A An introduction to congestion in the NEM

A.1 Introduction

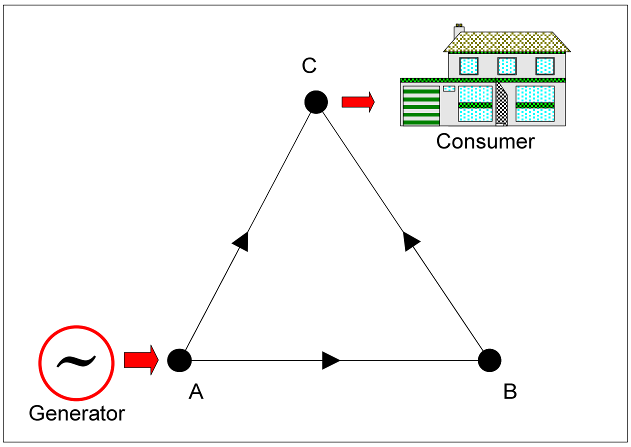
This appendix provides a simple introduction to congestion in the National Electricity Market (NEM) for readers who are new to the NEM and/or new to the topic of congestion.

A.2 What is congestion?

Electricity is transported from suppliers (generators) to consumers (retailers and large customers) along a transmission network. "Congestion" is what happens when there is a bottleneck somewhere on this network. That is, whenever a particular element on the network (e.g. a line or transformer) reaches its limit and cannot carry any more electricity than it is carrying already, it is "congested".

Electricity flows across the whole network, taking whichever paths are available. Using an example of a simple network (Figure A.1 below), power injected at point A (e.g. a power station) flows along *multiple* paths to where it is consumed at point C. This happens because power flows have to obey certain laws of physics.

Figure A.1 Flows across the network

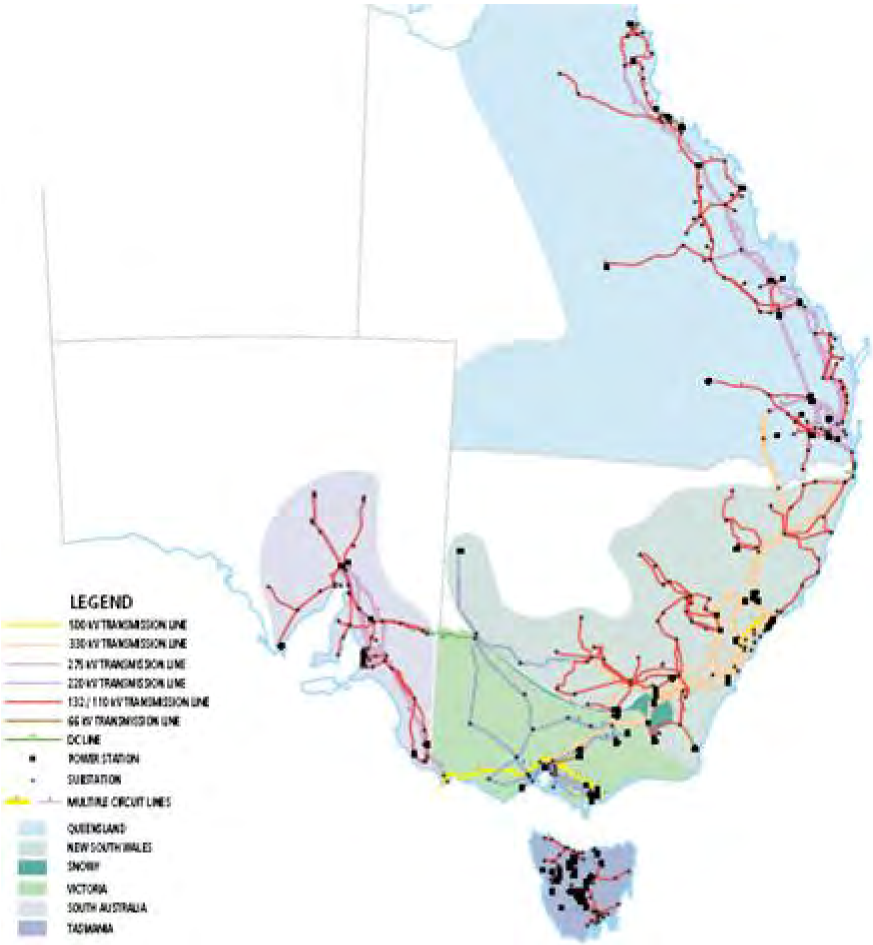


The flow of power across the network means that when a limit is reached on one part of the network, adjustments have to be made in generation and consumption across the network to ensure that the limit is not exceeded.

