COMMENTS ON THE FERC SMD NOPR
Docket RM01 12 000
November 15, 2002

By Shmuel Oren
University of California at Berkeley[1]

Executive Summary

The SMD NOPR is on the right track in terms of its overall objective and attempt to institute a sound science based approach to electricity market design. However, details matter and the objective of this note is to highlight three specific details in the SMD NOPR that require attention. The comment identifies potential problems associated with the specific design features and recommends remedies that will improve the design while keeping with the stated objectives of the NOPR.. I will focus on the following three aspects of the SMD

1. Virtual bidding in the day ahead market.
2. Management of congestion across ITP seams
3. Providing explicit economic signals for transmission investment.

Virtual bidding in the day ahead market provides a useful function in facilitating price convergence between the day ahead and real time markets. It also helps operation by enabling day ahead scheduling of transactions for parties that prefer real time settlements. However, virtual bidding is conducive to gaming of congestion when such bids are allowed to set the Day ahead LMP. The proposed remedies are to exclude virtual bids from the calculation of day ahead LMP and to use real time LMP for calculation of congestion charges imposed on bilateral transactions and for CRR settlements.

While congestion pricing and CRRs in the form of point to point or flowgate contracts are adequate instruments for allocating the use of transmission resources within a control area of an ITP, such mechanisms are insufficient for addressing loop flow problems across ITP seams. The problem is that price mechanism alone cannot guarantee that transactions scheduled in one ITP will not violate transmission constraints in the control area of an adjacent ITP. Physical flowgate rights can solve such seams problems by allowing transactions in one ITP control area to reserve capacity for loop flow that they induce in adjacent ITPs control areas. Such physical rights will only be available for cross ITP congestion management (replacing current TLR protocols). Pricing the physical flowgate rights at par with their financial counterparts ensures efficient allocation of transmission resources.

While LMP provide economic signals for efficient location of generation and load and for efficient dispatch of generation resources they do not provide adequate price signals for transmission investment. Flowgate pricing provides a more direct and explicit price

[1] This comment does not represent the position of the University of California or of any other entity with whom the author is affiliated.

signal for the value of transmission capacity enhancement. A congestion management and CRR system that accommodates both poin to point and flowgate rights and allows transmission companies to underwrite some of the CRR risk (by offering, for instance virtual flowgate capacity to the ITP) will go a long way toward producing explicit price signals and incentives for investment in the transmission infrastructure.

1. Preventing gaming through virtual bidding in the day ahead market.

The FERC SMD NOPR envisions a financially binding and physically feasible centralized day ahead market operated by the ITP. The physical feasibility of the day ahead schedules is intended to avoid situations that prevailed in the California PX where day ahead prices were based on schedules that ignored some of the known physical constraints and that created gaming opportunities such as the infamous DEC game. According to the SMD NOPR the day ahead security constrained bid based market clearing will result in day ahead LMP. These locational prices will be used for settlement of day ahead energy transactions, congestion charges imposed on bilateral transactions requiring transmission service, and settlements of CRRs. Both congestion charges and CRR settlements will be based on nodal price differences. The SMD also proposes that virtual bidding be allowed in the day ahead market. In other words, market participants will be allowed to submit bids for injection and withdrawal of power in the day ahead market that are not backed up by generation resources or load. The virtual bids are included in the day ahead optimal dispatch calculation and are indistinguishable from physical bids with respect to the determination of the day ahead LMP, meeting physical constraints and being subject to congestion charges. Such virtual bids represent pure financial obligations which are automatically offset by reverse transactions in real time resulting from non-execution of the day ahead virtual transaction. For example a 100MWh virtual injection bid in the day ahead market which is not delivered in real time will result in a negative generation imbalance of 100MWh in real time at the same location. The net effect of such a transaction is that the virtual bidder produces no energy but is credited or charged 100*(Day ahead LMP – Real time LMP) at the specific node. Similarly a virtual withdrawal bid of 100MWh in the day ahead will be offset by a negative load imbalance of 100MWh in real time and the bidder will be charged or credited 100*(Day ahead LMP – Real Time LMP) at the specific node. Such virtual bidding permits hedging or speculative arbitrage between the Day ahead LMP and Real time LMP when market participants have reason to expect a difference. Virtual bidding also allows market participants who choose to settle their transactions at real time prices to do so while still scheduling their transactions in the day ahead. The later can be accomplished by scheduling a physical transaction in the day ahead and at the same time scheduling a reverse virtual day ahead transaction. The two transactions cancel each other financially in the day ahead settlement whereas the non-execution of the virtual transaction in real time has the effect of pushing the settlement of the original physical transaction to real time. The arbitrage possibilities enabled by virtual bidding are useful mechanisms for improving efficiency in the energy markets and facilitating price convergence. Furthermore, the flexibility given to market participants to choose whether they want to settle their transactions at day ahead or real time prices is desirable from a systems operations perspective since it facilitates early disclosures of realistic schedules.

Unfortunately virtual bidding as envisioned in the SMD creates incentive for gaming which will distort price signals for transmission. The following will illustrate two such gaming possibilities

1. *Creating virtual congestion to increase CRR value.*

To illustrate such potential gaming consider a two node system with 900 MW inelastic demand and a 300 MW generator offering its power at $80/MWh at Node 1 and several generators with total capacity of 1200MW at Node 2 that offer their power at $40/MWh. Nodes 1 and 2 are connected by a transmission line with rated capacity of 1000MW. The CRRs for that capacity were auctioned off or grandfathered to existing users as hedges against congestion charges. Suppose that a market participant (generator, load serving entity or a speculator) acquired 400MW of CRRs. Given the assumed schedule the full load at Node 1 can be served by import generation from Node 2 and the LMP at both nodes will be $40/MWh in the day ahead as well as in real time. There is no congestion charge and the CRRs are worthless. Suppose, however, that the owner of the 400MW CRRs submits a virtual offer to inject 200MWh at $20/MWh at Node 2 and a matching virtual bid of 200MWh of load at Node 1. These virtual bids when added to the physical bids in the day ahead market will congest the transmission line. The resulting transmission constrained optimal dispatch will deploy the 200MWh virtual bid at Node 2 at $20/MWh, another 800MWh from the actual generators at Node 2 at $40/MWh and the remaining $100MW of generation at Node 2 at $80/MWh to serve a total of 1100MWh of real and virtual load at Node 1. The corresponding day ahead LMP will be $40/MWh at Node 2 and $80/MWh at Node 1. Consequently, the day ahead congestion charge for bilateral transactions from Node 2 to 1 is the nodal price difference of $40/MWh and the payoff to each MW of CRR is $40 per hour. The virtual bidder will receive in the day ahead 200*40 = $8000 for the virtual injection at Node 2 and will be charged 200*80=$16000 for the virtual load at Node 1. The difference of $8000 reflects the congestion charge for the virtual schedule of 200 MW from Node 2 to Node 1. The 400 CRRs held by the virtual bidder will be credited 400*40=$16000 resulting in a net gain of $8000.

