Prepared for the

Australian Energy Market Commission (AEMC)

Congestion Pricing Options for the Australian National Electricity Market: Overview

February 2008

Tendai Gregan[1] and E. Grant Read[2]

[1] Tendai Gregan is a Senior Advisor at the Australian Energy Market Commission (AEMC).

[2] E. Grant Read is the head of EGR Consulting Ltd, Adjunct Professor at the Department of Management, University of Canterbury, and a Senior Consultant with CRA International.

Contents

Introduction

A key design feature in any liberalised electricity market is determining how to reflect network congestion in the pricing of the wholesale electricity market. The regime for congestion pricing in Australia's National Electricity Market (NEM) has been under particular scrutiny over recent years through Rule change proposals affecting pricing in the Snowy Region, and through a general review of congestion management in the NEM undertaken by the Australian Energy Market Commission (AEMC) at the direction of the Ministerial Council for Energy (MCE).

We prepared this paper in the context of the AEMC's congestion management review and it reflects work undertaken with the purpose of informing the AEMC's Draft Report on the congestion management review, published in September 2007. The AEMC funded the work reflected in this paper.

The purpose of the paper is to set out a generic framework for describing the different ways in which network congestion might be reflected in the wholesale pricing and settlement arrangements. Its focus is therefore on enunciating what the different options are and what characteristics they have, rather than analysing the merits of particular options in particular circumstances. It will show that the same basic framework can be used to describe a wide range of options, from a single settlement price for the whole market through to a regime of nodal pricing. It considers the current arrangements in the NEM as an important reference point.

Our framework can also be used to describe, and facilitate, other options for managing congestion already allowed for in the National Electricity Rules that do not involve the introduction of explicit congestion prices, including:

1. Physical intervention in the market:
   - restricting interconnector flows ("clamping"); and
   - giving preference to either local or remote generation (or to one interconnector or another) when both affect a binding constraint.
2. Changes to the regional pricing structure:
   - shifting the location of the Regional Reference Node (RRN); and
   - changes to regional boundaries.
3. Constrained-on/off payments for the provision of Network Support and Control Services, particularly those services provided by gatekeeper generators.

Finally, the framework can also be used to describe a more radical re-definition of the Regional Reference Price (RRP), whereby the RRP is changed into a weighted average of the nodal prices in a region, rather than being the price at a single node in the region.

This common framework and terminology will hopefully help interested parties better understand, and engage in, current and future debates on this important issue of market design.

While this work has been undertaken in the context of the AEMC's congestion management review, and the AEMC has had regard to this work in the recommendations in its Draft Report, this paper represents the individual views of the authors.[3]

[3] The authors thank Colin Sausman, whose editorial work has greatly assisted in clarifying, sharpening and condensing this paper.

1 Theoretical Background

This chapter explains the theory of how we can use the language of congestion rents, and their allocation, to compare a nodally-priced market design with the NEM 's regional market design. This provides a basis for explaining how changes to the NEM's regional design would work. A more detailed exposition of this material can be found in Read (2007).[4]

A nodal market provides a useful benchmark for our purposes because the theoretical and practical properties of nodal markets are well understood, particularly the way in which the economic costs of network congestion are captured by nodal price differences, and the way in which "Financial Transmission Rights" (FTRs) can be used to hedge the associated risks. Describing the NEM arrangements as a restricted form of nodal pricing (with the restrictions relating to how congestion rents are allocated – as explained below) is probably less understood. This approach provides a number of useful insights, and a common language to discuss other options.

1.1 Nodally-priced market designs

In a nodally-priced market design, all participants face the same price at a particular node, whether buying or selling. Nodal prices will differ across the network when there are binding network constraints.[5] Where trading parties are located at different points on the network (i.e. in most instances), there is a price (or "basis") risk associated with network congestion—the risk that the price of the same commodity differs between two locations. We will refer to this as *congestion price risk*.

Congestion price risk could arise as a result of only one transmission limit binding. In a nodal market supported by an interconnected transmission network, this can create price differences between a number of different locations across a large physical area.[6] Additionally, congestion risk can arise from the cumulative effect of a number of different constraints binding.

The calculation of prices in a nodal market is generally supported by a Full Network Model (FNM), which represents the physical characteristics and limitations of all transmission elements in service (and hence constituting the network) at any point in time. The dispatch price at each network location reflects the marginal cost of supply at the location at that point in time, given the prevailing physical network (as characterised by the FNM). In a nodally-settled market, the locational marginal price used to determine dispatch volumes at each point on the network (i.e. the *dispatch*

[4] Read, E.G. 2007, *Network congestion and wholesale electricity pricing in the Australian National Electricity Market: An analytical framework for describing options*, Report to the Australian Energy Market Commission, EGR Consulting Pty Ltd, Christchurch, New Zealand, 8 November 2007. Available at: http://www.aemc.gov.au.

[5] In many markets, nodal prices also differ as a result of losses, but this will be ignored throughout our discussion.

[6] Specifically, the "spring washer effect" means that price differences will arise between any pair of nodes involved in any loop, which can be drawn through the binding limit.

price) is the same as the *settlement price* used to financially settle transactions at that location.

While this provides for cost-reflective settlement pricing at each location, it also creates a price risk management problem for market participants because they typically buy and sell energy at different points across the network, and the prices at these locations will generally diverge in the event of network congestion. Consequently, there is congestion price risk inherent in most trading positions. Market participants therefore have strong commercial incentives to hedge this risk.

The economic rents that accrue when constraints bind can be used as a basis for constructing financial instruments to help manage this price risk. In nodal markets, "Financial Transmission Rights" (FTRs) are the normal vehicle for providing such hedging. FTRs may not "be firm", but they are designed to hedge all congestion risk arising between two nodes, irrespective of which constraints may actually bind. Thus they are (normally) defined independently of network configuration, or of the capacity of any particular link.

In aggregate, these link rents equal the **Congestion Price (CP)**[7] (i.e. the reduction in the overall dispatch costs if the constraint could be marginally relieved) on the binding limit times the MW capacity of that limit. Thus they are sufficient to support a FTR between the two nodes either side of the limit corresponding to the volume of actual transfer across the binding link. Or (more usefully), they can support a wide variety of alternative patterns of "node-to-node" FTRs throughout the network, provided the net effect of all those FTRs implies a flow across the congested link (less than or) equal to the actual observed flow.

An FTR allocation satisfying this requirement is said to possess "revenue adequacy" in the FTR literature. A large number of alternative FTR assignments are possible because only a small proportion of the flow implied by some node-to-node "trades" may actually cross the congested link, and many of those flows typically cancel. Many practical issues arise, particularly determining what FTR volumes the network can support, and with what degree of firmness given the underlying network capacity is not firm. But, in principle, the theory of hedging in such markets is well understood. Our task, here, is to relate that theory to the congestion pricing framework we will use to describe other options.

1.2 The NEM market design

In the NEM, market participants are settled at the Regional Reference Price (RRP) for the Region in which they are located. The RRP is the marginal cost of supply at the relevant Regional Reference Node (RRN).

Conceptually, there is "a hub and spoke" structure underpinning the market design. RRPs are calculated for each RRN "hub" and this is the settlement price for generation and load located in the Region. A spoke to the RRN is, in effect, assumed to exist for any generation or load located within each Region. All inter-regional

[7] This is sometimes also termed the "shadow price of the constraint" in the literature.

trading occurs across notional links directly joining adjacent RRNs, and hence incurs a per-unit cost that is measured by the difference between the respective RRPs.

The NEM is dispatched, however, on the basis of minimising total dispatch costs, having regard to the marginal costs of supply at all locations on the network, not just at the RRNs. Market participants are therefore effectively dispatched on the basis of "Pseudo Nodal Prices" (implicit in the NEM Dispatch Engine (NEMDE) optimisation) but settled at the RRP. Differences between the dispatch price and the settlement price, and the behavioural incentives that such differences create, are commonly referred to as the "mis-pricing" problem.[8]

The NEM does not use a FNM. Rather, the physical limitations of the transmission network are expressed indirectly using "generic" constraint forms that are "oriented" to the respective RRNs. This permits the dispatch to reflect the physical network accurately, such that the dispatch cost can be minimised, and calculate the RRPs simultaneously. NEM constraints must be re-formulated if any of the RRNs change. As discussed in more detail below, this re-formulation of constraints affects the allocation of the economic rents associated with congestion.

A key step in understanding the NEM market design in the context of congestion pricing and rent allocation is to examine the relationship between RRPs, nodal prices and congestion rents. The following equation illustrates the relationship by deconstructing the NEM settlement price for market participant *i* into two component parts. First PNP_i is defined by:

$$PNP_i = RRP_{r(i)} - \sum_{k \in GC} weight_{ik} * CP_k$$

So, rearranging gives:

$$RRP_{r(i)} = PNP_i + \sum_{k \in GC} weight_{ik} * CP_k$$

Where: PNP_i is the **Pseudo Nodal Price** for *i*

$RRP_{r(i)}$ is the **Regional Reference Price** for the region in which *i* is located, *r(i)*.

$weight_{ik}$ is the weight assigned to *i* in generic constraint *k*

GC is the set of all such constraints

CP_k is the **Congestion Price** for generic constraint *k*

[8] See Biggar, D. 2006a, *Congestion Management Issues: A response to the AEMC*, Attachment to AER Submission on AEMC Congestion Management Issues Paper, AER, Melbourne, 12 April 2006, p. 1

This Pseudo Nodal Price is in fact identical to the nodal price which would be calculated from an equivalent nodal model.[9] So settlements based on Pseudo Nodal Prices are effectively nodal settlements. The difference between nodal settlement and NEM settlement is therefore the value of the weighted sum term in the above equation.

So what is the weighted sum? It is the total value of the intra-regional congestion price risk that would be faced by market participant in a nodal market.[10] When constraint k binds, it generates an economic rent. The total economic rent is equal to CP_k, the "price" of the constraint (i.e. the reduction in the overall dispatch cost if the constraint could be marginally relieved) multiplied by the megawatt volume at which the constraint binds (i.e. the constraint "right hand side" (RHS) in NEMDE terminology). When constraint k binds, this can be considered to represent the accumulation of a **Congestion Rental Fund (CRF)** for constraint k. But, because the weighted sum of participant terms on the "left hand side" (LHS) of a constraint equation must equal the RHS, when it binds, this same rent can also be expressed as a sum of participant rental contributions.

The term $weight_{ik}$ is thus a measure of the per-MW contribution of i to congestion in constraint k.[11] It can also be interpreted as the per-MW contribution that would be required from i to the associated CRF when constraint k binds if i were to be "exposed" to the CP in respect of constraint k. The weighted sum in the equation above is hence the total contribution that would be required of market participant i to all the different CRFs it is involved in, if all were priced. It should be noted that the set of CRFs that market participant i could possibly be involved in is equal to the set of generic constraints in which market participant i features. Thus, the first equation above demonstrates that settling market participant i at its PNP is arithmetically equivalent to it being settled at RRP and then making a payment equal to its total contribution to CRFs based on its actual dispatched volume.

Reversing this logic, the second equation above demonstrates that settling market participant i at the RRP is arithmetically equivalent to it being settled nodally and then <u>receiving</u> a payment equal to its total contribution to CRFs based on its actual dispatched volumes. In effect, this additional payment automatically "protects" market participant i from the intra-regional congestion price risk that it would otherwise face given its dispatched volume. Market participant i is (implicitly) granted a right to some congestion rents automatically as a consequence of being

---

[9] CRA introduced the term PNP to refer to a price calculated in this way, but with summation of congestion price terms only over the subset of constraints to which a congestion pricing regime might have been applied. In this paper, we will refer to that constraint sub-set as the "Managed" constraint set, and prices calculated in this way will be referred to as "Adjusted Nodal prices" (ANPs). Throughout this paper we will use the term PNP to refer to ANPs calculated to account for congestion on <u>all</u> binding constraints in *GC*, and also ignore the potential complication of what CRA called "non-NEO" effects, which could create discrepancies between generation and load prices, as discussed in Read (2007).

[10] This risk is "intra-regional" in the sense that it arises between the participant and its own Regional Reference Node. But it is aggregated across all constraints in which the participant is involved, both "intra-regional", and "trans-regional".

