Electricity Wholesale Markets: Designs Now and in a Low-carbon Future

*Richard Green**

This paper compares electricity wholesale markets in the United States and Europe. The Standard Market Design in the US involves an independent system operator, nodal pricing with financial transmission rights, and integrated markets for capacity and ancillary services. In Europe, there are national, or occasionally zonal, spot markets run by companies independent of the transmission operator, and of the latter's purchases of ancillary services. As the amount of low-carbon generation increases, prices and transmission constraints are likely to become more volatile, increasing the need to adopt an efficient market design. In most respects, the US standard market design is likely to give better results than the European models.

1. INTRODUCTION

It is a great pleasure to be able to contribute to this Special Issue of the Energy Journal in honor of David Newbery. I spent the first ten years of my career working with him on a series of projects funded by the Economic and Social Research Council, and he supervised my PhD. David was an excellent supervisor, generous with his time in discussion and in commenting on my work, happy to give me the freedom to follow my own ideas, but willing to make it clear when the time had come to concentrate on writing up. Working with David as a co-author allowed me to see a master at work. While I eventually left Cambridge, I am honored to be an Associate of the Electricity Policy Research Group that David leads, and I have been delighted to see how it has blossomed over the last few years. And while my own work with David has been almost entirely in the area of electricity, no-one should forget the other areas of economics that he has been working in – at a very high level – at the same time.

The Energy Journal, Special Issue. The Future of Electricity: Papers in Honor of David Newbery.

* Institute for Energy Research and Policy, University of Birmingham, Birmingham B15 2TT. Tel: +44 121 415 8216. Email: r.j.green@bham.ac.uk.

The aim of this paper is to consider the designs used in electricity wholesale markets from around the world, and ask how much market design matters for the move to a low-carbon electricity system. Casual observation quickly reveals that many different market designs have been adopted around the world. For the purposes of this paper, I will compare two main families, which I will describe as the United States and the European market designs. I do not want to suggest that all markets within the US, let alone all those within Europe, are the same, for there are many important differences within each family. It is the case, however, that European markets generally share some common features, and the US markets share a different set of common features. I want to explore the implications of those differences for market performance, now and in a future that will see a significant increase in renewable and distributed generation.

In the US, the main electricity wholesale markets are run by the Independent System Operator (ISO) responsible for transmission within each market area, while in Europe, separate companies usually run the wholesale markets and the transmission system. This division forces the European system operators to acquire ancillary services, such as reserve, separately from the energy market, while the US ISOs can acquire energy and most of their ancillary services in a single, co-optimized, process. In some US markets, in response to fears that this process would not give peaking generators sufficient revenues, capacity markets have been created to provide an extra revenue stream, while most European countries have no capacity market. The US markets can set separate prices for every point on the network, and use these to charge for transmitting energy from point to point, whereas European prices cover either a whole country or a large zone within one. This means that the European system operators have to take separate actions to deal with constraints on the transmission system (when present), whereas the main market prices signal these constraints, and the actions required, to US market participants.

Both systems are clearly capable of working, in the sense of allowing companies to trade electricity and deliver it to customers with a high level of reliability. The question I want to ask is which system is likely to work better, promoting more efficient outcomes, not just now, but in the future. It is hard to predict the future shape of the electricity industry, but it seems clear that it will be dominated by the need to reduce carbon emissions. There is a centralized route to doing so, based on nuclear power and large fossil-fuelled stations with carbon capture and sequestration. There is also a decentralized route, with renewable generation, typically dispersed in the areas where renewable resources are found, and distributed generators using combined heat and power (CHP) technology. In practice, it is likely that a mix of centralized and decentralized generation will be required, but that the future system will contain much more decentralized generation than at present. Furthermore, many of the decentralized generators will be intermittent – wind, wave and tidal generators depend on the availability of their power source, while CHP generators are most (technically) efficient when their electrical output follows their heat demand.

What will this increasing intermittency mean for the performance of different wholesale market designs? In the next section, I discuss the principles that a good wholesale market design should follow. Section 3 describes the common features of the European power markets, while the following section 4 discusses the key aspects of the US market design. Section 5 makes a first comparison between them. Section 6 considers how the challenges facing wholesale markets will change as the amount of low-carbon generation rises over the next decade or so, and section 7 considers the implications for our ranking. Section 8 sums up.

2. WHOLESALE MARKETS

First, however, we should define what we mean by the electricity wholesale market, and ask what we want it to do. "The wholesale market" is actually the combination of many separate ways of trading power, and the interactions between them are so important that no single market should be studied in isolation. In many countries, the key component, and the part that looks least like other commodities markets, is a day-ahead auction to sell power. Running the auction one day in advance gives sufficient time to plan the operation of (relatively) inflexible power stations, while still allowing reasonably accurate forecasts of demand and plant availability. These forecasts will be wrong, however, and so other markets are frequently used to make adjustments in real time, although in some countries, these adjustments are procured by the grid operator without the use of a market. Day-ahead prices are typically volatile, and companies wanting to avoid this volatility will try to trade most of their power well in advance on a forward market, which typically operates through over-the-counter trades rather than an organized auction. There may also be a long-term market for capacity or for other services that the grid operator wishes to procure well in advance.

When it comes to actual operation, the role of the system operator is critical. The electricity supply system is an extremely complex machine that must always be kept in balance, with generation equal to demand. If reliability is to be maintained, it is essential that some plants are able to increase output in the event of an unexpected increase in demand or a failure elsewhere on the system. Some of this response must be automatic, faster than human operators could issue instructions (regulation, or frequency control), and if it is called upon, other plants (reserve) must be instructed to increase output so that the margin needed for frequency control is restored as quickly as possible. Not all plants can provide these services – they must be physically capable of increasing output quickly, and many plants can only do this if they are already running part-loaded. That involves both an opportunity cost, in that they cannot sell the units they are not generating, and (usually) a reduction in the efficiency with which they generate the units that they do sell.

Geography also matters, at least in principle. Power flows according to Kirchhoff's Laws, and will use all the lines in a network between generators and loads. If one component fails, the flows will instantly re-direct themselves. If these redirected flows were to overload any component of the network, protection equipment (a fuse) would usually break the circuit in order to prevent damage to that part of the network, but this would lead to a further redistribution of the flows. It is very likely that this would lead to further overloads, more protection would be triggered, and there would be a cascading failure. The way to minimize the risk of such a failure is always to operate the system in such a way that any one failure will not lead to an overload – the so-called "N-1 condition." System operators have to respect both the thermal limits on each piece of the network – the amount of current that it can take without overheating – and also voltage limits. As the flows on an overhead line increase, the voltage at the far end will tend to fall, and so-called reactive power must be injected in order to keep the voltage at the correct level. This can be provided by generators (at the cost of a reduction in their output of "active" power), by some loads, or by devices installed on the network (static compensators).

I wrote that geography only matters "in principle" because if an area has a large enough amount of transmission capacity relative to generation and loads, then the flows on each line will be so low that the system will not risk breaching voltage or thermal limits, even if they increase after a failure. In that case, the system operators need not make any adjustments to the original dispatch to take account of transmission issues. More usually, however, generators which are too far from loads (at least in an electrical sense) may have to reduce output, in order to reduce the (potential) flow across critical network components. Those which are in a better position, relative to the load, would then be required to increase output to make up the difference.

How should we judge whether a market is successful? The Stanford Energy Modeling Forum set out six principles that should be followed when designing electricity transmission pricing (Green, 1997, p. 178). These were intended to cover both the use of system charges levied directly by the transmission owner, which normally provide the bulk of its revenues, and any transmission-related elements of wholesale prices. David and his co-authors have produced a very similar list (Brunekreeft et al., 2005, p. 75).[1]

The same principles can also be applied to wholesale markets, but the focus would shift from transmission to generation. Wholesale markets should:

1. ensure the efficient day-to-day operation of the generation sector;
2. signal the need for investment in generation and demand-side management;
3. promote efficient locational choices for these investments;
4. compensate (sufficiently) the owners of existing generation assets;
5. be as simple, transparent and stable as possible; and
6. be politically implementable.

1. Their list is equivalent (in their order) to points 1, 3, 2, 6, and 4 in the Stanford list, and the list for wholesale prices suggested here.

The first criterion has been strengthened, compared to the Stanford list (which suggested that transmission prices should "promote" efficient operation), since we are now considering the sum of wholesale trading arrangements, rather than a set of transmission pricing arrangements that might have only annual changes. In a market-based system, if the market arrangements do not lead to efficient operation, there will be no plan B. This essentially requires that the markets send the correct signals for operation, either through prices to which market participants respond, or through direct instructions from the system operator, coupled with payments that make participants willing to accept those instructions. The second principle considers long-term signals, and the need to ensure that generators (and those who would need to invest in order to provide demand-side response) are given enough incentive to do so. Since investment in generation involves significant time lags and very long-lived assets, wholesale markets will rarely extend far enough forward to give investors much certainty about the future revenues of a given project. What they can do, however, is send a strong signal about when to close capacity (a decision with a much shorter lead time), and provide generators with confidence that they will earn sufficient revenue if the market is not suffering from excess capacity.

