CEEPR Center for Energy and Environmental Policy Research

Short & long run transmission incentives for generation location

by

Ralph Turvey

06-004 WP March 2006

A Joint Center of the Department of Economics, Laboratory for Energy and the Environment, and Sloan School of Management

Short & long run transmission incentives for generation location

Ralph Turvey

London School of Economics and Frontier Economics

Introduction

This paper is about one aspect of Britain's electricity trading system, its advantages and its weaknesses concerning the incentives it provides or fails to provide for the location of generation. (Similar considerations apply to the location of loads, though these are less responsive to locational influences exerted by the trading system).

The optimal location of generation in the short-run is a matter of determining the unit commitment and dispatch of the existing generation park so as to minimise the cost of generation hour by hour, subject to security constraints and taking account of transmission losses. In the long-run, choices of the locational pattern of new plant construction and of the decommissioning of old plant should be influenced by their effects upon the cost of the transmission investment that they entail.

In systems with a gross pool, such as in New York, Ireland and New Zealand, there is a central dispatch. This, taking account of transmission losses and constraints, can produce locational marginal prices. Expectations concerning their future levels provide signals relevant to the location of new generation. Thus both in the short-run and in the long-run these systems provide locational incentives. In some of them, where the long-run incentives to investment provided by the uncertain prospect of future price spikes are deemed insufficient, capacity requirements are imposed upon (what in Britain are called) "suppliers". These too can embody a locational element, as in the LICAP arrangements in New York and proposed for New England.

In the British system, there is a net pool, and two cashout prices rather than one emerge from the Balancing Mechanism. This is Britain's version of what is elsewhere called the spot market, regulation market or real time market. Unit commitment is left to the generators, while National Grid, as system operator re-dispatches so as to preserve balance and to deal with transmission constraints. For the latter purpose, it constrains on here and constrains off there (though such actions may serve other purposes too). This costs it money, providing the occasion for it to weigh up the operating cost of dealing with constraints against the capital cost of removing them. But locational prices do not emerge from this process and no account is taken of locational differences in marginal losses. These are two defects of the short-run locational incentives provided by the British system.

On the other hand, the British system scores highly with respect to long-run locational incentives. Instead of providing these by participants' expectations of future locational differences in energy prices, and maybe capacity prices, Britain provides them through locational differences in the transmission costs borne by generators. National Grid's Transmission Use of System Charges vary locationally to reflect the results of an "Incremental Cost" analysis. But although these may be roughly right, National Grid's approach is imperfect, even though it has evolved to meet some past criticism[1]. This paper points to its remaining defects after first tackling the short-run issue of the treatment of losses.

Losses

The cost of transmission losses in Britain is recovered by adjusting the metered volumes of all transmission users by a uniform percentage to reflect the actual total losses incurred in each settlement period. In other words generator's injections are all treated as being larger than the withdrawals by the buyers by the same percentage. The settlement system does recognise that transmission losses could be allocated on a locational basis, but the locational transmission loss factors which would be used to calculate this are

---

[1] Ralph Turvey "NETA and Transmssion" (Beesley lecture 1999) in *Regulating Utilities* ed. Colin Robinson. Edward Elgar 2001

currently set to zero. Thus the bilateral market (and the Balancing Mechanism) disregards locational variations in loss factors, resulting in economic inefficiency.

Zonal transmission loss factors

Setting the transmission loss factors to reflect geographical variations in losses would be one way of ensuring that these variations were taken into account in generation scheduling, It was proposed (Modification P82) that average zonal transmission losses should be allocated for each Grid Supply Point group on an ex-ante basis throughout the year in England and Wales. This was fiercely resisted by some northern generators. It was rejected by the Balancing and Settlement Code Panel but nevertheless was approved by OFGEM. Following a subsequent public consultation on the application of Average Zonal Transmission Losses and on the basis of further work commissioned from OXERA, the Government then took "the view that any positive net benefit that might flow from the introduction of Average Zonal Transmission Losses is ambiguous, especially in light of the wide margins of the upper and lower limits of the costs and benefits."[2] The work which led the Government to take this view was summarised as follows:

"OXERA employed a load-flow model of the Great Britain electricity transmission networks in conjunction with a wholesale electricity market model to compare the different outcomes of zonal and uniform loss charging between the years 2005/06 and 2009/10."

