Energy solutions
for a changing world

Power Market Operations and System Reliability in the Transition to a Low-Carbon Power System

A Contribution to the Market Design Debate

Lead Author
Michael Hogan

May 2015

How to Cite This Paper

Hogan, M., Weston, F., and Gottstein, M. (2015). *Power Market Operations and System Reliability in the Transition to a Low-Carbon Power System: A Contribution to the Market Design Debate*. Brussels, Belgium: The Regulatory Assistance Project. Retrieved from http://www.raponline.org/document/download/id/7600

Electronic copies of this paper and other RAP publications can be found on our website at **www.raponline.org**.

To be added to our distribution list, please send relevant contact information to **info@raponline.org**.

Table Of Contents

Acronyms

CA ISO	California Independent System Operator
CO_2	Carbon Dioxide
ERCOT	Electric Reliability Council of Texas
GHG	Greenhouse Gas
GW	Gigawatt
IEA	International Energy Agency
IGCC	International Grid Control Cooperation
ISO	Independent System Operator
LMP	Locational Marginal Pricing
NYISO	New York Independent System Operator
RES	Renewable Energy Source

List of Figures

Introduction[1]

Major market economies have now articulated specific targets for reducing greenhouse gas (GHG) emissions: Europe remains committed to economy-wide reductions in GHG emissions of 80 percent to 95 percent below 1990 levels by 2050, and in October 2014 the European Council agreed on a 2030 emissions reduction target of at least 40 percent compared to 1990. Under a recent bilateral agreement with China, the US government has agreed to reduce economy-wide carbon dioxide (CO2) emissions to between 26 percent and 28 percent below 2005 levels by 2025. California, the eighth largest economy in the world, has committed to reducing GHG emissions to 1990 levels by 2020, with the longer-term goal of reducing emissions to 80 percent below 1990 levels by 2050.

It is well understood that rapid progress in the power sector is essential to achieving these objectives. But the power sector cannot sacrifice reliability in the process, and fortunately it need not do so: multiple authoritative studies have described power sector decarbonisation pathways based on existing technologies that can meet or exceed current reliability standards. And, as importantly, a secure, reliable transition to a low-carbon power system can be accomplished at a reasonable cost — indeed, at a cost little more than, and very possibly less than, the cost of "business as usual."

Framing solutions at a regional level offers governments one of the best options for maintaining security of supply at a reasonable cost. Market design, market rules, and market operations lie at the heart of this process, particularly in those regions where the power system has been organized around competitive markets. The intersection of energy and climate policy has spawned a lively discussion about markets and security of supply, a discussion that has tended to focus too narrowly on traditional notions of, and solutions for, the reliability challenge. Rather than begin from the question of which market design is best able to deliver a given quantity of investment, recent studies have illuminated the importance of asking first what kind of investment is best suited to the needs of a low-carbon power system. The least-cost reliability solution will be delivered not by a market that perpetuates investment in "more of the same" but rather shifts investment from a legacy resource mix dominated by inflexible baseload generation to one that can efficiently complement production from a growing share of variable resources.

The debate over energy-only vs energy-plus-capacity markets is important, but to some extent it misses the point. Both models, when implemented well, can ensure reliability, and each carries significant risks. But whether the cost of a low-carbon reliability solution is reasonable or costly will ultimately depend on whether the resulting mix of market instruments, market governance, and regulation adequately captures the need for an increasingly flexible system. A more comprehensive discourse is needed about how best to structure markets and pricing mechanisms (including those for renewable resources) to achieve climate, security, and economic goals during the transition to a low-carbon power system. This paper is meant to offer a starting point for that discussion.

1 An earlier version of this paper was written by the Regulatory Assistance Project (RAP) for Agora-Energiewende and is an extension of RAP's previous *Beyond Capacity Markets* work in Europe and the United States. It was presented for discussion purposes at a 16 October 2014 meeting of the Pentalateral Energy Forum, whose current members are Austria, Belgium, France, Germany, Luxembourg, and the Netherlands (with Switzerland as an observer). This paper adapts that document for a wider audience and represents solely the views of RAP. Lead RAP author: Michael Hogan, in close collaboration with Rick Weston and Meg Gottstein. Markus Steigenberger and Christian Redl of Agora-Energiewende also provided valuable input during the development of the earlier version of the paper.

1. Managing an Orderly, Low-Cost Transition

A low-carbon power system must continue to be a reliable power system in order to meet the expectations of industry and consumers. It should also continue to be an affordable power system for all and one that underpins economic competitiveness. As noted previously, significant GHG emissions reduction targets — and thus the implied drive toward a low-carbon power system — are currently enshrined in policy and legislation in many countries and regions, but it cannot be taken for granted. If the costs of meeting established reliability expectations were to rise significantly, the entire undertaking could come under intense political pressure. Fortunately there is a growing body of expert analysis targeting these important challenges, a growing consensus about the steps that can be taken — immediately — to deal with them, and more and more real-world experience to justify confidence in their efficacy. Although some of the strategies discussed here will involve a front-loading of costs with benefits accruing over time, each of them can be expected to contribute to reducing the overall cost of ensuring reliability in the transition to a low-carbon power system.

Under any realistic scenario, the low-carbon power system will rely on a significant share of renewable energy sources (RES), dependent predominantly on wind and solar radiation, the availability of which are variable and uncontrolled. As variable resources become major players in the energy mix, the cost and complexity of maintaining reliability can vary greatly depending on how the design and operation of the overall system evolves in response to the changing resource mix. It is now possible to envision a number of low-cost pathways based on different combinations of a set of mutually reinforcing options. One critical step is a deliberate investment shift toward more flexibility in the portfolio of generating resources — less inflexible baseload, more flexible mid-merit units. Other highly beneficial options include: the coordination of unit commitment, economic dispatch, and balancing over larger geographic areas; tighter integration of day-ahead, intra-day, and balancing market operations; incorporation of and frequent intra-day updating of state-of-the-art wind and solar forecasts in unit commitment and grid operations; and full participation of dispatchable demand-side flexibility (including demand-side energy storage) into balancing services[2], energy, and (if applicable) capacity markets.

Every one of these steps is technically and economically feasible today and, to varying degrees, can be observed in operation in competitive markets around the world. In fact, these measures interact with each other in important and positive ways. For instance, operation of balancing markets over wider areas reduces the need for added resource flexibility and vice versa. The question for decision-makers, then, is not whether the transition can be achieved reliably and at a reasonable cost, but rather which policy "levers" they choose to pull and in what combination to create such a pathway.

2 Balancing services are a suite of actions available to system operators to remediate imbalances on the system after trading on the energy market is closed shortly before real time. In some regions they are referred to as "ancillary services."

2. Evolving Solutions for Reliability During the Transition

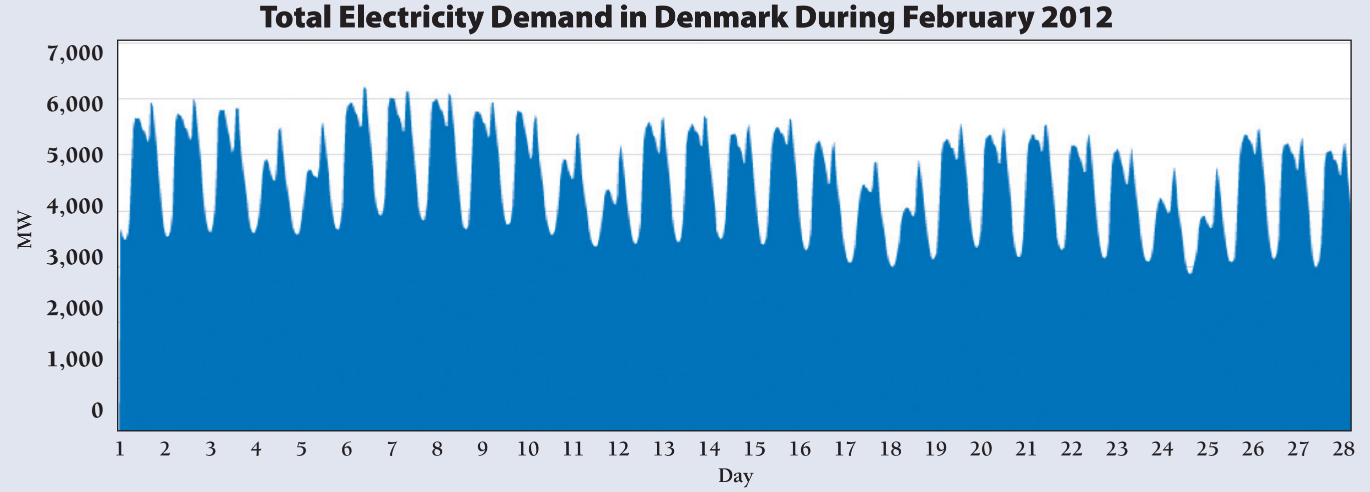
The central challenge facing energy policymakers and system operators is that of ensuring the availability of sufficient resources to meet demand for service at virtually all times and at a reasonable cost. This requires investment in both a quantity of resources as well as a least-cost mix of resource capabilities, a function that throughout most of the last century was carried out by central planners. Over the last 30 years, forces of economic and technological change, and environmental and public health policy, have transformed the energy landscape. As a result, the power sector is transitioning to one whose mix of resources and means of operation will differ greatly from that of the last century. This is leading to a reassessment of how best to ensure a reliable, least-cost power system. In other words, in the 21st-century power system, the question of *reliability* will remain in the forefront, but the nature of the solution must change as the penetration of variable[3] production increases.

A. Resource Adequacy is Not, and Never Has Been, Only About Capacity

Service reliability is established in two dimensions: (1) an *operational* dimension (typically referred to as *system security*), in which a combination of available resources is deployed to match expected demand in real time at the lowest reasonable cost; and (2) an *investment* dimension (typically referred to as *resource* or *generation adequacy*), in which investment is required to maintain, refresh, expand, and transform the portfolio of resources so that they will continue to be available as needed to meet future demand at the lowest reasonable cost. The growing reliance on variable renewable resources fundamentally transforms the system security dimension, placing greater emphasis on the ability of the remainder of system resources to complement renewable production efficiently and reliably. This can be seen clearly in Figures 1 and 2.

