Preliminary Draft

# A Spatial "Smart Market" for Electric Power and Transmission

by

Jeffrey K. MacKie-Mason

*University of Michigan, University of California Energy Institute and NBER*

July 1994
Current version: March 15, 1995

**Abstract.** Creating successful competitive electric power markets requires transmission pricing reform. Due to the physics of power transmission networks, efficient transmission prices must incorporate two types of loop flow externalities, and must be determined jointly with efficient power (nodal) prices. Prior proposed market schemes all involve an approximation to the efficient prices. I propose a real-time "smart market" that provides incentives for truthful revelation of the information required for finding the correct efficient prices.

**Address.** Prof. Jeffrey K. MacKie-Mason, Dept. of Economics, University of Michigan, Ann Arbor, MI 48109-1220. E-mail: jmm@umich.edu.

# A Spatial "Smart Market" for Electric Power and Transmission

Jeffrey K. MacKie-Mason

## 1. Efficient Short-Run Pricing

The entire institutional and market structure of the electric power industry is in turmoil in many countries. The U.K. has deregulated the power market, and created a central transmission pool (Newberry and Green (1995)). New Zealand has been implementing short-run marginal cost pricing for power (Ring (1994)). Regulators at both state and federal levels in the U.S. have been studying radical transformations. For example, the California PUC recently proposed retail competition among power providers (Blumstein and Bushnell (1994)).

Transmission pricing reform is essential to the creation of competitive power markets. A power system consists of supply (generator) and demand (load) nodes, connected by a transmission grid. By delivering power over the grid, multiple suppliers can compete to serve a given demand. However, building a transmission grid involves huge fixed costs relative to variable operating costs, and thus for the forseeable future crucial components of a grid will be natural monopolies. The regulation and pricing of transmission access will determine the success of power market competition.

Transmission grids are extraordinarily complex engineering systems. The delivery of power from one node to another may involve flows along several "parallel" paths; the actual flows are determined by physics, not economics or contractual arrangements. As a result, the flows induced by one power transaction may affect the costs incurred by others on the grid who are not participants in the transaction. Two types of external effects are common: flows induced by a transaction may (1) raise marginal transmission losses for other grid users; and (2) raise marginal generation costs if transmission line constraints become binding.

Power grid physics require that efficient prices for both power and transmission be found jointly as the solution to an optimization problem. The extent to which a regulatory reform obtains

I have benefitted from conversations with Ed Kahn, Bart McGuire and Matthew White. I undertook the research for this paper while visiting the California Energy Institute, University of California, Berkeley; I am grateful for the financial support provided by UCEI.

the benefits of an efficient power market depends on the closeness of the resulting transmission and power prices to the efficient solution. Numerous proposals for developing power markets that approximate the efficient solution have been made; two of the better known are Schweppe, Caramanis, Tabors, and Bohn (1988) and Hogan (1992).

In this paper I review the basic economics of efficient pricing for power and transmission, with a focus on short-term wholesale transactions. I emphasize the steep informational requirements for efficient pricing, especially in the wholesale markets that stand between the many non-integrated "local" grids. I then propose a feasible market mechanism that theoretically obtains the fully efficient set of power and transmission prices. The mechanism is based on a real-time "smart market" auction. The auction is designed to elicit truthful revelation of the information necessary to determine the efficient set of prices.

One useful role for the smart market model is to provide a meaningful benchmark against which to compare various approximate approaches to efficient pricing. The approach is useful because the smart market can feasibly be implemented: comparisons could be made through simulation studies. Since the smart market feasibly achieves the first-best outcome, comparisons would indicate how large are the efficiency gains potentially foregone by various approximation methods, and thus provide guidance on where it is worthwhile trying to improve these approximations.

For example, the quality of Hogan (1992)'s "contract network" approximation method depends on the availability of substantial (often private) information on costs and purchaser valuations of power. In his discussion of the availability of that information, Hogan assumes that the market structure will operate in the form of a tightly integrated pool, with control of the grid extending to the level of the individual generator. In fact, in many cases market reform is likely to occur first at the wholesale level, with the relevant transmission grid involving connections between various switchyards that provide tie lines to non-integrated utility grids. The information requirements for Hogan's approximate pricing method are much more severe when nodes connect not to specific loads or generators, but to other grids. The smart market mechanism I propose below is sufficiently general to induce revelation of the required information without much incremental difficulty.[1]

[1] Indeed, all or nearly all of the required information is already known to the participants, so revelation does not require additional effort on their part. The problem is to *induce* truthful revelation of the privately-held information.

Smart markets may increasingly be important not just as a hypothetical optimum, but as a practical alternative for actual power markets. As computing power rapidly becomes less expensive, it rapidly becomes more likely that a smart market would be not only feasible, but economically practical. Sooner or later it will make sense to implement smart markets in fact, rather than market structures that merely approximate the fully efficient outcome.

## 2. An Electric Power and Transmission System

An electric system essentially consists of three types of entities: generators, loads (sources of demand), and a transmission network or grid. The grid is represented by a constrained graph, $G(N, T, \Gamma)$, specifying a set of nodes, $i \in N$, connected by transmission lines, $(i, j) \in T$, with line flows $z_{ij}$. The vector of all flows is given by $z$. Flows are constrained by one or more physical restrictions, represented by vector functions $\gamma(z) \in \Gamma$. For example, a typical high-voltage power line is constrained by both minimal and maximal permitted flows.[2] The grid as a whole is constrained to be in energy balance, with the sum of power injections equal to the sum of extractions plus line losses.[3]

Each node or bus can support injections from power suppliers and extractions from power demanders. Injections and extractions are usually referred to as generators and loads in the literature, but I use the more general supply and demand terminology below because in a wholesale grid a node may be a tie-line bus that connects to another network. Such nodes can have power flows to or from the other grid, and are thus not specifically identifiable as a particular generator or load.