[11] This can be positive or negative, depending on the formulation of the generic constraint and the signing convention for market participant i's behaviour.

dispatched. We will call this the Implicit Dispatch Matching Allocation (IDMA) of Congestion Rental Rights (CRRs). Formally, the IDMA bundle of CRRs for participant *i* is a "bundled CRR", or BRR, defined by:

$$BRR_i(\,IDMA\,) = \left\{ CRR_{ik} \middle| CRR_{ik} = weight_{ik} * x_i \,, k \in GC \right\}$$

Where: x_i represents the net injection of participant *i* (generation or load)

CRR_{ik} is a CRR defining access to a share of CRF_k corresponding to CRR_{ik} MW of the RHS capacity of k[12]

The notional settlement payments to holders of IDMA CRRs are given by:

$$Settlement\_BRR_i(IDMA) = x_i * \sum_{k \in GC} weight_{ik} * CP_k$$

1.2.1 Generation and Load

The NEM pricing and settlements rules, in effect, confer IDMA status on all dispatched generation (and load). Thus, all dispatched generation (and load) receives (or pays) its own RRP in settlements, having been implicitly allocated CRRs that match its dispatch volume in all the CRFs relating to binding intra-regional and trans-regional constraints. This IDMA allocation of CRRs gives a load or generator financial access to settlements at the RRP for its entire dispatched volume. That is, under the NEM's regional settlement rules, participants are "protected" from congestion price risk within a region. Formally:

$$\begin{aligned} Zonal\ Settlement_i &= x_i * (RRP_{r(i)} - \sum_{k \in GC} weight_{ik} * CP_k) + x_i * \sum_{k \in GC} weight_{ik} * CP_k \\ &= x_i * RRP_{r(i)} \end{aligned}$$

Note, though, that the CRR volume assigned to the bundle by IDMA is determined ex post for each interval, depending on the actual generation or load volumes. This means that IDMA can not provide firm ex ante hedging for a specified MW level. In other words, it can hedge congestion price risk, but not congestion volume risk. This is an important limitation of the status quo. The point of expressing the status quo regime in terms of a congestion pricing framework is to provide a starting point for considering alternatives that can provide firmer and/or more flexible hedging options by <u>explicitly</u> defining and assigning CRRs in ways that differ from the status quo's implicit allocation.

Observing that under the status quo financial settlement is based on the RRP reinforces the potential desire to find alternative ways of assigning CRRs. The RRP is

---

[12] Read (2007) refers to CRRs of this form as "Constraint Rental Rights" (CRRs) or "Constraint Rental Contracts" (CRCs), if the MW value is fixed. Equivalents Rights/ contracts defined in terms of participant injection MWs are referred to as "Participant Rental Rights/Contracts". But multiplying/dividing by "weight" converts one form into the other, and the distinction is not important here.

what drives the behaviour of market participants and the volumes they wish to produce or consume. Put simply, at times when network constraints are close to binding, market participants seek to gain financial access to the RRP by altering their bids and offers in ways that ensure they are dispatched at desired volumes. Physically, this means they compete to avoid the impact that physical limitations on the network would otherwise have on their dispatch volumes. Financially, we can express this by saying that they compete to obtain CRRs, via the IDMA process, which will give them access to the RRP for their dispatch volumes.

The framework developed by Read (2007) and illustrated in this report allows us to describe alternatives to the status quo in terms of **exposing** various participants, or participant groups, to the congestion prices determined for what will be referred to as the **managed** constraint set. That is the IDMA process would no longer fully protect participants in relation to those constraints. Exposed participants would now face the pricing impact of congestion on those specific constraints. Such regimes can be operationalised by settling the energy market at the RRP, as under the status quo, but then requiring supplementary payments to/from the CRFs for managed constraints, as described above.

1.2.2 Interconnectors

We can think of a directional interconnector as a ″participant″ in this framework also, but the situation is a little more complex. Unlike loads or generators, there is actually an explicit "rental pool" for each interconnector, in the form of the inter-regional settlement residues (IRSR). Currently, IRSRs are auctioned through the Settlement Residue Auction (SRA) process. If an interconnector term involved in a binding constraint were exposed to the congestion price, the interconnector would be required to pay its share (positive or negative) of the congestion rent into the associated CRF. But the IDMA "protects" the interconnector, by granting it a bundle of CRRs matching its dispatched flow. This reverses the notional rental payments from the interconnector, leaving it with a net residue which is not notional – that is the IRSR.

The IRSR for each interconnector is normally thought of as reflecting the difference between the RRPs for the adjacent regions it connects. The IRSR can be determined by multiplying the interconnector flow by the relevant inter-regional price difference. That is:

$$IRSS_{ij} = FLOW_{ij} * (RRP_j - RRP_i) ,$$

Where $FLOW_{ij}$ is the flow on the directional interconnector from region i to region j.

But it can be shown that the inter-regional price difference itself is determined by the combined effect of all constraints in which $FLOW_{ij}$ is involved, and not just those relating to Cross-Border Limits.[13] Formally:

$$RRP_j - RRP_i = \sum_{k \in GC} CP_k * Weight_{ijk}$$

Where $weight_{ijk}$ is the constraint weight for flow from i to j on interconnector ij in constraint k.

If we imagine a CRF as existing for each such constraint, and note that the constraint weights applied to $FLOW_{ij}$ can just as easily be positive as negative, we have:

$$rent\ collected\ \ from\ \ ij\ \ for\ CRF_k = CP_k * weight_{ijk} * FLOW_{ij}$$

But $CP_k * weight_{ijk} * FLOW_{ij}$ is exactly the payout on a CRR defined as a right to a share of the CRF for constraint *k* corresponding to $weight_{ijk} * FLOW_{ij}$ MW of that constraint's RHS capacity. That is, it corresponds exactly to the CRR implicitly assigned to the interconnector by IDMA. In other words, this implicit CRR allocation exactly cancels the rental payment which would be required if the interconnector were exposed to congestion prices for all the binding constraints it is involved in.

Conversely, it is the IDMA allocation of CRRs to interconnectors that creates the IRSR pool. This means that the IRSR can be decomposed into components, which are derived from, and can be re-assigned to, CRF pools for all of the constraints which involve the interconnector. Formally:

$$IRSR_{ij} = FLOW_{ij} * \sum_{k \in GC} CP_k * Weight_{ijk}$$

We can thus see this IRSR pool as representing a <u>bundled</u> CRR, where the bundle is defined to include all constraint forms, for all possible network configurations, which may involve this interconnector. But, crucially, the CRR volume assigned to that bundle is determined differently for each interval, depending on the actual interconnector flow. As a result, it can not provide **firm** ex ante hedging for a specified MW flow.

---

[13] Remember that losses are ignored here. Cross-Border Limits are restrictions on the transfer of power between two regional reference nodes that arise solely from the (security constrained) physical capabilities of the interconnector at the regional boundary, and not as a result of restriction on the interconnector flow arising from transmission constraints within either of the adjoining regions. Cross Border Limits are also referred to as Pure Interconnector Limits (PILs). For example, a thermal limit across the 330kV transmission cut-set between Murray and Dederang will clearly limit power transfers from the Snowy region to Victoria. But Trans-regional Constraints, involving both interconnector and generation terms, also restrict inter-regional power flows, and thus influence the spread between RRPs. See Section 3 below.

This IRSR pool is then auctioned in proportional shares, which might be thought of as "scaled" CRRs. That is, the SRA sells ex ante rights to "bundled" CRRs that are scaled in proportion to the flows on the interconnector. [14]

Thus any IRSR pool with implicit CRR allocation can not provide either firm MW hedging or proportional hedging. Once more, this is an important limitation of the status quo.

1.3 A framework for congestion pricing options

The discussion above uses the reference points of a nodal market and the NEM market design to explain a generic framework for considering other options. More generally, by changing the IDMA status of NEM market participants, and using alternative ways of distributing CRRs for different sub-sets of CRFs, we can consistently describe a wide range of congestion pricing options.

Thus the point of re-expressing the status quo in terms of a congestion pricing framework is to provide a starting point for consideration of alternatives which can provide firmer and/or more flexible hedging by explicitly allocating the CRRs in ways which differ from the allocation implicit in the status quo. This is discussed in the sections that follow. For completeness, the paper concludes with a discussion of structural options which relate to the management of congestion risk but do not necessarily involve "congestion pricing" as such.[15]

[14] The underlying CRRs are not actually scaled in any particular predictable proportions, though, because the IDMA process implicitly defines the allocation of CRRs of differing MW volume all in proportion to a single $FLOW_{ij}$ variable. See Read (2007).

[15] Although these structural options are based on the same underlying theory, and can be also expressed in terms of CRR re-allocation.

2 Congestion pricing options involving interconnectors

2.1 Introduction

This hapter uses the framework and concepts of Congestion Prices, Congestion Rental Funds, and Congestion Rental Rights to describe the existing NEM Inter-Regional Settlement Residue arrangements – and to explain and consider alternative ways of identifying and allocating rents associated with constrained interconnector flows.

The options discussed in this chapter would not change the overall scope of congestion pricing in the NEM. Settlement at the RRP on dispatched volumes would continue for all generation and load. In practical terms, these options would therefore represent relatively small (but still significant) changes to the status quo compared to other options discussed later in this paper. Consequently, there are limitations to their likely economic impacts.

2.2 IRSRs

The SRAs enable market participants to buy shares in the out-turn value of the IRSR for each directional interconnector in the NEM. This can provide a degree of hedging against contract positions between the two Regions connected by the directional interconnector to which the SRA units relate. The payout to SRA unit holders is, however, uncertain ex ante, being dependent on actual flows and price differences during the relevant period.

We can also describe the value of each IRSR in terms of the value of CRRs for a number of CRFs. The flow across any given interconnector will appear in a number of different NEMDE constraint equations. Some of these will relate to limits at the boundary between Regions (Cross-Border Limits (CBLs)), while others will relate to more complex interactions between interconnector(s) and generators.[16] Any one, or several, of these constraints might bind in any given dispatch interval. When this occurs there will be a (non-zero) Congestion Price for that constraint, and the consequential accumulation of a CRF.

Now, imagine that "the interconnector" is allocated shares (or CRRs) in each of these CRFs, with the CRR set ex post to equal the contribution of the interconnector to the binding constraint, i.e. the interconnector flow multiplied by its constraint equation co-efficient. This is the IDMA approach to allocating CRRs discussed above, applied to interconnector flow terms. The cumulative value of these CRRs (over the relevant period) is precisely what holders of SRA units receive, in aggregate, from each IRSR fund. An SRA unit is effectively a share in this pool of CRFs.

The characterisation of the IRSR process above illustrates two quite distinct stages: (1) a process to determine the allocation of congestion rents to the different IRSR

[16] See Appendix C of AEMC (2007) *Congestion Management Review, Draft Report*, AEMC, Sydney, September. Available at: http://www.aemc.gov.au/electricity.php?r=20071010.173831.

pools; and (2) a process to determine the re-distributed value of the IRSR pools, i.e. through the SRA process.

2.3 Options retaining the existing number of inter-regional hedging funds

There is a class of options for change that focus solely on the process of constructing the funds which are subsequently re-distributed to SRA unit holders. Using the above terminology, this involves changing how different bundles of CRRs are allocated to the different inter-regional hedging funds.

However, as noted above, exposing interconnectors to congestion rents in this way does not change the settlement rules for other market participants, e.g. loads and generators. The IDMA method is assumed to continue to apply for all other market participants. This means that the total rent assigned to interconnectors can not change. Absent an external source of funding, an allocated CRR can only provide one interconnector with a larger claim on the CRF than the value of IDMA-determined share, if some other interconnector is assigned a lesser share. Therefore, we can only apply this type of option, if we limit it to instances where there is more than one interconnector flow term involved in a constraint. While this clearly limits its application, there are a significant number of such constraints in the NEMDE constraint library.[17]

Constrained flows on interconnectors might interact if interconnectors were allowed to form loops between regions in the NEM market structure. The current NEM market design has no such loops, but the underlying network reality involves cross-border loops.[18] But actually, the NEM market design means that interconnector interactions can occur even though the interconnectors do not physically interact at all, and are not involved in inter-regional loops. In fact, we would expect this to be normal whenever a region contains intra-regional network loops, and has more than one interconnector.[19]

The Snowy constraint situation, relating to line flow limits between Murray and Tumut, provides a case in point, and we use it to illustrate the implications of some of the concepts discussed here. This example has been produced by solution of a simplified LP dispatch model. It is adapted from an example in CRA(2004d), also discussed in CRA(2004c), and in an appendix to CRA(2004b). Loops within the NSW and Queensland (Qld) regions provide other examples, two of which are discussed in the above references. CRA(2003b) provides a detailed illustration of the application

[17] See Section C4 in Appendix C, AEMC (2007) CMR Draft Report.

[18] In fact there are cross border loops involving parallel interconnectors, such as QNI and Directlink, but such situations have only been allowed to develop where one of the parallel interconnectors is HVDC and they do not form a loop in between three or more regions in the NEM market structure. These particular situations were examined by CRA (2004c), and that analysis should be re-visited before coming to any conclusions with respect to the applicability of the concepts developed here to such situations. CRA(2004c), *NEM Interconnector Congestion: Dealing with Interconnector Interactions*. Released by NEMMCO, October 2004.

