a project-by-project approval, whereas revenue drivers can be structured to apply more generally and so may be more useful in releasing additional revenue in response to uncertainty; and

---

would be received in the first period, leading to a $6m ($12m - $6m) shortfall being borne by the TNSP.

- contingent project, as used currently, are only to adjust revenue upwards, ie not downwards - it is difficult to see how an equivalent mechanism to adjust revenue downwards would work in practice.

The difficulty for the AER in designing such mechanisms is to prevent cost shortfalls that are caused by TNSP capital inefficiency from being passed through to demand-side users. It is possible that this could be addressed through basing the pass-through amount on modelled costs rather than actual costs.

Further, the above incentive schemes would provide strong signals to TNSPs to manage the network consistently with the way in which capacity is valued by the market at any point in time. Exposing TNSPs to even *part* of the costs of network unavailability may have a large effect on TNSP behaviour.

8.3.4 TNSP risk premium

The OFA model would likely result in a change to a TNSPs risk profile:

- Since settlement shortfalls would be based on the spot market price, TNSPs would be exposed to spot market price movement – prices may range from -$1,000 to $12,900 per megawatt hour in the space of a day.

- In addition, in the absence of any of the above mechanisms to mitigate risk, to the extent that actual project costs differ from regulated access charges in the *current* regulatory period, TNSPs would be required to bear the difference in costs. Access charges could exceed or fall short of project costs; that is, the TNSP would be exposed to both upside and downside risk.

Given the potential change in the risk profile for TNSPs, it may be appropriate to allow TNSPs to recover some form of related compensation.

Allowing compensation for changed risk as part of the regulated rate of return may be problematic: the change in risk only relates to the expenditure that is related to the firm access standard: ie not that associated with meeting reliability standards.[17] A more appropriate compensation might be allowing TNSPs to recover some level of "insurance" relating to the risks associated with firm access. If an insurance allowance was deemed appropriate to be included in a TNSP's allowed revenue, then an associated cost should also be included in the access charge, paid by firm generators, to ensure that the cost is not paid by demand-side users.

8.3.5 Sharing of congestion risks

The incentive sharing factor can *conceptually* be anywhere between 0 per cent and 100 per cent, although there may be *practical* limits as to how high the factor could be commensurate with existing TNSP regulation and risk tolerance. However, the

17 Which could include "part" of an investment, where the investment enables the TNSP to meet both reliability and firm access standards.

incentive sharing factor does not represent the percentage of *total* congestion costs that the TNSP would bear, which would be substantially lower than this, because:

- the FAS explicitly recognises (through the definition of Normal Operating Conditions) that transmission capacity reduces under certain conditions. A TNSP is only exposed where capacity falls during normal operating conditions;
- conditions which are genuinely unmanageable and unforeseeable would be classed as Abnormal Operating Conditions under the FAS, under which TNSP has no exposure;
- TNSPs would be able to specify the earliest date that the access term could commence to give time for necessary network expansion;
- the risks of extreme NEM market conditions are already mitigated through administered pricing and market suspension;
- there would be an aggregate annual cap on TNSP payments under any incentive regime;
- payments under the operational incentive scheme are diluted through the incentive factor; and
- although a TNSP faces 100 per cent exposure to shortfalls in relation to short-term access, it is not obliged to offer short-term access and, if it chooses to do so, can price it at a level commensurate with the risks involved.

These factors mean that the TNSP risk is substantially lower – and more manageable – than if the TNSP had to bear *all* congestion costs. Most congestion costs will continue to be borne by firm and non-firm generators.

In any case, the AER is likely to take a cautious and prudent approach to revising the operational incentive regime on each regulatory reset, taking into account actual congestion and transmission conditions experienced in the previous regulatory period. For example, the AER might *back-cast* the new scheme, calculating the payments that the TNSP would have made had the revised regime applied in the previous regulatory period.

8.3.6 Issues around the short-term incentive scheme

An important driver for making TNSPs responsible for 100 per cent of their costs is in order to ensure that long-term firm access rights are unaffected by the issuance of short-term access.

For example, suppose that flowgate capacity is 100MW and target flowgate capacity from long-term access sales is just 90MW. Firm generators will receive their full entitlements and will not be scaled back.

Now suppose that a TNSP sells some short-term access, so that the target flowgate capacity increases to 120MW and so there is now a 20MW shortfall. In the absence of the short-term incentive scheme, but under the operational incentive scheme, a TNSP is only liable for a proportion of this shortfall. Suppose X is 10 per cent, say, so that TNSP support is 10 per cent x 20MW = 2MW. Effective flowgate capacity is just 102MW, so all firm generators – whether holding long-term or short-term access – will be scaled back, by a factor of 102/120.

Under the short-term incentive scheme, because the 20MW shortfall is entirely due to the issuance of short-term access, the TNSP would be liable for a 100 per cent of the shortfall. In this case, then, the TNSP support will be 20MW and the effective flowgate capacity is 120MW. Therefore, all generators – long-term and short-term – will be allocated their full target entitlements.

Now consider a different scenario where flowgate capacity is just 80MW, but target flowgate capacity is still 120MW, with 30MW of this due to short-term capacity. Therefore, of the 40MW capacity shortfall, 30MW will be subject to the short-term incentive scheme and 10MW subject to the operational incentive scheme. Therefore the TNSP support amount is:

TNSP support = 100 per cent × 30MW + 10 per cent × 10MW = 31MW

Effective flowgate capacity is now 80MW + 31MW = 111MW, so all generators will be scaled back by 111/120. However, long-term generators are not made worse off as a result of the short-term access: in its absence, they would have been scaled back by 81/90MW.[18]

Arguably, the short-term access generators should not be scaled back at all, and long-term firm generators should be scaled back by 81/90 irrespective of the short-term sales. However, the entitlement scaling algorithm in access settlement does not distinguish between short-term and long-term access, although it could potentially be made to do so.

Given their unlimited exposure, TNSPs would likely be risk averse in offering short-term access rights. Short-term access would therefore only be likely to be offered when the chances of transmission capacity falling short of what is needed to provide access are small.

In section 7.3.8 (Mezzanine ST Access) an alternative mechanism is described which would place risk relating to short-term access on the holders of the access instead of on TNSPs, and still keep long-term access holders unaffected by short-term sales. This alternative mechanism may encourage short-term access sales, because generators are typically not as risk averse as TNSPs. However, TNSPs would still have incentives to preserve the integrity of ST by not overselling: generators would not want to buy low quality access.

18 Since target flowgate capacity would have been 90MW and the TNSP support amount would have been 1MW.

8.3.7 TNSPs are rewarded for providing more transmission than FAS requires

As described, the incentive schemes appears *asymmetric*: the TNSP is *penalised* for a *shortfall* in transmission but not *rewarded* for a *surplus* of transmission.

It would be possible to design a symmetric regime, where a TNSP received a bonus for providing transmission capacity in excess of the FAS standard. The question that then arises is who should pay for these bonuses. The cost should not be borne by demand-side users, who would gain little benefit from the surplus transmission capacity. The beneficiaries are generators, who will receive a higher level of network access than they are entitled to. But generators already have the option of procuring additional firm access if they value it. Why should generators then also be charged for network access that they have not requested?

In any case, it is not necessary to redesign the incentive regimes in order to provide balanced incentives. Where a TNSP provides transmission capacity in excess of FAS requirements, it is able to sell short-term access, the proceeds from which it retains as profit. Thus, its reward for overachieving reflects the market value of that overachievement and is voluntarily paid for by those generators who value short-term access.

In summary, then, the regulatory regime as described does provide for symmetrical incentives on transmission provision. This occurs without any generator being charged for access that it has not requested.

8.3.8 Existing and discretionary spare capacity

A TNSP can sell short-term access when it already has spare capacity available, or where additional spare capacity can be made available through its operational actions. Although it seems appropriate that a TNSP is rewarded in the latter case, it is less clear that it should be rewarded in the former, given that the cost of that existing capacity is borne by users, not the TNSP.

However, even if it were considered theoretically appropriate to have an incentive scheme which distinguished between these two sources of spare capacity, it is not clear that this could practically be achieved. For example, where the rated capacity of a line is increased due to TNSP operational actions, who is to say – except for the TNSP – what the base rating would be in the absence of those actions? It is not even clear how this base level of capacity might be defined. After all, if a TNSP spent absolutely no operational expenditure on a line, its capacity and reliability would soon decline, but would this low level be appropriate as a baseline, particularly when the AER includes an opex allowance in the TNSPs revenue cap?

This issue would be considered further in OFA implementation.

8.3.9 RIT-T under OFA

The general structure of the planning process will not be changed under OFA. TNSPs will still be obligated to produce the following planning documents:

- Annual Planning Reports – which are detailed short-term plans for a particular region in the NEM, developed by the jurisdictional TNSP. These set out the current capacity and emerging limitations of the network under a range of different scenarios; and
- RIT-T – this is a separate and distinct process for individual investment decisions, which examines the costs and benefits of various options and establishes the one that maximises net market benefits. This must be applied to all augmentation investments with a value of over five million dollars.

Under OFA, the main changes to TNSP planning relate to the purpose of, and the analysis undertaken by TNSPs in, the RIT-T.

Under the current RIT-T investments may be undertaken to either meet reliability standards, or to deliver a net market benefit (ie economic expansion). Under the second type of investment, TNSPs are permitted to expand transmission even when this is not necessary to meet demand-side reliability standards. This occurs where the market benefits (from a prescribed list of benefits) exceed the expansion cost. More generally, where an expansion is required for reliability reasons, market benefits can be included in the net benefit calculation, which may affect the ranking of project options.

Under OFA, TNSPs will plan to meet two standards: FAS and demand side reliability standards. TNSPs will be required to:

- meet standards at "least cost";
- include benefits that accrue to other parties in the market, aside from generators, where these are material and not disproportionate in analysis; and
- consider options in other regions, including if they have a material effect on inter-regional capacity (ie cross-regional).

The philosophy of the intra-regional access product is that generators, rather than TNSPs, decide on the economic benefits associated with the expansion. More specifically, generators decide on the benefits associated with the firm access service that prompts expansion. As detailed, TNSPs then expand as necessary purely to maintain the FAS, not (explicitly) because of the economic benefits associated with the firm access service.

Accordingly, generators decide on the economic benefits associated with expansions. Generators consider private benefits to themselves and drive FAS investments through access requests. Users (or groups) have incentives to invest if the cost (to them) of augmentation is expected to be less than the continuing costs (to them) of the constraints that would otherwise be incurred.

This approach therefore reduces, perhaps significantly, the need for TNSPs to undertake centrally planned transmission investments. In order to avoid double counting of these benefits that are internalised to generators, benefits calculated under the RIT-T should only take into account those accruing to the TNSP, or the market more generally. Consistent with the current specification of the RIT-T these benefits should only be considered if they are material, and not disproportionate to the analysis.[19]

Table 8.3 sets out the market benefits that should still be considered under OFA.

Table 8.3 Benefits to be considered in the RIT-T analysis

Benefit	Currently considered	Considered under OFA
Changes in fuel consumption arising through different patterns of generation dispatch	✓	×
Changes in voluntary load curtailment	✓	✓
Changes in involuntary load shedding (unserved energy)	✓	✓
Changes in costs for parties, other than the RIT-T proponent, due to: • differences in the timing of new plant; • differences in capital costs; and • differences in the operating and maintenance costs	✓	×[20]
Differences in the timing of expenditure ie unrelated transmission investment	✓	✓
Changes in network losses	✓	✓
Changes in ancillary services costs	✓	✓
Competition benefits	✓	×

Only those categories of costs and benefits that are pertinent to generator access procurement decisions are excluded from the RIT-T. All other categories would remain. The exception to this rule is competition benefits, which will also be excluded from the RIT-T under the OFA model, for reasons set out in the next section.

19 Under OFA, this includes that calculating the benefits does not cause delays to the access being obtained.

20 Where the changes in costs for parties are for *other* TNSPs (eg due to the investment having a material inter-network impact) these would still be included.

8.3.9.1 Competition benefits

Competition benefits are currently required to be estimated under the RIT-T. While there may be circumstances in theory where the entry of a generator could result in these benefits,[21] it is not proposed that this would continue under the OFA framework, due mainly to the difficulties that would be associated with estimating them. The reasons for this are discussed further below.

This additional dispatch cost benefit does not accrue to the new generator and so will not be reflected in its request for firm access. As a result, this competition benefit will not be picked up under the proposed OFA model, unless competition benefits are included explicitly in the RIT-T assessment.

However, under the OFA model, intra-regional investments are not required to have an overall estimated positive market benefit in order for the investment to proceed. The TNSP is required to invest in the network to meet both firm access requests and load reliability standards, at the least overall net market cost. The inclusion of competition benefits for intra-regional transmission investment would therefore not affect the investment outcome.

The arrangements for inter-regional transmission investments differ from those for intra-regional transmission investments, in that under the OFA model benefits could influence whether or not inter-regional investments proceed. This may occur in circumstances where the estimated LRIC is not 100 per cent offset by "bids" from market participants.

As a consequence, if competition benefits are excluded from the RIT-T assessment, there could conceivably be circumstances in which the cost of the inter-regional transmission investment outweighs the benefits (as calculated under the RIT-T and reflected in generator bids for interconnector capacity), with the result that the investment would not go ahead, but where including competition benefits would result in an overall positive net market benefit, and the investment proceeding.

However, there are a number of practical difficulties involved with estimating competition benefits under the current market arrangements, which make such quantification potentially controversial and open to criticism. These practical difficulties include the need to identify what constitutes "realistic bidding" (such as Nash-Cournot type bidding, Bertrand bidding, or an alternative), and how such bidding will be affected by different transmission investment options.

Under the current RIT-T, competition benefits do not need to be identified separately. If a particular credible option would result in changes in generator behaviour changing generator investment/timing and/or dispatch outcomes also change, these "competition benefits" will be reflected implicitly through the market modelling of

21 That is, as a result of its impact on the bidding behaviour of existing generators leading to a reduction in dispatch costs associated with changes in the output of those existing generators, in addition to that due to the entry of the new generator.

these categories of market benefit, provided that assumptions of realistic generator bidding is used.

Under OFA the RIT-T would no longer include estimates of changes in generator investment/timing and dispatch costs, as these market benefits would instead be assumed to be reflected in the generator requests for firm access. As a consequence there would no longer be a need to undertake market modelling in order to identify generator dispatch benefits, or changes in generation investment.

However, if competition benefits are to continue to be included in the RIT-T, these competition benefits would need to be separately identified. Moreover, given that such overall market modelling would no longer be required as a matter of course, the quantification of competition benefits also implies the need for an additional, standalone market modelling exercise.

There would likely be problems in separately estimating "competition benefits" as part of the RIT-T under the OFA model. It is highly unlikely that any market modelling would exactly match the actual generator requests/bids received. Put another way, expected limitations in the market modelling mean that the market modelling would almost certainly not match reality in terms of which generators are dispatched. Indeed, if market modelling was expected to match generator bids in reality, then there would be no need for the OFA arrangements, as market modelling would be sufficient.

It appears difficult to reconcile this, ie to incorporate/calibrate the market modelling to reflect the actual generator dispatch revealed via the request/bidding process for firm access.

Therefore given that there may be difficulties with modelling these benefits, and that, even if this was possible, the process would likely be time consuming and extend the time necessary to undertake the entire RIT-T assessment, competition benefits should not be included in the inter-regional investment test.

8.3.10 Additional incentive schemes

There may be other additional incentive schemes that could potentially form part of the OFA model. These might not form part of the initial model – rather, these may be "add-ons" at a later stage.

Two of these incentive schemes are discussed below:

- long-term incremental access incentive scheme; and
- abnormal operating conditions incentive scheme.

8.3.10.1 Long-term incremental access incentive scheme

Different incentive frameworks correspond to lower or higher-powered incentives. Emerging evidence in Australia suggests that TNSPs respond effectively to relatively low-powered incentives (eg the Service Target Performance Incentive Scheme (STPIS) under which TNSPs are exposed to a very small amount of their maximum allowed revenue), ie low-powered incentives create large changes in TNSP behaviour. In general - in the first instance at least - incentive schemes should be relatively low-powered since this might significantly affect their behaviour, while still minimising risks on TNSPs.

There are some instances where more high-powered incentives might be appropriate. The presence of high-powered incentives also requires appropriate indicators and incentives to be in place to ensure that TNSPs do not sacrifice service quality in their drive to reduce costs.

In particular, one such instance would be when a new access request is received. This would be for incremental access, ie in addition to current access agreements. This would occur when a new generator wishes to become firm, or an existing generator desired a higher firm access service. Here, it would be important to incentivise TNSPs to make efficient trade-offs and so decisions about the best way to provide this additional access.

Under such a scheme, TNSPs would be exposed to 100 per cent of settlement shortfalls for some period of time following when the new access starts – perhaps five years.[22] Five years seems likely to be long enough to provide adequate incentives.

Following the conclusion of the five year period, the assets would be rolled into the asset base and subject to the operational incentive scheme as discussed above.[23]

Similar to the above operational incentive scheme, the aggregate amount of shortfall would be subject to caps. Caps would be set both on a short-term (eg monthly or quarterly basis) as well as on a long-term (annual) basis. The caps would be set based on a similar set of principles to those articulated above.

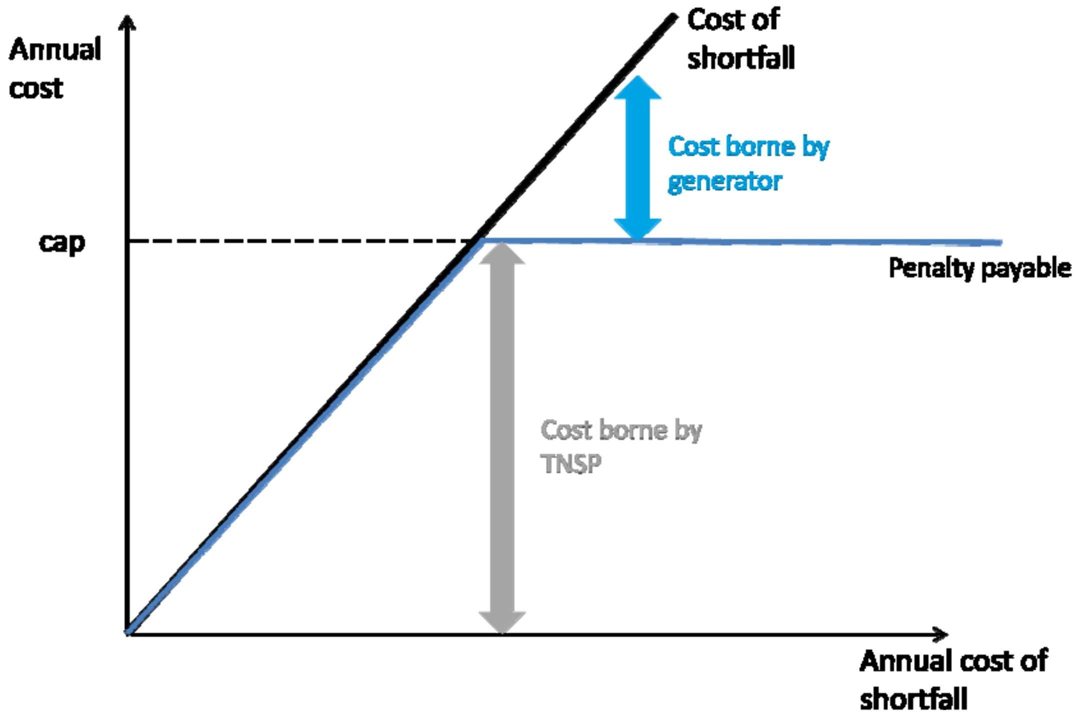
It is important to note that, if such a high-powered incentive scheme was warranted, it would only apply to a very small portion of the firm access provided.

Figure 8.4 illustrates the long-term incremental access incentive scheme. If firm access was not available to the generator, then the TNSP would be responsible for paying 100 per cent of the compensation to the generator - up to a cap.

22 The five year period could cover two regulatory periods, eg if access was released in year three of the first regulatory period, then five years would take it to year two of the second regulatory period.

23 To the extent that revenue received from access charges differed from the actual project costs that were rolled into the regulated asset base, then revenue from consumers would be used to offset the difference – that is, if access charges were less than the actual project costs consumers would fund this difference, whereas if access charges were greater than the actual project costs consumers would receive the benefit.

Figure 8.4 Long-term incremental access incentive scheme



The higher-powered incentives would more strongly incentivise the TNSP into making the right decision in how to provide the firm access service, choosing between:

- building new capacity now;
- making operational savings now, and building new capacity later;
- making operational savings forever;
- entering into network support agreements; or
- paying compensation.

This decision is very important and so it may be worth placing higher-powered incentives on TNSPs in order to encourage them to make the most efficient decision.

Furthermore, combined with this scheme, TNSPs could be subject to a regime to promote more timely release of access. A standard time would be set, in which most projects would be expected to be completed – although the times may differ for different projects, eg depending on estimated cost.

TNSPs could delay providing access within this time – provided they gained approval from the AER. If a TNSP considered there may be a delay in providing access (eg from delays in obtaining planning permissions) then it must obtain approval from the AER. It would have to justify why there is a delay through a written application. These would be circumstances where delay or failure to provide firm access is beyond the TNSP's control.

The AER would then approve (or not approve) this delay. If approval was not obtained, then the TNSP would be exposed to any settlement shortfalls.

An alternative to the TNSP gaining approvals from the AER each and every time there is a delay, which may result in large administrative burdens for the AER, would be to use a permit scheme. This is similar to a scheme used in the UK – described in Box 8.1.

Box 8.1: Permit scheme in the UK

National Grid Gas owns and operates the gas National Transmission System (NTS) in Great Britain. The UK system allocates capacity in the NTS between a small number of entry nodes and a national balancing point. Under this system National Grid Gas is obliged to release "incremental obligated entry capacity" for use by system participants, subject to a default investment lead time of 42 months.

National Grid Gas is given a number of opportunities or permits that can be used to extend the 42 month default lead time – these are defined in the license. This sets a limit on the number of months of allowed delay that can occur, subject to a total cap. These increases may be caused by the length of time required to obtain consents or construction challenges. The permits can be used without any justification to Ofgem (the regulator). The cap can also be increased by National Grid releasing incremental obligated entry capacity early.

When all of National Grid's permits have been exhausted, National Grid may only extend the 42 month default lead time for the release of capacity with the consent of Ofgem. This requires written application to, and approval from Ofgem.

8.3.10.2 Incentives outside the firm access standard

There is also likely a need to have incentives applying to TNSPs for performance outside the firm access standard, both:

- providing incentives to encourage TNSPs to provide an efficient level of access even outside normal operating conditions – wherever the cost of providing additional access is less than the benefit of doing so: eg by reducing the duration of planned outages when in abnormal operating conditions (and so forced outages apply); and

- providing incentives for TNSPs to minimise the duration of abnormal operating conditions: eg by reducing forced outage rates or return-to-service times.

However, TNSPs should face fairly low exposure under these schemes. Furthermore, any risk and reward balance for TNSPs in these circumstances should interact effectively with the other incentive schemes that exist.

9 Inter-regional access

9.1 Overview

The descriptions of OFA processes in the preceding sections relate entirely to *intra*-regional access: access from a generator node to the local RRN. However, the OFA model also establishes a framework for *inter*-regional access: access from the RRN of a neighbouring region to the local RRN.

Inter-regional access is included in the OFA model for two reasons. Firstly, many transmission elements provide a combination of inter- and intra-regional access and this combination is reflected in NEMDE by *hybrid* transmission constraints which include both generator and *interconnector* terms. To ensure that access settlement balances on *hybrid* flowgates, interconnector usage and entitlements must be defined, and interconnector access payments paid into or from access settlement. So long as there are hybrid flowgates, the inclusion of inter-regional access is unavoidable.

Interconnectors could play a purely passive role in the OFA model: accepting access on existing transmission capacity, but not seeking the additional access that would drive transmission expansion. However, just as there are efficiency benefits in allowing *generators* to decide their levels of *intra-regional* access, there are potential efficiencies from allowing interconnector stakeholders to decide *inter-regional* access. Generators and/or retailers can use an inter-regional access product to hedge against inter-regional price differences. This promotes competition between regions, placing downward pressure on prices.

Because the benefits of inter-regional access are potentially dispersed across a number of sectors, representatives of all of these sectors should – to the extent practical – be involved in a decision to expand interconnector capacity. This is achieved through a two-stage process. In the first stage, AEMO would run an auction for inter-regional access on interconnectors, offering access in quarterly blocks. Where auction demand appears sufficient to fund an interconnector expansion, AEMO uses the auction proceeds to request new long-term inter-regional access from TNSPs. In the second stage, the TNSPs assess whether the auction revenue is sufficient to cover the long-run costs associated with providing the new access, through a process analogous to a RIT-T.

Transitional inter-regional access may be allocated in the transition process, but only to the extent that this can be done without causing any additional scaling back of generator transitional access. Settlement payments arising from this transitional access will be paid into the inter-regional settlement residue, which will then be passed onto market participants through a process similar to existing Settlement Residue Auctions.

9.2 Design blueprint

9.2.1 Inter-regional terminology

Inter-regional access is similar to intra-regional access, but not the same. Interconnectors are like generators, but different in some important respects. Some additional terminology is needed to reflect these differences.

Interconnectors are conceptual dispatch entities in the NEMDE, corresponding to regulated interconnections between regions.[1] They do not submit dispatch offers; they are assumed implicitly always to have a zero network offer price.[2] But their dispatch by NEMDE can be considered to be analogous to the dispatch of generators, at least for the purposes of describing the OFA model.

A separate dispatch target is calculated by NEMDE for each DC regulated interconnector[3] and sent to the relevant TNSP, who operates the interconnector accordingly. Therefore, each physical DC interconnector is represented by a separate notional interconnector in NEMDE, even if they interconnect the same regions.

On the other hand, AC interconnectors are free-flowing: the interconnector dispatch will automatically reflect the imbalance between generation and demand in a region.[4]

There is therefore just one AC interconnector between each pair of neighbouring regions in the NEM, notwithstanding the fact that there may be several physical inter-regional AC transmission paths.

Each notional interconnector is treated the same in the OFA model, whether it is DC or AC .

Interconnector dispatch may be constrained in dispatch by NEMDE transmission constraints. The associated flowgates are referred to as inter-regional flowgates, if they are used only by interconnectors, or as hybrid flowgates, if they are used by both generators and interconnectors.[5]

Unlike with generators, there is no intrinsic maximum limit on the level of interconnector dispatch and so no corresponding concepts of registered capacity or offered availability. In practice, of course, there is always some flowgate that binds

---

1 Market interconnectors, or MNSPs, are treated like generators in the dispatch model, as discussed in section 2.3.10 (Interconnectors).

2 That is, they will be dispatched whenever the RRPs in neighbouring regions diverge. Put another way, the interconnector in one region can be considered to submit a generation offer price that is set equal to the RRP in the neighbouring region.

3 There are currently just two DC regulated interconnectors in the NEM: Murraylink and Directlink.

4 So, for example, if generation exceeds demand by 1,000MW in Queensland and 100MW is exported through Directlink to NSW, QNI will be dispatched for 900MW south and this amount is automatically exported on QNI to NSW.

5 Inter-regional flowgates are relatively rare and so it will be assumed that all flowgates relevant to interconnectors are hybrid flowgates.

when interconnector dispatch is high, so interconnector dispatch is not unlimited. However, there is no value, independent of NEMDE transmission constraints, which can be used as a proxy for availability.

A directed interconnector is a conceptual component of an interconnector. There is a pair of directed interconnectors for each interconnector, one in each direction. The concept is used in settlements, but not in dispatch, where the direction is signalled by the sign of the dispatch target.[6]

In AEMO settlement, a settlement amount is allocated to each directed interconnector. Currently, the amount allocated is based on the Inter-regional Settlement Residue (IRSR). Under the OFA model, the IRSR will be supplemented, in each settlement period and for each directed interconnector, by a payment from access settlement. The aggregate payment to each directed interconnector is referred to as the Directed Interconnector Payment (DIP).

In relation to settlement, directed interconnectors are analogous to generators, being the entities (albeit notional) that receive payments from AEMO. In the OFA model, this analogy extends to access: like a generator, a directed interconnector can have an amount of agreed access. Agreed access held by an interconnector is referred to as inter-regional access. In this chapter, where the distinction becomes important, a generator's agreed access is correspondingly referred to as intra-regional access.

9.2.2 Inter-regional access

Recall the description in section 2.2.2 (Dispatch and Network Access) of network access as a payment to a generator based on the difference between the regional price and the local price. Similarly, network access for a directed interconnector – inter-regional access – is also based on the difference between the "regional price" and the "local" price; but, in this case, the "regional price" is the RRP in the importing region and the "local price" is the RRP in the exporting region. The definitions of "importing" and "exporting" are predicated on the direction of the directed interconnector being considered. In this context, the directed interconnector is analogous to a generator that is located at the exporting region regional reference node but being dispatched to the importing region regional reference node.

As discussed in chapter 2 (Access), in the current NEM design, network access is linked to dispatch, so that a generator is paid this price difference on its dispatched output. The OFA model de-links access from dispatch and pays the price difference on the access level, irrespective of dispatch. Similarly, in the current NEM design, network access for a directed interconnector is based on interconnector dispatch but, in the OFA model, is based on the agreed access level.

6 AEMO uses the convention, which is also used in this report, that a positive dispatch target means a northerly flow direction and a negative dispatch target a southerly flow direction.

Algebraically, leaving aside losses, in the current NEM design, a directed interconnector is paid an amount equal to the inter-regional settlement residue (IRSR) which is defined as:

$$IRSR\$ = (RRP_M - RRP_X) \times IC \qquad (9.1)$$

Where:

RRP_M is the RRP in the importing region

RRP_X is the RRP in the exporting region

IC is the absolute level of interconnector dispatch

Importing and *Exporting* are predicated on the flow direction of the interconnector.

The payment is allocated to the directed interconnector which is aligned with the flow direction.

In the OFA model, access settlement is introduced which takes a form similar to generator access settlement:

$$\text{Access Pay}\$ = (RRP_M - RRP_X) \times (A - IC) \qquad (9.2)$$

where A is the access level, independent of dispatch. This means that the total payment to a directed interconnector is:

$$DIP\$ = IRSR\$ + \text{Access Pay}\$ = (RRP_M - RRP_X) \times A \qquad (9.3)$$

As with generator settlement, equations (9.2) and (9.3) are simplifications, since access settlement actually takes place on a flowgate basis. This is discussed in section 9.2.3 (Inter-regional hedging), below.

9.2.3 Inter-regional hedging

As discussed in section 2.3.3 (Financial Certainty for Generators), firm (intra-regional) access provides generators with a financial hedge against congestion risks. Similarly, firm inter-regional access provides directed interconnectors with a hedge against congestion risks. However, directed interconnectors are notional entities, not market participants; so, by itself, the provision of inter-regional access to interconnectors cannot provide any benefit to the market. This section explains how the OFA model is designed to ensure market participants can benefit from inter-regional access.

In the current NEM design, settlement payments to directed interconnectors are then allocated between those market participants who purchase, through the Settlements Residue Auction (SRA) the rights to receive a share of the IRSR.[7] The SRA rights might

7 In fact, anybody can participate in the SRA auction, but typically it will be a generator or retailer.

be purchased purely for their intrinsic monetary value, but more generally they are purchased for their value as inter-regional hedges (IRHs): hedges against exposure to inter-regional price differences that the purchaser has in its trading portfolio.

The firmness of the inter-regional hedging provided by current SRA rights is predicated on the firmness of the IRSR payments which, in turn, are predicated on the firmness of interconnector dispatch: in particular, if there is no interconnector dispatch, then the IRSR is zero and so there is no payment to SRA rights holders. Interconnectors, like generators, can be constrained off by NEMDE, due to the presence of generators competing with the interconnector for scarce transmission capacity.[8] Therefore, congestion affects the firmness of inter-regional access and, in turn, the firmness of inter-regional hedging provided by SRA rights.

