PRICING OF ANCILLARY SERVICES AND THE IMPACT OF WIND GENERATION ON THE CAPABILITY OF THE TRANSMISSION NETWORK

Darryl Biggar[1]

27 May 2010

It is widely recognised that increasing penetration of wind and other intermittent generation may increase the demand for balancing services. It is less widely recognised that increasing penetration of intermittent generation may reduce the operating capability of the electricity transmission network. In the Australian National Electricity Market (NEM), balancing services are purchased on a NEM-wide uniform-price basis. As a consequence sufficient spare capacity or operating margin must be maintained on the transmission network at all times to ensure that these balancing service can be 'delivered' to where it is needed. Any increase in the demand for balancing services results in an increase in the operating margin and therefore a reduction in the capability of the transmission network. This paper argues that this results in inefficiently low transmission capacity and inefficiently high costs of transmission congestion. Instead, overall efficiency should be maximised through localised procurement of balancing services, co-optimised with the transmission limits in the dispatch process. Furthermore, the resulting increased congestion costs should be passed back to the intermittent generators who cause this increased cost so as to ensure a level playing field across renewable technologies.

1. Introduction

The Australian National Electricity Market (NEM) encompasses the electricity supply industry in the five eastern and southern states of the country of Australia. In common with many governments around the world, the Australian Government has committed to policies which have had and will have the effect of significantly increasing the volume of energy in the NEM produced from renewable sources.[2]. In the short to medium term it is expected that most of this additional renewable generation will be in the form of wind generation. As of December 2009, there was around 1500 MW of installed wind generation capacity in the five regions of the NEM. This is forecast to increase to around 6000 MW within five years.

---

[1] Economist, Australian Energy Regulator (AER) and the Australian Competition and Consumer Commission (ACCC). The views expressed in this paper are those of the author and not necessarily those of the ACCC or the AER.

[2] These schemes included the Renewable Energy Target (RET, an obligation on retailers to purchase a proportion of their electricity from renewable sources) and the Carbon Pollution Reduction Scheme (CPRS, a cap-and-trade system for limiting carbon emissions). See http://www.climatechange.gov.au/government/initiatives/renewable-target.aspx and http://www.climatechange.gov.au/government/initiatives/cprs.aspx .

Figure 1: Forecast wind penetration in the NEM, as at January 2009

Source: Carbon Market Economics (2009)

As in many other countries around the world, this proposed large increase in the penetration of wind generation has raised concerns about the continued security and reliability of the operation of the network. This paper focuses on one aspect of that problem which has been largely neglected – the impact of such a large volume of intermittent generation on the operating capability of the transmission network.

This link between the volume of intermittent generation in the market and the operating capability of the transmission network arises from the way that balancing services are purchased in the NEM – particularly the fact that balancing services (which in the NEM are referred to as frequency control ancillary services or FCAS) are, under normal circumstances, procured on a NEM-wide uniform-price basis. As a consequence, sufficient spare capacity must be maintained on the transmission network at all times to allow the balancing services to be delivered to where they are needed.[3] Any increase in the demand for balancing services automatically results in an increase in the operating margin or headroom on the transmission network and therefore a reduction in the operating capability of the transmission network.

This reduction in the operating capability of the transmission network imposes a real economic cost, in the form of higher levels of transmission congestion, increased opportunities for market power, and increased need for new investment in transmission. Under the causer-pays principle, this cost should be allocated back to the underlying cause – the variability in output of intermittent generators. In the absence of a mechanism for pushing these costs back on to those responsible, wind generation is artificially favoured over other, more reliable forms of renewable generation. Put another way, there is an absence of a 'level playing field' between different renewable technologies.

The economic impact of intermittent generation technology on the capability of the transmission network could be reduced by moving to an approach under which balancing services are procured locally, using a form of nodal pricing for balancing services. In addition, as shown in this paper, the operating margin (or, equivalently, the capability of the transmission network)

---

[3] In practice, since the frequency of the interconnected power system is the same in all locations, balancing services do not strictly need to be "delivered". Nevertheless, the provision of balancing services changes flows on the transmission network in exactly the same way as if power was being delivered from the generator providing balancing services to the source of the imbalance. Therefore we can use the intuition that balancing services are being "delivered" to where they are needed.

should be co-optimised with the cost of the balancing services to maximise overall market efficiency.

The next section sets out, by way of background, how wind generation variability affects the need for balancing services. The third section discusses the extent of variability that has been observed in practice. Section four shows how an increase in the need for balancing services results in a reduction in the operating capability of the transmission network. Section five concludes.

2. Background

Wind generation, of course, is highly variable in its output. This variability or "intermittency" is the source of several problems with integration of large amounts of wind generation capacity into the electricity supply industry. For any given wind speed it is usually possible to produce *less* than the theoretical maximum of the generating station by turning off some turbines, or feathering the blades, effectively "spilling" some wind energy. However, at any given moment in time, for any given wind-speed there is a maximum output of energy that a generating station can produce. The moment-by-moment variation in the speed of the wind leads to moment-by-moment variability in the output of a wind generator.

In any electricity market the volume of electricity produced must be equal to the volume of electricity consumed at each instant in time. In a liberalised electricity market such as the NEM, the primary mechanism for balancing supply and demand over time intervals longer than the basic 5-minute dispatch cycle is the wholesale spot market and, in particular, the wholesale spot price. At times when the wind is blowing strongly and/or demand is low, the wholesale price is typically low, causing other generators to voluntarily reduce their output or to shut down altogether. At times when demand is high and the wind is hardly blowing at all, the wholesale price is typically high, inducing higher costs generators to come on-stream or to increase their output.

The greater the penetration of wind generating capacity in a market, the greater the volatility of the wholesale supply curve. This increase in supply volatility (coupled with the familiar day-to-day volatility in the wholesale demand for electricity), increases the volatility in the wholesale spot price (or prices). This increase in price volatility makes short-term forecasting of prices more difficult, which, in turn, makes it difficult for plants to make efficient start-up decisions (at least in those markets, such as the NEM, in which unit commitment decisions are decentralised to generators themselves). It also makes it more difficult for energy-limited plant (such as hydro storage generators) to make efficient production decisions. However, putting aside these short-term forecasting issues, in principle, variations in prices provide efficient signals to generators as to the efficient operation and investment decisions – including efficient signals as to the type, size, and location of new investment. Therefore, in the absence of other distortions in the market[4], these price-mediated responses by market participants to increased penetration of intermittent generation will be efficient. In principle, therefore, on the time scale of intervals longer than five minutes, in principle there is no need for any particular changes in the market in response to increased penetration of wind generation.[5]

However issues can arise in balancing supply and demand over a period of time shorter than the length of the basic dispatch cycle. In any electricity industry the balance between supply and demand must be maintained not just every five minutes but every few milliseconds. Therefore in the NEM, as in other liberalised electricity markets, the system operator must procure what are

---

[4] Such as binding wholesale price caps or floors.

