Electricity Wholesale Market Design in a Low Carbon Future

Draft 01/23/10

William W. Hogan, Harvard University[i]

For volume on

Harnessing Renewable Energy

Jorge Padilla (LECG, CEMFI, and CEPR) and Richard Schmalensee (MIT), Editors

This chapter discusses the difficulties associated with deploying renewable electricity over large scale electricity markets. It considers such issues as electricity system design, transmission networks, and the development of smart grids. There is great uncertainty about the appropriate investments and a consensus that achieving a low carbon future will require innovation and invention. Success is more likely if appropriate price signals and incentives appear to drive decentralized decisions of the many participants in electricity markets. This requires special emphasis on the interaction between wholesale market design and innovation with renewable energy and other low carbon technologies.

Introduction

Policy to convert the energy system to a low carbon configuration involves a central focus on the electricity sector. Electricity generation is a major use of fossil fuels and therefore accounts for a large fraction of the emissions of carbon dioxide; reducing carbon emissions in other sectors such as transportation may involve increased electrification; and electricity can and has been generated without carbon emissions. In addition, electricity generating facilities burning fossil fuels are large and easy to identify and thus a natural focus for regulation and policy to reduce carbon emissions.

Designing policy for a low carbon electricity sector depends on the nature of the technology and institutions in that sector. In many countries, the electric power system is already involved in a major restructuring process. Critical technological and institutional features of that reform process center on wholesale market design elements that have important implications for expansion of reliance on low carbon energy sources. The wholesale market covers bulk generation, dispatch and transport of electric power for ultimate delivery to retail customers. Although the details and jurisdictional rules differ in the US and the European Union, a common distinctive feature of the wholesale market is the interaction through the interconnected transmission grids that transcend states and countries. The complementary rules for retail supply or distribution of electricity are important, but the emphasis here is on electricity wholesale market design in a low carbon future. The main theme emphasizes the strong interaction between market design and policies for achieving a low carbon future.

Electricity System Fundamentals

All electricity systems share certain fundamental characteristics. Efficient power generation benefits from substantial economies of scale. The result of more than a century of evolution of sophisticated power systems is a fleet of relatively large generating plants.

To enjoy the benefits of these economies of scale, traditional power plants have not been located near the ultimate load. Moving power from the generating plants to the ultimate consumer requires high voltage transmission lines. Because of the nature of electricity transmission, the higher the voltage the lower the required current and the lower the resulting transmission losses. Transformers convert the relatively low voltage at the point of generation to allow for high voltage transmission to another set of transformers that lower the voltage for retail distribution to factories, businesses and homes.

Typically the local retail distribution system is simpler in its design than the transmission system. In particular, the distribution network is more decentralized or “radial” in the sense that electrical power problems in one distribution network, the principal cause of power outages, can be isolated from the rest of the system.

The interconnected transmission network is another story. Initially, the transmission network developed organically as individual companies sought to connect their generators to their loads. One of the features of the electric power system is the speed of adjustment when there is a change in load, generation or transmission connections. With present technology and little storage, to a first approximation electricity is the limiting case of same time production. The power produced and consumed must maintain instantaneous balance throughout the system, and any disruption of the transmission grid results in instantaneous redistribution of the flow of power in the system. The system is not controllable in the same ways as we imagine for a network of pipelines with valves and relatively slow moving fluids. Power flows at the speed of light.

Companies building bulk transmission facilities consequently relied on redundant pathways in the transmission grid to ensure reliability. In addition, the power system required standby generating capacity that could be called upon immediately in order to replace a lost generating unit. Since neighboring electricity companies would be unlikely to face disruptions at the same instant, redundant transmission and generation in one system could help other systems and lower the total cost of maintaining reliability. This confluence of technology and economics led to the gradual development of interconnected alternating current (AC) transmission grids linking loads to thousands of generators through tens of thousands of individual transmission links.

Throughout such an interconnected grid, all of the generators and load must be synchronized. The resulting ensemble has been called the largest machine in the world (Amin 2000, 264). For example, the North American continent from Mexico to Canada is dominated by only three synchronous grids in the Eastern Interconnection, the Western Interconnection, and the Texas Interconnection operating with limited exchange through direct current (DC) interconnections between them.[1] Similarly, a large European synchronized grid covers 24 European countries (ENTOS-E 2008, G2).

The complexity of a large synchronized grid is well known to electricity system operators but not by most customers. Traditionally, the electricity company was vertically integrated from generation through transmission to distribution. Customers faced a monopoly that provided the power and charged a price only loosely coupled to actual contemporaneous costs. The company would make its own investment plans for generation and wires, often through joint investments with other vertically integrated monopolies. Control of the interconnected transmission grid was handled through an array of bilateral and multilateral agreements among the members of the club, but with strict exclusion of potential new competitors. These club agreements worked reasonably well, although below the surface there were often difficult problems at the seams between regions or when one member of the group was "leaning" too much on the system.

The club agreements involved coordinated planning and expansion of the transmission grid, always an opaque and arcane process. It was difficult to define the resulting rights to transmit power and to determine how these rights would be allocated in the interconnected grid, and real-time dispatch of power plants had to be coordinated in order to maintain the reliability of the grid.

The rules and incentives for operation of this system were complicated by the nature of electricity regulation and its focus on cost recovery. The emphasis of regulation or government ownership was on control of the average costs, with relatively little attention paid to the role of marginal costs and the connection to incentives that could or should govern operations and efficient investment. This history would be a major hurdle to overcome in the development of electricity market reforms.

Economic efficiency always played a secondary role to maintaining system reliability, keeping the lights on. Gold-plating the system a little might produce some inconvenient questions on occasion, but major blackouts or even relatively minor but frequent local supply interruptions could limit a career and were deemed unacceptable. The common approach for the wholesale power system was and continues to be treating reliability as a constraint and economic efficiency as a goal to be sought subject to that constraint.[2]

In real-time, the variable cost of generating power ranges from virtually zero for run-of–the-river hydro power, to moderate for nuclear and coal plants, to expensive for natural gas and oil plants. With load changing rapidly over the day, and total generation at the trough during the night at as little as 30% of the peak during the day, it makes sense to choose the cheapest generators available to produce at any time (Monitoring Analytics 2009, 42).

The basic idea is to operate as closely as possible to the security constrained (respecting all the many reliability constraints on the grid) economic dispatch that provides the necessary ancillary services and minimizes the cost of meeting a given electric load at

[1] Alaska and Quebec have their own separated grids as well.

[2] Because of the lower availability of redundancy and the greater expense of reliability, a more nuanced tradeoff is seen at the retail distribution level.

any moment and over the course of the day. Under the traditional electricity model, the aggregate cost of this economic dispatch would be recovered through preset customer charges that muted or completely ignored the dramatic changes in opportunity costs over the day and the season. As a result, the traditional system often provided adequate cost recovery but did little to reveal or communicate incentives to guide either operations or investment. Rather, investment decisions relied upon and defaulted to the expert assessments of the established electricity companies and left few opportunities for innovation by new entrants.

Although the operating decisions were neither open nor transparent, a reasonable case can be made that traditional reliance on the framework of security constrained economic dispatch produced efficient operating decisions and certainly contributed to the reliability of the system. However, the signals sent for investment and innovation were an entirely different matter. The system depended on the planning decisions of the protected monopolies, and the lack of transparency coupled with major construction delays and cost increases precipitated deep concern with investment choices.

Electricity Market Design

As long as the natural benefits of economies of scale produced steady reductions in the real and nominal average price of electricity, as was true for decades, the vertically integrated model with franchise monopolies worked well. Rapid growth in electricity demand hid mistakes. Average costs were declining, and the big concern was to invest rapidly enough to stay ahead of growing demand. However, this system began to break down when the returns to increasing scale declined, electricity growth patterns changed, and mistakes were made that resulted in large cost overruns. The long running good news that new investment would lower average electricity costs was soon replaced with the bad news that new investment was more expensive and would raise average costs (Hogan 2009).

Electricity Market Reform

These changes accompanied a general interest in reducing reliance on regulation and turning more to market forces to guide investment and operations. For example, in the 1980s the United States (US) eliminated regulation of natural gas production, and there was a great boom of activity and improved economic efficiency. Policy-makers in the United Kingdom (UK) and Chile turned attention to similar reforms of electricity systems. In the US, successful experiments with limited introduction of generation investment by companies outside the monopoly club had demonstrated that new entrants could be accommodated without compromising efficiency or reliability. In 1992 the US adopted the Energy Policy Act (EPAct92) that included provisions for open access to the transmission grid.

Authorities in Australia, Argentina, most of Europe, New Zealand, the US, and elsewhere were experimenting with new structures for organizing the electricity system. Many recognized or assumed that economies of scale in generation had largely been exhausted and that new and more loosely regulated entrants were just as capable of building and operating successful power plants as had been the incumbent electric utilities. This precipitated the notion that vertical separation of generation, transmission, low voltage distribution and supply would be at least possible and perhaps desirable. Existing generators, perhaps divided and divested from existing electric companies, could compete against each other and against new entrants to construct power plants and sell electricity to others, including final customers. Local distribution facilities would continue to operate more or less under the old regulated monopoly model, albeit often as a wires-only business with electricity suppliers acquiring electricity from generators and selling on to final customers.

A key requirement for this vertical separation and competition in generation would be open access to the common integrated transmission grid and associated unbundling of critical services. The best form of organization for the common interconnected grid with the unbundling of services was not immediately apparent. In the US the emphasis was soon on transmission open access and non-discrimination in prices and services for all participants in the electricity market. As it turned out, these seemingly innocuous requirements would have profound implications for the nature of electricity market design.

