High level principles for guiding GB transmission charging and some of the practical problems of transition to an enduring regime

David Newbery
Electricity Policy Research Group
University of Cambridge
Revised 22 April 2011[1]

Executive summary

Ofgem launched Project TransmiT to review GB transmission charging and associated connection arrangements to seek advice on any changes needed to efficiently support the transition to a low-carbon energy sector. This report was commissioned by Ofgem to provide guiding principles for reforming the transmission charging arrangements in GB.[2] The urgency in the current review arises from the need to ensure that the large volume of new low-carbon generation seeking connection in the next decade, much of it intermittent wind, should be connected in a timely manner and in the right location. The Government is also consulting on Electricity Market Reform (EMR) and plans to publish its conclusions in a White Paper in Spring 2011, and clearly it would be helpful if Project TransmiT can be properly coordinated with EMR proposals.

The large increase in constraint costs caused by including Scotland in BETTA, which will likely be amplified by the recent "connect and manage" approach to bringing on new wind generation, makes a careful review of transmission charging particularly urgent. The existing grid and related charging regime were designed for an era in which new connections were primarily of base load fossil generation, guided by zonal annual Transmission Network Use of System (TNUoS) charges based on maximum generation, while the future will see large volumes of wind with a capacity factor closer to 25%, and a need for new flexible peaking plant, also with low capacity factors. Intermittency will lead to volatile prices and shifting local transmission constraints, which add to the pressures for reforming current transmission charging arrangements as well as the wholesale and balancing markets.

This paper argues that Locational Marginal Prices (LMPs) are the theoretically correct prices for transmission access, although the requirements of the EC Target Electricity Model may require LMPs within relatively unconstrained areas to be averaged to give zonal prices. Such zones could be created by market splitting on the Nordic model. In this model the System Operator (SO) attempts to clear the entire

---

[1] I am grateful for all the comments provided at and after the Ofgem seminar on 4th March.

[2] The scope of the work requested is provided at Appendix E

market at a single price in each hour, but whenever transmission constraints prevent a feasible dispatch at this price, the SO computes the market clearing price in each zone with supplies into that zone restricted to ensure feasibility for that hour.

LMPs or zonal prices will probably need additional locational signals to reflect any shortfall from the full long-term locational pricing signals that are needed. These long-term full locational charges would form the natural strike price for the long-term Financial Transmission Rights to be issued by the Transmission System Operator (TSO), replacing the TNUoS charges on Generators. The entire shortfall in revenue would be levied on Load, to bring GB into line with our interconnecting neighbours when market coupling is required by the Target Electricity Model.

The main guiding principle behind this conclusion is that charges should encourage efficiency in the location, type and timing of new generation investment, promote efficiency in the dispatch of all generation, and promote efficient and timely transmission investment. Efficient generation investment requires working back from the efficient charging regime for dispatch to determine what additional adjustments are needed to guide investment decisions. Efficient decentralised dispatch is most simply provided by Locational Marginal Pricing (LMP), sometimes called nodal pricing, in which each generator (and load) faces local prices for power, Conceptually, LMPs are the market clearing price at each node that balance demand at that node with the supplies that can be delivered to that node, taking account of transmission constraints and losses. LMP is widely used in the US and elsewhere, has been demonstrated to work well and avoids many of the inefficiencies and problems that other charging regimes encounter in liberalized markets.

If congestion is confined to a few key boundaries (such as the Cheviot boundary between England and Scotland) the LMPs within each zone defined by these boundaries may be sufficiently similar to allow zonal pricing. Zonal pricing is the currently preferred option in the EC Target Electricity Model that GB will be required to adopt in the near future and certainly before 2015. Large price zones are more attractive for traders as more generators and consumers face a common price, but as the size of the zone increases so do the redispatch costs of congestion management, and the optimal size of the zone (which might be a single node) involves balancing these costs and benefits.

LMPs provide useful information to guide investment decisions for generation and transmission, but are unlikely to be sufficient by themselves, as lumpiness and economies of scale in transmission, uncertainties about future generation investments and demand growth, and the difficulty of decentralising charges for reliability and quality of service mean that transmission systems will aim to be over-built relative to a smoothly and optimally adjusted benchmark that would make LMPs a good guide to investment. These nodal charges will likely need to be supplemented by additional locational charges and certainly by uniform (non-locational) charges to make up any short-fall in revenue. Where transmission constraints are unimportant within an area, LMPs would be similar, and in such cases all nodes in that zone could be faced with the same zonal price.

The starting point for the locational correction element would be an estimate of the network investments that are justified to accept generation at any node, based on some view of future demand and investment decisions by location, type and amount. That does not mean that it would be sensible to build sufficient transmission to allow all generators to deliver peak output, as total systems costs will typically be reduced by constraining off generation at some locations by installing less transmission capacity. This forward transmission planning exercise would allow an estimate of the deep connection charges of connecting new generation. Deep connection charges collect the full consequential incremental costs caused by accepting generation at that node. Given the instability and unpredictability of LMPs, investors will need the assurance of long-term contracts and these would naturally take the form of Financial Transmission Rights or FTRs for the desired level of Transmission Entry Capacity (TEC). These would entitle the holder to receive the difference between some reference nodal or balancing point price and the LMP in each half-hour, for the payment of a better estimate of the correct locational TNUoS charge (less the revenue recovery charge which would be an addition to the locationally varying charges).

There are two reasons for moving from the present shallow charging philosophy to one based on deep connection charges. The first, which is perhaps not very important, is that when GB is coupled to the Central West Europe market, GB generators will be disadvantaged in daily trading over interconnectors unless the average variable generation (G) charge is set to zero. Deep connection charging then remains as one possible way of providing new investors with the right locational guidance (although the current form of locationally varying TNUoS charges that average out to zero would also work). The second more important reason is that under the EMR, all new low-carbon generation investment will be under long-term contracts that should cover these deep connection charges, and thereby insulate existing generators from changes in their local TNUoS charges. In short, everyone is effectively insulated from any adverse effects of connection charges, while they remain to guide the locations at which the contracting party is willing to support new generation and make explicit the cost of so doing. It is important to stress that now the Government has accepted the need for long-term contracts to support low-carbon electricity, there is no reason to bias transmission charging principles to over-come any financial barriers that impede such generation from connecting.

The revenue recovery charge should be set to minimize distortions to operating, exiting and consumption decisions. The present system of Triad charging appears to do that well for Load, but the annual G charges could encourage premature exit, and would be avoided by long-term mortgage-like instruments.

The TSO would collect revenue from charges paid by Load (L) and those paid by Generation (G) – i.e. the wedge between what the generators receive and load pays for power. Demand charges would primarily be determined by cost causation, which normally means peak-load pricing, as the transmission system is typically sized for peak loads. In practical terms the current system of Triad charging for Load (charging in the peak hours) minimizes the distortions caused by purely revenue raising surcharges. The balance between revenue charged to L and G is currently 73:27, but in a closed system the balance is immaterial as long as it remains constant, as it is the total wedge between the prices received net of transmission charges by G and that paid gross of charges by L that matters in long-run equilibrium. Changes in that balance or significant changes in the level of G charges would, however, likely create winners and/or losers among incumbent generation, while not affecting their location or operating decisions, and it may be desirable to continue with the existing system for incumbents (perhaps converted to long-term obligations) and only change the charges for new generators.

In an interconnected system (in GB's case with Ireland, France and the Netherlands) there are good reasons for choosing a common system for at least the variable grid charges and possibly for new connections, while it is not clear what the impact would be on incumbents (and should be carefully investigated to test for its materiality). In this case our neighbours have zero G charges, and if we do not match that for our variable G charges, their exporting generators would be at an advantage selling into GB. A decision to levy all transmission charges on Load does not rule out locationally differentiated charges to Generation – all that is required is that the *average* generation charge is zero, with some positive and others negative. Whether that implies a GB average G charge of zero or whether it suffices to have a zero average G charge in the zones that are interconnected is an issue that will need to be addressed (and will be complicated by the new Carbon Support Price that disadvantages GB electricity exports to the Continent). The present TNUoS charges have negative prices in some zones (only paid if the generation is available when required), and that could also be the case with the nodal pricing proposals put forward here. Note that LMPs aid in ensuring that merchant interconnectors make efficient decisions, as zonal or regional price differences can be a misleading guide to the efficient nodal price differences at each end of the interconnector. As mentioned above, deep connection charging amortized over time and collected in a contractually fixed way would be an alternative to locationally varying TNUoS charges and has some advantages.

These pricing principles imply that generators should not disconnect prematurely solely because of the short-fall charges, as any contribution they make is worth having. That is an additional reason for replacing the annual TNUoS charge that can be avoided by disconnection with a longer-term contractual liability. Where it is not possible to retrospectively impose such contractual obligations (at the time when the G charges have to change anyway) they can be compensated through availability payments from the SO where their departure raises other costs. This might not be necessary in negative TNUoS zones. Similarly the short-fall charges to wind generation (and other intermittent generation) to the extent that it displaces fossil generation without the need for extra grid investment are largely an issue about who pays for renewables support and how. Imposing higher grid charges on wind (without recognising their lower capacity factor) will require higher subsidies in the long-term wind contracts. Properly calculated deep connection charges would handle this directly. Long-term contracts can absorb these charges, and these contracts should in any case be allocated to minimize the total system cost of meeting the renewables target.

Efficient transmission pricing is only one part of the policy bundle required to deliver the transition to a low-carbon energy system efficiently, for the form of support for currently non-commercial low-carbon generation may itself distort locational decisions. Part of the problem arises from the way in which European Directives address market failures, specifically for CO_2 emissions and providing support for renewable electricity. If these distortions could be corrected then it would be sufficient to concentrate on ensuring efficient price signals for generation and transmission, but otherwise the practical policy issue is whether to try and offset these distortions, or to ignore them and trust that they will or can be addressed separately. Fortunately the EMR approach of supporting new low-carbon generation with long-term contracts allows these issues to be separated from the design of an efficient transmission charging regime.

Finally, although in the short run the costs of retaining the current “Connect and Manage Socialised Cost” may appear modest they provide incentives to game transmission constraints and give rise to unnecessary windfall profits to favourably located wind farms, thus raising consumer costs unnecessarily. Inefficient pricing schemes could well have higher longer term costs that are hard to quantify. If we rapidly move to long-term contracts for renewables, and if, as they should be, they are locationally differentiated, then all the associated transmission charges can be included in the contract for appropriate compensation. By making these extra charges explicit, the decision on which renewables projects to accept for contracting can be made on sound cost-benefit principles. The combination of suitable contracting facilitated by the EMR and sensible transmission charging should achieve the desired efficient outcome without having to distort each part of that package.

The paper falls into two main parts and a conclusion. The first part sets out the high level principles and their application in the best of all worlds, in which the wholesale market is competitive and all market failures have been corrected (renewables support, etc) and externalities have been efficiently internalized (carbon pricing, other air pollutants, etc).

The second part faces up to some of the practical problems, in particular what to do if market failures are not properly addressed, and past decisions cannot readily be unwound. It raises questions of materiality – whether the benefits from change are sufficient to overcome the costs of disruption, and how these costs may be minimized. It points out the advantage of trying to move market signals closer to their correct level, if not now, then with clear indications that over time this will be the intention. Although the costs of inefficient signals may appear small at present, they may escalate as cumulative investment decisions lock the system into a more costly trajectory. In any case, it would be desirable to estimate the costs of mistakes under a full range of options for transmission pricing and rules for making transmission investment decisions, specifically whether to use market or shadow prices for the value of renewables output.

The final and third part summarises the case for making changes to the current transmission charging regime. The choice between LMP and zonal pricing should be guided by the balance between lower redispatch costs under the former and lower hedging contract transaction costs under the latter, although the choice may be imposed by the EC Target Electricity Model. In addition to nodal or zonal pricing, any required residual locational element should be determined as a deep connection charge on new generation (covered for low-carbon generation by long-term contracts). All remaining cost-recovery should be collected from Load, to allow us to harmonize with neighbours once we are required to adopt market coupling. Because the EC Target Electricity Model may restrict charging to be zonal, it is a matter of some urgency to establish the number of zones required – presumably comparable to the number already used for determining TNUoS charges, but a smaller number may allow for greater contract market liquidity without distorting dispatch decisions too much. It also concludes that the design of transmission charging is one part of a wider requirement for efficient wholesale and balancing markets and a more efficient renewables support mechanism, ideally making suitable changes to the 20-20-20 Directive, failing which seeking derogations to allow support to capacity not generation.

Table of Contents

PART 1 PRICING IN AN EFFICIENT WORLD

1. Introduction

Ofgem has recently launched Project TransmiT to review GB transmission charging and associated connection arrangements. The aim of the review is to ensure that Ofgem has in place arrangements that facilitate the timely move to a low-carbon energy sector whilst continuing to provide safe, secure, high quality network services at value for money to existing and future consumers. This paper sets out a broad brush approach for prioritising the objectives and setting out guiding principles to assist in setting the transmission charging arrangements in GB. These need to be designed to accommodate considerable changes in the type and location of new generation likely over the next decade, the prospect of possibly fundamental changes to the electricity market design that DECC launched in mid-December 2010, and EC requirements of the Target Electricity Model and its associated Network Codes that are to be implemented by 2014.

It is important to remember that setting out principles and objectives for transmission charging is only part of the task of ensuring that the electricity system is able to deliver the Government's stated objectives of a secure, sustainable and affordable energy system. Transmission access arrangements can facilitate or discourage the connection of renewables, but the prime task of ensuring that the Government's ambitions for renewable electricity generation are met lies with the Government in devising systems of financial support to enable renewables to compete with fossil generation, and in ensuring that planning applications are not unreasonably denied. At the time of writing, the Government was consulting on the Electricity Market Reform (EMR) to support its objectives of adequate low-carbon generation and security of supply.

The Budget of 23rd March 2011 announced the Carbon Price Support (CPS) mechanism designed to make low-carbon generation more attractive. It aims to remedy the current inability of the EU Emissions Trading System (ETS) to deliver an adequate, stable and credible future carbon price (actually a CO_2 price, which is a charge on the carbon in fossil fuels released on combustion). Figure 1 shows how the proposed CPS will provide a transition from the current very volatile CO_2 price to a path that should reach more sensible levels when the first new nuclear stations are commissioned towards the end of the decade. Clearly if the trajectory is continued and is considered credible by investors it should make nuclear investments commercially attractive. Note that the graphs shows the price (for an EU Allowance, EUA, of 1 tonne CO_2) is £ (not euros) at Jan 1 2011 prices (deflated by the RPI).[3] This

[3] It is not yet clear whether the CPS will be indexed to the RPI or CPI.

will provided additional (arguably excessive) support to existing renewables that current receive Renewable Obligation Certificates (ROCs), but the EMR proposes that in order to de-risk zero and low-carbon investment and make the CPS credible, the Government will provide long-term contracts. These contracts greatly simplify the task of delivering low-carbon investment without distorting the transmission charging arrangements and so the EMR is an essential complement to Project TransmiT.

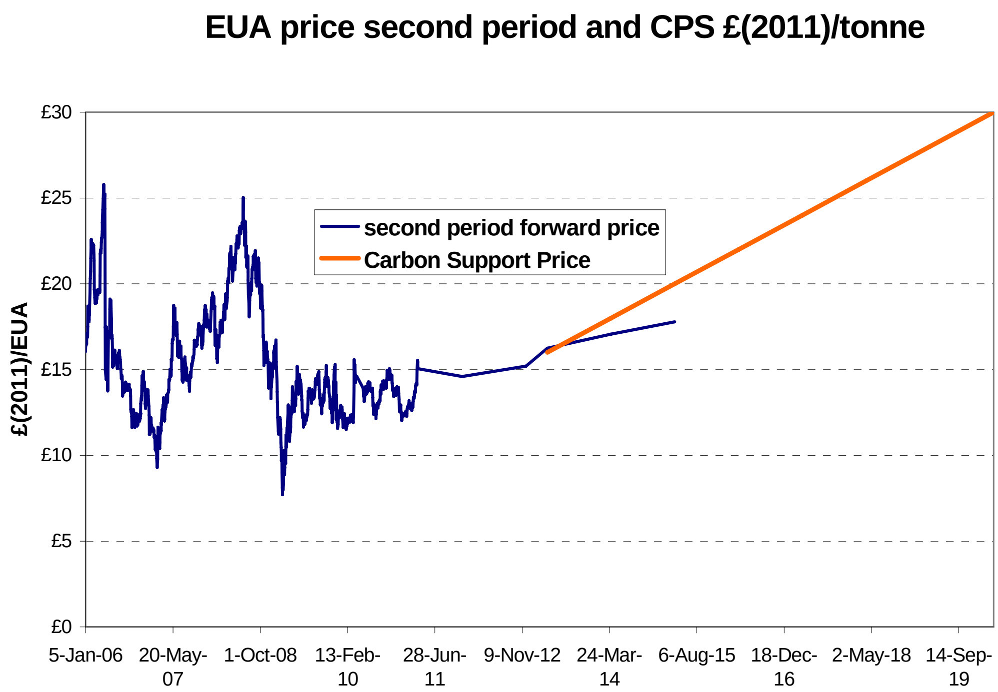

Figure 1 EUA prices at constant prices in £/tonne and the CPS
Sources: EEX and Budget Statement

One of the more serious problems prompting reform of transmission charging arrangements has been the perceived difficulty of securing timely grid connections for on-shore wind. The EMR proposes a support scheme for wind that would allow the design of transmission charging arrangements to be guided by sound economic principles without the requirement to introduce distortions to compensate for shortcomings in renewables support. It applies the sound principle that for each target there needs to be a suitable instrument that can be tailored to deliver that target without interfering with the pursuit of other objectives.

A long-term Feed-in Tariff on the German model (FIT) is the appropriate support for wind, and the FIT payments can cover the cost of any local transmission charges, while the decision to grant the FIT would naturally be conditional on delivering value for money. This would address one of the serious problems created by the current ROCs system and the various attempts to make transmission charging more supportive of wind, as will be explained in the second part of this report.

In addition to facilitating the connection of large volumes of renewable generation (mainly on and off-shore wind), the reforms to transmission charging need to take account of several other factors. Britain has three Transmission Owners (TOs, National Grid Electricity Transmission, NGET, in England and Wales, Scottish Power Transmission Limited, SPTL, and Scottish Hydro Electric Transmission Limited, SHETL, in Scotland), but from 1 April 2005, National Grid UK became the System Operator for the whole of the GB system. The TOs are private regulated monopolies, and thus delivering these objectives requires that the TOs and System Operator be provided with suitable incentives to deliver Ofgem's objectives. Affordability should be interpreted as efficiency (i.e. least cost given the other objectives, with poverty properly addressed through more general fiscal measures). Efficiency requires that the electricity generators are also provided with the right market signals, and face the right incentives (for e.g. emissions, provision of various services needed for secure, reliable and quality supply) and behave competitively.

Another consideration is that the UK is subject to EU legislation. The Third Package will have important implications for transmission, as will the evolution towards the Target Electricity Model for TSOs, under which markets (most of which are currently energy-only markets) will be coupled across a wide area (ideally all Europe), pricing zones will be defined by congestion rather than national boundaries, and the EU will evolve a system of day-ahead and intraday pricing that aims to deliver a single energy price for uncongested zones.

ENTSO-E is actively working to create the Single Electricity Market and is promoting market coupling to make better use of interconnectors and provide access to Member State markets, which, with Cross-Border Tariffication, will likely have implications for the transmission charges that can be levied on imports and exports - what in the US is described as the "seams" issue of dealing with trade across differing jurisdictions with differing tariff methods. Already interconnectors are covered under regulations 1228/2003 & 714/209, although Britned has sought derogations from these. As interconnectors are deemed to be neither generation nor load under EC rules transmission charging will need to reflect this change, although TSOs (or the owners on merchant interconnectors such as Britned) will enjoy any congestion rents arising from price differences across the interconnectors. Their guidelines (ERGEG 2011) were submitted to ACER in March 2011 (ACER 2011).

Finally, in addition to any EU legal conditions, there are UK legal and licence restrictions on what Ofgem and the industry can do and how they must behave. Ofgem and the electricity industry act under prevailing legislation and licence conditions, which may constrain what can be achieved without further legislative change or licence modification. In the present context, Parliament approved powers in the *Energy Act 2008* to enable the Secretary of State to intervene where the existing governance arrangements were deemed unable to reach a satisfactory solution to, in this case, the problem of facilitating timelier grid access for renewable electricity. A decision to use those powers was announced in July 2009, and the Government's preferred *enduring* (emphasis added) grid access reform ("Connect and Manage") was published in July 2010 as DECC (2010a). Changes to the industry codes and licences became effective from 11 August 2010, after which the Government notified the European Commission that these changes amount to a Public Service Obligation justified as in the public interest to support climate protection.

Any changes to licence conditions that affect *de facto* or *de jure* property rights will require a carefully designed set of transitional arrangements that protect these rights if lengthy and disruptive judicial proceedings are to be avoided. That means that changes to transmission charging principles that have been accepted and on which investors have made plans, such as the recent changes on improving grid access, create additional problems if further changes are to be made, as investors may have made decisions that might be adversely affected by the changes. The Government is anxious not to give the impression of an unstable regulatory or investment environment, particularly given the continual changes to which the industry has been exposed since privatization. The current consultation on Electricity Market Reform provides a good example of the difficulties faced in making a transition from Renewable Obligation Certificates (ROCs) to some form of long-term contract, where the Government has been at pains to reassure existing ROC holders that their rights will be protected if the present system is ended. The consultation will lead to new legislation, probably in 2012, and that would seem the opportune moment for changes to grid access that might better support an efficient transition to a low-carbon economy to be introduced.

