Comments on the California ISO

MRTU LMP Market Design

SCOTT M. HARVEY,
SUSAN L. POPE

LECG, LLC
Cambridge, Massachusetts 02138

WILLIAM W. HOGAN

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

Prepared for
California Independent System Operator

February 23, 2005

TABLE OF CONTENTS

Page

I. EXECUTIVE SUMMARY[i]

The California Independent System Operator (Cal ISO) proposal for its electricity Market Redesign and Technology Upgrade (MRTU) builds on basic principles of efficient use of electric networks and the associated locational marginal pricing (LMP). The present report reviews the details of the still evolving design to compare it against related features of other markets, identifies potential problems or internal inconsistencies, and suggests directions for future modifications. In addition to a review of the documents identified below, there has been extensive discussion with the CAISO as the design evolution has continued. The comments here reflect the MRTU design as specified in the documents we reviewed, as clarified in discussions with ISO staff.

The starting principles of the MRTU embrace the essential foundations of a successful electricity market design including bid-based, security-constrained, economic dispatch with locational prices, license plate access charges, bilateral schedules, financial transmission rights, a consistent network model for commercial transactions recognizing actual physical conditions, consistent day-ahead and real-time markets, unit commitment with simultaneous optimization of energy and ancillary services, and a multi-settlement system. The MRTU will be a major and important reform needed to address the difficulties inherent in the original market design that is to be replaced. The MRTU is also a complex package with many interconnected details developed through a lengthy process of analysis and interaction with stakeholders. The present report highlights problematic features of several of these details, ranging from serious matters that require immediate attention to improvements that should be considered for future implementation. The critical problems can be fixed to produce a highly effective market design.

Importantly, this evaluation is limited to the LMP market design and has not reviewed operational elements of the MRTU. In addition, this evaluation is limited to the conceptual description of the LMP market design and has not reviewed the proposed implementation of this market design.

The issues are listed in a rough order of priority:

1. The most problematic feature of the MRTU market design is the proposal for the nodal clearing of zonal load bids at load aggregation points (LAP), particularly when combined with the zonal settlement of nodally cleared load bids. As discussed in Section II.B and illustrated in Appendix I, these features of the market design would hinder Load Serving Entities (LSEs) from effectively managing their power costs in the day-ahead market because of the disconnect between the price bid by the LSE and the price used to clear the bid. Moreover, the reaggregation of the nodally cleared bids would provide inefficient bidding incentives for loads that would likely result in large uplift costs while leading to the kind of infeasible day-ahead schedules that have so burdened the Cal ISO markets in the past and that the MRTU LMP market design is intended to eliminate.

2. Reliance on highly aggregated load zones for pricing and congestion hedging combined with the nodal clearing mechanism would also undermine the effectiveness of otherwise attractive feature of virtual bidding in providing price convergence and

may cause competitive generators in constrained regions to withhold generation from the day-ahead market (DAM),[1] making this capacity available only in real time (as discussed in Section II.D). These problems are not inherent in the overall market design and could be avoided either by clearing the zonal load bids based on zonal prices (as is done in PJM and New York) or by disaggregating the zonal load bids into nodal bids that could be cleared and settled nodally. The use of highly aggregated zones is also likely to undermine the ability of the Cal ISO to award CRRs that effectively hedge congestion costs; and may lead to unintended cost shifts among transmission customers (as discussed in Section VIII and illustrated in Appendix VI). Solutions to these problems that have been employed by other ISOs are discussed.

Other features of the MRTU market design that ought to be addressed prior to implementation are:

3. The real-time congestion pricing mechanism for imports and exports (discussed in Section IV.B) is likely to produce a disconnect between the bids accepted in the hour-ahead scheduling process and real-time congestion prices because the constraints enforced in the hour-ahead scheduling process will not be reflected in the real-time dispatch. The NYISO encountered precisely this problem at start-up and the problem was not eliminated until the implementation of "ECA B" in Fall 2000, so it should be anticipated that similar problems would arise for the Cal ISO.

4. The mechanism proposed for determination of ramp limits for implementation of real-time constrained output generator (COG) pricing (discussed in Section V.B) could result in the calculation of inappropriately high prices during circumstances in which uneconomic gas turbines are operating as a result of either minimum run time or minimum-down time constraints. The NYISO also encountered this problem at start-up which required extensive price corrections until it was corrected in late July 2000. Whether similar patterns will appear in California depends on the relationship between the quantity of COG unit capacity located behind transmission constraints relative to the ramp rate of steam units on line behind those constraints and on dispatch procedures.

5. The failure to attempt to accurately reflect all costs (NOx allowances, current gas prices) in the calculation of start-up and minimum-load costs for the purpose both of clearing the day-ahead financial market and the reliability unit commitment (RUC) (Section VII.B) could lead to inefficiency, inflated resource adequacy costs and potentially compromise both gas and power system reliability. In particular, the use of bid-week gas prices for unit commitment and Pass 1 mitigation purposes. during periods in which the gas pipeline system is constrained, and high spot gas prices must serve to allocate gas to the highest valued uses could undermine both gas and power system reliability.

---

[1] The incentive would not arise from any ability to exercise market power but simply from the incentive of competitive generators to offer their capacity so as to be paid the market price.

6. The structure of demand response compensation (the difference between the nodal and LAP price, discussed in Section V.C) would not provide appropriate incentives for demand response in circumstances in which prices are high but there is little congestion. Further, the structure could give rise to incentives for behavior that would inflate costs to consumers during periods in which there is congestion but no shortages requiring demand response.

All of the problems identified have been successfully addressed in other LMP markets and can be readily addressed through changes consistent with the market designs that have been implemented in other regions.

Other potentially problematic features of the MRTU market design include:

7. A relatively low "soft" bid cap of $250 (discussed in Sections VII and IX) that could adversely impact reliability in California independent of LMP implementation by the Cal ISO. These problems would appear under circumstances of either high gas prices or capacity shortages, both of which could arise again during future low hydro conditions.[2] In these circumstances the $250 bid cap, whether it is hard or soft, will concentrate the impact of western capacity shortages on California consumers and likely limit supply during periods of high gas prices. When binding under these circumstances, the soft bid cap would compromise most of the transparency and beneficial incentives intended for the MRTU design.

8. Activity rules for bid reductions that encourage rather than deter extreme bids in real-time could be a problem. (See Section IV.B) The MRTU activity rules permit market participants to convert day-ahead schedules to self-schedules (with a price of -$30/MWh) in real-time but do not allow market participants to reduce their day-ahead offer prices by more moderate amounts to reflect costs that are sunk in real-time. This foreclosure of more moderate offer price reductions could potentially magnify the constrained off payments associated with real-time transmission outages that render day-ahead schedules infeasible as well as reducing the efficiency of the real-time dispatch.

9. A lack of a full multi-settlement system for ancillary services that optimizes real-time reserves and settles deviations from day-ahead schedules at real-time prices (discussed in Section VI.F) could raise consumer costs when reserves scheduled in the DAM must generate energy in real-time as a result of minimum run times, minimum down times or transmission constraints.

10. The use of extreme decremental (DEC) bids for Pass 1 schedules in Pass 2 of the DAM, with the intent of "minimizing" incremental (INC) adjustments and the use of forecast load in the market power passes of the DAM (Passes 1 and 2) and bid load in the scheduling and pricing pass (Pass 3) of the DAM (discussed in Section VII.C)

---

[2] While gas and power prices have been relatively low in California and the west over the past few winters, power prices have exceeded $250/MWh in New England on a variety of occasions during the winters of 2003-4 and 2004-5.

could render the RMR dispatch and local market power mitigation process ineffective in some circumstances. If RMR units in practice possess little or no local market power, this ineffectiveness would not adversely impact consumers.

11. The single pass mitigation structure for non-RMR units potentially possessing local market power (discussed in Section VII.D) will fail to mitigate the exercise of market power by non-RMR units that possess market power but face competition from high cost alternatives. Whether this feature would adversely impact market outcomes depends on whether there are any such non-RMR units that possess material local market power.

12. The availability payment for Residual Unit Commitment (RUC) capacity (discussed in Section III.C) will likely have unintended consequences if it becomes a mechanism for suppliers to recover the difference between actual market costs and those used to calculate the bid production cost guarantee for RUC units and in circumstances in which the Cal ISO forecasts a capacity shortage but there is no shortage either in the day-ahead financial market nor in real-time. There seem to be few reasons for the availability payment that would not be addressed by other modifications of the MRTU design.

While it would be desirable in principle to address these features of the market design, it is uncertain whether these latter six features of the market will, in practice, have much adverse impact in the near term. As above, most of these problematic features can be readily addressed by eliminating peculiarities of the MRTU market design and bringing it into closer alignment with the market designs in PJM and New York.

Section X discusses the relationship between the proposed MRTU market design and California Public Utility Commission (CPUC) resource adequacy proposals. The CPUC resource adequacy proposals were at a very early stage of development in these documents, but we did not identify any troubling inconsistencies between the MRTU market design and the general resource adequacy proposals that would hinder implementation of the kind of resource adequacy program outlined by the CPUC. Further, the CPUC Interim Opinion has generally identified the key complications in implementing a resource adequacy system, and many of the details remain to be developed.

These comments on the LMP market design elements of the California ISOs are based primarily on the following ten documents: the California ISO, Comprehensive Market Design Proposal, July 21, 2003 (hereafter CMD); the California ISO July 22, 2003 CMD Transmittal Letter (hereafter CMD Transmittal); California ISO September 17, 2003 Answer (hereafter Sept ISO); the FERC October 28, 2003 Order (hereafter Oct FERC); Cal ISO May 11, 2004 technical conference comments (hereafter May ISO); the California ISO June 2, 2004 Reply Comments (hereafter June ISO); the FERC June 17, 2004 Order (hereafter June FERC); California ISO, CRR Study 2, Final Scenario Assumptions, July 19, 2004 (hereafter CRR Study 2); California ISO, Congestion Revenue Rights Preliminary Study Report, October 1, 2003 (hereafter CRR Study 1) and the FERC September 20, 2004 Order on Rehearing (hereafter Sept FERC).

The comments on the MRTU market design and the CPUC resource adequacy proposals are based on the following documents: the January 22, 2004 Interim Opinion in Rulemaking 01-10-0924 (hereafter Interim Opinion); Comments of the California ISO on Workshop Report on Resource Adequacy Issues, July 14, 2004 (hereafter Cal ISO Workshop Comments); and Opening Comments of the California ISO Corp. on the Draft Decision of ALJ Wetzell Regarding Interim Opinion on Resource Adequacy, September 22, 2004 (hereafter Cal ISO Interim Opinion Comments).

Document	Reference
California ISO, Comprehensive Market Design Proposal, July 21, 2003	CMD
California ISO July 22, 2003 CMD Transmittal Letter	CMD Transmittal
California ISO September 17, 2003 Answer	Sept ISO
FERC October 28, 2003	Oct FERC
Cal ISO May 11, 2004 technical conference comments	May ISO
California ISO June 2, 2004 Reply Comments	June ISO
FERC June 17, 2004 Order	June FERC
California ISO, CRR Study 2, Final Scenario Assumptions July 19, 2004	CRR Study 2
California ISO, Congestion Revenue Rights Preliminary Study Report October 1, 2003	CRR Study 1
FERC September 20, 2004 Order on Rehearing	Sept FERC
January 22, 2004 Interim Opinion in Rulemaking 01-10-0924	Interim Opinion
Comments of the California ISO on Workshop Report on Resource Adequacy Issues, July 14, 2004	Cal ISO Workshop Comments
Opening Comments of the California ISO Corp. on the Draft Decision of ALJ Wetzell Regarding Interim Opinion on Resource Adequacy, September 22, 2004	Cal ISO Interim Opinion Comments
Draft Proposal for the Allocation of Congestion Revenue Rights to Merchant Transmission, August 6, 2004,	Cal ISO MT

II. DAY-AHEAD MARKET

A. General Structure

The Market Redesign and Technology Upgrade (MRTU) market design for the electricity markets coordinated by the California Independent System Operator (Cal ISO) builds on basic principles of efficient use of electric networks and the associated locational marginal pricing (LMP). The present report reviews the details of the still evolving design to compare against related features of other markets, identify potential problems or internal inconsistencies, and suggest directions for future modifications.

The context for the MRTU arises from the original California market design which included a number of critical defects. Most notable for the wholesale market were various features that reflected a commitment to market separation, simplified commercial transmission models and zonal pricing. Market separation referred to the principle that the commercial model for day-ahead scheduling and hour-ahead scheduling would be constructed largely independent of the realities of physical operations under the control of the Cal ISO. The commercial schedules and pricing would then be based on a simplified electric model that abstracted from the underlying constraints in the real electrical network. Zonal pricing allowed further claims for simplification by requiring only a few prices. The rationale for these features was that the Cal ISO should not be involved in operating day-ahead and hour-ahead energy markets, and the differences between the commercial schedules and settlements relative to the actual schedules required to maintain reliability could be resolved without significantly affecting the underlying incentives.

The outcome of these design principles was that commercial schedules in the day-ahead market were inconsistent with real-time requirements and operational constraints. In turn, this created inefficient incentives for competitive market participants, and would exacerbate any problems arising from an exercise of market power. The resulting incentives created the need for a system of rules and side-payments that eliminated the presumed simplicity of the commercial model, both for short-run operations and long-term investment. It is now recognized that the Cal ISO must administer certain critical markets. Therefore, the most fundamental wholesale market design reform requirement is to restore consistency between the commercial model and the physically reality, and between the day-ahead market and real-time operating constraints.

The starting principles of the MRTU embrace the essential foundations of a successful electricity market design including bid-based, security-constrained, economic dispatch with locational prices, license plate access charges, bilateral schedules, financial transmission rights, a consistent network model for commercial transactions recognizing actual physical conditions, consistent day-ahead and real-time markets, unit commitment with simultaneous optimization of energy and ancillary services, and a multi-settlement system. The MRTU will be a major and important reform needed to address the difficulties inherent in the original market design that is to be replaced. The MRTU is also a complex package with many interconnected details developed through a lengthy process of analysis and interaction with stakeholders. The present report highlights problematic features of several of these details, ranging from serious matters that require immediate attention to improvements that should be considered for future implementation. The critical problems can be fixed to produce a highly effective market design.

The presentation here begins with a discussion of the day-ahead market. However, throughout it must be understood that the day-ahead market cannot be considered in isolation from the rules for real-time operations and settlements. Participants will anticipate their ultimate treatment in the real-time balancing market when making decisions under the scheduling and bidding rules of the day-ahead market (DAM).

The MRTU day-ahead market will include several unit commitment and dispatch passes, some for the purpose of bid mitigation, some to determine day-ahead prices and schedules and some to ensure that sufficient resources will be available in real-time to meet the ISO's load forecast. This includes the pre-integrated forward market reliability and market power mitigation runs (Pre-IFM-RMPM) for competitive constraints (CC) and all constraints (AC). We briefly summarize the overall structure of the DAM here to clarify the discussion which follows. The details of the mitigation passes are discussion in Section VII.D below and the details of the reliability unit commitment (RUC) are discussed in Section III.

Pass 1A [Pre-IFM-RMPM-CC]: Market Power Mitigation Pass – Internal generation and import supply committed and dispatched to meet forecast load plus export bids, only monitoring competitive constraints

Pass 1B [Pre-IFM-RMPM-CC]: Market Power Mitigation Pass – Rerun Pass 1A (unit commitment and dispatch) with mitigated bids. Only bids that violate a conduct threshold are mitigated in this pass. Depending on the impact of mitigation on Pass 2 versus Pass 1 prices, the mitigated bids may remain mitigated or may revert back to the initial bids in the subsequent passes.

Pass 2 [Pre-IFM-RMPM-AC]: Local Market Power Mitigation Pass – Run a unit commitment and dispatch to meet forecast load, monitoring all constraints, based on the Pass 1 schedules. Reliability must-run (RMR) generation is available for dispatch.

Pass 3A: Market Schedule Pass – Internal generation and import supply is committed and dispatched to meet bid load, internal and export demand. RMR cost based bids for RMR dispatch levels from Pass 2, market and mitigated bids for other units as determined in Pass 1B or 2. Pass 3A determines the day-ahead market schedules.

Pass 3B: Market Pricing Pass – Final DAM Pricing Dispatch based on the Pass 3A unit commitment, mitigated bids and submitted demand schedules and bids. Some offer prices used in Pass 3A may not be eligible to set prices in Pass 3B. Pass 3B determines the DAM prices that are used for day-ahead settlements.

Pass 4: *RUC Pass* – Unit commitment and dispatch to meet forecast load, taking into account the day-ahead import and export schedules, minimizing the commitment cost (along with any RUC availability payments, where relevant) of adding capacity not committed in Pass 3A. Pass 4 determines the unit commitment, but does not determine prices or financial schedules (other than prices for RUC availability payments).

The proposed MRTU day-ahead market structure is workable and generally consistent with the structures used by PJM and New York. It has two features that deserve comment. First, an important feature of the structure of the proposed DAM is that after bids are mitigated in Passes 1 and 2, Pass 3 will recommit the market from the beginning based on the mitigated bids. This is an important difference between the Cal ISO market design and that employed in PJM that we believe is appropriate for the California market.[3] Second, the proposed MRTU market design will, like PJM, determine DAM schedules and market prices prior to commitment of RUC units. This differs from the approach in New York in which DAM schedules and market prices are determined after the commitment of RUC units. Both approaches have been shown to be workable and have advantages and disadvantages. An important advantage of the PJM approach is that if the ISO's load forecast is correct, the operation of the units added in the RUC pass will often be profitable at real-time prices so their commitment should give rise to relatively little uplift.[4] A potential disadvantage of the PJM approach, however, is that it may inflate ancillary services costs if ancillary services are scheduled in the DAM prior to the RUC commitment.[5] This issue is discussed in Section VI below.

The scheduling and pricing passes (3A and 3B) of the day-ahead market will minimize the as bid cost of meeting load based on a full network model, taking account of as bid energy costs, start-up costs, minimum-load costs, and ancillary services capacity charges.[6] The MRTU day-ahead market will simultaneously schedule energy, ancillary services and manage congestion.[7] There will be no distinction between inter and intra zonal congestion.[8] As clarified by the Cal ISO the DAM unit commitment will use all resources, regardless of which units are committed or dispatched in the market power mitigation passes. This is an important feature as

---

3 The proposed approach is also consistent with the structure of the NYISO DAM, which also recommits generation following the market power mitigation passes.

4 The DAM financial schedules would not include the minimum-load block of units committed in the RUC pass, so real-time supply at the DAM price would exceed the supply cleared in the day-ahead market. If the ISO's load forecast were accurate, however, real-time load would also exceed the supply cleared in the day-ahead market and the operation of the units committed in the RUC would generally be economic at real-time prices, as real-time prices should be higher than the prices in the day-ahead market. If real-time load were lower than the load forecast used for the RUC commitment or if additional low cost imports were to become available in real-time, there would be an increased potential for the RUC commitment to give rise to uplift costs on units committed in the RUC.

5 PJM avoids this outcome by not determining ancillary service prices in the day-ahead market. The NYISO avoids this outcome by clearing the DAM and scheduling ancillary services after committing units to meet forecast load and local reliability (the NYISO equivalent of the RUC commitment).

6 See CMD # 6, 32. The CMD originally provided that "The only resources considered for commitment in the DAM will be those committed in the pre-IFM-RMPM runs" (Passes 1 and 2 in the terminology adopted for this discussion). See CMD # 60. We understand that due to other changes in the DAM, the Cal ISO has dropped this restriction. Such a restriction could lead to peculiar outcomes. If it is not economic in Pass 3 to commit some units that were committed based on forecast load in Pass 2 and lower cost units are not permitted to be considered, DAM prices could be artificially high in some hours. This potential is clearest if gas turbines not dispatched in the pre-IFM RMPM runs are not permitted to be available for the DAM dispatch in pass 3. Gas turbine offer prices should cap prices in the DAM, but if they cannot be committed because they were not dispatched in the pre-IFM-RMPM, much higher bid curve portions on units committed could clear, while other units committed in Pass 2 would not be committed because they are uneconomic given bid load.

7 CMD # 6.

8 CMD # 17.

restricting the resources available in the scheduling and pricing passes (3A and 3B) can lead to unanticipated and undesirable outcomes that raise the cost of meeting load, regardless of whether Pass 2 is based on bid or forecast load. The commitment and dispatch of RMR units will be integrated into the DAM and security-constrained unit commitment (SCUC).[9] This is appropriate and should improve market performance.

Under the MRTU, energy bids in the day-ahead market for internal resources will include start-up cost, minimum-load cost and incremental energy curve and capacity bids for ancillary services.[10] The incremental bid curve for energy is a staircase function with up to 20 segments.[11] The unit commitment will take account of unit ramp rates, minimum run times and minimum down times.[12] These constraints must be the actual physical constraints of the unit.[13] Units may submit up to 10 ramp rates over the operating range of the resource for use in DAM and in real-time. Ramp rates will be fixed for the day and can only be changed when there is a change in the ramping capability of the unit.[14] Loads can submit price sensitive bids in the DAM reflecting their willingness to reduce consumption or their willingness to buy power at real-time prices.[15]

If the DAM commits a resource that was not self-committed, the resource is eligible for recovery of start-up and minimum-load costs, for the hours the unit would have been off-line based on its self-schedule.[16]

The CMD originally proposed that unit commitment would use a multi-day time horizon in order to take account of units with start-up times that are longer than one day.[17] The MRTU unit commitment process will now optimize over the operating day. Units with start-up times too long to be accounted for in the time frame of the day-ahead market will need to self-schedule the start-up and minimum-load blocks of these units in the DAM. Generators can generally internalize these multi-day commitment issues. There is, however, some interaction with the mitigation procedures for start-up costs, which is discussed in Section VII.B. An off-line process will be used to commit long-start-up time units when they are needed for reliability purposes over a multi-day time frame. This general approach to the commitment of units with long start-up times is reasonable and could be workable. However, the details should be reviewed as they are developed.

The closing time for the DAM will continue to be 10:00 a.m. The Cal ISO will produce a final DAM schedule before performing RUC. The Cal ISO will publish DAM schedules at end of RUC, in combination with RUC schedules around 1:00 p.m.[18] This appears to be a remarkably aggressive schedule given that there are three complete SCUC Passes and several additional

---

[9] CMD # 75, 146.

[10] CMD # 18, 25.

[11] CMD # 24.

[12] CMD # 57, 60.

[13] CMD # 105.

[14] CMD # 57.

[15] CMD # 127.

[16] CMD # 61.

[17] CMD # 60.

[18] CMD # 73, 74.

partial passes or dispatches to be completed in this three-hour interval, while presumably also allowing time for posting and a margin for resolving problems. The Cal ISO should leave the flexibility for longer processing period if operational tests indicate that this schedule cannot be maintained.

B. Pricing

1. Nodal Prices

The MRTU day-ahead market will calculate LMP prices at the nodal level.[19] Nodal prices will have three components, energy, transmission and losses.[20] Supply resources, generation and real-time dispatched demand reduction, will be settled at nodal prices.[21] Real-time demand reduction will buy power at zonal/LAP prices in the DAM, while exports will buy power at the nodal (scheduling point) price. Entities that can operate either as loads or generators (cogeneration and pumped storage hydro) will be treated as generators and schedule and settle at the appropriate nodal price in both the DAM and real time.

If there is insufficient supply to serve load in a constrained area, the pricing rules will set the market clearing price equal to the Damage Control Bid Cap.[22] Appropriate scarcity pricing is important to providing efficient incentives but it is important to be very careful in defining insufficient supply. It is not clear whether it is intended that the energy price would rise to the Damage Control Bid Cap only in the event that the “insufficient supply” is so extreme that the Cal ISO undertakes involuntary load shedding to maintain system stability or would rise to the Damage Control Bid Cap if the Cal ISO is unable to meet its reserve targets within that constrained region but is not required to shed load. Setting the energy and reserve price to the Damage Control Bid Cap only in the event of load shedding is not constructive from a resource adequacy perspective as the value of the lost load would likely greatly exceed $250/MWh. Conversely, if it is intended to set prices to the Damage Control Bid Cap in circumstances in which the Cal ISO is unable to meet its reserve targets within the constrained region but both avoids load shedding and violations of WECC reliability criteria, consideration should perhaps be given to whether the price should rise directly to the Damage Control Bid Cap in this circumstance, or whether different degrees of shortage entail varying reserve values that might be better expressed in a reserve demand curve such as that which the NYISO applies to its 30-minute reserves.

---

19 CMD # 6.

20 See CMD #14. The CMD does not appear to contain a definition of the LMP price, either in verbal terms, “the least costly means of obtaining energy to serve the next increment of load at each bus,” or in terms of an equation reflecting shift factors, constraint shadow prices and reference prices, but we understand that the standard definition of the calculation of nodal LMP prices is intended.

21 CMD # 123.

22 See Sept ISO, p. 79.

2. Losses

Under the MRTU, the cost of losses will be incorporated in the LMP prices, using a SCUC that models losses like NYISO.[23] This marginal loss pricing was approved by FERC. The loss residual will be credited to the congestion revenue right (CRR) balancing account. The loss residual will therefore help assure revenue adequacy for CRR settlements with any excess revenues being credited to the Participating Transmission Owners (PTOs), and eventually back to loads through reduced transmission access charges, using the same mechanism that would be applicable to congestion surpluses in the CRR balancing account.[24]

The Cal ISO has stated that allocating residual transaction by transaction is more accurate but complex.[25] However, the loss residual arises from the difference in average and marginal effects and is not driven by separable individual effects. Thus it is not made more accurate by focusing on individual transactions. Further, such a rule would create inefficient scheduling incentives. It is essential to avoid having such methods which conflate average and marginal effects imposed on the Cal ISO. Adding any loss residual to the CRR balancing account as a workable method of allocating these residual revenues has been approved by FERC.[26]

This is a reasonable method for allocating the residual. The critical consideration is to avoiding tying any credits for the loss residual to criteria that can be impacted by market participant actions so that the credit does not distort incentives. It is particularly important not to tie the credit to market participant schedules. The other issue is fairness, and under the MRTU surpluses in the CRR account will flow back to all loads, albeit indirectly through reduced access charges. If the loss residual is utilized to subsidize the award of a material amount of infeasible CRRs that are awarded to particular load serving entities, however, this could result in material cost shifts among transmission customers.

Market participants will not be able to explicitly self-provide losses under the MRTU, but can schedule generation in excess of load in order to, in effect, provide losses in kind.[27] Thus, there is no need to attempt to reconcile the irreconcilable difference between average and marginal losses. This is a reasonable approach that avoids creating artificial scheduling incentives that could distort the market and raise the cost of meeting load.

FERC directed the Cal ISO to explain how the allocation method for the loss residual will apply to entities that self-provide losses.[28] FERC again asked Cal ISO to explain how allocation method will apply to entities that self-provide losses in the June order.[29] It appears that FERC does not understand that the loss residual will simply be credited to the CRR balancing account. This ought to be clarified prior to the tariff filing.

---

23 CMD # 71.

24 See May ISO, p. 5, 70-76; Oct FERC ¶ 77; June FERC ¶ 142, 143.

25 May ISO, p. 73.

26 Oct FERC ¶ 78; Sept FERC ¶66; June FERC ¶ 144-146.

27 CMD # 72.

28 Oct FERC ¶ 78.

29 June FERC ¶ 148.

FERC encouraged the State of California to provide other rules for pricing and loss charges for wind generation.[30] These kinds of rules need to be carefully reviewed because they can lead to a revenue shortfall problem for the Cal ISO. Depending on the rules adopted, wind generators could have an incentive to submit financial schedules to the most distant load in the Cal ISO control area in order to maximize loss rebates. FERC is correct that since the output of such resources is not dispatchable, this will not distort short-run dispatch decisions. It could, nevertheless, be more costly than intended to consumers.

3. Constrained Output Generators

FERC directed the Cal ISO to develop a mechanism for constrained output generators (COG) to set LMP prices.[31] Under the MRTU, constrained output generators (gas turbines (GTs) able to operate only at full output) can set prices in DAM if they are needed to meet load.[32] Constrained output generators will be paid the day-ahead price for their schedule in the dispatch pass (Pass IIIA), and paid the real-time price for their real-time output in excess of their DAM schedule.[33] The proposed COG pricing was approved by FERC.[34]

This approach to pricing is workable and consistent with COG/fixed-block pricing in Eastern ISOs. COG pricing will, at times, result in higher prices than would otherwise be the case, but prices will better reflect the actual cost of serving incremental load, uplift will be reduced, and bidding incentives will be improved. Among other effects, COG pricing helps ensure that if GTs are scheduled in the DAM to support exports, the price paid by the export buyer will be sufficient to cover the costs of the GTs and will not give rise to uplift.[35] Implementation of this kind of COG pricing in the DAM is relatively straightforward by treating the units as unconstrained in the DAM pricing pass. The proposed scheduling rule will satisfy the revenue adequacy theorem applied to day-ahead and real-time schedules.[36]

---

[30] June FERC ¶ 153.

[31] Oct FERC ¶ 89.

[32] See May ISO, pp. 3-4, 58-61, Att A III.1, 2. It is stated in the September Answer (p. 61) that the NYISO's DAM pricing treats the minimum-load energy of generation committed in the RUC as flexible and able to be dispatched below minimum operating point. This is incorrect. The minimum-load blocks of generation committed in the NYISO RUC are treated as fixed blocks. It is gas turbines scheduled in the bid load pass that are treated as flexible/dispatchable in calculating prices.

[33] May ISO, pp. 58-61, Att A III.3.

[34] June FERC ¶ 121.

[35] Under the MRTU, however, the use of bid-week gas prices to commit generation in the DAM does give rise to the possibility that gas-fired generation may at times be uneconomically committed to support exports, with much of the resulting uplift borne by Cal ISO load.

[36] In the Midwest some transmission owners have quick starting gas turbines that have a minimum-load block and also a dispatchable range above the minimum-load block. We do not know if there are units with these characteristics located in the California ISO control area. If such units are present, restrictions would likely need to imposed on the relationship between the offer price for the minimum-load block and the dispatchable range or other adjustments made in the COG pricing system to better accommodate such units.

Like other units, COG units will be eligible for a bid production cost guarantee in the event they are dispatched but operate unprofitably.[37]

4. Zonal/LAP Pricing

Loads will buy power at load aggregation zone prices that are averages of the nodal prices over the service territories of the three investor-owned utilities (IOUs).[38] Except for load served by unconverted existing transmission contracts (ETCs), there will be three load aggregation zones for the purpose of load scheduling, bidding, and settlement, defined as the transmission service areas of the three California IOUs. "Virtually all loads within the ISO control area that are not served under ETCs will be scheduled, bid and settled at the level of the load aggregation zone in which they are located." This would include municipal utility and direct access load as well as load receiving retail service from the IOUs distribution utility.[39] This load aggregation pricing was approved by FERC.[40]

The ISO will assign load distribution factors (LDFs) to individual nodes within the load aggregation zones for running the DAM and establishing the final schedules.[41] The LDFs will vary by time period, e.g., business day, Saturday, and Sunday holiday and for different hours within the day (peak, non-peak, etc.). LDFs will be established and revised based on the ISO's state estimator.[42]

The exceptions to Zonal/LAP pricing for load in the day-ahead market are: 1) loads served under non-converted ETCs will be excluded from the load aggregation pricing system. These loads will schedule and settle at locations appropriate to their specific ETC rights.[43] 2) Entities that can operate either as loads or generators (cogeneration and pumped storage hydro) will be treated as generators and schedule and settle at the appropriate nodal price.[44]

---

37 CMD # 61, 106, 116.

38 CMD # 15, 84, 123.

39 See CMD # 64. The Cal ISO notes in several documents, for example its Sept 2003 Comments at p. 54, that this is consistent with Eastern ISOs. This is not accurate in the case of New York. There are multiple zones within the service territories of Con Ed, NYSEG and NIMO. For example Con Ed's service territory spans zones H, I and J. Both NYSEG and NIMO have load in many of the other zones.

40 Oct FERC ¶65.

41 "[T]he ISO will assign appropriate weights to each aggregation for the purpose of calculating aggregate prices as the weighted average of the nodal prices comprising each aggregation." CMD # 62. These weights will presumably be consistent with the amount of load clearing at each node in the nodal clearing process. See also CMD # 62.

42 CMD # 63.

43 The power delivered under ETCs will not be purchased in the spot market and will not pay congestion, so the price is irrelevant for these loads. This load will therefore not be included in the load weights used to calculate the zonal/LAP price. Entities holding ETCs can, however, choose to meet their load by purchasing power from the spot market, rather than using their ETCs. We anticipate that they would do so in circumstances in which the LMP price for the LAP in which they are located is lower than the LMP price at the generation used to serve their load, as their schedules would generate counterflow payments in these circumstances. This is discussed in general below.

44 CMD # 124.

The proposed LAP pricing has five features that are potentially problematic. These features are: (1) the nodal clearing of zonal bids; (2) reaggregation of nodal schedules into infeasible zonal schedules; (3) changes in nodal load weights between the DAM and real-time; (4) payments for zonal counterflow by vertically integrated LSEs; and (5) the LAP settlement rules for demand response. The nodal clearing of zonal load bids, the implied reaggregation of nodal schedules into infeasible zonal schedules, and the demand response pricing based on LAP pricing have consequences that need to addressed prior to MRTU implementation. In the discussion below we describe alternatives for correcting these features while retaining the LAP pricing system. The effects of changes in nodal load weights and payments for financial counterflow do not give rise to critical problems that need to be addressed prior to MRTU implementation but it is necessary that market participants understand these consequences which are intrinsic to a LAP pricing system and may impact choices regarding the degree of load aggregation. Each of these topics is discussed in greater detail below.

An important element of the MRTU design is the proposed clearing process for zonal/LAP load bids in the DAM. The apparent motivation for the proposed zonal/LAP bidding and nodal clearing mechanism springs from a perception that the proposed LAP zones are large and heterogeneous. Hence, the averaging of prices across the LAP eliminates significant price differences at individual nodes. This is a fundamental problem that preserves one of the principal defects of the previous market design that the MRTU was intended to eliminate.

Although scheduling coordinators will submit load schedules and bids at the level of the default aggregations, the DAM will perform congestion management and energy trading at the nodal level. Prior to running the DAM, load schedules and bids will be disaggregated to the nodal level using load distribution factors that are derived from the ISO's real-time state estimator. The DAM will then make adjustments to generation and load at the nodal level to clear congestion and execute energy trades.[45] It is proposed that zonal/LAP load bids will be evaluated and scheduled in the unit commitment and dispatch based on the zonal bid applied to the nodal prices for each node within the LAP.

It is essential to recognize that this is not the manner in which NYISO, PJM or ISO-NE clear zonal load bids. In the Eastern LMP markets, zonal load bids are cleared in the DAM against the zonal price. Thus, in New York if a LSE bids to buy 100 MW of power at any price less than $45/MW, it will buy 100 MW if the average zonal price is less than $45/MW and buy no power if the average zonal price exceeds $45/MW. This will not be the case under the methodology described in the CMD, where load would be cleared node by node at nodal prices, rather than being cleared based on the average nodal price reflected in the LAP price. The approach described in the CMD has some unattractive features. First, zonal bids that are less than the zonal price may not entirely clear in the DAM if some nodal prices exceed the zonal average price. Second, zonal bids may partially clear in the DAM even though the zonal price exceeds the bid (because some of the individual nodal prices would be less than the average zonal price). We understand that the Cal ISO would address this second situation by charging

---

[45] See CMD # 125. Similarly, the CMD stated that "The IFM will adjust schedules at the nodal level for clearing the energy market and managing congestion and to determine nodal prices." CMD # 62.

the buyer its bid rather than the zonal price (making up the difference between the Zonal/LAP price and its bid through an uplift payment).[46]

These features of the zonal/LAP clearing process mean that if there is congestion within a load zone/LAP, LSEs will not be able to use their bids in the DAM to efficiently limit the price they pay for power. If an LSE submits zonal load bids reflecting the expected average zonal price, its bids will only partially clear in the DAM because the nodal prices in the high priced portion of the zone will exceed the bid. This will leave the LSE exposed to real-time prices on the load that does not clear in the DAM. Alternatively, if an LSE submits zonal bids reflecting the expected price level in the high priced portion of the zone, the LSEs' bids may clear when the zonal price in the DAM exceeds the expected real-time zonal price and thus the cost of meeting load would be too high.

These inconsistencies in the DAM outcomes could be readily addressed within the constraints of a zonal/LAP load bid system by clearing the LAP bids against the LAP price, as do the other ISO coordinated markets, rather than clearing the zonal bids against the individual nodal prices within the zone. This approach would embody in the clearing rules the same assumption as in the bids; namely, that the zonal aggregate demand is allocated proportionally across the individual nodes. There would still be a potential loss of efficiency compared to a full nodal system, but the bids would be consistent with the DAM dispatch. Absent actual nodal load bids, it is not apparent that clearing the demand bids at the nodal level serves any purpose.

A zonal bidding system is less efficient that a fully nodal bidding system. However, within the context of a zonal bidding system, a concern has been expressed that clearing zonal load bids against a zonal LMP price would not clear the market efficiently compared to other clearing mechanisms for zonal bids. The apparent hope is that the efficiency of the nodal system might be achieved through a system which relies on zonal bids. The MRTU design does not accomplish that probably impossible task.

To address these issues, first it is important in this context to understand that while there is potential inefficiency in a zonal bidding system if other market participants have better information than the ISO regarding the nodal distribution of real-time load, the inefficiency does not arise from the zonal clearing. Appendix II illustrates the impact of ISO errors in forecasting the real-time nodal load distribution and it is seen that these errors can cause the ISO to run surpluses or incur deficits in its real-time settlements. If the ISO errors are centered around the true value and no other market participant has superior forecasts, the surpluses and deficits would roughly cancel out. There would nevertheless be a real resource cost of these errors because on some days too much generation will be committed and load could have been met at lower cost using fewer units. Conversely, on other days not enough generation will have been committed and too much load will be met with quick start units in real-time. These resource costs are simply the inevitable result of imperfect load forecasting.

---

[46] We understand that it is intended that the DAM LAP price would be calculated based on load weights determined by the quantities cleared at each node in the DAM rather than by the load weights used to disaggregate the zonal bids to the nodes. This understanding is reflected in the discussion and examples of LAP pricing in the appendices but is not fundamental to the conclusions.

The issue regarding zonal clearing of zonal load bids is different. If a price capped load bid does not clear because the load bid is lower than the zonal price, that means that it would cost more than the load bid to serve that load given the expected distribution of loads within the LAP. Partially clearing the price capped zonal load bid as if it cleared throughout the zone/LAP when it only clears at some low-priced nodes in the zone/LAP does not improve market efficiency. Such a clearing mechanism is simply clearing zonal/LAP load bids at a price lower than the cost of meeting that load. As illustrated in the examples in Appendix I, the result of that clearing mechanism would be to create real-time uplift, because the zonal/LAP load bids that would be cleared in this way in the DAM could not actually be supplied in real-time at the zonal/LAP price.

Another alternative for addressing this problem would be to require that all price capped load bids be submitted and cleared on a nodal basis and that only price taking self-scheduled load could be submitted on a zonal basis. With LDFs that accurately reflect the distribution implicit in the schedules, this would effectively mean moving to the equivalent of a nodal pricing system for load, although it might appear to be a zonal system to the self-scheduled load. Another alternative would be to accept nodal bids for loads but settle them at a zonal price. This would lead to a requirement for constrained-on and constrained-off payments. In effect all load would be treated like the demand side bids discussed below, rather than just "price capped" load bids and would present the same challenges in setting the base line allowed load bid as for the proposed demand side bidding.

A second problematic feature of the zonal/LAP pricing system is that the CMD provides that subsequent to the clearing of zonal load bids against nodal prices, the nodal load schedules clearing in the DAM will be reaggregated to create final load schedules for each scheduling coordinator at the default aggregation level, which will be settled at aggregate prices that are load-weighted averages of the constituent nodal prices.[47] This means that although zonal load bids would be cleared against nodal prices based on disaggregated nodal load representations, any portion of the zonal load that clears in the DAM at any node will be treated for settlements as if it cleared for the zone as a whole. This feature is one of the eight major implementation issues we have identified with the MRTU market design and is the most serious of these problems.

This manner of reaggregating nodally cleared load bids into zonal schedules will at best produce revenue inadequacy in the Cal ISO's real-time settlements because DAM schedules will be infeasible. Worse, this system would provide incentives for inefficient bidding behavior even in the absence of any market power. The incentives would be to exploit the inconsistency between the DAM schedules and settlements to potentially magnify the revenue inadequacy of the Cal ISO's real-time settlements, with potentially large uplift costs falling on LSEs that do not engage in these problematic bidding strategies. Moreover, it does not appear that it would be feasible to impose rules that would discourage such inefficient bidding incentives without seriously undermining the ability of LSEs to manage their exposure to high prices in the day-ahead and real-time markets.

---

47 CMD # 125.

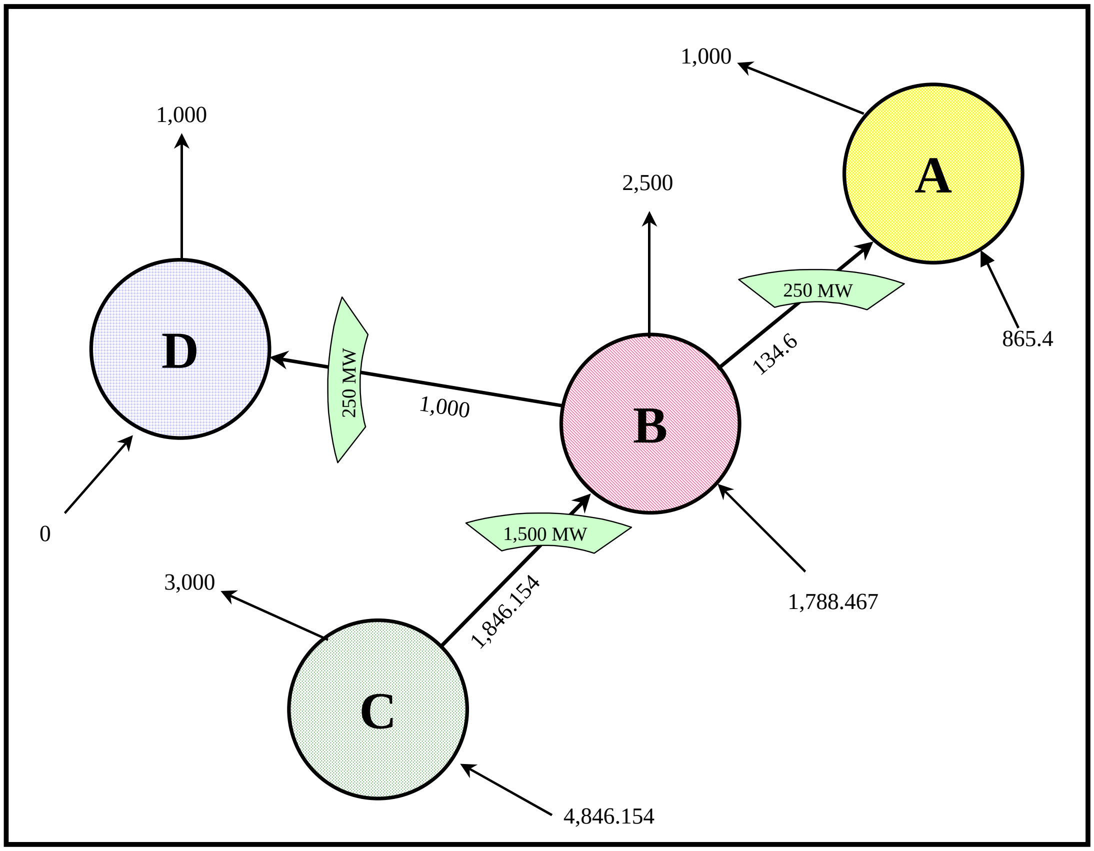
The revenue inadequacy arises because zonal load bids would in effect be cleared in the DAM using one set of nodal load weights and then settled in real-time using a different set of load nodal weights. Hence, the day-ahead market could be cleared as if all the load were in the unconstrained portion of the zone, and then settled in real-time as if a portion of the load cleared in the DAM were in the high priced constrained portion of the zone. Since generation would have been scheduled in Pass 3 of the DAM based on the nodal load weights used to clear the DAM bids, insufficient generation would be scheduled in the DAM to meet the load scheduled in the constrained portion of the zone. In essence, the DAM schedules would be infeasible because generation would have been scheduled to meet load based on the load distribution within the zone determined by the nodally cleared zonal bids, and then this load would be moved into the constrained portion of the LAP in the reaggregation process but the generation needed to meet this load would not have been scheduled. In real-time, the ISO would need to back down generation able to serve load at the nodes cleared in the DAM and dispatch up higher cost generation able to serve load at the high cost nodes where zonal load bids did not clear in the DAM. This buying back of low priced generation and purchase of high priced generation would create real-time uplift costs that would be borne by consumers.

To illustrate, suppose that there are two locations, A and B, in the same zone but with no transmission between them (to simplify the point). Load is price-sensitive ex ante (i.e., in the DAM) but fixed in real time. The demand curves of the LSEs serving the two locations are identical, with 100 MW at $60, 200 MW at $40, and 300 MW at $20. LSEs submit these two demand curves in the DAM as the load zone bids, which would be aggregated into a zonal demand for 200 MW at $60, 400 MW at $40 and 600 MW at $20. It is further assumed that there are unlimited $20 generators at A, and unlimited $60 generators at B. The LDFs for the regions are 50 percent.

Under the nodal clearing process the aggregate zonal demand would be allocated to the nodes based upon the LDFs for the regions. The demand for region A would clear 300 MW at $20, while the demand for region B would only clear 100 MW at $60. Hence, the nodal clearing price at A becomes $20 with load 300 MW, and $60 at B with load 100 MW. The load quantities also determine the generation schedules in the DAM. The DAM zonal price would be the weighted average load price, (300 MW * 20 + 100 MW * 60)/400 MW, or $30/MW. Under the nodal load reaggregation it would be assumed that 50 percent of the 400 MW of load was cleared in each region, so 200 MW would be scheduled at A and 200 MW of load at B, but 300 MW of generation would have been scheduled at A and only 100 MW at B. In real-time, the ISO would have to purchase 100 additional megawatts at bus B at $60, and sell back 100 MW of generation at A at a price of $20 to cover the DAM schedules. Hence the ISO would be revenue inadequate in real time. The problem would be more complicated with transmission included, increasing bids on the supply curves and asymmetric distributions of demand.

Thus, even if LSEs passively bid their expected real-time load into the DAM at the expected real-time zonal price and real-time load and prices were exactly equal to that expected by the LSE, the Cal ISO would be revenue inadequate in real-time if there was congestion, because not enough generation would have been scheduled in the DAM to meet load within constrained regions.[48]

Figure 1
INFEASIBLE SCHEDULES WITH NODAL CLEARING



This revenue inadequacy would reflect the fact that the schedules used for DAM settlements would be infeasible because the generation scheduled in Pass 3 to meet load at the low-priced nodes could not be dispatched to meet load distributed across the LAP based on the DAM settlement load weights if load bids were cleared nodally and then reaggregated.[49] This infeasibility is illustrated in Figure 1 (developed in Appendix I, Section C), in which the flows implicit in the DAM schedules produced by the MRTU LAP clearing and settlement rules exceed the limits on C-B and B-D.

This infeasibility of the DAM schedules would further exacerbate uplift costs because some of the infeasible DAM schedules would be attributable to generation in generation pockets with DAM schedules to meet inflated load within the pocket. Thus, if region C were a generation pocket with a single supplier, the C-B DAM schedules would be infeasible in real-time and the Cal ISO would have to accept DEC bids within the generation pocket. The load would be shifted out of the pocket in the settlement system and the generation would have to be

---

48 This potential is illustrated in Appendix I, Section B.
49 This is also illustrated in Appendix I, Section C.

dispatched down in real-time because there would be insufficient physical load within the pocket in real-time. The offer prices of the DAM schedules would automatically be reduced to -$30/MW in real-time and a single supplier within the generation pocket would not have to worry about lower cost redispatch being offered. The ISO would have to buy back the infeasible schedules at a high cost, just as under the zonal system today. Moreover, this outcome could not be addressed by conduct rules as the infeasible schedules are the result of LSEs bidding in their actual load and the MRTU rules would automatically reduce the DAM offer prices to -$30/MWh. These LAP settlement rules undermine one of the fundamental purposes of LMP implementation, the elimination of infeasible DAM schedules that can be exploited with low real-time offer prices, sometimes referred to as the "INC-DEC game."

The infeasibility of the Pass 3 DAM schedules would be identified in the RUC which would model forecast load at its expected nodal location so sufficient generation would be committed in the high priced regions to meet forecast load, but too much generation would have been scheduled in Pass 3 to meet load in the low priced region. Moreover, since LSEs could bid so as to cover their entire load in the DAM as described above, there would not necessarily be any real-time load imbalances to be assigned the RUC uplift costs. In fact, very little of the uplift costs would be attributable to the RUC commitment. The units committed in the high priced regions to meet forecast load would likely not incur large losses at real-time prices if the Cal ISO's load forecast were accurate, the uplift costs would be in the day-ahead market due to the infeasible schedules and allocated to all loads.

This revenue inadequacy could be exacerbated if LSEs were to bid to take advantage of the reaggregation of infeasible DAM schedules in submitting their zonal load bids. This would entail bidding the amount of the LSE's zonal load that is hedged with CRRs from low cost generation into the DAM at very high prices to ensure that it clears in the DAM. The LSE would then submit additional price capped load bids in the DAM capped at levels reflecting the expected price in the low priced portion of the zone, in an amount that exceeds the LSEs expected real-time load. The amount of price capped load bid into the DAM by the LSEs could be determined based on the expected nodal load weights such that when the zonal load bid was cleared nodally, the amount of load that cleared at the nodal price in the low priced region would be sufficient to cover the entire portion of the LSEs zonal load that was not hedged with FTRs from low price sources. This kind of bidding could produce very large revenue inadequacy in the Cal ISO's real-time settlements.[50] While the Cal ISO and FERC could attempt to deter such conduct by forbidding LSEs from bidding more than their expected load into the DAM, conduct rules of this sort could also prevent LSEs with better information than the Cal ISO from appropriately hedging themselves in the DAM.

The arbitrage by market participants could go further. LSEs could submit price capped load bids designed to clear bids in the DAM in excess of the LSEs real-time load, with the price capped bid structured to ensure that the load bids cleared only in the low priced nodes of the LAP. This excess load would then be sold back in real-time at the real-time LAP price, which would reflect actual congestion patterns and generation costs, giving rise to further uplift costs.[51]

---

[50] This potential is also illustrated in Appendix I, Section C.

[51] This kind of outcome is also illustrated in Appendix I, Section D.

Arbitrage by market participants could proceed until the LAP DAM price rose to the level of the expected real-time LAP price, eliminating arbitrage profits. Under the nodal clearing mechanism for zonal LAP bids, however, the DAM nodal generation prices and schedules produced by this arbitrage would likely be very different from the nodal prices and schedules expected in real-time. The result of such "perfect" arbitrage is illustrated in Appendix I, Section E, and discussed further in Sections II.D and IX.A.

These features of the proposed load zone pricing system are a critical problem that must be addressed before the MRTU market is implemented, but this problem can be readily solved. There are basically two ways to address the problem, both of which have been applied elsewhere and could be readily applied to Cal ISO markets. One approach would be to keep the proposed zonal/LAP pricing but to clear the zonal bids zonally based on zonal prices and ISO load weights. The other approach would be to further disaggregate the LAP either into smaller load zones or all the way to the node.[52] Under either approach, it would be essential that bids would be cleared, and settled, on a basis consistent with the level of aggregation at which the bid is defined.

As explained above, the zonal clearing of zonal bids need not lead to any inefficiency, other than that inherent in zonal aggregation, but highly aggregated load zones can expose the ISO to arbitrage and may reduce the effectiveness of virtual load bidding as discussed below.

A third potentially problematic feature of the zonal/LAP load pricing system is the proposed change in nodal load weights between DAM and real-time.[53] The potential problem can be best understood by assuming that the problems with zonal/LAP pricing discussed above were addressed by clearing zonal load bids zonally against the zonal price, which would eliminate both the nodal clearing process and the need for reaggregation. In addition, we assume for the purpose of this discussion that there is no demand response in real-time so real-time load does not depend on prices.[54] Under the proposed market design, the ISO would determine the nodal load weights used to calculate the zonal price and clear the market in the DAM. In real-time, the zonal price would be calculated based on the estimated real-time nodal load weights, which could be different from those anticipated by the ISO for the DAM. Importantly, an LSE that purchased 1,000 MW of power at the load zone price in the DAM would be perfectly hedged for 1,000 MW of consumption within the zone in real-time regardless of differences between the nodal distribution of load in real-time and that assumed in the DAM.

---

[52] We understand that there is a concern that the establishment of multiple load zones within the service territory of a single distribution company could give rise to inefficient incentives, particularly in a direct access (i.e., retail competition) environment.

The issue has arisen in other states which have shown that it can be addressed without the need for LAP zones that are coincident with service territories. As noted above, the service territory of Consolidated Edison of New York encompasses load Zones H, I and J and wholesale market prices are materially higher in Zone J than in the other zones. The New York PSC and ConEd have used a combination of a market supply charge and a monthly adjustment clause to maintain balanced retail access incentives despite ConEd-wide retail rates. The CPUC, the impacted utilities, and the Cal ISO may, therefore, find it helpful to examine the approaches taken in eastern states before concluding that the LAP approach is necessary.

[53] A related issue regarding CRR weights is discussed in Section VIII.

[54] We discuss the implications of relaxing this assumption below.

Under this settlement rule, any financial consequences of errors in the ISO's day-ahead assessment of nodal load weights will be included in real-time uplift and borne by all loads. The ISOs nodal load forecast cannot possibly always be right, but if it is not systematically high or low, the various errors would average out over time, sometimes resulting in real-time surpluses (when too much expensive generation was scheduled in the DAM)[55] and sometimes in real-time deficits (when too little expensive generation was scheduled in the DAM).

This is basically the same settlement rule used by the Eastern ISOs. It places the burden of estimating the nodal distribution of load on the ISO, simplifying the bidding process for LSEs. The critical premise of this approach is that the ISO has better information for estimating this nodal distribution of load than do any of the individual LSEs. If this premise is satisfied, the approach is reasonable. It is important, however, to recognize that this premise requires that the ISO utilize accurate forecasts of the real-time load distribution in calculating nodal load weights in the DAM. If the ISO uses simplified rules that result in predictable inaccuracies between the DAM and real-time load weights, the premise is not satisfied. If market participants have better forecasts of the real-time distribution of load than that used by the ISO to calculate DAM prices, the ISO's DAM prices will be subjected to arbitrage by market participants and the real-time settlements may be revenue inadequate on average.[56] Thus, if the ISO's nodal load weights were anticipated to include too little load in the high priced region of the LAP, market participants would anticipate that real-time load zone prices would exceed DAM load zone prices so they would buy extra load in the DAM which they could sell back in real-time. When the imbalances were settled in real-time at the real-time nodal weights, DAM buyers would in effect be selling back load in the constrained down region as load in the high priced region in real-time. The ISO, on the other hand, would be buying additional generation in the high priced region and buying back excess generation in the lowed priced region, giving rise to uplift and inflating costs for consumers.[57]

The magnitude of the uplift costs associated with arbitrage of predictable errors in the ISO's nodal load weights is likely to be dramatically lower than the costs associated with the currently proposed LAP bid clearing mechanisms. Thus, this is not a critical problem requiring market design changes prior to implementation of the MRTU. Instead, this issue is an area of concern that needs to be kept in mind as the general MRTU market design is carried forward into a software implementation to ensure that implementation decisions do not magnify these costs. The potential cost of this kind of arbitrage can also be reduced by defining load zones within which there is less congestion and less day-to-day variation in nodal load weights. Both factors would tend to favor defining smaller load zones than the LAP zones. Moreover, it is important to understand that the mere reality that the Cal ISO's forecast of nodal loads will not be perfectly accurate does not give rise to a market design problem. The existence of uncertainty is an operational and market reality. The potential market problem is not the possibility that the Cal ISO's forecast of nodal loads will be imperfect but that some market participants will be able to

---

55 Generators in the high priced region buy back their DAM schedules at high prices and the ISO schedules replacement generation in the low priced region at lower prices.

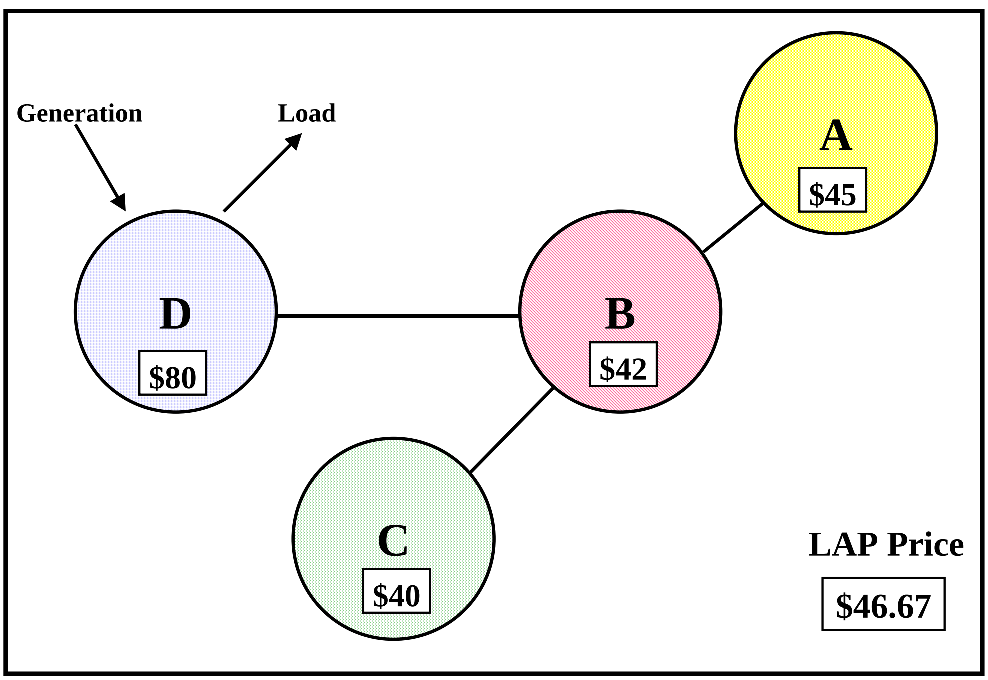
56 Market participants would not need to necessarily identify the underlying inaccuracy in nodal load weights to arbitrage these differences. Market participants could arbitrage these differences merely by observing the circumstances giving rise to differences in DAM and real-time prices.

57 This potential is illustrated in Appendix II.

develop better forecasts of real-time nodal loads than those employed in the DAM and will be able to use those forecasts to arbitrage the DAM as discussed above. The smaller the load zones, the less likely it is that such arbitrage will be a problem. In the case of nodal pricing, the problem disappears because there is only one location in each "zone" with an LDF that is necessarily 1.0.

A fourth problematic feature of the zonal/LAP pricing system concerns vertically integrated LSEs located within the high priced regions of the LAP. If such vertically integrated LSEs have generation sited to meet their load within the constrained region (one or more nodes with the same nodal price) in a nodal pricing system, there would be no congestion charges between the nodes at which the generation is located and that at which the load is located, as illustrated in Figure 1. Consider the D LSE located within constrained region D in which the LMP price is $80. The D LSE's load is also located within region D, so under a nodal system, the LSE would schedule its generation to meet its load and incur no congestion costs.

Figure 2
FINANCIAL COUNTERFLOW SCHEDULES



If the vertically integrated load's LMP price is averaged into the overall LAP for the A, B, C and D regions and the LAP price is lower than the nodal price at the location of the LSE's generation, the LSE will be paid counterflow charges for scheduling its generation to meet its load, although those schedules would actually have no impact on congestion. Thus, under the LAP pricing system, the D LSE would sell its generation in region D at $80 but buy power at $46.67 to meet its load in region D earning a counterflow payment of $33.33/MW, the cost of which would be borne by other LSEs. This cost shift would be avoided if the D LSE were assigned CRR obligations from its generation to the LAP load zone, but D LSE would not designate such CRRs in any voluntary process.

These payments for financial counterflow to the LAP are related to the issue discussed in Section VIII regarding the impact of LAP pricing on the allocation of CRRs. The issues are linked because to the extent that such vertically integrated LSEs receive counterflow payments for financial schedules that provide no counterflow on the real transmission system, the number of CRRs that can be allocated to other LSEs is reduced, raising their costs. These kinds of potential costs shifts could also be reduced or avoided by defining zones within which there is less congestion (and thus lower counterflow payments for scheduling power from generation located within the zone to meet the load zone load).

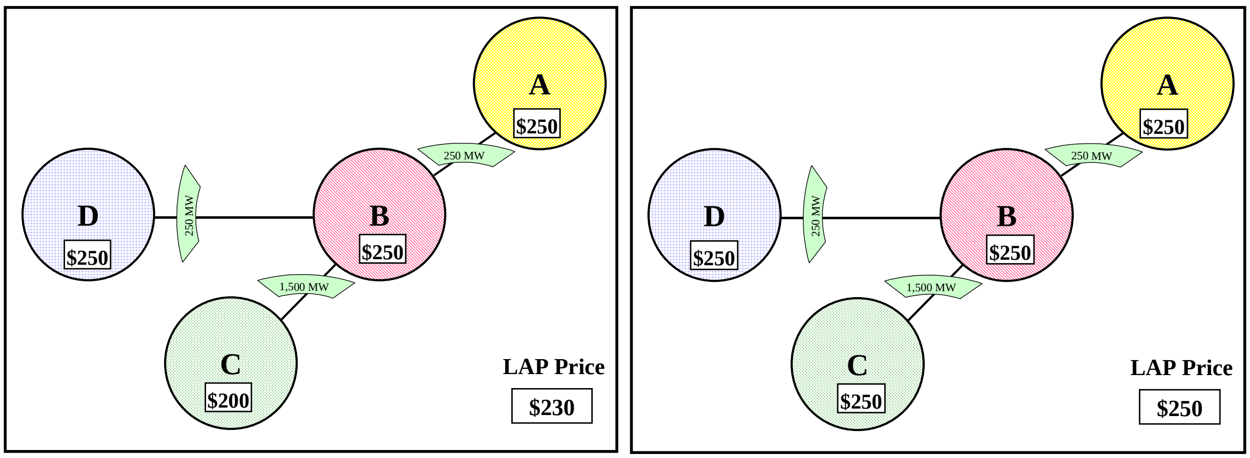
Another potential problem arising from the application of the LAP pricing system to vertically integrated LSEs with generation and load at the same location, is that it could in effect unmitigate otherwise mitigated market power. LSE's serving their own load with their own generation within transmission constrained regions generally have little or no incentive to attempt to exercise market power by withholding output as if they are successful in raising the locational price paid to their generation, they would also raise the locational price they pay to meet their load. If the proposed LAP pricing were implemented and LSEs decline to accept counterflow CRRs from their generation to the LAP, then vertically integrated LSEs might acquire an incentive to exercise market power that they would lack under a nodal pricing system.

A fifth difficulty of the LAP pricing system is the proposed treatment of price responsive load. Under the CMD proposal, long-lead time demand response could buy power at the LAP in the DAM up to their maximum MW of curtailment plus non-curtailable load and would sell the amount dispatched off at the nodal price in the DAM. This approach to demand response pricing has two problematic features. First, the supply of demand response under this pricing formula will be profitable whenever the nodal price exceeds the average LAP price. For consumers located within the constrained portion of a LAP, this may be the case for a large number of hours per year. Whenever this situation is expected to prevail, load providing demand response could offer its full amount of certified curtailment into the DAM (at a very low curtailment price, ensuring that it would be curtailed) even if it had no intention of consuming that amount of power. For example, a load in a constrained region could establish the reduced consumption from shutting a production line as certified curtailment and then bid this reduced consumption into the market to be dispatched at a price of $.10/MWh whenever the production line would not be running due to holidays, reductions in the number of shifts, maintenance or would be operating at less than the maximum level.

Loads in constrained regions at locations at which the nodal prices exceed the applicable LAP price therefore have an incentive to offer the difference between their actual consumption and their established maximum curtailment in the DAM at a load price that ensures that it is dispatched, whenever the nodal price is expected to exceed the LAP price. The required payments for energy not consumed could be expensive for consumers if these demand response resources have maximum curtailment quantities that materially exceed their average consumption.

Conversely, this pricing system provides very little incentive for day-ahead demand response if the system is expected to be strained but there is little congestion because prices are relatively high throughout the LAP. For example, instead of paying demand response the difference between its nodal price and the LAP price, it could be paid the difference between its nodal price and a threshold price defined for that node for consumption curtailments. Under such an approach there would be no payments for demand response on low load days because the nodal price would not exceed the threshold price. Conversely, such a system would permit payments to demand response resources located in regions with prices that are high but lower than the LAP price.

Figure 3
DEMAND RESPONSE PRICING INCENTIVES



Fundamentally, incenting demand response by paying it the difference between the LAP and the nodal price is unrelated to the value of the demand response. This pricing will not result in efficient demand response. Suppose the DAM price was $250 throughout the LAP as portrayed in the right panel in Figure 3, would demand response have no value, simply because the LAP and nodal prices were the same? Conversely, it also appears that this pricing system offers no incentive for demand response within the lower priced portions of the LAP, even when prices are high in absolute terms, because the difference between the Nodal price and the LAP price would be negative, as shown for region C in the left panel in Figure 3.

A second concern with this approach to demand response pricing are the inconsistencies arising from the assumption of price insensitive nodal load weights in the determination of the LAP price in the DAM and the assumption that demand response will in fact be price responsive. If the Cal ISO's nodal load weight forecast does not account for the impact of the demand response, this could give rise to an arbitrageable difference between the load weights used by the ISO in the DAM and those used to calculate the real-time LAP price. In addition, depending on how the various problems with the LAP bid clearing and settlement process described above are resolved, the application of these pricing and settlement rules to entities providing demand response, consuming power and perhaps selling power could provide additional opportunities for inefficient arbitrage that would impose uplift costs on consumers.

Most of these difficulties with the proposed approach could be readily avoided by requiring demand response to buy and sell at the nodal price. The simplest version of this approach has the limitation that loads that have not bought forward will have no incentive to conserve in the day-ahead market. This issue needs to be directly addressed, however, either by allowing such forward purchases or by defining a base level of consumption against which reductions are settled or some similar mechanism.

Overall, while there are some serious problems with the LAP pricing mechanism currently envisioned for the MRTU, these difficulties can readily be avoided by moving to a zonal pricing system more consistent with those employed by PJM and New York.

