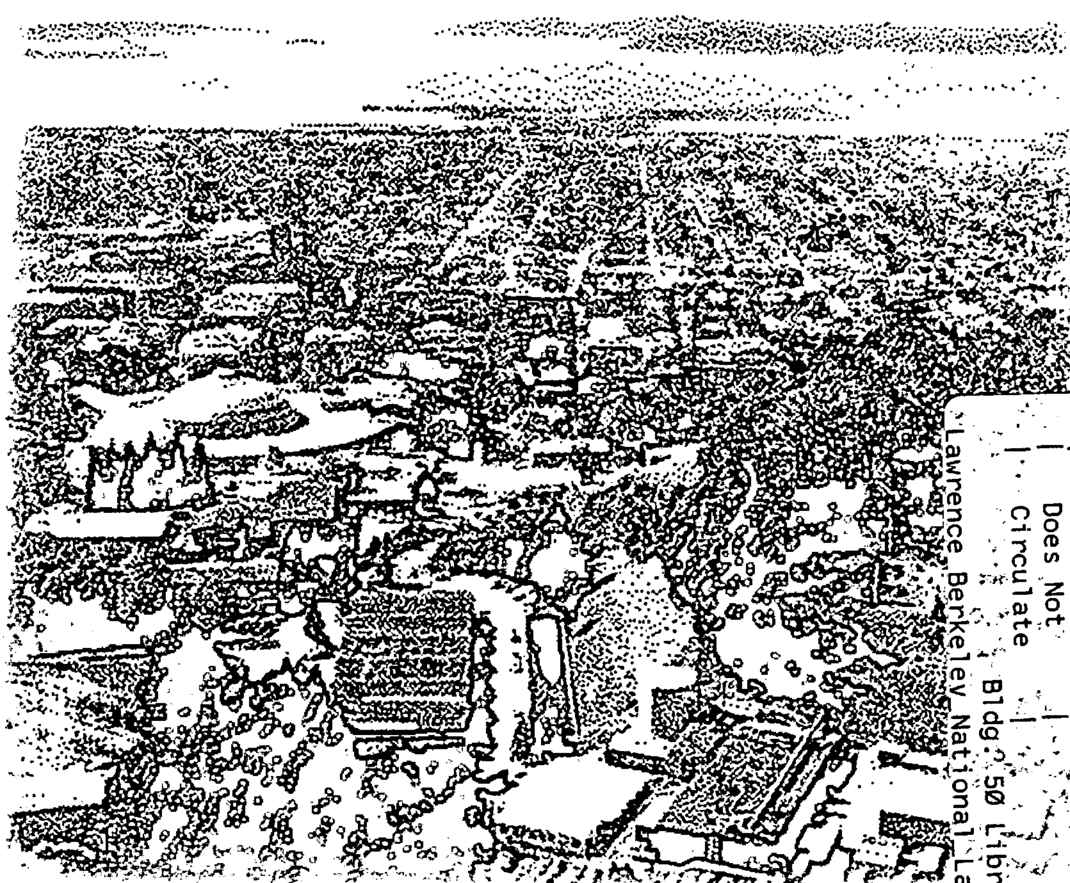
LBNL-39445
UC-900

Ernest Orlando Lawrence Berkeley National Laboratory

Analysis of the California WEPEX Applications to FERC

Steven Stoft
Environmental Energy Technologies Division

October 1996



DISCLAIMER

LBNL-39445
UC-900

Analysis of the California WEPEX Applications to FERC

Steven Stoft

Environmental Energy Technologies Division
Ernest Orlando Lawrence Berkeley National Laboratory
University of California
Berkeley, California 94720

October 1996

The work described in this study was funded by the Assistant Secretary for Policy and International Affairs, Office of Electricity Policy of the U.S. Department of Energy, under Contract No. DE-AC03-76SF00098 and by the University of California Energy Institute.

Table of Contents

Abstract

The major theme of this report is that the WEPEX Applications prevent the ISO from clearing the market, and that this is the root of the most important problems. The proposed rules at fault (1) prevent the PX and other Scheduling Coordinators from passing on all of their bids to the ISO, and (2) prevent the ISO from dispatching beyond the point at which congestion is eliminated. Although it is generally accepted that these restrictions prevent the ISO from achieving the least-cost dispatch, many other consequences of this market-clearing failure have not been widely recognized. These include sub-optimal dispatches by the PX when the system is uncongested, congestion charges that reward power flow in the congested direction, and incentives for Scheduling Coordinators to ignore known intra-zone congestion. But the most pernicious effect of failing to clear the market may be decreased system reliability.

Four minor themes will also be considered but in less detail. Most importantly we describe several examples of unequal treatment for the PX. Second, we will discuss the ambiguities introduced by zonal pricing. WEPEX's zonal system is based on a view of congestion that largely ignores loop flow and consequently has not been well defined. A replacement definition is offered. Third, WEPEX appears to have invented a new definition of the transmission congestion contract (TCC) that is based on actual instead of pre-specified flows. We will show that this ruins the incentive properties of TCCs, but that this problem is partially rectified by the market in TCCs. Lastly we will discuss losses. The WEPEX proposal intentionally avoids marginal-cost pricing of losses. It is known that this avoidance will cause some inefficiency, but what has not been recognized is that it significantly increases use of the power grid. Because this effect is greatest during peak usage, it will necessitate costly grid expansion.

Acknowledgments

The work described in this study was funded by the Assistant Secretary for Policy and International Affairs, Office of Electricity Policy of the U.S. Department of Energy, under Contract No. DE-AC03-76SF00098 and by the University of California Energy Institute.

The development of several examples provided by this paper was stimulated by John Chandley's work found in the California Energy Commission's Comments (1996) to the California Public Utility Commission.

The author would like to thank Bill Hogan, Bart McGuire, Shmuel Oren, Pablo Spiller, Jim Bushnell, Greg Basheda, Dave Meyer, and Chuck Goldman for many helpful comments and suggestions.

Terminology

This paper analyzes two applications filed with FERC by the California IOUs: the ISO Application and the PX Application. But because the ideas contained in these applications were developed by a much broader coalition known as the Western Power Exchange, or **WEPEX**, these will often be referred to as the WEPEX applications or the WEPEX proposal. This is not meant to imply that all WEPEX parties agree with the entire application.

The main actors described by the WEPEX applications are the independent system operator (**ISO**) which facilitates trading over the power grid and the scheduling coordinators (**SCs**) who are intermediaries between groups of traders and the ISO. One SC, the power exchange (**PX**), is of special interest because it is designated to provide a public spot market in bulk power. Other SCs, which I will term bilateral exchanges (**BXs**), are intended to facilitate trades among groups of bilateral traders.

Because this report makes use of much jargon that is specific to the California setting, a thorough glossary containing most nonstandard terms and all acronyms is included. The reader who is unfamiliar with the WEPEX debate may find it worthwhile to begin with the glossary.

Glossary

Access charge	A charge for transmission, paid by all entities withdrawing power from the ISO grid. It recovers the TOs' transmission revenue requirement.
The Applicants	PG&E, SDG&E, and SCE. (See also the Companies.)
Bus	A node. A point where transmission lines are joined electrically.
Bilateral exchange	A "non-PX scheduling coordinator." (This is non-standard notation.)
BX	Bilateral Exchange.
CEC	California Energy Commission.
CFD	Contract for differences. There are two polar versions, one fixes the generator's price, the other fixes the load's price. The generator's version is: the load pays the generator $Q_C \times (P_C - P_i)$, Q_C and P_C are the contract quantity and price, and P_i is the price at the generator's bus.
The Companies	PG&E, SDG&E, and SCE (See also "The Applicants")
CPUC	California Public Utility Commission
CTC	Competition Transition Charge. A non-bypassable charge to customers that provides full recovery of Applicants' stranded assets.
COB	California-Oregon Border. One of two delivery points for electricity futures traded on NYMEX. (See also Palo Verde.)
Dec	A generator's "decremental" bid for reducing generation. If a generator is decremented it, must buy replacement power at its Dec or the market price, whichever is less.
DRA	Division of Ratepayer Advocates. Part of the CPUC.
Edison	Southern California Edison Company.
FERC	Federal Energy Regulatory Commission.
FPA	Federal Power Act.
Feasibility rule	Rule for allocating TCCs to grid investors. Requires that total ISO-issued TCCs be equivalent to a feasible dispatch of the grid.
Forward markets	The day-ahead and the hour-ahead markets for energy.
Grid operations charge	The charge levied on SCs by the ISO for within-zone congestion. (See also "usage charge.")
Inc	A generator's "incremental" bid for generating additional energy.
IOU	Investor Owned Utility. (Generally one of the Applicants.)
IPP	Independent Power Producer. (Not a utility.)
ISO	Independent System Operator.
ISO Application	An application filed with FERC on 4/29/96 by the Companies under Section 203 of the FPA to transfer control of the

	Companies' transmission grids to an ISO. FERC Docket No. EC96-19-000.
Interface	The set of lines (path or network) connecting two zones.
NERC	North American Electric Reliability Council
Node	A bus. A point where transmission lines are joined electrically.
Non-spinning reserves	Load or off-line generation that can be ramped down or up in 10 minutes and maintained for 2 hours.
OASIS	Open Access Same-time Information System. The new name for TSIN.
OPF	Optimal Power Flow. The least-cost dispatch and the associated power flows on the grid. Also the process of finding this.
Palo Verde	One of two delivery points for electricity futures traded on NYMEX. See COB.
Replacement reserves	Load or Generation that can be ramped up in 60 minutes and can be maintained for 2 hours.
PG&E	Pacific Gas and Electric. 75TWhs (~$10B)/year. 18000 miles of 60+kV transmission. (ISO Application, §1)
Policy Decision	The CPUC's Policy Decision of December 20, 1995.
Preferred schedule	Suggested hourly dispatch of generation and load delivered to ISO by a scheduling coordinator. 24 hourly schedules are submitted a day in advance, and one updated hourly schedule is submitted each hour, 1 hour in advance
PX	Power Exchange. The provider of the nodal spot market.
PX Application	An application filed with FERC on 4/29/96 by the Companies under Section 205 of the FPA for authorization to make sales at market-based rates through the PX. FERC Docket No. ER96-1663-000.
Real-time market	Market in which the ISO balances generation and load second-by-second. (ISO App. §5.3.1.2)
PUC	Public Utility Commission.
SC	Scheduling coordinator.
Scheduling coordinator	The PX or any other party (BX) certified by the ISO to submit schedules. Schedules must be balanced (generation = load).
Scheduling markets	The day-ahead and hour-ahead markets (as opposed to real-time). Binding transactions are made in both.
SCE	Southern California Gas and Electric. 75TWhs (~$10B)/year. 5000 miles of 230+kV transmission. (ISO App. §1)
SDG&E	San Diego Gas and Electric Company. 15TWhs (~$2B)/year. 2000 miles of 69+kV transmission. (ISO App. §1)
Section 203	Part of FPA.
Spinning reserve	Generation capacity that is synchronized, responds automatically to system frequency, and can be ramped up in 10 minutes and maintained for 2 hours.

TCC	Transmission Congestion Contract. A financial contract for a specific power flow, Q, between two nodes, i and j, stating that the contract owner will receive Q×(Pj − Pi), where Pi and Pj are nodal spot prices.
TCR	Transmission Capacity Reservations. FERC's new proposal on implementing open access. Not directly addressed in the Applications.
TO	Transmission Owner. Initially, just the three Applicants.
TSIN	Transmission System Information Network. (See also "OASIS.")
UDC	Utility Distribution Company, the distribution part of the Companies.
Usage charge	A charge assessed by the ISO on the SCs for flows on congested inter-zonal paths and recovered by the PX from loads in the higher-priced zone. (ISO Application, 5.4.2.2.1.2) These revenues will be used to reduce the transmission revenue requirement. (See also "grid operations charge.")
WEPEX	Western Power Exchange. Voluntary association of the Companies, and all other parties interested in California restructuring. Final power of decision rests with steering committee which has one representative from each of the Companies.
WSCC	Western Systems Coordinating Council. A NERC regional council.
Zone	A portion of the ISO grid where congestion is expected to be small in magnitude or infrequently occurring and generally does not occur under normal conditions.
Zone hub	A central bus (node) at which the zonal price is defined. Generators are paid the hub price minus a loss adjustment.

Executive Summary

Over the last two years the California electricity industry has engaged in a vehement, if not always well informed, debate over the appropriate shape for a competitive electricity market. The central focus of this debate has been the nodal-pricing/bilateral-trading dichotomy. The current resolution of this debate, a compromise in which both market structures are allowed, is embodied in the WEPEX applications to FERC.

The two central features of this compromise are (1) a separation of the nodal spot market (PX) and the market for bilateral trades (BXs), and (2) the creation of zones. The separation of the markets is essential to the compromise, but the creation of zones is simply a failed attempt at simplification. The problems cataloged in this paper in no way indicate a problem with separating the two markets. However, the separation has been carried out in an unnecessarily bureaucratic fashion that will result in disputes and inefficiencies unless several remedial simplifications are implemented.

For a successful nodal/bilateral split, it is necessary to clear the nodal market and to allow bilateral traders the option of trading with the least possible interference from the ISO. As we will see, the ISO application prevents both the PX and the ISO from clearing the PX market and prevents the ISO from clearing the combined market. We will also find that nodal pricing has been distorted by the zonal approach and by the average-cost pricing of losses. Thus, true nodal-pricing is not available to participants in the PX. The bilateral markets are however quite free of interference. Bilateral traders are not required to submit Inc/Dec bids, and are allowed to self-provide ancillary services.

Most of the problems discussed in this paper result from complicating restrictions that are placed on the PX and ISO and which distort the nodal spot market. These do not seem to provide any advantage to the bilateral market, except perhaps by enhancing its relative appeal by damaging the PX. Ironically, this tendency towards complexity is exhibited most dramatically by zonal pricing which was adopted for the sole purpose of simplification. As we will see, under zonal pricing the ISO is still required to do a least-cost constrained optimization of the network, taking into account all submitted bids. The only simplification the ISO is allowed, and this simplification is an extremely minor one, has nothing to do with nodal pricing but is instead the restriction that prevents the ISO from clearing the market. Since prices are simply a byproduct of the ISO's constrained, least-cost dispatch, zones do not simplify the ISO's calculation. Beyond this, zones introduce major complexities because of the need to determine zonal boundaries and the need for the ISO to distinguish between intra-zonal and inter-zonal congestion.

The thrust of the examples presented in this paper should in no way be interpreted as an attack on the idea of a nodal-bilateral compromise. In fact the suggestions in this paper are designed specifically to make this compromise workable. Also this paper should not be seen as attacking the idea of using zones to simplify nodal pricing at the cost of some lost efficiency. Such a trade-off is eminently sensible; the point made here is simply that the

current proposal has not made such a trade-off. Instead, it has simply increased complexity and introduced new opportunities for dispute and litigation while decreasing efficiency as expected.

Six flaws in the WEPEX applications underlie the problems discussed in this paper, and they are listed here for quick reference. The second flaw, restrictions on the ISO's ability to redispatch, is mentioned most frequently.

- **The ISO's Redispatch Restriction:**
  The ISO may redispatch only to the point of relieving congestion but not beyond.

- **The Bid Restriction:**
  SCs may submit only bids from generators in their preferred schedules to the ISO.

- **PX Restriction:**
  The PX must dispatch around known constraints.

- **Load Price Averaging:**
  All loads in a zone share equally the cost of intra-zonal congestion caused by any generator-load pair in the zone.

