RAP®

DECEMBER 2017

Smart Non-Residential Rate Design

Optimizing Rates for Equity, Integration, and DER Deployment

Carl Linvill, PhD, Jim Lazar, Max Dupuy, Jessica Shipley, and Donna Brutkoski

Executive Summary

According to the Energy Information Administration, electricity use by non-residential customers accounts for nearly 66% of California's total consumption. Many of these customers are interested in adopting distributed energy resource (DER) technologies and many have access to sophisticated energy management and load control technologies, which means that these customers can be an important grid support resource. All utility customers stand to benefit if non-residential customers support a reliable, clean, and least-cost grid.

Current non-residential rate design, however, does not adequately encourage the deployment and use of non-residential customer resources in support of grid needs. Instead, current rate design encourages customers to control their own bills without synchronizing their consumption and production with the situation on the grid. Getting rate design right will ensure that price signals conveyed to the customer reflect what the power system needs. In other words, non-residential customer resources will become an important resource for integrating renewables and ensuring grid reliability. Well-designed price signals will induce cost-effective use of energy efficiency, self-generation, and demand response for the benefit of both the non-residential customer and all customers. Effective price signals will increase supply, decrease demand, and thus decrease market clearing prices for energy, capacity, and services.

Many California businesses, educational institutions, and city and county governments have commitments to the state's decarbonization goals, some have made commitments that go beyond state-level mandates. With well-designed rates, these leaders will have an economic incentive to make private investments that serve the public interest. These non-residential customers could become a large and beneficial contributor to least-cost, reliable decarbonization in California, but aligning their private choices with the public interest requires rate design reform.

Ascending clean energy technologies and aggressive California policy are changing the power system from one where we focused on ensuring adequate supply to meet anticipated demand, to one where active supply and active demand are optimized to ensure balance. Smart grid innovations allow utilities and customers to make more granular decisions about their energy use, while new storage technologies (including thermal storage) offer unprecedented opportunities to absorb variable energy production and shift usage. Wholesale markets are increasingly open to DER aggregators, introducing new value streams to customers who invest in DERs. Load control technologies and new end uses for electricity, especially, add new opportunities for system flexibility. For the past 125 years, the electricity industry has focused on controlling resources to match varying loads. In this new landscape, the challenge is increasingly to ensure that the power system is able to use demand and supply resources together to ensure reliability at least cost.

Rate design needs to embrace these changes—ensuring that customers have incentives to shift or control load and DER production when it benefits the system. Time-varying pricing (TVP) rate designs are necessary to better align private choice with the public interest.[1] Dynamic pricing options such as critical peak pricing (CPP) further refine price signals and are easy for customers to understand. More complicated dynamic rates like real-time pricing (RTP) can further refine price signals but require more sophisticated energy management, so are likely to be of interest to those organizations that have or hire sophisticated energy managers.

California's existing rate design evolved over decades, and transmission and distribution rates in particular have not been generally been updated to reflect the profound changes in customer loads, metering technology, and DER technologies. California is making changes like adapting the time-of-use (TOU) peak periods to match solar impacts, establishing default TOU rates, and encouraging movement toward coincident demand rates from non-coincident demand rates. Despite these interesting steps forward, California non-residential rate design has room for further improvement.

RAP's *Smart Rate Design for a Smart Future*[2] undertook an extensive discussion of residential and small commercial rate design, and identified three principles that should, in our opinion, apply to all customer classes:

- Principle 1: A customer should be able to connect to the grid for no more than the cost of connecting to the grid.
- Principle 2: Customers should pay for grid services and power supply in proportion to how much they use and when they use it.
- Principle 3: Customers who provide services to the grid should be fairly compensated for the value of what they supply.

In this paper, we propose smart non-residential rate principles that build off of these three. We propose:

---

[1] Throughout this paper we use the terms time-of-use (TOU) and time-varying pricing (TVP) to mean a volumetric price per kilowatt-hour that varies across the day, and that recovers both relevant capacity costs and relevant fixed and variable energy costs incurred to provide service in each time period.

[2] Jim Lazar and Wilson Gonzalez, "Smart Rate Design for a Smart Future," Regulatory Assistance Project, 2015, http://www.raponline.org/knowledge-center/smart-rate-design-for-a-smart-future/

- Non-Residential (NR) Principle 1: The service drop, metering, and billing costs should be recovered in a customer fixed charge, but the cost of the proximate transformer most directly affected by the non-coincident usage of the customer, along with any dedicated facilities installed specifically to accommodate the customer, should be recovered in a NCP demand charge.
- NR Principle 2.1: De-emphasize NCP demand charges except as noted in NR Principle 1. All shared generation and transmission capacity costs should be reflected in system-wide time-varying rates so that diversity benefits are equitably rewarded.
- NR Principle 2.2: Shift shared distribution network revenue requirements into regional or nodal time-varying rates. This recognizes that some costs are required to provide service at all hours, and that higher costs are incurred to size the system for peak demands.[3]
- NR Principle 2.3: Consider short-run marginal cost pricing signals and long-run marginal cost pricing signals together in establishing time-varying rates for system resources.
- NR Principle 2.4: Time-varying rates should provide pricing signals that are helpful in aligning controllable load, customer generation, and storage dispatch with electric system needs.
- NR Principle 2.5: Non-residential rate design options should exist that provide all customers with an easy-to-understand default tariff that does not require sophisticated energy management, along with more complex optional tariffs that present more refined price signals but require active management by the customer or the customer's aggregator.
- NR Principle 2.6: Optimal non-residential rate design will evolve as technology and system operations matures, so opportunities to revisit rate design should occur regularly.

RAP applied these principles to evaluate existing commercial rate designs at each of California's investor owned utilities. We found that if rate design is not changed to better align with these principles, California will continue to see underinvestment in DER resources and under-utilization of DER resources toward meeting California's policy goals.

RAP searched for rate design examples that better comport with these principles in California and elsewhere. The non-residential rate design we found that best comports with the principles and elements we have described above is that of the Sacramento Municipal Utility District. SMUD's non-commercial rate has a fixed charge to recovery customer-specific costs of billing, collection, and customer service; a site infrastructure cost ($/kW) to recover location-specific capacity costs; a super-peak demand charge ($/kW) to recover marginal T&D capacity costs associated with oversizing the system for extreme hours; and a TOU energy cost to recover all generation costs and

[3] One California municipal utility, for example, has TOU rates for commercial customers that include weekends as off-peak, but for residential customers, summer afternoons remain on-peak due to distribution system capacity constraints on residential circuits. The same concept could apply in different regions or nodes of a distribution system serving non-residential customers, where capacity constraints are reached at different times of the day or year.

remaining T&D costs. SMUD's rate sets it apart as an industry pace-setter, but we believe their rate design can be improved further.