[19] An exception is if those interconnectors all feed into the loop structures through a common section of unlooped network, which seems unlikely.

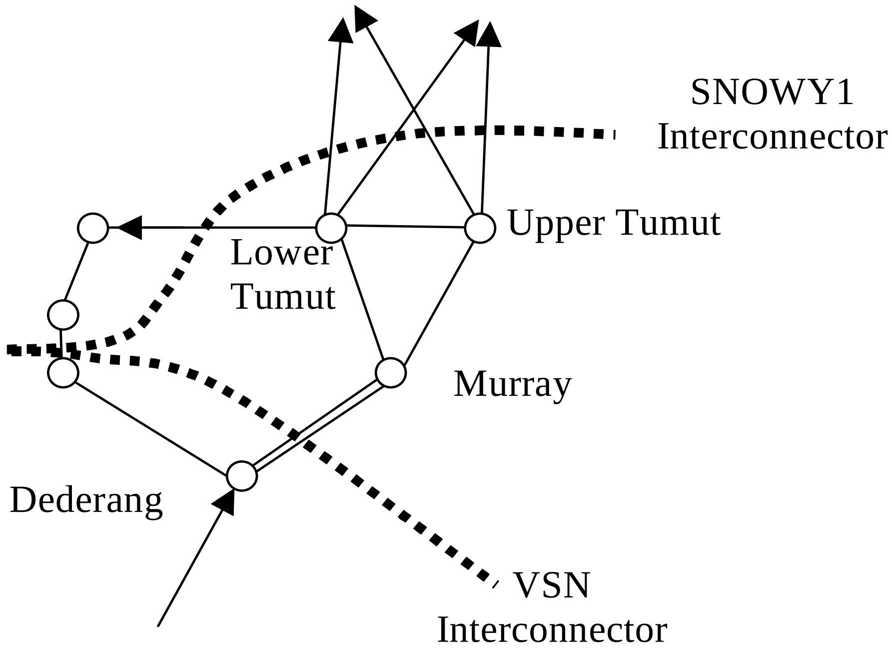
of the Constraint Support Pricing/Constraint Support Contract (CSP/CSC) proposal to an example based on the Tarong constraint in Southern Queensland. Another example occurs in the case of NSW, where the transmission rings around the Sydney/ Newcastle/ Wollongong area have NSW generation connected to them, as well as the Snowy 1 and Qld-NSW interconnectors.

In all of these circumstances, it would be possible to adopt an alternative to IDMA for allocating CRRs to interconnectors, as discussed in this section. This would involve aggregating the CRRs that would otherwise accrue to the interconnectors under IDMA, and using a pre-defined rule to re-allocate them between the interconnectors. Put another way, an (uncertain) amount of constrained capacity is being shared between the "interacting interconnectors", in accordance with defined rules, rather than being allocated on the basis of the dispatched interconnector flow volumes. This does not affect the dispatch, but it does involve a post-settlement financial adjustment to, in effect, transfer funds between the affected inter-regional hedging funds if the dispatched interconnector flows are different from the pre-defined shares.

The implementation of this type of arrangement would need a process for determining the rule for allocating CRRs to the interacting interconnectors. A discussion on the establishment of such rules or the merits of different approaches is outside the scope of this paper. For the purposes of the illustrative example below we assume that a simple "rule" has been agreed to allocates 75% of the available CRRs to SNY1 (interconnector between Snowy and NSW) and 25% to VSN (interconnector between Victoria and Snowy).

2.3.1 Illustrative example

We can consider a simplified (but realistic) example involving the Snowy Region to illustrate the discussion. The Snowy Region has generation at Upper Tumut, Lower Tumut and Murray, connected in a loop within the Region – with the RRN located at Murray. There is also a network loop involving Murray and Lower Tumut (and Dederang, in the Victoria Region), which passes through the New South Wales, Snowy and Victoria Regions. It is set out diagrammatically below.



In this type of setting, there could be a number of different types of constraint form in play. We will simplify this to three possible types: a Cross Border Limit (CBL) between Victoria and Snowy; a CBL between Snowy and New South Wales; and a "trans-regional constraint" capturing the interactions between generation output within the Snowy Region and flows across the two interconnectors. The trans-regional constraint is assumed to take the form:

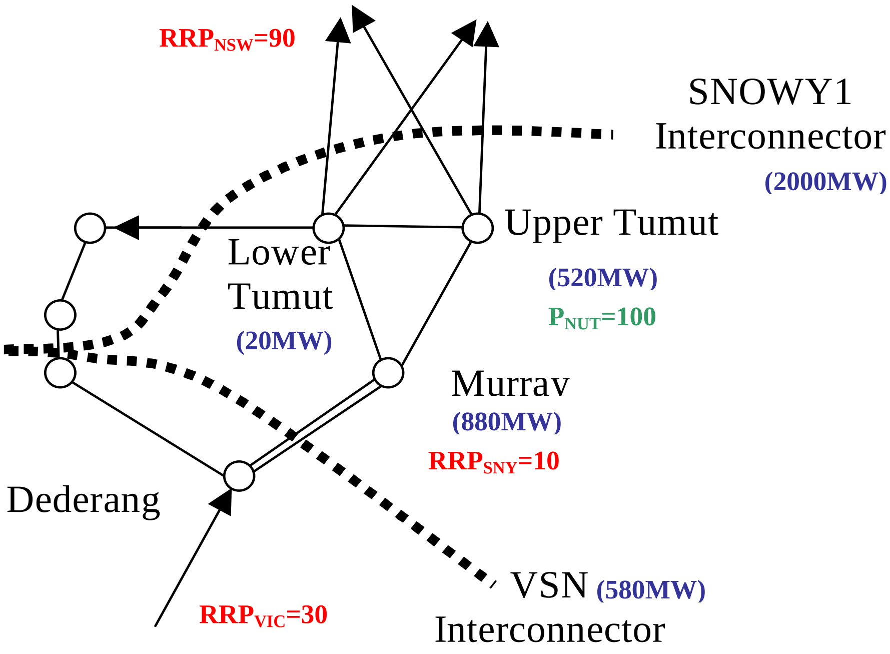
$$0.8*(SNY1-NLT) - 0.2*VSN - 0.9*NUT \leq 1000$$

This characterises a situation where increased output at Lower or Upper Tumut, or increased flow on VSN act to relieve the constraint (i.e. enable or "support" increased flow on SNY1), while increased flow on SNY1 acts to increase the likelihood of the constraint binding. The two CBL constraints are assumed to take the form:

$$SNY1 \leq 2001$$

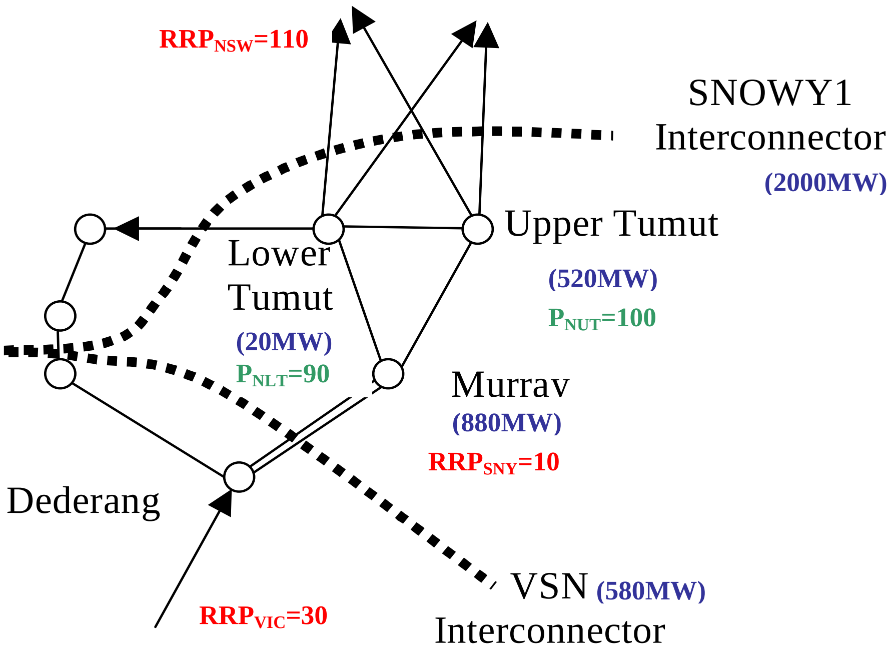
$$VSN \leq 1000$$

It can be shown that under reasonable bid and offer assumptions, a dispatch solution in these circumstances can deliver the following minimum-cost dispatch solution – in which only the trans-regional constraint binds.



The binding trans-regional constraint has a Congestion Price of $100. This is because relaxing the constraint by 1MW would allow (1/0.8)MW of flow from SNY to NSW, providing a gain of $80 (=$90-$10).

We can then allow for a second binding constraint. If the CBL between SNY and NSW was 2000MW, rather than 2001MW, then that constraint would bind also. This gives the following outcome:



Under these circumstances the value of the IRSRs will be as follows:

$IRSR_{SNY1}$ = Flow * Price Difference = 2000 * (110-10)

= $200,000

$IRSR_{VSN}$ = Flow * Price Difference = 580 * (10-30)

= -$11,600

There are two points worth noting here. First, under an efficient dispatch with no "dis-orderly" bidding, the IRSR for VSN is negative.[20] Second, the IRSR for VSN has accumulated funds (albeit negative) despite the fact that the CBL between Victoria and Snowy is not binding.

Now we can examine, precisely, how the IRSR funds are constructed. In our example there are three constraints – and therefore three possible individual CRFs. The VIC-SNY CBL does not bind, and so must have a CP of zero. The CP for the binding trans-regional constraint is still $100. But the extra $20 price differential between Lower Tumut and NSW, which have the same weight in the trans-regional constraint and so experience the same pricing impact from that constraint, is solely due to the additional binding constraint between them. In other words, the CP for this CBL is $20.

Thus the contributions of the exposed interconnector terms determine the value of the CRFs:

CRF_{Trans} = {(0.8*SNY1) – (0.2*VSN)} * $100

= $160,000 - $11,600

= $148,400

CRF_{CBL_SN} = SNY1 * $20

= $40,000

CRF_{CBL_VS} = VSN * $0

= $0

20 Dis-orderly bidding occurs when a generator seeks to align its dispatch volumes with the volumes it desires at the prevailing RRP (at which it will be settled). Dis-orderly bidding is a response to being constrained on or constrained off relative to the RRP. When a generator is faced with either being constrained off or constrained on, it has incentives to bid in a manner consistent with seeking to be dispatched at its desired volume (e.g. by bidding -$1,000), or seeking to avoid being dispatched at a greater volume than it wishes to be at the prevailing RRP (e.g. by bidding VOLL).

The IRSRs consist of the shares of the individual CRFs allocated to the relevant interconnectors. Hence the SNY1 IRSR consists of $160,000 from the CRF associated with the binding trans-regional constraint plus $40,000 from the CRF associated with the binding Cross-Border Limit constraint between Snowy and NSW. And the VSN IRSR consists of -$11,600 from the CRF associated with the binding trans-regional constraint.

We have formed three CRFs from the two IRSRs by "stripping out" the rents associated with the trans-regional constraint and placing them in a separate fund, CRF_{Trans}. And it would be possible to auction shares in each of these CRFs separately[21]. However, if we wished to retain the characteristic of a single pool of rents for each directional interconnector, then we could define a rule ex ante for allocating the contents of the trans-regional CRF pool back into interconnector-specific Inter-Regional Hedging (IRH) pools. In effect, the existing IRSRs are IRHs that are constructed using the IDMA method of allocating CRRs to decide how to distribute shares in the trans-regional CRF. Thus, the funds flowing to a particular IRSR depend on the dispatch outcome when the trans-regional constraint binds.

The trans-regional CRF of $148,400 at a CP of $100 represents 1,484MW of constrained RHS capacity. The IDMA approach, in this example, allocates 1,600MW to SNY1 and -116MW to VSN. This is sufficient to support 2,000 (=1,600/0.8) MW of flow on SNY1 and 580 (=-116/-0.2) MW of flow on VSN. The sharing out of available constrained RHS capacity, for the purposes of allocating congestion rents to IRSRs, depends on the dispatch outcome – and therefore is not known ex ante.