5. Trading Hubs

Under the MRTU, trading hubs may be defined as needed and appropriate. Initially NP15, ZP26 and SP15 will be modeled as trading hubs. The ISO will also create trading hubs corresponding to the three load aggregation zones (PG&E, SCE and SDG&E) to facilitate trades with load serving entities.[58] The nodes defining the trading hubs will not change over time. The trading hub LDFs used to calculate the weighted average prices will vary, however, and will be based on the total quantity of load that is scheduled at each node of the hub. In contrast to the load aggregations mentioned above, trading hub LDFs and prices will include load served under ETCs if such load is within the ISO control area.[59]

Under the CMD, trading hubs would just be scheduling points for inter-scheduling coordinator trades. That is, market participants would not be permitted to submit self-schedules or price capped bids or offers at these locations. That is, no virtual bids would be permitted at the hubs. If other design features are modified to accommodate virtual bidding, it does not appear that there is any compelling reason for allowing virtual bids at these trading points in addition to allowing virtual bidding at the LAPs, if these trading hubs are to merely be scheduling points for inter-scheduling coordinator trades.[60]

If these trading hubs continue to be defined exclusively as scheduling points for inter-scheduling coordinator trades pursuant to bilateral contracts, the varying hub LDFs will not give rise to ISO revenue inadequacy in the DAM or real-time because the ISO will have a zero net position at the hub in the DAM and thus cannot be financially impacted by changes in the node weights between DAM and real-time. We think it likely, however, that market participants will want to expand the role of these trading hubs in ways that will have the potential to give rise to ISO revenue inadequacy caused by the varying node weights.

First, it is likely that market participants will want to be able to submit price-capped load bids and supply offers at the trading hubs in the DAM. If the Cal ISO accommodates this, the Cal ISO will no longer be assured of a zero net position at the trading hubs in the DAM. Instead, the Cal ISO would be exposed to market participant arbitrage of differences in nodal weights

58 CMD # 65.

59 CMD # 65.

60 A market participant that wanted to cover a contract calling for delivery at SP-15 with a spot market purchase in the DAM, would submit a virtual demand bid at the SCE LAP, submit a transmission schedule from the SCE LAP to the SP-15 trading hub, and then have an inter-scheduling coordinator trade at the SP-15 trading hub to cover its contract. Alternatively, the market participants could turn the contract into a CFD at the SP-15 trading hub.

between the DAM and real-time[61] with successful arbitrage giving rise to uplift costs borne by load. If the changes in nodal weights between the DAM and real-time are, in fact, unpredictable, these differences should average out over time and not give rise to net uplift costs.

Second, it is likely that market participants will want to be able to acquire CRRs to and from the hubs. If the Cal ISO limits the acquisition of CRRs to and from the hub to the unbundling of CRRs defined from a generator to a load, then the CRR simultaneous feasibility test will not include any CRRs sinking or sourcing at the hubs and the CRR settlements will not be impacted by differences between nodal weights used to determine the hub in the simultaneous feasibility test and those used to settle the DAM. In addition, the ISO would need to forbid unbundled CRRs sourcing or sinking at the hub from being reconfigured in the auction. Thus, the hub would not be defined as a possible source or sink in the auction.

If the Cal ISO were, however, to permit market participants to purchase CRRs sourcing or sinking at these trading hubs in the auction, then the Cal ISO would no longer be assured that there would be no net injections or withdrawals at the hub in the auction solution, so nodal weights for the trading hub would need to be specified in the auction. If these weights are different from those used to settle the CRRs in the DAM, as they inevitably would be if the nodal weights varied from day to day in the DAM, then the Cal ISO's CRR settlements would also be exposed to arbitrage by market participants.

It does not appear necessary to us to subject loads to the potential uplift costs associated with trading hubs with changing nodal weight definitions. Given the likelihood that market participants will seek to allow bidding at the hubs and to buy CRRs to and from the hubs in the auction, it would be preferable to avoid the potential for revenue inadequacy and uplift costs by defining the trading hubs in terms of both fixed nodes and fixed nodal weights in the auction, the DAM and in real-time. This approach would be consistent with PJM's treatment of the western hub.

If the fixed hubs prove later to be insufficient for the market, it would be possible and easy to define new hubs in addition to the existing hubs. Preserving the fixed weights for the existing hubs would avoid interfering with contracts. And defining the new hubs with the same type of fixed weight rule would avoid the perverse effects of varying the weights under the same hub definition.

C. Self-Schedules

Under the MRTU, generator and load self-schedules will be submitted without associated energy bids.[62] Generator self-schedules will be handled in the unit commitment and dispatch that determines day-ahead schedules (Pass 3A) by assigning them a highly negative energy bid, such as -$1,000 but they will be price takers in most circumstances. Similarly, load self-schedules

---

61 Market participants would take net long positions at the hubs in the DAM when they anticipated that changes in nodal weights between the DAM and real-time will raise the real-time price and would take short positions when they expect the reverse.

62 CMD #11.

will be assigned a high positive bid, such as $1,000. FERC agrees with Cal ISO that self-schedules have to be price takers.[63]

The MRTU provides that if self-schedules have to be adjusted in the DAM because of insufficient market bids, prices at such locations will be determined administratively based on the damage control bid caps.[64] We assume that within the priorities defined by the Cal ISO this adjustment will be based on minimization of the as bid cost of meeting load, given the -$30 bids, rather than being based on prorata scaling. It is preferable not to have to rely on such non-market adjustments and the -$30/MWh bid floor should be low enough that non-market adjustments will rarely be required.

It is our understanding that generation self-schedules that are adjusted in the scheduling pass (Pass 3A) will be permitted to set LMP prices in the pricing pass (Pass 3B), based on an offer price equal to the damage control bid floor (i.e., -$30/MWh)), and load self-schedules that are adjusted (nodally) in the scheduling pass (Pass 3A) will be permitted to set LMP prices in the pricing pass (Pass 3B), based on a bid price equal to the damage control bid cap (i.e., $250/MWh). The CMD, for example, stated that when self-schedules must be adjusted, prices will be set based on the damage control bid caps.[65] It is our understanding that if self-schedules must be adjusted, the generation self-schedules clearing in the scheduling pass (Pass 3A) would be included in the pricing pass at -$1,000 but a portion of the segment curtailed in the scheduling pass (Pass 3A) would be included in the pricing pass at an offer price of -$30, thus enabling it to set prices.

It needs to be understood, however, that under such an approach -$1,000 self-schedule generation bids could be reflected in, but not set, prices that are less than -$30/MWh and self-scheduled load bids could cause (but not set) prices that exceed $250. If the -$1,000 offer price were on a radial line and there are no tradeoffs with other resources, then limiting the offer price to -$30 in the pricing pass would cause the LMP price at that location to rise from -$1,000 to -$30 in the circumstance in which self-schedules had to be adjusted in the scheduling pass. If there are tradeoffs with other generation that can be dispatched to provide counterflow, however, simply limiting the offer price of an adjusted self-schedule to -$30 in the price calculation pass, may not be sufficient to avoid prices far below -$30 at this location. This outcome would be appropriate in our view, because it reflects the costs imposed on the transmission system (and other customers) by the self-schedule, but it is not clear if this is what the Cal ISO intends.

If the Cal ISO were to use a self-schedule price such as -$1000 for unit commitment and dispatch, but then replace this with a -$30 price floor for the purpose of determining prices, there would be a possibility for market participants to be thereby insulated from the financial consequences of self-schedules that greatly raise the cost of meeting load. As discussed in Section VII.C, we believe that these potential problems are avoided in the MRTU design, which employs a bid floor but does not impose a price floor.

---

63 Oct FERC ¶ 145.

64 CMD # 31.

65 CMD #31.

In the event that self-schedules must be adjusted in the DAM, the CMD provided for the following scheduling priorities in clearing supply and demand and managing congestion: 1) supply and demand associated with ETC schedules; 2) self-scheduled demand associated with CRR schedules and self-scheduled supply from must take and must run resources (the portion with must take or must run status); 3) any other self-schedules; and (4) supply and demand with energy bids.[66]

Cal ISO clarified that the priority for CRRs was intended only for the load end, not the supply end, to avoid interfering with congestion management.[67] This approach would also avoid discrimination in generation market but obviously favors load with CRRs if the price rises to the price cap. FERC rejected the proposed scheduling priority for CRR holders because of a potential for discrimination in the generation market but did not explain what this potential was in light of the Cal ISO's clarification.[68] With this modification, all non-ETC self-schedules are on the same footing in the DAM. In particular, internal load and export self-schedules will have the same priority in the DAM.[69]

The Cal ISO will not reserve any internal transmission capacity for ETCs in the DAM beyond the capacity used by their day-ahead schedules.[70]

D. Virtual Bidding

The CMD did not explicitly provide for virtual bidding. It implicitly allowed a degree of virtual bidding because external loads could bid into the DAM and external supplies could be offered in the DAM and then zeroed out in the hour-ahead scheduling process. FERC agreed with intervenors regarding the benefits of allowing virtual bidding but initially did not require the Cal ISO to implement virtual bidding.[71] Concerned with the impact of the proposed lack of virtual bidding, FERC subsequently directed implementation of virtual bidding in the DAM, but allowed the possibility that virtual bidding could be implemented at a later date rather than upon start-up of the LMP market design.[72] It is understood that the Cal ISO currently favors a virtual bidding design whereby virtual load and supply bids will be zonal, but to date has not filed a specific virtual bidding proposal.

While implementation of virtual bidding is an important and desirable feature of the MRTU, its contribution to price convergence and market efficiency would be vitiated by the proposed nodal clearing of zonal bids and reaggregation. If virtual load/supply bids are cleared nodally like the zonal/LAP physical load bids as is being considered, all market participants will be able to arbitrage the zonal/LAP pricing system in the manner discussed in Section B4 above,

---

66 CMD # 33.

67 Oct FERC ¶ 182.

68 Oct FERC ¶ 184, 185.

69 This would change under current resource adequacy proposals, which would require self-scheduled exports to be supported by resources that are not committed to the Cal ISO resource adequacy program.

70 CMD # 66.

71 Oct FERC ¶ 151.

72 See June FERC ¶159, Sept FERC ¶73-76. The FERC order regarding virtual bidding is good in many respects. In particular, it is not clear that there still a need for availability bids for RUC units with virtual bidding.

which could be even more expensive for consumers than the basic LAP pricing system. Market participants would be able to do this by submitting low price capped virtual load bids in the DAM that would clear only at the low priced nodes in the LAP. In the settlement system, these virtual bids clearing only at the low-priced nodes would be converted to schedules cleared at both the low-priced and high-priced nodes. In real time, the market participant would sell back the power they bought in the DAM at the low-priced nodes at the higher real-time zonal/LAP price reflecting the higher real-time price at the constrained nodes. Market participants could also submit high priced virtual supply offers that would clear only at the high priced nodes but would then be spread to all nodes for the purpose of determining the real-time price. If the DAM price at the high priced nodes materially exceeds the average zonal real-time price, the market participants would make money from these virtual supply offers. Neither kind of virtual bid would provide the kind of DAM/real-time price convergence that FERC likely intended with its order. This is another piece of the first major implementation issue discussed above. This potential is illustrated in Appendix III.

The Cal ISO has also indicated that it is considering restricting virtual demand and supply bids to price taking offers. Such a restriction would avoid the kind of uplift creating arbitrage of the LAP pricing system described above, but it would also make virtual demand and supply bids useless for normal price converging arbitrage. In fact, it would require that market participants submit precisely the kind of inflexible virtual supply and demand bids that the NYISO does not permit out of a concern that such bids can never be motivated by legitimate arbitrage opportunities.

The problems discussed above are not inherent in virtual bidding but arise from the proposed clearing mechanism for zonal/LAP bids which is workable neither for physical nor virtual load bidding. If zonal/LAP virtual load bidding is permitted and cleared zonally as in New York and PJM, the uplift problems described above are avoided and virtual bidding will tend to provide improved price convergence between the day-ahead and real-time markets, improve market efficiency, and avoid a wide variety of problems in applying market power mitigation that arise when day-ahead prices do not reflect expected real-time prices. In addition, virtual load bidding will tend to produce a better relationship between bid load and real-time load, reducing the need to commit generation in the RUC.

The proposed nodal clearing process for zonal/LAP DAM load bids will likely combine with the broad LAP definition to undermine the effectiveness of virtual bidding in arbitraging differences between day-ahead and real-time prices, which will, in turn, adversely impact participation of generators in the day-ahead market. This is the second of the eight major implementation issues we have identified with the MRTU market design. As explained above, the nodal clearing prices for zonal/LAP bids will tend to cause zonal/LAP bids to clear only at the low-priced nodes within the LAPs. With little or no load clearing in the DAM at prices high enough to commit or dispatch high-cost generators in constrained regions within the LAP, DAM prices in the constrained regions will likely be well below real-time levels.

In Eastern LMP markets, generators and other market participants observing such a discrepancy can arbitrage the difference between DAM and real-time prices by submitting zonal (or, in PJM, nodal) virtual bids, so that the generator effectively is paid the real-time zonal price for its generation. Thus, the generator sells its output in the DAM at the DAM price but also buys power in the DAM through its virtual load bid. It is then paid the real-time price for the power it bought in the DAM. If nodal virtual bidding is permitted or if the load zone price is very similar to the generators' nodal price, virtual bidding allows generators to effectively sell their output at real-time prices while still committing their generation in the DAM. Such arbitrage also serves to bring DAM and real-time nodal prices together.

It appears that this kind of efficient arbitrage would not be feasible under the MRTU market design because the LAP zones are so large that the LAP price will bear little relationship to many nodal prices within the LAP zone. Generators seeking to ensure that they are paid the real-time LMP price for their output could therefore not accomplish this through virtual load bids at the LAP nor would the DAM nodal price reflect the real-time nodal price, because of the nodal clearing process. It should therefore be anticipated that not only will high cost generation in constrained areas not be scheduled in the DAM, even the low cost infra-marginal generation will not be scheduled in the DAM but will show up in real-time. These incentives will complicate the Cal ISO's RUC analysis, as discussed further in Sections III and IX.A.

The normal response of competitive suppliers to nodal DAM prices that are lower than expected real-time prices would be to offer their output into the DAM at offer prices reflecting expected real-time prices. Since the offer prices of such an infra-marginal generator would likely be mitigated in Passes 1 or 2 below the expected real-time price, it should be anticipated that infra-marginal non-RMR generation located within constrained regions would simply not offer its capacity in the DAM unless required under a resource adequacy contract.

If the LAP bid structure were retained but a conventional zonal clearing mechanism adopted for the zonal LAP bids, then arbitrage of by market participants of differences between the LAP price and the expected real-time price would not only drive together the DAM and real-time LAP prices but would also drive together the DAM and real-time nodal generation prices. Under a zonal clearing process in which load weights determined by the ISO were used to assign load to individual nodes but the zonal bids clear against the zonal price or not at all, arbitrage of differences between the DAM and real-time LAP would tend to drive DAM and real-time nodal prices together as well, as long as the load weights used by the ISO were consistent with market expectations. Thus, all of the potential problems with virtual bidding, as well as limitations on the effectiveness of virtual bidding in providing price convergence, can be avoided by clearing zonal load bids zonally.

E. External Schedules

Although the CMD proposed that the external network model will ultimately be a "closed loop" model that represents external electrical connections between the various inter-ties into the ISO control area, and thus allows the ISO to explicitly estimate and manage parallel path or 'loop flows' in coordination with other control areas in the region, apparently this is not planned for the initial implementation.[73] Instead, the Cal ISO apparently proposes to use a simpler "open

[73] CMD # 29, 30.

loop" representation of the external network, until such time as there is an effective coordinated western regional framework for day-ahead scheduling and congestion management."[74]

It appears to us that there are two pricing/scheduling issues regarding imports that need to be disentangled. First, the Cal ISO must continue to coordinate transaction scheduling on a contract path basis with adjacent control areas. Since these adjacent control areas will not schedule transactions in excess of the contract path limit, the contract path schedules submitted to the Cal ISO for check out that exceed these limits will not flow. Second, however, the Cal ISO needs to model the impact of the actual power flows resulting from import and export schedules on its internal transmission constraints. These actual impacts may not be related to the nominal contract path for these external schedules.

Consider three different ways of modeling these impacts and their implications.

Pure Contract Path.

Under the Pure Contract Path approach the Cal ISO would enforce the contract path scheduling limits on each "path" in the DAM, treating each path as a separate source and would also calculate path specific impacts of power flows on each contract paths on internal constraints, calculated as if that contract path were an open loop.

This approach addresses the contract path scheduling limits and ensures that they are not exceeded but has two problematic features. First, whether a transaction would clear in the Cal ISO DAM can depend on the contract path selected by the seller. For example, market participants might submit more import offers than could be scheduled over the LC1 contract path at the same time that the LC4 contract path was nearly empty. This outcome could be avoided if sellers were to schedule firm transmission in the neighboring control areas prior to offering the energy for sale in the DAM, but sellers may be reluctant to incur the costs of scheduling firm transmission into California prior to selling power in the DAM.[75] Alternatively, sellers that are unable to schedule imports in the DAM due to congestion on a particular path could nevertheless purchase transmission and schedule imports on alternative paths for delivery in real-time, but sellers might be unwilling to commit capacity day-ahead to support such real-time sales without locking in a day-ahead price. This need for day-ahead contracts to support unit commitment and

---

74 See CMD # 30. The rationale for this is that "forward scheduling in a manner that accounts for loop flows may create severe problems if the ISO were to adopt this feature ahead of its neighbors throughout the west." CMD # 30. We understand that the Cal ISO's interchange scheduling must respect the WECC's contract path scheduling practices and further that it is difficult under these scheduling practices for the Cal ISO to accurate assess the impact of changes in external schedules on the actual flows on internal Cal ISO transmission lines. Nevertheless, it seems to us that appropriate closed loop modeling by the Cal ISO for the purpose of modeling the impact of external schedules on internal transmission constraints would improve reliability even if not followed by other western control areas. The Cal ISO could continue to accept and check out contract path schedules as it does today; the Cal ISO would simply use a closed loop model to analyze the impact of these schedules on flows within the Cal ISO control area. Even an approximation of the closed loop effects would be better and would not be difficult to implement. If the problem is that the closed loop approach would produce results substantially different from the open loop calculation, that would reinforce the value of using something that is approximately right.

75 Market participants could submit schedules on multiple paths to ensure that some clear, but this approach carries the potential for multiple transactions to clear.

transmission scheduling can also be met under the MRTU design, if LSEs have the appropriate incentives to enter into such contracts. Such an inability to schedule transactions in the DAM due to an inability to guess the available scheduling path could reduce the supply of price-sensitive imports either day-ahead or in real-time.

Second, modeling the impact of import schedules on internal transmission constraints based on the scheduled contract path would permit market participants to schedule transactions on the path modeled as having the most favorable impact (the lowest congestion impact or the largest counterflow impact) on these internal constraints, and thus the highest price. This is efficient if the transactions on the various contract paths actually have different sources and thus different impacts on the internal constraints, but this is not efficient if the transactions on the various contract paths have the same actual source, and are merely scheduled differently to exploit the contract path fiction. Thus, while the cost of scheduling transmission makes it likely that imports scheduled on contract paths from the Pacific Northwest actually have a different source than transactions scheduled on paths from the Southwest, schedules on many of the contract paths from the Desert Southwest could all have exactly the same generation source and exactly the same impact on internal Cal ISO transmission constraints, regardless of the contract path identified for scheduling purposes.[76]

Combined Contract Path and Interface.

Under the Combined Contract Path and Interface approach the Cal ISO would enforce the contract path scheduling limits on each path in the DAM, treating each path as a separate source for scheduling and thus enforcing contract path scheduling limits but could calculate the impact of power scheduled on multiple paths as having the same impact on internal constraints. Thus, for example transactions scheduled on the contract paths LC1 and LC4 could be modeled as having the same source for analyzing the impact of those schedules on internal Cal ISO transmission constraints. This approach would have the same limitation as the first approach in potentially limiting the supply of imports in the DAM and thus raising consumer prices. It would, however, reduce the extent of the second problem, treating contract paths with identical sources as identical for pricing purposes. This would reduce the potential for inefficient scheduling practices that raise consumer costs.

We have not assessed the magnitude of the potential benefits of such modeling changes, which could be small if congestion impacts on internal transmission constraints as they are modeled under the pure contract path approach are in practice very similar or if the internal constraints are rarely expensive to solve with internal redispatch.

Pure Interface Approach

Under this approach the Cal ISO would define interfaces with similar generation sources and model all transactions scheduled on these interfaces as having the same impact on internal constraints. In addition, rather than enforcing individual contract path limits in the DAM, the Cal ISO would only enforce the total contract path limit, i.e., the sum of the schedules on the

---

[76] These considerations are discussed at greater length with examples in a paper prepared for the NYISO and ISO-New England, Scott Harvey, "Proxy Buses, Seams and Markets," May 23, 2003.

interface could not exceed the combined contract path limits for all of the contract paths included in the interface.[77] Thus, the Cal ISO would continue to require contract path scheduling in accord with WECC practices for transactions flowing in real-time, but would aggregate some paths for the purpose of clearing price capped import and export bids in the financial day-ahead market (which entails analyzing the impact of those transactions on internal constraints). This approach would facilitate market participants selling power in the DAM and then identifying the available contract path subsequently and scheduling firm transmission on that path to support the day-ahead financial schedule.

It is our understanding that the Cal ISO proposes to adopt the first approach under the CMD, which is basically the same approach employed for scheduling external transactions today. This approach has the limitations identified above, but we have not examined how significant these limitations are in practice.

External resources will not submit start-up cost or minimum-load bids,[78] but will be able to submit price-capped energy offers that can set prices in the DAM. The FERC approved Cal ISO proposal to permit imports to set market clearing prices in DAM.[79]

It is our understanding that under the MRTU, export schedules in the DAM do not need to be supported by a specific resource and will compete with California load on an equal footing in the DAM. Importantly, exports scheduled in the DAM will be treated on the same basis as control area load in the hour-ahead process. Suppliers therefore have no need to withhold generation from the RUC in order to support DAM exports and incur no export-related opportunity costs when unscheduled generation is committed on the RUC.

---

77 According to the Cal ISO, the DAM network model does include loops that go from an internal node to a boundary node. Thus an import schedule at the Palo Verde scheduling point flows on two paths, one from Palo Verde to Devoirs (SCE territory) and the other from Palo Verde to North Gila (SDG&E territory). The import limit at Palo Verde is on the net flow on the two branches together. Thus the impact of a schedule at Palo Verde on internal transmission is modeled along the lines mentioned here. However, the actual source of the import may not be the Palo Verde Nuclear units, and for all practical purposes may be in a third control area (if the source is in a third control area, the neighboring control area would require transmission reservation to Palo Verde from a boundary location with that third control area)

78 CMD # 18.

79 Oct FERC ¶ 90.

III. DAY-AHEAD RESIDUAL UNIT COMMITMENT

A. RUC Structure

The purpose of the day-ahead Residual Unit Commitment (RUC) is to ensure that sufficient resources are committed to reliably support the Cal ISO load forecast if the load that clears in the day-ahead market differs from the Cal ISO's forecast load.[80] This circumstance could arise as a result of LSEs underestimating real-time load, the Cal ISO mistakenly over-estimating real-time load, strategic behavior by LSEs seeking to impact the day-ahead price, LSE bidding or scheduling errors, or expectations that additional low-cost imports will be available in real-time. All bids in the day-ahead market will automatically roll over into the RUC. There will be no additional bid submission for the RUC. The RUC will be completed by 1:00 p.m. and the results published at 1:00 p.m.[81] Resources that do not participate in the day-ahead market will not be eligible to participate in the day-ahead RUC, but can participate in the hour-ahead scheduling process.[82] Resources may be issued commitment instructions in either the day-ahead RUC or the hour-ahead scheduling process, depending on their required start-up times.[83]

Units committed in the RUC will be selected based on system reliability on a nodal or local area basis. Local area needs will continue to be met with RMR resources, but non-RMR resources that participate in the DAM could be committed and dispatched to address local reliability needs in Passes 3 and 4 if they are the least-cost alternative.[84]

It is also our understanding that capacity does not need to be withheld from the RUC in order to support export schedules. Export schedules will clear in the DAM based on price. DAM export schedules will then flow in real-time.[85]

The three-part bids submitted in the DAM IFM will be used in the RUC process, including start-up cost, minimum-load cost and incremental energy bid curve as submitted in DAM, and as modified by bid mitigation.[86] With the change in the RUC decisions and objective function to scheduling capacity to minimize commitment costs only, the incremental energy bid curve would be modified in the RUC pass such that the energy offer prices of resources dispatched in Pass 3 or identified as meeting resource adequacy requirements would be set to

---

80 CMD # 8.

81 See CMD # 74. The bids are due in at 10:00 a.m., and Passes 1A, 1B, 2, 3A, 3B and RUC are all to be completed and posted by 1pm. As noted above, this is an aggressive schedule.

82 See CMD # 104 and May ISO, pp. 5-6, 79. The CMD originally proposed that all resources subject to Must Offer Obligations would be required to participate in the RUC procedure. CMD # 100. This was not approved by FERC. Under the FERC approved FOO discussed in Section VII, units that do not participate in DAM and RUC are obligated to be available in real-time, subject to the waiver process.

83 See CMD # 99. The CMD proposed that the SCUC used in the day-ahead RUC would utilize a multi-day time horizon. CMD # 60. This has been dropped and an informal process will be used to commit long-start-up time units to the extent this is necessary.

84 See also CMD # 110.

85 This would likely change with implementation of a California resource adequacy program as discussed below.

86 CMD # 105.

zero, while the energy offer prices of other resources offered in the DAM would be replaced with their RUC availability bids.[87]

Under the MRTU, the constraints such as minimum-load energy and minimum run time specified by suppliers must be the actual physical constraints of the resource not market based constraints. Import offers may not be resource specific and may not have start-up or minimum-load cost bids, except in the case of resources dynamically scheduled into the Cal ISO control area.

Units committed in a particular hour in the DAM would be treated as on-line during that hour in the RUC, so there would be no start-up or minimum-load costs associated with energy procured from those units. The commitment of units in the RUC for hours they were not scheduled in the DAM, will take account of those units' minimum-load costs in the additional commitment hours.[88] Load resources that are certified dispatchable in real time can commit to a real-time dispatchable reduction at a specified price per MW and be assured of recovery of start-up and minimum-load costs.[89]

This proposed structure of the Cal ISO RUC process under the MRTU appears to be workable and largely consistent with the similar processes that have been implemented in PJM and NYISO. In the discussion of MRTU, there has been a suggestion that the DAM and RUC unit commitment process be solved simultaneously. Such a simultaneous solution would in principle provide a lower cost unit commitment than the sequential approach employed by other ISOs. Development of such a simultaneous solution process is possible, in principle, and was considered in the early development in New York. The New York Power Pool (NYPP) and ABB, its software vendor, originally thought they would address the New York RUC through procurement of reserves and that it was simply a matter of increasing the reserve targets in the unit commitment so that the energy cleared in the DAM plus reserves equaled forecast load. At a late stage in software development, however, it was recognized that simply increasing the reserve target would not satisfy reliability criteria. The problem is that the standard security analysis logic for reserves verifies that reserves can be dispatched to meet load in the event of the relevant generation or transmission contingencies. The proposed reserve scaling logic would not provide assurance that a portion of the reserves could be dispatched to meet underscheduled energy demand and then other reserves dispatched to meet contingencies.

Directly solving the problem appeared to require the development of new algorithms in order to simultaneously optimize the unit commitment against bid load and forecast load. The structure of the New York RUC as a separate pass is a feasible solution to this problem that was developed within time constraints that foreclosed reliance on development of as yet unknown algorithms. By running a separate unit commitment and dispatch for the bid load pass and the forecast load RUC, the NYISO is able to enforce reliability criteria in both passes using standard security analysis software.

---

87 Resources committed to meeting the load of Cal ISO LSEs under resource adequacy contracts would not submit availability bids nor would they be paid an availability price.

88 CMD # 105.

89 CMD # 127.

The combined problem was formulated in a subsequent paper.[90] This paper demonstrated how the financial and reliability problems could in principle be jointly solved, but this formulation requires simultaneously solving and committing generation based on two representations of the network contingency constraints, one for the bid-in conditions and another for the ISO forecast. The paper further observed that solving the expanded problems appeared likely to require a major change in the tools and practices used to solve the unit commitment problem.[91]

While in principle PJM, ISO-NE and MISO and their software vendors all had time to try to develop new algorithms that would solve the problem more efficiently in a single step, it is instructive that in the end they all have chosen to use the sequential logic structure developed with the NYPP. Unless the difference between bid load and forecast load is sufficiently large to require the commitment of slow starting units day-ahead, there would be no cost savings in the unit commitment or the scheduling of ancillary services[92] from a jointly optimized solution. Absent such unit commitment savings, there would be little return to incurring the costs and delays associated with development of software to produce such a simultaneous solution. It should also be understood that were such a combined model to be implemented, Cal ISO load forecast errors could potentially have a material impact on day-ahead market prices, even if there were no reserves shortage or high prices in real-time. While there are considerations that would tend to favor allowing such Cal ISO load forecast errors to impact DAM prices, there are also considerations that might argue against this, particularly if RUC availability bids are permitted.

Overall, basing the Cal ISO MRTU design and implementation on the development of the algorithms and software required to handle this joint optimization of unit commitment would be a risky implementation strategy for uncertain and possibly modest improvements. Unless there are features of the Cal ISO resource mix or regional scheduling practices which analysis indicates would materially impact the efficiency of a sequential RUC process, it would be preferable to implement standard industry software and then assess whether there are material potential cost savings from developing more complex software. If the market performs well, bid load may be sufficiently close to forecast load that there would be little capacity committed in the RUC pass and therefore little potential for cost savings from a better RUC commitment requiring the development of improved software.

Another alternative approach to the day-ahead market that would entirely avoid the need for a RUC process would be to place responsibility for meeting load on the ISO, rather than the individual load serving entities. A fundamental premise underlying the structure of the Eastern RUC processes and similarly the structure of the proposed MRTU process is that it is the load

---

90 Michael D. Cadwalader, Scott M. Harvey, William W. Hogan and Susan L. Pope, "Reliability, Scheduling Markets and Electricity Pricing," May 1998.

91 Cadwalader, et al., 1998, p. 9.

92 As discussed further in Section VI below, the MRTU market design has the feature that ancillary services are scheduled day-ahead in Pass 3, which occurs prior to the RUC commitment. If additional capacity is often added in the RUC pass, the ancillary service schedules determined in Pass 3 would potentially not be the least cost solution in real-time. A simultaneous RUC and energy market solution could avoid these costs but they could also be avoided by employing a New York style DAM structure in which energy and reserves schedules are determined following the RUC commitment.

serving entities that are responsible for forecasting load and scheduling resources to meet this load. An alternative approach would be to place the responsibility for scheduling the day-ahead load of each LSE on the ISO, with the LSEs being responsible for the financial consequences of over or under forecasts by the ISO. This is a fundamentally different approach than that on which the Cal ISO and Eastern RUCs are based and would require fundamental policy decisions by the CPUC and others to implement.

B. RUC Target

The CMD originally proposed that the capacity target for the day-ahead RUC would be the next day's hourly load forecast plus reserve requirements, minus (1) final day-ahead schedule of energy plus ancillary service capacity; (2) a forecast of expected incremental hour-ahead schedule changes; and (3) a forecast of additional supplemental energy bids expected on the operating day. The ISO was to fine tune this estimation procedure to minimize over and under procurement. To the extent that metered subsystems within the ISO control area under-schedule in the day-ahead market but have designated adequate resources under their control to meet their own load and reserve needs, the RUC would not provide capacity to cover their share of the load forecast nor would RUC costs be allocated to them.[93]

This capacity target has since been revised to compare the Cal ISO's load forecast and estimated reserve requirements to DAM schedules and self-provided RUC.[94] In its May 2004 Comments the Cal ISO further provided that it would only schedule capacity, not energy, in the RUC.[95] This proposal was accepted by FERC.[96]

It is currently proposed in the MRTU that the Cal ISO will not buy import energy in the day-ahead RUC, only capacity.[97] This approach is more restrictive than that taken by the Eastern

---

93 CMD # 101.

94 The Cal ISO originally proposed that energy would be scheduled in the RUC, capped at the difference between the load cleared in the DAM and 95 percent of the next day's hourly demand forecast. Any remaining difference between the load cleared in the DAM and the ISOs load forecast would be covered by the unloaded capacity of units scheduled in the day-ahead IFM and RUC. CMD # 102. FERC accepted the Cal ISO RUC capacity target, but rejected the proposal to procure energy in the RUC. The FERC left open the possibility of the Cal ISO buying import energy in the RUC. Oct FERC ¶ 127.

95 May ISO, pp 3, 53, Att A II.10.

96 The elimination of energy procurement in the RUC obviated the need for a number of provisions in the CMD relating to the scheduling priority of RUC generation. "Minimum load energy from internal resources and energy procured from interties in the Day ahead RUC will have scheduling priority over incremental hour ahead energy schedules in the hour ahead IFM." See CMD # 103, June FERC ¶ 51.

Any energy procured in the day-ahead RUC i.e., the minimum-load energy of internal resources committed by RUC as well as energy procured form import suppliers will be submitted hour ahead as price taker. CMD # 109.

Any RUC energy not scheduled in the hour head market was well as additional hour-ahead RUC will be submitted as a price taker in real-time and will be eligible for RUC uplift. CMD # 109.

These provisions would have been problematic in the event that real-time load was less than the ISO's load forecast and the CMD proposal has been strengthened by eliminating them.

97 See May ISO, pp. 53-54. The CMD originally provided that the RUC would schedule energy from import suppliers, provided adequate capacity is available on the interties to accommodate the energy after the running

ISOs and there is a potential for this restriction to be problematic. Rather than thinking of this as an energy versus capacity issue, one should think of imports as units with a minimum-load block equal to their capacity. From this perspective, it is also apparent that there would be no discrimination for or against imports in using such an approach, as the treatment would be analogous to the treatment of internal units whose minimum-load block equals their capacity. Physical constraints determine the minimum-load blocks of internal units. External scheduling constraints that require advanced commitment to the scheduled amount have the same effect for imports. Like the minimum-load block of an internal generator, the Cal ISO could evaluate the scheduling of imports to meet the RUC target looking at the total energy cost of the imports. Imports would therefore normally be a very expensive way to schedule capacity, as obtaining the capacity requires paying the minimum-load cost on all of the capacity. Nevertheless, the scheduling of imports in particular hours could avoid the commitment of a steam unit with a substantial minimum run time that could cost more than the imports. While the Eastern ISOs rarely commit imports to solve forecast load imbalances, it does occasionally happen because it is cheaper than the alternative. Why should the Cal ISO impose such a restriction on consideration of imports if it raises the cost of the RUC commitment? Given the need to schedule transmission service for imports from much of the WECC, such imports may at times be more difficult to schedule in the hour-ahead process than day-ahead.[98] Because the scheduling of imports will likely typically not be a least-cost source of RUC capacity, this restriction probably would not have a substantial impact on RUC costs in the near term.

It was originally proposed that the RUC's objective function would be to minimize the total bid cost of procuring resources, including a bid based availability payment, and dispatching them for real-time energy to fully meet the real-time load forecast.[99] This objective was revised to the RUC minimizing the commitment costs for the capacity required to meet the Cal ISO load forecast.[100] This revision to only take account of commitment costs is appropriate and consistent with the approach of Eastern ISOs, and is intended to provide better incentives for day-ahead bidding.[101]

---

of the financial DAM. CMD # 103. This was dropped with the change in objective function (May ISO, pp. 40, 53-54).

98 It is not clear under the MRTU design whether generation scheduled in the RUC incurs any financial obligation analogous to a day-ahead schedule. Thus, if a generator is scheduled to operate at minimum-load in the RUC and subsequently trips off line, is the generator obligated to buy back energy equal to its minimum-load block in real-time. There is such an obligation in New York, because the minimum-load blocks of units added to meet forecast load are included in the DAM settlement. That is not the case in the proposed Cal ISO market design, so the obligation is unclear. Whether such an obligation exists is relevant to the discussion of availability payments below.

Such an obligation would appear to be necessary if the RUC were modified to include the scheduling of imports, as absent an obligation to perform, import suppliers could choose to deliver power in real-time only when power can be purchased in the hour-ahead scheduling timeframe at a price lower than their DAM offer price.

99 CMD # 105.

100 May ISO, pp. 40, 53, Att A II.10.

101 As noted in section II above, this change in the objective function was accompanied by the elimination of restrictions on the set of units available for commitment in Passes 3 and 4, so that the commitment in passes 3 and 4 is no longer limited to those committed in Pass 2. Retention of those restrictions would make the objective function in Pass 4 large3ly meaningless and would likely lead to unintended outcomes in terms of

C. Availability Bids

The CMD provided that unloaded capacity scheduled in RUC would receive a per MW availability payment for each MW of RUC procured capacity that was not awarded ancillary services or dispatched for energy in the hour-ahead or real-time. Resources could submit availability bids of up to $100/MW hour.[102] FERC rejected the $100/MW cap on availability payment, setting the cap at $250/MW.[103] The CMD also proposed that the RUC availability payment would be rescinded for each MW of RUC procured capacity that is scheduled or dispatched for energy or awarded ancillary services capacity in a subsequent market or if the unit is not available in real-time.[104] FERC rejected the proposal to rescind availability payment when the unit is dispatched.[105] The Cal ISO has indicated a preference to rescind the payment but has agreed to forgo this.[106] The resource will, however, lose its availability payment if it does not perform, i.e., becomes unavailable.

The CMD further proposed that the RUC availability payment would be made on an as bid basis and would be included in costs in calculating the bid production cost guarantee (BPCG).[107] The FERC ordered that the availability bids should set a locational market clearing price, rather than being settled on a pay as bid basis.[108] The Cal ISO agreed that RUC payments would be determined on a locational market clearing basis.[109] This locational RUC pricing of availability bids was approved by FERC.[110] It is not clear to us that a locational payment is necessary, given the possible competitive reasons for an availability payment. For example, if the availability payment compensates for the potential difference between DAM and real-time gas prices, why should this vary locationally based on electric transmission constraints, particularly since the transmission constraints would be reflected in the real-time electricity prices paid to the supplier. This issue is discussed further below.

Under the MRTU, RMR units will not be eligible to set or receive availability payments in RUC.[111] Self-Provided RUC will not be eligible to set or receive availability payments in RUC.[112] The procurement of RUC capacity and payment for availability would also be limited to the portion of capacity dispatched in the RUC Pass (Pass 4) that is on units with a start-time of

---

very high RUC availability prices. Since Passes 1 and 2 do not consider RUC availability bids in committing units (they currently dispatch generation to meet forecast load based on energy offer prices) restricting the units considered in pass 4 to those committed in pass 2 could leave the Cal ISO with no alternative to accepting very high RUC availability bids in pass 4, even when other units offered RUC capacity at much lower prices.

102 CMD # 107.

103 FERC also rejected the Cal ISO's proposed to cap RUC availability bids at $150 with a maximum energy plus availability payment of $250. May ISO, pp. 3, 26, 37, 42 Att A II.1, 4; June ISO, pp. 6-7. See also Oct FERC ¶ 123; June FERC ¶ 65, 66.

104 CMD # 107.

105 Oct FERC ¶ 123; June FERC ¶68; Sept FERC ¶ 30.

106 May ISO, p. 40-44.

107 CMD # 107.

108 Oct FERC ¶ 123.

109 May ISO, pp. 19, 38 Att A II.2.

110 June FERC ¶ 75.

111 May ISO, p. 40 Att A II.9.

112 May ISO, p. 40 Att A II.9.

five hours or more that must be committed in the day-ahead timeframe.[113] With the evolution of the RUC structure and availability bid, there is some ambiguity as to whether GTs and other quick starting units (hydro capacity?) would generally be eligible for the availability payment.[114]

With this background, the third major implementation issue we have identified with the MRTU market design concerns the role of the RUC availability bid. What are potential rationales for paying an availability charge for RUC capacity in addition to start-up and minimum-load costs and/or the real-time market price of power?[115]

1. If a unit were unable to sell into higher priced export markets as a result of providing RUC, such an availability payment as compensation for this opportunity cost would be warranted. We understand, however, that energy exports can be scheduled in the DAM without committing a unit and such exports are not subject to recall in RUC. Thus, there are no foregone energy export revenues as a result of committing a unit in RUC.[116] Generating capacity that is scheduled to provide RUC capacity therefore does not forgo any capacity value, because if the capacity value were expected to be greater outside California than inside California the capacity owner could merely bid its generating capacity into the Cal ISO market and self-schedule an export, assuring its ability to sell power outside of California in real-time if prices are higher outside California. It is important from a resource adequacy standpoint that it be understood that the proposed RUC design does not ensure that the Cal ISO will have adequate resources to meet the load of Cal ISO LSEs during a regional reserve shortage. On the contrary, as discussed further in Section VI.C below, the low level of the damage control bid cap practically ensures that California LSEs will bear the brunt of regional capacity shortages. The value of the RUC commitment is simply to ensure that LSE bidding errors do not result in situation in which capacity is available, i.e., there is no capacity shortage, but too little capacity is committed day-ahead, leading to a shortage in real-time that could have been avoided.

2. If units scheduled in RUC could incur gas costs that could not be recovered in real-time energy bids as originally proposed by the CMD, then an availability payment would be warranted. However, FERC has rejected the proposed restriction on bid

---

113 CMD # 99 and # 107.

114 GTs and other quick starting units would therefore need to incur manning costs and be available all the time. Presumably they would be able to use the waiver process to avoid these costs when they were not needed. DAM market operators would have discretion to issue day-ahead RUC commitment notices for quick-start units.

115 Another way of posing this question would be to ask why would a firm lacking market power require an availability payment in return for providing RUC capacity?

116 Similarly, if a seller were unable to offer capacity in subsequent RUC markets outside California as a result of committing the capacity in the California RUC market, there could be opportunity costs for commitment in California. Aside from the uncertainty as to whether any such markets will exist in the near future, if such a seller expected that a another market would be short in real-time, it would simply submit a bid to buy power in the Cal ISO DAM for export and then schedule this power for delivery into the other market in real-time, being paid the high real-time prices if the seller's expectation of a real-time capacity shortage were correct.

increases between day-ahead and real-time, so no such unrecovered costs arising from RUC status should exist.[117]

3. If the calculation of start-up and minimum-load costs did not cover the actual start-up and minimum-load costs, then an availability payment would also be warranted. The CMD proposal has some features concerning the calculation of the bid production cost guarantee that may tend to result in such a revenue shortfall for units committed in the RUC (rules regarding gas prices, the failure to take account of NOx allowance costs, etc.). But, this is not a good reason for including an availability payment. The use of availability bids to recover such costs leads to problematic outcomes when applied on a market-clearing basis. It would be better to simply appropriately compensate suppliers for their start-up and minimum-load costs and eliminate the need for the availability payment.

4. Gas turbines that need to be manned in order to be available for dispatch in real-time would incur these manning costs if they are identified as needed in the RUC, but if they are never dispatched in real-time they would earn neither energy nor reserve revenues.[118] This is a reasonable rationale for making an availability payment to such units, but it is unclear whether GTs would receive a day-ahead availability payment under the CMD. The CMD provision that day-ahead RUC would not issue commitment notices to quick starting units appears to eliminate RUC availability payments to these units and also may have the consequence that these units would not always be available in real-time when needed because the units might not be manned or the lack of a RUC DAM schedule would create an opportunity for undertaking minor maintenance that could make the unit unavailable for a few hours. It would appear preferable to clarify that all units that are identified as needed to meet forecast load in the day-ahead RUC dispatch (including GTs and hydro capacity) will receive notice that they will potentially be needed to meet real-time load. This would entail either making the units eligible for a RUC day-ahead availability payment that would recover manning costs, or some alternative cost recovery mechanism such as to provide GTs scheduled in the RUC with a bid production cost guarantee for these costs analogous to the bid production cost guarantee other units receive for start-up and minimum-load costs.

5. During periods in which winter gas balancing rules are in effect, it may be prudent from a gas and electric system reliability perspective for generators to buy gas in the day-ahead market to cover RUC schedules. If the ISO's load forecast is inaccurate,

---

117 June ISO, p. 6. This would be a bad rationale for an availability payment in any case. Why force suppliers to guess the difference between day-ahead and real-time costs. The difference is obviously not known at the time bids are submitted as the day-ahead price is the expected clearing price. How could the Cal ISO fairly mitigate such bids?

118 Some comments reflect an expectation that units dispatched in the RUC will almost certainly be committed in real-time. This is not necessarily the case and is particularly likely to be untrue for GTs which are high cost and need not be committed unless they are actually needed. Real-time load can exceed bid load while still falling short of the load forecast and leaving high cost units undispatched. Moreover, higher load in real-time may stimulate imports of additional power at prices lower than GT dispatch prices, yet the availability of the GTs may be essential in order to maintain reliability in the event additional imports are not available in real-time.

these forward purchases could be uneconomic, imposing costs on the resource supplier. The payment of an availability charge to compensate for these costs would be appropriate. As discussed below, an alternative approach to a RUC availability payment would be to make resource suppliers eligible for BPCG in the event that there were losses for this reason.

6. Exports scheduled day-ahead would pay DAM congestion charges while exports scheduled in real-time would pay real-time congestion charges. The need for units subject to the RUC to schedule exports in the DAM would require that they pay DAM congestion charges on exports in circumstances in which the California ISO was likely to be short of resources in real-time but prices were expected to be higher outside the California ISO control area. It is hard to conceive of circumstances in which these things are all simultaneously true. In particular, why would export congestion be expected to be higher in the DAM than it is expected to be in real-time, if prices are expected to be higher outside California in real-time.

7. Underbidding in the DAM by LSEs might depress DAM prices relative to real-time prices, but this would not impose any opportunity costs on units committed in the RUC because they would sell their output at the high real-time prices. Moreover, it does not make sense to attempt to reflect any such scarcity in RUC availability payments. All capacity scheduled in the DAM helps meet forecast load, so why should only the capacity scheduled in the RUC receive a scarcity payment? In addition, if the RUC does not include imports (see the comments above), the availability price could be high even when there is no real capacity shortage. More fundamentally, under the proposed MRTU market design any market participant that expects shortage conditions to exist in real-time can ensure that it earns the real-time price by submitting virtual load bids in the DAM.[119] It is possible that the ISO's load forecast could reflect shortage conditions that do not prevail in real-time, but it does not appear appropriate to compensate sellers for shortage conditions that exist neither in day-ahead financial markets nor in real-time but exist only in mistaken ISO forecasts.

8. If units scheduled in the RUC incurred a financial obligation to perform (i.e., to buy back some portion of their RUC energy at the real-time price if they failed to deliver), then an availability payment might be appropriate to provide compensation for incurring this financial obligation, but it does not appear to us that the CMD imposes such an obligation.

Not only are most of these rationales for an availability payment either irrelevant or of doubtful merit, they are inconsistent with the proposed mitigation mechanism, which takes no account of these kinds of costs. This is discussed further in Section F below.

The FERC September 2004 rehearing order refers to the RUC availability payment as "compensating" for the foregone opportunity to sell their product in a different market."[120]

---

119 Subject to the issues discussed in Section II.D.

120 Sept FERC ¶ 22.

Absent the other features of the MRTU mentioned above, there might be a case for such opportunity cost concerns. However, as noted above, with the rest of the MRTU market design in place, we have been unable to identify any opportunity costs that are incurred by suppliers designated to provide RUC capacity. To the contrary, if the resource owner is adequately compensated for its start-up and minimum-load costs (including manning costs for GTs and any gas availability costs) through a bid production cost guarantee, the other elements of the MRTU RUC structure provide a no-lose opportunity for resource owners. If the Cal ISO's load forecast is correct and LSEs have underbid actual real-time load, real-time prices may be high and the resource owner will be paid real-time prices for the capacity committed in the RUC. If the resource owner earns substantial profits as a result of having been committed by the ISO, it will keep its profits. If, on the other hand, the Cal ISO's load forecast is mistaken and real-time prices are not high enough to cover the resource owner's start-up and no-load costs, the resource owner will receive a bid production cost guarantee and be made whole for the costs it incurs. In the circumstance of a capacity shortage in the west, the RUC system is not a substitute for a resource adequacy system and will not prevent exports from being scheduled day-ahead when prices are expected to be higher outside California.

The discussion above suggests that under the current MRTU design, a principal reason that a firm lacking market power would submit a significant RUC availability bid would be to ensure that it would recover its actual minimum-load costs if it were committed in the RUC (i.e., compensating for the difference between its actual costs and those determined under the MRTU methodology). Thus, such a firm would submit a bid reflecting the difference between its actual minimum-load costs and those used to determine uplift payments under the MRTU. If firms bid this way, the sum of the minimum-load costs used by the Cal ISO and the generator's availability payment would correspond to a unit's actual minimum-load costs, so commitment of RUC capacity taking account of these availability bids would improve the economic efficiency of the unit commitment. This market design system, however, would have two unintended consequences. First, unlike a firm's minimum-load bid, the availability bid would be used to determine a market clearing payment to all generators. Thus, if firm A were a gas fired generator with a high NOx emission rates, it might submit a $4/MW availability bid for its minimum-load block capacity, reflecting its losses on this capacity if it were compensated based on the formula used by the Cal ISO to determine minimum-load costs, which ignores NOx allowance costs. Such a bid would not reflect market power, but merely the normal incentive of a competitive firm to not operate if its revenues will not cover its avoidable costs.

If this unit with the $4/MW availability bid were economic to commit in the RUC, this $4/MW availability bid could determine the availability payment to all generation committed in the RUC. The minimum-load payments to the other generation, would not necessarily be less than their actual costs, however, so no supplementary payment might be appropriate. Thus, the market clearing feature of the availability payment is not appropriate for availability bids motivated by such a cost reimbursement shortfall.

The second unintended consequence of this pricing system is that RUC suppliers would receive the RUC availability payment even when there is no shortfall in minimum-load cost compensation. The minimum-load bid is not used to determine generator revenues but is used to determine the unit commitment and costs for the purpose of a bid production cost guarantee. If real-time prices are very high because the ISO's load forecast is correct, real-time energy revenues may more than cover the actual minimum-load costs of the unit committed in the RUC, obviating the need for any payment pursuant to the bid production cost guarantee. Under the availability payment approach, however, the unit owner would receive the availability payment even if the unit more than covered its minimum-load and other costs in the energy market at real-time prices.[121] In principle, in equilibrium this expected benefit might be reflected in lower availability bids but this argument for the ability of markets to undo the problem should carry weight only if there are substantial other benefits in creating the problem in the first instance.

On the other hand, the payment of a RUC availability payment reflecting the manning costs or gas scheduling costs of the incremental RUC supplier would be an efficient outcome and it would, in principle, improve economic efficiency to take these costs into account in determining which resources should provide RUC capacity. The choice between an "as bid" or "market clearing" availability payment is complicated in this context. If the availability payment reflected an opportunity cost of the marginal supplier, it would likely be economically efficient to reflect this in a market clearing price. In this context, however, the most apparent rationales for substantial availability payments are to compensate for defects in the pay-as-bid mechanism for recovery of commitment costs.

Commitment costs are inherently unit-specific. The most prominent examples, start-up and minimum-load costs, are neither fungible nor homogeneous. Although it is possible to consider alternative pricing mechanisms that would permit recovery of start-up and minimum-load costs this entails much broader changes in market design than envisioned in connection with the availability bids for RUC capacity. It is a common feature of other electricity markets that where these commitment costs are recognized they are treated on an as-bid basis. This as-bid compensation presents some difficulties, but the general view is that compared to energy costs and prices, the relatively more transparent and less volatile commitment costs can be handled with simple rules such as keeping the start-up bids fixed for months at a time.

Without a clear delineation of the costs and associated incentive problems within the context of the rest of the MRTU that give rise to opportunity costs to be compensated through an availability bid, it is difficult to identify a role for locational or market-clearing availability payments for capacity commitment.

An alternative approach to RUC pricing would be to eliminate the RUC availability payment and to revise the pricing and uplift rules to ensure that suppliers committed in the RUC will be made whole for their actual costs.[122] With the implementation of virtual bidding in the DAM, which would allow market participants to arbitrage underbidding by load and reflect expected scarcity in DAM prices, the RUC should be the residual back up mechanism that it is intended to be in other control areas.

---

121 The original CMD proposal addressed this to a degree by rescinding the availability payment if the unit was dispatched in real-time, but that approach had the limitation that it could rescind the availability payment even when real-time prices were not high enough to allow the unit to recover its actual minimum-load costs.

122 In the case of GTs scheduled in the RUC and required to be manned, perhaps they should qualify for a manning cost guarantee analogous to the start-up and minimum-load guarantee for steam generators that are committed in the RUC.

If an availability payment is retained, the Cal ISO needs to revisit the mitigation approach, include imports in RUC to avoid inflated RUC costs, assess the costs the availability payment is intended to compensate for, and clear the availability bids in a manner consistent with the nature of the costs that are intended to be reflected in the RUC availability bid. The RUC availability bid mechanism is more likely to contribute to efficient outcomes if it serves to recover the actual incremental costs associated with providing RUC capacity, such as manning costs or gas scheduling costs, than if it is necessary for suppliers to use the availability bid to recover energy market costs that they would not be permitted to recover in their energy offer prices.

D. Uplift

Under the MRTU, units committed in the day-ahead RUC will be eligible for recovery of start-up and minimum-load costs, net of market profits during the commitment period, subject to restrictions on self-scheduling and the RUC procured capacity being fully available for and responding to ISO dispatch instructions. Market profits will include energy revenues, ancillary services payments and the RUC availability payment. The length of the commitment period will depend on operating characteristics of the unit, including start-up time, minimum run time, etc.[123]

FERC initially rejected the netting of start-up and minimum-load costs against market revenues.[124] The Cal ISO persevered in proposing netting.[125] FERC ultimately accepted netting of all as bid costs against all market revenues.[126] The Cal ISO's position is correct. All as bid costs need to be netted against all market revenues. The failure to net start-up and no load costs against market revenues in calculating uplift would lead to extremely problematic bidding incentives.

Resources that self-schedule energy or self-provide ancillary services in the day-ahead market will be viewed as self-committed and will not be compensated by the ISO for start-up and minimum-load costs for the hours in which they are self-scheduled or self-committed.[127] Long start time units are eligible for recovery of start-up and minimum-load costs only if they are not self-scheduled and are committed by the Cal ISO in the DAM or RUC.[128] A resource eligible for start-up minimum-load cost recovery may lose its eligibility if it self-schedules energy or ancillary services in the hour ahead or engages in uninstructed deviations outside the tolerance band.[129]

---

123 CMD # 106.

124 Oct FERC ¶ 115.

125 May ISO, p 39, 51-52, Att A II.6.

126 June FERC ¶ 44, Sept FERC ¶19.

127 CMD # 106.

128 June FERC ¶ 30.

129 CMD # 106.

E. Self-Provided RUC Capacity

Market participants, including Metered Sub Systems, can self-provide RUC by designating capacity that will be used to provide RUC.[130] Market participants will therefore be able to withhold capacity from Pass 3 under the must-offer obligation (MOO) or flexible offer obligation (FOO) by identifying it as capacity used to self-provide RUC capacity. This ability could be problematic if a unit that is designated to self-provide RUC were one of a limited set of alternatives for managing a transmission constraint. If this capacity were effectively removed from Pass 3 in this manner, the DAM price within the congested area could be set by price sensitive load or virtual supply bids. The self-provided RUC capacity would be available in real-time, and real-time prices would reflect this, but these real-time prices would need to be reflected in the DAM through virtual supply bids or price capped load bids, which would not be possible under the proposed zonal virtual bidding system.[131] We assume that this would not be a substantial problem in practice as units possessing locational market power and potentially able to materially impact DAM congestion prices through physical withholding would be RMR units and would not be permitted to thus withhold capacity from DAM by contracting to provide RUC to a third party. This should be made explicit by the Cal ISO.[132]

Under the MRTU, self-provided RUC will be subject to a delivery test. This deliverability test would verify that the RUC capacity can be dispatched to meet forecast load, while meeting normal reserve requirements. This should be readily testable in the normal RUC logic structure simply by inserting bids of -$30 for self- provided RUC in Pass 4.

FERC conditionally approved the Cal ISO self-provision rules, but required further details.[133]

Self-provided RUC will not be eligible to set or receive availability payments in RUC.[134] FERC has agreed.[135]

F. Mitigation

RUC availability bids of generation dispatched in Pass 2 will be subject to mitigation to a reference level based on accepted unmitigated availability bids in past 90 days. There would be

[130] May ISO, p. 55.

[131] Capacity not offered in the DAM pursuant to the FOO is similarly not available for congestion management in Pass 3, although it must be available in real-time. This is discussed further below.

[132] As noted above, the more likely outcome under the MRTU market design is not economic withholding of generation that elevates DAM prices above expected real-time prices but rather economic withholding of generation in the DAM because nodal DAM prices inside constrained areas are materially below expected real-time nodal prices.

[133] June FERC ¶ 57.

[134] May ISO, p. 40 Att A II.9.

[135] June FERC ¶ 45.

a common reference level for all units, with separate reference prices for on and off-peak hours.[136] The specifics of the mitigation process are discussed in Section VII.D2.

In addition to mitigation of RUC availability bids, it was originally proposed that incremental energy offer prices of capacity selected in the RUC scheduling process could not be increased in price once selected for RUC but could be decreased.[137] This restriction was rejected by FERC, which ordered that offers can be increased between the DAM and real-time for RUC units to reflect factors such as intra-day gas costs.[138]

It is preferable from an electricity market design, and gas system reliability standpoint to permit the RUC suppliers to increase their day-ahead offer prices to reflect intra-day gas market conditions. This ability to raise offer prices may complicate offer price mitigation for the real-time dispatch because of the need to take account of intra-day gas prices, but we understand that the Cal ISO proposes to undertake such mitigation in any case without regard to unit RUC status.

G. Recall

Export schedules cleared in the DAM are not subject to recall in RUC or in hour-ahead/real-time. FERC requested clarification of whether energy committed in the RUC can be sold via bilateral transactions in the hour-ahead market.[139] We believe that the answer is potentially "no" in the case of exports but "yes" for bilaterals serving load within the Cal ISO control area. Export schedules submitted hour ahead will have lower priority than internal California load for scheduling in the hour-ahead evaluation. Hour-ahead export schedules must be submitted with a price bid capped at $250, while internal California load will be modeled with a sink price bid in excess of $250 in the scheduling pass. Export load will be treated like internal California load only if the export was scheduled in the DAM, if the export is linked to an import scheduled in the hour-ahead evaluation, or if the export is linked to a generation resource not committed in the DAM or RUC.[140] These rules effectively ensure that capacity committed day-ahead to serve California load, either in the DAM or the RUC, is available to meet California load in real-time.[141] At the same time, however, the $250/MWh price cap and the lack of any resource adequacy mechanism providing a right of recall in the DAM will ensure that California is left short of capacity in the RUC if market prices rise above $250/MWh, either as a result of high gas prices or capacity shortages.

---

136 May ISO, pp. 50-51 Att A II.5.

137 CMD # 108.

138 June FERC ¶ 79,80, Sept FERC ¶22-24.

139 Oct FERC ¶ 123.

140 We believe this statement is consistent with the Cal ISO's intent. In particular, we do not believe that the Cal ISO intends market participants to schedule an import in the DAM and then link it to an export in the hour-ahead to create a wheel? Nevertheless, it should be recognized that if real-time prices in California are lower than elsewhere in the West, imports scheduled day-ahead will likely not flow in real-time, being sold instead into other higher priced markets.

141 This reflects a change from the September 2003 filing which provided that: "there is no obligation to serve California load (e.g., a resource, even it was committed in RUC) could schedule all of its power to Arizona and still be in compliance with the MOO," p. 72

The export recall issue would need to be revisited depending on developments in the resource adequacy requirements being developed by the CPUC.

H. Allocation of RUC Procurement Costs

Costs associated with the day-ahead RUC process will be borne first by scheduling coordinators whose metered load is not fully scheduled in the DAM. The ISO is to calculate a per MWh RUC charge by dividing total RUC procurement costs by the gross amount of RUC capacity and energy procured and somehow allocating this to load. It is proposed that this cost would be allocated to all virtual supply (which is by definition not delivered in real-time) and to each MW of metered load in excess of final day-ahead schedules. Changes in generation and load as a result of ISO dispatch instructions would not be counted as such deviations. Any excess RUC costs (i.e., costs arising from ISO load forecast error) will be allocated to all metered load plus exports.[142] This allocation methodology was approved by FERC.[143]

Apparently there would not be any geographic allocation of RUC costs. Thus RUC costs incurred to meet forecast load in San Diego would be allocated to real-time load imbalances throughout the state. This cost allocation methodology could be problematic given the load zone/LAP scheduling practices discussed in Section II.C above. With nodal clearing of zonal load bids, it s possible that LSEs in a particular LAP could fully schedule their load in the day-ahead market, yet substantial commitment costs could be incurred on their behalf in the RUC, because the Pass 3 schedules would not be feasible. These RUC costs would then be allocated to load and virtual supply imbalances throughout the state. If the nodal clearing procedure for zonal load bids were implemented, the Pass 3 schedules in the DAM would likely be infeasible, and additional units would be committed in Pass 4. The CMD provides that the procurement costs associated with this capacity would be divided by the amount of capacity procured but it is not entirely clear how the amount of capacity procured is defined. Would the capacity procured be defined as: (a) the difference between the ISO's forecast load and the load cleared in Pass 3; (b) the capacity of the minimum-load blocks of units committed in Pass 4; (c) the total capacity of all units committed in Pass 4, including capacity not eligible for a RUC payment; (d) the total capacity eligible for a RUC availability payment; or (e) something else?[144]

---

142 CMD # 111.

143 June FERC ¶58.

144 The Cal ISO offers as an example by way of clarification. "There are two sources of cost and corresponding MW quantities in RUC: (A) minimum load cost compensation (MLCC) for the minimum load MW of resources committed in RUC; (B) RUC availability payments, which are paid only for capacity slated in RUC above minimum load or above day-ahead Energy and A/S schedule. The unit rate ($/MW/hr) in question is computed by adding the $ amount of (A) and (B) and dividing by the higher of the sum of the MW quantities of (A)+(B), or the amount of underscheduled load (i.e., real-time metered minus scheduled load). If the MW of (A)+(B) is higher than (real-time metered minus scheduled load), this rate will not be enough to recover the $ amount of (A)+(B) from underscheduled load; the shortfall is then charged to all metered demand. Example: Assume that RUC Minimum load = 500 MW, RUC MLCC (after net of market) = $10,000; RUC capacity (above minimum load and/or above IFM schedule = 2,500 MW, RUC availability payment = $20,000. Assume the amount of underscheduled load is 2,000 MW and the total metered load is 40,000 MW. The charge per MW of underscheduled load is ($10,000 + $20,000)/max(500+2,500, 2,000) = $10/MW/hr. This brings only $10*2,000

Approach (a) could lead to very high per MW charges (thousands of dollars/MW) if units are committed in the RUC to solve infeasible schedules on days when bid load is close to forecast load. Approach (b) could lead to total RUC charges that greatly exceeded procurement costs because the denominator could be much lower than the difference between forecast and bid load. Approaches (c) and (d) could possibly result in RUC charges that exceed RUC procurement costs the denominator of (c) does not include unloaded capacity committed in Pass 3, while the denominator (d) does not include capacity that does not receive the RUC payment, however the more likely outcome under nodal clearing of zonal load bids would be that all of these approaches would allocate most RUC procurement costs to all load because the amount of capacity committed in pass 4 would be any measure greatly exceed the difference between forecast load and the load cleared in Pass 3. These complications associated with the allocation of RUC procurement costs are not intrinsic features of the MRTU market design but rather arise from the nodal clearing of the zonal LAP bids and its associated problems. These complications can be avoided by clearing zonal load bids zonally, in which case approaches a) and d) should yield similar results.

It should also be understood that if the load forecast used for the RUC commitment turns out to be accurate in real-time and there are no additional low cost imports available in real-time, then the uplift costs associated with the RUC procurement should generally be small, as operation of the units committed in the RUC should be economic at real-time prices, enabling these units to recover most or all of their commitment costs. Uplift costs could be incurred if the load forecast used for the RUC procurement turns out to be too high in real-time, and if so, some of these costs would be allocated to all load. If the load forecast used in the RUC commitment is accurate but uplift costs are incurred because low costs imports become available in real-time, then the uplift costs will be allocated to the LSEs that benefited from the low cost imports (i.e. the LSEs that did not buy power in the DAM and instead purchased power in real-time).

Metered subsystems can follow their own load, without incurring RUC costs, provided they establish resources in advance, schedule all load and exports in the DAM and meet a bandwidth requirement.[145]

= $20,000 when charged to the underscheduled load (Tier 1). The shortfall of $10,000 is charged to all metered load (Tier 2) at a unit rate of $10,000/40,000 = $0.25/MWh."

145 CMD # 112.

IV. HOUR-AHEAD FORWARD SCHEDULING AND RUC PROCESS

A. Hour-Ahead Schedules

It is proposed that under the MRTU there will be an hour-ahead scheduling process, rather than an hour-ahead market. Intertie scheduling and scheduling of other resources with binding predispatch instructions will be carried out in a process similar to that used by the NYISO. There will be no hour-ahead settlement.[146] This simplified hour-ahead process was initially accepted by FERC but is now being reexamined.[147]

Scheduling coordinators will submit bids and hour-ahead self-schedule changes for resources and imports as well as changes to wheeling schedules by T-75. There will be no need for load bids or load schedule changes in the hour-ahead process, as settlements will be based on real-time load. The Cal ISO will run the hour-ahead process to manage congestion and balance energy and rebalance ancillary services based on forecast load to the extent that additional ancillary service capacity is required. Hourly intertie schedules will be determined in this process based on the Cal ISO's load forecast.[148] This concept was accepted by FERC.[149]

The proposed Scheduling Priorities in the Hour-Ahead evaluation would be: 1) Final day-ahead schedules submitted without energy bids and other final day-ahead schedules that were converted to self-schedules after clearing the DAM; 2) hour-ahead deviations associated with ETC schedules; 3) hour-ahead self-scheduled deviations from must-take/must run resources; 4) all other hour-ahead self-scheduled deviations; and 5) hour-ahead supply and demand deviations with energy bids.[150]

Priority 4 will include internal California load and self-scheduled exports supported by wheels or self-committed generation not scheduled in the DAM or day-ahead RUC. Priority 5 will include all hour-ahead exports not supported by wheels or self-committed generation not scheduled in the DAM or day-ahead RUC. These load priorities will be relevant only if there are insufficient resources available at the bid cap price to meet bid (export) and forecast (internal) load.

