NERA
ECONOMIC CONSULTING

Valuing Load Flexibility and Resource Adequacy Mechanisms in the NEM

Prepared for the Energy Security Board

26 July 2021

CONFIDENTIALITY

Our clients' industries are extremely competitive, and the maintenance of confidentiality with respect to our clients' plans and data is critical. NERA Economic Consulting rigorously applies internal confidentiality practices to protect the confidentiality of all client information.

Similarly, our industry is very competitive. We view our approaches and insights as proprietary and therefore look to our clients to protect our interests in our proposals, presentations, methodologies and analytical techniques. Under no circumstances should this material be shared with any third party without the prior written consent of NERA Economic Consulting.

Project Team

George Anstey
Vakhtang Kvekvetsia
Will Taylor
Willis Geffert
Soren Christian
Sofia Birattari

NERA Economic Consulting
One International Towers
100 Barangaroo Avenue
Sydney NSW, Australia 2000
+61 2 8864 6535
www.nera.com

Contents

Executive Summary

Introduction

The Energy Security Board (ESB) of Australia has been tasked by the Council of Australian Governments Energy Council (COAG EC) to develop reforms for future pathways of the National Electricity Market (NEM) post 2025, as large and small scale renewable generation becomes more prominent. On 30 April 2021, the ESB released its Options Paper that proposes a range of reform pathways.[1]

In particular, the ESB sets out potential pathways under four categories:

- **Resource adequacy mechanisms (RAMs) and aging thermal retirement**, i.e. "arrangements [that] may be needed to deliver a more orderly exit of aging thermal generation and its timely replacement by a mix of new resources which will maintain reliability";[2]
- **Essential system services and scheduling and ahead markets**, or arrangements which ensure that system services are available and efficiently procured in light of less thermal generations;
- **Integration of distributed energy resources (DER) and flexible demand**, which could deliver benefits to customers providing DER as well as lowering total system costs; and
- **Transmission and access**, to ensure that transmission investment is sufficient to accommodate the energy transition.

To support these recommendations, we have been commissioned by the ESB to value the potential benefits of RAMs and DER integration (i.e. the first and third pathways above) to the total costs and operations of the NEM.

For each of these pathways, we have run a series of market simulations using the PLEXOS market optimisation platform, which allows us to identify total system costs, capacity expansion profiles and other outputs over a multi-decade horizon.

Our PLEXOS modelling is based closely on modelling carried out by the Australian Energy Market Operator (AEMO), particularly in its 2020 Integrated System Plan. We have appended AEMO's modelling with assumptions specific to this project and the ESB's advice. For instance, for both pathways, our models assume a higher uptake of electrification of industry and electric vehicles, as well as a more ambitious roll-out of solar power (both large scale and rooftop) and behind-the-meter storage. These additional assumptions come primarily from work we have conducted in parallel on behalf of the Australian Renewable Energy Agency (ARENA) regarding the value of flexible demand.

1 ESB (30 April 2021), Post 2025 Market Design Options – A paper for consultation.

2 ESB (30 April 2021), Post 2025 Market Design Options – A paper for consultation, p.7.

Integration of DER and Flexible Demand

Modelling set up

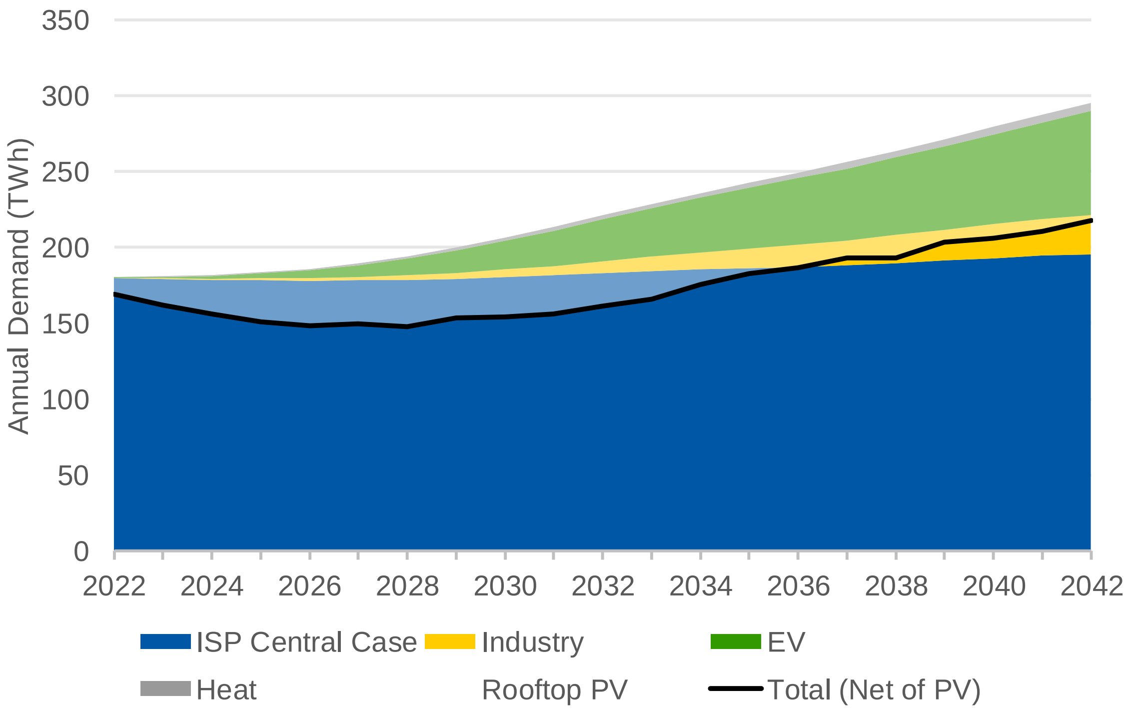
In estimating the value of integration of DER and flexible demand (collectively referred to as demand-side participation, or DSP), we run one model simulation in which DSP is enabled (the "Reform" scenario) and one in which it is not (the "No Reform" scenario). The difference in total system costs between them represents the value that DSP can provide to the system.

DSP provides value primarily through three channels:

- DSP can be activated at times of system shortage, circumventing the need to invest in new capacity to meet that demand, as we assume that DSP is provided to the NEM irrespective of any investment cost to the system; and
- DSP can reduce the cost of outages if it is available to avoid them when there would otherwise be insufficient capacity.
- DSP allows for more efficient use plant, through avoiding the fuel and variable O&M costs on inefficient plant, as well as the start up and shut down costs required to meet shifting load with physical generators. We do not capture the start up and shutdown savings in our modelling.

In both scenarios, we assume that the demand side is identical, with additional electrification from EVs and other sources irrespective of whether it is able to act flexibly. We show this assumption in Figure 1 below. We assume that rooftop PV output offsets end user demand rather than contributing as a generator within the NEM.

Figure 1: Annual Demand Assumption (TWh)



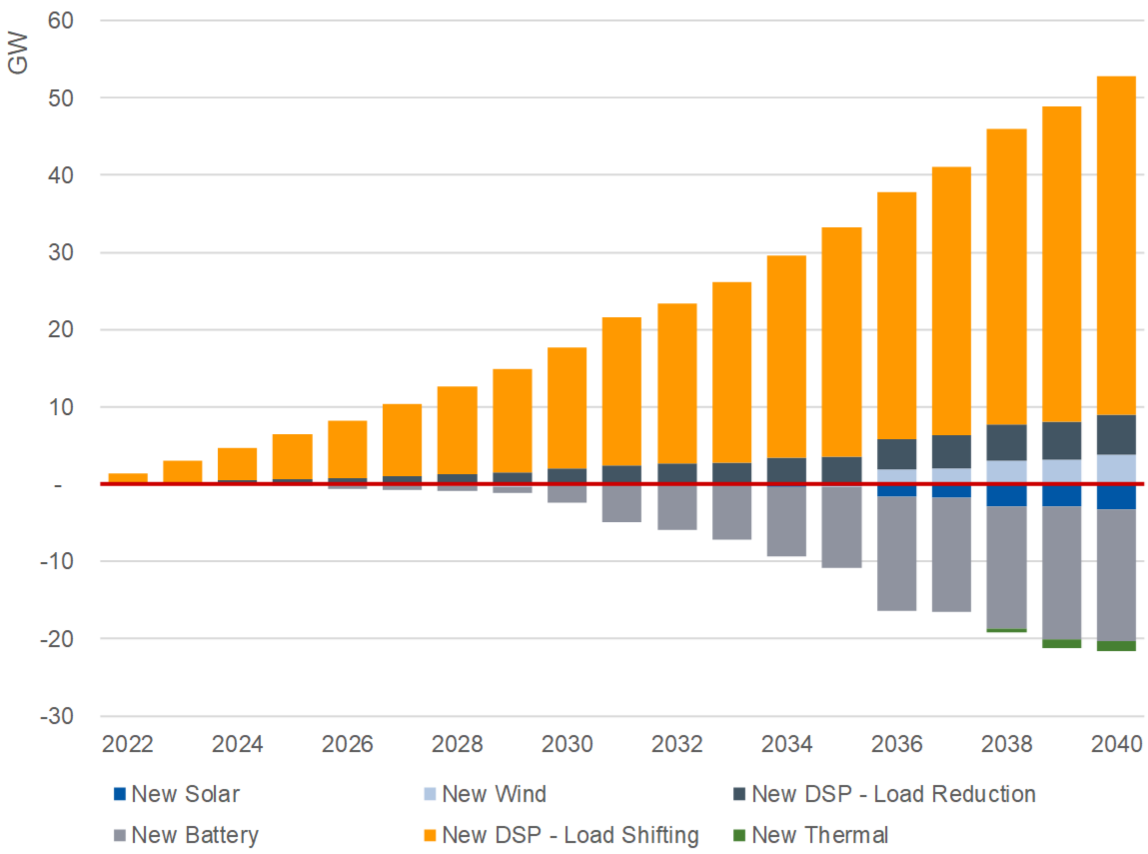
In the Reform scenario we assume that Industry, EVs and electrified heat is able to act flexibly. In addition, we model flexibility from a range of other sources which do not add new electricity demand to the system (e.g. a domestic air conditioning unit that is already electrified, but that does not act flexibly).

Some of these source act as load reducing objects, in that they do not require an increase in consumption at some other time in order to "make up" for the reduced load. These are equivalent to generators. Others act as load shifting objects, and require an increase in consumptions at other times to make up for reduced load. These are equivalent to batteries.

Modelling results

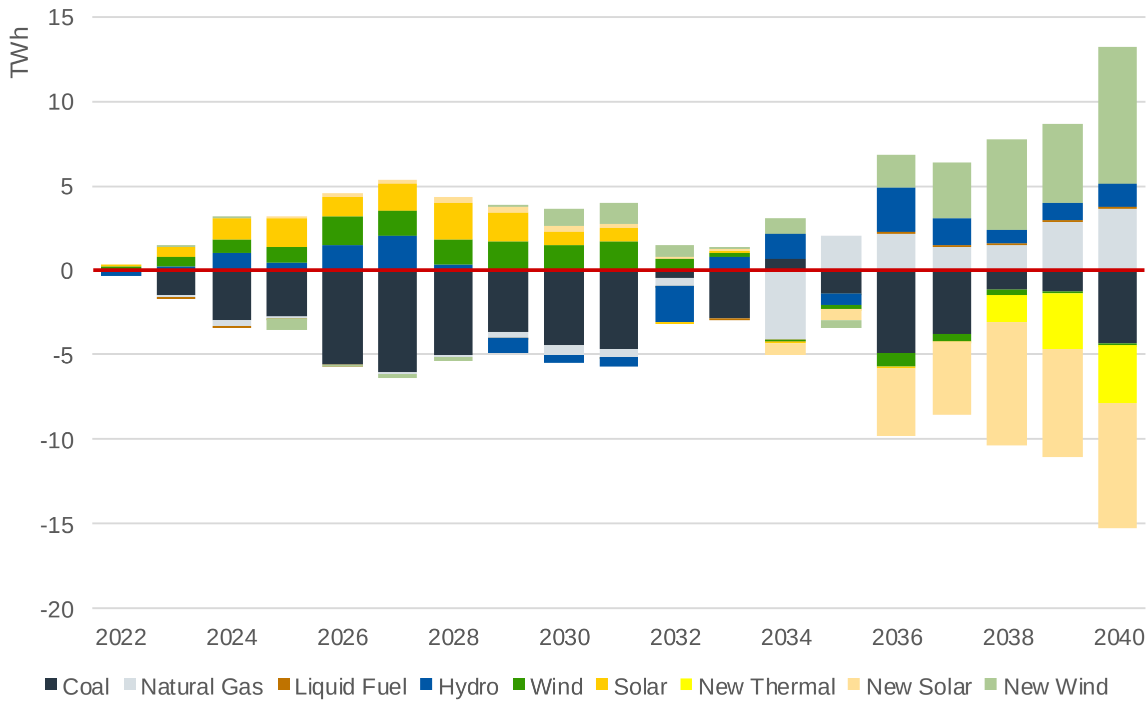
By enabling DSP, we change the pattern of capacity expansion, primarily by reducing the need to invest in large scale batteries. This reflects the fact that (a) the largest potential sources of load flexibility are battery-type objects (behind-the-meter batteries and electric vehicles), and (b) the primary investment required in the absence of load flexibility (as seen both in our No Reform run as well as AEMO's ISP modelling) is large scale batteries. Therefore, the types of DSP we introduce substitute for the types of new capacity required in its absence. We show the difference in capacity expansion profile in Figure 2 below.

Figure 2: Difference in Installed Capacity by Year, Reform minus No Reform (GW)



The make-up of electricity generation differs as well, as the presence of load flexibility in the Reform scenario allows for greater use of wind power and gas generation, relative to solar power and coal. We show the difference between the scenarios' generation profiles in Figure 3 below.

Figure 3: Difference in Generation from Generator Sources, Reform minus No Reform (TWh)



Source: NERA analysis of PLEXOS outputs

Combining the change in capacity expansion and the change in dispatch, we estimate the change in total system costs to 2040, omitting the final two years of our modelling horizon to account for "end-of-model" effects in the PLEXOS model. We present this in Table 1 below.

Table 1: Break-down of Cost Difference, Reform minus No Reform ($million, NPV discount rate 5.9%)

Cost Category	Cost	2021-28	2028-35	2035-40	Total	NPV
Thermal	Fuel Cost	-692	-830	578	**-945**	**-774**
Thermal	VO&M	-207	-182	-45	**-434**	**-282**
Thermal	Start & Shutdown	-	-	-	**-**	**-**
Thermal	FO&M	-0	-2	-36	**-38**	**-14**
Thermal	Annualised Build Cost	-	-	-403	**-403**	**-142**
Renewables	VO&M	49	9	117	**175**	**85**
Renewables	Start & Shutdown	-	-	-	**-**	**-**
Renewables	FO&M	-26	19	336	**330**	**116**
Renewables	Annualised Build Cost	-93	39	902	**849**	**288**
Storage	VO&M	-	-	-	**-**	**-**
Storage	FO&M	-30	-499	-1,037	**-1,566**	**-663**
Storage	Annualised Build Cost	-284	-4,863	-10,453	**-15,600**	**-6,579**
DSP - LR	VO&M	-4	-99	4,734	**4,631**	**1,629**
DSP - LS	VO&M	-	-	-	**-**	**-**
Total	**Total**	**-1,287**	**-6,408**	**-5,308**	**-13,003**	**-6,337**

Source: NERA analysis of PLEXOS outputs

As the table shows, we find that the Reform scenario could save $6,337 million in NPV terms and $13,003 million in undiscounted terms, to 2040 relative to the No Reform scenario. This is primarily due to capacity savings on large scale storage units, with some additional savings delivered through reduced fuel costs (representing the avoided use of thermal generation due to flexibility) and additional expenses in the Reform scenario from resorting to the most expensive demand response technologies in times of peak.