"Notwithstanding the many uncertainties attached to an exercise of this nature, OXERA's base case scenario estimated that the overall benefit in Net Present Value terms from the resulting changes in demand response and generator redispatch and relocation would possibly amount to £6-55 million . However, it was also proposed that the implementation and operational costs of AZTL might range between £3-31 million . It was also estimated that AZTL would cause a redistribution of almost £30 million from northern generators to southern ones, as the former would pay more of the cost of transmission losses while the latter would pay less. The opposite is true for electricity suppliers and, by implication, their customers. It is important to bear in mind that a redistribution is neither an overall benefit nor cost. [3]

---

[2] Government response to consultation on Transmission losses in a GB electricity market, June 2003

[3] Government response to consultation on Transmission losses in a GB electricity market, June 2003

The proposal has now been resuscitated as Modification Proposal 198, being put forward by RWE Npower in December 2005 and is to be examined by an assessment group which is to report in May[4]. It now applies to Great Britain, the introduction of the British Electricity Trading and Transmission Arrangements (BETTA) in April 2005 having extended the scope of the Code to incorporate Scotland,

Charging for losses

An alternative would be for the National Grid to bear the cost of losses, buying the energy for them and recovering these costs by geographically differentiated MWh charges for them. Such a locational transmission charge for losses could range from one which reflected average losses to one which reflected marginal losses.

This alternative was at one time favoured by National Grid, who believed "that the transmission system operator should be responsible for the costs of transmission losses and their subsequent charge-out in an economically efficient manner", although they were ambivalent about the use of average or marginal losses. They argued that "This option would also deliver the many benefits arising from the SO managing losses; including removing potential distortions from the Balancing Mechanism, helping the separation of transport and energy prices, consistency with European markets as well as consistency with the gas market."[5] By bearing both the cost of losses and those of operating its system and of undertaking reinforcements, the National Grid could be incentivised to choose the tradeoffs between these costs which would yield a minimum cost solution. The argument of course presupposes that the National Grid would buy the energy to meet losses at least as efficiently as suppliers,

Marginal vs average losses

The proposal which OXERA was commissioned to study was for a single transmission loss factor to be applied in each zone each year thus reflecting *annual averages of average losses*. It seems obvious that this was an inferior device. It would be much better to apply *time-varying marginal* loss factors, differentiated by time of day,

[4] ELEXON Document Reference: P198IR 6 January 2006

[5] NGC, Response To "Transmission Access And Losses Under Neta" Ofgem Consultation Document, May 2001, Appendix 4, July 2001

week and year, thus enabling the market to approximate locationally optimal commitment and dispatch. An example of such marginal loss factors is provided by Table 7,4 of National Grid's most recent Seven Year Statement. It shows the extreme cases to be that a small increase in generation in the North has an effectiveness of 94% in meeting demand across the system at the time of winter peak. Whereas a small increase in generation in the Peninsula has an effectiveness of 105% in meeting demand by virtue of reducing transmission power losses

If time-varying marginal loss factors which were determined ex ante, and thus known to users, were applied, generation would exceed demand by the excess of marginal over average losses. This could be:

- Accepted, allowing the National Grid to be a net seller of energy in practically all half hours, increasing the liquidity of the market and providing it with a source of revenue consistent with optimal dispatch. Without this revenue source, National Grid's user charges would have to be higher, providing unwanted incentives to users and so interfering with economic efficiency.
- Avoided, by scaling down all marginal loss factors by the same *absolute* amounts so that the absolute (locational) differences between them are preserved.

A similar choice applies under the first alternative noted above, namely purchases by National Grid to meet losses financed by time varying and locationally differentiated MWh charges paid by users. Setting these charges to equal the marginal cost of losses would yield a surplus over the cost of providing for losses. Charges reduced by the same absolute amount, preserving the differences between them, could avoid this.