The FERC suggested in its September Rehearing Order that the Cal ISO evaluate the costs and benefits of employing a financially binding hour-ahead market instead of the simplified

---

146 May ISO, pp 5-6, 79.

147 June FERC ¶ 93. Sept FERC ¶ 45-46.

148 The July 2003 filing proposed that the hour-ahead market would close at T-120, final hour-ahead schedules published at T-90, close the real-time market at T-60. This time line allowed a 30 minute rebid period between final hour-ahead schedules and the close of real-time bid submissions. CMD # 114.

ISO would then perform a real-time predispatch at T-45 to enable the ISO to give real-time dispatch instructions to units than cannot change operating levels in response to intra-hour dispatch instructions. Imports scheduled for the entire hour will be guaranteed their bid price but cannot set the five minute MCP. CMD # 116. This time line was replaced by the new combined process in the May 2004 filing. See also May ISO, pp. 80-81, Att A V.1,2.

149 June FERC ¶ 94.

150 CMD # 34, May ISO, p. 81.

hour-ahead scheduling process currently reflected in the MRTU market design.[151] As suggested by FERC, the Cal ISO could clear a financial hour-ahead market based on any additional bid load and then run a RUC process to ensure that sufficient resources were available to meet the Cal ISO's load forecast for the hour. FERC suggested that such an hour-ahead market might be helpful because of the variability of load in California and the importance that hour-ahead scheduling adjustments be accurate.[152] The proposed hour-ahead financial market would not be directly relevant to the accuracy of hour-ahead scheduling adjustments objective, because the hour-ahead schedules whose accuracy is important are those determined in the RUC process based on forecast load. Similarly, the FERC suggested that given the level of imports and the need for the Cal ISO to commit to a specific level of imports in the hour-ahead timeframe, it is important that hour-ahead schedules be accurate. Once again, the proposed hour-ahead financial market is not directly relevant to this objective, because the hour-ahead import schedules that need to be accurate are those determined in the RUC process to meet forecast load.

While an hour-ahead market would provide an additional market in which imports and exports could clear, market participants that wish to lock in the cost of imports or exports prior to real-time can do so by entering into bilateral contracts and scheduling these import or export transactions in the simplified hour-ahead scheduling process. The change in costs that could not be hedged bilaterally would be limited to the intra-hour change in congestion charges on transfer capability not scheduled in the day-ahead market. Hence, while there could in principle be benefits to providing incentives for bid-in load to provide better information for the load forecast, this indirect effect on scheduling based on the load forecast only an hour or so before real-time may not be very substantial. Unless carefully structured, the potential for import transactions to be cleared at low prices in a financially binding hour-ahead market due to low bid load while real-time prices are high, could discourage real-time imports unless the dual markets are carefully designed. Furthermore, clearing an hour-ahead ancillary services market in conjunction with the energy market as suggested by FERC[153]could cause problems because of the likely inconsistencies between the ancillary services schedules determined in conjunction with the load bid into an hour-ahead market and those based on forecast load for actual real-time operation.

The existence of both an hour-ahead market and a subsequent Hour-Ahead Scheduling Process or RUC process would introduce market design issues that would need to be satisfactorily resolved to avoid adverse impacts on market efficiency and reliability. Some of the significant market design issues that would arise with an additional hour-head market.

- Would import suppliers and export buyers be permitted to revise their bids and offers between the hour-ahead market and the hour-ahead scheduling process or RUC evaluation?

- Would internal generation suppliers be permitted to revise their bids and offers between the hour-ahead market and real-time?

---

[151] Sept FERC ¶ 45-46.

[152] Sept FERC ¶ 45.

[153] Sept FERC ¶ 46.

- Would virtual demand and supply bids be permitted in the hour-ahead market?

- Would LSEs be permitted to submit additional bilateral schedules in the hour-ahead scheduling process following the hour-ahead market?

- Would export buyers be permitted to schedule exports in the hour-ahead market that were not scheduled in the day-ahead market?

Some of the considerations involved in resolving these issues would be:

- If import suppliers were not permitted to offer additional supplies in the hour-ahead scheduling process or RUC, introduction of the hour-ahead market would move forward the effective deadline for scheduling imports, potentially reducing import supply offers.

- Such a structure could introduce incentives for LSEs to bid less than their expected load into the hour-ahead market, in order to price discriminate between suppliers selling power in the hour-ahead market and the real-time market. While the outcomes could be profitable for an individual LSE they could lead to a change in offer prices by suppliers or even a reduction in supply that would raise costs for the market as a whole.

- If suppliers were not permitted to revise their offers between the hour-ahead market and real-time and there were no virtual bidding in the hour-ahead market, suppliers would be likely to offer their output into the hour-ahead market at the expected real-time price, rather than at incremental cost, leading to market inefficiency, complicating market power mitigation and likely raising costs for loads and suppliers;

- If suppliers were permitted to revise their offers between the hour-ahead market and real-time, or if import suppliers were permitted to offer additional supplies in the hour-ahead scheduling process or RUC, this would expand the time interval required between the posting of schedules for the hour-ahead market and the running of the subsequent hour-ahead scheduling or RUC process.

- If export buyers were permitted to schedule exports in an hour-ahead market that preceded the hour-ahead scheduling and RUC process, underbidding by LSEs in the hour-ahead market could result in exports being scheduled at a level that results in reserve shortages in real-time, despite adequate resources committed in the day-ahead RUC.

In this regard it is particularly important to recognize that while some market participants may anticipate that with the introduction of an hour-ahead market hour-ahead prices would be systematically lower than real-time prices and see an advantage in such an outcome, this would be an unfortunate outcome. A market design in which hour-ahead prices were systematically lower than real-time prices would be precisely the kind of outcome that would need to be avoided in a market design that includes both an hour-ahead market and real-time settlement if the Cal ISO is to avoid adverse impacts on reliability. If, for example, the hour-ahead market were structured in such a way as to allow LSEs to price discriminate between the hour-ahead market and real-time, such a circumstance would serve to drive price sensitive supply offers out of the hour-ahead market, raising prices and making overall operating day supply and imbalance prices more volatile.

Among other purposes, the structure of the hour-ahead scheduling process is intended to assure that capacity scheduled in the day-ahead RUC is available to meet control area load and is not used to meet export demand if this would leave inadequate resources to meet control area load. This end would be accomplished by determining real-time exports (i.e., those not scheduled in the day-ahead market) in the hour-ahead scheduling process which will take account both of export demand and the Cal ISO's control area load forecast.

In the circumstance in which there is excess demand at the bid cap in the hour-ahead scheduling process, hourly schedules will be determined in part based on scheduling priorities. The proposed scheduling priorities for the transactions scheduled in the hour-ahead evaluation (imports, exports, wheel-throughs and schedules for units unable to follow real-time dispatch instructions) would be: 1) final day-ahead schedules submitted without energy bids; 2) hour-ahead deviations associated with ETC schedules; 3) hour-ahead self-scheduled deviations from must-take/must run resources; 4) all other hour-ahead self-scheduled deviations; and 5) hour-ahead supply and demand deviations with energy bids.

Priority 4 will include internal California load based on the Cal ISO's load forecast and self-scheduled exports supported by wheels or self-committed generation not scheduled in the day-ahead market or day-ahead RUC. Priority 5 will include all hour-ahead exports not supported by wheels or self-committed generation not scheduled in the day-ahead market or day-ahead RUC. These load priorities will be relevant if there are insufficient resources available at the bid cap price to meet the Cal ISO's load forecast.[154] In this situation, priority 5 exports will be scheduled to the extent that resources are available in addition to those required to meet the ISO's load forecast and day-ahead export schedules.

If export bids were cleared in a separate hour-ahead market that took account only of bid load, there would be a potential (if control area load bid into the hour-ahead market was less than forecast load) for exports to be scheduled in such an hour-ahead market supported by capacity committed in the day-ahead RUC, even if the scheduling of those exports left inadequate resources to meet control area load in real-time. To avoid this outcome in a market design including an hour-ahead market, it would be necessary to either impose other restrictions on exports scheduled in a hour-ahead market that could create seams and price disparities in the West during non-shortage conditions, or accept the possibility that capacity committed in the day-ahead RUC could be used to support exports during periods in which the Cal ISO control area was reserve-short.

The hour-ahead scheduling process is intended to avoid such an outcome because export schedules would be determined taking into account both export demand and the Cal ISO's load

[154] In fact, these priorities will only be relevant in the circumstance in which schedules in the hour-ahead scheduling process are cleared either at the bid floor or the bid cap.

forecast. Real-time exports would be scheduled in the hour-ahead scheduling process if they did not compromise the Cal ISO's ability to reliably meet control area load but exports supported by RUC capacity would not be scheduled under the hour-ahead scheduling process if the scheduling of those exports were expected to have adverse reliability impacts (i.e., if insufficient capacity were available at the bid cap to both meet export demand and control area load). This will be accomplished by assigning exports bids submitted in the hour-ahead scheduling process a lower priority than internal Cal ISO control area load.[155] This structure of the day-ahead market allows the Cal ISO to take account of export schedules in the day-ahead RUC commitment and ensures that sufficient capacity is committed to meet control area load, to the extent that sufficient capacity is available at the bid cap. At the same time, market participants would be able to schedule exports day-ahead that would not be subject to curtailment in real-time.

A second disadvantage of introducing an additional hour-ahead market is that because the hour-ahead market would be in addition to the other hour-ahead processes, it must precede them in time, requiring that the hour-ahead market be moved forward in time, relative to the proposed hour-ahead scheduling process, resulting in a greater time difference between such an hour-ahead market and real-time than between the proposed hour-ahead scheduling process and real-time. Under the proposed hour-ahead scheduling process, scheduling coordinators would submit bids and hour-ahead self-schedules and self-schedule changes for resources and imports, as well as changes to wheeling schedules, by 75 minutes prior to the operating hour. If this process were to be preceded by an hour-ahead market, the bid submission deadline for the hour-ahead market would need to be moved further forward in time. If market participants were provided an opportunity to rebid between the hour-ahead market and the hour-ahead scheduling process/RUC, the time frame for submission of bids and schedules to the hour-ahead market would certainly be two or more hours in advance of real-time.

A third disadvantage of adding an explicit hour-ahead market relative to the proposed hour-ahead scheduling process is that the introduction of a third settlement process (in addition to the day-ahead market and real-time imbalances) will increase the administrative costs of both the Cal ISO and its market participants. While there may have been a need to bear the administrative costs of a third settlement under the prior market design, as a result of the constraints placed on the real-time dispatch by the market separation doctrine, that is no longer the case. One of the potential cost savings from the introduction of LMP and elimination of market separation is elimination of these additional settlement costs. The administrative costs of implementing an additional market are not insignificant and need to be considered in choosing among these alternatives.

B. Pricing

While the hour-ahead scheduling process would implicitly and/or explicitly calculate prices, these hour-ahead prices would not be used for settlement purposes.[156] The Cal ISO has clarified

---

[155] Since exports scheduled in the Cal ISO's day-ahead market are not recallable by the Cal ISO in the hour-ahead process but can be cancelled by the market participant and sold into the Cal ISO market in real-time, market participants wanting to export firm power can do so under the proposed hour-ahead scheduling process by scheduling those exports in the day-ahead market.

[156] May ISO, p. 81, Att A V.2e.

that imports and exports will be scheduled based on offer prices in the hour-ahead evaluation but will settle at real-time prices. This approach is satisfactory if there is no transmission congestion on the interties in the hour-ahead evaluation but will likely lead to problems when the ties are congested. This is the third of the six major implementation issues identified with the MRTU market design. The potential problem is that given the Cal ISO's real-time modeling of tie lines as open loops for dispatch and pricing purposes, the dispatch of internal generation to manage tie-line flows will not be reflected in real-time prices.[157] This is analogous to the pricing rule the NYISO had at its start-up (as a result of a miscommunication of intended tie line modeling). This kind of pricing led to significant problems for the NYISO. In fact, this was one of the two big problems hindering the NYISO over its first six months of operation. The issue was that offer prices for imports in the NYISO hour-ahead process had nothing to do with the price the import supplier would be paid if the import was scheduled, so external suppliers had an incentive to bid - $1,000 to get themselves scheduled. This lead to a variety of difficulties and the problem was not solved until ECA B was implemented in fall 2000. It appears that the CMD pricing system for imports and exports is identical to that initially used by the NYISO. It therefore appears likely that the Cal ISO will run into similar problems and needs to modify the CMD to address this problem in some manner.[158]

---

[157] The Cal ISO has explained that internal generation is at times redispatched to manage tie line flows, but this redispatch is based on operator assessment of the effectiveness of this redispatch and managed by taking units out of merit and the cost of the redispatch would not be reflected in congestion prices on the ties.

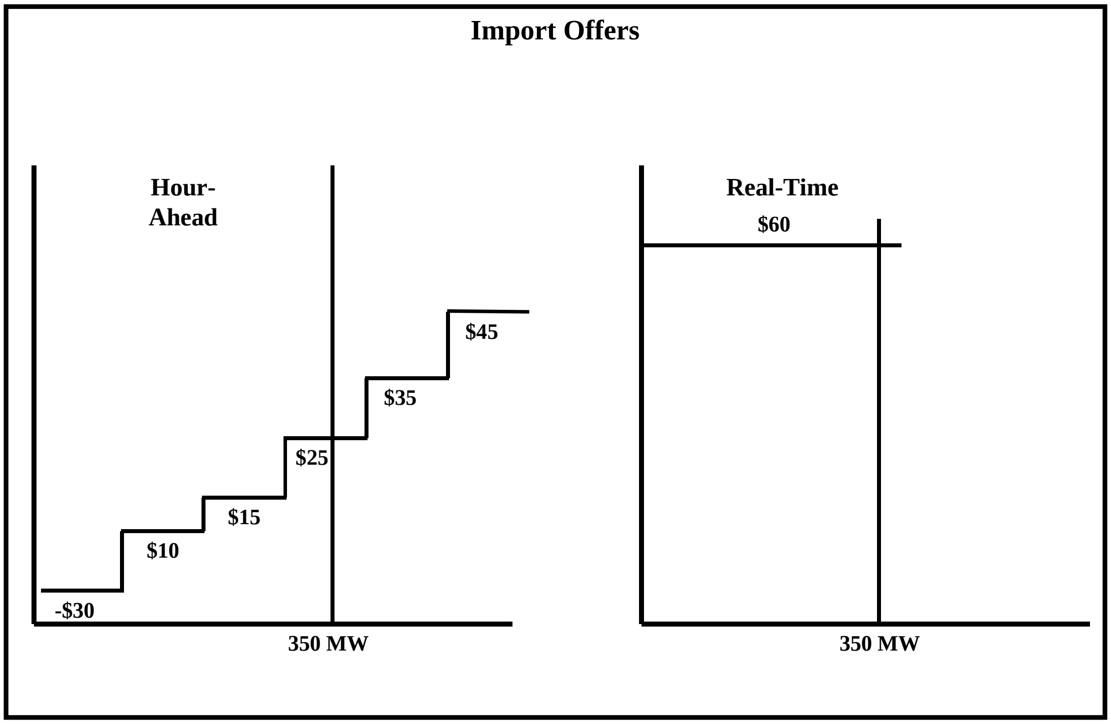
[158] The ECA B pricing mechanism was implemented by the NYISO to solve this problem in New York and was subsequently approved by New York market participants and the FERC. It has worked well to date. The original ECA B wording was:

i. These rules apply in hours that HAM import or export transactions are constrained in BME at an external proxy bus by a transmission limitation or a desired net interchange limit.

ii. For each hour in which one of these limits constrains net imports to NYCA from an external proxy bus in the HAM, the real-time settlement price at that external proxy bus will be the lesser or the real-time LBMP or the BME price at that external proxy bus.

iii. For each hour in which one of these limits constrains net exports from NYCA to an external proxy bus in the HAM, the real-time settlement price at that external proxy bus will be the greater of the real-time LBMP or the BME price at that external proxy bus.

Figure 4
CONGESTION PRICING FOR IMPORTS



If the real-time pricing system does not reflect the constraints that determine schedules in the hour-ahead process, there will be a disconnect between bids and revenues as portrayed in Figure 4. Bidding lower in the hour-ahead process would not reduce the price received when imports are constrained, the lower bids would only affect the probability of being scheduled.

Various elements of the MRTU provide that changes in offer prices in the hour-ahead process will be subject to activity rules. In particular, the offer prices of capacity scheduled to provide energy cannot be decreased. The MRTU would allow unloaded RUC capacity to decrease its offer price between day-ahead and real-time, and allow decremental offer prices for scheduled energy to be increased.[159]

The restrictions on offer price reductions between DAM and hour-ahead were approved by FERC.[160] It is assumed that the restrictions on reductions in day-ahead offer prices are intended to prevent market participants from reducing their offer prices to force the ISO to pay

---

159 CMD # 119.

160 The CMD originally proposed that generating units committed in the RUC also could not raise their energy bids between the DAM/RUC and real-time. CMD # 108. It was proposed that any cost difference would be recovered in the RUC availability bid. The FERC rejected this restriction on changes in energy bids between DAM and RUC. The FERC observed that recovering fuel costs in availability bids distorts, dispatch and scheduling of RUC, plus impossible for generators to estimate change in costs. June FERC ¶ 79, 80, Sept FERC ¶ 22-23. It seems to us that the FERC approach is preferable as it avoids problems that would arise under the original approach. The CMD also provided that energy offer prices for capacity scheduled to provide ancillary services could be increased, but FERC eliminated this restriction. Sept FERC ¶ 25, 26, 27. See also Oct FERC ¶ 137.

very low or even negative prices to back down their day-ahead schedules in the event that transmission outages make the day-ahead schedules infeasible. If these restrictions on offer price reductions were effective, they could raise consumer prices in the day-ahead market, however, by driving some resources that need to be able to schedule in this manner out of the market.

One category of resources that could be adversely impacted by such restrictions would be imports. Market participants selling power into California in the day-ahead market will likely then schedule transmission to deliver power into the market. Once this transmission is scheduled, the money spent on transmission service is sunk and the importer will want the power to flow even if the hour-ahead price falls slightly below the day-ahead offer price. If importers risk having transactions cut in the hour-ahead if the price falls even slightly below the day-ahead offer price, they may offer energy in the day-ahead market at higher prices than would otherwise be the case.

Similar incentives may apply to the output of cogeneration units whose offers in the day-ahead market may entail irreversible changes in production scheduling for the operating day. If they would potentially be dispatched off in the hour-ahead at only a slightly lower price than that accepted in the DAM, they may be unwilling to offer this capacity in the DAM.

In practice, these restrictions on offer price reductions appear to be rendered completely ineffective by the priority given day-ahead schedules in the DAM, which implicitly allow offer price reductions to -$30/MW for all day-ahead schedules.[161] In fact, it is understood that it is intended that, absent action by the resource owner, the real-time offer prices of any resource scheduled in the DAM will be set at -$30/MWh for the resources DAM schedule. This provision makes the activity rules on bid reductions meaningless and counterproductive. Rather than preventing reductions in offer prices to very low negative values, these activity rules actually require that resource owners seeking to reduce their offer prices slightly to reflect the kind of considerations noted above must reduce them all the way to -$30/MWh, and the rule serves only to prevent resource owners from selecting more moderate values. The activity rule precluding reductions in offer prices other than to -$30/MWh should therefore be eliminated.

Energy or capacity offered but not accepted in the day-ahead market can be offered in subsequent markets at higher or lower prices (subject to market power mitigation, etc.).[162] Generating units committed in the day-ahead market or RUC cannot decommit without reporting an outage to the ISO.[163]

C. Ancillary Services

The ISO will allocate DAM ancillary services capacity to the portion of the day-ahead energy curve that precedes the portion allocated to the day-ahead RUC.[164] The hour-ahead process will

---

[161] CMD # 34, May ISO, p. 81.
[162] CMD # 119.
[163] CMD # 120.
[164] CMD # 121.

produce advisory real-time ancillary services awards for any incremental capacity needed for load forecast changes or outages. The awards are advisory until finalized in the dispatch.[165]

The May 2004 filing appears to indicate that there would be no separate ancillary services capacity offer prices in the hour-ahead evaluation but we understand that capacity offer prices will be permitted for both reserves and regulation in the hour-ahead evaluation.

The May filing as noted above described the ancillary services awards determined in the hour-ahead process as advisory until finalized in the dispatch. It is understood that resources will be selected to provide reserves in the hour-ahead evaluation, taking account of both ancillary service capacity bids and energy offer prices. It is also understood that these schedules will be advisory in the sense that resources scheduled hour-ahead to provide reserves that end up being dispatched in real-time for energy as a result of capacity shortages or transmission constraints will be paid for energy only, not for reserves. This is reasonable.

An issue raised by the FERC Rehearing Order is the appropriate activity rule for capacity scheduled day-ahead to provide reserves. The CMD would have forbidden increases in real-time energy offer prices for such capacity but FERC has ordered that such increases be allowed. The Cal ISO has observed that once scheduled to provide reserves in the DAM, suppliers may have less competition in real-time. It should be kept in mind, however, that the DAM reserve optimization does not in any case consider the energy offer price of resources scheduled to provide reserves, except as an opportunity cost of using low-priced capacity to provide reserves. A supplier wanting high energy prices on its reserve segments could, therefore, self-schedule reserves and submit high energy prices on the capacity providing reserves without regard to the activity rules. Moreover, as FERC observed, there can be changes in fuel costs between DAM and real-time that ought to be reflected in real-time offer prices.

The underlying issue that this provision seeks to address is the potential for reserve suppliers that know they will have to be dispatched in the event of a generation or transmission contingency (depending on the probabilities involved) to submit inappropriately high energy offer prices on reserve capacity that would then be used to set prices during a contingency. This concern would be better addressed through appropriate market power mitigation and shortage pricing than through indirect restrictions on changes in offer prices, which do not address the real concern.

The Cal ISO should avoid unnecessarily limiting the resources available to provide ancillary services in real-time. All internal resources that are available for dispatch on a five minute basis should be available to be designated to provide reserves, up to the limits determined by their capacity and ramp limits, regardless of the ancillary services designations in the hour-ahead process. This would improve reliability by providing Cal ISO operators more flexibility in responding to real-time conditions and would assure that ancillary services and energy prices are rationally related to actual shortage conditions.

---

165 May ISO, p, 81, Att A V.2f; June ISO, p. 14.

D. Imports/Export

The hour-ahead process will determine the hourly predispatch of inter-tie bids.[166] The proposed hour-ahead process will also assure that RUC capacity is reserved for control area load and is not used to meet export demand.[167] The process accomplishes this because export schedules in the hour-ahead have lower priority than internal ISO load. Since exports schedule in DAM are not recallable in the hour-ahead process but can be cancelled, market participants potentially wanting to export firm power simply need to schedule these exports in the DAM.

E. Hour-Ahead RUC

The original CMD provided for a separate hour-ahead RUC process.[168] Under the current MRTU market design there will be a single process in which resources are scheduled against forecast load to determine import and export schedules and commitment of GTs, etc. There will no longer be a distinct hour-ahead RUC or hour-ahead RUC costs to be allocated. If the imports scheduled in the hour-ahead process and other commitment decisions (such as starting gas turbines) are economic at real-time prices, there will be no uplift costs to allocate. In practice, however, there will inevitably be hours in which the Cal ISO overestimates demand and imports scheduled in the hour-ahead process are uneconomic at real-time prices. These uplift costs will need to be allocated, presumably to all real-time load.[169]

---

166 May ISO, p. 82, Att A V.2g.

167 May ISO, p. 6.

168 Following each hourly run of the hour-ahead IFM the ISO will perform the hour-ahead RUC. ISO will assess whether committed resources and schedules are sufficient to meet real-time energy and reserve needs. If there is a short-fall, the ISO can commit additional resources that may require advance notice of an hour or more to be available for real-time operation. If the capacity procured in the day-ahead RUC appears excessive, the ISO will be able to cancel start-ups of RUC units. CMD # 113. Costs of hour-ahead RUC will be allocated first to real-time load in excess of final hour-ahead schedules. CMD # 111.

169 CMD # 111.

V. REAL-TIME DISPATCH

A. General

The real-time dispatch interval will be 5 minutes.[170] Incremental energy offer prices that were used to establish a resources final day-ahead and hour-ahead schedules will be available to the ISO to use in real-time for decremental adjustments to clear the imbalance energy market and to mitigate real-time congestion.[171]

B. COG Pricing

Gas turbines and other internal units that are lumpy with minimum operating level or minimum operating time constraints (constrained output generation) will be guaranteed their BPCG when they are dispatched by the ISO.[172] Because gas turbines generally have minimum run times and down times of one hour or more but are often dispatched in real-time in response to unanticipated changes in market conditions, they may end up operating for periods of time after real-time prices fall below their cost. The implementation of a BPCG for COG units is an important improvement in the Cal ISO market design that should improve market performance, as well as improve bidding and operating incentives. Moreover, COG units (GTs) will be treated as flexible for the purpose of setting real-time prices.[173] These resources would, however, only set the 5-minute prices when their energy is needed by the system, i.e., they would not set prices when they are on line as a result of a minimum run- or down-time constraint.[174] If a COG unit is no longer needed to meet load but is running due to a minimum-run time constraint, then it will not be able to set prices, but it would be eligible for a bid production cost guarantee if it was committed or dispatched by the ISO.[175] COG units will be block loaded in the real-time dispatch, either at zero or to its full capacity.[176] COG pricing was accepted by FERC, with the scope of COGs limited to gas turbines.[177] This general approach to COG pricing is reasonable, an important improvement on current procedures, and should improve the availability and performance of these units.

While this general approach to COG pricing is appropriate, experience in the NYISO has shown that the details of how COG pricing is implemented are very important. One implementation issue that is not discussed in the CMD is the determination of upper dispatch limits from unit ramp rates. A critical issue in implementing the COG pricing is how the upper dispatch limit of non-COG units is determined in the pricing dispatch. It is understood that it is

---

170 CMD # 14.

171 Subject, of course, to suppliers opting to convert accepted-day-ahead schedules to price-taking self-schedules. CMD # 122.

172 CMD # 61, 106, 116.

173 May ISO, p. 61, Att A III.3d,e.

174 CMD # 117.

175 May ISO, Att A III.2b.

176 May ISO, p. 60.

177 June FERC ¶121.

proposed that the upper dispatch limit of non-COG units would be based on their telemetered output plus their ramp rate. This is the fifth of the major implementation issues identified with the MRTU market design. The potential problem with this approach is that if some event has caused GTs to be running uneconomically in real-time, the actual output of the non-COG units will be backed down to accommodate the output of these uneconomic GTs. If the amount of uneconomic GTs running is large relative to the five minute ramp rate of the non-COG units that are on the margin, the calculation of whether the GTs are "needed" may find that GTs are "needed," but they would be "needed" only because the non-COG units have been backed so far down because of the uneconomic GT output that the non-COG units on the margin cannot ramp up fast enough in the price setting dispatch to replace all of the uneconomic COG units. As a result, prices could be determined by the offers of high-priced units that are not actually dispatched or by the offers of COG units that are actually not needed to meet load. The NYISO had problems with spurious price spikes arising from these kinds of situations until the NYISO changed the logic structure in the price calculation step in late July 2000.

It may be the case that the small amount of COG generation and the large amount of fast ramping hydro capacity in the Cal ISO control area will mean that this circumstance would rarely arise in California. However, New York also has a large amount of fast ramping hydro capacity and this problem still arose. This problem tends to arise when there are transmission constraints and both the COG generation and the non-COG generation backed down to accommodate the COG generation are on the same side of the transmission constraint. In this circumstance, the ramp rate of the non-COG generation that is backed down may not be large relative to the capacity of the out of merit GTs. For example, suppose that the San Francisco peninsula was constrained and Potrero 3 was on line meeting load at the margin at a higher offer price than that of generation outside the Peninsula. Then a line trip into the peninsula causes the Cal ISO to dispatch two of the Potrero GTs to eliminate the overload. After the GTs are on line, however, the transmission line comes back. Both Potrero 3 and the GTs are higher cost than the marginal unit located off the peninsula, but the GTs still have 45 minutes of their minimum run time to go, so Potrero 3 is backed down 100 MW to make room for the output of the GTs. The question is, what is the ramp rate of Potrero 3 relative to the output of the GTs. If Potrero 3's ramp rate were 1 percent of its capacity per minute, that would be 9-10 MW over a five minute dispatch interval. The two GTs have a combined output of around 100 MW, so the COG pricing logic would always find that the GTs are needed to meet load, but it is known that in fact Potrero 3 can meet all of the load, it only appears to be ramp constrained because it has been backed down so far to accommodate the uneconomic GTs. In the real-world dispatch, Potrero 3 would not be ramp-constrained. It would be backed down to accommodate the GTs and as long as the GTs were on-line covering their minimum run time, there would be no need to ramp up Potrero 3 and it would not be even close to being ramp-constrained in the physical dispatch, yet it could be ramp-constrained in the pricing dispatch.

This situation can arise whenever GTs are on the same side of a transmission constraint as the relatively cost steam units that would be backed down to accommodate them. A similar situation could potentially arise in San Diego, and the Cal ISO may be able to think of other possibilities involving Los Angeles area GTs.

The initial change in price calculation logic that the NYISO implemented in late July 2000 was to calculate the upper limit of units in the pricing dispatch based on their dispatch level in the prior pricing dispatch interval, rather than on their metered output. Due to some back and forth with FERC this logic has become more complicated over time.[178]

C. Demand Response

Demand response is power consumption that can be adjusted from a scheduled reference level in response to changes in power prices. Load with appropriate telemetry can set the real-time price under the MRTU.[179] ISO dispatched demand reduction would be paid the real-time nodal price, rather than LAP price and can be pre-dispatched like GTs and would receive a similar BPCG.[180] Demand reduction would settle at the appropriate nodal price in real-time, but would buy power in the DAM at the appropriate LAP price.[181]

Loads only need interval metering and an ability to receive and follow dispatch instructions to supply supplemental energy.[182] Loads providing demand response would be required to demonstrate their effective dispatch capability. Demand response bids would be at nodal level with a minimum size of .1 MW.[183]

The sixth of the major implementation issues identified with the MRTU market design is the proposed mechanism for demand response. Since demand response buys power at the zonal/LAP price in the DAM and sells power back at the nodal price, demand response at nodes within constrained regions have a money machine whenever their actual load is less than their allowed maximum demand response offer. The LSE providing demand response would merely buy power equal to its demonstrated dispatch capability at the LAP price in the DAM and bid demand response at a low enough price to ensure it is dispatched nodally down to its planned consumption in real time, earning the difference between the nodal price and the zonal price for doing nothing. This would be equivalent to the effect of virtual demand purchases at zonal prices in the DAM that are settled at nodal pricing in real-time.

A load's demonstrated dispatch capability is presumably limited by its maximum energy consumption but it may be economic to inflate this if the spread between the LAP and nodal price is material over a large number of hours. The implicit subsidy in buying at the LAP and selling at the nodal price could become expensive to other consumers. This cost could be exacerbated by some of the other market design features, such as the way LAP bids are cleared in the DAM, which would tend to magnify the difference between the DAM LAP price and the real-time nodal price.

Conversely, demand response resources would have little incentive to reduce load at times when congestion is low but prices high. Indeed, demand response loads in unconstrained portions of the transmission system might rarely have an incentive to provide demand response,

---

178 The details are discussed in testimony filed at FERC, as well as in the NYISO tariff. The clearest discussion is in Attachment B, Section IA1to the RTS tariff.

179 CMD # 127.

180 CMD # 127.

181 CMD # 124.

182 CMD # 127.

183 CMD # 128.

as the real-time nodal price would need to rise above the LAP price before it would be profitable for them to respond. If there is material congestion within the LAP, the real-time LAP price could be higher than the nodal price for these loads, diminishing their incentive to participate in such programs.

VI. ANCILLARY SERVICES

A. Procurement

Ancillary services procurement will be included in DAM and hour-ahead scheduling processes. The ancillary services products procured in these markets will be: regulation up, regulation down, spinning reserve, and non-spinning reserve.[184]

DAM scheduling will minimize the total as bid production cost of meeting load and ancillary services requirements. Resources that do not submit ancillary services offers, will not be considered for ancillary services procurement. Ancillary services prices will implicitly include the opportunity cost of providing ancillary services, rather than supplying energy, in the corresponding energy market (day-ahead or real-time, as appropriate). Ancillary services prices will be determined by the shadow price of the ancillary services requirement in the bid load dispatch, Pass 3B. Units committed to provide energy or ancillary services will be eligible for minimum-load cost compensation.[185]

Operating reserve ramp rates for spinning and non-spinning reserves shall be a single ramp rate, distinct from the operational ramp rate. This rate will be fixed throughout the day and can only change in response to a change in capability.[186] The regulation ramp rate shall also be a single ramp rate, submitted with the preferred schedule. A resource must have the same ramp rate for regulation up and down. The regulation ramp rate will be fixed for the operating day and can only change as a result of a change in capability.[187] An RMR unit can declare that its RMR contract ramp rate is equal to its operational ramp rate. If it does not so declare, then its RMR ramp rate will be used as its operational ramp rate.[188]

Participating loads may provide non-spinning reserves.[189] Loads providing non-spin will have relaxed telemetry requirements, 5-minute updates versus 4-second updates for generation.[190]

B. Ancillary Services Requirements

Ancillary services requirements will be determined by the ISO prior to the DAM, based on the ISO load forecast, firm net interchange[191] and anticipated real-time system conditions (i.e., reserves will depend on amount of thermal versus hydro generation scheduled).[192]

---

184 CMD # 48.
185 CMD # 50.
186 CMD # 57.
187 CMD # 57.
188 CMD # 59.
189 CMD # 49.
190 CMD # 127.
191 Actual net firm interchange will not be determined until the end of the bid load dispatch; an assumed level will be used in the unit commitment.

Ancillary services requirements may be determined for sub-areas of the ISO control area, which may result in different ancillary services clearing prices for these sub areas. Suppliers will be paid the appropriate sub-area ancillary services price, but loads would pay the average ISO wide ancillary services cost.[193] Higher quality services would substitute for lower quality services in ancillary services procurement if this reduced the total bid production cost of meeting load.[194]

The ISO will use its load forecast to establish the aggregate ancillary requirement schedule in the DAM.[195] This was approved by FERC.[196] The ISO would procure additional ancillary services in the hour-ahead scheduling process only if needed to supplement the day-ahead procurement due to unscheduled outages, changes in load forecast (which impacts the WECC reserve requirement), changes in the hydro-thermal mix (which impacts the WECC reserve requirement), failure of scheduling coordinators to meet their day-ahead commitment,[197] the dispatch of reserves to meet load, or changes in net firm imports.[198]

The ISO would procure additional operating reserves in real-time if needed to maintain required reserves when ancillary services capacity procured or self-provided in forward markets is dispatched in real-time for energy or is unavailable due to outages. Real-time ancillary services procurement will be part of the intra-hour short-term resource commitment procedure every 15 minutes and will use dynamic co-optimization of energy and ancillary services. Resources would be notified via the ISOs automatic dispatch system.[199]

The current "contingency only" flag for operating reserves (spinning and non-spinning) reserve would be retained, allowing operating reserve providers the ability to opt into the real-time imbalance energy dispatch or stay out of the imbalance energy dispatch to be reserved for contingency situations only.[200] This is a reasonable method for addressing the scheduling of energy limited units.

---

192 CMD # 51.

193 CMD # 51.

194 CMD # 51.

195 May ISO, pp. 4, 66-68, Att A IV.2.

196 See June FERC ¶ 107. The CMD proposed that the ISO could defer satisfying its total ancillary services obligation until the hour-ahead IFM, if it believed its load forecast was likely to change. Deferment was also to allow the ISO to adjust day-ahead ancillary services procurement to account for scheduling coordinator self-provision in hour-ahead market. CMD # 55. The ISO was also permitted to defer purchasing ancillary services in the DAM if it anticipated that the price of ancillary services might be lower in HAM. CMD # 48, 55. FERC approved the Cal ISO deferring ancillary services purchases for the purpose of price convergence, not for the Cal ISO to price discriminate and drive prices apart and proposed permitting suppliers to buy back ancillary services in real-time. Oct FERC ¶ 83. FERC also questioned what it meant for the ISO to defer purchasing ancillary services if it actually committed the capacity in the RUC. Oct FERC ¶ 84. This price based deferral of ancillary services purchases was dropped in the May ISO filing.

197 May ISO, p. 66, Att A IV.2.

198 This was approved by FERC. June FERC ¶ 107.

199 CMD # 56.

200 CMD # 54, Sept ISO, p. 164.

These elements of the market design will not of themselves lead to economic inefficiency, facilitate the exercise of market power nor motivate inefficient bidding strategies. Certain of these elements may cause problems in connection with other, more problematic features of the ancillary services markets, as discussed below.

C. Self-Provided Ancillary Services

Scheduling coordinators will have the option of self-providing ancillary services or relying on ISO procurement.[201] Scheduling coordinators can self-provide ancillary services by identifying units in the DAM that are capable of providing ancillary services and meet the ISO's locational requirements.[202] Resources that would be used to self-provide ancillary services resources must be identified and committed in the DAM.[203]

The Cal ISO position on requiring self-schedule resources to be identified in the DAM is correct. In addition to the reliability issues noted by the Cal ISO, allowing LSEs or scheduling coordinators to decide to self-provide ancillary services after the DAM market is cleared would enable them to self-provide in the hour-ahead whenever they can buy ancillary services cheaper than the price in the DAM, while relying on the DAM purchases when the hour-ahead price is higher. This would just leave other loads with the difference in cost between day-ahead and hour-ahead as the ISO would still need to buy all of the ancillary services in the DAM. Allowing such a practice would be particularly expensive under a single settlement system such as that currently proposed.

Under the MRTU, there will be no direct congestion charges on self-provided ancillary services within California, although the price paid for self-provided ancillary services may differ by location.[204]

D. Ancillary Services Imports and Exports

Under the MRTU, ancillary services may be provided by imports up to limits pre-established by the ISO. Imported ancillary services will require a transmission allocation in the DAM, which means that ancillary services capacity and energy would compete for transmission across inter-control area interfaces. If ancillary services imports contribute to congestion on an intertie, the supplier of the ancillary services import would be charged the applicable congestion usage charge.[205] The congestion charge (the difference between the price paid for internal and external

---

201 CMD # 48.

202 May ISO, p. 67-68 Att A IV.3.

203 See June ISO, pp. 9-10. It is assumed that the value in the DAM of the self-provided ancillary services will be credited against the ancillary services charges of the LSE, i.e., that self-provision will be financial. By "financial," it is meant that the value of the self-provided ancillary services would be credited against the scheduling coordinators' charges for ancillary services. This is similar to the approach for self-provision of losses. If something else is intended, this could be a problem area – in particular, quantity (MW) crediting would not be workable if ancillary service prices vary locationally.

204 Sept ISO, p. 100.

205 CMD #52.

ancillary services) would be the opportunity cost of scheduling ancillary services on the inter-control area interface.

The Cal ISO will support ancillary services exports through an "on demand obligation." Scheduling coordinators would submit on-demand obligations in the scheduling process and these obligations would be added to the scheduling coordinator's overall operating reserve obligation and to the ISO's operating reserve requirement at the relevant scheduling point. On demand obligations can be met through operating reserve imports at that scheduling point or operating reserves procured within the control area. On demand obligations would compete with energy schedules in the export direction in the IFM and may face congestion charges. On demand obligations would not create counterflows for energy in the security analysis.[206]

It should be recognized that the combination of self-scheduling reserves on a unit and scheduling operating reserve exports effectively withholds this capacity from the Cal ISO's congestion management in Pass 3 in a manner similar to self-provided RUC (discussed in Section III.E) or capacity not offered in the DAM under the FOO (discussed in Section VII.A below). Unlike self-provided RUC or capacity not scheduled in the DAM under the FOO, however, this combination of self-scheduled reserves and reserve exports could also effectively withhold the capacity from the Cal ISO's congestion management in real-time if the capacity were designated as contingency only reserves.[207] This possibility would not necessarily cause market problems. However, if a market participant possesses locational market power, such an ability to effectively withhold capacity from congestion management could result in inappropriately high prices within a constrained region. It is assumed that there are limits on the ability of RMR units to engage in such self-provision of ancillary services capacity and designate their capacity as "contingency only." If not, the Cal ISO may want to explicitly address this in future descriptions of reserve export rules.[208] Similarly, the Cal ISO should consider whether there are load pockets in which there is a realistic potential for material withholding of non-RMR capacity needed for congestion management as a result of this ancillary services scheduling process capacity supporting reserve exports.

E. Ancillary Services Pricing

Ancillary services prices in the DAM would implicitly include the opportunity cost of providing ancillary services, rather than energy. Ancillary services prices would be determined by the shadow price of the ancillary services requirement in the bid load dispatch, Pass 3B.[209]

---

206 CMD # 53.

207 We understand that the on demand obligation only requires the Cal ISO to provide reserves on some unit for export and does not allow the MP to designate the unit providing the exported reserves vs internal reserves. Thus, self-provided reserves would be withheld from the real-time dispatch and congestion management only if they were also designated as contingency only. It is not completely clear from the various filings whether it is the Cal ISO or the scheduling coordinator that determines which resources provide ancillary services exports and thus are not available for congestion management in real time. This should be clarified.

208 CMD # 51 refers to limitations on the amount of reserve procured by the ISO within constrained regions but does not appear to cover limits on the amount of self-provided ancillary services within such a region.

209 CMD # 50.

Although ancillary services suppliers may be able to submit capacity offer prices in the hour-ahead scheduling process and would be given advisory schedules, real-time ancillary services procurement and pricing would be in real-time at the 15 minute price.

At present, it is envisioned that there will not be a market clearing price for real-time ancillary services procurement, although this element of the market design is in flux. A resource designated to provide ancillary services in real-time would be paid its opportunity cost (real-time locational energy price minus its energy bid price) plus its hour-ahead capacity offer price.[210] If an off-line unit is committed by the Cal ISO to provide ancillary services, it would be eligible to recover its applicable start-up and minimum-load costs.[211]

The Cal ISO does not propose to pay a market-clearing price for ancillary services procured in real-time; instead each unit's payment for real-time ancillary services will be based on the nodal energy price at its location and its own energy offer price, i.e., its opportunity cost.[212]

The Cal ISO will charge regulating units the nodal energy price for imbalance energy and will not include their opportunity cost in the calculation of the market price of regulation. Regulation suppliers must therefore estimate their opportunity cost and incorporate this estimated cost in their offer prices.[213] This pricing system raises the question of how the Cal ISO would mitigate capacity offer prices for regulation if the offer price needs to account for these expected opportunity costs. This treatment would also lead to a degree of inefficiency in the least cost scheduling of energy and ancillary services in the hour-ahead scheduling process.[214] Why not include the calculated opportunity costs in the corresponding energy market in the calculation of the market price of regulation as is proposed for reserves?

Another element of the MRTU proposal that has been in flux is whether ancillary service suppliers will be permitted to submit capacity bids for reserves procured in the hour-ahead scheduling process. FERC's June Order required the Cal ISO to justify the absence of such capacity bids.[215] The proposed lack of capacity bids in real-time for reserves is consistent with the trend in other markets, such as the NYISO. As the Cal ISO has noted, if capacity is offered to be available in real-time to be dispatched for energy, then it is available to provide reserves. Nevertheless, as noted above one could determine a market clearing price based on opportunity

---

210 If hour-ahead ancillary service offers can include capacity prices, we presume that units would be paid their opportunity cost plus their capacity price.

211 CMD # 56.

212 See Sept ISO, p. 107. This consideration would not foreclose paying ancillary services providers the market clearing opportunity cost.

213 Sept ISO, p. 108-109.

214 If the Cal ISO schedules ancillary services to minimize the cost of meeting load and providing ancillary services, the opportunity cost of regulating units will be directly accounted for in the minimization through the impact on the total cost of shifting a unit from the energy dispatch to providing ancillary services. If the expected opportunity cost is also included in the units ancillary services capacity price, the opportunity cost is in effect double counted. The Cal ISO could try to modify the optimization logic to back this out, but this would have the potential to lead to very inefficient outcomes if the real-world opportunity costs are different from those expected by the regulation suppliers.

215 June FERC ¶105, 106.

costs. One important difference between the Cal ISO's reserve markets and those coordinated by Eastern ISOs is that the Cal ISO permits suppliers in other control areas to participate in its reserve markets. It is not the case for these import suppliers that capacity offered to supply ancillary services would also be available for dispatch by the Cal ISO in real-time if that capacity were not scheduled to provide ancillary services, nor is it true for these suppliers that no additional costs would be incurred in supplying real-time reserves. Suppliers located in other control areas must obtain transmission prior to the hour in order to supply ancillary services to California and capacity committed to provide contingency reserves to California would not be available for dispatch in its native control area. It therefore makes economic sense to permit such import suppliers to submit capacity bids, regardless of whether capacity offer prices are permitted for internal suppliers in the hourly scheduling process.

On the other hand, taking account of capacity offer prices for internal capacity that can either provide reserves or be dispatched for energy would complicate the real-time energy dispatch, as the incremental cost of dispatching a particular unit for energy would include its impact on the cost of reserves. This is a solvable problem and capacity bids will be taken into account in determining energy prices in day-ahead markets, but is an added complication in the real-time dispatch.

An underlying issue raised by the lack of hour-ahead/real-time ancillary services pricing is the lack of a multi-settlement system.

F. Multi-Settlement Issues

The Cal ISO proposes that the sale of ancillary services in the DAM would be a binding commitment. Hour-ahead buyback by suppliers of DAM ancillary services sales would be allowed only in the event of unplanned outages that render the originally sold capacity unavailable. In this case, the seller would buy back its day-ahead ancillary services capacity at the higher of the day-ahead or hour-ahead price.[216]

Suppliers would also be permitted to substitute different resources in hour-ahead for those scheduled in the DAM, if the alternate resources meet the ISO's performance and locational requirements and the unit has not been committed for another use in the DAM (i.e., scheduled to provide energy).[217] The Cal ISO proposal relating to resource substitution was approved by FERC subject to clarification of locational requirements.[218] The ISO would create a "no pay" charge to account for differences between the amount of capacity scheduled in forward markets and the amount available in real-time due to non-provision by the supplier.[219]

The Cal ISO will purchase additional ancillary services in the hour-ahead scheduling process evaluation to replace capacity that is unavailable to provide reserves due to outages or

---

216 See May ISO, p. 66. What is the rationale for charging a high price for unavailable ancillary services if the cost of replacing the ancillary services is low? Why not charge all suppliers that do not provide the scheduled ancillary services in real-time the hour-ahead ancillary services price?

217 May ISO, p. 5, 66,68, Att A IV.4.

218 June FERC ¶ 108,109.

219 CMD # 58.

having been dispatched for energy. There would not be a full two-settlement system however, as ancillary services capacity scheduled day-ahead would keep its capacity payment (apparently including the opportunity cost in the DAM) if it is dispatched for energy in real-time.

This ancillary services market settlement system is workable, but it should be recognized that it has features that would raise the cost of meeting load.[220] First, because it is not a full two-settlement system, ancillary services capacity that is scheduled to provide reserves day-ahead but then dispatched for energy in real-time (such as gas turbines dispatched following a contingency that are then no longer able to provide reserves as scheduled day-ahead or generation scheduled to provide reserves that must be dispatched in real-time to meet load within a load pocket) keeps its capacity payment and the Cal ISO will then pay again to acquire ancillary services in the hour-ahead process. Second, ancillary service procurement would not be reoptimized in the hour-ahead scheduling process; the process will be limited to replacing ancillary service capacity that will not be available in real-time. Thus, capacity that was scheduled to provide reserves in the DAM would not be available for economic dispatch in real-time if this would require scheduling different capacity to provide reserves. In particular, consider the capacity scheduled to provide reserves in the DAM based on its DAM energy offer price. If the supplier reduces its real-time energy offer price to such an extent that it would be infra-marginal in the real-time dispatch, the capacity would not be considered available for economic dispatch in real-time if such dispatch would require reserves to be shifted to units not designated to provide reserves in the DAM. Similarly, low energy offer price capacity that was nonetheless scheduled to provide reserves in the DAM as a result of ramp requirements, would not be available for dispatch in real-time if this would require reserves to be shifted to other units, such as units committed in the RUC. Hence, these constraints on reserve dispatch and pricing result in less than full joint optimization with energy dispatch in real-time.

It is particularly important to be cognizant of the fact that the RUC unit commitment occurs in Pass 4, after ancillary services are scheduled and priced in Pass 3. It is possible that the shadow price of rampable capacity (i.e., reserves) could be much lower in Pass 4 than it was in Pass 3. That is, capacity with relatively low energy prices may be scheduled to provide reserves in Pass 3 while higher cost capacity is scheduled to generate energy in order to meet the ramping requirement for reserves. In Pass 4, additional ramping capacity could be added, and the opportunity cost of the marginal supplier of reserves could be much lower than in Pass 3. Absent other disturbances or restrictions on shifting reserves between units, the availability of lower-cost reserve capacity would likely lead to reoptimization of ancillary services schedules in the hour-ahead scheduling process. Under the MRTU market design, however, there will be no reoptimization of reserves in an hour-ahead scheduling process, the capacity scheduled to provide reserves in the DAM will generally not be available for economic dispatch by the Cal ISO in real-time.[221] These restrictions on reoptimization of reserves in the hour-ahead

---

220 The NYISO relied upon a similar one-settlement system for several years although this was not originally planned. The NYISO interim one-settlement system had some features not present in the MRTU proposal that served to minimize excess ancillary services payments. These features are discussed below. The NYISO's one-settlement system was recently replaced by a two-settlement system.

221 The two exception would be: 1) if real-time demand is lower than forecast day-ahead, some capacity scheduled to provide reserves in the DAM could be dispatched for energy without either reducing the remaining real-time

scheduling process have the potential to raise the cost of meeting load both in the day-ahead market and in real-time, particularly if capacity capable of providing reserves capacity were committed in the RUC. The reoptimization of reserves in the hour-ahead scheduling process would entail double payment for reserves, however, so would be costly absent implementation of a full two-settlement system for reserves.[222]

This structure in which the RUC commitment occurs after ancillary services are scheduled is a noteworthy feature of the ancillary services market design that will likely raise the cost of ancillary services to load. If ancillary services are scheduled before the RUC is run, there is a potential for the shadow price of the marginal reserve capacity to be much higher in Pass 3 than it is in Pass 4 or will be in real-time.

PJM has a RUC commitment that follows its energy market as the MRTU proposes, but PJM does not establish day-ahead ancillary services schedules. The NYISO, on the other hand, determines the DAM ancillary services schedules after its RUC and local reliability passes with all units committed in these passes on-line. While the NYISO rarely commits units in the forecast load pass, it does commit units to meet the ConEd 2nd contingency requirements in the local reliability pass and these additions can radically lower the shadow price of reserves.

It may be in California that shadow price of reserves in Pass 3 will never be as high as it sometimes is in New York due to differences in the Cal ISO reserve requirements and the amount of hydro capacity available to provide reserves.[223] Moreover, if bid load is close to forecast load, there will be little change in the unit commitment between Pass 3 and Pass 4.[224]

---

reserves below MORC nor requiring the purchase of additional reserves; 2) if an individual reserve supplier is able to shift reserve schedules between units meeting Cal ISO ancillary service requirements.

222 Such reshuffling between the DAM and hour-ahead scheduling process would be particularly expensive under the Cal ISO pricing system absent a full two-settlement system as the MRTU apparently would pay ancillary services suppliers the full price of the ancillary services in the DAM if the capacity were dispatched for energy in real-time. The initial NYISO pricing system, which also was not a full two-settlement system, minimized the double payment due to shifts in reserves between day-ahead and hour-ahead by only paying DAM suppliers the market clearing availability payment based on their DAM schedules and then paying their real-time opportunity cost if they actually provided reserves in real-time. This settlement system had problems as well and was only utilized because of the cost of double paying for the opportunity cost of ancillary services. As noted above, the NYISO DAM structure scheduled reserves after the NYISO equivalent of the Pass 4 RUC commitment instead of before it.

In addition to the direct impact of paying opportunity costs in the DAM and again in real-time when ancillary services suppliers change, it may be difficult under the MRTU to minimize this double payment if resources were reoptimized. Under the NYISO system, since a unit was only paid its opportunity cost in real-time the optimization in the hour-ahead simply treated day-ahead availability bids as sunk costs and minimized opportunity costs. Under the MRTU system, the opportunity cost payment is also a sunk cost for the units scheduled day-ahead but the opportunity cost itself is not sunk because it impacts real-time energy prices as well.

223 Reserves tend to be expensive in the New York bid load dispatch when load is low and the NYISO reserve requirement is a larger proportion of total demand, because the spinning reserve requirement is a fixed amount. There are also locational constraints on where the spin must be carried. With spinning reserve defined as a percentage of load, no binding locational requirements, and more hydro, this may not be an issue for the California ISO.

224 Second contingency constraints that are enforced in the unit commitment for Pass 4 but not Pass 3 could also give rise to differences impacting reserve costs.

On the other hand, the combination of the normal changes in conditions between the DAM and real-time and the RUC/DAM structure could result in inefficiently high ancillary service costs in the DAM and either inefficiently high energy prices in real-time (if the Cal ISO does not shift reserves to resources with the lowest opportunity costs in real-time) or double payment of reserve costs. We have identified the issue to raise the question of whether this is likely to be a problem in the Cal ISO context.

An alternative approach would be to utilize a full two-settlement system for ancillary services such as the one the NYISO will implement in February 2005. Such a two-settlement system would eliminate the potential for consumers to pay either reserve opportunity costs or availability bids twice between the day-ahead and real-time markets. While such a two-settlement system obviously entails a second settlement process, it appears that it might considerably simplify the settlement process compared to the complexity of the current proposal. In this regard, it is important to understand that under such a two-settlement system the Cal ISO would buy ancillary services in the DAM at costs that would be borne by consumers, while the price at which ancillary services suppliers buy and sell ancillary services imbalances in real-time would be borne by the suppliers. Under a two-settlement system, consumers only exposure to real-time ancillary services prices arises in the circumstance in which real-time load is higher than anticipated and the Cal ISO purchases additional net ancillary services in the hour-ahead process.

It is important to understand that a implementation of a two-settlement system for ancillary services would not adversely impact the reliability of reserves as the designation of reserves would be determined by the Cal ISO under either the MRTU or a full two-settlement system. Implementation of a two-settlement system would simply reduce the costs to consumers arising from double payment for reserves and allow the Cal ISO to reoptimize reserves in real-time to reduce real-time energy prices without requiring further double payments for reserves.

VII. MARKET POWER MITIGATION

A. Must Offer Obligation

The CMD originally proposed that participating generators must bid or schedule their entire operable capacity into the day-ahead and hour-ahead forward markets; must be available for commitment by the ISO in the day-ahead and hour-ahead RUC, and must be available for dispatch in real-time to the full extent of their operable capacity.[225] FERC rejected this form of must offer obligation ("MOO") in the forward markets (DAM and RUC). Instead, it allowed the Cal ISO to require participating generators to offer capacity either in the DAM or in real-time. This is the FOO (flexible offer obligation).[226] FERC approved a 1/1/08 sunset date for FOO.[227]

Under the FOO, a generator can choose not to bid into the DAM and RUC, and if it sells the energy in another market, it has no further obligation. The generator would be subject to a must-offer obligation in real-time for any uncommitted capacity (i.e., capacity not scheduled to provide energy or ancillary services).[228] If a generator is on-line in real-time and has uncommitted capacity available, the generator is obligated to offer capacity it has not sold into other markets into the Cal ISO real-time market.[229] FERC allowed the Cal ISO to implement the flexible offer obligation if there was no resource adequacy mechanism in effect at the time LMP was implemented.[230]

FERC also accepted the Cal ISO proposal to grant waivers to units that are not bid into DAM and are not needed in real-time or to specify the hours for which they are needed.[231] Generators that bid into the DAM and RUC and are not accepted, are not required to start-up for the next day's real-time market.[232] If generating capacity subject to the FOO is not offered in the DAM/RUC, the capacity must be online in real-time unless it is released by the Cal ISO through the waiver process.

In our view, the FOO is in most respects equivalent to a MOO, because under a MOO a unit could self-schedule generation and schedule a matching export in the DAM, thereby taking its capacity out of the DAM market but leaving itself with a real-time obligation for its unit to be on line. Similarly, under FOO a unit could self-commit its generation, and use it to support an export scheduled in real-time. In either case, the generator would be on line in real-time but the

---

[225] See CMD #3. Under this MOO, the Cal ISO would generate Proxy Bids for Must Offer Resources that failed to bid in DAM/RUC. CMD # 26. Subsequently, the Cal ISO proposed a forward market must offer obligation that sunseted 1/1/08 or with full implementation of CPUC resource adequacy plan. May ISO, p. 16, Att A I.1. The MOO was to be replaced with a real-time MOO 1/1/08 or when CPUC plan is fully implemented May ISO Att A I.5.

[226] Oct FERC ¶227-228; Sept FERC ¶10-11.

[227] June FERC ¶28.

[228] Oct FERC ¶231.

[229] Oct FERC ¶ 230.

[230] June FERC ¶ 27.

[231] June FERC ¶ 29.

[232] Oct FERC ¶ 230.

net capacity would not necessarily be available to the Cal ISO to meet control area load either in the DAM or in real-time. Moreover, under either system, the unit could choose hour-ahead to sell its output into the Cal ISO market in real-time, rather than exporting the power. Under either system it could bid its unit to the Cal ISO's real-time market to be dispatched based on price.

A concern has been expressed that the FOO would result in an inefficient commitment of generation to meet real-time load because of the potential for capacity to not be offered in the day-ahead market but then be made available in real-time to comply with the FOO obligation. This concern is addressed by the Cal ISO's waiver process as noted above. More fundamentally, however, if the Cal ISO's day-ahead and real-time markets function efficiently, suppliers will, on an ex ante basis, find it most profitable to offer their capacity for scheduling in the day-ahead market.

The MRTU eliminates the one-part bids and non-chronological day-ahead market structure that at times deterred market participants from offering their supply in day-ahead markets under the prior market design. If the MRTU market design is successfully implemented and operates as intended, market participants will have no reason to take advantage of the flexibility offered by the FOO. This concern is therefore best addressed by focusing on MRTU implementation and eliminating market design elements that could distort day-ahead market outcomes and thus drive market participants out of the day-ahead market.

The only issue seen with a FOO relative to a MOO regards congestion management. For instance, if a unit located inside San Francisco were committed in the DAM with its net output offset by a matching export, no net capacity would be available to the Cal ISO control area, but the unit's DAM schedule would reduce congestion into the Peninsula regardless of whether the unit was nominally supporting an export or meeting SF load. If the unit did not bid into the DAM under the FOO, however, it would apparently not be modeled as generating energy to meet SF load in the DAM, even though it would be available and would provide congestion relief in real-time. Thus, it is possible that a generator with locational market power could raise prices inside a load pocket by withholding its output from the day-ahead market in this manner (this is very similar to the kind of potential withholding noted above relating to the self-scheduling of reserves in the DAM or self-provided RUC).[233] We presume, however, that generators with material locational market power are condition 1 or 2 RMR units and thus would be available for scheduling in the DAM, regardless of the MOO or FOO.[234]

---

233 This potential for elevated DAM prices would not arise with the proposed nodal clearing of load bids as that clearing process would always produce nodal prices in the constrained region that are lower than those that would prevail in real-time with little or no high cost generation committed in the DAM so high cost generation would not be committed even if it were not withheld. As discussed above, the proposed LAP nodal clearing of load bids and related lack of congestion in the DAM gives rise to other problems that need to be addressed. If the nodal clearing mechanism is replaced with a zonal clearing mechanism that permits consistency between DAM schedules and the real-time dispatch, the potential for physical withholding of congestion management by non-RMR units under the FOO would need to be evaluated.

234 If there are non-RMR units subject to the FOO that are potentially important to congestion management, the Cal ISO could potentially account for their real-time congestion management impact in the DAM by including them in the DAM dispatch but then either increasing the DAM reserve requirement or net exports by the amount that the unit is dispatched in the DAM, so that the RUC commitment is correct. This dispatch would then be zeroed

More generally, the nodal clearing process for LAP bids and the ineffectiveness of virtual load bidding over such broad zones in hedging real-time generator prices (see the discussion in II.D. above) may provide incentives for infra-marginal non-RMR generators in constrained regions to withhold their capacity from the DAM in order to ensure that they are paid the real-time price at their location. The capacity would be available in real-time but these incentives would complicate the ISO's RUC analysis. These potential problems should be significantly lessened if the nodal clearing process for LAP bids is replaced with a zonal clearing process.

B. Start-Up/Minimum-Load Costs

Under the MRTU, start-up and minimum-load cost offer prices may be either cost based or market based.[235] Importantly, a seller choosing to base its start-up and minimum-load offer prices on the administrative cost measure can reduce its start-up and minimum-load offer prices below the administrative cost measure down to zero at its discretion. This ability to reduce offer prices below the administrative cost measure is desirable because it permits resource owners to reduce their start-up cost bids early in the week (to reflect any likelihood that, once committed, the unit will remain on-line the rest of the work week, generating profits beyond the commitment day).

Market-based start-up and minimum-load bids will be fixed for a six month period, regardless of changes in energy prices.[236] The six-month restriction was accepted by FERC.[237] A similar restriction has been in place in PJM since implementation of market-based rates. The original motivation for this restriction in PJM was to avoid inefficient bidding strategies designed to exploit limitations of PJM's original one settlement system. In a competitive market, such a restriction could deter discounting of start-up costs related to the weekly cycle and expected weather conditions. Moreover, choosing to be subject to such a restriction may be so risky for a gas fired generator in today's volatile gas price environment that the market-based option would not be a real alternative for gas-fired generation. It is therefore important that the cost based measure that would be used to cap start-up and minimum-load offer prices accurately reflect actual costs.

Under the MRTU, the administrative cost based start-up offer prices will reflect start-up fuel and electricity requirements times the gas price and electricity price index. The start-up fuel and start-up auxiliary energy consumption will be actual physical parameters of the unit.[238] The physical parameters used for this purpose will take account of the state of the unit, i.e., whether it is warm or completely cold, etc.

It is understood that the gas price used to determine the cost-based start-up and minimum-load costs for the unit commitment and RUC, and to determine the incremental energy

---

out for settlement purposes. This would potentially be complex to implement and would need to be carefully modeled for each such unit.

235 CMD # 19.

236 CMD # 22.

237 Oct FERC ¶ 111.

238 CMD # 20, 21.

price mitigation in Passes 1 and 2 will be the monthly bid-week average price for that month. Units committed in the DAM and RUC based on the minimum-load costs calculated based on bid-week gas prices would then be reimbursed in the settlement process for start-up and minimum-load costs calculated based on the gas price in the day-ahead gas market for the operating day and the day following the operating day.[239] FERC has deferred ruling on 2 day average gas costs based on gas price indexes pending showing that indexes meet FERC criteria, but FERC has stated that the use of a 2 day average of gas price indexes has merit.[240]

The use of a bid-week gas price to determine unit commitment, day-ahead mitigation, and thus the day-ahead dispatch appears problematic and is the seventh of the major implementation issues identified with the MRTU market design. The circumstances in which gas prices tend to change significantly from day to day, and from bid week to the operating day arise in the winter during periods when winter gas balancing rules are binding because the gas supply system is under stress. If there are such large changes in gas prices between bid week and the operating day, it is not desirable to use an artificially low gas price in order to commit, or dispatch electric generating units. In fact, this appears problematic from the standpoint of gas system reliability and perhaps also electric system reliability. If gas prices someday again rise dramatically at the California border because the pipeline system is constrained and gas demand exceeds delivery capacity at the nominal pipeline tariffs, the high gas prices will serve to allocate the available gas to the highest valued uses. If the Cal ISO commits generation based on lower gas prices reflecting outdated bid-week market conditions, this increase in gas demand for electric generation would be inefficient and could potentially crowd out higher valued demands for that gas. In a worst case scenario, the Cal ISO could schedule generation in the DAM based on these artificially low gas prices instead of accepting offers to sell imported power. In real-time, this excess gas consumption by the gas fired generation could unnecessarily undermine gas system reliability by increasing gas consumption and this could in turn threaten electric system reliability if gas fired generation lacking dual fuel capability had to be curtailed to prevent critical drops in gas system pressure (as has happened elsewhere during the winter) and the power imports that were available day-ahead were no longer be available in real-time because the external units were not committed.[241]

Moreover, the use of artificially low bid-week prices to commit generation and mitigate offer prices in the DAM could result in California generation being committed to support export demand for power at artificially low prices in the day-ahead market at times when the California gas market is operating under stressed conditions that would be exacerbated by the uneconomic power exports.

---

239 Generators whose energy price offers are mitigated below the actual gas price would apparently not be made whole in the settlement process. May ISO, p. 52, Att A II.8.

240 June FERC ¶ 48.

241 The Cal ISO states that differences in gas prices may not greatly impact the commitment across gas units, which will be governed by relative gas consumption. As noted in the text, however, use of bid-week gas prices can materially distort choices between the dispatch of California units or imported power or even between units located on different pipeline systems in California and facing different gas prices. Even within a given region, the choice between units with different minimum levels and run times can be materially impacted by the level of gas prices used to calculate minimum-load costs.

It does not appear likely that the use of bid-week prices would serve to inflate electricity prices in a competitive market during months in which gas prices declined following bid week. Market participants selecting the cost based start-up cost option can choose to offer start-up and minimum-load offer prices that are less than the cost based measure, so declines in gas prices between bid week and the operating day should be passed through by suppliers lacking market power. In addition, the mitigation passes in the day-ahead market would of course not impact offer prices if the offer prices are below the reference prices calculated based on bid-week gas prices during periods in which gas prices declined subsequent to bid week.

Suppliers possessing locational market power, however, could presumably take advantage of such a system to submit offer prices at the cost cap based on bid-week gas prices rather than daily spot gas prices during periods in which spot gas prices were lower than bid-week prices. Thus, it appears that for units possessing locational market power, they would be able to submit bids based on bid-week gas prices to capture the difference between spot and bid-week gas prices when bid-week gas prices are higher than spot gas prices and would be compensated in the settlement system based on spot prices when spot gas prices are higher than bid-week gas prices.

There is no ideal solution to this problem because, as the Cal ISO has recognized, day-ahead gas price indexes are not available until after the Cal ISO's day-ahead market for the operating day would be complete.[242]

Overall, market efficiency and gas and electric system reliability will be best served by using gas prices most closely aligned with market prices for unit commitment and dispatch in the DAM. Basing commitment and market power mitigation on two day-ahead gas prices entails a loss of efficiency but would be better than basing these decisions on bid-week gas prices. Further, caution should be used in mitigating the offer prices of gas fired generation lacking market power and not needed to manage congestion during winter periods in which restrictive gas balancing rules are in effect. Thus, mitigating offer prices based on bid-week prices in Pass 2 would be less likely to have adverse impacts on gas and power markets than mitigation of offer prices in this manner in Pass 1. Mitigation of offer prices based on bid-week gas prices in Pass 1 potentially to support power exports would appear much more problematic than mitigation limited to generation that must be dispatched in Pass 2 to manage congestion on local constraints.[243]

Such lags in the determination of the gas prices used for mitigation would be less important in the summer because gas prices generally do not move sharply from day to day. On the few occasions on which gas prices do move suddenly during the summer, as a result, for example of events like the El Paso explosion, it would appear preferable that the Cal ISO use the

[242] This has not been a great problem in New York because most gas fired units have dual fuel capability and thus are mitigated based on oil prices during periods in the winter when gas prices are high, and oil prices usually change more slowly than gas prices.