Figure 1 shows gross demand on the Danish system in February 2012. Figure 2 shows net demand (gross demand less the contribution from zero-marginal-cost

Figure 1



Source: energinet.dk

3 The term "variable" used here refers to any generator whose ability to produce electricity — how much and when — is beyond the control of operators to a significant degree. The technical term often used for this is "intermittent."

Figure 2

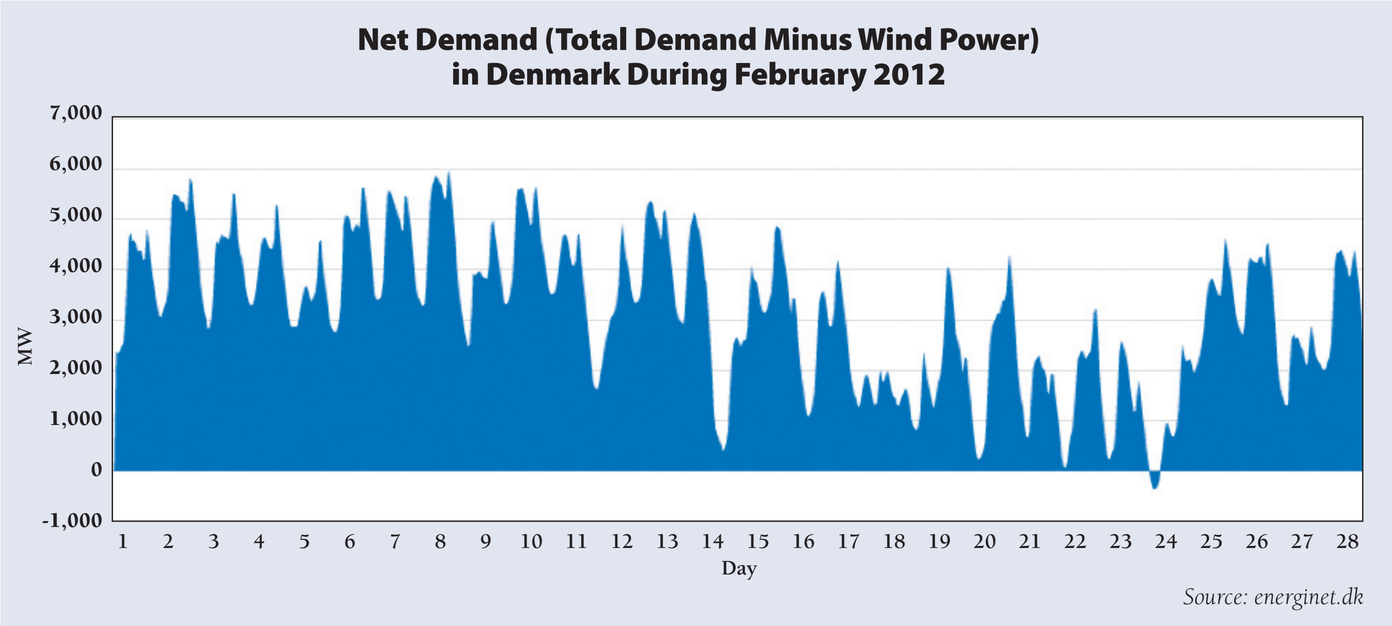


Figure 3

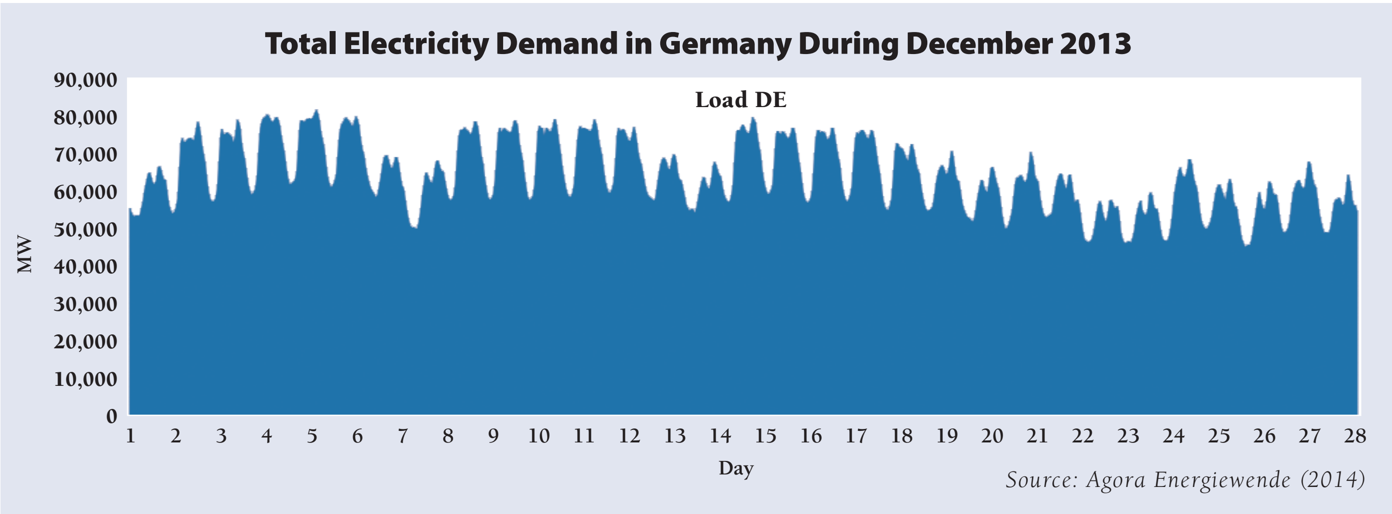


Figure 4

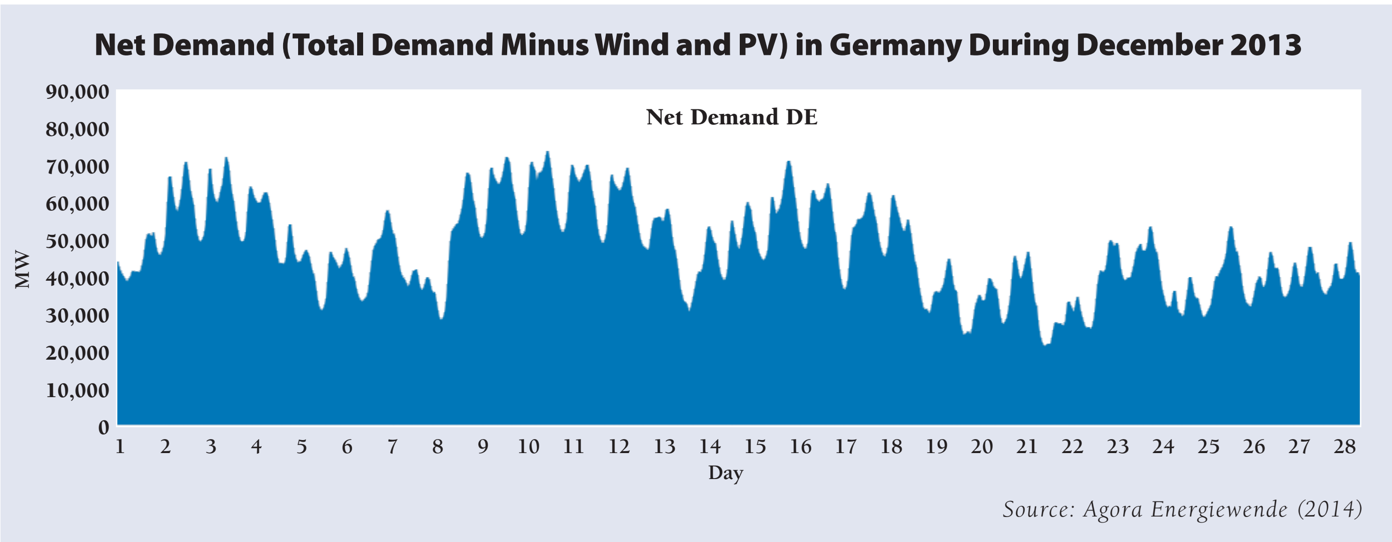
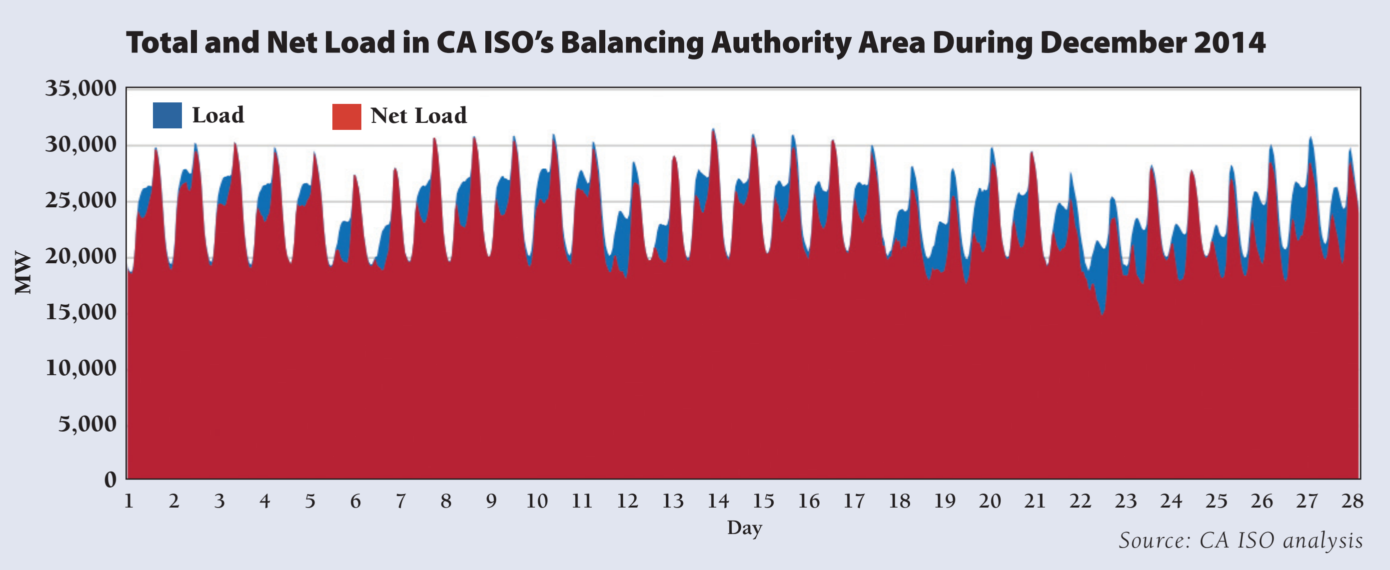