National Grid is currently provided by OFGEM with an incentive to reduce losses. This would fit badly with either alternative, for under both of them market participants would become incentivised to take account of locational differences in marginal losses in their current decisions, including their Balancing Mechanism bids. If the National Grid also faced incentives on the level of losses then they would take actions in the Balancing Mechanism even though participants had already internalised these costs. In any case, the incentives provided to National Grid require a rethink because minimising losses is not the same as minimising the costs of meeting the load which is what is wanted.

Conclusion re losses

The application of zonally varying average loss factors is now to be re-examined. It would be better still if the study covered time varying marginal transmission loss factors. Their introduction would of course have redistributive effects, some of the potential losers being more articulate than the potential gainers. It should not, however, be regarded as a zero sum game, since there would be real benefits to be weighed against real costs.

Connection charges

The problem of providing long-run *locational* cost messages and incentives for generation investment and disinvestment forms the subject of the rest of this paper. To the extent that locational choices are made with respect to loads the discussion applies to the demand side as well. What is at issue are choices between locations where new investment or closures will not exacerbate transmission constraints and locations where they will. In the latter case, new generation will impose a cost upon the transmission operator which will have to choose between more out-of-merit operation or transmission reinforcement.

With transmission use of system charges, users pay monthly or annually either according to their capacity or according to maximum metered flows. The schedule of charges applied in each particular case is pre-determined. Connection charges provide an alternative where charges are determined individually when users are connected. Thus they can provide a tailor-made locational differentiation, charges reflecting the estimated costs of accomodating additional generation in each particular case. In principle, therefore, they are the ideal solution as regards long-run locational incentives.

Connection charges are usually paid as a lump capital sum. They could, however, be spread out over time, being reckoned as the annuitised value of that lump sum. If the user closed down, payment to it of a credit reflecting any saving to the transmission operator could possibly be made, whereas in the case of annuitised payments users would automatically cease to pay upon disconnection.

Another choice lies between “shallow” and “deep” connection charges.

- Under a ‘deep’ approach, generators or new loads connecting to the transmission system would pay not only for the cost of the local connection but also for the incremental investment made on the wider transmission system to accommodate the additional generating capacity or load. Thus the generator or user would have to pay for all the additional transmission assets which connection would entail, including the costs of reinforcement at remote sites.
- Under a ‘shallow’ approach, new users would be required to pay only for the local assets specifically required to connect them to the transmission system and for their specific benefit. The costs of reinforcing the system beyond the connection assets would therefore be recovered as part of use of system charges.

There is obviously room for defining shallowness more or less widely as recent debate has demonstrated.

Because of the indivisibility of plant and equipment and because it may sometimes be sensible to increase capacity by more than is immediately required when this will cost less than would smaller successive increments, a “deep” cost of connection may be incurred to provide spare capacity over and above that immediately needed to service a new user. So, if deep connection charges are imposed, there will often be a third choice to be made: shall all of the reinforcement cost be borne by the new user, or shall the new user cover only part of it, leaving the rest to be covered by any subsequent new users? What if no reinforcement is currently necessary to accommodate a new user, but the arrival of that new user will bring forward the time when any subsequent appearance on the scene of further new user will necessitate upstream reinforcement?

If the chosen answer is to share out the cost of upstream reinforcements, then users arriving on the scene subsequently will have to contribute towards the cost of the past reinforcements which provided the spare capacity that they are now going to utilise. Hence when the advent of a new user does not entail costs (other than the shallow cost of the immediate connection) because sufficient capacity exists upstream to obviate any immediate need for reinforcement, this approach will require the new user to contribute towards past reinforcement expenditure.

The deep approach could have the merit of greater cost reflectivity, the connection charge depending upon the particular circumstances of the part of the network affected by the advent of a new user. Unfortunately, however, the third choice then has to be faced, involving difficult issues that are avoided with shallow charges. It is clearly acceptable to differentiate connection charges according to differences in the (shallow) cost of the immediate connection assets that are not shared. But can it be acceptable to differentiate them according to new users' estimated fractional contribution to either upstream reinforcement costs whose optimal scale may depend upon uncertain forecasts of future new generation and closures, or according to their contribution to load flows made possible by past reinforcement?