Resource Adequacy Mechanisms

Modelling set up

In estimating the value of a RAM, we quantify the value a RAM could deliver through two avenues:

- By sending a price signal for new investment that reflects the true cost of outages and shortages that new investment can avoid, a RAM should deliver more investment than would otherwise be underdelivered in the absence of one.
- By sending a signal to existing thermal generation to remain available to the end of their expected life, a RAM can avoid the unexpected early retirement of existing units.

We represent these two effects in the modelling as follows:

- We represent the price signal sent to investment in new capacity by varying the Value of Lost Load (VoLL) in the model, as well as the cost to dispatch the Reliability and Emergency Reserve Trader (RERT). When optimising the system, we allow RERT to dispatch at $1 below the relevant VoLL.
  - In the RAM scenario, we use a Value of Lost Load (VoLL) of $21,247/MWh, equal to the VoLL that AEMO uses in its ISP modelling. Because the ISP is calibrated to identify the optimal investment pattern, we treat this value as the true VoLL. Our

approach is a good proxy for the existence of a RAM, regardless of how such a mechanism is designed in practice.

– In the No RAM scenario, we use a VoLL of $7,500/MWh to optimise the system, considerably below the true VoLL of $21,247/MWh. This represents a world where the signals to invest in reliable capacity are inefficiently low due to explicit price caps (currently $15,000/MWh) and implicit price caps (e.g. investors may excessively discount revenues they could receive from rare, sporadic, high-value events). We undervalue VoLL and the cost of RERT dispatch when *optimising* the system, but when quantifying the total cost of the system, we use the true VoLL of $21,247/MWh and the cost of RERT of $21,246/MWh.

- With a RAM in place, we assume that thermal units close in line with the ISP Central Case assumptions, with the exception that coal plants retire in line with the ISP Step Change assumptions. We evaluate the case where retirement happens in a disorderly fashion by considering a scenario in which three large coal plants (Vales Point, Yallourn and Gladstone) close one year earlier than expected in the absence of a RAM to support their continued operation. We apply this adjustment only to the short-term operational model, meaning that we do not allow for the capacity expansion profile to anticipate these closures.

Therefore, our core RAM modelling runs comprise three rather than two runs:

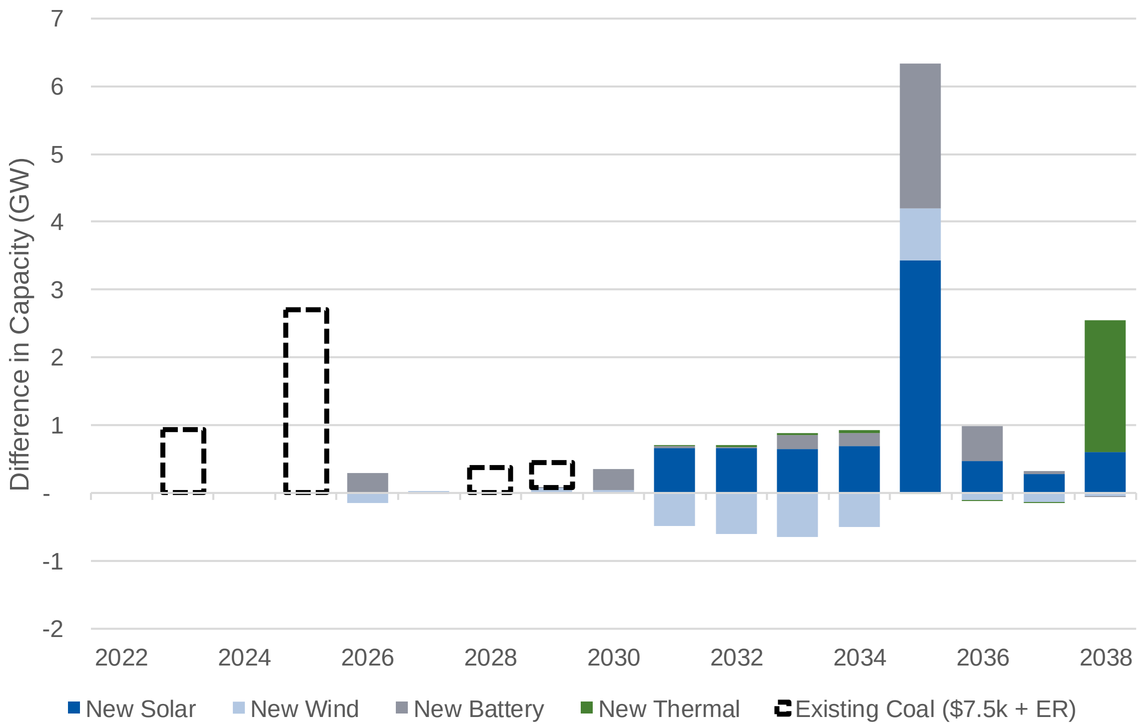
- A RAM scenario based on a $21,247/MWh VoLL. We refer to this as the “$21k” scenario.
- A No RAM scenario based on a $7,500/MWh VoLL. We refer to this as the “$7.5k” scenario.
- A No RAM scenario based on a $7,500/MWh VoLL and the unexpected early retirement of the three coal units described above. We refer to this as the “$7.5k + Early Retirement” (or “$7.5k + ER”) scenario.

Modelling results

We present modelling outcomes to the end of 2037/38, beyond which the modelled levels of unserved energy reflect end-of-period effects rather than actual changes in efficient capacity expansion and dispatch.

In Figure 4, we show the difference in the total capacity mix between the $21k scenario and the $7.5k scenarios. The RAMs scenario allows for greater commissioning of new solar capacity, often backed with storage, and less wind capacity. It also brings forward an investment in new thermal capacity by one year. In comparison with the $7.5k + Early Retirement scenario, it also allows for existing coal to stay on the system for longer.

Figure 4: Difference in Capacity, $21k - $7.5k

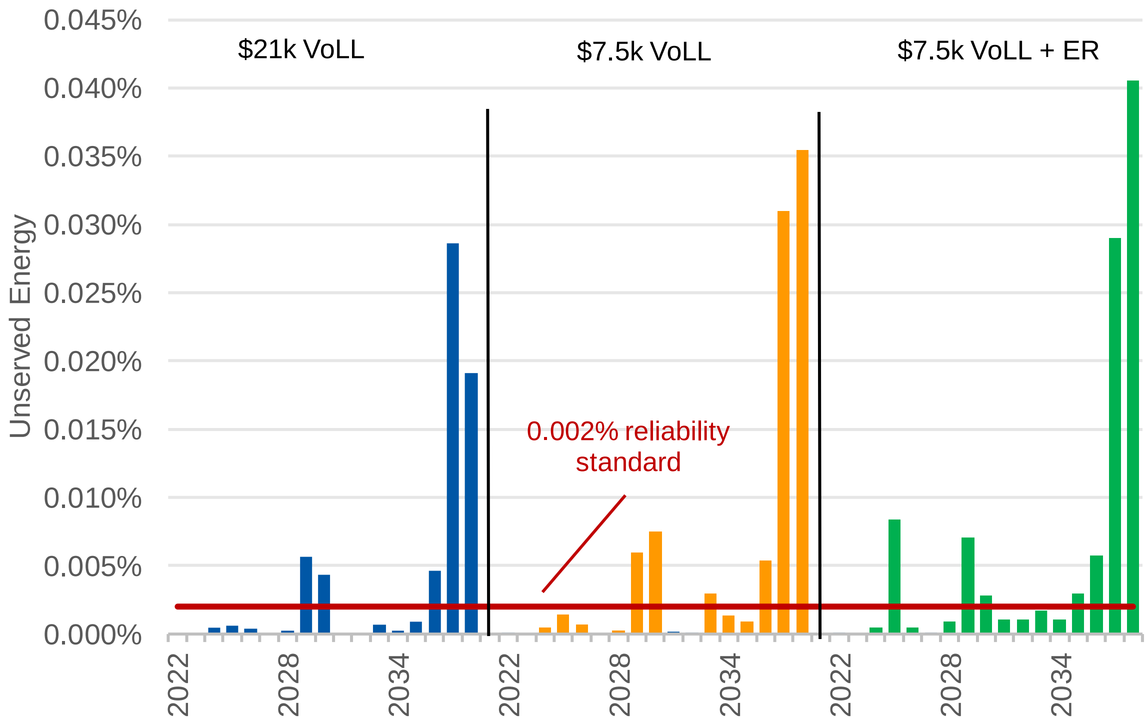


Source: NERA analysis of PLEXOS outputs

Under the modelled scenario, the existence of a RAM tends to increase investment, and so investment costs. This will be more than offset by savings in the avoided Unserved Energy (USE) and RERT dispatch. In Figure 5 below, we demonstrate that unserved energy is generally lower in the $21k scenario than in the $7.5k and, especially, the $7.5k + Early Retirement scenario.

In all cases, even with a RAM in place, we model excessively high levels of USE by the final few years of the modelling period. This is primarily a function of how PLEXOS sequentially optimises the system, which uses more simplified assumptions when planning capacity expansion than when choosing how to dispatch. Additionally, the "long-term" run likely gives excessive credit to batteries for meeting the reliability standard, which is not effective when a reliability event lasts longer than the storage duration of a battery. This is an area that requires further investigation which we were unable to conduct in the time available to us. We understand that the ESB will consider these specific modelling outcomes more fully during the detailed design phase of the RAM.

Figure 5: Unserved Energy



Source: NERA analysis of PLEXOS outputs

As described above, there are two primary drivers of value that a RAM could cause: an increase in overall capacity (and fixed O&M) costs more than offset by a reduction in outages or near outages (i.e. dispatch of RERT). There are other drivers of value, e.g. different dispatch patterns driven by different fuel sources, but these can be thought of as side effects of the core tradeoff between the cost of an outage and the cost of building to avoid it.

We combine these two effects (and others) in In this section, we quantify the change in overall costs that results from the introduction of a RAM. In all runs, we quantify the cost of USE and RERT based on the true VoLL ($21,247/MWh) or $1 below for RERT rather than necessarily the value that the model sees when optimising the system.

We show the savings in our core runs in Table 2 and Table 3 below, separately quantifying the value driven just by the signal to new investment from the value driven by the signal to both new investment and existing capacity.

Table 2: Total Cost Savings: $21k - $7.5k

Cost Category	Cost	2021-28	2028-35	2035-38	Total	Total - NPV
Thermal	Fuel Cost	- 13	- 96	- 60	**- 170**	**- 92**
Thermal	VO&M	16	- 16	- 6	**- 5**	**3**
Thermal	FO&M	- 4	- 0	6	**1**	**- 1**
Thermal	Annualised Build Cost	-	15	67	**82**	**33**
Renewables	VO&M	- 3	- 24	- 2	**- 29**	**- 15**
Renewables	FO&M	- 3	- 45	0	**- 48**	**- 24**
Renewables	Annualised Build Cost	- 17	- 118	11	**- 123**	**- 65**
Storage	VO&M	-	-	-	**-**	**-**
Storage	FO&M	5	28	11	**44**	**23**
Storage	Annualised Build Cost	47	445	247	**740**	**363**
System	USE, DSP, RERT	- 141	- 593	- 2,394	**- 3,129**	**- 1,354**
Total	**Total**	**- 113**	**- 404**	**- 2,120**	**- 2,638**	**- 1,130**

Source: NERA analysis of PLEXOS outputs

Table 3: Total Cost Savings: $21k - $7.5k + Early Retirement

Cost Category	Cost	2021-28	2028-35	2035-38	Total	NPV
Thermal	Fuel Cost	- 460	- 284	- 48	**- 793**	**- 520**
Thermal	VO&M	- 18	- 15	- 5	**- 37**	**- 23**
Thermal	FO&M	852	57	6	**915**	**714**
Thermal	Annualised Build Cost	-	15	67	**82**	**33**
Renewables	VO&M	- 7	- 21	- 4	**- 31**	**- 16**
Renewables	FO&M	- 3	- 45	0	**- 48**	**- 24**
Renewables	Annualised Build Cost	- 17	- 118	11	**- 123**	**- 65**
Storage	VO&M	-	-	-	**-**	**-**
Storage	FO&M	5	28	11	**44**	**23**
Storage	Annualised Build Cost	47	445	247	**740**	**363**
System	USE, DSP, RERT	- 563	- 567	- 2,757	**- 3,886**	**- 1,777**
Total	**Total**	**- 163**	**- 505**	**- 2,471**	**- 3,138**	**- 1,294**

Source: NERA analysis of PLEXOS outputs

A RAM that is calibrated to signal the true value of reliability can provide value to the system by signaling investment in new capacity necessary to avoid loss of load. In the absence of a RAM, explicit and implicit price caps may mean that investors may not see the value in investing in capacity that is societally optimal.

The value of a RAM increases if it is effective in maintaining existing plant to their anticipated closure date. The NEM could suffer high costs of outages and RERT dispatch in periods where a plant was expected to be available but it is not.

In summary, we estimate that a RAM could reduce total system costs by between $1,130 million and $1,294 million in NPV terms (with a 5.9 per cent discount rate), or between $2,638 and $3,138 billion in undiscounted terms, to 2038.

These results show the benefits of a RAM presuming that the market currently fails to deliver an efficient level of capacity (e.g. by assuming the presence of explicit and implicit price caps) and that the RAM is an efficient design. The benefits assessed do not include any estimate of the implementation costs of the RAM. In the absence of market failures in the current NEM design or given an inefficient design for the RAM, a RAM may not lead to social benefits. On the other hand, if the current distortions in the NEM design were larger than we assumed the benefits would be commensurately bigger.

Conclusion

In conclusion, we find that the integration of DER and flexible demand, and the introduction of a RAM, could reduce system costs to the NEM over the coming decades as renewable penetration rapidly expands:

- The integration of DER and flexible demand could reduce system costs by $6,337 million in NPV terms and $13,003 in undiscounted terms, to the end of the modelling period, driven primarily by the avoided need to build new storage capacity.
- The introduction of a RAM could reduce system costs by between $1,130 million and $1,294 million in NPV terms, or between $2,638 and $3,138 billion in undiscounted terms, to 2038. Additional capacity costs are more than offset by a reduction in dispatch costs and unserved energy.

To some extent, these benefits are overlapping and therefore cannot necessarily be added to one another. For instance, a RAM may encourage some of the same investment that is avoided by the existence of demand flexibility at times of system stress. Similarly, both interventions will tend to reduce the cost of shortages in the system.

On the other hand, neither conclusion cancels out the other: A RAM would likely reduce system costs with or without high integration of DER and flexible demand; and DER and flexible demand would likely reduce system costs with or without the existence of a RAM.

1. Introduction

The Energy Security Board (ESB) of Australia has been tasked by the Council of Australian Governments Energy Council (COAG EC) to develop reforms for future pathways of the National Electricity Market (NEM) post 2025, as large and small scale renewable generation becomes more prominent. On 30 April 2021, the ESB released its Options Paper that proposes a range of reform pathways.[3]

In particular, the ESB sets out potential pathways under four categories:

- **Resource adequacy mechanisms (RAMs) and aging thermal retirement**, i.e. "arrangements [that] may be needed to deliver a more orderly exit of aging thermal generation and its timely replacement by a mix of new resources which will maintain reliability";[4]
- **Essential system services and scheduling and ahead markets**, or arrangements which ensure that system services are available and efficiently procured in light of less thermal generations;
- **Integration of distributed energy resources (DER) and flexible demand**, which could deliver benefits to customers providing DER as well as lowering total system costs (we collectively refer to DER and flexible demand as Demand Side Participation, or DSP); and
- **Transmission and access**, to ensure that transmission investment is sufficient to accommodate the energy transition.

Following an industry consultation on each of the reform pathways, the ESB will present recommendations to Ministers in the middle of 2021.

To support these recommendations, we have been commissioned by the ESB to value the potential benefits of RAMs and DSP integration (i.e. the first and third pathways above) to the total costs and operations of the NEM.

For each of these pathways, we have run a series of market simulations using the PLEXOS market optimisation platform, which allows us to identify total system costs, capacity expansion profiles and other outputs over a multi-decade horizon.