In real time the virtual schedule is not executed so the virtual bidder is liable for a generation deficiency of 200MWh at Node 2 and a load deficiency of 200MWh at Node 1. These imbalances, however, decongest the transmission line so that the real time LMP at both nodes is set by the Node 2 generators to $40/MWh. Consequently the imbalance settlement for the virtual bidder is a wash and she gets to keep the $8000 in CRR revenues which are paid by those that were not fully hedged against congestion cost.

2. *Creating virtual counterflow to reduce congestion charges*.

Consider again a two node system as before but now the transmission line capacity is only 800 MW. Consequently the 900MW load at Node 1 will be served by 800 MWh imported from Node 2 at $40/MWh wile the balance of 100 MWh will be supplied locally at $80/MWh. The corresponding LMP are $40/MWh at Node 2 and $80/MWh at Node 1, whereas the congestion charge is the LMP difference of $40/MW/h. suppose now that 300MW of the load at Node 1 is served through a bilateral contract scheduled by a trader who purchases the power at Node 2 and resale it at Node 1. Assuming that the trader has no CRR's he will incur a congestion charge of 300*40=$12000/h. However, the trader can decongest the transmission line in the day ahead through virtual bidding of 100MWh injection at Node 1 at $35/MWh and a matching virtual load bid of 100MWh at Node 2. These virtual bids create virtual counterflow that decongests the transmission line and equalizes the LMP at both nodes to $40/MWh. The congestion charge goes to zero while the credit and charge for the virtual injection and withdrawal cancel out since the LMPs are the same at both nodes. The net result is a savings of $12,000 in day ahead congestion charges.

In real time the virtual transactions do not execute resulting in the real time LMPs reverting to $40/MWh at Node 2 and $80/MWh at Node 1. The trader will have to settle a 100MWh negative generation imbalance at Node 1 at a cost of 100*80=$8000 and a negative load imbalances of 100 MWh at Node 2 which earns him a 100*40=$4000 resulting in a net cost of $4000. In total the trader has saved 12,000 – 4000 = $8000 of congestion charges

One may argue that the two gaming possibilities described above will cancel each other since those who are not hedged against congestion charges will have the incentive to decongest the system with virtual bids while those who are overhedged will try to congest the system in order to increase the value of their CRRs. Unfortunately, such a scenario does not result in an equilibrium. In the case of arbitrage between day ahead and real time energy prices at a specific node the arbitrage will result in convergence of the two prices to a single price. In the case of congestion prices, however, the line is either congested or not and the congestion charge jumps discontinuously between zero and the difference between the prices of the last imported MWh and the cheapest local MWh (in our example ($40/MWh). Consequently, players in a repeated market will be induced to adopt mixed strategies that increase uncertainty in the relationship between the day ahead and real time markets. This defeats one of the primary objectives the of the day ahead market which is intended to give the system operator reliable advanced information that will enable better prediction of real time generation and load.

At PJM, where virtual biding is allowed, gaming of the first type described above is, apparently, suppressed through an administrative rule that requires FTR holders to forfeit their day ahead revenues if it turns out that the day ahead congestion upon which these revenues were based did not materialize in real time[2]. Such an approach is heavy handed with unclear consequences with respect to equity, competitiveness and correct price signals.

The straight forward approach to eliminate the type of gaming described above is to disallow virtual biding altogether. However, as discussed earlier virtual bidding does provide a useful mechanism for facilitating convergence between the day ahead LMP and real time LMP for energy trading. The following recommended remedies, would preserve

[2] Per Email communication from Andy Ott.

virtual bidding for energy trading but will neutralize the adverse impact of virtual congestion that can be created through balanced virtual schedules resembling ENRON's strategies in California.

1. **Exclude virtual bids from the calculation of the Day ahead LMP**.
In this framework virtual bids become price takers of the day ahead LMPs and are able to arbitrage the difference between the day ahead LMP and real time LMP but they cannot affect these prices. Of course this may result in virtual schedules that appear to violate the transmission constraints but that should not matter as long as the physical schedules satisfy the constraints. In implementing such an approach it is important to enforce strict limits on allowed deviations from physical schedules and impose proper penalties that will deter such deviations.

2. **Ex-post settlement of congestion charges and of CRRs based on real time LMP.**
This remedy amounts to having a single settlement for transmission and two settlements for energy. Under such a procedure a balanced virtual bid that attempts to arbitrage transmission charges between day ahead and real time will net zero whereas virtual bids at any particular location still enable arbitrage between the day ahead and real time LMP at that location. In the examples discussed earlier neither of the attempts to increase the payoff of CRRs by creating virtual day ahead congestion or reduce congestion payments by eliminating day ahead congestion will work if both the CRRs and the congestion charges are based on real time LMP. Bilateral transactions covered by CRRs will still be fully hedged. Furthermore CRRs can still be used for hedging day ahead market balanced transactions by employing a reverse virtual transaction which effectively pushes the settlement of the day ahead transaction to the real time market (as described earlier) and using the CRR to hedge the congestion cost in the resulting real time settlement.

2. Congestion management across ITP seams

The SMD NOPR envisions ITPs of sufficient scope to internalize all the externalities associated with the transmission system . In such an idealized ITP all constraints will be accounted for in the dispatch and all transmission right can be financial. In such an ITP all loop flows are internal to the ITP and the security constraint bid based dispatch assure that no constraints will be violated. Unfortunately technical and jurisdictional considerations makes such ITP utopia a remote reality and even under best case scenarios for the SMD we should expect the formation of multiple ITPs with seams and cross seams loop flows that must be coordinated. Current transmission load relief (TLR) protocols are economically inefficient in dealing with seams problems. The question is how to pass correct price signals among ITPs that will enable one ITP to correctly account for the impact of its loop flow on the transmission system of neighboring ITPs. In a pure DC system such impact might be handled through ex post financial settlements charging each ITP for the congestion created in neighboring ITPs. exchange of day ahead LMPs would allow each ITP to take such impact into consideration in its own security constrained dispatch. In such a scheme each ITP only accounts explicitly for constraints on transmission facilities under its own control and “prices out” the cost of congestion it creates in neighboring ITPs. If that scheme could work each ITP could hedge its out of area congestion cost by purchasing financial transmission rights, such as flowgate rights on elements impacted by its loop flow in other ITP control areas.