A congested element indicates that demand for electricity in the vicinity is equal to what the element can carry; taking into account power system security requirements. Demand must still be met, of course, so the additional electricity has to come along an alternative route and from an alternative source of supply—that is, another generator. When this situation occurs, it can affect the price at which electricity is supplied and the price that consumers must pay for it. Furthermore, congestion can make the price vary from one location to the next, whereas when there is no congestion (and assuming there are no electrical losses) the price is the same everywhere on the network.[44]

Congestion has commercial consequences and in particular creates risks for generators and retailers. It also affects the economic efficiency of the NEM as a whole.

Figure A.2 The NEM network



Data source: NEMMCO, “An introduction to Australia’s national electricity market”, June 2005, p. 30.

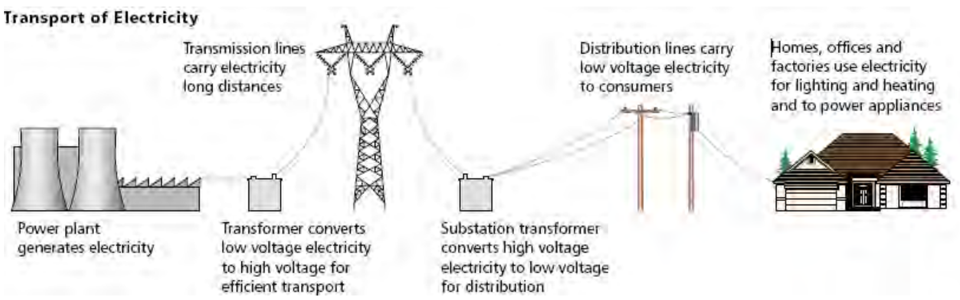
The transmission network extends from the east coast of Australia to mid-west South Australia, and from Far North Queensland all the way to Tasmania. Consumers can be supplied with electricity produced by any generator or combination of generators

44 Locational prices used to determine dispatch volumes take into account both the constraints and transmission losses. Transmission losses will cause a small variation in prices between locations even without any binding constraint on the network.

in the NEM. A consumer in Cairns, for example, can receive electricity produced by a generator in Victoria's La Trobe Valley.

This one interconnected physical network spans several distinct geographical areas. These "regions", as they are known in the NEM, are New South Wales, Queensland, Snowy[45], South Australia, Tasmania, and Victoria—almost, but not quite, identical to the States. This "one market, multiple regions" characteristic has important consequences when it comes to the impact of congestion on the NEM's financial market (discussed in subsection A.4.2).

Figure A.3 The transport of electricity: real-time supply and demand



Data source: NEMMCO, "An introduction to Australia's national electricity market", June 2005, p.3.

The network includes generation plants, transmission lines, transformers and distribution lines – the "supply" side – and consumers in homes, offices and factories – the "demand" side.

Consumer demand for electricity varies all the time; every time a light or air-conditioning unit or some other electrical device is switched on or off, demand changes. Because electricity cannot be stored, concurrently the supply of electricity varies all the time too, to keep pace with demand. So the amount of electricity being produced by generators and flowing along the network is never constant; it is always fluctuating. It is essential that supply matches demand at all times in order to maintain the quality of supply, to keep the lights on, and to maintain the integrity of the power system in the event of a security contingency (e.g. a lightning strike or plant breakdown).

Congestion is a consequence of the fact that electricity cannot be stored and therefore that supply and demand have to be kept in balance in "real-time".

A.3 Why does congestion occur?

Congestion occurs because there are physical limits to the network's ability to carry electricity. There are also security limits, designed to maintain the integrity of the

45 The Snowy region will be abolished from 1 July 2008. See AEMC 2007, *Abolition of Snowy Region*, Final Rule Determination, 30 August 2007, Sydney.

system. If there were no limits, there would be no congestion. The ability of the network to carry electricity is known as its "transfer capability".

Transfer capability

Transfer capability is not a simple concept. It is neither a single "amount", nor is it fixed. Instead it depends on a complex range of factors and varies from one moment to the next, dynamically responding to changing conditions.

Broadly, we can say that at any moment in time transfer capability is governed by:

- security and reliability parameters;
- patterns of generation and demand;
- ambient weather conditions;
- the availability of transmission elements (e.g. transmission lines and transformers being in service);
- the availability of contracted Network Support and Control Services (NSCS) (e.g. reactive power capability, network loading control); and
- the technical design limitations of individual network elements—their "capacity".

