Draft Report

Locational Signals for New Investment

Prepared for

Electricity Commission

August 2004

covec
expert analysis + practical advice

Covec is an applied economics practice that provides rigorous and independent analysis and advice. We have a reputation for producing high quality work that includes quantitative analysis and strategic insight. Our consultants solve problems arising from policy, legal, strategic, regulatory, market and environmental issues, and we provide advice to a broad range of companies and government agencies.

Covec develops strategies, designs policy, and produces forecasts, reports, expert testimony and training courses. Our commitment to high-quality, objective advice has provided confidence to some of the largest industrial and governmental organisations in Australasia.

Authorship

This document was written by Tim Denne and John Small. For further information email tim@covec.co.nz or phone (09) 916-1960

Disclaimer

covec
expert analysis + practical advice

Covec Limited Level 11 Gen-i tower 66 Wyndham Street
PO Box 3224 Shortland Street Auckland New Zealand
t: (09) 916-1970 f: (09) 916-1971 w: www.covec.co.nz

Contents

Executive Summary

This paper contributes to the Electricity Commission's development of a methodology for transmission pricing, and specifically the need to provide locational signals relating to new generation or load. It addresses:

- whether prices for the shared transmission network can and should be designed in a way to provide locational signals for significant new generation and load;
- the pros and cons of a locational charging approach; and
- whether signals for transmission alternatives are best dealt with in the Grid Investment Test (GIT), rather than in transmission pricing.

Costs Vary With Location

There are fixed and variable costs of transmission:

- The **fixed or embedded costs** are the costs of having a transmission network that meets quality standards, connects generators to load customers, and is of sufficient capacity to carry peak demand.
- The **variable costs** associated with transmission activity, change with the level and location of activity on the network, and include the costs of:
  - **Losses** that vary with the amount transmitted, the length of the connection and the quality of the asset;
  - **Congestion** that results in greater levels of loss and which limits the efficient connection of supply and demand, eg by requiring a longer transmission route; and
  - **Reinforcement or grid expansion** associated with significant increases in injection or off-take, such as because of new connections.

Not All Locational Costs are Signalled

Optimal location of activity that uses the grid could be achieved through signalling long run marginal costs (LRMC) of that activity. Currently there are externalities of congestion that have not been fully internalised, and this prevents LRMC signalling.

Nodal prices internalise the costs of losses and partially internalise the costs of congestion, in that they include the costs of out-of-merit-order dispatch. The ultimate cost of congestion is the cost of unserved energy (ie the value of lost load) and nodal prices can rise to this level only in limited circumstances—when a generator has market power at a congested node. They cannot when load customers are bidding for scarce supply at a congested node.

Our assessment of the current set of locational signals in nodal prices is summarised in the Table below. Generators have an incentive to enter where there is excess load; they can obtain congestion rent when they do so and where, without them, there would be unserved demand, can raise prices to the marginal willingness to pay of demand customers.

Table: Summary of Locational Signals from Nodal Prices

	Generation Entry	Load Entry	Load Reduction
Excess Load	• Generator can price to obtain congestion rent • Efficient locational signals present • Demand reduction response not signalled	• Current prices do not signal congestion • Network externalities of entry—potential for high prices or blackouts	• Current prices do not signal value of load reduction
Excess Generation	• Prices fall to SRMC • Efficient locational signals present	• Prices rise on entry • New load not rewarded	• Prices fall after demand reduction • Reward is limited

The current lack of a locational signal relating to congestion externalities is very largely limited to the demand side—location of new demand customers, demand growth and demand management measures. The key issues are that, post-entry, nodal prices change so that the desirable load is not rewarded, and pre-entry, prices do not rise to reflect marginal willingness to pay for electricity, so that excess demand is not discouraged.

Transmission charges include connection charges, which appear to effectively charge the incremental costs of new entry, and interconnection charges, which spread the costs of the core grid across all load customers; they provide no locational signals.

The Grid Investment Test (GIT) can assess where and when grid reinforcement is the best option to deal with congestion, but is somewhat limited in delivering transmission alternatives— investment in local generation, energy efficiency, demand-side management and distribution network augmentation. Demand side responses may be the least cost option but the market does not signal their value and, if the GIT suggests that demand side management is the least cost option, neither Transpower, nor any other participant, has the chief tool to trigger it—price.

There are distributional issues associated with the results of the GIT also. If grid reinforcement is identified as the least cost option, these costs, at least under current arrangements, are shared with all load customers. However, if the least cost option is new generation, the costs fall to a greater extent on those connected to the congested node(s). These distributional effects may have incentive effects as load customers connected to a crowded node will favour grid reinforcements, whose costs they can share, over generation options whose costs they would bear. This is especially important when the load customer may be the potential provider of the new generation, eg on-site cogen.

Options for Introducing a Locational Signal

It is likely to be easier to signal the cost of unserved energy in the market, or allow it to be signalled, rather than to signal the costs of one of the options that might be chosen to address the congestion problem. The cost of unserved energy will always set the ultimate cap on costs, ie grid reinforcements would never be made if their costs were greater. Thus if the market is able to signal these higher costs, it can incentivise lower cost solutions that will set market prices.

Our assessment is that these issues would be very largely addressed through demand-side bidding.

In the absence of such reforms, a locational element in interconnection charges might be used. However, the long-term signal that is required changes after the solution has been introduced. And specifically, keeping the locational price signal after new generation has been introduced will include double costs while retaining it for a load solution is entirely appropriate.

Locational elements in charges for core grid assets have been used in other countries, and notably in England & Wales and in Australia. But neither of these countries operate nodal pricing, so these charging regimes are addressing a much wider set of externalities. They use grid flow modelling to estimate the marginal impacts of additional injection or offtake at every system node, under a range of input assumptions. The impacts need to be combined with estimates of the costs of grid enhancement for specific parts of the grid or average grid components. The systems are complex and open to considerable debate over input assumptions.

A similar system was employed in New Zealand from 1996 to 1999 but was used to distribute costs, rather than to estimate marginal costs. It was not perceived as meeting its objectives of fairness and was never set up to achieve efficiency.

Options for New Zealand would include a grid flow method, more similar to that used currently in the two international examples, and to develop a scheme that essentially reinforces the nodal prices.

However, our assessment is that such complexity is not justified, given the existence of nodal prices and the limited extent of the locational signalling problem.

Recommendations

Our judgement is that introduction of demand side bidding would largely address the current locational problem—the un-internalised congestion externality. We recognise that there is a wide range of accompanying issues with such a proposal, beyond the scope of this short analysis.

In the absence of demand side bidding, our assessment is consistent with that of Frontier Economics, that the interconnection charges should spread core grid costs across all load customers and that the GIT has an important role in identifying locational signals. The GIT will not do much, in itself, to address this externality problem; it will enable assessment of when grid enhancement is the best option, but can do little to solve the demand side problems in the absence of price signals. This leaves the signal to load largely unaddressed, but the magnitude of the problem does not justify the introduction of a locational element in the interconnection charge.

1. Introduction

This paper contributes to the Electricity Commission's development of a methodology for transmission pricing.

Frontier Economics has provided the Commission with an initial issues paper[1]; It suggests that new load and generation assets should be charged in the same way as for current connected assets, ie that all are charged on the basis of their share of peak demand or peak injections. It also suggests that locational signals can be provided via the Grid Investment Test (GIT).