Such problems have led to the choice of shallow connection charging in Britain, the Regulator taking the view that

a deep connection charging methodology is more likely than a shallow charging policy to result in charges which could discriminate between similar customers depending on the time of their connection. The connection of a new customer in a given location may trigger the need for reinforcement of assets which would be shared by all local users. Under a deep connection policy, these charges would be charged to the new customer despite the fact that they will be shared by other users. Furthermore, given the lumpy nature of connection investments, subsequent new users may be able to connect at a relatively low cost. Such arrangements will act to distort competition by changing the cost base of otherwise similar users.[6]

In Britain, a shallow connection boundary based on single user assets has thus been chosen. Upon the connection to the grid of a new generator or new demand point, all assets which are shared or could be shared will now be charged for via use of system charges rather than connection charges. Substations (and associated site infrastructure and land), generation only spurs, and shared transformer circuits will be charged for via use of system charges.

National Grid's Use of System charges

Since the subject under discussion is locational incentives, what is of concern here is only the locationally varying element in National Grid's Use of System charges, i.e. *differences* in these charges between the 21 generation zones and the 14 demand zones. Each of the latter group together the Grid Supply points where a distribution network is

[6] GEMA, NGC's proposed GB electricity transmission charging methodologies, The Authority's decisions, December 2004, paragraph 3.28

fed, these being the groupings necessary for energy market settlement purposes. The generation zones, on the other hand, contain generation nodes whose marginal costs are close and which are geographically and electrically proximate.

National Grid's methodology

The Direct Current Load Flow model used to compute the marginal costs of investment in the transmission system requires the following inputs relating to system peak conditions:

- Nodal demand information
- Nodal generation information
- Transmission circuits between these nodes and their impedances
- The lengths of these routes, the proportions of each which is overhead line or cable and the respective voltage levels

The other input data are the cost of transmission reinforcement, expressed in terms of ratios of 132kV overhead line, 132kV cable, 275kV overhead line, 275kV cable and 400kV cable to the cost of 400kV overhead line. Here, costs are expressed as the annuitised values of the transmission infrastructure capital investment required to transport 1 MW over 1km with an addition for maintenance costs. These circuit cost ratios are used to derive what are called "circuit expansion factors", the analysis being rather quaintly expressed in terms of increases or decreases in units of kilometres (km) of the transmission system for a one Megawatt injection to the system.

For a completely accurate representation of power flows an AC load flow transport model would need to be used to consider voltage issues, but the difference in results would be small

The demand data for each node are based upon the Grid Supply point demands under Average Cold Spell conditions forecast by Users. The generation information, on the other hand is not a forecast of how peak demand will be met. Instead, it is arrived at by taking the sum of generators' export capacities and scaling them all down in the same proportion to the point where they will just meet the load. A load flow analysis can then derive the resultant pattern of flows based on the network impedances, assuming every circuit has as much capacity as needed.

This specifies an adequate baseline network which is just adequate to convey the flows resulting from the assumed levels and patterns of demand and generation. The model can then calculate for an injection of an extra MW of generation at each node accompanied by a corresponding additional one MW of demand at a reference node, the required increase or decrease in Megawatt-kilometres on all routes of the network. This allows calculation of marginal costs, though National Grid chooses to express the result in terms of marginal 400 kV MWkm by applying the circuit expansion factors. Marginal cost will, of course, be negative for additional injection in deficit zones. Marginal cost for demand is equal and opposite in sign to marginal cost for generation at that node. The choice of reference point affects the level of marginal cost but not the differences between nodes. MW-weighted averages of marginal costs for each group of demand and each group of generation nodes are employed to obtain zonal marginal costs.