Keeping the flow of electricity within a line's designated "capacity" is critical to the physical and operational security of the power system. Violating these limits may cause equipment damage, dangerous situations for the general public, or cascading load shedding that may ultimately lead to partial or full system shutdown. These limits can be broadly described as either "thermal" or "stability" limits:

- *Thermal limits* refer to the heating of transmission lines as more power is sent across them. The additional heat causes the lines to sag closer to the ground. The clearance above ground level must exceed certain minimum heights to ensure both public safety and power system security. Thermal limits also apply to other elements of the network, such as transformers.
- *Stability limits* refer to the need to keep the transmission system operating within design tolerances for voltage, with the ability to recover from disturbances, taking into account interaction control systems and other technical characteristics that are important to keep the power system intact. Stability limits tend to vary with the location and quantity of generation and demand, as well as with some other factors.

Congestion is therefore specific to a pattern of electrical flows, to the capability of the transmission system, and to a point in time. Congestion might emerge at a location in one five-minute dispatch interval, but disappear in the next interval. This might reflect, for example, changes in the patterns of generation or demand, or changes in transmission capabilities (e.g. as a line is brought back into service following maintenance).

In general, an enhanced ability to handle power flows means, other things being equal, a lower likelihood of network congestion occurring and hence reduced physical and financial trading risks for participants. These risks are discussed in detail in Appendix C.

In theory, congestion could be eliminated if enough money were spent on expanding the transmission network's infrastructure. However, the cost of doing this would outweigh the costs incurred from congestion itself. In this sense, congestion occurs not only because of the network's physical limitations, but also because of economic considerations of net costs and benefits. In other words, some level of congestion is in fact economically efficient.

A.4 How does congestion manifest itself?

The NEM has a physical side—the production, supply and consumption of a commodity—and a financial side—the transactions and contracts surrounding the buying and selling of that commodity. Congestion makes itself felt in both. Here we explain first how it affects the dispatch process (the physical side) and then how it affects the NEM's financial market.

A.4.1 Congestion and the dispatch process

As market and system operator, NEMMCO manages the process that determines which generators will be required to generate electricity and how much they will be required to generate in order to meet demand. This is the *dispatch process*. NEMMCO calculates dispatch every five minutes; these five-minute intervals are known as *dispatch intervals*.

In each dispatch interval NEMMCO's job is to achieve a *central dispatch objective* by calculating an optimal solution to a security-constrained dispatch problem, which contains a large number of variables, parameters, limits and constraints.

- The *central dispatch objective* is to meet demand using the "least-cost" combination of generation available.[46]

- The *variables* are the prices and quantities contained in the bids and offers submitted by market participants, as well as predefined parameters for maintaining system security and reliability.

- The *optimal solution* will therefore be to dispatch the "least-cost" combination of generation to meet demand, based on bids and offers, while remaining within the security and reliability parameters.

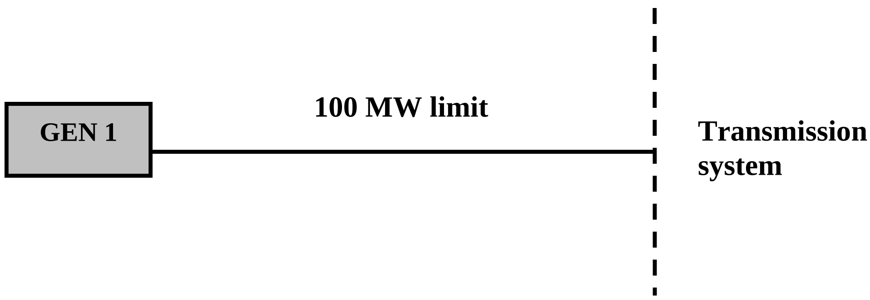
[46] Clause 3.8.1 of the Rules details the responsibilities of NEMMCO regarding the central dispatch process. This Rule states that the central dispatch process should aim to maximise the value of spot market trading on the basis of dispatch bids and offers. In practice this translates into the objective to meet demand using least-cost generation.

Because the calculations are so complex and have to be performed so rapidly, NEMMCO employs a *dispatch engine* (NEMDE) to do them. NEMDE is a computer program that uses industry-standard linear programming tools to optimise dispatch.