This paper canvasses some of the issues in more detail, and provides additional discussion on the potential use of locational pricing signals in transmission pricing. It addresses:

- whether prices for the shared transmission network can and should be designed in a way to provide locational signals for significant new generation and load;
- the pros and cons of a locational charging approach; and
- whether signals for transmission alternatives are best dealt with in the Grid Investment Test (GIT), rather than in transmission pricing.

In undertaking this review and analysis, this paper considers:

- the Commission's security of supply obligations under the Government Policy Statement (GPS) and the impact of locational signals on decisions to build new power stations;
- the strategic implications that locational signals might have on other market arrangements and the efficient operation of the market.

---

[1] Frontier Economics (2004) Transmission pricing methodology—Options and guidelines. Final Draft Issues Paper. June 2004.

2. Locational Costs and Locational Signals

This section sets out why location matters from the perspective of economic efficiency. It addresses the current set of market signals faced by new entrants and the extent to which they provide efficient locational signals.

2.1. How Costs Vary with Location

Our concern here is with the nature of the costs that a new entrant generator or load (demand) customer places on the transmission system.

There are fixed and variable costs of transmission:

- The **fixed or embedded costs** are the costs of having a transmission network that meets quality standards, connects generators to load customers, and is of sufficient capacity to carry peak demand.

- The **variable costs** associated with transmission activity, include the costs of:
  - **Losses** that vary with the amount transmitted, the length of the connection and the quality of the asset;
  - **Congestion** that results in greater levels of loss and which limits the efficient connection of supply and demand, eg by requiring a longer transmission route; and
  - **Reinforcement or grid expansion** associated with significant increases in injection or off-take, such as because of new connections.

The variable costs change with the level and location of activity on the network. This is illustrated with reference to a hypothetical grid example below.

Figure 1a shows a stylised grid with load at one end and generation at the other. The shared grid runs between nodes N1 and N2 through to N5 and N6. The link from N3 to N2 has greater capacity than the link to N1, and the link to N5 has greater capacity than that to N6. Both loads and both generators are assumed to be the same size.

Picture then Figure 1b in which a new generator or a new load customer has the option to connect at one of 5 locations—A to E. Other things being equal, the least cost connection point is clear in both cases. For the generator it is point B; here the connection is close to load customers (which minimises losses in transmission) and the connection is to an asset with spare capacity. For the load customer, the least cost connection is to point D; it is close to a generator (and we assume both generators have spare capacity) and again, the connection is to a transmission asset with spare capacity.

Below we assess the extent to which current prices can signal these optimal connection points.

Figure 1a: Stylised Electricity Network

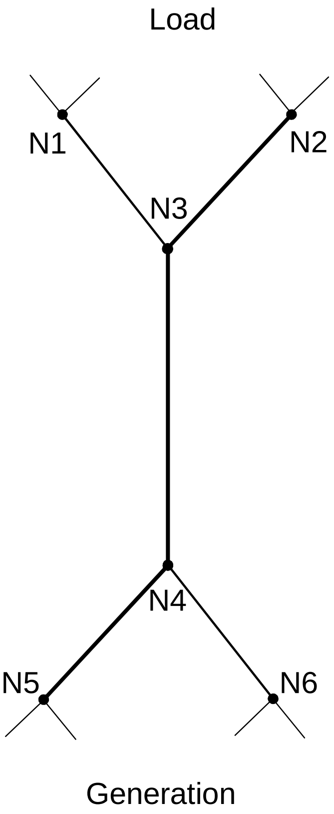
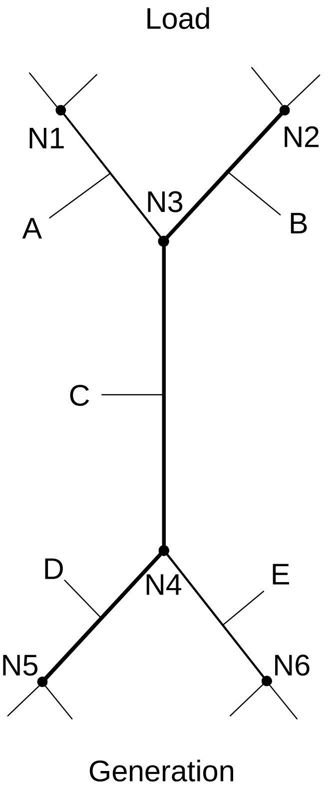


Figure 1b: Potential Connection points



2.2. Locational Signals in Current Prices

Market signals are provided to load and generation customers through transmission prices and via nodal pricing of electricity. Below we examine how these signals operate currently and the extent to which they signal optimal location.

2.2.1. Nodal Prices

Revenues obtained by generators, or prices paid by load customers, are determined in the wholesale market and may be set via a contract or the spot market price. All reflect the marginal costs of supply to some degree.

Nodal prices are determined as a result of supply side bids at different injection points modified by adjustments to take account of the costs of transmission losses and constraints, and given forecasts of demand. This provides price signals to generators and to load customers relating to the ongoing costs of drawing power from their connection at any point in the system. By connecting closer to load, a generator can bid its power in at a higher price and obtain some of the rent available because of the higher costs of transmission from other generators.

Losses

Nodal prices provide locational signals relating to transmission losses. Looking at Figure 1b, current generators might offer in at N5 or N6 at $20/MWh, but the costs of losses mean the potential price at N1 is $22/MWh. A new generator with a marginal cost of generation of $21/MWh might face a price penalty of only $0.2/MWh by locating at point A; however, it can bid in at $21.99/MWh and be dispatched. Losses are fully internalised via the nodal pricing system. The long term nature of this signal depends on the expected time pattern of prices and on changes elsewhere in the grid.

Congestion—Excess Load

Nodal prices give partial signals of the costs of congestion.

Considering excess load first, when part of the grid is congested, power flows through alternative routes to meet demand, and higher-priced generation may be dispatched. If the grid became so congested that demand could not be met, this would be a physical constraint and rolling blackouts would occur. Such congestion would not be priced in the market, ie nodal prices do not increase sufficiently to limit demand. This contrasts with what one might achieve in a more complete market.

In such a hypothetical market, in which the grid operator set prices efficiently, the costs of security were expressed in the market and demand was able to respond in real time, prices would rise to limit demand when congestion occurred. The nodal price would rise to the level of the maximum marginal willingness to pay of the demand customer. Frontier Economics suggests a value for New Zealand of $20,000/MWh[2] although this would vary by node.

If these circumstances were to persist, a new generator could enter and inject at the congested node, could price up to the costs of the alternative generator and, for the periods in which otherwise there would have been supply constraints, up to the load customer's willingness to pay. Generators might also anticipate growing demand and enter early to take advantage of potential congestion rents when demand grew. Further imagining this hypothetical market, the load customer might also have the option of investing in new transmission capacity where this is a lower cost option than the new generation.

This outcome can result to some extent currently. Further, the GIT will aim to reproduce some of these effects. In theory, a generator could come in to a crowded node and price as described above. It might not have complete knowledge of the demand customer's willingness to pay, as would be revealed in a market that was allowed to go to constraint, but it could develop an approximate estimate of that value and discover it later through market bidding. Alternatively, the load customer could build generation in full knowledge of its own willingness to pay.

