Cambridge Working Papers in Economics CWPE 0463

UNIVERSITY OF
CAMBRIDGE
**Department of
Applied Economics**

Electricity Transmission:

An Overview of the Current Debate

***Gert Brunekreeft, Karsten Neuhoff
and David Newbery***

The
Cambridge-MIT
Institute

*Massachusetts Institute of Technology
Center for Energy and
Environmental Policy Research*

CMI Working Paper 60

Cambridge Working Papers in Economics

UNIVERSITY OF
CAMBRIDGE
Department of
Applied Economics

Not to be quoted without permission

The Cambridge-MIT Institute

Massachusetts Institute of Technology
Center for Energy and
Environmental Policy Research

CMI Working Paper

Electricity transmission: an overview of the current debate

Gert Brunekreeft, Karsten Neuhoff and David Newbery[1],[2]

08 November 2004

University of Cambridge, Department of Applied Economics, Sidgwick Avenue, Cambridge CB3 9DE

Abstract: Electricity transmission has emerged as critical for successfully liberalising power markets. This paper surveys the issues currently under discussion and provides a framework for the remaining papers in this issue. We conclude that signalling the efficient location of generation investment might require even a competitive LMP system to be complemented with deep connection charges. Although a Europe-wide LMP system is desirable, it appears politically problematic, so an integrated system of market coupling, possibly evolving by voluntary participation, should have high priority. Merchant investors may be able to increase interconnector capacity, although this is not unproblematic and raises new regulatory issues. A key issue that needs further research is how to better incentivize TSOs, especially with respect to cross-border issues.

JEL classification: Electricity, Transmission, Regulation, Prices, Merchant Investment.

Keywords: L4, L5, L94

If the European Single Market is to be extended effectively to the electricity supply industry, then EU member states will need (amongst other issues) to make better use of transmission capacity, particularly interconnector capacity, to facilitate cross-border trade. At present international exchanges of electricity are less than 10% of total production in the old EU-15, and the poor correlation of spot prices between many neighbouring countries suggests that the national electricity markets are for the most part poorly integrated. This suggests either that cross-border flows are inefficiently impeded by the management of the existing interconnectors, or that there is insufficient interconnector capacity to allow price equalisation. Insufficient interconnector capacity may in turn result from a failure to undertake efficient investment in additional capacity, or because the expansion costs exceed the arbitrage benefits from unimpeded trade. More generally, market liberalisation allows consumers to choose their supplier, and hence would in a competitive market result in the cost-effective dispatch of existing generation capacity on a European scale. The growth in

[1] Tel: +44 1223 335474; fax +44 1223 335299. *Email address*: David.Newbery@econ.cam.ac.uk

[2] The authors gratefully acknowledge comments from Jean-Michel Glachant, Richard Green, Paul Joskow, Mike Metcalfe, Virginie Pignon and the participants of the CMI transmission workshop, held July 18th & 19th, 2003 in Cambridge (UK), who will find their contributions reflected. Support from the CMI project 045/P *Promoting Innovation and Productivity in Electricity Markets* and UK research councils ESRC/EPSRC under award number RG37889 is gratefully acknowledged.

intermittent generation makes closer European co-ordination more important. Wind power in particular creates new flow patterns across grids, which were primarily designed to provide secure supplies to moderately self-sufficient countries and are not optimally designed and controlled to handle these and other market-driven patterns. The same is true in the US, where decades of under-investment in transmission has revealed a system in stress just as the demand for more market-responsive electricity trading requires more capacity to interconnect regions under different Transmission Systems Operators (TSOs).

In response to the needs of liberalised markets and the stresses in transmission systems revealed by a series of high profile blackouts in the US and Europe, various bodies are now grappling with the problem of how to improve the operation and extension of the transmission network. In the US, the Federal Energy Regulatory Commission, FERC, pursued the implementation of a Standard Market Design and encouraging larger Regional Transmission Organisations to facilitate efficient trade over wider areas and transmission investment. In Europe, under prompting from the European Commission, the European TSOs are engaged in the Florence Process to develop rules for the better management of interconnectors and cross-border trade. The European Commission released a *Regulation on cross-border exchange* in 2003[3], which addresses these topics, although as we shall argue below, not completely satisfactorily.

In addition to these more visible inter-TSO issues of cross-border trade, investment, and the management of security, within each TSO's jurisdiction the transmission network is called on to perform an increasing role in supporting an effectively competitive power market. Both cross-border and internal transmission management demands raise a whole set of highly complex issues of both academic interest and policy relevance. The Cambridge-MIT-Institute Electricity Project held a two-day workshop in Cambridge, England in July 2003 to explore some of these issues, to delineate the present state of knowledge, and to identify important questions still remaining. This special issue of *Utilities Policy* provides a selection of papers reflecting discussions at that workshop. This paper gives an overview of the current debate, summarizing the discussions and indirectly introducing the remaining contributions.

Inevitably, the selection of issues is incomplete. It reflects what the participants of the transmission workshop considered to be the currently leading issues, and does not pretend to be exhaustive. We have tried to organize a large set of questions into a small number of overarching issues, set out in this overview paper. First, there is the important question whether locational marginal pricing (LMP) sets efficient long-term investment signals to generators and load, or whether additional locational differentiation of grid charges is necessary. This is discussed at length in section 1 below, and underlies part of the contributions of Boucher & Smeers and Perez-Arriaga & Olmos, since these deal with the pricing of (cross-border) transmission.

[3] Regulation (EC) No 1228/2003 of 26 June 2003 on conditions for access to the network for cross-border exchanges in electricity, OJ. L176/1, 15.07.2003, European Commission, Brussels. The regulation entered into force on 1 July 2004.

Second, and currently subject to intense European debate, the papers discuss congestion management and ask whether market coupling (also called market splitting) would be a workable model for Europe. The alternative, favoured by some parties, is a move towards increasingly refined coordinated auctions to manage cross-border trade within Europe. This is discussed in general terms in section 3, and examined in more detail in Boucher & Smeers, Perez-Arriaga & Olmos and Neuhoff & Newbery.

Third, the discussion on investment in transmission networks concentrates on the prospects and problems of merchant transmission investment. There is still active debate about their relative merits compared with regulated transmission investment, while the question of how they and regulated links should be regulated is the subject of the 2003 EU *Regulation on cross-border exchange*. While section 4 summarizes the main arguments for and against merchant transmission investment, Brunekreeft examines in detail the regulatory issues that emerge if merchant transmission investment takes place.

Fourth, the challenges of efficient network expansion and operation (including losses, congestion and balancing) imply the importance of understanding and designing incentive mechanisms for the system operator. One would like to avoid a system that provides perverse incentives for the operator to increase rather than decrease costs. Our understanding of these issues is still underdeveloped and deals more with what is wrong than how to solve the problem. Glachant & Pignon provide an insightful study of perverse incentives facing the Nordic TSOs, while Joskow summarizes experience and new developments in the US.

The reader should be aware of a institutional difference between the European and US situations, which is reflected in the paper. Throughout the paper we use the term transmission system operator “TSO” unless stated otherwise. Unfortunately, the name TSO is not unambiguous. A TSO comprises two parts: a transmission owner, TO, and a system operator, SO. These functions can be combined, as is usually the case in Europe, but can also be split, which is a trend in the US. Where the paper explicitly refers to the split (in a predominantly US context), we will use the term ISO (independent system operator) distinct from TO. An ISO is often the simplest solution to creating a wider market area without forcing the merger of the different grids into a single company.

1. Locational signals

The transmission network is a natural monopoly whose charges must be regulated. In an unbundled industry in which generators and consumers react to market signals the structure of network charges will have a potentially significant impact on network use and its development. It will affect the locational choices of new generation (and of energy intensive users), as well as influencing the bidding behavior of generators, and the willingness of neighboring electricity markets to trade and cooperate. Clearly, then, setting these charges at the right level is critical for ensuring the efficient use and development of the network and the wider electricity market. It is also one of the most challenging and difficult problems facing regulators. Ideally the structure of network charges should encourage:

- the efficient short-run use of the network (dispatch order and congestion management);

- efficient investment in expanding the network;
- efficient signals to guide investment decisions by generation and load (where and at what scale to locate and with what choice of technology – base-load, peaking, etc.);
- fairness and political feasibility; and
- cost-recovery.

Implementing an efficient scheme may require some political flexibility and compromise, as price changes typically create losers as well as gainers. The changes might reduce the profits of some private companies and/or a loss of power of some institutions (such as a former TSO within a larger dispatch area).

1.1 LMPs and deep connection charges[4]

For short-run optimal use of the network the benchmark is locational marginal pricing (LMP), also known as nodal spot pricing or a fully co-ordinated implicit auction. To achieve efficiency this requires that generators submit efficiently priced bids (i.e. a schedule of short-run marginal cost, SRMC, up to full capacity). The dispatch algorithm can then determine the efficient dispatch and the associated nodal shadow prices (which, if generators cannot increase output, can considerably exceed short-run marginal cost). Both generation and load would face these locational prices, although there would need to be additional grid connection charges to recover the balance of the regulated costs. If the bids are not set equal to marginal cost, scheduled flow patterns will be distorted. However, given that LMP results in a flexible allocation of transmission capacity and thereby creates the strongest net-demand response, it tends to mitigate the market power exercise of strategic generators more than other designs and hence distortions should be smaller than with other designs.

The concepts of efficient pricing have been developed in Bohn, Caramanis & Schweppe (1984), Read & Sell (1989) and the seminal contribution of Hogan (1992). Meanwhile, the LMP approach has been (or will soon be) implemented in some variation in several states in the US (e.g. PJM, New York, New England, Texas and California) and is a cornerstone of FERC's proposed Standard Market Design.

For short-run congestion management there is agreement that a system relying on LMPs works and is efficient (provided that bids are competitive). The more challenging question concerns the long-run effects of nodal pricing. The question whether differences between LMPs serve as the correct basis for rewarding new (merchant) transmission investment (i.e. the line owner) will be discussed in section 4. Here we discuss the question how (and how well) the LMPs work in giving investment signals for generation and load (i.e. the network users).

The question is closely related to the question whether LMPs recover all the costs of the network. If the LMPs recover all network costs then the LMPs unambiguously set the efficient investment signals for generation and load. In the long run, with optimal investment,

[4] It should be noted that we apply a broad definition of the term connection charges. These include annual charges to recover the fixed costs of the transmission network to the extent that these are not recovered by LMPs. Thus, our use of the term connection charges is broader than the charge for the original cost of being connection to the transmission system.

the difference between LMPs would reflect marginal network expansion costs. Unfortunately, for various reasons LMP do not recover all costs. Simulations suggest that even with optimal investment in generation and transmission long-run economies of scale allow only about 20-30% cost recovery (e.g. Perez-Arriaga et. al., 1995).