Unfortunately, as it was demonstrated by Oren and Ross in the attached paper, in an AC optimal power flow (OPF), even with perfect foresight of nodal or flowgate prices in the neighboring ITP control areas, the pricing out approach may lead to violation of flow limits in neighboring ITPs. Consequently, financial settlements across ITP boundaries are inadequate for preventing across the border constraints violation. From a technical perspective the problem arises due to the fact that when constraints are relaxed and added to the objective function of the OPF by imposing a cost on violation, the resulting objective function (i.e., the partial Lagrangian) may not be locally convex at the constraint solution even when the correct Lagrange multipliers (i.e., shadow prices on the relaxed out of area constraints) are used. Consequently, the relaxed optimal power flow (OPF), which, is to be solved by an ITP who prices but does not monitor the impact of its transactions on out of area constraints, may result in the violation of the relaxed constraints. Furthermore, the solution tend to be highly sensitive to small error in the prices imposed on the relaxed constraints unless the constraints are explicitly accounted for.

The above difficulty can be resolved through a combination of flow-based financial and physical transmission rights and financial point to point rights. Under such a system each ITP can issue financial transmission rights for hedging congestion cost on facilities within the ITP control area and in addition offer physical flowgate rights to neighboring ITPs who must cover all their out of area loop flows with such physical rights. The total of financial and physical flowgate rights and the intra ITP point to point financial rights must jointly meet simultaneous feasibility constraints. The pricing of the physical flowgate rights should be the same as for the financial flowgate rights. The difference between the financial and physical flowgate rights is that the financial rights only entitles the holder to the congestion revenues associated with the particular flowgate whereas the physical counterpart entitles the holder to the flowgate capacity. Consequently each ITP will deduct the quantity of physical rights sold on each flowgate from the available capacity used for the intra ITP dispatch and will augment the intra ITP constraints with constraints on the loop flow in neighboring ITPs not to exceed the acquired physical rights on the impacted lines. The procurement of physical flowgate rights across ITP boundaries can be done by individual market participants whose transactions create cross boundary loop flow or by each ITP on behalf of the market participants under its jurisdiction.

3. LMP do not provide adequate price signals for transmission investment

When generators offers reflect their short term marginal costs, the locational marginal prices resulting from a security constraints bid based optimal dispatch reflect short term marginal costs of delivering power at each node in the system. They provide correct incentives for short term operating and consumption decisions to generators and loads that can respond on a short time scale. In theory, if generators and loads can contract efficiently around the LMPs and properly hedge the congestion costs associated with implementing such bilateral contracts the LMP will lead to efficient long term price signals for generation investment and load location. The theoretical argument underlying the relevance of short term marginal pricing to long term investment is based on the premise that inframarginal competitive rents resulting from uniform market clearing prices in excess of marginal generation costs will finance startups and capacity payments that may not be reflected in short term cost based bids. The FERC SMD NOPR already recognized that short term energy prices do not provide adequate incentives for planning reserves and that additional measures such as capacity obligations are necessary to assure generation adequacy.

The SMD NOPR is deficient, however, in addressing the need for price signals that will provide incentives for transmission investment. LMP based marginal transmission pricing between any two nodes equate the short run marginal cost of transmission to the bid based marginal cost of counterflow that relieves congestion between the two points. While such a price signal may be adequate for short term redispatch decision and properly compensates generators that provide congestion relief it is totally inadequate for transmission investment decisions for several reasons.

- The LMP do not provide an explicit price signal that identifies the scarce transmission resources. While LMPs can be derived from the shadow prices on congested elements the opposite is not true. Even if we know all the nodal prices it is not possible to infer where the bottlenecks are without the explicit information on the shadow prices.
- Without direct trading of physical capacity rights or financial rights to flowgate capacity there is no explicit short term price signal for transmission capacity. The short term marginal value of transmission is provided as a financial derivative (locational swap) of energy prices.
- Even with explicit short term price signals for transmission capacity it is questionable whether such price signals are of any value for transmission investment.
- Marginal prices for transmission that are based on LMP or flowgate shadow prices are derived from scarcity. Such prices are useful as usage charges to induce efficient utilization of scarce resources, however they are not useful as investment incentives. In other words rewarding transmission with financial rights to scarcity rents is inadequate since by nature investment in transmission reduces or eliminates the scarcity rents feeding such financial rights. If transmission lines could be expanded incrementally, optimal expansion of line capacity would be such that the marginal expansion cost equals the expected scarcity rent as reflected by the shadow prices. Realistically, however, transmission investments are lumpy so market based signals for such investment that rely on LMP are unlikely to produce the desired outcome since such investment will eliminate the revenue stream that signaled the need for investment.

In sum the lack of explicit price signals identifying transmission investment needs, The fact that such signals are based on scarcity rents that the investment will eliminate and the lumpy nature of transmission investment makes LMP insufficient as incentives or price signals that will stimulate investment in the transmission system.

ATTACHMENT

Economic Congestion Relief Across Multiple Regions Requires Tradable Physical Flow-Gate Rights

by Shmuel S. Oren and Andrew M. Ross

Economic Congestion Relief Across Multiple Regions Requires Tradable Physical Flow-Gate Rights

Shmuel S. Oren, *Fellow, IEEE,* and Andrew M. Ross

***Abstract*—This paper is concerned with market-based protocols for relieving congestion caused by transactions outside the control area in which the congestion occurred. One approach, proposed by Cadwalader *et al.* is based on dual decomposition in which out of area congestion is "priced-out" and added to the optimal power flow (OPF) objective function of the control area operator while the prices are determined iteratively via nodal energy adjustment bids. The paper demonstrates through a simple three node example that even with "correct prices" on out-of-area congested interfaces, the augmented AC-OPF objective function of a control area operator might not be locally convex at the optimal solution and hence the control area's optimal dispatch may violate the thermal constraints on out-of-area interfaces. That conclusion supports the alternative "flow-based" approach that enforces thermal limits more directly, which is consistent with North American Electric Reliability Council's (NERC's) FLOWBAT proposal for interzonal transmission load relief (TLR).**

***Index Terms*—Duality, interconnected power systems, optimization methods, power distribution, power generation dispatch, power industry.**

I. Introduction

THE proliferation of competition in the electric power industry in the U.S. and around the world and the rapid growth of interregional trading of electric power require the development of procedures for coordinating congestion management across multiple control areas. U.S. Federal Energy Regulatory Commission (FERC) Order 2000, mandating the formation of regional transmission organizations, will undoubtedly accelerate interregional trading and increase the burden of interregional coordination of transmission use. The North American Electric Reliability Council (NERC) is in the process of implementing a multistage approach to transmission load relief (TLR) protocols aimed at keeping the use of the grid within secure capacity limits. There is general agreement that such protocols should promote efficiency and hence NERC envisions a gradual transition from the current approach of administrative curtailments to a market based approach [1]. However, the rate at which such a transition should occur and the ultimate market mechanism for interregional coordination are subject to debate.