[5] There are some complications in practice in the NEM since, although prices and dispatch are determined every five minutes, generators and loads are paid (or pay) prices which are set every half hour and are averaged of six consecutive five-minute periods.

sometimes called balancing services which will be called upon as needed during each five-minute dispatch interval to balance supply and demand.

An imbalance between supply and demand in the electricity industry results in a variation in the frequency around the target (which in Australia is 50 Hz). A shortfall in supply relative to demand (due to a reduction in generation or an increase in demand) results in a reduction in the system frequency. Conversely an increase in supply relative to demand results in an increase in the system frequency, until the balancing services come into play to restore the frequency to its target level. Since the balancing services are driven primarily by frequency deviations from the target they are known in the NEM as frequency control ancillary services (FCAS).

In the NEM, these balancing services are procured through a set of FCAS markets which operate in parallel to the primary market for the provision of electrical energy. There are separate FCAS markets for making up for a shortfall of supply relative to demand (these are known as "raise" services, and for reducing an excess of supply relative to demand (known as "lower" services). In addition there are separate FCAS markets for generator and loads prepared to make small on-going adjustments (known as "regulation") and markets for large adjustments in the event of significant supply-demand events in the NEM such as the loss of a large generator or a large load (known as "contingency" services). Some generators or loads can respond quickly to frequency deviations while others take longer to respond. As a consequence there are separate FCAS markets for different balancing services which differ in the time over which generators or loads commit to respond. Specifically, in the NEM there are FCAS raise and lower markets for six seconds, sixty seconds, and five minutes (combined with the 'regulation' raise and lower markets, there are a total of eight FCAS markets in the NEM).[6]

In the NEM these balancing services are procured every five minutes by the system operator (the Australian Energy Market Operator, AEMO). The volume of these services purchased in any given five minute interval depends on the overall system's sensitivity to supply-demand changes, the system operator's tolerance for frequency deviation and the size of the forecast maximum likely deviation of supply from demand during each five minute interval. The size of the maximum likely deviation of supply from demand depends on factors such as the size of the largest generating unit (since there is a small but non-zero probability that any given generating unit will be forced to disconnect from the network at any point in time), the accuracy of demand forecasts and, importantly, on the accuracy of wind generation forecasts.

Of course, the output of wind generators is not perfectly forecastable. Even with significant investment in improving forecasting technology there remains a risk that wind generation output will vary or "swing" by a material amount during any given five minute interval. The greater the installed base of wind generation in the NEM, the greater the possible variation of supply from demand.

As noted above, there are two different types of balancing services – the so-called 'regulation' service, which is designed to address the on-going moment-by-moment variation in the supply-demand balance during the dispatch interval, and the 'contingency' services which are designed to address a larger one-off event in the supply-demand balance.

The volume of 'regulation' balancing services required depends on the normal (i.e., highly probable) range of variation in the supply-demand balance during the dispatch interval. The greater the variability in the supply-demand balance during the dispatch interval, the greater the need for regulation services. An increase in the installed wind generation capacity leads to an increase in the unpredictability of the supply-demand balance, and therefore an increase in the requirement for 'regulation' balancing services equal to roughly four per cent of that capacity.

[6] See AEMO (2009).

This is a relatively small figure, but could still be material as the volume of installed wind generation becomes substantial.[7]

The volume of 'contingency' balancing services required to be purchased in a five-minute interval depends on the size of the largest credible contingency that could happen to the system in that five minute interval. The largest credible contingency is usually the loss of the output of the largest single generating unit. In the NEM the output of the largest generating unit is typically in the range 660-760 MW. As we will see below, there is a non-negligible probability that the output of wind generators will vary by more than fifteen per cent of the installed capacity in a five minute interval. At present, the largest credible deviation of wind output from target is smaller in size than the loss of the largest conventional generating unit. Therefore, at present, the requirement for contingency balancing services is set on the basis of the loss of the largest conventional generating unit. However, in principle, as the volume of wind capacity installed increases a point may be reached in the future where, at least on days of strong winds, the variation in wind generation output will be the largest credible contingency on the NEM, requiring an increase in contingency balancing services in addition to the increase in balancing services.

But just how much variability is there in the output of wind generation? This is explored in the next section before addressing the question as to how that variability in output affects the capability of the transmission network.

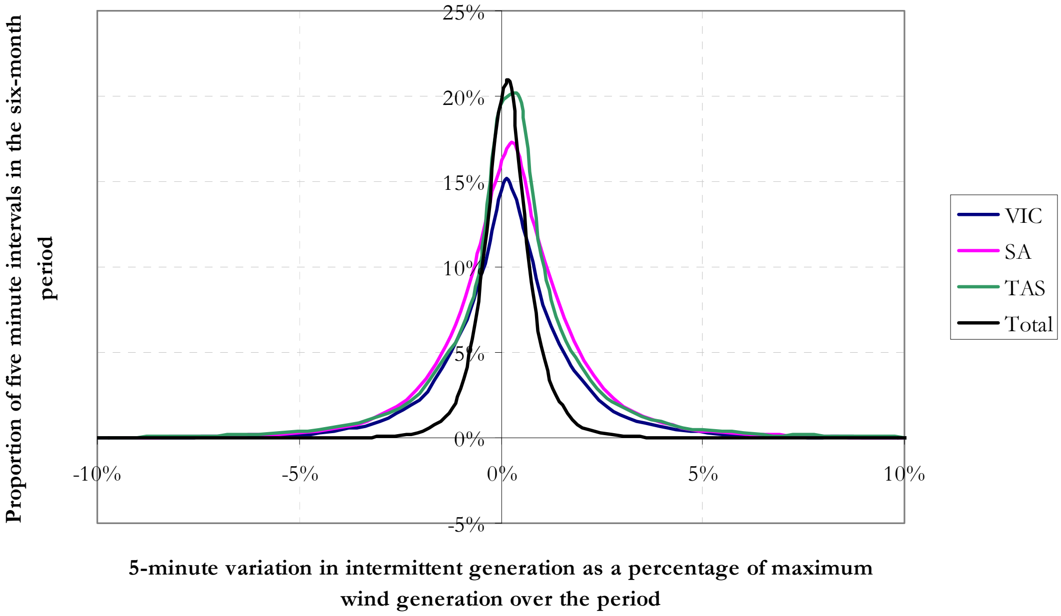
3. The variability in wind output in the NEM