- **Zonal Definition:**
  Zones are defined simply as areas with little congestion, not areas where marginal cost is nearly uniform.

- **Average-Cost Loss Pricing:**
  Total losses are divided among all generators.

Key Findings

The problems mentioned above are analyzed in Sections 2 through 7 which are described briefly below. Typically these sections contain an example which make the main points.

The PX-ISO Relationship (Section 2)

Both the PX and ISO can fail to achieve a reasonable dispatch of PX bidders because of a combination of the bid and the redispatch restrictions (see Example 1).

The disallowance of a second round of bidding also hinders efficiency. WEPEX noticed that a second round is helpful to the SCs for adjusting to unexpected congestion but failed to notice that it could be equally useful for adjusting to an unexpected lack of congestion.

The PX is shown to be disadvantaged relative to other SCs because of the PX restriction requiring it to account for known grid limitations.

The ISO's Limited Redispatch (Section 3)

Example 2 shows that the restriction on redispatch causes the ISO to leave too many lines at their security limits. In fact every line that is over-committed by the set of "preferred dispatches" received from the SCs must be left in this state.

Because such lines present a greater security risk than lines that have been dispatched well below their security limit, the redispatch restriction will cause the network to be less secure than if the ISO cleared the market and achieved the least-cost dispatch.

Congestion Management Under Zonal Pricing (Section 4)

A combination of diversity among SCs and the redispatch restriction imposed upon the ISO will sometimes cause inter-zonal congestion to be negatively priced (see Example 3). This rewards those who add to congestion. Surprisingly this makes sense given the WEPEX redispatch restriction because additional congestion gives the ISO greater scope for redispatch, thereby allowing it to clear the market more completely.

Example 4 shows that load-price averaging over zones will make it beneficial for SCs not to account for intra-zonal congestion in their dispatches. To do so would disadvantage their customers. Unfortunately the PX is required to, so its customers are put at a disadvantage.

Defining Zone Boundaries and Zones (Section 5)

If "interfaces between zones" are "defined by paths or networks for which the expected demand may often exceed the path or network ratings," then the definition of zones is self contradictory. In particular, because of loop flow, if congested paths are defined to be zonal interfaces, then some zonal interfaces must necessarily be uncongested paths (see Example 5A).

This contradiction is the result of WEPEX ignoring the problems of loop flow in its definition of zones. Example 5A is also used to develop a definition of zones based on nodal prices that resolves this contradiction, at least for the simplest examples of loop flow.

Example 5B further demonstrates the need for price-based zoning.

Transmission Congestion Contracts (Section 6)

Example 6 demonstrates that the way WEPEX appears to define TCCs leads to improper incentives for generation.

Losses (Section 7)

The WEPEX procedure for collecting losses will encourage excessive use of the transmission system and lead to significantly increased losses (see Example 7). Also, in order to compute the zero-loss-revenue prices required by WEPEX, it is necessary to compute optimal prices first, so the WEPEX procedure is more complex than optimal pricing.

1 Introduction

On December 20, 1996 the California PUC ordered California's investor-owned utilities (PG&E, SCE, and SDG&E) to file applications with FERC to establish a unified power grid operated by an independent system operator (ISO) and a public hourly spot market for bulk power operated by the Power Exchange (PX). The two applications were (1) the "ISO Application," filed under Section 203 of the Federal Power Act (FPA), to transfer operational control of the Company's transmission systems to the ISO, and (2) the "PX Application," filed under section 205 of the FPA, to establish the PX and allow sales at market-based rates. This paper analyzes the economic mechanisms contained in the two FERC Applications.

A large number of parties have collaborated directly, and indirectly through the WEPEX process, in the creation of the proposals presented in these applications. In many respects they represent the most advanced proposals put forth to date in the U.S. for creating a competitive and efficient market for the trading of electricity. However, the WEPEX process has occurred at breakneck speed with compromises required of many parties. As is inevitable in this environment, a number of mechanisms work at cross-purposes. These cannot be attributed to the intent of specific parties, and, even if that were the case, this paper would not be concerned with such attributions. The point of this study is to deepen public understanding of the mechanisms chosen by WEPEX to construct this market, and to examine contradictions and flaws before they are implemented.

1.1 Methodology

The mechanisms defined in the WEPEX application are extremely complex as is the actual market. In order to make progress in our analysis it is necessary to simplify. Fortunately any errors due to simplification will be biased towards making the mechanism look better than it is. For instance, I will generally assume that the market is perfectly competitive which causes bidders to bid their true marginal costs. Now it is certainly possible, if not likely, that this assumption is wrong, and so the WEPEX mechanisms will not behave as predicted when actually put to the test. But what discrepancy should we expect? If we find an example where, assuming perfect competition, the dispatch is wasteful, is it at all likely that in a less competitive market bidders would game their bids in exactly the way required to induce a perfectly efficient dispatch? Once a mechanism has been shown to fail in an idealized world, the burden of proof shifts to those asserting that this same mechanism functions as promised in the far hotter crucible of the real world.

In this spirit we will make many simplifications which probably eliminates the chance of finding many subtle problems. Nonetheless we will be able to discover a sufficient number of anomalies to keep our interest. The following simplifications will be used in the analysis which follows: (1) No market power, (2) No congestion, (3) No losses, (4) No spinning reserves needed, and (5) No startup or no-load costs.

Not all of these simplifications will be made at once; usually we will relax (2) or (3) to create just enough realism to illustrate the points of current interest.

2 The PX-ISO Relationship

Summary:

Example 1 demonstrates that both the PX and ISO can fail to achieve a reasonable dispatch of PX bidders because of a combination of bid and redispatch restrictions.

The disallowance of a second round of bidding also hinders efficiency. WEPEX noticed that a second round is helpful to the SCs for adjusting to unexpected congestion but failed to notice that it could be equally useful for adjusting to an unexpected lack of congestion.

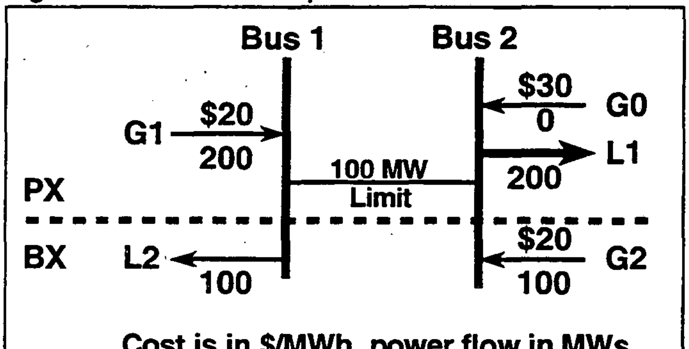
The PX is shown to be disadvantaged relative to other SCs because of the PX restriction requiring it to account for known grid limitations.

2.1 Example 1: A Sub-Optimal PX Dispatch

A number of parties in California have observed that neither the PX nor the ISO can discover the least-cost dispatch because neither has access to the complete list of bids.[1] The PX is not given the BX bids, and the ISO is given neither PX nor BX bids that fail to make their SC's "preferred schedule."[2]

Figure 1 shows an *optimal* dispatch of a simple power system with three generators, two loads and one transmission line. Note that this dispatch can be viewed correctly in two ways. First, we can say that 100 MW of power flows from G1 to L2, 100 MW flows from G1 to L1, and 100 MW flows from G2 to L1. This view is most closely aligned with the actual physics of the situation and shows that there are only 100 MW flowing on the line, so the congestion limit is not reached. Second we can use the electrical engineer's "superposition principle" to reflect

Figure 1. Least-Cost Dispatch



Cost is in $/MWh, power flow in MWs.

---

1 California Energy Commission's Comments on the WEPEX Applications. May 28, 1996. John D. Chandley Assistant Chief Counsel.

2 The author gave a simple example of this point to the full WEPEX transmission team on December 12, 1995, and that example will serve to illustrate here the CEC's point.

contractual arrangements and analyze the flows as 200 MW flowing from G1 to L1 in the PX and 100 MW flowing from G2 to L2 in the BX. In this case, we find that the flow from G2 to L1 cancels 100 MW of the flow from G1 to L1, for a net flow on the line of 100 MW, and again the contingency limit is satisfied.

The total cost of meeting the load is $6,000/hr ($20/MWh × 200 MW + $20/MWh × 100 MW). Note that the PX generator is fully supplying the PX load, while the BX generator is fully supplying the BX load. There has been no need for curtailment by the ISO in order to meet transmission contingency constraints. This is the optimal, least-cost dispatch (assuming G2 is limited to 100 MW).

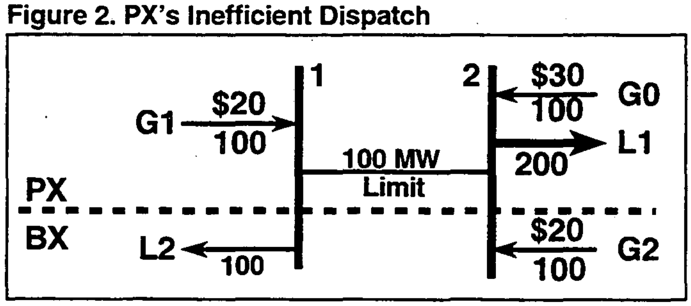
Counter Flows

It should be noted that counter flows, which are the basis of Example 1, are entirely a matter of interpretation. Any dispatch of more than minimal complexity can be interpreted as having counter flows or not, depending on which buyer is matched with which seller. In a competitive market this matching will be unpredictably rearranged, resulting in apparent counter flows in many new locations, although the dispatch is no different from what it was under regulation.

Now let us analyze what would happen under the Applicants' proposed market mechanism. First, the PX would submit its preferred (cost minimizing) schedule "taking into account known transmission constraints" (PX Application, 6.3.1). Clearly the transmission constraint will require the PX to dispatch 100 MW from Generator 1 and 100 MW from Generator 0. The BX will notice no such constraint and will simply make the obvious dispatch of 100 MW from G2 to Load 2.

When the ISO is presented with these schedules it will find zero flow on the congested line and thus no congestion. As required "if there is no congestion" the ISO "will inform the scheduling coordinators of ... the zonal market-clearing prices that are associated with the final day-ahead schedules" (ISO Application C.a.16). Because of this requirement, there will be no second round in the day-ahead auction, and, even if there were a second round, no additional information would be supplied by the ISO to the PX, so there would be no change in the PX bids. The result is the inefficient dispatch shown in Figure 2.

Figure 2. PX's Inefficient Dispatch



Note that in this dispatch, the cost of supplying the load is $7,000/hour (20/MWh × 100 MW + $20/MWh × 100 MW + $30/MWh × 100 MW). The faulty WEPEX mechanism has cost the PX's customers $1,000/hour. In addition, their loss is not the PX generator's gain. All that has happened is that a more expensive generator has been substituted for a less expensive generator. The loss is a dead-weight loss. Not only is this outcome wasteful, it is fully correctable within the PX without any use of BX bids.

2.1.1 *The Financial Magnitude of this Inefficiency*

It is impossible, at this time, to estimate the likely cost of the sub-optimal dispatches that would result from this aspect of the WEPEX proposal, but it may be helpful to consider the following calculations. Path 15 between Northern and Southern California is rated at 2,800 MW going south and 2,300 MW going north. Typically, PG&E transmits power north from the Diablo Canyon nuclear plant and SCE ships power south from the Northwest. Because these contractual flows partially cancel, it is possible for both parties to ship more than the path rating. (Of course what happens physically is that the Northwest power supplies PG&E, and the Diablo Canyon power goes south to SCE.) Because both PG&E and SCE are required to use the PX for the first five years, this should not cause a problem. But, if after the five-year transition period, PG&E later leaves the PX (or if some non-PX party gets involved sooner) then the PX might find it impossible to schedule SCE's full flow south because it could not take into account PG&E's northerly flow. Not accounting for PG&E's flow could easily reduce Path 15's capacity from SCE's point of view by 2,000 MW. Since the price differential between Palo Verde in the south and COB in the north is frequently as much at $5/MWh, this could result in a loss of $10,000/hr.

Such losses may occur frequently and on several different paths of this magnitude. In the long run, however, persistent problems of this type would force a realignment of contractual arrangements or would force parties to leave the PX. When this occurred, it would diminish the magnitude of this type of problem.

2.1.2 *Is There an Easy Remedy for this Problem?*

One might think that the ISO could just look at this situation and "do the right thing." After all, who would complain? But the Applications explicitly disallow this, saying "The ISO will only make scheduling adjustments which act to relieve congestion, and will cease making adjustments when congestion is relieved" (ISO Application, C.a.17). A second possibility is that the PX would predict the counter flow and take it into account when creating its preferred schedule. This is problematic because the PX application states that "These principles will ensure that the price determination by the PX is unambiguous" (PX Application, 6.4.1). This would seem to be a necessary requirement but one that is not compatible with the PX speculating about possible BX counter-flows when it determines which bids to place in its preferred schedule. That is probably why the ISO Application states that "The PX ... need not include estimates of congestion which might occur due to non-PX schedules" (ISO Application, C.a.8).

One partial remedy for this unacceptable outcome is to require the ISO to allow a second round of bidding even when it finds no congestion. Prior to this round it should inform all SCs of the available capacity of all lines. This procedure would be exactly analogous to the WEPEX requirement that in the case of congestion "to enable the market to effectively relieve potential transmission congestion, the ISO will identify the relative effectiveness of generation shifts in alleviating potential congestion" (ISO Application, C.a.17) In this case the ISO would simply report the capacity that would remain available on each line if the currently submitted schedules were taken as the final submissions. These can be unambiguously determined, so the PX would not be required to make estimates or to use its best judgment.

2.1.3 *Can the PX-ISO Communication Problem Be Completely Solved?*

We have now seen that there are two remedies for the particular problem demonstrated in our first example: (1) allowing a second round of bidding when there is no congestion, and (2) directing the PX to ignore all transmission constraints when it submits its preferred schedule. But the question remains: would these remedies completely eliminate this type of sub-optimal PX dispatching; and, if not, what would be required for a complete fix?

The answer is that these fixes are partial and that the only complete fixes require an effectively complete information transfer from the PX to the ISO or from the ISO to the PX. This could be achieved with many rounds of partial information transfer or with a single round of complete information transfer. The point is that cost-minimization is an extremely complex problem, and it cannot be accomplished by two parties with such partial information. The surprising thing is that a complete-information procedure was not proposed. It is simpler to specify, simpler to implement, much more effective, and it is transparent. Such a solution would look like this:

1) The PX submits its full set of bids to the ISO.
2) The BXs submit whatever bids they want to the ISO.
3) The ISO finds the least-cost feasible dispatch.

Any BX generator who wants to lock in a dispatch of Q MW and does not want to participate in ISO redispatching simply states this and does not submit Inc/Dec bids. The ISO will then accept this generator's dispatch as fixed unless it is necessary to change it for security reasons.

Why has such a simple and obvious solution not been adopted? There seem to be two popular arguments against this solution:

1) BX clients do not mind the ISO using Inc-Dec bids for relief of congestion, but they do not want them used beyond this for cost minimization.
2) Finding the cost-minimizing dispatch is simply too difficult.

These arguments have led first to the restriction on ISO redispatch cited above and second to zonal pricing. Section 4 examines in detail the consequences of the ISO restriction, and the following two sections takes up zonal pricing.

