# Transmission Pricing and Transmission Price Control*

*By Konstantin Petrov***

## Review of the Existing Pricing Arrangements

### Network and Connection Service

Order N 72-э/3 provides detailed discussion on the provision of transmission network service. However, little attention seems to have been paid to the transmission connection service. Separate treatment for transmission network and transmission connection service follows economic causality principles, namely to charge the costs to the party that has caused the costs. The pricing regime should follow the classification of transmission service elements and, therefore, distinguish between transmission use of network charges and transmission connection charges.

### Asset Valuation and Regulatory Asset Base

There are not clear statements with respect to the asset valuation and the establishment of regulatory assets base. Arguments about asset valuation vary from insisting ownership rights be recognised, to questioning whether any value should be attributed to sunken investment. The regulators usually endorse particular asset valuation methodology and the asset valuation issues must be considered with regard to the functional adequacy of regulated assets, market assets value, overall profitability of the regulated business and sustainable cash flows of the business as well as equity considerations. The objective of encouraging continuing investment in the regulated transmission service will require the FNC to be provided with an assurance that it will earn a reasonable (risk adjusted) return on its transmission assets. Therefore the treatment of assets and how their value should be adjusted over time reflecting the investment process (but also depreciations) is a crucial issue.

### Profit Allowances and Cost of Capital

Order N 72-э/3 suggests profit allowance for the regulated transmission service provider that includes three major components:

**(1)** for networks of 500 kV and above, the net profit (profit after tax) plus depreciations should be sufficient to fund planned investments;

**(2)** for networks of 330 kV and below, the net profit (profit after tax) plus depreciations should be sufficient to cover the debt service (including debt interests and principle); and

**(3)** a non-investment profit term that should suffice to cover dividends, reserves, social charges, etc.

Profits belong to shareholders and could be distributed or retained and used for financing of new investments, i.e., it is a decision of shareholders how to finance new investments. Establishment of predetermined liaisons between investment fund requirements and profit (dividend) distribution policy does not seem to be appropriate as it delegates to the regulator functions that naturally belong to the companies' management. The regulatory regime should specify a return on equity (as an element of Weighted Average Cost of Capital (WACC)) and consider taxes either as converting the after-tax in pre-tax rate of return or through explicit taxation allowances in the revenue requirements and after-tax rate of return. The decision on profit (dividend) distribution policy remains in the hands of the company's management that is compatible with the so-called "light-handed" regulatory policy.

The profit mechanism for networks of 500 kV and above implies the entire internal financing of all network investments. The profit is treated as a residual funding source, resulting from the difference between the investment plans and available cumulated depreciation volumes. Differently, the mechanism for networks of 330 kV and below calculates the profit level necessary to cover the debt service requirement resulting from debt funding of investments (net of depreciation). Presumably, this approach strives to ensure reliable funding arrange-

---

*To be continued in the next issue

**This publication is based on the work done by KEMA Consulting with the assistance of the European Union under the Tacis Program. The contents of this publication are the sole responsibility of KEMA Consulting and can in no way be taken to reflect the views of the European Union.

ments for network investments through funds generated internally on the basis of regulatory profit allowances. While this might be desirable from the perspective of keeping a reliable transmission network and sufficient level of investments, such an approach may easily lead to perverse incentives to the financial management and financial discipline of the FNC, e.g., excessive spending of depreciation funds and increasing the requirements to residual profit component (at the time point of assets replacement).

Ideally, investments in asset replacement should be financed through cumulated depreciations (valued at replacement cost to catch up asset price increase resulting from inflation) and investments in network extension should be funded by new external financing sources. Funding investments through profits that constitute part of the revenue requirements means that customers are forced to participate in the financing process. Once constructed, however, the assets should be depreciated over their life and the annual depreciation allowances should be included in the transmission network charges. Hence, the approach applied by Order N 72-э/3 implies double counting and charges transmission service users twice: via the residual profit component and the depreciation allowance.

There are no incentives for minimising cost of capital, i.e., optimising capital structure as the Order N 72-э/3 determines residually the required profit level and limits the decision freedom of the companies' financial management with respect to their capital structure. It is understandable that the FEC strives to ensure financial viability of the regulated network service providers and reliable service delivery. However, in keeping with an incentive-based approach, FEC may need to establish forward-looking benchmarks for calculation of the WACC based on the yardstick of an efficiently financed business in Russia. This would involve making assumptions regarding a normal, efficient capital structure, independent of the specific ownership and financing arrangements for the regulated network service provider. Accordingly, the WACC should be calculated on the basis of an efficient standard equity/debt ratio.

### Postage Stamp Versus Locational Pricing

Order N 72-э/3 applies a network service model (point-of-connection model) that allocates the transmission network cost to the connected transmission service users rather than to individual transactions. Order N 72-э/3 suggests cost allocation based on a postage stamp approach differentiated by transmission region and voltage level. Postage stamp tariffs have been applied in a number of countries with a varying level of inclusion of cost components. Depending on the size of the transmission network area, this approach might provide no incentives for proper location, since postage stamps will not differ in dependence of zone of connection nor of the distance involved in a transaction.

### Long-Term Versus Short-Term Locational Price Signals

The new wholesale market design in Russia establishes short-term market signals through different prices at the transmission network nodes. The nodal pricing regime manages congestion and sets prices for transmission losses and transmission constraints through a centralised energy market based on economic dispatch. Additionally, long-run signals could be established by locational differentiation of the transmission network cost across the system. The transmission service users pay transmission use of network charges only for their connection point to the transmission network, irrespective of the transaction concluded. This transmission use of network charge, however, will differ across the system, depending on where the transmission service user is connected to the transmission network.

### Pricing Concept

Order N 72-э/3 applies average cost pricing. Under considerations of cost recovery, the proposed average cost pricing may appear preferable. The average cost pricing reflects the cost of the existing transmission network assets. Such a pricing scheme, however, could fail to achieve allocation efficiency. In periods with excessive network capacity, transmission use of network charges would be inefficiently high, while in periods when the transmission network is heavily loaded prices would be inefficiently low. This may lead to an excessive network utilisation that would make an inefficient extension of the network necessary. An alternative solution could be to apply methods based on long--run marginal cost (LRMC). The purpose of transmission pricing methodologies based on LRMC is to signal to the transmission service users the costs of network expansion due to a marginal (incremental) increase in generation or demand.

### Cost Cascading and Tariff Design

Order N 72-э/3 cascades the transmission cost onto network voltage levels in proportion to the total connected load. Further, Order N 72-э/3 denominates transmission use of network charges in RUB per kW (demand charges). The arguments in favour of peak-based charging and/or cost allocation methods arise because transmission networks are dimensioned on the basis of reliably serving peak system demand and most of their costs are fixed. The exact form of a demand charge can vary. The peak demand can be defined as the single highest period of demand or supply over a number of years, within each year, or the average demand/supply over a small number of periods within the year. Alternatively, the peak demand can be defined as the contribution to system peak demand or regional peak demand, the customer's own peak, connected load or the peak for each supply point servicing a customer.

Simple aggregation of connected customer load for the purposes of cost allocation as suggested by Order N 72-э/3, however, may fail to reflect the actual use of the transmission network and the coincidence of the time occurrence of peak demand at each voltage level.

Using only the connected customer load but not the measured demand would decouple completely the chargeable basis (connected customer load) from the actual usage of transmission network service (measured demand) and entirely ignore the real time load flow and actual use of network. Therefore, sometimes combinations between measured demand and connected customer load are used (e.g., in the Netherlands). Further, imposing demand charges on the basis of a single measured demand value would make the total payment for transmission use of network charges dependent solely on this demand value and would expose the transmission service users to stochastic demand fluctuations. In order to avoid these adverse effects, additional chargeable demand values could be considered for the purposes of payments for transmission use of network charges.

Pure demand charges (even using more than one peak period to determine the payment liabilities) could create substantial differences for the payment of transmission service users having different load profiles. The effective payments by transmission service user with higher load factors could be significantly lower than those payments by transmission service users with lower load factors. In contrast to demand charges, an energy charge encourages under-consumption of electricity but is likely to have some positive impact on the energy market.

Energy charges might be desirable to meet some equity objectives as large customers pay a greater share of energy charge revenues than smaller customers. Energy charges achieve these equity objectives and are less likely to result in distortions in the energy market, since relative energy costs in different demand periods are not distorted. Energy charges might be more appropriate than demand charges, where there is excess network capacity and high demand charges would provide a perverse economic signal to restrict the economic usage, whereas on the contrary higher network utilisation should be encouraged.

Finally, fixed charges – appropriately applied – are less distortional than variable charges, are simple to apply, could be differentiated by size of customer, and would reduce the taxation impact on consumption. Fixed charges are the first best solution for sunken cost recovery under marginal cost pricing. However, imposing high fixed charges without any link to demand or energy will certainly face significant opposition from the customers.

**Payment Liability**

Order N 72-э/3 allocates the payment liability solely to load entities. Secure transmission network benefits both load entities and generators. Without a reliable transmission network, the generators (connected to the transmission network) are not able to feed in electricity into the transmission network. On the other hand, reliability is also important for load entities. A measurement for the value of reliability for load entities is the value of lost load. It could be argued that the foregone revenues of generators not allowed to generate electricity due to an unreliable transmission network are lower in relation to the value of lost load. However, such arguments would not change the principle that a fair allocation of payment liability will need to consider generation and load entities.