1.1 Caveats

This study is confined to transmission charging and does not explore such interesting issues as charging for distributed generation (DG) on Distribution Networks and what to do if DG results in negative loads at grid supply points. As the title indicates, it is concerned with principles, not in quantifying the impacts of different approaches, which would require a considerably more ambitious project. Clearly, a major consideration in making any reform is a careful cost-benefit study of the proposed reform, but before that can be done, one needs to identify promising options to quantify, and that is the role of this study. In the same vein, some reforms are likely to require changes to IT and management systems that we are not competent to cost.

One standard objection to any reform is that it will create uncertainty and thus discourage the very investments in low-carbon generation that are the central objective of the Government, and which TransmiT is aimed to facilitate. Such complaints should be examined carefully to see if they can be readily countered. The Government's EMR aims to offer long-term contracts for all new low-carbon generation and possibly also for reliability reserves. These contracts are intended to reduce risk and the procedure for securing such contracts will provide a large part of the reduction in uncertainty that any investment might face. Any changes in transmission charging should be taken into account in these contracts, so that reforms under TransmiT should not adversely affect any new contracted investment. Arguably that means that investors who plan to connect under present arrangements and receive ROCs could face increased uncertainty, and it might be desirable to offer them the option of contracting under the new arrangements as soon as possible – and as the present system is flawed that would be desirable anyway.

2. Objectives of transmission charging

A wider and more fundamental interpretation of Ofgem's requirements are that transmission charging and associated connection arrangements should ensure that investment in, and operation of, generation, transmission and distribution of electricity to final consumers be least cost (in present discounted value) consistent with meeting the legal requirements for a low-carbon energy sector that meets EU Renewables targets as set out in the 20-20-20 Directive, whilst continuing to provide safe, secure, and high quality electricity services to final consumers.[4] This objective includes aspects beyond the transmission system that may be relevant to transmission charging, and include systems operation and planning, as well as market design and regulation mentioned above.

For the purposes of clarifying the high level principles, it is convenient to assume that the rest of the relevant world is operated efficiently – that the wholesale market is competitive and that market prices are efficient, which requires that all market failures such as carbon pricing and providing renewables support have been corrected. These restrictive assumptions will be relaxed in the second part. Indeed, part of the purpose of the Government's Electricity Market Reform is to improve the wholesale market and renewables support mechanisms to deliver the low-carbon

---

[4] Least cost ensures that electricity is as affordable as possible. Affordable is the current preferred replacement for the previous energy policy objective of cheap. Given that domestic electricity prices relative to average wage costs are now only 10% of what they were in the inter-war period, and that electricity is so valuable in facilitating life as we know it, provided it is produced at least cost it must surely be affordable. There are important implications about how the additional costs of providing the public good of renewables support should be funded (see Newbery, 2011) that are not relevant in the current paper.

future while ensuring security of supply. It is not only logical but good policy design practice to ensure that each part of the system (transmission, markets, and renewables support) is appropriately adapted to its prime purpose, so that individual components (e.g. the transmission system) are not singled out to deliver all the policy objectives where other aspects (e.g., renewables support arrangements) are better suited to deliver these objectives (meeting the renewables target).

The most important task in managing the electricity system is to ensure that investment in generation, transmission, and increasingly distribution, should be of the right amount, in the right place, using the right technology, and delivered at least cost at the right time. The reason is simple – almost all such investments are very durable (20-60 years) and immovable (perhaps transmission can be relocated reasonably easily, but securing the way leaves is almost certainly very time consuming under present planning arrangements).

The choice of type and location of generation will affect the cost of operations, which could range from nearly zero (wind) to very high (peaking open-cycle gas turbines or coal-plant with CCS). The estimated cost of meeting our 2020 energy and renewable targets could be as high as £120 billion for the electricity sector alone over the next decade, so a cost over-run of 10% or a mistaken choice of investment could lead to extra costs of £ tens of billions, raising costs for decades to come. In contrast, the way generation is dispatched can in principle be changed quite quickly (relative to the life of the assets); although the way such operational decisions are made will likely influence future investment decisions and thus cast a longer shadow.

Once the investment is in place, it should be operated efficiently, which means ensuring dispatch is least cost subject to a number of important constraints. These include

- security constraints (e.g. N-1 so that any element of the system, from a generation set to an interconnector or transmission link, can fail without prejudicing security of supply). In addition to the N-1 standard (or higher) for system operation, reserve generation capacity is required to deliver the target probability of a loss of load; and
- specified quality of service standards (voltage and frequency stability, etc.).

These operational functions are the responsibility of the System Operator (SO) acting in accordance with the standards prescribed in its licence, and their achievement requires good market design as well as efficient system operation.[5]

---

[5] It is useful to distinguish various functions that are sometimes combined in one company. An SO need not own or be responsible for any transmission assets, and Independent System Operators (ISOs) are common in regional transmission organisations in the US. A TSO might be expected to also own, manage and be responsible for investing in and maintaining the transmission network, while a Transmission Owner (TO) would own, manage and be responsible for maintaining the transmission

Before 1990, when the CEGB was responsible for planning, building and operating transmission and generation, the location of power stations and the required transmission could be chosen (subject to site availability and suitability in terms of cooling water and available way leaves) to minimise the combined cost of generation and transmission to final consumers. Coal-fired stations are best located near pits or ports, as the cost of moving coal is higher than the cost of moving power (the transmission costs and the resistive losses in the lines). Gas-fired power stations (CCGT) were more footloose as the cost of moving gas is typically lower than the cost of moving power (subject to adequate gas transmission capacity, which can be reasonably easily expanded by increasing the number of compressors and pipes). CCGTs would logically be located closer to demand centres in deficit regions. Under some external pressure, the CEGB dispatched power to minimise the sum of generation and transmission costs, so that a more expensive station closer to demand might outrank a lower variable cost station farther away, particularly under heavy loads with higher grid losses (losses are proportional to the square of the current and hence demand).

Privatization, unbundling and liberalization changed all this, and risked losing these synergies, as power stations were now free to locate wherever they could secure a grid connection (and National Grid was required to make such connections available). As a result they might not locate efficiently from the system-wide perspective unless correctly guided by suitable transmission price signals. Moreover, the design of the wholesale market (the Electricity Pool) determined that all unconstrained in-merit dispatched generation should receive the same price (the Pool Purchase Price, PPP) regardless of location.[6] Transmission access was firm. The cost of grid losses and other ancillary services were smeared over all consumers regardless of their cost causation, leading to distorted short-run locational signals for dispatch, and potentially for generation location decisions. Providing locational signals to guide locational decisions was recognised as an important issue in the run-up to privatization, so that National Grid was empowered to provide suitable signals through its transmission access charging regime.

network, but not necessarily for system operation. New investment might be tendered competitively and built by different companies (as in the case of the offshore connections in GB), and one can imagine a separate body charged with planning and tendering for new grid investments. In developing principles for setting transmission charges these institutional details will be largely ignored although they are potentially very important for delivering efficient investment in response to the price controls, incentives and interactions with the regulator.

[6] In-merit here means all stations whose offers made them cheap enough to be dispatched to meet demand and unconstrained means that the transmission system had sufficient capacity to enable them to deliver the offered power. Constrained stations were paid a different price.

National Grid developed a system of zonal charging for Generation (G) through its Transmission Network Use of System (TNUoS) charges,[7] and demand or Load (L) that was intended to signal in which zones more generation would help to reduce transmission costs (higher L charges and lower G charges) and where more generation would give rise to higher transmission costs (higher G charges and lower L charges). The sum of the G and L charges collect about £1,600 million p.a. at present, nearly two-thirds of National Grid's regulated revenue. G charges were levied on transmission capacity,[8] provided that if these charges were negative (in severely deficit zones) they would only be paid to generation that was available at the average of the three highest metered volumes over the winter period.[9] Load was charged on demand in these Triad half-hours (the peak half-hour and two others separated by at least 10 days between November and February inclusive).

These locational signals were a considerable improvement on transmission charging in many other EU countries (although locational signals can be provided by Deep Connection charges and these are quite common in the EU). Only GB, Ireland, Norway and Sweden have annual locational charges, and these countries are among the few that levy charges on generation (G) – almost all countries impose 100% of the transmission charges on Load (L).

Figure 2 shows the range of locational transmission charging for the sum of G and L – which is not quite the same as the locational guidance given to generation and load to locate to relieve constraints. Thus in GB high values of G TNUoS charges are associated with low values of L charges, and vice versa, with the sum of the two varying far less than G charges alone. Nevertheless, Figure 2 is of some interest, and also shows the considerable range in these charges across countries, although the original source (ENTSO-E 2010) should be consulted to see what charges and voltage levels are included.

National Grid also collects Balancing Services Use of System (BSUoS) charges (£800m in 2010/11) which cover the costs of the day-to-day balancing of the network, including the costs of incentive schemes (National Grid, 2010). NETA and BETTA made some changes by trying to align balancing costs with cost causation for some system services, but did not address the lack of logic of a single GB price for raw power. Various attempts by Offer and Ofgem to include transmission losses ran into judicial problems because of disputes over the rights created at vesting. The Privatisation Prospectus stated that one of the Director General's first priorities was to review and revise the basis of charging for transmission losses to better reflect

---

[7] The appendix sets out the key principles for setting these TNUoS charges for generators and section 6.2 below discusses the logic of the existing methodology.

[8] Originally DNC, Declared Net Capacity, subsequently TEC, Transmission Entry Capacity

[9] See NGET's SYS 2010.

costs. The Scottish companies found it convenient to claim that they were entitled to the basis of charging that obtained at the time of privatisation, which was uniform recovery of transmission losses.

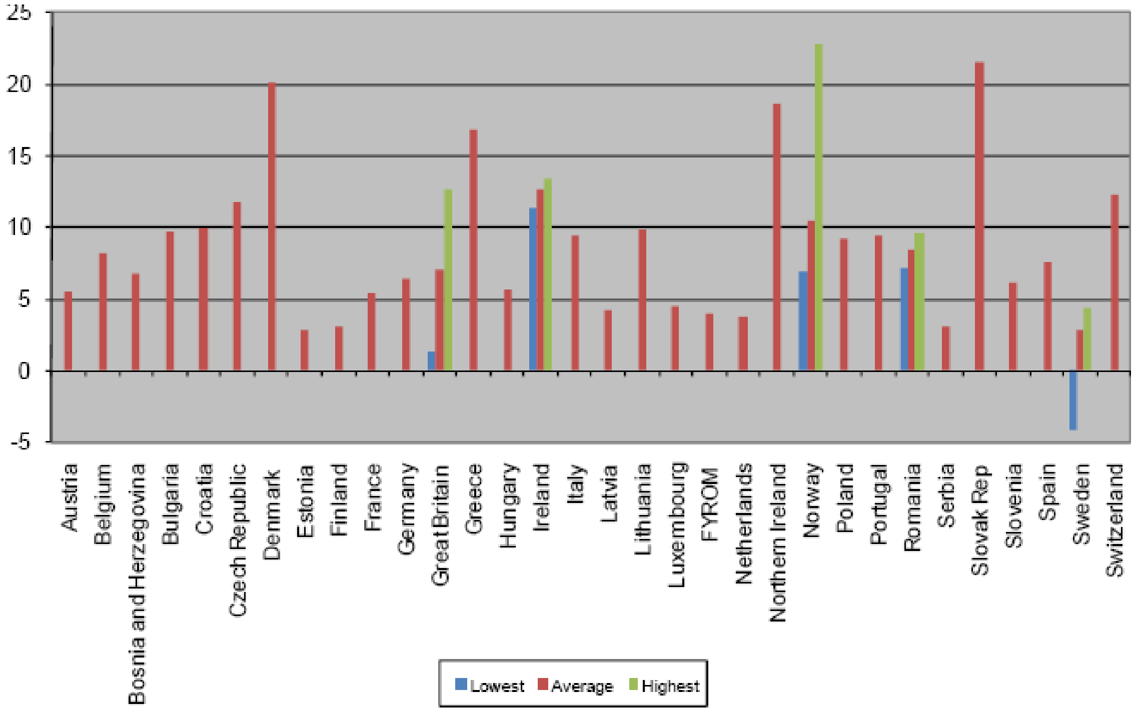

Figure 2 Comparison of transmission tariffs G+ L: impact of location for 2010, €/MWh.[10]
Source: ENTSO-E (2010)

3. What is the problem?

Transmission constraints were easy to manage when the CEGB controlled both generation and transmission, and dispatched stations centrally, and as the transmission system was strong, constraints did not appear to be a serious problem within England and Wales. Central dispatch was retained under the compulsory Electricity Pool, but generators were now free to submit offers based not on the

[10] The example taken for this comparison is the base case (5000 h utilization time that includes day hours of working days; the typical load considered is eligible and has a maximum power demand of 40 MW) for any transmission system user connected to the highest voltage level in each country. Other regulatory charges are included. Country Remarks: Greece: The 2010 tariffs have been estimated because the System Services part has components (for example the cost recovery, reserves) that are calculated ex post. The estimation of these values was based on the 2009 values. Italy: This figure includes the component to remunerate Terna for dispatching activities and an estimation for the year 2010 of the most relevant components of system services which are invoiced ex post as better explained in Country remarks at p18. Norway: based on estimations. Spain: Data for system services and losses are the same as in 2009 because estimation is not available. The access tariff taken approved for the first sixth months of 2010.

actual physical and cost characteristics of their plant, but with the duty to maximise shareholder profit. Offers, including start-up cost, might therefore be at prices above variable costs to generate additional profit. Some way to handle constraints was required, and the solution was to grant generators, who had paid for capacity-based firm access rights, compensation for any lost profit caused by not dispatching them ("constraining them off") when their offers were below the computed System Marginal Price (SMP). Plant that was required to generate in an import-constrained zone and which had offered above the SMP was paid its offer price. Where the generator anticipated that his plant would be essential to relieve an import transmission constraint, it might offer at a very substantially higher price than avoidable cost, leading to high constraint payments. The System Operator (SO) then recovered the cost of payments made to out-of-merit generators (and other ancillary payments) by adding a margin to all wholesale power to give the Pool Selling Price.

Figure 3 illustrates that transmission congestion costs were an issue in the early 1990's. Offer (and later Ofgem) created financial incentives for National Grid (now NGET) to reduce these costs with targeted investment in such cheap solutions as transformer upgrades and week-end maintenance Total costs declined to about one third their previous level by 1999.[11]

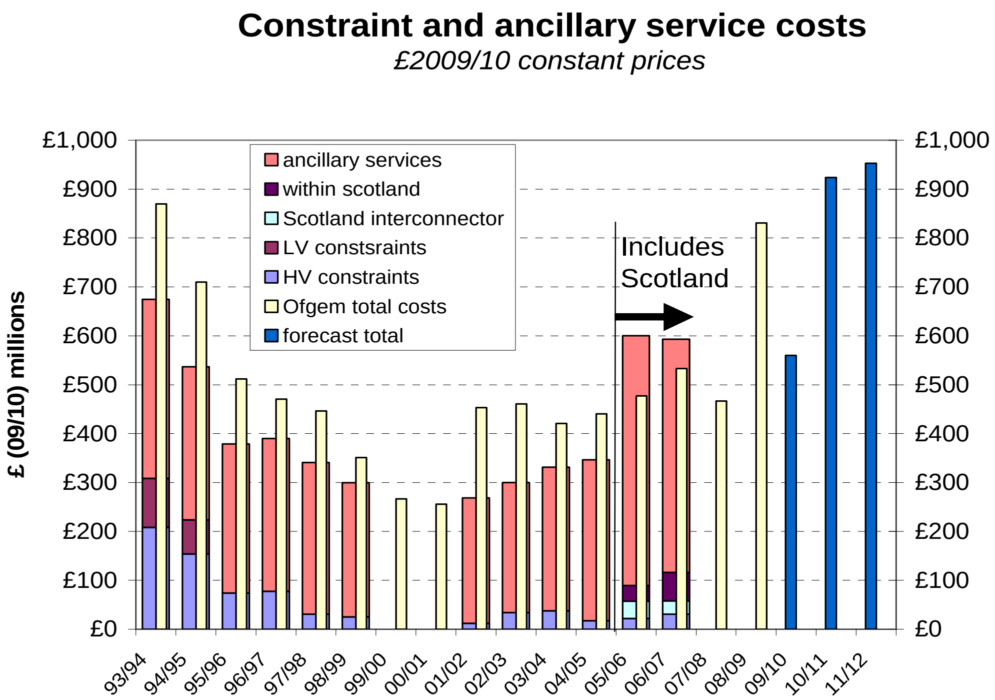

Figure 3 Congestion costs in the UK system
Source: Ofgem

[11] The narrow bars are recent figures for total costs from Ofgem, the wider bars are derived from an earlier consultation document dated 7 July 2007 containing more detail. It is not clear from the various sources whether all ancillary/balancing services are included in each series.

This was the experience against which Ofgem was required to implement NETA. There was a perception that congestion was no longer a serious issue in England and Wales, and that therefore all market participants could trade energy bilaterally without worrying about their location. Market participants have to announce their physical supply and demand decisions to the System Operator (SO) NGET 3½ hours ahead of dispatch (reduced to 1 hour before dispatch in July 2002). NGET offers firm transmission rights and so in the past only connected generators to the network for which there was expected to be sufficient transmission capacity. The SO uses bilateral contracts and the balancing mechanism to resolve remaining constraints, buying out (and hence compensating) those the system cannot accept.[12] Figure 3 shows that balancing costs increased sharply with the introduction of NETA in 2001 and then with BETTA. This access regime worked reasonably well up to the point where Scotland was integrated into the English and Welsh electricity market to turn NETA into BETTA.[13]

Figure 3 illustrates the disproportional increase of various costs associated with congestion relief and system balancing (note that the Scottish costs were not included in the period pre 2005/06, although in 2005/6 constraint costs in England and Wales were £19.6 m and for the whole of GB £80m). Since then constraint costs have risen from £70 million in 2007/8 to an outturn of £263 million in 2008/9 and £139 million in 2009/10 (NGET 2009b).[14]

Figure 4 gives more detail as to the source of these constraint costs. The "intact" constraint costs in Figure 4 are those that would persist even if the system were completely available, while "planned" constraint costs are those that arise because capacity is out of service for routine maintenance, refurbishment and upgrading. In 2008/9 these accounted for the bulk of the costs, but even so the intact costs rose appreciable at the Cheviot boundary between Scotland and England. Fault costs (that are caused through unplanned unavailability) are rising as the system ages and are increasingly under stress.

---

[12] NGET has incentives to manage investment efficiently and not over-build under price-cap regulation (in contrast to rate-of-return regulation) and so may even have a temptation to somewhat under-invest, as the costs of buying out constraints are largely passed through to users, while the cost of increasing transmission capacity fall on NGET (at least until the next price control period). At some stage, though, having a larger regulated asset value becomes an advantage, motivating the delivery of the assets paid for in the Price Control.

[13] The continued failure to include marginal transmission losses in the entry and exit prices was an exception to the claim that the system worked reasonably well.

[14] These figures were supplied by Ofgem and are slightly difference from those in NGET (2009b), which were used for the figure.

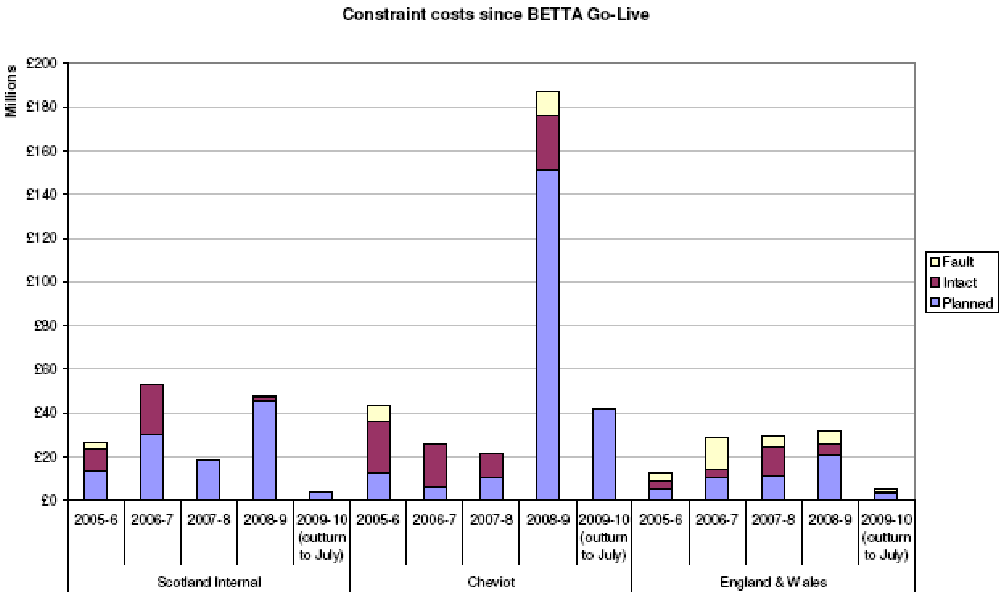

Figure 4 Constraint costs by type and location
Source: NGET (2009b, figure 5)

The assumption that the electricity network is mostly unconstrained no longer holds for three reasons.