Figure 2 below shows how the output of wind generation has changed from one five-minute interval to the next, expressed as a proportion of installed wind generation capacity, for the six-month time period from 1 November 2009 to 30 April 2010. Not all of the five-minute variation in output of wind generation is unpredictable. However, as a first approximation we will assume that the forecast wind generation output is just equal to the output in the previous five minutes, so that all of the 5-minute variation in output is deviation from forecast.

As can be seen, in those regions which have a larger volume of install wind generation capacity the variability in wind output is, with a high probability, less than two per cent of installed capacity. This is probably due to the greater geographic diversification of wind generation capacity in these regions. In the Tasmanian region of the NEM, the five-minute variation in wind output exceeded 5 per cent of the installed capacity in 3.5 per cent of the five-minute intervals in this six-month period.

[7] Other markets have also raised concerns about the impact of increased wind penetration on the need for balancing services. For example, in the case of the UK, see National Grid (2009, 2010).

Figure 2: Variability in intermittent generation November 2009 - April 2010



Based on a study of five South Australian (SA) wind farms in the six month period January-August 2006, Thorncraft et al (2007) report that if we seek to purchase sufficient regulation service to meet the likely variation in wind output in 99.9999 per cent of dispatch intervals (that is, a level which is likely to be adequate in all except one dispatch interval in ten years), we would need to purchase regulation lower service of about 25 per cent of the installed volume of wind capacity, and regulation raise service of about 22 per cent of the installed wind generation capacity.

In other words, for each 1000 MW of installed wind generation capacity in a region, approximately 250 MW of regulation service will be required each five minutes, in addition to the regulation service purchased to account for variations in demand, losses and so on. This compares with the current level of regulation service purchased which is normally in the range 120-130 MW (rising to 250 MW at times). The analysis above suggests that this requirement will need to be even higher in Tasmania where variation of more than 25 per cent of the installed volume is not all that rare.[8]

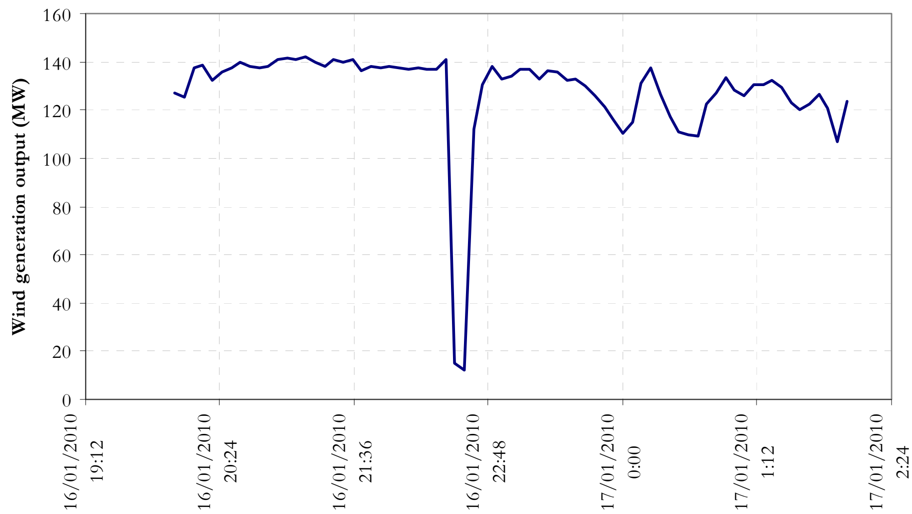
In addition to the high-probability variation in the output from the forecast, the system operator must purchase sufficient balancing services to handle the larger, low-probability variations in the supply-demand balance. The events which cause large variations in the supply-demand balance are known as 'credible contingencies'. As we have seen above, the largest credible contingency in the NEM at present is the loss of the largest single generating unit (typically in the range 660-700 MW). The question for us is whether or not, and under what circumstances, the variation in wind output might have an impact on the NEM to an extent comparable to the loss of the largest single generating unit?

---

[8] This link between increased wind penetration and the increased requirement for balancing services has been recognised for some time. Outhred (2006) cites a 2003 publication of the system operator in the NEM which notes that "as the amount of intermittent generation in the power system increases, there is likely to be an increase in the usage and cost of these ancillary services".

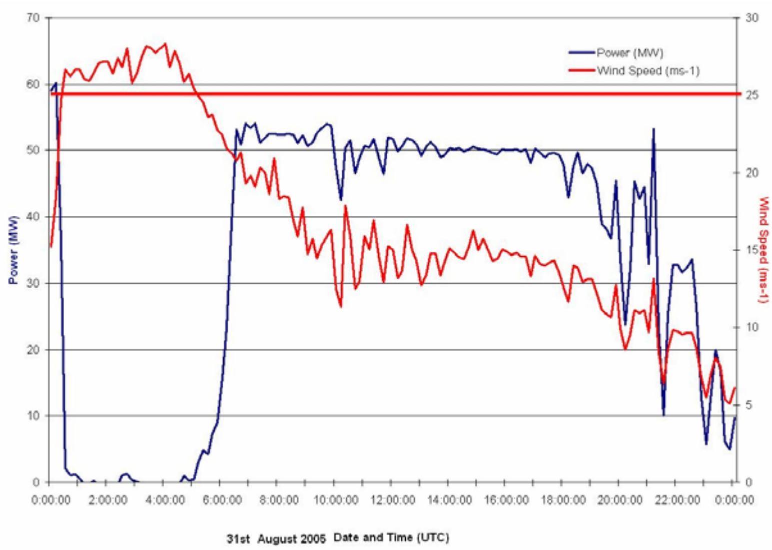
Examination of the wind generation output from the six-month period mentioned above reveals that there are several occasions when wind generation output has varied quite significantly from one five-minute interval to the next. For example, in Tasmania there have been 8 occasions in the six-month period when the wind output varied more than 100 MW in a five minute period (equal to just under half of the maximum wind output during this six-month period). One of these intervals is shown in figure 3 below. On this occasion, the wind output, which was in the vicinity of 140 MW, dropped rapidly to under 20 MW, before being restored a few dispatch intervals later.