**Price Control**

The major ideas of Order N 226 follow the logic of rate of return regulation. The document leaves open options for application of alternative methods; however, it does not elaborate further details concerning these alternative methods. Lack of transparency and clarity on timing and conditions for application of alternatives might increase significantly the regulatory risk in the Russian power sector. Under conventional rate of return regulation rates of regulated service, providers are reviewed on a regular basis and have to be adjusted (process called "regulatory review") to lower levels if cost savings have been achieved in the interim (since the last review). Regulated service providers thus only benefit from cost savings to the extent that regulatory reaction to cost savings is lagged (this lagged reaction is called regulatory lag). If the regulatory lag is short – one or two years – incentives for cost savings are suppressed almost completely. In such a framework, regulated service providers are said to operate dynamically inefficiently. Additionally, the rate of return regulation creates overcapitalization (over-investment in fixed assets) incentives and potentials for increase of capital costs. More advanced forms of price regulation (incentive regulation) have been developed and applied in other countries. These methods of regulation decouple totally or partially the allowed revenue/prices from the actual costs of the regulated service provider and provide efficiency increase incentives via retaining the achieved cost saving in the companies for a certain time period.

## Use of Network Charges

### Flat Versus Locational Charges

The new Russian wholesale market design is based on nodal pricing and internalises in this way the SRMC of transmission (marginal transmission losses and congestion).[1] Short-term market signals are established through different prices at the transmission network nodes.[2]

Additionally, long-run signals could be established by locational differentiation of transmission network costs across the system. In this case, transmission service users would pay transmission use of network charges only for their connection point to the transmission network, irrespective of the transaction concluded. This transmission use of network charge, however, may be different across the system, depending on where the transmission service user is connected to the transmission network.

Examples of locational pricing for the network infrastructure components could be found in the practice of England & Wales and Australia. The transmission use of network charge in England & Wales reflects the cost of installing, operating and maintaining the transmission system. It is calculated on the basis of the Investment Cost Related Pricing (ICRP) Method developed by National Grid Company (NGC) and using a LRMC pricing concept. Investment Cost Related Pricing builds upon a complete new network alongside existing rights of way and ignores real power flows. Instead, it bases its methodology on the transportation problem, a linear optimisation problem of the operations research literature. The use of system charge presents a zonal fee (14 zones exist), payable according to the location in the grid, rather than to the type of transaction. In Australia, the Cost-Reflective Network Pricing (CRNP) has been developed in preparation of the National Electricity Market (NEM). The CRNP method allocates approximately 50% of the annual revenue requirements of the transmission companies involved on a locational basis. The remainder is allocated on a postage stamp rate basis.

### Marginal Versus Average Cost Pricing

The approach of using marginal cost based prices as signals for efficient network extension attempts to replicate the outcome on the competitive markets. In these markets, producers sell at the competitive market price whenever it is equal to or greater than their SRMC of production. Short-run marginal cost of transmission are defined as the additional costs arising when one additional kWh is demanded and the installed capacity remains constant. Differently, LRMC take into consideration also the transmission network investments when one additional kWh is demanded.

Short-run marginal cost pricing typically implies that the marginal price for a non-congested transmission network is equal to the marginal transmission losses. Whenever the network is congested, SRMC will increase by consideration of opportunity cost. If a power system had a perfect network with no losses, or any kind of constraints limitations, at a given instant in time, the spot prices at all nodes of the system would be the same; i.e., there would be no spatial differentiation of spot prices. In any real network, the effect of losses and network constraints creates differences among the nodes and consumers pay, or generators are paid, different prices, depending on their locations. Those parties should be allowed access that can make most efficient use of the net-

[1] The underlying principle of SRMC nodal pricing method is that spot prices between two nodes should not differ for more than the SRMC of transporting the electricity. If a power system had a prefect network, with no losses or any kind of constraints, at a given instant in time, the spot prices at all nodes of the system would be the same, i.e., there would be no spatial differentiation of spot prices. In any real network, the effect of losses and network constraints creates differences among the nodes and consumers pay or generators are paid different prices, depending on their locations. The basic theory of real-time or spot-market pricing of electricity was developed by Vickery and Schweppe et al. As set forth by Schweppe et al., the optimal price for electricity is differentiated in space and time and accounts for the variable costs of producing any electricity at the time it is used and any added requirements to compensate for whatever transmission losses accompany the supply and delivery of the electricity used, and any generating or transmission capacity limitations that might influence the availability of supply as a function of time. See: Vickery, W. (1971): Responsive Pricing and Public Utility Services, Bell Journal of Economics and Management Science, Vol. 2, pp. 337–346; Schweppe, F. / Caramanis, M. / Tabors, R. / Bohn, R. (1988): Spot Pricing of Electricity, Kluwer Academic Publ., Boston, MA.

[2] In the framework of full nodal pricing, transmission constraints and transmission losses are integrated in the market price calculated for each node in the transmission network.

work in periods when network constraints bind. This will typically be the suppliers that can generate/procure electricity at lowest variable cost at the time of transmission constraints and that try to contract with the customers with the highest willingness to pay. The underlying principle of SRMC pricing method is that spot prices between two nodes should not differ for more than the SRMC of transporting the electricity.

An alternative solution could be to apply methods based on LRMC. The purpose of transmission pricing methodologies based on LRMC is to signal to the transmission service users the costs of network expansion due to a marginal (incremental) increase in generation or demand. The difference from the SRMC pricing concept is that in this case, the network assets are not considered static but rather dynamic, i.e., new investment resulting from incremental demand or generation is taken into account.

The average cost pricing reflects the cost of the existing transmission network assets. Under consideration of cost recovery, the average cost pricing may appear preferable. Such a pricing scheme, however, could fail to achieve allocative efficiency: in periods with excessive network capacity, transmission use of network charges would be inefficiently high, while in periods when the transmission network is heavily loaded prices would be inefficiently low. This may lead to an excessive network utilisation that would make inefficient network extensions necessary.

## Practical Solutions

### *Contract Path*

The contract path method requires the nomination of a particular grid supply and receipt point for a bilateral transaction and of a path between these two nodes. It addresses both the amount of contracted capacity to transport as well as the distance associated with the transport over the contract path. Power flows, however, in accordance with the physical laws, taking always the way of least impedance and not along a contract path, thus causing loop flows. The contract path method will thus not represent the physics of the system. It refers to the cost components of the particular contract path. The contract path method can be based on average cost (valued with rolled-in embedded costs or replacement costs) or marginal (incremental) costs. The contract path may allocate a share of the costs of the contract path to the transaction or draws on the incremental costs of providing the transaction.

### *Postage Stamps*

The postage stamp approach represents the opposite to the contract path method. Cost components of a number of a particular asset category or all categories are allocated to all customers on a pro-rata share. The allocation is normally done based on the individual share in the coincidental peak. The approach is used in a number of countries with a varying level of inclusion of cost components.

### *MW-Miles Method*

MW-miles methods could be used for pure bilateral transactions but also for multilateral transactions. To calculate the MW-miles, the power flows over the circuits have to be multiplied with the electrical distance of the circuits. All products are summed up to the overall number of MW-miles transported (total MW-miles transported). In case of a bilateral transaction, two load flow calculations will be necessary, without and with the transaction. The incremental effect of the transaction is estimated as a difference between the product of load flows and the affected lines without and with the analysed transaction (incremental MW-miles). The ratio between the incremental and total MW-miles can be used to allocate the transmission cost to the individual transaction.

A major problem in applying the MW-mile method on multilateral transactions (no direct relationship between a generating unit and a load can be identified) is the allocation of an increment of load at one bus to the increase in generation at the various buses and vice versa for an increment of generation. The MW-mile approach calculates the MW-miles between two nodes for a particular circuit. It does not provide immediately a method or a tool to allocate the MW-miles to a particular node.

### *Location Dependent Pricing*

In the context of the point-to-point service approach, the grid charges according to this solution will reflect the contract path and depend on the distance of electricity transport. In the network service-pricing scheme, locational differentiation of transmission fees is created by different nodal (or zonal) prices across the system. The transmission users pay a transmission fee only for their connection point to the grid, irrespective of the transaction concluded. This transmission fee, however, may be different across the system, depending on where the transmission user is connected to the network. Locational pricing is applied in the form of SRMC or LRMC pricing.

*Short-Run Versus Long-Run Price Signals*

To the extent that generators and large customers are less insulated from market signals, it should be recognised that SRMC may provide a relatively stable signal, even in the longer term. Generators and large customers receive some locational signals via nodal prices in the energy market. Other things being equal, the nodal prices are likely to encourage:

- large customers to locate in areas where nodal charges are low, and hence, indirectly, network augmentation costs are reduced; and
- generators to locate near nodes where nodal prices are high and hence where generation at such nodes is likely to reduce losses, congestion and associated network investment.

On the other hand, reliance on pure SRMC charging to signal and recover the costs of the network raises inter-temporal equity issues, given the long life of network assets. Long-run marginal cost charges are likely to be more equitable in this regard. Under full SRMC charging (full nodal pricing), network users would pay for investment via accumulated future losses and congestion rentals. Even if SRMC were sufficient to recoup the costs of investment, it is unclear whether those parties paying for short-term losses are the same as those benefiting from network expansion in the longer term. Therefore, additional locational signals could be established through locational pricing of transmission network infrastructure. However, one should be aware that a transmission pricing design based on a locational concept is characterised by significant complexity that will naturally limit transparency and practicability of the transmission pricing approach, and ultimately could hazard the successful implementation. Moreover such locational pricing will create geographical price differences (supplementary to the differences resulting from nodal pricing) and may face significant social and political resistance. If the FEC and/or FNC prefer to consider additional locational pricing for transmission infrastructure in order to establish long-term signals it is recommended to use locational pricing only for a portion of revenue requirements. Depending on the methodology applied, this portion will be either ex–ante determined (approach used in Australia) or will result from explicit modelling of LRMC (approach used in UK).

