Chapter 5:
Lessons for a cap-and-trade program

Dallas Burtraw
Senior Fellow, Resources For the Future

Alexander E. Farrell
Assistant Professor, Energy and Resources Group, University of California Berkeley

Lawrence H. Goulder
Professor and Shuzo Nishihara Chair in Environmental and Resource Economics, Stanford University

Carla Peterman
Energy and Resources Group, University of California Berkeley
OUTLINE

1. **Introduction**  
   1.1. California’s GHG Emission Reduction Goals

2. **Varieties of Emission Trading Systems**  
   2.1. Baseline and Credit Systems  
   2.2. Cap-and-Trade Systems  
   2.3. Open Market Trading

3. **Key Issues**  
   3.1. Mandatory Caps  
   3.2. Scope: Point of Regulation, Regulated Sources and Emissions  
   3.3. Allocation: The Initial Distribution of Allowances  
   3.4. Banking  
   3.5. Linkages Between Cap-and-Trade Systems  
   3.6. Offsets  
   3.7. Price Certainty  
   3.8. Environmental Justice

4. **The Performance Record**  
   4.1. U.S. Acid Rain Program (SO₂)  
   4.2. The Regional NOₓ Cap-And-Trade Programs  
   4.3. California RECLAIM  
   4.4. Existing GHG programs

5. **Recommendations for California**  
   5.1. Mandatory Caps  
   5.2. Scope  
   5.3. Allocation  
   5.4. Banking  
   5.5. Linkage  
   5.6. Offsets  
   5.7. Price Certainty  
   5.8. Environmental Justice
1 Introduction

There is broad scientific consensus that rising concentrations of greenhouse gases (GHGs) in the atmosphere have already caused perceptible changes in climate and will lead to further climate change in the future (Intergovernmental Panel on Climate Change 2001). The impact of climate change in California may be significant for water resources, agriculture, and ecosystems (Shaw 2002; Roos 2003; Hayhoe et al. 2004). Avoiding the most serious climate change impacts will be challenging; our economy is based in processes that produce greenhouse gases. Deep (greater than 75%) cuts in global GHG emissions will be required to stabilize the climate, while at the same time people in developing countries will be using greater amounts of energy as they modernize (Wigley et al. 1996; O’Neill and Oppenheimer 2002). Thus, industrialized economies like California will have to slash GHG emissions even more, perhaps by more than 90%.

Most of the energy sources we rely upon cause substantial releases of greenhouse gases. Even the large non-energy agriculture and forestry sectors of the California economy have the potential to significantly affect net carbon flux, in part by storing carbon in soils and biomass (e.g. standing forest). No environmental challenge, as a consequence, is more difficult to tackle. Confronting climate change will require significant changes from current practices, especially in the way that energy is consumed throughout the global economy, and in the direction of innovative activity (Azar and Dowlatabadi 1999; Hoffert et al. 2002). Such changes will not come about without some form of government action. This is due to the fact that avoiding environmental protection is a public good and is thus under-provided by markets. In addition, innovation designed to achieve public goods tends to also be under-provided by markets (Arrow et al. 1995; Norberg-Bohm 1999; Rubin et al. 2004).

Avoiding dangerous climate change is a great challenge that will require effective and economically efficient government action. In considering policy responses to climate change, the heterogeneity of the causes, as well as the wide range of potential solutions, makes a source-by-source government prescription a daunting and resource-intensive task. Yet, this same pervasiveness of the problem also demands a comprehensive response. A cap and trade (C/T) system is one of the most promising options to control GHG emissions in the State of California because this approach provides a broad market signal that can guide investments and economic behavior in an efficient way while allowing for considerable flexibility in the way any given emitter might respond (Stavins 2000; Nordhaus and Danish 2003).

C/T systems have been successfully applied for several different air pollutants and are attractive for a number of reasons (Stavins 2003). First, well-designed C/T programs have provided flexibility to regulated firms while protecting human health and the environment, and simultaneously promoting important social values, such as fairness (Farrell and Lave 2004). This is not an automatic feature of C/T systems, however. Some existing C/T systems have been flawed. Some pollutants (e.g. toxics) are poorly suited for a C/T program. Fortunately, all major GHGs are nontoxic. Second, C/T systems yield cost savings by establishing a market that encourages emissions abatement where it can be accomplished the most cheaply. C/T systems provide incentives to firms to improve environmental performance, rather than prescribing technical standards.

Third, C/T systems can make environmental protection look more like an ordinary business issue to managers and can allow them to apply risk management tools, perhaps in the
form of financial derivatives. Fourth, C/T systems can be easier to implement than traditional regulation because government does not need to use intrusive inspections or detailed permitting processes that require considerable industry knowledge, which many types of traditional regulation require. Instead of inspections, C/T systems require good quality monitoring and verification, which can be relatively easy to accomplish. Finally, C/T systems provide significant incentives for diffusion of technology. These systems also provide support for technological innovation, both managerially and technologically, although by themselves C/T systems are not thought to be adequate to fuel the research that is needed to address climate change in the long term. The benefits of a C/T system vary greatly with the details of the program design and implementation.

1.1 California’s GHG Emission Reduction Goals

This report examines the potential for a C/T system as part of a set of policies to implement Governor Schwarzenegger’s Executive Order S-3-05. Signed on June 1, 2005, the Executive Order established the following GHG emission reduction targets for California:

- by 2010, reduce GHG emissions to 2000 levels;
- by 2020, reduce GHG emissions to 1990 levels; and,
- by 2050, reduce GHG emissions to 80 percent below 1990 levels.

A C/T system may be an important part of the effort to reach those goals. However, other policies will be necessary to prepare the state to develop new technologies and build experience with them. Additional policies will be especially useful in promoting the invention of new, clean, and low-cost technologies that will be necessary to meet the 2050 target.

The C/T system might well expand over time. To meet the 2050 targets, virtually all GHG emissions in California may need to be managed in some way. In the near term, however, monitoring costs justify leaving certain sectors out of the C/T system, although not necessarily outside of California’s GHG policy entirely. For example, N₂O emissions from the agricultural sector might currently be too costly to monitor. As monitoring technologies improve, it will make sense to bring additional sectors into the program. A well-designed GHG policy can provide support and incentives to solve technology and policy design problems associated with gases and sectors that are hard to include in a C/T system. Such a program should have an architecture that will enable the inclusion of additional sectors and gases into the program when it is economically advantageous to do so.

This paper provides a strategy for the near term application of a cap-and-trade system to reduce GHG emissions in California. In the next section we summarize the evolution of C/T policies. In Section 3 we survey the concepts that form the cornerstones of C/T. Section 4 reports on the empirical experience. We conclude in Section 5 with recommendations for the design of C/T aimed at achieving the Governor’s goals in 2010 and 2020, and which can evolve in ways that will be necessary to progress toward more dramatic emission reduction targets by 2050.
2 Varieties of Emissions Trading Systems

Emission allowance trading provides a policy tool that allows society to achieve environmental goals at less cost than would other prescriptive regulations. Allowance trading leads to emissions reductions by those facilities that can achieve such reductions relatively cheaply.

2.1 Baseline and Credit Systems

Current C/T systems can be seen as extensions of earlier emissions trading systems based on facility-specific baselines that provided the opportunity for facilities to operate above or below their baseline by using credits (Hahn and Hester 1989). Baseline and credit systems can be further differentiated into systems in which the credits are considered permanent (commonly called Emission Reduction Credits, or ERCs) or one-time (commonly called Discrete Emission Reduction, or DER, credits).

The first application of emissions trading was developed by economists and policy analysts in EPA’s Regulatory Reform group in 1975–76, which sought a way to permit localities violating the air quality health standards to support economic development without further increasing emissions (Hahn and Hester 1989). To accomplish this they designed an ERC system, whereby new emitting sources could pay existing sources to reduce their emissions sufficiently to offset any increase in emissions. Hence these credits were labeled emission offsets. Under this program, the local government could obtain emission reductions from among existing sources (i.e. firms within the local government’s jurisdiction) and sell or give them away to newly established businesses, thereby offsetting emissions increases with reductions elsewhere.

This approach featured bilateral trades and no aggregate cap on emissions. The government had to approve every transaction and often mediated these transactions, which is very different from the C/T system we are proposing. Nevertheless, emission offsets provided a way for localities experiencing economic growth to avoid being constrained by clean air requirements. Variations on this theme include the “bubble” policy, whereby one could imagine placing a bubble over the multiple emission stacks of a plant and permit emissions of a particular type to be traded, (more precisely, allocated) among these stacks. EPA remained concerned only about the aggregate emissions from the metaphorical bubble. Bubbles could also cover smokestacks from multiple, co-located, facilities. In the 1977 Clean Air Act Amendments, Congress recognized the offset policy in law and also made it possible to “bank” emission reductions for later use.

An important feature of ERC systems is that they are voluntary (although the underlying regulations are not). In practice, this has meant that the incentives to create allowances and put them on the market have been weak (Hahn and Hester 1989). Firms generally do not go out of their way to create ERCs, because they do not see themselves as being in the ERC business and prefer to invest capital in their core activities. Firms often want to keep ERCs to support possible expansions of their own facilities in the future rather than selling them into a market (where potential competitors might buy them).

One problem is obtaining agreement between regulators, environmental advocates, and industry about the size of emission reductions, because it is often impossible to measure emissions directly and because disputes often arise over appropriate historical baselines. Firms must prove that the ERCs are “surplus” or “additional”—that is, that the actions that create the
ERCs would not have taken place anyway (Stavins 1997). This problem besets offsets as well, as discussed below.

Further, firms may harbor the concern that by voluntarily creating ERCs, they will provide regulators with additional information that emission control costs in that industry or process are lower than might be otherwise believed, leading regulators to mandate such emission reductions (Dudek and Palmisano 1988). Unfortunately, these problems can limit the number of available emission credits, making it hard for new entrants to buy credits to support new businesses. Additionally, the limited number of programs, the terms and conditions associated with ERCs, and the relatively small markets for them have hindered the development of risk management tools (e.g., futures and options) for them, limiting their utility. One of the reactions to this problem has been for local governments to obtain ERCs as part of the process approving their creation. These ERCs can be used to foster growth (and job creation) by giving them to new companies entering the area.

In the United States, ERC systems have had limited success. They have rarely been used elsewhere (Stavins 1997; Solomon and Gorman 1998). These systems have played a role in programs for federal air pollution control, as above, and are used by a few states. In California, the experience with ERCs has been mixed. The particularly difficult air quality challenges facing the state have led to more stringent rules governing ERCs for air pollutants than elsewhere, resulting in higher transaction costs and a less efficient outcome (Foster and Hahn 1995).

2.2 Cap-And-Trade Systems

In a C/T system, the government defines the regulated sources and the total amount of allowable emissions during a set period – the cap. Typically, the cap is set in mass units (e.g., tons), is lower than historical emissions, and declines over time. Environmental advocates consider the certain emission reductions implied by the cap to be the major virtue of this approach, however, it can also be a disadvantage. Caps are typically set in political processes that balance costs and benefits, and agreements on how to set a cap can be very difficult to change, even if the costs or benefits turn out to be very different from those assumed during the original negotiations. Thus, if emission reductions turn out to be less costly or more beneficial than originally thought (there is evidence for both in the case of sulfur dioxide, or SO₂), net social benefits would increase if the cap were lowered. On the other hand, a cap prohibits emission increases even in the event of unanticipated high cost. This has led some observers to suggest designing emission caps in ways such that the cap would respond to signals about the marginal costs of the program and be adjusted accordingly.