An alternative approach is to define the shares ex ante, which should be much firmer than allowing the dispatch to determine the shares.[22] For example, we could establish a rule which allocated 75% of the available RHS capacity to SNY1 and 25% to VSN, irrespective of the dispatch outcome. The choice of allocation shares is entirely arbitrary, and merely chosen for the sake of creating an example[23]. Under this scenario, the aggregated IRHs relating to SNY1 and VSN would be as follows:

$$IRH_{Sn\_N} = \$40{,}000 + \{(0.75*1484) * 100\}$$

$$= \$151{,}300$$

$$IRH_{V\_Sn} = \$0 + \{0.25*1484) * 100)$$

$$= \$37{,}100$$

If such a CRR allocation is agreed, then we can define the hedging available from the inter-regional hedging pools by the CRR allocation, not by the IRSR observed in the spot market. Thus it can be made firmer, ex ante, than under the statue quo.

---

21 This is what the Constraint Based Residues (CBR) proposal envisages, as discussed below.

22 For a fuller discussion of the concept of "firmness", see Read (2007).

23 A more extreme, but perhaps more obvious, example would be to assign a 0% share to VSN, thus eliminating any trans-regional congestion rent component, either positive or negative from that IRH.

The IRH will be positive, provided the CRF is positive, and the interconnector will receive a positive CRR allocation.[24] If constraints are expressed in standard < form, the CP will be positive, as will the CRF, provided what Read (2007) refers to as the *Protected RHS* is positive. The *Protected RHS* is the value of the RHS if all terms that are afforded IDMA status are moved from the LHS to the RHS. In the context of this section, the *Protected RHS* includes everything other than interconnector flow terms. If the *Protected RHS* is positive, then it will always be possible to create a positively valued IRH pool, if desired, for each interconnector.

Each CRR can still only be as firm as the *Protected RHS* of its own constraint. Thus, if only interconnector terms are exposed, CRRs will only be as firm as *Protected* RHS_F, formed by placing all other terms, including generator variables, on the constraint RHS. In other words, in this case, the *Protected RHS* would still reflect all variations due to generation, TNSP, ancillary service, and load terms. Compared to the status quo where <u>all</u> terms are effectively on the Protected RHS, the only uncertainty eliminated is the variations in the level of inter-regional flows.[25]

2.4 Options which increase the number of inter-regional hedging pools

There is another class of options that uses the same approach to de-construct the IRSR funds into the individual component CRFs, but it does not re-aggregate the CRFs to form the same number of inter-regional hedging funds as there are IRSRs. The purest example of this class of options is where the CRFs are not re-aggregated at all. Each CRF is treated as its own individual inter-regional hedging pool. In this context, in which only interconnector flows are exposed to CP, we can create a specific application of the more general Constraint-Based Residues (CBR) approach described by Biggar (2006b). In the example above, the de-constructed IRSRs produce the following three IRH pools:

IRH_{Trans} = \$148,400

IRH_{Sn_N} = \$40,000

IRH_{V_Sn} = \$0

The SRA process could be adapted to replace the auctions for the VSN and SNY1 IRSRs with three separate auctions for the IRHs relating to the two "pure" CBL constraints, and the trans-regional constraint. Market participants (or financial intermediaries) would need to decide how to construct the desired forms of inter-regional hedges from the three types of auction units (now) available.

---

24 Or, if both are negative.

25 Alternatively, partitioning the CRF pool, and hence implicitly the constraint capacity, in this way may be seen as conceptually equivalent to using CRCs with some parties to augment the firm hedging available to other parties, as discussed in Read(2007). In this case, though, we are not contracting with non-traders to provide more firm hedging to traders. Rather, we are contracting with one interconnector to provide more firm hedging to another interconnector, and vice versa. Just as in that discussion, firm contracting with one party actually also increases the firmness of hedging available to the other.

3 Congestion pricing options involving generators and interconnectors

3.1 Introduction

This chapter extends the discussion in the previous chapter by considering options that involve generators in the congestion pricing regime. Again, the framework and concepts of CP, CRF, and CRR provide a basis for the discussion. We will consider the inclusion of generators as an incremental step, i.e. we will continue to include interconnector terms. For completeness, it is feasible (although somewhat unlikely, from a practical perspective) to design options which involve generators, but do not involve interconnectors. Conceptually, including generators is a relatively straightforward extension of the options discussed in the previous chapter. However, practically it represents a more significant departure from the current NEM market design because, unlike interconnectors, generators are not currently involved in any kind of congestion pricing mechanism (other than the Snowy Trial).[26]

The options discussed in this chapter would not change the current NEM arrangements of settling loads on the basis of regional prices. The options would, however, expose at least some generators, in some form, to congestion pricing. In effect, this would represent a shift away from regional pricing for generators towards something more closely resembling nodal pricing.

If the options discussed in this chapter were applied to all generators in all circumstances, then this would represent generator nodal pricing. However, the framework is sufficiently flexible to describe options that could be applied in a much more focused manner (in the extreme, involving a single generator in a single constraint equation). Limited applications would appear to have more practical relevance to the NEM, given the current stated position of policy makers.[27]

3.2 What "exposure to congestion prices" means for generators

We explained above how to characterise the current NEM arrangements as involving all parties being exposed to congestion prices, but then implicitly receiving an automatic allocation of rights (CRRs) which "protects" them from certain forms of congestion price risk for their dispatched volumes. Again, this is the IDMA method of distributing CRRs. The practical effect for generators is to provide them with financial access to their local RRN for their dispatched volume.

---

[26] The Snowy CSP/CSC Trial is discussed in Appendix E of AEMC 2006, *Management of Negative Residues in the Snowy Region*, Final Rule Determination, 14 September, Sydney. Available at: http://www.aemc.gov.au/electricity.php?r=20051214.200416.

[27] Page 3 of the MCE's Terms of Reference for the Congestion Management Review stated that "no material efficiency benefits would be gained from a nodal pricing approach at this stage of market development."

We can illustrate the effect of IDMA for generators using the Snowy example discussed in the previous chapter. In the example, there were three constraint equations – but only one equation (the "trans-regional constraint") which involved any generator terms. In this equation, the two generators (Lower and Upper Tumut, NLT and NUT respectively) have negative co-efficients (-0.8 and -0.9, respectively). This represents a situation where increased output by a generator relieves the constraint. Additional generation output, in effect, "supports" additional interconnector flows.

We can now extend the scope of the CRF formed by the binding trans-regional constraint to include generator terms for NLT and NUT:

$$
\begin{aligned}
CRF_{Trans} &= \{(0.8*SNY1) – (0.8*NLT) – (0.2*VSN) – (0.9*NUT)\} * \$100 \\
&= \$160{,}000 - \$1{,}600 - \$11{,}600 - \$46{,}800 \\
&= \$100{,}000
\end{aligned}
$$

Hence, NLT and NUT are now exposed to the consequences of the trans-regional constraint binding. In this particular example, exposure is beneficial for NLT and NUT because the amounts they are "paying in" to the CRF are negative (-$1,600 and -$46,800, respectively). Put another way, exposure to congestion price risk, in the context of this constraint equation, would mean settlement at a price <u>higher</u> than the RRP. This illustrates a more general point that protection from congestion price risk pursuant to IDMA creates winners and losers, relative to being settled at a generator's RRP.

We note that exposing NUT and NLT thus reduces the value of the CRF (by $48,400). Intuitively, this reflects a reduction in the merchandising surplus pursuant to settling NUT and NLT at a price higher than the RRP (while not changing the prices paid by load), in these particular circumstances.

3.3 Management of congestion price risk for generators

We can envisage a congestion pricing option which specifies a number of constraint equations "to be managed", and exposes generator terms and interconnector flow terms within those constraint equations to congestion price risk. If any of the specified constraint equations were to bind, then a CRF would form for each constraint equation that bound. The dispatch volumes and co-efficients within the binding constraint equations for each of the involved parties would determine the contributions into each CRF. For generators, each contribution into a CRF (which could be positive or negative) would, in effect, be funded by variations away from the RRP in the effective price they receive in settlements. On a practical level, it is worth noting that a congestion pricing option focused on a particular area might involve creating CRFs for a (potentially large) number of individual constraint equations, e.g. reflecting differences in network capabilities due to outages or weather conditions.[28]

If generators, as a result of being involved in a congestion pricing scheme, face an additional price risk, then they will clearly have a commercial interest in the means to manage such price risk: that is in the acquisition of CRRs.

Two broad approaches to CRR acquisition may be seen as determining how such congestion pricing schemes might be designed in practice. First, a "passive" approach would simply provide interested parties with opportunities to buy (or sell) CRRs, i.e. shares of the (now explicitly valued) CRFs. Second, an "active" approach would allow, and perhaps require, a third party to determine an administrative rule for allocating CRRs to some or all parties involved in the congestion pricing scheme. Alternatively, this third party may be allowed to negotiate a CRR allocation between some or all at the parties involved in the congestion pricing scheme. The current NEM arrangements could be considered as a special case of "active" management, with the administrative rule adopted being IDMA-based allocation of CRRs.

The two approaches would both use CRRs relating to individual constraints as "building blocks", and could implement a congestion pricing scheme using derivatives products. Three obvious derivatives would be:

- "Bundles of CRRs" defined across a number of constraint equations, e.g. a set of all constraint equations of the same form (i.e. involved parties, and co-efficients) but different RHS limits, or a set of all constraint equations (of whatever form) relating to a particular physical limit;

- "Bundles of CRRs" defined over time, e.g. a share in the sum of the CRFs accumulating in respect of an individual constraint equation over a specified period of time; and

- Rights defined in fixed-megawatt terms rather than as shares (depending on the RHS limit when the constraint equation binds) of the rents associated with a constraint equation (or bundle of constraint equations) as with the FTRs traded in many nodal markets.

The following two sections discuss the design characteristics of "passive" and "active" congestion pricing options for interconnectors and generators in more detail. The starting point for both approaches is the same. From a practical perspective, one would need to identify a set of constraint equations and a set of participants involved in those constraint equations to include in the scheme. A CRF would form whenever one of the specified constraint equations bound. The contributions of the parties involved would determine the value of the CRF, as illustrated in the examples above.

---

[28] The same physical restriction, eg of flow on a particular line, may give rise to many different constraint "forms", with different "weights" depending on the network topology at the time, and different RHS volumes, depending on loads, for example. Differing weights give the constraint differing mathematical and hence hedging, properties, while differing RHS values alter the volume of hedging available, and hence the MW volume of any hedges defined as "shares". Thus perfect hedging can only be achieved with respect to this whole range of possible constraint forms if a perfectly hedged position can be achieved, ex ante, with respect to each individually.

Various schemes may then differ with respect to details such as the form of CRRs defined with respect to those CRFs, the degree of bundling etc. The main difference between a passive and active approach, though, is the means by which participants acquire CRRs. This relates to the reasons why a congestion pricing regime might be introduced in particular cases.

3.3.1 Passive management options

The two passive options specify a framework for determining how parties might obtain (shares in) the CRFs revealed by creating the congestion pricing scheme, but do not specify who those shares go to. It is for individual market participants to take active steps to procure what CRRs they need to manage their trading risks. The implementation of the scheme is limited to defining the product to be traded and providing the infrastructure for selling the product in the first instance.

The framework of CBR is an example. While the CBR proposal seems motivated in the context of a NEM-wide implementation, i.e. wider in scope than the options discussed in this chapter, the approach could equally be applied to a subset of participants involved in a subset of constraints. We will describe the regime in such terms for ease of comparison.

The CBR proposal is quite explicit on the form of the product being traded.[29] A separate product is envisaged for each CRF in each dispatch interval. This means that participants, who want, say, a hedge from their participant nodes to the RRN, would have to assemble an appropriate product from their component CRRs, or buy them from an agent who assembled them in this form.[30] The CBR proposal also defined CRRs as “shares” in the relevant CRFs, thus preserving strict revenue neutrality for each CRF in each dispatch interval. Thus they would not have fixed MW values, and would not provide hedging to match fixed MW trades.

The CBR proposal is also quite specific in stating that the only method for assigning its CRRs to participants would be by auction. We do note, though, that in many cases (e.g. the Snowy case above) there will be participants who can only meet their hedging requirements by selling CRRs into the auction, rather than buying them from it. In such circumstances, the auction would operate more like a two-sided open tender with some participants buying, and some selling, CRRs for the CRF. Apart from this, though, the nature of each CRF auction is simple, and each can be conducted independently of any others.

In practice, such a regime could be appended to the existing settlement arrangements relatively straightforwardly. In the event that any of the constraints binds, the generators and interconnectors will be settled at an effective price that includes the impact of congestion prices relating to the selected constraints, rather than settled at the RRP. That is, exposure revokes their IDMA allocation of CRRs that relate to the

---

[29] Biggar 2006b, *Solving the Pricing and Hedging Problems in the NEM Using ‘Constraint-Based Residues’*, Report to AEMC, 25 October 2006. (Available at: http://www.aemc.gov.au/electricity.php?r=20070416.124114).