A.4.1.1 Constraint equations and limits

The information characterising network capability and security and reliability parameters is contained in a set of "network constraint equations" within NEMDE. Each network constraint equation is a mathematical representation of the way in which different variables affect flows across particular transmission limits. A network constraint is thus a limitation imposed on the market dispatch relating to the physical capability of the transmission network in the relevant five-minute dispatch interval. The limits are derived taking into account power system security requirements. There is a separate constraint equation for each limitation imposed on the dispatch. NEMDE then solves the security-constrained dispatch problem, as described above.

To illustrate, consider the simplified example below where a generator is connected to the main interconnected transmission system by a single circuit that has a limit of 100 MW and there is no load connected between the generator and the main transmission system.



The constraint equation would be "formulated" by NEMMCO to ensure that the 100 MW limit on the line is not breached, i.e. that the output of generator GEN 1 does not exceed 100 MW. This constraint equation would therefore take the form:

GEN 1 < 100

However, NEMMCO would also need to provide for certain contingencies, such as when the transmission circuit linking GEN 1 to the main interconnected system is taken out of service for maintenance. In this contingency, the following constraint equation would be used:

GEN 1 = 0

To illustrate further, the simple example above can be extended to include load. If a load (LOAD 1) were located at the same location as GEN 1, the flow along the line with the limit of 100 MW would be determined by the generation output net of the amount of electricity consumed by LOAD 1. Hence, the constraint equation would take the form: [47]

GEN 1 – LOAD 1 < 100

In practice, the constraint equations need to reflect much more complicated sets of circumstances—for example, combinations of generation, loads and interconnector flows across multiple credible contingencies and allowing for electrical losses. There are also sets of constraint equations to ensure that system frequency is maintained within acceptable tolerances. However, the intuition behind the purpose of a network constraint equation still holds. It is a description—from the perspective of system security—of permissible combinations of variables that might influence electrical flows across a network element at a point in time.

As noted above, this “snapshot” provided by a constraint equation is dependent on the combination of transmission assets that are in service at the relevant time. The set of constraint equations reflecting a network configuration in the absence of any outages is referred to as a set of “system normal” constraints. In other instances, transmission outages might need to be scheduled to facilitate maintenance and other works on the transmission system. When this occurs, different sets of constraints need to be invoked in the dispatch process. In general, a separate constraint equation may be required for each potential contingency that materially impacts the permissible flow of electricity through a network limit, and it may sometimes be necessary for NEMMCO to build additional constraints to manage system security due to the occurrence of unusual network outage configurations. All the different constraint sets are contained in NEMMCO’s constraint library.

Form of constraint equation

In calculating the least-cost feasible dispatch, some factors can be adjusted or “controlled” through the dispatch, and other factors can be taken as given. The current convention for network constraints used in NEMDE is to include terms that can be controlled by NEMMCO through dispatch on the left hand side (LHS) of the equation, and terms that cannot be controlled by NEMMCO through the dispatch on the right hand side (RHS) of the equation. The limitation imposed on the dispatch is generally a requirement that the sum of the terms on the LHS cannot be greater than the sum of the terms on the RHS.

This is the so-called “fully co-optimised” form of constraint equation. Generator output terms and interconnector flow terms tend to appear on the LHS, while (non-dispatchable) load terms and terms relating to the limits of particular transmission elements tend to appear on the RHS.[48]

---

[47] By convention, load is expressed as a negative number, so strictly speaking the constraint equation would read: GEN 1 + LOAD 1 < 100.

[48] NEMMCO’s responsibilities regarding constraint formulation are set out in rule 3.8. Specifically, clause 3.8.10 states that NEMMCO must determine constraints on dispatch and that these must be

The extent to which increasing a particular term on the LHS utilises the limited flow allowed by a constraint, is reflected in its "coefficient" in the constraint equation. For example, a particular constraint equation may have two generator terms on the LHS, one with a coefficient of 0.3 and the other with a coefficient of 0.9. This means that the output of the generator with the 0.3 coefficient would utilise less of the allowable flow on the applicable network element(s) than the output of the generator with the 0.9 coefficient. This in turn implies that the generator with the 0.3 coefficient could produce more power without violating the constraint than could the generator with the 0.9 coefficient. A negative coefficient for a generator[49] means that its output helps relieve the constraint.

