A PLAUSIBLE CONGESTION MANAGEMENT SCHEME FOR THE INTERNAL ELECTRICITY MARKET OF THE EUROPEAN UNION

Working Paper IIT, Ref. Number: IIT-03-037A

(version 3.0, March 2004)[1]

Ignacio J. Pérez-Arriaga Luis Olmos

Instituto de Investigación Tecnológica (IIT)

Universidad Pontificia Comillas, Madrid, Spain

ignacio@iit.upco.es luis.olmos@iit.upco.es

http://www.iit.upco.es

Abstract – This paper proposes a scheme for the management of network congestion in the Internal Electricity Market (IEM) of the European Union. This scheme tries to combine the rigor in the treatment of the energy and transmission capacity transactions with the flexibility and pragmatism that are necessary to make the scheme compatible with the current diversity of regulatory approaches and market structures in the Member States participating in the IEM. First a reference scheme is presented with a complete formulation that jointly deals with the energy and capacity markets. Because of the implementation difficulties of this conceptually ideal approach a more pragmatic scheme is proposed instead. The core of this scheme is an explicit auction mechanism that must be run prior to any short-term (daily, typically) energy markets. In this auction, where only transmission capacity is traded, both bilateral contracts and energy bids to Power Exchanges can participate in order to acquire the capacity that is necessary to carry out their transactions. Some technical issues related to the practical implementation of the proposed approach are also examined; these include market liquidity, the financial or physical nature of the long-term contracts, the potential problems of "slicing" transmission capacity and the allocation of congestion rents. Market power issues are ignored.

Keywords: congestion management, transmission network constraints, coordinated auction, network utilization, power transfer distribution factors.

1 INTRODUCTION

Wholesale energy markets need transmission networks to materialize. Not all patterns of commercial transactions can be allowed, since many of them result in unacceptable operating conditions because of a diversity of security reasons. The most common of these unacceptable conditions happens when the resulting flow in a line exceeds the maximum value that the system operator considers to be secure. This limitation may be due to the thermal capacity of the line or perhaps because a higher value could result in some unstable situation for the power system as a

[1] First released in April 2003.

whole. When the flow in the line reaches its limit the line is said to be “congested”. Transmission networks may impose other types of limitations, such as upper or lower voltage limits, which typically require that some generators that are located at specific points in the network have to operate. Maintaining the power system in stable conditions may also exclude particular operating conditions. Although improperly, the term “congestion management” is used to describe the ensemble of actions that are needed to meet the security criteria imposed by the system operator. “Network constraint management”, or even better, “System constraint management” –since both generation and network are involved in most constraints from a system perspective- is a more adequate expression.

Congestion management in transmission networks is one of the areas where the power system academic community has been more prolific recently, see [1-10] as a sample of recent relevant literature on the subject. Despite this, the subject cannot yet be considered closed, particularly in the context of large multi-national markets, such as the Internal Electricity Market (IEM) of the European Union. In this case, any plausible approach must comply with sound engineering and economic requirements. On the other hand, it must also comply with the diversity of regulatory regimes and market structures in the IEM.

The contribution of this paper lies in the specification and critical examination of realistic alternatives of system constraint management that are presently available for implementation in the Internal Electricity Market (IEM) of the European Union (EU). The objective is not to propose the fanciest and more accurate model, since this would simply not be feasible in the present European context. Therefore, the considered set of alternatives does not include a full fledged centralized nodal pricing scheme encompassing the entire EU, because this is very far from being acceptable in the current European regulatory and political contexts. Therefore, the paper has to explore less complete but more viable approaches, while keeping full nodal pricing [11] as an ideal reference. This pragmatic approach resembles efforts to find an acceptable scheme for a Standard Market Design in the USA, see [12]. One has to keep in mind that a Regional Transmission Organization (RTO) in the USA may be of the same electrical size as an average European country. The regulatory diversity is very large among the IEM countries. The possibility of a more harmonized and compact regulatory scheme of the transmission activity is stronger in smaller multinational markets, such as the Central American Electricity market, see [13].

1.1 TERMINOLOGY AND APPROACH CLASSIFICATION

Congestion management methods can be classified with respect to the interface between the energy and capacity markets. *Implicit auctions* is a general term that applies to any congestion management method that is able to deal at the same time with the energy market and any market allocation mechanism of transmission capacity. Thus, the effect of network constraints and losses is internalized in the prices of energy. Nodal pricing, see [11, 14], also termed Locational Marginal Pricing (LMP), is the ideal implementation of this principle, with nodal differentiation of energy prices, which happen to be perfect short-term economic signals. Nodal pricing may be seen as the outcome of a joint competitive auction of energy transactions and transmission capacity rights.

A simplification of nodal pricing, -only legitimate when the pattern of network utilization meets certain conditions-, is *zonal pricing*, also termed *market splitting*. When serious congestions only happen between areas that are well meshed internally and losses are ignored it can be assumed that these areas are “single price areas” (SPAs) regarding congestion management [7, 8, 15]. In this case, the simplification may make sense. However, a second requirement must also be met. In order for the division in areas to be acceptable, a balanced transaction within a SPA should not significantly affect the flow over the inter-area links. In other words, any bilateral transaction within a SPA should not create loop flows outside this area which could contribute to congestions inter single price areas in any significant way. The selection of single price areas, whenever applicable, is not a trivial matter without practical consequences. The choice may imply a hidden priority in the scheduling of intra single price area transactions (as they would not be subject to coordinated congestion management procedures) and it will affect the validity of the results of congestion management.

The geographical pattern of single price areas, -when they exist-, may not respect the frontiers between control areas. Each control area is normally under the jurisdiction of a different Transmission System Operator (TSO) and its footprint often coincides with the political borders of a Member State in the IEM. For the sake of simplicity, we assume in this paper that each *Member State* has one or more *control areas*, each one under a different TSO. Each control area may comprise several *single price areas*, which are defined just for the purpose of congestion management. This can always be achieved, by conveniently subdividing any single price area, if necessary.

In power systems where network congestions are not serious or frequent, it may be acceptable to run the energy markets as if the transmission network had no limits. Once the outcome of the energy market is known, it can be checked if the pattern of commercial transactions violates any network constraint. Only when it does, the system operator needs to *redispatch* the transactions. That is to say, some modifications to the results of the market are necessary, -in principle those that incur in minimum additional cost-, so that any violation of network constraints is eliminated, see [16]. Redispatch can be accomplished with market-based mechanisms if the modifications are decided on the basis of bids sent by market agents indicating how much they ask for in order to change their market positions. A specific implementation of this method of *redispatch* is *counter-trading*, whereby the system operator nominates pairs of generators that modify their outputs so that a power flow is created that counters the excess flow that is causing the congestion. Obviously, one could generalize and say that any redispatch can be seen as a counter-trade, since any increase in the output of a generator has to be matched by a corresponding and identical (except for losses) reduction in the output of another generator. Typically, the extra costs of redispatch or counter-trade are socialized to all consumers, and any economic signals derived from congestion, -which could try to emulate nodal pricing-, are lost. Conceptually speaking, assigning the costs of redispatch to those market agents found "responsible" for the network constraint is possible. In this case, the economic signals would not be completely lost. This is a technically complex task nevertheless, see [17]. The results may not be conclusive and it has to be evaluated in each case if it is worth the effort.

The third approach assigns energy and transmission capacity in inverse order of redispatch, namely, network capacity first and energy afterwards. Prior to running the energy market, an *explicit auction* of network capacity takes place. Every transaction that might be using the transmission capacity likely to become congested has to participate in the auction and the right to use the transmission network is allocated to those agents who value it most. Once the transmission capacity rights have been assigned, the energy market can take place, but only those transactions with capacity rights across congested transmission, -or those who do not affect transmission constraints significantly-, can participate. Obviously, this method requires of some centralized coordination, see [2]. If they were not considered jointly, the complex flow patterns due to the multiple transactions might result in unexpected violations of network constraints.

This paper will not pay attention to those network constraints that only have a very local effect and whose solution is also clearly local, with almost no possibility of application of any market-based mechanism of resolution. This is the case of local voltage problems, whose solution typically involves starting up a particular generator or one among a very small number of them (often belonging to the same company).When dealing with this type of constraints, it is advisable to remunerate this service with a previously agreed administrative payment. These network constraints are customarily solved by the corresponding TSO, without having to recur to any kind of inter-TSO coordination. In this paper it will be taken for granted that these local actions have been applied and local network problems have already been solved.

Losses have been ignored in all the congestion management schemes to be presented below. They could have been included but this is not the issue of concern in this paper. Besides, it is not an issue that has drawn much attention in the IEM so far. How could one take losses into account? As mentioned before, nodal prices are the ideal reference model for short-term signals. Therefore, the combination of all the economic signals in system operation (energy, congestion and losses) should be equivalent to the use of full nodal prices. Then every agent should pay, per MWh of energy traded, for the marginal losses caused by a unitary transaction between the node where the agent is located and the reference node whose nodal or zonal price represents how much the agent charges or pays for his energy according to the adopted congestion management scheme. The consistency in the choice of reference node between the computation of marginal losses and the congestion management scheme is necessary for this decomposition of nodal prices into energy, losses and congestion to be valid, see [17].

This paper discusses different ways of implementing a congestion management scheme in the IEM. All the possibilities consider the existence of both bilateral bids for the capacity that is needed to execute a contract and unilateral bids to buy energy from or to sell energy to a power exchange. Each market agent has the possibility of choosing the type of transactions that better fits its needs. Thus, the resulting market is more flexible than those that only allow for one type of transactions or the other [18].

The paper ends up recommending the adoption of coordinated explicit auctions under the present regulatory conditions in the IEM. Section 2 presents the mathematical formulation of the implicit auction scheme, and it is shown that, although theoretically desirable, this approach does not appear to be presently viable for the IEM. Taking the implicit auction formulation as a reference, section 3 proposes a coordinated explicit auction of transmission capacity as a plausible approach to congestion management in the IEM. The paper examines which simplifications appear to be acceptable in the conceptually complete formulation so that a pragmatic implementation is possible. Section 4 analyzes some technical aspects that need to be addressed in the proposed implementation and section 5 presents the conclusions.

2 THE REFERENCE FORMULATION

A *coordinated implicit auction* of transmission capacity that is fully integrated with the energy short-term market of a power exchange appears to be the best solution from an efficiency point of view and from the point of view of market power (see [19] among others). We shall refer later to its practical implementation difficulties, at least in the IEM context.

The central part of this scheme is an IEM-wide auction that would take place one day ahead of real-time operation, -the usual time frame for short-term energy markets-, at a centralized auction house. In this auction both capacity and energy bids would be accepted for any given hour:

- Bids in €/kW from individual agents to buy or sell energy, without specifying on which power exchange they want to trade their energy. These are “unbalanced” or “one-sided” bids.
- Bids in €/kW to acquire the transmission right to buy energy from a certain geographical zone and sell it to another specific zone during the considered hour. These are “balanced” or “bilateral” bids.