There are two main contenders for market based TLR protocols. One approach, often referred to as the link-based approach, envisions a system of tradable physical transmission rights on the congestion prone links (flowgates). The impact of any bilateral transaction on these lines can be determined from the Power Transfer Distribution Factor (PTDF) matrix of the network. Under this scheme the TLR protocol requires that any transaction must be backed by flowgate rights in the amount of the flow it generate on each flowgate. The rights are traded in a market for transmission rights that operates in parallel with the energy market. Since counterflow on congested links enables more flow in the congested direction, counterflow should be regarded as virtual flowgate rights that can be traded as if they were physical rights. Revenues from the sales of counterflow rights will subsidize out of merit generation that helps relieving congestion. The theoretical foundation for this approach was developed by Chao and Peck [2] who have shown that the energy and transmission rights market will converge to the economic dispatch equilibrium and that the transmission rights prices on the congested links converge to the corresponding marginal values of additional capacity on these links under economic dispatch. Tabors [3] discusses an implementation scenario for a congestion management approach based on tradable physical rights that are initially awarded through an auction. An important feature of this approach is that feasibility of the flows is ensured by the number of physical rights issued for each line and the enforcement of the trading rules requiring that the appropriate rights back each transaction. The importance of this feature will become clear later in the paper. One should also point out that a rule reverting all unused rights to the system operator (within a reasonable time frame) would alleviate the classic concern about potential exercise of market power through withholding of rights. Alternatively, such withholding can be prevented by defining the transmission rights as financial rights with scheduling priority as the "Firm Transmission Rights" in California. Bundling financial rights with scheduling priority has the force of physical right but the scheduling priority can only be exercised through scheduling of a transaction and hence withholding is not possible. Further reasons to use the flow-gate approach are discussed in [4].

The other approach, due to Cadwalader *et al.* (CHHP) [5], is based on the nodal pricing paradigm in which transmission prices are calculated from nodal price differences for energy. The fundamental difference between this approach and the former is that here there is no direct trading of transmission rights and the price of transmission is determined by the energy market and adjusted through energy adjustment bids. The basic idea in this approach is to decompose the global optimal

Manuscript received November 11, 2000; revised July 17, 2001.
The authors are with the Industrial Engineering and Operations Research Department, University of California, Berkeley, CA 94720 USA (e-mail: oren@ieor.berkeley.edu; aross@ieor.berkeley.edu).
Publisher Item Identifier S 0885-8950(02)01018-0.

dispatch problem into subproblems corresponding to the different control areas. Each control area operator optimizes his dispatch by explicitly accounting for transmission constraints within his jurisdiction whereas constraints on lines he affects outside of his control area are "priced out" and accounted for as an added cost in his objective function. The prices for out-of-area transmission impact are exchanged among the control area operators through an iterative process in which energy prices and schedules are adjusted to account for the out-of-area transmission cost and each control area operator recalculates and reports the new transmission prices in his area based on the adjusted nodal prices. CHHP provide a very eloquent and thorough description of a market mechanism for implementing their proposed approach with a discussion of various practical implementation issues along with a detailed theoretical foundation. Unfortunately, as we shall see below, the CHHP approach has a basic theoretical flaw and if implemented without additional corrective measures it may result in dispatches that violate transmission constraints in neighboring regions. Such possible mishaps can occur when a control area operator uses an ac model (rather then a dc approximation) in optimizing the dispatch in its own control area. We present this effect in a simple three-node example under the assumption that the proposed market based TLR process finds the "correct" prices for transmission (i.e., the shadow prices that would results if the dispatch was optimized jointly). Our example demonstrates that even with correct prices for transmission the proposed TLR protocol can fail. One would expect that approximate prices that would result from a realistic adjustment process could make things worse.

II. Regional Decomposition Through Partial Dualization

The mathematical underpinning of the CHHP approach is the decomposition of the economic dispatch problem for the entire grid through a technique known as partial dualization or Lagrangian relaxation. The basic idea of this technique is to "price out" some of the constraints in a nonlinear constrained optimization problem and move them from the constraints set to the objective function. To illustrate this concept we consider a simple nonlinear optimization problem with two constraints

$$P1\colon \min_{\vec{x}} f(\vec{x})$$
$$\text{subject to: } g(\vec{x}) \le 0 \quad h(\vec{x}) \le 0.$$

Suppose that the vector $\vec{x}^*$ and corresponding Lagrange multipliers (shadow prices) λ_g^*, λ_h^* represent a local solution to $P1$, i.e., they satisfy first and second order optimality conditions. We can then formulate a new problem $P2$ in which the second constraint is relaxed and moved to the objective function ("dualized") as follows:

$$P2\colon \min_{\vec{x}} \{f(\vec{x}) + \lambda \cdot h(\vec{x})\}$$
$$\text{subject to: } g(\vec{x}) \le 0.$$

It can be easily shown that if $(\vec{x}^*, \lambda_g^*, \lambda_h^*)$ satisfy first order necessary (Karush/Kuhn-Tucker) conditions for $P1$ then $(\vec{x}^*, \lambda_g^*)$ will also satisfy these conditions for $P2$ when $\lambda = \lambda_h^*$. Thus, if second order conditions are also met then $(\vec{x}^*, \lambda_g^*)$ solve $P2$ for $\lambda = \lambda_h^*$. Satisfying second order conditions is assured when a dualized constraint is represented in a convex manner. In such happy circumstances it is possible to develop an iterative procedure that alternates between adjusting the multiplier λ in the objective function of $P2$ (dual iteration) and moving toward the solution $(\vec{x}^*, \lambda_g^*)$ (primal iteration), as CHHP proposes.

Unfortunately, when a dualized constraint is not represented in a convex manner, second order necessary conditions for $P2$ (even with the correct multiplier $\lambda = \lambda_h^*$) may not be satisfied by $(\vec{x}^*, \lambda_g^*)$. In such a case, $\vec{x}^*$ represents an inflection point of the objective function in $P2$.