Binding constraints

Congestion can be defined as occurring when there is a binding network constraint. A network constraint is considered to "bind" when it has a direct and limiting impact on the dispatch, meaning that the dispatch (and therefore electrical flows across the network) would be different if the constraint could be relaxed. This will occur when, based on bids and offers, the lowest-cost dispatch would result in the LHS of the constraint equation exceeding the RHS.[50] The dispatch engine automatically takes this into account and in effect scales back the combined output of the LHS terms to the extent required to avoid breaching the constraint limit, so that the LHS *is equal to* the RHS. In practice, there are several thousand constraints that are taken into account by NEMMCO in the dispatch process for any given dispatch interval, and any individual term (e.g. a generator, interconnector flow, or load) might be present in a number of different constraint equations. Further, at any given time, any number of constraint equations might bind.

Relieving congestion

Importantly, inherent within NEMDE is the notion that the marginal economic value arising from an incremental increase in network capability is the same as that arising from the same incremental reduction in generation (or load) that is contributing to the congestion. In other words, there are broadly two ways in which NEMDE can relieve congestion, both of which are of equal value in reducing the costs of dispatch:

- by changing the level of variables under its control (such as generation, dispatchable loads and interconnector flows) so that the RHS limit is not violated—that is, by adjusting one or more of the LHS terms in a constraint, such as by "constraining-on" or "constraining-off" particular generators (see section A.4.2.1 below); or

---

represented in a form that can be later reviewed. Also, clause 3.8.13 specifies that NEMMCO must publish the parameters used for modelling of the constraints.

49 Or for any other term that, by convention, is measured positively. For example, an interconnector by convention will be measured positively when flowing in one direction, and negatively when flowing in the opposite direction.

50 Most constraint equations are LHS < RHS. There are some LHS > RHS and LHS = RHS also.

- by raising the RHS limit of a constraint, thereby relaxing the constraint so that it no longer binds (or binds at a higher level). In order to raise the RHS of the constraint, it is often necessary to change the value of parameters that influence the RHS limit value. Many of these limit parameters are outside the control of the dispatch engine, and to change the parameter values some external action is required by NEMMCO or a Transmission Network Service Provider (TNSP) (e.g. network switching; using NSCS as discussed in Appendix E).

Characterising congestion

Capturing network capability through constraint equations illustrates the point that congestion is what occurs when a constraint equation binds. This provides two alternative ways of characterising congestion. From one perspective, congestion can be described by identifying particular transmission limits that have been reached (and likewise identifying the associated transmission equipment or circuits that cannot accommodate increased power flow). Hence, congestion can be viewed as occurring on a particular point (or across a particular "boundary") on the transmission system. From another perspective, congestion can be viewed as the constraining influence of a network limit on the optimality of generation dispatch. For constraint equations that contain at least one interconnector flow term (in the order of 75% of constraint equations in a normal dispatch interval), the binding of the constraint would affect generation dispatch in at least two regions of the NEM. This illustrates the point that constraints can have far-reaching effects on dispatch and therefore on pricing and settlement outcomes. This is discussed further in the following sections.

Appendix E provides further explanation of the types of constraints used in the NEM.

A.4.2 Congestion and the financial market

A.4.2.1 Wholesale pricing and settlement

The price at which a generator is prepared to supply electricity is its *bid* price.[51] The price it actually receives for this electricity, if it is dispatched, is the *settlement* price. They are usually different.

Generators bid every five minutes, at each dispatch interval. Bids compete against each other in *one* market—the NEM as a whole. In contrast, settlement prices are calculated by NEMMCO, every 30 minutes (a *trading interval*), and there is a separate settlement price for *each region*.

The settlement price in each region is known as the regional reference price (RRP). The RRP is the cost (based on bids and offers) of supplying an additional unit of

[51] In the NEM, the term generator "bid" has the same meaning as generator "offer", and "rebidding" has the same meaning as "re-offering".

electricity at a particular node in the region known as the Regional Reference Node (RRN).[52] The RRNs are generally located in the major load centres in each region, such as at or near the capital city. All generators in a region receive the applicable RRP on the volume of energy for which they are dispatched across the six dispatch intervals that comprise each trading interval (ignoring losses), regardless of whether or not they are located at the RRN. Similarly, all loads in a region pay the applicable RRP for the amount of electricity they consume in the relevant trading interval (again ignoring losses).

Inter-regional price separation

When congestion occurs, it can cause differences in the marginal cost of supplying energy at different locations. To the extent this leads to different marginal costs of supply at different RRNs, the result is that RRPs diverge.[53] This is typically reflected in cheaper generation being "backed off" in low-priced regions as a result of a constraint binding, and more expensive generation being dispatched in high-priced regions.

