(16)

MODERNIZING MEXICO'S ELECTRICITY SYSTEM:

SOME REFLECTIONS

Arnold C. Harberger

University of California, Los Angeles

August 1998

Introduction

This paper arose out of an all-day workshop held in Mexico City in June of this year. Following that exercise, I was asked to write a one-page summary of the main points I had set forth during the meeting. On this I expressed a certain reluctance, perhaps foreseeing what actually happened as I put pen to paper. In any case, the net result of my trying to set out the main thoughts I expressed is the paper before you. In it I have tried to be reasonably didactic, so that those not well versed in electricity economics can hopefully follow the arguments. At the same time I have tried to introduce enough real-world complications to help convince practitioners in the field that these thoughts are presented with real-world problems in mind, not just as classroom exercises.

I think the paper as it stands is a reasonably good introduction to the problems it treats, thinking always against the backdrop of the Mexican economy of today, and the electricity system that serves it. To me, the most obvious next steps are to try to calibrate the analytical framework here presented to the actual data of the Mexican electricity network -- data from recent years, projections of future demand, current capital costs of different types of new capacity as well as fuel costs for all the generating facilities now in service or projected for the near future. Such an exercise of calibration would help to sharpen the

analysis further, and make it more directly useful for decisionmaking by the Mexican authorities.

Competition in Electricity Generation

1. I start from the understanding that the long-term objective is a modern electricity system based on sound economics and building in ample participation of private-sector entities. Chile, Argentina and the United Kingdom are three countries which have seriously attempted to apply such a system. It is highly desirable that the Mexican authorities inform themselves concerning both the successes and the obstacles and difficulties encountered by these countries as they designed and implemented their modernization programs.

2. The key innovation in modern electricity economics is the idea that power generation can be made to function competitively. Whereas in earlier times electricity generation was classified along with its transmission and distribution as a "natural monopoly", the formation of extensive interconnected networks has rendered that classification a thing of the past for generation facilities. Somewhat paradoxically, scientific central management lies at the heart of any implementation of this modern idea.

3. The essence of scientific management is the activity of the "Load Dispatch Center". This center, which ideally should be a highly autonomous public body, takes control of the order in which different generating plants get plugged into (or out of) the system, so as to meet a demand which fluctuates greatly according to the hour of the day. the day of the week, the season of the year, etc., as well as (quite importantly in many cases) the weather at any given moment. The principle that should govern the "stacking" of the different generating units is the least cost principle. In simple applications this means using first the units with the lowest operating cost, and proceeding in sequence from low to higher

and higher operating cost per kwh, leaving the highest operating-cost capacity as the last to be turned on as demand increases, and the first to be turned off as demand declines. In more complicated cases, one must take into account the fact that some units can be almost costlessly and instantaneously turned on and off, while for other units turning them on and off entails extra costs, in addition to the operating (mainly fuel) costs while they are running.

4. In the "ideal" system, each generating plant, on joining the system, gives up the right to determine when it will be turned on and off. This right is ceded to the Load Dispatch Center. This does not mean, however, that the private enterprise that owns the generating plant cannot make contracts with its clients for delivery of energy at specified or unspecified times, with or without special conditions. All such contracts would in principle be honored under a modern system, but it need not be the contracting firm that does the generating. If generator G contracts with Client C to provide a certain amount of energy at, say, 5¢ (of dollar) per kwh, it may turn out that C would in fact get electricity generated by H, not G. This could happen if G's running costs were 3¢ per kwh, while H's were only 2¢. Then, so long as the cutoff point at the Load Dispatch Center turned out to be less than 3¢, G would not be generating at those times. But G's contract would be fulfilled using electricity generated by H. G would, in effect, buy this energy from H, paying, say, 2 1/2¢ (if that were the actual cutoff point). Then G would make more profit than it would have made by actually operating (in which case its costs would have been 3¢, not 2 1/2¢), while H would actually be earning a "rent to capacity" of 1/2¢ (= the cutoff point of 2 1/2¢ minus H's actual operating cost of 2¢ per kwh.

5. For such a system to function effectively, the individual private firms that are involved must have complete confidence in the technical capacity and in the total fairness of

the authority (the Load Dispatch Center) that manages the system. And since the system is not yet in place in Mexico, this complete confidence must somehow be established ahead of time -- i.e., before new private firms either make new investments in generating capacity or reach the decision to connect some existing generators (built for self-generation) to the national grid.

