AUSTRALIAN ENERGY REGULATOR

GPO Box 520
Melbourne VIC 3001

Telephone: (03) 9290 1444
Facsimile: (03) 9663 3699

www.aer.gov.au

13 April 2006

Mr Steven Graham
Chief Executive Officer
Australian Energy Market Commission
PO Box H166
AUSTRALIA SQUARE NSW 1215

Dear Steven,

Re: AEMC Congestion Management Review – Issues Paper

The AER welcomes the opportunity to comment on the issues paper recently released by the AEMC for its congestion management review.

The AER supports this Review and appreciates that the way in which congestion is managed has significant implications for the operation and success of the NEM. While the congestion management issues of most interest to the AER are related to our market monitoring and transmission regulatory functions, we are keenly anticipating the outcomes of this review, given the broader long-term implications for the NEM.

Clarifying the Problems

The AER understands that the issues associated with congestion in the NEM are complex and interdependent. Congestion problems need to be clearly identified and specified before the full array of possible solutions may be analysed. The AER has invited Dr. Darryl Biggar to provide an analysis of congestion management issues in the NEM. Dr. Biggar has an excellent understanding of the issues under consideration, having analysed these issues extensively during the ACCC's assessment of the CSP/CSC trial in the Snowy region. In addition, Dr. Biggar has looked into the current rule change proposal regarding the management of negative settlement residues in the Snowy region.

Dr. Biggar's analysis of congestion management issues is attached. In his paper, Dr. Biggar identifies the following two basic problems relating to congestion in the NEM:

1) When generation in a single region must be dispatched for a quantity of output which is larger or smaller than the amount that they have declared they are willing to produce at

the regional reference price, that generation will be 'constrained-on' or 'constrained-off' (respectively). Under the current arrangements in the NEM, without some additional mechanism, constrained-on or constrained-off generators have no incentive to submit an offer curve which accurately reflects their true marginal cost. This can result in higher cost generators being dispatched when lower cost generation is still available.

2) The current practice of dividing the total trading surplus into individual streams called inter-regional settlement residues does not guarantee that the individual streams of residues will be positive. In particular, as a result of loop flows between regions, negative settlement residues will accrue in some circumstances, even with fully efficient dispatch. In these cases, the actions by NEMMCO to limit the negative settlement residues will reduce the efficiency of dispatch.

Dr. Biggar's paper suggests six guiding policy principles for future management of congestion in the NEM:

- If there are two underlying problems, these should be tackled with two different policy solutions. For example, policies to manage congestion may need to be accompanied by separate policies to control the potential flow-on market power issues;

- Any future congestion management regime should involve a move towards finer geographic differentiation of prices for generators;

- To promote trade across different locations, settlement residues should be defined in such a way as to be as firm as the physical capability of the underlying network.[1] Dr. Biggar's suggestion is for NEMMCO to auction constraint-based residues.

- There are many reasons to suggest that it is preferable for consumers to also face finer geographic differentiation in prices (eg. allowing a demand-side response, curtailing generator market power). Those jurisdictions wishing to provide consumers with more accurate pricing signals at their location should not be constrained from progressing towards such an arrangement. However, given the relatively mild efficiency losses associated with regional pricing for consumers (ie. due to inelastic demand for smaller consumers), jurisdictions should maintain the option to decide whether or not they will move toward finer geographic differentiation of prices for consumers;

- While the welfare implications of a merger of regions are ambiguous, the division of an existing region can only improve welfare. The rules should provide that new region boundaries can be created, but existing region boundaries should not be removed; and

- If there is to be any free allocation or 'grand-fathering' of rights to generators under a congestion management regime (eg. CSP/CSC), no such rights should be allocated to new generators which are not already in existence in the NEM.

The AER endorses Dr. Biggar's analysis and proposed principles for guiding current and future congestion management policy in the NEM. However, it is noted that a broader set of issues exists around the allocation of rights which may require further consideration. Generally, the AER supports an incremental approach to congestion management, whereby policy options are outlined and prioritized clearly at the outset.

[1] The AER, through our continuing work on a service standard incentive scheme for TNSPs, understands that there are technical complexities involved in determining the appropriate level of capacity upon which to allocate rights to market participants

Quantifying Congestion Costs

The AER has been working with TNSPs and NEMMCO to develop measures of the impact of congestion in the NEM and considers that this work can assist the AEMC in determining the nature and materiality of constraints. The data pertaining to the Total Cost of Constraints (TCC)[2] and the Marginal Cost of Constraints (MCC)[3] is useful, particularly in quantifying the loss in dispatch efficiency that occurs when constraints bind.

One of the core areas of interest for the AER relates to transmission service standards. The AER has an established performance incentive scheme based on the AER's Service Standards Guidelines, which form part of its *Compendium of Electricity Transmission Regulatory Guidelines*.[4] Inter-regional and intra-regional constraints are included among five core performance measures upon which the incentive regime is based.

The AER considers that in time financial incentives based on the market impacts of constraints may be developed, and is currently working to develop a market-based service standards incentive scheme for these measures. The first step in creating such a scheme is the development of a Market Impact Transparency Report (MITR) which aims to:[5]

- identify the market impact of constraints;
- identify (to the extent possible) the root causes that affect the severity of transmission constraints; and
- publish information on the nature and market impact of transmission constraints, particularly the TCC and the MCC.

As mentioned above, the AER has already undertaken extensive work with the industry to develop measures for market impacts. Furthermore, the TCC and MCC methodologies used in the MITR have been independently audited, and the AER is satisfied that they are theoretically and practically sound.

It is anticipated that the MITR will be released in May 2006. When it becomes available, the AER would be happy to provide the full MITR data to the AEMC for the purposes of the congestion management review.

CSP/CSC Trial

The AER supports the framework outlined in the issues paper for assessing the success of the Snowy CSP/CSC trial, and encourages the AEMC to make this assessment a priority of the review. The AER is concerned that in the context of a 'staged approach' to congestion management (where region boundary change is a last resort option), few viable options would remain available should a CSP/CSC trial be implemented and found unsuccessful.

An objective, holistic assessment of the costs and benefits of the scheme is critical if congestion issues are to be meaningfully addressed in the NEM. The AER supports the AEMC as the appropriate body to undertake this assessment.

---

[2] The TCC is a measure of the reduction in the cost of dispatch as calculated by the NEMDE if all constraints in the network are relaxed.

[3] The MCC is the amount by which the total dispatch cost of energy would be decreased if a binding constraint was relaxed by a small amount.

[4] Issued in August 2005.

[5] AER Decision: *Market Impact Transparency Report* (p.2)

Holistic Approach

The AER notes that the AEMC is concurrently progressing a number of Rule change proposals which relate to congestion management in the NEM. As stated in previous submissions on these individual proposals, the AER encourages the AEMC to consider the extent to which some (or all) of these processes can be combined. Such an approach would promote a more comprehensive consideration of issues with broad market impacts, rather than the alternative of considering each of these processes separately. The AER believes that the congestion management review, which attempts to address the issues on a holistic basis, be given priority over the other concurrent proposals, which essentially deal with specific issues.

The AER believes that our relative expertise in these areas, as well as the MITR data on the cost of constraints may be of value to the AEMC in conducting this review. We look forward to discussing the ways in which the AER may best assist.

Yours sincerely,

Michelle Groves
Chief Executive Officer

CONGESTION MANAGEMENT ISSUES:

A RESPONSE TO THE AEMC

Darryl Biggar

12 April 2006

1. This paper is a response to the AEMC's request for submissions on the Issues Paper, released as part of their review of congestion management in the NEM. This paper seeks to clarify the nature of the underlying problem and to suggest several key principles which may guide the formulation of policy in this area going forward.

2. The key points of this paper are as follows:

- The development of good public policy in any field requires a clear and precise understanding of the underlying problem. Discussions on congestion management have tended to suffer from a lack of precision as to the nature of the underlying problem. In this paper, I highlight two basic problems arising from the current arrangements for handling congestion in the NEM. These two problems relate to the NEM's current arrangements for handling generators which are "constrained on" or "constrained off" and the current arrangements for handling settlement residues. I argue that all of the concerns which are raised by the AEMC flow from these basic problems.

- The first basic problem arises from the fact that, under the current NEM arrangements, certain generators are on occasions dispatched for a quantity of output which is larger or smaller than the amount they have declared they are willing to produce at the price they are paid. These generators have no incentive to submit an offer curve which reflects their true marginal cost. In fact, these generators may offer their output at the upper ($10,000) or lower ($-1000) limits on the allowed bids in the NEM. This distortion in the bidding behaviour of these generators will usually reduce the short-term efficiency of dispatch (higher-cost generation will be dispatched when additional lower-cost generation capacity is still available while meeting the transmission constraints). In addition, this distortion in bidding may lead to counter-price flows between regions which, under the current market arrangements, may require further inefficient market interventions. This distortion in bidding also alters the investment incentives on generators – generators will have too much incentive to invest in locations which aggravate transmission constraints and too little incentive to invest in locations which alleviate transmission constraints. This problem may arise from either intra-regional or inter-regional constraints.

- The second basic problem relates to the handling of the "trading surplus" that accrues to the system operator when there is a different price paid for electricity at different locations on the network. Access to this trading surplus is essential for market participants to hedge trading across different pricing regions. Ideally the trading surplus would be divided up into streams which allow for the creation of "firm" [1]

[1] Here, as throughout this paper, I am using "firm" in the sense of "as firm as the underlying physical network limits". This may be less than perfectly firm due to occasional network outages which reduce the capacity of the underlying physical network.

hedges. At present the market rules define a set of streams of payments known as "inter-regional settlement residues". However, it is easy to demonstrate that these inter-regional settlement residues are not "firm" (and may be negative) even in a fully efficient dispatch in a network with loop flow. Furthermore, these settlement residues are not firm (and may be negative) even in a fully efficient dispatch in a network with intra-regional constraints, with or without loop flow. This lack of firmness reduces the capacity of market participants to engage in inter-regional trading, allowing inter-regional forward-price differentials to persist. The first basic problem above makes this problem worse by increasing the likelihood of negative settlement residues. When negative settlement residues arise, NEMMCO is forced to intervene in the market, lowering the efficiency of dispatch and further reducing the usefulness of the settlement residues as a hedging instrument.

- In principle, it should be possible for the AER to quantify the magnitude of the short-term harm to dispatch efficiency resulting from these basic problems. This would involve answering the question: How much lower would be the total dispatch cost if constrained generators bid their marginal cost and if NEMMCO did not intervene to prevent negative settlement residues? The AER has been developing the tools to answer a very similar question for the purposes of calculating the total cost of transmission constraints. In principle these tools could be easily adapted to quantify the historic costs of the current arrangements for handling congestion in the NEM. This methodology, however, is not forward-looking and would not allow an assessment of the longer-term harm that results from a lack of firmness of settlement residues or the harm that results from inefficient generation or transmission investment decisions.

- All the solutions to the first basic problem identified above involve restoring the consistency between the price a generator is paid and the amount for which it is dispatched. In other words, all the solutions to the first basic problem above involve setting different prices for certain generators at different geographic locations within the same region. If we are to improve the current arrangements for handling congestion in the NEM, some move in the direction of finer geographic differentiation for (at least) generators is inevitable. The solutions considered by the AEMC have in common that they all increase the geographic differentiation of the prices paid to generators; they differ in other dimensions – such as the extent to which there is finer geographic differentiation in the prices to consumers, how the differentiated prices to generators are determined, and the handling of the trading surplus that results from introducing new prices for generators.

- The CSP/CSC proposal is one of the options being considered by the AEMC. Some CRA documents leave the impression that the CSP/CSC approach might be applied only to constraints which satisfy certain criteria (e.g., which are above some "threshold"). However, as CRA acknowledge, if certain constraints are ignored, the price paid to constrained generators (even after adjustment through the CSP) will still not be consistent with the quantity for which they are dispatched – so the incentive for distorted bidding will remain. The incentive for distorted bidding is only eliminated if the CSP/CSC approach is implemented for *all* binding constraints, in which case the price paid to each generator is the efficient locational price for that generator. The CSP/CSC proposal does make explicit the need (which arises under any solution to the basic problem) to determine each generator's "entitlement", "allocation" or "rights" to the trading surplus that results.

- Rather than comment in detail on the merits of various solutions, it seems more appropriate at this stage to seek agreement on certain key principles which could guide the design and assessment of options for handling congestion management in the next stage. In this paper I suggest the following principles which the AEMC could adopt to guide its decision-making going forward:

(a) The principle that two problems usually require two solutions. Under present arrangements one market distortion may be masking or offsetting the impact of another. In this case, fear of worsening the latter distortion may lead to resistance to correcting the former distortion – a case of two wrongs making a right. One policy instrument should not be made to serve two ends. If there are two underlying problems they should be tackled separately with two different policy solutions. For example, the current arrangements for managing congestion could be masking episodes of generator market power. In this case, improvements in the policies for handling congestion may need to be accompanied by explicit policies for controlling generator market power.

(b) The principle that as far as possible, the price paid to a generator should be such that the generator is willing to be dispatched for the quantity at which it is actually dispatched (in other words, the price paid and the quantity dispatched should be a price-quantity combination on the generator's offer curve). In the absence of this principle, constrained generators have no incentive to submit a bid which reflects their true marginal cost. Correcting this problem lies at the heart of solving the problem of inefficient bidding by constrained generation.

(c) The principle that the settlement residues should be defined in such a way as to facilitate hedging of price risk across different pricing regions. In particular, the settlement residues should be as "firm" as the capability of the underlying physical network. This implies, as a corollary, that these streams of residues would be positive. This would eliminate the need for ad hoc intervention by NEMMCO in the market to limit counter-price flows. Once the problem of handling constrained generation is corrected, the problem of negative settlement residues is essentially a financial problem and therefore requires a financial solution.

(d) The principle that it is preferable for electricity consumers – especially large consumers – to also face a geographically differentiated price but that jurisdictions could choose to impose geographic averaging for their consumers if they wish. Correct locational pricing of electricity for consumers enhances demand-side responsiveness (reducing the need for further generation and transmission investment) and improves the locational decisions of consumers while also reducing market power. However, these effects are currently limited due to the limited responsiveness of small consumers to the spot market price of electricity. If some jurisdictions have other objectives which involve geographic averaging of electricity prices to consumers they should be able to pursue this, especially for smaller consumers.

(e) The principle that new region boundaries can be created, but existing region boundaries should not be removed. In the absence of detailed and potentially contentious modeling analysis, policy-makers cannot be sure that a region boundary change will improve the efficiency of overall outcomes. However, in a competitive market a move to finer geographic differentiation of prices over time can only improve overall efficiency. Therefore, while the welfare implications of a merger of regions are unclear, the division of an existing region can only improve welfare. The rules should allow existing regions to be divided but no existing regions (or sub-regions) to be merged. This principle would allow the Snowy, VIC or NSW regions to be divided, but those new regions could not then be merged with southern NSW or northern VIC.

(f) The principle that whatever approach to handling congestion is chosen, no new generator (i.e., one that is not already in existence or "committed") will be offered or required to accept any allocation of rights. Every solution to handling congestion will require a decision on the allocation of rights to the resulting trading surplus. Even a decision to do nothing is implicitly a decision to allow those rights to be allocated as they are now (which, as we have seen, introduces an inefficient distortion in bidding behaviour). Linking the allocation of those rights to generator investment decisions will distort the location decisions of new generators. If there is to be any free allocation of rights to the trading surplus (such as a grandfathering of CSCs), such an allocation should not be granted automatically to new generators. In effect, the AEMC should announce a policy that whatever congestion management regime is put in place no rights or entitlements will be created for generation capacity which is not already in place or "committed".

