COMPETITIVE ELECTRICITY MARKET DESIGN: A WHOLESALE PRIMER

WILLIAM W. HOGAN

December 17, 1998

Center for Business and Government
John F. Kennedy School of Government
Harvard University
Cambridge, Massachusetts 02138

CONTENTS

COMPETITIVE ELECTRICITY MARKETS: A WHOLESALE PRIMER[1]

William W. Hogan[2]

A short-term electricity market coordinated by a system operator provides a foundation for a competitive electricity market. Combined with long-term contracts for generation and transmission congestion, the spot-market and competitive market pricing can support open access to the transmission grid.

INTRODUCTION

Electricity market restructuring emphasizes the potential for competition in generation and retail services, with operation of transmission and distribution wires as a monopoly. Network interactions complicate the design of the institutions and pricing arrangements for open access to the wires. The design of the institutions for the wholesale market can accommodate access for both wholesale and retail competition while recognizing the special requirements of reliability in the transmission grid.

A pool-based, short-term electricity market coordinated by a system operator provides a foundation for building an open access system. Coordination through the system operator is unavoidable, and a bid-based spot-market built on the principles of economic dispatch creates the setting in the wholesale market for competition among the market participants. The associated locational prices define the opportunity costs of transmission usage and support transmission rights without restricting the actual use of the system. A system of contracts can provide the connection between short-term operations and long-term investment built on market incentives.

The key elements of the wholesale market design include:

- A short-term spot market with bid-based security-constrained economic dispatch coordinated by the system operator.
- Spot-market transactions at market clearing locational prices to include marginal

---

[1] This paper combines and updates introductory descriptions from previous papers by the author on competitive market design and transmission pricing.

[2] Lucius N. Littauer Professor of Public Policy and Administration, John F. Kennedy School of Government, Harvard University, and Senior Advisor, Putnam, Hayes & Bartlett, Inc. This paper draws on work for the Harvard Electricity Policy Group and the Harvard-Japan Project on Energy and the Environment. The author is or has been a consultant on electric market reform and transmission issues for British National Grid Company, GPU Inc. (and the Supporting Companies of PJM), GPU PowerNet Pty Ltd, Duquesne Light Company, Electricity Corporation of New Zealand, National Independent Energy Producers, New York Power Pool, New York Utilities Collaborative, Niagara Mohawk Corporation, PJM Office of Interconnection, San Diego Gas & Electric Corporation, Transpower of New Zealand, Westbrook Power, Williams Energy Group, and Wisconsin Electric Power Company. The views presented here are not necessarily attributable to any of those mentioned, and any remaining errors are solely the responsibility of the author. (http://ksgwww.harvard.edu/people/whogan).

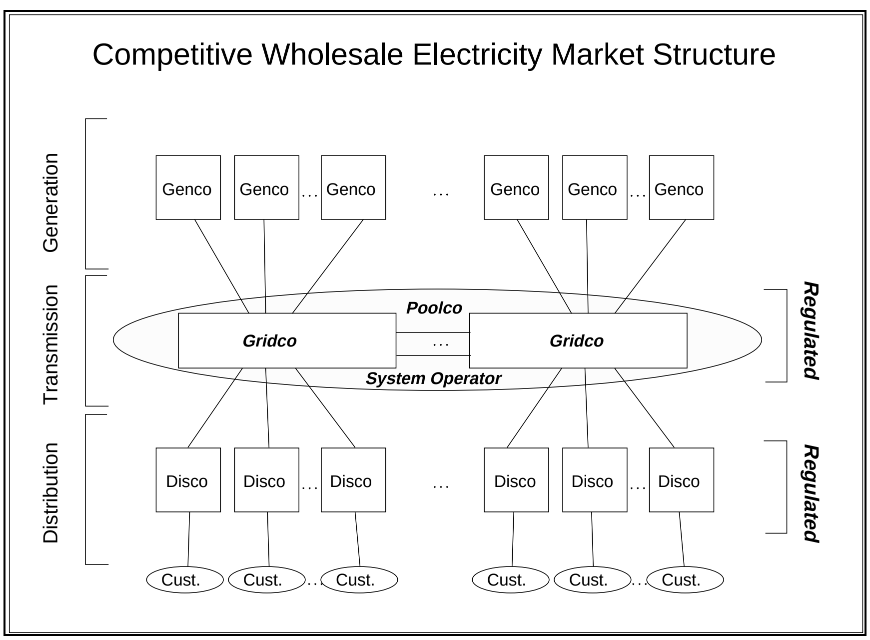
losses and congestion.

- Bilateral transactions with short-term transmission usage charges equal to the difference in the locational prices at source and destination.
- A two settlement system using day-ahead bidding, pricing and contracts, with real-time balancing at the real-time market prices.
- Transmission congestion contracts to allocate the benefits of transmission rights.
- Network access charges to cover the embedded costs of the grid and other fixed charges.
- Usage charges for loads to recover other unbundled ancillary service costs.

A sketch of the details outlines the basic elements of the competitive market design that would support both wholesale and retail transactions.

ECONOMICS OF A COMPETITIVE ELECTRICITY MARKET

A general framework that encompasses the essential economics of electricity markets provides a point of reference for evaluating market design elements. This framework for wholesale market provides the foundation for pricing and access, as well as extension to retail competition.



Restructuring of electricity markets typically emphasizes functional unbundling of the vertically integrated system. The usual separation into generation, transmission, and distribution is insufficient. In an electricity market, the transmission wires and the pool dispatch are distinct essential facilities. The special conditions in the electricity system stand as barriers to an efficient, large-scale market in electricity. A pool-based market coordinated by a system operator helps overcome these barriers.

Reliable operation is a central requirement and constraint for any electricity system. Given the strong and complex interactions in electric networks, current technology with a free-flowing transmission grid dictates the need for a system operator that coordinates use of the transmission system. Control of transmission usage means control of dispatch, which is the principal or only means of adjusting the use of the network. Hence, open access to the transmission grid means open access to the dispatch as well. There must be a system operator coordinating use of the transmission system. That this system operator should also be independent of the existing electric utilities and other market participants is attractive in its simplicity in achieving equal treatment of all market participants. The independent system operator provides an essential service, but does not be compete in the energy market. In the analysis of electricity markets, therefore, a key focus is the design of the interaction between transmission and dispatch, both procedures and pricing, to support a competitive market.

To provide an overview of the operation of an efficient, competitive wholesale electricity market, it is natural to distinguish between the short-run operations coordinated by the system operator and long-run decisions that include investment and contracting. Under the competitive market assumption, market participants are price takers and include the generators and eligible customers. For this discussion, distributors are included as customers in the wholesale market, operating at arm's length from generators. The system is much simpler in the very short run when it is possible to give meaningful definition to concepts such as opportunity cost. Once the short-run economics are established, the long-run requirements become more transparent. Close attention to the connection between short- and long-run decisions isolates the special features of the electricity market.

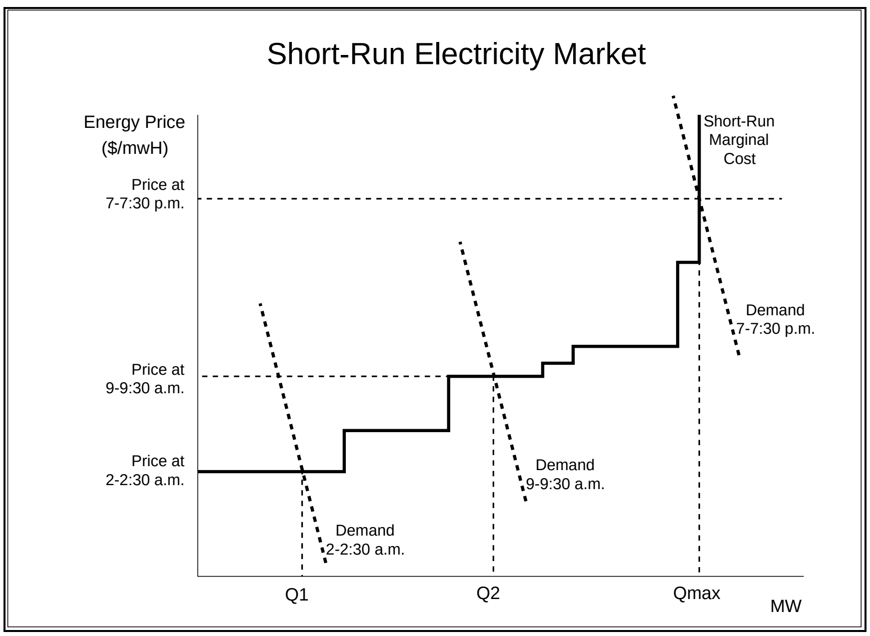
Short-Run Market.

The short run is a long time on the electrical scale, but short on human scale – say, half an hour. The short-run market is relatively simple. In the short run, locational investment decisions have been made. Power plants, the transmission grid, and distribution lines are all in place. Customers and generators are connected and the work of buyers, sellers, brokers and other service entities is largely complete. The only decisions that remain are for delivery of power, which in the short-run is truly a commodity product.[3]

---

[3] On the electrical scale, much can happen in half an hour and the services provided by the system include many details of dynamic frequency control and emergency response to contingencies. Due to transaction costs, if nothing else, it would be inefficient to unbundle all of these services, and many are covered as average costs in the overhead of the system. How far unbundling should go is an empirical question. For example, real power should be identified and its marginal cost recognized, but should this extend to reactive power and voltage control as well? Or to spinning reserve required for emergency supplies? The costs associated with delivery of electric energy include many services. The direct fuel cost of generation is only one component. In analyses of energy pricing, there is no uniformity in the treatment of these other, ancillary services. The typical approach formulates an explicit model approximating the full electricity system, computing both a dispatch solution and associated prices

Over the half hour, the market operates competitively to move real power from generators to customers. Generators have a marginal cost of generating real power from each plant, and customers have different quantities of demand depending on the price at that half hour. The collection of generator costs stacks up to define the generation "merit order," from least to most expensive. This merit order defines the short-run marginal-cost curve, which governs power supply. Similarly, customers have demands that are sensitive to price, and higher prices produce lower demands. Market participants using bilateral transactions provide schedules with any associated bids. Generators and customers do not act unilaterally; they provide information to the dispatcher to be used in a decision process that will determine which plants will run at any given half hour. Power pools provide the model for achieving the most efficient dispatch given the short-run marginal costs of power supply. Although dispatchable demand is not always included, there is nothing conceptually or technically difficult about this extension. The system operator controls operation of the system to achieve the efficient match of supply and demand based on the preferences of the participants as expressed in the bids.



This efficient central dispatch can be made compatible with the market outcome. The fundamental principle is that for the same load, the least-cost dispatch and the competitive-market dispatch are the same. The principal difference between the traditional power pool and the market solution is the price charged to the customer. In the traditional power pool model, customers pay and generators receive average cost, at least on average. However, as shown in the figure marginal cost

for the explicit variables. Everything that is not explicit is treated as an ancillary service, for which the assumption must be that the services will be provided and charged for in some way other than through the explicit prices in the model. Given the complexity of the real electric system, such approximations or simplifications are found in every model, and there is always a boundary between the explicit variables modeled and the implicit variables that will be treated as separate ancillary services. Payment for the ancillary services may be through an average cost uplift applied to all loads. Development of a full description of the interactions with ancillary services is beyond the scope of the present discussion. For the sake of the present discussion, focus on real power and assume that further unbundling would go beyond the point of diminishing returns in the short-run market.

implicitly determines the least-cost dispatch, and marginal cost is the standard determinant of competitive market pricing.

An important distinction between the traditional central dispatch and the decentralized market view is found in the source of the marginal-cost information for the generator supply curve. Typically, the cost data come from engineering estimates of the energy cost of generating power from a given plant at a given time. However, relying on these engineering estimates is problematic in the market model since the true opportunity costs may include other features, such as the different levels of maintenance, that would not be captured in the fuel cost. Replacement of the generator's engineering estimates (that report only incremental fuel cost) with the generator's market bids is the natural alternative. Each bid defines the minimum acceptable price that the generator would accept to run the plant in the given half hour. And these bids serve as the guide for the dispatch.

As long as the generator receives the market clearing price, and there are enough competitors so that each generator assumes that it will not be determining the marginal plant, then the optimal bid for each generator is the true marginal cost: To bid more would only lessen the chance of being dispatched, but not change the price received. To bid less would create the risk of running and being paid less than the cost of generation for that plant. Hence, with enough competitors and no collusion, the short-run central dispatch market model can elicit bids from buyers and sellers. The system operator can treat these bids as the supply and demand and determine the balance that maximizes benefits for producers and consumers at the market equilibrium price. Hence, in the short run electricity is a commodity, freely flowing into the transmission grid from selected generators and out of the grid to the willing customers. Every half hour, customers pay and generators receive the short-run marginal-cost price for the total quantity of energy supplied in that half hour. Everyone pays or receives the true opportunity cost in the short run. Payments follow in a simple settlement process.

Transmission Congestion.

This overview of the short-run market model is by now familiar and found in operation in many countries. However, this introductory overview conceals a critical detail that would be relevant for transmission pricing. Not all power is generated and consumed at the same location. In reality, generating plants and customers are connected through a free-flowing grid of transmission and distribution lines.

In the short-run, transmission too is relatively simple. The grid has been built and everyone is connected with no more than certain engineering requirements to meet minimum technical standards. In this short-run world, transmission reduces to nothing more than putting power into one part of the grid and taking it out at another. Power flow is determined by physical laws, but a focus on the flows – whether on a fictional contract path or on more elaborate allocation methods – is a distraction. The simpler model of input somewhere and output somewhere else captures the necessary reality. In this simple model, transmission complicates the short-run market through the introduction of losses and possible congestion costs.

Transmission of power over wires encounters resistance, and resistance creates losses. Hence the marginal cost of delivering power to different locations differs at least by the marginal effect on losses in the system. Incorporating these losses does not require a major change in the theory or practice of competitive market implementation. Economic dispatch would take account of losses, and the market equilibrium price could be adjusted accordingly. Technically this would yield different marginal costs and different prices, depending on location, but the basic market model and its operation in the short-run would be preserved.

Transmission congestion has a related effect. Limitations in the transmission grid in the short run may constrain long-distance movement of power and thereby impose a higher marginal cost in certain locations. In the simplest case, power will flow over the transmission line from the low cost to the high cost location. If this line has a limit, then in periods of high demand not all the power that could be generated in the low cost region could be used, and some of the cheap plants would be "constrained off." In this case, the demand would be met by higher cost plants that, absent the constraint, would not run, but due to transmission congestion would be then "constrained on." The marginal cost in the two locations differs because of transmission congestion. The marginal cost of power at the low cost location is no greater than the cost of the cheapest constrained-off plant at that location; otherwise the plant would run. Similarly, the marginal cost at the high cost location is no less than the cost of the most expensive constrained-on plant at that location; otherwise the plant would not be in use. The difference between these two costs, net of marginal losses, is the congestion rental.[4]

This congested-induced marginal-cost difference can be as large as the cost of the generation in the unconstrained case. If a cheap coal plant is constrained off and an oil plant, which costs more than twice as much to run, is constrained on, the difference in marginal costs by region is greater than the cost of energy at the coal plant. This result does not depend in any way on the use of a simple case with a single line and two locations. In a real network, the interactions are more complicated – with loop flow and multiple contingencies confronting thermal limits on lines or voltage limits on buses – but the result is the same. It is easy to construct examples where congestion in the transmission grid leads to marginal costs that differ by more than 100% across different locations.

If there is transmission congestion, therefore, the short-run market model and determination of marginal costs must include the effects of the constraints. This extension presents no difficulty in principle. The only impact is that the market now includes a set

---

[4] Losses are not always negligible and could be treated explicitly. However, it would add nothing to the discussion here. At locations distant from the constrained off or constrained on plants, prices can vary even more as it may be necessary to increase the use of expensive plants and decrease the use of cheap plants in combinations that create opportunity costs greater than the cost of any single plant.

of prices, one for each location. Economic dispatch would still be the least-cost equilibrium. Generators would still bid as before, with the bid understood to be the minimum acceptable price at their location. Customers would bid also, with dispatchable demand and the bid setting the maximum price that would be paid at the customer's location. The economic dispatch process would produce the corresponding prices at each location, incorporating the combined effect of generation, losses and congestion. In terms of their own supply and demand, everyone would see a single price, which is the locational marginal cost based price of power at their location. If a transmission price is necessary, the natural definition of transmission is supplying power at one location and using it at another. The corresponding transmission price would be the difference between the prices at the two locations.

This same framework lends itself easily to extensions including bilateral transactions. If market participants wish to schedule transmission between two locations, the opportunity cost of the transmission is just this transmission price of the difference between spot prices at the two locations. This short-run transmission usage pricing, therefore, is efficient and non-discriminatory. In addition, the same principles could apply in a multi-settlement framework, with day-ahead scheduling and real-time dispatch. These extensions could be important in practice, but would not fundamentally change the outline of the structure of electricity markets.

This short-run competitive market with bidding and centralized dispatch is consistent with economic dispatch. The locational prices define the true and full opportunity cost in the short run. Each generator and each customer sees a single price for the half hour, and the prices vary over half hours to reflect changing supply and demand conditions. All the complexities of the power supply grid and network interactions are subsumed under the economic dispatch and calculation of the locational prices. These are the only prices needed, and payments for short-term energy are the only payments operating in the short run, with administrative overhead covered by rents on losses or, if necessary, a small markup added to the cost of ancillary services and applied to all load. The system operator coordinates the dispatch and provides the information for settlement payments, with regulatory oversight to guarantee comparable service through open access to the pool run by the system operator through a bid-based economic dispatch.

With efficient pricing, users have the incentive to respond to the requirements of reliable operation. Absent such price incentives, the system operator would be required to restrict choice and limit the market, in order to give the system operator enough control to counteract the perverse incentives that would be created by prices that did not reflect the marginal costs of dispatch. A competitive market with choice and customer flexibility depends on getting the usage pricing right.

Long-Run Market Contracts.

With changing supply and demand conditions, generators and customers will see fluctuations in short-run prices. When demand is high, more expensive generation will be employed, raising the equilibrium market prices. When transmission constraints bind, congestion costs will change prices at different locations.

Even without transmission congestion constraints, the spot market price can be volatile. This volatility in prices presents its own risks for both generators and customers, and there will be a natural interest in long-term mechanisms to mitigate or share this risk. The choice in a market is for long-term contracts.

Traditionally, and in many other markets, the notion of a long-term contract carries with it the assumption that customers and generators can make an agreement to trade a certain amount of power at a certain price. The implicit assumption is that a specific generator will run to satisfy the demand of a specific customer. To the extent that the customer's needs change, the customer might sell the contract in a secondary market, and so too for the generator. Efficient operation of the secondary market would guarantee equilibrium and everyone would face the true opportunity cost at the margin.

However, this notion of specific performance stands at odds with the operation of the short-run market for electricity. To achieve an efficient economic dispatch in the short-run, the dispatcher must have freedom in responding to the bids to decide which plants run and which are idle, independent of the provisions of long-term contracts. And with the complex network interactions, it is impossible to identify which generator is serving which customer. All generation is providing power into the grid, and all customers are taking power out of the grid. It is not even in the interest of the generators or the customers to restrict their dispatch and forego the benefits of the most economic use of the available generation. The short-term dispatch decisions by the system operator are made independent of and without any recognition of any long-term contracts. In this way, electricity is not like other commodities.

This dictate of the physical laws governing power flow on the transmission grid does not preclude long-term contracts, but it does change the essential character of the contracts. Rather than controlling the dispatch and the short-run market, long-term contracts would focus on the problem of price volatility and provide a price hedge not by managing the flow of power but by managing the flow of money. The short-run prices provide the right incentives for generation and consumption, but create a need to hedge the price changes. Recognizing the operation of the short-run market, there is an economic equivalent of the long-run contract for power that does not require any specific plant to run for any specific customer.

