
Bernard C. Lesieutre is a Staff Scientist at Lawrence Berkeley National Laboratory in Berkeley, California, and is involved with studies of the electric power grid and electricity markets. He received his Ph.D. in Electrical Engineering from the University of Illinois and served on the faculty of Electrical Engineering at M.I.T. between 1993 and 2001.

Joseph H. Eto is a staff scientist at Lawrence Berkeley National Laboratory, where he manages the program office for the Consortium for Electric Reliability Technology Solutions. In 2002, he provided technical support to the U.S.

Department of Energy for preparation of the National Transmission Grid Study.

The work described in this article was coordinated by the Consortium for Electric Reliability Technology Solutions and funded by the Office of Electric Transmission and Distribution, Transmission Reliability Program of the U.S.

Department of Energy under Contract No. DE-AC03-76SF00098. This article is based on a longer technical report titled, *Electric Transmission Congestion Costs: A Review of Recent Reports (LBNL-54049)*, available at <http://certs.lbl.gov>.

When a Rose Is Not a Rose: A Review of Recent Estimates of Congestion Costs

Careful examination of congestion costs leads to some conclusions that are counterintuitive. For example, consumer costs may rise, not fall, as a result of reducing congestion, and minimizing consumer costs may not increase aggregate social wealth.

Bernard C. Lesieutre and Joseph H. Eto

I. Introduction

Transmission congestion results when there is not enough transmission capability to support all requests for transmission services, and in order to ensure reliability, transmission system operators must re-dispatch generation or, in the limit, deny some of these requests to prevent transmission lines from becoming overloaded. In other words, transmission congestion does not refer to deliveries that are simply held up or delayed (as in traffic congestion); it refers to requests for deliveries (transactions) that cannot be physically implemen-

ted as requested. The cost of transmission congestion, assuming that demand is fixed and must be met, is the net cost of the replacement power that must be supplied by other means (e.g., from generators located closer to the loads to be served) to make up for deliveries that cannot be executed as requested.

Recent published reports by independent system operators (ISOs) and others present widely varying estimates of congestion costs. The objective of this article is to discuss how these costs are calculated and how they should be interpreted and used. Understanding these reported

costs is crucial if they are to be used in making decisions about whether and how to address transmission congestion—e.g., by expanding/reinforcing the transmission system to increase its capability to deliver electricity or pursuing non-transmission strategies that accomplish the same end, such as increasing generation capacity closer to the load it serves or engaging in demand-side management.

To describe the differences among published estimates of congestion costs, we develop and motivate three ways by which transmission congestion costs have been calculated in restructured U.S. electricity markets. The assessment demonstrates that published transmission congestion costs are not directly comparable because they have been developed to serve different purposes: either as a measure of increased costs over a hypothetical uncongested system, or the settlement of costs incurred above energy costs in the actual system. More importantly, critical information needed to make them more comparable—for example, in order to evaluate the impacts of options to relieve congestion—is sometimes not available.

There is also a slight disconnect between congestion costs as typically viewed by ISOs and the quantity economists would designate as congestion costs. The former necessarily involves the system's daily and real-time settlement policies, while the latter tends to focus on changes in social

wealth due to congestion. Unfortunately, information that is not generally available publicly is required to evaluate changes in producer and consumer surpluses, and the ISOs, which have this information, must assume their markets are competitive in estimating these quantities.

Moreover, looking only at the total of producer and consumer surpluses will miss important concerns over

There is a slight disconnect between congestion costs as typically viewed by ISOs and the quantity economists would designate as congestion costs.

distributions of gains and losses, geographically and otherwise. As we will show, a plan to increase transmission capacity along a heavily congested path will likely reduce consumer costs and generator profit in the importing region, yet it will also likely increase consumer costs and generator profit in the exporting region. Such a plan may increase the total social welfare, but it may do so by increasing consumer costs to the benefit of suppliers.

This article demonstrates that there are a number of reasonable and different ways of accounting for and viewing the effects of transmission congestion. It is

necessary, then, for planners and policymakers to understand the differences when discussing options for transmission enhancement.

II. Transmission Congestion Costs in the Electricity Industry

Transmission congestion is not a unique feature of restructured electricity markets. Transmission system operators always have had to limit power flows (and will continue to have to do so) to maintain safe operating (or reliability) margins on the electricity grid.¹ To illustrate the relationship between the production cost impacts of transmission congestion and the evaluation of options to reduce these costs, we introduce as an example a two-area system to review how the “costs” of transmission congestion arise in a vertically integrated industry structure and in market-based “restructured” energy systems.

A. Two-area example

To present the methods and to discuss the issues that will appear throughout this article, we rely on an elementary and stylized example shown in **Figure 1**. To focus on congestion issues, we neglect other transmission line usage tariffs, generator limits, and subtle price differences that result from nominal losses in the system. We do, however, consider hard transmission capacity constraints that may arise from

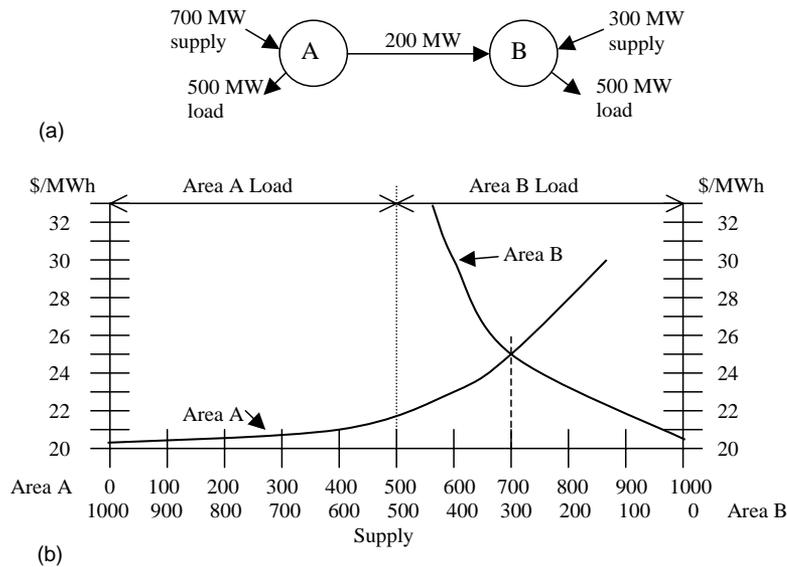


Figure 1: A two-area network (a) and corresponding aggregate supply curves for the generation in each area (b). Uncongested conditions are shown in this diagram. Without congestion, the intersection of the curves determines the dispatch, 700 MW from Area A generators and 300 MW from Area B generators

thermal, voltage, and stability considerations.

For the two-area system shown in Figure 1(a), each of the areas serves a totally inelastic load of 500 MW. A transmission line that connects the areas allows energy transfer between them. In Figure 1(b) we show the aggregate supply curves for the generators in each area; each curve plots the marginal cost of generation versus supply. The blocks of energy available to meet load are ordered from lowest price to highest price to obtain the supply curve. At any point along this curve, the marginal cost is the cost necessary to physically generate the next unit of energy for any given amount of supply. Looking from left to right on the plot in Figure 1(b), we can follow the generation marginal cost versus supply for generators in Area A. Looking from right to left, we follow the generation marginal cost versus supply for

generators in Area B.² The supply curves are presented in this fashion to illustrate the constraint that the total supply equals the total (inelastic) demand of 1,000 MW. The reader is reminded that there is 500 MW of demand in each area by labels over the top of the plot.