I firmly believe that it is a matter of very high priority to: a) establish as early as possible the precise legal and institutional framework within which private-sector electricity generators will operate; b) set in motion immediately, using the generating plants already in the interconnected grid, the precise mechanisms of load dispatch management that will prevail over the longer-term future; and c) make explicit immediately the method by which marginal running costs will be determined for each plant (this involves actually doing so for all existing plants), and how these marginal running costs will be adjusted as fuel prices change (this involves, among other things, establishing the precise formulas by which each plant's "marginal cost" would vary as fuel prices changed, the precise fuel or fuels which would be used for each plant's formula, and the precise method by which periodic quotations on such fuel prices would be obtained).

It is hard to imagine that potential private-sector operators would be very enthusiastic about investing in new capacity unless they are fully knowledgeable about how the system will work, and (equally important) fully convinced that the system will be operated in a fully transparent manner under well-specified and well-understood rules.

Linking Capital and Operating Costs

6. The economic pricing of electricity in a modern system is based on cost minimization. There has been some discussion of whether this cost minimization should be

based on short-run marginal costs (mainly fuel costs) or on some middle- or longer-run concept of marginal costs.

The short answer to this question is that the relevant costs are the short-run marginal costs of each plant. I will present here a numerical example using three types of thermal capacity.

Capacity Type	Capital Cost per KW	Running Cost Per kwh
G	$400	6¢
C	$600	4¢
B	$1000	2¢

I assume that the annualized return required for each type of capacity is 15% (in real terms).

Using the above data, we can obtain the optimal stacking pattern by solving two equations. First, we find the number of Hours H_1 for which the costs would be the same using capacity of type B or capacity of type C. The relevant equation is

$$(.15)(\$1000) + .02\ H_1 = (.15)(\$600) + .04\ H_2$$

$$.02\ H_1 = .15(\$400) = \$60.00$$

$$H_1 = 3000 \text{ hours}$$

This means that if we are building new capacity to be operated for more than 3000 hours, the least-cost way of doing so is by building type B ("baseload") capacity. If we want to introduce new capacity that is to be operated for less than 3000 hours, then we would probably turn to type C ("combined cycle") capacity. Why just probably? See below.

The second equation sets the point of cost equality between capacity of types C and G.

$$(.15)(\$600) + .04\ H_2 = (.15)(\$400) + .06\ H_2$$

$$.02\ H_2 = (.15)(\$200) = \$30.00$$

H_2 = 1500 hours

What can we say about tariffs in this simple system? In the first place, when only baseload capacity (B) is operating the wholesale tariff at a generating node should be 2¢/kwh. When capacity of types B and C are both operating, the relevant tariff (based on a marginal C-generating node) should be 4¢/kwh. When all three types of capacity are operating, the relevant tariff (based on a marginal G-generating node) should be 6¢/kwh.

7. The system as described so far is based only on marginal running cost. The question now arises of how to bring capital costs into account. The standard, and I believe correct answer is by way of a single peaktime surcharge to be applied in periods when all three types of generators are being used at (or near) full capacity.

Suppose that the system in question has a system peak of 1000 hours. Then the annualized capital costs of gas turbine (G) capacity would be spread over this number of hours. That would give us a peaktime surcharge of 6¢ per kwh (= \$60.00 ÷ 1000). Our load duration curve might look something like this:

SYSTEM LOAD DURATION CURVE

Let us now "test" the pricing format given above.

a) First, suppose that a "need" appears for new capacity to be operated for 4500 hours a year. Such capacity would be type B, and would presumably have the following pattern of output, classified by the system marginal cost (SMC).

1500 hours @ SMC = 2¢	$30.00
1500 hours @ SMC = 4¢	60.00
500 hours @ SMC = 6¢	30.00
1000 hours @ SMC = 12¢	120.00
Total Receipts	$240.00

These receipts would cover

Type B marginal costs (2¢ × 4500) =	$90.00
Type B annualized capital costs	
(= .15 × $1000) =	$150.00
Total Costs	$240.00

b) Now suppose the "need" is for capacity to be operated for 2000 hours. Our analysis tells us that this would be type C capacity. The accounting then is as follows.