CHHP have employed the partial dualization approach outlined above to decompose the total grid optimal power flow (OPF) problem into regional problems in which transmission constraints inside a region are considered explicitly while out-of-region constraints are priced (using market-based prices) and added to the regional objective function. This works if one assumes that all control area operators employ a DC approximation of the network in determining their own OPF and in representing out of area constraints. Indeed, the example used by CHHP is based on a DC model. However, their approach can be problematic if the control area operator employs more realistic AC-OPF models. This will be demonstrated in the subsequent sections by means of a simple three-node, two-region example.

CHHP argue that when the solution to first order conditions of the dualized problem is not a solution for the market equilibrium problem then "...The difficulties would extend beyond the mechanics of decomposition to call into question the existence of a competitive market equilibrium and might point to a greater role for more direct management of the grid and less reliance on markets."[1] It is important to realize, however, that the inadequacy of the partial dualization approach which we address is not inherent in the AC-OPF problem but rather in the decomposition approach and the associated market mechanism. To demonstrate this point it is useful to consider the alternative decomposition of the OPF problem that Kim and Baldick proposed [6]. In that approach the network is decomposed into subgrids by "cutting off" transmission lines along the borders, inserting dummy busses at each end of the cut, and adding coupling constraints forcing the cuts to match. The cuts are achieved mathematically by duplicating the border variables characterizing the line flow (i.e., real power, reactive power, voltage and phase). The coupling constraints forcing the border variables to match on both sides of a cut are dualized and this decomposes the total grid OPF problem into separate regional OPFs. The multipliers on the coupling constraints that enter the regional OPF objective functions are adjusted iteratively until the border variables on both sides of a cut match. CHHP highlight the conceptual similarity of their proposal to the Kim and Baldick approach. Furthermore, they argue that the computational success of the

[1]The possibility that partial dualization of transmission constraints in the AC case may result in market solutions that differ from the centralized OPF due to loss of local convexity in the dualized objective function was brought to the attention of the authors in a private communication but dismissed in a footnote which contained this quote.

Kim and Baldick algorithm supports their conjecture that their scheme should converge rapidly if price variables in other regions don't change much. However, the Kim and Baldick decomposition dualizes only linear constraints (simple equalities of the border variables); hence, adding the dualized constraints to the objective function does not affect its convexity. Then, first and second order optimality conditions for the dualized problem are satisfied by the solution to the original problem. Kim and Baldick use an augmented Lagrangian approach that adds to the dualized objective an additional quadratic penalty term on violations of the coupling constraint. Adding the penalty term is a standard method of convexifying the dualized objective function, thus preventing the problems discussed above and speeding up convergence [7]. Where Kim and Baldick augment the Lagrangian for linear equality constraints, the CHHP formulation uses nonlinear inequalities; the resulting problem does not lend itself to a distributed or market-oriented interpretation or implementation. Thus, we evaluate the CHHP system in its original form, without augmenting the Lagrangian. Reference [7] also proposes a convexification system for block-separable nonlinear inequality programs. Unfortunately, our problem is not block-separable, so we cannot take advantage of this procedure.

The remainder of this paper constructs a simple illustrative example that demonstrates in the context of a three node AC model how the CHHP decomposition scheme (which, as written, does not include an augmented Lagrangian) can lead to a severe violation of thermal constraints even when all out of area constraints are "priced" correctly.

III. AC Optimal Power Flow Model

We will use the power flow model from Wu *et al.* [8]. We have n nodes, and at each one we set a phase angle θ. The flow between node i and node j is $Y_{ij}\sin(\theta_i-\theta_j)$, where Y_{ij} is called the "admittance" of the line (if there is no (i,j) line we set $Y_{ij}=0$). The quantities q_i keep track of the net power flow at each node. If $q_i<0$ then node i is a net consumer of power; otherwise it is a net generator of power. Associated with each node there is a cost (benefit) curve $C_i(q_i)$ expressed as a function of net output that is convex and increasing. Negative cost represents consumption benefit. Our objective is to minimize the total cost. Each power line has a limited capacity M_{ij}. All power lines are symmetric. Our basic formulation is

$$\begin{aligned}&\underset{\vec{q},\vec{\theta}}{\text{minimize}}\ \sum_i C_i(q_i)\\&\text{subject to}\\&\forall i\text{:}\ \sum_j Y_{ij}\sin(\theta_i-\theta_j)=q_i\quad\text{(node flow)}\\&\forall i,j\text{:}\quad Y_{ij}\sin(\theta_i-\theta_j)\le M_{ij}\quad\text{(thermal).}\end{aligned}$$

This model ignores reactive power and resistive line losses, but it captures the essence of the planning problem. Because only the differences between phases are used, there is an extra degree of freedom, so we can choose one node's phase and set it equal to zero. We will find it more convenient to substitute out the q_i variables. We get this formulation

$$\begin{aligned}&\text{minimize}\sum_i C_i\left(\sum_j Y_{ij}\sin(\theta_i-\theta_j)\right)\\&\text{subject to}\\&\forall i,j\text{:}\quad Y_{ij}\sin(\theta_i-\theta_j)\le M_{ij}\quad\text{(thermal).}\end{aligned}\tag{1}$$

Denote the optimal solution to this problem $\vec{\theta}_A$, with Lagrange multipliers $\vec{\lambda}_A\le\vec{0}$. Even though the C_i are convex in their input q_i, here we see that the objective function is not globally convex in $\vec{\theta}$. Among other things, it has period 2π in each θ_i. However, it is locally convex near the feasible region, as pointed out by Chao and Peck [2]. The formulation (1) is equivalent to that used by CHHP, but they express the objective function and constraints in terms of the nodal injections/withdrawals q_i. While it is more intuitive to use q_i as decision variables, graphical illustration of the objective function and constraints set is more convenient in terms of θ_i, as we shall see below. There is no substantive difference between the two representations since the optimal injections/withdrawals can be readily calculate from the optimal phases through the relation $q_i=\sum_j Y_{ij}\sin(\theta_i-\theta_j)$.