[30] See Biggar 2006b.

set of managed constraints. However, the generators and interconnectors would still gain CRRs by IDMA for all the other binding constraints that are not included in the congestion pricing scheme. This is achieved by settling energy transactions at RRP, as under the status quo, then performing an explicit secondary settlement with respect to CP on the particular constraints which are being managed.[31]

There would be significant implementation issues, however, in establishing and settling the additional auctions for each of the CRFs (or bundles of CRFs, in non-CBR variants of this kind of approach).[32]

For generators who may be constrained off, exposure to CP represents a loss, because their current implicit CRR allocation allows them to access the RRP, which is higher than their "Adjusted Nodal Price" (ANP), given the set of constraints being managed under the CP regime.[33] Thus the CRRs which they would need to purchase, in order to be hedged with respect to the RRP, would have a positive value, and they would need to pay for the access they currently enjoy for free. This seems to be the situation implicitly envisaged as normal (for both generation and load) in the CBR proposal document (Biggar 2006b).

For generators who may be constrained on, though, exposure to CP represents a gain, because their current implicit CRR allocation forces them to accept the RRP, which is lower than their PNP. Thus the CRRs, which they would need to "purchase" in order to be hedged with respect to the RRP, would actually have a negative value, and they would only accept them if paid to do so. In other words, they would be in a position to sell positively valued CRRs into the auction process.

The Snowy example illustrates such a constrained on situation. For illustrative purposes we will first assume that interconnectors are not involved in the scheme (i.e. they receive the IDMA allocation of CRRs).[34] This means that the CRF is formed by the contributions from Lower (NLT) and Upper Tumut (NUT):

31 Participants that are not exposed to the congestion prices arising from the selected set of constraints including in the CP scheme continue to be fully protected from CP via the NEM's settlement rules. Protected participants receive implicit CRRs via an Implicit Dispatch Matching Allocation. The consequence of this is that protected participants continue to be settled at the RRP for their entired dispatch volumes.

32 If interconnectors are exposed in all constraints in which they are involved, there will be no residual IRSR to auction, since it will all be assigned to CRFs. Otherwise, there can be a residual IRSR, corresponding to the residual IDMA allocation of CRR for constraints which are not managed, and this would presumably be auctioned, too.

33 Or, in the limit, PNP, if all constraints are managed under the CP regime.

34 It is worth noting that, if interconnectors were really not involved (i.e. they were still "protected"), this trans-regional constraint would not be treated any differently, from a generator perspective, than an intra-regional constraint. Thus this same example may serve to illustrate either. All that differs, from a generator perspective, is that the Protected RHS of a trans-regional constraint will exhibit greater variation as a result of variation in interconnector flows. Such variation will obviously make it more difficult to achieve firm intra-regional hedging, and possibly much more difficult. But this is a difference in scale, not in kind.

$$CRF_{Trans} = \{-(0.8*NLT) - (0.9*NUT)\} * \$100$$

$$= -\$1,600 - \$46,800$$

$$= -\$48,400$$

The value of the CRF is negative because the two generators are "supporting" the interconnector flows observed in the dispatch. It is not clear why generators would be willing to supply such support under the status quo, unless their marginal generation cost was actually less than the SNY RRP. However, under a passive congestion management regime NLT or NUT might be willing to be settled at the Snowy RRP (which is lower than the settlement price they would otherwise see, because of their involvement in the congestion pricing scheme) if they were appropriately remunerated. This would involve being paid, rather than paying for, the CRRs to the CRF above.

Also, if the generators are being asked to sell CRRs to the CRF, and these CRRs represent the support they are agreeing to provide, we must ask how to fund this support. The logical answer is that, if generation is constrained on to support interconnector flows, then that support should be funded by payments received from traders buying inter-regional hedging. And, logically, this would be achieved using a two-sided auction process in which both generators and interconnectors were exposed to CP, and the volume of support provided by generators, and the volume of hedging on each interconnector, would all be determined by market interaction.

It seems obvious, though, that market power would be an issue in this situation. The volume of hedging available to participants wishing to trade between Victoria and NSW would depend significantly on the volume of CRRs which NTL and NUT collectively choose to sell in to the auction. And the price of hedging between Victoria and NSW would depend on the prices which NLT and NUT wish to charge for those CRRs. Hence, in part, the motivation for considering more "active" approaches, in which at least some of the CRR allocation arrangements might be negotiated in advance and, for example, made a condition of exposing parties who might be expected to enjoy market power under the CP regime.

3.3.2 Active management options

Differing from passive management options, active options have a party (or parties) take responsibility for specifying a set of administrative rules and/or negotiation processes for allocating CRRs (i.e. distributing the funds that accrue in the CRFs revealed by the introduction of the congestion pricing scheme). Such arrangements could be formalised in the National Electricity Rules (NER) as an adjunct to the settlement rules or could be given effect contractually outside the NER. They could apply only to a subset, or aspect, of the situations created by a CP regime. For example, they could apply only to specific "interconnector support" situations in which market power was an issue, and perhaps then only achieving a negotiated settlement with particular parties to supply CRRs to a CRF or IRH pool, or to allocate CRRs between IRH pools, to be auctioned subsequently as above.

As an example the CSP/CSC approach was developed over several years as a possible modification to the NEM's regional pricing and settlements regime. Early versions of this approach were the “gatekeeper” generator settlement models, developed during 2002 and 2003 to address situations where gatekeeper generators could potentially reduce the economic costs of dispatch by changing interconnector flows, but were not sufficiently motivated to do so under the NEM’s regional settlements arrangements.[35] The Snowy example above illustrates one possible case in point. The gatekeeper models were generalised and extended in 2004, and evolved into the CSP/CSC style congestion pricing options recommended to the MCE by consultancy firm Charles River Associates (CRA).[36]

Using the Snowy example, we can illustrate the application of an actively managed congestion pricing regime and how it compares with a passive regime. Extending the previous example, by removing IDMA-based allocation of CRRs from both the interconnector flow and generator terms involved, we create the following three CRFs, in the example dispatch interval:

$$CRF_{Trans} = \{(0.8*SNY1) - (0.8*NLT) - (0.2*VSN) - (0.9*NUT)\} * \$100$$

$$= \$160{,}000 - \$1{,}600 - \$11{,}600 - \$46{,}800$$

$$= \$100{,}000$$

$$CRF_{CBL\_SN} = SNY1 * \$20$$

$$= \$40{,}000$$

$$CRF_{CBL\_VS} = VSN * \$0$$

$$= \$0$$

The trans-regional CRF is the main point of interest, from a design perspective, because an active manage approach could involve the establishment of rules to determine how to distribute the CRF between:

- inter-regional and intra-regional hedging; and/or
- different generators within a region; and/or
- different interconnectors.

---

[35] See CRA(2003b), *Dealing with NEM Interconnector Congestion: A Conceptual Framework*. Release by the National Electricity Market Management Company of Australia (NEMMCO), March 2003.
See CRA(2004c).

[36] See CRA(2004a), *NEM Regional Boundary Issues: Theoretical Framework, Final Report*. Released by the Ministerial Council on Energy (MCE), September 2004.
See CRA(2004b), *NEM Transmission Region Boundary Structure, Draft Report*, September 2004. Released by the MCE, September 2004.
See CRA(2004d), *Review of NEM Transmission Region Boundaries: Consultation Draft*. Presentation to MCE on 19 & 20 October 2004.

CRA's work implicitly assumed that inter-regional hedging would be handled as under the status quo, by auctioning the modified IRSR pools. This would require establishment of a rule to allocate the residual CRF, after generator transactions are accounted for, to the relevant IRH pools, as discussed in the previous chapter. In the example, the total trans-regional CRF pool is $100,000 (i.e. 1000MW* $100/MW). The aggregate allocation to inter-regional hedging pools is $148,400, and the residual CRF relating to generator transactions is:

$$
\begin{aligned}
CRF_{Trans} &= \{-(0.8*NLT) - (0.9*NUT)\} * \$100 \\
&= -\$1,600 - \$46,800 \\
&= -\$48,400
\end{aligned}
$$

Possession of a CRR to this negatively valued CRF results in a generator receiving a lower settlement price than the RRP. Generators would not take on such an obligation voluntarily and, once exposed to CP, would have significant market power in any passive auction/tender process. The active management approach assumes that a process can be followed which results in generators agreeing to take on such an obligation. The obligation could be met physically by generating to the pre-agreed volume, or financially by making (or receiving) a difference payment if physical output is under (over) the contracted volume. CRA envisaged a negotiation process to agree on such obligations, with such agreement most likely being required as a condition of allowing these generators, who will gain from exposure to CP, to join the CP regime.

This begs the important design question as to which parties might negotiate such a solution with generators. Transmission Network Service Providers (TNSPs) could potentially assume such a role, and indeed might have a financial incentive to do so under some forms of revenue regulation—particularly when the situation involves interconnector support, which could be considered to be a close substitute for transmission investment. Similarly, a group of generators may have a financial interest in supporting increased (and firmer) interconnector flows—although negotiating between direct competitors for provision of a service is not without its complications. Further, the NER could oblige NEMMCO to examine the scope for such negotiated enhancements, although NEMMCO, by design, does not have the discipline of financial incentives in undertaking such a task.

For illustrative purposes, we will abstract from the practical question of how an allocation of CRRs to generators is agreed or determined, and assume an allocation is in place. The CRF of -$48,400 represents "support" of 484MW of additional interconnector flow. Put another way, if NUT and NLT had not generated, then the constraint would have bound at a level 484MW lower. For simplicity, we will assume that this is, actually, the agreed aggregate level of interconnector support. And we will assume that the provision of 450MW of support is allocated to NUT and 34MW to NLT. To provide 450MW of support, NUT must generate 500MW (= 450/0.9), and NLT must generate 42.5MW (= 34/0.8).

In our example, though, NUT actually generated 520MW and NLT actually generated 20MW.

So, NUT has provided more support than it was contracted to do. The extra 20MW it generated provided an extra 18MW (= 20*0.9) of support. At a CP of $100 this is worth $1,800. Conversely, NLT provided less support than it was contracted to provide. The 22.5MW it did not generate represented a shortfall of 18MW (= 22.5*0.8) of support. Similarly, worth $1,800. The supplementary settlement (over and above being settled at the Snowy RRP for their dispatch output) to give effect to the congestion pricing scheme would therefore transfer $1,800 from NLT to NUT.

This intra-regional revenue neutrality is an artefact of the construction of this particular example, though.[37] In general, we would expect that generators will provide more, or less, than the contracted support level, in aggregate. Thus they would make, or receive, corresponding payments form the CRF. Effectively, this means they are either selling, or buying back, incremental interconnector support/hedging capacity for the inter-regional trading market. Such incremental trading is quite appropriate, and will be appropriately driven by inter-regional prices at the time. Indeed, it may be argued that these price signals are all that is required, and that no active congestion management should be required at all. Leaving risk aside, this is correct, at the "first order" level[38]. But the point of prior contracting here, as is often the case, is not just to guard against risk due to random events, but also risk arising from the "second order incentives" which some potential support/hedging providers would otherwise have to restrict support, raising support prices and inter-regional price differentials, and reducing market efficiency.

Overall revenue neutrality could be imposed by expressing the CRRs as percentage shares of the CRF, as in the CBR-style approach. But this would mean parties would not know, ex ante, what output would meet their contractual obligations, and may not be able to physically match it.[39] Thus, when proposing "gatekeeper" arrangements, CRA assumed that particular parties would be paid to provide particular levels of MW "support" under particular circumstances. The contractual mechanism employed would involve a Fixed MW CRR (CSC in CRA's terminology) and create payments to/from the constraint CRF and create hedges which could be purchased by traders affected by the constraint. This would "firm up" the CRF/IRH pool, but that pool would still experience residual variability, due to variability in the terms not covered by the CSP/CSC arrangements—such as loads, transmission capacity, and perhaps ancillary service providers (i.e. from the terms on the "protected RHS" in our terminology).

The revenue generated from selling hedges would be a measure of the value added to the market, and should, in principle, be sufficient to cover the costs incurred by the arrangement. But this is arguably true of physical augmentation of interconnector capacity, too. Thus if active congestion management arrangements are seen as a substitute for physical network expansion, there is also no reason, in principle, why

[37] It ensures more direct comparability with the discussion of CBR above, where strict revenue neutrality for all constraints for all dispatch intervals is a requirement, and with the example in the previous chapter.

[38] That is, assuming that none of the parties involved believes it has any influence over the congestion price.

[39] It could also create some rather odd feedback situations with respect to interactions between contractual obligations of various parties.

they should not be evaluated, and ultimately funded, on a comparable basis. In particular, CRA saw no need to impose a strict revenue neutrality requirement on a period by period basis.