The details of the story and the alternative market approaches are many, and there is an extensive literature on alternative market designs (Sioshansi 2008). The purpose here is not to recite these details but to emphasize certain features that are salient for the discussion of a low carbon electricity system that relies heavily on technologies such as solar, wind and demand-side management, that differ substantially from the legacy portfolio of fossil-fuel generation.

Alternative Market Designs

For this purpose, we can divide electricity system market designs into three categories. First, there are legacy systems that continue to be organized around the model of vertical integration and closed transmission access.[3] This would include much of the Southeast regulated utilities and Northwest municipal and public power systems in the US. Second, there are organized counter-trade markets with aggregate zonal transmission systems that support sometimes separate trading platforms and scheduling procedures that are only partially integrated with system operations. The counter-trade refers to the activities of the transmission system operator who must conduct offsetting transactions that undo the schedules of market participants to bring the net schedules in line with the capabilities of the transmission system. This describes many systems including most of the electricity markets in the EU. Third, there are the organized spot markets under independent system operators (ISOs) that coordinate dispatch but do not own the grid and integrate real-time trading and dispatch with system operations. This integration includes the full locational granularity of the actual grid for determining energy flows and locational prices. This would include the multi-state Regional Transmission Organization (RTOs) or equivalent single-state ISOs that together cover approximately two-thirds of the US load.

To a significant degree the vertically integrated systems are similar to those of the past and we can think about how regulated utilities would be likely to adapt to a low carbon

[3] In the US case, vertically integrated utilities are subject to certain open access provisions, but these open access rules lack the critical ingredient of non-discriminatory access to economic dispatch.

future. The counter-trade systems include many market innovations and allow a great deal of flexibility in trading arrangements, operations and investment incentives. However, at their core these markets are incompatible with the requirements of system operations. The aggregate trading arrangements ignore many of the relevant constraints on the grid. The associated nominal schedules are thus often not collectively feasible, and counter-trade is necessary to undo what the nominal market has done. Counter-trade, or some other system redispatch intervention like it, is necessary whenever the market schedules do not reflect system operation reality. Furthermore, workable counter-trade market designs and interventions are necessarily discriminatory in selectively choosing and paying for redispatch of some generators but not others similarly situated, as shown by experience in the US (Hogan 2002). This counter-trade design has implications for the treatment of renewables and other low carbon energy sources. In order to understand the implications, it is convenient first to sketch out the essential features of the RTOs and organized spot markets that integrate market design and system operations.

Integrated Locational Electricity Market Design

The US RTOs have regular market monitoring organizations that prepare extensive evaluations of the design and operation of their markets.[4] The critical design criterion is to structure the spot market products and services as much as possible to conform to actual system operations, and then to price those products and services to reflect the marginal costs that would drive market clearing competitive prices.

The initial focus is on the spot market. The forward markets play a critical role, but forward market participants will look ahead to the anticipated operation of spot markets. Thus efficient design of forward markets depends in significant measure on a good design for spot markets.

Further, a focus on competitive market design is not intended to overlook the possibility of the existence and exercise of market power. Rather, a good competitive spot market design greatly simplifies possibly necessary regulatory intervention to mitigate market power. For electricity systems, a necessary condition for good market design is good design of the spot market.

When closely tied to system operations, good spot market design can be integrated with bilateral or multilateral trading. The organized spot market does not preclude decentralized trading. Transactions through the spot market can be gross (meaning without separate bilateral transactions and schedules) or net (meaning to cover only imbalances relative to bilateral transactions and schedules). The differences are largely semantic, despite heated arguments in the past for the merits of gross or net market design.

The principal market design problem that arises in all electricity systems follows from the nature of the interconnected transmission grid. In addition to being large, interconnected, fast responding, and contingency constrained, electricity transmission grids have an unusual feature that follows from the need for fast response. The redundancies and contingency constraints arise because it is not possible to use valves to control the flow of

---

4 See RTO annual reports at (http://www.hks.harvard.edu/hepg/rlib_rp_RTO_ISO_reports.html).

power. To a good first approximation, once the pattern of load and generation is set, the flow of power through the grid is determined by physical laws that distribute the power over the system to (more or less) minimize losses. It turns out that power flows to varying degrees on every available path between generation and load. This means that control of power plant dispatch and control of transmission flows are two sides of the same coin. Changing the dispatch changes the power flows, and changing the power flows requires changing the dispatch.

This technological fact has profound implications. For example, it means that it is impossible to operate the system with only decentralized decisions about generation and load. There must be central coordination of everything and central control of enough of the dispatch to meet the requirements of system operations. The parallel flow of power means that any trading system must confront material externalities as the real flows for one transaction interfere with the real flows of other transactions. In the US the electric utilities struggled for decades to design a workable system of transmission flow accounting that would support physical transmissions rights, and never succeeded in working around the inconvenient fact that power flows everywhere (Lankford et al. 1996).

The design solution to this otherwise intractable problem is to recognize that the framework of security constrained economic dispatch, already familiar to system operators, provides the foundation for a coordinated spot market (Schweppe et al. 1988). This spot market inherently respects the requirements of system operations, and provides a convenient connection to electricity markets and forward trading.

Recognizing that there must be a system operator, this design asks for bids and offers for power at the physical locations, or schedules between locations, or schedules with bids and offers for deviations for the schedules. Thus the structure of the information provided to the system operator is the same as under a vertically integrated system, only now the engineering estimates of loads and costs are replaced by the bids and offers for power.

As before, the system operator selects the changing pattern of economic dispatch to meet load and provide ancillary services subject to the (many) security constraints and the complex flows of power on the transmission grid. In economic terms, the independent system operator internalizes the major transmission externalities so that transmission and transaction schedules can be point-to-point, and be silent on the path between the points.

The quantity result of the security constrained economic dispatch for any interval is a schedule of generation and load at each electrical location. Hand-in-hand with the quantity dispatch is a set of market-clearing prices for settlement purposes that capture the system marginal cost of meeting increased load or decreased generation at each location. The term of art is locational marginal pricing (LMP). These locational marginal prices provide an immediate definition of the appropriate spot-price of transmission between any two locations: a straightforward arbitrage argument dictates that the competitive spot-price of transmission between any two locations is the difference in the energy spot prices at the locations.

When there is no congestion in the system, the market clearing prices would be the same as the normal single market clearing price of simpler models, except for differences in marginal transmission losses. But when transmission constraints are binding, the locational prices differ according to the marginal impacts of activity at each location on system congestion. These congestion induced differences can be surprisingly large, and sometimes counterintuitive. Simultaneous market clearing prices can often be well above the marginal cost of the most expensive generation running at some locations, and well below marginal cost or even negative elsewhere (Hogan 1999, 4). This is a material effect of the complicated electrical interactions made visible in the spot market but previously hidden in system operations.

Once this security constrained economic dispatch framework is adopted, with the accompanying locational prices applied in market settlements, it is a simple matter for market participants to schedule or otherwise arrange transactions. For example, a bilateral transaction can be scheduled between any two locations, for a charge at the difference in locational prices. Any imbalances in actual delivery would be paid for at the relevant locational prices. With security constrained economic dispatch and locational prices, the imbalance market merges with the spot market. All parties can participate, so there is open access. And all parties face the same scheduling rules and resulting prices, so there is non-discrimination. There is no provision for counter trading through market positions taken by the system operator, because there is no need for counter trades to bring the spot market into conformance with system operations. Hence, there is less concern with conflicts of interest for the system operator. By construction, security constrained economic dispatch is consistent with system operations and the spot market.

In addition, the existence of the spot market and locational prices provides the foundation for the financial transmission right (FTR) (Hogan 1992). In the simplest form, an FTR calls for paying to the right holder the (congestion) difference between the corresponding locational prices. Absent grid outages, the FTR is a long-term instrument that provides a perfect hedge for the short-term congestion fluctuations in transmission spot prices. With a matching FTR between generation and load, a supplier can sign a fixed price contract for delivered power and be sure that the system would always allow fulfillment of the contract through either physical delivery or offsetting purchases and sales with no incremental charges for volatile transmission congestion. This workable financial transmission right provides the functionality sought without success in unworkable systems of physical transmission rights.

There were early concerns that this integrated locational model could not work very well with very many locations. The US went through the agony of forcing aggregations across zones to reduce the granularity of the market representation. Zonal aggregation subject to rules of open access and non-discrimination was tried and abandoned in PJM, New England, California and Texas (Hogan 2002, 12-13; Potomac Economics 2009, viii). This zonal aggregation inherently suppressed the detail of real system operations, and thereby lost the efficient pricing signals, which in turn required complex rules to bridge the constructed gap between the spot market and system operations. Painful experience made it clear that the simplest system is to go as far as possible to match temporal and locational granularity with real operations and the associated system opportunity costs. The integrated locational design with very many locations works. For example, in the Mid-Atlantic states, PJM updates the coordinated dispatch and prices every five minutes for approximately 8,000 locations. This granularity means that geographically proximate locations that are electrically distant can have very different prices, and the different prices reinforce rather than contravene the imperatives of efficient and reliable system operations. One of the regular reports from system operators newly adopting this model is that it greatly simplifies and improves dispatch operations because market participants have strong economic incentives to follow dispatch instructions.