Figure 3: Wind generation output, Tasmania, 16 January 2010



One possible explanation for this sudden variation in wind output has been suggested by Thorncraft et al (2007) – that on days of very high winds wind farms may choose to cut output rather than risk damage to turbines due to over-speeding. Thorncraft et al (2007) provide an example of a Tasmanian wind-farm on a day of very high winds. This wind farm shut down its turbines suddenly when the wind speeds rose above 25 m/s, restoring them when the wind speed subsequently dropped below 25 m/s. The result was a rapid fall in the output of the wind farm.

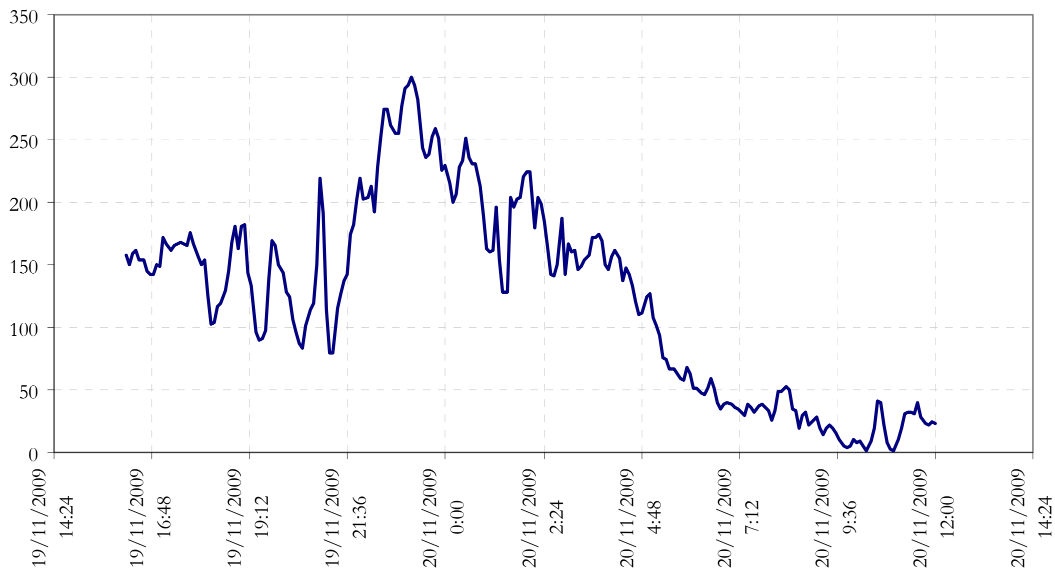
Figure 4: Variation in output of wind farm due to high winds 31 August 2005



Source: Thorncraft et al (2007)

Relatively large swings in the output of wind farms have also been observed in other NEM regions. Figure 5 shows the output of wind generation in South Australia on 19th November 2010. On this occasion, the wind output peaked at 220 MW in the dispatch interval ending 20:55, but reduced to 80 MW fifteen minutes later (a variation of 140 MW), including a drop of 80 MW in one five-minute interval.

Figure 5: Wind generation output, SA region of the NEM, 19th November 2009



Even relatively large swings in the supply-demand balance can be accommodated by the overall electricity supply system provided that there is sufficient regulation balancing services available at all times. In order for these swings to have an impact on the need for contingency balancing services, these variations would have to occur over a time frame too short for the regulation services to respond – that is on a time scale of less than, say, ten seconds.

Thorncraft et al (2007) report that the output of a wind generator can swing by 10 per cent over a ten second period with a probability of 0.0001 per cent. A swing of this magnitude in such a short time frame might require contingency FCAS services – that is, might reasonably be classified as a "credible contingency". However, ten per cent of the installed wind generation capacity is not likely to exceed the output of the largest existing generating unit in the NEM for at least the next five years. Therefore there is not likely to be a need for additional contingency FCAS services in the medium term.[9]

As an aside, it is worth noting that the need for balancing services may depend on the output of specific generators. For example, where the largest credible contingency is the loss of the largest generating unit, the volume of balancing services required will increase as the output of the largest generating unit increases. Similarly, where the largest credible contingency is a rapid drop in wind output, the volume of balancing services required will depend on the output of wind generators at that point in time. It is interesting to note that circumstances may arise where it

[9] We can also observe the impact of generation uncertainty on transmission flows directly by observing the variation between target and measured output over each interconnector at the end of each five-minute interval. The difference between these two measures is an indicator of the magnitude of the uncertainty in the market – due to errors in forecasting demand, contingencies on the network, and the uncertainty in wind generation. Measurement of the deviation between target and out-turn flows on the VIC-SA interconnector shows that the forecast error has been increasing over time, probably due to increased wind penetration.

makes sense to reduce the output of the generator(s) that are driving the need for the balancing services in order to reduce the cost of purchasing balancing services. Put another way, it may make sense to "co-optimise" the output of the generator(s) giving rise to the need for the balancing services with the cost of purchasing balancing services. Specifically, it may make sense to limit the output of wind farms precisely when the cost of procuring additional ancillary services is high. The next section provides a worked example where this trade-off is efficient. This co-optimisation currently occurs in the NEM.

4. Regulation requirements and transmission operating limits

Increased penetration of intermittent generation increases the demand on balancing services. But how does that increased demand for balancing services affect the operating capability of the transmission network?

We have seen that the system operator must procure sufficient balancing services so that any variation between out-turn supply and out-turn demand can be offset during the five-minute dispatch interval. As long as constraints on the transmission network can be ignored, these balancing services can be purchased anywhere on the network. In principle, a generator in South Australia could stand ready to increase its output in response to a shortfall in output from a generator in Queensland – many thousands of kilometers away.