[243] The potential for mitigation in Pass 2 based on bid-week gas prices of the offer prices of generation possessing market power resulting in the scheduling of uneconomic power exports could be avoided by putting a floor under the Pass 2 offer price mitigation such that no offer prices would be mitigated below the Pass 1 LMP price at that location.

most recent spot prices that would reflect whatever has occurred for the purpose of unit commitment and dispatch, again from the standpoint of economic efficiency and both gas and electric system reliability.

FERC initially ruled that in state gas transportation costs would not be eligible for cost recovery in the minimum-load cost because it would be a demand related cost.[244] The Cal ISO, however, proposed that minimum-load costs would include intra-state gas transportation and municipal use fees.[245] FERC then approved inclusion of auxiliary power costs in start-up costs and inclusion of intrastate gas transportation costs, and municipal use fees in minimum-load cost.[246] The calculation of start-up and minimum-load costs will apparently not include NOx allowance costs or other environmental charges.

The administrative measures of start-up and minimum-load costs should attempt to accurately reflect these costs. It is not in the interest of consumers or market efficiency to cap these offer prices at levels that do not reflect actual costs. If resources are committed at minimum-load based on understated costs and not made whole through energy prices or a bid production cost guarantee, this will adversely impact resource supply by inducing high-cost marginal units to shut down or otherwise withdraw from the DAM. If units are committed based on understated costs but then made whole based on actual costs that are materially higher than those used to choose between competing supplies in the DAM, the cost of meeting load may be increased.[247]

Another issue with the proposed cost-based cap on minimum-load and start-up offer prices, is that it appears that the cost based measure of start-up and minimum-load costs will not account for the minimum-load costs of units started late in the day that will need to remain on into the early morning hours to satisfy minimum run time requirements. It is our understanding that the bid production cost minimization will be applied to the operating day and thus will not include the losses such a unit would incur if it had to stay on line through the night, as a result of a late-in-the-day commitment. The NYISO has had problems with this happening, and has provided increased flexibility in start-up cost bidding to allow generators to factor these costs into their start-up offer prices for late in the day starts.[248] Failure to account of these costs in some manner could lead to peculiar and inefficient outcomes in the unit commitment process.

---

[244] Oct FERC ¶ 112.

[245] May ISO, pp. 39, 52 Att A II.7,8.

[246] June FERC ¶ 46.

[247] Under a resource adequacy system such inefficiency would inflate the costs of resource adequacy contracts to Cal ISO LSEs.

[248] Allowing resource owners to submit higher start-up cost offers for late-in-the-day starts does not allow the exercise of market power because any inappropriately high offers for late-day starts would, in essence, be mitigated simply by starting the unit earlier. Allowing high offer prices for late-in-the-day starts avoids the possibility that units will be inefficiently committed because not all costs were considered.

C. Price Caps

The CMD originally proposed that the nodal prices used for settlement would be capped at the damage control bid cap. Any resulting shortfalls were to be recovered through uplift charges.[249] FERC rejected the $250 cap on each nodal price for aggregation and the proposed uplift.[250] Pursuant to the FERC order, prices will be calculated based on capped bids without further price capping.

The bid cap is "soft" in that cost-justified bids in excess of $250/MWh would be permitted, but would not be used in the market-clearing price calculation. An important implication of the $250/MWh soft bid cap in combination with other elements of the market design is that in the event of a shortage of capacity in the WECC, the LSEs served by the Cal ISO would bear that shortage, as LSEs in other control areas would be able to enter into bilateral contracts for power in the day-ahead market and self-schedule exports that could not be recalled to avoid load shedding within California in real-time.[251] This possibility is one reason for implementing, in conjunction with the current MRTU market design and price cap levels, a resource adequacy mechanism that includes a real-time recall right. Even with a recall right supported by a resource adequacy requirement, a $250/MWh bid cap would tend to concentrate the impact of western resource shortages in California. Such a low bid cap would permit the Cal ISO to be outbid for imports day-ahead and the relatively low real-time imbalance price for imports scheduled day-ahead that do not flow in real-time would also tend to leave the Cal ISO short when outages lead to tight market conditions outside California in real-time.

Under the MRTU, there will also be a soft bid floor of -$30/MWh, but there will be no price floor.[252] The elimination of the originally proposed CMD price floor and replacement with a bid floor is appropriate. The likely motivation for such a price floor is to limit the ability of generators to submit real-time bids that require the Cal ISO to buy back day-ahead schedules at large negative prices in the event that a transmission outage makes the day-ahead schedules infeasible. On the other hand, a price floor on generator prices could also insulate generators at locations having very negative impacts on constraints from the financial consequences of their day-ahead and real-time schedules. Similarly, as observed above in the discussion of self-schedules, there is a possibility that a difference between the large negative price used for scheduling these transactions and a -$30/MWh damage control bid floor could lead to inefficient scheduling practices. It should be kept in mind that if the price falls to -$150/MWh somewhere

---

249 See CMD # 16. This would also have capped the prices charged to transmission customers as well as the price paid to suppliers. This would have been an invitation for inefficient and unintended outcomes and is best eliminated.

250 Oct FERC ¶66.

251 Allowing high cost supplies to be paid more than $250/MWh in Cal ISO spot markets does not address the reliability problem, as if the competitive market price exceeds $250/MWh, inframarginal suppliers would have an incentive under a soft bid cap to export their power so as to be paid the competitive market price. In practice, to avoid shortages, the Cal ISO LSE's would have to pay the market price for all power not covered by bilateral contracts.

252 The negative bid cap is a soft bid cap and entities with cost justification can submit lower numbers. June FERC ¶ 66. These lower numbers will be used in the DAM and dispatch to determine schedules but not prices. These lower numbers will be used in the DAM and dispatch to determine schedules but not prices. CMD #37l, Sept. ISO, p. 141.

while the bid floor is -$30/MWh, this means that every MW of generation injected at the location with the -$150/MW price would cause the Cal ISO to incur more than $150 of out-of-merit costs somewhere else in the system. Large negative prices therefore can serve a useful role in incenting generators to reduce injections that greatly raise the cost of meeting load. In general, the potential for a -$30 price should be sufficiently painful to get the attention of most suppliers and motivate them to submit price-sensitive schedules. The exception would be in the case of a seller that by self-scheduling generation that greatly exacerbates a constraint, the seller could cause the ISO to buy the seller's power at another location at a very high price. It is not known if this kind of scenario is likely on the California grid. However, this kind of circumstance is a reason to have bid floors but not price floors.

D. Energy Bid Mitigation

1. Overview

Under the MRTU, market power mitigation will consist of a) damage control bid cap at $250 for energy and ancillary services capacity (the energy bid cap is a soft cap and sellers can be paid more if cost justified, but higher bids will not be used to set prices; while the bid cap for ancillary services is a hard bid cap); b) soft bid floor at -$30; c) extension of current Cal ISO real-time Automated Mitigation Measure (AMP) to the day-ahead market, RUC and imports; and d) RMR contracts and PJM style mitigation for local market power.[253]

FERC deferred ruling on the mitigation of import offers until after a technical conference and does not appear to have returned to this issue.[254] Mitigating import offer prices based on historic offer prices would simply deter import suppliers from offering price sensitive supply offers at times when their supply offer would be higher than the reference price. This could reduce the elasticity of import supply into the California market during high load conditions, which appears problematic. It would be preferable to not apply mitigation to import or export bids and offers.

2. DAM Mitigation

It is understood that under MRTU, the DAM bid mitigation structure will be as follows:

Pass 1A [Pre-IFM-RMPM-CC]: Market Power Mitigation Pass – Generation committed and dispatched to meet forecast load, only monitoring competitive constraints. Competitive constraints include generation pockets, Path 15, Path 26, interties.[255] Market bids and self-schedules would be used for non-RMR and Condition 1 RMR units. Condition 1 RMR units would be included only if they submit bids. Condition 2 RMR units will not be considered.

---

[253] CMD # 13, 37, 38. This soft bid cap would be problematic if gas prices were to rise to the point that potentially marginal units have costs that exceed the bid cap. This issue is discussed further in Section IX.C.

[254] Oct FERC ¶ 90.

[255] CMD # 40, 41.

Pass 1B [Pre-IFM-RMPM-CC]: Market Power Mitigation Pass –Pass 1A (unit commitment and dispatch) would be rerun with mitigated bids for bids that violate the conduct threshold. If bid mitigation does not produce a material effect on market prices, the offer prices used in Pass 1A would not be mitigated. Otherwise, mitigated bids would be used for all capacity of the mitigated units in Passes 2 and 3.[256]

It is understood that all bids in excess of the conduct threshold would be mitigated if the impact threshold is violated anywhere. The NYISO has found it appropriate to also include a "No Harm" criterion, under which the original unmitigated bids are utilized if the result of mitigation would be to raise the total amount paid by load, regardless of whether prices fall at a few buses. Was there some reason that the "No Harm" test is not included in the MRTU design?

As noted above, generation offer prices could be mitigated in Pass 1B in order to schedule generation to support exports so it is desirable from the standpoint of California load that generation offer prices not be mitigated below the actual level of the gas prices and NOx allowance (or other environmental costs) of California generation.

Pass 2 [Pre-IFM-RMPM-AC]: Local Market Power Mitigation Pass –The unit commitment and dispatch would be run to meet forecast load, enforcing all constraints.[257] The schedules from 1A or 1B would be the starting point:

DEC bids: Accepted bids from generation and import schedules from Pass 1 would be replaced with very high negative penalty DEC bids.[258]

INC Bids: INC bids for condition 2 RMR units would be set at the level specified in schedule M of the RMR agreement up to the full available capacity. INC bids for Condition 1 RMR units would be set at the lower of the RMR price or the submitted market bids. INC bids for all non-RMR units would be mitigated or unmitigated bids from 1A/1B as appropriate.[259] RMR contract start-up and minimum-load costs would also be included for all RMR units not committed in Pass 1.

Under the MRTU, if an RMR unit's schedule in Pass 2 exceeds the unit's schedule in Pass 1, the unit's Pass 3 RMR schedule will be the schedule in Pass 2. The RMR unit's market bid would be retained in Pass 3 above the level the unit is dispatched in Pass 2. Non-RMR units would also be subject to local market power mitigation if their Pass 2 schedule were higher than their Pass 1 schedule.[260] Only the portion of the bid curve dispatched in Pass 2 would, however, be subject to mitigation for the exercise of local market power.[261] The offer price mitigation for local market power would be based on the higher of the highest accepted portion of the bid curve

---

[256] CMD # 40-41.

[257] CMD # 40-41.

[258] "To preserve the relative merit order of these bids, the penalty price will be applied as an adder to the original bid" (i.e., original bid -$10,000). ISO will assign higher DEC penalty prices to Must take and must run resources consistent with treatment of self-schedules. CMD # 42.

[259] CMD # 42.

[260] CMD # 44.

[261] Sept ISO, p. 75, 78.

(in Pass 1A/1B) or the unit's default bid. A unit's offer price would therefore not be mitigated in Pass 2 below the level of its bid accepted in Pass 1.[262] It is not fully resolved how the offer prices of non-RMR generation would be mitigated in Passes 2 and 3. The CMD states that non-RMR generation will be committed and dispatched in Pass 2 based on its Pass 1 bids.[263] If these units were dispatched in Pass 2 based on these unmitigated offer prices, these offer prices would apparently then be subject to mitigation in Pass 3 for the quantity dispatched in Pass 2.[264] This is similar to the offer price mitigation structure applied in PJM. It is important to recognize that under this approach to offer price mitigation if non-RMR units are not committed or dispatched to manage congestion in the local market power pass (Pass 2), their offer prices would not be mitigated in the scheduling and pricing pass (Pass 3).[265]

The rationale for replacing the actual offer prices of generation dispatched in Pass 1A or 1B with large negative numbers in the Pass 2 test for local market power is not apparent. In the current market design, these Pass 1A and 1B schedules have no significance with respect to the market dispatches 3A and 3B. Passes 1A and 1B serve purely to test for the marketwide exercise of market power. Forcing the Pass 1A or 1B schedule into the Pass 2 dispatch in this manner gives rise to the possibility that the congestion pattern and unit commitment analyzed in Pass 2 would be very different from that which would be present in Passes 3A and 3B and in real-time. Such differences would give rise to the possibility that the generation mitigated in Pass 2 would in some circumstances not be committed and dispatched in Pass 3A, while non-RMR and Condition 1 RMR generation not committed or dispatched in Pass 2 (and thus not mitigated) would need to be committed and dispatched in Pass 3A.[266]

The proposed structure of the local market power mitigation pass under MRTU gives rise to three concerns. First, because of this structure it is possible that the local transmission constraints binding in Pass 2 would be different from those binding in Pass 3. Moreover, even if the same constraints were binding, the relative constraint shadow prices could be very different, implying different locational prices. As a result, generation possessing local market power may not be dispatched in Pass 2, could therefore be unmitigated, and potentially able to exercise market power in Pass 3. More likely, generation possessing locational market power could be dispatched in Pass 2 but would be dispatched to a lower level than would be appropriate based on actual market bids and offers in Pass 3 so that unmitigated offer prices would set prices at the margin in Pass 3.[267] Moreover, these circumstances in which mitigation would not be triggered could be quite predictable (and thus capable of being exploited by firms possessing location

[262] Sept ISO, p. 75, 78

[263] CMD #42.

[264] CMD #44.

[265] This feature of the MRTU methodology would likely work as intended if the generation subject to mitigation literally has no substitutes but would operate in a manner that is probably not intended for units competing with high cost alternatives. This is the fourth concern discussed below and it is illustrated in Appendix VIII.

[266] The circumstance in which the unit commitment within the constrained region is different in Pass 3 than in Pass 2 implies the existence of alternatives. This does not necessarily, however, imply the existence of competition as both units could have the same owner.

[267] Restricting the resources that can be committed and dispatched in Pass 3 to those committed and dispatched in Pass 2 would prevent unmitigated units from setting prices in Pass 3 but would further raise the cost of meeting load as if the resources that would be dispatched absent such a restriction are removed, even higher cost resources would be needed to meet load.

market power) as they could arise as a result of consistent differences between the actual offer prices of generation in Pass 1 and the extreme DEC bids utilized in Pass 2.[268]

A second concern is that treating the Pass 1 unit commitment, which is based on a consideration only of the competitive constraints, as fixed in Pass 2, could cause the Pass 2 unit commitment to be quite different from the overall least cost unit commitment. If the market power mitigation in Pass 2 is not based on a least cost unit commitment, there is further reason for concern that units possessing locational market power would be unmitigated and potentially able to exercise market power in Pass 3.

Of particular concern in this regard are the implications of this treatment of Pass 1 dispatch schedules for the scheduling of RMR condition 2 units. RMR condition 2 units are not eligible for dispatch in Pass 1 and Pass 1 schedules are treated as fixed in Pass 2 through the inflexible DEC bids discussed above. The unit commitment in Pass 2 may fail to commit RMR condition 2 units that are the least-cost alternative for managing the constraints that will be binding in Pass 3, because those constraints were not be binding in Pass 2 (as a result of the use of these extreme DEC bids). This would be particularly problematic if the commitment and dispatch of RMR condition 2 units to manage local congestion in Pass 3 were to be limited by the dispatch in Pass 2, which is not clear from the discussion in the various filings. It appears, however, that condition 2 RMR units that are not committed in Pass 2 could not be committed in Pass 3, no matter how extreme the congestion in Pass 3. Restrictions of this type on the commitment and dispatch of RMR units may have been necessary and appropriate within the framework of market separation and in the absence of a unit commitment process but they are not appropriate in the context of an ISO-coordinated day-ahead unit commitment process and could lead to unanticipated outcomes. The extreme DEC bids for Pass 1 schedules are presumably intended to ensure that RMR units are only dispatched to manage local transmission constraint. The fundamental problem with this approach is that the local constraints that are binding in a dispatch based on these extreme DEC bids need not be the same constraints that are binding in a dispatch based on the actual Pass 3 offer prices, so the approach does not necessarily identify the units whose dispatch is needed to manage congestion in Pass 3.[269]

---

268 This potential is illustrated in Appendix IV.

269 More generally, the question should be asked of why it is appropriate to artificially withhold Condition 2 RMR units from the market, even if there is no congestion. The initial answer might be that this withholding is required by the terms of the RMR contracts, but how can there be a FERC approved contract that requires physical withholding of available infra-marginal capacity merely because a particular transmission constraint is not binding?

If an RMR condition 2 unit is the least-cost method of meeting load, then it should be committed in Pass 3. These units are effectively earning a regulated rate of return and should be committed like a regulated unit, when the market price exceeds their cost, regardless of whether the market price is high due to local congestion, congestion on competitive constraints, or westwide shortage conditions. An apparent requirement that all non-RMR generation be dispatched before RMR condition 2 units appears problematic, even absent any market power. Why should prices be potentially set by a high heat rate, high emissions cost unit when load could be met at lower cost by a more efficient RMR condition 2 unit, all of whose fixed operating costs are being borne by consumers?

The eligibility of condition 2 units for commitment and dispatch would benefit consumers because the RMR condition 2 unit would be committed and dispatched only if it were lower cost than the alternative. If the

Finally, a third concern is that the structure of the local market power mitigation pass (Pass 2) will not necessarily preclude the exercise of market power by non-RMR units that are the least-cost method for managing congestion on local transmission constraints but have high-cost alternatives. These high-cost alternatives could either be higher-cost units at a similar location or comparable units at a less favorable location. The significant feature of the Pass 2 mitigation in this regard is that, under this PJM-style mitigation, non-RMR units are subject to offer price mitigation in Pass 3 only to the extent that they are actually dispatched in Pass 2. If there is a high-cost alternative to dispatching a particular non-RMR unit, a unit with inflated offer prices would not be dispatched, or would not be dispatched at the competitive level in Pass 2 if the unit submitted offer prices that exceeded those of its high-cost alternative.[270] In this circumstance the unit with the inflated offer prices would either not be mitigated at all or would only be mitigated over a portion of its bid curve in Pass 3 under the current mitigation structure. In this situation, the unmitigated portion of the bid curve would effectively economically withhold capacity and allow the market price to be set by the offer price of the high-cost alternative.

This lack of effective mitigation for non-RMR units in this circumstance may be appropriate as all units possessing material market power may be subject to RMR contracts and thus dispatched based on their RMR contract price in Pass 2. If this were the case, however, there would be no need for mitigation of the offer prices of non-RMR units and this entire process could be deleted from the MRTU. It is important that all concerned understand that the MRTU market power mitigation mechanism for non-RMR units can only serve to cap prices at the level of the next highest cost alternative, not at the level of the mitigated offer prices of the non-RMR unit that chooses to inflate its offer prices.

An alternative approach to mitigation of non-RMR units that has been considered is New York style mitigation in which there would be two or more passes within Pass 2, allowing a conduct and impact test to be applied. Under a New York conduct and impact approach, non-RMR units would be committed and dispatched based on their unmitigated offer prices in Pass 2A. Offer prices that failed the conduct test would then be mitigated and the commitment and dispatch process would be repeated based on the mitigated offer prices. An impact test would then be applied to determine whether the mitigated offer prices would be applied in Pass 3. If mitigation were applied it would be applied to the entire capacity of the mitigated unit.

The difference between the application of PJM and New York style mitigation when there are multiple alternatives can be illustrated with a simple example. Suppose that within a load pocket there is a steam unit with 400 MW of capacity having a reference price of $50/MW

---

resource owner were allowed to keep any profits arising from such dispatches, both consumers and the resource owner would be better off from relaxing this restriction.

While the FERC's July 8, 2004 order on amendment 60 expresses some concerns regarding the dispatch of RMR condition 2 units, it appears that these concerns arise from the context in which RMR condition 2 units would be dispatched out-of-merit and would not set the market price. It does not appear that the FERC would be opposed to procedures that allowed RMR condition 2 units to be dispatched and set the LMP price like any other unit based on their contractual dispatch price. Indeed, FERC appears to have just the concern expressed here, that this restriction could prevent units that are high cost, but lower cost than the alternatives, from being used to meet load.

270 This potential is illustrated in Appendix VIII.

and 200 MW of gas turbines having a reference price of $150/MW. Under PJM style mitigation if load were 300 MW and the steam unit submitted a market offer price of $200/MW, the gas turbines would be dispatched in Pass 2 for 200 MW and the steam unit for 100 MW. The steam unit would therefore be mitigated to $50/MW in Pass 3 over 100MW and the gas turbines would set the market price at $150/MW.

Under New York style mitigation, the gas turbines would set the price at $150/MW in Pass 2A, while the steam unit would be dispatched to meet all of the load in Pass 2B, setting price at $50/MW. If the $100/MW price difference between Pass 2A and Pass 2B failed the impact test, the entire capacity of the steam unit would be mitigated in Pass 3, setting the price at $50/MW. Of course, if there is no need for mitigation of the offer prices of non-RMR units because units possessing significant location market power have been designated as RMR units, then there is no need for this complexity in the mitigation passes.

These concerns do not reflect fundamental limitations of the MRTU market design but arise from specific avoidable features of the local market power mitigation pass (Pass 2). An alternative structure of the local power mitigation pass that would avoid these problems would: 1) Base the Pass 2 unit commitment and dispatch on the bids used in Pass 1, mitigated as determined by Pass 1B. 2) Repeat the entire unit commitment process in Pass 2, rather than starting with the Pass 1 unit commitment. The schedules determined in Pass 1A and 1B would therefore have no significance, except in terms of determining mitigation.

It has been suggested that the intent of the use of extreme DEC bids in Pass 2 is to allow Pass 2 to dispatch the minimum quantity of INC instructions needed to eliminate the infeasibility of the Pass 1 outcome, so that the dispatch of INC instructions in Pass 2 only solves the constraints imposed in Pass 2, without re-optimizing the complete unit commitment and dispatch. If so, this rationale reflects a very problematic continuation of the "market separation" fallacy which has so plagued the Cal ISO in the past. Perhaps the motivation for this structure is solely the terms of the RMR contracts, but this interpretation of the contracts also appears to reflect the legacy of market separation. Market separation required dispatching generation resources based on a posited distinction between energy and congestion management. However, under an LMP market there is no apparent rationale for applying a contract that physically withholds capacity when congestion is not present even though the price of energy on that uncongested day is higher than on a day when congestion is present. This distinction means that a Condition 2 unit might be available for dispatch on one day when there is congestion and the LMP price at its location is $85 and not available on another day when there is no congestion and the price is $155.

Given the problems in accurately reflecting costs in the determination of mitigated offer prices, there are legitimate reasons for concern about unnecessary mitigation of offer prices. To the extent that Pass 2 is only dispatching RMR units based on contractual dispatch prices, that do not understate market costs inappropriate mitigation should not be a major concern.

The use of extreme DEC prices based on the Pass 1 dispatch, however, does not make sense from the standpoint of market power mitigation and appears instead to reflect a continuation of the market separation doctrine. The Pass 1 dispatch is simply a hypothetical dispatch absent "local" transmission constraints. There is no reason to attach special significant to this dispatch or to constrain subsequent dispatches to be consistent with it. Importantly, it is understood that the dispatch in Pass 3 that determines the actual schedules and LMP prices will be completely unaffected by the Pass 1 dispatch solution. There is no reason therefore to constrain the dispatch of RMR units in Pass 2 based on the Pass 1 dispatch. And, as shown in Appendix IV, imposition of such a the constraint could undercut the Cal ISO's ability to mitigate the exercise of market power by RMR units.

Pass 3A: Market Pass – *G*eneration committed and dispatched to meet bid load. The final DAM dispatch is based on submitted demand schedules and bids. RMR cost based bids will be used for RMR dispatch levels determined in Pass 2. Market and mitigated bids will be used for other units as determined in Pass 1B or 2.[271] Pass 3A determines the DAM schedules.

Pass 3B: Market Pass – Final Pricing Dispatch based on mitigated bids and submitted demand schedules and the unit commitment determined in Pass 3A. Some offer prices included in Pass 3A may not be eligible to set prices in Pass 3B, such as offers covered by the soft bid cap. Pass 3B determines the DAM prices.

Pass 4: RUC Pass – Unit commitment and dispatch to meet forecast load based on a criterion of minimizing incremental commitment costs.

An additional peculiarity of this mitigation structure is that the offer price mitigation analysis is based on unit commitment and dispatch to meet forecast load in Passes 1 and 2, while Pass 3 in which prices are determined is based on bid load.[272] While prices would generally be higher and more units committed in Passes 1 and 2 (which meet forecast load) than in Pass 3, it would not necessarily always the case that the units committed in Passes 1 and 2 would include all of the units that would be the least cost commitment to meet bid load in Pass 3. In particular, it is possible that gas turbines could be the least cost method of meeting load in Pass 3 yet not have been dispatched or subjected to mitigation in Pass 1 or 2.[273]

Another significant feature of this mitigation structure concerns the mitigation of RUC availability bids. It is understood that the RUC availability bids of generation capacity segments dispatched in Pass 2 but not in pass 1 would be subject to mitigation. At present, this mitigation would be based on reference prices determined by past accepted RUC bids, with a $100/MW conduct threshold. The logic behind this mitigation approach is likely that the generation segments dispatched in Pass 1 lack locational market power and generally face substantial competition while generators committed in Pass 2 may possess locational market power.

---

[271] CMD # 45.

[272] CMD # 43. The CMD originally restricted the set of units available for commitment in Pass 3 to those committed in Passes 1 and 2. Such a restriction would serve to further increase the cost of meeting load as units that were not dispatched in Passes 1 or 2 will only be dispatched in Pass 3 if they are lower cost than the alternative, so excluding them would further raise the cost of meeting load.

[273] It was explained in the CMD that this approach avoids sub-optimal commitment decisions that can arise by committing resources in two stages, first in the DAM based on load schedules and bids, and then again in the day-ahead RUC based on the ISO load forecast. In the current DAM structure these passes determine bid mitigation. They do not determine the unit commitment, so this rationale is not applicable.

First, the purpose of offer price mitigation is to prevent the exercise of market power, not to depress offer prices below incremental cost. One of the possible rationales for RUC availability bids and payments is to permit suppliers to recover costs not included in the allowed minimum-load cost bid. But the proposed mitigation mechanism whereby the RUC availability bids of capacity dispatched in Pass 2 would be mitigated to a reference price based on the average of the highest non-mitigated availability bid accepted over some prior period has no relationship to the magnitude of any understatement of minimum-load costs for the purpose of the bid production cost guarantee. If RUC availability bids were constrained only by the $100/MWh conduct threshold, there would be a potential for units possessing locational market power to raise RUC availability bids above a competitive level, although it is possible that in practice all units possessing such locational market power would be RMR units that would not be able to submit availability bids. Second, if the conduct threshold were eliminated or greatly reduced to foreclose the potential exercise of market power by units possessing locational market power, generation capacity committed in Pass 4 could have its RUC availability bid mitigated to minimal levels even if the availability bid were calculated solely to recover the difference between calculated minimum-load cost with their associated bid production cost guarantee and the unit's actual minimum-load cost. Similarly, if the conduct threshold were eliminated, RUC bids designed to recover manning costs for gas turbines or gas availability charges would potentially be mitigated to zero without regard to whether they would merely recover actual incremental costs of being designated to provide RUC capacity.

If generating units scheduled to provide RUC possess locational market power, their RUC availability bids need to be mitigated to prevent the exercise of this market power. The proposed MRTU mitigation mechanism for RUC availability bids, however, makes it difficult to avoid mitigating availability bids that merely recover costs without exempting availability bids that greatly exceed costs.

This potential problem could ideally be avoided if LSEs bid their expected load into the DAM so that there would be no need for a RUC commitment inside a load pocket in which suppliers might possess market power, but the LAP structure and Cal ISO reliability needs may make this outcome unlikely if not infeasible.[274]

The underlying problem is that the current method of calculating the bid production cost guarantee can cause suppliers lacking market power to seek to recover the difference in the availability payment, which greatly complicates mitigation, or even distinguishing cost reflective offers from those reflecting the exercise of market power. Moreover, RUC availability bids that are intended solely to recover minimum-load costs that are not reflected in the Cal ISO calculation (i.e., availability bids that reflect purely competitive behavior by a generator

[274] It might be rational for LSEs to bid less than their expected real-time load into the DAM in the expectation that additional low cost imports would probably become available in real-time. This potential would not exist within a transmission constrained load pocket, however, so LSEs should be motivated to bid their expected real-time load within such load pockets into the DAM, limiting the need for capacity to be committed in the RUC process within load pockets in which some suppliers may possess market power. Unfortunately, the highly aggregated LAPs that are currently proposed for MRTU implementation would prevent LSEs from bidding their full expected real-time load into the DAM within such transmission constrained load pockets and a bidding a lower proportion of their expected real-time load into the DAM for regions that they expect will in part be served by real-time imports.

committed in Pass 1) could set a market clearing price that would be paid to a substantial amount of generation that does not face such a short-fall in recovering its minimum-load costs. In Pass 4 the Cal ISO would select the least costly RUC capacity, so resources that have high RUC availability bids because they would be substantially under compensated by the RUC uplift formula might not be selected in the RUC pass and thus would not determine the market clearing RUC availability price, but this outcome is far from assured in practice. Units might submit high RUC availability bids precisely because the BPCG calculation would substantially understate their actual costs, yet have total commitment costs that are infra-marginal.[275] Overall it appears that the better course would be to fully compensate generation committed in the RUC for their start-up and minimum-load costs and to either completely eliminate the RUC availability bid or substantially reduce the costs that suppliers would need to seek to recover in the RUC availability charge to costs that are at least roughly measurable (i.e., gas turbine manning costs and possible gas availability costs during winter conditions).

3. Hour-Ahead Mitigation

Under the MRTU, the Cal ISO may undertake additional of offer prices mitigation in the hour-ahead scheduling process. First, the Cal ISO would determine any incremental RMR requirements based on SCUC and dispatcher knowledge, insert the RMR price for these incremental quantities, and notify RMR owners through an "ex post dispatch notice." Second the Cal ISO will perform a dispatch based on forecast load to determine the application of any global or local market power mitigation.[276]

It is not explained in the discussion of real-time mitigation what gas price information would be utilized for such real-time offer price mitigation. While mitigation would ideally be based on intra-day gas prices, the intra-day gas market is thinner than the day-ahead market. Moreover, power market offer prices may be based on expected gas purchase prices, while actual intra-day gas market transactions may occur after generation has been dispatched (i.e., the transactions replace gas that is burned so as to cover DAM schedules later in the day).[277]

---

275 These RUC availability bids could become extremely problematic if the unit commitment in Pass 4 were restricted to the units committed in Passes 1 and 2 as originally envisioned in the CMD. Passes 1 and 2 dispatch generation to meet load based on energy bids alone, so the RUC availability bids are never considered in the commitment process. Imposing such a restriction on the unit commitment in Pass 4 could force acceptance of very high RUC availability bids in circumstances in which much lower cost RUC was available on other units. Such a disconnect between the commitment criteria in Passes 1 and 2 and the clearing process in Pass 4 appears likely to encourage suppliers to submit inflated RUC availability bids, even suppliers that would lack any market power in an efficiently structured market.

276 The CMD proposed a number of dispatches to trigger and apply market power mitigation. CMD #39-46.

277 Thus, generators may submit offer prices based on the expected intra-day gas price but not purchase gas until they have actually been dispatched sufficiently far above their day-ahead schedules for a sufficient number of hours to warrant such purchases. The purchase prices may therefore be observable only after the time that mitigation would be applied. The issue of intra-day gas pricing is really a winter balancing rule issue and perhaps the Cal ISO can simply mitigation by setting up different mitigation systems for when a generator is subject to winter balancing rules and when it is not. If summer balancing rules are in effect, perhaps FERC would permit the Cal ISO to mitigate based on some ratio like 1.05 percent of the prior day price on the premise that spot gas prices normally do not change dramatically from day to day in the summer. Unfortunately, in these

The final hour-ahead run will be based on demand bids and schedules, utilizing any additional bids mitigation and incremental RMR dispatch resulting from 1A, 1B. This final hour-ahead run would determine GT start decisions, the schedules of off-dispatch units and import/export schedules.[278]

The Cal ISO will dispatch the system in real-time based on the offer price mitigation determined in the hour-ahead process. There will be no additional real-time mitigation following the hour-ahead scheduling process.[279]

E. RMR

RMR units will be committed and dispatched in the DAM based on the lower of their submitted market based offer prices or the RMR contract variable cost to the extent that RMR resources are needed to manage local transmission constraints. These offer prices will be used to set market prices.[280] This use of RMR offer prices to set market prices is appropriate.

Condition 2 generating units under the RMR contract are prevented under the RMR contract from market transactions absent an RMR dispatch, although they can be called as PGA units in an out-of-market transaction or System Emergency circumstances.[281]

As discussed at length above, there are good reasons that all RMR units should be modeled in the unit commitment and dispatch in Pass 3 based on the costs and dispatch prices in their RMR contracts and their contractual offer prices should be permitted to set LMP prices.

F. LAP Structure

There may be circumstances in which vertically integrated LSEs have market power in managing congestion on some local constraints because they are the only entity owning generation within their service area. Under a nodal pricing system, vertically integrated LSEs with generation and load at the same location will typically be unable to benefit from the exercise of market power because raising the offer prices of their generation will also raise the LMP price at which they

---

days of volatile international oil prices that might not always be valid and gas prices could change by more than 5 percent in response to events like the El Paso explosion.

The winter intra-day gas price for mitigation would be harder to measure. An upper bound could perhaps be calculated based on the winter balancing rule penalty, but this would be too high on days when intra-day gas demand is lower than expected day-ahead.

278 CMD # 46.

279 The CMD proposed that in real-time the ISO would determine any incremental RMR requirements based on SCUC and dispatcher knowledge, insert the RMR price for these incremental quantities, and notify RMR owners through an ex post dispatch notice. The ISO would then perform a real-time pre-dispatch based on forecast demand to determine the need for any additional global and local market power mitigation. Any additional bid mitigation and RMR dispatch is incorporated in the real-time market. CMD # 47. This structure appears to have required multiple passes in the real-time dispatch which would have complicated real-time operations. This complexity is avoided with the simplified H/A market structure.

280 CMD # 146.

281 Sept ISO, p. 115.

buy power so there is no net impact.[282] Under the proposed LAP pricing system, however, such entities may have an incentive to exercise market power because the LAP price they pay for power could rise much less than their increase in generation revenues, were they to raise their offer prices. These incentives would be eliminated if the vertically integrated LSE held CRRs from its generation to the LAP, but in the circumstance that the generation price exceeds the LAP price these would be counterflow CRR obligations that the LSE would not voluntarily agree to hold, as discussed in Sections II and VIII. Such LSEs would seek to hold CRRs, but they would seek to hold CRRs from low priced locations to the LAP, not from high priced nodes at which the price is impacted by their generation offer prices.

This potential incentive for the exercise of market power by vertically integrated LSEs already exists to some degree under the current zonal structure to the extent it has not been mitigated by RMR contracts. It is not currently proposed to assign LSEs counterflow CRRs from high-price locations to the LAP pricing zone under the MRTU. As a result of the application of aggregate LAP pricing zones without the assignment of counterflow CRRs, there could be greater benefits to the exercise of market power by such vertically integrated LSEs under the MRTU market design than is presently the case. This could exacerbate existing market power problems or give rise to problems where none currently exist.

G. Conclusion

There are several elements of the MRTU market design that could potentially permit the exercise of locational market power. Whether any of these possibilities will give rise to substantial problems cannot be assessed in the abstract but depends on whether there are non-RMR suppliers that possess material locational market power, how some of the critical deficiencies of the MRTU market design (such as the nodal clearing process and settlement rules for zonal/LAP bids) are addressed, and the mix of unit characteristics within load pockets.

[282] It would still be necessary to carefully monitor their offer prices for efforts to shift costs onto other LSEs through uplift charges.

VIII. CONGESTION REVENUE RIGHTS (CRRs)

A. CRR Definition

Under the MRTU market design CRRs will be source-to-sink congestion hedging financial instruments. In general, the sinks and sources for CRRs may be single network nodes or sets of nodes, such as trading hubs or load aggregation zones (LAPs). As in PJM and New York, CRR holders will be paid the difference between the congestion component of the LMP price at the sink and source of the CRR in the day-ahead market.

CRRs will be defined primarily as "obligations," so that the CRR holder will be obligated to make a payment when the congestion price difference is in the opposite direction of the CRR.[283] An exception is that ETC rights holders who convert their rights to CRRs will be offered a choice between CRR options and obligations in the initial CRR allocation.[284] In addition, some Cal ISO proposals appear to contemplate the award of options to parties that fund merchant transmission expansions and to new PTOs.[285]

Documentation and data should be available at this time to allow the Cal ISO to assess the technical feasibility of offering CRR options. Software is available to perform a CRR auction and simultaneous feasibility test for a combination of CRR options and obligations using a DC model, and PJM has been offering FTR options and obligations in its auctions for over a year. Unless the Cal ISO foresees a need to use an AC Optimal Power Flow (OPF) for the simultaneous feasibility test, the Cal ISO would be able to use the same or similar software to implement the testing of options that is proposed in CRR Study 2. Thus, information should be available to allow assessment of the technical feasibility of awarding CRR options, as well as the quality of the approximations that would be used in implementing a simultaneous feasibility test for CRR options on the grid coordinated by the Cal ISO, to the extent that these have not already been determined.

Aside from the choice between a DC and AC OPF, the other technical issue that could arise in implementing CRR options in California is a likely increase in the solution time for the CRR auction software. When and if CRRs options are offered, this will likely cause an increase in the number of bids and in the number of transmission constraints binding in the CRR auction solution, which may increase the number of iterations required to reach a solution. This impact will need to be anticipated and tested, taking into account the size of the California transmission model and the number of contingency constraints that are monitored. A number of solutions

---

283 FERC has encouraged the Cal ISO to develop additional types of CRRs, such as "options" if they are valued by Cal ISO market participants. Oct FERC ¶177. The Cal ISO has generally taken the view that CRR options may be offered in the future once the market gains experience with obligations, and following a determination of the technical feasibility of implementing CRR options and a showing that the benefits of offering options are demonstrably greater than the costs. Sept ISO, p. 91.

284 CMD # 78.

285 See CRR Study 2 at pp. 8 and 17 and "Draft Proposal for the Allocation of Congestion Revenue Rights to Merchant Transmission," August 6, 2004, hereinafter Cal ISO MT, p. 3. The issues that arise from the choice of whether to allocate options or obligations to various groups of market participants (holders of ETC rights, new PTOs, non-ETC LSEs, etc.) are discussed in Section C.

may be considered to address concerns about the model solution time, including PJM's approach of limiting the set of allowed source and sink locations for FTR option bids. Thus, the technical issues that need to be investigated prior to implementing CRR options appear to be reasonably well identified: the quality of the software approximations, and the software solution time. If the Cal ISO offers options to ETC rights holders who convert their rights to CRRs, to new PTOs or to parties that fund merchant transmission expansions, implementing the simultaneous feasibility test for these awards will require the Cal ISO to implement the simultaneous feasibility test for options. In this event, the incremental costs of testing and implementation for offering options more broadly, such as in the CRR auctions, would be low.

The Cal ISO plans to offer CRRs of two term lengths, annual and monthly, and distinct CRRs will be issued for the on-peak and off-peak periods.[286] Annual CRRs will be available on a rolling two-year basis, i.e., each annual release of CRRs will make specified quantities of CRRs available for each of the following two years; these quantities may differ for year 1 versus year 2. The ISO will release a fixed percentage of the transmission capacity as annual CRRs for a particular operating year, after accounting for the impact of ETCs on the available capacity of the grid. The Cal ISO has proposed to use 75 percent of transmission system capacity to support annual CRRs, and 25 percent of capacity to support monthly CRRs; these proportions may be adjusted based on the outcome of the Cal ISO's CRR Study 2 study and LMP testing. Thus, the Cal ISO will use 75 percent of the transmission capacity to support annual CRRs for the first year and, separately, a half (37.5 percent) of the capacity to support CRRs for the second year. In the beginning of the second year the difference between 75 percent of transmission capacity at the time and the amount of capacity released in the first year will become available to support additional annual CRRs.[287] The volume of CRRs released for the following year (third) would be based on 37.5 percent of the relevant transmission capacity estimated for that year. The proposed procedures will enable participants to obtain CRRs that are valid for the first year following the allocation procedure as well as, separately, CRRs that are valid for the second year following the allocation procedure.

The current proposal and procedures provide a reasonable approach to making both short-term and somewhat longer-term CRRs available to market participants. The availability of CRRs for up to 2 years in the future exceeds the term for which FTRs are currently available in PJM. Moreover, the procedures for releasing longer-term CRRs could readily be extended to even longer terms (e.g., 15 percent of forecast capacity could be made available for year three) if this were determined to be desirable in the future. The availability of one-month CRRs matches the shortest term available in PJM and New York.[288] The proposal to initially release 37.5 percent of forecast year 2 transmission capacity as year 2 CRRs, 75 percent of forecast year 1 transmission capacity (including long-term outages) as year one CRRs and 100 percent of monthly transmission capacity (including longer-term and shorter-term outages) as monthly CRRs appears to be reasonable. The Cal ISO's approach to reserving some transmission

---

286 In CRR Study 2, all CRRs had a one-month term and annual CRRs were strips of 12 monthly CRRs. CRR Study 2 pp. 4, 5. "[A]n annual term CRR is comprised of potentially 12 one-month CRRs with the availability of these CRRs being one year." CRR Study 2, p. 5. See also CMD # 91 and # 79.

287 CMD # 91.

288 No region has yet offered weekly CRRs auctions, although some market participants have requested such an enhancement to improve their ability to implement short-term changes in their hedging positions.

capacity for allocation and auction monthly, in order to accommodate the impact of short-term transmission outages on available transmission capacity, is also reasonable. The intention is to reserve sufficient capacity when determining the annual awards that transmission outages not included in the transmission model used for the annual auction can be included in the monthly transmission model without causing infeasibility of previously awarded annual CRRs. The Cal ISO has indicated an intent to adjust the 75 percent/25 percent conceptual target if appropriate.[289]

The CMD proposed that scheduling priority would apply to the demand side of CRR schedules.[290] The FERC has determined, however, that the Cal ISO has not justified the provision of energy scheduling priority to the holder of a CRR, so this element of the CMD has been dropped from the MRTU and is not considered in this discussion.[291]

B. CRR Sources and Sinks

Under the MRTU, a CRR source location may be a single injection node, an inter-tie point or a Cal ISO-defined trading hub, including the load aggregation points used for pricing wholesale power purchases. CRRs allocated to LSEs will have a sink location corresponding to one of three standard LAP zones: PG&E, SCE and SDG&E. These LAP zones must be used as CRR sinks for CRRs allocated to LSEs but not for the CRRs used to reserve capacity for ETC loads. The Cal ISO contemplates using seasonal load distribution factors for on-peak and off-peak periods to assign load to nodes within these LAP zones for purposes of the CRR Study 2 simultaneous feasibility test.[292]

In addition to CRRs with a pre-determined source specified by the requestor the Cal ISO also will offer Network Service CRRs (NS-CRRs) that are to provide an LSE with "an optimal congestion hedge at least cost."[293] To obtain a NS-CRR, an LSE will specify a set of injection nodes or inter-ties and assign nodal quantity bids or priorities to indicate its preferred distribution of CRRs over these nodes, as well as acceptable adjustments in case the preferred distribution is not feasible. The CRR allocation procedures will provide the preferred distribution, if possible, or can optimize the distribution. Once a NS-CRR is issued the distribution factors for the

---

289 The Cal ISO states that 75% annual/25% monthly are conceptual targets at this time because "concrete data are not yet available upon which to base a determination of optimal release quantities." CMD Transmittal, 77. A number of considerations factor into the decision concerning the percentage of the transmission system to allocate and auction as CRRs for different future time periods, in addition to the goal of maintaining simultaneous feasibility that is discussed in CRR Study 2. While contracting flexibility is valuable to LSEs in adapting to changes in market conditions, LSEs seeking to enter into long-term energy contracts to hedge their cost of meeting load, perhaps by supporting the construction of new generation resources, need a mechanism that allows them to acquire CRRs, or an entitlement to CRRs, having a similar duration.

A two-year allocation horizon for CRRs may provide a reasonable balance for the initial implementation of the MRTU market design, enabling LSEs to observe how the market operates during this start-up period. Once the market is implemented and operating, however, the Cal ISO needs to implement a mechanism that would make at least a portion of transmission transfer capability available to support long-term CRRs.

290 CMD # 89.

291 Oct FERC ¶184.

292 CRR Study 2, p.5.

293 CMD # 88.

injection nodes are fixed. NS-CRRs subsequently may be unbundled into single injection node CRRs, consistent with the distribution factors defining the NS-CRR.

While the proposal to offer NS-CRRs appears to be reasonable, it is not clear how the distribution of each NS-CRR over a set of source nodes could be optimized in the simultaneous feasibility test run for CRR allocation. Specifically, it is unclear how the software can optimize based on the adjustment bids for the distribution of NS-CRR rights over injection nodes at the same time that it is optimizing the allocation of different CRRs based on the rankings that will be used to distinguish the priority of non-converted ETC rights, converted ETC rights, and non-ETC rights (4 levels). It appears that the CRR allocation optimization will need to be based on a single cardinal preference/priority ranking for each CRR.[294] If the availability of optimized NS-CRRs is important to particular market participants, this feature of the market design should be retained if it can be implemented. If it does not have strong support, it is a complication that could be eliminated; neither PJM nor NY offers the equivalent of optimized NS-CRRs.[295]

In addition to the problematic features of the nodal clearing mechanism for LAP load bids discussed in Section II.B, the use of aggregated LAP zones has two undesirable features from the standpoint of CRR allocation. First, the use of highly aggregated load zones to define the load zone sinks for LSE CRRs is likely to lead to unsatisfactory tradeoffs between Cal ISO revenue adequacy and hedging by LSEs. The Cal ISO has recognized that when a simultaneous feasibility test is performed for CRRs defined to broadly aggregated load zones, the resulting set of feasible CRRs is likely to understate the actual ability of the existing transmission system to hedge congestion.[296] This understatement would occur because transmission constraints within the aggregated load zone can result in differences in the proportion of load that can be met with imports across different areas within the aggregated load zone. In essence, when CRRs are defined to the LAP, the most limiting transmission constraint into any sub-region of the LAP limits the quantity of CRRs that can be awarded from a given source to the LAP. A result of this underallocation is that relatively less of the congestion rents accruing under the MRTU market design would be allocated through CRRs to LSEs serving load within the LAP and relatively more would accrue as excess congestion rents. These excess congestion rents would be allocated to the PTOs on a pro-rata basis in proportion to their overall transmission revenue requirement, and would then serve as an offset in the determination of the transmission access charges paid by load. If there are constraints within the aggregated load zones, the allocation of CRRs to aggregated load zone sinks will make it difficult or impossible to assign all of the congestion rents to loads in the form of useful CRR hedges.

The potential for underallocation of CRRs defined to broadly aggregated load zones has been observed in the operation of the NY markets, where the allocation of ETCNL was defined

---

294 In CRR Study 2, the Cal ISO stated that the NS-CRR software functionality would probably not be available for the second CRR study.

295 The perceived need for NS-CRRs appears to be related to the specific rules used in the process through which LSEs submit requests for CRRs for each of the four proposed sub-priority levels, as discussed in Section E. Thus, if the requests for priority level 2 CRRs can be adjusted after information is made available concerning the level 1 awards made to each LSE (and so on), then each LSE will have more ability to shape the portfolio of injection nodes in its final CRR allocation, decreasing the need for NS-CRRs.

296 Feb CRR Study 2, p. 14.

to load zones. ETCNL in New York is essentially a source to sink auction revenue right for native load, meaning that it is a financial right funded by the TCC auction revenue, where the payment is equal to the auction value of a TCC from the ETCNL source to sink.[297] Because the auction revenue paid to ETCNL is the same as that paid for the corresponding TCC sold in the auction, the NYISO applies a simultaneous feasibility test to the ETCNL prior to each six-month auction; this tests the feasibility of the proportion of initial ETCNL allocations being valued in that auction, in combination with all other unexpired TCCs and rights.[298] In the simultaneous feasibility test, ETCNL is modeled identically to TCCs; thus, 100 MW of ETCNL from Niagara to zone J would be modeled as a 100 MW injection at Niagara and a 100 MW withdrawal distributed over the nodes in zone J. If the ETCNL and unexpired TCCs and rights are not simultaneously feasible, the ETCNL is pro-rated to restore feasibility.

In New York, the load zones do not correspond to the retail service territories of the original transmission-owning utilities, but are more narrowly-defined regions. Hence, the ETCNL for ConEdison load is defined to sink in either zone H, I or J, rather than in a broadly defined ConEdison zone. Nevertheless, in applying the simultaneous feasibility test, the NYISO found that the zonal definition of the ETCNL allocations would have led to a material decrease in the feasible ETCNL into the New York City zone (J) as a whole due to transmission constraints that limit the delivery of power to load pockets within zone J. To avoid the consequent unnecessary pro-rationing of ETCNL, the NYISO applies the simultaneous feasibility test for ETCNL to each of the separate nodal sinks for the ETCNL, so that ETCNL to a zone as a whole does not need to be pro-rated if the infeasibility exists for only the portion of the load that is situated within a load pocket. Revenue adequacy is preserved because the NYISO then calculates the value of such node-to-node ETCNL in the six-month auctions based on the nodal sink quantities and prices for the ETCNL, not based on the load zone price and nodal sink quantities. It is worth noting that the need for disaggregation of sink locations has existed in New York, even though the load zones are defined at a more granular level than the retail service territories proposed under the MRTU. Additional insight into the potential for underallocation of CRRs as a result of highly aggregated LAP load zone definitions is provided by the illustrations in Appendix VI, Section H.

The Cal ISO's comments regarding CRR Studies 1 and 2 are consistent with our expectation that the determination of the number of CRRs that can be awarded under the simultaneous feasibility test can be sensitive to load aggregation. In the initial CRR allocation

---

297 Because the ETCNL has the same value as the corresponding TCC (CRR), transmission customers that are credited with the ETCNL revenues receive the same ex ante value that they would receive if they were allocated the corresponding TCC (CRR). Allocating ETCNL is therefore financially equivalent to allocating CRRs but better accommodates the sale of long-term CRRs.

298 There is a direct analogy between the simultaneous feasibility test run on a set of CRRs/TCCs to be settled in a day-ahead forward market, and the simultaneous feasibility test run on a set of auction revenue rights and CRRs/TCCs that are valid at the time of an auction. Infeasibility of the CRRs/TCCs that are settled in the day-ahead forward market may mean that there is revenue inadequacy, i.e., insufficient congestion revenue will be collected in the forward market settlements to pay the congestion rents due to the holders of the CRRs/TCCs. Similarly, if the outstanding auction revenue rights and CRRs/TCCs are infeasible at the time of an auction, it means that the auction may be revenue inadequate, in which case there would be insufficient auction revenue to pay the source to sink auction price to all holders of auction revenue rights and holders of CRRs/TCCs that sold their rights in the auction.

study, the Cal ISO attempted to address this limitation of aggregated load zones by breaking some of the LAP regions down into smaller load groups for the purpose of the simultaneous feasibility test in the CRR allocation process.[299] It is proposed that this would also be done in the second study.[300] As explained above, the approach of defining CRRs to load groups that are smaller than the LAP zones is reasonable if there are transmission constraints within the large LAP zones.[301]

However, the approach of disaggregating the load zones into smaller load groups for the purpose of applying the simultaneous feasibility test and awarding CRRs is only consistent with the revenue adequacy of Cal ISO settlements, however, if CRRs are awarded to the same disaggregated load group regions to which the simultaneous feasibility test has been applied, and thus CRR payments in the day-ahead settlements are calculated based on the price for the disaggregated load group regions. If CRRs that are feasible only if defined to specific nodes or smaller load group sinks were nevertheless awarded to sink at the broadly aggregated LAP zones, and CRR payments were made based on the aggregated LAP zones' prices, the Cal ISO would be awarding CRRs that would not satisfy the simultaneous feasibility test, and would be obligated to payments that would preclude revenue adequacy. It appears that this is the approach proposed for implementation under the MRTU, although the matter continues to be discussed and revenue inadequacy will be assessed during CRR Study 2.[302] This approach to testing simultaneous feasibility will inherently award source to aggregated load zone CRRs that are in fact infeasible if defined to the LAP zone, rather than to the smaller load group regions. Preservation of revenue adequacy in the award of CRRs requires that if the simultaneous feasibility test is applied to CRRs defined to disaggregated load groups, then the CRRs awarded are also defined to these same disaggregated load groups.[303] Further explanation of the potential

---

299 CRR Study 1, p. 28.

300 CRR Study 2, p. 14.

301 "Without breaking down the load aggregation areas into load groups, any downward adjustments made to bid injections at the nodal level by the SFT necessary to achieve simultaneous feasibility could translate into major curtailments of CRRs at the higher load aggregation level since the load distribution factor associated with each injection or withdrawal is fixed." CRR Study 2, p. 12. As the Cal ISO observes: "The purpose for disaggregating the four load aggregation areas into smaller load groups was to alleviate constraint violations encountered during the SFT in a more efficient manner and thus allow a larger number of CRR MW Obligations to 'clear' the market."

302 The Cal ISO comment in the initial study that "[t]he resulting cleared bids were subsequently 'reassembled' to arrive at the total quantity of cleared bids from the original source to the original load aggregation areas" leaves some ambiguity as to what was done and whether the final awards satisfied the simultaneous feasibility test. It was unclear from the original description whether the reassembly was to be limited by the extent that there were sufficient cleared bids to each of the smaller load groups composing the larger areas, with some remaining CRRs defined with the smaller load groups as sinks, or whether all of the awards to the smaller load groups would in some manner be transformed into awards to the larger load aggregation regions. Subsequent discussions have clarified that the Cal ISO intends to apply the latter approach.

303 The amount of the revenue inadequacy resulting from CRR reaggregation can be calculated as the difference between the CRR congestion rents paid for the CRR set defined to load groups and the congestion rents paid for the reassembled CRR set defined to the aggregated load zones. Estimates of this difference could be calculated from the Cal ISO's CRR and LMP studies. Irrespective of such estimates, however, it would not be advisable to implement a market design that builds in infeasibility through the reassembly of CRRs. Such a system risks substantial pro-rating of CRR congestion rents in the future, which is a cost that would be borne by all CRR holders, not just the LSEs benefiting from the reaggregation. If such an approach were adopted, a separate

for revenue inadequacy under the proposed allocation methodology is provided by illustrations in Appendix VI, Section I.

To achieve the goal of maximizing the extent to which LSEs can hedge themselves against congestion costs through CRR ownership while maintaining Cal ISO revenue adequacy, it is desirable that all load zones be sufficiently disaggregated for the purpose of the simultaneous feasibility test to assure that the award of CRRs is not materially limited by constraints internal to the load zone, and it is necessary that CRR awards and settlements be based on the same disaggregated sink definitions that are used for the simultaneous feasibility test. Moreover, to avoid unintended cost shifting it is also desirable that all load zones be treated symmetrically.[304] As disaggregation increases, so does the quantity of feasible CRRs that may be initially allocated to the LSEs serving load within a zone. If different standards are applied to disaggregation of CRRs sinking in different LAPS, cost shifting may occur because the LSE serving load in a less disaggregated zone would potentially receive fewer of the (nodally) feasible CRRs in its allocation than would an LSE serving load in a more disaggregated zone, and the residual auction revenues and congestion rents resulting from the unallocated transmission capacity for the less disaggregated region would be shared with the LSEs serving load in other regions through a proportional credit against the PTOs' transmission revenue requirements.

A second problem that arises from the use of aggregate load zones in the context of CRR allocation is that it leads to differences between the load zone price and generator price for generation and load at the same location. These artificial price differences will impede hedging by LSEs serving load with their own generation, lead to cost shifts among LSEs, and exacerbate the underallocation of CRRs arising from reliance on aggregate load zones for pricing. The Southern Cities previously observed that if the Cal ISO bills their loads (which would be within the SCE load aggregation zone) based on the average price for the entire SCE load aggregation zone, it is unclear how they would receive the economic value of their resources.[305] The Southern Cities' apparent concern was that LSEs serving their load with their own generation at the same location as their load could be required to pay congestion charges from their generation to the aggregated load zone location if their generation is located at a lower priced location than the LAP zone as a whole, even though there is no actual congestion impact in meeting load with generation at the same location. The assessment of such congestion charges would lead to a potential cost shift.

The Cal ISO has pointed out that the potential cost shift for LSEs serving their load with generation at the same location that could be associated with aggregated LAP pricing could be avoided by awarding such LSEs CRRs from their generation to the aggregated load zone location. While such an approach might substantially reduce the cost shifts associated with introduction of aggregated LAP pricing, this approach could also lead to additional cost shifts

---

mechanism would likely need to be considered for funding the revenue inadequacy resulting from reaggregation.

304 It appears, for example, that in CRR Study 1 the PG&E load zone was disaggregated into 26 zones averaging less than a fifth of the size of the SDGE aggregated load zone which was not disaggregated.

305 Sept ISO, p. 56.

and would not address some of the other problems created by aggregated LAP zones applied to vertically integrated LSEs.

First, if vertically integrated LSEs with generation sited at their load were allocated CRRs from the LSE's generation to the LAP zone in which the LSE's load were located these CRRs would likely be valuable even in hours in which the LSE would not need a hedge between local generation and the aggregated load zone price. In order to provide a full hedge for the LSE, CRRs from its generation to the aggregated load zone would need to be allocated for the level of its peak load. At levels of load less than peak load, however, the MW quantity of these CRRs would be unchanged and the LSE would be paid congestion rents for the full quantity of these CRRs even though the nominal congestion charges that it would incur in meeting its load with its generation at the same location would fall with the level of its load. If an LSE were allocated sufficient CRRs from its generation to the aggregated load zone to hedge the congestion charges from its generation to the LAP for its peak load, then it would likely receive a windfall during lower load conditions, when it would earn congestion rents that were not needed to offset congestion charges between its generation and the LAP. This outcome is illustrated in Appendix VI, Sections A-E.

One could attempt to avoid such a windfall by allocating LSEs fewer CRRs than required to fully hedge the congestion costs between their generation and their load at the same location at peak load, in the expectation that these congestion costs would be offset by excess CRR revenues during lower load conditions. A difficulty with this approach is that the LSE would then not be hedged against congestion between its generation and load at the same location. If the actual congestion levels or frequency of high load conditions requiring the dispatch of this generation were to differ from those expected in assigning CRRs, the LSE could be adversely impacted. It is important to keep in mind that the purpose of a hedge is to reduce the financial impact of congestion charges when congestion differs from that which was expected.

A second difficulty in applying such a system of aggregated load zones to vertically integrated LSEs is that some CRRs from local generation to the aggregated LAP zone could be counterflow CRRs within the settlement system, so they would not be voluntarily requested by the LSEs.[306] The failure to award these counterflow CRRs, however, could make many other CRRs from generation to the aggregated LAP zones infeasible. This is illustrated in Appendix VI, Section F. Based on the current MRTU description it does not appear that such counterflow CRRs would be involuntarily assigned to an LSE with generation and load at the same location during the CRR allocation process.

A third problem in applying such a system of aggregated load zones to vertically integrated LSEs is that there is a potential for CRRs from local generation to the LAP to be infeasible, even though there is actually no congestion from the local generation to the actual load. This can occur in the simultaneous feasibility test because CRRs from local generation to the LAP could overload constraints on deliveries to other regions in the LAP, even though there are actually no deliveries from the local physical generation to the physical load in these other regions. This potential is illustrated in Appendix VI, Section G.

[306] A counterflow CRR is a CRR for which the congestion rents have an expected value that is less than zero because the price at the sink is lower than the price at the source.

Overall, the allocation of CRRs to aggregated LAP zones with internal transmission constraints and vertically integrated LSEs has the potential to lead to unintended costs shifts and to limit the ability of market participants to effectively hedge congestion risk.

C. CRR Allocation

The Cal ISO filings provide information on the general objectives and procedures that are planned for allocating CRRs, observing that the details of the allocation will be developed consistently with the FERC White Paper. The FERC has determined that the allocation proposal is reasonable[307] but has requested further details.[308]

The Cal ISO has stated as a general principle that it intends "to allocate quantities of CRRs that are adequate to fully protect loads from congestion costs, provided these quantities are simultaneously feasible."[309] This statement of purpose for CRR definition and allocation could be important in explaining the CRR allocations and in guiding the CRR implementation process, but it may be problematic if "fully protect loads from congestion costs" is interpreted to mean assured access to low cost generation. In the past, it has been necessary to dispatch generation out of merit to meet load in California and these costs have been borne by California rate payers both under utility operation and under Cal ISO operation of the transmission system. It would be unrealistic to offer as a goal an objective of hedging all load to the lowest cost generation in the region, or even the lowest cost generation inside California.[310] A goal of hedging all load relative to the cost of generation at *some* location, however, is attainable. Thus, it is possible to hedge all loads for particular patterns of system utilization, i.e., to some generator, provided these quantities are simultaneously feasible.

The description of CRR Study 2 describes the CRR allocation goal in different terms: "adequate hedging of congestion costs over the course of the year, rather than trying to cover Load Serving Entity (LSE) schedules on a MW basis in each hour."[311] As explained in the February CRR study, this approach would attempt to allocate CRRs such that an LSE's expected congestion charges over the year would be equal to its expected CRR revenues.

While the decision rule described by the Cal ISO in discussing CRR Study 2 could be employed as a criterion for allocating CRR revenue, it should not be described as an approach for assigning hedges against congestion charges. The purpose of hedging is to provide market participants with a tool for managing the congestion payments and charges that actually occur, not just those that are expected ex ante. By design, CRRs provide a hedge under which congestion charges associated with meeting load with specific generation resources do not vary

---

307 Oct FERC ¶171.

308 Oct FERC ¶172 and June FERC ¶168.

309 CMD Transmittal, p. 74 and CRR Study 1, pp. 5, 14.

310 The Cal ISO's discussion of its decision to abandon its former goal of hedging all load net of local generation, suggests that the first interpretation of the Cal ISO goal is correct, but this goal is infeasible. See CMD Transmittal, p. 75, CMD #83.

311 Feb CRR Study 2, pp. 2, 3. Similarly, the Cal ISO stated that "using an instrument such as CRRs to provide a 'full' hedge against congestion costs requires thinking about congestion costs on an average basis for a given period of time, rather than on an hour-by-hour basis." Sept ISO, p. 85.

as a result of unanticipated changes in congestion patterns, such as those arising from variations in load patterns or changes in relative fuel prices. This is not the same as assigning CRRs such that the congestion charges will be offset by congestion revenues if congestion patterns are as expected. Many different sets of CRRs may have the same *ex ante* expected value as the congestion charges that an LSE expects to pay in meeting its load. However, these sets of CRRs can differ tremendously in the degree to which the payments that they provide to an LSE will vary in step with changes in the congestion charges that the LSE must pay. CRRs that more closely match the LSE's schedules will provide the better hedge against unexpected changes in congestion patterns, meaning that the CRR revenue will vary directly with changes in congestion charges. It is important to keep in mind that the purpose of holding CRRs is to be hedged against unexpected changes in congestion charges. If the purpose were merely to distribute revenues reflecting expected congestion, all CRRs could simply be auctioned and the auction revenue could be assigned to the LSEs consistently with their expected congestion charges. The Cal ISO recommendation in CRR Study 2 that LSEs request CRR sources reflecting the actual sources of supply they use to serve their loads is consistent with the allocation of CRRs that reduce risk by hedging congestion costs.[312] It is also anticipated that having the LSEs request CRRs in spatial patterns that match their actual use of the grid should allow more of these CRRs to be simultaneously feasible, compared to allowing all LSEs to request any CRR they want.

A workable goal for CRR allocation might be stated as to allocate quantities of CRRs that achieve an appropriate balance across parties and that adequately protect a specified pattern of load and generation from net congestion costs (congestion charges net of payments), provided these quantities are simultaneously feasible. This appears to be consistent with what the Cal ISO has done in its CRR studies, which is to allocate CRRs to the extent possible to LSEs, consistent with their actual sources of supply.

Under the MRTU, CRR obligations would be allocated to all native load within the Cal ISO control area that pays the embedded costs of the transmission grid. The allocation to loads would be based on the historic level of load, the geographic distribution of load, and the anticipated distribution of a load's supply resources. In general, the LSEs that serve such loads would be recipients of CRRs on behalf of the loads. This proposal for allocating CRRs to native load is, at a high level, reasonable and consistent with the approaches used in New York and PJM. Issues will continue to arise as detail is added to further define the meaning of terms and phrases such as "anticipated distribution of a load's supply resources." The definitions used in the allocation process would determine which parties receive allocations of relatively more valuable CRRs and which receive less. The equity tradeoffs involved in each decision, such as the implications of basing the allocation on the system peak load versus on each LSE's peak load, will need to be evaluated by market participants.

The MRTU has spelled out some of the details of the CRR allocation. First, it has been proposed that loads such as those of the State Water Project that are not formally served as retail customers by a LSE would receive CRRs. The Cal ISO has also stated that CRRs would not be allocated to parties historically engaging in short-term wheeling transactions that do not serve native loads internal to the Cal ISO control area (except for the case of ETCs, which are long-

[312] CRR Study 2, p. 8.

term contracts).[313] Parties that wish to hedge the congestion costs associated with short-term wheel-through schedules may acquire CRRs in the auction or secondary market.[314] The CRR allocation will account for loads served under ETCs.[315]

Second, under the MRTU, there would be no reduction in the quantity of CRRs allocated to an LSE due to the LSE's ownership of local generation.[316] That is, the allocation of CRRs to an LSE would not be net of local generation unless a lower level is requested by the LSE in question for "whatever" reason. This is significant and a consequence of the aggregated LAP pricing zones. It was observed in Section C that due to the aggregate load zones used for load pricing, an LSE could require CRRs from generation to load located at the same location in order to hedge itself against congestion charges arising from the LAP pricing system. It would therefore not be appropriate to reduce the CRRs allocated to an LSE based on its local generation under the LAP pricing system since even an LSE with generation at the same location as its load would need CRRs from its generation to its load in order to be hedged for changes in congestion charges. Conversely, LSEs with generation and load at the same location would not be required to accept counterflow CRRs from their generation to a lower priced LAP, but could instead designate CRRs from lower priced sources. As discussed in subsection VIII.C above, the ability of LSEs with generation and load at the same location to choose whether or not to accept CRRs is likely to result in lost shifts and magnify the impact of the aggregate load zone definition in limiting the ability of LSEs to effectively hedge congestion charges.