In principle, all transactions in the IEM would have to participate in the auction, as all of them have the potential to create network constraint problems. However, if the single price area (SPA) assumption is acceptable, only inter-SPA transactions need to be accounted for in the formulation of the model. We assume that the corresponding TSOs will take care of redispatching in order to remove any congestion internal to the SPAs. They will also be in charge of solving any minor inter-SPA congestion that might happen because of the assumption that intra-SPA transactions have no effect outside the SPAs. Using an accurate model of the IEM makes the computation of the impact of each possible CBT transaction on the transmission network flows possible. The accuracy of the results would improve if the TSOs could provide an estimation of the patterns of generation and demand within each SPA[2].

An optimization algorithm determines the set of bids that maximizes global welfare and is compatible with the network constraints. Only these bids are accepted whereas the remaining ones are discarded. This certainly constitutes a level playing field for the two types of bids.

[2] One argument that has been frequently employed, in the IEM context, is that the use of single price areas is preferred to the use of all network nodes because of confidentiality reasons. In this way individual transactions cannot be easily identified when collecting the data that are needed for congestion management.

In addition to this short-term energy and capacity auction, long and medium term capacity auctions could also take place. These would be auctions where agents involved in bilateral contracts as well as those planning to trade their energy on a certain power exchange could obtain the transmission rights they need to secure the price they will pay to access the grid. Either the total network capacity or a fraction of it (a "slice") could be offered in these auctions. An accurate representation of the IEM network is necessary also in this case. All market agents, as well as speculators, could bid at these longer-term auctions, whose prices should therefore be guided by the expected future real time prices of energy and network capacity. The same assumptions regarding the SPAs are applicable here.

The following subsection presents a concise mathematical formulation for this implicit auction reference model termed Model I. But Model I, although efficient, is difficult to implement in the context of the IEM. Model II, developed in section 2.2, attempts to improve the practicability of the model at the cost of sacrificing the efficiency somewhat. It still may not be viable in the European context. The following two sections describe both models and subsection 2.3 still another alternative scheme, Model II*.

2.1 MATHEMATICAL FORMULATION OF THE SHORT TERM ENERGY AND CAPACITY AUCTION: MODEL I

The formulation basically corresponds to that of an optimal power flow, see [20]. It has been studied in [21]. For the sake of simplicity, congestions in transmission lines will be the only network constraints considered here. Power Transfer Distribution Factors (PTDFs) are used to quantify the impact that a per unit transaction i between two nodes (single price areas, as an approximation), or the transaction corresponding to a one-sided bid of an agent j at a power exchange, have on every line k of the network. It is assumed that a reasonable number N of *single price areas* exists in the complete IEM. In this case, the potential number of sets of PTDFs to be computed is NxN, where N is much smaller than the total number of nodes. Each set corresponds to each pair of single price areas and consists of one PTDF for each line or corridor of interest for congestion management. If the number of SPAs were very high it would be preferable to use a complete network model, where the intra-area networks could be somewhat simplified[3]. Confidentiality reasons may lead to the

[3] A formulation that is based on PTDFs has been used throughout this paper. The main ideas, however, also apply to a formulation that is based on a complete load flow model for the entire network (see section 2.1.2) and where the individual injections and withdrawals taking place within a SPA could be somehow distributed pro-rata among the different nodes of the SPA. This formulation would require a full network representation but its size does not depend on the number of SPAs or the number of

modification of the actual load and generation patterns within each SPA. If the assumption that intra-area transactions do not significantly affect the flow over congested links is accurate enough this should not cause further problems. However, if SPAs of enough size do not exist, confidentiality of commercial transactions (if this is a real problem) may be incompatible with an efficient dispatch. Computer models for large-scale systems are available to perform these tasks, see [22] for instance.

The bids of individual agents and transactions represent their willingness to pay and are supposed to respond to their respective utility functions.

In order to be able to evaluate the impact on the network flows of each individual unbalanced transaction, a *slack* or reference node is necessary as the counterpart of the transaction. In the formulation below a single slack node responds to all the unbalanced transactions, as all of them are cleared at a single central clearinghouse. The implications of the choice of the slack bus will be discussed later.

Model I:

$$\max_{p_{bi}, p_{uj}} \sum_i B_{bi} p_{bi} + \sum_j B_{uj} p_{uj}$$

$$s.t.$$

$$-\overline{F}_k^- \le \sum_i PTDF_{kbi} p_{bi} + \sum_j PTDF_{kuj} p_{uj} \quad ,\gamma_k \qquad \forall k = 1,2,...number\ of\ corridors \qquad (1)$$

$$\sum_i PTDF_{kbi} p_{bi} + \sum_j PTDF_{kuj} p_{uj} \le \overline{F}_k^+ \quad ,\mu_k \qquad \forall k = 1,2,...number\ of\ corridors \qquad (2)$$

$$\sum_j p_{uj} = 0 \qquad global\ balance\ equation \qquad (3)$$

$$p_{bi} \le \overline{P}_{bi} \quad ,\beta_{bi} \qquad \forall i = 1,2,...number\ of\ balanced\ transactions \qquad (4)$$

$$p_{uj} \le \overline{P}_{uj} \quad ,\beta_{uj} \qquad \forall j = 1,2,...number\ of\ unbalanced\ transactions \qquad (5)$$

$$p_{bi} \ge 0 \quad ,\alpha_{bi}, \alpha_{uj} \quad \forall i,j\ balanced\ and\ unbalanced\ transactions \qquad (6)$$

where:

i: set of balanced transactions (contracts) between agents located within different SPAs.

j: set of unbalanced transactions.

k: set of congested lines or corridors.

---

potentially congested lines. Besides, it can accommodate the formulation of complex system network constraints more easily, see section 4.4.

B_{bi} : price bid (€/kW) per kW of capacity in the considered hour, which is offered by the bilateral transaction i so that it can be executed.

B_{uj} : price bid (€/kWh) per kWh of energy offered by agent j in the considered hour when either buying or selling power to a power exchange. It is positive if j is a load and negative if it is a generator.

p_{bi} : scheduled power of the balanced transaction i.

p_{uj} : scheduled power that is purchased (respectively sold) from a power exchange by agent j (j is involved in an unbalanced transaction). The variable is positive if j is a load (purchases electricity) and negative if it generates electricity.

$PTDF_{kbi}$: power transfer distribution factor of balanced transaction i with respect to line k.

$PTDF_{kuj}$: power transfer distribution factor of unbalanced transaction j with respect to line k when the single slack node previously chosen for the entire system acts as the counterpart.

$\overline{F}_k^{\,+}$: maximum flow allowed by the system operator in the 'positive' direction over line k. Its value must internalize any security criteria that might be considered by the system operator (as a result of a contingency analysis, for instance)[4].

$\overline{F}_k^{\,-}$: maximum flow allowed by the system operator in the 'negative' direction over line k.

$\overline{P}_{uj}$: maximum quantity that the agent j is willing to generate or consume.

$\overline{P}_{bi}$: contracted power in balanced transaction i.

2.1.1 Discussion of the model

The objective function to be maximized in Model I makes sure that those individual agents or transactions who value most the utilization of the network have access to it, regardless of whether they try to execute a bilateral contract or to bid into a power exchange. Welfare maximization is therefore so achieved.

---

[4] This may result in conservative, and therefore inefficient, estimates. Alternatively, additional constraints corresponding to contingency situations may be included. The alternative formulation that is commented in a previous footnote may help here, see [6].

Equations (1) and (2) represent the restrictions that limit the flow on every line of the system (only those lines that might become congested need to be accounted for). All inter-SPA transactions, -even the one-sided ones, with the help of the slack bus associated to the central auction house-, have been expressed in a bilateral format. Equation (3) forces the set of accepted unbalanced bids to be balanced as a whole.

Equations (4) and (5) limit the amount of power that can be scheduled for each transaction. Finally, (6) makes sure that all the scheduled quantities corresponding to a balanced transaction are either positive or zero. Losses have not been taken into account in this streamlined formulation.

The amount that each transaction must pay for its use of the network is determined by its scheduled power, its PTDF in any congested line and the dual variables γ_k and μ_k of equations (1) and (2). Thus, the mathematical expression of the charge that a bilateral transaction *i* would have to pay for the capacity right over line *k* is:

$$C_{ki} = (\gamma_k + \mu_k) PTDF_{kbi} p_{bi} \tag{7a}$$

Equation (3) implies that the selection of the single slack node as the reference for each unbalanced transaction is irrelevant under the point of view of which bids are accepted in the auction, since the accepted bids could now be paired (arbitrarily) as bilateral transactions where the slack node would not have any function to play. Why is it then necessary?. Because the formulation we have used is based on the concept of PTDF and in order for the impact of a transaction on the flows to be expressed in terms of PTDFs this transaction must be balanced. The location of the slack node determines the PTDFs of each unbalanced transaction and therefore its responsibility in the payments for capacity rights. The charge an unbalanced transaction *j* would have to pay for the capacity right over line *k* is:

$$C_{kj} = (\gamma_k + \mu_k) PTDF_{kuj} p_{uj} \tag{7b}$$

One may feel uneasy by learning that the arbitrary choice of the reference node (perhaps at the physical location of the central clearinghouse, although not necessarily) determines the congestion management payments for the unbalanced transactions. However, this effect is counteracted by the differences in energy prices that should take place among the different SPAs as a result of the existence of congestions. For instance, assume that a generator *g* presents an unbalanced bid *j* to the central auction house and it is accepted and paid the corresponding marginal price, which is the price of the SPA where the slack node has been chosen[5]. Bid *j* also has to pay the capacity charge to access that SPA. The net result is equivalent to the generator selling its output at the energy price of the SPA where it is physically connected. In Model I the resulting energy prices for the unbalanced transactions at the different SPAs (i.e. the energy price at the reference node plus the corresponding transmission capacity charge) are equivalent to those obtained in a scheme of zonal pricing. Bilateral transactions should mimic those zonal prices when trying to arbitrage energy prices. Then, the choice of the reference node should not have any practical implications on the physical dispatch of generation and the net remuneration of each generator.

Any payments for the use of the congested lines should not result in additional income for the TSOs that operate these lines, nor for the owners of the lines. Both the system operators and the transmission owners (with some rare exceptions, as it is the case of merchant lines) are regulated entities that are subject to price control due to the monopolistic nature of the transmission activity. Under ideal conditions (if economies of scale were not present in the planning of the grid and investments were continuous) revenues resulting from the application of nodal prices would suffice to pay for the construction and maintenance costs of the grid. In real-life systems congestion rents only pay for part of these costs, around 20% according to some estimates in [23]. Regardless of how much they recover it seems clear that the congestion rents should then be used to reduce the authorized revenue for transmission owners that has to be recovered via network tariffs.