As in other markets, these inter-regional price differences play an important signalling role in the NEM. In the short-term, they provide signals to generators in higher-priced regions to produce more and to loads in higher-priced regions to consume less. In the longer term, price differences can encourage efficient decisions by market participants concerning when and where to invest in generation and load assets.

Inter-regional price differences also create financial trading risks for participants. If a generator in one region contracts with a party in another region (and references the contract to that other region), the generator will have to manage the risk that the price in its region may differ from the price in the other party's region. The risk of this price difference occurring is called "basis risk". Basis risk and the way it is presently managed in the NEM is discussed below in section A.4.2.2. Appendix C also provides more detail.

Intra-regional "mis-pricing"

In the NEM's regional pricing and settlement structure, congestion can also cause differences in the marginal cost of supply within a region, i.e. between the RRN and other nodes in the region.[54] The marginal cost of supply at each node other than the RRN is referred to as the local or "pseudo" nodal price, and this is calculated as a by-

---

52 In order to calculate RRPs, each constraint must be "correctly orientated" towards the relevant RRN. Constraint equations are correctly orientated if and only if there are no terms involving the RRN in any region in any constraint equation.

53 Price divergences can also be caused by electrical losses and frequency control effects, in the absence of any binding network constraints, but the focus throughout this report is on price divergences caused by network congestion.

54 Appendix C explains how the extent of mis-pricing for any particular point on the network, at any particular point in time, can be calculated.

product of the dispatch process. In other words, participants are effectively dispatched on the basis of a comparison between their bid or offer price and their local nodal price. In general, if their bid or offer price is less than their local nodal price, they will be dispatched to the corresponding volume; if their bid or offer price is greater than their local nodal price, they will not be dispatched. There are exceptions, however (see the discussion about dispatch risk below).

To the extent that congestion causes divergences between the RRP and local nodal prices in the NEM, this impact is not reflected in differences in the prices paid or received by participants located at those other nodes in the region. As noted above, all generators (and loads) within a region receive (and pay) the same price (the RRP) for the energy they are dispatched to produce (and consume) within a trading interval regardless of whether their implied local nodal price is the same as their RRP. This disjunction between the implied nodal prices yielded by the dispatch process and the RRPs used for settlement is commonly referred to as "mis-pricing". Mis-pricing can create risks for participants and promote behaviours that reduce economic efficiency. This is discussed further in Appendix C.

Dispatch risk

Mis-pricing gives rise to physical or "dispatch" risk for generators, because it means that generators may not be dispatched even if they are willing to supply power at or below the prevailing RRP. This could lead to generators having to make "unfunded difference payments" on their contracts. Alternatively, generators could be dispatched even if they are not willing to supply electricity at or below the RRP.

A generator can be required to generate a volume of output that is different to the volume it would wish to generate given the prevailing settlement price (i.e. the RRP). In such situations, generators are referred to as being "constrained-on" or "constrained-off":

- a generator is said to be "constrained-on" when it is dispatched for a quantity that is *greater* than the amount it is willing to produce at the (settlement) price it is paid;

- a generator is said to be "constrained-off" when it is dispatched for a quantity that is *less* than the amount it is willing to produce at the (settlement) price it is paid.

The main risk for a constrained-on generator is that it incurs a loss on the additional output it is required to produce. This might be a *direct* loss, such as where the constrained-on generator is paid less than its avoidable fuel cost of production. Or it might be an *indirect* loss, such as where an energy-constrained generator is required to forego the opportunity to generate at times when it is more profitable.

The main risk for a constrained-off generator is that it is prevented from earning the RRP on the volume of output it would wish to generate at that price. To the extent such a generator is financially contracted, it may be required to make cash difference payments on its contracts that are not funded by its revenues in the spot market. If this occurs at times of very high prices, the cost can be substantial. However, even if a generator is not contracted, being constrained-off implies that it has foregone revenues that it could otherwise have earned if it were not constrained-off.

When a generator is constrained-on, it is said to be "negatively mis-priced", because its settlement price (the RRP) is less than the nodal price used to determine its dispatch volume. Conversely, a constrained-off generator is said to be "positively mis-priced", because its settlement price (the RRP) is greater than the nodal price used to determine its dispatch volume.