Figure 5



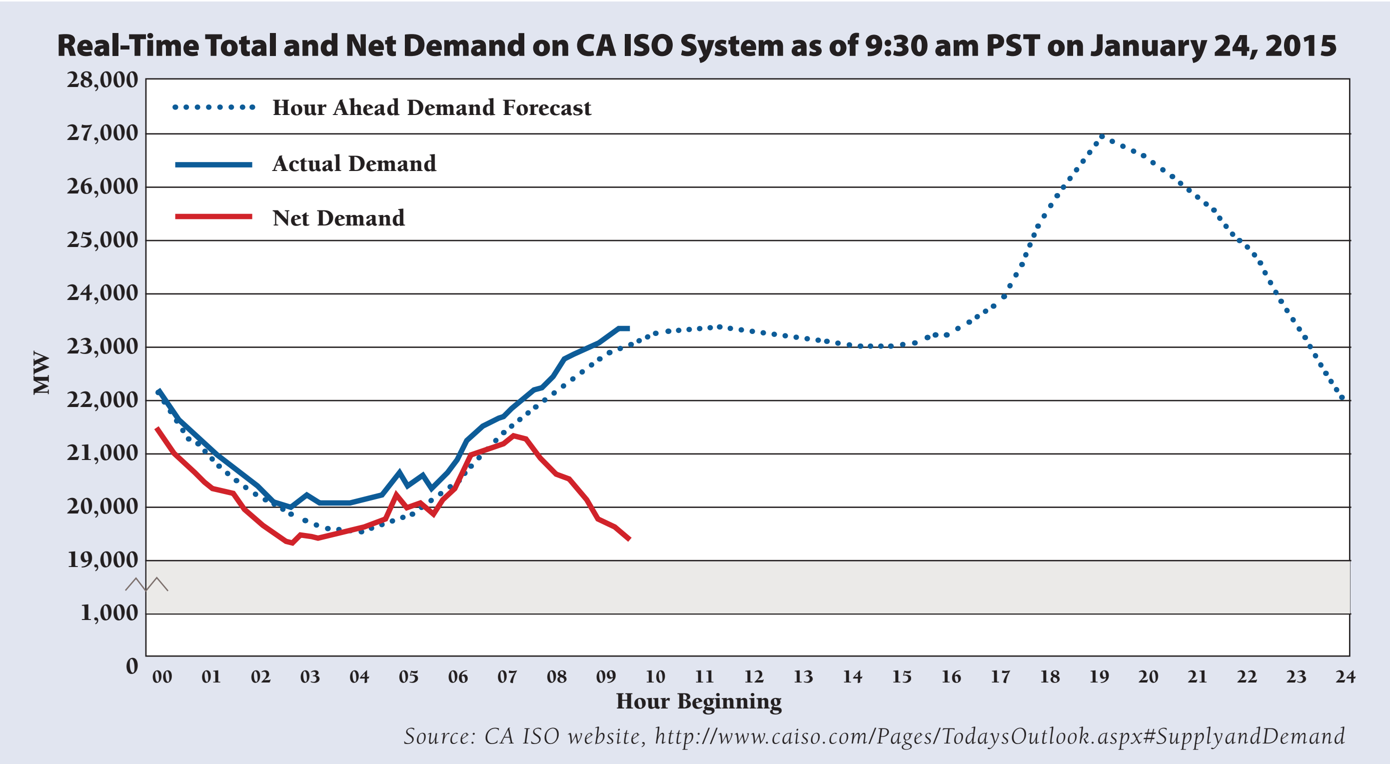
renewables) over the same period, net demand now representing the task facing dispatchable[4] resources.

Similarly, Figures 3 and 4 show gross and net demand prevailing in the German power system during December 2013. Although the daily spread between peak and minimum gross demand amounted to some 20 gigawatts (GW), the spread between peak and minimum net demand amounted to some 40 GW during the observation period.

Figure 5 provides another illustration of gross and net load (i.e., demand) for the month of December 2014 in the balancing authority area operated by California's Independent System Operator (CA ISO).[5]

In Figure 6, one can see more clearly the hourly fluctuations between gross and net demand that occur

Figure 6



4 "Dispatchable" refers to system resources that are generally available to be turned on or off, down or up as and when needed, within limits specific to each resource.

5 Approximately 80 percent of California's end-use demand is within the CA ISO balancing authority area. The values reflected in Figures 5 and 6 exclude demand that is offset by behind-the-meter installations, such as rooftop solar panels, to which the CA ISO has no visibility. Source: CA ISO.

on the CA ISO system — in this case, during the early morning hours of 24 January, 2015.

One implication of the situation captured in these figures is that a more flexible mix of resources, including both supply- and demand-side resources, capable of shifting up or down in synch with the less controllable shifts in variable renewable production, will have far higher asset utilization rates and require far less redundancy (and therefore far less investment) than a less flexible mix of resources.[6] Although resource adequacy has never been only a matter of the quantity of resources, now more than ever the answer to the question "How much?" depends on the answer to the question "What type?"

B. Capacity and Capabilities

Where generation and supply are competitively provided, decisions about when to invest, how much to invest, and what to invest in are, in principle, left to the market. A debate is underway in many markets about whether to rely solely on energy markets or to adopt some combination of energy markets and capacity remuneration mechanisms. Where capacity mechanisms are proposed, they are often designed as "single product" mechanisms addressing only the quantity of capacity, based primarily on the claim that the need to invest in resource flexibility can be left to the energy market. The problem with this is that the price signals missing from the energy market to drive investment in capacity are the same price signals the energy market is meant to use to remunerate investments in greater resource flexibility. In both cases investors in energy markets are meant to rely on the expression of "scarcity value" in the pricing of energy and balancing services to do so. That is, as demand approaches the limits of the options system operators have to meet it, the value of energy can increase beyond the short-run marginal cost of generation, reflecting a combination of the increasingly costly actions taken by the system operator and the value to consumers of uninterrupted service.[7]

The principles underlying the theory of competitive wholesale electricity markets rely on individual instances of scarcity pricing to express a shortage of responsive resources in the short term and the accumulation of scarcity pricing incidents, when they become sufficiently frequent, to express a shortage of investment in such resources in the long term. Shortage or scarcity value is often misunderstood to mean infrequent price "spike" events appearing abruptly and at extreme levels. In fact as will be discussed in Section 5 below, shortage value typically emerges often and gradually such that with timely information market participants can respond to emerging shortages without the need for extreme price spikes. However, in many markets **scarcity value is suppressed through administrative interventions[8] or poor market implementation.** Where this is the case the absence of proper scarcity pricing devalues investment in *both* firm capacity *and* increased operational flexibility (or "capabilities"). Where such administrative distortions are allowed to persist, some form of supplemental mechanism in support of investment may be appropriate.

Such supplemental mechanisms have been implemented or proposed in a number of markets in what are often referred to as capacity remuneration mechanisms, or capacity markets. These mechanisms have typically been designed to address only the quantity of MW of generating capacity (or an equivalent quantity of demand reduction) available to system operators. It is becoming clear that assurance of a given quantity of MW of capacity on a system is insufficient to assure security of supply at a reasonable cost. In addition to the simple quantity of capacity it is necessary to assess the operational attributes of the available capacity. The

6 See, e.g.: International Energy Agency. (2014). *The Power of Transformation;* US Dept. of Energy/National Renewable Energy Laboratory. (2012). *Renewable Energy Futures;* North American Electric Reliability Corp. (2009, April). *Accommodating High Levels of Variable Generation;* Hinkle, Pedder, Stoffer, et al, GE Energy. (2011, 8 March). *Contributions of Flexible Energy Resources for Renewable Energy Scenarios;* European Climate Foundation. (2012). *Power Perspectives 2030.* McKinsey & Co., KEMA, Imperial College London, Regulatory Assistance Project, E3G.

7 Until we can enable more active involvement of customers in purchasing decisions, the security constraints used by system operators tend to serve as a proxy for the value of lost load; system operators in the more advanced energy markets are translating these security constraints into scarcity pricing in the day-ahead and intra-day energy and balancing markets.

8 In some cases such interventions are appropriate to mitigate the potential for abuse of market power, and where market power continues to be a legitimate concern it will limit the scope for reliance on unconstrained energy market pricing. A determination that there is effective competition and the introduction of effective market monitoring are necessary preconditions for the removal of these market power mitigation measures.

European Commission, in recent guidance to member states regarding capacity mechanisms[9], suggested that the existing approach to assessing security of supply "may be insufficient to tackle the challenges of the future in a fully satisfactory way."

To make the point more directly, the North American Electric Reliability Corporation, a leading authority on power system reliability, has stated explicitly that resource adequacy cannot be determined by measuring capacity as an undifferentiated commodity, but rather the adequacy of system resources can only be determined with reference to their operational characteristics.[10] Similarly, the Council of European Energy Regulators has recently issued recommendations for adequacy assessments including the necessity to consider explicitly flexibility, resource needs disaggregated by time period, and demand-side flexibility.[11]

That this is crucial is borne out repeatedly in power system experience, where the great majority of generation-related system reliability events occur during periods when total de-rated ("firm") generating capacity on the system comfortably exceeds total demand.[12]

In short, one cannot ensure resource adequacy by intervening in the market to support investment in capacity indiscriminately without also addressing the fact that the very same "missing" scarcity value also distorts the relative value of more flexible capacity. On the contrary, as the share of variable renewable production increases, a single-product capacity mechanism may only reinforce the mismatch between the inflexibility of the current portfolio and what will be needed to ensure least-cost system security going forward.

C. Major Consequences for "Adequate" Investment Levels

Leaving this problem to be addressed later will lead to poor asset utilization and an unstable investment environment, necessitating additional investment costs for consumers that could have been avoided. The consequences of failing to value resource flexibility fully and in a timely fashion were addressed in the recent International Energy Agency (IEA) study on renewables integration.[13] In analysing a system well advanced in the low-carbon transition, the study highlighted dramatic differences between a system in which the mix of resources shifts in response to the growing role of variable renewables and one that continues to invest in "more of the same." The results are captured in the graphs in Figure 7 for a system undergoing a transition from 0 percent to 45 percent variable renewables.

