Secure Energy Transitions in the Power Sector

International Energy Agency

Electricity Security 2021

INTERNATIONAL ENERGY AGENCY

The IEA examines the full spectrum of energy issues including oil, gas and coal supply and demand, renewable energy technologies, electricity markets, energy efficiency, access to energy, demand side management and much more. Through its work, the IEA advocates policies that will enhance the reliability, affordability and sustainability of energy in its 30 member countries, 8 association countries and beyond.

IEA member countries:

Australia
Austria
Belgium
Canada
Czech Republic
Denmark
Estonia
Finland
France
Germany
Greece
Hungary
Ireland
Italy
Japan
Korea
Luxembourg
Mexico
Netherlands
New Zealand
Norway
Poland
Portugal
Slovak Republic
Spain
Sweden
Switzerland
Turkey
United Kingdom
United States

The European Commission also participates in the work of the IEA

IEA association countries:

Brazil
China
India
Indonesia
Morocco
Singapore
South Africa
Thailand

International Energy Agency
Website: www.iea.org

Abstract

Electricity is an integral part of all modern economies, supporting a range of critical services from healthcare to banking to transportation. Secure supply of electricity is thus of paramount importance. The structural change from an electricity system based on thermal generation powered by fossil fuels towards a system based on variable renewable energy continues apace at various stages across the globe. Digitalisation tools such as smart grids and distributed energy resources, along with the electrification of end uses put electricity increasingly at the forefront of the entire energy system. As a result, governments, industries and other stakeholders will need to improve their frameworks for ensuring electricity security through updated policies, regulations and market designs. This report details the new approaches that will be needed in electricity system planning, resource adequacy mechanisms, incentives for supply- and demand-side flexibility, short-term system balancing and stability procedures. It provides examples and case studies of these changes from power systems around the world, describes existing frameworks to value and provide electricity security, and distils best practices and recommendations for policy makers to apply as they adjust to the various trends underway.

Acknowledgements, contributors and credits

A cross-agency group of experts from the Directorate of Energy Markets and Security; the Directorate of Sustainability, Technology and Outlooks; and the Strategic Initiatives Office of the International Energy Agency (IEA) prepared this report. Edwin Haesen, former Head of the System Integration of Renewables (SIR) Unit, and César Alejandro Hernández Alva, Head of the Renewable Integration and Secure Electricity (RISE) Unit, led and co-ordinated the study. Keisuke Sadamori, Director of Energy Markets and Security, provided expert guidance and advice.

Keith Everhart, Craig Hart, Zoe Hungerford, Divya Reddy and Peerapat Vithayasrichareon were lead authors of this report. IEA colleagues Enrique Gutierrez Tavarez, Stefan Lorenczik and Gergely Molnar also contributed to the analysis. Anna Kalista provided essential support.

IEA colleagues including Dave Turk, Laszlo Varro, Paolo Frankl, Laura Cozzi, Peter Fraser, Tom Howes, Brian Motherway, Aad van Bohemen, Brent Wanner, Nicole Thomas, Randi Kristiansen, Sylvia Elisabeth Beyer, Edith Bayer, Michael Waldron, Kathleen Gaffney, Sara Moarif, Vanessa Koh and Clémence Lizé provided valuable input, comments and feedback.

Justin French-Brooks was the editor of this report. Thanks go to colleagues in the Communications and Digital Office (CDO) for their help in producing the report and website materials, particularly, Jad Mouawad, Head of CDO, and Tanya Dyhin, Astrid Dumond, Christopher Gully, Jethro Mullen, Isabelle Nonain-Semelin, Julie Puech, and Therese Walsh.

The IEA held a high-level workshop on Electricity Security in Paris on 28 January 2020. The participants offered valuable insights and feedback for this analysis. Further details are available at www.iea.org/events/iea-electricity-security-workshop.

The authors would like to thank the many external experts who provided valuable input, commented on the analysis and reviewed preliminary drafts of the report. They include: Enrique De Las Morenas Moneo, Francesca Gostinelli and Viviana Vitto (Enel); Hans Martin Füssel, Mihai Tomescu and Stephane Quefelec (European Environment Agency); Hannele Holttinen (IEA Wind Technology

Collaboration Programme Task 25); Jochen Kreusel and Alexandre Oudalov (ABB); Manuel Baritaud (European Investment Bank); Patrik Buijs (Elia Group); Stephen Woodhouse (AFRY); Michael Hogan (Regulatory Assistance Project); Lwandle Mqadi (Eskom); Laurens de Vries (Delft University); Doug Arent and Martha Symko-Davies (National Renewable Energy Laboratory); Takashi Hongo (Global Strategic Studies Institute); Sushil Kumar Soonee (Power System Operation Corporation Limited); Jean-Michel Glachant (Florence School of Regulation); Masato Yamada and Tusitha Abeyasekera (MHI Vestas); Kerry Schott (Energy Security Board, Australia); and Russ Conklin, Fowad Muneer and Carolyn Gay (United States Department of Energy).

Comments and questions on this report are welcome and should be addressed to EMS-RISE@iea.org.

Executive summary

Building on the Paris Agreement signed in December 2015, governments are showing even greater commitment towards reducing greenhouse gas emissions, including reaching net zero in many economies by mid-century. Variable renewable energy (VRE) is driving the ongoing decarbonisation of the power sector, and also reshaping the operation of the electricity system.

Policies will need to be updated accordingly to provide the necessary level of electricity security. In order to address risks to adequacy in the long term and operational security in the short term, regulations and market designs must ensure that all system resources are compensated in accordance with their system value. New products will need to be created in response to the development of flexibility sources such as dispatchable generation, demand response, storage, digitalisation and interconnections. Policy makers should particularly focus on the role distributed energy resources, energy source diversity, energy efficiency and fuel security will play in ensuring a secure and resilient power system at lowest cost to consumers.

Electricity mix trends

The clean energy transition will bring a major structural change to electricity systems around the world. Variable renewable generation has already surged over the past decade. The trend is set to continue and even accelerate as solar photovoltaic and wind become among the cheapest electricity resources due to technological advancements and cost reductions and contribute to achieving climate change objectives. In the International Energy Agency Sustainable Development Scenario, the average annual share of variable renewables in total generation reaches 45% by 2040.

Such rapid growth in VRE will help alleviate traditional fuel security concerns, but it will **call for a fast increase of flexibility in power systems**. On the other hand, conventional power plants, which provide the vast majority of flexibility today, are stagnating or declining, notably those using coal. On the demand side, electrification will increase demand for electricity, and technology and digitalisation are enabling a more active role for consumers as part of more decentralised systems.

Power system planning

Traditional frameworks for ensuring electricity security will not be sufficient in the face of these changes. The challenge for policy makers and system planners is to update policies, regulation and market design features to ensure that power systems remain secure throughout their clean energy transition.

Experience in a number of countries has shown that variable renewables can be reliably integrated in power systems. Many countries and regions in many parts of the world have succeeded in this task using different approaches and taking advantage of their flexibility resources. They leave to the world a large set of tools and lessons to be integrated into the policy maker toolkit.

Gas security will become increasingly relevant to electricity security. Gas-fired plants will play an expanded role in the provision of adequacy, energy and flexibility in their power systems and thus it will be crucial to ensure that gas will be deliverable when needed in instances of high electricity and gas demand combined with low availability of variable renewables.

Delivering resource adequacy

Making the best use of existing flexibility assets and ensuring these are kept when needed should be a policy priority. This will require market and regulatory reforms to better reward all forms of flexibility and careful adequacy assessments of the impact of decommissioning plants of dispatchable supplies.

However, going forward, new additional flexibility resources need to develop in parallel with expanding solar and wind, especially in emerging and developing economies that are facing strong electricity demand growth. Maintaining reliability in the face of greater supply and demand variability will require greater and more timely investments in networks and flexible resources, including demand side, distributed and storage resources, to ensure that power systems are sufficiently flexible and diverse at all times. Development of low carbon fuels in conventional power plants could also help achieve system stability and decarbonisation.

Notably, current investment trends do not support such requirements and will need to be upgraded accordingly, sooner rather than later. Grids are particularly concerning, as annual investment has declined by 16.3% since 2015. Grids also require long-term planning, have long construction lead times and often face social acceptance issues.

Ensuring power system flexibility

Building new assets to provide needed adequacy and flexibility will require an update of current market designs. Increased reliance on renewables will augment the need for technologies that provide flexibility and adequacy to the system. This will include storage, interconnections, natural gas-fired plants in many regions, and demand-side response enabled by digitalisation. Updated approaches to planning will also be necessary, with more advanced probabilistic analyses that account for and enable contributions from all available technologies to adequacy.

Providing system balancing services

Balancing is one of the key processes to ensure security of supply. System operators will be faced with new challenges associated with the energy transition, as **the factors that affect the need for system balancing become more complex and interdependent**. Increasingly, system operators are turning towards market-based mechanisms in order to provide these services at the lowest possible cost, as well as implementing systems that allow for more dynamically managed system requirements such as reserve quantity and short pricing intervals.

Stability procedures should be modernised

Growth in variable renewables and decentralised power sources present technical challenges for the system to **maintain a state of operational equilibrium and withstand disturbances** which can compromise electricity security. Ensuring electricity security means that these technical challenges are addressed through appropriate innovative solutions in system planning, operation and services. For example, declining system inertia can be addressed through new system services such as fast frequency response or new infrastructure such as synchronous condensers. It is therefore essential for policy makers to take action, establishing the necessary logical steps in the system planning process to consider system reliability in an appropriate manner. This may involve a review of connection requirements (including grid code revision and mandatory system service), revised operational practices and innovative market-based solutions such as expanded system services markets.

Summary of recommendations

The world's electricity systems are experiencing profound change. The traditional power sector, where dispatchable sources were controlled centrally with very little reaction from consumers, is gradually transforming into one where variable sources increase their share of the energy mix every year, and where an increasing number of actors and devices actively participate and interact with the power system. This report investigates how policy makers should update existing planning, investment and operational frameworks to maintain the reliability that society requires as its reliance on electricity grows, while accomplishing a successful transition to a low-carbon power system.

This report covers the areas of system adequacy, investment signals in market design, flexibility, real-time system balancing and stability challenges. Electricity security calls for the application of five steps for managing the transition of electricity systems, as follows.

Institutionalise

Policy makers, regulators and system operators need to allocate appropriate responsibilities and incentives to all relevant organisations within their jurisdiction and ensure these organisations co-ordinate their work in practice.

Policy makers, regulators and system operators should:

Recommendations to ensure adequacy

- Provide enough forward visibility of the policies affecting the power sector – considering inputs from other authorities and stakeholders during the decision-making process.
- Regularly assess the market design to ensure that it is bringing the adequacy, flexibility and stability services needed for the secure operation of the system.
- Implement planning frameworks that allow for co-ordination across jurisdictions and across the electricity value chain including, inter alia, grid planners and operators, project developers, large consumers and city planners. This can allow for better development of planning criteria and for long-term investment (such as in transmission grids and flexibility requirements) to better follow market and policy signals.
- Set reliability targets to reflect the changing topography of electricity systems, including the re-dimensioning and expansion of reserves.

Recommendations to ensure system balance and stability

- Provide a clear framework to provide every power sector stakeholder with a clear set of obligations to prevent threats and to react in exceptional circumstances.
- Assign responsibilities for co-ordinated action between the operators of the transmission and distribution systems, including where systems are interconnected.

Identify risks

Policy makers need to ensure that operators of critical electricity infrastructure identify, assess and communicate critical risks. Policy makers should:

Recommendations to ensure adequacy

- Conduct regular adequacy of supply assessments, including appropriate methodologies adapted to VRE variability and all system uncertainties.
- Include gas-related contingencies in their adequacy assessments in jurisdictions relying on gas-fired plants as a flexibility resource.
- Review standards and metrics used in adequacy assessments to ensure all relevant outage risks are captured, including average and extreme events.
- Ensure that both the technical and market operations of the gas and electricity systems are well co-ordinated, particularly where natural gas is used extensively for heating.

Recommendations to ensure system balance and stability

- Consider all flexibility sources as options to satisfy adequacy in planning, including electricity networks and distributed energy resources (energy efficiency, demand response and distributed generation and storage).

Manage and mitigate risk

Policy makers and industry have to collaborate to improve readiness across the entire electricity system value chain. Policy makers should:

Recommendations to ensure adequacy

- Set rules that reward resources for their actual contribution to secure operation, instead of an expected or average contribution.
- Guide investment frameworks consistent with policy uncertainty in adequacy assessments.
- Mitigate the impact of events (crisis or actual contingency) by assessing and reforming adequacy mechanisms when temporary or structural out-of-the-market measures are applied, to guarantee secure operation.

- Assess where increased diversity of the resource mix could ensure resilience against social, geopolitical, market, technical and environmental risks.
- Create market and investment frameworks, including the required digital environment, to enable distributed energy resources to effectively participate in markets and contribute to system adequacy and flexibility needs.

Recommendations to ensure system balance and stability

- Develop grid codes to future-proof connection requirements, while continuously updating and amending them as the needs of the electricity system evolve.
- Review and adapt historic load-shedding plans in the context of embedded generation, digitalisation of the entire value chain and greater economically viable demand response.

Monitor risks and track progress

Policy makers need to ensure mechanisms and tools are in place to evaluate and monitor risks and preparedness, and to track progress over time. This is important at the operational level for individual utilities, as well as at the level of policy makers and regulatory authorities who need to understand if strategic objectives are met. Policy makers should:

Recommendations to ensure adequacy

- Perform resilience tests and keep track of power system reliability.

Recommendations to ensure system balance and stability

- Review substantial events like outages to learn lessons and adapt policies.
- Mandate common planning procedures and information-sharing tools in interconnected systems.

Respond and recover

This entails ensuring resilience goes beyond preventing incidents and includes effectively coping with attacks. Policy makers need to enhance the response and recovery mechanisms of electricity sector actors. They should:

Recommendations to ensure system balance and stability

- Set an emergency response framework with clear responsibilities and liabilities.
- Execute regular response exercises to capture lessons learned and adapt practices.
- Stimulate information logging and sharing to facilitate analysis of actual incidents.

Introduction

Electricity systems are in transition, with new technologies playing a central role

Over the past decade, global attention on the need to mitigate greenhouse gas emissions has increased, as reflected and reinforced by the signing of the Paris Agreement in December 2015. At the same time that policy makers' focus on decarbonisation has grown, technological advancement and cost reductions have led to renewed momentum behind clean energy technologies. The convergence of these trends has created highly favourable conditions to transition the global energy system toward low-carbon technologies. Nowhere is this energy transition more apparent than in the power sector, where wind and solar generation, in particular, have surged globally based on impressive technology gains and falling costs.

These forms of variable renewable energy (VRE) have unique characteristics that are not only driving the ongoing decarbonisation of the power sector, but are also reshaping the operation of the electricity system.

Simultaneously, larger volumes of dispatchable generation, namely coal, nuclear and oil, are facing retirement, especially in advanced economies. These forms of generation have historically underpinned electricity security. The energy transition is therefore transforming the fuel mix in the power sector and raising new concerns about electricity security, as the frameworks and tools for ensuring electricity security face new conditions and require the adjustment of current practices as well as new rules.

Moreover, the energy transition is about much more than just VRE. Again driven by technological progress and decarbonisation agendas, the electricity sector is experiencing an increase in digitalisation as tools such as smart grids and smart meters are deployed to achieve decarbonisation and energy efficiency goals. The rise of distributed energy resources is enabled by digitalisation and is driving decentralisation of the power system. These resources include rooftop solar installations, batteries and demand-side response devices, such as water heaters. This decentralisation has the potential to upend the balance between the transmission and distribution sectors and encourage consumers to play a larger role in the future electricity system's operations.

The energy transition includes a trend toward increased electrification in end-use sectors such as transport and heating, with the potential to drastically alter the balance of supply and demand for electricity and to put electricity increasingly at the forefront of the entire energy system. As such, the notion of energy security for policy makers will entail paying greater attention to electricity security in particular.

These transformations will fundamentally alter the electricity mix and the way the sector is governed, planned and operated from an electricity security perspective. Considerations include increased geographical integration and managing co-ordination among various energy segments within the system. The transition requires changes in technical specifications, operational practices and market design.

This report offers practical guidance to decision makers responding to emerging risks

This report offers practical guidance to energy policy makers and other stakeholders on how to deliver a clean energy transition in the electricity sector in a secure manner. The following questions are addressed:

- How can we measure security of supply to identify systemic flexibility issues and trigger policy action?
- How can market design or other policy measures ensure adequate investment in capacity and flexibility that is cost-effective and in line with sustainability targets? Which flexibility providers need to be kept in the system and what future sources can be tapped into? How do we ensure timely investment?
- Given the central role of electricity in various end uses and the linkages with other energy sectors, does our view of electricity security capture all the uncertainties and all the indirect security risks, in particular at the gas-electricity supply nexus?
- Higher shares of VRE require new technical concepts. How can policy makers guide innovation, facilitate the scaling up of new solutions, steer markets to deliver them and ensure grid codes are future-proof?

Policies need to address risks to adequacy in the long run and operational security in the short run

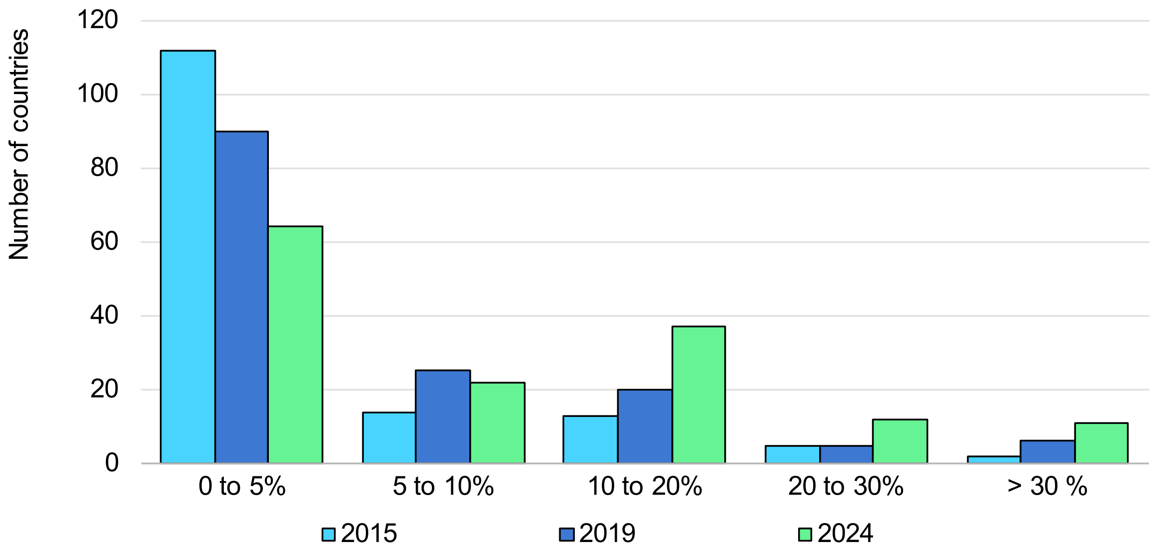
Policy is key to guiding the processes and responsibilities needed to appropriately identify risks. Policy makers need to update regulations and market design features to provide the incentives for the necessary level of security during the energy transition. System operators will need to create new products to remunerate resources and manage this evolution. On both the supply and demand sides, policies need to ensure that the contribution of all system resources to system adequacy is remunerated. And sources of flexibility such as dispatchable generation, storage, demand response, digitalisation and interconnection need to be compensated for their contributions to system balance. Particular attention needs to be paid to distributed energy resources and their ability to reduce system costs like grid investment while also enhancing resilience to extremely high-impact events such as system blackouts. Also, policy makers should value the contribution of system resources to goals such as energy source diversity, energy efficiency, resilience and fuel security in a clear and transparent manner in their planning processes.

Trends in the electricity mix

Growth of VRE accelerates the power system transition

Driven by technological advances and supportive policies, the share of the electricity mix provided by VRE has increased substantially in recent years. In 2015 the number of countries where VRE had an annual generation share greater than 5% was just over 30. By 2019 this number had increased to nearly 50 countries. The share of VRE in many countries or regions is expected to rise from 5-10% to 10-20% over the next five years (Figure 1). Regions with shares of 20-40% are also expected to increase significantly. In the Sustainable Development Scenario of the IEA World Energy Outlook, the share of VRE in a number of regions, including the People's Republic of China ("China"), India, Europe and United States, is set to be higher than 30%.

Figure 1. Share of variable renewables in the global electricity mix



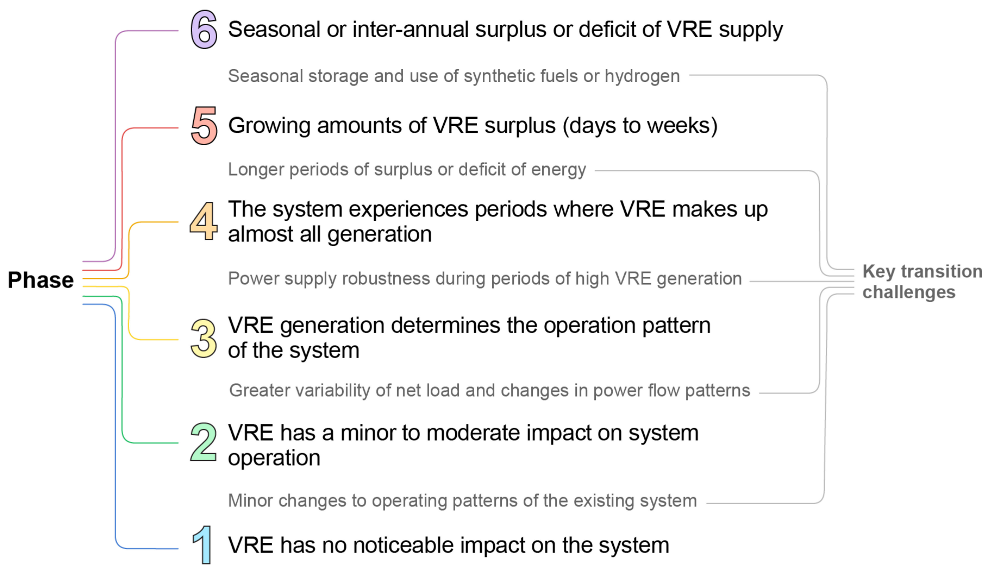
Source: IEA Renewables 2020.

Integration of VRE creates distinct sets of issues at different stages

The rapid growth in VRE penetration in electricity systems around the world raises questions about how to ensure cost-effective and secure integration. The challenges related to VRE integration are context specific. No two systems are the same in terms of legacy infrastructure, solar and wind resources, and flexibility resources. It is almost impossible to derive simple rules linking, for example, a certain annual share of VRE with a specific integration activity or cost. The integration challenges can be categorised according to the potential impacts on system operation, which depend on characteristics such as the size of the system, the technology mix, operational practices and standards, demand patterns and market and regulatory design.

The IEA uses a framework of phase categories to capture the evolving impacts, relevant challenges and priority of system integration tasks to support the growth of VRE. The framework specifies six phases of VRE integration covering the main issues that are experienced (Figure 2). A system will not transition sharply from one phase to the next. The phases are conceptual and intended to help set the order of priority for institutional, market and technical activities. For example, issues related to flexibility will emerge gradually in Phase Two before becoming the hallmark of Phase Three. Two countries may be in different phases even though they share a similar annual VRE share of electricity.

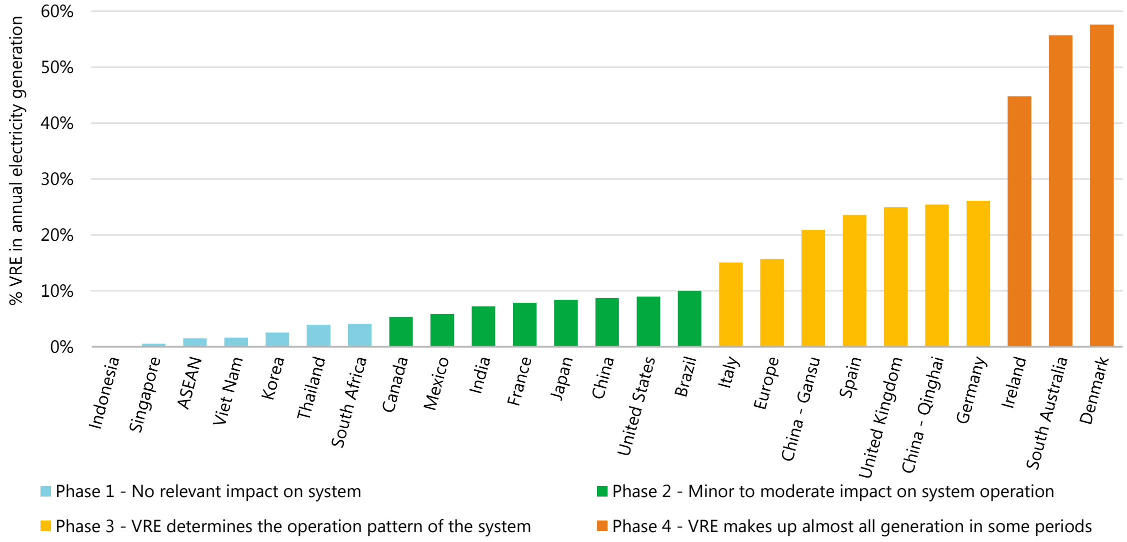
Figure 2. VRE integration phase assessment



Source: IEA Status of Power System Transformation 2019.

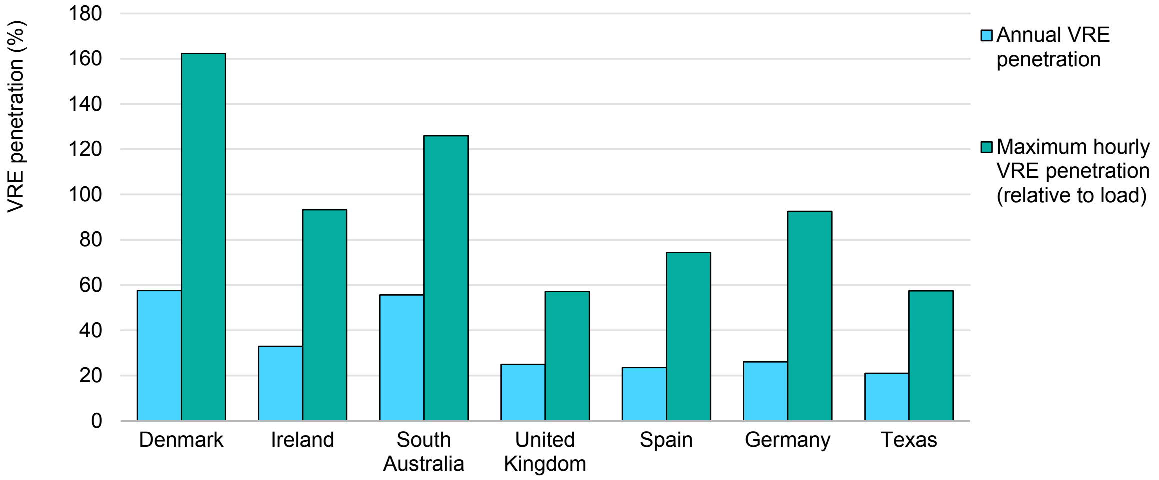
Presently, Phase 4 is the highest VRE integration phase that has been achieved, as for example in Denmark, Ireland and South Australia where instantaneous VRE penetration can be higher than 100%. Many other systems are still in Phases 1 and 2 and have up to 5-10% shares of VRE in annual electricity production. The general trend is clearly one of higher phases of system integration in most countries.

Figure 3. Annual VRE share and corresponding system integration phase in selected countries/regions, 2019



The annual share of VRE provides a general picture of the contribution that VRE sources make to the power system. Much experience has been gained in VRE-leading regions on how to cope with the uncertainty and variability of solar PV and wind. It is also worth highlighting that high instantaneous VRE penetration levels can be an indicator of electricity security challenges. This can be of critical concern to system operators. When VRE production reaches a very high share or even exceeds total system demand in certain periods (as has occurred in regions such as Denmark and South Australia), system stability becomes an electricity security concern as the power system's ability to respond to unexpected events may change. Instantaneous VRE penetration can vary significantly during the year and does not necessarily correlate with demand patterns due to a number of context-specific factors. In many power systems, a maximum permitted level of instantaneous VRE penetration has been set, following detailed studies, to ensure the stability of the system. For example, Ireland has set a maximum level of instantaneous non-synchronous generation (VRE and interconnector imports) to ensure operational security.

Figure 4. Maximum half-hourly or hourly VRE penetration in selected regions, 2019



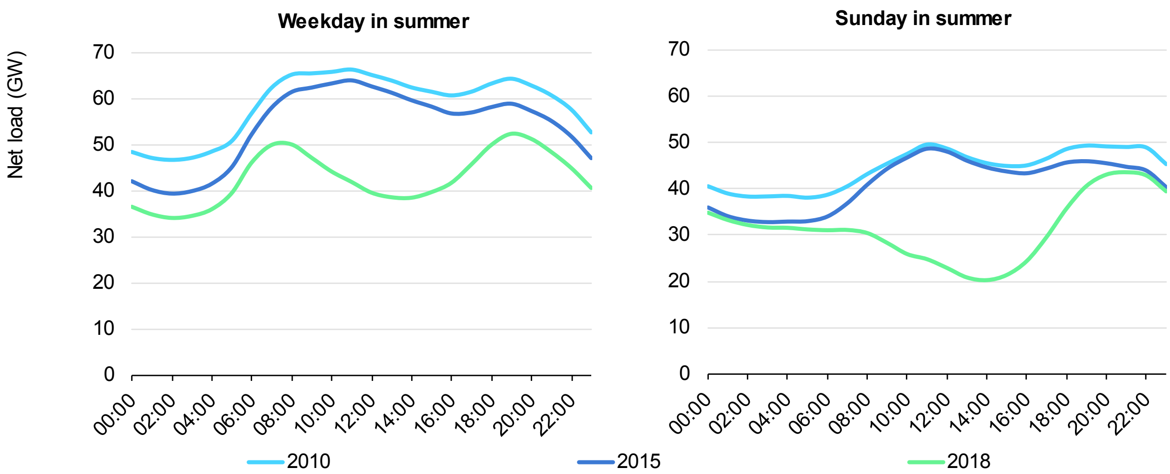
Sources: Based on data from AEMO (2020); ENTSO-E (2020); EIA (2020).

Growing levels of solar PV and wind increase the variability and uncertainty of electricity supply

Electricity systems are always designed to cope with variability and uncertainty. Historically, variability came mainly from the demand side, while uncertainty was instead a supply-side issue caused by the sudden loss of a large generator or transmission asset. Requirements for flexibility are evolving, particularly as the share of wind and solar PV increases. Their output is constrained by the instantaneous availability of wind and solar irradiation. This makes them both **variable** and partly **uncertain**: variable because the available output varies over time depending on the availability of the primary resource (wind or sun); and uncertain as the available output cannot be perfectly forecasted, especially not at longer lead times.

High VRE penetration will have an additional impact on the combined variability and uncertainty that the entire power system needs to cope with. As the system-wide variability needs to be balanced, VRE output is often subtracted from the demand profile to form what is known as a “net load curve”. Many countries are experiencing significant changes in their net load, both the profile shape and magnitude, due to higher VRE generation. In Germany, for example, the variability of daily net load has been increasing over the past several years.

Figure 5. Average hourly net load in Germany on weekdays and Sundays, 2010, 2015 and 2018

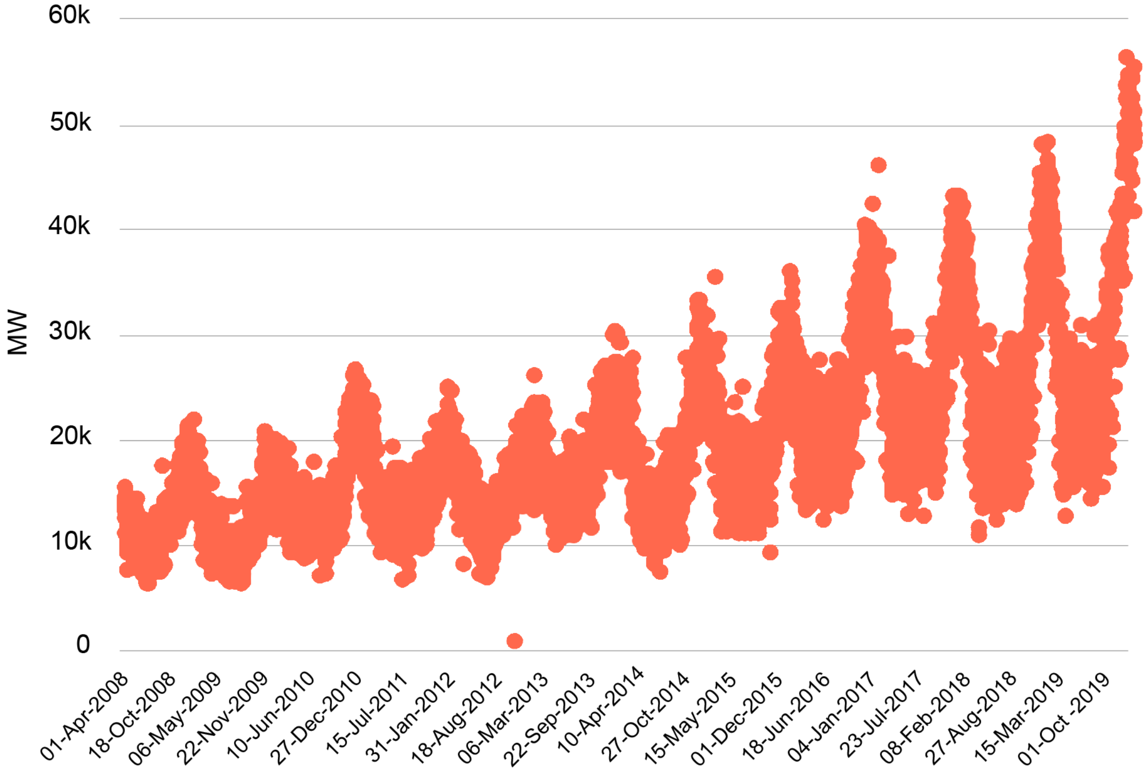


Source: Based on data from ENTSO-E (2020).