The second major feature of C/T systems is the trade of emission allowances. Each allowance entitles its owner to emit a given quantity of a pollutant. The regulating authority decides which sources will be regulated and the size of the cap for those sources. It then creates emission allowances equal to the size of the cap and a mechanism for distributing allowances to regulated entities, called the allocation mechanism (discussed in more detail below). Then, the government requires regulated facilities to measure (or estimate) their emissions and periodically (e.g. annually) to surrender emission allowances equal to their emissions. The government will also set procedures for emissions monitoring, establish rules for how allowances may be used, and apply enforcement measures (if needed).
Through trades of emissions allowances, sources or facilities for which the cost of reducing emissions is particularly high can purchase additional emissions allowances, thereby relaxing their obligations for emissions reductions. Likewise, facilities that have relatively low costs of emissions reductions will find it profitable to sell some of their allowances. This obliges them to take on greater abatement responsibilities, but the proceeds from the sale will more than compensate for the added abatement costs. Both buyer and seller benefit from such trade, which is the feature that the business community tends to prefer. At the same time, because the market leads to greater reductions by low-cost emitters, the overall economic cost of achieving emissions reductions is lowered as well. The business community finds these cost reductions to be the major virtue of C/T systems.

Trades in emissions allowances do not change the overall amount of emissions; but they reallocate emissions-reduction efforts across firms. So long as the damages from emissions are the same regardless of the source, the trades have no impact on overall damages. In the case of GHGs, emissions from all sources are indeed equivalent because GHGs are uniformly mixing in the atmosphere.

The regulating authority controls the trading of emissions allowances differently in various C/T systems. The government usually acts as the accountant for C/T systems by establishing a registry for participants and assigning a serial number for each allowance. In some programs participants must report the size of transactions and the names of the buyer and seller, in other programs reporting transactions is voluntary, but the number of a specific allowance is identified at time of compliance. There is usually no requirement that market participants disclose the price at which a sale was made, nor any requirement that they inform government of the trade in a timely manner. This lack of information can limit the transparency of the market, as participants may delay reporting trades in order to conceal strategic information, but these characteristics are no different from other product markets. Brokerage and consulting firms complete the picture by providing services to market participants, including developing derivative commodities (e.g. options) and providing information about the markets. Simplicity in market design and competition among brokers has tended to keep transaction costs low in emission allowance C/T markets, which has helped lead to their success.

Some key features of C/T systems are worth noting. First, a cap on total emissions means that as an economy grows, new emissions-control technologies or emission-free production processes will be needed. Some observers worry that a fixed emissions cap is a limit to economic growth. While there have been modeling studies on this issue, there appears to be no analysis of the effect of existing emission trading programs on economic growth (Hoffert et al. 1998; Energy Information Administration 2001; Ono 2002). However, all existing programs to date may simply be too small to have a noticeable effect on growth, so such a study may not be feasible at this time. Significant GHG emission reductions might have far larger effects. Second, the standardization of C/T allowances allows for larger emissions markets and permits brokers to offer derivative securities based on them. This has proved important since the ability to use derivative securities like options and futures greatly enhances the flexibility a firm has in planning its operations.

The environmental performance of cap and trade programs have been quite favorable compared to the performance of more prescriptive approaches to regulation (Ellerman et al. 2003). Experience with the SO2 program indicates that the opportunity for banking led to early over-compliance and accelerated emission reductions compared to the anticipated schedule for
reductions. Implementation of cap and trade programs typically happens quickly, especially compared to the timeline required for many prescriptive policies. The SO₂ program has achieved virtually 100% compliance, as has the NOₓ trading programs. The exception to 100% compliance has been the RECLAIM program in southern California, which saw emissions in excess of the caps during the California electricity market disruptions of 2000. Those excess emissions were ultimately “paid back” through subsequent emission reductions.

Emission trading is particularly well suited for greenhouse gases because they are uniformly mixed and do not require limitations on trading to minimize “hot spots.” A key component to the success of emission trading in other programs is strong monitoring and enforcement provisions, including reasonably accurate emissions measurement, automatic excess emissions penalties that are not subject to appeal or waivers (Swift 2001; Ellerman, Joskow et al. 2003), and public access to emissions and trading data through the use of information technology and the Internet (Kruger et al. 2000; Tietenberg 2003).

One of the central issues in designing a cap-and-trade policy is the mechanism for the initial distribution of allowances to the regulated entities or potentially to other parties. This mechanism must be viewed as fair and equitable, balancing the diverse interests of the regulated entities as well as the public interest.

There are two principal methods of allocation: free allocation and auctions. In free allocation, allowances are given to the regulated entities or others at no cost based on a formula or prescribed method for determining the distribution. In an auction, the regulating authority sells allowances to the highest bidders. Of course, mixed (or hybrid) approaches can be used. The auction approach has potential efficiency advantages, related to benefits from the use of auction revenues, impacts on electricity prices, and impacts on innovation. These are discussed further below.

If an auction is not used, the allowances would be distributed for free according to some agreed-upon process. The most common way is often called “grandfathering,” which bases free distribution to incumbent firms on a historic measure of performance. The measure might be related to inputs (in electricity this is the heat content of fuel used at a facility) or to output (in electricity this would be electricity generation). Typically, the allocation takes into account facility-specific factors, for instance, mass emissions (tons per year) or emissions rate (tons per unit of output). This type of allocation is used in the SO₂ emission trading program and many others. Free distribution of allowances conveys a valuable asset to the recipient, worth potentially billions of dollars annually. Hence this approach is relatively popular with incumbent firms. However, this method produces no revenues for the government and is problematic for new entrants.

In some previous programs firms have expressed concern about the liquidity of the allowance market, and whether allowances would be available for new sources. To address this issue, a portion of the emission cap can be set aside for free distribution to new sources. Alternatively, as in the SO₂ program, a share of the cap could be set aside to be auctioned off, so that all sources, new and existing, could bid.

An important alternative to grandfathering is free allocation based on constantly updating the measure on which allowances are distributed, perhaps as a rolling average of performance.
several years previous to the year of the allocation. Variations of this approach are used in some states for \( \text{NO}_x \) emission trading. This alternative is discussed further below also.

### 2.3 Open Market Trading

There have been attempts to allow DERs to be used in C/T systems—a concept called “Open Market Trading” (OMTR) (Ayres 1994; U.S. Environmental Protection Agency 1995). Advocates of this approach typically look to create DERs in the mobile source sector (by buying and scrapping old vehicles, or paying for upgrades to cleaner vehicles) to sell to stationary sources. Although the open market trading approach has been attempted several times, it has usually failed, often due to disagreements over credit certification requirements (National Healthy Air License Exchange 1995; Goffman 1997; EPA - Office of the Inspector General 2002). For instance, a prominent OMTR program in New Jersey collapsed in late 2002 after years of development. It appears that the costs of adequate monitoring and verification for the use of DERs in C/T systems may be high enough to largely eliminate the value of OMTR programs.

---

1 For example, an updating approach in year X may depend on a measure of electricity output in year X-3.
3 Key Issues

Several major issues are involved with implementing and operating an emissions trading program, and with managing a business under an emissions trading program.

3.1 Mandatory Caps

As noted previously, there are a variety of less formal approaches to incentive-based environmental regulation that do not involve mandatory emission caps. These include offset and banking programs, open market trading programs and performance standards. There have been many examples of successful efforts using these less formal approaches, although most are characterized by costs that are well above the least cost compliance activities that have been identified in simulation models. Moreover, these programs have been widely criticized for having large transaction costs and creating uncertain investment climates. An environmental criticism of these programs has been that the emission reductions are not certain except when there is a mandatory emission cap.

The EPA’s Regulatory Reform group recognized the virtue of emission trading when imposing a mandatory emission cap when it sought ways of introducing flexibility to allow economic growth without increasing emissions in areas that were in non-attainment with air quality standards. Among all the approaches that have been tried, cap and trade programs built on mandatory emission caps have generated the most attention because the approach has successfully bridged the concerns of the environmental and business communities, achieving significant emission reductions in a way that saved costs for business.

Efforts to improve public and corporate knowledge about GHG emission reduction possibilities and encourage their voluntary adoption have great appeal in principal. They foster responsible citizenship, tend to empower those with the most detailed information about cost-effective steps, and can encourage innovation. However, the practical, on-the-ground results of education, information, and voluntary approaches have been limited, and many of the supposed benefits of specific voluntary approaches may have occurred anyway due to other forces and trends (Wilbanks and Stern 2002; Harrison and Antweiler 2003). The empirical evaluation of these programs is weak. One reviewer thought that this was a systematic problem that reflected, “...a pathology of unclear targets and the kinds of monitoring, verification, and public reporting needed to support program evaluation” (General Accounting Office 1994; Harrison 2002 p. 277). For instance in reviewing the EPA’s 33/50 program, the reviewer noted that much of the reductions in emissions occurred before the program’s inception, and that non-participants made similar improvements. Without better program review, it would be hard to understand what effect such a program had.

In a book devoted to evaluating the prospect of corporate environmental management systems (EMSs) for public policy purposes, the idea of EMSs substituting for public policy is rejected in favor of considering how the two can work together (Coglianese and Nash 2001). While the authors find that EMSs are very useful in responding to public policy, they suggest some caution before accepting EMSs as a basis for public policy. They argue that whereas “many managers find EMSs to be an important and helpful tool, public policy designed to encourage or incorporate EMSs into regulatory strategy may steer both firms and agencies in the wrong direction. EMSs may incorporate goals that represent manager’s aspirations, not society’s. This approach may distract agencies from more promising policy options. The incentives
agencies provide may encourage EMS adoption, but not the underlying managerial dedication that makes such systems meaningful. Furthermore, EMS-based policy initiatives may consume resources of firms and agencies without leading to results that would justify the commitment of these resources.” (Coglianese and Nash 2001 p. 226)

Given the importance and scale of the challenge facing California in avoiding dangerous climate change, EMS and other voluntary strategies are too weak and unreliable to form the basis for public policy. Nonetheless, real world experience shows that there are specific steps that can make education, information, and voluntary approaches more likely to have a positive impact. Such approaches should neither be ignored, nor relied upon solely. Mandatory policy such as an emission cap has to be a centerpiece of policy if the state intends to reach the Governor’s goals.

3.2 Scope: Point of Regulation, Regulated Sources and Emissions

The government has several options concerning the point of regulation (upstream or downstream), the sectors to be covered, and the gases to be included.

Upstream vs. downstream and sectoral coverage. Many analysts have noted that the ideal program for GHG trading would be economy-wide, rather than limited to specific sectors (e.g., electric power) (Kruger 2005). This is because of the prevalence of CO₂ in virtually every economic sector and the efficiencies that arise by equalizing marginal costs across the entire economy. To facilitate an economy-wide system, these analysts have argued that CO₂ emissions should be regulated “upstream” (i.e., by producers or processors of fuel) rather than “downstream” (i.e., direct emitters such as power plants and industrial facilities) to capture the largest percentage of emissions and to encompass the fewest number of sources. A hybrid system with both upstream and downstream elements is also possible (Environmental Law Institute 1997).

Since carbon dioxide emissions are largely determined by the carbon content of fossil fuels, one way to introduce an upstream system is to regulate the supply or use of fossil fuels as they enter the California economy. Under this approach, CO₂ related to petroleum use would be accounted for through caps on petroleum input to California oil refineries and storage facilities; CO₂ associated with natural gas use would be accounted for through caps on natural gas throughput to pipelines and on the natural gas input to California load-serving electric power facilities; and CO₂ stemming from coal use would be covered through caps on coal input to California load-serving electric power facilities.

In contrast, a thoroughly downstream approach might attempt to monitor economic activity on a decentralized basis and to associate emissions with certain activities and fuel inputs. A sector-based approach is somewhat intermediate. The program could include some sectors and not other sectors.