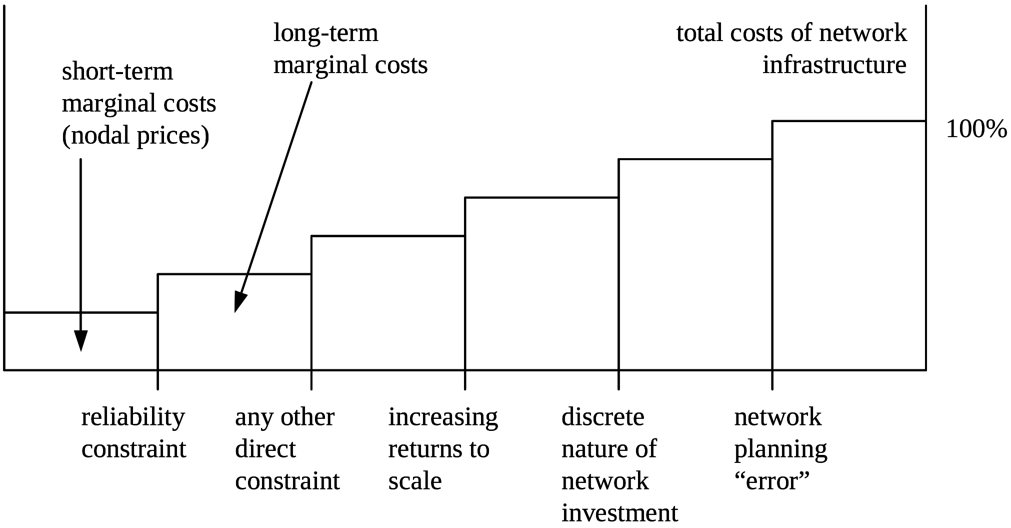

Figure 1: Cost drivers of a transmission network
Source: Perez-Arriaga, 2003

The various causes creating the wedge between LMP revenue and total network cost are illustrated in figure 1. Some of these sources of shortfall arise from a failure to properly charge for other attributes of the transmission system, many of which are quasi-public goods such as reliability. Others follow from the traditional reasons why SRMC pricing falls short of full cost recovery, such as economies of scale and lumpiness in expanding both generation and transmission to deliver the same standard of reliability. Offsetting this to some extent, properly charging for the marginal cost of losses (which are twice the average cost) will make a modest net contribution to system revenue (equal to the average cost of the losses). Losses can be significant compared to congestion charges, and estimates suggest that the total loss factor of moving power from a distant source to a load can be 20-30% (of the average cost of the power moved). Thus in New York State, the loss of moving power from the western part to New York City can be 20% or more and in the Western part of the US, where distances are longer, the loss factor can be 25-35% (Liu and Zobian, 2002).

Finally, as the costs of black-outs caused by inadequate transmission are very high, and the costs of somewhat over-building the network are rather modest, those charged with ensuring reliable supplies are likely to err on the side of too much rather than too little spare capacity. This further depresses the scarcity value of the network, lowering the dispersion of LMPs and hindering cost recovery (although it has an additional benefit in increasing the effective size of the market within which each generator bids, and hence reduces market power).

It follows that cost recovery requires a supplementary pricing scheme, either demand-based such as Ramsey pricing, or by deep connection charges, and/or a two-part pricing scheme. The question is, should these additional elements be locationally differentiated? If so, then LMPs do not provide all the relevant long-run locational signals, but if not, then LMPs are efficient short-run prices and the remaining problem is purely one of cost-recovery.

The natural approach is to draw on the standard model of competitive pricing, which states that an efficient equilibrium can be supported by a complete set of competitive prices if production sets are convex (i.e. there are no economies of scale and all externalities and public goods are also properly priced). If, as is the case with the network, parts of the production set are not convex, provided that the output decisions of the non-convex part are efficiently chosen (e.g. by a central planner or TSO), the remaining production decisions can be decentralized by the competitive prices, and one can also associate efficiency prices with the non-convex production decisions. (The distinction is that these efficiency prices would not guide profit-maximizing agents to the efficient non-convex production point and in that sense would not be competitive prices.)

The first point to note is that some of the problems identified in figure 1 are problems of non-convexity, but some may arise from not properly pricing reliability (and the resulting risks of blackouts). The question is whether an optimal *reliability-constrained* dispatch, where reliability includes all the relevant risks of line failure and generator outage, is accurately priced and included in the nodal prices (together with the cost of ancillary services such as spinning reserve at each location where it is required).[5] If so, then we can apply the competitive theorems. If not, then there will need to be additional charges. If it is possible to properly value the impact of any additional generation (G) or load (L) at any node on system reliability, then these charges on G and L should be included to give the proper nodal prices for analysis.

1.2 The central planned benchmark

Assume that the network is (centrally) planned, and the TSO accurately forecasts the (optimal) sequence of generation and network investments into the future. At each date, given accurate forecasts of load (amounts and location, which may depend on the evolution of nodal prices), the optimal dispatch is computed and nodal prices derived. In this case, LMPs do reflect the marginal costs of the network (*irrespective* of fixed infrastructure costs). The standard theories of competitive pricing imply that a complete set of prices (for each location at each time and date) will give efficient investment signals for generation (and load) provided these do not experience economies of scale. Investors will make their location and plant choices based on the present value of selling at current and future nodal prices (see O'Neill, 2003).

If there are economies of scale in generation there may be problems of coordinating the size and location of generation and transmission investments. In the EU and US, though,

[5] Baughman et. al. (1997) show that it is formally possible to integrate these components in transmission prices.

the optimal scale of generation plant is small compared with existing demand. Combined cycle gas turbines (CCGT) reach minimum economic scale in the range 50-350 MW, compared to typical system peak demands of 20-50,000 MW. The likely costs of mistakes (the wrong size at the wrong place) should therefore be small, although the extra transmission costs of mistaken G investment decisions could be significant. That is a possible subject for further research through simulating possible system evolution paths and comparing their costs and whether, given the price signals, generators are indeed led to locate efficiently (Baldick & Kahn, 1993).[6]

If these problems of coordination are indeed unimportant (and in any case where there are no economies of scale in G) then any additional cost-recovering charges need not be locationally differentiated, considerably simplifying their determination. While this appears to be a useful benchmark, it remains to check whether the discrepancy between our strong assumptions (perfect foresight, competitive pricing of generation) and reality are serious enough to invalidate the claim that the only spatial differentiation should come from the short-run nodal prices.

1.3 Problems with the benchmark nodal price solution - uncertainty

There are a number of sources of possible distortion to efficient decision-making based on these prices. The first is that future LMPs are not known and may be hard to predict. Investors deciding where to locate new generation based on current LMPs may choose the wrong location. Given the lumpiness of transmission investment large deviations of current LMPs from their long run equilibrium value can be expected. Of course, future electricity prices are also unknown, and may be very volatile on an hourly, seasonal and annual basis. The standard solution to the volatility of wholesale prices is a contract between buyers and sellers. When wholesale prices are high, sellers gain but buyers lose, and *vice versa* when wholesale prices are low. Each party reduces its risk by signing a contract, and the basic wholesale contract is a contract for differences (CfD). This specifies a strike price, s, a market price, usually a spot price in a formal wholesale market, p, and an amount, M. The generator receives the spot price, p, from the wholesale market, and $(s - p)M$ from the counterparty to the CfD. If the generator sells M in the spot market, his revenue is completely pre-determined at sM, and correspondingly for the buyer. The CfD will need to define the location where the wholesale price is set (a National Balancing Point, or Pool Price, or swing bus). A generator located at some other node will face a nodal price that may differ substantially from the price specified in the CfD, and will continue to be exposed to risk (often called *basis risk*).

The solution is an additional contract, a Financial Transmission Right or FTR (also called Tradable Congestion Contract, or TCC), that pays the holder $p - p_n$ per unit at node n. If the generator holds M TCCs at node n and a CfD for M, and generates M, then his revenue

[6] In practical terms, the major constraint on location is more likely to be planning permission both for generation and transmission, so that in many countries new generation will be on existing sites (that have grid connections and access to cooling water), making use of the existing grid. It remains an interesting and potentially important question how much these politico/environmental or NIMBY (not in my back yard) constraints cost.

will again be *sM*. The prices of the contracts may be positive or negative, depending on the strike prices and the forecast underlying prices, but will be known at the time of entering the contract, eliminating price risk.

The logical solution to the problem of the unpredictability of future nodal prices is for the TSO to offer a long-term FTR to the reference node (where energy is priced and traded). Its price (or value) is the present value of the predicted shortfall of nodal prices at that location relative to the reference node over the life of the investment. (In practice this would be issued as a debt instrument that could be liquidated at a constant yearly rate over the contract life.)

1.4 Commitment problems

The second problem is that the TSO may find it difficult to commit to its transmission expansion schedule regardless of generator decisions. Thus the least-cost expansion plan may involve generators investing inside an import-constrained zone, where competition may then become rather intense. A company may prefer to locate instead in an export-constrained zone, predicting that the TSO will have to invest in extra transmission. Once the generator has made its decision, the least-cost way for the TSO to fulfill its license obligations to deliver security of supply and adequate capacity may be to expand transmission capacity. This combination of the G and T investment would be more expensive than the least-cost expansion plan, but given the G investment, might be socially preferable to not investing in transmission. If the TSO had been able to commit to not investing in additional transmission, the low prices in the export-constrained zone might have deterred the generator from its investment and encouraged the least-cost solution.

This problem suggests the attractions of deeper (and hence spatially variable) grid connection charges for new G. In exchange the new G would receive a FTR. The effect should be that the TSO can use FTRs to commit to future nodal prices and hence can effectively deter inefficient locational decisions. This is not an additional spatial differentiation charge on top of the LMPs but merely a contract for those LMPs.

There is a related issue that falls part-way between uncertainty and the problem of predicting regulatory actions and the consequent nodal prices. Joskow and Tirole (2004) show that the relationship between efficient prices and optimal prices during reserve deficiencies are extremely sensitive to discretionary actions by the TSO, and to that extent hard to predict. Again the solution would seem to be for the TSO to issue FTRs to reduce this uncertainty.

1.5 Market power

The third problem is that in a fully liberalized market, generator market power can distort bidding and hence the calculation of nodal prices (which become locational *market* rather than *marginal* prices). Distorted prices may induce an inefficient pattern of investment. Different systems of regulation react to this problem in different ways. The US approach which requires regulators to ensure that prices are “just and reasonable” can lead to such heavy-handed regulation as requiring generators to submit cost-based bids if their market bids distorts dispatch and pricing beyond specified bounds. A more relaxed approach would assume that provided there are no artificial entry barriers, the exercise of market power will be self-limiting, inducing competitors to enter at high-priced nodes, and permanently reducing future revenues. Since entry at modest scale was economic with gas-fired generation, ensuring that entry is contestable was perceived to be sufficient to ensure workably efficient outcomes.

More to the point, if generators hold contracts equal to their planned output, they will have no incentive to misrepresent their bids. In a network with transmission constraints this requires that generators hold transmission contracts to complement energy contracts with counter parties at other locations. Joskow and Tirole (2000) assess how such transmission contracts can impact the exercise of market power and Gilbert et.al. (2004) show how auction design and restrictions on ownership can reduce the exercise of market power by strategic generators. Since generators may choose to contract for hedging reasons, the problem may not be too serious, provided shortages (that greatly amplify market power) are not readily predicted. Where there are predicted and potentially lengthy shortages (e.g. a systematic shortfall in capacity that will require new build that cannot come on stream for some considerable time) then market power may spread to the contract market. Price caps on contracts are typically far less distorting than on spot markets, and a requirement to offer such capped contracts defensible.

Note that most of the alternatives to LMP are potentially worse at dealing with the locational distortions caused by market power. The original English electricity Pool offered a single price and firm transmission rights, so that plant that could not be dispatched because of constraints would be paid its theoretical lost profit (Pool price *less* bid) to not generate. A generator assured that he is not required and facing little local competition might then bid very low to maximize income. It may then pay to locate in an export-constrained zone to enjoy these profits, even though this is exactly the wrong place to locate. The counterpart is that generation in import-constrained zones can bid high (and be paid its bid price if constrained to run out of the unconstrained merit order) and will therefore be more strongly motivated to locate in such zones by the presence of market power.