The third principle also relates to signaling, but to its spatial aspect, and has been downgraded slightly, since other mechanisms, and in particular charges paid to the transmission company, are also available to send these signals. Even so, if the wholesale market rules do not penalize a generator located on the wrong side of a transmission constraint – and some designs do not – then the transmission charges will have more work do to in providing the correct incentives. The fourth principle relates to the second function of prices, which is to distribute resources between economic agents, and refers to the need for generators (and others) to receive sufficient remuneration for the services they provide. Sufficient must be defined carefully, however, for revenues can be too high as well as too low. If generators are able to exert market power to increase their revenues, the industry is performing poorly, although the extent to which specific market designs can enhance or reduce market power is an open question (Newbery and McDaniel, 2003; Fabra *et al*, 2006). Similarly, if the market is grossly over-supplied with capacity, then not all of that capacity *should* be able to recover its full costs. If it were able to do so, the incentive for investment in that market would be too strong. There would be little incentive for investment, however, if stations that the market needed were unable to cover their costs from their market revenues.[2]

The fifth principle, that of simplicity and transparency, has been expanded to include stability, and an important rider added. The electricity industry is inherently complex, and no legislator can repeal Kirchhoff's laws. A trading system that ignores these complexities is doomed to failure, but one that is perceived to be

2. An anonymous referee asked whether the use of the term "existing generators" was meant to imply a distinction between incumbents and entrants. It is not – the wording was chosen to mimic the earlier list, and retained on the basis that as soon as an entrant is in a position to earn wholesale market revenues, its plant can be counted as "existing".

more complicated than it needs to be will not be attractive to potential participants – which is important if the market is voluntary, or new investment is needed. Since overly frequent rule changes add to the complexity of the market, and can act as a deterrent to entry by smaller companies (Baldick and Niu, 2005), stability can be interpreted as applying to the market rules. It should not be too easy to change the rules of the market, but neither should it be too difficult, and this requires appropriate governance procedures. A second dimension of stability is in terms of market outcomes, and prices in particular. The challenge here is that when market conditions change, so should the outcomes. Hedging contracts, however, allow market participants to stabilize their costs and revenues while still facing efficient prices at the margin.

Finally, the market must be able to attract, and retain, the support of politicians and other stakeholders, which the Pool of England and Wales clearly failed to do (Newbery, 1998). Prices that appear unnecessarily high or volatile will sacrifice political support, and may lead to unfortunate interventions, such as those in California (Blumstein et al, 2002). Since the Californian debacle, political support for liberalized power markets has been eroded in much of the US, with eight states suspending their restructuring programs.[3] The European Commission has been more successful in persuading Member States to continue with liberalization. However, the debates about whether or not to liberalize have generally been at a level far removed from the details of wholesale market design.[4] And it is to those details that we now turn.

3. MARKET DESIGN IN EUROPE

The world's first fully-functioning liberalized electricity markets were European. The Electricity Pool of England and Wales was established in 1990, and the Norwegian spot market, which has gradually expanded to become Nord Pool, in 1991. The Spanish electricity market was established in 1998, and most other countries in Western Europe followed within a few years. This evolutionary, and national, approach means that there are many differences between European markets, and exceptions to the general pattern that I will portray. Nonetheless, the main European markets do share some common features.

The first feature is that the main "spot" market takes the form of a voluntary day-ahead auction, which is not run by the system operator but by an independent power exchange. Second, these auctions trade power for delivery at a notional point within a country, or within a sub-national region, rather than at specific nodes on the electrical network. Third, most of these markets trade en-

3. Fourteen states and the District of Columbia had active programs in April 2008, and twenty-eight had never implemented an active restructuring program (source: Energy Information Administration).

4. In fact, California had a market design much closer to those in Europe than the Standard Market Design of the north-eastern US, although the details of wholesale trading rules were not one of the fundamental causes of the crisis there.

ergy, with no special arrangements for remunerating capacity. Fourth, the system operator has to buy reserve and deal with transmission congestion. Fifth, there is usually a balancing market in which generators' outputs (and sometimes customer loads) are adjusted to keep the system in balance – this market is sometimes used to help deal with congestion. Sixth, prices established in the balancing market form the basis of the amounts charged to participants for their imbalances – the difference between their contracted positions (counting self-generation, longer-term contracts, the results of power exchange trading, and trades in the balancing market) and their physical position. A seventh feature of the European markets is that many of them are now cooperating in market-coupling arrangements to ensure the most efficient use of the interconnectors between countries.

Most European countries have a voluntary day-ahead auction, coupled with bilateral trading and self-supply. Bids to buy and sell power are submitted to the market operator each morning, and it will calculate the intersection of supply and demand for each hour of the following day. Some markets allow more complicated bids, in which a generator can ask to sell for a number of consecutive hours or not at all, or specify a minimum income before it is willing to generate. These bids reflect the costs and operating limitations of large thermal power stations, for which each start represents an investment of time and cost. The British Isles contain the two main exceptions to this rule at the time of writing. Great Britain itself does not currently have a day-ahead auction, relying on continuous bilateral trading until Gate Closure when the system operator takes over the task of balancing the system, although the Futures and Options Association (2007), representing a large number of energy traders, is assessing proposals for the organization of a day-ahead auction and a clearing house for bilateral contracts. The All-Island Market covering Northern Ireland and the Republic of Ireland is a mandatory market in which all generators above a *de minimis* size must participate. This market has many similarities to the former Pool in England and Wales, in that generators are allowed multi-part offers, with prices for energy and for starting up their plant, which are rolled into a single System Marginal Price.

All European markets follow either a market-wide or a zonal approach to price setting. In the former, traders simply have to be able to deliver power, or take delivery of it, at some point on the system covered by the market, which is often bounded by national frontiers. Any restrictions on delivery that might be due to transmission constraints are ignored in the day-ahead market. Under the zonal approach, traders must specify the location of their generators and loads. Norway usually has three zones, with borders that are set from time to time by the system operator, reflecting anticipated constraints on the grid. Italy has seven geographical zones, with five running from north to south along the mainland, and the two islands of Sardinia and Corsica. While the market operator starts by calculating a common national price, taking all offers and bids together, if the amount of power to be produced in a zone at this price exceeds the export capacity of the lines connecting that zone to the rest of the country, then the local price is reduced until the power flow is expected to respect transmission limits. Similarly, a higher local price will be set in zones that would have inadequate production at a common national price.

Two European markets, the Iberian market covering Spain and Portugal and the All-Island Market, include capacity payments in their market designs. The original capacity payment, in England and Wales, was set each half-hour on the basis of half-hourly demand and generators' availability over the previous eight days, and was extremely volatile. In Spain, the capacity payment for consumers is set by the government, and then distributed among the generators participating in the day-ahead auction.[5] In the All-Island Market, the total payment is equal to the fixed cost of a peaking plant, less any revenues from ancillary services and from selling energy at more than its variable cost, multiplied by a capacity requirement based on the forecast peak demand (All Island Project, 2007). The annual sum is then allocated to months, and then to half hours on the basis of forecast demand, the forecast loss of load probability, and the ex-post loss of load probability. Half-hourly capacity prices are obtained by dividing the amount of money for that half-hour by the capacity available in it. This produces a compromise between reflecting the actual system conditions (sending the signal that best reflects marginal cost) and the anticipated conditions (which will be less volatile).

Most European markets operate without capacity payments. Implicitly or explicitly, they are based on the idea that generators will be able to recover enough revenue from energy prices, and payments for ancillary services, alone – either in the short-term markets, or in contracts which are ultimately linked to the prices there. For a peaking generator, considered as a separate unit, this requires either high peak prices or significant payments for being available as reserve capacity (which may, as in Great Britain, be determined through market-based mechanisms such as auctions). Most European countries have just a few large generating firms at the moment, however, and these may be willing to consider their portfolios of plant as a whole. In this case, revenues for the portfolio may be adequate, even if some stations would not cover their costs from the wholesale market, given the pattern of prices. In fact, many European utilities are integrated between generation and retailing, and might not mind if the transfer price between generation and retailing made their generation appear unprofitable, provided that their retail revenues exceeded their total costs.[6] It remains to be seen whether this approach could be sustainable in a more fragmented market.

Europe's electricity system could not be operated on the results of the day-ahead markets, and bilateral trading, alone. First, these may violate transmission constraints, even in a zonal market. Even if prices have been adjusted to limit flows between zones to the available capacity, there may be some lines

5. Since only generators participating in this auction could receive capacity payments, it is not surprising that there was practically no bilateral trading in Spain (Crampes and Fabra, 2005) – at least until this rule was changed!