Third, the Cal ISO also expects CRRs to provide a financial hedge against the congestion costs associated with trading Ancillary Services (Operating Reserves) across interties. Day-ahead ancillary service schedules across inter-ties will incur congestion charges just like day-ahead energy schedules.[317] Parties will not be allocated CRRs to offset the congestion costs associated with ancillary services, but they may obtain these through the auction or secondary market. The Commission has found this proposal to be acceptable.[318]

Fourth, the allocation of CRRs will include both on-peak and off-peak CRRs and it is understood that LSEs could submit different annual requests for the on-peak and off-peak period, and each LSE could have different maximum MW annual allocations for on-peak and off-peak hours.[319] This appears to be a reasonable approach, since the load distribution will impact the relative allocation of CRRs among LSEs. The CRR allocation would be conducted monthly, as

313 Sept ISO, p. 98.

314 The eligibility of entities that serve load outside of Cal ISO but have contributed to the embedded cost of the Cal ISO control area to receive an allocation of CRRs is being considered. "The CAISO, in discussion with CRR Stakeholders, has decided to conduct a sensitivity study that will consider allocating CRRs to CMUD and other entities identified to serve load outside of ISO control area and have made a significant contribution to the embedded cost of the ISO control area." CRR Study 2, p. 8. We agree that this issue should be considered, as there is a substantial difference between the embedded cost contribution of a party taking annual wheeling service, versus a party taking daily service. Another fact to be considered is whether long-term firm wheeling service in some cases provides counterflow for other service that will receive an allocation of CRRs.

315 CMD Transmittal, p. 74.

316 Sept ISO, p. 94; CMD Transmittal, p. 75.

317 CMD #94.

318 Oct FERC ¶188.

319 CMD Transmittal, p. 78.

well as annually, which would permit LSEs to hedge short-term load growth by adjusting allocation quantities in the monthly CRR allocation process as needed.[320] It is understood that LSEs would be permitted to adjust their monthly on-peak and off-peak CRR requests prior to the monthly allocation, up to their monthly maximum allocations, rather than having their monthly requests determined from annual requests.[321]

The Cal ISO likely will need to address concerns that arise if different LSE's peak electricity usage occurs during different months or seasons of the year, because the simultaneous feasibility test may not be satisfied if different LSEs base their nominations on load profiles for different time periods. Will the Cal ISO determine each LSE's maximum annual requests on a coincident or a non-coincident basis (it is assumed that this will be done separately for the on-peak and off-peak periods)? In the Midwest, the potential tradeoff between equity and feasibility led the parties involved to consider performing the annual allocation based on monthly load distributions and monthly CRR requests. It is not clear what is intended in this regard in California. Annual CRR requests in CRR Study 2 consist of a set of 12 monthly requests, but it is also stated that annual awards will be of a constant MW amount and it is understood that the current intent is for fixed MW allocations over the year for the on-peak and off-peak periods in the annual allocation process. While there are advantages to performing the annual allocation at a monthly level of analysis, it may be burdensome for the Cal ISO and ultimately expensive for market participants compared to a system in which annual CRR requests are evaluated in a single annual model run, or one in which there are separate requests and model runs for just summer and winter periods. LSEs would receive an additional monthly allocation of CRRs on top of their annual allocation, so it may not be essential that the annual allocation be performed at a monthly unit of analysis.

Fifth, entities eligible for a CRR allocation will submit their nominations by specifying source, sink, MW quantity and time of use. The MW quantities will be capped by an upper bound determined by the 0.5 percent exceedence level of the monthly load duration curve of the eligible entity.[322] CRRs will be allocated to LSEs based on the priority level of their rights. The three broad priority levels are: (1) Unconverted ETC Rights; (2) Converted Rights (ETCs that convert to CRRs and new PTOs); and (3) LSEs (including metered subsystems and municipal utilities). In addition, a four-level priority approach will be applied to LSEs. Thus, the upper bound of each LSE's nomination quantity (in megawatts) will be divided by four. Along with each CRR request, the LSE will provide a tag with a sub-priority from 1 to 4, with 1 being the highest sub-priority. The total nominations for each sub-priority may not exceed the sub-priority megawatt upper bound.

Finally, under the MRTU the differentiating factor in allocating CRRs of the same priority type would be each CRR's effectiveness in alleviating transmission constraints. Overall, the objective of the CRR allocation is stated to be to maximize the quantity of allocated CRRs in

---

[320] For the CRR Study 2 study, one month in duration CRRs could be requested for each of 12 months of the study period. See also Sept ISO, p. 96.

[321] Possible difficulties with allowing this flexibility but maintaining the priority system are discussed in a footnote to Section VIII.E.

[322] CRR Study 2, p. 21.

terms of MW, taking into account the priorities associated with different CRR types.[323] The meaning of this objective function is not entirely clear. What choice would the software make between allocating 100 CRRs with priority 2 versus 205 CRRs with priority 4? Which would be chosen? One interpretation of the statement of the simultaneous feasibility test objective is that the software would, first, allocate as many CRRs as possible with priority 1, second, as many CRRs as possible with priority 2, and so on. A second interpretation is that the allocation software would maximize the sum of (MW awarded)/(priority level), and other interpretations may be possible.[324] However, this second approach does not strictly maintain the priority system; i.e., it does not ensure that all possible priority 1 CRRs are allocated prior to allocating any level 2 CRRs.

The following paragraphs describe and comment on some of the details of the steps that the Cal ISO proposes to use to allocate CRRs.

Step 1. Account for the impact of non-converted ETCs on transmission capacity.

The MRTU would include non-converted ETCs in the simultaneous feasibility test by designating CRRs that have ETC load locations as their sinks, rather than LAP zones. These CRRs would be represented in the simultaneous feasibility test used to allocate CRRs, but would not be allocated to the ETC customer. The CMD discusses the possibility of designating these CRRs as obligations, but the CRR Study 2 scenario specification indicates the intention to explore modeling ETCs as options rather than obligations in the simultaneous feasibility test. The issues to consider in deciding whether to model these ETCs as options or obligations are: first, if power only flows under the ETC when the ETC is in the direction of congestion, but the direction of congestion varies within the allocation period, then the simultaneous feasibility test needs to model the ETCs as options in reserving capacity to support the ETCs; and, second, if power always flows under the ETC regardless of the direction of congestion, then it would be appropriate to model the ETCs as obligations if the ETC holder's schedules are assumed to use

---

[323] CRR Study 2, p. 12.

[324] There are different software requirements for the two possible interpretations of the allocation objective given above, which each have advantages and disadvantages. The first interpretation would implement the simultaneous feasibility test for the CRR allocation so as to allow LSEs to modify their requests for CRRs of a lower quartile priority based on the outcome of the allocation of CRRs for a higher quartile priority. Thus, if LSE A received relatively few of the CRRs that it most needs in the first quartile allocation, it could choose to request these same CRRs in the second quartile allocation step. This type of process may help LSEs to obtain the hedged that they most need or that reflect a certain distribution of generation sources. On the other hand, an allocation process that runs a simultaneous feasibility test separately for each priority level may be administratively burdensome (especially if new or revised requests are accepted between each simultaneous feasibility test run and the process has to be run every month). This approach would necessarily maximize the allocation of CRRs of a given priority level, even if this causes a much larger megawatt number of CRRs of a lower priority level to be infeasible. This outcome would not (necessarily) occur if the allocation software were implemented in a single step that would, for example, maximize the sum of (MW awarded)/(priority level). Such a one-step allocation process would not allow an LSE to make its requests for lower quartile CRRs contingent on what it was awarded for higher quartiles Implementation of contingent CRR bids by introducing additional linear constraints on the simultaneous feasibility test was discussed in the Midwest. This is theoretically possible, but could be a potentially large complication and would need to be discussed with software developers.

its ETC rights to obtain this transmission service.[325] Alternatively, if an ETC holder would be permitted to schedule transactions using its ETC rights in the direction of congestion and to schedule counterflow transactions in the market without buying additional transmission service, then the ETC rights would need to be modeled as options.

The CAISO proposes to assess the impact of non-converted ETCs on the overall allocation of CRRs by requiring that ETC holders provide a description of their normal use of the grid under their ETC rights, with specific quantities of generation and load at each location. In CRR Study 2, ETC schedules were requested from the scheduling coordinator for each ETC schedule. Although the documents state that in the event that all ETC reservation patterns are not simultaneously feasible, the algorithm would make curtailments based on global ETC priorities, as agreed upon by the relevant PTOs, to achieve simultaneous feasibility of the ETCs.[326] The Cal ISO has clarified that the current plan is to evaluate the feasibility of ETCs, Converted Rights and LSE CRR requests simultaneously.[327]

Step 2. Allocate CRR obligations or options to ETCs that convert to CRRs and to new PTOs.

As in the previous step, entities converting ETCs to CRRs and new PTOs would be asked to provide the ISO with their normal grid use patterns. While the ISO has stated a preference for providing CRR obligations to these entities, in CRR Study 2, it is assumed that new PTOs will receive options. It is important to recognize that, given a choice, parties receiving an allocation of CRRs always will choose 100 MW of options from A to B rather than 100 MW of obligations between the same two points; even the transmission service for a baseload unit that always operates. The ISO will need to consider the extent to which assigning options to these parties would entail payment for counterflow that would not receive payments if scheduled under the ETC, and thus the conversion to options would cause other ETC rights and non-ETC CRR requests to become infeasible.

The Cal ISO continues to work out details concerning the modeling of new PTO rights. CRR Study 2 states: "the transmission that the new PTOs have rights on and that lies outside of the control area will be modeled in the Full Network Model ("FNM"). The Cal ISO will continue to work with the five new PTOs to determine the correct modeling of Point-to-Point CRRs on this transmission."[328]

Some issues remain to be worked out for Step 2. It is understood that when existing transmission contracts allow a choice among several source locations, the ETC or new PTO will be allowed to choose the source up to the limits of its rights, and that ETCs that are wheeling contracts will be included in the allocation. What validation procedures will be used to check the

---

[325] Thus, since the ETC would be used for transmission service, the ETC holder would not be paid for the counterflow.

[326] CMD #93.

[327] It likely would not be useful to run a simultaneous feasibility test for the ETC rights on a stand-alone basis, since the feasibility of the ETC rights can only be determined after taking into account any counterflow provided by requests for CRR obligations. For instance, if counterflow from generation serving other load was assumed in the original evaluation of transmission service for an ETC customer, the ETC rights might not be feasible in a simultaneous feasibility test that does not take account of this counterflow.

[328] CRR Study 2, p. 8.

"normal grid use patterns" submitted by the converting ETCs and new PTOs when they request CRRs? It is understood that the general intent is to use historical data, supplemented by information on new energy contracts or plants, but that the actual procedures are still being determined.

Step 3. Allocate CRRs to non-ETC loads.

Under the MRTU, LSEs and other eligible loads would be asked to provide the Cal ISO with data showing the grid usage patterns they normally rely upon to meet their needs.[329] Municipal utilities, metered subsystems and direct access providers would be included in this step of the allocation process.

For CRR Study 2, the Cal ISO guidelines state that "a consistent pattern should exist between the CRR source-sink request and the actual or historic supply sources that the requestor uses to serve load."[330] Pumped load was asked to use an average water year. In this regard, note that pumped storage may need two different sets of CRRs: one for when it is pumping, and another for when it is generating; these may arise naturally if LSEs are allowed to request different CRRs for the on-peak and off-peak periods.

For this third step of the allocation, there is an issue of whether loads with bilateral contracts must provide a specific source location, or may specify a trading hub source; to address this issue, the Cal ISO planned scenarios to explore the alternative approaches in CRR Study 2. This seems to be a reasonable approach, but the Cal ISO should be wary about basing its evaluation of impacts on non-binding requests that parties make for the purpose of the CRR study. It should be anticipated that if there is a choice the non-ETC load will choose the alternative that consumes the most capacity over congested interfaces and/or supplies the least amount of valuable counterflow.

A few additional questions and issues arise in considering Step 3 of the CRR allocation process.

First, it may take some time to reach agreement on the definition of the sources that will be allowed for the non-ETC CRR requests. The following types of questions will arise, if they have not already. Must the generator sources be owned or under a long-term contract to the LSE? What CRR sources will LSEs be allowed to nominate for load that they have typically served with short-term transmission contracts, such as imports?[331] What if the LSE has a supply contract that will expire shortly – what generation source will it be permitted to nominate instead? The answers to these questions will regulate the degree to which non-ETC loads will be able to shape their CRR requests in order to obtain relatively more valuable CRRs, and also affect the degree to which the initial requests are infeasible as a whole.

---

[329] CMD #93.

[330] CRR Study 2, p. 10.

[331] It is understood that the current intent is to allocate CRR import capacity on each intertie based on each LSE's historical use of that tie to serve load.

Second, there is an important question of whether non-ETC loads will be allowed to nominate a quantity of CRRs that is less than their load. It appears that this would be allowed, since an LSE could request a lower allocation of CRRs for "whatever" reason.[332] A related question is whether LSEs will be allowed to exclude certain generators (such as peaking units) as sources in the set of CRRs that they nominate. Transmission from particular generators to non-ETC load may provide important counterflow in the simultaneous feasibility test and may have been assumed to be present when new PTOs and/or ETC customers were awarded their transmission service in the past. CRR Study 2 will examine this question. Moreover, as discussed in Section B, this question could be particularly important for LSEs meeting load with generation at the same location. There is a likelihood under the proposed system of zonal load aggregation that the number of CRRs that can be awarded to the load zone will be sensitive to the designation of counterflow CRRs from generation in constrained areas to the load zone, even if that generation does not produce any physical counterflow but only serves to meet load at the same location as the generation.

If LSEs using generation to meet load at the same location are not required to take CRRs from their generation to the load zone, the CRRs requested by others (converting ETCs, new PTOs or even other non-ETCs) may be infeasible. On the other hand, assigning CRR obligations to LSEs based on the use of local generation to meet peak load could impose substantial cost shifts during non-peak conditions.

The market design does not appear to require LSEs to nominate a certain level of CRR obligations, or CRR obligations reflecting the use (in some reasonable proportions) of all generators used to serve their peak load; this flexibility is consistent with the approach in PJM. It could, however, lead to the need to severely curtail non-ETC CRR requests, even if these are made for CRR obligations.

The CMD states: "In the event that not everything is simultaneously feasible [following the requests for CRRs by non-ETC loads] the ISO would curtail non-ETC load or LSE CRR requests first, and preserve converted ETC CRR obligations as far as possible, to provide converted ETCs a higher degree of certainty of receiving their desired CRRs as a benefit for converting. CRR obligations allocated to non-converted ETCs would maintain the highest degree of protection in this process."[333] The amount of curtailment of non-ETC load will depend on whether the CRR requests of ETCs and new PTOs are required to be obligations (as assumed by the previous quote), in which case they will provide counterflow, or are permitted to be options (which appears to be the assumption elsewhere).

Overall, it appears that there are a variety of reasons under the MRTU that a non-ETC LSE will not be required to nominate CRRs from all of the generating units that the LSE has historically used to serve its load. CRR Study 2 will explore the implications of modeling CRR allocations to non-ETC loads as options versus obligations, which will provide some information about the importance of counterflow from non-ETC CRRs to the feasibility of other parties' CRR requests. If the results of this analysis are to be reliable, it will be necessary to examine the

---

[332] CMD #83.

[333] CMD #93.

value of CRRs from generation to load at the same location when defined to the LAP and assume that no counterflow CRRs (i.e., CRRs requiring payments) will be designated.

Another potential problem area in the allocation process is the proposal that the allocation of CRRs would be prorated based on relative constraint impact. Such an allocation rule could have surprising unintended consequences, particularly when combined with CRRs sinking at LAPs. This rule could cause the entire proration of CRRs to fall on a single LSE whose requested CRRs have the largest constraint impact, even if the generation source for the requested CRRs has historically been used to meet load and the need for proration arises from the CRR requests of other LSEs. Even more troubling, it is possible for CRRs from generation to the LAP to be curtailed to zero under this rule even if the generation and load are actually at the same location and the physical injections have no impact on the constraint that causes proration to be applied. Moreover, if LSEs have complete freedom in designating CRR sources in the allocation process, this criteria could lead to designation of CRRs that have little to do with hedging LSE load from generation to the LAP.

Step 4. Perform an auction for the rest of the CRRs, as discussed in the following section, D.

At each of the four steps of the allocation process described above, the proposal calls for a simultaneous feasibility test to be run, after fixing the allocation of rights and CRRs that resulted from previous steps. (See Section E.) It appears to us, however, that if some of the CRRs in Step 3 are obligations, a meaningful simultaneous feasibility test can only be run after Step 3.[334] The CRRs modeled in Step 1 as a proxy for ETCs may not be feasible on a stand-alone basis if they rely on counterflow from other CRR requests. Similarly, the Step 1 and Step 2 CRRs, encompassing converting and non-converting ETCs and new PTOs also may not be feasible without the counterflow from non-ETC CRR requests in Step 3. Thus, the purpose of running a simultaneous feasibility test after each step is not clear, since all of the CRRs and all of the counterflow need to be in the model before the feasibility of any ETC right or CRR request can be determined. It does not appear to be workable, for instance, to run a stand-alone simultaneous feasibility test for the ETC rights in Step 1, since any curtailment that results from this simultaneous feasibility test might need to be reversed after taking into account the counterflow from the CRR requests in Step 3.[335]

A final aspect of CRR allocation under the MRTU is that CRRs would be reallocated to follow the load if load switches to another LSE.[336] The proposal that CRRs be reallocated from LSE to LSE with changes in load has the potential to require the Cal ISO to develop and administer complex rules to govern this load following process. While these costs may not be material if there are no such shifts in load between LSEs for the foreseeable future, it is not clear why such shifts might not occur. Moreover, application of this approach would in the long-run hinder, if not foreclose, development of a market for long-term CRRs, adversely affecting forward contracting and generation development.

---

[334] A simultaneous feasibility test and pro-rating could occur after each step if all rights and CRR requests took the form of CRR options, since in this case there would be no interactions between steps caused by CRR counterflow.

[335] It is understood that the current intent is to run a single, simultaneous feasibility test at the end of Step 3.

[336] CMD # 81.

While the reallocation of CRRs among LSEs so as to follow loads as they shift between LSEs under retail access programs is consistent with the objective of assigning the economic value of CRRs to the loads that pay the embedded costs of the transmission system, the complexity of administering such a system in a retail choice environment should not be underestimated. Rules would need to be established to define the obligations of the LSE losing loads and the Cal ISO settlement process would need to support these rules. As retail load moves from LSE to LSE from month to month the CRRs to be reassigned would likely be measured in fractional MW. In addition, each LSE serving load within a given load aggregation region may have CRRs from different sources. Thus, as each LSE gains and loses retail customers it could be gaining and losing distinct CRRs in various proportions. For these reasons, this CRR reassignment process could become unwieldy and expensive from an administrative standpoint for all concerned. Moreover, CRRs allocated on an annual or multi-year basis may be sold by the LSE prior to the time that load switches, requiring the assignment of negative CRRs to the losing LSE to offset the reallocated CRRs. Some of the potential administrative difficulties of a system in which CRRs follow load are illustrated with a simple example in Appendix VII.

The long-run problem that will be created by such a process of reassigning CRRs to follow load is that no entity would have a long-term entitlement to CRRs. Lacking such a long-term entitlement to CRRs, LSEs would be unable to sell CRRs in the auction to support the sale of long-term CRRs to other LSEs wishing to enter into long-term power purchase contracts. This potential inability to acquire long-term congestion hedges could interfere with implementation of resource adequacy programs and would be a step backward from the long-term contract path rights available elsewhere in the WECC. Most of these problems would be avoided if the CRRs were treated as assigned auction revenue rights with the revenues used to reduce PTO transmission access charges, and all the actual CRR hedges for LSEs would be obtained from the auction.

D. CRR Auction and Secondary Markets

Under the MRTU, the transmission capacity that is left after CRRs are allocated to LSEs will be made available to support the sale of additional CRRs in an ISO coordinated auction.[337] Qualified bidders, including CRR holders, may participate in the auction as buyers and sellers. The revenue from the sales of CRRs by a CRR holder will be paid to the selling party.[338] The net auction revenue (after paying selling CRR holders) will be allocated to the Participating Transmission Owners (PTOs), in proportion to their transmission revenue requirement, to be applied to reducing their transmission revenue requirement.[339]

Some features of the CRR auction are not yet specified:

337 CMD # 92. Thus, if CRRs within a particular LAP were not fully allocated because of intra-LAP constraints, auction participants would buy node-to-node CRRs across the undersold constraints and the resulting auction revenues would be shared by the customers of all PTOs.

338 Oct FERC ¶165.

339 CMD # 92.

- Will on-peak and off-peak CRRs be auctioned simultaneously for a given period of time (month or year), or will two auctions be held sequentially?[340]

- Will the bidding for annual CRRs be for a set of 12 monthly bids, or a single annual bid? Can parties buy monthly CRRs in the annual auction?[341]

- Can auction bids be for point-to-point, point-to-load zone and point-to-hub CRRs?

How these details are resolved will affect a variety of other elements of the MRTU market design.

One critical ambiguity is what is intended by the description that the "ISO will run a CRR auction to allocate any transmission capacity that remains after loads and converted ETC holders received their shares." At the beginning of the allocation period, LSEs will have had the opportunity to designate CRRs using 100 percent of system transfer capability for the next month, 75 percent for the rest of the allocation year and 37.5 percent of capability for the following allocation year. We anticipate that it is intended that only the portion of system transfer capability that has been made available for designation by LSEs would be made available for sale in the auction, unless it is intended that the auction would be limited to monthly CRRs.[342] If literally all remaining transfer capability were made available in the auction, few valuable CRRs would remain to be allocated to LSEs in future periods.

The Cal ISO's July 22, 2003 filing states that the auction will be run to "maximize the auction proceeds."[343] This is not the appropriate objective function for the auction from an economic perspective, and is not the objective function used in the auction of financial rights by PJM and New York. The appropriate objective function is to maximize the value of all accepted bids to buy CRRs, less the value of all accepted offers to sell CRRs, subject to the requirement for simultaneous feasibility of the awarded CRRs. The valuation of accepted bids would be made at the offer price made to buy or sell the CRR. In economic terms, this objective function equates to maximizing the "social welfare" or "gains to trade" of the auction, which is the sum of the value implicitly placed on auction purchases by the buyers (as communicated by their bids), less the sum of the reservation prices implicitly placed by sellers (as communicated by their offers to sell). A basic principal of economics is that markets should be structured to the extent possible to achieve a result that maximizes social welfare. Competitive markets achieve this

---

340 Based on CRR Study 2, it appears that the auctions will be sequential. Cal ISO MT, p. 5.

341 In CRR Study 2, there will be 24 allocation periods (12 months * two time-of-use periods). Here, it appears that there must be separate monthly bids. (Cal ISO MT, p. 5)

342 It would be possible to run long-term auctions that reserve capacity to support future CRR allocations to LSEs by scaling up the allocations for the partially allocated periods and extending these allocations forward into future years. Such a procedure would ensure that the auction would not interfere with the ability of LSEs to designate their current CRRs in future years. These methods would not ensure, however, that LSEs would be able to designate different CRRs in future years.

There is a fundamental tension between accommodating long-term CRR auctions to support long-term congestion hedges for LSEs and allocating CRRs to LSEs on a short-term basis. This conflict can easily be fully resolved by allocating LSEs auction revenue rights rather than CRRs.

343 CMD #93.

outcome, as opposed to monopolistic markets. The objective function of "maximizing auction proceeds" would require the Cal ISO to economically withhold transfer capability from the auction if this would sufficiently raise auction CRR prices to increase net auction revenues. The Cal ISO should be acting as an auctioneer in running the CRR auction, using an objective function that maximizes social welfare and mimics a competitive market outcome, rather than an objective that entails attempting to exercise market power. Both the PJM and New York FTR/TCC auctions use the objective function of maximizing the value of all accepted bids to buy CRRs, less the value of all accepted bids to sell CRRs.

In addition to the allocation process and auction run by the Cal ISO, market participants may trade CRRs through a secondary market. In secondary market transactions, CRRs may be unbundled into any specific hours of the day, days of the week, or seasons, and NS-CRRs may be unbundled into their separate injection nodes consistent with the distribution factors defining the NS-CRR. Secondary market trades must be registered with the Cal ISO's Secondary Registration System. The Cal ISO does not itself intend to facilitate secondary market CRR trades other than through its auction. This proposal is reasonable and consistent with the practices in other regions.

The MRTU proposes that no limits will be imposed on CRR holdings. This policy would be reconsidered if there were evidence of gaming or the exercise of market power based upon possession of excessive amounts of CRRs.[344] In addition, the Cal ISO should consider imposing creditworthiness standards on those purchasing CRRs in the auction and transacting in CRRs. These standards may be important since CRR auction markets allow a market participant to be paid today in return for holding a counterflow CRR that will likely require future payments. ISO revenue adequacy requires that such purchasers make good on subsequent payments of negative congestion rents billed to this CRR in future day-ahead markets. If the party holding such a counterflow CRR were to liquidate or go bankrupt after the auction, there could be an impact on Cal ISO revenue inadequacy. Both PJM and New York have credit standards for financial rights holders.

E. Simultaneous Feasibility Test

The Cal ISO will perform a simultaneous feasibility test to determine the CRRs to be awarded through each step of its allocation and auction process.[345] The Cal ISO will run one optimization/simultaneous feasibility test process for allocating annual-term CRRs (one test each for on-peak and off-peak) and one for allocating monthly-term CRRs (one test each for on-peak and off-peak).[346] A simultaneous feasibility test also would be run whenever new transmission capacity is added or removed, so as to determine the incremental CRRs to be awarded.[347]

---

[344] CMD Transmittal, p. 78.
[345] CMD # 77.
[346] CRR Study 2, p. 10.
[347] CMD # 96.

For the annual simultaneous feasibility test, the Cal ISO will assume that all lines are in-service, except if major long-term outages are scheduled.[348] Planned outages would be included in the simultaneous feasibility test for the monthly CRR allocation and auction models but the criteria for inclusion have not yet been developed.

There should be a close correspondence between the transmission model, contingency set and constraint representation used in the simultaneous feasibility test and those used in the day-ahead market. Modeling differences, such as assumptions concerning PAR settings or unscheduled flow over the Cal ISO control area, should be identified and investigated to determine their potential impact on revenue adequacy. Some aspects of the grid configuration, such as lines in service, will vary over the month, so reasonable assumptions must be made in the simultaneous feasibility test, as recognized by the Cal ISO.

For purposes of CRR Study 2, the Cal ISO proposes to remove the transfer capability of transmission that is within the Cal ISO control area, but not under the control of the Cal ISO (called Transmission Ownership Rights) from the network model used for the simultaneous feasibility test by defining point-to-point CRR options at the source and sink of the line. This would be a reasonable approach provided that the options correspond to the schedules that can be expected over the non-ISO transmission lines and that the Cal ISO has no responsibility for managing congestion on these lines

The Cal ISO has stated that the transmission model used to perform the simultaneous feasibility test for the allocation in CRR Study 2 was created by scaling down the network capacity to 75 percent of that defined by the full network model (after removing capacity associated with Transmission Ownership Rights**)** and the operating constraints being used.[349] This model was used for determining the allocation of the long-term CRRs in the study. For each month studied, "an upper bound based on historical load data" was calculated.[350] Network capacity was not scaled down for the short-term monthly CRR allocation in the study; instead "for each month of the short-term allocation a CRR upper bound based on forecasted load data" was calculated.[351]

There are a number of ambiguities concerning the scaling down methodology that have implications for other aspects of the allocation and auction process, including its implementability. Some of the more basic questions are:.

---

348 CRR Study 2, p. 6.

349 CRR Study 2, p. 5. It is not clear what is intended from an implementation standpoint. It should be kept in mind that while pre-contingency limits can be adjusted by deducting ETC flows from the limit and multiplying the reduced limit by 75 or 37.5 percent, this methodology is not workable for post-contingency constraints as the ETC flows on a monitored element will be different for different contingencies, so no single limit adjustment would be appropriate for all contingencies.

The NYISO has developed a methodology for allocating or auctioning only a portion of transfer capability by scaling up bids and offers, rather than scaling down limits.

350 CRR Study 2, p. 5.

351 CRR Study 2, p. 5.

- Will the annual allocation process include 75% or 100% of the CRRs used to account for non-converted ETC schedules? Will it include 75% or 100% of the CRRs requested by new PTOs and by entities with ETCs that convert to CRRs?

- If the annual allocation includes only 75% of an entity's annual CRR request, will the entity be able to choose which particular CRRs (source to sink) are included in the 75% set or, alternatively, will a flat 75% of each annual CRR request be included, with the remaining 25% carried into the monthly auction? If so, does this mean that parties will be required to specify their monthly CRR requests at the time of the annual allocation?[352] Among other things, the answers to these questions affect the relative priority of the allocations made to different groups of customers, as customers with higher priority will have an advantage if they can pick *which* of their point-to-point CRR requests to include in the 75% annual allocations.

- How much transmission capacity will be available to support the annual auction of CRRs? The transmission system capacity used for the annual allocation and auction will be scaled down to 75%, while the CRR requests made for the annual allocation will include, at a minimum, 75% of the CRRs used to account for non-converted ETCs, 75% of the CRRs requested by new PTOs and by entities with ETCs that convert to CRRs, and 75% of the requests of non-ETCs. Moreover, it appears to us that in order to insure that sufficient transmission capacity is reserved for non-converted ETCs, 100% of these will need to be reserved in the annual allocation, rather than 75%.[353]

F. CRR Settlements

The congestion rents collected for each hour in the day-ahead market will be used to fund payments to CRR holders for that hour. Any hourly surpluses or deficits will be accumulated in a single balancing account. Funds from the balancing account would be disbursed at the end of each month to CRR holders that were not compensated fully during the month.[354] In addition, there will be a yearly clearing of the balancing account, in which any surplus funds in the account at the end of the year will be allocated to CRR holders in proportion to their gross

---

[352] It appears to us that under a scaling down methodology in which 75% of the CRR requests are included in an annual allocation for 75% of system capacity and 25% of the allocation occurs monthly, the implementation must be undertaken by representing 75% of each annual CRR request in the annual simultaneous feasibility test and 25% of each annual CRR request in the monthly simultaneous feasibility test. If this restriction were not imposed (i.e., monthly requests could be modified, or could be a non-proportionate subset of the annual request), then the ISO could not be sure that it was appropriately reserving transmission capacity for the monthly allocations of new PTOs and converting and non-converting ETCs. Thus, it might inadvertently allocate such transmission capacity to non-ETCs as annual CRRs, or sell it in the annual auction, not knowing it would later need this capacity to honor the monthly requests from higher-priority groups.

[353] Similarly, 100 percent of new PTO allocation requests and converting ETC requests may need to be considered at the time of the annual allocation in order to preserve the priority of these customer classes over the non-ETC customers. Otherwise annual CRRs could be allocated to non-ETCs that cause the monthly allocations to new PTOs and non-converting ETCs to be infeasible.

[354] CMD #90.

unrecovered annual shortfall. Any remaining surplus at the end of the year after fully funding all CRRs will be paid to PTOs. If the balancing account is short at the end of the year, no additional payments or charges will be made.[355] It is inferred that if the congestion rent collections are insufficient to fully fund payments to CRR holders in the hour, these payments will be pro-rated down proportionately in the hourly settlements but this does not appear to be explicitly spelled out and ought to be clarified.[356] This is a reasonable proposal which is consistent with the CRR settlement mechanism in PJM.[357]

G. CRRs for Third-Party Transmission Expansions

The MRTU provides that in the case of market-based transmission upgrades, the parties bearing the cost of the upgrade will receive CRRs reflecting the added transfer capability if they are not recovering their investment through an access-charge-based rate of return, a transmission credit, or direct payment from an existing PTO.

The Cal ISO's draft proposal for the allocation of CRRs to merchant transmission describes three different types of transmission upgrades, and their eligibility to receive CRRs:

- ***Upgrades Associated with Large Generation Interconnections (greater than 20 MW)***

The facilities added between the point of interconnection and the new generating facility are called interconnection facilities. Any modifications or additions made to the ISO-controlled grid beyond the point of interconnection, so as to accommodate the new generator, are called network upgrades. Network upgrades will be eligible for CRR allocations, but not interconnection facilities.[358]

There are two types of upgrades that are not explicitly associated with the interconnection of a large generator.

- ***Reliability-Driven Upgrades Not Associated with Large Generation Interconnection***

Reliability driven upgrades are upgrades required to ensure system reliability according to the Applicable Reliability Criteria. It is assumed that reliability driven upgrades will be made by PTOs, who will recover the cost of the upgrades through their regulated rate of return. To the

---

355 CMD Transmittal, p. 72. It does not appear to us that the balancing account would ever be "short"; the minimum balance would be zero, since it is used only to accumulate excess congestion revenue.

356 Since the loss residual will also be credited to the CRR balancing account (May ISO, pp. 70-76), there is a low probability of an overall annual shortfall but such a shortfall could arise during a particular hour or month.

357 In the future the Cal ISO may consider a more direct allocation of hourly congestion rent shortfalls and surpluses to the parties responsible for the changes in transfer capability causing such shortfalls or surpluses. This may become important in providing efficient maintenance incentives if some of these events have a relatively large dollar impact on the CRR settlements. In the CMD transmittal letter (pp. 72-73), the Cal ISO discusses why there will not be separate balancing accounts for each TO. They recognize that separate accounts might be used to provide an incentive for PTOs to manage transmission maintenance more effectively, as each PTO could be made to make up its shortfall in funds at the end of each year.

358 Cal ISO MT, pp. 1-2.

extent that the costs of these upgrades are recovered in regulated rates, the upgrade will not be eligible for an allocation of CRRs.

The likely intent of this rule is to avoid assigning CRRs to entities that do not bear the costs of the transmission investment. This rule also appears, however, to imply that the PTO ratepayers who do bear the costs of the transmission investment would only receive a portion of the benefits, through their entitlement to a share of the annual congestion rent residual. This rule would be appropriate if the costs of the transmission investment were allocated to all ratepayers on the same basis as their entitlement to the annual residual, but it is not clear that this will be the case. Shouldn't the economic value of the CRRs produced by a transmission investment by a PTO be assigned to that PTO's ratepayers?

- ***Economically Driven Upgrades Not Associated with Large Generation Interconnection***

Economically driven upgrades not associated with large generation interconnection are transmission assets that are put under the control of the Cal ISO, but are not needed to ensure system reliability. If the sponsors for these upgrades do not recover their investment cost under a FERC-approved rate of return or direct payment from a PTO, they are eligible to receive CRRs for the increase in transfer capability provided by the upgrade.[359]

The general principle under the MRTU is that the project sponsor of such economically drive upgrades will receive CRRs commensurate with the amount of transfer capability added to the system.[360] The quantity of CRRs allocated to the party responsible for the increased transmission capacity will be capped by the increase in transfer capability.[361] Any CRRs awarded for a transmission expansion will be required to be simultaneously feasible in combination with all previously awarded CRRs, which will maintain the Cal ISO's revenue adequacy.[362]

The Cal ISO lists a number of principles that it proposes to apply to the allocation of CRRs to eligible owners of new merchant transmission ("MT owner"):

- First, "CRRs will be allocated to the MT owner only after the MT upgrades have been energized and in operational control of the CAISO."[363] This is an appropriate approach because the final impact of the new upgrade on transmission capacity may differ from prior estimates made at the time that the upgrade was proposed or constructed. Such differences may occur because of changes in the specification of the new facility, or because the order in which upgrades are completed differs from the order in which they were proposed, or differs from the order in which they crossed any of a number of regulatory, financial and contractual hurdles. The CRRs that are incrementally feasible for a given upgrade will depend on the final specifications of the upgrade and the actual order in

359 Cal ISO MT, pp. 1-2.
360 Sept ISO, p. 103.
361 Cal ISO MT, p. 1.
362 CMD Transmittal, p. 79.
363 CMD Transmittal, p. 80, Cal ISO MT, p. 2.

which upgrades are completed. The CRRs that are estimated to be feasible based on estimated completion dates on related projects may be infeasible based on the actual completion order, leading to revenue inadequacy if CRRs were awarded based on these estimates.

- Second, "once the CAISO has included the MT related transmission upgrades in the FNM, these upgrades need to be consistently modeled in the FNM in all subsequent CRR Allocations/Auctions and other CRR related processes."[364] This is the correct approach for maintaining revenue adequacy and equity in the CRR allocations.[365]

- Third, "the terms of the CRRs that are allocated to the MT owner should be good for the life of the transmission facility."[366] This feature of the market design is reasonable and consistent with the approaches in other regions. Parties that make such long-term investments need to receive long-term property rights. It can be challenging, however, to design a process for determining incremental long-term CRRs awards for new merchant transmission owners that appropriately takes into account the rights of those paying the embedded cost of the existing transmission system. In order to determine long-term *incremental* CRRs for the new merchant transmission owner, it is necessary to define the long-term claims to CRRs on the existing system, since the award of incremental CRRs for the expansion will likely cause parallel flows on the existing system as well. A difficulty is that the long-term claims of existing ratepayers are not well defined under the MRTU. PJM addresses this issue in allocating incremental FTRs to merchant transmission by assuming that currently outstanding CRRs are a proxy for the long-term CRRs desired by existing transmission rate payers. Thus, the allocation of long-term CRRs is required to be simultaneously feasible on an incremental basis, taking into account currently outstanding CRRs. This is a reasonable approach, but it works better the longer the term of the entitlements to use of the existing system that are modeled in this process. An additional test might be run to insure that the incremental CRRs are feasible in the absence of currently outstanding CRRs, i.e., that they do not rely on counterflow by CRRs that might not be nominated in the future.[367]

---

364 Cal ISO MT, p. 2.

365 The CMD states that when "new transmission capacity is added or removed, the CAISO will review the impact of the change on the system network to determine the appropriate amount of new capacity to be released in subsequent CRR allocations and auctions." (CMD #78) The meaning of this statement is not clear. When new transmission capacity is added or removed, changes need to be made to the transmission grid model used in the CRR auction and allocation processes that are identical to the changes that will be made in the transmission grid model used in the IFM. This may include adding or changing monitored contingencies and constraints. When new transmission capacity is added or removed it is not necessary to determine a specific increase or decrease in the capacity that will be offered as CRRs; this follows implicitly from the changes that the ISO makes to the transmission grid model.

366 Cal ISO MT, p. 2.

367 There is an outstanding question of the impact on the project sponsor's long-term CRRs from the retirement of existing transmission capacity, of a transmission derating or an increase in external loop flow. Would the CRRs

- Fourth, "if the incorporation of MT related transmission upgrades causes previously awarded CRRs to become infeasible, it is the responsibility of the MT owner to provide counter flow CRR Obligations to relieve the infeasibility only for the terms of those CRRs that were deemed infeasible."[368] It is appropriate to require that the merchant transmission investor accept counterflow CRRs to maintain the feasibility of outstanding CRRs. This provision appears, however, to mean that after the expiration of the term of the CRRs that would have been rendered infeasible by the merchant transmission upgrade the merchant transmission owner would no longer be responsible for any costs that the upgrade imposes on existing rate payers, in terms of decreasing the capacity of the existing transmission system. If this is the intent of this provision, it is not apparent why it is appropriate to provide that the new merchant transmission owner does not have a continuing obligation to compensate transmission customers for the adverse impact that its new facility has on the CRRs that may be allocated to existing ratepayers or auctioned for the benefit of these ratepayers, relative to those that were allocated prior to the upgrade. The Cal ISO should reconsider this provision.

- Fifth, "the CRRs allocated to the MT owners should not at any time become a revenue liability to the MT owner; except in the case of those counter flow CRR Obligations provided to the MT owner to relieve any infeasibility caused by the MT upgrade."[369] This is a reasonable provision that, as proposed by the Cal ISO, is most easily implemented by awarding CRR options, rather than CRR obligations, to merchant transmission owners.

The proposed methodology for allocating CRRs to merchant transmission investors contained in Section 4 of the Cal ISO draft proposal for allocation of CRRs to merchant transmission addresses a number of the complications in assigning CRRs for transmission upgrades. Some additional questions and considerations that the Cal ISO needs to consider as it continues to develop this methodology include:

- The document does not state how many CRRs (source/sink pairs) the merchant transmission owner may nominate during the allocation process. A strength of the Cal ISO proposal appears to be that it will allow the merchant transmission owner to nominate multiple CRR source-sink pairs simultaneously. This is appropriate because in a contingency constrained network it may be difficult for the merchant transmission owner to predict exactly which source/sink CRRs will be incrementally available after the upgrade. It is appropriate that the proposal does not to place any limitations on the location of the sources or sinks for CRR requests for upgrades that are not associated with large generation interconnections.[370]

allocated to the merchant transmission sponsor be guaranteed while the transmission capacity available for the allocation and auction of other CRRs declines?

[368] Cal ISO MT, p. 3.

[369] Cal ISO MT, p. 3.

[370] Cal ISO MT, p. 6.

- Step 4 of Section 4.1 describes how “Capacity CRRs” will be estimated to reserve CRRs that are requested by the merchant transmission owner, but were incrementally feasible on the transmission grid prior to the owner’s upgrade. The documents indicate that this set will be determined so as maximize the MW quantity of CRRs.[371] The Cal ISO might consider, alternatively, identifying this CRR set by maximizing the value, based on the market clearing prices from the last annual auction.

- The proposed use of optimization software, using the nominated CRRs as control variables (as well as Fixed CRRs and Capacity CRRs with large penalty factors), to determine the CRRs allocated to merchant transmission owners is reasonable. The optimization software would determine the best set of feasible CRR allocations based on the merchant transmission owner’s nominations, rather than requiring the merchant transmission owner to guess which CRRs will be incrementally feasible.

  However, it appears that the intention is to optimize the MW quantity of the allocation. This will lead to an allocation that favors nominated CRRs that have the lowest shift factors over the constraints that bind in the security analysis. A drawback is that this allocation may not maximize the value of the awarded CRRs and could lead to anomalous outcomes. As an alternative, the Cal ISO should consider using an objective function that would maximize the value of the CRR set allocated to the merchant transmission owner based on either: the market clearing prices from the last annual auction, or preference values provided by the merchant transmission owner.

- Would the proposed process allow the merchant transmission owner to voluntarily request to be allocated additional counterflow CRR obligations so as to increase its allocation of nominated CRRs?

[371] Cal ISO MT, p. 3

IX. INTERACTIONS

The discussion of the MRTU LMP market design in Sections I through VIII above has been organized around distinct market elements. Several features of the MRTU market design have effects that are important but have not been fully addressed in the preceding sections because the effects involve multiple features of the MRTU market design. We revisit those features in this section.

A. LAP Pricing and Nodal Clearing

We discussed in Section II.B the potential problems arising from the nodal clearing mechanism for zonal load bids. The likely outcome of this clearing mechanism is that no high cost generation would be scheduled to operate within constrained regions in Pass 3 of the DAM. Aside from the ISO revenue adequacy consequences discussed in Section II.B, this clearing mechanism would mean that the DAM nodal prices paid to generators within constrained regions would be systematically less than expected real-time prices.

Low cost generators (i.e., infra-marginal) located within these constrained regions would predictably seek to be paid the real-time price at their location. Because the LAP regions are so large, virtual load bids for the LAP zone in which the low cost generators are located would not enable the generators to receive the real-time price at their location, but only the far lower real-time LAP price. Such generators might respond to such price discrepancies between DAM and real-time by submitting DAM offer prices that reflected expected real-time prices, but as noted above such offer prices would likely be mitigated in Pass 2 of the DAM, and the generator would have to sell its output at the artificially low (because of the absence of congestion) nodal DAM price. One consequence of this would be that while high cost generation within constrained areas might be offered in the RUC, low cost generation within the constrained area would not be offered in the DAM at all, although it would be available in real-time.

One problem for the Cal ISO in such an environment would be that although the Cal ISO would know that this capacity would be available to relieve congestion in real-time (even if the power were exported, the power would relieve congestion), the capacity would not be offered in the DAM. In such a circumstance, the Cal ISO therefore might need to commit capacity in the RUC somewhere to maintain its ability to meet aggregate real-time load (because capacity not offered in the DAM might be exported), but the Cal ISO would not need to commit generation in the RUC to manage congestion. Modeling this situation would require scheduling the generation on in the RUC so that the flows solve congestion but then inserting another load to offset the generation and restore the actual load generation balance. It may not be possible to address this problem without exposing the Cal ISO to arbitrage as well complicating the RUC.

B. LAP Pricing, CRR Allocation and Vertically Integrated LSEs

Another theme that runs through several sections of this report is the problematic impact of LAP pricing for CRR allocation and revenue adequacy. These problems will not arise if the nodal clearing mechanism for zonal load bids is maintained, as under that clearing mechanism there is unlikely to be any congestion in the DAM and all congestion costs will likely be incurred as

RUC costs and real-time uplift (because DAM schedules would be infeasible). If the nodal clearing problem were addressed, however, by zonally clearing zonal LAP bids but the LAP structure retained, then there would probably be congestion in the DAM and the availability of CRRs to hedge these costs would be important.

If the simultaneous feasibility test is applied to the award of LAP CRRs, this will limit the CRRs awarded to those satisfying the most binding transmission constraint, artificially limiting the ability of LSEs to hedge their congestion costs. This outcome would not arise if infeasible LAP CRRs are nevertheless awarded as described under the CMD, but this would likely lead to substantial uplift costs because the CRRs would not satisfy the simultaneous feasibility test and the Cal ISO would therefore not collect enough congestion rents to pay CRR holders.

Another feature of the LAP pricing design is that LSE generation located at the same physical location as LSE load would potentially face apparent congestion between its generation and its load even though the congestion cost is not real (the generation and load are at the same location) and arises only because of the LAP design. The impacted LSEs could be hedged against this artificial congestion by awarding them CRRs from their generation to the LAP. The ability to award such CRRs could be limited by transmission constraints between that generation and the LAP, despite the fact that there would be no congestion between the physical generation and physical load. This circumstance could result in vertically integrated LSEs being unable to acquire the CRRs needed to fully hedge themselves for congestion costs between generation and load at the same physical location. Other vertically integrated LSEs with generation located within high cost constrained areas might benefit, because they would decline to accept CRRs from their generation to the LAP and would be paid counterflow for dispatching their generation to meet their load at the same physical location.

These impacts are cost shifts rather than market inefficiencies but they will affect the willingness of vertically integrated LSEs to join the Cal ISO market. In addition, unless such vertically integrated LSEs are assigned CRRs from their generation to the LAP, they may have an incentive to exercise market power that would not be present under a nodal pricing system.

C. Ancillary Services and Soft Bid Caps

The relatively low level of the proposed bid cap ($250/MWh) raises the possibility that the costs of incremental generation, and market prices, could exceed the bid cap as they often did in 2000-2001. In this circumstance, the soft bid cap could cause energy-limited resources to be dispatched in place of units submitting offers above the bid cap, which could lead to reliability problems. It may be that limited energy units will be able to avoid being inefficiently dispatched for energy by making use of the contingency only flag for spinning and non-spinning reserves. It is not clear, however, whether energy-limited units would be able to submit offers that would enable them to be scheduled to provide reserves if the soft bid cap is in effect. Absent other provisions, scheduling reserves to minimize the as-bid production cost of meeting load would result in energy-limited units offered at $250/MWh being scheduled to provide energy, while reserves would be scheduled on high-cost units offering energy at prices in excess of $250/MWh.

The best way to avoid these problems would be to ensure that the bid cap is raised whenever gas prices rise to a level that would cause the cap to be binding. Other alternatives could be to allow owners of limited energy units to self-schedule these resources to provide reserves or to enforce the bid cap only in Pass 1B and not apply it to capacity scheduled to provide reserves in Pass 1A.

X. RESOURCE ADEQUACY

The general framework of the CPUC resource adequacy proposals addresses three related, but distinct issues. The first is the desirability of LSEs hedging themselves through forward contracts against substantial short-term changes in the market price of power. The second is the desirability of having sufficient capacity available to avoid involuntary load shedding during short-term demand spikes or during a typically large generation or transmission outages. The third is of special importance to the west. This is the desirability of having sufficient energy generating capacity available to avoid involuntary load shedding during periodic low hydro generating conditions.

Each of these goals interacts with the MRTU market design in somewhat different ways and is discussed separately below. Overall, we have not identified any fundamental inconsistencies between the MRTU market design, the general framework described by the CPUC, or the need to effectively address the limitations of eastern installed capacity (ICAP) markets.

A. Short-term Capacity Shortages

1. The Problem

The traditional resource adequacy problem in the electric utility industry arises from the uncertain character of electric demand, the limited ability to store electricity to make up shortages, and the uncertain nature of outages. These properties of the electric industry require that generating capacity be maintained to meet variations in expected load and in the expected availability of generation. In the short-term, this role is met by operating reserve requirements which assure that the required capacity is on-line in real-time. In the longer-term it is necessary to assure that enough capacity is built to maintain short-term reliability, this is the resource adequacy problem. The resource adequacy problem is an economic problem because keeping enough generation available to meet peak load during these extreme conditions is expensive.

Under the vertically integrated utility model, resource adequacy standards (i.e., deciding how much capacity to maintain) were resolved between the individual utility and its regulators. The consequences of inadequate utility resources to meet utility load were straightforward. The utility that lacked sufficient generation to meet its load had to buy energy and schedule transmission to import additional power or it would have to undertake involuntary load shedding. The determination of which LSE would shed load during shortage conditions was easy; this was the utility that was short of power or did not have firm transmission service to deliver power to its load. The Midwest in 1998 saw very high power prices for transactions between vertically integrated utilities seeking to meet reserve requirements and avoid load shedding during shortage conditions. The lesson is that when the distribution company was the entity responsible for dispatching generation it could be held accountable for having sufficient generation to meet load and would balance the high cost of maintaining excess generation against the high costs of buying energy during shortage conditions to avoid load shedding.

The need for resource adequacy mechanisms, such as installed reserve requirements, the precursor of ICAP systems, initially arose in the Northeast from the implementation of economic dispatch which eliminated the link between an entity's generation and load. Individual utilities bought and sold power through the pool and their generating units followed pool dispatch instructions. An individual utility might be a net buyer during a shortage not because it was short of capacity, but merely because that utility's generation was the lowest cost source of operating reserves or regulation. This operating environment led to rules providing for shared responsibility for load shedding within the impacted region of the pools, rather than attempting to assign responsibility to the generation-short distribution company.[372]

Maintaining the capacity needed to meet peak load on a one-day-in-ten-year reliability criterion is very expensive on a per MWh basis. The cost of maintaining seldom used capacity can often far exceed $250 per MWh of generation by the units. Moreover, because marginal capacity will almost never be used, maintaining this capacity can materially raise LSE's overall cost of meeting load. Shared responsibility for load shedding in pools therefore gave rise to the prospect that individual utilities would choose to reduce their costs by not incurring the high cost of maintaining the capacity needed to meet their peak load at conventional reliability levels, knowing that most of any resulting load shedding would be borne by the customers of other utilities. Installed reserve requirements, the pre-cursor of ICAP markets, arose in part to ensure that all pool members incurred the cost of maintaining the capacity needed to meet peak day load on a reliable basis.

Importantly, these earlier reliability structures did not historically rely on prices to allocate energy during shortages within the pool or reserve sharing group. Energy was bought and sold at cost based rates that did not reflect the value of energy or capacity during these shortage conditions. This was the crux of the incentive problem as the price of energy during shortage conditions was far less than the cost of the capacity required to make that energy available. Thus, the resource adequacy problem is to ensure that adequate revenue opportunities exist for generation that is needed to serve load to recover its going forward fixed costs and that adequate incentives or regulations exist for ensuring that needed investments in transmission and new generation are made.

The context in which the CPUC and Cal ISO are addressing resource adequacy is similar to the problem in the Northeast. Prior to 1998, the individual California utilities had an incentive to contract for sufficient resources to meet their customers load, because if they were unable to buy sufficient power to meet their customers load, the utility would have to impose involuntary load shedding on its customers. With the implementation of the Cal ISO this link between utility purchases and customer load shedding was eliminated, so that load shedding was imposed geographically during 2000 and 2001 to reflect the location of shortages and transmission constraints but without regard to which LSEs had paid for enough power to meet their customers load and which had not.

---

372 Of course, to the extent that only a single distribution company served load within the constrained region in which there were inadequate resources available to meet firm load, the load shedding would fall entirely on the responsible distribution company. This will not necessarily be the case, however.

One alternative for maintaining reliability within ISO coordinated markets of the Northeast pools when the pools transitioned to ISO dispatched open access markets was to maintain the reserve requirements of the power pools in some form as a reliability mechanism. The need for such a reliability mechanism was increased by the $1,000/MWh bid cap, the imperfect shortage pricing that existed at start up of the PJM and NYISO. Absent a bid cap, generators could in principle be induced by the prospect of high spot prices during shortages to build and make available enough capacity to maintain reliability even without contracts with LSEs or an ICAP market. But the $1,000 bid cap materially reduced the incentive to keep capacity available in order to supply power during shortages and made it unlikely that spot prices alone would keep the necessary capacity available. The need for a general resource adequacy requirement was reinforced further by the intent of several states to utilize transmission open access to support retail access programs, which would give rise to free-rider problems that would undercut the efforts of individual LSEs to contract for sufficient generation to maintain reliability.

The essence of existing ICAP systems is that LSEs are required to contract for a specified amount of capacity that must be made available in day-ahead markets and is subject to recall by the ISO during shortage conditions, without regard to the price of power in external markets. The ICAP contract therefore assures that the contracted capacity is available to the ISO for commitment and is available to the market rather than exported in real-time during peak conditions. Significantly, ICAP systems do not directly hedge LSEs against the cost of buying power during shortage conditions. The existing ICAP systems may indirectly reduce energy prices by assuring that sufficient capacity is built and kept in operation to avoid capacity shortages. If a capacity shortage nevertheless develops, the impact of ICAP systems on the cost of power is limited to assuring that the power is available at prices no higher than the market price cap.

Thus, a key link between the CPUC resource adequacy mechanism and the MRTU market design for addressing short-term reliability needs is that the resource adequacy mechanism needs to establish a recall right for the resources covered by the mechanism, without regard to external market prices or whether exports were scheduled in day-ahead markets. However, under the current MRTU market design, exports of energy that are scheduled in the day-ahead market are not subject to recall by the Cal ISO in real-time, even to avoid load shedding. This means that as noted in Section II, in the event of regional power shortages, external systems can enter into bilateral contracts at prices in excess of $250/MWh, schedule these exports in the DAM, and be assured that these exports will flow in real-time.

With implementation in California of an ICAP type resource adequacy system with recall rights, power exports scheduled in the DAM would be subject to recall unless they were supported by the capacity of an on-line unit that is not a Cal ISO ICAP resource. In addition, the CPUC resource adequacy mechanism will need to provide a replacement for the must-offer waiver process, to ensure that needed capacity is offered for commitment.

Installed capacity systems have several potential limitations as a solution to the resource adequacy problem that should be kept in mind in designing the California resource adequacy system and these limitations are generally recognized in the Interim Opinion:

- An ICAP system ensures that the electricity market clears by keeping in operation generating capacity that otherwise cannot recover its costs in the energy and ancillary services markets. The cost of keeping this capacity available may exceed its actual value to consumers.

- A set of rules is required to define the types of facilities that constitute qualifying capacity.

- A set of rules is required to govern the location of qualifying capacity.

- A set of rules is required to govern generator operational availability.

- A set of rules is required to govern the treatment of imports.

- There is a potential for free-riding by any loads not required to maintain installed reserves.

- Low energy prices mean that there will be too little incentive for loads to become price-responsive in real time unless this incentive is built into the ICAP system.

- Absent additional rules, an ICAP system ensures the availability of capacity but does not ensure that energy is available in any particular quantity at any particular price from this capacity.

- There is a potential for a short-term ICAP system to become little more than a second payment for energy.

- There is a potential for the exercise of market power that can be difficult to address without undermining other policy goals (reliability, retail access).

2. Market Power

Existing ICAP systems are sometimes seen as a method of addressing locational market power. This is not the case, at least in the short term. If a resource owner has locational market power in short-term energy markets, then it will also have locational market power in ICAP markets. If a resource owner has the ability to profitably drive the energy price to $5,000/MWh by economically withholding its capacity from the energy market, then it will in general also have the ability to profitably drive the ICAP price to a comparable level by withholding capacity from the ICAP market.

It might be argued that there is less potential for the exercise of locational market power in long-term ICAP contracts than in energy spot markets because the ability of LSEs to enter into long-term contracts with generation entrants will often preclude or at least constrain the exercise of market power in such long-term markets. This assessment is likely accurate regarding the competitiveness of long-term generation markets, but it confuses the difference between ICAP and energy markets with long-term vs short-term contracting. The same competitive pressure from entrants exists in long-term energy markets and any lessening of competitive pressure from entrants in short-term energy markets would also exist in short-term ICAP markets. LSEs that are contracting for ICAP a day at a time or a month at a time are not insulated from the exercise of locational market power merely because they are buying ICAP rather than energy.

On the other hand, economic withholding becomes progressively more difficult to identify as the timeframe moves further away from real time. In a centrally dispatched system such as in PJM and New York, generation that is available (taking account of ramp constraints, deratings and environmental limits) and not generating energy or providing reserves or regulation in real time despite market prices that exceed its incremental /opportunity costs is economically withheld. While the application of this criterion can be complex for units managing energy or fuel limits and during periods of volatile gas prices, it is generally possible in LMP markets based on market clearing prices and cooptimization of energy and reserve markets to clearly identify substantial economic withholding in real-time. When we move to the day-ahead timeframe, economic withholding is somewhat less clear cut as a competitive seller would not offer to sell power in the DAM for less than the expected real-time price, regardless of its incremental costs, and the expected real-time price is not observable by the market monitor or regulator. In well designed LMP markets, however, market participants expecting high real-time prices can use virtual load bids to arbitrage any difference between day-ahead and real-time prices while making all of their capacity available for commitment in the day-ahead market and for dispatch in real time.

In ICAP markets, it is harder to identify economic and physical withholding. In an ICAP market, the long-run floor on ICAP prices is provided by the avoidable costs of a unit that would not be recovered in energy and reserve margins. The avoidable costs of a unit can be roughly estimated based on historic costs as can past energy and reserve margins, but these past margins are not necessarily a good measure of current expectations. The shorter the timeframe the ICAP payments applies to, the harder it can be to distinguish economic or physical withholding from an unwillingness to keep money-losing capacity available or to sell for less than the market price.

An ICAP owner might keep a money losing unit on line for a period of days despite zero daily ICAP prices but it would not agree to a forward commitment to keep the unit available for a sustained period of time as an ICAP resource for a zero ICAP price.[373] While the value of the recall right should place a floor under ICAP prices, it is not a ceiling. This is one of the disconnects between daily pricing of ICAP and the term character of capacity decisions that complicates market power analysis as well as the effective functioning of ICAP markets.

Alternatively, one could measure net energy market revenues after the fact and provide generation having locational market power with a guaranteed cost recovery with performance incentives. A locational ICAP system, including such after-the-fact adjustments, has been proposed by NEPOOL.[374]

---

[373] As noted above, the ICAP price is also bounded by the price at which the units ICAP capability could be sold for in adjacent regions or the value of being able to sell non-recallable power into adjacent regions.

[374] Steven E. Stoft, Prepared Direct Testimony on Behalf of ISO New England, Inc., FERC Docket ER03-563-030, August 31, 2004.

3. Deliverability

Existing ICAP deliverability tests are a central issue in implementing ICAP systems in decentralized electricity markets, particularly with respect to the ability of new generators to participate in the ICAP market. PJM, NEPOOL and the NYISO rely on locational pricing for congestion management. This has enabled all three ISOs to adopt a "minimum interconnect" standard for generators selling energy into the market. A new generator satisfies the "minimum interconnect" standard if it is able to deliver its power to the transmission grid without adversely affecting reliability and its interconnection (at zero energy dispatch) does not reduce transfer capability.

LMP pricing in energy markets provides new generators with incentives to site themselves efficiently, without restricting competition. Congestion impacts are reflected in the locational energy prices and thus in the revenues of both incumbents and entrants. Generators that locate at places where they cannot be dispatched because of transmission constraints will earn very low energy margins, incenting new generation to locate where capacity is needed and energy prices higher. Generators may receive ICAP payments, however, whether they operate or not, so there may be no locational price signal in the ICAP market absent some form of deliverability requirement.[375]

Absent any form of deliverability requirement there is a potential for ICAP capacity to be developed in locations at which it is cheap to construct, even if, because of transmission constraints, the capacity adds little to the amount of power that can be used to meet load under stressed system conditions. The more important the ICAP payment is as a source of generator revenue, the greater the potential incentive problem. Thus, if almost all of the net margin of the marginal generator is derived from the energy market, it is less important to impose deliverability requirements in the ICAP market as capacity that is not dispatchable to meet load under stressed system conditions will likely be uneconomic to build or keep in operation regardless of whether a deliverability test is applied for ICAP purposes. The larger the proportion of revenues of the marginal generator that are derived from the ICAP market, however, the greater the potential, absent an ICAP deliverability requirement, for construction of generation that is not cost effective in terms of its contribution to regional reliability.

The Northeast ISOs have struggled with how to apply some form of deliverability test to sellers in the ICAP market and have taken different approaches. Such a test should meet at least three objectives.

- No barriers to entry: The deliverability test should preserve the condition for efficient entry that the entrant's full generating costs need only be less than the avoidable generating costs of the incumbent.

- Permit long-term ICAP contracts: The deliverability test should permit long-term bilateral contracts for ICAP. This requires that ICAP sellers be able to hedge

[375] Deliverability requirements can take many forms, ranging from the locational ICAP requirements of the NYISO to the CETO/CETL tests of PJM.

themselves against the financial impact of entry by competitors on their deliverability.[376]

- Reflect reliability criteria: The deliverability test needs to ensure that capacity eligible for ICAP payments is able to make an appropriate contribution to reliability under stressed system conditions.

In PJM existing ICAP suppliers are grandfathered so the failure of ICAP resources to collectively satisfy the deliverability test does not affect the ability of incumbents to supply ICAP; it only excludes competition from entrants. This grandfathering of incumbents allows generators to enter into multi-year ICAP contracts but it violates the efficient entry condition. This kind of grandfathering of incumbents is a critical issue in evaluating deliverability.

Locational ICAP systems avoid the grandfathering of incumbents in constrained-down regions. One limitation of a locational ICAP system is that the ICAP requirement in small load pockets combined with a price cap essentially amount to an administratively determined capacity payment, with very little role for markets. The reality is that the Con Ed locational ICAP payment generally clears at the price cap set in the Con Ed divestiture contracts. This kind of outcome is even more likely in smaller load pockets with even fewer competing suppliers. The NYISO ICAP demand curve addresses this situation to a degree by allowing locational ICAP prices to vary in a range with changes in supply but the basic reality is that the ICAP payment is largely administratively determined, not market determined, in non-competitive load pockets. In extreme cases, if there is only one supplier in a load pocket, the supplier must effectively be treated like a regulated utility, assured of recovering its going-forward costs and of a return of and on any incremental investments. An ICAP market does not solve this problem.

The CPUC and Cal ISO have recognized the existence of the deliverability test problem in resource adequacy markets. As long as all capacity satisfying the resource adequacy requirement is available for commitment and dispatch by the ISO, the CPUC and the Cal ISO are correct that deliverability tests should be applied at the level of the Cal ISO rather than the utility.[377] The MRTU market design is generally consistent with this approach to deliverability.

4. Outage Performance

A second performance issue arising under ICAP systems is that rules are needed to ensure that the capacity receiving ICAP payments is sufficiently reliable that it is available to meet load under stressed system conditions. This has been addressed by the development of the unforced capacity (UCAP) system in the Northeast.

The UCAP systems calculate an ICAP requirement based on projected generation availabilities based on historical performance and then scale down the amount of ICAP provided by each supplier based on its historical performance. Under the existing UCAP systems, the

---

[376] This is not the same as preventing entry. It is analogous to the ability of generators to hedge themselves against congestion in energy markets through CRR ownership.

[377] Comments of the California Independent System Operator on Workshop Report on Resource Adequacy Issues, Cal ISO, July 14, 2004, p. 27-28.

UCAP capacity of each seller is fixed prior to each auction based on the forced outage performance of that seller's units (their equivalent forced outage rate EFORd) during a prior period. Under all three systems,

$$UCAP = ICAP * (1 - EFOR_d)$$

Units with poor historical forced outage performance therefore earn less money in the ICAP market in the future, motivating them to maintain high levels of availability.

One negative side effect of UCAP systems is that generators may be reluctant to declare outages because of the impact on their ICAP revenues. Instead, they will simply undergenerate and drag on the system. It is therefore important to either have significant penalties for failing to follow dispatch instructions or some other system of performance. Outages and deratings that occur on a high load day are unfortunate but their reliability impact is exacerbated if the system operator is not informed of them and the units do perform as instructed.

A second issue is whether the EFORd index provides sufficient performance incentives for baseload units.[378] The EFORd UCAP systems employed in the Northeast essentially cause an ICAP supplier to receive the ICAP payment in proportion to its availability. Thus, if an ICAP unit were on line 6,650 hours, and out due to forced outage in 350 hours, it would have a 95 percent EFORd rating and would be paid for 95 percent of its capacity (or one can think of this unit as being paid 95 percent of the ICAP price). This would be the case whether the 350 hours of forced outage occurred in the spring when the price of power was $12/MWh and the outage had no reliability impact or if the 350 hours of forced outage occurred in July, the average LMP price was $500 and the outage resulted in load shedding.

Similarly, under most existing ICAP systems the incremental value of staying on line over a day is relatively small. For a unit with around 7,000 hours on line and out of service, the impact on the EFORd of a 24-hour forced outage would be a little more than .3 percent, so would cost a little less than $350/MW for a New York City unit with a $100,000/MW UCAP price. The UCAP system by itself therefore provides baseload units with relatively little incentive to make themselves available under stressed market conditions.

One motivation for the implementation of shortage pricing rules in markets with ICAP systems is to address the potential incentive problem by attempting to ensure that the marginal ICAP supplier recovers a meaningful proportion of its going-forward costs from the energy market during shortage conditions. Units whose 350 hours of forced outage fall during the summer peak risk failing to recover their going-forward costs. Another approach would be to tie the ICAP payment to generator availability during stressed system conditions, measured in some appropriate manner such as participation in the day-ahead market on days with high DAM prices or reserve shortages.

---

[378] A related issue which does not appear to be a problem would be a potential for market participants to simply not offer capacity in real-time without declaring a forced outage. The ISO rules appear to deter such behavior. ISO New England market rules call for imposing a sanction of an amount up to the deficiency charge and imposing a financial sanction equal to the corresponding real-time LMP price. New England Power Pool Market Rule 1, Appendix B, p. 307.