The graphs depict a system in which the share of variable RES has grown to 45 percent, under two scenarios. In the "Legacy" scenario, the incumbent mix of thermal generation capacity (baseload, mid-merit, and peak) has remained essentially unchanged through the transition. Most of the non-renewable energy production comes from inflexible baseload plants, with flexible mid-merit plants producing a much smaller amount. Baseload plants that traditionally saw capacity factors in the 90 percent range are now running only 62 percent of the time, whereas mid-merit plants that typically ran approximately 40 percent of the time are now seeing only 11 percent capacity factors, in both cases insufficient to support investment without some form of supplemental

9 European Commission. (2013, 5 November). *Generation Adequacy in the Internal Electricity Market - Guidance on Public Interventions* [Commission staff working document]. Retrieved from http://ec.europa.eu/energy/sites/ener/files/documents/com_2013_public_intervention_swd01_en.pdf

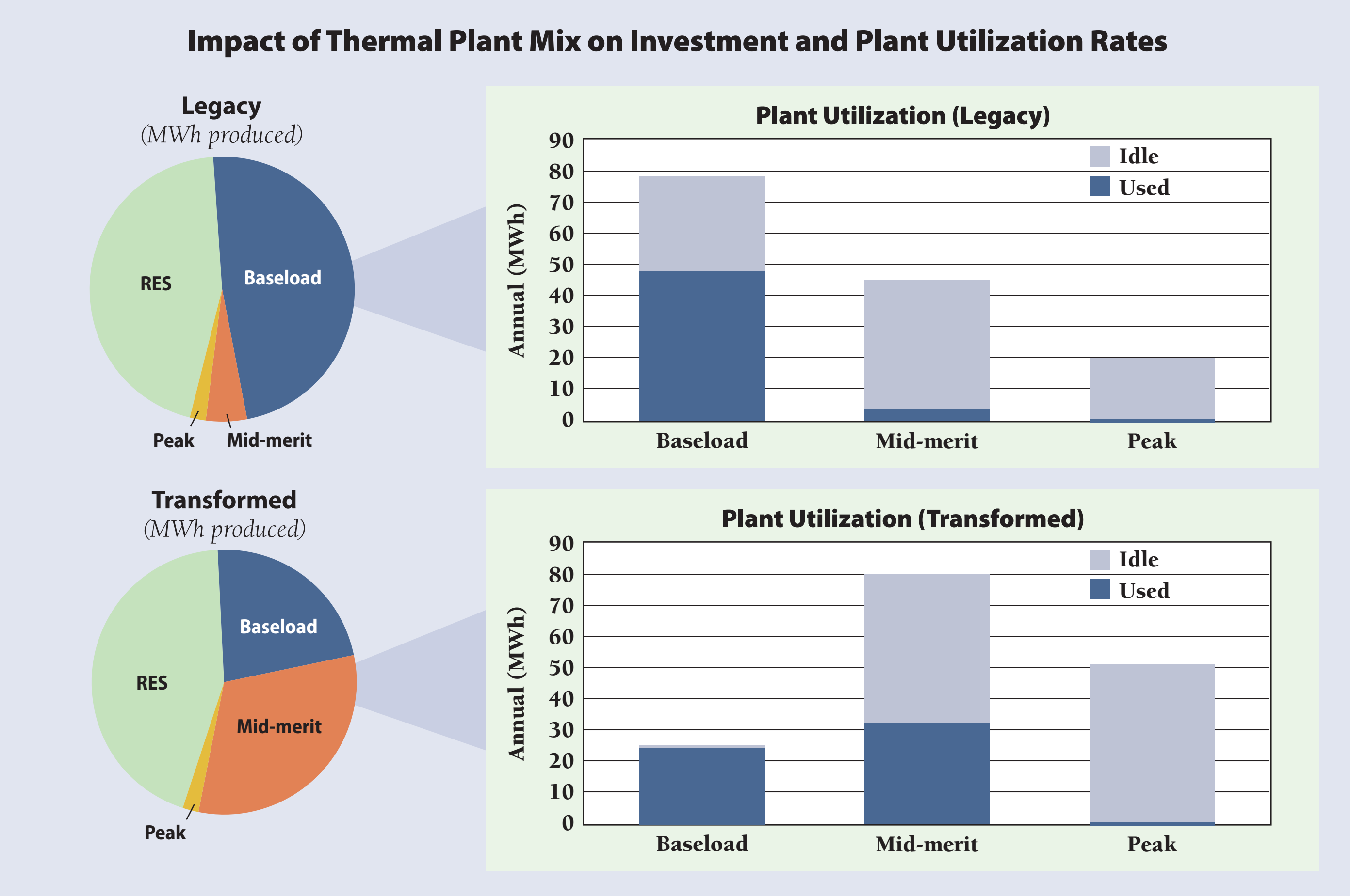
10 NERC. (2011, 26 January). *Balancing and Frequency Control,* pp 40-41. Retrieved from http://www.nerc.com/docs/oc/rs/NERC%20Balancing%20and%20Frequency%20Control%20040520111.pdf

11 CEER. (2014, 8 October). *Recommendations for the Assessment Of Electricity Generation Adequacy.* Ref: C13-ESS-33-04. Retrieved from http://www.ceer.eu/portal/page/portal/EER_HOME/EER_PUBLICATIONS/CEER_PAPERS/Electricity/Tab3/C14-ESS-33-04%20Recommendation_Generation%20Adequacy%20Assessment_final_081014.pdf

12 An oft-cited claim of supply shortage occurred in Germany in February 2012, when unexpectedly high demand forced system operators to exercise all balancing options. The Ministry of Economics has recently confirmed [BMWi, 2014: *Ein Strommarkt für die Energiewende. Diskussionspapier des Bundesministeriums für Wirtschaft und Energie* (Grünbuch)] that the system had more than enough capacity at the time, but because of distorted incentives in the balancing and energy markets suppliers deliberately under-nominated day ahead, leaving the system operator with too few available resources committed on the day. Data from ERCOT, the market operator in Texas, show that between 2006 and 2011, 75 percent of supply-related emergencies took place outside of the peak demand season and nearly half occurred during non-peak hours.

13 IEA. (2014, February). *The Power of Transformation: Wind, Sun and the Economics of Flexible Power Systems,* pp. 162-164.

Figure 7



assistance. This is just the sort of dire picture often painted of a system with high shares of variable RES. What the IEA analysis and other recent analyses demonstrate is that there is nothing inevitable about this outcome. In the "Transformed" scenario, the mix of thermal resource types has been re-balanced in response to the growth in variable resources, with investment shifting to more flexible plant. Slightly more energy is now produced by flexible mid-merit plant than by baseload plant. The remaining baseload plants are back at more than 90 percent of capacity, whereas the mid-merit plants are back to approximately 40 percent of capacity.

Taking the analysis further, these results imply much different levels of investment. Using conservative assumptions for the cost to build these types of plant,[14] the Transformed scenario delivers the same amount of energy to the same reliability standard, but with more than 40 percent less investment required to do so. **In short, a more flexible mix of dispatchable resources, capable of shifting operations up and down in synch with the less controllable shifts in variable renewable production, will have far higher asset utilization rates and require far less redundancy (and therefore far less investment) than a less flexible mix of thermal resources.**

14 We assumed an average of €3,500/kW new-build costs for baseload, €1,300 for mid-merit, and €350/kW for peaking.

3. Addressing the Investment Challenge – Conventional Generation

Reliability, then, rests on a foundation of investment in system resources. The adequacy of that investment derives from *both* an adequate quantity of resources *and* an adequate mix of resource capabilities (or operational attributes). In competitive electricity markets there are broadly two approaches to delivering an adequate portfolio of resource investment: "energy-only" markets that rely on the outturn prices for energy and balancing services to drive investment, and energy markets to which an administrative mechanism has been appended that is meant to set a value for capacity over some future period of time.

Energy-only markets are capable in theory of delivering an economically efficient resource adequacy solution, but there are risks inherent in relying solely on energy-only markets, and often, in practice, shortcomings in the implementation of such markets. Implementation problems can include price-distorting power mitigation measures, such as price caps, necessitated by a failure to limit and monitor market power adequately, or energy pricing that is insulated from supply-demand conditions in the balancing market. Capacity mechanisms can create greater certainty around the availability of a given level of resource investment, but as administrative mechanisms they carry their own implementation risks. Particularly in the context of the low-carbon transformation, these risks include setting the capacity margin higher than is actually needed, as well as the ability or willingness of market administrators to capture sufficiently the portfolio of resource attributes best suited to assuring at reasonable cost the resource adequacy these mechanisms are intended to address. Put simply, given the practical constraints on administrative solutions and the human and political factors that come into play, a capacity mechanism can lock in too much investment, and in the wrong mix of resources, which in turn can lead to yet more overinvestment and needless escalation of the costs of the transition.[15]

Reasonable people can disagree about the wisdom of and need for adopting capacity mechanisms. We do not take a position on that question here. Whichever route is ultimately chosen, the important question remains the same: Is the market driving investment toward a portfolio of resources best suited to provide an acceptable level of reliability at least cost as we decarbonize the power sector? There is now sufficient experience in markets around the world with both approaches to draw some valuable insights regarding their proper place.

A. The "No Regrets" Approach

Frustration with the practical and political challenges of perfecting the operation of energy-only markets has led in a number of market areas to the adoption of capacity mechanisms as a way of reducing the risk of under-investment. In virtually every instance these have initially been designed as "single product" mechanisms that treat capacity as an undifferentiated commodity, both in order to reduce complexity and in the belief that the energy market will direct investment toward the right mix of resources. Reducing unnecessary complexity is laudable when designing administrative mechanisms, especially ones as inherently complex as these, but as we have already discussed, there are fundamental flaws in

15 Although beyond the scope of this paper, there is a growing body of academic analysis demonstrating that reliability standards traditionally applied in many market regions have little or no objective basis in economic analyses of the value of reliability. See, e.g.: Brattle Group. (2012, 1 June). *ERCOT Investment Incentives and Resource Adequacy,* page 100 and following. As a result, resource adequacy assessments have often led to questionable conclusions about the required level of investment and the performance of energy-only markets in delivering it. The resulting tendency to use capacity markets to lock in uneconomic capacity investments is of particular concern in a low-carbon power system.

the assumption that an energy market deemed incapable of delivering the right *quantity* of resource investments can nonetheless be relied upon to deliver the right *mix* of resource investments.