2.2 The PX Is Disadvantaged

Is the PX treated equally to other SCs? No, the PX is at a clear disadvantage simply because of the special restrictions placed on the way it formulates the schedules it must submit to the ISO. One PX restriction that may well be unnecessary, and even harmful, is the completely innocuous sounding requirement that in formulating its preferred schedule the "PX will ... minimize the total cost to PX buyers of meeting the bid-in demand, subject to producing a reliable and feasible schedule (taking into account known transmission constraints) ..." (PX Application, 6.3.1). Thus the PX cannot submit a schedule that violates known transmission constraints. The BXs are under no such obligation and can very well use this to their advantage. So if a BX were given the same bids as given to the PX above, they could submit the optimal schedule (as shown in Figure 1) as their proposed schedule, and they would be successful in doing so. By bidding in apparent violation of a transmission constraint, they would supply their load customers with cheaper power. There would be no losers. But, the PX will not be able to withstand the competition from BXs with fewer bidding constraints.

3 The ISO's Redispatch Restriction

Summary:

The PX is shown to be disadvantaged relative to other SCs because of the PX restriction requiring it to account for known grid limitations. Example 2 shows that the redispatch restriction causes the ISO to leave too many lines at their security limits. In fact every line that is over committed by the set of "preferred dispatches" received from the SCs must be left in this state.

Because such lines present a greater security risk than lines that have been dispatched well below their security limit, the redispatch restriction will cause the network to be less secure than if the ISO cleared the market and achieved the least-cost dispatch.

In the appendix to the ISO Application, the Applicants are quite emphatic about limiting the ISO's redispatch when it manages congestion. First they say that:

> "... the ISO may only adjust the scheduled output of generation based on the price bids, and then only as required to relieve congestion." (ISO Application, C.a.17)

Then, they quickly reiterate that:

> "The ISO will only make scheduling adjustments which act to relieve congestion, and will cease making adjustments when congestion is relieved."

This is stated in full three times, first for the advisory redispatch of the day-ahead market, then for its final redispatch, and then once again for the hour-ahead market.

This rule has been in contention for over a year, and it is clear that the Applicants have come down firmly in favor of so restricting the ISO and against allowing it to find the least-cost dispatch by clearing the market. (Numerous references to market clearing prices indicate that they did not intend to prevent market clearing, but they apparently forgot this inevitable consequence of preventing voluntary trades.) This does not mean they wish the ISO to ignore cost, rather they want it to dispatch to least cost, subject to a restriction that prevents it from going all the way to a cleared market. We will now investigate that restriction and its implications.

Clearing the Market

The Applications speak often of the "market-clearing price," yet we have just seen that the ISO is not allowed to clear the market. This confusion deserves a closer look. "Clearing" a market means making all available profitable trades. Once these have been made, trade ceases and the market is "clear." If all profitable trades have been made, then all remaining bids to sell must be greater than all remaining bids to buy.

Inc and Decs

This definition needs some explanation in the WEPEX context because of the unconventional use of incremental bids (Incs) and decremental bids (Decs), and because throughout the WEPEX Applications demand bids are generally ignored. While the WEPEX restriction on redispatch would prevent some trades with demanders, for the sake of comparability we will continue to ignore demand. Sticking to the supply side, we can still apply the concept of "market clearing" because a Dec is really a bid to buy replacement power. A trade is made by a Dec generator curtailing generation and buying replacement power at the Dec price (or at the market price, whichever is less). So long as there is an Inc that is greater than some Dec, a profitable trade can be made either by the ISO or through bilateral agreement. So as long as there are some Decs that are greater than some Incs the market has not been cleared.

The Market-Clearing Price

Although it is possible to clear the market by arranging many pair-wise trades, it is also possible to clear it by establishing a single market-clearing price (MCP). This is a price at which the amount of power offered exactly equals the amount of power demanded. (Often there will be some bids at the MCP; it is simplest to count these only to the extent they are needed to match other accepted bids.) Once all accepted trades are executed, no pair of traders will remain who can trade profitably because all remaining offers to sell (Incs) will be greater than the MCP and thus will be greater than all remaining offers to buy (Decs).

Marginal Cost and the Least-Cost Dispatch

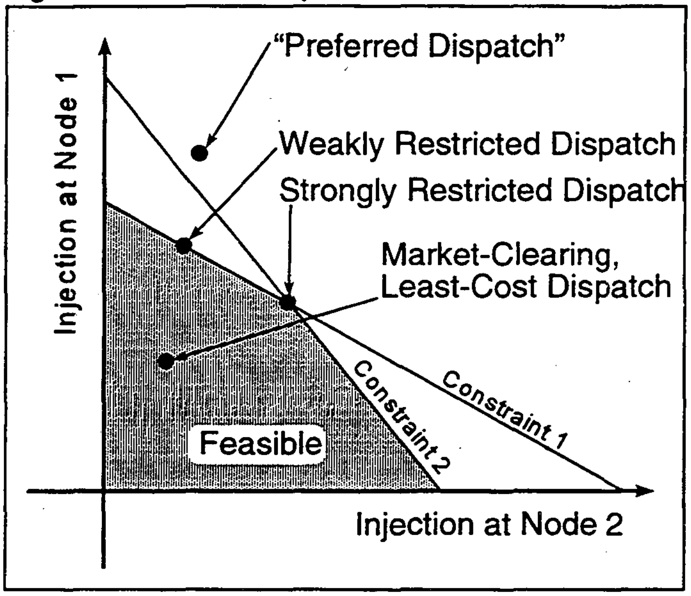
If traders have no market power, they will bid their marginal cost (or, in the case of load, their marginal benefit). This is because changing to a different bid would not change the MCP but it might cause their bid to be accepted at an unprofitable MCP or to be rejected at a profitable MCP. Consequently all unaccepted Incs will represent a marginal cost above the MCP, and all unaccepted Decs will represent a marginal cost below the MCP. Thus any additional kW traded would cost more to generate than would be saved by the decrementor. So the dispatch that corresponds to the market-clearing price is a least-cost dispatch. Conversely when the ISO is forbidden to make all profitable trades, the dispatch will not be least-cost.

The restriction as stated is that the ISO can only make adjustments to relieve congestion. But such a restriction must be restated with mathematical precision if it is to be implemented. To do this we need the help of a diagram showing grid constraints and feasible and infeasible dispatches. To avoid entanglement in the complexities of linear programing, I will not construct such a diagram from a specific network but will instead use one that qualitatively resembles one from a typical three-line, three-bus grid. Note that this results in a two dimensional graph because injections at any two nodes determine the flow at the third node. Thus every point in this diagram defines a dispatch, and every dispatch has a cost, so every point corresponds to a cost. At the origin of this graph where injections at Nodes 1 and 2 are zero, the implied injection at Node 3 must be very large in order to cover the load which we assume is fixed. Thus the cost minimum will not usually occur at the origin but at some point further out which describes the optimal mix of generation.

3.1 Example 2: The ISO's Redispatch Restriction

As shown in Figure 3, the "preferred dispatch," consisting of the sum of all submitted preferred schedules, is infeasible. In fact it violates two security constraints. Being linear constraints, these are represented by straight lines which define pairs of injections (at Nodes 1 and 2) that load a particular line to its security limit.

Figure 3. Restricted Dispatch



Because the preferred dispatch is not in the feasible set the ISO must redispatch. Let us assume that, as in our first example, that the least-cost dispatch is well inside the feasible set as shown. In this case there is meaning to the restriction that the ISO must not redispatch beyond the point at which congestion is relieved.

The Weak Interpretation

Now, for the ISO to honor the redispatch restriction, it must stop its redispatch when it comes to the boundary of the feasible set, otherwise it has done more than relieve congestion. But this leaves it with many choices because there are many points on that boundary and the ISO could reach most of them. This is where the injunction to minimize cost is helpful. Of all the points on the feasible boundary, only one will be least cost. In this case it is the point labeled as the "weakly restricted dispatch" in Figure 3. So the most literal interpretation of the restriction would bring the ISO to this point.

But is this what the Applicants had in mind? There are two cases to consider: (1) the least-cost dispatch is interior to the feasible set, or (2) the least-cost dispatch is on the boundary of the feasible set. In the second case, this interpretation places no restriction whatsoever on the ISO. It simply dispatches to least-cost exactly as it would under nodal pricing and, because some lines are still borderline-congested, it can claim to have "ceased making adjustments when congestion was relieved." Thus whenever the least-cost dispatch leaves at least one line at its security limit, the weak interpretation of this restriction has absolutely no effect.

In Case 1, the restriction is binding, but its effect will typically be very small. The ISO redispatches to a point where only one constraint remains binding, all the others having been more than satisfied. For instance, in our example, Constraint 2 has been more than met; the ISO went well beyond relieving congestion on the line corresponding to Constraint 2. One would expect that in a large network where many lines may have their constraints violated by the uncoordinated SC schedules, there will be at least one constraint on a rather small and inconsequential line.[3] According to our present interpretation, the ISO is free to go well beyond relieving congestion on all other lines, provided it leaves that one small line just barely congested. Since the big money is riding on the large lines, this will very likely correspond to the least cost solution. In other words, the ISO will be free to go all the way to least-cost, nearly clearing the market, provided it leaves one tiny line in a remote part of the system at its security limit. Of course, there will be some cases where there is no small congested line that the ISO can take advantage of and then even the weak interpretation will have a significant impact.

Economically speaking, accepting the weak interpretation would be good news. It means the restriction would cause very little waste, and the problems engendered by a market that has not been cleared would be minimized. Unfortunately, this lack of impact probably indicates that this is not at all what the Applicants had in mind. We turn now to the more likely interpretation.

The Strong Interpretation

We must look for a stronger restriction than the restriction that only one line must remain at its limit. The only such interpretation that seems logical is the strongest version possible: that no over-loaded line should be brought below its security limit by the ISO redispatch. In other words the ISO must "cease making adjustments when congestion is relieved" on a line-by-line basis.

This is indeed a requirement with meaning. It will cause the ISO to do something very different from clearing the market with a least-cost dispatch. But it has its costs both in terms of system security and in terms of missed opportunities for profitable voluntary trades. In the current example, it corresponds to exactly one point on the feasibility frontier, the intersection of the two constraints. Remember that this is not the least-cost point on the feasibility boundary, that point was the "weakly restricted dispatch." In general, if there are N busses and k violated constraints in the "preferred dispatch", the ISO will have $N-k-1$ degrees of freedom within which to minimize cost. However, in the end the ISO will be required to leave k contingency constraints at their limiting values. This will not always be possible for reasons discussed in Appendix A.

3 Note this is made even more likely by the fact, demonstrated in the next section, that for intra-zone congestion SCs will make no attempt at congestion management.

The main points of this section are: (1) that the ISO restriction has not been fully specified; and (2) that when it is, it will either prove to be nearly meaningless or quite restrictive. In the latter case, which seems more likely, the ISO will be required to leave the system with as many lines at their security limits as it had overloaded constraints in the "preferred dispatch." This will sometimes keep the dispatch quite far from least-cost, and the market quite far from clear.

3.2 The Redispatch Restriction Is a Security Risk

The requirement that the ISO must "cease making adjustments when congestion is relieved" (ISO Application, C.a.17, C.a.20, and C.b.24) has one unintended physical side affect that deserves attention. It will cause lines to be operated exactly at their security limits in many cases when it would actually be cheaper to operate them in a less fully loaded and therefore safer manner. That operating near a security limit reduces grid reliability is amply demonstrated by the fact that the Pacific Intertie was derated to 3,660 MW from 4,880 MW because of the last network collapse in the West. The closer the grid is to its operating limits, the easier it is for chain reactions of line outages to start and to propagate. By requiring the ISO to stop its redispatch as soon as it finds an operating point at which the grid is just barely within security limits and to go no farther in relieving line loadings, the ISO Application inadvertently increases the risk of and likely severity of network collapse.

3.3 How Complex Is Limited Redispatch?

At the end of Section 3 we noted two popular arguments against a full-information solution. The first led to the redispatch restriction, and the second, complexity, lead to zonal pricing. Although the strongest rebuttal to the second argument is the extraordinary complexity of zonal pricing itself, the complexity of redispatch also deserves attention.

A major theme in the California debate over nodal pricing has been the feasibility of computing what is known as an optimal power flow (OPF). This refers to the least-cost dispatch and its associated power flows on the grid. Locational marginal costs which can be used as prices are necessarily generated as a byproduct of this calculation. Traditionally utilities have only computed standard power flows given the operation of the generators, state of the grid, and setting of phase shifters. These are all that are needed to check feasibility. The economic part of the problem, minimizing the cost of generation, was solved by operators using merit order lists and generation cost information on a more ad hoc basis. This resulted in an approximate solution to the OPF problem. The difficulty with using an OPF is that it requires a (computerized) search process in which a guess is made as to the best dispatch, then a standard power flow is computed and checked for feasibility. Then a second guess is made with the intention of either reducing constraint violations or generation costs. Although very sophisticated procedures have been developed for making this sequence of guesses, the process of finding an OPF can still require that hundreds of standard power flows be computed.

Two other worries concerning OPFs have surfaced during this debate: an old one that the search process may sometimes and perhaps often converge to the wrong answer; and a new one. The old one seems to have been resolved by the new generation of programs pioneered by Alex Papalexopoulos at PG&E, but it may reemerge if OPFs are extended to the unit-commitment problem (Oren et al., 1996). The new one, which has gained attention in the recent debate, is that the prices associated with the OPF are in some way "too sensitive" to computational details and thus "not meaningful." Because the cause of this supposed price sensitivity has never been explained, not much can be said in response. But the one piece of evidence for this claim has been a report by Singh, Papalexopoulos and Hou (1996) showing that prices vary dramatically from bus to bus throughout much of the Western grid.[4]

Now these twin arguments of computational complexity and price "sensitivity" have been used as the only two technical arguments in support of moving to a "simpler" zonal system. Having analyzed the ISO's dispatch problem, we are now in a position to evaluate these arguments. First note that as we reviewed the ISO's dispatch problem above we did not encounter any reference to zones. In fact the only way in which zones play a role in this calculation is through the requirement that the ISO clearly distinguish between redispatch that is done for the purpose of relieving inter-zonal congestion and redispatch that is attributable to intra-zonal congestion. This cannot make life simpler for the ISO. In fact it may be a requirement that cannot be met.

The only reason the WEPEX ISO's redispatch problem differs from the standard OPF problem is because of the redispatch restriction. Putting this additional constraint on the ISO can actually simplify the problem slightly because it replaces inequality constraints (make sure a line is not overloaded) with equality constraints (make sure a line is exactly fully loaded). This reduces the problem's degrees of freedom. But this simplification could only be significant when a large percentage of the lines are congested, a situation that is extremely risky but fortunately almost never arises. There are also circumstances, discussed in appendix A, which make it impossible to find a feasible dispatch. In such cases the ISO must make an arbitrary decision which would be susceptible to dispute.

The bottom line is that zonal pricing has in no way simplified the OPF problem and the ISO's redispatch restriction only simplifies the problem in proportion to the number of lines that are

4 It should be noted that PG&E has put forth in public two rationales for favoring zones: (1) prices vary erratically (much more than necessary to split zones) from node to node; and (2) there is essentially no congestion within large zones. But both cannot be true because price variation is simply the result of congestion.

congested. Also the supposed nodal price "instability" problem, if it exists, cannot possibly be mitigated with regard to generators by either zonal pricing or the redispatch restriction.[5]

The Role of the Real-Time Market

The rules that have been analyzed so far apply to the forward markets but not to the real-time market which is unfortunately largely ignored by the Applications. But the ISO Application does tell us (C,c,29) that "The ex post price of imbalance power in each hour will be based on the marginal price of imbalance energy ... used by the ISO to balance the grid during that hour."