With no capacity limit on transmission between areas, we observe that the unconstrained, least-cost dispatch will be 700 MW for generators in Area A and 300 MW for generators in Area B. The generators in Area A serve their native load of 500 MW and export 200 MW of power to Area B.

In Figure 2, a transmission constrained system illustrates how production costs are increased due to the constraint, and how production costs would change if the constraint were relieved (i.e., if the system returned to its original, uncon-

strained state, as described in Figure 1). In this system, transfers between Areas A and B are limited to 100 MW. To respect this constraint, load in Area B must rely on more expensive generation in its own area rather than lower-cost generation that could be dispatched from Area A if there were no constraint. In this example, total production costs exceed those of the earlier example in which there was no limit on the amount of power that could be transferred from Area A to Area B. The difference in production costs between the constrained and unconstrained examples is a measure of the cost of congestion.

B. Congestion costs in a vertically integrated industry

In a vertically integrated industry, the same firm owns and operates all generation, transmission, and distribution. The cost of maintaining adequate safety margins, reflected by the 100 MW transmission constraint in Figure 2, is simply “rolled in” as one element of the cost of providing reliable electric service to the entire franchise of retail customers (in this case, all customers in Areas A and B). The shaded area under the two curves represents the total production cost of energy to supply these loads.

Figure 2 also illustrates the decrease in production costs (i.e., the production cost savings) that would result from increasing the transfer limit between Areas A and B from 100 to 200 MW. Given

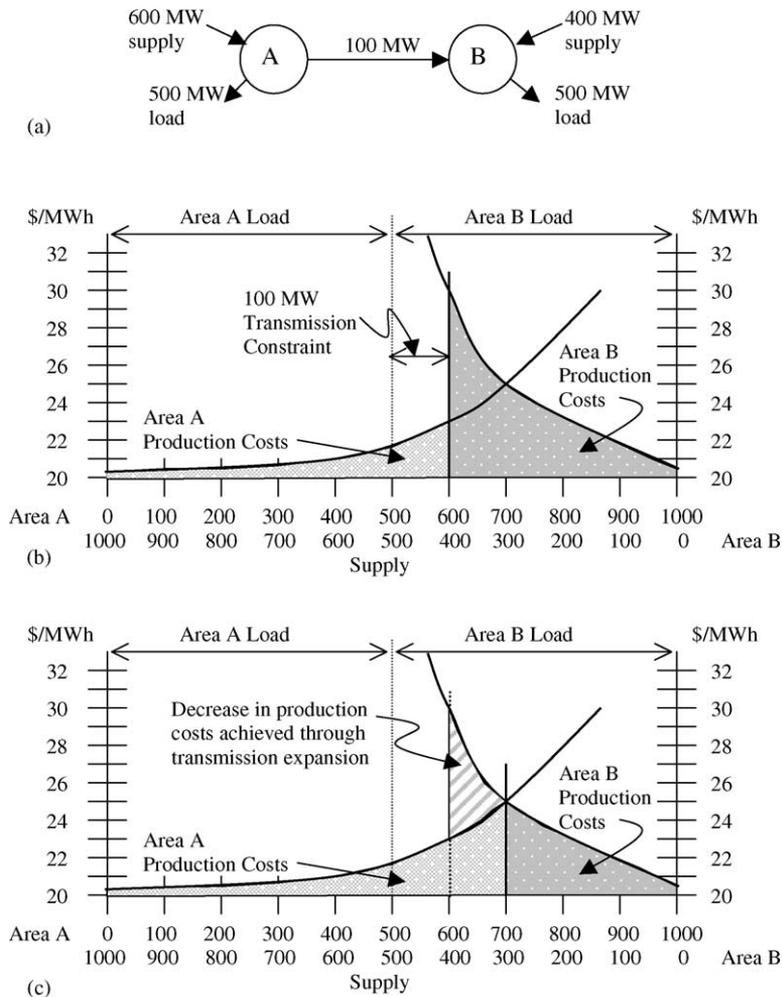


Figure 2: The shaded regions represent production costs for (b) the constrained case and (c) the unconstrained case. Note that production costs are lower in the unconstrained case

appropriate regulatory or other incentives (e.g., mandates from governing bodies), vertically integrated firms assess strategies that could lower the total cost of providing electricity service (these strategies include building new generation closer to the load served, building new transmission lines to access lower-cost generation in locations that are remote from the load served, or undertaking demand-side programs to reduce loads). The decrease in production costs illustrated in Figure 2 would be a measure of the value of one or

more of these strategies from the standpoint of reducing congestion.

The area representing the decrease in production costs also represents, in this case, the increase in consumer surplus between the congested and uncongested scenarios. From the perspective of Area A generators, the supply curve of generation for Area B looks and acts like a demand curve for generation supplied by Area A. The consumers in Area B are evidently willing to pay at least the price represented by the demand curve;

and consumers benefit from the lower-priced power made available through transmission capacity enhancement (once rates are adjusted to reflect lower costs). The measure of this benefit is the change in consumer surplus. Because this graph shows only production costs, there is no producer surplus. Although this discussion is couched in the language of a vertically integrated firm, it also has a bearing on transmission planning in restructured electricity markets. We will return to this topic in Section III when we discuss policy decisions to increase social wealth through the increase in the sum of consumer and producer surplus.

C. Transmission congestion cost calculations in restructured electricity markets

In a restructured electricity market, generators are owned by many different firms, the transmission system is operated (if not owned) by a separate business entity, and distribution is provided to many, distinct franchises of customers. In this market, the "cost" of maintaining safe transmission operating margins can be defined in a variety of ways. Each definition reflects the design objectives and cost-recovery policies of the particular market. A critical element is specifying how the costs of safe operating margins are recovered from or paid to customers receiving electricity service and/or are paid to or recovered from the generators.

Although these costs are defined differently in different markets, they are usually referred to using the same term: "transmission congestion costs."

The discussion that follows is based on the understanding that a principal objective of electricity industry restructuring is to rely on competition among power suppliers to lower costs to consumers. This basic principle implies that higher energy prices are an incentive for the introduction of new, lower-cost sources of supply. Transmission can be used to enlarge the sources of supply available to meet demand for electricity. Transmission congestion in this context is the consequence of a physical limit on the transmission system's capability to accommodate all requests to using lower-cost, remotely located generation rather than more expensive local generation to meet demand.

All restructured U.S. electricity markets rely on offer-based, centralized, wholesale trade in which generators offer to sell blocks of electricity at fixed prices. Market makers (currently ISOs) clear the market by arranging the offered quantities from least to most expensive and accepting all offers necessary to meet demand. If the market is competitive, the supply curve created by aggregating generator offers should closely approximate the system marginal production cost of generation.