Several indicators point to increased flexibility challenges. The net load ramping on the system (both hourly and sub-hourly) is a robust indicator of the flexibility requirement from a variability perspective. Another important indicator is minimum net load. Both ramping and minimum net load need to be met by flexibility sources, which currently consist mainly of conventional generation and to some extent demand response, while storage is increasingly emerging as an option in various power systems.

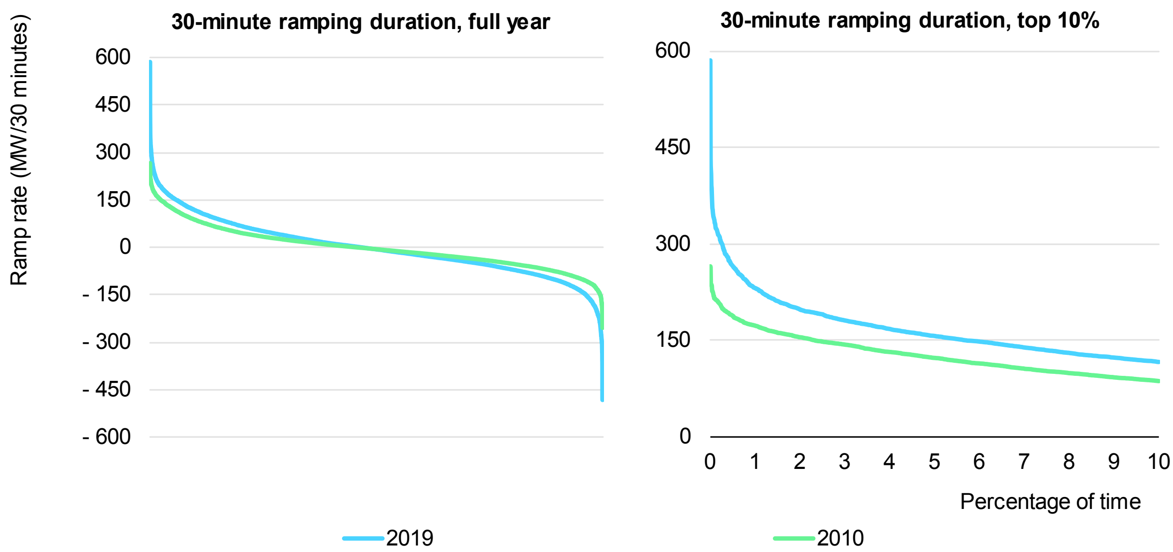
Variability and uncertainty trends are observed in many systems, including for example India and Australia where flexibility requirements associated with meeting higher ramping and a wider spread between minimum and maximum load during the day have become much more evident over the past decade. In India the variability of net load has been increasing in recent years because of a growing share of VRE and to some extent increased air-conditioning usage, as indicated by the increasing gap between minimum and maximum demand. Conventional thermal power generation represents a major flexibility resource in India, where many plants have been retrofitted to achieve higher ramp rates and lower minimum operating levels.

Figure 6. Daily difference between the maximum and minimum demand in India's electricity system, 2008-2019



Source: POSOCO (2020), Flexibility Analysis of Thermal Generation for Renewable Integration in India.

Figure 7. Net load ramps at 30-minute intervals in South Australia, 2010 and 2019



Source: Based on data from AEMO (2020), Market data NEMWEB.

South Australia has more than 40% annual VRE penetration at present and has seen larger net load variability, as indicated by much steeper 30-minute ramps. The maximum 30-minute net load ramp in 2019 doubled compared to 2010. In 2019, the highest 30-minute ramp rates accounted for close to 50% of demand, compared to just 20% in 2010.

Higher VRE penetration also brings additional uncertainty into system operation and planning. The forecasting of a VRE plant generation profile has to manage potentially high levels of uncertainty. This uncertainty of VRE is mainly linked to the accuracy of meteorological forecasts. Using advanced forecasts in grid operations leads to operational cost savings. It can help predict the amount of wind or solar energy available and reduce the uncertainty of the available generation capacity, while reducing the amount of conventional generation that must be held in reserve.

Flexibility needs to be scaled up substantially in the coming decades

Integrating larger shares of VRE requires sufficient system flexibility to keep supply and demand in balance in a cost-effective and reliable manner (which may imply sufficient reserves). Understanding the flexibility requirements across timescales can inform policy makers and developers on which actions are best suited to enhancing system flexibility. This can lead to effective utilisation of, and levels of investment in, different flexibility resources given their potential contribution.

In addition to the short-term timescales (seconds to hours), which are generally associated with meeting peak load, minimum net load and system ramps, flexibility requirements in the medium- and long-term timescale (hours to seasons and years) are also important. The critical factor in these timescales is the amount of energy that can be called upon to supply electricity during periods of high demand, and options to help utilise generation during times of low demand or high VRE production. Storage options such as pumped storage hydro, power-to-fuel conversion and batteries, as well as distributed energy resources (including demand response, distributed generation and electric vehicles) have the potential to provide flexibility services, but their respective effectiveness varies across different timescales.

Requirements for system flexibility can be assessed according to the variations in net load over a day, a week and eventually also across seasons.[1] Relevant indicators that can be used to assess flexibility requirements include the flexibility capacity (MW) needed to cope with net load variations, ramp capabilities (MW/min) and flexibility volume (MWh). These indicate the extent to which flexibility options would be needed on a daily and weekly basis to cope with net load profile variability.

A power system today typically requires flexibility resources to provide a buffer spanning a period of between one and six hours to meet its daily and weekly flexibility requirements.

These indicators provide insight into whether a system requires more short-term or long-term flexibility resources, or in instances where storage is being considered, the order of magnitude of volumes that are relevant. If the analysis is conducted for many climatic years and load scenarios, probabilistic insights can provide an indication of how often specific flexible capacities and volumes may be needed.

Flexibility requirements in these daily and weekly timescales are increasing in many systems with a relatively high share of VRE, particularly in relatively small isolated systems and those with limited interconnection capacity, such as South Australia and Ireland. For systems with greater interconnection, such as Denmark, it is less challenging to integrate high shares of VRE.

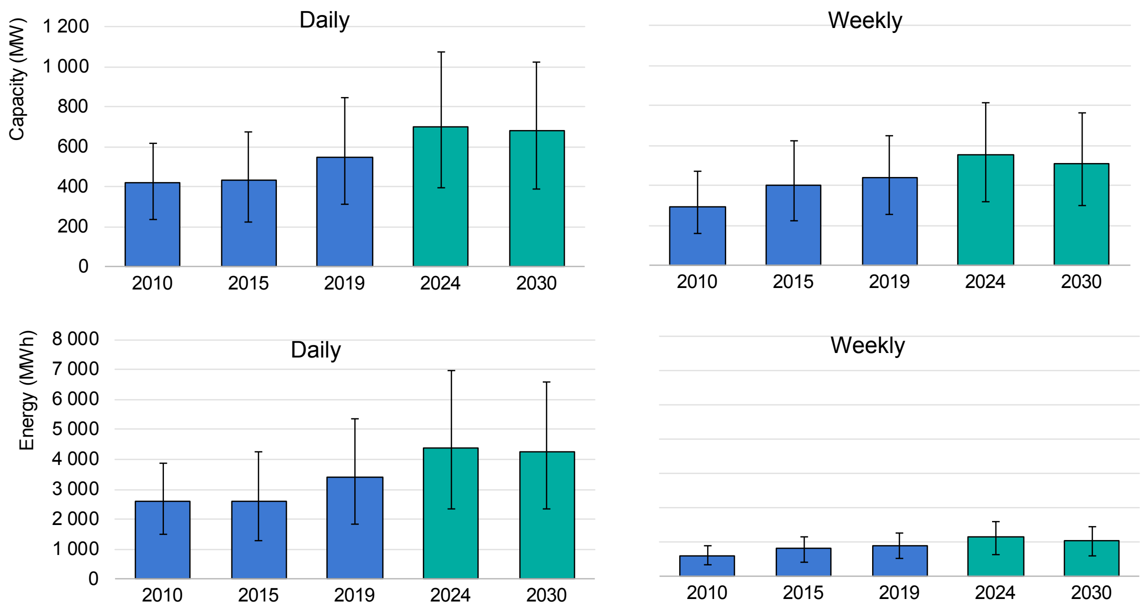
In South Australia, which is part of the National Electricity Market (NEM), flexibility needs have grown over recent years, driven by the relatively high share of VRE, which is already beyond 40%. Large grid-scale battery storage is one of the main flexibility resources in the state, consisting of Hornsdale Power Reserve and the Dalrymple battery system, installed in late 2017 and 2019 respectively. With their capability to provide fast frequency response, these resources have contributed to improving system security and also driving down the price of frequency control ancillary services (FCAS), which was a major issue due to the considerable share of inverter-based generation. The contribution of grid-scale batteries to system security and FCAS prices is particularly evident during large system events, which cause the islanding of the South Australian region from the rest of the NEM, such as during the events in November 2019 and January 2020.

[1] Daily flexibility is computed by looking at the variation of maximum and minimum hourly net load each day from the daily average, resulting in 365 values assuming one year of climatic data. Weekly flexibility is computed for each week throughout the year based on the variation of maximum and minimum average daily net load from the weekly average, resulting in 52 values for each climatic year.

Demand response is another flexibility resource that is playing an increasing role in maintaining the reliability of the system by providing FCAS and emergency reserve, as evidenced during a number of events such as extreme weather events and the islanding of the South Australian system in January 2020.

The daily and weekly flexibility requirements to accommodate VRE generation in South Australia have been increasing over the past decade. This is due not only to the rising share of VRE, but also its unique characteristics of being an isolated system relying on interconnectors. Its flexibility requirements are expected to grow over the next five years with the continued increase in VRE penetration. However, as VRE penetration is projected to be lower in 2030 under the central scenario, this could potentially lead to reduced flexibility requirements.

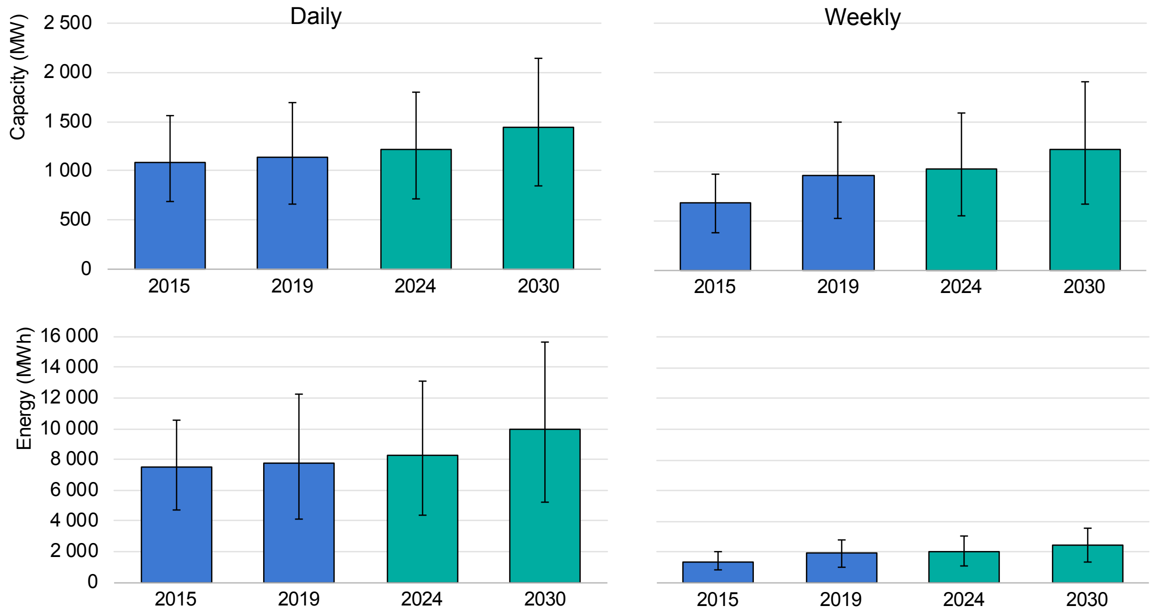
Figure 8. Flexibility requirements in South Australia, 2010-2019 and 2024-2030



Source: Based on data from AEMO (2020), Market data NEMWEB.

With an annual share of wind energy close to 30%, Ireland is one of the countries leading the way to increase system flexibility to reliably and cost-effectively support wind integration. To ensure a secure electricity system, Ireland initially set the maximum instantaneous share of non-synchronous generation (mainly wind) at 50% in 2010, but has succeeded in raising it to 75% in 2020 through a series of measures. These include the development of enhanced operational practice and system services arrangements, such as a fast frequency response market product, aimed at addressing short-term flexibility by incentivising capable technologies to contribute frequency response. Flexibility requirements in Ireland are expected to grow over the next decade with the rising share of wind generation.

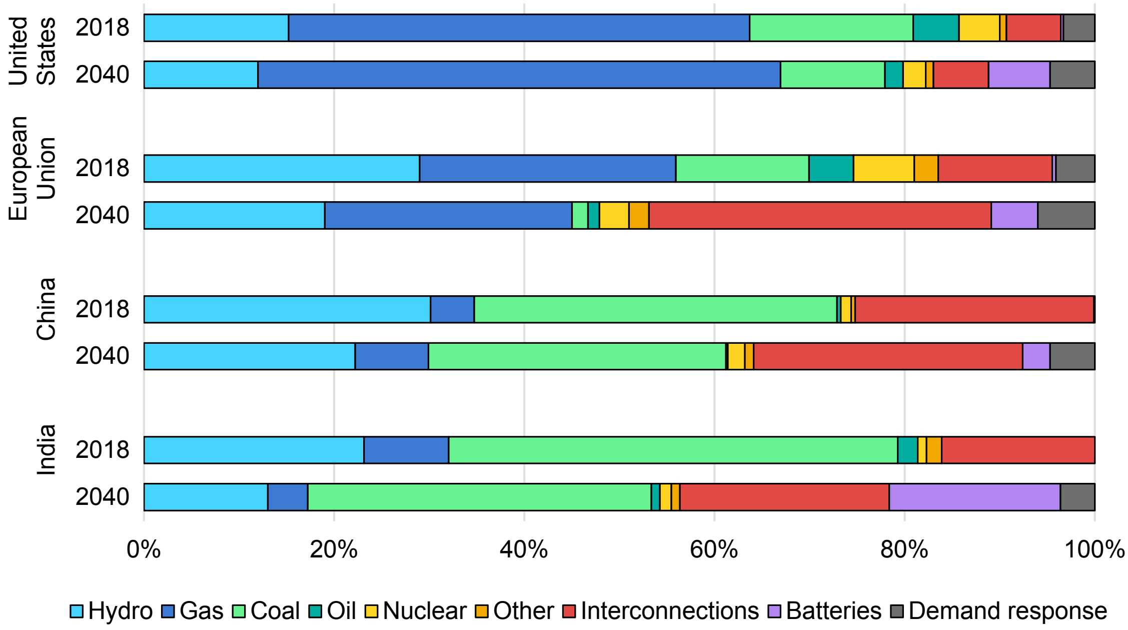
Figure 9. Flexibility requirements in Ireland, 2015-2019 and 2024-2030



Source: Based on data from ENTSO-E (2020).

With the projected increase in VRE penetration in many countries over the coming decades, the need for flexibility will rise. The IEA World Energy Outlook Stated Policies Scenario (STEPS) shows regional trends to 2040 for the combination of flexibility resources that is required. Conventional power plants and networks have long been the major source of flexibility around the world, and they are still expected to play a role in 2040, particularly in emerging economies such as China and India. However, emerging flexibility resources, such as batteries and demand response, including from new electrification loads such as electric vehicles, will become prominent flexibility resources in 2040. Advanced VRE resources can also potentially contribute to system flexibility, which has led increasing number of countries to introduce market reforms and regulations that activate flexibility from VRE resources.

Figure 10. Relative capacity of non-VRE flexibility resources in different regions in the Stated Policies Scenario, 2018 and 2040

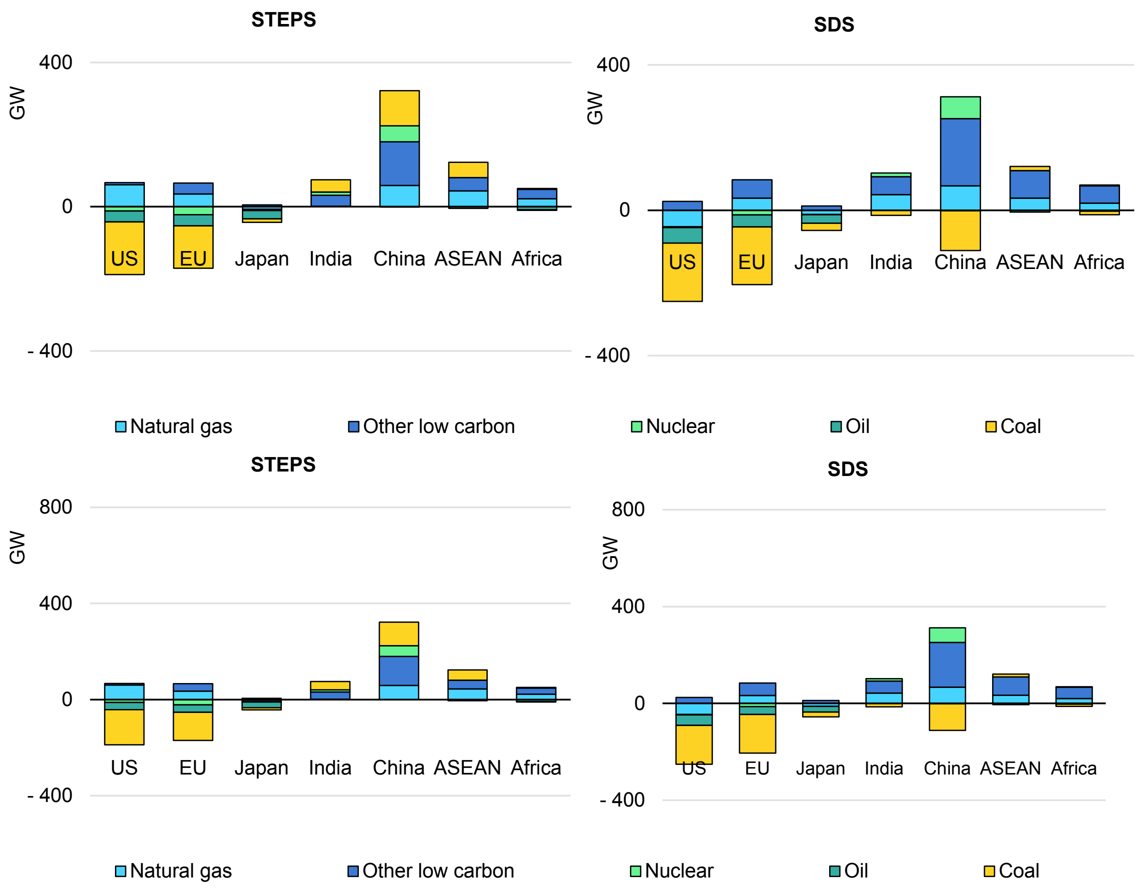


Note: Wind and solar generation can also provide flexibility in the form of short-term balancing services, as well as contributing to adequacy requirements.
Source: IEA World Energy Outlook 2019.

As VRE increases and dispatchable thermal generation declines, flexibility sources need to be scaled up

The coming decade will see an unprecedented shift in the electricity mix and in the way the electricity sector functions. The IEA Stated Policies Scenario (STEPS) shows how advanced economies will see a sizeable retirement of coal, oil and nuclear generating capacity, reducing the amount of dispatchable capacity available in their systems (Figure 11). In a path compatible with the Paris Agreement, such as the IEA Sustainable Development Scenario (SDS), the retirement of dispatchable fossil fuel fired resources is more pronounced, while the greater deployment of dispatchable low-carbon sources is needed.

Figure 11. Dispatchable generation capacity, net additions by technology in the 2020-2030 period



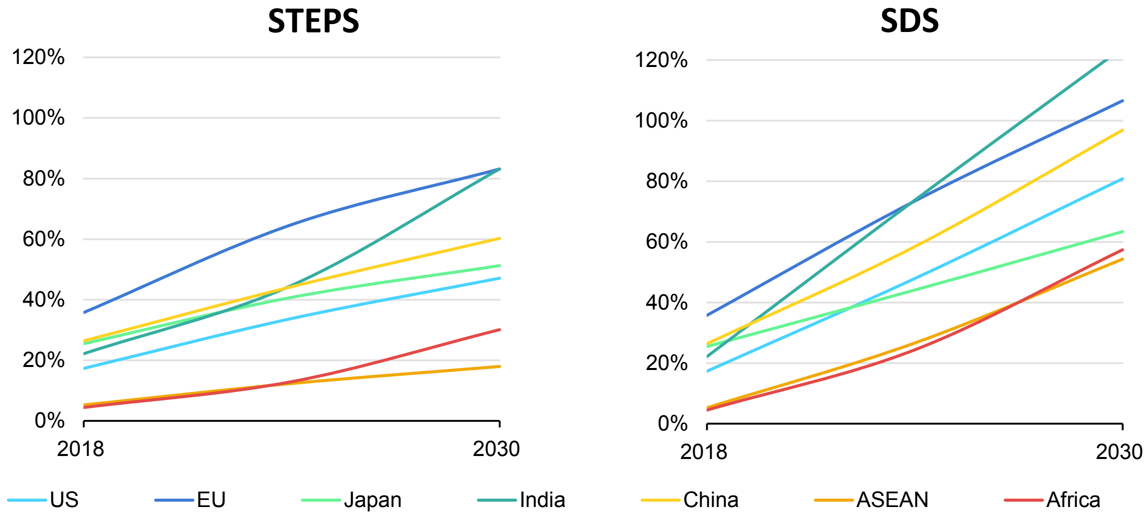
Note: Other low-carbon includes geothermal, concentrated solar power, biomass and marine.
Source: [IEA World Energy Outlook 2019.]

Under present policies, these retirements may be partially compensated by additional gas-fired generation and, to a lesser extent, low-carbon dispatchable sources. If additional policies are implemented to curb projected carbon emissions towards a path that is in line with the Paris Agreement, the speed and level of retirement of fossil fuel thermal generation are most likely to be even stronger.

The outlook to 2030 is very different for countries witnessing large electricity growth rates. Countries such as India and China, as well as those in the ASEAN region and Africa, still rely on a wider array of dispatchable thermal capacity, particularly coal and nuclear, and under present policies will continue seeing net additions despite the increasing amount of renewable resources.

Regardless of whether countries experience large reductions in dispatchable capacity or net additions, most countries and regions will experience important changes in the way their power systems operate due to the large additions of VRE generating capacity. Even countries expected to see net additions in dispatchable thermal capacity will experience a growing share of VRE in the generation mix.

Figure 12. VRE capacity as a percentage of dispatchable generating capacity in STEPS and SDS up to 2030



Source: IEA World Energy Outlook 2020.

VRE capacity as a percentage of dispatchable capacity, as shown for various countries and regions, provides a good measure of the extent to which a system's structure will change as a result of VRE integration, even if it does not account for intra-regional trade and interconnections. It is an initial indicator of the extent to which its operations will need to adapt. VRE itself is not a direct substitute for dispatchable technologies, but rather one part of a portfolio of resources needed to fulfil the power system's requirements. Increasing shares of VRE require the timely and proportional deployment of flexible resources, alongside a paradigm shift in how a flexible system is designed and operated in a cost-effective and secure manner.

Low-carbon dispatchable generation sees limited growth

The shift to VRE is increasing attention on the need for dispatchable generation. The key characteristic of dispatchable generation is its capability to modify output as required by the system. Dispatchability comes in various grades, however. Wind and solar PV evidently depend on the availability of wind and solar irradiation. They can be combined with storage, joined together in a larger portfolio, combined with other power sources, and given instructions to reduce output, but all this still makes them only partially dispatchable.

There are other low-carbon power plants that are largely dispatchable and are widely deployed, including reservoir hydropower and bioenergy. They still see further growth in most projections, but are often limited, respectively, by local hydro site suitability and climatic conditions, and by bioenergy supply levels.

Other low-carbon supply resources that are considered dispatchable include concentrated solar power, nuclear and thermal gas or coal generation with carbon capture, use and storage (CCUS). They have the technical capability to vary their output to accommodate changes in demand and supply, but also have limitations. Ideally they should be operated as baseload plants given their high capital intensity and low operating costs. Operating them at their rated power level on a continuous basis is usually more cost-efficient and simpler.

Nuclear power currently accounts for the largest source of low-carbon generation. While some designs can be operated flexibly, historically nuclear energy has been considered an inflexible resource in most countries due to safety requirements. In countries where nuclear power provides a considerable share of electricity generation, such as France, the flexible operation of nuclear units has been applied with load-following capabilities, but this still needs to be complemented with other more flexible options.

CCUS has the potential to play an important role in the energy transition, reducing emissions across the global energy system. The technology needs to be expanded at a significant scale to decarbonise the electricity sector. To date, experience with flexible operation is limited as there are not many large-scale CCUS projects in existence around the world.[2] Techniques are available to potentially enhance the flexibility of CCUS-equipped thermal plants, enabling them to be fully dispatchable and provide flexibility services.

For some of the largest electricity-consuming regions, such as the United States, the European Union and Japan, net additions of low-carbon dispatchable capacity may not make up for the amount of nuclear capacity being retired. Although nuclear power generation is expected to increase globally by around 30% to 2040, its share of supply is expected to remain relatively similar to present levels. The growth in nuclear generation is mainly seen in emerging economies, while it faces

[2] There were two large-scale CCUS projects using post-combustion capture technology applied to coal-fired power plants: the Boundary Dam project in Saskatchewan, Canada, and the Petra Nova Carbon Capture project in Texas, United States, with annual capture capacities of 1.0 $MtCO_2$ and 1.4 $MtCO_2$, respectively. The Petra Nova Carbon Capture project has been mothballed since May 2020.

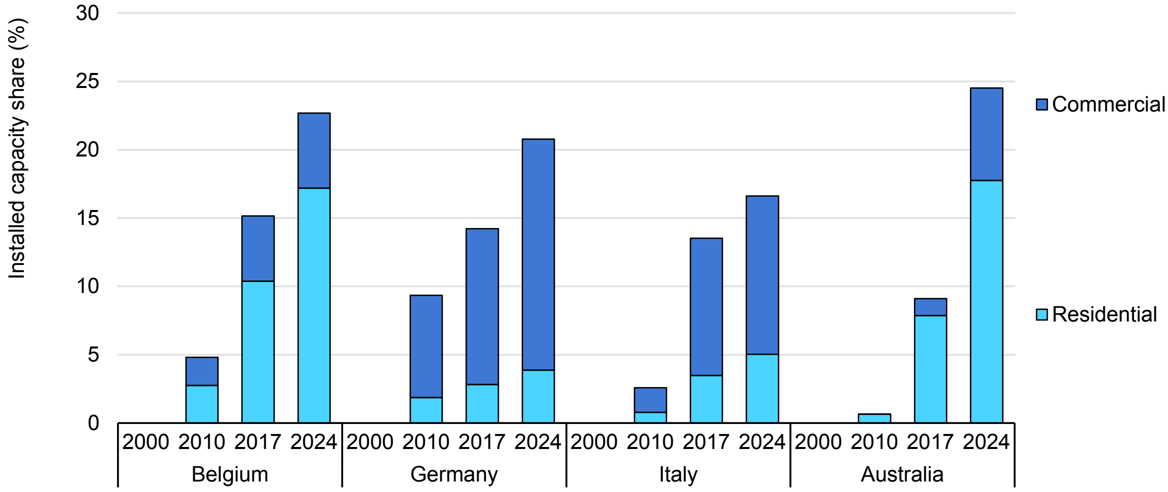
retirements in advanced economies. In the period to 2040, CCUS could play an increasingly important role as a dispatchable low-carbon generation, reaching 5% of the generation mix under the IEA Sustainable Development Scenario (SDS), with 320 GW of coal and gas plants being equipped with CCUS.

The uptake of distributed energy resources means electricity security is becoming a greater concern in distribution systems

A large increase in distributed energy resources is another key aspect of the energy transition, with implications for electricity security. One aspect of this is the growth in distributed generation, particularly solar PV in the form of rooftop and small-scale residential, commercial and industrial solar PV installations. These can affect the system at both a local and system level. In addition, an increase in smart capabilities is creating an important opportunity for the demand side to actively provide system services that help to maintain electricity security. Smart capabilities are becoming commonplace in existing loads such as air conditioners and in new loads resulting from the electrification of different sectors such as electric vehicles.

Impressive growth in distributed PV systems has been seen in many countries in recent decades. In 2000 most countries had no distributed PV generation, while by 2024 some countries should have more than 20% of their generating capacity in distributed systems. As a result, in some time periods distributed PV will provide a large share of generation and its responses during stress situations affect the whole system. While full decentralisation of electricity supply is not anticipated, except on dedicated sites, this is still a substantial shift happening at a rapid pace, which requires attention to ensure secure operations.

Figure 13. Distributed PV capacity as a share of total installed capacity, 2000-2024



Sources: AEMO (2020b); ENTSO-E (2019); IEA (2019d).

The growth in distributed PV has implications not only for the local grid, but also at the system level since it can make up a very large share of generation in sunny periods. At the level of the distribution grid, local generation affects power flows and can require both changes in operational practices and potential infrastructure upgrades to ensure local power quality. At the system level, the cumulative capacity of many small generators can become comparable to the capacity of large, centralised plants and even potentially come to represent the single largest contingency in the system.

As a result, distributed generators become a critical element in security of supply, both in terms of their protection settings and the need to be considered in setting operational reserves. Appropriate settings ensure distributed generation can support the system and minimise these risks. The shift to greater distributed generation also requires more active co-ordination and communication between distribution and transmission system operators to facilitate the visibility, control and management of distributed resource impacts and avoid conflicts of interest. Co-ordination between distribution and transmission system operators also allows the flexibility available within the transmission and distribution grids to be better exploited to provide a secure, resilient and cost-effective system.

The combined trends of digitalisation and electrification also bring challenges and opportunities that are focused on the distribution grid. Electric vehicle uptake has the potential to impose large additional requirements on the local grid, while at the same time, as a resource, they come with great potential for advanced charging control and management to limit additional costs and provide benefits both locally and to the larger system. Fully unlocking the potential contribution of electric vehicles to the electricity system, for example through smart charging or vehicle-to-grid flows, is expected to come with additional infrastructure costs that need to be balanced against this contribution. Enhanced data collection and control technology also opens up opportunities at the distribution level, allowing advanced data-driven operations and planning. In addition, distributed energy resources such as battery storage can be deployed as grid assets, as an alternative to other infrastructure reinforcements.

Overall, distributed energy resources will both provide new opportunities and increase overall system complexity at the distribution grid level, and investment in traditional infrastructure will be complemented by a move towards smart grid technologies and new solutions. Annual average distribution grid spending over the next decade to 2030 increases to over USD 300 billion per year in the IEA World Energy Outlook Stated Policies Scenario (STEPS), 60% higher than in 2020, and would need to rise even further in the Sustainable Development Scenario (SDS) despite slower electricity demand growth.

The way forward for our electricity security frameworks

The energy transition will require new approaches to maintaining system adequacy, balancing and stability.

An integrated approach to power sector planning can help to manage increasing system complexity

Traditional network planning processes primarily focused on expanding supply infrastructure (generation, transmission and distribution networks) to meet projected electricity demand growth over the next 20 to 30 years. However, the power sector landscape is changing. This is largely due to increases in the uptake of VRE and distributed energy resources, including demand-side participation, and the electrification of transport and heat. Power sector planning needs to become more sophisticated by taking into account the role and impact of these developments. A well-integrated planning approach that considers these factors will help identify pertinent options for future power systems in a timely manner.

Electricity networks continue largely to be viewed as natural monopolies that need to be regulated. In unbundled systems the planning processes for generation and network investment are considered separately even though their respective contributions to meeting system flexibility needs are strongly interlinked. In systems with integrated utilities, the uptake of VRE is frequently much slower or VRE capacity is developed by independent power producers, which again decouples generation and network planning.

While distribution networks have historically depended on power supplied by the transmission network, and distributed it to consumers, the situation is changing since more generation resources are being added to the distribution network locally at low- and medium-voltage levels. Where deployment of many smaller VRE plants is concentrated geographically, reverse flows from the distribution network up to the transmission level become increasingly common, and congestion on distribution networks can grow and must be managed securely. Most distribution networks are physically able to manage two-way flows of power, although a number of upgrades and operational changes in voltage management and protection schemes can be necessary.

Closer co-ordination between transmission system operators (TSOs) and distribution network operators is important to deal with this change. Policy makers can help ensure that transmission and distribution planning processes are better integrated with generation planning, particularly as the latter begins to take system flexibility into consideration. Appropriate planning rules for expansion in the electricity sector – covering grid, new generation, storage and other flexibility options – will play a crucial role. Importantly, policy makers and regulators will also need to empower grid operators to include multi-year planning and investment related to climate change adaptation within network planning. This will allow grid operators to undertake costly infrastructure upgrades to bolster resilience.