Below, some methods for locational transmission pricing used in different countries are described, namely:

- Investment Cost Related Pricing applied in England and Wales (zonal pricing based on LRMC);
- Cost-Reflective Network Pricing applied in Australia (zonal pricing based on LRMC);
- Pricing based on marginal transmission losses applied in Norway (zonal pricing based on SRMC);
- Full nodal pricing applied in New Zealand, Singapore (nodal pricing where transmission constraints and losses are handled on the market).

*Investment Cost Related Pricing*

The use-of-system charge is set by NGC for making available its transmission system for the bulk transportation of electricity. The use-of-system charge reflects the cost of installing, operating and maintaining the transmission system. It is calculated on the basis of the ICRP method developed by NGC and using the LRMC pricing concept. The charges are levied on both generators and demands but the charge rate varies for each. A rate is determined and set for generation or demand in each area. These prices vary with location and are derived based on an analysis of capacity requirements. In the north of England, significant amounts of coal-fired generation have traditionally been available while most of the demand is in the south. Therefore, NGC charges so as to discourage generation in the north of England while encouraging demand there. Thus, the generation charge is relatively high while the demand charge is relatively low. Conversely, in the South of England, the generation charges are low, and in some cases negative, so as to encourage generation at locations where the economic incentives would not otherwise encourage generation. In these same locations, demand charges are very high so as to discourage the excess off-takes. This location signal is required because the spot prices in the UK pool do not vary with location, and hence cannot convey a signal as to the relative value of new generation capacity or load in different locations.

Investment Cost Related Pricing builds upon a complete new network alongside existing rights of way and ignores real power flows. Instead, it bases its methodology on the transportation problem, a linear optimisation problem of the operations research literature. The concept of ICRP was introduced by the NGC in England and Wales in 1993/94. It is applied to use of system charges for the transmission network (275 – 400 kV). Investment Cost Related Pricing represents a quasi-LRMC approach intended to provide locational signals based on the cost of network expansions. The concept comprises three steps and can be summarised as follows:

- In a preliminary step, NGC derives the "expansion constant" as an annuitised value of investments in new transmission capacity per MW and km;[3]
- The second step represents the core of the methodology. Investment Cost Related Pricing assumes the building of a complete new network alongside existing rights of way. In addition, the method ignores real power flows but creates a new set of optimal flows by solving a linear optimisation problem (transportation problem), based on the "triad demand" of each load connection point and the registered capacity for each generator;[4]
- The third step determines the incremental cost for 1 MW of additional generation or load at each network node. For this purpose, incremental costs are derived starting from an arbitrarily chosen slack node with an assigned incremental cost of zero. For the sake of simplicity, geographically adjacent nodes with similar incremental cost are finally aggregated into zones.[5]

The ICRP thus derives the quasi-LRMC for 1 MW of network capacity at system peak load for each node of the network. The resulting incremental charges are higher (lower) for generation (load) in the north and actually negative for generation in some southern zones.[6] As revenues from this locational tariff element amount to less than a quarter of total cost, they are supplemented by a uniform postage stamp that is supposed to cover the security cost of the network.

The major advantage of the ICRP is that it is focused on the explicit determination of long-run average incremental costs (an approximation of LRMC) of the network, i.e., the method aims to provide a realistic picture of the LRMC using a direct approach for their calculation. Moreover, the method provides rather stable figures regarding the transmission charges and, thus, does not lead to unnecessary price fluctuations, because the results depend only to a limited degree on the actual system conditions and costs.

On the other hand, the method uses a substantial approximation of the network topology for the purposes of the LRMC derivation. Further the method opens some contentiousness over data because of the approximate assumptions in the generation dispatch, an increase in demand of 1 MW at node *i* is assumed to be supplied out of the reference node and for an increase of generation at node *i* vice versa. The ICRP ignores the physical properties in the existing capacity of the transmission network but does preserve the geography of the system rights of way. Flows are represented by directed flows of material as on a rail, road or pipeline and ignore more or less the differences related to the nature of electric power. Finally, the ICRP explicitly ignores the existence of scale economies (assuming instead that the costs of new construction are totally linear in the length of the line and the megawatts of capacity) and lumpiness (assuming instead that additional transmission capacity can be built in any size increment desired).[7]

*Cost-Reflective Network Pricing*

The CRNP cost allocation is a nodal based method and it requires all of the costs to be attributed to the links ("lines") between the nodes on the network. Some assets are wholly dedicated to providing network service to a participant or group at a single point. These assets can be allocated to the participant directly at the node. Other assets throughout the network are shared among the users and the relative use by each participant must be determined. The shared network costs are allocated amongst users based on the marginal change in network element current flows as a result of an increment of user load at each bus. The generation source for each load is defined using the "electrical distance" as a measure of the capability of a generator to supply each load point. Using this approach, a greater proportion of load at a particular location is deemed to be supplied by generators that are electrically closer than those that are electrically remote. The generator to load allocation is carried out according to relative fault contributions by each generator to a three-phase fault at each load point.

The following critical points have been pointed out by NECA's[8] review on the CRNP transmission pricing approach. It can lead to perverse pricing signals because it seeks to reflect total, rather than marginal, costs to

[3] To reflect the different cost of lines and cables, cable lengths are adjusted (increased) in relation to the cost difference between lines and cables. All investment costs are valued at replacement cost.

[4] Generator ratings are uniformly scaled down so that total installed generation equals total demand. Triad demand is defined as the half-hour during the year with the highest demand for electricity and the next two highest half-hours, which are at least 10 days apart from the highest half-hour and from each other.

[5] There are currently 15 generation zones and 12 demand zones.

[6] However, these bonuses are not paid out unless there is a guaranteed availability.

[7] Some additional critical points raised by Ofgem are as follows: (1) the charging basis is hard to justify. Either the security charge should be differentiated by location or the expansion constant should be revised upwards; otherwise, there is a danger that NGC's charges artificially stimulate the demand for more transmission lines; (2) the suppliers pay 75% and generators 25%. An adjustment towards increase of generators' contribution may be appropriate; and (3) the practice of scaling generation by the ratio of peak demand to registered generation capacity has the effect of reducing the zonal differentials for generation charges, compared with those for demand charges, so distorting the locational signals.

[8] NECA is the Australian National Electricity Code Administrator.

customers. This means that CRNP takes no account of the level of spare capacity on the system in setting prices. Therefore, if the system is at full capacity, CRNP will produce a lower unit price compared to a situation where there is spare capacity. Such pricing signals are the opposite of those that one would expect to see in a competitive market. Secondly, it allocates costs on the basis of load flows in an attempt to identify the users of particular assets on the system. Some of these load flows may, however, be subject to significant changes on a periodic basis. Such changes can create volatile use of system charges as the cost allocation methodology reallocates charges between customer groups.[9]

*Pricing Based on Marginal Losses*

Both previous methods represent different variations of LRMC pricing. The use of marginal losses, on the other hand, serves as a proxy for SRMC pricing.

Marginal losses are typically calculated for different nodes (or zones) of the network by means of a power flow analysis. In the case of Norway, these losses are then expressed as a percentage of load or generation at each node, the so-called loss coefficient. However, as marginal losses depend on the loading of the network, they may change considerably over time. For this reason, loss coefficients in Norway are differentiated between high- and low-load hours,[10] and are updated several times a year (every eight weeks). Loss coefficients may be both positive and negative, but are subject to a maximum cap of ±10%. All losses (i.e., the product of load or generation and the corresponding loss factor) are valued at the NordPool spot price for the relevant hour. Finally, energy-related payments for marginal losses only represent one part of the Norwegian tariff system, while the residual network costs are recovered by means of other tariff elements.

Apart from locational marginal pricing for wholesale spot markets, the use of marginal losses represents the second dominant application of SRMC pricing. While the approaches discussed above focus on investment decisions, this model primarily aims at short-term, i.e., operational or dispatch decisions. However, provided that marginal losses remain stable over an extended period of time, this method will simultaneously influence investment decisions as well. In summary, marginal loss pricing therefore clearly has the ability to improve short-run allocative efficiency but it may not necessarily provide correct long-term signals.

To achieve the method's full benefits, loss coefficients would basically have to be calculated in real-time. In practice, this is not possible which explains the decision by Statnett to provide estimated data for predetermined time periods. Due to the variation of losses over time, this adjustment causes a deviation between the standardised loss coefficients and actual losses. For this reason, pricing signals from marginal loss coefficients in Norway will not generally be truly identical to the system's SRMC, thereby somewhat reducing the scope for achieving allocative efficiency. The same basically holds true for the limitation of marginal losses to ±10%, which will provide insufficient (i.e., too small) pricing signals at all places where marginal losses are capped.

A similar dilemma can be observed with regards to the price of marginal losses. In theory, these would also have to be determined in real-time. In Norway, hourly spot prices from NordPool arguably serve as a good proxy. But this approach depends on the existence of an (liquid) electricity spot market. An alternative approach has been implemented in Sweden, where the energy price of losses is determined for one year in advance (separately for four time periods), based on the bulk-purchasing rate that Svenska Kraftnät has to pay to its suppliers. But this again introduces strong elements of average cost pricing and reduces the scope for allocative efficiency.

Like all other methods of marginal cost pricing, this approach does not allow the full recovery of network cost by the marginal element.[11] Nevertheless, marginal loss payments actually lead to an income in excess of the total cost of network losses, called transmission loss rentals,[12] allowing a slight reduction in payments for the residual cost.

Compared to the approaches for LRMC pricing discussed above, marginal loss charges will be much more volatile. This does not harm short-run allocative efficiency as it will provide adequate signals for dispatch decisions. But the lack of stability can be detrimental with regards to long-term investment decisions, especially when marginal losses fluctuate considerably or, in the worst case, may be both positive and negative at the same lo-

[9] This point mentioned above under the NECA's review is related to load dynamics and price stability. This argument is strongly correlated with the Australian regulatory price control that is based on the revenue cap regulation. As the allowed revenues are determined for the duration of the regulatory period, load changes could cause recalculation of charges and redistribution of revenue collected from the customers. In this sense, the argument is relevant for the inter-temporal charge dynamics and the regulatory price control modus. The regulators often used methods imposing side constraints on the price changes in order to limit the price fluctuations.