The market clearing price is set based on the last accepted offer and is "uniform";

that is, each accepted offerer is paid the same price regardless of the original offer made. The market clearing price for a given region is the LMP of electricity for that region.³ In other words, the market clearing price is, to a first approximation, the cost of producing one more (or one less) MWh of electricity in that region. This fact can be observed by reviewing the supply curve of generators' offers where we can see that the market clearing price is the marginal cost of supplying (or more accurately, marginal willingness to supply, if offers differ from costs) one additional (or one less) MWh of electricity beyond the amount used to set the market clearing price.

Figure 3 illustrates the application of these principles to a restructured market without congestion. This example uses the same physical circumstances presented in Figure 1, which illustrated least-cost dispatch for an uncongested, vertically integrated market with two areas, A and B. The generators in Area A serve their native load of 500 MW

and transport 200 MW of power to Area B. The LMP for all generators is \$25/MWh. Under this unconstrained operating condition, the total dispatch costs paid to the generators are equal to:

$$(700 \text{ MW})(\$25/\text{MWh}) + (300 \text{ MW})(\$25/\text{MWh}) = \$25,000/\text{h}.$$

These funds come directly from the loads:

$$(500 \text{ MW})(\$25/\text{MWh}) + (500 \text{ MW})(\$25/\text{MWh}) = \$25,000/\text{h}$$

where consumers in both areas pay an energy price equal to \$25/MWh.

Figure 4(a) shows a congested system in which the line connecting the Areas A and B is at its capacity limit of 100 MW. Because of this transmission constraint, Area A generators produce a total of only 600 MW: 500 MW serve the load in Area A, and 100 MW exported to Area B. With the decrease in generation in Area A compared to the uncongested case, the marginal cost of generation offers for Area A (i.e., the

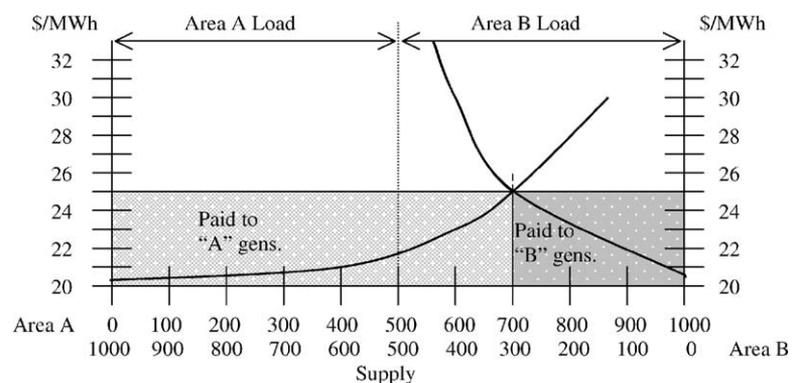


Figure 3: The shaded regions show the dispatch costs paid to the generators in each area from a uniform clearing price market without congestion. Assuming the loads are charged the same price, the same shaded area indicates the revenue collected from the loads

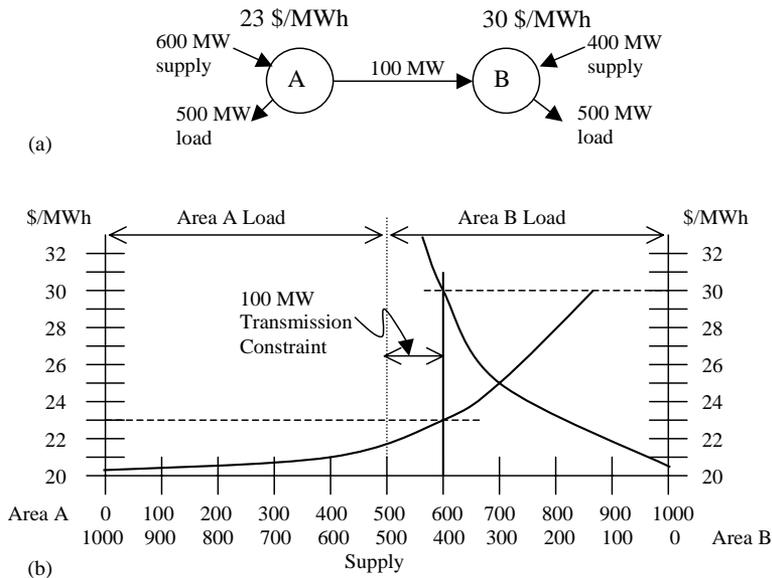


Figure 4: A constrained two-area system (a) and corresponding aggregate supply curves (b). The 100-MW transmission capacity constraint limits Area A generation to 600 MW. This necessitates a corresponding increase in generation in Area B over the unconstrained condition

LMP for Area A) has decreased slightly to \$23/MWh. That is, the marginal cost to physically supply the next unit of energy from Area A generators is equal to \$23/MWh. In Area B, where the generation has increased over the uncongested case to 400 MW, the marginal cost of generation offers (i.e., the LMP for Area B) has increased to \$30/MWh. This is observed on the supply curves shown in Figure 4(b).

The examples of transmission congestion described above enable us to examine methods for collecting payments from loads and making payments to generators. How much the generators are paid and what revenues are collected from the loads depend on the market design and policies of the transmission system operator. In addition to the change in production cost measure of congestion, we have

identified three generic approaches that are used to address congestion costs in restructured U.S. electricity markets:

1. Uplift charges
2. System redispatch payments
3. Congestion revenues.

These three methods are explained in the following subsections in relation to the example two-area system shown in Figure 4 and described above.

1. Uplift charges. Uplift charges suppress the differences in LMPs between the two areas in our example system and instead rely on a single uniform price for energy throughout the system. This method is closest in spirit to the situation in a vertically integrated industry because it equates congestion costs with the change in dispatch payments associated with taking additional

generation on the congested side of an interface “out of merit order,”⁴ which is sometimes called the cost of “redispatch.”

In the uplift charge approach, the system operator declares a uniform market clearing price, which we assume is equal to the price in the unconstrained system in our example. Any generator dispatched “out of merit order” because of congestion is paid its actual offer price, but these payments do not influence the market clearing price. In our example, the dispatch costs paid to the generators are represented by the total shaded areas in Figure 5 and are approximately equal to:

$$\begin{aligned}
 &(600 \text{ MW})(\$25/\text{MWh}) \\
 &+ (300 \text{ MW})(\$25/\text{MWh}) \\
 &+ (100 \text{ MW})(\$27.50/\text{MWh}) \\
 &= \$25,250/\text{h}.
 \end{aligned}$$

The uplift costs resulting from congestion in this case are equal to \$250/h. The source of these funds is ultimately the loads (consumers), which share the costs equally in this market design. The combined energy and uplift costs (price) to the loads is:

$$\begin{aligned}
 &(\$25,250/\text{h})/(1,000 \text{ MW}) \\
 &= \$25.25/\text{MWh},
 \end{aligned}$$

or just slightly higher than the uncongested price of \$25/MWh. With the declared market clearing price equal to the uncongested price, the apparent “congestion cost” under the uplift approach is the difference in generation offers over and above the market clearing prices of the unconstrained system (Figure 2(c)) and the constrained system (Figure 2(b)).