There are limits to this response. There will be some demand response to persistently higher prices, which the new entrant generator needs to predict. And here there is scope for a stand-off. The level of demand, post entry, may be less than would justify entry. But, because of the absence of demand side bidding, the market cannot signal the price that would result in this reduced demand prior to entry.

---

[2] Frontier economics (2004) Draft Grid Investment Test. Final Draft Discussion Paper.

Poorly defined property rights are likely to restrict private investment in transmission enhancement; here the GIT has its role in estimating when such investments are optimal, and in ensuring that transparent processes can constrain inappropriate investments in costly new generation while not disincentivising those that were socially optimal.

Thus, while not providing a complete solution in all cases, the existing nodal pricing system combined with a well functioning GIT do provide (at least qualitatively) appropriate signals to potential investors in new generators to locate plant efficiently.

Rather than generation entry, if we imagine a new load customer connecting to a point that is already congested, or an existing load customer expanding its activities (or even gradually expanding demand across many customers), the situation is rather different. Such additional load faces potentially high costs, either from outages or from new entrant generation, as discussed above. But there are also network externalities, as these costs are also passed on to other load customers.

More importantly, because there is no demand side bidding, existing load customers cannot signal their willingness to pay for scarce electricity at that node in a way that provides potential new load with an appropriate price signal. Such a signal would allow a potential new load to assess whether it values electricity more highly than incumbents. For example, suppose that incumbent loads had a marginal willingness to pay for electricity of $20,000/MWh, but a new entrant load would be willing to take an interruptable contract and had a marginal willingness to pay of only $200/MWh. The new load might well enter in a way that led to short term blackouts. New generation or new grid reinforcement under the GIT might enter, but the lower cost option might be to discourage this load or make it interruptible.

A similar situation applies in respect of the incentives for load reduction at a crowded node. This is not fully signalled by the current system of nodal prices because of the externalities involved. Individual customers considering load reductions do not factor into their decisions the benefits such reductions would bring to other load customers.

In principle, where there is excess load, the existing market arrangements can signal the value of new generation, but it fails to signal the costs of additional load or the value of reduced load.

Congestion—Excess Generation

Grid congestion can also occur as a result of excess generation at or near a node. In this case generators are constrained from supplying demand because of physical constraints in the system, nodal prices fall to the short run marginal costs (SRMC) of supplying the constrained demand.

If a new load customer is to enter, because there is no demand side bidding in the current market, they will be constrained from obtaining this low price (or even a price just above current SRMC). They receive the new nodal price that results after their load is added to the system below the constraint (which will be higher than the pre-entry price). This problem might be overcome or mitigated through contracting, but the market would tend to reward the generators that did not contract.

The costs of the load customer not locating close to a surplus of generation are that it purchases electricity at a higher price elsewhere and places further demand on the grid.

If a new generator enters an area already constrained through excess generation, they will perform in a merit order against other generators below the constraint. They have the correct market signals to discourage entry.

As for excess load, the market is likely to give efficient signals discouraging new generation but the absence of demand side bidding limits the signals to new load customers.

Table 1 summarises the impacts described above.

Table 1: Summary of Locational Signals from Nodal Prices

	Generation Entry	Load Entry	Load Reduction
Excess Load	• Generator can price to obtain congestion rent • Efficient locational signals present • Demand reduction response not signalled	• Current prices do not signal congestion • Network externalities of entry—potential for high prices or blackouts	• Current prices do not signal value of load reduction
Excess Generation	• Prices fall to SRMC • Efficient locational signals present	• Prices rise on entry • New load not rewarded	• Prices fall after demand reduction • Reward is limited

The externality problem is very largely limited to the demand side.

2.2.2. Transmission Prices

Entrant **generators** must pay the costs of connection to the core grid. This includes[3]:

- **Connection charges**—these are charged to all connected customers (load and generation) and cover the costs of connection to the core grid assets. Where connection assets are shared by more than one customer, the costs are shared in proportion to their peak injection (or peak offtake for load customers). Connection charges are calculated annually and are specific to the connection; the total cost recovered in the year to 31 March 2004 is $99.04 million.

- An **HVDC charge** if they are located in the South Island. It is used to recover the annual costs of the inter-island link between Benmore and Haywards. It is recovered as a flat rate ($/kW) applied to all electricity injected into the grid in the

---

[3] Transpower New Zealand Limited (2002) Pricing for Grid Connection Services.

South Island. In the year to 31 March 2004 the rate is \$19.08/kW, to recover \$63.2 million.

Load customers are charged for the costs of the core grid (HVAC assets). It is recovered by a flat rate **interconnection charge** (\$/kW) applied to all electricity taken from the grid. In the year to 31 March 2004 the rate is \$50.57/kW to recover \$310.9 million.

Because the revenue requirements used to determine charges are estimates, actual revenue requirements may differ. There is an **economic value adjustment charge** used to adjust the total revenue earned from customers. The economic value adjustment is calculated annually and allocated to all customers in proportion to their connection, interconnection and HVDC charges.

For connection charges, in general the new entrant pays for the incremental costs of connection, wherever the connection is. However, if there are shared assets, the costs are shared and the efficiency impact of the change will differ between short and long run considerations. In the short run, if there is no spare capacity and thus additional costs are incurred, incumbents subsidise the connection costs of the new entrant. If there is spare capacity, the new entrant must share the costs currently paid by incumbents. This could make a new entrant indifferent between connecting to an asset with spare capacity and constructing a new asset.

Efficiency gains of this approach may be greater over the long run; building ahead can enable the network owner to achieve economies of scale. Charging entrants for a proportion of the total costs of an asset should result in final costs to these entrants being lower than in parts of the network with lower capacities.

Connection charges tend to separate connection assets from core grid assets by the fact that they are used by a single customer. This approach to connection charges is similar to that used in other countries (see Section 3) and seems an appropriate approach.

Currently the transmission pricing system provides no locational signals relating to the main shared part of the network; the interconnection charges are levied only on load customers and charge rates do not vary by location. In Figure 1b, there is no incentive, on the basis of interconnection charges, for a new generator to locate at point D, rather than E. And for a new load customer, there is only a very small incentive to locate at B—its share of the increased total costs of the shared network.

2.3. Grid Investment Test

The Grid Investment Test (GIT) is used to ensure that investments in the grid are made consistent with economic efficiency and that grid reinforcements are compared with transmission alternatives[4]. The GIT considers investments that can be justified on the basis of their net benefits, as opposed to investments required to maintain grid reliability.

---

[4] Defined as alternatives to investment in the grid, including investment in local generation, energy efficiency, demand-side management and distribution network augmentation

The benefits that the GIT needs to assess are of avoiding the costs of:

- Supply interruption or unserved energy; and
- Out of merit order dispatch.

It needs to consider the best means to solve these problems, or indeed whether the benefits of any option justify the costs. Once the ideal option is decided there are some limits to the response. Whereas Transpower can deliver a transmission upgrade, it must rely on others to supply new generation, demand side management and/or shifts in the location of new load.