In the OFA model, the payment to directed interconnectors, the DIP, depends upon the agreed access, not the dispatch, of the interconnector, as is seen from equation (9.3). Thus, when re-allocated to market participants, the DIP can provide firmer inter-regional hedging than currently. The rights to receive a portion of the DIP are referred to in the OFA model as Firm Interconnector Rights (FIRs) and are discussed further in section 9.2.5 (Firm interconnector rights and inter-regional expansion).

9.2.4 Inter-regional access settlement

Interconnectors make use of the same shared network capacity as generators. Therefore, they need to be included in access settlement on the same basis as generators. That means that access settlement for interconnectors occurs on every binding hybrid flowgate. If a directed interconnector's flowgate use is higher, or lower, than their flowgate entitlement then they will pay into, or receive from, access settlement, respectively.

However, there are some special considerations required for interconnectors, as discussed below.

9.2.4.1 Hybrid flowgates and directed interconnectors

The transmission constraint corresponding to a hybrid flowgate will include both generator terms and interconnector terms on the LHS: for example:

$$\alpha_1 \times G_1 + \alpha_2 \times G_2 + \alpha_{IC} \times IC < FGX$$

Where:

G_1 = dispatched output of generator 1

G_2 = dispatched output of generator 2

8 In fact, interconnectors are worse off in this respect than generators, because they are not able to rebid to -$1,000 when congestion occurs and so will typically be constrained off before generators.

IC = dispatched flow of interconnector in a northerly direction

α_1 and α_2 are the participation factors for the generators

α_{IC} is the participation factor for the interconnector[9]

FGX is the flowgate capacity.

In NEMDE, α_{IC} may be positive or negative, indicating whether the constraint limits the amount of inter-regional flow north or south, respectively. For the purposes of the OFA model, a directed interconnector is considered to participate in a flowgate only where it has a positive participation, based on the sign of the interconnector constraint co-efficient, α_{IC}. Mathematically, this means that:

- on *northerly flowgates*, with $\alpha_{IC}>0$, the *northerly interconnector* is considered to participate, with participation $+\alpha_{IC}$ and flowgate usage $\alpha_{IC} \times IC$; and
- on *southerly flowgates*, with $\alpha_{IC}<0$, the *southerly interconnector* is considered to participate, with participation $-\alpha_{IC}$ and flowgate usage $-\alpha_{IC} \times IC$.

Note that, although participation is always positive, usage may be positive or negative.

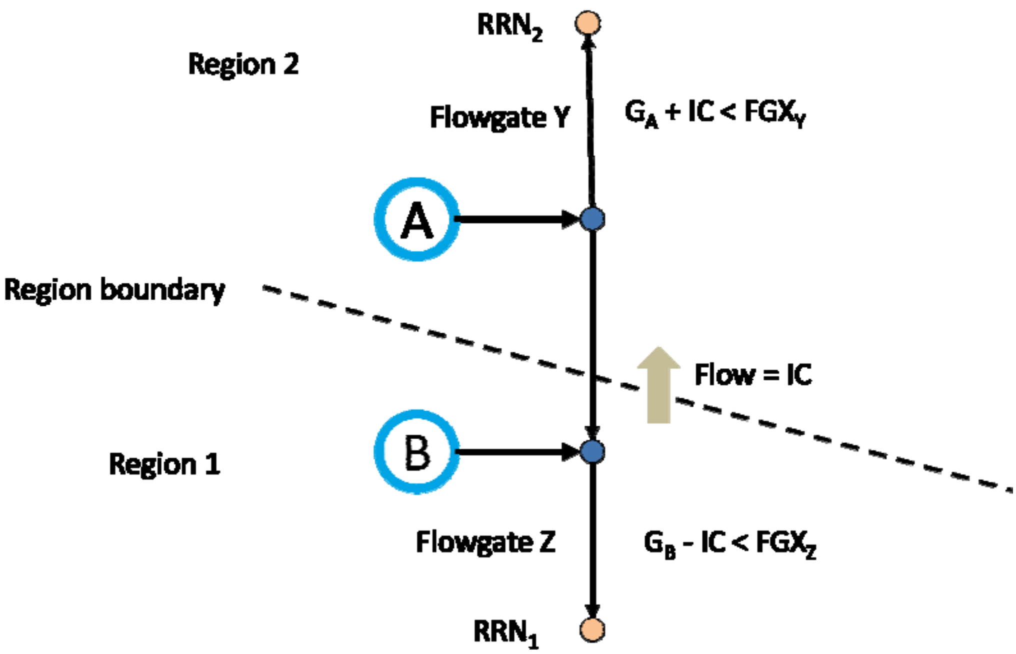
This convention avoids having to address issues that arise with negative participation.[10] It also ensures that inter-regional access is effective in supporting inter-regional hedging, irrespective of the direction of interconnector congestion or interconnector flow, discussed further in section 12.3.3 (Mixed Interconnector Constraints).

An example of northerly and southerly constraints on an interconnector is illustrated in Figure 9.1, below. The northerly interconnector participates in the northerly flowgate, Y, and the southerly interconnector participates in the southerly flowgate, Z.

9 Again, using AEMO's northerly convention.

10 Discussed in section 2.2.9 (Flowgate Support).

Figure 9.1 Northerly and southerly interconnector constraints



The two directed interconnectors are settled separately, based on the entitlements and usages for the respective flowgates that they participate in, and directed interconnector payments are determined accordingly.

9.2.4.2 Entitlements

Like generators, directed interconnectors may have some agreed (inter-regional) access.[11] A directed interconnector that participates in a flowgate will be allocated an entitlement on that flowgate based on its agreed access level.

Unlike with generators there is no concept of interconnector availability or interconnector registered capacity, prompting a slightly different method for setting entitlement targets:

- interconnector firm target entitlement is based solely on agreed access and participation;[12]
- interconnector super-firm target entitlements are defined to be zero;[13] and
- interconnector non-firm target entitlements are defined to be zero.[14]

---

11 Access procurement for interconnectors is discussed further in sections below. Note that a directed interconnector will typically have a different amount of agreed access to its twin in the other direction.

12 Recall that generator firm targets are capped by registered capacity, for which there is no interconnector equivalent.

13 Recall that generator super-firm targets only arise when agreed access is higher than registered capacity.

Firm actual entitlements for each directed interconnector are then calculated in the same way as for generators, with hybrid flowgate capacity allocated between directed interconnectors and generators. Because of the directed interconnector's zero non-firm target entitlement, it is possible that there will be some residual flowgate capacity even after all firm and non-firm targets have been fully met. This residual is allocated to the relevant directed interconnector as a non-firm actual entitlement.

For example, consider a radial hybrid 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 and 200MW of flowgate capacity remains. This 200MW is allocated to the interconnector as a non-firm actual entitlement.

9.2.4.3 Existing Settlement

Under existing settlement, the total inter-regional settlement residue on an interconnector that is available to be paid to the corresponding directed interconnectors is:

$$IRSR\$ = IC \times (RRP_N - RRP_S) \qquad (9.4)$$

Currently, the entire residue is paid to the directed interconnector whose direction is aligned with the interconnector flow; ie

If $IC>0$ then $IRSR\$_N = IRSR\$$, $IRSR\$_S=0$

If $IC<0$ then $IRSR\$_N = 0$, $IRSR\$_S = IRSR\$$

Where:

$IRSR\$_N$ = payment to northerly interconnector

$IRSR\$_S$ = payment to southerly interconnector

Whenever the flow direction changes, the payment stream will switch to the other directed interconnector. Under the OFA model, this allocation needs to change so that it is independent of interconnector flow direction.

The total existing IRSR$ can be broken down into payments associated with each congested flowgate on which the interconnector participates. Payment from each congested flowgate is then allocated according to the direction of the flowgate: payments from northerly flowgates go to the northerly directed interconnector; payments from southerly flowgates go to the southerly directed interconnector. Each flowgate payment takes the form:

---

14 Recall that generator non-firm targets are based on any excess of availability over agreed access, but there is no interconnector availability concept.

$$\text{IRFSR\$} = \text{U} \times \text{FGP} \tag{9.5}$$

Where:

IRFSR$ = is the inter-regional flowgate settlement residue for the flowgate

U is the directed interconnector's use of the flowgate

FGP is the flowgate price

The IRFSR$ payments sum to the IRSR

$$\text{IRSR\$} = \sum \text{IRFSR\$}$$

This result is demonstrated mathematically in section 12.2.11 (IRSR Allocation).

The allocation of the IRSR between the two directed interconnector is now dependent only on the directions of the congested flowgates and not on the direction of the interconnector flow, per se.

9.2.4.4 Access settlement

Access settlement for directed interconnectors uses the same formula as for generators: FGP x (E-U). The access settlement payment is paid to – or by – the relevant directed interconnector.

Thus, the total amount paid to a directed interconnector for each flowgate in which it participates is

$$\begin{aligned} \text{DIP\$} &= \text{IRFSR\$} + \text{access settlement\$} \\ &= \text{FGP} \times \text{U} + \text{FGP} \times (\text{E-U}) \\ &= \text{FGP} \times \text{E} \end{aligned} \tag{9.6}$$

This is a generalisation of equation (9.2) above. Again, for each flowgate, the payment will be made to the directed interconnector that is aligned with the flowgate direction, irrespective of the interconnector flow direction.

9.2.5 Firm interconnector rights and inter-regional expansion

9.2.5.1 Allocation of directed interconnector payments to market participants

In the current NEM design, AEMO regularly auctions the rights to receive the IRSR through Settlement Residue Auctions (SRAs), with the auctions covering quarterly periods for the next three years. Each purchaser of an SRA right for a particular directed interconnector receives a corresponding share of the IRSR in each settlement period. Fees paid to AEMO in the SRA are passed onto TNSPs, who return this money to demand-side users through reduced TUOS prices.

In the OFA model, SRA rights are replaced by similar financial instruments, called Firm Interconnector Rights (FIRs), which provide rights to receive a portion of the DIP rather than the IRSR. In each settlement period, a FIR holder is paid, for each flowgate that the relevant directed interconnector participates in:

$$\text{FIR Payment\$} = FGP \times \alpha_{IC} \times \text{FIR amount} \times k_F \qquad (9.7)$$

Where:

FIR amount is the MW amount of the FIR holding

α_{IC} is the participation factor of the directed interconnector in the flowgate

k_F is the firm scaling factor for the flowgate: the ratio of actual to target firm entitlements

Payments are made from the DIP of the relevant directed interconnector.

This payment formula is designed to achieve two objectives:

- the FIR provides a firm inter-regional hedge for the holder; and
- so long as the total amount of issued FIRs does not exceed the agreed access level on the directed interconnector, the total FIR payments never exceed the DIP in a settlement period.

These two features are discussed in section 12.11.1 (Inter-regional hedging using FIRs) and section 12.11.2 (Backing of FIRs by the directed interconnector payment), respectively.

Any residual DIP, remaining after FIR payments are made, is paid directly to TNSPs.[15]

9.2.5.2 Backing of FIRs by inter-regional access

FIRs are issued to market participants by AEMO through regular auctions, similar to the existing SRAs. The following sections provide an illustrative description of a possible auction design. The particular details of the auction design would be developed during OFA implementation. Importantly, the auction should create the right signals but also avoid creating opportunities for gaming by participants. The auctions would be designed to both allocate existing interconnector access, and prompt the requesting of new inter-regional access as required.

There is an important difference between existing SRA auctions and the FIR auctions proposed in the OFA model: in the FIR auctions, AEMO is not constrained to only issue FIRs in relation to *existing* inter-regional access, but may also request *new*

15 The residual DIP will come from three sources: any unsold FIRs; any non-firm interconnector entitlements; and the effect of inter-regional losses. The method for allocating the residual DIP between the two relevant TNSPs would need to align with the agreed allocation of FIR auction revenue, discussed in section 9.2.4.4 (Access settlement).

inter-regional access from TNSPs where there is sufficient market demand to justify this request.

To ensure that FIR payments are able to be funded entirely from the DIP, AEMO is not permitted to issue, in aggregate, a MW amount of FIRs on a directed interconnector that exceeds the level of inter-regional access on that interconnector. That access may be from:

- existing access: access previously issued; or
- new access: requested with TNSPs immediately following the current FIR auction.

Therefore, at an auction, AEMO may issue FIRs relating to:

- any unsold capacity from existing inter-regional access, ie "unsold baseline inter-regional access";
- new short-term inter-regional access, ie "short-term inter-regional access"; and
- new long-term inter-regional access requested after the auction, ie "incremental long-term inter-regional access".

These three components are explained further below.

The existing baseline inter-regional access comes from both transitional inter-regional access and any additional long-term inter-regional access that TNSPs have agreed to provide following earlier FIR auctions (that is, incremental inter-regional access following its commissioning).

For example, on a particular directed interconnector, there may be 100MW of transitional access and 200MW of additional inter-regional access, giving a total of 300MW of existing or baseline access. If only 250MW of FIRs are currently on issue, then there is 50MW of unsold capacity, which AEMO can use to back an issue of a further 50MW of FIRs at the auction, without needing new, incremental access from TNSPs.

Unsold baseline capacity may be provided for terms extending out to a specified long-term horizon: for example 10 years out.[16] This access is priced at a level reflecting LRIC, which is determined through separate processes undertaken by AEMO and TNSPs, as discussed further below.

New inter-regional access may be short-term or long-term.

Short-term inter-regional access is similar to short-term intra-regional access.[17] It is access that a TNSP is able to provide without needing to undertake any network

16 The specific term would be decided during OFA implementation.

17 Discussed in section 7.2.4 (Short-term Firm Access Issuance).

expansion. This might be because a TNSP is able to undertake operational actions which increase the effective capacity provided by existing network assets.

Short-term inter-regional access may be provided for terms up to three years out. Because it would typically not involve capital expenditure, it is not priced at LRIC, but rather at a price chosen by the TNSP, again similarly to short-term intra-regional access.

Incremental long-term inter-regional access is provided through network expansion. It would be offered for a term starting at 3-years out - reflecting the required lead time for investment. This access is priced at a level reflecting LRIC, which is determined through separate processes undertaken by AEMO and TNSPs, as discussed further below.

9.2.5.3 FIR auction schedule

It is proposed that AEMO holds quarterly FIR auctions.

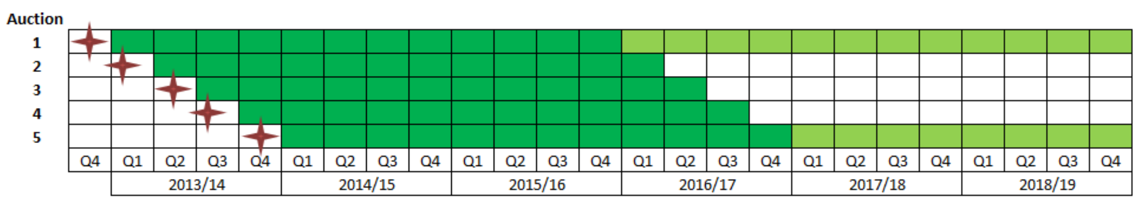
An example auction schedule is illustrated in Figure 9.2 below.

Auctions 1 and 5 would auction off all three types of access identified above - unsold baseline inter-regional access, short-term inter-regional access and incremental long-term inter-regional access - over a long-term period. The auction would be held in Quarter 4 of the year preceding when access is first offered. The dark green represents the quarters where unsold baseline inter-regional access and short-term inter-regional access are auctioned, while the light green represents the quarters where potential increased capacity and unsold long-term inter-regional capacity is auctioned.

This only occurs annually since the selling of long-term FIRs can potentially lead to substantial new capital expenditure by TNSPs, and so the auction process is correspondingly more complex. The objective of the auction design is to ensure that the cost to the TNSPs of expanding inter-regional access is funded by the auction payments for the corresponding long-term FIRs.

The auctions in between these full auctions (ie auctions 2, 3 and 4) will auction off the first two types of access over the upcoming three-year period - unsold baseline inter-regional access and short-term inter-regional access. These auctions can be considered analogous to the current SRA auctions.

Figure 9.2 Example auction schedule



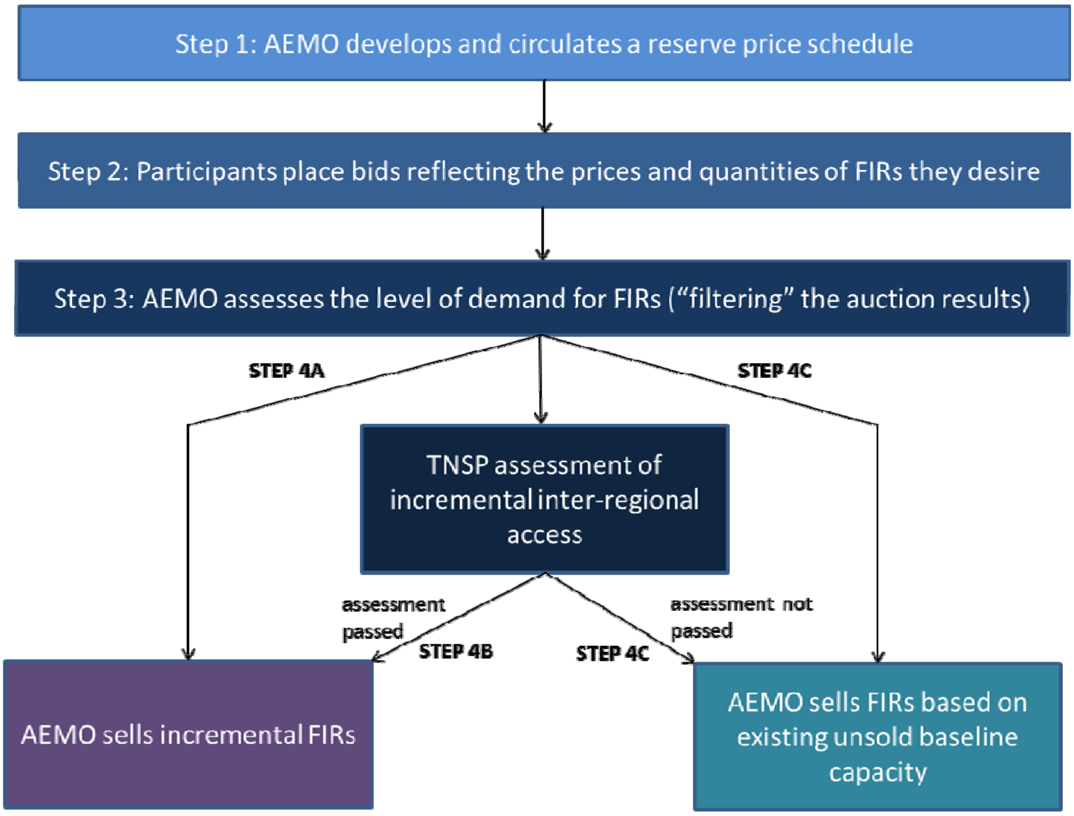
The distinction between short-term and long-term FIRs will be transparent to auction participants. Once purchased, all FIRs operate in exactly the same way, paying out according to the formula presented in equation 9.4.

9.2.5.4 FIR auction process

The FIR auction is composed of the steps set out below (and in Figure 9.3 below):

- Step 1: AEMO develops and circulates a reserve price schedule to the registered auction participants;
- Step 2: participants place bids reflecting the prices and quantities of FIRs that they wish to purchase in each quarter on each directed interconnector;
- Step 3: AEMO assesses the level of demand for FIRs, and the amount of new, incremental inter-regional access (if any) to be requested from TNSPs;
- Step 4A: if the level of demand meets a level of supply for incremental inter-regional access, then AEMO sells this amount of FIRs, with TNSPs providing the incremental inter-regional access;
- If the level of demand does not meet a level of supply for incremental inter-regional access, then either:
  - AEMO directs TNSPs to undertake an assessment of the access request - if assessment is passed then incremental FIRs will be released (Step 4B); or
  - if AEMO does not direct TNSPs to undertake an assessment, or the assessment is not passed, then AEMO sells a lower amount of FIRs that is able to be backed by the existing unsold baseline capacity alone (Step 4C).

Figure 9.3 FIR auction process



Any new access provided by TNSPs is funded by auction participants through an appropriate share of the auction revenue being passed through to them. AEMO does not itself fund the new access, nor does it have significant discretion about how much access should be requested: that would be decided through a mechanistic review of the FIR bids received from the market. It effectively acts as an agent for the auction participants, collating their bids and forwarding their requests for new access to the TNSPs.

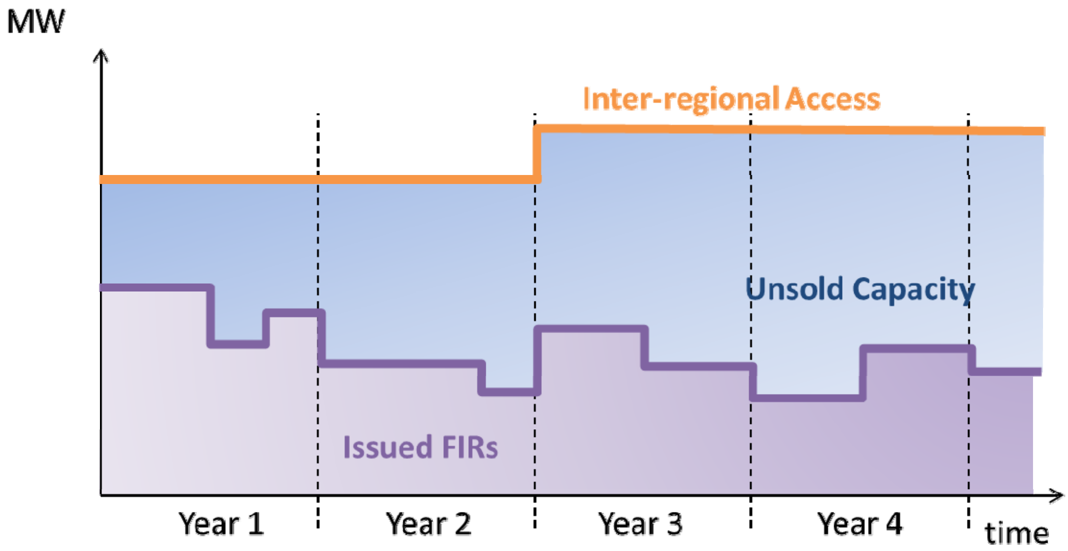
In Step 1, AEMO develops and circulates a reserve price schedule to the registered auction participants. Reserve prices would be set each quarter over the auction timescale taking into account:

- the level of unsold capacity in each quarter;
- an objective of maximising the realisable value of the unsold capacity; and
- the access prices payable to TNSPs for new inter-regional access.

To estimate the access prices payable, AEMO uses the standard LRIC pricing model.

Reserve prices will reflect either the value of unsold capacity or the cost of new access, depending upon how the corresponding FIRs are backed. Quantities of unsold capacity depend upon quarterly levels of FIRs sold in previous auctions and so may vary each quarter, as illustrated in Figure 9.4. Quarterly reserve pricing will reflect this variation.

Figure 9.4 Unsold capacity



The pricing of short-term access is discussed further in the section below.

Step 2 involves participants placing bids reflecting the prices and quantities of FIRs that they wish to purchase in each quarter on each directed interconnector. Bids for FIRs (for all three types of access) would all be in the same form. As noted above, the distinction between long-term and short-term rights should be transparent to auction participants.

In Step 3, AEMO assesses the level of demand for FIRs, and the amount of new, incremental inter-regional access (if any) that needs to be requested from TNSPs. In doing so, AEMO would take into account how much access would be needed to back the FIRs issued and the need to cover costs.

If the level of demand (based on the bids received) met a level of supply for incremental inter-regional access (based on the estimated LRIC price), then AEMO would sell this amount of FIRs, with TNSPs providing the incremental inter-regional access (Step 4A).

If the level of demand did not meet a level of supply for incremental inter-regional access, then AEMO could direct TNSPs to undertake an assessment of the access request, with the assessment determining whether or not incremental inter-regional access would be released.

The TNSPs would consider whether to offer the incremental access based on whether the dollar amount offered plus any positive externalities is considered sufficient to justify the cost of any network expansion (immediate or future) that would be required to back the new inter-regional access. The process for that decision is discussed further in the section below.

If this assessment is passed, then the new inter-regional access is provided and all the cleared FIRs are sold to the successful bidders (Step 4B).

If the assessment is not passed, then no new, incremental inter-regional access is provided and only the cleared FIRs that can be backed by the existing unsold capacity are sold (Step 4C).

Alternatively, AEMO would not direct TNSPs to undertake an assessment and simply sell a lower amount of FIRs (Step 4C), which is able to be backed by the existing unsold baseline capacity alone. This would occur when demand appeared likely to be insufficient to justify a further assessment by TNSPs.[18]

9.2.5.5 TNSP assessment of access request

As discussed above, relevant TNSPs would undertake a joint assessment to evaluate the costs and benefits of providing the requested access if requested by AEMO. This would occur where there was there a reasonable level of demand for access, which did not fully meet a level of supply for incremental inter-regional access.

This joint assessment can be considered analogous to the RIT-T. TNSPs would provide the incremental access if the cost associated with providing access was less than the benefits associated with the provision of access.

There are several important differences in the inter-regional situation which justify the need for an additional evaluation process in certain circumstances:

- there may be positive externalities (spin-off benefits) associated with the *particular* network expansion which is triggered by the new access: these create an effective reduction in the cost of access provision;[19] and
- because market participants are unlikely to bid for all of the FIRs associated with the new access over the auction timeframe, the amount demanded is likely to be less than the standard LRIC price, even though this is the basis used by AEMO for setting reserve prices.

These differences are considered in turn below.

The positive externalities include changes in the cost to the TNSP of maintaining demand-side reliability standards or the firm access standard. For example, an increase in inter-regional access on the NSW-Victoria interconnector may involve expanding network capacity in south-western NSW and/or north-west Victoria. These expansions may improve reliability for local demand in these zones and therefore substitute for other reliability projects - immediate or future - that a TNSP would otherwise need to undertake to maintain DSRS for this demand. It could also substitute for projects to meet firm access for other generators.

18 For example, if bidder demand was primarily just in the first year, say, it is unlikely that the associated auction revenue would be sufficient to satisfy TNSPs: one year of FIR revenue would be insufficient to cover the cost of the new ten or twenty year access.

19 These positive benefits are more likely to arise in inter-regional expansions, than in intra-regional expansions.

In relation to the second difference, when AEMO sets the auction reserve prices, it takes the inter-regional LRIC as the starting point and converts this into quarterly reserve prices using some form of amortisation. For example, suppose the LRIC for 100MW of 10-year access is $100m, or $1m/MW. Assuming a zero discount rate for simplicity, AEMO might convert that into a quarterly reserve price of $25,000/MW. Generally, the auction will not clear the full 100MW across all 10 years. It might clear, say, 100MW for the first five-years and 50MW for the remaining five years. If the clearing prices equal the reserve prices,[20] total revenue is then $75m: ie it covers only three-quarters of the LRIC, since only three-quarters of the potential FIRs have been sold.

In light of these differences, TNSPs will make a customised calculation of LRIC to be used as the “cost” in the investment test. This would typically involve a detailed analysis of the "with new access" and "without new access" expansion paths, and a comparison of the corresponding expansion costs. In other words, TNSP would undertake detailed expansion planning in order to estimate any positive externalities that may arise, as well as to more accurately assess the costs of specific expansion paths.

This analysis (both for costs and for positive externalities) undertaken should be proportionate to the magnitude of the new access request, ie large access requests should require more comprehensive and detailed analysis.

This different method for calculation costs is required in certain circumstances since the LRIC pricing methodology only models the cost of expanding thermal transmission capacity and assumes that this simply involves duplicating existing transmission paths. This is a reasonable approximation of intra-regional expansion, but inter-regional expansion will often involve entirely new transmission paths (eg Murraylink) or changes to network control equipment to address stability constraints. These possibilities are not represented in the LRIC method and may give rise to a different cost.

In the example above, the standard LRIC of new access is $100m but the amount offered, based on the incremental FIR auction revenue, is only $75m. Having carried out the detailed analysis, the TNSPs might find that the expansion cost is just $90m (compared to the $100m estimated in the LRIC pricing model) and that there are positive externalities worth $20m. In this case, the effective cost ($70m) is less than the amount offered ($75m) and the TNSPs would approve the request.

The evaluation of costs and benefits in the request assessment has some similarities to the evaluations that are undertaken in a RIT-T. However, the RIT-T applies to a particular project (or set of project options), and decides which project should proceed; whereas this process assesses a series of projects over time, based on an expansion path.

20 Depending upon the auction design, the clearing price may be equal to or higher than the reserve price. It must never be lower.

9.2.5.6 Time required for the TNSP assessment process

FIR auction bids are binding: auction participants are obliged to purchase the FIRs that are cleared. Where new long-term inter-regional access is involved, clearance could be contingent on the TNSPs assessment.[21] The TNSPs' assessment process is complex and may take some time. It would be unreasonable to expect market participants to tolerate indefinitely the contingent liability associated with binding, but not yet cleared bids. Therefore, there should be a time limit on the assessment process, with FIR bids automatically lapsing after this time. A 12-month time limit is suggested.

The analysis involved has a similar level of complexity to the RIT-T, because it requires the identification and analysis of actual expansion projects. Furthermore, because the two processes have the same ultimate objective – to ensure that network expansion is efficient and that the consumer does not end up paying for unnecessary or stranded assets – the need for transparency and consultation is similar. Currently, the RIT-T process as set out in the rules takes approximately 18 to 24 months. For the assessment to be completed within 12 months, some process streamlining may be needed.

We understand that the majority of time associated with preparing a RIT-T currently relates to complex energy market modelling, used to assess changes in generator fuel costs and location patterns. However, energy market modelling would no longer be required for the assessment, since costs and benefits relate only to networks.

It would be possible, and sensible, for TNSPs to make a head start on the assessment by undertaking some analysis before a formal access request is received from AEMO. For example, if the trend of earlier auctions, coupled with the findings of the NTNDP is that there is growing interest in new access on a particular interconnector, TNSPs could prudently commence analysis on expansion options knowing that this work would likely become relevant at some time in the future.

We note that the main drivers for the move to the RIT-T was to better allow for stakeholder consultation, and also to better facilitate interest in non-network options. Arguably, in the context of inter-regional access, these objectives are at least partially achieved within the FIR auction and so are less important in the assessment process. For example, a market participant that is considering bidding for FIRs will likely have already explored other options for hedging inter-regional risk.

9.2.5.7 Interaction between Short-term FIRs and Long-term FIRs

TNSPs would offer additional, quarterly, short-term access through AEMO. Ideally, this would be at access prices of their own choosing - consistent with short-term intra-regional access.

Since most interconnectors make use of the network of two TNSPs, the relevant TNSPs would need to reach agreement on the volume and price of offered short-term access

21 Unless, the level of demand meets a level of supply for incremental inter-regional access, in which TNSPs simply provide this access.

and how the proceeds would be divided between them. Each TNSP would need to ensure that it would be able to maintain FAS should the offered access be requested by AEMO, including through making any necessary operational changes.

However, there may be difficulties designing an auction to sell both short-term FIRs (where prices are ideally set by TNSPs) and long-term FIRs (where prices are derived under the LRIC methodology). This interaction, and the implications for auction prices, would be considered further in OFA implementation.

Ideally AEMO would design the auction rules:

- relating to short-term FIRs to maximise the expected proceeds, eg through setting a reserve price or by only selling a portion of the available capacity in each quarter; and
- relating to long-term FIRs to ensure that the cost to the TNSP of expanding inter-regional access is funded by the auction payments for the corresponding long-term FIRs.

9.2.5.8 Auction settlements

On completion of the FIR auction clearance, settlement must take place. Settlement could be immediate, but is likely to be delayed, at least for longer-term access and FIRs. For example, market participants might be required to pay for their FIRs immediately preceding the quarter in which they apply.

AEMO receives payments from successful bidders for FIRs, based on the auction clearing prices. This revenue is passed in its entirety to TNSPs.[22]

The settlement payments to the TNSPs would need to be divided into different buckets to reflect different regulatory treatments. The buckets would relate to:

1. FIR sales backed by existing unsold capacity;
2. FIR sales backed by new short-term access; and
3. FIR sales backed by new long-term access.

The first and third of these buckets would be retained by TNSPs, similar to access revenue from sales of long-term intra-regional access.