In this report, we set out our findings in each of the two tasks as follows:

- In Chapter 2, we describe our overall PLEXOS modelling framework, which we use in both workstreams;
- In Chapter 3, we set out our approach to modelling the benefits of DER and present our results; and
- In Chapter 4, we set out our approach to modelling the benefits of RAMs and present our results; and
- In Chapter 5, we conclude.

[3] ESB (30 April 2021), Post 2025 Market Design Options – A paper for consultation.

[4] ESB (30 April 2021), Post 2025 Market Design Options – A paper for consultation, p.7.

2. Technical PLEXOS Modelling Background

In this chapter, we provide a technical summary on how we build up our baseline model, before augmenting it to include sources of DSP and RAMs. This chapter proceeds as follows:

- In Section 2.1, we describe our core PLEXOS model, taken from published PLEXOS models run by the Australian Energy Market Operator (AEMO);
- In Section 2.2, we describe adaptations to the model to include further sources of demand beyond what is included in AEMO's models; and
- In Section 2.3, we describe our approach to approximating short-term price volatility, which is not possible to model in PLEXOS but is nonetheless important to quantifying the value of RAMs and DSP.

2.1. We Use PLEXOS to Model the Benefits of System Flexibility

We model the primary benefits of demand flexibility reform using the specialist energy modelling software PLEXOS. Broadly speaking, we build a "baseline" model based on modelling assumptions published by AEMO, which we then augment to quantify changes in outcomes due to the presence of higher DSP or RAMs. For each workstream, we run one or more "counterfactual" runs without the policy change of interest and compare to the results where the policy is implemented.

Modelling the market using PLEXOS has a number of key advantages for quantifying the benefits of flexibility:

- PLEXOS is an industry-leading platform for modelling electricity markets for which we and stakeholders already have access to published versions run by AEMO for the NEM, namely the Integrated System Plan (ISP) and the Electricity Statement of Opportunities (ESOO) models;
- The model optimises the long-term least cost expansion planning of generation and the short-term optimal dispatch patterns. Accordingly, a single modelling engine can be used to estimate both the short term and dynamic benefits;
- As a publicly-recognised modelling platform, stakeholders have much greater clarity and understanding of our results than if we were to use a bespoke, proprietary algorithm. (At NERA we have our own proprietary modelling tools, however reliance on PLEXOS is much more transparent and therefore appropriate for modelling with implications for a wide range of stakeholder groups); and
- Reliance on a standardised modelling logic already run and validated by AEMO reduces the need for checking and auditing all of the inputs and operation in our baseline model runs.

Broadly speaking, PLEXOS optimises over two time horizons sequentially:

- A "Long-term" (LT) run optimises the capacity expansion of the whole system, particularly the construction of new generating or storage capacity and, if allowed by the model set-up, the retirements. In our case, however, we include pre-determined retirement dates of existing capacity based on AEMO's assumptions. To determine the optimal expansion, the optimiser compares forecast demand (plus any reserve requirements) to existing sources of generation and candidate sources of generation, and determines the least cost way of meeting the demand and reserve requirement.

Often this happens simultaneously over a long horizon, though not necessarily the full horizon. In our case, as in the published ISP model, we break the 21-year horizon into three blocks of seven years each, so PLEXOS first identifies the least-cost way to meet demand in the first seven years, then the cheapest way to meet demand in the second seven years, and so on. In each step, it takes as given any existing capacity that is not scheduled to retire, including the capacity it has chosen to build in previous steps.

- A "Short-term" (ST) run takes as given the available capacity as optimised in the LT run above, and determines the least-cost way (in NPV terms) of dispatching the system to meet demand in every period (half-hour in the NEM). In our models, we optimise the entire 21-year period in a single step, though it is possible to estimate this sequentially in shorter steps.

We run a 21-year PLEXOS simulation based initially on a combination of AEMO's 2020 ISP Central Case model and its 2020 ESOO model, with a simulation period beginning on 1 July 2021 and ending on 30 June 2042. Unless otherwise specified, references to a single year in the modelling horizon refer to the financial year ending in June of that calendar year (e.g. "2032" = July 2031-June 2032). Due to "end-of-period" effects in PLEXOS (where PLEXOS assumes the world ends at the end of the modelling horizon), we truncate our outputs after 2040 for the DSP task and after 2038 for the RAMs task.[5]

The two sources and our uses of them differ as follows:

- **The 2020 ISP model** is based on the same 21-year horizon we use, and is primarily used by AEMO for determining optimal long-term expansion. AEMO publishes its PLEXOS modelling files and results for the LT run of its Central Case simulation, which we use as a starting point for our modelling. We make some amendments to include assumptions from the AEMO's Step Change simulation alongside the Central Case assumptions.

  From this model, we draw all of our long-term assumptions, aside from those we specifically augment as described further below. More specifically, we use the following, which are all included in the published ISP assumptions book:[6]

  – Embedded DSP, grouped into five "bands", with prices from $300 to $14,700/MWh. In the Counterfactual run of the DER workstream and in all runs of the RAMs workstream, we leave this DSP enabled. In the "High DER" model runs, we remove this DSP and replace it with explicit assumptions around the source of flexibility.

  – Technical characteristics of existing generators and storage units, including their capacities, heat rates (if thermal), retirement dates, and fixed and variable O&M (FOM and VOM) costs;

  – Technical characteristics and costs of candidate new entrant plants, by technology, state and year. We use assumptions from the ISP Step Change for this input. The

5 The end-of-period effects are particularly acute for Unserved Energy (USE) and Reliability and Energy Reserve Trader (RERT) dispatch, which are more directly relevant for the RAMs analysis, so we truncate a larger period for this task.

6 AEMO (July 2020), 2020 Integrated System Plan, Assumptions Book.

costs include build costs (annuitised using AEMO's assumed weighted average cost of capital, or WACC, of 5.9 per cent), FOM, VOM, and heat rates which, in combination with fuel costs, determine the short-term running costs;

- Output traces for solar and wind plants, by half hour and Renewable Energy Zone (REZ). These are expressed as a proportion of the maximum, and so are multiplied by the relevant capacity in that year and REZ to determine the actual output in each half-hour period. As in the ISP model, these are based on a "rolling reference year", such that the production patterns actually observed between 2010 and 2019 are assumed to repeat sequentially.
- Build limits on new units by technology (for renewable energy) and in total, by location.
- Fuel costs by fuel type and year;
- In determining whether to build capacity to meet demand, or instead to allow demand to be unmet, the model assumes a Value of Lost Load (VoLL) of $21,247/MWh. We use this assumption in all scenarios except the No RAMs scenarios, in which we use a lower value of $7,500/MWh.
- End-user demand by year and state. This is based on AEMO's 10 per cent Probability of Exceedance (POE10) case, i.e. AEMO believes that there is a 10 per cent chance that actual demand will exceed this level. On a half-hourly basis, end-user demand is shaped using the same "rolling reference year" discussed above, so that consumption patterns (rather than its absolute level) matches the patterns actually observed between 2010 and 2019. As discussed in Section 2.2, we overlay this with additional sources of demand to reflect a more ambitious path of decarbonisation and electrification.
- Renewable Energy Targets (RET) by state and year, for which we use the ISP Step Change assumptions. For example, renewable energy output in Victoria must be 40 per cent of total Victorian generation by 2025, and 50 per cent by 2030. The intervening years are interpolated between those targets.
- Reserve margins by state.

▪ The **2020 ESOO Model**. This model captures many of the short-term operational constraints relevant to the NEM, such as short-term thermal constraints and contingencies between different units. However, AEMO's assumptions only cover 10 years, from 2021 to 2031, so we are only able to use assumptions from the ESOO which can be extrapolated over the full 21-year horizon. We take the following properties from the ESOO model:
  - Forced outage rates for existing thermal generators;
  - Outage factors as a proportion of rated capacity;
  - Minimum time to repair;
  - Minimum stable generation levels, for all thermal generation except CCGT and gas-fired steam turbines;

▪ Additionally, we use inputs to AEMO's Time-Sequential Model for the following:
  - Minimum stable generation for CCGT and gas-fired steam turbines;
  - Ramp up and ramp down rates.

2.2. Adjustments to Reflect Ambitious Electrification and Decarbonisation

ESB has asked us to introduce additional assumptions to reflect an ambitious level of electrification and decarbonisation relative to the assumptions embedded in the 2020 ISP Central Case scenario. These adjustments allow us to capture the value of DSP and RAMs in a world where they are more likely to be needed and is more in line with decarbonisation objectives that the ESB's pathways work towards.

Some of these come from the ISP Step Change scenario, as described above:

- Lower capital costs for new renewable generating technologies; and
- More ambitious renewable energy targets by state.

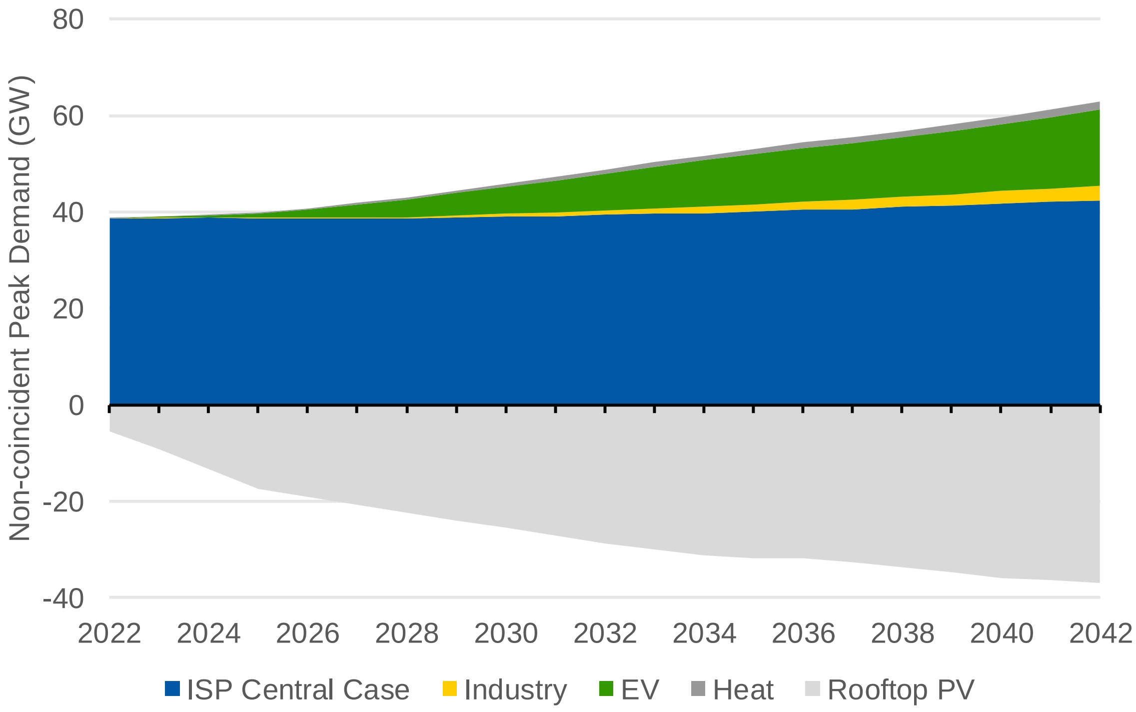
Beyond these assumptions which derive from the 2020 ISP, we draw on assumptions that we have developed through a parallel piece of work we are conducting on behalf of the Australian Renewable Energy Agency (ARENA) to assess the value of load flexibility. In that piece of work, which is still in draft form as of July 2021, we have drawn on techno-economic analysis conducted by our partners at Energy Synapse, which has identified a range of ambitious but plausible levels of electrification of the following sources:

- Electric vehicles (EVs). In an ambitious state of the world, we assume there could be around 14 million EVs in the NEM region by 2042. Some of these are already embedded in the ISP Central Case forecast, so we add an overlay to demand based on the *incremental* uptake of EVs. This adds 18 GW in demand above the level embedded in the ISP Central Case, shaped according to AEMO's trace of EV charging.
- Residential and commercial heat. In an ambitious state of the world, we assume there could be up to 2 GW in additional demand from electrified heat. This additional demand is shaped based on the trace of system demand in the winter only.
- Industrial process heat. In an ambitious state of the world, we assume there could be 3 GW in additional electrification from industry, operating at baseload.

Additionally, we assume that there could be 38 GW of distributed, behind-the-meter PV generation (i.e. rooftop solar) by 2042 beyond what is embedded in the ISP demand forecast already. Because these small-scale PV owners are unlikely to participate actively in the NEM, we net this capacity *off* of system demand, shaped according to AEMO's rooftop PV trace.

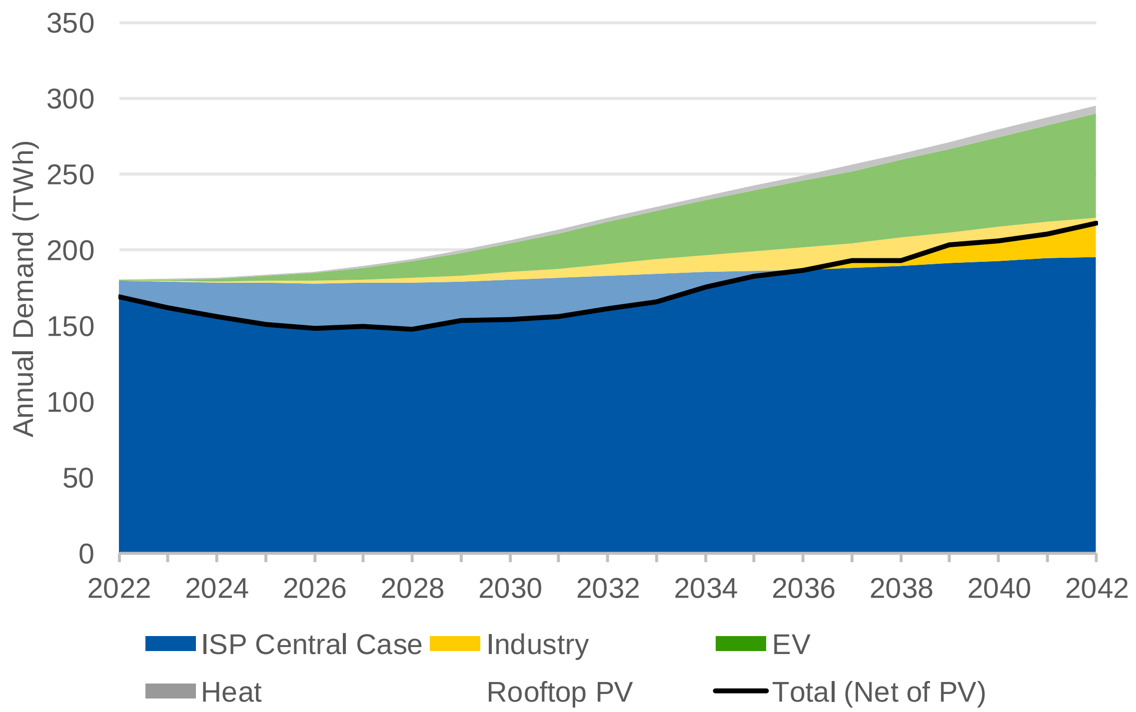
In Figure 2.1 below, we show the peak demand assumption from each source, plus the baseline ISP Central Case demand. Note that each area represents the peak demand of that source in isolation. The peak demand of one source may not happen at the same time as another, and so the values cannot simply be added together to derive system peak demand.

Figure 2.1: Non-coincident Peak Demand Assumption (GW)



In Figure 2.2, we show the annual demand from each source, in TWh. Because these are annual totals, they can be added across sources, and we explicitly net off Rooftop PV in this chart.

Figure 2.2: Annual Demand Assumption (TWh)



2.3. Short-run Price Volatility

When PLEXOS optimises dispatch in the ST runs, it starts with the cheapest units (in terms of their short-run costs) and brings in capacity from other units until generation output matches demand in each period, subject to operational constraints such as ramp rates and minimum stable loads.