What exactly are the problems with the current arrangements for handling congestion?

3. The development of good public policy in any field requires a clear identification and specification of the problem. Without a clear understanding of the problem, it is not possible to properly identify the full range of solutions or to assess the merits of those solutions. In the field of congestion management many papers (including, especially, papers by CRA) have been somewhat vague about the precise underlying problem. As a result, these papers have been correspondingly vague as to the design and/or assessment of the merits of the various solutions. In my view, reaching a consensus as to the underlying problem would constitute material progress in resolving these issues. A substantial part of this paper therefore, is devoted to setting out the problem(s) with the current arrangements.

An introduction to congestion management

4. To begin, however, it is necessary to return to first principles. What exactly is congestion management? How is it handled in the NEM?

5. The term "congestion management" refers to the set of policies in an electricity industry which ensure that the overall industry is operated in a way which is consistent with the physical limits of the transmission network. The term also includes the set of mechanisms that are used by market participants to hedge the financial risk of trading across the physical limits of the transmission network.

6. The easiest way to understand the impact of physical limits of the transmission network is to imagine how an electricity industry would be operated in the absence of any transmission limits.

7. Consider the task of a system operator who has control of a large number of generating units and who must give orders to those units (known as "dispatch targets") as to how much to produce, subject to the overall constraint that the total amount of electricity

produced must equal the total demand for electricity (known as the "load") at every point in time.

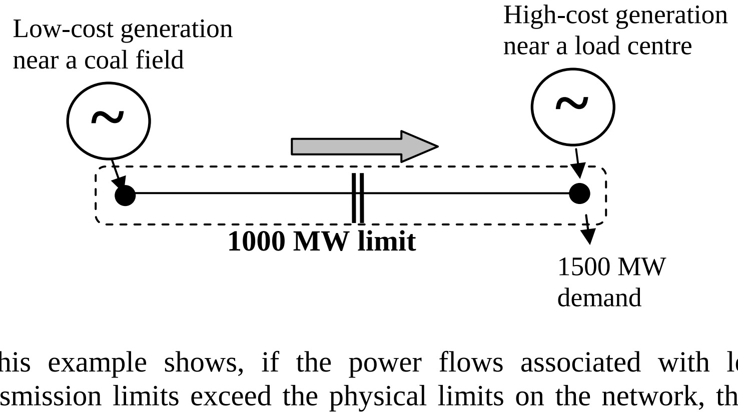
8. If the system operator is seeking to maximise overall economic welfare, the system operator will choose the "least-cost dispatch" – that is, the combination of output targets (one for each generating unit) which minimizes the total cost of generating sufficient electricity to meet the total demand.

9. It is easy to verify that, in the absence of any limits on the transmission network, the dispatch which minimizes the total cost of generation has the property that there is a common system-wide marginal cost. Each generator, no matter where it is located, is dispatched up to the point where its own marginal cost is equal to the common system-wide marginal cost. The system-wide marginal cost is also the price paid by consumers for consuming the electricity. That price is chosen so that the total quantity of electricity produced precisely matches the total quantity of electricity consumed.

10. This least-cost dispatch will correspond to a certain pattern of power flows over the transmission network. The relationship (or "mapping") from any given dispatch of the generators to the resulting power flows over the network, depends on the topology and physical characteristics of the network, the location of the generators and load, and the physics of AC power flows.

11. It may arise that the power flows associated with this least-cost dispatch ignoring transmission limits will exceed the physical capabilities of one or more elements of the transmission network. These physical limits on the transmission network arise from the need to prevent overheating on individual network elements (such as transmission lines or transformers) or from the need to ensure that the system is stable in the face of small disturbances.

12. For example, suppose that an electricity network consists of a group of low-cost "remote" generators located next to a coal field and a group of high-cost "local" generators, located next to a major load centre. Suppose that there is a transmission line connecting the remote generators with the load centre, with a capacity of 1000 MW. Suppose that the demand at the load centre is 1500 MW. In this case, since the remote generators are lower-cost, the least-cost dispatch ignoring transmission limits would involve dispatching the remote generators to produce the full 1500 MW of load. But this would give rise to a flow on the transmission line equal to 1500 MW. Since the limit on this line is only 1000 MW, the line would quickly be in danger of overheating. At best, the system operator would have only a few minutes to modify the dispatch so as to bring the flows on the transmission line back to its limit of 1000 MW. To do this, the system operator will have to back off remote generation at the coal field and turn on some high-cost generation at the load centre.



13. As this example shows, if the power flows associated with least-cost dispatch ignoring transmission limits exceed the physical limits on the network, the system operator

must choose another dispatch with a higher total cost. The system operator does this by increasing the output of some generation with a marginal cost which is higher than the common system-wide marginal cost and reducing the output of some generation with a marginal cost which is lower than the common system-wide marginal cost.

14. Since demand for electricity is almost entirely inelastic (that is, insensitive to the price) the amount by which the higher-cost generation must be "turned on" precisely matches the amount by which the lower-cost generation is "turned off". As a result, the total cost of dispatch is higher than before. The amount by which the total dispatch cost is raised as a result of constraints on the transmission network is a measure of how much these constraints have reduced overall economic welfare.[2]

15. These basic principles (that least cost dispatch in the absence of constraints implies a common system-wide marginal cost, and that least-cost dispatch in the presence of transmission constraints implies that the local marginal cost will be higher in some locations and lower in other locations) apply in *any electricity industry, no matter what its structure*.

16. In particular, these principles apply whether the electricity industry is vertically integrated, or vertically separated; whether the industry has a single generating firm or many independent firms; whether the industry has full locational marginal pricing or a single geographically-uniform price for electricity. All electricity industries, no matter how they are organised, must address the problem of transmission constraints. Efficient handling of those constraints requires that more expensive (i.e., higher marginal cost) generation must be turned on at the margin and less expensive (i.e., lower marginal cost) generation turned off. "Congestion management" is not a problem unique to a liberalized electricity market – it arises in any electricity industry.

17. There are, however, key differences between an integrated electricity industry and a liberalized electricity market. In an integrated electricity industry, the central management can use its powers to acquire information from each generator about its marginal cost, and can use its power to direct each generator to produce the appropriate quantity at the appropriate time. In a liberalized electricity market, the system operator cannot compel generators to provide information or compel them to produce electricity – instead the system operator must create an environment under which generators *voluntarily* reveal their marginal cost and *voluntarily* choose to produce to meet the system operator's dispatch targets.

18. In the NEM, all generators submit offers, signaling how much they are willing to produce at each price, to a central computer system. Under the philosophy of the NEM, these offers are assumed to reflect the short-run marginal cost of each generator. The central computer system calculates the lowest-cost dispatch, based on these generator offers, consistent with the constraints on the transmission network. Generators are paid an amount equal to the price in their region times the amount at which they are dispatched.

19. This system will deliver the economically efficient dispatch outcome provided that each generator submits an offer curve which accurately reflects its own marginal cost. This raises the question: Under what conditions does each generator have an incentive to submit an offer curve which accurately reflects its own marginal cost curve?

20. The answer is that each generator has an incentive to submit an offer curve which accurately reflects its own marginal cost (at least, at the margin) if and only if the following conditions hold for each generator:

---

[2] This is the basic economic argument for the "TCC" measure of transmission congestion.

(a) Given the price paid to that generator for its output, the generator is dispatched for an amount of output which it is willing to produce (as signaled by its offer curve) at that price. In other words, the generator is dispatched for a price-quantity combination which lies on its offer curve; and

(b) There is adequate competition between generators at the electrical location of that generator.

21. The following box shows that when these conditions hold, each generator will be induced to submit an offer curve equal to its marginal cost (at the margin):

Box 1: When will generators submit a bid equal to their short-run marginal cost (at the margin)?

Suppose that the two conditions in the text hold – that is: (a) given the price the generator is paid, each generator is dispatched for an amount which is on its offer curve; and (b) there is adequate competition between generators at each electrical location on the network. Under these conditions it is an equilibrium for each generator to submit an offer curve equal to its short-run marginal cost curve.

Suppose not. Suppose that a generator submits an offer above its short-run marginal cost curve. The generator then learns the price it will receive and the output for which it is dispatched. Under the assumption of adequate competition at its electrical location, the generator will have little impact over the price it is paid. Since this price is above its marginal cost curve (at the amount for which it is dispatched) it can increase its profit by increasing its output above the amount for which it is dispatched. Therefore, it cannot have been an equilibrium for it to submit an offer above the SRMC in the first place. [3]

Similarly, suppose that the generator submits an offer below its short-run marginal cost curve. As before, the generator will subsequently learn the price it receives and the amount for which it is dispatched. Again, assuming adequate competition, the generator's own offer will have little impact over the price it is paid. Since this price is below its marginal cost curve at the quantity for which it is dispatched, it can increase its profit by reducing its output. Therefore, it cannot have been an equilibrium for it to submit an offer below the SRMC in the first place.

Since this applies for every generator, we can conclude that, under the conditions above, every generator has an incentive to submit an offer curve which reflects its SRMC at least for that range of prices which may emerge with some positive probability.

Basic Problem #1

22. The conditions in paragraph 20 above are fulfilled in the case where there is both a separate price computed at each separate geographic location in the transmission network (known as "locational marginal pricing"[4]) and there is adequate competition at each location. In this case, the NEMMCO central computer (known as the "dispatch engine") calculates the efficient price at each node and then dispatches each generator to a quantity where the corresponding price on the offer curve is equal to that local price. Competition between generators at each location ensures that generators do not have any incentive to inflate their offer curves.[5]

---

[3] Recall that a particular offer of a generator is an equilibrium if, given the offer curves of the other generators (which, in this case, determine the regional reference price and the dispatch of this generator), the generator's own offer maximises its profit. Therefore to show that an offer is not an equilibrium we only need to show that the outcome does not maximise the profit of the generator in question.

[4] Also known as "full nodal pricing".

[5] Generators with market power raise their offer curve above their marginal cost in a process which is known as economic withholding.

23. The NEM does not at present use locational marginal pricing. Instead the NEM uses a system of "regional" or "zonal" pricing under which the price paid for generation is forced to be the same throughout a region of the NEM.

24. The conditions in paragraph 20 above would still apply in the case of a "regional" or "zonal" electricity market provided that the transmission constraints in the network were such that no generator in any region were ever required to be dispatched with a marginal cost (as revealed in its offer curve) which is higher or lower than the Regional Reference Price (RRP) in that region. Since the RRP is the price paid to all generators in a region, if all generators are dispatched to a quantity where their marginal cost (as revealed in their offer curve) is equal to the RRP then, by definition, each generator is dispatched for a price-quantity combination on its offer curve. Provided there was effective competition between generators in each region, each generator would be induced to submit an offer curve which accurately reflected its marginal cost (at least at the margin).

25. But what if this condition does not hold? In other words, what if one or more generators in a region must be dispatched with a marginal cost higher or lower than the RRP? In this case, in equilibrium, such generators will not submit an offer which accurately reflects their short-run marginal cost curve, even if there is effective competition between generators at that location.

26. Let's suppose that a generator is dispatched for an amount which is larger than the quantity where its marginal cost curve equals the RRP. In this case the cost saving by reducing output by one unit (which is the marginal cost) exceeds the reduction in revenue that results (which is the RRP)[6]. As a result, the generator could increase its profit if it reduced its output. In other words, the assumption that bidding short-run marginal cost is an equilibrium is demonstrated to be false.

27. In this case, the generator will attempt to reduce the amount for which it is dispatched by raising its offer above its marginal cost curve (or, in various other ways, pretending to be unavailable or unable to increase output). If there are a number of generators all of which are attempting to limit the amount for which they are dispatched, each has an incentive to bid their entire output (or, at least the output above the quantity at which they are willing to be dispatched at the RRP) at the price ceiling, known as VoLL[7].

28. A generator which is dispatched for a quantity which is greater than the amount it is willing to produce at the price it is paid, is said to be *"constrained on"*.[8] Such generators naturally respond by seeking to reduce the amount for which they are dispatched by raising their offer curve, or pretending to be unavailable. They will raise their offer curve to the point where either (a) the generator's bid is equal to VOLL, (b) load is shed or (c) other, more

---

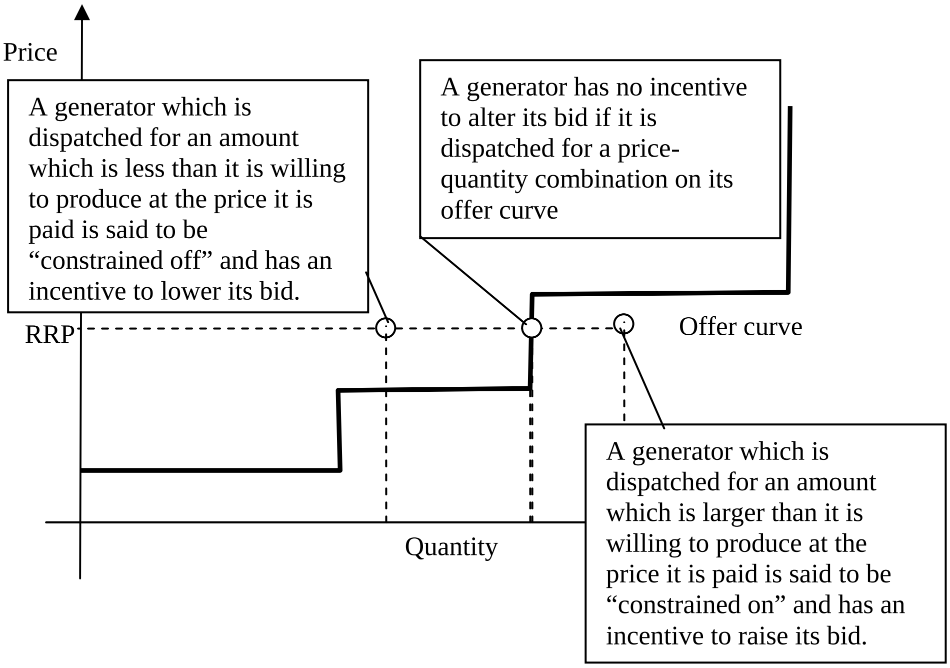
[6] Strictly speaking, we also need to establish that the bidding of this generator will not affect the regional reference price. This is clear in the case of an intra-regional constraint without loop flow. It is not so clear in the case of loop flow.

[7] VoLL stands for "Value of Lost Load" and is intended to represent (in some models) the cost (per MW) to electricity users of losing electricity supply.

[8] The AEMC paper suggests that "a generator could be considered constrained-on when its offer price is more than the regional reference price and it is dispatched to meet demand" and that "a generator might be considered to be constrained off when its offer price is less than the regional reference price, but it is not dispatched". These definitions are contained in the more general definitions in the text. The AEMC also goes on to say that "when a constraint binds, the price at the regional reference node is a combination of the offer prices of those generators whose output is increased to ensure that the network remains within operating limits. In such cases multiple generators may be marginal so that the regional reference price is not set solely by the highest offer price of an individual unit constrained on to manage the congestion". I suggest that these sentences need to be reconsidered.

expensive generation is turned on, to restore the system balance. This is illustrated in the following diagram.

Figure 1: Constrained-on and constrained-off generation



29. The opposite effect occurs for generation which is dispatched for a price-quantity combination which is above its offer curve. Suppose that a generator is dispatched for a quantity which is lower than the quantity at which its marginal cost curve is equal to the RRP. In this case, an increase in output will increase the revenue of the generator (which is given by the RRP) by more than the additional cost of the output (which is given by the marginal cost). This implies the generator could increase its profit by increasing the amount for which it is dispatched. In other words, bidding an offer curve equal to its marginal cost could not have been an equilibrium.