In many jurisdictions, increasingly integrated and co-ordinated planning frameworks have played a critical role in the cost-effective and secure transition to a new electricity mix. Integrated planning exercises can still be useful in unbundled systems to inform project developers, grid operators and authorities. Co-ordinated and integrated planning practices that are emerging can be broadly categorised into:

- Integrated generation and network planning and investment.
- Interregional planning across different balancing areas.
- Integrated planning across a diversity of supply and demand resources (and other non-wire alternatives).
- Integrated planning between the electricity sector and other sectors.

Security can be enhanced by regional co-operation

Co-operation between neighbouring electricity systems is essential to electricity security. In synchronous interconnected systems such as the east and west interconnections in the United States and the Continental European grid, each system needs to maintain the required frequency by balancing loads, resources and net interchange. At a basic level, defining common terms and understanding the processes used by each system will enhance the ability to communicate. Using common definitions for terms like “emergency condition” allows for efficient responses by neighbouring systems during reliability events.

At a deeper level, mutual assistance agreements that compel operators to respond to emergency conditions in neighbouring systems can prevent cascading outages. Outage co-ordination optimises transmission and generation during an outage to ensure that sufficient resources are online. Congestion management enhances efficiency by more closely aligning the physical state of the grid to energy markets. More broadly, regional co-ordination and co-operation provide the electricity system with diversification of demand and generation patterns, allowing efficient allocation between neighbouring systems that co-ordinate. For this reason regional co-operation both within and between countries has significant economic and security benefits. These issues have been analysed in greater detail in the IEA's Integrating Power Systems across Borders report.

Table 1. Interregional co-ordination tasks in select interconnected systems

	United States (MISO/PJM)	Europe	India	China
Reliability co-ordination	✔	✔		
Transmission planning co-ordination	✔	✔	✔	✔
Congestion management	✔	✔		
Market co-ordination	✔	✔	✔	
Outage co-ordination	✔	✔		
Adequacy assessment		✔	✔	

Note: The Midcontinent Independent System Operator (MISO) and PJM Interconnection (PJM) are independent system operators comprising utilities in the East and Midwest regions of the US and the Canadian province of Manitoba.

Interconnected European and US regions already co-ordinate their activities extensively. In India and China some regional co-ordination is also ensured between state/province-level control centres. For example, in the United States, co-ordination between the Midcontinent Independent System Operator (MISO) and PJM Interconnection (PJM), two independent system operators with a combined peak load of 268 GW, is governed by a joint operating agreement that has been in place since 2004. While the agreement initially only covered reliability-related issues, it has been extended to include economic co-ordination, including procedures to manage congestion at flowgates affected by both systems through economic redispatch.

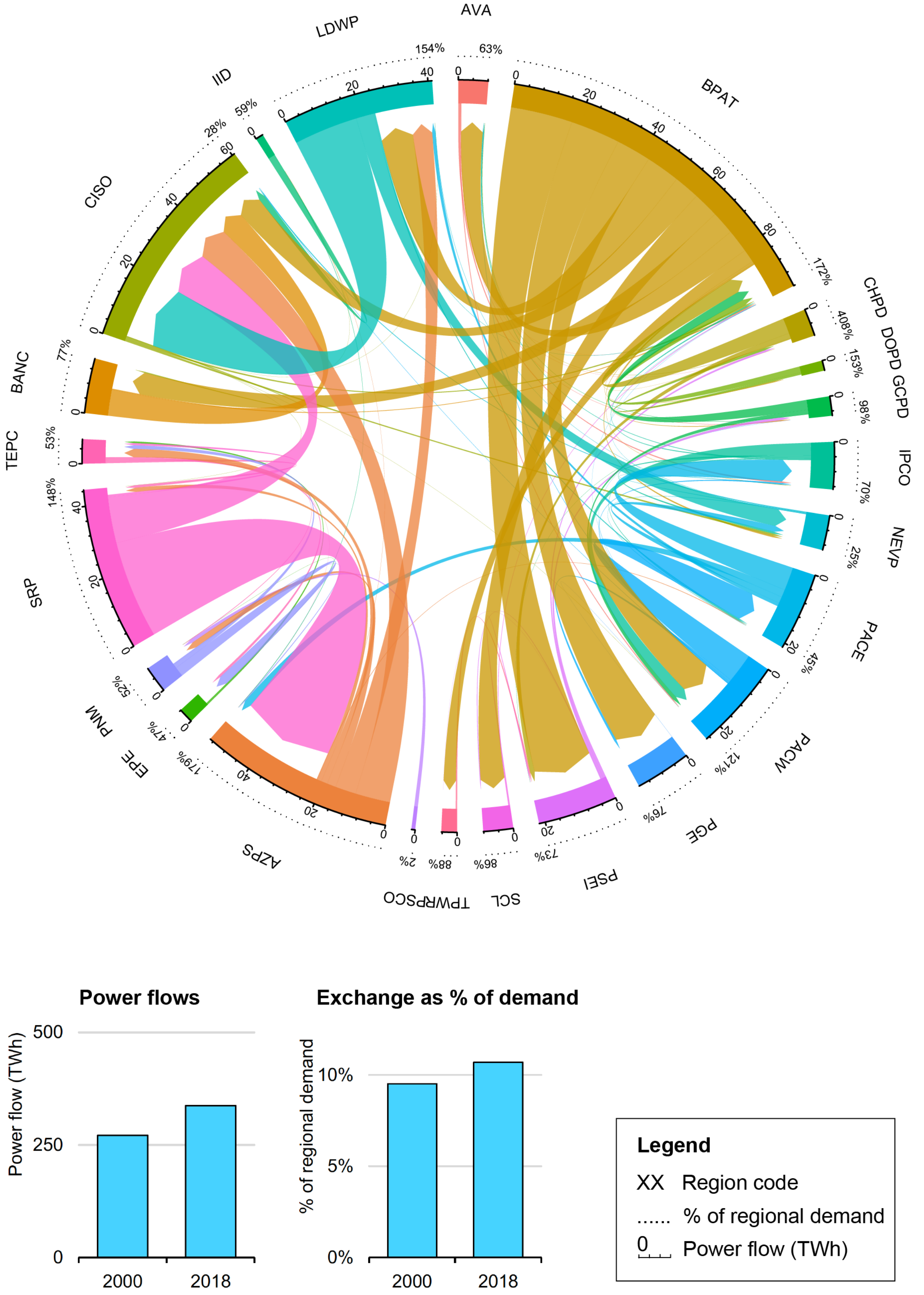
In Europe co-ordination has evolved over time since the legislative third energy package in 2009 initiated the first series of Europe-wide network codes, establishing – among other roles – Regional Security Co-ordinators (RSCs). The RSCs have several tasks to ensure efficient market functioning, and monitor secure operations in support of national TSO actions. In the latest European legislative package (the so-called Clean Energy Package), the RSCs are enhanced to become Regional Co-ordination Centres (RCCs), which provide them with more tasks and greater authority in system operations.

In other regions across the world, markets are developing to allow for regional co-ordination. One such example is India, which has introduced a national spot market. Progress is slow, however, given that India is still heavily reliant on long-term physical power purchase agreements (PPAs), with only around 4% of consumption traded on power exchanges in 2017/18. 2020 has seen an increase in trades, as well as the introduction of a new real-time market in June. When regions are reliant on long-term physical PPAs, the result is less efficient dispatch. Once nationwide spot markets become more liquid, we can also expect interregional security co-ordination to become more critical, as has been the experience in the United States and Europe. A similar trend is seen in China, where the system is moving from administration-based dispatch to market-based trading, and is accommodating larger interregional flows.

Cross-border integration brings economic and ultimately security benefits, which should increase in future systems

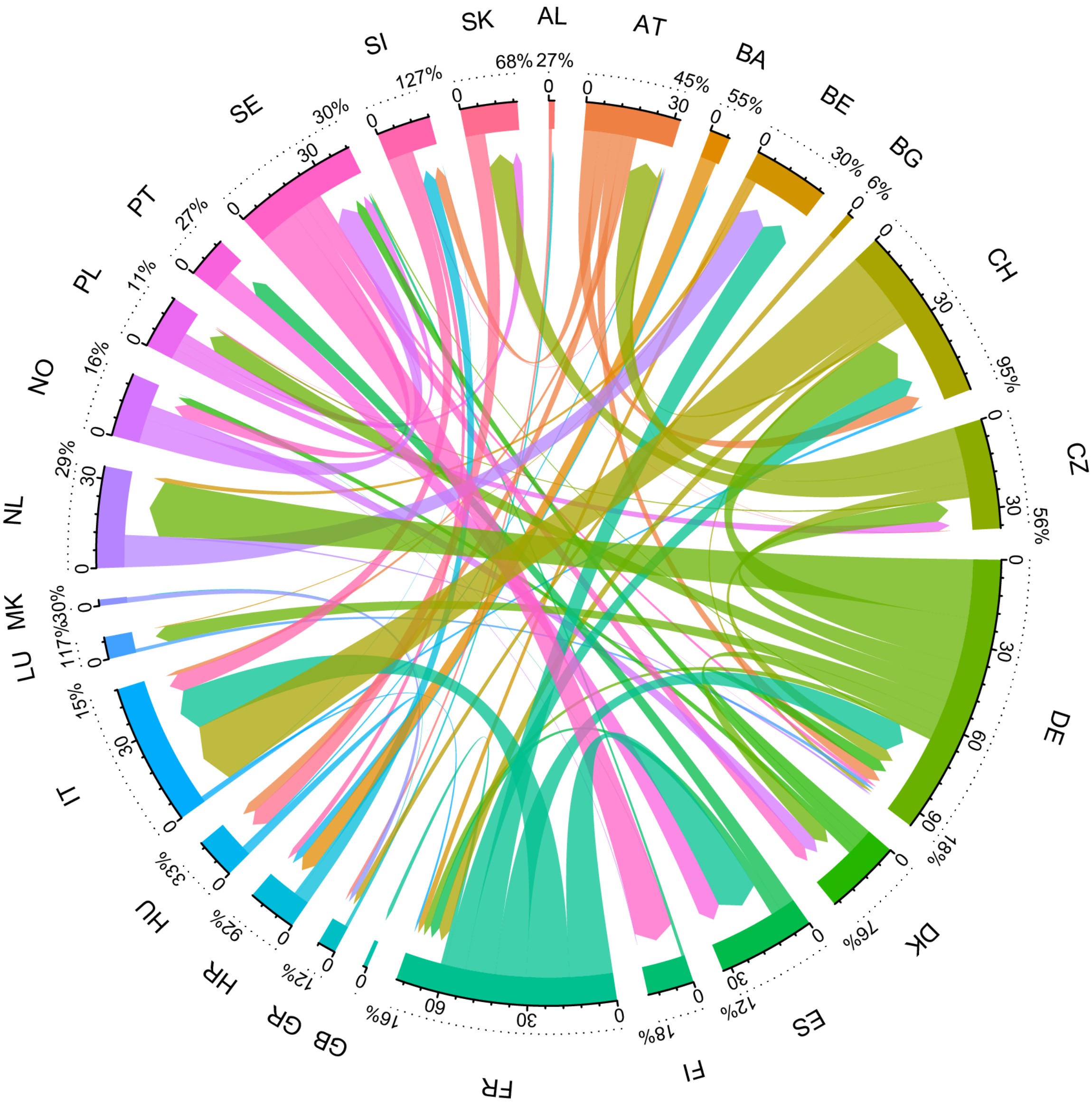
The energy transition has magnified the economic and energy security benefits of regional co-ordination and integration. The generation of clean energy – such as wind, solar PV and hydro – and the prevailing demand profile are highly location-specific and vary by region according to factors including weather conditions and electricity end use. Many regions have sought interconnection with neighbouring regions, where possible, to combine their complimentary systems, increase security and create aggregating effects – and eventually bring down overall system costs.

Figure 14. Cross-border exchanges in Europe, 2018



Sources: UCTE Statistical Yearbook 2000, ENTSO-E Statistical Factsheet 2018.

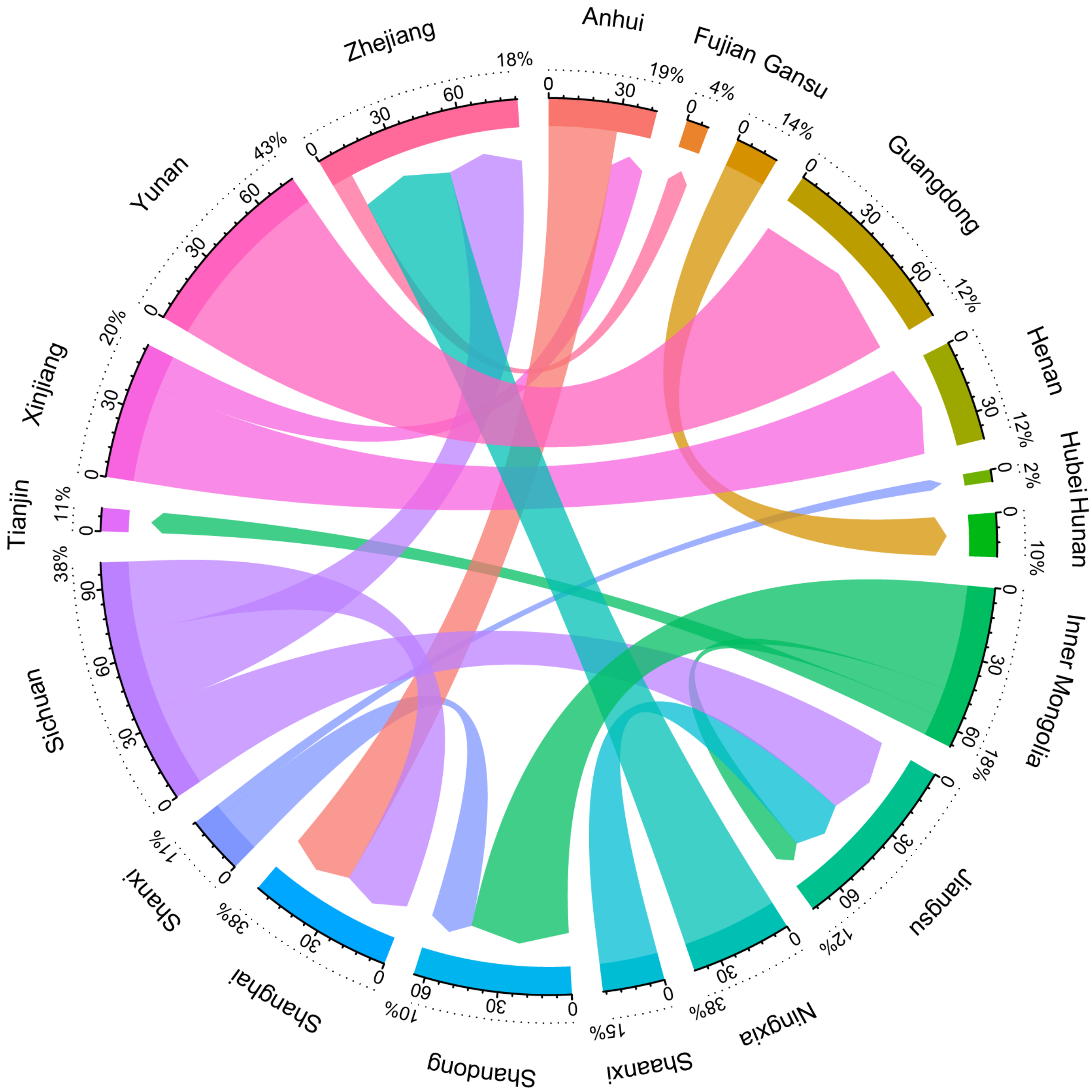
Figure 15. Interregional exchanges in US Western Interconnection between balancing authorities, 2018



Note: Balancing authorities: Avista Corporation (AVA), Arizona Public Service Company (AZPS), Balancing Authority of Northern California (BANC), Bonneville Power Administration (BPAT), Public Utility District No. 1 of Chelan County (CHPD), California Independent System Operator (CISO), PUD No. 1 of Douglas County (DOPD), El Paso Electric Company (EPE), Public Utility District No. 2 of Grant County, Washington (GCPD), Idaho Power Company (IPCO), Imperial Irrigation District (IID), Los Angeles Department of Water and Power (LDWP), Nevada Power Company (NEVP), PacifiCorp East (PACE), PacifiCorp West (PACW), Portland General Electric Company (PGE), Public Service Company of New Mexico (PNM), Public Service Company of Colorado (PSCO), Puget Sound Energy, Inc. (PSEI), Seattle City Light (SCL), Salt River Project Agricultural Improvement and Power District (SRP), Tucson Electric Power (TEPC), City of Tacoma, Department of Public Utilities, Light Division (TPWR).

Source: EIA Open Data.

Figure 16. Chinese interprovincial ultra-high-voltage flows, 2018

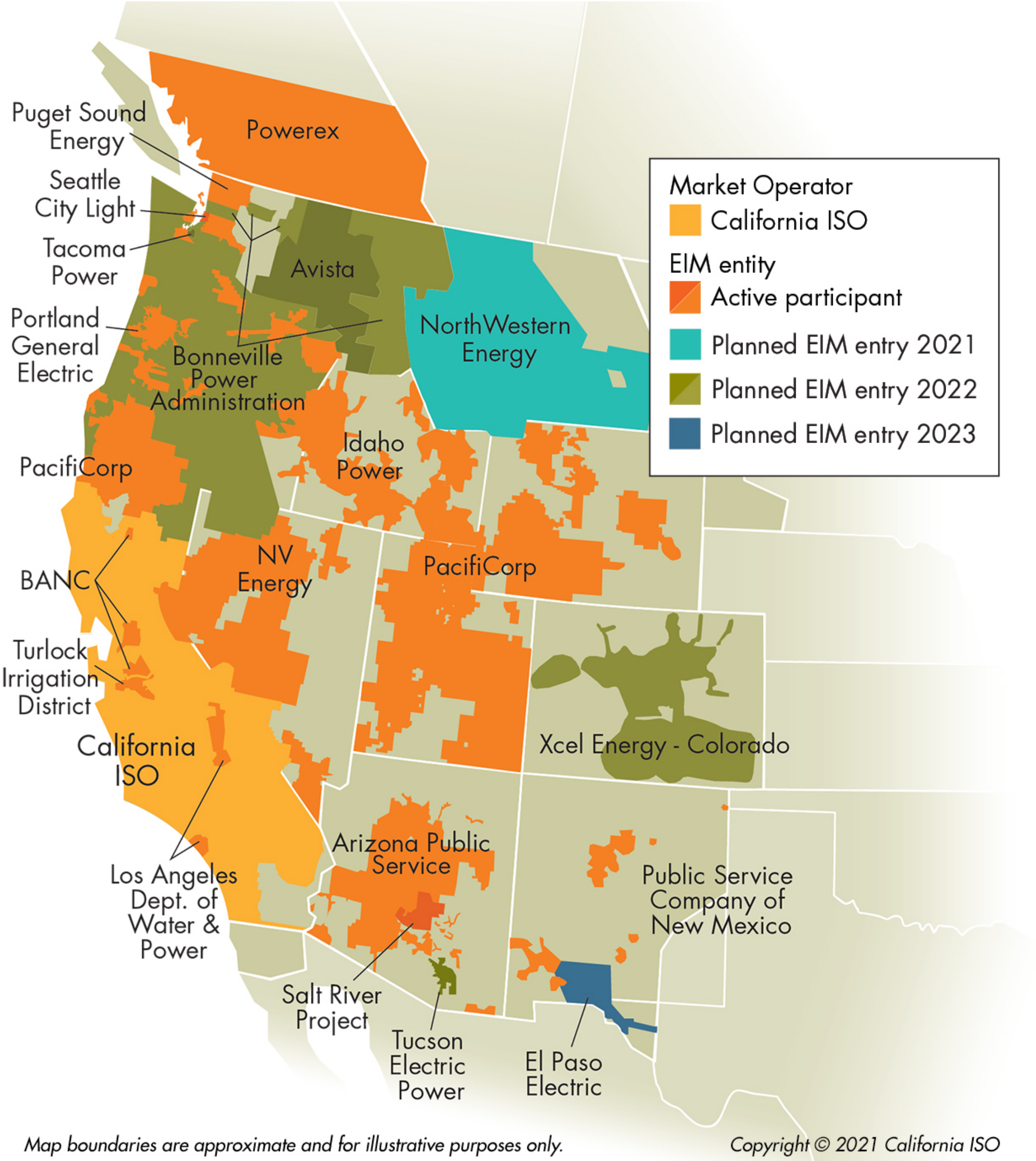


Source: NEA 2020.

Integrating systems that possess complementary characteristics has shown to increase electricity security and reduce the cost of system balancing through diversification. Since the early 20th century, California, the Pacific Northwest and Canada have exchanged surplus energy arising from differing load and generation resource availability patterns in each region. With 7 900 MW of interconnection capacity between the regions, seasonal exchanges create more balanced portfolios of resources and have deferred costly investment in generation resources.

California energy demand peaks in the summer at a time when hydropower availability is high in Canada and the Pacific Northwest due to mild temperatures and accumulation of spring snow melt runoff into storage reservoirs on the Columbia River System. In the winter, Canada and the Pacific Northwest are peaking when energy demand is relatively low in California. Puget Sound Energy of Washington State and Pacific Gas and Electric (PG&E) of California have agreed to exchange 300 MW of firm capacity and 413 000 MWh on a seasonal basis. Puget receives energy between November and February, and PG&E receives energy between June and September. This accounts for 4.4% of overall winter demand and up to 16% of hourly demand in Puget's balancing area. By formalising this exchange via contract, each party is able to secure transmission rights for the energy on a firm basis and can also credit the capacity received in its resource plans, reducing the need to procure gas peaking resources to meet expected demand.

Figure 17. Western Energy Imbalance Market membership



Notes: EIM = energy imbalance market; ISO = independent system operator. The California ISO includes the balancing areas for Pacific Gas and Electric (PG&E), Southern California Edison (SCE) and San Diego Gas and Electric (SDGE).
Source: CAISO.

The Western Energy Imbalance Market (WEIM) has extended this co-operation between balancing areas in the United States and Canada. Created by the California Independent System Operator (CAISO) in 2014 to optimise resources between CAISO and PacifiCorp, which operates in Utah, Oregon and extreme Northern California, the membership has increased to 11 today, with 8 other balancing areas planning to join by 2022. By then, the footprint will extend to 10 US states and Canada. While not a true regional ISO with fully optimised energy and ancillary services markets, the WEIM operates 15-minute and 5-minute real-time dispatch markets that allocate available transfer capability on an economic basis.

Between its inception in 2014 and the end of 2019, the participants in the WEIM estimate gross economic benefits of USD 861 million due to more efficient dispatch. Additionally, the WEIM claims environmental benefits from the reduction of renewable curtailment in the amount of 433 120 tonnes of CO_2 over the same period.

Interregional electricity planning enhances system flexibility

Electricity planning was traditionally confined to established single-utility areas. Expanding the planning perspective across multiple areas can provide greater flexibility through resource diversification at lower cost. Greater geographic diversification of generation sources will lead to less variability and uncertainty in supply. For example, wind generation spread out over large areas will see complementary infeed profiles due to spatial patterns in wind speed and thus lower variability at aggregate level. Larger integrated power systems also increase security, but are more complex from the perspective of system operation.

Greater interregional planning, progressive market integration, co-ordinated balancing action and reliance on interconnection for electricity security are emerging across various jurisdictions. Planning that expands across balancing areas or national jurisdictions can lead to more efficient use of existing generation and electricity networks and minimise the cost of expansion. Fully reaping the benefits of regional integration of electricity markets does require co-operation at both regulatory and operational levels, which can be challenging. Efficient development of cross-border infrastructure requires, for example, the development of common methodologies for cost-benefit tests. Market integration involves careful attention in adapting processes and may for some systems (and customers) result in higher wholesale prices even when the overall societal benefit is positive. Eventually, higher levels of cross-border integration in markets, operation and the electricity mix results in a critical trade-off for policy makers: the extent to which they wish to stay in control solely of national electricity security and investment, or the extent to which they accept dependence on neighbouring systems at the benefit of lower overall costs.

Many neighbouring TSOs have already started to co-ordinate power system planning to optimise the use of resources and benefit from increased flexibility.

Interregional co-ordination is evident in the European Union, South Asia, the Association of South East Asian Nations (ASEAN) and the United States (IEA, 2019f). For example, in the European Union, the European Network of Transmission System Operators for Electricity (ENTSO-E) develops ten-year network development plans at the European level based on scenario inputs, assessments and studies to identify the projects that can contribute towards meeting EU energy policy goals. Those are then used as the basis for national network development plans.

Policy makers play a central role in facilitating electricity planning

Policy, market and regulatory frameworks have a critical role in guiding electricity planning and investment decisions as the energy transition takes place and electricity systems respond to growing threats from cyber-attacks and climate change. Electricity planning should consider adopting a certain level of flexibility in its approach to adapt to fast-changing circumstances, but, more importantly, it should align itself with key high-level political objectives and agendas at national, regional and international levels. Steps for policy makers to facilitate long-term planning that strives to minimise negative economic and operational impacts include:

- Encourage integrated planning across power market segments (generation and electricity network) and economic sectors (electricity and other sectors).
- Synchronise the building of new electricity networks with VRE deployment by linking incentives for new lines to policies that support renewables investment.
- Expand electricity planning and regulatory co-ordination to cover larger geographical areas, either across balancing areas or national jurisdictions.
- Address flexibility requirements in long-term planning processes by considering the full suite of flexibility resources, including power plants, electricity networks, storage and distributed energy resources.
- Update metrics and methodologies used in long-term planning to account for increased uncertainty in load profiles and generation needs.

Monitoring the system: Adequacy assessments with large shares of renewables

Long-term adequacy assessments provide critical information for power systems in transition

The adequacy of an electricity system refers to its ability to meet all levels of load under expected conditions. The focus of adequacy analysis ranges from relatively short-term (e.g. the coming winter or summer) to long-term assessments looking decades into the future. Adequacy is a key indicator to identify potential future reliability risks. Therefore, generation and transmission adequacy are typically assessed relative to a system-wide reliability standard, which quantifies the likelihood of not meeting all electricity demand, such as loss of load expectation (LOLE).

Energy transitions have implications for the way we think about adequacy. First, transitions are linked to a change in the technology mix. Emerging technologies such as VRE sources need to be evaluated differently from conventional technologies when understanding system adequacy. This section explores that in further depth. Second, new technology – combined with new business models – is expanding the range of options for meeting adequacy requirements. Third, energy transitions affect how adequacy is provided, including the mechanisms and incentives that support capital investment. This is discussed in the following section.

Planning is a key component in delivering adequacy for all systems, even when they have different institutional arrangements. This ranges from fully centralised approaches, such as in an integrated utility context, to systems fully relying on markets or mechanisms complementing energy-only markets. Adequacy assessments play an important role in informing the planning decisions of system operators, market actors and policy makers.

They are useful tools in the case of capacity shortfalls to understand risk levels and the effects of various mitigation options. Some systems face overcapacity rather than capacity shortfalls, particularly in a number of emerging economies due to investment in generation capacity to stimulate economic growth. While in the case of overcapacity an adequacy assessment to identify security risks may seem superfluous, it can actually provide useful information to identify the level of overcapacity when clear reliability targets are set (see case study on China's power system). Adequacy assessments using scenario analysis for different future pathways also have a role in guiding investment frameworks consistent with both development and policy uncertainty. Conducting assessments on a frequent basis is another important consideration, to allow emerging developments to be constantly integrated.

Two central aspects of long-term adequacy are explored in more detail here in the context of energy transitions: the metrics applied; and the assessment methodologies. Energy transitions with a higher share of VRE require new tools and assessment approaches to ensure all resources can be represented appropriately. Regulators and system operators should consider all flexibility sources as options to satisfy adequacy in long-term planning, and accurately integrating all resources into adequacy assessments is a critical step to allow this.

The standards and metrics used in adequacy assessments need to be reviewed

While the metrics used for adequacy assessments may appear complex, most are based on a limited number of considerations. Reserve margins have for a long time been the dominant approach in the electricity sector. They refer directly to a capacity margin that can be relied upon above projected peak demand, and can account for contingencies such as outages and unavailability of variable generation. Corrections can be made to this margin to cover demand response and interconnection flows. Essentially, adequacy is about the availability of the system to supply demand. Therefore, most other metrics used describe how not all load may be supplied, with two main parameters:

- the dimension of load not served, by focusing on either magnitude (amount of energy not supplied) or frequency (how often periods with unserved load occur)
- the likelihood of unserved load, by focusing on an “average” case, or explicitly evaluating “tail risks” (high-impact, low-frequency occurrences).

Table 2. Various reliability metrics

Reliability metric	Dimension	Probability
Reserve margin	Available capacity margin (% of peak demand)	None
Expected energy not served (EENS)	ENS magnitude (MWh)	Average value
Loss of load expectation (LOLE)	ENS frequency (hr/year)	Average value
P95 (95th percentile)	ENS frequency (hr/year)	Tail risk, 1/20-year event
Loss of load probability (LOLP) for a specified ENS volume	ENS frequency (hr/year)	Probability of a specified ENS level

Note: ENS = energy not served.

These metrics all provide some information on the risk level in the system. They are all an indicator of the likelihood of lost load, and primarily give a signal as to whether the system has sufficient capacity. It is important to recognise that the analyses these metrics are based on do not provide any prediction of how often load shedding or even full blackouts will occur. They are a measure of sufficiency in the system. Typically, power systems will have emergency mechanisms in place to help manage the system under conditions of stress. Similarly, inadequacy is not the only trigger for possible load shedding. For a full reliability assessment, other types of analysis remain essential, and the resources required for providing other services such as operating reserves must be set aside in adequacy assessments to avoid double counting.

Currently, there is little standardisation between regions on metrics adopted in adequacy assessments, with many countries still using reserve margins and others progressing towards the adoption of a range of probabilistic metrics. Notably, a standardisation initiative is ongoing in Europe. Most countries using probabilistic metrics use only a single measure, with the consequence that for a single region all the major dimensions (magnitude and frequency, average and tail risk) are not typically considered. While performance in one measure is to some degree indicative of performance in other measures, the relationship between the metrics may not be straightforward and may also reveal differences across regions. Some metrics such as LOLE also have an economic relevance, allowing the cost of energy not served to be compared against the cost of new capacity.

In the context of energy transitions with a higher share of renewables and more flexibility enabled, moving from reserve margin approaches to probabilistic metrics is highly recommended – so that stakeholders can monitor metrics that encompass both the magnitude and frequency of energy not served, and consider both average and more extreme system conditions.

Many regions are moving from reserve margin approaches to metrics that consider full load and supply profiles

Adequacy assessments are routinely undertaken in large power systems today, contributing to the suite of power system planning processes that ensure a reliable system. These assessments inform system operators, policy makers and all market actors on how reliability obligations can be met.

Regions often have multiple adequacy assessments that focus on different time horizons and which can be based on different methodologies. For example, the

Electric Reliability Council of Texas (ERCOT) has both a long-term adequacy assessment that examines a ten-year time horizon, the Capacity, Demand and Reserves report, and a shorter term assessment that focuses on the coming summer or winter, the Seasonal Assessment of Resource Adequacy.

The Australian Energy Market Operator (AEMO) similarly performs the Electricity Statement of Opportunities with a ten-year horizon, as well as the Energy Adequacy Assessment Projection that focuses on supply risks, with a two-year horizon. AEMO in addition performs two Projected Adequacy of System Assessments on a continual basis, one with a one-week horizon – published every two hours – and one with a two-year horizon – published every week.

Figure 18. Adequacy assessments undertaken in various systems

Region, organisation	Study	Horizon →	Regional standard
Australian NEM, AEMO	Energy adequacy assessment projection (EEAR)	2 years	Probabilistic 0.002% EENS
	Electricity statement of opportunities (ESOO)	10 years	
Belgium, Elia	Strategic reserve volume (SRV) evaluation	3 years	Probabilistic: 3 hr LOLE, 20 hr P95
	Adequacy and flexibility study	10 years	
Texas, ERCOT	Seasonal assessemt of resource adequacy (SARA) (operational)	6 m	Deterministic: 13.75% planning reserve margin
	Capacity, demand and reserves (CDR) report	10 years	
South Africa, Eskom	Medium-term system adequacy outlook (MTSAO)	5 years	Probabilistic: EENS < 20 GWh*
EU, ENTSOE	Mid-term adequacy forecast (MAF)	10 years	No common standard (Probabilistic: considers LOLE, EENS)

* The MTSAO also considers capacity factors of peaking and high-cost-base generation.
Sources: AEMO (2020a), (2020b); Elia (2019), (2020); ENTSO-E (2020), Eskom (2020), ERCOT (2021).

Some assessments have a legal basis, such as those produced by system operators in Australia, Belgium and Texas, and by ENTSO-E and regional co-ordination centres in Europe, while others are for information purposes only. Many are required on an annual basis, while some additional assessments are performed on an ad hoc basis. Adequacy assessments can also be used to examine different development pathways or scenarios for both electricity demand and supply. Elia, for example, does this using scenarios of additional capacity in the rest of Europe in its Adequacy and Flexibility Study. Including scenario analysis in adequacy assessments is an important activity to guide investment needs and account for policy and development uncertainty. While assessments differ substantially between regions, there is a general trend towards simulation approaches with increasing levels of detail in multiple dimensions.

A move from deterministic methods, such as the planning reserve margin, towards probabilistic approaches, such as probability convolution or Monte Carlo simulations, is one way in which adequacy assessments have evolved in recent years. For example, ENTSO-E in Europe has been adopting this approach since 2016 and is still actively developing its methodologies. As energy transitions advance, older approaches become less fit for purpose, in particular because of their limitations in evaluating systems with high levels of VRE, more demand response and greater interregional dependency.