Consider the case first of no transmission congestion. In this circumstance, except for the small effect of losses, it is possible to treat all production and consumption as at the same location. Here the natural arrangement is to contract for the deviation from the equilibrium price in the market. A customer and a generator agree on an average price for a fixed quantity, say 100 MW at five cents. On the half hour, if the pool price is six cents, the customer buys power from the pool at six cents and the generators sells power for six cents. Under the contract, the generator owes the customer one cent for each of the 100 MW over the half hour. In the reverse case, with the pool price at three cents, the customer pays three cents to the system operator, which in turn pays three cents to the generator, but now the customer owes the generator two cents for each of the 100 MW over the half hour. Familiar in other commodity industries, this is known in the electricity market as a "contract for differences."

In effect, the generator and the customer have a long-term contract for 100 MW at five cents. The contract requires no direct interaction with system operator other than for the continuing short-run market transactions. But through the interaction with system operator, the situation is even better than with a long-run contract between a specific generator and a specific customer. For now if the customer demand is above or below 100 MW, there is a ready and an automatic secondary market, namely the pool, where extra power is purchased or sold at the pool price. Similarly for the generator, there is an automatic market for surplus power or backup supplies without the cost and problems of a large number of repeated short-run bilateral negotiations with other generators. And if the customer really consumes 100 MW, and the generator really produces the 100 MW, the economics guarantee that the average price is still five cents. Furthermore, with the contract fixed at 100 MW, rather than the amount actually produced or consumed, the long-run average price is guaranteed without disturbing any of the short-run incentives at the margin. Hence the long-run contract is compatible with the short-run market.

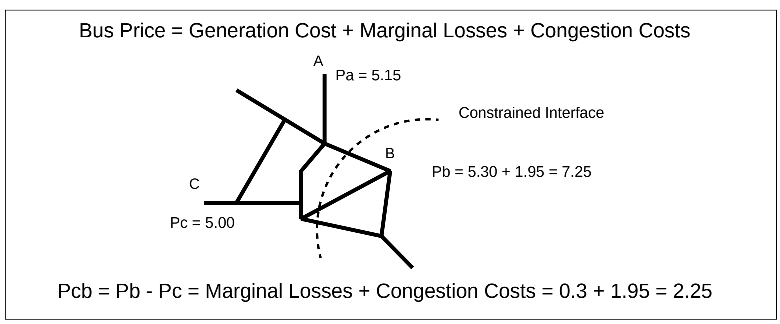
The price of the generation contract would depend on the agreed reference price and other terms and conditions. Generators and customers might agree on dead zones, different up-side and down-side price commitments, or anything else that could be negotiated in a free market to reflect the circumstances and risk preferences of the parties. Whether generators pay customers, or the reverse, depends on the terms. However, system operator need take no notice of the contracts, and have no knowledge of the terms. Such contracts for differences have become common in restructured electricity markets.

In the presence of transmission congestion, the generation contract is necessary but not sufficient to provide the necessary long-term price hedge. A bilateral arrangement between a customer and a generator can capture the effect of aggregate movements in the market, when the single market price is up or the single market price is down. However, transmission congestion can produce significant movements in price that are different depending on location. If the customer is located far from the generator, transmission congestion might confront the customer with a high locational price and leave the generator with a low locational price. Now the generator alone cannot provide the natural back-to-back hedge on fluctuations of the short-run market price. Something more is needed.

Transmission congestion in the short-run market raises another related and significant matter for the system operator. In the presence of congestion, revenues collected from customers will substantially exceed the payments to generators. The difference is the congestion rent that accrues because of constraints in the transmission grid. At a minimum, this congestion rent revenue itself will be a highly volatile source of payment to the system operator. At worse, if the system operator keeps the congestion revenue, incentives arise to manipulate dispatch and prevent grid expansion in order to generate even greater congestion rentals. System operation is a natural monopoly and the operator could distort both dispatch and expansion. If the system operator retains the benefits from congestion rentals, this incentive would work contrary to the goal of an efficient, competitive electricity market.

A convenient solution for both problems – providing a price hedge against locational congestion differentials and removing the adverse incentive for the system operator – is to re-distribute the congestion revenue through a system of long-run transmission congestion contracts operating in parallel with the long-run generation contracts. Just as with generation, it is not possible to operate an efficient short-run market that includes transmission of specific power to specific customers. However, just as with generation, it is possible to arrange a transmission congestion contract that provides compensation for differences in prices, in this case for differences in the congestion costs between different locations across the network.

NETWORK TRANSMISSION CONGESTION CONTRACTS



- DEFINE TRANSMISSION CONGESTION CONTRACTS BETWEEN LOCATIONS.
- FOR SIMPLICITY, TREAT LOSSES AS OPERATING COSTS.
- RECEIVE CONGESTION PAYMENTS FROM ACTUAL USERS; MAKE CONGESTION PAYMENTS TO HOLDERS OF CONGESTION CONTRACTS.
- TRANSMISSION CONGESTION CONTRACTS PROVIDE PROTECTION AGAINST CHANGING LOCATIONAL DIFFERENCES.

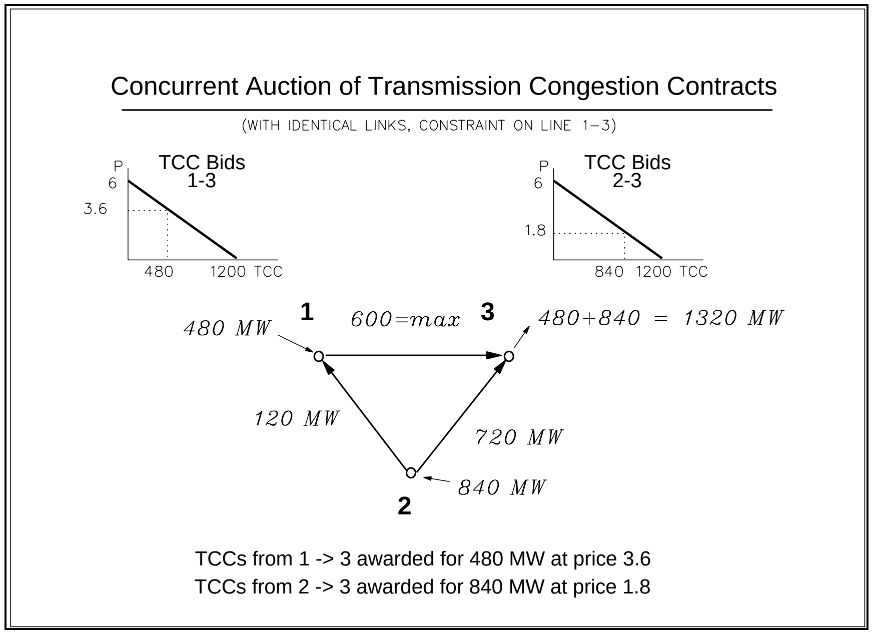
As illustrated in the figure, a transmission congestion contract defines a hedge for differences in locational prices. Everyone who uses the transmission system pays according to the spot-market locational prices. For example, generation at A receives 5.15 cents. Load at B pays 7.25 cents, with the price including marginal losses and congestion. Transmission between C and B would pay 2.25 cents. Included in this transmission charge would be a congestion payment of 1.95 cents. A transmission congestion contract between C and B would result in a payment of the 1.95 cents to the holder of the contract. Hence, if the participant actually transported the power, the transmission congestion contract would just balance the congestion charge for the quantity covered by the contract. And if the holder of the transmission congestion contract does not transport power, the result is the same as selling the transmission right to the actual user.

The transmission congestion contract for compensation would exist for a particular quantity between two locations. A generator at location C might obtain a transmission congestion contract for 100 MW between the generator's location and the customer at location B. The right provide by the contract would not be for specific movement of power but rather for payment of the congestion rental. Hence, ignoring losses, if a transmission constraint caused prices to rise to 6.95 cents at the customer's location, but remain at five cents at the generator's location, the 1.95 cents difference would be the congestion rental. The customer would pay the pool 6.95 cents for the power actually taken. The pool would in turn pay the generator five cents for the power supplied in the short-run market. As the holder of the transmission congestion contract, the generator would receive 1.95 cents for each of the 100 MW covered under the transmission congestion contract. The generation and the load would see the right incentives on the margin. The transmission congestion contract revenue would allow the generator to pay the difference under the generation contract so that the net cost to the customer is five cents as agreed in the bilateral power contract. Without the transmission congestion contract, the generator would have no revenue to compensate the customer for the difference in the prices at their two locations. The transmission congestion contract completes the package described as a "contract network."

Note that the transmission congestion contract is defined as a point-to-point comparison of locational prices. There is no reference to the individual transmission links or the path by which the power may move between two points. This is an important distinction that separates the transmission congestion contract from "link-based" rights that otherwise confront difficult anomalies and perverse incentives. For example, it is entirely possible to construct examples of valuable transmission links that have no net power flow, making them appear worthless for many definitions of link-based rights. But this condition would not impact the definition or the value of transmission congestion contracts.

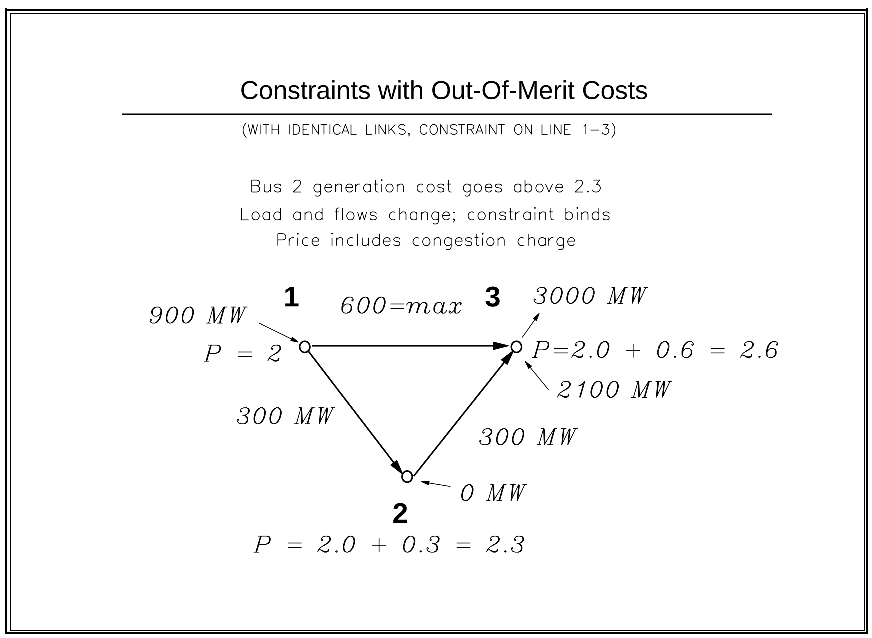
The allocation of transmission congestion contracts among market participants could arise in many ways. For a market in transition during a restructuring process, the allocation might be intended to reflect explicit or implicit historical rights. For new transmission, there could be a negotiation process to select and award one of the many possible combinations of feasible contracts. In addition, some or all of the transmission congestion contracts for the grid could be defined and awarded through an open auction. The collective bids would define demand schedules for contracts. The concurrent auction would respect the transmission system constraints to assure simultaneous feasibility.

For example, the accompanying figure illustrates a hypothetical auction with two demand curves for transmission congestion contracts on a simple network with three



locations. The optimal award that maximizes the benefit[5] of the contracts includes 480 MW from bus 1 to bus 3 and 840 MW from bus 2 to bus 3.

With spot market locational prices, the transmission congestion contracts provide price protection. Even with changing load patterns, the congestion revenues collected by the system operator will be at least enough to cover the obligations for all the contracts.



Consider a typical dispatch with the transmission constraint binding. Suppose the solution is as in the accompanying figure, with a total load of 3000 MW at bus 3. The generation at bus 3 accounts for 2100 MW, with the balance of 900 MW coming from bus 1. The assumed market clearing prices at buses 1 and 3 are 2 cents and 2.6 cents, respectively. From these prices, and the knowledge that there is

---

[5] The benefit is defined as the consumer welfare, the area under the demand curves. Note this is not the same as the value of the transmission congestion contracts at the market clearing price. The model is the simplified DC-Load model.

only one binding transmission constraint, we can compute the implied equilibrium prices at all other buses.[6] In this case bus 2 has a price of 2.3 cents. Note that the market equilibrium for this dispatch is quite different from the pattern of flows envisioned in the allocation of the transmission congestion contracts. However, the different pattern of use, and the associated spot prices, present no difficulty for the system operator. As along as the allocation of the transmission congestion contracts defines a set of inputs and outputs that would be simultaneously feasible, then the revenue collected from the spot prices for congestion will be sufficient to compensate the holders of the contracts for the obligations at the same set of spot prices.[7]

The resulting calculation is illustrated in the accompanying table. The payment obligations for the inputs and outputs at the spot prices define a set of net payments for the actual loads. The payment obligations for the transmission congestion contracts equal the net revenues.

System Operator Revenues

	Quantity	Price	$
Bus 1	900	2	($1,800)
Bus 2	0	2.3	$0
Bus 3	2100	2.6	($5,460)
Bus 3	-3000	2.6	$7,800
TCC 1-3	480	0.6	($288)
TCC 2-3	840	0.3	($252)
Net Total			$0

In effect, the system operator collects the congestion rents from the users of the system and distributes these same rents to the holders of the transmission congestion contracts. In this case, with the simplified assumptions and the same binding constraint, the payments exactly balance. In general, excess congestion rents may remain after paying all obligations under the transmission congestion contracts. These excess rentals

[6] William W. Hogan E. Grant Read, and Brendon J. Ring, "Using Mathematical Programming for Electricity Spot Pricing," International Transactions in Operational Research, Vol. 3, No. 3/4, 1996, pp. 209-221.

[7] For a further discussion of power flows and spot pricing, see William W. Hogan, "Contract Networks for Electric Power Transmission," Journal of Regulatory Economics, Vol. 4, 1992, p. 214.

should not remain with the system operator, but could be distributed according to some sharing formula to those paying the fixed costs of the existing grid or along with the payments under transmission congestion contracts.[8] Many variants are possible, allowing great flexibility in developing and trading contracts. The contract network can allow great commercial flexibility while respecting the reality of the actual network in determining the locational prices.

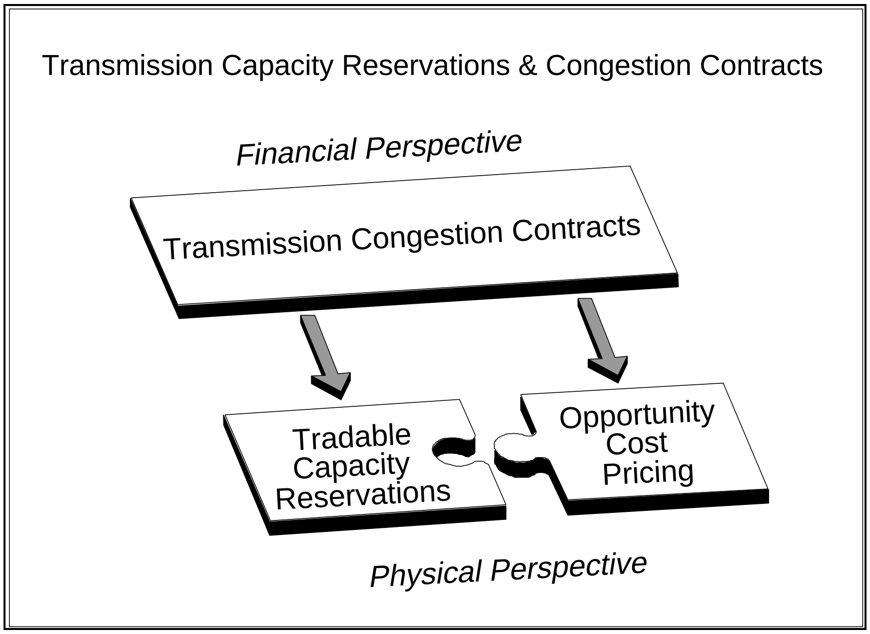
When only the single generator and customer are involved, this sequence of exchanges under the two types of contracts may seem unnecessary. However, in a real network with many participants, the process would be far less obvious. There will be many possible transmission combinations between different locations. There is no single definition of transmission grid capacity, and it is only meaningful to ask if the configuration of allocated transmission flows is feasible. However, the net result would be the same. Short-run incentives at the margin follow the incentives of short-run opportunity costs, and long-run contracts operate to provide price hedges against specific quantities. The system operator coordinates the short-run market to provide economic dispatch. The system operator collects and pays according to the short-run marginal price at each location, and the system operator distributes the congestion rentals to the holders of transmission congestion contracts. Generators and customers make separate bilateral arrangements for generation contracts. Unlike with the generation contracts, the system operator's participation in coordinating administration of the transmission congestion contracts is necessary because of the network interactions, which make it impossible to link specific customers paying congestion costs with specific customer receiving congestion compensation. If a simple feasibility test is imposed on the transmission congestion contracts awarded to customers, the aggregate congestion payments received by the system operator will fund the congestion payment obligations under the transmission congestion contracts. Still, the congestion prices paid and received will be highly variable and load dependent. Only the system operator will have the necessary information to determine these changing prices, but the information will be readily available embedded in the spot market locational prices. The transmission congestion contracts define payment obligations that guarantee protection from changes in the congestion rentals.

Equivalent Transmission Interpretations.

The transmission congestion contract can be recognized as equivalent to an advantageous form of point-to-point "physical" transmission right. Were it possible to define usage of the transmission system in terms of physical rights, it would be desirable that these rights have two features.

First, they could not be withheld from the market to prevent others from using the

[8] The allocation of excess congestion rentals is part of the specification of the transmission congestion contracts. Although important in practice, it would have little impact on the present analysis.

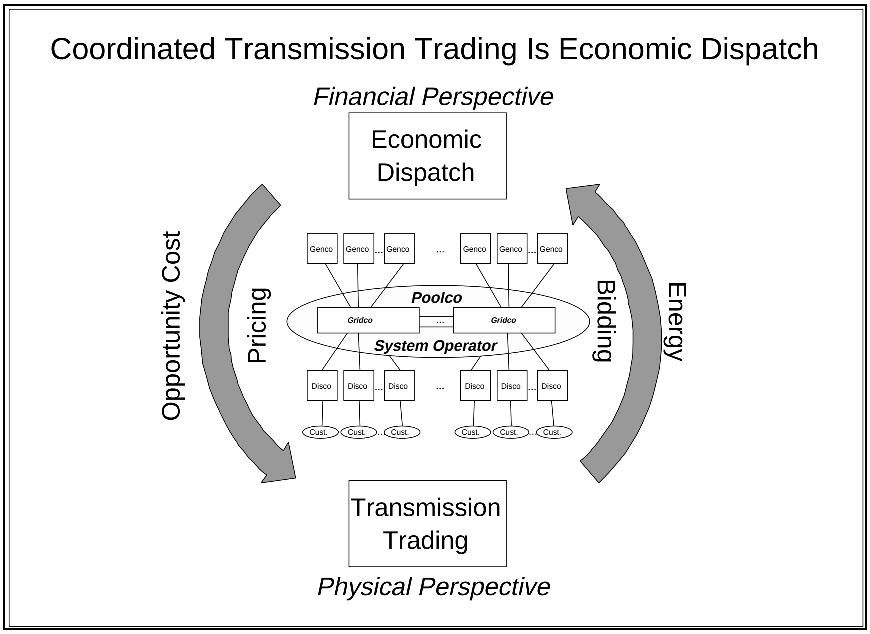


transmission grid. Second, they would be perfectly tradable in a secondary market that would support full reconfiguration of the patterns of network use, at no transaction cost. This is impossible with any known system of physical transmission rights that parcel up the transmission grid. However, in a competitive electricity market with a bid-based economic dispatch, transmission congestion contracts are equivalent to just such perfectly tradable transmission rights.