Let the set of buyers and sellers be given by $B$. Each agent $b$ can inject or extract real power $y_{ib}$ at node $i$; power $y_{ib}$ is positive if it is injected, and negative if it is extracted. An agent $b$ receives

[2] The constraints are usually not hard and fast, the way that a pipeline's diameter is a fixed physical constant. A line with a design maximum of 1000 mW, for example, will be able to carry more than 1000 mW, and will not instantly fail if called upon to do so. However, standard network engineering practice dictates that thermal limits and others be treated as hard constraints, and I will do so in the analysis below.

[3] The power flow balance equations in an alternating current network have both real and complex components. The complex component is known as "reactive" power; the reactive power balance gives rise to constraints in the form of voltage magnitudes. I will follow the common practice of ignoring the reactive power contraints in the model, which is equivalent to assuming that they are never binding. However, it probably is not appropriate to ignore reactive power constraints in the design of an actual network dispatch mechanism; see Hogan (1993) and Kahn and Baldick (1994) for a debate on the importance of these constraints. The model in this paper can be generalized easily to include reactive power contraints.

a gross benefit (before any financial transactions) from $y_{ib}$ equal to $u_b(\vec{y}_b)$, where $\vec{y}_b$ is the vector of $b$'s power flows at all $i \in N$ nodes, and $u_b(\cdot)$ is assumed to be concave.[4] The benefit function is denominated in monetary units, and thus can be thought of as the participant's willingness to pay for power received, or willingness to sell for power provided.

Notice that benefits are written as a general function of all of the power injections and extractions in which the agent participates; this allows for the possibility that the value of injecting $y_{ib}$ at node $i$ depends on how much power $y_{jb}$ the agent injects at node $j$. The computation of an efficient network allocation would be simpler if no such interactions were permitted, which would be equivalent to assuming that agents are independent and participate at no more than one node. However, the more general situation is quite reasonable for a wholesale market, as will be demonstrated in Section 4 below.

Let us now explicitly specify some of the line flow constraints, generically represented above as $\gamma(Z)$. To begin with, we can dispense with constraints on the paths that particular flows follow through the network. Power flows in a grid are homogeneous. Thus, participants care only about their own power extractions and injections, not where they originate or terminate.[5] Indeed, it is generally not possibly to route electrons along a particular origin/destination path.[6] Therefore, there are no routing constraints that must be satisfied in an electrical network. There is, however, an energy balance contraint: total injections must equal total extractions plus total line losses:

$$\sum_i \sum_b y_{ib} - L(Z) = 0 \qquad \text{(energy balance)} \tag{1}$$

where $L(Z)$ is a function describing the total losses in the system as a function of line flows. Using a direct current (DC) load flow approximation (which ignores reactive flows), Caramanis, Bohn,

[4] I am implicitly assuming here that participants have quasiliniear utility, $U_b = I_b + u_b(\vec{y}_b)$, with $I_b$ equal to the budget available for other purchases. In this case, there are no interactions between the value they receive from their power flows and any other economic transactions they may undertake. This seems reasonable for a wholesale market.

[5] Assuming that participants care only about homogeneous extractions or injections conflicts with current *institutional* arrangements in many regional grids. Wholesale power transactions are often arranged through bilateral contracts betweeen buyers and sellers located at separate nodes. I am assuming that there is no intrinsic reason to care about the identity of one's trading partner, given the homogeneity of power on the grid and the impossibility of directing the path that given electrons follow. If there are external benefits from bilateral agreements to inject and extract wholesale power, I assume that side contracts can implement the benefits of such agreements while the actual power flows are bought and sold at each node in transactions with the "pool" or node manager.

[6] Phase shifters can be used to regulate the amount of power that can flow along certain links, but it is not possible to tag electrons and control their path.

and Schweppe (1982) have shown that $L(z) = L(y) = Hy$, where $H$ is a rectangular matrix function of technological parameters of the network, and $y$ is the vector of net injections (gross injections minus gross extractions) at the nodes in the network. That is, line flow losses can be expressed as a linear function of net injections.

In addition to the systemwide balance, flows on each line also face contraints. In particular, thermal limits impose a maximum flow on a line,

$$z_{ij} \leq z_{ij,max} \qquad \text{(line maximum)}. \tag{2}$$

Some lines may also be constrained to carry a minimal load (or be placed out of service),

$$z_{ij} \geq z_{ij,min} \qquad \text{(line minimum)}. \tag{3}$$

For simplicity I will ignore the other constraints that occur in an electric grid, such as voltage magnitudes and various stability parameters. I am not excluding limits on generation capacity: those are incorporated in the benefit (cost) functions of participants who supply power: a capacity constraint would imply an infinite cost of providing power in excess of that contraint. Likewise, firm system requirements for reserve power can be incorporated through the addition of an artificial participant who receives an infinite benefit from a fixed quantity of power at each node.

A vector of net injections $y$ such that conditions (1), (2), and (3) are satisfied is $feasible$. The general problem of system efficiency requires maximizing the sum of participant benefits subject to the solution being in the feasible set.

### 3. Loop Flow Externalities

It might seem that wholesale electricity pricing is a straightforward problem: calculate the marginal cost of supply (including transportion costs, in the form of line losses) at each node, and sell available capacity at that price. However, the supply curve at each node depends on the injections and extractions at every other node in the network, and the engineering constraints on the lines. This is so because the flows of electricity through a grid are determined by Kirchoff's law, which more or less requires that electrons follow the path of least resistance.[7] An additional injection or

[7] In fact, reactance and resistance together determine flows.

extraction by a participant at one bus can affect the marginal losses throughout the system, and the extent to which various line contraints are binding, as line loads adjust in conformance to Kirchoff's law. Thus, one participant's action can affect marginal cost for all others. These external impacts are known in the electric industry as *loop flow* externalities.

An example may help make the loop flow problem clear (see panel (i) in Figure 1). We can ignore line losses and focus just on line load constraints. Suppose that a firm ($B_1$) wants to purchase 500MW of power at bus $B$. Suppose that the unit cost of supply at bus $C$ is constant at 7.5, while the constant unit cost of supply at bus $A$ is 10. The generator at bus $C$ has a maximum output of 500MW. Clearly, the efficient supply solution is for the firm at $C$ to inject 500MW.