[10] Day time versus nights and weekends.

[11] Statnett recovers only about 15% of its total network-related cost from marginal loss payments.

[12] This can be explained by the fact that marginal losses grow faster than load. Marginal losses will therefore be higher than average losses, resulting in a net income from marginal loss payments for the TSO. For the same reason, marginal losses cannot directly be applied under a system of self-procurement, i.e., where network users have to supply/purchase a certain amount of energy to compensate network losses. In that case, loss coefficients must either be based on average losses or scaled down to avoid overcompensation and minimise system imbalances.

cation over time.[13] Under these circumstances, a user may no longer be able to assess the impact of marginal losses on his long-term cost.

*Full Nodal Pricing*

The nodal pricing manages congestion and sets transmission prices through a centralised energy market based on economic dispatch. The basic idea of the nodal pricing approach is to organise the market as a pool in which generators (and ideally loads) submit hourly bids for node specific injection and withdrawals of power to the pool operator with full co-ordination and price setting authority. The pool operator minimises the total system's gain from trade (demand bids less generation offers) subject to transmission and reliability constraints. The price at each node is then set to the incremental offer or bid price of the most expensive unit generated or consumed at that node. These nodal prices become the hourly prices charged to loads and paid to generators at the respective nodes. When there is no congestion and losses, all nodal prices are in theory identical.

As prices rise, suppliers will provide more if the price exceeds the cost of providing an additional unit, while consumers will consume less as the price rises above the benefit gained from the last unit being consumed. Dispatchers can use this mechanism to ensure that supply and demand are matched most efficiently. Full dynamic SRMC-based spot pricing promotes efficiency by signalling the true marginal cost of producing and transporting power across the network. In this case, the spot price at a particular location will equal the cost which would actually be incurred by supplying one more unit of demand at that location, taking account of the necessary adjustments in generation patterns to re-dispatch generators to avoid interregional constraints, make up losses and maintain system security.

The nodal pricing approach is also able to accommodate bilateral transactions.[14] Such a transaction is then subject to an ex-post transmission charge that equals the opportunity cost of the transaction, i.e., the cost difference of selling the power to the pool at the injection node price and buying it back at the withdrawal node price. Thus, the transmission charge between any pair of nodes is set ex-post to the nodal price difference between the nodes. The cost of transmission, therefore, varies between each pair of locations and is only known to trading parties after the fact.

This regime provides an explicit and accurate signal to market participants of the behaviour required to achieve an efficient dispatch and accounts explicitly for the available transmission capacity. Therefore, if transmission flows between two regions reach the system's flow capacity; the locational price in the receiving region will rise relative to the price in the sending region, which encourages more expensive generators in the receiving region to provide for additional local demand. A revenue (congestion surplus) is earned on the constrained transmission capacity because the transmitted power is sold in the receiving region at a higher price than that for which it is purchased in the sending region.

**Revenue Requirements and Cost Recovery**

***Establishment of Revenue Requirements***

Revenue requirements are equivalent to the justified costs that should be allowed to be recovered through selling of the regulated transmission service. The terms "economically justified" and "eligible cost" are widely used by the regulatory authorities. Eligible costs should include the reasonable efficient O&M and capital costs (including depreciation and return on assets). The O&M costs are often considered the "cash outlay" costs of an infrastructure business. Recovery of these costs does not provide any return to the infrastructure owner, as they are paid out in the form of salaries, ongoing O&M costs, emergency service costs, etc. These costs allow the business to provide and maintain service. On the other hand, the inclusion of capital costs in the revenue requirement formula recognises the owner's investment in the regulated utility and the capital-intensive nature of network infrastructure businesses.

The regulation should recognise the importance of recovering sufficient level of O&M and capital cost. However, it is important that the regulated network service providers do not incur excessive or unnecessary costs in providing their services. It is in the interests of all concerned that the FNC and interested parties are able to examine the level of current and forecast costs, and are able to compare those costs with other similar entities (abroad).[15]

[13] This may, for instance, be the case under seasonally changing flows patterns.

[14] A physical bilateral transaction can be scheduled as if the injection submitted a zero offer and the load submitted an infinite bid.

[15] Benchmarking is sometimes used by the regulators to disclose the efficient cost that should be allowed to the regulated service providers. The output of the benchmarking process is the relative degree of inefficiencies of a particular regulated network service provider in comparison with the best industry practice. Electricity networks use a wide range of inputs (capital, labour) to provide services to customers. While all network service providers use broadly the same inputs, some providers may use proportionately more of some inputs and less of others. The mix of inputs used depends, among other things, upon management practices and the operating environment. Similarly, the nature of services provided by networks varies according to the nature of customer demands. For example, some network service providers may need to maintain significant network capacity to transport power to a small number of customers while others may serve a large number of customers with limited and highly variable demand. The benchmarking technique must be able to accommodate these different supply and demand conditions. Further, the technique must recognise that it is possible to operate at equal efficiency while accepting different inputs and supplying different outputs.

Establishment of revenue requirements (including capital and O&M costs) is a topic that should be discussed for the purposes of network pricing design (cost allocation and tariff structure) and network price control concept (price regulation).

*Asset Valuation*

The inclusion of capital costs in the revenue requirement formula recognises the owner's investment in the regulated utility and the capital-intensive nature of network infrastructure businesses. Failure to include adequate capital related costs as part of the revenue requirement of the regulated business risks a reduction in investment in the industry. This could ultimately lead to reductions in cost coverage and quality service levels, hence to a reduction of the security of supply in the medium and long term. Fundamental to the measurement of capital costs in the revenue requirement is an assessment of the regulated business' capital investment and the establishment of a regulatory asset base (asset value that is used for the calculation of return on assets). Arguments about asset valuation vary from insisting ownership rights be recognised, to questioning whether any value should be attributed to sunken investment. The regulators usually endorse particular asset valuation methodology. These asset valuation issues must be considered with regard to the functional adequacy of regulated assets, market assets value, overall profitability of the regulated business and sustainable cash flows of the business as well as equity considerations.

The review process did not find explicit statements in the Russian documents of how the opening asset value has been computed for the purposes of price determination. Based on these investigations, one might assume that the existing assets are valued on a historic cost basis. The historical cost methodology values assets at their original purchase price. This approach has the advantage that it is administratively efficient and can be easily audited because the data should be available from financial statements. Further, it is relatively inexpensive since it does not require the explicit computation of regulatory asset values and it is objective because it relies on actual data rather than judgements. However, the method exhibits some disadvantages. Historic costs may understate asset prices in times of inflation and overstate asset prices in times of technological change. Secondly, it may lead to unstable prices (e.g., prices may rise when new, more expensive assets replace existing assets). Thirdly, data may be inadequate (especially for assets that have been acquired a long time ago). Historic costs are generally applied in the regulatory price control in the USA. Also, a number of regulatory authorities in Europe applied current book values from financial statements of the regulated service providers for asset valuation purposes in the regulatory accounting.

On the presumption that some compensation for asset or capital devaluation through inflation must be conveyed to the investors in order to allow for its long-run operation (new investments have to be paid for in current values and must be partly financed through revenues in the past), inflation compensation can in principle be achieved through the inclusion of asset replacement costs in the regulatory schemes. The replacement cost methodology calculates the cost of replacing an asset with another asset (not necessarily the same) that will provide the same services and capacity as the existing asset.

One interpretation of depreciated replacement costs is that it is a valuation methodology consistent with the price charged by an efficient new entrant into an industry. Hence it is consistent with the price that would prevail in the industry in a long-run equilibrium. A second interpretation is that replacement costs reflect the price that a firm with a certain service requirement would pay for existing assets in preference to replicating the assets.

The main economic principle for assessing the economic value of any assets is that their value to investors isequal to the net present value of the expected future cash flows generated by those assets. The practical difficulty in making this assessment for regulated monopoly businesses is that the future revenue derived from the assets is itself determined by the regulator – hence the issue of circularity associated with the use of discounted future cash streams as a methodology to value sunk assets. This potential circularity could be eliminated by the use of a replacement cost approach. The value of a network is the sum of the depreciated replacement cost of the assets that would be used if the system were notionally reconfigured so as to minimise the forward looking costs of service delivery.[16]

The application of top-down calculations based on the discount cash flow approach would be hardly possible because of lack of a stable price starting point in Russia.[17] Moreover, the application of final consumer prices as a starting point for

[16] In the Australian regulatory practice, this approach is called "optimised depreciated replacement cost." In this way, additional emphasis is put on the regulatory option to eliminate redundant network assets. See: IPART, Independent Pricing and Regulatory Tribunal of New South Wales (1999): Draft Statement of Principles for Regulation of Transmission Revenues, May.

[17] The bottom-up calculations give the revenues that the company should be allowed to recover if its asset values have been fixed. If its prices have been fixed, then the calculations can be reversed (top-down) to derive the implied asset value, as discussed in the previous sections (using recoverable amount approach). This is effectively how the value of a company in a competitive market would normally be estimated by stock analysis deciding whether to buy or sell shares. In this way, the regulatory asset value is set equal to the predicted market value of the company (i.e., the present value of the expected stream of free cash flows).

discounted cash calculation could lead to low assets value if these prices are not cost-reflective. Obviously, this will be the logical outcome since the revenue streams that could be earned with the particular assets will depend on the predetermined starting prices used in the calculation. Further, even if there is a political will to correct the starting point of calculation, the determination of the normative level of these starting prices will be associated with a number of methodological difficulties mostly related to the "circularity" problem. The "circularity" problem results from the fact that on one side, the regulatory assets value depends on the revenues (prices) as they determine the cash inflow in the discount cash flow calculation. On the other side, the common logic of the regulatory process requires the regulatory asset value as an input in order to determine the revenues and prices.[18]

All the issues described above necessitate the application of proxy approaches for determination of regulatory asset base. Generally, one will need to compare the usefulness of the historic and replacement cost approaches. The application of historic costs for the establishment of the opening value of the regulatory asset base is also unlikely to be appropriate due to the strong inflation figures in the past. If in Russia the assets have been revaluated (in many of the Central and Eastern European countries, the assets have been revaluated), the current balance sheet values reflect more the replacement costs (at the time of revaluation) than the historic costs.