The second bucket would be retained by the TNSP, to encourage it to release short-term access, similar to the approach for short-term intra-regional access. Unlike for intra-regional access, there is a clear definition of existing unsold (ie spare) capacity, and short-term access is explicitly additional to this. However, in practice it is likely

---

22 AEMO may be entitled to levy auction fees to cover its reasonable costs, but these would be levied on bidders over and above the FIR payments.

that there would be some spare capacity on certain hybrid flowgates without the TNSP taking additional operational measures.

Most inter-regional access is provided jointly by two or more TNSPs, so the settlement would need to be divided between TNSPs in some way. Currently, SRA proceeds are paid entirely to the TNSP of the importing region on each interconnector. However, it is not clear whether that approach should continue,[23] when the cost of network expansion to provide inter-regional access may be borne in part by the exporting TNSP.

9.2.5.9 Lead time for new long-term inter-regional access

There is a lead time for TNSPs to expand network capacity to cover the processes of network planning, planning approvals, RIT-T assessment, construction and commissioning. When new inter-regional access triggers immediate network expansion – as it typically will – the expansion lead time will give rise to a corresponding lead time for provision of inter-regional access.

The auction design described above proposes a 3-year lead-time for incremental long-term inter-regional access.[24] FIRs sooner than 3-years out can, instead, be backed either by short-term inter-regional access or unsold existing inter-regional access. The appropriate lead time would need to be considered further in implementation. Potentially, lead times could vary for different tranches of capacity or for different interconnectors, on advice from TNSPs or the NTP.

9.2.6 Inter-regional and intra-regional interactions

9.2.6.1 Firm access standard

A TNSP is required to maintain capacity on flowgates in its region at or above the target flowgate capacity. That target will be driven by the agreed access levels of generators and directed interconnectors that participate in the flowgate. In some cases, generators in a neighbouring region may participate.

Thus, FAS requirements are driven by a combination of local intra-regional access, remote intra-regional access and inter-regional access. Notwithstanding the source of the FAS requirement, the TNSP retains sole responsibility for maintaining it.

The issuance of inter-regional access (whether in transition or through future inter-regional expansion), will mean that inter-regional transmission capacity must be maintained and may not be degraded through TNSPs using the capacity to provide

---

23 Except, perhaps, in relation to transitional access.

24 This is consistent with current practice. For example, ElectraNet and AEMO are assuming a three-year lead time for obtaining planning approvals, and constructing the relevant assets necessary to upgrade the interconnector.

new intra-regional firm access to generators connecting on inter-regional transmission paths.[25]

Although inter-regional expansion would commonly be a joint project between two TNSPs, FAS obligations would nevertheless fall solely on the TNSP in whose region the congested flowgate is located.[26]

9.2.6.2 Intra-regional access pricing

Inter-regional access would be included in the baseline flows in the access pricing methodology in the same way as intra-regional access is. Element flows associated with this access would be based on a load flow from one RRN to the other. Generally, flows on an element will be increased by interconnector flows in one direction and decreased by opposite flows, so only the former flow need be modelled.

A forecast of growth in inter-regional access needs to be included in the LRIC baseline. This would be based on information in the NTNDP as well as on the outcomes of recent FIR auctions.

Conversely, current and forecast firm generation would be included in the LRIC baseline for pricing inter-regional access.

9.2.6.3 Intra-regional Regulatory Investment Tests for Transmission

In some cases, a network expansion project required to provide or maintain intra-regional access may create new spare capacity on hybrid flowgates that facilitates the provision of additional inter-regional access.[27] The value of this would be assessed by the TNSP and this will be predicated on the TNSP's forecast demand for inter-regional access. Again, that forecast is likely to be based on information in the NTNDP and from recent FIR auctions.

Conversely, the RIT-T for a inter-regional network expansion project may need to take account of forecasts of firm generation.

The NTP could look at the interaction between these inter-regional expansions and intra-regional expansions as part of its expanded NTP role. This could form part of the information it is required to assess when evaluating cross-regional investment planning.

---

25 If a new firm generator did connect on an inter-regional transmission path, the TNSP would be required to expand transmission so that the interconnector firm access was not affected.

26 Where the location is unclear – for example in the case of stability constraints – the FAS obligation would need to be allocated and managed through some agreement between the two TNSPs.

27 This would affect the level of short-term inter-regional access that a TNSP could offer in the FIR auction and also the LRIC of long-term inter-regional access.

9.2.6.4 Short-term access

The OFA blueprint proposes similar, but separate, processes for issuing short-term intra-regional access (section 7.2.4 (Short-term Firm Access Issuance)) and short-term inter-regional access (section 9.2.5 (Firm interconnector rights and inter-regional expansion)).

The similarities are:

- each is issued through a quarterly auction;
- each relies on the TNSP to offer spare network capacity into the auction, including discretionary capacity that is contingent on some TNSP operational changes; and
- each provides upside to a TNSP by allowing it to retain the auction revenues.

A key difference is in the access term: as proposed in this OFA blueprint, short-term intra-regional access is only issued for the next quarter, whereas for inter-regional access the term is the next twelve quarters.

Depending upon how a TNSP chooses to release capacity, this may effectively give first refusal to inter-regional purchasers, in relation to spare capacity on hybrid flowgates. That might effectively prevent generators who use those flowgates from purchasing short-term intra-regional capacity.

Even if the access terms were aligned, the chronological requesting of the two auctions could largely determine how short-term hybrid flowgate capacity is assigned.

Alternatively, the two auctions could be combined into a single auction and so the capacity would go to the highest bidder, whether this was inter-regional or intra-regional. However, recognising that the auctioning of short-term FIRs is already combined with the auctioning of long-term FIRs, it may be difficult practically to link with the short-term intra-regional auction as well. This is a matter that should be considered further during OFA implementation.

9.2.7 Transitional inter-regional access

The scaling stage of the transition process (described in section 10.2.2.2 (Stage 2: Access Scaling)) will assume that there is zero inter-regional TA. Once that stage is complete, the maximum possible level of FAS-compliant inter-regional TA will be calculated through a similar process.[28] That level of transitional inter-regional access will be allocated in year one.[29] Unlike generator TA, inter-regional TA will not be sculpted back, but will instead remain at its initial level indefinitely.

28 Discussed in further detail in section 12.6.2 (Transitional Access Scaling).

29 Which might be zero, if hybrid flowgate capacity has been fully allocated to generator TA.

The TA would be used by AEMO to back the issuance of FIRs, as described in section 9.2.5 (Firm interconnector rights and inter-regional expansion). Over time, this TA would be supplemented by new long-term inter-regional access as discussed above.

9.2.8 Summary

The OFA model places market participants in the driving seat in relation to inter-regional expansions, just as it does for generators in relation to intra-regional expansions. However, it also recognises that there may be significant externalities associated with inter-regional expansion: benefits accruing to parties other than those who receive the settlement payments associated with inter-regional flows. The access pricing and FIR auction processes described allow these externalities to be factored into the expansion process.

9.3 Design issues and options

9.3.1 Non-firm target entitlement for interconnectors

Because there is no inter-regional equivalent of offered availability for interconnectors, their non-firm target entitlement is set to zero. An alternative approach would be to base the non-firm target entitlement on some nominal interconnector capacity which could be determined by AEMO, TNSPs or the NTP.

There are conceptual and pragmatic reasons for rejecting this alternative approach. Pragmatically, it is not clear on what basis the nominal interconnector capacity would be set: inter-regional capacity is varying continually, driven by changing transmission, generation and demand conditions.

Conceptually, it is not clear what benefit a non-zero non-firm target entitlement would create. It does not affect settlement payments to holders of firm interconnector rights (since these are based on firm entitlements) and would just feed some extra payments into the residual DIP paid to TNSPs, which will nevertheless remain highly non-firm and be of little benefit to TNSPs or demand-side users. At the same time, the change would reduce non-firm entitlements for generators, which would impose some additional costs and risks on those generators.

For these reasons, the preferred approach is to have zero target non-firm entitlements for interconnectors.

9.3.2 Counterprice flows and negative IRSR

Recall that, in the current NEM design, the settlement payment into the IRSR for an interconnector is

$$Pay\$ = (RRP_N - RRP_S) \times IC$$

Where:

IC is the interconnector dispatch in a northerly direction (negative if southerly)

RRP_N is the RRP at the northern end of the interconnector

RRP_S is the RRP at the southern end of the interconnector

Generally, interconnector flow direction is from the lower price to the higher price region, ensuring a positive settlement payment. However, it is possible for the interconnector to flow in the opposite direction, called a counterprice flow, leading to a negative IRSR. To offset this settlement deficit, the TNSP in the importing region is required to make a corresponding payment into AEMO settlement.

AEMO is required to clamp interconnectors[30] during periods of counterprice flow to prevent the negative residues from growing too large and materially affecting TNSPs and, ultimately, demand-side users.

In the OFA model, the payment to a directed interconnector, for each congested hybrid flowgate in which it participates, is (from equation 9.3 above):

$$\text{Pay\$} = \text{FGP} \times \text{E}$$

Where:

FGP is the flowgate price

E is the actual entitlement of the interconnector on the flowgate

The FGP is always positive. The entitlement, E, can never be negative: even if the interconnector has zero firm access, E, at worst, equals zero. Therefore, negative payments to directed interconnectors can never occur and, correspondingly, there is never a need for AEMO to clamp interconnectors or for TNSPs to make good any settlement shortfall.

The removal of negative residues will improve financial certainty for TNSPs and demand-side users, who will no longer be obliged to pay money into settlement during counterprice flows.

Counterprice flows are not inconsistent with dispatch efficiency. For example, there might be a situation of a low-cost generator in one region that cannot be dispatched towards its local RRN because of congestion but can, instead, be dispatched through the interconnector into a neighbouring region. Its dispatch is consistent with dispatch efficiency, so long as it is lower cost than the generation it is displacing in that neighbouring region. A high RRP in its own region (which would cause the interconnector flow to be counterprice) is irrelevant to dispatch efficiency.

30 Ie introduce additional, artificial inter-regional constraints into NEMDE so as to prevent the counterprice interconnector dispatch.

Thus, interconnector clamping can potentially be detrimental to dispatch efficiency. Consequently, by removing the need for such clamping, the OFA model can improve dispatch efficiency.

9.3.3 Pricing of inter-regional access

As discussed in section 9.2.5 (Firm interconnector rights and inter-regional expansion), reserve prices for long-term FIRs in the FIR auction are set by AEMO, based on the standard LRIC pricing method applied to inter-regional access. This creates consistency with the intra-regional access process.

The LRIC curve (meaning the LRIC expressed as a function of access amount) may in practice be a variety of shapes, which has implications for setting the prices within the auction:

- If the LRIC curve were monotonically upward sloping then there should be a conventional multi-tranche auction process where the LRIC is represented by a reserve price for each tranche and tranches would be cleared in turn so long as there was sufficient remaining demand above the tranche reserve price.
- If the LRIC curve were non-monotonic, the prices would require more consideration. Potentially, tranches can be defined so that the price of each tranche increases monotonically.[31]

Another potential problem is the term of LT inter-regional access. The LRIC method takes term into account, both by attributing "incremental usage" on each transmission element only for the term of the access request, and by assuming some long-term forecast flow growth on each element, so that any transmission capacity used by the agreed access during its term would not be completely stranded at access expiry but would have some "terminal value" which is then discounted from the access charge in the LRIC calculation. A similar approach could be taken for LT inter-regional access.

As our auction process proposes, the aggregate level of LT inter-regional access sold could potentially vary quarterly during the access term. The LRIC could price this variation, but in terms of the auction the LRIC-based reserve price must be set in advance. One potential resolution is to force participants to bid a "flat" amount for the full auction term. An alternative would be to have an iterative auction process, where prices vary over the auction term, reflecting the varying quantities.

9.3.4 Sharing with customers

A similar process applies in UK gas transmission (Box 9.1). Here, National Grid Gas decides whether to allocate capacity via a NPV test. This allows some of the risk associated with building the investment to be borne by customers. However, a different approach is proposed here, since it is counter to the overall rationale for OFA ie the market signalling the need for investment. Further, the 50 per cent sharing factor

31 This approach is adopted in the UK by National Grid Gas.

was set - it is understood - arbitrarily. Given this, we have therefore favoured the above approach (at least in the first instance).

Box 9.1: UK incremental entry capacity release for gas pipelines

National Grid Gas owns and operates the gas NTS in Great Britain. This has a similar process for allocating and expanding network capacity, as that described above for allocating and expanding inter-regional access. The UK system allocates capacity in the NTS between six large entry nodes, and a notional balancing point. This occurs through market participants signalling demand through auctions for entry capacity - quarterly blocks are sold on an annual basis.

This occurs through a four stage process, specifically:

- National Grid Gas holds a Quarterly System Entry Capacity (QSEC) auction - National Grid Gas publishes a price schedule for the auction, which sets out prices for releasing additional capacity. Bidders bid the quantity of entry capacity they want at each price, in each quarter over the auction period (16 years).
- National Grid Gas allocates existing capacity - where the aggregate quantity of bids at the reserve price is less than existing available capacity, then existing capacity is allocated.
- National Grid Gas decides whether to allocate incremental capacity via a NPV test - where a minimum quantity of additional entry capacity is demanded in any quarter, then National Grid Gas will consider whether to release this capacity by undertaking a NPV test. Here, National Grid Gas calculates the NPV of revenue from bids over 8 years. If the NPV is greater than, or equal to, 50 per cent of the "estimated project value" then it passes the NPV test. It is understood that 50 per cent is set on an arbitrary basis; however, this does facilitate risk sharing with consumer.
- National Grid Gas then decides whether or not to release incremental capacity - National Grid Gas is obliged to offer incremental capacity, but it is not obliged to build capacity to meet demand. It can either:
  - invest to increase NTS capacity - in which case it must make a proposal to Ofgem (the regulator) to do so; or
  - accommodate the increased obligations by better utilising the existing network eg substituting across the network or contracting for the capacity; or
  - buy-back capacity to meet its obligations.

9.3.5 Priority of transitional access allocation

It is proposed to give priority to existing generators in the TA scaling process and only to provide inter-regional TA in the event it is possible to do that, and remain FAS compliant, without clawing back TA from generators.

This approach is not predicated on any policy priority of inter- versus intra-regional access, but intended simply to reflect the de facto situation in the status quo. Currently, when there is congestion on hybrid flowgates, the affected generators can maximise dispatch by bidding -$1,000. Since interconnectors are unable to bid in this way, generators are always dispatched at the expense of interconnectors. AEMO intervenes to prevent material counterprice flows, which guarantees interconnectors zero dispatch but no more.

Thus, in the status quo, where access is restricted, generators receive a priority allocation and interconnectors are guaranteed only a non-negative allocation. The TA allocation design reflects the reality of this situation, irrespective of its merits.

9.3.6 Sculpting of inter-regional transitional access

It is proposed that, unlike with generator TA, inter-regional TA is not sculpted back over time but remains at its year one level in perpetuity. This approach is based on the fact that the reasons for sculpting of generator TA do not apply to interconnectors. Those reasons were:

- to reflect the future closure of power stations – and associated ending of de facto access – in the continuing status quo counterfactual; providing finite-life power stations with perpetual access would give them an unjustified windfall gain; and
- to prevent access hoarding creating a barrier to new entry or prompting unnecessary expansion.

The first reason does not apply to interconnectors, since they have no closure date. The second reason does not apply because the FIRs associated with the inter-regional TA are periodically auctioned and so cannot be hoarded over the longer-term.

10 Transition

10.1 Overview

Transition processes would apply in the early years following, implementation of the OFA model, with the objectives of mitigating the impact of its introduction and ensuring that affected parties have time to develop their capabilities for operating in the new regime without being exposed to undue risks in the initial period.

The main transition mechanism will be the allocation of *transitional access* (TA) to existing generators. TA acts identically to other firm access, except that it does not need to be procured from a TNSP and no access charges apply.

The transitional allocation process will have four stages. Firstly, generators' access *requirements* – the level of firm access they would need to have unfettered access to the RRN – are estimated, based on historical generation patterns. Secondly, these access requirements are *scaled back* to the extent necessary to ensure that the *existing shared network is FAS compliant*. Thirdly, this scaled access level is *sculpted* back over time, so that transitional access reduces over a number of years and eventually expires. Finally, an *auction* will be established to allow generators to sell some of their transitional access or buy additional transitional access from other generators.

10.2 Design Blueprint

10.2.1 Transition Objectives

The objectives of the transition process are:

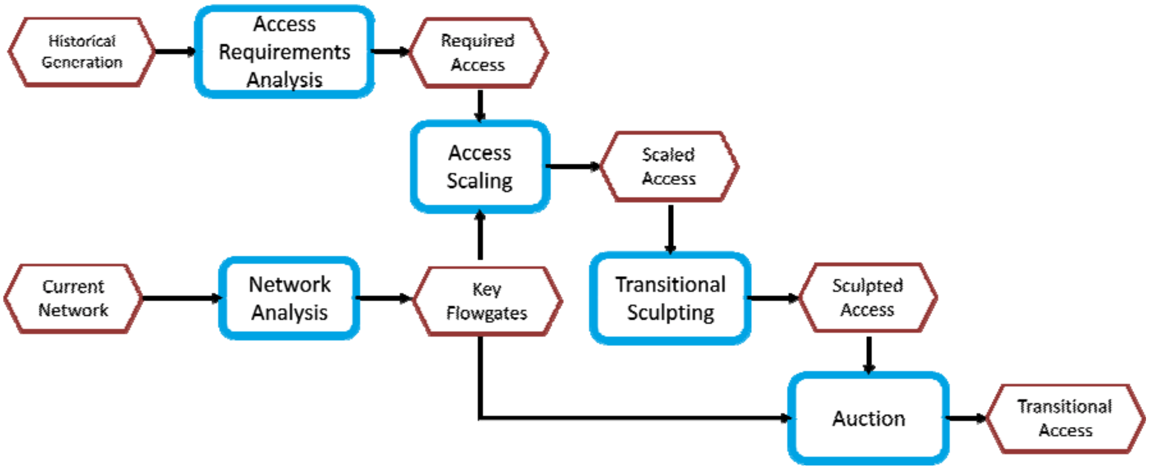
- to mitigate any sudden changes to prices or margins for market participants (generators and retailers) on commencement of the OFA regime;
- to encourage and permit generators – existing and new – to acquire and hold the levels of firm access that they would choose to pay for;
- to give time for generators 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 agreed access that could create dysfunctional behaviour or outcomes in access procurement or pricing.

Importantly, the transition process will not delay or dilute the efficiency benefits that the OFA model is designed to promote.

10.2.2 Transitional Access Allocation

The process for determining levels of transitional access consists of four stages, as illustrated in Figure 10.2 below. These are described in turn.

Figure 10.1 Transitional access allocation processes



10.2.2.1 Stage 1: Access Requirements

A generator's access requirement is the maximum level of access that a generator needs to give it full access to the RRN. As noted earlier, access agreements will be based on the peak and off-peak periods used in forward trading. Generator access requirements will therefore be expressed in terms of these periods, as summarised in Table 10.1 below.

Table 10.1 Generator access requirements

Gen Type	Peak Access Requirement[1]	Off-peak Access Requirement[2]
Baseload	Generator capacity	Generator capacity
Mid-merit	Generator capacity	Zero or minimum generation[3]
Peaking	Generator capacity	Zero
Intermittent	Generator capacity	Generator capacity
MNSP	Capacity in peak flow direction	Capacity in off-peak flow direction

1 Peak and off-peak times are aligned with forward contract convention.

2 Peak and off-peak times are aligned with forward contract convention.

3 Depending on whether the generator historically shuts down off-peak or runs at minimum generation level.

Gen Type	Peak Access Requirement[1]	Off-peak Access Requirement[2]
Interconnector	Zero[4] access in each direction	Zero[5] access in each direction

Therefore, this stage of the transition process aims to determine generator types – and hence their access requirements – by analysing historical output and bidding data.

10.2.2.2 Stage 2: Access Scaling

In this stage, access requirements will be scaled back as necessary to ensure that they do not lead to FAS breaches, given the capacity of the existing shared network. Scaling will be based on the following principles:

- Transitional access should be *maximised*, subject to *not* causing FAS breaches.
- Scaling should be *robust*: small changes in flowgate formulations should not cause large changes in scaling.
- Scaling should be *efficient*: extra access should not be granted to one generator if this disproportionately restricts the access that can be provided to other generators.

The algorithm for the scaling process will be conceptually similar to a dispatch algorithm: with access level analogous to dispatch level and a common requirement that aggregate access/dispatch must not cause transmission/flowgate constraints to be violated. In this context, the relative level of access will be determined by the quasi-bids which are entered into the scaling process.[6]

The access scaling process is described further in section 12.6.2 (Transitional Access Scaling).

10.2.2.3 Stage 3: Access Sculpting

The transitional access levels determined by stage 2 will set the level of access provided in year one: the year of OFA commencement. These levels will be sculpted back over time, following a profile illustrated in Figure 10.2 below.

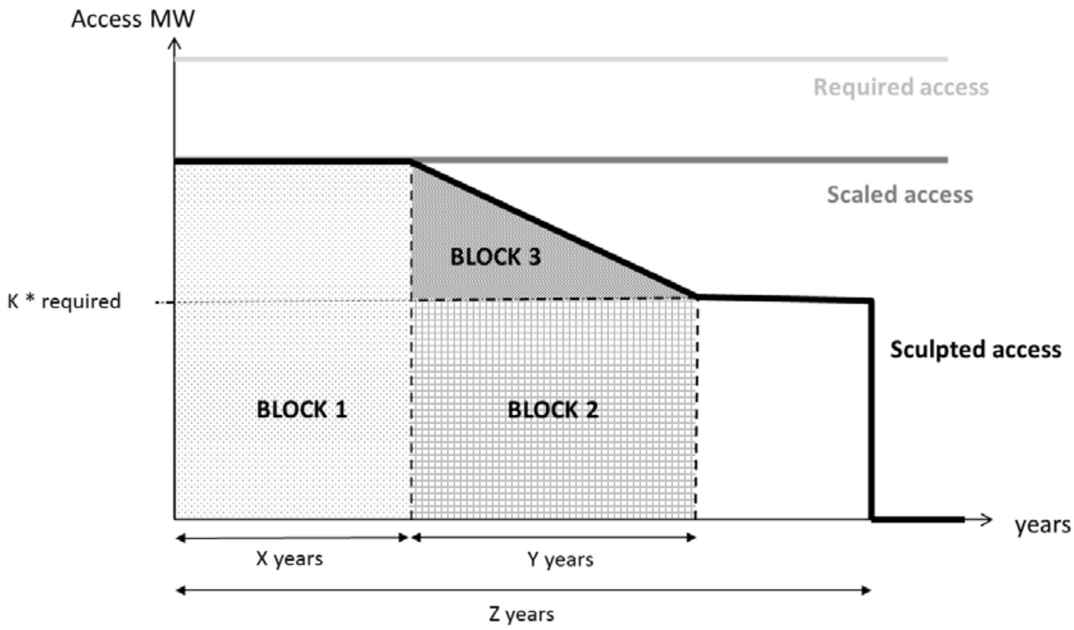
---

4 Zero access means that IRSR is compensated for any counter price flows (see section 9.2.4 (Inter-regional Access Settlement)).

5 Zero access means that IRSR is compensated for any counter price flows (see section 9.2.4 (Inter-regional Access Settlement)).

6 Sufficient additional notional load would be added at the RRN to ensure that generators were "dispatched" as high as the transmission capacity permitted.

Figure 10.1 Sculpting of transitional access for a Power Station



All power stations are provided with a minimum X+Y years of access. X is a learning period; Y the period needed to ensure a gradual transition. Younger power stations are provided with longer terms, Z years, where Z is a proxy for residual power station life. Z is specific to each power station, whereas X, Y and K are common to all generators.

The values of X, Y, Z and K (the access sculpting factor) would be determined during optional firm access implementation.

10.2.2.4 Stage 4: Access Auction

An *auction* will be established that permits generators to bid to top-up their TA level, by buying from other generators who choose to offer to *sell* some of their TA. Auction participation will be voluntary and bids and offers unregulated, except that a generator is not permitted to offer more TA than it has been allocated.

The auction would be similar to the ST auction described in section 7.2.4 (Short-term Firm Access Issuance). The main differences are:

- longer-term products would be traded, as discussed below;
- the auction clearing and settlement would be undertaken by AEMO;
- TNSPs would not offer any firm access into the auction; and
- any net revenue recovered from the auction would be passed through to demand-side users.

The only constraint placed on the clearing of bids and offers[7] is that the final, post-auction allocation is *FAS-compliant*. So long as compliance is based on the same set of flowgate constraints as in the stage 3 scaling process, auction settlement will clear.[8]

Auctioned products would be based on three blocks, shown in Figure 10.2:

- Block 1 is an X year block of constant volume;
- Block 2 is a Y year block of constant volume; and
- Block 3 is a Y year triangular block of declining volume.[9]

These three blocks are the generic components of all generator TA allocations for the first X+Y years. The auction design is described further in section 10.3.

Settlement of the auction would be through AEMO acting as a clearing agent: thus buyers would pay AEMO, who would pass payments onto sellers. Potentially, settlement could take place progressively over the term of the TAs, so in this case appropriate prudential arrangements would need to be established and applied.

10.2.3 Summary

The transition process will help to ensure that, from day one of the OFA regime, existing generators will hold agreed access amounts that provide them with firmness of access to the RRN similar to the *de facto* access they enjoy currently. Aggregate access holdings will initially be commensurate with transmission capacity but, as these are sculpted back over a number of years, transmission capacity will be freed up to support new access issuance: to existing or new entrant generators.

10.3 Design Issues and Options

10.3.1 Implications for Competition and Contestability

It has been argued that transitional access for existing generators creates a barrier to entry of new generators and so will lessen the degree of future generation competition in the NEM. This argument rests on two premises:

- that new generators must pay access charges to obtain firm access, whereas existing generators do not, placing the new generators at a competitive disadvantage; and
- that existing transmission capacity is allocated to transitional access, meaning that new generators must rely on transmission expansion.

---

7 Apart from the ones that apply to all auctions: that bids and offers are only cleared if the clearing price is below or above the bid or offer price, respectively.

8 See section 12.6 (Firm Access Allocation and Auctions) for an explanation of this.

9 Ie 100% of nominal volume in year one, (1-1/Y) ×100% in year two, (1-2/Y) ×100% in year three etc.

Having to bear costs that your competitors do not does create a competitive disadvantage. But such inequalities will always exist in competitive markets: for example, new generators must incur the capital cost of building a power station, whereas for existing generators, the costs are sunk; some generators may have higher debt or a higher cost of capital than others; and so on. Other things being equal, lower cost generators will be more profitable than higher cost generators, but this by itself does not disrupt competition. New generators will enter when market prices cover their entry and operating costs. This implies that market prices will have to rise somewhat to cover the cost of access charges. New generators will not enter the market until they expect the wholesale market prices to be high enough to cover their costs, and since new generators will also have to pay for firm access, market prices must be somewhat higher. However, TUOS prices will be lower than in the status quo counterfactual since generators are now bearing some of the costs of the transmission network. This lower TUOS price, would likely offset the higher wholesale prices for consumers.

It is true also that new generators may have to rely on transmission expansion. In some parts of the network, further transmission expansion may be impractical and so new generators may be effectively prevented from siting in these locations. However, there will always be plenty of other locations where new generators *can* and *will* locate.

In any case, over time, as TA is sculpted back, existing generators will increasingly bear the cost of access charges and spare capacity on the existing transmission network is likely to become available.

10.3.2 Reason for Sculpting Back Transitional Access

In principle, transitional access could be provided in perpetuity. However, this would compensate existing generators by more than is necessary to address any transitional impacts. A generator would continue to financially benefit from perpetual transitional access even after its power station had been closed - by selling the access to another generator, or by using it to provide access for a future power station - when there is no possible continuing impact from the OFA model on the earnings from that (now closed) asset.

The provision of ongoing free access would reduce the demand for new, paid firm access and so consumers would need to pay higher TUOS charges to make up the resulting shortfall in TNSP revenue. In short, perpetual transitional access would create an unnecessary wealth transfer from consumers to existing generators.

10.3.3 Choosing the Sculpting parameters

The sculpting approach used in stage 3 of the transition process is likely to be refined and changed during the OFA implementation process, with more specific parameters being defined.

Nevertheless, it is an important principle of the transition process that the same sculpting parameters (ie X, Y and K) are applied to each existing generator. This principle has the benefits of simplifying the process, avoiding debates about how one generator might be *slightly* different to another, and allowing for an auction process in which just the three standard blocks are auctioned.

11 Behaviour and Outcomes

11.1 Overview

The OFA model relies on changes in the market design and regulations to encourage changes to generator behaviour that promote market efficiency. Generator behaviour is driven by many factors, including the behaviour of other, competing generators. Poorly designed markets can easily encourage - unintentionally - behaviours that diminish market efficiency. Such behaviours may also arise in well-designed markets, if they not fully competitive.

This section examines, qualitatively, the drivers for generator behaviour in three areas:

- forward contracting;
- access procurement; and
- generator bidding.

The analysis is necessarily theoretical and identifies behaviours that generators might engage in under the OFA model, consistent with commercial objectives such as maximising profit and minimising risk. There is no empirical assessment of actual generator behaviour under current arrangements, or consideration of whether current behaviour is consistent with the behavioural models presented in this chapter.

For tractability, the analysis assumes moderately-sized, independently-acting generators optimising their short-term positions at the margin. Thus, behaviours inconsistent with these assumptions are not identified or examined: for example, portfolio behaviour, tacit collusion and so on.

11.2 Forward Contracting and Access Procurement

11.2.1 Overview

Access procurement is a form of forward contracting. In the current NEM design, a generator sells forward contracts in order to provide a stable revenue stream, hedged against the volatility of RRP. Similarly, an access agreement hedges a generator against the volatility of congestion prices: ie differences between RRP and LMP.

It is useful to draw analogies between forward contracting (RRP hedging) and access procurement (LMP hedging). The next section reviews drivers for forward contracting in the current NEM design, and the following section draws analogies between this and access procurement.

11.2.2 Forward Contracting in the current NEM Design

There are several drivers for a generator's forward contracting level:

- *Hedging value:* the forward price is much more stable than the RRP;
- *Forward premium:* since a forward contract also provides hedging value to risk-averse retailers, forward prices may be set at a premium to fair value; and
- *Limiting downside risk:* a generator who is unable to generate (due to unit outages or congestion) during periods of high RRP may be short and so exposed to large financial losses.

The first two factors encourage a generator to be *highly* contracted in order to maximise the hedging and premium value. However, these drivers are tempered by the last factor, which encourages a generator to choose a *lower* contracting level. Empirically, generators typically contract to a high percentage of generation capacity, but leave some uncontracted capacity in order to manage *short risks*.

11.2.3 Generator Margins under OFA Model

When a generator, under the OFA model has both sold forward contracts and procured access, its short-term operating profit, or margin, will be determined according to the formula below:

Margin = forward revenue + spot revenue + access revenue – access charge - generation costs

$$= F \times (FP - RRP) + G \times RRP + (A - G) \times (RRP - LMP) - AC - G \times C$$

Where:

FP = forward price

F = forward level (MW amount of forward contracts sold)

A = access level

G = dispatched output

C = marginal generating cost

AC = access charge

This equation can be arranged to:

$$\text{Margin} = F \times (FP - C) + (A - F) \times (RRP - C) + (G - A) \times (LMP - C) - AC \qquad (11.1)$$

In periods when there is no congestion, $RRP = LMP$ and so equation (11.1) reduces to:

Margin = F × (FP– C) + (G – F) × (RRP – C) – AC

During periods of congestion, the downside risks associated with high RRP are driven by the middle term in equation (11.1), and are managed by ensuring that $A \geq F$ over these periods. When there is no congestion, the risk is managed by ensuring that $G \geq F$, as in the current NEM design.

During normal operating periods, it is sufficient to procure agreed access at the forward level; or, conversely, to limit forward sales to the agreed access level. However, since access may be scaled back outside of normal operating periods, a generator may decide to procure a higher agreed access to cover this risk.

In summary, a generator will decide on its forward contracting based on similar considerations to now. It will then procure agreed access commensurate with its risk appetite. The more risk averse it is, the higher level of agreed access, in order to ensure that it has enough access outside of normal operating conditions.