Pattern of output

500 hours @ SMC = 4¢	$20.00
500 hours @ SMC = 6¢	30.00
1000 hours @ SMC = 12¢	120.00
Total Receipts	$170.00

The corresponding table for costs of the type C plant is

Type C marginal costs (4¢ × 2000) =	$80.00
Type C annualized capital costs	
(= .15 × $600) =	$ 90.00
Total Costs	$170.00

c) Finally, suppose the "need" is for capacity to be operated for just 1200 hours. This would, according to our analysis, be type G capacity. Here we have the following pattern of output

200 hours @ SMC = 6¢	$12.00
1000 hours @ SMC = 12¢	120.00
Total Receipts	$132.00

The breakdown on the cost side is

Type G marginal costs (6¢ × 1200) =	$72.00
Type G annualized capital costs	
(= .15 × $400) =	$ 60.00
Total Costs	$132.00

These examples strike some people as nothing less than a miracle. With just one peaktime surcharge applied only during the hours of system peak, we manage to cover the capital costs not just of the gas turbine (type G) peaking capacity, but of all the other types of capacity as well.

But, miracle or not, it is nonetheless true. One gets into considerable trouble if one thinks of charging, for each type of capacity, "marginal costs" that provide for recovery of the capital costs of that type of capacity.

8. The precise expected payoff of ex ante capital costs, described in the preceding section, is characteristic of mature or fully-equilibrated electricity systems. Suppose we draw two horizontal lines in the load duration curve, one reflecting a duration of 3000 hours, the other of 1500 hours. These two lines divide the area under the curve into three parts. The area under the bottom (3000-hour) line is supposed to be filled by baseload (type B) capacity. The area between the 3000-hour and the 1500-hour lines is supposed to be generated by combined cycle (type C) capacity. And the area above the upper line is supposed to be generated by gas turbine (type G) capacity. To do this obviously requires having the "right" amount of each type of capacity. A system might have too much or too little of any type of capacity. And when this is true the relative abundance or shortfall of different types of capacity provides a signal for investment policy in the coming years.

For example, Mexico emerged from the debt-crisis years of the early and middle 1980s with a tremendous shortfall of combined cycle capacity. This was the product of two causes: a) a half-decade or more in which fiscal constraints virtually cut off investment in new electricity capacity, and b) great advances in combined-cycle technology, which in effect expands the range of load durations for which combined cycle is the most appropriate capacity to employ.

To illustrate how this works, simply modify the data for type C capacity so that the marginal running cost is 3 1/2¢ rather than 4¢ per kwh. Now the two equations for H_1 and H_2 are as follows:

$$90.00 + .035\ H_1 = 150.00 + .02\ H_1$$

$$.015\ H_1 = 60.00$$

$$H_1 = 4000$$

$$60.00 + .06\ H_2 = 90.00 + .035\ H_2$$

$$.025\ H_2 = 30.00$$

$$H_2 = 1200$$

So this simple technological change has made combined cycle capacity "the best" for load durations of 1200 to 4000 hours, as compared with a range of 1500 to 3000 hours before the change.

This in turn means that new investments in B and G capacity are unlikely to be called for for quite some time. Thus it is not surprising that I was informed, sometime around 1990, that the Federal Electricity Commission's investment plan called for practically nothing but additional combined cycle capacity, all the way up to something like the year 2000 (which may well have been its investment horizon at that time).

The basic lesson here is that if the actual installed capacities of different types do not correspond to the ranges generated by comparisons of investment-cum-running costs, then this imbalance itself sends a strong message about the type or types of capacity on which the investments of the near future should be concentrated.

Obviously, when a technological change of the type treated here takes place, this also implies a corresponding change in the step function representing system marginal costs, and in the pattern of electricity tariffs, especially if these are designed (as they should be) to reflect system marginal costs.

9. In this section I illustrate the problem created by "full-cost pricing". Suppose that the plants of type B typically operated (on average) for 5000 hours while those type C typically worked 2500 hours and those of type G 1200 hours. Full-cost pricing would then mean charging

$$2¢ + \frac{\$150.00}{5000} = 5¢/\text{kwh for energy coming from type B capacity}$$

$$4¢ + \frac{\$90.00}{2500} = 7.6¢/\text{kwh for energy coming from type C capacity, and}$$

$$6¢ + \frac{\$60.00}{1200} = 11¢/\text{kwh for energy coming from type G capacity.}$$

If these tariffs were treated as marginal costs, we would end up grossly over-remunerating the capacity of types B and C. Thus, if a rate of 5¢ were charged in the hours when the marginal capacity was of type B, a rate of 7.6¢ when the marginal capacity was of type C, and a rate of 11¢ when the marginal capacity was of type G, then the previous tables for types B and C capacity would have to be modified as follows:

1500 hours @ 5¢ =	$ 75.00
1500 hours @ 7.6¢ =	$114.00
1500 hours @ 11¢ =	$165.00
Total Receipts	$354.00
Compared with total cost of	$240.00

For Type C capacity

500 hours @ 7.6¢ =	$ 57.00
1500 hours @ 11¢ =	$165.00
Total Receipts	$222.00
Compared with total cost of	$170.00

Readers can experiment with different ways of assigning a charge of more than system marginal cost to each hour, and they will find they all lead to the same sort of problem -- for all except the peaking-type capacity, total receipts from such a pricing policy would always

exceed total costs.