6. If it is easier to enter generation than mass-market retailing, this would act as a barrier to entry. However, it might encourage entrants to sell power to large industrial customers, unless their prices were based on the wholesale price.

within zones that would be overloaded (or potentially overloaded in the event of a contingency) if the auction results were used for the dispatch. Instead, the system operator must counter-trade, reducing generation on the wrong side of the transmission constraint, and increasing it elsewhere. These trades may be made in a balancing market, run by the system operator close to real time. This market takes bids and offers to reduce and increase generation (and to increase or reduce demand) to keep demand and generation in balance. Most markets calculate a single price and use it for all the trades in a given period, although the British Electricity Transmission and Trading Arrangements (BETTA) pay each trader their own offer price. Counter-trading to resolve constraints will involve a net cost – the generator required to increase output will be offering a price which exceeds the market price in its zone (or it would already have been scheduled), while the generator which has to reduce output will be bidding the market price or less to buy back its power. The system operator will also procure reserve capacity, to ensure that it does have enough balancing resources to call on in the event of sudden changes in demand or plant failures.

Prices from the balancing market are typically used to settle differences between the sum of participants' trades and their actual positions. Exactly how this is done varies from country to country. Some countries have a single price for both participants with a surplus of power and those with a deficit, while others charge more for a deficit than they pay for a surplus.[7]

The final issue to consider is what happens at the border between countries. The world's first multi-national electricity market was set up as long ago as 1996, when Sweden joined the Norwegian market to form Nord Pool, subsequently expanded to include Finland and both (electrically separate) halves of Denmark. Nord Pool operates as a zonal market, calculating a System Price which applies whenever none of the borders (between countries, or between zones within Norway) is congested, and separate prices for each side of a congested border. This should automatically ensure that power flows from low-priced to high-priced areas, producing an efficient solution (within the limitations of the zonal model).

Capacity on other European borders has not been allocated in such an efficient manner. In some cases, legacy contracts between incumbents still control access to the interconnector capacity, although the European Commission has been trying to eliminate these contracts as part of its liberalization program (Green, 2008). The European Regulation on Cross-Border Trading called for market-based access to interconnectors, and some countries organized regular auctions to allocate capacity. When system conditions are such that the efficient direction of power flows is predictable, the need to make separate trades in each national market and in the auction for interconnector capacity is not a serious

7. The difference between the price for a deficit and that for a surplus may encourage participants to try to balance their positions in advance. In practice, however, the penalty for a deficit (measured as the difference between the imbalance price and the price in the day-ahead market) can be far higher than that for a surplus, and this asymmetry can produce an incentive to aim for a surplus of power in real-time – something which cannot be achieved by all participants simultaneously.

obstacle. When conditions are changing and the efficient direction for trade is not predictable, however, the need to trade in two separate power markets and acquire capacity on the interconnector may prevent some efficient trades from taking place – particularly when the markets operate on different timescales. The European Commission (2007) has shown that nominations to send power on the cross-channel interconnector between England and France (which cannot be disrupted by transit flows to third countries) frequently go from the market which is more expensive to the one which is cheaper at the time of nomination.

One solution to this problem is market coupling. Traders still bid to separate national markets, but at approximately the same time, and the markets then coordinate their auctions. By comparing the net exports from each market at various prices, the market operators discover whether there is a common price that would allow generation to equal demand without violating any cross-border (or cross-zonal) transmission constraints. If there is, then this is set as the price in each market. If not, then each market (or zone) will have its own price, and power will be exported from low-priced markets, up to the limit of the interconnectors. A congestion rent will be created, since there are net sales of power in the low-priced exporting market, and net purchases in the higher-priced importing area. This rent is paid to the holders of cross-border transmission capacity, either the transmission system operators, or companies that have acquired contracts to use the interconnectors. The Trilateral Market Coupling between France, Belgium and The Netherlands started operation in November 2006. The German-Austrian market EEX is due to couple with this market, and has also started to couple with Nord Pool. At times, this could produce a single day-ahead price for power throughout North-West Europe. This would be a notable achievement, symbolic of the European Commission's desire to have a common market for electricity.

4. MARKET DESIGN IN THE UNITED STATES

In the United States, regulatory policy is a matter for the states, although the Federal Energy Regulatory Commission (FERC) has powers over high-voltage transmission where this involves trade between states. In the last few years of the last century, roughly half of the states chose to adopt electricity markets, although some of them reversed their decision after the California electricity crisis of 2000-1. Three main market designs have been put into practice. California chose to follow a version of the European model, separating the market operator(s) and the system operator, and setting prices for two main zones within the state. The reform collapsed when rising input prices, high demand levels (for part of the period), low imports and the exercise of market power drove wholesale prices well above the still-regulated level of retail rates, bankrupting one of the major retailers (Joskow, 2001). Texas also adopted a zonal market, but run by the state's independent system operator. In the north-east, however, three markets have versions of what I will call the Standard Market Design. This name comes from FERC's (2002) attempt to have markets of this kind adopted throughout the country, an attempt which failed in the face of non-liberalized states' desire to stay with a regulated system.[8] The markets are New England, New York and PJM, originally centered on Pennsylvania, New Jersey and Maryland, but now covering a much wider area. California and Texas are now both changing their markets to adopt versions of the Standard Market Design. There are important differences between the various market rules (for example, Helman et al (2008) use the differences between PJM and New York to illustrate some of the necessary design choices), but these markets have shown "convergence … on important common design elements" (*ibid*, p. 184).

I will start my description of this design by contrasting it with the common features of the European markets. First, the main "spot" market also takes the form of a voluntary day-ahead auction, but one which is run by the system operator. The system operator takes into account all the constraints it faces, ensuring that the market results respect limits on the transmission system. This may mean that prices differ (if there is congestion) at every node on the network, giving a second difference with the European model. These prices are also used to set fees for other participants who move power over the network. Third, some of these markets have adopted capacity markets to deal with what is known as the "missing money" problem, and ensure peaking generators are remunerated. The fourth feature identified in Europe is also a point of difference, in that the system operator's auctions cover reserve and ancillary services as well as energy, and the system operator minimizes the total cost of buying everything that it needs to meet demand, rather than simply minimizing the cost of energy. As in Europe, there is a balancing market operating close to real time (point five), and some markets offer other opportunities to trade between the day-ahead auction and this real-time market. As with the first point, however, there is a difference, in that the real-time markets in the US set nodal prices and cover the full range of services. The sixth point is where there is least difference with Europe, as imbalances are also cashed out at prices established in the real-time market. The US markets have no formal equivalent of market coupling, the seventh feature I identified in Europe, but participants located outside a market area can offer to sell power into it, and a nodal price for this power will be calculated at the border. Table 1 summarizes these differences and similarities.

This market design is based on the work of Fred Schweppe and his coauthors (Bohn et al (1984), Schweppe et al (1988)). They showed how spot prices, varying over space and time, could reflect the marginal cost of generating power and delivering it to any point on the network. This gave the basis for an electricity wholesale market, even though their model of retail pricing, using a variety of adjustments to ensure that the industry recovered the amount of money actually due to it, was firmly within the paradigm of regulation. It is possible to have a competitive wholesale market without opening retailing to competition, and states such as Vermont and West Virginia are in this position.

8. In general, states which remain regulated are those that had below-average electricity prices, and hence saw less to gain from changing their system (White, 1996).

Table 1. Key Features of Wholesale Market Designs

	"European Markets"	**"US Standard Market Design"**
Day-ahead market run by:	Separate company (typically)	System operator
Energy prices are:	National or zonal	Nodal
Capacity paid for via:	Some capacity payments, no system-wide markets	Some capacity markets
Ancillary services:	Separate procurement by system operator	Integrated into spot market
Real-time market:	Run by system operator, setting national or zonal prices	Run by system operator, setting nodal prices
Imbalances cashed out:	Via real-time market	Via real-time market
Inter-market transactions:	Market coupling increasingly common	Outside participants submit bids/offers at borders

The Standard Market Design is built around voluntary auctions in which the Independent System Operator calculates the optimal dispatch, based upon generators' offers and bids from the demand side (Helman et al, 2008). Generators and retailers that have already traded bilaterally must submit these trades to the ISO as part of the dispatch, and have the option of submitting prices at which they are willing to change their plans, if doing so allows the ISO to find a more efficient solution. The ISO's dispatch uses a single optimization to cover its requirements for energy and for reserve and other ancillary services, and also has to respect security constraints on the transmission system. Using a single optimization is more efficient than trying to acquire the different products separately: in its early days, the California ISO was required to buy a fixed quantity of low-quality (long notice) reserves even if high-quality (short notice) reserves were available at a lower price (Wolak *et al.*, 1998). With a single optimization, the ISO can automatically substitute among the various products to meet energy demands and reliability constraints at the lowest possible cost.

The ISOs actually run several markets over different time scales. The main energy market is the day-ahead market, held each day to set prices and quantities, and determine the optimal dispatch for the following day, on the basis of the information then available. Over the following hours, new information will arrive, and there will be opportunities to trade in response to it. Some ISOs run intra-day markets, and all have a real-time market in which generation is adjusted to meet demand. The quantities set in the day-ahead market are financially firm, however. Generators (retailers) are committed to either deliver (accept) the power or to buy (sell) it back at a later trading opportunity. If a generator neither delivers the power nor buys it back, it will face an imbalance and have to pay the price set in the ISO's real-time market. Most of these markets accept bids and offers until close to real time,[9] and use them to dispatch generators to meet the actual load. LMPs are set in the same way as in the day-ahead market, although reflecting the different constraints the ISO faces in real time. In the day-ahead market, the ISO can commit generators with long start-up times, whereas only generators actually running, or able to start quickly, can provide energy or ancillary services in the real-time market.