To calculate the reserve margin, every technology requires some estimate of its availability during the system peak. For conventional dispatchable technologies this estimation is based on the outage rate and can be relatively straightforward, although there is a tendency to overestimate capacity requirements when each generator's availability is considered separately. A simplistic estimate of the contribution of variable renewables to the peak will be highly inaccurate and typically results in underestimating VRE contribution and therefore conservative estimates of dispatchable capacity requirements (discussed in more detail below). This is misleading and should be avoided if a truly cost-effective and secure system is the target. Low availability of VRE during longer specific periods may be a realistic situation and requires dedicated analysis. Even sources considered dispatchable may have constraints on fuel supply and operations that deserve attention (see case study on the gas/electricity interdependency in northwestern Europe). Reserve margin approaches tend to take a simplified approach to demand, focusing on the peak period, which limits the ability to account for all demand response possibilities. The contribution of storage to reliability in a capacity margin approach is also challenging to assess, as is the impact of regional grid interconnections.

With advances in computing power and data availability, newer assessment methods provide a more accurate and detailed picture of system adequacy. This can include "probability convolution", in which the availability of different parts of the system are evaluated together. As this approach becomes highly complex for large systems, there are many advantages in moving towards Monte Carlo simulation techniques. These allow a very complex system to be sampled and assessed many times to identify the likelihood of various reliability levels and other indicators. Each sample can be a full-year scenario run of the entire system with selected load, generation and outage profiles, which provides a number of indicators including reliability-related indicators such as the energy not served or the number of hours per year with insufficient capacity. Over a number of samples, the intricate correlations between all parameters can be captured and a probabilistic view of the reliability indicators emerges.

Probabilistic simulation offers many opportunities to enhance adequacy modelling

Policy makers, regulators and system operators should consider implementing Monte Carlo-based simulations, as they open the door to wide-ranging improvements in the accuracy and comprehensiveness of the assessment. These include:

- Options for a more detailed and accurate assessment of the contribution of VRE based on site-specific parameters and correlations over larger areas.
- The ability to better represent planned and non-planned generation and transmission outages and their interactions.
- The possibility of assessing adequacy across larger regions, allowing the potential adequacy contribution and interaction of interconnected areas to be accounted for.
- Improvements in the inclusion of system reserve margins and main system contingencies.
- The opportunity to represent load variability beyond simply considering a single peak period, to account for price-responsiveness of various demand sectors, and include explicit demand response options.

In a simulation approach, generation and transmission outage statistics can be combined with a detailed representation of the physical topology of the system. Possible outcomes can be simulated for many different random patterns of outages of all technologies. This provides a more accurate answer compared to considering each piece of infrastructure separately, and also gives a picture of the statistical spread of possible outcomes.

Simulations also allow for deepening the analysis further, for example to account for time-based variation of the likelihood of generator or transmission outage based on the local weather conditions. Larger regions can also be represented, allowing the reliability contribution of interconnected areas to be understood. Different areas may be stressed under similar conditions, in which case they make limited contribution to reliability. However, as areas have a different energy mix, load profile and see a smoothing effect in aggregated VRE, interconnections will often provide increased adequacy. Operating reserve margins are another area that deserves close consideration, including dynamic reserve sizing that takes into account load, variable generation and the largest contingencies in the system.

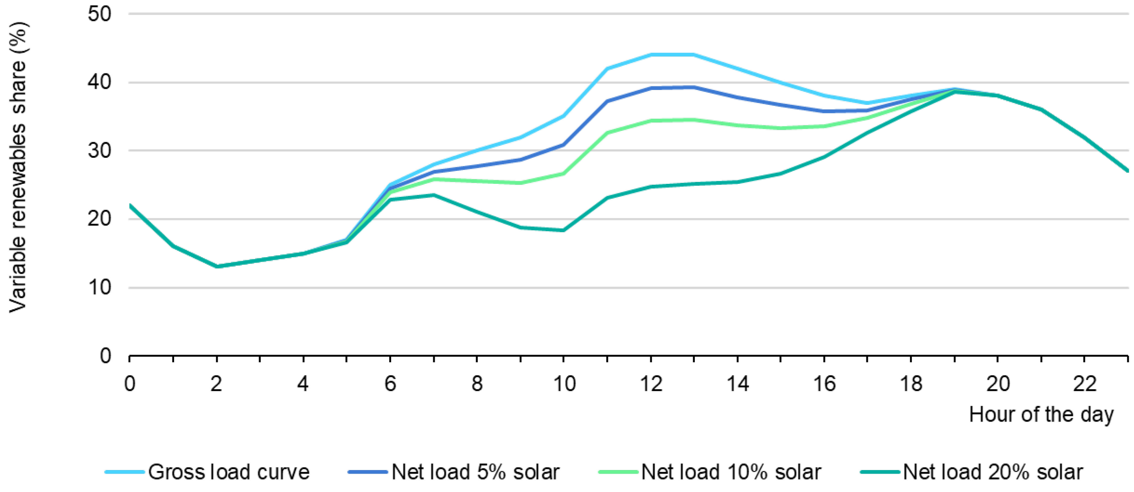
Variable renewables can contribute to system adequacy

The impact of VRE on system adequacy should not be oversimplified and requires dedicated detailed analysis. Variable renewables can, and typically do, make a contribution to generation adequacy, but the size of this contribution is far more complex to assess than for a conventional dispatchable plant, given both the variability and uncertainty inherent to VRE. It is possible to assess the capacity contribution of VRE generators as a single fixed estimate (for example, 20% of installed wind capacity is available during peak demand), such that it can be used in a traditional reserve margin assessment, however this must be calculated separately for each VRE share. If possible this approach should also take into account many years of weather data. This approach still does not provide all the benefits of a full probabilistic assessment.

For an isolated solar plant or wind farm, its contribution to the system peak load – or its "capacity value" – is first dependent on both the availability of wind or sun, and second on the time of day of the peak itself. There can be elements that drive a correlation between the system peak and variable renewable supply. For example, a summer afternoon peak driven by air-conditioning demand might tend to be complemented by the availability of solar generation, and conversely might in some regions tend to correlate with low wind availability. There is also a large random element that cannot easily be captured without considering many possible outcomes of both demand and variable renewable supply.

In addition, a layer of complexity is added by the interactive effect of multiple wind and solar plants. To understand this, it is necessary to look at the impact of wind and solar on the load that needs to be met by dispatchable generators. This impact can be analysed by means of the net load profile. This is formed by subtracting the VRE profile from the load at each time point. This highlights how VRE alters the shape of the remaining load profile that dispatchable generators or other flexibility options such as energy storage need to meet. If enough wind and solar generators have high output during the time of the absolute system peak load, the peak in the net load will move to another time. The peak contribution of new renewable generators should be assessed in relation to their contribution to this new peak.

Figure 19. Illustrative gross and net load curves with increasing solar penetration



Source: IEA own depiction of a system with midday peak.

An interesting result of this interaction is that VRE generators can have both negative and positive impacts on the capacity contribution of other VRE generators. In the case of increasing solar penetration, for example, if the net load peak is moved from the daytime to the evening, then in some regions where wind generation is stronger in the evening this will actually increase the capacity contribution of wind. Conversely, once the net load peak occurs after sunset, the contribution of additional solar PV generation to capacity requirements will fall to zero.

This complexity illustrates why the capacity value of wind and solar can only be determined in the context of a specific share of both technologies, and must be recalculated for a new scenario with different shares. This is important to ensure that the contribution of VRE is neither underestimated – leading to unnecessary investment in additional capacity – nor overestimated – leading to insufficiency and an unreliable system. In the case of hydropower, the overall contribution is highly dependent on the specific type of facility as well as water availability, which can vary strongly year by year and which can be assessed in simulation approaches to show the complete system interaction.

Demand characteristics and overall system flexibility potential need to be well understood to assess the correct system adequacy level

An important area that adequacy analysis needs to capture is the demand side. This is increasingly important as energy transitions enable increased opportunities for flexible demand through digitalisation and expanding business models.

The simplest and most common approach to the demand side in adequacy assessments takes into account only the anticipated value of the demand peak, as well as some room for error in this estimate (e.g. a high and a low demand growth case).

A next step is to study the interaction of the demand profile across a year or longer with other parts of the system (renewables availability, generator outage patterns). This is to more comprehensively assess times when the system may be under stress, and constitutes a real and important improvement for system planning.

Traditional approaches have treated electricity demand as an exogenous input that must always be managed with flexible supply, effectively assuming demand to be completely inflexible.

There is considerable scope for a more detailed analysis of the demand side, which can greatly deepen the understanding of future system needs, including:

- A better understanding of the growth trajectory for specific demand sectors, so that changes in the total demand can be anticipated more accurately, including sector coupling or potential electrification of demand sectors such as transport and heating.
- Improved tracking of self-consumption of distributed generation so that its impact on the demand profile can be anticipated.
- A better understanding of the hourly patterns of demand, down to relationships with weather and detailed end uses.
    - A better understanding of the impact of energy efficiency measures on both total and peak demands.
- Building on these steps, a more profound integration of demand response so that the system flexibility potential can go beyond traditional peak-shaving approaches.

The key limitation for extensive demand-side modelling is data availability. Simple adequacy assessments rely on historical load profiles with a projected overall load growth and possible weather-related corrections. The more a system relies on variable supply sources, the more improving the understanding of demand behaviour through data-driven insights becomes a top priority, including:

- More disaggregated demand profiles, including sectoral but also specific appliance and end-use monitoring.
- Information on appliance stocks and characteristics.
- Information on demand drivers such as GDP and per-capita income at sufficient spatial resolution.
- Price responsiveness and the full potential of explicit demand response.

As energy transitions enable more responsiveness at the grid edge, the behaviour of demand during system stress situations, and opportunities for demand to actively participate in meeting adequacy needs, need to be well understood to come to correct conclusions on overall adequacy levels.

Better understanding reliability levels in China's power system transformation: An advanced system adequacy case study

An advanced probabilistic analysis case study for China provides more in-depth insights into the methodology and possible applications of adequacy assessments for systems with a higher share of VRE and flexible resources. It highlights the key questions policy makers, regulators and utilities can face in their power system transformation. This case study builds on the regional model of China developed by the IEA in the World Energy Outlook 2017 and further developed in the extensive China Power System Transformation assessment in 2019. The case shows that in the present projections for China's electricity mix in 2035, up to about 5% of total generation capacity could be reduced without jeopardising system adequacy, albeit increasing the LOLE slightly, simply by adjusting planning reserve requirements.

The analysis identifies five main points of attention:

- Probabilistic analysis gives a much more detailed picture of the reliability of a power system than can be achieved with traditional reserve margin approaches. This allows countries to more precisely aim for their target level of reliability. Based on a 15% reserve margin approach and current policy trajectories, China would have close to 1 290 GW of fossil fuel plants currently in operation by 2035. This assessment shows that 165 GW – equivalent to around 5% of total capacity – could be eliminated without breaching a reliability target of 3 hours LOLE per year at regional level.
- While metrics such as LOLE and expected energy not served (EENS) broadly track one another, the relationship is not predictive and some differences in results will occur depending on the metric. Notably, the reliability targets set by policy makers have a practical impact as they are translated into specific infrastructure

requirements and investment plans. Thus, policy makers need to understand that small differences in target setting can result in substantial differences when assessing necessary levels of flexible capacity.

- Inter-annual variation in large and rare weather conditions has an important impact on both average and tail-end (less probable but high-impact) statistics; it is therefore essential to consider a long historical record, and equally important to anticipate changes in climate patterns to undertake informative assessments.
- Water availability is a key factor for hydro-rich regions and deserves dedicated analysis methodologies.

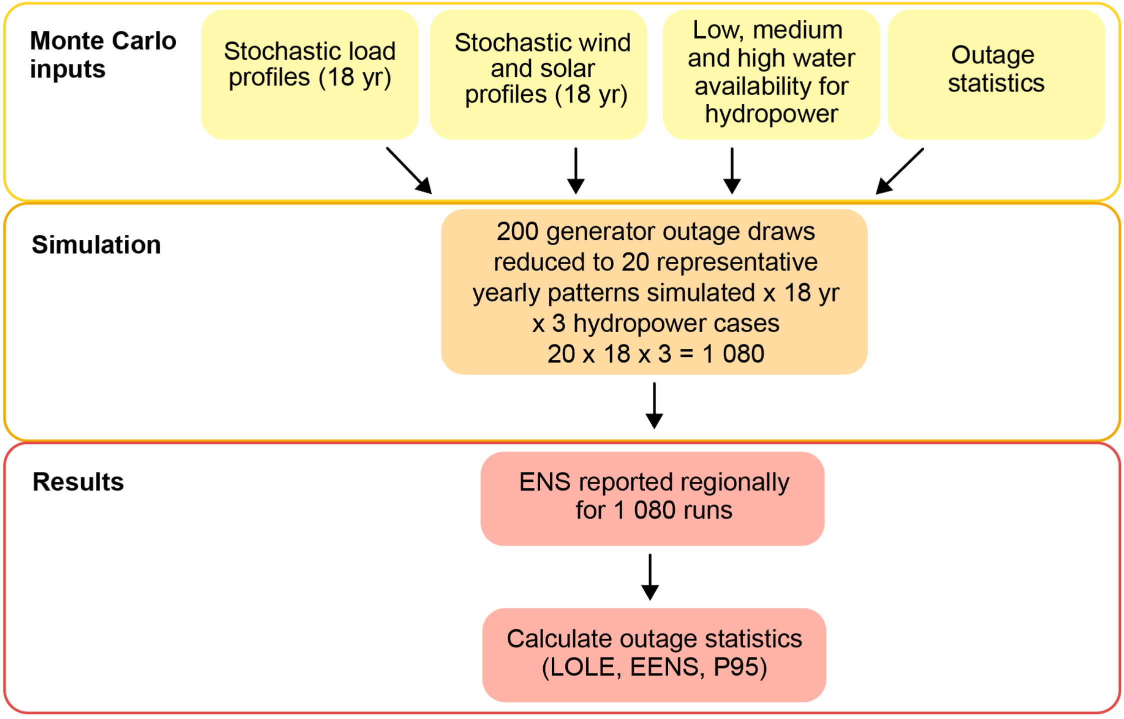
Detailed modelling results are discussed below to support and expand on these elements.

The China model includes six regions for load and generation, as well as the transmission between them, based on the World Energy Outlook 2018 projections for 2035. Demand is based on a combination of bottom-up and top-down end-use projections, enabling demand response to be modelled regionally according to regional end-use availability.

The case study considers stochastic models based on both the New Policies Scenario (NPS, now labelled the Stated Policies Scenario, STEPS) and the Sustainable Development Scenario (SDS) projections from the World Energy Outlook 2018. This enables an assessment of adequacy for different potential development pathways, but also provides an illustration of differences in the results based on the increased penetration of VRE in the Sustainable Development Scenario.

The generation of sample years for the Monte Carlo simulation is based on the key elements of a probabilistic adequacy assessment methodology: weather-based demand, solar and wind generation, hydropower availability and a large set of random outage patterns, statistically reduced for computational feasibility. For this analysis weather data of the period 2000-17 is used.

Figure 20. Schematic for China adequacy assessment showing main inputs, simulation step and outputs



Note: Reference to "generator outage draws" refers to a randomised outcome of the time and duration of outages for generators in the model.

For the first stage of analysis, LOLE is assessed for the base New Policies Scenario, which follows nationally set policies. The overall reliability of the system from an adequacy perspective is very high, with an LOLE value of < 0.1 hours per year. This is not surprising as the capacity determination is based on current reserve setting in China, which uses a relatively conservative 15% reserve margin. This is at the upper end of the standard range, which is more applicable to predominantly thermal systems, and most regions in China have even greater margins in practice.

It is important to note that the LOLE is the level of expected unserved energy from an adequacy perspective and is not a direct prediction of outages. LOLE is not directly linked with the system average interruption duration index (SAIDI) or the system average interruption frequency index (SAIFI), which measure service interruption from a customer perspective, and high reserve margins are not a substitute for addressing reliability risks directly. It is typically expected that system operators would have a range of mechanisms at their disposal to avoid outage during actual operations.

To identify capacity that could be removed from the system while still retaining a level of reliability that can be considered acceptable, an LOLE range of 2-3 hours per year for each region is then targeted. Flexible fossil capacity is removed from the system so that the highest observed LOLE in any region sits within this range. On this basis it is possible to remove 165 GW of firm capacity from the New Policies Scenario, or 5% of total capacity. This substantial capacity reduction highlights the large financial savings available using more detailed adequacy analysis and targeted reliability insights. For some regions, such as China's North West Region, the LOLE for the analysis tends to remain well below the target as there is no remaining flexible fossil capacity available for removal to bring the LOLE into the target range.

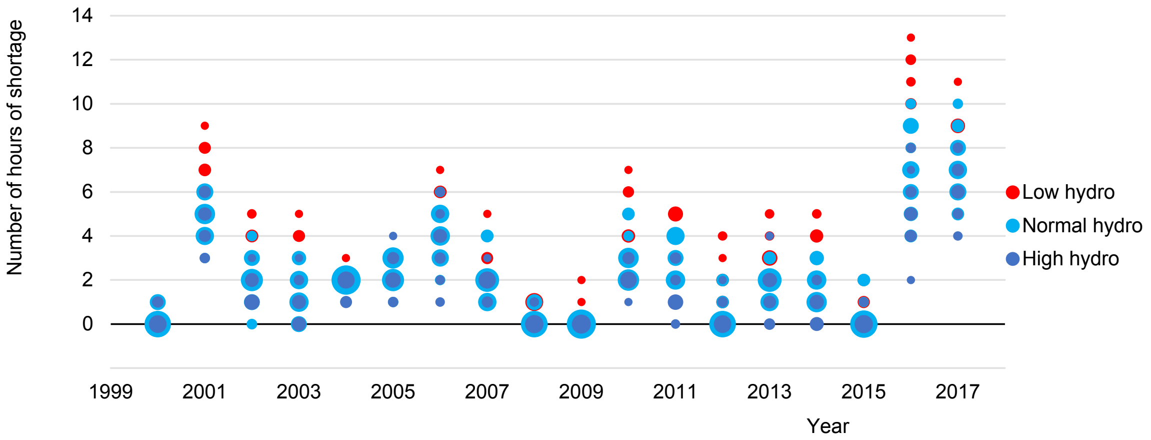
Table 3. Regional reliability under the New Policies Scenario 2035 with capacity trimming

Region	LOLE (hours)	P95 (hours)	EENS (%)
Central Region	2.5	7	0.0011
Eastern Region	2.5	8	0.0011
North Central Region	2.2	8	0.0008
North East Region	0.4	2	0.0001
North West Region	0.1	1	0.0000
Southern Region	1.8	5	0.0004
Capacity reduction	165 GW (or 5%) 127		

On both LOLE and EENS, all regions stay within reliability thresholds used in internationally – e.g. 3 hours LOLE in Europe and 0.002% EENS in Australia.

The LOLE value for each region is calculated by averaging the number of hours during which there was unserved energy, or the loss of load hours, for each of 1 080 different samples (based on the weather year, hydro condition and outage pattern). Looking instead at the individual samples of the number of hours experiencing shortages that go into this average across the different climate base years, there are substantial inter-annual variations. This illustrates how this type of analysis can capture the impacts of different weather patterns that may occur, including both the relationship with load and renewables generation. Sample weather years with a higher occurrence of challenging weather events (e.g. high cooling demand coinciding with low solar production or high heating demand during low wind periods) will have within them more samples where high LOLE values are seen. Because of the long-term variability of weather patterns, even within a relatively long time period (e.g. 18 years used in this analysis), it is challenging to ensure that the probability of unusual weather periods will be captured accurately, such as the more challenging 2016 and 2017 weather profiles, which resulted in much higher loss of load values across the modelling samples than most other years.

Figure 21. Number of hours experiencing supply shortages in each climate year and stochastic sample for China Eastern Region (NPS-based scenario in 2035 after capacity reduction)



Notes: Number of hours experiencing shortages is the number of hours in a simulated year that any loss of load is experienced. The bubble size is proportionate to the number of stochastic samples out of the 1 080 Monte Carlo runs experiencing a given number of hours of shortage.

Relatively rare, but challenging, weather patterns have a strong influence on the average metrics. For example in the case of the Eastern Region, if the year 2016 is excluded from the results the LOLE would drop from 2.5 to 2.2 hours per year. This underlines the importance of assessing as long a time frame as possible with an extensive climatic database, such as the ENTSO-E Mid-term Adequacy Forecast assessment, which uses 35 years of weather data for European analyses. Considering the impacts of climate change will be increasingly critical for adequacy assessments in the coming years, although this is outside the scope of this case study.

This inter-annual variation also highlights the impact of climate trends that need to be taken into account in longer-term adequacy assessments. If the weather influence, mainly on demand, observed in 2017 (for example) were to become substantially more common, then the LOLE for the same capacity assumptions could rise above the threshold of the standard. Variability between low, normal and high hydro scenarios also shows the sensitivity of the results to assumptions on water availability for hydro-rich regions, which will change over time. For this case study water availability is approximated using typical river estimates across a five-year time frame; however, the importance of this assumption indicates the need for a more detailed analysis across a longer time period, ideally capturing any correlations with other weather variables, to ensure a robust assessment.

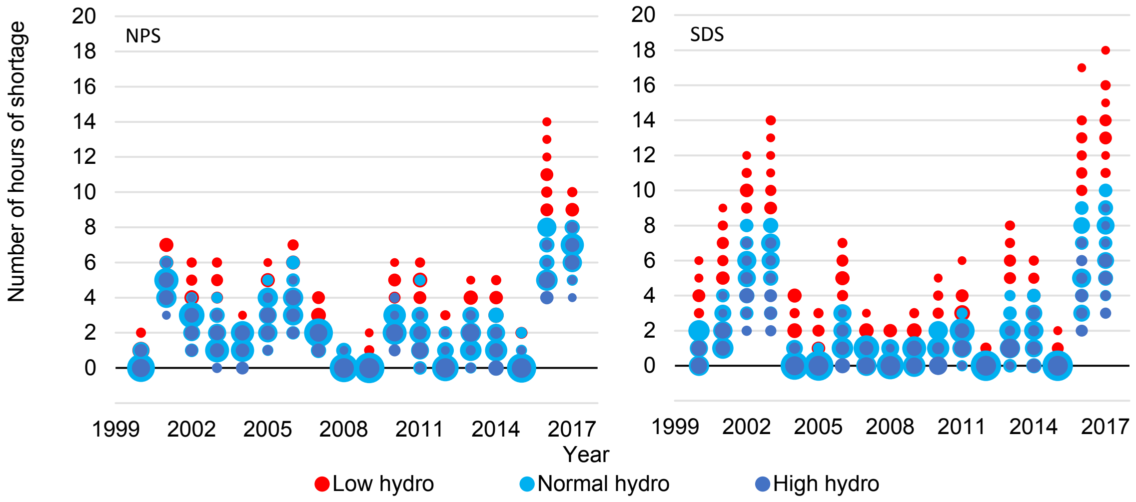
The same type of assessment is also undertaken for the Sustainable Development Scenario-based analysis, for which the base scenario also has a reserve margin of 15% and LOLE < 0.1 hours. The analysis reveals a similar potential to reduce fossil capacity by 160 GW while keeping all regions' LOLE under 3 hours per year, equivalent to around 4% of total capacity.

Table 4. Regional reliability under the Sustainable Development Scenario 2035 with capacity trimming

Region	LOLE (hours)	P95 (hours)	EENS (%)
Central Region	2.4	8	0.0010
Eastern Region	2.3	9	0.0009
North Central Region	2.6	9	0.0010
North East Region	2.2	9	0.0020
North West Region	0.0	0	0.0000
Southern Region	2.1	8	0.0007
Capacity reduction	160 GW (or 4%) 127		

It is also clear from these results that the LOLE and EENS metrics will not give the same results for a given threshold. For example, the Central and Eastern Regions have a slightly higher LOLE than the North East Region, while their EENS is around half that of the North East Region. As with the New Policies Scenario, LOLE and EENS for all regions remain within internationally applied standards.

Figure 22. Annual distribution of number of hours experiencing shortages by sample in the China Central Region for the New Policies Scenario and Sustainable Development Scenario



Taking the Central Region, which has very similar LOLE values of 2.5 hours in the New Policies Scenario and 2.4 hours in the Sustainable Development Scenario, we can also compare the distribution of the number of hours experiencing shortages across samples and see that a wider spread is observed in the Sustainable Development Scenario for this slightly higher level of reliability. Thus while the mean value, i.e. LOLE, is almost the same, the more extreme samples (e.g. a low hydro year with a more difficult climate year and outage pattern) have a larger number of hours with supply shortfall than that seen in the New Policies Scenario.

Gas deliverability is increasingly relevant to electricity security

Economic factors and decarbonisation policies are expected to drive reductions in the oil, coal and nuclear power fleets in various jurisdictions around the world, creating the conditions for an expanded role for gas-fired plants in the provision of adequacy, energy and flexibility in their power systems.

This would increase the reliance of the power sector on natural gas generation, a fuel that for many countries has to be transported over long distances, sometimes crossing multiple jurisdictions, and whose logistics will become more and more relevant for security of supply in the power sector. In this context, analysing the future reliability of gas supply is increasingly important, and also complex, due to the interplay with the electricity sector and the increasing influence of stochastic weather effects in both systems. This report presents the main aspects that policy makers should consider to assess electricity security risks from the increased reliance on natural gas for power. To illustrate them we present a case study in the European context, showing that the amount of gas that can be delivered in stress periods will become a critical parameter for electricity security assessments in the coming years.

The energy transition in Europe sees large increases in variable renewables, while coal and nuclear sharply decline

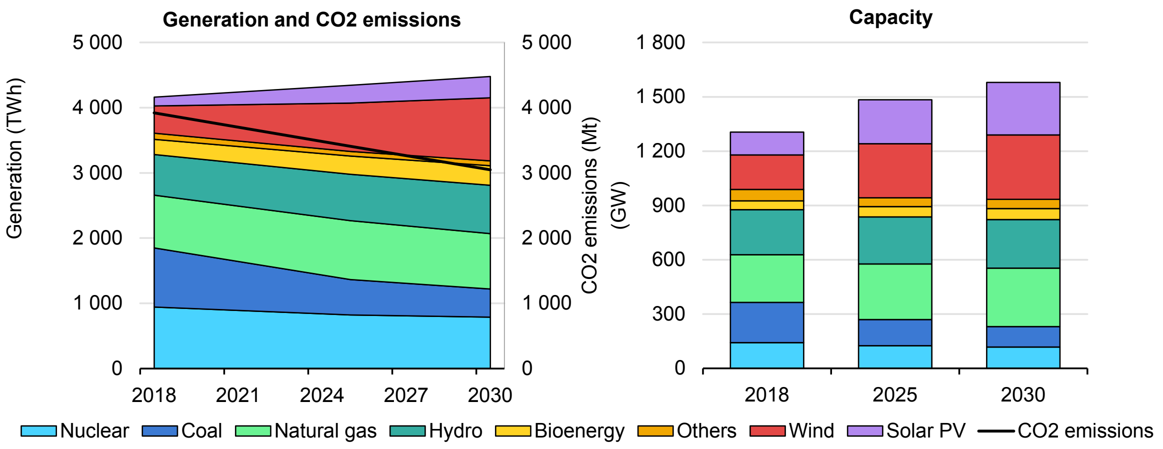
The European energy system is undergoing a profound change towards fewer emissions and a higher share of renewables. This transition is being led by the electricity sector, which is destined to expand its role in satisfying energy demand while at the same time significantly reducing CO_2 emissions – by around 30% to 2030 compared to 2018 levels.[3] The impacts of the Covid-19 outbreak in the form of changes in behavioural patterns and economic activity, and government measures to stimulate the economy, might accelerate this change.

Developments on the demand side are being shaped in particular by new applications such as electric vehicles (see the IEA Global EV Outlook), increased electric heating and efficiency measures. The total net effect of these changes is expected to increase total electricity demand by around 4%.

The supply side is seeing a shift towards variable renewables, with solar PV and wind set to double their current annual supply. At the same time, coal-fired and nuclear-based generation are due to face large drops compared to today's levels (down by 64% and 23% respectively) – with their combined capacity being almost halved. Only gas-fired plants are expected to see relatively stable generation levels, and a capacity increase of around 20% to compensate for the retiring nuclear and coal capacity.

[3] Values for 2030 based on the Stated Policies Scenario (STEPS) of the IEA World Energy Outlook 2019.

Figure 23. Electricity generation, emissions and capacity in Europe under the Stated Policies Scenario (2018-2030)



Given these changes in the electricity system, especially with respect to new sources of consumption and a profound shift in the mix of flexible plants, the role of gas-fired power plants in providing flexibility to a power system with a high share of variable renewables will become increasingly important, creating a more intimate link between security of electricity supply and natural gas deliverability.

Northwest Europe[4] is a region where a growing reliance on gas-fired power plants – and hence on the deliverability of the gas system – is expected to be very noticeable in the coming years. The following pages shed light on security of supply issues in this context and demonstrate how a stress test can be applied to evaluate the security of electricity supply given its interdependency with the natural gas system.

Stress testing should account for multiple interrelated factors that affect the performance of critical system elements

This section presents a framework of stress factors that can be applied to the electricity and gas systems to assess their resilience in meeting high demand under challenging supply conditions.

The analysis can be subdivided into three steps: first, the construction of an extreme weather scenario; second, the assessment of how the electricity system

[4] Defined in this context as: Belgium, France, Germany, the Netherlands and the United Kingdom.

behaves in the context of this scenario; and third, the response of the gas supply system to the requested gas demand.

The weather will have an increasing impact on both the electricity and gas systems as renewable electricity sources increase their share in the energy mix. In particular, two aspects are of uttermost importance: first, with the rising significance of electricity-based heating and cooling, electricity demand will be increasingly influenced by high or low temperatures; and second, as a result of the rising importance of variable renewables, available generation capacity will vary greatly with the availability of wind and sunshine.

In addition to weather impacts, several additional aspects have to be considered when analysing security of supply in the electricity system. Dispatchable power plants such as nuclear, coal- and gas-fired plants can be unavailable for several reasons, including scheduled and unexpected outages, fuel supply issues (e.g. low water levels in rivers used for coal transport) or operational restrictions (e.g. limits on cooling water temperature for nuclear plants). Electricity storage is characterised by physical limits on the storable electricity (ranging from a few seconds to several days of continued generation). This is why the proper way to assess a stress situation is to look not only at the absolute peak scenario, but also at several consecutive days of high residual demand. This similarly applies to demand flexibility, which is expected to significantly grow in the future. Demand is expected to be more flexible with respect to the time of consumption, but shifting is likely to be limited in time, since (for instance) the user of an electric vehicle might avoid charging during peak times, but might not be able to avoid charging at all for a whole week.

In interconnected electricity systems, import capacity is typically able to contribute to some extent to security of supply. Especially in systems with a large share of variable renewables, connecting large territories can help smooth weather-dependent generation – and even guarantee a certain minimum availability. This is why attention has to be paid to carefully selecting the appropriate geographical scope.

In many electricity systems the energy transition goes hand in hand with reduced fuel diversity and less flexible electricity generation, increasing the reliance of the power system on gas-fired power plants during peak demand with simultaneously low wind and solar generation. This makes it necessary to assess the deliverability of the gas system, especially when considering declining domestic gas production and potential gas storage site closures.

Gas markets in the northern hemisphere have typically strong seasonal demand patterns, with consumption during the winter season driven up by the heating requirements of the residential and commercial sectors and indirectly for power generation to satisfy electric heating demand. In addition, shorter-term variability of demand (volatility) is usually present, driven in the heating season by variations in temperature and across the year by the fluctuating needs of the power sector.

The flexibility needed to satisfy natural gas demand can be met through a combination of upstream and downstream flexibility tools available to market participants.

Production flexibility associated with gas fields close to demand centres can be an important source of additional gas supply deliverable in a relatively timely manner. However, it is important to note that only a limited number of gas fields have the geological conditions that enable them to provide a “production swing” large enough to meet seasonal demand variations. Moreover, production flexibility tends to decline once gas fields enter their depletion phase.

Available spare capacity in import pipelines gives market participants the option to exercise upward nomination rights and/or purchase gas volumes above the level of their contractual terms. However, import pipelines are typically run at maximum capacity through the heating season and hence can provide only limited additional upward flexibility.

Spare regasification capacity can provide additional importation capability. However, as highlighted in the IEA Global Gas Security Review 2018, in practice it takes at least several days to bring an additional LNG cargo to the market where it is required.

Midstream interconnectivity with adjacent markets allows participants to purchase additional gas volumes from their respective hubs. One potential limitation is that geographically close market areas usually experience similar weather (and hence market) conditions, for example limiting the volume of freely available gas volumes during simultaneous cold snaps. In addition, gas volumes “stored” within pipelines can provide short-term flexibility to meet intraday variations in demand. However, this storage is not available in periods when a pipeline delivers gas at its maximum technical capacity. Gas storage sites located close to consumption centres can provide timely response to arising flexibility requirements on the demand side. The withdrawal rate of gas storage sites tends to decline with the decreasing level of working gas in stock, given the lowering pressure level within the storage space. Depending on the individual system, this puts additional stress on the gas system at the end of the heating season, when the withdrawal rates at gas storage sites are typically 25% lower than their maximum technical withdrawal capacity.

Extreme weather events present both supply and demand challenges

As the electricity systems in our focus region, consisting of Belgium, France, Germany, the Netherlands and the United Kingdom, are characterised by peak demand in winter (driven by electric and gas-fired heating demand), our test case focuses on an especially cold low-wind period in late winter – a weather phenomenon known as cold anticyclones. The scenario is based on weather data from the past 30 years.