5. Availability Limitations

The existing ICAP markets in the Northeast focus on transmission system deliverability and operational forced outages and deratings to measure generator availability under stressed system conditions. Experience has shown, however, that generation may be deliverable and in perfect operating condition yet unable to meet load under stressed system conditions because of other availability limitations. There are four basic problems that can produce this result: fuel availability, energy limits, restrictive start-up conditions, and restrictive availability conditions. The first three of these limitations have figured prominently in reliability crises over the past several years in the Northeast, California and Texas, while the fourth is of increasing importance as renewable resources are added to the ICAP resource mix.

a) Fuel Availability

While we often think of the summer peak as the time of maximum stress on the transmission and generation system, several reliability crises have arisen in recent years during the winter months. First, most of the load shedding in California during the 2000-2001 crisis occurred during the winter and spring, not the summer. Second, the last time load shedding was necessary on a wide scale in PJM was during the winter of 1993-1994. Third, this past winter NEPOOL came uncomfortably close to requiring load shedding during a winter cold spell. Fourth, Ercot's worst recent reliability crisis came during the winter of 2003, not during its summer load peak.

A problem common to all but perhaps the PJM case was that high demand for electricity was accompanied by a high demand for gas for space heating. This high demand for space heating drove gas prices to very high levels, greatly raising the cost of electricity from gas-fired generation and limiting its availability.

Aside from the price impact of the gas shortages, there are three areas of potential reliability impacts of gas shortages under ICAP systems. First, there can be times that gas-fired generation at some locations simply cannot consume additional gas at any price, because any higher burns would drop pipeline gas pressure below the critical level leading to generation trips and immediate load shedding.

Second, most existing ICAP systems do not require gas-fired generation to contract for firm gas transmission with either the local distribution company or the interstate pipeline. Under traditional LDC pipeline curtailment rules, a lack of firm gas transmission service would mean that a gas-fired generator would not be able to use gas to generate electricity during periods of gas curtailments. In practice in many regions today including California, gas availability is determined by the market, not curtailment rules. In these areas, a generator lacking firm gas transmission service can still generate during periods of gas shortage by buying gas at the market-clearing price.[379] The gas system today is often balanced by customers choosing not to

[379] Thus, while New England gas LDCs and interstate pipelines curtailed non-firm transmission customers during the January 2004 cold spell, generators lacking firm transmission were still able to obtain gas by purchasing it from entities that had firm transmission. See ISO New England, Inc., Market Monitoring Department, "Interim Report on Electricity Supply Conditions in New England during the January 14-16, 2004 `Cold Snap'," May 10, 2004 (hereafter ISO-NE May 2004). Similarly, although there was generally no non-firm transmission

buy gas at high prices rather than by curtailment priorities. Conversely, a generator having firm gas transmission service may sell its gas on days with high gas prices if the electricity price is not sufficiently high to warrant operation.[380] Overall, physical curtailment is only a concern today in areas with generation served under traditional curtailment rules (at the LDC level). Generators are of course not precluded under an ICAP system from contracting for firm gas transmission service, but an ICAP system may diminish their incentive to do so. The crux of an ICAP system is that energy market revenues under shortage conditions are limited by price caps and marginal capacity is kept available by the ICAP payment. If the ICAP payment does not depend on having firm gas supply, the energy market revenues may not be sufficient to cover the cost of contracting for firm gas supply and generators may not do so.[381]

Third, gas market price volatility under stressed market conditions may cause gas fired generation lacking dual fuel capability or hedged gas supply to withdraw from the day-ahead electricity market. As noted above, any individual gas fired generator can in principle be assured of obtaining gas under market based gas systems by offering to pay the market clearing price of gas. The generator would then be able to supply electricity at a cost commensurate with the market price of gas. There is nevertheless a reliability problem. In aggregate it is not true that gas fired generators collectively can acquire all the gas they need at the market clearing price as the gas demand of non-generators may become highly inelastic in the short run as generators increase their gas consumption at the expense of other consumers.[382] This may lead to extreme gas price volatility that may drive unhedged generators out of the gas market.

The potential for gas price volatility to reduce generation supply has several elements. First, consider the position of a gas fired generator operating in a region whose DAM closes after the gas market closes. If the generator purchases gas in the regular day-ahead gas market before offering supply in the day-ahead electric market, the generator risks buying high priced gas that turns out to be uneconomic in the power market. Indeed, this would be likely if gas fired generators collectively offered whatever it took to buy gas in the day-ahead gas market and then sold their excess gas in the in-day gas market. Alternatively, a generator could offer electricity at a high price in the electric market and then buy gas in-day to cover this position if the generator's offer cleared in the day-ahead electricity market. If gas prices are extremely volatile and the gas market thin and illiquid, however, this strategy could also be quite risky with a substantial risk of having to pay a much higher than expected gas price in order to cover the electric market

---

service available to California on El Paso or Transwestern during the winter of 2000-2001, customers lacking firm transmission service could readily acquire daily, bid-week, or term gas at the California border (SoCal Border pricing point).

380 While ISO-NE found that gas-fired generation with firm gas transmission was somewhat more available than gas-fired generation lacking firm transmission, the difference was not dramatic (56 versus 42 percent) and may have been due to other factors (i.e., the gas-fired generators may have had firm gas contracts because they are cogeneration units that operate without regard to the electricity market). See ISO-NE May 2004, pp. 68-69, 72, 141.

381 The MAPP reserve sharing program requires either dual fuel capability or firm gas supply and firm gas transportation for capacity to qualify as ICAP during the winter season. MAPP Reliability Handbook, Section 3.4.7.2.1.

382 The price responsiveness of non-electric generator gas demand may also vary substantially from LDC to LDC. Some gas LDCs may serve a lot of price-sensitive industrial demand that will reduce consumption as prices rise, while other LDCs may serve largely residential demand that is price-inelastic in the short run.

position. The same situation would arise if the day-ahead electric market cleared prior to the day-ahead gas market, the generator could sell power before buying gas, but the generator would then risk not being able to purchase gas at a price low enough to make money generating electricity.[383] The best strategy might therefore be to offer power in the day-ahead market at $1,000/MW and to only run to cover this position if it is possible to buy gas at a sufficiently low cost. If the gas price is too high, the generator simply would not run.[384]

It may at first appear that these reliability impacts are addressed by the must offer requirement of ICAP systems, but this is not the case. A common feature of the NYISO, ISO-NE and PJM ICAP markets is that ICAP resources are obliged to offer their capacity into the day-ahead market of the control area for which they are an ICAP resources.[385] Gas-fired generators lacking dual fuel capability, lacking firm gas supply, or unwilling to risk purchasing extremely expensive gas are therefore required by the market rules to offer their ICAP capacity in the day-ahead market. If these units instead take a forced outage, they suffer a revenue impact in the next ICAP auction. We noted above, however, that the financial impact of such outages could be very small for a baseload unit and much lower than possible losses from selling uneconomic power in the day-ahead market.

More significantly, from a reliability perspective, this behavior is not consistent with the reliability analysis on which the resource adequacy analysis is based. Control area ICAP requirements are based on probabilistic analyses of available generation, transmission and load. Critically, the reliability analyses assumes that forced outages are independent events. It is unlikely in these reliability analyses that a large number of generating units will suffer a forced outage on the same day, so many units with low forced outage rates enable the control area to be confident of satisfying its one-day-in-ten-year reliability criteria. If the “forced outage” is actually a failure to offer capacity due to lack of gas supply combined with a lack of dual fuel capability, then the “forced outages” are not appropriately modeled as independent, instead they are highly correlated across many gas fired units lacking dual fuel capability and the control area may have a much higher reliability risk than indicated by the probabilistic analysis used to develop the resource adequacy requirement. Moreover, as noted above, the actual ICAP revenue impact of a 72-hour forced outage on a baseload gas unit would be very small, providing little incentive to incur substantial costs or risks in order to be available.

The mere fact that reliability problems can emerge is not necessarily a limitation of an ICAP system as these reliability problems could simply be an unavoidable real-world possibility. The underlying problem is that an ICAP system lacking appropriate performance incentives is less effective than forward energy contracts and uncapped day-ahead and real-time prices in

[383] The market power mitigation system could be another source of risk if it does not track contemporaneous gas prices and generators buying gas on day t to cover generation on either day t or t+1 risk having their offer prices mitigated based on the gas prices reported for day t-1.

[384] These kinds of concerns appear to have reduced the supply of gas-fired generation in New England during the January 2004 cold snap. Gas was available for purchase at a price, but the intra-day gas market was thin, and the price volatile and unpredictable. As a result, much gas-fired generation was unavailable due to a “lack of gas supply.” See ISO-NE May 2004, pp. 44, 49-51, 56-65, 104-106.

[385] NEPOOL Manual for Market Operations p 2-11. PJM Operating Agreement, Section 1.10.1A, Day-Ahead Energy Market Scheduling, Sheet 93; NY ICAP Manual, Section 4.8, p. 4-14; NYISO Services Tariff, Section 5.12.7, Sheets 135c, 135d.

providing an incentive for market participants to address these reliability problems. There are at least three such actions that gas-fired generators could potentially take to improve overall reliability. First, gas-fired generators could develop and maintain dual-fuel capability. Second, they could inject gas into production area storage, or contract for LNG deliveries into storage making more gas available at times when the pipeline system is constrained. Third, they could contract for new gas transmission capacity into the region, increasing delivery capacity. None of these actions will be incented by a normal ICAP system.

If permitted by environmental constraints, the ideal solution to winter gas supply reliability and market impacts of gas shortages is the development of gas-fired generation with dual fuel capability. At the time of ICAP market implementation in the Northeast, a substantial proportion of the former utility gas-fired generation had dual fuel capability and routinely switched to oil during periods of high gas demand. It is not clear, however, whether the ICAP-based reliability mechanisms in the Northeast will sustain this capability. The ICAP systems currently do not require dual fuel capacity and a considerable proportion of the gas-fired generation that has been built in the Northeast lacks dual fuel capability, having neither permits nor oil capable burners. Even in those cases in which generating units were permitted as dual fuel, the generators have not in all cases either installed liquid storage or filled the storage. Worse, from a reliability perspective, there is a prospect for material amounts of the existing dual fuel capable generation being shut down and replaced with gas-only generation having a lower heat rate.

If there are environmental restrictions on fuel switching by gas-fired generation, the gas market price can rise far above the cost of oil before fuel switching occurs. The electric system may still be reliable in principle if fuel switching can occur if sufficient gas is not available and load shedding would be required,[386] but market prices can become extremely high under such rules, as was seen in California. Because gas demand may not be highly elastic during winter conditions, the gas price can become so high that electric generating companies are reluctant to buy gas at those prices out of a fear that they will not be able to recover those costs in the power market.

Beyond the mere possession of dual fuel capability, reliability during winter conditions can also be impacted by the amount of oil fuel in storage. Possession of dual fuel capability by gas-fired generation does not help reliability if the fuel oil stocks are not sufficient to keep the generation burning oil for very long. Problems with oil stockpiles have been an issue in most winter reliability crises.

In PJM during 1994 frozen coal piles were accompanied by frozen rivers and ice covered roads that hindered resupply of oil stocks, leading to rolling blackouts by Pepco, PSE&G, Baltimore Gas & Electric, Jersey Central Power, and Vepco. Similarly, ISO-NE reported losing at least 100 MW of oil-fired generation during the January 2004 cold snap due to lack of fuel and

---

[386] It is important for the ISO to coordinate operations with the affected gas distribution companies in these circumstances. Unanticipated ramp ups of electric generation can cause reliability problems on the gas distribution system and dual fuel units cannot instantly switch between gas and oil.

additional outages of oil-fired and dual fuel units may have been related to petroleum fuel constraints.[387]

A second potential incentive problem is that the ICAP systems such as those in place in the Northeast do not require generators to put gas in consumption area storage to meet generation demand when the gas pipeline system is constrained, but availability of storage gas can be important in meeting load.

While generators helped finance new gas pipeline construction in California beginning in 2001, these contracts were driven by high energy prices and forward supply contracts not an ICAP system. A new generator that will obtain most of its revenues from the ICAP market or from summer operation, will have little incentive to contract for firm gas transmission capacity to ensure availability of low cost gas in the winter.

There are several ways to address these kinds of fuel supply constraints. One approach would be to add fuel availability requirements to the ICAP program. MAPP has such requirements in its reserve sharing program. The MAPP reserve sharing program requires either dual fuel capability or firm gas supply and firm gas transportation for capacity to qualify as ICAP during the winter season.[388] In addition,, MAPP requires that units have sufficient fuel storage to enable the unit run during the 4 peak hours five days in succession.[389] A second approach would be to apply some form of derating to resources based on their availability during peak conditions. Thus, gas-fired generation that is unavailable during the winter peak might suffer a derating in addition to the random outage derating reflected in the EFORd. In reviewing the January 2004 cold snap, ISO-NE has mentioned the possibility of adjusting its UCAP rating system to more heavily weight outages during peak periods and has included such a feature in a recent locational ICAP filing.[390] This approach will not be simple to implement, however, without unintended consequences. The reliability issue is not the proportion of time the units are unavailable but the correlation across units and with stressed system conditions. Moreover, if the gas supply for generation during the winter peak is limited, tying generators' receipt of ICAP revenues (which may be thousands of dollars per MW) to whether they outbid other customers for gas during the winter gas peak has the potential to have many unintended consequences on both the gas and electric markets. The derating would therefore be best implemented across all gas generation lacking dual fuel capability, with a collective availability used for this capacity in assessing resource adequacy.

A third approach would be to allow high electric prices during winter reserve shortage periods to incent gas fired generators to maintain dual fuel capability, while high gas prices would incent market participants to fill storage and contract for firm gas transmission. It is doubtful that generators will incur such costs in the future unless they are hedging a forward power contract.

---

387 ISO-NE May 2004, pp. 31, 94-96, 99.

388 MAPP Reliability Handbook, Section 3.4.7.2.1.

389 MAPP Reliability Handbook, Section 3.4.7.2.1.

390 ISO-NE May 2004, p. 144.

b) Start-Up Conditions

The ICAP systems in the Northeast require that ICAP resources offer their capacity in the day-ahead market. It is not always understood, however, that the generators may accompany those offers with start-up times greater than 24 hours so that the capacity is effectively not available for the next day.[391] This kind of offering behavior is likely for rarely used units that are not expected to operate in the near future and are therefore not manned. In these instances, the long-start up times are submitted to enable the plant operator to recall employees and then start the units. The same kind of behavior prevailed prior to deregulation. Some of the problems during the 1993-1994 PJM cold snap were due to misforecasting of demand and an inability to get units not normally used in the winter manned and on-line. The problem also arose in Texas during the winter of 2003 when some new combined cycles were not available because they had not prepared for such low temperatures, while units normally used for summer peaking had been mothballed for the winter and could not come on-line in time when weather forecasts changed.

Long start-up times can have reliability implications, however, as was seen on May 7 and 8, 2000 when a sudden change in weather forecasts in the northeast caused PJM, New York and New England to be generation short. The NYISO tracked the changes in the weather forecast and by Sunday had a high load forecast for Monday but the NYISO could not commit units with 72-hour start times (on Sunday to be available on Monday), yet these units qualified for full ICAP payments.

Some of these reliability problems arising from demand surprises are unavoidable, but they can be exacerbated by an ICAP system. If most of a high cost rarely operated unit's revenues come from the ICAP market and those ICAP market revenues will be received even if the unit is unmanned and requires a 72-hour start-up notice, the unit owner would have little incentive to staff the unit during normally low load periods and the unit may not be available to meet reliability surprises in the fall, winter or spring. Conversely, if that unit were dependent on high prices in the energy market during shortage conditions for its revenues, or if it had signed forward call contracts that it needed to cover, the unit owner would be more willing to incur higher costs in order to make the unit available to operate on a shorter term basis. These incentives could be better aligned by reducing resource adequacy payments to units that are not offered commitment in the DAM during stressed system conditions because of long start-up times.

c) Restrictive Availability Conditions

A final limitation of ICAP systems is their application to resources with restrictive availability conditions, such as wind and solar. While energy limited units have the ability, given an appropriate market design, to ensure that their limited energy is used during peak load conditions, wind and solar units are subject to random availability limits. The treatment of wind units under ICAP systems is particularly problematic as the availability of wind energy is likely

[391] This capacity can be committed by the ISOs but only if they foresee a reliability problem several days in advance.

to be inversely correlated with peak demand. Thus, wind energy output at many projects is likely to be least on hot humid windless summer days when air conditioner load is at its highest.

While the non-availability of wind energy can be treated as forced outages under ICAP systems, this is not sufficient for the purpose of reliability analysis. As noted above with regard to fuel availability, forced outages are treated as random events in the reliability analysis used to develop ICAP requirements, but this treatment will not be accurate if wind energy non-availability is not random but correlated with high demand conditions and correlated across wind units.

The common feature of all of these energy limitations is that existing ICAP systems do not provide enough incentive for capacity to be available during peak conditions. This is intrinsic to ICAP systems that are not tied to availability during stressed system conditions. ICAP systems necessarily pay the generator less for being available during the peak shortage hours than the generator would receive during shortage conditions under an energy only market with a price cap based on the value of lost load. Most existing ICAP systems attempt to compensate for this incentive problem by requiring that resources receiving compensation for providing ICAP demonstrate their capacity. This approach works well for many thermal resources as if the capacity is available, forced outages will be random and this random risk can be analyzed. This approach is not adequate, however, to deal with fuel availability limits and start-up conditions in particular, as resource availability under strained system conditions depends on choices made by the resource owner, and resource owners will not incur the efficient level of costs to maintain availability if they realize limited returns from incurring those costs.

6. Retail Access

Under power pool operation, both PJM and New York had installed capacity requirements imposed on each of the transmission owner LSEs belonging to the pool. Each transmission owner had an obligation to serve load in its service territory so they were able to anticipate their installed capacity requirements and to add capacity when needed to meet load growth. The potential for the exercise of market power by generation-long utilities was constrained by the ability of pool members to anticipate installed capacity requirements and add quick-start capacity to meet ICAP requirements.

This system is fundamentally altered by retail access as LSEs under retail access systems do not know what load they will be serving on any future date and generally only have short-term contracts with their retail customers. Moreover, there is the potential for individual LSEs to satisfy their individual ICAP requirements by releasing customers if ICAP prices are high, suddenly shifting an ICAP or other resource adequacy requirement to the POLR provider.[392] In an environment with retail choice, revenue adequacy markets must, therefore, incorporate mechanisms to accommodate load switching between LSEs without undermining the reliability role of the ICAP requirement. Critically, retail access requires a mechanism for settling day-to-day imbalances in ICAP positions as load shifts between LSEs. The need for a daily balancing system in the ICAP market is potentially problematic as one does not take capacity out of service

[392] See Interim Opinion, pp. 41-43.

or put it in service on a day-by-day basis. The market price of capacity in a daily market absent the exercise of market power, is likely to either be zero or rise to the level of the deficiency payment. Moreover, daily imbalance pricing in the ICAP market can lead to supplier behavior that turns ICAP market prices into a mirror of energy market prices (eliminating the smoothing role) and lead to LSE behavior that undermines the reliability function of the ICAP market.

Furthermore, the volatility of ICAP prices can be exacerbated by daily deficiency charges. Under ICAP systems, LSEs that do not contract for sufficient generation to meet their ICAP requirement must pay a deficiency charge. If the price of ICAP is volatile and varies day by day and LSEs have the option of paying daily deficiency charges instead of buying ICAP, LSEs will have an incentive to buy ICAP when the daily prices is less than the deficiency charge and to pay the deficiency charge when the ICAP price would be higher. This can produce a shortage of ICAP during precisely the days on which it is needed.

The underlying problem that retail access poses for resource adequacy systems is that most retail access systems are fundamentally inconsistent with long-term commitments, which undercuts reliance on the resource adequacy market to support entry and reliance on entry to keep resource adequacy markets competitive. Unfortunately, retail access programs pose similar problems for both ICAP and energy only markets so the problem cannot be addressed by switching to energy-only markets.

Residential customers are unlikely to be willing to sign long-term energy contracts that lock in payments for ICAP over a five to ten year period. Given the typical rate at which people change houses, residential customers signing 5 to 10-year power contracts could constantly be faced with buying out uneconomic contracts or trying to capture the value of in the money contracts from the new owner or renter. This approach is simply not attractive to residential customers so contract duration is actually one year or less, too short to support long-term ICAP or energy purchase contracts. This concern would be reduced for larger commercial and industrial accounts for whom long-term contracts under retail access may pose fewer problems.

Short-term sales contracts with residential consumers could nevertheless in principle support long-term fixed price purchase contracts by LSEs. Oil companies, for example, have supported the construction of crude oil and refined product pipelines through long-term take-or-pay contracts with the pipeline, despite no contract at all with retail motorists. While long-term fixed price ICAP or power purchase contracts would increase the risk of retail access providers, the riskiness of these contracts could be managed by periodically entering into long-term contracts for only a portion of the retailers customer demand.

Such a retailer would lose money when the market price of ICAP or power fell below its long-term contract price, but it would make money when the market price of ICAP or power rose above the long-term contract price. By entering into a temporally diversified set of power purchase contracts, such a retailer could limit its risk expose to sudden swings in market prices.

There appear, however, to be three features of retail access markets that undercut long-term ICAP or energy contracts that are not hedged by long-term customer contracts. First, such long-term contracts would only be economic if the losses incurred when the market price of ICAP is below the contract price were offset by profits when the market price of ICAP is above the contract price. If the retailer were a regulated utility that cannot retain such a difference between the contract price and the market price, then the risk of loss with no offsetting possibility of gain would preclude entering into either long-term ICAP or energy contracts.

Second, unregulated LSEs may be unable to benefit from low contract prices during periods of high ICAP prices if the regulated price to beat does not rise commensurately. Retail rate regulation policies that tend to keep the price to beat too high when ICAP and energy prices are low and high low when ICAP and energy prices are low tend to discourage long-term ICAP and energy contracts. With this retail price structure it is more profitable for unregulated LSEs to shed customers back to the utility when ICAP prices rise than to enter into long-term contracts that hedge ICAP costs.

Finally, the third factor is the risk of regulatory change. ICAP is an artificial product. LSEs may be deterred from entering into long-term ICAP contracts by the risk of regulatory changes. Particularly problematic would be regulatory change which retains the ICAP system, and thus does not trigger clauses terminating payments if the ICAP requirement is terminated, but dramatically reduce the price of ICAP. ISO NE's decision to dramatically reduce the ICAP deficiency payment without eliminating the requirement would be an example of this kind of risk.[393] Another example of regulatory risk is the introduction of locational ICAP in New England. A Boston LSE that had entered into a 10-year ICAP contract would find itself no longer fully hedged.

7. Energy and Capacity Imports

In determining the level of capacity that LSEs need to contract for to maintain short-term reliability it is necessary as the CPUC points out to account for the level of imports that will be available from adjacent systems as a result of load diversity.[394] Moreover, as the CPUC further observes this assessment needs to be made at the level of the Cal ISO, or perhaps the state, not at the level of the individual utility, to ensure that multiple LSEs do not attempt to rely on the same load diversity to reduce their capacity needs.[395] This is the approach taken in defining the aggregate ICAP target in Eastern markets. The ISO makes an assessment of likely peak loads and then subtracts from that the level of imports are expected to be available to meet those peak loads. This remaining responsibility is then assigned to LSEs.

A further element of an ICAP system is the need to account for capacity imports. Traditionally, the power pools assumed that a certain amount of power would be available on some interface under stressed conditions and netted this from the collective pool capacity requirement. To avoid double counting of the import power relied upon, PJM for example, imposed a CBM margin which made most of the external transfer capability unavailable to

---

393 Of course, there is also the risk of tightened requirements and FERC in fact ultimately restored a substantial deficiency payment.

394 Interim Opinion, p. 31-32.

395 Interim Opinion, p. 32.

support firm imports.[396] With the development of explicit recall rules, this logic is less compelling as external ICAP must be dedicated non-recallable capacity.

The NYISO places an overall limitation on the amount of the ICAP requirement that can be met with external resources (2,755 MW) and then places additional interface by interface limits on external resource ICAP imports.[397]

ISO New England also makes the transfer capability of each interface, net of grand fathered agreements and less any tie line benefits assumed in calculating the ICAP requirements, available to support ICAP imports.[398]

One problem that has manifested itself with respect to ICAP imports is the need for more detailed security analysis of import capability, as in some cases transmission maintenance outages can dramatically reduce transfer capability and render ICAP undeliverable. Thus, it is probably not appropriate to make the entire N-1 transfer capability on an external interface available to support ICAP imports, as a single maintenance outage could make much of this ICAP unavailable. There may therefore be a movement toward using N-2 transfer capability to define limits for external ICAP.

A further and rather contentious issue is the treatment of units seeking to split their capacity between pools. One problem with splitting unit capacity between markets is that it complicates monitoring of compliance with outage and derating rules, which are potentially subject to circumvention if a market participate can choose how to assign outages between multiple ICAP markets. Split ICAP units also complicate market power mitigation and enforcement of the DAM bidding requirement, as software needs to account for distinct physical unit upper limits and ICAP upper limits.

B. Forward Hedging

As explained above, ICAP systems tend to constrain energy prices by avoiding shortage conditions and limiting prices to the price cap if regional shortages occur, but they do not directly hedge LSEs against high energy prices, particularly those arising from changes in fuel costs.

Hedging against changes in power prices requires entering into forward contracts for energy. One of the ambiguities in the CPUC resource adequacy proposals is that it is unclear whether they are intended to require that LSEs enter into forward energy purchases or call contracts on capacity or something else. The interim order requires "utilities to forward contract

---

[396] The PJM CBM margin was quite different from CBM margins used in the Midwest. The PJM CBM margin only reserved the sale of this capacity as firm transmission service, thus making it unavailable to support firm imports for ICAP. In real-time all of the transfer capability was made available for use to support non-firm transmission. The PJM CBM margin was simply a mechanism for managing reserve requirements. There was no need to restrict use of CBM capacity to support imports in real-time as these imports met PJM load just as well as emergency imports. See PJM OATT, Attachment C, Sheet 280.

[397] NYISO Installed Capacity Manual, Section 2.7, and Attachment B.

[398] NEPOOL Manual for ICAP, Attachment G, Section 1.5.

for 90 percent of their summer monthly peak needs."[399] This requirement could be interpreted in a number of ways. One interpretation might be that this is intended to require that LSEs contract for energy in all peak hours equal to 90 percent of the monthly peak load. Such contracts would greatly exceed the LSEs actual energy demand, however, in most peak hours. A second interpretation might be that this is intended to require that LSEs contract for energy equal to 90 percent of the average energy consumption over the peak hours. This interpretation would have the LSE buying substantially more than 10 percent of its peak load in the spot market during high priced hours, however, because the average peak hour consumption is typically well below the consumption in the highest hours of the month. A third interpretation might be that it is intended that LSEs contract for energy covering 90 percent of the energy over each hour of the expected load duration curve for the month. Since hourly consumption is not known, this would require that a portion of the purchases be in the form of call contracts of some sort. It is also possible that these comments might be intended to require only that LSEs enter into some kind of ICAP type recall contract for 90 percent of expected monthly peak load. These differences are important and need to be clarified.

Forward energy contracts serve to lock in several different costs. First, they lock in the return to capital which may reduce the cost of financing new generation, reducing prices to consumers. Second, forward contracts can serve to lock in the fuel, emission allowance and variable O&M component. This can be important in providing short-term price stability, but attempting to lock in energy costs over the term of a multi-year contract may raise rather than lower the risk of the contract for the seller and raise the expected price paid by loads.

It is desirable that LSEs contract forward to lock in the price of power for at least a portion of their load, locking in the energy component as well as the capital return prior to the day-ahead market. As observed by the CPUC, however, it is unlikely to be cost effective for LSEs to lock in the energy required to meet 100 percent of their load prior to the day-ahead market. Given the inability to accurately forecast load more than a day or two in advance, such a goal would mean overcontracting for power in forward markets and then selling energy in the day-ahead or real-time markets on most days. It would likely be lower cost for loads to lock less than 100 percent of the potential real-time load in forward markets and to buy in spot markets. Short-term resource adequacy, however, requires the ability to meet 100 percent of potential real-time load, without necessarily locking in the cost.

C. Sustained Energy Shortages

One important difference between western resource adequacy needs and those of the eastern ISOs is the need in the West to account for the impact of long-term annual variations in the quantity of hydro generation.

In eastern markets, the analysis of ICAP requirements and reliability is focused on having enough capacity available to meet demand during short-term system peaks or in the event of short-term generation outages. Thus, an ICAP reliability analysis normally does not ask whether there is enough energy available to meet load over the year. An important feature of the western

---

[399] Interim Opinion, p. 30.

electricity crisis over the period 2000-2001 was that it evolved from a capacity shortage in the summer of 2000 into an energy shortage during the winter of 2000-2001. A different kind of analysis than is typically undertaken in ICAP modeling would be required to assess the risk of energy shortages arising from low hydro conditions in the west.

One way to address these risks would be to take the hydro generation cycle into account both in estimating the quantity of imports that are assumed to be available to meet short-term load spikes and in assessing the ability of the other ICAP resources to provide enough energy to meet load over the year. Given the magnitude of the potential variations in available energy and capacity over the hydro cycle, it would be very expensive to western consumers to maintain through an ICAP type mechanism enough excess generation to make up for the loss of hydro generation during low hydro years. The interim order does not clearly address the treatment of the western hydro cycle.

If such adverse hydro conditions can be accurately forecasted, a lower cost approach might be to rely in part of bringing high cost mothballed generation back in service or raising energy prices to increase conservation during low hydro years.

---

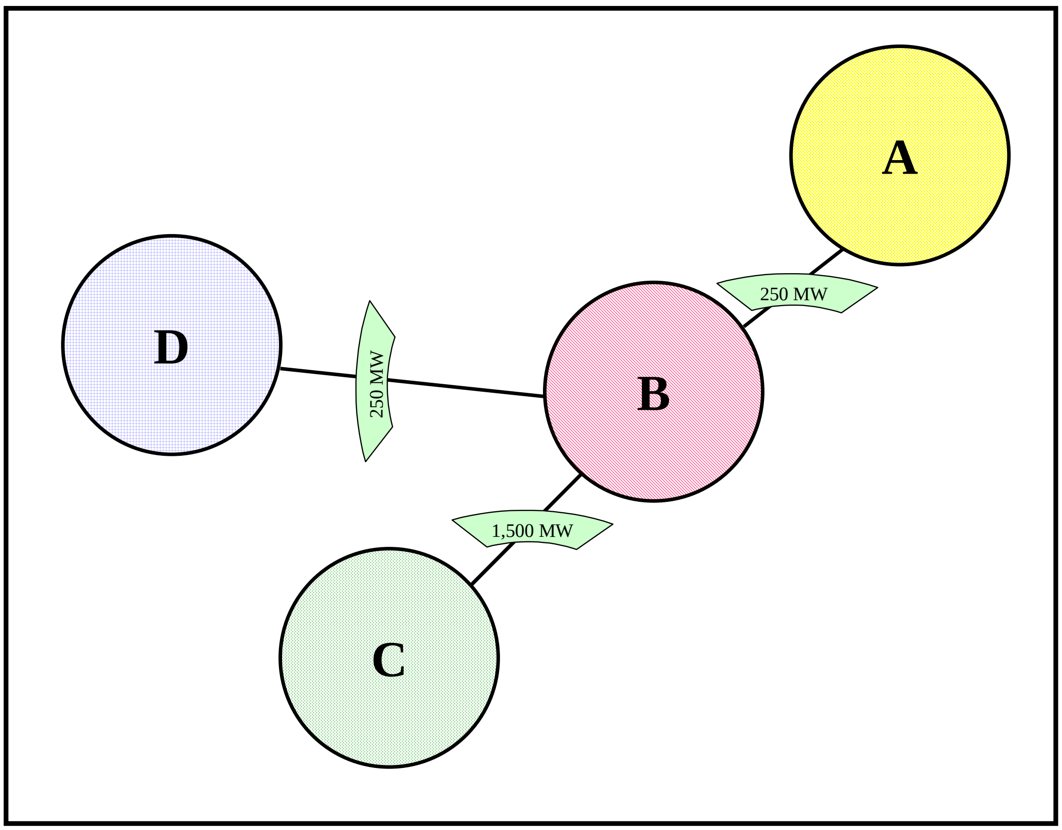
[i] Scott Harvey is a Director and Susan Pope is a Principal at LECG, LLC. William W. Hogan is the Lucius N. Littauer Professor of Public Policy and Administration, John F. Kennedy School of Government, Harvard University and a Director of LECG, LLC. This paper draws on work for the Harvard Electricity Policy Group and the Harvard-Japan Project on Energy and the Environment. The authors are or have been consultants on electric market reform and transmission issues for Allegheny Electric Global Market, American Electric Power, American National Power, Australian Gas Light Company, Avista Energy, Brazil Power Exchange Administrator (ASMAE), British National Grid Company, California Independent Energy Producers Association, California Independent System Operator (Cal ISO), Calpine Corporation, Central Maine Power Company, Comision Reguladora De Energia (CRE, Mexico), Commonwealth Edison Company, Conectiv, Constellation Power Source, Coral Power, Detroit Edison Company, Duquesne Light Company, Dynegy, Edison Electric Institute, Edison Mission Energy, Electricity Corporation of New Zealand, Electric Power Supply Association, El Paso Electric, GPU Inc. (and the Supporting Companies of PJM), GPU PowerNet Pty Ltd., GWF Energy, Independent Energy Producers Assn, ISO New England, Maine Public Advocate, Maine Public Utilities Commission, Midwest ISO, Mirant Corporation, Morgan Stanley Capital Group, National Independent Energy Producers, New England Power Company, New York Independent System Operator, New York Power Pool, New York Utilities Collaborative, Niagara Mohawk Corporation, NRG Energy, Inc., Ontario IMO, Pepco, Pinpoint Power, PJM Office of Interconnection, PP&L, Public Service Electric & Gas Company, Reliant Energy, Rhode Island Public Utilities Commission, San Diego Gas & Electric Corporation, Sempra Energy, SPP, Texas Utilities Co, TransÉnergie, Transpower of New Zealand, Westbrook Power, Williams Energy Group, and Wisconsin Electric Power Company. The views presented here are not necessarily attributable to any of those mentioned, and any remaining errors are solely the responsibility of the authors. (Related papers can be found on the web at www.whogan.com).

Appendix I
Nodal Clearing and Settlement Process for Zonal/LAP Bids

A. Introduction

This appendix illustrates the potential unintended consequences of the proposed nodal clearing and settlement process for zonal/LAP load bids in the day-ahead market. The example does not rest on the possession of market power by any market participant but illustrates the impact of competitive profit-seeking behavior in conjunction with the proposed settlement rules. Suppose the aggregate load zone/LAP includes regions A, B, C and D, as portrayed in Figure I-1.

Figure I-1
AGGREGATED LOAD ZONES



Let us further suppose that the locations A and D in Figure I-1 both have 1,000 MW of peak load, 1,000 MW of generation, and 250 MW of transfer capability from B to A and B to D. While peak load in these regions can in principle be met without relying on imports, the import capability allows LSEs serving load at A and D to take advantage of low cost energy available on the spot or term market. In addition, we assume that there is 2,500 MW of peak load in region B, 3,000 MW of peak load in region C and transfer capability of 1,500 MW from C to B; see Table I-2. There is a single LAP covering regions A, B, C and D.

Table I-2 NODAL LOAD		
Region	**Load**	**Import Capability**
A	1,000	250
B	2,500	1,500
C	3,000	0
D	1,000	250
Total	7,500	

Table I-3 portrays a set of assumed generation costs, with high cost generation required at the margin to meet load in regions A and D, which are for simplicity of the example assumed to be radially connected as shown in Figure I-1.

Table I-3 GENERATION COSTS				
Region	**Capacity (MW)**	**Generation (MWh)**	**Cost ($/MWh)**	**Total Cost ($)**
A	500	250	45	11,250
	500	500	40	20,000
Total A	1,000	750		31,250
B	300	0	44	0
	650	150	42	6,300
	600	600	40	24,000
	750	750	35	26,250
Total B	2,300	1,500		56,550
C	3,500	2,000	40	80,000
	2,500	2,500	35	87,500
Total C	6,000	4,500		167,500
D	500	250	80	20,000
	500	500	60	30,000
Total D	1,000	750		50,000
Total	10,300	7,500		305,300

The expected real-time dispatch, regional prices, and load weighted cost are summarized in Table I-4.

Table I-4
EXPECTED REAL-TIME PRICES AND DISPATCH

Region	Load MW (A)	LMP Price $/MWh (B)	Load Cost $ (C)	Generation Dispatch MW (D)	Net Imports MW (E)
A	1,000	45	45,000	750	250
B	2,500	42	105,000	1,500	1,000
C	3,000	40	120,000	4,500	-1,500
D	1,000	80	80,000	750	250
Total	7,500		350,000	7,500	0
LAP Price			46.66667		

Load Weight: (C) = (A) * (B)
Net Imports: (E) = (A) – (D)

B. Passive LSE Bidding

We initially assume that the LSE's serving load in the LAP bid the entire expected real-time load, 7,500 MW into the DAM, at slightly above the expected real-time zonal/LAP price, as shown in Column C of Table I-5. In New York or PJM, all of this load would clear at the zonal/LAP price, which would be $46.67/MWh as shown in Table I-4.

Table I-5
BASE CASE DAM SETTLEMENTS
7,500 MW Bid in at $50

	Gen Offer Price $/MWh	Load Weights (B)	Bid Load MW (C)	Dispatched Generation MW (D)	Cleared Load MW (E)	LMP Price $/MWh (F)	Load Cost $ (G)	Gen Revenues $ (H)
A	40-45	0.133333	1,000	750	1,000	45	45,000	33,750
B	35-42	0.333333	2,500	1,500	2,500	42	105,000	63,000
C	35-40	0.4	3,000	4,500	3,000	40	120,000	180,000
D	60-80	0.133333	1,000	0	250	42	10,500	0
Total		1.0	7,500	6,750	6,750		280,500	276,750
LAP Price							41.55556	
Congestion Rent							3,750	

(G) = (F) * (E)
(H) = (F) * (D)

Under the proposed MRTU nodal clearing process, however, the bid load would be assigned to the four regions based on the ISO-determined load weights portrayed in column (B). The bid load allocated in this manner would then be cleared nodally. All of the bid load allocated to regions A, B and C would clear, but only 250 MW of the load allocated to region D would clear, as load levels in excess of 250 MW would require dispatch of generation at prices of $60/MWh or higher and thus price-capped load bid in at $50/MWh would not clear in region D above 250 MW as shown in columns (D) and (E) of Table I-5.

Total load of 6,750 MWh would therefore clear in the DAM at a LAP price of $41.56. Total load payments would exceed generator revenues as shown in columns (G) and (H) of Table I-5, causing the ISO to collect congestion rents of $3,750.

As shown in Appendix VI, 1,875 MW of CRRs could be awarded from generation in zone C to the LAP. The settlements for 1,875 MW of Zone C to LAP CRRs for the day-ahead market portrayed in Table I-5 are shown in Table I-6.

Table I-6
CRR SETTLEMENTS

CRRs	Source	Sink	Sink Price $/MW	Source Price $/MW	CRR Value $/CRR	CRR Payments $
1,875	C	LAP	41.55556	40	1.555556	2,916.667
Total						2,916.667

The CRR value is $1.56/MW, entailing total payments of $2,916.67, reducing the DAM congestion surplus to $833.33 as shown in Table I-7.

Table I-7
NET DAM CONGESTION SETTLEMENTS

Congestion Rents	3,750
CRR Payments	-2,916.667
Net Surplus	833.3333

In real-time, generation would be dispatched to meet the entire 7,500 MW of real-time load, and an additional 750 MW of high cost generation not scheduled in the DAM would be dispatched in region D and paid the real-time nodal LMP price, $80/MWh, as shown in columns (E) and (F) of Table I-8. The LAP price would be $46.67, so the ISO would sell an additional 750 MW of power in region D at the real-time LAP price, collecting $35,000 in net real-time revenues, but would pay $60,000 for the dispatch of generation in region D.

Table I-8
REAL-TIME SETTLEMENTS

	Gen Offer Price $/MWh (A)	LMP Price $/MWh (B)	Load MW (C)	Load Cost $ (D)	Gen Dispatch MW (E)	Gen RT-DAM MW (F)	RT Gen Revenues $ (G)
A	40-45	45	1,000	45,000	750	0	0
B	35-42	42	2,500	105,000	1,500	0	0
C	35-40	40	3,000	120,000	4,500	0	0
D	60-80	80	1,000	80,000	750	750	60,000
Total			7,500	350,000	7,500		60,000
LAP Price				46.66667			
RT Load Imbalance		46.667				-750	-35,000
RT Revenue Shortfall							25,000
Combined Market Shortfall							24,166.67

The ISO would therefore incur a real-time revenue shortfall of $25,000, considerably in excess of the $833.33 of excess congestion rents collected in the DAM, leading to a combined shortfall of $24,166.67. The fundamental problem giving rise to the shortfall is that the LSE load bids failed to clear in the DAM in region D, but the real-time settlements system effectively treats the DAM load bids as if they cleared proportionately throughout the load zone, and thus the real-time imbalance is also spread proportionately throughout the load zone and settled at the real-time LAP price. Instead of the underbid LSEs having real-time imbalances in region D, which they would have to settle at $80/MWh if each region were a pricing zone, they settle their imbalances at the $46.67/MWh LAP price. The ISO, however, must settle imbalances in the generation market at $80/MWh.

C. Inflated Load Bids

The potential ISO revenue shortfalls under the MRTU LAP settlement rules could be even larger than suggested by the passive bidding example. If LSEs within the LAP recognize that their LAP bids are cleared nodally, they can reduce their costs and avoid exposure to real-time prices by bidding more than their expected load into the DAM, at prices designed not to clear at the high priced nodes within the LAP. The aggregate load cleared in the DAM could correspond reasonably well to their expected real-time load, yet no generation would have been scheduled on high-cost units in constrained areas.

Table I-9 portrays such a circumstance in which LSEs within the LAP bid 8,365.385 MW of load into the DAM. We assume that 1,875 MW of load hedged by CRRs is bid in at $250/MWh (because it is hedged), and the remaining 6,490.385 MW is bid in at a price cap of $50/MWh as shown in columns (C) and (E) of Table I-9. The 1,875 MW of load bid in at $250 clears in each zone. The remaining 6,490.385 of bid load clears in zones A, B, and C as shown in column (F), but not in zone D as the nodal price would need to reach $60/MWh for generation to be dispatched to meet load above 250 MW in Zone D. It is also shown in column (G) that the total load cleared in the DAM is 7,500 MW at a zonal price of $41.55/MWh, generating DAM congestion rents of $3,750.

Table I-9
DAM SETTLEMENTS
INFLATED LOAD BIDS

	Gen Offer Price $/MW (A)	Load Weights (B)	Bid Load @ $250/MW MW (C)	Cleared Load MW (D)	Bid Load @ $50/MW MW (E)	Cleared Load MW (F)	Total Cleared Load MW (G)	Dispatched Generation MW (H)	LMP Price $/MWh (I)	Load Cost $ (J)	Gen Revenues $ (K)
A	40-45	0.133333	250	250	865.3847	865.3847	1,115.385	865.38467	45	50,192.31	38,942.31
B	35-42	0.333333	625	625	2,163.462	2,163.462	2,788.462	1,788.4617	42	117,115.4	75,115.39
C	35-40	0.4	750	750	2,596.154	2,596.154	3,346.154	4,846.154	40	133,846.2	193,846.2
D	60-80	0.133333	250	250	865.3847	0	250	0	42	10,500	0
Total		1.0	1,875	1,875	6,490.385	5,625	7,500	7,500.0003		311,653.9	307,903.9
LAP Price										41.55385	
Congestion Rent										3,750	

Table I-10 shows the payments to the 1,875 MW of Zone C to LAP CRRs, are $1.55/CRR and total $2,913.46.

Table I-10
CRR SETTLEMENTS

CRRs	Source	Sink	Sink Price $/MW	Source Price $/MW	CRR Value $/CRR	CRR Payments $
1,875	C	LAP	41.55385	40	1.553846	2,913.462
Total						2,913.462

The net DAM congestion settlements therefore have a surplus of $836.54 as shown in Table I-11.

Table I-11
NET CRR SETTLEMENTS

Congestion Rents	3,750
CRR Payments	-2,913.462
Net Surplus	836.5385

In real-time the ISO would need to dispatch additional high cost generation in region D, while backing down low cost generation in regions A, B and C as shown in column (F) of Table I-12, but there would be no net load or generation imbalances between the DAM and real-time. There would therefore be no load buying power at the real-time LAP price, but the ISO would incur an additional $28,846.14 in generation costs to dispatch generation up in region D and down in the lower cost regions as shown in Table I-12. The real-time shortfall would rise to $28,846.14, leaving a combined market shortfall of $28,009.6.

Table I-12
REAL-TIME SETTLEMENTS
INFLATED LOAD BIDS

Region	Gen Offer Price $/MWh (A)	LMP Price $/MWh (B)	Load MW (C)	Load Cost $ (D)	Gen Dispatch MW (E)	Gen RT-DAM MW (F)	RT Gen Revenues $ (G)
A	40-45	45	1,000	45,000	750	-115.3847	-5,192.31
B	35-42	42	2,500	105,000	1,500	-288.4617	-12,115.39
C	35-40	40	3,000	120,000	4,500	-346.154	-13,846.16
D	60-80	80	1,000	80,000	750	750	60,000
Total			7,500	350,000	7,500	-0.000333	28,846.14
LAP Price				46.66667			
RT Load Imbalance		46.66667				0	0
RT Revenue Shortfall							28,846.14
Combined Market Shortfall							28,009.6

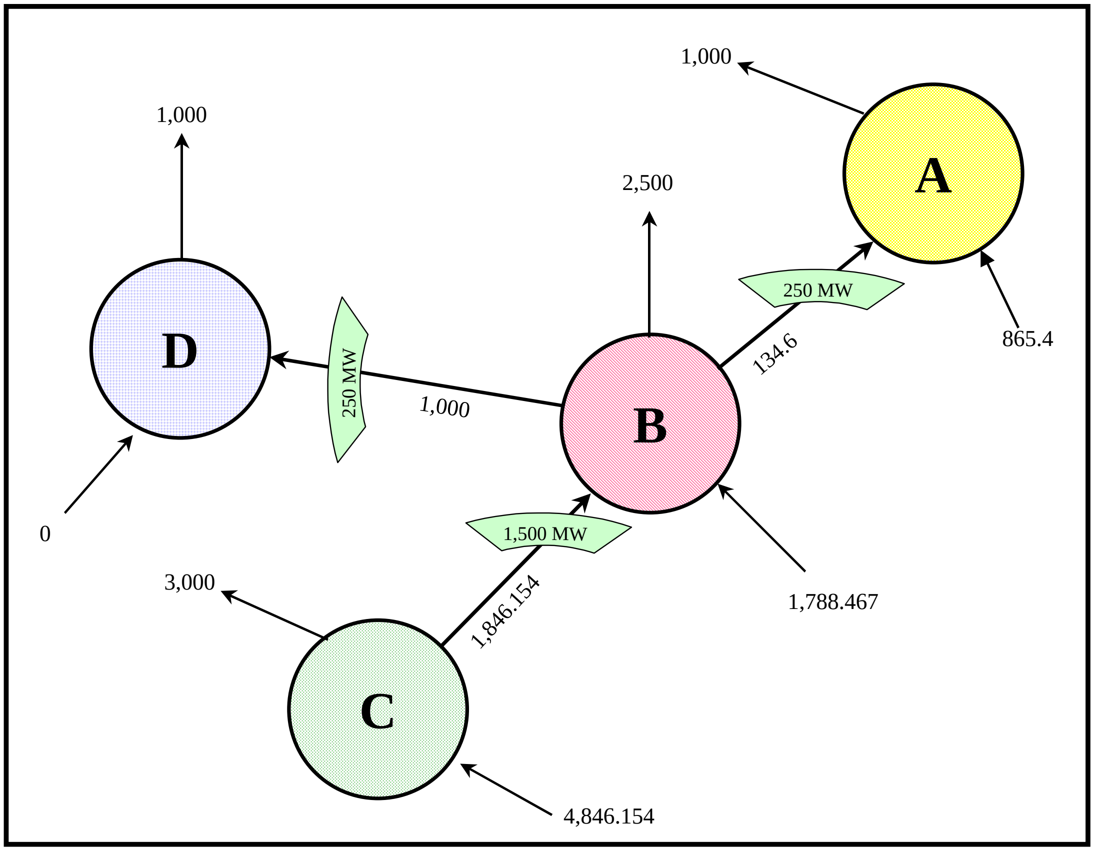
Another way of looking at this problem is that the DAM schedules are not feasible. Table I-13 shows the generation and load schedules in the DAM settlements. The load quantities are determined by the real-time load weights applied to the total LAP quantity cleared in the DAM. The generation quantities are determined by the DAM generation schedules.

Table I-13
DAM SETTLEMENTS

	Load Weights	Settlement Load	Settlement Generation
A	.13333	1,000	865.38467
B	.3333	2,500	1,788.4617
C	.40	3,000	4,846.154
D	.1333	1,000	0
		7,500	

Figure I-14 portrays the transmission flows associated with these DAM settlement quantities.

Figure I-14



It can be seen that the flows on both B-D and C-B exceed the limit. The DAM financial schedules are therefore infeasible and will lead to revenue inadequacy in real time.

It is important to recognize that while region C has been described as a large region with many competitive suppliers, the LAP pricing system would apply in the same way if region C were a generation pocket containing a single generator and accounting only for 1 percent of LAP load. The LAP clearing and settlement rules would produce the same kind of infeasible DAM schedules portrayed in Figure I-14 but, absent competition, the single supplier in region C could allow its DAM schedules to be scheduled with -$30/MWh DEC bids in real-time, greatly increasing the Cal ISO uplift costs in the example. If region C is a generation pocket, the situation portrayed in Figure I-14 is no different from the infeasible DAM schedules that give rise to the behavior referred to as the "INC/DEC game" under the current market design.

D. Excess Load Scheduling in DAM

The settlements portrayed in Tables 9 and 12 are not the worst case outcome under the LAP pricing system. It should be noted that the DAM LAP price in Table I-9 was only $41.55, while the real-time price was $46.67. Every MW in excess of the real-time load that could be purchased day-ahead could therefore be sold back in real-time for a profit, as long as the DAM LAP price is less than the real-time LAP price. LSEs can therefore find it profitable under the proposed LAP system to clear load in the DAM that is in aggregate in excess of their expected real-time load. To illustrate this, suppose that the LSEs in regions A, B, C and D again submit 1,875 MW of load bids at a price of $250/MWh, but now submit 8,000 MW of load bids at a price of $45/MWh as shown in Columns (C) and (E) of Table I-15.

Table I-15
BASECASE DAM SETTLEMENTS
EXCESS LOAD SCHEDULES

	Gen Offer Price $/MWh (A)	Load Weights (B)	Bid Load @ $250/MW MW (C)	Cleared Load MW (D)	Bid Load @ $45/MW MW (E)	Cleared Load MW (F)	Total Cleared Load MW (G)	Dispatched Generation MW (H)	LMP Price $/MWh (I)	Load Cost $ (J)	Gen Revenues $ (K)
A	40-45	0.133333	250	250	1,066.667	1,000	1,250	1,000	45	56,250	45,000
B	35-42	0.333333	625	625	2,666.667	2,666.667	3,291.667	2,291.6667	44	144,833.3	100,833.3
C	35-40	0.4	750	750	3,200	3,200	3,950	5,450	40	158,000	218,000
D	60-80	0.133333	250	250	1,066.667	0	250	0	44	11,000	0
Total		1	1,875	1,875	8,000	6,866.667	8,741.667	8,741.6667		370,083.3	363,833.3
LAP Price										42.33556	
Congestion Rent										6,250	

(G) = (D) + (F)
(J) = (G) * (J)
(K) = (H) * (I)

Only 1,000 MW of generation is available to be dispatched in region A, so only 1,250 MW of load clear in that region , and the $3{,}291\frac{2}{3}$ MW of load bid in region B requires the dispatch of even higher cost generation than in prior examples, raising the nodal price to $44/MWh. Overall, 8,741.67MW of generation clear at a DAM LAP price of $42.34/MWh. At the higher nodal prices, congestion rents collected in the DAM rise to 6,250, and CRR payments rise to $4,379.17, as shown in Table I-16.

Table I-16
CRR SETTLEMENTS

CRRs	Source	Sink	Sink Price $/MW	Source Price $/MW	CRR Value $/CRR	CRR Payments $
1,875	C	LAP	42.33556	40	2.335558	4,379.171
Total						4,379.171

Overall, therefore, the ISO DAM congestion account shows a surplus of $1,870.83.

Table I-17 NET CRR SETTLEMENTS	
Congestion Rents	6,250
CRR Payments	-4,379.171
Net Surplus	1,870.829

In real-time, however, actual load is only 7,500 MW, and the zonal/LAP price is $46.67. The ISO has to dispatch up high cost generation in region D while backing down substantial cheap generation in regions A, B and C. Because the ISO backs down more generation than it buys, it receives net payments of $22,500 from generators in real-time as shown in column G of Table I-18. LSE DAM load schedules, however, are also well above real-time loads, however, so the ISO must buy back 1,241.67 MW of DAM load schedules at the real-time price of $46.67, requiring payments to LSEs of $57,944.4, leading to a net revenue shortfall in real-time of $35,444.44 and a combined DAM and real-time shortfall of $33,573.62.

Table I-18
REAL-TIME SETTLEMENTS
EXCESS LOAD SCHEDULES

	Gen Offer Price $/MWh (A)	**LMP Price $/MWh (B)**	**Load MW (C)**	**Load Cost $ (D)**	**Gen Dispatch MW (E)**	**Gen RT-DAM $ (F)**	**RT Gen Revenues $ (G)**
A	40-45	45	1,000	45,000	750	-250	-11,250
B	35-42	42	2,500	105,000	1,500	-791.6667	-33250
C	35-40	40	3,000	120,000	4,500	-950	-38,000
D	60-80	80	1,000	80,000	750	750	60,000
Total			7,500	350,000	7,500	-1,241.667	-22,500
LAP Price				46.66667			
RT Load Imbalance		46.66667				1,241.667	-57,944.44
RT Revenue Shortfall							35,444.44
Combined Market Shortfall							33,573.62
(D) = (C) * (B) (G) = (F) * (B)							

E. Perfect Arbitrage

The examples above illustrate problems with nodal clearing of zonal bids in the DAM. The examples all also show a gap between the prices in the DAM LAP and the LAP in real time.

Hence, there is an arbitrage opportunity. The lack of equilibrium in the examples, however, does not cause the real-time shortfall in the ISO's settlements. The same revenue shortfall would be present even with perfect arbitrage. This can also be readily illustrated using the preceding example. Suppose that the generation supply curve in regions A, B, and C was more steeply upward sloping so that the market clearing price in each region was $46.67 at the quantities cleared in the prior example, Table I-15. This situation is portrayed in Table I-19 below.

Table I-19
BASECASE DAM SETTLEMENTS
Perfect Arbitrage

	Gen Offer Price $/MWh (A)	Load Weights (B)	Bid Load @ $250/MW MW (C)	Cleared Load MW (D)	Bid Load @ $45/MW MW (E)	Cleared Load MW (F)	Total Cleared Load MW (G)	Dispatched Generation MW (H)	LMP Price $/MWh (I)	Load Cost $ (J)	Gen Revenues $ (K)
A	40-45	0.133333	250	250	1,066.667	1,000	1,250	1,000	46.67	58,337.5	46,670
B	35-42	0.333333	625	625	2,666.667	2,666.667	3,291.667	2,291.6667	46.67	153,622.1	106,952.1
C	35-40	0.4	750	750	3,200	3,200	3,950	5,450	46.67	184,346.5	254,351.5
D	60-80	0.133333	250	250	1,066.667	0	250	0	46.67	11,667.5	0
Total		1	1,875	1,875	8,000	6,866.667	8,741.667	8,741.6667		407,973.3	407,973.6
LAP Price										46.67	
Congestion Rent										0	

In this hypothetical all buyers in the day-ahead market pay $46.67 for power and the generators scheduled in regions A, B and C are all paid $46.67 for power. There are no congestion rents and CRRs between the regions and the LAP have no value as shown in Table I-20.

Table I-20
CRR SETTLEMENTS

CRRs	Source	Sink	Sink Price $/MW	Source Price $/MW	CRR Value $/CRR	CRR Payments $
1,875	C	LAP	46.67	46.67	0	0
Total						0

In real-time the ISO must dispatch up high cost generation in region D and back down lower cost generation just as in the preceding example, as well as buying back excess load schedules at the real-time price. Table I-21 shows that the perfect arbitrage in the DAM has no impact on the ISOs real-time settlement short-fall which is exactly the same as in Table I-18 above.

Table I-21 REAL-TIME SETTLEMENTS							
	Gen Offer Price $/MWh (A)	**LMP Price $/MWh (B)**	**Load MW (C)**	**Load Cost $ (D)**	**Gen Dispatch MW (E)**	**Gen RT-DAM MW (F)**	**RT Gen Revenues $ (G)**
A	40-45	45	1,000	45,000	750	-250	-11,250
B	35-42	42	2,500	105,000	1,500	-791.6667	-33,250
C	35-40	40	3,000	120,000	4,500	-950	-38,000
D	60-80	80	1,000	80,000	750	750	-60,000
Total			7,500	350,000	7,500	-1,241.667	-22,500
LAP Price				46.66667			
RT Load Imbalance		46.66667				1,241.667	57,944.44
RT Revenue Shortfall							35,444.44
Combined Market Shortfall							35,444.44

F. Conclusion

The actual operation of the LAP clearing and settlement mechanism would have more complex effects than illustrated in these examples. The failure to schedule generation in zone D in Pass 3 could cause higher cost generation to serve real-time load than would otherwise have been the case, raising the real-time price above $80/MWh. LSEs could further reduce their costs by bidding in load at prices even below the expected LAP price, thus capping the amount they pay.[1] As more and more phantom load was bid into the DAM in the load price zones B and C, the clearing prices in those zones would rise towards the actual real-time price diminishing the incentive to schedule additional phantom generation. Overall, however, the proposed clearing and settlement rule would lead to inefficient arbitrage that raises rather than lowers the cost of meeting load.

[1] Thus, in the example portrayed in Table 13, an LSE could have bid load into the DAM at $41/MWh. This load bid would have cleared only in zone C and the LSE would have been charged only $41/MWh in the DAM for these load schedules.

Appendix II
Arbitrage of LAP Load Weights

The same simple transmission system used in Appendix I can be used to illustrate the impact of arbitrage by LSEs of errors in Cal ISO load weight forecasting. To focus on this issue, we assume that the problems associated with the proposed clearing and settlement mechanisms for zonal/LAP load bids are addressed by clearing zonal/LAP load bids against the zonal/LAP price.

Tables II-1 through II-3 illustrate the settlements under such a system if the nodal load weights used by the ISO in the DAM accurately reflected the real-time distribution of load. Table II-1 shows 7,500 MW of load bid into the DAM at $50/MWh, all of which clears at the zonal/LAP price of $46.67/MWh, as shown in Table II-1.

Table II-1
BASE CASE DAM SETTLEMENTS
7,500 MW Bid in at $50

	Gen Offer Price $/MWh (A)	Load Weights (B)	Bid Load MW (C)	Dispatched Generation MW (D)	Cleared Load MW (E)	LMP Price $/MWh (F)	Load Cost $ (G)	Gen Revenues $ (H)
A	40-45	0.133333	1,000	750	1,000	45	45,000	33,750
B	35-42	0.333333	2,500	1,500	2,500	42	105,000	63,000
C	35-40	0.4	3,000	4,500	3,000	40	120,000	180,000
D	60-80	0.133333	1,000	750	1,000	80	80,000	60,000
Total		1.0	7,500	7,500	7,500		350,000	336,750
LAP Price							46.66667	
Congestion Rent							13,250	

(G) = (E) * (F)
(H) = (D) * (F)

CRR payments to CRR holders would be $6.67/CRR totaling $12,500, out of total DAM congestion rents of $13,250, leaving a DAM congestion rent surplus of $750, as shown in Table II-2.

Table II-2 BASE CASE CRR SETTLEMENTS						
CRRs	**Source**	**Sink**	**Sink Price $/MWh**	**Source Price $/MWh**	**CRR Value $/CRR**	**CRR Payments $**
1,875	C	LAP	46.66667	40	6.666667	12,500
Total						12,500
Net DAM Congestion Settlements:						
Congestion Rents:						13,250
CRR Payments:						-12,500
Net Surplus:						750

In real-time, we have assumed that the ISO's load forecast is exactly right, so there are no deviations between the DAM generation schedules and the real-time dispatch. There are no net ISO revenues in real-time, so the combined DAM/Real-time settlements equal the $750 DAM surplus as shown in Table II-3.

Table II-3 BASE CASE REAL-TIME SETTLEMENTS							
	Gen Offer Price $/MWh (A)	**LMP Price $/MWh (B)**	**Load MW (C)**	**Load Cost $ (D)**	**Gen Dispatch MW (E)**	**Gen RT-DAM MW (F)**	**RT Gen Revenues $ (G)**
A	40-45	45	1,000	45000	750	0	0
B	35-42	42	2,500	105,000	1,500	0	0
C	35-40	40	3,000	120,000	4,500	0	0
D	60-80	80	1,000	80,000	750	0	0
Total			7,500	350,000	7,500		
LAP Price				46.66667			
RT Load Imbalance		46.66667				0	0
RT Revenue Shortfall							0
Combined Market Shortfall							-750

We can now illustrate the impacts of errors in the ISO's estimate of real-time load weights by examining the impact of load weights for region D in the DAM, which are too low or too high, with offsetting errors in region C. Table II-4 shows the real-time settlements in the event that the real-time load in Zone D is lower than forecast by the ISO in the DAM (850 MW instead of 1,000 MW). The ISO will be able to dispatch down expensive generation in region D and dispatch up cheap generation in region C, while load charges remain unchanged. This produces a $6,000 ISO revenue surplus in real-time as shown in Table II-4.

Table II-4
REAL-TIME SETTLEMENTS
LOW LOAD IN ZONE D

	Gen Offer Price $/MWh (A)	LMP Price $/MWh (B)	Load MW (C)	Load Cost $ (D)	Gen Dispatch MW (E)	Gen RT-DAM MW (F)	RT Gen Revenues $ (G)
A	40-45	45	1,000	45,000	750	0	0
B	35-42	42	2,500	105,000	1,500	0	0
C	35-40	40	3,150	126,000	4,650	150	6,000
D	60-80	80	850	68,000	600	-150	-12,000
Total			7,500	344,000	7,500		-6,000
LAP Price				45.86667			
RT Load Imbalance		45.86667				0	0
RT Revenue Shortfall							-6,000
Combined Market Shortfall							-6,750

Conversely, Table II-5 shows the real-time settlements in the event that that the ISO's DAM load forecast for region D was 150MW too low and its forecast for region C 150 MW too high. In this case the ISO will need to dispatch up expensive generation in region D in real-time while dispatching down cheap generation in region C. DAM and real-time aggregate load are again equal so there are no net load settlements, and the ISO incurs a $6,000 revenue shortfall in real-time.

Table II-5
REAL-TIME SETTLEMENTS
HIGH LOAD IN ZONE D

	Gen Offer Price $/MWh (A)	LMP Price $/MWh (B)	Load MW (C)	Load Cost $ (D)	Gen Dispatch MW (E)	Gen RT-DAM MW (F)	RT Gen Revenues $ (G)
A	40-45	45	1,000	45,000	750	0	0
B	35-42	42	2,500	105,000	1,500	0	0
C	35-40	40	2,850	114,000	4,350	-150	-6,000
D	60-80	80	1,150	92,000	900	150	12,000
Total			7,500	356,000	7,500		6,000
LAP Price				47.46667			
RT Load Imbalance		47.46667				0	0
RT Revenue Shortfall							6,000
Combined Market Shortfall							5,250

If these kind of over and under load forecasts averaged out and over time and were unpredictable, they would not give rise to ISO revenue inadequacy as the gains and losses would roughly offset as suggested by the example.

This happy outcome may not be sustainable, however, if the errors in the ISOs load weights are predictable to market participants. If the errors can be predicted, LSEs will underbid in circumstances in which they believe the ISO is likely to over forecast load in the high cost regions and they will overbid load in circumstances in which they believe the ISO is likely to under forecast load in the high cost regions. This arbitrage would be profitable for market participants but give rise to ISO revenue inadequacy. The likely magnitude of the revenue inadequacy will be far less, however, than that associated with the LAP clearing and settlement system that is currently proposed.

To illustrate this potential, we consider first the circumstance in which LSEs believe that the ISO's load forecast for the high cost region zone D is likely to be too high. In this circumstance it will be profitable for the LSEs to reduce their purchases in the DAM and purchase more power in real-time (because the DAM LAP price will be higher than the real-time LAP price). Table II-6 illustrates a case in which LSEs reduce their DAM purchases to 7,000 MW, compared to their expected load of 7,500 MW.

Table II-6
BASE CASE DAM SETTLEMENTS
UNDER BIDDING BY LSEs TO ARBITRAGE LOAD WEIGHTS

	Gen Offer Price $/MWh (A)	Load Weights (B)	Bid Load MW (C)	Dispatched Generation MW (D)	Cleared Load (E) MW	LMP Price $/MWh (F)	Load Cost $ (G)	Gen Revenues $ (H)
A	40-45	0.133333	933.3333	683.3333	933.3333	45	42,000	30,750
B	35-42	0.333333	2,333.333	1,333.333	2,333.333	40	93,333.33	53,333.33
C	35-40	0.4	2,800	4,300	2,800	40	112,000	172,000
D	60-80	0.133333	933.3333	683.3333	933.3333	80	74,666.67	54,666.667
Total		1.0	7,000	7,000	7,000		322,000	310,750
LAP Price							46	
Congestion Rent							11,250	

Because there is no longer congestion between B and C, DAM congestion rent collections decline and the surplus in the DAM congestion account falls to zero.

Table II-7
CRR SETTLEMENTS

CRRs	Source	Sink	Sink Price $/MWh	Source Price $/MWh	CRR Value $/CRR	CRR Payments $
1,875	C	LAP	46	40	6	11,250
Total						11,250
Net DAM Congestion Settlements:						
Congestion Rents:						11,250
CRR Payments:						-11,250
Net Surplus:						0

In real-time, load is 7,500 MW so the ISO has to dispatch 500 MW more generation than scheduled in the DAM which is sold at the real-time price, $45.87. Because the ISO dispatches down expensive generation in region D and replaces it with cheaper generation elsewhere, the ISO has a surplus in the real-time dispatch. The surplus has fallen from $6,000 in Table II-4 to $5,600 in Table II-8 because of the reduced DAM purchases by LSEs. The reason for the drop is that DAM purchases are 500MW less than in the base case, and the ISO sold this power for $45.87 in real-time, but the cost of buying this power is $46.67, a loss of $.80/MW on each of the 500 MW. Overall, however, the ISO is still not only revenue adequate, but shows a real-time surplus.