There are two transmission paths between $C$ and $B$; one path goes through bus $A$ (see the first panel in Figure (1)). The lines between each bus have identical length, and there are no losses. Assume that initially there are no other supplies or demands on this simple network. Resistance is quadratic in line flows, so Kirchoff's law can be stated as[8]

$$\min z_{ab}^2 + z_{ac}^2 + z_{cb}^2$$

$$\text{subject to} \qquad z_{ca} = z_{ab}$$

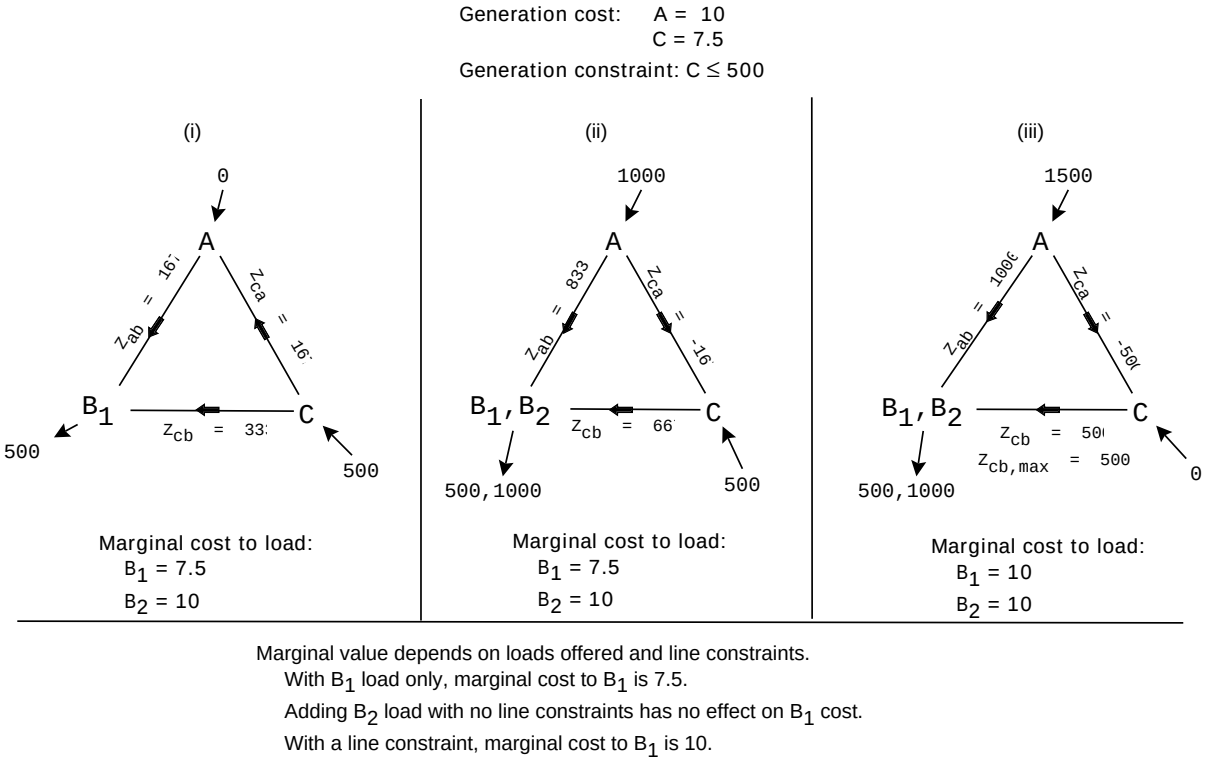
$$z_{ab} + z_{cb} = 500.$$

Since the path through bus $A$ is twice as long, its marginal resistance is twice as great. The result is a 333MW flow along the link from $C$ to $B$, with the remaining 167MW flowing along the indirect path ($CAB$). The marginal cost of generating power and delivering it to the customer at $B$ is 7.5.

Now suppose that a new user emerges, $B_2$, who wants to purchase 1000MW, also at bus $B$ (see panel (ii) in Figure 1). Since generator $C$ is already injecting its maximum output, it is necessary to inject power from the more expensive generator at bus $A$. What are the flows? Minimizing resistance subject to the efficient injections and extractions puts 833MW on the link from $A$ to $B$, with the other 167MW flowing from $A$ through $C$ to $B$; thus the flow from $C$ to $B$ is 667MW. Since $B_1$'s supplier is still injecting 500MW at the low marginal cost of 7.5, there is no external impact from $B_2$'s demand on $B_1$'s cost of power.[9] $B_2$, of course, incurs a marginal cost of 10.

[8] I am using the DC load approximation, and thus ignoring reactive power.

[9] If we accounted for line losses, there *would* be a loop flow externality in this case: the added flow on line $CB$ would raise the marginal loss, thus raising the marginal delivered power cost to customer $B_1$.




Marginal value depends on loads offered and line constraints.
With $B_1$ load only, marginal cost to $B_1$ is 7.5.
Adding $B_2$ load with no line constraints has no effect on $B_1$ cost.
With a line constraint, marginal cost to $B_1$ is 10.

