Opinion on "Load-Based and Source-Based Trading of Carbon Dioxide in California"

by
Frank A. Wolak, Chairman
James Bushnell, Member
Benjamin F. Hobbs, Member
Market Surveillance Committee of the California ISO

November 27, 2007

1. Introduction

California Assembly Bill 32, the "The Global Warming Solutions Act of 2006", established a goal of reducing the state's greenhouse gas (GHG) emissions to 1990 levels by the year 2020. The California Air Resources Board (CARB) is charged with developing the necessary measures to achieve that target. CARB is cooperating with the California Public Utilities Commission (CPUC) and the California Energy Commission (CEC) to evaluate alternative mechanisms for achieving that goal in the electricity sector. On November 9, 2007, the Administrative Law Judges in the CPUC-CEC joint proceeding on GHG issues[1] issued a ruling requesting comments on issues relating to the type of greenhouse gas regulation that should be applied to the California electricity supply industry.

This opinion responds to that request. The Market Surveillance Committee (MSC) has previously been involved in discussions of the development of GHG policies for the power sector. In particular, the MSC held a meeting at CARB's offices in Sacramento on June 8, 2007 to discuss the interaction of GHG policies and short-term electricity markets in the western United States (US) and the impact of GHG policies on procurement by the state's load-serving entities (LSEs) and other, non-LSE retail providers of electricity, such as municipal utilities, which we will refer to generically as LSEs. In this opinion, we only address a subset of the questions in the ALJs' Ruling, emphasizing the question of the point of compliance with a state-imposed GHG emissions cap. In particular, we address the economic efficiency implications for the California electricity market of alternative points of compliance.

There are essentially four broad alternatives for implementation of AB 32 within the California electric sector. The first is to regulate emissions by placing a reporting and compliance obligation on LSEs. Under this "load-based" approach, LSEs would have to demonstrate that the power they have purchased represents a mix of sources that achieves a specified target, either in terms of tons per year or in terms of carbon intensity.[2] The second is to implement a "pure" source-based cap and trade system similar to other cap-and-trade systems in

[1] Administrative Law Judges' Ruling Requesting Comments On Type And Point Of Regulation Issues, dated November 9, 2007 , issued by ALJ Charlotte F. TerKeurst (CPUC) and ALJ Jonathan Lakritz (CEC) in CPUC R.06-09-004 and CEC Docket No. 07-OIIP-01 (the ALJs' Ruling).

[2] Because electricity in a looped transmission network flows according to the laws of physics, it is physically impossible to determine the GHG emissions caused by each MWh of electricity consumed by each load-serving entity. For this reason, a load-based system must use an administrative procedure to assign GHG emissions to each MWh of electricity consumed in California.

other parts of the world.[3] A third approach would be to implement some hybrid cap-and-trade system that would effectively act like a source-based program for plants within the state, but still try to capture the emissions impact of imports in some fashion. The “first-seller” approach is the most widely discussed of this general hybrid concept.[4] The last alternative would be to focus AB 32 implementation efforts on mechanisms other than cap-and-trade. In that case, California’s participation in a cap-and-trade system would be implemented in concert with a regional or federal program, rather than preceding it.

A choice between these approaches should take into consideration various economic and environmental goals. These include efficiency of system dispatch and the performance of wholesale and retail electricity markets, the efficiency of investment in new generation facilities and energy efficiency technologies, consumer costs, administrative simplicity, and effectiveness in achieving the GHG reduction goals set forth in AB32. Because GHGs are global pollutants, perhaps the most important consideration is compatibility with possible west-wide or federal GHG regulations that might be adopted in the near future. Even if California were to reduce its GHG emissions by, say, half, this would reduce world GHG emission by less than one percent. Consequently, a key measure of the success of any state-level GHG emissions regulation is the extent to which other states and jurisdictions adopt it.

While we believe that there are advantages and disadvantages to each of the approaches described above, in this opinion we wish to emphasize two points. First, an often-claimed advantage of the load-based and hybrid approaches–that they regulate the GHG content of imported electricity–is grossly overstated. Although firms would not be able to avoid compliance by physically moving their sources of production out of the State (“leakage”), they would be able achieve much the same ends by “reshuffling” their purchases of imported energy to originate from clean sources.[5] In fact, reshuffling is in many ways a less costly strategy for circumventing environmental regulation than is leakage.[6]

The second point that we wish to emphasize is that the first option, a load-based cap-and-trade system, is clearly and substantially inferior to the other options. We believe that the load- and source-based approaches are similar in some respects, but that the load-based approach is

---

[3] A source-based approach places compliance responsibility on the facility that is emitting the pollution (the source). In a source-based system each facility would need to acquire emissions permits to offset their total emissions.

[4] A first seller is an entity that first brings power into the California market. All generation units located in the California ISO control area are first sellers of electricity. So in this sense, the first-seller approach is a source-based approach because it is straightforward to determine the GHG emissions per MWh of energy produced from the technical operating characteristics of the in-state generation unit. However, for imports of electricity, the first seller is the entity importing the power into the state. In this case, an administrative procedure must be designed to assign a GHG emissions rate per MWh of energy imported into California for each importing entity. In this sense, the first-seller approach functions like a load-based mechanism because there is no unambiguous method to determine the GHG emissions caused by the electricity sales into California.

[5] Several options for mitigating reshuffling have been raised, but they remain among the most controversial and legally vulnerable aspects of the overall cap-and-trade design.