11.2.4 Relying on Non-firm Access

In principle, a generator could rely on non-firm access to cover its short risk during congestion, which is similar to what a generator must do currently. However, in the current NEM design, many generators will be able to rely on some level of *de facto* access (through being dispatched) which is proportionate to their availability, through the mechanism of disorderly bidding.[1] In the OFA model, if *all* generators using a congested flowgate are *non-firm*, there is a similar pro rata outcome: since they will all receive non-firm entitlements proportionate to availability.[2] On the other hand, if all other generators around it are *firm*, a lone non-firm generator may get little or no access during congestion.

On this basis, one can then envisage two possible equilibrium scenarios:

- if all generators are *non-firm*, each might consider contracting forward without procuring access and thus taking the risk of *pro rata* non-firm access during congestion; or
- if most generators are *firm*, a non-firm generator is unlikely to take the risk of congestion and will procure access to cover its forward position.

So, two alternative *equilibria* are that:

- all generators are firm and new generators choose to be firm; or
- all generators are non-firm and new generators choose to be non-firm.

1 Assuming that constrained generators have the same participation factors. Where participation factors vary, generators with lower participation are preferentially dispatched, as discussed further in section 11.6.2 (Optimal Output under the Current Arrangements).

2 The difference is that in the status quo the pro rata access relies on the generator being dispatched, but in the OFA model it does not.

Games where optimal strategies are to do the same as everyone else are referred to as *coordination games*.[3]

In fact, there are two factors reducing the chances of a non-firm equilibrium:

- transition arrangements mean that most generators will start with high levels of agreed access and these levels will only be sculpted back slowly; and
- in a non-firm scenario, access prices will be low,[4] and so access procurement will be encouraged.

Therefore, it seems more likely that generators will procure access to cover their forward positions.

11.2.5 Summary

In the current NEM design, a generator's preferred forward level is limited by a concern not to be short – as a result of unit outages or being constrained-off – during periods of high RRP. In the OFA model, the congestion risk is removed, so long as the agreed access level equals the forward level. Thus, other things being equal, there will typically be increased forward contracting under the OFA model.

11.3 Free Riding

11.3.1 Overview

The discussion on forward contracting in the previous section is predicated on there being congestion risks that a generator needs to protect itself against. However, if congestion risks are low, a generator need not procure access. In effect, the generator can obtain the benefits of the transmission network (in removing congestion and ensuring that it receives RRP for its dispatch) without having to pay for it: it is *free riding* on transmission capacity that others have paid for.[5]

In the OFA model, transmission is expanded as necessary to meet both FAS and demand-side reliability standard obligations. Therefore, spare transmission capacity (ie over and above that required by firm generators) will be available only due to:

- *legacy:* transmission capacity built prior to OFA commencement;
- *lumpiness:* spare transmission capacity created during FAS-driven expansion; and

3 One example of a coordination game is deciding which side of the road to drive on.

4 Because the access pricing methodology will see a lot of spare transmission capacity and so substantially discount the costs of any future expansion caused by new access.

5 This free riding strategy is subtly different to the non-firm strategy discussed above. The non-firm strategy relies on obtaining non-firm access to congested flowgates, whereas free riding relies on flowgates being generally uncongested.

- *reliability:* transmission capacity provided to meet demand-side reliability standard obligations.

These possibilities are discussed in turn below.

11.3.2 Legacy

There is some congestion in the NEM currently, but it is unclear whether that congestion risk would be material enough to deter free-riders. In any case, the transition arrangements mean that existing generators will obtain free access initially and will not be making access procurement decisions for a number of years.

A new generator is likely to make access procurement decisions on a timeframe commensurate with its asset life: eg 30 years. Whilst legacy transmission capacity may be important in the early years of that timeframe, it becomes increasingly irrelevant further ahead.

In summary, the driver for free-riding will not be congestion materiality on OFA commencement but rather anticipated congestion 10 or 15 years out.

11.3.3 Lumpiness

Lumpiness of expansion applies at an *element* level, not at a *nodal* level. Thus, a generator seeking new access may cause a lumpy expansion of some transmission elements between that generator and the RRN but, at the same time, spare capacity on other elements that are *not* expanded will be reduced.

For example, consider a simplified scenario of a generator using three transmission elements *in series*.[6] Suppose that:

- each element has a lumpiness of 500MW;
- a new generator seeks 200MW of access; and
- initial spare element capacities are as shown in Table 11.1 overleaf.

6 Meaning that the path from the local node to the RRN is made up of the three elements, end to end.

Table 11.1 Impact of lumpy expansion on spare access

Element	Spare capacity prior to new Access	Expansion?	Spare capacity after new access
A	130	Yes	430
B	260	No	60
C	390	No	190
Node	130		60

The access request exhausts spare capacity on element *A,* prompting expansion and creating additional spare capacity. However, spare capacity is eroded on elements *B* and *C* but no expansion is prompted. The spare nodal capacity - the minimum of the spare capacities on the three elements - has actually decreased rather than increased, despite the lumpy expansion on element *A*.

Therefore, apart from special cases where a generator node is close to the RRN, it may not be the case that additional spare access would be created as the result of a lumpy expansion. Thus, it appears plausible at least that lumpiness of transmission expansion will not create significant opportunities for free-riding.

11.3.4 Reliability

A TNSP will be obliged – in order to maintain demand-side reliability standards – to provide *reliability access* to some non-firm generators if aggregate agreed access is less than peak demand. If a non-firm generator were confident that it would be provided with reliability access it may choose to free ride.

However, whenever there is some generation capacity margin - aggregate peak generation capacity exceeds peak demand - not all non-firm generators will need to be provided with reliability access. For example, suppose that:

- peak demand is 10GW;
- firm generation capacity is 9GW; and
- non-firm generation capacity is 3GW.

Total generation capacity is 12GW, giving a 2GW capacity margin. Reliability access must be provided to 1GW of non-firm generators; so each non-firm generator has, broadly, just a one-in-three chance of successfully free riding.

A TNSP is required to meet reliability standards at least cost. In the context of reliability access, that would mean choosing the non-firm generator which has the lowest expansion cost associated with reliability access. Typically, this would be a generator located closest to the RRN. Thus, a generator remote from the RRN may be ill-advised to take the free-riding gamble.

Is it plausible that there will be a substantial "capacity margin" of generation capacity over peak demand? Where the development of peak generation capacity (eg open cycle gas turbines) is primarily driven by peak RRPs, the answer might be no, since as soon as there is a significant capacity margin, peak RRPs will rapidly decrease. However, in the context of carbon pricing and the growth of intermittent renewables, it is possible that generation investment will be driven by factors (REC or carbon prices) other than peak RRPs and so a large capacity margin is plausible, at least.

11.3.5 Access Pricing and Congestion

The access pricing methodology takes account of spare transmission capacity and trends in demand for firm access and TUOS. Where there is limited congestion and significant free riding, these factors will both lead to lower access prices. Thus, free riding will be *self-regulating*, in the sense that high free riding will reduce access prices and so encourage more access procurement.

11.3.6 Short-Term Access

It is proposed that a TNSP is able to sell short-term access through an auction process, where there is spare network capacity. Auction clearing prices may be substantially lower than standard access prices, allowing generators to acquire cheaper access. Thus, a possible alternative free-riding strategy might be to not buy *long-term* firm access but instead rely on buying cheaper *short-term* firm access.

However, since the access is short-term, access would need to be renewed frequently, leaving a generator at risk of volatile auction prices. That risk may deter use of this free-riding approach.

11.3.7 Summary

It is possible that low levels of congestion will promote free riding, as generators choose to bear modest congestion risks rather than pay for access. However, this scenario relies on there being an enduring surplus of transmission capacity over the FAS requirement across the long time frame over which generators' access decisions are likely to be made. The most plausible source for this surplus would be the demand-side reliability standards. However, an individual non-firm generator can rely on being provided with reliability access only if the generation capacity margin remains low (meaning that TNSPs are required to provide reliability access to all non-firm generators) or if its advantageous location means the cost of providing it with reliability access is low compared to other non-firm generators.

Generators might delay procuring access until congestion becomes more material. However, a trend of growing congestion and access procurement will lead to growing access prices. Therefore, that strategy might mean a generator taking increased risks for no financial gain: it simply pays a higher access price later rather than a lower access price earlier.

In summary, a qualitative analysis suggests that the risk of substantial and enduring free riding is modest. However, that conclusion would need to be confirmed through quantitative modelling to determine the relative levels of congestion costs and access charges, which will be the main driver of free riding behaviour.

11.4 Generator Bidding in the Absence of Congestion

11.4.1 Overview

In the absence of congestion within a region, there is no access settlement and no difference in generator payments or incentives between the current arrangements and the OFA model.[7] The purpose of this section, then, is to introduce some important bidding issues and concepts in the relatively straightforward and familiar context of an uncongested region. It is not to assess the benefits of the OFA model compared to the status quo (since there is no difference), nor specifically to examine and critique current generator bidding behaviour.

In the NEM, generators can be expected to *bid* (ie submit dispatch offers) in a way that maximises their profitability. In this chapter, such bidding is referred to as *optimal bidding*.

A generator's profitability depends upon its output and on the spot price it receives. Therefore, a change in output (for example, through a rebid) affects profitability in two ways:

- *directly:* because of the change in output; and
- *indirectly:* due to the change in the market price caused by the change in output.

The size of the indirect component depends upon the degree of *influence* that the generator has (through changing its output level) on the market price. If this effect is small, the direct effect dominates and the generator seeks simply to maximise that component. That is done through maximising output when the market price exceeds generator marginal cost and minimising output when the market price is below marginal cost. Such a bidding strategy is referred to as *in-merit* bidding.[8] In an

7 Except that, under the OFA model, it is possible that a generator could bid to create congestion, even where the region is initially uncongested. This possibility is not analysed in this section. Once the congestion was created, bidding could then be expected to follow the strategies described in section 11.5 (Bidding Strategies under a Radial Constraint) or 11.6 (Bidding Strategies under a Loop-flow Constraint).

8 A generator is said to be *in-merit* when the market price exceeds the short-run marginal generating cost. Therefore, the in-merit bidding strategy is to only run when in-merit.

uncongested region, a generator bids in-merit by submitting dispatch offers in which the offer prices reflect marginal generating costs.[9]

On the other hand, where a generator has a degree of pricing influence and the indirect component becomes material, a generator's optimal bidding strategy will move away from in-merit bidding.

11.4.2 Choosing the Optimal Output

Operating margin for a generator when a region is uncongested is:

$$\begin{aligned} \text{margin} &= \text{forward contract revenue} + \text{NEM revenue} - \text{generating costs} \\ &= F \times (FP - RRP) + G \times RRP - G \times C \\ &= F \times (FP - C) + (G - F) \times (RRP - C) \\ &= \text{forward margin} + \text{regional margin} \qquad (11.2) \end{aligned}$$

Where:

FP = forward price

F = forward level (MW amount of forward contracts sold)

G = dispatched output

C = marginal generating cost

Note that the forward margin is independent of output or RRP and so optimal bidding aims simply to maximise the regional margin.

Bidding behaviour depends primarily on three factors:

- marginal generating costs;
- forward level; and
- the sensitivity of RRP to generator output changes (other things being equal).

The relationship between bidding and these variables is illustrated in Figure 11.1.

9 When there is congestion, such dispatch offers might not achieve in-merit output, for reasons discussed in section 11.5 (Bidding Strategies under a Radial Constraint).

Figure 11.1 Illustration of regional bidding

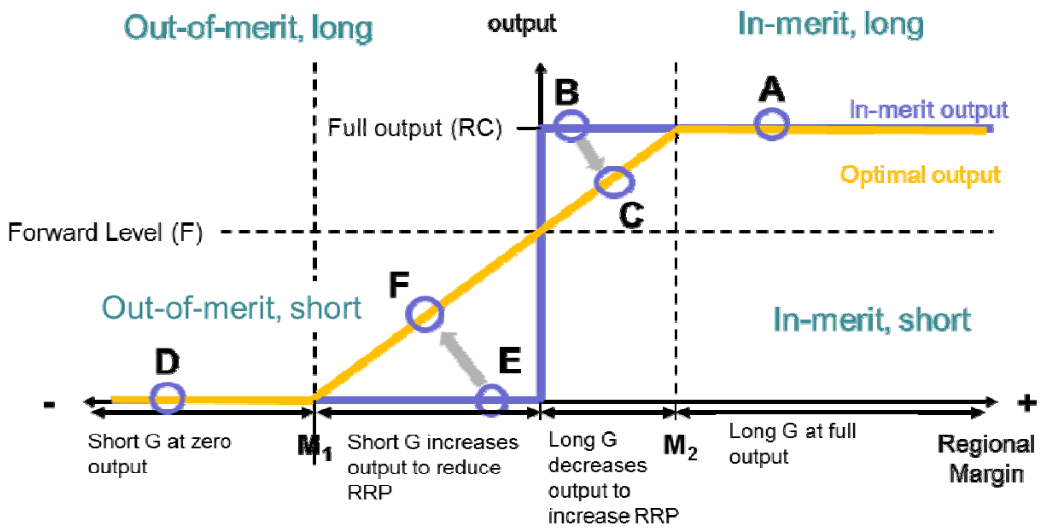


Figure 11.1 plots output (on the y-axis) against *regional margin* (on the x-axis). Regional margin is defined as the difference between RRP and marginal generating cost, *C*. If the margin is positive, or negative, the generator is said to be *in-merit*, or *out-of-merit*, respectively. If the margin is zero, the generator is said to be *at-the-margin*

On the graph, output is compared to the forward contract position or *forward level*. If output is greater than, or less than, the forward level, the generator is referred to as *long*, or *short*, respectively. It will be seen from equation (11.2) that a *long* generator benefits if RRP *increases*. Similarly, a *short* generator benefits if RRP *decreases*.

There are two curves on the graph. The *in-merit output* curve is the output of a generator bidding in-merit. Such a generator simply generates at full output if it is in-merit and doesn't generate if it is out-of-merit.[10]

The *optimal output* curve is the output level under optimal bidding. The optimal output level shown in Figure 11.1 has the following characteristics:

- *Away* from the margin – where the generator is either substantially in-merit or substantially out-of-merit – the optimal output is the same as the in-merit output.

- *Close* to the margin, the optimal output slopes gradually from zero output to full output as the margin changes, in contrast to the step change seen in the in-merit output.

The reasons for these characteristics can be understood by examining a generator who initially bids in merit and then considers whether it could improve its profitability by changing its output.

Consider a generator that:

10 The simplifying assumptions here will quickly be seen: dynamic and minimum output constraints are ignored.

- has a generation capacity = RC (in MW);
- has a regional margin = M (in $/MWh);
- has a forward position = F (in MW) which is less than its capacity (F<RC); and
- is bidding in merit.

Suppose that RRP has a *pricing sensitivity*, P, meaning that a 1MW increase in output by a generator (through rebidding), leads to a price reduction of $P/MWh, other things being equal. In reality, because generators are required to bid a constant offer price over each offer band, RRP will not respond *smoothly* to small output changes: it would typically not respond at all until there is a sufficient change to cause NEMDE to dispatch a different offer band or to cause other generators to rebid. However, for the sake of the analysis, the RRP impacts are assumed to be smooth and proportionate to output change.

Pricing sensitivity is inversely related to *supply elasticity*, S, which is the MW increase in output, across all generators, for a unit increase in price. If a generator rebids to increase its output by 1MW, the output from all other generators must reduce by 1MW and a price decrease of 1/S is required to induce this. Thus:

$$P = 1/S$$

If supply is perfectly elastic, meaning that S is infinite, then P=0. The *lower* the supply elasticity is, the *higher* the value of P will be, other things being equal.[11]

Since the generator is assumed initially to bidding in-merit, there are two situations to consider:

- *positive margin: M>0* and so the in-merit output is equal to RC and the generator is long (since *RC>F*); and
- *negative margin: M<0* and so the in-merit output is zero[12] and the generator is short (since *F>0*).

These are considered in turn.

In the *positive margin* case, if the generator (initially at full output) *reduces* its output by 1MW:

- the *direct* impact on profitability is the loss of 1MW of margin, costing $M per hour; and

11 However note that elasticity is typically expressed in relation to percentage volume and price changes, whereas P and S are expressed in terms of absolute changes in MW and $/MWh.

12 If a generator has a minimum output above zero it could be assumed to be operating at this level instead. The analysis is very similar under that alternative assumption.

- the *indirect* impact is to *increase* RRP by \$P/MWh, benefiting *P x (RC-F)* per hour, since the generator is long and so benefits from higher RRP.

Therefore, the overall impact is beneficial if:

$$M < P \times (RC-F) \tag{11.3}$$

There is therefore a breakpoint in behaviour when the margin is:

$$M_2 = P \times (RC - F) \tag{11.4}$$

If the margin is *greater* than M2 (illustrated on the graph by point A), then any reduction in output will be costly to the generator and its optimal output is full output, the same as *in-merit* output. If the margin is less than M2 (point B on the graph) then there is some benefit from reducing output.

The decrease in output leads to higher RRP (and hence higher M) and lower output, G. Thus, the direct cost of a further *reduction* in 1MW has increased and the indirect gain has decreased. Eventually, at the point at which:

$$M = P \times (G - F) \tag{11.5}$$

No further gains are possible and so the output level, G, is optimal (point C on the graph). The transition from in-merit output to optimal output is shown in the graph by a grey arrow.

Now consider the *negative margin* case, which is the opposite of the positive margin case. If the generator (originally bidding at cost and so at zero output) *increases* its output by 1MW:

- the direct impact on profitability is the addition of 1MW of negative margin, costing -\$M per hour (recalling that M is negative); and

- the indirect impact is to reduce RRP by \$P/MWh, benefiting P x F, since the generator is short and so benefits from lower RRP

Therefore, the overall impact is beneficial if:

$$-M < P \times F$$

ie if:

$$M > - P \times F \tag{11.6}$$

There is a breakpoint in behaviour at margin:

$$M_1 = -P \times F \tag{11.7}$$

If the margin is less (more negative) than M1, shown on the graph as point D, then any increase in output will be costly to the generator and its optimal output is zero, the same as its in-merit output. If the margin is greater (less negative) than M1 then there is some benefit from reducing output (point E on the graph).

The increase in output leads to lower RRP (and hence lower, more negative M) and higher output, G leading to a reduced short position F-G. Thus, the direct cost of a *further* increase in 1MW has increased and the indirect gain has decreased. Eventually, at the point at which:

$$M = P \times (G - F) \tag{11.8}$$

no further gains are possible and so the output level, G, is optimal (point F on the graph). The transition from in-merit output to optimal output is shown in the graph by a grey arrow.

If, for simplicity, it is assumed that the RRP sensitivity, P[13], is constant over the output range of the generator, the optimal output curve between M2 and M1, given by equations (11.5) and (11.8), is a straight line given by equation:

$$G = M/P + F \tag{11.9}$$

Or, re-arranging 11.9 and recalling that S is in the inverse of P

$$G = F + S \times M \tag{11.10}$$

This line has slope equal to S. If supply is highly elastic (S is high) then the slope is steep and optimal output will be similar to in-merit output. If supply is inelastic, the slope is flatter and supply will be closer to the forward level. Thus, regional optimal bidding – in this highly simplified model - can be summarised as follows:

- generators that are *away from the margin* will bid in merit;
- bidding of generators *close to the margin* will depend upon the level of supply elasticity;
- optimal bidding moves a close-to-the-margin generator's output closer to its forward level, compared to in-merit bidding; and
- the more *inelastic* the supply curve, the greater the likelihood and impact of *away-from-merit* bidding.

At any point in time, there will be:

- *possibly some short generators slightly out-of-merit:* bidding higher than in-merit output to reduce RRP;
- *possibly some long generators slightly in-merit:* bidding less than in-merit output to increase RRP; and

13 Or, equivalently, the supply elasticity, *S*.

- all remaining generators bidding in merit.

The impact on RRP of generators bidding away from merit will depend upon the relative influence of the short and long generators. RRP cannot go both up *and* down. It is like a tug of war, with long, close-to-the-margin generators trying to pull the price up and short, close-to-the-margin generators trying to pull it down. Therefore, the RRP under optimal bidding might be *higher* or *lower* than it would be if all generators simply bid in merit, depending upon the margins and influence of the long and short generators.

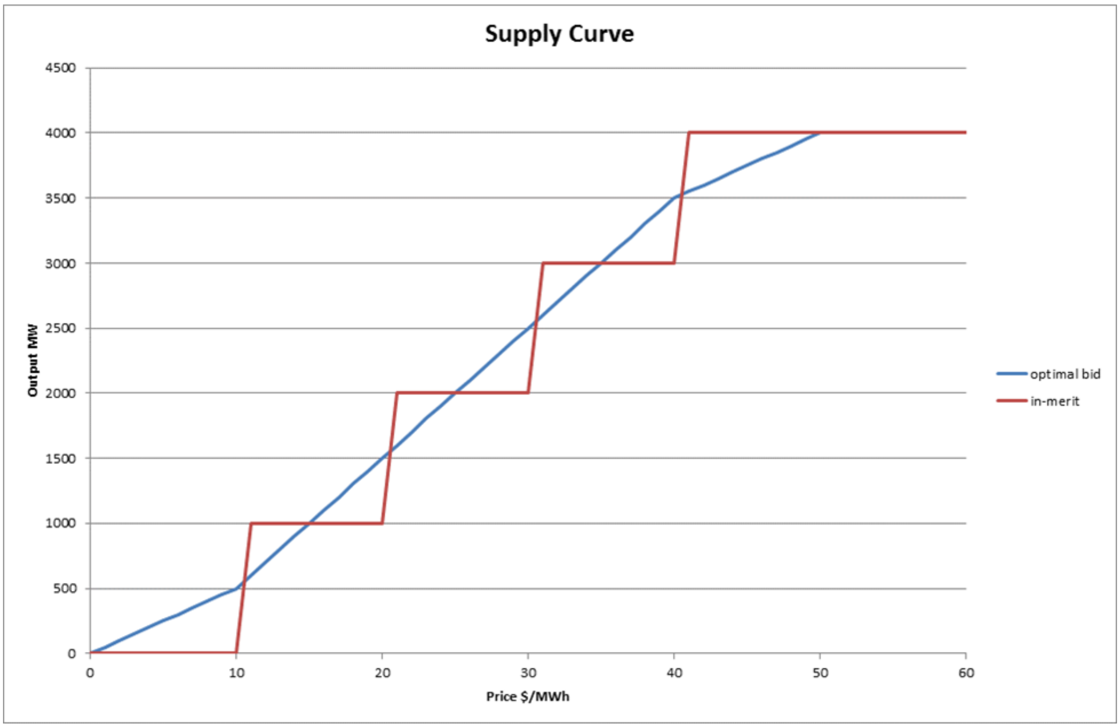
11.4.3 Supply Elasticity

One might think that supply elasticity is dependent on generator bidding. If this were the case then, since generator bidding is dependent upon supply elasticity, the analysis becomes circular and it becomes unclear whether the algebra presented above is meaningful.

However, although bidding has some impact on supply elasticity, the impact is limited. This is because the total amount of supply provided by an individual generator – being its generating capacity – does not change, just the way this is capacity spread across a price range. The aggregate effect of this is shown in Figure 11.2, below, where the supply curve from bidding in merit is compared to that when generators bid strategically.[14] It is seen that, apart from each end, the supply elasticity in each case is similar.

14 It should be noted that the axes in Figure 11.2 are reversed compared to a conventional supply curve graph. This is to emphasise that, for each individual generator, price is the independent variable and output is the dependent variable, being a function of price as expressed by the optimal output curve.

Figure 11.2 Supply elasticity



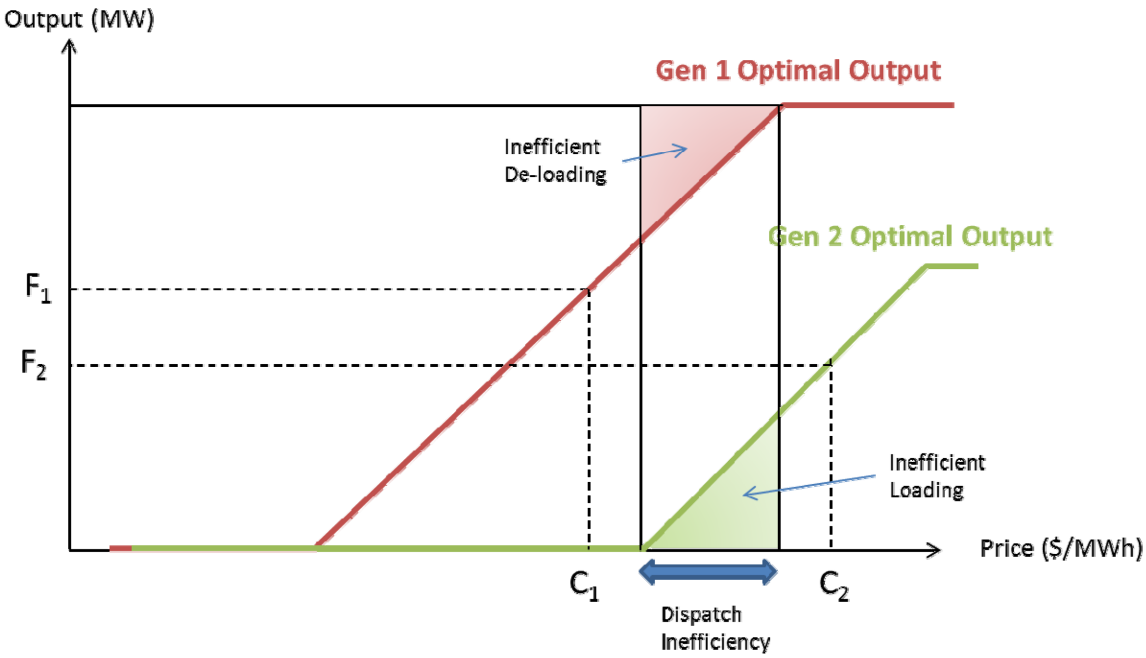
Thus, apart from at the top end, it can reasonably be assumed that the supply elasticity is independent of generator bidding. At the top end, because of the inter-dependence between supply elasticity and bidding strategy, there is a possible positive feedback effect: if generator optimal output curves flatten, this decreases supply elasticity, causing optimal output curves to flatten further and so on. This effect may underlie the bidding strategies seen from some generators in the NEM of bidding a small part of their generating capacity at very high offer prices.

11.4.4 Generator Bidding and Dispatch Inefficiency

Dispatch efficiency is a component of productive efficiency in the generation market. Dispatch is *efficient* if, to the maximum extent possible, generators are dispatched in merit-order, with the lowest cost generator fully dispatched before higher cost generators are dispatched. Dispatch is inefficient to the extent this ideal is not achieved.

In an uncongested region, dispatch inefficiency occurs if a generator is dispatched above minimum output and, at the same time, a lower-cost generator is dispatched below maximum output. Leaving aside ramp-rate limits on generator output, this should only occur where the sloping optimal output curves of the two generators overlap, as illustrated in Figure 11.3. The cost of the inefficiency is related to the amount of overlap, multiplied by the difference in cost between the generators. The flatter the optimal output curves, the greater the potential for dispatch inefficiency.

Figure 11.3 Dispatch inefficiency



11.4.5 Summary

A simple theoretical analysis suggests that generator optimal output curves should be sloping, rather than the step-changes associated with perfectly competitive, in-merit, bidding. The *slope* is in proportion to supply elasticity. The *intercept*, a generator's optimal output when RRP equals its costs, equals its forward level.

Optimal bidding will cause some dispatch inefficiency when two generators' optimal output curves overlap, causing a higher cost generator to be dispatched before a lower cost generator, for some portion of their output. Dispatch inefficiency is likely to worsen as supply elasticity falls, other things being equal.

The impact of optimal bidding on RRP is uncertain, since optimal bidding generally causes a smoothing-out of the supply curve rather than a raising or lowering. This can be envisaged as a tug-of-war between long and short generators close to the margin. However, at the top end of the supply curve, RRP is likely to be increased by optimal bidding.

11.5 Bidding Strategies under a Radial Constraint

11.5.1 Features of a Radial Constraint

Under a *radial constraint* all participation factors are either 1 (for *constrained* generators) or zero (for *unconstrained* generators). Radial constraints are relatively straightforward to analyse because the output of constrained generators cannot affect RRP.

This can be illustrated by dividing total generator output into constrained and unconstrained quantities:

G_C = total output of constrained generators

G_U=total output of unconstrained generators

The radial constraint means:

$$G_C = FGX$$

and the regional demand constraint requires that

$$G_C + G_U = D$$

Where:

FGX is the flowgate capacity

D is the regional demand

Putting these two equations together:

$$G_U = D - FGX$$

It is seen that G_U is independent of the output of the constrained generators, and that any increase in demand must be met entirely by an increase in unconstrained generation. Thus, RRP depends only on the bidding of the unconstrained generators and cannot be affected by the bidding or output of the constrained generators, so long as the constraint remains binding.

11.5.2 Optimal Output under the Current Arrangements: Disorderly Bidding

Recall from equation 11.9 that under the current arrangements, the optimal output for a generator is given by:

$$G = M/P + F \qquad (11.11)$$

P is the pricing sensitivity, which is the change in RRP caused by a 1MW decrease in the generator's output. As demonstrated above, the pricing sensitivity for a constrained generator is zero. Therefore, such a generator aims to run at maximum output if the margin is positive and minimum output if the margin is negative. That is to say, the optimal bidding strategy is to bid in merit.

In an unconstrained region, a generator with an in-merit bidding strategy would submit cost-reflective offer prices in its dispatch offer. But when there is congestion, a generator with such a dispatch offer could soon find itself being constrained-off. That is essentially because NEMDE does not dispatch against RRP but against LMP; that is to say, a generator is only dispatched by NEMDE if its offer price is below LMP, irrespective of the level of RRP.

For example, consider a generator with a cost of $30 which, wishing to bid in-merit, submits an offer price of $30. During congestion, whilst RRP remains at $40, say, LMP might fall to $20, say, causing the generator to lose dispatch.

To counter this anomaly, a generator that is bidding in-merit will need to reduce its offer price, further forcing down LMP and causing other generators also to reduce offer prices. This quickly leads to the familiar *disorderly bidding* where all constrained in-merit (against RRP) generators bid at the market floor price of -$1,000/MWh.

It should be clear from the discussion above that, in this situation, the bidding strategy itself is not disorderly; indeed, this is one of the few situations where all (constrained) generators will have in-merit bidding strategies, which is the competitive ideal that is generally sought by market designers. Rather, it is the tactical *application* of that bidding strategy which is *necessarily* disorderly, because of the anomaly of nodal dispatch and regional pricing.[15]

Unconstrained generators will continue to be dispatched against RRP and so follow the optimal regional bidding strategy described in section 11.4.2 (Choosing the Optimal Output), above. However, because constrained generators no longer respond to changes in RRP, the regional supply elasticity as reduced and the optimal output is given by adapting equation 11.10 to reflect this change in supply elasticity:

$$G = F + S_U \times M \qquad (11.12)$$

where:

S_U is the supply elasticity provided by the unconstrained generators

Because it involves only a subset of the generators in the region, the unconstrained supply elasticity, S_U, will generally be lower than the regional supply elasticity, S, which involves *all* generators. Define the *gearing factor*, γ, as:

$$\gamma \equiv S / S_U$$

Then equation 11.12 above can be re-expressed as:

$$G = F + S / \gamma \times M \qquad (11.13)$$

The gearing factor is named because it represents the degree by which the pricing influence of an unconstrained generator is *multiplied* during congestion compared to the uncongested situation: recalling that pricing influence in an uncongested region is the inverse of regional supply elasticity.

Under a radial constraint, the gearing factor for unconstrained generators is greater than one. For a *minor constraint*, involving only a minority of generators in a region, supply elasticity is not affected a great deal and gearing is only slightly higher than one. Under a *major constraint*, involving a majority of generators, supply elasticity is reduced substantially and so gearing may be much greater than one.

