Nodal Prices and LRMC charging

Executive Summary

1. The scheduling, pricing and dispatch model (SPD) sets nodal prices in the New Zealand electricity market to signal costs of transmission limits (congestion) as well as energy costs and losses. The focus of this paper is on congestion.

2. Currently, transmission constraints (congestion) are priced node-by-node in the wholesale market. The congestion price for a circuit rises above zero to ration use of the circuit to capacity when users demand for electricity would otherwise cause the circuit to exceed its capacity.

3. Nodal pricing is widely regarded as highly efficient, in terms of signaling marginal costs of using scarce capacity and so constraining grid use to capacity. Where there is congestion, congestion prices rise. Where there is no congestion, congestion prices are low or zero to reflect that there is no opportunity cost in using the grid (ie, to reflect that one party's use of the grid does not prevent another party from using the grid).[1] This is efficient.

4. Provided there is sufficient price-sensitive load and generation at each node, nodal prices are by themselves sufficient to ensure that the use of the grid is constrained to its capacity and no other form of load control is required.

5. Some submissions on the TPM have argued that nodal pricing is by itself insufficient to ensure efficient use of and investment in the grid. Instead, they argue, nodal pricing needs to be supplemented by a long run marginal cost (LRMC) charge. Under LRMC-based pricing, users of the grid are charged a fee which is proportional to the costs of new capacity that will be needed to serve increased demand. This fee is targeted at periods of peak demand and increases as the line becomes increasingly congested.

6. Although the LRMC charge is targeted at peak demand, it has no effect when nodal prices are sufficiently high, since it is offset by a reduction in the nodal price.[2] It can only have an effect by reducing the use of the grid to below its capacity; that is by causing inefficient use of the grid[3]. This inefficiency is a price some proponents of an LRMC charge are prepared to pay for what they see as its benefits.

7. Several issues have been raised in support of the need for an LRMC charge. These are:

    a. in practice, **nodal prices do not and cannot rise high enough** to reflect the impacts of additional demand on congestion because:

        i. when capacity is very limited, **un-priced demand reductions are used to manage congestion**

        ii. **high nodal prices will draw negative public and political reactions** and are therefore not a sustainable approach to signaling costs of use

---

1 Except losses, which are ignored here for simplicity.

2 This reduction in the nodal price would occur "automatically" as demand would reduce during periods when the LRMC charge is in operation.

3 This assumes that nodal prices are efficient relative to the other practical options available.

iii. users never see the full costs of their actions because **investment is usually triggered 'early'**, before nodal prices have risen to levels commensurate with signaling that additional investment would be beneficial

b. **consumers don't actively monitor nodal prices** and so will not respond to changes in prices.

c. **consumers don't accurately anticipate nodal prices** so they make long-lived investment decisions which will eventually cause congestion and higher prices, but they do not take that into account because they do not know that this will happen.

d. **consumers cannot coordinate to take actions that affect the timing (and so the cost) of future transmission investment,** so although they may be aware of it, they do not take account of the impact of their decisions on the need for future transmission investment.

8. We have assessed all these claims. In summary, we have concluded that in most of the situations where we have considered the case for a LRMC charge, the case does not stand up. The one possible exception is the argument that an LRMC charge might be needed to ensure that consumers take into account the effect of their own decisions on future transmission investment (the issue discussed in paragraph 7.d). If they could coordinate to do so, this may lower transmission costs and so overall costs. However, even in this case it remains questionable whether the LRMC-based charge would improve efficiency in practice: this would need to be tested through cost benefit analysis. In this executive summary, we first discuss the situation in paragraph 7.d where an LRMC charge might be justified, and then the other situations discussed in paragraph 7. The main paper discusses these issues in the same order as paragraph 7.

Consumers can't coordinate actions to affect the timing of transmission investment

9. The fourth concern raised above is that consumers cannot coordinate to take actions that affect the timing (and so the cost) of future transmission investment. As a result, even though they may be aware of it, they do not take account of the impact of their decisions on the need for future transmission investment. As a result, they make long-lived investment decisions which will eventually cause congestion and higher prices, but they do not take that into account because they cannot coordinate to take any action to avoid the costs. If these long-term effects were signaled to consumers via an additional price that applied consistently during peak periods they would make different decisions and those decisions would lead to more efficient outcomes compared to relying on nodal prices.

10. We think this concern is potentially valid in theory; however, even in this case it remains questionable whether the LRMC-based charge would improve efficiency in practice.

11. The potential issue can be seen by assuming there is a single user of the new transmission investment who can buy an energy-using technology (which makes immediate transmission expansion efficient) or an energy-saving technology (which allows deferral of expansion for a number of years). Then the user's decision is efficient because the user alone affects the timing of the transmission investment. The user's decision is to buy the energy-using technology and expand transmission if the lower cost of buying the energy-using technology (including capital and energy costs) and the benefit the user gets from the increased capacity (ie, lower congestion charges and increased use) is less than the present value of the savings from deferring transmission investment.

12. Where there are multiple users, each user correctly assumes that they do not much influence the timing of the investment. As a result, their private calculation does not take account of:

    a. the present value of the deferred transmission costs

    b. the benefit that user gets from the early expansion of the transmission investment (in terms of lower congestion charges and consequent increased use of energy).

13. The omission of these two terms can cause individual users' decisions to deviate from the efficient decision discussed in paragraph 11.

14. The annualised value of the savings from deferring the transmission investment is just the LRMC of the investment. So (ignoring the term in paragraph 12.b), imposing an LRMC charge makes it profitable to invest in the energy-saving technology if it is efficient to do so. That is, it results in the joint optimisation of their own investment decisions and the transmission investment decision(s).

15. However, there are other considerations that need to be taken into account, which mean that an LRMC charge will over-signal the benefit of deferring grid investment and mean that calculating the appropriate charge is much harder. These include the effect of nodal prices; the benefit users get from altering their use to avoid future transmission charges, and the benefit the user gets from an early expansion of capacity. In addition, as discussed before, an LRMC charge has the effect of inefficiently reducing grid use below capacity.

16. The LRMC charge would only improve efficiency to the extent that it is a better approximation to the annualised cost of the actual investment eventually undertaken, adjusted by the considerations discussed in paragraph 15, than not imposing any additional charge and to the extent that the benefit exceeded the cost of reducing grid use. Given the difficulties in estimating LRMC and in making many of the adjustments discussed above, this issue is not clear cut.

17. The rest of this executive summary covers the other issues in paragraph 7 where we consider a LRMC charge is not justified.

Nodal prices cannot rise high enough

18. It is suggested that in practice, nodal prices do not and cannot rise high enough to reflect the impacts of additional demand on congestion. There are three reasons that are commonly advanced. However, in each of these cases, we consider that it is unlikely that an LRMC charge would improve efficiency.

19. First, it is suggested that when capacity is very limited, unpriced demand reductions are used to manage congestion. This may be an issue at present since SPD cannot solve for congestion prices when there is not enough price sensitive demand or generation at a node to allow demand to be reduced below capacity. However, this should be resolved by the introduction of real time pricing, since this will enable users to react to actual prices, and it should result in more load reacting to prices by reducing its demand.

20. However, it may be prudent to provide for a temporary demand control charge in the TPM in case real time pricing does not function as expected or to ensure that any failure to curb peak demand associated with removing the RCPD charge can be contained.

21. Second, it is suggested that high nodal prices will draw negative public and political reactions and are therefore not a sustainable approach to signaling costs of use. This seems unlikely, since the electricity market has been in place for some time, and the high nodal prices that have occurred appear to have been broadly accepted. In part this may be because price sensitive

customers can shield themselves from nodal prices with hedging products and because most households are on fixed price variable volume contracts.

22. Third, it is suggested that users never see the full costs of their actions because investment is usually triggered 'early', before nodal prices have risen to levels commensurate with signaling that additional investment would be beneficial. If this is so, it is because there is some mechanism, other than nodal prices, that is triggering the investment. The appropriate policy solution is not to increase the nodal price with an LRMC charge, but to address the problem that is causing the early investment. If an area-of-benefit charge is introduced, it should help in this regard, since it will encourage the supposed beneficiaries of the investment to oppose inefficiently early investment.

23. So in each of these cases, it is unlikely that an LRMC charge will improve efficiency. Moreover, even if it were decided that an additional price signal was desirable, it would be more efficient to impose it through SPD, since that would restrict the additional signal to the times when capacity is tight and would so not inefficiently discourage grid use when there is plenty of capacity.

Consumers don't actively monitor nodal prices

24. The second concern raised is that consumers don't actively monitor nodal prices and so will not respond to changes in prices.

25. It is not clear why consumers would respond to an LRMC charge when they would not respond to nodal prices. One argument is that it is more stable than nodal prices. However, this ignores the fact that stable and predictable prices can emerge efficiently through the workings of the electricity market. For example, if mass market consumers value stability, they can enter into a fixed price variable volume contract with retailers. Provided the retailer finds it profitable to provide such contracts, that is efficient.