Some contend that a more geographically aggregated system would be easier and cheaper to implement. This might be correct if there were an available alternative that could avoid the need to attend to all the detail in real operations.[5] However, there is no such alternative available. If the system operator is applying the basics of security constrained economic dispatch, then all the essential information on transmission lines and multiple locations must already be available. The distinctive feature of the integrated locational market design is to make the locational marginal opportunity costs visible and to apply these as the prices used in the settlements system. The incremental costs of using the full granularity are modest, or even negative when considering the costs of imposing rules and procedures to overcome the effects of artificial aggregation.

Aggregate Regional Counter Trade Market Design

The counter-trade markets differ from the integrated locational markets described above in a variety of ways (Boucher and Smeers, 2002). A typical feature is an attempt to separate spot market trading from the actual details of system operations. The energy product is defined according to inputs and outputs at some level of regional aggregation. Trading reaches an equilibrium in these aggregate products, which is turned into a set of schedules for implementation by the transmission system operator. The nominal equilibrium schedules determined in the market are seldom if ever strictly feasible for implementation on the actual grid, since they ignore many or all of the constraints on grid operations. The transmission system operator has available essentially two broad strategies for bridging the gap between the nominal market equilibrium and physical system through applying either ex ante limits or ex post counter trading.

An ex ante limit would be to define different regional aggregations for the energy products and pretend there is a simple pipeline between the two regions. Transfers through the pipeline would be limited to a conservative estimate of the real capacity, and transfers from anywhere in the first zone to anywhere in the second zone would be (artificially) treated as having the same effect. The conservative capacity cushion would help limit problems caused by trades that would appear identical in the pipeline model but would have different impacts on capacity in the real system. Of course, this approach ensures underutilization of the grid by all parties that must use this mechanism for trading. The same conservative assumptions create opportunities for transmission owners to exploit transmission capacity that is in effect withheld from the rest of the market.

Other forms of ex ante rules required by the disconnect between market design and system operation may include restrictions on the lead time for changing energy

[5] Likewise, administrative aggregation without changing the underlying grid does not increase competition or reduce market power.

schedules. Physical operation of plants and control mechanism often allow for changes up to minutes before real-time, but formal market schedules may foreclose changes an hour or more ahead of operations. This scheduling lag could be particularly important for highly variable energy resources.

Despite the conservative capacity restrictions and other ex ante rules, market schedules may still violate the physical limits of the system. Even though generation and load may balance within the boundaries of an aggregate zone used to define a common energy product , for instance, intra-zonal power flows may violate transmission constraints. There is no way to set an ex ante rule that maintains the aggregation of the zone and reflects the impacts of the intra-zonal impacts on the transmission system. The trade-determined schedules cannot be honored without some redispatch of the system, and this provides a need for counter trade. Typically the initial schedules are honored, but the system operator arranges a generation redispatch that involves the equivalent of purchases and sales of energy at different locations in the system and that creates counter flows that relieve the transmission constraints. In principle and in practice, this counter trade redispatch may involve the very trades and generators that created the potential overload. In effect, a generator may be paid in the first instance to provide energy as part of the nominal market equilibrium schedule and then be paid not to generate the same energy in order to accommodate the counter flow and relieve the transmission constraints.

As might be expected, the incentive effects of these aggregation and counter-trade arrangements can be perverse in the extreme. The zonal aggregation models in PJM, New England, California and Texas all faced these perverse incentives, which precipitated abandonment of the zonal aggregation and counter-trading market design.

Other markets have implemented counter-trade systems that work much better. For instance, the EU universally employs aggregate regions for defining energy products. The EU uses pipeline models with conservative transmission capacity assumptions. All these workable systems find it necessary to dispense with principles of non-discrimination. The transmission system operator is able to make one-off arrangements for counter trades that relieve transmission constraints and balance the system without having to offer the same deal to every other participant in the market. This reduces the ability of generators to extract payments for not running. While necessary, this selective participation in the market by the transmission operator creates its own set of incentive problems.

The cost allocation mechanisms for these counter-trade systems are varied. In some cases, such as in the UK, the transmission system operator absorbs the costs under a price cap regime and, therefore, internalizes incentives to minimize the counter-trade costs. However, this same system requires the transmission system operator to take on the sole responsibility for transmission expansion, as the aggregate price signals in the market do not provide sufficient incentive or information to guide detailed transmission investments.

Market Design for a Low Carbon Future

In the US installations of wind energy are disproportionately found in the RTO markets because of the greater ease of integration (EPSA 2008). As would be apparent from the summary of key features presented here, there are net benefits to moving away from either vertical integration or regional aggregation with counter trade and towards the integrated locational market design. Similar conclusions can be found in other important reviews looking towards a low carbon future in the US (Joskow 2008; Helman et al. 2010) and the UK (Grubb et al. 2008). As summarized by the International Energy Agency in its review of market experience across its member countries, "[L]ocational marginal pricing (LMP) is the electricity spot pricing model that serves as the benchmark for market design – the textbook ideal that should be the target for policy makers. A trading arrangement based on LMP takes all relevant generation and transmission costs appropriately into account and hence supports optimal investments." (IEA 2007, 16)

Connection Standards

The rules for generator grid interconnection play a major role in affecting the integration of new capacity. In the case of the integrated locational market design there is fundamental simplicity in theory that can be reasonably approximated in reality. In the other market designs, system connection rules confront the challenge of designing protocols to overcome the perverse incentives created in the market.

The simplicity of the integrated locational market design is illustrated in the rules for "energy resource" interconnection in the PJM system (PJM 2009). An investor that has a site with the required permits and wants to build a new facility to sell energy into the PJM market has to meet only a minimal set of requirements. Primarily these are technical constraints to ensure that the connection to the grid will satisfy electrical characteristics to ensure that connection and charging of the line, with no material power generation, do not compromise the rest of the system. In effect, the generator interface is an extension cord to the rest of the grid that has to meet certain standards, and the generator has to pay for the equipment and extension line.

This form of interconnection allows the generator to produce, buy, and sell energy in the spot market at the locational marginal price at its point of interconnection. There is no guarantee about the price or the profitability of the energy produced, of course, nor any guarantee that access to the grid at the point of connection provides transmission rights to hedge the cost of delivery to other locations in the grid. When the transmission system is constrained, the locational marginal price at the point of connection might be very low, the generator may choose not to run. The generator can have the system operator make these decisions in real-time by simply making an offer to produce with a minimum acceptable price. There would be no need to arrange advance schedules or bilateral transactions.

Of course, the generator could participate in forward markets, arrange long-term contracts, and purchase FTRs to hedge transmission costs, and so on. But these are not necessary conditions of the connection rule, and immersion in trading would not change the incentives for the price-taking generator from making simple offers into the real-time market to allow the system operator to take the energy when available and when the locational price is at least equal to the minimum bid.

As discussed below, in PJM most generators choose a more complicated form of interconnection in order to participate in the capacity market that operates on top of the energy market. However, the simple ideal of the energy-only interconnection is available and is used by some generators.

The counter-trade markets face a quite different situation when considering the rules for connecting new generators. The generators do not see the costs of transmission constraints obscured by zonal aggregation of energy product definitions. Hence, they may choose to locate where costs are really higher but are imposed on someone else. Furthermore, the typical counter-trade market design carries with it an explicit or implied obligation to accept the generators' energy output and deliver it to some loads. In effect, in the terms of the integrated locational market design, new generators acquire a bundle of transmission rights by virtue of their interconnection. The connection of new generators may imply that the transmission owner must make sometimes substantial transmission investment to support these new rights. Similarly, for all generators, but particularly for highly variable wind and solar plants, there must be increased operating reserves or other controls needed to support energy delivery. The transmission system operator will have to make these investments or impose constraints on the interconnection.

This gap between the incentives for the new entrant and the real costs of interconnection creates a need for rules and regulations to overcome the incentive effects. For example, new generators may be required to make so-called deep transmission investments elsewhere in the network in order to accommodate their new production without undoing the implicit rights of other generators. Alternatively the costs of reliability and transmission upgrades to support the expanded set of transmission rights are imposed on the transmission system operator or somehow collected in socialized charges applied to all customers.

Smart Grids and Smart Pricing

The notion of the smart grid typically implies a mix of control technologies, information systems, and information distribution that allow for much finer control of all the devices connected in the largest machine in the world. This would include more flexible controls in the transmission grid, better monitoring of grid and distribution flows, real-time status and fault monitoring, automatic meter reading and, importantly, real-time automation of end-use devices. Signals from the central system operator would communicate through the network to change the settings on transformers deeply embedded in the grid or trigger your settings to turn down the air conditioner in your home.

The evolution of the smart grid will be particularly relevant in the development of dispersed generation, end-use efficiency and load management. An important question is what information is provided and who decides how to use it. Consumers are unlikely to surrender their ability to control the temperature in the house or to decide whether to run a factory. However, it is easy to imagine widespread adoption of smart devices that implement instructions as to when to curtail usage and when to run as needed.

For these smart systems to be successful, consumers must receive a signal that aligns their incentives to behave in a way that reflects the opportunity costs in the system. This is a challenge in the vertically integrated and aggregated counter-trade systems, since under these market designs prices facing the consumer do not reflect the granular detail of system operations. Command and control under the vertically integrated model and counter trading under the aggregated market models can in principle provide an efficient dispatch for generators and respect the transmission constraints, but they do not by themselves align the incentives of customers.

By contrast, the integrated locational market design provides a straightforward alignment of incentives. With a retail rate design that passes through the locational price of energy, the smart device would have smart prices that aligned with the incentives of the consumer and the system's costs. Some distortions would remain to the extent that distribution or other customer service fixed costs are collected in variable rates, but the distortions would be small given the large changes in energy prices that would or should be the norm during peak periods.