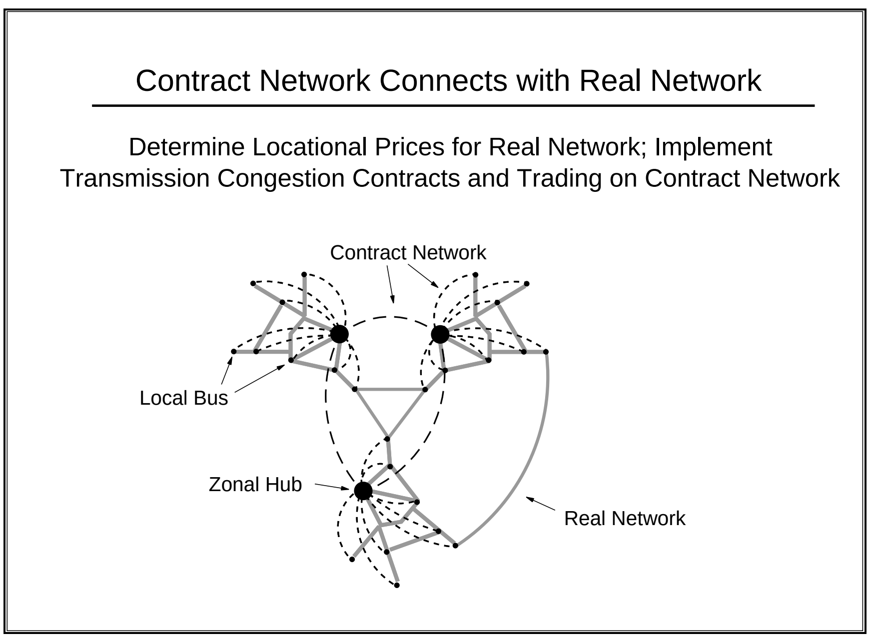
With a competitive market and tradable physical transmission capacity reservations, any use of the system not matched by a reservation would be settled at opportunity cost prices determined by the final dispatch or actual use of the system. Trading of physical transmission capacity reservations must be coordinated through the system operator. Under competitive conditions, the price that would be paid for transmission capacity would be the opportunity cost, which is the same as the difference in the locational prices. If there were no transaction costs, therefore, this physical perspective would be indistinguishable from the financial perspective of transmission congestion contracts based on the same opportunity cost.

The physical perspective may be more intuitive. The financial perspective with transmission congestion contracts as perfectly tradeable rights is easier to implement and has lower transaction costs.

Hence, we can describe transmission congestion contracts either as financial contracts for congestion rents or as perfectly tradable point-to-point physical transmission rights. If the capacity for transmission congestion contracts has been fully allocated, then the system operator will be simply a conduit for the distribution of the congestion rentals. The operator would have no incentive to increase congestion rentals: any increase in congestion payments would flow only to the holders of the transmission congestion contracts. The problem of supervising the dispatch monopoly would be greatly reduced. And through a combination of generation contracts and transmission congestion contracts, participants in the electricity market can arrange price hedges that could provide the economic equivalent of a long-term contract for specific power delivered to a specific customer.



Further to the application of these ideas, locational marginal cost pricing lends itself to a natural decomposition. For example, even with loops in a network, market information could be transformed easily into a hub-and-spoke framework with locational price differences on a spoke defining the cost of moving to and from the local hub, and then between hubs. This would simplify without distorting the locational prices. As shown in the figure, a contract network could develop that would be different from the real network, without affecting the meaning or interpretation of the locational prices.



With market hubs, the participants would see the simplification of having a few locations that capture most of the price differences of long-distance transmission. Contracts could develop relative to the hubs. The rest of the sometimes important difference in locational prices would appear in the cost of moving power to and from the local hub. Commercial connections in the network could follow a configuration convenient for contracting and trading. The separation of physical and financial flows would allow this flexibility.

Creation or elimination of hubs would require no intervention by regulators or the system operator. New hubs could arise as the market requires, or disappear when not important. A hub is simply a special node within a zone. The system operator still would work with the locational prices, but the market would decide on the degree of simplification needed. However, everyone would still be responsible for the opportunity cost of moving power to and from the local hub. There would be locational prices and this would avoid the substantial incentive problems of averaging prices. The hub-and-spoke approach appears to give most of the benefits attributed to zones without the costs, and it implies that the system operator works within a locational pricing framework.

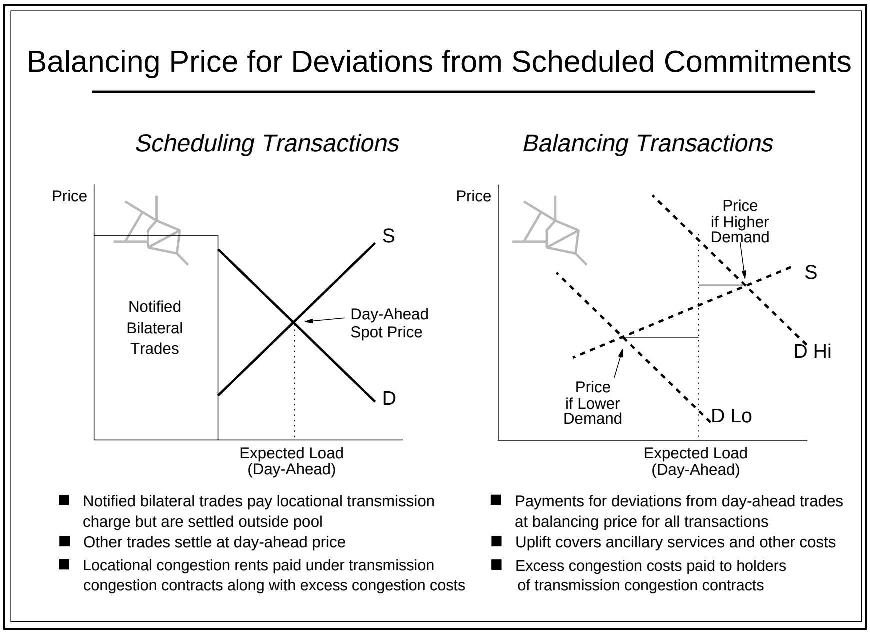
A version of this hub-and-spoke system appeared in the market launched in the Pennsylvania-New jersey-Maryland (PJM) system in April 1998. Although the market participants could create hubs without the participation of the system operator, there was a popular request that the system operator identify and post hub prices. In order to reduce the price volatility that might be present in selecting a single location as a hub, the system operator responded to the request to create hubs consisting of a fixed-weight average of a number of underlying locations. Hence, purchases and sales at the hub are equivalent to a portfolio of purchases and sales at the underlying locations, with the portfolio composition in terms of the fixed weights. Likewise, transmission between any location and the hub is the equivalent of transmission between the given location and the portfolio of locations that make up the hub. This system is working now.

Scheduling and Balancing.

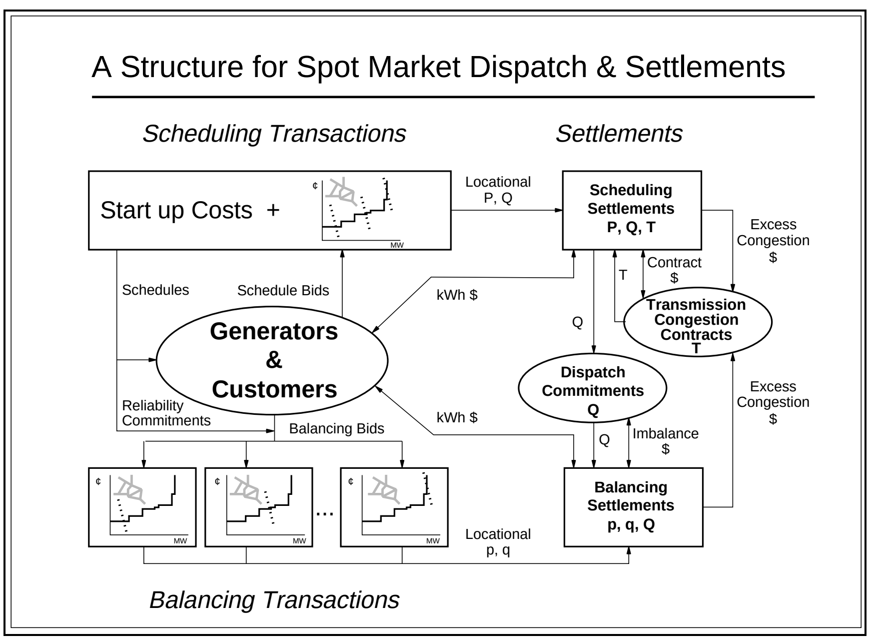
Implementation of the short-run dispatch market could take many forms. In principle, the coordinated dispatch might be left to only the final hour, with all other commitments and trades developed through decentralized transactions, notifying the system operator of the schedules only in time to complete the final balancing. At the other extreme, with long lead times for changing the configuration of generation or the patterns of loads, it might be preferred to include unit commitment decisions over weeks or even longer periods.

In practice, most countries or regions adopt or recommended a procedure that falls somewhere in between these alternatives. A two-settlement system with day-ahead bids is especially attractive in providing greater opportunity for participation by flexible demand without requiring real-time dispatch capabilities. The system operator accepts bids and nominations for scheduled dispatch, say for a day ahead, and determines an appropriate market clearing equilibrium and associated payment settlements. This schedule then defines a set of commitments for delivering and taking power in the short-run dispatch. In the event, the actual dispatch will differ from the scheduled commitments, and appropriate balancing settlements would be arranged.

These connected scheduling and balancing settlements present no major difficulties, but there are a few points that need clarification to maintain consistent payments and incentives. The basic schematic appears in the accompanying figure. Participants in the market submit scheduling bids for the day ahead for both supply and demand. These bids may include start up costs, ramping rates and any of range of plant and load characteristics. The system operator utilizes all this information to define a least-cost dispatch over the day that matches scheduled load and generation. The results is a set of market clearing prices (P) and quantities (Q) that define the schedule. In addition, the system operator arranges for any necessary reliability commitments, such as for spinning and standby reserves to meet any expected deviations in load from the day ahead schedules.



The equilibrium prices and quantities differ by location and represent immediate firm commitments. Based on these commitments, payments would be made through a settlements process. Conceptually these settlements could occur the minute the schedule is determined; in practice, the settlements would occur after the fact. However, to preserve the consistency of the incentives and payments, the settlements must be based on the scheduling prices and quantities. These settlements will include payments under long-term transmission congestion contracts shown by the symbol "T" in the schematic. Holders of transmission congestion contracts would receive or pay



the appropriate amounts of congestion cost differentials between locations for the contracted quantity under the transmission congestion contract. Under some fixed sharing rule, the transmission congestion contract holders, or the presumptive parties responsible for paying the fixed charges of the transmission grid, would also share in any excess congestion payments after settlement of the locational differences.

The schedule and the associated dispatch commitments (Q) would provide the reference point for the actual dispatch. In principle, bids in the scheduling market could be revised to create balancing bids for increments and decrements against the dispatch commitments. Again the system operator would find the least-cost dispatch, hour by hour, based on the actual conditions and the final balancing bids. The result would be an actual dispatch with associated equilibrium prices (p) and quantities (q). These prices and quantities would differ to a degree from the schedules, with the "imbalances" (q-Q) settled at the balancing price of "p". In this balancing settlements, the dispatch commitments are conceptually similar to the transmission congestion contracts that apply in the scheduling settlements. And just as for the transmission congestion contracts, after settling all the imbalances at the market clearing price "p", there could be some excess congestion payments that would be disbursed to the users and not kept by the system operator. In part, this excess would be used to reduce user payments for overhead and ancillary services, not shown in the schematic, or rebated again to those who bear the fixed costs of the grid.

The precise treatment of the excess congestion rentals is not important, other than to disburse them to the users and not the system operator in a way that creates no incentives for an inefficient dispatch. What is important, however, is to settle the scheduling markets and balancing markets with their own internally consistent prices (P,p) and quantities (Q,q). This is required, for example, to avoid the initial over-the-day gaming problems created in England and Wales by settling based on scheduled prices (P) and actual quantities (q).

Long-Term Market Investment.

Within the contract environment of the competitive electricity market, new investment occurs principally in generating plants, customer facilities and transmission expansions. In each case, corresponding contract-right opportunities appear that can be used to hedge the price uncertainty inherent in the operation of the competitive short-run market.

In the case of investment in new generating plants or consuming facilities, the process is straightforward. Under the competitive assumption, no single generator or customer is a large part of the market, there are no significant economies of scale, and there are no barriers to entry. Generators or customers can connect to the transmission grid at any point subject only to technical requirements defining the physical standards for hookup. If they choose, new customers or new generators have the option of relying solely on the short-run market, buying and selling power at the locational price determined as part of the half-hourly dispatch. The system operator makes no guarantees as to the price at the location. The system operator only guarantees open access to the pool at a price consistent with the equilibrium market. The investor takes all the business risk of generating or consuming power at an acceptable price.

If the generator or customer wants price certainty, then new generation contracts can be struck between a willing buyer and a willing seller. The complexity and reach of these contracts would be limited only by the needs of the market. Typically we expect a new generator to look for a customer who wants a price hedge, and for the generators to defer investing in new plant until sufficient long-term contracts with customers can be arranged to cover a sufficient portion of the required investment. The generation contracts could be with one or more customers and might involve a mix of fixed charges coupled with the obligations to compensate for price differences relative to the spot-market price. But the customer and generator would ultimately buy and sell power at their location at the half-hourly price.

If either party expects significant transmission congestion, then a transmission congestion contract would be indicated. If transmission congestion contracts are for sale between the two points, then a contract can be obtained from the holder(s) of existing rights. Or new investment can create new capacity that would support additional transmission congestion contracts. The system operator would participate in the process only to verify that the newly created transmission congestion contracts would be feasible and consistent with the obligation to preserve any existing set of transmission congestion contracts on the existing grid. Unlike the ambiguity in the traditional definition of transmission transfer capacity, there is a direct test to determine the feasibility of any new set of transmission congestion contracts for compensation–while protecting the existing rights–and the test is independent of the actual loads that may develop. Hence, incremental investments in the grid would be possible anywhere without requiring that everyone connected to the grid participate in the negotiations or agree to the allocation of the new transmission congestion contracts.

Access Fees For Embedded Cost Recovery.

There are substantial economies of scale in transmission expansion.[9] The total congestion revenue from transmission usage charges, therefore, generally would be less than the cost of the grid. This means that transmission pricing cannot rely solely on congestion payments to recover the full costs of the existing transmission grid. The

[9] Ross Baldick and Edward Kahn, "Network Costs and the Regulation of Wholesale Competition in Electric Power," Journal of Regulatory Economics, Vol. 5, 1993, pp. 37-384. The problems created by economies of scale are addressed in the context of transmission pricing in the New Zealand by E. G. Read and D. P. M. Sell, "Pricing and Operation of Transmission Services: Short Run Aspects," Report to Trans Power, Canterbury University and Arthur Young, New Zealand, October 1988; E. G. Read, "Pricing of Transmission Services: Long Run Aspects," Report to Trans Power, Canterbury University, New Zealand, October 1988; E. G. Read and D. P. M. Sell, "A Framework for Transmission Pricing," Report to Trans Power, Arthur Young, New Zealand, December 1988.

incentive of avoiding future congestion payments may prompt transmission investment, but the ex post congestion payments could be lower after the investment. Hence, the competitive market model outline here relies on pricing according to two-part prices including a fixed and variable charge. This approach contrasts with the alternative of a one-part transmission charge with a single price based on usage. Application of two-part prices is especially apt in the case of transmission, where it is difficult to produce an acceptable definition of use that is tied to any subset of the network. Any attempt to recover all or some of the fixed costs of the network based on variable charges confronts the problem of defining who is using which facility, and why. In a free-flowing electrical transmission grid, power may flow in ways that are beyond the control or intent of the user. And if as can be the case, the short-run opportunity costs are significantly different than some pro-rata allocation of the long-run capital costs, a one-part tariff that averages fixed costs along with variable costs can produce inefficient incentives and contentious debate over complex and changing allocations. However, in the competitive market model outlined here these problems for transmission are handled through the effective application of a two-part price. Short-run variable costs of transmission use are paid through the automatic pricing of the pool operations. Long-run fixed charges are agreed to under contract as part of the process of deciding on new investments in the grid.

The access fees can differ by location, but would be under the "license plate" model. Customers would pay only once for access to the grid. Prices for actual use of the grid would based in locational opportunity costs in the spot market.

Security Concerns and Capacity Reserves.

Secure economic dispatch coordinated through the system operator respects the current contingency limits on power flows and generation limits. In the short-run, the effect of security constraints in the system is to limit transmission and force out-of-merit generation which gives rise to short-run congestion costs. These short-run marginal costs are the true opportunity costs of "security," and in the competitive market customers face these opportunity costs. In the long-run, investment in the grid is undertaken when customers find it economic to reduce these congestion costs and the cost of losses. In this sense, evolution of the grid would be determined by the market. Since all demand is always met, at the short-run locational marginal cost, there is no non-price curtailment of demand and no need to build excess capacity to provide a reserve margin. There may be high prices at times, and price will rise until demand can be met at marginal cost. But there is no separate or additional security or reliability problem. Hence, security in the short-run is maintained through the security constrained dispatch, and security in the long-run is priced and provided through the market for long-run investments to increase generation and transmission capacity.

CONCLUSION

A pool-based, short-term electricity market coordinated by a system operator provides a foundation for building a wholesale market with open access and tradeable transmission property rights in the form of transmission congestion contracts. Coordination through the system operator is unavoidable, and spot-market locational prices define the opportunity costs of transmission that would determine the market value of the transmission rights without requiring physical trading and without restricting the actual use of the system. In this setting, these transmission congestion contracts are equivalent to perfectly tradeable physical rights.

APPENDIX

Competitive Electricity Market Pricing

INTRODUCTION

Examples of pricing in networks illustrate the issues and the use of least-cost dispatch with accompanying transmission congestion contracts under the pool-based model.[10] Pricing in a competitive electricity market is at marginal cost. The many potential suppliers compete to meet demand, bidding energy supplies into the pool. The dispatchers choose the least-cost combination of generation or demand reductions to balance the system. This optimal dispatch determines the market clearing prices. Consumers pay this price into the pool for energy taken from the spot market and generators in turn are paid this price for the energy supplied.

Inherently, energy pricing and transmission pricing are intimately connected. The FERC has outlined objectives for transmission pricing that would be compatible with a competitive market.[11] A series of examples of pricing in the competitive electricity market model illustrates the determination of prices under economic dispatch in a network and relates transmission constraints to congestion rentals that lead to different prices at different locations. These fundamentals provide the building blocks for an energy and transmission pricing system that addresses the several requirements of the FERC outline.

SHORT-RUN TRANSMISSION PRICING

A system operator (SO) can implement a pricing regime to support the competitive market. This pricing and access regime can accommodate both a pool-based spot market and more traditional "physical" bilateral contracts. The key is in how the SO provides balancing services, adjusts for transmission constraints and charges for transmission usage. The SO would match buyers and sellers in the short-term market. The SO would receive "schedules" that could include both quantity and bidding information. For the participants in the pool, these schedule-bids would be for loads or generation with maximum or minimum acceptable prices. For the self-nominations of bilateral transactions, the schedule-bids would be for transmission quantities with increment and decrement bids for both ends of the transaction. These incremental and decremental bids would apply only

---

[10] These examples illustrate the elements of locational marginal cost pricing. They are adapted from W. Hogan, "A Wholesale Pool Spot Market Must Be Administered by the Independent System Operator: Avoiding the Separation Fallacy," The Electricity Journal, December 1995, pp. 26-37; W. Hogan "Transmission Pricing and Access Policy for Electricity Competition," The Harvard-Japan Project on Energy and the Environment, February 1996.

[11] Federal Energy Regulatory Commission, "Transmission Pricing Issues," Staff Discussion Paper, Inquiry Concerning the Commission's Pricing Policy for Transmission Services Provided by Public Utilities Under the Federal Power Act, Washington, DC, June 1993, pp. 7-8.