1.6 Pricing for supply security

A fourth source of potential distortion already noted is a failure to properly value and price the impact of investments on security and quality of supply. Joskow and Tirole (2004) explore this issue in some depth, and the challenges faced in adapting the old engineering-based approach to cope with liberalised markets. Typically, and reasonably, given the asymmetry in social costs between under- and over-investment, investment standards are set conservatively, so there is a tendency to over-invest in security. If the operational standards would be equally conservative, requiring the same amount of reserve capacity to satisfy security criteria, then the locational price differences would reflect the additional security requirement. If, however, the spare capacity required at the investment stage exceeds the operational reserve capacity, then the nodal price differences will be reduced and the locational signals muted, while at the same time the cost-recovery problem exacerbated.

There are several possible solutions. One might be to add an additional locational charge for security and quality, if location makes an important difference to either security or quality of supply. The second is to make the fixed charge spatially variable to restore the price signals for G location. Britain, which has not adopted LMP, has annual zonal charges for being connected that give (and are designed to give) quite strong locational signals. This is perhaps the most important qualification to the claim that nodal prices provide all the locational variation that is required. However, while the annual charges should properly be based on capacity (or peak demand), the nodal prices will charge according to energy, and these short-run energy price signals should not be distorted by the corrections required to signal investment location.

1.7 Deep connection charges

A related question is how to charge new G where their entry requires expanding and/or upgrading the grid (either locally, by a shallow connection, or in some other parts of the network by a deep connection upgrade). It is tempting to argue that as LMPs will not recover all network costs, so those who cause grid investment should pay the extra cost of that investment (through deep connection charges). It is standard practice for G or L to pay for the assets required to connect to the nearest suitable grid connection point, and this carries the logic beyond that point where reinforcement is required. If so, then it might also be reasonable for this charge to take the form of a long-term transmission contract to the reference node, transferable to other generators at that node if the generator chooses to disconnect. The problem would not arise in the absence of all the other distortions, and the relevant distortion here is a combination of not knowing the future evolution of the whole system (G, L and T) and the tendency in the presence of uncertainty to err on the side of security and overcapacity.

It is a natural step to argue that as the LMPs do not reflect all network costs, the incentives for generation investment may be distorted, because the generator would not bear all the costs it causes. If connection charges should reflect cost causation, one is led to propose *deep connection charges*. The principle was adopted in Britain for setting the entry and exit prices of the gas National Transmission System, which asked of each entry and exit point what would be the additional investment required to support a significant unit increase in injections or withdrawals. These costs were then translated (rather crudely) into regulated charges, which, however, applied to all requiring entry or exit at that node. The system of setting prices for entry was then subsequently replaced by auctions, which are the natural counterpart to the electricity system LMPs.[7]

[7] Auctions for (and subsequent trading in) explicit entry rights seem feasible for some gas networks (and are used in the UK), where there are only a few relevant entry points and sufficiently many suppliers wishing entry access. Gas networks have considerable flexibility, and can usually manage with daily balancing, rather than minute by minute for electricity. In contrast, physical transmission rights for electricity might be too complex to allow for sufficient liquidity for efficient pricing and trading.

Although deep connection charges appear attractive at first glance, they are not unproblematic. Whereas it is possible to reasonably simulate the post-investment situation (as in PJM in the US), it is not at all clear what the costs caused by the new connection actually are or should be. If, for example, the upgrades take account of indivisibilities to over-build ahead of future demand, what fraction of the cost is attributable to the present connection? If capacity is typically oversized because of the asymmetry in cost of under or over-building, is it reasonable to charge this general security benefit to a new entrant? Finally, and decisively in the British case, if there are benefits to encouraging entry to mitigate market power, then deep connection charges discriminate against entry (and additional G capacity) and lose some of their attraction.

Shallow charges promote new entry, which may appear attractive to those wishing to promote renewables or additional generation capacity to improve security of supply. The issue is rapidly gaining relevance with the development of large-scale off-shore wind energy in the UK, Denmark and northern Germany. Exactly how such generation should be charged has attracted considerable attention in Britain, where NGC has estimated that the grid reinforcement costs to handle modest amounts of extra wind power in Scotland might be £205-525 million, or 75-85% of the total cost of all new wind power. As the Government document notes "Given NGC's shallow connection charging policy, these costs would not form the basis of connection charges, but would rather be additional capital expenditure to be recovered from all users through transmission network use-of-system charges (TNUoS)." (DTI, 2002, ch 8).

These charges are derived from a simplified DC investment cost model of a the grid, and are spatially differentiated, but are levied on the highest injection declared for the year. As such they may not properly reflect the load patterns of wind compared to conventional generation, and may not be, as arguably they ought, a fair reflection of the annual implied nodal price differences from a reference node. This suggests that the current charging methodology overstates the regional access price differences for intermittent wind power. In response, there have been suggestions that wind power should face a locationally undifferentiated grid charge, but the regulator, Ofgem has correctly argued that this would not encourage renewables to locate in the least cost (generation *plus* transmission costs) location. The discussion illustrates that existing transmission charging regimes are typically developed for existing technologies, and may require adjustment to provide a technology-neutral treatment of new technologies.

To summarise, in the absence of LMP, there is a strong case for a locational element to grid charges, and these should be computed to guide location decisions to minimize the present discounted cost of all G and T investments required to maintain reliability and security standards. The logic of British zonal charges was to roughly reflect the incremental cost of transmission investment to support additional generation in each zone in a way that was auditable by any potential investor. The latter requirement may have distorted the signals from their efficient level, but the gains in transparency and accountability, and the difficulty of predicting the relevant efficient charges in an uncertain world, probably outweighed the disadvantages. Of course, the difficulties could be avoided by using LMPs and FTRs, but these face one final problem.

1.8 LMPs without a Pool price

If all generation has to submit bids at each node, then it is straightforward to compute the LMPs. The English Electricity Pool could have computed such LMPs from the data submitted each day, but was replaced by the New Electricity Trading Arrangements in 2001. At that point the pool and compulsory bidding ended, to be replaced by bilateral and OTC markets, a voluntary day-ahead power exchange, and a short term Balancing Mechanism to elicit bids and offers to balance the system and reconcile constraints. The latter was a pay-as-bid auction with two imbalance prices (for being short or long) set at the average price of the bids or offers accepted. As such the balancing mechanism is ill-suited to producing a single set of LMPs, first because it trades only a tiny fraction of output, second, because it lacks a single locational marginal price comparable to the day ahead price of LMP, and third, because it does not provide sufficient notice about congestion pattern to influence ramping decisions.

The Nordpool and PJM experience shows that LMP (or, more precisely, market coupling) works if all parties that are transacting across regional boundaries (or nodes at PJM) have to submit bids to the energy spot markets in the respective areas or have to notify bilateral transactions whose transmission is charged the day ahead spot-price difference between the injection and exit node. Hence all contracting positions are effectively financial contracts, which are settled at the spot prices at each region. For these prices to provide full hedging, the spot markets have to be cleared by one central algorithm that arbitrages the markets subject to the available transmission capacity between the regions.

It would therefore seem to follow that LMPs sit most comfortably with a compulsory gross pool, although the Nordic market manages dynamic market splitting without a compulsory pool.[8] That said, NordPool is far more liquid than any other electricity power exchange, and there is a well-defined price to use for zonal differentiation.

[8] Nordpool uses the term market splitting to deal with the case in which constraints prevent the Nordic market clearing at a single price, at which point the market is split into pre-defined zones in each of which there is a single market clearing price. In this paper we use the term market coupling to cover this case as well, even though the markets that have been combined under a single system operator may need to be subdivided into separately priced zones. The full description is given in section 3 below.

2. Recovering the short-fall in transmission revenue

One very simple way to demonstrate the short-fall between LMP revenue and that required for the regulated transmission business is to imagine that each node normally has an equal generation and load. As LMP prices would be the same for G and L, there would be no net revenue, and the security benefits of interconnecting the nodes would be unrewarded. Of course, G and L are most unlikely to be balanced hour by hour, even if their capacities are the same, and cheaper G will export, lowering LMPs and producing net revenue. Nevertheless, all the evidence suggests a substantial short-fall that will need to be recovered by non-locational charges, and these will have to introduce a wedge between the G and L charges at each node.

The regulator will have determined the allowable revenue for the transmission system, and hence the size of the shortfall after taking account of LMP revenue and those for ancillary and other grid services. The revenue will have to be allocated to capacity (or power) of G and L. This immediately raises several questions: does it matter how this wedge is allocated between G and L, and on what basis should the G and L charges be set?

The first point to note is that in a competitive and isolated system, the proportions charged to G and L make no difference, as the final price paid by the consumer will be the G cost plus the T charge. If the fraction of the T charge t paid by G is α, and the generator's efficient bid is b then the wholesale price will be $b + \alpha t$, to which the L will pay an additional $(1- \alpha)t$ to give a delivered price of $b + t$. The appendix shows that this continues to be true even if the generators have market power and bid above their efficient price. Clearly, if two interconnected systems choose a different allocation there will be distortions. If, for example, one system places all the grid charges onto L and the other onto G, then the first system will have a comparative advantage selling to customers in the second, unless the interconnector levies a suitable charge. Harmonising the G:L balance therefore becomes important in interconnected systems, and there is some attraction in levying all the grid charges on consumers. However, this requires care in systems without the locational signals contained in LMP prices, where the correct interpretation is that the weighted average of G connection charges is zero, preserving any locational differentials in G charges that are required. Thus in Britain for 2004/5 the annual zonal G tariffs range from £10.7/kW to -£6.8, a range of £17.5/kW. This is exactly half the interest and depreciation on a CCGT plant costing £310/kW at 10% interest and hence a strong locational signal, which requires some negative G charges.[9]

The next question is to determine the basis on which the charges should be set. Ideally, as the charges are effectively taxes to recover the shortfall, they should be minimally distorting, and independent of any actions that those connected might take. With that in mind, let us consider the British system as an example. G pays according to Transmission Entry Capacity (TEC) connected to the system on April 1 each year, and L pays according to demand taken at the "triad" – the three half-hours of system peak demand separated by 10

[9] To ensure that the plant provides useful power when needed, the payment to the plant depends on its output at the system peak (strictly, at the triad described in the text).

days, an amount that is determined after the event. Both suffer from potential limitations. Consider first the question whether an annual fixed charge discourages rarely run peaking plant from the (potentially considerable) annual connection charge. It is most likely to be required at the triad, in which case consumers will pay the same grid charges regardless of how the G charge is allocated over different generators. If, however, the peaking plant pays the full G charge, investment in or keeping available of peaking plants requires higher or more frequently peaking prices to recover this cost and this will be passed on to consumers. The consumer price will be higher and hence demand lower in these periods, and that will have some effect on choice. As such it will be distortionary.

The question of how to set the least-distortionary form of taxation was addressed and answered by Ramsey in the 1920s. In its simplest form a Ramsey tax is a mark-up on the efficient price inversely proportional to the demand elasticity.[10] Ramsey pricing is now accepted as the best feasible way of collecting such shortfalls in revenue (at least provided the resulting charge does not bear more heavily on the poor than alternatives).[11] Stoft (2002, p. 413) notes that Ramsey pricing is theoretically optimal but is "complex and controversial and will not be pursued." He argues instead for spreading the charges "as thinly and evenly as possible", but does not say whether this should be proportional to TEC (as in Britain) or output (MWh).

If an annual grid connection charge for rarely-run plant discourages that plant's availability, then it distorts production decisions and fails the Ramsey test. What is required is a cost-recovering charge that is based on the physical characteristics of the plant rather than use (e.g. thermal efficiency, fuel, etc. which will determine its merit order) and possibly age (that may be correlated with efficiency and hence also merit order). However, such charges are likely to distort the choice of plant type and availability decisions.