26. In addition, it seems likely that over time aggregators will emerge to manage load on behalf of consumers.

Consumers don't accurately anticipate future nodal prices

27. The third concern raised is that consumers don't accurately anticipate nodal prices so they make long-lived investment decisions which will eventually cause congestion and higher prices, but they do not take that into account because they do not anticipate that this will happen. It is argued that if these long-term effects were signaled to consumers via an additional price that applied consistently during peak periods they would make different decisions and those decisions would lead to more efficient outcomes compared to relying on nodal prices and on consumers' ability to predict the effects of their decisions on congestion and on nodal prices.

28. It does seem unlikely that mass market consumers would anticipate that congestion charges and transmission charges are likely to rise in the future. However, it is important to distinguish between these two prices. If the congestion price is currently near zero but is expected to rise in the future, that is eventually likely to trigger congestion-relieving investment. At that point, the congestion price will again fall to zero. So if the user makes an investment decision now assuming that the congestion price will remain near zero, the decision is likely to be relatively efficient. In this case, imposing an LRMC charge would encourage them to expect a higher congestion charge in future and lead them to make what is from their perspective an inefficient investment (ie, ignoring the problem of co-optimisation with transmission investment discussed under the heading *Consumers can't coordinate actions to affect the timing of transmission investment* above).

Conclusion

29. In most of the situations where we have considered the case for a LRMC charge, the case for an LRMC charge does not stand up. Typically, the best solution is to rely on nodal prices and instead focus on improving the responsiveness of demand and supply to nodal prices. If an additional price signal to that currently provided by nodal prices is required, the most efficient way of imposing the charge is by imposing it as an additional constraint in SPD.

30. The one situation in which an additional charge related to LRMC may in principle justified is to ensure that consumers take into account the effect of their own decisions on the need for and timing of future transmission investment. In this case, the calculation of the charge is quite complex, which makes it more questionable whether the LRMC-based charge would improve efficiency in practice. In addition, this efficiency gain has to be compared with the efficiency loss resulting from the reduction in current use and inefficient short-life investment that the LRMC charge causes.

31. In addition it may be prudent to provide for a temporary demand control charge in the TPM (not related to LRMC) in case real time pricing does not function as expected or to ensure that any failure to curb peak demand associated with removing the RCPD charge can be contained.

Purpose

32. Wholesale electricity prices (nodal prices) signal costs of congestion in New Zealand transmission networks. But are those signals sufficient to ensure efficient use of and investment in the grid? Nodal prices also signal energy costs and losses, but the focus of this paper is on congestion[4] and congestion prices (also sometimes called nodal transport charges).

33. This paper discusses the benefits of locational marginal prices (LMP), how they operate at present under nodal pricing[5] in the New Zealand wholesale electricity market. The paper also lays out arguments for and against supplementing nodal prices with a long run marginal cost (LRMC) charge. Though the two are related in some ways, the latter is less efficient in most circumstances (leaving aside transactions costs).[6]

Background

34. The Authority published the document *Transmission Pricing Methodology Review: LRMC charges: working* paper on 28 July 2014. Many parties, including Transpower and distributors, supported the introduction of a long run marginal cost (LRMC) charge.

35. The Authority published a proposed TPM, which is set out in the *TPM Issues and Proposal: Second Issues Paper* on 17 May 2016 (second issues paper) and in the *TPM Second Issues Paper: Supplementary Consultation Paper* on 13 December 2016 (supplementary consultation paper). This proposed TPM is called the 2016 TPM proposal in the remainder of this paper.

36. The 2016 TPM proposal allowed for the possibility that an additional type of charge – called a LRMC charge – may be applied to support wholesale 'nodal' prices in order to more effectively signal congestion costs.

37. The supplementary consultation paper stated that:

> *'The Authority sees the LRMC charge as a price that reflects the opportunity cost of the current use of a scarce resource – the existing grid. The user who benefits from the grid pays the LRMC charge not because future investment is required but because the opportunity cost of their use of the existing grid is the cost of denying another user the use of the existing grid.'*

It stated that an LRMC charge could only be included in the TPM if nodal pricing was insufficient to ensure efficient grid use and if it was at least as good as using other forms of grid support

---

4 Congestion prices, in general terms, are prices which rise or fall to reflect increases or reductions in use of a resource when there is limited capacity and additional use has negative effects on service quality or quantity for all people using the resource. In the context of electricity networks, congestion prices reflect costs from using grid circuits to transport electricity. Congestion prices ration the use of the circuits to their capacity and signal costs to all users from increased losses of energy [Comment: and the cost of more expensive supply that must be used to supply demand because of the constraint) as demand levels increase. This note is mostly concerned with situations where use of the grid needs to be kept within absolute limits (i.e. rationed) rather than losses.

5 As is discussed later in the paper, locational marginal pricing, which is the basis for setting nodal prices in New Zealand, is widely regarded as highly efficient. It is of course feasible to have nodal pricing based on less efficient methods. But in New Zealand the term 'nodal pricing' has become synonymous with locational marginal pricing.

6 For clarity, the discussion abstracts from the transactions costs of applying nodal prices. There is a potential trade-off between the benefits of applying nodal prices deeper in to the network and the costs of calculating and applying the prices. This is unlikely to be relevant to transmission, but may become relevant if consideration was given to extending nodal pricing into distribution networks.

arrangements to limit grid use. The supplementary consultation paper guidelines did not put any other constraints on the design of the charge.

Several submissions on the Authority's 2016 TPM Proposal have promoted the view that nodal prices on their own are not sufficient to support efficient use of or efficient investment in the grid.

Nodal pricing is an effective method for pricing congestion

38. Currently, transmission constraints (congestion) are priced node-by-node in the wholesale market. Wholesale prices are set to reflect the cost of increasing generation and/or reducing demand so as to avoid violating physical and reliability constraints (limits) on the use of the grid.

39. This method of pricing is called nodal pricing or locational marginal pricing. It does four related jobs. It:

    a. **identifies the value,** to users, of access to electricity and to transmission capacity. Higher prices indicate that users who purchase electricity place a high value on the ability to use electricity while users who value use of electricity less highly go without

    b. **ensures users face the costs of using the grid** such as the cost of choosing to consume during periods of constrained capacity (in terms of the effects this has on the service quality for other users)

    c. **allocates available capacity to the highest valued uses**, or at least reduces risks that quantity-based rationing[7] will reduce service to high-valued uses

    d. **signals the efficiency of new investment** in increased transmission capacity or increased demand (eg, a new factory or a new heat pump). The higher and the more frequently that prices rise, the more likely it is that it will be efficient to undertake investment to increase capacity (because the benefits of investment - such as lower nodal prices - outweigh the costs) and the less likely it will be efficient to invest in energy using equipment.

40. The system operator's Scheduling, Pricing and Dispatch (SPD) model is used to set nodal prices. It schedules generation and instantaneous reserve offers and demand-side bids, to minimise total system costs, subject to various constraints. Among these are 'security constraints', which are required to enable the system operator's security policies to be met. These constraints reflect congestion and reliability-related rules as well as physical limits and, and there is, in practice, little difference between the two.

41. SPD is used to calculate congestion prices and those prices then signal the cost grid users impose on others by using the grid (ie, the cost of congestion).[8] If demand is high and there are limits to transmission capacity, this can cause higher priced generation to be dispatched instead of cheaper generation – because the lower priced generation can't be dispatched without compromising security constraints. The increase in costs caused by having to dispatch higher priced generation is the cost of the transmission limits. It is reflected in the difference in the

---

7 That is, planned and unplanned load shedding and electricity outages resulting from users desired consumption exceeding the grid's capacity.

8 This description applies to all users. Congestion is a problem relating to collective action. When congestion occurs there is not automatically any obvious identifiable 'additional' user. Congestion pricing, in effect, defines the 'additional user' as the one least willing to pay for the last unit of scarce capacity and so the user who does not get to use that unit of scarce capacity.

nodal price between where the load connects and the low price where the low cost generation connects.

42. It is also possible for local load in the constrained area to submit a bid to reflect the price at which it is willing to forgo using electricity. If it is not willing to pay higher prices, then the cost of the transmission limit will be signaled by the value of the demand bid (if it is a dispatchable demand bid).[9]

43. Few consumers currently bid dispatchable demand. But this is expected to change with improvements in wholesale pricing (discussed in the next section) and wider use of rapidly improving demand management technologies. In addition, many consumers do monitor the evolution of wholesale prices and when constraints cause prices to rise they can reduce their demand. Reductions in demand are then reflected by a decline in prices. In this way price 'discovery' and cost signaling is currently a multi-period process.[10]

44. Provided there is sufficient price-sensitive load and generation at each node, nodal prices are by themselves sufficient to ensure that the use of the grid is constrained to its capacity and no other form of load control is required.

45. Nodal pricing[11] is widely regarded as highly efficient, in terms of signaling marginal costs of using scarce capacity and so constraining grid use to capacity. Where there is congestion, congestion prices rise. Where there is no congestion, congestion prices are low or zero to reflect that there is no opportunity cost in using the grid (ie, to reflect that one party's use of the grid does not prevent another party from using the grid).[12] This is efficient.

46. Congestion prices, in SPD, are also intimately connected to investment costs. That is, the benefit of transmission investment is signaled by congestion costs. This is because costs of congestion and benefits of investment are symmetric. Congestion pricing is based on the cost of

---

9 Dispatchable demand bids are offers load makes to reduce demand if prices rise to a particular price level. Consumers who offer dispatchable demand bids need to have demand which is capable of being reduced on direction (i.e. when being dispatched) and are subject to specified metering requirements so that demand reductions can be verified. The term 'dispatch' is used for these demand bids as well as for generation offers because reducing demand is approximately equivalent to increased supply.