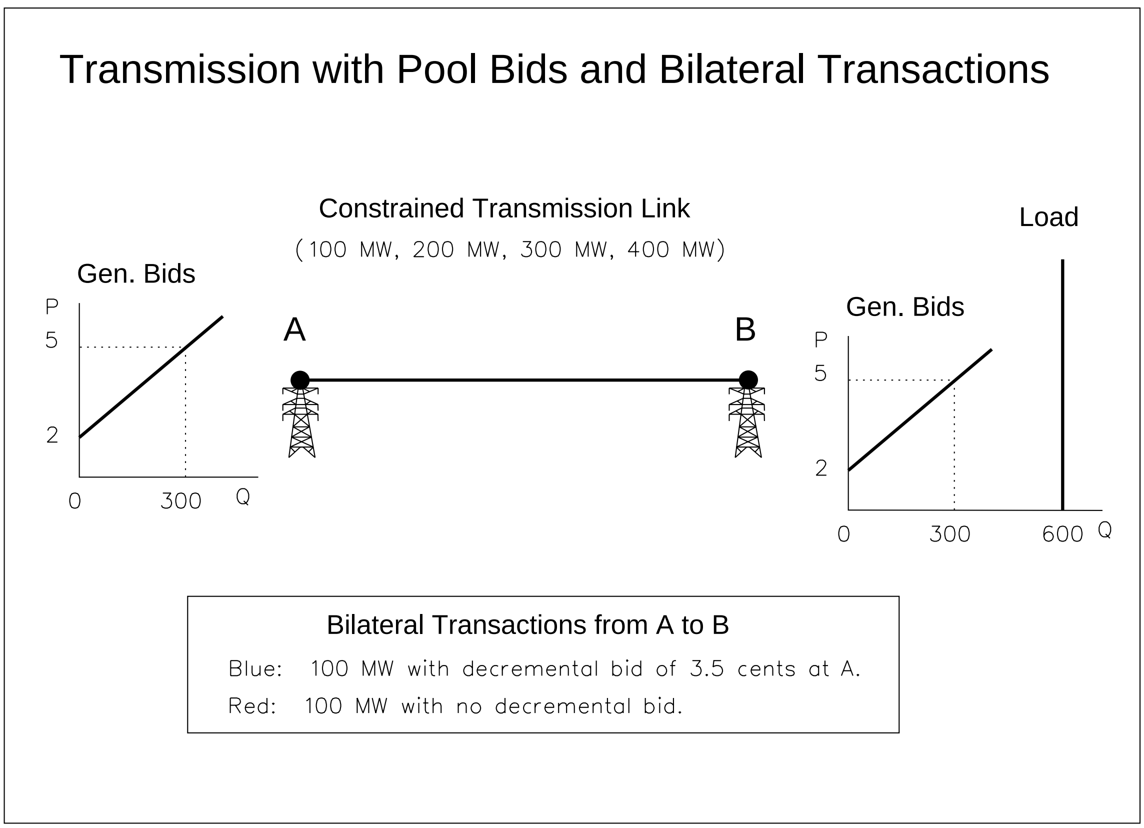
for the short-term dispatch and need not be the same as the confidential bilateral contract prices.

The responsibility of the SO would be to integrate the schedules and the associated bids for deviations from the schedules to find the economic combination for all market participants. This range of schedule-bids would be more varied and flexible, giving everyone more choices.

Transmission Pricing Examples

A set of examples can illuminate the treatment of spot-market transactions and bilateral transactions, under the SO's responsibility to achieve an economic dispatch. These examples are simple, but they capture the essential points in terms of the alternatives available for bilateral transactions. The test of no conflict of interest and non-discrimination is that, other things being equal, there should be no incentive in the dispatch or pricing mechanism to favor either the spot market or the bilateral transaction.

Figure 11



For simplicity, we ignore here any complications of market power or long-run issues, such as the creation of transmission congestion contracts, and focus solely on the short-run dispatch and pricing issues. A market with a single transmission lines, as shown in the accompanying Figure 11, allows an illustration of the basic principles. What is less obvious, however, is that these same principles in no way depend on the special case of a single transmission line. Unlike many other approaches, such as ownership and physical control of the line, or the contract-path fiction, as expanded below in further examples for a grid, these pricing principles extend to a framework to support open access in a complicated network.

The assumptions include:

- Two locations, A and B.
- Total load is for 600 MW at location B. For simplicity, the load is fixed, with no demand bidding.
- A transmission line between A and B with capacity that will be varied to construct alternative cases.
- Pool bid generation at both A and B. To simplify, each location has the same bid curve, starting at 2 cents/kwh and increasing by 1 cent/kwh for each 100 MW. Hence, a market price of 5 cents at A would yield 300 MW of pool-based generation at that location. Likewise for location B.
- Two bilateral transaction schedules, Blue and Red, each for 100 MW from A to B. Each bilateral transaction includes a separate contract price between the generator and the customer; the SO does not know this contract price.

    Blue provides a (completely discretionary) decremental bid at A of 3.5 cents. In other words, if the price at A falls to 3.5 cents, Blue prefers to reduce generation and, in effect, purchase power from the pool. Blue may do this, for example, if the running cost of its plant is 3.5 cents, and it would be cheaper to buy than to generate.

    Red provides no such decremental bid, and requests to be treated as a must run plant.

The SO accepts the bids of those participating in the spot market at A and B and the bilateral schedules. The load is fixed at 600 MW. The bilateral transactions cover 200 MW, or the person responsible for the bilateral transaction must purchase power at B to meet any deficiency. The remaining 400 MW of load must be met from the spot market to include production at A or B, and use of the transmission line.

In determining the economic dispatch, the system operator treats the pool generation bids in the usual way. The Blue bilateral transaction is treated as a fixed obligation, with the 3.5 cent decrement bid as an alternative source of balancing adjustment at A. The Red bilateral transaction is treated as a fixed obligation, with no such balancing adjustment.

Power Flows and Locational Prices					
	Alternative Cases				
Link Capacity A to B	MW	400	300	200	100
Total Load at B	MW	600	600	600	600
Price at A	cents/kwh	4	3.5	3	2
Price at B	cents/kwh	4	5	6	7
Transmission Price	cents/kwh	0	1.5	3	5
Pool Generation at A	MW	200	150	100	0
Pool Generation at B	MW	200	300	400	500
Blue Bilateral Input at A	MW	100	50	0	0
Red Bilateral Input at A	MW	100	100	100	100

Assuming that the net of the fixed obligations with no balancing adjustments is feasible, which is the interesting case, we can vary the capacity on the link to see the results of the economic dispatch and the payments by the participants. The examples cover four cases, starting at 400 MW of transmission capacity, and reducing in increments of 100 MW. The details are in the accompanying table.

400 MW. In the case of 400 MW of transmission capacity, the economic dispatch solution is just balanced with no congestion. Everyone sees the same price of 4 cents. The payments for each party include:

- Pool Generation at A: Paid 4 cents for 200 MW.

- Pool Generation at B: Paid 4 cents for 200 MW.

- Pool Load at B: Pays 4 cents for 400 MW.

- Blue Bilateral: Pays zero cents for transmission of 100 MW.

- Red Bilateral: Pays zero cents for transmission of 100 MW.

Everybody is happy.

300 MW. In the case of 300 MW of transmission capacity, the economic dispatch solution encounters transmission congestion, and the prices differ by location. The price at A drops to 3.5 cents, and the price at B rises to 5 cents. The opportunity cost of transmission is 1.5 cents. The payments for each party include:

- Pool Generation at A: Paid 3.5 cents for 150 MW.

- Pool Generation at B: Paid 5 cents for 300 MW.

- Pool Load at B: Pays 5 cents for 400 MW.

- Blue Bilateral: Pays 1.5 cents for transmission of 50 MW. Blue makes up the remaining 50 MW obligation at B at a price of 5 cents.

- Red Bilateral: Pays 1.5 cents for transmission of 100 MW.

Everybody would prefer less congestion, but everyone is paying the opportunity cost of the transmission congestion. Note that at these prices, Blue is indifferent to bidding in its generation at 3.5 cents in the pool at A, or continuing as a bilateral transaction. Further, note that the SO reduced both pool and Blue transactions. There is no artificial bias induced by the SO fulfilling the directives of the economic dispatch.

200 MW. In the case of 200 MW of transmission capacity, the economic dispatch solution encounters more transmission congestion, and the prices differ more by location. The price at A drops to 3 cents, and the price at B rises to 6 cents. The opportunity cost of transmission is 3 cents. The payments for each party include:

- Pool Generation at A: Paid 3 cents for 100 MW.

- Pool Generation at B: Paid 6 cents for 400 MW.

- Pool Load at B: Pays 6 cents for 400 MW.

- Blue Bilateral: Prefers not to generate and has no transmission. Blue makes up the 100 MW obligation at B at a price of 6 cents.

- Red Bilateral: Pays 3 cents for transmission of 100 MW.

Everybody would prefer less congestion, but everyone is paying the opportunity cost of the transmission congestion. Note that at these prices, Blue is better off than if it had actually generated. Of course, Blue would still be indifferent to bidding in its generation at 3.5 cents in the pool at A, or continuing as a bilateral transaction. Further, note that the SO reduced both pool and Blue transactions. There is no artificial bias induced by the SO fulfilling the directives of the economic dispatch.

100 MW. In the case of 100 MW of transmission capacity, the economic dispatch solution encounters transmission congestion to the point of eliminating everything other than the must run plant, and the prices differ more by location. The price at A drops to 2 cents, and the price at B rises to 7 cents. The opportunity cost of transmission is 5 cents. The payments for each party include:

- Pool Generation at A: No generation.

- Pool Generation at B: Paid 7 cents for 500 MW.

- Pool Load at B: Pays 7 cents for 400 MW.

- Blue Bilateral: Prefers not to generate and has no transmission. Blue makes up the 100 MW obligation at B at a price of 7 cents.

- Red Bilateral: Pays 5 cents for transmission of 100 MW.

Everybody would prefer less congestion, but everyone is paying the opportunity cost of the transmission congestion. Note that at these prices, Blue is better off than if it had actually generated. Of course, Blue would still be indifferent to bidding in its generation at 3.5 cents in the pool at A, or continuing as a bilateral transaction. Further, note that the SO reduced both pool and Blue transactions. There is no artificial bias induced by the SO fulfilling the directives of the economic dispatch.

Power Flows and Locational Prices					
	Alternative Cases				
Link Capacity A to B	MW	400	300	200	100
Price at A	cents/kwh	4	3.5	3	2
Price at B	cents/kwh	4	5	6	7
Transmission Price	cents/kwh	0	1.5	3	5
Payments to Independent System Operator					
Pool Load at B (400 MW)	cents (x1000)	1,600	2,000	2,400	2,800
Contract Load at B (200 MW)	cents (x1000)	0	0	0	0
Generation at A	cents (x1000)	(800)	(525)	(300)	0
Generation at B	cents (x1000)	(800)	(1,500)	(2,400)	(3,500)
Blue Transmission	cents (x1000)	0	75	0	0
Blue Imbalance at B	cents (x1000)	0	250	600	700
Red Transmission	cents (x1000)	0	150	300	500
Red Imbalance at B	cents (x1000)	0	0	0	0
Net to SO	cents (x1000)	0	450	600	500

The net spot-market payments that are made to and from the SO are summarized in the accompanying table. Note that the cases of transmission congestion include net payments to the SO. These net payments are equal to the value of the constrained transmission capacity. These are the congestion payments which would be redistributed through a system of transmission congestion contracts, as illustrated below in further examples.

Implications

These examples for a single, isolated line are simple, but they capture the essential features. These features generalize to a more complicated network under the economic dispatch model in the sense that participants can provide bids at their discretion. Some of the bids can be "must run." The locational prices are easily determined from the economic dispatch considering all the bids and schedules, not just those included in the power exchange. And although everyone would prefer a less congested system, all users would pay the short-run opportunity costs of their contribution to the congestion. Other things being equal, there would be no bias between spot market and bilateral transactions.

Note that if Blue and Red did not pay the opportunity cost of transmission, there would be a substantial bias in favor of the bilateral transactions. Furthermore, the locational prices are consistent with the efficient competitive outcome, as is best illustrated by Blue's willingness to adjust a bilateral transaction.

Contrary to the argument above, that the SO would have a bias in favor of spot market transactions, the treatment of the Red bilateral transaction might lead to an accusation that there is a reverse bias in favor of the bilateral transaction. However, there are two important features of the pricing and access rules that run counter to this assertion.

First, the spot market participants could achieve the same result by bidding in generation at A at a zero reservation price, or lower. In fact, in performing the economic dispatch, the SO treats the Red transaction as just this type of bid. Under these circumstances, the price at A could drop to zero, or lower, with a corresponding increase in the opportunity cost of transmission.

Furthermore, suppose that Red's true short-term generation cost is 3 cents, but it refused to make a decremental bid to the SO. Then in the 100 MW case above, Red would have acted irrationally and would be worse off than if it offered such a decremental bid. It can also be shown that the cost thus imposed on Red is at least as large as the total cost imposed on everyone else in the market. Thus Red would pay for its own mistakes; the effect would be a net gain for the other generators and load (although there could be winners and losers, in aggregate everyone else would win).

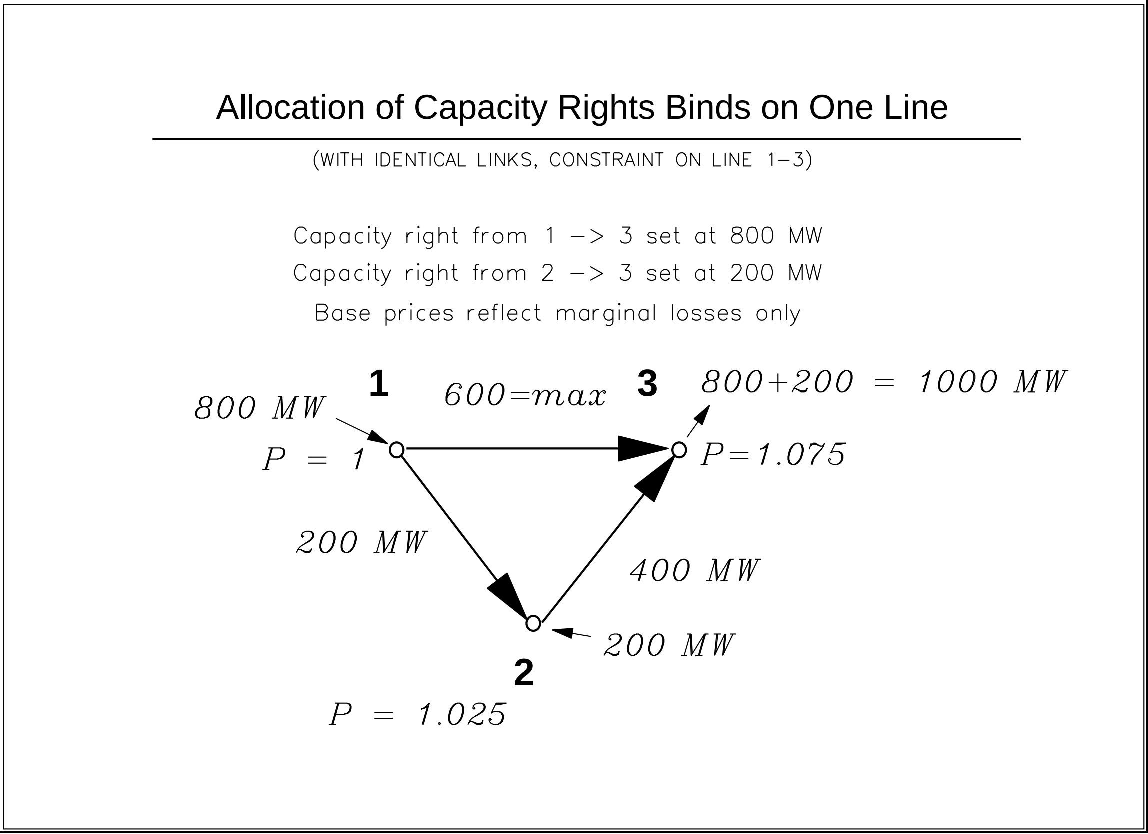
Failure to offer a bid-based economic dispatch will return us to the complications and fictions of the contract-path world of old, and the many artificial arbitrage opportunities that create profit by creating confusion. This would not be good public policy.

CONTRACT NETWORK EXAMPLES

For purposes of further illustration, consider the case of a three bus network with identical lines and identical thermal limits on each line. A three bus network is the minimum case needed to observe the network interaction effects of loop flow. Here we use the DC-Load approximation for real power only, and ignore contingency constraints. Reactive power and contingency constraints can be included without changing any of the fundamental points examined here.[12]

[12] W. W. Hogan, "Contract Networks for Electric Power Transmission," Journal of Regulatory Economics, Vol. 4, No. 3, September 1992, pp. 211-242.

Figure 12

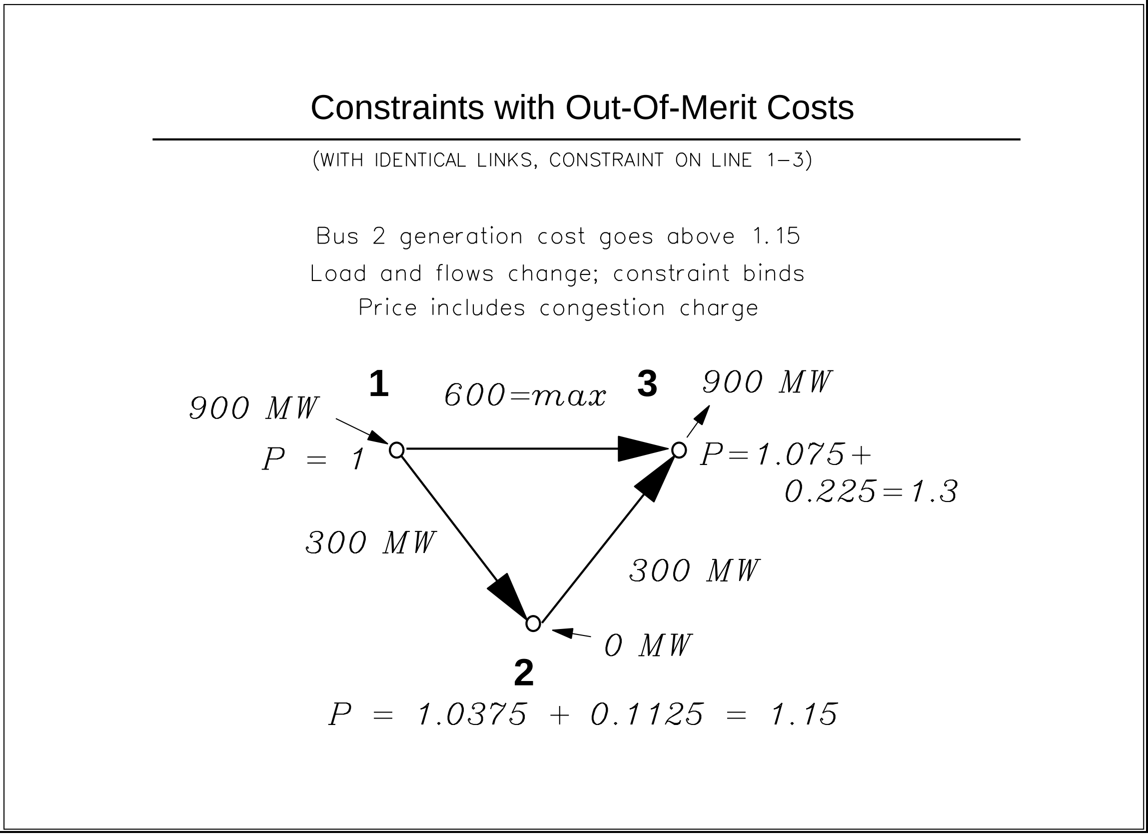


An alternative base case model and an allocation of congestion contracts are shown in Figure 12. Here we assume that the desired transmission congestion contracts are for 800 MW from bus 1 to bus 3, and 200 MW from bus 2 to bus 3. Or, an equivalent definition is that the customer at bus 3 has the contract to purchase 800 MW at bus 1 and 200 MW at bus 2. The simultaneous allocation of these contracts is feasible, but it does hit the thermal transmission constraint of 600 MW on the line between bus 1 and bus 3.

In Figure 12 the prices calculated for this dispatch are shown relative to the price at bus 1. In this instance there is no congestion and the prices cover only the cost of generation at bus 1 and the marginal cost of losses. In this simplified case, the equilibrium required is that the marginal losses are linear in the flow on any link and are the same along any parallel path. Hence the marginal loss of one additional megawatt from bus 1 to bus 3 is 0.075, whether by the path $1 \rightarrow 3$ or via $1 \rightarrow 2 \rightarrow 3$. There is no additional congestion cost, and hence there is no payment from the pool under the congestion rental contracts. Equivalently, the customers at bus 3 buy their 800 MW from bus 1 and 200 MW from bus 2, just as specified in the transmission congestion contracts.