These calculations assume that the transmission system is sized so that it is just adequate to accommodate the flows resulting from the postulated peak generation and demands. But since security standards require that it should still be able to accommodate them in the event of any one of a number of plausible single and double circuit outages, the zonal marginal costs are all multiplied by 1.8. This multiplier is derived from the (surprisingly uniform) ratios for all nodes between the results of a load flow model that allows for such contingencies and the simple model which does not. The model that allows for contingencies calculates load flows along each circuit for each of a series of plausible outages and finds the maximum flow along each circuit. For each node the sum of the derivatives of these maxima with respect to net injection at that node, each multiplied by circuit equivalent length, then yields a marginal cost expressed in circuit equivalent length. It is these "secured" marginal costs which average 1.8 times the marginal costs computed for zero contingencies, (oddly denoted "intact" marginal costs by National Grid).

In order that National Grid can recover the total revenue allowed to it by the Regulator, a constant non-locational residual tariff for generation and demand, which includes infrastructure substation asset costs, is calculated and added to the marginal costs to determine National Grid's Use of System tariff.

There has been much discussion of how the tariff is apportioned between demand and generation. Demand charges are based on the average of half-hourly metered or profiled demand during the Triad. (This is the half hour of system peak demand and the two half hours of next highest demand separated from it and from each other by at least 10 days, between November and February.) Generator charges in positive charging zones are based on their highest Transmission Entry Capacities applicable for the year. In negative charging zones, charges are based on a triad-type average of metered volumes. A proposal to charge only demand was turned down by the Gas and Electricity Markets Authority even though it recognised that once the market has adjusted, the change in apportionment would make no difference, since in all cases the amount of the tariff will constitute a gap between the net sum received by the generator from the net cost to the purchaser. However, consumers might lose during the adjustment period,[7]

National Grid in its role as System Operator keeps the electricity system in balance (known as "energy balancing") and looks after security of supply ("system balancing"). These balancing costs include the costs it incurs in constraining off some generation and constraining on generation elsewhere in order to deal with transmission congestion. They are recovered from customers in proportion to their Metered Volumes adjusted for transmission losses. Thus despite a locational element in these costs there is no locational element in the way they are recovered. National Grid has justified this on the grounds that "it is not possible to take balancing actions on a "one action, one reason" basis. For example, resolution of transmission constraints is one of many reasons underlying SO balancing actions and allocating the cost unambiguously is impossible."[8]

Modeling incremental capacity costs

Incremental capacity costs are expressed as the costs per MWkm of newly built lines and cables of different voltages although additional capacity is not always provided by constructing new circuits. Conductor re-profiling (allowing the conductors to operate at a higher temperature), re-conductoring and voltage uprating constitute alternative extensively used methods of adding to capacity. National Grid have, however, produced

---

[7] OFGEM letter 31.10.2005

[8] GB Transmission Charging: Final Methodologies Consultation, 20 August 2004, page 50.

an analysis which shows that, allowing for that fact, the cost of new build does provide a reasonable estimate of the cost of providing incremental capacity.[9]

An alternative approach to cost analysis has the desirable property that the estimates resulting are consistent with the way in which reinforcements are actually decided and costed. It proceeds by using engineering estimates of the costs of reinforcements that are actually planned or of the reinforcements that would be chosen in order to deal with alternative potential load increments.

GB SYS Boundary	Required reinforcement (as identified in phase 1 of Transmission Investment for Renewable Generation)	Annuitised distance -related cost including O&M £/kW/yr	Generation Tariff differential between receiving and sending end £/kW/yr
B3 Sloy Export (SHETL)	275/132kV substation at Sloy & 132kV line Works	2.6	2.4
B4 (Also B1, B2) SHETL-SPTL	Beauly – Denny 400 kV line (220Km)	17.1	9.2
B5 North – South (SPTL)	3 Mechanically Switched Capacitors Series reactor @ Windyhill Switchgear replacement @ Easterhouse and Clyde Mill substation	1.1	0.8
B6 SPTLNGC	Reconductor Eastern Interconnector, Upgrade Western Interconnector. New substations.	11.2	3.9
B7 North - Midlands	Predominately reactive compensation	3.4	3.9

[9] GB Transmission Charging: Use of System Charging Methodology Revised Proposals Consultation Version 2, 20 December 2004, section 5.5