**Figure 1**. Example of Loop Flow Externalities

Now suppose that there is a line load limit of 500MW on the line from $C$ to $B$. Clearly, the power flows in panel (ii) are infeasible. In order to reduce the flow on the line from $C$ to $B$, it is necessary to reduce the injection at $C$ and compensate with a larger power injection at $A$. Indeed, the feasible solution is to generate 1500MW at $A$ and nothing at $C$.[10] The marginal cost of production for both purchasers is now 10. Since $B_1$'s demand for 500MW could be provided by the generator at $C$ with the 500MW line constraint if $B_2$'s demand were zero, it is clear that the demand by $B_2$ has an external impact on the cost of power to satisfy $B_1$'s load. This result is known as "out-of-merit-order" dispatch, which is to say that a lower marginal cost generating unit is idle (or underutilized) when a higher marginal cost facility is in use, in order to jointly satisfy

[10] This is the solution to the following optimization problem that represents the application of Kirchoff's laws: $\min z_{ab}^2 + z_{ac}^2 + z_{cb}^2$, subject to $y_a + y_c = -1500$, $z_{cb} = y_c - z_{ca}$, $z_{ab} = y_a + z_{ca}$, and $z_{cb} \leq 500$.

the flow demands and the line constraints.[11]

What does the loop flow externality imply for power transactions? The answer depends on the allocation of property rights over the transmission lines in the grid. For example, consider the simplest case (based on the example in Figure 1): the generator at node $C$ owns links $CA$ and $CB$, and $CB$ has a 500MW thermal limit on its load. Then, as demand $B_2$ induces an increase in output from generator $A$ (compare panel (i) to panel (iii) in Figure 1), generator $C$ can demand compensatory payments for the higher costs induced by the use of its lines, thus placing the full incremental cost of the power on demand $B_2$. Both demanders face the incremental cost that they place on the system, and thus make efficient decisions about how much to demand.

However, transmission links are a bottleneck facility, or a natural monopoly. Thus, if the line owner had full private property rights, it could charge monopoly prices for the use of its facilities, which would induce another (different) inefficient outcome. Due to concern about monopoly pricing of bottleneck facilities, the FERC has historically regulated transmission line prices and access, thus establishing a public property right over the use of lines. Although users of a transmission line generally are required to pay *some* compensation to their private owners, the FERC has used pricing methods that bear no resemblance to the shadow (incremental cost) value of the lines. The most common pricing method is "postage stamp" pricing, which assigns a fixed charge per megawatt transported, independent of the *distance* transported (which determines the amount of induced transmission loss) and independent of loop flow considerations. Thus, a generator at node $A$ can "free-ride" on line $CA$: the new load $B_2$ may not pay its full incremental cost. Part of the cost is instead borne by the generator at $C$.

## 4. Complexity Due to Non-Integrated Local Grids

The difficulty of obtaining an efficient allocation in the wholesale grid depends on whether such an allocation creates external effects in the periphery, outside the wholesale grid. The wholesale allocation may induce external effects if a peripheral grid connects at more than one tie bus, and is

[11] When line losses are considered, they provide another source of loop flow externality. Consider the example from panel (ii) in Figure 1: the power injected at node $A$ flows in part over link $CB$, raising the total flow on $CB$ and thus raising the marginal line loss (since losses are convex in flows). Thus, $B_2$'s demand raises $B_1$'s costs even if merit-order dispatch is maintained.

subject to loop flow effects (from either or both of line losses and line constraints).[12]

These external effects can substantially complicate the participant benefit functions. Suppose participant $b$ is a power utility with its own transmission grid that connects to the wholesale grid at two nodes, $i$ and $j$. Then $b$ is in a position to buy or sell at two nodes, and has a benefit function $u_b(y_{ib}, y_{jb})$. If there were no external effects in $b$'s local grid, then the benefit derived from transactions at each wholesale node would be independent, and we could write $u_b(y_{ib}, y_{jb}) = u_{ib}(y_{ib}) + u_{jb}(y_{jb})$.

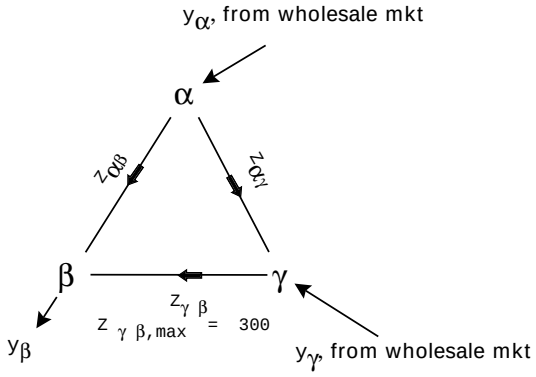



**Figure 2**. Example of Loop Flow Externalities

To make the point clear, consider the following example, illustrated in Figure 2. We are now considering a local grid that is not part of the wholesale grid. It has three nodes ($\alpha$, $\beta$, and $\gamma$); there are tie lines from the wholesale grid to nodes $\alpha$ and $\gamma$. There is a 300MW thermal constraint on line $\gamma\beta$. For simplicity, suppose the local grid has no local generating capacity, but must purchase all power from the wholesale grid. There is a single load, at node $\beta$, denoted $y_\beta$. The demand at $\beta$ is for

$$\begin{cases} \text{0 MW if } 10 < p \\ \text{600 MW if } 7.5 < p < 10 \\ \text{900 MW if } p < 7.5. \end{cases}$$

[12] There is no important distinction that divides the transmission network into a single wholesale grid and multiple peripheral (local) grids. The point is to understand the complications when the grid subject to a market allocation mechanism is multiply connected to other grids (which may themselves be allocated by market mechanisms). Thus, the analysis applies for any general pattern of interconnected grids.

Denote the injections by $(y_\alpha, y_\gamma)$. Flows are determined by applying Kirchoff's law.[13]

$$\begin{aligned} \min\ & z_{\alpha\beta}^2 + z_{\gamma\beta}^2 + z_{\gamma\alpha}^2 \\ \text{subject to}\quad & y_\alpha + y_\gamma = -y_\beta \\ & z_{\gamma\beta} = y_\gamma - z_{\gamma\alpha} \\ & z_{\alpha\beta} = y_\alpha + z_{\gamma\alpha} \\ & z_{\gamma\beta} \leq 300. \end{aligned}$$

By substituting the constraints into the minimand and solving for $z_{\gamma\beta}$ we obtain

$$z_{\gamma\beta} = (y_\gamma - y_\beta)/3 \qquad \text{subject to } z_{\gamma\beta} \leq 500 \tag{4}.$$

The other flows $(z_{\gamma\alpha}, z_{\alpha\beta})$ can be found by solving the constraints in the flow problem above.