It might be objected that uncertainty about future grid charges could discourage generation investment. That could be solved by offering a contract for the expected life of the station, whose annual payments are specified in advance and are independent of whether or not the station continues to operate (or be connected). Apart from the difficulty and costs to obtain the required credit guarantees it would still be necessary to differentiate charges according to Ramsey principles, in order to persuade investors to install peaking plant, but it would reduce the risk of *ex post* opportunism in resetting the grid charges.

The case for setting the L charges on the basis of the triad dates back from the period when transmission and generation were both provided by the Central Electricity Generation Board, which charged a two-part tariff for both generation and transmission capacity (based on peak or triad demand) and energy (the system marginal price). It is less clear that the capacity element of the grid alone is best measured by the triad demand – indeed some of the heaviest uses of the grid arise off-peak when cheap power is wheeled over longer distances to the load. Peak-load plant being footloose can usually be located closer to demand. Nevertheless, Ramsey

10 This is true provided there are zero cross-price elasticities and no externalities, otherwise the formula is more complicated.

11 Or, correctly, providing the regulator setting the charge is required to set the charges solely on efficiency grounds, leaving the tax system to address any distributional concerns.

pricing suggests that if demand is less elastic at the peak then grid charges can be concentrated there, although it seems unlikely that the ratio of the price elasticity at other times to at the triad is almost infinite (which is what would be required to justify loading all the charges on those three half-hours).

To summarise, in an unbundled and liberalised electricity industry, cost-recovering grid charges risk distorting choices of plant type and operation. We seem to be some way short of an agreed and feasible methodology for such charges, although we are also unclear how serious the costs of any potential distortions are. The need to finance grid reinforcements for new wind power may encourage more research on this topic.

2.1 Cross-border tarification (CBT)

A postage stamp is a charge for Use of the System (UoS) which is independent of distance; it is a common approach to apply a system of connection fees (energy or capacity based) which are not transaction related. The promotion of competition is often argued to justify a postage stamp charge for network use, and within a country this may be reasonable if internal congestion is minor. The Netherlands justifies such a charge on the grounds that the country is a "copper plate" with no internal congestion, although there are clearly some parts where only one generation company can deliver power at certain times of the day. In other words, a grid which is adequate for a socially optimized dispatch may allow pockets of market power that can exploit the potential to congest internal links, and in such cases LMP charges have a (possibly minor) beneficial impact on market power. If the postage stamps differ across regions they are called license plate fees.

A major advantage of both systems is that they avoid pancaking the fees of several TSO regions involved in a transaction (i.e. stacking each region's fees on top of each other). As such it removes inefficient protection of local generators and thereby promotes competition. Pancaking is clearly to be avoided if possible, but it is less clear how to efficiently charge for transmission where several networks are interconnected, and how to compensate the network owners so that they will be willing to trade and invest. Consider countries (and thus networks) A, B and C. If A exports to country C and most of the flow goes through country B, then country B is a transit country. Although the trade (from A to C) relies on network B, with network costs charged to the generator and final user, A and C do not actually pay for the network in B. Horizontal inter-network compensation arrangements aim to repair this flaw and are discussed, in the European context, under the header cross-border tarification, CBT, (and in the US as the "seams issue").

The compensation issue is irrelevant for efficiency as far as sunk costs are involved, although if the compensation is based on flow patterns, then it might induce network operators to distort power flows from the optimal dispatch. Inter-network compensation becomes important whenever it affects investment decisions. In the European example the national regulator typically regulates the national network and there is no counterpart to FERC, which in the US has jurisdiction over inter-state (and thus inter-TSO) flows. The regulator in region B might not support a beneficial network expansion in country B if all the costs fall on G and L in B and the consumers and generators benefiting in A and C make no contribution. Furthermore, the details of the compensation rule will affect grid revenues for A, B and C and thereby affect the use-of-system (UoS) charges required to make up the shortfalls in each area. These may then affect investment decisions for generators and load. Whether this is empirically relevant depends on the amount of the compensation payments. As long as the design of cross-border compensation payments remains unsettled, it will be difficult to justify interconnector investments, as these become sunk and less relevant to bargaining over the determination of the payments once made.

Vázquez, Olmos and Pérez-Arriaga (2002) discuss various schemes for charging network users for that (substantial) part of the total cost not recovered from congestion charges or charges for losses (which typically make a profit).

2.2 The marginal participation rule or area of influence rule

The method of marginal participation attempts to estimate how flows respond to a change in injection (or withdrawal) of 1 MW at any node, with a view to charging agents at that node their share of the costs of the links on which flows change. Boucher & Smeers (2003, p. 20) stress the difficulty of the calculations and the arbitrariness of such cost-allocation rules and question the efficiency effects in general. Vázquez et. al. (2002) point out that the choice of the "slack bus" effectively determines from which node the withdrawal (or injection) of 1 MW comes, and hence strongly influences the resulting flows. As a consequence the principle of the marginal unit can lead to unreasonable results if the marginal value in a line is significant although the "real" flow is negligible (Vázquez et. al., 2002).

However, it is important to note that irrespective of the choice of slack bus, if both generation and load are exposed to the same locational component, then the net total payment for a balanced transmission is not influenced by the choice of the slack bus. The choice of the reference bus only determines the allocation of costs between generation and load. Vázquez et. al. (2002) demonstrate this conjecture, and their argument is set out in the appendix.

There are additional refinements of the marginal participation rule that are likely to be even more complicated to compute, even if they are conceptually more satisfactory. Logically, any additional injection has to be matched by withdrawals of the same total amount (less losses) but not necessarily at a single node (let alone an arbitrary slack bus). Conceptually, an increase in power in a competitive system would cause prices to change by different amounts at different nodes, and these price changes would stimulate changes in generation and consumption adding up to the required amount. The associated set of withdrawals would define the marginal changes in power flows and hence the relevant grid charges. It then remains to determine the fair amounts to charge for the use of these lines, and here we come back to the original problem that efficient pricing of lines leads to under-recovery, while average charges remove the efficiency justification.

The obvious objections to the marginal approach are that it is too complicated, apparently arbitrary, and likely to either under-reward additional investment, or over-charge relative to the efficient solution. That suggests the need for a simpler, defensible rule that would be acceptable to the various TSOs who have to reach agreement.

2.3 The average participation rule or tracing

This rule allocates responsibility for the costs of actual flows on various lines from sources to sinks according to a simple allocation rule, in which inflows are distributed proportionally between the outflows (Vazquez et.al., 2002, p. 5). The main attractions of tracing are that the rule has some theoretical backing based on the Shapley value (cf. Kattuman et.al., 2004) and does not require the choice of a slack node. The drawbacks of tracing are first that aggregation of users can lead to counterintuitive results: If generation and load or different nodes are aggregated, then they are exposed to different tariffs. Second, the choice of the allocation rule is decisive but apparently arbitrary (apart from the Shapley justification).[12]

The marginal participation rule is applied in Argentina (cf. Woolf, 2003, pp. 262 ff.), where it works reasonably well because of the predominantly radial system. In Argentina, the main load centre is Buenos Aires, which is the slack node, and the power plants are built near faraway gas fields. The lines are mainly radial lines from the generation centres to the load centre, and thus the allocation of cost is relatively straightforward.

While this suggests that the choice of mechanism should depend on the nature of the network under consideration, in Europe and the US, within-country grids are typically densely meshed and the marginal approach is likely to be too complex and contentious. If the TSOs can be persuaded of the fairness of the tracing approach, and if it leads to (or can be combined with) efficient interconnector investment, then it has a great deal of merit. The first step in Europe would seem to be to set up an independent technical group, provide it with all the relevant data for the entire European network and the dispatch pattern, to see if they are able to identify robust candidates (on social cost-benefit terms) for interconnector investment, and to work out the implied patterns of costs and benefits to see if each TSO benefits from the set of such investments under the proposed cross-border tariffing rule. While it may be too much to expect each individual investment to be Pareto improving, there is more chance that this would be true of a large enough set, given the poor current state of interconnector capacity.

3. Cross-border congestion management

The TSOs as well as market actors have stated that congestion occurs at least occasionally at almost every European border (Haubrich et al., 2001).[13] There is general agreement that market-based mechanisms are required to allocate scarce transmission capacity on congested links. The theoretical optimal solution is nodal pricing, successfully

---

[12] The Shapley allocation results from the exercise of the bargaining power of the various coalitions of countries or regions involved in setting the cost allocation rule, according to some reasonable *a priori* restrictions on how to measure bargaining power and how to divide the gains from cooperation.

[13] Haubrich et al. (2001) assessed the following corridors: Spain-Portugal, France-Great Britain, France-Belgium, France-Germany, Netherlands-Belgium/Germany, Germany-Denmark, Germany-Sweden, France/Switzerland/Austria(/Slovenia)-Italy, Austria-Switzerland, Austria-Germany, Norway-Sweden. But interconnections with central European countries (e.g. Hungary) are also frequently constrained.

applied in parts of the US East coast.[14] However, it requires centralization of system responsibility, which is currently perceived as difficult to achieve. System operators are responsible for system balance and constraint management of their regional/national network, and they argue that only they can provide the required system security. Not surprisingly, they want to retain this authority. System operators currently determine bilaterally the amount of transmission capacity made available between neighboring countries for commercial transactions. If demand exceeds the available capacity, then auctions can be (and often are) used to make this capacity available to the market.[15]

A shared perception is that the current decentralized auctions of international transmission capacity fail to make effective use of a highly meshed and integrated electricity network.[16] For example, if electricity is transmitted from France to Germany then energy not only flows along the direct links between the countries, but also across the interconnector from France to Belgium and then via the Netherlands to Germany. Currently the Belgium system operator does not receive timely information on German-French flows. He therefore has to be conservative in issuing commercial transmission rights, e.g. between France and Belgium, because he has to anticipate the largest possible impact the unknown flows from France to Germany could have on the network. Most of the times the flows will not have this large impact, so the network is underutilized even when there is a scarcity of commercial transmission rights between France and Belgium. This could induce the system operator to apply less conservative estimates, which could jeopardize system security if large transmissions both between France and Germany and France and Belgium coincide. The problem is made more difficult as the share of intermittent generation from wind increases, where generation can change dramatically on time scales that are short compared to the allocation procedure for interconnection capacity.

Figure 2 illustrates two similar solutions that are suggested to resolve the problem of international flows: a coordinated auction for all European interconnector capacity and market coupling between European electricity markets.

The coordinated auction consists of three steps: First, each system operator has to inform the central auctioneer about the transmission capacity available for commercial flows on its transmission network. Second, market participants submit bids for transmission rights between any two countries. Third, the auctioneer allocates available commercial transmission capacity to the bidders using an algorithm analogous to nodal pricing. Based on the submitted information this offers a well-defined and non-disputable allocation. The auctioneer effectively considers the interests of both the bidders for transmission capacity from France to Belgium and France to Germany when determining the optimal set of international transmission rights he issues. The auction revenue can either be allocated to some sharing

[14] PJM and New England. Note that although PJM has adopted nodal pricing to make better use of the transmission system within PJM, this does not solve the problem of trade across the boundaries with other TSOs – the so-called "seams" issue, nor does it ensure that efficient interconnector investments will be undertaken across these seams. See Joskow (this issue).

[15] Examples are Germany-Netherlands or UK-France. See e.g. Newbery and McDaniel (2003).

[16] ETSO (2001), European Commission (2003), Boucher & Smeers, (2003)

rule, or attributed to the individual network constraints according to their scarcity value calculated in the auction process.