The application of replacement costs will require special attention. As already mentioned, the replacement cost methodology calculates the cost of replacing an asset with another asset (not necessarily the same) that will provide the same services and capacity as the existing asset at the time of valuation. The application of replacement costs should be checked by FEC with respect to the following major issues:

! Is FEC competent to impose revaluation of assets for regulatory purposes and does it contradict with the national statutory accounting?

! Is it feasible to revalue the assets?

! If price should be raised after the assets revaluation is it politically and socially acceptable?

*Depreciation*

To an accountant, the term 'depreciation' means a systematic allocation of the cost of an asset to the accounting periods in which the asset provides benefits to the entity. This allocation is designed to mirror the consumption of the service potential or economic benefits associated with an asset over its useful life, resulting from both use and obsolescence. The purpose of provisions in accounting is to ensure that the cost of the flow of services provided by capital assets is met in the price of these services, and additionally to build up funds for the replacement of these assets.[19]

In the companies' accounting, a number of approaches are used for constructing the depreciation schedule. These include:

! Nominal (or historic cost) accounting method – usually straight line or diminishing balance reductions[20] in the original cost over time;

! Current cost accounting method (based on replacement cost estimates) – again straight line or diminishing balance depreciation based on the age of the asset is applied; and

! More flexible arrangements whereby depreciation is adjusted to complement other components of return so that the revenue stream mirrors the behaviour of an annuity.

While the assets valuation aspects have been discussed in the previous section, here an explanation will be given on the establishment of depreciation profiles. In traditional regulatory frameworks, straight-line depreciation is the norm and the use of straight-line depreciation by FEC would be consistent with those frameworks.[21] In common with the traditional approach used in valuation, this approach calculates the write-down of the gross asset value to obtain the depreciated asset value, by assuming a linear relationship between accumulated depreciation and the age of the asset relative to its expected economic life. This approach amounts to assuming

[18] Even if a privatisation value for the Russian network service providers would exist, this would not solve the circularity problem. The regulatory framework should pre-date the privatisation process and, on the other hand, the privatisation value is needed for the determination of the regulatory asset base. Moreover, experience shows that in many cases, the regulators and energy policy makers seek to establish a pragmatic approach that leads directly to the same quantitative answer instead of searching for economically "perfect" solutions with a low degree of implementation potentials.

[19] See IPART, Independent Pricing and Regulatory Tribunal of New South Wales: (1999) Rolling Forward the Regulatory Assets Base in Electricity and Gas Industry, Discussion Paper, January.

[20] Often the companies' accounting uses accelerated depreciation methods. The mostly used such method is the method of declining balance. According to this method, a fixed percentage of the written value of the asset is charged as depreciation each year. The effect of this is that decreasing amounts are charged each year in contrast with the straight-line method that produces an equal charge each year. The fixed percentage to be used is the percentage that should be deducted from the written down value each year so that, over the life of the asset, the total installed cost is reduced to the net scrap value.

[21] Straight-line depreciation is not immune to critique and FEC and FNC should be aware of this. One of the major issues regarding the determination of regulatory depreciation is the need to achieve a time profile of revenues that is economically efficient. It could be argued that straight line depreciation fails to capture the important features of economic depreciation that are evident from the sale value of assets or the pricing of products over the life-cycle of productive assets.

that the economic depreciation of an asset is equivalent to straight-line depreciation of that asset. It makes sense that the approach to depreciation taken within the cost valuation methodology matches the one utilised within the regulatory framework.

*Return on Assets*

Regulated service providers compete for finance with companies operating in competitive markets and thus have to accept these conditions. Equity and debt finance will only be available to utilities that agree to credit conditions given to firms that operate in competitive industries and have a comparable credit ranking. Equity finance will only be available if profitability (consisting of a dividend component and market value growth component) can be expected that covers the risk-free rate of interest (i.e., yield of long-term credible government bonds) and a risk premium (where risks and risk premium can be expected to lie significantly below corresponding values in competitive industries, provided that the political and regulatory environment is predictable and stable). Investors will be interested in engaging in the regulated industry, if projects allow them to meet financial requirements. These financial requirements are measured against the benchmark of earnings to be made in other product markets, in industries in other countries or in the international capital markets.

Given the capital-intensive nature of electricity network businesses, the return on capital component of the regulated revenue could account for 50% or more of annual aggregate revenue. As relatively small changes to the rate of return can have a significant impact on the total revenue requirement and ultimately on end-user prices, it is important that the regulator sets the rate of return at a level that reflects a commercial return for the regulated businesses. Setting a rate of return below the cost of funds in the market could make continued investment in developing the network difficult or unattractive for the owner. This would create pressure for the regulated service providers to reduce maintenance and capital expenditure below optimum levels and undermine the quality of service offered to users. Conversely, if the rate of return was set too high by the regulator, the regulated businesses would earn a return in excess of their cost of capital. This would distort price signals to consumers and investors, resulting in a misallocation of resources and sub-optimal economic outcomes.

The WACC is a commonly used method for determining a return on an asset base. Weighted Average Cost of Capital is determined in the regulatory regimes as the weighted average of the cost of each individual component of the capital structure weighted by its share.

$$RR = WACC_t = EP_t \times ROE_t + DP_t \times DI_t$$

where:

$WACC$ – Weighted Average Cost of Capital for period *t*;

$RR$ (%) – Rate of Return for period *t*;

$ROE$ (%) – Return on Equity for period *t*;

$EP_t$ – proportion of capital comprising equity and equal to (value of equity)/(value of equity + value of debt) estimated at the end of period *t*;

$DI_t$ (%) – cost of debt for period *t*; and

$DP_t$ – proportion of capital comprising debt and equal to (value of debt)/ (value of equity + value of debt) estimated at the end of period *t*.

Traditional industry use of WACC is to determine the value of the cash flows resulting from an investment to assess its profitability. It is a common practice in industry to determine the after-tax cash flows and apply a nominal after-tax WACC discount factor to those cash flows in order to develop a present value. In the regulatory environment, WACC is applied to an asset base in order to determine the cash flows a business will receive. A number of regulatory authorities have determined that a pre-tax real WACC should be applied to an inflated value of asset base in order to determine ongoing returns. The use of pre-tax WACC is used due to the reversal usage of the WACC formula and the tax affairs being viewed as a matter for the business and not the regulator. Some regulatory offices use after-tax WACC and consider taxes explicitly in a separate position similarly to O&M cost.

*Operation and Maintenance Costs*

Operation and maintenance costs are the costs incurred by the FNC in maintaining and operating the transmission network assets to the technical standards used in Russia and the requirements of the Technological Rules. Operation and maintenance costs will cover those cost elements that are spent for accounting purposes.

***Treatment of Transmission Losses and Congestion Surplus***

As a result of the application of nodal pricing, in which generators are paid at nodal prices of generation buses and consumers are charged at marginal costs of load buses, a surplus (transmission loss and congestion surplus) is collected for the owners of the transmission system.

Settlement is performed on the basis of the actual amounts of electricity generated by generators, and actual amounts of energy consumed by loads, as measured by meters distributed throughout the network. Because nodal prices are used in the dispatch and settlements processes, the amount paid by purchasers of electricity (loads) through the settlements process is normally higher than the amount paid to generators. In other words, the settlements process normally results in net income to the market operator (ATS) referred to as the "surplus." A number of solutions are conceivable.

*Reduce Transmission Network Charges*

The first solution is to use the surplus to reduce transmission charges, and hence is indirectly returned to end consumers. This mechanism for distributing the surplus is transparent as it will be part of the FNC's regulated revenue, can be universally applied, and is equitable, as consumers will receive a benefit via reduced transmission charges. It is also efficient in that it will have a minimal impact on the market while retaining economic signals. However, this approach has been criticised in some countries for adding complexity to the market design.

*Return the Surplus Directly to Market Participants*

An alternative suggestion is that the surplus be immediately returned to spot market participants. Accordingly, the surplus would be distributed to certain retailers and/or generators in proportion to the extent and nature of their participation (energy, capacity) in the spot market at the time that the surplus was accumulated. The rationale for this approach is that giving the surplus to market participants will allow them to commercially trade the surplus in the market. It could be argued that this is a more viable option than delivering the surplus to the FNC who is a non-commercial player. It is not directly obvious why generators or retailers have a legitimate claim to what is essentially a revenue related to transmission service functions. Nor is it apparent where the commercial incentive for generators and retailers to pass the surplus on to end use consumers would come from. Returning the surplus to market participants may be a simple option, but given the above concerns over incentives it is doubtful whether this option meets the efficiency or transparency objective. In addition, the assignment of the surplus to generators or retailers raises a number of equity issues that would need to be resolved. Finally, such approach would probably dilute locational signals resulting from the nodal pricing. On the other hand, the latter could be a desirable effect at least for the demand side where most probably the nodal pricing would face some social or political resistance.

*Use the Surplus to Relieve Network Constraints*

The accumulation of a surplus provides a signal that it may be economically viable to augment the networks and hence remove constraints and this could be another solution. However, this option raises a number of issues, such as who becomes the owner of the new transmission lines and is thus entitled to the remaining and future surplus, as well as the regulated income. Moreover, augmentation of the transmission network may not be always the optimal solution.