[6] With leakage, firms have to physically change the sources of production from whatever they were before the environmental regulation took effect. Assuming that firms were buying power from the cheapest sources, changing the mix of generation would have to involve increasing costs. Under reshuffling there could be *no* change in the mix of generation at all, only a realignment of the transactions that define who is buying power from which source.

distinctly inferior in others.[7] In particular, we argue that the two systems are essentially the same on the issues of determining the GHG content of power imports and incentives for investments in energy efficiency and renewable energy. However, in terms of administrative complexity, adverse impacts on the efficiency and costs of dispatching generation units to meet load in California energy and ancillary services markets, and compatibility with likely federal GHG legislation, a load-based system has serious disadvantages compared to any of the other options. Contrary to some claims, we believe that resulting cost of energy to consumers would likely be *higher* under a load-based cap. We discuss each of these issues below. The Appendix summarizes a simple model that demonstrates the equivalence of the two systems in terms of total cost of energy to final consumers–under assumptions that ignore the potential higher consumer costs of a load-based approach due to inefficient generator dispatch in the California day-ahead and real-time markets.

2. The Issues of Imports and Compatibility with Federal Legislation

All options face the same challenge in achieving the goal of reducing total GHG emissions from sources that serve California's electricity demand. The California market is embedded in the much larger western North American market. When only a subset of loads or generation units in this larger market are subject to regulation, a local GHG emissions reduction goal can be frustrated by increases in imports and thus unregulated GHG emissions from elsewhere in the larger market. This has been an issue with state-level regulation of GHG elsewhere in the U.S.[8]

Further, efforts to prevent increased imports from unregulated regions (GHG "leakage") or to incent emissions reductions elsewhere in the west by identifying sources of power for imports and their emissions are likely to be ineffective, regardless of the administrative procedures used to identify specific generation sources. This is because the depth of the west-wide market and the amount of "clean" generation available is such that there is likely to be more than enough clean generation that can be assigned, on paper, to California imports, without actually changing system operations, or investment, in the west. This has been called the "contract shuffling" problem.[9] Markets for electric power will tend to identify and use the cheapest sources of electricity; prohibiting or penalizing imports that, in name, are connected with dirtier sources are unlikely to result in their being dispatched differently, if they are indeed the cheapest power source in the region not subject to GHG limits. Consequently, any policy—load-based or source-based—that addresses only California emissions, or attempts to prevent

[7]See D. Burtraw, "State Efforts to Cap the Commons: Regulating Sources or Consumers," Resources for the Future, Nov. 9, 2007 for a related and, in some cases, more extended discussion of several of these issues.

[8]For instance, it has been estimated that all of the nominal CO_2 emissions reductions that would occur by expanding the eastern states' Regional Greenhouse Gas Initiative to Maryland would be offset by greater CO_2 emissions elsewhere in the Eastern Interconnection. However, interactions with non-Eastern markets through emissions allowances markets together with purchases of non-power emissions offsets means that the net effect of Maryland joining RGGI is an overall decrease in emissions. (See University of Maryland, Resources for the Future, The Johns Hopkins University, and Towson University, *Economic and Energy Impacts from Maryland's Potential Participation in the Regional Greenhouse Gas Initiative*, Submitted to the Maryland Dept. of Environment, http://cier.umd.edu/RGGI/documents/UMD_RGGI_STUDY_FINAL.pdf, Jan. 2007, Section 9.3.3.)

[9]J. Bushnell, C. Peterman, and C. Wolfram, "California's Greenhouse Gas Policies: Local Solutions to a Global Problem?" CSEM Working Paper 166, University of California Energy Institute, April 2007.

leakage by administrative procedures that identify sources of imports, is very likely to have its environmental goals frustrated by the inability of a California-only policy to alter operations or investment decisions elsewhere in the western North American market.[10]

This conclusion means that a fully effective GHG policy for the electric sector must cover the bulk of the western US market. This implies that a California policy under AB32 should be viewed as an initial step, and that a major goal of that policy should be to facilitate the establishment and implementation of federal or other west-wide policies, rather than to act as an obstacle to such policies. Precedent, as well as the preponderance of proposed federal legislation, indicates that source-based trading of emissions allowances will likely be the basis of any federal regulation of power sector GHG emissions.[11] The emissions accounting and other mechanisms associated with a California load-based system would, at best, be sunk costs that would be abandoned if a federal source-based GHG trading system is adopted. At worst, the existence of an incompatible state-level system could delay or increase the cost of implementing the federal system.

3. The Relative Cost of Load-Based and Source-Based Trading Systems

Another way in which load-based and source-based systems are similar is in the resulting cost of energy to consumers. Experience in the European Union with CO_2 trading in the power sector has shown that high prices for CO_2 translate into higher prices of electricity, and that many generators enjoy the benefits of increased revenues.[12] Arguments have been made that a load-based system would avoid these problems in California, at significantly less expense, in terms of consumer payments, than would a source-based system that achieves the same level of total GHG emissions.[13]

These arguments are incorrect. Even assuming that the same generation sources are used to serve demand under both systems, we demonstrate in the Appendix that a source-based system in which LSEs sign contracts with individual generators to minimize the cost of serving load, while meeting a GHG constraint, results in the *same* cost to load as a source-based system in which generators maximize profit, subject to a cap-and-trade system for GHGs, with the same constraint. This conclusion assumes that, in the latter system, consumers will be allocated all

[10] An effective change in dispatch or investment in western markets not subject to GHG regulation, or the prevention of leakage via increased imports, might be accomplished by a significant regulatory intervention. For example, credit for clean generation outside of California might only be granted to new renewable generation investments that are not counted towards any state's renewable portfolio generation; such investment would be unlikely to have occurred otherwise, and so would represent a real change in system investment and operations outside of California. However, such rules would represent a significant regulatory intervention in market processes, and the emission benefits could arguably have accomplished at less cost through more ambitious renewable energy goals without a cap-and-trade system for GHG emissions.

[11]These precedents include the federal SO_2 and NO_x trading systems under Title IV of the Clean Air Act Amendments, the NO_x SIP call, and, most recently, the Clean Air Interstate Rule.

[12] J.P.M. Sijm, K. Neuhoff, and Y.Chen, "CO_2 Cost Pass Through and Windfall Profits in the Power Sector," *Climate Policy*, Vol 5, No. 1, 2006, pp. 49-72.