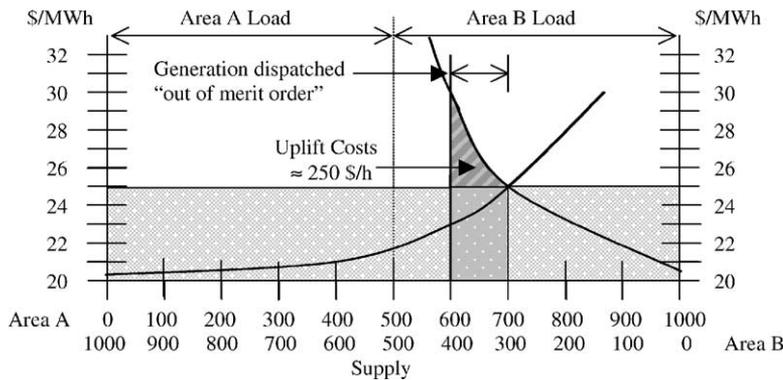


Figure 5: The shaded area represents the dispatch costs paid to the generators in both areas. Note that 100 MW of power come from generators in Area B dispatched “out of merit order” (their higher-cost supply would not be used in the unconstrained case)

One of the drawbacks of the uplift approach is that it does not, by itself, provide a strong economic incentive to site new generation where it is most needed. This lack of incentive may inhibit the long-term benefits of competition among suppliers because new generation will not necessarily be developed in areas where it could drive prices down.

2. System redispatch payments. The uplift charge approach suppresses differences in LMP between two areas separated by a transmission constraint. Acknowledging the differences in LMP between areas is the basis of the system redispatch payments approach to determining congestion costs. This approach is also related to measurement of transmission congestion costs in a vertically integrated industry in that it focuses on the question: How does the total cost of system dispatch change because of the congested interface between two regions? However, in restructured markets, the uniform

market clearing price, set at the distinct LMP for each region, is substituted for the supply curve of generation costs that would be used in the vertically integrated structure. The cost of system redispatch is determined by comparing total dispatch payments to suppliers in the two regions to the total dispatch payments assuming no constraint.

The change in revenue paid to generators for dispatch away from the uncongested case is easily calculated:

$$\begin{aligned}
 &(600 \text{ MW})(\$23/\text{MWh}) \\
 &+ (400 \text{ MW})(\$30/\text{MWh}) \\
 &- (700 \text{ MW})(\$25/\text{MWh}) \\
 &- (300 \text{ MW})(\$25/\text{MWh}) \\
 &= \$800/\text{h}.
 \end{aligned}$$

This quantity is represented graphically in **Figure 6**.

3. Congestion revenues. The congestion revenues method for calculating congestion costs is based on the policy common among ISOs of paying generators a region-specific LMP for their output regardless of the region where the generation is used to serve load, and, at the same time, charging all loads within a given region a single, region-specific LMP. The difference between what is collected from loads and

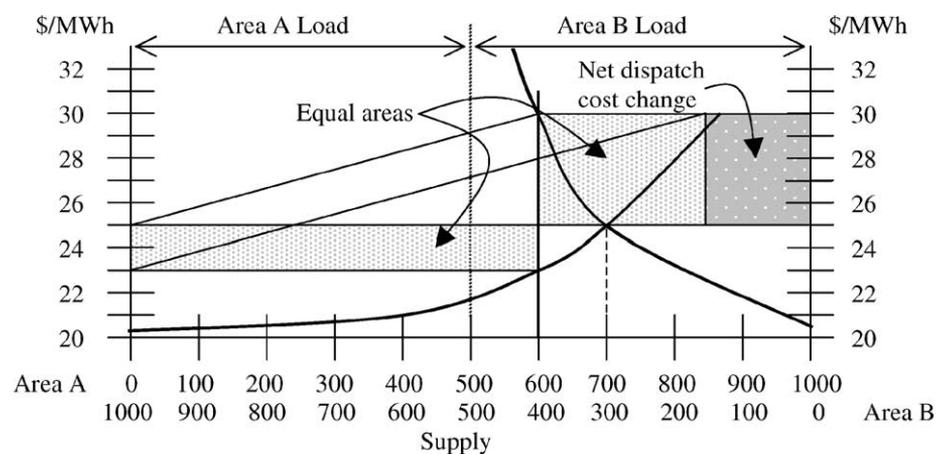


Figure 6: The difference between the shaded areas represents the change in dispatch payments to the generators relative to payments in the uncongested case. In Area B, the price increased relative to the uncongested price; in Area A, the price decreased. The change in dispatch payments to generators because of congestion may be positive, as in our example, or negative. The change in dispatch payments to generators is the darker shaded area on the right edge of the plot. This area is obtained by mapping the area corresponding to the reduced dispatch costs in Area A into the area corresponding to the increased generation costs in Area B. The remaining area on the right represents the net generation dispatch payment increase for this example

what is paid out to generators is used to compensate the holders of the rights to transfer power from one region to the other.⁵

In the constrained example, the LMP in Area A is \$23/MWh and the LMP in Area B is \$30/MWh. The dispatch payments to generators using LMPs and a uniform price auction are equal to:

$$(600 \text{ MW})(\$23/\text{MWh}) + (400 \text{ MW})(\$30/\text{MWh}) = \$25,800/\text{h}.$$

Using LMPs, the money paid by the loads exceeds that paid to the generators. The revenues collected from the loads are equal to:

$$(500 \text{ MW})(\$23/\text{MWh}) + (500 \text{ MW})(\$30/\text{MWh}) = \$26,500/\text{h}.$$

The difference between what the loads pay and what the generators receive results entirely from congestion. Thus, it is natural for the ISO refer to this difference as "congestion costs." This differ-

Table 1: Dispatch, Dispatch Costs, and Revenues for the Unconstrained and Constrained Systems

	Unconstrained System			Constrained System		
	MW dispatched	Dispatch costs paid to generators	Revenues received from loads	MW dispatched	Dispatch cost paid to generators	Revenues received from loads
Area A	700	\$17,500/h	\$12,500/h	600	\$13,800/h	\$11,500/h
Area B	300	\$7,500/h	\$12,500/h	400	\$12,000/h	\$15,000/h
Total	1,000	\$25,000/h	\$25,000/h	1,000	\$25,800/h	\$26,500/h

ence also equals the sum of congestion charges along each of the constrained transmission lines. The congestion charge is the product of the power flow and the LMP differences along the line and is represented graphically by the shaded area in Figure 7. A quick check confirms that the difference between load revenues and dispatch payments:

$$\begin{aligned} \$26,500/\text{h} - \$25,800/\text{h} &= \$700/\text{h} \\ \text{is equal to congestion charges:} \\ (100 \text{ MW})(\$30 - 23/\text{MWh}) &= \$700/\text{h}. \end{aligned}$$

The congestion charges of \$700/h, paid by the loads, are distributed according to the market settlement policy. Each ISO has a policy for the disbursement of these funds. Typically, they are allocated to owners of the congestion revenue rights (CRRs) for the congested transmission path between areas.

Table 1 summarizes the system conditions for the uncongested and congested cases described above. The data include MW of generation dispatched, dispatch payments to generators, and revenues received from loads. In the constrained case, we can observe increased dispatch costs as well as revenues from loads that exceed dispatch costs.