The smart grid needs smart pricing. Market models that suppress price information, especially in prices of final delivery to the consumer, would require new and sometimes complex rules to undo the consequent perverse incentives. By contrast, the integrated locational market design intentionally links the market and operations in order to provide the right price signals.

Resource Adequacy

A central function for vertically integrated electric utilities has always been to invest in anticipation of electricity demand. The club members of vertically integrated companies developed rules for investing while sharing the responsibility to maintain reliability. In practice, this amounts to a responsibility to build or contract for energy generation capacity adequate to provide a specified planning reserve margin (e.g., 15% of estimated peak load) that would ensure adequate capacity in the event of normal outages for maintenance and some temporary equipment failures.

This obligation continues for the vertically integrated utilities. However, the situation has become more complicated in the market models where generation has been separated from transmission, distribution and final supply to the customers. The success of these models in achieving resource adequacy is neither certain nor resolved. In the early days of restructuring, at least in the US, there was a surge of new construction of more efficient generating plant. This created excess capacity. When coupled with bad market design and rising oil and natural gas prices, this capacity was not economic, and the result was a major contraction in investment and significant financial losses for some generators. In one sense, this was the intended consequence of electricity restructuring in shifting the burden of risks to voluntary investors rather imposing them on customers. Of course, the excess investment was not a good thing even if a good deal of the problem was caused by the incentives of bad market design rather than the reliance on markets per se.

The result has been a concern that in the future there would not be adequate investment in generating capacity in electricity markets. This concern has been strongly reinforced by the regular demonstration that the returns to investors in the energy and ancillary services prices have not been adequate to support new investment.[6] Although it is the anticipated returns that should govern, not the realized returns in a market with excess capacity, the low realized returns are sobering. The consensus analysis is that energy prices in many electricity markets have been too low, and are especially low during periods of capacity scarcity such as during the peak hours of the day.

The response to this concern has been varied. In some markets, notably New York, PJM and ISONE, a "capacity" product and an associated capacity market have been created. The essence of the capacity market has been to construct a forecast of future loads and investments and then conduct an auction for existing and new generators, or demand response, to offer "capacity" to meet those requirements. The full description of how to define capacity and how to ensure its future deliverability on the grid is another task. However, it is clear that this capacity construct is an attempt to recreate something like the responsibility of the vertically integrated utility, but now under the management of the independent system operator.

The difficulties of defining and measuring capacity for new low carbon technologies like wind, solar, and load management are particularly relevant. As discussed below, the variable availability of renewable resources and the problem of defining baseline load make it hard to to map them into the equivalent savings in conventional fossil fuel generation capacity. In part the problems are inherent in capacity markets where the product is not easy to define or measure, but the difficulties are exacerbated by the substantial differences in the technologies.

At the other end of the spectrum has been the recognition that inadequate scarcity pricing is the root cause of the deficient incentive for investment.[7] In markets such as Australia and Texas there are mechanisms to allow prices to rise during times of scarcity on the assumption that this will be sufficient to meet the needs of resource adequacy.

There is nothing fundamental that requires a choice between improved scarcity pricing and adopting well designed capacity markets. Better scarcity pricing would provide benefits in improved operations as well as market incentives for investment. Furthermore, better scarcity pricing in markets would complement the designs of capacity markets to account for actual or expected scarcity payments (Hogan 2006). The design of better scarcity pricing systems is an active agenda item in the RTO markets in the US (NYISO 2008; ISONE 2006; MISO 2009).

The concern with resource adequacy may be more acute in the US than in the EU market design models. All of the US models include procedures to monitor and mitigate the exercise of market power, primarily through the application of offer caps. These procedures, coupled many other market design nuances, accumulate to depress energy market prices. By comparison, the EU market is more opaque in this dimension, more inclined to structural interventions such as divestiture, and less inclined to directly intervene in the bidding process. The incentive to exercise market power is material, and

---

[6] PJM reports that on average between 1999 and 2008 a combustion turbine would have earned 43% of its fixed charges in energy net revenue (Monitoring Analytics 2009, 143).

[7] The causes of inadequate scarcity pricing are many, including but not limited to regulatory offer caps on supply (Hogan 2005, 27-33).

the opportunity may be greater in the EU context. The net result is a more mixed picture about the adequacy of incentives for new generation investment. Furthermore, at the moment, the more aggressive policies in some EU countries for introducing renewable energy sources, particularly wind, have produced excess capacity and raised total system costs, but the increase in capacity depresses energy prices as they appear in the nominal market (Traber and Kemfert 2009).

Infrastructure Investment

Providing generation resource adequacy is only part of the necessary development of the electricity system. Other necessary infrastructure development includes transmission grid expansion and deployment of system upgrades in the form of better information, smarter controls, and smarter metering. In the vertically integrated electric system these investment decisions were and remain primarily the responsibility of the electricity monopoly, subject to regulatory approval. The restructured markets, however, present a different mix of choices, and in many respects responsibility for infrastructure investment is unresolved.

A principal purpose of development of electricity markets was to provide better incentives and opportunities for decentralized decisions with a corresponding allocation of risk and reward for investments in generation plants and end-use efficiency. In an idealized electricity market structure, energy market prices would be sufficient drivers of generation and end-use investment. As discussed under resource adequacy, the less than ideal reality has not yet achieved this goal for generation, and similar problems arise in providing the incentives for end-use investment.

The case for transmission and related infrastructure investment is even more of a challenge. The transmission grid and the associated dispatch functions of the system operator continue as natural monopolies. Even in the idealized version of electricity market design, it could be difficult to fashion a regime that relied on market based transmission infrastructure investment (Hogan 2009, 152-158). Accordingly, it is widely (but not universally) assumed that some residual monopoly, either a transmission owner, a system operator or a hybrid special purpose organization, must make infrastructure investment decisions, cause the expansion to be implemented, and then use the compulsion through regulation to make the erstwhile unwilling beneficiaries pay.

If the beneficiaries were willing to pay, they could in principle organize to make the transmission investment without the coercive power of regulation. In some cases market participants do make their own decisions to invest and acquire incremental transmission rights (Hogan 2009, 156). However, in other cases it may well be that no coalition of beneficiaries could be assembled to undertake desirable transmission investment. In these cases, a transmission company, the system operator and the regulator must cooperate to implement investment plans.

Designing a hybrid system that could accommodate both market-based and regulatory driven infrastructure investment is a major challenge for liberalized electricity markets. In some regions, the assumption is that such a hybrid system is impossible and only the transmission company can or will make investments. In other regions, there is a struggle to find a workable compromise. An interesting case is the hybrid transmission expansion framework adopted under the New York Independent System Operator (NYISO), modeled after the transmission investment experience in Argentina. In essence, the NYSIO model allows for decentralized investment when the beneficiaries voluntarily assume the responsibility to pay. In cases where the dispersion of benefits makes it difficult or impossible to assemble a coalition of the beneficiaries, the system operator and regulators can go forward subject to a super majority endorsement where the beneficiaries vote and pay for the investment in proportion to the estimated benefits (NYISO 2007, 14-15; Bhudraja et al. 2008). An important feature of both investment avenues is that beneficiaries pay and the costs are not socialized, which provides good incentives and is compatible with the framework for investment in generation and end use without creating conflicting subsidy regimes.

The ability to develop an adequate hybrid infrastructure investment framework depends on the rest of the market design. In systems with broad energy regions with a common nominal energy product and a single price, the details of transmission congestion are hidden and market signals are weak (between regions) or absent (within regions). In these cases there is no incentive for market participants to invest and a transmission company monopoly may be the only workable system. For example, in England & Wales the single transmission owner National Grid is responsible for designing, implementing and charging for transmission investment under the supervision of the regulator. In the US there is a mixture of evolving infrastructure investment schemes. New England follows close to a central planning and full cost socialization system, whereas New York, PJM, Texas, the Midwest and California have various hybrid models in place or in flux.

The details of each of these systems can be complicated, but a general principle connects to the underlying market design. By construction, the more aggregated the market design the less articulated are the investment incentives for transmission. Within aggregated zones, there is no incentive for market participants to invest because transmission constraints are assumed away. Between aggregate zones, transmission investment incentives may exist, but the price information in the market would only dimly reflect the underlying reality of the transmission interactions. Only in the case of the fully locational integrated market designs could energy prices provide accurate information about transmission congestion that provides a framework for awarding feasible transmission rights. But these market designs provide efficient infrastructure investment decisions only when the investment itself has relatively limited impact on market prices. Larger investments that could materially change expected market prices may then require a hybrid framework like the New York model.

Green Energy Resources

As discussed elsewhere in this volume, green energy resources include a diverse array of technologies. Nuclear power and large-scale hydro provide energy without producing material carbon emissions, and their impacts on the electrical system are well understood. Newer low or zero carbon generation technologies such as wind, solar thermal, solar photovoltaic, tidal system, biomass, and fossil fuels with carbon capture and storage offer a variety of non-traditional operating characteristics that could affect electricity systems and markets. Energy efficiency and load management investments could materially alter the load profile and interact strongly with the portfolio of generation and transmission assets.