Two alternatives are possible at this point: either use them to reduce network access tariffs line by line, i.e. devoting the difference in prices between both ends of a line times its flow to reduce the amount paid by agents that otherwise would have to afford the whole cost of the line, or use them to reduce network access tariffs in the system as a whole. The first alternative relies on the idea that those agents paying for a line seem to be entitled to receive, to the extent possible, the benefits created by the line, including the congestion rents. However, distributing congestion rents for each line in proportion to the fraction of the cost of the line afforded by each agent is equivalent in practical terms to granting transmission rights over the line in the same proportion. The initial allocation made of transmission rights would be discriminatory and, even more important, potentially harmful from an efficiency point of view. Recent studies conducted at the University of Cambridge (see [24]) conclude that, under ideal circumstances, traders put a higher value on

[5] The energy price corresponding to the SPA where the slack node is located can be computes as the dual variable of the global balance equation (3).

transmission rights that those agents whom ownership of these rights encourages them to abuse their market power. Consequently, traders will outbid these agents in a single price marginal capacity auction. Additionally, we may be able to define transmission rights such that their value does not increase with the exercise of market power by their owner. However, none of these measures, aimed at fighting market power, would be of any help if, prior to any auction, congestion rents resulting from the auction were assigned line by line in proportion to the cost of each line born by each agent. What is more, this allocation scheme could significantly alter the short-term economic signals that market agents receive. Agents with the ability to affect unilaterally energy prices may see how the value of their portfolio of transmission rights changes with their decision to increase or decrease their power output. Some agents' variable costs would depend on the rights they own. Potential problems like these may be fixed, at least partially, if the cost of a line works as a limit to the congestion rents received by market agents who pay for the line.

As mentioned above, the second possible approach would be to use congestion rents resulting from the sale of capacity over a line for the first time to pay every line in the system in the same proportion. In this case we would solve all the problems mentioned above, while at the same time contributing to the recovery of the grid sunk cost and eliminating any perverse incentive for the TSO or transmission owners. This second alternative may be seen as less fair than the first one nevertheless. This issue is dealt with in Article 6 of the new Regulation that has been recently approved by the EU Commission and Parliament, see [25].

Note that the use of Model I for congestion management is not in conflict with the existence of shorter-term energy markets, such as intra-daily markets, regulation markets or markets for operating reserves. Market agents may use these markets to better adjust their positions to their actual interests or possibilities in real time or to participate in the provision of ancillary services.

Despite the simplicity and compactness of Model I formulation, -which ensures that congestions are managed with maximum efficiency (within the approximations of the single price area assumption), one has to admit that Model I demands a level of centralization of the IEM that appears to be unacceptable under the present regulatory conditions. Note that in Model I there is no other power exchange than the central auction house.

The next subsection presents an alternative formulation of Model I which is based on a full nodal representation of the grid. This formulation may be of special interest when the number of Single Price Areas is so high that the resulting number of PTDFs to be computed becomes unreasonably high.

2.1.2 NODAL FORMULATION OF MODEL I

In this section an alternative formulation of the reference auction model is presented. The formulation presented below corresponds to the classical optimal load flow where some modifications have been included to take into account the possible existence of bilateral (multilateral) contracts.

$$\max \; B^t p$$

$$s.t$$

$$SHy + y = 0 \tag{8}$$

$$y = Zp \tag{9}$$

$$Hy \le \overline{F} \tag{10}$$

$$\underline{F} \le Hy \tag{11}$$

$$p \le \overline{P} \tag{12}$$

$$\underline{P} \le p \tag{13}$$

$$Up = 0 \tag{14}$$

where B is the vector of unbalanced bids by agents, p is the vector of scheduled powers corresponding to each one of these bids, S is the incidence matrix that contains the topology of the grid (which line is connected to which node), H is the transfer admittance matrix which allows us to express the line flows as a function of the net bus injections, Z is the matrix that provides the specific location (node) for each bid and U is the matrix that relates each bid by an agent to the bilateral transaction, if any, the bid belongs to. Finally, $\overline{F}$, $\underline{F}$, $\overline{P}$ and $\underline{P}$ are the limits to the line flows and the power scheduled for each unbalanced transaction respectively.

Discussion of the model:

The objective function we are maximizing is exactly the same as the one in Model I. The balanced transactions have been decomposed into two or more different unbalanced transactions, each of them offering to either withdraw or inject power into the grid. Of course, the aggregate generation corresponding to the unbalanced transactions being part of a balanced transaction must equal the aggregate demand.

Equation (8) is the balance equation for each of the nodes in the system. Its dual variable is the nodal price at the corresponding node. Equations (10) and (11) limit the flow over each line in both directions. Equations (12) and (13) ensure that the amount of power granted to each offer in the auction is within the limits established by the agent. Finally equation (14) guarantees the equilibrium between generation and demand for each balanced transaction.

As said earlier, one can make a case for this 2nd formulation of Model I when there is a high number of SPAs. The size of the optimization problem to be solved when the original formulation, based on the computation of PTDFs, is used grows proportionally to the square of the number of SPAs whereas that of the nodal formulation model remains unchanged.

Because of the implementation difficulties with Model I, section 2.2 presents a new model which tries to preserve as much as possible the current functions and independence of the individual power exchanges within an implicit auction scheme.

2.2 MATHEMATICAL FORMULATION OF THE SHORT TERM ENERGY AND CAPACITY AUCTION: MODEL II

In Model II every unbalanced transaction has to specify the power exchange *e* that will receive its offer to buy or sell power. Each power exchange is assigned a different reference node. Thus, every market agent that engages in a bilateral contract or that bids to a power exchange chooses where to trade its energy. The impact on the network flows of each individual unbalanced transaction can be accurately quantified. Some solutions similar to the one proposed here have already been presented in the academic literature, see [8] for instance. The selection of the reference nodes, even if distributed slack nodes are used (see [17]), do not affect the price paid or charged to each agent for its energy. In the case of an agent bidding at a power exchange, this reference node is the slack node chosen for this power exchange. Given a set of bids by agents, the algorithm maximizes the global value of the accepted bids subject to the specified network constraints. The set of accepted bids is independent of the reference node used for each power exchange. The different components of the final price of energy that each agent sees (energy and network congestion), are coherent and always refer to the node where the agent is located. However, the decision by agents on whether to bid to a power exchange or another is relevant with regard to the final outcome of the optimization process. Unless perfect arbitrage exists, the set of accepted bids is dependent on where agents address them. We cannot talk about nodal prices because an offer from an agent at a specific node may get different prices depending on where it is addressed. As explained below, the liquidity of the trade at each power exchange may be a matter of concern under the conditions expressed in Model II.

Model II:

$$\max_{p_i, p_j} \sum_i B_{bi} p_{bi} + \sum_j \sum_e B_{uje} p_{uje}$$

$$s.t.$$

$$-\overline{F}_k^{\,-} \le \sum_i PTDF_{kbi} p_{bi} + \sum_j \sum_e PTDF_{kuje} p_{uje} \quad ,\gamma_k \qquad \forall k = 1,2,....number\ of\ corridors \tag{15}$$

$$\sum_i PTDF_{kbi} p_{bi} + \sum_j \sum_e PTDF_{kuje} p_{uje} \le \overline{F}_k^{\,+} \quad ,\mu_k \qquad \forall k = 1,2,....number\ of\ corridors \tag{16}$$

$$\sum_j p_{uje} = 0 \qquad set\ of\ balance\ equations \quad ,\tau_e \qquad \forall e = 1,2,....number\ of\ power\ exchanges \tag{17}$$

$$p_{bi} \le \overline{P}_{bi} \qquad ,\beta_{bi} \quad \forall i = 1,2,...number\ of\ balanced\ transactions \tag{18}$$

$$\underline{P}_{uje} \le p_{uje} \le \overline{P}_{uje} \qquad ,\beta_{uje} \quad \forall je = 1,2,...number\ of\ unbalanced\ transactions \tag{19}$$

$$p_{bi} \ge 0 \qquad ,\alpha_{bi} \quad \forall i\ balanced\ transactions \tag{20}$$

where:

je: set of unbalanced transactions that bid at power exchange *e*

B_{uje}: bid price (€/kWh) per kWh of energy offered by agent *j* in the considered hour when either buying or selling power to power exchange *e*. It would be positive in case *j* is a load and negative in case it is a generator.

p_{uje}: scheduled power that is purchased (resp. sold) from power exchange *e* by agent *j*. It is positive when agent *j* purchases power and negative when it sells power.

$PTDF_{kuje}$: power transfer distribution factor of unbalanced transaction *j* with respect to line *k* when a prescribed slack node representing the power exchange *e* is used as the counterpart.

$\overline{P}_{uje}$, $\underline{P}_{uje}$: maximum and minimum quantity that the agent *j*, who offers its energy at power exchange *e*, is willing to generate or consume.

2.2.1 Discussion of the model

Most constraints and the objective function to be maximized in Model II are the same as in Model I, but now there are additional constraints whose impact on the global economic efficiency needs to be evaluated.

Again, all transactions, -even the one-sided ones, with the help of the slack bus associated with each power exchange *e*-, are expressed in a bilateral format. The set of restrictions represented by equation (17) forces a global match of all the unbalanced bids addressed to each power exchange *e* and subsequently accepted. This bilateral association of each unbalanced transaction and a power exchange is critical in deciding how much it should pay to go through the congested network, since the value of the PTDFs directly depends on the origin and end of each transaction.

Employing a slack node for each power exchange may have some implications. In order for formulations of Models I and II to be equivalent in practice, market agents that send unbalanced bids under Model II should be able to accurately predict the resulting price of energy in the different power exchanges. This is due to the fact that in Model II the unbalanced bids are not jointly considered as in Model I, but they are separately matched at the power exchange where they have been sent. Errors in estimating energy prices at the different power exchanges by the market agents who send unbalanced bids may result in the dispatch outcome of Model II not being the same as that of the reference Model I. Note that under Model I unbalanced transactions are not forced to bid into a specific power exchange. For instance, a bid to a power exchange *e* may be rightly rejected but would have been accepted at a different power exchange *f* in Model II, and also in Model I.

If market agents were perfectly informed, due to arbitrage of prices, they would receive the same price signals as under a centralized auction. As a result, the resulting pattern of generation and load would be the same also and so the use made of the grid. The network congestion rent that each agent would have to pay for the capacity required to carry out its transaction would depend on the power exchange to which it had addressed the bid. However, the price it would charge or pay for the energy would be the energy price at the corresponding power exchange. Consequently, the net payment (either positive or negative) would be independent of the selection of the particular power exchange where the unbalanced bid has been sent.

Note that, properly speaking, the concept of price areas does not exist in Model II. Two different power exchanges could theoretically be located quite near each other and within the same SPA. However, if agents are not capable of arbitraging prices appropriately, the price of energy in both markets could be very different. At the same time, agents connected to the same node of the grid could see different energy prices depending on where they submit their offers.