National Grid has provided such estimates for five reinforcements of pinch points in the grid and compared them, excluding the costs corresponding to substation works, with the differences in the model-derived tariff as shown above:[10]

The approximation is not particularly good. Note that the reinforcements are proposed for network boundaries representing transmission pinch points (flowgates). The capability of a boundary is the maximum transfer across the boundary that can be tolerated for a particular pattern and level of demand and generation without breaching security criteria. This means that following any of a large number of plausibly possible contingencies, such as fault outages of transmission circuits, there are no overloaded items of transmission equipment, no unacceptable voltages, and all demand is supplied (save for interruptible demand). An engineering analysis of marginal costs node by node would proceed by calculating Power Transfer Distribution Factors — the incremental flows through each part of the transmission network that would result from a postulated increment in nodal generation or demand and an increment of opposite sign at the reference node. Each such cross-border incremental flow which would lead to a flow across the boundary (plus an allowance to take account of non-average conditions e.g. relating to power station availability, weather and demand) in excess of that boundary's capability would then be multiplied by the unit costs of reinforcing that boundary's capability.

Looking at the matter in this way immediately reveals four more problems or complications. The first is that as well as expenditure on lines and cables, the cost of reinforcement includes expenditure on substations, which are excluded from National Grid's incremental cost calculations. The second is that reinforcements are costed as if an extra one MW were involved whereas in fact they are lumpy investments that accommodate hundreds of additional MW. The third is to remember that increments can be negative, i.e. that decrements should also be considered because decremental cost may differ from incremental cost. Lastly, there is the point that the model is designed to provide market participants with a cookery book formulation of tariff determination.

---

[10] *op.cit.* Table 2

Transformers & switchgear

Substation assets are not included in the cost calculations. The reason for this lies in the nature of the cost model which is formulated entirely in terms of circuit lengths by circuit type. National Grid has sought to justify this by asserting that "generally substation costs are non-locational"[11]. Yet, as the examples above demonstrate, expenditure on transformers, switchgear and load flow devices does form part of reinforcement costs. Estimation of marginal costs made by applying Power Distribution Transfer Factors to engineering estimates of the cost of relieving constraints would obviously include them.

The cost model used has evolved from an even simpler initial version which was a pure "transport" model, unrelated to the way National Grid describes possible needs for reinforcement in its Seven Year Statements. These Statements describe the ability of the British Transmission System to accommodate further generation and demand in different zones in terms of the thermal and voltage limits on the bulk transfer of power across certain system boundaries. "17 boundaries have historically reflected some of the main weaknesses on the interconnected system. Such weaknesses can lead to the need to restrict power flows across the system; possibly through the potentially uneconomic constrained operation of generating plant. Alternatively, transmission weaknesses may be removed through some form of transmission reinforcement. Although the most critical boundaries may not now be precisely the same as those studied, the 17 boundaries which have been used remain relevant for illustrating system trends and limitations." "The 17 zones have been grouped into five opportunity groups, namely: VERY LOW, LOW, MEDIUM, HIGH AND VERY HIGH. These categorisations are intended to provide a broad indication of the relative level of possible opportunities for connection within individual zones, or groups of zones, without the need for further major inter-zonal transmission reinforcement."[12] These, implicitly, are statements about marginal costs.

Indivisibility

The second problem with National Grid's cost analysis is that it assumes transmission capability to be exactly proportioned to what is dictated by the assumed

[11] GB Transmission Charging: Final Methodologies Consultation, 20 August 2004

[12] National Grid, 2005 Seven Year Statement

pattern and level of generation and loads and by security requirements. Line costs are expressed per Megawatt kilometre and it is implicitly assumed that any necessary amount of Megawatt kilometres can be installed. Yet investment in lines and cable, and in the substation capacity that is omitted from the analysis, is unavoidably lumpy: equipment comes in standard sizes and additions to capacity may sensibly be sized to allow for future load growth beyond that which has necessitated them. The result is that many individual lines, transformers and switchgear are underutilised so that up to some point they can accommodate increased load flows without any need for reinforcement. The consequence is that an increased flow of x MVA over a particular component of the grid may lead to no reinforcement if x does not exceed a certain amount but will lead to a reinforcement which augments capacity by a larger amount than x if it does. In other words to allow the analysis to run in terms of *marginal* costs it has to pretend that there is no indivisibility in transmission plant and equipment.