10 Currently prices in the wholesale market are only indicative when dispatch takes place. Prices are then subject to verification processes and not finalised until 2 days later. However, consumers do respond to these indicative prices. The System Operator also publishes forecast prices which consumers may respond to.

If and when real time pricing is introduced, a reduction in demand driven by indicative 5 minute prices will be reflected in final prices for the trading period. This provides dynamic information about the value users put on accessing additional electricity and so the value of new investment to relieve the constraint causing the high prices. Indeed retailers have been known to increase retail prices to reflect spot market costs associated with transmission constraints (as discussed in para 6.1.4 of the Electricity Commission's cost benefit analysis of locational rental allocations in 2009 – see https://www.ea.govt.nz/dmsdocument/911). This parallels the way they increase retail prices to reflect rising long run energy costs. So the price signal is passed through the system.

11 More precisely, 'locational marginal pricing', which is the basis for the SPD model's methods, is widely regarded as highly efficient. It is of course feasible to have nodal pricing based on less efficient methods. But the term 'nodal pricing' has become synonymous with locational marginal pricing.

12 Except losses, which are ignored here for simplicity.

demand that has not been met.[13] The benefits of investment are, approximately, the value of the demand that can be served with additional capacity.[14]

47. There are other ways to signal opportunity costs of using the grid but using LMP is the best way. This is because it imposes a price which constrains use of the grid only when a constraint would otherwise be violated and only to the extent necessary to ensure that the constraint is not violated. So it allows users to make the best possible use of the grid consistent with that constraint. As one expert has noted 'the LMP system already in place represents the 'gold standard' for appropriate marginal pricing incentives'. The expert, Professor Bushnell[15], goes on to say that 'an additional surcharge to LMPs under the guise of long run marginal cost, or benefits-based charging, is only justified on efficiency grounds if there is a fundamental problem with the LMPs, or the ways network users, and planners, respond to them. If this were the case, the proper response is to address the source of the problem (LMP penalties, the transmission planning process) rather than attempt to correct one distortion by adding another one'.

48. The current RCPD charge is seen by some as a method for signaling costs of congestion, or at least the cost of additional demand on potential new investment to relieve congestion. This charge is very blunt. It is not related to the cost of congestion and is not necessarily applied when congestion is present – because, for example, it is applied at a highly aggregated level (region wide). So, RCPD is at best a very poorly targeted congestion charge. RCPD charges do, however, influence demand during times when congestion may be present. Distribution companies and other consumers explicitly use demand control (eg, distributors use ripple control) to reduce their exposure to these charges.

Are there limits to nodal prices?

49. Several issues have been raised in support of the need for additional pricing signals[16]:

    a. in practice, **nodal prices do not and cannot rise high enough** to reflect the impacts of additional demand on congestion because:

        i. when capacity is very limited, **un-priced demand reductions are used to manage congestion**

---

13 See the Appendix for a brief graphical description of the connection between congestion prices and investment costs when investment proceeds efficiently.

14 Benefits of investment can also include reduced losses. These benefits are proportional to degree of congestion but can also accrue to users where there was no congestion. However, these benefits are very small compared to the value of avoided congestion. For the North Island Grid Upgrade the total benefits from avoided losses were estimated to be in the order of 3,000 to 7,500 GWh (present value, 6% discount rate) rate – see www.ea.govt.nz/dmsdocument/169 pp.25-27 for information on the full range of potential loss reductions under a range of different scenarios about how the electricity market might evolve. At a long run marginal cost of energy of $70/MWh this equates to between $21 million and $53 million, present valued, over a 50-year project life (discounted at 6% per annum). And most of those benefits accrue to users of the grid during peak times when losses are largest. For a comparison on the relative size of investment benefits, Transpower estimated that the North Island Grid upgrade would reduce the cost of unserved energy by $27.3 billion, in present value terms, over the project evaluation period.

15 James Bushnell Equity and Efficiency Implications of New Zealand's Transmission Pricing Methodology Options, August 2015 https://www.ea.govt.nz/dmsdocument/19786.

16 These issues, or similar expressions of them, have been raised by various parties (eg, Axiom Consulting Ltd, for Transpower) in submissions on the Authority's 2016 TPM proposal.

ii. **high nodal prices will draw negative public and political reactions** and are therefore not a sustainable approach to signaling costs of use

iii. users never see the full costs of their actions because **investment is usually triggered 'early'**, before nodal prices have risen to levels commensurate with signaling that additional investment would be beneficial

b. **consumers don't actively monitor nodal prices** and so will not respond to changes in prices.

c. **consumers don't accurately anticipate nodal prices** so they make long-lived investment decisions which will eventually cause congestion and higher prices, but they do not take that into account because they do not know that this will happen. If these long-term effects were signaled to consumers via an additional price that applied consistently during peak periods they would make different decisions and those decisions would lead to more efficient outcomes compared to relying on nodal prices and on consumers' ability to predict the effects of their decisions on congestion and on nodal prices.

d. **consumers cannot take actions that affect the timing (and so the cost) of future transmission investment,** so although they may be aware of it, they do not take account of the impact of their decisions on the need for future transmission investment. As a result, they make long-lived investment decisions which will eventually cause congestion and higher prices, but they do not take that into account because they cannot take any action to avoid the costs. If these long-term effects were signaled to consumers via an additional price that applied consistently during peak periods they would make different decisions and those decisions would lead to more efficient outcomes compared to relying on nodal prices.

Real-time pricing will improve congestion price signals

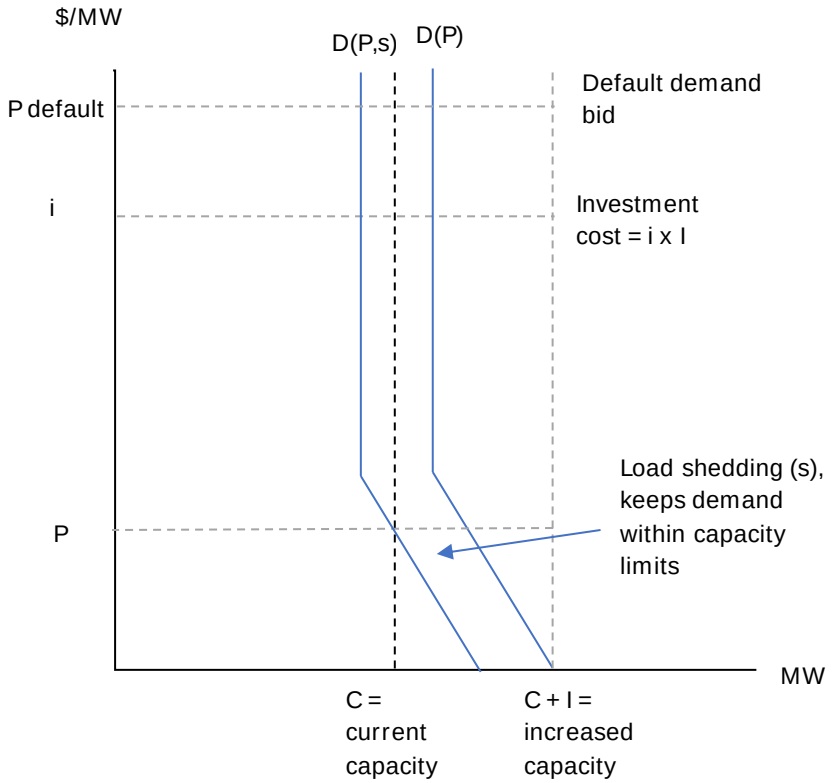
50. The first problem, of un-priced demand reductions, is a potential issue but one that should be resolved soon.

51. SPD cannot solve for congestion prices in situations where demand needs to be dispatched but there are no demand-side bids in large enough amounts to ensure system security. In these situations, demand is reduced through managed reductions in load – such as distribution companies using ripple control to reduce demand. This situation is depicted in Figure 1. This shows the demand curve, D(P), for electricity at a node in MW as a function of the price of electricity in $/MW. The capacity C of the grid to service the node is represented by the vertical dotted line. The demand curve after the managed reduction in load is represented by D(P,s). The demand curve is depicted as having a sloped portion and a vertical portion to reflect demand bids that do not have prices attached to them and so will be served at any price, if possible (i.e. the vertical portion of the curve).

52. In the stylised example in figure 1, demand is depicted as shifting back by more than is necessary to keep within capacity constraints. This is roughly what happens in practice, because managed demand reductions are not very precise. This imprecision and a lack of price responsive demand also mean that the price (P) which is struck is likely to undervalue the costs of congestion to those consumers whose load was shed. That is, they would have been prepared to pay more than P for the unused capacity, but cannot do so, which is inefficient. In

figure 1 the cost of investment is also included to emphasise that the option to invest (to increase capacity from C to C+I at a unit investment cost of i) should be preferred to load shedding if the cost is lower than the cost of load shedding.