One important goal for revision of non-residential rate design should be to better adapt to the incorporation of customer resources, such as thermal or electrical storage, customer provision of ancillary services through smart inverters, and customer load control for peak load management. The general framework of the rate design we propose directly compensates many of these through simple, clear, and compensatory TOU rate elements:

Table ES-1. Proposed Illustrative Rate Design for Non-Residential Consumers

	Production	Transmission	Distribution	Total	Unit
Metering, Billing			$100.00	$100.00	Month
Site Infrastructure Charge			$2/kW	$2/kW	kW
Summer On-Peak	$0.140	$0.020	$0.040	$0.20	kWh
Summer/Winter Mid-Peak	$0.100	$0.015	$0.035	$0.15	kWh
Summer/Winter Off-Peak	$0.070	$0.010	$0.020	$0.10	kWh
Super Off-Peak	$0.030	$0.010	$0.010	$0.05	kWh
Critical Peak	Maximum 50 hours per year			$0.75	kWh

This design is generally similar to SMUD's, with three important differences. First, it is unbundled between generation, transmission, and distribution to enable more granular application. Second, rather than have a super-on-peak demand charge, those costs are reflected in a critical peak price for up to 50 hours per year. The amount recovered is similar to that for SMUD's super-peak demand charge, but converted to an hourly rate to directly track high-cost hours and to enable better customer response as system conditions change. Third, we have introduced a super off-peak rate, consistent with the recommendation of CAISO. We have intentionally left the definition of time periods unstated, as these will be specific to particular utilities and particular nodes within each service territory, and will change over time as loads and resources evolve.

RAP also reviewed a number of real time pricing tariffs and, while we did not identify one in particular that we would classify as best practice, we did identify lessons learned from Texas, Illinois, Georgia, and Maryland that will be useful to the CPUC as it considers RTP optional tariffs. We suggest designing an RTP option that builds from our TOU plus CPP recommendation, and propose the following simple initial design:

- A wholesale energy cost component, charged on a per kWh basis, that fluctuates hourly. This would be based on the relevant CAISO zonal locational marginal price and would replace the "production cost" component of our recommendation above.
- Transmission costs and distribution costs would be collected in the same way that they are

collected under our recommendation above, as would any generation capacity costs that aren't accounted for in wholesale rates.

Note that this design would not achieve the full benefits of an ideal RTP approach. In particular, this would not include comprehensive price signals reflecting conditions on the local distribution network. Instead, the hourly pricing innovation here is increased exposure of end users to existing CAISO wholesale prices. Over time, as California introduces new approaches that animate the value stack for resources at the distribution level, new rate designs will be able to incorporate more complex and comprehensive RTP components.

Table of Contents

I. Introduction

Designing non-residential rate tariffs well is essential to California meeting its policy objectives at least cost. The consumption price signals conveyed through tariffs inform consumer choices on consuming and conserving energy and on investing in energy efficiency measures. If prices are too low when supply or delivery capacity is constrained, non-residential customers will conserve too little and invest too little in energy efficiency. If prices are too high during periods of oversupply, non-residential customers will forestall beneficial load shifting and beneficial electrification that could have helped with renewable integration. Since about two-thirds of all electricity consumption happens in the non-residential sector in California, failing to get non-residential prices right will make achieving California's carbon reduction, clean energy, energy conservation, and energy efficiency goals noticeably more expensive than necessary.

The price signals conveyed through tariffs also communicates the value of producing distributed energy resources (DER) energy and services. Getting these price signals right can encourage private investment that addresses distribution and bulk electric system needs. Some non-residential customers stand ready to help California meet its goals, so failing to get these price signals right will deprive the state of significant private investment. Private sector innovators like Stone Edge Winery, the signatories to the Corporate Energy Buyers Principles, and cities represented by the Climate Mayors have committed capital toward the clean energy future that can help California achieve its goals at least cost by leveraging private investment.[4]

Consumer and prosumer choices also affect needs on the distribution and even the wholesale electric system and affect the system costs required to ensure sustained reliable service. Utility and system operator assessments of system need and evaluation of alternatives to address identified needs are affected by non-residential consumer choices, which in turn are affected by the price signals embodied in the tariffs.

California's aggressive carbon reduction and clean energy goals have depended on and will continue to depend on wise public and private consumption, production, and investment choices to keep the cost achieving California's goals as cost-effective as possible. The fundamental goal of rate design is to align prices and costs, so that customers who use more and cause system costs pay for what they use, and those who constrain their use and reduce system costs receive appropriate savings.

Section II explains that while the fundamental goals of designing non-residential tariffs have not changed, underlying technological and policy changes have shifted far enough that fundamental changes in tariffs are necessary. Existing tariff designs generally do not support current technological capabilities as well as they could, do not reflect the changing needs of the electric system as well as they could, and they do not support wise public and private investment choices toward meeting California policy goals as well as they could. The effective adoption and use of

[4] See "Sonoma Vineyard Evacuated in Recent Wildfires Highlights Microgrid Benefits," *California Energy Markets*, November 3, 2017, p. 3; Corporate Renewable Energy Buyers' Principles (http://buyersprinciples.org/for-energy-buyers/); and the Climate Mayors website (http://climatemayors.org/) for stories of how private clean energy investment is being deployed and could potentially support the grid..

ascending technologies is creating a space for more cost-effective implementation of California's decarbonization and clean energy policies in the power, buildings, and transportation sectors, but rate design needs to be changed to support their adoption.[5] Section II provides several concrete examples that demonstrate how rate design needs to change to align with current and emerging technologies and policies.

Section III builds on Section II and presents a set of principles and regulatory policy recommendations to guide non-residential rate design. These principles and policy recommendations are the core of the paper, and we apply them in Section III to propose a prototypical TOU rate design and a RTP rate design that adhere to the principles and recommendations.

In Section IV, we turn to exploring how current rate designs are counterproductive to achieving California's goals. We contrast the current rate design landscape in California with rate designs that reflect the principles. We call out problems with current rate designs and lessons that can be learned from SMUD's relatively well-implemented non-residential rate design. We then provide examples demonstrating how current rate designs adversely affect several important DER technologies.

In Section V, we turn to a number of examples from around the country that offer important lessons in the possibilities for better rate design. California is a leader in implementing policies that support DER adoption and power sector carbon reduction. Nonetheless, it is useful to observe some successful innovations from other parts of the country and the world. The examples presented reinforce the principles and recommendations that conclude the report.

Section VI concludes with some summary recommendations on implementing the principles in California and presents two prototype rate designs.