4. Combining system redispatch payments and congestion revenues. The congestion costs paid by loads will necessarily include changes in generator dispatch payments because loads are charged for energy they use. These congestion costs will also include the net congestion revenues resulting from the difference between the prices paid to generators and the

$$(100 \text{ MW})(\$30 - 23 /\text{MWh}) = \$700 /\text{h}.$$

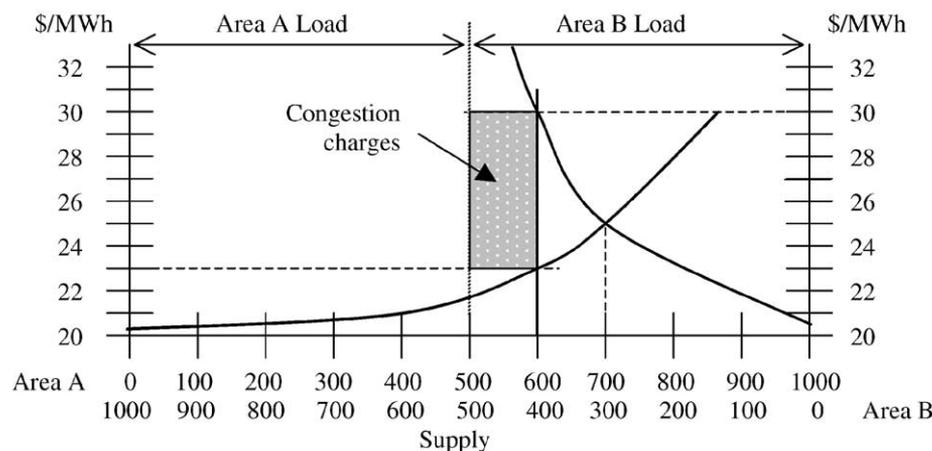


Figure 7: The shaded area represents the congestion charges. In this case, 100 MW of power is transferred from Area A where the generators are paid \$23/MWh to Area B where the loads pay \$30/MWh for the same energy. The congestion charges are the excess payments received from the loads over the dispatch payments to the generators

Table 2: Comparison of Congestion Costs Associated with System Redispatch Payments, Congestion Revenues, and Total Costs to Loads

	"Congestion Costs" for Previous Example
System redispatch payments	\$800/h
Congestion revenues	\$700/h
Total costs to loads (without CRRs) compared to an uncongested case (Method 1 + Method 2)	\$1,500/h

congestion costs. The table indicates which of the three generic cost-calculation methods described was used to determine the congestion costs in each report. The table shows that no single method or set of methods was used consistently in all of the studies. Instead, the reported

prices charged to loads. The net congestion revenues that must be recovered, however, depend both on the total congestion revenues and the amount and allocation of congestion revenue rights.⁶ Thus, there are two elements of congestion costs:

1. System redispatch payments resulting from congestion (see Figure 6), and

2. Total congestion revenues (see Figure 7).

A third element may be inferred if information is available about CRRs:

3. Total costs to loads resulting from congestion, which equals the sum of system redispatch payments and congestion revenues, less any funds returned to the loads through an allocation of CRRs or other means.

These elements correspond to methods of calculating congestion costs, which in turn correspond to different congestion costs that could be reported for our two-area example above, as shown in Table 2.

D. Summary of congestion cost review

Table 3 summarizes the findings from our review of published

Table 3: Summary of Congestion Costs Reported by ISOs, DOE, and FERC

	Period	Congestion Costs	Congestion Cost-Calculation Method(s)
PJM [1]	1999	\$53 M	Congestion revenues
PJM [1]	2000	\$132 M	
PJM [1]	2001	\$271 M	
PJM [2]	2002	\$430 M	
ISO-NE [3]	5/99–4/00	\$99 M	Uplift charges ⁷
ISO-NE [3]	5/00–4/01	\$120 M	
ISO-NE [4]	2003	\$50–300 M	System redispatch payments
CAISO [5]	2000	\$391 M	Congestion revenues
CAISO [5]	2001	\$107 M	
CAISO [6]	2002	\$42 M	
CAISO [7, 8]	2005	\$–7.47 to 306 M	System redispatch payments + congestion revenues
NYISO [9]	2000	\$1,240 M	System redispatch payments (est) + congestion revenues
NYISO [9]	2001	\$570 M	
NYISO [10]	2000	\$517 M	Congestion revenues
NYISO [10]	2001	\$310 M	
NYISO [11]	2002	\$525 M	
FERC [12]	6/00–8/00	\$891 M	System redispatch payments (partial) + congestion revenues
DOE [13]		\$157–457 M	System redispatch payments + congestion revenues

[1] PJM Interconnection, *State of the Market Report 2001* (PJM, 2002).

[2] PJM Interconnection, *State of the Market Report 2002* (PJM, 2003).

[3] ISO New England (ISO-NE), *Annual Markets Report* (ISO New England, 2002a).

[4] ISO New England, *RTEP02* (ISO New England, 2002b).

[5] California Independent System Operator (CAISO), *Market Analysis Reports* (CAISO, 2000, 2001a).

[6] CAISO, *2002 Annual Report on Market Issues and Performance* (CAISO, 2003).

[7] CAISO, *Path 15 Expansion Economic Benefit Study: Phase II* (CAISO, 2001b).

[8] CAISO, *Potential Economic Benefits to California from Expanding Path 15* (CAISO, 2001c).

[9] *New York Congestion and Physical Constraint Cost Estimates* (POWERGEM, 2002).

[10] *2001 Annual Report on the New York Electricity Markets* (Patton and Wander, 2002).

[11] *2002 State of the Market Report: New York Electricity Markets* (Patton, 2003).

[12] Federal Energy Regulatory Commission (FERC), *Electric Transmission Constraint Study* (FERC, 2001).

[13] U.S. Department of Energy (DOE), *National Transmission Grid Study* (DOE, 2002).

costs are derived using very different methods. Because of these differences, direct comparisons of these reported costs should be made with caution. In the next section, we discuss some key issues associated with making meaningful comparisons of reported congestion costs.

III. Key Issues for Accurately Determining and Interpreting Congestion Costs

This section reviews six key observations that emerge from our assessment of published transmission costs. The first two identify additional information that is needed to determine congestion costs on a more comparable basis. The last four are important issues to consider when including congestion costs as part of a process of evaluating strategies to reduce transmission congestion.

1. *Information on operation of congestion revenue rights markets is needed to assess the impacts of congestion revenue costs on consumers.*

Some congestion revenue charges incurred by consumers may be offset by CRRs that return these costs. Information about the CRR process, which differs among ISOs, is necessary to assess the final impact of congestion revenue charges on consumer costs.

In NYISO, CRRs (called transmission congestion contracts, or TCCs) are auctioned, and

the proceeds are given to transmission owners. In PJM, load-serving entities may request CRRs (called fixed transmission rights, or FTRs) up to their peak load. (Remaining CRRs are available through a monthly auction.) The effects of these CRRs and how they are obtained must be accounted for in determining the final



impact of congestion costs on consumers.

2. *Information on generators' offers is needed to assess system redispatch payments.*

ISOs' reports of congestion costs, which refer only to costs handled by each ISO's settlement process, do not permit direct evaluation of system redispatch payments that might be influenced by relief of transmission congestion. In addition, generator offer curves are needed to estimate system redispatch payments that are realized in payments to generators. This information is generally not available, yet it is needed in order to assess the impacts of options to relieve

congestion. The congestion cost impacts of system redispatch can only be determined by comparing dispatch in a congested case to dispatch in an uncongested case. These prices can only be approximated (with an upward bias), based on information on market clearing prices.