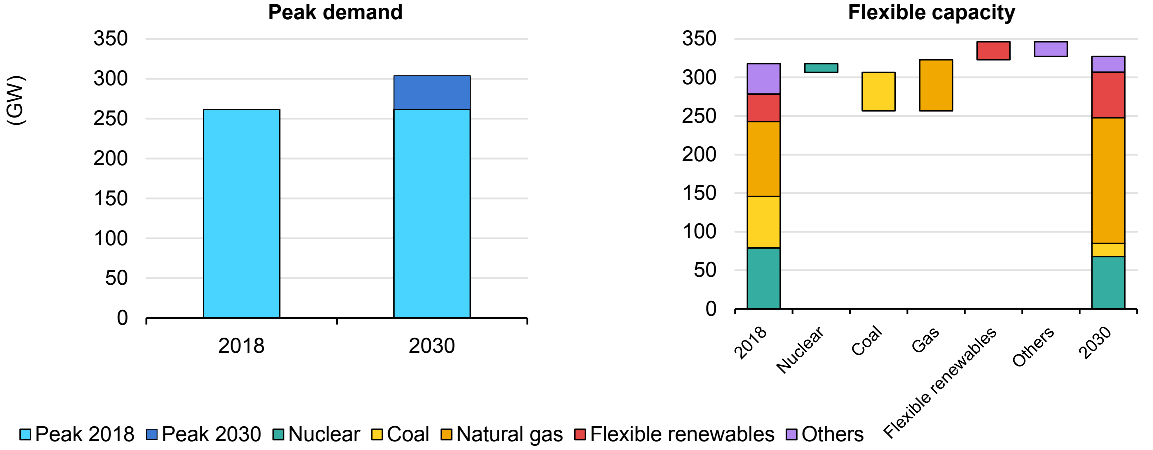
The resulting weather scenario consists of a week with temperatures on average of -3°C across all five countries (-5°C on the coldest day). The average wind availability over the whole week is 18% (benefiting from the higher availability of offshore wind plants), and falls to 6% on the day with the lowest wind speeds.

An intimate link exists between natural gas deliverability and electricity supply security

To assess electricity security during a one-week stress situation, we first use a European electricity model that embeds northwest Europe in the broader European context and thereby accounts for potential imports and exports between countries. By optimising electricity supply and demand, flexibility options such as storage and demand flexibility are taken into consideration.

In this scenario, electricity demand in northwest Europe follows a similar trend to the European Union: growing demand for electric heating and vehicles on the one hand, and declining demand due to improving energy efficiency in electrified end uses on the other. In addition to overall demand, the impact of these changes on peak demand are important, especially so for the purpose of analysing security of supply. Due to a combination of exceptionally cold temperatures and the changing demand structure, we observe approximately 15% higher peak electricity demand in 2030 than in 2018.

Figure 24. Northwest Europe's peak electricity demand and flexible generation fleet in 2018 and 2030



This increase in peak demand will have to be met by a changing power plant fleet. Due to the choice of the stress situation, most variable generation capacity (wind and solar PV) will not be available – which leaves it to flexible capacity sources to serve the majority of demand.

Almost half of northwest Europe's nuclear, coal- and lignite-fired power generation capacity is set to close by 2030, largely driven by phase-out policies. With the exception of Germany, which is set to reduce its coal-fired capacity to about one-third of today's level, the region will be coal-free. Due to rising CO_2 emission prices in recent years, some older and inefficient plants might leave the market for economic reasons, hastening the exit of these resources.

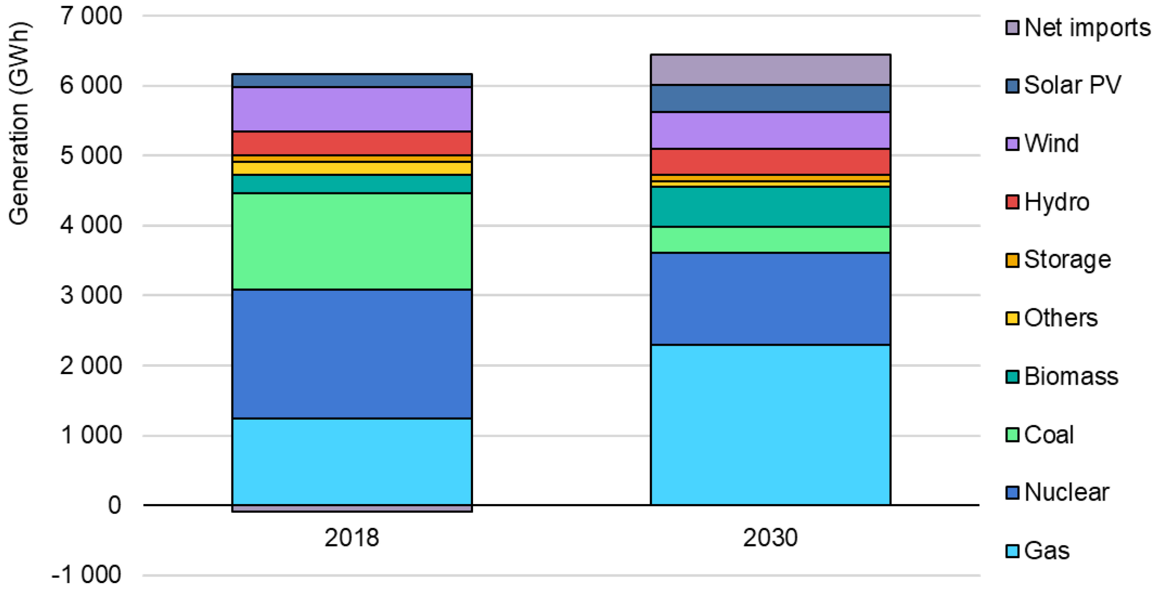
The expected role of nuclear power in northwest Europe in 2030 is strongly influenced by policy. A phase-out of nuclear plants in Germany will be completed in 2022, and a planned phase-out in Belgium is to be completed by 2025. Although France has one reactor under construction, it has established limits to this technology's share of the power mix and a dozen units are expected to close by 2035. While the United Kingdom has two units under construction, and is considering more, all but one of its currently operating nuclear plants are slated to close by 2030.

As a consequence of power plant closures, northwest Europe's flexible power generation capacity could be dominated by gas-fired plants, accounting for about 50% of the thermal generation fleet by 2030 (vs 30% in 2018).

During a stress situation with high electricity demand and low generation from variable renewables, the reliance on gas-fired power plants to serve peak demand will be significant, making natural gas deliverability a critical component of electricity supply security.

Our simulation suggests that during a cold snap occurring in 2030, gas-fired generation would become crucial to meeting peak demand – accounting for 36% of total annual supply and generating about 85% more than during the peak in 2018, with a corresponding increase in demand for natural gas.

Figure 25. Northwest Europe's peak supply days in 2018 and 2030



Average gas demand decreases, but supply margins remain tight in peak demand periods

The growing role of gas-fired power plants in providing flexible power supply, and meeting peak demand in an increasingly VRE-dominated power system, creates the need to assess gas deliverability from an electricity supply security point of view.

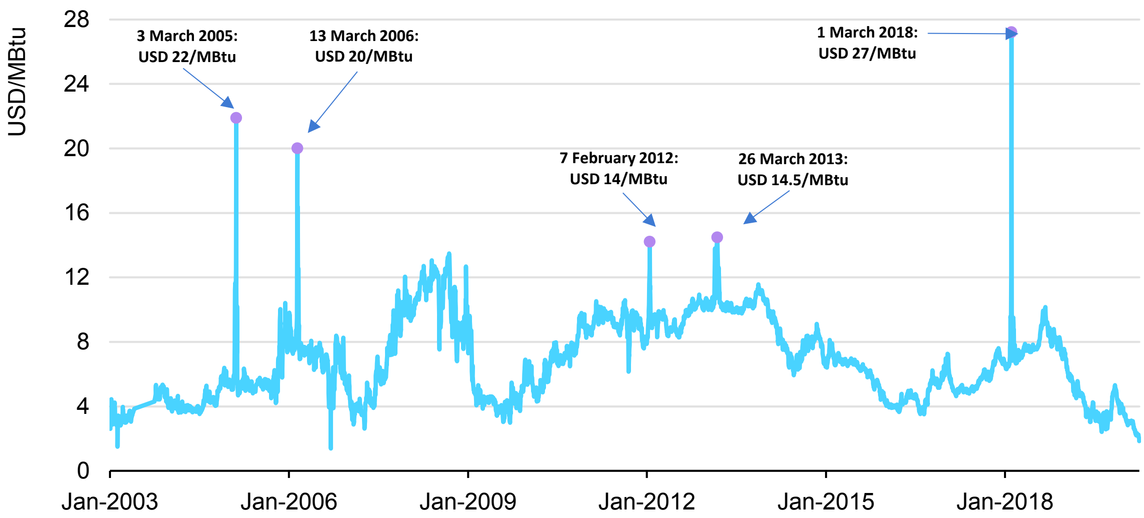
Northwest European gas consumption shows a strong seasonal pattern. Demand during peak days can be twice as high as the annual average, primarily driven by space heating requirements in the residential and commercial sectors.

The reliability of the northwest European gas system has been historically underpinned by upstream flexibility, a high degree of midstream interconnectivity and vast underground gas storage – allowing it to meet demand requirements characterised by a high degree of seasonal and short-term fluctuations.

Nevertheless, the system has come under stress on multiple occasions during the past decade, resulting in sharp price spikes on the Dutch Title Transfer Facility (TTF) – northwest Europe's leading benchmark gas hub.

Price spikes have occurred almost exclusively during the period of late winter and early spring. In all cases, they have been triggered by late cold snaps at a time when the deliverability of the gas system has been lower than during the official winter season (December-February), primarily due to lower storage withdrawal capacity.

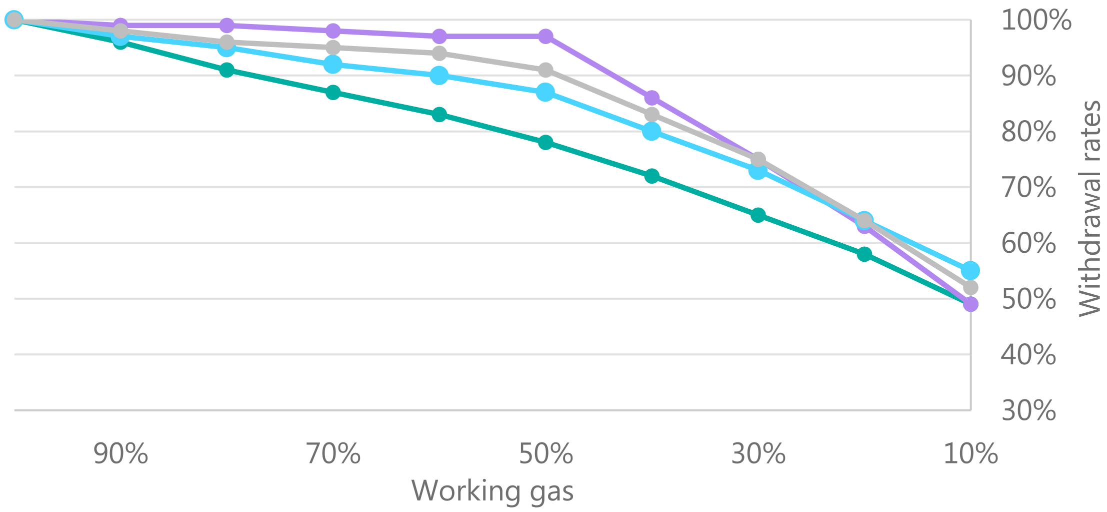
Figure 26. Price spikes on the TTF (2003-2020)



Note: MBtu = million British thermal units.

Storage sites have typically low fill levels by late winter/early spring, which – as a result of lower pressure levels in the reservoir – naturally decreases withdrawal rates. In northwest Europe, storage sites with inventory levels at 30% would see their withdrawal rates decline by approximately one-quarter compared to their maximum technical withdrawal capacity. This lower deliverability during periods of unexpected and unseasonably high demand can result in tight market conditions driving up intraday and day-ahead gas prices.

Figure 27. Typical storage delivery rates in northwest European gas markets

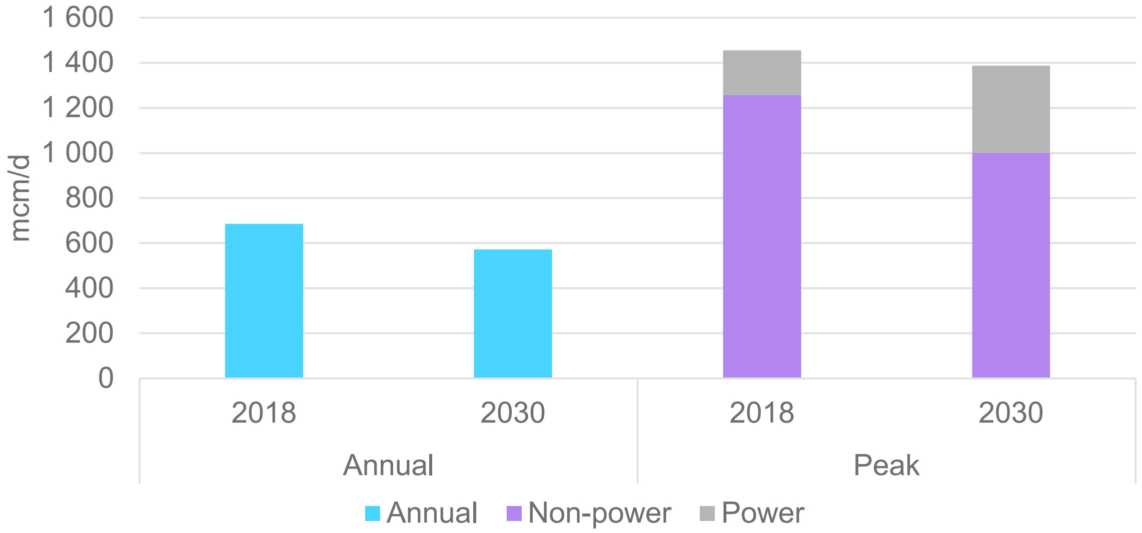


Natural gas demand is expected to decrease in all sectors over the next decade. In the residential and commercial sectors the improving energy efficiency of the building stock will push down overall space heating requirements, while the gradual electrification of heating will further diminish demand for natural gas directly used for space heating purposes. Gas-to-power demand is expected to decline with the strong build-up of renewables, while industrial demand is set to decrease due to efficiency gains in industrial processes and gradual fuel switching. Altogether, natural gas demand is expected to fall by over 15% by 2030 compared to 2018.

Despite declining overall gas consumption, our simulations show that the "peakiness" of gas demand will remain. In fact, the peak/average demand ratio will increase by close to 15% compared to 2018, limiting the decline in peak demand to less than 5%. In absolute terms this translates into a decline of almost 70 mcm/d.

However, the structure of peak demand will change and will be increasingly driven by the power sector. Our simulations indicate that gas-to-power demand would almost double on a peak day in 2030 compared to 2018 under the weather conditions described in our stress scenario. This is due to two factors. First, the intensifying electrification of space heating indirectly supports demand for gas-to-power through the heating season. Second, gas-fired power generation will be increasingly responsive to the variations in electricity supply from renewable energy sources, serving as a back-up in periods of low VRE output.

Figure 28. Northwest Europe's natural gas demand, 2018 vs 2030

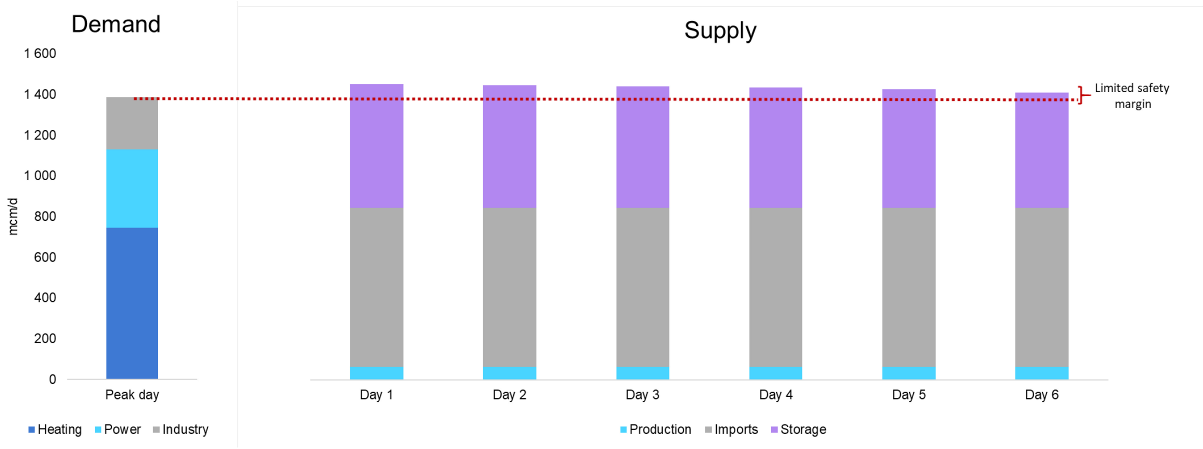


To assess the gas deliverability of the northwest European gas system on a peak day, we made the following assumptions:

- Considering current infrastructure developments and the upstream production capabilities of key gas exporters to northwest Europe, we expect that the region's peak import capability will remain similar to today's levels.
- Northwest Europe's indigenous gas production is expected to fall by over 60%, from just above 70 bcm in 2019 to below 25 bcm by 2030. This will be primarily driven by the closure of the Groningen field in the Netherlands by mid-2022 and by gradually depleting fields in the North Sea. As a consequence, daily domestic output is expected to fall by over 120 mcm/d to below 70 mcm/d and will be primarily supplied by mature, less-flexible offshore fields.
- A number of storage sites are expected to close by 2030. As a consequence of the phase-out of L-gas consumption in Belgium, northern France and northwest Germany by 2030, the *raison d'être* of L-gas storage sites located in France and Germany will disappear (in particular for those not connected to the Dutch gas network). Moreover, the expiry and non-renewal of service contracts puts at risk certain storage sites. In total, we expect that the closure of storage sites could reduce the maximum deliverability of northwest European storage by ~15% (or 150 mcm/d).
- Imports from other parts of Europe, including reverse flows from Italy and Eastern Europe, are typically limited (close to zero) during cold spells that are simultaneously present in adjacent European energy markets.

Under our stress scenario – a late cold snap combined with low wind speeds – the gas system would still be able to meet demand requirements, but with a safety margin barely exceeding peak demand levels.

Figure 29. Northwest European gas demand vs deliverability of the system



Creating effective markets to encourage investment in security

Historically most electricity systems were designed, built and operated with thermal and hydro generation as the primary supply technology options located at the transmission level, with end-user demand being inflexible. The energy transition brings a wider range of technological options, including VRE, various types of demand-side response and battery storage, which can be located both at the transmission and distribution levels.

Maintaining security of supply while achieving power system decarbonisation at an affordable cost will require system operators to take advantage of the technical capabilities that different technologies can provide to meet the adequacy, flexibility and clean energy needs of the system. **The energy transition is likely to place stress on existing market frameworks, reducing some existing income streams and creating the need to ensure that all required services are compensated.**

This means that policy makers and regulators should review their current market and investment frameworks. In an evolving system the services needed to maintain security of supply will also evolve, and more than ever it is necessary to ensure that power sector participants are given the right incentives to supply important services – including not only clean energy, but also net peak load capacity, flexibility and reserves.

The electricity sector is not a typical commodity market, but is driven by significant regulatory and administrative intervention

Power sector investment decisions are driven by significant regulatory or administrative intervention, as important technical characteristics differentiate the sector from others that do not rely on these rules to attract investment.

The first important characteristic is that, at present, consumers cannot express their willingness to pay for electricity in real time by responding to price signals, for example reducing or shifting their consumption when supply is scarce. Reliability criteria are used as a substitute for this lack of responsiveness. Enhancements in power sector modelling have improved reliability criteria compared to the simple metrics used in the past, but demand is still largely modelled as insensitive to market prices. When supply does not match demand plus the required reserve margin, system operators reduce the consumption of customers through load shedding to preserve the balance and avoid system-wide outages.

The second important characteristic is that electricity is provided through a shared asset, the grid, which requires instantaneous matching of supply and demand. This means that all users face substantial social costs beyond their own non-consumption in the case of a partial or full collapse of the system. This is contrary to other industries, where lack of supply of a good to a single customer does not jeopardise its availability to the rest.

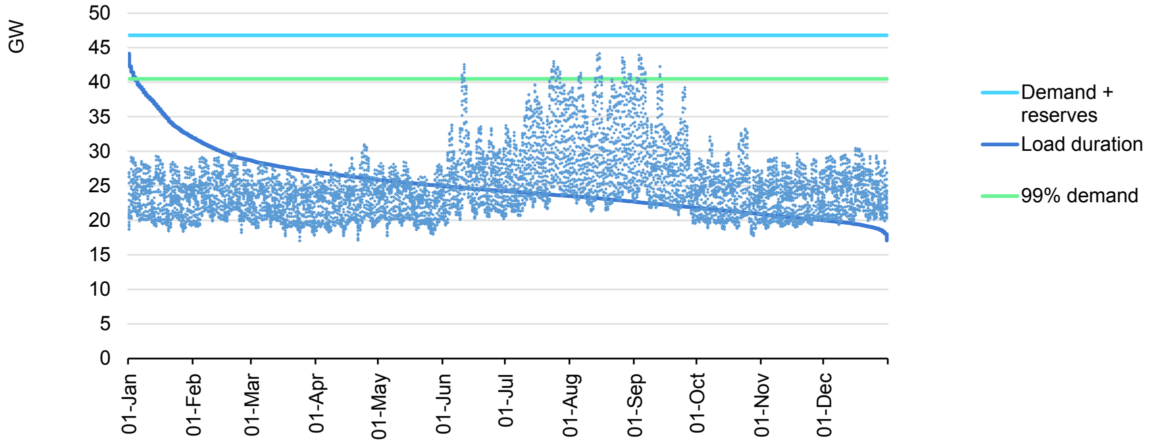
These characteristics have been long recognised in electricity systems, regardless of whether they are centrally planned or operated through market mechanisms. That has led to complex institutional frameworks to attract investment and set rules to ensure that resources are available to the system when needed. Secure and affordable energy transitions call for the constant review of these investment and market frameworks to check they are fit for purpose.

The power system needs to be designed to handle stress situations that arise only a few hours per year

Any system, regardless of whether it is centrally planned or market driven, needs to identify a cost-effective portfolio of resources to fulfil energy, system adequacy and flexibility needs, either via a planning authority's analysis or through price signals. Market-based systems require an extra step to ensure that necessary investments materialise and that existing plants remain available to the system for adequacy purposes despite reduced revenues from the energy market. Most investments make economic sense due to load growth, fuel switching or efficiency gains of new entrants and new technologies. During energy transitions policy makers, planners and system operators will need to consider not only the economics of the plants providing the bulk of the energy, but also the economic implications of dispatchable resources serving a more diversified array of needs. These can range from dispatch to cover peak load and net peak load in power systems with high VRE shares to providing reserves during stress periods.

Policy makers need to fully appreciate that a generator's ability to provide electricity for a few critical hours per year is a different service compared to providing electricity during normal times, as the main contribution is to ensure system adequacy. Traditionally this was measured as the generation capacity being available during peak demand periods. For example, the California Independent System Operator (CAISO) has a footprint of annual demand that reached 45 GW for just a few hours in summer 2019. In this situation 2.6 GW (equivalent to 5.7% of peak demand) was required to generate for fewer than 100 hours per year, and an additional 4 GW (8.8%) was required to provide the 6% of reserves needed year-round to handle contingencies. Therefore, at least 14.6% of available resources rely mostly on revenue streams linked to their availability and generation during a very small number of critical hours.

Figure 30. CAISO hourly electricity demand, 2019



Source: CAISO (2020).

Regulators create reliability standards in an attempt to balance security of supply and system costs

The amount of dispatchable resources available to the system is a product of the market and investment framework in place, and relevant authorities such as regulators require this to be adjusted accordingly to achieve a certain level of security of supply. But to do that effectively, policy makers need to define an acceptable level of interruptions for that system.

Although most of these standards are set in parameters that are non-intuitive, it is relevant for system operators and planners to express this tangibly, if society and policy makers are to understand the consequence of their choices.

Many standards are inherited from the era of vertically integrated systems, with heuristic rules defining an acceptable level of interruptions based largely on deterministic approaches (as discussed in the previous section on system adequacy assessments). Other standards are a product of the explicit balancing of the costs associated with power interruptions and the benefits of reducing investment in capacity or retiring underutilised resources. Such historical standards will consider load shedding or curtailed consumption as the most expensive action that can be taken as the last resort to maintain security of the system. It essentially also puts an acceptable price on system adequacy.

Great Britain, for example, uses an LOLE of three hours per year as a reliability standard. This is based on the average of nine possible cases, considering various estimates of value of lost load (VoLL) values and costs of new entry. A reliability standard of a three hours/year LOLE is also applied in many other systems.

Table 5. Definition of LOLE reliability standard in Great Britain based on the average of nine scenarios centred on VoLL and the cost of new entry

Equilibrium reliability standard in LOLE (hrs/yr)		Cost of new entry (GBP/kW)		
		Low GBP 31.89	Central GBP 47.18	High GBP 66.21
VoLL (GBP/MWh)	35 490	0.90	1.33	1.87
	16 940	1.88	**2.78**	3.91
	10 290	3.10	4.59	6.43

Source: DECC (2013).

An essential input to such an approach is a reasonably accurate estimate of the VoLL in the system, as well as the cost of new entry, the potential for supply and demand flexibility and especially the willingness-to-pay of consumers, which is evolving radically.

It is important for policy makers to understand the implications of the inherited reliability standards, their risks, costs and benefits, and to assess if these parameters, essential for the correct functioning of the power sector, are consistent with the cost to society of a certain level of supply interruptions. In this process, systems operators should make clear what the implications of the existing standards are, how the assessment accounts for average or extreme conditions, and what are the consequences of alternatives regarding the likelihood and depth of potential interruptions. While these reliability standards deserve to be set on clear economic considerations, it is helpful when they are expressed in meaningful terms, for instance, “the set standard allows that 10 000 households might lose supply for 2.5 hours every 10 years”.

As electricity becomes more important in our economies, such reliability standards applied in the electricity system need to be reviewed so that they remain consistent with the cost to society of power interruptions, as well as taking into account the societal acceptability of outages with respect to both average and extreme events.

Choosing a reliability metric that targets the optimal level of supply security

Every system situation can be measured using various reliability metrics. However, there is no simple one-to-one translation between the various metrics. Each system has different demand and supply characteristics, which means that it is not always the same metric that becomes the critical or relevant factor for system planning.

Analysis of the ERCOT system in Texas demonstrates the implications of different levels of reliability expressed through different metrics. In this exercise, different levels of dispatchable resources are associated with different levels of security of supply. Loss of load expected events (LOLEV), also known as loss of load frequency, is defined as the number of events in which system load is not served in a given time period. For example an LOLEV of 0.1 (1 interruption event on average every 10 years) is achieved with a 13.5% reserve margin. At the same time a loss of load hours (LOLH) standard of one day every ten years (2.4 hours per year) results in a reserve margin of around 9%. This shows how one (event) in 10 years planning standards are much more conservative in this specific case. It is important to be aware that relationships between metrics observed in this dataset will not hold for all systems, and thus one metric cannot be used to systematically predict another.

Table 6. The relationship between different reliability metrics in the ERCOT system and associated system reserve margins

Reserve margin (%)	Total annual loss of load			Average outage event		
	LOLEV (events/yr)	LOLH (hour/yr)	EUE (MWh)	Duration (hours)	Energy lost (MWh)	Depth (MW)
6%	2.33	8.35	17 015	3.59	7 315	2 038
7%	1.68	5.81	11 263	3.46	6 714	1 938
8%	1.18	3.95	7 198	3.34	6 086	1 824
9%	0.81	2.61	4 426	3.21	5 444	1 698
10%	0.54	1.67	2 610	3.08	4 805	1 562
11%	0.35	1.03	1 468	2.94	4 182	1 421
12%	0.22	0.61	778	2.80	3 571	1 277
13%	0.13	0.33	374	2.61	2 919	1 118
14%	0.07	0.16	148	2.34	2 117	903
15%	0.03	0.07	48	2.09	1 409	673
16%	0.02	0.05	80	3.40	5 295	1 558

Note: EUE = expected unserved energy. This is the same as the expected energy not served (EENS) described in the previous section.
Source: ERCOT (2018).

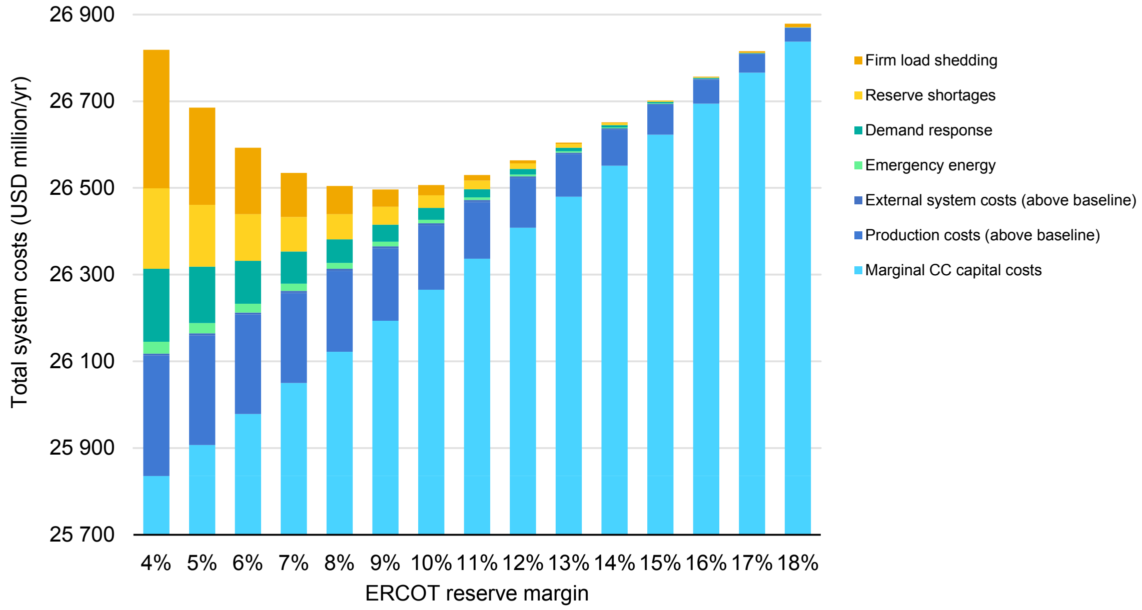
"One event in ten years" standards require larger reserve margins than those for "one day in ten years". The reliability standard chosen should be consistent with society's preferred level of reliability.

In the choice of the right metric, special attention should also be given to the duration and the amount of energy not supplied due to these interruptions in the context of the country characteristics. For instance, even if there are few events, the impact of longer interruptions could be unacceptable for a country with highly electrified space heating that regularly observes peak demand during winter, since longer interruptions in extreme weather situations could entail safety risks.

The next step is to understand the costs and benefits associated with these standards, also taking into account their level of social acceptance. Higher reserve margins come with higher costs, as more resources are paid to stay online, even if they are seldom used. But there is a benefit to be recognised, as they provide a certain level of insurance. Even when some reliability levels can only be sustained with capacity at very low utilisation rates, the costs of overcapacity to society are often compensated by the large economic value of avoiding interruptions. Furthermore, other costs can be reduced and compensated, specifically the additional fixed costs linked to a higher level of planning reserves. This means that beyond an optimal level, the net cost of additional capacity in the system increases very slowly as reduced interruptions bring additional benefits.

But this happens only up to a point. For instance, in the analysis done in 2018 for the case of ERCOT, adding capacity to the point where there are no more load shedding events would require a reserve margin of 12%. This is only 3% percentage points higher than the overall economic optimal reserve level at 9%, and slightly below the reserve margin of 13.5% that would be enforced by a “one event in ten years” planning standard.

Figure 31. ERCOT case of total system costs versus planning reserve margins



Notes: Total system costs include a baseline of total system costs that do not change across reserve margins.
Source: ERCOT (2018).

Markets need to be designed to bring the level of investment needed to meet the reliability standard

Once the reliability standard is defined, the investment and market framework needs to be carefully designed in such a way that the necessary investment materialises for the secure real-time operation of the system. Traditionally, systems would bring in dispatchable thermal and hydro capacity to cover peak demand, reserves and some level of flexibility. In systems with large shares of VRE, more digitalised solutions and increasing distributed energy resources, it becomes more important – but also more complex – to assess what is an efficient portfolio of flexible technologies.

Efficient use of resources will result in plants with the highest variable costs only being used during stress hours. This implies that the revenues needed to cover the fixed costs of keeping existing plants online or bring investment in new plants should principally come during these limited hours of operation, with possible additional compensation based on their availability. The mechanisms designed to achieve this vary according to authority and regulator decisions and the risk borne by investors.

Many types of investment frameworks exist across jurisdictions, but they can generally be categorised into three types:

a) Centrally planned systems. These exist in jurisdictions where vertically integrated utilities perform all activities across the electricity value chain, including generation, system operation, transmission, distribution and retail. Central planning also exists in jurisdictions where various regulated utilities share resources through a wholesale market ("gross pool") to optimise the short-run costs of electricity supply. Consumers cover the full investment risk. Reserve margins are defined directly according to the chosen reliability standard.

b) "Energy-only" markets exist in many liberalised markets where revenues come only from selling energy and ancillary services. In some market designs, such as Texas and Australia's NEM, prices are set at the estimated VoLL during the hours of scarcity, i.e. when the system operator is not able to procure appropriate levels of operating reserves to ensure security of supply. In these markets, the total amount of available resources is a result of individual investment choices made by market participants reacting to incentives and parameters defined by regulators, and not directly by an authority or a planner.

c) Energy and capacity markets combine energy revenues with capacity products, which ensure a separate stable signal is given for investment in operating reserves. The main objective generally is to recognise that resources provide some insurance or hedge against load shedding even if they are not dispatched. While there has been strong debate for decades in the electricity sector about whether energy-only markets can deliver security of supply, in many cases the pragmatic reality is that a capacity mechanism should be seen as insurance at extra cost that policy makers can choose to take out against load shedding. Capacity markets are more common in wholesale markets where price caps are implemented well below the VoLL level, using the argument that price caps protect vulnerable customers against price volatility, as well as in cases where a missing money problem is identified for new investment. Capacity markets can also be centralised by region, or leave load-serving entities to procure resources on a bilateral basis, with decentralised mechanisms providing advantages in accounting for demand-side flexibility.