In each half-hour period, the resulting electricity price will be the short-run cost of the most expensive unit that is required to meet demand in that period, i.e. its fuel costs plus its variable O&M costs. Based on the technical and cost assumptions derived from AEMO's models, this means that the price of electricity almost never rises above $300/MWh, roughly equal to the short-run costs of the most expensive units in the NEM.

In actuality, prices very frequently rise above this level for a range of possible reasons, including: (i) strategic bidding from operators who believe they may set the price in a particular period;[7] and (ii) costs of unanticipated operation (e.g. start costs that must be recovered over a small number of operating hours). In Table 2.1, we show the frequency with which prices exceeded each price level in the 2020 calendar year.

Table 2.1: 2020 Actual NEM Price Volatility

Price Threshold	NSW	QLD	VIC	SA	TAS
300	52	45	31	99	19
500	42	21	17	15	12
1,000	37	12	17	15	10
2,000	31	6	14	10	10
5,000	16	0	11	5	0
10,000	7	0	8	2	0

Source: NERA analysis

For instance, in South Australia in 2020, there were 99 half-hour periods when the market price was above $300/MWh (i.e. 0.006 per cent of all periods); 15 when the market price was above $500/MWh (and also above $1,000/MWh) (0.001 per cent of all periods); and so on.

This short-run price volatility is not possible to simulate directly in PLEXOS, but is important to estimating the value of load flexibility, particularly with respect to the avoided costs of dispatching expensive generators. This is because many of the sources of flexibility we model are more expensive than $300/MWh, and so PLEXOS would never choose to operate them compared to the apparently cheaper costs of the most expensive generators included in our assumptions. In reality, these sources are cheaper than the highest prices actually observed, and so would operate and shave the peak prices accordingly.

We therefore introduce short-run price volatility as follows:

- From a baseline run, we identify the periods with the highest short-run prices in each year. In particular, we select the number of periods based on how frequently prices rose

[7] Due to short-term operational constraints, some operators may hold short-term market power. These operators could take advantage of this market power by offering very high prices for that period.

above $300/MWh in 2020. For example, we choose the 52 most expensive periods in New South Wales in each year, the 45 most expensive periods in Queensland, etc.

- In these periods only, we introduce plant-specific adders to the Variable O&M (VOM) of gas plants, ranging from $300 to $10,000/MWh. We apply a different adder to each plant, such that all gas units have a lower adder than liquid fuel units, and more efficient plants within a technology have a lower adder. Once the units are ordered, we apply an adder that increases in even intervals from $300 to $10,000/MWh.
- Therefore, in these hours, the model may be forced to call upon one or more units that is more expensive than it usually appears, causing a spike in the power price. The size of the spike may be mitigated if DSP can be called in place of the inflated price of thermal plants, representing one form of value provided by DSP.

3. Integration of DER and Flexible Demand

DER and flexible demand (collectively DSP) can provide value to a system and reduce total system costs because they can provide similar services as physical generating or storage capacity. As a result, by having a large volume of DSP available, the NEM could avoid investing in new physical capacity (and hence save on the investment costs). It could also avoid the cost of dispatching expensive capacity or of disconnecting load, although the cost of dispatching DSP (if available) is not necessarily lower than the cost of dispatching expensive physical capacity (if built). Finally, DSP can reduce the need to frequently start and stop physical capacity as demand fluctuates. However we do not capture these cost savings in this report.

In this chapter, we describe our approach, results and conclusions with respect to the integration of DSP. This follows closely with a subset of our parallel piece of work for ARENA. We estimate the value of flexible demand coming from a wide range of sources, some of which are associated with the additional electrification described in Section 2.2:

- Residential heating, air conditioning, hot water and pool pumps;
- Commercial heating, air conditioning and hot water;
- Electric vehicles;
- Behind-the-meter (BTM) storage; and
- Electrified industry.

This chapter proceeds as follows:

- In Section 3.1, we describe our modelling set up in detail;
- In Section 3.2, we present the results of our modelling; and
- In Section 3.3, we draw conclusions from our modelling.

3.1. Description of Modelling

Fundamentally, our analysis is based on a comparison between a world with DSP (“Reform” scenario) and without DSP (“No Reform” scenario).

In the No Reform scenario, we keep all of the assumptions set out in Chapter 2, including the additional electrification and rooftop PV roll-out, but do not allow any of it to behave flexibly. For example, we assume a high uptake of EVs, but that users cannot deviate from the charging profile assumed in the ISP’s EV trace.

In the Reform scenario, we allow a wide range of DSP sources to participate on the supply side of the market, based on the findings of Energy Synapse’s techno-economic analysis. In its work on behalf of ARENA, Energy Synapse conducted an extensive study involving, desk-based research, bilateral interviews with a range of industry stakeholders, and a round-table workshop with a wider set of industry stakeholders. The output of Energy Synapse’s work was a projection of what level of electrification and demand response could be possible from each industry.

All sources of flexibility, if enabled, act as a virtual generator or storage unit, from the perspective of the PLEXOS model. In reality, a residential air conditioning system may *reduce* its load in response to a price signal, *reducing* system-wide demand, but its demand reduction is equivalent to a physical generator providing the load that would otherwise be required to meet the unabated demand.

We distinguish between load reducing and load shifting sources of flexibility. A load reducing source would turn down its demand, without a compensating increase in demand at some other time. For example, an electrified metal refinery that generally runs baseload would not be able to "catch up" on any of its lost electricity consumption in other hours. We treat these sources as virtual generators.

A load shifting source, by contrast, would increase its consumption in other periods to compensate for the foregone consumption in the period when it is called. For example, if an electrified water heater switches off in a given period, it will have to run in other hours that it otherwise would not have in order to keep the water hot. We treat these sources as virtual storage units, and they behave similar to batteries.

In some cases, we apply restrictions on when or how often load reducing and load shifting resources are available. For example, residential heating could not act flexibly in the summer because the underlying activity (the heating itself) is switched off in the summer. We also include assumptions on the price required to be paid to load flexibility resources, equivalent to the variable O&M of operating a physical generator or battery.

In the remainder of this section, we set out the supply-side characteristics of the load reducing and load shifting resources.

3.1.1. Residential and commercial

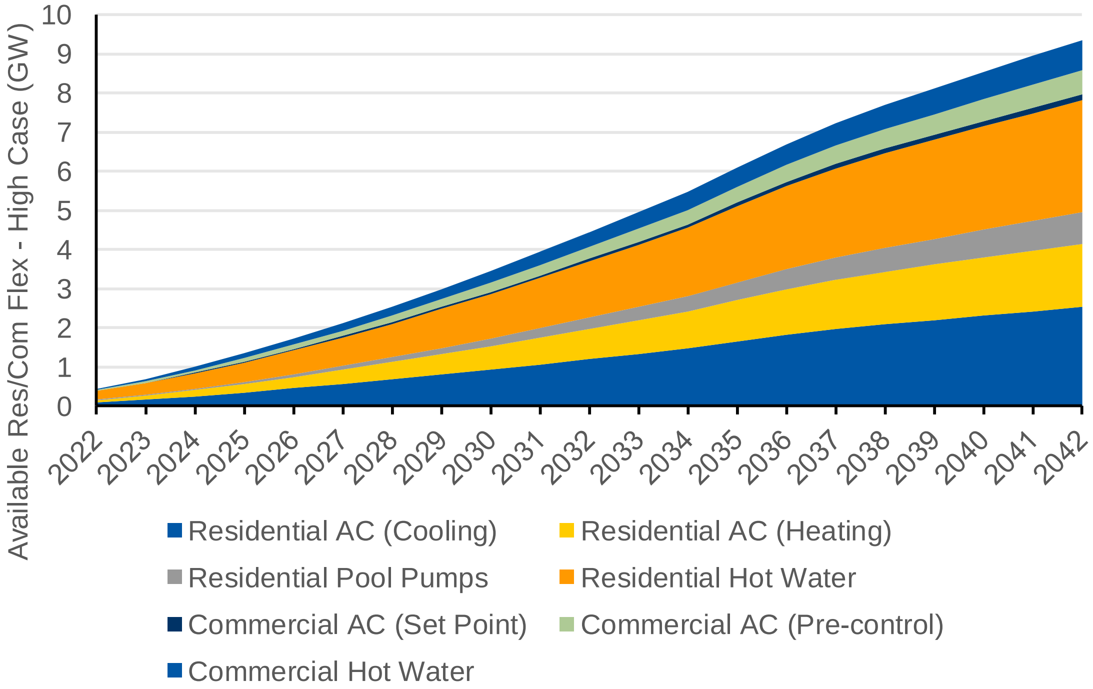
Below, we describe our approach to modelling residential and commercial sources of flexibility. With the exception of electrified residential heat, these sources of flexibility would come about by enabling the flexibility of existing load, either through upgraded equipment or new commercial arrangements.

We consider the following sources of flexibility in this category:

- Residential air conditioning (cooling), i.e. summer only;
- Residential air conditioning (heating), i.e. winter only;
- Residential pool pumps;
- Residential hot water;
- Commercial air conditioning (set point), i.e. setting the thermostat at a different temperature;
- Commercial air conditioning (pre-cooling/pre-heating); and
- Commercial hot water.

In Figure 3.1, we show the maximum available capacity for flexibility of each of the sources of demand. The figures present the maximum available capacity in a given year, but in a given half-hour period, less may be available if there are seasonal or time-of-day restrictions on availability. We present these as NEM totals, but in practice we implement them on a state-wide basis.

Figure 3.1: Available Residential/Commercial Flexibility, Base and High Case (GW)



As the figure shows, the largest source of flexibility by available capacity is residential hot water, followed by residential air conditioning (cooling). Commercial buildings do not provide much flexibility by comparison.

Residential and commercial sources of load flexibility vary in their technical characteristics (how they can operate, when and for how long), and their cost characteristics. We set these out below:

Table 3.1: Technical characteristics – Residential/Commercial

		Response duration	Seasonal restriction	Time of day restriction	Price 1 ($/MWh)	Price 2 ($/MWh)	Price 3 ($/MWh)
Res. AC (Cooling)	Reducing		Summer only		$500	$5,000	$10,000
					20%	60%	20%
Res. AC (Heating)	Reducing		Winter only		$500	$5,000	$10,000
					20%	60%	20%
Res. Pool Pumps	Shifting	8 hours (summer) 2 hours (winter)			$0	$300	
					25%	75%	
Res. Hot Water	Shifting	8 hours			$0	$500	
					25%	75%	
Comm. AC (Set Point)	Reducing		Winter = 80% of Summer	8 am - 11 pm	$500	$1,000	
					50%	50%	
Comm. AC (Pre-control)	Shifting	1 hour	Winter = 80% of Summer	8 am - 11 pm	$500	$1,000	
					50%	50%	
Comm. Hot Water	Shifting	8 hours			$0	$500	$1,000
					25%	37.5%	37.5%

3.1.2. Electric vehicles

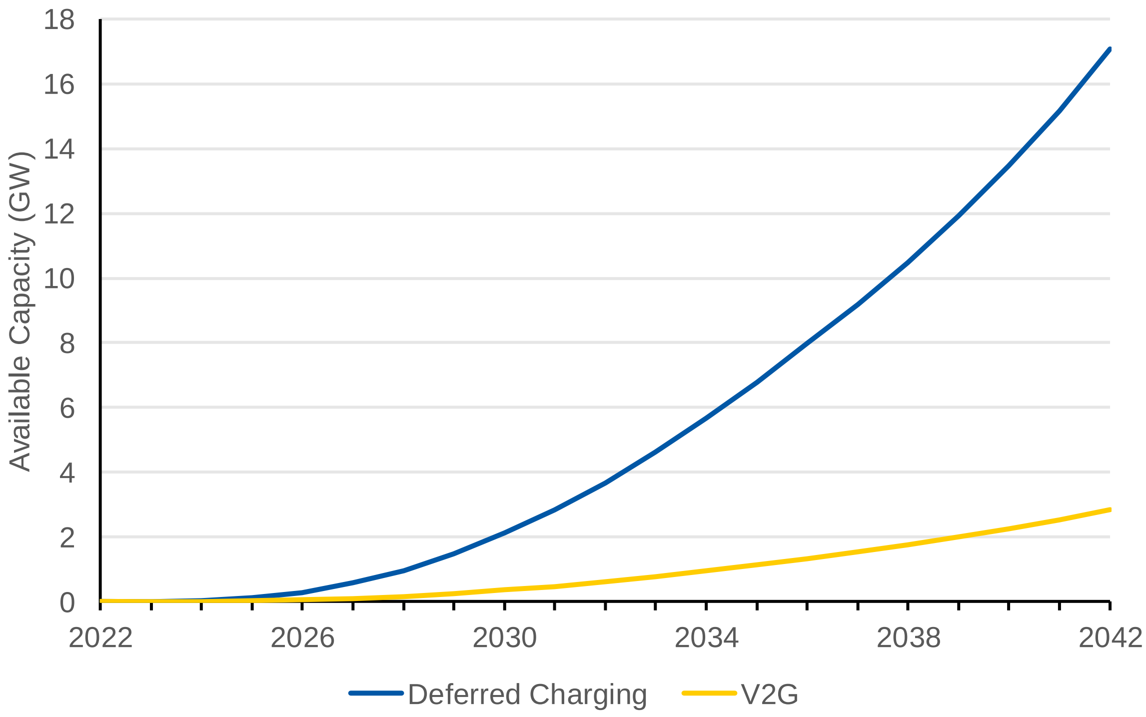
In this section, we describe our approach to modelling EVs, which could provide flexibility services either through deferred charging or through vehicle to grid (V2G). A high uptake of EVs would represent new demand on the system.

We assume that there will be around 14 million EVs in the NEM region by 2042, contributing around 18 GW of peak demand. We assume that EV flexibility is only available between 6 pm and 8 am, when EVs are likely to be plugged in at home. We distinguish between two types of flexibility services, both of which act as load shifting objects in PLEXOS:

- **Deferred charging**, where an EV charges at a different time than it otherwise would. For example, if someone returns home from work at 6 pm and plugs in their vehicle, it ordinarily would charge immediately for 2-3 hours, though this is during a period of system peak demand. If enabled to provide such flexibility, it could instead charge later in the night when electricity is cheaper. We assume that deferred charging can be provided for 2 hours in a day, based on the amount of time it would ordinarily take to charge the vehicle. This can be shifted to any time before 8 am.
- **V2G**, where an EV uses its actual battery to discharge to the grid, to be recharged later. For example, if someone returns home from work at 6 pm and plugs in their vehicle, it could immediately discharge its remaining charge to be recharged to its original state (and eventually to full charge) later. It could continue to provide this service throughout the night. Note that the vehicle in this example is simultaneously also deferring its charging, so it can provide both sources of flexibility in our model. We assume that V2G can be provided for 2 hours at a time.

We assume that 90 per cent of EVs can provide deferred charging by 2042, and 15 per cent of EVs can provide V2G (all of which also provide deferred charging). We assume that 0 per cent can provide either service at the beginning of the modelling period, and that the percentage increases linearly to 2042. Therefore, as shown in Figure 3.2, the available capacity grows increasingly quickly towards the end of the period as (a) more people have EVs; and (b) a greater percentage of them are available to provide flexibility.