The ISO therefore needs to ensure in advance that enough generators will be running to meet its demands for energy and for reserves. Most of the ISOs do so by running a reliability unit commitment at the day-ahead stage. This is a back-up to the day-ahead market, recognizing that some of the apparent supply in that market may come from financial traders without the ability to actually deliver power – their strategy is to buy it back before delivery is due. This is one of a number of deviations from the "pure" market results that reflect the physical realities of running an electrical system in a reliable manner. It is quite possible that the units committed at this stage will not earn enough from selling power (or reserve) to cover the cost of starting up. To ensure that they are nevertheless willing to do so, the ISO will compensate them for any costs involved that they cannot recover from market prices. The market prices may also fail to reflect (e.g.) start-up components of some scheduled generators' offers, and extra payments allow the ISO to offer a revenue sufficiency guarantee, paid for by an uplift charge on demand.

The ISO calculates a Locational Marginal Price (LMP) for every point on the network, equal to the marginal cost (according to the generators' offers) of providing power there. One point on the network will be designated as the reference node (or slack bus) and "system lambda", the cost of providing additional energy, is calculated for this point. This gives the reference node its LMP. For all other nodes, the cost of power is equal to the cost at the reference node, plus the cost of transmitting power from the reference node to the node in question. This consists of marginal losses (positive if an increase in demand at the node would lead to an overall increase in flows, and negative if enough flows would decrease) and the marginal cost of constraints. Until 2007, PJM did not include marginal losses when calculating LMPs, so that in hours when no constraints were binding, prices could be the same all across the market. It would only take one binding constraint, however, to make prices differ at every node. The congestion component of each LMP then equals the shadow cost of the constraint (the amount by which the cost of generation could be reduced if the capacity across the constraint could be increased slightly), multiplied by the incremental flow over the constraint that would be caused by an increase in demand at that location. These incremental flows can be summarized by a set of power transfer distribution factors (PTDFs). Since a change in demand at each node in the network will cause a slightly different pattern of flows, each node has its own PTDF and a single constraint will give

9. PJM requires bids and offers to reach it by 18.00 the previous day. Submitting offers in advance means that generators will be less responsive to changing conditions, neither signalling their own circumstances nor raising prices if the system faces unanticipated problems during the following day.

every node a different price. (In the case of two nodes that are close together and a long way from the constraint, the difference may be trivial.) If there is more than one binding constraint, the marginal cost of each of them will be included when the LMPs are calculated.

Generators receive the LMP for their own location. In principle, loads would pay their own LMP, and when power is transmitted from one node to another, the difference between the LMPs acts as a transmission charge – generators sending power across a congested line from a low-priced node to a high-priced one are paying the marginal cost of transmission. To ensure that there is no artificial reason to avoid the ISO markets, even traders who are not participating in them still have to pay transmission charges based on the LMPs. In practice, retailers buying power in the ISO markets often pay the load-weighted average of the LMPs across a number of nodes, sometimes covering the territory of more than one utility. This will tend to reduce the volatility of the prices they pay, and makes it easier to set a uniform tariff for their customers, while potentially weakening the price signals for demand response. It does not change the total amount received by the ISO.

The locational elements of LMPs can be volatile, and Financial Transmission Rights allow the holder to receive the difference between the LMP at one location and at another, or at a hub, the load-weighted average of LMPs across a wide area. The combined revenue stream from selling at the LMP, plus that received from an FTR for that node, is equivalent to selling at the hub price and gives the generator a hedge for transmission costs. As long as the ISO sells no more FTRs than the physical capacity of the transmission system, it will receive revenues from nodal price differences that exceed the payments it has to make through the FTRs, and is also hedged.

In the textbook model of an electricity market, peak prices would be high enough to cover the fixed cost of providing the capacity needed to meet the peak demand (including reserve) (Stoft, 2002). These prices would often be set by the demand side, as consumers with appropriate (hourly) meters show their willingness to reduce consumption as the price rises. In practice, many consumers do not have these meters, or do not receive the price signals. Furthermore, the system operators frequently take actions which have the effect of reducing market prices at times of system stress (Joskow, 2008). They can reduce the amount of reserve that they are holding, albeit at an increased risk of a system collapse. They may reduce the system voltage, which lowers demand. The operators may also be able to take energy "out of the market" from stations with which they have contracts, thus avoiding the need to buy higher-priced power in the day-ahead or real-time market and raise prices there. Finally, most US markets have offer caps as a remedy for system-wide market power, although these caps rarely bind.[10]

10. As well as a blanket cap of $1000/MWh (in most markets), which is rarely observed as a price, there may be tighter caps on units which face little competition because of transmission congestion. New York and PJM have both had to relax these caps because of their impact on generators located in load pockets that were persistently having their offers revised down to the level of their variable costs (Helman et al., 2008, p. 235)

If prices do not rise sufficiently above their variable costs, generators will not be able to recover their full costs from the energy markets alone. The remedy adopted in the north-eastern US is to have a separate market for capacity, operating on a much longer timescale than the day-ahead markets.[11] In fact, the New England capacity market runs three years in advance, so that entrants can effectively compete with incumbents, having time to build a peaking generator if they succeed in the auction. (They can also opt to receive the resulting price for five consecutive years, reducing their revenue risk.) The ISO determines demand, which is price-responsive and location-specific. This reflects the fact that the value of capacity, just like that of energy, depends on where it is sited. The price of capacity can therefore be higher in regions that are transmission-constrained and short of plant. In the earlier market designs, a largely fixed amount of plant (since the markets did not allow time for new build) was competing to meet a fixed demand, sending prices either to zero or to a price cap. The redesigned markets have price-responsive demand curves which are set so that if there is just sufficient capacity to meet the standard administratively-set capacity margin (typically 15% on top of expected peak demand), then the price will just equal the expected cost of new entry, net of energy and ancillary service revenues. With more capacity, the price will be lower, while the demand curve slopes up towards a capped level (1.5 times the cost of new entry) if less capacity is offered.

The market designs take account of the fact that peaking generators should be able to make a surplus from peak energy and ancillary services prices that exceed their fuel costs. In PJM, the net cost of new entry is calculated (and fixed in cash terms) after subtracting an estimate of this surplus, based on the market prices of the previous three years. In New England's capacity market, however, the out-turn energy and ancillary services prices (net of fuel costs) are used to reduce the eventual payments to a successful seller. This makes the payments from the capacity market more variable, but this should be exactly offset by the rents received in the energy market, providing a more stable income stream overall. It also provides a hedge for consumers, and reduces the incentives for generators to exercise market power at peak times, since changes in peak prices are offset by changes in capacity payments (Joskow, 2008).

5. COMPARING MARKET DESIGNS

How do our two families of market designs compare to each other, considered against the six criteria of section 2? Both designs clearly work, in the sense of allowing the continued operation of the electricity system. The question is whether one design works better.

Our first criterion is whether the market design leads to efficient short-run operation. In making any comparisons, we must take account of market-specific

11. Texas and California have not adopted capacity markets, and it is not universally agreed that they are needed to ensure generators can recover their costs (Wolak et al., 2007).

features that can affect performance. There are no examples of a market moving directly from a "European" design to the Standard Market Design (or vice versa) that would allow us to compare their performance in similar operating conditions, although California has effectively made this journey in stages. Sioshansi et al. (2008) accordingly have to use a simulation to compare two market designs in the setting of New England. They compare a dispatch following the Standard Market Design with one where generators individually choose outputs to maximize their profits, given prices announced by an auctioneer (equivalent to the European model), and find that the latter gives production costs that are 4% higher than the former. The Standard Market Design coordinates the dispatch of generators to minimize costs in a way that simple price signals are unable to replicate.

When the centralized coordination of the Pool in England and Wales was replaced by the much less centralized NETA (New Electricity Trading Arrangements), there were concerns that generators were keeping more plant partly-loaded, with a reduction in operating efficiency (ILEX, 2002). These concerns can be quantified (albeit imperfectly) with UK-wide quarterly data on the major power producers' fuel inputs and outputs. The last year of the Pool (to the end of March 2001) saw similar fuel prices and electricity outputs to the first year of NETA (from April 2002), and David has pointed out that the generators had little market power by this time (Newbery and McDaniel, 2003). Using the system-wide (quarterly) ratios of fuel input to electricity output from 2000-1, we obtain a counter-factual fuel requirement for the 2001-2 outputs that is 0.8% lower than the actual inputs. This equates to an additional fuel cost of £26 million, on a base of £3.5 billion.[12] This combines an indication of the cost of separating the markets for energy and for ancillary services with the impact of the initial rules of NETA (later amended), which gave generators particularly strong incentives to avoid negative imbalances and hence keep plant part-loaded.