In theory, an indirect calculation of system redispatch costs can be calculated with knowledge of total congestion revenues and the price of energy for a hypothetical uncongested system. The LMP is sometimes represented as a sum of energy, congestion, and loss components. If the energy component is chosen to be the uncongested energy price, then the total premium paid by loads due to congestion and losses is easily calculated (the product of demands and the difference between LMP and energy price). This less the congestion revenues yields the system redispatch payments. However, the energy component of the LMP is not typically taken to be the uncongested price. In the original theoretical development of spot prices for electricity and in the practical implementation at the NYISO, the energy component of the LMP is chosen to be the LMP at a specified reference bus.⁸ In PJM no energy component is reported. Since losses are neglected in PJM, congestion revenues can be calculated directly from differences in LMPs. Plans for the new market design in California state that the energy component of the LMP will equal the uncongested price. If it is

implemented in this manner, then a direct calculation of the sum of congestion revenues and system redispatch payments will be possible. Subtracting congestion revenues, one will be able to determine the redispatch payments.

3. *Many studies presume that generator offers reflect competitive market conditions.*

While information on generator offer curves is needed to estimate system redispatch payments, using offer curves may understate the full value of relieving congestion if the market is not competitive (i.e., generators can unfairly raise their offers above their marginal cost of production). Because relieving congestion will reduce the ability of generators to exercise market power, it is appropriate to consider this effect when assessing the benefits of options to relieve congestion.

The examples presented in Section II assumed that the generator supply curve reflected competitive behavior by generators to make offers at their true costs of production. However, if generators are able to exercise market power, congestion cost methods based on generator offers will tend to overstate the costs of congestion relative to what would be found in a more competitive market. In the next point in this section we will observe that the differences between the two CAISO studies of the costs and benefits of relieving congestion along Path 15 hinge on different assessments of the extent

of market power and the effectiveness of congestion relief (alone) in reducing the ability of generators to exert market power.

In the absence of mitigating measures (e.g., taxation), overstated congestion costs will be borne by customers. Overstated congestion costs may also distort assessments of the value of



options to reduce congestion. Transmission expansion may increase competition, causing generators to change their offer curves to represent true costs.⁹

4. *Customer costs may rise as a result of reducing congestion.*

Generally speaking, reducing congestion will likely increase electricity prices in exporting regions, which in turn is expected to be offset by lower prices in the importing regions. Yet, whether total average consumer costs increase or decrease with transmission capacity expansion, which reduces congestion, depends on price changes and levels of demand in each region. In planning studies of decreased

congestion conducted for California, an overall increase in consumer costs is observed in many scenarios. The examples we described earlier support the common assumption that reducing congestion leads to lower consumer costs. However, as noted in the second observation in this section, knowledge of generators' offers is needed to determine the full effect of reductions in congestion. In fact, costs to consumers may actually increase depending on the aggregate supply curve of generators' offers for each area. See the appendix to this article for an example illustrating this possibility.

5. *Minimizing consumer costs may not increase aggregate social wealth.*

Economic theory suggests that policy decisions should consider the effect of transmission congestion relief on the sum of consumer surplus and producer surplus.¹⁰ Yet, as noted in the previous finding, there are instances when relieving congestion would increase rather than lower consumer costs. Thus, although economic theory suggests that maximizing aggregate wealth is an appropriate policy objective for relieving congestion, the potential need to trade off this objective with that of lowering consumer costs must also be considered.

In this context, we refer to the example introduced in the appendix, in which transmission expansion causes an increase in consumer costs. A careful examination of the net change

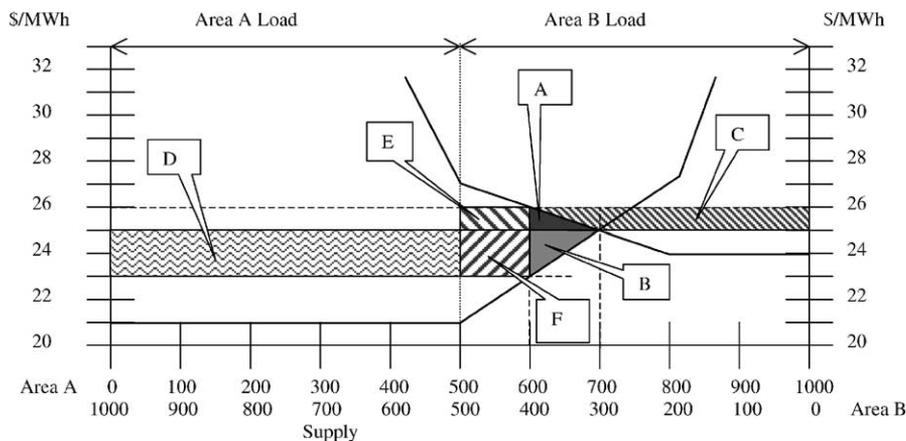


Figure 8: Transfers and changes in consumer and producer surplus from an increase in transmission capacity to eliminate congestion: (A) increase in consumer surplus, (B) increase in producer surplus, (C) transfer of producer to consumer surplus, (D) transfer of consumer to producer surplus, (E) consumer surplus from congestion revenues, (F) producer surplus from a portion of congestion revenues

in consumer and producer surpluses reveals that relieving congestion will increase the aggregate wealth.¹¹ A detail diagram displaying changes in consumer and producer surpluses and the wealth transfers involved is shown in **Figure 8**.

In **Figure 8** the areas labeled “A” through “F” represent the following changes:

A. An absolute increase in consumer surplus due to increased transmission capacity;

B. An absolute increase in producer surplus due to increased transmission capacity;

C. A transfer from producer surplus to consumer surplus;

D. A transfer from consumer surplus to producer surplus;

E. Consumer surplus from a portion of congestion revenues;

F. Producer surplus from a portion of congestion revenues.

Assuming the congestion revenues are disbursed to market participants in a

manner described by ISO policy, the change (increase) in the sum of consumer and producer surplus is represented by the sum of Areas A and B. All other areas represent transfers of wealth between participants.

The appropriate distribution of wealth, however, is a not a topic for which economic analysis of the type used in the discussions so far is well-suited to address; it is instead a matter of equity and political economy. For example, in the current case, elimination of congestion involves a net transfer of wealth from consumers to producers. When there are numerous consumers and few producers, this distribution might not be desirable; this is one reason why other issues, such as consumer costs, are considered. Alternatives such as maximizing aggregate wealth followed by a system of transfers to achieve some designed equity are attractive in theory, but

face many practical and political barriers.

6. *There is no standardized conceptual framework for studies of congestion costs.*

Our survey of ISO and government reports on congestion costs, summarized in **Table 3**, reflects a diversity of practices which makes it difficult to compare reported congestion costs across institutions. This is hardly surprising, as there are several perspectives from which the economic and financial impacts of congestion can be viewed. It would be useful for future reports to state clearly the different purposes to which the various measures apply and then apply the measurement concepts in more standardized fashion.