As a consequence, in markets that have accumulated the longest experience with capacity mechanisms, authorities are finding it necessary to revisit this issue. Several have recently begun to introduce differential levels of compensation for different types of resources based on their operational capabilities, even in markets where the penetration of variable resources remains relatively modest.[16] As the share of production from variable resources becomes more significant, with the attendant transformation in the optimal mix of conventional resource capabilities, the shortcomings of single-product capacity mechanisms become more and more apparent. As various markets consider the merits of adding capacity mechanisms to their power markets at the same time as they incorporate more and more variable resources, capacity will become an increasingly differentiated product, and market interventions will have to reflect that in some fashion if they are to continue to serve their intended purpose of ensuring resource adequacy at a reasonable cost.

Although the relative simplicity of single-product capacity mechanisms is ultimately unsustainable, there will be practical constraints, such as the need for adequate liquidity, on the degree of complexity that can or should be designed into these administrative mechanisms. Design complexity should also be limited by the recognition that the search for precision will quickly overtake the capacity of administrators to determine with any certainty what the right future mix of resources will be. For this reason, the responsibility for shaping investment toward the optimal mix of resource capabilities can never be left entirely to capacity mechanisms. However beneficial a well-designed capacity mechanism might be in reducing uncertainty, it will always be a comparatively crude administrative tool. Although it is important to harness such mechanisms in support of the need to shift investment toward more flexible system resources, they cannot fully replicate the role energy-only markets can and should play in doing so. Once again, in many of the more advanced wholesale electricity markets there is evidence that this is the case. Aggressive reforms designed to improve scarcity pricing in both energy and balancing services markets have recently been introduced in several large competitive markets, not only in energy-only markets but also more notably in markets firmly committed to the use of capacity mechanisms.[17]

The need to ensure a least-cost mix of resources drives the need to improve the effectiveness of energy and balancing market price signals. The "no regrets" option is to redouble efforts to bring the operation of energy markets more into line with their theoretical potential, regardless of what decision might be taken about capacity mechanisms.

Should a capacity mechanism be adopted, the hard-won lessons of real-world experience tell us that it must sooner or later be designed to recognize the difference in value between different types of resources in delivering least-cost reliability. As more variable resources enter the system, the value of discriminating explicitly in favour of operational flexibility will be impossible to ignore.

B. The Geography of Adequacy

When considering how much of a margin in capacity resources over and above demand is enough, the issue of cross-border integration must also be mentioned. It is well established that, all else being equal, the quantity of resources required to meet a given resource adequacy standard is reduced as the size of the market in which it is applied (in terms of both area and demand) is increased. The broad benefits of true cross-border market integration to low-cost decarbonization are dealt with later in this paper, but it is worth considering here that assessments of resource adequacy within artificially delimited footprints (e.g., within political borders) will

16 For information on ISO New England's January 2014 proposal, see http://isonewswire.com/updates/2014/1/22/spi-news-iso-ne-submits-proposal-to-strengthen-performance-i.html; for a description of PJM's August 2014 proposal, see http://www.pjm.com/~/media/documents/reports/20140820-pjm-capacity-performance-proposal.ashx.

17 For a brief description of shortage pricing improvements in the PJM market, see http://www.pjm.com/~/media/about-pjm/newsroom/fact-sheets/shortage-pricing-factsheet.ashx; for the UK market, see https://www.ofgem.gov.uk/electricity/wholesale-market/market-efficiency-review-and-reform/electricity-balancingsignificant-code-review; for the NYISO market, see http://www.nyiso.com/public/webdocs/markets_operations/committees/bic/meeting_materials/2014-11-12/agenda_09_Comprehensive%20Shortage%20Pricing.pdf

inevitably lead to the need for more investment than if security of supply and resource adequacy were managed over larger geographic footprints. Although this has always been true to some extent, recent studies have demonstrated that the benefits of geographic aggregation become overwhelming where there are significant shares of variable resources in the relevant market areas.[18]

The greatest benefit accrues where responsibility for system security and resource adequacy is vested in a single balancing authority across the relevant footprint. Where that is not possible for whatever reasons, much of the benefit can be realized through the adoption of market mechanisms designed to integrate unit commitment, dispatch, and balancing operations in real time among multiple balancing control areas. This latter option will effectively have been implemented in the EU if and when balancing markets are fully coupled, complementing the progress already made in coupling energy markets.

Action by groups of US states or European member states would do much to demonstrate the benefits and accelerate progress in other regions. As the share of variable resources continues to grow, the alternative becomes increasingly unattractive. Continuing the practice of managing resource adequacy within artificially constrained jurisdictional boundaries will only exacerbate over-investment and raise the cost of reliability over the course of the transition and thereafter.

18 See, e.g.: Booz & Co, et al. (2013, September). *Benefits of an Integrated European Energy Market.* Retrieved from http://ec.europa.eu/energy/infrastructure/studies/doc/20130902_energy_integration_benefits.pdf; NREL. (2010). *Combining Balancing Areas' Variability.* Retrieved from http://www.nrel.gov/docs/fy10osti/48249.pdf; NREL. (2013). *Examination of Potential Benefits of an Energy Imbalance Market in the Western Interconnection.* Retrieved from http://www.nrel.gov/docs/fy13osti/57115.pdf; and IEA. (2014). *The Power of Transformation.* Retrieved from http://www.iea.org/w/bookshop/add.aspx?id=465.

4. Addressing the Investment Challenge – Renewable Energy Sources

The challenge — and imperative — of continued investment in RES presents a special case. All prudent power sector decarbonization pathways rely on sustained commercial deployment of renewable generation at some significant level. Among the decarbonization technology options, renewables are unique in the extent of progress made in recent years in addressing performance and cost challenges. Were deployment of renewables to grind to a halt simply because it is incompatible with current market conditions, the dramatic and hard-won progress of recent years in building a commercial industry, so crucial to the long-term competitiveness of a low-carbon energy system, will have been for naught. In recognition of this, the European Union recently renewed its commitment to RES deployment, adopting a (minimum) 27 percent EU target for RES penetration by 2030 against the current 2020 target of 20 percent. This translates to a share of 45 percent to 53 percent[19] renewables in the power sector by 2030 compared to an expected 34 percent by 2020. Many other jurisdictions are similarly adopting or updating ambitious commitments to RES deployment.

Yet despite the dramatic progress made over the past two decades, investment in renewables — in particular variable renewables — still faces challenges in the current market environment. In many cases the total cost of renewable power is at or near (in some cases even below) the level of the average prices with which they should be expected to compete, but other factors come into play. Variable renewables selling into the energy market tend to earn less than the average market price because prices tend to be lower during periods when renewable energy is most available and higher when renewable energy is less available. In contrast, more dispatchable resources have the opportunity to earn higher prices during periods when variable renewable production is low and demand is high. Known changes in market operation and the overall portfolio of system resources — which could reduce overall price volatility and balancing costs to the benefit of all investment, not just variable renewables — have been slow in coming and will take years to fully materialize, where they are currently being pursued at all. (See Section V for a brief discussion of some of these opportunities.) Therefore as the "booster stages" of deployment support near the end of their role for many well-developed renewable technologies, but with some external hurdles yet to be overcome, there are still unanswered questions as to what forms the "intermediate stages" of support will or should take. Two key areas to explore are examined below.

A. Rethink Deployment Support Policies

Because of the variable nature of most of the primary renewable sources, only a limited percentage of their capacity can be relied upon to be always available and therefore they are not likely to benefit significantly from capacity-based interventions, at least not in any of the forms currently used or under serious consideration. As for support specifically for renewables technology investments, most recent proposals are still essentially subsidy mechanisms designed to bring market signals more strongly into play, including auctions and "feed-in premiums" as well as variations on quota-based tradable certificate programs.

It may be better to look at moving more clearly beyond a subsidy paradigm. As the costs of many key RES technologies have declined, the need for support to compensate for above-market costs has declined as well and, with the expectation of higher carbon prices and yet further cost reductions, should continue to decline.

19 EC. (2014). *Impact Assessment Accompanying the Communication. A Policy Framework for Climate and Energy in the Period From 2020 up to 2030.* The RES-E range mentioned refers to the scenarios with a GHG reduction of 40 percent.

"Priority dispatch" was intended to protect variable RES from curtailment for convenience, but with RES now well established in most markets, its very low short-run marginal cost means it is likely to be dispatched virtually whenever it is available. Despite arguments for and against priority dispatch on both sides of the RES debate, it is not entirely clear what incremental impact (in either direction) priority dispatch has at this stage.

Improved market operations and more responsive demand should also reduce the need for insulation of variable RES from curtailment and balancing risk, which will be a critical factor supporting the trend in policy toward increasing RES exposure to balancing responsibility. As described previously, however, the challenge of earning disproportionately low energy prices will continue for some time into the future, essentially setting the bar for RES investment higher today and through the medium term than it is likely to be once the system is better able to absorb the swings in variable renewable production. During this transition period the support required for RES investment will increasingly have less to do with subsidy paradigms and more to do with alternative revenue mechanisms that enable variable RES to realize close-to-average market prices. There is creative thinking yet to be done on this front.

B. Actively Manage the Impact on Supply and Demand

One additional challenge facing investment in all resource categories, including RES as they become more exposed to market price levels, is the trend in developed economies toward overcapacity and stranded assets as policy-driven deployment of RES capacity goes forward. RES support policies, particularly feed-in tariff schemes, have until now tended to promote investment in new renewable capacity without regard to the resulting impact on the balance between the demand for productive capacity and the supply of productive capacity and with no obvious plan or policy to deal with any surplus or stranded capacity that might be created as a result. The substantial oversupply of capacity now seen in many markets, the result of the combined effects of the deep recession, successful efficiency measures, ill-timed investment in new fossil-fired plants, and aggressive renewable support mechanisms, is one of the primary causes of the instability currently plaguing the conventional generating sector in many markets, particularly in Europe, although it is certainly not the only cause.