[5] We introduce the term to encompass emerging technologies (i.e., new technologies that are not yet deployed but have future promise) and technologies that have been successfully deployed but will assume much broader adoption over time). Ascending technologies are collectively the transformative technologies that are moving the power sector toward a multi-way transactive sector. The text box in the next section provides examples that distinguish emerging from ascending technologies.

II. Rate Design Foundations, Ascending Technologies and California Policies

Rate design needs to change to align price signals with current technological and policy realities. This is not to say that rate design principles developed over the last 60 years do not provide useful lessons. So before turning to how ascending technologies and California policies are driving changes in rate design, we offer a very brief survey of the time-honored principles of rate design. This section then turns to highlighting how changes in technology are driving changes in rate design and a section that looks at how California policies are also driving change. This section is intended to set the stage for a discussion of rate design principles in the next section.

Rate Design Foundations

Rate design has multiple goals, including the fundamental goal of communicating cost information and aligning cost causation with prices, but also preventing price discrimination that would not occur in competitive markets, ensuring rates are fair within and among customer classes, ensuring the utility a reasonable opportunity to recover their allowed revenues, and supporting other regulatory and policy goals. Attaining these goals requires that cost allocation among classes of customers is done correctly, according to sound economic principles, and that alignment of cost causation with price paid is carefully considered. These goals sound straightforward, but their practical implementation has never been easy. For example, the terms “cost-based” and “cost-reflective” pricing have many meanings. There are as many ways of determining utility cost as there are analysts doing cost studies. And within each major approach to cost determination—including embedded cost studies, marginal cost studies, and incremental cost studies—there are many different methods used.

Fortunately, several authors of seminal texts have proposed regulatory and rate design principles that have stood the test of time, and they can guide us today in seeking to align prices and costs. We have provided a brief appendix summarizing their contributions and directing readers to useful literature reviews. For example, the observation that “the single most widely accepted rule for the governance of the regulated industries is to regulate them in such a way as to produce the same results as would be produced by effective competition, if it were feasible,” helps to guide our thinking even in the far more complex world we face today.[6] However, the fundamental task of this paper is to review these regulatory and rate design principles in light of the fundamental changes we see today.

Public policy and increasing customer demand for clean energy are affecting the mix of resources on the electric system and changing what it means for supply and demand to be in balance. Traditional rate design presumed that supply needed to chase and meet demand. However, the ascendance of variable renewable energy technologies and DER technologies means that (1) both supply and metered demand have become much more time-variant, (2) DER placement and

---

[6] Alfred Kahn, *The Economics of Regulation*, 1988. p. 17.

operation have location-specific effects on the distribution system that need to be accounted for, and (3) rate design now must recognize that supply and demand resources can be optimized to ensure system balance and cost minimization. Traditional rate design principles are still useful, but they have different implications in light of these changes. In this paper, we focus on how non-residential rate design needs to change to account for greater time-varying supply and demand and to support the optimization of system supply and demand.

The next two subsections summarize how ascending technologies and California policies are creating a need for fundamental change in non-residential rate design.

Ascending Technologies and Their Impact on Rate Design

Ascending technologies driven by technological advances, changes in the market structure of the industry in the western states, and changes in the expected end uses of electricity are combining to make the next decade one of likely monumental innovation.

Smart grid innovations and “big data” are an important element of this, allowing both utilities and consumers to have much greater understanding of their consumption and greater ability to make more granular decisions on production, consumption, storage, and conservation. Innovation in renewable energy has brought the cost of new wind and solar generation below the total costs of new natural gas power plants, and will likely soon fall to the level of the operating costs of existing plants.[7] Improvements in electricity storage technology, and deployment of new technologies for thermal energy storage, offer an opportunity to absorb variable energy production and to shift energy usage as never before.

Market changes are equally seismic in scale. The integration of the western energy grid through the evolution of the Energy Imbalance Market is offering the entire western region an opportunity for economic savings. A full western regional system operator, a much-discussed possibility, would create further benefits. Renegotiation of treaties with Canada will change the flexibility of the Canadian and US hydro systems to respond to new demands imposed by variable wind and solar resources. The rapid expansion of community choice aggregators is accelerating the shift in diminishing the role of California’s investor-owned utilities as integrated power suppliers/purchasers.

Evolving markets for aggregators in providing demand response, storage, and delivering customer supplied energy to the grid is creating many new opportunities for customers to make self-oriented investments and enjoy market benefits from the operation of these resources. Expanded technology for load control, particularly of thermal loads (including water heating, space conditioning, and refrigerated warehouses), is creating system flexibility not previously available. New end uses for electricity—primarily from electrification of transportation, water heating, and space heating loads traditionally served by direct use of fossil fuels—are creating new market opportunities, but also potentially add new opportunities for system flexibility. If vehicles, water heaters, and cold storage

[7] See Michael Liebrich, Bloomberg New Energy Finance, “Trends in Clean Energy and Transportation,” CAISO Stakeholder Symposium, October 18, 2017.

warehouses can be charged when the sun is shining or the wind is blowing, and charging curtailed when other loads demand service—that is, if these end uses are electrified in a beneficial fashion[8]—then the system gains immensely in its ability to absorb new low-cost variable renewable energy sources.

Taken together, these changes present challenges, but at the same time present plentiful opportunities to manage the challenges. For the past 125 years, the electricity industry has focused on controlling resources to match varying loads; now the challenge is increasingly to ensure the power system is able to adjust loads to match resources. Measures to deal with this include demand response, which including storage and controllable load can be used to ensure that system capacity is not exceeded by system demand, and to ensure that available resources are productively utilized. Time-varying pricing is a crucial tool in this load shaping effort.

Any new approach to rate design needs to embrace these changes. Rate design should ensure that customers make efforts to shift load to where it is valuable, and to control load when it is worthwhile to do so. The historical focus on annual, monthly, or even daily load factor becomes irrelevant in a system where the resources themselves do not produce steadily. Time-varying pricing, to reflect production, transmission, and distribution service costs, will be essential. Because the underlying costs of providing electricity vary hourly and seasonally, it is impossible for the customer to see to an appropriate price signal without that signal also varying over time. As smart technologies take hold, the connection between customer usage patterns and underlying costs will become apparent. As this happens, it is inevitable that time-differentiated pricing will become more widespread. Options such as critical peak pricing and demand response are simple, understandable, deployable, and effective. Real time pricing, where prices change as often as hourly, is another time-varying pricing tool that will be a viable option for organizations with sophisticated energy managers and customers with end uses, such as many EV charging systems, that can automatically respond to real-time prices.

Ascending and Emerging Technologies

The smart thermostat is a real product available in the market today, but only a small percentage of customers have installed them. It enables energy efficiency savings and peak demand savings. Nest proved that this technology could deliver significant load relief during the 2017 total solar eclipse. This is an **ascending technology**: available to deploy, and potentially rapidly deployed with economic and programmatic support.