Then, why should one use Model II, where the rigidity of the unbalanced bids that have to be addressed to specific power exchanges could impair the achievement of full economic efficiency? The interest of Model II resides in that it seems to be more amenable to a viable implementation in the IEM context. An IEM system-wide nodal energy pricing scheme, or a global IEM-wide energy auction that is run by a single power exchange as in Model I, are far from what appears to be achievable in the European Union in the short and medium terms.

Model II relies on arbitrage among market agents in order to achieve an economically efficient solution. Perfect arbitrage will never take place and there will be some loss of efficiency. This is the price that must be paid so that agents may bid anywhere and existing power exchanges maintain their individual functionality. Model II increases flexibility and decentralization at the expense of some efficiency loss.

The following is a description of what could be a pragmatic blueprint for implementation of Model II:

i) Define a number N of single price areas (SPAs) that could comply with the single price area assumption with sufficient accuracy just for the purpose of the bid matching algorithm. Locate each power exchange at a single node within a representative SPA, or maybe distribute it among several nodes of that SPA, see [17]. Model II only considers balanced transactions between agents located within different SPAs and all the unbalanced transactions.

ii) A central auction house runs the joint energy and transmission capacity auction of Model II, despite the existence of several power exchanges. All transactions that are scheduled by Model II are firm and they have to be physically executed, including the bilateral contracts that have obtained the transmission rights to be carried out. The set of scheduled unbalanced transactions for any power exchange must be balanced.

iii) The central auction house communicates the results of Model II to all the power exchanges and the TSOs. At this point, each power exchange could run its individual auction. There are two possible options for this auction: the first one would be to include only the generation and demand within the SPA where the reference node is located. This auction could not modify the scheduled transactions involving agents located outside this SPA. The second option would be to organize an auction where agents from any part of the region covered by the corresponding power exchange could take part. This last option would be even more appealing in those cases where the power exchange is somehow associated with a control area, such as OMEL in Spain or APX in The Netherlands. In this case, the last auction could be jointly applied to the several SPAs within the control area, therefore increasing the liquidity and reducing the effect of the balancing requirement in a single SPA that was discussed before. The result of this auction would probably be unfeasible since some of the transactions involved would make use of the capacity of the congested lines. A balancing or other short term market could then be run at control area level to solve the infeasibilities caused by the previous auction within the region covered by the control area. The ensuing potential problem is that the consequences of the auction may affect the flows in SPAs external to the considered control area, since the auction allows transactions between different SPAs. This relaxation procedure may make sense specially when applied to "regional markets", such as Scandinavia, UK plus Ireland or the Iberian Peninsula. The same approach would work when the region's geography or the topology of the network causes the impact of transactions between SPAs within the region on flows outside it to be negligible.

After portraying Model II, one has to face again the same important question as in Model I: Is there any possibility that the coordinated implicit auction in Model II is adopted for congestion management in the IEM? It seems also very unlikely. Again Model II requires a high level of centralization of the energy market in the IEM, a level that most Member States will not probably be willing to accept now.

Therefore, although a centralized IEM-wide auction house for the short-term markets of energy and transmission capacity could conceptually look like an appealing proposition, it appears unlikely that such a scheme could be adopted. This is why an alternative and more plausible approach is proposed in the next section.

*2.3 AN EVOLUTION OF MODEL II TO ALLOW POWER EXCHANGES PLAY A MORE SIGNIFICANT ROLE: MODEL II**

Before moving to the approach in next section, this subsection discusses one more version of the coordinated implicit auction algorithm in which the individual power exchanges play a more significant role. The first step consists of an energy only auction at power exchange level, where the network constraints would be ignored. The results from each one of the auctions (as many as power exchanges) are accepted and left unchanged in the following steps. The second step is actually a version of Model I where the non-accepted bids in the first step are used. The obvious problem is that there is no guarantee that any possible congestion created in the first stage of the process can be fixed with the means available in the second step. The alternative is to unfreeze the results in the first step, so that they could be modified in the second step, but then we are back in the centralized situation where the individual power exchanges become irrelevant.

Other schemes based on the implicit auction approach may be possible. EuroPEX, the European Association of Power Exchanges, has recently proposed a scheme based on a series of day-ahead bilateral implicit auctions between pairs of countries sharing a common congested border. Then, an iterative procedure follows, so that the non-accepted transactions in one auction may be used by another one, see [26]. The detailed mathematical formulation of this algorithm, as well as its properties of convergence and its practical viability, remains to be put forward.

3 A PLAUSIBLE ALTERNATIVE TO THE REFERENCE CASE

The basic idea is to try to avoid the centralized energy auction, -because of the aforementioned implementation difficulties-, while maintaining as much as possible the good properties of the scheme in Model I. We present here a coordinated explicit auction of transmission capacity as a suitable alternative to implicit auctions. The European Association of Transmission System Operators (ETSO) has already proposed this type of scheme, see [2].

The formulation of the explicit auction also adopts a breakdown of the IEM system into single price areas as Models I and II do. Defining a different reference node (either a physical or a virtually distributed node) for each power exchange *e* makes it possible to compute the impact on the network flows of the bids at each one of them.

The essential part of this scheme is an IEM-wide explicit auction of transmission capacity that must take place at a central auction house prior to the opening time of the short-term (day-ahead, typically) energy market. In this auction, only bids for transmission network capacity are accepted. The format is the same for both the bilateral contracts (balanced transactions) and the energy bids at power exchanges (unbalanced transactions). All of them are bids, expressed in €/kW for the considered hour, made by agents in order to market their energy at a specific power exchange (assigned to a specific geographical area) or to exercise a bilateral contract between two areas.

The product that is auctioned is the right to establish a transaction between two single price areas. The transaction can be a bilateral contract or a transaction at a power exchange. Therefore, the market participants do not have to acquire separate transmission rights for individual lines that could become congested (flowgate rights), but a point-to-point right (single-price area to single-price area, if the approximation is an acceptable one), which will implicitly include any necessary transmission rights on the affected lines. This is why the formulation below is expressed in terms of transactions and not in terms of transmission rights on individual lines. This format of the auction is much more convenient for the market participants.

3.1 MATHEMATICAL FORMULATION OF THE EXPLICIT AUCTION: MODEL III

In contrast to the implicit auction in Model I, now only bids for capacity are accepted. The explicit auction assigns the available capacity in such a way that the agents who are granted additional capacity are the ones who value it most. The set of accepted bids must be feasible. The formulation proposed here is very similar to that of the implicit auction in section 2.1.

Model III:

$$\max_{p_i, p_j} \sum_i B_{bi} p_{bi} + \sum_j B_{uj} p_{uj}$$

$s.t.$

$$-\overline{F}_k^- \leq \sum_i PTDF_{kbi} p_{bi} + \sum_j PTDF_{kuj} p_{uj} \quad ,\gamma_k \qquad \forall k = 1,2,....number\ of\ corridors \tag{21}$$

$$\sum_i PTDF_{kbi} p_{bi} + \sum_j PTDF_{kuj} p_{uj} \leq \overline{F}_k^+ \quad ,\mu_k \qquad \forall k = 1,2,..number\ of\ corridors \tag{22}$$

$$p_{bi} \leq \overline{P}_{bi} \quad ,\beta_{bi} \qquad \forall i = 1,2,..number\ of\ balanced\ transactions \tag{23}$$

$$p_{uj} \leq \overline{P}_{uj} \quad ,\beta_{uj} \qquad \forall j = 1,2,..number\ of\ unbalanced\ transactions \tag{24}$$

$$p_{bi}, p_{uj} \geq 0 \quad ,\alpha_{bi}, \alpha_{uj} \qquad \forall i, j\ balanced\ and\ unbalanced\ transactions \tag{25}$$

where B_{uj} is the bid price in (€/kW) offered by agent j when bidding for the capacity it needs to access the power exchange where it plans to trade its energy. Agent *j* offers this price per kWh of energy to be produced or consumed in the considered hour. Note that this meaning is different from the definition of B_{uj} in Model I.

The notation that has been used here is completely analogous to that of Model I with the sole exception that in Model III agents are bidding only for capacity and all transactions are therefore balanced. As said before, Model III does not incorporate energy bids.

Due to the fact that only balanced transactions are considered (since all bids are bids for capacity), we do not need to include the set of restrictions associated with the balance of generation and load for each power exchange. By using as a counterpart the reference node of the corresponding power exchange, the central auctioneer computes how much capacity an agent needs to acquire on each link so as to market its energy at this power exchange. In principle, the transactions that are scheduled in Model III have to materialize physically in the short-term energy market (see the comments on the relation with the short-term energy market in subsection 3.1.1).

As in Model I, TSOs determine the limits expressed in equations (21) and (22) on the basis of the characteristics of the network and the expected pattern of generation and load in the IEM, as well as any potential contingencies that may occur in the system. The a priori declaration of all transactions that can affect inter-SPA congestions will reduce the uncertainty in the specification of these limits to the line flows. The precision of this estimation will also depend on the adopted number of single price areas: the more the areas the higher the accuracy.

3.1.1 Discussion of the model

This subsection discusses several issues of much relevance in the implementation of Model III.

Coordinated versus individual auctions

Because of the lack of an agreed IEM-wide method for congestion management, explicit auctions for transmission capacity are presently being held, -or have been proposed-, for several frequently congested inter-connectors between Member States [27]. This lack of coordination may easily result in a loss of efficiency, and even into unfeasible flow patterns, since avoiding the mutual interdependence among different parts of the network is simply not possible. A coordinated allocation algorithm, such as the one in the preceding section, is the adequate response to this problem.

Assumptions behind the definition of the SPAs

As with Models I and II, the successful implementation of Model III critically depends on whether we are able to find a set of SPAs that satisfactorily meet the basic conditions expressed in section 1.1. Otherwise, we would be facing an unsolvable dilemma: nodal pricing would be the only alternative left but it may be too centralized to be acceptable for the IEM.

Relation with the short-term energy market

The purpose of a coordinated explicit auction of transmission capacity is to facilitate *a priori* that short-term energy markets do not run into problems of unfeasibility because of network congestion. This requires a careful design of the coupling between the explicit auction of capacity and the short-term energy markets, -which typically consists of a day-ahead energy market, typically followed by even shorter-term markets-.

The first point to be noticed is that, unlike Model I, the combination of an explicit capacity auction and a short-term energy market, -where only the inter-SPA transactions that were scheduled in the previous capacity auction are allowed to bid-, is not a fully complete market for energy in the sense indicated by [4]. The fact that separate auctions are employed to optimize the allocation of transmission capacity and energy may entail some loss of efficiency and even the risk of unfeasibility. This combination of a coordinated explicit auction followed by independent dispatches for each country is less efficient than Model I. Additional constraints appear in this second approach in order to ensure that the resulting final dispatch is feasible. The following paragraphs explain this idea with more details.