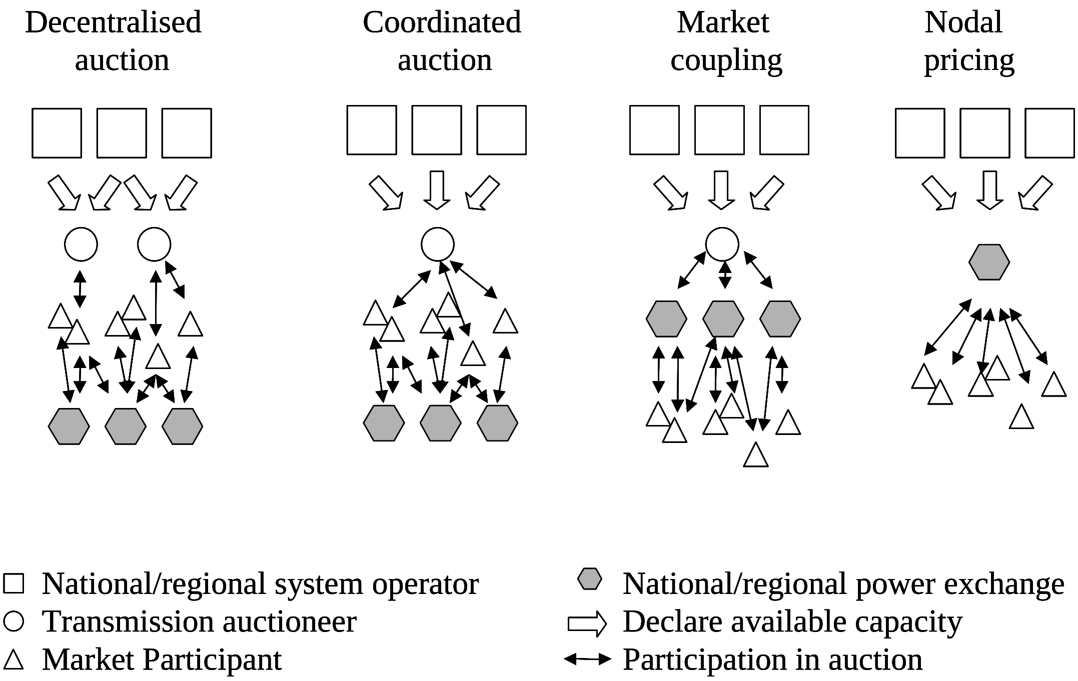

Figure 2: Different allocation mechanisms for scarce transmission capacity

The market coupling (sometimes called market splitting) approach works in a similar way. Once again, each system operator has to inform the central auctioneer about transmission capacity available for commercial flows on the transmission network. However, instead of market participants bidding for cross-border rights, this time the national power exchanges submit bids to the central auctioneer. Each national power exchange would add all bids and offers to create a net demand curve, which is submitted to the central auctioneer. The net-demand curve specifies at which price the national market would be balanced (the current market clearing price) and what amount of energy would be available for exports at higher prices or required from imports at lower prices. Analogous to the coordinated auction the system operator uses the typical nodal pricing method to determine the optimal use of the commercially available capacity between countries. He issues transmission rights to the national power exchanges to implement this solution. The national power exchanges then clear the local power exchange given net imports and exports and set the local prices. The revenue the central auctioneer receives from the transmission rights can be used as described in the coordinated auction.

The advantage of market coupling is that energy transmissions are determined after generators and demands have submitted their information to national power exchanges. This allows for the use of all available information and improves on the efficiency of production and allocation decision. For instance Joskow (2003, p. 21) notes that where energy and transmission-capacity markets were not integrated (like California and Texas) congestion costs appeared to be too high, making a case in favor of integration. Neuhoff (2003) uses the explicit auction between Germany and the Netherlands and the market coupling between Sweden and Northern Norway for a test that agrees with the hypothesis that explicit auctions allow for more exercise of market power because generators face lower effective demand elasticities. In a numerical model of the meshed network of the Benelux countries, France and Germany, Ehrenmann et.al. (2003) show that market coupling would reduce prices relative to a coordinated auction of interconnectors. A potential additional benefit of market coupling is that all transmission allocation is firm so that counter-flows allow an effective use of the network. This is also possible if physical transmission rights are formulated as obligations, but currently their implementation makes them look more like options than obligations and it is unclear whether this situation will change. Options reduce the amount of capacity that can be safely allocated relative to obligations which can offset other flows.

Both proposals for coordinated auctions and for market coupling currently discussed in Europe rely on the assumption that congestion is an international issue and ignore congestion within countries. Creating only zonal prices (Sweden) or facilitating bilateral contracting by ignoring transmission constraints in the energy market (as in UK) creates perverse incentives for the location of new generation, as noted above. Thus a generator in the frequently export-constrained North of Sweden/UK first gets the national energy price, and then further profits in the balancing market when his re-dispatch bid is accepted by the SO that has to resolve constraints. The lack of explicit congestion treatment thus gives the wrong investment signals (Neuhoff, 2002). In addition generators may bid strategically to create congestion, as occurred in the US with zonal pricing, but it would equally apply to coordinated auctions.

Another way to describe the relative merits of market coupling over coordinated auctions is that because all markets are cleared simultaneously, the auctioneer can make more transmission available and hence reduces the extent to which individual generators are able to exploit congestion constraints, which fragment markets and increase volatility. Harvey and Hogan (2003) address the question of whether to subdivide zones to address internal congestion. They show that the impact of market power is weakly smaller if zones are split up instead of using the system operator to re-dispatch generation to resolve the constraint and maintain a single price. Splitting up zones in an explicit auction design requires that any transaction between the subzones is exactly matched by corresponding physical transmission contracts, increasing contract complexity and reducing liquidity. With market coupling subdividing zones is less critical. First, financial transmission contracts would only be required if the risk of price differences is perceived to be significant. Secondly, financial transmission contracts between the subzones could be defined over longer periods, as most of the price risk can be hedged even if not every energy transaction is exactly matched by transmission contracts. This suggests that market coupling is preferable.

Finally, from the perspective of market participants financial transmission contracts offer the same services that they would expect from physical transmission contracts. Assume a German generator sells to a Dutch industrial load and the price difference is hedged with a transmission contract. How will they subsequently use the contract? If they own a physical transmission contract, they will nominate the generation and exports in Germany, present the transmission contract, and nominate the imports and demand in the Netherlands. With a financial transmission contract, they will offer generation at any price in Germany and the bid for the corresponding load at any price in the Netherlands. They will be exposed to the price difference of the two markets, which is exactly covered by the financial transmission contract. They can actually improve on this with financial contracts, as bidding avoidable cost for generation allows a cost saving if the market-clearing price in Germany is below the avoidable cost, for the load will still be served in the Netherlands.

3.1 Implementation

Why are European governments not willing to implement nodal spot pricing (or the simplified version of market coupling) as the logical congestion management scheme? Pérez-Arriaga & Olmos (this issue) argue that the number of control areas in Europe (17 to 27 depending on where the borders are drawn) is significantly smaller than in the US with some 200. This may reduce the need for a single integrated solution in Europe. Moreover, as the control areas are larger and their boundaries tend to coincide with country borders, the European member states may have a stronger political voice than the states in the US (though see Joskow's article in this issue for a more pessimistic view on the US). Lastly, because the interconnections between control areas are the 'weak links' of most transmission systems, the large number of control areas in the US resulted in larger benefits to efficient congestion management.

Apart from this political constraint the previous arguments all seem to favor market coupling rather than a coordinated auctions for transmission capacity. However, the implementation of market coupling requires the coordination of a large group of stakeholders. Initially TSOs were concerned about the implications for security of supply of market coupling. Now it is acknowledged that market coupling allows TSOs to retain the same operational autonomy to ensure security of supply. Organizations involved in energy trading might prefer auctions, as they provide trading opportunities and the uncertainty involved increases trading margins. However, firms (among them TSOs) interested in merchant transmission investment and generation companies considering the location of new investments are affected by the likely instability of future regulation if an unsatisfactory market design (such as coordinated auctions) is implemented. Moving towards the most efficient solution thus reduces regulatory risk.

Given that a full European nodal approach is currently some distance from political reality, could an incremental approach via regional market coupling be pursued? Could market coupling be introduced into smaller areas and then gradually extended e.g. starting with the Benelux and then extending to neighboring countries? This might require that most expansions are Pareto improvements for the relevant decision groups. Neuhoff (2003) shows that coupling of the Belgian and Dutch markets should not increase prices on either side if the same constraints on the exercise of monopoly remained after market coupling is implemented. This is, however, a strong assumption, discussed by Neuhoff and Newbery in this issue.

Many of the principles set out in the FERC SMD are economically sensible[17] and where implemented could provide lessons for Europe, particularly if their experience demonstrates the superiority of a market design that could be lends itself to replicated in Europe and is compatible with efficient market integration. The main concern is to avoid choices in the short run that make it difficult to move towards an efficient system of cross-European pricing and use of the full interconnection capacity.

4. Merchant transmission investment

Merchant (also called market-based) transmission investment is a relatively new development currently under discussion in many parts of the world. Merchant transmission investment refers to investment in transmission facilities remunerated by arbitrage between differently priced regions. Such merchant investment is not eligible for regulated network connection and UoS charges. The US and Australia already have some experience, and several possible merchant lines are under active discussion in Europe.

Apart from the well-known arguments for allowing market forces to operate where possible, there are three quite specific arguments for allowing merchant transmission investment. First, vertically integrated utilities have poor incentives to invest in interconnector capacities, because they increase the competition facing their own generation markets. Second, regulatory uncertainty impedes investment by regulated private transmission companies. Gans & King (2003) quote a discussion of the Australian Productivity Commission on "regulation holidays" for *risky* new significant investment. The argument is that a regulator cannot credibly commit to "allow high profits" if *ex post* the state of the world turns out to be good, but will happily allow the company to suffer any losses.[18] Hence, given uncertainty, the *ex ante* expected rate of return will be lowered, depressing investment. In the discussion in Australia, a commitment to refrain from regulation for a predetermined number of years is seen as a possible way out, opening the option of merchant investment.

The third argument is a public-choice argument. If permission to build a line is required on the two ends of an interconnector, the authorities would each need to be convinced that the extra charges need to finance a regulated investment can be justified on cost-benefit criteria as benefiting their jurisdiction. Merchant transmission investment mitigates this problem because it does not require such a cost-benefit test (although regulators may still deny interconnection if they believe it reduces social welfare in their own jurisdiction). The key argument is that the requirement of economic approval allows other goals to enter the discussion or can easily be abused.

It is important to distinguish carefully between the developments in the US on the one hand and Europe and Australia on the other. The nodal pricing approach in the US allows a

[17] Although many have been criticised and may not be politically feasible.
[18] A similar problem arises in the US where investment must be "prudent" and "used and useful" to qualify for receiving regulated revenues, but in that case it only earns the normal rate of return.

more refined merchant system with incremental Financial Transmission Rights, FTRs[19] than the zonal approaches in Europe and Australia. A properly chosen set of point-to-point incremental FTRs can internalize network effects caused by the loop flows of the new investment (cf. Hogan, 2003 and Kristiansen & Rosellón, 2003). However, such schemes create considerable complexity in their implementation, because they require the allocation of residual rights created by the merchant investment. This requires several years of multi-round auctions to identify the demand for transmission rights in the absence of merchant investment. Only then can the merchant investment projects receive transmission rights created by their addition to capacity. Moreover, incremental FTRs require centralized allocation of transmission rights and are against the spirit of decentralized market-driven decisions (Joskow & Tirole, 2003).

Without a well-defined nodal pricing scheme, such a refined point-to-point incremental FTR scheme is not possible. This suggests restricting merchant transmission investment in Europe and Australia to high voltage DC interconnection between different systems (and to leave other types of investment to the regulated Transmission Owner). The market-based revenues in Europe and Australia are derived from the link-based differences in (spot) prices between the two ends of the line. Merchant transmission investment in Europe and Australia should therefore be restricted to network expansion (interconnection) whereas in the US merchants are also allowed to invest in network deepening projects to some extent if they wish so (c.f. Brunekreeft, 2003).