The expansion of the PJM market to the west in 2004 combines the impact of moving from bilateral trading to the Standard Market Design, and expanding the area which an ISO is responsible for. Exports from the American Electric Power region to the original PJM region rose from an average of 1,550 MW to 2,300 MW after it was integrated into PJM, while exports from the adjoining area of First Energy, which was not integrated, hardly changed (EASI, 2005). Before the integration, these exports were often limited because they would have breached the flow limits on medium-voltage lines outside the PJM (original) area.[13] After the integration, it became apparent that it would be economic to relax these constraints with a relatively small increase in generation at a station (or stations) close to them on the "downstream" side, given the prices offered for doing so. Before these stations joined PJM, there was no clear way for them to signal their willingness to increase output, nor any responsibility for the system opera-

12. The thermal efficiency of coal-fired generation fell by 1.3%, but that of gas-fired generation rose slightly, by 0.25%, taking the weighted average of the quarterly figures. The fuel costs are evaluated at the average prices paid by the major power producers.

13. I am indebted to Matthew White for this information.

tors to take action to make higher exports possible. Mansur and White (2008) estimate that the higher exports raised the annual gains from trade between the regions from $135 million to between $293 and £316 million. The price spreads between hubs in the traditional PJM and in the new area fell by between 35% and 49% of their pre-integration value for peak power, and by between 37% and 81% for off-peak power.

The gains in PJM come from participants trading more efficiently in a centralized market than they could via bilateral contracts, and from the increased transmission capacity that could be made available once the ISO had a mechanism to do so. In the European markets, the growth of market coupling should ensure the efficient use of given cross-border transmission capacities. What it may not do is maximize those capacities. Glachant and Pignon (2005) show that within the zonal Nordic market, the Swedish system operator had an incentive to declare that congestion was happening at the national border, rather than within Sweden. In the former case, it would be resolved by setting different prices for the different countries within Nord Pool, creating a surplus for the transmission companies. In the latter case, the Swedish transmission operator had to incur the cost of resolving the constraint by counter-trading within the country. Market coupling could create the same incentives across Europe, unnecessarily limiting the gains from trade, compared to a nodal system.

Our second criterion is that the market design should signal the need for investment in generation and in demand reduction. The Standard Market Design has a clear mechanism for doing this, with the capacity auctions that have been implemented in the north-eastern markets. The European design relies on energy prices sending a signal that is sufficiently strong, and early, to persuade generators to add capacity when it is needed. The debate on whether a capacity market (or other mechanism) is needed in principle is unresolved. Joskow (2008) argues that we need one in practice, because of the lack of demand response to set prices, and actions taken by system operators that depress the market price of energy. Roques (2008, p.182) suggests that it would be better to address these problems "at source, not by overlaying offsetting regulations." Green (2005) compares the relative marginal costs and benefits of generating capacity, which imply that in an uncertain world, a quantity instrument could be expected to produce better results, on average, than relying on prices. Many European countries had adequate generating capacity when they opened their markets a few years ago, and so have not yet needed to test their ability to create investment.[14] With an industry dominated by oligopolists that cannot afford the political embarrassment of a shortage of capacity, we may well get investment when we need it, but it could be despite the European wholesale market design, rather than because of it.

The third criterion was that the market design should promote efficient locational choices for investments. The Standard Market Design does this ex-

14. In Sweden, the system operator had to pay for additional generation capacity outside the main market when reserve margins grew low around 2000 (Damsgaard and Green, 2006).

plicitly – generators will earn more in areas where they are badly needed, from both the energy and the capacity market, and will earn less in areas with a surplus of power and export constraints. In European markets, generators inside an import constraint can earn more, through counter-trading, than those in an area with surplus power, and so the market design does help to promote investments in these areas. At the same time, however, a generator inside an export-constrained region will be compensated for any power it is unable to sell, whereas it would be paid less under the Standard Market Design. The European design for wholesale markets therefore gives no signal that some areas should be avoided. In some countries, however, the tariff for using the transmission system might send such a signal, allowing us to reach an efficient outcome nonetheless.

The fourth criterion was that the market design should adequately compensate the owners of generation assets. The verdict here has to be linked to that on the second criterion. If the Standard Market Design needs a capacity market to ensure that the generators have sufficient revenues to invest in new generation when it is required, then the European markets are in danger of providing insufficient revenues – unless market power, or revenues from elsewhere in the vertical chain, make up the difference.

The European markets probably do better on the fifth criterion, however, that of simplicity and stability. Their main energy auctions would certainly appear simpler to outsiders than the geographically differentiated US markets. For generators, however, faced with trading energy and ancillary services separately, the integrated US market may provide a more straightforward solution. Stability comes in part from the ability to hedge prices through forward trading, and this will be easier with a single nationwide price than with a large number of location-based prices. There is a solution to this within the Standard Market Design, however, as it is possible to trade electricity within each US market at one or more hubs. The hub price is the load-weighted average of a number of nodal prices, giving results that reflect general conditions in the market but are not made volatile by transmission effects at individual nodes. The hub therefore provides a good basis for forward trading. In 2007, the volume of financial contracts traded against the PJM market was 3.4 times the final sales of power in the area, and while this was below the level of Nord Pool (5.6 times), it was hardly illiquid.[15] A generator that needs to hedge the difference between its nodal price and that at the hub can do so by acquiring financial transmission rights.

Stability also refers to the need for rules that do not change too often. In the US, the two markets that did not start with a version of the Standard Market Design (California, and Texas) have made the major change of adopting it. In the north-eastern markets, the capacity markets have seen significant changes with the adoption of locational prices and price-sensitive demands. The European power

15. The figure for PJM refers to financial contracts traded by the ICE and Nymex (source: ICE), while the figure for Nord Pool includes financial contracts traded on its futures market, and those contracts submitted to the company for clearing (source: Nord Pool).

exchanges have seen fewer changes, although the introduction of market coupling could have significant impacts on prices. The way in which many system operators buy reserve and balancing has changed significantly, with pressure from the European Commission to move to market-based, rather than tariff-based, balancing.

The final criterion was that the market should be politically implementable. The European design scores better here – versions have been adopted in much of the European Union, as Member States comply with the EU's policy on liberalizing the electricity industry. This may be related to its "incremental" nature – once electricity companies have started to trade power bilaterally within a country, it is relatively easy to add a power exchange, and for a system operator that was using tariffs to balance the system to start a real-time energy market. In other words, the European market design does not need much support from politicians to evolve within a country's electricity industry, once a policy of liberalization has been chosen. While the Standard Market Design has become the dominant market design in those US States that have adopted electricity markets, the Federal Energy Regulatory Commission has failed in its attempt to impose it throughout the country. This gives us an unfortunate result – the European design appears to be less efficient, but is more likely to receive enough political support to be implemented. Table 2 sums up the relative advantage of our two, slightly stylized, families of market designs.

Table 2. Electricity Market Design: The Balance of Advantage

Criterion	**"Standard Market Design"**	**European Markets**
Efficient Operation	Integration of energy and ancillary services, and transparent pricing for transmission, raises efficiency	
Investment signals	Capacity markets provide revenue adequacy	Some commentators doubt that these extra revenues are really needed
Locational signals	Locational prices reward investments in "helpful" locations and penalize those in "bad" ones	Transmission tariffs may provide alternative signals
Remuneration of investments	Capacity markets provide revenue adequacy	Some commentators doubt that these extra revenues are really needed
Simplicity and stability		Slightly more hedging takes place; fewer major rule changes (in most markets)
Political acceptability		Have been adopted throughout the EU, with high-level support for liberalization; market design a low-key issue

Should electricity markets in Europe attempt to move towards the more efficient Standard Market Design? We should not forget the transactions costs involved. There are the one-off costs of adopting a new set of rules and re-writing computer systems, and the ongoing costs of running the market. With more prices to keep track of, the latter may rise (although the system operator may need to do no more calculations than are already required to keep the system stable), although if an integrated market replaces separate trading of energy and ancillary services, this should lead to some cost savings. These ongoing costs are therefore unlikely to offset the benefits, but a major change to the market rules is expensive – NETA cost £700 million (Newbery, 2005). However, it is more likely to be worth making the change if likely developments (such as a move towards low carbon electricity generation) in the electricity industry increase the relative advantage of the Standard Market Design. We thus turn to these.

6. A LOW-CARBON ELECTRICITY SYSTEM

The European Union and some American States have adopted aggressive targets for reducing carbon emissions. Energy efficiency, heat from biomass and bio-fuels in transport can make significant contributions, but it seems inevitable that much of the reduction will have to come from the electricity industry. Fuel switching from coal to gas, building more efficient power stations, nuclear power and carbon capture and storage could all contribute to this reduction without changing the industry's current model of a largely centralized system. The European Union, however, has also set a target for 2020 of producing 20% of its energy from renewable sources. This could involve renewable electricity generation of 1250 TWh a year (Pöyry, 2008), nearly three times the current level of renewable generation in the EU (which was 440 TWh in 2005). Wind output might rise from around 60 TWh a year to nearly 350 TWh a year. Electricity generation from biomass, presently at a similar level to wind generation, might rise to nearly 450 TWh a year. Much of this could be in small plants, providing combined heat and power.