Having listed above the properties of the different methods, it is preferred to apply the first method where the surplus is used to reduce the transmission charges. If the surplus was to be paid to generators, or refunded to purchasers, then the knowledge that these payments or refunds were to be made could be taken into account when developing bidding strategies, thereby diluting the effect of the marginal pricing concept on which the market is based. With the first option, the FNC can use the settlements surplus it receives to reduce the charges it imposes for the use of the transmission network. These reduced prices should ultimately filter through to reduced electricity prices for consumers. The Australian Federal Regulator ACCC in its determination of 10 December 1997 stated that the proposal to distribute the settlements surplus to end use consumers via reducing network charges is the most transparent, equitable and efficient distribution of the surplus.

***Forms of Cost Recovery***

In systems that present increasing returns to scale, such as the transmission business, this surplus is not sufficient to finance the system operation and development.[22] The collected surplus covers a small

[22] Essentially, a transmission pricing methodology based on a multilateral concept must respond to the following issues: identify the system that must be paid, the costs that must be covered and the allocation of payments among the different agents. Such charges for transmission infrastructure have to cover all the capital and operating costs of the transmission network and are consequently not marginal costs for the system as a whole (sunk cost). Although depending on the chargeable basis in the transmission pricing regime, they could be perceived as marginal by the transmission service users.

part of the revenue requirements, the percentage varying depending on the system (see Table). A supplementary mechanism for cost recovery is necessary. For example, in South America, a two part tariff is used with the second part, in the form of transmission tolls, being added to the marginal cost income to fully finance the system. These tolls, or wheeling rates, are charged to users of the system, definition of users and allocation of charges varying from country to country.

Table. **Cost recovery under SRMC pricing**

| Country | Pricing Concept | Cost Recovery |
|---|---|---|
| Australia | Long-Run Average Cost | 50% (predefined, no explicit calculation of LRMC) |
| New Zealand | SRMC | 10% |
| Norway | SRMC | 17.2% |
| Chile | SRMC | 10% |
| U.S. Estimate | SRMC | 5% –20% |
| Bolivia | SRMC | 3.6% |

***Source:*** *Read, Transmission Pricing in New Zealand, 1997; Glende & Westre, Transmission Pricing in Norway; Rudnick, Presentation: Latin American Experience in the Restructuring of Electric Power; 1998. Federal Energy Regulatory Commission, The Recovery of Fixed Transmission Costs, 7 December 1997, NARUC-DOE National Electricity Forum; Powerlink; OFGEM and NGC Reports; ACCC and IPART Reports.*

*Normative Solutions*

The ideal cost recovery regime would involve a “perfect tax” which, in order to avoid distorting incentives in the electricity sector, should be levied on some basis other than electricity consumption. Thus, it could be argued, the efficiency of the electricity sector could be enhanced if the government were simply to write off the value of existing assets, so that pure marginal costs could prevail. This is obviously unrealistic, but the argument is also flawed, in two respects. First, this solution could only be implemented by raising general taxation, which will increase distortions elsewhere in the economy. Second, the precedent could have unfortunate repercussions with regard to future expectations of cost recovery, and price levels, in the electricity sector.

Economic theory suggests that normally only exactly one system of relative prices between all goods leads to maximum efficiency and it occurs when prices are equal to marginal cost. This rule will be labelled “allocatively efficient pricing” and provides the best possible outcome that maximises aggregate consumer welfare. However, in the case of decreasing long-run average cost function (typical case in the electricity transmission networks), with prices equal to marginal costs, revenues fall short of total costs by an amount equal to the fixed costs. Therefore if the regulated service provider is required to be self-financing, it is necessary to choose a different set of higher prices. The Ramsey prices[23] are the “second best” prices that succeed in raising enough revenue to cover total cost with the smallest possible sacrifice in consumer welfare. The general result of Ramsey pricing rules is that the departure from the marginal cost pricing (best solution) should be inversely proportional to the price elasticity of demand of the relevant product. The principal idea is that the demand structure should not be heavily distorted by the allocation of the remaining cost.[24] Ramsey prices are applied in a limited manner as the price elasticity of demand is not known.

*Locational Versus Non-Locational Cost Recovery*

On one hand, it could be argued that sunk cost recovery should be designed so as not to send any locational signals since there is no economic point in sending locational signals, based on historical costs, to parties who have already located. The signals they require are already conveyed by the locational spot prices and by the associated new investment regime. On the other hand, the application of locational pricing for the existing transmission is frequently justified by the need to establish long-term signals. This argument is also supported by the fact that such locational pricing models aim either to calculate explicitly or approximately the level of LRMC of transmission through partial allocation of sunk cost on a locational basis.

Determining the appropriate balance of these charges should be driven by the following considerations:[25]

! longer-run cost signalling objectives;

! cost attribution objectives, but these require an investigation of the relationship between LRMC and average network costs and may lead to different balances across regions;

! the extent to which greater reliance on locational charging is likely to lead to network bypasses; and

! equity (or cost averaging) objectives across the network resulting from social goals or other political constraints.

[23] See: Ramsey, F. (1927): A Contribution to the Theory of Taxation, in: Economic Journal, Vol. 37.

[24] Additional results on Ramsey pricing may be found in: Baumol, W. J. / D. Bradford (1970): Optimal Departure from Marginal Cost Pricing, in: American Economic Review, Vol. 60, No.3.

25 In Australia, the National Electricity Code determines a 50/50 split between locational (cost reflective network pricing) and postage stamped charges. This was initially determined as a pragmatic approach for estimating LRMC across the shared network. Later, it was discovered that this approximation may hold for South Australia, Victoria and New South Wales but is more likely an underestimate for Queensland where projected network investments were high.

*Non-Locational Post Stamps*

This method is a further simplification of zonal pricing where the zones are defined solely on the basis of voltage level without using further criteria for locational differentiations resulting from load flow and allocation of generation and demand. According to this method, the revenue requirements are allocated to each voltage level (but not to node) using a cost cascading approach; i.e., the cost of a certain voltage level is allocated to the cost of the lower voltage level and the consumers connected to this voltage level. The transmission tariffs are differentiated by voltage level (additionally could include differentiation for types of customers and time of use).

To ensure that costs are properly allocated, the costs should be classified into categories that reflect the main factors determining the level of overall costs of providing the relevant service. The cost allocation procedure should provide an estimate of the cost associated with the delivery of the products provided by the regulated service provider and should establish a basis for pricing these products. Generally, the cost allocation process includes the following steps:

! definition of cost categories;

! definition of cost pools (cost centres);

! allocation of costs of each category to cost centres;

! definition of products delivered by the regulated service provider; and

! allocation of cost attributed to each cost centre to the relevant products.

This is also the process followed by the Order N 72-э/3 where each transmission region is priced separately and voltage levels are defined as separate cost pools (cost centres). Once allocated to the voltage levels, costs of each regional transmission network are cascaded downward in the system in order to derive the network cost to be allocated to customers connected to the respective voltage level.[26]

*Locational Methods*

Use of Distance Capacity Related Approach

For any off-take or injection, the resulting MW·km are calculated. For off-takes it is assumed that the generation has been injected at the reference node and for injections vice versa. The MW·km of each injection and each load over all nodes are summed up to calculate an aggregate MW·km utilisation of the network. The revenue requirements are divided through the aggregate MW·km in order to obtain a specific MW·km price. The MW·km of all injections and off-takes are valued with this specific MW·km price and then divided through the relevant usage quantity (registered capacity or peak load) in order to obtain a capacity related price for generation and demand at all nodes.

Load Flow Based Participation Factors

The algorithm of average participations is simple and robust but it uses a very simplified representation of the network (only the first Kirchhoff's law). However, this simplification may be justified, since the criteria to allocate the complementary charge are also fairly loose and mostly of a qualitative nature.[27] The method deals separately with injections to and withdrawals from the network. First, the responsibility of the historical flow in each line and at a given time is fully allocated to the injections at the nodes. Next, the flows are completely allocated again to the withdrawals from the network. In this way, there is the option of assigning different global weights to injections and withdrawals when computing the final coefficients of allocation of flows to generators and loads. Differently from the average participation factors, the incremental participation factors are defined by means of a sensitivity analysis that relates a variation in the injected/ejected power in a busbar to a variation in the flow through a certain transmission line. It considers that a change in the injected/ejected power in any bus is absorbed by a similar negative change in the reference bus, while injections from generators and loads in other buses, as well as losses, are maintained constant.[28]

---

[26] Simple aggregation of connected customer load for the purposes of cost allocation, however, would fail to reflect the actual use of transmission network and the coincidence of time occurrence of peak demand at each voltage level (assuming that cost should be allocated in proportion to the contribution to the peak demand at each network voltage level). The time of occurrence of peak demand at a particular transmission network voltage level could be different from the time of occurrence of peak demand at an upper transmission network voltage level. Consequently, the contribution of this particular network voltage level to the peak demand at the upper network voltage level will be less than its individual peak demand. Therefore, there is a need for coincident metering.

Cost allocation according to energy consumption might be more easily implemented as the energy measurements exist and the energy consumption on the different network voltage levels could be easily aggregated. It could be argued that demand is the relevant criteria for network planning and the customers with low utilisation of their connected load will pay lower amounts for the use of transmission network in comparison to those that they should have paid in case of demand allocation criteria. It is correct; however, that the economic benefits of proper use of "causality principles" will depend on the availability and reliability of the demand data. It is doubtful whether demand criteria would lead to better allocation when the input information is not available or distorted.

[27] Robustness, i.e., little volatility with respect to input data or the absence of arbitrary decisions such as the choice of a slack bus that influence the results in other methods, is a very desirable characteristic.