Measures of changes in payments to generators due to congestion, such as uplift charges and system redispatch payments, partially reflect transfers between producers and consumers. To the extent that the planner cares about the allocation of social surplus between producers and consumers, a plan that maximizes the sum of producer and consumer surplus and a system of transfers that achieves the desired allocation between the groups, in principle, can be implemented. For this reason, many economists (Joskow and Tirole, 2003; Barmack et al., 2003) focus on “redispatch costs.” This measure reflects the social costs of congestion, i.e., the change in

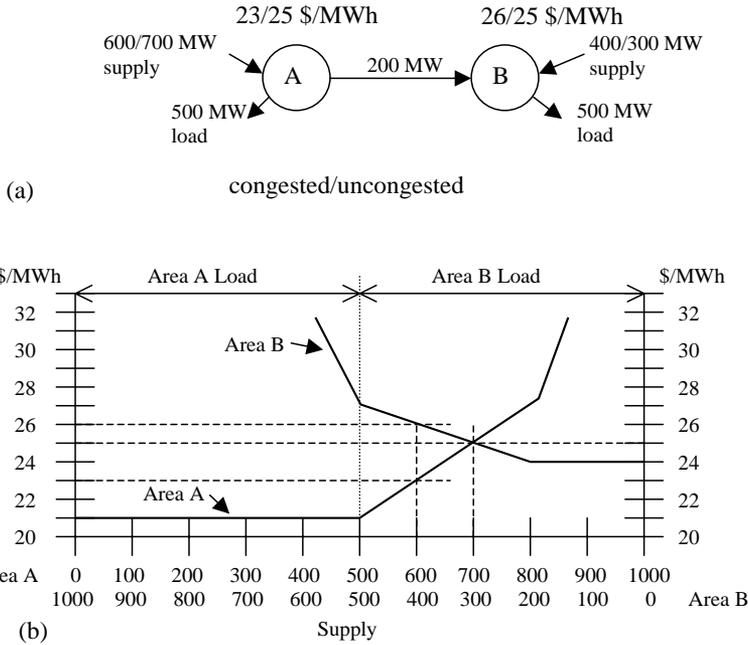


Figure 9: A two-area system in which transmission expansion to eliminate congestion will increase total costs

curve in Area A has a more pronounced positive slope. With a 100 MW transmission line capacity limit, the total generation dispatch payments are equal to:

$$(600 \text{ MW})(\$23/\text{MWh}) + (400 \text{ MW})(\$26/\text{MWh}) = \$24,200/\text{h}.$$

The revenues collected from the loads using LMPs are:

$$(500 \text{ MW})(\$23/\text{MWh}) + (500 \text{ MW})(\$26/\text{MWh}) = \$24,500/\text{h}.$$

When the transmission capacity is increased to allow unconstrained

production costs from out-of-merit-order dispatch due to congestion, and nets out transfers between producers and consumers. This measure is particularly valuable in transmission expansion studies. Very few existing studies focus on this measure, in part because it is difficult to calculate using readily available data.¹² ■

Appendix A

The counterintuitive result that reducing congestion can increase consumer rates can be illustrated with a simple example based on those presented in the body of this article by assuming a different shape for the supply curves of generator offers, as shown in Figure 9. The supply curve in Area B is mildly sloped in the operating region, and the supply

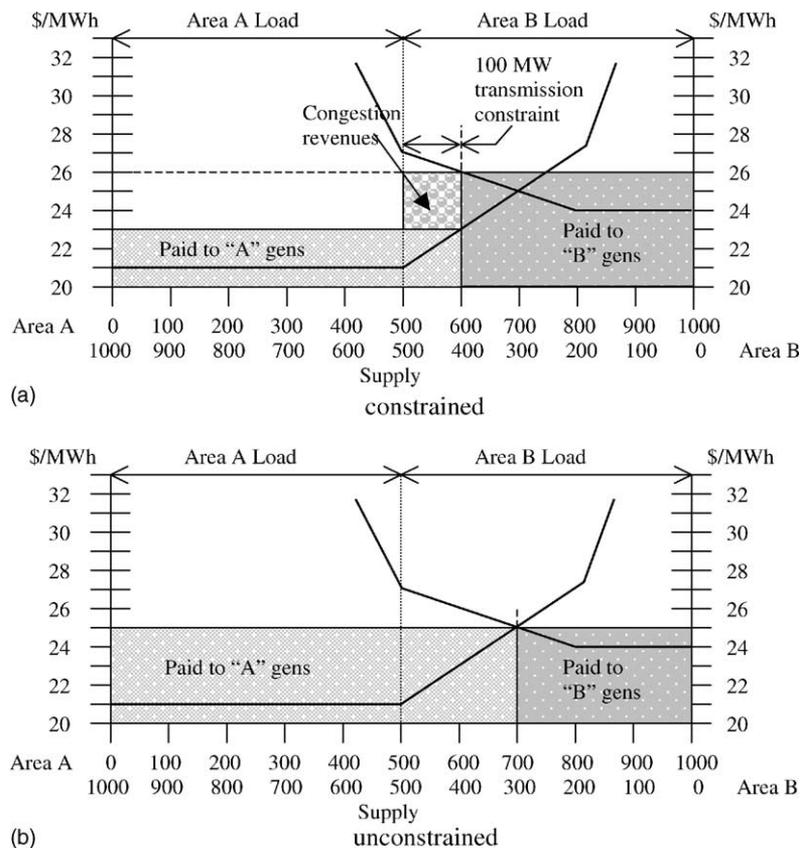


Figure 10: (a) The dispatch payments to generators and the additional congestion charges incurred by loads represent the total costs to consumers. (b) Congestion charges are zero with the expansion of transmission capacity; however, dispatch payments to suppliers have increased so that the overall cost to consumers has increased

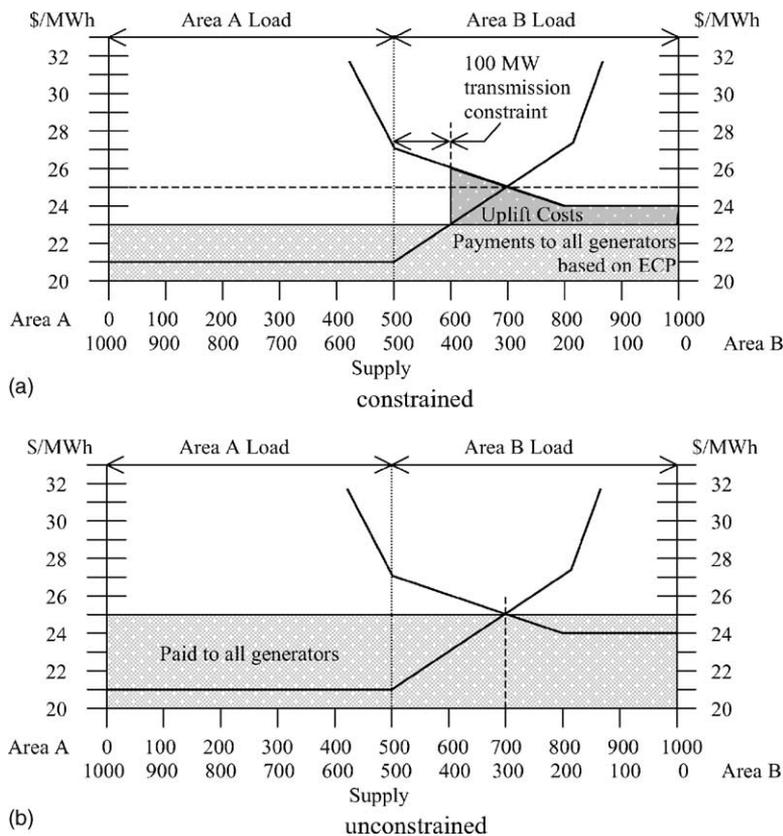


Figure 11: (a) In this congested case, all generators in Area B are dispatched “out of merit order” because lower-cost generation is available. The market clearing price is set to \$23/MWh, the cost of the last unit of power dispatched in merit order. (b) In the uncongested case, all generators are paid the market clearing price, and there are no uplift costs. In this example, the transmission constraint and choice of market clearing price effectively reduce payments to generators and lower total costs to consumers