Figure 3.2: EV Available Capacity (GW)



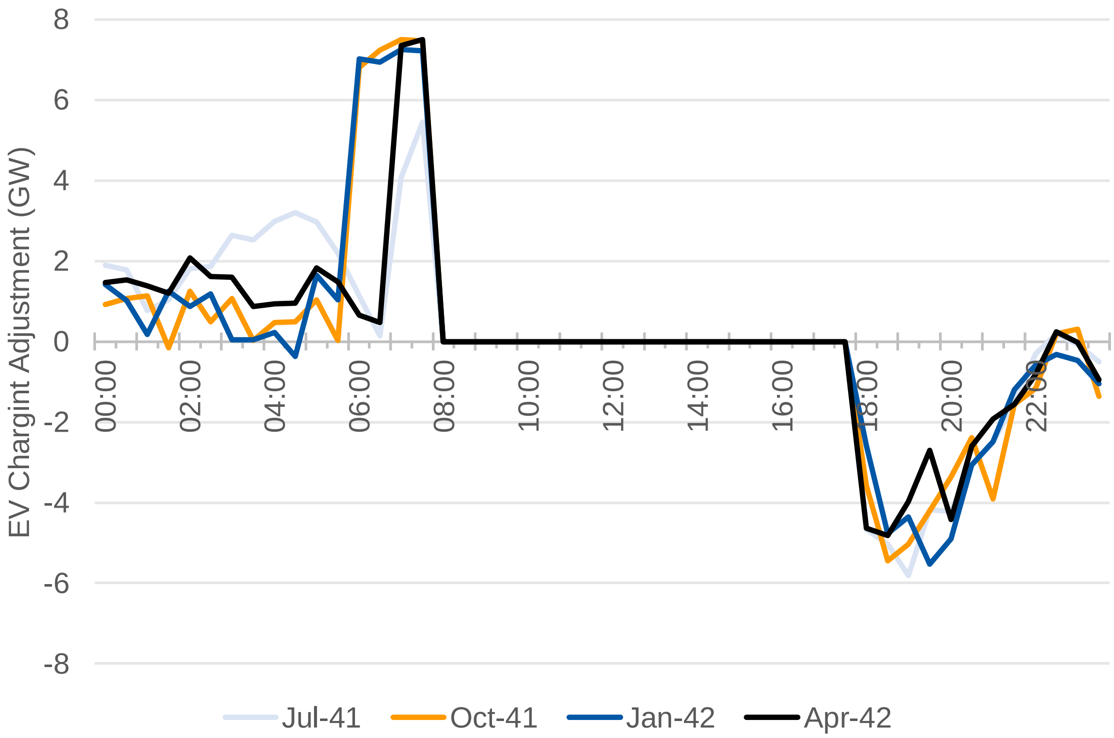
For both Deferred Charging and V2G, we assume that there is no cost to operate, save for the cost of inefficiency in charging (95 per cent) and discharging (95 per cent). If the price spread between the period of charging and the period of discharging is large enough to cover the c. 10 per cent efficiency losses, EVs will participate.

In our initial modelling work, we have found that PLEXOS is unable to fully optimise the behaviour of EVs while ensuring that generation is sufficient to meet end user demand. In short, if we allow EVs to act flexibly as a load shifting object, PLEXOS does not recognise the need to generate the electricity to "charge" the EVs in the first place.

To correct for this, we have derived optimised EV charging and discharging patterns from an initial set of runs where the raw generation output is insufficient to meet final demand. Whilst the final modelling results cannot be used to derive the value of flexibility, they nonetheless can be used to model optimal EV charging and discharging behaviour.

In Figure 3.3 below, we show the average monthly charging/discharging pattern for four months in New South Wales in 2041/42. This trace represents an estimate of how EVs could act when allowed to participate in demand response, as they do in the "original" model run from which the trace derives. We can see from the chart that throughout the year, it is optimal for EVs to discharge (or defer their charging) in the evening towards the time the sun sets, and charge in the early morning after sunrise (and PV comes online), but before the time we assume that customers have to leave the house. This trend is more pronounced in the summer months when the sunrise is earlier.

Figure 3.3: Monthly Average Half-Hourly Charging/Discharging of EVs (GW)



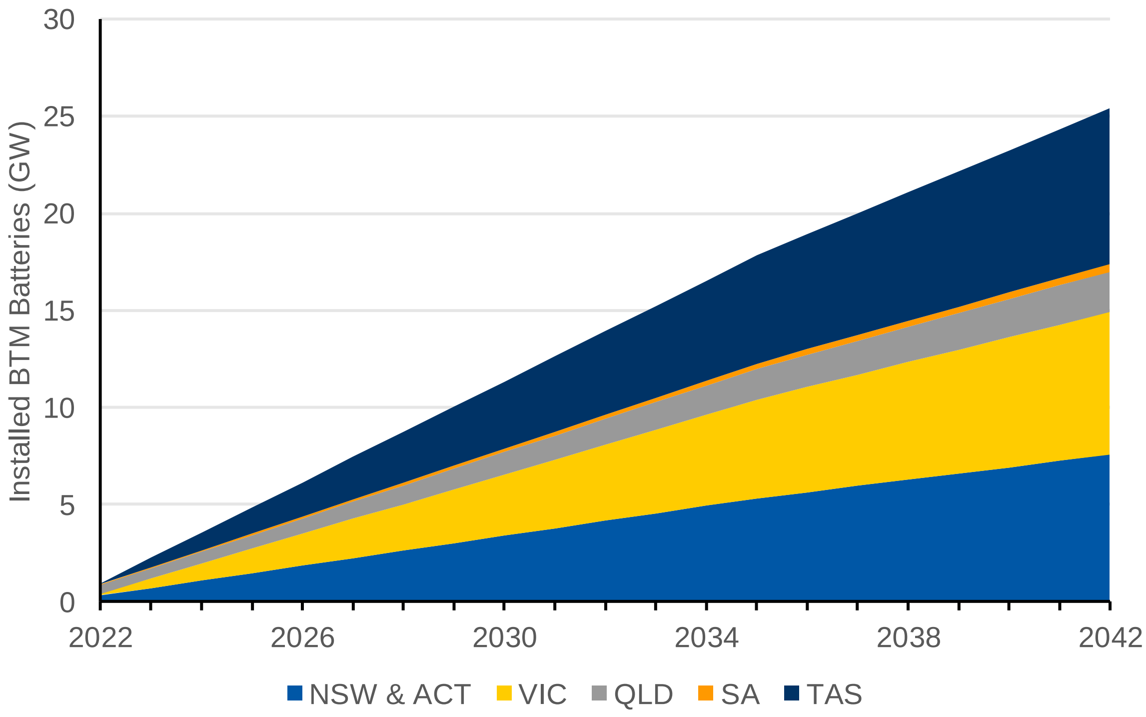
Note: Charging is positive (demand increases) and discharging is negative (demand decreases).

In the Reform scenario, we shift the EV charging trace by this amount. In other words, total system demand in that case is lower in the evening, slightly higher in the middle of the night, and considerably higher in the earliest daytime hours.

3.1.3. BTM batteries

Next, we consider the potential for BTM batteries, which can be called upon to provide load shifting services. In Figure 3.4 below, we show our assumed level of uptake of BTM batteries.

Figure 3.4: Installed BTM Battery Capacity – Energy Synapse High Case



BTM batteries act as storage units in PLEXOS, with a response duration of 2 hours with no other restrictions on use in aggregate. We assume that BTM batteries have no VOM and will operate anytime the price differential exceeds their storage losses.

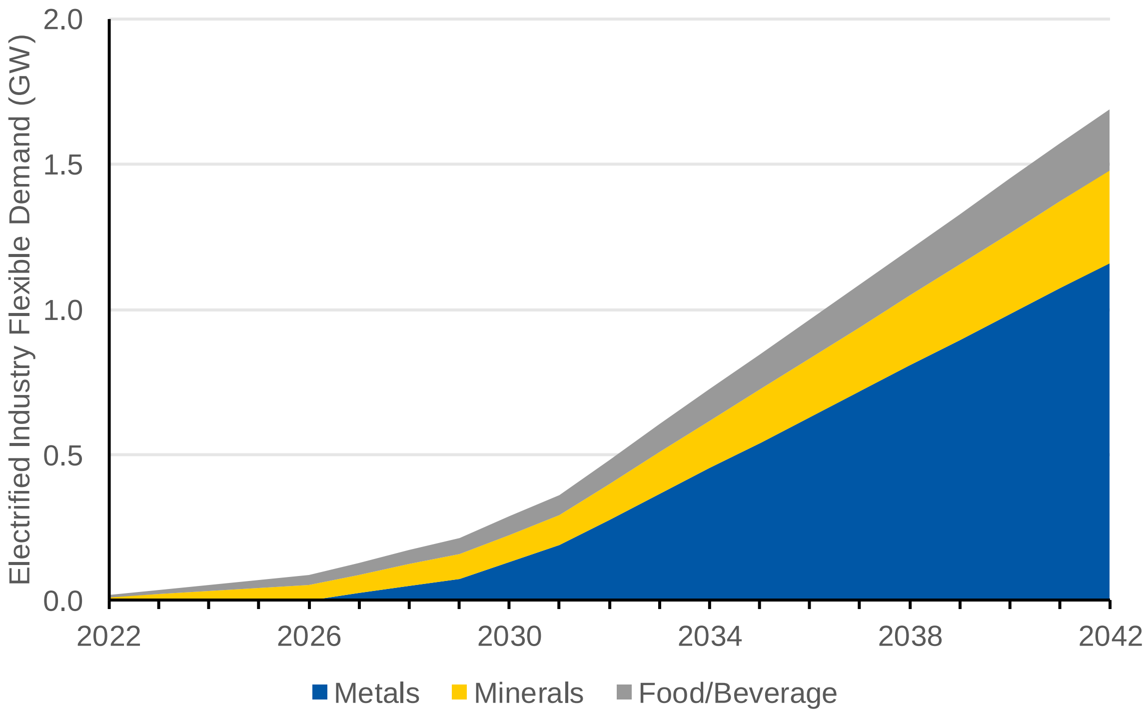
3.1.4. Electrified industry

Industrial facilities which currently rely on steam for process heat may electrify, providing both an additional load source and a source of load flexibility. We consider separately metals processing, minerals processing and food/beverage production.

We assume that up to 3 GW of industrial processes could be electrified by 2042, but not all of this is available to act flexibly, because that would require industrial processes to stop operating completely when called upon, which could be damaging to the commercial or mechanical operations of an industrial facility.

In Figure 3.5, we present the amount of load that is available to act flexibly, around half as much is electrified.

Figure 3.5: Electrified Industrial Flexible Demand



We assume that metals and minerals processing are load reducing objects while food/beverage production is a load shift object. In the case of food/beverage production, we assume that it is possible (but not necessary) for facilities to install boiler equipment that could make them indifferent to fluctuations in energy flows, and would then provide flexibility free of charge.

We summarise our technical assumptions below:

Table 3.2: Electrified Industry Technical Assumptions

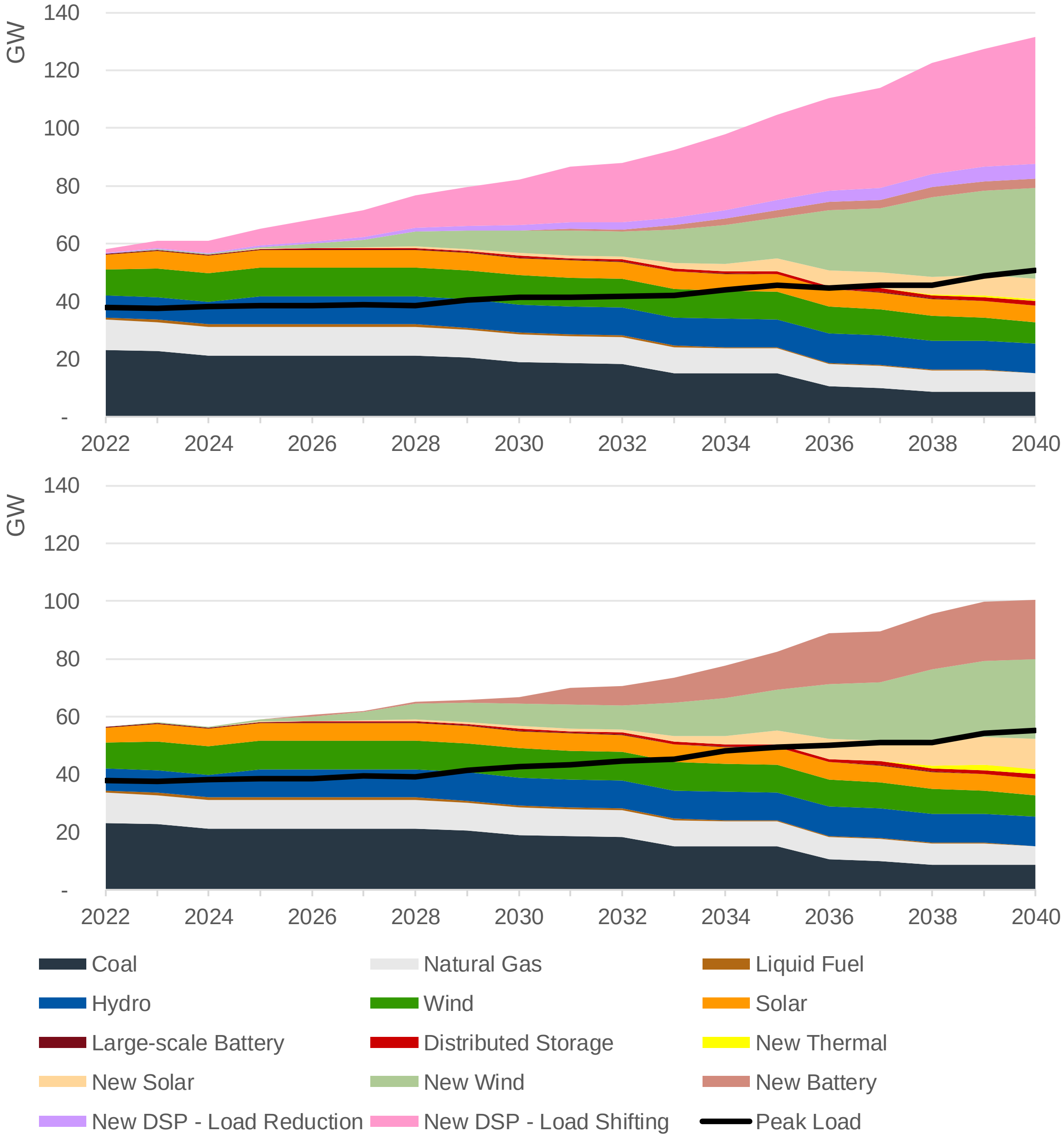
		Response duration	Price 1 ($/MWh)	Price 2 ($/MWh)	Price 3 ($/MWh)	Price 4 ($/MWh)	Price 5 ($/MWh)
Metals	Reducing	4 hours	$250	$1,500	$2,500	$5,000	$13,000
			18%	25%	25%	25%	7%
Non-metals	Reducing	4 hours	$1,000	$5,000	$10,000	$10,000	$10,000
			100%				
Food/Beverage	Shifting	4 hours	$0	$500			
			25%	75%			

3.2. Modelling Results

Using the above assumptions of enabled load flexibility in the Reform scenario, we compare with a modelling run where such flexibility is not available. We present results through 2040, omitting the final two years to exclude "end-of-period" effects in PLEXOS.