Desirable policy for pursuing low carbon technologies depends in large part on the associated choices in market design. In some cases, such as nuclear and hydropower, the technology is familiar and well understood. In other cases, such as carbon capture and storage (CCS) or expanded co-firing of biomass, costs may be different but there is nothing obviously fundamentally different from existing fossil fuel generation that could alter market design. In other cases, such as wind or solar, the operating characteristics are significantly different. At small scale, almost any new generation technology can be absorbed, but large-scale integration of the newer low carbon technologies is more complex. And some anticipated changes, such as the development of smart grids, smart appliances, and electrification of the transportation sector, could require or facilitate major changes in system operations and market design.

Uncertainty and Technology Choice

If you know what to do, do it![8] If we knew which technologies to embrace, which infrastructure investments to support, and how opportunities would evolve, then the low carbon investment problem would be simple, and the vertically integrated monopoly electric utility model would have much to recommend it. Regulators would continue to oversee the investment choices and set rates for cost recovery. With little uncertainty the task would be relatively easy, risks would be low, and the lack of efficient incentives for market participants would be less problematic.

Unfortunately, we don't know exactly what to do. By the late 1980s, partly as a result of changes in technology, vertically integrated electric utilities around the world were saddled with high costs and unwanted assets. The resulting pressure led to reform of electricity markets to change the incentives, locus of decisions, and allocation of risks. Part of the animating idea was to promote innovation, entry and flexibility in the electricity system. New technologies and new operating procedures would arise as a result of market incentives and decentralized decisions. Incentives would be better aligned to support innovation in both end use and generation facilities, and innovation would increase. Rewards would go with the risks, and the market participants would do a better job of euthanizing bad investments as well as capitalizing on what works.

The more innovative the idea, the less likely it would be adopted under the regulated vertically integrated organization. Regulation works better when the choices are highly standardized and easy to evaluate. But when the product includes many moving parts or involves a great deal of uncertainty, it can be much harder to write down the rules. In these cases, business judgments can be made and the uncertainty can be priced, with the associated risks and rewards. For example, consider the case of the Enernoc and its demand product that provides operating reserves in markets such as PJM.[9] The de facto

---

[8] "If you know what to do, do it! If you don't know what to do, decision analysis can help you decide." Aphorism from Professor Ron Howard of Stanford University.

[9] http://www.enernoc.com/.

product provided to PJM behaves essentially like reliable standby generating reserve that can be called upon to meet gross load or reduce net load at a set of locations. The actual service, hidden from view of the system operators, is a complicated array of contracts and operating agreements with end users that allow a variety of interventions with various technologies to reduce or temporarily shift electricity load. These contracts are negotiated privately, and voluntarily, to share risks and rewards. There is no need for regulators to approve the complex and varied terms and conditions of the individual contracts. Enernoc takes a reasonable business risk, constrained in its returns through the competition from others in providing similar services. PJM gets innovative access to cost effective operating reserves that fit naturally under the market design. The market design stimulated a market response, but the rules did not prescribe the result.

This broad motivation for electricity market reform is greatly reinforced by the challenges of developing a low carbon electricity system. Carbon emissions have a long duration in the atmosphere and will affect the climate for centuries after their sources have been retired, and generating plants may operate for many decades. Hence we are interested in impacts over very long time scales, and we are hoping for and expecting dramatic innovations in technology and operating procedures. Over this time scale, there is enormous uncertainty about what will work and how the system will operate in practice. The biggest surprise would be if there were no major surprises.

In this setting, it is even more important to engage the creativity and innovations of the many current and potential market participants rather than to rely on the wisdom of a few central planners. If we knew what to do, the central planners could execute. But since we don't know exactly what to do, we need an electricity market design that facilitates learning.

In the face of great uncertainty, the signals and incentives should be as technology neutral as possible. Rather than prescribing or constraining the choices of technology, market operations and prices should reflect the real costs incurred. To the extent possible, externalities should be internalized. For example, rather than mandate technologies, there should be a price on carbon that should affect technology choices.

Among the three market models, the vertically integrated model is most like central planning, and the integrated locational market design is most attuned to reflecting real costs associated with system operations. The aggregated counter-trade models fall somewhere in between, with more opportunities for decentralized investment decisions but more muted incentives to guide those investments.

The need to support innovation, pursuing technologies, investments and consumption choices not yet envisioned by anyone—including the central planner—provides a powerful motivation for getting the signals and incentives aligned to support decentralized decisions. The more the electricity market design relies on socialized costs, muted incentives and regulatory mandates, the more the central planner assumes the burden for making investment decisions and applying regulatory compulsion to recover the costs.

Investment and Operating Cost

There would not be much policy concern with low carbon technologies if they did not appear to be more expensive than fossil-based alternatives. With a price on carbon emissions, the cost differential between fossil fuel plants and low carbon alternatives may be overcome. However, there may remain material differences in costs for nascent technologies entering the market.

The response to any remaining cost differential would depend on the nature of the higher costs and the structure of the market design. If there is a price on carbon but the price is not sufficient to induce adoption of a particular technology, there could be several explanations. For example, it may be that the technology is simply too expensive compared to other alternatives. In this case there is no real problem and the failure to embrace the particular low carbon investment is the correct solution.

Another possibility is that the price of carbon is too low. The best response would be to raise the price of carbon. Absent an adequate carbon price, if there is a policy decision to invest in low carbon technologies that are not cost competitive, the nature of the cost differential would have different impacts under alternative market designs.

An alternative case would be that the price of carbon is appropriate, the current cost of the new low carbon technology makes it not competitive, but learning by doing could lower the costs enough to make the low carbon technology competitive. Again the impacts will be different under alternative market designs.

A market design based on a vertically integrated monopoly will face familiar issues of technology adoption and investment in all three cases. The regulator would oversee investment choices and provide rates for cost recovery. Low carbon technologies could be adopted as a matter of policy, despite their costs. The particular nature of the cost differentials would not be a first order concern.

In the other market models, relying more on prices and other market incentives, the nature of the cost differential would matter more. With higher costs, adoption will require some form of subsidy. If the higher cost is in the form of high investment costs, there are many ways to provide tax credits or other capital subsidies. Once the subsidy is provided, the new generation plant or end-use investment can participate in the ongoing electricity market on the same basis as other investments.

If the source of the higher cost is a higher operating cost the interaction with the market design may be more immediate. Consider for example, an end-use technology like load management to reduce consumption when the avoided cost of electricity is very high. If the market design does not provide the appropriate price signals that capture the avoided cost in the system, often over relative short intervals, there may not be enough incentive to pursue load management even if the up-front investment cost is low or subsidized. Here both the temporal and locational granularity of prices would be important. Similar challenges would arise for generation investments like pump storage, flywheel storage, or pondage hydro, that could smooth out effective load curves but which face high effective operating costs. If the market design does not offer sufficient granularity of price determination, then other cost effective low carbon technologies would be too expensive to operate or would require regulatory mandates for adoption.

Intermittency and Reliability

Of the many technical characteristics of certain low carbon technologies, the feature which draws the most attention is intermittency of the supply and the implications for reliability and meeting system load. The canonical concern is for wind and solar photovoltaic (PV) panels. Once constructed, the variable operating cost of these generation plants is effectively zero. But if the wind doesn't blow, or the sun doesn't shine, there is no power generated.

This intermittency gives rise to two related but different problems. The first is the availability of the power when it is needed. With traditional nuclear, hydro, and fossil fuel plants there is the possibility of scheduling required maintenance to minimize the impact on peak hours, and the random outages that cannot be controlled have a relatively low probability. By contrast, the implied effective capacity factor of 30% or below for individual renewable facilities is equivalent to random outage probabilities of 70% or more. The correlation of these outages with peak and near peak loads would be important, but the basic intermittency is a hurdle for some low carbon technologies.

Another aspect of variable availability of low carbon technologies is the speed with which they come on or, more importantly, go off the system. This "ramp rate" can present material operational challenges for the system. For small scale penetration of the low carbon technologies, there is not much more than including this variability as part of the natural variability of residual load. But on a large scale the challenge would be different depending on the dispatchability of the plant.

A common answer is that the individual variability of solar or wind facilities is less of a problem when there is sufficient regional diversification of sources where winds speeds are not correlated. It is true that the correlation reduces across wider areas (Holttinen et al. 2009).[10] However, exploiting this portfolio effect depends importantly on the configuration of the transmission grid as discussed below.

In principle, for the integrated locational market design, variability in production does not present much of a design problem. Security constrained economic dispatch would impose limits on the output of the plants and reduce prices for energy the more that capacity has to be reserved to meet the potential changes in production. For example, this might involve restricting output from wind facilities when the wind is blowing hard to keep the combined output of the wind farm below the level that the system could handle if the wind suddenly drops. The net effect would be to reduce the locational price of energy and increase the locational price of operating reserves. Most of the cost of variability would then fall on the wind supplier, with the associated incentives to invest in storage or other backup options.

Of course, this appeal to the theory of integrated locational market design contradicts the common assumption that windmills should generate whenever the wind is blowing.

[10] This portfolio effect does not eliminate the night when the sign is down nor necessarily eliminate the cases when even a large area can be becalmed and remove all the wind generation (NERC 2009, 16).

Although this is not correct as an economic matter, because at times free energy is too expensive, there is widespread concern that investment in low carbon technologies should be combined with other investments to ensure that the power can flow when it is available.

The market design implications of the intermittence of supply are familiar. The more granularity, over time and location, in the prices and dispatch, the easier it is to integrate low carbon technologies. The better the prices reflect avoided costs, the easier it would be to integrate low carbon technologies, especially those with high operating costs.

The greater variability of supply for key low carbon technologies and increased end-use response become more and more important the greater the investment in smart grids. As discussed above, smart grids require smart prices. Smarter prices would better reflect scarcity conditions, locational differences, and dynamic response.