Any agent who wants to bid at a power exchange located in any external single price area first has to acquire the corresponding physical transmission right in the capacity auction. Depending on the model formulation, the commitment of the agent to comply with the results of the auction will vary. Two possibilities are considered here:

Alternative A: "Use-it" (obligations)

This alternative is consistent with the formulation proposed for Model III in equations (21) to (25). Once an agent has obtained a right in the auction, it must use it, i.e. it has to carry out the physical transaction. Otherwise, the resulting pattern of transactions (where counter-flows are not only possible, but frequent) probably would not be feasible. Then, the agents that have been granted some transmission capacity in the explicit auction must make sure they are also scheduled in the energy market (in some power exchanges there are shorter-term markets, -intra-daily markets or regulation markets-, where it might be possible to adjust the commercial position in order to meet the commitment from the explicit auction). If agents do not comply with this requirement, they face administrative economic penalties. This rule is stronger than the frequently mentioned "use-it-or-lose-it" rule, which does not protect against the counter-flow problem. The rule proposed here has also been termed an obligation contract, see [5]. It basically guarantees the feasibility of the outcome of the energy market when capacity rights granted in both directions over a link are netted out as in Model III (see equations (21) to (25)).

The "use-it" rule, combined with the uncertainty about market prices, may sometimes result in some loss of efficiency when compared with the optimal outcome of Model I. The reason is simple. A generator, for instance, that has acquired in the capacity auction the transmission right to sell into a power exchange located outside its single price area, is now committed to deliver this energy regardless of the market price at the point of delivery. To avoid being penalized, the generator then will have to bid its energy in the short-term energy market at a very low price. Then the strategy of the generator when bidding in the explicit capacity auction should be to internalize in its bid its best estimate of the final price of energy in the power exchange. Therefore, in this process all the risk is shifted to the capacity bid. Given the unavoidable errors that will be committed by the agents when predicting the market energy price, some efficiency losses will inevitably occur. Note that the energy transactions among SPAs are scheduled in the explicit capacity auction the day before the short-term energy market takes place. This scheme seems to lack an acceptable level of flexibility.

The "use-it" rule has the purpose of guaranteeing feasibility while maximizing the use of the available transmission capacity between SPAs. It achieves this objective at the expense of incurring into some loss of efficiency, as explained before. However, since the explicit capacity auction must take place before any of the energy auctions at the power exchanges opens, unexpected events, such as plant or transmission line outages may take place in between. Hence, feasibility cannot be guaranteed one hundred per cent. Besides, the assumption that transactions that are internal to single price areas do not have any implication on the considered network congestions is only an approximation. Consequently, the TSOs will have to monitor these effects and apply ad hoc measures (such as some amount of redispatch or counter-trade) to solve any unfeasibility problems that may arise because of these reasons. In order to reduce the unfeasibility risk, the TSOs may decide to add some safety margins in the definition of the limits $\overline{F}_k$ in equations (21) and (22). The final implication of all this is an additional loss of efficiency.

The major advantage of alternative A is that it preserves the independence in the functioning of the several power exchanges, since each one will separately receive and handle (ignoring network constraints) the energy bids from those agents within the same SPA and from those who have previously acquired the corresponding transmission right.

However, the liquidity of the short-term energy market of any power exchange will be too small in general. Since all accepted capacity bids have to become energy transactions because of the "use-it" rule, only energy bids from within the SPA where the slack node of the power exchange is located will be able to emerge as new in the short-term energy market. These energy bids have to balance the, in general, unbalanced set of capacity-transformed-into-energy accepted transactions that were successfully addressed to the considered SPA in the explicit capacity auction.

A first implication is that the SPA where the slack node of a power exchange is located should be chosen so that it contains a large number of active agents. In conclusion, the number and size of the SPAs are critical factors for the successful implementation of Model III. If the SPAs happen to be very small, in order to make the implementation of this scheme possible, a mostly bilateral market will have to be devised where there would be almost no room for power exchanges.

Alternatively, energy and capacity could be marketed bilaterally during a continuous trading window. Offers addressed at energy trading hubs would match others on a continuous basis too, i.e. without having to resort to a centralized auction for the energy bought and sold at the hub.

Accordingly, there would not be a single price for each trading hub. This market design would provide some level of integration between the energy and capacity markets and may probably solve liquidity problems exhibited by Model III.

Other possibilities may be considered. In the first one, agents from several SPAs could jointly respond in the short term energy market in each power exchange. In this way, the unbalanced set of capacity-transformed-into-energy accepted transactions from the explicit capacity auction might be successfully balanced without any liquidity problems. However, this may result in undesirable effects on the inter-SPA congestions, as explained in section 2.2.1, where a similar approach was proposed. These infeasibilities might be solved by resorting to a pan-European balancing market.

A second possibility, which also makes use of obligations, consists of making SPAs coincide with the current countries in the capacity auction. Energy auctions for each country would follow where the internal grid of the country would be considered. Finally, a balancing EU wide market would take place so that market agents can adjust their positions and also solve infeasibilities.

Alternative B: "Use-it-or-lose-it" (options)

A third possibility involves applying the milder "use-it-or-lose-it" rule to the explicit auction that is run the day before the short term energy market and also to all those explicit auctions taking place before that. There are several ways to proceed in this case:

Version B1: allowing for the netting of flows

If the algorithm in Model III is applied, flows in opposite directions over a line are netted out. However, because of the looser "use-it-or-lose-it" rule, this design does not comply with the principle of revenue adequacy. Therefore, the resulting pattern of flows may not be simultaneously feasible[6] [5, 28]. Again, several options are possible.

- Just before the short term energy market, the central auctioneer could run an explicit auction where the "use-it" rule would apply. This auction could make use of any capacity reserved for transactions that were not finally carried out. Holders of rights obtained in previous auctions, where the 'use it or lose it' rule applies, would be paid according to the congestion rents obtained from the last capacity auction, where transmission rights are firm. After the

[6] In order for the outcome of the auction to be feasible, model III should include as many additional constraints as possible combinations of the granted options. Each one of the constraints will ensure that the corresponding combination of transmission rights is feasible. But this would make the resulting auction much more difficult to solve.

short-term energy market, intra-daily energy markets (as in Spain) or a balancing market run at regional level, could help the agents to achieve balanced commercial positions.

- Another possibility would be to carry out the energy dispatch for each power exchange without considering the grid. Finally, a balancing market could be used to solve grid problems.

Version B2: flows in opposite directions are dealt with separately

Giving more flexibility to the bids for the transmission capacity needed by agents to access the energy markets of individual power exchanges reduces the risk of economic loss for these market participants. If, in this case, we net the power flows over each line, the possibility of ending up with an unfeasible schedule increases. In order to avoid this, the formulation of Model III can be modified so that flows in opposite directions are dealt with independently. As a result, the capacity of the transmission network would not be fully used in general. This may be quite conservative in a well-meshed network, where counter-flows frequently take place. Due to the fact that the algorithm does not allow for flows in opposite directions to cancel out, the "use-it-or-lose-it" rule can be safely applied. Any agent not willing to use the capacity rights it has obtained in the explicit auction of the preceding day, will simply communicate its decision to the system operator so that somebody else can use the spare transmission capacity. The scheme may be enhanced by incorporating a minimum bid price in the capacity auction in order to deter agents from trying to accumulate network capacity that they do not need. Capacity that might be sold later at abusive prices.

The Balancing Market

As already mentioned in the previous paragraphs, running an additional balancing market may solve the problems that have arisen with alternatives A and B. This balancing market would involve running a Model I type of auction (joint auction of energy and transmission capacity) at a central auction house. In the balancing market adjustments to the positions resulting from the scheduled bilateral contracts and the energy markets at the power exchanges would take place in order to solve infeasibilities. The "use-it" rule would apply in this market and, therefore, commitments would be firm. The transmission capacity that is left unused by those with transmission rights acquired in previous auctions where the 'use-it-or-lose-it' rule is applied should be made available to any other agents.

Since Model I is a complete market model (if the assumption about the single price areas is ignored), there will be nothing to object to the final result of the balancing market. The problems of this design option, as with Model I in the reference case, lie in the implementation. On one hand, running a third auction is an inconvenience, even if it is done automatically, without requiring additional bids from the market participants. But, most importantly, this Model I-type third auction, unless subject to many restrictions that would reduce its usefulness or employed just to solve the infeasibilities resulting from the previous markets, would make the previous explicit capacity auction and mainly the energy auctions at the power exchanges to appear as useless. Market agents could always resort to this third auction to meet their objectives. The current power exchanges might be reduced to the lesser role of "scheduling coordinators", as in the failed Californian model, see [29].

Long and medium term capacity auctions

Most market participants may wish to hedge against the uncertainty that the appearance of network constraints may introduce in the congestion charges corresponding to their transactions, whether these transactions are bilateral trades or bids at power exchanges across congested lines. This can be accomplished by allowing the explicit capacity auctions to take place also at longer time intervals, such as one or several months and maybe up to a few years.

The format of the longer-term capacity auctions would also be the one described in Model III. In order to promote the liquidity of this market, the transmission rights to be negotiated could be financial instead of physical.

However, during the short-term explicit capacity auction held just before the short-term energy market, all transmission rights have to become physical, since these rights are required to materialize bilateral contracts or to bid at power exchanges outside the single price area of an agent. Section 4.1 explains how to perform this conversion together with a detailed examination of the convenience of the use of one type of rights or the other.

With regard to the organization of these auctions, different options are possible. First, the central auctioneer could auction the entire transmission capacity of the network already at the first long-term auction. Subsequent auctions would then offer the opportunity for agents to adjust their positions buying some more transmission rights or selling some of them. Second, the TSO could "slice" the total transmission capacity into several fractions, which would then be put on sale in separate auctions. It would reserve one last fraction of the transmission capacity for the last short-term explicit capacity auction. Section 4.3 will discuss the pros and cons of both alternative approaches.

Acceptability of the proposed scheme

The preceding discussion assumes that large enough SPAs exist. The proposed scheme (alternative A) consists of an explicit auction of transmission capacity and subsequent separate energy auctions at the different power exchanges subject to the “use-it” rule. TSOs will cope with any minor infeasibility problem by introducing any required adjustments (this is what they have been doing for many years with an excellent security record, although, admittedly, with a lower level of cross-border trade).

This scheme preserves the operational independence of the power exchanges. The potential loss of economic efficiency appears to be small in theory. There may be serious problems of market liquidity although the academic community and industry players have provided several ideas in order to cope with this issue.

The issue of which organization will run the explicit capacity auction remains open. Given that the specification of Model III and the interpretation of its results is mainly a technical problem, -although with economic implications-, the European Association of System Operators (ETSO) seems the most suitable entity to perform this task. ETSO would be subject to the supervision of the national regulatory authorities and their association, the Council of European Energy Regulators (CEER).

Allocation of the congestion rents

The criteria for the assignment of the congestion rents should be the same described for the reference case in section 2. How to provide incentives for the System Operator in order to reduce the congestion management costs without, at the same time, encouraging it to reduce the level of system security remains an open issue.