Figure 1: 'infeasibilities' solved with managed load reductions



53. The Authority is currently considering improvements to wholesale pricing which will improve the efficiency of pricing including the accuracy of congestion pricing signals. These improvements include:

    a. prices being struck in real time[17], improving participants' ability to respond with certainty to price signals, and

    b. introducing default demand-side bids. These will better signal costs of curtailing demand when capacity constraints bind, node-by-node[18], and will encourage increased demand-side bidding because pay-offs to demand-side bidding will increase (consumers will, for example, face the prospect of higher prices during periods of congestion but will have improved information upon which to react and avoid these prices).

54. These improvements would ensure that nodal prices better reflect the scarcity of capacity – depicted in figure 1 by the introduction of a default demand-side bid price (P default). .

---

[17] As is explained earlier, there is currently a lag between real-time decision making and prices being finalised, two days later.

[18] Currently, if demand cannot be served, scarcity prices come into play to signal the cost of lost load but they are only activated when there is a national or island-wide event, rather than node-by-node.

Participants in the wholesale market will know that the default bid can set prices. They will then factor this into their own bids and offers and can avoid this high price – thus increasing demand response and the speed of demand response.

55. That said, it is worth considering whether options are needed for additional pricing mechanisms to control demand as a transitional measure or as a back-stop if the real-time pricing initiative does not function as expected.

56. The 2016 TPM proposal proposed removal of explicit (RCPD) peak pricing from transmission pricing. This would mean a change to the pricing signals which distributors currently see. Most distributors do not currently face wholesale energy prices.[19] As transmission customers they do, however, currently face peak pricing signals from Transpower.

57. So, removal of the RCPD peak prices is likely to raise concerns that distributors will not be incentivised to manage demand on their networks in the way many currently do to avoid high charges at peak. This concern is understandable. It would be desirable for the Authority to facilitate arrangements between distributors and those who would benefit from using ripple control (such as retailers and end consumers) to moderate peak nodal prices so as to ensure that ripple control is used to the benefit of end consumers. Failing this, it can be expected that as technology improves, it will increasingly allow aggregators to work on behalf of end consumers to bid demand response into the wholesale market. This should be facilitated by the introduction of real-time pricing. In fact, real time pricing should improve the volume and efficiency of demand response by allowing market participants to decide for themselves the value they place on access to an additional unit of electricity and by ensuring that the market filters out less valued uses of electricity before more valued uses (as opposed to the current administratively and retrospectively determined peak demand charges).[20]

58. In the meantime, there is a potential transitional issue. While it can be expected that distributors and other providers of load response, such as aggregators, will respond eventually, it seems unlikely that they will respond immediately if a new TPM is introduced. This could lead to an unexpected increase in demand and increased administrative load shedding. This risk may be overstated, since transmission users will still face nodal prices and distributors will have an incentive to avoid inefficiently interrupting their customers' supply. However, to take account of the risk of an unexpected demand surge, a new TPM could include a temporary (but potentially long-lived) demand control charge to replace the RCPD charge. Presumably this charge would initially look similar to the previous level of the RCPD charge. If this were phased out over time, it would allow Transpower to monitor developments and respond in real time to any unexpected increase in demand resulting from its being phased out.[21]

---

19 They do face price signals for interruptible load, as part of the instantaneous reserve market. We understand that some distributors offer interruptible load and so do face wholesale prices.

20 This of course relies on the system operator being prepared to forego administrative demand reductions and let nodal prices work.

21 Transpower has expressed concern that the removal of the RCPD charge may lead to unexpected demand shifts that lead to instability in the grid. The Authority commissioned Concept Consulting Ltd to investigate this concern. Concept found that at an island-wide level, demand was likely to remain within capacity. However, this doesn't rule out the possibility of problems at a more granular level. Introducing a supplement to nodal prices outside SPD that phases down is a way of minimising any risk of disruption at a relatively low cost.

Prices are not going to be unmanageably high

59. The second concern outlined in paragraph 49 is that high nodal prices will draw negative public and political reactions and are therefore not a sustainable approach to signaling costs of use.

60. Concerns about political sustainability are not well articulated. Little if any evidence is ever provided to support this view.

61. It is true that, in overseas jurisdictions, periods of high electricity prices have led to governments imposing price caps on electricity prices.

62. It is also true that nodal prices can be very high for very short periods ("the price-spike issue"). In principle, this may call into question the sustainability of such prices due to the costs being highly concentrated and highly visible while the efficiency gains of well targeted prices accrue much more widely and over time and are less visible.

63. However, the electricity market has been in place for place for some time, and the high nodal prices that occasionally occur appear to be broadly accepted.

64. In part, this may be because many mechanisms have been put in place to allow parties to manage their exposure to nodal prices. These include hedges, financial transmission rights and stress testing. These allow users to transfer the risk of high prices and the need to adjust use to mitigate the high prices to those on the other side of the hedge who are prepared to accept the volatility of nodal prices. Risk transfers like this are efficient. With these products in place, parties who choose not to make use of them have only themselves to blame for their exposure to nodal price volatility.

65. In effect, most retailers currently offer hedges to consumers (especially to small consumers such as households). They do this by giving consumers access to fixed price variable volume contracts. If consumers choose fixed price contracts because they don't like price variability and retailers can profitably provide them, then that is efficient. The availability of such contracts should substantially mitigate any concerns about political sustainability of high nodal prices.

66. The price-spike issue will also become less apparent as real-time pricing is developed and as technology is introduced which allows price-sensitive consumers to respond more to real-time price spikes. As they respond, the volatility of prices will decrease.

Early investment could be incorporated into nodal pricing constraints

67. The third concern identified in paragraph 49.a is that users never see the full costs of their actions because investment is usually triggered 'early', before expected nodal prices have risen to levels commensurate with signaling that additional investment would be beneficial.[22] This leads to the claim that there is a 'missing price signal' that can be provided by long-run marginal cost (LRMC) charges.[23] The mechanics of this 'missing price signal' is illustrated in figure 2 below. As in figure 1, this shows the demand curve, D(P), for electricity at a node in MW as a function of the price of electricity in $/MW. The capacity C of the grid is represented by the vertical dotted line. This leads to the nodal price P.

---

[22] Some argue that economies of scale necessarily mean that investment must take place before nodal prices have risen to levels that would signal additional investment is beneficial. As is discussed further below, this is not correct. Provided demand and supply at each node is sufficiently responsive to nodal prices, there is no technical reason why nodal prices can't continue to rise to restrict demand to capacity until new investment is justified.

[23] The 'un-priced' administrative load reduction discussed before is another reason for this 'missing price signal'.

68. It is assumed that for some reason an investment in increased capacity is triggered at this point. Incremental investment costs (i) are higher than the nodal price at which investment is triggered. The benefits of investment – in terms of demand which can be served – is smaller than the cost of investment.[24] Nonetheless, the investment proceeds.

69. It is often claimed that the 'missing price signal' reflects uncounted or unmeasured benefits from investment to ensure system security constraints. This claim is mistaken. Nodal prices take account of system security constraints. Even if it did not, the SPD model could be modified to incorporate any missing constraint and so efficiently incorporate the effect of the constraint in nodal prices. As such the capacity limit depicted in figure 2 includes limits based on system security constraints.

70. Rather, investment may be undertaken sooner than is efficient, ostensibly because the cost to consumers of investment errors is asymmetric. That is, costs of inefficiently low investment are larger than the costs to consumers of inefficiently early investment. Investment in transmission networks is often large (lumpy) with long lead times from when a decision is made to invest to when new capacity comes on-line. Perfectly timed investment is not possible, so investment decisions need to take account of this inevitable uncertainty and factor it into the costs and benefits assessed for the investment.

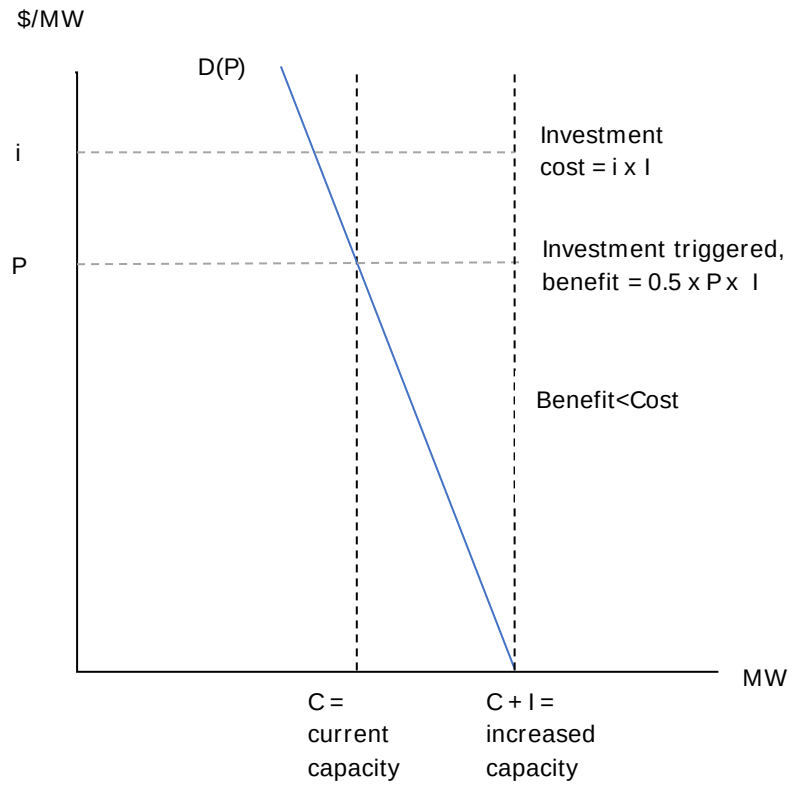
71. It is not clear why investment should be inefficiently early.[25] Even if the costs to consumers of unduly low investment are very high in some circumstances, then in the benefit calculation to determine whether investment is justified, scenarios in which this occurs will reveal large benefits to consumers from undertaking the investment.