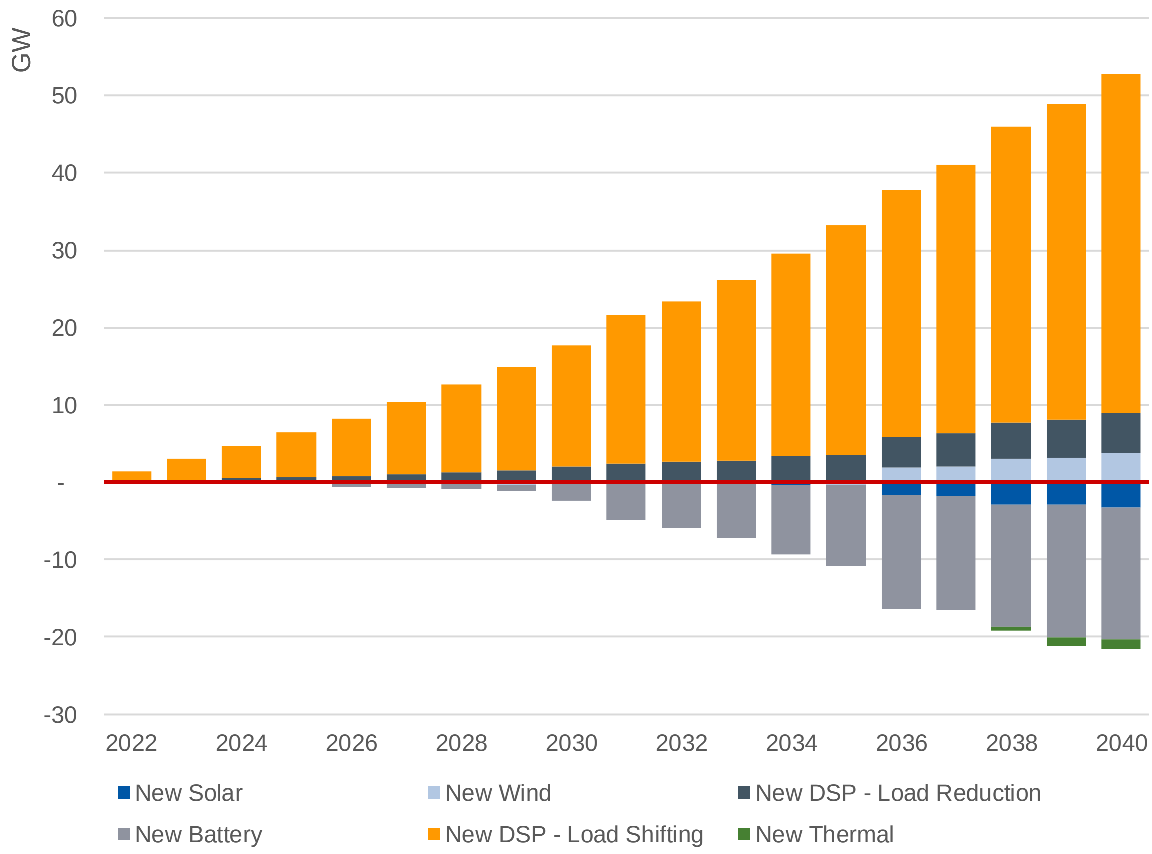
We show the capacity mix in each run in Figure 3.6 below, and the difference between the runs in Figure 3.7. As the figures show, the Reform scenario primarily allows the system to avoid investing in new storage capacity. By the end of the modelling period, the Reform scenario avoids the need for roughly 20 GW of storage capacity, as well as 1 GW of new thermal capacity.

Figure 3.6: Capacity Mix by Year, Reform (Top) v. No Reform (Bottom), GW



Source: NERA analysis of PLEXOS outputs

Figure 3.7: Difference in Installed Capacity by Year, Reform minus No Reform (GW)

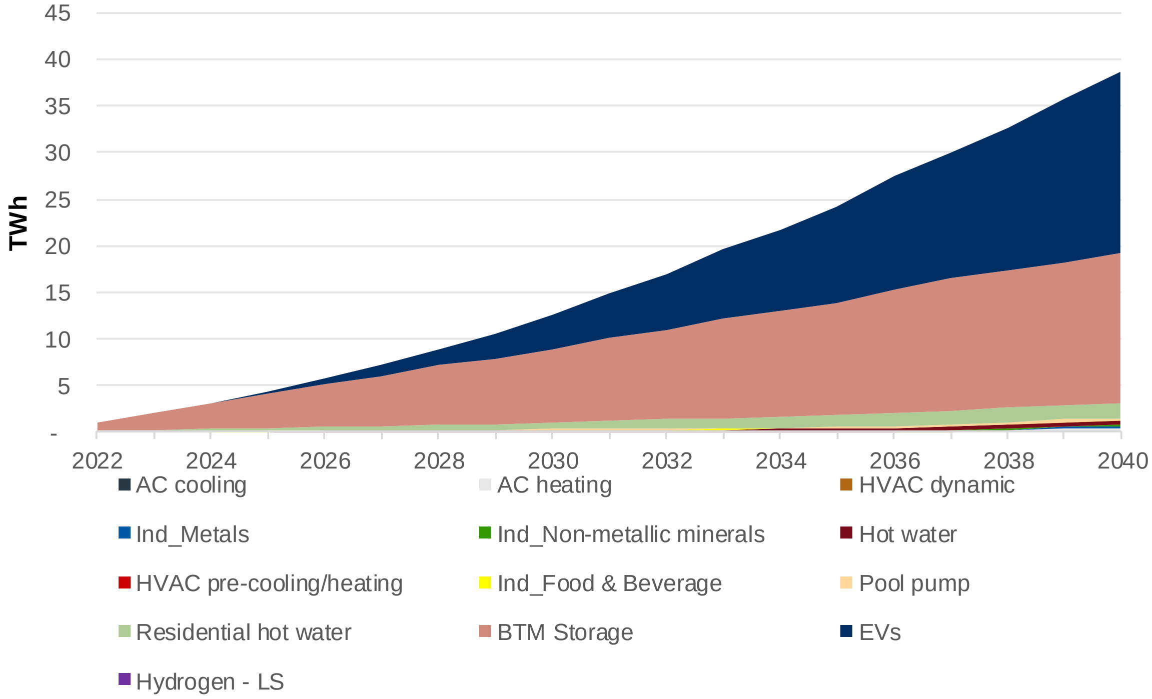


Source: NERA analysis of PLEXOS outputs

In addition to the differences in capacities, determined by the long-term expansion modelling, the capacity that does exist acts differently.

In Figure 3.8 below, we show how load flexibility services are called upon in the Reform scenario. As the figure shows, this is primarily provided by load shifting sources like BTM storage and EVs, as load reducing services tend to be more expensive and restrictive in when or how often they can operate.

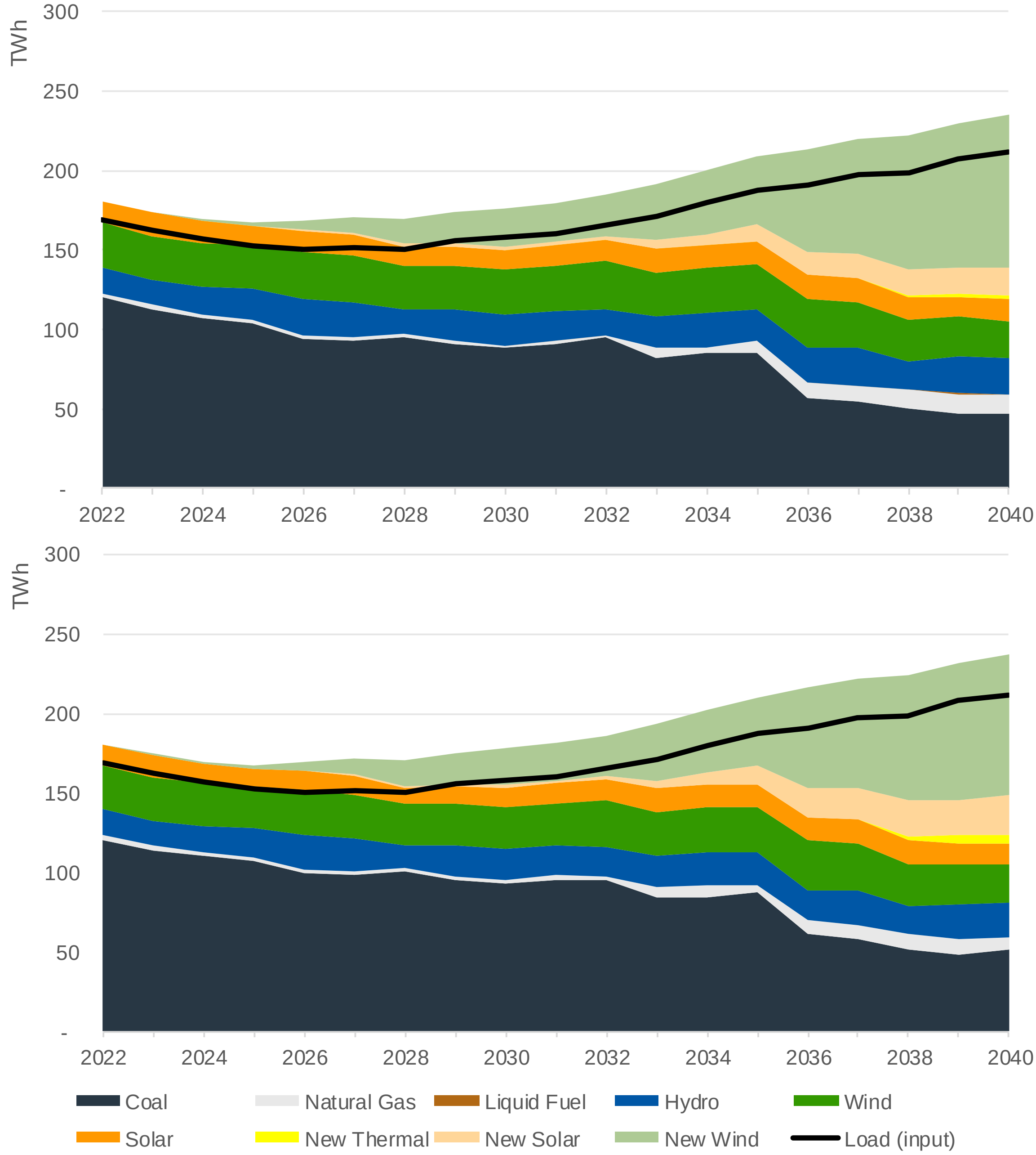
Figure 3.8: Dispatch of Flexibility Sources



Source: NERA analysis of PLEXOS outputs

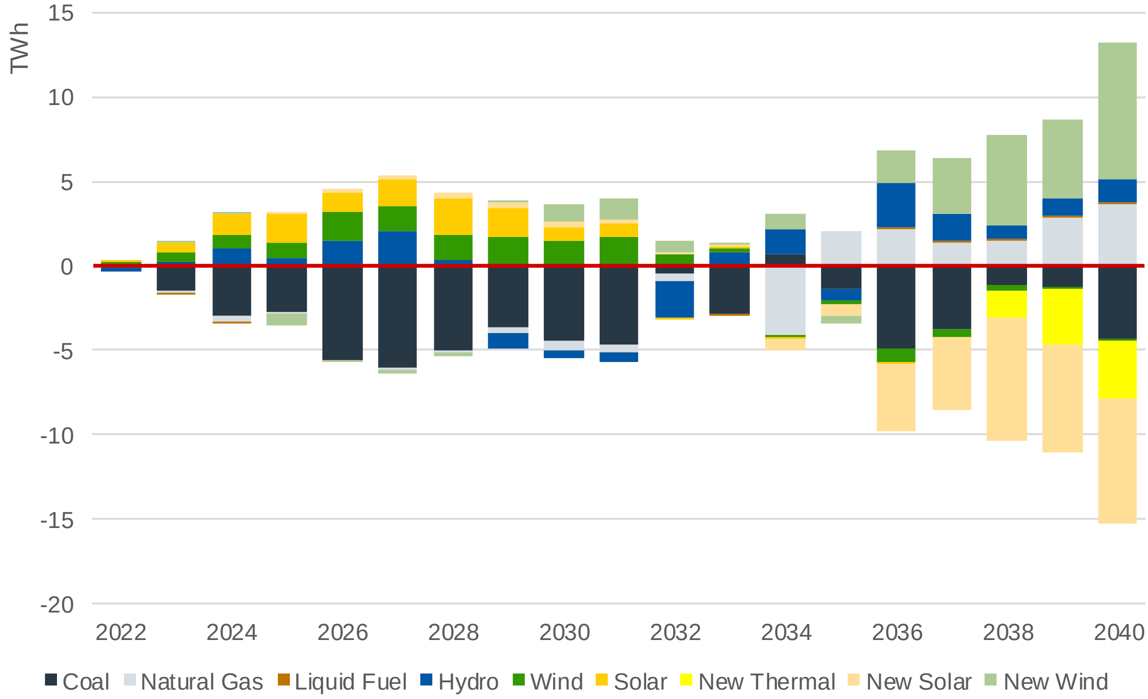
In Figure 3.9 and Figure 3.10 below, we show how output from generators differs between the two runs. As the figures show, the total amount of electricity generated is roughly similar in the two cases. However, the make up of that electricity generation is different, as the presence of load flexibility allows for greater use of wind power and gas generation, relative to solar power and coal.

Figure 3.9: Generation from Generator Sources, Reform (Top) v. No Reform (Bottom) (TWh)



Source: NERA analysis of PLEXOS outputs

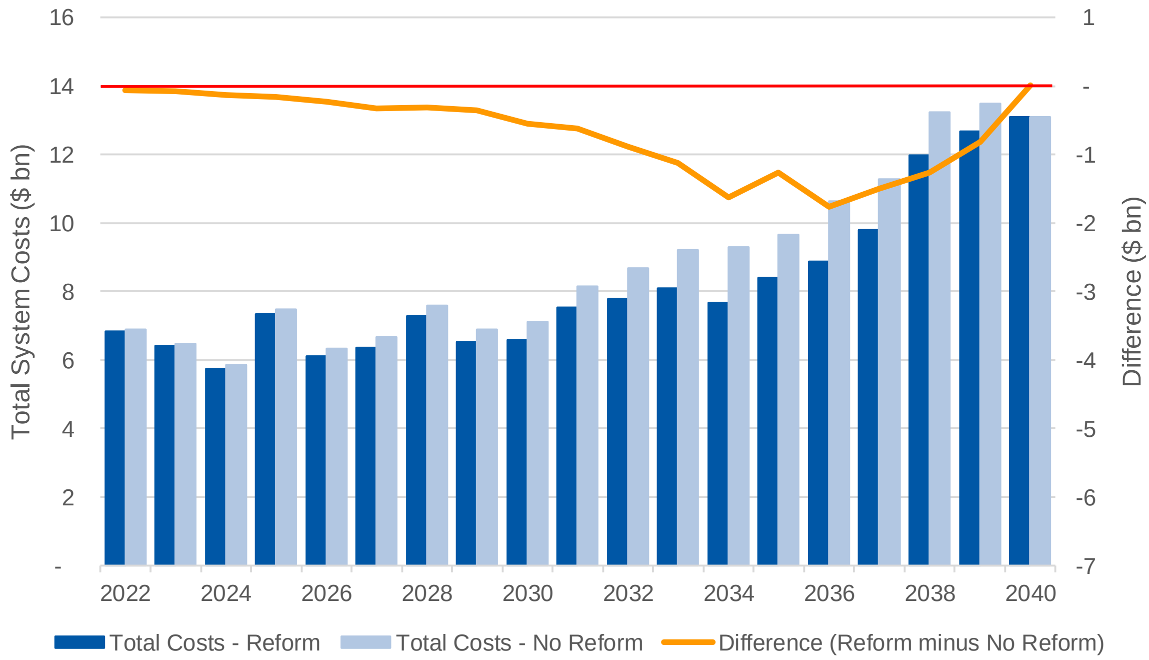
Figure 3.10: Difference in Generation from Generator Sources, Reform minus No Reform (TWh)



Source: NERA analysis of PLEXOS outputs

Finally, we quantify the savings resulting from the changes in capacity expansion and dispatch of resources. In Figure 3.11 below, we show the total NEM system costs under the Reform and the No Reform scenarios, and the difference between them by year. Table 3.3 breaks down the costs by category and shows undiscounted and discounted savings. Note that these do not include transmission costs, which are outside of the scope of our PLEXOS representation of the NEM.