If the requirement for strict revenue neutrality were to be further relaxed, e.g. requiring revenue adequacy over a period of time, then all CRRs could be expressed in terms of megawatts volumes—recognising that in any given dispatch interval there may be net surpluses or deficits remaining in the CRF after settling the CRR claims. In that case all the same "revenue adequacy" issues arise as in FTR markets. While fully firm hedging may never be possible, hedging can probably be made much firmer than under the status quo, and than under any regime which imposes strict revenue neutrality.

3.3.3 Issues Arising

There are several points to note here:

- First under either a CBR-style or CSP/CSC-style approach, the aggregate hedging supported by the available IRH pools can only be as firm as the *Protected RHS*, as defined by the terms that are subject to IDMA of CRRs in respect of each of the CRFs within the scope of the congestion pricing scheme. The purchase of a "share" in one of the IRH, or CRF, funds will hedge the price risk on a greater or lesser MW volume, depending on the dispatch volumes of the other terms on the *ProtectedRHS*.

- In practice, there might be a wide range of ("state-contingent") constraint forms relating to trans-regional effects, e.g. reflecting different outage conditions. The options discussed in the first section deal with this by "bundling" all individual CRFs relating to different binding constraint forms back into IRH pools (as occurs today in constructing the IRSRs). While individual auctions could be held for each possible constraint, this could increase the number of auctions substantially. This could be viewed as a strength (in terms of flexibility for market participants) or a weakness (in terms of complexity).

- In principle, by increasing the "dimensionality" of the auction process to match the full dimensionality of the underlying constraint structure, this approach provides greater flexibility for participants to tailor hedging to their requirements. In particular, a CBR style regime does away with the need to form any prior agreements with respect to the allocation of CRRs to IRH pools.

- By way of contrast, a CSP/CSC style regime reduces the dimensionality of the hedging problem faced by participants, but requires a prior agreement with respect to the allocation of CRFs between IRH pools, thus restricting the market's flexibility to choose the balance between hedging across particular interconnectors, for each trading interval.

- Cases can be readily constructed in which either the aggregate CRF value for some constraints will be negative[40], or particular market participants will need to buy negative shares (i.e. sell shares) in some positively valued CRFs in order to obtain hedging. While it is entirely feasible to auction a liability (it is generally called a "tender" rather than an "auction"), this would add a new dimension to the wholesale market trading arrangements as discussed below.

- A CBR-style regime which relies entirely on auctions, and in which some parties must effectively buy rights off a small group of sellers might also reveal some issues of localised market power. Thus, in reality, it may not be possible to completely do away with the concept of negotiated allocations, either between interconnectors, or (as discussed later) between interconnectors and market participants. Hence, in part, the motivation for discussion "allocation" and/or "negotiation" mechanisms under the CSP/CSC style approach.

- Given the number of constraints that may be involved, implementation of an "**integrated network-based auction**" (INBA) may be desirable and offers the promise of combining the simplicity of a CSP/CSC style hub-spoke structure, without imposing a requirement for prior negotiation of rules to allocate CRRs between interconnectors, for example. As discussed by Read (2007) the underlying instruments in such an auction could be CRRs, but market trading would probably be restricted to FTR-like instruments in hub/spoke format.[41]

---

40 See the "cross border load" example in Read (2007)

41 Thus it would be analogous to Flow Gate Right (FGR) auctions, which have been proposed in the literature, and should be possible, given the analogy to FTR auctions.

4 Wider extensions of congestion pricing options

4.1 Introduction

The previous two chapters emphasised options that could apply to generators and/or interconnectors on a localised basis, as a supplement to the basic framework of a regionally settled market. This chapter considers the wider application of options to include other types of market participants (e.g. loads, TNSPs, and providers of ancillary services), and/or to apply to the market as a whole rather than on a localised basis.

4.2 Widespread application to generators

If all generators in the NEM were exposed to CP in every constraint, this effectively introduces generator nodal pricing and settlements.[42] At the same time, introducing explicit CRRs for generators would provide generators with financial access to settlements at the RRP (to the extent covered by the explicit allocation of CRRs).

As far as generators are concerned, this market structure is really no different from a nodal market, with one important difference: if all load is still settled at RRP, load will presumably want to purchase energy contracts referenced to RRP. So, unlike a nodal market, in which participants may be interested in purchasing FTRs from any node to any node, there seems likely to be little interest in trading anything but node-to-hub, and hub-to-hub hedging instruments.

A CBR style approach would theoretically allow for more flexible trading arrangements. In fact, the CRRs traded under such a regime would be equivalent to what have been called financial "Flow Gate Rights" (FGRs) in nodal markets.[43] But, with loads settled at RRP, the only obvious advantage of that greater generality is the flexibility it allows to alter the balance between intra-regional and inter-regional trading, as discussed previously.

Leaving aside the question of whether load should be exposed to congestion pricing (see below), a critical issue is whether widespread implementation of congestion pricing in the form discussed here would be a better or worse way of implementing a fully nodal market for generation, if that were considered desirable. CRA argued that the theoretical equivalence between the various regimes meant that key aspects of that question could be decoupled, and considered separately. Expanding on their logic:

[42] With the caveat that the PNPs implied by the CP regime will only be equivalent to nodal prices in a nodal pricing model if the constraint coefficients for generator terms reflect only what CRA has termed NEO effects. That is, the constraint coefficients must only reflect the impact of energy injection at the nodal location and not, for example, unit inertia. This would mean that generation was being treated symmetrically with load, as it is in nodal markets.

[43] Originally, this concept was intended to be "physical", but was later generalised to allow financial contracting. It has not, so far as we are aware, been implemented though, in either form.

- First, it could well be desirable, for various technical reasons, to replace NEMDE with a Full Network Model, irrespective of any changes to market design. Such a model would be used to dispatch generation on a nodal basis, but the results could still be interpreted in terms of the theory developed here, with both load and generation primarily settled at RRP prices, and congestion pricing schemes introduced on a small or large scale as desired.

- Second, if a comprehensive nodal pricing regime is preferred, it would seem logical to use the prices from a nodal market model directly, rather than constructing PNP indirectly as discussed here.

- Third, if a Full Network Model were employed for dispatch purposes, essentially the same Full Network Model could be used to clear an FTR market, as is common in nodal markets elsewhere. As noted, participants may only wish to trade node-to-hub and hub-to-hub FTRs, which are equivalent to the bundled CRRs proposed under the CSP-CSC type approach. But there is no obvious reason why a general model integrated network based auction can not clear a market in such a restricted set of instruments, as suggested above.

- Fourth, the alternative paradigm would be to employ a nodal model, but then create a market in CBR-type instruments. As noted above, these would effectively be FGRs in a nodal market context. Such instruments have been proposed elsewhere, and such a market seems possible. But it is worth noting that the international consensus, so far, has overwhelming favoured trade in "bundled' node-to-node FTR instruments, rather than pure constraint-based FGR instruments. As we understand it, a major reason for this preference rests on the observation that, while the two approaches may seem to have similar complexity under certainty, the number of FGR type instruments which need to be traded multiplies rapidly under uncertainty.

In conclusion then, a standard implementation of a fully nodal generation market with FTRs is probably preferable to a comprehensive implementation of either CSP/CSC or CBR approaches. Importantly, it has the additional advantage of being a well established international paradigm.

Conversely, if such an implementation is not considered to be justified, it seems unlikely that comprehensive implementation of either CSP/CSC or CBR approaches would be justified either. Thus this logic leads us back to CRA's conclusion that the advantage of the CSP/CSC approach was precisely that it could be applied in a limited fashion, e.g. to contract for "interconnector support", without undue disruption to the NEM structure.

The logic for implementation of a CBR-style regime on such a limited basis seems less clear, since it seems designed more as a comprehensive alternative to nodal pricing with FTRs. In particular, it seems less suited to the kind of active congestion management which may be required to deal with some of these localised situations in which a few participants may have a significant influence on a particular constraint. But there are other situations in which it may make a useful contribution, e.g. to provide an alternative structure to the current SRA process.

4.3 Congestion pricing options involving load

The discussion in this paper assumes, unless otherwise stated, that loads are not exposed to congestion pricing. Markets do exist, which expose generators to nodal prices while load faces a regional or market-wide price. Determining the most sustainable pricing regime, or RRP level, for load is another design issue. Markets like Singapore, for example, rely on averaging nodal prices to form a hub price that is used to settle loads.

Another way to extend the application of congestion pricing options, though, would be to expose loads to congestion pricing also. A congestion pricing scheme involving all market participants would expose generation, interconnectors flows and loads to congestion prices. Involving loads would not create any more CRFs, because only one is required for each binding constraint. But it would effectively expose loads to nodal pricing, and imply the need for loads to somehow acquire CRRs to hedge their locational risk.

It is commonly said that FTRs, in nodal markets, can be made as firm as the RHS of the constraints defining the network model in the market-clearing formulation. But the constraints employed in those formulations have load as a variable, on the LHS of the constraint equations. Thus, in our terminology, the Protected RHS in those markets consists of just the TNSP network capacity terms, plus those representing any network support ancillary services. Exposing loads to congestion prices therefore increases the firmness of traded hedging instruments. It removes the implicit assignment to loads, via IDMA, of a component of hedging firmness—hence increasing what is available to trading participants in aggregate.

CRA noted the possibility of involving loads in a CSP/CSC regime, but did not consider it to be politically acceptable. As we read it, the full CBR approach advocated in Biggar (2006b), does involve loads though.

4.4 Exposure of transmission and/or ancillary services to congestion price risk

The options discussed in this section, and in earlier sections, provide a framework for describing options which are designed: (a) to change the market environment so that market participants can be exposed to the pricing implications of congestion risk; (b) provide instruments and market structures to allow them to better manage the consequences of congestion risk; and (c) actively to manage the level of congestion.

The original motivation for the work which led to CSC/CSP was focused on the last. The potential for congestion pricing options to be used to manage the levels of congestion can be illustrated further by considering their direct application to parties who are in a position to reduce, or manage congestion, particularly TNSPs and ancillary service providers.

Recall that NEMDE constraints involve:

- interconnector and generator terms, treated as variables, on the LHS; and
- TNSP (line capacity), load and ancillary service terms, treated as constants, on the RHS.

Up to this point, we have discussed "exposing" interconnector, generator, or possibly load terms, in some combination, to congestion pricing. That is, we have discussed placing these terms, alone or in combination, on what Read (2007) describes as the "exposed LHS" of managed constraints, while placing all other terms on the "protected RHS".

Generalising that discussion, though, there is no reason, conceptually, why the TNSP or ancillary service terms could not be shifted to the exposed LHS of some or all constraints. Practically, this would mean that:

- Payments would be paid between those parties and the CRF for each binding constraint in which they were exposed, just as for the generator/interconnector terms discussed previously.
- Contracts could be entered into with those parties which would typically offset the gross payment implied by CP exposure, just as for the generator/interconnector terms discussed previously.

Such contracts would probably be expressed in terms of the net MW contribution of the relevant party to the RHS of the constraint, and expressed in fixed MW terms, since the intent is not to provide hedging to these parties, but to contract for specific congestion-reducing behaviour.

In situations where network support ancillary services were contracted by or through a TNSP, these incentives could be passed through by applying them first to the TNSP, then leaving the TNSP to contract with the ancillary service provider, using a contractual form which could, but might not narrow the incentives the TNSP faces from the CPS.

The net effect would be that, if the TNSP or ancillary service provider matched their contracted volume of line capacity, or ancillary service provision, they would have no net exposure to CP. If they exceeded the contract quantity, they would be rewarded by CP times the extra provided. If they fell below the contract quantity, they would be penalised by CP times the shortfall.

In practice, such exposure could be full or partial. For example, a partial exposure level could be set within the framework of the regulatory Service Target Incentive Scheme (STIS), applying to the TNSP. Thus a TNSP with, say, 10% exposure to CP on a particular constraint, would be in a moderated version of the situation described above. If they matched their contracted volume of line capacity, they would have no net exposure to CP. If they exceeded the contract quantity, they would be rewarded by CP times the 10% of the extra provided. If they fell below the contract quantity, they would be penalised by CP times 10% of the shortfall.

The residual "exposure" (90% of the total in this example) would still be reflected in variability in the hedging provided to trading participants, as is the case for 100% of the capacity variation now. The direct financial impact may be only a marginal improvement to hedging firmness. But the indirect impact would be to motivate TNSPs to make physical capacity firmer, which would enhance real market value, and also make hedging firmer, perhaps significantly so.

Conversely, we note that it is often argued that TNSPs can be exposed to the market value implications of congestion risk, because network capacity varies for a wide variety of reasons outside TNSP control. In NEMDE, the capacity apparently defined by a constraint RHS even varies as a function of load, and possibly with factors such as the number if units generating. But the mechanism described here could be applied only to specific components which determine the RHS, such as the availability of specific transmission elements, for which the TNSP can reasonably take responsibility. The implications of increasing TNSP risk exposure in this way would have to be accounted for in the overall TNSP regulatory framework, though.