For new generation this might not matter. As discussed above, new generators entering a crowded node can price to extract congestion rent. We noted that this has some limits where there is expected to be a significant demand response to price, but in general, the signal of a crowded node can trigger investment in new generation. The same does not apply on the demand side. Demand side responses may be the least cost option but the market does not signal their value and, if the GIT suggests that demand side management is the least cost option, neither Transpower, nor any other participant, has the chief tool to trigger it—price.

There are distributional issues associated with the results of the GIT also. If grid reinforcement is identified as the least cost option, these costs, at least under current arrangements, are shared with all load customers. However, if the least cost option is new generation, the costs fall to a greater extent on those connected to the congested node(s). These distributional effects do not matter from the national perspective, but they do when they have incentive effects which distort decisions away from the optimum. Load customers connected to a crowded node will favour grid reinforcements, whose costs they can share, over generation options whose costs they would bear. This is especially important when the load customer may be the potential provider of the new generation, eg on site cogen.

2.4. Definition of the Locational Problem

The current charging scheme provides mixed signals for the efficient location of new load or generation.

The costs of grid losses are fully internalised via nodal prices. This provides signals for static and dynamic efficiency.

The costs of congestion are partially internalised. There are two types of congestion costs, those of:
- out-of-merit-order dispatch; and
- unserved energy.

The analysis above suggests that the costs of out-of-merit-order dispatch are signalled in the market, but only to generators. The costs of unserved energy are not signalled completely. The costs of unserved energy can be signalled by a new entrant generator to a node with excess demand, but they cannot be signalled by existing load customers in a way that might limit demand or signal the high costs of entry to new load.

In an ideal world we would see nodal prices settle at a price based on the interaction of:

- demand (or cost of unserved energy); and
- supply which includes current nodal price plus the costs of grid reinforcement or the lowest cost transmission alternative.

The lack of demand side bidding means the price signals provided to the demand side are weak if not absent. Where new generation enters, its cost is signalled. But the cost of grid reinforcement is not. We turn to options for signalling these costs in the next section.

3. Interconnection Pricing Options

This section discusses the range of pricing alternatives available for the shared grid. It provides examples from other countries to illustrate how these have been used in practice and includes a discussion of an approach previously used by Transpower in New Zealand. It places the discussion of overseas examples in the context of overall approaches to transmission pricing, including connection charges and pricing of losses and congestion.

Two main approaches are used to set transmission prices:

- Cost-recovery approaches that define a required revenue and then allocate this to customers, sometimes using locational signals; and
- Efficient pricing approaches that split the variable and fixed cost elements of the transmission system and seek to apply a marginal cost based approach to changes in use of the transmission system.

3.1. Cost Recovery Options

The main cost recovery options are[5,6]:

- **Postage stamp** techniques which typically divide total costs of operating the system by MW of demand or supply. This average cost is used to allocate the revenue requirement. The rate is updated annually to take account of changes in costs and load, including the costs of new transmission investment. The postage stamp approach can be used on a zonal basis, ie by defining load and costs into zones within the total network.
- **Contract path**—a flow of power from injection to exit is agreed between the transmission owner and the customer and provides the basis for an estimation of use of system. Compared with the postage stamp approach, this takes greater account of location by measuring a distance between injection and consumption.
- **MW-mile or MW-km** method—transmission charges are based on the physical distance between entry and exit (or relative to some agreed reference point) and the MW of transmitted power.
- **Power flow based**—this approach uses power flow analysis to estimate the extent of use of each network facility by each customer. Costs are allocated according to power flow.

Although these labels are used in various jurisdictions, there is a sense in which there are two main options, one using a postage stamp and the other being based on some estimate of network usage.

---

[5] Jian Yang and Max D Anderson (1999) A comprehensive dynamic pricing method for the use-of-transmission system charges in the context of power systems deregulation. http://www.power-gateway.com/papers/Potent99/potent99-long.pdf

[6] Thilo Krause (2003) Evaluation of Transmission Pricing Methods for Liberalized Markets – A Literature Survey. Internal Report EEH_PSL_2003_001. EEH Power Systems Laboratory. Swiss Federal Institute of Technology

The postage stamp has no real locational element, or does only to the extent that costs are split into regions or zones and costs attributed accordingly. Methods that measure use of the transmission system provide a greater degree of locational signal but depend critically on the way in which generation is mapped to load to estimate distance or use of system.

3.2. Efficient Pricing Methods

Efficient pricing methods seek to identify the change in the total costs associated with a change in quantity of load on the system or its location, and hence change in the pattern of flow. Prices defined in this way do not cover all costs so there is a need for additional charges to cover the remaining costs of transmission.

3.2.1. Short Run Marginal Cost Approaches

The short run marginal cost (SRMC) approach is an extension of nodal pricing. If p_i and p_j are the prices at two nodes, i and j, then the marginal transmission price is the difference between these, ie $p_i – p_j$; this is a measure of how much it costs to accept an input at i and deliver it to j.

To make this practical, generation needs to be mapped to load, or differences need to be defined relative to some reference node. Then, the sum of the differences in price will add to zero and will be equal and opposite for generation and load at any point in the system. Because of this, they might be used to redistribute a sum amongst participants. Or they might be used to ensure recovery of some variable element of costs, but not the fixed (embedded) element. This method is explained further below.

3.2.2. Long Run Marginal Cost Approaches

Long run marginal cost (LRMC) approaches take account of the need for additional system capacity to cater for additional demand on the system; LRMC incorporates capital and variable costs.

One suggested approach is “over a long time horizon of several years, all transmission expansion projects are identified and costed. This cost is then divided over the total power magnitude of all new planned transactions to calculate the marginal reinforcement cost”[7].

Alternative approaches use modelling to estimate required investments. Again, we consider below a variant of this approach which might be feasible in New Zealand.

3.3. Overseas Examples

There are a number of practical examples from other countries that provide useful information. Table 2 summarises the payment responsibilities for the fixed and variable/incremental components of network costs. It includes a summary of pricing for

---

[7] Shirmohammadi D et al (1991) Cost of transmission transactions: an introduction. Power Systems, IEEE Transactions, 6(3): 1006-1016 in: Thilo Krause (op cit)

connection and the approaches used to cover the costs of losses and of congestion, as this provides some context for the interconnection charging approach.

Table 2: Payment Responsibility for Network Fixed and Variable/Incremental Costs

Country	Fixed/ Embedded Cost	Incremental Costs		Reinforcement
		Losses	Congestion	
New Zealand	Loads	Loads via loss-adjusted marginal (nodal) prices	Loads via loss-adjusted marginal (nodal) prices	Connection costs attributable to individual loads/generators covered by responsible participant. Network upgrades treated as embedded costs.
California	Loads (including exports)	Generators (including imports) directly pay for losses by having their delivery reduced based on nodal loss factor	Scheduling coordinators (SCs) pay for congestion charges. These charges are assigned to generators and loads using specific arrangements among SCs and these entities, eg generators and loads settling based on zonal prices	Connection costs are covered by the responsible participant. Rolled in network upgrade costs normally are treated as embedded costs
England & Wales	Generators (27%) and demand customers (73%)	Generators and loads pay by having their metered volumes adjusted based on losses	Charges are paid equally between generators and loads based on their metered volume during the relevant half-hour	Connection charges are paid by responsible participants. Contribution to system costs has a locational element
Australia	Loads in the region	Generator and loads settle based on marginal price for energy, which include losses. The difference is paid to transmission service providers to offset embedded costs	Similar to losses	Network upgrades are normally included in embedded costs and treated accordingly. Connection costs are recoveed from responsible participants. Merchant transmission costs recovered from all participants through marginal nodal prices.
Norway	Divided among generators and loads on MW basis	Generators and loads including imports and exports	Generators and loads pay for congestion through their transmission agent or directly through zonal prices	Connection costs are paid by responsible participants. Rolled-in network upgrade costs are generally rolled into embedded costs
US FERC	Loads	Generators and loads share costs through loss-adjusted locational marginal prices (LMPs)	Generators and loads share costs through congestion adjusted LMPs	Rolled in network upgrade costs treated as embedded costs. No final policy on connection costs.