Of course, a change in the economics of generation could induce transmission congestion with the associated differences in prices across locations. In Figure 12, it was

Figure 13



economic to generate power at bus 2 and the actual economic dispatch is the same as the dispatch with simultaneous use of the allocated congestion contracts. In Figure 13, the assumed conditions change with an increase in the running cost of power at bus 2 and the need to use expensive generation at bus 3. If the gross load at bus 3 is still 1000 MW, then part of the load must be met with local generation, which costs 1.3, including a congestion rental of 0.225. At this price for bus 3 and with these loads and flows, the price at bus 2 is determined by the equilibrium conditions of optimal economic dispatch. The easiest way to verify the equilibrium prices is to assume that an additional 2 MW of power could be supplied at bus 2. This would allow a reduction of 1 MW at bus 1 and an increase of 1 MW delivered to bus 3 that could displace the expensive generation at bus 3. The flow from 1→2 would reduce by 1 MW, with a corresponding 1 MW increase along 2→3. The flow along 1→3 would still be at the limit of 600 MW, and total losses would be the same. Hence the net savings would be 1 unit at bus 1 and 1.3 units at bus 3, for a total of 2.3 units. This implies a price for the incremental generation of 2 MW at bus 2 of 2.3÷2 = 1.15, which is the equilibrium price. Along with the marginal losses, this total opportunity-cost price implies a bus 2 congestion price of 0.1125.

By assumption, the cost of the generating plant at bus 2 is above 1.15, so the plant does not run. Now all the power transmitted is generated at bus 1, and only 900 MW can be transmitted. The thermal constraint of 600 MW on the line between bus 1 and bus 3 is binding. All the users of the grid pay or are paid these prices for the actual dispatch. In addition, the holders of the point-to-point transmission congestion contracts receive payments from the pool operators.

The resulting payments are shown in Table IV. Hence the owner of the 800 MW contract from bus 1 is paid the congestion rental difference from bus 1 to bus 3 of 0.225 for the full 800 MW, requiring a payment from the pool of 180. Likewise, the owner of the 200 MW contract is paid the difference in the congestion rental between bus 3 and bus 2, or 0.1125, for the full 200 MW contract and a payment of 22.5. Both users actually buy power from the pool at bus 3 for the price of 1.3. For the customer with the 800 MW contract to purchase at bus 1, the payment of 180 is the total value of the congestion price differential between bus 1 and bus 3. And for the customer with the contract for 200 MW at bus 2, the payment of 22.5 is the total value of the congestion price differential between bus 1 and bus 2.

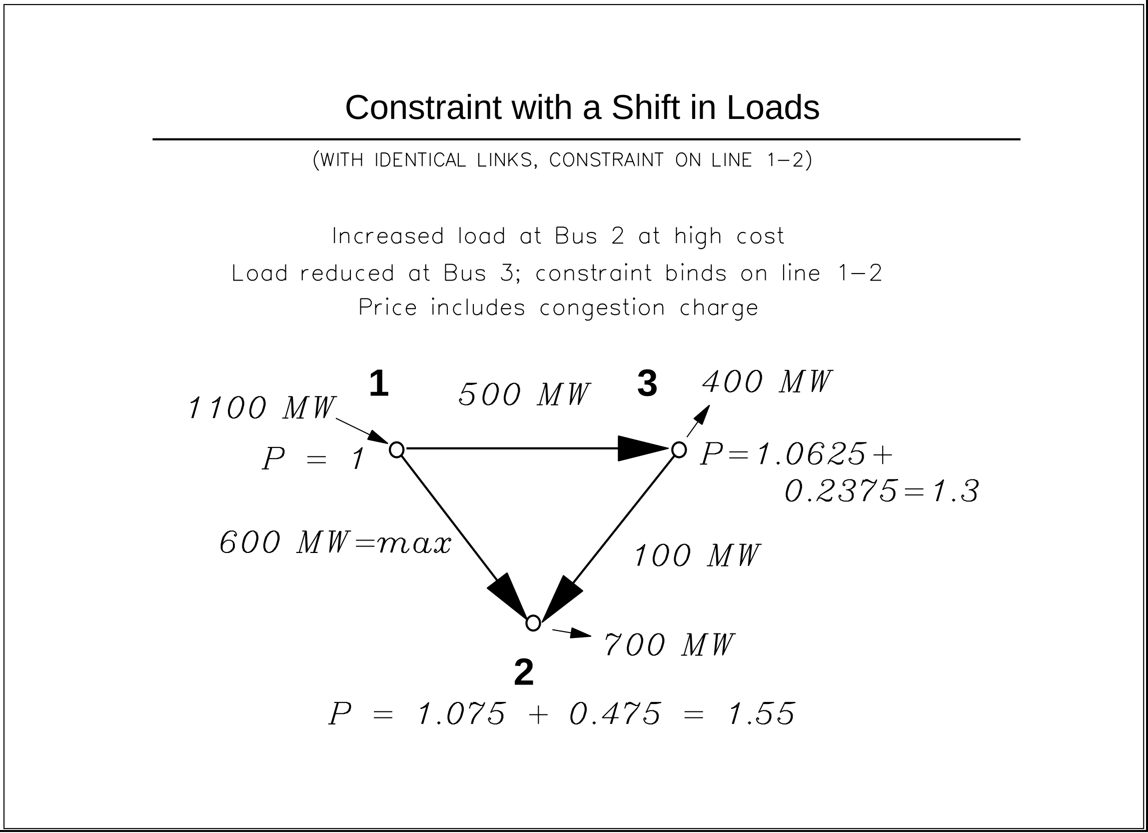
By purchasing 200 MW at bus 3 at a price of 1.3 and then applying the transmission congestion payment of 22.5, the holder of the 200 MW transmission contract can in effect purchase 200 MW at the price at bus 2 and pay only the marginal losses to move the power to bus 3. Although the actual dispatch is different than the simultaneous allocation of congestion contracts, the payments to the congestion contract holders provide the guarantee in effect that the congestion contract holders can purchase power at the price of power at another location. This holds true even if no power was actually generated at that location, as here for bus 2. Furthermore, specific performance to actually generate and transmit the 800 MW and 200 MW according to the congestion contracts would not be feasible under this economic dispatch. Only by foregoing the advantages of the economic dispatch, and increasing total costs, could the specific plants be used for specific customers. The transmission congestion contracts guarantee the economic value of the transmission, but do not determine the actual flows.

Table IV: Congestion with Out-Of-Merit Costs

Congestion Contracts	Q (MW)	Congestion Price Difference	Direct Receipts	Excess Share	Excess Rentals	Net Receipts
1->3	800	0.225	180	80.00%	0	180
2->3	200	0.1125	22.5	20.00%	0	22.5
Total			202.5		0	202.5

The example in Figure 13 finds the pool dispatcher collecting congestion rentals from the actual users and paying the same rentals to the owners of transmission congestion contracts. Because the same transmission constraint limits both the actual dispatch and the initial allocation of transmission congestion contracts, there are no excess congestion rentals. All the congestion revenue collected is required to compensate the holders of the point-to-point transmission congestion contracts.

Figure 14



An alternative case is shown in Figure 14. In this case the economics of load and dispatch have changed significantly. Power still costs 1.3 at bus 3, but now the net load is reduced to 400 MW. There is a big net load at bus 2, and the equilibrium power cost there is at 1.55. The relatively cheap generation at bus 1 is used at the level of 1100 MW, which causes a shift in the flows. Now the dispatcher has no problem with a thermal limit on the line between bus 1 and bus 3, but the line between bus 1 and bus 2 has reached a similar thermal limit at 600 MW. This transmission constraint induces the indicated bus prices and congestion rentals.

To verify the equilibrium price at bus 2, given the prices at buses 1 and 3, note that 1 MW of additional load at bus 2 would be met by an increment of 2 MW in generation (or an equivalent reduction in load) at bus 3 and a reduction of 1 MW of generation at bus 1. The net change in generation cost would be 2*1.3 -1 = 1.6 units. The flow on line 1→2 would be unchanged, but the flow on line 1→3 would be reduced by 1 MW and the reverse flow along 2→3 would increase by 1 MW. Since these lines have different marginal losses, there would be a net reduction in losses of 0.05. Hence, the total increase in cost for an additional MW at bus 2 would be 1.6 - 0.05 = 1.55, the price at bus 2.

Again the pool pays or is paid the short-run prices for power at each of the locations. And again the pool makes payments to the holders of the point-to-point transmission congestion contracts. The summary of the various payments appears in Table V. For the customer with the contract for 800 MW between bus 1 and bus 3, the congestion differential is 0.2375 and the total payment from the pool is 190. This allows the contract holder to purchase 800 MW at bus 3--with some of that power necessarily generated by plants located at bus 3--pay the price of 1.3, and after netting out the payment of 190 from the pool, effectively purchase the 800 MW at bus 1 and pay only marginal losses to move the power to bus 3.

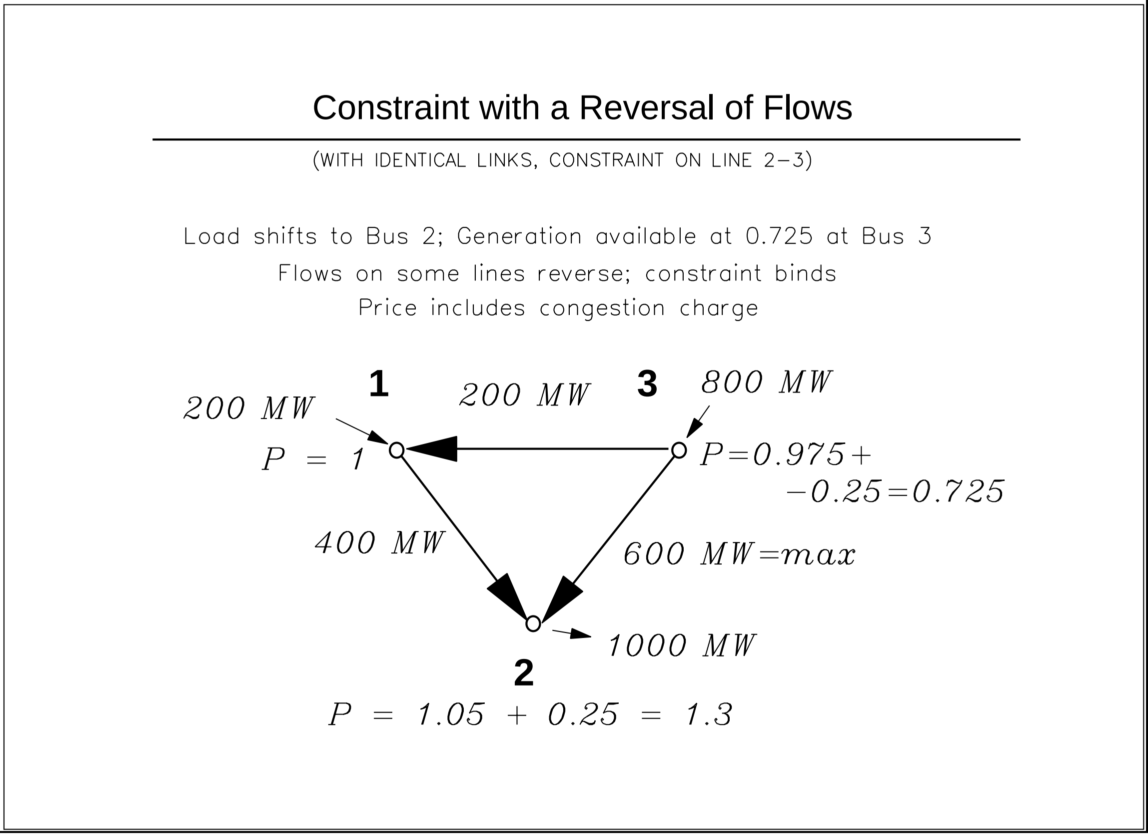
Similarly, the customer with the contract for 200 MW from bus 2 to bus 3 can purchase 200 MW at bus 3 at the price of 1.3. However, this price is lower than the price at bus 2, and the difference in congestion rentals is now negative, at -0.2375. Under the terms of the point-to point contract, this customer must make an added payment of 47.5 to the pool. When coupled with the purchase of 200 MW at bus 3, this is equivalent to purchasing 200 MW at bus 2 at 1.55 and then moving to bus 3 paying only the marginal losses (in this case the marginal losses also would be negative between bus 2 and bus 3). The final effect is as promised under the transmission contract of the customer at bus 3 to purchase 200 MW at bus 2.

Table V: Constraint with a Shift in Loads

Congestion Contracts	Q (MW)	Congestion Price Difference	Direct Receipts	Excess Share	Excess Rentals	Net Receipts
1->3	800	0.2375	190	80.00%	228	418
2->3	200	-0.2375	-47.5	20.00%	57	9.5
Total			142.5		285	427.5

In this case of a shift in loads, with a new transmission constraint binding, the pool can make all the necessary payments to the holders of the point-to-point transmission congestion contracts, but the payments out amount to only 190 - 47.5 = 142.5. However, the total for the congestion rentals paid by the users of the grid is 700*0.475 + 400*0.2375 = 427.5. There remain excess congestion rentals of 285. Assuming that the fixed charge payments are proportional to the transmission congestion contracts, one way to dispose of these excess congestion rentals would be to pay them out to the transmission congestion contract holders in the ratio of 80 to 20. Hence the transmission contract holder from bus 1 would receive an additional payment of 228, for total receipts of 418. Likewise, the transmission contract holder from bus 2 would receive 57 from the excess congestion rentals for total receipts of 9.5.

Figure 15



In Figure 14 there is a shift in load and economics, and one of the transmission contract holders, for power from bus 2, is required to make addition congestion payments to the pool. With enough of a change in the loads and transmission flows, it is possible that everyone with a transmission contract holds them in the reverse direction, and in this case the payments under the sharing of excess congestion rentals take on added importance. For example, consider the conditions depicted in Figure 15. Here the economics of the dispatch and load have changed even more dramatically compared to the initial allocation of transmission congestion contracts. Now there is low price at bus 3 and a net input of 800 MW, and the higher price is at bus 2 with a net load of 1000 MW. The flows on the links from bus 3 are now reversed.

The prices at the buses include a positive congestion component at bus 2 and a negative congestion impact at bus 3, all relative to bus 1. As before, to verify the equilibrium price at bus 3, given the prices at buses 1 and 2, note that 1 MW of additional load (or an equivalent reduction in generation) at bus 3 would be met by an increment of 2 MW in generation at bus 1 and a reduction of 1 MW of generation at bus 2. The net change in generation cost would be 2*1 - 1.3 = 0.7 units. The flow on line 2→3 would be unchanged, but the reverse flow on line 1→3 would be reduced by 1 MW and the flow along 1→2 would increase by 1 MW. Since these lines have different marginal losses, there would be a net increase in losses of 0.025. Hence, the total increase in cost for an additional MW at bus 2 would be 0.7 + 0.025 = 0.725, the price at bus 3.

Once again, the users of the grid pay or are paid according to these short-run marginal cost prices. The pool collects the payments and, in turn, makes the necessary payments to the holders of the transmission congestion contracts. In this case, both the customers with congestion contracts to bus 1 and those with contracts to bus 2 face negative congestion rent differentials. Hence the customer with congestion contracts of 800 MW from bus 1 sees a differential of -0.25, and makes a total additional payment to the pool of 200. With the purchase of 800 MW at bus 3 at the price of 0.725, these combined payments are equivalent to a purchase of 800 MW at bus 1 and then moving to bus 3 at the cost of marginal losses.

For the customer with a 200 MW contract to bus 2, the congestion price difference is -0.5, and the direct payment to the pool is 100. These payments from the contract holders to the pool add to the total congestion rentals collected by the pool from the actual users of the grid, who pay -800*(-0.25) + 1000*0.25 = 450. In all, as summarized in Table VI, there are 750 units of excess congestion rentals. As before, these congestion rentals could be distributed according to the share in the fixed cost allocation. In the present example, this would provide a payment of 600 to the customer with transmission congestion contracts of 800 MW from bus 1, for net receipts of 400; and 150 for the customer with contracts of 200 MW at bus 2, for net receipts of 50.

Table VI: Constraint with a Reversal in Loads

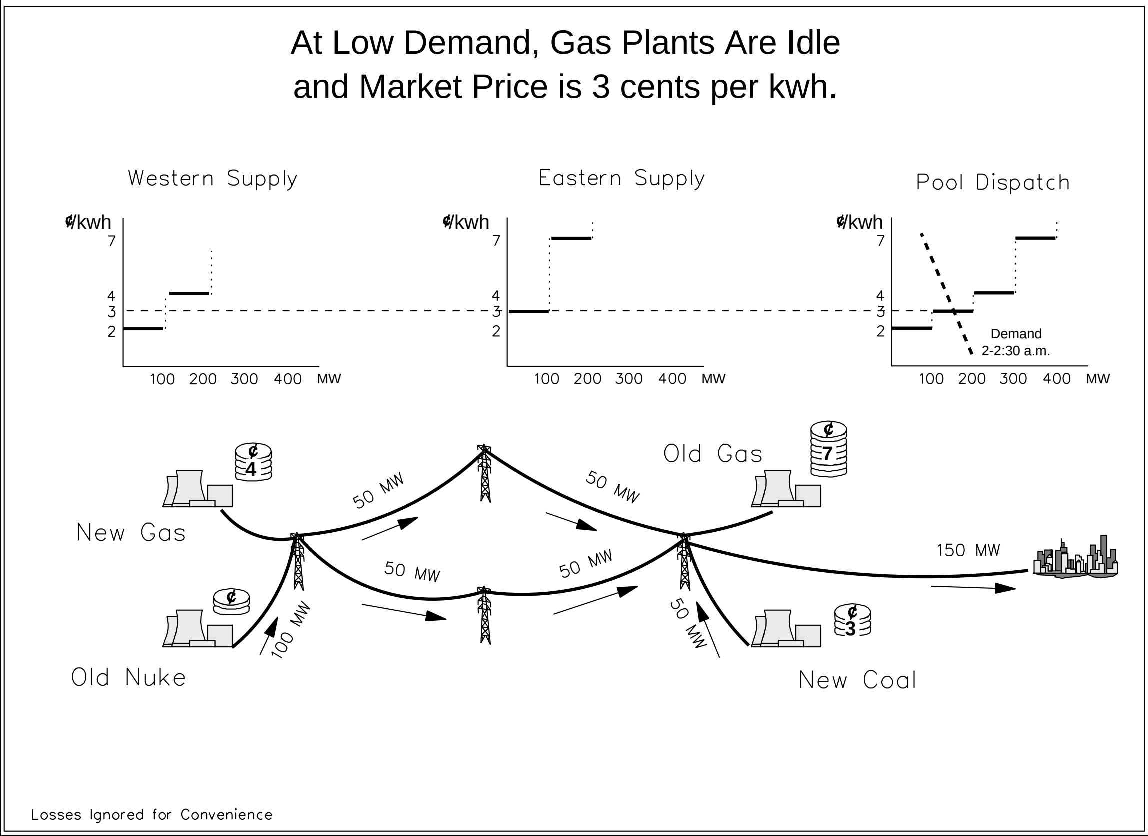
Congestion Contracts	Q (MW)	Congestion Price Difference	Direct Receipts	Excess Share	Excess Rentals	Net Receipts
1->3	800	-0.25	-200	80.00%	600	400
2->3	200	-0.5	-100	20.00%	150	50
Total			-300		750	450

In all cases, the net effect of economic dispatch, marginal cost pricing, and assignment of transmission congestion contracts is to collect congestion costs from the actual users of the grid and pay the congestion costs to those who bear the burden of the fixed charges. The transmission congestion contracts based on price differences compromise two forms. First, point-to-point transmission congestion contracts can be offered which provide the economic equivalent of a customer at one location always having the effective contract to buy delivered power at the cost at a distant location plus the marginal losses. Second, to the extent that there are excess congestion rentals, these rentals can be distributed according to some agreed formula. In aggregate, the congestion rentals paid are always adequate to honor the point-to-point transmission congestion contracts, and sometimes there can be additional rents that could be dispersed according to a sharing formula. In some instances, the congestion payments under the point-to-point contracts can be negative, but only when the economics of the dispatch have switched to provide the contract holder, who has access to cheap generation, the money from an operating margin through the pool dispatch that can fund the congestion payments. The transmission congestion contract is analogous to a futures contract which provides a perfect hedge for the cash market in transmission. The pool dispatcher and operator of the settlements system is taking no financial risk in providing these price guarantees, and the actual dispatch is not constrained by the transmission congestion contracts. The dispatcher always has the freedom to provide the most economical generation possible given the current costs, bids, and system constraints.