In the absence of sufficiently granular prices and quantity definition, incentives for low carbon investment diverge from the real costs of system operations. The details will be different in different markets, but the greater the gap the more the reliance on regulation and central mandates to support low carbon technologies.

Green Infrastructure Investment

The places where the wind blows and the sun shines are not always where people congregate. Hence, from a strictly technical perspective, some of the best locations for siting low carbon technologies like wind and solar energy are far from loads. West Texas has a great deal of wind, but the load is in Dallas and Houston. The result is a substantial need for transmission investment. In Texas there has been a policy decision to move ahead with transmission expansion to enable greater utilization of the wind resources and to socialize the cost through additional charges to all customers.[11]

In other regions, the justification for additional transmission and the associated cost allocation are more problematic. In the case of New York, the large potential wind resources are in the Northwest region of the state, and the major load center is around New York City. There is controversy about whether to build new transmission and who should pay. The integrated locational pricing design sends clear signals about the potential benefits of expanded transmission, but so far the benefits have not been seen as worth the costs. Under the innovative New York transmission investment rules, the beneficiaries must pay and must approve in advance according to super-majority voting. In the case of the proposed New York Regional Intertie, designed to exploit upstate wind, the developers withdrew the plan on the grounds that there would not be sufficient support from the putative beneficiaries to obtain approval (Puga and Lesser 2009).

One of the challenges for transmission investment is the intermittency of the wind and solar energy production. A standard argument is that diversification of sources of supply across large regions where wind and cloud cover are not perfectly correlated would provide a sufficient portfolio effect to mitigate the intermittency. What this argument must confront is the level of transmission investment required. For the diversification

[11] Texas Utility Code § 36.053 (d), Texas Statutes - Section 36.053.

argument to hold fully, transmission capacity must be sufficient to accommodate the maximum output of all of the sources. In the case of dedicated long distance lines to bring power to load, each line must be sized to be fully utilized only a fraction of the time. Otherwise the full diversification benefit would not be realized. This surge capacity investment in transmission, and planned underutilization, could materially add to the costs of interconnection for remote low carbon energy technologies (NAS Committee 2009, 298).

The more the market design adheres to the granularity of time and location, the greater will be the transparency of incentives for transmission investment. The more the market design aggregates across time and location, the more hidden will be the benefits of transmission. The higher the degree of aggregation, the greater will be the tendency to socialize the cost of transmission.

Movement towards the integrated locational market design facilitates a beneficiary-pays system of transmission expansion and supports other features of market investment decisions. Without a beneficiary pays protocol, there is no principled answer to why the subsidies to transmission should not be extended to demand and production alternatives that are partial substitutes for transmission expansion. A beneficiary pays protocol for transmission is an important long term component of market design. Of course, as ready examples in Texas and New England demonstrate, the use of an integrated locational market design for the energy market is no guarantee that transmission investment will be market driven and not socialized.

Green Policy Mandates

Many policies have arisen or might arise to support development and deployment of low carbon technologies. In some cases, these policies would complement broader market decisions. In other cases, the efforts to internalize carbon externalities might conflict with market design features.

Carbon Emission Caps

By far, the most important component of an efficient and effective policy to internalize the climate impacts of carbon emissions is to put a price on carbon, as has been done in the EU and is likely in the US. For all the usual reasons, the carbon pricing regime should to the extent possible be economy wide and global in its coverage (Stavins 2007). The details of the policy choice of a carbon tax versus a cap-and-trade proposal are important and raise many other issues.

A carbon cap with trading of emission permits is widely accepted as being more politically feasible because of the greater opacity of the revenue transfers inherent in allocation of the permits. A principal feature of a good allocation design is to break any connection between actual energy production and the aggregate allocation of the permits.[12] In this event, free permit allocations involve a transfer of wealth but provide no reduction in the cost of carbon emissions on the margin.

A key element of a workable policy will be the credibility of the carbon regime and the exposure to future political risk (Neuhoff et al. 2009). Major investments in expensive low carbon technologies would be profitable only if sustainable public policy ensures that there is a significant implicit or explicit price on carbon emissions that is expected to high enough for long enough to recover the substantial investment.

Sustainability of public policy will depend in part on compatibility with the basic market design. For example, an economy wide cap on carbon emissions would be largely technology neutral and would integrate well with efficient market operations. Investors in a particular technology would not need to worry as much that their specific venture would be subject to idiosyncratic subsidy changes. In order to change the economics of the individual investment, the carbon cap would need to be changed for all emissions, and this is less likely than a change in a targeted subsidy. By contrast, the boom and bust cycle of solar deployment in the US, driven by the dependence on periodic reauthorization of targeted tax credits for solar investments, illustrates the risk of developing a sustained commitment to investment based on special purpose support (NAS Committee 2009, 308). Comprehensive carbon emission caps and good market design would reinforce the sustainability of the system.

Research, Development and Demonstration

The usual arguments imply that there should be major spillover benefits to research, development and even early demonstration (RD&D) efforts to discover and refine better low carbon technologies. Instituting policies to put a price on carbon would reinforce incentives for private RD&D, but there is no reason to believe that the full externalities of the public good aspect of RD&D would be overcome by putting a price on carbon. There are major challenges in mounting an RD&D program to support energy innovation (Anadon and Holdren 2009). However, the challenges are not much affected by different wholesale electricity market designs.

Infant Industries and Learning by Doing

Commercialization beyond RD&D and learning by research interact with deployment and learning by doing. A major argument in favor of public investment to support early adoption of low carbon technologies is the ability to capture the benefits of early investment which accumulate to reduce the going-forward cost of the technology. The externality arises when the benefits of information about the success of the technology and the improved understanding of how to build and operate cannot be captured by the investor. Thus learning by doing is a type of infant industry argument, where early support is needed to launch the technology but once mature the technology is able to compete on its own without further subsidy.

[12] Initial designs in the EU included allocations for new power plants as a set aside under the total cap. This creates the incentive to build new plants and raises the cost for existing plants by reducing the net cap for the existing fleet.

An interesting feature of the application of learning by doing analysis to energy technologies is the substantial nature of the putative benefits. High learning rates generate such substantial reductions in total electricity generation cost that most of the benefit comes from improved efficiency of the electricity system, and carbon emission reduction benefits are a small component of the gain. In some cases, if the claimed learning rates are true, it would pay to provide public support for the technology even if carbon emissions posed no cost (van Bentham et al. 2008).

This argument about large potential cost reductions might apply to wind and solar generation technologies that are alternatives to fossil fuel generation but cannot apply to CCS technologies. Although there may be substantial learning to come in CCS design and deployment, the remaining costs will always be in addition to the cost of fossil fuel generation. Even if CCS were effectively free, it would not be cheaper than using the fossil fuel plant without CCS. Hence, all the benefits for CCS come from the carbon reduction.

Policies to support learning by doing should be structured to recognize the distinction between supporting learning about the technology (through early cumulative investment) and supporting the technology (by offering sustained subsidies with little or no further learning). Quantitative targets for low carbon technology investment may have little connection with the public benefits of learning. For instance, the case of the California solar rooftop initiative is instructive. Launched with a promise to install solar PV on a million rooftops by 2015, the idea was to both reduce carbon emissions and capture the benefits of learning by doing. However, while the assumed learning rate is an input, the necessary scale of the investment should be an output of the analysis and not a constraint on the policy. In the California case, further analysis revealed both the implied trajectory of the optimal subsidies and investment profiles as a function of the assumed learning rate. A central conclusion was that the optimal subsidies would be associated with investment at only about a quarter of the scale of the nominal million rooftop target. Higher learning rates needed less initial subsidy, and lower learning rates might not produce enough benefit to justify any material subsidy (van Bentham et al. 2008).

Proper design of support policies to capture the benefits of learning by doing interact with market design in straightforward ways. The more transparent the market design, and the clearer the market signals, the easier to achieve the benefits of the learning, and to know whether the benefits are appearing. In the energy market, the better the price signals, to include scarcity costs, the better the opportunity for private investment to drive the benefits of lower costs.

Renewable Energy Standards

The competition between incentives and targets appears most starkly in the case of renewable energy standards.[13] Motivated by the market model of the vertically integrated electric utility, with a single buyer, the idea is to promote penetration of low carbon technologies by setting mandates for the percentage of electricity generation that must be provided by defined categories of renewable or low carbon energy sources. Such goals

13 Also known as renewable portfolio standards.

supported by feed-in tariffs have been adopted as targets throughout the EU and as quantity mandates in many states in the US. Experience with this policy instrument presents a number of difficulties, some of which interact with the nature of the market design.

The standards often lack a principled basis. A cap on carbon emissions, where the emissions of carbon can be measured and controlled, is technology neutral. By contrast, renewable energy standards require selecting the acceptable technologies. The result has been a substantial degree of political bargaining with variation across jurisdictions revealing the fundamental disagreements about goals and means. For instance, in the US state programs define a wide array of more than a dozen possible renewable technologies, yet "...states agree on only three technologies; biomass conversion, solar photovoltaic and wind." (Michaels 2008, 10)

The renewable energy standards also suffer from confusion about objectives and a poor mapping between policy and goals. Requiring minimum penetration standards to support early development of expensive or risky renewable technologies is akin to the quantity target and subject to the comments above about the mismatch between learning-by-doing justification and quantitative targets.