As well as reducing supply elasticity, congestion generally reduces aggregate supply, as generators are constrained off. This will force NEMDE to dispatch generators further up the supply curve. Typically, the supply curve is convex, meaning that supply elasticity *decreases* as output increases. Again, lower supply elasticity means higher pricing sensitivity.

As pricing sensitivity increases, unconstrained generators' optimal output curves will flatten and a generator close to the margin will seek to move its output closer to its forward levels. If most of these generators are long, this will lead to a further reduction in both supply and supply elasticity, amplifying and reinforcing the impact on RRP.

11.5.3 Optimal Output under the OFA model: Local Bidding

The optimal bidding strategies of unconstrained generators and constrained generators under the OFA model are considered in turn below.

The local price of unconstrained generators equals RRP, meaning that they have zero access settlement. Furthermore, as discussed above, because output of the constrained generators is limited by flowgate capacity and cannot affect RRP, the level and sensitivity of RRP remains driven solely by the supply of the unconstrained generators. Thus, for unconstrained generators, nothing changes under the OFA model, and so they will continue to have the geared, regional bidding strategy that was described in the previous section.[16]

For a constrained generator, LMP is not equal to RRP and so its revenue – and hence it optimal output – will be different under the OFA model. However, the associated bidding strategy (referred to here as *local bidding*) closely resembles – at least qualitatively – the regional bidding described earlier, for reasons explained below.

Recall the operating margin of a generator under OFA, as discussed in chapter 2 (Access).

Margin = dispatch margin + access margin

= G × (LMP-C) + A × (RRP - LMP)

= A × (RRP-C) + (G-A) × (LMP - C)

= regional margin + local margin (11.14)

Where

G = output level

15 And these tactics may also extend to other bidding aspects, such as changing ramp rate limits.

16 This ignores possible portfolio effects where a generating company owns both a constrained generator and an unconstrained generator. A change to the output of the constrained generator could affect the optimal output of the unconstrained generator and hence RRP. Portfolio effects are not considered here.

A = access level

C = marginal generating cost

The structure of equations (11.2) and (11.14) is very similar. Because a constrained generator cannot affect RRP, its regional margin is independent of output and so local bidding aims simply to maximise the local margin. We see that local margin in (11.14) has the same structure as regional margin in (11.2), with LMP replacing RRP and access level replacing forward level.

Therefore, similar to regional bidding, local optimal bidding depends upon three factors:

- marginal generating costs;
- access position; and
- the *local* supply elasticity: the MW amount of increase in local generation for a given change in LMP (other things being equal):

Correspondingly, for the purposes of local bidding, we say that a generator is long if output exceeds access, and *short* if access exceeds output.[17]

Because the margin drivers between the regional and local situations are very similar, optimal bidding is also very similar. Assuming again that supply elasticity is constant over a generator's output range, the optimal output is given by the equation:

$$G = A + S_L \times M = F + S/\gamma \times M \qquad (11.15)$$

Where:

S_L is the local supply elasticity

$\gamma = S/S_L$ is the gearing factor

Only constrained generators will respond to LMP changes, and so local supply elasticity is the supply elasticity provided by constrained generators: ie

$S_L = S_C$

$S = S_U + S_C$

Where:

S_C=supply elasticity from constrained generators

17 This *long* and *short* terminology is used in this section for consistency with analogous terminology in relation to RRP. Each compares generation output against a contracted level. It should not be confused with the terminology of access-long and access-short used in other chapters, which has the opposite meaning: eg *access long* is where access *exceeds* generation.

S_U = supply elasticity from unconstrained generators

S = regional supply elasticity

Note that the optimal bidding strategy for constrained generators is to respond to changes in LMP, not RRP, as expressed in equation 11.15, above. So the constrained supply elasticity is the change in output for a given change in *LMP*. However, supply elasticity is still fundamentally dependent on the capacity and cost of the relevant generators, for the same reasons set out in section 11.4.3 (Supply Elasticity) above.

Local optimal bidding is presented in Figure 11.5.

The optimal output is again plotted against margin, but now it is the *local* margin, LMP-C, rather than the regional margin. The reasons for the shape of the curves is exactly analogous to the regional situation and do not need to be explained a second time.

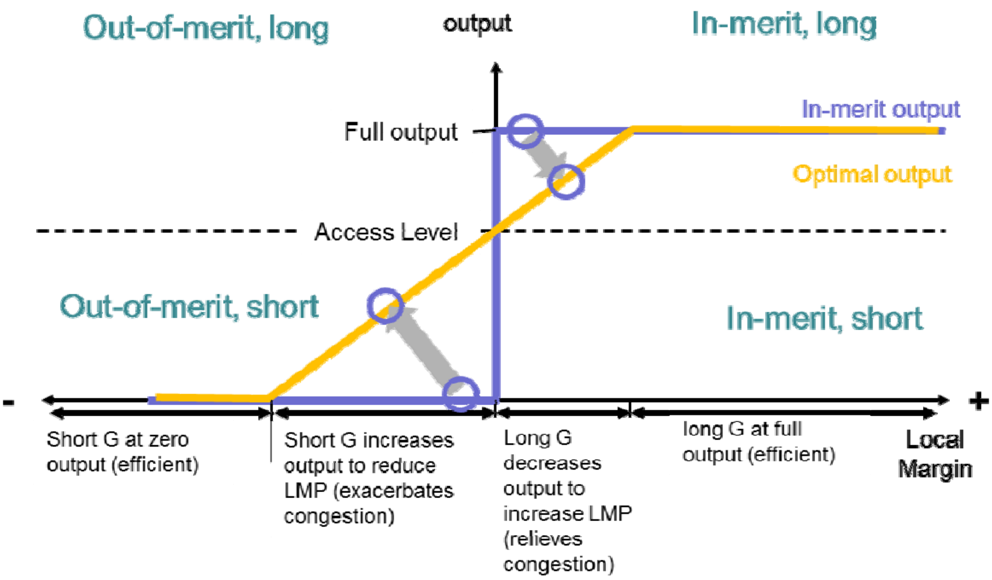
Given the similarity of the graphs, we can draw very similar conclusions about local optimal bidding:

- generators that are **locally** *away from the margin* will bid in merit (against LMP);
- bidding of generator **locally** *close to the margin* will depend upon local supply elasticity;
- optimal **local** bidding moves a generator's output closer to its **access** level compared to in-merit bidding; and
- the more *inelastic* the **local** supply curve, the greater the likelihood and impact of away-from-cost bidding.

The differences from the regional conclusions are highlighted in **bold** font. It is important to note that optimal output under local bidding is a function of *LMP*, which is also the basis on which NEMDE dispatches generators. So the existing anomaly between bidding and dispatch under the current arrangements[18] – which gives rise to disorderly bidding – is resolved in the OFA model. In particular, a generator bidding *in-merit against LMP*, simply submits cost-reflective offer prices.

18 As discussed in the previous section.

Figure 11.4 Illustration of local bidding



Similar to the regional bidding situation, we can think of a tug of war metaphor where long close-to-the-margin generators seek to increase the LMP and short close-to-the-margin generators seek to decrease the LMP, recalling that:

- short means *output* < *access:* eg firm, constrained-off generators; and
- long means *output* > *access:* eg non-firm, dispatched generators.

Increasing the LMP is equivalent to *reducing* flowgate prices: ie seeking to relieve congestion. Similarly, decreasing the LMP is equivalent to increasing the flowgate prices: ie seeking to exacerbate congestion.

11.5.4 Impact of Optimal Local Bidding

Recall from section 11.4.2 (Choosing the Optimal Output) that a generator will only bid away from cost where its margin against the market price is in the range:

$$M_1 < M < M_2$$

The values of M_1 and M_2 were presented for the regional context in equations (11.4) and (11.7) above. The corresponding equations for local optimal bidding are:

$$M_1 = -P \times A \qquad (11.16)$$

$$M_2 = P \times (RC-A) \qquad (11.17)$$

Where

P is the local price sensitivity (the inverse of the local supply elasticity)

A is the level of access

RC is the registered capacity

The size of the range between M_1 and M_2 is:

$$M_2 – M_1 = P \times RC \tag{11.18}$$

Assuming, still, for simplicity that P is a constant value, the RHS of equation (11.18) is equivalent to the impact that a generator can have on the *local* price by increasing from zero output to full output. A similar result would be obtained for the regional case. In practice, P is not a constant value, but qualitatively this result is likely to be broadly true: that the price range of a generator's away-from-cost bidding will be proportionate to the generator's pricing influence: the extent to which it can influence local or regional prices.

Local supply elasticity will be lowest – and so pricing influence is highest – under minor constraints with few constrained generators. Thus, optimal bidding departs most significantly from in-merit bidding under minor constraints and the impact on dispatch inefficiency is likely to be proportionately higher for such constraints. On the other hand, because only a few generators are involved, the materiality of the inefficiency (as a proportion of *total* regional dispatch costs) may remain low.

On the other hand, for major constraints involving a majority of generators in the region, local supply elasticity is similar to supply elasticity in an unconstrained region and so the materiality of dispatch inefficiency is also likely to be similar.

The overall impact on LMP[19] depends upon whether short or long close-to-the-margin generators have more *pull* in the tug of war. The design of the OFA model ensures that long and short generators are broadly matched since – for a radial congested flowgate at least - total access equals total generation equals flowgate capacity.[20] So:

$$0 = \text{total generation - total access}$$

$$= \textstyle\sum_i G_i - \sum_i A_i$$

$$= \textstyle\sum_i (G_i - A_i)$$

meaning that individual generator long and short positions exactly offset each other.[21] However, that does not mean that the short and long positions of close-to-the-margin generators exactly offset each other.

An important feature of local pricing impacts are that they do not affect customers, who continue to pay regional prices in the OFA model.[22] Therefore, optimal bidding is

19 Noting again that behaviour of constrained generators cannot affect RRP.

20 Since entitlement equals access.

21 This result does not necessarily apply in the regional market, where total forward contracts only equal total generation, if retailers – the buyers of those forward contracts – are 100 per cent hedged.

22 Although in some cases loop flows effects could mean that local bidding will distort regional prices to some extent.

a zero sum game between local generators. One important consequence of this is that a generator with a local monopoly has no game to win, because there is no other generator for it to take money off. It is indifferent to the local price.

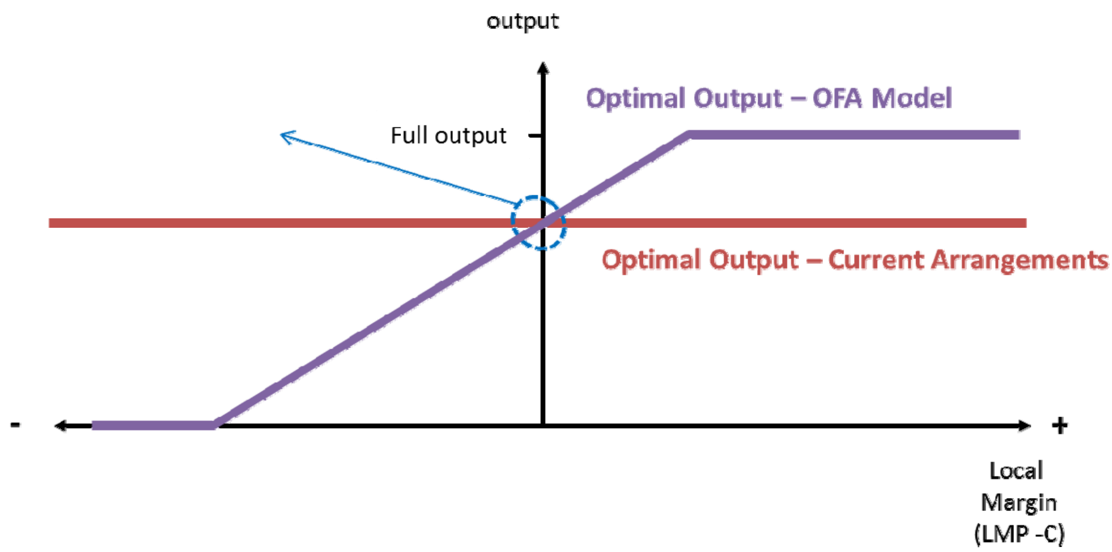
11.5.5 Comparison of the OFA model with the current arrangements

As noted above, the behaviour of unconstrained generators does not change under the OFA model. Therefore, any change in dispatch efficiency is associated with constrained generators.

Under the current arrangements, all constrained, in-merit (against RRP) generators bid a common price of -$1,000 and so NEMDE is unable to choose between them. Thus, each will be dispatched in proportion to its availability, with the cost of generation playing no role. This is similar to the dispatch outcome that would occur if every generator bid a *completely flat* output curve, with the intercept being that same proportion of availability.

Compare this to what would occur under the OFA under a scenario where all generators were firm. Each would have access scaled back in proportion to capacity. Each generator will then bid a *sloping* output curve, with the intercept being the scaled access level. Thus, compared to the current arrangements, the intercept is the same but the slope is greater. Thus, irrespective of how low supply elasticity is, any dispatch inefficiency will always be less than the corresponding dispatch inefficiency under the current arrangements: a *reduced* slope is always better than *no slope at all*.

Figure 11.5 Radial constraint bidding



Admittedly, this comparison does not allow for the fact that, under the current arrangements, only in-merit (against RRP) generators will bid to be dispatched. In contrast, under the OFA model, *all* generators (including out-of-merit generators) may have some agreed access level. That means that, in principle, some out-of-merit generation might be dispatched under the OFA model, in contrast to the status quo.

However, in practice, at least when congestion is severe, there would be such a large gap between LMP and RRP that this is unlikely to occur: because even a generator with a sloping output curve will have zero output when the (negative) local margin is worse than LMP-RRP.

The above scenario does not allow for the likelihood that some generators will opt to be non-firm. During congestion, non-firm generators will run at a lower output than firm generators, other things being equal. This could lead to lower-cost *non-firm* generators being incentivised to reduce dispatch and being replaced by higher-cost *firm* generators in dispatch, worsening dispatch efficiency. However, it is not anticipated that there will be any behavioural correlation between cost and firmness, so it is just as likely that the reverse occurs: ie that higher cost *non-firm* generators are incentivised to reduce dispatch and thus improve dispatch efficiency. So it is not clear that the *mixing* of firm and non-firm generators will systematically worsen dispatch efficiency compared to the alternative scenario of *all* firm generation.

Because the major impacts on dispatch efficiency are likely to be from major congestion involving a large proportion of regional generation, it is worth focusing on outcomes under these conditions. Such major congestion is likely to give rise to high RRPs, both because of the reduced supply available to the region, and because the lower supply elasticity from unconstrained generators exacerbates the impact of optimal regional bidding on RRP.

Under the current arrangements, both these factors mean that there will be a large amount of constrained generation – potentially with quite diverse costs – engaged in disorderly bidding.

On the other hand, under the OFA model, because local supply elasticity is likely to be relatively higher under major congestion, optimal local bidding will be closer to in-merit (against LMP) and so dispatch inefficiency may be relatively low. Despite the high RRP, LMP behind the constraint is likely to remain low, so high cost constrained generators will stay out of dispatch.

11.6 Bidding Strategies under a Loop-flow Constraint

11.6.1 Features of a Loop-flow Constraint

In general, constraints are not radial, but instead involve generators with participation factors that are neither one nor zero. This more general constraint may be referred to as a *loop-flow constraint*, because it arises where one or more generators lie on a looped path of transmission lines which also contains the RRN and the congested flowgate. Participation factors can be anywhere between minus one and plus one. Only positive participation factors are considered in this section. Negative participation factors are considered in section 11.7 (Bidding Strategies for flowgate support generators).

11.6.2 Optimal Output under the Current Arrangements

Under the current arrangements, optimal bidding of constrained generators will be similar to that under radial congestion. The difference is that the pricing influence of a constrained generator may be geared – ie scaled up or down – depending upon its participation factor. This gearing effect is explained below.

Define two participation factors:

> α_m = participation factor of the *marginal constrained generator*
>
> α = participation factor of the generator whose bidding we wish to analyse

The marginal constrained generator is the one whose output NEMDE changes to balance the flowgate constraint.

Suppose that our generator rebids to decrease its output by 1MW. This causes a reduction in flowgate usage of α, which allows the marginal generator to increase its output by an amount $\alpha m/\alpha$.

For example, suppose that:

> α_m=0.5
>
> α = 0.2
>
> *FGX* = 1,000MW

If the flowgate is congested before the rebid, then total flowgate usage (FGU) must equal 1000MW. If our generator rebids to reduce output by 1MW, its flowgate usage reduces by 0.2MW, and so FGU is now 999.8MW. The spare flowgate usage allows NEMDE to increase the output of the marginal generator[23] by 0.4MW, which increases flowgate usage by 0.2MW and brings FGU back to 1000MW.

In this example, one constrained generator reduces output by 1MW and another increases output by 0.4MW, so the net *reduction* in the aggregate output of constrained generators is +0.6MW. Thus, the output of *unconstrained* generators must *increase* by 0.6MW, which will lead to an increase in RRP. Thus, the 1MW reduction in output from our generator has led to an increase in unconstrained generator output of just 0.6MW.

In general, the net reduction in constrained generator output prompted by a 1MW rebid down is:

> net reduction = $1 - \alpha/\alpha_m$

---

23 Since the generator is marginal against LMP, which is less than RRP, it is worthwhile increasing the output of this generator, which costs LMP, to offset the output of the marginal unconstrained generator, which by definition costs RRP.

Where:

$$\gamma \equiv (1 - \alpha/\alpha_m) \times S/S_U \tag{11.19}$$

In section 11.4.2 (Choosing the Optimal Output), the optimal regional bidding of a generator was expressed in terms of forward level, regional margin and pricing sensitivity:

$$G = F + M/P$$

Under a loop-flow constraint, a constrained generator continues to be paid RRP and so its optimal output is determined by the same formula. All that has changed is the pricing sensitivity. Therefore, optimal output for a constrained generator is:

$$G = F + S/\gamma \times M$$

Where:

F is the generators forward level

M is the regional margin = RRP – generating cost

γ is determined by equation 11.19 above

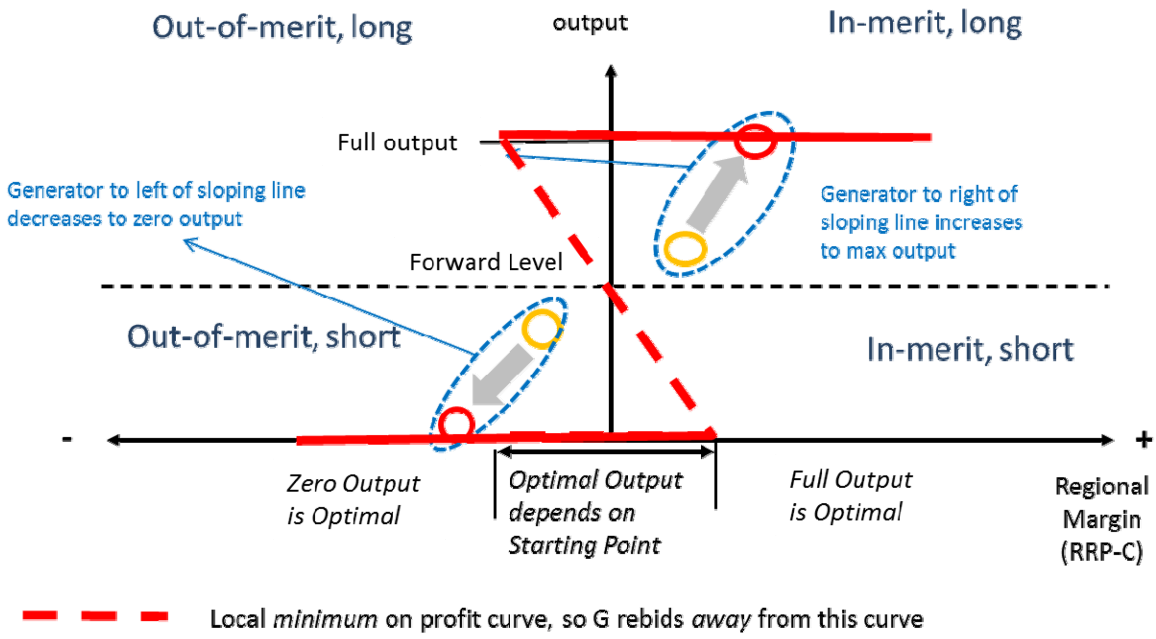
There are a number of cases and categories to consider:

Table 11.2 Optimal Output under the Current Arrangements Scenarios

Case	*α*	*γ*	*comment*
unconstrained	0	1	normal optimal regional bidding
low participation	$0<\alpha<\alpha_m$	$0<\gamma$	normal pricing power, with gearing possibly greater than or less than one
high participation	$\alpha>\alpha_m$	<0	*reversed* pricing power
radial constraint	$\alpha=\alpha_m=1$	0	zero pricing power: in-merit bidding
support generator	$\alpha<0$	>1	*enhanced* pricing power

Apart from the negative participation generator (which is discussed in a later section below), the most counter-intuitive result is the high participation generator which has a *reversed* pricing power: meaning that an *increase* in its output actually causes an *increase* in RRP. The optimal output for such a generator is illustrated in Figure 11.6.

Figure 11.6 Reserve Pricing Power



For such a generator, the sloping section of the optimal output curve represents a local profit *minimum* and a generator will always seek to move *away* from it. This will typically mean that a generator starting to the left of the line will keep decreasing its output until its hits minimum output[24] and a generator starting to the right of the line will keep increasing its output until it hits maximum.[25]

When congestion commences, there are two effects on a constrained generators gearing and hence its bid output:

- the lower local supply elasticity causes a flattening of the output curve; and
- the participation effect causes a steepening[26] or even a reversing of the slope.

The onset of congestion is likely to cause an increase in RRP, driving all generators that are already operating further into merit: ie to the *right* of the optimal output graph. Thus, the high participation generators (with reverse pricing power) will rebid to full output. Given their high participation factors, this higher output will exacerbate the congestion. Since congestion is best relieved by off-loading high participation generation, these generators will find themselves constrained off and quickly rebid lower offer prices, triggering disorderly bidding.

Given the equal offer prices submitted under disorderly bidding, AEMO will seek to dispatch the lowest participation generators first, since these use less of the scarce flowgate capacity. By definition, the marginal generator is then the last one dispatched.

24 Or until the flowgate congestion is removed or the marginal generator changes.

25 This is a little confusing, as it implies two different possible outputs (either maximum load or minimum load) for the same margin level, depending upon the starting point.

26 Apart from for support generators.

Generators with a higher participation factor than the marginal generator, despite having reverse pricing power and so wishing to be fully dispatched, will be fully constrained off.

Having been forced away from their optimal output curves and into disorderly bidding, constrained generators will be generally unresponsive to changes in RRP. Thus, pricing sensitivity of RRN will be determined solely by the supply elasticity from unconstrained generators, similar to the situation under a radial constraint: ie

$$\gamma = S/S_U \qquad (11.20)$$

11.6.3 Looped constraints under the OFA model

The profit equation for a generator under the OFA model is:

Profit = RRP margin + forward income + access settlement

= G × (RRP – C) + F × (FP-RRP) + (A-G) × (RRP-LMP)

= F × (FP-C) + (A-F) × (RRP-C) + (G-A) × (LMP-C)

= forward profit + regional profit + local profit

Previous sections have considered:

- an uncongested region, where regional and local margins collapse into a single term: *(G-F)* × *(RRP-C);* and
- a radial constraint, where the regional margin is outside the influence of the constrained generator and so only the local margin is variable.

In these two scenarios, there is just one variable term, which depends on RRP or LMP, respectively. Under a loop-flow constraint, both regional margin and local margin are variable and so both RRP and LMP effects need to be considered. This makes analysis of bidding strategies rather more complicated.

Because both regional margin and local margin are variable, a generator will seek to maximise the sum of these. Thus, it is trading off:

- the *direct* change in local profit from a change in its output: this depends upon the *local* margin, LMP-C;
- the *indirect* change in *local* profit from the change in LMP caused by the change in output; and
- the *indirect* change in the *regional* profit from the change in RRP caused by the change in output.

Similar to the discussion in section 11.4.2 (Choosing the Optimal Output), profit is maximised where the gain from the direct effect exactly matches the losses from the indirect effect. However, there are now two indirect effects, so the optimal output equation becomes:

$$M = P_{RRP} \times (A\text{-}F) + P_{LMP} \times (G\text{-}A) \qquad (11.21)$$

Where:

$M = LMP - C$ = local margin

P_{RRP} = sensitivity of RRP to change in generator output

P_{LMP} = sensitivity of LMP to change in generator output

If, for simplicity, we assume that P_{RRP} and P_{LMP} are constant over the output range of the generator, then equation 11.21 above shows that the optimal output G is a linear function of margin, with inverse slope equal to P_{LMP}, similar to the optimal output under a radial constraint. However the *intercept* level – the optimal output when the margin is zero – has changed. In the radial constraint, the intercept level equalled the access level A. In the loop constraint, the intercept is moved from the access level by an amount which will be labelled δ:

$$\text{Intercept} = A + \delta$$

Giving an optimal output curve:

$$G = (A + \delta) + M/P_{LMP} \qquad (11.22)$$

Using the same definition of gearing and supply elasticity as in previous sections, this equation can be rewritten as:

$$G = A + \delta + M \times \gamma/S$$

Where:

$$\gamma = S/P_{LMP}$$

By combining equations (11.21) and (11.22) above it can be shown that:

$$\delta = (F\text{-}A) \times P_{RRP}/P_{LMP} \qquad (11.23)$$

Therefore, in scenarios where P_{LMP} is large compared to P_{RRP} (ie under minor constraints), δ will be small and the intercept will be close to the access level. On the other hand, under major constraints, the local and regional price sensitivities are likely to be similar, meaning that the intercept is closer to the forward level, F.

In the previous section, it was demonstrated that, under the current arrangements, gearing was a function of participation. This is also the case under the OFA model, but the relationship is more complex and the algebra required to derive it is not presented here. In summary, gearing is determined by the formula:

$$\gamma = 1 + \varepsilon^2 \qquad (11.24)$$

Where:

$\varepsilon = (\alpha - \alpha_{mean})/SD_{\alpha}$

α is the participation factor for the relevant generator

α_{mean} is the average participation factor of marginal generators[27]

SD_{α} is the standard deviation of participation factors for marginal generators

Marginal generators are those that are operating on the sloping part of their optimal output curve. Mean and SD are calculated using weighting factors based on the supply elasticity of each marginal generator.

ε is a measure of the extent to which the generator's participation factor is different to the *mean* participation of all marginal generators, expressed in terms of the number of standard deviations from the mean. It is clear from equation 11.24 that γ is always greater than one. In particular, no generator has negative gearing: ie reverse pricing power.

The equation above applies to all generators in the region, including *unconstrained* generators, which of course have a participation factor of zero. The calculation of average and standard deviation also includes these zero participation generators. Thus, for *minor* constraints, the mean participation factor and also the standard deviation are likely to be both low. This means that generators with high participation factor will have the higher gearing and may in some cases have *very high* gearing. These generators will tend to operate close to their intercept levels which, as noted above, will be close to the access level. Generators with low participation factor will be close to the average, have low gearing, and so operate closer to in-merit levels.

On the other hand, for major constraints, the mean participation is primarily determined by the participation of constrained generators. In this situation, gearing factors are generally lower, with those generators having the more extreme participation factor – whether very high or very low - having the highest gearing levels.

The gearing of unconstrained generators, also given by equation 11.24 above, will generally be lower than under the current arrangements (as given by equation 11.20). This is because constrained generators remain responsive to changes in RRP, through the corresponding changes in LMP, and so *add* to the supply elasticity provided by the unconstrained generators, reducing the price sensitivity of RRP.

11.6.4 Comparison of the OFA model with the current arrangements

On a loop-flow constraint, efficient dispatch requires that both participation and cost are taken into account when dispatching generators. Other things being equal, the lower the cost and/or the lower the participation, the higher the dispatch should be.

---

27 Similar to α_m in the previous section, but allowing for multiple marginal generators.

Under current arrangements, participation but not cost are taken into account, because all generators bid at the same offer price. Under the OFA model, other things being equal, a generator with lower participation and/or lower cost will have a higher local margin and so have a higher optimal output. However, the responsiveness to margin depends upon gearing. Thus, as with a radial constraint, higher gearing can lead to reduced dispatch efficiency.

Again, dispatch efficiency under major congestion will be the most critical test of the OFA model. Under such conditions, gearing of both constrained and unconstrained generators will be relatively low: mainly because of a higher SDα under this scenario Thus, not only will the dispatch of constrained generators be more efficient than under disorderly bidding, but the dispatch of unconstrained generation will be more efficient too, because of their lower gearing under the OFA model.

11.7 Bidding Strategies for flowgate support generators

11.7.1 Introduction

Flowgate support generators have *negative* participation in a flowgate and so their output *relieves* congestion. Under the OFA model, flowgate generators are paid RRP for output, as they are under the current arrangements. Nevertheless, the different treatment of constrained generators under the OFA model can have an indirect impact on support generator bidding strategies.

By definition, flowgate support generators cannot exist under radial constraint, where all participation factors are +1. Therefore, the analysis in this section builds upon the analysis of loop-flow constraints in the previous section.

The OFA model places incentives on TNSPs to efficiently manage flowgate capacity, including through negotiating network support agreements with support generators. Clearly, this will have a direct effect on their bidding. However, this factor is not considered further in this section.

11.7.2 Flowgate Support generators under current arrangements

The situation for a support generator is structurally similar to that for a conventional constrained generator. It has an optimal output curve driven by a gearing factor which is determined by the formula derived in section 11.6.2 (Optimal Output under the Current Arrangements):

$$\gamma = (1 - \alpha/\alpha_m) \times S/S_U$$

Because α is less than zero, gearing is now greater than S/S_U. In some circumstances – where α is highly negative and α_m is low – gearing can be *much greater*. This leads to a much flatter optimal output, which remains close to the generators forward position, even where the generator is substantially away from the margin: whether in-merit or out-of-merit. Thus a flowgate support generator with high forward cover will have a high output and so help to relieve congestion. On the other hand, a support generator with low forward cover will reduce output and so exacerbate congestion.[28]

In the NEM, generators have a right *not* to be dispatched, so flowgate support generators do not get caught in the disorderly bidding spiral that prevents constrained-off generators from achieving their optimal output. A support generator can simply bid zero offer price (say) and limit its availability to its optimal output, rebidding to change this as optimal output changes.

11.7.3 Flowgate Support generators under the OFA Model

Because support generators earn RRP under the OFA model, the analysis of bidding strategy is quite similar. The major difference is that the higher supply elasticity at the RRN under the OFA model (as noted in section 11.6.3 (Looped constraints under the OFA model) will *reduce* the gearing of the flowgate support generator compared to the current arrangements.

11.7.4 Comparison of the OFA model with the current arrangements

Under both regimes, a flowgate support generator has an enhanced level of gearing compared to an unconstrained generator. However, because under the OFA model an unconstrained generator's gearing is lower, the flowgate support generator's gearing is similarly reduced. This mitigates the problem of a flowgate support generator with a low forward level choosing a low output and so exacerbating congestion.

11.8 Impact of Bidding Strategies on Hybrid Flowgates

11.8.1 Features of a hybrid flowgate

A hybrid constraint is one in which generators *and* one or more interconnectors participate. Interconnectors do not explicitly bid into NEMDE and are dispatched by NEMDE in a region as though their offer price were the RRP of the neighbouring, interconnected region. Thus, they cannot engage in strategic bidding and are constrained to effectively "bid" in-merit.

11.8.2 Current Arrangements

Under disorderly bidding, as other generators rebid to -$1,000 – but an interconnector continues to *bid* in-merit - an interconnector *appears* to NEMDE to be at least $1,000/MWh more expensive to dispatch than the constrained generators. As a result, it either becomes marginal or is dispatched to minimum output. The minimum output for an interconnector is *usually* negative: ie the interconnector flow direction is

---

28 Incidentally, this suggests that a portfolio generator who has constrained-off generation and also support generation is likely to increase output on its support generator, to offset the loss of output from the constrained-off generators, and so relieve congestion.

switched to *exporting*.[29] Since the congestion in the region often causes a high RRP, this exporting flow will often be *counterprice*, giving rise to a negative settlement residue.