Voluntary, Real-Time Redispatch Could Improve Efficiency

From this we may surmise that if the expensive generators simply fail to meet their obligations, they will not be penalized beyond having to buy replacement power at the market price. Consequently a least-cost dispatch could be achieved in the real-time market through optimal non-cooperation.

For example, if the \$30/MWh generator, G0, in Example 1 simply refuses to generate, the ISO will be forced to buy imbalance energy which would probably be available from G1 for \$20/MWh. Certainly G0 will be motivated to implement this strategy because it will make \$10/MWh for doing nothing, while if it generates it makes nothing (marginal cost equals price).

Improvements of this type will be limited by the real-time nature of this market. First, generators will encounter unit commitment problems unless they can accurately predict the real-time market price well in advance, and second, it will simply be very difficult for traders to solve the optimal power flow problem in real time without the benefit of the ISO's coordinating function.

Gaming

Improvement in dispatch will not undo wealth transfers caused by the ISO's incorrect dispatch. For instance the load in Example 1 that was charged \$30/MWh instead of \$20/MWh will still have to pay \$30/MWh. The new dispatch will simply transfer some of this wealth from the ISO to G0. This means that G0 can expect to turn a handsome profit in the real-time market because of flaws in the forward markets. This will motivate G0 to accentuate those flaws if possible. It has also been noticed belatedly by many WEPEX participants that prices in the forward markets will differ from those in the real-time markets because of the inclusion of start-up and no-load costs in the forward bids and forward pricing. These will not affect the real-time price. It is believed that arbitrating these markets will provide yet more opportunities for gaming.

5 Of course loads do pay zone-average prices, so they could be shielded from meaningless price fluctuations if these exist.

4 Summary: Pricing Congestion with the Zonal System

Summary:

Example 3 shows that a combination of diversity among SCs and the redispatch restriction imposed upon the ISO will sometimes cause inter-zonal congestion to be negatively priced. This rewards those who add to congestion. Surprisingly this makes sense given the WEPEX redispatch restriction because additional congestion gives the ISO greater scope for redispatch, thereby allowing it to clear the market more completely.

Example 4 shows that load-price averaging over zones will make it beneficial for SCs not to account for intra-zonal congestion in their dispatches. To do so would disadvantage their customers. Unfortunately the PX is required to, so its customers are put at a disadvantage.

The Applicants have proposed two quite different schemes for dealing with congestion, one for inter-zonal and one for intra-zonal congestion. We will discuss inter-zonal congestion first because it is supposed to be far more important and common than intra-zonal congestion. This type of congestion must be dealt with three times for each hour: in the day-ahead market, in the hour-ahead market; and in the real-time market. The day-ahead market determines the ISO's congestion charge, and so it is the market of primary concern to us.

4.1 Pricing Inter-Zone Congestion

We have already discussed how the ISO redispatches to relieve congestion; it is now time to examine the pricing ramifications of congestion. The WEPEX rules for pricing inter-zonal congestion work as follows:

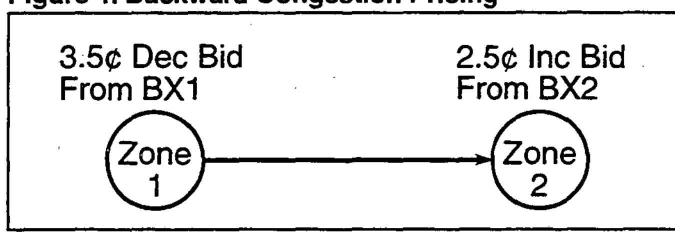
1) The PX and BXs submit preferred dispatch schedules and Inc/Dec bids of the included generators. (Inc/Decs are optional for BX generators.)
2) Assuming there is congestion, the ISO "determines an advisory redispatch which would eliminate the potential congestion at the least cost." (ISO Application, C,a,17)
3) The ISO provides the SCs with the "transmission congestion prices which would be charged to all power flowing across the congested interfaces."
4) After repeating steps (1) and (2) (ISO Application, C,a,19), "the ISO will again ... develop a final redispatch which would eliminate the potential congestion at the least cost. This would result in congestion prices ..."

Section 5.4.2.2.1.2 of the ISO Application states that "The congestion cost between two zones is defined as the difference in the marginal, or market-clearing price for energy in the two zones." The PX Application (6.4.2) provides a little more help by defining the marginal generators as "those units which are the sources of the last increments of generation." We will use this definition in the next example.

4.2 Example 3: Backwards Pricing of Inter-Zone Congestion

We have seen that the ISO is enjoined from clearing the market, and we have studied that process a little in the context of the network as a whole. We will now examine how this lack of market clearing will be distributed over zones and its consequences for the determination of inter-zonal congestion prices.

Figure 4. Backward Congestion Pricing



The basic insight we will gain is this. Because a market that has not cleared does not have a single market clearing price but instead has a range of available offer prices on both sides of the market, the ISO will have its pick of prices even at the so-called margin. So in the zone where it needs to redispatch up (Zone 2) it will look for and find low prices, and in the zone where it needs to dispatch down (Zone 1), it will look for and find high prices. If it does not have to redispatch too much generation and so does not run out of these bargain prices, then its marginal incremental price will be lower than its marginal decremental price. A little thought shows that this is exactly the opposite of what is needed for standard congestion pricing. We now demonstrate this with an example.

Assumptions for Example 3:

1) There are three scheduling coordinators: the PX, BX1, and BX2.
2) There is no congestion in the individual schedules submitted to the ISO.
3) All three SCs have cleared their markets.
4) The clearing prices are: BX1 = 3.5¢, PX = 3¢, BX2 = 2.5¢.
5) The ISO finds that the combined schedules cause congestion from Zone 1 to 2.

Now there is some controversy around one assumption of this example. That is the assumption of a single market clearing price at each individual SC. Some believe that the marginal Inc would be greater than the marginal Dec at each SC, and some believe it would be the reverse.[6] This does not matter at all! Anyone who is worried by assumption 3 may

6 BXs can use whatever system pleases them; it is quite likely that some BXs will choose to find the least-cost pattern of generation for their loads. To do this they will decrement any generator whose Dec is higher than some Inc, and they will increment the generator with the cheaper Inc. This will result in the lowest remaining Inc being just a hair below the highest remaining Dec.

reconstruct this example with his or her choice of marginal Incs and Decs, and so long as those of BX1 are both greater than the PX's bids and those of BX2 are both less than the PX's bids, nothing will change.

The first job of the ISO is to determine a new feasible dispatch "which would eliminate the potential congestion at the least cost." Because the congestion prevents sufficient power from reaching Zone 2, the ISO must dispatch additional generation there. Because all three SCs are present in both zones, the least-cost generator that the ISO can find in Zone 2 will be a BX2 generator with a Inc of about 2.5¢. The ISO will dispatch this generator. Similarly in Zone 1, the ISO will curtail a BX1 generator with a Dec of about 3.5¢.[7]

The second job of the ISO is to determine a congestion price. As noted at the end of the last section "The congestion cost between two zones is defined as the difference in the marginal, or market clearing, price for energy in the two zones." Because the ISO is not allowed to find the market clearing price, we must assume that it will use the "marginal" price in the sense defined by the PX Application, 6.4.2, i.e., that the "marginal generators are those units which are the sources of the last increments of generation." With this definition, the Zone-1 price would be 3.5¢, the Zone-2 price would be 2.5¢ and the price of congestion would be negative 1¢/kWh. The congestion cost is negative because the cost in the upstream market is always subtracted from the cost in the downstream market.[8] Under nodal pricing this would always give a positive value in a congested one-line network because congestion is always a cost and never a benefit. How is it that the zonal procedure gives this inverted result?

The problem with the ISO's procedure is again simply that it is prevented from clearing the market. Because the ISO cannot redispatch past the point of relieving congestion, it cannot clear the market in either zone. If it did clear the market, either it would relieve the congestion and the prices would converge, or the market-clearing price at the upstream end would be lower than at the downstream end indicating a positive charge for congestion. With the market failing to clear in both zones, the ISO is be able to find generators offering power both above and below the market-clearing price. This is why the ISO is able to find, and therefore must use, a high-priced generator at the downstream end and a low-priced generator at the upstream end. And this is why the congestion charge is likely to be negative.

These bids are entirely voluntary and therefore must reflect profitable opportunities for the parties involved. That the ISO is forbidden to make these matches seems entirely perverse and protects no one. While preventing valuable voluntary trades is wasteful, the operational

7 Clearly the BXs might not have generators available in the required zones with exactly these Incs and Decs. But if they both represent a large number of generators it is likely that they will have ones quite close to these values. In any case it is entirely possible that they would, and we will assume so for purposes of this example.

8 This rule arises from the fact that in a nodal spot market one transmits power from A to B by selling at A and buying at B. If the price at A is lower, this transaction is costly which is equivalent to paying a positive congestion charge.

problem with preventing the ISO from clearing the market is that it cannot determine a satisfactory price for congestion.

4.2.1 *Round Two of the Day-Ahead Auction*

Because the information flow between PX and ISO has been deliberately curtailed, the Applicants have designed a two-round auction. In some circumstances this does allow a partial correction of mistakes that are made in the initial round because the ISO reports tentative congestion results to the PX. This gives the PX, and other SCs, some additional market information.

This second auction round would affect our example as follows. The PX would be encouraged to ship more power from Zone 1 to Zone 2 because the ISO is paying for increased congestion on this path (the congestion charge was negative). This will, of course, force the ISO to engage in even more redispatching on the second round. But contrary to the standard intuition, this redispatch *makes money for the ISO* while improving efficiency. Because the ISO is only allowed to redispatch to the point at which congestion is relieved, increased congestion allows it to go farther towards clearing the market. This is a strange world in which congestion is actually beneficial because it allows the ISO to more fully overcome the Applicants' prohibition against clearing the market.

4.2.2 *A Quandary for the PX*

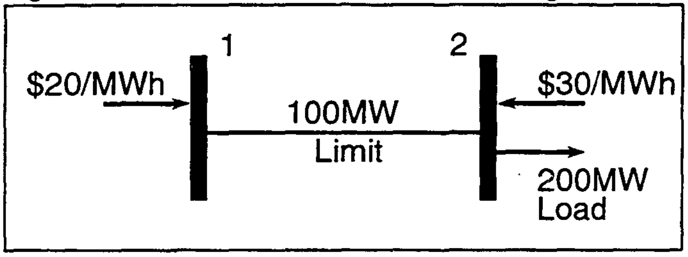
After the ISO prices congestion, the PX must price energy in the two zones. We now investigate how this would be done. The PX Application (6.4.2) defines the marginal generator as being selected from the limited set of generators included in the "preferred schedule." But the preferred schedule is a schedule of only PX generators and so does not include any BX generators.

Thus, in our example, the redispatched generators would not be included as possible marginal generators. Because of this the PX's marginal generators would be the same after the ISO manages congestion as before, and because they were not redispatched by the ISO, they will have the same marginal costs as before. In this case the PX Application (6.4.3) implies that the PX generation price and the PX load price will both be unaffected by the congestion.

This is only a small problem in this example because the congestion charge is negative, so the PX is paid by the ISO for the congestion. But this is still a problematic outcome because the PX is making money on the dispatch, and no provision for the use of these funds has been made.

4.3 Example 4: SCs Will Not Manage Intra-Zone Congestion

Figure 5. SCs Will Not Clear Intra-Zone Congestion



Because the method for relieving intra-zonal congestion is admitted by the ISO Application to be suboptimal and to send incorrect signals to grid users, there is little use in giving examples of its suboptimal outcomes. The WEPEX plan is to keep these inefficiencies small by dividing zones whenever "the price differences that grid users would see would be ... sufficient to send useful price signals to grid users regarding location, consumption, and demand-side management." However two points are worth noting. First, the ISO suffers from the same prohibition on market clearing with respect to intra- as to inter-zone congestion, and therefore the same congestion cost inversion will often occur. Second, the averaging of congestion costs over all loads will prevent the SCs from making any attempt to schedule around intra-zone congestion. We now explore this second point.

4.3.1 *SCs Will Not Manage Intra-Zone Congestion*

Any given SC is responsible for only a fraction, F, of the load in a given zone, and according to the ISO Application it will bear only that fraction of the zone's congestion costs assigned by the ISO. Consequently if the SC recognizes that a certain line within a zone will be congested and considers whether it should redispatch to relieve that congestion it will discover that it is always cheaper to leave it to the ISO. If it spends $X relieving congestion, this cost will be fully born by its load; while if it leaves the problem to the ISO, the ISO will have to spend $X (or possibly less) and only the fraction F of $X will be charged to the SC. Consider the following simple example.

Say the intra-zone line from Bus 1 to Bus 2 is used only by a certain BX1 and has a security limit of 100 MW. As shown in Figure 5, there is a cheap $20/MWh generator at Bus 1 and a more expensive generator, $30/MWh, next to a 200 MW load at Node 2. BX1 has a choice. It can dispatch 200 MW from Bus 1 and turn the problem over to the ISO or it can solve the congestion problem by dispatching 100 MW from each of its generators.

If it does its own re-dispatch it will pay $5,000/hr for 200 MW and pass this cost on to the load. If it turns the problem over to the ISO, the ISO will reduce the dispatch of Generator 1 by 100 MW and charge BX1 $2,000/hr, and it will dispatch Generator 2 at 100 MW and pay BX1 $3,000/hr. Now, BXs define their own internal procedures but one very likely procedure would be for BX1 to pass these charges and payment through to the relevant generators. This would result in exactly the same payments to the generators and the same dispatch as if BX1 had solved the congestion problem itself.

The only difference is that the ISO has now made a net contribution of $1,000/hr. So BX1 is, as a whole, much better off. But where does this $1,000/hr come from? It is collected proportionally from all of the loads in this zone. Thus if BX1 represents 25% of the zone's load, it will have to pay $250/hr, and will pass this cost on to its load. With the ISO handling redispatch, BX1 pays $4,000/hr to Generator 1 for its preferred (uncongested) dispatch, plus it pays the ISO's congestion charge of $250/hr for a total cost of $4,250/hr. This cost will be passed on to the load which will save $750/hr over the BX1 redispatch method.

This example invites several observations. First, because BXs are free to adopt the strategy of letting the ISO spread around their congestion costs, the PX must be allowed to do the same or it will be at a disadvantage. Thus the PX must not be required to take into account known transmission constraints when making its preferred schedule. Second, note that the load in this example was subsidized by all other loads in the zone, and that the smaller the SC, the larger the subsidy. Third, note that the rule that limits the ISO to consider only bids of generators in the preferred schedule makes it necessary for BX1 to dispatch a very small amount of Generator 2 in order to get its bids considered by the ISO.

5 Defining Zone Boundaries and Zones

Summary:

If "interfaces between zones" are "defined by paths or networks for which the expected demand may often exceed the path or network ratings," then the definition of zones is self contradictory. In particular, Example 5A shows that, if congested paths are defined to be zonal interfaces, then some zonal interfaces must necessarily be uncongested paths.

This contradiction is the result of WEPEX ignoring the problems of loop flow in its definition of zones. Example 5A is also used to develop a definition of zones based on nodal prices that resolves this contradiction, at least for the simplest examples of loop flow.

Example 5B further demonstrates the need for price-based zoning.