[28] Such have been applied in South America (and also in New Zealand). The New Zealand approach determines grid "usage" by tracing flows, as does the "area of influence" regime employed in a number of South American countries, where it is also common to charge generators rather than consumers for transmission assets. In Chile, the area of influence is defined as the set of assets upon which flows are deemed to be directly and necessarily affected by the connection of a generator to the grid. Generators pay for the lines to which they contribute a positive flow in proportion to their usage of the capacity of those lines. They are then free to use those lines, but a separate toll, similar to a wheeling fee, is charged for trading outside the area of influence.

Use of Fault Level Contribution Matrix Approach

Similar to the method of incremental participation factors, this method is also based on a sensitivity analysis. It uses the 'electrical distance' represented through the impedance of the circuits in order to pair generation and load. A way to pair generation and loads using the "electrical distance" is the application of Fault Level Contribution Matrix. By faulting each bus, the fault current supplied by each generating unit is determined depending on the impedance of the circuits. After pairing the generation and load, a full system load matrix is constructed, showing how the generation nodes are allocated to the load nodes. The matrix is multiplied with a sensitivity matrix that gives the information on how the circuit loading changes in case of changes in different network nodes. The figures obtained are used to allocate the circuit cost to the various nodes.

***Cost Allocation under Time Dependence Aspects***

Primary responsibility for the total network capacity required lies with those consumers who make demands for energy at times that coincide with the demands made by many other consumers. The degree of coincidence dictates the level of the peak demand on the network and the sharpness, or otherwise, of the peak obtained when demand is plotted against time. Because the time at which the peak demand occurs may change from day to day and month to month, it could be argued that it is better to allocate network costs over a number of hours rather than to a single hour. This is particularly the case if the actual peak observed tends to be broad rather than narrow in time. Further, the cost allocation may not be proportional to the energy delivered during these periods because the probability of not meeting the load in the demand peak periods is greater than in the demand off-peak periods. Therefore, the proper economic pricing signals would be ensured when larger portions of the revenue requirements are allocated to the hours of peak demand.

The time relationship of the network costs is not as immediately obvious as the variation with time of the cost (prices) of the generation. However, the overall capacity and, hence, capital and maintenance cost of the networks is dictated by the peak loading to which they are designed. This, in turn, depends on the coincidence in time of the demands made by individual consumers. When this is taken into account, responsibility for the costs of these networks can be allocated to particular time periods. Development of tariff structures includes reaching a decision on the periods at which time of use tariffs will be made available to the network users – viz time of day and season of year. It is also necessary to define which hours of the day are to be included in each period and if the periods should change in time or duration from season to season.[29]

The time of use tariffs vary with the time of delivery and allocate transmission network cost (revenue requirements) to the different periods of time. The allocation rules could use the probability of not meeting the maximum demand in the relevant voltage level or simple criteria based on the load profiles on each voltage level. In a simple case, it could be found that there is not a reasonable expectation of demand exceeding the available network capacity during any off-peak period (e.g., night, weekend, etc.) However, as to the remaining hours the probability of not keeping the demand supply balance is higher, hence a greater portion of the network cost could be allocated to those hours.[30] The time of use tariff would require time-dependent metering to register the energy or demand served in the tariff time periods.

***Payment Liability***

Secure transmission network benefits both the load serving entities/direct retail customers and the power plants. Without a reliable transmission grid, the generation units (connected to the transmission grid) are not able to feed in electricity into the transmission grid. On the other hand, reliability is also important for load-serving entities and directly connected customers. A measurement for the value of reliability for load customers is the value of lost load. It could be argued that the foregone revenues of generators not allowed to generate electricity due to an unreliable transmission network are lower in relation to the value of lost load. Another argument to allocate a higher portion of transmission service payments to load could be justified by the fact that load is less elastic and pricing under the aspects of allocative efficiency should consider allocation according to the inverse price elasticity of demand for transmission service.

Also on the balancing market (superimposed on the day-ahead market) where the generators might be allowed to react using incremental/decremental bids when the transmission system precludes them from generating. This supports the proposi-

[29] In the case that a pure demand charge is applied, incentives for the customers would appear to be to avoid paying charges for transmission service by limiting his usage to only off-peak hours.

[30] In the practice, the so-called "threshold approach" is used. The costs are allocated to the time periods where demand exceeds a pre-determined threshold level. The threshold is set as a percentage of the maximum system demand.

tion that the generators benefit from availability of a secure transmission network. It would be inappropriate for generators to be recompensed in such a way, were the transmission system not planned and operated to defined security standards and were generators not to bear part of the burden of having a secure transmission network. On the other side, imposing payments on the generators for full compensation of revenue requirements on a locational basis would not bring additional incentives in term of locational efficiency, as part of transmission network costs (assets and non-assets related costs) is related to a function that might not exhibit a locational nature and provides equivalent benefits to all transmission service users without any differentiation on their location.[31] From this point of view, advantages of joint payment liabilities might be justified with the fact that the forwarding of costs incurred by the generators for payments of those costs of the transmission network would reflect the individual bidding strategies of the generators in the wholesale market and accordingly their willingness to reduce some of the expected production surpluses. The effect might be limited as the proportional payments would be more or less passed through in the bids of the generators. However, recovery of sunken transmission cost through the energy market prices will certainly have an adverse affect on the efficiency signals delivered by the wholesale market prices.

It is proposed that both generation and load are liable for transmission service, the prevailing part to be charged to the load nodes (e.g., 25% to generators and 75% to the load). However, this should only be applied if the FEC and/or FNC deem it appropriate to introduce locational pricing for the transmission network infrastructure. If this is not the case and the postage stamp approach is used, it is more meaningful to assign the costs entirely to the load entities.

***Transmission Tariff Design***

The Order N 72-э/3 details the transmission use of network charges in RUB/kW. Economic theory suggests that charges would be better based on peak usage, rather than energy.[32] Cost allocation according to the peak demand seems to reflect the nature of the transmission network costs (most of them are fixed) as the transmission network is designed to meet the system peak.[33] Also from an economic point of view capacity has no value unless it is fully used. Thus, at off-peak times, no one who would be prepared to pay for more capacity since none is required, while at peak-times capacity is more likely to be fully utilised, and users are likely to be prepared to ration capacity. However, using only one peak period makes the whole payment liability dependent on one single value and exposes the service users to rather stochastic characteristic of demand. Moreover, a user may substitute other forms of supply (e.g., via embedded generation) to lower peak usage and hence charges. This may be inefficient, though, as the user is incurring costs and using potentially expensive resources instead of using an asset for which the costs are the sunk.[34] In order to avoid these negative effects additional system peak loads are considered in the transmission price design (e.g., load triad used by NGC in UK).

Using only a demand charge (even when using more than one peak period to determine the payment liabilities) could create a substantial difference in the payment of transmission service users having different profiles. The effective payments of customers with higher load factors will be much lower than the customers with lower load factors. Therefore for equity reasons, the tariffs could consider the introduction of energy-dependent charges and fixed charges[35] and the related non-demand criteria in the process of tariff design development. Variable energy-based charges achieve these equity objectives and could result in limited distortions in the energy market, since relative energy costs in different demand periods are not distorted. These charges might be more appropriate than peak-based charges, where there is an excess network capacity.

[31] Part of the costs falling under this category could be identified explicitly, e.g., the costs of the communication lines, network planning, general overhead, etc. On the other hand, part of the costs is included in the network infrastructure in form of additional investments necessary to meet contingency standards. In the Australian pricing practice, the costs that could be individually identified are separated and form an explicit charge called Transmission Common Service Charge. In the UK pricing practice (as already described), the reverse approach is used; i.e., the transport related costs are explicitly calculated by means of modelling (ICRP approach).

[32] From an economic perspective, the actual allocation of sunk costs is largely an arbitrary decision, provided that allocation is done in such a way as to avoid any impact on future decision-making. In particular, this implies that any charge levied to recover such costs should be "fixed" in the sense that it should not depend on any present, or future, actions of the parties involved. If these charges vary in any way with current or future "usage," then grid users may change their usage, or build new assets – such as embedded generation – to avoid the charge, even though this makes no difference to the costs, which must be recovered. These users would incur additional costs in order to gain a net advantage for themselves, but leave others to cover the cost of the assets already committed, thus causing an overall loss to the sector as a whole.

[33] Network asset costs are related to peak capacity, and if costs must be allocated in some way, an allocation on the basis of contribution to peak demand may be appropriate. The exact form of a peak charge can vary: (1) the peak can be defined as the single highest period of demand or supply over a number of years, within each year, or the average demand/supply over a small number of period within the year. An energy charge is then simply the peak charge averaged over all periods; and (2) a peak can be defined as the contribution to system peak demand or regional peak demand, the customer's own peak or the peak for each supply point servicing a customer.

[34] If peak usage does not change significantly over time, then a peak usage charging regime corresponds closely to the ideal of a fully fixed transmission charge.

[35] Fixed charges – appropriately applied – are less distortional than variable charges, are simple to apply, could be differentiated by size of customer, and would reduce the taxation impact on consumption.

However, when the charges are based on energy, then the user faces a network cost that depends upon its throughput of the network, even though throughput may have no impact on the level of costs to be recovered. Consequently, the user sees a higher cost per unit of delivered energy than it would if charges were truly fixed. As a result, the user may have an incentive to reduce energy consumption below the level implied by SRMC alone. This would be inefficient, as reducing the usage of the transmission assets produces no saving in the cost of the asset.

We suggest two tariff design options: (1) peak-based regime, where the peak usage is appropriately defined so as to minimise the attractiveness of inefficient charge avoidance; and (2) combined approach using energy, demand and fixed charge. The formulas below are designed for joint payment liability, i.e., load and generators pay for transmission service but could be simply rewritten only for load.