Under a looped constraint, generators are dispatched based on participation as well as offer price. Because of its higher effective offer price, even an interconnector with a low participation factor is likely to be constrained off before the -$1,000-bidding generators.

This situation can impact substantially on supply to the RRN and hence on the level of RRP. For example, suppose that an interconnector with participation factor of 0.2 is displaced in the merit order by generators with participation factors of 0.5, and the interconnector flow is reversed from 1,000MW importing to 1,000MW exporting. Supply to the RRN has reduced by 2,000MW but flowgate usage has only reduced by 2,000MW x 0.2 = 400MW. This means that the constrained generation can only increase output by 800MW (since this gives additional flowgate usage of 800MW x 0.5 = 400MW). The shortfall of 1,200MW must be replaced by increased output from unconstrained generators. This may have a substantial impact on both dispatch efficiency and RRP.

11.8.3 OFA Model

Under the OFA model, the need and incentive for disorderly bidding is removed. Although there will be some strategic bidding – as described in section 11.6.3 (Looped constraints under the OFA model) – the impact of this is fairly symmetrical, with firm generators seeking to increase output (compared to the efficient level) and non-firm generators seeking to decrease output.

Furthermore, if there is some level of agreed inter-regional access on the interconnector, the access levels of firm generators will be commensurately lower, since total access must always equal flowgate capacity. Thus if, under a high gearing scenario, constrained generators all have output close to their access levels, this should leave some *headroom* for the interconnector, up to the inter-regional access level.

Thus, it seems unlikely that constrained generators will have a strong incentive to maximise their output and squeeze the interconnector into a counterprice flow. In particular, under severe congestion, high participation generators will see a very low local LMP and are likely to back-off to minimum output irrespective of their gearing or access level. Therefore, the problem of extreme RRPs caused by inefficient counter-price flows is unlikely to arise.

11.8.4 Comparison of the OFA model with the current arrangements

Under the current arrangements, where there is congestion on a hybrid flowgate, the problem of disorderly bidding of constrained generators and higher gearing of unconstrained generators is further exacerbated by inefficient dispatch of the

29 Assuming that, prior to the congestion, the interconnector was importing. In situations where an interconnector was previously *exporting*, the distortion introduced by disorderly bidding is unlikely to affect interconnector dispatch.

interconnector due to the bidding anomaly that the interconnector, unlike generators, cannot bid at -$1,000. The resulting distortion to dispatch can be particularly severe on constraints where interconnector participation is low.

Under the OFA model, this bidding anomaly becomes much less significant, since disorderly bidding does not occur and generators will not wish to be dispatched substantially out of merit.

11.9 The Central Queensland Constraint

11.9.1 Features of the Constraint

A recent example of a major binding looped hybrid constraint in the NEM is the Central Queensland Calvale-Wurdong constraint ("Central Queensland constraint"), which has given rise to disorderly bidding, counterprice flows and high RRPs. It is understood that this constraint has the following features:

- a low participation interconnector;
- a high participation flowgate support generator;
- a large amount of constrained generators (in CQ and SWQ zones);
- some high participation constrained generators (in CQ); and
- congestion is triggered by relatively small changes in flowgate capacity: associated with a *dynamic* rating of the associated transmission line.

The contract cover of the flowgate support generator is not known. However, if this were fairly low[30] this constraint has all of the qualities that would give rise to dispatch inefficiency and high RRP under the current arrangements.

11.9.2 Current Arrangements

The analysis above suggests the following behaviour and outcomes during congestion on the Central Queensland constraint.

Firstly, congestion quickly leads to disorderly bidding, for the reasons described in section 11.6.2 (Optimal Output under the Current Arrangements). Because there is the possibility for quite strong positive feedback – as the reverse pricing power of the in-merit, high participation generators leads them to seek maximum output and thus exacerbate the congestion – disorderly bidding can be triggered by quite modest congestion.[31]

---

30 And based on the analysis above, the forward level can be inferred from the output of this generator during the congestion.

31 There is also an incentive for generators to act quickly to maximise or maintain their output, before the rebidding of low ramp rate limits effectively freezes dispatch at pre-existing levels.

Secondly, despite it having a low participation, the interconnector's offer price disadvantage sees it quickly dispatched towards minimum output and, most likely, counterprice flow, as described in section 11.8.2 (Current Arrangements).

Thirdly, the consequent loss of supply at the RRN leads to a substantial increase in RRP (section 11.8.2 (Current Arrangements)), as well as a sharp decrease in, already low, supply elasticity of unconstrained generation, due to the convexity of the supply curve (section 11.4.3 (Supply Elasticity)).

The high RRP, coupled with the counterprice flows, may give rise to high negative settlement residues. AEMO will endeavour to clamp these flows, but this may be ineffective if low ramp rate limits have been bid by the constrained generation.

Given its enhanced gearing, together with the very low supply elasticity of unconstrained generation, the flowgate support generators is likely to generate close to its forward level. If the generator has a low contract cover, the resulting low output further reduces flowgate capacity and exacerbates all of the factors listed above.

Even if this is not an accurate description of *actual* behaviour in Queensland, it is an interesting hypothetical example of feedbacks and amplifications that might occur due to the interaction of interconnectors, constrained-off generators and flowgate support generators on looped constraints.

11.9.3 OFA model

Drawing on the analysis above, it is anticipated that the following outcome would have occurred under the OFA model.

As a major constraint, the gearing levels of constrained generators are unlikely to be high (section 11.6.3 (Looped constraints under the OFA model)). To the extent that there is pricing power, generators will tend to move towards their access level (section 11.6.3 (Looped constraints under the OFA model)) with high participation and/or non-firm generators being the main ones backed-off. Given the fairly modest congestion initially (due to changing line rating rather than a line outage), moves in generator output from the pre-congestion level may be relatively small.

Given its low participation and low cost, the interconnector is unlikely to be backed off. However, even if counterprice flows were to occur, this would not lead to negative settlement residues or AEMO intervention (section 9.2.4 (Inter-regional access settlement)).

Given the more efficient dispatch of the interconnector, the supply shortfall at the RRN will not be exacerbated and so extreme RRPs are likely to be avoided.

Given the lower pricing power of the flowgate support generators than under the current arrangements, both intrinsically to the OFA model and because of lower RRPs, the flowgate support generator will bid closer to in-merit against RRP. Thus, to the extent that the congestion causes high RRPs, the flowgate support should be dispatched and should so relieve the congestion.

11.9.4 Comparison of the OFA model with the current arrangements

By introducing local bidding and removing disorderly bidding, the OFA model ensures that all generators are responsive to changes in RRP and congestion, and do not operate too far from their in-merit level. This provides three benefits. Firstly, improved dispatch efficiency of constrained generators. Secondly, lower RRP price sensitivity and thus more efficient bidding and dispatch of unconstrained and flowgate support generators. Thirdly, more efficient dispatch of the interconnector, as it no longer has to compete with generators bidding -$1,000.

11.10 Differences between the OFA Model and Nodal Pricing

In a nodal pricing market design, all generators are paid LMP and all customers pay LMP. Although *theoretically* efficient,[32] nodally priced markets have a number of shortcomings:

- generators with local market power can increase prices to customers and increase the market price of congestion hedges;[33]
- retailers and generators find it difficult to contract forward, since each faces a different spot price;
- congestion hedges may be unavailable or may not be effective in managing congestion risk; and
- as a result, generators and retailers typically vertically integrate along zonal lines in order to naturally hedge congestion, leading to reduced retail competition.

Although LMPs are implicitly used in the OFA model, as discussed, the OFA model is very different to a nodal pricing design. Specifically:

- customers pay regional prices, not local prices;
- generators *never* get paid higher than the regional price, which limits their pricing influence;
- generators can obtain firm access to the regional price by agreeing access with TNSPs;
- TNSPs are obliged to offer firm access at regulated prices;
- TNSPs must expand the network to accommodate firm access; and

32 Under strict assumptions such as perfect competition and generator and customer risk neutrality.

33 Financial contracts that hedge LMP price differentials.

- increased severity of congestion is likely to prompt additional access procurement, leading to network expansion and thus a reduction in congestion: ie congestion levels are self-regulating.

As a result:

- optimal bidding is as likely to cause LMPs to reduce as increase;
- optimal local bidding is a zero sum game between generators, with customers largely unaffected; and
- firm access will give generators increased confidence to locate remotely from their customers.

In summary, there are fundamental differences between the OFA model and a conventional, nodally priced market design. Generator behaviour under the OFA model should not be inferred from analysis or experience of nodal markets, but rather by specific analysis of the elements and drivers present in the OFA model.

12 Technical Detail

12.1 Overview

This section provides some additional design detail for the OFA model. This detail is not really necessary for understanding the OFA model design. However, it helps in explaining how the model achieves its goals and objectives, how it is applied in unusual situations, or how the design would be implemented in practice.

12.2 Flowgate Pricing and Local Pricing

12.2.1 Overview

There is something of a disconnect between the high level description of the OFA model (see chapter 2 (Access)) and the more detailed design (as described in chapter 4 (Access Settlement)). 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 ensures the *no regrets* principle: that all generators are paid, net, at least their offer price irrespective of their level of access or entitlements.

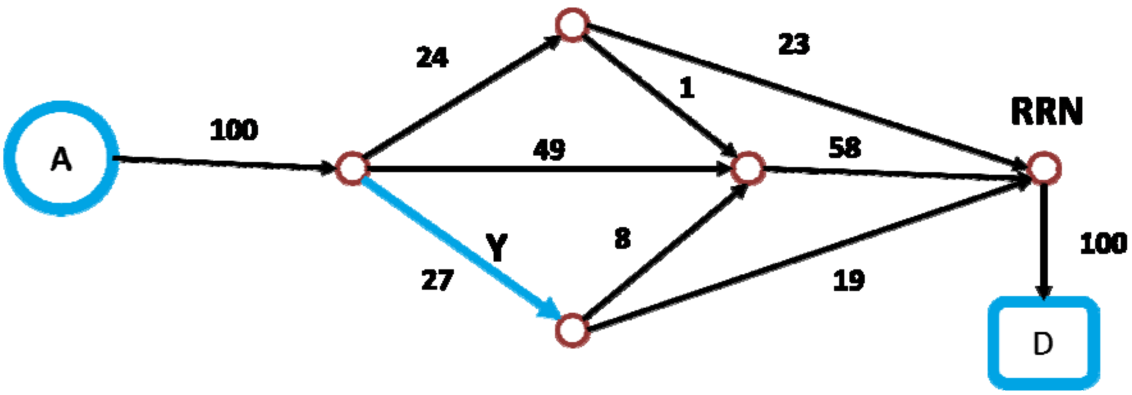
12.2.2 Flowgate Participation

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.
- *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 fluid flow through a network of pipes. Figure 12.1, below, illustrates a transfer of power through a network, from a generator node to the RRN, based on the lossless DC approximation:

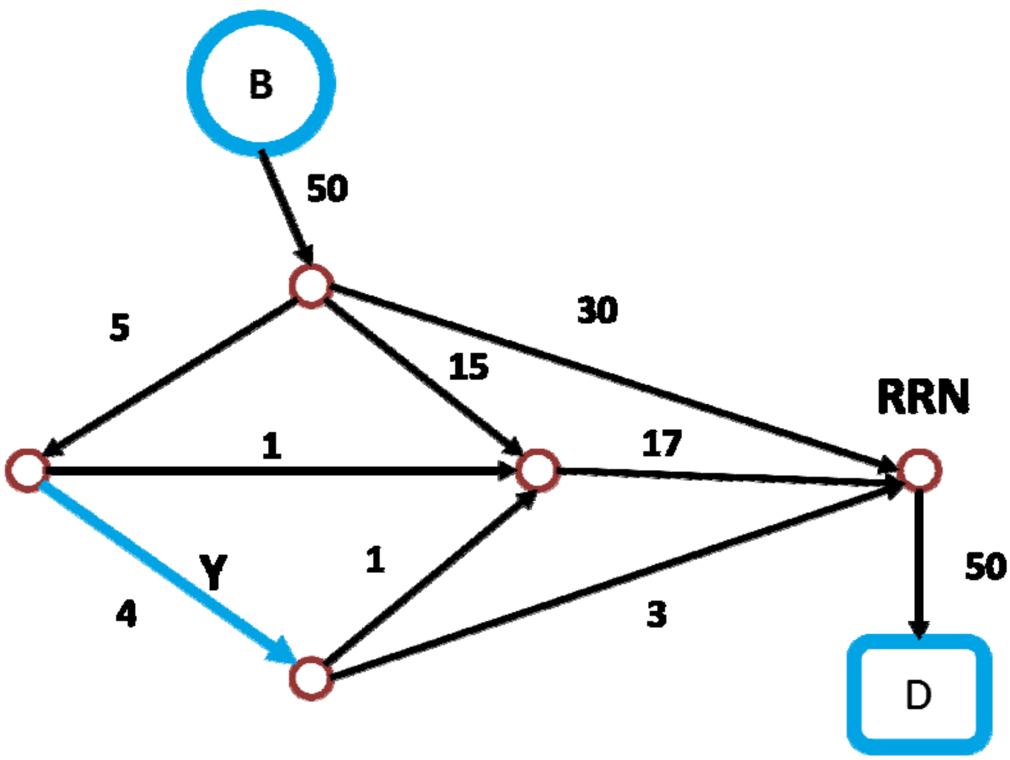
Figure 12.1 Load flow example 1



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 flowgate, labelled *flowgate Y*, is marked on the figure. 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 potentially be congested in either direction, meaning there are two (directed) flowgates on each line.

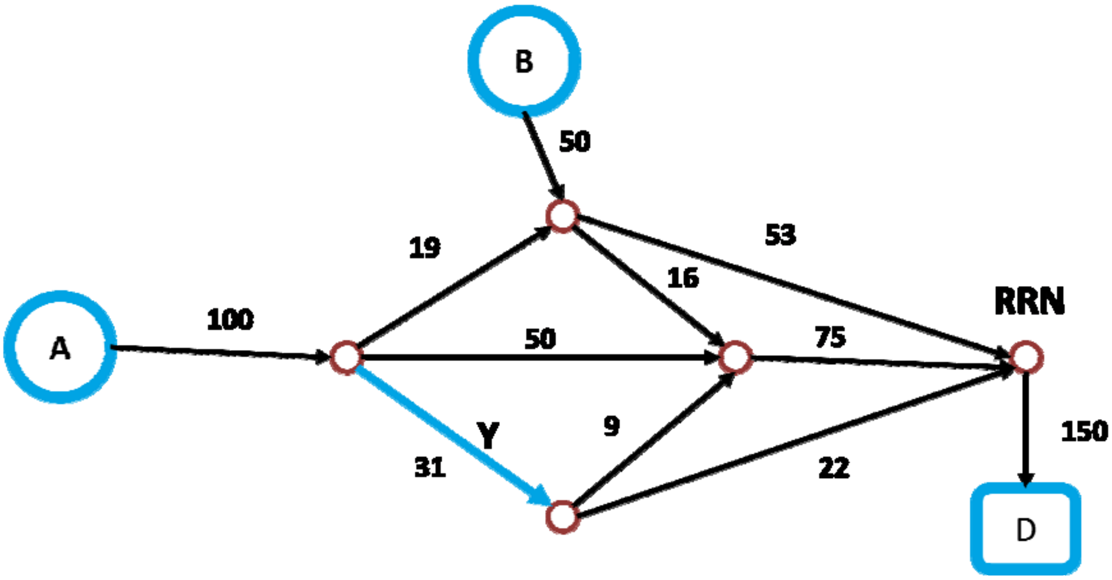
A similar load flow for generator B is presented in Figure 12.2, below. Because the generator connects at a different node, it has a different participation factor. The participation of generator B in flowgate Y is 4MW divided by 50MW, or 8 per cent.

Figure 12.2 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 the load flows presented in Figure 12.1 and Figure 12.2 are superimposed, a third load flow is created as shown in Figure 12.3.

Figure 12.3 Superimposed load flows



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% x 50

In general, the flow on the flowgate is:

$$\text{Flowgate Y Flow} = 27\% \times G_A + 8\% \times G_B \qquad (12.1)$$

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 *B* 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:

$$\text{Flowgate Y flow} = 27\% \times G_A + 8\% \times G_B \leq TX_Y \qquad (12.2)$$

Where:

TX_Y is the transmission capacity of the line of flowgate *Y*

This inequality has the familiar form of the transmission constraints used in NEMDE. The general version of this inequality is:

$$\sum_i \alpha_{ik} \times G_i \leq TX_k$$

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*

12.2.3 Transmission Capacity versus Flowgate Capacity

Suppose that there is some local demand at nodes A and B. For our simplified model, the impact of demand on network flows is exactly the same as negative generation: ie it is the *net injection*, *G-D*, at each node that determines flows; generation of 100MW will create the same flow as generation of 130MW and local demand of 30MW (since 130 - 30 = 100).

When local demand is included, inequality (12.2) becomes:

$$27\% \times (G_A - D_A) + 8\% \times (G_B - D_B) \leq TX_Y \qquad (12.3)$$

Where:

D_A is the local demand at node *A*

D_B is the local demand at node *B*

Demand is not dispatched by NEMDE and so it is treated as a constant and moved to the RHS of the inequality. Therefore, equation (12.3) becomes:

$$27\% \times G_A + 8\% \times G_B \leq TX_Y + 27\% \times D_A + 8\% \times D_B \equiv FGX_Y$$

Where:

$$FGX_Y = TX_Y + 27\% \times D_A + 8\% \times D_B \qquad (12.4)$$

FGXY is the *flowgate capacity* for flowgate Y. It is seen that flowgate capacity is a combination of the transmission capacity and local demand. The general form of equation (12.4) is:

$$FGX_k = TX_k + \sum_i \alpha_{ik} \times D_i$$

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

It is seen from the above formula that flowgate capacity will vary as local demand varies. When TNSPs are managing their networks so as to maintain FAS, they must take account of local demand as well as transmission availability.

12.2.4 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*. 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. That is because, despite their quite different origins, they have the same general form in NEMDE as thermal constraints. Indeed, they must take the same form, because that is the only form of equation that NEMDE is designed to solve.

This algebraic commonality between thermal and stability constraints 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.

12.2.5 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.[1] Therefore:

---

1 Or, alternatively, the decrease in economic dispatch cost allowed 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.

$FGP = c(ED_2) - c(ED_1)$

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 chosen originally 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.
- *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.[2]

12.2.6 Local Prices

The local price[3] at a node is defined as *the marginal value of generation at a node*. As discussed previously, the impact of +1MW of generation is identical to -1MW of demand. Therefore, the local price is also the *marginal cost of supplying demand at a node*.

Using the same definition of marginal value as above, this means it is 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 local price 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.
- 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.

---

2 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.

3 Also referred to as locational marginal price, LMP or nodal price.

Thus, the local price defined above is a *clearing price*: the generator is dispatched if its offer price is below the local price and not dispatched if its offer price is above it.

12.2.7 Marginal Generators

Suppose that in an economic dispatch there is a *part-loaded* 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? Well, 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 local price is:

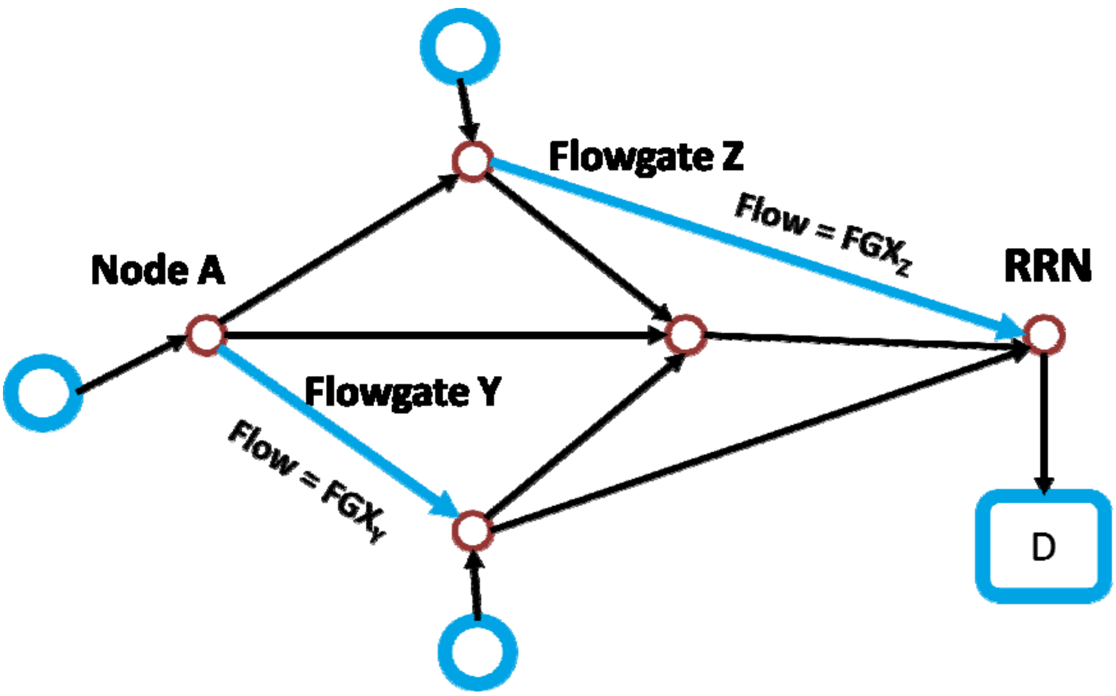
$$P_B = C_B$$

In general, whenever there is a part-loaded generator,[4] the generator's offer price sets the local price. Such a generator is referred to as marginal.

12.2.8 Relationship between Local and Flowgate Prices

Consider now the economic dispatch shown in Figure 12.4, below. Flowgates Y and Z are both congested and have corresponding flowgate prices FGP_Y and FGP_Z.

Figure 12.4 Economic Dispatch Example



[4] Or strictly a generator that is part-loaded within a dispatch offer band.

What is the local price at node A? The marginal value of generation at node A can be examined by considering changing economic dispatch in Figure 12.4 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

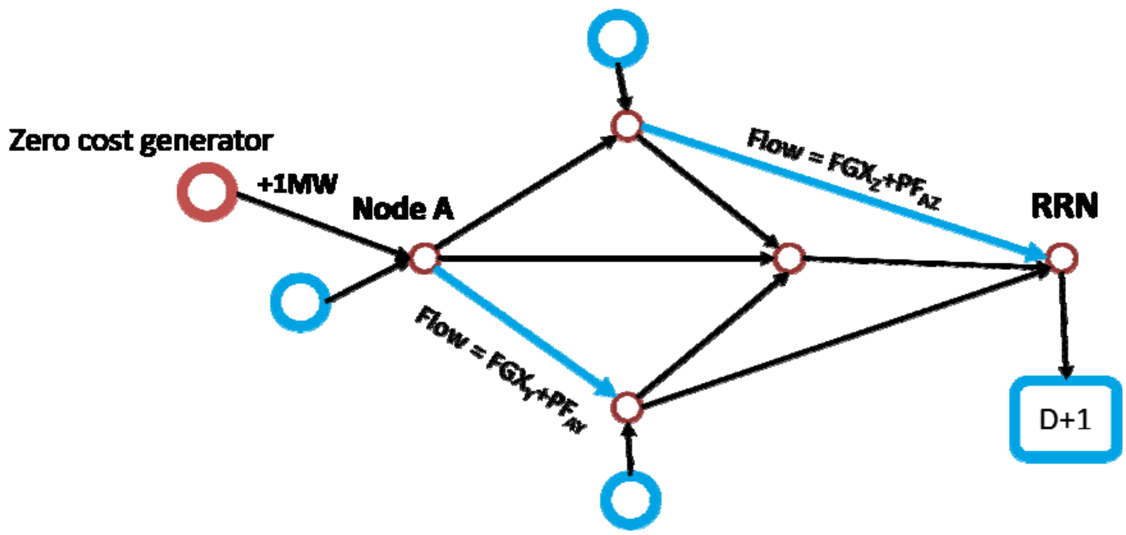
α_{AZ} is the participation of a generator at node A in flowgate Z

$Flow_Y$ 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 12.5:

Figure 12.5 Adjusted Dispatch (Infeasible)



This dispatch is *not feasible*, since the flowgate flows on Y and Z now exceed the flowgate capacity.[5] 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

5 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 local price:

- The extra 1MW of generation at node A *decreases* dispatch costs by P_A.
- The extra 1MW of demand at node R *increases* the dispatch cost by P_{RRN}.

P_{RRN} is the local price 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:

$$RRP - P_A = \alpha_{AY} \times FGP_Y + \alpha_{AZ} \times FGP_Z \qquad (12.5)$$

Rearranging this equation, PA is defined by the formula:

$$P_A = RRP - \alpha_{AY} \times FGP_Y - \alpha_{AZ} \times FGP_Z \qquad (12.6)$$

The example considers the situation of two congested flowgates, but the analysis applies irrespective of the number of congested flowgates. In general, then, the local price is defined by the formula:

$$P_i = RRP - \sum_k \alpha_{ik} \times FGP_k \qquad (12.7)$$

Where:

P_i is the local price 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 all congested flowgates or, equally, over all flowgates, recalling that the price of uncongested flowgates is zero.

12.2.9 Inter-regional Price Difference

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 (12.5) becomes:

$$RRP_2 - RRP_1 = \alpha_Y \times FGP_Y + \alpha_Z \times FGP_Z \qquad (12.8)$$

The participation factors α_Y and α_Z represent the amount of flow through flowgates Y and Z for a flow from RRN1 to RRN2. Such a flow is referred to in the OFA model as the *directed interconnector* from region 1 to region 2. Generalising equation (12.8) gives:

$$RRP_N - RRP_S = \sum_k \alpha_k \times FGP_k \qquad (12.9)$$

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 directed interconnector in flowgate *k*

12.2.10 Generator Access Settlement

Suppose that a non-firm generator at node *i* with zero entitlement on every congested flowgate is dispatched to a level *G*. In normal settlement it is paid *RRP* ×*G*. In access settlement it will pay an amount on flowgate k equal to:

$$\text{Access Pay\$}_k = U_k \times FGP_k$$

Where:

U_k = the generator's usage of flowgate k

Recall that usage is defined by:

$$U_k = \alpha_{ik} \times G$$

Therefore, the total payment is:

$$\text{Total access pay\$} = \sum_k \text{Pay\$}_k$$

$$= \sum_k \{(\alpha_{ik} \text{ x } G) \times FGP_k\} = G \text{ x } \sum_k (\alpha_{ik} \times FGP_k)$$

The generator's net payment in settlements is then:

$$\text{Net Pay\$} = G \times RRP - G \text{ x } \sum_k (\alpha_{ik} \times FGP_k)$$

$$= G \times \{RRP - \sum_k (\alpha_{ik} \times FGP_k)\}$$

$$= G \times P_i \qquad \text{(from equation (12.7))}$$

So, a generator with zero entitlement gets paid its *local price*, P_i. Since the generator is dispatched, the local price must be higher than its offer price, so the generator does not *regret* being dispatched.[6]

6 Assuming that its offer price is no lower than its generating cost.

In general, generators – even non-firm ones – will have *some* entitlements on congested flowgates and will receive a price higher than the local price.

Now suppose instead that the generator has an agreed access level A and receives its firm target entitlement on each congested flowgate:

$$E_k = \alpha_{ik} \times A$$

Its total payment is:

$$\text{Pay\$} = RRP \times G + \textstyle\sum_k\{(E_k\text{-}U_k) \times FGP_k\}$$

$$= RRP \times G + \textstyle\sum_k\{(\alpha_{ik} \times A - \alpha_{ik} \times G_k) \times FGP_k\}$$

$$= RRP \times G + (A\text{-}G) \times \textstyle\sum_k(\alpha_{ik} \times FGP_k)$$

$$= RRP \times G + (A\text{-}G) \times (RRP - P_i) \qquad \text{(from equation (12.7))}$$

$$= P_i \times G + A \times (RRP - P_i)$$

which is the formula for access presented in equation (2.1) in chapter 2 (Access).

Thus, access settlement correctly implements the principles set out in chapter 2 (Access).

12.2.11 IRSR Allocation

Ignoring losses as usual, the IRSR accruing on an interconnector is:

$$IRSR = IC \times (RRP_N - RRP_S)$$

Where:

IC is the interconnector flow in the northerly direction.

RRP_N is the RRP in the northern region

RRP_S is the RRP in the southern region

Substituting for the inter-regional price difference from equation (12.9), we get

$$IRSR = IC \times \textstyle\sum_k (\alpha_k \times FGP_k)$$

$$= \textstyle\sum_k \{(\alpha_k \times IC) \times FGP_k\} \qquad (12.10)$$

Now, recall from section 12.4 that:

- the northerly interconnector is considered to participate in flowgates where the participation factor is positive; and

- the southerly interconnector is considered to participate in flowgates where the participation factor is negative.

We will refer to these two subsets of flowgates as sets N and S respectively. Therefore, access settlement payments are:.

$$\text{Access Pay\$N} = \textstyle\sum_{k\epsilon N}\{(E_k - U_k) \times FGP_k\} \qquad (12.11A)$$

$$\text{Access Pay\$S} = \textstyle\sum_{k\epsilon S}\{(E_k - U_k) \times FGP_k\} \qquad (12.11B)$$

Where

E_k = is the entitlement on flowgate k for the relevant directed interconnector

U_k = is the usage on flowgate k for the relevant directed interconnector

Access Pay$\$_N$ is the access settlement payment to the northerly interconnector

Access Pay$\$_S$ is the access settlement payment to the southerly interconnector

Using those flowgate subsets, the Existing IRSR can be divided between the two directed interconnectors. Equation (12.10) can then be written as:

$$\text{Existing IRSR} = \textstyle\sum_{k\epsilon N}\{(\alpha_k \text{ x IC}) \text{ x } FGP_k\} + \sum_{k\epsilon S}\{(-\alpha_k \text{ x } -\text{IC}) \text{ x } FGP_k\}$$

$$= \text{Existing IRSR}_N + \text{Existing IRSR}_S$$

Where:

$$\text{Existing IRSR}_N \equiv \textstyle\sum_{k\epsilon N}\{(\alpha_k \times \text{IC}) \times FGP_k\} = \sum_{k\epsilon N}(U_k \times FGP_k) \qquad (12.12A)$$

$$\text{Existing IRSR}_S \equiv \textstyle\sum_{k\epsilon S}\{(-\alpha_k \text{ x } -\text{IC}) \times FGP_k\} = \sum_{k\epsilon S}(U_k \times FGP_k) \qquad (12.12B)$$

Note that this allocation of the IRSR between the two directed interconnectors is *not* the same as in the *current* NEM design, which allocates *all* of the IRSR to the directed interconnector which flows into the higher-priced region. There will need to be a change to existing NEM settlements to change the allocation to the one defined in equations (12.12) above.

In the OFA model, the existing settlement payments into the IRSRs are supplemented by the access payments shown in equations (12.12), meaning that total IRSR is:

$$\text{IRSR\$}_N = \text{existing IRSR\$}_N + \text{Access Pay\$}_N$$

$$= \textstyle\sum_{k\epsilon N}(U_k \times FGP_k) + \sum_{k\epsilon N}[(E_k - U_k) \times FGP_k]$$

$$= \textstyle\sum_{k\epsilon N}(E_k \times FGP_k) \qquad (12.13A)$$

And similarly:

$$\text{IRSR\$}_S = \textstyle\sum_{k\epsilon S}(E_k \times FGP_k) \qquad (12.13B)$$

We can consider equations (12.13) under two special situations. Firstly, if an interconnector has zero entitlement on all flowgates, the corresponding IRSR will be zero. This is the worst case scenario for IRSR: it can never be *less* than zero, because neither entitlements nor flowgate prices can be negative.

Secondly, consider a northerly interconnector, with agreed access, A, that receives its firm target entitlement, and no non-firm entitlement, on each flowgate, ie:

$E_k = \alpha_k \times A$

Then from equation (12.13):

$\text{Net Pay } \$ = \sum_{k \in N} (\alpha_k \times A \times FGP_k)$

$= A \times \sum_{k \in N} (\alpha_k \times FGP_k)$

A corresponding result holds for a southerly interconnector.