---

[24] The benefit of the investment is the amount users are collectively prepared to pay for using the additional investment, which is the area of the triangle bounded by P and by C and C+I on the horizontal axis or ½ * P * I, on the assumption that the same generators continue to supply the energy.

[25] If nodal prices are allowed to operate, consumers should rarely have their access to electricity rationed. More precisely, there should be a large decline in administrative load control and outages, because in situations which currently cause load control or outages, nodal prices will restrict demand to capacity. That is, some consumers will choose not to consume as prices rise. There will still be administrative load control and outages, because the cost of providing 100% reliability would outstrip the benefits. The calculation to decide whether to undertake new investment should include trading off the benefits of reducing outages against the cost of avoiding them.

Figure 2: Investment triggered before benefits outweigh costs



72. But assume for the moment that investment is inefficiently early. Then the trigger for the early investment cannot be nodal prices or other charges for use of the grid, since these prices simply restrain grid use and do not themselves trigger investment. Rather there must be something else triggering the investment despite nodal prices signaling that the investment is inefficient. The appropriate policy response is to rectify the problem that is triggering the investment, not to inefficiently restrict grid use.[26]

73. The current TPM does not provide any incentive for Transpower to propose efficient investments. It relies on the Commerce Commission investment approval regime to ensure Transpower's investment proposals are efficient. However, under the 2016 TPM proposal, the beneficiaries of an investment would have to pay for it, so they would likely oppose it if they expect their charges to exceed the benefit they will get from it. This should help discourage inefficiently early investment.

74. However, if despite this, it is considered that there is a need to provide a price signal to consumers to discourage use at peak and so early investment, this issue does not need to be addressed with an LRMC charge. A more efficient way to reflect the 'missing price signal', rather

[26] The current Investment Test for the core grid may result in investment being triggered inefficiently early. This test relies on an administrative test to determine when reliability investment is required. If that test does not in fact reflect the economic benefit of reliability, Transpower may propose and the Commerce Commission may approve a reliability investment that the supposed beneficiaries of the investment are not prepared to pay for (ie, that is inefficient). The Investment Test is based on, and so the investment results from, the grid reliability standard that the Authority is responsible for.

than imposing a price, would be to tighten capacity constraints when calculating nodal prices. [27] As such, the cost of congestion would be directly incorporated into nodal price discovery processes.

75. If the Authority decided that this was desirable, it would require a Code amendment to allow for changes to system constraints (see next section).

76. The LRMC solution to this 'missing price' problem is to levy an administrative charge that reflects the (discounted) cost of the investment that will be triggered and the timing of that investment.

77. LRMC charges are often proposed as a way of promoting efficient use of the grid. The purpose of these charges is to provide grid users with a signal of the costs they create by choosing to use the grid at times of limited capacity.

78. Under LRMC-based pricing, users of the grid are charged a fee which is proportional to costs of new capacity that will be needed to serve increased demand. This fee is targeted at periods of peak demand and increases as the line becomes increasingly congested.

79. From the view point of nodal pricing in the wholesale market, this has the effect of causing users to adjust offers. This is depicted in figure 3 which shows demand shifting inwards to $D(p,i^*)$ as a result of imposing the LRMC charge i^*.

80. There are several consequences of imposing LRMC charges.

81. First, the addition of the estimated efficient LRMC charge (i*) means that customers' willingness to pay (bids) for capacity and for electricity is obscured. It is only possible to observe demand given the LRMC charge imposed to reflect expected costs of transmission investment. Expected costs of transmission investment should reflect demand. But they do not. They reflect demand given the LRMC charge. Actual willingness to pay is no longer directly identifiable from looking at demand bids. This undermines the value that is gained from having nodal prices which reveal consumers' actual (minimum) willingness to pay for electricity and transmission services.[28] This matters because accurate measures of consumer demand and willingness to pay are important inputs into investment decisions.

82. Second, the charge will be subject to forecast errors. This is because it needs to be set based on forecast demand growth and the forecast cost of investment to meet that demand growth. This would not normally be a problem if the LRMC charge was a better approximation to the 'correct' price than not having it. This is not the case as the next paragraph indicates.

83. Third, where the LRMC charge is less than what the nodal price would have been in its absence (called the 'default nodal price' in what follows), the LRMC charge has no effect at all, except to reduce the nodal price by the extent of the LRMC charge. The LRMC charge only has an effect on demand when it is more than the default nodal price, and it has more effect the lower the

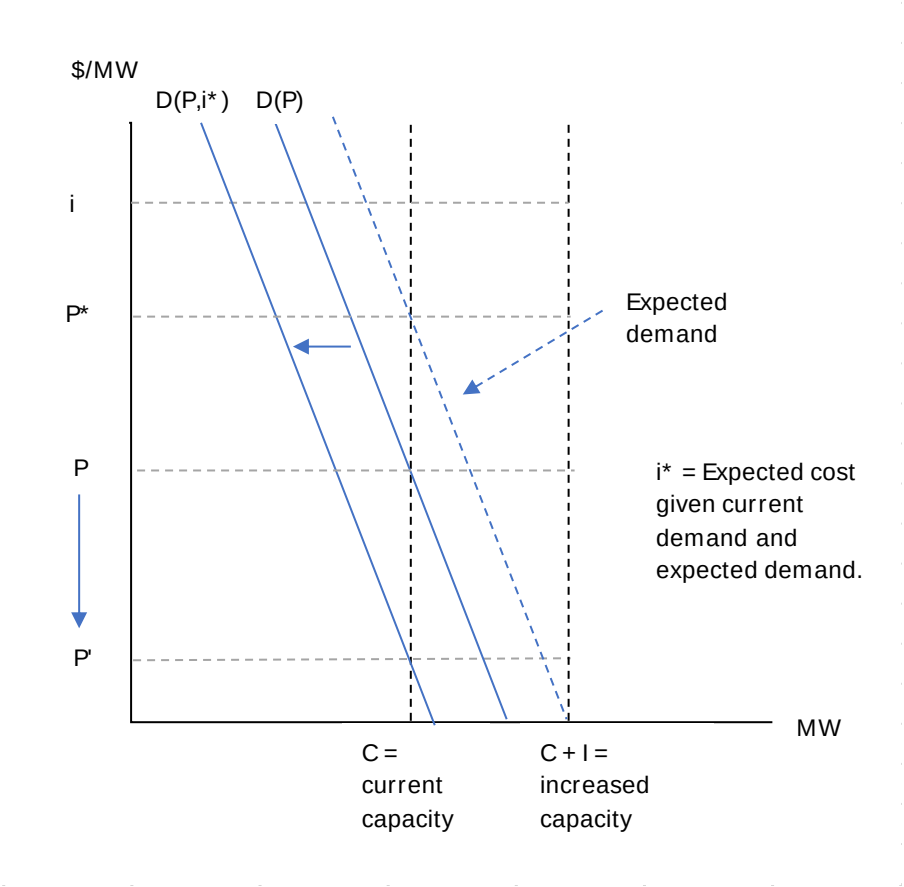
---

27 In principle, this could be done either by tightening quantity constraints or by raising prices – via constraint violation penalties or perhaps increasing the value of default demand bids when real time pricing is operating.

28 If all users respond to LRMC charges the same way that they respond to nodal prices it may be possible to accurately infer underlying (pre-LRMC) demand and willingness to pay – working back from what is observed in the wholesale market and the value of LRMC charges. But this is likely to be inaccurate if consumers adjust demand, to avoid LRMC charges, based on expectations that the market will strike some pre-set peak demand conditions or metrics (such as a coincident peak) that trigger imposition of the LRMC charge. Uncovering demand and willingness to pay would then require understanding consumers' underlying expectations about their exposure to LRMC charges.

default nodal price.[29] In these cases it reduces the use of the grid to below capacity. That is, it inefficiently reduces use of the grid. In particular, if demand is in fact never going to reach the investment trigger point, then the LRMC charge would be highly inefficient because it would reduce demand for no reason at all.

Figure 3: Effects of LRMC charges on demand and nodal prices



84. This inefficiency in use is a price that some proponents of the LRMC charge are prepared to pay to signal that there will be higher prices in future caused by congestion and future investment. This is discussed further later in the paper. Other proponents of an LRMC charge also see the reduction in nodal prices caused by the LRMC charge as a virtue, because it obfuscates how high nodal prices actually are, and so mitigates the problem discussed under the heading

85.

86. *Prices are not going to be unmanageably* high above.