Or consider the justification that the technology would never be competitive on its own, and the minimum standard is intended to reduce carbon emissions. While this approach may have an appeal in the absence of a carbon cap, the argument no longer applies in the presence of a cap. As soon as there is a binding cap on carbon emissions, an RES produces no incremental carbon emission benefits. If the cap is binding, the effect of an RES is to require more expensive investments that reduce carbon emissions and therefore lower the cost for everyone else and every other technology in meeting the carbon cap. Hence, the proposed 33% RES by 2020 in California should raise costs of control in California, to the benefit of the rest of the country.[14] But once a cap-and-trade system is in place, carbon emissions would be the same as without the RES. In effect, the RES lowers the market price for carbon emission permits and raises the social cost of reducing carbon emissions but does not lower carbon emissions.[15]

In market designs without vertical integration, RES presents the challenge of how to implement and enforce the technology constraint. One approach is to create RES certificates and require all buyers or all generators to redeem a certain percentage of RES generation credits or pay a tax.[16] The price of the credits is then a factor in the price of electricity and alters the economic dispatch. However, there is an imperfect connection with carbon emissions, since low carbon technologies like nuclear are often excluded from RES credits. Thus under the preferred cap-and-trade system there would be parallel trading systems to control emissions and promote renewableness. There is nothing inherently infeasible about having multiple permits systems. For example, the carbon

---

[14] California Executive Order, November 17, 2008, http://gov.ca.gov/executive-order/11072/.

[15] The argument is slightly different for a carbon tax, where the price of carbon is fixed. Then an RES adds to costs by requiring use of more expensive low carbon technologies, but it does reduce carbon emissions.

[16] Renewable Energy Certificates System. (www.recs.org).

permit would be layered on top of tradable sulfur dioxide emission permits. But the added complexity does raise a question about the incremental benefits of RES mandates.

Green Uneconomic Dispatch

Another approach to promoting green energy resources would be to mandate priority for low carbon technologies to be selected first in energy dispatch (Dubash 2002, 25). This idea presents more complications than might first appear because of the nature of the transmission system interaction with electricity market design.

The complex interaction of power flows in the transmission system can produce cases where actions at a location to reduce energy use or increase emission-free electricity production can actually increase carbon emissions in the overall system (Rudkevich 2009). Meeting electricity load at certain locations creates counter flow in the grid that reduces congestion. Reducing that counter flow can create a requirement for a multi-plant redispatch that increases overall carbon emissions. The technical condition is similar to the effects of transmission counter flow which can drive locational energy prices negative even when all generation offers are positive. Hence the system impacts on carbon emissions are not always easily predictable.

This might support a call for a green energy dispatch protocol in place of economic dispatch, but the interaction with energy market design would be disruptive. In the case of the vertically integrated monopoly there would be less of a problem. By assumption, the monopoly can internalize all the untoward effects in the production and delivery chain, and ignore the internal transfer prices.

By contrast, the market structures now based on bid and offers in a framework of security constrained economic dispatch would confront a substantial disconnect between the economic prices applied in the settlements system and the opportunity costs of changes in the dispatch. As we have seen from much experience, whenever market prices used in settlements systems diverge materially from the costs of real system operations, arbitrage opportunities and temptations are created that could undermine the dispatch and the market.

It would be possible to impose a constraint on emissions in the dispatch (over very short intervals) and then apply the usual principles of economic dispatch subject to the emissions cap. This would produce consistent market prices with scarcity components reflecting the cost of carbon emissions implied by the constraint. In effect, this would replace an economy-wide cap on emissions, allowing implicit trading across time and space, with a new cap-and-tax system with short-term caps and implicit taxing at the implied variable carbon price. Absent a formal electricity-only-cap-and-trade system that allowed intertemporal balancing, the system operator would have to decide on the limit for virtually every dispatch period. It would be reasonable to anticipate that the implicit (very) short-term carbon price would be highly volatile.

A far simpler approach would be if the cap on aggregate emissions and the implied carbon price in the economy-wide cap-and-trade system were sufficient to internalize the cost of carbon. Then the carbon emission permit costs, along with all other costs for fuels and other emissions, would be treated the same in the security constrained economic dispatch. In the integrated locational market design the associated locational marginal prices would include the effects of carbon emissions. There would be no additional need for an incremental green dispatch that could break the connection between market design and real operations.

Feed-In Tariffs

Green energy mandates that require system operators to accept energy from approved technologies at approved fixed prices, known as feed-in-tariffs, have been widely adopted in the EU. Not surprisingly, with high enough prices, such mandates have been highly successful at increasing the market share of approved sources such as wind and solar installations (Muñoz et al. 2007, 3104). However, the structure of the feed-in tariff creates collateral damage similar to the effects of green uneconomic dispatch, and it is not clear whether the model is consistent with developing sustainable growth of low carbon investments.

Set aside the question that the new technologies are expensive and the expense is becoming a matter of policy concern in the EU.[17] The feed-in-tariff interacts poorly with the rest of the market design in terms of system operations, price signals, and infrastructure investment.

The problem with system operations appears because the system operator must-take the energy input, even when the resulting power flow might foreclose other generation alternatives, except in emergency conditions. In the case of wind, for example, the variable cost of wind energy may be zero, but this does not mean that it is always cost-effective to accept the wind energy. There are conditions on the grid where the value of the wind energy is negative. In effect, the feed-in-tariff forces wind energy onto the grid and drives the implicit (in the EU) or explicit (in the US RTOs) locational price of energy substantially below zero.[18]

Since the mandate is to take the green energy, the burden of investing in transmission connections and upgrades falls on someone else, either the transmission company or through some method of cost socialization through the system operator. Besides the obvious total cost inefficiencies, this cost socialization distorts the choices among renewable energy resources, as discussed below.

The simple alternative to the feed-in tariff would be some mix of investment tax credit or direct subsidies that reduce the cost of investment (but not operations). In order to provide incentives for operations, collection of the subsidy could be conditioned on offering the capacity in the real time market, whereupon the usual principles of security constrained economic dispatch would apply.

---

17 “The new German coalition government will cut feed in tariffs for solar power by more than the planned 8% reduction for 2010, according to a 128-page policy document published on Friday. The tariffs will now be reviewed every three years,” ENDS Europe, Monday 26, October 2009, (http://www.endseurope.com/22460).

18 For example, this is a regular event in PJM.

Note that conditioning the subsidy on offering the energy to the market is not the same thing as conditioning the subsidy on the energy actual produced. System operators have good experience in discriminating good faith offers of energy, and need not actually dispatch the plant whenever it is offered to verify the bona fides of the plant. But if the subsidy is conditional on actually producing the energy, the effect would be to convert the fixed investment subsidy into a variable opportunity cost. In this case, the wind or solar plant could profitably enter a negative bid (bounded below by minus the variable subsidy value) to ensure it was dispatched. This would distort prices and exacerbate the problem of the technology subsidies sending the wrong signals to the market.

Transmission Investment

A major challenge remains in defining the appropriate protocols for transmission investment. The general problems of constructing a satisfactory hybrid system, as outlined above, are magnified by the perception of many that a major expansion of transmission capacity is required to build the long-distance connectors between loads and the major potential sources of renewable energy (NAS Committee 2009, 296).

The principal implication for green energy mandates would be to decide on the degree to which transmission costs are considered in the competition among green energy resources. Which major expansions of transmission systems are necessary is far from obvious. Solar rooftops installations may require relatively little, and certainly different transmission investments than distant wind or solar installations. Renewable energy sources are partial substitutes for each other, and there is no reason to believe that every renewable option should be pursued. It follows that the model for transmission investment and cost allocation could have a major impact on the choice of renewable technologies. If transmission investment cost is socialized, for everything or even only for renewable sources, then distant wind will look cheaper and nearby rooftop solar more expensive, even when the opposite ranking may be true.

An obvious principle to apply to green transmission investment, as for any transmission investment, would be to adhere to a beneficiary pays system (Hogan 2009, 152-158). If the requirement to pay for transmission makes a preferred renewable technology more expensive, then this can be addressed directly in the balance of subsidies. In the presence of a carbon cap, with the cap accepted as the proper balance of carbon costs and benefits, then there is no good reason to treat green transmission investment any differently than other transmission investment.

Conclusion

Policy supporting low carbon technologies and end-use choices in the electricity sector interacts with electricity market design. There is great uncertainty about the appropriate investments and a consensus that achieving a low carbon future will require innovation and invention. Success is more likely if appropriate price signals and incentives appear to drive decentralized decisions of the many participants in electricity markets. Placing a price on carbon is a critical step. Getting the resulting incentives right will be easier the closer the electricity market design reflects the reality of electricity system operations. The experience with electricity market alternatives points to the integrated locational electricity market design as the only workable model that is compatible with open transmission access and non-discrimination.

References

M. Amin, *Automation, Control, and Complexity: An Integrated Approach*, Samad & Weyrauch (Eds.), John Wiley and Sons, 2000, pp. 263-286.

Laura Diaz Anadon and John P. Holdren, “Policy for Energy Technology Innovation,” in Kelly Sims Gallagher (ed.), *Acting in Time on Energy Policy*, Brookings Institution Press, Washington DC, 2009, pp. 89-127.

Arthur van Benthem, Kenneth Gillingham and James Sweeney, “Learning-by-Doing and the Optimal Solar Policy in California,” *The Energy Journal*, Vol. 29, No. 3, 2008, Pages 131-152.

Jacqueline Boucher and Yves Smeers, “Towards a Common European Electricity Market Paths in the Right Direction Still far from an Effective Design,” *Journal of Network Industries*, Vol. 3; No. 4, 2002 pages 375-424.