30. A generator which is dispatched for a quantity which is less than the amount it is willing to produce at the price it is paid is said to be *"constrained off"*. Such generators respond by trying to increase the amount for which they are dispatched by lowering their offer curve, or manipulating their bid in such a way as to maintain a higher level of output.[9] They will lower their offer curve to the point where either (a) the generator's bid is equal to $-1000, or (b) other, cheaper generation is turned off to restore the system balance.

31. Why is it a problem for a generator to submit an offer curve which does not reflect its short-run marginal cost? There are three primary consequences:

(a) First, if a generator does not submit an offer which accurately reflects its marginal cost, the overall efficiency of the dispatch will normally be reduced. More expensive generation may be used when less expensive generation is available, for two reasons:

(i) First, suppose all the generation at the "constrained" location has an incentive to offer its output at either VoLL or $-1000. In this case, the dispatch engine is no longer able to distinguish the relative cost of these generators – instead,

[9] In the past, NECA (now part of the AER) has prosecuted generators which have manipulated their bids in this way for violating the National Electricity Rules.

it simply dispatches each generator an equal amount (a "pro-rata" share of the total output required at that location), even if the marginal costs of those generators differ. As a result, some higher-cost generation will be turned on, even when some lower-cost generation capacity remains available.

(ii) Second, depending on the network topology, it may be that generators in other regions compete with generation at the constrained location. In this case, if the constrained generators are bidding VoLL, they will appear as relatively more expensive than their competitors in other regions. This will cause the dispatch engine to decrease the output of the constrained generators and increase the output of their rivals in other regions. If the constrained generators are, in fact, relatively lower cost generation, this is an inefficient outcome. The opposite effect happens when generators are constrained off. Examples of this outcome can be seen in the appendix in scenarios C and F.

(b) The second primary consequence of inefficient bidding incentives is that in the longer term, investment decisions will be distorted. In an efficient market, the "price-duration" curve (that is the proportion of the time the price spends above any given level) determines both the incentive for investment in generation capacity and the mix of different types of investment (i.e., whether baseload, mid-merit or peaking generation).

These mechanisms are distorted at locations where generators are constrained on or off. For example, consider a generator which is investing in a location which is periodically constrained off. Suppose that this generator has a marginal cost which is higher than the other generators at this location. This generator will know that if it locates in that region, when the constraint binds, it will have the same chance of being dispatched as the other generators even though it is relatively higher cost. The reduction in overall efficiency is a cost which is not fully borne by this generator (instead, it partially falls on other generators). The private incentive for investment exceeds the social benefit from that investment. In general, generators will have excessive incentives to expand capacity in locations which would aggravate existing constraints.[10]

Conversely, consider the decision of a generator which is considering investing in a region which is periodically constrained on. Suppose that this generator has a marginal cost which is lower than the other generators at the constrained location (but still above the RRP). There is a social benefit from this generator investing in this location (it reduces the cost of the additional generation required to meet the constraint) but this benefit is not captured by this generator. In general, generators will have inadequate incentives to invest in locations which would alleviate existing constraints.[11]

There may also arise distortions in network investment decisions. In some circumstances (such as in the case of MNSPs), the price differences between two regions is taken as a signal for the need for new investment between those regions. However, as we have seen, inter-regional price differences can arise even when there

[10] If the generator can choose other locations in the network, it will likely choose a location where it will not be constrained off at all. However not all generators can choose their location (some must be located next to fuel sources). The point here is that a generator which is making a decision whether to invest or not may find it privately beneficial to invest even when it is not socially beneficial.

[11] As an example, in their submission to the AEMC chapter 6 review a group of southern generators point to potentially inefficient investment in wind farms and gas-fired generation in south-east South Australia. TRUEnergy et al, submission to the AEMC, December 2005.

are no capacity constraints between the regions involved – in this case the investment in expanding the inter-regional network capacity is a pure social waste. To make matters even worse, in the event of counter-price flow, an MNSP might choose to invest in parallel with an existing inter-connector, but carrying power in the opposite direction, thereby increasing the flow on the interconnector and increasing the need for an augmentation.

(c) The distortion of dispatch arising from inefficient bidding incentives causes one further problem. As explained further in the next section, under the current NEM network design, power normally flows from regions of low prices to regions of high prices. However, the distortion in dispatch arising when some generators are not bidding their true marginal cost increases the likelihood that power will flow between regions in the opposite direction – from a high-priced region to a low-priced region. These "counter-price flows" are incompatible with the current design of the market (for the reasons set out in the next section). As a result NEMMCO must intervene to limit such counter-price flows. These NEMMCO-invoked limits may further reduce the efficiency of dispatch. There are examples of this effect in the appendix.

32. This, then, is the first of the two key economic problems arising from the current arrangements for handling constrained generation in the NEM. We might call this basic problem #1:

Basic Problem #1:

When generation in a single region must be dispatched with a marginal cost which is higher or lower than the regional reference price, that generation will be "constrained on" or "constrained off" (respectively). Under the current arrangements in the NEM, without some additional mechanism, constrained on or constrained off generators have no incentive to submit an offer curve which accurately reflects their short-run marginal cost[12]. There are three primary consequences:

(i) the short-term efficiency of dispatch is reduced – more expensive generation is turned on when less expensive generation is available. In an extreme case, NEMMCO could be forced to shed load even though generation is available.

(ii) In the medium term, investment decisions are distorted – there is under-investment in generation capacity (relative to the efficient level) in locations which would alleviate transmission constraints; and over-investment in generation capacity (relative to the efficient level) in locations which aggravate transmission constraints; furthermore, the mix of peaking and baseload generation is distorted.

(iii) This distortion to dispatch may cause power flows between regions to be counter-price, forcing NEMMCO to further intervene in the market with a further reduction in the efficiency of dispatch, and further reducing the usefulness of settlement residues as a hedging device.

33. It is worth noting that, although this problem is sometimes referred to as the problem of handling intra-regional constraints, this problem is not solely due to the presence of intra-regional constraints. In fact, the problem would arise *even if there are only inter-regional constraints* – when there are loop-flows between regions. The appendix to this note provides several examples of how the basic problem above leads to dispatch inefficiency and possibly negative settlement residues both with and without intra-regional constraints. In my view, this

[12] As an aside, it is worth noting under certain network configurations, this incentive to bid inefficiently may be offset if the generator holds a share of the inter-regional settlement residues.

problem is better described as the problem of handling "constrained generation" rather than the problem of "intra-regional constraints".[13]

34. This problem – that the existing arrangements create incentives for constrained generators to bid inefficiently – is noted by the AEMC. The AEMC writes (in section 4.2.2.1):

> "Congestion can cause inefficient dispatch by affecting a participant's incentives to bid in relation to their true costs as part of the dispatch process. The current design and operation of the NEM is predicated on the assumption that market participants have incentives to reveal costs in their bids and offers, and that dispatching the system based on those bids and offers is therefore efficient or will tend toward efficient levels.
>
> However, these incentives may be weakened if there is a binding intra-regional constraint. In simplified form, if there is a binding intra-regional constraint between a generator and the regional reference node, the generator cannot influence the price set at the regional reference node. This will mean that the price the generator receives is unlikely to be influenced by its bidding behaviour.
>
> This can result in a situation where participants offer at low prices, even submitting negative offers, as they compete to be dispatched, in the knowledge that their offers are unlikely to affect the price they receive. These perverse commercial incentives can result in inefficiencies in dispatch and pricing, which can then result in more long-term inefficiencies"[14]

35. As just noted, although this extract only mentions intra-regional constraints, this basic problem is not solely due to intra-regional constraints – the same problem can in principle arise as a result of pure inter-regional constraints (when there are loop flows, as in the example in the appendix). Second, this extract discusses the incentives of generators which are constrained off, but not the incentives of generators which are constrained on. Just as generators which are constrained off have an incentive to submit negative offers, generators which are constrained on have an incentive to submit their output at VoLL.

36. How frequently is this occurring in the NEM? This is a somewhat tricky question to answer for the reasons set out in the box below. However, it is clear that at least some generators are constrained on or off very frequently.

Box 2: How frequently are generators constrained on or off in the NEM?

Just how often generators are constrained on or off in the NEM? This is not exactly the answer to the question: "how material is this problem?" (For a discussion on how to answer that question see the box

[13] It is worth noting, if it is not already obvious, that constrained generators cannot be located at the regional reference node. The regional reference price is defined to be the marginal cost of supplying another unit of node at the regional reference node. So, by definition, each generator located at the regional reference node must be dispatched to the point where its marginal cost is equal to the regional reference price.

[14] AEMC (2006), page 27. The issue of inappropriate bidding is also mentioned by CRA in their reports but it not clear that CRA identify this as a central problem. In addition, CRA seem to view inappropriate bidding as primarily a market power question to be dealt with through competition law. Their theory paper ("NEM Regional Boundary Issues – Theoretical Framework", Sep 2004) mentioned "unproductive bidding wars" for constrained-off generation but does not mention constrained-on generation. Another paper ("NEM – Transmission Region Boundary Structure", Sep 2004) raises concerns about "inappropriate bidding" but seems to take the view that this inappropriate bidding is a matter for competition authorities (see page 7-8).

on quantification below) but it still provides a useful measure as to whether this problem occurs frequently or infrequently and which generators are most affected.

One way to attempt to measure how frequently generators are constrained on or constrained off might simply be to determine how frequently generators are dispatched for a quantity which is higher or lower than the quantity they are willing to produce at the price they are paid. There is a problem with this approach, however – as we have seen, when a generator is constrained on, it will raise its bid in an attempt to reduce the amount for which it is dispatched given the price it is paid. The generator will continue to reduce the amount for which it is dispatched until it reaches the quantity it is willing to produce given the price it is paid. In other words, if this generator is successful at reducing its output it will no longer appear as being constrained on or off.

As a result, if we ever observe a generator being dispatched for an amount above or below the amount which it has declared it is willing to produce at the price it is paid, it must only be because (a) the generator has been unable to increase/reduce the amount it is dispatched as much as it would like, despite its bid being at the bid ceiling (VOLL) or floor ($-1000); or (b) the generator is subject to a direction or a network support agreement which provides it additional revenue over-and-above the revenue it receives from the spot market; or (c) the amount the generator is constrained on or off is too small or too short-lived to justify a re-bidding effort. Putting aside separate arrangements (such as directions or network support agreements), a test which looks for generators dispatched for an amount above or below the amount they are willing to produce at the price they are paid is likely to underestimate the true impact of the problem of constrained generation in the market.

Nevertheless, it appears that at least some generators are regularly constrained on or off in the NEM. Analysis of the data collected for the TCC calculation for 2003/04 reveals that generators in northern Queensland (specifically Mt Stuart units 1 and 2 and Yabulu) were “constrained on” for a total of around 160 hours even though the generators involved were bidding all of their output at VoLL. The quantity of electricity purchased from these generators totaled almost 9000 additional MWh over-and-above what these generators were apparently willing to produce at the RRP.

In the specific instance of Mt Stuart 1 and 2, the problems arising from the current arrangements in the NEM for handling congestion are mitigated through a network support agreement with these generators. In fact, for at least 142 hours these generators were operating in accordance with a network support agreement with Powerlink. In principle, such an agreement could ensure that these generators are paid an amount which covers their marginal costs at those times when they produce and provides enough compensation towards their fixed costs to ensure that these generators have an incentive to continue to operate at this location in the NEM in the long run.

Another generator which was regularly constrained on this period was “SNUG1” which was constrained on for 3.6 hours. On 28 Oct 2003, for example, SNUG1 was constrained on when the RRP in SA was $20-25 despite its bidding at VOLL. On 24 August 2003, APS was constrained off to 95 MW, at a time when the RRP was $16, despite bidding $-986. On 18 December 2004, Hazelwood was briefly constrained off around 4 pm. At that time, the local price at the Hazelwood node was $4.09, even though the regional reference price was $50.44.

Recall, however, that this methodology will not detect all episodes when generators are constrained on or off – rather, this methodology will only detect those episodes for which the affected generators were not willing or able to fully counteract the impact of the constraints through changing their bids. The distortion to bidding caused by the current arrangements for handling congestion in the NEM could, in principle, be occurring far more frequently than suggested by this methodology.

37. One final point is also worth making. As noted in the AEMC paper, it is not necessarily the case that inefficient bidding incentives will lead to inefficient dispatch. If a generator which is constrained on bids at VOLL, and is not dispatched at all, this does not necessarily result in any inefficiency in dispatch – if the constrained on generator is, in fact, higher cost than the other generation which replaces its output. Similarly, if a generator which is constrained off, bids at $-1000 and is dispatched, there is no necessary reduction in the efficiency of dispatch – if the constrained generator is, in fact, lower cost than the generation

it replaces. In general, however, since we cannot know a generator's "true" short run marginal cost, we cannot know whether the resulting dispatch is efficient or not.

38. In the case where only generators at the constrained location are affected by the constraint, then if (a) all those generators had the same marginal cost (for example, if there is just one such generator) and (b) if NEMMCO can force those generator(s) to produce (or not to produce, respectively)[15], then there is no loss in short-term dispatch efficiency at all, no matter what these generator(s) bid (since the total amount for which they must be dispatched is independent of their bid). There still remains, however, the problem of achieving the correct incentives for generation investment at that location.[16]

Basic Problem #2

39. The second basic problem with the current congestion management arrangements in the NEM relates to the handling of the "trading surplus" that accrues to NEMMCO as a result of its trading of electricity at different prices at different locations.

40. In the previous section I observed that efficient management of congestion in a liberalized electricity market requires setting different prices for electricity in different locations on the network. In effect, we can imagine that the system operator "buys" electricity from generators at one price and "sells" electricity to load at another price. The theory of transmission pricing shows that, under efficient dispatch in a market with correct locational pricing[17], the system operator always makes a net profit from buying and selling electricity in this way. This "profit" is variously known as the "trading surplus", "merchandising surplus" or "settlement residues".

41. It is very important for this trading surplus to be made available to market participants as a tool for hedging the financial risk of trading across two locations with potentially different prices for electricity. Without access to this trading surplus, market participants which are trading electricity across locations with different prices face the risk of price separation between those locations. In the absence of any mechanism for hedging this risk, market participants will require compensation for bearing that risk in the form of a risk premium. This will limit the volume of trade between different locations – leading to permanent forward or "contract" price differentials across locations – larger than can be justified by the physical transmission limits between those locations alone.

42. Ideally the trading surplus would be divided up into streams which satisfy two conditions:

(a) These streams would, as far as possible, facilitate the writing of "firm"[18] hedges between two differently-priced locations; and

(b) These streams would be positive, for reasons discussed below.

---

[15] If NEMMCO cannot force a constrained-on generator to produce there is a risk that load will need to be shed. This, of course, is a yet another form of dispatch inefficiency.

[16] In practice, this problem is often addressed through some form of "network support agreement" – the relevant TNSP will seek to reach an agreement with a generator in a strategic location to compensate that generator for producing more (or less) at certain times which partly offsets the incentives-for-investment problem.

[17] As we will see later, in a market with geographically-averaged prices for consumers this result does not necessarily hold.

[18] Again, as firm as the physical capability of the underlying network.

43. The concept of "firmness" of a given stream of residues relates to the quantity of firm inter-regional hedges that can be written using those residues as backing and the remaining risk that must be borne by the holder of the residues. The concept of firmness is discussed further in the attached box.

Box 3: What is "firmness" and why is it important?