A. Feasible Region

To visualize the feasible region for the power flow problem in the phase variables (θ_i), we will use a three-node network as in Fig. 1. As mentioned above, we can set one of the phase angles to zero, so we will make $\theta_3=0$. We also take arcsines on each side of the thermal constraints

$$|\theta_i-\theta_j|\le\arcsin(M_{ij}/Y_{ij}).$$

In Fig. 2, we plot the feasible regions represented by these constraints in the (θ_1,θ_2) plane. These combine to give a feasible region that will often be an irregular hexagon, but could be an irregular quadrilateral if one of the constraints is never binding. This shape repeats every 2π units in each direction, but we will ignore those other regions since they will result in the same power flows. For a network with n nodes, the feasible region is a polyhedral solid in $n-1$ dimensions with faces that are at 45 and 90° angles to the coordinate axes.

B. Partial Dualization

In order to solve this problem across multiple regions, we partition the tie lines into two sets, which we call Mine and Other. We explicitly monitor the thermal constraints on the lines that are Mine, while we relax the constraints on the Other lines using the CHHP partial dualization approach. The new formulation is

$$\begin{aligned}&\text{minimize}\sum_{\forall i} C_i\left(\sum_{\forall j} Y_{ij}\sin(\theta_i-\theta_j)\right)\\&\quad-\sum_{(i,j)\in\text{Other}}(\vec{\lambda}_A)_{ij}\cdot(Y_{ij}\sin(\theta_i-\theta_j)-M_{ij})\\&\text{subject to}\\&\forall(i,j)\in\text{Mine:}\quad Y_{ij}\sin(\theta_i-\theta_j)\le M_{ij}\quad\text{(thermal)}\end{aligned}\tag{2}$$

As shown, we know that any solution to (1) is a first-order stationary point of (2). But, a minimum point of (2) might not match the optimal point for (1), since the dualized constraints are not in a convex form. This difficulty is not inherent in the problem but rather in the implementation of partial dualization. For instance, the difficulty would disappear if instead of dualizing the constraints on line flow (following CHHP) we would dualize the equivalent constraints on the phases θ_i which do represent convex sets. However, that would change the interpretation of the Lagrange multipliers to prices on phase angles rather than on flows and this interpretation would not lend itself to a market-based implementation of the dual iteration.

IV. Example

To illustrate the potential problems with partial dualization, we present a small three-node example in Fig. 1. We first consider the centralized OPF for the entire network. The two degrees of freedom in this problem are the phases θ_1 and θ_2 (in radians). Fig. 3 shows the level sets of the cost function and the feasible region for the two phase variables described by the inner hexagonal region. The optimal solution is at the lower left corner of the rectangular constraint, i.e., the smallest feasible values of θ_1 and θ_2. The corresponding optimal injections (withdrawals) and resulting line flows are shown on the network diagram in Fig. 4. We now calculate the Lagrange multipliers corresponding to the optimal point and dualize the constraints for the (1, 3) and (2, 3) lines. This bypasses the iterative market process needed to converge on the correct price for transmission and assumes that the correct prices have been found. Pricing out the two power lines removes the rectangular constraints from the previous picture, leaving the diagonal strip as the feasible region. Fig. 5 shows the level sets of the objective function and the relaxed constraint set in the dualized problem. Since the optimal solution to the original problem is interior to the remaining constraint, it is a stationary point of the dualized objective function as expected. However, this is an inflection point (which can be identified by the cusp in the level sets) rather than a local minimum. The minimum of the dualized problem, on the other hand, has moved outside the old rectangle to the point shown by the arrow where the constraint boundary is tangent to the level set of the dualized objective function. This means that if the operator has no information about the capacity of the two transmission lines that are not under his control and is only given price information about the marginal cost of using these lines, he will choose to operate at the new optimal point that is outside the feasible region prescribed by the priced-out constraints. The implication of the new optimum (i.e., the partially dualized solution) with respect to injections (withdrawals) and flows are given on the network diagram in Fig. 6. The flows on lines (1, 3) and (2, 3) exceed their thermal limits by 60% and 92%, respectively.

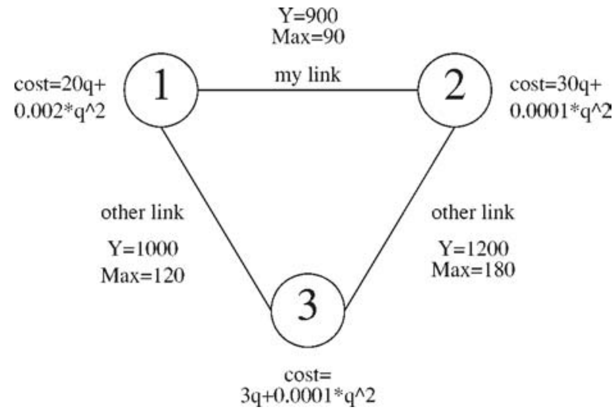

Fig. 1. Particular network for our counterexample.

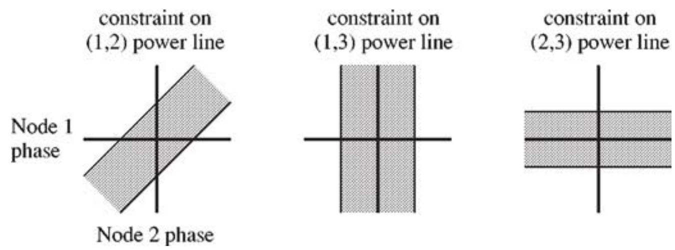

Fig. 2. Feasible regions for the three thermal constraints.

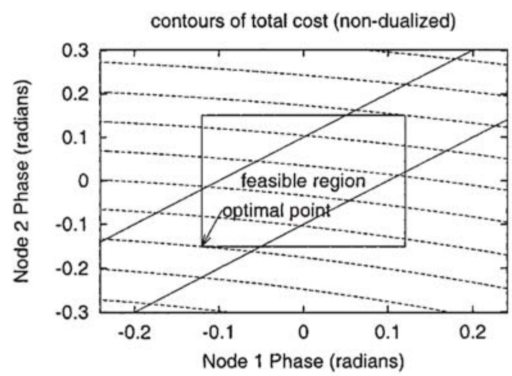

Fig. 3. Feasible region and contours of the true cost function.

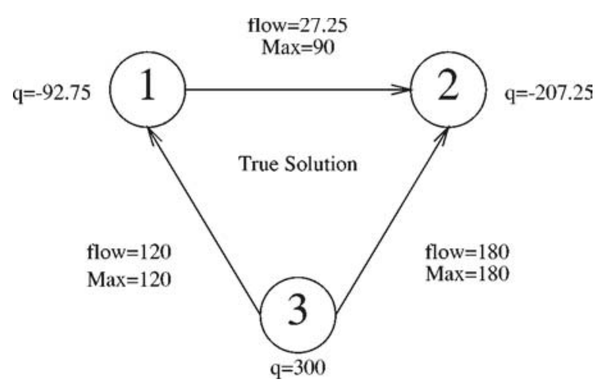

Fig. 4. Optimal flow for our example.