Back-of-the-envelope profitability calculations suggest that merchant transmission investment is unlikely to be adequate by itself, because the risk premium will be high. At best, we are more likely to see a mix of regulated and unregulated lines being developed. Hogan (2003) emphasizes the danger of the slippery slope: i.e. the regulatory alternative (which gives access to regulated and hence secure revenues) tends to crowd out the unregulated alternative. To avoid this Hogan argues for the following sharp distinction: socially beneficial but commercially unprofitable projects (large when compared to the relevant market) qualify for the regulated option while everything else should be left to the market. Littlechild (2004) expresses the same concern following his analysis of a recent application of the regulatory test in Australia.

Joskow & Tirole (2003) and Joskow (2003, p. 54 ff.) are quite sceptical about market-based transmission investment and list a number of inefficiencies. These include market failures arising from investment lumpiness, market power in generation, improperly defined (non-contingent) property rights, and network effects. Joskow & Tirole conclude that merchant line investments are likely to be inefficient if a split in ownership (TO) and operation (SO) of transmission lines results in governance problems and incentive asymmetries with regard to dispatch and maintenance.

Two further problems follow from the perception that the European market design is still evolving. First, merchant investments may lock in the current situation (decentralized power exchanges) and may make changes to future regulatory intervention difficult. This is a

[19] Sometimes called TCCs. They can be used to define so-called Auction Revenue Rights (ARRs).

serious concern given the ongoing discussion about optimal pricing schemes and the expected changes of network and generation technology in response to CO_2 reduction objectives. If it is proposed to change the system of pricing (e.g. from a postage stamp to nodal pricing), then merchant investors might object and delay desirable changes. Perhaps the right of conversion to a regulated status as in Australia can relieve the problem, although not without some moral hazard. Second, without well-defined reference prices at both ends of the new interconnection, FTRs are not feasible. Changes to the definition and/or scope of the reference prices might either create uncertainty about the *ex ante* value of such FTRs, or again provoke opposition to desirable changes, e.g. towards a more refined nodal system, or within a market coupling approach a move towards endogenous zoning.

In Australia, investors have the option to be unregulated and rely solely on the price differentials between the two nodes or rely on regulated revenues, which partly consist of regulated connection charges. In order to qualify for regulated revenues the investment has to pass the "regulatory test". The test is passed if the investment has the highest net present value of market benefit with regard to possible alternatives. The main advantage of an explicit test is that obvious detrimental investments can be checked. The main disadvantage is that it inevitably introduces an arbitrary and bureaucratic element in an otherwise market-driven environment. Littlechild (2004) is critical of a recent application of the regulatory test and subsequent approval of a regulated project in the so-called SNI case in Australia. In a recent review, the Australian Competition and Consumer Commission raised interesting questions concerning the regulatory test (ACCC, 2003). One question concerned the alternatives to be examined. First, were new power plants alternatives to new lines? Second, the test included modeled projects, "likely to be commissioned". This seems reasonable but opens up gaming possibilities: firms can "model" fake projects.

The ACCC also questioned the measurement of social costs and benefits. First, new lines will in general have an effect on the competitiveness of the generation market on both sides of the line. A large part of the impact will be transfers from generators to consumers or vice versa (depending on the direction of flows), and the change in deadweight loss arising from more intense competition is likely to be very small (as demand elasticities are so small). There is a feeling that only to count the direct deadweight loss reduction understates the competition effects that should be taken into account in the social cost benefit analysis. For example, a more competitive market is likely to induce less "excess entry" and require less regulatory intervention (which, as California demonstrates, can be hugely costly). The latter argument is partly circular if it is assumed that the regulator also works with social cost benefit criteria. Alternatively, weighing consumer interests more heavily than shareholders' reinforced the claim that an unweighted deadweight loss analysis underestimates the competition effect. The question of how to measure the effect on and value of increased competitiveness remains open. Second, new lines will have network effects. For instance, a new line can increase the reliability of the network, but might also require a network upgrade. The issue relates strongly to the discussion on deep connection charges above. Defining the project to be evaluated then becomes somewhat arbitrary.

The jury is still out on whether a regulatory test is workable at all and if not, would a tender for constructing and operating the (unregulated) line be a feasible alternative? It will be interesting to compare Australian and American (e.g. PJM) experiences.

Unregulated merchant transmission investment raises a set of regulatory questions (cf. Brunekreeft in this issue). The answers to these question depend strongly on whether the institutional background is US-nodal-like or Europe/Australia-zonal-like. Institutionally the difference is reflected in the light-handed regulatory framework in Australia (called *safe harbors*) as opposed to the more heavy-handed situation in the US with *open season auctions,* while the regulatory discussion in Europe is ongoing. The issues are reflected in for instance the discussion around the merchant project BritNed which aims to interconnect the UK and the Netherlands, in which case the regulators should set up a regulatory framework.

The first question to address is whether there should be ownership restrictions on who is entitled to build and operate the line. A regulated private TSO owning an unregulated (and connected) merchant line may have incentives to distort dispatch and other investments in the network. Should there be a restriction on ownership of the line by dominant generators? The Australian safe harbors prescribe a limit of 35% of control of generation capacity on either side of the line. As a rule, competitiveness on the generation import market will be increased even if the dominant generator owns the line. The argument then is that the competitiveness could have been increased even more if a third party owned the line. It is disputable whether this justifies an intervention. One is inclined to leave the issue to ad-hoc control by competition authorities. However, in many cases competition laws seems poorly equipped to avoid new investment by a dominant generator and *ex post* control of abuse of market power is a hazardous task as compared to an *ex ante* control on market structure.

The second question to address is the design of the access regime. Should the line owner be allowed to participate in using the line or should this be fully separated? Should the line owner be free to determine who will be entitled to use the line, or should a non-discriminatory open-access regime apply? Again, the choice is between light- and heavy-handed regulation, which depends strongly on the underlying institutional framework. Applying insights from the theory on vertical relations, a regulated access regime seems redundant if the line revenues are unregulated. The unregulated line owner will normally have an incentive to maintain competitive pressure among line users. Application of the essential-facilities doctrine in competition law suffices to tackle abuse should it occur. The situation changes if the line revenues are regulated, either explicitly or implicitly by a tender for capacity. In those cases, an access regime is necessary to avoid leverage of market power.

Third, should a part of the capacity be reserved for the short-run spot market instead of selling all capacity in long-term contracts? To share the risks, merchants are likely to sell off long-term contracts, possibly even before making the sunk investment. This may impede competition if at a later stage new firms would be excluded from using the line, although there is likely to be a short-term capacity market to allow contract holders to balance their position.

Fourth, should there be a provision against capacity withholding? Attractive though a must-offer (or, use-it-or-lose-it) rule may be *after* the investment, *prior* to the investment such a rule adversely affects the investment decision, which may offset any gains of applying the rule. Brunekreeft and Newbery (2004) identify situations, in particular demand uncertainty and demand growth, when the adverse effect on investment (*ex ante*) offsets the positive effects on the line use (*ex post*). In contrast if the incentive to withhold capacity is the result of pre-emptive investment, a must-offer provision is more likely to enhance welfare the lower fixed investment costs. Overall, especially with high fixed investment costs, it is likely that application of a must-offer provision in case of unregulated merchant investment decreases social welfare.

5. Incentivizing system operators: Regulatory frameworks and experience

The England and Wales electricity market provides an excellent example of the benefits of exposing the TSO (NGC) to an incentive scheme to reduce constraint costs. Constraint costs fell from over £200 million per year in the early years of the Pool to about £25 million per year after the incentives were introduced. Incentives to reduce congestion costs are particularly important in England and Wales, because the entire region is treated as a single price area. Hence all transmission constraints have to be resolved by the TSO using bilateral contracts. Such a setting allows generators with market power to play the inc-dec game (i.e. bidding increments or decrements of generation).[20] In the UK transmission constraints are rare and mainly caused during maintenance periods. Incentivising the TSO to reduce these periods, reduce bottlenecks, and engage in clever contracting to limit the extent to which generators can play the inc-dec game is crucial.

The success of the scheme hinges on the ability of the regulator to expose the TSO to incentive schemes, which is only feasible if the TSO has an adequate asset base, typically the network, to bear the financial risks involved. The English approach will also require careful adaptation before such a scheme can be applied to continental TSOs with strong interconnections to neighboring countries. The potential problems can be illustrated by two observations. Glachant & Pignon (this issue) argue that the combination of inter-regional constraints and redispatch to resolve constraints within regions can create incentives for the TSO to distort dispatch from the optimal choice. If less transmission capacity is declared available for inter-regional capacity than in many network configurations (e.g. in Sweden) some of that capacity can reduce constraints within a region. If a TSO is exposed to some of the costs for re-dispatching generation to resolve constraints within regions, then it benefits from understating available inter-regional transmission capacity.

Boucher & Smeers (2003) and Patton (2002) point out another cross-border incentive problem. They argue that a TSO in one area may have poor incentives to inform a neighboring TSO, or worse, the TSO may even have an incentive to resolve its own problems

[20] For example in California, generators first nominated excessive energy flows so that the system operator had to contract, typically with the same generators, to re-dispatch in order to avoid violations of transmission constraints. This profitable game is possible because every one is granted unlimited access even if transmission rights are scarce. The issue is avoided if scarce transmission capacity is either allocated as physical transmission contracts or charged through a nodal pricing mechanism.

at the expense of the other TSO ("shifting congestion to the neighbors"). These arguments are particularly important for the European market. It seems necessary that any incentive scheme within a country takes explicit account of cross-border effects (or is combined with a suitable cross-border incentive).

The theory of the design of a comprehensive and consistent regulatory framework for TSOs is not well developed, with the exceptions of Vogelsang (2001), Leautier (2000 and 2001), Yoon & Ilic (2001). As will be explained further below, NGC 's incentive mechanism, aiming at reducing balancing costs and transmission losses, is a practical attempt of an incentive scheme that works reasonably well for improving the efficiency of use of the existing assets (transmission and generation). The scheme works by setting a target level, with revenue deviations from the target are partly borne by NGC. Caps and collars secure that the risks are manageable. Transmission losses can be reduced by transmission expansion and upgrades, and so short-run incentive may have favorable long-run effects, but the proper design of an incentive scheme that would deliver efficient transmission investments is a considerably more challenging task.

Vogelsang (2001) develops a two-part price cap where the variable part reflects congestion charges and the fixed part the fixed network costs. The integrated TSO balances the two elements: grid expansion requires investment but lowers congestion and vice versa. The integrated TSO will have an incentive to expand capacity if marginal congestion costs exceed the costs of expansion. In contrast, if TO and SO are split, it seems natural that the SO would be responsible for network operation and congestion and the TO for network expansion and maintenance. With a TO/SO split, a grid expansion lowers congestion, which are foregone revenues for the SO. Whether or not the TO has sufficient incentives to invest depends on the cap. In any case, balancing between congestion costs and grid expansion in the case of a TO/SO split seems to be coincidental. The argument is more important for the split TO/SO cases in the US than the integrated cases in Europe.

Note how the regulatory scheme for NGC can be seen as two different baskets: one for transmission network use-of-system and one for balancing services use-of-system. Following the Vogelsang approach and much in line with most of the price-cap literature one might suggest integrating both parts into one basket and have the firm decide on a proper balance. However, in the presence of vertically integrated TSOs such an approach seems inappropriate. Different tariff structures (e.g. capacity versus energy based) bias the choice of generation technologies. Therefore the vertically integrated company will choose a tariff structure to support technologies dominated by companies affiliated with the network operator, technologies consistent with the current operational paradigm or technologies that might allow better reaction to inherently incomplete operational or investment incentives. In distribution tariffs this has been a frequent complaint of distributed generation.