But the presence of transmission constraints complicates this problem. Let's suppose that there is a sudden shortfall in supply in region A. If we are not to shed load in region A[10], we must increase the supply into region A – either by increasing the output of generators in region A or by increasing the output of generators in other regions. But increasing the output of generators in other regions increases the flow on the transmission lines into region A. If the transmission lines into region A are already operating at their physical limit, no more power flow into region A can be sustained. Instead, all the shortfall in supply in region A must be made up by an increase in the output of generators in region A. When the transmission lines into region A are operating at their physical limits, generators located outside region A cannot provide "raise" balancing services within region A at all.

Although there are mechanisms in the NEM which allow the system operator to purchase FCAS locally under certain circumstances, this is not how the NEM usually operates.[11] Instead, FCAS services are purchased on a NEM-wide basis. That is, if there is a credible risk of the loss of a generator in region A, the balancing services needed to cover that risk may be purchased from anywhere else in the NEM. This policy is only sustainable, of course, if the transmission lines are never operated up to their physical limits. Instead, sufficient "headroom" must be maintained on every interconnector such that, in the event of any credible contingency, sufficient additional power can be delivered over that transmission line so as to restore the balance between supply and demand until at least the next dispatch interval.

Suppose, say, that there is a credible risk of a shortfall of supply of, say, 200 MW in region A (perhaps due to the variability in the output of wind). Suppose therefore, that the system operator purchases 200 MW of raise service. Under the current NEM rules this 200 MW could be supplied by any generator located anywhere in the NEM. As a result the transmission lines in the NEM cannot be operated up to their physical limit. Put another way, an "operating margin" or "headroom" of 200 MW on the interconnector is required, reducing the operating capability

---

[10] Load shedding as a means to balance supply and demand is typically only used as a last resort.

[11] The only circumstances in the NEM under which FCAS services are procured locally within each region are when the loss of the interconnector is considered a credible contingency. In this case, if the interconnector were to fail entirely all the resulting imbalance of power has to be made up by FCAS services procured locally (since, in the absence of the interconnector FCAS services can obviously not be delivered to where they are needed).

of the transmission network. The larger the regulation raise volume required, the larger the operating margin and therefore the lower the operating capability of the transmission network.

This problem of higher-FCAS-requirements leading to lower-transmission-capability is not just a consequence of increasing wind penetration. Any other change in the market which increases the FCAS requirements will have the same effect. Prior to 2007 the largest single generating units in the NEM were 660 MW units in NSW. In 2007 a 760 MW unit was installed at Kogan Creek in Queensland. The establishment of this unit increased the size of the largest credible contingency and therefore increased the size of the FCAS requirements, resulting in a reduction in the transmission capability, including a reduction in the flow limits on the NSW-Queensland interconnector at times.

This reduction in the operating capability of the transmission network has real economic costs. In increases the incidence of transmission congestion, forcing out-of-merit dispatch and thereby increasing the cost of generating sufficient power to meet demand. In addition, increased transmission congestion also enhances the scope for market power which further increases the dispatch costs, distorts price signals, and may reduce the liquidity of financial markets. Finally, enhanced congestion increases pressure to expand the transmission network which, although not as costly as installing a new generating station, still typically involves hundreds or sometimes billions of dollars.

This problem has, to an extent, been recognised by the rule-making authority in the NEM (the Australian Energy Markets Commission, AEMC). In the "Central Dispatch and Integration of Wind and Other Intermittent Generation" final rule determination the AEMC made the following comments:

> "**Reduced market efficiency due to higher operating margins**
>
> The network constraint equations used in NEMDE to control network flows to be within secure transfer limits include a safety margin to account for measurement errors and other uncertainties due to inaccuracies of forecast demands on network flows. These safety margins need to be sufficiently large to allow for the errors and uncertainties but the presence of a safety margin does reduce the network transfers associated when the associated constraint equation is binding.
>
> A reduction in the transfer capability through an increase in the operating margin means that at the times when the constraint is binding a higher cost generator must operate at an otherwise increased output, with an equivalent reduction in the output of a low cost generator, thus increasing the cost of dispatching generation to meet the load.
>
> The presence of large non-scheduled intermittent generation is likely to increase the uncertainty in the network flows, thus increasing the operating margins, reducing the transfer allowable capability and increasing costs of dispatching sufficient generation to meet demand. Where the affected network is an interconnector the reduction in network transfer capability may reduce the firmness of the hedges funded by the associated inter-regional settlements residues."[12]

Partly in response to concerns over the impact of intermittent generators, the AEMC introduced a new class of generation participant in the NEM known as a "semi-scheduled generator". A semi-scheduled generator is required to submit an offer curve just like any other generator, and to follow the dispatch targets set by AEMO (to the extent they are able). However, the total available output of these semi-scheduled generators is determined by a separate wind forecasting

---

[12] AEMC (2008), page 12.

system. In practice this means that semi-scheduled generators are not allowed to produce more than the target set by AEMO, but may produce less.

Preventing wind farms from producing above their target reduces the likelihood that wind farms will produce more than their target but does not reduce the likelihood that wind farms will produce less than their target. This reduces the uncertainty in transmission flows somewhat, but does not eliminate it entirely. System security problems still remain and must still be taken into account through higher operating margins.

5. The solution: nodal purchasing and co-optimisation

The previous sections have established that, when FCAS requirements are purchased "globally" (that is, on a NEM-wide basis) there arises a link between the volume of FCAS required and the required operating margins on the transmission network. But this raises a question: How can we ensure that the overall total cost of transmission congestion plus the cost of procuring balancing services is minimised?

There are three possible approaches:

(a) The first possibility, like the status quo, purchases all balancing services globally, on a NEM-wide basis. This approach reduces the operating capability by whatever amount is necessary to allow the balancing services to be delivered to where they are needed. This approach minimises the cost of procuring the balancing services, but may result in very high congestion costs arising from the low level of inter-regional transfer capability.

(b) The second possible approach, which is the opposite extreme, is to set the operating limit of the transmission lines equal to their physical maximum and to purchase all balancing services locally. This approach reduces the congestion costs to the minimum but may result in very high costs of balancing services – particularly in regions with a shortage of generators able to provide regulation services.

(c) The third possible approach is to co-optimise the operating limit of the transmission lines with the costs of congestion and the costs of procuring balancing services. For example, it may be that some proportion of the required balancing services can be cheaply procured locally, reducing the size of the operating margin, and therefore reducing the size of the congestion costs.

This paper argues that the last approach is the most efficient. As the penetration of wind generation increases, increasing the need for balancing services (especially regulation service), the operating limits on the transmission network, and the generators selected to provide balancing services, should be chosen as to minimise the sum of the cost of transmission congestion and the cost of procuring balancing services.