V. Price Adjustment

In deriving the optimal solution to the partially dualized (relaxed) optimization problem we have used the "correct" Lagrange multipliers based on the optimal solution to (1). Nevertheless, one may still wonder whether further adjustments to these Lagrange multipliers can restore feasibility of the relaxed solution with respect to the violated thermal constraints. Such adjustments can be implemented in a market-based framework through two means. One is a direct adjustment to the link-based

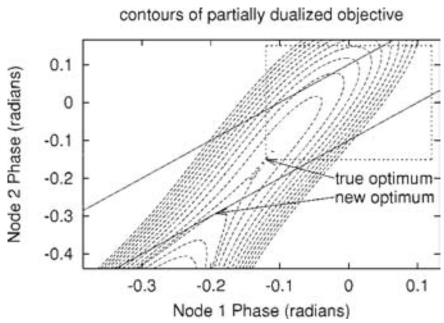

Fig. 5. Contours of the partially dualized cost.

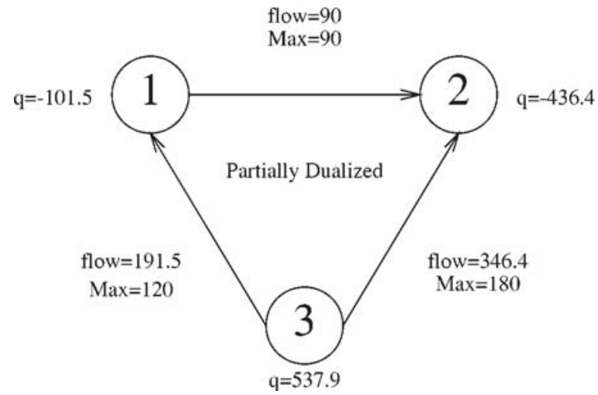

Fig. 6. "Optimal" (but actually infeasible) flows for the partially dualized problem.

prices, which equal the Lagrange multipliers corresponding to the violated thermal constraints. The other is energy adjustment bids at the nodes, as proposed by CHHP. The energy adjustment bids can be obtained through a linear transformation of the adjustment to the Lagrange multipliers using the power transfer distribution factors (PTDF) corresponding to the local linearization of the constraints. Hence, without loss of generality we may examine the effect of direct adjustment to the Lagrange multipliers. In the context of Lagrangian relaxation approaches to optimization, which is what the CHHP proposal is trying to emulate through market processes, Lagrange multipliers corresponding to relaxed constraints are commonly adjusted by means of subgradient steps. A subgradient step amounts to incrementing each of the Lagrange multipliers corresponding to a violated relaxed constraint by an amount that is proportional to the violation. In mathematical terms the Lagrange multipliers are adjusted via the iteration

$$\vec{\lambda}' = \vec{\lambda} + \alpha \cdot \vec{g}^{+}(\vec{x})$$

where $\vec{g}^{+}$ is the violation of the constraints (or zero if a constraint is not violated), evaluated at the current approximation to the solution—which in our case is the false optimum obtained from doing partial dualization with the correct Lagrange multipliers. In our example, since the constraint violations are large, the step size α can be very small and still have a noticeable effect.

In our example, only λ_{31} and λ_{32} are nonzero and so are the corresponding adjustments. Table I gives the values of the adjusted Lagrange multipliers corresponding to several values of α. The graphs in Figs. 7–10 illustrate the level sets of the dualized objective function resulting from using the modified Lagrange multipliers in pricing out the out-of-area thermal constraints. We also show in Table I the corresponding percentage changes to the Lagrange multipliers.

Note that the largest change is under one percent. Since, as indicated before, the corresponding adjustments to nodal energy prices are just linear transformations of the Lagrange multipliers adjustments the adjustments to the nodal energy prices will also be under 1%. The graphs illustrate that the dualized objective function is highly sensitive to changes in the Lagrange multipliers, and small adjustments to these multipliers (or equivalent nodal energy adjustment bids) will radically alter the shape of the objective function and the location of the optimal solution to the relaxed problem. In view of these sensitivities, even a computer implementation of the

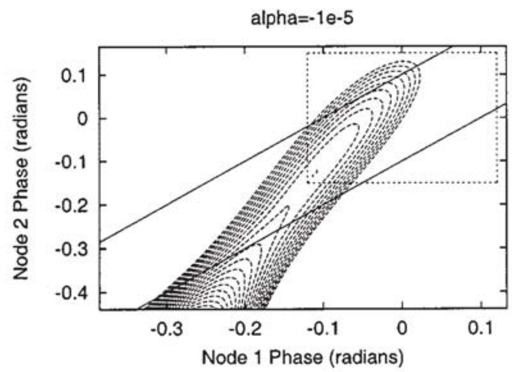

Fig. 7. Contours for $\alpha = -10^{-5}$ are almost unchanged from $\alpha = 0$. A local optimum is barely visible.

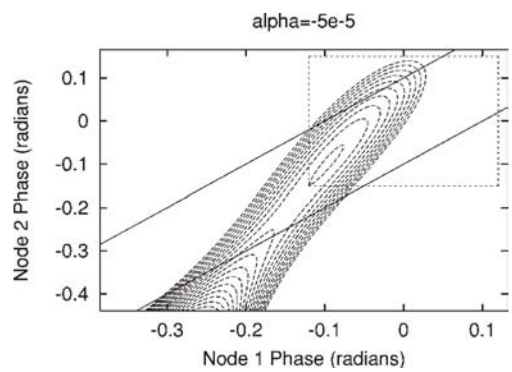

Fig. 8. Contours for $\alpha = -5 \cdot 10^{-5}$. Now a feasible local optimum is visible.

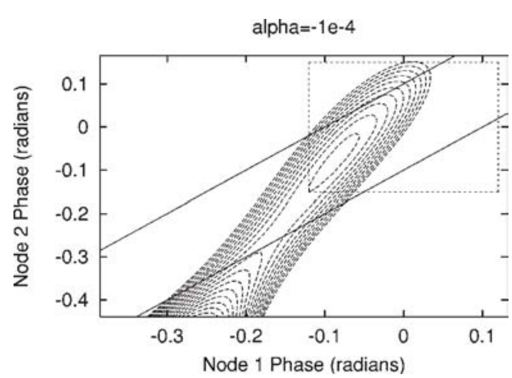

Fig. 9. Contours for $\alpha = -10^{-4}$. Optimum has moved away from true optimum.