Although an upstream approach has advantages, a difficulty with this approach, many economists warn, is that the price effect of a cost based on the carbon content of fuel would have little effect on individual driving habits or on vehicle choice. Even though technology exists that has a relatively short payback period, there appear to be practical hurdles in bringing it to market. This has motivated alternative approaches to improve efficiency in the transportation sector, such as AB1493, which regulates emissions from tailpipes. This has led many observers to suggest that the transportation sector could be excluded, at least initially, from a C/T policy in the state.
**Gases covered.** In principle, a trading system should not only be economy-wide but it should consider multiple greenhouse gases to capture the widest array of cost-effective sources. For example, Reilly et al. find that inclusion of all six greenhouse gases regulated under the Kyoto agreements could provide increased emissions reduction at a lower cost, although this study does not consider the costs of monitoring the various gases (Reilly et al. 2003). When several gases are covered in a C/T system, it is necessary to determine what constitutes equivalent reductions across different gases. Under the Kyoto Protocol, this is accomplished by calculating the global warming potential (GWP) of each gas. These GWPs serve as an “exchange rate” to set equivalencies for the six gases regulated under the Kyoto agreement.

**Electricity-sector issues: generator-based vs. load-based regulation and leakage.** A concern with a regional cap-and-trade program that might cover only California or the western region, and that might cover only some sectors of the economy, is the effect that the program will have on the level of CO₂ emissions outside the region. This is commonly called emissions leakage. If the program covers the power sector and thereby raises the cost of generating electricity from fossil sources in the state, then sources outside of the State gain a cost-advantage. Many of these outside sources are fossil-fired. Thus, if the California program caused generation from these outside sources to rise, the program’s intended impact on GHG emissions would be compromised. California currently imports about 20% of its electric power, although roughly 50% of the CO₂ emissions associated with electricity consumption in California result from these electricity imports.

The choice of point of regulation has important implications for the degree of emissions leakage. In the electricity sector, two options are generator-based and load-based regulation. Under the generator-based regulation, the emissions restrictions apply to the generators. Thus, each generator’s emissions are restricted to the amount authorized by the number of allowances it owns. Under load-based regulation, emissions restrictions apply to load-serving entities (LSEs), that is, the utilities or facilities that buy power and then distribute it to industrial, commercial, and residential purchasers. The LSE is responsible for compliance and surrendering an emission allowance for each ton of emission. The generator remains responsible for monitoring and reporting its fuel use.

In the northeast states, which have been considering the Regional Greenhouse Gas Initiative (RGGI), several analyses have been conducted to anticipate the leakage problem. In the absence of the policy, that region currently imports about 18% of its power according to modeling by ICF and RFF. Analysis in RGGI has focused strictly on generator-based regulation and has shown a range of estimates about leakage, from nearly zero to nearly complete. These have depended on assumptions about policies outside the region in the U.S. and Canada, and also it depends on the expectations of firms about future climate policy and how they take these expectations into account in capacity planning. An overall estimate of leakage in RGGI might be 20-30% of emission reductions in the region (Burtraw et al. 2005). The amount of leakage is reduced somewhat if emission allowances are distributed on an updated measure of performance because this provides an incentive to expand electricity generation inside the region.

If the GHG policy places enforcement at the generator level it would be difficult to deal with the leakage problem. This is because, under the law, California could only directly regulate the generators located within the State. Hence, under generator-based regulation a California program would likely lead to considerable increases in generation from outside the State. In contrast, load-based regulation can significantly reduce the potential for leakage. Under this
form of regulation, each load-based facility would be responsible for the emissions associated with the generation of its electricity, regardless of whether this power was generated within or outside of the State. Hence, this form of regulation does not enable facilities to directly escape the regulation by purchasing more power from outside of the State.

The LSE could directly account for emissions from direct purchases from generation facilities. Some of these might include out-of-state generators. A small remaining portion of power is purchased off the spot market. One cannot directly associate an emission source with this power; rather, it is the average of sources generating in the region that are not under contract. To account for emissions from these sources an accounting procedure would be developed, presumably by a state authority. Although regulation of LSEs provides a better handle for capturing out-of-state emission sources, it does introduce a dilemma. LSEs and generators will have an incentive under the program to dedicate relatively clean non-emitting sources to contracts with LSEs and to shuffle dirtier sources to customers outside of California (Center for Clean Air Policy 2005). The contract-shuffling problem could pose a major challenge to accurate accounting of emissions and it constitutes an important potential source of leakage, but it will not directly undermine the incentives for construction of new clean facilities.

In this context, it is worth noting the relevance of the Interstate Commerce clause, which prohibits state policies that discriminate between in-state and out-of-state enterprises. Load-based regulation does not discriminate, since it treats in-state and out-of-state generators on an equal bases.

Another factor that can affect the extent of emissions leakage is the way allowances are initially distributed. If emission allowances are distributed based on an updating measure or performance, such as electricity sales under contract with in-state generators or non-emitting sources out of state, then the recipient LSE has an incentive to expand these sales in order to earn a larger allocation (Fischer 2003; Fischer and Fox 2004; Burtraw, Kahn et al. 2005).

Leakage also can occur as a result of the decision to self-generate electricity and bypass the electricity grid. Distributed generation is a generic term that is used to describe the emerging technologies that allow parties to generate electricity away from central power stations. Many observers find there are promising environmental and technological options emerging through distributed generation. However, it is also the case that these facilities may be sources of GHG emissions that would not be captured by a system that targeted just grid-connected electricity generation. This is especially important as a result of our recommendation to regulate at the sector level. In contrast, a comprehensive fuels approach to an emissions program could capture the emissions from distributed generation more easily than a downstream system that regulates on a sector-specific basis. One remedy to this problem would be to account for emissions associated with fuel use at facilities that connect to the electricity grid for back up reliability, and to require reporting of fuel use as part of providing backup reliability connections to the grid.

3.3 Allocation: The Initial Distribution of Emission Allowances

An important design choice is how to initially distribute the allowances to emitters or regulated facilities. Two options are free allocation and auctioning.

Auctioning has potential advantages in terms of economic efficiency, for several reasons. First, the revenues received from an auction can be used to finance government activities, thereby reducing the need for the government to rely as much on ordinary taxes. To the extent
that the government can do this, it will avoid some of the excess burden or distortionary cost that these ordinary taxes would otherwise produce. This efficiency-benefit from using auction revenues to finance cuts in distortionary taxes has been termed the revenue-recycling effect. (Goulder et al. 1997; Parry 1997; Goulder et al. 1999). In contrast, free allocation of emissions allowances yields no revenue. Hence it cannot yield this efficiency benefit.

Second, auctioning may lead to efficiency benefits related to electricity pricing. Retail electricity prices for most customers in California are regulated. Prices are set based on the cost of service for load serving entities. Electricity prices equal average cost but differ from marginal cost systematically. Burtraw et al. show that the approach to the initial distribution of emission allowances can have a large effect on electricity price and on the efficiency of the program (Burtraw et al. 2001). An auction approach tends to lessen the difference between price and marginal cost, while free allocation tends to amplify this difference. As with the tax interaction effect, the efficiency costs are significant. Free allocation in regulated markets can lead to social costs that are 2-3 times greater than if allowances are auctioned, simply due to interactions with regulatory institutions in the electricity market.

Finally, auctioning may stimulate greater innovation and may lead to more efficient investments in technology (Milliman and Prince 1989; Kerr and Newell 2003; Popp 2003). Real-world complexities, however, such as multiple distortionary taxes and policies, monopoly power, and differences among regulated firms complicate the issue, making the optimal choice less clear (Babiker et al. 2003; Fischer et al. 2003; Parry 2005). The same is true for the specific method used to initially distribute grandfathered allowances (those given away for free to pre-existing units), whether based on heat input, generation output, or historic emissions. Output-based allocations are likely to be more efficient, though this depends on specific circumstances (Babiker, Metcalf et al. 2003).

On the other hand, free allocation is a form of compensation to firms that are affected by the regulation because the allowances are a valuable asset. To receive these allowances for free, rather than having to purchase them, reduces the regulatory burden for the firm. Indeed, several studies show that freely allocating all carbon emission allowances to U.S. fossil fuel suppliers generally will cause those firms to enjoy higher profits than in the absence of C/T policy; and freely allocating a small fraction of the allowances may be sufficient to keep profits from falling (Bovenberg and Goulder 2001; Burtraw et al. 2001; Palmer et al. 2005). This leaves a significant portion of allowances to be given away to other affected parties, or to be auctioned with revenues directed to other purposes. These considerations reveal a potential trade-off between efficiency and political feasibility: auctioning tends to be more cost-effective, while free allocation has distributional consequences that may reduce political resistance.

The other important effect of allocation hinges on the way in which free allocation is implemented. If allocation hinges on a historic measure such as heat input or electricity production then there is no incentive at the margin to change economic behavior associated with the allocation per se; the incentive of the program is created strictly from the opportunity cost of the allowances. However, if the metric for allocation is updated over time, then the firm may recognize an incentive to change its behavior in order to change its allocation in a future period. One important aspect of this incentive is that if the firm is rewarded only for electricity sales, then it will strive to expand sales potentially to the detriment of conservation efforts. Consequently it is important that the LSE earn allowances for the provision of electricity
services, which would include both sales and “nega-watts”, that is, sales that are avoided through conservation and efficiency investments.

If LSEs are the point of regulation, then the upstream electricity generators providing power into the wholesale power market will perceive their change in relative costs under free allocation almost the same as they would under an auction. The LSE would be expected to recognize the opportunity cost in its contracting procedures. However, the LSE’s electricity customers would see different effects under free allocation than under an auction. If allowances were distributed for free to the LSE their opportunity cost would not be directly reflected in electricity price. Customers would see some change in price due to the change in the relative costs of electricity suppliers, but this would be a second order effect compared to the direct value of the emission allowances. On the other hand, under an auction electricity customers would see the full effect opportunity cost of allowances in their electricity price.

Revenues associated with either a tax or C/T program for GHGs could have significant impacts on the distribution of household income due to two other effects. First, by creating emission allowances, government creates a new asset, which has value and can be taxed. To whom that value is assigned and how it is taxed are inevitable policy choices that face governments that use market based instruments. In grandfathering schemes, in competitive markets the value (e.g., “rents”) of the allowances will ultimately accrue to shareholders. This creates the most significant effect; because stock ownership is highly skewed to the top income quintile (which owns about 60% of all shares), well-off households can actually end up better off after the imposition of a C/T policy than before, while other households are worse off and the economy as a whole is less efficient. However, in regulated markets the value of the allowances accrue to ratepayers. Second, lower-income households tend to spend a higher fraction of their incomes on energy and energy-intensive goods, increasing the burden on them directly. Thus, they tend to bear the largest burden of emissions control policy, as a percentage of income.

3.4 Banking

Banking provides inter-annual flexibility within a C/T program. The price of emission allowances is expected to rise over time simply due to pressures on the allowance market stemming from economic growth. However, it is likely that the price of allowances will rise at less than the rate of interest. That means that, other things equal, a firm is better off delaying compliance activities and putting its financial resources into some other productive investment. The main purpose of the bank in this case is to provide some insurance against price volatility in the allowance market, which itself could stem from changes in fuel prices, weather or economic activity. And in this case the steady state size of the allowance bank would be relatively small.

However, in practice the activity in the bank will depend on other factors, including specifically the initial allocation of emission allowances and the stringency and timing of emission targets. If the allocation of emission allowances will decline over time, their increasing scarcity value will add to the value of a bank. Allowance price over time may rise at less than or equal to the rate of interest, but it is unlikely to increase at more than the rate of interest. The reason, in this case, is that firms would be better off to draw on their resources to make early compliance investments in order to bank allowances. In the overall market this activity would be expected to increase until the short run price of allowances increased to a point such that the path of allowance prices over time equaled the rate of interest (Rubin 1996; Schennach 2000).
The primary role of the emission bank is not environmental because it has little effect on the pace of global warming. A key component of how GHG contribute to global warming is the long time that GHG are resident in the atmosphere. It depends very little on exactly when emissions occur, at least over the time frame of the first few years of the C/T program. Hence, the role of banking is not environmental but primarily economic.