4 ADDITIONAL ISSUES

Once a conclusion has been reached regarding the most appropriate congestion management scheme for the IEM, -within the regulatory environment that is foreseeable for the near and medium-term future-, this section studies in more detail some technical issues to be solved before arriving at a successful implementation. Four topics are addressed: i) the nature -whether financial or physical, flowgates or point-to-point, options or obligations- of the transmission rights to be auctioned [7, 28, 30], ii) who are the agents or transactions that should be involved in a congestion management mechanism, iii) the convenience of the division of the total transmission capacity into slices or fractions to be separately assigned in the several long-term auctions and iv) the criteria for allocation of congestion rents that are derived from complex or unusual network constraints.

4.1 THE NATURE OF THE TRANSMISSION RIGHTS

Options versus obligations

This is very much related to the debate on "use-it" versus "use-it-or-lose-it". An *option* transmission right gives its owner the possibility to exercise it, i.e. to use the corresponding transmission capacity or receive the corresponding congestion rents, but it does not entail the obligation to do so. On the contrary, with an *obligation* the agent has a commitment to make use of the right. The main consequence of choosing one type or the other in Model III is that, with options, it is not possible to net power flows over a line if we want to preserve the simultaneous feasibility condition. This is because there is no guarantee that the corresponding transactions will ever take place [5, 28]. If we want to apply netting of flows, while using options, an additional mechanism such as a very liquid balancing market should be put in place in order to solve infeasibilities. Therefore, unless obligations are used, it will be impossible to optimize the utilization of the grid. When options are employed, separate limits have to be placed for the total amount of transactions that cause flows in one direction and in the opposite direction, since netting of flows cannot be ensured. On the other hand, options give agents more flexibility to adjust their transmission rights to the commercial positions they hold. The adopted scheme should keep a balance between the higher use of the grid allowed by obligations and the stronger flexibility that options give to the agents. Normally, the more flexible a market is the more liquidity it has.

Flowgate versus point-to-point rights

There is abundant academic literature both in favor and against these two approaches [5, 7, 9, 28]. Despite some questionable advantages that some authors see in the use of flowgate rights, this paper makes a choice for point-to-point rights. These could be either physical or financial in the long term and, in the case of separate markets for energy and capacity, they will have to become physical before the operation. Node-to-node rights (or area-to-area rights, if the single price area assumption is acceptable) confer an unambiguous right to execute a transaction (or to capture a congestion rent, as a difference between the two nodal or area prices) between the corresponding nodes or areas, regardless of any changes in network configuration, new constraints, contingencies, PTDFs, etc. All these changes, however, will affect in general the fraction of the capacities of flowgates (i.e. individual congested lines) that are needed to execute a given transaction. Therefore, a collection of flowgate rights will not necessarily match the flows caused by a bilateral transaction nor will it fully hedge against the uncertainty in the price difference between two nodes or single price areas.

Both Models I and III are based on transactions between single price areas, and therefore make use of point-to-point transmission rights. However, equations (21) and (22) explicitly compute the utilization of each line by each transaction. This is of great help to contest a frequently heard criticism about point-to-point transmission rights. Many people argue that the use of point to point rights would cause lack of liquidity in the auctions since very few agents would be interested in buying or selling a capacity right precisely between nodes m and n of the transmission network. The point is that formulation in Model III implies that, when the agents explicitly bid for point-to-point rights, the model is implicitly exchanging flowgate rights. It must be acknowledged that, contrary to what happens with point-to-point rights, flowgate rights might not need to be mediated through a centralized institution. Consequently, flowgate rights may make secondary capacity trading a bit more flexible than what can be achieved with point-to-point rights. This advantage of flowgate over point-to-point rights does not offset the positive general effect of the use of point-to-point rights, however.

Physical versus financial rights

Financial transmission rights allow the owner to collect the corresponding congestion rents. In the case of a point-to-point contract from node j to node k, the congestion rent is equal to the contracted capacity (MW) times the difference between the energy prices at nodes (or single price areas) j and k. This difference may be negative, thus implying a payment instead of a credit[7]. A physical transmission right entitles the owner to use the network between the two considered points or areas, so that it can inject electricity at one node and retrieve it at the other [10, 28]. Interestingly, according to this definition, a physical transmission right makes the collection of congestion charges by the right owner conditional upon whether it actually carries out the corresponding physical transaction[8]. Apparently, physical rights do not give the owner of the rights any additional economic advantage, since owners of financial rights might as well sell the power at one node and

---

[7] In some regulatory contexts the owner of a line is entitled to collect any congestion rents that accrue from the use of the line, regardless of the direction of the flow. In this case the "congestion rents" will be always positive.

[8] There are no universally accepted definitions for these concepts since different specific implementations create new definitions. In an alternative definition, both financial and physical transmission rights allow the owner to collect the corresponding congestion rents. On top of the right to collect (or pay) the congestion rents, a physical transmission right would entitle the owner to use the network between the two considered points or areas, so that it can inject electricity at one node and retrieve it at the other.

purchase it at the other node, with the congestion rent providing a perfect hedge against the price difference.

Ownership of a physical transmission right makes sense in two instances. The first one is precisely congestion management: In the short-term the system operator has to make sure that the ensemble of transactions results in a feasible pattern of flows. One way of achieving this is forcing the agents involved in each transaction across single price areas to have the corresponding physical transmission right, either as an option or an obligation (preferably). Only agents with the physical capacity to inject or withdraw power can be allowed to own physical transmission rights. One can see that the difference between physical and financial rights is very subtle, since financial rights can also be forced to be "dispatchable", i.e. compatible with the physical limitations of the network, see [5]. The main difference is that financial rights do not imply anything in terms of physical use of the grid and, therefore, do not constitute a valid option in the very short term unless a final integrated energy and capacity market is run afterwards.

The second instance refers to the need for physical transmission rights in the long term, which becomes apparent only when they are associated with firm energy contracts, i.e., contracts where the producer provides a physical guarantee of supply to the consumer. In the case of a power shortage (otherwise the guarantee of supply can be easily obtained via third parties), this physical guarantee is only possible if the generator has the ability to actually produce and if the contract includes a physical transmission right between the producer and the consumer. Otherwise, supplying the demand in the contract would probably involve redispatching other generation units. When the provision of long-term security of supply entails an economic compensation, the economic value of physical transmission rights may increase beyond the expected value of congestion rents.

In the proposed implementation of congestion management for the IEM, the capacity rights that are auctioned in the long-term can be financial, -although they should be dispatchable in principle, since they will have to become physical in the short-term-. In fact, the owner of a long-term financial transmission right between nodes k and j can always exchange it for the same amount of the physical right between the same two nodes, in the short term market. In order to acquire them he only needs to bid very high in the short-term explicit auction for physical rights. The amount of previously negotiated financial rights should be "dispatchable" so that there is no possible conflict if owners of financial rights want to exchange all of them for physical rights in the short-term. In the context of the explicit mechanism proposed here, it is necessary that all rights that exist after the last capacity auction are physical. In other words, agents must be committed to carry out their physical transactions and make use of their rights if we want to optimize the use made of the interconnection capacity and also ensure that the final dispatch is feasible.

4.2 WHICH MARKET AGENTS SHOULD BE INVOLVED IN CONGESTION MANAGEMENT?

This paper has assumed without discussion that the agents that must take part in the management of network congestions are those that engage in transactions (bilateral contracts or purchases from or sales to power exchanges) that "cross the borders" of single price areas. Remember that "single price areas" are an artificial construction based on the assumption that serious congestions that need to be considered by a multinational congestion management scheme only happen between them. Each single price area can be approximately considered as a single node for congestion management purposes. As indicated before, single price areas do not need to coincide with political borders or with the control areas that system operators use to monitor and control the scheduled flows at the interconnections and the frequency of the system. It is convenient, however, that they coincide.

It follows the reasoning that supports our preference for Model III and the choice of the transactions that "cross the border" between areas as the only ones to be involved in congestion management:

- With nodal pricing all transactions are implicitly involved. In what follows, we assume that the simplification of using single price areas is a valid one, it is accepted and used.
- We also assume that the agents act with economic rationality and that they have complete information about the opportunities to buy from and sell to all the agents in the system. In this case, two consequences follow:
    i) The flows over the lines, and therefore the network congestions, do not depend on the particular commercial transactions, as any set of commercial transactions would result in the same final dispatch of generation and demand (any accepted set of commercial transactions would be optimal).
    ii) The same conclusion also applies to the flows crossing the interconnectors between control areas. These flows are controlled by the system operator of each control area, who must make sure that the net outflow from the control area is equal to the global imbalance between production and consumption of all the scheduled commercial transactions. According to the assumption above, this imbalance should not depend on the specific

commercial transactions that have been agreed. Any other economically justified set of commercial transactions would have yielded the same net imbalance.

- Congestions do occur and some transactions have to be curtailed or modified in order to keep the flows within secure limits. Which transactions or agents should be affected by these changes? According to what has been discussed above, all market agents should contribute to the solution if a global nodal pricing scheme is used to handle the complete energy market. However, this is not by far the situation in the IEM.
- The pragmatic solution adopted in this paper implies defining a subset of agents and transactions that can solve the problem while leaving the remaining agents and transactions undisturbed. The chosen subset consists of all the commercial transactions (bilateral contracts and bids to power exchanges) that involve agents located in different single price areas (i.e. transactions that "cross" the borders of single price areas). All remaining transactions stay within a single price area and therefore, according to the underlying assumption, they do not contribute to congestions. Therefore, the chosen subset must be sufficient to manage congestion on the grid (located between areas).
- The choice of the subset of transactions to be considered has an effect on the amount of congestion costs that each market agent has to pay. This choice must be in accordance with the place where each market agent decides to trade its energy. Each market agent has to pay the congestion costs associated with the transaction between the place where it is located and the one where it buys or sells its energy. Thus, the market agents' decision on where to trade their energy determine the subset of market agents to be involved in the congestion management mechanism.
- But, what about the economic consequences for generators j and m of being chosen or not to solve the network constraints? If j is chosen it will have to pay the congestion fee, while generator m will not. And conversely. However, under a sound regulatory scheme in the IEM, the economic consequences are exactly the same in both cases. Why? Assume there are only two areas A and B. If the interconnection between A and B is congested, energy prices p_A and p_B in A and B will differ. Suppose generators j and m are located in area A. If generator j, bidding at the power exchange of area B, is chosen to manage the congestion, it will pay a congestion charge that amounts to $p_B - p_A$ and will sell its power at a price p_B. Accordingly, agent j will face a net selling price of p_A. This is the same price that generator m will obtain by selling its power within area A and, therefore, not being involved in the management of

congestion. This result can be easily generalized to include bilateral contracts and more than 2 areas.

In conclusion, choosing the transactions that cross the borders of the SPAs as the only ones to be involved in the congestion management scheme of Model III is legitimate. It makes intuitive, physical and economic sense. It is not discriminatory under a sound regulatory system and minimizes the involvement of market agents in the congestion management procedures.