5 Policy options within regional pricing market designs

This section reviews different policy approaches that have been advocated or adopted in regional pricing market designs (such as the NEM) to deal with the "mis-pricing" of generation without introducing explicit congestion pricing of the kind discussed in the previous three chapters. We focus, in particular, on examples from the NEM.

The framework used above to describe various congestion pricing options is used here to describe other options for managing congestion that are already allowed for in the NEM's Rules, but which do not involve explicit congestion pricing:

1. constrained-on/off payments for the provision of network support and control services, in particular gatekeeper generation;

2. physical interventions, such as giving preference to local or remote generation where both affect a binding constraint or giving preference to one interconnector over another when both impact on a binding constraint; and

3. changes to the regional pricing structure, through either shifting the location of the RRN within a region or changing regional boundaries.

In addition, the framework can also be used to describe a more radical re-definition of the RRP than that arising from shifting regional boundaries or the location of the RRN within a region.

5.1 Constrained-on/off Payments

Many zonal markets have explicit "constrained-on" and "constrained-off" payments. These are a form of "side payment" or "financial incentive" that is designed to better reflect the economic externality benefits (e.g. lower congestion cost, higher reliability or security) arising from a generator (or load) altering its production/consumption level to relieve a binding constraint.

A participant or location specific value for the constrained-on and constrained-off payments can be calculated in a number of ways, including:

(a) based on the bids and offers of the affected parties;

(b) based on a negotiated/arbitrated value; or

(c) based on the Congestion Price at that location, which reflects the economic value of relieving the constraint.

The CSP/CSC and CBR approaches discussed in this report represent variations on option (c). As they stand, though, the NEM's Rules explicitly exclude constrained-off and constrained-on payments as part of energy settlements (Clause 3.9.7(b)). Despite this, there are a range of other mechanisms available in the Rules for making payments to constrained-on and constrained-off generators under specific circumstances. The agreed value of such payments is generally based on negotiations (including tenders for network support) or failing that, arbitration.

There is allowance for constrained-on compensation to be paid when NEMMCO formally directs a market participant to increase its generation level (clause 3.12.11). NEMMCO calculates the value of the constrained-on compensation using a cost-based formula (clause 3.12.11), with any dispute about the compensation amount being referred to an independent expert arbitrator (clause 3.12.11A).

The Rules provide for contracts with market participants for the provision of Network Support and Control Services (NSCS). Such contracts, which can be with a TNSP or NEMMCO, can oblige generators to be constrained-on or constrained-off in some circumstances. The Rules also provide for contractual agreements between combinations of generators, TNSPs and/or distribution network service providers (DNSPs) in the context of negotiated access agreements.

In 2002, there was a proposal to extend the use of "compensation" payments to manage a specific form of network congestion that, though prevalent in the NEM, was not being addressed using the Rules' existing compensation mechanisms. The specific form of congestion was a binding constraint involving both generation in a region and flow on a single interconnector—one of the forms of a "trans-regional" constraint. This type of binding constraint requires NEMDE to trade off generation output and flow on a single interconnector in order to manage the level of flow through a congested part of the network.

This trade-off can result in "gatekeeper" effects, whereby changes in the output level of the generator can increase (i.e. "support") flow capacity on the interconnector or decrease (i.e. "block") flow capacity on the interconnector. These gatekeeper effects can affect the efficiency of dispatch and the firmness of inter-regional settlement residues used for inter-regional trading.

Specifically:

1. A **positive gatekeeper** can *support* interconnector flows. An increase in the output of a positive gatekeeper generator facilitates an increase in the interconnector flow level, which enables a reduction in the total costs of dispatch across the NEM. This reduction in dispatch costs occurs because the higher interconnector flow displaces generation in the importing region that has a higher marginal resource cost than generation at the exporting end of the interconnector.

2. A **negative gatekeeper** can *block* interconnector flows. An increase in the output of a negative gatekeeper generator facilitates a decrease in the interconnector flow level, which causes an increase in the total costs of dispatch across the NEM. This increase in dispatch costs occurs because the displaced interconnector flow is replaced by generation in the importing region that has a higher marginal resource cost than generation at the exporting end of the interconnector.

Examples of positive gatekeepers in the NEM include:

- Southern Hydro generation (in Victoria) relieves a transformer constraint at Dederang, thereby enabling higher levels of flow along the Snowy-Victoria directional interconnector. Under certain network conditions, a small increase in generation by Southern Hydro's Kiewa, Eildon or Dartmouth plant can enable a significantly larger increase in interconnector flow from NSW/Snowy, which overall can be considerably cheaper than sourcing additional supply from within Victoria.[44]

- Upper and Lower Tumut generation supporting flows on the Snowy-NSW interconnector into NSW.[45] Without the output of Lower and Upper Tumut generation, the maximum volume of interconnector flow on the Snowy-NSW interconnector is around 1,350MW. The maximum flow from Snowy into NSW is over 3,000MW when Lower and Upper Tumut generate at sufficiently high levels.

An example of a negative gatekeeper is:

- Generation output at Liddell and Bayswater power stations displacing Queensland-NSW interconnector flows, when there is a binding constraint within NSW between Liddell-Newcastle and/or Liddell-Tomago.

The gatekeeper analysis recognised that under the existing regional pricing and settlement Rules, positive and negative gatekeeper generators may not have an incentive to generate at levels that are consistent with supporting greater interconnector flows—and thereby minimising the total costs of dispatch. This is because, under the NEM's regional settlement Rules, a positive gatekeeper could find itself being constrained on and a negative gatekeeper constrained off. In these circumstances, the gatekeeper generators would have incentives to change their offers in ways that align their dispatch volumes with the volumes they are prepared to sell at the prevailing RRP used in settlements.

Two approaches were investigated which involved some form of "cost-based' compensation to generators that were constrained on/off in order to provide interconnector support:

(a) a "pay-as-bid" methodology, in which the cost of constraining generation on/off to support an interconnector was determined from its market offers; or

(b) a methodology based on external estimates of underlying costs.

Approach (b) might be preferred if approach (a) seemed likely to provide incentives for grossly distorted offers.

But both of these cost-based compensation regimes were considered to be inferior to a "congestion priced" approach, which this paper focuses on. A number of reasons for preferring a congestion pricing approach over a cost-based compensation regime include:

---

[44] These conditions relate to pre- and post-contingency losses of any of the three 330/220kV Dederang transformers.

[45] See CRA(2004c), pp. 34-47.

- less complexity, given that congestion prices are already calculated by NEMDE;
- consistency in treatment under congestion pricing approaches – with payment always based on the "value" of the service being provided;
- potential for compensation rules to be "gamed" (i.e. exploited);
- the need to allocate responsibility and rules for setting compensation amount; and
- congestion pricing can be generalised to a wide variety of situations, such as where support on one interconnector is provided by another interconnector, not a generator.

However, while congestion pricing options were preferred to compensation-based models, gate-keeper type models are still a relevant policy option.

5.2 Physical interventions

Another class of policy option involves physical interventions in the dispatch process to provide for (or against) particular outcomes in particular circumstances. An example of such an intervention in the NEM is NEMMCO's procedure of "clamping" interconnector flows as a means of managing the incidence and magnitude of electrical flows between regions in a counter-price direction, and the consequent accumulation of negative inter-regional settlement residues. There are a number of reasons why counter-price flows might emerge, only some of which relate to "mis-pricing" of generators and the incentives that can create to submit bids into the dispatch that do not reflect costs. Counter-price flows can, in other circumstances, be entirely consistent with an efficient dispatch. The Snowy Region example used to illustrate congestion pricing options in the previous chapters demonstrates this.

Four physical options have been trialled in the NEM or discussed in the CRA work and through the process of consultation under CMR:

(a) giving preference to an interconnector over local generation, when both are affecting a binding trans-regional constraint that is resulting in the accumulation of negative inter-regional settlement residues;

(b) giving preference to local generation over an interconnector, when both are affecting a binding trans-regional constraint that is resulting in the accumulation of negative inter-regional settlement residues;

(c) giving preference to one interconnector over another interconnector, when both are affecting a binding trans-regional constraint that is resulting in the accumulation of negative inter-regional settlement residues; and

(d) restricting interconnector flows to stop negative settlement residues accruing (i.e. "clamping").

These approaches force the market to produce certain dispatch outcomes, regardless of participant bids and offers, and do not address the underlying behavioural incentives created by the NEM's basic regional pricing and settlements structure. This may be considered undesirable, in theory, but it is an empirical question whether, in practice, the transaction costs involved in implementing such a "fix" may be low enough to enable it to deliver greater net benefits compared to other approaches, which may be more theoretically elegant.

Physical options are designed to ration constrained physical capacity on a basis other than the value of bids and offers. We can illustrate the general framework by considering the example of two parties (interconnectors or generators, it does not particularly matter) involved in the same constraint.

We can envisage the available capacity being notionally partitioned as illustrated in Figure 1. Here the partitioning of constraint capacity is represented by point X on the constraint line, with the co-ordinates being projected to the axes to define SHARE1 and SHARE2. These can be thought of as defining either minimum or target capacity levels to be made available to each generator/interconnector involved, expressed in terms of participant MW generation/flow, rather than constraint capacity.

While it is unclear what various parties may intend when they speak of "giving priority" to a particular generator/interconnector, it is useful to consider what might, or should be done when the unconstrained dispatch solution lies at the various points shown in Figure 1:

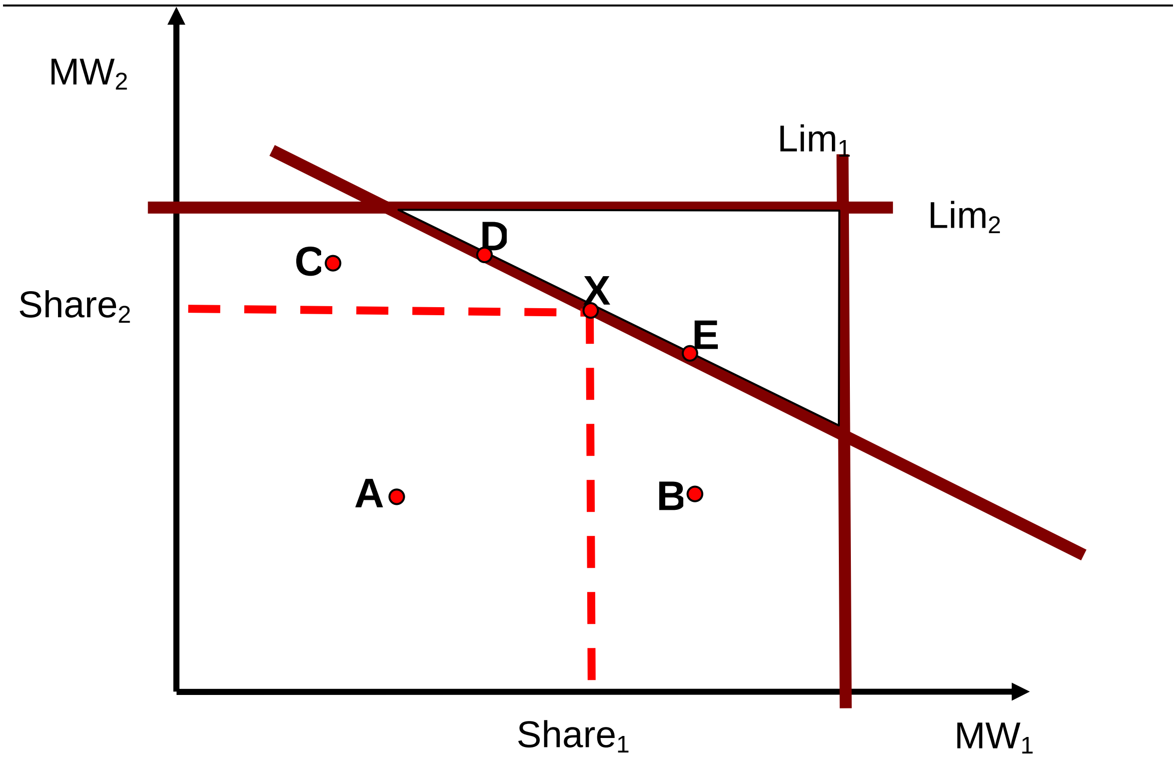
- At point A, no action seems necessary. Both generator/interconnectors are using less than the agreed capacity, but presumably this is because the market does not want to use the full capacity. Thus, there is no point in artificially forcing the dispatch up to the agreed capacity levels. There is also no issue of rental allocation because the constraint is not binding and CP = 0.