Suppose that a generator at location A wishes to sell a contract to sell 100 MW of power to a retailer at location B over the period of one year. This generator is then exposed to several risks, including the risk that price differences will arise between location A and location B. To perfectly hedge this risk, this generator would like an instrument with a payoff equal to the price difference between A and B times 100 for each trading interval in the year.

If A and B are located in different pricing regions, there may be an inter-regional settlement residue defined between those two regions. An inter-regional settlement residue yields a stream of payments is equal to price difference between A and B times the flow on the interconnector between region A and region B. Under certain conditions (in the absence of constrained generation and in the absence of loop flow) it can be shown that the price-difference between two regions is zero unless the interconnector between the two regions reaches its limit.

Let's suppose this limit is fixed and constant at 2000 MW. The generator could then purchase a 5% share of the inter-regional settlement residues and be perfectly hedged – since either the price-difference is zero, or the settlement residues pays out 5% of the price difference times 2000, which is exactly the hedge the generator requires.

Suppose, however, that on occasions the capacity of the interconnector must be reduced (perhaps due to maintenance, or outages). Suppose that on these occasions, the capacity of the interconnector is reduced from 2000 MW to 1000 MW without notice to the market. In this case, the inter-regional settlement residue is no longer a perfectly firm hedge. The generator might try to compensate for this by buying a slightly larger share of the settlement residues. However unless those outages can be predicted with certainty the generator will continue to bear some risk. In this case the settlement residues are not perfectly firm but are *as firm as the physical capability of the underlying network*. If the underlying network is not perfectly "firm" it is not possible for the system operator or any other market participant to offer perfectly firm hedges without taking on some risk.[19]

As we will see later, for various reasons related to the definition of the inter-regional settlement residues themselves, there are many occasions when price differences may arise between regions even when the interconnector between the two regions is not at its physical limit. For example, price differences between the regions might arise even when the interconnector is nowhere near its physical limit, but only flowing at 500 MW or 200 MW or some other figure. In this case inter-regional settlement residues defined in this way are not at all "firm" and are only of limited usefulness at hedging inter-regional price risk.

To make matters even worse, as can be seen in the appendix, price differences may also arise between regions even when the flow on the interconnector is in the *opposite* direction. This "counter-price flow" gives rise to negative settlement residues, which create further problems for the system.

Intuitively, we can see that the "firmness" of a given stream of residues is related to the variability in the ratio of the total settlement residues to the price difference between two regions. This ratio (which reflects the flow on the interconnector) should ideally be a fixed, positive value when price-differences arise. However, in practice, this ratio is highly variable, unpredictable and may even be negative. This lack of firmness arises primarily from the way the inter-regional settlement residues are presently defined.

[19] Some commentators argue that the TNSP should take on that risk, as an incentive device to induce the TNSP to deliver high levels of availability. I will not discuss this issue further.

44. Under the present market design, the trading surplus accruing to NEMMCO arises primarily from price-differences between regions. At present, this trading surplus is divided up into streams known as "inter-regional settlement residues" or IRSRs. The IRSR between two neighbouring regions is defined as the price-difference across those regions times the flow on the interconnector between those regions.[20]

45. Under the present NEM arrangements, NEMMCO auctions these streams to market participants in what amounts to a fixed-for-floating swap.[21] However, under the present arrangements, there is no mechanism for NEMMCO to fund its losses when the residues on any one stream turn out to be negative. Therefore, each stream of residues must yield a positive payoff each period.[22]

46. But, under what conditions do the "inter-regional settlement residues", as currently defined in the NEM, satisfy the two conditions above (that they facilitate writing "firm" inter-regional hedges and that they are positive)? It turns out that the IRSRs only satisfy these conditions under strictly limited conditions. In fact, it is easy to show that the IRSRs are non-firm hedging instruments in a variety of situations.

47. Specifically, as I show in the appendix, IRSRs are non-firm (and may be negative) even under fully efficient dispatch in the presence of constrained generation – and this applies whether the binding constraint is inter-regional or intra-regional and whether the network is radial or meshed. Even in the absence of constrained generation – that is, even under fully efficient dispatch – IRSRs will still be a non-firm instrument in a meshed network. The distortion to dispatch efficiency brought about by constrained generators not bidding their true marginal cost only makes this problem worse.

48. The problem lies in the definition of the IRSRs. Defining IRSRs as equal to the price difference between regions times the flow between those regions will only yield a "firm" hedging instrument and a positive stream of residues if the following conditions are satisfied:

(a) there are no generators constrained on or constrained off (in the sense discussed earlier); and

(b) there are no electrical loops between regions.

49. These conditions are not satisfied in the NEM, and are even less likely to be satisfied in the future, as intra-regional constraints become more important and as the NEM becomes more "meshed".

50. This lack of firmness of IRSRs causes two primary problems. The first is that the risk of inter-regional trading is increased, increasing the barriers to inter-regional trade and allowing long-term price differentials between regions to persist.

---

[20] The interconnector is not necessarily a single transmission line but may consist of many lines flowing roughly in parallel. In this case the flow on the interconnector is defined to be the sum of the flows on these individual lines.

[21] We can imagine that NEMMCO engages in a fixed-for-floating swap – it swaps the "floating" payment corresponding to the residues for a fixed payment from a market participant equal to the auction proceed. We would expect that the fixed payment should be roughly equal to the expected value of the floating payment. This principle still applies if the residues are negative. In this case, NEMMCO would be in effect swapping the floating negative residues for a fixed payment to the market participant.

[22] NEMMCO has considered offsetting negative residues at certain times with positive residues elsewhere – either at the same time on other interconnectors or on the same interconnector at other times. However, except in certain special circumstances, this practice of "smearing" the negative residues will not normally yield a stream of payments which is useful for writing firm hedges.

51. Furthermore, when IRSRs threaten to become negative, NEMMCO is forced to intervene. Under the present market arrangements, NEMMCO cannot allow negative settlement residues to persist on any one stream of residues. When significant negative IRSRs threaten to accrue on any stream of residues, NEMMCO intervenes in the market to limit the negative settlement residues. It does this in two ways: either by (a) restricting flows between regions; or (b) effectively merging two regions. However, as we have seen, negative settlement residues arise even under fully efficient dispatch in a variety of circumstances. In these cases, NEMMCO's intervention to restrict inter-regional power flows or to merge regions can only reduce the efficiency of dispatch and thereby distort pricing and investment decisions.

52. The current practice in the NEM of paying all generators in the same region the same price, even when they are constrained on or off (the first basic problem above), further exacerbates these problems – increasing the likelihood of negative settlement residues and limiting the ability of constrained generators to hedge their risks.

53. Furthermore, under the current arrangements, generators which are constrained on or off cannot effectively hedge the financial risks which they face.[23]

54. This then is the second basic problem with the current arrangements for handling congestion in the NEM:

Basic problem number 2:

The current definition of the streams of payments known as "inter-regional settlement residues" does not guarantee that the individual streams of residues will either (a) facilitate the writing of "firm"[24] contracts across price regions; or (b) be positive. In particular, the IRSRs will be non-firm – even under efficient dispatch – when some generators must be constrained on or off. They will also be non-firm even without constrained generation in a network with loop flow.

This lack of firmness is an obstacle to trading between pricing regions. In addition, in extreme cases, NEMMCO is forced to intervene to limit the accumulation of negative settlement residues. The actions by NEMMCO to limit the negative settlement residues further reduce the efficiency of dispatch.

These problems are made worse in the NEM by the current practice of paying all generators in the same region the same price even when some generators are constrained on or off (the first basic problem above). This policy worsens the problem of negative settlement residues and effectively dissipates the intra-regional settlement residues in a manner which does not allow constrained generation to effectively hedge its risk.

[23] According to one view, intra-regional settlement residues already exist under the current NEM arrangements, but these residues are dissipated in a way which does not facilitate hedging. We could imagine that, under the current arrangements, constrained generators are, in fact, paid their correct locational price – but then are simultaneously given a share of the resulting settlement residues equal to the difference between that locational price and the regional reference price times the output for which they are dispatched. From this thought-experiment we can see that, at present, the intra-regional settlement residues that arise when generators are constrained are currently being paid back to the constrained generators – but in a manner which both distorts their bidding incentives and does not allow them to effectively write firm hedges.

[24] Again, as firm as the physical capability of the underlying network.

55. These issues are, to an extent, recognized in the AEMC paper. Section 4.3.2.2 notes that:

> "Typically, the IRSR has a positive value, indicating that the price paid in an importing region exceeds that paid in the exporting region. However, the IRSR can sometimes be negative, indicating the reverse – that is, power flows are counter to the price difference, from a high price region to a low price region.
>
> The IRSR units do not provide a firm financial hedge against inter-regional price risk. This is because the IRSRs that accrue are a function of the direction and flow on the link over time and the units sold at the IRSR auctions are an entitlement to a proportion of the residues that accrue. If the flow on an interconnector is limited to less than the nominal rating used as the basis of the SRA, the amount of IRSR to be shared among IRSR unit holders will be reduced on a pro-rata basis, even though their financial exposure from price separation may be unchanged.
>
> Some participants have expressed concerns that the non-firm nature of IRSRs reduces their effectiveness as an inter-regional hedging tool. Similar concerns may arise in relation to the limited duration of the units that are auctioned (quarterly), and the limited degree to which they can be purchased on a forward basis, compared to the longer term nature of many bilateral financial contracts".

56. As I discuss below, this second basic problem requires a fundamental re-think of the way we define settlement residues.

Other problems?

57. Before going on to discuss the way forward, it is worth exploring whether there are any other potential problems in the NEM which should be mentioned.

58. One potential problem is the problem of generator market power. Earlier I noted that one of the conditions for efficient bidding is that there should be adequate competition between generators at each electrical location on the network. In the presence of transmission constraints, these electrical locations can become quite small – isolating one or just a few generators. In this circumstance, such generators can have substantial market power.

59. It is possible that the existing arrangements in the NEM are masking some of that market power – under the existing arrangements, generators which are constrained on are paid no more than the RRP, no matter how much market power they might have. The solutions considered by the AEMC (and discussed further below) will likely involve paying such generators a price which is higher than the RRP. In this context, certain generators may be able to increase their locational price through the exercise of market power.

60. There is a sense in which generator market power is a third "basic problem" with the current arrangements for handling congestion in the NEM. However, the problem of market power is not directly related to congestion management – it arises whenever a generator can, by varying its output, change the price that it receives. The exercise of market power already occurs in the NEM. Improvements in the handling of congestion management may require improved procedures for handling generator market power, but otherwise the two issues are conceptually quite separate.[25]

[25] In their theory paper, CRA raise the issue of allocating CSCs in a way to mitigate market power. However, generators can presumably trade CSCs to obtain any final allocation they desire. It is hard to see how an initial allocation of CSCs will have any final impact on a generator's market power.

61. Another issue which is raised by the AEMC relates to the issue of formulation of constraint equations. The debate over congestion management in the NEM has, on occasion, been associated with the debate over the appropriate formulation of constraint equations in the NEM. It may be worthwhile to spell out the linkages between these two policies here.

62. As noted earlier, the philosophical approach of the NEM is that generators are dispatched according to the costs they themselves reveal in their bids subject only to the physical limits of the transmission network. The physical limits are represented in the "dispatch engine" in the form of mathematical equations known as "constraint equations".

63. Under the philosophical approach of the NEM, all of the constraint equations should be formulated in such a way as to put all generators in the NEM on an "equal footing" with other generators subject only to the physical limitations of the transmission network – whether those generators are located in the same region, in a neighbouring region, or at the other end of the NEM. Put another way, the constraint equations should be formulated in such a way as to accurately and correctly reflect the underlying physical limits and capacities of the network – and nothing else. This has been described as the "Direct Physical Representation" or "Option 4" approach to constraint formulation.

64. However, because of the problem of inefficient bidding and possible negative settlement residues noted earlier, the market outcomes when all the constraints are formulated under the "option 4" methodology are, on occasions, manifestly inefficient. In this case, market participants may be tempted to argue that certain constraints be formulated in a way which favours some generators over others, in an attempt to restore efficiency.

65. This is an example of using two wrongs to make a right. The current arrangements in the NEM cause constrained generation to bid inefficiently. This "wrong" could in principle be offset by another "wrong"- that of formulating the constraint equations in a way which favours some generation over others. As I argue later, rather than seek to offset one "wrong" with another, it is preferable to address each problem individually. The problem of inefficient incentives on constrained generation should be put right using one or more of the solutions discussed below; at the same time, the constraint equations should be formulated in a way which reflects only the physical limits on the network.

66. The AEMC has rightly asked the question as to how best to quantify the costs of the two basic problems identified above. The box on the following page discusses how tools being developed by the AER may provide some insight in answering this question.[26]

[26] It is also worth pointing out that NEMMCO, in fulfilment of its obligations under the rules, publishes, for each dispatch interval, the results of an alternative run of the dispatch engine in which "intra-regional constraints" are relaxed. This so-called "BNC" run is intended to provide some indication of the extent to which the dispatch of generators is being affected by intra-regional constraints. In principle, by valuing these change in dispatch we could obtain a measure of the costs of intra-regional constraints. However, there are a number of problems with this approach: (a) the ramp rate limits are not relaxed, limiting the extent to which the BNC dispatch can depart from the dispatch in the production run; (b) the approach ignores inter-regional constraints which could result in constrained generation; and (c) as discussed in Box 2, the approach will completely ignore generation which is bidding inefficiently but which is successful at preventing its dispatch from being affected by the intra-regional constraint.

Box 4: Quantifying the costs of the current congestion management arrangements

The AEMC has rightly asked the question of how to quantify the costs associated with the current arrangements for handling constrained generation and the lack of firm (and occasionally negative) settlement residues.

The basic problems identified above will cause problems in both the short-term and the longer-term. In the short-term there will be a loss of short-term dispatch efficiency. In the longer-term there will be a persistence of inter-regional price differentials which are unrelated to the underlying physical constraints, and the investment decisions of generators and transmission companies will be distorted. Of these two impacts, the longer-term investment impacts are relatively difficult to measure, but in principle it is straightforward to assess the impact of these problems on the short-term efficiency of dispatch.

To determine the impact of these problems on the short-term efficiency of dispatch we need to answer the question: how much lower would be the total dispatch cost if constrained generation and negative settlement residues were handled efficiently?

This question is similar to the question which the AER is seeking to answer in computing the cost of transmission constraints (the "TCC"). The TCC is the answer to the question: how much lower would the dispatch cost be if we relax all the transmission constraints in the network? The AER is currently in the process of preparing the first set of results of the TCC for the 03/04 financial year.

As noted above, the impact of the current arrangements is that constrained generation does not bid in a way which reflects its short-run marginal costs and, in some cases, NEMMCO is forced to intervene with unnecessary additional constraints to control counter-price flows between regions. In order to determine the cost of the current congestion management regime we need to determine how much lower the cost of dispatch would be, assuming that all constrained generators bid at their marginal cost and eliminating any constraints imposed by NEMMCO to manage negative settlement residues.

It is not always easy to determine the short-run marginal cost of a generator – in particular, this problem is notoriously difficult for a hydro generator. However, in principle it might be possible to come up with an estimate of the short-run marginal cost of a set of constrained generators and then to assess how much lower the dispatch cost would have been if those generators had bid efficiently. At the same time we could relax the constraints that NEMMCO imposes to manage counter-price flows.

This analysis would give a picture as to the magnitude of the costs of the current congestion management arrangements in the past. However two facts should be borne in mind:

(a) First, the costs of the current arrangements are likely to evolve over time. As demand patterns change and new constraints emerge it is likely that intra-regional constraints will become more important in the future rather than less important.[27] Developments such as new generation and the commissioning of Basslink will change the pattern of flows on the existing network. The AEMC, in making decisions as to how to proceed on congestion management should base its decision on a forward-looking future (rather than historic) estimate of the costs of the current regime.