This increase in wind and other distributed generation poses significant challenges for the electricity system. Wind power is not controllable, except to the extent that electricity can be spilled by rotating a turbine's blades and reducing its efficiency if the system is unable to accept its full power output. The level of output depends on the wind, and this is highly variable. In Western Denmark, the total wind output during 2007 was equal to 26.3% of electricity consumption, but for 10 per cent of the hours in the year, it was less than 2.7% of the consumption in that hour, and for 10 per cent, it was more than 61.8%. In 50 hours (out of 8760), wind output exceeded the local demand (source: Energinet). There can also be significant fluctuations from hour to hour – in Western Denmark, wind output changed by at least ten percent of capacity in 5 per cent of the hours in 2007. The Danish system operator is able to keep its system in balance by exporting power when it has a surplus, but this depends on its neighbors' ability to absorb the fluctuations. As the proportion of intermittent plant in neighboring systems increases, each country will have to manage a greater share of its own fluctuations.

This means that the load on thermal plants becomes far more variable. At the operational level, more plants must be kept in reserve to cope with the risk of a sudden loss of power when wind speeds fall (or exceed the point at which the turbine must shut down for its own safety). This is exacerbated if the trading system requires generators to submit their plans a long time in advance. Bathurst et al. (2002) show that wind generators unable to predict their output accurately were exposed to significant imbalance penalties in the early days of NETA. Holttinen (2005) shows that a generator in the Nordic market, normally trading between 12 and 36 hours before delivery, would increase its net income by 4% if it could trade between 6 and 12 hours in advance, and by 8% (in total) if it could trade just one hour in advance. Müsgens and Neuhoff (2006) show that trading closer to real time would also reduce the cost of thermal power, since fewer stations would be started up, only to find that previously unexpected wind power substitutes for their output.

This will tend to make prices more volatile. Another area that already has a significant amount of wind capacity is West Texas – 5.5 GW out of a total zonal capacity of 10.5 GW. ERCOT's balancing market energy prices in that zone ranged from –$999.01/MWh to $1533.51/MWh during 2007. They were negative in 175 of the 15-minute periods for which the market sets prices, and fell below –$50/MWh in 50 of these. There were 184 occasions when the price changed (up or down) by more than $500/MWh between adjacent periods, 369 times when it changed by more than $200/MWh, and 492 when it changed by more than $100/MWh.[16] Balancing market prices will always be more volatile than prices set further in advance, and the volumes exposed to them will be small. Nonetheless, as the proportion of wind capacity increases, the volatility of day-ahead prices will also increase. This will increase the need for hedging contracts, maybe of novel designs (Newbery, 2008).

Moving to investment, more capacity is needed in total, to compensate for the risk that the wind stations are not producing at the times of the system peak. These costs are estimated in a number of studies, summarized by Gross *et al.* (2006). Their conclusion for Great Britain was that the cost of intermittency would be in the range of £5-8/MWh of intermittent output, at least while wind power supplied less than 20% of annual demand. In the market context, however, the volatility means that prices, and particularly peak prices, will become more volatile. This could well reduce the attractiveness of keeping peaking plant open at the very time when a greater reserve margin becomes necessary.

The other aspect of variable wind outputs is geographical. For obvious reasons, wind stations are mostly built in places with high average wind speeds, and these are not spread evenly over Europe. In Great Britain, most wind farms are in the north and west, remote from the load centers, with a few more helpfully located in East Anglia. Plans for future stations show an even stronger bias towards locations in Scotland. If these stations are built, then the transmission

16. I would like to thank an anonymous referee for pointing me towards this data.

system in its present state will not be able to accommodate all of their output on a windy day. One approach (and that still formally enshrined in the industry's rules) is not to connect any wind farms until the transmission system *can* accommodate all their output, but this has the weakness that constraints that might only bind for a few hours a year are then causing either a delay in generation investment or excessive investment in transmission.[17] Strbac *et al* (2007) estimate that with 10 GW of wind capacity in Scotland, the economic level of transmission capacity on the border with England would be 5.4 GW, compared to the 7.6 GW that the present planning standards require. This would, however, entail a significant increase in transmission constraints – which has been accepted in Greece as the most efficient way to allow more wind power (Kabouris and Vournas, 2004).

Distributed CHP plants, unlike wind, are controllable, but achieve their maximum levels of technical efficiency if they are scheduled to run when there is a demand for heat, which need not coincide with the needs of the grid. Hawkes and Leach (2007) find that residential micro-CHP plants, with an electrical output of only 1 kWe, may find it most economic to follow the greater of the home's heat or electricity demands, although Peacock and Newborough (2007) show that these plants could be operated in a way that reduces loads on their local distribution system. Larger plants with heat storage can also synchronize their electricity generation with the grid's needs, but still need an incentive to do so. Gordijn and Akkermans (2007) discuss several business models for distributed generation (DG), highlighting the need for the model to be attractive both to the generator-customer and to the electricity industry. They find that the business case is significantly helped by regulatory stability and by the ability "to sell and trade directly on a power exchange market" (p. 1188). This contrasts with what Strbac (2007) calls the "'fit and forget' approach" to connecting distributed generation, in which "no real attempt has been made to integrate DG in system operation" (p. 1143). Strbac regards it as "an economic imperative that DG participates in the provision of ancillary services needed for secure and reliable operation of the power system" (p. 1147). In Denmark, CHP plants that previously received a fixed tariff for their power are now exposed to the spot market, and some are installing immersion heaters to allow them to consume surplus power on windy days when they still need to provide hot water (Franken, 2006).

It is unlikely to be realistic for many distributed generators to participate directly in the market – many of the owners are small companies or individuals, without the capacity to actively engage in the electricity market. Instead, they should trade through intermediaries, generally the companies that would also sell

17. In Great Britain, the difficulty of getting planning permission means that the consequence has generally been a delay in investment, but in Texas, generators were allowed to build wind farms in an area (McCamey) with good wind resources, but behind a local constraint that could only be relaxed with an investment of $150 million (Texas PUC, 2003).

them power at times when their generation is inadequate for their own needs.[18] This means that when we think about the requirement for the market to be as simple as possible, we can do so in relation to the capacity of a specialized participant, and not a typical individual. The market rules should allow many small sources to be aggregated (since system operators would be incapable of dealing with each individually). When the rules create an anomaly, such as (for example) paying very different prices for generators small enough to be treated as negative demand, and those large enough to be treated as generation, there must be clear boundaries to show how any given generator will be treated. These can minimize attempts at regulatory arbitrage, even though they may then influence investment decisions designed to stay on the more favorable side of the boundary.

While these developments will make the electricity system more challenging to operate, other changes will help. Better communication technologies can allow for more demand-side response, with the consumer's active involvement (inspired by smart metering) or without (e.g. via frequency-responsive devices for heating or cooling, perhaps linked to more thermal storage). Electric vehicles could be charged up when electricity is (relatively) abundant and might even supply power back to the grid when it is short (Kempton and Letendre, 1997). Improvements to network and control technology may increase transmission capacity. The question is whether one market design will be better suited to adapt to, and shape, these developments.

7. IMPLICATIONS FOR MARKET DESIGN

What do the combination of high levels of wind energy and distributed CHP imply for the optimal wholesale trading system? Our first criterion, operational efficiency, will become more challenging to achieve as the loads on thermal power stations become more variable. The amount of reserve needed will increase – in Great Britain, National Grid's requirements may double or triple from their current 3.5 GW (Bennett, 2008). Wind generators, and others, will wish to adjust their trading position close to real time, to take account of the latest weather forecasts. Since many wind stations will be remote from demand, power flows over the grid will also become more variable, and (unless large amounts of additional network capacity are added) there will be more transmission constraints. These factors will increase the importance of scheduling plants as efficiently as possible, and maximizing the available transmission capacity. The Standard Market Design is likely to increase its superiority over the European market design in this context. The purchase of energy and reserves is integrated, the location of transmission constraints is transparent, and the process is repeated close to real time, allowing for adjustments in response to new information. Small distributed gen-

18. In Great Britain, a company called Flexitricity (http://www.flexitricity.com) does not trade or supply electricity, but acts as an intermediary to signal the opportunity for consumer-generators to cut load or increase output, sharing the benefits with the consumer and their retailer.

erators may be able to provide useful ancillary services, and a market design that allows them (or their agents) to sell both energy and ancillary services through the same mechanism is more likely to encourage this than separate markets.

The second criterion, that there be adequate incentives to invest in generation (and demand response) capacity, will also become more challenging. More capacity will be needed in total, but much of it will only generate (or be used for reserve) occasionally, when high demands and low winds coincide. The European market design relies on high energy prices at these times to provide sufficient incentive for investment. The risks – for generators and consumers – involved in this approach will rise as prices become more volatile. Hedging via long-term contracts would reduce these risks, and many European electricity markets are associated with liquid contract markets. The Standard Market Design has a quantity target, with the risk that regulators will set this at the wrong level and that consumers will end up paying for a different standard of security than they would have freely chosen – typically a higher one. The use of a downwards-sloping demand curve for capacity does tend to reduce the cost of such errors, however, since both the price and the level of capacity can adjust. Since the debate over the best way to remunerate generation capacity is unresolved, the fact that the issue is becoming more important does not affect the balance between our two families of market designs. The incentives for demand-side response, however, may be stronger in a price-based system where it allows the customer to internalize the benefits of avoiding short periods of high energy prices than in one based on a capacity market which is only accessible to large (or aggregated) customers.