Vikram Budhraja, John Ballance, James Dyer, Fred Mobasheri, Alison Silverstein, and Joseph H. Eto, “Transmission Benefit Quantification, Cost Allocation and Cost Recovery,” California Energy Commission, Energy Research and Development Division, 2008.

Committee on America’s Energy Future (NAS Committee), *National America's Energy Future: Technology and Transformation*, National Academy of Science, Washington, DC, 2009.

Navroz K. Dubash, “The Changing Global Context For Electricity Reform,” Chapter 2 in in *Power Politics: Equity and Environment in Electricity Reform*, World Resources Institute, Washington, DC, 2002.

ENTSO-E, Regional Group Continental Europe, Statistical Yearbook, 2008.

Electric Power Supply Association, “EPSA Statement on Renewable Energy in Organized Electricity Markets,” Press Release, Washington, DC, January 10, 2008.

Michael Grubb, Tooraj Jamasb and Michael G. Pollitt, “A Low-Carbon Electricity Sector for the UK: What Can Be Done and How Much Will It Cost?”, in Michael Grubb, Tooraj Jamasb and Michael G. Pollitt (eds.) , *Delivering a Low-Carbon Electricity System*, Cambridge University Press, 2008.

Udi Helman, Harry Singh, and Paul Sotkeiwicz, “RTOs, Regional Electricity Markets, and Climate Policy,” in Fereidoon P. Sioshansi, Generating Electricity in a Carbon-Constrained World, Academic Press, Burlington, MA, 2010, pp. 527-563.

William W. Hogan, “Contract Networks for Electric Power Transmission,” *Journal of Regulatory Economics*, 1992, Vol. 4, pp. 211-242.

William W. Hogan, “Getting the Prices Right in PJM: Analysis and Summary: April 1998 through March 1999, The First Anniversary of Full Locational Pricing,” Harvard University, available at (www.whogan.com), April 2, 1999

William W. Hogan, “Electricity Market Restructuring: Reforms of Reforms,” *Journal of Regulatory Economics*, Kluwer Academic Publishers, Vol. 21, No. 1, 2002, pp. 103-132.

William W. Hogan, "On An 'Energy Only' Electricity Market Design For Resource Adequacy," Center for Business and Government, John F. Kennedy School of Government, Harvard University, September 23, 2005, (available at www.whogan.com).

William W. Hogan, "Resource Adequacy Mandates and Scarcity Pricing ('Belts and Suspenders')," comments submitted to the Federal Energy Regulatory Commission, Docket Nos. ER05-1410-000 and EL05-148-000, February 23, 2006, (www.whogan.com).

William W. Hogan, "Electricity Market Structure and Infrastructure," in Kelly Sims Gallagher (ed.), *Acting in Time on Energy Policy*, Brookings Institution Press, Washington DC, 2009, pp. 128-161.

Hannele Holttinen, Peter Meibom, Antje Orths, Bernhard Lange, Mark O'Malley, John Olav Tande, Ana Estanqueiro, Emilio Gomez, Lennart Söder, Goran Strbac, J Charles Smith, Frans van Hulle, "Impacts of Large Amounts of Wind Power on Design and Operation of Power Systems, Results of IEA Collaboration," 8th International Workshop on Large Scale Integration of Wind Power into Power Systems as well as on Transmission Networks of Offshore Wind Farms, Bremen, Germany, Oct. 2009.

International Energy Agency, *Tackling Investment Challenges in Power Generation in IEA Countries: Energy Market Experience*, Paris, 2007

ISONE: FERC Electric Tariff No. 3, Market Rule I, Section III.2.7, February 6, 2006.

Paul Joskow, "Challenges for Creating a Comprehensive National Electricity Policy," Speech given to the National Press Club, Washington, DC, September 26, 2008, (http://www.hks.harvard.edu/hepg/Papers/Joskow_Natl_Energy_Policy.pdf).

Craig B. Lankford, James D. McCalley, Narinder K. Saini, "Bibliography On Transmission Access Issues," *IEEE Transactions on Power Systems*, Vol. 11, No. 1. February 1996.

Robert J. Michaels, "A Federal Renewable Electricity Requirement, What's Not to Like?," Cato Institute, Washington DC, No. 627, November 13, 2008.

MISO: FERC Electric Tariff, Volume No. 1, Schedule 28, January 22, 2009.

Monitoring Analytics, LLC, Independent Market Monitor for PJM, *2008 State of the Market Report for PJM*, Volume 2: Detailed Analysis, March 2009.

Miquel Muñoz, Volker Oschmann, and J. David Tábara, "Harmonization of Renewable Electricity Feed-in Laws in the European Union," *Energy Policy*, Vol. 35, 2007, pp. 3104-3114.

Karsten Neuhoff, Sam Fankhauser, Emmanuel Guerin, Jean Charles Hourcade, Helen Jackson, Ranjita Rajan, John Ward, "Structuring International Financial Support to Support Domestic Climate Change Mitigation in Developing Countries," Cambridge, United Kingdom, September 2009.

New York Independent System Operator (NYISO), Inc Docket No. OA08-13-000, "Order No. 890 Transmission Planning Compliance Filing," Cover Letter Submitted to Federal Energy Regulatory Commission, December 7, 2007, pp. 14-15.

NYISO: Ancillary Service Manual, Volume 3.11, Draft, April 14, 2008, pp, 6-19 to 6-22.

North American Electric Reliability Corporation, *Special Report: Accommodating High Levels of Variable Generation*, Princeton, April 2009.

PJM, "PJM Open Access Transmission Tariff, Subpart A – Interconnection Procedures," Sixth Revised Volume No. 1, Para. 36.1, June 8, 2009.

Potomac Economics, Ltd., Independent Market Monitor for the ERCOT Wholesale Market, *2008 State Of The Market Report For The ERCOT Wholesale Electricity Markets*, Texas, August 2009.

J. Nicolas Puga, Jonathan Lesser, "Public Policy and Private Interests: Why Transmission Planning and Cost-Allocation Methods Continue to Stifle Renewable Energy Policy Goals," Electricity Journal, Elsevier, December 2009, pp. 7-19.

Aleksandr Rudkevich, "Economics of CO2 Emissions in Power Systems," CRA International, February 15, 2009. Available at (www.hks.harvard.edu/hepg/).

Fereidoon P. Sioshansi, *Competitive Electricity Markets: Design, Implementation, Performance*, Elsevier, Oxford, 2008.

F. C. Schweppe, M.C. Caramanis, R.D. Tabors, and R.E. Bohn, *Spot Pricing of Electricity*, Kluwer Academic Publishers, Norwell, MA, 1988.

Robert N. Stavins. "A U.S. Cap-and-Trade System to Address Global Climate Change," Hamilton Project Discussion Paper, 2007-13, The Brookings Institution, Washington, DC, October 2007.

Thure Traber and Claudia Kemfert, "Gone with the wind? - Electricity Prices and Incentives to Invest in Complementary Thermal Power Plants under Increasing Wind Energy Supply," German Institute for Economic Research, (DIW Berlin), January 2009.

Endnote

[i] William W. Hogan is the Raymond Plank Professor of Global Energy Policy, John F. Kennedy School of Government, Harvard University and a Director of LECG, LLC. This paper draws on work for the Harvard Electricity Policy Group and the Harvard-Japan Project on Energy and the Environment. The author is or has been a consultant on electric market reform and transmission issues for Allegheny Electric Global Market, American Electric Power, American National Power, Aquila, Australian Gas Light Company, Avista Energy, Barclays, Brazil Power Exchange Administrator (ASMAE), British National Grid Company, California Independent Energy Producers Association, California Independent System Operator, Calpine Corporation, Canadian Imperial Bank of Commerce, Centerpoint Energy, Central Maine Power Company, Chubu Electric Power Company, Citigroup, Comision Reguladora De Energia (CRE, Mexico), Commonwealth Edison Company, COMPETE Coalition, Conectiv, Constellation Power Source, Coral Power, Credit First Suisse Boston, DC Energy, Detroit Edison Company, Deutsche Bank, Duquesne Light Company, Dynegy, Edison Electric Institute, Edison Mission Energy, Electricity Corporation of New Zealand, Electric Power Supply Association, El Paso Electric, GPU Inc. (and the Supporting Companies of PJM), Exelon, GPU PowerNet Pty Ltd., GWF Energy, Independent Energy Producers Assn, ISO New England, Luz del Sur, Maine Public Advocate, Maine Public Utilities Commission, Merrill Lynch, Midwest ISO, Mirant Corporation, JP Morgan, Morgan Stanley Capital Group, National Independent Energy Producers, New England Power Company, New York Independent System Operator, New York Power Pool, New York Utilities Collaborative, Niagara Mohawk Corporation, NRG Energy, Inc., Ontario IMO, Pepco, Pinpoint Power, PJM Office of Interconnection, PPL Corporation, Public Service Electric & Gas Company, Public Service New Mexico, PSEG Companies, Reliant Energy, Rhode Island Public Utilities Commission, San Diego Gas & Electric Corporation, Sempra Energy, SPP, Texas Genco, Texas Utilities Co, Tokyo Electric Power Company, Toronto Dominion Bank, Transalta, Transcanada, TransÉnergie, Transpower of New Zealand, Tucson Electric Power, Westbrook Power, Western Power Trading Forum, Williams Energy Group, and Wisconsin Electric Power Company. The views presented here are not necessarily attributable to any of those mentioned, and any remaining errors are solely the responsibility of the author. (Related papers can be found on the web at www.whogan.com).