(b) Second, this approach does not and cannot measure the other sources of harm from the existing arrangements, such as the consequences for inefficient location decisions by generators or consumers, or even inefficient investment decisions.

[27] CRA (2004a) write that "We … note that future development of the network is likely to lead to more occasions when [the existing] approach will reduce efficiency of dispatch, especially as more and larger loops are created in the network due to normal expansion" CRA (2004a), page 6.

Solutions

67. In this paper I have taken the view that it is too early to express detailed views on the pros and cons of the different potential options. Rather, I have suggested below a set of principles which could guide AEMC decision-making as they seek to develop and assess specific options. However, it may be worth making a few comments first on the nature of any potential solutions.

68. We saw above that the first basic problem is that constrained generators do not have an incentive to bid their short run marginal cost. In order to restore the incentive to bid short run marginal cost, we need to achieve two things:

(a) First, since the problem arises from a mismatch between pricing and dispatch for certain generators, this relationship between pricing and dispatch must (naturally) be restored. Each generator which is constrained on or constrained off must be paid a price (at the margin) at which it is willing to produce the quantity for which it is dispatched – in other words, the combination of the price paid for the output of the generator and the amount at which is dispatched must be a price-quantity combination on its offer curve.

In other words, *every solution to the first basic problem identified above involves some form of geographic differentiation of the prices paid to generators*.[28] There is no getting around this. In effect – although in different ways and with different implications – it is possible to assert that *every* solution to this basic problem will involve some form of move towards correct locational pricing of electricity for generators.

(b) Second, each generator must, as far as possible, face adequate competition at its electrical location (or, where that is not possible, some other form of control on a generator's market power).

69. The solutions to this first basic problem raised by the AEMC include:

(a) Increasing the number of regions;

(b) Network support agreements;

(c) Payments to constrained-on or constrained-off generators;

(d) The CSC/CSP mechanism (discussed further below).

70. All of these solutions have in common that they increase the degree of geographic differentiation of the prices paid to generators – generators which are currently constrained on or constrained off will, under these arrangements, receive a price (at the margin) which is higher or lower than the RRP they receive today. All solutions to the first basic problem of congestion management will involve a move to more finely-differentiated prices to generators.

71. These different approaches do, however, differ in important ways. In particular, these different approaches differ in:

[28] For completeness, I note that it is possible to construct scenarios in which a reformulation of the constraint equations is a plausible solution, but these scenarios are exceptional and involve the system operator making judgements as to the marginal cost of different generators. In my view any attempt to reformulate the constraint equations should be dismissed as a solution.

(a) whether or not prices to *consumers* are geographically differentiated. In fact all of the approaches above except (a) (increasing the number of regions), will leave prices paid by consumers unchanged. As discussed later, this is, in principle, less efficient than region boundary division but in practice, given the very low elasticity of demand for electricity, the efficiency cost of this is likely to be small in the short term.

(b) how these prices are computed (whether these prices are computed by the existing NEM dispatch engine, by a separate auction mechanism, negotiated with the generator, or computed mechanically through the marginal values of the binding constraints);

(c) the handling of the "trading surplus" arising from constraints and, in particular, the extent to which generators are automatically granted a share of this surplus. Since all of the solutions to the basic problem involve finer geographic differentiation of prices to generators, it follows that there will arise new sources of trading surplus which should be made available to market participants in some way.

Region division

72. It should be clear that the inconsistency between pricing and dispatch underlying the first basic problem above could be restored by increasing the number of regions – ideally to the point where there are no constrained generators, so that every generator is paid a price at the margin, consistent with its dispatch.

73. More generally, under the assumption of effective competition between generators at all locations, the division of a region can only ever improve (or leave unchanged) the efficiency of dispatch. In other words, even without further analysis, under this assumption, a region division can only ever improve welfare.

74. The same cannot be said, of course, of a merger of regions (or parts of regions). If, in some market outcomes, the merger of two regions would force a uniform price on two generators who, in an efficient dispatch, would be dispatched to a different marginal cost, then the problems identified above will recur, with the corresponding reduction in the efficiency of dispatch. A merger of two regions can only ever reduce (or, at best, leave unchanged) the efficiency of dispatch. On this basis, I argue below that we should adopt the principle that only divisions of existing regions should be permitted (except perhaps under exceptional circumstances). This would ensure a move in the direction of finer geographic differentiation of prices over time.

75. Should we prefer the division of existing regions over the other solutions to the first basic problem above? The key difference between this solution (region division) and the other solutions is that region division (unlike the other solutions) also increases the geographic differentiation of the prices paid by consumers of electricity. All the other solutions above involve geographic averaging of the prices paid by consumers.

76. There are strong reasons to prefer an approach under which consumers also face the correct locational price:

(a) The first reason relates to the implications for demand-side response to prices. As long as consumers face a geographically-averaged price for electricity they have no incentive to respond to the local price by increasing or reducing their demand, even when it is efficient to do so. Demand-side responsiveness has several well-known advantages. In particular, it reduces the need for investment in augmenting the network and in peaking generation capacity. Demand-side responsiveness also reduces the geographic variation in prices paid to generators.

In addition – and in some circumstances this can be very important – demand-side responsiveness can significantly curtail the opportunities for exercise of market power. The market power of a generator depends strongly on the elasticity of the demand curve. Increasing the elasticity of demand for electricity can significantly reduce market power. As noted earlier, it is likely that the existing arrangements are masking the market power of some generators. A move to finer geographic pricing for generators may therefore require additional policy tools to control that market power. Allowing demand-side response by consumers could be a key tool for controlling potential market power.

In the short-term, the responsiveness of electricity demand to price is limited, but there are some well-known instances of large electricity consumers already responding to price spikes and this is likely to grow over time with the roll-out of interval meters and as consumers make choices about switching to other fuels.

(b) The second reason relates to consumer location decisions. As long as consumers face a geographically-averaged price for electricity they have no incentive to make efficient location decisions. An aluminium smelter might, for example, choose to locate in northern Queensland (close to say, a bauxite mine) even though that decision could impose significant costs on the network. For many consumers, of course, electricity costs are not sufficiently large as to affect their location decisions. However, for the largest electricity consumers – in other words, precisely those consumers who we would want to make efficient location decisions – geographic averaging of electricity prices may induce inefficient location decisions.

(c) There are a couple of other, slightly more technical, reasons for preferring an approach under which consumers face the correct locational price.

(i) The first relates to the handling of losses. At present in the NEM, inter-regional losses are correctly computed dynamically in each five-minute dispatch interval. Intra-regional losses, however, are handled through an approximation using static “marginal loss factors” which are only updated once every year. This approximation yields to small inefficiencies in dispatch. A region division, by improving the handling of losses on certain lines, would reduce the reliance on this approximation and increase the efficiency of dispatch.

(ii) The second technical issue relates to the overall residues or “trading surplus”. In a network in which consumers face a geographically-averaged price there is no guarantee that the overall “trading surplus” of the system operator will automatically be positive. The reason is that geographically-averaging the price leaves consumers in constrained-on locations better-off than under full locational marginal pricing. In effect they receive an “implicit subsidy” per MWh equal to the difference between the regional reference price and the local price.

For example, suppose that an intra-regional constraint requires generation in far-north Queensland to be dispatched at a time when local demand in far-north Queensland is 200 MW. The local marginal price for generation may be, say, \$100/MWh, even though the regional reference price is, say, \$20/MWh. Let’s suppose that the line into far-north Queensland is at its limit and is carrying 100 MW. The total surplus accruing to NEMMCO is the residue from the constrained line (\$100-\$20 times 100 = \$8000/hour) less the "implicit subsidy" for consumers in far-north Queensland equal to \$100-\$20 times 200 = \$16,000/hour. In this example, NEMMCO makes a loss of \$8000 per hour.

Note that this problem already arises in the NEM, in the context of network support agreements. At present, network support agreements are funded through the regulated revenue stream of TNSPs. The same mechanism could, in principle, be used to finance the short-fall in the trading surplus that results from geographic averaging of consumer prices. A move to finer locational pricing for consumers would, however, solve this problem without requiring external sources of financing.

77. As this discussion makes clear, region boundary division – which allows finer geographic differentiation of prices to generators and consumers – is clearly the preferred "first best" outcome for resolving the first basic problem above. However, given the relatively limited responsiveness of consumers to electricity prices in their consumption or location decisions, geographic averaging of prices to consumers is acceptable in the interim, until a region division could lead to a more permanent solution to the problem.

78. If geographic averaging of prices for consumers must be maintained, it is preferable for that averaging to be limited to those consumers who are least responsive to the local spot market price for electricity. Large electricity consumers, who could be expected to change their consumption or their location in the face of electricity prices, should face the correct locational price. Geographic averaging of electricity prices should be limited to only the small residential or small business consumers. At the least, demand-side bidders in the NEM (such as the large pumps) should probably receive the local generator price.

79. Note that there is no reason why different parts of the NEM could not determine their own policies regarding geographic averaging for consumers. Some areas of the NEM could pursue finer geographic pricing for both generators and consumers, seeking the efficiency gains that result; while other areas of the NEM could maintain geographically-averaged prices for consumers, choosing to tolerate the resulting inefficiencies and funding any shortfalls that result. As long as generators face the correct locational prices at the margin, there does not seem to be any further reason to require consistency in pricing for consumers across the NEM.

80. Before leaving this topic, it is worth considering the role of investment in solving the first basic problem above. The AEMC in its "staged approach" to congestion management places "boundary change" as the last option after many intermediate stages including investment in transmission. Should investment options be exhausted before considering a region boundary change?

81. Note first that this staged approach discussed by the AEMC is not consistent with the "cycle of phases" approach discussed by CRA in their theory paper[29] in which investment was a last resort option, only after boundary change options have been exhausted.

82. The AEMC, in arguing for a staged approach to congestion management, comments that "boundary change can result in high costs. It may require amendments to systems operated by market institutions and market participants. It is also likely to be disruptive to

[29] CRA (2004b), page 61.

participant hedging arrangements. Many participants enter hedges of several years duration, which could be affected by a change to region boundaries".[30]

83. However, it is worth noting that the transactions and adjustment costs are at least partly linked to the introduction of new pricing arrangements. If, as argued here, the NEM introduces new pricing arrangements for generators, a proportion of those adjustment costs will be incurred whether or not a region boundary change is involved. The remaining adjustment costs can be mitigated through a suitable period of advance notice of a forthcoming boundary change, allowing participants time to adjust their hedging arrangements.

84. In assessing the role of investment in managing congestion, the key question is not whether region boundary change involves costs – *but whether it involves costs larger than the physical costs of the investment*. At a rough guess, it would seem that the adjustment costs and transactions costs associated with region boundary change would be unlikely to amount to more than millions of dollars, whereas an investment in augmenting the network is likely to cost tens or hundreds of millions of dollars. A region boundary change involves, at one level, a mere change in the rules governing the market with no need for any physical changes in the network at all. Achieving the same improvement in the efficiency of dispatch through a network augmentation could easily cost tens of millions of dollars. Viewed in this light, it seems inappropriate for region boundary change (or, as I argue, region division) to be given a lower priority than investment.

The CSP/CSC scheme

85. The CRA CSP/CSC scheme has been raised by the AEMC for, at least, interim congestion management before a region division occurs. The CSP/CSC scheme is a mechanism which:

(a) Pays a constrained generator a price which reflects some or all of the binding constraints at a given point in time. This price is referred to by CRA as the "Pseudo-Nodal Price" (or PNP[31]); and

(b) Pays that (and other generators) a fixed allocation of the proceeds of the resulting "trading surplus" (this is the CSC component); this amount may be negative – in other words a generator might have to make a payment to the system operator.

86. There are two concerns which could be raised with respect to the CSP/CSC scheme. The first relates to the choice of constraints which will be covered by the CSP. CRA in their earlier presentations and documents leave the impression that the system operator could choose which constraints would be handled with a CSP/CSC; small or infrequent constraints could, it is implied, be ignored while larger constraints could enter into the CSP/CSC congestion management regime.[32]

87. It seems important to emphasise that unless the CSP/CSC mechanism covers *all* binding constraints, the pseudo-nodal price will be above or below the correct locational price. As a result, the same issue of inefficient bidding incentives will arise – generators will continue to be dispatched to a quantity which is above or below the quantity at which they are

[30] Page 37.
[31] See CRA, (2005), "Constraint Support Pricing: Implementation of Snowy Proposal", March 2005, page 11.
[32] For example, CRA's "Theory Paper" states "These CSCs behave very much like FTRs, except that they apply selectively to particular constraints", (page iii).

willing to be dispatched at the price they are paid – even after that price is adjusted using the CSP. This issue is recognized by CRA in their March 2005 paper on the Snowy Proposal. In that paper they noted that "the pseudo-nodal price for participant *p* will be the same as the nodal price for its node if and only if CSPs are applied to ALL constraints affecting *p*"[33]

88. In other words, if the problem of inefficient bidding is to be resolved, the price a generator receives under the CSP/CSC scheme *must* be equal to the full locational marginal price. It seems to me that there are no short-cuts here. Although the system operator could, in principle, ignore small or infrequent constraints, the resulting CSP/CSC mechanism will not solve the problem of inefficient bidding – thereby calling into question the value of a move to the CSP/CSC mechanism in the first place. Any mechanism for solving the problem of inefficient generator bidding *must* move directly to full locational pricing for generators.

89. The second concern with respect to the CSP/CSC scheme relates to the allocation of shares in the trading surplus. CRA recognize this will be contentious, and devote 20 pages of their Theory Paper to this issue. Are there any principles which we can note to guide this allocation decision?

- First, it is worth noting that this is not a decision that can be avoided. Any decision (including a decision to do nothing about congestion management) is a decision which involves an implicit allocation of rights.[34] Under the present arrangements, CSCs are, in effect, allocated in proportion to the dispatch target of each generator. This gives rise to the current inefficient bidding incentives. Correcting these incentives will involve, in effect, separating the CSC allocation from the dispatch target of each generator. But the decision as to the allocation still remains.

- Second, CSCs should not be granted automatically to new generators. In effect, the AEMC should announce a policy that whatever congestion management regime is put in place no rights or entitlements will be created for generation capacity which is not already in place or "committed". Just as the current policy of linking the CSC allocation to the dispatch target creates inefficient incentives for generators, linking the CSC allocation to *any* action on the part of the generator will also distort generator incentives (whether that action is a location, expansion, contraction, or fuel choice decision).[35]

- Third, if any CSCs are allocated for free to existing generators it is very important that these rights be tradeable, so that new generators (who may not be automatically allocated any rights) are able to acquire the hedging instrument they need to establish in a particular location. If CSCs were essentially non-tradeable, incumbent generators (who received these rights automatically) would have a significant competitive advantage over new entrant generators. For any proposed allocation methodology it

[33] CRA, "Constraint Support Pricing: Implementation of Snowy Proposal", page 11. Emphasis in the original.

[34] At present, CSCs are effectively allocated one-for-one with each unit of dispatch. This creates incentives for constrained off generators (for whom CSCs have positive value) to seek to be dispatched to inefficiently high levels. It also creates an incentive for constrained on generators (for whom CSCs have a negative value) to seek to avoid being dispatched to the efficient level.