[13] See for instance, Synapse Energy, "Exploration of Costs for Load Side and Supply Side Carbon Caps for California", Joint *En Banc* Hearing of PUC and CEC on Point of Regulation in the Electricity Sector (R.06-04-009), August 21, 2007, available at www.synapse-energy.com (accessed Nov. 16, 2007).

emissions allowances, which they then can sell to generators. However, as we argue in Section 5 below, we believe that the perverse incentives created by regulatory efforts to assign the emissions of specific sources to LSEs will lead to the deployment of a less efficient generation mix to serve demand. Overall, this would result in the load-based system leading to higher, *not* lower, energy cost to consumers.

The higher wholesale electricity profits that result from implementation of a GHG trading system, whether load-based or source-based, have at least two possible sources. One is the value of the emissions allowances themselves ("the allowance rents"), which is the allowance price times the number of allowances. If allowances are given for free to generators, this value increases generator profits. On the other hand, if allowances are given to load, and then sold to generators (perhaps via an auction) for use in a source-based system, with the proceeds returned to consumers, then these rents will, to some extent, offset the price increases resulting from the cap-and-trade mechanism. These rents are also retained by consumers under a load-based system.

The other possible source of higher profits resulting from GHG emission permit trading is what might be called the "rents of clean generation." In a source-based system, generation units with low emissions will benefit from higher energy prices because the price increases will exceed the expense of allowances. The entire industry will benefit, on average, if the average emissions rate for all generation units is less than the marginal emissions rate that causes the price increase. This profit is retained by generation unit owners in both source-based and load-based systems. This is because, in load-based systems, as the Appendix shows, LSEs will pay more for electricity from cleaner generators, because that generation is more effective in helping the LSEs meet their emissions constraint. In a competitive market, the difference in prices that LSEs would pay to different generation units will equal the difference in their emissions rates times the implicit cost of emissions to the LSE. It turns out that this implicit cost will be the same as the price of allowances in a source-based system with the same GHG target.

If this clean generation is independently owned, the "rents to clean generation" would be retained by generators under either load- or source-based systems. However, within California, a significant fraction of this generation is owned by utilities, so any such additional profits to those plants could be returned to consumers. Meanwhile, the portion of those rents accruing to new renewable sources, given effective competition in that sector, would translate into lower subsidies from LSEs under California's renewable portfolio standard (e.g., lower renewable energy credit prices), also resulting in a return of some of those rents to consumers. These returns of "rents to clean generation" would not be affected by the existence of a load-based or source-based system.

Therefore, under an assumption of comparable production sources under the two systems, the load-based system yields no advantages in costs over a source-based system (with allowances owned by consumers). This is because consumers would, in both cases, retain the allowance rents as well as the portion of rents to clean generation that accrue to utility-owned and new renewable generation. However, if the load-based system leads to a distortion of the mix of production sources, it would yield higher costs than a source-based system.

4. Impact on LSE Incentives for Energy Efficiency and Renewable Energy

It has sometimes been argued that a load-based system will result in a greater incentive for LSE investments in energy efficiency and renewable technology than a source-based system. The argument is that the load-based system "paints a target on the back" of the LSE, making it more accountable for its carbon footprint. This effect is speculative, and we doubt that it would be significantly different from the incentives provided by source-based regulation, for the following reasons.

First, California investor-owned utilities are already subject to an extensive regulatory system that arguably provides more incentives than any other state for investment in energy efficiency and renewables. These incentives include procurement priorities that place efficiency at the top of the list among all resources; a charge on all California electricity consumers to fund cost-effective energy efficiency programs; the decoupling of utility revenues from sales; and the rate-of-return incentives adopted by the CPUC in September 2007. With implementation of AB32, carbon costs will be included as part of the "avoided energy costs" in the "Total Resource Cost Test" used to identify beneficial efficiency programs under California's rules; as a result, more energy efficiency programs will become cost-effective. This will be true under either load- or source-based programs.[14] California's many regulatory incentives will then motivate the state's utilities to pursue many, if not most, of those opportunities. We see no reason why California's regulated utilities will be more likely to pursue these newly cost-effective programs under one emissions regulatory system than another.

Second, California's LSEs are being required to account for and report the GHG emissions associated with their contracts no matter what sort of GHG regulatory system is implemented under AB32. There will be public visibility and pressure to pursue energy efficiency to lower emissions under either load-based or source-based systems.

Third, California already has ambitious renewable energy goals for its LSEs. It seems unlikely to us that a load-based trading scheme would motivate LSEs to exceed the 20% renewable goal for 2010..

5. Using the California ISO Markets to Enhance GHG Regulation

An important way in which load-based and source-based systems differ is in how they interact with the new markets that are to start operation next year under the ISO Market Redesign and Technology Upgrade (MRTU).[15] Under the MRTU design, the ISO's new day-ahead market will perform two functions in an integrated manner:

(1) provide a market for wholesale buyers and sellers to transact, and
(2) schedule the use of the transmission grid to deliver energy to consumers.

---

[14]The Appendix provides an example the equivalence of avoided energy costs under load- and source-based systems.

[15]Note that the concerns we express in this section about the load-based system would also arise if, instead, GHG trading were to be implemented under the present California ISO markets or, indeed, under any real-time or day-ahead market that mixes different sources of power.

In contrast, under today's market structure, these two functions are separated, so that participants can only schedule their use of the transmission grid, but cannot engage in energy trading through a formal day-ahead market. To maximize the benefits that market participants will receive from the integrated day-ahead market that will exist under MRTU, they must submit "economic bids" to this market; that is, specific quantities of energy they want to buy or sell at specific prices, rather than "self-scheduling" all or the great majority of their energy by submitting only their desired quantities without prices. As the volume of self-schedules relative to economic bids increases, the efficiency of the economic dispatch declines, and this is the main concern about how the choice of GHG regime interacts with the ISO markets. As we explain below, the load-based approach will encourage self-scheduling in conflict with the efficiency potential of the MRTU markets, whereas the source-based approach will encourage economic bidding, thus utilizing MRTU's new economic dispatch in concert with GHG regulation to achieve the desired environmental objectives.