Economic Dispatch on a Grid

The three bus and three line example is the minimum network needed to illustrate the effects of loop flow and the impact of locational prices in a network. The results for the three link loop can be quite different than those found for a single transmission line or a radial connection. Analogies built on the case of a single line can be misleading. However, the analysis of the three bus case extends to more complicated networks, with one additional and important amendment. In the three bus case, it may be easy to fall into the trap of assuming that transmission congestion contracts are connected with individual lines between buses, since there is no difference in the geography of point-to-point definitions when every bus is connected to every other by a direct line. In a more complicated network, the transmission congestion contracts can be defined quite separately from the map of the individual lines.

Figure 16



Consider the simple market model in Figure 16, which will serve as the starting point for a set of a succeeding examples for a grid that move to a grid with multiple loops. In this market there is one load center, a city in the east, supplied by generators located far away in the west, connected by transmission lines, and by local generators who are in the same region as the city customers. The plants in the west consist of an

"Old Nuke" which can produce energy for a marginal cost of 2 ¢/kWh and a "New Gas" plant that has an operating cost of 4 ¢/kWh. These two plants each have a capacity of 100 MW, and are connected to the transmission grid which can take their power to the market in the east.

The competing suppliers in the east are a "New Coal" plant with operating costs of 3 ¢/kWh and an "Old Gas" plant that is expensive to use with a marginal cost of 7 ¢/kWh. Again these eastern plants are assumed to have a capacity of 100 MW. The two plants in the west define the "Western Supply" curve, and the two plants in the east define the corresponding "Eastern Supply" curve. These supply curves could represent either engineering estimates of the operating costs or bids from the many owners of the plants who offer to generate power in the competitive market. For simplicity, we ignore transmission losses and assume that the same supply curves apply at all hours of the day.

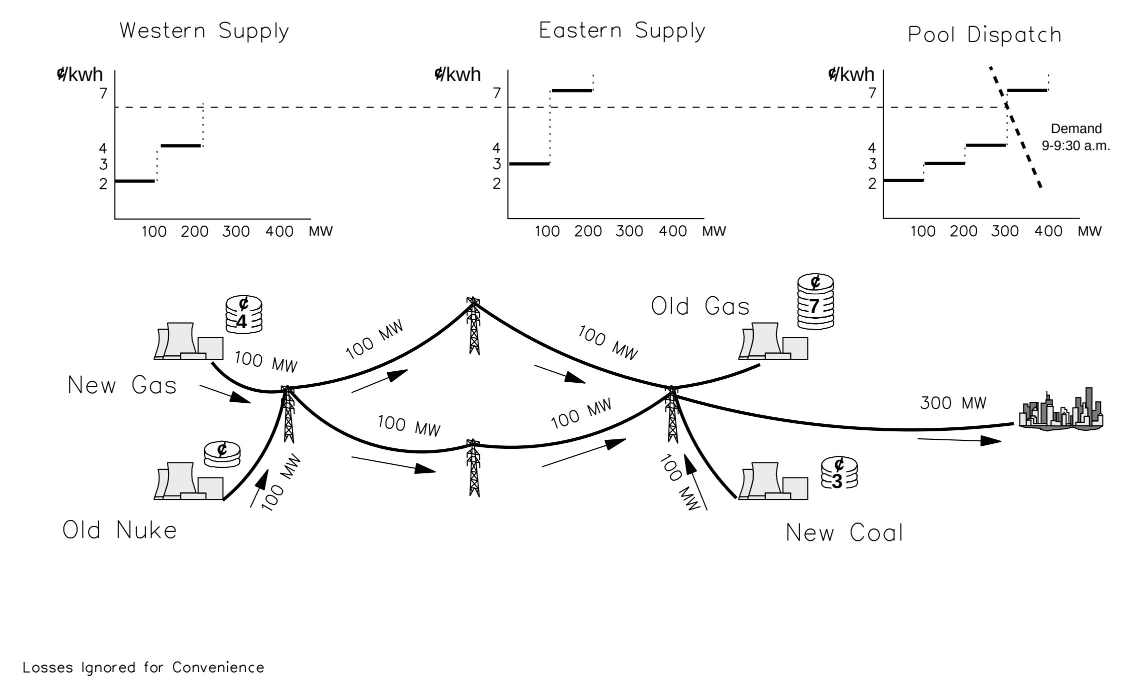
Under low demand conditions, as shown in Figure 16 for the early hours of the morning, the supply curves from the two regions define an aggregate market supply curve that the pool-based dispatchers can balance with the customer demands. The aggregate market supply curve stacks up the various generating plants from cheapest to most expensive. The pool-based dispatchers choose the optimal combination of plants to run to meet the demand at this hour. In Figure 16, the result is to provide 150 MW. The inexpensive Old Nuke plant generates its full 100 MW of capacity, and the New Coal plant provides another 50 MW. The New Coal plant is the marginal plant in this case, and sets the market price at 3 ¢/kWh for this hour. Hence the customers in the city pay 3 ¢/kWh for all 150 MW. The New Coal plant receives 3 ¢/kWh for its output, and this price just covers its running cost. The Old Nuke also receives 3 ¢/kWh for all its 100 MW of output. After deducting the 2 ¢/kWh running cost, this leaves a 1 ¢/kWh contribution towards capital costs and profits for Old Nuke owners.

In this low demand case, and ignoring losses, there is no additional opportunity cost for transmission. The 100 MW flows over the parallel paths of the transmission grid. But there is no constraint on transmission and, therefore, no opportunity cost. Hence the price of power is the same in the east and in the west. In the short run, there is no charge for use of the transmission system.

If demand increases, say at the start of the business day, the system operator must move higher up on the dispatch curve. For example, consider the conditions defined in Figure 17. This hour presents the same supply conditions, but a higher demand. Now the pool-based dispatchers must look to more expensive generation to meet the load. The Old Nuke continues to run at capacity, the New Coal plant moves up to its full capacity, and the New Gas plant in the west also comes on at full capacity. The New Gas plant in the west is the most expensive plant running, with a marginal cost of 4 ¢/kWh. However, this operating cost cannot define the market price because at this price demand would exceed the available supply, and the system operator must protect the system by maintaining a balance of supply and demand.

Figure 17

At Morning Demand, New Gas Plant is Dispatched
and Market Price is 6 cents per kwh.



In this case, the result is to turn to those customers who have set a limit on how much they are willing to pay for electric energy at that hour. This short-run demand bidding defines the demand curve which allows the system operator to raise the price and reduce consumption until supply and demand are in balance. In Figure 17 this new balance occurs at the point where the market price of electricity is set at 6 ¢/kWh. Once again, the customers who actually use the electricity pay this 6 ¢/kWh for the full 300 MW of load at that hour. All the generators who sell power receive the same 6 ¢/kWh, which leads to operating margins of 2 ¢/kWh for New Gas, 3 ¢/kWh for New Coal, and 4 ¢/kWh for Old Nuke.

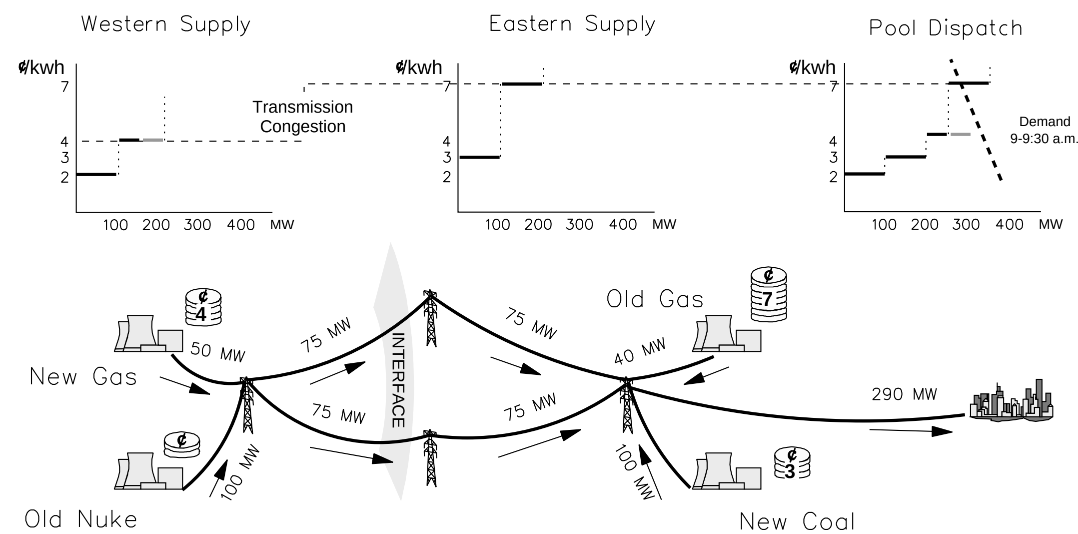
Once again, the pool-based dispatch in Figure 17 depends on excess capacity in the transmission system. The plants in the western region are running at full capacity, and the full 200 MW of power moves along the parallel paths over the grid to join with New Coal to meet the demand in the east. There is a single market price of 6 ¢/kWh, and there is no charge for transmission other than for losses, which are ignored here for convenience in the example.

Transmission Constraints

With the plants running at full capacity, there might be a transmission constraint. To illustrate the impact of a possible transmission limit, suppose for sake of discussion that there is an "interface" constraint between west and east. According to this constraint, no more than 150 MW of power can flow over the interface.

Figure 18

Morning Demand & Transmission Interface Constraint Yield Congestion; Market Price is 7 cents in the East and 4 cents in the West.



Transmission Constraint Creates 3 cent Congestion Rental.

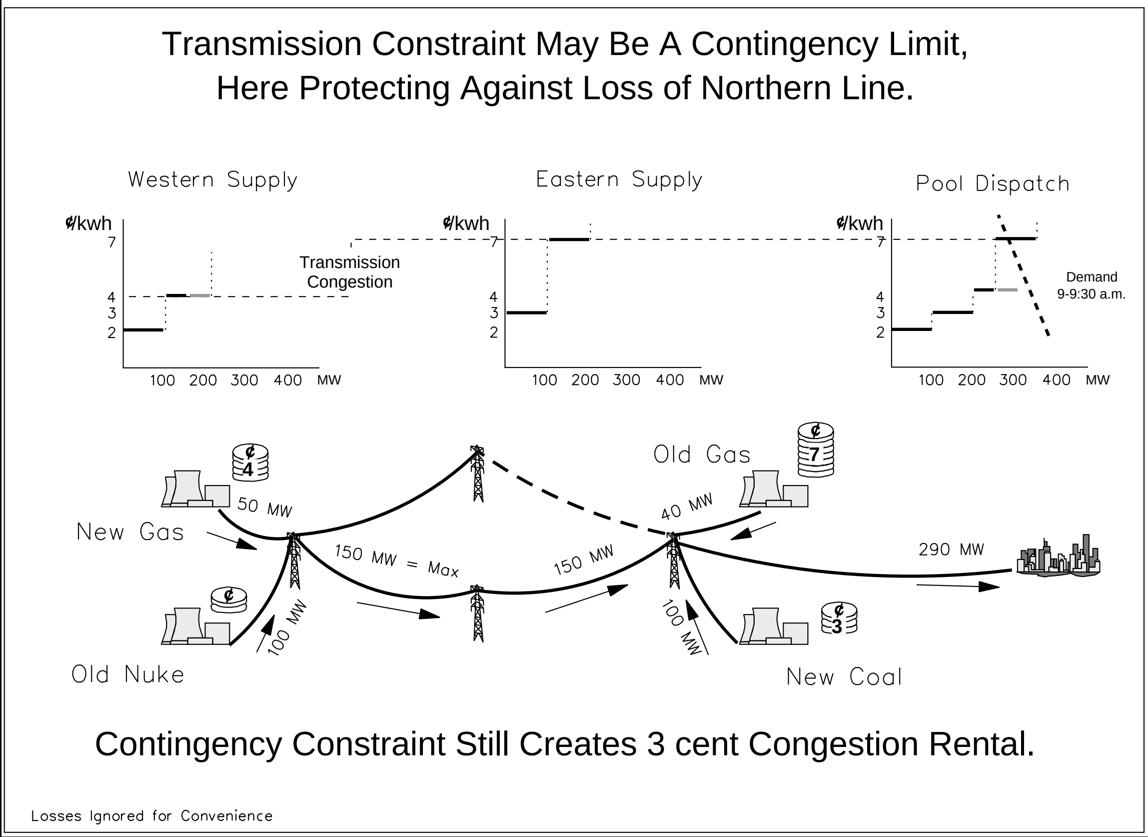
Losses Ignored for Convenience

As shown in Figure 18, this transmission constraint has a significant impact on both the dispatch and market prices based on short-run marginal costs. In Figure 18 the level of demand from the city in the east is assumed to be the same as in the case of Figure 17. However, now the pool-based dispatcher faces a different aggregate market supply curve. In effect, only half of the New Gas output can be moved to the east. To meet the demand, it will be necessary to simultaneously turn off part of the New Gas output and substitute the more expensive Old Gas generation which is available in the East. This new dispatch increases the market price in the east to 7 ¢/kWh and necessarily induces a further reduction in demand, say to a total of 290 MW. The New Coal and Old Gas plants receive this full price of 7 ¢/kWh for their 140 MW, which provides a 4 ¢/kWh operating margin or short-run profit for New Coal and allows Old Gas to cover its operating costs.

In the western region, however, a different situation prevails. The transmission interface constraint has idled part of the output of the New Gas plant. Clearly the market price in the west can be no more than the operating cost of the plant. Likewise, since the plant is running at partial output, the market price can be no less than the operating cost of 4 ¢/kWh. This is the price paid to New Gas and Old Nuke, which covers New Gas operating costs and provides Old Nuke an operating margin of 2 ¢/kWh.

The 3 ¢/kWh difference between the market price in the east and the market price in the west is the opportunity cost of the transmission congestion. In effect, ignoring losses, the marginal cost of transmission between west and east is 3 ¢/kWh, and this is the price paid implicitly through the transactions with the system operator. Electricity worth 4 ¢/kWh in the western region becomes worth 7 ¢/kWh when it reaches the eastern region.

Figure 19



The transmission "interface" constraint is a convenient shorthand for a more complicated situation handled by the pool-based dispatchers. The interface limit depends on a number of conditions, and can change with changing loads. Typically it is not the case that there is a 75 MW limit on one or both of the parallel lines through which power is flowing in the grid. In normal operation, it may well be that the transmission lines could individually handle much more flow, say 150 MW each or twice the actual use. At most normal times, the lines may be far from any physical limit. However, the pool-based dispatchers must protect against contingencies--rare events that may disrupt operation of the grid. In the event of these contingencies, there will not be time enough to start up new generators or completely reconfigure the dispatch of the system. The power flow through the grid will reconfigure immediately according to the underlying physical laws. Hence, generation and load in normal times must be configured, and priced, so that in the event of the contingency the system will remain secure.

For instance, suppose that the thermal capacity of the transmission lines is 150 MW, but the pool-based dispatchers must protect against the loss of a northern transmission line. In this circumstance, the actual power flows may follow Figure 18, with 75 MW on each line, but the pool-based dispatchers must dispatch in anticipation of the conditions in Figure 19. Here the northern line is out, and in this event the flow on the southern line would hit the assumed 150 MW thermal limit. This contingency event may never occur, but in anticipation of the event, and to protect the system, the system operator must dispatch according to Figure 19 even though the flows are as in Figure 18. In either case, the transmission constraint restricts the dispatch and changes the market prices. The price is 4 ¢/kWh in the west and 7 ¢/kWh in the east, with the 3 ¢/kWh differential being the congestion-induced opportunity cost of transmission. This "congestion rental" defines the competitive market price of transmission.

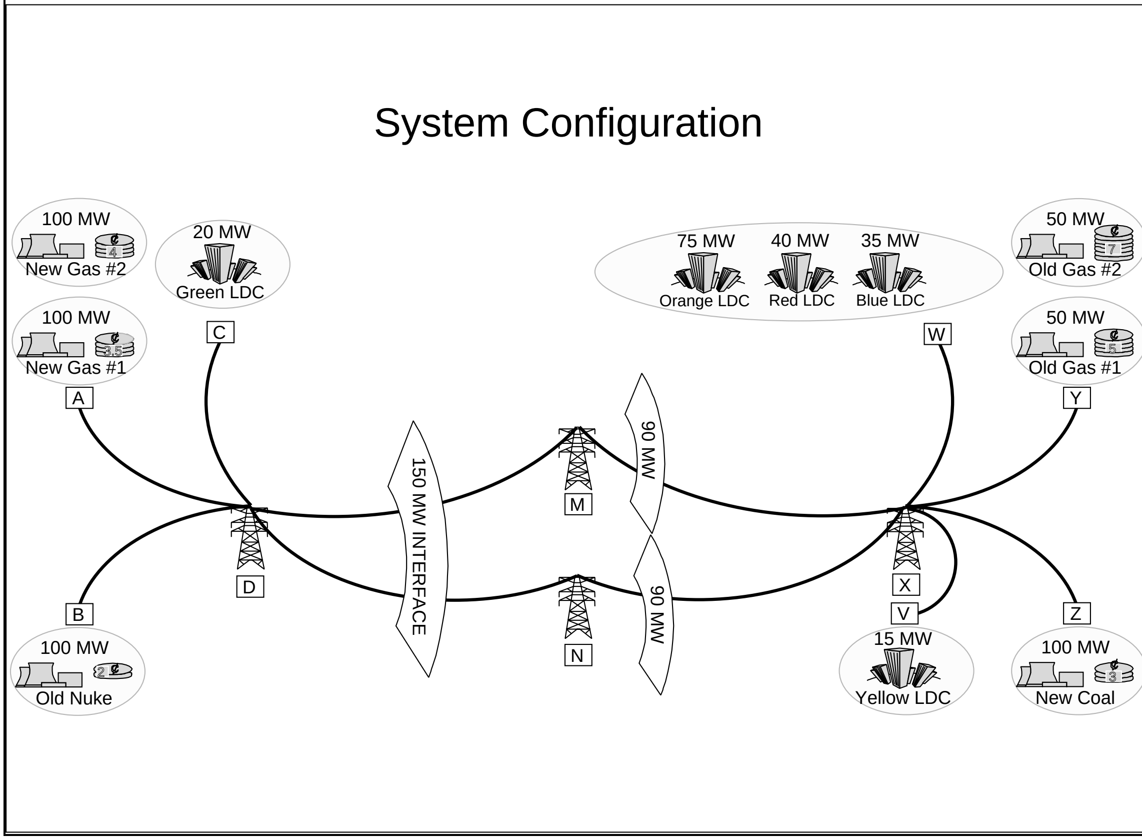
Buying and selling power at the competitive market prices, or charging for transmission at the equivalent price differential provides incentives for using the grid efficiently. If some user wanted to move power from east to west, the transmission price would be negative, and such "transmission" would in effect relieve the constraint. The transmission price is "distance- and location-sensitive," with distance measured in electrical rather than geographical units. And the competitive market prices arise naturally as a by-product of the optimal dispatch managed by the system operator.

The simplified networks in Figure 16 through Figure 19 illustrate the economics of least-cost dispatch and locational prices. However, these networks by design avoid the complications of loop flow that can be so important in determining prices and creating the difficulties with physical transmission rights. The extension of these examples and the basic pricing properties to more complicated networks includes the possibility of inputs and load around loops in the system. Here assume a transmission system as before but with the basic available generations and loads as shown in Figure 20. These generators define a basic supply configuration with quantities and prices, coupled with the associated loads, and all with the following characteristics:

- Generation available at four locations in the East (Y, Z) and West (A, B).