[35] CRA in their "Theory Paper" notes that "If the newcomer sites in an area where the CSP impact is positive, it will have no desire for a CSC clawing back that advantage. Ant it would be economically inefficient to impose one, because that would negate the economic signal which correctly incentivises them to locate in that position. … If the newcomer sites in an area where the CSP is negative, it will certainly want a CSC to offset that disadvantage. But it would be economically inefficient to provide one, because that would negate the economic signal which correctly incentivises them not to locate in that position" (page 59).

will be important to assess the extent to which the rights could be traded between generators.

90. Beyond these principles, economic theory can provide relatively little guidance as to how to allocate these rights. One possible principle is that, in order to minimize regulatory risk, these rights should be allocated in a manner which respects the legitimate expectations of generation investors in the NEM. It could be argued that, since the code makes clear the ability to change region boundaries frequently (annually), new investors could not assume that they would have long-term rights to sell their output at the current regional reference nodes. This would suggest that generators should not receive an allocation of rights which makes them better off than if there had been a region boundary change.

91. On the other hand, a case could be made that the obstacles to region boundary changes that have arisen in practice have been so significant as to make a region boundary change a remote possibility in the medium term, so that a generator could have a reasonable expectation it could continue to sell its output at the regional reference price in the medium term. This would suggest that generators should receive an entitlement which reflects their current right to sell their output at the regional reference node.

92. Another possible principle relates to the implication of an allocation of rights for the incentives to oppose a region division. Constrained-off generation which receives an allocation to sell a share of its output at the regional reference node has a strong incentive to oppose a region division (which would effectively extinguish those rights). It could be argued that since, in the long-run, a region boundary division is the most efficient way to manage congestion, no generator should be given a reason to oppose a region division.[36] On the other hand, such a generator would then have a strong incentive to oppose the introduction of any congestion management arrangements in the first place. It could be argued that grandfathering existing rights is necessary to obtain the consensus needed to bring about reform of the NEM's approach to congestion management.[37]

93. This issue of the allocation of rights arises under any new arrangement for handling congestion in the NEM. It will be a key issue for the AEMC to address. One possible compromise approach might be to grandfather existing rights but only for a limited period of, say, five years.

94. The AEMC raises the question of what lessons can be learned from the trial of the CSP/CSC concept in the Snowy region. I have not conducted any new analysis on this issue for the purposes of this paper. However I note that one of the implications of some earlier work I carried out is that it may be difficult to detect any improvement in overall dispatch efficiency resulting from the Snowy trial. In that earlier work it was shown that, at least in the case of northerly flows (which is the predominant flow direction) the dispatch of Tumut was not manifestly inefficient. I showed that the dispatch at Tumut would, in fact, be efficient if Snowy Hydro purchased a certain quantity of the inter-regional settlement residues between Snowy and NSW.[38] To the extent that there was little discernible inefficiency in dispatch

[36] CRA in their "Theory Paper" notes that "The most obvious implementation of a pricing regime would use a reference point of 0 MW generation, thus implicitly assuming that, until they generate, participants have neither rights, nor obligations" (page 27).

[37] In an earlier paper (Biggar, Darryl, "Understanding Constraint Support Pricing / Constraint Support Contracts", November 2004) I argued that provided each generator was given an entitlement equal to its output under the current arrangements, no generator will be left worse off than under the status quo. Therefore, in principle, no generator would have an incentive to reject the implementation of this package. However, of course, having received the entitlements, some generators would then have an incentive to oppose a move to a region division.

[38] See Biggar, Darryl, (2005), "The Consequences of a Move to Nodal Pricing at the Tumut Nodes in the Snowy Region".

before the trial, it may be hard to detect any improvement in dispatch efficiency as a result of the trial (there still could be other benefits, such as improvements in the incentive to invest at Tumut).

95. Let's turn now to look at the second basic problem identified earlier. What is the range of potential solutions to this problem?

96. The AEMC raises a number of possible options for improving the firmness of settlement residues.[39] These options include requiring NEMMCO to offer firmer settlement residue products or requiring TNSPs to take on some responsibility for ensuring firmness of settlement residues. However, I hope it is clear from the discussion here and in the appendix that the problem does not lie with the way the existing IRSRs are packaged, or primarily with any failure of reliability on the part of TNSPs. The problem primarily lies with the existing definition of settlement residues which yield a non-firm instrument under certain network outcomes.

97. The most obvious solution, therefore, is a redefinition of the nature of the settlement residues.[40] One possible approach is linking settlement residues to individual constraints and not to interconnectors. Transmission pricing theory shows that the total settlement residues associated with any one constraint (a) can be used to yield a firm hedge (strictly speaking, as firm as the underlying physical network) and (b) will always be positive. Therefore, if settlement residues are linked to specific constraints they will always be positive and could be used as the basis of a "firm" hedge.

98. The box below briefly outlines the "constraint-based residues" approach. New constraint-based residues could be created (and auctioned) over time, with no need to immediately drop the existing IRSRs. These constraint-based residues would in principle apply equally for intra-regional and inter-regional constraints and in radial or meshed networks.

99. In any case, it is essential that the NEM make available a financial instrument which correctly facilitates "firm"[41] hedging across regions. At this stage it appears that the primary options are: a move to full firm financial transmission rights (FTRs)[42], the establishment of CSCs, and a move to constraint-based residues.

100. At this stage, without the benefit of further detailed analysis, it seems to me that the best way forward for the NEM would be an immediate adoption of locational pricing for generators coupled with the auctioning of any constraint-based residues (whether intra-regional or inter-regional) that arise.[43] This approach would have two clear advantages:

- First, this approach would resolve the problem of congestion management in a simple manner for the long-term, leaving only the question of region boundary division to be determined over time.

---

[39] Section 5.3.2.

[40] It is also known that the CSC mechanism yields, in some cases, a firmer inter-regional settlement residue. I am not sure the extent to which this is a general result. This is a good question for further research.

[41] As always, as firm as the capability of the underlying network.

[42] Firm financial transmission rights are rights to a stream of payments related to point-to-point price differences. A key issue with FTRs is whether or not the total sum of all FTRs between market participants is "simultaneously feasible". To assess this essentially requires the system operator to run a second market alongside the spot market under which the prices of FTRs are determined by the physical state of the network at any point in time.

[43] These policies may also need to be coupled with policies to control market power on some generators.

- Second, this approach is the natural extension of the existing market arrangements. In fact, to the extent that the assumptions underlying the original market arrangements (that is, no constrained generation and no loop flows between regions) continue to hold this proposed approach would correspond perfectly with the current market arrangements – that is, there would be a uniform price in each region and the constraint-based residues would be identical to the existing inter-regional settlement residues. However, to the extent that constrained generation becomes more of a problem over time, or to the extent that the NEM becomes more "meshed", the proposed approach would allow a natural and long-term solution to the problem of congestion management, no matter how significant these problems become in the future.

Box 5: What are constraint-based residues?

Constraint-based residues are an alternative way of dividing up the total surplus into streams of residues. As already noted, this approach is equivalent to the existing inter-regional settlement residues when there are only inter-regional constraints (and no constrained generators).

In brief, constraint-based residues would operate as follows: There would be a separate stream of residues or "fund" for each possible binding constraint. Let's suppose that a constraint takes the following generic form: the left-hand side of the constraint is simply a linear combination of the output of different generators and the right-hand-side is a simple fixed limit.

In this case the total surplus accruing to the system operator when that constraint binds is simply the "marginal value" of the constraint times the limit. This amount would then be paid into the residue "fund" for that constraint. Notice that this amount is fully "firm" (that is, as firm as the physical capability of the network) and is always positive.

Market participants could then create their own "firm" hedges by purchasing a share of that stream of residues. The proportion that any one generator would choose to purchase would depend on the extent to which its own output affects the constraint – this depends in turn on the coefficient on that generator in the constraint equation. This coefficient is easily determined from public information, so each generator can determine precisely how much of each constraint-based residue stream it needs in order to obtain a firm hedge. The theory of constraint-based residues is explained further in Biggar (2005).[44]

The Way Forward

101. The AEMC Issues Paper touches on a very large number of issues, many of which have already been the subject of extensive lengthy debate. Any specific policy development arising from this issues paper will likely require extensive further consultation and debate in its own right. Rather than focus on specific policies at this stage it seems more appropriate to lock in key principles which will guide future policy-making in this area.

102. I suggest the following foundational principles as a guide to future policy development in this area:

(a) The principle that two problems usually require two policy solutions;

(b) The principle that given the price that a generator is paid, it should be dispatched for a quantity at which it is willing to produce given that price;

[44] Biggar (2005), "Managing Negative Settlement Residues on the VIC-SNOWY interconnector", May 2005

(c) The principle that going forward, settlement residues should be defined in such a way as to allow the writing of hedges which are as firm as the underlying physical network between any two locations with different prices;

(d) The principle that while geographically differentiating prices to (at least large) consumers is preferable, jurisdictions should be allowed to maintain geographically averaged prices for (small) consumers if they wish.

(e) The principle that although regions may be divided, no regions shall be merged.

(f) The principle that whatever mechanism is chosen for allocating or auctioning rights to the trading surplus, no entitlement or allocation should be granted to a new generator.

Two problems usually require two policy solutions

103. Under present arrangements, one market distortion may be masking or offsetting the impact of another. In this case, fear of worsening the latter distortion may lead to resistance to correcting the former distortion – a case of two wrongs making a right. One policy instrument should not be made to serve two ends. I suggest that the AEMC adopt the policy that if there are two underlying problems they should be tackled separately with two different policy solutions. For example, the current arrangements for managing congestion could be masking episodes of generator market power. In this case, improvements in the policies for handling congestion may need to be accompanied by explicit policies for controlling generator market power.

Consistency between pricing and dispatch

104. I have argued many times in this paper that there should be consistency between pricing and dispatch – in other words, the combination of the price a generator is paid and the quantity it is dispatched, should be a price-quantity combination on the generator's offer curve. In the absence of this principle, constrained generators have no incentive to submit a bid which reflects their true marginal cost. Correcting this problem lies at the heart of solving the problem of inefficient dispatch. This is not the only policy action necessary to solve the problem of inefficient dispatch (it may also be necessary to improve competition between generators or otherwise control generator market power), but it is a critical first step and an important key principle going forward.

"Firm" inter-regional hedging instruments

105. The present definition of settlement residues yields a non-firm instrument for hedging in a range of common market outcomes. Facilitating inter-regional trading and preventing negative settlement residues are primary objectives for any regime for improving congestion management. The settlement residues should be defined in such a way as to be as "firm" as the capability of the underlying physical network. This implies, as a corollary, that these streams of residues would be positive. This would eliminate the need for ad hoc intervention by NEMMCO in the market to limit counter-price flows.

Geographically-differentiated prices for consumers

101. In principle, it is preferable for electricity consumers – especially large consumers – to also face a geographically differentiated price. Geographic differentiation of pricing of electricity for consumers enhances demand-side responsiveness (reducing the need for further generation and transmission investment) and improves the locational decisions of consumers while also reducing market power. However, these effects are currently limited due to the limited responsiveness of small consumers to the spot market price of electricity. If some jurisdictions have other objectives which require geographic averaging of electricity prices to consumers they should be able to pursue this, especially for smaller consumers.

Region division, not region boundary change

106. Region boundary change has been a highly contentious issue in the NEM. For any given region boundary change it is exceedingly difficult to work out the full implications without detailed and potentially contentious market modeling. However, in a competitive market a move to finer geographic differentiation of prices can only improve welfare[45]. Therefore, I suggest that the AEMC adopt the principle that there should be no reduction in the degree of geographic differentiation of prices over time. Specifically, there should be no mergers of existing regions or groups of nodes as long as there remains a possibility that it would be efficient to price those regions or groups of nodes separately under some outcomes of the market.

107. Under this rule, existing regions could be divided but regions or nodes which are currently separately priced could not be brought together. This approach would almost certainly rule out some of the current proposals for region boundary changes in the Snowy region, which involve merging the Snowy region with parts of VIC or NSW. Under this principle, the Snowy region could be divided into new regions, but those regions could not then be merged with NSW or VIC.[46]

108. This approach is consistent with the longstanding view of the ACCC that the NEM should be moving towards full locational pricing. In my view, we should not talk about a "region boundary change" but a "region division".

No allocation of rights to new generators

109. If there is to be any free allocation of rights to the trading surplus (such as a grandfathering of CSCs), such an allocation should not be granted automatically to new generators. Just as the current policy of implicitly linking the CSC allocation to the dispatch target creates inefficient incentives for generators, linking a possible CSC allocation to *any* action on the part of the generator will also distort generator incentives (whether that action is a location, expansion, contraction, or fuel choice decision). In effect, the AEMC should announce a policy that whatever congestion management regime is put in place no rights or entitlements will be created for generation capacity which is not already in place or "committed".

110. The proposal I put forward earlier (a move to full locational pricing for generators coupled with auctioning of constraint-based residues) would satisfy these principles.

[45] Putting to one side the adjustment costs of market participants as they adjust their trading systems and hedging arrangements to the new prices.
[46] As long as there is some prospect that it would be efficient to price those new regions separately in some state of the market.

Appendix:

111. The purpose of this appendix is to illustrate the key points in the text using a series of simple network diagrams. Specifically, these examples show that:

(a) The current approach to defining settlement residues leads to settlement residues which are not as "firm" as the underlying physical network and which may be negative. These settlement residues are therefore of reduced usefulness as a tool for hedging. This arises even under fully efficient dispatch in a network with loop flow; and even under fully efficient dispatch in a network with regional pricing when there is constrained generation, with or without loop flow. The inefficiency in dispatch that arises when constrained generators do not bid their true cost makes this problem worse.

(b) The current approach to pricing constrained generation induces these generators to inefficiently bid in a way which does not reveal their true cost, reducing the efficiency of dispatch.

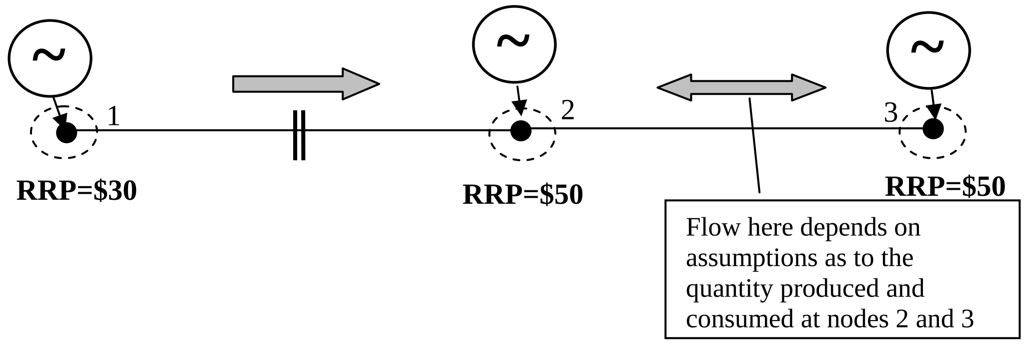
112. All of the examples that follow use a simple three-node network. In the first set of examples these three nodes are connected in a line, in a simple linear or "radial" network. In the second set, I consider a simple "meshed" network – that is, a network with "loop flow".

Scenario A: Linear network, full locational pricing

113. Let's assume first that we have a simple linear network with three nodes and each of these nodes is in its own region, with no intra-regional constraints. We can use this outcome as a benchmark for comparison with the outcomes under regional pricing with and without policies for managing congestion.