Another way in which the TSO can be motivated to increase the availability of capacity is the use Financial Transmission Rights (FTRs) as applied in the US. The idea would be to issue FTRs corresponding to the capacity that the TSO promises to make available. Much on the same line as Contracts for Differences, the TSO would be exposed to the additional congestion costs if it delivers less transmission capacity on which he will receive congestion payments than contracted in the FTRs for which he has to pay the nodal price difference. The same mechanism ensures that the TSO benefits from providing additional capacity. Such a scheme has been proposed for the British gas National Transmission System (also now owned by NGT, the owner of NGC). The concept might be more difficult in a meshed network, because several sets of FTRs are mutually feasible and it is difficult to determine which of these sets to issue. The success or failure of such a scheme is likely to depend on the fine details.

Alternatively, in a setting with a TO/SO split, the SO function might be auctioned off, which may introduce some competitive pressure (cf. Demsetz, 1968). Demsetz highlights the virtues of franchise bidding for natural monopolies as a substitute to regulation. In response, Williamson (1976) is more sceptical, but his critique highlights the problems of selling sunk assets (with a longer life than the auctioned period). In particular, if the sale price cannot credibly be arranged in advance, a subsequent hold-up may lead to underinvestment (Williamson, 1976, p. 87). Since the SO does not own the assets, the Williamson (1976) critique does not apply. The main assets are the idiosyncratic skills of staff in the SO-department, which may not give rise to such problems. An SO/TO split does, however, have the disadvantage that the asset-poor SO cannot be exposed to strong incentive schemes because it may go bankrupt if revenues become negative if targets are not achieved. However, if the cap-and-collar scheme as for NGC works sufficiently well it will reduce this risk. Nevertheless, an asset-poor SO can only be exposed to a fraction of constraint costs and therefore will not fully internalize these costs in his decision process.

The next subsection discuss a sample of issues of the regulatory framework and experiences for the TSO NGC in the UK. The situation in the US is discussed in the paper by Joskow in this issue.

5.1 UK: National Grid Company[21] planning of new investment and charges

The foreword of NGC 's 2003 7-year statement states that the "broad approach to planning the future development of the transmission system is to maximize the utilization of the existing system" and to "continue to maintain or improve its reliability and availability". In practical terms, investment in the grid concerns not so much new lines, but instead incremental reinforcements and upgrades. There are two main determinants of investment: supply and demand growth,[22] and the need to meet the (regulated) quality standards.[23] Retaining quality requires maintenance and upgrading of the grid. NGC works with three criteria: availability, reliability (MWs loss of supply), and quality of service, measured by deviations from target voltage and frequency levels.

---

[21] The NGC documents can be downloaded from NGC's websites: http://www.nationalgrid.com/uk/. It should be noted stressed that several issues are currently under revision.

[22] Cf. Security and quality of supply standard (setting out connection requirements) and NGC's 7-year statement.

[23] Cf. "Security and quality of supply standard, Report to DG of Ofgem under special condition AA2" of the Transmission Licence and the Grid Code.

To finance these activities, NGC has two classes of network charges: connection charges for the grid connection, and Use-of-System charges. The latter can be subdivided into transmission network UoS charges (TNUoS) and Balancing System UoS charges (BSUoS). Note that NGC 's use of the term connection charges is narrow as compared to the use elsewhere in this paper, where connection charges includes TNUoS. Connection charges are only for costs directly associated with connection of a user at an entry/exit point of the grid. NGC 's (or rather, the regulator, Ofgem's) approach is relatively shallow and relate mainly to the costs of assets which can be attributed directly to a user or a subset of users. Substation assets and generation-only-spurs fall under connection assets (as opposed to TNUoS), and off-shore connections from wind farms would also be borne by the generator. All else is considered TNUoS. Significant new connections will normally have an impact elsewhere in the network and may require upgrades. Shallow charging does not attribute these costs to the new connection.

TNUoS charges cover the cost of installing, operating and maintaining the transmission system. The charges are regulated by a price-cap mechanism. The system relies on the marginal expansion costs for an optimized network (using the Investment Cost Related Pricing (ICRP) transport model). The marginal costs of required investment are calculated for marginal generation and load increases at each node. The nodes are attributed to the 15 generator zones and 12 demand zones using weighted averages. An additional term, called the Security & Residual Tariff, which is not locationally differentiated, serves to ensure adequate revenues. The key elements are that the so-defined charges are differentiated in two respects. First, the generator-load split was set at approximately 27:73 in 2003. Second, under zonal TNUoS pricing, generators in the north pay high G charges while load face low L charges, and conversely in the south, reflecting the excess number of generation capacity in the north.

Finally, the Balancing Service UoS charges cover losses and balancing services, where NGC is incentivised to reduce their costs. If NGC beats the target, it can retain 60% of the savings, whereas if the balancing costs are higher than the target level it will have to share 50% of the additional costs up to a cap of plus £60 million and a collar of minus £45 million.

6. Conclusions and future research agenda

LMPs are unlikely to recover fixed network costs and additional charges are required. Deep connection charges could cover some of these additional costs if they can be properly identified and are mainly required to compensate for the difficulty of reflecting all the other attributes of transmission service (particularly reliability) in the nodal prices. Provided these signals are accurate, the combination of LMPs and deep connection charges should provide efficient generation (and load) locational investment signals. The alternative of zonal charges may give reasonable investment signals but is less well suited to efficient network use and congestion management.

Nodal spot pricing is considered to be the most efficient congestion management system. It is therefore disappointing if understandable that Europe retains a zonal (often country-wide) approach and (initially at least) seems to prefer co-ordinated explicit auctions to the intermediate state of market coupling. At least the policy advisers seem agreed to press for as close an approximation to nodal pricing as politically feasible and to aim for a transitional system that allows the final step towards nodal pricing to avoid a lock-in to an inappropriate intermediate stage. In the meantime, market coupling may be preferred for network zones with identifiable single price areas. Provided stepwise increases of the market coupling areas offer improvements to both participating areas, this gradual approach seems promising, as might be the case in Benelux.

Financing new transmission investment in interconnectors remains problematic. First, it is not yet clear whether merchant transmission investment is a reliable option, and if so, whether it would be preferred to regulated investment. Any assessment depends quite strongly on specific cases and on the institutional framework. Notably, the nodal approach in the US allows a refined payment scheme with incremental point-to-point FTRs. The zonal approaches in Europe and Australia have to rely on link-based spot price differences. This suggests that the role of merchant investors could be greater in the US while it should be restricted to DC interconnectors in Europe.

Second, where merchant transmission investment takes place it raises new regulatory questions. The regulatory approaches in the US, Australia and Europe are strikingly different. Whereas the regulatory framework in the US is relatively stringent, in Australia it is light handed, following the link-based zonal approach. In Europe the issue still has to be settled, but it seems natural to follow the Australian approach as long as European congestion management is zonal.

The workshop identified a range of topics meriting future research, and the following list is not exhaustive. The welfare benefits of reducing market power by transmission investment appear modest if measured in the conventional deadweight loss way as demand is very inelastic, but this may seriously underestimate the social benefits. Some concern was expressed that a shortage of transmission capacity could panic regulators or politicians in taking very expensive actions during crisis periods of high prices, as happened in California. Increasing the resilience of the system by possibly some over-investment might reduce the risk of expensive and inappropriate responses to high prices. If the benefits from increased competition are significant, then it becomes important to compare alternative ways to achieve the same result, at perhaps lower cost than “excess” transmission investment.

Many of the proposed methods of creating strong price signals also create prices risks that need hedging contracts such as FTRs. In addition, FTRs can play an important role in ensuring that investments are efficient. Such markets are new and need more study, perhaps supplemented by experimental economics tests of FTR auctions. For pricing to achieve its full effect, the demand side should be involved, and there are outstanding questions on the costs and benefits of real time pricing and demand price responsiveness. In some cases this should be almost costless, e.g. for large customers with interval metering, but the benefits need further documentation (see Borenstein and Holland, 2004).

Incentivizing the TSOs remains an underdeveloped area both in theory and practice. The main problem to be avoided is that a TSO benefits from creating congestion. An incentive scheme should reward the TSO for reducing congestion and balancing costs by carefully balancing daily operation and network investment. The use of redefined FTRs may be helpful here; if successful, this would yet give another reason for moving towards a nodal pricing scheme. There is also a clear need for more research on the benefits of incentive-based regulation of (asset owning) TSOs, compared to alternative models (such as not-for-profit ISOs for which such incentives may have to be considerably weakened). The relative merits of state-owned and private transmission companies also remain largely untested. State ownership (as in many EU countries) may make it easier to move towards an integrated regional system (as in Nordpool), although the private NGT in Britain has experienced the considerable change brought about by ending the Pool, and is currently involved in integrating the English and Scottish systems. More problematic are private vertically integrated companies such as RWE, E.On and Electrabel, where private ownership requires consensual restructuring, often with substantial compensation.

The question of how best to judge the desirability of significant new transmission investment in interconnected grids under different TSOs remains empirically and institutionally unresolved. Further progress will require TSOs to pool transmission and flow data and commission the necessary modeling work, and there are some encouraging signs that ETSO is moving in this direction. In this connection the whole question of assessing and then charging for off-shore wind, and resolving the deep versus shallow connection charge question is becoming urgent. Finally, we need more quantitative assessments of the difference between implicit and explicit markets (market coupling vs. auctions) for Europe.

References

ACCC, 2003, *Discussion paper: Review of the regulatory test*, 5 February 2003, Australian Competition & Consumer Commission.

Baldick, R. & Kahn, E., 1993, Transmission planning issues in a competitive economic environment, *IEEE Transactions on Power Systems*, Vol. 8, No. 4, pp. 1497-1503.

Baughman, M.L., Siddiqi S.N. & Zarnikau, J.W., 1997, Advanced Pricing in Electrical Systems: Part I : Theory and Part II : Implications, IEEE Transactions on Power Systems, vol.12, n°1.

Bohn, R.E., Caramanis, M.C. & Schweppe, F.C., 1984, 'Optimal pricing in electrical networks over space and time', *Rand Journal of Economics*, Vol. 15, No. 3, pp. 360-376.

Borenstein, S., & Holland, S., 2004, 'On the Efficiency of Competitive Electricity Markets With Time-Invariant Retail Prices', *Rand Journal of Economics*, forthcoming

Boucher , J. & Smeers, Y., 2003, 'The European regulation on cross border trade: Can one do without a standard market design?', *mimeo*, March 2003, University of Louvain, Belgium.

Boucher , J. & Smeers, Y., this issue.

Brunekreeft, G., 2003, 'Market-based investment in electricity transmission networks: Controllable flow', CMI EP 29, University of Cambridge, *forthcoming Utilities Policy*.

Brunekreeft, G., this issue (also: 2004, 'Regulatory Issues in Merchant Transmission Investment', CMI EP 38, University of Cambridge).

Brunekreeft, G. & Newbery, D.M., 2004, ‘Should merchant transmission investment be subject to a must-offer provision?’, *mimeo*, DAE, University of Cambridge.

Bushnell, J.B. & Stoft, S.E., 1997, ‘Improving private incentives for electric grid investment’, *Resource and Energy Economics*, Vol. 19, pp. 85-108.

Demsetz, H., 1976, ‘Why regulate utilities?’, *Journal of Law and Economics*, Vol. 11, pp. 55-66.

DTI, 2002, *Future Offshore*, London: Department of Trade and Industry

Ehrenmann A., & Neuhoff K., 2003, ‘A comparison of electricty market designs in networks’, CMI EP 31, University of Cambridge.