- Load in the East, consisting of the Yellow LDC at V and the Orange, Red and Blue LDCs at W.

Figure 20



- Load in the West, consisting of a Green LDC at C.

- Interface constraint of 150 MW between bus D and buses M and N.

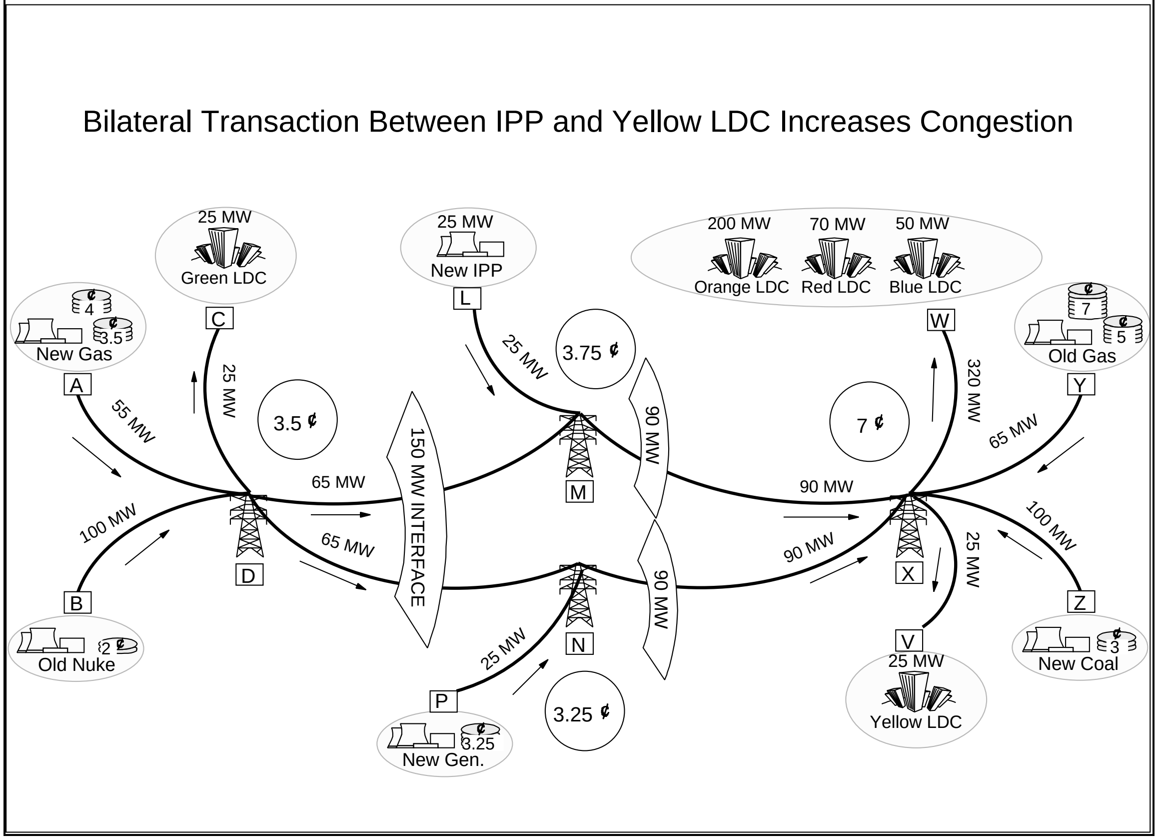
- Thermal constraints of 90 MW between M and X and between N and X.

- The New Gas and Old Gas generating facilities each consist of two generating units whose marginal costs of production differ.

Loads in Figure 20 are illustrative and will vary systematically in each example. For convenience, losses are ignored in all examples.

The first example to introduce the effect of loop flow involves a new sources of supply at a location on the loop. Here a low cost, large capacity generator becomes available in Figure 21 at bus "P." An IPP at bus "L" has bid in a must run plant at 25 MW, having arranged a corresponding sale to the Yellow distribution company at bus

Figure 21

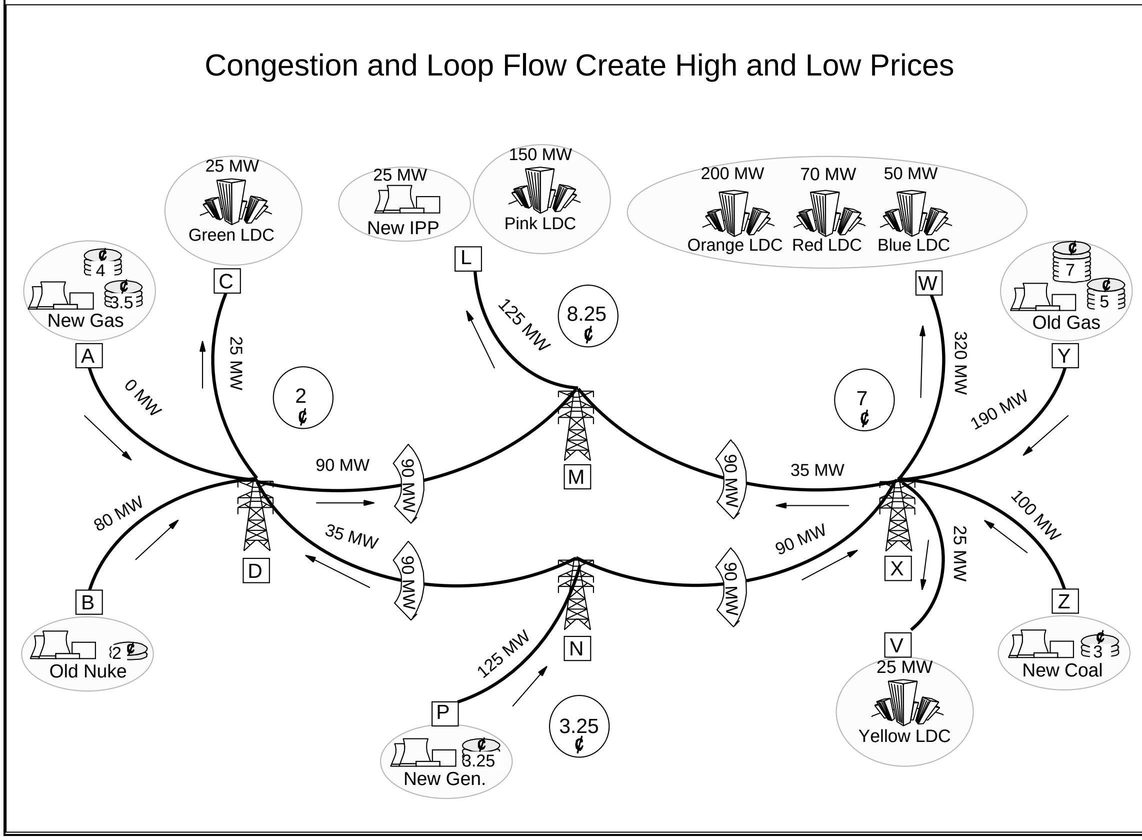


"V". Were it not for the IPP sale, more power could be taken from the inexpensive generators at bus "P" and at bus "A". However, because of the effects of loop flow, these plants are constrained in output, and there are different prices applicable at buses "D", "M", "N", and "X".

In a further example the constraints are modified to replace the interface limit with limits on the flows on individual lines. Here every line in the main loop is constrained by a thermal limit of 90 MW, replacing the interface limit. With these constraints in Figure 22, an added load of 150 MW at bus "L" alters the flows for the market equilibrium. In this case, the combined effect of the increased load and the constraints leads to a price of 8.25¢ per kWh at bus "L". This illustrates that it is possible to have market clearing prices at some locations that are higher than the 7¢ marginal running cost of the old gas plant at bus "Y", the most expensive plant in the system. The interaction of the network constraints is such that with a reduction of load at bus "L" it would be possible to reduce out put of the most expensive plant by even more, and make up the difference with cheaper sources of supply, causing the high price for load at bus "L".

Changing the network further adds new loops and even more examples of the effect on prices and dispatch caused by the network interactions. In this case, a new line has been added to the network in Figure 23, connecting bus "N" to bus "M". This line is

Figure 22

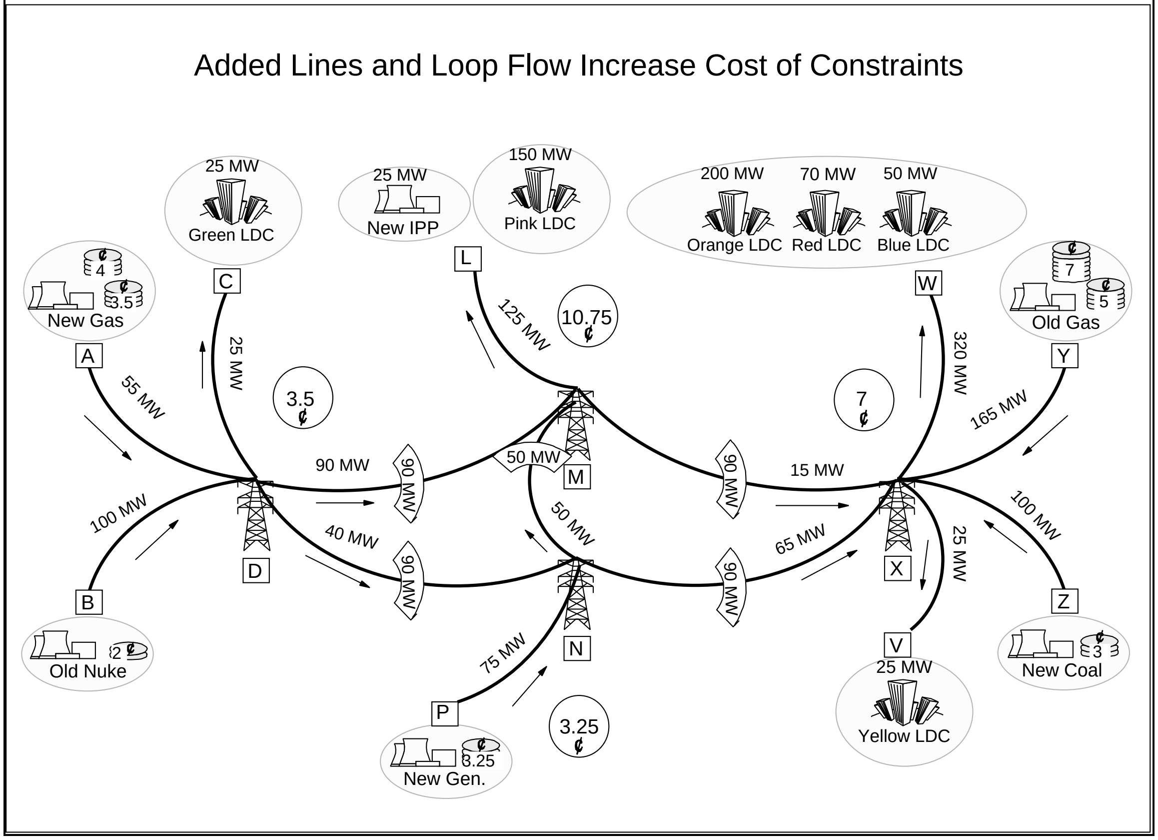


assumed to have a thermal limit of 50 MW. The new line adds to the capability of the network in that the new pattern of generation lowers the overall cost of satisfying the same load. The total cost reduces from $20,962.50 in Figure 22 to $19,912.50 in Figure 23. Although the average cost of power generation fell, the marginal cost of power increased at bus "L", where the price is now 10.75¢ per kWh. The new loop provides more options, but it also interacts with other constraints in the system. This set of interactions is the cause of the high price as it appears at bus "L".

As a final example that confirms the sometimes counterintuitive nature of least-cost dispatch and market equilibrium prices, add a new bus "O" between bus "M" and bus "N" in Figure 24, and lower the limit to 30 MW between bus "O" and bus "M". Bus "O" has a small load of 15 MW. The increased load of 15 MW at bus "O" actually lowers the total cost of the dispatch, as reflected in the negative price. Each additional MW of load at bus "O" changes the flows to allow a dispatch that lowers the overall cost of meeting the total load. The optimal solution would be to pay customers at "O" to accept dump power, thereby relieving congestion elsewhere and providing benefits to the overall system.

This final example, therefore, illustrates and summarizes the types of interactions

Figure 23

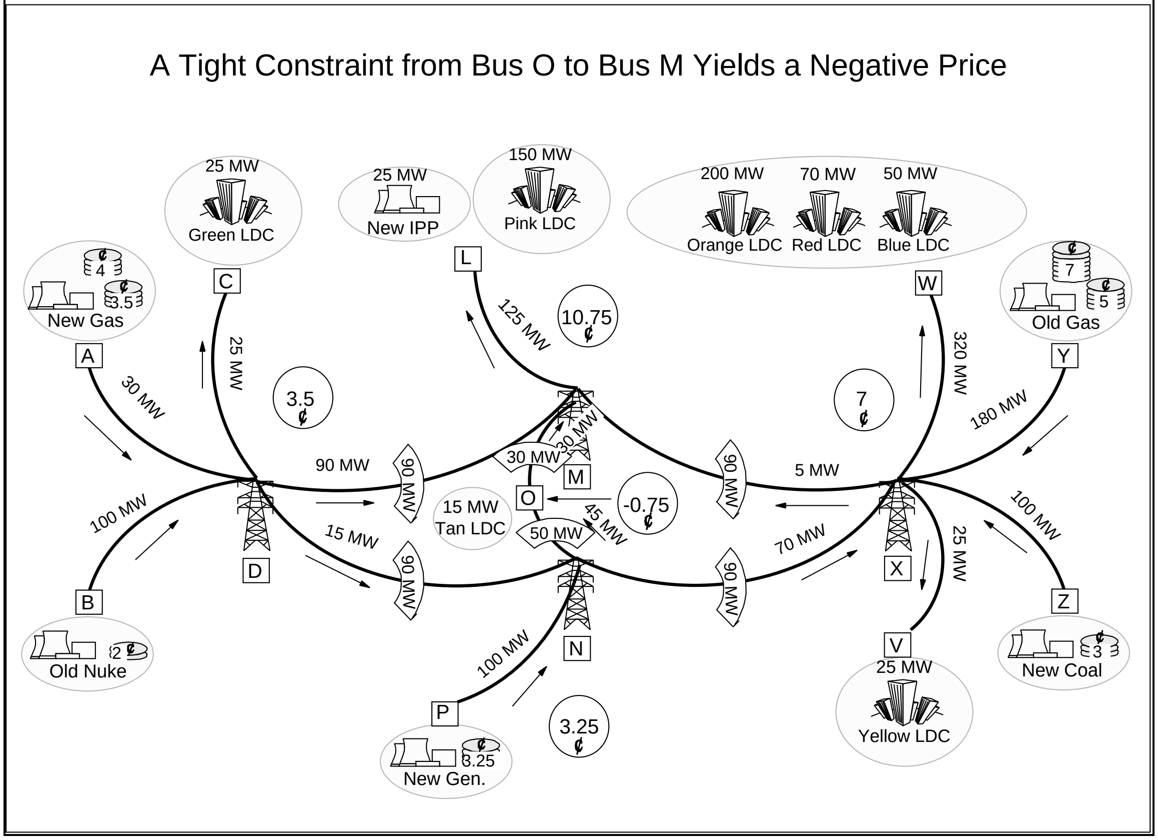


that can develop in a network with loop flow. Power can flow from high price nodes to low price nodes. The competitive market clearing price, equivalent to the marginal costs for the least-cost dispatch, can include simultaneously at different locations prices higher than the cost of the most expensive generation and lower than the cost of the cheapest generation source.

Zonal Versus Nodal Pricing Approaches

Application of the principle of locational pricing implies that transmission congestion would lead to many prices. Even with only a single constraint, there could be a different price at each location. The use of locational prices has been described as being too complex, with the implication that an alternative approach would produce a simpler system. A common response to this assertion has been to recommend a "zonal" approach that would aggregate many locations into a smaller number of zones. The assumption has been that this would tend to reduce complexity. However, in the presence of real constraints in the actual network, the zonal approach may not be as simple as it

Figure 24



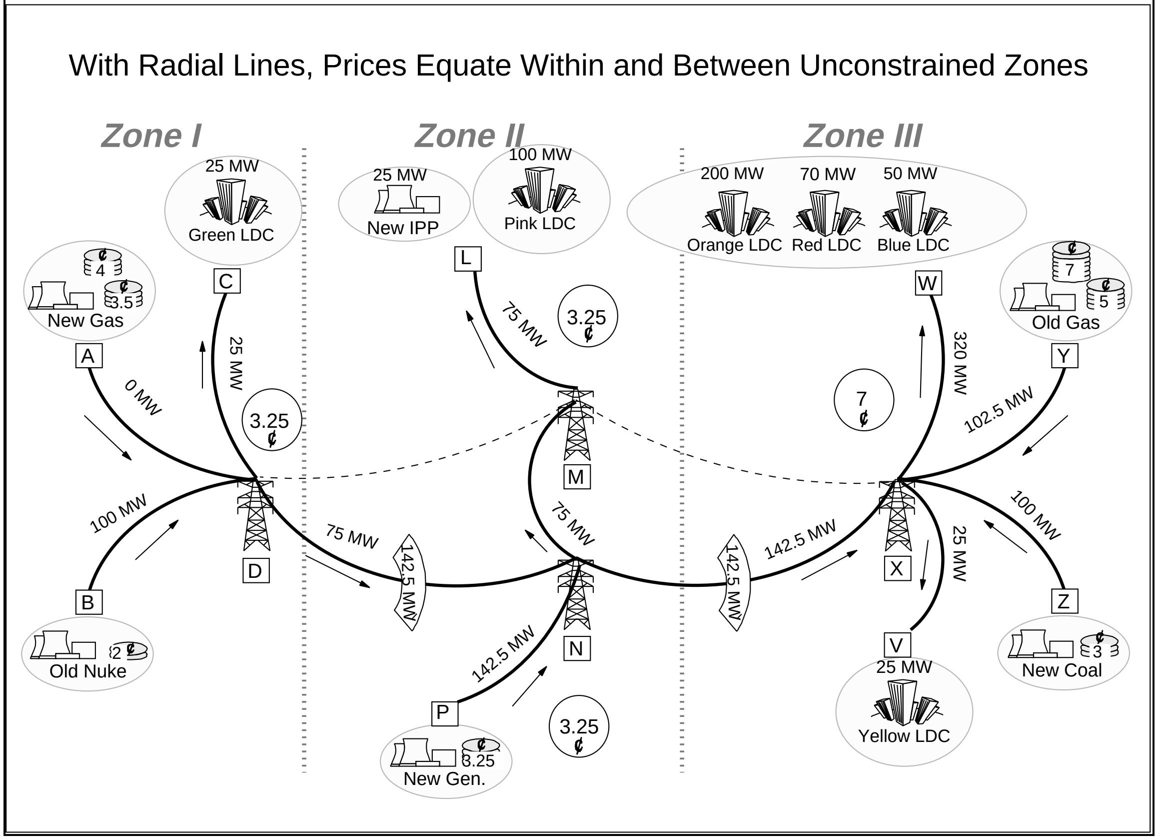
might appear without closer examination.[13]

The difficulties would arise in the context of a competitive market where participants have choices. If the actual operation of the network system does not conform to the pricing and zonal assumptions, there will be incentives created to deviate from the efficient, competitive solution. In the presence of a vertical monopoly that can ignore the formal pricing incentives, this has not been a problem. But under the conditions of a market, where participants will respond to incentives, the complications created by a zonal approach may be greater than any complications that would exist with a straight locational approach to pricing and transmission charging.

Consider the simplified example in Figure 25. The network has been constructed so that there are only radial connections. With strictly radial connections, locations within and between unconstrained zones would have a common price. Hence, aggregation of locations offers an apparent simplification by reducing to a few distinct zones. This motivation from a typical radial examples leads to the assumption that in general there could be areas in a real network that would have the same prices and, therefore, these

[13] For a similar analysis with similar conclusions, see Steven Stoft, "Analysis of the California WEPEX Applications to FERC," Program on Workable Energy Regulation, University of California, October 8, 1996.