The residential ice-storage air conditioner has been produced in very limited quantities by Ice Energy, for a pilot deployment overseas. It will enable load shifting and peak demand savings. It is not available for purchase in California now. The technology and control systems are being perfected as this is written. This is an **emerging technology**: it could be available for future deployment with economic and programmatic support.

[8] For more on the topic of beneficial electrification, see Keith Dennis, Jim Lazar, and Ken Colburn (July 2016), "Environmentally Beneficial Electrification: The Dawn of 'Emissions Efficiency,'" *Electricity Journal* 29/6, 52–58; and RAP's blog series: http://www.raponline.org/?sfid=5489&_sf_s=Beneficial Electrification&_sft_category=blog&sort_order=date+desc&post_date=01092017+02092017

Example 1: From Load Factor to Load Shape

An example illustrating how optimizing supply and demand resources affects pricing is the needed shift in focus from “load factor” to “load shape.” “Load factor” is the ratio of average demand to peak demand. Historically, utilities and rate design have focused on improving the “load factor” of individual customers, with the expectation that this will improve the load factor of the system and thereby improve the utilization of capital investments in production, transmission, and distribution capacity. This made sense when all resources were dispatchable by injecting more fuel and a high system load factor was a primary economic planning criteria, but in a world of variable renewable energy supply, focusing on load factor without considering load shape is a serious mistake. A low-load-factor customer with irregular usage, but at off-peak times, is a beneficial load to the system because that customer increases system utilization without adding to system peak; an example is a high school football stadium, with usage only in the evening hours and mostly in the autumn. A high-load-factor customer with continuous usage, on the other hand, is always imposing a load at system peak times. Thus, focusing on load factor without considering load shape can lead to rate design decisions that are out of line with cost causation.

Precisely because of situations like this example, analysts have begun to focus on “load shape,” meaning the distribution of the loads across the day, month, and year. Loads that predominantly occur during off-peak periods are more desirable (lower-cost to serve) than loads that are continuous and thus occur at the time of the system peak or distribution system peak. The advent of electric vehicle charging, customer electricity storage, ice and chilled-water storage for air conditioning, and other tools to shift load mean that some controllable but intermittent loads are more desirable—and potentially lower-cost to serve—than stable and continuous loads.

A focus on load shape is particularly important when it comes to the issue of demand charges. A demand charge measured on a customer’s highest 15-minute non-coincident peak (NCP demand charge), will encourage a customer to reduce its own individual peak, regardless of the correlation with the system peak. A demand charge measured on the customer’s highest usage during the expected system peak period (CP demand charge) will encourage a customer to have a high load factor relative to that customer’s peak demand within the system peak window. This means that the customer likely will use similar levels of power throughout the system peak period. The effect of the demand charge price signal is to reduce benefits of diverse loads. More dynamic pricing methods can better match price to system impact than either NCP or CP demand charges. For example, a time-of-use (TOU) volumetric charge will apply equal weighting for capacity cost recovery to each hour within the peak period. This will encourage customers who may need high levels of power at 5 p.m. to decrease that usage at 6 p.m., as it is hourly use, not maximum use, that drives nearly all of their bill.

The only costs that are “caused” by an individual customer’s NCP demand are those near the point of delivery, where the shared distribution circuit ends and the individual customer connection occurs. Typically, the only system components sized based on the customer NCP are the final line transformer[9] and the service wire to the meter. Everything upstream is sized based on the

[9] In the large non-residential sector, dedicated customer line transformers are the norm; for residential and very small commercial customers, shared transformers, with diversity considerations between the multiple customers, are common.

combined usage of many customers, and once upstream of the distribution substation, the correlation of local demand with system demands becomes quite close.

Example 2: Metering and Information Technology Enable Billing that Recognizes Load Diversity Benefits

Historically, system capacity costs have been recovered through non-coincident demand charges that measure each customer's individual highest usage during a month, regardless of whether the usage is coincident with the system peak. This measurement was used as a proxy for that customer's contribution to system capacity costs driven by coincident peak demands. Demand charges were implemented in this way due to the limitations of metering technology.[10] Until electronic interval meters became widespread, mechanical meters only recorded total kWh consumed and the maximum demand that occurred during the billing period. This shortcut was implemented partially because of the inability of traditional electric meters to provide more granular data on customers' usage. The availability of smart meters, and the dynamic data they provide, is now giving utility regulators the ability to focus rate designs on time-varying volumetric (per kWh) charges and dynamic pricing.

Using a non-coincident demand charge fails to bill customers accurately in two ways. First, a customer's maximum demand may occur during off-peak hours rather than during system peak periods. Though this method is roughly accurate for many large commercial customers because their highest usage *usually* coincided roughly with the system peak, even that is not always the case and is still an approximation of those customers' contributions to the incurrence of capacity costs. For smaller and more intermittent users, such as schools, churches, sports stadiums, and emergency facilities, there may be very little correlation between individual customer maximum demand and system demand.

Second, billing customers based on their individual maximum demands may unfairly allocate costs to customers with more variable demands due to the benefits of load diversity and the related ability of multiple customers to share capacity. A simple example this diversity is a school, with usage occurring primarily during the week, and a nearby church, with loads on Sunday that are much higher than other days. Clearly these two customers can share generation, transmission, and network distribution capacity; if each pays the same demand charge as a continuous-use customer (like a 24-hour mini-mart), the school and church are being overcharged, and the continuous-use customer undercharged. Conversely: a religious campus that contains both a church and a school that takes power through a single meter avoids this rate design problem. Table 4 on p. 28 presents a quantitative example that demonstrates this situation.

Rate designs that focus on non-coincident peak demand charges have the effect of focusing customers on optimizing load management to reduce their bills in ways that do not contribute to controlling system peak demand. Instead, customers will work to levelize their own demand relative to their individual peak. NCP charges also have the effect of shifting costs, without justification from cost causation, from continuous-demand customers (who are always drawing

---

[10] Yakubovich et al., 2005, "Electric charges: The social construction of rate systems," *Theory and Society*, 34: 579–612.

power at the time of the peak) to more sporadic customers (whose usage varies, meaning that multiple customers can share the same capacity).

With smart meters and dynamic data regulators can consider more accurate coincident peak and off-peak demand charges, not just the single hour (or 15 minutes) of usage typically used to compute non-coincident demand charges, but these are a second-best approach in light of load diversity.[11] The availability of more precise usage data makes demand charges a largely antiquated approach for all customer classes.