- At points B and C, no action seems necessary, either. Although one generator/interconnector is using less than the agreed capacity, and the other is using more, the market has freely chosen this dispatch, and the "excess" dispatch of one generator/interconnector is not at the expense of the other, which is still free to increase dispatch at will. Again, there is no point in artificially forcing the dispatch either up or down to the agreed capacity levels, and there is no issue of rental allocation because CP = 0.

- At points D and E, though, one or the other generator/interconnector is using more than its agreed capacity share and this is at the expense of the other generator/interconnector. Thus, presumably, a regime based on giving physical "priority" to specific generator/interconnectors would attempt to force the dispatch solution back from such points to, or at least towards, the agreed point, X.[46]

[46] A more moderate regime could allow a "dead band" of acceptable balances, and only take action when the balance strays outside of those bounds. Or it might only require action under specific circumstances, such as a negative IRSR. This does not change the comments made here about the kind of mechanism that may be employed to give effect to such a regime.

Figure 1 Physical Prioritisation



Practically, it is possible to force the solution to X, if that point is feasible, given the network configuration and generation capacity offered, by either:

- constraining the generator/interconnector whose dispatch is above its capacity assignment down to its SHARE; or
- constraining the generator/interconnector whose dispatch is below its assigned capacity up to its SHARE.

Either way, the additional constraint to impose this requirement would be a simple bound, and it would itself be binding.

Thus the intervention constraint would have a positive shadow price, and imposing the constraint on an interconnector would increase the inter-regional price difference if the interconnector flow was forced down, and decrease the inter-regional price difference if the interconnector flow was forced up.[47] This situation arises when a NEMMCO uses a "clamping constraint" to restrict flows on the VIC-Snowy interconnector, for example.

47 A forcing constraint imposed on generation, if properly oriented, should not directly impact on the RRP. There may be an indirect effect though, as the overall regional supply/demand balance will be affected.

The choice between these alternatives is important because the pricing implications are very different. Intuitively, it seems most reasonable to think in terms of forcing one generator/interconnector dispatch down in order to free capacity for the other. But, mathematically, the choice seems arbitrary, and it is not obvious how such a limit should be oriented, for example.[48]

Forcing the market in this way obviously imposes economic costs, as measured by the market objective function. For this reason, options of this kind have been judged to be undesirable in past studies, with "fully optimised" (i.e. "Option 4" or "Option 5") constraint forms preferred.[49] But physical intervention of some kind has still been advanced as an option, particularly where it is believed that the market objective function, which is defined by participant offers, does not properly reflect the underlying economics.

If priority is treated as being absolute, then the balance target should be treated as a hard constraint, the shadow price on which could become very high, if the market really wants to achieve quite a different dispatch balance on the basis of bids and offers.[50] If the prioritisation rule is not to be a "hard" constraint, then a design issue is to establish a method of defining how seriously it should be taken, and how high the CP should be allowed to rise before the dispatch is deemed to be close enough to being in the desired balance. And what economic cost should the market bear in order to achieve this goal?

Another important design question is whether, and how, to compensate parties affected adversely by any such physical interventions. Clearly there are a range of possible options, including no compensation (as is the case with NEMMCO's currently practice of "clamping"). Congestion pricing options, in contrast, resolve this issue through the use of Congestion Prices and contracts. Active constraint management options do raise some similar issues. But rather than partitioning the physical capacity, such an option may partition the corresponding CRF. And rather than forcing the market dispatch to match that partitioning, the contracts merely give participants relatively gentle "second order incentives" to move dispatch in that direction.

5.3 Changes to regional pricing structures

Two types of structural change are allowed for in the NEM Rules, and could be used to facilitate congestion management, by aligning the market structure more closely with the physical constraint structure. Basically, they involve changes to the RRP faced by some, or all participants in a region:

1. ***Changes to regional boundaries***— this directly changes the settlement price faced by both generators and loads that are assigned to new regions. The region

48 These issues would require further consideration, if this kind of option were to be pursued further.

49 See discussion in CRA(2002), "Network Constraint Formulation: Impact on Market Efficiency", released by NEMMCO, January 2002 and Ministerial Council on Energy (2005), "Statement on NEM Electricity Transmission", May 2005, p.5.

50 In the limit, it could reach VoLL if it is not possible to meet the target with available capacity.

change also alters the implicit dispatch matching CRRs granted to participants as part of the settlements process, under the IDMA process. The new CRRs relate to the new RRN, rather than the old RRN, and the volume of new CRRs that are implicitly granted via the dispatch process depends on the pattern and incidence of physical constraints between a connection point and the new RRN, taking into account bids and offers.

2. ***Shifting the location of the RRN within a region, leaving the boundaries of the region unchanged***. This requires a transformation of all the constraints involving participants within the region, and/or interconnectors to/from the region, so that they are correctly oriented towards the new RRN. The closest NEM example is switching the Snowy RRN from Murray to Dederang, and re-orienting all the relevant constraints towards the Dederang node, changes the formulation of the constraints, and by so doing changes the nature of the implicit CRRs granted to generators in the Snowy region and to IRSR unit holders on the VIC-Snowy and Snowy-NSW interconnectors.

These structural options alter incentives, but are not really "physical options" because dispatch is still in accordance with offers. Structural options are really examples of CRR re-arrangement because they change the financial access of market participants to settlement at a given RRP.

The possibility of using CRRs to re-allocate congestion rents between market participants and inter-regional hedging pools means that, if desired, one could maintain any agreed measure of access to the original RRN, simply by identifying and explicitly allocating the constituent CRR streams that were formerly allocated, implicitly, to those participants. Thus, the transition to a new region boundary configuration could be actively managed through the implementation of a congestion pricing prior to the region changes taking full effect. This was one motivation for the "CSC" part of CRA's CSP/CSC proposal, developed in the context of a review of regional boundary issues.

If region boundaries were maintained, but an alternative RRN adopted, then the IDMA-based allocation of CRRs would change. An appropriately-designed congestion pricing regime could give effect to this type of change. It might be thought that, since the "re-oriented" constraints are now different, the price structure would also be fundamentally different. But this is not the case. What really changes when the RRN is shifted is that all participants in that region effectively receive a new CRR allocation, giving access to the new RRN. Equivalently, they all now face a new RRP, representing the marginal cost of supply to that node.

It may also be worth noting that the net effect of de-allocating a CRR to/from the old RRN, and then allocating one to/from the new RRN is equivalent to allocating a transmission right from the old RRN to the new RRN, at least with respect to this particular constraint. And the same would apply to any allocated CRR. Thus any transition of this nature can be characterised in such terms, and could actually be implemented that way, too.

More importantly from a policy perspective, the commercial position of any participant can, if desired, be protected for some period after any such change by allocating rights of this nature, thus effectively allowing CRRs to the old RRN to remain effective for the agreed period.

5.4 Changing the definition of the RRP

The framework in the paper can also be used to describe a more radical re-definition of the RRP than that arising from either a shift in regional boundaries or the location of the RRN.

Under the existing NEM Rules, the RRP is defined as a nodal price near the largest load (or generation) centre in a region (i.e. the RRN's price). The Rules do not contemplate other ways to define the RRP, however.

The RRP could be re-defined to be a synthetic trading hub price, equal to the weighted sum of all nodal prices in a region. A shift to a hub price like this opens up a range congestion management options, including having generators being settled on the basis of their nodal prices but loads continuing to be settled at the RRP. In such an arrangement, the hub price, like the existing RRN, would be used for trading purposes, but some form of explicit CRR would be used to manage risks arising from differences in nodal prices and the hub RRP. See Read (2007) for further discussion.

Appendix A – Dealing with negatively-valued CRRs

The class of options discussed above uses interconnector shares (CRRs) in rental funds (CRFs) to construct IRH pools. Theoretically, the aggregate CRF will only have negative value if the "Protected RHS" of the constraint is negative; that is if the constraint represents an obligation to met a requirement (e.g. to maintain supply to some load), rather than a limitation on a network resource.[51] In the later case, a positively valued nett CRF pool will only be available for hedging purposes if some market participant can be induced to accept a CRR which has a negative value. More generally, there will be many cases where the supply of positively valued CRRs to participants who require them to hedge reasonable trading positions will be heavily dependent on other participants being ready to accept CR allocations of negative value or, equivalently, to sell positively valued CRRs into the pool.

In this context, negatively valued CRRs will be relevant whenever flow on one interconnector "supports" the flow on another, in the least cost economic dispatch based on bids and offers. In fact, this situation occurs with respect to the Snowy example discussed earlier, where flow on the VIC-SNY interconnector increases the effective transfer capacity between SNY and NSW, thus supporting flow on the SNY-NSW interconnector.

Under the status quo the allocation of negatively-valued CRRs as a result of IDMA reveals itself through a net reduction in the relevant IRSRs. In the example above, the allocation of negatively-valued CRRs to VSN when the trans-regional constraint binds (and when the *ProtectedRHS* is positive) means that the total value of the VSN IRSR is lower than it would otherwise be. The VSN IRSR may only become negative in a few cases, but the essentially the same phenomenon is actually occurring whenever a constraint of this form (i.e. with a negative weight for this interconnector term) binds. Thus, once the constituent parts of the IRSR were deconstructed to the level of the individual CRFs, transactions involving systematically negative CRRs would be a feature of the regime.

A.1 Auctioning negatively valued CRRs

As noted above, it is feasible to auction off CRRs that are expected to have a negative value. It does, however, means that prospective buyers will submit negative bids, i.e. they will require payment to take on the (expected) liability.

A.2 Allocating negatively valued CRRs

We can consider the issue of allocating negatively valued CRRs by extending the Snowy example used above. In the example, a positive CRR allocation was assumed for the SNY-VIC interconnector, and this produced a positive inter-regional hedging

[51] Here we assume that all constraints are presented in a standardised "less than" form, so that obligations are expressed by a requirement for some weighted sum of variables to be less than a negative value. But this assumption only affects the form, not the substance, of the argument here.

pool after de-constructing and re-constructing the IRSR pool as explained above. The Congestion Pricing Trial in the Snowy Region, and in particular the refinement to the original Trial, introduced as the "Southern Generators' Rule", [52],[53] is a practical example of an allocation rule. In that case, the ex ante agreed allocation to the VIC-SNY IRH pool is set to zero, so that it is left empty after being de-constructed and re-constructed.

In principle, we could construct more extreme cases where the CRR allocation was negative.

This may seem inappropriate, but it would not actually be unreasonable, or meaningless, though, given the supporting role that flows on this interconnector play with respect to overall VIC-NSW transfer capacity.

But responsibility for playing such a support role can not ultimately rest with an interconnector. It must ultimately rest with generation, which must be induced to generate so as to create a cross-border flow, which provides the required support. The obligation here is not just to generate, but to "trade" in the counter-price direction. More exactly, since these are financial contracts, generators must commit to making the specified contributions to the CRF pool whenever the constraint binds, whether or not they are actually generating. This, in turn, will incentivise them to generate closer to the required level in order to match their contract positions.

Thus, if a negative CRR allocation were made to the VIC-SNY IRH pool, the implication is that, in aggregate, the SRA process would be looking for participants willing to take on the obligation of trading in a counter-price direction, from VIC to SNY (and then perhaps on to NSW) by accepting hedges with negative values. Participants would obviously expect payment to take on those obligations. They would do this by offering to sell positively valued CRFs into the pool.[54] It does not mean that hedging is not available, at positive cost, in the opposite direction, just that the aggregate of all sales and purchases must match this total.

Of course, the status quo implicitly assigns negatively valued CRRs in this situation too, or equivalently makes a negative assignment of positively valued CRRs. The difference here would be that the aggregate volume of such CRRs allocated would be firm, and would thus provide this volume of firm support to the SNY-NSW interconnector.

The other situation in which such negatively valued allocations could arise would be due to a negative Protected RHS, perhaps reflecting a situation in which the interconnector(s) involved in a constraint had an obligation to support a locational load. An example would be a situation where loads in Northern NSW would have to

---

52 AEMC 2006, *Management of negative settlement residues in the Snowy region,* Final Rule Determination, 14 September 2006, Sydney. (Available at http://www.aemc.gov.au/electricity.php?r=20051214.200416).

53 *National Electricity Amendment (Management of negative settlement residues in the Snowy Region) Rule 2006 No. 14*. (Available at http://www.aemc.gov.au/electricity.php?r=20051214.200416).

54 Parties doing this would presumably be incentivised to match these inter-regional hedges to energy contracts, buying or selling in each region as appropriate.

be supplied by QLD generation via QNI in the event that flows from Central NSW were limited below the Northern NSW demand level. This could result in negative residues accruing on flows from QLD to NSW when the QLD RRP was above the NSW RRP, but flows on QNI were counter-price in order to supply Northern NSW loads. This situation is discussed in more detail in the context of congestion pricing options involving generators.[55]

[55] See Section 6.4 of Read (2007).