An example

Some further understanding of this problem and the proposed solution can be gained with the aid of a simple network model with two-nodes and a single transmission link.

In this model the transmission link has a physical limit of 1000 MW. There are three reliable generators. The first reliable generator (G1) is located at node 1 and has a variable cost of $10/MWh. The second reliable generator (G2) is located at node 2, has a capacity of 400 MW and a variable cost of $100/MWh. The third reliable generator (G3) is located at node 2, has a capacity of 800 MW, and has a variable cost of $120/MWh. There is also 800 MW of intermittent generation at node 2 with zero variable cost. This intermittent generation is assumed (for simplicity) to be either "full on" or "full off" (i.e., producing at full capacity, or unable to produce at all).[13] There is a fixed, constant, load of 1800 MW at node 2 (so power always flows from node 1 to node 2). The probability of this full in intermittent generation is assumed to be 1-in-100 (1 per cent). This network is illustrated in Figure 6 below.

Figure 6: Simple two-node illustrative network

Since the risk here is a sudden loss of the output of the intermittent generator, there is a need for a 'raise' service. The amount of regulation raise service required at any given moment in time is equal to the output of the intermittent generation at that time. All three reliable generators are capable of delivering a raise regulation ancillary service. G1 is assumed to be able to deliver a given MW quantity of regulation raise service at a cost of $5/MWh. G2 can deliver regulation raise service at a cost of $20/MWh and G3 can deliver regulation raise service at a cost of $200/MWh.

The key parameters of the generators are summarised in the following table:

Table 1: Summary of generators in the simple network

Generator	Type	Variable cost	Capacity	Cost of providing regulation raise service
G1	Reliable	$10/MWh	Unlimited	$5/MWh
G2	Reliable	$100/MWh	400 MW	$20/MWh
G3	Reliable	$120/MWh	800 MW	$200/MWh
G4	Intermittent	$0/MWh	800 MW	Does not provide this service.[14]

Status Quo

Let's first consider the case where there is a uniform price for ancillary services, and the transmission line capacity is "backed off" whenever necessary to allow those ancillary services to be provided where they are needed. Since the generator at node 1 is the cheapest supplier of regulation, it will always be selected as the provider of regulation service, and the transmission line operating limit will be set at a low enough level to accommodate that regulation supply. Since,

[13] As noted earlier, swings of 25 per cent of wind farm capacity happen with a probability of once in ten years, so this scenario might correspond to a wind farm capacity four times the size of the variation considered here.

[14] In fact, wind generators could in theory provide some regulation or FCAS services.

in this example, it turns out that the intermittent generator is producing at its maximum (which is 800 MW), 800 MW of regulation service is required, so the transmission line must maintain 800 MW of headroom, reducing the output of the generator at node 1 to 200 MW. The remaining output is made up by the reliable generators at node 2. Because G1 has been backed off, the total cost of generating to meet demand is quite high at $90,000/h. But the cost of providing the regulation service is relatively low at $4,000/h. The overall expected total cost is $94,080/h.

Local procurement of regulation service

Now consider the case where regulation service is sourced locally, so that there is no need to "derate" the interconnector to match the increase in intermittent generation. Since G1 is the cheapest generator, it is always run at is maximum of 1000 MW. As long as the wind output is less than 400 MW, the overall efficient dispatch is to run G3 at 400 MW and to make up the difference with G2. This leaves G2 in a position where it can provide raise regulation service in the event that the wind output at G4 fails. The higher the wind output, the more dispatch at G2 can be displaced (which saves $100/MW as long as the wind is blowing) but increases the cost of procuring regulation (which increases at the rate of $20 per MW of wind output). So, overall, as long as the wind output at G4 is less than 400 MW, it is efficient to allow the wind generators to produce more.

However, when the wind output increases past 400 MW, G2 can no longer provide all the required regulation service. Instead, the regulation service must be also provided by G3 – so G3 must be backed off. Now, for each additional MW of wind generation at G4, G3 must be backed off by 1 MW. This saves $120/MW in generation costs but adds $150/MW in regulation costs. Now the additional costs of regulation exceed the savings in generation, so it is not efficient to allow the wind generators to produce more than 400 MW.

In other words, when the regulation service must be entirely procured locally, this is likely to increase the cost of procuring regulation services. In this example, this increased cost of procuring regulation service outweighs the cost saving from using wind energy – as a consequence it is not efficient to allow the wind generators to produce as much as they are able. It is more efficient for the market as a whole to restrict the output of wind generation in order to limit the requirement for raise services.

Overall the efficient dispatch is for G1 to produce 1000 MW, G2 to produce nothing in the energy market, but to offer 400 MW of raise service, G3 produces 400 MW of energy and provides no raise service and the intermittent generation G4 is limited to producing 400 MW. Overall the cost of generation is much lower (due to the increased reliance on cheap generation at G1), and the cost of regulation is higher than in the status quo case above.

Co-optimisation of regulation procurement and transmission limits

In fact, as this paper argues, we can do better than either fully local or fully global procurement of balancing services. By allowing the headroom on the interconnector to be co-optimised with the procurement of regulation service (and the volume of regulation service required) we find that it is efficient to allow G1 to produce 600 MW (retaining room for it to provide 400 MW of raise service). The resulting saving on regulation service is larger than the additional generation cost arising from not being able to produce using cheap generation at G1. The overall optimum is to produce the energy requirement using G1 and G3, and to produce the regulation requirement using G1 and G2. The overall social cost is significantly lower than under either of the above approaches alone.

Table 2: Optimal dispatch of energy and regulation services under different approaches

		Global procurement	Local procurement	Co-optimisation
Generation output (MW)	G1	200	1000	600

	G2	400	0	0
	G3	400	400	400
	G4	800	400	800
Regulation provided (MW)	G1	800	0	400
	G2	0	400	400
	G3	0	0	0
Cost of generation ($/h) (no contingency)		$90,000	$10,000	$54,000
Cost of generation ($/h) (with contingency)		$98,000	$98,000	$98,000
Cost of regulation ($/h)		$4,000	$8,000	$10,000
Total expected cost ($/h)		$94,080	$66,400	$64,440

The above example uses the simplest possible two-node, one-link network. However, the principles extend to networks with multiple nodes and loop flows. Importantly, however, in the case of a meshed network, in the event of a need to procure raise services – the most efficient way to provide those services may involve some generators being dispatched to provide raise services and other generators being dispatched to provide *lower* services. Intuitively, the reason is that combining raise and lower services can allow the raise requirement to be met while minimising the impact on transmission flows – thereby allowing transmission lines to be operated with lower headroom.