4.3 ASSIGNMENT OF TRANSMISSION CAPACITY TO THE SEVERAL EXPLICIT AUCTIONS

We now want to answer the question of whether the total amount of scarce transmission capacity should be up for auction in the long term or a fraction of it should be reserved for the short-term.

There are two main issues for discussion: the influence that the decision could have on the prices of transmission capacity and the difficulty in "slicing" the transmission capacity in practice.

The first concern is the effect that the decision on how much transmission capacity to auction in the long run (perhaps in several instances), and how much to reserve for the short run, could have on the resulting prices of transmission capacity. We should always keep in mind that, when bidding for transmission rights, agents are predicting the difference in prices in the short term energy market between the two nodes or areas the right is defined between. Any significant difference between transmission prices in the long and the short run will be arbitraged. Therefore, unless market agents are ill informed about the likely evolution of market prices in the future or, for some reason, there is a shortage of agents who are willing to participate in the auctions, the prices of transmission rights should be neither much higher, nor much lower, than the corresponding congestion rents.

This helps in deciding how to slice the capacity among the different auctions. According to equations (21) and (22) in Model III, we should be able to determine how much of the total capacity of each line must be up for sale at each auction. Fortunately, if the value put on transmission capacity is stable, constraining more some parts of the network than others when slicing the transmission capacity will not have a dramatic impact on capacity prices nor will it significantly alter the physical use made of the grid. Given that the same product (the transmission capacity over a line for a specific hour or number of hours) is auctioned in the long and short term, transmission capacity price for an hour should always be the same regardless of when the auction takes place. For this reason, the system operator could slice transmission capacity so that this price remains as stable as possible. In other words, if the value of transmission capacity over a line in a long-term auction is higher than in the short term, some of the capacity reserved for the short term should already be made available in the long term. Correspondingly, when capacity prices are higher in the short term, some more capacity should be put on sale for the first time just before the energy market. A general recommendation, already expressed in section 3.1.1, is that the system operator should in any case be conservative when auctioning capacity in the long and medium-term. A security capacity margin should be always available in the short-term auction just in case the conditions of operation of the system change unexpectedly.

Market power issues, in relation with management of transmission constraints, is a subject that is beyond the scope of this paper. There is a great amount of academic literature [19, 24] which examines the implications that the election of one congestion management method or the other may have on the incentives that agents with market power will have to exercise it.

In conclusion, it seems advisable to start auctioning slices, or maybe the totality, of transmission capacity in the long and medium term. This will help the market agents to hedge against the volatility induced by the appearance of transmission constraints and to establish long and medium term supply contracts. Transmission capacity would be progressively auctioned till everything is available to the market just before the short-term energy market takes place.

4.4 ECONOMIC IMPLICATIONS OF COMPLEX OR UNUSUAL CONSTRAINTS

As indicated before in this paper, not all network related constraints are "congestions" and some of them cannot be expressed mathematically in the same format as equations (21) and (22) in Model III. This section will examine these "non-congestion" constraints and the following two questions will be specifically addressed:

- Who should be considered responsible for the cost of these restrictions?
- Who should receive the congestion rents or the proceedings from the auctions?

It follows a new formulation of Model III which includes a new type of constraint, -a reliability constraint. For the sake of simplicity, all transactions in the model are balanced.

Model IV:

$$\max \sum_j B_j p_j$$

$$s.t.$$

$$-\overline{F_k}^- \leq \sum_j PTDF_{j,k} p_j \qquad ,\gamma_k \qquad \forall k = 1,2,..\ number\ of\ corridors \qquad (26)$$

$$\sum_j PTDF_{j,k} p_j \leq \overline{F_k}^+ \qquad ,\mu_k \qquad \forall k = 1,2,..\ number\ of\ corridors \qquad (27)$$

$$p_j \leq \overline{P_j} \qquad ,\beta_j \qquad \forall j = 1,2,..\ number\ of\ restrictions \qquad (28)$$

$$p_j \geq 0 \qquad ,\alpha_j \qquad \forall j = 1,2,..\ number\ of\ restrictions \qquad (29)$$

$$fiab(p,\overline{F}) \leq 0 \qquad ,\xi \qquad reliability\ restriction \qquad (30)$$

The reliability restriction represents a complex relationship among the magnitude (power) of some of the bilateral transactions (p), and the power flow limits for some of the lines in the system ($\overline{F}$). The following expression can be derived (see Annex A).

$$\xi \sum_j \frac{\partial fiab(p,\overline{F})}{\partial p_j} p_j = -\xi \sum_k \frac{\partial fiab(p,\overline{F})}{\partial \overline{F_k}} \overline{F_k} \qquad (31)$$

The left hand side of equation (31) represents how much each one of the transactions j must pay because of its responsibility in causing the reliability constraint to be binding.

The right hand side of the same equation represents how much the owner of each one of the lines k (or the owner of the corresponding transmission rights) should receive under the concept of "reliability constraint rent".

One must be aware that some simplifications have been introduced in Annex A when deriving equation (31). In our model, power transfer distribution factors (PTDFs) of the flows over the lines in the system with respect to the power injection at each node do not change when the capacity of a line is marginally increased. Actually, given a generation and load pattern, the distribution of flows over the grid would vary if a line is reinforced since its impedance would change also. Some additional terms should therefore be included in equation (31) and the problem would become non-linear and more difficult to handle. As a consequence, the value of marginally increasing the capacity of a line is not equal to the sum of the shadow prices associated with the restrictions on the amount of power flowing over this line. In any case, when jointly taking into account the results arising from recent studies (see annex A) and those presented in Annexes A and B, we can conclude that the result expressed in equation (31) remains valid. Annex A also assumes that the expression of the reliability constraint is linear.

These assumptions may be a bit optimistic about the real functioning of the system. As a matter of fact, non-linear effects may be important. Then, we should distinguish between payments and charges related to each constraint in the system. If payments and charges obtained according to the marginal theory are not the same for a particular constraint, additional uniform charges levied on market agents at system level could be used to reconcile both.

5 SUMMARY AND CONCLUSIONS

The adoption of a specific scheme of congestion management is not independent of the specific regulatory context of the power system where it will be implemented. This is even more critical when different regulations are involved, as it is the case of regional or multinational markets.

This paper has examined different possibilities for the implementation of a congestion management scheme in the Internal Electricity Market (IEM) of the European Union. A conceptually ideal model consisting of a joint energy and capacity auction totally equivalent to an optimal power flow serves as a reference since it is the most efficient approach. Given that it is not viable in the present European context, two alternative approaches are tested by comparing their efficiency and technical feasibility against the reference model, and also by examining its political acceptability under the present European and national regulations. The first one is an integrated transmission and energy auction where each market agent decides where it wants to trade its energy. Despite giving agents more flexibility it seems still far from being applicable in Europe.

The second approach consists of a coordinated explicit auction of transmission capacity followed by separate energy auctions at the different power exchanges, subject to the "use-it" rule. This is the one that this paper recommends. A number of possible variations on this basic approach may overcome some of its possible shortcomings, like the lack of flexibility. Options could be employed instead of obligations for instance. Another possibility would be to define larger SPAs. A European-wide balancing market could then solve any infeasibilities. In the end, TSOs will have to cope with any minor unfeasibility problems by introducing any required adjustments by redispatch or counter-trade. This approach mostly preserves the operational independence of the power exchanges, which will continue functioning basically as most of them are doing now, i.e. without paying much regard to network constraints. The potential loss of economic efficiency appears to be small. A complete conceptual mathematical formulation of the model has been provided, although the paper indicates that some variations are also possible in this respect.

The paper also provides guidelines concerning some of the practical problems to be found during the implementation of the method: regarding the type of transmission rights to be auctioned we suggest point-to-point rights which could be financial in the long-term and should become physical after the last capacity auction. Both options and obligations are possible although we make a final choice for obligations. With respect to the division of transmission capacity into slices to be auctioned in several auctions with different time horizons, we propose auctioning part of the transmission capacity in the long term and putting the rest of it up for sale progressively till everything is available in the short-term. A fraction should be always reserved for the short-term in case the conditions change unexpectedly.

However, the successful application of the proposed approach, as well as of the other alternatives discussed in this paper, depends on whether single prize areas can be defined within the IEM that comply reasonably well with the specified requirements. Validation of this hypothesis therefore remains an issue of utmost importance.

ACKNOWLEDGEMENTS

This work is part of a long-lasting effort on transmission pricing and regulation at the IIT from the Comillas University in Madrid. The authors are in debt for the comments and discussions provided, before and during the realization of this paper, by fellow researchers at IIT: Carlos Vázquez, Michel Rivier, Fernando Perán and F. Javier Rubio-Odériz. The paper has also greatly benefited from the interaction with a number of people who are involved in the discussions on the design of the Internal Electricity Market in the European Union, specially Carlos Solé and Alberto de Frutos (CNE, Spain), Carlos Sallé (Iberdrola, Spain), María Luisa Llorens (REE, Spain), Pierre Bornard and Damian Cortinas (RTE, France), Michel Massoni and Damian Caby (CRE, France), Yves Smeers (University of Louvain, Belgium), Matti Supponen (GD TREN, EU Commission), Rudi Hakvoort (Delft University, The Netherlands) and Michael Metcalf (NGC, UK).

REFERENCES

1. O'Neill, R.P., et al., *The Joint Energy and Transmission Rights Auction: A General Framework for RTO Market Designs*. 2001, Office of Markets, Tariffs and Rates (OMTR), Federal Energy Regulatory Commission.

2. *Co-ordinated Auctioning. A market based method for transmission capacity allocation in meshed networks*. 2001, ETSO (European Transmission System Operators). Web page: http://www.etso-net.org/.

3. Henney, A., *What the US Could Learn from Western Europe and Elsewhere.* Electricity Journal, 2002. **15**(10): p. 53-64.

4. Boucher, J. and Y. Smeers, *Towards a common European Electricity Market-Paths in the right direction... still far from an effective design*. 2001, Web page: http://www.ksg.harvard.edu/hepg/Standard_Mkt_dsgn/Smeers_Interconnections1_4jni_3.do1.pdf: Brussels.

5. Hogan, W.W., *Financial Transmission Right Formulations*. 2002, Center for Business and Government, John F. Kennedy School of Government, Harvard University: Cambridge. http://www.ksg.harvard.edu/hepg.

6. Ruff, L.E., *Competitive Electricity Markets: Why They Are Working and How to Improve Them*. 1999, n/e/r/a: http://www.ksg.harvard.edu/hepg.

7. Chao, H.P., et al., *Flow Based Transmission Rights and Congestion Management.* The Electricity Journal, 2000. **Oct.**: p. 38-58.