Since there are no line losses, the local utility is indifferent between whether the power arrives at node $\gamma$ or node $\alpha$, as long as the imports do not violate the thermal limit on line $\gamma\beta$. By rearranging the result in (4) we can state the feasible bundles of imports as

$$\begin{aligned} y_\gamma \leq & 900 + y_\beta \\ y_\alpha = & - y_\beta - y_\gamma, \end{aligned}$$

expressed in terms of the total load, $-y_\beta$.

Subject to the flow feasibility conditions, we can now state the utility's willingness to pay for power as follows:

$$u(y_\alpha, y_\gamma) = \begin{cases} 10(y_\alpha + y_\gamma) & \text{for } 0 \leq y_\alpha + y_\gamma \leq 600 \\ 1500 + 7.5(y_\alpha + y_\gamma) & \text{for } 600 \leq y_\alpha + y_\gamma \leq 900. \end{cases}$$

(Given the thermal constraint it is not feasible to deliver more than 900MW to load $\beta$.)

We can see the interdependence between the valuation or wholesale transactions at the two different nodes. First, the quantity the utility will purchase at one node depends on the quantity it is able to purchase at other nodes. Second, the value at one node depends on the amount purchased at other nodes. Solving for efficient power and transmission prices in the wholesale grid requires complete knowledge of complex valuation functions for power at multiple nodes.

[13] We are minimizing network resistance, which is quadratic in line flows (DC flow model). The first constraint is the energy balance: injections equal extractions. The second constraint is the line balance on line $\gamma\beta$: the flow from $\gamma$ to $\beta$ equals injections at $\gamma$ minus net flows from $\gamma$ to $\alpha$ (if power flows from $\alpha$ to $\gamma$ then $z_{\gamma\alpha} < 0$ and this flow is added to injection $y_\gamma$ to find the total load on line $\gamma\beta$). The third constant is the corresponding line balance on line $\alpha\beta$.

## 5. Spot Pricing Solution

The efficient allocation in the wholesale grid is obtained by maximizing the sum of user valuations, subject to the engineering constraints of the network. In this section I show how such an allocation would be obtained by a central planner with full information. This solution is essentially a review of the nodal spot pricing results developed by Bohn, Caramanis, and Schweppe (1984). I then comment on different approaches to decentralizing the spot pricing equilibrium, as motivation for the auction mechanism proposed in the next section.

Let participants be indexed by $b = 1, \ldots, B$, and nodes be indexed by $i \in N$. Consider a slight modification of the bidder benefit function given above:

$$\begin{aligned} & u_b(\vec{y}_b) \\ \text{s.t. } & \underline{y}_{bi} \leq y_{bi} \leq \bar{y}_{bi} \qquad \forall i \in N \end{aligned} \tag{5}$$

The change is to specify a set of "capacity" constraints separately from the utility function. This is merely for expositional convenience: below it will help to have explicit shadow prices (Lagrangian multipliers) associated with the capacity constraints. Notice that since $y < 0$ indicates a load, or demand, $\underline{y}_{bi}$ might well be negative, and represent the maximum short-run load that participant $b$ is capable of taking at any price.

Consider an omniscient central planner, that I will refer to as the central dispatch agency (CDA). In particular, the CDA knows all of the participant benefit functions, $u_b(\cdot)$, as well as the engineering constraints in the grid. It is the former information requirement that seems implausible; the information problem will be discussed later. To maximize system efficiency, the CDA solves:

$$\begin{aligned} \max_{\{y_{bi},\theta,\eta,\mu\}} & \sum_b u_b(\vec{y}_b) \\ & + \theta \sum_b \sum_i y_{bi} - L(Y) && \text{(energy balance)} \\ & + \sum_l \eta_{l,min}(Z_l - Z_{l,min}) && \text{(line minima)} \\ & - \sum_l \eta_{l,max}(Z_l - Z_{l,max}) && \text{(line maxima)} \\ & - \sum_b \sum_i \mu_{ij,max}(y_{bi} - \bar{y}_{bi}) && \text{(capacity maxima)} \\ & + \sum_b \sum_i \mu_{ij,min}(y_{bi} - \underline{y}_{bi}) && \text{(demand maxima)} \end{aligned}$$

Now suppose that we define $\eta_l \equiv \eta_{l,min} - \eta_{l,max}$, and

$$p_i^* = \left(1 - \frac{\partial L(Y)}{\partial y_i}\right)\theta + \sum_l \eta_l \frac{\partial Z_l}{\partial y_i}$$

The $p_i^*$ are the shadow prices for delivered power at each node. These shadow nodal prices are obtained by solving the first-order conditions of the system above:

$$p_i^* = \frac{\partial u_b}{\partial y_{bi}} + \mu_{bi}$$

where $\mu_{bi} \equiv \mu_{bi,min} - \mu_{bi,max}$. Thus, for suppliers the $p_i^*$ will always be greater or equal to the supplier's marginal willingness-to-sell. For demanders, $p_i^*$ will always be greater than or equal to the demander's marginal willingness-to-pay. Thus, the CDA can assign the allocation $\{y_{bi}^*\}$ and charge each participant $\sum_i p_i^* y_{bi}^*$, and everyone should be willing to participate.

So far the only participants considered have been the buyers and sellers of power at the wholesale buses. What about the willing participation of the transmission line owners? Hogan (1992) has shown that the owner of a line from node $j$ to node $i$ will be a willing participant if she is paid

$$T_{ji}^* = p_i^* - p_j^*$$

per unit of power carried on that line. From the definition of $p_i^*$ we see that

$$T_{ji} = p_i - p_j = \left(-\frac{\partial L(y)}{\partial y_i} + \frac{\partial L(y)}{\partial y_j}\right)\theta + \sum_l \eta_l \left(\frac{\partial Z_l}{\partial y_i} - \frac{\partial Z_l}{\partial y_i}\right).$$

The first term in this expression is the transmission loss along line $ji$. The second term is the shadow value that represents the congestion rent, or scarcity value of the line given the flow conditions in the grid.[14]

[14] Hogan (1992) shows that in a DC network, the congestion payments collected are exactly equal to the congestion payments made, and that the loss payments exceed the value of the losses incurred, so that there is some net contribution to the fixed costs of the transmission grid. Unfortunately, this happy outcome cannot be guaranteed in real, AC grid.