This simplified classification hides a wide variety of designs and implementation strategies, which are beyond the scope of this report. Readers are referred to the IEA reports Repowering Markets and Status of Power System Transformation 2018 and 2019 for further context. Nevertheless, within each of these three categories, most designs share a similar division between decisions that are taken administratively or by regulators and those that are left to market forces.

Table 7. Administrative and regulatory decisions in various adequacy mechanisms

Administrative/regulatory decision	Centrally planned	Energy + centralised capacity market	Energy + decentralised capacity market	Energy only
Reliability standard	●	●	●	●
Energy prices in periods of stress		●	●	●
Level of operating reserves	●	●	●	●
Peak demand forecast	●	●	●	
Defining technologies capable of delivering the product	●	●	●	
Product definition	●	●	●	
Amount of capacity to be procured/capacity demand curve	●	●		
Technology/fuel	●			
Location	●			
Size	●			

Note: Within market frameworks administrative decisions may have differing attributes for specific aspects of the market and characterisations here are illustrative of typical cases only.

A main point of attention for policy makers in any of the three types is to ensure all flexibility in the system is recognised and can provide system services in an economically sensible manner to minimise overall system costs. In a liberalised market, whenever a decision is taken to add capacity mechanisms, careful consideration needs to be given to all possible cost-related impacts, the true verifiable reliability improvements and the potential for infrastructure lock-in (i.e. committing the system to long-lived assets that may only be required for a short time).

Over time many systems worldwide have evolved from a centrally planned system to some level of unbundling and market implementation. The question of whether to resort to an energy-only market or to complement with capacity revenues remains a challenge in many jurisdictions. Energy transitions with particularly large shares of VRE often result in a net load curve that shows fewer hours of high demand to be met by non-VRE sources and lower average net load demand overall. While this may signal pressure on the economic viability of thermal plants, some systems address this by providing sufficient levels of price transparency and scarcity pricing, and incentivising various types of flexibility.

Market implementation across NERC regions shows how different systems deliver different margins

So that policy makers and regulators can assess the effects of potential changes to market and investment frameworks to bring a desired level of security of supply, it is useful to see the performance of different regimes in other jurisdictions.

Table 8. Adequacy approaches across NERC regions with range of reserve margins shown by type of market design

	Anticipated reserve margin	Reference reserve margin
Energy-only markets		
ERCOT (Texas)	8.5%	13.75% (non-binding)
Alberta	25.8%	10%
	Reserve margin surplus (-5.25%, +15.8%)	
Decentralised capacity markets		
MISO (Midwest USA)	19.3%	16.8%
CAISO/MX (California, Baja California)	22.4%	12%
	Reserve margin surplus (2.5%, 10.4%)	
Centralised capacity market/long-term contracting		
PJM (Northeast USA)	29%	15.9%
New England	30.7%	18.3%
New York	24.8%	15%
Ontario	28.6%	14.9%
	Reserve margin surplus range (12.4%, 13.8%)	
Centrally planned		
Florida Reliability Coordinating Council	24.9%	15%
MRO Sask Power	13.4	11%
SERC (Southeast USA)	28%	13.15%
SPP (Southwest USA)	31.8%	12%
Northwest Power Pool (Northwest USA)	31%	20%
	Reserve margin surplus range (2.4%, 19.8%)	

Source: North American Electric Reliability Corporation (2019).

The review of reliability approaches in the jurisdictions that are part of the North American Electric Reliability Corporation (NERC) provides a unique opportunity to compare how different market designs and reliability standards interact. Within the NERC region, systems generally fulfil the reliability standards chosen by policy makers, no matter if they have centrally planned systems, capacity markets or energy-only markets. The ERCOT market in Texas stands out as having opted for a reference target instead of a binding reliability standard and at the same time for having the lowest margin in North America. Assessments of this market point to a margin slightly higher than the economic optimal level, although well below the more conservative one event in ten years standard.

North America's experience also shows that centrally planned systems with some type of capacity payment, in particular those who rely on centralised procurement, tend to have higher excess capacity, as do systems with heavily contracted long-term schemes and vertically integrated utilities. The extra margin overall is between 6% and 12% above levels that would be sufficient to satisfy conservative reliability standards (one event in ten years). Overcapacity is not exclusive to systems with capacity markets or that are centrally planned. Alberta, the other North American energy-only market, also presents a significant degree of overcapacity at 25.8%, which is 15.8% above the reference margin.

Although every system has managed to comply with the reliability standards defined by policy makers, different mechanisms to ensure adequacy have different advantages and drawbacks. Implementation shows that there is no perfect system, and that the strengths of each of them should ideally be incorporated in the designs to ensure security of supply. The following sections describe briefly the strengths and weaknesses of energy-only markets and systems relying on capacity markets, to guide policy makers who need to continuously reconsider their optimal market design strategy in the face of energy transitions.

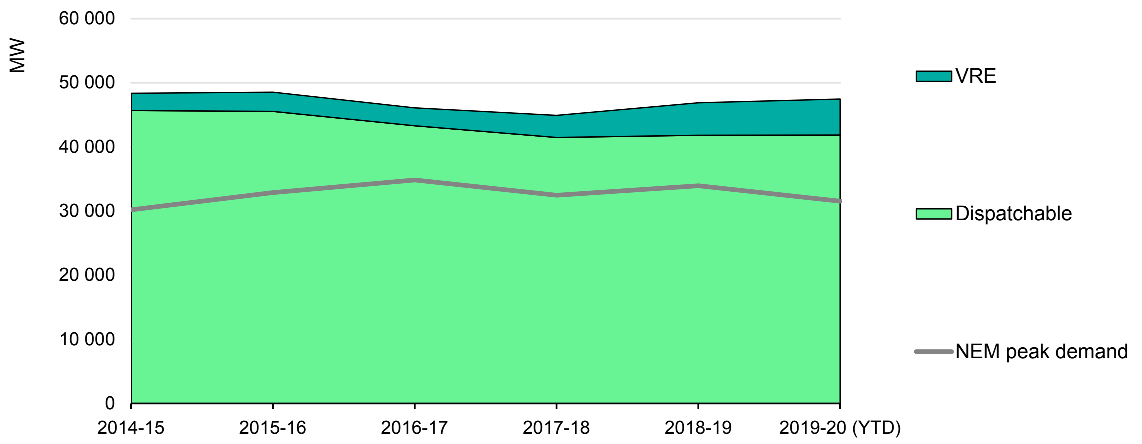
Overall, energy transitions put pressure on every market framework, either by reducing the revenues that thermal units need to recover fixed costs, or by making it harder to assess which technologies are actually contributing to adequacy. Those developing market designs will benefit from incorporating the strengths of other existing market designs into their framework.

Australia's NEM shows how an energy-only market provides adequacy with increasing shares of variable renewables

The "missing money problem" refers to the idea that in an energy-only market, only the operating or variable costs of the marginal sources of generation are directly compensated, leaving investment costs unaccounted for, or "missing". The National Electricity Market (NEM) in Australia provides a good example of the strengths and weakness of an energy-only market. Australia created the NEM in

1998 with an innovative market design where all the revenues in the spot market are from the provision of energy and reserves. The incentives to both keep old capacity in the system with a very low utilisation factor and attract new investment come from spikes in spot prices during hours when the system operator is not able to procure enough reserves to address contingencies. These spikes, by design, are allowed to rise to a maximum price that is itself set as an estimate of the VoLL (AUD 14 000/MWh), which is many orders of magnitude higher than the variable cost of energy and which creates rents needed to cover the non-energy-related costs of the plants.

Figure 32. Australia's NEM peak demand and generation capacity (2014-2020)



Source: IEA using data from AEMO and AER.

This market design was implemented in a system that had clear targets and certificate mechanisms to bring renewable generation to 20% by 2020. This model successfully handled the transition over many years from administrative decision-making on investment to a market design with one of the lowest levels of administrative and regulatory intervention worldwide.

The Australian NEM market design demonstrated its advantages in having wholesale energy prices that reflect the state of stress of the system. This allowed every resource in the system to support overall adequacy.

It pushes the demand forecast implications onto market participants, retailers and generators, creating incentives for retailers to avoid being uncontracted in advance in case of a hike in demand.

It rewards VRE and storage sources for their actual contribution to meeting peak demand, without any assumption on their "capacity contribution" in advance and without any need for complicated market product definitions and qualification criteria.

It creates high incentives for all resources to react to situations of stress, including demand flexibility. When AEMO triggers the market price ceiling of AUD 14 000/MWh, between 0.5% and 2.1% of overall demand is removed. This may seem a fairly small share of demand, but it is still a relatively large proportion of the 6% of reserves that are required to re-establish the secure operation of the system.

Table 9. AEMO estimates of demand response (MW) activated at various price triggers, 2016-2017

Price trigger	NSW	QLD	SA	TAS	VIC Not summer	VIC Summer	Total
AUD 300/MWh	38.3	27.3	15.4	4.9	76.7	76.7	162.6
AUD 500/MWh	50.2	27.9	16.6	4.9	79.0	79.0	178.6
AUD 1 000/MWh	53.2	28.6	17.2	4.9	81.5	81.5	185.4
AUD 7 500/MWh	61.0	82.6	88.1	15.2	85.0	141.8	388.7
Market price ceiling (AUD 14 000/MWh)	248.5	147.5	120.2	43.0	85.0	141.8	701.0

Notes: NSW = New South Wales; QLD = Queensland; SA = South Australia; TAS = Tasmania; VIC = Victoria.
Source: AEMO (2016).

Despite having ensured adequacy for over 20 years, the NEM's design as an energy-only market has been subject to criticism. It is claimed that the market design does not provide enough revenues to keep timely investment flowing. The market is said to have a "missing money" problem, meaning that revenues in the spot market do not justify investment, even when the system needs it.

Specific events created pressure to review the investment model. These include:

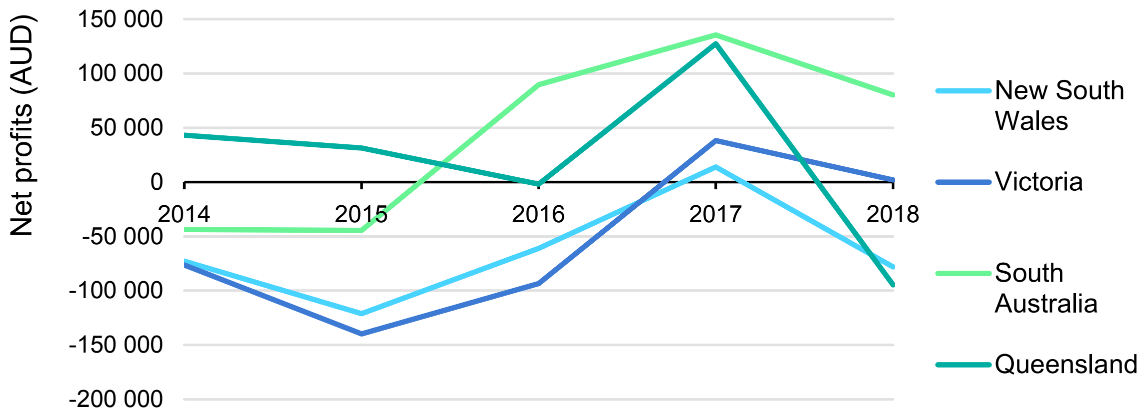
- The retirement of the Northern and Playford coal plants in 2016, whose 786 MW covered 21% of South Australia's dispatchable generation. At that moment, Northern station had been in operation 31 years, while Playford had for 50 years.
- The retirement of the Hazelwood coal plant in 2017, after 46 years in service, covering 14.7% of the dispatchable capacity in Victoria.

- A blackout event in South Australia in 2016, triggered by extreme weather and aggravated by inappropriate interconnection standards, which attracted attention to the whole investment and regulatory framework, including the overall market design.

We undertook an exercise to illustrate the underlying investment incentives in the Australian framework. Analysis shows the net profits from a hypothetical generator: one open-cycle gas turbine (OCGT), which would traditionally be a "peaker" plant, in each of the Australian states. The net profits of the plant are the revenues from the energy market minus all the variable and fixed costs of the plant, including investment and finance in the period 2014-18, with actual gas and electricity wholesale prices for each region.

This hypothetical generator would have lost money or be close to break-even in three out of four states every year, with the exception of 2017, in spite of wholesale market prices doubling from 2014 to 2018.

Figure 33. Net revenues for a hypothetical OCGT plant in each of Australia's NEM four states over the course of 2014-2018



Such losses should not necessarily be interpreted as a flaw in the market design. They can also be a signal that the system has sufficient supply, and no new investment is required. When the system has enough resources to manage peak capacity, losses to potential investors are a signal to market participants to delay investment. In fact, overall, the NEM had enough dispatchable capacity to cope with peak demand during that period.

The specific case of South Australia shows how the Australia NEM energy-only design reacts to scarcity, and at the same time provides many insights into how future electricity systems may evolve during energy transitions. In 2016 the remaining coal-fired power capacity in South Australia, representing 18% of the dispatchable capacity in the state, was retired. In spite of a coincident hike in gas prices, the reduction in supply due to this retirement resulted in increased wholesale prices, and improved the profitability in the market. South Australia could be seen as a test lab for the future of many power systems with a high VRE share and expecting retirement of thermal plants. In this case, the retirement of coal plants effectively restored profitability from the end of 2016 onwards.

This increase in profitability eventually attracted investment, as in the third quarter of 2019 a 210 MW gas-fired facility commenced operations in South Australia, alone increasing by 7% the dispatchable capacity of the state. This investment would ideally have happened in sufficient time to substitute for coal retirements, avoiding unnecessarily long periods of scarcity. Such a long period of scarcity and lost profits for generators represent a net economic cost to the system as a whole. The delay may be explained by the uncertainty investors face, exacerbated by lead times for construction and permitting, which can be very long.

This hypothetical supplier operating pattern shows how dispatchable plants might operate in an environment with high VRE capacity. The hypothetical plant in South Australia shows a very different operating pattern in 2014 compared to 2018. In 2014 it would have been “in the money” for 77.5 % of the hours of the year, while in 2018 this reduced to 65%. The rest of the time prices would have been lower than the variable costs, or even negative. Besides this reduced number of operating hours, a larger share of the profits would have been earned in a smaller number of hours, since the share of total profits made in the 100 hours with the highest demand increased from 32% in 2014 to 38% in 2018.

Energy transitions will intensify this trend, with dispatchable resources earning most of the revenue needed to cover their fixed costs during a smaller number of hours. While the NEM experience shows this can work, the trade-off for policy makers is whether they want investors to be exposed to a higher risk, potentially delaying investment, or whether they prefer predefined certainty on the system margin, and whether customers will accept long periods of high electricity prices. In the case of the NEM, supplementary measures have been adopted to ensure an acceptable level of reliability – the Retailer Reliability Obligation was adopted in 2019, requiring companies to hold contracts or invest directly in generation or demand response to support reliability.

Capacity products can close the revenue gap and provide investment signals and certainty

Policy makers, regulators and system operators need to assess whether their current market design remains compatible with the desired level of security of supply as the share of VRE increases. The example of South Australia shows how thermal plants will depend on a reduced number of operation hours to recover their fixed costs. The example of ERCOT shows how even market designs where extremely high electricity prices are considered acceptable may not be compatible with the most strict or conservative reliability standards, such as one event in ten years.

If policy makers believe that tighter standards better reflect the cost to society of supply interruptions, relevant authorities such as regulators should change the market and investment framework to tackle the associated missing money problem. This problem arises when the large amount of dispatchable resources needed to fulfil such reliability standards means that prices rarely increase above variable costs, leaving insufficient revenues to pay for the fixed costs of new entrants.

The creation of a capacity product is one of the mechanisms that has been implemented to overcome this and to guarantee that the electricity system will have enough resources to cope with extreme episodes. They have the advantage of providing a tool to ensure a minimum amount of dispatchable resources in the system and, when well designed, provide a market-based mechanism to minimise the cost of fulfilling a certain level of security of supply.

Although there is a long list of definitions of a “capacity product”, these payments are mostly linked to the amount of capacity that a resource provides, or is expected to provide, during an event of distress, which is often the peak demand of the system.

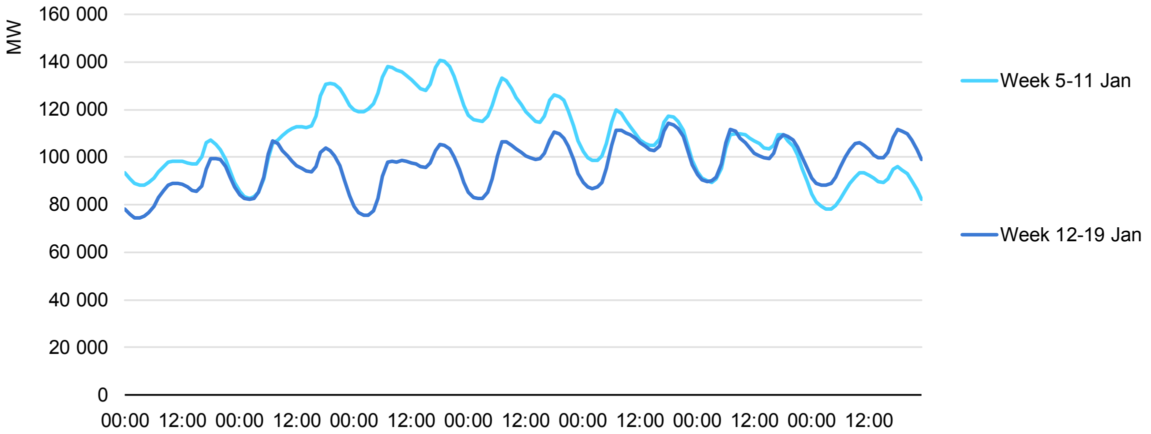
In practice, capacity payments have the objective of complementing energy revenues to a point where marginal resources (peakers) are able to recover both their variable costs and the costs associated with maintaining their availability all year. In most markets where the system operator procures the capacity, auctions engaging generators and other resources take place three to four years in advance of the delivery year, in order to give investors time to build and refurbish new facilities, or to bring mothballed plants back into production. For an ageing plant nearing decommissioning, this period means that once the plant no longer clears a capacity auction, it will still have to meet the obligations it entered into in previous auctions to deliver capacity before retiring. This gives time for new plants to be built and thus provides a mechanism for an orderly substitution of plants without the associated hikes in prices.

While the advantages and trade-offs may seem straightforward, a divergence in the design of capacity markets shows the practical difficulties that regulators and system operators find in designing rules capable of providing the right investment and generation/consumption incentives for their specific situation.

The past is not always a good predictor of the future: Lessons from the 2014 polar vortex in PJM's capacity market

A good example of the difficulties in designing a proper capacity market is provided by an earlier approach adopted by PJM, which is the largest market in the United States and whose capacity product is considered a pillar in the functioning of the wholesale market. The capacity market design in PJM was tested during the initial days of January 2014, when the northeast of the United States experienced an exceptional meteorological phenomenon. A polar vortex brought extreme cold weather and increased the demand for electricity in an unexpected way across the PJM region. At maximum demand, load reached 140 GW on 7 January at 6 pm, which was 33% higher than under normal conditions.

Figure 34. PJM hourly demand in January 2014



Source: IEA analysis with PJM data.

During that event, the system had 20 hours of alerts due to distress in the system, in spite of the 177.5 GW committed in the capacity market, which should have been sufficient to face the maximum demand of 140 GW plus operating reserves.

This lack of performance caused a careful review of the capacity market design. Unfortunately, the flaws found in the PJM market are still common to many capacity markets around the world and are likely to intensify during energy transitions in which most sources cannot be considered dispatchable 100% of the time due to weather conditions, fuel supply issues and other reliability constraints.

As in most systems, PJM held an auction in which it is defined in advance which resources will need to be available or generating during any stress period. In reality, whether a plant is generating or available depends on a series of factors that are not fully known until the time of delivery, such as outages in the case of thermal plants or the sun and wind conditions in the case of VRE.

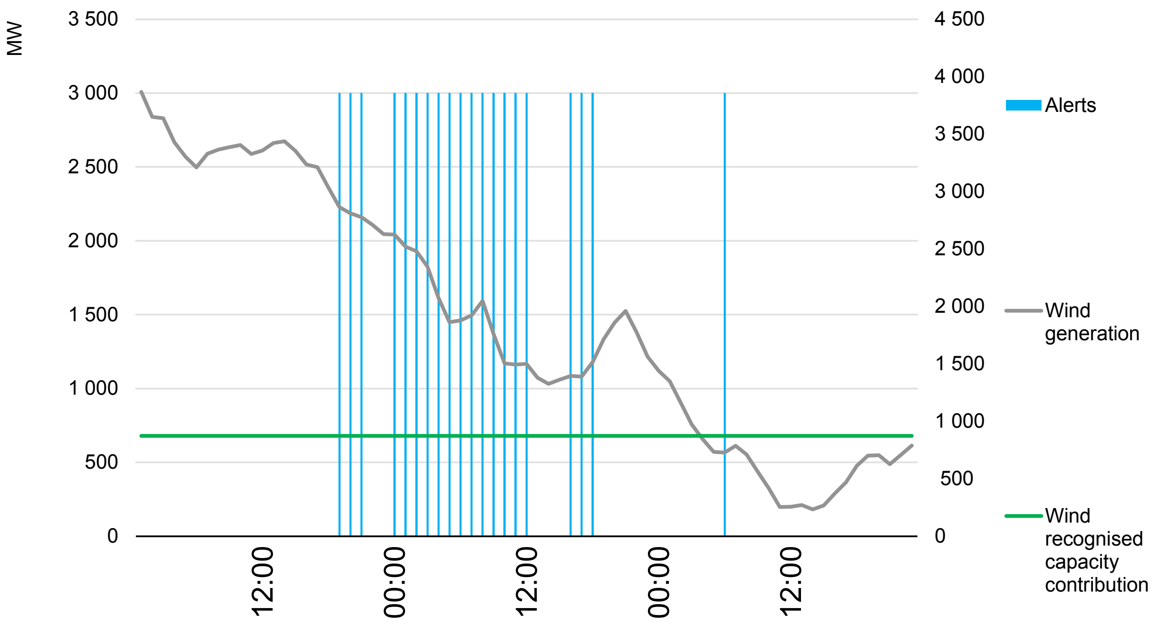
In traditional thermal and reservoir hydro-based systems, the relevant point in time to make this assessment of the necessary capacity and the availability of providers would have been during the system peak hours. In the case of PJM this period is in summer. However, as the polar vortex event showed, contingencies can arise at any moment in the year.

The capacity that a resource is able to provide is forecasted using an average performance of the asset at peak demand. However, stress periods are unlikely to have the same characteristics as peak demand, as they can appear at random moments throughout the year. Furthermore, historical average performance may not match with what happens during a stress period. During the polar vortex, 22% of the 177 457 MW of resources committed in the capacity market were in forced outage, which is well above the historical average of 7%. The extreme weather impacted fleet availability either by affecting the equipment itself (42% of the outages) or by fuel shortage (24%) often linked to lack of firm capacity in gas pipelines. This means that these resources, mostly thermal, were considered ex ante to have a higher capacity contribution than they actually delivered at the time it mattered most.

Resource performance can be not only overestimated, but underestimated too. PJM’s wind infeed was compared against the hours when the system operator issued alerts in the period between 6 and 8 January. Contrary to thermal resources, which were experiencing forced outages due to the extreme weather, wind resources performed better than the recognised capacity contribution in 19 of the 20 critical hours.

A mismatch between the expected capacity contribution and the observed contribution is one of the main problems in many capacity mechanisms. This can have questionable impacts when the revenues flow to committed resources such as generators or demand-side resources that do not respond during the contingencies, while resources actually available, such as VRE during the polar vortex, are not recognised for their contribution. Energy transitions will not make the design of capacity markets easier. Although resources such as VRE, demand flexibility and energy storage will gain relevance and may provide a significant contribution to system adequacy, their nature and characteristics present a challenge for those designing capacity products that fairly recognise the real contribution of each technology to the adequacy of the system.

Figure 35. PJM area wind generation and critical hours over the period of 6-8 January 2014 during the polar vortex



Source: PJM (2014).

An example of this problem arises when defining the number of hours for which a resource should be able to respond in order to be considered a capacity provider. Although many alerts last less than 3 hours, in the worst moment of the 2014 polar vortex there were alerts that lasted for 13 consecutive hours. The question of how many hours a resource should be able to react – either by reducing consumption or by increasing generation – has been relevant since demand response was included as part of the resources that can receive a capacity payment.

This creates the need to consider carefully the set of scenarios for which the system should be covered. There is value in keeping the process technology-neutral and allowing more providers to offer the service to ensure cost-efficiency. Different customers have different abilities to adapt their infeed for a specified duration. The definition of a capacity product requires standardisation that may prevent a significant amount of potential resources from participating in the capacity market, depending on the selected characteristics. This becomes especially relevant for demand response and also battery storage, due to their relatively short duration periods, even if they are able to provide significant relief to the system during critical hours.

While capacity mechanisms can be seen as a form of insurance for the system at a fairly limited cost, the exact design parameters require detailed techno-economic assessment and many high-impact assumptions. These will require careful consideration by regulatory authorities to ensure cost-efficiency.

Energy transitions change the meaning of and the solution to scarcity concerns

Both the Australian and the PJM cases, while having been successful in achieving the level of adequacy required by policy makers and regulators, and both having been successful in bringing investment to their own systems, show the limits of the existing market designs. In the Australian case, being capable of effectively mobilising all existing assets via high-value incentives at times of stress barely brings enough investment to cope with peak demand, and not always in a timely manner. Major market and policy uncertainty might further delay investment in dispatchable capacity under this regime. Conversely, the former capacity market design in PJM shows that simply having the assets interconnected to the system is not sufficient. Moreover, most capacity market designs are having problems convincing stakeholders that the contributions of VRE, batteries and demand-side response to covering peak demand are fairly recognised.

Beyond the design of the products, energy transitions will make these issues more pronounced for various reasons and will also create new ones. Peak demand periods will not necessarily correspond with moments of scarcity for dispatchable resources, which means that systems should consider a transition from products based on peak demand towards a definition based on “net peak load”, which will better reflect real scarcity in a world with a large share of variable renewables. In systems experiencing annual peaks during daytime in summer and seeing increasing shares of solar PV, the late afternoon hours could present a critical shortage of reserve capacity rather than a peak in demand, causing scarcity periods to be more distributed throughout the year. With a large share of VRE, contribution to peak demand is no longer the only relevant metric to be used as a qualifier in a capacity auction. The same complex interplay of resources that needs to be addressed in an adequacy assessment will have to be correctly covered in capacity product design, if policy makers choose to take this option.

Regulators and system operators willing to maintain the reliability our societies are used to at an acceptable cost need to adapt existing designs. This includes providing substantial technology-neutral incentives at moments of distress – as a true "energy-only" market does, while implementing better mechanisms for investment to arrive in a timely manner – as capacity markets do.

Ensuring power system flexibility

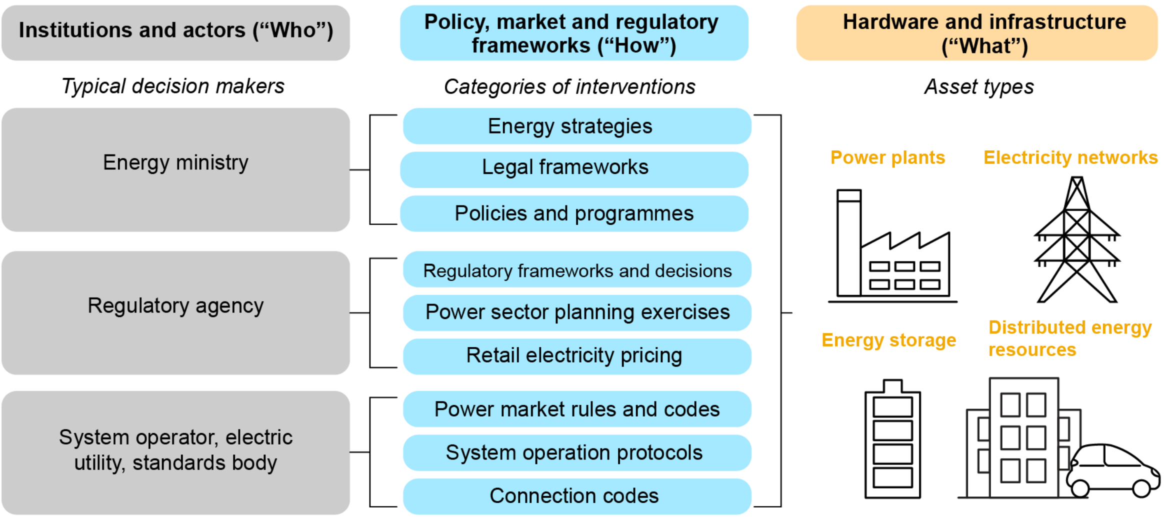
Power system flexibility is necessary to ensure electricity security in modern power systems and enable the higher shares of VRE that are crucial to a successful clean energy transition. It represents the key characteristics needed to handle variability and uncertainty in the system. Declining flexibility during the energy transition can contribute to reduced operational security and system adequacy.

Managing this variability and uncertainty in electricity supply and demand by matching generation and demand – over timescales ranging from several years to seasons, days, hours, minutes and seconds – will require new intervention in technical standards and economic incentives to influence the deployment of the most appropriate investments in hardware and infrastructure to manage this risk.

System flexibility solutions to increasing variability and uncertainty need to address the questions of who, how and what

Power system flexibility encompasses all resources of the power system that allow for its efficient and reliable operation with growing shares of variability and uncertainty. It is composed of three main layers: (a) the roles and responsibilities of the different entities providing or managing system flexibility (the "who"); (b) the policy, regulatory and market frameworks (the "how"); and (c) the technical options, i.e. the hardware and infrastructure (the "what").

Figure 36. Layers of power system flexibility



Source: IEA Status of Power System Transformation 2019.

The institutional layer encompasses the roles and responsibilities of various actors and stakeholders that can participate in providing system flexibility, relating to power system operation and planning. Hardware and infrastructure encompass the technical resources that provide physical power system flexibility – both the physical equipment itself and the flexibility services the equipment provides. There is a range of policy, market and regulatory instruments that play unique, and often complementary roles, in boosting system flexibility. These measures can be grouped into several categories of intervention for decision makers to consider.

Various policy measures can support the cost-effective deployment of flexible hardware and infrastructure, which encompass the technical resources that provide system flexibility. These flexibility resources include power plants (both conventional and VRE), electricity networks, energy storage and distributed energy resources.

Different flexibility options (both physical assets and operational practices) offer timescale-specific capabilities. As power systems transition toward higher phases of system integration, resources from all four categories of flexibility hardware can work in concert to maintain the resilience of the power system.

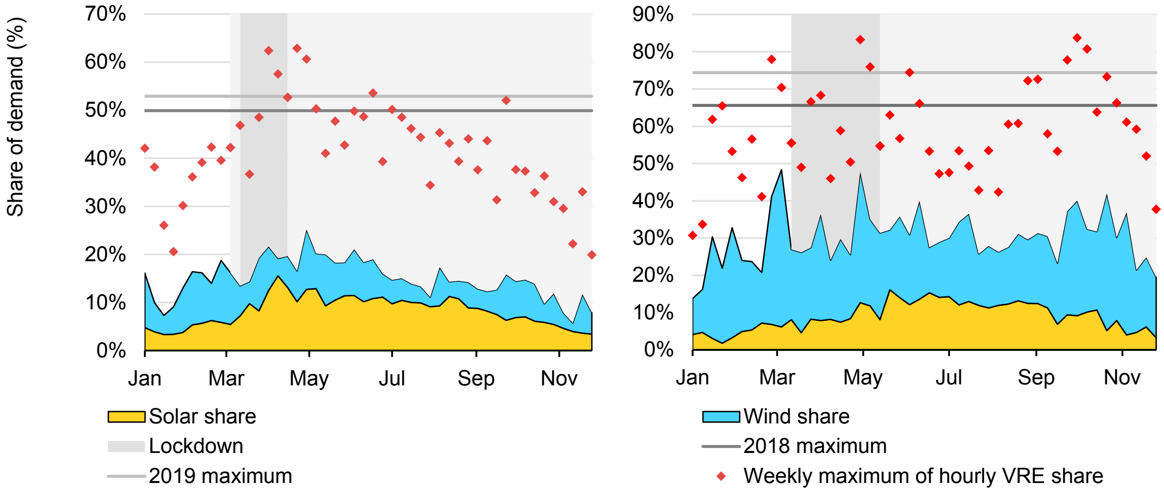
The availability of these technical options requires appropriate policy, market and regulatory frameworks to influence power system stakeholders to invest in and operate hardware and infrastructure to achieve flexibility targets. These frameworks influence the deployment and operation of system flexibility resources through: (a) technical rules and standards for hardware and power system operation; (b) economic incentives influencing the operation of hardware and infrastructure; and (c) economic incentives and planning protocols influencing investment in new hardware and infrastructure.