This should be used as an opportunity rather than endured as a crisis. As policy-driven investment in RES continues to grow at a rapid pace in many markets, faster than the growth in demand, surplus capacity steadily accumulates, with the legacy share of inflexible baseload generation increasingly ill-suited to the needs of the system. There should be a plan to remove the resulting surplus in baseload generating capacity from the market in an orderly and equitable fashion. This can create a more stable marketplace for those resources remaining on the system and facilitate the transition to a more appropriate mix of resources.

Unfortunately in some regions the onset of oversupplied markets seems to have come as a surprise, and the belated response from some governments has been to consider supporting redundant generation financially through fixed capacity-related payments. This may create challenges for continued support for RES investment. Artificially depressed energy prices inflate the perceived gap between average wholesale power market prices and the prices paid for RES supply, generating unwanted political blowback against RES support policies. The failure to assist redundant baseload generation in exiting the market also impedes the critical transformation of the conventional generation portfolio into a smaller, more nimble fleet dominated by flexible mid-merit generation.

The future policy framework around renewable deployment, whatever form it takes, needs to do a better job of dealing with this issue. In particular, policymakers and regulators need to consider how best to deliberately and selectively remove surplus baseload thermal generating capacity from the market.

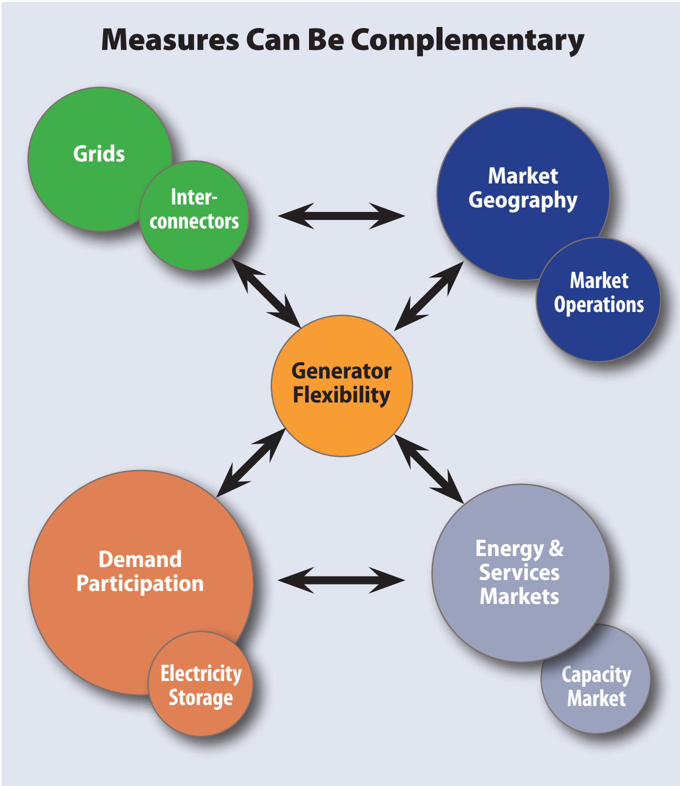
5. Market Structure, Market Rules, Market Governance

Section 3.A addresses the importance of improving the effectiveness of wholesale electric energy markets regardless of whether or not a capacity mechanism is appended to the market. If and when introduced, capacity mechanisms need to be "smart" — they need to recognize that flaws in the implementation of energy markets that risk underinvestment in capacity will also risk underinvestment in operational flexibility — but they cannot fully substitute for improvements in the functioning of energy markets. This section will look at practical options for making energy markets more effective.

The basic failing of many energy markets is often referred to as "missing money," referring to income required to support needed investment that is not available in the energy market. Typical causes of "missing money" are various forms of price suppression, either through administrative interventions (such as price caps), lock-in of an oversupply of production capacity (for example via capacity markets with excessive reserve margin requirements), or poorly designed market rules (such as balancing services mechanisms that do not reflect the real-time value of the service). This "missing money" problem, where it exists, affects both the quantity and the capabilities of capacity resources. There are other obstacles to the effectiveness of energy markets that are not so much failings as they are missed opportunities. These missed opportunities include the failure to exploit non-traditional resources such as dispatchable demand-response that can compete very favourably with generation, as well as advances in market administration (discussed below) that can more effectively absorb the impacts of variable production.

Measures are available to governments, regulators, and system operators to redress these failings, and these measures can interact in a complementary manner to improve the flexibility of wholesale markets. The value of each of the system flexibility options described throughout the paper is interdependent on the extent to which others are implemented (see Figure 8). Some are designed to restore the expression of scarcity value and, at the same time, to make it more reliable, less volatile, and less extreme. In so doing they provide a more attractive basis for investment and give more robust expression to the value of flexibility. Other measures are designed to activate the potential for demand as a flexible resource alongside flexible supply-side resources, critical to enabling consumers to mitigate the impact of more dynamic scarcity pricing. And some measures are designed to reduce the overall need for more flexible resources in the first place. We discuss some of the more prominent examples of each type of measure below.

Figure 8



A. Mitigating the Need to Increase Resource Flexibility

Much has been said and written about the value of increasing the flexibility of dispatchable generation in the transition to a low-carbon power system. Although that is clearly an important factor, an exclusive emphasis on increasing the flexibility of dispatchable generation can lead to an overly costly transition. By implementing a range of comparatively low-cost measures, market authorities can moderate the need for greater generator flexibility and extend the timeframe for transitioning away from the installed base, thereby reducing upfront capital intensity.

1. Larger Balancing Control Areas

Increasing the size of balancing control areas reduces the need for more resource flexibility. Larger control areas are beneficial in any case, but where the share of variable production is significant, the benefit can be especially large. In most cases the size and the frequency of swings between resource surplus and resource scarcity can be reduced dramatically.[20] The benefit derives from three main sources: (1) increasing the size of the control area reduces the impact of any single system event and affords the control area authority a more diverse portfolio of resource options with which to maintain system balance; (2) demand across large geographic areas is generally not well correlated and thus the natural variability of demand cancels out to some extent; and (3) the variability of variable renewable resources is generally not well correlated over large geographic areas, reducing the variability of supply.

The most direct way to access these benefits, and the one that maximizes the benefits available, is simply to consolidate multiple contiguous control areas under a single balancing authority. The regional independent transmission system operator model found in parts of North America and Australia is a good example. Where full integration of area control under a single regional authority is not feasible for whatever reason, much of the benefit can be accessed through virtual consolidation. The integration of balancing markets in Europe as proposed under the Target Model would be a major step toward the real-time consolidation of balancing markets across national borders. An important step toward this objective is the International Grid Control Cooperation (IGCC). The IGCC is an initiative led by the four German TSOs[21] to integrate markets for certain types of reserves across multiple control areas, "… to exploit synergies [as] in a single fictitious control area, without giving up the proven structure of control areas. It also enables a flexible response in case of network bottlenecks."[22] Another example of this is the emerging Energy Imbalance Market in the Western Interconnect of North America.[23]

2. Faster Markets

Yet another way that energy and balancing services markets can be structured to reduce the need for additional flexibility is to make them "faster." Fast energy markets are those in which the dispatching of system resources takes place as close to real time as possible, and where dispatch schedules are updated at multiple points throughout the day based on updated weather forecasts. Frequent rescheduling at shorter market intervals reduces the range of uncertainty about real-time outcomes between dispatch schedules and thereby reduces the need for system reserves. Resources can be dispatched in smaller increments during periods when system net demand is ramping up or down to a significant extent. Uncertainty is further reduced by frequent and more sophisticated weather forecasts. In the most advanced energy markets, the system is dispatched at five-minute intervals based on state-of-the-art weather forecasts that are at most a few hours old, whereas in many traditional markets the dispatching of market resources takes place only once an hour and is based on day-ahead weather forecasts. The need for more flexible system resources acting in fast-response mode can be reduced dramatically by adopting faster market processes.

Centrally dispatched energy markets, common in many regions, are particularly well suited to adopting

20 NREL. (2012). *Energy Imbalance Markets.* Retrieved from http://www.nrel.gov/docs/fy12osti/56236.pdf

21 In addition to the four German control areas, the IGCC currently includes Denmark, Switzerland, Austria, the Netherlands, Belgium, and the Czech Republic.

22 IGCC. (2014). *Information on grid control cooperation and international development.* Version: 10/04/2014. Retrieved from https://www.regelleistung.net/ip/action/static/gcc?gcc=&language=en

23 The EIM is an initiative to integrate balancing markets across nearly 40 contiguous control areas in the western United States and Canada without going to full control area consolidation. It currently includes California ISO, Pacificorp, Sierra Pacific, and Nevada Power.

state-of-the-art fast market processes. Comparably fast markets are more difficult to implement in decentralized dispatch models such as in Europe, but it is still possible to do so. One of the key challenges in decentralized markets is to ensure that the quality and timeliness of information flows between the power exchanges and the system operator is sufficient to allow grid constraints and changes in trading positions to be resolved in the shortest possible amount of time.

B. Tapping the Potential for Demand-Side Flexibility

A number of measures can effectively reduce the need for increased generator flexibility by increasing the opportunity for demand to respond in real time to uncontrolled swings in supply. The keys to accessing this potential are to offer dynamic pricing (preferably real-time pricing) to those wishing to participate and to remove barriers to participation by demand in day-ahead and intra-day energy markets. In many energy markets it has long been possible for large industrial customers to participate directly, although often in very rudimentary ways. But pushing this direct market participation model to the much larger and more diverse pool of residential and small commercial customers is challenging on a number of levels, including the fact that such customers will in most cases have neither the capacity nor the willingness to take action themselves in real time to respond in any reliable or enforceable fashion.