The appendix to this paper sets out a formal mathematical model of the simultaneous procurement of energy and FCAS services in a simple network. The appendix shows that local procurement of FCAS services, combined with co-optimisation of the procurement of FCAS services with transmission limits, results in the most efficient overall outcome.

The model in the appendix focuses on the short-term dispatch problem, taking existing generator locations and capacities as fixed. In the longer term local procurement of FCAS services would have the further benefit of incentivising efficient generator location decisions – that is, rewarding generators for locating in regions where their ability to provide FCAS services enhances transmission capability.[15]

Of course, even in the efficient outcome, overall FCAS requirements may be higher and transmission limits may be lower (and the congestion costs higher) than in the case where only reliable generation capacity is present in the market. The presence of intermittent generation increases the overall costs on the market. These costs should be passed back to the wind generators themselves.

In principle these costs should be apportioned according to the extent to which wind generators deviate from their targets, so as to reward those wind generators who invest in equipment or processes to stabilise their output around the target.[16] It is conceivable that some wind generators

---

[15] At present, transmission network operators in the NEM are able to directly enter into arrangements with individual generators under which the generators are rewarded for providing FCAS services which enhance network capability. However the scope for such arrangements depends on the willingness of the transmission operator to detect the need and to enter into contractual arrangements. Under the proposal in this paper, the market arrangements would automatically provide location signals of this kind.

[16] Under the present arrangements in the NEM wind generators do face some of the costs of their unreliability – by being required to pay a share of the regulation costs that is related to the deviation in their

might even be able to take advantage of particular local conditions to give rise to a wind output which is "counter-cyclical" – that is, which deviates from target in a manner which is opposite to the majority of the wind farms. Generators should be rewarded for being able to follow a target (and should be even more rewarded for providing counter-cyclical output). In principle, of course, the same rewards and penalties should be applied to other renewable and non-renewable generators. The precise methodology for allocating the costs incurred as a result of a higher regulation FCAS requirement are beyond the scope of this paper.

In the absence of a mechanism for imposing the costs of intermittency back on intermittent generation, intermittent generation would be inefficiently favoured over other forms of renewable generation.

Conclusion

This paper argues that, as the penetration of intermittent generation increases, the current arrangements in the Australian NEM will not result in an overall efficient mix of renewable generation and efficient use of and investment in the transmission network, for the following reasons:

(a) The current policy of procuring balancing services on a NEM-wide basis combined with the policy of maintaining sufficient headroom on transmission links, results in lower transmission network limits than is strictly necessary and lower than is efficient – that is, in principle it should be possible to raise the operating limits on the transmission network, with consequent benefits for the entire market, without any additional capital expenditure, merely by introducing local procurement of FCAS services combined with co-optimisation of the FCAS markets with transmission limits.

(b) Wind and other intermittent generation does not currently face the full impact of their entry decision on the operation of the NEM. Specifically, under the current market arrangements increased wind penetration will ultimately give rise to lower transmission limits. The cost of lower transmission limits (in the form of increased congestion and market power) is not passed back to wind generators.

These concerns may not yet be material in practice. In recent years AEMO has increased the headroom on the VIC-SA interconnector by 35 MW. This amount is still relatively small relative to the volume of installed wind generating capacity. If this adjustment is typical, the volume of installed wind capacity in SA would have to double or triple before there was a material impact on transmission flow limits.

Nevertheless in the next few years a very substantial volume of wind generation is forecast to enter the Australian NEM. In my view, consideration should be given to moving to localised purchase and pricing of regulation services combined with co-optimisation with transmission network limits. Further work could determine precisely how the costs of these regulation services and/or increased transmission congestion would be passed back to the causers, and the implications of this approach for the firmness of inter-regional settlement residues. Further consideration of these issues is likely to prove necessary to achieve a smooth and efficient transition to a low-carbon electricity generation sector.

---

output from the target. However, those regulation costs are sourced on a NEM-wide basis and therefore are likely to be, on average, lower than is efficient.

Appendix

Let's suppose we have a electricity network with *n* generators which (without loss of generality) are located at up to *n* different nodes. There is assumed to be just one intermittent generator (labelled generator 0) which, for simplicity is assumed to be located at the reference node (or swing bus). Each generator can offer to produce either electrical energy or raise service. When the generator *i* produces electrical output at the rate g_i it incurs costs at the rate described by the cost function $c_i(g_i)$. Similarly, when it offers to provide the amount of raise service Δg_i, it incurs the cost $b_i(\Delta g_i)$. Each generator has a maximum level of output that it can produce denoted $\overline{g}_i$. With some probability *p*, the capacity of the intermittent generator may drop to the level $\overline{g}_0^L \leq \overline{g}_0$. There is a load at the swing bus equal to *d*.

The social objective function is to find the choice of energy targets for each generator $\{g_i\}$ and the level of raise service provided by each generator $\{\Delta g_i\}$ which minimises the cost of generation plus the cost of regulation:

$$SWF = (1-p)\sum_{i=0}^{n} c_i(g_i) + p\sum_{i=0}^{n} c_i(g_i + \Delta g_i) + \sum_{i=1}^{n} b_i(\Delta g_i) \qquad \ldots(1)$$

Subject to the following constraints: First, the energy balance constraint requires that the total output matches the total demand both with and without the contingency. Without the contingency the energy balance constraint is simply the requirement that total supply must equal total demand:

$$\sum_{i=0}^{n} g_i = d \qquad \ldots(2)$$

With the contingency, the energy balance constraint states that the total supply of each generator plus any raise service offered must be equal to total demand:

$$\sum_{i=0}^{n} (g_i + \Delta g_i) = d \qquad \ldots(3)$$

Given equation (2) above, we can write equation (3) in the following form (which we might label the FCAS balance equation). (Here Δg_0 is a negative amount representing the size of the fall in output of the intermittent generator when the contingency occurs).

$$\sum_{i=0}^{n} \Delta g_i = 0 \qquad \ldots(4)$$

In addition for each reliable generator, the total output (including the raise service) must not exceed its total capacity $g_i + \Delta g_i \leq \overline{g}_i$ and the total amount of energy and raise service offered must be positive $g_i \geq 0$, $\Delta g_i \geq 0$. In the case of the intermittent generator the corresponding constraints are that the total output without the contingency must not exceed the total capacity $g_0 \leq \overline{g}_0$ and the total amount of energy produced in the contingency must not exceed the reduced capacity $g_0 + \Delta g_0 \leq \overline{g}_0^L$.