This problem could be solved by paying each type of capacity a single price for all the hours of its operation -- i.e., 5¢ to type B, 7.6¢ to type C, 11¢ to type G. But this means that different prices are being paid for different kwh produced and consumed at the same time. And related to this, it means inefficiencies on the demand side to the extent that the users of time-priced energy are required to pay more than system marginal cost.

Qualifications for Cases With Substantial Hydro Capacity

10. Special problems arise in countries with substantial volumes of hydroelectric power: The main such problem arises out of the unpredictability of nature. Thus, gas turbine capacity, which in a purely or nearly purely thermal system would be dedicated to meeting daily and seasonal peaks, must take on the added role of an "insurance policy" against drought. In such cases, the system peak (or demand peak), which in our previous example was 1000 hours per year, is no longer the relevant criterion for determining when a peakload surcharge should be applied. The short answer in this case is that most of the daily and seasonal fluctuations in demand will occur at times when a substantial fraction of gas turbine capacity is idle. Why is it idle? It is being held as a reserve against the contingency of a very dry year, or, even more relevant in some cases, against the contingency of two or three dry years in a row. When this is the case, the daily and seasonal peaks would be using, in our example, 6¢ energy rather than 12¢ energy.

How does the "theory" of marginal cost pricing handle such a situation? Very simply. It asks the same question as before. When is the thermal capacity of the system being fully (or nearly fully) utilized? These are the hours which, when demand grows, will give rise to a need to build more thermal capacity. And it is therefore to these hours that the

capital costs of expanding such capacity should be assigned.

Hence, in place of the 1000-hours-per-year peak in our earlier example we might have a 3000-hour-a-year use of our full gas turbine capacity, say once in 20 years, and maybe a 1500-hour-a-year use on two other occasions in a 20 year span. Total full-capacity use would be 6000 hours, in 20 years, or 300 "hours per year", on average. Our annualized capital cost of $60.00 would now have to be spread over 300 hours, leaving a peakload surcharge of 20¢ per kwh, not 6¢ as in the previous example.

This situation is part of the "fate" of countries with substantial amounts of hydropower. The logical implication would be that a peaktime charge (f.o.b. generating nodes) of 26¢ per kwh would apply in the periods of "thermal peak" -- 3000 hours once in 20 years, and 1500 hours twice during the same span.

11. Nobody that I have talked to on this subject considers that such high rates, concentrated in just a few dry or very dry years, should actually be implemented. Electricity users are the opposite of enthusiastic about the idea. And power company representatives take an equally dim view.

Probably the best approach in this case is to collect a fixed charge from electricity users. If levied as a demand charge, this could equal $60.00 per KW per year in the case that all the maximum demands coincided. It could be lower in the case of noncoincident peak demands. The sum total of the amounts collected, under this demand charge, could then be distributed to the generating plants at the rate of $60.00 per KW of capacity per year. This would put us in the situation illustrated in Section 7, in which receipts and costs would precisely balance, for new investments embodying the best currently-available technology. Moreover, it would permit actual tariffs per kwh to be equal, in principle, to marginal

running costs for all hours in all years. The only deviation from optimal policy would be during the critical hours of full capacity use (3000 + 1500 + 1500 in the example just presented), for which economic analysis would call for a peaktime surcharge to deter demand (and cover capital costs). But we started this section with the "fact" that no major interest group supports such a policy. This rejection condemns us to seek a second-best solution, for which the one just presented (the "demand charge") is an excellent candidate.

An alternative second-best solution would be to charge an "insurance premium" on all kwh sold during a year. The purpose would be to amass precisely the same total sum as would be collected by the "demand charge" method -- i.e., enough to give a fixed payment of (in our example) $60.00 per year to each KW of capacity in the system. The amount of this "insurance premium" can be calculated as follows $60 ÷ (8760 × system load factor). Thus, if the system load factor was .5, the surcharge would be about 1.37¢ per kwh. This would introduce a small "triangle" of efficiency cost for all kwh produced by the system, except for the critical 6000 hours during the specially dry years, when the 1.37¢ surcharge would represent a small step in the right direction.