First, the inclusion of Scotland means that a region that traditionally was exporting to the South and had been facing export constraints is now part of the same bilateral market. Before BETTA physical capacity on the Cheviot boundary was limited and the Scottish TSOs had to manage their system internally to limit flows, but under BETTA Scottish generators are allowed to sell as much as they like to England, and the SO has to buy off excess supply.

Second, the generation pattern is gradually evolving, with shifts between fuels, and new locations of demand and generation. New generation in export regions, particularly wind in Scotland, contributes to increasing levels of network congestion and different times of the day when congestion becomes problematic. The Large Combustion Plant Directive limits the hours of non-compliant but flexible old coal-fired generation for which the grid had been designed on the assumption that they would run base load. The larger the volume of generation that has to be constrained off and replaced by other plant, the higher will be the cost per MW replaced, which amplifies the total cost, as Figure 5 shows. Note that 70% of the time the system is long and that 80% of imbalances lie between -670 MWh and +270 MWh. Ten percent of the time the spread is more than £50/MWh.

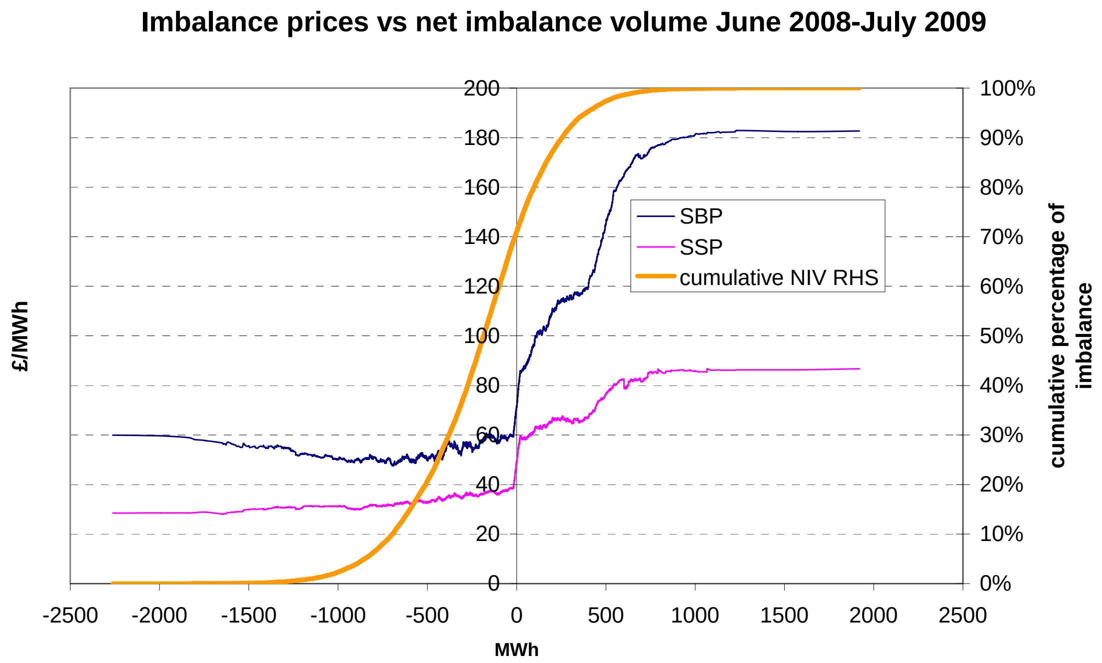

Figure 5 Bid-offer spread in the Balancing mechanism (moving averages)
Source: Elexon data[15]

Third, much of the new renewable energy connected to the network has low capacity factors. As a result much of the existing generation is not entirely replaced, but will remain connected to the network to complement the output of intermittent renewables. When the wind blows, then coal and gas power stations reduce their output, and vice versa. This means that in an efficient system, the network capacity no longer needs to closely match generation capacity.

As the simplified illustration in figure 6 shows, suppose there is an existing 1 GW coal power station connected to a transmission link of 1 GW capacity serving a peak load of 1 GW, and suppose a 1 GW wind farm is contemplating connection to the same node as the coal station. Whenever there is wind, the coal power station should reduce output, as wind is then the lower cost option to produce energy. Overall the ratio of capacity to peak demand will increase, but the main determinant of transmission capacity is peak demand, not total generation capacity and there is no need for an grid expansion (in this simplified model – if there were more load further downstream it might be worth expanding the upstream link).

[15] I am indebted to Ben Hall and Nigel Cornwall for providing this data in user-friendly form

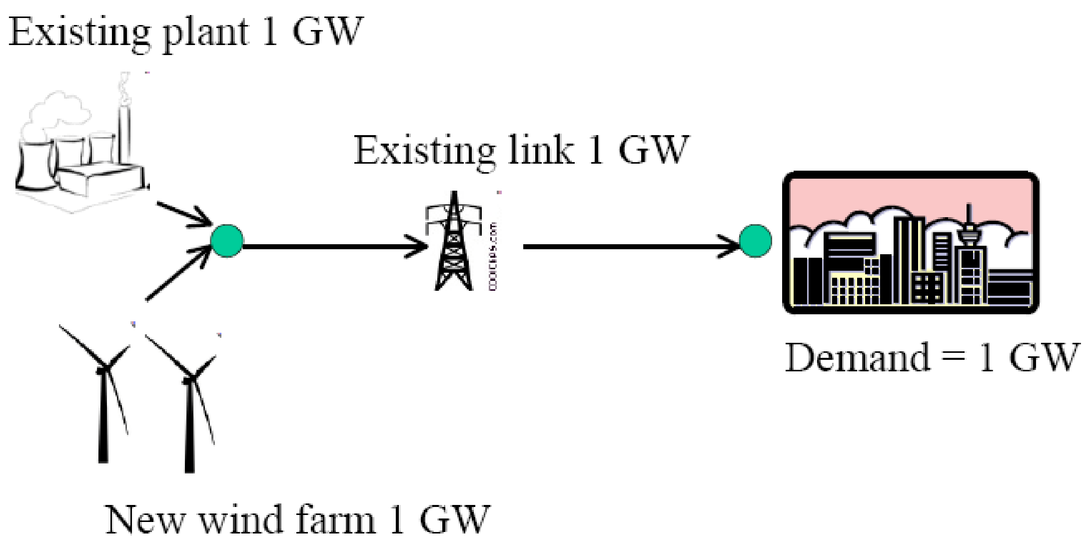

Figure 6 Allocating transmission capacity to match demand

One of the obvious drawbacks of the previous transmission arrangements was that National Grid would only connect generation if it had enough capacity to allow it to generate at its requested TEC. In figure 6, National Grid would not be able to ensure that existing and new generators could both produce at full capacity until it expanded capacity by an extra 1 GW (which might not make sense, is in this scenario).

A more sensible approach would be to allow the new wind farm to connect and to share the transmission capacity. Such capacity sharing would have been theoretically possible under the previous regime. Assume a coal power station has the historic access rights to the network. A wind turbine in the same constraint zone could negotiate spot (i.e. on an hour by hour basis) with the coal power station whether to buy the access rights and displace the coal station. In practice, the owner of the coal power station would use its negotiating power to sell these access rights at a likely high price, capturing much of the value of the wind turbine. Equivalently, the fossil generation company could buy out the wind farm and manage the combined capacity, or sign a long-term contract to do so. A far-sighted wind farm developer would negotiate the contract before building or face a rather unattractive offer later and might be deterred from the lengthy process of securing planning and other access rights as a result.

In practice a spot market solution would be difficult, given the current design of the GB Balancing Mechanism (which, as its name implies, is not a market, in contrast to the former Pool). The Balancing Mechanism (BM) accepts the bids (to buy power or reduce supply) and offers (to increase output or reduce demand) from generation (and load) across Britain. Its function is to ensure system balance in real time, for which it needs to be able to call on offers to increase output or bids to reduce output. It has always been plagued with low liquidity,[16] such that prices can respond dramatically if the system becomes (or is expected to become) appreciably short, as figure 5 shows. This liquidity would be further reduced if physical access products were to be distinguished by location and traded.

The recently implemented solution to the problems facing wind farms seeking to connect to the system but unable to secure the necessary Transmission Entry Capacity (TEC) has been to require NGET to "connect and manage", while continuing to socialize the resulting balancing and constraint costs. The result would be that if the wind farm were built and connected, it should always gain priority access in any efficient dispatch, and in practice NGET would either have to constrain off the coal station via the balancing mechanism whenever the wind farm generated, or would have to sign a cheaper long-term contract to dispatch the coal station.

There remain important issues to do with how much it would be reasonable to charge the two generators for grid access, as they both benefit from access to the link but might have very different capacity factors. Under the present system the coal station would be compensated for any loss of profit so would presumably still be charged the full TNUoS, so perhaps the wind farm should be charged for the lost profit only of the coal station. Whether this would give the right location incentives for the wind is less clear and will be considered below. Locational Marginal Pricing (LMP, described below) would automatically deliver an efficient dispatch, with the LMP equal to the demand node price in windless conditions (less any marginal losses) and at the short-run marginal cost (SRMC) of coal (at its reduced level of output) when the wind blows. The advantage of connect and manage is that it in this case it delivers the correct operating decisions and overcomes the barriers to negotiated solutions. The key questions are whether it gives the right investment guidance and whether it encourages generator behaviour that raises consumer costs.

4. High level principles

Transmission charging (together with the design of the wholesale and balancing markets) has to fulfil two major functions with two very different time scales. Prices need to guide the long-run investment decisions and the short-run operating decisions. Once an investment is made, the long-run guidance role ends (except if it affects exit), but daily operation decisions continue to be affected by price signals. It would clearly be desirable in many cases to be able to disentangle these two different functions, so that short-run pricing is not unduly influenced by the need for good

[16] Arguably intentionally, to encourage agents to contract ahead of time to avoid penalty prices in the BM.

long-term decisions. In some cases there is no conflict between the two, but when there are reforms or market design changes there may be a conflict between continuing to honour the expectations of earlier investors while adapting price signals to changed circumstances. It is worth noting that the present method for setting TNUoS charges does not provide any assurance about the stability of charges,[17] and there are concerns that the large increase in transmission investment will adversely affect existing generators, an issue that is considered further below.

There is a very simple device for making this separation, through long-term contracts which give rights that can be exercised in the daily markets. Gas transmission pipelines from wells to markets are often financed by long-term contracts that entitle the subscribers to capacity rights that can then be traded in shorter-duration markets. The same could also be devised for generation connections to the transmission system. A generation station would expect to operate for 20-60 years, depending on type, perhaps with the expectation of a gradual diminution in capacity factor as it is superseded by more efficient plant. Transmission is similarly long-lived, and if additional reinforcements are required to accept the new generation, the logical contract would be a long-term one that would amortise the fixed transmission costs over an agreed number of years (say 15-25), and would be a long-term mortgage liability on that station, to be paid regardless of whether or how much it produced. That payment would entitle the station to a specified level of TEC to the grid for an agreed period (arguably the expected life of the station or of the assets invested to accommodate entry) and would be transferable to other users at that node. This is discussed further in section 6.1.

This issue becomes important when the transmission charging regime needs to change – for example to adapt to the Target Electricity Model or EU rules on interconnector charging – or when new kinds of generation are to be connected, for example intermittent wind with low capacity factors. It may well be desirable to indicate a new set of locational signals to guide investment in this new regime without having to worry about how they will impact existing plant, whose location decisions have already been made and cannot be changed. Such changes can be very significant. Thus SSE (2010) claims that investing in offshore DC links to provide more transmission capacity from Scotland to England (the "bootstraps") could raise the annual TNUoS charge on Peterhead from £22m to £55m p.a.[18]

---

[17] The draft TNUoS tariffs for 2011/12 show G changes ranging from +£0.75/kW to -£1.80/kW across zones, and L changes from £1/kW to over £3/kW caused by changes in demand and generation.

[18] If the clean spark spread is £5/MWh, the capacity is roughly 1,200MW and the capacity factor is hypothetically 75%, the annual gross profit would be about £40 million, so the increase in TNUoS would be very large compared to this level of profit.

It is helpful therefore to separate out two quite different functions of transmission charging. The first, and most important one that will occupy most of our attention, is to ensure that the system gives the correct guidance to new investment, as that is where the greatest gains or mistakes will be made. The second is to deal with existing generation which has largely implicit contracts. Had they been explicit contracts of the kind argued for above, then they could be largely ignored in designing the short-term price signals, and it is convenient to imagine that these implicit contracts have been formalised and can be put to one side. The process of formalising them is then a separate and financially hugely important task that can be addressed once the form of the new market design and transmission regime have been determined.

When determining transmission charges that will guide investment decisions in an unbundled and liberalised electricity market, the logical sequence is to start with how an existing transmission system should be priced to deliver efficient short-run price signals to guide the operation of generation and ancillary services[19] and the demand side. The next step is to consider whether these short-run pricing rules are sufficient to guide longer-run decisions (specifically on generation location and type, but also on demand location and on investments in demand management, transmission and interconnection). If not, then additional spatial signals will be needed to encourage efficient location investment decisions. Once these variable price signals are determined (those that vary with time and place of use), the final step is to compare the revenue these will deliver against that allowed by the regulator. The revenue allowed in the Price Control cover the costs of running the transmission system, including the major part, the interest and depreciation the allowed capital value (the Regulatory Asset Value). If, as seems likely, there is a shortfall, then the balance of allowed revenue needs to be collected in the least distortionary way, and the design of these additional charges will require some care.

To summarise, 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, wind, etc.);
- legality, fairness and political feasibility; and
- cost-recovery.

[19] See Appendix D for the description of ancillary services and the current system of charging,

Existing generation needs to be motivated to generate efficiently and not to exit prematurely, but otherwise the charging regime should be such as if it had inherited a suitable long-term access contract, under which it would not expect to receive windfall gains from changes in transmission charges, nor to be subject to adverse diminutions in its contractual rights. That is not to say that changes in market prices (e.g. for fuels) that are commercial risks should not be passed through to the generators, but that they have a reasonable expectation of being shielded from regulatory changes that are intended to address different problems.

5. Optimal short-run use of the existing network and nodal pricing[20]

For short-run optimal use of the network, the benchmark is locational marginal pricing (LMP), also known as nodal (spot) pricing. The simplest way of thinking of LMP is that it would be the price that equilibrates supply and demand at that node (i.e. the point at which a generator injects into the grid or a large load receives power) rather than in the market as a whole (i.e. GB under BETTA). Demand at the node is well-defined, but the cost of supply might have to include the cost of delivery from some other node if the nodal demand exceeds the supply connected at that node. Under LMP the price at any node is made up of three components: energy, congestion and losses. The energy component is the cost to serve the next increment of demand at that node, from the least expensive generating unit that still has available capacity. However, if the transmission network is congested, it may not be possible to deliver the next increment of energy from the least expensive unit on the system because it would violate transmission operating criteria. The congestion component, or transmission congestion cost, is the difference between the energy component of the price and the cost of providing the cheapest energy that can actually be delivered at that location. The congestion component can also be negative in export-constrained areas where there is more generation than demand. Losses are the marginal loss over the extra transmission from the supplying generator, where the marginal loss is twice the average loss in that line. A brief summary of the concept (from which this short description is taken) and the way it operates in the North East of the US is given in Appendix C.[21]

To achieve efficiency this requires that generators submit efficiently priced offers (i.e. a schedule of short-run marginal cost, SRMC, up to full capacity, together

[20] This and later sections draws heavily on Brunekreeft, Neuhoff and Newbery (2005).

[21] The concepts of efficient pricing have been developed in Bohn, Caramanis & Schweppe (1984), Read & Sell (1989) and Hogan (1992) and largely adopted by the US FERC in its Standard Market Design (SMD).

with start-up and other relevant costs and constraints).[22] The dispatch algorithm can then determine the efficient dispatch (allowing for reliability and transmission constraints and transmission losses) 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 charges to recover the balance of the regulated costs. If the offers are not set equal to marginal cost, scheduled flow patterns will be distorted. However, given that nodal pricing results in a flexible allocation of transmission capacity and thereby ensures that generation at each node faces as much direct competition as possible from other generators connected to the system, it mitigates the market power exercised by strategic generators more than other designs and hence distortions should be smaller than with other designs (Green, 2007 suggests that this benefit could be large as it could raise consumer welfare by 1.3% of generator's revenue or £200 million at current electricity prices).

Nodal pricing is the natural counterpart in a meshed transmission network to competitive pricing in a market, where if each agent offers goods at marginal cost, the result will be the efficient market equilibrium. Just as these competitive prices can be found as the set of shadow prices associated with maximising some weighted sum of individual utilities, so the shadow prices computed from the dispatch algorithm gives a set of nodal prices that will lead to an efficient dispatch, provided they are based on the correct generator costs.

While LMP may appear to be a highly artifical construct ("system lambdas" sounds very abstract), they are just a by-product of the Optimal Power Flow (OPF) optimisation tool used by system operators since the 1960's in centrally dispatched power systems. The first step is normally to find the least cost set of generators that can meet demand without taking account of transmission constraints (and this gives the System Marginal Price, SMP, often used as a reference price). In the second step

---

[22] It might be thought that if LMPs are based on SRMCs, then the generators will not recover their fixed costs. If there were no transmission constraints, the System Marginal Price (SMP) will be higher than the SRMC of all except the most expensive generator required to operate, and this will produce a contribution to covering the fixed costs of all these generators. Under admittedly ideal conditions such pricing will exactly cover every generator's full costs over the course of the year. (These conditions require that the mix of various types of plant is optimal and in long-run equilibrium and that all available plant receives a capacity payment equal to the excess of Value of Lost Load over the SMP multiplied by the Loss of Load Probability, as in the former Electricity Pool.) Departures from these conditions might result in under or over-payment of full costs. Bohn et al (1984) demonstrated that the same is true under LMP in constrained networks given efficient scarcity (capacity) pricing as in the pool model. Certainly the US has been sufficiently concerned at the problem of "missing money" that they have introduced capacity payments to make up the short-fall, and the current EMR is consulting on whether they would also be desirable in the UK.

the OPF tool is used to find the least cost re-dispatch of generators to relieve congestion by based on their cost characteristics. LMP has been successfully implemented in a wide range of electricity markets, most notably in the PJM Interconnect, a market that has evolved from its original Pennsylvania, New Jersey and Maryland base to cover an area with three times the GB installed capacity (see Appendix C). As time passes and the evidence of its success grows, it is increasingly deployed in other states across the US, partly as a response to the urgings of FERC's proposed Standard Market Design, and partly as other solutions are revealed to have costly flaws.[23] For short-run congestion management there is agreement that a system relying on LMPs works and is efficient (provided that offers are competitive), and more generally, avoids the problems faced by all other approaches in liberalized systems.

The next question is whether these LMPs contain all the necessary spatial information to guide long-run investment location decisions. There are several reasons to doubt this, but before examining them, we need to understand what other services the transmission system provides, and how or whether they can be properly charged, as it is the set of all locational charges that will influence these investment decisions.

5.1 Complications: charging for reliability and quality of service

Consumers need reliable and high quality supply. Quality of service (QoS) applies to all customers connected to the relevant part of the network, and as such cannot readily be subject to market forces – it is hard to imagine different consumers being able to buy different qualities of service from the same node, for example. It will therefore be laid down in grid codes and the licence conditions for the System Operator, and these will have implications for the amount and type of grid investment needed to deliver these standards.

Reliability is in theory potentially different and one can imagine a market for reliability. Reliability requires adequate capacity and transmission, so that if one generator fails locally, or one grid link fails, then power can be delivered from other sources or through other links. In theory reliability is a private good, in the following sense. If each consumer could specify the maximum price he is willing to pay to continue to receive supply (the Value of Unserved Energy or Value of Loss of Load,

[23] Arguably LMP is more important in the US than in Europe, as the US transmission links between the large number of smaller utility control areas are considerably weaker than across comparable geographical scales in Europe, so transmission constraints became more obvious after liberalization widened markets beyond these utility areas, so that nodal prices can differ widely over modest distances. As increased volumes of wind generation connects to currently underserved parts of the GB system, so GB may experience similar problems in the future, and the aim of TransmiT is to provide a future-proof solution.

VOLL), and if each consumer could be disconnected when the price reached that level,[24] the Transmission System Operator, TSO, could shed load as the price rose above each customer's VOLL and thus always balance supply with the demand that consumers were willing to pay for.

Over longer time scales the TSO would decide whether to expand the grid capacity to deliver to each node according to whether the improved reliability offered matched aggregate consumers' willingness to pay for that reliability. The grid charges paid by consumers at each node would include the cost of meeting the chosen level of reliability. Similarly, the TSO would announce charges at each node that would encourage generators to make the right decisions about the amount of local capacity to supply (and its reliability) to increase the reliability of local and overall supply. These would be longer-run spatially varying charges (set to guide investment decisions), supplemented through the balancing and/or reserve market to induce optimal availability in real time.