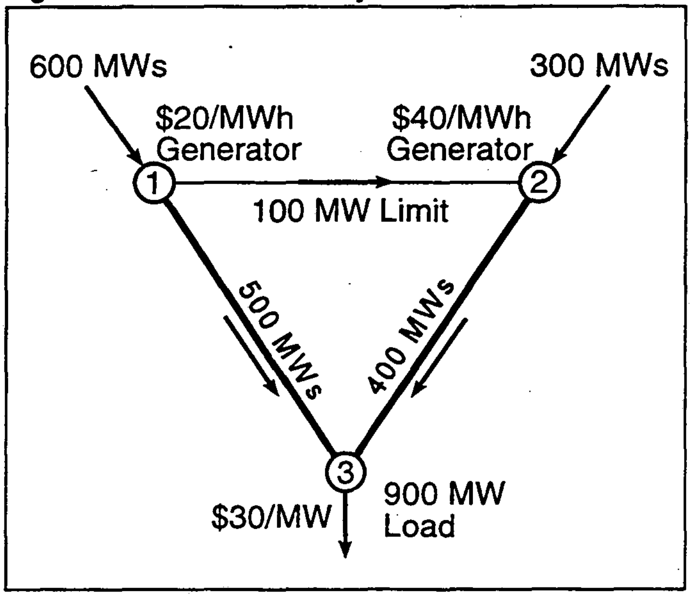
To apply the definition of a zonal interface, it is necessary to specify how congested a line must be before it becomes an interface. This is done by WEPEX's "5%" rule which we will examine before we attempt to understand the interface definition.

The WEPEX rule for splitting a zone is to split it if the cost of congestion over the course of a year is at least 5% of the TO's access charge times the capacity of the rated path. Let us consider these numbers. The average access charge will be about \$20/kW-yr[9] (ISO Application, 5.4.2.1.8), which is 0.23¢/kWh, or \$2.3/MWh for a line at full capacity all the time. If we assume the line in question is on average loaded to 1/2 capacity, because we do not expect congestion all the time, then the charge per kWh of transmitted power is about \$4.6/MWh. Five percent of this is 23¢/MWh (notice this is MWh not kWh). Since the cost of generation is in the neighborhood of \$23/MWh or more, we are talking about a cost of 1% or less. So the rule for splitting zones is roughly that a new zone should be created if the cost of congestion is 1% of the cost of generating the power affected by the congestion.

Note that a 1% cost of congestion means that on average the price is wrong by 1%, but it does not indicate a social (dead-weight) loss of 1%. The dead-weight loss is generally proportional to the square of the pricing error, so for a 1% pricing error one would typically

[9] I will interpret this as a cost-per-year per kW of peak demand. If this value represents a cost per kWh transmitted, then I have underestimated the strictness of the 5% rule.

Figure 6. The Zone Boundary Problem



expect a much smaller dead-weight loss.[10] Although congestion costs may be incorrectly measured and partially hidden by gaming as suggested by the CEC,[11] it seems unlikely that these effects would be large enough to cause any serious economic inefficiency by delaying the creation of new zones.[12] We now turn to the problem of defining boundaries when zones are created.

5.1 Example 5A: Contradictions in the Definition of Zones

The ISO Application defines a zone to be a portion of the ISO grid within which congestion is expected to be small. It then specifies the 5% rule for congestion cost on a congested path. It further states that "Interfaces between zones, on the other hand, are defined by paths or networks for which the expected demand may often exceed the path or network ratings" (ISO Application, 5.4.2.2.1.1). While nothing appears to be amiss with such a definition, close inspection soon shows that it is based on naive notions of power flow. This definition will be seen to be self contradictory in the presence of loop flows.

10 As an example of the relationship between mispricing and dead-weight loss, consider the case of a market where both supply and demand elasticity are 1. In this case, for small changes in price the loss is approximately the square of the price deviation. Thus if price is set 10% too high, then the deadweight loss will be about 1% of cost. From this we see that the most damage would be done in a market where congestion was occasional but severe. Following the above example, we see that if the lines were congested for only 5% of the power flow which is 2.5% of the year (about 220 hours), then the price during that time period would have to be about 20% higher if congestion costs are to amount to 1% of total cost. (This calculation employs the fact that the line is typically half loaded but when congested it is fully loaded.) Now let us assume that due to cost averaging the price seen by the consumer is wrong by 20% for 5% of the power flow. This might induce a $(20\%)^2$ or 4% dead-weight loss, for 5% of the power flow. Four percent of 5% is 0.2%. Thus the deadweight loss from failing to split a zone before the WEPEX criterion is reached is probably well under 1% of the cost of the power flow on the congested path.

11 These comments, originally made to the CPUC, were filed as an appendix to SDG&E's June 28th *Explanatory Statement* to FERC, which it has now requested to withdraw.

12 There are two possible exceptions to this conclusion. First, because the WEPEX criterion considers only the total congestion cost and ignores its time profile, it would be possible to have very short-duration but very expensive congestion that caused significant inefficiency without exceeding the 5% total limit. Second, ignoring "non-normal conditions" (ISO Application, 5.4.2.2.1.1) may ignore conditions that can and should be influenced by a price signal.

The problem is most easily explained by way of an example, and this time the example will need to encompass "loop flow," so we will need to abandon our single line network. The grid in this example consists of three lines, one of which has a contingency limit low enough to bind, and two lines that have, for all practical purposes, unlimited capacity. The load presents a fixed demand of 900 MW which would be most cheaply supplied by Generator 1, whose cost is a flat $20/MWh. But this is impossible because 300 MW, one third of the power, would flow the long way from 1 to 3, and this would exceed the capacity limit of 100 MW on Line 1-2.

The least-cost dispatch requires that Generator 2 be dispatched to produce a counter flow on Line 1-2 sufficient to cancel all but 100 MW of the flow from Generator 1. A 300 MW flow from 2 to 3 will produce a 100 MW counter flow on Line 1-2. This will cancel all but 100 MW of a 200 flow on Line 1-2 produced by Generator 1's dispatch of 600 MW to the load at 3. This dispatch uses the maximum generation that can be accepted from Generator 1 and is therefore least-cost. If Load 3 were to demand 1 more MW, the cheapest way to obtain it would be to dispatch both Generators 1 and 2 up by 0.5 MW for a total cost of $30/hr. These two resulting flows on Line 1-2 would just cancel, thereby keeping the line within its limit. Thus the marginal cost of power at Node 3 is $30/MWh.

Now we have an example of a network in which Path 1-2 must be an "interface between zones" but Paths 1-2 and 1-3 must not. This means that Busses 1 and 2 are in different zones, but Busses 1 and 3 are in the same zone and Busses 3 and 2 are in the same zone. This is simply impossible.

This contradiction arises because loop flow has been ignored in WEPEX's design of zones. This is made clear by the third sentence in Section 5.4.2.2.1.2 of the ISO Application, "Determining Congestion Management Costs Between Zones" (a section with which SDG&E explicitly "does not join"). We read as follows:

> "With congestion, however, the market clearing prices will not be the same -- prices in the net export zone(s) would be lower than they would have been absent the congestion and prices in the net import zone(s) would be higher, ..."

This statement is true for radial networks which have no possibility of loop flow. But it is contradicted by Example 5A, one of the simplest possible examples of loop flow. This is seen by noting that without congestion, Bus 2 could be supplied with $20/MWh energy from Bus 1 and that Bus 2 is a net export zone. So we see that prices in this net-export zone are much higher then they would have been absent congestion. The flow from Bus 2 to Bus 3 is the famous "uphill" flow which is discussed at great length by Wu et al. (1994).

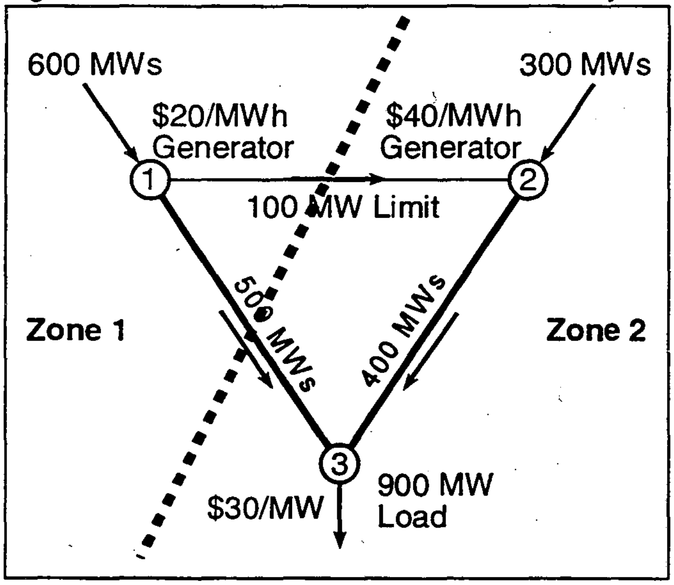
This flaw in the WEPEX definition does not indicate that zones cannot be well defined, it simply indicates that they have not been. As we will now see there is a strong possibility that a useful and consistent definition of zones can be based on nodal prices instead of on the physical congestion of lines or paths. There are now three possible solutions to the problem of defining zones in example 5A: divide the busses between zones this way [1,3 | 2], or this way [1 | 3,2], or this way [1 | 2 | 3].

Before evaluating these three possibilities we turn to nodal pricing to find the optimal level of congestion costs that should be imposed on the load. We have already computed the nodal spot prices, and they are displayed in Figure 6. Congestion prices are given by nodal-spot-price differences according to the following simple formula:

$$\begin{aligned}\text{Congestion Charge} &= \text{power flow} \times (\text{destination price} - \text{origin price}) \\ &= Q \times (P_D - P_O)\end{aligned} \tag{1}$$

So the congestion cost from 1 to 3 is (\$30−\$20)×600 = \$6,000/hr. A similar calculation give the congestion charge from 2 to 3 as −\$3,000/hr, so the total congestion charge imposed by the ISO is \$3,000/hr. It must be remembered that this is the charge collected by a nodal-pricing ISO. It does not include the extra cost of out-of-merit generation which should also be charged to the load indirectly through its nodal price. This cost is an additional (\$40−\$20)×300 = \$6,000/hr.

Figure 7. A Possible but Incorrect Zone Boundary



Now let us look at the WEPEX system of congestion management. Beginning with a "preferred dispatch" that does not take account of congestion, we find the load paying \$20/MWh for 900 MW from generator 1. The ISO then redispatches the system by decrementing Generator 1 by 300 MW and incrementing Generator 2 by 300 MW. This costs the ISO \$12,000/hr for Generator 2 and earns the ISO \$6000/hr from Generator 1. The SCs representing those generators would be billed and paid by the ISO, and they would in turn pass on these bills and payments to the generators. Because the ISO at first ends up short by \$6,000/hr, the SCs will also end up short \$6,000/hr. They will in turn charge the load for this redispatch expense. This matches the out-of-merit part of the nodal spot price. Now to complete the match, we need a congestion charge of \$3,000/hr, and we will look for a zone definition that provides this.

Bus 3 Defined to Be in Zone 1

The simplest possibility to reject is that of putting Bus 3 into Zone 1. In this case the net flow is 300 MW from Zone 2 to Zone 1, so the origin price is \$40 and the destination price is \$20. This would result in a congestion charge of 300×(\$20−\$40) = −\$6,000/hr. This is both the wrong sign and wrong magnitude.

Bus 3 Defined to Be in Zone 2

Putting Load 3 into Zone 2, as shown in Figure 7, gives a more reasonable answer. The net flow is 600 MW from Zone 1 to Zone 2, so the congestion charge is 600×(\$40−\$20) = \$12,000/hr. At least this charge has the right sign, although it is four times too much.

Creating Three Separate Zones

Clearly the only remaining hope of assigning the correct congestion charge is to make each bus its own zone. This time there are two net flows. From Zone 1 to Zone 3 there is a net flow of 600 MW with a corresponding congestion charge of \$6,000/hr, and from Zone 2 to Zone 3 there is a net flow of 300 MW with a corresponding congestion charge of minus \$3,000/hr. The total congestion charge is then \$3,000/hr, which is just what is needed.

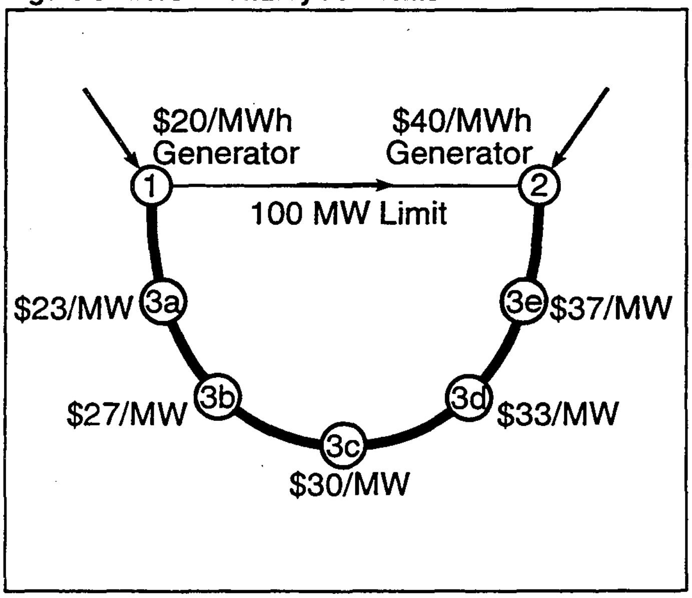
We have now established the following:

1. The WEPEX definition of zonal interfaces and thus of zones is self contradictory.
2. A zonal interface should be redefined as a path that exhibits a spot price difference between its ends due to congestion somewhere in the network.
3. With this definition, zonal congestion pricing can work perfectly in at least some networks with loop flow.

5.2 Example 5B: Many Zones from One Congested Path

The need for a price-based definition of zones is further illustrated in the following somewhat more complex example. This example simply adds a series of busses to the previous example.
All of these busses lie along the uncongested path from Bus 1 to Bus 2, and all of them can be thought of as load-only busses, though they could have a small amount of

Figure 8. More Boundary Problems



generation without changing anything of substance. The marginal cost of delivering power to each of these nodes changes gradually as we move from Node 1 to Node 2.

How should zones be assigned in this case? With only one line where "demand may often exceed the path or network ratings," there is only one interface between zones according to the WEPEX definition. This would imply that there can be only two zones. This would force the assignment of all 3-type busses to either Zone 1 or Zone 2. This can be done by placing a zone boundary through the uncongested path. Although there is no particular rationale for locating such a boundary it could easily be done arbitrarily, but this leads to a particularly dispute-prone situation. Say the boundary runs between 3c and 3d. The load at 3c will not have to pay any congestion charge while the load at 3d will have to pay half of it. Because they are so close electrically and have very similar optimal nodal prices, this would be viewed as unfair, arbitrary and capricious by the load at 3d. Litigation could easily be the next step, and the ISO will find it impossible to give a convincing rationale for its choice of zone boundary.

Because this argument applies to any choice of boundary for a two-zone division of these busses, we must look for a multi-zone solution. Given the indicated price differentials, every single line would probably pass the 5% test and should therefore be declared a zonal interface. This would make every bus a zone. If, however, two busses had a sufficiently small price differential they would be lumped into a single zone.

To conclude, it appears that a zonal system could be implemented without serious economic drawbacks, but the system should be based on congestion as determined by the marginal costs derived from a least-cost dispatch, i.e. from optimal nodal prices. If this is done, and if the rule for new zone creation is stringent, as the WEPEX rule appears to be, then a zonal system might approximate a nodal-pricing system quite closely. But it should be remembered that this new price-based definition of zones has only been tested on a single example. Nonetheless, the new definition appears promising.