As pointed out in the previous section, LSEs will pay cleaner generation a higher price than high-emissions generators in a power market that is subject to a load-based GHG compliance mechanism. Such a market is incompatible with the ISO day-ahead and real-time markets for energy and ancillary services, because MRTU will make no distinction between generation units having different emission rates in its bidding and dispatch processes. It will not be possible for the ISO to define, for example, different locational marginal prices (LMPs) for dirty and clean power at each bus, or to explicitly consider relative emissions rates in deciding what units will be chosen to provide, say, spinning reserves or residual unit commitment services. MRTU will not be able to accommodate demand bids that express a higher willingness to pay for low emissions power. Power and ancillary services from various sources with various emissions rates will be inextricably mingled within the ISO markets. The best that can be done is to calculate an average or marginal emissions rate associated with ISO power delivered at different times and locations.

That the California ISO markets will not differentiate sources of power by their emissions is a problem only with load-based systems where LSEs must track the emissions associated with different sources of power. By contrast, in a source-based system, compliance is at the point of production, and the opportunity cost of allowances is internalized into the bids submitted by suppliers to the ISO markets. Efficient compliance with emissions caps can then be attained, without complicating the ISO markets.

We believe that a load-based system, in which LSEs must track, using a pre-specified administrative procedure, the emissions associated with all their energy transactions, will pose a grave danger to the efficiency and competitiveness of the California short-run markets. This is because LSEs participating in ISO markets will be buying power, and generators will sell power in the ISO markets based on some average or marginal emissions rate that is administratively determined. As a result, generation unit owners that can command a premium for their units in the bilateral market, because of the unit's low emissions, will selectively avoid providing bids to the ISO markets, leaving just the high emission generation units willing to accept the ISO prices, which would reflect average emissions. In this sense the "dirty" generation would chase out the "clean."[16] Low emission sources will tend to self-schedule, in order to secure higher prices. Another reason why more self-scheduling is likely to occur is because each LSE will be trying to self-manage its supply portfolio to stay within their emissions limitation. Assuming that compliance will be based on actual output, as opposed to contracted supply, LSEs will seek to protect their portfolio from being re-dispatched in the ISO markets, by submitting self-schedules.

As a result, the amount of market bids that the ISO will have available to manage congestion and to optimize total system dispatch will be severely limited. The ISO markets for energy and ancillary services will become significantly thinner. For instance, hydropower, which has zero GHG emissions, would likely be less willing to provide spinning reserve to the ISO because it would not want to earn the (relatively) low energy prices it would gain if it is dispatched in the ISO's markets, thereby giving up more lucrative "clean" energy prices in the bilateral market. Likewise, highly efficient combined cycle units would be less likely to bid into the ISO's real-time markets. Having less resources available for real-time system operation would increase costs for the ISO for any redispatch that must occur between day-ahead and real-time to manage congestion, to accommodate demand forecast errors, and to adjust for unexpected equipment failures. With fewer resources bid into the ISO markets, the likelihood of schedule curtailment would increase, as would the stress on system operators as they try to keep the system balanced. Furthermore, thinner markets would likely also be less competitive markets. Ultimately, all of these increased costs would be passed on to consumers.

Thus, a load-based system would conflict with the goal of more competitive energy and ancillary services markets in California, and with the goal of creating liquid and deep markets day-ahead and in real-time in order to lower operation system costs and maximize the ability of the ISO operators manage unforeseen contingencies.[17] In contrast, a source-based policy and MRTU would work together to lower the costs of meeting GHG goals and California's need for power.

It is important to recognize that the problems created by a bias against pool-based markets grow larger as the geographic scope of a load-based cap-and-trade system grows. Although the environmental regulatory problem that motivates the load-based scheme (i.e. regulating imports) grows less significant as more states participate, the *economic* consequences of the environmental regulation can grow more serious. Many recognize that the western market

---

[16]This is, in a very general way, analogous to the infamous "dec" game in zonal power markets, in which intrazonal congestion was ignored day-ahead, but resolved in real-time by "inc"ing costly generation in load pockets and allowing cheaper generation in generation-rich areas to buy out of their day-ahead commitment at a low price. The result was that day-ahead markets would receive an excess of undesirable generation (from generation-rich areas) while the most desirable generation (in load pockets) would stay away, awaiting higher prices in real-time. This increased congestion and consumer costs.

[17]The difficulties that arise in the ISO MRTU markets if power and emissions attributes are bundled in a load-based system can be avoided if emissions attributes are unbundled and traded separately between generators and LSEs (see M. Gillenwater and C. Breidenich, "Internalizing Carbon Costs in Electricity Markets: Using Certificates in a Load-Based Emissions Trading Scheme," Unpublished manuscript, Science Technology and Environmental Policy Program, Princeton University, Princeton, NJ). However, that unbundling proposal can be shown elsewhere to be economically equivalent to source-based trading of allowances with allowances allocated free to generators according to their sales; see Section A.5 of the Appendix. Such a system would be costly to consumers, while being more complex to administer than a source-based system.

is currently a patchwork of less than ideally coordinated trading rules and protocols. Overcoming these "seams" issues continues to be an important concern. By creating an institutionalized bias against pool-based markets, the west could be turning its back on the opportunity to better unify its regional markets for energy and GHG emissions permits.[18]

6. Concluding Comments

Our recommendation against adopting a load-based program for regulating the emissions of greenhouse gases associated with electricity consumption in California should not be interpreted as implying that we necessarily favor the immediate implementation of source-based trading in the state. The very likely advent of federal GHG regulation in the next few years means that there are advantages to deferring implementation of a formal trading system in California until the form of federal regulation becomes clear. Given the ambitiousness of California's existing renewable energy and energy efficiency programs, we believe that most of the GHG reductions that would be achieved in the power sector under an emissions cap (either load-based or source-based) would likely result from those programs in any event. We believe that it is crucial that a level playing field ultimately be established that would reward all measures for reducing CO_2, because measures such as improving the efficiency of fossil-fuel power plants might be cost-competitive with investments in renewables and energy efficiency. However, because California's dependency on imported power raises doubts about the environmental integrity of a California-only GHG trading system, it is difficult to justify the cost of establishing a sophisticated trading system (either load-based or primarily source-based) that might be abandoned quickly in the face of federal preemption.