In practice it is not yet possible (and may never be acceptable for many consumers) to decentralize these reliability choices to all consumers and they are therefore also set in grid codes, which apply to all consumers (except for a modest number of large consumers offering demand reduction services). Consequently, reliability has similarities to quality of service in that it has many characteristics of a public good, and is a responsibility of the TSO. Smart metering may in future allow more consumers to participate in meeting these standards (or in selecting their own preferred reliability standards) but that will still leave the remainder to be handled collectively.

The effect of setting quality and reliability standards for the whole grid, rather than in response to individual demands, combined with an asymmetry of outcomes from under or over-supply (costly disruption from under-supply, small extra cost for over-supply) is that the TSO will aim at grid capacity that is likely to be oversized relative to a perfectly informed economically justified size. This will reduce the range of spatial variation in the LMPs once the system has been adjusted to deliver the reliability and quality requirements, but if additional generation wishes to connect before any transmission upgrades are made to absorb extra injections at the same reliability and quality standards, the LMPs could still change markedly from their current values. The absence of much current variation in LMPs does not mean that all locations are equally attractive for new generation. Instead investment decisions would need to be made on the basis of a whole time sequence of LMPs, leaving the current LMPs as primarily a guide to (or reflection of) optimal dispatch.

[24] This could be done in a decentralised way with two-way smart meters, or via contracts with high penalties for not disconnecting/reducing load in response to automated calls.

5.2 Further complications: lumpiness, uncertainty and scale economies

There are further complications that might make short-run LMP an imperfect guide to long-run investment decisions. The future levels of generation and demand (and their location) are uncertain over the expected future life of new grid investments (60+ years), and these grid investments are lumpy, if only because there are economies of scale once the sites for the towers have been secured. Both factors make it sensible to over-size lines to allow for future uncertain needs, so that as with the asymmetry argument above, the TSO will aim to over-provide capacity (and the regulator will likely also accept that judgment for similar reasons).

If links are typically oversized, then LMPs will vary less across nodes than would otherwise be the case (marginal losses will be lower and transmission constraints implying very high or low nodal prices less frequent). As the TSO collects revenue from the difference between injection and withdrawal LMPs, this revenue will be less than in a continuously optimally sized network. Simulations suggest that even with optimal investment in generation and transmission the combination of these various factors allow only about 20-30% cost recovery (e.g. Perez-Arriaga et. al., 1995).

The various causes creating the wedge between LMP revenue and total network cost are illustrated in figure 7. 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.

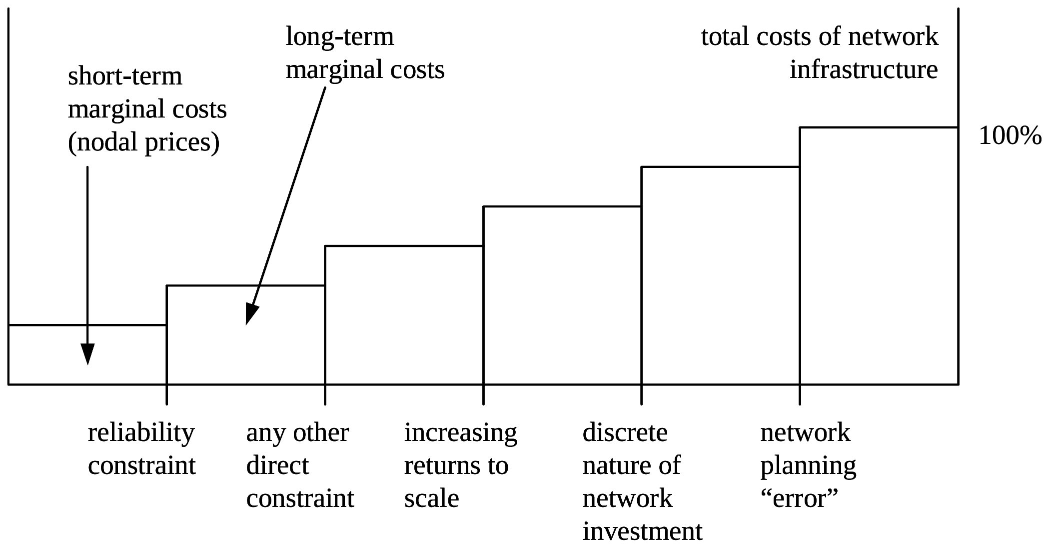

Figure 7: Cost drivers of a transmission network
Source: Perez-Arriaga, (2003)

One simple way to see how reliability, if not charged, undermines cost recovery is to consider a network that connects nodes, each of which has enough local generation to meet the local demand and in equilibrium the same marginal generation cost. In such a world local generation could optimally meet local demand at each node with no flows on the links. In that case all LMPs would be the same and there would be no difference between the prices paid by L and facing G, so there would be no revenue. The network would exist purely for reliability and QoS reasons, and could almost certainly be justified on that ground, but there is no payment under pure LMP to cover its costs.

Offsetting this tendency to under-recovery 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). That might be the case in moving wind power from the Northwest of Scotland down to London. The effectiveness of marginal generation due to transmission losses varies from less than 90% in Scottish and Hydro's zones to 112% in the Thames Estuary (for the projected seven year period covered by the current SYS in National Grid 2010 at Table 7.4) – a range of more than 22%, although average GB grid losses are quite low (less than 3% of peak load (table 7.3).

5.3 The choice between nodal and zonal charges

Nodal pricing has the advantage of providing generation with the correct price signal to guide its optimal short-run dispatch, but they introduce a new source of risk, as the local value of energy will now differ, probably by varying amounts, from any reference price that would be used in determining the price for consumers. This risk will need to be hedged, just as the volatility of the uniform GB wholesale prices need to be hedged.

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, although variations around M will be guided by the spot price, not the contract price, ensuring efficient short-run decisions. The CfD will need to define the location where the wholesale price is set (a National Balancing Point, or Pool Price, or swing bus, or System Price). 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, and described above as, *basis risk*).

The solution to this basis risk 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 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.

While this appears to solve the problem facing generators at each node, it creates another set of problems, because there will be a potentially very large number of idiosyncratic FTRs between each node and the reference node (or system price) and very few agents interested in any one of these financial contracts. One potentially attractive solution is to group nodes into zones within which there are few congested links, but between which congestion is important, at least in some hours. This approach has been adopted in Nord Pool, which starts by determining a System Price ignoring all internal congestion (much as in the Pool described earlier). If this dispatch is feasible, all zones have the same price, but if not then plant is re-dispatched to satisfy the transmission constraints, and the prices that clear each zone or set of mutually unconstrained zones will differ. This process of starting with the whole region and then splitting the market into separate price zones when necessary is naturally enough called market splitting (and is closely related to the Target Electricity Model's market coupling in which previously separately dispatched regions are linked together),

There is now a clear trade-off, for the fewer the number of separate price zones, the fewer FTRs or CfDs (defined from the zonal price to the System Price) are needed to provide hedging, but the higher the costs of re-dispatch, which is completely avoided under the dispatch supported by LMPs. These costs can be amplified by the opportunistic bidding behaviour of generators, who may play the "inc-dec" game. That is, anticipating that some internal transmission link will be constrained, they offer a large volume of power at a very low price, and then have to be constrained off (their large increment has to be decreased, hence inc-dec) and paid their lost profit, equal to the zonal price less their very low offer price.

One of the key design issues is whether and if so how large a zone will be justified, bearing in mind these trade-offs. That will depend on the robustness of the transmission network, the bidding rules and the effectiveness of regulatory oversight. Thus in the Single Electricity Market in Ireland, generators are required under the bidding codes to offer at short run variable cost, based on audited cost estimates and indexed to spot fuel prices, in which case such inc-dec games would be ruled out, and the constraint costs would be the real costs of meeting the security constrained level of demand.

6. Longer-run price signals to guide investment

If LMP prices fail to cover costs, and particularly if a connection is oversized because of lumpiness, it seems unlikely that a new generator connecting to a node and subsequently facing a time series of LMPs over future dates would face the true costs (and benefits) it imposes on the system. A new generator will cause changes in the pattern of LMPs that are not themselves sources of market failure (assuming all previous and subsequent location decisions had perfect foresight about their LMPs).[25] However, a new generator's actions might also affect system reliability and/or the QoS and hence either precipitate the need for more or less transmission investment to maintain the required standards of reliability and QoS. As a high level principle any such costs should be charged to the new entrant, although it might be difficult to make accurate calculations about future costs and appropriate charges.

6.1 Deep vs. shallow connection charges

This raises the issue of deep vs. shallow connection charging. Deep connection charging would charge entrants for the full additional consequential costs they impose on the network (in terms of reinforcement investments), part of which would be covered by the LMP charges the entrants would pay at the node. Shallow connection charging (the current system) just charges for any assets needed to connect the generator or load to the grid (which for offshore wind could be very substantial). A surprising number of EU countries have deep or at least partially deep connection charges, as shown in Table 1.

Theoretically, deep charging gives the correct information that, allowing for the revenue from the predicted future LMPs, would give the correct investment signals in a world of perfect certainty and no lumpiness. In practice, matters are more complex. If future generators might choose to locate at or near the node to which our generator is to connect, then the TSO should think ahead and over-size

[25] Just as any change in demand or supply of a good in a competitive market can change prices, these new prices remain the efficient prices and so there is no market failure – the price changes are distributional externalities that merely redistribute income.

the reinforcement, given lumpiness, economies of scale, asymmetry (extra investment cost vs. the downside cost of under-investing and risking more blackouts) and uncertainty. Then the question is what fraction of the reinforcement might reasonably be charged to the first generator to connect – too high and he might not locate there, invalidating the assumption on which the charge were made, too low and future locators might find charges too high and decide against connection, with the same result.

Table 1 EU Countries with Deep or Partially Deep connection charges

Source: ENTSO-E (2010)	
Austria	Grid user builds own connection line. If grid reinforcements are necessary the user has to pay for this
Croatia	
Estonia	All the equipment belonging to the connection + all reinforcements needed prior to the connection are included in the connection fee.
Germany	**Deep** (customers) **shallow** (power plants)
Hungary	**Partially Deep** Maximum of 70% of investment costs for customers and 100% for generators; or generators build own connection line. If the generator used at least 50 % of renewable energy for its production per year, it pays only 70 % of the investment costs, and if this value is at least 90 %, it pays only 50 % of the investment costs.
Ireland	**Shallow to Partially Deep**. The connection charge is based on the Least Cost Technically Acceptable shallow connection method. However the Least Cost Technically Acceptable shallow connection method depends on the availability of appropriate transmission infrastructure in the area e.g. voltage level etc. Charges can also include station common costs or station extension costs (if higher). Demand customers pay only 50% of the charge, generators 100%.
Latvia	**Deep.** Grid users builds own connection line. All connection equipment and reinforcement are included in the connection fee.
Lithuania	**Partially Deep** (20% of investment costs for customers and 100% for generators)
Romania	
Serbia	**Shallow:** Generators and distributors pay for connection lines aimed at meeting security criteria (the most frequent case is the building of 'in-out' connection toward an existing line) and for substation. **Deep**: Industrial customers, in addition to payment for connection lines and substations, have to pay connection fees aimed at supporting further network development. Connection fees are 43 € per approved power in MW. **Note**: Generally, in 110 kV network, grid users keep ownership over 110/x kV substations
Slovak Rep	**Partially Deep**. Distribution companies pay 40% charge, TSO pay 60 % charge. Direct customers connected on the TSO pay 100% charge.
Slovenia	
Sweden	

Consider, for example, a location with a large excess of local G over L distant from demand centres (e.g. Scotland), and suppose that the needed (and economically justified) transmission investment is being held up by planning complaints. The resulting short-run nodal price difference from the generation node to the demand located outside the export constraint may be well above the long-run marginal cost (LRMC) of expanding transmission capacity – in other words with a locally undersized export network, LMPs may be too low so that generators receive too low a nodal price because of constraints on grid expansion.

If the TSO were able to predict the date of commissioning of the reinforcement, he could publish the future estimated LMPs, which would signal to those planning to connect to the node whether to wait until nodal prices had risen, or whether to go ahead and accept a period of low prices. Arguably, some part of the investment that would benefit the new generator should be collected from that generator, if the line would otherwise not be reinforced, as the generator will benefit from the higher future LMPs. If the line is such that there is no longer an export constraint then the LMPs at each end will be almost the same (except for transmission losses) and so the TSO would receive no extra revenue on its investment, which would need to be recovered from other users.

This is illustrated in Figure 8 where the existing power plant and transmission link (continuous arrows) to the market at right have been suitably sized. A large number of wind farms would find the current LMP of £38/MWh very attractive, but if they were to connect before any transmission upgrades, the LMP would fall to £10/MWh (LMP_1), while the wind would only be economic at a local LMP of £25/MWh or higher. However, an integrated planner choosing generation and transmission investment would find both investments together profitable, as the extra transmission (dotted arrows) cost of £14/MWh would be less than the wind profits of £15/MWh now that its power is worth £40/MWh (LMP_2). Lumpiness or the expectation of future additional wind farms leads to a transmission investment capacity that eliminates all congestion. At these new LMPs the TSO makes no profit and others have to pay for the reinforcement.

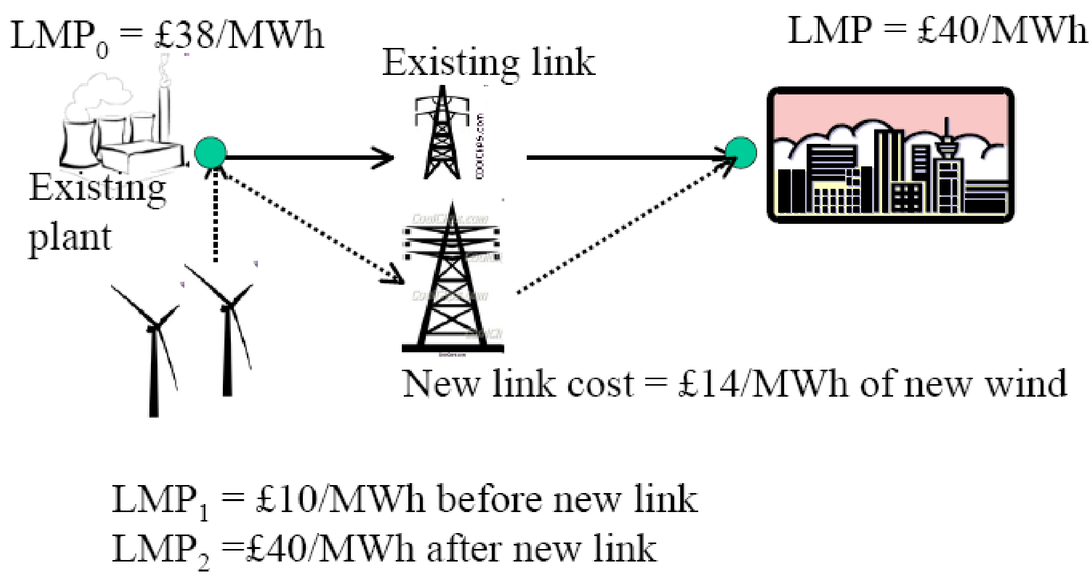

Figure 8 Wind and transmission investments

If deep charging is contemplated, the practical question is whether there are any cases that are sufficiently well-defined and where the LMPs give inaccurate locational signals, that justify additional quasi-deep charging, and if so whether rules for how to do this could be devised that would be accepted (in the courts) as fair and non-discriminatory, but understanding power flows in a meshed network and the case for reinforcements would likely be difficult for judicial appraisal. On the other hand if all transmission investments have to be justified on their LMP revenues many sensible investments will be blocked. A good regulator would examine the economic case for both the generation and associated transmission investment and would consent the former and require the latter to be undertaken if justified, although perhaps an independent Transmission System Planning Authority (TSPA) might be better placed to do that. Such a TSPA might also actively identify suitable locations for new wind farms and pro-actively seek planning permission, then auctioning off the sites with planning and connection agreements in place.

Some (perhaps most) of these problems can be circumvented by offering long-term contracts for accessing the network, as argued at the start of this section. These, together with standard energy contracts, will be needed in any case to reduce risk and the possible disincentive to efficient decision-making that might occur if it were just left to investor's predictions of likely future LMPs (and other locational charges). Investors deciding where to locate new generation based on current LMPs may choose the wrong location, if, as will often be the case, current LMPs differ substantially from their long-run equilibrium. Of course, future electricity prices are also unknown, and may be very volatile on an hourly, seasonal and annual basis, but the contracts would hedge the *basis risk*, that is the difference in the local nodal price and the National Balancing Point used to determine the average wholesale price, removing locational risk and just leaving the normal commercial risks, as described above, where the appropriate financial hedge is a Financial Transmission Right, or FTR.

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.) Once this is accepted, these contract prices can be designed to reflect the necessary investments to deliver reliability and QoS, as well as signaling enduring constraints or high levels of losses because of distance from demand centres. The logical place to start in setting the charges that would insure against variations in LMPs might be the annual TNUoS charges for that zone (rebated by any other fixed charges that would be added back), if this were done on an annual basis. However, before deciding on this approach, one would need to be reassured that the TNUoS charges were efficiently determined, and there are good reasons for being doubtful that this is the case, discussed in the next section. A preliminary but important step before offering long-term contracts would be to move to a set of defensible TNUoS charges and an improved methodology (or at least set of principles) under which they could be predicted.

If (as might be the case for investment decisions) a longer term guaranteed price were required, this could be based on a realistic estimate of the likely future LMPs and other location specific charges. The advantage of contracting on the basis of forecast TNUoS is that they are already accepted as legitimate charges under National Grid's Price Control and so would mesh nicely with the regulatory regime. However, TNUoS is based on Investment Cost Related Pricing (ICRP), which has a number of problems, among which is that it is based on the fiction that the network can be instantly reinforced to accept increments of generation and load, and so will underestimate any congestion before the upgrade is commissioned. That might be fair given that it shifts the cost of delays to shareholders in National Grid and then on to consumers, but it may encourage over-hasty connections under "Connect and manage".

6,2 Investment Cost Related Pricing (ICRP)

ICRP assumes the system can be adjusted optimally given the existing way-leaves, and as such it sets TNUoS charges that amount to the average cost of connecting all that generation in the zone. That has the immediate effect that existing generation is liable for increased charges whenever transmission to their zone has to be increased by links that are of higher average cost per MW, as would be the case if new links were undergrounded or put off-shore, as has been proposed for reinforcing links within Scotland and to England. It has been criticized on these grounds by SSE (2010), who also claims that ICRP was designed for the highly meshed England and Wales transmission system (275/300kV) and is unsuitable for Scotland that is more radial and lower voltage (132/275kV) with more single circuit connections. SSE particularly objects to the implied ICRP TNUoS for connections in the Western Isles of £76/kW, in Orkney of £42/kW and in Shetland of £100/kW. As all these connections would presumably require off-shore DC links that are likely to be expensive, these criticisms are not obviously valid, as such costs may indeed be the extra costs of accommodating extra surplus generation in those locations.

SSE's more telling objection is that if NGET were to increase export capacity to England through offshore DC links ("bootstraps") then the ICRP for the North of Scotland would double, as noted above in the case of Peterhead power station. This is where the averaging looks inappropriate. If incumbent Scottish generation would not on its own justify this expensive off-shore reinforcement, then it is not obvious why it should be charged extra when it is built. Had generation enjoyed long-term contracts the prices presumably would have reflected the status quo ante, and new generation would have had to pay the incremental cost of the extra export capacity – and that might have given a clearer cost-related locational signal. Such problems would be largely avoided with correctly calculated deep connection charges.

It might be argued that these deep connection charges would massively discourage wind generation, which it has also been argued will have to be built in Scotland and her islands if the Government is to meet its 2020 renewables target. Fortunately that need not be the case, as under the Electricity Market Reform, new low-carbon generation would be offered long-term contracts, and these contracts would cover all the costs needed, including extra deep connection charges.

There is another reason why deep charging appears more suitable than the present method. Some types of generation (presumably nuclear stations) require more transmission circuits and more secure connections than others, and under nodal pricing this might give more attractive LMPs than for those plants for which weaker links would be cost-effective. Under deep connection charges the former stations would pay more up-front (or annually via their mortgage arrangements) in return for lower transmission charges under LMP than the latter.[26]

[26] This point was made in a submission by rewneableUK to Ofgem dated 11 March 2011.

In addition to this problem of averaging, Turvey (2006, 2011) notes a number of other criticisms of ICRP, on which perhaps the most important is that it fails to properly compute the full cost of accepting more power injected in any zone, by ignoring the extra investments in transformers, switchgear and load flow devices currently excluded. As a result of this and other problems, the current methodology does not appear to give very close estimates of the actual costs incurred in making reinforcements. There are precedents for changing the model used and basing it on a more careful engineering approach. The then owner of the gas high pressure transmission system (now part of National Grid) was required to give more rapid estimates of the cost of transmitting gas from entry to exit points, and commissioned the development of a model, Transcost, that could replicate reasonably well the calculations of the very complex planning model used to design expansions.

Whether it would be worth improving the ICRP methodology if the whole basis of determining transmission charges for generation were about to change is moot. It seems likely that averaging combined with increases in the average cost of new transmission would make future G charges higher under the current methodology, so incumbent generators might be quite attracted to a set of contracts that insulated them from future cost increases caused by future costly reinforcement, but that would require a shift to deep or at least deeper charges for new entrants.