Our third criterion was that the wholesale market should promote efficient location decisions. Again, this will become more challenging, with the need to accommodate renewable generators when the best resources are far from demand. In most cases, the quality of the resource will offset the higher transmission costs, making the remote location the most efficient one to choose. This will increase the flows on the transmission system, and the costs of transmitting power from other generators in the same area. This makes it more important that conventional generators with a good choice of sites are given strong incentives to avoid high-cost locations. The Standard Market Design, with its nodal prices that reflect the cost of losses and congestion, provides such an incentive. Most European markets do not, although, like the Standard Market Design, they ensure that generators in particularly favorable locations can earn more.

Our fourth criterion is that generators should be able to cover their costs from wholesale market revenues. As in the discussion above, it is linked to the second criterion, and the investment signals that the different market designs send. With volatile wholesale prices, the issue becomes more challenging, but the question of the superior design remains open.

The fifth criterion considers simplicity, stability and transparency. With (potentially) many more small generators, giving them access to simple ways of trading becomes more important. This does not necessarily mean that the wholesale market must be simple, however, if intermediaries can buy power from the

Table 3. Electricity Market Design: The Changing Balance of Advantage?

Criterion	"Standard Market Design"	European Markets
Efficient Operation	The need to integrate energy and ancillary services, and send accurate signals on transmission costs, will be greater	
Investment signals	More capacity will be needed, and it may be more risky to rely on volatile energy prices	The debate on generators' need for a capacity market is not settled. Demand response may receive stronger signals
Locational signals	Greater need to signal the gains from offsetting renewable capacity in remote areas with other plant closer to loads	
Remuneration of investments	More capacity will be needed, and it may be more risky to rely on volatile energy prices	The debate on generators' need for a capacity market is not settled
Simplicity and stability	No change in relative ranking	No change in relative ranking
Political acceptability		Need to overcome opposition to locational prices from renewable generators in remote places.

small generator and sell it into the wholesale market. Spot market prices will be more volatile, low when the wind is strong, and high when there is little wind. This will increase the need for hedging. Since both market designs allow this, as discussed above, the impact of low-carbon energy on this criterion does not change their relative ranking.

The final criterion concerns political acceptability. The key issue here is that if European countries were to move towards the Standard Market Design, there would be winners and losers. Generators located a long way from demand would typically lose, as might customers located a long way from generation. The customers may not know that they are about to lose from a change in the electricity wholesale market; the generators will. A change that reduces the revenues of renewable generators can be portrayed as preventing action against climate change: the First Minister of Scotland, Alex Salmond, has described transmission charges there as "intolerable" for this reason (Scottish Government, 2008). This does not mean that progress is impossible, however. David Newbery has pointed out in many contexts that if a change leads to a more efficient outcome, then it is possible to design a mechanism that will allow the change to take place, compensate the losers, and still leave a surplus. In this case, Scottish (and other) generators could be given Financial Transmission Rights that would "lock in" their expected revenues from the current system, but give them the correct incentives to respond to nodal prices if these were adopted (Newbery and Neuhoff, 2008). Future investment decisions, though, could be made on the basis of the correct price signals. Designing the detailed transition mechanisms would be a challenging task, but one which gives hope that the Standard Market Design could be implemented in Europe.

As we move towards a low-carbon electricity system, the balance of advantage between the two families of market designs will therefore change, in ways that Table 3 summarizes.

8. CONCLUSIONS

This paper compares two families of electricity wholesale market designs against criteria intended to promote economic efficiency and political acceptability. In making that comparison, however, it is important to remember how much the designs have in common. In almost every case, participation in any centralized day-ahead market is voluntary, and most power is traded bilaterally through long-lasting contracts, or kept within an integrated company. In real time, generators respond to the system operator's instructions and are paid for doing so, receiving their own offered price or that of a more expensive unit that was also needed. The main differences are in whether the main energy market is run by the system operator that also procures ancillary services, in the treatment of transmission effects (whether the markets are national/zonal or nodal), and in the presence or absence of a capacity market.

In principle, there are three choices here, and many of the eight combinations that result are feasible. For example, the Pool in England and Wales had a system operator that procured both energy and ancillary services, but in a non-spatial market, and the "Texas nodal" re-design does not include a capacity market. However, nodal pricing would not be possible unless the energy market was run by the transmission system operator. Even then, it would be feasible for the system operator to separate the purchase of energy and ancillary services, but there would be little reason for doing so. The aim should be to make the choices likely to give the best results according to the criteria of section 2, now and in a low-carbon future.

Our first choice is thus whether the energy markets should also be linked to ancillary services, and run by the system operators. There is clear evidence that this would improve operating efficiency, and that the gains will increase as we move towards a low-carbon energy system. The existing electricity market operators would have to move towards trading longer-term contracts, as it is unlikely that two nearly simultaneous day-ahead auctions in the same country could successfully co-exist. Resistance from these market operators is likely to be one obstacle to this change; another comes from the vertical integration between transmission and generation in many countries. An integrated market operator would have strong incentives to favor its own generation, and would not be desirable. However, the European Commission's Third Energy Package proposes that each Member State should create an independent system operator (some already have independent transmission companies), and two large energy companies in

Germany have agreed to divest transmission assets in order to settle anti-trust cases.[19] With full separation between generators and system operators, European countries could allow the latter to run efficient markets for energy and ancillary services. The case for doing so now is strong; with rising demands for reserve and other ancillary services in a low-carbon electricity system, it will get stronger.

The second question is whether the energy market should be zonal or nodal. Again, short-run operating efficiency points squarely towards the choice of a nodal market. Nodal prices show where the network is congested and provide natural incentives for generators (and consumers) to take actions to resolve the congestion. Increasing amounts of renewable generation in remote locations will lead to greater power flows and more congestion, raising the importance of this issue.

The disadvantage of nodal prices is that they will create winners and losers, and individual nodal prices will be much more volatile than those set for a whole zone. Financial Transmission Rights could resolve both of these problems. Transitional contracts can compensate the potential losers, while the system operator can reduce its own risks by selling FTRs that replace a volatile stream of nodal price differences with a fixed payment. As long as both kinds of FTR are for fixed volumes of electricity, however, they preserve the incentive to respond to the nodal prices.

The third question is whether a separate capacity market is needed. The evidence here is unclear. Some countries and states have capacity markets; others do not. In principle, an energy-only market may be able to provide sufficient revenues to compensate generators for the capacity needed to give consumers the standard of security that they would (collectively) like. In practice, market imperfections, such as a lack of demand-side response that feeds through to the wholesale market, may prevent prices from rising to the level needed to ensure this. Would it ever be possible to remove these imperfections, or should we accept the need for an offsetting intervention? In this case, what design will minimize the distortions it brings to the energy market, and how can we reduce the risk that consumers will be asked to pay for a gold-plated security standard?

The move to a system with a high proportion of intermittent generation will make paying for rarely-used capacity both more important and more challenging. In principle, the energy-only market could still provide the correct signals and revenues: the pattern of prices might change, but all of the generators in the efficient capacity mix would cover their costs. In practice, the shortfall arising from market imperfections might be greater. My own belief, given the large volumes of investment needed in many European countries,[20] is that a capacity market should be part of the solution. The costs of under-investment rise more rapidly

19. E.ON offered to sell its electricity transmission network in February 2008, and RWE offered to divest its gas transmission in May. Vattenfall Europe, which did not face an anti-trust action, is also planning to sell the electricity transmission network that it owns in Germany.

20. The International Energy Agency (2006) predicts that the EU will need to spend $925 billion (at 2005 prices) on building 862 GW of capacity between 2005 and 2030.

than those of over-capacity, and it is thus better to risk a surplus than a shortage. Nonetheless, if the market follows the New England design, refunding the actual peak revenues from the energy and ancillary services markets to the holders of capacity certificates, this should minimize the additional cost to consumers.

This paper has argued that Europe could gain if countries were to move from their current electricity markets towards versions (not necessarily identical) of the Standard Market Design used in the US. I would not wish to under-estimate the political difficulties of doing so, or the transactions costs of changing the systems. Those costs will be substantial, and could offset the gains from doing so, if they were small. The available evidence suggests that the gains, in efficient operation and investment decisions, are large, and likely to increase as the proportion of intermittent generation rises. Operating a low-carbon electricity system will be a major challenge, and we should use the best tools available to do so.

ACKNOWLEDGMENTS

This research is supported by the Engineering and Physical Sciences Research Council and our industrial partners through the Supergen FlexNET consortium, grant EP/E04011X/1. I would like to thank Michael Pollitt, Matthew White, two anonymous referees and participants at the Norel conference on electricity markets, Saltsjöbaden, June 2008, for helpful comments. The views expressed are mine alone.