Ensuring electricity system resilience to a system shock such as Covid-19

From an operational perspective, critical system conditions are often associated with periods of very high demand or low generation availability (VRE, hydro or thermal). The Covid-19 pandemic reminds us that various other shocks can happen to the system and for which resilience is essential, both in the short and long term. The Covid-19-related lockdowns not only broadly reduced economic activity, resulting in lower demand, but also changed the typical patterns of demand in many regions. While most systems worldwide did not show reliability issues, the Covid-19 lockdowns were a stress test that exposed systems to new operational and planning challenges in rare but extreme events at various timescales. These events deserve further attention for future planning.

The lockdown periods saw lower demand and higher-cost thermal generation being pushed out of the merit order, due to the impacts of volume and price. In terms of volume, VRE penetration increased as wind and solar PV infeed is driven by priority dispatch, fixed PPAs and low short-run marginal costs. In combination with continuously increasing VRE capacity over recent years, many systems broke records for instantaneous VRE penetration levels during the lockdown, as seen in Spain and Italy. Higher VRE penetration caused greater variability in the net load profile, which affected the flexibility requirements to meet minimum net demand and system ramp. Wholesale electricity prices have fallen in many regions during lockdowns.

Figure 37. Weekly VRE share in 2020 and during Covid-19 lockdown in Spain (left) and Italy (right)

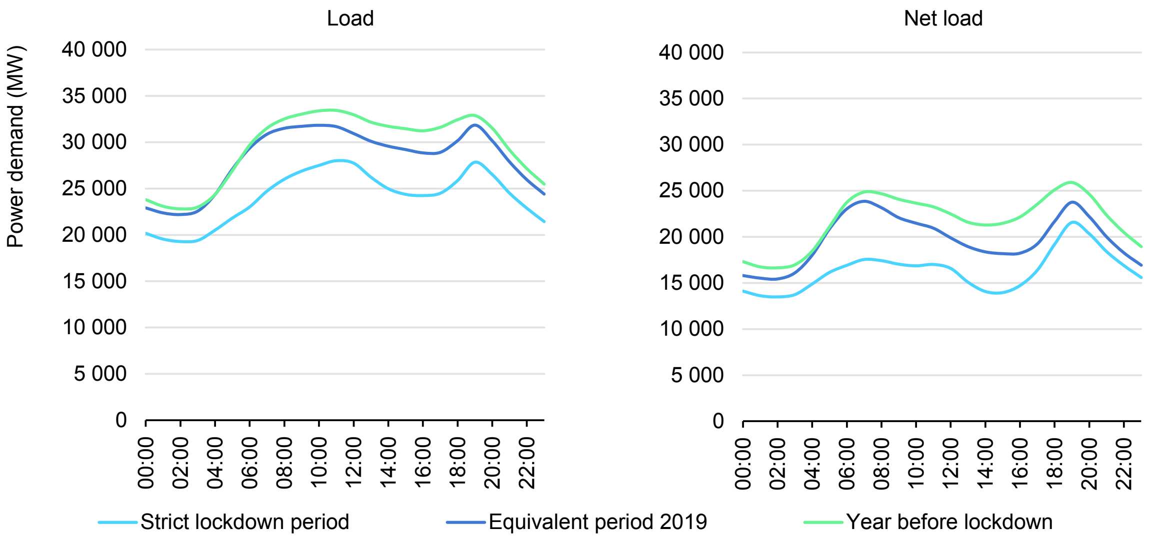


Source: Based on data from ENTSO-E (2020).

New highs in the share of VRE were seen during the lockdown periods as demand fell and higher-cost thermal generation was pushed out of the merit order.

The average daily load profile during lockdown was significantly reduced relative to normal situations in most power systems around the world. This lower demand and higher VRE penetration has implications for real-time system operations, longer-term operational planning and maintenance, and the economic viability of new investments. Our analysis here focuses mostly on real-time system operations and what lower demand actually means for required system flexibility.

Figure 38. Average weekday load (left) and net load (right) in Spain during the full lockdown period compared with the same period in 2019 and the entire year leading to the lockdown

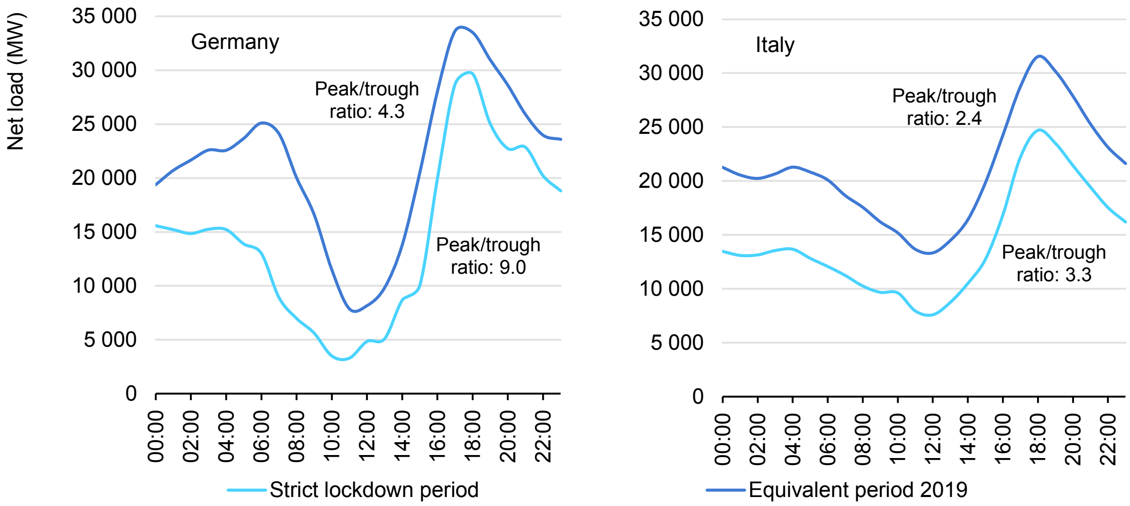


Source: Based on data from ENTSO-E (2020).

Overall reductions in demand, and consequently net demand, during lockdowns increase the spread between minimum and maximum demand relative to the daily peak load. As this spread increases, conventional plants need to operate more flexibly. Plants that are needed to cope with net peak demand may have technical difficulties or be missing economic incentives to reduce output to sufficiently low levels during the minimum, which could lead to renewables curtailment. Under normal conditions such plant inflexibility is already a reason for renewables curtailment and thus a barrier for further renewables deployment in some systems, especially those with less mature market conditions. Shocks that lower demand could aggravate the situation. No clear evidence is available yet that curtailment increased during the Covid-19 pandemic.

Some systems saw new record lows in minimum net load and lower amounts of synchronous generation, but no system reportedly had issues with low inertia or frequency management during such situations. However, systems like that of Great Britain had to advance new procurement schemes for flexibility services to cope with such situations, especially as even lower demand and higher PV penetration over the summer could have been more challenging for the system.

Figure 39. Hourly net load profile for days with the largest difference between the minimum and maximum load requirement, normalised to the daily peak load, for Germany (left) and Italy (right)



Source: Based on data from ENTSO-E (2020).

In most systems, demand fluctuation during the day has been dampened because of lower demand rise in the morning. The decrease in midday demand can also lead to a greater ramp from a very low minimum during the day into the evening peak, which can be more pronounced in systems with considerable share of solar PV.

Despite the operational challenges caused by Covid-19, most power systems are still resilient to handle such shocks by relying on the existing flexibility resources. To ensure reliable system operations, one of the main challenges for many operators has proven to be the human factor. Systems are often designed, planned and operated to cope with various technical potential contingencies. The impact of social distancing and avoiding infection resulted in many operators having to postpone maintenance work and ensure critical staff remain sufficiently distanced or even stay in the control centre.

In addition to the operational issues, lockdowns also forced many planned works to be delayed. This complicates operational planning of the system. A pronounced challenge to ensure system security was visible in France. EDF, as the operator of France's nuclear fleet, had to postpone maintenance and refuelling to later in the year, resulting in lower plant availability during autumn and winter. This could result in potential adequacy issues for the French system, especially in the instance of a cold spell.

In the longer term, periods of lower demand can create revenue issues for all utilities, which may translate into problems of economic viability, both to keep existing plants open and invest in new flexible resources of any type. Some of these issues have also become more apparent in many emerging economies, such as overcapacity caused by overinvestment and existing PPA contracts, which could affect investment decisions in the power sector in the coming years.

Covid-19 highlights the importance of considering such shock scenarios in planning at all timescales. Basic emergency response planning has to cover not only asset and system performance, but also the human factor. Systems and markets need to be ready to handle greater flexibility needs when ramping requirements or minimum net load conditions become more severe. In the long-run most systems anticipate electricity load growth and investment in low-carbon sources. Policy makers should be prepared to intervene in order to minimise the excessive investment gap caused by shocks like that of the Covid-19 pandemic.

Increased uncertainty calls for state-of-the-art forecasting methods

System operators strive for accurate forecasts when determining unit commitments (in central dispatch systems) and reserve requirements. This can minimise ramping requirements that have to be met by fossil fuel and nuclear plants and the need for operating reserves, potentially creating significant system cost savings. Given that VRE forecasts are more accurate closer to real time, power plant schedules should ideally have the option to be updated accordingly. Otherwise, a power plant that may be technically capable of supplying flexibility may be prevented from doing so due to a binding schedule. In market-based systems this uncertainty is lessened by clear balancing responsibilities, close to real-time options for trading and efficient balancing mechanisms. This has led to the development of shorter gate-closure periods in many short-term power markets. For example, Germany has systematically developed its intraday market to facilitate shorter-term trading closer to real time, which has allowed intraday trade to more accurately reflect the profiles of solar PV generation.

Combining the operation of adjacent power systems or balancing areas into a larger region, where this is possible, can unlock the benefit of smoothing out both the variability and forecast errors associated with VRE. The same also counts for load profiles, as the aggregation effect of multiple loads is always taken into account in transmission and especially distribution grid planning. When forecast errors for different VRE sites are not perfectly correlated, an increase in sample size tends to decrease the error. Therefore, forecasts for larger areas are inherently more accurate and relative uncertainty of VRE production is smaller.

Updating system balancing mechanisms

Continuous review is needed to ensure a continuous supply/demand balance

Regardless of the institutional set-up of a specific electricity system, one of the key processes to ensure security of supply is system balancing. Electricity systems need a continuous balance between supply and demand. Mechanisms are needed to correct imbalances resulting from inaccurate forecasting, imperfect information, outages or any other reason. Balancing actions during the energy transition need to adapt to new challenges, as the factors that affect the need for balancing become more complex and interdependent. Variable renewables increase the uncertainty levels in the system, but can also be made more firm and provide balancing services to the system under the right framework. This section addresses specifically how balancing is approached in market-based set-ups where there is an increasing share of variable renewables.

Balancing mechanisms are increasingly market-based

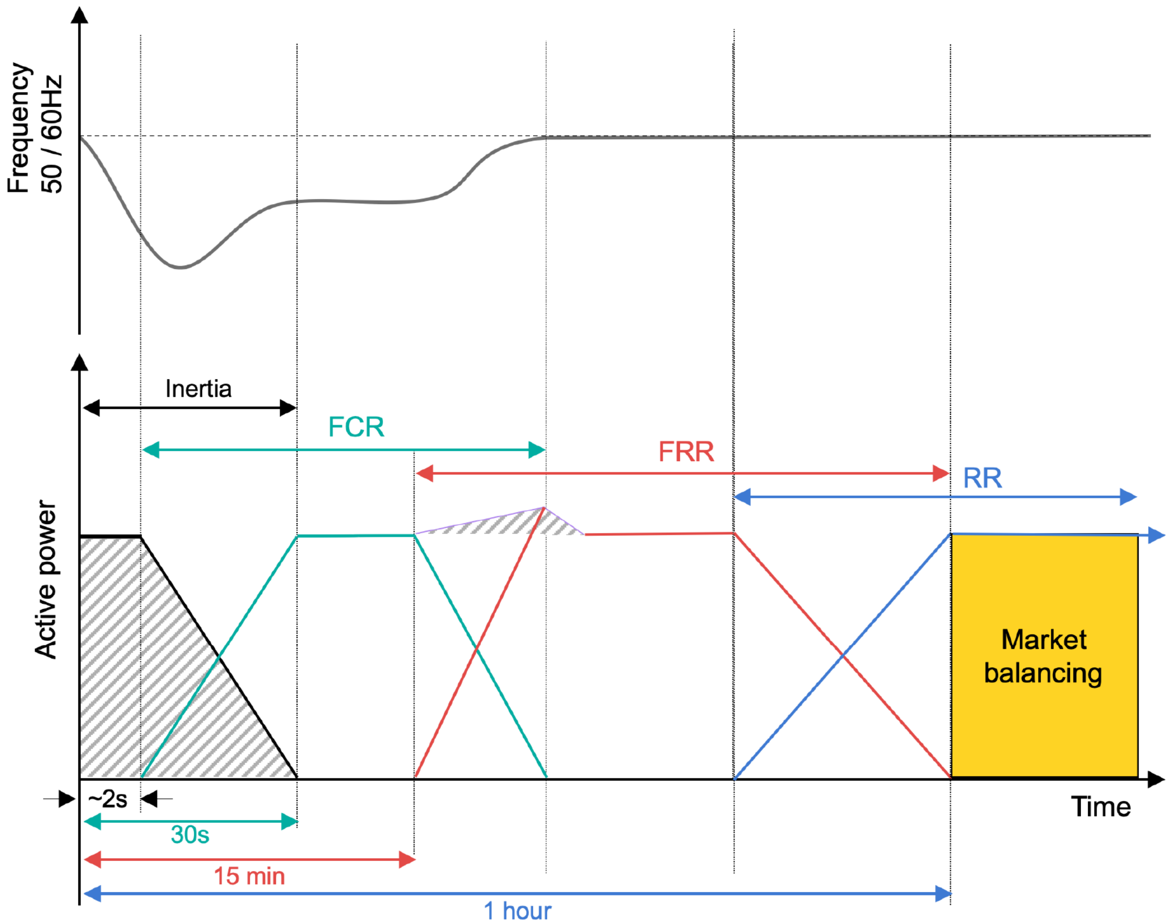
Vertically integrated utilities have traditionally provided balancing services at cost-based rates, but increasingly they are open to competition as regulators recognise the efficiency gains associated with market-based systems. Regulators need to ensure the appropriate level of reliability and thus have a critical role in the design of products that are needed to meet balancing requirements. Criteria include the type, sizing, timing, location, market access, control and pricing for procurement, delivery and settlement of each balancing product. Creating markets involves considering trade-offs between security of supply, economic efficiency and facilitation of market participants. Key steps in this design challenge comprise the following:

- reach agreement on key balancing market criteria and variables
- start from power system and market conditions
- consider future power system and market developments
- strive for appropriate incentives for market participants
- reduce uncertainties through in-depth analysis and monitoring of performance.

Balancing the system relies on a series of complementary reserve mechanisms

Balancing products are differentiated by response time and duration, and the definitions vary by region. For example, in Europe balancing products are categorised as frequency containment (2 second response, 30 second duration), frequency regulation (30 second response, 15 minute duration), and replacement reserves (15 minute response, 1 hour duration). Europe recently announced the ambition to harmonise and consolidate a legacy of diverse processes. Response time criteria limit the types of resources that are able to participate in each category, but can better ensure that the necessary capacity is available to meet the physical demands of the system. Other systems use similar forms of reserves, which can be complemented with additional types such as very fast reserves needed for system conditions.

Figure 40. Types of balancing reserve by response time



Notes; FCR = frequency containment reserve; FRR = frequency regulation reserve; RR = replacement reserve.
Source: Kaushal and Van Hertem.

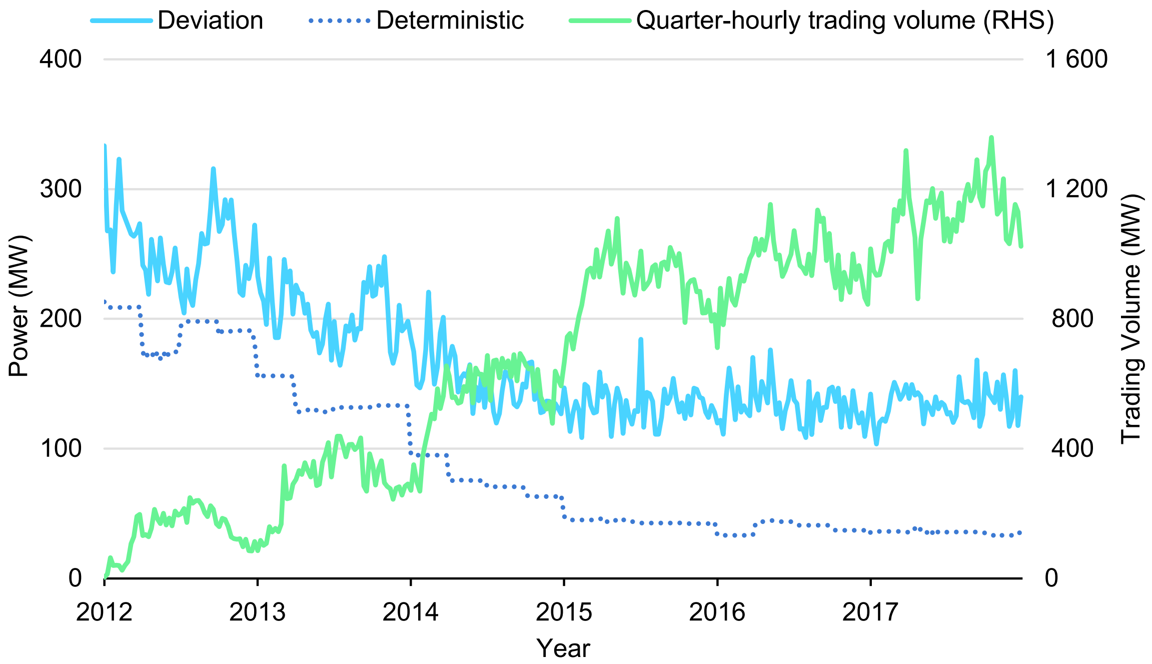
Dimensioning of the system's reserve needs affects total costs and delivered reliability

The sizing of reserve requirements follows two main approaches – deterministic and probabilistic. The deterministic approach defines reserve levels according to specific potential events, such as the largest single contingency (N-1), or a percentage of peak load, and is usually static across time. Probabilistic sizing considers a broader set of events and the correlations between them to assess the level of reserves necessary to meet a prescribed reliability standard. These levels can be static or can dynamically change over time based on system conditions to optimise costs. Increasing size requirements generally brings increased reliability, but at the expense of economic efficiency.

Higher granularity of reserve products and closer to real-time procurement can improve cost efficiency

Timing refers to both the balancing reserves interval and the forward time of procurement (time between gate closure and delivery). This varies significantly by region. In most US ISOs, the market operator procures balancing reserves centrally and co-optimised with energy in short intervals (5 or 15 minutes) close to delivery (within the hour). By contrast, European system operators generally acquire balancing capacity for a variety of intervals before energy is traded in day-ahead markets. A longer timescale increases certainty about the availability of reserves, but at the expense of price efficiency as resource owners require a premium to commit reserves prior to the energy market, decreasing the potential pool of economic capacity. A shorter interval decreases the amount of reserves required due to schedule changes, but also complicates procurement. In Germany the increase in shorter-interval (15 minute) trading led to a decrease in required reserves. In particular, the need for reserves due to scheduling changes, represented by the dotted blue line, was lessened significantly.

Figure 41. Effect of quarter-hourly trading on the need for reserves in Germany, 2012-2017



Source: Koch and Hirth (2019).

As with energy markets, balancing procurement benefits from larger regional integration

The area for which the operator defines reserve requirements is key. A larger geographic scope of reserve procurement increases economic efficiency by lowering volume requirements, as imbalances can net out over a larger area. It increases allocative efficiency, enhances competition and increases accessibility for VRE and demand response. In the case of the European internal energy market, the benefits of co-ordinated balancing were estimated as a saving of EUR 36 billion over 10 years against the initial baseline through regional procurement of balancing reserves. A main challenge to achieving co-ordinated balancing is the harmonisation of processes, contracts and communication systems. Regional integration has its limits, as systems need to acquire a certain share of reserves locally due to transmission constraints that could otherwise affect deliverability.

Balancing markets should tap into all relevant flexibility sources

Market design rules need to address the access requirements placed on resources that affect their ability to participate in balancing, including minimum resource size, technical capabilities and resource type. Minimum unit size requirements can be a function of technical issues such as market operation software limits or dispatch algorithm-processing times, but should be reviewed in light of technical advances in these areas.

Regulatory authorities around the world, such as FERC, European national regulators and the EU Agency for the Cooperation of Energy Regulators (ACER), are encouraging market operators to soften some of these requirements so as to facilitate participation from demand response and distributed energy resources. This could significantly increase supplies available for balancing and reduce costs. In addition to increasing supply, these assets can in fact provide faster and more accurate responses than traditional sources of balancing in many cases. However, relaxing technical requirements has the potential to decrease reserve quality as many of these units lack the detailed telemetry and metering capabilities that are required under current market rules. Jurisdictions are making positive progress in facilitating resource aggregation to meet minimum size requirements and provide some economies of scale, including FERC, which recently passed Order 2222 for this purpose.

Balancing prices need to incentivise total system cost efficiency

The pricing of reserves can be cost- or market-based. The goal of market-based pricing should be to reduce the cost of procuring needed reserves and to provide transparency to the process. Within market-based pricing, there are approaches that either charge all users of reserves that same price, or vary the price of procuring reserves according to the factors affecting the activation of reserves. Charging a mark-up to parties responsible for creating imbalances can induce those entities to improve their own portfolio-balancing procedures. VRE sources and load-serving entities might invest in better forecasting methodologies or develop more flexible supply and demand options as a result. But these charges should still reflect the costs imposed on the system. Punitive pricing, which in this case would mean prices above rates that reflect system costs, can instead cause participants to withhold generation from the market so they can use it to balance their portfolios. While this would help the entity avoid punitive balancing charges, it can increase the price that others need to pay for reserves by withholding potentially economic sources of balancing.

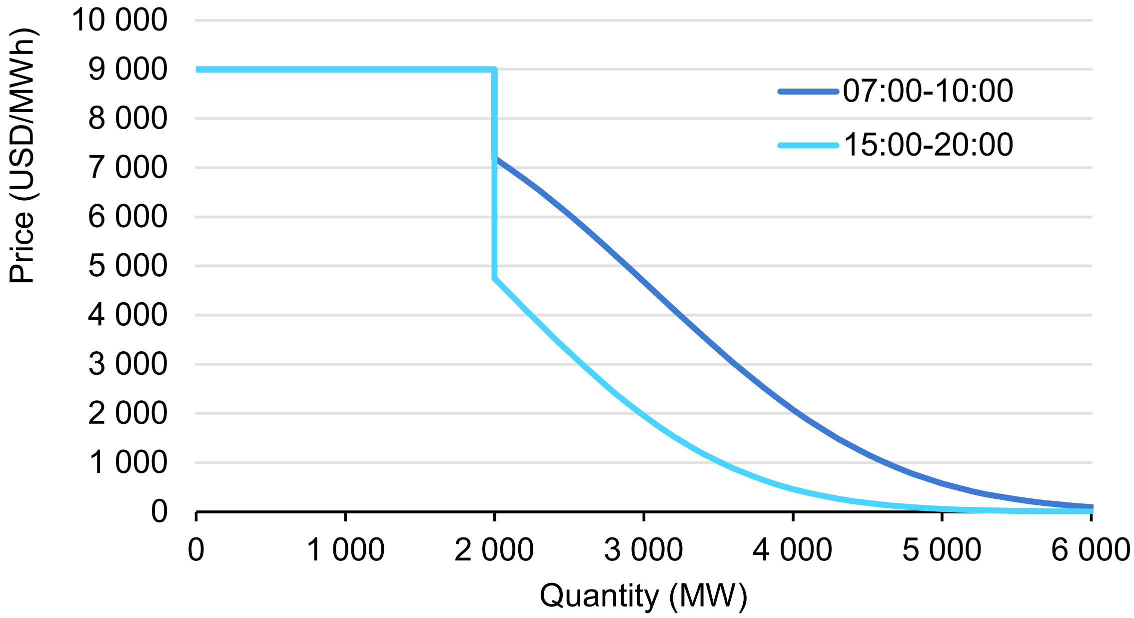
System operators take emergency action when reserves are expected to drop below minimum levels or frequency deviates from prescribed tolerances. From least to most severe, operators call for voluntary load reduction, economic demand response, emergency demand response, emergency generating capability, reserve shortage, voltage reduction, and finally manual load shedding.

Markets such as that managed by ERCOT in Texas and the Australian NEM presently rely on high wholesale energy prices during times of scarcity to signal to loads and flexible generation when it is valuable to increase supply or reduce demand. In addition, Australia has the Reliability and Emergency Reserve Trader (RERT) mechanism, allowing the market operator AEMO to contract for medium- and short-term additional electricity generation reserves (as from November 2017, a maximum of ten weeks ahead of a shortfall). Typical resources that can be procured for RERT include unscheduled load that can be curtailed and restored on request from AEMO, including large industrial loads or aggregated smaller loads, and unscheduled generation assets such as standby diesel generators.

ERCOT introduced an operating reserve demand curve as an administrative mechanism to improve wholesale pricing during scarcity, increasing the prices according to an administratively predetermined schedule as the available operating reserve decreases. This curve places an economic value on flexible capacity in cases of capacity shortage, progressively increasing the price up to the VoLL (USD 9 000/MWh). Dynamically pricing reserves through such a downward sloping demand curve can increase security and economic efficiency by acquiring reserves at greater than the minimum required amount when the marginal value of additional reserves exceeds the cost. The operating reserve demand curve is constructed by assessing the expected economic benefits of avoiding loss of load on the system. It should be noted that this kind of assessment is subject to numerous parameter choices, such as the VoLL and the probability of shedding load, which are difficult to measure and vary greatly between systems, timescales and individual customers.

This “reliability price adder” ensures that reliability actions taken by ERCOT are actually reflected in reserve and energy prices. In the absence of this adder, reliability actions would supress prices. Scarcity prices create the incentive for all market participants to provide capacity when it is most needed by the system, and these high prices are contributing to the recovery of the fixed costs of capacity that is rarely used.

Figure 42. Operating reserves demand curve in ERCOT



Source: ERCOT.

Modernising system stability procedures

A shift in supply sources and new technologies affects system operations and stability

The advent of more variable renewables, demand response and storage affects how the system can be kept in a secure state at all times. The challenges are related to the stability of the power system, which reflects its ability to withstand disturbances on very short timescales. For example, when a larger thermal generator fails or a fault is seen on the system, a stable power system will only see a small deviation from its nominal operational settings. In contrast, in less stable systems these disturbances may trigger a number of significant impacts that can compromise electricity security. Stability becomes one of the key concerns as systems approach Phase 4 of system integration (as discussed in the section on power system flexibility).

System stability studies are inherently highly technical analysis performed continuously by planning agencies, system operators and developers of new generation plants and industrial sites. Therefore, the shifting trends in the make-up of electricity systems may necessitate a change in the connection and operational requirements of all actors, with cost implications. To ensure these trends move towards a more sustainable system in both a cost-effective and reliable manner, it is critical that policy makers such as governments and regulatory authorities take action.

This section summarises the main aspects being considered in various systems and highlights best practices to address them in the context of the energy transition.

Growth in variable and decentralised power sources can create new congestion scenarios, calling for innovative planning, operations and services

Electricity systems were designed for many decades on the basis of large, centralised power plants (thermal and hydro), which feed into a transmission grid that connects industrial loads and supplies small-scale users via the distribution network. Transmission grids were designed in such a way that they could accommodate flows between power plants and main load centres within a specific region without structural congestion. Distribution networks cover smaller individual loads but in aggregate, and in comparison with transmission grids, have a much larger asset base. Such distribution networks typically evolved according to anticipated load growth, with a reasonable level of over-dimensioning to handle unexpected events. Historically, congestion on the grid was relatively exceptional and mostly avoided by grid design. It is often disregarded as a critical security issue, with congestion management being part of the business-as-usual practices of system and market operators.

In the power system transformation being seen across the globe, the generation mix is changing substantially. Renewable sources such as onshore wind, large-scale PV and especially offshore wind are often connected at locations far away from conventional generation or loads. Conventional thermal generation is being phased out or facing an unclear future. Part of the replacement generation – small-scale PV and wind – is increasingly connected at the distribution level. All this affects the main flows within a system and as a consequence can create new potential congestion areas and operational needs.

To mitigate structural congestion and minimise new transmission costs, system planners can incentivise the construction of new renewable plants in preferential locations or closer to demand centres and relevant network infrastructure. These locations may have the incentive of faster connection procedures or lower connection costs. Examples are the renewable energy zones applied in Texas and Australia (see section on system planning).

A higher share of decentralised and variable sources makes it essential that both distribution and transmission system operators have sufficient visibility and controllability to suit system conditions, especially during emergency conditions. These sources should also be incentivised to make production forecasts as accurately as possible.

New generation and net demand patterns would, in theory, require new grid investments and can in some cases even warrant new overall designs. As grid development is capital-intensive and in many parts of the world faces long lead times, other options for optimal grid build-out and operation become important. They include the more efficient utilisation of existing grid assets by means of improved operational planning, dynamic asset ratings, grid controllability of flexible alternating current transmission systems (FACTS), and phase shifters.

Optimised system designs to mitigate future congestion can also include increased use of high-voltage direct-current corridors, higher voltage levels and new conductor types. In the long run – as electricity becomes a more important part of final energy demand and further electricity grid development may hit practical limits – other flexibility options may cost-effectively increase in scale, including flexible energy storage (batteries) and transport (hydrogen). Emerging smart-grid technologies are already contributing to optimal grid management and enhancing reliability, and will continue to grow in importance.

System reserves evolve continuously, but require fundamental rethinking to cope with future system conditions

Electricity systems are operated with a set of operating reserves available for activation at all times. This ensures that forecast errors, portfolio imbalances and contingencies such as plant and equipment failure can be addressed to ensure system-wide supply/demand balance (see section on Updating system balancing mechanisms). Such reserves can be seen as a safety cushion for the system. In past decades these operational reserves were dimensioned to cope with a defined large-scale contingency (or reference incident) and other system parameters. Reserves were usually provided by large thermal plants (often combined-cycle gas turbines [CCGTs]), pumped hydro storage and large industrial sites. Liberalised markets worldwide are allowing stakeholders to make progress on ensuring the volume of contracted operational reserves, their procurement and activation are optimised.

Even with this progress in market design it is essential to keep other system trends in mind, in particular the nature of future providers of reserves and, more fundamentally, which type of contingencies the system should be designed to withstand. Systems that relied on coal- or gas-fired plants for reserves may see these providers leave the system. Experience from market reforms applying technology-neutral criteria and more time-granular ancillary products with shorter lead times shows a substantial increase in reserves provided by demand, storage and aggregators. Furthermore, VRE itself should be able to provide specific operational reserves (mainly downward), although in practice it happens rarely because qualification criteria are still tuned to dispatchable thermal plants.

A more significant question for future electricity security is the contingency that the system should be designed to withstand. A typical design criterion in many systems is the maximum loss of the single largest asset (generator or network component). It places a threshold on the maximum size of a single generator or interconnector to prevent the worst possible imbalance in case of failure. In some cases so-called system split scenarios set the relevant design criteria. Policy makers pay great attention to reliability standards such as LOLE targets. Essentially the reference incident a system is designed to withstand dictates the volume and type of investments than can be made in grids, reserves and the connection capabilities of grid users. With the advent of new large direct-current corridors and offshore wind farms, some system plans promise new electricity “hubs” in the grid with multi-gigawatt interconnections. New technologies may result in a system that is inherently more intelligent and controllable. However, it could also result in a system with larger contingencies to take into account and thus make the system more fragile.

In this journey it remains essential that policy makers take logical steps in system planning that specifically considers reliability:

- The future energy mix and system design strategy should be envisaged.
- Reliability targets need to be defined that link to and cover for a set of reference contingencies.
- This results in system designs and reserve needs that can cope with these contingencies.
- These needs have to be translated into grid investment options and reserve products. They may be complemented with the setting of mandatory capabilities for new grid users.

Maintaining frequency quality faces challenges due to new supply sources and market design criteria

Frequency is the pulse of a power system. A power system in balance will operate at its nominal frequency, either 50 Hz or 60 Hz. Deviations from that balance – by a sudden drop in either supply or demand – will appear very quickly as a deviation in frequency. Only small deviations from the nominal frequency are tolerated (typically -5% to +3%), and balancing services and operational reserves are in place to restore any imbalance and thus the system frequency as rapidly as possible.

In a number of systems, frequency has been managed within a very narrow range for many decades, allowing for only fractions of a percent in deviation for limited time. The presence of system inertia – in the form of large rotating machines – has been a critical factor in keeping system frequency stable. In cases of small imbalances, the difference in energy is taken from or stored in the kinetic energy from these rotating machines. This causes a respective slowing down or speeding up of all rotating machines, which compensates for the energy balance but still results in a quasi-instantaneous frequency deviation. This deviation is then monitored and automatically adjusted by operational reserves from power plants, which provide frequency regulation services via automatic generation control.

Greater interconnection of systems has allowed power systems to access a larger reservoir of inertia and has improved frequency stability. For example, the improvement in the Indian power grid has been particularly dramatic since the synchronisation of regional power grids in the early 2000s. This allowed for a strong reduction in frequency deviations, which provides more secure system operation and is beneficial for the performance of many connected loads and generators.