This is wrong but, I now argue, it is nevertheless unavoidable!

Consider what, in principle, is an appropriate concept of the cost of an increment of x MVA in the peak flow along a particular line or cable or through a substation?

1. If it does not necessitate immediate reinforcement, it will nevertheless bring forward the time in the future when any subsequent growth in that flow will require reinforcement. Hence the opportunity cost of x is the expected probability of this multiplied by the increase in the present worth of the cost of that future reinforcement due to bringing it forward.
2. If it does necessitate immediate reinforcement and this raises capacity by more than x MVA, it may put off the time in the future when any subsequent flow growth will require reinforcement. Hence from the cost of the reinforcement must be deducted the product of the expected probability of this and the decrease in the present worth of the cost of that future reinforcement due to postponing it.

These are concepts without appeal to a practical mind. Yet, in practice, account does have to be taken of concepts as difficult as the above even if they are expressed in another way. Consider the choice between a small and a large reinforcement when it is clear that the latter would turn out to have been cheaper if possible subsequent further demands or new generation did in fact materialise. The decision involves risk, so it is necessary to decide who is best placed to bear it.

The DNO, generator or large consumer responsible for the increased flow which triggers off the reinforcement could pay the whole cost (either as a lump sum or, annuitised, as part of its annual Transmission Use of System Charges). It could then be partly repaid by further users if and when they came along.

Alternatively, National Grid could bear the whole cost. In this case, both the DNO, generator or large consumer responsible for the increased flow, and further users if and when they came along, would share out that cost in the charges they pay. These would then equal unit cost grossed up so that the expected present worth of the revenue they would generate would cover that whole cost, thus allowing for the futurity and the uncertainty of the arrival of further users. This amounts to making an allowance for the expected average degree of underutilisation of such assets. The charges could take the form either of a lump sum or, annuitised, as part of annual Transmission Use of System Charges

It seems reasonable to say that which alternative is to be preferred depends on whether National Grid, the DNO or sometimes the generator is better placed to foresee future developments. In all cases, charges should be based on unit costs multiplied by coefficients exceeding unity to allow for the fact that their indivisibility causes assets to be underutilised on average.

This does not mean that these coefficients can be computed by estimates of the extent of spare capacity currently existing in the system, for there are, of course, other causes of the underutilisation of assets. The major one is the provision of redundancy in order to meet security standards. This is taken account of in a simple way in National Grid's calculations by the 1.8 multiplier. A much lesser one is that long run changes in the location of generation and loads may have falsified the expectations upon which some circuits or substations were designed. Thus National Grid was right to abandon[13] its

[13] National Grid, GB Transmission Charging: Final Methodologies Consultation, 20 August 2004. Page 46

original idea of treating circuits identified as having spare capacity as less costly to invest in because there was a buffer before new investment would be required. This had been previously modelled in their transport model by reducing the length of these routes to 75% of the original length to reflect the reduced cost. This failed to recognise the cost of the corresponding spare capacity provided in sizing new capacity.

Reversibility

Discussion of optimal pricing in terms of the cost of increments makes it too easy to forget that increments can be negative. The principle that users should be required to cover the costs of adding to the network imposed by additional flows is matched by the principle that the saving in their charges if they reduce flows should reflect the resultant saving in National Grid's costs. National Grid's model, with its assumption of divisibility, yields marginal cost estimates which apply equally to marginal increments and marginal decrements.