If it is decided that regulation of the GHG emissions of the California power sector should proceed immediately, despite these concerns, we strongly recommend that a source-based system be implemented, rather than a load-based system. We conclude that a load-based system, rather than lowering energy costs to California consumers relative to a source-based system, would likely result in higher costs. At best, the load-based system is no less expensive to consumers than the source-based approach, if both result in efficient dispatch and emissions allowances are allocated to LSEs. However, the load-based approach poses significant risk to dispatch efficiency by discouraging cleaner sources from submitting bids the California ISO's day-ahead and real-time markets, thereby decreasing the flexibility and competitiveness of those markets. In contrast, a source-based system utilizes those markets to help achieve the GHG policy objectives more effectively and efficiently.

Therefore, the speculative benefits of a load-based system, in terms of possibly greater incentives for energy efficiency or renewables, cannot be justified in light of the additional administrative complexity and cost of such a system, the threat that it would pose to the competitiveness and efficiency of the ISO-administered markets under MRTU, and the

---

[18] This is not just an academic question. Research on the expansion of the PJM market has demonstrated a significant change in the operations of power plants in the eastern U.S. (see E. Mansur and M. White, "Market Organization and Efficiency in Electricity Markets", Working Paper, April 2007 (http://www.som.yale.edu/faculty/etm7/papers/mansur_white_pjmaep.pdf). It appears that the previous wholesale market regime, which is comparable to much of the western U.S. today, was not taking full advantage of the efficiencies offered by the network. Increased efficiency can produce both economic and environmental benefits.

additional difficulties that would arise when transition to a federal cap-and-trade system would occur.

Technical Appendix[19]

This Appendix provides a demonstration of the general result described in Section 3: that a load-based system results in the same consumer costs as a source-based system in which allowances are sold by consumers to generators. This demonstration is based on simplified models of the power and emissions markets under load- and source-based policies (Sections A.1 and A.2), followed by an analysis of their general properties (Section A.3). A numerical example is then given that illustrates the principle (Section A.4). That example also illustrates the equivalence of avoided costs (for use in the "Total Resource Cost" benefit-cost test for energy efficiency programs in California) under the two policies. Finally, in Section A.5, we show the economic equivalence to source-based trading of the Gillenwater and Breidenich (2007) (*op. cit.*) proposal to unbundled emissions attributes from power in a load-based system, pointing out that it involves significant subsidies of producers**.**

A.1 Model of a Load-Based Equilibrium

The model for a load-based system consists of models of consumer and supplier decision making, which combined with a market clearing condition defines the market equilibrium. The consumer model includes a constraint on the emissions resulting from the consumer's portfolio of supply contracts.

<u>Consumer Model:</u> One single LSE serving the market is assumed; this model can be readily generalized to multiple LSEs, and the fundamental results do not change. The single LSE acts as a price taker with respect to the price of electricity (i.e., does not exercise unilateral market power).

x_{Li} = MW purchases from supplier i by the LSE in the load-based equilibrium
p_{Li} = \$/MWh price paid (assumed fixed by LSE) for power from supplier i in the load-based equilibrium
E_i = ton/MWh emissions rate for supplier i
L = MW load for LSE (a single hour is assumed for simplicity)
K = tons/MWh emission cap for load

The model is:

MAX -Expenditures = $-\Sigma_i\, p_{Li}\, x_{Li}$
subject to:
$\Sigma_i\, E_i\, x_{Li} \le K\, L$ (shadow price α_L)
$\Sigma_i\, x_{Li} = L$ (shadow price β_L)
$x_{Li} \ge 0$ all i

[19] The model discussed in this memo is an elaboration of models in B.F. Hobbs, "An Analysis of the Breidenich/Gillenwater Proposal for Load-based Trading of CO_2 rights," Unpublished manuscript, The Johns Hopkins University, June 7, 2007, and in Y. Chen and A. Liu, "Economic and Emissions Implications of Load-based, Source-based, and First-seller Emissions Trading Programs under the California AB32", Draft, University of California Merced, November 2007.

That is, the LSE minimizes the cost of meeting power demand and the emissions constraint by choosing which suppliers to buy power from. (The objective is phrased as a maximization so that the dual variable of the emissions constraint is nonnegative.)

Producer Model: There is one plant per producer, with a constant marginal cost and a fixed capacity. Each producer is a price taker.

y_{Li} = MW sales from producer i to LSE
C_i = \$/MWh marginal cost of production for producer i
CAP_{Li} = MW generation capacity for producer i

The producer's problem is:

$$\text{MAX } \text{profit}_{Li} = (p_{Li} - C_i)y_{Li}$$

subject to:

$y_{Li} \leq CAP_i$ (shadow price μ_{Li})
$y_{Li} \geq 0$ all j

Market Clearing: This ensures that supply and demand for energy from each producer are equal, and mathematically generates the market clearing price.

$x_{Li} = y_{Li}$ for all i (shadow price p_{Li})

Equilibrium Model: The equilibrium model consists of the first-order conditions for each of the market participant's optimization problems (Kuhn-Tucker conditions) together with the market clearing conditions, yielding a "square" system (as many conditions as unknowns). The unknowns are $\{x_{Li}, y_{Li}, p_{Li}, \alpha_L, \beta_L, \mu_{Li}\}$. It can be shown that low emission producers get a premium for their power representing the value to LSEs for meeting the emissions constraint. Furthermore, if there is more than one LSE, each LSE will pay the same price for power from a given producer (of course, transmission constraints are disregarded).