Source: Modified from Shirmohammadi D (2003) A comparison of Transmission Pricing Methods in Electricity Markets Worldwide (Work from Past Group 38) CIGRE Conference Brasilia-September 23, 2003.

In these markets, most include some means for passing on the costs of losses and congestion via spot prices. Also connection costs are widely attributed to new connectors. However, there are few examples of deeper reinforcement costs being allocated other than in the same way as fixed or embedded network costs, ie via a charge that smears costs across all customers.

The two main examples where there is a locational element of the main network costs are England & Wales and Australia, and these approaches are outlined in more detail below. Note, neither of these countries has nodal pricing of power, so transmission pricing is being used to correct a wider set of network externalities than is required in New Zealand.

3.4. England and Wales

In England and Wales, the National Grid Company (NGC) levies Transmission Network Use of System (TNUoS) charges, which cover long term costs of transmission assets, and Balancing Servce Use of System (BSUoS) charges, which cover the short term costs of balancing the system; there is also a separate charge for the Scotland-England interconnector[8]. The principle underlying the charging approach is that services are priced to reflect the incremental costs of supplying them at different locations[9]. The basis for the TNUoS charge is outlined below.

3.4.1. TNUoS Charges

TNUoS charges are levied on generator and load customers. They vary by location and have two elements:

- a charge to reflect the long-run marginal cost of a change in generation or demand at a particular point on the network;
- a charge to reflect the overall costs of providing a network, ensuring NGC recovers its total revenue requirement.

3.4.2. Locational Element

The locational element is derived by NGC using a (DC Loadflow Investment Cost Related Pricing or DCLF ICRP) transport model that uses a stylised representation of available transmission routes to estimate the difference in marginal cost (measured as additional km of transmission line) of an additional 1 MW of generation or demand at each node on the network. The model uses a set of inputs representative of peak conditions and based on NGC's Seven Year Statement, which sets out projected peak demand and new entry (based on contracted capacity) at different grid supply points (GSPs).

In the model, circuits with spare capacity are assumed to be less costly to invest in, because of the buffer before new investment is required. This is modelled by reducing

---

[8] Ofgem (2003) Transmission charging and the GB Wholesale Electricity Market.

[9] National Grid Company (2004) The Statement of the Use of System Charging Methodology. Effective from 1 May 2004. Issue 4 Revision 1.

the length of these routes to 75% of the original length (or by applying 75% of the expansion constant).

For the basis of calculation, all costs are estimated relative to a reference node[10]. For a MW of injection at each node, there is an assumed offtake of 1MW at the reference node. An estimate is made of the kilometres of extra transmission capacity required to maintain grid security standards. The marginal km cost for demand is equal and opposite to the km cost of injection at each node.

Nodal results can be positive or negative, depending on whether an additional MW increases or decreases overall utilisation of the routes in the transport model. The nodal results are grouped to form charging zones, based on the weighted average of all relevant nodes within a zone; there are 15 generation zones and 12 demand zones.

The system ensures that 73% of total use of system revenue is recovered from demand and 27% from generation. Locational tariffs account for approximately 25% of total revenue required.

Generation charges are calculated on the basis of either transmission entry capacities (if the zonal TNUoS charge is positive) or peak winter output (if the zonal TNUoS charge is negative). In 2003/04 they varied from negative £10.54/kW (Inner London) to £9.07/kW (North).

Charges for half-hourly metered demand are calculated annually, based on consumption levels during the three periods of peak system demand ('the Triad').

3.5. Australia

The pricing approach aims to reflect the additional transmission costs of additional use (or a decrease in use) of the transmission network by customers at any connection point. In the same way as the England and Wales system, there are two component charges:

- a Customer Transmission Use of System (TUOS) **usage** charge which is location-specific and relates to the costs associated with a particular connection point; and
- a Customer TUOS **general** charge, which recovers part of the balance of aggregate required revenues.

The usage component of TUOS costs for a connection point are established using the cost reflective network pricing (CRNP) method or the modified cost reflective network pricing (MCRNP) method, outlined below in respect of each connection point.

CRNP determines the costs of individual elements of the network, and the proportion of each network element used to transmit to each point on the network. Costs allocated to each point are the costs of every network element attributed to each point. These costs are then allocated to load customers on the basis of peak offtake at that point.

---

[10] Pelham GSP

Under the CRNP, the cost of each network element is determined by taking a proportion of its optimised replacement value in the network valuation using the annual cost fraction (the ratio of the annual revenue requirement to the total replacement value of the network valued on a deprival basis).

Generation is paired to load to estimate use of network elements. An electrical distance[11] is measured, in which a greater proportion of load at a particular location is assumed to be supplied by generators which are electrically closer than those which are electrically remote. The use made of any element by a particular load is the ratio of the flow on that element resulting from the supply to this load to the total use of the load made by all loads and generators in the system.

The following steps are then carried out for different conditions.

The change in flow in each network element (line or transformer) is calculated for a small change in each load (with corresponding changes in the generation supplying that load). These sensitivities (partial derivatives) are weighted by the load magnitude to give a MW "flow component" magnitude in each element due to each load for that hour.

The relative use of network elements by each load is obtained from the MW flow components above. In this calculation only flow components in the direction of the prevailing line flow are used. When all operating conditions have been completed the cost of the individual network element to be allocated to loads based on this locational allocation (ie half the annual cost of the element) is shared on a pro-rata basis with the maximum flow component that each load has imposed on the element for the operating conditions considered.

Finally the network costs are allocated to each load. This process involves summing, across all network elements, the product of their relative utilisation by that load and the allocated cost of each element.

3.6. Transpower's 1990s Transport Methodology

Transpower introduced a transport methodology for core grid pricing in 1996[12]. It was withdrawn when the methodology was revised in 1999. Some customers under long term contracts remained on it until their contract expired; all have now expired.

Power flow modelling was used to estimate the proportion of each grid asset that was used by each customer. It did not measure marginal or incremental costs. Rather it used proportional use to distribute a proportion of costs. It allocated the costs of optimised assets rather than the physical grid.

---

[11] In electrical engineering terminology the "electrical distance" is the impedance between the two locations, and this can readily be determined through a standard engineering calculation called the "fault level calculation" (Australian National Electricity Code, Schedule 6.4). (http://www.neca.com.au/files/necacode/)

[12] Conrad Edwards, Transpower, personal communication

The methodology used the "top 25% average" load flow as the basis for allocating 50% of the costs of each asset in the core[13] grid, the other 50% being allocated through the "Access" charge to all offtake customers by peak offtake. Core grid was defined as not connection or HVDC

The "top 25% average" load flow[14] was determined annually. Costs for each asset of the core grid were allocated in proportion to the portion of power flow attributable on this basis to a particular offtake customer.