*Characterizing the Solution*

These shadow prices can be used to decentralize the wholesale market (if they can be determined). Suppose that each participant is allowed to anonymously buy and sell any amount of power at any node, for price $p_i^*$. Participants will solve

$$\begin{aligned}\max_{y_{bi}} u_b(\vec{y}_b) - &\sum_{i \in N} p_i^* y_{bi} \\ &- \sum_i \mu_{ij,max}(y_{bi} - \bar{y}_{bi}) \\ &+ \sum_i \mu_{ij,min}(y_{bi} - \underline{y}_{bi})\end{aligned}$$

which has the same first-order conditions as the CDA's problem above. In this hypothetical wholesale market, it is easy to see why the transmission prices $T_{ji} = p_i^* - p_j^*$ make the line owner indifferent between using the line himself and renting it to others. Suppose the owner has a contract to deliver power from $j$ to $i$. He could take the power at $j$, and use his line to deliver it to $j$. Alternatively, if someone else uses the line, he could take the power at $j$ and resell it on the wholesale market for $p_j^*$, then buy power at bus $i$ for $p_i^*$ and deliver it, at a net cost or $p_i^* - p_j^*$, which is exactly the amount that $T_{ji}$ provides in compensation for foregoing use of the line.

The fundamental problem with this decentralized solution is that full implementation places a heavy information burden on the CDA. How is it to determine the participant benefit functions, $u_b(\vec{y}_b)$? The $\{p_i^*\}$ vary over both time and space; variations over time can occur at fairly high frequency, as Bohn et al. (1984) have shown.

Suppose the CDA asked the market participants to report their benefit functions so that it could determine the efficient allocation? The problem is that participants would know how their reported preferences would be used to set prices, and thus they would have an incentive to misreport in order to manipulate the prices. For example, generators might tend to underreport the availability of capacity, in order to raise spot prices and then deliver more power at those prices.

Hogan (1992) proposes basing prices on *ex post* information: the actual injections and extractions at each node. That is, the CDA operators would solve the allocation problem centrally, without explicitly maximizing the efficiency of the network. The presumption is that through experience and communications with various generators and loads, the CDA can come close to the efficient allocation without formally solving for the market solution. Then, after the dispatch was

determined, consistent prices can be determined: *i.e.*, those prices that would have generated the actual allocation. The quality of these *ex post* estimates will depend, however, on the quality of certain types of information that the CDA must still obtain, such as the marginal cost of generation at each bus, and bounds on the marginal valuation of load at each bus (from, for example, the prices in interruptible contracts).

There are two classes of problems with Hogan's method. The first is that it still imposes a rather heavy information burden on the CDA. How closely the calculated bus prices agree with the true consistent prices depends on the quality of information provided to the CDA. Some of that information is not easy to observe in ordinary operations. In particular, as discussed above in Section 4, a participant's marginal valuation at one node may depend in complex ways on the prices at other nodes. It may not be straightforward in a wholesale grid to identify the marginal generator providing power at a given node, since that node may attach to another grid; likewise it may be difficult to quantify the marginal load valuation. For this reason, Hogan's method seems better suited to pricing for an integrated pool (in which the CDA controls distribution across the entire grid) than to a wholesale market that interconnects to other grids with their own dispatch operators.[15]

The second general problem with Hogan's method is that prices set *ex post*, even if they equal the true prices that would have generated the observed system allocation, do not provide participants with the correct *ex ante* incentives that would induce an *efficient* system allocation. If participants could determine on their own the efficient system prices, and acted on that basis (in terms of loads offered and demanded), then their expectations would be fulfilled and the *ex post* prices would rationalize the efficient equilibrium. But of course, the participants are if anything less able than the CDA to correctly determine the efficient prices *ex ante*. Thus, their participation will generally be based on expected prices that are inconsistent with an efficient allocation, and the resulting allocation will be inefficient. How serious this problem is depends, of course, on how accurately participants can forecast the efficient system prices. The more frequently that the system allocation is changing, the less accurate we would expect price expectations to be. Particularly during periods

[15] There is another information problem with Hogan's method: if participants are asked to provide the needed information, there is again a likelihood that reporting will be distorted to manipulate prices.

of system stress, the participants may make their economic decisions based on expected prices that lead to a quite inefficient allocation.

As with any scheme that approximates competitive market pricing, we should expect the Schweppe et al. and Hogan methods to only approximate the efficient allocation. Is it possible to do better? That depends on how much efficiency is lost relative to an alternative scheme, and how much more (or less) costly it is to implement the alternative scheme. This is fundamentally an empirical question, and the answer is likely to differ across different power systems.[16] However, the first necessary step is to have a feasible alternative mechanism. Such an alternative is proposed in the next section.

## 6. An Efficient, Decentralized Auction Mechanism

The Hogan implementation of spot pricing approximates prices via *ex post* inference, by relying on the ability of the CDA operators to expertly approximate the efficient allocation without complete information or *ex ante* pricing signals to participants. An alternative is to elicit the necessary information to establish efficient *ex ante* prices. I shall explain how this can be done through a "smart market" auction mechanism.

Suppose that the CDA operates an auction, and solicits the participants to announce benefit functions $r_b(\vec{y}_b)$. The CDA then solves the allocation programming program given above, replacing the unknown true benefit functions ($u_b$) with the announced functions ($r_b$). If bidders truthfully reveal their benefit functions ($r_b = u_b$) then the resulting allocation would be fully efficient. The programming problem is sufficiently general to accomodate loop-flow externalities within the wholesale grid, as well as complex bidder contingencies and interactions that might result from loop flow considerations in their local, attached grids.