87. Much has been written about the merits of LRMC pricing and various practical difficulties and design choices.[30] But few proponents reflect on the fundamental information problems that

---

29 In the Authority's LRMC working paper, it was proposed that the LRMC charge would be applied only when the grid was expected to be congested. However, unless the LRMC charge was applied half-hourly, like nodal prices, it would likely also apply in half-hours when the grid was not congested. Moreover, it is difficult to see how you would apply the charge half-hourly and only to the extent that the grid was congested without replicating nodal pricing.

30 The details do not need to be repeated here. Recent contributions on the conceptual merits can be found in the reports by Axiom Consulting Ltd, for Transpower. Discussion of both the merits and the practical details can be found in a report by Sapere for Transpower https://www.transpower.co.nz/industry/transmission-pricing-methodology-tpm/sapere-lmrc-pricing-paper-february-2018 . The Sapere report acknowledges that in principle SRMC can constrain

come with LRMC charges and the fact that they do not reflect efficient costs of investment but rather estimates of efficient costs of investment. These charges are invariably based on a highly imperfect 'price' discovery process. LRMC pricing is a one-way street. Pricing proceeds by assuming that the future demand, technology, and investment costs are known. Of course, they are not known. In contrast, nodal pricing reflects actual demand and supply conditions as expressed by generators and consumers.

Consumers are no more likely to react to an LRMC charge than nodal prices

88. The fourth concern expressed in paragraph 49 is that consumers don't actively monitor nodal prices and so will not respond to changes in prices. The claim is made that LRMC charges can improve the clarity of price signals to consumers, beyond what they would see in nodal prices. This argument relies on two conditions being fulfilled:

    a. that users will perceive the LRMC charge despite not perceiving nodal prices

    b. that supplementary prices will make pricing more efficient.

89. It is unlikely that either of these conditions would hold. In the first instance, if mass-market consumers do not perceive and react efficiently to congestion costs in nodal prices it is unclear that they would perceive and react to an LRMC charge bundled along with distribution prices and energy prices. Furthermore, for the reasons outlined in the previous section, an LRMC charge is likely to have either no effect or is likely to inefficiently suppress demand.

90. One argument sometimes advanced for an LRMC charge is that it is more simple, stable and predictable than nodal prices, and so mass market consumers are more likely to respond to it. It is not clear why this would be so. In any case it ignores the concern that this response by consumers will be inefficient unless the period happens to coincide with what would be high nodal prices.

91. In addition, stable and predictable charges for mass market consumers can emerge efficiently from the workings of the electricity market. As is discussed in paragraph 65 above, mass market consumers may choose to enter into a fixed-price variable-volume contract with retailers (and currently, most do). Some consumers may well end up facing prices which have the same or similar characteristics to an LRMC charge, such as peak-period pricing. So simplified prices, reflecting costs of peak demand or congestion, can emerge without being part of a transmission pricing methodology. In this case, the price risk is borne by some other party, who is therefore

---

the grid to capacity. It states at page 7 that 'In principle it is possible to charge simply on the basis of SRMC, relying on high SRMCs during congested periods to drive investment in capacity to relieve such congestion." However, it assumes that nodal prices will not be allowed to rise to this level. It states that SRMC does not provide a good signal for long term investment "due to the fluctuating nature of SRMC" (page 6). These issues are noted in paragraph 49 and our views on them are discussed in this paper. In particular, Sapere's point about the inefficient investment signal is incorrect. In industries with large lumpy investments there is often a trade-off between short run efficiency and dynamic efficiency – because investors cannot recover the costs of their investment if they have to price at (efficient, short run) marginal cost. So investment may be undersupplied. The case might then be made for a prospective investment-related charge (an LRMC charge) to promote investment. But in the case of transmission investment this is dealt with via overall allowable rates of return and the regulated price-quality path which allows Transpower to price its services above short-run marginal cost. In the Authority's 2016 TPM proposal, these charges are levied in a way that does not cause short run prices to deviate from efficient levels. With the investment decision dealt with, dynamic efficiency, in terms of pricing for using the grid, requires only that short-run efficiency is achieved over time.

incentivised to respond to the scarcity signaled by high nodal prices. If such arrangements emerge, they are likely to be efficient.[31]

92. In addition, as discussed above, it can be expected over time that aggregators will emerge to undertake load control on behalf of consumers. This is also efficient.

93. More sophisticated consumers can and anecdotally do respond to nodal prices. This includes the parties on the other side of the fixed price variable volume contracts discussed in paragraph 91.

Consumers may not correctly anticipate nodal prices

94. The fifth issue identified in paragraph 49 is that consumers don't accurately anticipate nodal prices and they make long-lived investment decisions which will eventually cause congestion and higher prices, but they do not take that into account because they do not know that this will happen. If they were exposed to a LRMC price signal based on future investment costs, then they would make different decisions and those decisions would lead to more efficient outcomes compared to relying on nodal prices and on consumers' ability to predict the effects of their decisions on congestion and on nodal prices. The claim is made that LRMC charges imposed before the relevant circuits become congested can improve the clarity of price signals to consumers, beyond what they would see in nodal prices.

95. Mass market consumers are unlikely to devote time to forecasting future transmission charges. Retailers will face strong incentives to accurately anticipate future transmission charges when they set retail prices, but are unlikely to look much beyond the contractual term of their contract with their consumers.

96. Accordingly, if nodal prices and transmission charges are expected to rise some years into the future, it can be expected that mass market consumers will not anticipate the increases. If mass-market consumers have to make major long term investment decisions (eg, how well to insulate a new home, or whether to invest in a heat pump) and they use the current price as an indicator of future prices, then imposing a LRMC charge before grid circuits become congested may well lead them to make different (and proponents would argue) more efficient decisions. Proponents further argue that the efficiency benefits may be individually small, but collectively will be substantial.

97. The case for a forward-looking investment charge is often made by claiming consumers need to accurately forecast transmission costs and charges to make efficient investment decisions. In particular, if they are poor forecasters, the grid owner or some other central authority could do that forecasting for them and levy a LRMC charge which signals those costs, thereby improving the efficiency of demand-side investment decisions.[32]

98. However, it is important to distinguish between the congestion charge and the transmission charge the user faces. After the new investment is commissioned, the congestion charge a

---

31 The point is that mass-market consumers do not need to face nodal prices for efficiency. It is necessary for retailers to face nodal prices, but if they can profitably offer consumers fixed price, variable volume contracts by absorbing or passing off that risk, that is likely to be efficient.

32 For major users, it seems that much the same effect on expectations of future prices could be obtained without the inefficiency in current use by simply providing consumers with the information about what future investment is expected to do to transmission charges and when. This would also allow customers to make assessments in real time of the impact of changes in market conditions.

consumer faces is likely to reduce to near zero – the same as it was before the relevant circuits first became congested. That is, a naïve consumer that uses the current congestion charge in this circumstance to predict future prices will be making an unbiassed prediction of the future congestion charge they will face.

99. One thing that this naïve consumer will get wrong is the increase in the congestion charge leading up to the new investment. This is not normally an issue identified by proponents of an LRMC charge. However, attempting to deal with this is not easy, because:

    a. when the grid is uncongested, the consumer will under-estimate the future congestion charge during the period that the grid is congested (before the congestion-relieving investment is made)

    b. when the grid is congested and new investment is imminent, the consumer will over-estimate the future congestion charge during the period that the grid is uncongested (after the congestion relieving investment is made).[33]

100. Another thing that the consumer will get wrong is their share of the cost of the new investment. However, since this is a fixed charge independent of use, it should not affect consumers' major long-term investment decisions.[34] That is, the consumer who naively uses the current congestion charge as a forecast of the future congestion charge will make an investment decision that is appropriate once the new transmission investment is in place.

101. More sophisticated consumers will be able to assess potential increases in congestion charges, including with reference to developments in the FTR market and trends in locational price differentials.

102. Furthermore, the introduction of an area-of-benefit charge is intended to provide additional incentives for more sophisticated consumers to consider the effects of their decisions on transmission investment and to scrutinise investment decisions. This reduces problems for market durability which can arise when people are surprised by costs of investment.

103. However, one thing that both naïve and sophisticated users will not take fully into account is the effect of their decision to invest in a more or less energy efficient investment on the need to invest in future in new transmission investment. Small individual users cannot much affect the timing of new investment. However, if users could coordinate, it may be that it may pay them to collectively invest in more efficient technologies because that would avoid or defer future transmission investment and so save them transmission charges in future. This would be the

---

[33] How important these biases are in practice is debatable, however, because it depends on how much each consumer's investment decisions are impacted by the change in energy price caused by changing congestion charges and how long those investments last compared to

a. for paragraph 65.a, the time until the grid becomes congested and the time until the new investment is made

b. for paragraph 65.b. how much congestion prices have yet to rise and how long they stay high (ie, until the new grid investment is made).

If it were considered important enough to justify a charge in case a. and a subsidy in case b., the charge or subsidy would relate to the expected change in congestion charges rather than LRMC.

Whether the decision would be more efficient overall is more problematic however. This is because any gains made from better investment decisions have to be offset by the cost of inefficiency in use caused by the tax or subsidy analogous to that discussed in paragraph 83.