operation, the total dispatch payments to the generators are:

$$(700 \text{ MW})(\$25/\text{MWh}) + (300 \text{ MW})(\$25/\text{MWh}) = \$25,000/\text{h},$$

which are more than the \$24,200/h paid when transmission was congested. Because there are no congestion charges in the unconstrained case, the revenues collected from the loads also equal \$25,000/h, an increase from the congested value of \$24,500/h. All of these costs are depicted in Figure 10.

Enhancing transmission capacity,

in this example, increases costs to consumers in Area A and decreases costs to consumers in Area B. Similarly, the overall profits to producers in Area A increase while those in Area B decrease.

Without knowledge of the relevant supply curves for different areas separated by congested interfaces, it is impossible to determine whether transmission expansion will actually reduce consumer costs.

A similar situation arises for markets that use uplift costs to pay for generation

dispatch out of merit order because of congestion. Depending on how the market clearing price is determined, transmission expansion could increase total costs to consumers. In the specific example we considered in Figure 5, we assumed a market clearing price set equal to the unconstrained system price. More typically, however, the market clearing price is set at the marginal cost of the last generator chosen for supply in merit order. When this happens, the constrained system can have lower total dispatch payments to generators and lower total costs to consumers. This is illustrated in Figure 11.

References

- M. Barmack, P. Griffes, E. Kahn, and S. Oren (2003), *Performance Incentives for Transmission*, ELEC. J. April 2003, at 9–22.
- California Independent System Operator (2000), *Market Analysis Reports*, Jan. 2000–Dec. 2000.
- California Independent System Operator (2001a), *Market Analysis Reports*, Jan. 2001–Dec. 2001.
- California Independent System Operator (2001b), *Path 15 Expansion Economic Benefit Study: Phase II—Year 2005 Prospect*, Sept. 24, 2001.
- California Independent System Operator (2001c), *Potential Economic Benefits to California from Expanding Path 15—Year 2005 Prospect*, Sept. 24, 2001.
- California Independent System Operator (2003), *2002 Annual Report on Market Issues and Performance*.
- U.S. Department of Energy (2002), *National Transmission Grid Study*.

- Federal Energy Regulatory Commission (2001), *Electric Transmission Constraint Study*.
- S. Hunt (2002), *MAKING COMPETITION WORK IN ELECTRICITY* (New York: John Wiley).
- ISO New England, Inc. (2002a), *Annual Markets Report, May 2000–April 2001, Section 2, Technical Review*, Sept. 12, 2002.
- ISO New England, Inc. (2002b), *RTEP02*, Nov. 7, 2002.
- P. Joskow and J. Tirole (2003), *Merchant Transmission Investment*, available at <http://econ-www.mit.edu/faculty/pjoskow/papers/Merchant.pdf>.
- D.B. Patton and M. Wander (2002), *2001 Annual Report on the New York Electricity Markets*, June 2002.
- D.B. Patton (2003), *2002 State of the Market Report: New York Electricity Markets*, April 2003.
- PJM Interconnection, LLC (2002), *PJM Interconnection State of the Market Report 2001*, Report of the PJM Market Monitoring Unit, June 2002.
- PJM Interconnection, LLC (2003), *2002 State of the Market Report*, Report of the PJM Market Monitoring Unit, March 2003.
- POWERGEM (2002), *New York Congestion and Physical Constraint Cost Estimates*, Project report prepared for the New York ISO, Aug. 2002.
- A. Siddiqui, E. Bartholomew, C. Marnay, and S. Oren (2003), *The Effectiveness of Transmission Congestion Contracts in Hedging Transmission Congestion Risk in New York State*, Eighth Annual Power Conference on Electric Industry Restructuring, Berkeley, CA, March 14, 2003.

Endnotes:

1. The nature of electricity and the interconnectedness of the networks over which it flows means that it is extremely important to maintain safe operating margins through the coordinated efforts of all interconnected transmission

system operators. Failure to operate the entire network a coordinated manner increases the likelihood that loss of one or more elements (e.g., a generator or transmission line) could lead to a catastrophic, cascading blackout of the entire network.

2. It is useful to observe that, viewed from the perspectives of generators in Area A, the supply curve of generation in Area B looks like and in fact acts as a “demand curve” on



generation supplied from Area A. This presentation of a reverse supply curve to investigate the value of transmission appears in several places in the literature including Hunt (2002), Joskow and Tirole (2003), and Barmack et al. (2003).

3. This discussion suppresses important technical details regarding differences between zonal and nodal pricing. Locational marginal prices, for the purposes of this report, refer only to the broad principle of prices that vary according to location within the grid.

4. The term “out of merit order” has its origin in the unit commitment problem, in which decisions are made about which units to operate. Those chosen for reasons other than least cost (e.g., reliability, congestion) are said to be dispatched out of merit order.

5. These rights are called congestion revenue rights (CRRs) in the FERC Standard Market Design (SMD) Notice

of Proposed Rulemaking (NOPR). They are called financial transmission rights (FTRs) in the more recent FERC white paper on Wholesale Power Market Platform.

6. There is ongoing debate in the industry regarding the adequacy of various methods for redistributing congestion charges collected by ISOs. In an omniscient allocation of congestion revenue rights to loads, all revenues might be returned. Empirical analysis of CRRs has shown that actual performance is mixed (Siddiqui et al., 2003).

7. ISO New England’s congestion cost-calculation method was modified in March 2003.

8. There are technical reasons for the use of reference bus to aid in the solution of the optimization problem that determines the dispatch and LMPs.

9. Expansion solely for this purpose would have to be weighed against other policy measures that might be adopted to mitigate market power.

10. The producer surplus is the difference between revenues received for production and actual production costs. The consumer surplus is the difference between the willingness to pay for a product and the actual payment.

11. Assuming a strictly inelastic load, the willingness to pay and consumer surplus are ill-defined. The change in consumer surplus, however, is calculable. In our and similar examples in the literature, the willingness to pay is at least that defined by the supply curve of the importing region, and is treated as such in Joskow and Tirole (2003) and Barmack et al. (2003). Assuming any willingness to pay above this, even infinite, does not affect the calculation of the change in surplus.

12. To the extent that this measure is used at all, it is usually in a planning context. For example, see <http://www.aiso.com/docs/2003/08/25/2003082516110324793.pdf>.