6 Transmission Congestion Contracts

Summary:

Example 6 demonstrates that the way WEPEX appears to define TCCs leads to improper incentives for generation.

The ISO Application (5.4.2.4) leaves the introduction of TCCs "contingent on market demand" and specifies that they will be offered "until independent marketers develop to meet demand for such price-hedging instruments." There is some uncertainty about whether TCCs will in fact be introduced in California. But because they would be useful and may be introduced, we will review the way they are defined by the WEPEX Applications, and the implications of this definition.

The WEPEX Applications appear to include the following specifications:

1. The purpose of TCCs is to meet the needs of new direct-access customers.
2. Existing customers should obtain full market value from the sell-off of TCCs.
3. TCCs will hedge inter-zonal congestion and be associated with "inter-zonal interfaces."
4. If new zones are defined, new TCCs will be offered.
5. "The ISO will define and administer TCCs which will be allocated to the parties ... paying the costs of new inter-zonal transmission facilities."
6. "The holders of these TCCs may use up to the amount of the incremental inter-zonal capacity associated with the new facilities, without paying the usage charge ..."

TCCs are designed to accomplish the goals of (1) hedging new long-term generation contracts, and (2) encouraging efficient investment in the grid. An often ignored benefit TCCs is that they accomplish these goals without any adverse affect on the incentives produced by nodal pricing. We will show the WEPEX definition interferes with this property.

6.1 Hedging and Investment

TCCs are designed to encourage market entry by facilitating the financing of new IPPs. TCCs serve this function by allowing a new project to write a long-term contract at a fixed price without risk to either party. Without this hedge, small firms will be less competitive, and entry into the generation market will be diminished. The details of this effect are discussed more fully in Appendix A.

One difficulty with using WEPEX TCCs as a hedge is that these TCCs will not cover intra-zonal congestion. This may or may not be a substantial problem depending on the flexibility and speed with which new zones are established when needed. It is likely to be a perceptual problem for IPP projects that are trying to get project financing because banks will find it hard to estimate the extent to which unhedged intra-zonal congestion might affect the cost of transmission.

The benefits of TCCs to new direct-access customers, mentioned by WEPEX, is much less clear. TCCs can certainly hedge customer risk as well as they hedge supplier risk, but the burden of this risk on customers is very slight for two reasons. First, congestion costs are a very small part of most customers' portfolios. Second, congestion costs are not well correlated with the rest of a customer's portfolio, so this risk is largely diversified, and thus deserves a very small risk premium. Within the context of restructuring, this benefit of TCCs to customers is probably too small to deserve any attention at this time.

6.2 TCCs and Grid Investment

WEPEX envisions a combination of market-driven and regulatory procedures for determining grid-investment. If TCCs were allocated to investors according to the "feasibility rule," they would provide an environment that minimizes the reliance of economically beneficial projects on the backstop regulatory process. Without TCCs to protect an investor's right to the grid, market-driven investment would be significantly diminished. Similarly, without TCCs to screen out projects that are beneficial to the proposer but detrimental to other market participants, there will be a growing demand for regulatory review of the market implications of all proposed grid "expansions." This topic is covered in more detail in Appendix B, and by Bushnell and Stoft (1996a and 1996b).

Another difficulty that could affect both market entry and grid investment is attributable to the use of zones when defining TCCs. When zones are divided this may change the definition of an "inter-zonal interface" for which a TCC had previously been specified. This could have an unpredictable effect on the protection provided by a TCC to a grid investor. This problem could easily be overcome by specifying TCCs between pairs of busses, as was the original intent. These bus pairs could easily correspond quite closely to inter-zonal interfaces but, unlike zones, busses cannot be redefined out of existence, so there would be no problem with honoring old TCCs when zone boundaries are changed.

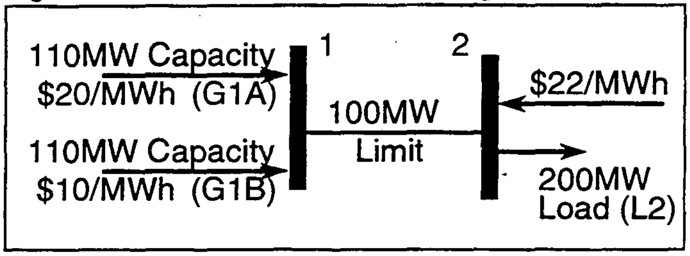
6.3 Example 6: Wrong Incentives from WEPEX TCCs

A dispatch incentive problem will be caused by Point 6 described above; i.e., that:

> "The holders of these TCCs may use up to the amount of the incremental inter-zonal capacity associated with the new facilities, without paying the usage charge ..." (ISO Application, 5.4.2.4)

This implies, though perhaps not intentionally, that if a TCC owner uses only half of the designated capacity of her TCC, then she will not be paid by the ISO for any congestion charge on the unused portion. This has in fact been a policy advocated by many WEPEX participants. Unfortunately this destroys the incentive properties of zonal pricing for the TCC owner. This can be seen in another two-bus example.

Figure 9. WEPEX TCCs Give the Wrong Incentive



We begin the example with only one generator, G1A, at Bus 1. This generator has a capacity of 110 MW and a constant marginal cost of $20/MWh. It owns a TCC for 100 MW from Bus 1 to Bus 2 and uses it to cover its full output. The price at Bus 1 will be $20/MWh. (It is best to imagine that G1A is actually a set of competitive generators.) G1A earns nothing on its sale of power, but earns a profit of $2/MWh on its TCC. Now, as shown in Figure 9, a second generator, G1B, enters at Bus 1 with a lower cost of $10/MWh. What will happen?

Correctly Specified TCCs

The price at Node 1 will be bid down to $10/MWh as G1B takes over the market. This will not hurt G1A because it is fully covered by its TCC. In fact, G1A will now make $12/MWh of flow on the line. But the crucial point is that G1A has stopped producing and G1B, a much cheaper generator has replaced it. Because the TCC payment rule has been correctly specified, this is also the least-cost dispatch.

WEPEX TCCs

But what if the TCCs had been misspecified to covering only actual use of the line? Then, if G1A relinquished its market share to G1B as it would under real TCCs, it would no longer make any profit. It would instead lose its $2/MWh hour profit. But, because of its TCC, it can afford to keep it's market share even if it has to bid $9/MWh to do so. Say that it does. Then it keeps its production level at $100 MW and loses $11/MWh on production and sale but makes $(22−9) = $13 MWh on its TCC for a net gain of $2/MWh. Thus the TCC has protected it from the new entrant but at the cost of producing a very inefficient dispatch. Clearly G1A would keep producing, and we would not achieve the least-cost dispatch. Thus if WEPEX means what it appears to mean, its TCCs will cause a misallocation of resources. There is no need for this. The altered definition provides no advantage; it is simply a mistake.

Can the Market for TCCs Correct This Incentive Problem?

It has been suggested that this problem with WEPEX TCCs only arises because the generator owns the TCC instead of the load owning it, and there is considerable truth to this notion.[13] Because the demand of loads is less elastic than the supply of generators, mispricing energy to loads causes much less damage than mispricing it to generators. Nonetheless, as long as some loads have a price elastic demand, examples similar to the present one can be constructed for loads, so the WEPEX definition will result in some misallocation of resources.

Probably more important than the dead-weight loss caused by the WEPEX definition is the complexity it adds to TCC ownership. Say generator G1A has built the line from Bus 1 to Bus 2 and accepted a 100 MW TCC from the ISO for doing so. As we have seen, this TCC will be less valuable than a standard TCC and will induce inefficient behavior if it is held by G1A after G1B enters the market. But these problems can be largely corrected by G1A selling its TCC to Load 2 and writing a contract for differences (CFD) with Load 2. This CFD should specify that L2 will pay G1A $\$(P_C - P_1) \times 100$ MW, where P_C is the contract price, and P_1 is the price at Bus 1. Because G1A will receive $P_1 \times 100$ MW for selling 100 MW at Bus 1, its total receipt will be $P_C \times 100$ MW, which means G1A is paid the contract price independent of the market price.[14] But the incentive property of this contract is seen by considering G1A's profit: $\pi = Q \times (P_1 - C) + (P_C - P_1) \times 100$, where Q is the quantity generated by G1A and C is its marginal cost. Clearly if the spot price P_1 falls below marginal cost, G1A will find it profitable to reduce output. This is the right incentive.

The conclusion of this analysis is that WEPEX TCCs will do only a little damage, provided loads exhibit a very small elasticity of demand and provided further that generators do not hold TCCs but instead sell them to their corresponding load and simultaneously sign an appropriate CFD. Thus the market for TCCs can correct largely, but not completely, the incentive problems introduced by the WEPEX redefinition of TCCs.

13 I would like to thank Gregory Basheda of U.S. DOE for pointing this out.

14 L2, on the other hand, must pay $P_C \times 100$ MW + $(P_2 - P_1) \times 100$ MW which means that L2 is affected by the difference in price between the two busses. For this reason L2 needs 100 MW of TCC to be fully hedged.

7 Losses

Summary:

Example 7 demonstrates that the WEPEX procedure for collecting losses will encourage excessive use of the transmission system and lead to significantly increased losses. Also, in order to compute the zero-loss-revenue prices required by WEPEX, it is necessary to compute optimal prices first, so the WEPEX procedure is more complex than optimal pricing.

The WEPEX treatment of losses is largely motivated by a concern that the ISO not "over-collect" for losses. This is tempered by an understanding that marginal-cost pricing leads to a least-cost dispatch. The difficulty is that because losses are nonlinear (quadratic) in power flow, marginal-cost pricing leads to "over-collection." The intent of the compromise is to scale back the marginal losses proportionally for all generators to the point where "over-collection" is just avoided. It is generally recognized that this will sacrifice some efficiency. The Applications do not explain why it is acceptable for the ISO to "over collect" for congestion but not for losses. Certainly the additional "over-collection" could be used to reduce the access charge just as the present "over-collection" is used.

This section will begin by explaining how losses are treated under marginal-cost pricing (least-cost dispatch) and why this treatment leads to "over collection." Then it will deduce the specific method of assigning losses that fits the WEPEX description and is most sensible. Then it will show that the WEPEX system, while unlikely to lead to serious mispricing, may well induce large and unnecessary flows that could overburden the transmission grid.

What is "Over Collection?"

Power flows interact. If one person can inject 100 MW at A and take out 90 MW at B, then two people cannot do this simultaneously because each increases the losses of the other.

Say that, as above, a 100 MW injection at A causes a 10% loss. Given this flow, if I want to transmit 9 kW to my partner at B, I must put in 10 kW at A. Unfortunately, measurement will prove that there is another 1 kW missing that I have not made up. WEPEX will collect this loss by increasing loss factors of all traders from 10% to 10.001%. Since this affects me, I will really have to make up 1,000.1 Watts which is my loss plus a tiny bit of the 1 kW loss I caused others.

Charging for marginal losses would require me to make up my 1 kW loss, plus pay an amount equal to the value of the 1 kW loss that I caused to others. This is what WEPEX terms "over collecting" for losses. (Note that others make up for their own losses including the part caused by my use of the line, so it is not possible for me to physically make up for the loss I cause others; I must pay for that damage financially.)

Throughout this section we will completely ignore congestion. This greatly simplifies the examples and the discussion without in any way impairing the validity of the points being made.

7.1 Loss Pricing Under Least-Cost Dispatch

The first point to be understood is that marginal cost pricing does in fact "over-collect" for losses in the sense that the ISO collects more than enough to buy the necessary replacement power. This can be demonstrated with a one line example and an explicit algebraic formula for losses. For our example's loss equation we choose:

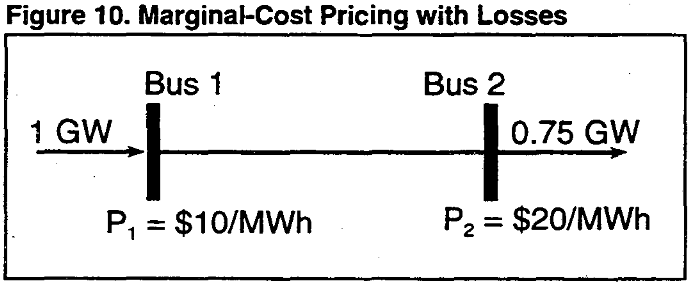
$$W_2 = W_1 - 0.25\,W_1^2, \quad \text{where power, W, is measured in GW} \tag{2}$$

Equation (2) gives the power delivered to Bus 2 (W_2) as a function of the power injected at Bus 1 (W_1), with power measured in GW. Because in equilibrium we know that the marginal cost of power delivered directly to Bus 2 will equal the marginal cost of power delivered directly to Bus 1 and transmitted to Bus 2, we can deduce the following equation relating prices at Bus 1 and Bus 2.

$$P_1 = \frac{dW_2}{dW_1} P_2 = (1-0.5\,W_1)\cdot P_2 \tag{3}$$

This equation tells us that if putting an extra kW in at Bus 1 yields only a half kW at Bus 2, then an extra kWh generated at Bus 1 is worth only half what an extra kW at Bus 2 is worth. More specifically, equation (3) tells us that for a 1 GW flow into Bus 1, the price at Bus 1 will be half that at Bus 2. Equation (2) tells us that 1 GW in at Bus 1 yields 0.75 GW out at Bus 2 for a loss of 0.25 GW. So if the price were, say, \$10/MWh at Bus 1, the ISO would pay that generator \$10,000 for a 1 GW flow and would sell the 0.75 GW at Bus 2 for \$20×750 = \$15,000. The ISO would net \$5,000/hr, even though losses have been covered. This "over collecting" for losses is what the WEPEX system seeks to avoid.

Figure 10. Marginal-Cost Pricing with Losses



Although this system looks odd to the uninitiated, the following points should be made in its favor. The price at each node does equal the cost of delivering the final (marginal) MW. By charging these prices, both buyers and sellers are motivated to behave in such a way that they could not both be made better off by any other trade. It is electricity that is being priced, not losses, and this is exactly the way a competitive market prices goods.

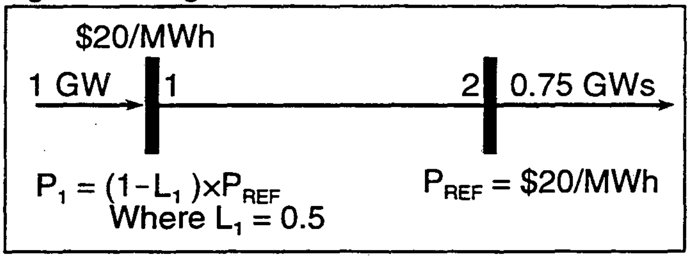
7.2 How Losses Are Treated by WEPEX

Section 5.4.2.3 of the ISO Application makes two crucial points when defining the treatment of losses. These are:

1. The ISO assigns total system losses to individual generators by computing "marginal loss factors."
2. The ISO sums each SC's generation losses and assigns this sum to the SC.

These clear points are accompanied by some hints about how the assignment process should work. First we are told that each generator will be assigned a marginal loss factor, and second we are told that this is "defined to be the marginal impact of that generator's output on total system transmission losses." Unfortunately this definition is not as precise as it sounds. We are not told where the generator's output leaves the system, and so we do not know precisely how to compute such a marginal impact.