[34] This assumes that the fixed charge is not variabilised by distributors. Any variabilisation of fixed charges by distributors will undermine the efficiency of the TPM. This should be taken account by the Distribution Pricing Principles.

case if the present value of the additional cost of investing in the energy saving technology is less than the present value of the transmission costs that would be saved.

104. This is discussed more in the next section.

Consumers can't coordinate their actions

105. The final issue identified in paragraph 49 is that consumers cannot take actions that reduce the cost of future transmission investment to them because they cannot affect the timing of the investment. So although they may be aware of the potential need for future transmission investment, they do not take account of the impact of their investment decisions on the need for the future transmission investment. As a result, they make long-lived investment decisions which will eventually cause congestion and higher prices, but because they are small, they do not take that into account. If these long-term effects were signaled to consumers via an additional price that applied consistently during peak periods they would make different decisions and those decisions, proponents argue, would lead to more efficient outcomes compared to relying on nodal prices alone.

106. The situation here is that consumers make long-term irreversible investment decisions (eg, about installing a low-efficiency heater instead of a more energy-efficient heater, such as a heat pump or air-conditioner, providing the same benefits) before the congestion charge has risen as a result of congestion. It could be, for example, that it if all consumers could agree to install a heat pump, that would sufficiently reduce peak demand to defer or avoid a transmission investment. Furthermore, it may be the case that the net benefit that the consumer gets from installing the heat pump instead of installing the low-efficiency heater is positive, once the cost of transmission is taken into account. This net benefit calculation would incorporate:

    a. the benefit of lower energy use, of lower nodal prices in the short-term and of the deferred transmission costs, and

    b. the cost of installing the heat pump relative to the low-efficiency heater.

107. However, if users cannot coordinate, each user will know that if they install a heat pump, it will have no effect on the timing of the transmission investment. Therefore, in undertaking the net benefit calculation, they will set the benefit of deferred transmission costs to zero. Furthermore, in assessing the benefit of the heat pump, they will assume that:

    a. the nodal transport charge after the transmission investment will be near zero, for the reasons outlined in the previous section

    b. they cannot avoid their share of the transmission cost for the new transmission investment, because by design it is intended to be unavoidable (although they can alter what their share is, as is discussed below).

    In other words, they will ignore these charges in undertaking the net benefit calculation.

108. The argument for imposing an LRMC charge in this instance is that it variabilises an approximation to the annual cost to the user of the future fixed transmission charge. The user will face this charge until the transmission investment is actually made. This means the user's individual private benefit calculation will take it (and so the benefit of deferring transmission investment) into account. This calculation will be the same as the net benefit calculation in paragraph 106, except that the term "benefit of deferred transmission costs" will be replaced by the term "LRMC charge".

109. This can be seen by assuming there is a single user of the new transmission investment who can buy an energy-using technology (which makes immediate transmission expansion efficient) or an energy-saving technology (which allows deferral of expansion for a number of years). Then the user's decision is efficient because the user alone affects the timing of the transmission investment. The user's decision is to buy the energy-using technology and expand transmission if the lower cost of buying the energy-using technology (including capital and energy costs) and the benefit the user gets from the increased capacity (ie, lower congestion charges and increased use[35]) is less than the present value of the savings from deferring transmission investment.

110. Where there are multiple users, each user takes account of the cost of the two technologies in the same way as the single user. However, because the user correctly assumes that they do not much influence the timing of the investment, the private calculation does not take account of:

    a. the present value of the deferred transmission costs

    b. the benefit that user gets from the early expansion of the transmission investment (in terms of lower congestion charges and consequent increased use of energy).

    The omission of these two terms can cause individual users' decisions to deviate from the efficient decision discussed in paragraph 109.

111. The annualised value of the savings from deferring the transmission investment is just the LRMC of the investment. So (ignoring the term in paragraph 110b), the LRMC charge makes it profitable to invest in the energy-saving technology if it is efficient to do so.

112. The purpose of the charge is to encourage investments by consumers (for example, purchase of energy-efficient appliances) that will allow transmission investment to be deferred. So, although the charge would need to be applied before the grid became congested, it would be targeted at uses that would likely lead to future transmission investment (that is, current investments that will be operating at peak up until the time the new investment takes place).

113. In practice, this means it would likely be focused on current peak demand, so as to encourage measures to reduce the growth in peak demand. Even so, as discussed in paragraph 83 above, it would only have an effect on grid use when the grid is not congested.

114. In addition it means that the charge should not be applied too far in advance of the proposed transmission investment, since it would clearly be inefficient to apply it to investments that did not survive long enough to affect the timing of the transmission investment. Clearly, this means that there would have to a compromise between affecting long-lived investments (such as house insulation) and short-lived investments.

115. Subject to the considerations discussed below, the result of applying the LRMC charge is that both a naïve user and a rational forward-looking user make investment decisions that result in the joint optimisation of their own investment decisions and the transmission investment decisions. In particular, it results in the user taking into account the time profile of the congestion charge, which falls to zero immediately after the investment is made, after rising from zero as the time of the investment approaches.[36]

---

35 The benefit from transmission expansion is explained in footnote 23.

36 If the user is not forward looking, they may over-react to the LRMC charge by assuming that the nodal transport charge would be positive even after the investment is made. This may lead them to install more energy efficient equipment than can be justified by the nodal transport charge.

116. However, there are other considerations that need to be taken into account.

117. First, as discussed previously, the LRMC charge necessarily reduces grid use below capacity. This loss of use is inefficient, and must be offset against the efficiency gains form co-optimising investment. In particular, it incentivises inefficient investment in short lived assets that are replaced before the transmission expansion is actually built. Given this, it is likely that the optimal charge should be between LRMC and zero.

118. Second, when the nodal transport charge rises to reflect increasing congestion, it signals to the user the benefit of investing in more energy-efficient technology, in addition to efficiently constraining grid use to capacity. This means that the total annual LRMC charge must be reduced by the total annual nodal transport charge that a user pays as the grid becomes congested.

119. Third, the benefit the user gets from an expansion in capacity, as discussed in paragraph 109 and 110, should also be taken into account. This starts at zero when the grid is uncongested and rises over time as the grid becomes congested. Since the LRMC charge on its own is intended to reflect the full cost of the future investment, the LRMC charge would need to be reduced by the annual value of the congestion charge, so the total charge equals LRMC. .

120. Fourth, although a sophisticated user cannot alter the timing of the transmission investment, under a beneficiaries-pay approach they can alter their share of the transmission charge by reducing the benefit they get from the proposed investment. For example, they may install battery equipment or distributed generation behind the meter to reduce the benefit they get from the new transmission investment. This incentive is stronger than is efficient[37] and works in the same direction as the LRMC charge. It also gets stronger as the time of the proposed investment approaches. It may therefore reduce or eliminate the need for an LRMC charge, particularly close to the time the new investment is commissioned.

121. The LRMC charge would only improve efficiency to the extent that it is a better approximation to the annualised cost of the actual investment eventually undertaken, adjusted by the considerations discussed immediately above, than not imposing any additional charge. Given the difficulties in estimating LRMC and in making many of the adjustments discussed above, this issue is not clear cut.

Situations where additional charges may be warranted

122. The discussion above indicated that there may potentially be situations where additional charges are required to signal costs of congestion and the impacts that demand may have on system costs. These situations would arise because nodal pricing may not operate as effectively as it needs to. This includes the situation outlined in paragraph 49.d and the transitional issue discussed in paragraph 58. That is, the situations where an additional charge may be justified are:

    a. to encourage users to co-optimise their investment with future transmission investment

[37] It is too strong because the user saves the average reduction in cost rather than the marginal reduction in cost as a result of their actions. This is the reason the marginal price signal for new investments was proposed on page 105 of the second issues paper. If that method was used to adjust the user's share of benefits, the incentive the user faced would be efficient and the rationale for applying the LRMC charge would continue to apply.

b. as a (possibly extended) transitional measure, to allow time for distributors and/or aggregators to start managing load on behalf of consumers.

These are discussed further below.

123. For all the others issues outlined in paragraph 49, there either does not seem to be an issue, or, if there is one, there are better policy options than introducing an additional charge to resolve it. In particular, if a change is required, the objective should be to integrate any additional price measures within the overall efficient pricing mechanism of nodal prices. This is because nodal prices are the only mechanism that can restrict grid use to capacity while never inefficiently restricting use when the grid is below capacity. This could involve changes to SPD and in some situations a Code amendment may be required.

Co-optimisation of user investment and transmission

124. There does seem to be a potential case for a charge based on LRMC to encourage users to co-optimise their investment with future transmission investment. However, the potential benefit of this would need to be traded off against the distortion to current grid use it would involve. In addition:

a. it would need to be reduced for several reasons including the increasing size of the congestion charge as the time of new transmission investment approaches.

b. for all users and especially for sophisticated users, it is likely that the charge would reduce and may become zero as the time of the investment approaches.

125. This, together with the difficulty of ensuring the estimate of the LRMC charge is reasonably accurate, means that although there is a potential case for an LRMC charge to encourage users to co-optimise investment, there is a very real risk of getting it wrong. There is therefore a risk that the implementation of the charge in practice would be less efficient than not implementing it. A careful analysis would therefore be desirable before it was introduced.