Banking also can have a political role. The presence of banked allowances represents a financial asset for a company. The value of that asset is wrapped up in the solvency and longevity of the program. Firms that hold allowance banks are more likely to be interested in the long run success of the program because that is the only way they will realize the value of the allowances in their bank. Consequently banking can contribute to the political stability of a program, as well as to its economic performance.

Finally, it is worth noting that full inter-temporal flexibility would allow for emission borrowing as well as banking. Borrowing would allow a firm to borrow from a future year allocation of emission allowances and effectively to incur a debt that has to be repaid in the future. In principle, full inter-temporal flexibility could increase the efficiency of the emission trading program. However, in practice it also could undermine the program by creating a large debt along with calls from industry for debt relief. In effect, borrowing has an opposite political influence of banking. While the emission bank creates a vested interest in the success of the program, borrowing creates a vested interest in the program’s demise.

3.5 Linkage

Most observers appreciate that the emission reductions that are achieved in California will by themselves have little effect on the future of the earth’s climate. The success of California’s GHG program is intertwined with the success of other programs and the expansion of efforts to reduce GHG emissions to a global scale.

Linkage with other programs is a principle way in which California’s initiative can be felt beyond California’s borders. Linkage can provide expanded opportunities for cost savings when the marginal cost of emission control differs in different regions. When firms do business in different regions of the country, linkage can also provide a way for firms to rationalize their compliance activities. Most importantly, linkage can help to demonstrate the political will that may catalyze national action on climate policy. National action is preferable because it is the proper level of jurisdiction for addressing what is fundamentally a global environmental problem.

To illustrate the economic benefits of linkage with programs outside California consider Figure 5.1, which depicts the marginal cost schedules for emission reductions in two states with GHG cap and trade programs. For simplicity in this example, assume emissions from each state economy are equal and each would like to cut their emissions in half. The marginal costs of emission reductions ($/ton) are plotted for State A (blue) from left (zero emission reductions) to right (complete elimination of GHG emissions) and for State B (brown) from right to left. Because their emissions are equal in size and reductions of one half from each state are desired, emission reductions equal to the width of the Figure 5 are desired.
If the emission reduction goals are accomplished in the two states individually, quantities $Q_A^1, Q_B^1$ and prices $P_A^1, P_B^1$ will result. The figure illustrates the case where there are different control costs in the two states. When the states meet their targets individually, the emission reduction quantities are equal in the two states, but the marginal costs in State B are much higher than in State A. If the states link their programs, quantities and prices both shift to $Q_A^2, Q_B^2$ and $P_A^2, P_B^2$. Now the prices are equal, having gone up in State A and down in State B. Trade in allowances causes revenue to flow from State B to State A. The total cost of GHG emission reductions, which is the area under the marginal cost curves, will be lower in the linked case. However, the emission reductions are identical, illustrating how emission trading can maintain environmental effectiveness while improving economic efficiency.

There are numerous issues that complicate the opportunity for linkage with other programs. Perhaps primary among them is the relative stringency of the programs that are to be linked. Stringency may be perceived as the emission reductions, the emission intensity, the overall emissions, or some other metric. But from a practical perspective the only comparison that will matter is the relative price of an allowance, which in turn should reflect the marginal costs of emission reductions. If one program has relatively lower marginal costs, it would be expected to be an exporter of emission allowances and its marginal costs, and consequently the allowance price in that program, would be expected to rise. This should occur until the programs have a common marginal cost.

Consequently, for the government to establish an emission ratio that values emission allowances from different programs at different rates is essentially a guess as to the relative marginal costs of the program. In effect, this is like fixing the exchange rate for currency in monetary policy. Arbitrage should work to see emission reductions continue until the exchange rate is exactly realized.

Some features can undermine the opportunity to link programs. Some examples include topics to be addressed subsequently, including the opportunity for use of offsets and the use of mechanisms such as a symmetric safety valve to provide greater certainty in allowance prices. Both of these mechanisms open the door for the introduction of allowances in excess of allocation to regulated sources under a C/T program.
3.6 Offsets

Emission offset projects provide an opportunity for entities to purchase lower-cost emissions reductions from verified projects outside of the cap and trade system. Offsets serve as an additional tool for flexibility and compliance for entities as well as to promote economic development in the offset locales. A successful offset program requires a simple, transparent, and credible verification process. Such a verification process minimizes fraud and potential for fictional emissions reporting. Verification protocols can be project, sector or technology based. Project based protocols have been adopted by the Kyoto Protocol (Clean Development Mechanism), the California Climate Action Registry (CCAR) for forest projects, Climate Trust, and the GHG Protocol Initiative. Sector and technology based protocols are less common, but CCAR has developed a sector protocol for the California power sector and plans to develop similar protocols for the cement, gas transmission and distribution, and oil sectors.

Project based protocols have the potential advantage of increased credibility due to tailored consideration of individual project factors and merits. This degree of specification, however, has practical limitations. This process is time intensive, especially if inadequately staffed, slows done the verification process, and can result in high transaction costs and delays in program implementation. Such delays not only reduce the number of offsets available, but hamper opportunities to promote new investments, initiate learning by doing, or harvest low cost emission reductions. Such is the case with the Kyoto Protocol CDM mechanism. To date, 35 CDM projects have been approved, but over 400 projects are still awaiting approval. Calls have made for streamlining the CDM process with standardized baselines (Sharma and Shrestha; Begg and Van Der Horst 2004; Sathaye et al. 2004). Moreover, due to the arbitrary nature in which GHG emissions reductions and baselines can be determined, the more detailed project-based verification system does not guarantee increased credibility (Parkinson et al. 2001).

Since sector and technology-specific protocols are broader in scope and applicability, they do not suffer from the same implementation constraints. By streamlining the verification process, they facilitate a greater number of offsets qualifying for use. Since verification would be done at the aggregate level, there is more transparency regarding protocol design. Transparency increases because a single, sector-wide baseline-setting exercise is easier to review (Lazarus et al. 1999). Moreover, pre-determined offset protocols reduce uncertainty about the financial pay-off for investment thus reducing transaction costs. A sector-based approach also facilitates feasibility analysis for potential project sponsors, thus potentially leading to greater investment. Credibility of a sector-based approach can be enhanced by regular reviews (every three years) of the protocols. Similar to the sector based approach; there has been some advocacy for “multi-project baselines” for CDMs. Under this system, within a given sector, a project earns credits against an agreed-upon metric within the sector (e.g. the plant on the operating margin, or the plant on the built margin or some combination of these. (Kartha et al. 2004).

A sector-based system can still offer some flexibility regarding project credibility by awarding reductions at rates other than one-for-one to reflect uncertainty about the baseline. Technology based rates that are valued less than one-for-one acknowledge that many investments would have occurred in the absence of the emission trading program, although it is difficult and expensive to know which ones and to what degree. Predetermined offset rates that are quality adjusted can balance considerations of credibility, high transaction costs and implementation feasibility.
3.7 Price Certainty

A fundamental feature in the design of incentive-based environmental policy is the choice between quantity and price instruments. The C/T program is a quantity based instrument because it aims to provide certain emission reductions by capping the amount of emissions that can occur. A consequence of this strategy is that the marginal cost of emission reductions may vary in unanticipated ways, if for no other reason than variability in climate and factor markets (including fuel markets) cause the costs of economic activity to vary over time. Conversely, a price-based mechanism could set a specific cost for emissions in the form of an emission fee. This approach would provide certainty with respect to the marginal cost of emission reductions, but the quantity of reductions and actual emissions could not be forecast with certainty.

Price variability in the C/T policy has implications for the economic cost of the program and for its political viability. That adjustment could not happen in the SO2 trading program because the emission target was determined in statute and was inflexible. Similarly, if costs turn out to be higher than expected, the emission cap might be adjusted upward to relieve the unanticipated cost pressure on affected firms and on the economy (Pizer 2002). These types of adjustment mechanisms are called safety valves. A symmetric safety valve would set a floor and a ceiling for allowance prices and the number of emission allowances would be adjusted in a negative or positive direction whenever the floor or ceiling was achieved.

The idea of a price ceiling safety valve has been criticized because it is thought to undermine the incentive for innovation, because a high allowance price would reward innovation. However, a price floor safety valve would reward innovation because the price would set the minimum value for emission reductions in the future. A symmetric safety valve that provided price certainty by setting a ceiling and a floor on the allowance price would be effectively neutral with respect to the incentive for innovation.

A more challenging aspect of the safety valve idea is that it would make linkage with other state or regional programs difficult. In particular, a price ceiling safety valve would mean that the emission cap could be exceeded. This possibility would make it difficult to link with other programs that did not have a safety valve.

3.8 Environmental Justice

A primary concern from the standpoint of environmental justice is that a cap and trade system should avoid inequitable regional environmental or economic impacts or risks to public health. Due to the non-localized impacts of GHG emissions, a GHG trading system should not inherently create pollution “hot spots” – as can be a concern with other pollution trading systems. Regardless, care should be taken in program design to avoid emitters increasing toxic emissions due to incentives or requirements for GHG emissions. Moreover, a CA program should prohibit the trading of toxics, even if they are GHGs. The most important GHGs including carbon dioxide, methane, nitrous oxide, sulfur hexafluoride, perfluorocarbons, and hydrofluorocarbons are nontoxic. In addition, small new units for electricity generation, often referred to as distributed electricity generation units, may be located closer to population centers than larger centralized units and should be required to offset any increases in risks from criteria pollutants.

Transparent system design, protocols, and emissions monitoring are necessary elements for avoiding inequitable impacts. Establishing an environmental justice oversight activity for the entire statewide climate change program with subpoena power, but not enforcement, can assist in
ensuring this transparency and public accountability. In the event that there are some inequitable impacts, auction revenue can be dedicated to compensating affected communities. The inclusion of “no toxic trading” and “anti-backsliding” provisions lessen the need for project-specific environmental justice review, which could slow innovation and limit cost-saving trades. Establishing a mechanism for discovering environmental justice violations should provide an incentive for cautious and responsible action.

The economic aspect of environmental justice raises the question of the incidence of burden in the economy associated with the GHG mitigation program. Specific communities may be especially hard hit by GHG policies, especially where those communities are home to fuel mining or other economic activities that have a high level of associated emissions. In addition, the GHG policy will affect consumers through changes in the price of products and services. This impact is likely to disproportionately affect lower-income households, when measured as a share of household income, because the poor typically drive older less efficient vehicles, live in older housing units and in general spend a higher fraction of their income on energy services. The poor may be the first to feel the effects of any negative effects on the state’s economy that result from the GHG mitigation policy. However, we note also that the poor may be among those who are most adversely affected by global warming. One way to mitigate the effect on lower income households is to design a program that is as efficient as possible, thereby imposing the least possible cost on the economy in general. A second way is to use the C/T program to raise revenues that can be allocated toward softening the blow on communities that are especially disadvantaged by the GHG program.
4 The Performance Record

The previous C/T programs have addressed the concepts reviewed above in various ways. The overall experience with formal C/T programs has been quite favorable. In this section we consider that experience and apply these lessons to California’s GHG program.2

4.1 U. S. Acid Rain Program (SO2)

The best-known C/T system is the EPA’s Acid Rain Program for SO2 emissions from coal-fired power plants, based on Title IV of the 1990 Clean Air Act Amendments. The program is national in scope, there is no geographic restriction on trading, and it aims at relatively deep cuts in emissions. The aggregate level of emissions over a multi-year time horizon is fixed, but the annual emissions may vary because of the provision that allows banking of allowances for use in subsequent years. The annual allocation of emission allowances was set at two levels. Phase I began in 1995 and affected the 110 dirtiest coal-fired electricity-generating facilities, including about 374 generating units. Virtually all of the Phase I units are located east of the Mississippi River. Phase II started in 2000 and covered all other coal-fired electricity-generating facilities with a capacity greater than 25 megawatts, and smaller facilities using fuel with a relatively high sulfur content, totaling about 1,420 generating units. In addition, the allocation to Phase I sources was reduced by slightly over half at the onset of Phase II. By 2010, annual emissions were expected to represent a 50% reduction from a business as usual baseline.