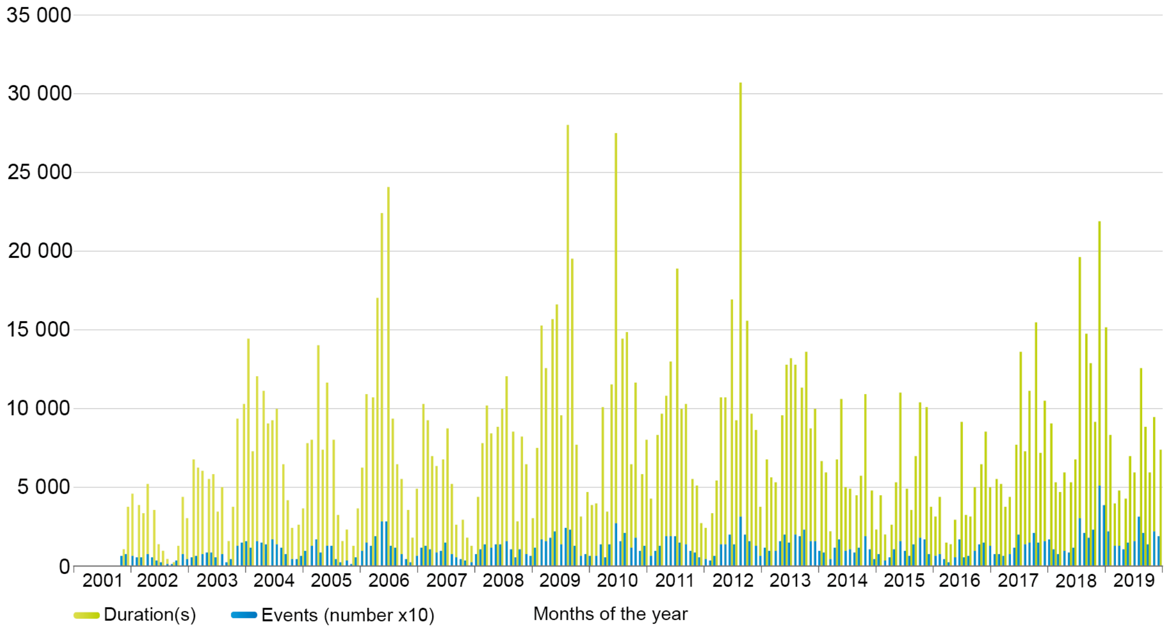
An often-debated topic in recent years is whether modern electricity systems can transition from conventional thermal plants to significantly more solar PV or wind from the perspective of frequency stability, as some system inertia is removed. While there is greater risk exposure when synchronous generators retire from the system, the underlying issues of a transition to converter-based resources, and the solutions to address them, are more complex.

Another reason for frequency deviations is observed specifically in unbundled market settings where so-called deterministic frequency deviations occur. These are frequency deviations at the change of market intervals (e.g. each hour in the case of hourly bids). Such deterministic deviations can be addressed by specific market design improvements, such as moving to shorter market intervals, limiting the ramping of portfolios or individual units, and limiting net position changes across market intervals at the zonal level.

The large continental European system has recently experienced a number of large deviations and their duration has increased again after seeing stable frequency until 2017. The majority of these deviations are due to changes in generation schedules based on hourly blocks, which creates sudden imbalances. This highlights how electricity market rules need to take into account the physical characteristics of the system, and the security implications. Alternatively, more reserves can be activated, which limits the deviations without solving the root cause.

This topic of frequency quality is not always characterised as a security issue. Nevertheless, given the critical role of maintaining frequency quality and the adverse effects that deviations can have on the reliability of equipment and possible disconnection, it needs to be taken into account in power systems undergoing a transition.

Figure 43. Frequency deviations in continental Europe larger than 75 MHz



Source: ENTSO-E (2019), Continental Europe Significant Frequency Deviations – January 2019.

Voltage management remains crucial in more intensely loaded systems with a changing supply mix

If frequency is the pulse of the electricity system, voltage is its blood pressure. Secure operation of grid and user assets relies on voltage profiles being kept within set target levels. Voltage stability is a critical security consideration in system planning and operations. Many large-scale outages in recent decades can be attributed to voltage instability (e.g. Brazil in 2009), often triggered by another incident.

Historically voltage levels were maintained by reactive power control of large thermal or hydro plants and in some situations dedicated equipment belonging to grid operators. Voltage management principles also relied on the system flows being mostly unidirectional from large central plants to users down at lower distribution levels.

As the energy mix shifts to new sources with large capacity in remote locations, as well as large numbers of small-scale units embedded in the distribution system, the approach to voltage management and maintaining voltage stability experiences significant impacts. Bidirectional energy flows can create overvoltage situations at lower voltage levels. This can be aggravated as more networks are moved undergrounded, as they have a different reactive power profile that can trigger overvoltage, especially in low-load situations. Voltage stability issues have become much better understood in the past three decades. Particularly in emerging economies, rapidly increasing demand may not always be followed by grid investment. In such cases voltage instability triggers need to be continuously evaluated in operational and grid planning.

Tools and knowledge are available for managing voltage levels in transformed power systems. Solutions are also needed that depend on policy maker and regulatory authority decisions. Such hardware and software solutions can be either network operator-driven, such as dedicated voltage control devices, or can be imposed on grid users, such as generators and distribution network operators, as a requirement to provide voltage control capabilities. Such decisions need to be made with fundamental regulatory criteria in mind, such as long-term objectives for system reliability, total system costs for end users, and cost allocations between regulated entities and market actors.

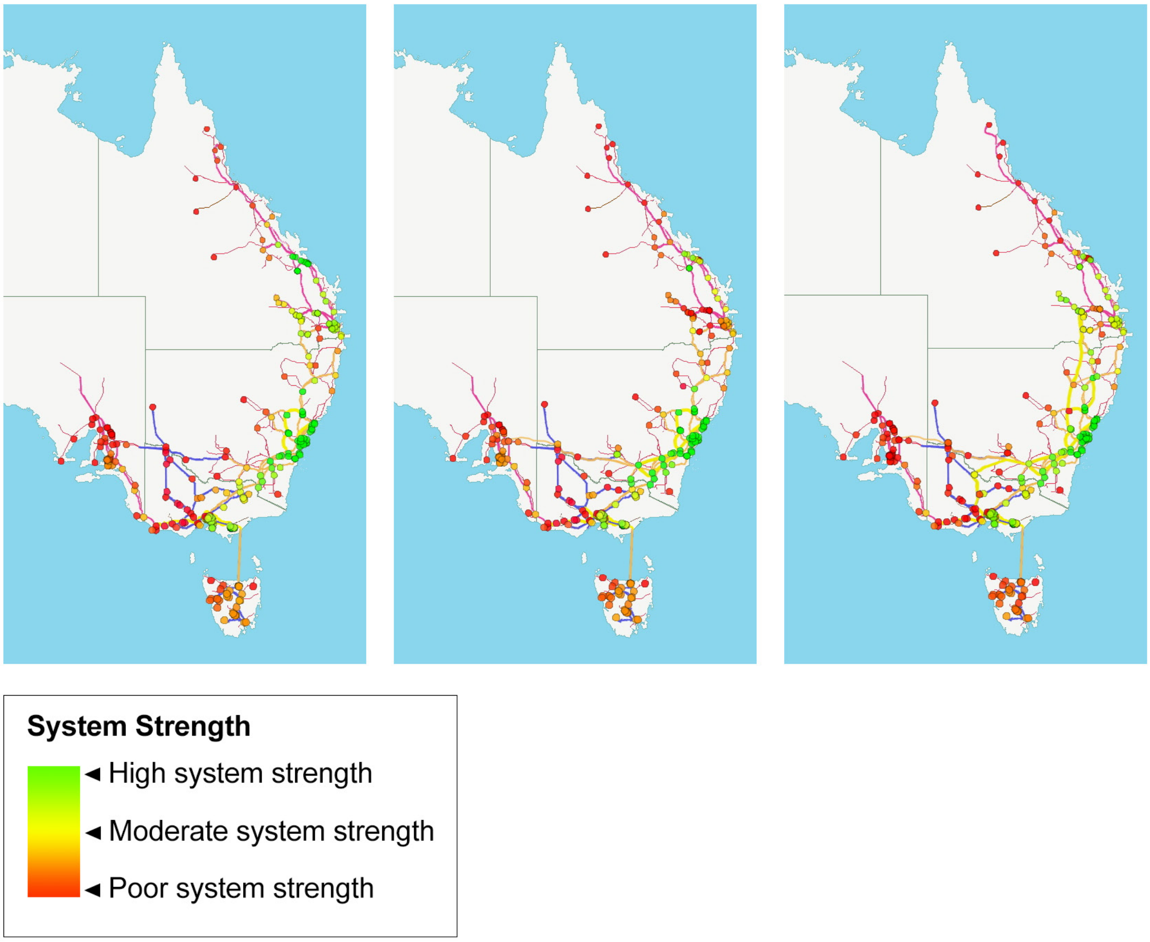
A transformed power system needs a renewed approach to ensure secure faults

Electricity systems cover large areas and are prone to fault situations with resulting short-circuit currents. These include equipment failure, trees hitting overhead lines, construction works compromising underground cables, and so on. Much like a fuse box at home, the electricity grid relies on detection of short-circuit currents to isolate faults to protect the system from damage. Transmission grids typically have a meshed design, so that a faulted line segment or generator can be isolated while still supplying all grid nodes via redundant paths. Distribution networks are normally operated in a radial manner, which implies that in the case of a fault, the nodes behind a fault will all be disconnected from supply. In addition to current-related fault detection, the system also relies on protection settings related to, among other criteria, frequency and voltage. These are meant to disconnect generation and demand to either protect the assets themselves or the wider system when faults occur.

With a changing energy mix these principles have had to be reviewed over the past decade and will require continuing attention in the future. System protection normally implies that, in case of a fault, there is a high enough short-circuit current that can be detected by protection relays and activate a disconnection. These high currents are intrinsically provided by rotating electrical machines such as thermal generators or hydro plants, as well as electrical motors. When conventional generation is replaced by solar PV and wind turbines on part of the system, these new sources can provide energy, and to some extent, support frequency management and voltage control, but they do not provide intrinsic high currents in case of nearby faults. This lower short-circuit current level – in combination with the aforementioned system inertia level – is characterised as weaker “system strength”. This can be interpreted as systems having weak, fuzzy or unstable voltage levels whenever disturbances occur.

Systems with lower system strength rely on new types of fault detection methods and protection schemes. These systems ultimately require other sources that fill the gap caused by lower system strength, by delivering very fast frequency response as close to inertia-like response as needed, as well as higher currents when faults occur. These currents, needed to trip protection in a correct and fast manner, can be provided by dedicated assets such as synchronous condensers or by dedicated controls from solar PV, wind, battery and HVDC converters. For example, the low system strength in the South Australian grid has resulted in the placement of four synchronous condensers in those areas where system strength has dropped to critical levels.

Figure 44. System strength levels in South and East Australia in the Central Scenario in 2020-2021 (left), 2029-2030 (middle) and 2034-2035 (right)



Source: AEMO (2020), 2020 Integrated System Plan.

A further issue that deserves attention with an increasing share of renewables is that of protection settings of individual grid users. Generators of all sizes and large demand connections have protection devices to disconnect them according to a variety triggers, which can signal a system issue that jeopardises the safety of staff and equipment. These triggers include (among others) frequency level, rate-of-change of frequency, voltage level, voltage dip profile, voltage vector jumps, voltage balance, currents and others. The energy transition has several impacts. Large shares of small-scale generation can have a wide system impact if they all disconnect at the same time. Therefore, their disconnection settings should not be based solely on local effects, but also on the overall impact on system security.

A typical example of this issue is that of solar PV units in Germany, which were installed about a decade ago with disconnection triggers at 50.2 Hz to comply with local distribution safety criteria during a system disturbance. While such a rule allows secure system operation at low penetration levels of solar PV, it can pose a threat at higher levels. Whenever the entire system frequency rose to 50.2 Hz – which is outside a normal range, but a likely event – the simultaneous disconnection of many solar PV units would result in a sudden loss of many gigawatts, an event larger than operating reserves are designed to manage. This necessitated an update of grid codes for new units, retrofits to old units and additional reserve procurement to cope with the remaining old unit settings, which were implemented by various countries. At the same time, new options are being explored for protection strategies in distribution grids.

Erroneous simultaneous tripping of many solar PV units (1 200 MW) also occurred in the Southern California system in August 2016. Wildfires caused several transmission faults, which were secured. Voltage transients in the system were identified by the PV inverters, and as very fast strong frequency deviations could be interpreted as islanding of the system, the solar PV units simultaneously disconnected. As a lesson learned from this situation, grid codes were updated to allow for a time delay in the protection setting.

Another scenario occurring in systems with high shares of PV and wind has signalled that fault ride-through settings need revision. Whenever a transmission fault is cleared, the nodes in the vicinity of the fault see a voltage dip for a short period of time. It is essential that a sufficient number of generators remain connected during such a secured fault. A mass disconnection could also result in a substantial supply/demand imbalance and a possible frequency collapse. Worldwide grid codes are being updated to strengthen fault ride-through settings for smaller generation. This essentially results in a trade-off between risks and costs among grid users, distribution system operators and transmission system operators.

These are just some examples of how grid code review processes, even when covering detailed technical aspects, require regulatory oversight and policy maker action.

Policy makers need to steer a review of connection requirements, operational practices and market-based solutions

Electricity security as defined in this report is a broad concept of various policy, economic and technical measures to cope with a wide range of risks. The notion of system stability is narrower and refers to the property of the electricity system to maintain a state of operational equilibrium and to recover from disturbances on very short timescales. It is an intricate electrical engineering notion that has received wide-scale attention in the scientific community for decades. From a system analysis perspective, the impacts of an energy transition are well-documented, identifying the issues that are likely to come up and the solutions that work well. Ensuring electricity security implies that these technical issues and solutions are addressed in clear regulatory processes. This involves following a set of logical steps:

1. Identify electricity scenarios resulting from present and possible future policies.
2. Set reliability targets and identify critical contingencies that need to be handled securely by the system.
3. Assess system needs to meet these targets, in the short term and the long term. These needs can include active or reactive power infeed responses to various triggers or ride-through capabilities.
4. Anticipate long-term needs by means of dedicated research and demonstration. This may also result in mandatory connection requirements to future-proof the system.
5. For more short-term system needs, analyse optimal delivery of these needs by either market services, mandatory services or dedicated grid operator assets.
6. For market services and mandatory response services, set clear technical specifications and compliance procedures.
7. Set up relevant market schemes and grid codes.
8. Regularly review and improve.

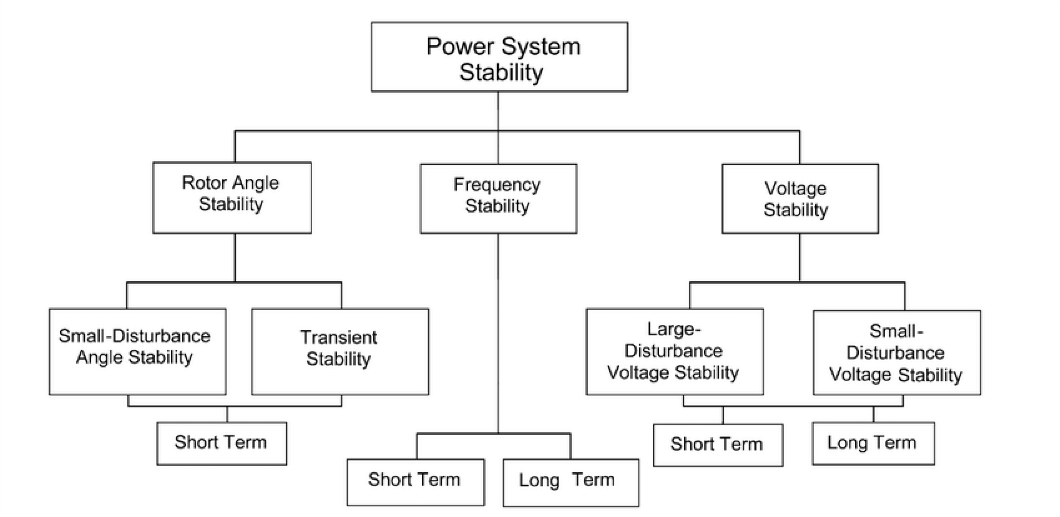
This requires active direction from policy makers and regulatory authorities to ensure that the objectives of security, cost-effectiveness, and a fair allocation of costs and benefits are met.

Box 1. A range of stability phenomena remain as relevant as ever

Electricity systems are highly non-linear and depend on multiple variables. The analysis of such systems requires detailed modelling tools. A proper understanding of system stability, and as such the main risks and solutions, benefits from a structured categorisation of typical stability problems.

The question of defining and classifying power system stability has been addressed by the International Council on Large Electric Systems (CIGRE) and the Institute of Electrical and Electronics Engineers (IEEE) task forces. Their reference classification of stability problems covers rotor angle, voltage and frequency stability. An additional differentiation is that of short-term instability phenomena occurring in a timespan of seconds and long-term instabilities due to a cascading effect, which can take minutes to unfold. The framework also applies definitions for reliability and security that are slightly different from those applied in this report and cover purely technical aspects. A full description of these stability phenomena goes beyond the scope of this report, but can be summarised as follows.

IEEE/CIGRE classification of power system stability



Source: IEEE/CIGRE (2020).

Rotor angle stability refers to the ability of synchronous generators to remain synchronised after a disturbance. A case where synchronism is lost can result in the generator tripping. Any system is designed to cope with the simultaneous disconnection of one or even multiple large units. Therefore rotor angle instability is rarely the sole cause of an outage. Still, such disconnection can trigger further issues such as cascade tripping, voltage instability or frequency instability.

Voltage instability (also called voltage collapse) can occur after the tripping of one or more lines or generators, and is intensified in heavily loaded systems due to the reactive behaviour of the grid itself. Some outages of past decades due to voltage collapse are the blackouts in Belgium in 1982, Sweden/Denmark in 2003 and the Philippines in 2013.

Frequency instability refers to issues in cascading disconnections that increase the supply/demand mismatch and can be associated with insufficient reserves, ineffective protection settings and limited co-ordination. Ultimately a strong frequency drop can trigger a first stage of demand disconnections, as experienced in continental Europe in 2006 and Great Britain in August 2019.

Box 2. Cascading effects trigger frequency instability and demand disconnection in Great Britain – the outage of 9 August 2019

During the late afternoon of Friday, 9 August 2019 about one million people in Great Britain lost power for nearly an hour. Many residential customers, business and rail commuters were affected. The outage was triggered by several almost simultaneous generator trips after a lightning strike hit a transmission line. Initially an offshore wind farm and a CCGT tripped due to erroneous control software settings and issues in a steam bypass system. In addition, a share of embedded generation at distribution level tripped immediately due to voltage transients caused by the large plant tripping and rate-of-change of frequency protection settings.

This combined outage was about 1.4 GW and went beyond the safety standards of the system. All available fast reserves were activated and could arrest the frequency drop at 49.1 Hz. Unfortunately another turbine at the same CCGT subsequently tripped. This resulted in a total supply loss of 1.6 GW. No more reserves were available in the system as this situation went far beyond what the system is designed for. When frequency hit 48.8Hz a first stage of demand was disconnected, which proved to be sufficient to restore the supply/demand balance and prevent a full system collapse. Fast recovery measures were undertaken by the transmission system operator and various distribution network operators to restore power.

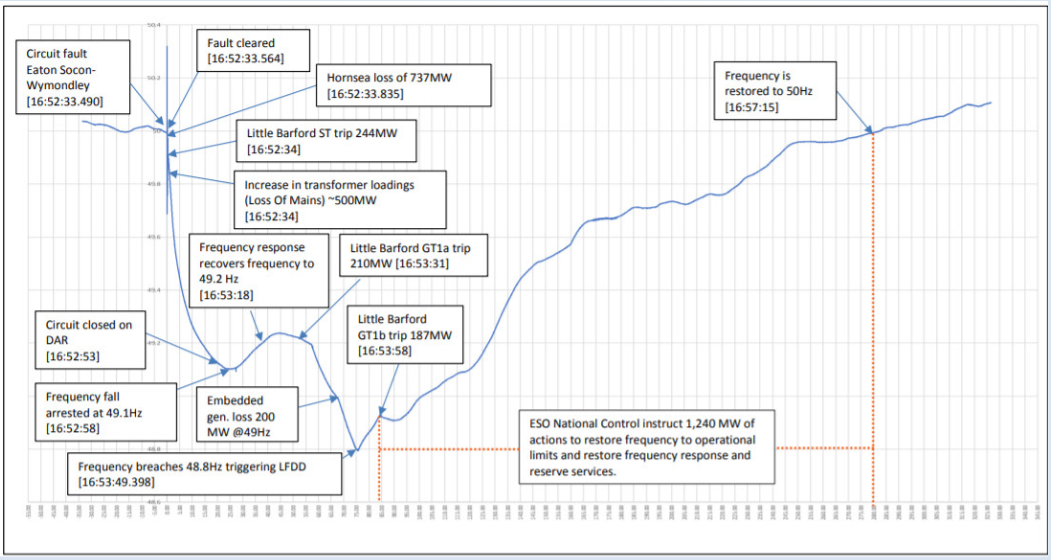
The outage evidently received widespread attention due to the number of affected people. The incident gave several insights and lessons learned:

- Reliability standards do not and should not capture all possible events. Two quasi-simultaneous large outages were beyond reasonable safety standards. A higher sizing of reserves to cope with such a large loss of generation may have prevented the demand disconnection, but would have come at higher costs.

Reliability standards are meant to strike a balance between such costs and improved security levels.

- Embedded generation tripped in large amounts. A programme was already underway to review and update protection settings. When embedded generation takes a large share of the total electricity mix, its disconnection is no longer just a local distribution issue – mass tripping can have system-wide impacts. The incident gave reason to review the timescales of the programme. It highlights how countries that are just starting to see substantial VRE shares should set connection requirements that are future-proof.
- Load shedding is an effective tool to prevent full system blackout, albeit a last-resort mechanism with substantial economic impact and public backlash. It is essential that supply is restored as soon as possible by fast and appropriate co-ordination and communication between key authorities and system operators. Authorities need to oversee which parts of the system are shed by setting clear criteria. Grid users need to be encouraged to take appropriate measures to cope with short-duration outages.

Sequence of events of the August 2019 outage in Great Britain



Source: National Grid ESO (2019), Technical Report on the events of 9 August 2019.

Box 3. The effect of high shares of converter-based sources

Historically, the rotating inertia of large generation was relied on as a stabilising property of power systems. In conventional generation – coal, gas, nuclear and hydro – the rotating parts used for power generation spin at the same electrical speed across an interconnected power system such that all generators contribute to keep each other in balance. This can be compared to cyclists on a tandem bike working together. The common speed of such synchronous generators defines the power system frequency, which at all times is to be kept close to the nominal level.

Synchronous generators can easily pick up lost output in the case of a generator outage. In such a situation, and provided that the disturbance is not significant relative to the remaining synchronous generation, the stored kinetic energy of the rotating parts will be partially transferred into electrical energy and fed into the system. In classic power systems dominated by synchronous generation, it is easier to maintain the frequency following a disturbance as each individual generator will have limited kinetic energy available to transfer to compensate for a disturbance. It will see limited slowdown and thus limited reduction in frequency.

Unlike synchronous generation, which directly injects AC power into the grid, VRE generates electrical DC power and is connected to the power system through power electronics called converters. These sources do not provide inertia like large conventional plants. As their share increases in the system, this leads to lower overall system inertia which creates the risk of stronger frequency instabilities in case of an incident.

A related issue is that converters also provide less short-circuit current and thus create issues for proper system protection schemes. The combination of low inertia and low short-circuit power is referred to as a system strength issue.

Recent years have seen much attention being given to new technologies for VRE integration to cope with this challenge. Research shows that 100% converter-based systems without any conventional generation could be operated by requiring and defining grid-forming controls within converters. This refers to the capability of the converters to:

- Behave as close as possible to an ideal voltage source, meaning that their provision of voltage is not sensitive to current changes.
- Synchronise with other voltage sources such as thermal generators, or act as a reference for other grid-forming units.
- Operate in islanded mode.
- Protect against current surges to avoid instability.

Grid-forming controls are deployed in several applications already, but not yet at large scale in an interconnected power system where the number of infeed sources changes dynamically. Experiences include:

- Microgrids: Grid-forming controls have been shown to operate small islanded power systems for hours or even days. A key difference to a large interconnected power system is that there is no meshed topology and there is upfront certainty on the size and location of all other generation sources.
- Offshore wind: HVDC links for offshore wind farms include a station with grid-forming capabilities. In this arrangement, the converter has a much larger dimension compared to individual wind turbines and controls their output.
- Uninterruptible power supply: Critical infrastructure such as hospitals and data centres often have local units to ensure continuity of supply. These are of small size and do not operate in a large meshed network.

Stability issues will need resolution before a 100% instantaneous penetration level of VRE is reached. Such very high instantaneous VRE penetration is already seen in parts of an interconnected system, such as Denmark and South Australia. A future power system with annual VRE levels of 24% or 40% globally, as projected in the World Energy Outlook Stated Policies Scenario and Sustainable Development Scenario respectively, will see moments in time across large areas where such high instantaneous infeed levels are reached. A clear roadmap needs to be implemented in time for systems with ambitious renewables projections. This should include the following elements:

- Assess the stability limits of non-synchronous generation early in the planning process, and monitor system inertia behaviour closely in real time.
- Initially, in the absence of specific technical solutions, a high infeed of PV and wind may need to be curtailed and covered by conventional generation (often fossil-fuel based). This may be reasonable when curtailment levels remain very limited. If there is a policy ambition or market willingness to ramp up VRE deployment, such curtailment may no longer be a cost-effective option in the long run and can put a barrier to achieving low-carbon targets.
- Synchronous condensers may be a viable option to ensure stability, and also when system strength becomes an issue. These assets are similar to large generators as they have rotating mass, high fault current levels and reactive power control, but they have no prime mover and do not generate active power. Such synchronous condensers can be new equipment or repurposing of phased-out thermal plants. They have already been implemented in Denmark and South Australia to cope with high levels of VRE and mitigate system strength problems.

- New fast frequency services ("synthetic inertia") can, in combination with real-time assessments, limit curtailment. These services can be procured from a variety of assets (VRE, batteries, electric vehicle chargers, and others).
- Grid-forming solutions can be applied to push the level of instantaneous VRE penetration higher. Further innovation and demonstration projects to increase technology readiness levels for their large-scale implementation will become key. A roadmap is essential to graduate from demonstration projects to the widespread deployment of grid-forming controls. Pilot regions can be considered that are already at the forefront of the development of renewable energy sources.
- Eventually, grid-forming solutions require planning to ensure technical specifications are clear, reasonably harmonised, checked for interoperability and enforced in appropriate regulatory procedures. The cost-optimal solution – synchronous condensers, fast frequency reserves and other similar services, or grid-forming converters – depends on the local system context, in particular the prevailing VRE levels.

Source: IEA-RTE webinar on "Technical conditions for 100% converter based infeed", 25 March 2020.

References

AEMO. (2016). *Demand side participation*. *September*. https://www.aemo.com.au/-/media/Files/Electricity/NEM/Planning_and_Forecasting/NEFR/2016/Demand-Side-Participation---2016-National-Electricity-Forecasting-Report.pdf.

AEMO. (2020a). 2020 Electricity Statement of Opportunities. https://aemo.com.au/en/energy-systems/electricity/national-electricity-market-nem/nem-forecasting-and-planning/forecasting-and-reliability/nem-electricity-statement-of-opportunities-esoo.

AEMO. (2020b). Energy Adequacy Assessment Projection. https://aemo.com.au/en/energy-systems/electricity/national-electricity-market-nem/nem-forecasting-and-planning/forecasting-and-reliability/energy-adequacy-assessment-projection-eaap#:~:text=The%20Energy%20Adequacy%20Assessment%20Projection,National%20Electricity%20Market%20(NEM).

AEMO. (2020c). Draft 2020 Integrated System Plan (ISP) Generation Outlook. https://www.aemo.com.au/energy-systems/major-publications/integrated-systemplan-isp/2020-integrated-system-plan-isp.

AEMO. (2020d). Market Data NEMWEB. http://www.nemweb.com.au/.

DECC. (2013). *Annex C: Reliability Standard Methodology*. *July*, 1–12. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/223653/emr_consultation_annex_c.pdf.

EIA. (2020). *EIA Independent Statistics and Analysis*. https://www.eia.gov/opendata/.

EirGrid Group. (2019). *All-Island Generation Capacity Statement 2019-2028*. https://www.eirgridgroup.com/site-files/library/EirGrid/EirGrid-Group-All-Island-Generation-Capacity-Statement-2019-2028.pdf.

ELIA. (2019). Adequacy and flexibility study for Belgium 2020-2030. https://www.elia.be/-/media/project/elia/elia-site/adequacy-studies/adequacy-studies/20190628_elia_adequacy_and_flexibility_study_2020_2030_nl_archive.pdf

ELIA. (2020). The need for a strategic reserve for winter 2021-22. https://www.elia.be/en/electricity-market-and-system/adequacy/strategic-reserves.

ENENTSO-E. (2017). MID-TERM ADEQUACY FORECAST 2017 Edition. https://eepublicdownloads.entsoe.eu/clean-documents/sdc-documents/MAF/MAF_2017_report_for_consultation.pdf.

ENTSO-E. (2020). Mid-term adequacy forecast. https://www.entsoe.eu/outlooks/midterm/

ESKOM. (2020). Medium-term system adequacy outlook. https://www.eskom.co.za/Whatweredoing/SupplyStatus/Pages/SupplyStatusT.aspx.

ERCOT. (2021). Resource Adequacy. http://www.ercot.com/gridinfo/resource.

European Commission. (2016). The role and need of flexibility in 2030: focus on energy storage. In *Dg Ener* (Issue August). https://doi.org/10.2833/639890.

FERC. (2019). *PSE Annual Electric Balancing Authority Area and Planning Area Report*.

Hogan, W. W. and Pope, S. L. (2019). *PJM Reserve Markets: Operating Reserve Demand Curve Enhancements*. 87. https://sites.hks.harvard.edu/fs/whogan/Hogan_Pope_PJM_Report_032119.pdf

IEA. (2017a). *Getting Wind and Sun onto the Grid A Manual for Policy Makers*. https://www.iea.org/reports/getting-wind-and-solar-onto-the-grid/.

IEA. (2017b). *Status of Power System Transformation 2017: System integration and local grids*. https://webstore.iea.org/status-of-power-system-transformation-2017.

IEA. (2018). *Status of Power System Transformation 2018: Advanced Power Plant Flexibility*. https://www.iea.org/reports/status-of-power-system-transformation-2018/.

IEA. (2019a). *Integrating Power Systems across Borders*.

IEA. (2019b). *Renewables 2019*. www.iea.org/renewables2019.

IEA. (2019c). *Southeast Asia Energy Outlook 2019*.

IEA. (2019d). *Status of Power System Transformation 2019: Power System Flexibility*. https://www.iea.org/reports/status-of-power-system-transformation-2019.

IEA. (2019e). *World Energy Outlook 2019*. https://webstore.iea.org/world-energy-outlook-2019.

IEA. (2019f). *World Energy Outlook 2019*. https://www.iea.org/reports/world-energy-outlook-2019.

IEA. (2020a). *CCUS in Power*. https://www.iea.org/reports/ccus-in-power.

IEA. (2020b). *World Energy Outlook 2020*. https://www.iea.org/reports/world-energy-outlook-2020.

Kaushal, A. and Van Hertem, D. (2019). An overview of ancillary services and HVDC systems in European context. *Energies*, *12*(18), 1–20. https://doi.org/10.3390/en12183481.

Koch, C. and Hirth, L. (2019). Short-term electricity trading for system balancing: An empirical analysis of the role of intraday trading in balancing Germany's electricity system. *Renewable and Sustainable Energy Reviews*, *113*(June), 109275. https://doi.org/10.1016/j.rser.2019.109275.

Newell, S., et al. (2018). *Estimation of the Market Equilibrium and Economically Optimal Reserve Margins for the ERCOT Region 2018 Update*.

North American Electric Reliability Corporation. (2019). *Summer Reliability Assessment*. *June*.

POSOCO. (2020). *Flexibility Analysis of Thermal Generation for Renewable Integration in India*. https://www.researchgate.net/publication/341755366_Flexibility_Analysis_of_Thermal_Generation_for_RE_Integration_in_India.

PSE. (2017). *2017 PSE Integrated Resource Plan*. *November*.

Resource Adequacy. (n.d.). Retrieved July 7, 2020, from http://www.ercot.com/gridinfo/resource.

The Brattle Group, Sargent & Lundy. (2018). *PJM Cost of New Entry: Combustion Turbines and Combined-Cycle Plants with June 1, 2022 Online Date*. April 19. https://www.pjm.com/~/media/committees-groups/committees/mic/20180425-special/20180425-pjm-2018-cost-of-new-entry-study.ashx.

van der Veen, R. A. C., and Hakvoort, R. A. (2016). The electricity balancing market: Exploring the design challenge. *Utilities Policy*, *43*, 186–194. https://doi.org/10.1016/j.jup.2016.10.008.

"*This publication has been produced with the financial assistance of the European Union as part of the Clean Energy Transitions in Emerging Economies programme. This publication reflects the views of the International Energy Agency (IEA) Secretariat but does not necessarily reflect those of individual IEA member countries or the European Union (EU). Neither the IEA nor the EU make any representation or warranty, express or implied, in respect to the publication's contents (including its completeness or accuracy) and shall not be responsible for any use of, or reliance on, the publication.*"

The Clean Energy Transitions in Emerging Economies programme has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952363

IEA Publications

International Energy Agency

Website: www.iea.org

Contact information: www.iea.org/about/contact

Typeset in France by IEA – Avril 2021

Cover design: IEA

Photo credits: © ShutterStock