California Policies

California has been leading the nation in adapting utility regulation to the challenges and opportunities posed by DERs. The CPUC DER Action Plan directs utilities to plan for DER growth, preserve a role for third-party players, and utilize DERs for system operations. Senate Bill 350, a 2015 law, codified the state's 50% renewable energy mandate for 2030 (second highest in the nation, with Hawaii at 100% renewables by 2045), established a 2030 GHG reduction target of 40% below 1990 levels, and required a doubling of energy efficiency savings by 2030. The bill also specifically requires an integrated resource plan (IRP) process in which the CPUC is to recognize linkages to the transportation and building sectors and "to identify optimal portfolios of resources," including DERs, to meet California's policy goals. California's policy goals recognize and seek to leverage the ascending technology trends identified earlier in this report.

In November 2016, the CPUC released a DER Action Plan for aligning the Commission's support of DERs with the ongoing work of staff across numerous proceedings.[12] The DER Action Plan recognizes the need to remove unintentional barriers to DER deployment behind the meter and in front of the meter, and clarify the DER value proposition. In fact, the action plan specifically requires the review of non-residential demand charges and the consideration of changes to these customers' rate designs, specifically for the alignment of pricing with the DER vision. The Action Plan further directs that, "by 2017, consider changes to non-residential rate design including modification of demand charges." [13]

Rate design is debated and determined in Commission processes separate from the detailed planning processes in an IRP. But rate designs directly affect IRP assumptions and modeling results. Rate design impacts load growth and consumer behavior. These, in turn, are important assumptions that go into IRP models and influence utilities forecasts of their future needs. Pricing also affects customer decisions about whether to adopt DERs, which similarly influences the outcomes of IRP modeling. With good rate design, DER portfolios and customer load management activities can help to meet system needs and achieve benefits for customers. Rate design that does

[11] Mathematically, a demand charge spread over demand in all on-peak hours is equal to a time-varying energy charge recovering the same costs over the same hours, but such a granular demand charge has no advantage relative to dynamic pricing so these highly granular demand charges are not useful.

[12] The CPUC endorsed the DER Action Plan on November 10, 2016, and subsequently produced "California's Distributed Energy Resources Action Plan: Aligning Vision and Action," May 3, 2017.

[13] Ibid, see Action Item 1.9 on p. 3.

not properly reflect changing costs and advancing customer-side DER technologies will lead to an IRP that does not acknowledge trends already underway. That could lead to unnecessary cost burdens on society, and could fall short of optimally advancing state policy goals around DERs.

California Policy Example 1: Rate Design Affects Renewable Integration

The DER Action Plan states, "Senate Bill 350 requires the Commission to implement an integrated resource plan (IRP) process to identify optimal portfolios of resources to achieve the state's GHG goals and meet the challenge of renewable integration, and DERs will play an important role."[14]

In addition to California policy driving a need to reexamine rate design practices, there is a need to provide customers the opportunity to contribute to the cost-effective operation of the power system. Specifically, rate design affects how much and when customers choose to consume power, and influences whether they opt to install on-site generation or storage. These choices, in turn, affect whether customers can contribute to the least-cost integration of renewable energy. Improper rate design can lead to overconsumption at times when the system is already stressed or under consumption when there is an abundance of solar. It can also lead to a lack of incentive for commercial and industrial customers to invest in storage, which could be a cost-effective means for integrating California's abundant rooftop solar resource. This means that utilities must bear the entire cost of integrating variable generation, which could come at a higher cost to the system and all customers. For example, non-coincident demand charges can actually exacerbate integration challenges and cause costs for other customers by encouraging consumption at times that increase system stress.

California Policy Example 2: Rate Design Affects Beneficial EV Charging

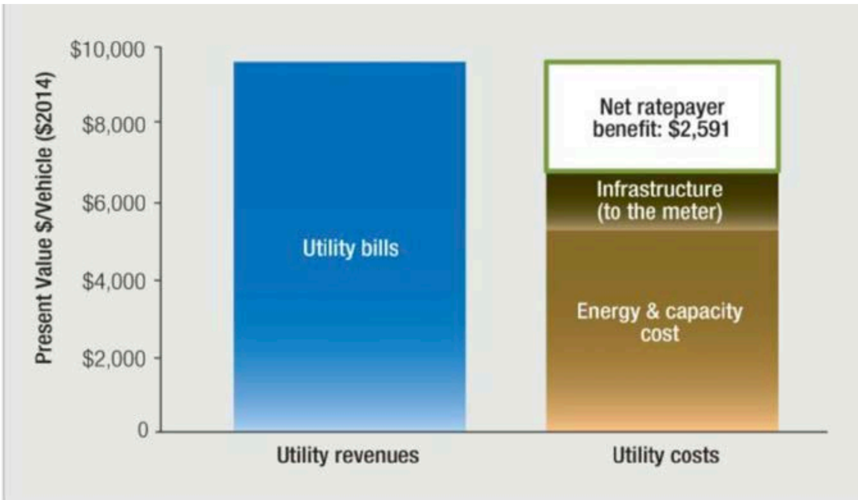
A cost-effective electricity sector can lead to least-cost power for all customers. Rate design has ramifications for this outcome as well. Commercial customers can be incented to participate in the cost-effective supply and storage of power, as well as to consume power in a way that is beneficial to the grid, which will help lower costs for all customers. For example, electric vehicles with smart charging capabilities will help reduce costs for all customers.

An analysis by Energy and Environmental Economics (E3) of EV adoption scenarios in California highlighted the significant utility system benefits from DERs (see Figure 1 on the following page). Utilities' cost to serve EV charging load was found to be less than the revenue they would be bringing in from those customers, meaning a net benefit to the utility system and to all ratepayers (not just EV drivers). Off-peak charging of EVs increases the utilization of the transmission and distribution system, lowering the average cost to serve all customers. Controlled charging enables greater integration of variable renewable energy resources.

---

[14] Ibid, p. 1.

Figure 1. Light-Duty Plug-In Electric Vehicle Load at California Utilities, Under TOU Rates



It is important to note that system and EV user benefits greatly increase with (though are not entirely dependent on) the ability to move the majority of EV charging to charge off-peak times. This point argues for appropriate rate design that includes time-of-use elements—through peak and off-peak energy—which will more accurately communicate system costs to consumers and reward those who respond by shifting their demand to low-cost hours. Dynamic pricing elements will encourage controlled charging, as customers use programmable and interactive charge controllers to avoid high-cost hours and take advantage of low-cost hours.