ETSO, 2001, “Position paper on congestion management”, April 2001.

European Commission, 2003, *Strategy Paper: Medium term vision for the internal electricity market*, June 23, 2003, DG Tren, Brussels.

Gans, J. & King, S., 2003, ‘Access holidays: The panacea for network infrastructure investment?’, *mimeo*, March 2003, University of Melbourne.

Gilbert, R., Neuhoff, K. and Newbery, D., 2004, ‘Allocating transmission to mitigate market power in electricity networks’, *RAND Journal*, Winter 2004 vol 35, no.4l.

Glachant, J.-M. & Pignon, V., this issue (also: 2003, ‘Nordic electricity congestion's arrangement as a model for Europe: Physical constraints or operators' opportunity’, *Working paper*, DAE WP 0313 /CMI WP 20, University of Cambridge / MIT).

Haubrich H.J., Zimmer C., von Sengbusch K., Li F., Fritz W., Kopp S., Schulz W., Muesgens F., Peek M. (2001) Analysis of Electricity Network Capacities and Identification of Congestion, Final Report for the European Commission, Aachen.

Harvey, S. M. and Hogan, W.W., 2000,. ‘Nodal and zonal congestion management and the exercise of market power’, Harvard University, http://ksghome.harvard.edu/ ~.whogan.cbg.ksg/.

Hogan, W.W., 1992, ‘Contract networks for electric power transmission’, *Journal of Regulatory Economics*, Vol. 4, pp. 211-242.

Hogan, W.W., 2003, ‘Transmission market design’, presented at Electricity Deregulation: Where to from here? Conference at Texas A&M University, April 4th, 2003.

Joskow, P. and J. Tirole, J., 2000,. ‘Transmission rights and market power on electric power networks’, *RAND Journal of Economics* 31(3), 450-487.

Joskow, P. & Tirole, J., 2003, ‘Merchant transmission investment’, *mimeo*, May 2003, MIT, Cambridge, Mass., forthcoming *Journal of Industrial Economics*.

Joskow, P. & Tirole, J., 2004, ‘Reliability and competitive electricity markets’, CMI EP 53 University of Cambridge

Joskow, P., 2003, ‘The difficult transition to competitive electricity markets in the U.S.’, *mimeo*, May 2003, MIT, Boston.

Joskow, P., this issue.

Kattuman, P. A., Green, R.J. and Bialek, J.W., 2004, ‘Allocating electricity transmission costs through tracing: a game-theoretic rationale’, *Operations Research Letters*, vol. 32, no. 2, pp 114 – 120.

Kay, J. A. and Keen, M.J., 1983, ‘How Should Commodities Be Taxed? Market Structure, Product Heterogeneity and the Optimal Structure of Commodity Taxes’, *European Economic Review*; 23(3), December, 339-58.

Kristiansen, T. & Rosellón, J., 2003, ‘A merchant mechanism for electricity transmission expansion’, *mimeo*, HEPG, Harvard University.

Leautier, T.-O., 2000, ‘Regulation of an electric power transmission company’, *Energy Journal*, Vol. 21, No. 4, pp. 61-92.

Leautier, T.-O., 2001, ‘Transmission constraints and imperfect markets for power’, *Journal of Regulatory Economics*, Vol. 19, No. 1, pp. 27-54.

Liu, L., Zobian, A., 2002, ‘The Importance of Marginal Loss Pricing in an RTO Environment’, *Electricity Journal*, Vol. 15, No. 8, pp. 40-45.
Littlechild, S., 2004, ‘Regulated and merchant interconnectors in Australia: SNI and Murraylink revisited’, CMI EP 37, University of Cambridge.
Neuhoff, K., 2002, 'Optimal congestion treatment for bilateral electricity trading', Working Paper, CMI EP 05, University of Cambridge.
Neuhoff, K., 2003, ‘Integrating transmission and energy markets mitigates market power’, Working Paper, CMI EP 17, University of Cambridge.
Neuhoff, K. and Newbery, D.M., this issue.
Newbery, D.M. and McDaniel, T., 2003, ‘Auctions and Trading In Energy Markets - An Economic Analysis’, Ch 10, pp195-234 in *CRI Regulatory Review2002/ 2003* P. Vass (ed.) Bath: Centre for the Study of Regulated Industries.
O’Neill, R.P., Helman, U., Hobbs, B.F., Stewart, W.R. & Rothkopf, M.H., 2003, ‘A joint energy and transmission rights auction: Proposal and properties’, *mimeo*.
Patton, D.B., 2002, ‘Comments of the MISO independent market monitor on the RTO confguration’, unpublishled, downloaded from www.ksg.harvard.edu.
Pérez-Arriaga, I.J., Rubio, F.J., Puerta, J.F., Arceluz, J. & Marin, J., 1995, ‘Marginal pricing of transmission services: An analysis of cost recovery’, *IEEE Transactions on Power Systems*, Vol. 10, No. 1, pp. 546-553.
Pérez-Arriaga, I.J. & Olmos, L., this issue.
Pérez-Arriaga, I.J, 2003, ‘The EU Internal Electricity Market: Current State of Debate’ presentation at the CMI Transmission Workshop, 18/19 July 2003, Cambridge.
PJM MMU, 2003, *State of the market 2002*, PJM Market Monitoring Unit, March 5, 2003.
Read, E.G. & Sell, D.P.M., 1989, ‘Pricing and operation of transmission services: Short run aspects’, *in: Turner, A., (ed.), 1989, Principles for pricing electricity transmission*, Transpower New Zealand.
Read, E.G., 2002, ‘Financial transmission rights for New Zealand: Issues and alternatives’, *Study prepared for New Zealand Ministry of Economic Development*, Draft 2, April 2002.
Stoft, S., 2002, *Power system economics: Designing markets for electricity*,Wiley-Interscience.
Vazquez, C., Olmos, L. & Perez-Arriaga, I.J., 2002, ‘On the selection of the slack bus in mechanisms for transmission network cost allocation that are based on network utilization’, *ITT Working Paper*, IIT-01-108A, IIT, Universidad Pontificia Comillas, Madrid.
Vogelsang, I., 2001, ‘Price regulation for independent transmission companies’, *Journal of Regulatory Economics*, Vol. 20, No. 2, pp. 141-165.
Williamson, O.E., 1976, ‘Franchise bidding for natural monopolies - in general and with respect to CATV’, *Bell Journal of Economics*, Vol. 7, N0. 1, pp. 73-104.
Woolf, F., 2003, *Global transmission expansion: Recipes for success*, CMS Cameron McKenna.
Yoon, Y.T. & Ilic, M.D., 2001, ‘Price-cap regulation for transmission: Objectives and tariffs’, *IEEE*, 2001, pp. 1052-1057.

Appendix: A neutrality result for allocating grid charges between G and L

The regulator will determine how much revenue the regulated grid can collect, and thus the shortfall after receiving any LMP revenue. This revenue will have to be collected from G and L in some proportion. The balance between G and L clearly does not matter in a perfectly competitive and integrated or isolated market (where the division is the same for all agents), provided that the difference between prices at different nodes is preserved (Kattuman, Green and Bialek, 2004; Vázquez, Olmos and Pérez-Arriaga, 2002).

The natural next question is whether this neutrality result carries over to the case of imperfectly competitive generators. Fortunately the answer is yes. To see this, consider a general model of imperfect competition, and consider generator i competing with other generators j, where the reference price is p and the transmission charge to deliver to demand or load is t. Let the fraction paid by G be α, and by L, 1-α. Suppose total demand is $Q = \Sigma q_j$, and the inverse demand schedule is $H(Q) = p + (1-\alpha)t$, which depends on the delivered price including that part of the transmission charge paid by L. The generator's problem is to choose q_i to maximise profits, where the cost of generator i producing q_i is $C_i(q_i)$, given the conjectural response of other generators to his choice:

$$\text{Max } \pi_i = (p - \alpha t)\, q_i - C_i(q_i).$$

The first order conditions w.r.t. q_i are

$$p - \alpha t - C_i'(q_i) + q_i H'(Q)\left(1 + \sum_{j \neq i} \frac{dq_j}{dq_i}\right) = 0, \tag{1}$$

where the terms dq_j / dq_i are the conjectured responses of other generators to increased output q_i. Cournot conjectures would make these all zero, Bertrand would make the sum -1 in the case of unconstrained constant cost output, and so on. Now the first term in (1) can be rearranged from the inverse demand schedule $H(Q) = p + (1-\alpha)t$:

$$p - \alpha t = H(Q) - t,$$

which is independent of the sharing parameter α. It follows that the choices of output q_i determined by (1) are also independent of the shares of total cost borne by G and L, provided that the mechanism for collecting the charge is independent of the level of (nodal) prices. It is easy to see that the neutrality result would not hold if $\theta \equiv t/p$, and producers receive $p(1-\alpha\theta)$ and consumers pay $p(1+(1-\alpha)\theta)$, for in that case α would not be eliminated in the first order conditions.

This result parallels the neutrality result derived by Kay and Keen (1983) for the incidence of specific taxes, which is the same whether paid by producer or consumer (in a rational and frictionless, but not necessarily perfectly competitive, world).

The choice of the "slack bus" in the marginal participation rule

Vázquez et. al. (2002) point out that the choice of the "slack bus" effectively determines from which node the withdrawal (or injection) of 1 MW comes, and hence strongly influences the resulting flows. As a consequence the principle of the marginal unit can lead to unreasonable results if the marginal value in a line is significant although the "real" flow is negligible (Vázquez et. al., 2002). This might appear to make the marginal participation rule arbitrary, but Vázquez et. al. (2002) demonstrate that irrespective of the choice of slack bus, if both generation and load are exposed to the same locational component, then the net total payment for a balanced transmission is not influenced by the choice of the slack bus as follows.

Let $A_{i,j}$, represent the marginal participation sensitivities (by how much flow through line j increases in response to an injection of 1 MW at node I, given the choice of the slack bus where the withdrawal will be made), and suppose that the costs of each line j to be recovered are C_j. Total injection at i is g_i and withdrawal is d_i , so net injection is (g_i - d_i). Suppose the original choice of slack bus gives rise to a per MW charge of T_i at node i. They also show that if the choice of the slack bus is changed then the effect is to add a fixed term X_j to each line's sensitivity factor which is independent of the node. Thus the general case where the slack bus (and hence the set of X_j 's) has yet to be determined can be found as follows. The resulting per MW charge for net injections at i is T'_i ($\boldsymbol{X}$), where

$$T_i'(X)\cdot(g_i-d_i)=\sum_j\left(C_j\cdot\frac{(A_{i,j}+X_j)(g_i-d_i)}{\sum_k(A_{k,j}+X_j)(g_k-d_k)}\right).$$

Ignoring losses, if the system is in balance, then $\Sigma_k(g_k - d_k) = 0$ and hence $\Sigma_k X_j (g_k - d_k) = 0$. Therefore the total charge to the agents at node i is

$$\sum_j\left(C_j\cdot\frac{A_{i,j}(g_i-d_i)}{\sum_k A_{k,j}(g_k-d_k)}\right)+(g_i-d_i)\cdot\sum_j\left(\frac{X_jC_j}{\sum_k A_{k,j}(g_k-d_k)}\right)=T_i+(g_i-d_i)\cdot t\,.$$

The effect of changing the slack bus is to change *all* nodal charges by an additive element t per MW. We can immediately deduce from the earlier argument on nodal prices with imperfect competition that *in a balanced system*, the choice of the slack bus (which is equivalent to choosing the proportion of the total cost allocated to G and L) will make no difference to the nodal prices, *even in the presence of imperfect competition*.