Finally, we have the transmission constraints. Let's suppose that we have a DC-linear load flow approximation to the AC network. Taking node *n* as the swing bus, the flow on line *l* can then be represented as a linear combination of the generator injections as follows: $\sum_{i=1}^{n-1} h_{li} g_i$ where h_{li} is the matrix of power transfer distribution factors. If we let the limit of the flow on line *l* be F_l, we have two additional sets of constraints: The flow on any line without the contingency must not exceed the limits:

$$\sum_{i=1}^{n-1} h_{li} g_i \leq F_l \qquad \ldots(5)$$

Similarly, the flow on any line post-contingency must not exceed the limits:

$$\sum_{i=1}^{n-1} h_{li}(g_i + \Delta g_i) \leq F_l \qquad \ldots(6)$$

The objective of this appendix is to show the main results claimed in the text – in particular that local (nodal) purchasing of FCAS services, combined with co-optimisation of the transmission limits will result in more efficient outcomes than either of the alternative approaches: global purchasing of FCAS services (combined with increased headroom on the transmission network) or local purchasing of FCAS services (and reduced headroom on the transmission network).

First, let's show that, even in the absence of transmission constraints, it may make sense to reduce the output of the intermittent generation in order to save on FCAS costs (even where there is a resulting increase in energy costs). From the first order conditions for the optimisation problem above we see that it is efficient for each generator to be dispatched for energy up to the point where its marginal cost is the same and equal to the wholesale market price, unless the generator is also dispatched to provide raise services – in other words, unless the generator's capacity constraint is binding. In this case the generator can be dispatched to a level of output corresponding to a lower marginal cost. But, for the intermittent generator, the capacity constraint is, in effect, always binding – since it is the fall in capacity of this generator which drives the need for raise services in the first place. If the intermittent generator's output did not change there would be no need for regulation services. In other words, the generator which is "causing" the need for FCAS services will always be dispatched to a point on its marginal cost curve which is lower than the wholesale market price. In some circumstances, this will mean that the intermittent generator will have its output reduced relative to the level of output it would normally be dispatched at the wholesale market price.

How would the above optimisation problem change if (as in the present arrangements in the NEM) all FCAS services were required to be purchased at a uniform price from any location in the NEM? This could only arise if we reduce the flow limit on any line from F_l to $F_l + \max_i(h_{li})\Delta g_0$ (recall that $\Delta g_0 < 0$) (in other words, increase the headroom to $-\max_i(h_{li})\Delta g_0$). Now, we can observe that:

$$\sum_{i=1}^{n-1} h_{li}\Delta g_i \leq \max_i(h_{li}) \sum_{i=1}^{n-1} \Delta g_i = -\max_i(h_{li})\Delta g_0 \qquad \ldots(7)$$

If we increase the headroom in this way, equation (7) shows that equations (5) and (6) are automatically satisfied for any choice of the FCAS dispatch. In other words we have shown that (a) we can achieve global purchasing of FCAS services provided we allow headroom on the *l*th interconnector equal to $-\max_i(h_{li})\Delta g_0$; and (b) the resulting transmission constraints are more restrictive than the optimisation problem above and therefore results in a higher social cost and a less efficient overall dispatch.

Finally, how would the above optimisation problem change if we required all FCAS requirements to be purchased locally? Since the FCAS requirement occurs at the swing bus in this problem, under this restriction only generators located at the swing bus would be able to provide FCAS services. This amounts to introducing additional constraints on the problem above, namely that $\Delta g_i = 0$ for all generators not located at the swing bus.

Since local purchasing of FCAS further restricts the set of constraints set out above it also results in a higher social cost and a less efficient overall dispatch. We have therefore proved the result that local (nodal) purchasing of FCAS services, combined with co-optimisation of transmission limits, results in the most efficient overall dispatch.

Notice that the analysis above focuses only on the procurement of regulation raise service. It is easy to demonstrate that, in some circumstances, the most efficient dispatch of regulation services may involve *both* the purchase of raise and lower services even if the FCAS requirement arises from the potential for a sudden *shortfall* in supply. This might happen for example, in a network where the intermittent generation is located on a transmission loop with at least one network link operating at its physical limit. In this case the most efficient response to a sudden shortfall in supply at one node may be to simultaneously ramp up generation on one side of the transmission constraint and ramp down generation on the other side of the transmission constraint.

References

Australian Energy Market Commission (AEMC) (2008), "Rule Determination: National Electricity Amendment (Central Dispatch and Integration of Wind and Other Intermittent Generation) Rule 2008", 1 May 2008, available at: http://www.aemc.gov.au/Media/docs/Final%20Rule%20Determination-340d1503-3f23-407d-83f7-4cacbab44eca-0.pdf

Australian Energy Market Operator (AEMO) (2009), "Operating Procedure: Frequency Control Ancillary Services", Document No. SO_OP3708A, 1 July 2009 available at http://www.aemo.com.au/electricityops/so_op3708av012.pdf

Carbon Market Economics (2009), "Renewable Generation Projections: 2009-2028: Revised Final Report", Carbon Market Economics Pty Ltd, January 2009, available at: http://www.aemo.com.au/planning/419-0032.pdf

National Grid, (2009), "Operating the Electricity Transmission Networks in 2020: Initial Consultation", June 2009

National Grid, (2010), "Operating the Electricity Transmission Networks in 2020: Follow up report", February 2010

Outhred, H. (2003), "National Wind Power Study: An estimate of readily accepted wind energy in the National Electricity Market", November 2003

Outhred, H. and I.F. MacGill (2006), "Integrating Wind Energy in the Australian National Electricity Market", presented at the World Renewable Energy Conference, August 2006.

Thorncraft, S.R., H.R. Outhred, D.J. Clements, F.K. Barker, (2007), "Operation of Ancillary Service Markets with Large Amounts of Wind Power", 26 April 2007