Figure 11. Marginal Loss Factors



The name "marginal loss factor" is probably our best clue as to what the WEPEX authors had in mind. This is a term used in the decomposition of nodal prices, and it is this term to which SDG&E assumes that the ISO Application refers (see SDG&E Statement, Appendix D). We now investigate this loss factor and its possible uses.

What we are looking for is a method of using the "marginal loss factor" to assign physical losses to generators so that the assignments sum exactly to system losses. Note that this method replaces the use of prices with physical quantities.[15] Thus if there is a loss between Zones 1 and 2 (but no congestion), the ISO would assign the same zonal price to each but would assign different physical losses to each. Later we will see that this is *equivalent to a particular pricing system* but it has been neither conceptualized nor presented in those terms.

To understand the marginal-loss factor, consider an uncongested but lossy network. Marginal-cost pricing, as explained above, allows us to compute nodal prices which will differ from bus to bus because it is more costly to ship power to some locations than to others. These prices may be expressed by arbitrarily picking a reference bus, naming the price at that bus P_{REF}, and giving all other nodal prices relative to that reference price. Because it is the

[15] This is the WEPEX system by which the ISO assigns losses and not the PX system for allocating losses. The PX system does use prices.

deviations from the reference bus price that are caused by losses, the price at Node i is typically expressed as $P_i = (1-L_i)\times P_{REF}$, where L_i is termed the "marginal loss factor at Bus i."

Using loss factors, the prices in the last example would be expressed as shown in Figure 11. Having now defined the marginal loss factors, we need to specify how they would be used to assign physical losses. This assignment is based on an equivalence between assignments of physical loss and price reductions.

Consider the example in Figure 11 again. We could reinterpret Generator 1's loss factor of 0.5 to mean that 50% of any power it injects into the grid will be counted as *uncompensated* "makeup power" to cover losses and then set the price at Bus 1 to \$20/MWh. This would make no difference to Generator 1. It would still receive \$10 for every MWh injected. This is the physical loss equivalent of the nodal spot price. We formalize this concept for future reference;

The loss equivalence principle: Receiving price P and being required to count L% of output as uncompensated loss replacement is equivalent to receiving price $P\times(1-L\%)$ and having no responsibility for physical loss replacement.

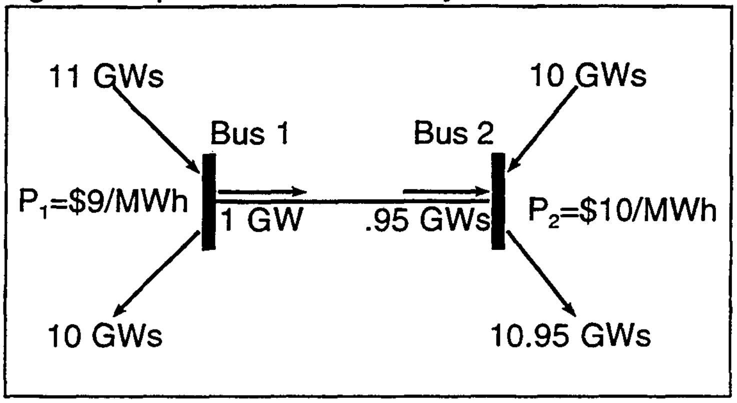
Notice that the physical system just described cannot be the system proposed by WEPEX because too much power would be injected. In the present example, if the traders submitted a balanced dispatch of 0.75 GW, the marginal loss factor would require the generator to supply an extra 0.75 GW to cover losses. This is physically too much. WEPEX wants the sum of a "balanced dispatch" plus loss replacement to result in a true power balance for the grid. This requires an adjustment of the marginal loss factors. In this case a loss factor of 0.25 would do the trick. The Bus-1 generator would then inject 1 GW, specify a balanced trade with the load of 0.75 GW, and use 25% of its injected power to cover losses.

As SDG&E has pointed out, there are several ways this adjustment could be made, but only one holds any real promise. This is to replace the set of optimal nodal prices $\{P_{REF}, P_i = (1+L_i)\times P_{REF}\}$ with $\{P^*, P_i = (1+\alpha L_i)\times P^*\}$ where α is chosen so that the ISO's net revenue collection would be zero, and P^* is chosen so that the load flow is in balance.[16] Note that it would appear that this system is ambiguous because we can choose any bus as the reference bus. Of course choosing a different bus results in different loss factors, L_i. But it can be shown algebraically that no matter what reference bus is chosen, the set $\{P^*, P_i = (1+\alpha L_i)\times P^*\}$, with α and P^* chosen as specified, is the same. Therefore this system of choosing a new set of zero-loss-revenue (ZLR) nodal prices is unique.

16 Actually both are chosen simultaneously to satisfy both conditions but it is easy to see that load balance can be adjusted with P^*, and revenue balance can be adjusted with α.

What can be said about these ZLR prices? First, if we rank busses in order by price, we get the same ranking as with optimal prices. Second, ZLR prices will probably always have less spread than optimal prices. Third, since these prices are not optimal, they will not result in a least-cost dispatch.

Figure 12. Optimal Prices In a Lossy Network



To apply these results to the WEPEX loss assignment system we need only show that ZLR pricing is equivalent. This needs elaboration because the WEPEX system does not make direct use of these prices but instead uses quantities. Here we refer back to our loss-equivalence principle which assures us that the ZLR prices are equivalent to physical loss factors of αL_i used as described above. This equivalent physical system will have the properties required by the WEPEX Application, the ISO will have no "over collection", and the power flows will balance.

We have now specified the exact system that should be used if WEPEX wants to achieve its stated goals with the least damage to economics and fairness. The WEPEX system is the physical loss-replacement system that is equivalent under "the loss equivalence principle" to ZLR pricing. It is now time to investigate its implications, and we will do so by examining ZLR pricing.

7.3 Example 7: WEPEX Loss Factors Increase Line Usage

The following basic example shows how the WEPEX loss-coverage system works and its effect on trading in the WEPEX market and use of network facilities.

Figure 12 shows an optimally dispatched, one-line, lossy network. The dispatch problem is specified by supply and demand functions at each bus and by the loss equation for the network. Demand is taken to be fixed at 10 GW and 10.95 GW as shown, while the other equations are as follows:

$$Q_1^S = 2P_1 - 7 \ , \quad Q_2^S = 2P_2 - 10 \ , \quad W_2 = W_1 - 0.05W_1^2 \qquad \textbf{(4)}$$

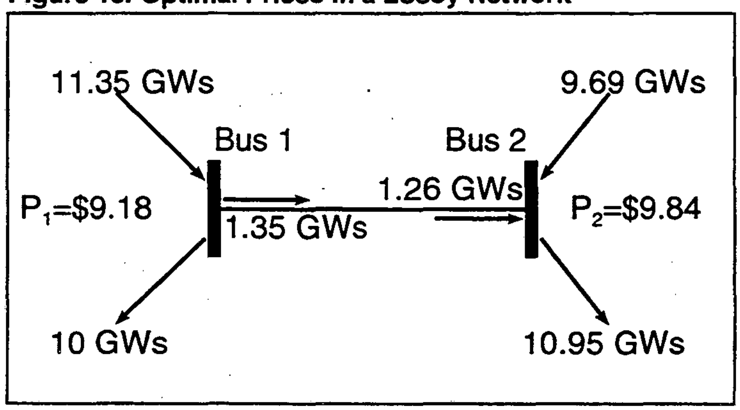
where W_1 is the power flowing into the line at Bus 1 (measured in GW) and W_2 is the power flowing out of the line at Bus 2. The supply and demand equations can be checked by direct substitution. By using Equation (3), the loss-price equation for this example can be seen to be $P_1 = (1 - 0.1\,W_1) \times P_2$. With $W_1 = 1$, this indicates that $P_1 = 0.9 \times P_2$ which completes the check of the optimality conditions.[17]

Now that we have the optimal pricing solution, the next step is to find WEPEX's ZLR prices. This is not a trivial process but because we have only two prices to find, and because ZLR prices are determined by evaluating two free parameters, α and P^*, we can ignore this loss-factor parameterization and simply treat the two prices as free parameters. This is a trick that could not be used with more than two busses. Having said this, there are still two conditions that must be satisfied by the prices: the power balance condition; and the zero-loss-revenue condition. These can be expressed as follows:

$$\begin{aligned} \text{Balance:}\quad & W_2 + Q_2^S = 10.95 \\ \text{ZLR:}\quad & W_1 \cdot P_1 = W_2 \cdot P_2 \end{aligned} \qquad \textbf{(5)}$$

Note that these equations must be used in connection with the equilibrium conditions in Equation **(4)**. Note also that the ZLR equation has been simplified by observing that no net revenue is collected on the power that simply flows in and out of Bus 1, or in and out of Bus 2. It is only on power flow from Bus 1 to Bus 2 that the ISO can make money. Solving these equations is a bit messy, so the details will not be presented here, but the interested reader can easily check the results by substitution. The solution is shown in Figure 13.

Figure 13. Optimal Prices In a Lossy Network



The first thing to observe is that the ZLR prices are not terribly different from the optimal prices. The price change at both nodes is well less than 5% and thus somewhat less than might be expected. This may come as a surprise to the generator at Bus 1 whose price was low by $1 in the optimal dispatch and who may know the rule that, under marginal-cost pricing, the ISO typically "over collects" by an amount equal to the value of the loss. Consequently this generator might have expected an increase in price of 50¢, not 18¢. Two things must be remembered: (1) when all the nodal prices change, all generators change behavior, and (2) the ISO's revenue is reduced both by raising the price at low-price, net-generation nodes and lowering the price at the high-price, net-load nodes.

17 Solving for these prices and quantities is a little difficult, but the reader only needs to check, supply, demand and price equations, all of the optimality conditions, to check that the presented solution is correct.

From an efficiency point of view it is good that prices are not drastically changed from their optimal values. But the prices changes do have one dramatic consequence. By raising the price at sources and reducing the price at sinks, the ZLR prices act to increase transmission loads. In this example it can be seen that the flow on the line increases 35%. Because losses are quadratic, this causes an increase in losses of 83%. In Section 6 we noted that grid costs are approximately $20/kW-yr. Thus an increase of 350 MW at the system peak would require a more robust grid with an increased cost of approximately $7 million/year. Long-distance flows in California are many times the 1 GW value in this example, so this figure should be scaled up appropriately. This cannot be considered an estimate but it is not implausible. The excess lost power is 42 MW. Assuming this example represents only about 1/5 of the California flows and that these flows only occur 6 hrs/day, the extra power losses from ZLR prices could have a value of 5×(42 MW)×($20/MWh)×(2,190 hr/yr), which is $9 million/year.

Several factors make these estimates very uncertain. First, I have used supply elasticities of about 2 because I believe that supply is somewhat elastic. In fact it could be much more elastic than assumed, but a supply elasticity of 1, for example, was found to give only a 21% increase in line flow. Second, the initial line loss is crucial. The value shown is generally considered to be on the high side, but it may not be when the system is heavily loaded and when the full path from generation bus to load bus is considered. A third factor is demand elasticity which is quite low in the short run and was assumed to be zero in this example. But if the WEPEX system is implemented it will be viewed as a long-run policy, and the long-run demand response will become applicable. This may well be near one. Finally it should be noted that this effect is strongest exactly during the system peak when it is most damaging.

Before leaving this example we should note where the income transfers occur. The generators at the lossy bus gain, and their loads lose. The generators at the lossless bus lose and their loads gain. But of course we know net losses outweigh net gains because the ZLR dispatch is not least-cost.

Given that ZLR pricing is more complex (since it requires the computation of optimal prices before ZLR re-pricing can even begin) and that it makes the dispatch less efficient, causes increased line losses, and requires a stronger power grid, one might well ask why anyone should propose it. The ISO Application gives two answers. "First, it sends efficiency-enhancing, marginal cost-based signals to users of the transmission grid. Second, it ... does not require over-collecting and rebating revenues associated with losses." The Applicants have carefully worded the first reason as "efficiency-enhancing" instead of "efficient" because they know that optimal pricing is efficient but that ZLR pricing is less so. The second "reason" is no reason at all; it is simply one group's preference stated as a rationale. And, we have already seen that the ISO will be doing exactly this type of "over-collecting" and "rebating of revenues" for congestion pricing.

8 Conclusion

California's WEPEX has proposed a compromise system for trading bulk power on a consolidated and publicly administered transmission grid. This compromise is intended to allow both nodal spot-market trading (in the PX) by those who want the convenience and openness of a public market and private bilateral trading (in the BXs) by those who believe this gives them more flexibility in contract design. The intention is good, and if wisely implemented, this compromise could provide not only an efficient marketplace but an extremely useful comparison of the two trading models. If the PX and BXs are treated equitably, then the market will determine which system it prefers, and this should give some indication of which system is more efficient.

Unfortunately the compromise is a classic example of design-by-committee and bears the inevitable scars of such a process. Fortunately, the underlying system is essentially sound and all of the necessary remedies are simplifications. This underlying system, which would produce the fair and workable compromise sought by WEPEX, should work as follows:[18]

1) The PX accepts bids from nodal traders and submits all of them to the ISO.
2) The BXs accept proposed trades and bids, processes them according to its own rules, and submits the result to the ISO.
3) The ISO finds the least-cost dispatch and curtails proposed trades when necessary for security reasons.
4) The BX's are given one chance to voluntarily curtail trades that have unacceptably high congestion costs and losses (curtailments can be converted to Dec bids).
5) The ISO computes a final least-cost dispatch, curtailing only when necessary.

I have avoided specifying this system in detail and do not suggest that the WEPEX system be revised to such an extent. Instead the examples in this paper suggest changes that would help restore the clarity and efficiency of the fundamental system while modifying the WEPEX proposal as little as possible. These changes are:

1) Allow the PX and BXs to submit all of their bids to the ISO.
2) Allow the ISO to minimize cost using all PX bids.
3) Allow the ISO to minimize cost using all submitted bids. (Includes 2.)
4) Redefine zonal interfaces based on differences in marginal cost.
5) Use the standard TCC definition (not one based on actual usage).
6) Allow cost minimization to correctly account for losses.

The examples presented illustrate the need for each of these proposed changes.

---

18 Note that this is only a basic trading system. I have addressed neither the unit commitment problem nor the problem of ancillary services. It will be well worth investigating these once the basic system is corrected.

Example 1: A Sub-Optimal PX Dispatch

Example 1 demonstrates that both the PX and ISO can fail to achieve a reasonable dispatch of PX bidders because of a combination of bid and redispatch restrictions. At times this could cost PX users $10,000/hr on a single congested path; however no true estimate can be made without a more complete description of the real-time market than is currently available. Probably the more serious problems associated with this example are the unfairness to the load which is forced to pay for congestion which does not exist, and the disadvantage of the PX relative to the BXs. This problem would be completely solved by Changes 1 and 2 above.

Example 2: The ISO Redispatch Restriction

Example 2 shows that the redispatch restriction causes the ISO to leave too many lines at their security limits. The redispatch restriction has costs both because the resulting dispatch is not minimum cost, and because it reduces the system's security level. At this time, no plausible calculations can be done regarding the magnitude of these costs. Although this problem will be partially solved by traders working around the ISO's dispatch,[19] because of unit-commitment problems and the difficulties of trading in real time one must expect a significant part of this problem to remain.

Unnecessarily dispatching lines at their security limits increases the chance of and likely severity of system collapse. Because this problem can be solved without any increase in complexity and with a decrease in generation costs, change number 3 is recommended: that the ISO redispatch all the way to least cost.