12. Another complication that emerges when hydro capacity is significant is how such capacity fits into the so-called "stacking pattern". The easy answers here concern: a) run-of-the-stream capacity (de pasada) and b) daily reservoirs. For the first, it is clear that its natural place is at the base of the stacking pattern. For the second, its natural place is at and around the daily peak (i.e., the water accumulated during off-peak hours is released during peak hours).

Complications arise with larger dams, mainly those built for seasonal or inter-annual storage. The first lesson concerning these dams is that when they are full, they function like

run-of-the-stream capacity, and their proper place is at the base of the stack, and that when they are nearly empty, their natural use is for peaking purposes. This gives us a clue to the issue of pricing electricity coming from these dams: the price should vary, depending on how close to full (or how close to empty) the dam is. Unfortunately, it also varies seasonally, mainly depending on prospective stream flow. An empty dam at the beginning of the rainy season is not as much of a problem as an empty dam at its end.

The bottom line on storage hydro is that there should be for each dam a schedule showing how the "marginal cost" of electricity generated by it will vary according to: a) the height of water in the dam, b) the season of the year (expected stream flow) and c) the prices of the key fuels used in the thermal part of the system.

One very key economic proposition that follows from this analytical framework is that, except when a dam is being used for genuine peaking purposes, it should be used at full capacity or not at all. This follows from the fact that the water to be released from a given dam on any given day has an economic price, linked to external conditions plus the water level of that dam at that time. The water level will not change significantly in any one day. Hence, if the economic price of energy from that dam on that day is 4.3¢ per kwh, the dam should not be operating at all during hours in which the system marginal cost is 4.2¢ or less, and should be operating at full capacity before any plant with marginal cost of 4.4¢ or more is turned on.

A final note on hydro power. There is no reason whatsoever for the capital cost of any hydro project to come into play in calculating its effective marginal cost or its proper place in the stacking pattern at any moment. This is a perfect example of a case where "sunk costs are sunk".

The Obsolescence Factor in Thermal Costs

13. This is a good point at which to bring up the question of whether sunk costs are sunk for thermal capacity as well. The answer is, of course, always going to be yes, but there is an important appendix to be added in the case of thermal capacity. That is that any new investment should properly give rise to an expected flow of returns that will repay the costs of that investment, accumulated at a real rate of return reflecting the appropriate opportunity cost of capital. Hence if, in response to growing system demand, investments in new 200 MW combined cycle units are called for, then similar units that are already in the system should be receiving equivalent economic rents to capacity. It is in this sense that, with appropriate marginal cost pricing, one is likely to find in a growing electricity system a fair number of existing units that are yielding, on their historical investment amounts, a rate of return at or near the prevailing opportunity cost of capital.

But the more basic truth is that such situations are the product of a particular dynamic setting -- i.e., a growing demand met by investment in new capacity which is quite similar to certain components of already-installed capacity. The more pervasive lesson is that sunk costs are always sunk, and hence there should be no presumption that "old" investments will yield a target rate of return on their historical costs.

This is a hard message to sell in the public utilities sector, where so much of past regulation has been focused on guaranteeing (or almost guaranteeing) a given rate of return on historical cost. But it is a message that should not only be "sold", but also implemented. Basically the setting of electricity rates at any given time should be forward-looking (not backward-looking to historical costs). They should focus on the prospective flows of returns to investments that are to be made in the immediate future.

One thing that should redound to the benefit of past investments is an allowance for future obsolescence on the investments now being made. It is very likely that the project profile of a typical thermal plant will be characterized by a declining stream of benefits through time. This may stem partly from greater maintenance costs (including downtime) for older units, but I believe it mainly comes from the prospect that, just as new units currently being installed are likely to be lower in the stacking pattern (i.e., used for more hours per year) than older, somewhat similar units, so these currently-new units are likely to be displaced upward in the stacking pattern as still newer units of similar type come on line 5, 10, or 15 years hence. This future displacement is expected to occur because history tells us that technological advances will continue, even though we may not be able to describe at this moment what their precise shape will be. They key point is that if they are genuine advances, they will be cost-reducing in nature.