REFERENCES

All-Island Project (2007). *Annual Capacity Payment Sum: Final Value For 2008*, [online], available from http://www.allislandproject.org/en/capacity-payments-consultation. aspx?article=4c42e409-1082-4b9c-b9f3-b3e0ac03f564 .

Baldick, R. and H. Niu (2005). "Lessons Learned: The Texas Experience." In Griffin, J.M. and S.L. Puller, eds., *Electricity Deregulation: Choices and Challenges*, Chicago, The University of Chicago Press, pp. 182-224.

Bathurst, G.N., J. Weatherhill and G. Strbac (2002). "Trading Wind Generation in Short Term Energy Markets" *IEEE Transactions on Power Systems*, 17(3): 782-789.

Bennett, C. (2008). "Evidence given to the House of Lords Economic Affairs Committee inquiry on the Economics of Renewable Energy", uncorrected transcript, accessed at http://www.publications.parliament.uk/pa/ld/lduncorr/econ240608ev6.pdf.

Blumstein, C., L. Friedman and R.J. Green (2002). "The History of Electricity Restructuring in California" *Journal of Industry Competition and Trade*, 2(1-2): 9-38.

Bohn, R.E, M.C. Caramanis, and F.C. Schweppe, (1984). "Optimal Pricing in Electrical Networks over Space and Time", *The Rand Journal of Economics,* 15(3): 360-76.

Brunekreeft, G., K. Neuhoff and D.M. Newbery (2005). "Electricity transmission: An overview of the current debate" *Utilities Policy* 13(2): 73-93.

Crampes, C. and N. Fabra (2005). "The Spanish Electricity Industry: Plus ça change …", *The Energy Journal*, 26(Special Issue): 127-153.

Damsgaard, N. and R.J. Green (2005). *Regulatory Reform in the Swedish Electricity Industry – Good or Bad?* SNS Occasional Paper No 95, Stockholm, SNS Förlag.

Energy Security Analysis, Inc (2005). "Impacts of the PJM RTO Expansion: A report prepared for PJM", Wakefield, Ma., ESAI.

European Commission (2007). *DG Competition Report on Energy Sector Inquiry, 10 January 2007*, Brussels, European Commission.

Fabra, N., von der Fehr, N-H M., and D. Harbord (2006). "Designing Electricity Auctions" *RAND Journal of Economics* 37(1): 23-46.

Federal Energy Regulatory Commission (2002). *Notice of Proposed Rulemaking, Docket No. RM01-12-000, issued July 31, 2002* (the "Standard Market Design"), Washington DC, Federal Energy Regulatory Commission.

Franken, N. (2006). "The Danish Solution," *New Energy* 06(6): 22-25.

Futures and Options Association (2007). *Request for Proposals for the Delivery of Clearing and Market Services for the UK Wholesale Power Market*, 13 June. London, Futures and Options Association, accessed at http://www.foa.co.uk/power/mdp_rfp2007.pdf.

Glachant, J.-M., and V. Pignon, (2005). "Nordic Congestion's Arrangement as a Model for Europe? Physical Constraints vs. Economic Incentives", *Utilities Policy* 13(2): 153-162.

Gordijn, J, and H. Akkermans (2007). "Business Models for Distributed Generation in a Liberalized Market Environment", *Electric Power Systems Research* 77: 1178-1188.

Green, R.J. (2005). "Markets and Electricity" *Oxford Review of Economic Policy* 21(1): 67-87.

Green, R.J. (2008). "EU Regulation and Competition Policy among the Energy Utilities" Department of Economics Working Paper 08-01, University of Birmingham.

Gross, R., P. Heptonstall, D. Anderson, T.C. Green, M. Leach and J. Skea (2006). *The Costs and Impacts of Intermittency: An assessment of the evidence on the costs and impacts of intermittent generation on the British electricity network*, London, Imperial College.

Hawkes, A. and M. Leach (2007). "Cost-effective Operating Strategy for Residential Micro-combined Heat and Power", *Energy*, 32(5): 711-723.

Helman, U, B.F. Hobbs and R.P. O'Neill (2008). "The Design of US Wholesale and Ancillary Service Auction Markets: Theory and Practice." In F.P. Sioshansi, ed. *Competitive Electricity Markets: Design, Implementation, Performance*, Oxford, Elsevier.

Holttinen, H. (2005). "Optimal Electricity Market for Wind Power", *Energy Policy*, 33(16): 2052-2063.

International Energy Agency (2006) *World Energy Outlook*, Paris, International Energy Agency.

Joskow, P.L. (2001). "California's Electricity Crisis", *Oxford Review of Economic Policy*, 17(3): 365-388.

Joskow, P.L. (2008). "Capacity payments in imperfect electricity markets: Need and design" *Utilities Policy*, 16(3): 159-170.

Kabouris, J. and C.D. Vournas (2004). "Application of Interruptible Contracts to Increase Wind-Power Penetration in Congested Areas", *IEEE Transactions on Power Systems*, 19(3): 1642-1649.

Kempton, W. and S.E. Letendre (1997). "Electric Vehicles as a New Power Source for Electric Utilities" *Transportation Research Part D: Transport and Environment,* 2(3):157-175.

Mansur, E. and M. White, (2008). "Market Organization and Efficiency in Electricity Markets" mimeo, Department of Economics, Yale University.

Müsgens, F., and K. Neuhoff (2006). *Modelling Dynamic Constraints in Electricity Markets and the Costs of Uncertain Wind Output,* Electricity Policy Research Group working paper EPRG 0514, University of Cambridge.

Newbery, D.M. (1998). "The Regulator's Review of the English Electricity Pool" *Utilities Policy*, 7(3): 129-142.

Newbery, D.M. (2005). "Electricity liberalisation in Britain: The quest for a satisfactory wholesale market design" *The Energy Journal*, 26(Special Issue): 43-70.

Newbery, D.M. (2008). "Reduce the risk of nuclear investment" *Financial Times*, 9 January.

Newbery, D.M. and T.M. McDaniel (2003). "Auctions and Trading in Energy Markets – An Economic Analysis" In P. Vass, ed. *Regulatory Review 2002/2003* Centre for the Study of Regulated Industries, University of Bath, UK, pp. 195-234.

Newbery, D.M. and K. Neuhoff (2008). "Market design for large shares of renewables: time and space" presented at the EPRG Spring Research Seminar, 16 May 2008.

Peacock, A.D. and M. Newborough (2007). "Controlling micro-CHP Systems to Modulate Electrical Load Profiles" *Energy* 32(7): 1093-1103.

Pöyry (2008). *Compliance Costs for Meeting the 20% Renewable Energy Target in 2020: A report to the Department for Business, Enterprise and Regulatory Reform* Oxford, Pöyry Energy Consulting.

Roques, F.A. (2008). "Market design for generation adequacy: Healing causes rather than symptoms" *Utilities Policy*, 16(3):171-183.

Schweppe, F.C, M.C. Caramanis, R.D. Tabors, and R.E. Bohn (1988). *Spot Pricing of Electricity* Boston, Kluwer.

Scottish Government (2008). "Transmission of renewable energy", news release of 18 April 2008, accessed at http://www.scotland.gov.uk/News/Releases/2008/04/18085238.

Sioshansi, R., S. Oren and R. O'Neill (2008). "The costs of anarchy in self-commitment markets." In F.P. Sioshansi, ed., *Competitive Electricity Markets: Design, Implementation, Performance*, Oxford, Elsevier, pp. 245-266.

Stoft, S.E. (2002). *Power System Economics: Designing Markets for Electricity* Chichester, Wiley.

Strbac, G. (2007). "Electric Power Systems Research on Dispersed Generation", *Electric Power Systems Research,* 77: 1143-1147.

Strbac, G., D. Pudjianto, M. Castro, P. Djapic, B. Stojkovska, C. Ramsay and R. Allan (2007). *Transmission Investment, Access and Pricing in Systems with Wind Generation* DTI Centre for Distributed Generation and Sustainable Electrical Energy.

Texas Public Utility Commission (2003). *Order Adopting New §25.501 as Approved at the August 21, 2003 Open Meeting*, Project No 26376, Austin, Texas PUC.

White, M. (1996). "Power Struggles: Explaining Deregulatory Reforms in Electricity Markets", *Brookings Papers on Economic Activity, Microeconomics,* 1996: 201-250.

Wolak, F.A., R. Nordhaus, and C. Shapiro (1998). *Preliminary Report On the Operation of the Ancillary Services Markets of the California Independent System Operator (ISO) Prepared by the Market Surveillance Committee of the California ISO*, Folsom, California, accessed at http://www.caiso.com/docs/1998/12/18/19981218064938135 77.pdf.

Wolak, F.A., J. Bushnell and B.F. Hobbs (2007). *Final Opinion on "Long-Term Resource Adequacy under MRTU",* Market Surveillance Committee of the California ISO, November 5, 2007, Folsom, California.