In the common case where all binding hybrid constraints on an interconnector are in the same direction (northerly say)

$\text{Net Pay } \$ = \sum_{k} (\alpha_k \times FGP_k)$

$= A \times (RRP_N - RRP_S)$ (from equation (12.9))

Thus, in this case, the access settlement payment converts the *existing IRSR*, which is proportionate to the interconnector *flow, IC*, and the inter-regional price difference, into an IRSR which is proportionate to the interconnector *access* level, A and the inter-regional price difference. A similar result again holds for the southerly interconnector, which may have a different agreed access level to the northerly interconnector. In general, the firm IRSR will be allocated to the directed interconnector in the direction of the binding constraints: ie which is *directed towards the higher price region*. The allocation does not depend upon the direction of interconnector *flow*.[7]

The more complex case where there is mixed congestion on the interconnector (both northerly and southerly binding flowgates) is discussed in section 12.3.3 (Mixed Interconnector Constraints) below.

12.3 Mixed Constraints

12.3.1 Overview

For generators and interconnectors, OFA access settlement treats flowgates differently depending upon whether the party 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.

7 Recalling that the interconnector flow is defined as the flow across the regional boundary, not the flow through the hybrid flowgates.

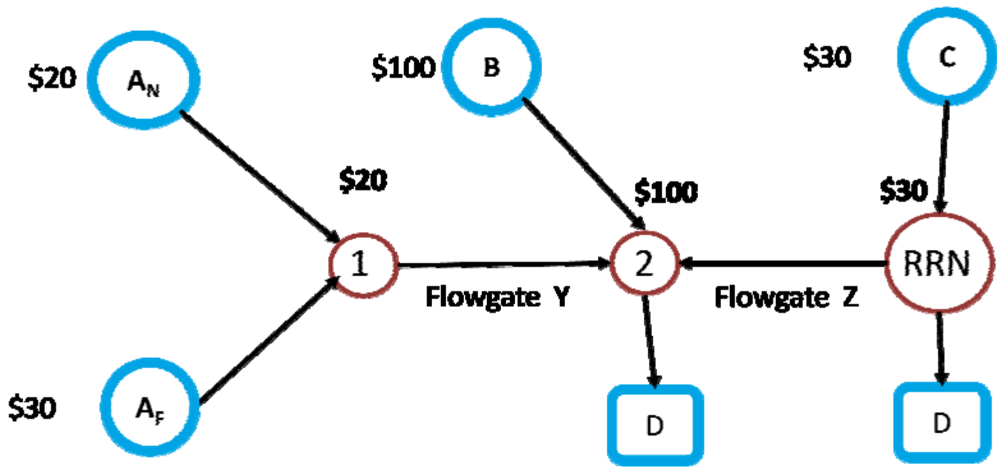
For generators, this scenario may be a largely theoretical situation which never, or rarely, occurs in practice. However, if quantitative modelling were to show that it could occur frequently, some changes may be needed to the OFA model design to address the issues arising.[8]

12.3.2 Generator Situation

Figure 12.6, below, presents a simple scenario where a generator participates in two binding flowgates.[9] Generators A_N and A_F each have positive participation in flowgate Y and a negative participation in flowgate Z, both of which are binding.

Issues arising under mixed constraint situations are discussed further below.

Figure 12.6 Mixed generator constraints



Assume that generator A_N is non-firm, and receives zero entitlement on flowgate Y. It is therefore charged the flowgate price on its output.[10] However, because it is a flowgate support generator for flowgate Z, it receives a negative entitlement and so has zero access settlement on the flowgate. Therefore, payments to generator A_N are:

$$\text{Pay\$}_N = \text{RRN settlement\$} + \text{Flowgate Y settlement\$} + \text{Flowgate Z Settlement\$}$$

$$= \$30 \times G - (\$100 - \$20) \times G + 0\$ = -\$50 \times G$$

Generator A_F receives an entitlement on flowgate Y which (since the generator is not dispatched) pays $80 ($FGP_Y$) multiplied by the access amount. Generator A_F is not dispatched and so receives no negative entitlement on flowgate Z. Therefore, generator

8 Which would be along the lines discussed in section 2.3.9 (Flowgate Support and Constrained-on Generators).

9 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.

10 Since the flowgate is radial, usage equals output.

AF is compensated at *$80* for being constrained off, despite its opportunity costs being only $10.

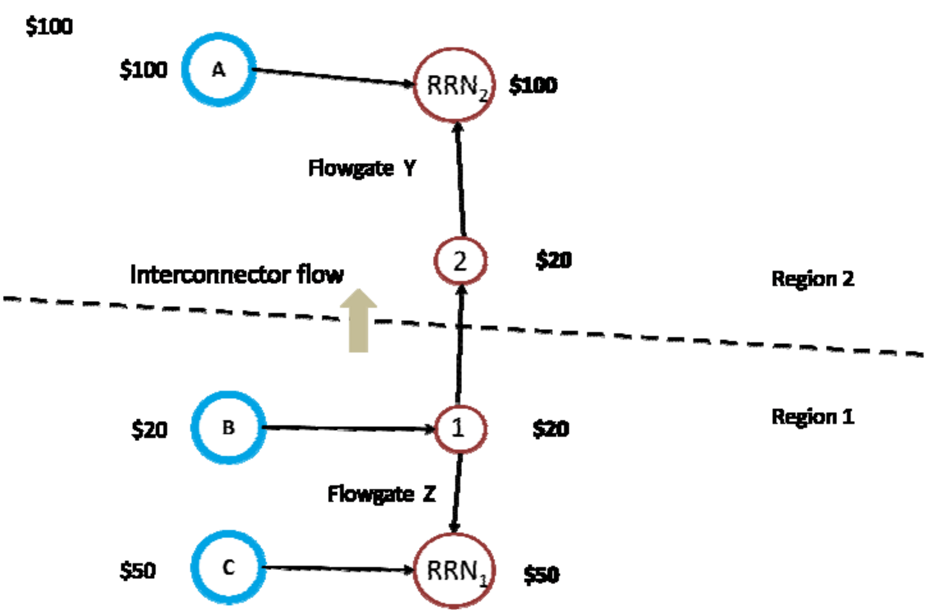
So, in this example, a non-firm generator is not just paid less than its *offer price*, it is paid less than *zero*. A firm generator, on the other hand, receives substantially more compensation for being constrained off than is justified. The root of the problem is not in the OFA model, *per se*, but in the treatment of flowgate support generators in the current NEM design,[11] where the *no regrets* principle does not apply for constrained-on generation.

Generator A_N in this example would clearly reduce its offered quantity until either flowgate Y was no longer binding, 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.

12.3.3 Mixed Interconnector Constraints

Figure 12.7, below, presents an example of an interconnector facing mixed flowgates.

Figure 12.7 Mixed interconnector constraints



Recall that directed interconnectors never have negative participation in flowgates, because each directed interconnector is deemed to participate only in positive-participation flowgates. Therefore, in the example, the northerly interconnector participates in flowgate Y and the southerly interconnector participates in flowgate Z.

11 Which the OFA model retains.

As explained in section 12.2.11 (IRSR Allocation), 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
- 200MW for the southerly interconnector

and assume also that firm entitlement targets are met and that there are no non-firm entitlements. In this situation, IRSR allocation is as follows:

$$IRSR\$_N = A_N \times FGP_Y = 100MW \times (\$100-\$20)$$

$$IRSR\$_S = A_S \times FGP_Z = 200MW \times (\$50-\$20)$$

Where:

A is the agreed access

FGP is the flowgate price

suffixes N and S refer to northerly and southerly, respectively

suffixes Y and Z refer to flowgates Y and Z respectively

The IRSR payouts are based on price differences of $80 and $30, respectively, neither of which matches the inter-regional price difference of $50. That might suggest that these IRSR payouts – and the associated SRA and FIR instruments – do not act as effective inter-regional hedges.

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 IRSR 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 IRSR will provide an effective southerly inter-regional hedge at such times.

Recall that, under the OFA model, IRSR allocation and amount do not depend upon the direction of interconnector *flow*.[12] In the situations described above the flow could remain northerly or turn southerly; it would not affect the IRSR 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.

---

12 The flow across the regional boundary.

12.4 Thirty-Minute Settlement

12.4.1 Overview

Access settlement uses trading intervals (TIs), which are 30-minute periods, for settlement. However, the settlement calculation is based on dispatch information which is defined by dispatch interval (DI): a 5-minute period. There is therefore a need to define settlement algebra that reconciles these two bases.

This is not a new problem: current NEM settlement addresses similar issues. Indeed, it is important to ensure that, as far as possible, access settlement adopts the same approach to addressing the issue, so as not introduce new basis risks resulting from DI-TI discrepancies.

12.4.2 Access Settlement Approach

Table 12.1 below shows the dispatch data used by access settlement and how this is converted from 5-minute dispatch information to 30-minute settlement inputs.[13]

Table 12.1 Conversion of DI data to TI data

TI Settlement data	Processing of dispatch data
Flowgate applied in NEMDE	if transmission constraint applied in one or more DIs
Congested Flowgate	if applied constraint is binding in one or more DIs
Flowgate Participation	from static data in constraint library
Generation Dispatch	average of DI dispatch across TI
Flowgate Usage	participation x TI dispatch
Flowgate Support	negative of TI flowgate usage, for flowgate support generators
Flowgate Capacity	aggregate of TI flowgate usages
Flowgate Price	average of DI flowgate prices[14]
Availability	average of the DI offered availabilities or UIGFs
TNSP Support	discussed in next section below

13 Note that this processing calculates MW rather than MWh values, so 30-minute settlement amounts need to be divided by two: Pay$ = price ($/MWh) x volume (MW) / 2.

14 Flowgate price is taken to be zero in any DIs where the corresponding transmission constraint is not applied in NEMDE.

Averages will be simple, unweighted, arithmetic means. Each average is on a TI basis, meaning that it is the average of six DI-based values. This averaging is consistent with the approach currently taken in settlements: eg half-hourly spot prices are based on the average of dispatch prices.

The major complexity in converting from DI to TI is that transmission constraints may be applied in NEMDE – or be binding in NEMDE – only for a subset of DIs in a TI. This issue is addressed by using the calculated LHS of the constraint (ie aggregate usage) as a proxy for the *RHS* (flowgate capacity).

A flowgate is considered to be congested in a TI if the corresponding transmission constraint is binding in a single DI. In other DIs within the same TI, the constraint may not actually be binding. In fact, it may not even be applied in NEMDE.[15] In these situations:

- *aggregate usage* < *flowgate capacity* in DIs where a constraint is not binding; and
- possibly, *aggregate usage* > *flowgate capacity* in DIs where a constraint is not applied in NEMDE.

Given these possibilities, a constraint could be binding in a TI, but have aggregate usage either higher or lower than the *true* flowgate capacity.

This is not a problem for settlement balancing: since flowgate capacity is *defined* to be equal to aggregate flowgate usage for congested flowgates, settlement must balance. It may affect generators somewhat, since their entitlements will be affected. However, it is accepted in the current NEM design that there will be some basis risks arising from the 5-30 discrepancy. It is not expected that the issue above will significantly exacerbate those risks.

On the other hand, this issue *could* materially affect TNSPs, since it might cause FAS breaches to be flagged (and possibly penalised) even when there was no actual breach: or vice versa. Therefore, a different measure of flowgate capacity will be used in that context, as discussed in the section 12.4.4 (FAS Monitoring Approach), below.

12.4.3 Alternative Access Settlement Approach

The settlement approach described in the previous section can be summarised as: calculate 30-minute average dispatch outcomes and then apply the access settlement algebra to these 30-minute outcomes to calculate a 30-minute settlement amount.

An alternative - essentially reversed - approach would be: first to calculate DI settlement amounts based on the unprocessed DI-based information from dispatch; then to aggregate the DI settlement amounts across each TI to calculate 30-minute settlement amounts.

15 Ie where transmission conditions change midway through a TI, causing new constraints to be introduced into NEMDE and other constraints to be removed from NEMDE.

The former approach is analogous to existing settlement, where amounts are based on the product of 30-minute average dispatched output and 30-minute average spot price. Aligning the access settlement approach and existing settlement approach ensures that access settlement acts as the best hedge against dispatch uncertainty. Recall equation 2.8:

$$\text{Pay\$} = \text{RRP} \times \text{G} + (\text{RRP} - \text{LMP}) \times (\text{A-G}) \qquad (2.8)$$

The second term (from access settlement) acts as a hedge against the dispatch-related uncertainty of the first term (from existing settlement). The hedge arises because both terms contain the same RRP term. However, if access settlement were to apply the alternative approach described in this section, the second RRP term would, in effect, be somewhat different to the first RRP term, creating additional basis risk[16] for generators.

12.4.4 FAS Monitoring Approach

FAS monitoring requires flowgate capacity to be compared to target capacity. The calculation of target capacity has previously been discussed in section 5.2.4 (FAS is flowgate specific). This section considers the calculation of flowgate capacity. As discussed above, the measure of flowgate capacity used in access settlement would not be appropriate for use in FAS monitoring.

The materiality of a FAS breach is proportionate to the flowgate price. Therefore, DIs in which the flowgate price is high should contribute most to the estimation of flowgate capacity for the TI. For this reason, for the purposes of FAS monitoring, the TI flowgate capacity is defined as the FGP-weighted-average of the DI-based FGXs:

$$\text{FGX}_{\text{TI}} = \textstyle\sum_i (\text{FGP}_i \times \text{FGX}_i)/\textstyle\sum_i \text{FGP}_i$$

where:

FGP_i is the flowgate price in a dispatch interval i

FGX_i is the flowgate capacity in a dispatch interval i, calculated from the constraint LHS as above

FGX_i is the measure of flowgate capacity in a TI, used in FAS monitoring

The summation is over all DIs in the TI

Since a flowgate that is congested in the TI must be congested in at least one DI, the sum of the FGPs must be greater than zero and so the weighted-average is well-defined. The weighting ensures that DIs in which the flowgate is either not binding or not applied in NEMDE are ignored in the TI measure, since FGP_i equals zero in these DIs. Thus the issues discussed in the previous section do not arise.

16 There is already basis risk introduced by the 5-30 minute averaging process in existing settlement, but this access settlement approach would further exacerbate it.

12.5 Transmission Losses

12.5.1 Overview

Transmission losses have been ignored in the description of the OFA model and, in particular, in access settlements. That is because the OFA model deals only with congestion and the impact that it has on access. Nevertheless, it is worthwhile considering the impact that losses have under the OFA model and ensuring that they do not affect the principles and objectives underlying the OFA model.

12.5.2 Marginal Loss Factors

In the current NEM design, intra-regional transmission losses are represented through static marginal loss factors or *MLFs*, which are defined for each node in a region. The nodal MLF represents the additional generation that must be dispatched by NEMDE to supply an extra 1MW of demand at the node:

Incremental Dispatched Generation in NEMDE = Incremental Local D × MLF

The incremental generation is the amount extra dispatched by NEMDE. However, MLFs similarly apply to generation at a node:

Aggregate Dispatched Generation in NEMDE = $\sum_i G_i \times MLF_i$

Where:

G_i is the amount of generation dispatched at node i

MLF_i is the MLF at node i

For example, if the aggregate generation target in NEMDE was 1,000MW this could be met by 1,000MW of generation dispatched at the RRN (which, by definition, has an MLF of one) or by 990MW of generation at a node with an MLF of 1.01 (990MW x 1.01 = 1,000MW).

Because 990MW of generation at the latter node is worth the same as 1000MW of generation at the RRN, offer prices are adjusted accordingly: if both generators in the example offered at the same price, NEMDE would dispatch the generator at the local node since a lower quantity is required. In general, the cost assumed by NEMDE for dispatching a generator is:

Assumed Cost = Offer Price/MLF

NEMDE then dispatches generators in ascending order of *assumed cost*, not offer price.

Finally, to reflect their value in the market, generator payments are adjusted by MLF

Pay$ = RRP × Dispatch Output × MLF

Thus, if the RRP is \$20, then the generator at the RRN would be paid:

$$Pay\$ = \$20 \times 1{,}000MW \times 1 = \$20{,}000$$

A generator dispatched at the local node to meet the same demand would be paid:

$$Pay\$ = \$20 \times 990MW \times 1.01 = \$20.2 \times 990MW = \$20{,}000 \qquad (12.14)$$

The payments are the same, since each generator makes the same contribution to meeting demand.

MLFs can be greater or less than one; output, offer price and RRP may be scaled up or scaled down accordingly.

12.5.3 Local Prices in the Absence of Congestion

It is seen from equation (12.14) that the value to dispatch of a generator at a local node is RRP x MLF. Thus, in the absence of congestion, the *local price* is RRP x MLF, not RRP.

If a generator's offer price is C, the cost assumed by NEMDE is C/MLF and so it will be dispatched if:

$$C/\ MLF < RRP$$

Or, equivalently, if

$$C < RRP \times MLF$$

Thus, in the absence of congestion, the local price RRP ×MLF is a *clearing price*: a generator will be dispatched if its offer price is below this and not dispatched if its offer price is higher.

As noted above, a generator is paid the local price RRP ×MLP for dispatched output.

12.5.4 Local Prices during Congestion

In section 12.2.8 (Relationship between Local and Flowgate Prices) an expression was derived for the local price in terms of RRP and flowgate prices:

$$P_i = RRP - \textstyle\sum_k \alpha_{ik} \times FGP_k \qquad (12.15)$$

The derivation was based on the fact that if a generator participating in congested flowgates increased its output there would need to be some redispatch to bring flowgate usage back to flowgate capacity. Thus, the second term on the RHS of equation (12.15) reflects the additional costs associated with congestion.

Now that the local price in the absence of congestion has been more accurately defined to reflect losses, this congestion cost should be subtracted from that revised price formulation, ie:

$$P_i = RRP \times MLF - \sum_k \alpha_{ik} \times FGP_k \qquad (12.16)$$

In section 12.2.10 (Generator Access Settlement) it was demonstrated that, in the OFA model, a generator with zero access is paid its local price, as expressed in equation (12.15). In that case, losses were ignored and the RRP term on the RHS of equation (12.15) is the assumed payment from existing NEM settlements. If losses are included, then payment under existing settlement is RRP x MLF and so the aggregate payment to a generator without access is the RHS of equation (12.16).

Furthermore, because NEMDE takes account of both losses *and* congestion, as previously described, the local price defined by equation (12.16) is a clearing price: NEMDE will dispatch a generator if its offer price is below that local price.

Therefore, in summary, when losses are considered:

- the expression for local price is given by equation (12.16), whether or not there is congestion;
- the local price is a clearing price: a generator is dispatched by NEMDE if its offer price is below the local price;
- a generator with zero entitlements is paid the local price; a generator with some entitlements is paid higher than the local price; and
- thus, the no regrets principle applies: any generator that is dispatched will be paid at least its offer price.

The access settlement process correctly applies the principles of the OFA model in the presence of intra-regional losses.

12.6 Firm Access Allocation and Auctions

12.6.1 FAS Compliant Access Allocations

Access allocation refers to a process for allocating firm access to generators such that, taking account of *aggregate* access, the TNSP is FAS compliant. There are three access allocation processes in the OFA model:

- transitional access scaling (section 10.2.2.2 Stage 2: Access Scaling);
- transitional access auction (section 10.2.2.4 Stage 4: Access Auction); and
- short-term access auction (section 7.2.4 (Short-term Firm Access Issuance)).

In section 5.2.3 (FAS Implications for Flowgate Capacity), the FAS requirement was interpreted as being equivalent to requiring that the following load flow is feasible:

- all firm generators are dispatched at their agreed access levels;

- all local demand is supplied, based on expected demand levels for the particular study; and
- residual demand (ie total firm generation minus total local demand) is supplied at the RRN.

Thus, access allocation is analogous to a dispatch process in which:

- access allocated to each generator is analogous to its dispatched output;
- the FAS restriction is analogous to the flowgate constraints placed on dispatch;
- dispatch offers are set in the context of the objectives of the particular allocation process.

It is worth noting that, whereas real-time dispatch only includes current transmission constraints, an access allocation process must include *all* constraints that might possibly arise during NOC over the relevant access period.

12.6.2 Transitional Access Scaling

In the dispatch analogy introduced above, TA issuance is analogous to generation dispatch and so the TA allocated to an individual generator is analogous to its dispatch level. What is missing from the analogy are dispatch *offer prices*, which are needed to determine which generator gets dispatch priority. Let the analogy to these under transitional access scaling be referred to as *TA offer prices*. The lower the TA offer price for a generator, the more TA the generator will be issued. Of course, the generator cannot set its own offer price, as each generator would then bid -$1,000 to maximise its TA.

That scenario is similar to the disorderly bidding seen in the current NEM design and its implications for TA issuance are worth exploring further. Under disorderly bidding, as in all situations where offer prices are equal, generators with the same participation factor in a binding constraint will be scaled back *pro rata*: each is dispatched in the same proportion to its availability. However, when generators have differing participation factors, other things being equal, generators with *lower* participation factors are dispatched first, even if the difference in participation is minimal: eg if there were two generators bidding at -$1,000 having participation factors of 0.30 and 0.31, the first generator would be *fully* dispatched before the second generator had any dispatch. In the context of TA issuance, such an outcome might appear *efficient* (in that it maximises the aggregate amount of TA issued), but that efficiency is predicated on the accuracy of the constraint formulation and, specifically, the participation factors. If the participation factors were, instead, 0.31 and 0.30, the efficient outcome would be reversed.

To avoid such sensitivity of outcomes, it would be possible to *taper* the TA offer prices so that minor differences in participation factor only caused minor differences in TA issuance. An example of a tapered offer price: 10 TA offer bands are used, each applying to 10 per cent of a generator's registered capacity; offer pricing in each band is the same for each generator: eg $10, $20,…,$100.

With a tapered offer, the impact of a minimal difference in participation factor would only be noticeable in one offer band: in the example above, the generator with 0.30 participation might be dispatched in 8 bands, say (and so receive TA for 80 per cent of capacity), whilst the generator with 0.31 participation might then only be dispatched in 7 bands (corresponding to 70 per cent of capacity). So, lower participation will lead to higher TA issuance, but in a proportionate rather than abrupt way.[17]

Using tapered TA offer prices would preserve the broad efficiency of the TA allocation but make it more *robust*: ie less sensitive to inaccuracies or uncertainties in constraint formulation. The appropriate trade-off between efficiency and robustness could be achieved by varying the gradient of the pricing taper: eg using $10, $11, $12… instead of $10, $20, $30….

It is proposed that interconnectors would be given lower priority in the TA allocation process. This would imply setting TA offer prices for interconnectors much higher than those for generators,[18] so that even with a low participation, an interconnector would not be allocated TA in preference to a generator.

12.6.3 Access Auctions

An auction creates a *change* in access allocation from pre-auction to post-auction levels. It is assumed that the pre-auction allocation will be FAS compliant. The auction process must ensure that the post-auction allocation is also FAS compliant: ie that the analogous dispatch is *feasible*.

So long as pre-auction and post-auction FAS compliance is assured, it is not necessary to place additional constraints on the auction: for example, that total purchases match total sales at a node or across a zone. It is not even necessary to ensure that the auction clears *financially* – ie that there is no auction settlement deficit. For reasons explained below, an auction surplus is assured, so long as the same set of flowgate constraints is placed on the post-auction allocation as was placed on the pre-auction allocation and that both allocations comply with these constraints.

In an auction, participants submit bids and offers, which are price-quantity pairs representing maximum quantities they wish to buy or sell, and maximum or minimum prices at which they wish to buy or sell, respectively. For an access auction, in the dispatch analogy, these are converted into equivalent *dispatch offers*, using the process described below. An economic dispatch is then calculated, using these dispatch offers together with the necessary flowgate constraints. To ensure that a maximal amount of

17 Obviously, the impact could be made even smoother with, say, 100 offer bands.

18 This is where the analogy with dispatch breaks down. Interconnectors do not have explicit offer prices in dispatch. However, a high interconnector offer price could be modelled in dispatch by assuming that the local RRP was very high ($10,000 say) and assuming that the remote RRP was

access is allocated, a large notional demand is included at the RRN,[19] together with a large notional RRN generator bidding at a high price that sets RRP.

The value of agreed access is based on the difference between RRP and the local price: the lower the local price, the higher value. To reflect this, the amount paid by a generator at the auction is:

Auction Payment$ = TA purchased at node × (RRP – P)

Auction Payment$ = Access Purchased at node × (RRP – P)

Where the regional and local prices, together with *dispatched output*, are calculated in the *auction dispatch*.

Similarly, if the generator sells into the auction, the amount received by the generator would be:

Auction receipt$ = Access sold at node × (RRP – P)

The RRP level is arbitrary – it is the *difference* that is important – so can be set at an arbitrary level: $100, say. In dispatch, a generator is dispatched if its offer price is below the local price. In the analogy, a generator that is dispatched in the auction is allocated post-auction access. So, how should a generator's dispatch offer be set?[20]

Consider first of all a generator who has 100MW of access pre-auction and offers to sell all of it if the auction clearing price (RRP minus local price) is higher than $40. That is equivalent to wishing *not* to be dispatched if the local price is lower than $60 (since $100-$60 = $40). Therefore, the dispatch offer should be 100MW at $60, leading to the following possible outcomes:

- if the local price is <$60 (and so the clearing price > $40) the generator is not dispatched, has 0MW after the auction, meaning it has sold 100MW; and
- if the local price is >$60 (meaning clearing price<$40) the generator is dispatched to 100MW, meaning it has kept its initial holding.

Similarly, if a generator with no access pre-auction wishes to buy 50MW if the clearing price is lower than $25, its dispatch offer should be 50MW at $75. The possible outcomes are:

- if the local price is <$75 (and so the clearing price > $25) the generator is not dispatched, has 0MW after the auction, meaning it has bought nothing; and

---

only slightly lower ($9,000, say) so that local generation would always be dispatched in preference to the importing interconnector.

19 Note that there is no need to include local demand, because the effect of this is already captured in the flowgate constraints.

20 In practice, each generator would submit conventional bids or offers and these would be converted into TA offer prices in the auction process.

- if the local price is >$75 (meaning clearing price<$25) the generator is dispatched to 50MW, meaning it has purchased 50MW at the auction.

More generally, a generator making multiple bids or offers at a node can be represented in dispatch by a generator with multiple offer price bands; or, alternatively, as multiple generators.

12.6.4 Auction Settlement Surplus

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.

Since both pre-auction and post-auction holdings must be FAS compliant, Q and R both represent dispatch outcomes which must be feasible on the transmission network. Thus, they must both obey each relevant flowgate constraint, k, and so:

$$\sum_i \alpha_{ik} \times Q_i \leq FGX_k$$

$$\sum_i \alpha_{ik} \times R_i \leq FGX_R$$

Now, the payment by each generator into the auction is the quantity purchased multiplied by the clearing price:

$$Pay\$_i = (R_i - Q_i) \times (RRP - P_i) \qquad (12.17)$$

Recalling the formula for local price:

$$P_i = RRP - \sum_k \alpha_{ik} \times FGP_k \qquad (12.18)$$

Equation (12.17) becomes:

$$Pay\$_i = (R_i - Q_i) \times \sum_k \alpha_{ik} \times FGP_k \qquad (12.19)$$

where the FGPs are the flowgate prices determined in the *R* dispatch.

Aggregate payments into the auction are determined by summing equation (12.19) across all generators:

$$\text{Total } Pay\$ = \sum_i Pay\$_i$$

$$=\sum_i \sum_k \{(R_i \times \alpha_{ik} \times FGP_k) - \sum_i \sum_k (Q_i \times \alpha_{ik} \times FGP_k)\}$$

$$=\sum_k FGP_k \times \{\sum_i \{(R_i \times \alpha_{ik}) - \sum_i (Q_i \times \alpha_{ik})\}$$

Now, flowgate prices are only greater than zero if the relevant constraint is binding so:

$$FGP_k > 0 \text{ implies } \sum_k (Ri \text{ x } \alpha_{ik}) = FGX_k$$

Therefore:

$$\text{Total Pay\$} = \sum_k \{FGP_k \times [FGX_k - \sum_i (Q_i \times \alpha_{ik})]\} \qquad (12.20)$$

Now, because the Q dispatch is feasible:

$$\sum_i (Qi \times \alpha_{ik}) \leq FGX_k$$

Therefore, the difference term in equation (12.20) is never negative and, since flowgate prices are also never negative:

$$\text{Total Pay\$} \geq 0$$

That is to say, total payments made by purchasers in the auction will always equal or exceed total payments made to sellers. There will be a surplus only if one or more flowgates that are binding in the R dispatch are not binding in the Q dispatch. If the same set of constraints is binding in both dispatches, then the surplus will be zero. If there is a surplus in the TA auction, it would be passed to demand-side users via a TNSP: similar to the SRA arrangements. Any settlement surplus in the ST auction is paid to the TNSP, as discussed below.

One way of understanding the algebra above is to think of nodal access as a bundle of *flowgate entitlements*. An access holding, A_i, at node *i*, provides a generator with an entitlement amount, $A_i \times a_{ik}$, 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 for each entitlement, implying a total payment:

$$\text{Pay\$}_i = \sum k\ [(R_i - Q_i) \times \alpha_{ik}] \times FGP_k$$

$$= (R_i - Q_i) \times \sum_k \alpha_{ik} \times FGP_k$$

Which is the same amount that it pays based on the nodal clearing prices, as derived in equation 12.19, above.

Since all flowgate entitlements 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.

Because flowgate prices are always positive, an auction deficit can only arise on a flowgate if the total purchases on that flowgate are *negative*: ie if there are net *sales* in aggregate. But in this case, the total post-auction holdings must be less than total pre-auction holdings which, in turn, must be less than or equal to flowgate capacity. Since post-auction holdings are less than flowgate capacity, some of the flowgate capacity is effectively *passed-in*. Since there are no reserve prices in the auction, this must mean that the flowgate price is zero: it is analogous to the flowgate being uncongested in the post-auction dispatch, *R*. This means that, even where there are net sales of entitlements on a flowgate, there is still no auction deficit.

On the other hand, if there are net *purchases* on a flowgate, it is possible that the available flowgate capacity is fully auctioned and so the flowgate price will be positive and there will be an auction surplus.

12.6.5 TNSP Auction Sales

Under the ST access issuance process (section 7.2.4 (Short-term Firm Access Issuance)) a TNSP uses an auction to sell additional, ST access. The amount of ST access it is able to sell depends upon the difference, on each relevant flowgate, between the *expected* flowgate capacity over the ST access term, and the *target* flowgate capacity under the current (pre-auction) access allocation.

The target flowgate capacity is set by the formula:

$$FGT_k = \sum_i Q_i \times \alpha_{ik}$$

Where:

FGT_k is the target flowgate capacity on flowgate k and

Q_i is the pre-auction access holding of generator i

Equation 12.20 above now becomes:

$$\text{Total Pay\$} = \sum_k FGP_k \times (FGX_k - FGT_k) \quad (12.21)$$

That is to say, payment to the TNSP equals the product of flowgate price and flowgate capacity "released". The release might be because some flowgate capacity is currently underutilised, because the TNSP finds additional flowgate capacity that can be made available for the short-term period or because the TNSP is able to undertake some expenditure to increase flowgate capacity.

The auction formulation described above means that all of the available flowgate capacity will be released no matter how low the flowgate price. A TNSP might instead wish to place a reserve price on the flowgate capacity, so that the additional capacity is only released if the flowgate price exceeds that reserve price.[21] This can be achieved by adding additional constraints in the auction dispatch taking the form:

$$\sum_i (R_i \times \alpha_{ik}) \leq FGT_k$$

21 This might be desirable, for example, where a TNSP needs to ensure that the cost of any capacity expansion will be covered by sales revenue.

These constraints would be permitted to be violated, with the constraint violation penalty set based on the flowgate reserve price. The constraint violation penalty is therefore defined as:

$$CVP_k = RP_k \times (\sum_i R_i \times \alpha_{ik} - FGT_k)$$

Where:

CVP_k is the constraint violation penalty: zero if the constraint is not violated

RP_k is the reserve price set for flowgate k

In dispatch, the constraint violation penalty is added to the objective function: ie the cost that the dispatch calculation is seeking to minimise. Thus, the constraint is only violated if the value of doing so exceeds the violation penalty: ie the flowgate price exceeds the reserve price.[22]

12.6.6 TNSP Auction Buybacks

In principle, a TNSP could use the ST auction to buy back some firm access in the situation where the existing access allocation breaches FAS: ie

$$FGX_k < FGT_k$$

For one or more flowgates k.