Thus when we (in our example above) employ a factor of .15 to represent the gross-of-depreciation rate of real return to capital, we really should say "gross of depreciation plus obsolescence". Such a rate might consist of an opportunity cost of capital equal to 10% per year, and a rate of physical depreciation (or greater maintenance cost) of, say 3% per year, plus a rate of obsolescence of, say, 2% per year.

Thus, if we were to conduct ex post cost-benefit analyses of a number of different thermal projects, we would likely find that some ended up yielding more, others less than the real opportunity cost of capital. And it would be a good guess that those that yielded less would be the ones that had suffered most from obsolescence. In the life of an already constructed thermal plant, this suffering is not measured in units of thermal efficiency or even of real economic cost per kwh. Rather, it is reflected in a plant's being squeezed out

by newer, better plants, so that it functions in its later years for a far smaller number of hours than was initially expected. It is also reflected in the retirement of older, less efficient capacity at an earlier-than-expected date. Both of these forces are represented in the formula

$$B_{jt} = \Sigma_i H_{ijt}(C_{it}-C_{jt})$$

Here B_{jt} is the flow of real benefits to plant j at time t. It consists of a weighted sum of cost advantages $(C_{it}-C_{jt})$, representing the economic rents that plant j is earning, per kilowatt hour, in the hours when plant i is the marginal plant actually operating. The H_{ijt} are the kilowatt hours actually produced by plant j; summed over i, they add up to the total energy output of plant j in year t. As any thermal plant ages, two things happen -- it operates for less total hours (thus producing less total energy), and the average cost advantage $(\bar{C}_{it} - C_{jt})$ that is has over the marginal plant in operation at each moment of time grows smaller and smaller as older, less efficient plants are one after another retired from the system.

An ex post evaluation would tend to reveal a normal rate of return if, for project j, the time path of $\Sigma_i (H_{ijt}(C_{it}-C_{jt})$ turned out to be more or less and anticipated. If this time path were above or below what had been expected, the ex post rate of return would tend to be above or below "normal".

14. The main benefits of competition in electricity generation come when entry is truly free, and the only guarantees given to the investing company are that its new plant j will be plugged into the system ahead of all higher cost plants, and that at such times it will receive a price equal to system marginal cost. If this SMC includes a peaktime surcharge sufficient to cover the annualized capital cost of new peaking (typically gas turbine) capacity, that is all the commitment that the system needs to make to the owners of plant j. In case a demand charge is levied to cover the annualized fixed costs of peaking capacity, new capacity

hooked into the system would be entitled, in place of a peaktime surcharge, to a fixed charge per KW of capacity ($60.00 per year, in the example above), on top of economic rents based on its marginal running cost advantage. (Plants would lose this fixed charge when they were retired from the system.)

Free entry of the kind just described is what guarantees that the system (or better, its customers) will benefit quickly from new advances in generating technology. I have no doubt, for example, that if this system had been in effect for the past decade or so, the revolution represented by combined cycle technology would have had a much more rapid diffusion through the system. The gains thus created would have come at the expense of older plants which would have suffered from a more rapid obsolescence and a lower ex post rate of return.

It is important to realize that the scenario just described is an essential part of the picture. If the older plants were guaranteed a "full" rate of return on their investments, the most likely outcome would be that the entry of new, more efficient capacity would be artificially delayed until the old plants had lived through a comfortable and profitable middle age and had reached the normal time of retirement. If no delay occurred, then some way would have to be found to pad the receipts of the older plants, so as to give them the guaranteed rate of return. That money would have to come either from electricity customers or from taxpayers. Neither of these is a pleasant thought.

I conclude that one must avoid rate-of-return guarantees if the true benefits of competitive electricity generation are to be achieved. What problem might appear, if no guarantee were given? Clearly, private investors might then insist on an expected rate of return that was somehow "too high". And how would that come about? Most likely, from

uncertainty and lack of transparency concerning the way in which the "privatized" system would be operated in the future. This is the reason I have placed so much emphasis on the need for creating an atmosphere of clarity, confidence and certainty about how the new system will operate. This is a task of highest priority, and the time to start is yesterday!

A Note on Transmission Pricing

15. Some economists have expressed the view that the pricing of transmission services can follow a principle similar to the one we have described for generation. In this section I attempt to explain why I do not believe this is the case.