6.3 Peak pricing

The capacity of the transmission network for delivering power to consumers is likely to be driven by peak demand and it therefore makes sense to levy the marginal cost of extra capacity to the hours in which that extra capacity is needed (subject to not driving the previous peak demand below some other non-peak demand). That principle is followed by Triad charging of demand, which encourages demand reduction at these peak hours and hence signals the need for less generation (they will not be able to secure such high peak prices) and transmission (which will be sized to predicted peak loads). If demand is properly charged for transmission capacity in the peak hours, generation that is available in the peak hours will be paid a premium price (the SRMC of generation will be highest then, as will be the necessary scarcity premium needed to deliver the required reserve margin). The issues about properly incentivizing adequate generation and implications for the timing of different grid charges are considered below.

The transmission needed to cope with generation may not necessarily be sized to handle peak demand flows, as the pattern of generation may be far more unequal in off-peak periods and give rise to large off-peak flows across certain boundaries, which may need to be sized to handle these rather than peak hour flows.

As such they ought to be picked up by (suitably augmented) LMPs and correctly calculated ICRP-based TNUoS charges.

7. Recovering the short-fall in revenues from efficient charges

The arguments above suggest that the spatially varying pattern of efficient charges is likely to fall considerably short of delivering the allowed transmission revenue, so additional charges will be needed to make up the shortfall. The aim is to design a set of additional charges that gives rise to the least distortion to generation and use decisions. This is the classic Ramsey pricing problem that was developed first to characterize the least distorting tax system, and then by Boiteux when he was head of EdF to set tariffs for a natural monopoly whose marginal costs were below average costs. Ramsey pricing would include a mark-up on the efficient price (SRMC) that would be inversely proportional to the demand elasticity – higher mark-ups where demand is less elastic, lower mark-ups when demand is more elastic.[27] Thus the old CEGB bulk supply tariff collected the difference between average and marginal cost in a fixed charge, on the argument that the demand for connecting to the system and enjoying the benefits of buying electricity at its SRMC was highly inelastic, whereas generation in most hours was more price-responsive.

While it may be difficult to accurately measure the required elasticities of demand to determine these mark-ups, Ramsey principles are nevertheless useful in a number of cases, such as the bulk supply tariff for the CEGB and others considered below. In other cases, where there is no compelling evidence that elasticities vary markedly, equal mark-ups can be defended. The usual objection to Ramsey pricing is that it is exploitative, extracting maximum income from consumers who have least opportunity to evade the charge by changing their behaviour. While that may be a valid political or rhetorical point, it is precisely because the aim of these purely revenue-recovery charges is to avoid changing behaviour that the approach is economically rational. An implication is that once the locational aspects of charging has been optimally set to induce efficient decision making, the subsequent revenue-raising charges should not unduly distort those efficient price signals. That means that the revenue-raising element should not vary spatially (although it might vary by time of use, as it does with the Triad charges).

The starting point for collecting any revenue shortfall is to create a wedge between the G and L charges at each node, so that the price of taking power from a node is equal to the price paid to the generator plus the margin between the G and L prices, although again taking account of Ramsey principles as far as possible in

[27] This is a simplification that holds if demands depend only on their own price and not on relative prices. The correct general rule is that mark-ups should be chosen to lead to an equi-proportional reduction in demands – hence lower mark-ups on elastic actions.

determining when and how to insert this wedge. Clearly, for this to work it should not be easy for G and L to agree ways of avoiding the use of the grid – excessive differences between G and L charges may encourage inefficient self-generation by load unless the charges were carefully designed. Indeed, Triad charging for load means that there are incentives for load shifting through embedded generation, and penalties for its failure (the full charge would then be levied). Providing the charge for L correctly reflects the risk that the grid capacity might be called on and hence would need to be provided, the signals should be correct. Again, the design of the charge (and when it is levied) should reflect the full cost of providing the service.

7.1 Allocating transmission costs to generation and load

Creating a wedge between G and L charges immediately raises several other 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?

This is not entirely straightforward, as it depends whether the charges depend on output or are fixed (e.g. per unit of TEC). If the charges were solely on output, then in a competitive and isolated system, the proportions charged to G and L should make no difference, as the final price paid by the consumer will be the G cost plus the transmission charge. If the fraction of the transmission charge t paid by G is α, and the generator's efficient offer 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 $b + t$. Brunekreeft, Neuhoff and Newbery (2005) showed that this continues to be true even if the generators have market power and offer above their efficient price.

This argument needs some care, as it supposes that the G and L charges occur at the same moments, and may not apply if the timing of the payments differs. For example, suppose that the G charges were collected on MWh and the L charges were levied on peak demand, then there would be a difference in the hourly pattern of final prices (generation plus transmission) if all charges were levied on G (uniform per MWh) compared to them all being levied on L (high in peak hours, lower in off-peak hours). In practice G is levied on TEC, and this is a fixed charge for the generator that they have an incentive to recover in the least distorting way, following good Ramsey principles, so if that is also true of the way L affects demand, they should be jointly set at the right level and it should not matter what the division is between them.

Although this principle is useful in theory, there is an important reason why it cannot be readily applied in practice. The argument works well in an isolated system in long-run equilibrium in which all investment decisions were properly informed about the future trajectory of these charges. If, to take an obvious example, the level of G charges were suddenly changed (e.g. raised because of a massive rise in the regulatory Asset Value as a result of the planned transmission investments, or lowered if we are required to harmonise with the Continent under the Target Electricity Model) then existing generators will not be able to change their behaviour or pricing decisions (assuming that the charges are not so onerous that they cause exit). Let us consider the example of harmonising with the continental preference for zero G charges and all transmission charged to Load. Existing generators are presumably already pricing in the wholesale market to maximise their profit and these actions will not be affected by any changes in fixed costs, so the wholesale price will not change. Load will now experience an increase in peak charges, and demand might fall somewhat as a result. Whether this fall will reduce generator profits by exactly the amount they gain in reduced G charges is unclear, but given the way the Triad works (only for three half-hours and retrospectively determined) it seems unlikely. In the long-run with new entry based on the new charging arrangements, the original equilibrium should be restored, but in the short to medium run there may be winners and losers.

These problems would be avoided if the generators signed long-term transmission access charges which included whatever was their predicted contribution to recovering the short-fall on transmission, with all annual adjustments thereafter made on Load, and in contracts for new generation. As we do not have such a system it becomes important to estimate the materiality of windfall gains and losses caused by large changes in the level of G charges on existing generators. It then becomes a regulatory and perhaps legal issue whether such changes will be allowed to have negative impacts (or even whether they should be considered together with changes, such as the carbon Price Support and the ETS that give rise to windfall gains, to reach an overall assessment of the appropriateness of making a whole set of reform changes.

What we can conclude, though, is that the division of variable transmission charges (i.e. per MWh, like BSUoS) between G and L will have material consequences if electricity is traded with other charging jurisdictions that apply different G:L ratios.If, for example, one system places all the variable 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 variable G:L balance therefore becomes important in interconnected systems, and there is some attraction in levying all the variable grid charges on consumers (if only to gain export advantage).

It is noteworthy that most EU countries impose all the charges on L, and that of our immediate neighbours, France charges 100% to L and the Netherlands charges 98% to L. Admittedly Norway charges G 35% and we might interconnect to Norway in the future, but Denmark (the other partner in a future North Sea Grid) only charges 2-5% to G (ENTSO-E, 2010). As we are likely to be required to meet the EC’s market coupling demands by 2014 or earlier, this issue is likely to become very important (ERGEG, 2011).

However, such harmonization requires care in systems without the locational signals contained in LMP prices, where the correct interpretation is that the weighted average of variable G charges should be zero, preserving any locational differentials in these G charges that are required. If all variable G charges are non-locational (as at present in GB) then there is less concern about variations in fixed G charges, as these primarily impact the profits of the existing generators, not their operating decisions. If the GB price is set by imported electricity, then reducing fixed G charges just gives windfall gains to incumbents. However, if variable grid charges are modified to contain a locational element (and congestion relief is essentially a locational issue) then again it is the average such charge that matters.

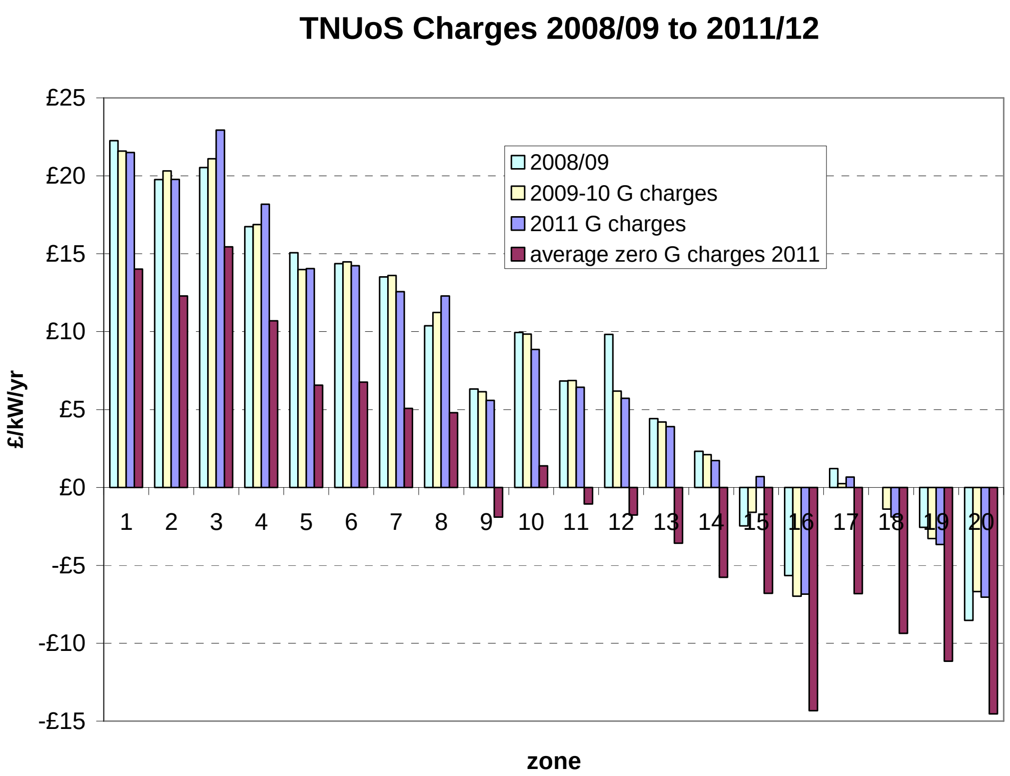

Figure 9 The effect of adjusting TNUoS G charges to average zero
Source: National Grid

The variation in fixed grid charges remains very important for investment decisions. Thus in Britain for 1 December 2010 to 31 March 2011 the annual zonal G tariffs range from £24.3/kW to -£4.9/kW, a range of £29.9/kW. Recent forward clean spark spreads have been about £5/MWh, which at 6000 hours per year, is £30/kW, so the spatial difference in grid charges is comparable to the returns to capacity and hence a strong locational signal. If the average fixed charge were set to zero, this would require some large negative G charges, as shown in figure 9.[28]

There would seem to be no reason to do this for incumbents, but if there were a different set of fixed G charges for new entrants (as under deep connection charging, which would need to be embodied in long-term contracts) then reducing the average to zero would encourage generation investment here rather than abroad (and partly offset the advantage that foreign generators have selling to us if our carbon price is higher than theirs, as it might be under the Carbon Price Support – see figure 1).

The corresponding L charges for half-hourly metered customers range from £3.84/kW (Northern Scotland) to £24.7/kW (in London), with an energy consumption charge for those two zones of 0.58p/kWh (£5.8/MWh) to 3.29p/kWh (£32.9/MWh) for non-half hourly metered customers,[29] again giving a strong variation across the country.[30] The sum of G and L charges are more uniform across the country. The zonal or nodal variation can be achieved around any average charge in theory, although it might seem easier when all (or most of) the G charges are positive. Arguably, the European Target Electricity Model provides a strong argument for actively considering zonal or nodal charges combined with deep connection charges (suitably amortized over a sensible station life) to avoid the difficulties with zero G charges as there is no reason not to charge all the cost-recovery element to Load. Note that this might also solve the apparent problem that under proposed EC rules, interconnector connections to the network can no longer be charged as generation – if the average G charge is zero then this might not be a problem. At present the zones in the south and east where IFA and Britned land have below average G charges so these would become negative, and arguably these interconnectors would be overcharged, rather than undercharged. This might be partly or wholly offset if congestion boundaries created local zonal prices where the interconnectors land, as required under the EC Target Electricity Model.

The next question is to determine the basis on which the charges should be set. Ideally, assuming there is efficient nodal pricing, these additional uniform charges are effectively taxes to recover the shortfall. As such 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

---

[28] To ensure that the plant provides useful power when needed, the payment to the plant could depend on its output at the Triad or other critical hours. Note the considerable variability in charges for some zones from year to year (although such comparisons should be made with care as there are definitional and sometimes zonal changes periodically).

[29] The quite complex formulae are set out in chapter 3 of National Grid (2010).

[30] Although as these customers are unlikely to migrate in response to zonal electricity tariffs, the variation is presumably to make them consistent with the half-hourly metered demand charges.

according to Transmission Entry Capacity (TEC) connected to the system in each year, and L pays according to demand taken at the Triad – the three half-hours of system peak demand separated by 10 days, an amount that is determined after the event. While the demand charging methodology seems appropriate (and accords with the peak pricing rule), the G (or TNUoS) charge suffers 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, 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. That means that less peaking capacity is needed, and if the rarely run plant is otherwise mainly required for balancing, then its costs will be recovered through peak pricing and balancing charges (if there is a separate imbalance market, as under BETTA) or possibly through capacity availability payments (as under VOLL LOLP in the earlier Electricity Pool). The fact that the plant is only used for peaking does not seem an argument against an annual fixed charge. The high variation in TNUoS gives peaking generation even stronger signals to locate in load pockets, and thereby to avoid high charges (and even to be paid to be available when needed).

Ramsey pricing requires the mark-up over the efficient short-run price (LMP) to be higher the less responsive decisions are to that price. These decisions may not be easy to identify, but it is nevertheless worth thinking carefully about how to identify these cases and what pricing implications follow from that. The key issues will be the choice of type of generation. In-merit base-load firm power (low variable cost fossil and nuclear generation) will select Transmission Entry Capacity (TEC) equal to its maximum exports to the grid, as it will be running at that capacity much of the time, and it will therefore be efficient to charge annually for the TEC. Higher variable cost plant that is near retirement may be sensitive to the level of annual TEC charges, as if these are set at the same level as other base-load generation, they may leave too little remaining profit to justify remaining open. If they are paying their full variable costs (for locational injections and for all environmental harm) then any contribution they make to cover the short-fall in transmission revenue is worth having, and hence any positive charge is better than losing that revenue source.

Put another way, low cost base-load plant has a very inelastic demand for TEC, as not to buy it would mean giving up high profits, while plant on the margin of closing has a very elastic demand for TEC, as a small increase might make it cease to operate at all. On the Ramsey rule the former should pay more than the latter, which should be set at no higher than the level that induces premature exit. One solution is for the generator to sign a long-term fixed price contract for these short-fall charges for an agreed period (related to the expected life of the plant). These charges, if levied annually, would then be an inescapable liability and would not affect exit decisions (at least, assuming the plant owner remained credit worthy or had posted a bond).

An annual TNUoS charge proportional to the TEC requested penalises intermittent generation that has a low capacity factor and which cannot choose whether to run if conditions (e.g. wind) are not favourable. On-shore wind power may have a capacity factor of 20-30% and so its average use of the transmission system is only that fraction of its full capacity. The question is: what is the right revenue short-fall charge to levy for such generation? The answer will partly depend on the efficiency of dispatch. Thus if all generators face LMPs, and if the system of renewables support does not distort operating decisions (e.g. takes the form of the proposed Contract for Differences or Feed-in Tariffs under consideration in the Electricity Market Reform consultation[31]) then wind will displace almost all generation with any but the lowest variable costs, as the variable cost of wind is largely O&M, with no fuel cost. If there is no more wind capacity behind an export constraint than thermal plant, then presumably no extra transmission capacity is required, so the only issue is how to levy the additional charges (to make up the revenue short-fall, which are uniform across the country) in ways that does not discourage the efficient level of wind generation. In one sense it may not make much difference, if any discouragement is compensated by higher subsidies to meet particular renewable targets, although the extra subsidy itself may be more or less distortionary than collecting it via grid charges.

If the shortfall is levied on TEC uniformly across the country it will not lead to incorrect wind farm locational decisions, but it will make wind seem relatively more expensive than if, for example, the shortfall were levied on average capacity, or on the deemed firm capacity that the wind farms offer (the extent to which other generation can be scaled back and still maintain system reliability and quality). That should not mean this it discourages wind and hence is at variance with the requirement to facilitate low-carbon and renewables connection, because it does not affect the underlying costs of operating the system and supporting renewables, only the explicit subsidies that wind would need (rather than implicit subsidies through cross subsidies through transmission charging).

---

[31] Available at http://www.decc.gov.uk/en/content/cms/consultations/emr/emr.aspx

Once wind requires extra transmission capacity then there is a case for charging the deep connection charges caused (if that principle is generally accepted for all new generation) or possibly the ICRP (bearing in mind its limitations noted above) on this extra transmission, in proportion to the wind capacity that causes it (allowing of course for the amount collected via LMPs). Such a system would encourage wind to locate where it can displace thermal power without requiring extra grid reinforcement, bearing in mind that the low land cost and high wind availability may still make remote locations attractive even with higher grid charges (lower LMPs). Again, under the EMR proposals for long-term contracts for renewables, deep connection charging would merely change the place that the costs show up, changing the balance that is collected from grid users (i.e. consumers) to that collected from renewables subsidies (which will probably be levied on those same consumers).[32]

7.2 Disconnection charges

At present, and consistent with the philosophy of shallow connection charging, generators are only charged for the amount of TEC requested, which falls to zero when they disconnect, leaving only a liability for any payment still owing from the past. But disconnecting generation from some nodes may require extra investments in generation or transmission to retain the same security and reliability standards, and this raises the question of the appropriate design of the connection agreement in the first place (or for new generation). The format of other contracts can illuminate the issues raised. Thus an option to export over an interconnector, without the requirement to deliver, is more costly as it prevents an offsetting import over that interconnector that would release more export capacity (netting the flows). Similarly, an option to inject power into a grid if available is more costly to the TSO than accepting an obligation to remain for a specified number of years and to be available to generate when required (i.e. when the nodal price or value exceeds the cost of supply). One can imagine generators being offered either contract with the firm contract being cheaper but containing termination penalties. Just how valuable this would be to a TSO looking over long time periods with uncertain future fuel prices, technologies, and demands is unclear and might be small for most plant, but might be important for plant designed to provide back-up or reserves in specific locations – although that might best be handled by be-spoke contracts with the SO who needs

[32] Socialised transmission charges for extra costly grid reinforcements to connect wind have the advantage that they would likely be charged on Ramsey principles, which tend to be efficient though perhaps rather inequitable. Placing the renewables subsidies on consumer bills might be worse on both scores, while collecting the required revenue from general taxation would be economically better but politically worse.

these services (i.e. capacity availability payments for the option the SO can exercise of calling on that generation when needed).

7.3 Guiding transmission investment decisions

LMPs offer useful information about the desirability of grid investment, and some have argued that it allows the transmission system to become contestable, with grid expansion expected to be provided by merchant investment. There are considerable doubts about this, not least because of the problems of cost recovery identified above, there is a further potential problem arising from Braess' paradox if the existing system is not fully contracted with TCCs (Bushnell and Stoft, 1996). Braess' paradox states that a profitable investment connecting two nodes in a meshed network can reduce the overall efficiency of the whole network and thereby reduce social welfare, as illustrated in Appendix B. Nevertheless, even if merchant investment may not deliver all socially valuable transmission investments, the LMPs provide useful information to the TSO, the regulators, and other parties who might contest for the right to a regulated short-fall revenue, and as such these price signals are valuable. Parail (2010) argues, on the basis of evidence from Norned and simulations for Britned, that merchant interconnectors can deliver most of the potential gains from interconnecting different jurisdictions, but these interconnectors are DC links that avoid the problems of loop flows that can cause the Braess Paradox.

The issue of interconnectors provides another potentially powerful reason for nodal pricing, particularly where these are merchant undertakings. The fact that two regions (e.g. the Netherlands and Norway) have sufficiently different prices to justify the investment does not necessarily mean that the relevant nodes at which the interconnectors land will have the same LMP differences, and that is what should guide the investment decision. Indeed, constraints within the country may limit the ability of the interconnector to export or import power, and/or may force expensive reinforcement investment by one of the countries (although typically regulators have to give permission for the line and could determine whether it were still economic given the extra costs to be incurred, which might depend on whether it were able to levy an appropriate charge to cover any costs).

8. Should generation lead transmission or transmission lead generation?

There is an important philosophical difference between the "planned" and "market-led" approach that emerged very clearly at the Ofgem discussion of this and the other reports on 4th April 2011. Very roughly the "planned" approach would be closer to the old CEGB approach in which transmission and generation are planned together to minimise the total system cost of delivering power to consumers. If one considers that the "guiding intelligence" is or should be located in the TSO function, then such a body (or better still, in an independent Transmission System Planning Authority or TSPA), then its task would be to work out where the optimal location of all new generation should be, given the feasibility and cost of connecting them to the transmission system. It would then indicate sites at which such generation would be granted access, and might reinforce this message by attractive nodal prices and/or connection charges there and unattractive prices elsewhere).