Table II-8
REAL-TIME SETTLEMENTS – UNDERBID DAM LOAD – LOW LOAD AT D

	Gen Offer Price $/MWh (A)	LMP Price $/MWh (B)	Load MW (C)	Load Cost $ (D)	Gen Dispatch MW (E)	Gen RT-DAM MW (F)	RT Gen Revenues $ (G)
A	40-45	45	1,000	45,000	750	66.66667	3,000
B	35-42	42	2,500	105,000	1,500	166.6667	7,000
C	35-40	40	3,150	126,000	4,650	350	14,000
D	60-80	80	850	68,000	600	-83.3333	-6666.67
Total			7,500	344,000	7,500	500	17,333.33
LAP Price				45.86667			
RT Load Imbalance		45.86667				-500	-22,933.33
RT Revenue Shortfall							-5,600
Combined Market Shortfall							-5,600

Now, however, let us consider the converse situation in which LSEs anticipate that the ISO's DAM load weights in region D are too low and they bid additional load into the DAM. In this circumstance it will be profitable for the LSEs to increase their purchases of power in the DAM and sell excess power in real-time (because the DAM price will be lower than the real-time price). This is illustrated in Table II-9 in which LSEs schedule 8,000 MW in the DAM, to cover 7,500 MW of expected real-time load.

Table II-9
DAM SETTLEMENTS
OVERBIDDING BY LSEs TO ARBITRAGE LOAD WEIGHTS

	Gen Offer Price $/MWh (A)	Load Weights (B)	Bid Load MW (C)	Dispatched Generation MW (D)	Cleared Load MW (E)	LMP Price $/MWh (F)	Load Cost $ (G)	Gen Revenues $ (H)
A	40-45	0.133333	1,066.667	816.6667	1,066.667	45	48,000	36,750
B	35-42	0.333333	2,666.667	1,666.667	2,666.667	42	112,000	70,000
C	35-40	0.4	3,200	4,700	3,200	40	128,000	188,000
D	60-80	0.133333	1,066.667	816.6667	1,066.667	80	85,333.33	65,333.333
Total		1.0	8,000	8,000	8,000		373,333.3	360,083.33
LAP Price							46.66667	
Congestion Rent							13,250	

Because the system remains constrained and the same units remain on the margin in the example, DAM congestion rent collections and CRR prices are the same as in the base case and the ISO has a $750 surplus in the DAM congestion account.

Table II-10
CRR SETTLEMENTS

CRRs	Source	Sink	Sink Price $/MWh	Source Price $/MWh	CRR Value $/CRR	CRR Payments $
1,875	C	LAP	46.66667	40	6.666667	12,500
Total						12,500
Net DAM Congestion Settlements:						
Congestion Rents:						13,250
CRR Payments:						-12,500
Net Surplus:						750

In real-time, load is 7,500 so the ISO has to dispatch 500 MW less generation than scheduled in the DAM and the overscheduled LSEs sell back the power they purchase at $46.67/MWh for $47.47. Because the ISO dispatches up expensive generation in region D and dispatches down cheaper generation elsewhere, the ISO has a deficit of $6,400 in the real-time dispatch. The deficit has risen from $6,000 in the no arbitrage case (Table II-5) to $6,400 (Table II-11) because of the increased DAM purchases by LSEs which are sold back at a real-time LAP price that exceeds the price in the DAM, realizing an arbitrage profit of $.80/MWh.

Table II-11
REAL-TIME SETTLEMENTS – OVERBID DAM LOAD – HIGH LOAD AT D

	Gen Offer Price $/MWh (A)	LMP Price $/MWh (B)	Load MW (C)	Load Cost $ (D)	Gen Dispatch MW (E)	Gen RT-DAM MW (F)	RT Gen Revenues $ (G)
A	40-45	45	1,000	45,000	750	-66.667	-3,000
B	35-42	42	2,500	105,000	1,500	-166.667	-7,000
C	35-40	40	2,850	114,000	4,350	-350	-14,000
D	60-80	80	1,150	92,000	900	83.33333	6666.667
Total			7,500	356,000	7,500	-500	-17,333.33
LAP Price				47.46667			
RT Load Imbalance		47.46667				500	23,733.33
RT Revenue Shortfall							6,400
Combined Market Shortfall							5,650

Misforecasts by the ISO are inevitable. The important issue is not the misforecasts, but whether they are predictable to market participants. The surpluses and deficits in the no-arbitrage case (Tables II-3 and II-4) offset each other (+$6,000 - $6,000 = 0), but that is not the case with arbitrage (Tables II-8 and II-11), ($5,600 - $6,400 = -$800)).

Appendix III
Nodal Clearing of Zonal Virtual Load Bids

This appendix illustrates the potential unintended consequences of applying the proposed nodal clearing and settlement process for zonal/LAP load bids in the day-ahead market to virtual load and supply bids.

The illustration is based on the same simple transmission system and generation resources used in Appendix I.

As shown in Appendix I, if the LSE's serving load in the LAP bid the entire expected real-time load, 7,500 MW into the DAM, at slightly above the expected real-time zonal/LAP price under the proposed MRTU nodal clearing process the bid load would be assigned to the four regions based on the ISO-determined load weights portrayed in column (B). The bid load allocated in this manner would then be cleared nodally. All of the bid load allocated to regions A, B and C would clear, but only 250 MW of the load allocated to region D would clear, as load levels in excess of 250 MW would require dispatch of generation at prices of $60/MWh or higher and thus price-capped load bid in at $50/MWh would not clear in region D above 250 MW as shown in columns (D) and (E) of Table III-1.

Table III-1
BASE CASE DAM SETTLEMENTS
7,500 MW Bid in at $50

	Gen Offer Price $/MWh	Load Weights (B)	Bid Load MW (C)	Dispatched Generation MW (D)	Cleared Load MW (E)	LMP Price $/MWh (F)	Load Cost $ (G)	Gen Revenues $ (H)
A	40-45	0.133333	1,000	750	1,000	45	45,000	33,750
B	35-42	0.333333	2,500	1,500	2,500	42	105,000	63,000
C	35-40	0.4	3,000	4,500	3,000	40	120,000	180,000
D	60-80	0.133333	1,000	0	250	42	10,500	0
Total		1.0	7,500	6,750	6,750		280,500	276,750
LAP Price							41.55556	
Congestion Rent							3,750	
(G) = (F) * (E) (H) = (F) * (D)								

A total load of 6,750 MWh would therefore clear in the DAM at a LAP price of $41.56. Total load payments would exceed generator revenues as shown in columns (G) and (H) of Table III-1, causing the ISO to collect congestion rents of $3,750.

In real-time, generation would be dispatched to meet the entire 7,500 MW of real-time load, and an additional 750 MW of high cost generation not scheduled in the DAM would be dispatched in region D and paid the real-time nodal LMP price, $80/MWh, as shown in columns (E) and (F) of Table III-2. The LAP price would be $46.67, so the ISO would sell an additional 750 MW of power in region D at the real-time LAP price, collecting $35,000 in net real-time revenues, but would pay $60,000 for the dispatch of generation in region D.

Table III-2
REAL-TIME SETTLEMENTS

	Gen Offer Price $/MWh (A)	LMP Price $/MWh (B)	Load MW (C)	Load Cost $ (D)	Gen Dispatch MW (E)	Gen RT-DAM MW (F)	RT Gen Revenues $ (G)
A	40-45	45	1,000	45,000	750	0	0
B	35-42	42	2,500	105,000	1,500	0	0
C	35-40	40	3,000	120,000	4,500	0	0
D	60-80	80	1,000	80,000	750	750	60,000
Total			7,500	350,000	7,500		60,000
LAP Price				46.66667			
RT Load Imbalance		46.667				-750	-35,000
RT Revenue Shortfall							25,000
Combined Market Shortfall							24,166.67

The ISO would therefore incur a real-time revenue shortfall of $25,000, considerably in excess of the $833.33 of excess congestion rents collected in the DAM, leading to a combined shortfall of $24,166.67.

In this appendix, we modify the example developed in Appendix I to illustrate the potential incentives of virtual loads and suppliers and the impact on consumer costs of virtual load and supply bidding under the proposed settlement rules. Table III-3 portrays a situation in which LSEs within the LAP bid 7,500 MW of load into the DAM. We assume that 1,875 MW of LSE load hedged by CRRs is bid in at $250/MWh, and the remaining 5,625 MW is bid in by LSEs at a price cap of $50/MWh as shown in columns (C) and (E) of Table III-3. In addition, virtual loads bid in another 2,000 MW at $41/MWh. The 1,875 MW of LSE load bid in at $250 clears in each zone. The remaining 5,625 MW of LSE load clears in zones A, B, and C as shown in column (F), but not in zone D as the nodal price would need to reach $60/MWh for generation to be dispatched to meet load above 250 MW in Zone D. Finally, the virtual load bid in at $41/MWh clears only in zone C. It is seen in column (I) that the total load cleared in the DAM is 7,550 MW at a zonal price of $41.39/MWh, generating DAM congestion rents of $3,750.

Table III-3 BASE CASE DAM SETTLEMENTS 2,000 MW Virtual Load													
	Gen Offer Price $/MW (A)	Load Weights (B)	LSE Bid Load @ $250/MW MW (C)	Cleared Load MW (D)	LSE Bid Load @ $50/MW MW (E)	Cleared Load MW (F)	Virtual Load @ $41/MW MW (G)	Cleared Load MW (H)	Total Cleared Load MW (I)	Gen Dispatch MW (J)	LMP Price $ (K)	Load Cost $ (L)	Gen Revenues $ (M)
A	40-45	0.133333	250	250	750	750	266.6667	0	1,000	750	45	45,000	33,750
B	35-42	0.333333	625	625	1,875	1,875	666.6667	0	2,500	1,500	42	105,000	63,000
C	35-40	0.4	750	750	2,250	2,250	800	800	3,800	5,300	40	152,000	212,000
D	60-80	0.133333	250	250	750	0	266.6667	0	250	0	42	10,500	0
Total		1.0	1,875	1,875	5,625	4,875	2,000	800	7,550	7,550		312,500	308,750
LAP Price												41.39073	
Congestion Rent												3,750	
Total Bid Load					9,500								

Table III-4 shows the payments to the 1,875 MW of Zone C to LAP CRRs, are $1.39/CRR and total $2,607.62.

Table III-4 CRR SETTLEMENTS						
CRRs	Source	Sink	Sink Price $/MW	Source Price $/MW	CRR Value $/CRR	CRR Payments $
1,875	C	LAP	41.39073	40	1.390728	2,607.616
Total						2,607.616

The net DAM congestion settlements therefore have a surplus of $1,142.38 as shown in Table III-5.

Table III-5 NET CRR SETTLEMENTS	
Congestion Rents	3,750
CRR Payments	-2,607.62
Net Surplus	1,142.384

In real-time the ISO would need to dispatch additional high cost generation in region D, while backing down low cost generation in regions A, B and C as shown in column (F) of Table III-6. In addition, the ISO would buy back 50 MW of DAM load at the real-time price. The ISO would incur an additional $28,000 in generation costs to dispatch generation up in region D and down in the lower cost regions as shown in Table III-6 and then pay $2,333.33 to buy back the excess DAM load. The real-time shortfall would rise to $30,333.33, leaving a combined market shortfall of $29,190.95. The overall market shortfall would include the cost of $312.58 (800 * $.39073) of uplift payments to the virtual loaders whose bids cleared at $41, despite a zonal/LAP price of $41.39.

Table III-6
REAL-TIME SETTLEMENTS
INFLATED LOAD BIDS

Region	Gen Offer Price $/MWh (A)	LMP Price $/MWh (B)	Load MW (C)	Load Cost $ (D)	Gen Dispatch MW (E)	Gen RT-DAM MW (F)	RT Gen Revenues $ (G)
A	40-45	45	1,000	45,000	750	0	0
B	35-42	42	2,500	105,000	1,500	0	0
C	35-40	40	3,000	120,000	4,500	-800	-32,000
D	60-80	80	1,000	80,000	750	750	60,000
Total			7,500	350,000	7,500	-50	28,000
LAP Price				46.66667			
RT Load Imbalance				50 @		46.66667	2,333.333
RT Revenue Shortfall							30,333.33
CRR Settlements							-1,142.38
Uplift on Virtual Bid							312.582
Combined Market Shortfall							29,503.53

Table III-7 shows the profitability of the virtual bids that would buy power at $41 in the DAM and sell it back at $46.67, even though LSEs bid their entire load into the DAM at a price in excess of the real-time LAP price.

Table III-7
VIRTUAL BIDDER PROFITS

DAM Purchases	-800	41	-32,800
Real-Time Sales	800	46.66667	37,333.33
Net Profit			4,533.333

Unlike the examples in Appendix I, the potential for this kind of arbitrage by virtual bidders does depend on the existence of disequilibrium. If virtual and LSE bidders were permitted to submit sufficient load bids in excess of real-time physical load to bring DAM and real-time prices into equilibrium as shown in the example in Section E of Appendix I, the kind of arbitrage portrayed in the example in this appendix would not be feasible.

If one attempted to address the uplift costs arising from the flawed nodal clearing process by restricting LSE load bids to prevent equilibrium, these restrictions could produce the kind of arbitrage portrayed here.

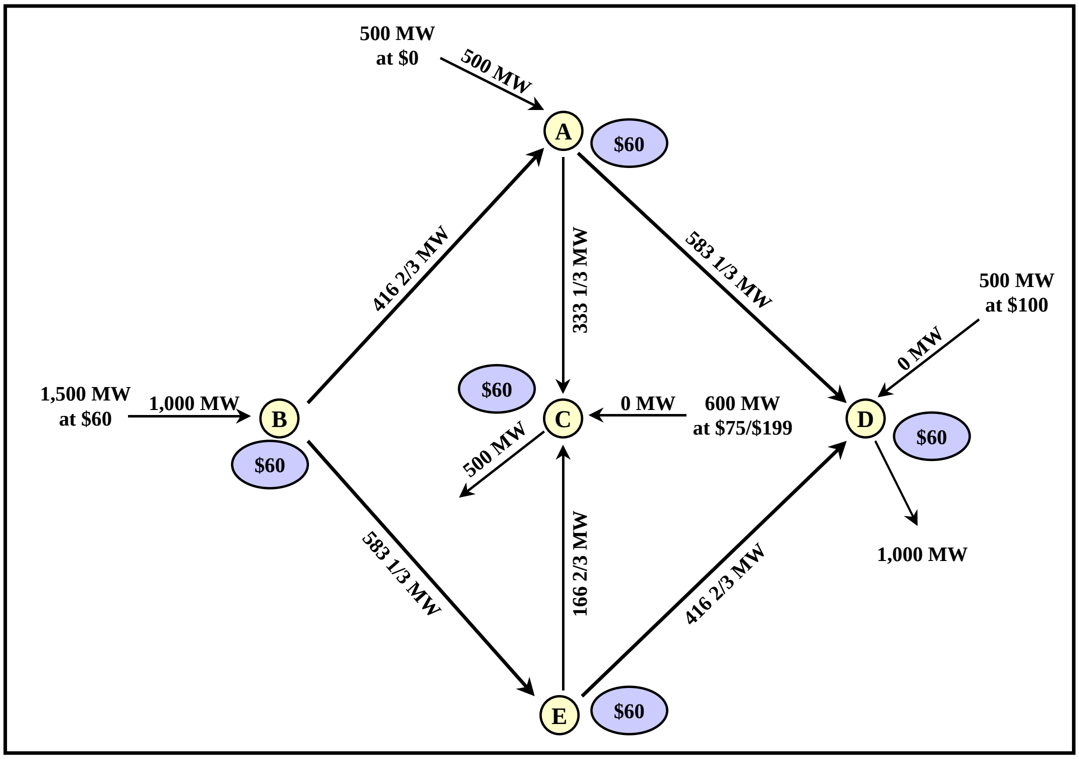
Appendix IV
MRTU Local Market Power Mitigation

This appendix illustrates the potential for the use of extreme decremental offer prices in the local market power dispatch (Pass 2) to undercut the effectiveness of the local market power mitigation mechanism.

Figure IV-1 illustrates the Pass 1 dispatch in which generation would be dispatched without regard to local transmission constraints. For the example, it is assumed that the limits (on lines A-D and A-C) are local transmission constraints. If the system is dispatched without regard to the constraints, all demand can be met at a price of $60/MWh, with 500 MW dispatched at A and 1,000 MW at B.

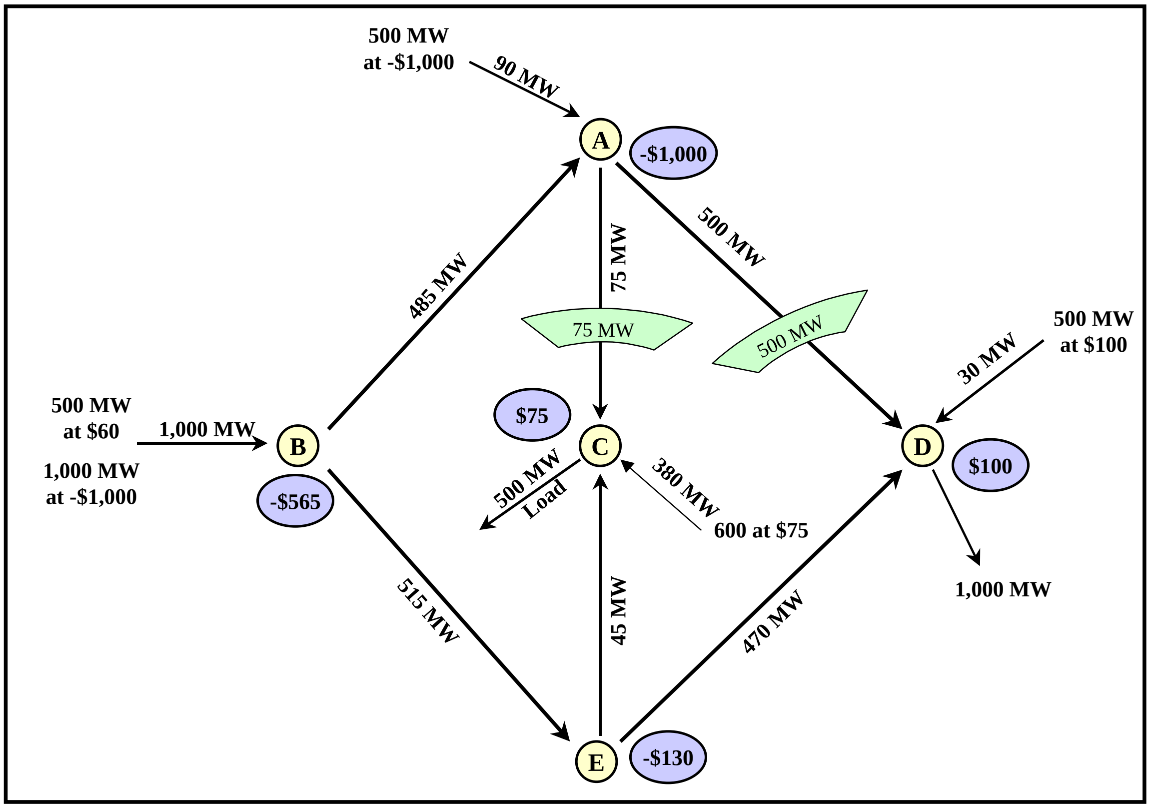
It is assumed that the generation at C is bid into the market at $199/MWh in Pass 1A and mitigated to $75/MWh in Pass 1B. Since the LMP price at C is $60/MWh in Passes 1A and 1B, the generation would not be dispatched in either pass and thus would not be subject to mitigation in this pass (whether or not the generation at C were RMR generation).

Figure IV-1
PASS 1A, 1B DISPATCH



Now suppose that the MRTU-local market power mitigation methodology were applied in Pass 2 by reducing the offer prices of the generation dispatched in Pass 1 to -$1,000/MWh, while enforcing the local transmission constraints. 30 MW of the generation at D would be dispatched at the mitigated RMR price of $100/MWh, while the generation at C would be dispatched for 380 MW at its $75/MWh mitigated RMR offer price, as shown in Figure IV-2. It is important to note that the prices at A, B and E are negative and the price at B is far below the actual offer price of the generation at that location.

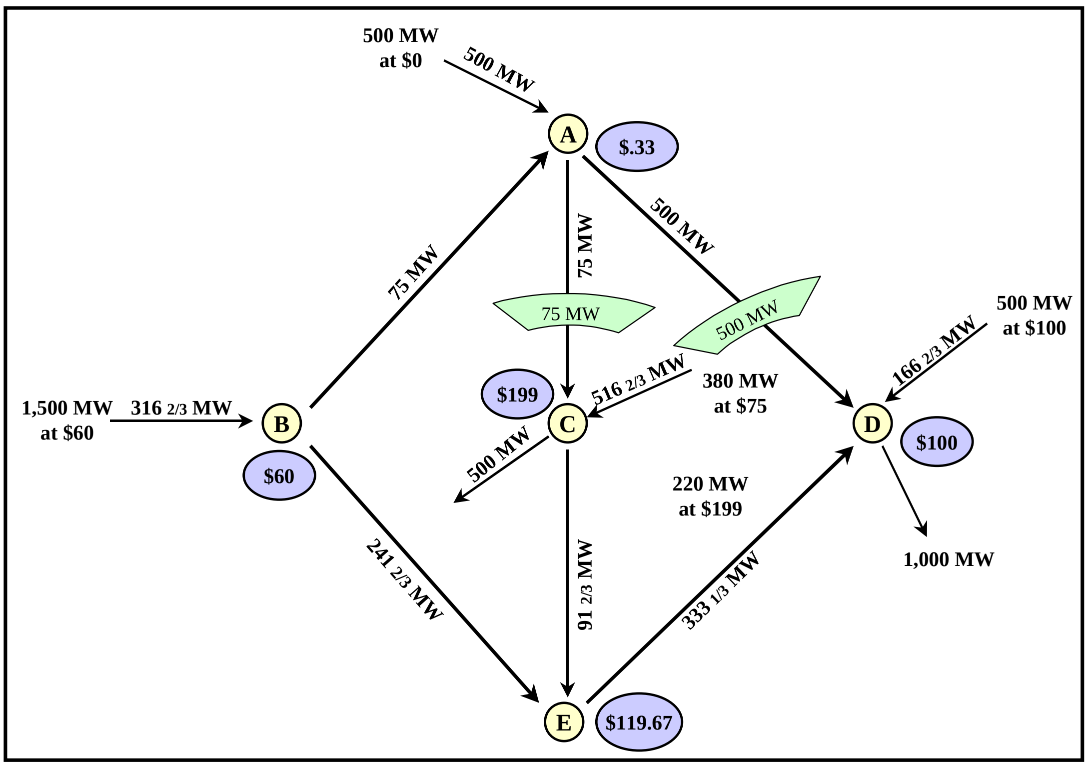
Figure IV-2
PASS 2 DISPATCH



In Pass 3, generation at A and B is offered at the same prices as in Pass 1, the generation at D offers 500 MW at the $100/MWh RMR price, 380 MW of generation at C is offered at the mitigated price of $75/MWh, and 220 MW at $199/MWh. It can be seen in Figure IV-3 that when the system is dispatched based on the actual bids of suppliers at A and B the LMP price at C is no longer $75/MWh as in Pass 2, but rises to $199/MWh. Because the local market power mitigation was applied in Pass 2 based on the extreme DEC bids, it was ineffective in preventing the exercise of market power by the RMR generator at C.

Similarly, had there been RMR generation at E, it would not have been dispatched in the local market power mitigation pass (since the price at that location was -$130/MWh), yet the price at E in Pass 3 is $119.67/MWh.

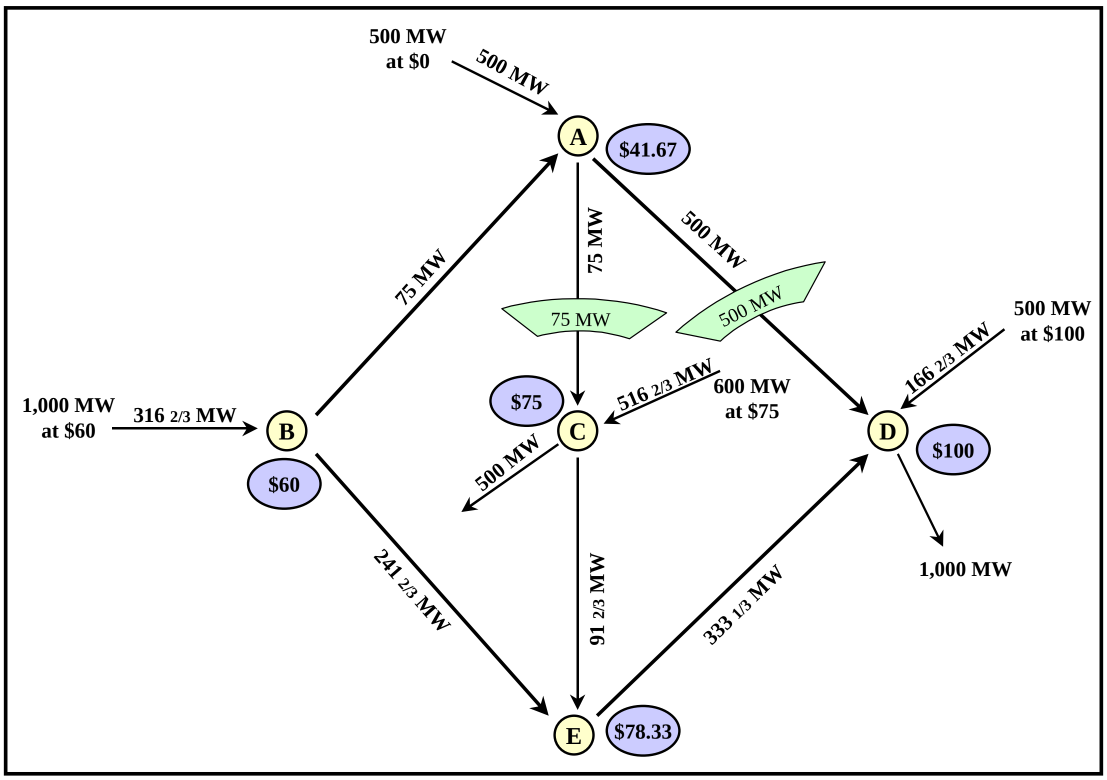
Figure IV-3
PASS 3 DISPATCH
MRTU MITIGATION



The generator at C is able to exercise market power despite the presence of market power mitigation because the Pass 2 dispatch did not use the actual offer prices of the suppliers at A and B but instead used DEC prices of -$1,000/MWh as provided for in the MRTU

If the local market power mitigation in Pass 2 were based on the actual offer prices of the generation at A and B, then the generation at C would have been dispatched at its mitigated offer price and LMP prices at D would have been $75/MWh, as shown in Figure IV-4, rather than $199/MWh. In addition, if Pass 2 were based on actual offer prices generation at E with offer prices less than $78.33/MWh would have been mitigated and dispatched.

Figure IV-4
PASS 2 AND 3 DISPATCH
USING ACTUAL BIDS



The total cost borne by load can be calculated either as total payments by load at the LMP price less congestion rents (which directly or indirectly flow to load) or as total payments to generation (which is payments by load less congestion rents). Under the MRTU market power mitigation approach, as portrayed in Table IV-5, the total cost to load would be $138,650 compared to only $95,251.67 under a standard market power mitigation approach.

Table IV-5

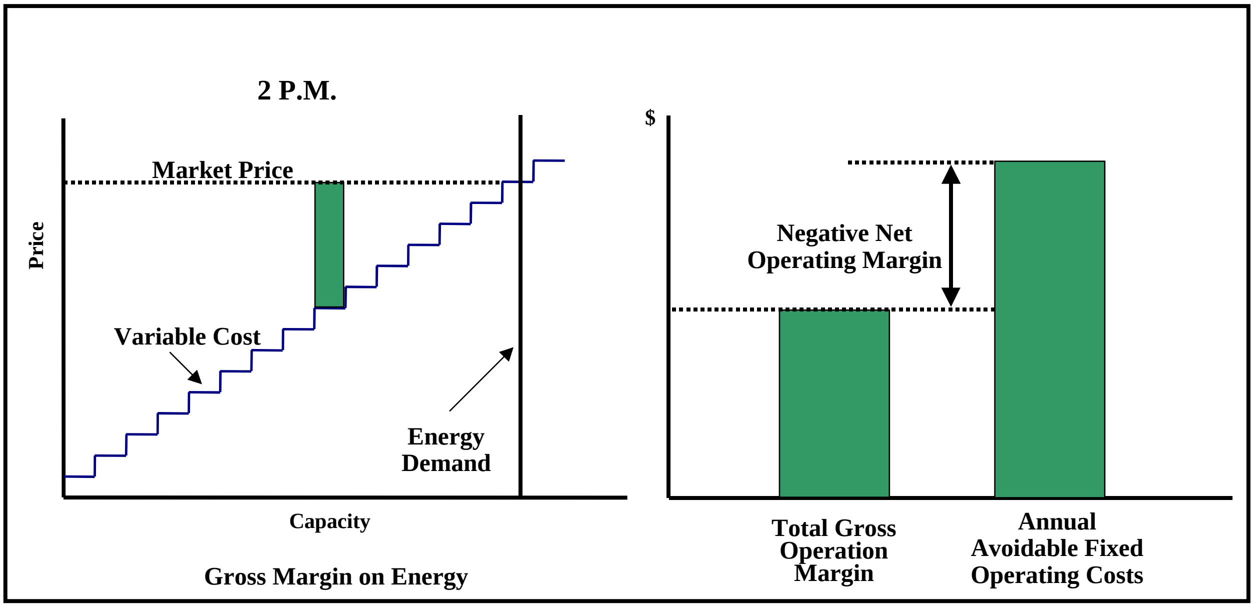
	MRTU Mitigation			**Standard Mitigation**		
	Generator Output	**LMP Price**	**Generator Revenues**	**Generator Output**	**LMP Price**	**Generator Revenues**
A	500	0.33333	166.665	500	41.67	20835
B	316.66666	60	19000	316.66667	60	19000
C	516.66666	199	102816.67	516.66667	75	38750
D	166.66666	100	16666.666	166.66667	100	16666.667
Total	**1500**		**138650**	**1500**		**95251.667**

Appendix V
Resource Adequacy Systems

A. ICAP Market Systems

Price equal to the short-run marginal cost of the marginal supplier is a basic short-run equilibrium condition. With the introduction of market-based marginal cost pricing in energy markets, infra-marginal generation earns revenues on sales of energy and ancillary services, earning margins equal to the difference between its revenues and the variable costs incurred in generating energy, as portrayed in Figure V-1. Generation also incurs fixed costs, some of which can be avoided if the generation owner chooses not to make its capacity available for operation (i.e., if the capacity is either mothballed or closed permanently). In the absence of an ICAP or installed reserve requirement, generation owners will not choose to keep capacity in operation for dispatch unless their gross operating margin exceeds their avoidable fixed operating costs.

Figure V-1
GENERATOR OPERATING MARGINS



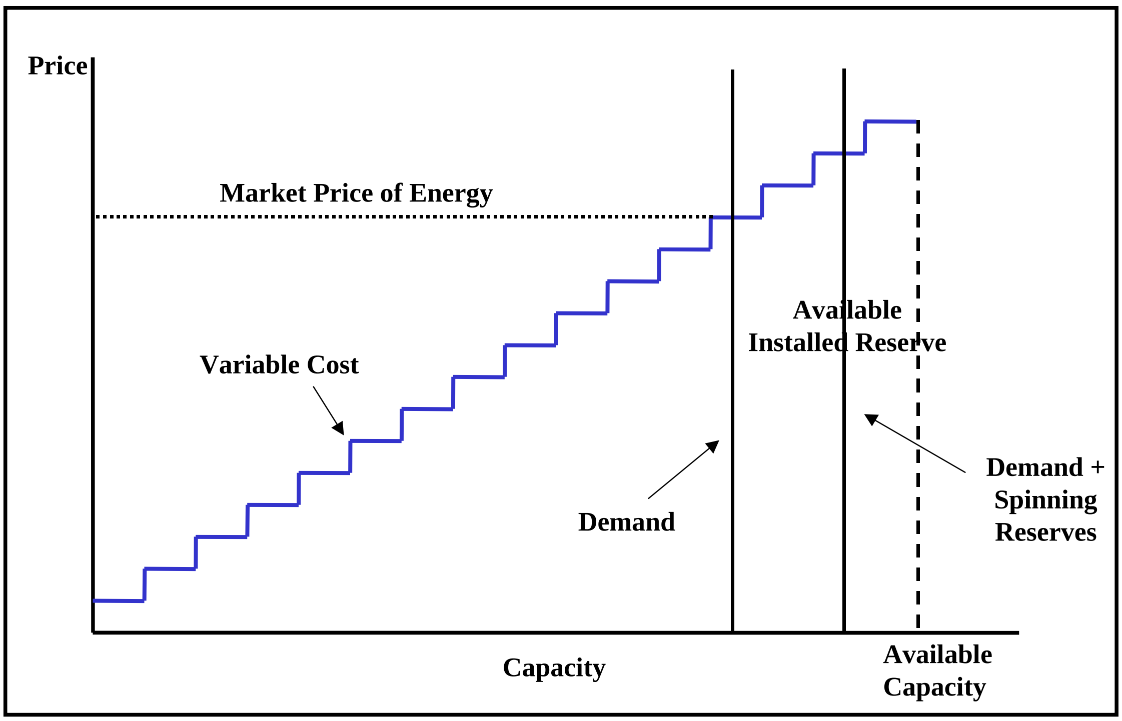
Under an ICAP system, a capacity requirement is established such that there is almost always sufficient capacity to avoid involuntary load shedding.[1] The wholesale energy market therefore usually clears at the intersection of demand and the variable cost (dispatch) curve, as portrayed in Figure V-2. If the energy market design does not include demand bids or some other form of demand curve to represent scarcity pricing, the price of energy would be generally set by the incremental cost of the energy generated by marginal units. The price may not be high

[1] The number of hours of reserve shortage (i.e., emergency state operation) is relatively low but the frequency of reserve shortage exceeds the one-day-in-ten-year load shedding standard. Only severe reserve shortages result in load shedding.

enough often enough to cover the full cost of keeping these marginal units in operation over the year (i.e., the units will have a negative net operating margin as portrayed in Figure V-1), unless the energy price is extremely high during hours of reserve shortage.

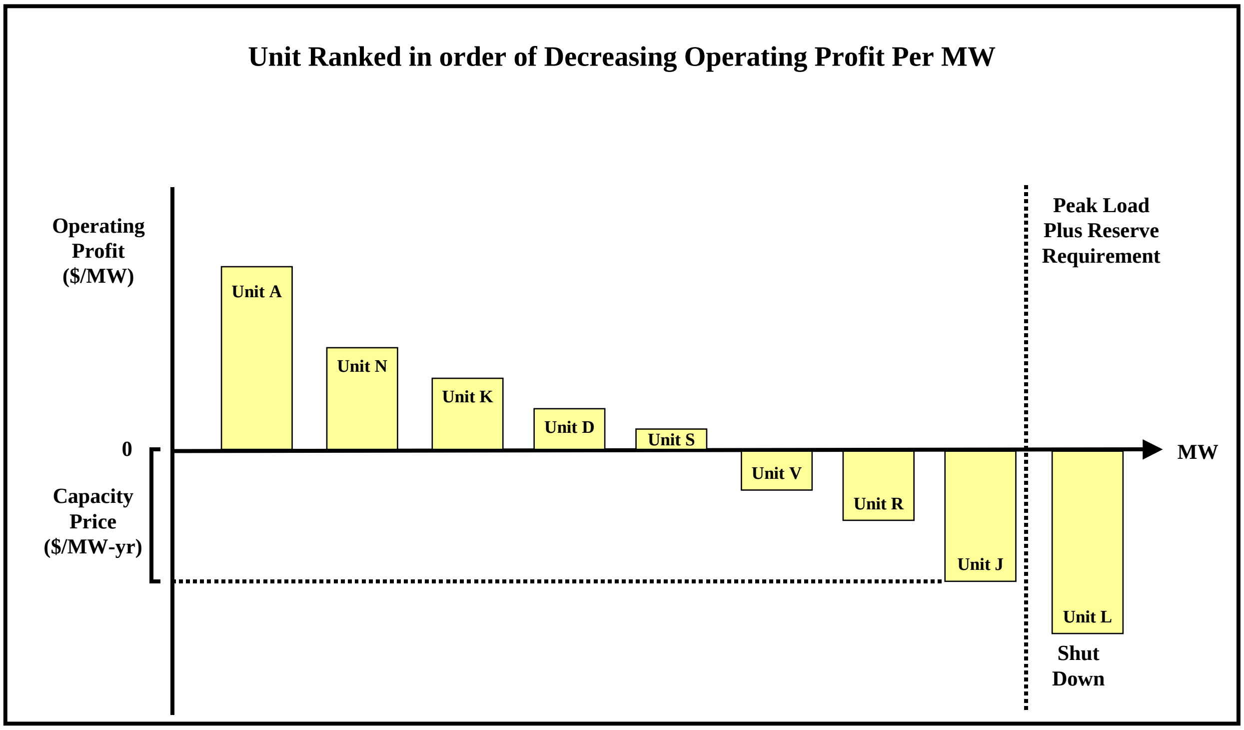
Under an ICAP system, the negative net operating margin is addressed by imposing a market-wide ICAP requirement symmetrically on all load-serving entities within the market. LSEs may not have the traditional obligation to serve, but under an ICAP system they must demonstrate reserves sufficient to meet the installed capacity requirement for their customers. If the amount of generation required to be available under the installed capacity requirement exceeds the amount of generation that would have been available in the absence of such a requirement (i.e., the amount justified by energy and reserve market revenues alone), a market for capacity is created. Thus, with such a binding ICAP requirement, capacity takes on value in and of itself. Marginal units, unprofitable on the margins they earn on energy sales and ancillary services, would demand a capacity payment in return for agreeing to make themselves available for operation and allowing the contracting LSE to satisfy its ICAP requirement.

Figure V-2
ENERGY MARKET PRICES WITH INSTALLED CAPACITY REQUIREMENTS



To keep capacity open under an ICAP system, the owner of the marginal unit requires a capacity payment of at least the difference between its avoidable fixed operating costs and its net margin on energy and ancillary services sales (i.e., it must recover the negative net operating margin portrayed in Figure V-1 on an expected value basis). Competition among capacity owners and with entrants should cause the market-clearing capacity payment to approximate the per-MW payment that would induce just enough generation to remain available to enable the ICAP requirement to be met. Under a market-based ICAP system, all generating capacity contracting to provide installed reserves are paid the market-clearing price of capacity, as portrayed in Figure V-3. Between the capacity payments they receive and their margins on energy sales and ancillary services, all units providing the capacity needed to meet the ICAP requirements would earn enough to cover their avoidable fixed operating costs and thus would remain available.

Figure V-3
DETERMINATION OF MARKET PRICE OF CAPACITY



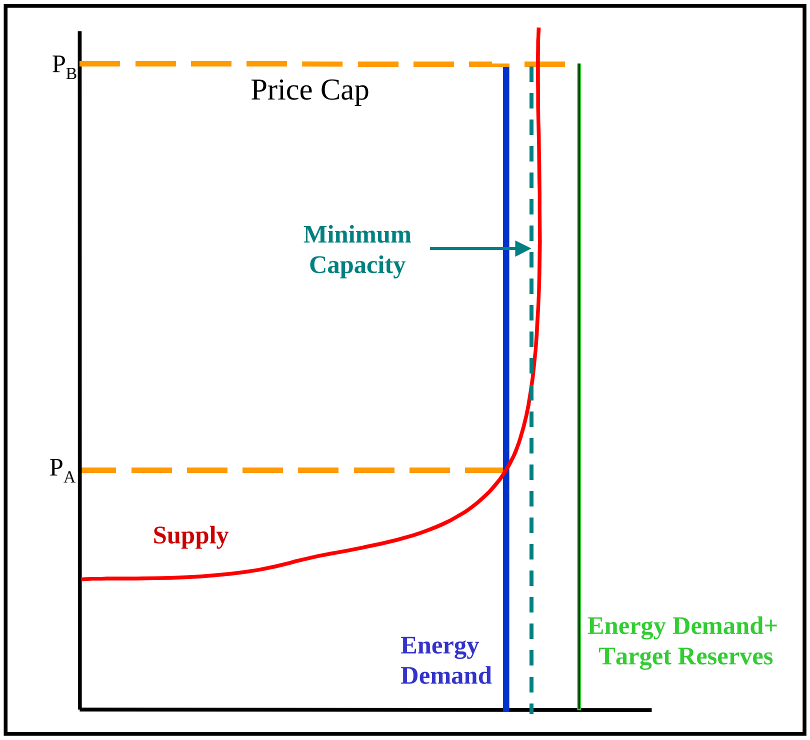
With an ICAP requirement, the capacity payment is determined by the per-MW payment required to enable the marginal unit (Unit J in Figure V-3) to at least break even versus the payment required to keep the next most expensive unit in operation (Unit L). Because the market can meet the ICAP reserve requirement without Unit L, the market-clearing capacity payment would be insufficient for it to cover its anticipated operating losses and Unit L will be closed. Between the capacity payments they receive and their margins on energy and ancillary services sales, each of the other units remaining open would make more than enough to cover their avoidable fixed operating costs.

B. Energy-Only Pricing

An alternative to an ICAP system in maintaining resource adequacy is to structure energy and ancillary service markets such that the marginal generator is able to recover its going-forward costs in energy and ancillary service prices. In principle, the changes needed to implement such an energy market-based resource adequacy system are to implement shortage pricing that causes the prices of energy and ancillary services to rise to a sufficiently high level during shortage conditions that the marginal capacity supplier required to meet established reliability criteria is able to recover its going-forward costs during these shortage hours.

For vertically integrated utilities such an energy only market could operate much like an ICAP system. The system operator/ISO could determine the shortage prices that it estimates would be required to keep sufficient capacity available to meet the target level of reliability and could even inform the vertically integrated utilities of the implied reserve margin. The shortage pricing would support the implied reserve margin as the pricing system would assure that there would be enough hours with high prices to justify keeping the target level of capacity in operation. Nevertheless, there are some risks in this market design and these risks are magnified in markets with unintegrated retailers and suppliers.

Figure V-4
SUPPLY AND DEMAND IN A SHORTAGE



The fundamental problem with this market design is that with vertical demand, the market price of energy and reserves only exceed the incremental costs of the marginal generator when the control area is reserve short. As long as these reserve shortages are small, they will have little impact on reliability, so reliability would be consistent with a shortage pricing system. Thus, as portrayed in Figure V-4, the price would rise to the price cap when reserves fell below the target level, but involuntary load shedding would occur only when capacity fell below the minimum capacity level. The problem is that actual peak load is uncertain, as is the available capacity (due to random outages). Thus, the more often the system is expected to be in a reserve short condition, the greater the potential for bad luck in terms of weather or outages to throw the system into the range in which involuntary load shedding is required.

From this perspective, there are three basic risks in relying on an energy-only pricing system defined in this manner:

- Miscalculation of the cost of capacity by the pool.
- Miscalculation of expected prices by LSEs/suppliers.
- Required shortage frequency.

Each of these risks is discussed below.

1. Miscalculation of Capacity Costs

Under an ICAP system the system operator determines the reserve margin and ICAP requirement through Monte Carlo type analysis of reliability under stressed system conditions. Importantly, the calculated reserve margin does not depend on the cost of having capacity available during stressed system conditions. Instead, the system operator determines the physical capacity requirements and the cost of keeping this capacity in operation is, in principle, determined in the market by the supply decisions of resource providers.

Under an energy pricing system driven by shortage pricing, however, the amount of capacity that will be made available by resource suppliers in response to any set of shortage prices depends on the cost of having this capacity available during those shortage conditions. If the system operator or reliability authority misunderstands the cost of having capacity available or miscalculates the revenues generated by marginal capacity during non-stressed conditions, then a given set of shortage prices may result in more or less capacity being available than expected by the system operator, resulting in a different level of reliability than planned for by the system operator. Since the system operator does not participate in commercial markets there is a potential under this approach to energy-only pricing for the system operator to significantly misassess the cost of having generating capacity available during peak conditions. If there is a strong link between the shortage costs used by the system operator and the actual reliability value of capacity during those conditions, this error would be less important in terms of its impact on consumer welfare as the prices would reflect the value of the capacity. Absent such a strong link, there is a potential for misestimation of capacity costs by the system operator to lead to a material difference between the actual and intended level of reliability.

2 Miscalculation of Expected Prices

While suppliers have a good sense of the overall cost of keeping their capacity available, both suppliers and LSEs may have difficulty projecting expected annual net revenues based on the system operators shortage cost rules. The expected price level would depend on both generation and transmission outage probabilities as well as the supply of imports. If suppliers and LSEs have different expectations than the system operator about the frequency and degree of shortage conditions, then they may not provide the anticipated level of capacity in response to a given set of shortage prices, even if the system operator accurately assesses the cost of providing this capacity. While the system operator could make public its profile of simulated shortage prices, it is not clear that market participants would necessarily find it commercially reasonable to rely on these results. A further potential source of divergent expectations is the assessment by resource suppliers of the likelihood that regulators will permit high market prices, even during shortage conditions. Thus, if the nominal price cap were $10,000/MW, the system operators' assessment might be that the marginal supplier would recover its entire net operating revenue shortfall of $50,000/MW during eight hours of shortage conditions in which the price of power was projected to exceed $5,000/MW. If the resource owner does not believe that it would be permitted to earn more than $1,000/MWh during these market conditions, the revenue assessment of the system operator and market participant would be radically different and much less capacity might be forthcoming than assumed by the system operator. In principle market participants should over time be able to assess the accuracy of the system operator price forecast, as well as the price level regulators would allow, but the reality is that the forecast by resource suppliers and the system operator will be based on expected conditions. Even an accurate forecast may only average out to reflect actual prices over a period of a number of years, and conditions may be changing more rapidly than the actual outcomes converge on the forecast. It may therefore be difficult or impossible for market participants to distinguish whether price estimates are biased ex ante or are accurate estimates of volatile conditions and prices.

3. Required Shortage Frequency

Under an energy-only pricing system, there is a very explicit tradeoff between the expected price level during shortages and the number of shortage hours required to recover a given net operating cost shortfall. The greater the number of hours of reserve shortage, however, the greater the likelihood that the shortage in some hours will be sufficiently severe as to require involuntary load shedding. Thus, the lower the price cap in the energy market, the larger the number of shortage hours required to recover a given operating cost shortfall, and the more likely that load shedding will be necessary during some of the shortage hours.

As suggested in the illustration above, with a price cap of $10,000/MWh and effective shortage pricing, a small number of shortage hours would be sufficient for the marginal generator with an incremental running costs of $100/MWh to recover $50,000/MW in going-forward costs. Given this small number of shortage hours, the probability distribution of demand and supply surprises might yield a one-day-in-ten years probability of such a large capacity shortage that load shedding was required.

Suppose, on the other hand, that the price cap were $1,000/MWh. The most that the marginal generator could recover during a shortage hour would be $900/MWh. The number of shortage hours required for the marginal generator to recover its going forward costs on an expected basis would be around 55 hours per year. A capacity balance tight enough to produce 55 hours per year of shortage conditions, however, would likely have a much greater risk of requiring load shedding than if only 8 hours were expected, and the increase in likelihood might be non-linear.

At the extreme, suppose the price cap was set at $250/MW as proposed under the MRTU. In this circumstance around 333 hours of reserve shortage would need to be expected on an average annual basis for such a marginal generator to recover its going forward costs in energy prices alone. Such a high frequency of reserve shortages would in turn produce a very high probability of involuntary load shedding during some hours of particularly large shortages.

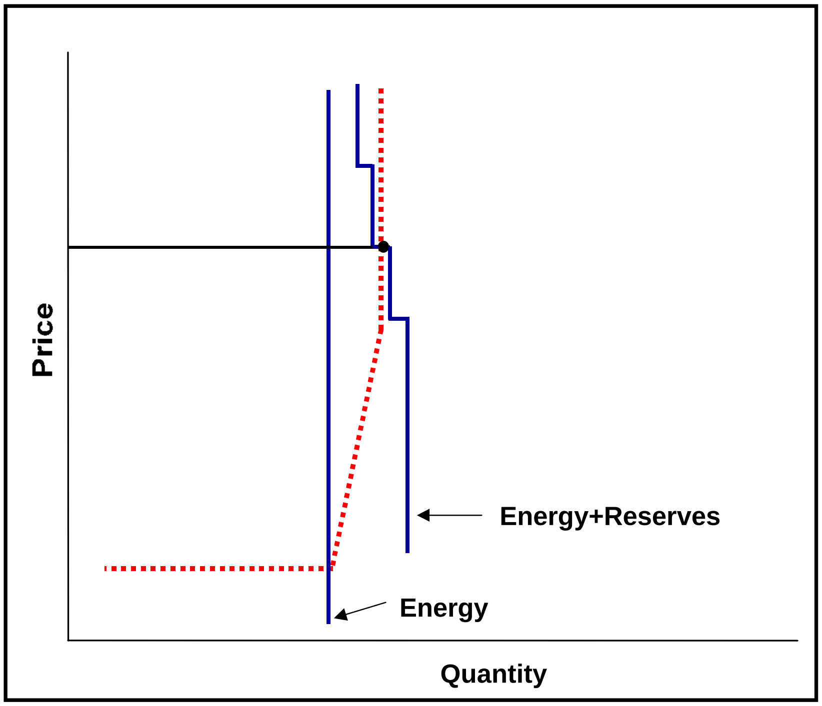
An energy-only pricing system is therefore workable only if prices are very high during shortage conditions. This is problematic from at least two perspectives. First, if even small reserve shortages result in prices of $5,000/MWh or more, there would potentially be a greatly increased incentive for energy suppliers to physically withhold capacity in order to produce a reserve shortage and drive prices to extremely high levels. Second, the variability of supply and demand conditions will make it likely that resource suppliers will not recover their going-forward costs evenly year to year, but rather the recovery will be concentrated in particular years.[2] This price pattern could sustain multi-year energy contracts that would recover generator going-forward costs. Under retail access systems, however, there may not be many multi-year energy contracts and thus most customers would be exposed to high energy prices during the one year in five or six in which suppliers recover their going forward costs. This implies large variations in retail prices that may not be realizable within the regulatory structure.

Two elements of the NYISO market design address these limitations of a energy-only market system. First, the explicit reserve markets of the NYISO provide an additional relatively stable income stream for the marginal ICAP resource, which should be a 10-minute combustion turbine located east of Central East. Nevertheless, the expected reserve market earnings of around $10,000/MWyear fall far short of what is required to keep the marginal unit in operation.

[2] Thus, a marginal generator with a going forward cost of $50,000/MW year might anticipate recovering $15,000/MW year in most years but recovering $200,000/MW year every five years or so.

Second, the reserve demand curve implemented by the NYISO prior to the summer of 2002 addresses the market power problem by in effect making the residual demand curve facing a supplier with market power more price elastic than would otherwise be the case as shown in Figure V-5. In addition, the demand curve somewhat raised energy prices.

Figure V-5
NYISO RESERVE DEMAND CURVE



Nevertheless, the marginal generator east of Central East will not be able to recover its going-forward costs in NYISO energy and reserve markets unless the price cap is materially raised.

4. Reliability Consequences

In a market with a substantial amount of price sensitive load these sources of error in assessing the frequency of shortages would not be of great importance from a reliability standpoint as the errors would result in variations in market prices but firm load would be met. Similarly, if there is adequate price-responsive load, even frequent reserve shortages need not lead to involuntary load shedding. If there is little or no price sensitive load, however, then both these kinds of errors or a price cap would translate into differences in the frequency of reserve shortage conditions and inefficiently high probabilities of involuntary load shedding.

The potential for all of these kinds of miscalculation and the recognition that there would be little or no price sensitive load in the short run, as well as a reluctance to allow extremely high energy prices, lay behind part of the initial reluctance to rely on energy only pricing to maintain reliability for the initial implementation of LMP markets in New York and PJM.

C. Energy-Only Pricing with Price-Responsive Real-Time Load

The reliability risks associated with energy only pricing are likely to be greatly reduced or even eliminated if the real-time demand for power is price responsive. In an energy-only pricing system with price-responsive real-time load, market clearance and reliability would be ensured by price-responsive real-time customer demand, without the need for administratively determined installed reserve requirements and without undue reliability risk.

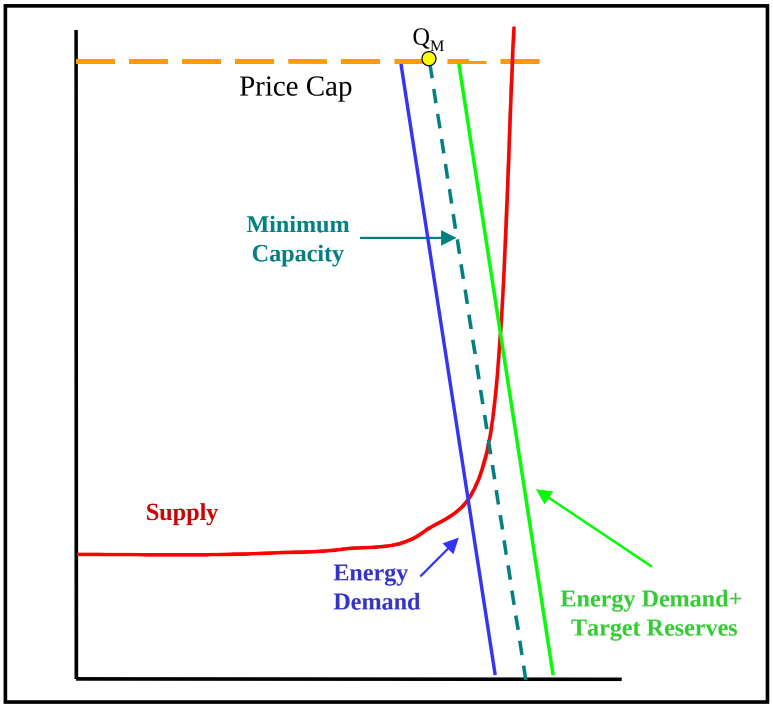
- Operating reserve margins would be maintained with price-responsive load reducing consumption in response to high prices.

- There would be no administrative reserve requirement or capacity payment. Long-term capacity decisions would be left to market incentives.

- The electricity market would clear while providing reliability, through operating reserve standards, energy pricing and market-determined installed reserve levels.

The crux of such a system is that instead of shocks such as unexpected weather or bad luck in terms of generation outages translating into reduced operating reserves and higher load shedding risk, these shocks translate into higher prices and voluntary load reductions. Thus, the demand curve portrayed in Figure V-6 could shift out considerably, or outages considerably shift in the supply curve without the minimum level of energy demand plus minimum reserves (Qm) exceeding the available capacity.

Figure V-6
ENERGY PRICING WITH PRICE-RESPONSIVE LOAD



Peak consumption would be lower under an energy-only market than under an ICAP system because consumers could avoid paying energy whose true cost of production exceeds its value to them.

Energy-only pricing systems based on price-responsive load still have several potential limitations relative to an ICAP system. First, these pricing systems can maintain reliability only if there is in fact sufficient load responding to short-term price signals to enable the market to clear during shortage conditions while maintaining reliability levels and generating substantial price cost margins for the available generation resources. Thus, there have to be truly effective demand response programs that can be relied upon to produce real load reductions during high load conditions. Second, practical experience suggests that only limited demand response is available at low energy prices. State and federal regulators must be willing to allow real-time energy prices to rise well above the incremental cost of the marginal generator during shortage conditions in order for demand response to be effective in maintaining reliability. The presence of price-responsive load does not change the reality that the marginal generator must be able to recover its going-forward fixed operating costs. While the presence of substantial price-responsive load greatly reduces or eliminates the potential reliability risks arising from misjudgments under an energy-only market, it does not solve the political problem of high energy prices, and unless demand is very price-elastic, the recovery of fixed costs may be concentrated in particular years, creating political and regulatory risk under retail access systems lacking multi-year energy contracts to hedge prices. Third, customers lacking real-time meters will be unable to avoid paying for power that may, at times, be expensive.

Appendix VI
Aggregate Load Zones and CRR Allocation

A. Introduction

The reliance on LAP zones for pricing has the potential to give rise to a number of problems, particularly when applied to the allocation of CRRs to LSEs serving load from local generation. This appendix illustrates several difficulties relating to CRR allocation that would arise from the use of aggregate load zones for pricing. These difficulties arise because the pricing point for the generation and load would differ, even when their electrical location is essentially the same, and also arise from the zonal representation of load in the SFT.

If the pricing points for generation and load at the same location are different, a LSE serving its load with its own generation at that same location would need to be allocated CRRs from its local generation to the LAP zone pricing point in order to be hedged against congestion charges. The first problem (illustrated in Sections B through E) is that these CRRs are likely to be valuable even in hours in which the LSE does not need a hedge between local generation and the aggregated load zone price, leading to a potential windfall for the LSE during lower load conditions. A related problem illustrated in Section F is that some CRRs from local generation to the aggregate LAP zone will have negative expected values and, thus, will not be voluntarily requested by market participants. However, failure to assign these "counterflow" CRRs can lead to the infeasibility of other CRRs that other LSEs need to hedge congestion charges from their generation to their load. A third problem, illustrated in Section G, is a potential for CRRs from the local generation to the LAP to be infeasible, even though there is actually no congestion between the generation and the actual load. The examples below thus show that the need to designate CRRs between locations that are actually identical will create trade offs between hedging congestion costs for some LSEs and avoiding cost shifting to other LSEs that may be insolvable.

Sections H and I illustrate another set of problems that arise from the zonal representation of sinks in the SFT used to perform the simultaneous feasibility test for CRR allocation. Section H illustrates the underallocation that can result from intra-zonal constraints if CRRs sink at large aggregate load zones. Section I shows the revenue inadequacy that can arise from attempting to address this underallocation by disaggregating the load zone sinks for purposes of the SFT, but subsequently reassembling the disaggregated CRRs into LAP zones for calculating settlements.

These issues are illustrated below by comparing hedging and CRR allocation under a system based on disaggregated CRRs versus a system based on CRRs that sink at aggregated LAP zones. For the purpose of the examples it is assumed that regions A and D in Figure VI-1 are both served by vertically integrated LSEs with 1,000 MW of peak load, 1,000 MW of generation, and that there is 250 MW of transfer capability from B to A and B to D. While peak load can in principle be met without relying on imports, the import capability allows the LSEs at A and D to take advantage of low cost energy available on the spot or term market, as well as to conduct generation maintenance and meet load when generation is not available due to forced outages. In addition, it is assumed that there is 2,500 MW of peak load in region B, 3,000 MW of peak load in region C and transfer capability of 1,500 MW from C to B.

Figure VI-1
ILLUSTRATIVE ASSUMPTIONS FOR LAP ZONE

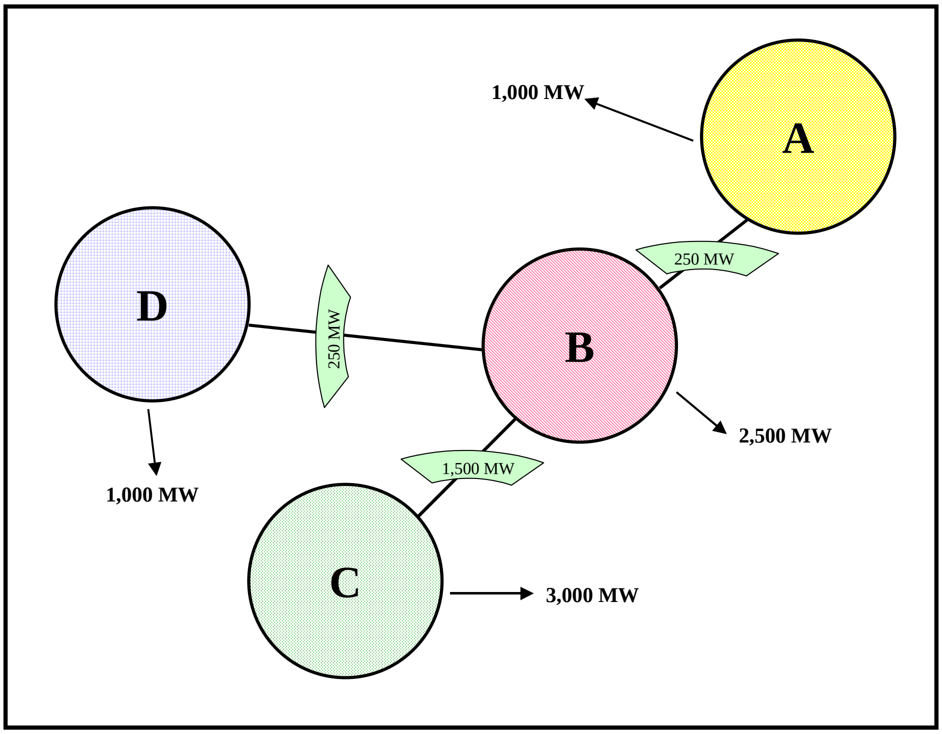
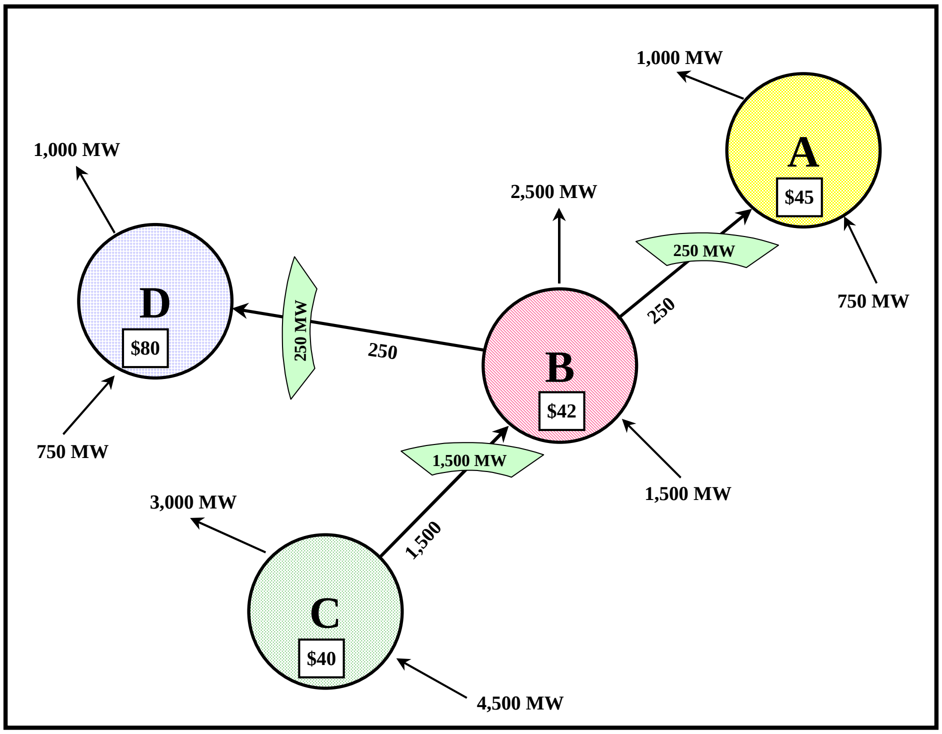


Table VI-2 portrays a simple set of assumptions regarding generation costs and the least cost dispatch, which determine the nodal prices for load shown in Table VI-3. As shown in Figure VI-4, the generation dispatched on the margin determines the nodal price in each of the four regions, which are for simplicity assumed to be radially connected. All radial constraints bind in the dispatch, with imports from generation in region C serving load in regions A, B and D. High cost generation is required at the margin to meet load in regions A and D.

Table VI-2 SCENARIO I GENERATION COSTS				
LSE	**Capacity (MW)**	**Generation (MW)**	**Cost ($/MW)**	**Total Cost ($)**
A	500	250	45	11,250
	500	500	40	20,000
Total A	1,000	750		31,250
B	300	0	44	0
	650	150	42	6,300
	600	600	40	24,000
	750	750	35	26,250
Total B	1,750	1,500		56,550
C	3,500	2,000	40	80,000
	2,500	2,500	35	87,500
Total C	5,000	4,500		167,500
D	500	250	80	20,000
	500	500	60	30,000
Total D	1,000	750		50,000
Total Production	10,300	7,500		305,300

Table VI-3 SCENARIO I NODAL PAYMENTS BY LOAD			
LSE	**Load (MW)**	**Nodal Price ($/MW)**	**Total ($)**
A	1,000	45	45,000
B	2,500	42	105,000
C	3,000	40	120,000
D	1,000	80	80,000
	7,500		350,000

Figure VI-4
SCENARIO I LEAST COST DISPATCH – PEAK LOAD



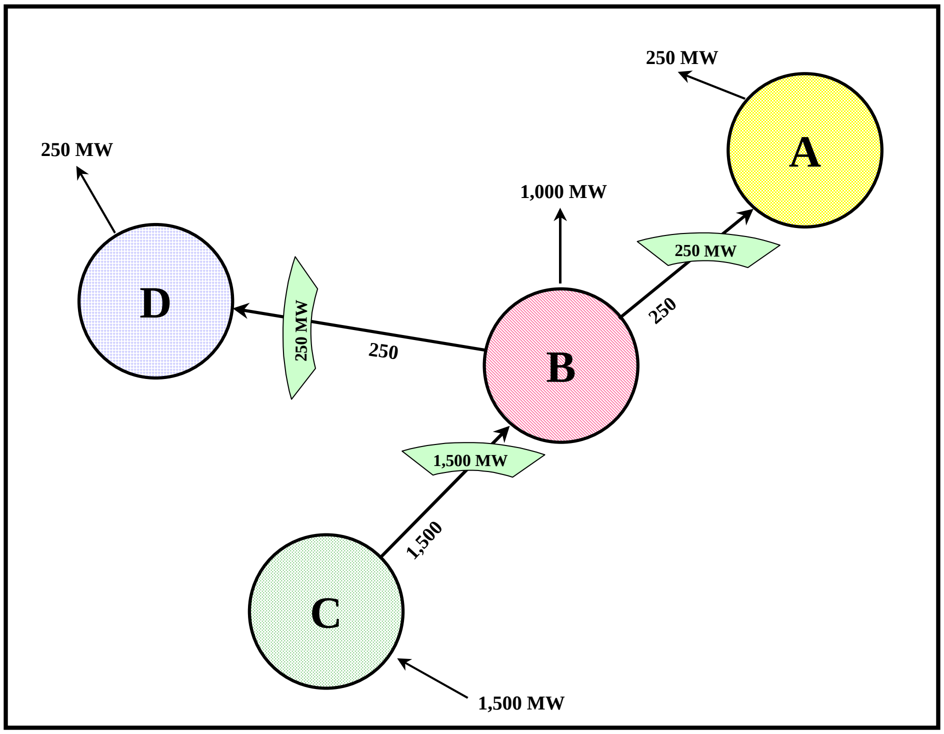
The resulting generation revenues and nodal prices are summarized in Table VI-5.