Figure 25



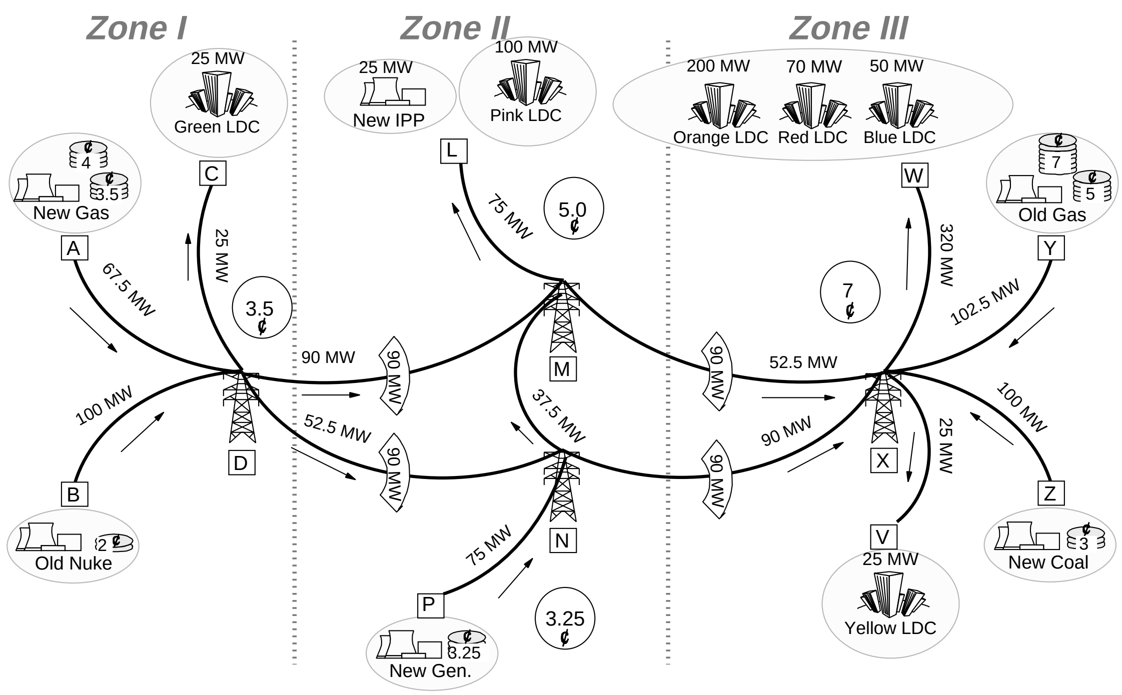
locations could be aggregated into zones that would be simpler for participants in market operations.

There are two problems with this line of argument. First, if the multiple locations truly do have the same prices, then there is no need to aggregate into zones. The point of the aggregation was to reduce the number of prices, and in the case where the assumption of common prices holds, aggregation would be unnecessary.

Second, the definition of a zone, which appears easy in the case of a radial network, becomes more problematic in the case of a more realistic network with loop flows. The radial examples can be a poor guide to thinking about interactions in networks. For example, it is often argued, or assumed, that congestion or differences in prices between zones would be caused only by transmission constraints that could be defined for lines that connect the zones. Furthermore, it is often assumed that differences in prices within zones can only be caused by congestion on lines within the zone. Under these simplifying assumptions, therefore, it is assumed that zones can be well defined and that what happens within a zone can be treated independently of what happens between zones, or independently of what happens in other zones. When we move beyond the radial examples, however, these assumptions and the associated conclusions can be false.

Figure 26

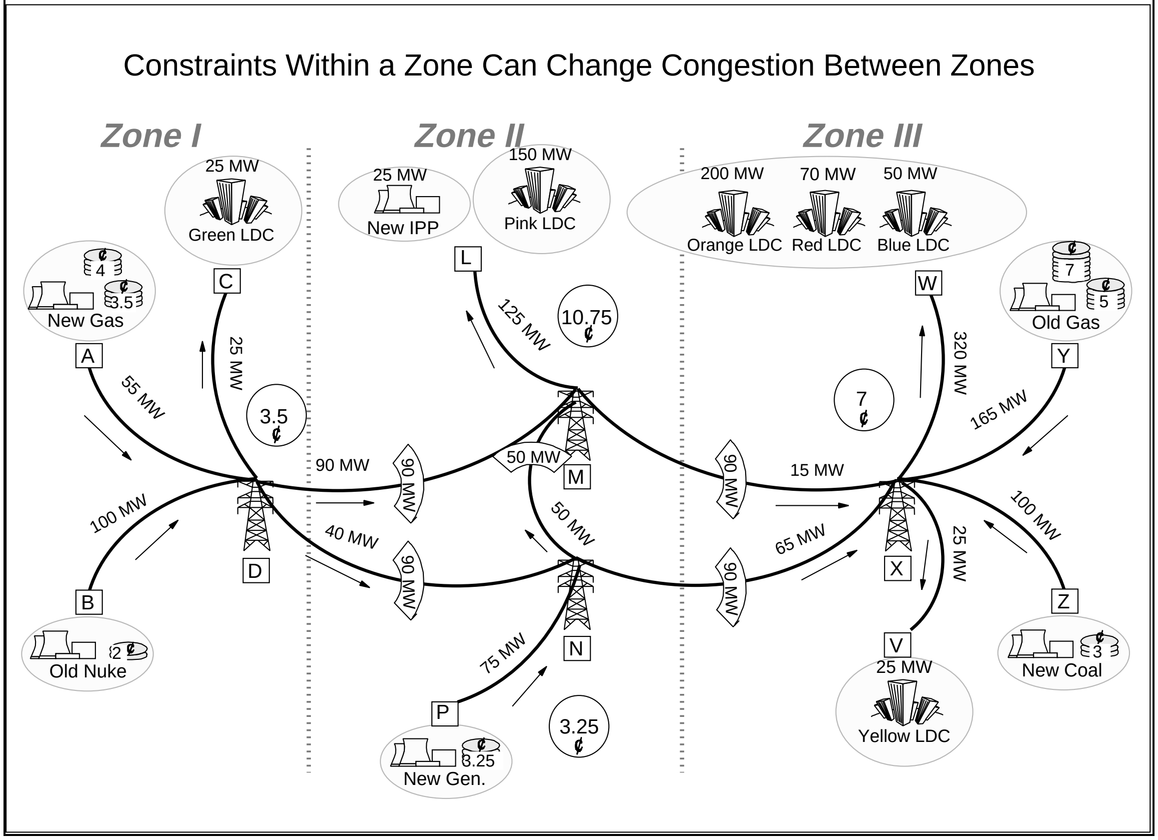
Constraints Between Zones Can Change Congestion Within Zones



With the more typical case of loops in a network, prices could differ within and between "unconstrained" zones due to the indirect effects of "distant" constraints. Consider the slightly modified example in Figure 26. In this case, the zones developed from the radial analogy produce a very different outcome from the assumptions derived from the radial case in Figure 25. In this example, the prices within "Zone II" differ, but there is no binding constraint in the zone. The lines within the zone are operating below their thermal limits. The difference in prices between buses M and N arises not due to constraints within the zone but because of the loop flow effects interacting with the binding constraints between the zones. Apparently the determination of prices within a zone can not be made independent of the effects on constraints outside the zone.

A symmetric result appears in Figure 27 with a different pattern of loads and flows. In this case, there is no constraint binding between Zones II and III, but the price in Zone III differs from the prices in Zone II. Again this effect cannot be seen in radial networks, but it is easy to create in real networks with loop flow. The price in Zone III differs from all the other prices in part because of the interaction with the constraints in Zone II. In a sufficiently interconnected network, these examples suggest that a wide variety of pricing patterns would be possible. In fact, with loop flow, it is possible for a single binding constraint to result in different prices at every location in the system, reflecting

Figure 27



the fact that every location has a different impact on the constraint.

Aggregation into zones may add to complexity and distort price incentives. The assertion that conversion to zones will simplify the pricing problem is not supported by analysis of the conditions that can exist in a looped network. Furthermore, aggregating networks presents a number of related technical problems that follow from the fact that exact aggregation requires first knowing the disaggregated flows. In other words, the first step in calculating consistent aggregate flows and prices is to calculate the disaggregated flows and prices. Hence aggregation produces no savings in computation, and no additional simplicity. If no price dispersion exists, no aggregation is necessary. And if price dispersion does exist, aggregation only sends confused price signals. In the end, the simplest solution may be to calculate and use the locational prices at the nodes, without further aggregation.

Transmission Congestion Contracts

The congestion rental received by the system operator provides the key to defining property rights in the transmission grid. In the face of transmission constraints, prices will be more volatile and it will not be possible for a generator to provide guaranteed price stability in the form of a long-term contract with a customer. Furthermore, customers and generators will have an incentive to expand the grid only if they recognize and believe that after making an investment in the grid there will be some protection against any future congestions costs. A simple way to define the property right and provide this guarantee is to assign the congestion rental not to the system operator but to the holder of the transmission congestion contract. In the transmission constrained cases of Figure 18 or Figure 19, the congestion contracts for a total 150 MW of power might have been held by customers in the city, or by generators in the west. In either case, this right is defined only as the right to collect the congestion rental. The generators and customers would not control the use of the grid. The system operator would determine the efficient pattern of use through economic dispatch. The system operator would collect the congestion payments from the actual users of the grid and pay them in turn to the holders of the transmission congestion contracts.

With this definition of transmission congestion contracts, it is an easy matter for generators at Old Nuke and New Gas to arrange long-term contracts that provide price stability for customers in the city. For example, the owners of Old Nuke may have acquired power contracts for 100 MW, and signed long-term contracts that guaranteed to provide power delivered to the city at a price of 5 ¢/kWh. In the case of low demand as in Figure 16, the short-run price is only 3 ¢/kWh, which customers pay and generators receive through the system operator. Separate from the system operator, the customers pay Old Nuke the difference of 2 ¢/kWh owed under the contracts. If demand shifts to the higher case in Figure 17, the market price is 6 ¢/kWh, and again the customers pay and generators receive this short-run price through the system operator. In this event, the generators separately pay the customers the difference of 1 ¢/kWh required under the long-term contract.

When transmission constraints bind as in Figure 18 or Figure 19, the price paid by the customers to the system operator is 7 ¢/kWh, and the price received by the generators from the system operator is 4 ¢/kWh. If the generators own the transmission congestion contracts, then the system operator pays the generators an additional 3 ¢/kWh which allows the generators in turn to pay the customers the 2 ¢/kWh difference agreed to by contract. The owners of Old Nuke are always making an operating margin of 3 ¢/kWh, and the customers are always receiving the equivalent of 5 ¢/kWh electricity. Likewise, if the customers own the transmission congestion contracts, the customers receive the 3 ¢/kWh from the system operator and in turn pay 1 ¢/kWh to the generators. Again the owners of Old Nuke are always making an operating margin of 3 ¢/kWh, and the customers are always receiving the equivalent of 5 ¢/kWh electricity. Furthermore, in either case the system operator ends up with no transmission congestion rentals; the system operator serves only to pass through the congestion costs from the actual users of the grid to the holders of the property rights in the economic interest of the transmission grid.

This system of transmission congestion contracts and payments back and forth may seem unnecessary and cumbersome in the case of the simple system of Figure 16 through Figure 19. After all, couldn't the pool-based dispatchers in effect assign the generation from Old Nuke to the long-term customers in the city? In principle, this specific performance model--assigning particular generation to particular users--is possible in this simple case, but it does not generalize into the more complicated reality of an interconnected grid with many different sites of load and generation, real and reactive power, thermal and voltage limits, and multiple contingencies. The examples in Figure 21 through Figure 24 illustrate the difficulties attendant to the network interactions. It is impossible in a real system to meaningfully assign any particular sources and destination of electricity, and attempts to do so can only serve to compromise the efficiency objective of maintaining an optimal dispatch which may require only partial use of plants in constrained regions, violating the assumptions of specific performance. However, the payments of congestion rentals from the system operator to the holders of point-to-point transmission congestion contracts do generalize to the more complicated case, and allow optimal dispatch for efficiency while accommodating long-run contracts for price differences and congestion rentals, contracts that provide both stability and the essential protection of investment in the network.

These point-to-point price protection transmission contracts defined in alternative equivalent ways, with various advantages for implementation and interpretation. For example:

- **Difference in Congestion Costs**. Receive the difference in congestion costs between two buses for a fixed quantity of power.

- **Purchase at a Distant Location**. Purchase a fixed quantity of power at one location but pay the price applicable at a distant location.

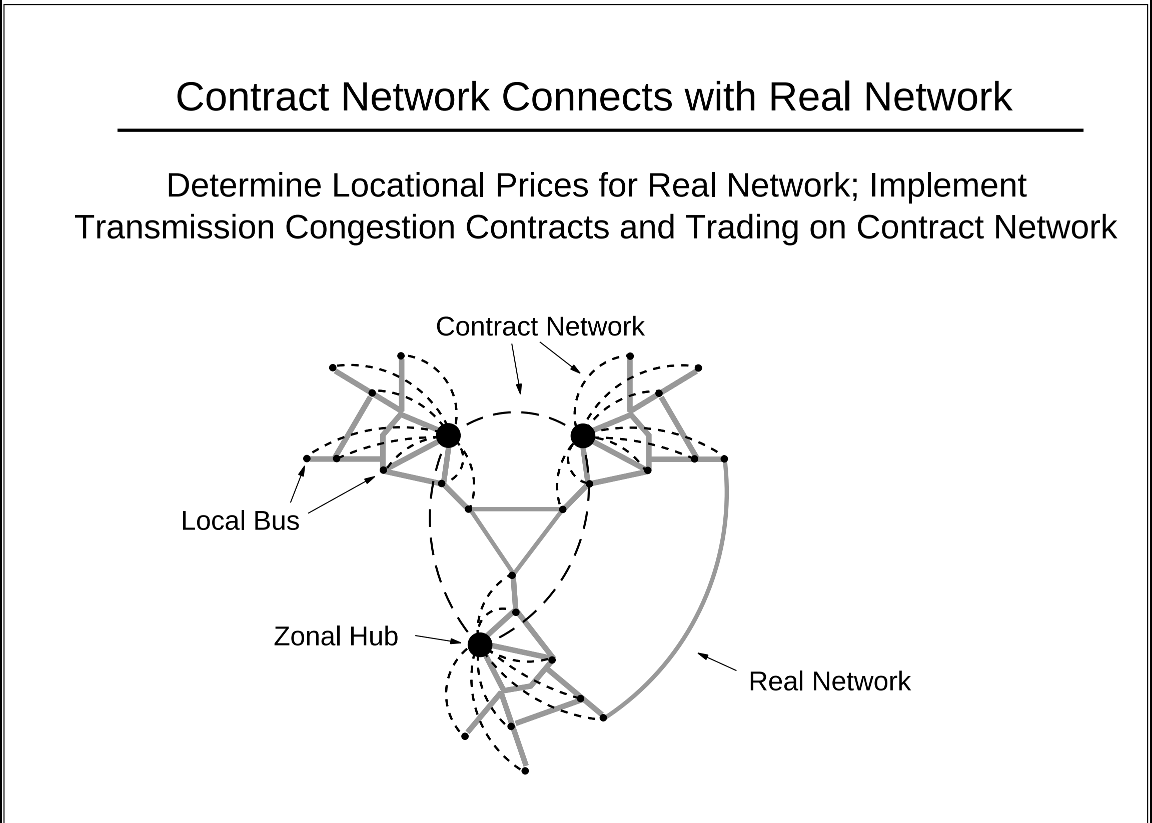
- **Dispatch with No Congestion Payment**. Inject and remove a fixed quantity of power without any congestion payment.

The total quantity of these transmission congestion contracts can be defined for a given configuration of the network, and the congestion contracts guaranteed for any pattern of loads in the network. In a real system, the transmission congestion contracts would respect all the constraints in the grid, including contingency constraints for thermal limits on lines and voltage limits at buses.

The point-to-point congestion contracts can be provided by the pool and allow protection of new investment in generation and transmission. The pool takes no risk in offering these transmission congestion contracts because the revenue collected from users paying at short-run marginal cost is always great enough to honor the obligations under the congestion contracts. Under certain circumstances, the revenues collected by the grid will be greater than the obligations on the point-to-point congestion contracts, and there will be "excess congestion rentals." For example, if there are no point-to-point contracts assigned, then none of the congestion rental would be paid out under such contracts. In the more interesting cases, if point-to-point congestion contracts are assigned up to the limit of some transmission constraint, it is possible that some other network constraint will be binding in the actual dispatch. In these cases, there will be more than enough revenues to honor the rental or purchase rights, and the pool will face the added task of allocating the excess congestion rentals.

The natural assignment of any excess rentals is to those who are paying the fixed charges. There is no single rule for allocating the rights to the excess congestion rentals. The rule examined here is to share the revenues according to the share of the fixed charges. Hence, the owners of the grid who commit to pay the fixed charges have access to two types of well defined and tradeable transmission congestion contracts: the point-to-point rental contract and a share in any excess congestion rentals.

Figure 28



As shown in Figure 28, the contract network must anchor to the same locations, but the point-to-point contracts can follow a very different geography. Market hubs can arise and be included, with the contract connections in the network following a configuration convenient for contracting and trading. The separation of the physical and the financial flows allows this flexibility with the congestion revenues always sufficient to cover the obligations under transmission congestion contracts, no matter what the resulting pattern that appears in the least-cost dispatch.

To illustrate this conclusion for a more general network, consider again the example network in Figure 24. For this example, consider the extreme case where the market has elected to use bus "O" as the market hub, with transmission congestion contracts all defined relative to this hub. Generators may have the contracts to get to the market at "O". Customers may have similar contracts to get from the market at "O" to their own locations. The individual transmission congestion contracts may embody flows which would never be individually feasible, especially given the limits on the lines connecting "O" and the unusual conditions in this extreme case. As long as the collective flows under the contracts would be feasible, however, the congestion costs collected will always be sufficient to meet obligations under transmission congestion contracts, even in this extreme case.

Here Table VII offers with two feasible sets of transmission congestion contracts (TCCs) for a hub at "O". If the only inputs and outputs in the system consisted of those with TCC 1 at 180 MWs in at "D", 30 MWs in at "M", 30 MWs in at "N", 60 MWs out at "O" and 180 MWs out at "X", the flows would be feasible even though the individual contracts appear to require flows that are not feasible. Similarly for TCC 2. With these feasible transmission congestion contracts, alternative dispatch cases illustrate the impact of the changing congestion payments.

Table VII: Transmission Contract Allocations

From-To	TCC 1 (MW)	TCC 2 (MW)
"D-O"	180	160
"O-X"	180	160
"M-O"	30	10
"N-O"	30	70

The results summarized in Table VIII capture the outcomes for economic dispatch with the only change being three different assumptions about the load at bus "L" in Figure 24, ranging from 0 MW to 150 MWs, with all other conditions the same. The table shows the corresponding prices at key buses, ignoring losses, and the associated collection of congestion rents. These rents are compared in turn with the obligations under the point-to-point contracts of the two sets of transmission congestion contracts. In the case of no load at "L", the congestion payments amount to $6300. Under the TCC 1 the obligation is also $6300; under TCC 2 the point-to-point obligation is $5750, leaving an excess of congestion payments to be disbursed through a sharing formula.

Table VIII: Transmission Congestion Payments

Load at "L"	Bus Prices ¢/kWh					Total Rents $	TCC 1	TCC 2
MW	"D"	"M"	"N"	"O"	"X"			
0	3.50	3.75	3.25	3.50	7.00	6300	6300	5750
50	3.50	5.58	3.25	4.15	7.00	6300	6138	6084
150	3.50	10.75	3.25	-0.75	7.00	10950	1650	1650

As the load at bus "L" increases, dispatch reconfigures and prices change. Power flows are different than in the transmission congestion contracts. However, in every case and for each set of TCCs, the congestion rentals equal or exceed the obligations under the point-to-point contracts. The transmission congestion contracts can always be honored, no matter what the pattern of load. In some instances, there will be excess congestion rentals to disburse, but transmission congestion contract holders will always be able to hedge power contracts without requiring physical transmission rights and without compromising the least-cost dispatch.

- end -