The program has strict monitoring provisions for both SO2 and NOx. Importantly, facilities regulated by the Acid Rain Program must still comply with health-based CAC SO2 regulations that prevent hotspots from developing, although these restrictions have not affected the market.

The SO2 C/T program initially distributed allowances based on a historic measure. The law initially distributes emission allowances free of charge to each affected power plant unit based on its heat input during a historical base period (1985–1987), multiplied by an emissions rate calculated such that aggregated emissions equal the target emissions cap, with some modifications (Joskow and Schmalensee 1998). A small portion (2.8%) of allowances are withheld from the market and auctioned, with revenues from the auction returned to industry.

The Acid Rain Program has been a success in several ways. First, substantial emission reductions have occurred, as shown in Figures 2 and 3. Emissions of regulated sources have declined substantially since their peak in the early 1980s. In part this is due to the availability of low-sulfur coal (from the Powder River Basin) across much of the country (Ellerman and Montero 1998; Ellerman et al. 2000). However, this process was accelerated and extended by the Acid Rain Program, as emission reductions continued to occur in spite of increasing coal use. Total emissions in 1995, the first year of the program, were 11.87 million tons—25% below 1990 levels and more than 35% below 1980 levels. Although emissions from the Phase I units remained relatively flat between 1995 and 1999, emissions at the unconstrained Phase II plants rose, causing total emissions to climb up to 13.1 million tons in 1998 and 12.5 million tons in 1999. From 1990-2002, overall SO2 emissions declined by about one-third, while coal-fired generation increased by more than 20%. During the less-stringent Phase I (1995–1999) regulated

2 This review draws on a number of sources including Burtraw et al. (2005).
sources over-controlled and banked more than a year’s worth of emission allowances, and began to draw them down in more stringent Phase II, following a relatively efficient path over time (Carlson et al. 2000; Schennach 2000; Ellerman 2003).

A number of authors have examined the question of whether geographic trading of SO$_2$ emission allowances has led to hot spots or in other ways introduced emission profiles that have harmed individual communities or populations. The trading approach without safeguards against the creation of hot spots does not necessarily guarantee that some communities are not adversely affected due to trading. However, the empirical and simulation literature indicates that virtually every community in the country enjoyed substantial environmental benefits as a result of the program (Burtraw and Mansur 1999; Swift 2000; Swift 2004). The primary reason was that emission trading was not a policy goal for its own sake. Rather, it was a tool that was used to achieve the goal of important aggregate emission reductions. The cost savings associated with emission trading were part of a political compromise that enabled greater emission reductions to be achieved than would have found political acceptance if other regulatory approaches had been used (Kete 1992).

Despite its success in reducing emissions, the program has been criticized for its provision that allows units that were not covered by the program in Phase I to voluntarily participate. This provision is thought to have introduced adverse-selection, meaning that units that would have reduced emissions anyway, perhaps due to the decrease in the cost of low sulfur coal, opted into the program. It is thought this loophole led to an overall increase in emissions of one to two million tons (Montero 1999).

![U.S. Acid Rain Program](image)

**Figure 5.2. SO$_2$ emissions from sources regulated by the Acid Rain Program**

The second way in which the program has been viewed as a success is due to the significant cost savings for SO$_2$ control compared to command-and-control polices. In the first five years, emissions trading reduced compliance costs by about one-third to half, estimates of the annual savings range from $350 million to $1,400 million (Ellerman, Joskow et al. 2000). In
their econometric analysis, Carlson et al. 2000 find annual savings of $250 million during the first years of Phase I, and project savings of $784 million in Phase II, or about 43% of estimated total compliance costs under a uniform standard regulating the rate of emissions at a facility. Compared with an alternative counterfactual policy that forces a specific technology (flue gas desulfurization, or “scrubbers”) to achieve the same level of emissions, cost savings of the program are estimated to be almost $1.6 billion per year (1999 dollars). This is not to say that the cost of SO$_2$ control has been cheap. In 1995 annual costs were about $726 million, and capital costs for scrubbers in Phase I alone are estimated at $3.5 billion. Much of the savings that have been realized are not due to trading of allowances per se, but from the flexibility in compliance that allowed firms to find their own least-cost approach. Nonetheless, the allowance market has provided firms with an option in case other compliance activities are delayed or in cases of demand spikes due to weather (Burtraw 1996).

An important development in the implementation of the SO$_2$ program was that Midwestern power plants that were designed to burn high-sulfur local coal were adapted to burn low-sulfur Western fuel (from the Powder River Basin) just as it was becoming cheaper due to railroad deregulation. Sales of low-sulfur coal (defined as less than 0.6 pounds of sulfur per million Btus) increased by 28% between 1990 and 1994 while prices fell by 9%. During the 1980s, while potential acid rain policies were being debated, the near universal assumption was that coal-fired power plants in the Midwest, the most important sources of SO$_2$, would find it cheapest to control emissions by continuing to burn high-sulfur Midwest coal and use scrubbers (Schmalensee et al. 1998). This line of thinking arose partly due to technical reasons (it was thought that Western low-sulfur coals were incompatible with Midwestern boilers), partly economic reasons (transportation costs had previously precluded Western coals from being competitive in the Midwest), and partly political reasons (Midwestern legislators sometimes attempted to protect Midwestern coal mining jobs by mandating the use of high-sulfur, in-state fuels) (McCarthy 1992).

Following this line of thinking, some power plants ordered scrubbers in the early 1990s. Others had installed scrubbers already, in response to state emission control regulations. But as Western low-sulfur coal began to emerge as a better option, and as the courts began to overturn protectionist state legislation, some scrubbers were canceled (Kolstad 1990; Smock 1991; Greenberger 1992; Kuehn 1993). Nonetheless, the erroneous expectations about the need for scrubbers led to more scrubbing capacity than was needed for compliance (Ellerman and Montero 1998). The growing realization of this result drove allowance prices down sharply, from around $350/ton in the first trades in 1992, to under $100 by late 1995—well below prior predictions (See Figure 5.3 below).

Another important element of cost savings was the decentralized aspect of innovation. The C/T program provided an incentive for experimentation and learning for each plant manager. For instance, blending of coals with different sulfur contents has enabled much greater reduction than had been anticipated before Title IV. It is important to note that this capability did not emerge alone from firm-level responses but as an R&D effort led by the Electric Power Research Institute (EPRI). Another form of innovation was improvement in scrubber performance, spurred in part by the flexibility associated with allowance trading. This created a form of competition among various forms of compliance, which provided incentives to reducing scrubbing costs. Keohane (2003) and Taylor (2003) both find that abatement costs per ton of removal have fallen substantially, especially in retrofitted scrubbers installed for compliance in the SO$_2$ program. Part
of the decline in scrubber costs was due to improved performance, which enabled an increase in the utilization of scrubbed units in Phase I (Ellerman et al. 2000; Carlson et al. 2000). Increased utilization is important to reducing the average cost of scrubbing because it spreads capital costs over a greater number of tons reduced. Before the SO\textsubscript{2} program, scrubbers did not exhibit reliability rates sufficient to achieve the current level of utilization. Popp finds that the move to cap-and-trade regulation for SO\textsubscript{2} in the late 1990s was accompanied by an improvement in the SO\textsubscript{2} removal efficiency of scrubbers (Popp 2003).

![Graph showing SO\textsubscript{2} emissions and electricity generation from coal](image)

**Figure 5.3.** SO\textsubscript{2} emissions (left) and electricity generation from coal (right)

Finally, there were also exogenous technical changes that clearly would have occurred in the absence of the program. Carlson et al. show that technical improvements, including overall generating efficiency, lowered the typical unit’s marginal abatement cost function by almost $50 per ton (1995$) of SO\textsubscript{2} over the decade preceding 1995 (Carlson, Burtraw et al. 2000).

To what extent can the program take credit for the cost savings resulting from technological change? Some of these changes such as changes in transportation cost of low sulfur fuels and exogenous technical change are not directly attributable to the design of the SO\textsubscript{2} program. However, we note that every other title in the 1990 Clean Air Amendments used a prescriptive approach to regulation that does not afford firms the flexibility to choose their compliance strategy. In contrast, the SO\textsubscript{2} trading program in Title IV granted firms the flexibility to capitalize on these exogenous changes in technology and relative fuel prices.

Although there is ample empirical evidence of cost savings, this does not mean the allowance market was perfectly efficient. There is evidence that the market did not lead firms to achieve their emissions reductions at minimum cost, i.e., to perform efficiently in the early years,
and that some opportunities for cost savings were not realized. Carlson et al. find that, in the first two years of Phase I, marginal costs differed among facilities and actual compliance costs exceeded the least-cost solution by $280 million in 1995 and by $339 million in 1996 (1995$). Roughly speaking, this would erode almost all of the potential gains from least-cost compliance. In contrast, Ellerman et al. provide an ex post cost estimate that is only about 3–15% above the modeled estimate of least-cost compliance in Phase I (Ellerman, Joskow et al. 2000). Several studies point to state public utility regulations and other state laws as influences that have tended to undermine the efficiency of the SO\textsubscript{2} market, leading many firms to pursue a policy of “autarchy” (no trade) and self-sufficiency in compliance in the first years of the program (Winebrake et al. 1995; Bohi 1997; Fullerton et al. 1997; Ellerman, Joskow et al. 2000; Swift 2001; Arimura 2002). Of particular interest to state policymakers was to promote the use of high sulfur, in-state coal, even though sometimes this did not represent the least cost way to reduce emissions from the perspective of ratepayers.

A third way in which the program has been viewed as a success is in the performance of the SO\textsubscript{2} market. The time path of allowance prices is reported in Figure 5.4. The market is not overseen by financial regulators such as the Securities and Exchange Commission or the Commodity and Futures Trade Commission. Nonetheless, this market has functioned well, with few (if any) controversies or problems. There are up to several dozen trades each day, resulting in from 20,000 to 100,000 allowances trading hands each week. Several different organizations monitor the market closely, some of which publish regular (daily or monthly) reports. After the year that an allowance is issued, all vintages of allowances are priced the same, because there are no restrictions on banking, which helps smooth the operation of the market. There is also a forward market that prices allowances to allocate in future years at different prices.

A new, more stringent SO\textsubscript{2} cap was proposed as part of the Clean Air Interstate Rule (CAIR) in 2004 and was adopted in early 2005. The CAIR changed the market dynamic for SO\textsubscript{2} allowances because it applies an exchange rate at which SO\textsubscript{2} allowances can be used for compliance that changes over time, beginning in 2010, placing a premium on current year vintage allowances. Market participants began to account for this in 2004, which partly explains the increase in SO\textsubscript{2} prices seen at the end of the data presented in Figure 5.4. Concerns about meeting the CAIR cap, along with high natural gas prices, have caused SO\textsubscript{2} prices to rise even further, in the fall of 2005 the price has risen to over $1,200 per ton for a current year vintage allowance. This provides an example of how allowance markets help industry respond and plan for changes in regulation or fuel prices without causing disruptions in production or distortions in other markets.