However, there were a number of problems that made this methodology unpopular with Transpower's customers:

- customers with unvarying loads had costs that varied year by year because of changes elsewhere in the grid;
- there were some sections of the grid that effectively had no, or very little, flow. Small changes in injection/offtake could change the direction of power flow with the result that the cost of the assets would be allocated to completely different parties, contributing to the volatility; it was thus very sensitive to injection/offtake assumptions; and
- customers connected directly to a very expensive core grid asset (eg a small customer connected directly to 220kV) often faced much higher charges than a customer of the same size connected to smaller assets.

The rationale for the transport methodology was equity rather than efficiency, and the customers did not consider it equitable in practice. This was a major reason for its withdrawal in 1999.

3.7. Lessons for New Zealand

Most countries have no locational element in their transmission pricing. Those that do are countries that have not used nodal pricing.

There is some experience with estimating the proportional use of the existing assets as the basis for distributing the costs of the grid, eg on the basis of distance or using load flow methods, such as Transpower's 1990s method. But this is not an efficient way to cover fixed costs, nor does it provide locational signals that are related to costs.

Some countries have called their approaches marginal or incremental cost-based and set out to measure the impacts on total costs of marginal changes to the system in the form

---

[13] Note that the definition of core grid assets in the 1996 methodology is different from the definition of grid assets in the current pricing methodology because connection assets were defined differently in the 1996 methodology form the way they are currently.

[14] The top 25% average demand for a load is the average of the demands recorded for the load coincident with the top 25% of system demand (sum of all offtake). The top 25% generation was determined in a similar manner and adjusted to meet system losses. A load flow solution for the network was then determined and a portion of power flowing on each circuit was attributed to each connection point.

of additional injection or offtake at individual points on the network. These methods are computationally complex. And in practice, they do not measure marginal costs. Rather they measure incremental physical impacts on congestion. This is combined with estimates of the costs of grid reinforcement to provide a long run signal.

The England & Wales and Australian systems have separated out a variable cost element of the shared network. Allocation of costs to different parts of the network is undertaken using flow models. These enable an estimation of the impacts on total system costs of marginal changes in supply or demand at any point. The estimate of marginal costs is used to cover some costs of the core grid. There is a separate charge which spreads costs across all targeted customers.

The examples discussed above provide some possible methodologies for including a locational signal in transmission pricing in the form of a cost of reinforcement. However, the cost of grid reinforcement is not necessarily the minimum cost of mitigating congestion, since additional injection or offtake might perform the same role.

If the market was able to discover the long run cost as the lowest cost of these options, there would be little cause for concern. However, the systems discussed here pass on the long run signal, regardless of the actual market response.

In the next section we explore options for New Zealand, including whether a more efficient market signal can be given.

4. Possible Options for New Zealand

4.1. Energy Pricing or Transmission Pricing

In this review we have examined whether optimal locational signals are given to new investors connecting to, or having an impact on, the transmission grid. International examples have illustrated how a locational element in transmission pricing can be used to provide these location signals; these have suggested that the optimal signal is achieved via charging the long run marginal cost of grid enhancement.

However, grid reinforcement is only one option. It is used to address an underlying problem; that of congestion—there are also problems of grid reliability but these are fundamentally linked. The long run costs of congestion are the costs of out-of-merit-order dispatch and of unserved energy. These are more readily signalled in the market and in a way that can lead to optimal entry of new generation, load or demand side management, and against which the GIT can evaluate the relative costs of grid reinforcement.

The cost of unserved energy will always set the ultimate cap on costs, ie grid reinforcements would never be made if their costs were greater than the value of lost load. Thus if the market is able to signal these higher costs associated with unserved demand, it can incentivise lower cost solutions that will set market prices.

Our assessment is that this is achievable. But it requires a reform of price setting in which demand players can bid in the spot market.

In the absence of such reforms, a locational element in interconnection charges might be used.

What is largely missing is a signal that encourages optimal location of load and which provides ongoing reward for its location choice. However, as the GIT acknowledges, there are many potential solutions and, if a locational element of interconnection charges is used to encourage optimal location, the way that it should function, and particularly its duration, will differ with the solution.

- If generation solves a congestion problem, its costs will be reflected in nodal energy prices and ongoing locational grid pricing will add additional costs. Ideally locational pricing should discontinue on entry.
- If grid enhancement solves an excess load (generation) congestion problem, nodal prices will fall (rise) post entry to reflect lower levels of congestion. The congestion of the specific node changes relative to the average, and locational signals should ideally change.
- If demand side management solves a congestion problem, locational pricing should continue to reward those measures.

Below we examine two possible approaches to locational pricing:

- one that uses information available via current nodal prices; and
- another that measures marginal impact using flow modelling.

In both cases there would be a need for two elements of the interconnection charge. A fixed element that does not vary with changes to the pattern of grid connections, and an adjustable element that does. The approaches considered in this section would apply to the adjustable element.

There are two general issues that are relevant for both approaches considered here. The first concerns the need to estimate which proportion of total costs is covered by the adjustable element of tariffs. The second is whether discounts for one party should imply increases for some other party.

On this second issue, the relevant choice is between adjustments that involve transfers between participants, and those that do not. Transfers are unpalatable in this context for several reasons, not least being that they would create additional externalities (unpriced spillover effects) between participants; connection decisions of one party would affect other parties' liability for grid charges. Since the total cost of the grid needs to be funded however, the only alternative to a system of transfers is to secure some external funding for the locational incentive scheme.

Loss and constraint rentals may provide a suitable external source of funds. These are a by-product of efficient short-run pricing of electricity, and need to be disposed of in a way that does not create undesirable incentives. For example, it has been argued that such rentals should not be retained by Transpower on the grounds that this would create an incentive on Transpower to maximise losses and constraints.[15] The use of transmission rentals to fund an scheme that is designed to incentivise efficient alternatives to grid expansion would seem entirely reasonable.

This approach would also place a natural upper limit on the share of total grid charges that could be made adjustable. It may not be necessary or desirable to use all of the transmission rentals for this purpose; the appropriate scale of any incentive scheme remains to be determined.

4.2. Nodal Pricing-based Approach

In the NZ electricity industry, locational signals are provided to generators and offtake customers through the nodal pricing of electricity. Under this system, the price of power at any node of the national grid is intended to approximate the marginal cost of power for users, and the marginal value of power in the case of generators. Nodal prices are highly volatile, changing frequently through each day. As a consequence, their value as signals has sometimes been regarded as primarily affecting short-run decisions.

---

[15] We find this argument less than compelling, especially since Transpower is now subject to an overall price-cap threshold regulatory regime administered by the Commerce Commission.

It could however be argued that nodal prices do not actually provide effective short-run signals for many users. For practical reasons, final prices are not published until the next day, so participants need to react to forecast prices taking account of the possible forecast errors. Secondly, much of the final load has no real-time knowledge even of forecast prices and no short-run exposure to the final nodal prices. Finally, on the supply side of the industry, vertical integration and the trend towards balancing of expected load and dispatch provide generators with so-called “natural” hedges against the impact of the spot prices on which nodal prices are based. Energy suppliers may therefore more concerned about the gap between their generation cost and their retail prices than with nodal prices as such.