In general, volume or dispatch risk caused by mis-pricing can result in:

- constrained-on generators being incentivised to make offers up to the maximum price of $10 000/MWh (or bidding unavailable[55]); and
- constrained-off generators being incentivised to make offers down to the market floor price of -$1 000/MWh (or bidding inflexible).[56]

These sorts of behaviour are referred to as "dis-orderly" bidding. Clearly, such offer prices would not reflect generators' underlying resource costs of production. In an environment of such "dis-orderly bidding", the economic efficiency properties of the bid-based merit-order dispatch approach used in the NEM may be undermined. For example, a generator with a resource cost of $30/MWh that seeks to avoid being constrained-on by offering its capacity at $10 000/MWh may cause the dispatch of a generator with a resource cost of $50/MWh. This leads to a short-term loss of economic welfare to the market of $20/MWh multiplied by the output of the higher-cost generator. Similarly, a generator with a resource cost of $100/MWh may avoid being constrained-off by offering its capacity at -$1 000/MWh, thereby displacing a generator with a resource cost of $30/MWh. This behaviour would cause a welfare loss of $70/MWh over the displaced output.

To the extent that generators cannot manage their dispatch risks by bidding in a dis-orderly manner, they may be inclined to reduce their overall level of financial contracting and/or increase contract premiums. Given that a large proportion (if not all) of most generators' contracts are made with counterparties within their own region (i.e. settled at their local RRP), this could lead to reduced contract competition within that region. The result may be higher retail prices and reduced consumption, reducing allocative efficiency.

In the longer term, dispatch risk caused by mis-pricing may distort generators' locational investment decisions under the NEM's regional wholesale pricing and settlement arrangements. For example, to the extent a proponent of a generation project believes it can manage dispatch risk through dis-orderly bidding, it may be tempted to invest in a relatively high-cost plant in a congested part of the network. Alternatively, if dis-orderly bidding is unlikely to enable a prospective generator to

55 "Bidding unavailable" means a generator is unable to be dispatched (supply output) to meet demand.

56 The extent to which this extreme "dis-orderly" bidding behaviour will occur depends on the extent to which a generator's offer price affects the RRP that it is paid. The smaller the influence of a generator's bid on the RRP, the less inhibited it will be about, say, bidding at -$1 000/MWh.

manage dispatch risk, then even an efficient new entrant may be deterred from investing. Either way, dynamic efficiency would be compromised.

For these reasons, the extent of mis-pricing may provide a useful indication of the potential productive and dynamic costs of congestion. Estimates of the incidence and materiality of congestion in the NEM are discussed in Appendix B.

Importantly, a key implication of both basis risk and dispatch risk is a reduction in generators' willingness to contract. In the case of basis risk, the unwillingness largely relates to inter-regional contracting. In the case of dispatch risk, the unwillingness largely relates to intra-regional contracting.

A.4.2.2 Financial trading

Derivative trading

The NEM is a "gross pool" market, in that virtually all electricity must be bought and sold through the wholesale spot market operated by NEMMCO.[57] Therefore, participants tend not to enter contracts for the physical delivery or receipt of power. However, participants do enter financial contracts in order to hedge their exposures to volatile wholesale spot prices. Financial contracts are used to set or limit the price ultimately paid for and received for wholesale electricity in the NEM by retailers and generators, respectively.

Generators are exposed to the risk of *low* spot prices, so they need to manage cash flows to meet financial obligations relating to operational and maintenance costs, fuel costs and financial charges. Retailers are exposed to the risk of *high* spot prices, so they need to manage their gross margin, i.e. the difference between the price at which they purchase energy and the price they charge customers for the energy they consume. These risks are largely inverse, creating a potential for generators and retailers to hedge their spot price exposures by entering financial contracts with one another.

For example, swap contracts allow participants to agree on a fixed "strike price" that is based on the RRP in a particular region. Where the RRP in a trading interval is *above* the strike price, one counterparty (typically the generator) will make "difference payments" to the other counterparty (typically the retailer or large customer). Where the RRP is *below* the strike price, the retailer (typically) will make difference payments to the generator. As in other financial markets, many other types of contracts exist, such as caps and collars.

There are two options for entering into contracts in the NEM:

- *over the counter (OTC) contracts* involve entering into a bilateral agreement with a known counterparty. OTC transactions can either be negotiated directly with

---

57 The Rules provide certain exemptions, largely related to on-site generation.

other market participants (retailers or generators), or arranged via a broker who offers contracts with standard terms and conditions; and

- *exchange traded contracts* involve entering into a standardised contract with an exchange, such as the Sydney Futures Exchange (SFE). The exchange stands between the buyers and the sellers of futures contracts, so that the buyers and sellers do not trade directly with each other.