8. Stoft, S., *Congestion Pricing with Fewer Prices than Zones.* The Electricity Journal, 1998. **11**(4): p. 23-31.

9. Tabors, R. and M. Caramanis, *Real Flow, A Preliminary Proposal for a Flow-based Congestion Management System*. 2000, Web page: http://www.ksg.harvard.edu/hepg/flowgate/Real%20Flow.pdf: Cambridge, MA.

10. Chao, H.-P. and S. Peck, *A Market Mechanism for Electric Power Transmission.* Journal Of Regulatory Economics, 1996. **10**: p. 25-29.

11. Schweppe, F., et al., *Spot pricing of electricity*. 1988: Kluwer Academic Publishers.

12. *Notice of Proposed Rulemaking, Remedying Undue Discrimination through Open Access Transmission service and Standard Electricity Market Design*. 2002, Federal Energy Regulatory Commission: Washington.

13. *Reglamento Transitorio del Mercado Eléctrico Regional*. 2002, Comisión Regional de la Interconexión Eléctrica, Web page: http://www.omca.net/documentos/Rtmer01.pdf.

14. Hogan, W.W., *Contract Networks for Electric Power Transmission.* Journal of Regulatory Economics, 1992: p. 211-242.

15. Christie, R.D. and I. Wangensteen, *The Energy Market in Norway and Sweden: Congestion Management.* IEEE Power Engineering Review, 1998: p. 61-63.

16. *Evaluation of congestion management methods for cross-border transmission*. 1999, ETSO (European Transmission System Operators). Web page: http://www.etso-net.org/.

17. Rivier, M. and J.I. Pérez-Arriaga, *Computation and decomposition of spot prices for transmission pricing.* 11th PSCC Conference, 1993.

18. Singh, H., S. Hao, and A. Papalexopoulos, *Transmission Congestion Management in Competitive Electricity Markets.* IEEE Transactions on Power Systems, 1998. **13**(2).

19. Neuhoff, k., *Integrating Transmission and Energy Markets Mitigates Market Power*. 2002, Cambridge, DAE. Web page: http://www.econ.cam.ac.uk/electricity/publications/wp /index.htm.

20. Wood, A.J. and B.F. Wollenberg, *Power Generation, Operation and Control*. 1984, New York: John Wiley & Sons.

21. *Reconciliation of market splitting with coordinated auction concepts*. 2002, European Transmission System Operators, http://www.etso-net.org/.

22. *Open Access Technology International (OATI). Web page: http://www.oatiinc.com*.

23. Rubio, F.J. and J.I. Pérez-Arriaga, *Marginal Pricing of Transmission Services: A Comparative Analysis of Network Cost Allocation Methods.* IEEE Transactions on Power Systems, 2000. **15**(1): p. 448-454.

24. Gilbert, R., K. Neuhoff, and D. Newbery, *Allocating Transmission Capacity to Mitigate Market Power in Electricity Networks*. 2002, Cambridge-MIT. Web page: http://www.econ.cam.ac.uk/electricity/.

25. *Regulation of the European Parliament and of the Council on conditions for access to the network for cross-border exchanges in electricity*. 2003. Web page: http://europa.eu.int/comm/energy/electricity/legislation/amending_legislation_en.htm.

26. *Using implicit auctions to manage cross-border congestion: Decentralized market coupling*. 2003, EuroPEX (European association of Power EXchanges) at the 10th Meeting of the European Electricity Regulatory Forum. Web page: http://www.europex.org/pages/943372/index.htm: Rome.

27. *Analysis of Electricity Network Capacities and Identification of Congestion*. 2001, Institute of Power Systems and Power Economics (IAEW) and CONSENTEC Consulting fur Energiewirtschaft und -technik. Report for the European Commission, Directorate-General Energy and Transport: Aachen.

28. Ruff, L.E., *Flowgates vs. FTRs and options vs. Obligations*. 2000: http://www.ksg.harvard.edu/hepg.

29. Joskow, P.L. and E. Kahn, *A Quantitative Analysis of Pricing Behavior in California's Wholesale Electricity Market During Summer 2000.* The Energy Journal, 2002. **23**(4): p. 1-35.

30. *Key Concepts and Definitions for Transmission Access Products*. 2001, ETSO (European Transmission System Operators). Web page: http://www.etso-net.org/.

ANNEX A

The Lagrangian of the optimization problem that has been termed Model III is:

$$L = -\sum_j B_j p_j + \sum_k \gamma_k \left(\sum_j -PTDF_{k,j} p_j - \overline{F_k}\right) + \sum_k \mu_k \left(\sum_j PTDF_{k,j} p_j - \overline{F_k}\right) + \sum_j \beta_j \left(p_j - \overline{P_j}\right) + \\ + \sum_j -\alpha_j p_j + \xi\, fiab(p, \overline{F}) \tag{32}$$

Thus, according to the Karush-Kuhn-Tucker optimality conditions:

$$\frac{\partial L}{\partial p_j} = -B_j - \sum_k \gamma_k PTDF_{k,j} + \sum_k \mu_k PTDF_{k,j} + \beta_j - \alpha_j + \xi \frac{\partial fiab(p, \overline{F})}{\partial p_j} = 0 \tag{33}$$

The relationship between the bid of agent *j* and the dual variables of the different constraints is:

$$B_j = \overbrace{-\sum_k \gamma_k PTDF_{k,j} + \sum_k \mu_k PTDF_{k,j}}^{\text{associated with restrictions on the maximum line power flow}} + \overbrace{\beta_j - \alpha_j}^{\text{associated with the restriction on the maximum amount of power agent j wants to transact}} + \overbrace{\xi \frac{\partial fiab(p, \overline{F})}{\partial p_j}}^{\text{associated with the feasibility constraint}} \tag{34}$$

where α_j and β_j represent the difference between agent *j's* bid and the price agent *j* must pay for each MW of power it is allowed to transact because of the line and reliability constraints.

The total volume of payments from all agents as a result of the auction is:

$$\sum_j C_j p_j = \sum_j p_j \sum_k -\gamma_k PTDF_{k,j} + \sum_j p_j \sum_k \mu_k PTDF_{k,j} + \sum_j p_j \xi \frac{\partial fiab(p, \overline{F})}{\partial p_j} \tag{35}$$

where C_j is the price that agent *j* must pay per MW of its transaction p_j.

An alternative expression can be obtained for the first and second terms on the right side of equation (35):

$$\sum_j p_j \sum_k -\gamma_k PTDF_{k,j} = -\sum_j \sum_k p_j \gamma_k PTDF_{k,j} = -\sum_k \gamma_k \underbrace{\sum_j p_j PTDF_{k,j}}_{-\overline{F_k}\ \text{if}\ \gamma_k \neq 0} = \sum_k \gamma_k \overline{F_k} \tag{36}$$

$$\sum_j p_j \sum_k \mu_k PTDF_{k,j} = \sum_k \mu_k \underbrace{\sum_j p_j PTDF_{k,j}}_{\substack{\overline{F_k}\ \text{if}\ \mu_k \neq 0 \\ \text{(the line is congested} \\ \text{in the negative direction)}}} = \sum_k \mu_k \overline{F_k} \tag{37}$$

The revenue that the owner of a line k should receive as a result of the auction per MW of capacity of the line (either because the line is congested or because the reliability constraint depends on it) is, according to marginal pricing principles:

$$VM_k = -\frac{\partial L}{\partial \overline{F_k}} = -\left(-\gamma_k - \mu_k + \xi \frac{\partial fiab(p, \overline{F})}{\partial \overline{F_k}} \right) = \gamma_k + \mu_k - \xi \frac{\partial fiab(p, \overline{F})}{\partial \overline{F_k}} \tag{38}$$

As mentioned in section 4.4, we have assumed here that the power transfer distribution factors do not change when the capacity of the line marginally increases, which is not strictly true. However, if we consider the expression that relates the capacity of a line and its admittance to be linear and the intercept is zero, the marginal value of a line associated with the restrictions on its power flow is the difference of nodal prices between both ends of the line times its flow. Then, Annex B shows that the total amount of congestion rents to be paid to line owners because of these restrictions would be, according to the aforementioned expression for the marginal value of a line, equal to the first two terms of the right hand side of equation (39), i.e. they would be coincident with those computed assuming PTDFs independent of the change in the capacity of a line. Both assumptions are approximations. In reality, line impedances vary with an increase in the transmission capacity of the line but the impedance change is not a linear function of the change in the capacity. Whether one approximation or the other is better remains to be proved.

Finally, the total amount of revenues that must be paid to the owners of all the lines in the system is:

$$\sum_k \overline{F_k} \frac{\partial L}{\partial \overline{F_k}} = \sum_k \overline{F_k} \gamma_k + \sum_k \overline{F_k} \mu_k - \xi \sum_k \frac{\partial fiab(p, \overline{F_k})}{\partial \overline{F_k}} \overline{F_k} \tag{39}$$

Since the total volume of revenues must equal the total volume of payments, it follows that:

$$\xi \sum_j \frac{\partial fiab(p, \overline{F})}{\partial p_j} p_j = -\xi \sum_k \frac{\partial fiab(p, \overline{F})}{\partial \overline{F_k}} \overline{F_k} \tag{40}$$

ANNEX B

We consider here a DC model of the grid and disregard losses, as we have done in the rest of the paper. This annex proves that the total congestion payments due to transmission right owners that are related to the set of constraints on the flow over the lines of a system are equal to the sum over the set of lines in the system of the marginal value of the capacity restrictions for each line (i.e. the dual variable of the corresponding constraint) times the capacity of the line.

Equation (41) holds for every congested line k:

$$\overline{F_k} = \sum_i PTDF_{ik} p_i \tag{41}$$

According to traditional nodal pricing equations, see [11]:

$$\Delta\rho_i = \sum_k PTDF_{ik} (\gamma_k + \mu_k) \tag{42}$$

where $\Delta\rho_i$ is the difference between nodal prices at the withdrawal and injection points for transaction *i* caused by congestion in the grid and γ_k and μ_k are the dual variables of the restrictions to the flow over line *k*. Furthermore, if we agree that the marginal value of a line equals the difference of nodal prices between both ends times its flow, the total payments owed to the owners of the lines due to the existence of restrictions on their flows is:

$$cong\_payments_1 = \sum_i p_i \Delta\rho_i \tag{43}$$

By substituting (42) in (43) we obtain the following expression for these congestion payments:

$$cong\_payments_1 = \sum_k ((\gamma_k + \mu_k) \sum_i p_i PTDF_{ik}) \tag{44}$$

Additionally, the sum over all the lines in the system of the shadow price of the restrictions limiting the flow over each line times the capacity of the line is:

$$cong\_payments_2 = \sum_k (\gamma_k + \mu_k) \overline{F_k} \tag{45}$$

Finally, by substituting (41) in (45) we obtain:

$$cong\_payments_2 = \sum_k (\gamma_k + \mu_k) \sum_i p_i PTDF_{ik} \tag{46}$$

Therefore, when all constraints limit the flow over individual lines, total congestion payments to all lines can be equivalently computed with both formulations.