How to obtain truthful revelation? I propose the use of a "smart market."[17] The idea of a smart market is to combine modern computing power with preference revelation theory in order to obtain

[16] For example, the severity of the approximation errors in the Hogan method depend on the ease of measuring marginal load valuations at constrained busses, and of measuring marginal generation costs. As suggested above, this problem should be much less severe in an integrated pool than in a wholesale market that interconnects various separate grids.

[17] The idea for a smart market originated with Vernon Smith and Charlie Plott [cite]. Hal Varian and I have proposed a simple smart market for allocating spot delivery priorities on the Internet (MacKie-Mason and Varian (1993), MacKie-Mason and Varian (1994)). I have developed a smart market scheme for reserving resources in advance in a network running a connection-oriented protocol like Asynchronous Transfer Mode (ATM) (MacKie-Mason (1994)).

efficient allocations. The market is "smart" because the "auctioneer" clears prices not by simply matching the highest bid to the lowest ask, but by repeatedly solving a complex programming problem that can handle multiple dimensions, varying quantities, and externalities.

The smart market is a "generalized Vickrey auction" (GVA). The Vickrey auction is an elegant scheme for assigning a good to the agent who places the highest value on it, when individual valuations are private information. The idea is to solicit bids and award the good to the highest bidder, but charge the second highest bid as the price. It turns out that bidding one's true valuation is a dominant strategy for each agent. The generalized Vickrey auction extends the idea to allocate multiple units of a good, multiple goods, and goods with externalities (so that agents care about how much others are consuming).

To understand the intuition for a Vickrey auction, suppose there is one unit of one good, and two agents. Efficiency requires that the agent with the higher valuation, $v_i$ receive the good. If agents announce a bid $b_i$, then agent 1's expected payoff is

$$\Pr[b_1 > b_2][v_1 - b_2].$$

Suppose that agent 1 announces the truth, $b_1 = v_1$. Then agent 1 always gets the good when the payoff is positive ($v_1 - b_2 > 0$), and never when it is negative. Clearly this is a dominant strategy. Since both agents will tell the truth, the good always goes to the agent with the higher true valuation.

This example illustrates the intuition for how the Vickrey auction works. First, the probability that an agent wins the auction depends on her bid, but the payoff she receives does not. Second, since the agent's bid cannot affect the payoff, the payoff should be structured so that truthtelling always wins the auction when the payoff is positive, and always loses when the payoff is negative. Truthtelling then dominates, and the resulting allocation is Pareto efficient.

The payoff structure that induces truthtelling in the one good, one unit Vickrey auction is second-pricing. More generally, the payoff structure that induces truthtelling is to give each participant his or her utility plus all of the social surplus of the other agents measured using reported utility functions. If the agent reports truthfully, then the agent's payoff is precisely equal to total social welfare from the auction, which is the desired maximand. Since a truthtelling agent receives precisely what the auction is maximizing, there can be no better strategy than truthtelling.

Of course, giving each agent all of the social surplus clearly is very costly! Therefore, the auction design usually requires a payment by each agent. If that payment depends only on the

reports made by *other* agents it will not change truthtelling incentives. This is the role that the "second-price" plays in the Vickrey auction: it takes away the value of the good to the highest value user other than the agent in question. Thus in the two agent example, the winning agent gets her surplus less the second agent's surplus; the losing agent gets the winning agent's surplus less the winning agent's surplus, or zero.

The generalized Vickrey auction for the power wholesale market follows this intuition, but applies to problems in which agents can have preferences over more than one good, more than one unit of the goods, and over the quantities of the goods that are consumed by other agents (externalities). Recall that the objective function is

$$\max_{\{y_{bi}\}} \sum_b u_b(\vec{y}_b)$$

subject to the network engineering and capacity constraints in conditions (1) – (5). The the GVA for the problem is:

1. Each bidder $b$ reports a bid function, $r_b(\vec{y}_b)$.
2. The CDA computes

   $$\{\vec{y}_b^*\} = \arg\max \sum_b r_b(\vec{y}_b),$$

   subject to constraints (1) – (5). The CDA then assigns the vector $\vec{y}_b^*$ to each bidder $b = 1, \ldots, B$, and computes

   $$W_{-b}(Y^*) = \sum_{a \neq b} r_a(\vec{y}_a^*).$$

3. Bidder $b$ receives net payoff

   $$u_b(\vec{y}_b^*) + \left[W_{-b}(Y^*) - G_b(r_{-b})\right], \tag{6}$$

   where $G_b(r_{-b})$ is an arbitrary function of the *other* other bidders' reported benefits.

Truthtelling is the dominant strategy for this auction. The "second-price" analogue for a payment $G_b(\cdot)$ by each agent is to charge each agent the total social surplus that would be possible if that agent did not participate in the auction at all. The result, then, is that the net payoff received by agent $b$ is the net increment in total surplus that his participation creates. This payment would be

$$G_b(r_{-b}) = \max_{\{y_{ai}\}} \sum_{a \neq b} r_a(\vec{y}_a)$$

subject to the same constraints (1) – (5), except that they now are applied only to participants $a \neq b$.

With this definition of the payment $G_b(\cdot)$, user $b$'s payoff in (6) has two nice interpretations. First, the expression in square brackets is the increase (decrease) in the welfare of all other users that is created by user $b$'s participation. Thus, for participating in the market, $b$ receives her own direct utility plus the net increase in the social welfare of others that she creates by her participation. If her injections and extractions affect marginal transmission losses or affect binding thermal line constraints, then she has a negative effect on the value of the grid to others, and she pays precisely the incremental reduction in grid value to others that her participation causes.

Alternatively, if we regroup the expression (6),

$$\left[u_b(\vec{y}_b^*) + W_{-b}(Y^*)\right] - G_b(r_{-b}),$$

it becomes clear that $b$ is receiving the net increment in total social surplus that her participation creates. Since she gets all of the incremental surplus, she wants to truthfully reveal her preferences. If her participation creates a negative congestion externality, then the net increment in total surplus is less than her direct utility gain, and she pays a congestion tax.