The vast majority of trading in electricity derivatives by volume occurs using OTC contracts rather than exchange traded contracts.[58]

Basis risk

Where participants in the NEM have entered financial contracts that are settled against RRPs in other regions, they are vulnerable to differences between their local RRP and the RRP at which those contracts are settled. These differences in RRPs are caused by differences in the marginal cost of supply at different RRNs. This gives rise to financial trading or basis risk. These trading risks largely derive from participants' entry into financial derivative contracts.

For example, a generator in Victoria is settled in the spot market at the Victorian RRP. However, if the generator has entered into a swap contract with a retailer in NSW and this contract is settled at the NSW RRP, the generator faces a risk that Victorian and NSW RRPs could diverge due to binding constraints. If the NSW RRP rises above the Victorian RRP, the generator could be in a position where it has to make difference payments (equal to the difference between the NSW RRP and the strike price of the swap multiplied by the contract quantity) to the NSW retailer, even though the generator has only received the (lower) Victorian RRP on its actual output.

The extent of such inter-regional basis risk depends on the frequency of binding constraints that affect flows between regions and the divergence between regional prices at these times. However, to the extent it arises, basis risk may deter participants from entering contracts with counterparties in other regions. Ultimately, because most retailers typically seek to be fully hedged against spot price volatility, reduced contract competitiveness could be expected to lead to higher contract premiums and higher retail prices. This, in turn, could lead to lower electricity consumption than would otherwise be the case, harming allocative efficiency.

Reduced contract competitiveness could also reduce dynamic efficiency in the longer term by distorting generation and load investment incentives in terms of the timing and location of new plant. For example, higher retail electricity prices could deter or delay investment in new load projects and could encourage generation proponents to invest before it is efficient to do so.

58 However, between 2002/03 and 2006/07, there has been a significant increase in the total volume of exchange traded contracts and the relative decline in the proportion of broker traded OTC contracts. See PriceWaterhouseCoopers (2006), "New Perspectives on Liquidity in the Financial Contracts Electricity Market", PWC, Sydney, November.

Tools are available to enable participants to hedge the inter-regional price differentials caused by congestion. When RRPs diverge, inter-regional flows create Inter-regional Settlement Residue (IRSR) funds, which are equal to the difference between the RRPs of the destination (i.e. importing) and source (i.e. exporting) regions, multiplied by the volume of flow and time duration.[59] Settlements of inter-regional power flows are made from the IRSR funds. Shares to a proportion of the IRSR fund for each directional interconnector are regularly sold at Settlement Residue Auctions (SRAs). Participants can acquire IRSR units to hedge the basis risk of contracts referenced to a different region's RRP.

However, as discussed in Appendix C, IRSR units do not typically provide a firm (i.e. reliable) hedge against contract exposures arising as a result of inter-regional price differentials. To the extent that IRSR units provide an imperfect hedge for basis risk, the actual or potential presence of congestion may deter participants from contracting across regional boundaries and/or demanding higher contract premiums.

An alternative means of managing basis risk is for participants to enter bilateral contracts with a participant in another region. This is equivalent to participants "backing out" of their inter-regional basis risk exposures.

A.5 Summary: the consequences of congestion

In summary, congestion has direct and indirect as well as short- and long-term consequences.

A.5.1 Direct consequences

Higher prices

The most direct impact of congestion is that more expensive generators have to be dispatched to meet demand than would otherwise be the case. This increases the price of electricity for both wholesale and retail customers.

System security issues

Congestion increases the likelihood of system security and supply reliability problems, which then have to be resolved by NEMMCO (the system and market operator).

---

[59] Clause 3.6.5 defines settlement residues due to network losses and constraints. This includes the process for settlement residue distribution and recovery. Rule 3.18 identifies the Settlement Residue Auction as the process by which NEMMCO auctions off rights to these residues, which are allocated to regulated directional interconnectors in the NEM. This rule also sets out the concepts, general auction rules, persons eligible to participate in the auction, auction proceeds and fees and the responsibilities of the Settlement Residue Auction Committee.

A.5.1.1 Indirect consequences

Trading risks for participants

Congestion increases the trading risks—both physical (dispatch risk) and financial (basis risk)—faced by market participants. In response, participants engage in strategies and activities to manage these risks. This leads to participant behaviours that reduce the economic efficiency of the NEM in both the short- and long- terms. For example, "dis-orderly bidding" by generators compromises productive efficiency.

Uncertainty for investors

Congestion can weaken economic signals that support efficient investment decisions by generators and large industrial and commercial users about where to invest in transmission and/or generation. This is a longer-term effect of congestion, which compromises dynamic efficiency in the NEM.