The "market-led" approach would allow generators to choose where to locate, preferably informing the TSO in advance, and the TSO would then respond with suitable reinforcements and new transmission lines to accommodate this investment. It would be market led in the sense that the TSO would indicate the pricing implications of any locational choice by new generation. Each has its merits and limitations. The first approach assumes that the TSPA can collect as good generation siting information as developers, and/or has a comparative (possibly legal) advantage is securing site planning permits. It also assumes that transmission decisions, for which planning approval is particularly difficult and time-consuming, are best made with a system-wide view of where such investments are feasible.

The market-led approach is based on the belief that developers are better at locating viable generation sites, and that they can put pressure on the TSO to respond more actively to overcoming transmission problems than if these are left to the initiative of the TSO. If so, that is a compelling argument itself for a TSPA. Whether there is any practical difference between the two philosophies in the UK which finds it difficult to build anything anytime soon anywhere is a moot point.

9. Summary of high level principles

To summarise, in an unbundled and liberalised electricity industry, in which the market structure and the various support schemes for low-carbon and renewable generation are efficiently designed, the high level principles for transmission access pricing are that they should encourage:

- the efficient short-run use of the network (dispatch order and congestion management);
- cost-effective methods of hedging risks facing generation and load;
- 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, wind, etc.);
- cost-recovery; and
- legality, fairness and political feasibility

Efficient short-run prices for accessing the transmission system imply LMP or nodal pricing. Cost-effective risk hedging instruments involve a trade-off between liquidity and re-dispatch cost. Liquidity requires a single strike price for an adequate number of generators and consumers, and that requires nodes to be aggregated into zones, but the larger the zone, the higher will be the internal re-dispatch costs. Efficient generation investment decisions are likely to require some further locational adjustments to any short-run price signals, and both are likely to be best delivered by a long-run contract for desired TEC– an FTR or TCC – that would likely be set at a properly computed deep connection charge. These will not generate sufficient revenue to cover the regulated revenue allowed, and will need to be supplemented by additional charges to make up the short-fall set on Ramsey principles. These cost-recovering grid charges risk distorting choices of plant type and operation unless carefully designed to minimise the change in behaviour away from the efficient choice. Peak load pricing of demand and access pricing for controllable (i.e. not intermittent) generation that is not in danger of exiting satisfy these conditions. Otherwise there is a case for charging on the basis of effective capacity, rather than TEC, provided the entrants pay the excess of deep connection charges over any other payments they make. There is also a case for new generators signing a long-term contract setting out their liability for future charges and their associated FTRs, that would decouple payments by generators that have already connected, and future charging regimes, to avoid legacy rights restricting future changes.

PART 2 TRANSMISSION PRICING IN A SECOND-BEST WORLD

10. Dealing with market and government failures

Transmission pricing is only likely to support efficient decision making if other relevant markets give efficient price signals. When it comes to generation there are important market failures that need correction, as most fossil generation emits air pollutants like SO_2, NO_x, particulates, and CO_2. Ideally all these harmful emissions should be charged at their marginal social cost, but in the EU only CO_2 is currently priced under the EU ETS. Unfortunately the ETS is deeply flawed and the current CO_2 price is too low, unstable, and fails to offer assurance for investors in low-carbon technology (DECC, 2010c). The other pollutants are dealt with by setting emission performance standards. The Large Combustion Plant Directive will force non-compliant generators to close by 2016.

The Budget of 23rd March 2011 announced the Carbon Price Support (CPS) that is intended to correct the under-pricing of CO_2 (see figure 1) and new plant will clearly have to be compliant with the other emissions standards, which leaves the way in which renewables are supported as a potential government failure with market consequences.

10.1 The 20-20-20 Renewables Directive

The EU 20-20-20 Renewables Directive requires the UK to deliver 15% of total energy from renewable sources by 2020. The least-cost way of achieving this is to deliver a high (30-35%) share of electricity from renewable sources, largely wind. The logic of the 20-20-20 Directive is not to reduce the EU's CO_2 emissions, whose level is already predetermined by the ETS cap. Instead it is primarily designed as a demand-pull instrument to encourage a substantial increase in investment in renewable energy, which is expected to lower the cost of future renewables through learning–by-doing, as illustrated in figure 10. Each doubling of installed capacity results in a cost reduction of a certain percentage, which might be expected to continue with further doublings of capacity, perhaps at diminishing rates. Individual manufacturers of the new technologies cannot fully internalize learning externalities, which spill over and enable other manufacturers to build future capacity at lower cost. As such the learning induced by the renewables investment demand is a public good and therefore a legitimate reason for public support. The case for EU action is that if successful it will encourage other countries to adopt these technologies when their costs fall sufficiently and mitigate CO_2 emissions with universal benefit.

The Directive shares the burden of the extra cost of supporting this deployment across Member States by specifying targets for each that balance their ability to deliver the targets at least EU cost with their ability to pay – richer MSs will be responsible for a higher share than would be indicated on overall cost minimization arguments.

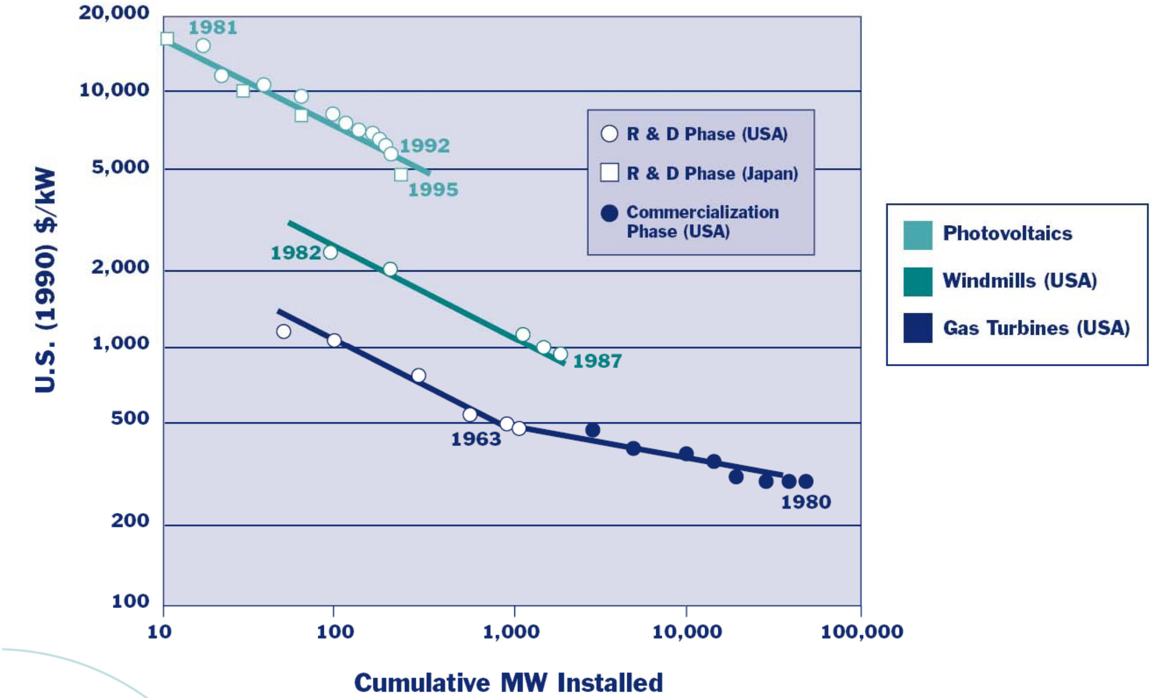

Figure 10 learning curves for generation technologies
Source: Nakicenovic et al (1998).

The relevance for GB transmission access is that much of the anticipated transmission investment is required to allow the large volume of on- and off-shore wind generation to deliver its power. The logic behind the 20-20-20 Directive is that the key to achieving the learning benefits is largely in the production of wind-turbines and their installation in favourable locations. That includes better modelling and predictions of local wind conditions leading to better siting, improvements in site management, as well as learning how to negotiate local opposition, including the planning hurdles. Provided the installations are reliable and that operations and maintenance (O&M) costs have been efficiently reduced, there is no obvious reason to subsidize their operation, as the learning benefits that justify their support lies in the steps leading up to and including installation, not subsequent operation. The logical form of subsidy for on-shore wind would be an agreed payment per kW of *available* capacity.[33] In the case of wind located behind an export constraint the logical

[33] The availability payment may be designed to be higher in more valuable hours, and lower in less valuable hours, as was delivered under the Pool capacity payment scheme that paid in proportion to the Loss of Load Probability, and/or may be restricted to unconstrained hours, and may take account of location to extract rent, as discussed below.

form of capacity payment would be to concentrate them on unconstrained hours, up to some maximum level.

The Electricity Market Reform proposes that all low-carbon generation will be offered long-term contracts, although it is silent on how these will be administered. Logically the form of contracts should be tailored to suit the specific technology and its location, and that will be the working assumption in what follows. It then becomes necessary to distinguish between the counter-party to these contracts (the Non-fossil Fuel Contracting Agency, NoFFCA), who would pay the generator the contract price but receive the relevant energy price and be liable for transmission charges, and the generator, who would receive the contract price.

In the case of on-shore wind, NoFFCA would receive the spot LMP electricity price and if centrally dispatched would offer at its avoidable O&M cost, and receive the spot price whenever this were greater (almost all the time), plus the availability capital subsidy payments. This could be easily de-risked through a standard Power Purchase Agreement (PPA), which would specify the payment per unit of available capacity and would pay the avoidable operating costs (normally fuel and O&M, but here fuel costs would be zero), or less ambitiously through a long-term FTR. NoFFCA would then offer a Feed-in Tariff (FIT) to the on-shore wind farm with the right incentive properties (availability, production when profitable) but with sufficiently low risk to enable the developer to finance the project almost entirely from cheap debt, and with just enough profit to induce the development, but no more.

While the current Electricity Market Reform recognizes the need for long-term contracts to de-risk nuclear power and renewables and also to provided additional support to renewables, the way in which that support is provided may interact with the transmission access regime. This matters because project TransmiT also supports the same goal of reducing carbon intensity and so must take account of the way in which the market reforms support carbon reductions (and vice versa – transmission charging should inform the contract design details). It raises the delicate question of whether it would be desirable to introduce intentional distortions into the transmission charging regime to offset distortions that for various reasons cannot be changed in the renewables support regime (the standard "second-best" problem).

If the 20-20-20 Directive had been properly designed it would have required the delivery of a specified amount of renewable energy *capacity* rather than output, to the extent that the learning benefits are almost exhausted once the capacity is in place.[34] For immature technology like off-shore wind and any marine power, where

[34] This would have been more complicated as different technologies have different capacity factors, and thus offer to the rest of the world differing opportunities of the ultimate goal of decarbonising electricity. If the EU had pursued this line of argument it might also have had to address the issue that

there remain considerable obstacles to overcome to ensure reliability at acceptable cost, there is a case for supporting output as well, as that will incentivize learning how to deliver more reliable operation. The evidence from early off-shore Danish wind farms (Horns Rev 1) suggests that early reliability problems were severe and considerably more durable (and expensive) turbines and gear boxes were required to deliver lower life-time costs. On-shore wind appears to have now reached a high level of operational reliability, and so it would seem more appropriate to reward capacity availability.

Supporting renewable capacity availability rather than output would reduce the high incentive to locate in windy but inaccessible locations, because the current ROC support scheme over-rewards output relative to capacity, and interacts with the current "Connect and manage" access regime, described in the next section.

10.2 "Connect and manage" and current arrangements

After the earlier Transmission Access Review (Ofgem, 2008), DECC and Ofgem introduced an interim "Connect and Manage" until an "enduring" solution could be finalized. Parliament approved powers in the *Energy Act 2008* to enable the Secretary of State to intervene where the existing governance arrangements were deemed unable to reach a satisfactory solution to the problem of facilitating timelier grid access for renewable electricity. A decision to use those powers was announced in July 2009, and the Government's preferred *enduring* grid access reform ("Connect and Manage Socialised Cost model") was published in July 2010 as DECC (2010a). Changes to the industry codes and licences were to become effective from 11 August 2010, after which the Government will notify the European Commission that these changes amount to a Public Service Obligation justified as in the public interest to support climate protection. DECC (2010a) noted that

> "Ofgem raised several other concerns about the preferred Connect and Manage model including their view that other models of access reform had not been fully assessed, that DECC's targeted intervention would create uncertainty as it would leave a number of complex issues outstanding, that charges faced by generators could be volatile and unpredictable and that the model might not deliver the achievement of the Government's carbon targets. In particular, Ofgem considered that the proposed approach was unlikely to help offshore wind generation connect earlier given its view that offshore works were unlikely to be completed much more quickly than wider reinforcement works. It also expressed concern that overselling of capacity may mean 'constraining off' some low carbon generation. Both Ofgem and

---

not all renewables are equally valuable to develop, as their prospects and ultimate resource base differ widely. The resulting Directive was arguably the simplest compromise to deliver EU-wide support to drive down the cost of renewables.

Consumer Focus felt that the proposed model would not provide clear signals for Transmission Owners to identify areas of the network requiring reinforcement."

DECC argued that their commissioned analysis by Redpoint suggested that although other variants including locational BSUoS could result in lower costs, their extra complexity could not justify the modest savings. Although the costs would be higher to consumers under their socialized model, the extra costs were only 20p/household per year, although with 28 million households (who only take about one-third the total supply) that is not necessarily all that small.

Clearly socializing the costs under the Connect and Manage Socialised Cost (CAMSOC) model introduces additional distortions, and the next step is to see how they interact with the particular problems of supporting renewable electricity as required by the 20-20-20 Directive. In fairness to CAMSOC it is worth asking how it compares with the alternatives suggested here, and whether any adverse effects could be offset through appropriate contract design under the EMR.

Consider the case of an existing fossil generator, G, of 1 GW with a MC of £20/MWh located at node A in a region with no demand as shown in Figure 11. A wind farm, W, with a peak output of 1 GW then connects to the same node A, and is provided with a FIT paying £70/MWh. The export link is upgraded to a deemed optimal capacity 1.2 GW.

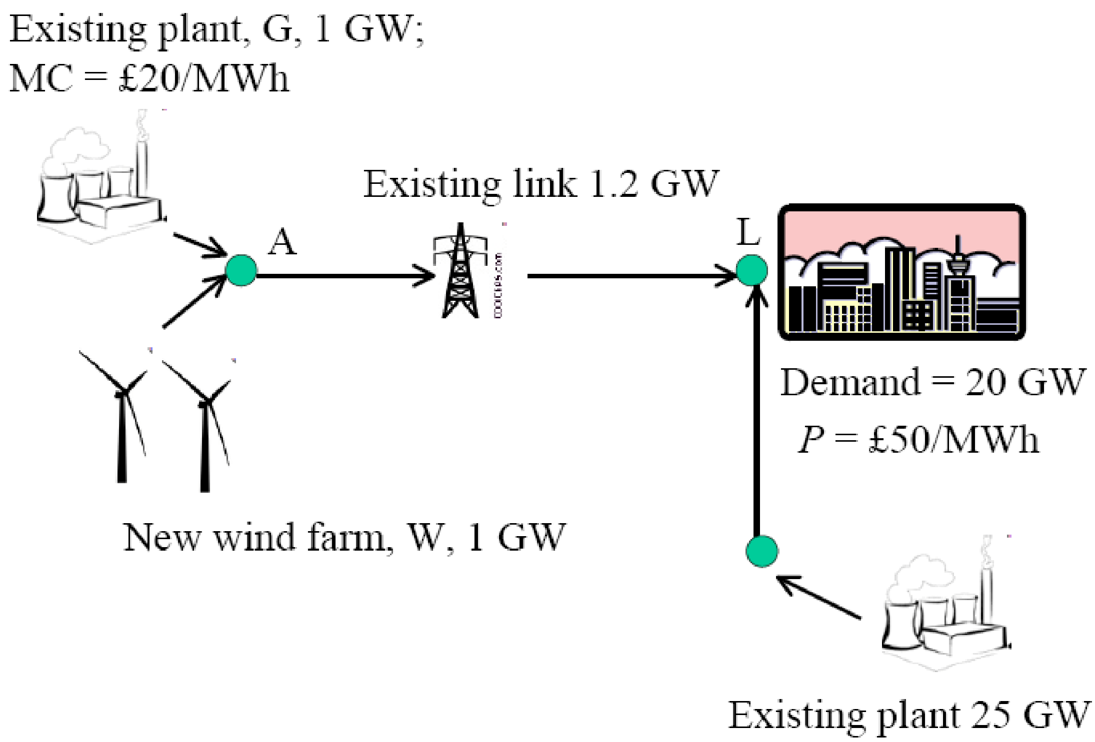

Figure 11 Connecting a wind farm behind a transmission constraint

Suppose for convenience the price in the demand zone L is always the same at £50/MWh. Suppose that the existing generator G has a long-term FTR on the link of 1 GW and a 1 GW CfD with a strike price of £50/MWh, and we have nodal (in this case the same as zonal) pricing. Now consider three scenarios – (i) no wind, (ii) low wind (output of the wind farm = 250 MW) and (iii) high wind (1 GW). In all cases with complete contract cover G offers its 1 GW of power at node A at £20/MWh.

In case (i) the transmission link AL is unconstrained, and the nodal/zonal price is £50/MWh, the FTR pays nothing, G delivers 1 GW over the link and receives £50/MWh from the CfD and has no transactions with Load, giving a profit of 1000 x £30/MWh = £30,000/hr.

In case (ii) the link is constrained, the nodal/zonal price falls to £20/MWh; G generates 950 MW, and makes no profit or loss on generation. G receives £30/MWh on the 1 GW FTR, delivers 950 MW to Load and buys and sells the balance at zero net cost, making profit of 1000 x £30/MWh entirely from the FTR. The wind farm receives £70/MWh on 250 MW and the counterparty, NoFFCA, would effectively pay a subsidy of £50/MWh.

In case (iii) the link is fully constrained, the nodal/zonal price falls to £20/MWh; G generates 250MW, and makes no profit or loss on generation. G receives £30/MWh on the 1 GW FTR as in case (ii) and buys and delivers the entire CfD volume at zero net cost, with exactly the same result. The FIT counterparty meets the subsidy cost of 1 GW x £50/MWh.

If we had a BETTA type arrangement in which all nodes received the same price of £50/MWh, then G does not need an FTR. If unconstrained in (i) there is no difference. In (ii) the constraint bites and G bids very low (just above zero) to be constrained off at £50/MWh, making £50/MWh profit on its constrained off volume (50 MW) and receiving £30/MWh profit on 950 MW, making an excess profit of £20 x 50 = £1,000/hr. The same is true in (iii) except that G is now constrained off for 800 MW and makes excess profits of £20 x 800 = £16,000. With a wind capacity factor of 30% G increases its profits by 16%, and that extra cost to consumers. CAMSOC is therefore potentially considerably more expensive to consumers than zonal pricing and congestion management as set out in ACER (2011), and which may be required for GB congestion management if accepted by the EC.

The inc-dec game being played in this example could be avoided if the generator could be required or induced (e.g. by contract) to always bid is marginal cost into the balancing mechanism. In that case the generator would be paid the necessary balancing-down payment, which ought to be the difference between the relevant price (e.g. the zonal price in a Target Electricity Model-compliant update to

BETTA) and its marginal cost, making up the lost profit it would otherwise have enjoyed. This would have been the LMP in the LMP model (assuming in both cases honest information is made available to the SO), and had the generator held a FTR, it would have received the same sum, equal to the zonal or reference price less the LMP. Of course, in both cases there are issues with ensuring that the constrained off payments or determination of LMP are based on the correct generation cost (SRMC), although whether one or other system (CAMSOC or LMP) encourages more market manipulation is clearly relevant to choosing between them.

That leaves the main potential distortion of CAMSOC that it encourages over-hasty wind building in some locations, when perhaps it would have been better if they had chosen to enter somewhere else. If contracts are offered that pay amounts that are site-specific (as they should be to avoid unnecessarily over-rewarding wind in favoured locations) then it should be possible to guide location choices, again given good information, supporting the case for a TSPA to be involved in contract design.

10.3 Complications caused by the 20-20-20 Directive

The Impact Assessment for Connect and Manage (DECC, 2010b) estimates that the extra congestion costs caused by locating in Zone R1 (the northernmost Scottish zone) allocated to those new wind generators causing the congestion, would vary between £37 and £41/MWh up to 2017, falling thereafter to £35/MWh for 2017-20, compared to zero in many other zones. If wind farms had to pay the extra costs they visit on the system, as they would under LMP (amplified by deep connection charging in the presence of lumpiness), then they would almost certainly choose not to locate in such highly constrained zones, as the value of the extra generation valued at spot prices rather than spot + ROC prices (which figure 12 shows almost double the price received) would not cover these congestion costs. Assuming the capacity incentive were sufficient to deliver the target rate of renewable investment, this would then be delivered at lower cost by reducing the demand for potentially uneconomic transmission investment.