A.2 Model of a Source-Based Equilibrium

Notation for this model is the same as for the load-based equilibrium, except that the subscript "S" is substituted for "L" on all variables.

Consumer Model: The model for consumers is a simplified version of the load-based model:

$$\text{MAX } -\text{Expenditures} = -\Sigma_i \, p_{Si} \, x_{Si}$$

Subject to:

$\Sigma_i x_{Si} = L$ (shadow price β_S)
$x_{Si} \geq 0$ all i

Producer Model: This model differs from the load based one because it includes the expense of allowances in the profit function.

MAX $\text{profit}_{Si} = (p_{Si} - C_i - \alpha_S E_i) y_{Si}$

Subject to:

$y_{Si} \leq CAP_i$ (shadow price μ_{Si})
$y_{Si} \geq 0$ all j

Market Clearing Condition:

$x_{Si} = y_{Si}$ for all i (shadow price p_{Si})
$\Sigma_i E_i y_{Si} \leq K L$ (nonnegative shadow price α_S)

The emissions price can be positive only if the emissions constraint is binding. The total amount of allowances is assumed for the sake of comparison to be the same as the sum of maximum emissions by the LSEs under the load-based model.

Equilibrium Model: The equilibrium model consists of the first-order conditions for each of the market participant's optimization problems (Kuhn-Tucker conditions) together with the market clearing conditions, yielding a "square" system (as many conditions as unknowns). The unknowns are $\{x_{Si}, y_{Si}, p_{Si}, \alpha_S, \beta_S, \mu_{Si}\}$. It can be shown that energy prices p_{Si} paid to all producers i are equal (unlike the load-based case).

A.3 Theoretical Equivalence of the Models

Results for the relationship of the two models can be shown as follows.

The first set of results concerns the relationship between the equilibrium values of the price and quantity variables:

$p_{Li} = p_{Si} - \alpha_S E_i$ for all producers i
$\{x_{Li}, y_{Li}, \alpha_L, \beta_L, \mu_{Li}\} = \{x_{Si}, y_{Si}, \alpha_S, \beta_S, \mu_{Si}\}$

That is, the load-based price for energy from a producer equals the source based price minus a penalty for its emissions. This penalty is the per ton shadow price of emissions (which is implicit in the LSE's maximization problem) times the emissions rate. This result is shown in two steps. The first step is to substitute in the source-based equivalents for the load-based variables in the load-based equilibrium conditions, and showing that the source-based variables satisfy those equilibrium conditions. The second step is to go the other way: substitute $p_{Li} + \alpha_S E_i$ for p_{Si}, and $\{x_{Li}, y_{Li}, \alpha_L, \beta_L, \mu_{Li}\}$ for $\{x_{Si}, y_{Si}, \alpha_S, \beta_S, \mu_{Si}\}$ in the source-based equilibrium conditions; it turns out that the load-based variables satisfy those conditions. Thus, prices (adjusted for emissions in the case of energy prices) and the quantities for the load-based equilibrium and source-based equilibrium are the same. (More generally, if the equilibrium for one of the models is not unique, then this is also true for the other model, and each solution to one has an equivalent solution to the other.)

The second set of results concern the equivalence of the consumer payments under the two systems. In particular, because $p_{Li} = p_{Si} - \alpha_S E_i$, the consumer payments minus allowance rents (assumed to accrue to consumers) under the source-based system equal the consumer payments under the load-based system. The demonstration is as follows:

Source-based energy payments minus allowance rent

$$= \Sigma_i p_{Si} x_{Si} - \alpha_S K L = \Sigma_i p_{Si} x_{Si} - \alpha_S \Sigma_i E_i y_{Si} = \Sigma_i (p_{Si} - \alpha_S E_i) x_{Si}$$
$$= \Sigma_i (p_{Li}) x_{Li} = \text{Load-based energy payments}$$

The second step is true even if the emissions constraint is not binding, because $\alpha_S = 0$ in that case.

By the same logic, it can be shown that generator profits are the same under a load-based or source-based system, assuming that under the latter consumers are allocated the allowances initially, and sell them to producers. Thus, the "Rents to Clean Generation" that generators earn in the source-based system are also retained by generators in the load-based system. An assumption that load would gain those rents under the load-based system is incorrect.

A.4. Numerical Example

Consider an isolated power system (no imports) in which there are two load serving entities (1 and 2) three generation companies each with different types of generation: A,B, and C.

- The load serving entity has constant load L = 2000 MW. Under the load-based system, it is obliged to buy power contracts that, on average, have an emissions rate of 0.55 tons/ MWh.
- Generation type A has emissions E_A of 0 tons/MWh, marginal cost C_A = 0\$/MWh (wind or hydro), and capacity CAP_A = 500 MW.
- Generation type B has emissions E_B of 0.6 tons/MWh, marginal cost C_B = 80\$/MWh (wind or hydro), and ample capacity CAP_B (no limit).
- Generation type C has emissions E_C of 1 ton/MWh, marginal cost C_C = 40\$/MWh (wind or hydro), and ample capacity CAP_C (no limit).

The solution to the load-based equilibrium model from Section A.1 results in the following generation, cost, and prices:

- MW generation y_{Li} from companies i=A,B,C: y_{LA} = 500 MW, y_{LB} = 1000 MW, y_{LC} = 500 MW. These also equal MW purchases by the LSE (x_{LA}, x_{LB}, and x_{LC}, respectively).
- Prices p_{Li} paid by the LSE for each type of generation i=A,B,C: p_{LA} = \$140/MWh; p_{LB} = \$80/MWh; p_{LC} = \$40/MWh.
- The total paid for power by the LSE is \$170,000, or \$85/MWh. This is also the value of β_L, the shadow price of the load constraint. However, the marginal cost of serving load for the LSE is the sum of this shadow price plus K times the shadow price of the LSE's emissions constraint (K*α_L, 0.55 tons/MWh *\$100/MWh), or \$140/MWh.
- Only generator A makes a profit (of \$140/MWh*500 MW, or \$70,000).