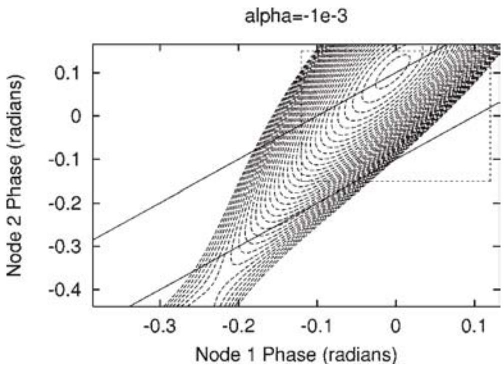

Fig. 10. Contours for $\alpha = -10^{-3}$. Optimum has moved much farther away from true optimum.

TABLE I
PERCENT CHANGE IN λ FOR VARIOUS VALUES OF PENALTY α

α	$\bar{\lambda}'$		percent change
0	λ_{31}	-7.2090	-
	λ_{32}	-34.7308	-
-10^{-5}	λ_{31}	-7.2097	0.00991
	λ_{32}	-34.7325	0.00479
$-5 \cdot 10^{-5}$	λ_{31}	-7.2126	0.0496
	λ_{32}	-34.7391	0.0240
-10^{-4}	λ_{31}	-7.2161	0.0991
	λ_{32}	-34.7474	0.0479
-10^{-3}	λ_{31}	-7.2805	0.991
	λ_{32}	-34.8972	0.479

Lagrangian relaxation approach with high numerical precision may face difficulties converging to a feasible optimal solution of the original problem. Achieving such convergence through a market-based implementation of the price adjustments is unconscionable.

VI. CONCLUSION

It is by now widely recognized that the physical realities of electric power systems limit the variety of market designs that will support decentralized operational paradigms and competition while respecting the secure operational limit of the system. Just as loop flow has been recognized as a phenomenon that must be contended with in the design of market-based congestion management protocols, nonlinearities due to the AC characteristics of the system must be recognized in the design of interregional TLR approaches. We have demonstrated that simple relaxation and pricing of thermal limits on out of region transmission lines can result in violation of such constraints. Furthermore, the solution is highly sensitive to the prices, so any attempt to correct such infeasiblities via market-based price adjustments will result in unpredictable outcomes. These observations raise serious questions about the viability of mechanisms that treat transmission constraints indirectly through adjustments in the energy market. This is a direct consequence of the basic conclusion indicated by our examples which is that price mechanisms without some form of direct quantity controls are insufficient to ensure feasible usage of transmission lines. By contrast, market designs based on direct trading of link-based physical transmission rights (or scheduling priorities) in parallel with energy markets are not prone to such constraint violations. In such designs adherence to the secure limits is enforced by the number of rights issued whereas the market determines the value of these rights which are required to support energy trades. This link-based approach is in line with NERC's proposal for a gradual transition from a TLR protocol based on administrative curtailment of transactions to a market-based protocol in which flow-based rights on congested flowgates can be traded among competing users (the FLOWBAT approach).

Drawing policy conclusions based on stylized examples is always a precarious undertaking, although in the electric restructuring arena it is a way of life. Our example is by no means a representation of real electric power systems and admittedly was not easy to find. In fact, we identified this example using a systematic random search of cost functions and network parameters. It is also not clear whether in more complex networks the phenomenon we illustrated disappears or becomes more prevalent. Nevertheless, our counterexample identifies a flaw in the theory underlying the CHHP proposal and suggest caution in adopting such an approach. It also highlights a relative strength of the competing approach based on physical tradable flowgate rights, which is not subject to the same problem.

REFERENCES

[1] NERC Transaction Reservation and Scheduling Self Directed Work Team, "Discussion paper on aligning transmission reservations and energy schedules to actual flows," NERC, Princeton, NJ, Nov. 1998. Distributed via the NERC process.

[2] H. P. Chao and S. Peck, "A market mechanism for electric power transmission," *J. Reg. Econ.*, vol. 10, pp. 25–59, July 1996.

[3] R. D. Tabors and R. Wilson, "Auctionable capacity rights and market-based pricing," unpublished, Apr. 21, 1998.

[4] H. P. Chao, S. Peck, S. Oren, and R. Wilson, "Flow-based transmission rights and congestion management," *Electricity J.*, vol. 13, no. 8, pp. 38–58, Oct. 2000.

[5] M. D. Cadwalader, S. M. Harvey, W. W. Hogan, and S. L. Pope. (1999, Oct.) Coordinating congestion relief across multiple regions. [Online]. Available: http://www.ksg.harvard.edu/whogan.

[6] B. H. Kim and R. Baldick, "Coarse-grained distributed optimal power flow," *IEEE Trans. Power Syst.*, vol. 12, pp. 932–939, May 1997.

[7] J. Eckstein, "Parallel alternating direction multiplier decomposition of convex programs," *J. Optim. Theory Applicat.*, vol. 80, no. 1, pp. 39–63, Jan. 1994.

[8] F. Wu, P. Varaiya, P. Spiller, and S. Oren, "Folk theorems on transmission access: Proofs and counterexamples," *J. Reg. Econ.*, vol. 10, no. 1, pp. 5–23, July 1996.

Shmuel S. Oren (M'72–SM'99–F'01) received the B.Sc. and M.Sc. degrees in mechanical engineering from the Israel Institute of Technology (Technion), Haifa, and the M.S. and Ph.D. degrees in engineering economic systems from Stanford University, Stanford, CA, in 1965, 1969, and 1972, respectively.

Currently, he is Professor of Industrial Engineering and Operations Research at the University of California, Berkeley, and a former Chairman of that department. He is the Berkeley Site Director of PSerc, a multiuniversity Power Systems Engineering Research Center sponsored by the National Science Foundation and industry members. His academic research and consulting focus on planning scheduling and economic analysis of power systems, and in particular on issues concerning market design for competitive electricity. He published numerous papers in this area and has been a consultant on electricity restructuring issues to numerous public and private organizations in the United States and abroad.

Andrew M. Ross received the B.S. degree in mathematics from Harvey Mudd College, Claremont, CA, and the M.S. degree in operations research from the University of California, Berkeley, in 1996 and 1997, respectively. He is currently pursuing the Ph.D. degree in industrial engineering and operations research at the University of California, Berkeley

His teaching interests are in linear and nonlinear programming, while his research focuses on nonhomogeneous queueing systems that are common in the telecommunications industry.