III. Principles for Smart Non-Residential Rate Design

In *Smart Rate Design*, RAP published an extensive discussion of residential and small commercial rate design, but identified three principles that should, in our opinion, apply to all customer classes:

1. **Principle 1: A customer should be able to connect to the grid for no more than the cost of connecting to the grid.** That is, the fixed charges should not exceed the recovery of customer-specific costs such as the final transformer (for secondary voltage customers), plus the service drop, metering, billing, and basic customer service expenses. Distribution circuit costs should not be included in the cost to connect.
2. **Principle 2: Customers should pay for grid services and power supply in proportion to how much they use and when they use it.** That is, the costs of power supply, transmission, and distribution costs should generally be recovered through volumetric charges linked to season, time of day, and other cost-reflective metrics.
3. **Principle 3: Customers who provide services to the grid should be fairly compensated for the value of what they supply.** While customers purchase *from* the grid on a "cost of service" basis, those who are suppliers *to* the grid should be treated like other suppliers, with compensation based on the value of what is supplied. This may be significantly different than the cost of service retail price, reflecting the characteristics of the products and services supplied.

Principle 1–connecting to the grid for no more than the cost of the grid—is relevant for non-residential rate design. The cost of dedicated facilities for the customer and the cost of facilities directly sized to the non-coincident peak (NCP) demand is normally much greater than for residential and small commercial customers so the demand charge takes on more significance.

- Non-residential (NR) Principle 1: The service drop, metering and billing costs should be recovered in a customer fixed charge but the cost of the proximate transformer most directly affected by the non-coincident usage of the customer along with any dedicated facilities installed specifically to accommodate the customer should be recovered in a NCP demand charge.

Principle 2—recovering costs based on how much and when energy is used—is the most important of these in the California non-residential rate design context, because the costs of the shared generation, transmission, and network distribution constitute the clear majority of utility costs. NR Principle 2 includes several parts.

- NR Principle 2.1: De-emphasize NCP demand charges except as noted in NR Principle 1. All shared generation and transmission capacity costs should be reflected in system-wide time-varying rates so that diversity benefits are equitably rewarded.
    - Rationale: Non-residential customers often place diverse demands upon the electric system that can use resources in a complementary manner.
- NR Principle 2.2: Shift shared distribution network revenue requirements into regional or

nodal time-varying rates. This recognizes that some costs are required to provide service at all hours, and that higher costs are incurred to size the system for peak demands.[15]

- Rationale: The distribution network is a shared resource where incremental investment is driven by system stress conditions. System stress conditions should align with time-varying rates that reflect the degree of stress being placed upon the system, and time-varying rates rather than coincident peak demand charges should be relied upon to communicate system stress.

- NR Principle 2.3: Consider short-run marginal cost pricing signals and long-run marginal cost pricing signals together in establishing time-varying rates for system resources.
    - Rationale: The resources and load shape needed in five to ten years should be a focus, including extensive workplace charging of EVs, higher levels of PV and wind, and eventually rising natural gas prices. Customers are making investments in durable technologies that will become part of the power system for 20 years or more, so ratemaking should be forward-looking. Short-run marginal costs, like locational marginal prices (LMP), convey important system stress price signals that should be recognized. However, long-run marginal cost price signals are also relevant in establishing time-varying rates, because new resources have durable value that displaces the need for future infrastructure and system resources and thus is responsible for avoiding certain future costs.

- NR Principle 2.4: Time-varying rates should provide pricing signals that are helpful in aligning controllable load, customer generation, and storage dispatch with electric system needs.
    - Rationale: Customers will dispatch their resources to manage their bills, but aligning price with system needs will prompt customers to dispatch their systems in a manner that supports system need. Rate design tools like critical peak pricing, time-of-use pricing, and dynamic pricing can support aligning customer resource dispatch with system needs.

- NR Principle 2.5: Non-residential rate design options should exist that provide all customers with an easy-to-understand default tariff that does not require sophisticated energy management, along with more complex optional tariffs that present more refined price signals but require active management by the customer or the customer's aggregator.
    - Rationale: The non-residential class of customers is very broad, and customers on the medium to small side of the spectrum may not have the desire or capacity to manage a complex tariff. Other customers do have that capability or will be willing to pay an aggregator to manager their consumption and resources to handle more complex pricing structures like granular dynamic pricing.

[15] One California municipal utility, for example, has TOU rates for commercial customers that include weekends as off-peak, but for residential customers, summer afternoons remain on-peak, due to distribution system capacity constraints on residential circuits. The same concept could apply in different regions or nodes of a distribution system serving non-residential customers, where capacity constraints are reached at different times of the day or year.

- NR Principle 2.6: Optimal non-residential rate design will evolve as technology and system operations matures so opportunities to revisit rate design should occur regularly.
  - Rationale: Over time, locational pricing will become more granular, perhaps with nodal pricing down to the level of the substation or the feeder. The opportunity to meet wholesale electric system needs with aggregated DERs is expanding, but is not yet matured. The opportunity to meet distribution system operator needs with direct or aggregated customer loads and DERs will mature over time. Each of these changes will affect opportunities and will affect optimal pricing signals so the rate designs implemented today should take advantage of the system as it exists but at the same time recognize it is a system in transition.

Principle 3, on DER compensation, is worth considering whether non-residential tariffs require more specific treatment to implement Principle 3, but it is beyond the scope of this paper.

One can apply these costing principles to the various components of the electric utility system in a structured manner, with rate design elements tracking the function and nature of each component of the system. Shared assets should generally be recovered on a volumetric, TOU, or RTP basis, so that shared capacity costs are paid by all consumers using that capacity on an equitable basis. The costs associated with customer-specific investments should accrue to the specific customer. Table 1 shows an example of how the different components of the system may logically track into rates.

Table 1: Tracking Electric System Elements into Rate Design

Category	Characteristics	Notes
Generation Capacity	Shared systemwide	TVR recovery appropriate across all hours when resources provide service
Generation Operating	Shared systemwide	TVR recovery appropriate
Bulk Transmission	Shared systemwide	TVR recovery appropriate
Network Transmission	Shared regional/nodes	Nodal TVR recovery appropriate
Substations	Shared local/nodes	Nodal TVR recovery appropriate
Distribution Circuits	Shared local/nodes	Nodal TVR recovery appropriate
Final Line Transformer	Dedicated or shared multi-customer	Customer-specific $/kW
Secondary Service Lines	Dedicated or shared multi-customer	Customer-specific $/kW
Meters	Customer-specific	Portion of meter costs attributable to DR, EE, loss reduction
Billing/Collection	Customer-specific	Billing frequency is volume-related

The application of the above characteristics to rate design raises an important issue: the question of whether rates should vary by sub-region, distribution circuit, or “node” within a system. We will not address these issues in depth in this paper, but a brief note on each is helpful.

Important Caveats and Future Considerations

We are attempting to stay focused on the immediate rate design issues that need to be confronted in California today. However, there are several ancillary issues that should be considered by readers today but may more appropriately be dealt with at a later time or in a different CPUC proceeding.