One important feature of the program, especially from a political economy perspective, was the ability to bank allowances without restriction. This feature proved valuable to the political success of the program. Once firms had built up a bank of unused allowances, they had a vested interest in maintaining the value of those banked credits and thus in furthering the program itself. Most observers agree that the trading program met remarkably little litigation and other resistance from industry in part because of the valuable assets that industry recognized it would realize if the program came to fruition.
Another important feature of this market was the auction of set-aside allowances (2.8% of total annual allocations) starting in 1992 and 1993, several years before the first compliance year. Although the design of these early auction markets has been criticized, they were valuable because they helped with price discovery in an untested and highly uncertain market. These auctions also allowed for new entrants and the public to participate in the market. (Several school and public interest groups bought allowances in order to retire them.) This system also had an attractive program for early emission reductions, while some state programs encouraged early investment in emission control equipment (scrubbers), which made some allowances excess and created natural sellers. All of this helped the industry build up a bank of allowances before the first compliance year, reducing regulatory and price uncertainty.

In another way, however, the initial distribution of allowances in the SO₂ program has been questioned. As described previously, several economic analyses have shown that the way in which allowances are distributed initially can strongly affect the economic efficiency of cap and trade systems. An auction approach can generate revenues that can be used to reduce the need for other distortionary taxes. Goulder et al. examine the SO₂ program using both analytical and numerical general equilibrium models and find that the tax interaction effect adds 70% to the estimated program compliance costs (Goulder, Parry et al. 1997). Burtraw and Palmer estimate the failure to raise revenue to offset distorting taxes through the use of an auction squanders the savings in compliance costs that could be achieved by a flexible tradable allowance system (Burtraw and Palmer 2004). These analyses assume the electricity sector is competitive, but at this juncture this does not describe over one-half of the country, including the area where coal is the primary fuel for electricity generation. Nonetheless, this is theoretical evidence that the initial distribution of allowances can play an important role.
Several important findings about the operation of allowance markets and industries regulated by a C/T system have emerged from the Acid Rain program. First, market participation and compliance strategies have evolved, from an autarkic approach in which firms trade allowances among their own units and bank their own allowances, towards a greater and greater reliance on the market (Swift 2001). Second, allowance prices have shown considerable volatility, and the lower bound of emissions prices have been shown to be equal to the marginal cost of operating emission control devices (scrubbers) (Ellerman and Montero 1998). Third, the relatively few units that did install scrubbers increased their utilization and lowered their emissions beyond the original design specifications as a result of the incentives offered by the allowance market (Taylor et al. 2003). Fourth, there is widespread evidence of productivity improvements and innovation. Whether these innovations are directly attributable to the SO$_2$ program or not, it is clear that the design of the program gave firms the flexibility to capitalize on these changes in ways that would have been much less likely under other approaches to regulation.

4.2 The Regional NO$_x$ Cap and Trade Programs

Further insights into how best to design a GHG cap and trade program for California, and how this might develop into a multi-state program, can be gained by examining some of the state-based cap and trade programs that have been successfully implemented.

The 1990 Clean Air Act Amendments created the Ozone Transport Commission (OTC) covering the northeastern and Mid-Atlantic States (National Research Council - Committee on Tropospheric Ozone Formation and Measurement 1991; Ozone Transport Commission 1994; U.S. Environmental Protection Agency 1994; Tikalsky et al. 1995). In the 1990s the OTC states took steps that led to the creation of a regional, five-month summertime C/T program covering the electricity sector (Farrell et al. 1999; Farrell 2000). In 2004 this program was superseded by a program covering a larger region including 19 states and the District of Columbia (U.S. Environmental Protection Agency 1998; Arrandale 2000; Farrell and Keating 2002). At the national level, the NO$_x$ SIP Call is expected to lead to reductions of 22% from an annual baseline level of 5.4 million tons in 2007. National summer-season emissions are expected to fall by 40% from 2.4 million tons in 2007. In the SIP Call region, the program is expected to lead to annual reductions of 34% from projected baseline levels of 3.51 million tons in 2007 and summertime emission reductions of 62%, from 1.5 million tons to 0.56 million tons.\(^3\)

In many ways the NO$_x$ program shares features in common with the SO$_2$ program. Enforcement relies on the use of continuous emissions monitors, as does the SO$_2$ program. However, the ways in which the NO$_x$ program differs provide the greatest opportunity for learning. One important way the programs differ is the restriction on banking under the NO$_x$ program.\(^4\) The program imposes a rule that limits the number of allowances if withdrawal from

---

\(^3\) U.S. EPA 1998b, Figure 2-4 and Table 2-4. One reason these numbers are approximate is that the reductions pertain to EPA’s original program that targeted 22 states and the District of Columbia.

\(^4\) The rationale for disallowing banking and establishing such a rigid system was that the environmental problem that being addressed (regional smog) was a relatively short-term (1-5 day) phenomenon and that banking over longer time periods (several years) could result in short-term increases in emissions that would lead to unhealthy air quality and defeat the purpose of the cap and trade system. As a cumulative environmental phenomenon that does not involve toxics or criteria pollutants, mitigating climate change through the control of GHGs does not have this problem. See the discussion on environmental justice below for further elaboration on this issue.
the bank is more than 10% of the aggregate annual allocation. This effectively limits the practical size of the bank.

The other key way in which the programs differ is that the authority for implementation in the NO\textsubscript{x} program rests with the states rather than in federal legislation. The program is based on a model rule, but there are considerable differences in the state laws and regulations implementing the trading program, specifically in allowance allocation mechanisms and special programs that provided additional allowance allocations for the adoption of specific technologies.

Performance of the market differs from the SO\textsubscript{2} program because while the SO\textsubscript{2} program experienced a decline in allowance prices in the early years, there were early price spikes in the NO\textsubscript{x} programs. These spikes are attributable to the absence of unrestricted banking and the absence of a generous initial allocation as characterized the SO\textsubscript{2} program. Consequently, at the start up of the program there were buyers looking for sellers, and few sellers to be found. An additional source of market volatility in the early part of the NO\textsubscript{x} Budget Program arose from uncertainty about the effectiveness of the primary strategies for compliance, which were load shifting and small operational modifications. The market could anticipate the cost of retrofit technology, but the performance of the operational strategies chosen was relatively unknown. Eventually it was recognized that operational strategies exceeded performance expectations (Farrell 2000). Within a fairly short period of time, prices in the program stabilized in the vicinity of initial forecasts.

Overall cost of the OTC program in just the northeastern and Mid-Atlantic states are estimated to be about $1 billion (1996$) for the second phase (1999–2002). This represents an estimated $900 million in savings over the assumed alternative command-and-control approach, which assumes boiler level caps equal to its allocation under the cap-and-trade approach. There are no ex post estimates of the total abatement cost of the program as it was implemented.

In the larger 19 state region, it is estimated that annual cost of abatement controls to achieve the reductions required by the program will be $2.1 billion (1997$) (Burtraw et al. 2001). Assuming that the controls installed for seasonal compliance will be employed annually, they find that, using 2008 as an example, the program will reduce annual emissions within the region from 3.45 million to 2.43 million tons. The authors do not model an alternative command-and-control approach. Finally, we note again that the method of initially distributing emission allowances is thought to be important, but there is no ex post analysis. Goulder et al., analyzing a hypothetical NO\textsubscript{x} cap-and-trade policy, find that general equilibrium costs are substantially higher when emission allowances are given away for free than when they were to be auctioned (Goulder, Parry et al. 1999).

There has been interest in the possible creation of geographic or temporal hot spots in the NO\textsubscript{x} program as there was in the SO\textsubscript{2} program, but the evidence suggests that hot spots have not materialized. Swift finds very little emission shifting, as emission reductions in most states (especially large ones) were close to the average (Swift 2004). Moreover, Swift (2004) finds that the largest emitters prior to implementation of the program had disproportionately large reductions in emissions, suggesting that areas most greatly affected by NO\textsubscript{x} emissions have realized the greatest benefits.

Moving from spatial to temporal shifting of emissions, Swift shows the program resulted in lowering NO\textsubscript{x} emissions both in total and on high emissions days. Farrell (2003) uses more
sophisticated techniques to show specifically that average and peak emissions have been lowered in equal proportion, alleviating concerns about temporal hotspots. Given that reductions occurred both in total and on high emissions days, one may begin to draw the conclusion that cap-and-trade programs are perhaps more effective than rate-based standards in consistently reducing emissions regardless of short-term changes. However, it cannot be ignored that the rate-based standards are still effective during the ozone season and perhaps prevent particularly excessive emissions in a short-term period. Farrell suggests that this combination of regulations may be preferred to unrestricted allowance trading (Farrell 2003).

4.3 California RECLAIM

Another relevant example of a C/T program is California’s Regional Clean Air Incentives Market (RECLAIM), a C/T program for SO$_2$ and NO$_x$ emissions from industrial sources such as power plants, refineries, and metal fabricators (Mueller 1995; Lents and Leyden 1996; Klier et al. 1997; Thompson 2000; Israels et al. 2002). RECLAIM does not permit banking, since government regulators felt this would compromise its environmental integrity. The only form of inter-temporal flexibility is that some sources reporting emissions from January to December and others from July to June. Sources in these two groups, or cycles, are provided allowances that may only be used in the compliance year they are allocated. So, allowances may not be banked, at least not in a straightforward manner. Because sources in the different cycles may trade with one another, the staggered allowance allocation provides only limited opportunity for inter-temporal shifting of allowance use (emissions) if sources in separate cycles continuously swapped their allowances. This proved to be a significant problem for RECLAIM, and events in the RECLAIM program illustrate the benefits of allowing banking. Fortunately, banking is relatively unimportant for stock pollutants like GHGs, making it easier to impose effective and efficient emission trading programs.

When RECLAIM was first implemented in 1994, the cap was generous, allowing for an increase in emissions over historical levels for many sources, but it declined steadily each year, aiming at an overall reduction of about 75% by 2003. For the first several years, the RECLAIM market functioned well, with readily available allowances at low prices. However, emissions in 1993–1998 did not decline and firms generally did not make investments in emission controls, even though the cap was declining from a generous initial level. Many market participants did not perceive a need to install emission control equipment. Although the state regulatory agency amply warned participants of a looming problem, many firms were unwilling to take appropriate actions because of the limited benefits from early action, since allowances were not bankable. The result of the inability to bank allowances was that firms adopted a “just in time” emission reduction strategy, delaying investments as long as possible.

However, mostly due to poor design of the deregulation of the electricity sector in California and due to a low level of precipitation that led to low availability of hydroelectricity, in the year 2000 many less frequently used generators in the Los Angeles area were pressed into extensive service. In 2000, electricity generators were allocated 2,350 tons of credits (14% of the total allocation), purchased about 2,250 tons of credits, but emitted 1,100 tons over their total holdings (Coy et al. 2001; Ellerman, Joskow et al. 2003). By early 2000 it had become clear to even the most shortsighted that emissions would exceed allocations, which was a problem because RECLAIM had no banking provision, and prices for NO$_x$ allowances skyrocketed to over $40,000/ton. Electricity companies, which were making record profits at the time, could
Lessons for a cap-and-trade program

afford these prices, but other companies in the RECLAIM market could not. Thus, the
RECLAIM cap was broken, and several firms were significantly out of compliance and paid
record fines. This is the first time a C/T system had failed in this way. Facing significant political
pressure, the state regulatory agency decided essentially to go back to a CAC approach for
electric power plants by requiring them to submit compliance plans. This is particularly
important in that most cost savings in C/T systems come from the ability to innovate in
compliance strategy, not from buying or selling allowances. In addition, state regulators
separated power companies from the rest of the RECLAIM market and subjected them to a high
tax for emissions not covered by allowances. For other participants, RECLAIM proceeds as
before and allowance prices have moderated.