Figure 3.11: System Costs Comparison by Year ($million)



Source: NERA analysis of PLEXOS outputs

Table 3.3: Break-down of Cost Difference, Reform minus No Reform ($million, NPV discount rate 5.9%)

Cost Category	Cost	2021-28	2028-35	2035-40	Total	NPV
Thermal	Fuel Cost	- 692	- 830	578	**- 945**	**- 774**
Thermal	VO&M	- 207	- 182	- 45	**- 434**	**- 282**
Thermal	FO&M	- 0	- 2	- 36	**- 38**	**- 14**
Thermal	Annualised Build Cost	-	-	- 403	**- 403**	**- 142**
Renewables	VO&M	49	9	117	**175**	**85**
Renewables	FO&M	- 26	19	336	**330**	**116**
Renewables	Annualised Build Cost	- 93	39	902	**849**	**288**
Storage	VO&M	-	-	-	**-**	**-**
Storage	FO&M	- 30	- 499	- 1,037	**- 1,566**	**- 663**
Storage	Annualised Build Cost	- 284	- 4,863	- 10,453	**- 15,600**	**- 6,579**
DSP - LR	VO&M	- 4	- 99	4,734	**4,631**	**1,629**
DSP - LS	VO&M	-	-	-	**-**	**-**
Total	**Total**	**- 1,287**	**- 6,408**	**- 5,308**	**- 13,003**	**- 6,337**

Source: NERA analysis of PLEXOS outputs

As the table shows, we find that the Reform scenario could save $6,337 million in NPV terms to 2040 relative to the No Reform scenario, and $13,003 million in undiscounted terms. This is primarily due to capacity savings on large scale storage units, with some additional savings delivered through reduced fuel costs (representing the avoided use of thermal generation due to flexibility) and additional expenses in the No Reform scenario from resorting to the most expensive demand response technologies in times of peak.

3.3. Conclusions on Integration of DER and Flexible Demand

The integration of DER and flexible demand can reduce system costs by circumventing the need to invest in new capacity that would otherwise serve a similar purpose. In our No Reform scenario, as well as in AEMO's ISP modelling, we model large investments in battery storage capacity, especially in the 2030s as existing thermal capacity begins to retire. This capacity investment is avoided through the availability of load shifting DSP, primarily BTM storage and EVs.

In principle, DSP could also reduce dispatch costs by limiting the system's reliance on expensive generation. In practice, however, the system in the absence of DSP does not rely heavily on thermal generation to meet peak demand. It instead discharges battery storage, which is costless to operate except for the cost of charging it (i.e. when it is sunny or windy) and any efficiency losses it incurs. Therefore, because we are primarily substituting one load-shifting resource (large scale batteries) for another (distributed batteries), we do not see substantial reductions in dispatch costs.

In total, we estimate that DER and flexible demand could reduce system costs by $6,337 million in NPV terms to 2040 and $13,003 million in undiscounted terms.

4. Resource Adequacy Mechanisms

As the NEM rapidly decarbonizes and increases the penetration of renewable energy on the grid, there is a risk that such rapid changes will result in insufficient reliability on the system. Accordingly, one of the ESB's pathways to reform is the introduction of RAMs, which signal and incentivise investment in reliable capacity.

The ESB has not set out the precise design features of a RAM, but any design would need to signal the value of providing capacity at a time of system shortage. In other jurisdictions, this is often represented by a capacity remuneration scheme or a reliability option.

In practice, we represent this in our PLEXOS model as a difference in the Value of Lost Load (VoLL). If the model sees a high VoLL, it will prioritise building more capacity to ensure that it does not often get hit. If the model sees a lower VoLL, it may determine that system costs can be minimised by allowing for some unserved energy (USE) in place of building new capacity. If the VoLL that the model sees is lower than the true value of lost load, then the model will tend not to minimise system costs.

In short, in the presence of a properly calibrated RAM, we would expect to see an additional cost of investment relative to the absence of one, but this would be more than offset by savings in short term costs like the value of unserved energy and dispatch of the Reliability and Emergency Reserve Trader (RERT).

Additionally, the price signals provided by a RAM may be necessary to ensure that existing thermal capacity stays connected to the end of its economic life. We therefore include in our analysis a sensitivity where, in the absence of a RAM, three large coal plants close one year earlier than anticipated.

This chapter proceeds as follows:

- In Section 4.1, we describe and justify our modelling approach in greater detail;
- In Section 4.2, we present our modelling results; and
- In Section 4.3, we conclude.

4.1. Description of Modelling

4.1.1. VoLL and Unserved Energy

The primary difference between our "RAM" and "No RAM" runs is in the treatment of the VoLL seen by PLEXOS and used for optimisation:

- In the RAM scenario, we use a VoLL of $21,247/MWh, equal to the VoLL that AEMO uses in its ISP modelling. Because the ISP is calibrated to identify the optimal investment pattern, we treat this value as the true VoLL. If a RAM is correctly calibrated to send an efficient signal that represents the true value of reliability, then this will be equivalently achieved if PLEXOS optimises the trade-off between the true value of lost load against the cost of investing in new capacity to avoid it. Our approach therefore represents a good proxy for the existence of a RAM, regardless of how it is designed in practice.

- In the No RAM scenario, we use a VoLL of $7,500/MWh, considerably below the true VoLL of $21,247/MWh. This represents a world where the signals to invest in reliable capacity are inefficiently low, for three reasons:
  - *Explicit price caps*. The electricity price in the NEM is currently capped at $15,000/MWh. With no additional source of remuneration in the NEM, an investor in new capacity could not expect to ever receive a higher price higher than this.
  - *Implicit price caps*. Even with an explicit price cap of $15,000/MWh, an investor may be skeptical that prices could reach that level repeatedly without it becoming politically untenable.
  - *Capital market imperfections*. Even if a new generator could credibly earn the market price cap frequently enough to earn back its investment cost, these occurrences may be infrequent and random enough that they do not deliver assurances to satisfy lenders and investors that an asset is bankable.

  To capture these three effects, we select a VoLL equal to half of the market price cap. As a result, we expect the No RAM scenario to deliver less investment in exchange for greater levels of unserved energy.

Because the VoLL in the No RAM scenario is inefficiently low, we must correct for it when quantifying the total system costs. If we did not, we might find that the No RAM scenario was preferable to the RAM scenario simply because the cost of losing load is lower, but of course the value to society remains the same. Therefore, in quantifying total system costs in the No RAM scenario (after PLEXOS optimises based on the lower VoLL), we value all unserved energy at the full VoLL of $21,247.

4.1.2. RERT

Before load is actually switched off in practice, AEMO will dispatch RERT, an emergency reserve priced at the market cap. AEMO contracts with providers of RERT in advance of an anticipated use (usually industrial sites), regardless of whether it actually calls upon it. These contracts typically apply only to a single anticipated need. For example, in advance of an anticipated need in New South Wales on 4 January 2020, AEMO contracted 368 MW of Short Notice Reserve, for a term of 5.5 hours.[8]

We rely on AEMO's historical contracting activity to represent the availability of RERT going forward. In particular, we have reviewed AEMO's annual reports since 2018/19 and selected the highest single level of contracted RERT by state. We assume that AEMO could contract this amount if needed in future:

- New South Wales: 520 MW
- South Australia: 227 MW
- Victoria: 596 MW

AEMO did not contract any RERT in Queensland and Tasmania in this period, but that does suggest that it could not have if it required any. On ESB's advice, we have assumed that AEMO could procure 500 MW of RERT in each Queensland and Tasmania.

8 AEMO (August 2020), Reliability and Emergency Reserve Trader (RERT) End of Financial Year 2019-20 Report, p.4.

In order to ensure that RERT dispatches in our PLEXOS model before load is disconnected, we set the price at $1 below the relevant VoLL, i.e. $21,246/MWh in the RAM scenario and $7,499/MWh in the RAM scenario. Similar to our valuing of unserved energy, we assume that the true value of RERT is $21,246/MWh in both the RAM and the No RAM scenario, and hence we use this value when calculating total system costs.

4.1.3. Retirement Schedule

The value that a RAM can deliver depends on whether there is any shortage of firm capacity. If there is a surplus of capacity, then investment decisions are unchanged by a RAM – the cheapest capacity is the one that already exists. Therefore, to fully capture the value of a RAM, we assume that existing coal plants retire at a rate in line with the ISP Step Change, faster than the rate found in the ISP Central Case.

Another benefit of a RAM is that it could prevent the early closure of existing thermal capacity, as the owners of these units would receive additional revenue for participating in the RAM. We evaluate the case where retirement happens in a disorderly fashion by running an additional No RAM scenario where three large coal plants retire a year earlier than anticipated:

- Vales Point, a 1,320 MW coal plant in New South Wales, retires from 2022 to 2024 instead of 2023 to 2025;
- Gladstone, a 1,680 MW coal plant in Queensland, retires from 2022 to 2024 instead of 2023 to 2025; and
- Yallourn, a 1,450 MW coal plant in Victoria, retires from 2024 to 2028 instead of from 2025 to 2029.

We only change this assumption in our short-term operational run, after PLEXOS has optimised the construction of new capacity. This reflects the risk that, in the absence of a RAM, existing thermal capacity could close when it has not been planned for. In these periods of insufficient capacity, we would expect to see higher short-term costs through unserved energy and RERT dispatch.

Therefore, our core RAM modelling runs comprise three rather than two runs:

- A RAM scenario based on a $21,247/MWh VoLL. We refer to this as the “$21k” scenario.
- A No RAM scenario based on a $7,500/MWh VoLL. We refer to this as the “$7.5k” scenario.
- A No RAM scenario based on a $7,500/MWh VoLL and the unexpected early retirement of the three coal units described above. We refer to this as the “$7.5k + Early Retirement” scenario.

In Figure 4.1 below, we present the amount of available existing coal capacity by year under the two specifications above (i.e. with and without early retirement).

Figure 4.1: Coal Retirement Schedules

Source: NERA analysis of PLEXOS outputs

4.2. Modelling Results

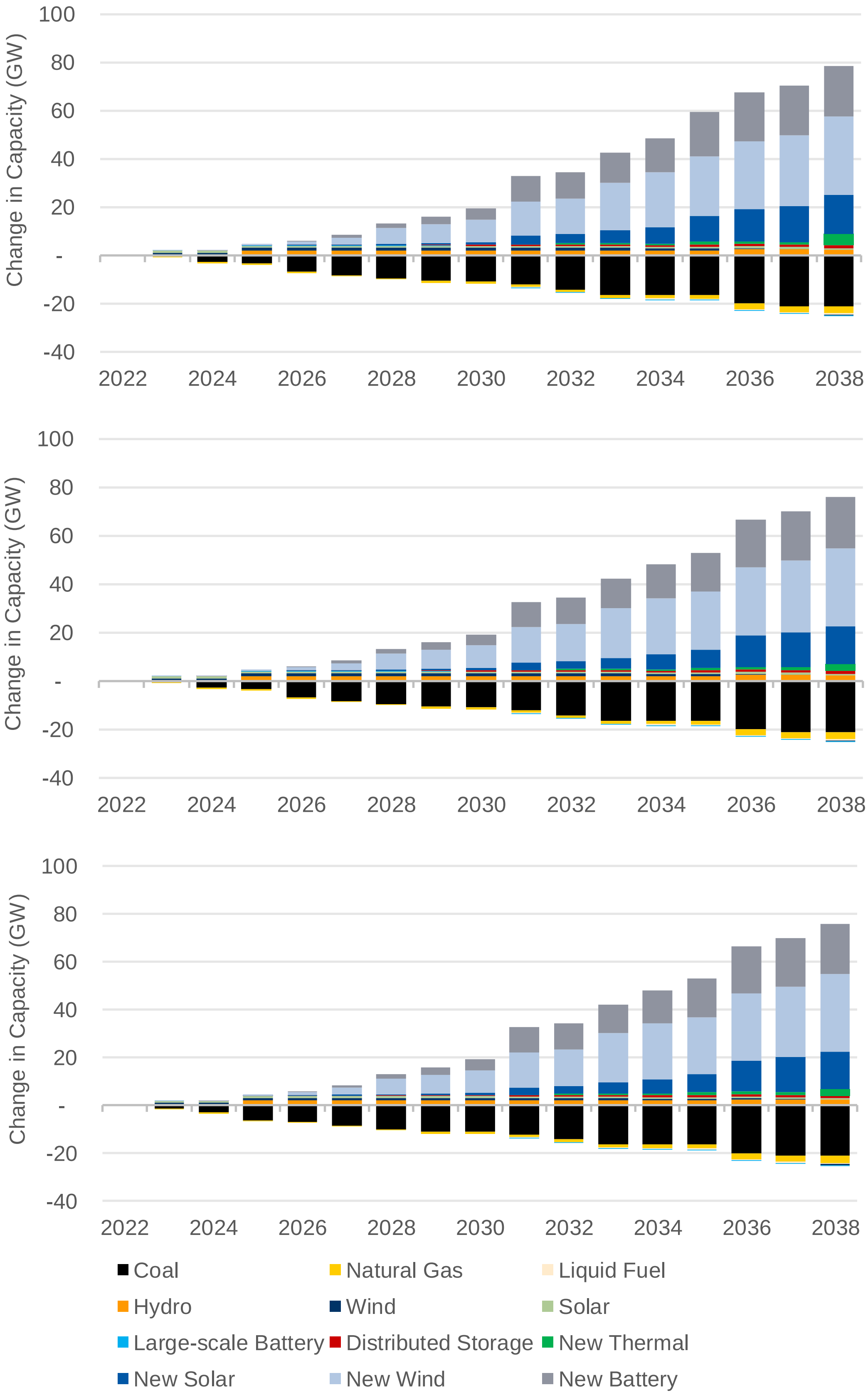
In this section, we present the findings of our modelling runs, comparing the $21k scenario (i.e. with a RAM in place) with both the $7.5k scenario and the $7.5k + Early Retirement scenario, thus separately identifying the value of a RAM from (i) signaling efficient investment in new capacity; and (ii) maintaining existing thermal capacity to its planned retirement date.

We present results through 2038 to exclude "end-of-period" effects in PLEXOS. We truncate our results by more in this task than in our modelling of DER and flexible demand because much of the value of a RAM is driven by savings in USE and RERT, and these results are particularly sensitive to end-of-period effects.

4.2.1. Differences in Capacity Mix

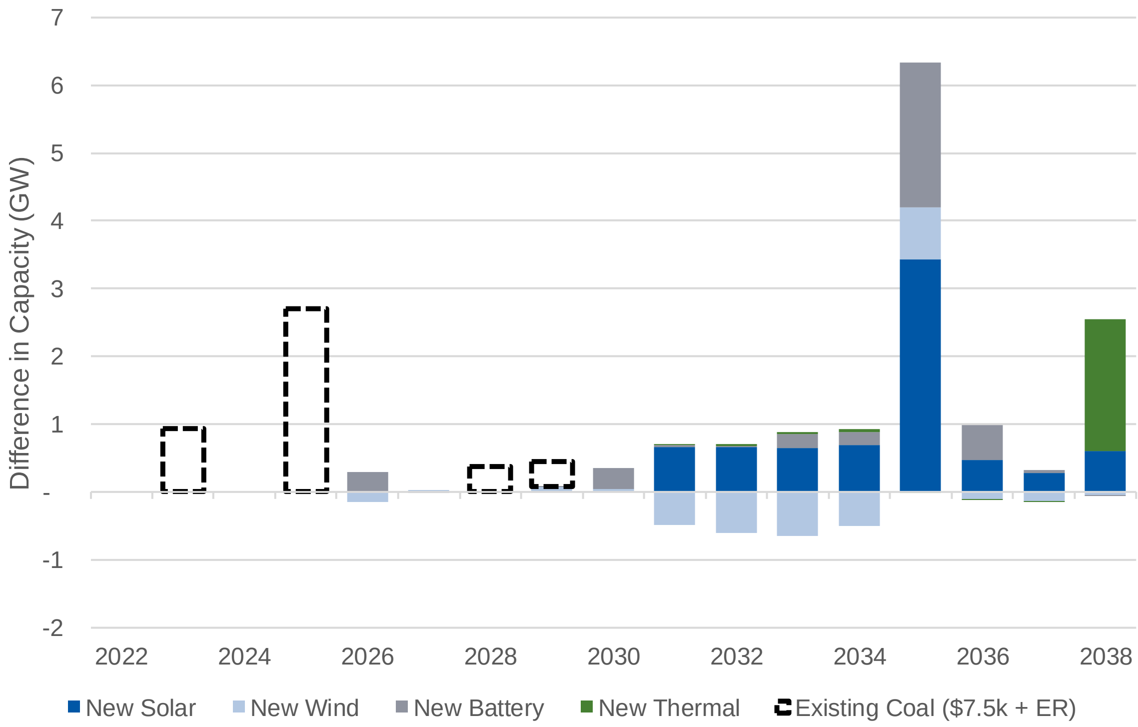
In Figure 4.2, we show the change in capacity from 2022 levels in each of the three scenarios. As the figures show, we see substantial new investment in solar, wind and battery storage, replacing coal capacity, irrespective of the scenario.