In this case, equation 12.21 shows that net auction payments might not be positive: since the component relating to the problematic flowgate will be negative. Thus, the TNSP may have to pay money into the auction for it to financially clear. This is analogous to the TNSP buying back firm access: although, being a TNSP, it does not actually *hold* that bought-back access in any sense; rather, it just has simply reduced the overall amount of access on issue. The TNSP incentive scheme may mean that the TNSP considers itself better off paying money to buy back access than being exposed to financial penalties for breaching FAS.

Similarly to setting a reserve price on auction sales, the TNSP could set a limit on the price that it would be prepared to pay to buyback access. This would be done analogously to the reserve price setting described above: by creating additional constraints in the auction dispatch that could be violated subject to a constraint violation penalty. In this case, permitting constraint violation *would* be necessary to ensure a feasible dispatch outcome.

In general, it is quite possible that, in the ST auction, a TNSP could be releasing capacity on some flowgates and, at the same time, buying back capacity on other flowgates.

---

22 Constraint violation penalties are normally included in dispatch to ensure that a feasible dispatch can be found. However, in the auction context, the pre-auction allocation is always feasible, so feasibility concerns do not arise.

12.7 Entitlement Scaling

12.7.1 Overview

Section 4.2 (Design Blueprint) qualitatively describes the process for setting target and actual entitlements on flowgates. The sections below provide a mathematical description of this process, together with numerical examples.

12.7.2 Target Entitlements

Entitlement targets are based on the decomposition of agreed access into three components:

$A_F = \min(AA,RC)$

$A_{NF}=\max(Avail-AA,0)$

$A_{SF} = \max(AA-RC,0)$

Where the subscript:

F refers to the firm access component

NF refers to the non-firm access component

SF refers to the super-firm access component

and:

A = access component

AA = agreed access amount

Avail = generator availability

RC = generator registered capacity

Note that:

- access components are non-negative;
- the sum of the firm and non-firm components must equal or exceed availability;
- A_{NF} and A_{SF} cannot both be non-zero.

The *entitlement target* for each component is calculated by multiplying the access component by the generator's participation factor for the relevant flowgate:

$ET_F = \alpha \times A_F$

$ET_{NF} = \alpha \times A_{NF}$

$ET_{SF} = \alpha \times A_{SF}$

Where:

α = flowgate participation factor

ET is the entitlement target

A numerical example is provided in Table 12.2 below to illustrate the target setting process. Note that each generator belongs to a different access category.

Table 12.2 Calculation of Target Entitlements

Generator	Nodal Values					α_i	Flowgate Values		
	AA	RC	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.

12.7.3 Actual Entitlements

For a flowgate to be congested, there must be the *potential* for total flowgate usage to be greater than flowgate capacity. Recall that flowgate usage is:

$$U_i = \alpha_i \times G_i \qquad (12.22)$$

Where:

U_i = flowgate usage of generator i

G_i = dispatch output of generator i

For congestion, there must be some possible set of generator outputs, G_i, such that:

$$\sum_i (\alpha_i \times G_i) = \sum_i U_i > FGX \qquad (12.23)$$

Now since dispatched output can be no higher than availability:

$\sum_i ET_{Fi} + \sum_i ET_{NFi} \geq \sum (\alpha_i \text{ x } Avail_i)$ (since $A_F + A_{NF} \geq Avail$ as noted above)

$\geq \sum (\alpha_i \text{ x } G_i)$ (since $Avail_i \geq G_i$)

$> FGX$ (from equation (12.23))

Therefore, if a 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.

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 entitlements, and a single *non-firm scaling factor* is applied to all non-firm entitlements;
- firm entitlements are only scaled back when non-firm actual entitlements have been scaled back to zero; and
- 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 no higher than the firm target entitlement.

The formulae for determining actual entitlements, based on target entitlements, are presented in Table 12.3, below.

Table 12.3 Formulae for actual entitlements

Symbol	Meaning	Calculation
k_F	firm scaling factor	using a goal seek algorithm
k_{NF}	non-firm scaling factor	$k_{NF} = (FGX-\sum EA_F)/ \sum ET_{NF}$
EA_F	actual firm entitlement	k_F x ET_F
EA_{NF}	actual non-firm entitlement	k_{NF} x ET_{NF}
EA_{SF}	actual super-firm entitlement	$\min\{ET_F-EA_F, k_F \text{ x } ET_{SF}\}$
EA	actual (total) entitlement	$EA_F + EA_{PF} + EA_{SF}$

To illustrate these formulae numerically, actual entitlements are calculated, from the targets presented in Table 12.2, 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 12.4, below.

Table 12.4 Actual entitlements under two capacity scenarios

	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	108	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 scenario 2, no scaling back of firm entitlements is necessary and so some non-firm entitlements are provided.

Note that total firm and non-firm target entitlements equal 1250MW, meaning that if flowgate capacity exceeded 1250MW there would be no congestion.

12.8 TNSP Planning and Operations under the Firm Access Standard

12.8.1 NOC1 FAS Obligation

As described in section 5.2.3 (FAS Implications for Flowgate Capacity), a TNSP is required to ensure that, at each NOC1-tagged congested flowgate:

Effective Flowgate Capacity ≥ Target Flowgate Capacity ≡ $\sum_{\alpha i} A_i$

Where:

α_i = participation factor of generator i

A_i= agreed access of generator i

The summation is over all *access* generators: ie *support* generators are excluded

Effective flowgate capacity is the RHS of a flowgate constraint, adjusted to include flowgate support. It is therefore determined by three factors:

- transmission capacity;
- local demand;[23] and
- the dispatch of flowgate support generation.

Consider a dispatch study set up as shown in Table 12.5 below.[24]

Table 12.5 Dispatch Study for analysing NOC1 Obligation

Generation Side	Demand Side
Dispatch of firm access generation at agreed access level	Supply forecast nodal demands
Dispatch of an assumed level of flowgate support generation	Balance of demand located at RRN

The balance of demand is included at the RRN to ensure that total generation equals total demand. That balance could be positive or negative.[25]

If that dispatch is feasible, that implies that no transmission constraints that would be placed on the dispatch are violated: since otherwise a change to the dispatch would be required. The form of those constraints would be:

$$\sum \alpha_i G_i \leq \text{flowgate capacity}$$

The LHS consists of access generators, dispatched at their agreed access level and support generators dispatched at an assumed support level. Thus:

$$\sum \alpha_i G_i = \text{Target flowgate capacity} - \text{flowgate support}$$

The constraint equation can therefore be rewritten as:

$$\text{target flow capacity} - \text{flowgate support} \leq \text{flowgate capacity}$$

And so:

$$\text{target flowgate capacity} \leq \text{flowgate capacity} + \text{flowgate support}$$

The RHS of this equation is the *effective flowgate capacity*, defined above. Therefore, the flowgate constraints in the dispatch study ensure that:

$$\text{effective flowgate capacity} \geq \text{target flowgate capacity}$$

23 As discussed in section 12.2.3 (Transmission Capacity versus Flowgate Capacity).

24 There is a potential difficulty with how this study is defined in that generators that have positive participation in one binding constraint may have negative participation in another constraint and so need to be dispatched at two different levels. Thus, different dispatch studies may be required to consider these constraints separately.

25 Adding a negative demand is equivalent to adding generation at the node.

Therefore, a *sufficient* condition for a TNSP meeting its FAS obligation under NOC is that the dispatch of generation described in Table 12.5 is feasible and secure on its transmission network.

This load flow condition is *sufficient* but not *necessary*. The FAS only requires that sufficient flowgate capacity is provided at *congested* flowgates, whereas the load flow condition ensures sufficient capacity at all flowgates. If a TNSP could be confident that a particular flowgate would never be congested, it would not explicitly need to maintain capacity on that flowgate.

In particular, the adding of balancing demand at the RRN is liable to cause congestion – in the load flow study - *locally* to the RRN which could never arise in practice. Such congestion can safely be ignored by the TNSP.

12.8.2 Scenario Analysis

The load flow described above depends upon assumptions on:

- nodal demands;
- dispatch of flowgate support generation.

Therefore, this load flow condition would need to be checked under a range of conditions, such as:

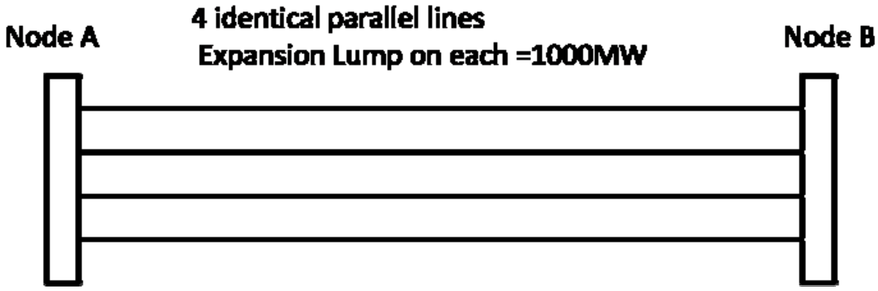
- high and low demand levels;
- low level of flowgate support.

Where low levels of flowgate support created problems in meeting the load flow condition, a TNSP may have to enter into network support agreements with the relevant generators.

12.9 Meshedness in Access Pricing

As discussed in section 6.2 (Design Blueprint) a meshedness factor is applied to transmission lumpiness in the access pricing model. This is to reflect the fact that, in a heavily meshed part of the network, lumpiness of transmission expansion is less relevant and so effective lumpiness is lower. This effect is illustrated in Figure 12.8 below.

Figure 12.8 Lumpiness and meshedness



Because the four lines are identical and the access pricing model considers each element separately, the model will schedule expansion of all four lines at the same time, meaning that the total transmission capacity between the two nodes is increased in lumps of 4,000MW. Clearly, it is more realistic to model expansion of this capacity at 1,000MW at a time, which is achieved by setting the effective lumpiness of each individual line at 250MW. In general, the pricing model defines lumpiness as:

lumpiness = expansion lump / meshedness

In the figure, the expansion lump is 1,000MW and the meshedness is 4, meaning that each line is assumed to have lumpiness of 250MW.

That *meshedness=4* in the example is apparent, but how should meshedness be determined in a general situation? It will be seen, for the simple example network, that if 100MW is injected in at node A and withdrawn at node B, then a 25MW flow is generated on each line. The ratio of the injection to the line flow gives the line's meshedness (100/25 = 4). This formulation can be applied to meshedness for all network topologies, ie:

meshedness = (MW of injection and withdrawal at each end of the line)/(lineflow)

The higher the meshedness, the more alternative paths there are, or the higher the admittance of these alternative paths relative to the line admittance.[26] It is straightforward to calculate meshedness on general network topologies using this definition.

12.10 Annual Payment Profiling

12.10.1 Overview

As discussed in section 6.2.5 (Payment Profiling Algorithm) there will be a need to specify an algorithm for converting the lump-sum access charge calculated by the

26 Admittance is a measure of how easily power flows through a line.

access pricing methodology into a stream of annual access payments which, at a specified discount rate, have an NPV equal to the calculated lump sum.[27]

The algorithm would specify the *default* payment profile; customisation of this payment would be permitted in accordance with the TNSP's customisation policy.

Regulatory issues associated with the payment profiling are discussed in section 12.2 (Flowgate Pricing and Local Pricing). In that section it is noted that risks arise from differences between the payment profile and the profile of carrying costs of regulated assets that is applied under revenue regulation. The risk is borne by the TNSP in the first regulatory period of the access agreement and borne by demand-side users in subsequent regulatory periods. Risks can be mitigated by minimising these differences. This objective is considered and applied in this section.

12.10.2 Revenue Regulation and Carrying Costs

Under the building block approach to determining regulated revenue, the allowed annual revenue associated with an asset in the Regulatory Asset Base (RAB) is:

$$\text{Real Revenue} = \text{WACC} \times \text{DRC} + \text{D} \qquad (12.24)$$

Where:

WACC is the *real* regulated rate of return for the TNSP

DRC is the depreciated replacement cost for the asset

D is the annual depreciation

The depreciation schedule is typically *straight-line* over the asset life and the DRC is generally approximated by the historical asset cost (in real terms). These assumptions are used in the discussion below, although it would be possible to adopt different assumptions.

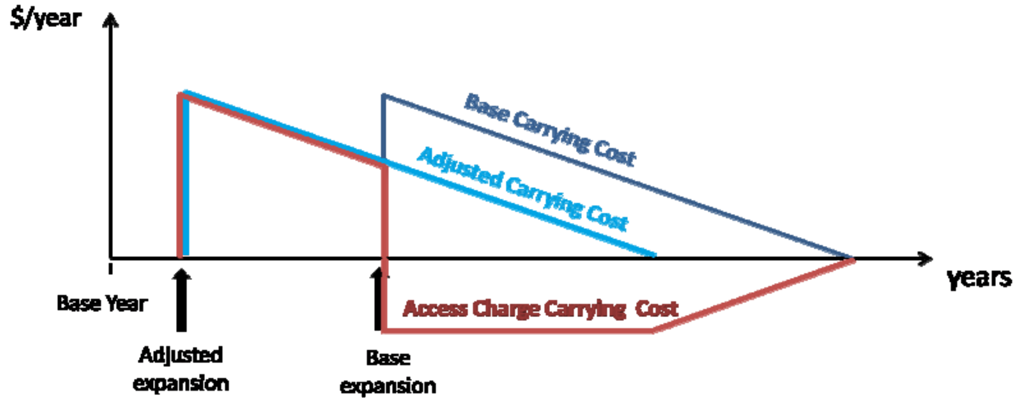
An expansion plan in the access pricing methodology is a set of expansions on the shared network, with each expansion taking place at a defined time and for a defined capital cost. The annual carrying cost of each expanded asset can then be determined by applying equation (12.24) and these costs can be summed across all assets in the expansion plan to give an annual carrying cost for the plan.

The LRIC methodology calculates the access charge as the difference in NPV between two expansion plans: the adjusted expansion plan and the base expansion plan. Thus, an annual carrying cost for the access charge can be determined by taking the difference between the annual carrying cost for the adjusted plan and that for the baseline plan.

27 The discount rate would probably be based on a forecast of regulated WACC.

An example of this is illustrated in Figure 12.9, below, in relation to a very simple access charge that is based on the advancement of a single planned expansion.[28]

Figure 12.9 Example of access charge carrying costs



Since the NPV of the carrying cost of each asset in each plan equals the NPV of the capital cost, by aggregation the NPV of the access charge carrying cost equals the access charge lump-sum. That equivalence makes the carrying cost a possible annual payment profile. However, it is clear from Figure 12.9 above that such a profile would be unacceptable. It would imply high positive charges in the early years followed by negative charges (ie rebates) in later years. Given the long asset life, the rebates could continue for years – decades in fact – after the end of the access agreement.[29]

Nevertheless, this analysis does illustrate how the costs for a TNSP in providing access might be substantially *front-ended*: ie occurring primarily at or near the start of the access agreement.[30] Ideally, the payment profile should endeavour to reflect this front-ending, whilst avoiding the problem of annual rebates being required after the end of the agreement.

Since it is discrepancies in the *first* regulatory period which are most problematic (since these affect the risk-averse TNSP) a possible approach would be to set annual payments equal, in NPV terms, to annual carrying costs in in the first regulatory period, and then base them on a conventional straight-line repayment plan for the remaining years of the agreement, in order to provide the correct NPV in aggregate. Such a profile is illustrated in Figure 12.10, below, for the simple access charge described in the previous figure: in the figure, the NPV associated with the payment profile during the first period is equal to the NPV associated with the carrying charge.

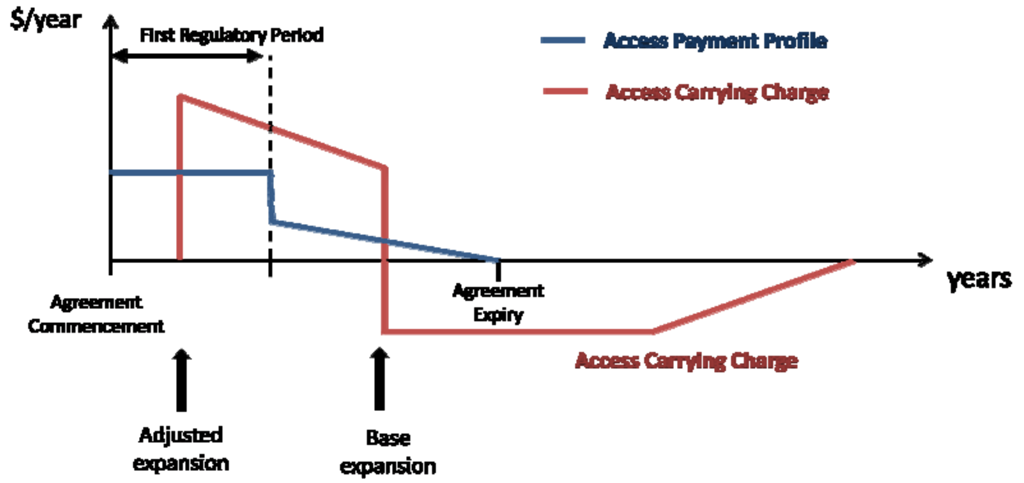
---

28 There could be many other expansions in the plans, but if the timing of expansions is the same in the two plans, their carrying costs will cancel out.

29 In practice the rebates would continue until the year in which the two plans converged following the expiry of the access agreement.

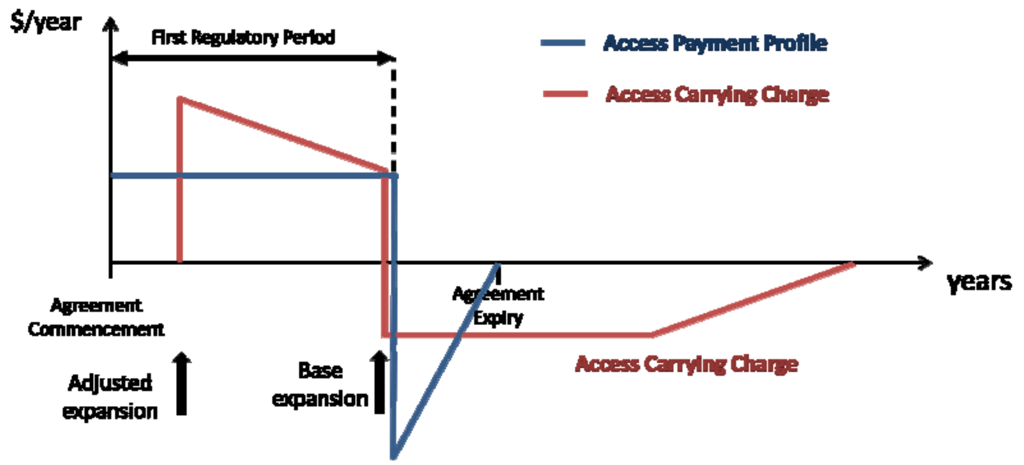
30 This will not always be the case: if there is substantial spare capacity, for example, any advanced expansion may not occur until many years after agreement commencement.

Figure 12.10 Possible access payment profile



However, this payment profile may not be reasonable if the NPV of carrying costs after the first regulatory period is negative, since this would mean *overcharging* for access in the first regulatory period and then *rebating* this overcharge subsequently. This situation would arise in the example shown in Figure 12.11, below, which is the same as the previous example, except that the first regulatory period is twice as long.

Figure 12.11 An impractical access payment profile



12.10.3 Indexing of Annual Payments

The revenue formula provided in equation (12.24) is in *real terms* and so the revenue allowance is indexed by CPI. Any annual payment profile based on carrying costs would similarly be defined in real terms and indexed by CPI.

In the revenue formula, WACC is adjusted on each regulatory reset. Correspondingly, any payment profile would be indexed by WACC. The payment profile would be defined by a repayment profile, R, from which the annual payment is determined using the formula:[31]

$$PAY_t = (AC - CR_t) \times WACC_t + R_t$$

Where:

AC is the lump sum access charge

WACC is the regulatory WACC applying in year t

R_t is the repayment amount in year t

CR_t is the cumulative repayment for year t

$CR_0 = 0$

$CR_t = CR_{t-1}+R_t$

$R_1+R_2+\ldots+R_T = AC$

Year T is the final year of the access agreement

The repayment profile, *R*, could be set in order to keep the TNSP whole in the first regulatory period, as discussed in the previous section. However, leaving that objective aside, any repayment schedule could potentially be applied, using the formula above.

12.11 Additional information on directed interconnector payments

Section 9.2.5 (Allocation of directed interconnector payments to market participants) discussed how FIRs provide rights to receive a portion of the DIP rather than the IRSR. The payment formula is designed to achieve two objectives:

- the FIR provides a firm inter-regional hedge for the holder; and
- so long as the total amount of issued FIRs does not exceed the agreed access level on the directed interconnector, the total FIR payments never exceed the DIP in a settlement period.

These two features are discussed in greater detail below.

12.11.1 Inter-regional hedging using FIRs

Recall that the payment to an FIR holder in respect of a flowgate is:

$$\text{FIR Payment\$} = FGP \times \alpha \times \text{FIR amount} \times k_F \qquad (12.25)$$

31 The formula may change slightly, depending upon whether payments are made at the start or end of each year.

Summing this over all congested flowgates and assuming, for simplicity, that k_F is the same on every flowgate:

$$\text{Total FIR payment} = \text{FIR amount} \times k_F \times (\textstyle\sum_j FGP_j \times \alpha_j)$$

Where:

FGP_j is the flowgate price on flowgate j

α_j is the participation of the interconnector in flowgate j

Assuming that all congested flowgates on the interconnector are binding in the direction of the directed interconnector being considered:[32]

$$\textstyle\sum_j FGP_j \times \alpha_j = RRP_M - RRP_X \qquad (12.26)$$

Where:

RRP_M is the RRP in the importing region

RRP_X is the RRP in the exporting region

The result in equation (12.26) is demonstrated in section 12.2.8 (Inter-regional Price Difference).

Therefore, the total FIR payment is:

$$\text{Total FIR payment} = \text{FIR amount} \times k_F \times (RRP_M - RRP_X)$$

This means that the FIR provides the holder with an inter-regional hedge (IRH) with volume

$$\text{Hedging Volume} = \text{FIR amount} \times k_F$$

In the more general case where the firm scaling factors are different on each flowgate

$$\text{Hedging Volume} = \text{FIR amount} \times \text{average}(k_F)$$

Where:

average(k_F) is a weighted-average of the k_Fs on each flowgate

The FAS requires that, during NOC, there is no scaling back: ie $k_F = 1$. Therefore, so long as the TNSPs are FAS-compliant, a FIR provides a perfect inter-regional hedge during normal operating conditions.

By comparison, the hedging volume provided by an SRA right in the current NEM design is proportional to the interconnector flow and goes to zero when the flow is

32 The situation of mixed constraints where this is not true is discussed in section 12.3.3 (Mixed Interconnector Constraints).

counterprice. Thus, the hedging firmness of an FIR will be far superior to the hedging firmness of an SRA right.

12.11.2 Backing of FIRs by the directed interconnector payment

The contribution to the DIP from congested hybrid flowgate in which a directed interconnector participates is given by the formula

$$\text{IRSR pay\$} = \text{FGP} \times \text{E} \qquad (12.27)$$

The entitlement is the sum of firm and non-firm entitlements[33] meaning that

$$\text{E} \geq \text{firm actual entitlement} = k_F \times \text{firm target entitlement} \qquad (12.28)$$

Where:

k_F is the firm scaling factor

Recall from section 4.2.3 (Target Entitlements) that

$$\text{firm target entitlement} = \alpha \times \text{AQ} \qquad (12.29)$$

Where:

α is the participation of the directed interconnector in the flowgate

AQ is the agreed access amount for the directed interconnector

From equations (12.27), (12.28) and (12.29)

$$\text{IRSR pay\$} \geq \text{FGP} \times k_F \times \alpha \times \text{AQ} \qquad (12.30)$$

Now, recall from section 9.2.4 (Firm interconnector rights) that the FIR payment on a flowgate is

$$\text{FIR Payment\$} = \text{FGP} \times \alpha \times \text{FIR amount} \times k_F \qquad (12.31)$$

Summing this payment across all issued FIRs

$$\text{Total FIR payment\$} = \sum(\text{FGP} \times \alpha \times \text{FIR amount} \times \text{kF})$$

$$= \text{FGP} \times k_F \times \alpha \times \sum \text{FIR amount} \qquad (12.32)$$

It will be seen by comparing equations (12.31) and (12.32) that

$$\sum \text{FIR amount} \leq \text{AQ}$$

is a sufficient condition for ensuring that

33 Because an interconnector has no availability, it can have no super-firm entitlement.

DIP > Total FIR payment$

for each directed interconnector on each congested flowgate and, therefore, in aggregate.

A OFA Model Glossary

Table A.1 OFA Model Glossary

Defined Term	Meaning
access	*network access*
access charge	a charge payable by a generator to its local TNSP in return for receiving firm access
access settlement	a new AEMO settlement process in the OFA model through which *access-long* generators receive payments and *access-short* generators make payments
access-long generator	(for a *generator*) being dispatched at a level below its *access* level, thus entitling it to payments from *access settlement*
access-short generator	(for a *generator*) being dispatched at a level above its *access* level, thus obliging it to payments into *access settlement*
agreed access (amount)	the nominal amount of access specified in an *firm access agreement*, which may vary between peak and off-peak periods
abnormal operating condition	a transmission operating condition specified as an abnormal operating condition in the firm access standard
availability	for a conventional generator, the offered availability; for an intermittent generator, the Unconstrained Intermittent Generation Forecast
capacity shortfall	the difference in a settlement period between *target flowgate capacity* and actual *flowgate capacity* when the latter is less than the former
congested flowgate	a *flowgate* whose capacity is fully utilised in dispatch and which is causing dispatch to be constrained
constrained off	(for a *generator*) dispatched below its *preferred output*; a firm, constrained-off generator will typically be access-long and so entitled to payment from access settlements
constrained on	(for a *generator*) dispatched above its *preferred output*
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
dispatch access	the right to be dispatched in NEM dispatch at a specified MW level in accordance with a dispatch offer and paid the local price on dispatched output
effective flowgate capacity	the amount of capacity that is allocated between *generators* in the entitlement scaling algorithm; equals the *flowgate capacity*

Defined Term	Meaning
	plus the *flowgate support* plus the *TNSP support*
embedded generator	a distribution-connected *generator*
exporting region	the region from which a *directed interconnector* withdraws power
firm access (service)	a transmission service provided to generators that have a *firm access agreement* with their local TNSP, up to the level of the *agreed access* amount
firm access agreement	an agreement between a TNSP and a generator which specifies the service parameters for the provision of *firm access* service
firm access level	(for a *generator*) the lower of the generator's *agreed access* and *capacity*
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 a *firm access agreement* and an *agreed access amount* equal to its capacity
firm interconnector	an *interconnector* for which AEMO holds some *agreed access* in trust
firm interconnector right	a right to receive a specified proportion of the IRSR proceeds of a *firm interconnector*
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 capacity	the maximum aggregate usage of a *flowgate* allowed in dispatch. The RHS of the corresponding NEMDE transmission constraint
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 LHS of the corresponding NEMDE transmission constraint
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
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
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	a power station, or the generating company responsible for the power station, depending upon the context

Defined Term	Meaning
generator node	the transmission node at which a *generator*, or the distribution network used by an *embedded generator*, connects to the shared transmission network
hybrid flowgate	a *flowgate* in which *generators* and *interconnectors* both participate
importing region	the region into which a *directed interconnector* injects power
interconnector	a notional entity that is dispatched by NEMDE to transfer power from one RRN to a neighbouring RRN across a regulated interconnector
inter-regional access	network access provided to a directed interconnector, from the RRN in the exporting region to the RRN in the importing region
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 fund, held in trust by AEMO, into which, or from which, settlement payments relating to *directed interconnectors* are paid
intra-regional access	*network access* provided to a *generator*, from its *generator node* to the *RRN* in its *local region*
local region	(of a generator or access agreement) the region in which the relevant generator node is located
local price	the marginal value that a generator at a node provides to economic dispatch; the locational marginal price
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
network access	the right to be paid in AEMO settlement the difference between RRP and LMP for a specified MW level
non-firm access	the access received by generators that do not have an *firm access agreement* and so do not receive a *firm access service*
non-firm access level	(for a *generator*) the difference between *availability* and *agreed access amount*, when the former takes a higher value
non-firm generator	a generator without a *firm access agreement*.
normal operating condition	a transmission operating condition specified as a normal operating condition in the firm access standard
part-firm generator	a generator with a *firm access agreement* and an *agreed access amount* less than its capacity

Defined Term	**Meaning**
preferred output	the quantity of a *generator's* availability that is offered at or below the RRP
regional price	the price paid to a dispatched generator in regional settlement; the Regional Reference Price
reliability access	the peak-time access provided to a *reliability generator*
reliability generator	a *non-firm generator* who nevertheless receives peak-time access as a result of a TNSP expanding transmission to meet a *demand-side reliability standard*
remote region	a region other than the *local region*
service parameters	values contained in an *access agreement* specifying *access amount*, term, location and so on.
settlement residue auction	the auction through which AEMO sells SRA rights
SRA right	the right to receive a specified proportion of the *inter-regional settlement residue* for a specified *directed interconnector*
super-firm access level	(for a *generator*) the difference between *agreed access amount* and *capacity*, when the former takes a higher value
super-firm generator	a generator with a *firm access agreement* and an *agreed access amount* greater than its capacity
target firm entitlement	(for a *generator* on a *congested flowgate*) the product of the *firm access level* and the *participation factor*
target flowgate capacity	the minimum amount of *flowgate capacity* that a TNSP must provide on a congested flowgate to comply with FAS
target non-firm entitlement	(for a *generator* on a *congested flowgate*) the product of the *non-firm access level* and the *participation factor*
target super-firm entitlement	(for a *generator* on a *congested flowgate*) the product of the *super-firm access level* and the *participation factor*
TNSP support	(at a flowgate) the absolute value of the negative *flowgate entitlement* allocated to TNSPs pursuant to a financial quality incentive regime
transitional access	a level of *firm access service* that is allocated to existing generators at the commencement of the optional firm access regime and for which no *access charge* is payable

Table A.2 Abbreviations

Abbreviation	Meaning
AARR	Aggregate Annual Revenue Requirement
AEMO	Australian Energy Market Operator
AFMA	Australian Financial Markets Association
AOC	Abnormal Operating Condition
CPI	Consumer Price Index
DNSP	Distribution Network Service Provider
DSRS	Demand-side Reliability Standards
FAS	Firm Access Standard
FG	Flowgate
FGP	Flowgate Price
FGX	Flowgate Capacity
FIR	Firm Interconnector Right
FTR	Financial Transmission Right
IRH	Inter-regional Hedge
ISRS	Inter-regional Settlement Residue
LHS	Left-hand Side
LMP	Local Marginal Price
LRIC	Long Run Incremental Cost
LRMC	Long Run Marginal Cost
MP	Market Participant
NEM	National Electricity Market
NEMDE	NEM Dispatch Engine
NOC	Normal Operating Condition
NPV	Net Present Value
NTP	National Transmission Planner
OFA	Optional Firm Access
RHS	Right-hand Side

Abbreviation	Meaning
RRN	Regional Reference Node
RRP	Regional Reference Price
SRA	Settlement Residue Auction
TA	Transitional Access
TFR	Transmission Frameworks Review
TNSP	Transmission Network Service Provider
TUOS	Transmission Use of System
UIGF	Unconstrained Intermittent Generation Forecast
WACC	Weighted Average Cost of Capital

Table A.3 Algebraic Variables used in Equations

Variable	Meaning
A	access level
C	marginal generating cost
E	flowgate entitlement
F	forward quantity
FGP	Flowgate Price
FGX	flowgate capacity
FP	forward price
G	dispatched output
IC	dispatched interconnector flow
LMP	local marginal price
Pay$	Settlement Payment. Positive value means payment *from* AEMO settlement
RRP	regional reference price
U	flowgate usage
α	flowgate participation factor