To access this much larger potential it is essential to open energy market access to demand aggregation, in which consumption by a number of individual consumers, or more effectively by individual loads at consumer premises, is managed under contract to a single service provider in return for whatever form of compensation the aggregator and consumers agree. The aggregator then uses the demand under contract to sell the equivalent of energy production into the market. (Aggregators can and do also use demand response to supply various balancing services, a source of flexibility that will be discussed later, as well as the capacity value of demand response, which can often compete successfully with the cost of an equivalent amount of generating capacity.[24])

Aggregation can be carried out by any qualified commercial entity, including competitive electricity suppliers. Experience demonstrates the clear benefits of ensuring that the opportunity to manage consumers' energy services be fully open to competition from both traditional and non-traditional enterprises, meaning that market power must be strictly regulated. One valuable step in this direction, particularly where actual or virtual vertical integration is still a market reality, would be to separate the roles of electricity supplier and Balancing Responsible Party.[25] Demand aggregation is a separate service, entered into at the customer's discretion, in which the service provider essentially steps into the customer's shoes and manages the interface between primary energy supply and the provision of various energy services. There is no good reason suppliers must play this role — although where there is effective competition they should have the right to do so — nor is there a reason they should have to retain responsibility for the balancing issues that may arise as a result. In assuming the role of managing energy services, the aggregator can and should also assume balancing responsibility for the supply procured on behalf of that customer. The current bundling of the roles in some markets creates an unnecessary barrier to market entry by placing service providers (i.e., demand aggregators) in a position of either being in direct competition with incumbent suppliers or being required to negotiate balancing exposure remedies with such suppliers.[26]

Another form of responsive demand can be accessed by making combined heat and power facilities more flexible in response to the needs of the power system. This typically involves the incorporation of thermal energy storage systems so that the provision of heating (or cooling) when demanded by customers can be physically decoupled from the operation of the combined heat and power plant for the production of electricity. This same

24 See: Hurley, Peterson, & Whited. (2013). *Demand Response as a Power System Resource.* Synapse and The Regulatory Assistance Project. Retrieved from http://www.raponline.org/document/download/id/6597.

25 Balancing Responsible Party is the term used in the European Network Codes for that party responsible for any imbalances between the final demand by a given consumer in any given settlement period and the supply committed on that consumer's behalf. Different terminology will be found in other market areas.

26 For a discussion of the roles of suppliers, BRPs, customers, and grid operators in the German context, see: The Regulatory Assistance Project. (2013). *Nachfragesteuerung im deutschen Stromsystem – die unerschlossene Ressource für die Versorgungssicherheit.* http://www.raponline.org/document/download/id/6658

application of distributed energy storage is technically feasible and can be applied inexpensively directly to thermal appliances at customer premises, yet another source of demand flexibility that could be dispatched to match the needs of the power system.

C. Making the Value of Resource Flexibility More Visible

Although responsive demand and improved market operations can reduce the cost of integrating low-carbon resources, resource flexibility is still expected to become more valuable as the share of RES increases in the transition to a low-carbon power system. In many current markets, however, the cost of inflexibility is obscured behind a range of market distortions, the very same distortions that obscure the value of investments in the underlying capacity. Making the cost of inflexibility transparent will be crucial in allocating capital efficiently to the right mix of investments.

1. Fully Price All Energy Market Balancing Decisions

It is a common misconception that energy markets are meant to set prices based on the short-run production cost of the marginal generating resource. In fact, they are meant to set prices based on the short-run value of whatever marginal action is required to balance supply with demand.

A major cause of "missing money" is that many of the actions taken to maintain the balance between demand and supply, in particular actions with the highest marginal value, take place outside of the energy market. These are actions taken by the system operator (or by market stakeholders at the request of the system operator) within the system balancing mechanism, the market operating regime that begins once control over market operations is turned over the system operator (the timing of which varies from market to market, typically from fifteen minutes to one hour before real time). System operators typically deploy these resources at a long-term contract cost that has no relationship to the instantaneous value of the resource when it's needed, or they require system resources to provide them at no charge, as a supplement to the resources deployed in the energy market. In this way the true short-term value of providing critical balancing services such as reserves at times when resources are constrained are often obscured. This in turn obscures the opportunity cost of supplying energy rather than providing reserves. Energy markets would be far more efficient at signalling the need for investment, and in particular the need for investment in greater resource flexibility, if the true cost of balancing the system at times of scarcity were reflected in clearing energy prices. One approach to doing so is by co-optimising energy and balancing markets, which forces the value of energy and the value of balancing services to converge more dynamically in real time.

Some markets have already begun to move in this direction. In North America, the PJM, New York Independent System Operator (NYISO), and the Electric Reliability Council of Texas (ERCOT) markets have all adopted measures that will allow energy market clearing prices to be set by an expanded set of balancing actions including deployment of demand-side resources. In Europe the United Kingdom's Ofgem has recently adopted similar measures as part of their Electricity Balancing Significant Code Review.

If a sufficiently large and diverse portfolio of end-use loads can be activated (including via demand aggregation) and factored into the setting of energy market clearing prices, the demand curve will become more sloped than vertical. In other words, consumers will begin to acquire the ability to express, directly or indirectly, the actual value of real-time access to electricity to power the various energy services they consume. This is an important step in making scarcity pricing more real, less volatile, less extreme, and more predictable, and therefore a more robust investment signal. It is also an important step in making consumers less vulnerable to the abuse of generator market power, in turn making it possible for governments and regulators to be more amenable to the relaxation of energy market price caps. The ability to relax price caps is a critical step in enhancing the efficiency with which energy markets support investment, lowering consumer energy prices overall. For the purposes of this paper, the important point is that it will make much more visible the value of investments in more flexible resources capable of efficiently complementing production from variable renewables.

2. Fully Price Scarcity in Balancing Services

Given historical limitations on consumers' ability to express more fully and accurately the true, more granular value of reliability, that value has traditionally been expressed by proxy via the security constraints adopted by system operators to keep the system in balance after gate-closure. Put simply, system operators determine for each scheduling interval how much and what type of fast-responding resources must be kept in reserve in order to reduce the likelihood of failure to a level that reflects the accepted standard for service reliability.

As described earlier, in most cases the market value of these reserves is set by a long-term contract price that bears no relationship to their value at the times they are deployed, or they are available to the system operator at no charge and their value is set accordingly. In reality, at times when demand is approaching the limits of supply and more and more resources are scheduled to produce energy, the supply of these balancing services available to the system operator, although still sufficient to meet demand, can fall below the level required to meet security constraints. In those moments the value of additional supply can rise dramatically. If this value is not made visible by the balancing mechanism it deprives the market of one of its most important and reliable means of expressing scarcity value. In addition to distorting balancing service price signals directly, the failure to price the demand for balancing services correctly deprives the energy market of the information it needs to gauge the opportunity cost of selling energy rather than selling reserve services, thereby distorting price signals in the energy market.

This situation can be remedied and several markets have moved or are moving to do so. Markets including PJM, ISO New England, and the GB market in the United Kingdom have adopted different versions of an administrative remedy sometimes referred to as an Operating Reserve Demand Curve. The ERCOT market in Texas is in the process of adopting such a mechanism. A detailed explanation is beyond the scope of this paper, but in short this mechanism establishes a value for balancing reserves based on the same "value of lost load" methodology that lies at the heart of the resource adequacy process. The value runs along on a continuum that rises from a variable cost-based value when reserves are plentiful, up to the maximum value of energy in the market once all reserves are exhausted.

As with other measures described here, this mechanism is intended not just to restore scarcity pricing to the energy market, but also to do so in a way that reveals scarcity value in a continuous manner whenever scarcity begins to emerge. In so doing these measures reflect more realistically how resource scarcity manifests and reveal incremental market signals for responses to scarcity long before demand reaches the limit of available resources.

3. Open Balancing Markets to Non-Traditional Service Providers

Improving the real-time price signals for energy and balancing services will be of only limited value, indeed it may simply increase costs, if the market for those services is effectively closed to the most flexible resource options available. Many balancing services mechanisms lock in the resources they expect to need well in advance of real time under procurement processes that, not surprisingly, reflect the characteristics of the conventional supply-side resources from which they've traditionally been obtained. In so doing they exclude resources that don't fit that particular template. An example is that many system operators conduct solicitations for primary, secondary, or tertiary reserves for terms of months or even years at a time. Many resources, such as various demand response opportunities, capable of providing balancing reserves as well as or better than traditional generators and at a lower price, are excluded because they are seasonal rather than annual resources. There are numerous similar examples of what is often unintentional discrimination. Aggregators of such non-traditional sources are excluded from markets, in some cases explicitly so. In other cases procurement processes effectively exclude variable renewable resources from providing the service by imposing unnecessary requirements on suppliers.

Many markets, for example some of the large organized markets in North America, have successfully adopted procurement for certain reserves categories that replicate the rhythm of the daily energy markets, obtaining the reserves they need from the most effective and economic resources available at the time. As the share of variable resources grows on the system, regulators and market operators can avoid unnecessarily high integration costs by combining proper price signals for the value of balancing services with open access for the most cost-effective balancing service providers available.

4. Locational Pricing

As noted earlier, the diversity of loads and supplies that comes with actually or virtually increasing the size of a balancing area will, all else being equal, reduce the quantity of resources required to meet a given resource adequacy standard and also reduce the amount of flexibility needed to integrate a specified amount of variable resources. This does not mean that the energy clearing price across the enlarged market is, or even should be, uniform. Limitations in the ability of the system to move lowest marginal-cost power throughout the system at any and all times — that is, “congestion” — will necessarily result in differences in the costs to serve load at different times and places. Those cost differences (the “costs of congestion”), if visible, can have very real and important effects on decisions to invest — on both the types of resources to be deployed and where to site them. As production from variable renewables increases, these differentiated local impacts will increase as well. As renewables become more exposed to market conditions, the benefits of addressing locational issues head-on will outweigh any short-term disadvantages that might arise from doing so. Locational price differences in and of themselves are of limited value in driving investment owing to their transitory nature, but making them visible makes more tangible and urgent the case for proceeding with the investments needed to resolve them.