First, there is no doubt that local peak demand on the distribution system is not perfectly correlated with system peak demand and using local demand and local DER resources to address local distribution system stress will be a consideration to be dealt with soon. Dealing with local system stress in an equitable way is a complex issue. Historically this has not been done, with a single "on-peak" period applied system-wide. This is generally out of a sense of "perceptions of equity and fairness," one of Bonbright's key ratemaking principles. Adding differentiation by sub-region may be perceived as discriminatory (even though it is sometimes justifiable discrimination), but there is no question that distribution system nodal differences in power cost exist, and no question that peak periods sometimes vary from circuit to circuit. The CPUC has begun dealing with this for larger distribution upgrades by requiring utilities to hold non-wires alternatives competitive RFP opportunities for certain upgrades. Addressing this issue and ensuring consistency with non-residential rate design is an important issue but is beyond the scope of this paper.

Second, ensuring revenue adequacy for California's investor owned utilities so that safe, reliable and affordable grid service persists is an important goal of rate design that we did not explicitly mention in a principle. First of all, a myth exists that demand charges are likely to support revenue adequacy better than energy based charges; the text box to the right seeks to dispel this myth. Recovering costs when prices are set at short-run marginal costs has been a persistent challenge; Borenstein (2016) explains the problem well.[16] Further exacerbating this problem is the prevailing trend of diminishing variable costs (see text box below). RAP believes that a guiding principle for ensuring revenue adequacy for the utility while establishing a level playing field for new DERs is to align these goals with long-run marginal cost price signals, and this is a point of disagreement with Borenstein.[17]

Revenue Adequacy: Demand Charges vs. Energy Charges

Some rate analysts have opined that demand charges produce a more stable revenue stream than volumetric energy charges, and raise this as an objection to moving towards time-varying pricing.

RAP believes this concern is unfounded, for two reasons. First, customer NCP and CP demand tends to be highly weather-sensitive, while about 70% of energy consumption is for uses other than space conditioning, and thus not weather-affected. Second, California has a decoupling mechanism that ensures that revenue stability is not an issue for its utilities from year to year.

Looking ahead, however, we believe that time-varying volumetric pricing will be easier for consumers to understand, and thus to estimate savings that can be achieved through electrification of existing fossil energy end uses. This may help open new market opportunities for California electric utilities, and contribute to the achievement of California's energy policy goals.

[16] Severin Borenstein, "The Economics of Fixed Cost Recovery by Utilities," *Electricity Journal,* 29 (2016), 5-12.

[17] One important role of regulation is to create a level playing field between the incumbent utility's pricing and the market-derived pricing for

It is not necessary to consider this debate further here, as that is likely to detract from the central purpose of this paper, which is to provide concrete guidance on how rate design should be structured in the near future. RAP has written on this issue before, and we recommend the Weston (2000) appendix on the economics of regulation as a good resource on this view.[18]

The third caveat we wish to raise here is implications of a more transactional grid for pricing and cost recovery. Ascending technologies are setting the stage for a more transactional grid where there will be multilateral sales and purchases happening on the grid involving many transactional parties. Ensuring that all beneficiaries of the transactional grid contribute to the maintenance and improvement of a transactional grid will be paramount. Traditional models of cost causation are not well-adapted to cost recovery of transactive grid investment. This topic is beyond the scope of this paper but highly relevant to the longer-term future of rate design in California. We commend readers interested in these issues to Cazalet et al (2016).[19]

Diminishing Variable Costs Force Us to Reconsider Rate Design

California is at the forefront of a global trend in which renewable resources and technology for energy storage and load control is replacing the use of fossil fuel generation to meet varying customer requirements through the day and the year.

Many decades ago, cost allocation was relatively simple, and directly fed rate design: Nearly all generation was local (oil and gas), long distance transmission lines had not been deployed, and fuel costs were more than half the total cost of service for California electric utilities. High load factors were considered desirable.

The first and second oil embargoes of 1973-74 and 1978-79 began a transition, and the more recent directives of the California legislature to reduce carbon emissions and rely on renewable energy sources, along with decline in cost of renewables has redoubled the pace of change. As non-fuel resources are substituted for older resources, within a few years it is likely that, variable costs will be no more than 20% of the cost of service.

The challenge for cost allocation and rate design is to ensure that costs incurred to serve specific uses—baseload versus peak demand, for example—are assigned to the right customer classes in the cost allocation process, and targeted at the right periods of usage in the rate design process. Traditional embedded cost and marginal cost methods do not do this.

We can no longer rely on the notion of “fixed” or “variable” costs for guidance in designing rates. Solar and wind projects are “fixed” costs incurred to reduce emissions from burning fuel, long-considered a variable cost. Capital-intensive long-distance transmission may be built to integrate capital-intensive variable renewable resources, and while these are both “fixed” costs, they may have little or no role in meeting peak demand.

The challenge today is to ensure that costs are assigned to the purpose for which they are incurred, and those costs are recovered over the appropriate time periods when those resources provide service. The focus in this paper on shifting from demand charges to time-varying usage charges recognizes this evolution.

competitive alternatives such as on-site generation, storage, and efficiency. In RAP’s view, rates based on short-run marginal cost do not do this; only rates that reflect the full long-run incremental costs of electricity supply and distribution perform this function.

[18] Frederick Weston, “Charging for Distribution Utility Services: Issues in Rate Design,” Regulatory Assistance Project, 2000.

[19] Cazalet, De Martini, Price, Woychik, and Caldwell, “Transactive Energy Models,” prepared by the NIST Transactive Challenge: Business and Regulatory Models Working Group, September 2016.

IV. Non-Residential Rate Design in California Today

This section looks at current California IOU commercial and industrial rate design in practice, with a focus on the question of whether rates are correctly aligned with cost causation and how the alignment with cost causation can be improved. We look at rate design for generation, distribution, and transmission service components and consider whether and how well the rate designs line up with the principles outlined above. In particular, we consider whether rates for each component are time-dependent in California.

Generation

The determination of generation costs reflects energy costs (which fluctuate throughout the day with CAISO wholesale electricity market prices) and additional capacity costs incurred in resource procurement. The cost of compliance with resource adequacy and flexible capacity requirements is included in the cost of bilateral contracts and embedded, as needed, in long-term power purchase agreements.

Table 2 on p. 24 presents a SCE, PG&E, and SDG&E rate design for reference. Rate design for generation for medium and large commercial and industrial customers of California IOUs is typically as follows:

- Medium/large commercial customers of IOUs face both per/kWh and per/kW rate components related to generation.
- The per/kWh components are TOU.
- The per/kW components reflect generation capacity costs and are almost always recovered using coincident demand charges, with the demand charges applying only in the summer peak period for some utilities.