Some might claim that had a prescriptive program been in place in the Los Angeles area
the compliance difficulties of 2000 would not have occurred. But there were command-and-
control policies affecting the electricity generators in urban areas elsewhere in California. During
the disruption generators in other cities violated annual restrictions on \( \text{NO}_x \) emissions, operating
hours, and fuel use. Where allowed emissions are a function of the amount of electricity
generated or fuel used, these emission increases did not attract attention because with greater
utilization these sources could legally emit more (and without commensurate reductions
elsewhere or at another time, as would be true under emission trading). It seems unlikely that a
command-and-control policy would have responded any better to the challenges faced in Los
Angeles in 2000 (Burtraw et al. 2005).

Recently, the SCAQMD adopted provisions that bring the electricity generators back into
the RECLAIM market. In addition, the total annual allocation is expected to fall about 11.7%
from 2004 levels in 2007 and an additional 10.8% in 2008. To address spikes in allowance prices
the new changes include a provision similar to a safety valve. If allowances go above \$7.50 per
pound, the SCAQMD could decide to increase annual allocations (up to 2004 levels) in the
following year. But, unlike the safety valve, the decision to do so is at the discretion of the
SCAQMD adding some uncertainty to the market. In addition, sources may petition the
SCAQMD for an exemption from the additional allowance reductions based on their adoption of
stringent abatement controls. However, any reductions not taken from the allocations to sources
that receive the exemption must be made up by sources that do not receive the exemption.
Therefore no source knows exactly what its allocation will be starting in 2007.

Several key lessons emerge from the RECLAIM experience. First, because they force
firms to gather more information and make more decisions, MBIs may be more difficult for
firms to understand and manage than CAC programs, even if they have lower costs. This is
especially true for smaller companies, some of whom may even have increases in monitoring
costs required by RECLAIM that were greater than the savings in control costs. Second, in some
cases the optimal strategy may be non-compliance, placing more emphasis on the design of
penalties. Third, emission markets are no different from others; they are volatile (especially
when it is not possible to store the commodity, like electricity).

4.4 Existing GHG programs

The largest C/T program in the world, and the most ambitious effort to date to regulate
GHG emissions, is the European Union Emissions Trading System (EU ETS), which began
operation in January 2005. After one year of operation there is an active market in emission
allowances with several trading exchanges and firms providing services. The allowance pool
already has an annual value of approximately € 38 billion. The first year of implementation has had its difficulties, however, and foremost among these has been the initial distribution of emission allowances. As we have noted, the way that emission allowances are distributed initially provides incentives that affect economic behavior. The initial allocation is left to member states, which must develop their own National Allocation Plans (NAPs) to be approved by the EU Commission. EU rules specify that in the first period (2005-2007) only a small percentage of allowances may be auctioned, and at least 95% of allowances must be allocated for free. In the second period (2008-2012) the amount given away for free must be at least 90%.

Beginning in 2008 the EU will be faced with Kyoto compliance obligations. Individual member states may find some advantage by including facilities inside the trading program, or not doing so and regulating the sources separately. In particular, there has been controversy whether the first phase of allocation has provided too much flexibility allowing some Member States gain a competitive advantage over others. There has also been considerable debate about whether free allocations to the electric power sector created an economic windfall for the sector.

Also troublesome are the issues of reporting, monitoring, and verification of emissions. Monitoring is left largely to the individual member states, under guidelines considerably more flexible than rules used in the US sulfur dioxide and nitrogen oxides programs. Emissions are to be verified independently, either by agencies within the EU member states or by third party firms that specialize in that work and that are certified by the member states. The EU program adds a layer of complexity by allowing each member state to maintain its own registry, although nations may join together to share registries. In contrast, in the United States, even the multi-jurisdictional nitrogen oxides trading program, which includes 21 states and the District of Columbia, has one centralized registry run by the EPA to track allowances. Among the 25 different EU member states, there are dramatically different legal systems, enforcement cultures, and administrative capabilities. In most of the 10 countries joining the EU this year, environmental institutions have been weak historically.

A final source of controversy has been higher than expected allowance prices. As of November 2005, prices hovered around € 22 per allowance. Reasons for the higher prices include high natural gas prices, lack of liquidity from the CDM because of a cumbersome approval process, and delays in the start of trading by some of the new EU countries like Poland and Czech Republic, who have some of the lowest cost emission reductions. In addition, restrictions on banking between the first two phases of the program and uncertainty about the post-2012 period may be inhibiting cost-effective longer-term investments in greenhouse gas mitigation.

In the U.S., two western states, Oregon and Washington, have enacted carbon standards to reduce GHG emissions. Since 1997, Oregon has had facility based carbon standards. Specific standards include: 0.675 lbs CO$_2$ / kWh for base load natural gas plants, 0.675 lbs CO$_2$/kWh for non-base load (peaking) power plants (all fuels), and 0.504 lbs CO$_2$/horsepower-hour for non-generating energy facilities (all fuels). Oregon has not yet set a CO$_2$ emissions standard for base load power plants using other fossil fuels. Rules allow base load gas plants that have power augmentation equipment to meet both the base load and non-base load standards for the respective parts of the plant. Generating plants have the option of offsetting part or all of their excess carbon dioxide emissions through guaranteed cogeneration. Emitters can also propose carbon dioxide offset projects which either they themselves, or a third party, will manage.
Alternatively emitters can provide funds via the "monetary path" to the Climate Trust for such offsets.

In 2004, Washington enacted a carbon standard (Senate Bill 3414) that requires new fossil-fueled power plants with a generating capacity of 25 megawatts or more to offset or mitigate 20% of the CO₂ emissions the plant produces over 30 years. This requirement also applies to existing plants that increase production of CO₂ emissions by 15%. Similar to the Oregon system, emitters can pay third parties (at a rate of $1.60 / MT assuming 60% run-rate) to undertake mitigation activities. The new law provides that the fee can be adjusted biennially, but the adjustment cannot exceed 50 percent of the current rate. Although there is interest in potentially integrating the Oregon and Washington programs with California there are important impediments.

Also relevant to California is the Clean Development Mechanism (CDM), which was created by the Kyoto Protocol and intended to provide less costly options for developed countries to achieve emissions reductions through investment in developing country projects. The success to date of the program has been mixed. Although approved projects have provided less costly compliance options, the pool of available projects has been limited due to the slow approval process and high transaction costs. As we noted previously, the number of accredited projects to date is 35, but approximately 400 projects are still waiting approval. Projects with emission reductions less than 0.2 Mt CO₂e are not economically viable at current CER prices (Michaelowa et al. 2003; Krey 2005). The CDM has failed to incentivize certain types of projects - specifically, renewable electricity generation and investments in rural areas. Renewable electricity generation has been limited because the payoff of reducing GHGs with higher GWPs has been much higher to date as participants seek the easiest compliance options. Rural areas have less than the expected number of projects because the baseline consumption in these areas is low to begin with, resulting in more costly projects. There is concern that the CDM will decrease incentives for investing in renewables at home, displacing the local benefits of such technologies (del Rio Gonzalez et al. 2005). This may be less of an issue in California where an RPS is already in place, but suggests that a greenhouse gas policy alone should not be assumed to take the place of a renewables standard. Additionality has also arisen as a concern with the CDM process. If a project would have occurred even without crediting, then it is part of the baseline and should not be entitled to benefits under the baseline and credit program (Sugiyama 2001).
5 Recommendations for California

This section provides the core recommendations for a cap-and-trade system for GHGs in the State of California. We view the California program as having great value on a stand-alone basis. At the same time, we would hope that this program could serve as a catalyst for a national program.

5.1 Mandatory Caps

Recommendation: California should require reductions of GHGs under a mandatory emission cap. Meeting the Governor’s targets will require policies that are environmentally effective, economically efficient, and conducive to technological innovation. Policy design should follow best practices. Education and information programs can serve as supplements, but they cannot accomplish emission reductions by themselves.

Justification: There is no evidence that voluntary measures provide sufficient incentives for emission reductions beyond what is otherwise required. To achieve the stated goals, mandatory emission caps will be required.

5.2 Scope

Recommendation: The program should aim for broad coverage, both in terms of the parts of the economy covered and in terms of the GHGs included. All gases and economic activities that can be monitored at low cost should be included. Because of monitoring costs, it may not be advisable for the program to cap emissions of certain gases from certain sectors in the short term, but coverage can be expanded over time as monitoring technology improves and a wider range of gases can be monitored at low cost.

Justification: The wider the range of sectors covered, and the broader the array of GHGs included, the greater the opportunities for achieving the state-wide emissions targets at the lowest cost. A broad program yields numerous compliance opportunities in the short term and provides strong incentives for innovation, thereby allowing for further cost-reductions over time.

5.2.1 A sector-based approach is preferable to an upstream approach for a state program.

Justification: In principle, an upstream approach has many attractions. To the extent that all activities that lead to emissions are regulated upstream, there is no need for further regulation downstream. However, in practice the sector-based approach is superior for California’s state-level program. The sector approach would dovetail with other state policies concerning transportation, such as policies to reduce emissions of GHGs from vehicle tailpipes authorized by AB1493. In contrast, the upstream approach would lead to redundant regulation of the transportation sector. In addition, the sector approach helps prevent GHG leakage associated with coal-fired electricity.
5.2.2 In the electric power sector, emissions caps should be established based on the carbon content of the fossil fuel inputs. Electricity generators would be responsible for measuring and reporting their fuel use.

Justification: Carbon dioxide is the principal greenhouse gas that is released by this sector, and it is straightforward to determine emissions of this gas based on carbon content of fossil fuels. When and where feasible, other relevant gases (e.g. sulfur hexafluoride) released by this sector should also be covered by emissions caps in this sector.

5.2.3 In the electric power sector, the point of regulation for compliance should be all public and private load serving entities (LSEs).

Justification: A significant share of California’s electricity (approximately 20 percent) is generated outside of the state, and regulation of those generators is infeasible. If only in-state generators were covered, serious emissions leakage would result. In contrast, a load-based cap covers emissions generated both in and out of the state, thereby reducing the leakage problem. Also, a load-based cap may promote a wider range of possible mitigation options (e.g., demand-side management, energy-efficiency improvements) than would apply under a generator-based cap.

5.2.4 In the natural gas sector, emissions caps should be established based on the carbon content of the natural gas.

Justification: Carbon dioxide is the principal greenhouse gas that is released by this sector, and it is straightforward to determine emissions in this sector.

5.2.5 In the natural gas sector, all natural gas LSEs should be covered, with the exception of emissions associated with stationary sources (e.g. electric power plants and refineries) that are regulated directly.

Justification: Natural gas consumption by residential and commercial gas users accounts for over 10% of statewide GHG emissions. Including these emission sources can significantly lower the costs of the program. An LSE-based cap for natural gas is preferred to upstream regulation because it will parallel the electricity sector approach and thus yield greater consistency in the program design. In addition, a cap for gas LSEs will incorporate distributed electricity generation – that is, generation by decentralized units as well as consumption that is not managed by an electricity LSE. Otherwise, distributed generation would have an unjustified advantage.

5.2.6 Emissions caps should apply to carbon dioxide emissions, and other GHGs as feasible, from all major stationary sources, including cement makers, refineries, landfills, and other manufacturing.

Justification: Major stationary sources contribute an important share (approximately 20 percent) to the state’s GHG emissions. Including these sources in the program is administratively feasible and their inclusion would lower significantly the cost of meeting the Governor’s targets. Moreover, including these sectors offers additional flexibility in the timing of investments.
5.2.7 Transportation emissions should be excluded from a cap and trade program.