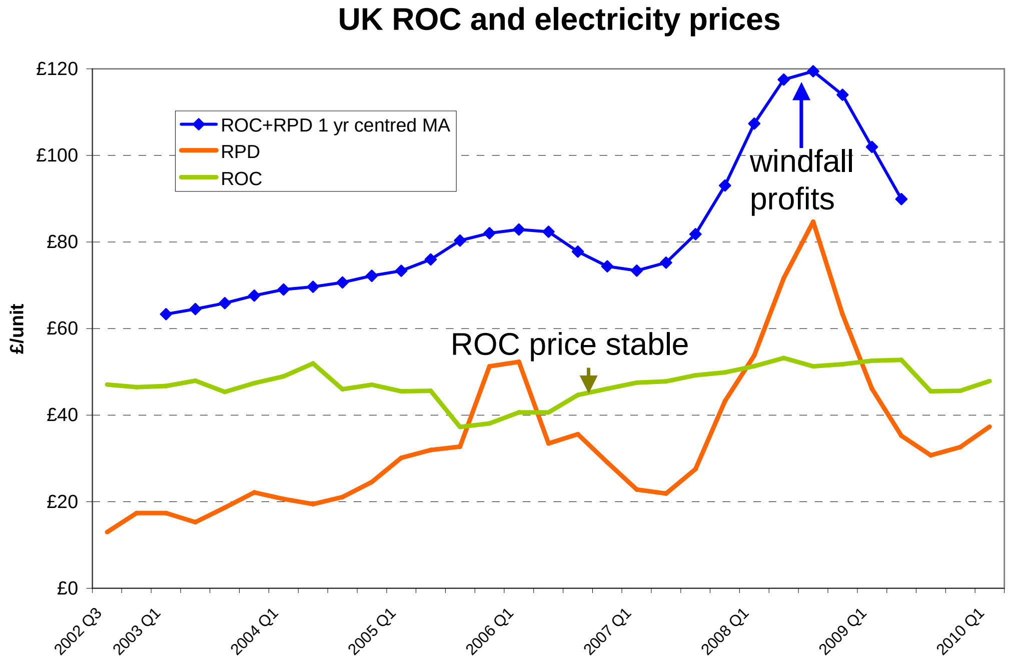

Figure 12 UK prices faced by on-shore wind farms
Sources; UKPX, Ofgem

The Impact Assessment observes that the form in which the Connect and Manage regime would operate under the variants considered (but *excluding* full constraint cost charging) will charge less for any constraint costs than the extra revenue earned from a windy distant location, and therefore will not affect location decisions, but only the division of income between different wind farms and generators and consumers. Even if that were correct (and the next paragraphs considers reasons for doubting this) these revenue allocation choices are material considerations. The energy policy objectives include affordability, so the cost to consumers is relevant. The approach proposed in GB is almost the exact opposite of that in Germany, where FITs are adjusted so that wind farms in windy areas receive *less* than those in less windy areas. The aim is to offer a just sufficient FIT to make the investment more attractive in better locations without handing out unnecessary rents to those in such locations, which would otherwise raise the cost to consumers. The proposed CAMSOC approach is to socialize all additional balancing (i.e. congestion) costs so that those in windy locations will earn more and be cross-subsidized by those in less windy but unconstrained locations.

Figure 13 illustrates the interaction between the forms of transmission charging (and investment decisions) and the way in which renewables are supported.

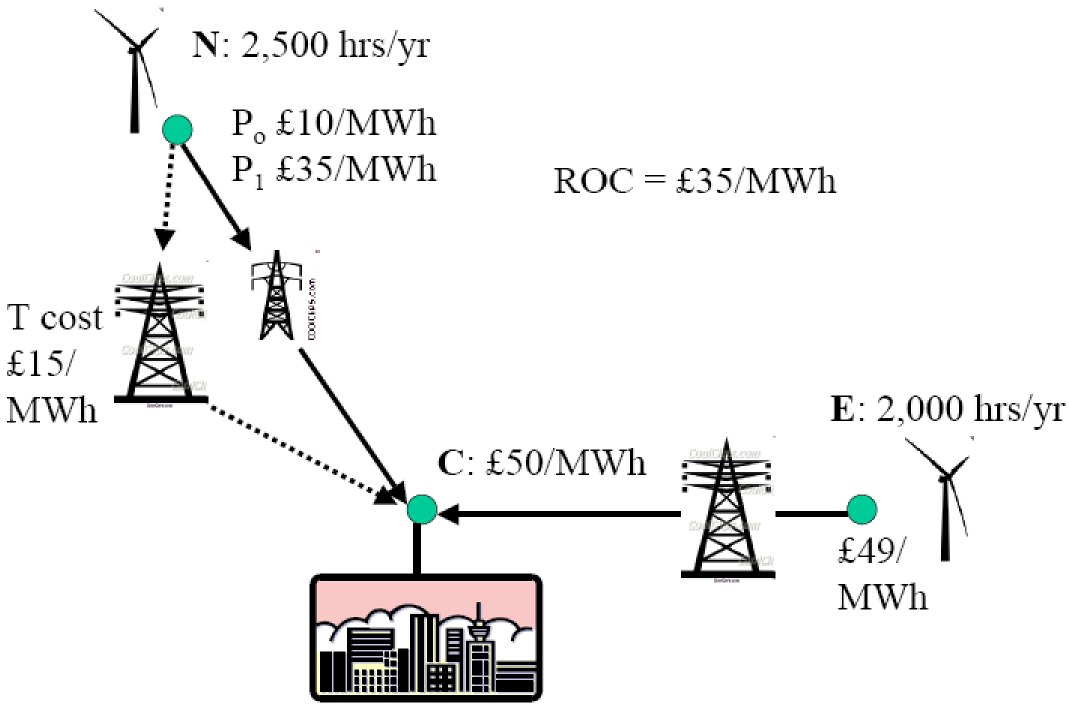

Figure 13 Location choices under LMP and spot pricing for wind

If a wind farm locates at E (right) it can use an existing strong transmission link and under LMP receive £49/MWh and would run 2,000 hrs/yr. If it were to locate at N, which currently has a weak congested link, the nodal price would be £10/MWh but at N it would enjoy more wind and operate 2,500 hrs/yr. If the ROC price is £35/MWh, the wind farm at E would receive £168/kW capacity per year and at N would receive £112.5/kW per year, and so even with ROCs but LMP would still locate at E. If congestion costs (and losses) were socialized (and for convenience fall on consumers), it would earn £40+35/MWh and £150/kW/yr at E, but £40+35/MWh and £187.5/kW/yr at N, and so would locate there, as it would even without ROC payments (£80/kW/yr at E vs. £100/kW/yr at N).

Suppose now that a stronger link can be built to N, which would cost the equivalent of £15/MWh for the new wind.[35] If this cost were visited upon the new wind at N (LMP, possibly with deep connection charging) the effective revenue at N would rise to £35/MWh or £87.5/kW/yr (without ROCs) compared to £98/kW/yr at E, and that would signal the socially correct decision not to invest in either transmission to N nor wind farms at N. Shallow transmission charging with LMP would almost certainly lead to a higher nodal price at N and thus signal an apparently preferable location for the wind farm, while any ROCs or socialization

[35] The numbers are purely hypothetical and it may be that the investment costs of transmission to northern Scotland are lower than the correctly valued additional wind output there.

would overwhelmingly favour N (and under CAMSOC, too soon, before the transmission upgrade is in place).

Thus the way in which renewable are supported and any distortions introduced into their pricing will interact with the way in which their transmission connection is managed and charged. It is an interesting question whether the UK should accept the flawed implementation of the 20-20-20 Directive (supporting output not investment) and optimize within that, or seek the underlying benefits intended and use them in its support and charging design. It might be worth exploring the possibility of seeking a reinterpretation of the Directive that allows capacity calculations rather than output in estimating compliance. The EMR should therefore consider how the long-term contracts it proposes might best be designed to deliver efficient amounts and locations of wind at least cost. If it does so, then there is considerably less pressure placed on adapting and possibly distorting transmission charges in order to counter the perverse locational incentives of supporting wind energy production (as under the ROC system).

Either way, and whatever final decision is made about transmission charging, there is an important question about how the Transmission Owner should be incentivized to invest in transmission, and specifically what prices to use in determining the value of any investments – prices inclusive of ROCs, or without, or the prices that might be included in contracts. There are already precedents for using shadow prices (e.g. for CO_2 in UK appraisals of utility investments in the water industry) and there are reasons to doubt that the ROC-inclusive electricity price is a good measure of the value of renewables output facilitated by a transmission investment. Indeed, the intention of phasing out ROCs and replacing them by long-term contracts under the EMR is a recognition of that fact.

11. Pricing in the absence of LMP charging

It may be that GB could more readily/cost-effectively adopt market splitting on the Nordic model rather than full nodal pricing, and this would be agreeably consistent with the European Target Electricity Model, which envisages market coupling with price zones defined by congestion, rather than national, boundaries (ACER, 2011). This would be a close approximation to nodal pricing if there are only a few major transmission constraints and if the nodal prices within each split area were close to each other, but that is a testable hypothesis that should not be pre-judged. There are other factors to take into account when deciding on the granularity of locational pricing, for as noted above, the smaller the price zone (in the limit at each node), the fewer participants will be interested in trading contracts at that zonal price, so contract markets will be very illiquid. The larger the zone, the more participants and the greater will be contract market liquidity. Higher liquidity reduces transaction costs, enhances competition, creates greater price transparency and credibility, assists entry, and has thus genuine value to set off against the extra re-dispatch costs of larger zones. Nodal pricing is likely to leave would-be contracting parties unhedged and exposed to basis risk, which is the difference in the local and reference prices.

Whether this is a material problem is again an empirical question. The Nordel market defines a System Marginal Price (SMP, i.e. the price ignoring all transmission constraints), and this is the standard market price used in Contracts for Differences (CfDs). Generators and Load in each zone can buy CfDs between their zone and the SMP, but these are not very liquid and have higher transaction costs, so they may find them unattractive and choose not to be entirely hedged – they will remain exposed to basis risk. If that average price difference is predictable and deviations from that average modest, then the cost of this basis risk is small and the problem unimportant, if not then it may be desirable to define a number of reference prices to reduce that risk or find ways of reducing the transaction costs of the interzonal CfDs. One problem in Nordel is that these inter-zonal CfDs are not issued by the TSO, the natural counter-party to interzonal flows, and that may increase their transaction costs as the issuers lack the offsetting revenue from the interzonal price differences, and so they are bearing the basis risk.

If, for various reasons, it is thought to be impossible to adopt market splitting or LMPs (and they are widely used elsewhere, as well as greatly facilitating efficient trade over interconnectors) then the entire burden of signaling where to locate will fall on the annual TNUoS fixed charges, and we are back to the present methodology based on zonally varying Investment Cost Related Pricing (ICRP), supplemented by a uniform charge to recover the short-fall from the allowed revenue. In that case while investment decisions may be given some guidance (and longer-term contracts would be desirable for some agreed level of TEC around which annual adjustments at the annual charging rate can be made to improve this guidance), short-run dispatch decisions are likely to be distorted, as discussed below when considering the particular issues that arise with intermittent wind.

If all the locational signals are to be provided through TNUoS then the way in which the charges are determined might need revisiting, for at present they are derived from the ICRP methodology that has potentially serious shortcomings discussed in section 6.2. If congestion is a material element for a period until new transmission investment is commissioned, then the ICRP will at best provide the correct locational signals after that date and before then perhaps an additional congestion element should be added. This would be a change of methodology that would disadvantage incumbents, and since it would make no difference to their location decisions (but might affect their disconnection decisions), any congestion element should be combined with a contract for the requested TEC, with congestion charged above that.

If new connections were confronted with this congestion augmented TEC, then the wind farm might choose to delay until the transmission were built, although the charging would seem both discriminatory and hostile to new renewables. If the developer were choosing freely between this and less constrained locations, and if the choice would not affect the rate of renewable investment, then encouraging a different sequence of development would deliver the same rate of investment at lower cost. If the developer just delayed and as a result the amount of renewables was lower, then it might be more costly to stimulate the extra renewables to make up the short-fall.

Another way of justifying the current ICRP plus CAMSOC is that by giving a subsidy to investment (in some distant constrained locations) it moves the form of support for the renewables closer to the efficient format of supporting capacity rather than output (but the example above shows that it can also lead to mistakes). It also puts the cost burden for planning and other delays on voter/consumers, which might seem fairer and might even encourage them to change the planning process.

PART 3 THE CASE FOR CHANGE

The central question is whether there is a case for changing the GB transmission charging model, taking account of international best practice and current thinking. This report has argued that current thinking supports LMP, at least as a starting point. Experience in the US has demonstrated that LMP is not only theoretically sound but practical. The US federal regulator, FERC, recommend its use as best practice in its Standard Market Design (SMD). The main contender to LMP is zonal pricing, in which zones are defined such that congestion within the zones is modest and the costs of re-dispatch similarly small and more than offset by the gains from increased contract and market liquidity. The EU is following the Nordic example of zonal pricing in its Target Electricity Model, on the assumption that within countries transmission grids are strong and congestion currently modest. The US was driven to nodal pricing by the collapse of zonal pricing, in part due to gaming (the "inc-dec" game), and this was arguably the result of a relatively under-developed transmission system between the large number of utilities. Once markets were liberalized and trade between utility areas was freed, congestion became a more serious problem and started the process that culminated in nodal pricing.

BETTA has demonstrated that Scotland should at the least be a separate price zone given the congestion at the Cheviot boundary. In addition, the expected rapid increase in on and off-shore wind makes it likely that there will be increasing strains placed on the transmission system, both in connecting the new wind farms, and delivering its power to final consumers. Congestion is likely to be more variable over time and space, and its predictability day ahead will be considerably less than with conventional generation. It may be that Europe and GB find that wind precipitates increasing congestion within its price zones, and that the natural end-point of market splitting is nodal pricing as in the US.

Increasing volumes of intermittent wind may provide a strong case for reforming the wholesale and balancing markets (something perhaps surprisingly absent from the so-called Electricity *Marke*t Reform). This could evolve without major legislative changes if the System Operator were to encourage a significant fraction of flexible generation to join a voluntary centrally dispatched pool, where the SO would also logically take control over the intermittent generation. If so, and best international practice – the US SMD – supports that reform, then there will be important implications for the way the SO procures balancing services. As these are locationally specific, and as there is a strong case that at any location there should be a single balancing price, this would support a change to LMP pricing, at least for BSUoS. To repeat, the case for central dispatch is greatly increased as the share of wind increases, as it greatly simplifies the short-term scheduling of flexible generation to maintain system balance. The SO would receive short-term and real-time wind forecasts from every wind farm and manage the system with considerably better information than is the case in a self-dispatched decentralised system such as BETTA. In addition, a pool model gives the necessary price discovery and transparency for a liquid market in Contracts-for-Difference of the type favoured by the Government's EMR.

The scale of this change should not be exaggerated, as it could evolve from the present trading arrangements by reforming the Balancing Mechanism into a proper balancing market and intra-day market. That is almost certainly required to handle increasing volumes of wind efficiently, and would simplify contracting for wind under the EMR. The counter-party to the wind FITs (NoFFCA?) would benefit from a liquid intra-day market, which would remove the requirement that wind farms need to find buyers for their unpredictable power. A liquid market would provide better price discovery for any low-carbon CfDs that will be required under the EMR, and if that market provides the reference price, more generation may find it simpler to sell directly into this voluntary pool-type market. All this could be done through the normal process of modifying the Balancing Mechanism, without the major legislative change that drove NETA, and possibly without much additional changes to IT and settlement procedures.

Without a pool or its evolutionary equivalent, the SO is likely to require a larger share of flexible generation, particularly in export constrained zones, to be under contract and dispatchable, to ensure system stability at least cost. LMP pricing could be used to determine short-term transmission charges to include in the contract terms, which will likely need to be supplemented with some further (annually fixed capacity related) locational adjustments for guiding generation investment decisions that makes up the shortfall between the LMP element and a properly computed deep connection charge. The major change in this second element is to relate the locational guidance to deep rather than shallow connection costs, although its form would be similar to a properly computed ICRP-based element of TNUoS (and reforming the methodology behind ICRP might be necessary anyway if it is retained). In addition, NGET and Ofgem should agree on the design of suitable long-term access contracts for desired TEC – an FTR or TCC – that provides assurance and a predictable time trajectory for the average of these spot and annual charges.

The case for deep connection charges for new generation is becoming urgent, for there are justifiable concerns that the costs of major new transmission investments such as the "bootstraps" connecting Scotland to England offshore (very expensive) and the off-shore wind connections would be added to the total transmission revenue to be collected from existing generators and load, with the potentially adverse effects noted in the case of Peterhead, discussed in sections 4 and 6.2. Existing generators facing much higher G charges feel aggrieved at having to pay for new entrants. If the average G charge is set to zero, then this problem is alleviated, but whether the spatial variation in G TNUoS charges would remain correct would need investigation. Deep charging also makes explicit the transmission costs associated with new low-carbon generation that will be given long-term contracts under EMR. It is therefore simple in theory to add these costs to their cause and cover them in the contract, without any discouragement to that new low-C investment. The alternative is in effect a stealth tax and as such dangerously attractive to a cash-strapped Treasury.

If deep connection charging is ruled out (because, for example, there is not sufficient evidence that it would improve locational decisions or for cosmetic financial reasons), then the transition from the present system is relatively straightforward, with annual or longer-term grid entry contracts on TEC for existing generation effectively replicating possibly redesigned TNUoS charges. New entrants would be confronted with LMPs but offered suitable FTRs to hedge their temporal risks (but not location choice). On the other hand, the Target Electricity Model for market coupling combined with the preference of most of our partners for charging all transmission to Load might argue for deep connection charging (quite common in the EU) combined with (small?) pricing zones, with all the cost-recovery charges levied on Load.

The FTRs for new generation would be sold at prices equal to the approved regulated locational transmission charges and should not result in large transfers to or from the TSO, but the FTRs granted to existing generation will likely require payments to generators as they will find it profitable to offer lower supplies when displaced by wind, and will enjoy the lost profit of not selling at the GB electricity price. This is in effect what happens under CAMSOC, so it does not represent a major (financial) change, and might even be cheaper if generators are at present exploiting congestion constraints. Such constrained-off payments would previously have been financed from BSUoS charges, but there is a question whether this would be appropriate given that they are part of the package of encouraging new wind generation onto the system. Ultimately the cost would fall on electricity consumers unless the Government made a (desirable) major policy change to recoup the extra cost of renewables from general revenue (or even from the EU ETS generation auction revenue and that from the new Carbon Price Support). Unless that happens it probably makes little material difference to the nature of consumer bills whether they are recovered through BSUoS or charges to cover the cost of renewables support. However, there is a strong case for clarity in cost allocation so that subsequent policy decisions are guided by more accurate and informative information about costs.

LMP pricing alone will not generate sufficient revenue to cover the regulated revenue allowed, and will need to be supplemented by additional charges to make up the short-fall. These non-spatial cost-recovering transmission charges should be designed on Ramsey principles to avoid distorting choices of plant type and operation. If all charges fall on Load, then the existing Triad methodology does that pretty well, regardless of the final choice of how to set G charges.

One problem that considerably complicates the setting of efficient transmission charges is that the various market interventions to support the public good of renewables deployment (ROCs, FITs, etc) may themselves introduce distortions that can interact with any transmission charging regime to distort location and grid investment decisions. The Government will need to decide whether to accept the implication of the 20-20-20 Directive as implemented and provide support for renewable generation, or whether, more logically, to seek a change or derogation that allows compliance to be determined by available capacity rather than output. Some sceptics might even argues that as we are unlikely to meet our renewables targets we shall be in the mode of explaining how we delivered what we actually managed to do and that is a good moment to defend the methodology we chose to justify the form and financial amount of renewables support.

Either way the locational signals for, and renewables support to, new wind farms should encourage the least total system cost of delivering either capacity or generation, and should provide the smallest incentive (or excess profit) to induce the investment at the right location, once the system of FITs/CfDs has been agreed. The present system over-rewards costly distant locations and over-rewards renewables in favoured (e.g. windy) locations, rather than minimising consumer costs and making electricity more affordable.

The current system of transmission charging seems to be some way short of this ideal, although it is unclear how serious are the costs of any potential distortions at present, and whether future generation investment could be guided to efficient locations through the design of the FIT/CfD contract. If not, then the cost of inefficient location and dispatch will likely rise with growing wind penetration.

The cost of financing the very significant and costly grid reinforcements for new wind power is high compared to the cost of the research needed to estimate the cost savings from efficient charging. The question to be addressed in this quantified, model-based research is to measure the size of the difference between the total system cost under the current system and the proposed model of efficient short-run pricing (LMP) combined with longer-run charging according to cost causation, supplemented by minimally distorting uniform charges to recover any shortfall and correct pricing of externalities and public goods. There is therefore a strong case for simulating alternative charging models, perhaps using some of the scenarios for future renewables investment, and under varying assumptions on the ease or difficulty of overcoming planning objections. The research should ideally also examine the wind resource in different locations to estimate the advantages of accessing lower cost if less windy locations earlier.