Transitional measure

126. As is discussed earlier in paragraph 58, there may also be merit in allowing for the possibility of a supplementary transmission charge (different from the LRMC charge) which can ameliorate the initial impact of removing the RCPD charge and address potential temporary limitations on nodal pricing for signaling costs of congestion.

Appendix A Short-run congestion costs and long-run investment costs

A.1 This appendix discusses the relationship between short-run congestion costs and investment costs.[38]

A.2 This is important for clarifying that there is in most cases no in-principle reason for having prices which reflect incremental investment costs. If investment proceeds efficiently, users of the grid will face costs which reflect investment costs.

A.3 Arguments in favour of prices based on incremental investment costs are essentially pointing to the need to resolve practical pricing problems. Those problems, to the extent they exist, in most cases may be better resolved by addressing the source of the practical problems (insufficiently cost-reflective congestion prices) rather than adding additional charges based on error-prone estimates of investment costs.

Congestion costs signal efficient investment costs

A.4 Figure 4 is adapted from a paper by Nobel laureate Oliver Williamson (1966)[39] who considered optimal peak-load pricing with lumpy investment (and hence periods of constrained capacity).

A.5 Short-run costs of supply are zero except where capacity is constrained – at which point supply costs are infinite and the efficient price level can only be determined based on the price that is needed to keep demand inside the capacity limit (eg, C). This is the so-called shadow price of the constraint.[40]

A.6 The diagram includes a per-unit price of capacity denoted i, which, in this depiction, is equal to long-run average costs. This per-unit price captures the opportunity cost of investment.[41]

A.7 In figure 4, there is a once-and-for-all change in demand equal to an amount z (a static partial-equilibrium perspective). It is efficient for this shift in demand to trigger an investment in capacity if the change to total surplus (the change to consumer surplus plus producer surplus) from investment is larger than the cost of investment. In figure 4 by construction AB is just larger than DE. This means that the shift in demand (from D(p) to D(p) + z) is sufficient to meet this condition. This is because the change in consumer surplus together with the loss in producer surplus (when prices fall from P to P' and the quantity supplied increases from C to C+I) is fractionally larger than the cost of the new investment I*i.

A.8 For any positive change in demand x above D(p) but less than D(p)+Z (ie, z>x>0) a new investment in capacity would create a positive change in electricity consumers' surplus. But until demand shifts the full amount z – and prices reach P' – the investment is not socially beneficial. That is, for any given change in demand, the cost of investment and the benefit to consumers from relieving the constraint both need to be weighed against each other to

---

38 It is important to be aware that 'long-run' does not necessarily mean a long time. In this context it is the fact that capacity changes (up or down) and investment can take place that distinguishes the long run from the short run.

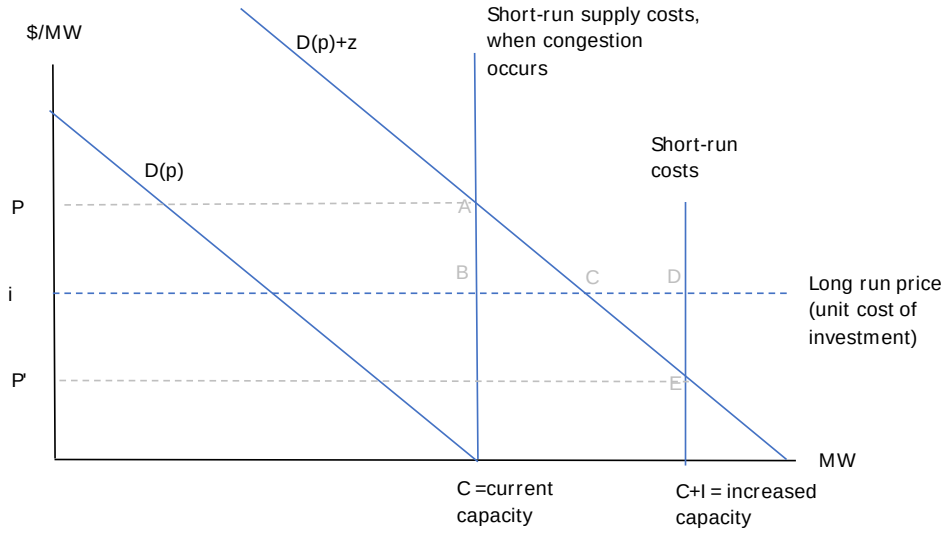
39 Williamson, O. (1966). Peak-Load Pricing and Optimal Capacity under Indivisibility Constraints. The American Economic Review, 56(4), 810-827. Retrieved from http://www.jstor.org/stable/1813529

40 Williamson's initial set-up and, he argues, his addition to the literature involves an explicit social welfare framework in which the constraint on maximising total surplus is the condition that (in the absence of new investment) demand remain less than some maximum level C.

41 This opportunity cost is the economy-wide price. Note that while investment is lumpy in this sector, investment elsewhere may not be.

determine the efficient rate of investment. And the price needs to rise sufficiently high – above unit investment costs – before the investment is efficient.[42] This then also means that consumers who face congestion prices will face congestion prices equal to the cost of investment assuming that investment is efficiently timed and not carried out too quickly. And consumers may face costs that are inefficiently high if investment occurs too slowly.

Figure 4: Connection between congestion costs and investment costs over time



LRMC was a solution when short-run pricing was not an option

A.9 Ralph Turvey, one of the most cited proponents of LRMC pricing, had no problem with the general analysis in figure 4. Commenting on the work of Williamson (1966) he said:

> 'The basic notion which he and his predecessors put forward is fully accepted, given his assumptions. This is that the optimum requires price to exceed marginal running cost in periods when demand is high by amounts which will both restrict demand to capacity output in all of those periods and which sums up over them to equal the marginal cost of capacity. In other periods, price must equal marginal running cost' (page101).[43]

A.10 However, Turvey promoted LRMC pricing because he thought that short-run congestion pricing was an impractical 'ivory-tower' solution:

> 'In purest principle, prices should be varied at very short notice to meet short-run fluctuations in demand curves…. But it is scarcely worth pursuing this thought, on two grounds. The first … is the cost of charging customers along these lines. The second is that consumers may prefer a simple tariff.' (page 105)

---

42 Note that this is the average price for (say) a year compared to the annualised cost of investment. The peak price would be higher again.

43 Turvey here slightly mis-states Williamson's position, which is that the strict equality between prices and cost of capacity only holds at the point where investment is justified.

A.11 Turvey stated that an electricity tariff can have no more than four different prices per year, 'except for very large consumers where the expense of recording load hour by hour can be borne.'

A.12 Turvey's objection is clearly outdated. Locational marginal prices are feasible, so there is no need to approximate them with an LRMC charge. Or, alternatively, an LRMC-based charge may be useful when locational marginal prices are not feasible.

With efficient congestion pricing, other charges make much less sense

A.13 This point has been made forcefully by Professor James Bushnell in his comment on the Authority's options working paper (also referenced in the section of this note discussing the effectiveness of nodal pricing). Bushnell states:

> 'However, what's missing from the EA's discussion is the fact that in those previous applications peak-load pricing needed to be applied in the absence of any other pricing that could properly capture the costs of capacity in a dynamic way. With regards to transmission, however, New Zealand already has a mechanism for the real-time pricing of scarce capacity. It is not efficient to leave that constraint out of the nodal prices and instead recover through a surcharge, even if that surcharge is targeted at hours of peak flow. If some of those peak flow hours are not congested, then the transmission asset will go underutilized and inefficiencies (both long-run and short-run) will result. Long-run efficiency cannot be improved by imposing a series of inefficiencies in the short run.
>
> One justification for a transmission surcharge under this framework would be if the penalty price for violating the relevant constraints were not set high enough in the nodal pricing algorithm. This would create artificially constrained scarcity prices analogous to those produced in energy markets with relatively low price caps. Following that comparison, the LRMC proposed here is very analogous to a capacity payment in a price-constrained energy market.
>
> I am not aware of claims that the penalty prices in New Zealand's market are artificially low, as evidenced from the general lack of support for capacity markets and level of comfort in the current energy-only paradigm for generation investment. Adding additional charges to the LMP would then constitute pancaking capital costs on top of scarcity charges that are already implicit in the nodal prices, leading to inefficient under-utilisation of resources. Investment planning in turn would have to consider the value of new transmission assets that would themselves be potentially under-utilized because of future LRMC charges. Another possible justification for pricing above the efficient scarcity level in the current period would be if consumption decisions for future periods were based upon current prices.
>
> In summary, the very purpose of LMP is to properly capture in real time the shadow costs of all relevant network constraints. Transmission investment decisions should in turn ideally be based upon the stream of values provided by that asset. If investments are being driven by a constraint, it should be represented in the LMPs.…
>
> …. aspects of the proposal reflect good intent to provide appropriate marginal signals, but continue to reflect the EA's reluctance to acknowledge that the LMP system already in place represents the 'gold standard' for appropriate marginal pricing incentives. In particular, the desire to add an additional surcharge to LMPs under the guise of long-run marginal cost, or benefits-based charging, is only justified on efficiency grounds if there is a fundamental problem with the LMPs, or the ways network

users, and planners, respond to them. If this were the case, the proper response is to address the source of the problem (LMP penalties, the transmission planning process) rather than attempt to correct one distortion by adding another one.'