The main source of the problem is the huge economies of scale that exist in the building and running of transmission lines. When old lines have reached their capacity, new lines ultimately have to be built. They consist of miles and miles of towers strung with high-tension cables. The problem is that the investment in such a system is enormously lumpy. A new transmission line connecting say, the center of Mexico with Monterrey, might be adequate to meet the increments of transmission needs for the next decade or two. Maybe the stringing of cables might be done in two or three steps over time, but without doubt the cost of towers has to be treated as one huge lump.

What this means is that most of the time transmission lines will be operating with lots of idle capacity. Bottlenecks will surely develop from time to time, but they would most appropriately give rise to ad hoc "rationing surcharges" to help surmount the bottleneck period. I do not believe a case can be made that is similar to the one supporting peaktime surcharges (as the standard means of covering the capital costs of new capacity) in a thermal generating system facing a steadily growing demand.

The big picture as far as transmission is concerned is that it costs more to bring energy to distant places than to near ones, and that this extra cost includes not only transmission losses but also the capital cost of the transmission lines themselves. Hence if energy is transported from the center of Mexico to the Monterrey area, then electricity should cost more in Monterrey than in the center. And if there is further transmission from Monterrey to Hermisillo, then customers in Hermisillo should be made to feel the economic cost of building and maintaining those transmission lines. Transmission costs should not be financed by a uniform charge spread equally over all the kwh produced in the entire system (i.e., in all regions of Mexico), or (as a lump sum demand charge) over all the KW of maximum demand in the entire system. Somehow, the subset of customers benefitted by each given transmission line should be made to perceive its cost.

The support for this view comes directly out of the decisionmaking process within the electricity network itself. When a need appears for a major expansion of deliveries of energy to Hermosillo, there are two obvious options: a) install a new thermal plant there, and b) draw the additional energy from the grid. At that point the cost of building the transmission lines is juxtaposed to the cost of local generation. If the grid wins, it means that electricity c.i.f. Hermasillo is cheaper sourced from the grid than produced locally. Obviously, this means that Hermosillo customers can get their energy cheaper, buying from the grid and paying transport costs, than by paying the standard economic rates that would apply if the corresponding capacity were built locally.

16. The question of how to charge for the costs of transmission is not easy, because some degree of inefficiency is inevitably involved. Perhaps the easiest place to start is with transmission losses. These are true losses, which must be taken into account by the load

dispatch center in its hourly allocations of sources to uses of energy. Thus, nearly all the time the system marginal cost of energy at a delivery node will be the system marginal cost at the point of origin, augmented to account for transmission losses. Energy whose SMC is 4¢ per kwh at the point of origin will have a SMC of 5¢ at the destination node, if 20% of the energy sent is lost in transmission.

But taking account of transmission losses in this way does nothing at all to reflect or cover the capital costs of transmission lines. A ray of hope appears when one considers those moments when particular transmission lines reach full capacity use. At such times it is natural and correct for the Load Dispatch Center to incorporate peak use surcharges in its calculations of system marginal cost. These will help determine the way in which thermal plants, both at origin and destination nodes, will be programmed in order to avoid a system overload.

Peak use surcharges for transmission lines are thus, in my view, an essential component of the economic management of the system. But they are more a management tool than a major source of revenue. Though they would be incorporated in system marginal costs at the affected nodes, they are very unlikely to provide anywhere near the sum needed to defray the capital cost of the system's transmission network.

Thus we are left with the need to find other ways of covering these costs. I think this problem is quite analogous to the problem of collecting for the capital costs of peaking capacity under the dry-season scenario. But whereas in that case I lean toward a demand charge based on maximum use, in the present case I lean toward a surcharge (or better, a series of surcharges) per kilowatt hour used. This spreads the cost as widely as possible among the actual users of transmitted energy. Thus, it generates a lot of little triangles of efficiency

cost rather than a few big ones. This creates a presumption of lower total efficiency cost, following the principle that the size any given efficiency cost triangle varies with the square of the distortion in question.