Example 3: Backwards Pricing of Inter-Zone Congestion

Example 3 shows that a combination of diversity among SCs and the ISO's redispatch restriction will sometimes cause inter-zonal congestion to be negatively priced. This rewards those who add to congestion. Surprisingly this makes sense given the WEPEX redispatch restriction because additional congestion gives the ISO greater scope for redispatch, thereby allowing it to clear the market more completely. This example does not show a separate flaw, but instead shows an unexpected consequence of the redispatch restriction. This anomalous pricing runs so counter to the expectations of the Applicants that it will undoubtedly be addressed soon after the ISO begins operation if not before.

---

19 Expensive generators can simply fail to meet their contractual obligations and pay for replacement power in the real-time market.

Although an ad hoc fix (such as declaring congestion costs to be zero whenever they are computed to be negative) is possible and even likely, Change 3 above would solve this problem in a simple and efficient manner.

Example 4: Scheduling Coordinators Will Not Manage Intra-Zonal Congestion

Example 4 shows that load-price averaging over zones will cause the SCs not to account for intra-zonal congestion in their dispatches. This is because to do so would disadvantage their customers. This flaw will disadvantage the PX, which is required to take into account intra-zone congestion. The cost of this flaw cannot be estimated because it depends on the amount of intra-zonal congestion, which in turn depends on details of the grid and on the effectiveness zone creation. To remedy this bias against the PX, Changes 1 and 2 should be adopted.

Examples 5A & 5B: Contradictions in the Definition of Zones

Example 5A shows that, because of loop flow, if congested paths are defined to be zonal interfaces, then some zonal interfaces must necessarily be uncongested paths. This illustrates a contradiction in the WEPEX definition of zones caused by a failure to account for loop flow. Fortunately, Example 5A also demonstrates how a definition of zones based on nodal prices can resolve this contradiction, at least for the simplest examples of loop flow.

Another question is whether new zones will be created to the extent they are needed. Section 6 estimates that the dead-weight loss from failing to split zones is probably well less than 1% of the cost of generation because the WEPEX "5%" rule on zone creation is actually quite strict. If such a strict rule is accompanied by Change 4, the redefinition of zonal interfaces, the zonal system should be workable and reasonably efficient. This conclusion still needs to be checked on more complex examples.

Example 6: Wrong Incentives from WEPEX TCCs

Example 6 demonstrates that the way WEPEX appears to define TCCs leads to improper incentives for generation. The WEPEX definition seems to specify that payment is received on a TCC only if the owner actually transmits power on the specified path. This definition ignores the incentive properties of TCCs, and consequently subverts those properties. Fortunately the market for TCCs will partially fix this problem by inducing generators to sell their TCCs to loads. Since loads are much less price sensitive, the damage from incorrect price signals will be far less. The fix is quite simple, as specified by Change 5, the standard TCC definition should be used.

Example 7: WEPEX Loss Factors Increase Line Usage

Example 7 demonstrates that the WEPEX procedure for collecting losses will encourage excessive use of the transmission system and lead to significantly increased losses. In order to compute the zero-loss-revenue prices required by WEPEX, it is necessary to compute optimal prices first, so their procedure is considerably more complex than optimal pricing.

The excess losses on a 1 GW flow could be as much as 40 MW or $800/hr at $20/MWh. Although California uses 30 or 40 times this much power, most flows have much lower losses than in the example. In the long run a larger cost may be the cost of upgrading the network to accommodate the increased flow that results from encouraging lossy suppliers. This cost is estimated to be several times larger. Again this problem can be averted by implementing Change 6. This would require the standard marginal-loss calculation, which in any case must be done as the first step in the WEPEX calculation.

References

Bushnell, J., Stoft, S. 1996a. "Improving Private Incentives for Electric Grid Investment." Forthcoming in *Resources and Energy Economics.*

Bushnell, J., Stoft, S. 1996b. "Electric Grid Investment Under a Contract Network Regime." *Journal of Regulatory Economics*; 10:61-79.

CEC Comments 1996. California Energy Commission's Comments on the WEPEX Applications. May 28. John D. Chandley, Assistant Chief Counsel.

Hogan, W. 1992. "Contract Networks for Electric Power Transmission." *Journal of Regulator Economics* 4(3): 211-242.

ISO Application 1996. Joint Application of Pacific Gas and Electric Company, San Diego Gas & Electric Company, and Southern California Edison Company for Authorization to Convey Operational Control of Designated Jurisdictional Facilities to an Independent System Operator. April 29. FERC Docket No. EC96-19-000.

Oren, S., Spiller, P., Varaiya, P., Wu, F. 1994. "Nodal Prices and Transmission Rights: A Critical Appraisal" UCEI POWER Working Paper, PWP-025, Dec.

Papalexopoulos, Alex D. Undated. "Challenges to an On-Line OPF Implementation." PG&E.

PX Application 1996. Joint Application of Pacific Gas and Electric Company, San Diego Gas & Electric Company, and Southern California Edison Company for Authorization to Convey Operational Control of Designated Jurisdictional Facilities to an Independent System Operator. April 29. FERC Docket No. ER96-1663-000.

SDG&E Statement: Explanatory Statement of San Diego Gas & Electric Company. Filed with FERC June 28, 1996. (Now withdrawn.)

Singh, H., Papalexopoulos, A. and Hao, S. 1996. "Power Auctions and Network Constraints." PG&E.

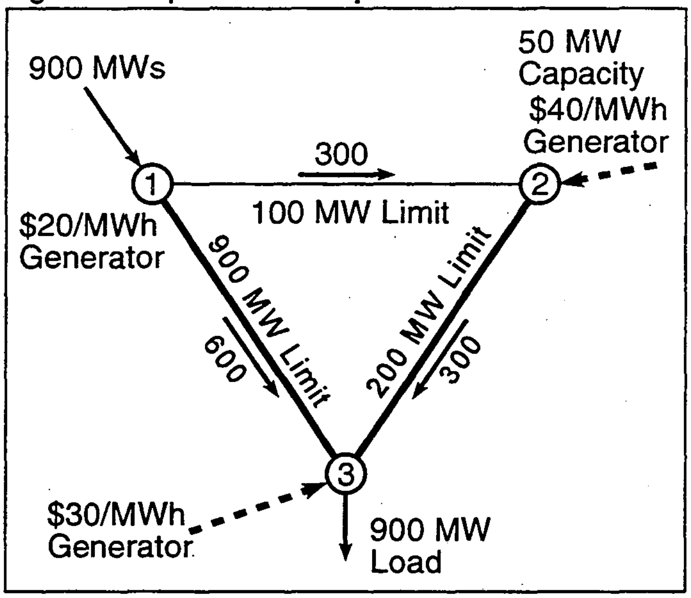
Wu, F. and Varaiya, P. 1995. "Coordinated Multilateral Trades for Electric Power Networks: Theory and Implementation." UCEI POWER Working Paper, PWP-031, Jun.

Wu, F., Varaiya P., Spiller P., Oren, S. 1994. "Folk Theorems on Transmission Access: Proofs and Counter Examples" UCEI POWER Working Paper, PWP-023, Oct.

Appendix A: When WEPEX Makes Redispatch Impossible

The strong interpretation of the ISO redispatch restriction (which is probably the one intended) requires that the ISO leave a certain set of lines exactly at their security limits. These are the lines that are at or above this limit in the preliminary "preferred dispatch" of the SCs. Instead of the usual inequality constraints, these lines must satisfy equality constraints. To satisfy these, k, equality constraints, the ISO can control power injections at N busses. Since any network with more than one "loop" has more lines than nodes, it is possible to have $k > N$. In this case the ISO must solve a system of equations with more equations (k) than variables (N). Barring unlikely coincidences, this is impossible. Because real networks are sparse (i.e. with N nodes, they have many fewer than the maximum possible $N \times (N-1)/2$ lines), and given that we do not expect a large fraction of the lines to be overloaded in the preliminary dispatch (although little is know about such dispatches), it seems likely that $N > k$ may never occur. But, there are two other possible causes of inconsistent redispatch restrictions.

Figure 14. Impossible Redispatch



A second possibility is that the equality constraints do not meet at a single point.[20] Of course this happens in the case just described, but it could also happen if one constraint is parallel to and inside of another constraint. This could happen, for example, if two lines with different limits connect the same two nodes, as well as in more complex circumstances. In this case, the inner constraint cannot be satisfied without passing the point at which congestion is relieved on the outer constraint. This contradiction could be resolved by allowing the ISO to "more than relieve congestion" on any constraints that are strictly dominated by other constraints.

[20] I would like to thank Bill Hogan for pointing this out.

A third possibility is the most likely. Consider the preliminary dispatch shown in Figure 14. Here we have a preliminary dispatch of 900MW from the cheapest generator violating two flow constraints, i.e., on line 1-2 and on line 2-3. (Line impedances are equal.) According to the strong interpretation, the ISO should leave both of these lines at their security limits. This is a perfectly feasible dispatch of the network (G1 = 400, G2 = 100, G3 = 400 MW), but it cannot be accomplished because of G2's capacity limit of 50 MW.

Now we can see the ISO's real dilemma. It cannot leave both lines at their security limits, so it must make an arbitrary choice. Should it come as close as possible to congesting line 2-3 as required and dispatch generator G2 to its maximum capacity, or should it ignore line 2-3 and achieve the least cost dispatch? The first choice seems closer to the spirit of the restriction, but the second choice certainly makes more economic sense.

Appendix B: TCCs

B.1 Hedging Long-Term Generation Contracts

The CEC is concerned with market entry, and TCCs are designed to encourage this by facilitating the financing of new IPPs. TCCs serve this function by allowing a new project to write a long-term contract at a fixed price without risk to either part. Without this hedge, buyers are reluctant to enter a long-term fixed price contract, and without such a contract generators cannot get project financing which is financing that uses the project's assured-income stream as collateral for a loan. Since project financing, because it is low risk, is cheaper than equity or bond-market financing, TCCs make new IPP financing cheaper. For very small companies without the necessary assets to utilize the financial markets effectively, TCCs and project financing make entry possible.

The absence of TCCs will hinder entry, especially by small firms, but the nature of IPPs is changing rapidly, and the trend towards larger firms will undoubtedly continue. For this reason, the absence of TCCs should not have dire consequences. Nonetheless, market power remains one of the largest factors jeopardizing the success of a free electricity market, so any facilitation of entry should be most welcome.

One factor pointed out by the CEC and not explicitly recognized by WEPEX is the nature of the TCC risk premium. TCCs could be issued either by the ISO or by the private market. But, because the ISO suffers no risk when it backs TCCs, and because a private originator does, the ISO can create TCCs more cheaply. This should translate into savings for the generators who buy TCCs. The reason for this savings is as follows. Generator are hedgers and so are willing to pay the TCC's expected value (generally positive) plus a risk premium. If TCCs are in very short supply, the generators will be exposed to more risk and will be willing to pay a greater risk premium. If the market is private, the supply of TCCs will shrink to the point where generators are willing to pay a high enough risk premium to bring forth that level of supply from risk-averse TCC underwriters. But if supply in the ISO market is determined by the feasibility rule (explained shortly) the supply of TCCs will be sufficient to satisfy all hedgers (those actually using the grid). In this case the price should approximately equal the expected TCC values and should certainly be lower than in a private market.

B2. TCCs and Grid Investment

WEPEX envisions a combination of market-driven and regulatory grid-investment procedures. In this section we will consider the market-driven expansions that happen without reliance on the backstop regulatory process, and we take it as given that WEPEX wishes to provide an environment that minimizes the reliance of economically beneficial projects on the backstop regulatory process.

TCCs can serve this function well, though not perfectly. To make maximum use of the market for grid investment, it is necessary to reward investors with TCCs. The proper matching of reward (or in case of perverse investments, punishment) with physical grid expansion is crucial. This is slightly complicated by the fact that a grid's transfer capability is extremely hard to quantify. A single grid is capable of innumerable different dispatches, and any modification of the grid is almost sure to add very many new feasible dispatches and may well eliminate numerous others. As long as the added feasible dispatches are more valuable than those eliminated, the expansion deserves some reward.

An effective rule for rewarding investment was proposed by Hogan (1992) and has been assessed by Bushnell and Stoft (1996).TCCs and Incentives for Private Grid Investment This rule, the feasibility rule for TCC allocation, works by using an analogy between TCCs and an actual dispatch. Just as a dispatch can be viewed as a set of power flows between grid nodes, so TCCs specify a particular power flow between two nodes. Thus a set of TCCs corresponds naturally to a set of power flows; in other words, to a dispatch of the system. If the dispatch corresponding to a set of TCCs is a feasible dispatch, then we define that set of TCCs to be a feasible set of TCCs. If T is the complete set of TCCs that has currently been issued by the ISO, the feasibility rule for allocating TCCs to investors is this:

Feasibility Rule for TCC Allocation: The investor must take, at his choosing, some set of TCCs, t, such that $T+t$ is a feasible set of TCCs.

Bushnell and Stoft (1996) have shown that, in general, investors will find it most profitable to choose a set of TCCs that match their expected use of the grid, and that if all grid users have TCCs that match their grid usage, no detrimental grid expansion would be profitable. TCCs also allow an investor to capture some of the positive externalities of her investment. For instance, if after building the line, it turns out that another generator can make better use of it and the line is congested, then the investor can allow this generator to bid up the congestion charge and capture the use of the line. In this case the TCC will net her more profit than she would have made by generating.

There are several problems with this use of TCCs to facilitate the market for grid investments. First, it does not fully overcome the free-rider effect. Second, the matching of TCCs to grid use may be quite imperfect, which will keep TCCs from being a perfect discouragement to detrimental expansions. Third, the grid, and thus the feasible set of dispatches, is constantly changing due to both deliberate and accidental outages.

The free rider problem will probably need to be overcome with a backstop regulatory mechanism similar to the one proposed by WEPEX. Matching may well be good enough so that, given the cost of transmission projects, almost all detrimental projects are prevented. Nonetheless, this problem deserves further investigation. Temporary grid modifications could cause the ISO to collect less revenue than necessary to cover TCCs, but this effect is minor and at other times the ISO will run a surplus.

The problem of changes in the grid is not a problem for the definition and allocation of TCCs, but it could be a problem for the ISO's finances. After each upgrade of the grid, there must be a well-defined set of feasible dispatches, given that the grid is fully operational. This is necessary for all of the dispatch procedures under consideration. This definition of feasibility will be used for the "Feasibility Rule for TCC Allocation," so this rule is unambiguous. Fortunately, when TCCs are allocated by this rule and the grid is fully operational, the ISO is guaranteed to have adequate revenue from congestion charges to cover the cost of all the TCCs it has issued. But when the grid has a reduced capacity due to an outage this is not necessarily the case. Fortunately there are many factors that mitigate this problem. First, the grid's capacity is defined by taking into account contingencies in such a way that if only one contingency arises, no curtailment will be necessary. If no curtailment is necessary, then the ISO's revenue will still be accurate. Second, the ISO will often have positive net revenues when TCCs fail to match usage. Third, there is no reason that other sources of funds, such as sales of TCCs for the existing grid, could not be used to cover shortfalls in ISO revenue. The net result is that this problem should be minimal.

ERNEST ORLANDO LAWRENCE BERKELEY NATIONAL LABORATORY
ONE CYCLOTRON ROAD | BERKELEY, CALIFORNIA 94720

Prepared for the U.S. Department of Energy under Contract No. DE-AC03-76SF00098