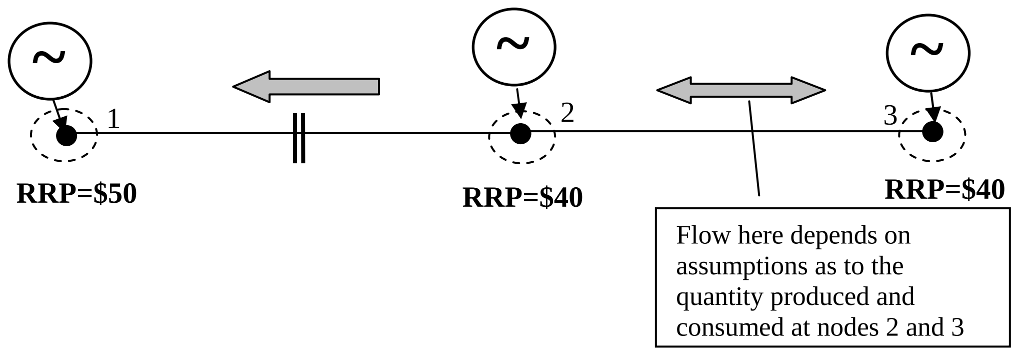
114. As the diagram below shows, let's assume that the configuration of generation and load across the nodes is such that the flow on the line between node 1 and node 2 has reached its limit in the direction of node 2. This leads to price separation between node 1 and nodes 2 and 3. Since there are no constraints between nodes 2 and 3 they both share the same price. For concreteness, let's say that the regional reference price at node 1 is $30 and the regional reference price at nodes 2 and 3 is $50.

115. The flow on the line from node 2 to node 3 cannot be determined without further assumptions. The flow on the line from node 2 to node 3 depends on whether or not there is a net export or import of electricity from node 3. We could say that node 3 is a "generation centre" if there is a net export of power from node 3. Similarly, we could say that node 3 is a "load centre" if there is a net import of power into node 3. Whether or not node 3 is a generation centre or a load centre will depend on the configuration of generation and load at node 2 and node 3. If, at the price of $50, the supply and demand for electricity at node 3 just balance, node 3 will be neither a generation centre or a load centre and the flow on the line from node 2 to node 3 will be zero.



116. There are two potential inter-regional settlement residues (IRSRs) – the residues from the interconnector between node 1 and node 2 and the residues from the interconnector between node 2 and node 3. In this scenario the residues between node 2 and node 3 are zero (as we would expect since this interconnector is assumed to be not operating at its physical limit) and the residues between node 1 and node 2 are equal to the price difference ($20) times the flow on the line from node 1 to node 2. In both instances, these inter-regional settlement residues are fully "firm" and can be used to perfectly hedge the risk of trading between any pair of the nodes.

117. For completeness, let's also consider a network where the flow is in the opposite direction over the constrained link, as illustrated below. As before, the flow on the line from node 2 to node 3 depends on assumptions about the configuration of generation and load at nodes 2 and 3. As before, provided there is adequate competition at each node each generator bids its true marginal cost and the dispatch is fully efficient. The inter-regional settlement residues are fully firm.



118. The following table summarises these results for this first scenario:

Case:	**Scenario A**
Network structure:	Linear, three nodes.
Congestion management arrangements:	• Separate pricing in each region (with no intra-regional constraints – i.e., full locational marginal pricing) • Inter-regional settlement residues defined as the price difference times the flow between regions.
Dispatch efficiency:	Fully efficient dispatch
Firmness of settlement residues / negative residues	• Inter-regional residues fully firm; negative settlement residues on individual streams will not arise. • Overall surplus is positive;

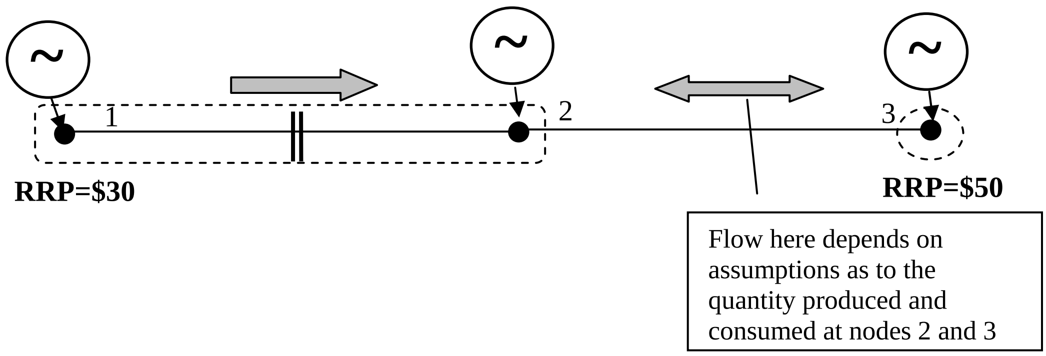
Scenario B: Linear network, regional pricing, dispatch inefficiency corrected

119. Now let's assume that nodes 1 and 2 are placed in the same region, so that the constraint between nodes 1 and 2 becomes an intra-regional constraint. Let's assume that node 1 is the regional reference node.

120. In this scenario I will assume that we have a mechanism which correctly prices the generation at node 2 so that the generators at node 2 have an incentive to efficiently bid their true marginal cost. In particular I will assume that we have introduced locational pricing for generators within each region. As a result, since the local price for generation at node 2 is $50 (as we saw in the previous scenario), generators at node 2 are paid a price of $50 for their output, even though the regional reference price is $30.

121. A mechanism of this kind gives rise to a trading surplus. As I have argued in the text, this surplus needs to be made available to market participants to allow them to hedge the risk of trading across the intra-regional constraint. I will assume that this intra-regional trading surplus is made available to market participants in the form of an *intra*-regional settlement residue defined as the price difference between node 2 and node 1 times the flow between node 1 and node 2.

122. As before, the flow on the line from node 2 to node 3 depends on factors such as the capacity of the generation at node 2 relative to the load at node 2. We can make this power flow in any direction depending on the assumptions we choose.



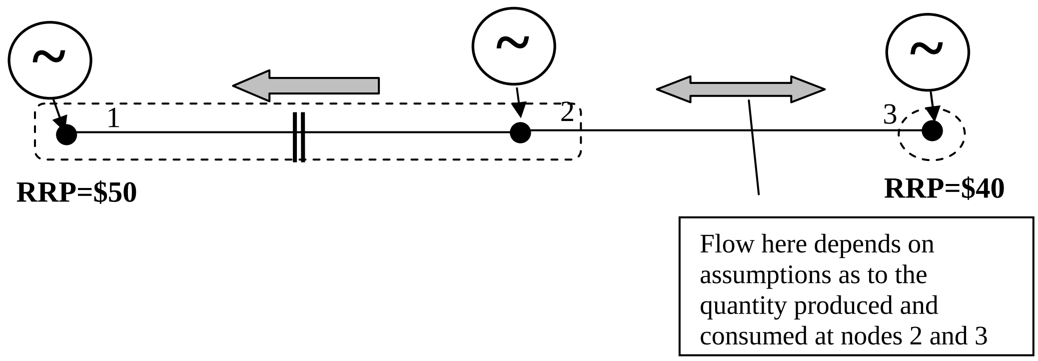
123. As before, the dispatch in this example is fully efficient. Let's explore therefore, the usefulness of the settlement residues as a hedging instrument. Consider first the inter-regional settlement residues. The inter-regional settlement residues are equal to the price difference between node 3 and node 1 times the flow between node 2 and node 3. It is immediately apparent that this residue is far from "firm". This residue may be positive, zero or even negative depending on the flow between node 2 and node 3 (recall that the flow between node 2 and node 3 depends on demand and supply factors at nodes 2 and 3 which have nothing to do with the constraint between nodes 1 and 2).

124. For example, suppose that the configuration of generation and load at nodes 2 and 3 is such that node 3 is a net generation centre. Let's assume that node 3 produces 100 MW more electricity than it consumes. This leads to a flow on the line between nodes 2 and 3 of 100 MW, in the direction of node 2. This counter-price flow results in the accumulation of ($50-$30) times 100 = $2000 of negative settlement residues each hour.

125. What about the intra-regional settlement residues? Since the intra-regional settlement residues in this network are zero when the flow between node 1 and 2 is not at its limit; and are equal to the price difference times the flow when the flow between node 1 and node 2 is at its limit, these intra-regional settlement residues are a fully firm instrument for hedging transactions between nodes 1 and 2.

126. It is worth noting that even though there is no inefficiency in dispatch in this example, there is no guarantee that the total settlement residues accruing the system operator are positive. To see this, note that the total surplus is equal to the price difference on the constrained line times the flow (which is always positive) less the “implicit subsidy” to consumers located at node 2, which is equal to the price difference times the demand at node 2. If the demand at node 2 exceeds the flow on the line from node 1 to node 2, the total surplus will be negative (as noted earlier, this is a very similar issue to the issue of funding of network support agreements).

127. Very similar results arise when the flow on line between nodes 1 and 2 is at its limit in the opposite direction. As before, the flow on the line from node 2 to node 3 depends on assumptions about the demand and supply at nodes 2 and 3. If node 3 is a load centre then the flow on the line from node 2 to node 3 must be in the direction of node 3 – even though this flow appears to be counter-price. Obviously, by changing these assumptions we could make the flow on the line from node 2 to node 3 in the direction of node 3 zero, or in the direction of node 2. It is clear from this example that the inter-regional settlement residues are not at all firm and may be negative. In this case, however, the total surplus is increased by the “implicit tax” on the consumers at node 2, so in this case the total surplus is always positive.



128. The following table summarises these results for this second scenario:

Case:	**Scenario B**
Network structure:	Linear, three nodes.
Congestion management arrangements:	• Separate pricing in each region; • Inter-regional settlement residues defined as the price difference times the flow between regions; • Intra-regional pricing arrangement to correct the incentive for inefficient bidding at node 2 • Intra-regional settlement residues defined as the price difference between the constrained node and the regional reference node times the flow.
Dispatch efficiency:	Fully efficient dispatch
Firmness of settlement residues / negative residues	• Inter-regional settlement residue not at all firm; negative inter-regional settlement residues may arise; • Intra-regional settlement residue is firm; • The overall residues accruing the system operator are positive when generation at node 2 is constrained off and may or may not be positive when generation at node 2 is constrained on.

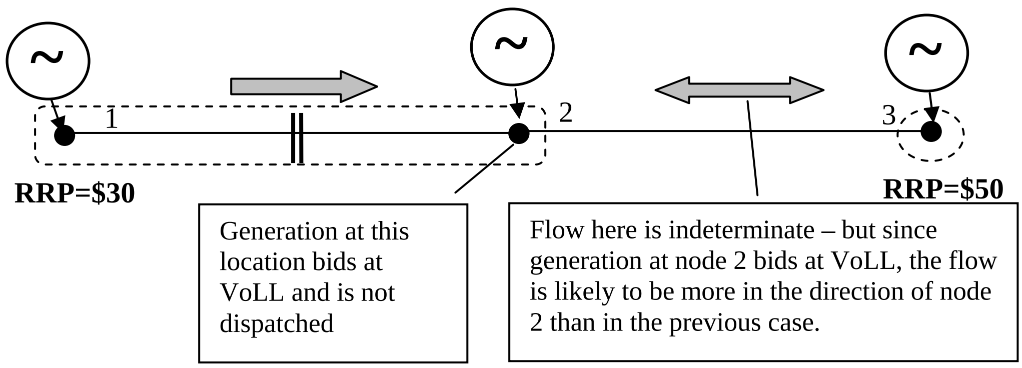
Scenario C: Linear network, regional pricing, dispatch inefficiency not corrected

129. Now let's assume that there is no mechanism for correctly pricing the generation at node 2, so that generation at node 2 is constrained on or off and therefore has no incentive to bid in a way which reflects its marginal cost.

130. Let's assume first that the constraint is in the direction of node 2. Let's assume that all the generation at node 2 has a marginal cost above the regional reference price of $30 but below the local price of $50. In this case, generation at node 2 (which is only paid the RRP of $30) will attempt to not be dispatched. It will do this by bidding its output at VoLL.

131. Since the bid of these generators at node 2 is above the efficient local price, these generators will not be dispatched, leading to a reduction in the output at node 2. This is made up by an increase in the output of generators with a higher marginal cost at node 3 (potentially increasing the regional reference price at node 3). Since generation with a marginal cost below $50 has been replaced by generation with marginal cost at or above $50, there is a reduction in the efficient of dispatch.

132. As before, the flow on the line from node 2 to node 3 depends on factors such as the capacity of the generation at node 2 relative to the load at node 2. However, since we have effectively removed the generation at node 2 from the market, in all possible scenarios this flow is likely to be more in the direction of node 2 than before. Put another way, since the regional reference price at node 3 either stays the same or increases, node 3 is more likely to be a net exporter of power than before. In other words, counter-price flows are even more likely in this scenario than when the dispatch inefficiency is corrected.



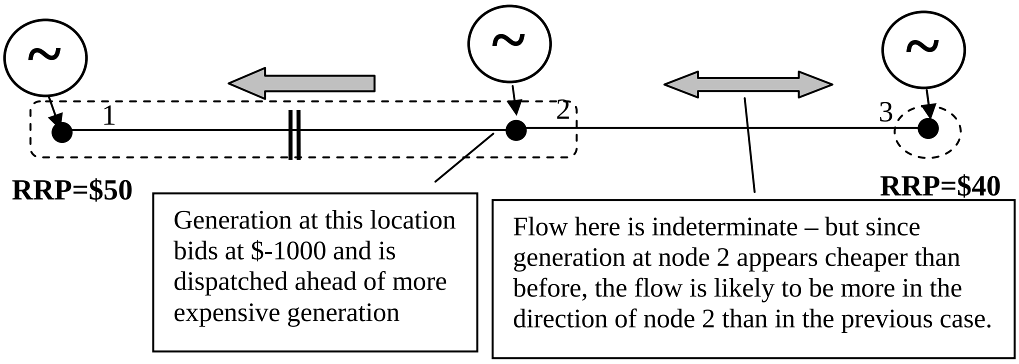
133. As just noted, the dispatch in this network is inefficient; what about the usefulness of the settlement residues as a hedging instrument?

134. As before, the inter-regional settlement residues are not at all firm. In fact, since the flow on the line from node 2 to node 3 is more likely to be in the direction of node 2, negative settlement residues are even more likely to arise than before.

135. Since generation at nodes 1 and 2 in the first region is paid the same price there are no intra-regional settlement residues. Overall the total residues accruing to the system operator can be negative.

136. The opposite pattern results when the flow on the line from node 1 to node 2 is at its limit in the opposite direction. In this case generation at node 2 may be constrained off. As a consequence, it has an incentive to bid its output at $-1000. If generation at node 2 is more expensive than generation at node 3, the impact of this bidding at $-1000 is that the dispatch engine will dispatch generation at node 2 ahead of generation at node 3. Since higher-cost

generation is displacing lower-cost generation, this reduces the efficiency of dispatch. In addition, this increases the flow on the line from node 2 to node 3 in the direction of node 3, increasing the likelihood of counter-price flows.



137. Again, we find that the inter-regional settlement residues are not at all firm and the overall residues may be negative.

138. The following table summarises these results for this third scenario:

Case:	**Scenario C**
Network structure:	Linear, three nodes.
Congestion management arrangements:	• Separate pricing in each region; • Inter-regional settlement residues defined as the price difference times the flow between regions;
Dispatch efficiency:	Potentially inefficient dispatch
Firmness of settlement residues / negative residues	• Inter-regional settlement residue not at all firm; negative inter-regional settlement residues may arise (and are more likely to arise than in scenario B); • No intra-regional residues exist. • The overall residues are equal to the IRSR and are more likely to be negative than in scenario B.