My preference here is also influenced by the procedures that might be followed in billing for the surcharge. As I perceive it, electricity users would in principle pay an energy charge per kwh that was somehow related to system marginal cost (see Section 17). Then on top of this they would be billed for a series of transmission line costs. The costs of the line from Monterrey to Hermosillo would be spread among Hermosillo customers (and others along the line). The costs of the line from the center to Monterrey would be shared by those in Monterrey and along that line, plus all those (including those on the Hermosillo line) receiving energy from transmission lines fanning out from Monterrey to more distant distribution nodes. In this way, the prorated costs of building each major transmission line segment would be spread among essentially all those who share in its benefits. It is probably better to prorate costs on a per-kilowatt-hour basis rather than simply a certain amount per year. Otherwise the few who are using as new line in its early years would be very heavily burdened compared to the many who would be using the same line in its later years. Thus, the capital costs (K) of a new transmission line would be divided by the total number (T) of kwh that are expected to be sent over that line in its economic lifetime. Then during the j^{th} year of operation of the line each user would pay a surcharge of $(K/T)(1+r)^j$ per kwh carried over that line. The amount to be collected in year j would be $C_j(K/T)(1+r)^j$, where C_j is the total number of kwh expected to be carried over the line in year j. The per kwh surcharge on a customer's bill would be $(C_j/H_j)(K/T)(1+r)^j$, for the line in question. Customers in Hermosillo would probably end up paying a series of such surcharges --

perhaps one from Grijalva to the center, and from the center to Monterrey, and one from Monterrey to Hermosillo.

On the question of how to deal with cases where the actual transmission traffic is more or less than estimated, my instincts are to eliminate the charges for any old lines, once their capital costs have been "paid off". Perhaps one could require that, say, 1.2 times capital costs be paid off before a charge is eliminated, with the additional 20 percent being set aside for covering shortfalls on lines whose traffic did not grow as expected.

Sometimes, transmission lines carry energy in one direction during certain hours, and in the other direction at other times. My suggestion would be to count traffic in both directions and to assign the cost $(K/T)(1+r)^j$ per kwh transmitted in each respective direction to its corresponding set of beneficiaries.

Wholesale and Retail Electricity Pricing

17. By now there is, I believe, a quite general acceptance of the idea that the relevant economic cost of electricity varies by hour of the day and season of the year. Hence, if it could be done costlessly, something like a continuously varying time pricing of energy would be optimal. Unfortunately this is not (yet) the case, so in fact we have time pricing for only a subset of customers. But readers should realize that the costs are not outrageous. At our home in California, for example, we have for several years been paying different rates for peak and off-peak energy use.

What peak/off-peak pricing does is approximate a system marginal cost function that has many steps, with a much simpler approximation that has only two steps. Sometimes this leads to trouble, as when the institution of a peak charge, say from 7 to 10 p.m., creates a new peak, say from 5 to 7 p.m. There is no mystery to this so-called "shifting peak"

problem. The answer is to make a better approximation to the underlying system marginal cost curve. Very early in its applications of time pricing, Electricité de France found that simple peak/off-peak pricing would do the trick in summer, but that one needed a superpeak/peak/off-peak schedule in order to handle the problem in winter.

Personally, I don't think it will be very long before instantaneous time pricing will be economically feasible, and that is the direction in which we should probably be thinking. That would face most users with the true system marginal cost of each unit of energy they use, plus whatever surcharges might be in order to cover fixed costs of transmission lines, standby capacity for dry years, administrative and distribution expenses, etc. The issue of whether these latter charges should be levied in a maximum demand basis, or as a per kwh surcharge, should be dealt with as a problem in "second-best" applied economics on a case-by-case basis.

Before committing myself completely to continuous time pricing as a long term goal (and theoretical ideal), let me mention a possible qualification. With continuous time pricing, users would typically not know in advance what price they would be paying per kwh at each future moment in time -- they would only know what prices they had in fact paid at particular times of day on particular past days of the week (or year), and they might also receive from the electricity company a prospective schedule of likely rates to apply at different hours and days of, say, the next month. This puts the risk of rate variation (around the expected time-price for each hour) squarely on the back of the consumer.

Now consumers may not appreciate this risk, and may in fact be willing to pay the electricity company to assume it for them. There is nothing wrong, in such a case, for the risk to be transferred to the company, with the customers being faced with a known schedule

of time prices for energy, instead of with the certainty that they will never pay either more or less than the true system marginal cost at each moment.

The great likelihood is that the more sophisticated pricing schemes are likely to be preferred by large industrial customers, while simpler schemes will better suit the preferences of residential and small commercial users. This fits well with the way electricity systems have evolved up to now, with time pricing in particular working its way gradually down from the big industrial (wholesale) customers toward customer groups with smaller and smaller average consumption (i.e., retail customers). Some such differentiation among the billing methods used for different classes of customers is likely to be a permanent feature of electricity systems. Modernization just brings about changes in the way these systems are managed, and in the sense that modern pricing systems have much closer links to good economic analysis than the older systems they are in the process of replacing.