One problem with the GVA is that it requires substantial computation. If the network planner simply knew the correct utility functions, as I assumed in section 5, the computational burden would be modest. If the utility functions $u_b(\cdot)$ were linear, then the problem would be equivalent to a linear programming problem, which could be quickly solved. With nonlinear, but strictly concave utility functions, the problem becomes somewhat more costly, but a solution is guaranteed in finite time.

Unfortunately, when we turn to the GVA, with unknown participant benefit functions, the problem becomes considerably more costly. The GVA mechanism itself requires only a single solution of the network optimization, plus some ancillary linear computations. However, the GVA with a payment in the suggested form of $G_b(\cdot)$ requires that the network problem be re-solved once for every $b$ (with a minor reduction in dimensionality by the elimination of one participant). Since some price $G_b(\cdot)$ would certainly be necessary for any practical GVA, it appears that the computation may be on the order of $B$ separate network optimization problems.[18] It is clear that

[18] The problems are not entirely unrelated of course: the solution of each of the $B$ reduced grid allocations with one participant eliminated should be substantially easier given the solution to the initial problem with all participants active. The initial solution is likely to give good starting values and clues about which constraints are binding.

the granularity of the allocation problem is a crucial decision variable: the more disaggregated are participants, the larger the computational burden becomes in two dimensions: the solution time for the network optimization is likely to increase polynomially in $B$, and the number of such optimizations increases linearlly in $B$.

Fortunately, not all of the computation must be undertaken "on-line", in real time. Only the efficient allocation is required for the operation of the grid. The extra computations required for the payments to and from various participants can be undertaken off-line, with some lag, without affecting the truthtelling incentives. Thus, it seems reasonable to think that the GVA can be solved with about the same frequency and resource commitment as the optimal dispatch problem for a grid.

## 7. Remarks

There are a number of difficult questions to address before implementing a smart market. A continuously updating auction would not be feasible with current technology, given the computation complexity and cost. Therefore, it is necessary to decide how frequently the market should clear: should bids be for hourly power? Half-hourly? What arrangements should be made for futures contracting? In addition, I have assumed that benefit functions are time-separable. In fact, buyer and seller valuations are not independent over hours of the day. For example, a generator's optimal commitment plan could be affected by spot wholesale prices, due to fixed spin-up costs and the contraints imposed by minimum downtime for generators. In theory it would be more efficient if futures prices for $T$ periods ahead where simultaneously determined with the spot prices, and then recalculated every bidding period.

Another issue that needs to be addressed is cost recovery. It can be shown that the efficient spot prices (that include congestion rent) provide the correct marginal signal for generator and transmission line capacity expansion. However, the revenues of the auction may not cover total costs. How then should costs (especially fixed costs) be recovered?

These considerations are quite important for designing a functioning, feasible market. However, these same issues must be addressed by other market mechanisms, as well: they are not peculiar to the smart market. For example, although the payments in a generalized Vickrey auction are not guaranteed to balance the budget, the same is true of the other market-based proposals that use

approximations to efficient pricing. The problem is caused by the high fixed costs in the industry, not by the particular market mechanism implemented. Nonetheless, it may be useful to pursue the question of cost recovery in the context of designing smart markets: perhaps an incentive-compatible, efficiency-preserving mechanism can be developed for cost recovery that preserves the efficiency benefits of the spatial smart market for power and transmission.

## 8. Numerical Example.

## References


Blumstein, C., and Bushnell, J. (1994). A reader's guide to the Blue Book: Issues in California's electric industry restructuring and regulatory reform. *Electricity Journal*. Co-authored with the Program on Workable Energy Regulation (POWER), University of California Energy Institute.

Bohn, R. E., Caramanis, M. C., and Schweppe, F. C. (1984). Optimal pricing in electrical networks over space and time. *Rand Journal of Economics, 15*(3), 360–376.

Caramanis, M. C., Bohn, R. E., and Schweppe, F. C. (1982). Spot pricing of electricity: Practice and theory. *IEEE Transactions on Power Apparatus and Systems, PAS-101*(9), 3234–3245.

Hogan, W. (1993). Markets in real electric networks require reactive prices. *The Energy Journal, 14*(3), 171–200.

Hogan, W. W. (1992). Contract networks for electricity power transmission. *Journal of Regulatory Economics, 4*, 211–242.

Kahn, E., and Baldick, R. (1994). Reactive power is a cheap contraint. Tech. rep., University of California Energy Institute, Berkeley.

MacKie-Mason, J. K. (1994). Reserving resources in an integrated services information network: A smart market mechanism. Tech. rep., University of Michigan, Dept. of Economics.

MacKie-Mason, J. K., and Varian, H. (1993). Some economics of the Internet. Tech. rep., University of Michigan. Available from ftp://gopher.econ.lsa.umich.edu/pub/Papers.

MacKie-Mason, J. K., and Varian, H. (1994). Pricing the Internet. In Kahin, B., and Keller, J. (Eds.), *Public Access to the Internet*. Prentice-Hall, Englewood Cliffs, New Jersey. Available from ftp://gopher.econ.lsa.umich.edu/pub/Papers.

Newberry, D. M., and Green, R. (1995). Regulation, public ownership and privatisation of the english electric industry. Tech. rep., Dept. of Applied Economics, Cambridge University. Forthcoming in *Internation Comparisons of Electricity Regulation*, Richard Gilbert and Edward Kahn, eds.

Ring, B. (1994). *Short Term Pricing in Decentralized Power Systems*. Ph.D. thesis, University of Canterbury, Christchurch, New Zealand.

Schweppe, F. C., Caramanis, M. C., Tabors, R. D., and Bohn, R. E. (1988). *Spot Pricing of Electricity*. Kluwer Academic Publishers, Boston.