Figure 4.2: Capacity Relative to 2022 – 21k (top), 7.5k (middle), 7.5k + ER (bottom)



In Figure 4.3, we show the difference in the total capacity mix between the $21k scenario and the $7.5k scenarios. The RAMs scenario allows for greater commissioning of new solar capacity, often backed with storage, and less wind capacity, while bringing forward investment in new thermal capacity from 2039 to 2038. In comparison with the $7.5k + Early Retirement scenario, it also allows for existing coal to stay on the system for longer.

Figure 4.3: Difference in Capacity, $21k - $7.5k (+ ER)

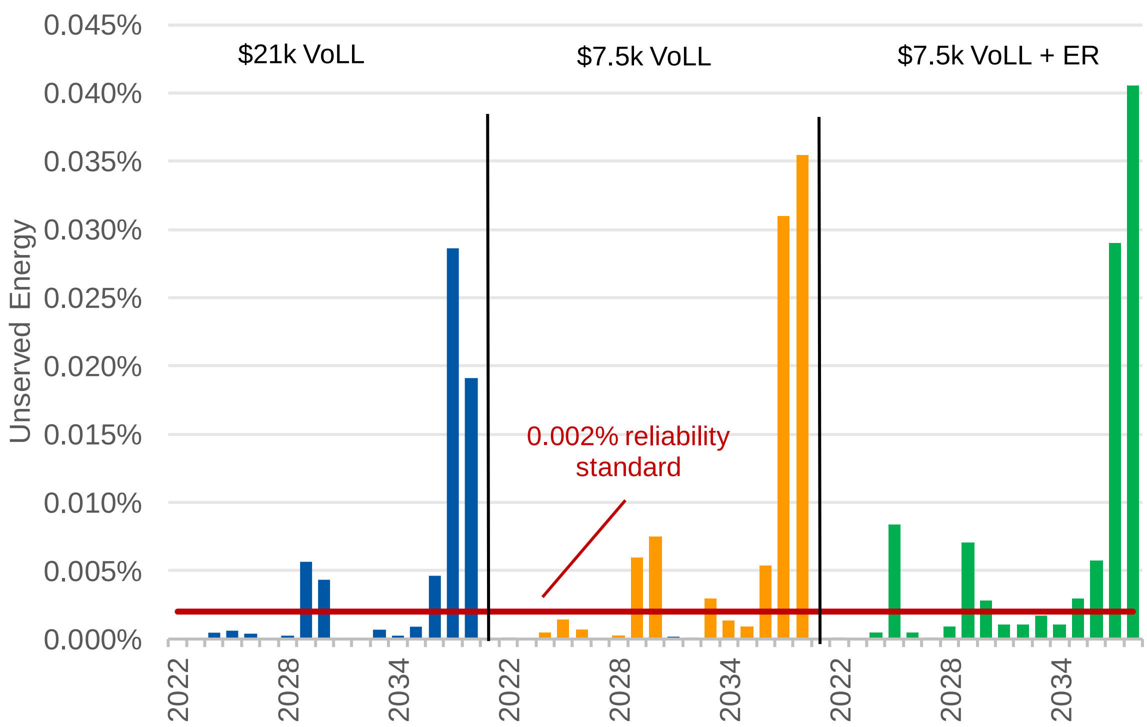


Source: NERA analysis of PLEXOS outputs

4.2.2. Differences in USE

While the existence of a RAM will tend to increase investment costs (because it encourages more investment), this will be more than offset by savings in the avoided USE. In Figure 4.4 below, we demonstrate that USE is generally lower in the $21k scenario than in the $7.5k and, especially, the $7.5k + Early Retirement scenario.

Figure 4.4: Unserved Energy



Source: NERA analysis of PLEXOS outputs

In all cases, even with a RAM in place, we model excessively high levels of USE by the final few years of the modelling period. This is primarily a function of how PLEXOS sequentially optimises the system, rather than an outcome we would expect in reality. When choosing which units to build in the "long-term" run, PLEXOS compares a small sample of demand forecasts (e.g. 1 in every 8 hours) with general characteristics of each candidate technology. When choosing which units to dispatch in the "short-term" run, PLEXOS optimises dispatch over all periods, while also creating random plant outages.

Additionally, when planning and building to meet the reliability standard, PLEXOS gives full credit to the reliability of batteries and hence chooses to build lots of them. However, batteries are limited in duration before they need to charge, meaning they will not be available to meet high demand if it lasts longer than the storage duration of a battery.

This is an area that requires further investigation which we were unable to conduct in the time available to us. We understand that the ESB will consider these specific modelling outcomes more fully during the detailed design phase of the RAM.

4.2.3. Total System Costs

In the previous two subsections, we describe and quantify the two primary drivers of value that a RAM could cause: an increase in overall capacity (and fixed O&M) costs more than offset by a reduction in outages or near outages (i.e. dispatch of RERT). There are other drivers of value, e.g. different dispatch patterns driven by different fuel sources, but these can be thought of as side effects of the core tradeoff between the cost of an outage and the cost of building to avoid it.

In this section, we quantify the change in overall costs that results from the introduction of a RAM. In all runs, we quantify the cost of USE and RERT based on the true VoLL ($21,247/MWh) or $1 below for RERT rather than necessarily the value that the model sees when optimising the system.

We show the savings in our core runs in Table 4.1 below. This represents the savings of a RAM due to the incentive to invest in new capacity, i.e. without considering the risk of early retirement. As the table shows, there is an increase in total capacity costs of around $700 million in NPV terms (with a WACC of 5.9 per cent). This is more than offset by reductions in USE and RERT (as well as expensive demand-side participation, to a much lesser extent), as well by a reduction in fuel costs. The net effect is a savings of $1,130 million in NPV terms, and $2,638 million in undiscounted terms.

Table 4.1: Total Cost Savings: $21k - $7.5k

Cost Category	Cost	2021-28	2028-35	2035-38	Total	Total - NPV
Thermal	Fuel Cost	- 13	- 96	- 60	**- 170**	**- 92**
Thermal	VO&M	16	- 16	- 6	**- 5**	**3**
Thermal	FO&M	- 4	- 0	6	**1**	**- 1**
Thermal	Annualised Build Cost	-	15	67	**82**	**33**
Renewables	VO&M	- 3	- 24	- 2	**- 29**	**- 15**
Renewables	FO&M	- 3	- 45	0	**- 48**	**- 24**
Renewables	Annualised Build Cost	- 17	- 118	11	**- 123**	**- 65**
Storage	VO&M	-	-	-	**-**	**-**
Storage	FO&M	5	28	11	**44**	**23**
Storage	Annualised Build Cost	47	445	247	**740**	**363**
System	USE, DSP, RERT	- 141	- 593	- 2,394	**- 3,129**	**- 1,354**
Total	**Total**	**- 113**	**- 404**	**- 2,120**	**- 2,638**	**- 1,130**

Source: NERA analysis of PLEXOS outputs

The savings from a RAM grow slightly if we assume that a RAM is necessary to prevent early retirement of coal plants. We show this in Table 4.2 below. While the additional build costs are the same (because investment in new capacity is the same), the avoided cost of reliability events is larger, especially in the early period when these early retirements happen. In this scenario, we find that a RAM would save $1,294 million in NPV terms to 2038 and $3,138 million in undiscounted terms.

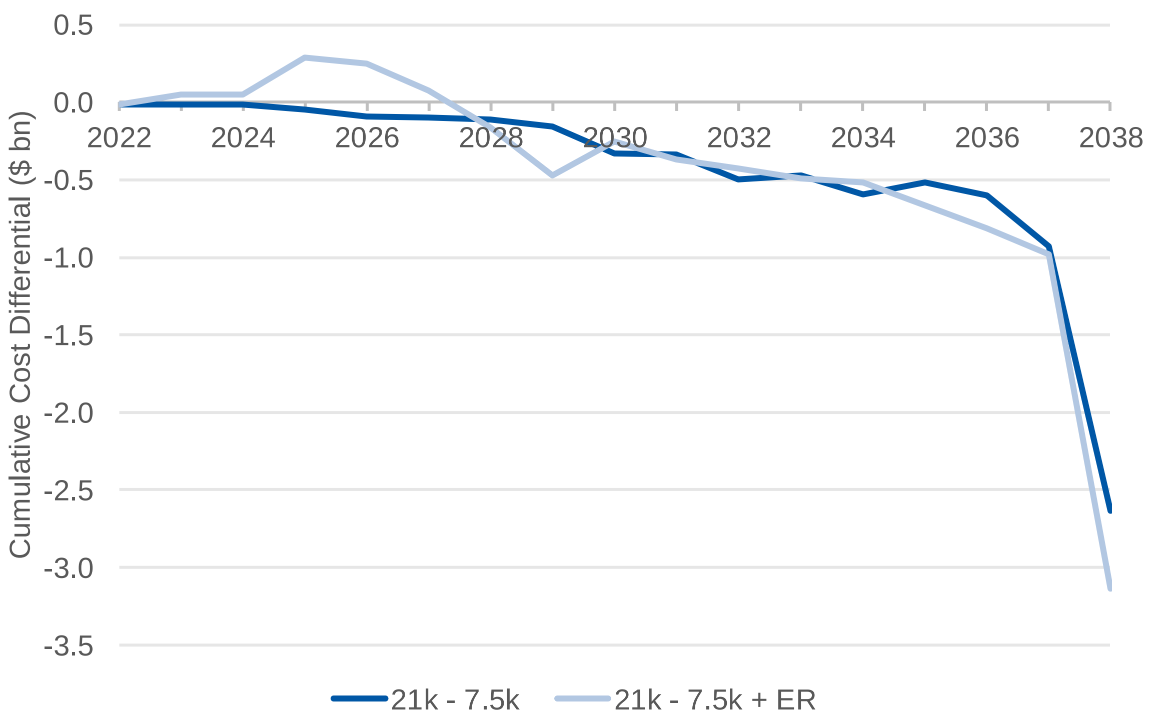
Table 4.2: Total Cost Savings: $21k - $7.5k + Early Retirement

Cost Category	Cost	2021-28	2028-35	2035-38	Total	NPV
Thermal	Fuel Cost	- 460	- 284	- 48	**- 793**	**- 520**
Thermal	VO&M	- 18	- 15	- 5	**- 37**	**- 23**
Thermal	FO&M	852	57	6	**915**	**714**
Thermal	Annualised Build Cost	-	15	67	**82**	**33**
Renewables	VO&M	- 7	- 21	- 4	**- 31**	**- 16**
Renewables	FO&M	- 3	- 45	0	**- 48**	**- 24**
Renewables	Annualised Build Cost	- 17	- 118	11	**- 123**	**- 65**
Storage	VO&M	-	-	-	**-**	**-**
Storage	FO&M	5	28	11	**44**	**23**
Storage	Annualised Build Cost	47	445	247	**740**	**363**
System	USE, DSP, RERT	- 563	- 567	- 2,757	**- 3,886**	**- 1,777**
Total	**Total**	**- 163**	**- 505**	**- 2,471**	**- 3,138**	**- 1,294**

Source: NERA analysis of PLEXOS outputs

We present the cumulative difference in cost on an annual basis in Figure 4.5 below. As the figure shows, savings are particularly large in 2038 (due to large amount of USE in all scenarios in that year), though due to the 5.9 per cent discount rate, these savings have less of an effect on NPV costs as they do on undiscounted costs.

Figure 4.5: Cost Difference by Year



Source: NERA analysis of PLEXOS outputs

4.3. Conclusions on Resource Adequacy Mechanisms

A RAM that is calibrated to signal the true value of reliability can provide value to the system by signaling investment in new capacity necessary to avoid loss of load. In the absence of a RAM, explicit and implicit price caps may mean that investors may not see the value in investing in capacity that is societally optimal.

In the context of the NEM, we conclude that the value of a RAM depends in large part on the assumptions around the anticipated and unanticipated retirement of existing thermal capacity. In an extreme case, where we expected to have a surplus of reliable existing capacity beyond the end of the modelling horizon, we would find no value of a RAM because it would not change investment decisions (no one would invest in new reliable capacity as it is not needed). In our modelling runs, we begin to see a need for new reliable capacity around 2030, at which point the value of the RAM starts to appear.

The value of a RAM increases if it is effective in maintaining existing plant to their anticipated closure date. The NEM could suffer high costs of outages and RERT dispatch in periods where a plant was expected to be available but it is not.

In summary, we estimate that a RAM could reduce total system costs by between $1,130 million and $1,294 million in NPV terms (with a 5.9 per cent discount rate), or between $2,638 and $3,138 billion in undiscounted terms, to 2038.

These results show the benefits of a RAM presuming that the market currently fails to deliver an efficient level of capacity (e.g. by assuming the presence of explicit and implicit price caps) and that the RAM is an efficient design. The benefits assessed do not include any estimate of the implementation costs of the RAM. In the absence of market failures in the current NEM design or given an inefficient design for the RAM, a RAM may not lead to social benefits. On the other hand, if the current distortions in the NEM design were larger than we assumed the benefits would be commensurately bigger.

5. Conclusion

In conclusion, we find that the integration of DER and flexible demand, and the introduction of a RAM, could reduce system costs to the NEM over the coming decades as renewable penetration rapidly expands:

- The integration of DER and flexible demand could reduce system costs by $6,337 million in NPV terms and $13,003 in undiscounted terms, to the end of the modelling period, driven primarily by the avoided need to build new storage capacity.
- The introduction of a RAM could reduce system costs by between $1,130 million and $1,294 million in NPV terms, or between $2,638 and $3,138 billion in undiscounted terms, to 2038. Additional capacity costs are more than offset by a reduction in dispatch costs and unserved energy.

To some extent, these benefits are overlapping and therefore cannot necessarily be added to one another. For instance, a RAM may encourage some of the same investment that is avoided by the existence of demand flexibility at times of system stress. Similarly, both interventions will tend to reduce the cost of shortages in the system.

On the other hand, neither conclusion cancels out the other: A RAM would likely reduce system costs with or without high integration of DER and flexible demand; and DER and flexible demand would likely reduce system costs with or without the existence of a RAM.

Qualifications, assumptions and limiting conditions

This report is for the exclusive use of the NERA Economic Consulting client named herein. This report is not intended for general circulation or publication, nor is it to be reproduced, quoted or distributed for any purpose without the prior written permission of NERA Economic Consulting. There are no third party beneficiaries with respect to this report, and NERA Economic Consulting does not accept any liability to any third party.

Information furnished by others, upon which all or portions of this report are based, is believed to be reliable but has not been independently verified, unless otherwise expressly indicated. Public information and industry and statistical data are from sources we deem to be reliable; however, we make no representation as to the accuracy or completeness of such information. The findings contained in this report may contain predictions based on current data and historical trends. Any such predictions are subject to inherent risks and uncertainties. NERA Economic Consulting accepts no responsibility for actual results or future events.

The opinions expressed in this report are valid only for the purpose stated herein and as of the date of this report. No obligation is assumed to revise this report to reflect changes, events or conditions, which occur subsequent to the date hereof.

All decisions in connection with the implementation or use of advice or recommendations contained in this report are the sole responsibility of the client. This report does not represent investment advice nor does it provide an opinion regarding the fairness of any transaction to any and all parties.

NERA Economic Consulting
One International Towers
100 Barangaroo Avenue
Sydney NSW, Australia 2000
www.nera.com