Justification: A cap for individual vehicles is impracticable because of difficulties of monitoring mobile-source emissions. Moreover, allowance costs that are likely in a combined (transportation plus stationary) system are likely to be far too low to affect automaker or consumer decisions (either in vehicle characteristics or vehicle use), making such a system largely ineffective in reducing transportation sector GHGs. In addition, emissions from the transportation sector will be limited (although not capped) by the policies to reduce emissions of GHGs from vehicle tailpipes authorized by AB1493, which will affect vehicle technologies by introducing lower GHG technologies. Because such a large fraction of other transportation modes (e.g. air, rail, shipping, ground freight) are interstate or international, it would be difficult to control them.

5.2.8 To facilitate making the C/T program broader over time, the program should include incentives to develop improved monitoring technologies.

Justification: Rewards for improved monitoring technologies can help ensure that, over time, technology will improve to reduce monitoring costs, thereby making it advisable to include gases or economic activities not originally covered by the cap.

5.2.9 The cap and trade program should establish an “on-ramp” for GHG emissions that are not covered in the initial program, and the remainder of the regulatory system should provide incentives to develop the technologies (e.g. monitoring) and procedures to enable all GHG-emitting economic activities in the state to take advantage of the efficiency and flexibility of the allowance market.

Justification: While all GHG emissions contribute to climate change, not all sectors of the California economy are large enough, have adequate emissions monitoring capabilities, or have reasonable near-term options for reducing their emissions to warrant inclusion in the initial program. However, a cap and trade program can lower the costs of meeting overall emission reduction goals for the entire economy while providing significant flexibility for individual firms and facilities. Therefore, a cap and trade system should work with other parts of California’s GHG policy to enable and encourage all gases and all sectors to eventually participate.

5.3 Allocation

**Recommendation:** The initial distribution of emission allowances must balance efficiency and equity concerns.

5.3.1 The initial distribution of allowances should embrace both efficiency and equity dimensions by involving both an auction and free allocation.

Justification: The auction approach offers greater efficiency benefits, since it yields revenues that can be devoted to cuts in pre-existing distortionary taxes and electricity prices that are closer to society’s opportunity costs. This means that an auction will minimize the total cost of meeting the Governor’s GHG reduction goals. On the other hand, free distribution of a portion of allowances can help reduce the burdens on regulated facilities or electricity consumers, depending on how they are distributed and to whom. The
combination of auctioning and free allocation offers a reasonable balance of the competing goals of efficiency and equity.

5.3.2 The free allocation should be based primarily on a recent measure of electricity output or its thermal equivalent (to account for cogeneration) and the measure should be updated regularly (e.g., rolling averages of the previous 3 years). A portion of the free allocation could be based on a historical measure.

Justification: The initial distribution of the free portion of the allowances should be based on updating measures. Updating gives rise to lower consumer costs than allocation based on historic emissions or output, and is expected to lessen the leakage problem. However, a historical approach to allocation (i.e., grandfathering) of some allowances may be justified as compensation to severely affected firms.

5.3.3 Over time, the role of the auction should grow at a predetermined rate.

Justification: The portion of allowances that are distributed for free should be higher at the start of the program to reduce initial cost increases while giving the economy time to realign investments and processes to reduce carbon emissions. Over time, meeting the Governor’s targets will require that GHG emission reductions become deeper and occur across a broader portion of California’s economy. As this occurs, the overall efficiency of the program will become increasingly important, leading to the increased importance of auctioning. A pre-determined rate of growth in the relative share of allowances distributed by the auction will allow parties plan for this transition.

5.3.4 There is no need for a set-aside for new sources.

Justification: Unlike allocation based on historic measures (grandfathering), which favors existing sources, auctions and free allocation based on updated measures, as well as offset provisions, allow all sources an opportunity to obtain allowances at equal cost.

5.3.5 The initial distribution based on updating should account for an LSE’s new investments in energy conservation and allocate emission allowances in an equivalent way to the LSE’s efforts at demand reduction and to its overall share of electricity consumption.

Justification: Using updated measures for the initial distribution provides an incentive to expand electricity sales. Investments that provide energy services in lieu of electricity generation through conservation and efficiency should qualify for having displaced electricity sales.

5.3.6 Auction revenue should be dedicated to compensating affected communities, supporting technological innovation and diffusion, and reducing pre-existing taxes in the state.

Justification: Compensating communities for job losses or other significant economic effects, especially in the short run during the initiation of the program, would appeal to standards of fairness and would help to build political consensus. Reducing distorting taxes and investing in research, development, and deployment of low carbon technologies can address efficiency goals. These investments can help make the entire California economy more competitive, especially in low carbon technologies, potentially creating new export products.
5.4 Banking

Recommendations: Unlimited use of banked allowances should be a central design feature of the California program.

Justification: Banking reduces the overall cost of emission reductions and helps to avoid short-run volatility in allowance prices.

5.4.1 Firms should be encouraged to accumulate an emission allowance bank through early emission reductions. Generous credit should be granted for early emission reductions.

Justification: The accumulation of an emission bank vests interested parties in the longevity of the program because it gives them an asset that has value only if the program comes to fruition. In addition, given the challenges of maintaining economic growth in a carbon-constrained economy, banking has the highly desirable property that it accelerates technological diffusion and experience with low-carbon technologies.

5.4.2 The State should defend the legitimacy of banked emission allowances when linking the California program with other programs and its potential transition to a national program.

Justification: Certification of early reduction credits and commitment to the solvency of those credits in either a state or federal program rewards industry in California for reductions and places them in an advocacy position with respect to a national program, within which their investments retain value.

5.4.3 Borrowing should not be allowed.

Justification: Borrowing has the opposite effect of banking in that it allows the delay of diffusion of technology. It may also create pressure for future waivers of the emission cap through relief of debt borrowed against the cap. Better alternatives to borrowing are (i) to allocate generously in the first years of the program and provide incentives for accumulation and maintenance of an allowance bank, (ii) to allow offsets to be used, and, (iii) to link the California system to other GHG cap and trade systems.

5.5 Linkage

Recommendations: California should facilitate linkage of its GHG cap and trade program with others and promote symmetric treatment in buying and selling allowances.

Justification: Interaction with other programs could expand opportunities for cost savings and innovation by allowing investments in GHG reductions to flow freely across state boundaries. Interaction could also offer a market for Californian GHG emission allowances and help build political momentum for national policy.

5.5.1 Only other programs that have mandatory emission caps with a positive price for allowances should be considered for linkage.

Justification: Linkage with voluntary programs will erode investor confidence in the solvency of the California allowance market. Offset provisions should be used if linkage with voluntary programs is to be considered.
5.5.2 California should develop explicit conditions for linkage with other programs. These conditions would include common protocols and requirements for monitoring and enforcement.

Justification: The interaction with other programs could erode confidence in the value of compliance activities and investments in California if careful protocols for monitoring and enforcement are not maintained.

5.5.3 Only allowances from programs that have no safety valve, or a safety valve price greater than the expected market equilibrium price in a linked GHG allowance market, qualify for use in the California system.

Justification: A safety valve price that is less than the price at which allowances trade in California will cause emissions to effectively exceed the cap (and therefore fail to meet the Governor’s targets) while providing less incentive for innovation and deployment of low-GHG technologies.

5.5.4 Linkage with programs that have different provisions for offset credits should be preceded by a negotiated attempt to establish consistent common policies. A less desirable alternative would be to establish trading ratios.

Justification: The Commerce Clause would make it difficult to preclude the use of equivalent emission allowances from two state programs, and a trading ratio might be able to make two somewhat different allowances equivalent. However, the trading ratio undermines the simplicity and transparency of the C/T program and should be avoided if possible.

5.6 Offsets

Recommendations: Emission offsets provide an opportunity for cost-savings and economic development, and thus should be included in the program. However, the program should establish conditions for such offsets that reduce the prospects for fictional emissions reductions and inefficient revenue transfers.

5.6.1 Accreditation should be sector-specific or technology-specific. The accreditation of project specific investments can be a supplement but should not be the core of the offset program.

Justification: The major difficulty with offset programs is in establishing an emission baseline against which emission reductions can be measured. Evaluation and accreditation of individual projects on a project specific basis imposes substantial transaction costs for all parties and undermines the opportunity to promote new investments or harvest low cost emission reductions.

5.6.2 Simple and transparent criteria should be established to accredit offsets. Credit for offset investments should be pre-approved. Characterization of qualifying investments and their offset rate should be regularly reviewed every three-years.

Justification: The accreditation of sector-specific or technology-specific investments can accelerate investments and the realization of emission reductions from the program. Accreditation should be aimed at generic categories of investment rather than specific
projects. Predetermined offset values will reduce uncertainty and lower transaction costs for investors.

5.6.3 Offsets from credited projects should not be limited in quantity, but they can be adjusted for quality. Credits may be awarded for emission reductions from broad categories of technology or investment at rates less than one-for-one.

Justification: The offset program provides an incentive for innovation and investment in clean technology and provides an incentive to harvest low-cost emission reductions outside of the capped sector. Offset values can vary from one-for-one, as described in the text of this report, to reflect the uncertainty about the emission baseline. Technology based rates that discount the emissions reductions from offset projects acknowledge that many such projects would have occurred anyway even without the emissions trading program (a problem sometimes called “additionality”). Trying to determine the magnitude of this problem and accepting only projects that are additional to what would have happened anyway is difficult and expensive. On the other hand, pre-approval provides an element of certainty that can help projects overcome financing hurdles. Predetermined offset rates that are quality adjusted can balance these considerations. The extent to which out-of-state offsets should be incorporated is not clear, but a limitation on this approach might be considered (a concept sometimes called “supplementarity”).

5.7 Price Certainty

Recommendation: California should not use safety valves.

Justification: The presence of a safety valve ceiling on the price of allowances would undermine the cap, incentives for technological innovation, as well as the ability to link the program with other programs. The three flexibility mechanisms already included here -- banking, linkages, and offsets -- provide substantial opportunities to guard against the risk of high allowance prices while encouraging innovation.

5.8 Environmental Justice

Recommendations: The program should be designed in a manner that avoids inequitable regional impacts or risks to public health. Therefore, the program should:

5.8.1 Prohibit the trading of toxic GHGs. (Note that the most important GHGs – carbon dioxide, methane, nitrous oxide, sulfur hexafluoride, perfluorocarbons, and hydrofluorocarbons – are nontoxic.)

Justification: The trading of toxics can cause hotspots or disproportionate health impacts that are counterproductive and unfair.
5.8.2 Include provisions to prevent sources of toxic emissions from avoiding mandated reductions of, or causing increases in, toxic emissions due to incentives or requirements for GHG emissions.

Justification: There should be no tradeoffs accepted between public health and climate change mitigation, both are important goals and any potential conflicts should be minor and must be solved rather than negotiated around.

5.8.3 New distributed electricity generation units must offset any increases in risks from criteria pollutants and should not be permitted to emit measurable amounts of toxics.

Justification: Small-scale distributed generation units may be located closer to population centers than larger, centralized units, and will tend to have exhaust plumes that do not disperse as rapidly, yet cogeneration (or combined heat and power) applications may reduce net GHG emissions. The differences in the health effects of mass emissions should be accounted for.

5.8.4 Establish an environmental justice oversight activity for the entire statewide climate change program with subpoena power, but not enforcement.

Justification: There may be significant environmental justice issues associated with climate change, especially in the area of impacts. If the “no toxic trading” and “anti-backsliding” provisions are included there is no need for project-specific environmental justice review, which could slow innovation and limit cost-saving trades. However, establishing a mechanism for discovering environmental justice violations will provide an incentive for caution and responsible action.
References