To the extent that nodal prices do affect market behaviour, their pattern over periods ranging from hours to months is more likely to be influential than their pattern at higher frequencies. For example, industrial users with time-of-use pricing may manage their daily load profile with a view to cost minimisation. Firms with seasonally varying usage, such as food processors, may adopt a similar approach over the course of a year.

Nodal prices could also affect long-run decisions over the location of generation and load. Other things being equal, new industrial plant would choose a location with low prices, while new generators would have the reverse incentive. Even though other things rarely are equal, one would expect that the reasonably persistent pattern of (relative) nodal prices throughout New Zealand would have some marginal effect on locational choices, notwithstanding the fact that these choices will themselves alter (post-entry) nodal prices in ways that tend to weaken the locational signal.

More generally, the pricing of electricity and its transmission is imperfect. Even ignoring the (DC) simplifications inside the nodal pricing algorithm, this system has mixed effects. It is an imperfect guide to short-run decisions and it is likely to have some influence over long-run decisions. This fairly obvious point has an important implication for all decisions over industry design and/or reform. According to the theory of second-best , unless it is possible to design a perfect (first best) system, there is no reason to expect that aligning any component of that system with first best principles will make the situation better overall.

This implies, for example, that first best principles about the role and treatment of sunk costs may be completely inappropriate. What matters for efficiency is the manner in which all of the components of the system interact to influence behaviour. The theoretical attractiveness of any part of the system is a secondary consideration. Nevertheless, for practical analytical purposes, there is considerable value in dissecting the issues involved in industry reform, provided the evaluation of options is conducted with reference to broader objectives.

4.2.1. Nodal Price Spreads

Our analysis identifies some circumstances in which there may be a role for transmission charges that influence the long-run decisions of load customers. The maximum cost concept for this purpose is the long-run marginal cost of grid capacity but there may be cheaper options that defer grid augmentation. Participants contributing to the deferral of capacity expansion could gain a benefit that depends on this LRMC, while those accelerating the need for expansion could face a cost that depends on LRMC.

The development of any mechanism of this type would require an understanding of the relationship between short-run energy costs and long-run grid expansion costs. A first step in developing such an understanding is to recognise that nodal prices reflect the short run marginal cost (and/or value) of electricity at each place and time. This cost depends on the values of, and interplay between, three variables:

- price offers received from the available generation;
- the total load and its spatial distribution; and
- the capacity of the transmission system.

The first two of these exhibit substantial short-run variation. The wholesale spot market is designed so that they interact through a reasonably efficient market, with the result that the price at any space/time point is equal to the short-run marginal cost of the highest cost plant required to meet total system load.

The third determinant of nodal prices, the capacity of the transmission system, evolves much more slowly and can be regarded as controlling the efficiency of the energy trading market. The greater the capacity of the transmission system, the more competitive is the power market because the lower are the marginal costs (caused by losses and constraints) of transporting electricity. This makes remote generation a better substitute for local generation, which mitigates local market power and stimulates the competitive process.

If we view the transmission grid as controlling (or at least influencing) the efficiency of the spot market, then the geographical spread in nodal prices indicates the extent to which it is successful. Let p_i and p_j denote the prices at nodes i and j at some time, and define the spread between these nodes as

$$S_{ij} = p_i - p_j.$$

Generation at node i is better able to compete with generation at node j (in both markets i and j), the smaller is S_{ij}. This implies that long-run decisions over the construction of new grid assets, new generation and new load will move the system towards a more efficient configuration if they reduce the set of spreads S_{ij} in some aggregate way.

Arguably, the variance of the S_{ij} also affects market efficiency, along with the mean value. The greater is the variability in the spread, the less certainty exists over at least one of the prices. And since the volumes injected into or drawn from the grid need to be chosen by participants, unpredictable price spreads can lead to large revenue transfers. Grid constraints will be an important determinant of the variance of the spreads.

4.2.2. Multi-lateral Spreads

If nodal prices are to be capable of supporting a LRMC-based charge, a method of summarising the information relevant to each node will be required. For each of the N nodes of the grid, there are (N-1) nodal price spreads at any time. We could aggregate these spreads by simply adding them, resulting in a node-specific spread factor f_i at each point in time.

$$f_i = \sum_{j=1}^{N} S_{ij}$$

At any time, the sum of these factors across all N nodes is zero. This is because $S_{ij} = -S_{ji}$. Alternative definitions of a multi-lateral spread are also possible, including methods that would weight each spread by some measure of the power flowing through the nodes at each end of the spread. However for the purpose at hand, the multilateral f factors defined above appear to be adequate.

The fact that the f factors sum to zero across the nodes means that they can in principle support a reallocation of interconnection charges among nodes. Suppose that I_i is the interconnection charge for node i under the existing price setting method. We could use the f factors to penalise nodes with high (multi-lateral) spreads by defining a new charge as $I^*_i = I_i f_i$. These new charges would recover exactly the same amount of revenue as the existing charges. This would magnify the location signal that is already present to some extent within the nodal price. This is not the only possible approach, but it does provides a basis for proceeding with the exposition of this method.

The f_i factors are an aggregate of spreads between one node (i) and all other nodes. We can place them all into a vector, denoted $f = (f_1, f_2, \ldots ,f_N)'$. The inner product of this vector $V = f'f$ is the sum of squared f_i factors. This is a useful measure of the overall variance of nodal prices in a geographical sense. If the locational signals are working well, the decisions of end-users will tend to reduce V over time.

4.2.3. Criterion for Adjustment

If a significant user connects to the grid or relocates to another node, the geographical pattern of nodal prices will change and this will induce a change in the pattern of multi-lateral spread factors. New load added near nodes that have a surplus of generation, and new generation added near nodes with a surplus of load, will both tend to reduce the absolute values of the multi-lateral spreads at nearby nodes. Based on our analysis above, such a reduction is beneficial to the system as a whole.

However the fact that the f_i's sum to zero means that when such beneficial changes occur in the pattern of connections, the multi-lateral spread must increase in absolute value for some other node(s). Thus, a locational signal that was automatically reflected through to all nodes on the basis of changes in the pattern of f_i factors would create externalities within the system, even if the overall spare capacity of the grid was increased.

Such externalities will always be created if the required revenue is simply reallocated among participants. As noted above, the only apparent alternative is to fund the incentive scheme externally, such as through the use of the existing transmission rentals.

If such external funding is available, it is possible to consider indicators other than the f_i factors as a basis for incentivisation. The geographical variance of nodal price spreads across the system, V, is an attractive option. Under this approach, users would effectively receive a payment[16] for connecting new facilities when this had the effect of reducing the overall price variance.

At this stage it is not clear whether incentive payments should be based on node-specific information such as the f_i factors, or on overall system information such as V. Fortunately, we can still describe the other feature of this system without relying on any particular choice of indicator.

4.2.4. Timing Issues

If a significant new load or generator is connected to the system today, the relevant efficiency indicator (f_i or V) will undergo a step change. The objective is to induce market participants to locate these new connections in ways that increase efficiency rather than decrease it. Since change is sought, incentive payments need to be linked directly to changes. This suggests that a one-off payment, linked to the change in the relevant indicator, could be appropriate.