All this overlooks the asymmetry between increments and decrements which is particularly obvious when the indivisibility and durability of plant and equipment is recognised. Putting in additional capacity entails a cost, while using existing capacity less may provide no saving, at least immediately. Only when the time comes to replace old switchgear or an old circuit, may it be possible to do it more cheaply than if flows had not diminished. (Nor will O & M costs fall much if at all while the old switchgear or circuit continues in use.) Otherwise the cost saving resulting from a decrement will be as large as the cost of providing for an increment only when the occurrence of a decrement obviates the need to reinforce to cope with an increment in the load flow resulting from the actions of some other user. Thus the appropriate message and incentive given to users contemplating increased generation or demand should be much stronger than that given to users contemplating reductions. Yet Transmission Use of System Charges provide symmetrical incentives for increments and decrements.

Once again, the only practicable procedure is to take an average — of what is appropriate for users who may expand their activities and of what is appropriate for users who may contract their activities. If it seems equally likely that users will expand and that users will contract, then a simple average would be appropriate. Thus if marginal cost downward is practically zero, charges should be set at half marginal cost upward, plus the uniform addition necessary for National Grid to earn its allowed revenue. If a simple look at marginal cost downward shows that it is one half of marginal cost upward, then the locational differences in charges should be set at ¾ of locational differences in marginal cost upward. Locational differences in National Grid's Use of System Charges are biased upward by its failure to take any account of this point.

Replicability

National Grid makes its model and data available so that market participants can check its mathematics, understand its results and examine the effect upon charges of any change in the assumed level and pattern of generation and demand which they choose to postulate. This is desirable in itself, but prevents the National Grid from using more realistic data which embody its commercially confidential knowledge and its uncertain but sensible forecasts about the commissioning and decommissioning of generation plant and about fuel costs.

The generation and load data used come from National Grid's SYS. Customer-based demand forecasts are used which show stronger growth than National Grid's own 'base' projections. As regards generation, existing plant and new generation for which an Agreements has been made are included, while speculative new projects, potential closure of existing stations or other developments that may have been discussed with the relevant customer are not included without the agreement of the customer.

There are therefore two artificialities in the way load flows are projected. One is that, instead of an approximation to merit order generation, all generators are assumed to produce at the same percentage of their capacity. The second is that, as National Grid explicitly recognises, "the 'SYS background' does not necessarily represent the most likely outturn. For example, it is reasonable to suppose that new applications for power station connections will be received, some power stations will close and some contracts for generation projects may be modified or terminated. This may lead to the need to vary the planned future development of the transmission system to meet changing system requirements."[14]

---

[14] National Grid, 2005 Seven year Statement, section of Introduction entitled The GB SYS Background.

Taking the first point on board, National Grid has recognised that, in principle, a merit order approach would ”better fit the principle of the model, namely an analysis of the peak system conditions“ though they “have some concerns regarding such an approach, not least how to ensure transparency, and how to avoid the potential for volatility year on year” but ”undertake to consider this idea in the future”.[15]

Regarding the second point, National Grid has taken account of its current views on the various generation and demand uncertainties and used Monte Carlo analysis to derive probabilistic ranges of net transfers across each boundary, year by year. The analysis randomly selects generator openings and closures, balancing the probable generation capacity against probable peak demand and probable plant margin. The results show that for some boundaries the SYS transfers generally lie towards the top of the likely transfer range whereas there are others where the SYS transfers generally lie towards the bottom. But it is the relationship of these probabilistic transfers to boundary capabilities which indicates which boundaries would probably require reinforcement because of the increased flows across them resulting from a postulated incremental flow between any given node and the reference node. So the model errs in treating any increase in flow across any boundary at the time of peak demand on the GB system as a whole as imposing a reinforcement cost.

Conclusions

A proper treatment of losses is perfectly possible.

The existence of different ways of providing long-run locational incentives is illustrated by the contrast between the Standard Market Design in the USA with the British system. Under the former, the relative merits of different locations for future investment are illuminated by long-run expectations about *future short-run* market results. In Britain, on the other hand, locational messages and incentives are conveyed by National Grid’s transmission tariffs reflecting *long-run* cost estimates. Thus National Grid does the right thing with its Use of System charges, although it does it imperfectly.

[15] National Grid, GB Transmission Charging: Final Methodologies Consultation 20 August 2004