Table VI-5 SCENARIO I GENERATION REVENUES			
LSE	**Generation (MW)**	**Nodal Price ($/MW)**	**Generation Revenues ($)**
A	750	45	33,750
B	1,500	42	63,000
C	4,500	40	180,000
D	750	80	60,000
Total	7,500		336,750

B. Hedging with Nodal CRRs

We first consider CRR allocation and hedging under a system based on CRRs defined nodally or to disaggregated load zones. For purposes of this example, it is assumed that the LSEs serving load in regions A and D are assigned 250 MW of CRRs from C to A and C to D respectively, while the LSE serving region B is assigned 1,000 C to B CRRs. As shown in Figure VI-6, these CRRs are simultaneously feasible.

Figure VI-6
FEASIBLE NODAL CRR ALLOCATION



The feasible CRRs take on the values shown in Table VI-7 under the scenario I dispatch assumptions.

Table VI-7
SCENARIO I NODAL CRR VALUES

LSE	CRRs	Congestion Rent ($/MW)	CRR Revenues ($)
A	250 C to A	5	1,250
B	1,000 C to B	2	2,000
C	0	0	0
D	250 C to D	40	10,000
Total			13,250

The net cost to the LSEs serving load in regions A, B, C and D is shown in Table VI-8. The net cost of serving load is equal to the payments by load at the nodal price, minus generation revenues (including those from power sales), plus the cost of generation (including the costs incurred for power sales), minus CRR revenue. The net cost of serving load can equivalently be expressed as the cost of own-generation used to serve load, plus the cost of purchased power, plus congestion charges, minus CRR revenue.

Table VI-8
SCENARIO I NET COST TO LOADS – NODAL HEDGING

LSE	Payments by Load ($)	CRR Revenue ($)	Generation Revenue ($)	Generation Cost ($)	Net Cost to Load ($)
A	45,000	-1,250	-33,750	31,250	41,250
B	105,000	-2,000	-63,000	56,550	96,550
C	120,000	0	-180,000	167,500	107,500
D	80,000	-10,000	-60,000	50,000	60,000
Total	350,000	-13,250	-336,750	305,300	305,300

C. Hedging with Zonal/LAP CRRs

The next step in the example is to assess whether the nodal cost of meeting load shown in Table VI-8 can be replicated under a system of zonal price aggregation, in which all LSEs pay a load-weighted average zonal price. The answer, as illustrated by the examples below, is that this replication can be made to occur for a specific snapshot of load and generation through an allocation of CRRs. One problem that arises, however, is that this same set of CRRs will not replicate the nodal pricing outcome in hours in which the actual distribution and level of load and generation do not match those implicit in the snapshot that is the basis for the CRR allocation.

Table VI-9 shows the payments for load for the LSEs serving regions A, B, C and D at an aggregated zonal price of $46.67/MW calculated for all four regions. Regions A, B and C pay higher prices for their load under a zonal pricing system than under a nodal pricing system, while region D pays less. However, the generation costs and revenues for each LSE are unchanged in moving from the nodal to the zonal CRR system, since the generation dispatch is unchanged. Thus, prior to taking into account CRR revenues, the net cost to LSEs A, B and C has increased under LAP, while the net cost to LSE D has fallen. The costs for LSEs A, B and C increase because they are incurring new congestion charges (the difference between the zonal price for load and the nodal price for generation) in delivering their power from their generation to their load under LAP, even though their generation and load are electrically at the same location.

Table VI-9
SCENARIO I NODAL VERSUS ZONAL PAYMENTS BY LOAD

LSE	Load (MW)	Nodal Price ($/MW)	Payments at Nodal Prices ($)	Payments at Zonal Price ($)	Difference
A	1,000	45	45,000	46,666.67	1,666.67
B	2500	42	105,000	116,666.7	11,666.7
C	3,000	40	120,000	140,000.0	20,000
D	1,000	80	80,000	46,666.67	-33,333
Total	7,500	46.66667	350,000	350,000	0

The effect of the artificial congestion charges arising from the zonal price aggregation can be offset by assigning the LSEs serving load in regions A, B, C and D additional CRRs from their generation to the aggregate load zone, as shown in Table VI-10. Hence, for example, LSE A receives an additional 750 MW of CRRs from region A to the LAP zone. In addition, the sink location for its other 250 MW of CRRs is changed from region A to the LAP zone to match the pricing point for its load.

Table VI-10 shows that this allocation of CRRs from generation to the aggregate load zone is simultaneously feasible. In the simultaneous feasibility test, a CRR sinking in the aggregate load zone is assumed to be distributed across the four regions in proportion to peak load: region A, 0.1333 (1000/7500); region B, 0.3333 (2500/7500); region C, 0.4 (3000/7500); and region D, 0.1333 (1000/7500). Each CRR sinking at the aggregate load zone will be modeled in the SFT as sinking 13.33 percent at location A, 33.33 percent at location B, 13.33 percent at location C, and 40 percent at location D. Thus, in the SFT, shown in Table VI-10, a CRR sinking from C to the load zone is shown as having a shift factor of 0.1333 across the B to A constraint, 0.1333 across the B to D constraint, and 0.6 (0.1333 + 0.1333 + 0.3333) across the C to B constraint. Table VI-10 shows that the CRRs to the LAP zone are feasible given the distribution factors used to represent zonal load in the SFT.

Table VI-10
FEASIBLE CRR ALLOCATION TO LAP ZONE

LSE	**CRR**	**B to A Constraint (Limit = 250 MW)**		**B to D Constraint (Limit = 250 MW)**		**C to B Constraint (Limit = 1,500 MW)**	
		CRR Shift Factor	**Flows (MW)**	**CRR Shift Factor**	**Flows (MW)**	**CRR Shift Factor**	**Flows (MW)**
A	250 C to Zone	0.133333	33.33333	0.133333	33.33333	0.6	150
	750 A to Zone	-0.86667	-650	0.133333	100	-0.4	-300
B	1,000 C to Zone	0.133333	133.3333	0.133333	133.3333	0.6	600
	1,500 B to Zone	0.133333	200	0.133333	200	-0.4	-600
C	3,000 C to Zone	0.133333	400	0.133333	400	0.6	1,800
D	250 C to Zone	0.133333	33.33333	0.133333	33.3333	0.6	150
	750 D to Zone	0.133333	100	-0.86667	-650	-0.4	-300
Total			250		250		1,500

Table VI-11 shows that the congestion rent that the LSE serving each region earns for its additional CRRs to the load zone exactly offsets the change in the payments by load (Table IV-8) in moving from a nodal to a zonal pricing system. Note that the additional CRRs assigned to the LSE serving region D have a negative value, because the LAP price is lower than the nodal price at the location of its generation.

Table VI-11
SCENARIO I LOAD ZONE CRR ASSIGNMENT AND VALUE

LSE	CRRs	Congestion Rent ($/MW)	CRR Revenues ($)
A	250 C to Zone	6.666667	1,666.667
	750 A to Zone	1.666667	1,250
Total A	1,000		2,916.667
B	1,000 C to Zone	6.666667	6,666.667
	1,500 B to Zone	4.666667	7,000
Total B	2,500		13,666.67
C	3,000 C to Zone	6.666667	20,000
D	250 C to Zone	6.666667	1,666.667
	750 D to Zone	-33.3333	-25,000
Total D	1,000		-23,333.3
Grand Total	7,500		13,250

The generation revenues and costs are unchanged by the zonal price aggregation, so it can be seen in Table VI-12 that if CRRs are assigned to offset changes in the payments by load in moving from nodal to zonal pricing, then the net cost of meeting load for each region is identical to the cost under a nodal pricing system.

Table VI-12
SCENARIO I NET COST TO LOADS – ZONAL HEDGING

LSE	Payments by Load ($)	CRR Revenue ($)	Generation Revenue ($)	Generation Cost ($)	Net Cost to Load ($)
Net Cost to A	46,666.67	-2,916.67	-33,750	31,250	41,250
Net Cost to B	116,666.67	13,666.67	63,000	56,550	96,550
Net Cost to C	140,000	-20,000	-180,000	167,500	107,500
Net Cost to D	46,666.67	23,333.33	-60,000	50,000	60,000
Total	350,000	-13,250	-336,750	305,300	305,300

D. Hedging with Nodal CRRs – Low Load

Unfortunately, the impacts of zonal price aggregation are more problematic than suggested by the outcome above. The first problem is that the zero cost-shifting outcome shown in Table VI-12 only occurs when the actual day-ahead load is equal to the load (assumed here to be peak) for which the zonal price hedges were allocated. At any other load level, even if the load weights were perfect, the zonal price aggregation would lead to cost shifts. To illustrate this, assume in scenario II that the load in each region is 80 percent of the load assumed in the first scenario, as shown in Table VI-13.

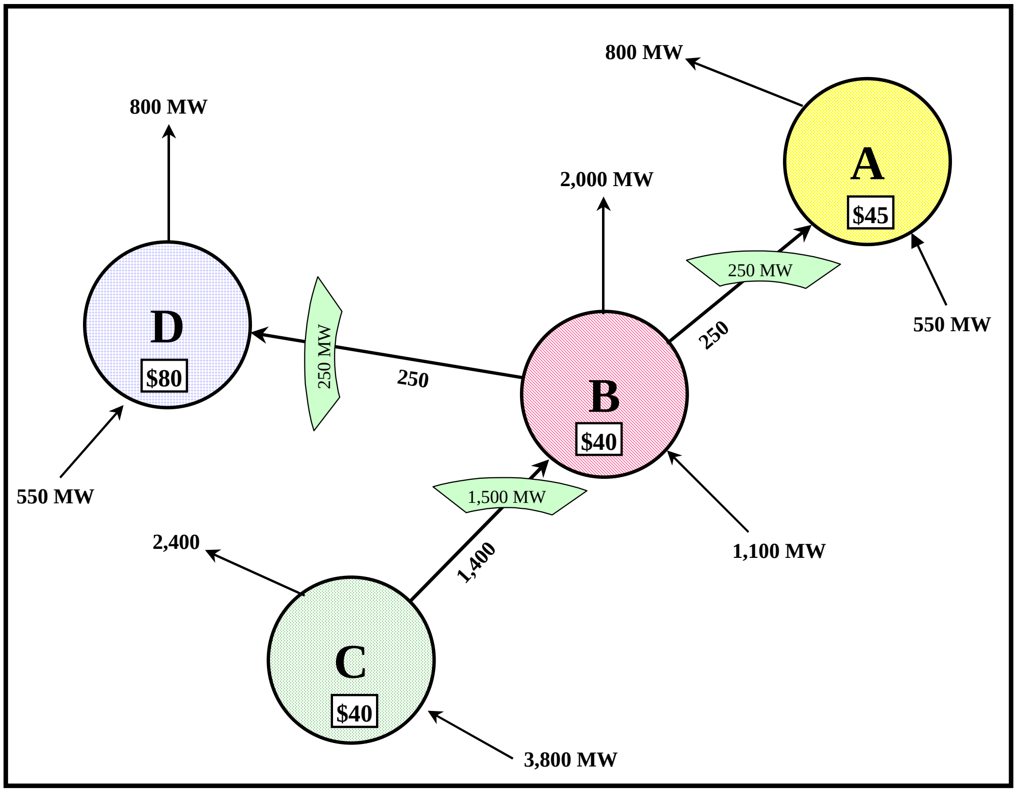
Table VI-13
SCENARIO II NODAL PAYMENTS BY LOAD – LOW LOAD

LSE	Load (MW)	Nodal Price ($/MW)	Total Cost ($)
A	800	45	36,000
B	2,000	40	80,000
C	2,400	40	96,000
D	800	80	64,000
Total	6,000		276,000

In Scenario II, less generation must be dispatched to meet load, as shown in Table VI-14. As shown in Figure VI-15, and the radial constraints from B to A and B to D are binding. The nodal price falls to $40 in region B because, on margin a $40/MW generator would be dispatched to meet incremental load in this region.

Table VI-14 **SCENARIO II GENERATION COSTS – LOW LOAD**				
LSE	**Capacity (MW)**	**Generation (MW)**	**Cost ($/MW)**	**Total Cost ($)**
A	500	50	45	2,250
	500	500	40	20,000
Total A	1,000	550		22,250
B	300	0	44	0
	650	0	42	0
	600	350	40	14,000
	750	750	35	26,250
Total B	2,300	1,100		40,250
C	3,500	1,300	40	52,000
	2,500	2,500	35	87,500
Total C	6,000	3,800		139,500
D	500	50	80	4,000
	500	500	60	30,000
Total D	1,000	550		34,000
Total Production	10,300	6,000		236,000

Figure VI-15
SCENARIO II LEAST COST DISPATCH – 80% LOAD



Generation revenues are portrayed in Table VI-16.

Table VI-16
SCENARIO II GENERATION REVENUES – LOW LOAD

LSE	Generation (MW)	Nodal Price ($/MW)	Generation Revenues ($)
A	550	45	24,750
B	1,100	40	44,000
C	3,800	40	152,000
D	550	80	44,000
Total	6,000		264,750

CRR revenues, which depend only on the nodal prices, change in this example because of the change in the nodal price at B; these revenues are shown in Table VI-17.

Table VI-17 SCENARIO II NODAL CRR VALUES – LOW LOAD			
LSE	**CRRs**	**Congestion Rent ($/MW)**	**CRR Revenues ($)**
A	250 C to A	5	1,250
B	1,000 C to B	0	0
C	0	0	0
D	250 C to D	40	10,000
Total			11,250

Thus, the cost of meeting 80 percent of peak load at nodal prices falls relative to the cost of peak load as shown in Table VI-18.

Table VI-18 SCENARIO II NET COST TO LOADS – NODAL HEDGING AND LOW LOAD					
LSE	**Payments by Load ($)**	**CRR Revenue ($)**	**Generation Revenue ($)**	**Generation Cost ($)**	**Net Cost to Load ($)**
Net Cost to A	36,000	-1,250	-24,750	22,250	32,250
Net Cost to B	80,000	0	-44,000	40,250	76,250
Net Cost to C	96,000	0	-152,000	139,500	83,500
Net Cost to D	64,000	-10,000	-44,000	34,000	44,000
Total	276,000	-11,250	-264,750	236,000	236,000

E. Hedging with Zonal/LAP CRRs – Low Load

The next step in the example is to illustrate the cost of meeting 80 percent of peak load for each LSE under a system of zonal price aggregation, rather than under nodal pricing. For scenario II, the zonal price and payments by load are shown in Table VI-19.

Table VI-19
SCENARIO II NODAL VERSUS ZONAL PAYMENTS BY LOAD – LOW LOAD

LSE	Load (MW)	Nodal Price ($/MW)	Payments at Nodal Prices ($)	Payments at Zonal Price ($)	Load Zone LDF Weights
A	800	45	36,000	36,800	0.133333
B	2,000	40	80,000	92,000	0.333333
C	2,400	40	96,000	110,400	0.4
D	800	80	64,000	36,800	0.133333
Total	6,000	46	276,000	276,000	1.0

Given this LAP zone price and the generation prices from Table VI-16, Table VI-20 shows the value of the additional aggregate LAP CRRs that were allocated so as to hedge peak load in scenario I.

Table VI-20
SCENARIO II LAP CRR ASSIGNMENT AND VALUE – LOW LOAD

LSE	CRRs	Congestion Rent ($/MW)	CRR Revenues ($)
A	250 C to Zone	6	1,500
	750 A to Zone	1	750
Total A	1,000		2,250
B	1,000 C to Zone	6	6,000
	1,500 B to Zone	6	9,000
Total B	2,500		15,000
C	3,000 C to Zone	6	18,000
D	250 C to Zone	6	1,500
	750 D to Zone	-34	-25,500
Total D			-24,000
Grand Total	7,500		11,250B

The resulting cost of meeting each load with a system of zonal price aggregation is shown in Table VI-21. In scenario II, the zonal price aggregation reduces the cost of meeting load for the LSEs serving load in regions A, B and C, and raises the cost of the LSE serving load in the high cost region D, relative to a less aggregate pricing system. This result may at first seem counter-intuitive, as the LSE that loses under zonal /LAP aggregation is the LSE whose wholesale price is reduced by the aggregation. The reason for this is that under LAP pricing an additional allocation of CRRs was made to hedge the congestion costs between generation and load that are at the same location (i.e., the A to load zone, and B to load zone CRRs), so as to provide the same net cost of serving peak load as under a nodal pricing system. While these additional CRRs were allocated to provide a congestion hedge for the peak level of load, they will also accrue value at lower load levels. Hence, in the example, under zonal aggregation the LSE serving load in region A receives $200 more in congestion rent for its CRRs than the congestion costs that it actually incurs at the 80 percent load level. Similarly, the LSE serving load in region B receives congestion rents that exceed its congestion costs by $3,000, and the LSE serving load in region C receives a net of $3,600. The example shows that if an LSE is allocated sufficient CRRs from its generation to the aggregated LAP load zone to hedge its peak load, then it may receive a windfall during lower load conditions, when it will earn congestion rents that are not needed to offset congestion charges between its generation and its load that are actually at the same location.

The CAISO could attempt to avoid such windfalls by allocating such LSEs fewer CRRs than are required to fully hedge the LSEs' congestion costs at peak load, in the expectation that these congestion costs would be offset by excessive CRR revenues during lower load conditions. The difficulty with this approach is that the LSE's customers are then not fully hedged against congestion. If the actual congestion levels differ from those expected by the CAISO, the LSE's customers may be adversely impacted.

Table VI-21
SCENARIO II NET COST TO LOADS – ZONAL HEDGING AND LOW LOAD

LSE	Payments by Load ($)	CRR Revenue ($)	Generation Revenue ($)	Generation Cost ($)	Net Cost to Load ($)
Net Cost to A	36,800	-2,250	-24,750	22,250	32,050
Net Cost to B	92,000	-15,000	-44,000	40,250	73,250
Net Cost to C	110,400	-18,000	-152,000	139,500	79,900
Net Cost to D	36,800	24,000	-44,000	34,000	50,800
Total	276,000	-11,250	-264,750	236,000	236,000

F. Counterflow CRR Allocations to LAP Zones

Another dimension of this problem is that the LSE serving load in region D pays for the windfall received by the LSEs serving load on regions A, B and C through the payment of a total of -$24,000 in congestion rents, i.e., its net payment into the congestion account for holding counterflow CRRs. This payment exceeds D's decrease in congestion costs under LAP pricing, leading to a net increase in its costs of $6,800. As part of the settlements, D pays -$25,500 (see Table VI-20) in congestion rents for its CRRs from region D to the aggregate load zone. These CRRs are counterflow FTRs under zonal price aggregation and have large negative values that are offset by counterflow revenues for the generation at high load (i.e., the difference between the high price paid for D's generation and the low aggregated price paid by its load), but these CRR payments are not offset by counterflow generation revenues at lower load levels when less generation is dispatched. At lower loads, the quantity of LSE D's load met at the lower zonal aggregation prices falls, but the quantity of CRRs is fixed, raising its net cost of meeting load.

The anomalous result for the LSE serving load in region D suggests the next difficulty with zonal aggregation, which is that it implicitly entails the assignment of counterflow CRRs to LSEs with load in high cost regions. In both scenarios I and II the CRRs from D to the aggregate LAP zone have negative values, even though the generation in region D is by definition at the same location as LSE D's load. LSE D would be much better off if it did not accept these counterflow CRRs during the CRR allocation process and, if given a choice, it presumably would not do so. For LSE D, the cost of accepting the counterflow congestion hedge is likely to exceed any risk reducing value.

The difficulty with this outcome is that in order for the allocated set of CRRs for the other LSEs to be simultaneously feasible, LSE D needs to accept the counterflow CRRs from D to the aggregate load zone. In the zonal aggregation these counterflow CRRs are necessary not only to the simultaneous feasibility of the C to load zone CRRs held by D, but also of the C to load zone CRRs held by the LSEs for the other regions. Table VI-22 shows the implied shift factors of the various generator to aggregate load zone CRRs assigned for the zonal pricing example. For example, because 1/7.5 of the load for the aggregate load zone is in region A, a CRR from generation at A to the aggregate load zone would produce counterflows across the B to A constraint of .8667 MW per MW of CRR from A to the aggregate load zone. With the assumed allocation of CRRs, the flows across each of the constraints, B to A, B to D and B to C associated with the CRRs is less than or equal to each of the line limits so the allocation satisfies the SFT. Table IV-22 shows that in this SFT the D to aggregate load zone CRRs provide counterflow on constraints C to B and B to D.

Table VI-22
FEASIBLE SFT, WITH COUNTERFLOW FROM ALL CRRs

		B to A Constraint (Limit = 250 MW)		**B to D Constraint (Limit = 250 MW)**		**C to B Constraint (Limit = 1,500 MW)**	
LSE	**CRR**	**CRR Shift Factor**	**Flows (MW)**	**CRR Shift Factor**	**Flows (MW)**	**CRR Shift Factor**	**Flows (MW)**
A	250 C to Zone	0.133333	33.33333	0.133333	33.33333	0.6	150
	750 A to Zone	-0.86667	-650	0.133333	100	-0.4	-300
B	1,000 C to Zone	0.133333	133.3333	0.133333	133.3333	0.6	600
	1,500 B to Zone	0.133333	200	0.133333	200	-0.4	-600
C	3,000 C to Zone	0.133333	400	0.133333	400	0.6	1,800
D	250 C to Zone	0.133333	33.33333	0.133333	33.3333	0.6	150
	750 D to Zone	0.133333	100	-0.86667	-650	-0.4	-300
Total			250		250		1,500

Because the D to aggregate load zone CRRs likely have a negative value, though, the LSE at D would prefer not to accept such D to load zone CRRs; it would likely be better off if it were unhedged. Table VI-23 shows the SFT test for the remaining CRRs if LSE D were permitted to choose not to accept the D to aggregate load zone CRRs. It shows that the allocation of CRRs in Table VI-10 would substantially overload the B to D and C to B limits, absent the counterflow provided by the D to aggregate load zone CRRs.

Revenue adequacy could be restored by proportionately prorating down the CRRs in Table VI-23 to bring the CRR flows below the rating limits enforced in the SFT. However, Table VI-24 shows that a very substantial prorationing of CRR allocations would be required to satisfy the SFT and that the number of congestion hedges available to the LSEs at A, B and C would be dramatically reduced relative to the earlier example. As a result, these LSEs would be exposed to substantial congestion risk as well as cost shifting.

Table VI-23
INFEASIBLE SFT, WITH NO COUNTERFLOW FROM COUNTERFLOW D TO ZONE CRRs

		B to A Constraint (Limit = 250 MW)		**B to D Constraint (Limit = 250 MW)**		**C to B Constraint (Limit = 1,500 MW)**	
LSE	**CRR**	**CRR Shift Factor**	**Flows (MW)**	**CRR Shift Factor**	**Flows (MW)**	**CRR Shift Factor**	**Flows (MW)**
A	250 C-Zone	0.133333	33.33333	0.133333	33.33333	0.6	150
	750 A-Zone	-0.86667	-650	0.133333	100	-0.4	-300
B	1,000 C-Zone	0.133333	133.3333	0.133333	133.3333	0.6	600
	1,500 B-Zone	0.133333	200	0.133333	200	-0.4	-600
C	3,000 C-Zone	0.133333	400	0.133333	400	0.6	1,800
D	250 C-Zone	0.133333	33.33333	0.133333	33.3333	0.6	150
	750 D-Zone	0.133333	0	-0.86667	0	-0.4	0
Total Flow			150		900		1,800

Table VI-24 SFT – NO COUNTERFLOW: REDUCED CRR ALLOCATION							
		B to A Constraint		B to D Constraint		C to B Constraint	
LSE	CRR	CRR Shift Factor	Flows	CRR Shift Factor	Flows	CRR Shift Factor	Flows
Limit			250		250		1,500
A	69.44444 C-Zone	0.133333	9.259259	0.133333	9.259259	0.6	41.66667
	208.3333 A-Zone	-0.86667	-180.556	0.133333	27.77778	-0.4	-83.3333
B	277.7778 C-Zone	0.133333	37.03704	0.133333	37.03704	0.6	166.6667
	416.6667 B-Zone	0.133333	55.55556	0.133333	55.55556	-0.4	-166.667
C	833.3333 C-Zone	0.133333	111.1111	0.133333	111.1111	0.6	500
D	69.44444 C-Zone	0.133333	9.259259	0.133333	9.259259	0.6	41.66667
	0 D-Zone	0.133333	0	-0.86667	0	-0.4	0
Total Flow			41.66667		250		500

G. Infeasibility of Zonal/LAP CRR Hedges Its Load

A final problem arising because the use of LAP causes the pricing point for local generation and load to be different is that the CRRs that an LSE needs to hedge congestion charges between its generation at the same location may be infeasible, even though there is actually no congestion between the local generation and the actual physical load. Suppose that the load for the LSE for region A were to increase from 1,000 to 1,150 MW. Due to the binding transmission constraint of 250 MW between regions B and A, the increment of 150 MW would be served through the dispatch of an additional 150 MW from LSE A's $45/MW generator in region A. In order for LSE to be fully hedged against the congestion charges accruing from the difference in the pricing points for its local load and generation, it would need to be awarded an additional 150 MW of CRRs from region A to the aggregate load zone.

The difficulty, as shown in Table VI-25, is that it is not feasible to award LSE A any additional CRRs from region A to the LAP zone. The table shows that each additional CRR from region A to the LAP zone causes an additional 0.1333 MW of flow on the B to D constraint in the SFT. The additional 150 MWs of CRRs needed to hedge LSE A against the LAP would cause the 250 MW B to D constraint to be overloaded in the SFT by 20 MW = 150*0.133333. LSE A could not be awarded the additional 150 MWs that it needs to provide a hedge between its generation in A and its load, when priced at the LAP. Even though A's generation and load are physically at the same location, the LAP system produces the perverse result that in order to receive the CRRs that it needs, LSE A would need to expand the transmission system from B to D. Under LAP pricing the LSEs in the example could not achieve the same hedge against congestion charges as under nodal pricing, where there would be no difference between the pricing point for LSE A's load and local generation.

A second possible result in this situation is that a prorating system could be used in the CRR allocation process, so that all LSEs would receive a proportionately reduced number of CRRs that have positive shift factor on the B to D constraint. Thus all LSEs would be exposed to congestion risk to a degree.

Table VI-25
INFEASIBLE SFT, WITH ADDITIONAL 150 MWs OF CRRs FOR LSE A
(Counterflow from all CRRs)

		B to A Constraint (Limit = 250 MW)		B to D Constraint (Limit = 250 MW)		C to B Constraint (Limit = 1,500 MW)	
LSE	**CRR**	**CRR Shift Factor**	**Flows (MW)**	**CRR Shift Factor**	**Flows (MW)**	**CRR Shift Factor**	**Flows (MW)**
A	250 C to Zone	0.133333	33.33333	0.133333	33.33333	0.6	150
	750 A to Zone	-0.86667	-650	0.133333	100	-0.4	-300
(new CRR)	150 A to Zone	-0.86667	-130	0.133333	20	-0.4	-60
B	1,000 C to Zone	0.133333	133.3333	0.133333	133.3333	0.6	600
	1,500 B to Zone	0.133333	200	0.133333	200	-0.4	-600
C	3,000 C to Zone	0.133333	400	0.133333	400	0.6	1,800
D	250 C to Zone	0.133333	33.33333	0.133333	33.3333	0.6	150
	750 D to Zone	0.133333	100	-0.86667	-650	-0.4	-300
Total			120		270		1,440

H. Underallocation of CRRs resulting from LAP Sinks

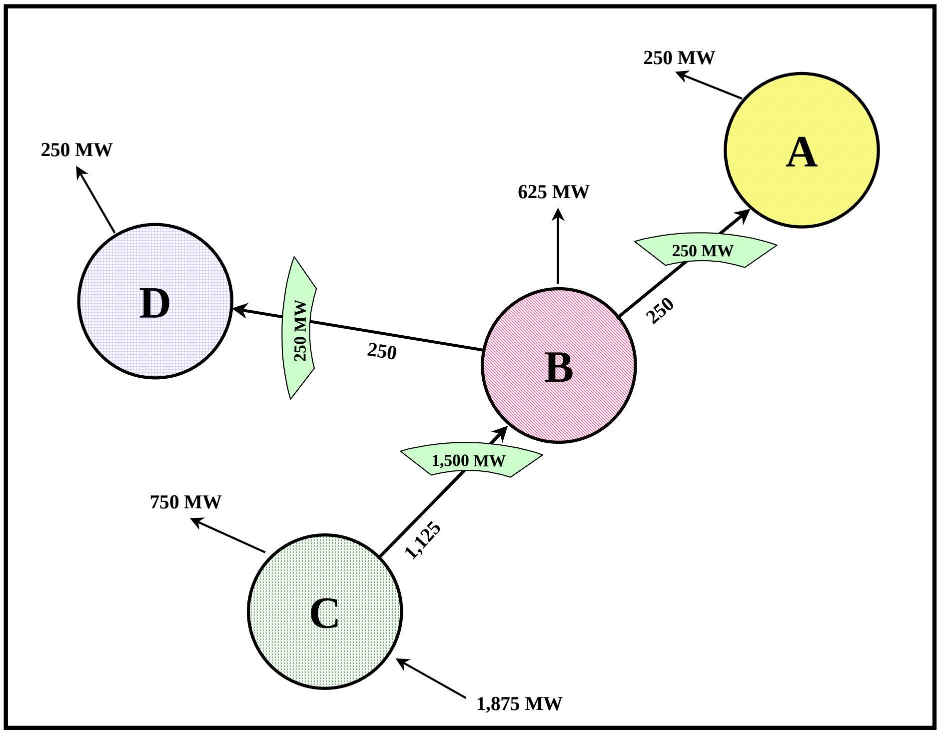
The example also can be used to illustrate the underallocation that can result from requiring CRRs to sink at large aggregate load zones. Suppose that the LSEs for regions A, B, C and D all continue to be part of a large LAP zone and are all serving their load through a combination of spot market purchases and longer-term purchases from generation located in region C. The only CRRs that the LSEs request are therefore from region C to the LAP zone; they do not request any CRRs from local generation to the LAP zone. Real-time peak load is distributed across the four regions in the same proportions as in the preceding examples: region A, 0.1333 (1000/7500); region B, 0.3333 (2500/7500); region C, 0.4 (3000/7500); and region D, 0.1333 (1000/7500). Each CRR that an LSE requests to the LAP zone will be modeled in the SFT as sinking 13.33 percent at location A, 33.33 percent at location B, 13.33 percent at location C, and 40 percent at location D.

Now suppose each of the LSEs A, B and D were to request a quantity of C to LAP CRRs equal to its load – thus A requests 1,000 MW; B requests 2,000 MW; and D requests 1,000 MW. In the SFT, 4,500*0.1333 MW = 600 MW of CRR flows would be modeled as sinking at location A, 4,500 * 0.3333 = 1,500 MW modeled as sinking at location B, 4,500*0.4 = 1,800 MW modeled as sinking at node C and 4,500*0.1333 = 600 MW modeled as sinking at node D. The ISO would find, however, that only 250 MWs of CRR flows are feasible from C to location A (42 percent of request), 1,500 MW are feasible to location B (100 percent), and 250 MW are feasible to location D (42 percent). Differences in the feasibility of CRRs sinking at different nodes result from transmission constraints internal to the aggregated load zone, specifically the 250 MW limits from B to A and from B to D. If CRRs are defined and evaluated to an aggregate load zone sink, the most limiting of the internal constraints would limit the quantity of CRRs awarded and the LSEs as a group could only be awarded 42 percent of their aggregate request for CRRs. Thus, 4,500 MW * 13.33% * 42% = 250 MW is the maximum quantity of CRRs that may feasibly sink at location A or D. Thus, the LSEs can only be awarded 4,500 * 42% = 1875 MW of CRRs in total to the aggregate zone. As shown in Table VI-26 and Figure VI-27 below, this quantity of CRRs from C to the LAP will exhaust the transmission from B to A and from B to D, but will leave some room on the constraint from C to B.

Table VI-26
FEASIBLE QUANTITY OF CRRs FROM C TO THE LAP

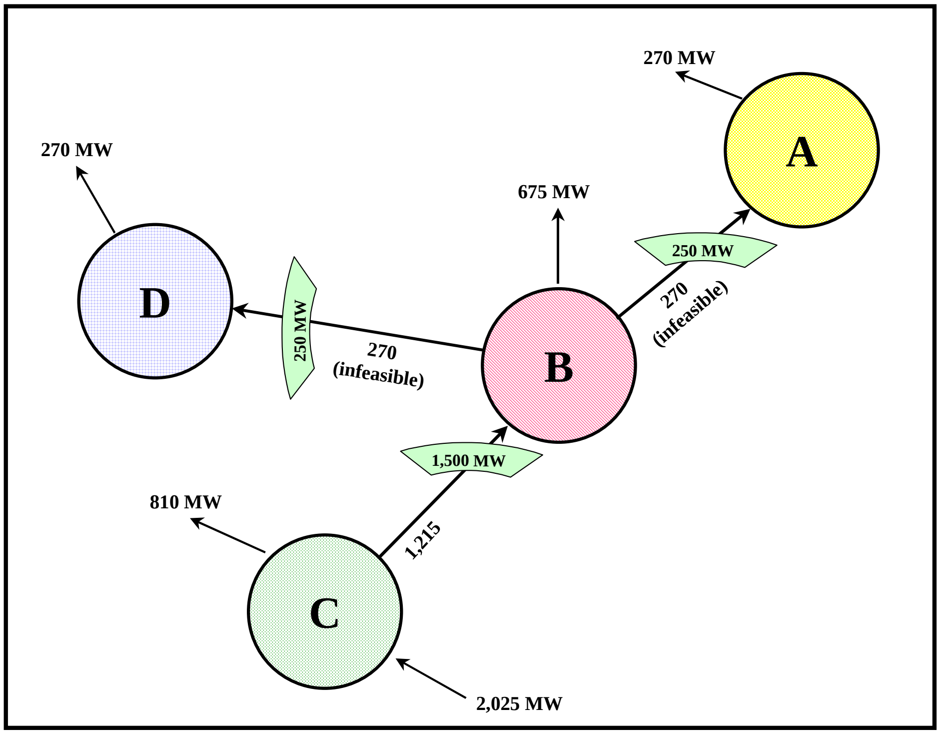
CRR	B to A Constraint (Limit = 250 MW)		B to D Constraint (Limit = 250 MW)		C to B Constraint (Limit = 1,500 MW)	
	CRR Shift Factor	Flows (MW)	CRR Shift Factor	Flows (MW)	CRR Shift Factor	Flows (MW)
1,875 C to Zone	0.133333	250	0.133333	250	0.6	1,125

Figure VI-27
SFT FOR 1,875 MW OF C TO LAP CRRs



It would not be possible to award any more CRRs to the aggregate load zone without violating the simultaneous feasibility test and revenue adequacy condition, even though additional CRRs would be feasible to specific nodes within the LAP. For instance if the LSEs were awarded 45 percent of their request (2,025 MW), in the simultaneous feasibility test the flows would exceed the 250 MW feasible to location A: 4,500 MW * 13.333% *45% = 270 MW, as shown in Figure VI-28 below.

Figure VI28
INFEASIBLE SFT FOR 2,025 MW OF C TO LAP CRRs



The 1,875 MW of feasible CRRs from C to the LAP do not provide the LSEs in the example with as much hedge against congestion charges as the CRRs that could be assigned on a nodal basis in the example in Table VI-7. This is because, in comparison with the nodally-assigned CRRs, the CRRs from C to the LAP do not fully use the transmission capacity that is available on the C to B constraint. This is shown in Table VI-29. In the table, the CRR settlements are calculated using the prices presented in scenario I and compared to the settlements previously shown for nodally-defined CRRs. While the C to LAP CRRs could be allocated to benefit particular LSEs relative to a system that awards less aggregated CRRs, overall the LSEs will be less well hedged with the LAP CRRs.

Table VI-29 SCENARIO I – VALUE OF CRRs DEFINED TO LAP VERSUS NODES						
	Nodal Sinks		LAP Sinks		CRR Revenues ($)	
LSE	CRRs	Congestion Rent ($/MW)	CRRs	Congestion Rent ($/MW)	Nodal	LMP
A	250 C to A	5			1,250	
B	1,000 C to B	2			2,000	
C	0	0			0	
D	250 C to D	40			10,000	
Total			1,875 C to LAP	6.66667	13,250	12,500

In this example, an additional 375 MW of CRRs would be feasible from location C to location B, since the C to B rating limit is 1,500 MW. Table VI-30 shows that the 1,875 MW of C to LAP CRRs and the 375 MW of C to B CRRs are feasible together.

Table VI-30 FEASIBLE CRRs FROM C TO THE LAP AND FROM C TO B						
	B to A Constraint (Limit = 250 MW)		B to D Constraint (Limit = 250 MW)		C to B Constraint (Limit = 1,500 MW)	
CRR	CRR Shift Factor	Flows (MW)	CRR Shift Factor	Flows (MW)	CRR Shift Factor	Flows (MW)
1,875 C to Zone	0.133333	250	0.133333	250	0.6	1,125
375 C to B	0	0	0	0	1.0	375
Total		250		250		1,500

The 375 MW of CRRs that are not allocated from C to B (because they do not have a LAP sink) have a congestion rent of 2 $/MW * 375 MW = $750 in scenario I, which is the difference in the value of the CRRs defined nodal and those defined to sink at the LAP, as shown in Table VI-29.

I. Revenue Inadequacy from Inconsistent Reassembly of CRRs

The underallocation of CRRs that may result from the use of load zone sinks may be addressed by disaggregating the definition used for CRR sinks in the SFT. The disaggregation of load zones is only consistent with revenue adequacy, however, if CRRs are awarded and settled based on the same disaggregated load regions to which the SFT has been applied. The previous example can also be used to illustrate the revenue inadequacy that may result if the CRR settlements are made based on an infeasible reassembly of the CRRs that were actually determined to satisfy the SFT.

In the example, 1,875 MWs of C to LAP CRRs and 375 MW of C to B CRRs are found to be simultaneously feasible. No revenue inadequacy would result if the LSEs in the example were actually allocated 1,875 MWs of C to LAP CRRs and 375 MW of C to B CRRs. Infeasibility would occur, however, if the C to B CRRs were instead "reassembled" and 2,250 MW of CRRs were awarded to the aggregated load zone. With the load distribution factors in the example, this would mean that 2,250 * 13.333% = 300 MW of flows would implicitly cross the B to A and B to D constraints in the SFT, as shown in Figure VI-31 and Table VI-32.

Figure VI-31
INFEASIBLE SFT FOR CRR REASSEMBLY

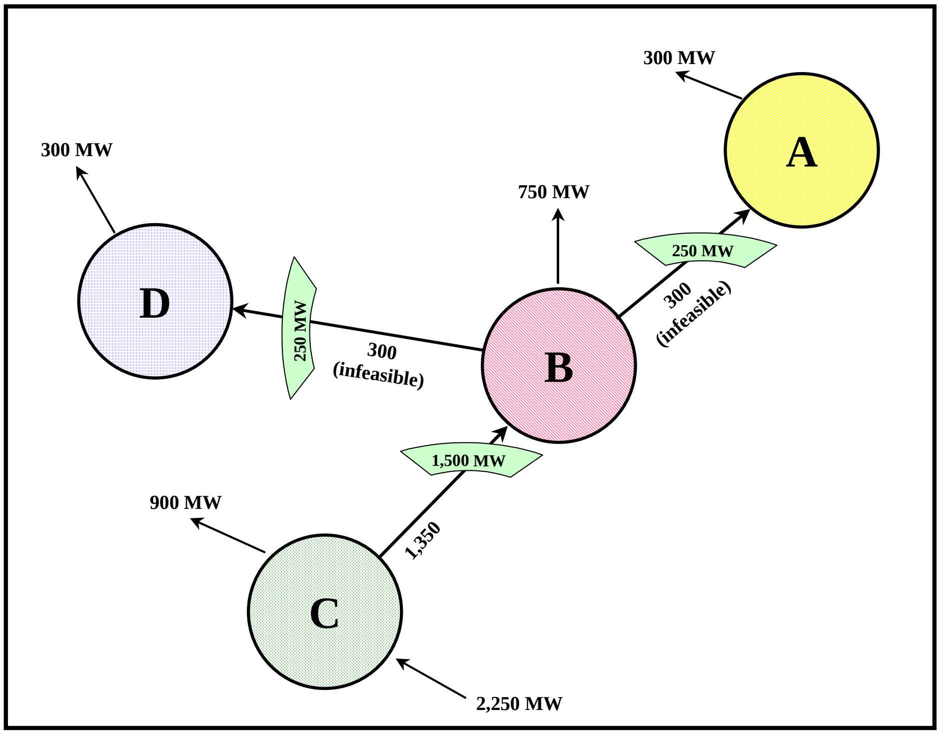


Table VI-32 INFEASIBLE REASSEMBLY OF CRRs FROM C TO THE LAP AND FROM C TO B						
	B to A Constraint (Limit = 250 MW)		B to D Constraint (Limit = 250 MW)		C to B Constraint (Limit = 1,500 MW)	
CRR	CRR Shift Factor	Flows (MW)	CRR Shift Factor	Flows (MW)	CRR Shift Factor	Flows (MW)
1,875 C to Zone	0.133333	250	0.133333	250	0.6	1,125
375 C to Zone	0.133333	50	0.133333	50	0.6	225
Total		300		300		1,350

The 300 MW implicitly crossing the constraints exceeds the 250 MW rating limits for B to A and B to D. Thus, revenue inadequacy may occur if the ISO were to pay congestion rents for 2,250 MW of CRRs from C to the LAP.

Thus, the total payments by load at the zonal price would be $350,000 as shown in Table VI-9, while generator revenues would be $336,750 (Table VI-8) producing congestion rents of $13,250. Settling 2,250 MW of C to LAP CRRs, however, would cost $15,000 ($6.67 * 2,250), producing a CRR shortfall of $1,750.

Appendix VII
Load-Following CRRs

This appendix illustrates some of the potential administrative difficulties of a system in which CRRs follow load. Suppose that there were 12,000 MW of peak load priced at load zone B, and that the four LSEs serving load at B were in aggregate assigned: 1,000 A-B CRRs, 500 C-B CRRs, 2,000 D-B CRRs, 750 E-B CRRs and 25 F-B CRRs.

These CRRs could then be assigned to each LSE in proportion to its share of the 12,000 total load in the zone. Thus, the Blue, Red, Green and Yellow LSEs would be assigned CRRs as portrayed in Table VII-1.

Table VII-1
UNIFORM CRR ALLOCATION TO LSEs

LSEs	CRRs					Load
	A-B 1,000	C-B 500	D-B 2,000	E-B 750	F-B 25	12,000
Blue	20.83333	10.41667	41.66667	15.625	0.520833	250
Red	62.5	31.25	125.0	46.875	1.5625	750
Green	875.0	437.5	1,750.0	656.25	21.875	10,500
Yellow	41.66667	20.83333	83.33333	31.25	1.041667	500
	1,000.0	500.0	2,000.0	750.0	25.0	12,000

Suppose that Blue gained 50 MW of load from Green. If the CRRs were reallocated to follow this shift in load, then Blue would gain CRRs as portrayed in Table VII-2 based on a 50/12,000 increase in its load ratio share of the load in zone B.

Table VII-2
BLUE CRRs AFTER GAINING CUSTOMERS FROM GREEN

	CRRs					Load
	A-B 1,000	C-B 500	D-B 2,000	E-B 750	F-B 25	12,000
Initial CRRs	20.83333	10.41667	41.66667	15.625	0.520833	250
Gain from Green	4.166667	2.083333	8.333333	3.125	0.104167	50
Final Blue CRRs	25.0	12.5	50.0	18.75	0.625	300

If the Blue LSE then lost 25 MW of load to Red and 10 MW of load to Yellow, the CRRs would again be reallocated using the same fractional allocation rule, as shown in Table VII-3. Blue would lose 25/12,000 of each of its CRRs to Red and 10/12,000 of each of its CRRs to Yellow. Similar calculations would be required to determine the increase in CRRs for Red and Yellow. Hence, this method for reallocating CRRs among LSEs would require the CAISO billing and settlement to track and settle fractional CRRs that result from shifts in load, perhaps even on a daily basis.

Table VII-3
BLUE CRRs AFTER LOSING CUSTOMERS TO RED AND YELLOW

	CRRs					
	A-B 1,000	**C-B 500**	**D-B 2,000**	**E-B 750**	**F-B 25**	**Load 12,000**
Blue CRRs	25.0	12.5	50.0	18.75	0.625	300
Loss to Yellow	2.083333	1.041667	4.166667	1.5625	0.052083	25
Loss to Red	0.833333	0.416667	1.666667	0.625	0.020833	10
Final Blue CRRs	22.08333	11.04167	44.16667	16.5625	0.552083	265

Load switching among LSEs presents challenges just in accounting correctly for the settlements for power delivered to a customer at a fixed location. It would be a substantial increase in complexity to have to account for an increasingly complex pattern of CRRs from many possible generating points. Moreover, such an approach would not even provide an allocation of CRRs that would match an LSE's hedging preferences. LSE's would face a constant reallocation of their CRRs, and the administrative cost to them of tracking and trading to assemble such fractional CRRs into useful hedges from their specific generation sources may even exceed the hedge value of the CRRs. In effect, the result of reallocating CRRs to follow shifts in load would be the costly development of a complicated administrative procedure for tracking and reassigning the initially allocated CRRs, followed by additional LSE or customer costs for seeking out different patterns of CRRs for hedging and for trading and managing their CRR portfolio.

The administrative process required to track CRRs is likely to be even more complicated than that shown in the example in Tables VII-1 through VII-3; this example is simplified by the feature that each LSE's holdings of each CRR are proportional to its retail load. This simplification arises because of the assumption that each LSE is initially assigned the same proportion of each CRR. This symmetry will probably not exist under the CAISO's proposed CRR allocation methodology, as each LSE would be able to designate different sources for their initial allocation CRRs. Variations in the share of each CRR initially allocated to each LSE would further complicate the tracking process.

Table VII-4 portrays an initial allocation in which various LSEs have selected CRRs with varying sources. The total quantity of CRRs allocated is 3,500 MW, which is the same as before. However, neither each LSE's share of this total, nor its share of the CRRs between any two locations is necessarily equal to its load ratio share for zone B.

Table VII-4
NON-UNIFORM DESIGNATION OF CRRs

LSEs	CRRs					Load 12,000
	A-B 1,000	C-B 500	D-B 2,000	E-B 750	F-B 25	
Blue	0	0	0	100	0	250
Red	250	0	0	0	0	750
Green	750	250	2,000	650	25	10,500
Yellow	0	250	0	0	0	500
Total	1,000	500	2,000	750	25	12,000

Suppose again that the Green LSE loses 50 MW of load to the Blue LSE. In this case, the reallocation of CRRs from Green to Blue must be based on Green's unique allocation of CRRs. Table VII-5 shows that as in the first example it continues to be necessary for the CAISO to track and settle fractional CRRs. Now, in addition, the allocation of CRRs is not a fixed function of Blue's additional load ratio share (e.g., 50/12,000) of the total CRRs between any source and sink, but would vary for each CRR. For instance, it would gain (50/10,500*750) of Green's CRRs from A-B versus (50/10,500*250) of Green's CRRs from C-B.

Table VII-5
BLUE CRRs AFTER GAINING CUSTOMERS FROM GREEN

	CRRs					Load 12,000
	A-B 1,000	C-B 500	D-B 2,000	E-B 750	F-B 25	
Blue Initial	0	0	0	100	0	250
Gain from Green	3.571429	1.190476	9.52381	3.095238	0.119048	50 Green-Blue
Final Blue CRRs	3.571429	1.190476	9.52381	103.0952	0.119048	300

If Blue then lost 25 MW of load to Yellow and 10 MW Red, it would lose a share both of the CRRs it was initially assigned and those it acquired in reassignments from Green as shown in Table VII-6. Thus, Blue would lose (25/300*103.0952) CRRs from E-B to Yellow. It is apparent that there would likely be an increased need to track and settle very small fractions of CRRs if LSEs were to initially acquire varying amounts of CRRs with different sinks, which is the current proposal.[1]

Table VII-6 BLUE CRRs AFTER LOSING CUSTOMERS TO YELLOW AND RED						
	CRRs					**Load 12,000**
	A-B 1,000	**C-B 500**	**D-B 2,000**	**E-B 750**	**F-B 25**	
Blue	3.571429	1.190476	9.52381	103.0952	0.119048	300
Loss to Yellow	0.297619	0.099206	0.793651	8.59127	0.009921	25 Blue-Yellow
Loss to Red	0.119048	0.039683	0.31746	3.436508	0.003968	10 Blue-Red
Final Blue CRRs	3.154762	1.051587	8.412698	91.06746	0.105159	265 Blue

These complexities of the CRR allocation and reallocation process are potentially avoidable if the CAISO were to account for load shifts by reallocating the economic value of a *given set* of CRRs (i.e., dollars), rather than the CRRs themselves. The economic value of CRR awards can be measured by the market clearing prices in the proposed monthly CRR auction. Thus, the economic value of the CRRs could be made to follow load through cash payments based on the monthly auction market value of the assigned set of CRRs. This process would be administratively much simpler, since it is much easier to track and reassign fractional dollars than fractional CRRs. Moreover, the process would make it straightforward for LSEs to acquire the hedges that they need, after taking into account changes in their load and generation position. The LSEs would choose which CRRs they want to acquire in an auction, rather than having to trade and reconfigure the fractional CRRs that they receive through an administrative reallocation. The funding for the CRRs that they acquire would be provide in whole or part by the payment they receive for their proportionate share of the monthly auction value, which is derived from their initial CRR allocation and subsequent shifts in load. Each LSE would be in the same net position at the end of the month after receiving an allocated share of CRR dollars, rather than CRRs. The only difference is that the LSEs administrative costs would be lower for engaging in CRR trading to achieve its desired hedge position.

---

[1] In the documents reviewed, the CAISO does not state whether CRRs allocated on an annual versus monthly basis would be reallocated jointly or whether the monthly allocations would be reassigned first or some other procedure followed. Taking account of this distinction would further complicate the reallocation of CRRs to follow load.

The CAISO should reconsider whether the objective of assigning the CRR value of the transmission system to those that pay the embedded costs of the grid is most efficiently attained through the allocation of CRRs to load or the allocation of the market value of CRRs to load through a system based on auction revenue rights. The second approach appears to be a much more workable way to accommodate a reallocation of the benefits of CRRs in response to shifts in load among LSEs. The only distinction between allocating auction revenue rights (ARRs) to LSEs, rather than allocating the actual CRRs is in the treatment after the allocation. The initial allocation of ARRs could look like the allocation of CRRs; the ARRs could be defined as source to sink rights and follow the CRR allocation procedure proposed by the CAISO to match a particular pattern of generation and load. However, once the auction occurs, the ARRs would define the allocation of the auction revenue but would not remain as actual CRRs in the hands of the party receiving the allocation, unless the party chooses to buy the CRR in the auction. Thus a party can choose to buy the CRR corresponding to its ARR, or can choose to sell it and buy a different CRR that better matches its load/generation portfolio. This process would automatically allocate and reallocate money as load shifts between LSEs and would substantially reduce the administrative burden of tracking shifts in CRRs between LSEs.

An additional consequence of rules providing for the allocation of CRRs to LSEs, rather than CRR economic values, is that they limit the duration of forward auctions for CRRs. If CRRs must be retained by the CAISO for allocation to LSEs on an annual, biannual or monthly basis, then they cannot be sold in multi-year auctions. This limitation would be avoided if LSEs were allocated the financial proceeds of the auction (i.e., ARRs), rather than CRRs. In this case, long-term CRRs could be sold in forward auctions and the revenue could be held in escrow and later distributed to LSEs through the CAISO settlements based on the load distribution at a future point in time.

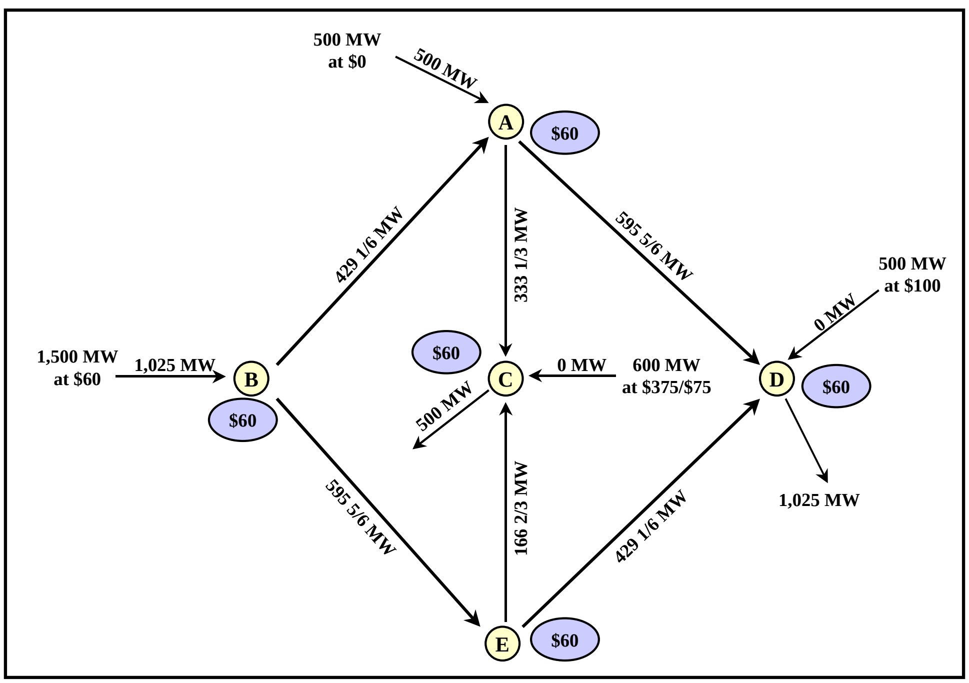
While it is likely reasonable to limit the duration of the CRRs sold in the initial CRR auctions, since market participants may not have a good basis for assessing CRR market values until an LMP market is actually in operation, it would be helpful in the long-run to build a system that could allow market participants to acquire long-term source to sink CRRs in an auction. LSEs seeking to enter into long-term contracts to hedge their energy cost may in the future want to have the opportunity to acquire a long-term hedge of the congestion costs associated with energy deliveries under that contract. If there is no possibility of offering a portion of the CRRs through a long-term auction, LSEs will have no means of acquiring such a long-term hedge other than by contracting with local generation.

Appendix VIII
Pass 2 Mitigation Structure

The example in Appendix VIII illustrates the potential for the peculiar structure of the local market power mitigation pass to undermine the effectiveness of the market power mitigation. For clarity, that Appendix portrayed the application of the local market power mitigation mechanism to RMR units. This appendix illustrates additional potential limitations of the MRTU local market power mitigation mechanism if there are generators possessing material local market power that are not subject to RMR contracts and that are intended to be subjected to market power mitigation in Pass 2, based on a PJM-style mitigation system.[1]

This example is based on the same transmission grid and bids employed in Appendix VIII. As before, Figure VIII-1 illustrates the Pass 1 dispatch in which generation would be dispatched without regard to local transmission constraints. It is again assumed that the limits (in lines A-D and A-C) are local transmission constraints. If the system is dispatched without regard to the constraints, all demand can be met at a price of $60, with 500 MW dispatched at A and 1,000 MW at B. It is assumed that the generation at C is bid into the market at $375 in Pass 1A and mitigated to $75 in Pass 1B. Since the LMP price at C is $60 in Passes 1A and 1B, the generation would not be dispatched in either pass and thus would not be subject to mitigation in this pass.

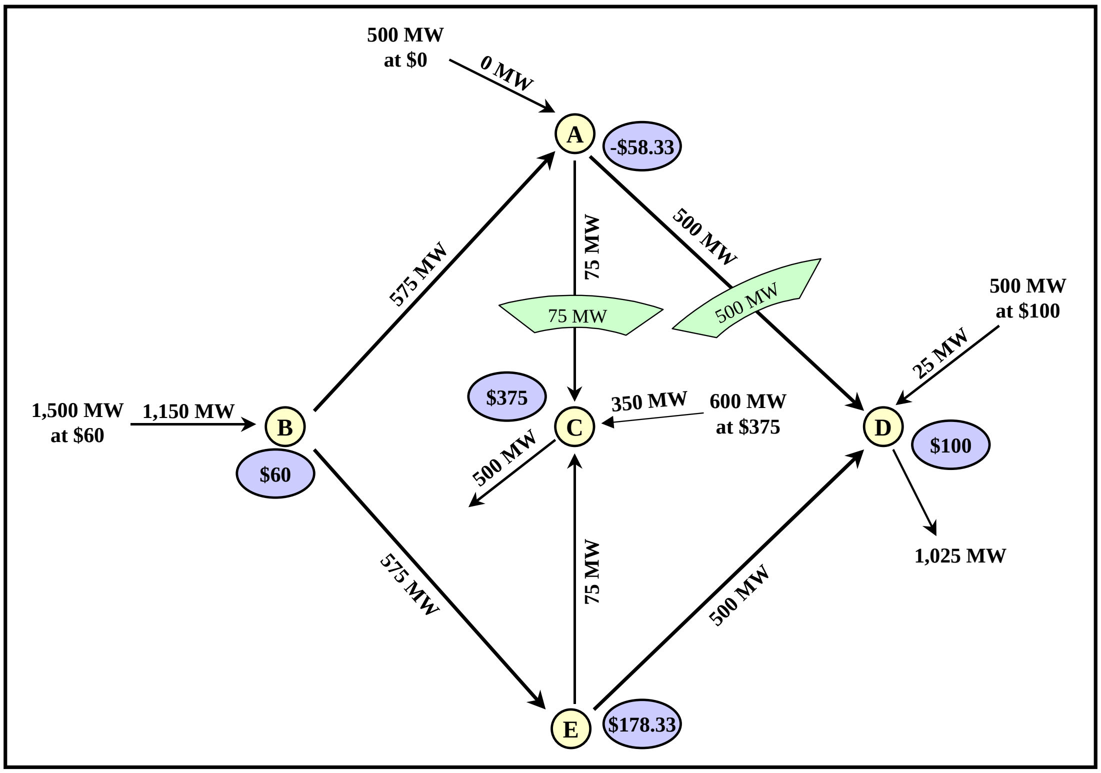
Figure VIII-1
PASS 1A, 1B DISPATCH



[1] Under a PJM-style mitigation system, units subject to mitigation are first dispatched based on their unmitigated offer prices and then subjected to mitigation on the portion of their offer curve that is dispatched.

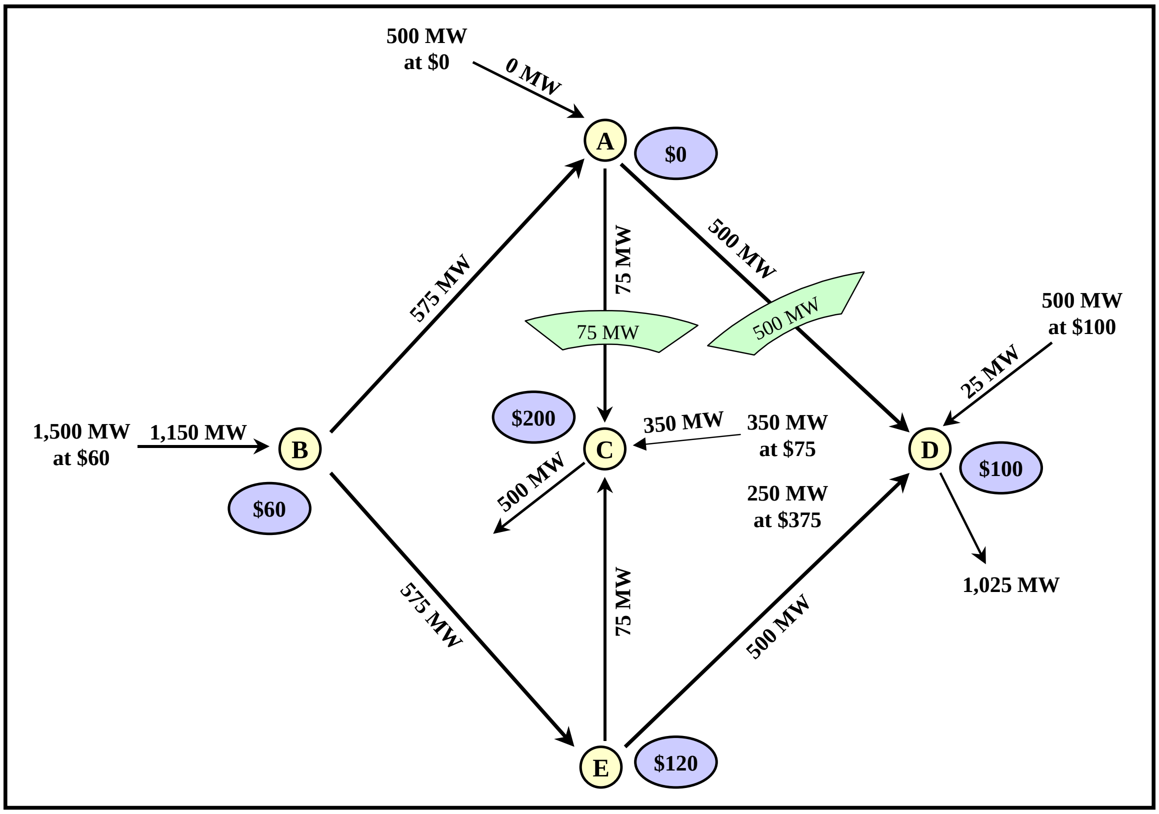
For the purpose of the example in this appendix, we will assume that the methodology for Pass 2 is changed so that the Pass 2 dispatch is based on the actual offer prices of the generation dispatched in Pass 1, rather than having those schedules set to -$1,000/MWh. In this example, however, we assume that the generation at C is not RMR generation. It is therefore offered into Pass 2 at its unmitigated Pass 1 offer price, rather than at an RMR price. With this change, the Pass 2 dispatch is as shown in Figure VIII-2. It can be seen that the generator is dispatched for 350 MW in Pass 2 and this amount of its capacity would therefore be subject to mitigation in Pass 3.

Figure VIII-2
PASS 2 DISPATCH



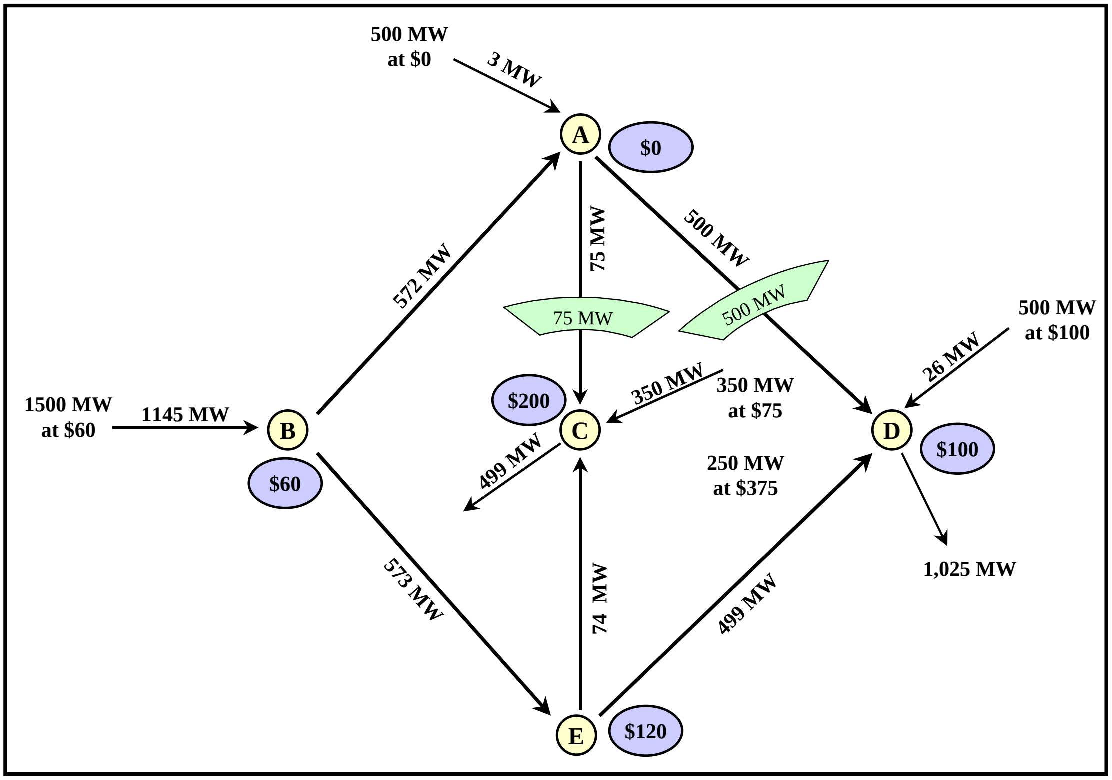
In Pass 3, the offer prices of the generator at C would be mitigated to $75/MWh for the first 350 MW of output, but the remainder of the unit's output would be available only at $375/MWh. More important, the 350 MW output at C reflects the tradeoff between generation at A, B, C and D for a $375/MWh price at C. The price of power would be $200 at C in Pass 3, as shown in Figure VIII-3, even though 350 MW of power were offered at $75/MWh.

Figure VIII-3
PASS 3 DISPATCH



This is illustrated in Figure VIII-4 which portrays the dispatch to meet only 499 MW of load at C. The total cost of meeting load falls by $200: -5 MW at $60 at B (-$300) +1 MW at $100 at D (+100), +3 MW at $0 at A (+$0). So even though 350 MW of power is offered at C at the mitigated price in Pass 3, the price at C is $200 because generation at C is not marginal and the price at C is determined by the tradeoffs with other alternatives. The marginal cost of those alternatives was determined in Pass 2 based on a $375/MWh offer price for generation at C.

Figure VIII-4
COST OF INCREMENTAL LOAD



The proposed mitigation procedure also can result in a large difference between the cost of meeting the last MW of load ($200 to meet the 500th MW at C in Figure VIII-4) and the next MW of load ($375 to meet the 501st MW at C).