***Option 1. Demand-Based Charges***

The charges reflecting the residual costs will be calculated as:

$$UNC_g = \frac{CA_g}{RC_g} \text{ (RUB/kW/year)},$$

$$UNC_l = \frac{CA_l}{AL_l} \text{ (RUB/kW/year)},$$

where:

- $CA_g$ – revenue requirements to be recovered through transmission use of network charge from generators;
- $CA_l$ – revenue requirements to be recovered through transmission use of network charge from load;
- $UNC_g$ – transmission use of network charge for generation;
- $UNC_l$ – transmission use of network charge for load;
- $RC_g$ – total rated generation capacity;
- $AL_l$ – system peak load or average of a number of system peak loads (for pre-defined assessment period);[36]

[36] The assessment period is defined as the settlement period in which peak load occurred and a given number of settlement periods either side of the settlement period in which the peak load occurred. This number could vary; e.g., in the UK, the system peak demand is supplemented by two additional system peak demands (load triad) with at least ten days time difference to avoid peak demand sequences. In New South Wales, IPART proposed in March 1996 that demand charge (25% of the network costs are covered by the demand charge) will apply to 10 winter daily demand days and 10 scaled summer daily demands. The summer demands are scaled up by the ratio of winter/summer demand for the system. The rationale for this price structure is that the need for augmentation of TransGrid network may be attributed roughly equally to summer and winter limitation.

[37] The distribution ratio shown in the example should be considered as an example and not as a prescriptive coefficient.

***Option 2. Multi-Component Tariff Design***

The total costs allocated to nodes comprise three component prices:[37]

$CA_g \rightarrow CA_g^{(\text{fixed})}$ (e.g., 25%); $CA_g^{(\text{demand})}$ (e.g., 50%); $CA_g^{(\text{energy})}$ (e.g., 25%)

$CA_l \rightarrow CA_l^{(\text{fixed})}$ (e.g., 25%); $CA_l^{(\text{demand})}$ (e.g., 50%); $CA_l^{(\text{energy})}$ (e.g., 25%)

The demand, energy and fixed transmission use of network charges are derived as follows:

*Generation Nodes*

$$UNC_g^{(\text{demand})} = \frac{CA_g^{(\text{demand})}}{RC_g} \text{ (RUB/kW/year)}$$

$$UNC_g^{(\text{energy})} = \frac{CA_g^{(\text{energy})}}{E_g} \text{ (RUB/kWh/year)}$$

$$UNC_g^{(\text{fixed})} = CA_g^{(\text{fixed})} \text{ (RUB/year)}$$

*Load Nodes*

$$UNC_l^{(\text{demand})} = \frac{CA_l^{(\text{demand})}}{RC_l} \text{ (RUB/kW/year)}$$

$$UNC_l^{(\text{energy})} = \frac{CA_l^{(\text{energy})}}{E_l} \text{ (RUB/kWh/year)}$$

$$UNC_l^{(\text{fixed})} = CA_l^{(\text{fixed})} \text{ (RUB/year)}$$

**Connection Charges**

Separation of transmission network service from transmission connection service follows the economic causality principles, namely to charge costs to the party that has caused these costs. Accordingly, the transmission pricing should distinguish between transmission use of network charges and transmission connection charges. The transmission connection charge should cover only the cost of the transmission connection assets and should be charged directly to the respective connected party (generators and load entities, called transmission connection service users). On the other hand the transmission use of network charge should recover the costs (including depreciation, O&M costs and return on the transmission network assets[38]) necessary for secure transport of power.

Connection charges are levied when a user first connects to the system or requests a substantial change to its capacity connection. This applies to both generation and demand, although the treatment of these two types of connections may vary. Connection charges should cover the cost of providing the assets necessary to make the connection, including a reasonable rate of return on those investments, plus, possibly, ongoing maintenance costs. In general, these charges should be designed to recover in full from the user, often as an up-front, one-off, charge, the relevant costs of making the connection.

In deciding on the relevant connection charges that should properly be charged to the connected party, a fundamental problem is related to the split of assets between the "connection" and the "core transmission network." Generally, two extreme positions can be considered. In the first, known as shallow charging, the connected party is required to fund only the assets specifically required to connect it to the system and for the specific benefits of this particular user. The second, known as deep charging, levies all the costs consequential on making the connection on the connected party, including the costs of incremental investment in the wider system, whether or not these relate to the network local to the connection point. The advantages and disadvantages of these approaches are considered below.

### *Shallow Connection Charges*

Under this approach, the connected parties pay for only those assets located in the immediate vicinity of their point of connection. These are the assets required to make a physical connection from the existing system to the connected party. Any costs associated with the network augmentation are regarded as resulting from general demand or generation growth and are recovered from all users through transmission charges.

There is inevitably scope for defining shallow connection assets more or less narrowly. For example:

- a "very shallow" approach would exclude from connection assets all spurs between the generator or customer and the main system. It would therefore include all substation assets (mainly, transformers, switchgear and busbars) and exclude all overhead lines or cables, other than those which constitute substation assets;
- a "less shallow" approach might include all single and multiple spurs that serve to connect the generator or customer or groups of generators and customers to the core transmission network, in addition to substation assets;
- an intermediate position might be one in which "generation only spurs" are included in the definition of connection assets, to give potential generators adequate locational signals. Spurs connecting customer exit points to the main system would be classified as "core transmission network" assets, on the grounds that the case for locational signals is stronger for generators than for customers.

Shallow connection charges tend to minimise the charges required to connect to the system, thus encouraging the development of demand and generation connections. Although this requires a precise definition of assets belonging to the "connection" as opposed to the "core transmission network," it is possible to create a set of clear and explicit rules to cover most situations.[39] Consequently, shallow charging makes it easier for users to estimate connection themselves and thus tends to be more transparent.

The usual critic against shallow connection is that connected parties receive limited locational signals about the full incremental cost (not only the cost of direct connection assets but also the costs of augmentation of the core transmission network) resul ting from their connection. Charging the costs of augmentation of the core transmission network via the general transmission tariffs may encourage inefficient entry of new users as the cost caused by their connection will be recovered by the existing transmission service users. Finally, shallow connection charges socialise the risk of stranded network assets that may occur if a transmission service user leaves the system.[40]

[38] Transmission network assets include only core transmission network assets and exclude transmission connection assets.

[39] Generally, the transmission connection assets on the generation side (entry transmission connection assets) will include: transmission switchgear plant (circuit breakers and isolators) and associated plant; station establishment and building cost; land. The transmission connection assets on demand side (exit transmission connection assets) will include transmission switchgear plant (circuit breakers and isolators) and associated plant; transformers that supply the subtransmission voltage level and associated switchgear; station and establishment and building cost; land. Taking this into account, the transmission connection charges should cover the connection cost required to provide physical access to the transmission network. This, in turn, consists of the following services: (1) the design, project management and provision of connection assets (including purchase and installation); (2) repair, maintenance of connection assets to ensure that they remain fit for use throughout their replacement period, including the replacement of failed assets during the replacement period; and (3) the removal of the connection assets either at the request of the user or at such time as the connection becomes unsafe.

[40] This could occur if costs resulting from certain connection had not been recovered and become stranded after the user's decision to leave the system. However, the same is also true for deep connection charging, unless all investment costs for the deep connection are immediately covered by an up-front payment.

***Deep Connection Charges***

In this approach, the connected parties pay for full incremental costs incurred by the transmission network service provider as a result of their connection, including investments at places remote from the connection site. Examples of this are where the system needs reinforcement at a higher voltage level than that of the connection as a direct result of the connection, or where equipment has to be upgraded, perhaps to increase transmission capacity between different parts of the network, or to cater for increased fault current levels. In other words, the connected party is required to pay for all the transmission assets which would not be required if the particular user did not exist, including the cost of reinforcements at remote sites.

Therefore, deep connection charges tend to be higher than under a shallow connection policy. On the contrary, deep charging tends to reduce the level of transmission charges for all users (excluding the connected party). The major advantage of deep connection charges is that they provide strong locational signals for new connections. This helps to improve allocative efficiency, although many customers, especially domestic consumers, will generally be unresponsive to locational signals. Nevertheless, it is fair to say that deep charging inhibits the same positive properties as locational transmission charges.

The main drawback of a deep connection charging policy is that these charges are difficult and arbitrary to apply in practice. While it is relatively easy to define the required assets in the immediate vicinity of the connection, this is not the case for reinforcements throughout the system. Most importantly, this relates to problems caused by the significant and "lumpy" nature of reinforcement that can be driven by a small increase in connection capacity. For example, if a new connection requires network reinforcement and the capacity upgrade, because of lumpiness, exceeds the new load, significant redundant capacity will occur, which is not immediately required by new or existing users. In consequence, the resulting costs may be much higher than would be required to meet the new connected load.

In the sense that some remote reinforcements can be argued to be of benefit to a great number of users, deep connection charges may not be cost-reflective. In order to avoid discrimination, this may require an allocation of these costs, which will be arbitrary by nature, to the connected party and the existing users. Deep connection charges are therefore neither transparent nor predictable for users.

Deep connection charging may discourage new connections because of the higher level of initial charges that connected party will need to pay. Deep connection charging requires an explicit redistribution mechanism among the connected parties. If the first user is charged the full deep connection costs the spare capacity gain will bring benefits to the later connected parties. These later connected parties should reimburse part of the full deep connection costs to the first user.

Similarly, if applied for the first time, deep connection charging will probably be discriminatory between existing network service users and new entrants. While it could consistently be applied to all new connections, this would be impossible for existing users, given the historic nature of the system. In other words, it would be impossible to determine for each existing connection what remote reinforcement costs were necessary in the past to accommodate those connections and to derive appropriate charges in that case. □