Generally speaking, *locational pricing* is the cost of most efficiently supplying an increment of load at a particular place, while satisfying all operational constraints.[27] Put another way, it is the means by which least-cost system operation (i.e., merit order dispatch) is achieved when the bulk power grid is congested. As competitive wholesale electricity markets evolved over the past several decades, two general approaches to locational pricing emerged. The first, nodal pricing (also referred to as locational marginal pricing, or LMP), calls for the calculation of prices at every “node” on the transmission grid. A node denotes a place where supply (generation or an import) is injected onto the grid or where demand is withdrawn from it. Depending on the size of the grid, there can be many hundreds or even thousands of nodes. Nodal pricing is in effect in Argentina, Chile, New Zealand, Russia, Singapore, and several regions of the United States (New England, New York, PJM, and Texas).[28]

The second is zonal pricing or market splitting, which reduces the number of locational prices to be determined by aggregating nodes into larger areas (zones) of uniform pricing. Ideally, the zones are configured so as to minimize intra-zone congestion and thereby minimize the congestion costs that are hidden in the zonal price. In much of Europe today, each country is effectively its own zone. Countries that have several zones include Australia (each state is a zone), Denmark, Norway, and Sweden.[29]

In principle, the concept of locational prices is straightforward. In practice, their calculation is complicated. They consist of at least two elements, the cost of energy and the cost of congestion; the eastern US markets recognize a third (and very real) cost, that of line losses. The system operator calculates prices first for the day-ahead market. They derive from the clearing of supply and demand at every node or zone, subject to meeting reliability criteria. In a system without constraints (and ignoring line losses), the prices will equalize across all nodes or zones. The market clearing price will equal the bid price of the marginal unit (the last unit to clear in the market), because its price represents the marginal cost to serve the next increment of demand.[30] Where there are constraints, however, the locational prices will diverge. In the constrained areas, the prices will rise because higher cost resources will be called on to serve the local load. That increase in price is the “congestion component” of the locational price. The price in the unconstrained areas will again equalize, but they may in fact be lower than they would have otherwise been, if there is now an excess of generation caused by the inability to serve across the constraints (i.e., the congestion component is negative).[31]

It is on the basis of these day-ahead prices that the financial obligations of sellers and buyers are set and the next day’s dispatch determined. Buyers pay the locational prices at their nodes or zones and sellers are paid the prices at theirs. If there is no congestion, the total payments of the buyers will equal the total receipts of the sellers (including the costs of line losses). If there is congestion, the buyers’ payments will exceed the amounts paid to sellers; the difference is the cost of congestion and is typically used to compensate those

27 Litvinov, E. (2011). *Locational Marginal Pricing.* ISO-NE, WEM-301, pp 70-71.

28 Holmberg, P., & Lazarczyk, E. (2012, April). *Congestion Management in Electricity Networks: Nodal, Zonal and Discriminatory Pricing.* University of Cambridge, Electric Policy Research Group, CWPE 1219 & EPRG 1209, p. 4.

29 Ibid.

30 Strictly speaking, the LMP at a location is defined as a change in the total cost of production associated with meeting an increment of load at that location.

31 Note that this description has ignored line losses. In fact, in unconstrained systems, LMPs will vary from node to node, to the extent that line losses vary on different parts of the transmission system. New England, New York, and the PJM systems adjust their LMPs to reflect line losses. Texas does not (which means that the costs of line losses are shared equally among all nodes).

who purchased hedging instruments (sometimes called "financial transmission rights") to manage the risk of exposure to these costs.

In the United States, locational prices are also calculated in real time, typically at five-minute intervals, to re-optimize dispatch and to determine the value of incremental supply for balancing and other services (including flexibility). These real-time prices are also used to determine, for the purpose of financial settlements, whether committed supply is performing as required. This has to do with protecting against the exercise of market power (which resources in constrained areas often possess) and making sure that prices reflect the true costs of congestion.

Locational pricing is seen to have two sets of benefits. In the short term, it improves economic efficiency by revealing the cost of congestion and thus ensuring that, given the physical limitations of the network, demand is met at the lowest total operating cost. In the longer term, it reveals the value of solving congestion problems and thereby lowers obstacles to efficient new investment — generation, transmission, demand response, and energy efficiency — in the constrained areas.

In this way, overall economic efficiency is enhanced. And, insofar as the locational prices reflect the value of lost load, investment in flexibility will also be encouraged. This in turn transforms the debate over grid expansion, and the alternatives to it, from one about cause to one about effect, because it exposes where congestion can be found and how much one is willing to pay to get rid of it. There might still be arguments about the best ways to solve the problems, but there will be little question about their existence and cost.

6. Conclusion

We set out to describe how wholesale electricity markets can be adapted to match the needs of a decarbonising power system to the prevailing expectations for power system reliability. Our argument is that a secure, reliable transition to a low-carbon power supply can be accomplished at a reasonable cost. Indeed, preserving reliability at a reasonable cost through the transition is essential to sustaining strong political support for the project.

Market rules, market design, and market operations are at the centre of this process. We looked at some of the impacts of high shares of variable production on the nature of the resource adequacy challenge and on how markets determine the quantity and quality of investment that will be required to meet it. We addressed in particular the need to expand the investment problem set to include not just the quantity of capacity but also the operational capabilities needed to deliver a least-cost reliability solution during the transition to a low-carbon power system. We looked at the rationale for adopting capacity markets and discussed why they will not deliver resource adequacy at least cost if capacity is valued as an undifferentiated commodity. We showed that even “smart” capacity mechanisms would need to rely on the improved operation of energy markets. As such, improving energy markets constitutes a “no regrets” measure.

We considered the special case of continued investment in renewable capacity. Given the commitment to a low-carbon power sector, there are sound security of supply and economic reasons why investment in the deployment of key renewable technologies must continue, but the nature of support for that investment, as with investment in conventional resources, is currently the subject of considerable discussion. We looked at the evolving nature of the challenges facing investment in renewables, particularly variable renewables, and offered some suggestions for what should drive renewable support policy going forward, including the need to be more deliberate and selective in dealing with oversupply of baseload generation as the share of variable RES grows.

Looking beyond the various debates about energy market interventions, we returned to the underlying energy market itself. We reviewed a range of practical options available to governments, regulators, and system operators to better match the structure of the market to the needs of a low-carbon power sector, and to restore the functioning of energy and balancing services markets as close as possible to their full potential.

A decarbonized power sector is at the heart of delivering on the climate policy goals being adopted by a steadily expanding set of jurisdictions. The challenges in delivering on this objective in a secure and affordable manner are real. So are the many options available to overcome those challenges. Making good choices begins with sound fundamentals and a holistic approach. We have attempted to provide some of that foundation here as a context within which to evaluate a number of key choices facing policymakers and regulators.

Related RAP Publications

Beyond Capacity Markets
http://www.raponline.org/document/download/id/4854

Delivering a reliable power supply to consumers has always been a central objective of market design and various solutions to this challenge have been adopted in Europe and elsewhere. In the context of wholesale power markets, these solutions have generally taken the form of creating market rules to pay for firm capacity alongside energy-only prices. A closer look at the new reliability challenges associated with meeting Europe's decarbonization targets suggests that our collective thinking will need to evolve "beyond capacity markets" in order to address them. In particular, the power system will need resources capable of rapidly changing output or flexing demand frequently and continuously throughout the year around the energy availability from intermittent renewables. Based on these requirements and related considerations, RAP sets out in this paper a set of market design principles that can be used to assess the suitability of proposals intended to deliver system reliability. The proposal to introduce a capacity payment mechanism being considered for Great Britain is reviewed against these criteria.

What Lies "Beyond Capacity Markets"?
http://www.raponline.org/document/download/id/6041

"Beyond Capacity Markets" discussed reasons why, in power systems with large shares of variable renewable production, existing capacity market models may no longer be up to their intended task of driving the investments required to deliver least-cost reliability. This paper follows up on that analysis by addressing the question of what regulators and policymakers should do in response to these concerns.

Around the world, the ultimate aim of those involved in regulating a monopoly power sector or designing and overseeing competitive electricity markets is to find the set of rules and practices that efficiently and reliably delivers the right amount and the right mix of resources. Many different approaches have been taken and all have been subject to multiple revisions. The next challenge is to understand and address how the growing share of variable renewable production will require us to rethink our current practices. While many of the discussion points of this paper apply equally to all industry structures, our primary focus is on adapting competitive wholesale power markets to deliver their intended economic efficiency and reliability outcomes under this new resource paradigm.

Reliability has always involved two dimensions, but they have traditionally functioned in different timescales. Resource adequacy – access to enough firm resources to be able to meet the highest expected level of demand – has dominated planning at investment timescales. In contrast, system quality – the right mix of resource capabilities deployed to ensure that in every moment supply can be balanced with demand – has been the focus of services markets that have functioned primarily at operational timescales. While existing capacity market models are concerned only with resource adequacy, it is the system quality dimension that is fundamentally transformed by rising shares of variable renewables, making resource flexibility increasingly an investment consideration as well as an operational one.

This paper offers two market design models to deliver least-cost reliability in power systems with an increasing share of variable renewables. The paper also lays out the critical implementation steps as well as a decision framework for regulators and policy makers to use in thinking through the design and implementation of appropriate market mechanisms.

Energy Efficiency Participation in Electricity Capacity Markets: The US Experience

http://www.raponline.org/document/download/id/7303

Motivating power producers to plan for and develop enough capacity to meet the needs of customers during times of peak demand has been a challenge for power planners. A substantial portion of system capacity is needed for a relatively small number of hours each year. One example from New England showed that the 50 hours of the year with the greatest demand (i.e., the top 0.6 percent of the hours of the year) have been responsible for seven to ten percent of system capacity needs. Some have expressed concern that high wholesale electricity prices in those peak hours may not be enough to ensure that sufficient capacity will be made available to meet reliability requirements. To address that concern, several U.S. Regional Transmission Organizations—sometimes called Independent System Operators (ISOs)—have created what are commonly called "capacity markets," in which payments are made for commitments to provide electric capacity during the time of future system peaks and capacity shortage situations. This paper summarizes the rules governing how efficiency resources participate in the ISO New England and PJM capacity markets, the results of that participation, and lessons learned to date.

Aligning Power Markets to Deliver Value

http://www.raponline.org/document/download/id/6932

Wholesale markets will play a key role in driving investment in the flexible resources needed to ensure reliability as the share of intermittent renewable resources grows. The author identifies three areas where power markets can adapt to enable an affordable, reliable transition to a power system with a large share of renewable energy. These are a) recognize the value of energy efficiency, b) upgrade grid operations to unlock flexibility in the short-term, and c) upgrade investment incentives to unlock flexibility in the long term.

RAP®

Rue de la Science 23
B – 1040 Brussels
Belgium
Tel: +32 2 894 9300
www.raponline.org