Scenario D: Meshed network, locational pricing

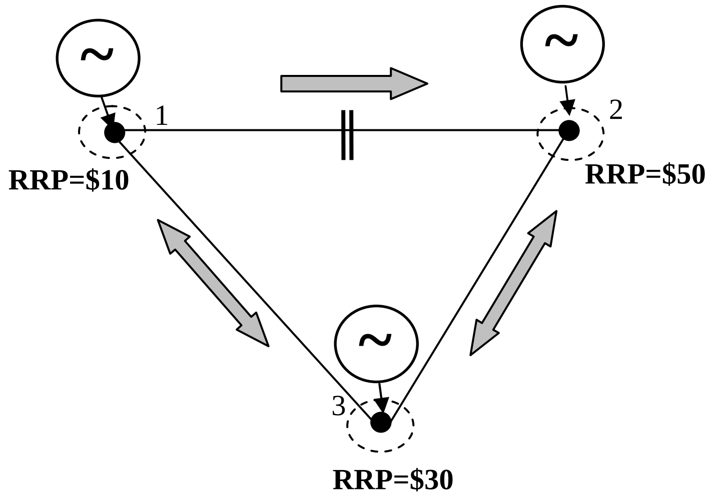
139. Now let's examine the outcomes in a looped network with three nodes, as illustrated below. Again, we will start by assuming that each of these nodes is in its own region, with no intra-regional constraints. As before, this outcome can be compared with the outcomes under regional pricing with and without policies for managing congestion.

140. Let's first assume, as the diagram below shows, that the flow on the line between node 1 and node 2 has reached its limit in the direction of node 2.

141. If we assume that all three transmission lines in this network have identical electrical characteristics, it follows automatically from the theory of transmission pricing that the correct locational price for electricity at node 3 (which is opposite the constrained line) must equal the average of the price for electricity at nodes 1 and 2 (which are at each end of the constrained line). In this example, let's assume that the local price at node 2 is $50 and the price at node 3 is $30. Therefore the price at node 1 must be $10. The resulting dispatch is fully efficient.

142. As before, the flow on the unconstrained lines cannot be determined without further assumptions. It turns out that if the net production at node 2 exceeds the net production at

node 3 (for example, if node 2 is a major generation centre and node 3 is a load centre) then the flow from node 2 to node 3 will be in the direction of node 3 – which is the counter-price direction.



143. In this example the dispatch is efficient. There are, however, significant problems with the use of inter-regional settlement residues as a hedging device. The inter-regional settlement residues between nodes 1 and 2 is equal to the price difference times the flow between node 1 and 2. This is fully firm, as required. However, the inter-regional settlement residues between nodes 1 and 3 and nodes 2 and 3 are not at all firm – the flows in each case are indeterminate and depend on other factors which are independent of the constraint on the line from node 1 to node 2. In fact, as just mentioned, under certain assumptions the flows on the line from node 2 to node 3 can easily be counter-price – and, in fact, will *always* be counter-price if node 2 is a generation centre and node 3 is a load centre.

144. Although the individual settlement residues may be negative, in this network the overall surplus is always positive.

145. The following table summarises these results for this fourth scenario:

Case:	**Scenario D**
Network structure:	Meshed, three nodes.
Congestion management arrangements:	• Separate regions with no intra-regional constraints (full locational marginal pricing); • Inter-regional settlement residues defined as the price difference times the flow between regions;
Dispatch efficiency:	Fully efficient dispatch
Firmness of settlement residues / negative residues	• Inter-regional residues are only firm on the constrained inter-connector. *Other inter-regional residues not at all firm*. Negative settlement residues may arise (and in some cases will always arise); • Overall surplus is positive.

Scenario E: Meshed network, regional pricing, dispatch inefficiency corrected

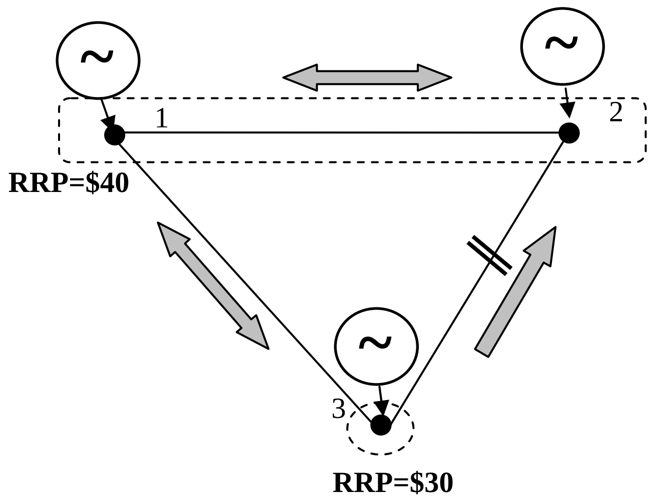
146. Now let's take the network of the previous scenario and assume that nodes 1 and 2 are placed in the same region. There are now two regions in our network. Let's assume that node 1 is the regional reference node for the first region. The "interconnector" between the

two regions is now defined to be the sum of all the flows between the two regions – that is the sum of the flow on the transmission line from node 1 to node 3 plus the flow on the transmission line from node 2 to node 3.

147. Let's assume first that the constrained line is the line between nodes 2 and 3. Notice that there are no intra-regional constraints in this network.

148. As before, the theory of transmission pricing tells us that the price at node 1 (which is also the regional reference price) must be equal to the average of the price at node 3 and the "price" at node 2. Let's assume that the price at nodes 3 and 2 are $30 and $50 respectively, so that the price at node 1 must be $40.

149. In this scenario let's assume that we have introduced some mechanism which correctly prices the (marginal) generation at node 2 so that the generators at node 2 have an incentive to efficiently bid their true marginal cost. As before, such a mechanism gives rise to an intra-regional trading surplus. Let's assume that we make this available to market participants in the form of intra-regional settlement residues equal to the price difference between nodes 1 and 2 times the flow between nodes 1 and 2.

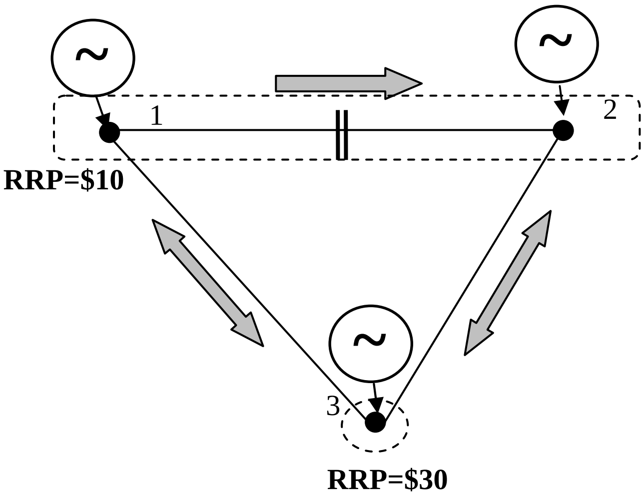


150. The dispatch in this example is fully efficient. Let's explore therefore, the usefulness of the settlement residues as a hedging instrument. Consider first the use of the inter-regional settlement residues. The inter-regional settlement residues are equal to the price difference between node 3 and node 1 times the sum of the flow between node 1 and node 3 and between 2 and node 3. Since the sum of all the injections at every node must be zero, the sum of the flow between node 1 and node 3 and between node 2 and node 3 is precisely equal to the net withdrawal at node 3. In other words the net flow between the regions depends (as before) on whether node 3 is a load centre or a generation centre.

151. It is immediately apparent that this inter-regional residue is far from "firm". Even though the flow on the transmission line from node 3 to node 2 is at its capacity limit, the flow on the line from node 1 to node 3 is indeterminate. As a result, the inter-regional settlement residue may be positive, zero or even negative depending on the net injection or withdrawal at node 3. For example, suppose that there is a net withdrawal of electricity at node 3 (i.e., node 3 is a load centre). In this case, the flow between the regions will be in the direction of node 3, giving rise to counter-price flows and negative inter-regional settlement residues.

152. In addition, in this example, the intra-regional residue will also not be firm. There is a price difference between node 1 and node 2, but the flow between these nodes is indeterminate (and, indeed, could be also counter-price). As before, the total surplus is equal to the price difference times the flow on the constrained line less the "implicit subsidy" to consumers at node 2. The total surplus may be negative if the load at node 2 is large relative to the flow on the constrained line.

153. Now let's consider the case of an intra-regional constraint in a looped network. Let's assume now that the constrained line is the line between node 1 and node 2. Again, let's assume that we have some mechanism for correcting the price at node 2.



154. As before, the dispatch is fully efficient, but there are problems with the settlement residues. As before, the net flow between the regions is equal to the net injection/withdrawal at node 3. Since this is indeterminate the inter-regional settlement residues are non-firm. However, the intra-regional settlement residues in this example are fully firm.

155. The following table summarises these results for this fifth scenario:

Case:	**Scenario E**
Network structure:	Looped, three nodes.
Congestion management arrangements:	• Separate pricing in each region; • Inter-regional settlement residues defined as the price difference times the flow between regions; • Intra-regional pricing arrangement to correct the incentive for inefficient bidding at node 2 (note that this is required whether the binding constraint is an inter-regional or intra-regional constraint). • Intra-regional settlement residues defined as the price difference between the constrained node and the regional reference node times the flow. (Note that this gives rise to a non-zero intra-regional settlement residue whether the binding constraint is an inter-regional or intra-regional constraint)
Dispatch efficiency:	Fully efficient dispatch
Firmness of settlement residues / negative residues	• Inter-regional settlement residue not at all firm; negative inter-regional settlement residues may arise; • Intra-regional settlement residue may or may not be firm depending on whether the constraint is intra-regional or inter-regional. • Total surplus may be negative in the case of constrained-on generation.

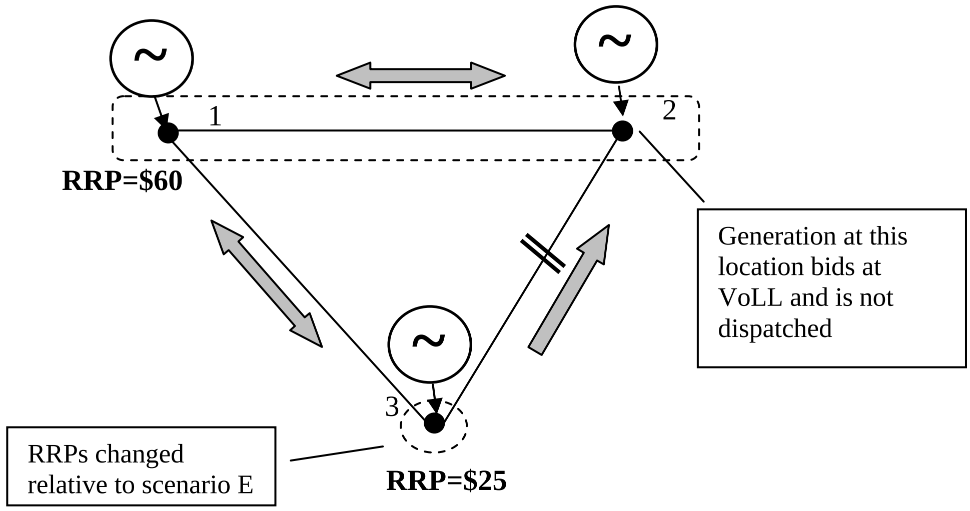
Scenario F: Looped network, regional pricing, dispatch inefficiency not corrected

156. For the final scenario let's take the previous scenario and assume that we have not corrected the dispatch inefficiency that results from regional pricing with an inter-regional or intra-regional constraint.

157. As before, let's assume first that the constrained line is the line between nodes 2 and 3, with the flow at the limit in the direction of node 2. As before, the price at node 1 (which is also the regional reference price) must be equal to the average of the price at node 3 and the "price" at node 2.

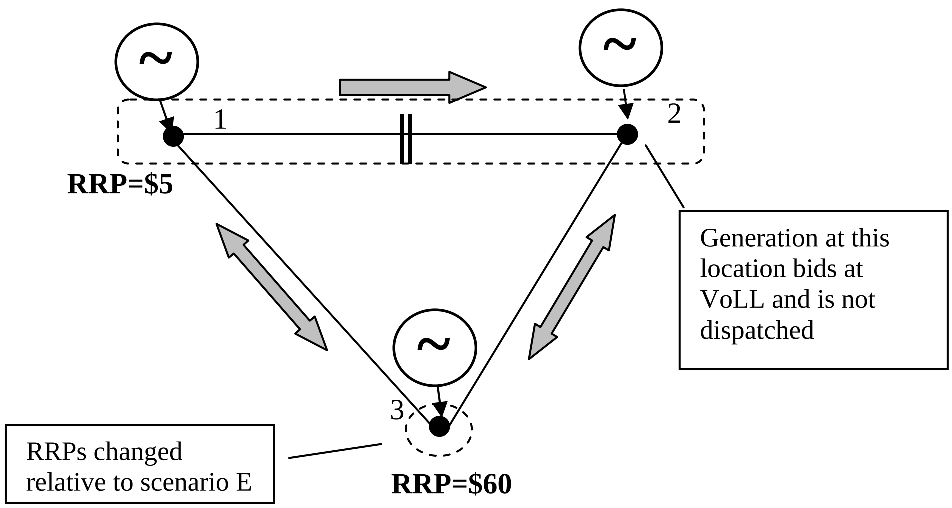
158. Under these assumptions, since the generation at node 2 is only paid the regional reference price, it is constrained on. It therefore has an incentive to submit a bid equal to VoLL, effectively removing this generation from the market. The reduction in output of this generation must be made up by an increase in the output of generation at node 1 (and possibly a reduction in the output of generation at node 3). This will likely increase the marginal cost of the last units dispatched at node 1, increasing the regional reference price and reducing the efficiency of dispatch (relative to the efficient dispatch as in scenario E).

159. This example shows how inefficient dispatch can arise under the current arrangements for congestion management, *even in the absence of intra-regional constraints*.



160. As in the previous scenario, the inter-regional settlement residues are not firm and may be negative. If fact, since the effect of the inefficient bidding is to raise the regional reference price in the first region and to lower the regional reference price at node 3, node 3 is more likely to be a load centre and therefore the flows on the interconnector are more likely to be in the direction of node 3 – increasing the risk of inter-regional settlement residues. There are no intra-regional settlement residues for hedging intra-regional dispatch risk in this scenario.

161. Now let's consider the case of an intra-regional constraint in a looped network. Let's assume now that the constrained line is the line between node 1 and node 2, with the flow in the direction of node 2. As before, the generation at node 2 is constrained on. As before, the generators at node 2 have an incentive to bid at VoLL. This effectively removes these generators from the market – their output is replaced by increased generation at node 3. For the reasons given above this can lead to inefficiency in dispatch.



162. Since the regional reference price at node 3 is increased, node 3 is more likely to be a net exporter, so the flow on the interconnector is more likely to be away from node 3 – this increases the risk of negative settlement residues.

163. Of course, it is also possible to configure these examples so that the remote intra-regional generation is constrained off. As before, this leads to an inefficient in dispatch and the inter-regional settlement residues are non-firm.

164. The following table summarises these results for this sixth scenario:

Case:	**Scenario F**
Network structure:	Looped, three nodes.
Congestion management arrangements:	• Separate pricing in each region; • Inter-regional settlement residues defined as the price difference times the flow between regions;
Dispatch efficiency:	Inefficient dispatch due to inefficient bidding at the remote intra-regional node (node 2) whether the binding is an inter-regional or intra-regional constraint.
Firmness of settlement residues / negative residues	• Inter-regional settlement residue not at all firm; negative inter-regional settlement residues may arise (and are more likely to arise than in scenario E); • Intra-regional settlement residue may or may not be firm depending on whether the constraint is intra-regional or inter-regional.