Thus, cleaner generation gets a premium. The premium results from the value it provides to the LSE by making it easier for the LSE to achieve its emissions target; an LSE is willing to pay more for power that is cleaner. As shown in Section A.3, it turns out that the "shadow price" of the LSE's emissions constraint— $100/ton—*equals* the price of emissions allowances in the source-based example, below.

Now consider a source-based system in the emissions cap is 1100 tons, and consumers own the allowances. It will result in the following equilibrium using the model of Section A.2:

- MW generation y_{Si} from companies i=A,B,C: y_{SA} = 500 MW, y_{SB} = 1000 MW, y_{SC} = 500 MW. These also equal MW purchases by the LSE (x_{SA}, x_{SB}, and x_{SC}, respectively).
- The price for power is $140/MWh for all producers.
- The price for allowances is $100/ton. So the total allowances rent is $110,000. As a result of this price of allowances, the net marginal cost for B's generation is $140/MWh (=$80/MWh for fuel + 0.6 ton/MWh*$100/ton for allowances), which is the same for C (=$40/MWh for fuel + 1.0 ton/MWh*$100/ton). Neither B nor C earn any operating profit, as price equals their marginal cost.
- On the other hand, A's marginal cost is $0, as it has neither fuel costs nor emissions; therefore, it will produce at its 500 MW capacity, and earn $70,000 in profits ($140/MWh*500 MW).
- The LSE pays $280,000 for its power ($140/MWh*2000 MW). But since consumers own the allowances, they get the allowances rent ($110,000, e.g., from auctioning the allowances), so the net cost to the LSE is $170,000 or $85/MWh.

Thus, the two systems (load-based and source-based/consumer-owned allowances) result in the *same* cost to load. The "Rent to Clean Generation" in both cases accrues to Generator A (the cleanest generator). Generator A earns this rent in the source-based case because it earns the full power price without having to pay for allowances. It earns it in the load-based case because LSEs are willing to pay a premium for its power relative to higher-emissions sources.

Under the California "Standard Practice" for benefit-cost analysis of demand-side programs,[20] "utility avoided costs" quantify the utility's energy cost savings resulting from changes in load. In the models of this Appendix, this equals the per unit reduction in cost to the LSE resulting from a change in load, accounting for all cash flows. In the load-based model of Section A.1, a unit decrease in L lowers the right-hand side of the LSE's emissions constraint by K units (in tons/MWh) and the right-hand side of the LSE's load constraint by 1 unit. This results in a cost savings (change in the LSE's objective function) of $K\alpha_L$ and β_L, respectively, or a total of 0.55 tons/MWh*$100/MWh + $85/MWh = $140/MWh. In the source-based model of Section A.2, a unit decrease in L affects only the right-hand side of the LSE's load constraint (by 1 unit); since its shadow price is β_S = $140/MWh, the cost-savings to the LSE is the same as in the load-based model. Thus, the "utility avoided cost" is the same under the load-based and source-based models, as we argued in Section 4, *supra*.

[20] "California Standard Practice Manual: Economic Analysis of Demand-Side Programs and Projects," October 2001, ftp://ftp.cpuc.ca.gov/puc/energy/electric/energy+efficiency/em+and+v/Std+Practice+Manual.doc

A.5. Analysis of the Gillenwater and Breidenich (2007) (*op. cit.*) Load-Based Proposal

The load-based system Gillenwater and Breidenich (*op. cit.*) propose unbundles emissions and power so that the ISO would not have to track emissions associated with power sources. The proposal has the first seller unbundle the GHG emissions rate from power production; then the seller can sell the GHG rights to load, and power to whomever it wants. The rights are called "Generation Emission Attribute Certificates" (GEAC), have units of energy (MWh), and have the additional attribute of the actual emissions rate of the seller.[21] Thus, the GEACs are a differentiated commodity. Each load-serving entity (LSE) is responsible for buying enough GEACs to meet its load, and the total emissions associated with the GEACs it buys must be no more than the LSE's emissions limit. The idea in the proposal is captured in the following equilibrium model, consisting of a consumer model, a producer model, and market clearing conditions. There are two sets prices that clear the market: p_i (the \$/MWh price of power from producer *i*) and β (the \$/ton price of CO_2 credits implied by the trading of GEACs).

Consumer Model. The LSE has the following optimization problem. Choose (1) the amount of electricity x to buy and (2) the amount of GEACs z_i to purchase from each producer i in order to maximize net benefits of consumption, subject to regulatory constraints concerning the amount and mix of GEACs that each LSE has to buy.

$$\text{MAX} \; -\Sigma_i\, p_i\, x_i - [\Sigma_i\, \beta\, (K - E_i) z_i]$$

subject to:

$$\Sigma_i\, z_i - \Sigma_i\, x_i = 0$$
$$\Sigma_i\, x_i = L$$
$$\Sigma_i\, E_i\, x_i \leq K\, L$$
$$x_i \geq 0 \quad \forall i$$

The notation is the same as in the rest of the Appendix, with the addition of a new decision variable z_i, equal to the number of GEACs that the LSE buys from producer i. The next to last constraint says that emission-weighted GEACs can't exceed the target rate times consumption.

The pricing rule for GEACs embodied in the LSE's objective is that a GEAC from producer i would have price $\beta(K - E_i)$ \$/MWh. This is a reasonable interpretation of Gillenwater and Breidenich (*op. cit.*)'s statement that consumers are willing to pay more for cleaner certificates. The rule follows from the reasonable expectation that a consumer should be willing to pay a premium for a certificate that makes it easier to comply with its emissions constraint (*i.e.*, a GEAC whose $E_i < K$), while a consumer would have to be bribed to accept a certificate that makes it more difficult to comply with that constraint (*i.e.*, a GEAC whose $E_i > K$) and, further, the amount of payment should be proportional to the difference between E_i and K. Thus, low emission producers would be paid handsomely for their GEACs, while a coal plant might have to pay LSEs to take the GEACs off its hands.