There are several advantages in using a one-off payment, as distinct from a sequence of payments over several years. Perhaps the most compelling is that doing so would provide a more certain and immediate impact on the investor's calculation of project economics, and therefore be more likely to influence the location decision. If payments are spread over several years, there will generally be some risk of not receiving them, or not receiving all of them. This would dilute the incentive power associated with any level of funding.

A second advantage of paying once is that doing so allows subsequent investors to secure their own incentives without fear of some share of the funding being secured by investors who entered earlier. Such a concern might arise if, for example, a factory that was built last year argues that is about to increase production and is therefore entitled to additional funding. For these reasons, we favour a "pay once" system.

A further timing issue relates to whether incentive payments would be based on ex-ante predictions or one ex-post measurements. This involves a slightly awkward trade-off. There is a risk of some inaccuracy in predictions of the way nodal prices will change in response to major new connections, and this counts against the ex-ante method. On the other hand, this approach would have a considerably stronger incentive effect, since potential investors would know the payment in advance of commissioning their plant.

---

[16] The term "payment" should not be interpreted literally here. We have in mind a system of discounts off the standard interconnection charges, and use 'payment' as shorthand for such discounts.

On balance, we prefer the ex-ante approach. However we note that it would be important if using this method to ensure that the incentive payment was structured in such a way that any delays in plant commissioning would also be reflected in the payment.

4.2.5. Scale Issues

Perhaps the most difficult problem with the nodal pricing based method is to link changes in the spread indicator to the long-run marginal cost (LRMC) of grid expansion. There are two main components to this problem: estimating LRMC and aligning this with the incentives being designed.

There are likely to be significant problems in estimating the LRMC at the nodal level of aggregation. There are several reasons for this. One is simply the complexity of physical flow paths through the grid and the difficulties that loop flows and constraints have on predicting these paths. Secondly, but relatedly, to the extent that there are "pinch points" within the grid, the LRMC of grid expansion will depend on the cost of expanding capacity through these sections of the grid. Finally, grid investment is far from "marginal". Capacity enhancements are only available in discrete and relatively large lumps.

If the problem of estimating LRMC can be solved, the linkage with a nodal pricing indicator should also be possible. The general approach would be to view the indicator as an index, estimate the extent to which the index would move in response to a known capacity increment in the grid, and associate with that change in the index a monetary value equal to the cost of the capacity increment. If there was a linear relationship between grid investment costs and the nodal pricing indicator, this exercise would only need to be undertaken once and extrapolated to all other cases. Even if there is no predictable (i.e. location independent) relationship of this type, repeated applications of the same principle could be used to estimate the appropriate scale of incentive required.

4.2.6. Eligibility Issues

A question arises as to who should be eligible for incentive payments, and whether to penalise parties that increase load on the grid. The main options are to make it an opt-in system or a compulsory one. Additionally, there is likely to be a need to define some threshold beyond which a change is deemed to be sufficiently important to warrant inclusion in the incentive scheme.

Under the opt-in approach, the Commission would call for applications for incentive payments, assess these and determine the appropriate scale of incentive for each applicant. This would have a simplicity advantage for the Commission and could work well provided the incentives were sufficiently attractive. The main disadvantage is that those parties connecting in the "wrong" locations would not reveal themselves if doing so would make them liable for incentive-based penalties.

4.3. Flow Modelling-Based Charges

The alternative approach is to develop a system of charges based on an estimation of the impacts of marginal changes in supply and demand at every node in the system.

One approach might be to use Transpower's scheduling pricing and dispatch (SPD) model, currently used to calculate nodal prices, to estimate the existing level of spare capacity of each kilometre of network. This would produce a total equal to the sum of MW-kilometres, ie the sum of spare capacities for each kilometre. Modelling injections of additional MWs via generation at each node, or additional demand at each node, would be used to estimate the change in total capacity kilometres.

This would provide a factor for each node that might be positive or negative. And which would be equal and opposite for injections and off-takes. These factors might be used alongside a cost of reinforcement of the grid to provide a MW of capacity for one kilometre—most likely this would be the per MW costs for a weighted average size of network connection (although arguably locational signals should reflect the costs of reinforcement of specific grid components). Noting the signals for generation that exist at the moment, the signals might be limited to demand off-takes only.

The MW-kilometre factor times the MW-kilometre cost provides a cost for attachment to each node of the network and would apply to existing and new entrants per MW of peak demand. The application to incumbent loads provides efficient signals because it provides incentives to shave peaks where this would yield benefits to the network.

This is an approach similar to that applied elsewhere, eg in the England & Wales and Australian systems. It is different from the approach previously used in New Zealand because it is used to estimate a marginal effect, and not to allocate required revenue; in practice this means that it applies to a smaller quantity of revenue.

There are many practical difficulties with this approach. These include the need to agree a set of assumptions for analysis, including such factors as hydro inflows, availability of key thermals (such as Huntly) and future projections of new entry. Reaching agreement on such assumptions, which would be likely to have significant impacts, would be challenging, to say the least. In addition the SPD model is not currently set up to undertake such analysis and it would take some time to do so.

That said, this approach is used elsewhere and it is clearly possible.

One of the important decisions is the frequency with which the assessment of the values for each individual node was undertaken. For loads it is important that a long term signal is provided. This would suggest that any assessment of locational prices would be for multiple years.

5. Conclusions

Optimal location of activity in the grid is achieved through signalling long run marginal costs (LRMC) of that activity. The LRMC of congestion is not necessarily the cost of grid reinforcement, but the market can discover LRMC below a maximum equal to the cost of unserved energy. Currently there are externalities of congestion that have not been fully internalised that prevent the discovery of LRMC.

It is likely to be easier to signal the cost of unserved energy in the market, or allow it to be signalled, rather than the costs of one of the options that might be chosen to address the congestion problem. The cost of unserved energy will always set the ultimate cap on costs, ie grid reinforcements would never be made if their costs were greater. Thus if the market is able to signal these higher costs, it can incentivise lower cost solutions that will set market prices.

Currently the lack of a locational signal relating to congestion externalities is very largely limited to the demand side—location of new demand customers, demand growth and demand management measures. The key issues are that, post-entry, nodal prices change so that the desirable load is not rewarded, and pre-entry, prices do not rise to reflect marginal willingness to pay for electricity, so that excess demand is not discouraged.

Our assessment is that these issues would be very largely addressed through demand-side bidding.

The GIT will not do much, in itself, to address this externality problem. It will enable assessment of when grid enhancement is the best option, but can do little to solve the demand side problems in the absence of price signals.

Without pricing reforms, a locational element in interconnection charges might be used. However, the long-term signal that is required changes after the solution has been introduced. And specifically, keeping the locational price signal after new generation has been introduced will include double costs while retaining it for a load solution is entirely appropriate. Two options were explored in Section 4, both of which are complex and bring large uncertainties regarding the efficiency of the signal. It is unlikely that such complexity is justified, given the extent of the locational signalling problem.

Our assessment is that introduction of demand side bidding is the better solution. In its absence, our assessment is consistent with that of Frontier Economics, that the interconnection charges should spread core grid costs across all load customers and that the GIT has an important role in identifying locational signals. This leaves the signal to load largely unaddressed, but the magnitude of the problem does not justify the introduction of a locational element in the interconnection charge.