If nodal pricing is ruled out, then it becomes important to try and correct for distortions to short-run dispatch decisions by at least having local marginal loss charges (as in the Republic of Ireland) and more directed constraint charging, combined with deep connection charges (possibly reforming ICRP and including marginal congestion costs) recovered through annual TNUoS charges, but the complexity of this seems unnecessary as it can so easily be avoided by nodal pricing. Zonal pricing of the form envisaged by the Target Electricity Model might be a satisfactory compromise, has advantages for contract liquidity, and might be required in any case. It is worth noting that Sweden attempted to operate as single price zone, but was forced to move to multiple price zones after a DG Comp investigation that found the Swedish System operator's congestion management actions had spill-over impacts on Denmark (who brought the complaint). That route of enforcing a more efficient dispatch in a meshed system that impacts other Member States may not apply in GB, which is not part of the Continental AC system, but presumably those adversely affected by costly and unnecessary re-dispatch actions might similarly appeal to Ofgem for a more satisfactory set of congestion-defined zones.

A significant change from the present system will require a carefully designed set of transition arrangements that protect existing property rights (to firm grid access on terms set out in the grid principles in force at the time these rights were granted, while recognising that TNUoS charges can and do change quite significantly and at relatively short notice). At present that presumably means starting from the current CAMSOC arrangements, although zonal pricing would surely reduce the costs of doing so and might be an EU imposed solution required in any case.

As there is a clear need for FTR/TCCs in any case, these are likely to form part of that solution, but will need careful calculation to avoid overcompensating existing property rights. Again, interactions with the changing plans for renewables support and carbon pricing under the EMR with their attendant risks of creating windfall profits complicate the issue.

To conclude, it may be that a large part of the potential gains from better transmission charging can be achieved through incremental changes to the balancing and intra-day market, combined with the form of zonal pricing to be determined in the EC Target Electricity Model, and through the contracts for new low-carbon generation. Here the main message is that these contracts should be locationally specific and take full account of the costs that the new generation imposes on the system, The discussions over the exact form of the EC Target Electricity Model are continuing, with a preference for the CWE/Nordel model of zonal prices (through market coupling or market splitting, as there is a strong push for wide area single price zones to facilitate trading on local power exchanges). In the US, zonal pricing rapidly collapsed into nodal pricing, and there must be some concern that the same would happen in Europe with increasing wind penetration and slow transmission expansion, although the underlying market structure and system of regulation is rather different. Large zones facilitate trading ahead of dispatch, but run the risk of costly re-dispatch actions. It remains to decide where to strike that balance, which will have implications for the way in which the GB market will need to adapt, and how transmission charging should be adapted to support overall system efficiency.

References

ACER (2011) *Framework Guidelines on Capacity Allocation and Congestion Management for Electricity*; Draft for Consultation, DFGC-2011-E-003, 11 April 2011

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.

Brunekreeft, G., Neuhoff, K. and Newbery, D. (2005) 'Electricity transmission: an overview of the current debate', *Utilities Policy,* **13** (2), June, 73-94

Bushnell, James B. and Steven Stoft (1996), "Electric Grid Investment under a Contract Network Regime," *Journal of Regulatory Economics*, 10, 61-79.

DECC (2010a) *Government Response to the technical consultation on the model for improving grid access*, available at http://www.decc.gov.uk/en/content/cms/consultations/improving_grid/improving_grid.aspx

DECC (2010b) *Proposals for improving grid access Impact Assessment*, 27 July available at same website

DECC (2010c) Electricity Market Reform: A consultation document, Dec, available at http://www.decc.gov.uk/en/content/cms/consultations/emr/emr.aspx

ENTSO-E (2010) *Overview of transmission tariffs in Europe: Synthesis 2010*, available at https://www.entsoe.eu/fileadmin/user_upload/_library/Market/Transmission_Tariffs/20100914_Transmission_Tariffs_Synthesis_2010.pdf

ERGEG (2011) Final Draft Framework Guidelines on Capacity Allocation and Congestion Management for Electricity Ref: E10-ENM-20-03, 3 February available at http://www.energy-regulators.eu/portal/page/portal/EER_HOME/EER_FWG/Electricity/Congestion%20Management/Overview/E10-ENM-20-03_FG%20CACM_3-Feb-2011.pdf

Green, R.J. (2007) "Nodal Pricing of Electricity: How Much Does it Cost to Get it Wrong?", *Journal of Regulatory Economics*, 31, 2, 125-149

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

Liu, L., Zobian, A., 2002, 'The Importance of Marginal Loss Pricing in an RTO Environment', *Electricity Journal,* Vol. 15, No. 8, pp. 40-45.

Nakicenovic, N., A. Grübler, and A. McDonald, eds., *Global Energy Perspectives* (CUP, 1998)

National Grid (2010) *Statement of the Use of System Charging Methodology,* April, at http://www.nationalgrid.com/NR/rdonlyres/590ACE8D-01BE-4ED9-A028-3736D28A7CCE/41181/UoSCMI6R0v10Final.pdf

NGET (2009a) *Consultation Document GB ECM-18 Locational BSUos Charging methodology*, Re-issue 23 March 2009

NGET (2009b) *National Grid Electricity Transmission System Operator Incentives for 1 April 2010*, 9 Sep, available at http://www.nationalgrid.com/NR/rdonlyres/5BCE1A3B-D7BC-4F1B-8DFF-68B6162643DA/36862/NGETSystemOperatorIncentivesfor1April2010Consultat.pdf

Ofgem (2008) 'Transmission Access Review – Final Report', http://man270109a.decc.gov.uk/en/content/cms/what_we_do/uk_supply/network/trans_access/trans_access.aspx

Newbery, D.M. (2011) ` Financing research, development, demonstration and deployment of low-carbon energy', mimeo, Cambridge, 9 January

Parail, V. (2010) 'Properties of Electricity Prices and the Drivers of Interconnector Revenue' EPRG Working Paper 1033 available at http://www.eprg.group.cam.ac.uk/wp-content/uploads/2010/11/ParailCombined2EPRG10331.pdf

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, 2003, 'The EU Internal Electricity Market: Current State of Debate' presentation at the CMI Transmission Workshop, 18/19 July 2003, Cambridge.

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.

SSE (2010) SSE's TransmiT response submitted to Ofgem on 17 November

Turvey, R (2006) Short & long run transmission incentives for generation location, MIT CEEPR Working paper 0604

Turvey, R. (2011) "Project TransmiT: comments on the draft academic reports on charging models" submitted to Ofgem, April

Appendix A Extracts from the current Grid charging methodology

The methodology is set out in National Grid (2010), and usefully describes some of the principles in chapter 1, from which these extracts are taken.

> 1.6 The underlying rationale behind Transmission Network Use of System charges is that efficient economic signals are provided to Users when services are priced to reflect the incremental costs of supplying them. Therefore, charges should reflect the impact that Users of the transmission system at different locations would have on the Transmission Owner's costs, if they were to increase or decrease their use of the respective systems. These costs are primarily defined as the investment costs in the transmission system, maintenance of the transmission system and maintaining a system capable of providing a secure bulk supply of energy.

Appendix B Braess' Paradox

The general proposition is that adding additional capacity (for example by interconnecting two nodes) to a network in which agents selfishly and individually optimize may reduce overall performance. The equilibrium concept appropriate to a decentralized network is Nash equilibrium – each agent does the best for himself given the actions of all other agents. The result is most readily illustrated for a congested road network, which is taken from the Wikipedia entry at http://en.wikipedia.org/wiki/Braess%27s_paradox

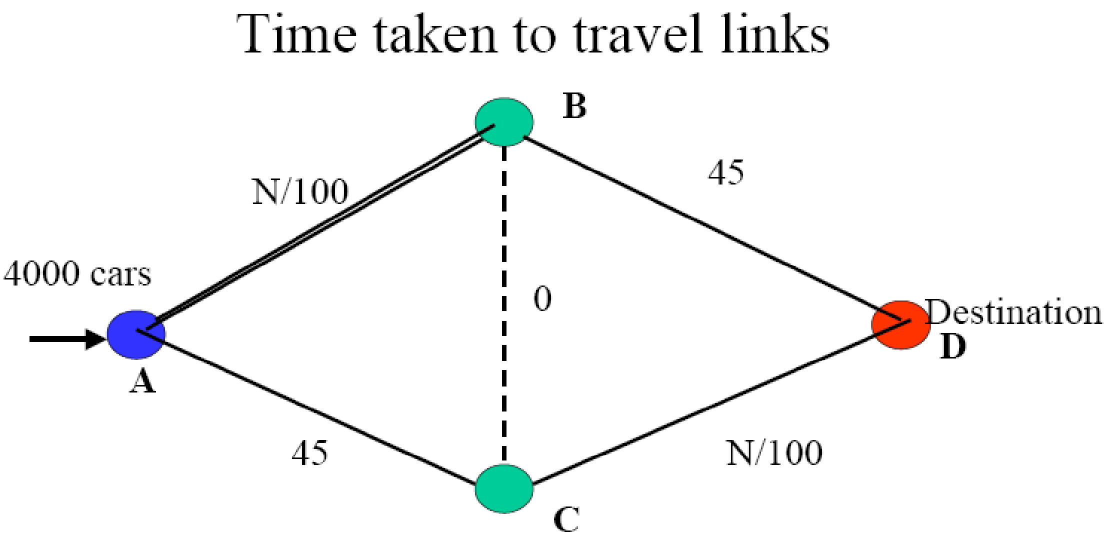

Figure 14 Braess' paradox for transport network

In figure 14 cars enter at the left node and choose routes that are privately optimal to reach their destination at the right. Two roads are uncongested but rather inferior and each link takes 45 minutes, while the other two are potentially congested with travel time taking N/100 minutes, where N is the number of cars on that link. Without the dashed line the equilibrium is one in which 2000 cars take the upper two links ABD and 2000 cars take the lower two links ACD, with total time 45+ 2000/100 = 65 minutes. If the two intermediate nodes B and C are connected by a short route (a bridge over a river, say) that takes essentially no time to traverse, then initially cars will want to travel ABCD expecting that AB and CD will each only take 20 minutes, lowering their travel time to 40 minutes. In fact all cars will route ABCD and so AB will take 4000/100 = 40 minutes and so will CD, giving an overall journey time of 80 minutes, clearly worse than before. Note that both before and after the link the system was in Nash equilibrium (for transport networks often called a Wardrup equilibrium) in which no individual car can find a route taking less time.

Appendix C What Is "Locational Marginal Pricing"?[36]

The "Locational Marginal Price" (“LMP”) or "Locational Marginal Pricing, also referred to as "Nodal Pricing," is a market-pricing approach used to manage the efficient use of the transmission system when congestion occurs on the bulk power grid. The Federal Energy Regulatory Commission (FERC) has proposed Locational Marginal Price as a way to achieve short- and long-term efficiency in wholesale electricity markets.

Marginal pricing is the idea that the market price of any commodity should be the cost of bringing the last unit of that commodity - the one that balances supply and demand - to market. In electricity, LMP recognizes that this marginal price may vary at different times and locations based on transmission congestion. With Locational Marginal Price, market participants will know the price of hundreds of locations on the system

Electric grid congestion develops when one or more restrictions on the transmission system prevent the economic, or least expensive, supply of energy from serving the demand. For example, transmission lines may not have enough capacity to carry all the electricity demand required to meet the demand at a specific location. This is called a “transmission constraint.” Locational Marginal Price includes the cost of supplying the more expensive electricity in those locations, thus providing a precise, market-based method for pricing energy that includes the “cost of congestion.”

LMP provides market participants a clear and accurate signal of the price of electricity at every location on the grid. These prices, in turn, reveal the value of locating new generation, upgrading transmission, or reducing electricity consumption—elements needed in a well-functioning market to alleviate constraints, increase competition and improve the systems’ ability to meet power demand.

Calculating LMP

Unlike the original market in New England, in which there is only one energy clearing price, under SMD, prices are calculated at three types of locations: the node, the load zone and the hub. Offers and bids are submitted, markets settle, and LMPs

[36] from http://www.demandsidemanagement.com/locational_marginal_pricing.htm

are calculated at these locations. Under SMD, prices are first calculated at more than 900 locations, called nodes, throughout New England. Nodes represent places on the system where generators inject power into the system or where demand, or load, withdraws from the system. Each pricing node is related to one or more electrical buses on the power grid. A bus is a specific component of the power system at which generators, loads or the transmission system are connected. These location-specific prices are made up of three components: energy, congestion and losses. The energy component (or marginal cost) is defined as the cost to serve the next increment of demand at the specific location, or node, which can be produced from the least expensive generating unit in the system that still has available capacity. However, if the transmission network is congested, the next increment of energy cannot be delivered from the least expensive unit on the system because it would cause overloading on the transmission system or violate transmission operating criteria, such as voltage requirements. The congestion component, or transmission congestion cost, is calculated at a node as the difference between the energy component of the price and the cost of providing the additional, more expensive, energy that can be delivered at that location. The congestion component can also be negative in export-constrained areas where there is more generation than demand.

All transmission systems experience electrical losses, which occur as electricity is sent over transmission lines and accounts for a small percentage of electricity from generators. Nodal prices are adjusted to account for the marginal cost of losses. If the system was entirely unconstrained and there were no losses, all of the LMPs would be equal and would reflect only the energy price. The lowest possible cost generation could flow to all nodes over the transmission system. Generators are paid nodal LMPs. SMD market rules assure that generators recover their as-offered or bid-in costs, including start-up and no load costs for all energy generated. If a generator operates “in-merit,” most of its compensation will be from the energy market, unless the energy revenues are insufficient to cover its costs. If higher priced generation is dispatched to relieve congestion, the higher cost for this generation is borne by the location in which it occurs through higher LMPs that those locations must pay. In the original market, these costs are absorbed by all load, or demand, across the ISO’s zone regardless of their areas’ contribution to the transmission constraint.

Load Zone
Under SMD, demand, or load, will pay the price calculated for eight load zones, or aggregations of nodes. New England will be divided into the following zones: Maine, New Hampshire, Vermont, Rhode Island, Connecticut, Western/Central Massachusetts, North-eastern Massachusetts (which includes Boston) and South-eastern Massachusetts. The eight load zones under SMD coincide with the eight reliability regions in New England. Reliability regions reflect the operating characteristics of and the major transmission constraints on the transmission system.

The prices calculated for load zones are a load-weighted average of the nodal prices located within each zone. They still reflect the cost of congestion and represent a true cost for delivering power by location. But because they are an aggregation of nodes, zonal prices are less volatile than nodal prices. The New England market is likely to move to a nodal pricing system for load and generation. Load zones are being implemented as a temporary means to help market participants transition from the old market design to SMD. To move to a nodal system, more detailed metering of the 900-plus nodes is needed. Hub In addition to the nodes and zones, a hub has been defined as a single trading location in which the average price is not affected significantly by congestion. It provides a stable pricing location for energy transactions within New England, which serves to enhance transparency and liquidity in the marketplace. The hub is calculated as an average of the prices at all of the nodes defined of the hub. These nodes are electrically connected and are located in an area that has little congestion within it and therefore has a price that reflects the overall energy price.

Short-term, long-term and real-time benefits of Locational Marginal Price
Locational Marginal Price is a market-based means of pricing the efficient use of the transmission system when constraints prevent economically priced power from flowing to where it is needed. In the short-term, LMP improves the efficiency of the wholesale electricity market by ensuring that the cost of congestion is reflected in electricity prices and ensures that the least-cost supply of electricity is delivered while respecting the physical limitation of the transmission network. In the long-term, LMP helps relieve congestion by promoting efficient investment decisions. Because LMP creates price signals that reflect the locational value of electricity, participants can readily determine areas of congestion and will see the value of investing in generation, transmission and demand response programs.

Appropriately located generation additions, transmission and demand response will increase the competitiveness of the New England market. Greater access to a larger number of competing suppliers helps to enforce market discipline without resorting to administratively applied market power remedies. Increased access to energy from lower-cost generators or imported power will ensure robust, competitive prices. And increased competition from strategically located lower-cost units and demand response will benefit much of New England, as the transmission grid is utilized more efficiently. Ultimately, increased competition should result in a more efficient wholesale energy market with lower costs.

Appendix D Balancing Services

The System Operator needs to secure various ancillary services to ensure the safe and secure operation of the transmission system, and should be incentivised to secure these at least cost through short and long term contracts and purchases. These have the nature of public goods, in that the resulting security and quality of service is provided to all connected to the system within some geographical area. The current charging methodology is set out in National Grid (2010), from which the following has been extracted:

"All CUSC Parties are liable for Balancing Services Use of System charges based on their energy taken from or supplied to the National Grid system in each half-hour Settlement Period.

BSUoS charges comprise the following costs:

(i) The Total Costs of the Balancing Mechanism
(ii) Total Balancing Services Contract costs
(iii) Payments/Receipts from National Grid incentive schemes
(iv) Internal costs of operating the System
(v) Costs associated with contracting for and developing Balancing Services
(vi) Adjustments
(vii) Costs invoiced to National Grid associated with Manifest Errors and Special Provisions.
(viii) BETTA implementation costs

Present arrangements for securing ancillary services[37]

The services that we procure, as GBSO, in order to operate the transmission system constitute Balancing Services.

Balancing Services include:

- Ancillary Services;
- Offers and bids made in the Balancing Mechanism; and
- Other services available to National Grid which serve to assist us in operating the transmission system in accordance with the Electricity Act 1989 or the Conditions in an efficient and economic manner.

Ancillary Services, under the Grid Code, can be Part 1 System Ancillary Services, Part 2 System Ancillary Services or Commercial Ancillary Services. Part 1 System Ancillary Services are those which Users are required to have available in accordance with the Grid Code. Part 2 System Ancillary Services are those optional

[37] Extracts from National Grid's Seven Year Statement 2009

services (e.g. black start capability) set out in the Grid Code, which the User has agreed to have available. Commercial Ancillary Services are other optional services (e.g. hot standby) described in the Grid Code, which the User has agreed to have available.

Balancing Mechanism offers and bids are commercial services offered by generators and suppliers and procured through arrangements set out in the BSC. They represent the willingness to increase or decrease the energy output from BM Participants in exchange for payment.

Other Services refers to commercial services that can be entered into with any party, which are classified neither as Ancillary Services nor BM offers or bids. These services can be provided by parties who are not authorised electricity operators. This category would include any service provided by parties that are not signatories to the BSC and may also include the procurement of energy ahead of BM timescales.

For further information on Balancing Services, please see the following website:-http://www.nationalgrid.com/uk/indinfo/balancing

National Grid has actively encouraged and facilitated market arrangements for the provision of ancillary services. Whilst BSUoS charges are levied on all BSC signatories, the provision of ancillary services is not limited to those signatories. Accordingly, the provision of such services is open to any party who can provide a service, including embedded generation, cost-effectively.

System operators at the national control centre use ancillary services. They are only able to call-off a limited number of service blocks in the short period of time available. Thus, for practical reasons, de-minimis sizes are specified for control use. These are:

- frequency response : 3MW each dispatch instruction
- reserve : 3MW each dispatch instruction
- reactive : $^{+}$/- 15Mvar at station terminals
- black start : must be capable of charging circuit

… Our previous duty to purchase ancillary services economically and to despatch plant in accordance with a merit order has been replaced by a general duty to operate the transmission system in an efficient, economic and co-ordinated manner through the procurement and utilisation of Balancing Services including Balancing Mechanism bids and offers. Our GBSO Incentive Scheme normally covers this duty.

Appendix E Scope of work requested by Ofgem

Ofgem has recently launched Project TransmiT, which is an independent and open review of transmission charging and associated connection arrangements. The aim of the review is to ensure that we have in place arrangements that facilitate the timely move to a low carbon energy sector whilst continuing to provide safe, secure, high quality network services at value for money to existing and future consumers.

We are commissioning three short reports on optimal charging arrangements from independent academic experts in the area of network charging. In particular, we are looking for views on what an efficient charging regime might look like for GB electricity and gas networks given the new challenges we face today. These independent reports will be used to stimulate further debate within the shareholders and to inform our own policy development.

The focus of each report will be on electricity transmission charging, although the principles will be considered in the wider context of both gas and electricity transmission.

We expect the report to draw on relevant international best practice and latest academic thinking. The report will consider all aspects of transmission arrangements that are relevant to the allocation of costs arising in transmission, including: investment in transmission assets, costs of transmission congestion and transmission losses, costs for purchasing ancillary services required for safe and secure operation of the transmission system.

We are looking for views on:

a) appropriate guiding principles for transmission charging that are consistent with meeting the objectives set out above;
b) the broad building blocks of a suitable target charging model that would best achieve the objectives as a whole, taking into account any trade-off amongst these objectives, for example:
   - economic efficiency vs facilitation of carbon reduction;
   - long-run investment efficiency including both transmission and generation vs short-run operational efficiency; and
   - requirements for a self-contained system vs those relevant for closer integration of other European systems cross-border.

c) the interdependencies between the proposed charging model and other aspects of the regulatory regime for electricity and, where relevant, gas networks, including cross-European regulatory and policy developments. Where possible, the report should also provide views on the extent to which these help or hinder under the existing GB arrangements.

Deliverables

In line with our high level timetable for Project TransmiT, which includes a milestone of publishing Ofgem's recommendations in Summer 2011, we require the following deliverables from the advisors:

- An initial note by mid December 2010 on high level principles for transmission charging;
- A first draft report by early February 2011 to be presented to a workshop with key experts;
- A draft final report by end March 2011 that has taken into account the comments from Ofgem and the workshop; and
- A final report submitted by end April 2011 that can be published.