---

[21] These are sometimes also called these "Tradable Emission Attribute Certificates."

Other pricing schemes are possible. In particular, make the price of a GEAC from producer i equal to $\beta(H–E_i)$, where H is some arbitrary or default emissions rate; if H is high enough, then all GEACs would have a positive price. However, in a closed power market in which load L is fixed, in equilibrium, this would just serve to lower the price of electricity by $\beta(H-K)$ \$/MWh; this would yield the *same* generation and consumption solution and consumer costs as using H=K. However, in a world in which L is not fixed (due not only to price elasticity, but also due to customer switching among LSEs), the pricing rule $\beta(K–E_i)$ is arguably the most sensible one in terms of ease of administration (since LSEs would then not need to be involved in the system; see below).[22] It is shown below that having $K>E^T$ would be equivalent to taxing consumption by a fixed per MWh rate, and that would be a much simpler implementation of this system than asking consumers to track purchases of GEACs.

<u>Producer Model.</u> Each producer i has problem of choosing the amount of generation y_i [MWh] in order to maximize profit.

$$\text{MAX} \quad (p_i - C_j)y_i + \beta(K - E_i)y_i$$
$$\text{subject to: } y_i \geq 0$$

If it is a clean producer, it gets paid for credits ($K - E_i > 0$), but if it is dirty, it has to pay consumers to take the credits off its hands ($K - E_i < 0$).

<u>Market Clearing Conditions.</u> There are two market clearing conditions. First, for energy, generation = consumption.

$$x_i = y_i \text{ for all i } \quad \text{(shadow price } p_i\text{)}$$

Second, the amount of GEACs produced by each producer has to equal the amount sold.

$$x_i = z_i \text{ for all i } \quad \text{(shadow price } \beta(K - E_i)\text{)}$$

The market equilibrium model consists of combining the first-order conditions of the consumer and producer models with the market clearing conditions.

<u>Example.</u> A consumer has a load of 1 MWh, and two producers are available: A, which has high emissions (E_A = 1 ton CO_2/MWh) and B, which has low emissions (E_B = 0.5 ton/MWh). The emissions rate target is K = 0.75 tons. The marginal cost of A is \$40/MWh, and B's marginal cost \$70/MWh. The equilibrium is $p_A = p_B$ = \$55/MWh, and β = \$60/ton. Producer A has to bribe consumers to take its credits, while producer B gets paid. There is only one electricity price and the ISO does not have to track different "flavors" of electricity.

Interestingly, the consumer pays *nothing* on net for its GEACs; it pays \$60*(0.75-0.5) = \$15 for 0.5 GEACs from producer B, but is paid \$60*(1-0.75) = \$15 for the 0.5 GEACs it accepts from producer A. As is pointed out below, this is no coincidence; each LSE pays \$0 for

[22] Having a higher default emissions rate ($H > K$) would result in payments, on net, from consumers/LSEs to producers. As shown later in this Appendix, having $H = K$ results in zero net payments.

its GEACs. So there is no point to having them participate in the market if another mechanism can be devised to keep track of emission rates; it turns out that one can can be easily devised that only involves producers.

Reduction of the Gillenwater and Breidenich (op. cit.) Load-Based Proposal to a Cap-and-Trade System with Free Allocation of Allowances to Producers. The above model simplifies considerably if it is recognized that the assumed pricing rule will result in consumer's emissions constraint being binding in an optimal solution. Substituting the emissions constraint into the LSE's demand constraint yields

$$\Sigma_i E_i x_i = K \Sigma_i x_i$$

which implies that the objective function term $\beta [\Sigma_i (K - E_i)x_i]$ is identically zero. This means that each LSE pays nothing, on net, for its GEACs. Thus, there is no need to have load participate in this market. The potential complications of having not only to monitor producer emissions but also track producer sales of GEACs to LSEs serves no purpose and can be avoided. This nominally load-based trading system is actually a source-based trading system with the following properties:

1. An elastic cap that is proportional to the emissions rate times total production
2. Free allocation of allowances to producers in proportion to their output

The free allocation means that producers retain the allowances rents under this system.

What if instead of pricing rule $\beta (K - E_i)$ the rule was more generally $\beta (H - E_i)$, with H being a "default emission rate" that differs from the target emissions rate K that the LSE must attain? If the equilibrium price β was unchanged (which might not be the case if demand is elastic), the difference in consumer payment compared to the objective in the above LSE model would be :

$$\beta [\Sigma_i (H - E_i)x_i] - \beta [\Sigma_i (K - E_i)x_i] = \beta (H - K)L$$

That is, this would be equivalent to taxing the consumer by amount β (H–K) per MWh; producers would receive a payment of this amount per MWh generated (assuming no losses). Note that this subsidizes energy production.[23] Therefore, there is no theoretical reason to set up an elaborate load-based accounting system to implement a system with a default emission rate $\neq$ K; one can just use an energy tax and pass its proceeds to generators, or use the energy tax proceeds for other purposes.

The above analysis, strictly speaking, only applies to a closed (no imports) system. Gillenwater and Breidenich (*op. cit.*) propose that it be applied to a system with power imports

[23] Further, assuming L is fixed (perfectly inelastic), then in equilibrium, the tax payments by consumers would be returned to them in the form of lower power prices, and nothing would be accomplished—the net costs to consumers would be exactly the same. So there would be a reason to do this only if the taxes were used for some purpose other than a subsidy to producers. (In the case of a consumer subsidy paid by producers, power prices would be raised instead.)

by allowing producers outside California to voluntarily join the system; there would be an incentive to do so if a producer's emissions E_i were less than the target H. However, this system is subject to the same difficulties concerning contract shuffling as the other systems, as we discuss in Section 2.