Transmission

Transmission costs are determined in a process where each transmission owner seeks approval from FERC (Federal Energy Regulatory Commission) for a "transmission revenue requirement" (TRR).[20] This revenue requirement is then allocated to end users. Here we focus on the allocation of transmission costs to the IOU distribution companies and their customers.

As with the distribution case, some transmission assets (costs) are "shared" and some can be attributed to a particular user or group of users. Currently there is a process for allocating "regional" (high-voltage) and "local" (low-voltage) transmission costs, with the "local" costs being allocated to end users in each given locality. All high-voltage assets are lumped together into an

[20] For detailed discussion of transmission revenue requirements and allocation to distribution companies, see CAISO, "How Transmission Cost Recovery Through the Transmission Access Charge Works Today," 2017.

aggregate revenue requirement (known as R-TRR) and a "postage stamp" rate (known as the regional transmission access charge or R-TAC) is determined by dividing the R-TRR by the aggregate forecasted gross loads of all transmission owners. A local L-TAC is also calculated on a similar basis but unlike the R-TAC, the L-TAC is specific to each participating transmission owner.

The R-TAC and L-TAC are assigned to the IOU distribution companies. In particular, the R-TAC is allocated across different IOU distribution companies based on the gross load of each.[21] These distribution companies, in turn, collect from LSEs, and ultimately from end users. (The LSE may be the distribution company itself, retail service providers, or community choice aggregators.) Once the IOUs' TRRs are determined, they are allocated to customer classes (residential, small commercial, medium and large commercial, etc.) based on 12-coincident peak (12-CP) methodology that accounts for the customer class's contribution to system peak demand on the transmission system.

Finally, end-user retail transmission rates are calculated by dividing the class allocated TRR by the class billing determinant (either kWh sales, for smaller customers, or the sum of class non-coincident kW, for medium and large commercial customers). This approach results in flat (non-time-dependent) volumetric retail transmission rates for residential and small commercial customers, and (non-time-dependent) non-coincident transmission demand charges for SCE's medium and large retail commercial customers.

Therefore, while the allocation of TRR to a class of customers' accounts for the customer class's contribution to system peak demand on the transmission system, the individual customer's transmission rates do not reflect the individual customer's contribution to coincident peak demand.

End users of an IOU distribution company face the same retail transmission rates and rate design, regardless of LSE type. This rate design is approved by FERC.[22]

In summary, medium/large customers of California IOUs face transmission charges that are structured largely as non-coincident peak charges.

The CPUC has recently encouraged some IOUs to file with FERC to reduce reliance on maximum non-coincident demand charges to recover transmission costs—and move toward time-dependent rates.[23] Among the IOUs, SDG&E has taken early steps in this regard: In 2008, FERC approved SDG&E's request to have a portion of transmission revenue collection from medium/large commercial customers moved from non-coincident peak charges toward seasonally differentiated coincident peak charges. Recently, the CPUC required SDG&E to perform studies of its transmission rate design to determine if further changes are warranted.[24]

---

[21] CAISO, 2017.

[22] The involvement in retail rate design by FERC is an unusual situation that stems from California's particular restructuring situation. In other states, FERC is not typically involved with retail rate design. See: "What FERC Does Not Do" at https://www.ferc.gov/about/ferc-does.asp.

[23] See CPUC, Decision 14-12-080, p. 21, http://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M143/K631/143631744.PDF; or CPUC, Decision 17-08-030, p. 92, ordering paragraph 34.

[24] In addition to these costs, there is an adjustment made called the EPMC, which is an important aspect of implementing the NERA

Distribution

For distribution network costs, two types of marginal costs are developed: marginal customer access costs (MCACs) and marginal distribution demand costs (MDDCs).

- MDDCs reflect common (shared) demand-related costs (e.g., utility wires, line transformers[25], substations). Marginal distribution demand cost ($/kW/year) is estimated based on NERA methodology.[26]
- MCACs reflect customer-specific costs, such as final line transformer investments, customer hookup costs, and customer service costs (meter reading, billing, etc.).

As with the generation component, an EPMC is calculated. Crucially, the EPMC for a given customer class is calculated as a function of customer class's annual non-coincident demand (kW) along with customer-class-specific MCAC.

Rate design for the distribution component includes the following:

- Monthly per-meter fee that reflects a portion of the MCAC;
- Per kW (non-coincident) demand charge for medium and large customers plays a larger role at SCE but PG&E and SDG&E and the NCP charge for all three is differentiated by:
  - season (summer vs. winter) and
  - voltage level;
- Per kW (coincident) demand charge for medium and large customers at PG&E and SDG&E is differentiated based on:
  - season (summer vs. winter) and
  - voltage level.

In short, medium and large customers of IOUs face charges substantially based on non-coincident peak use for shared distribution costs, while these have little cost-causative impact on distribution system investment. While the California IOUs also collect part of their distribution revenues via CP demand charges, there is good reason to believe that NCP demand charges are overused.[27]

California's existing rate design has evolved over decades with recent changes include a shift in TOU peak hours, the increasing use of CP demand charges rather than NCP demand charges, the shifting of demand charges to volumetric charges under the Option R, and, most significantly, the shift to mandatory default critical peak pricing and mandatory TOU rates by 2019. Despite these significant changes, non-residential rate design has retained features from the earlier era.

The rest of this section identifies a number of situations where rate design may need to change to match the current technological and policy context. Some are situations faced in many places, while a couple of others are specific to the California situation.

---

methodology. Detailed discussion is beyond the scope of the report, but suffice to say that the EPMC is a multiplier that is used to scale up marginal costs. It is worth noting that for distribution costs, the multiplier is driven by the NCP demand charge.

[25] Other than the final line transformer, which is closest to the customer and considered customer-related.

[26] See Appendix B for a discussion of the NERA Marginal Cost Methodology.

[27] The recent SDG&E decision, D.17-08-030, recognized this and shifted some of SDG&E's distribution NCP demand charges to CP demand charges (OP 17).

Table 2. Comparison of California IOU Rates for 500 kW Secondary Voltage Customer[28]

	SCE (Sched TOU-8)	PG&E (Sched E-19)	SDGE (Sched Al-TOU
Generation			
per kWh component			
Summer On-Peak	$0.07	$0.15	$0.12
Summer Mid-Peak	$0.05	$0.11	$0.11
Summer Off-Peak	$0.03	$0.08	$0.08
Winter On-Peak	-	-	$0.11
Winter Mid-Peak	$0.05	$0.11	$0.09
Winter Off-Peak	$0.04	$0.09	$0.07
per kW component			
Summer On-Peak	$18.92	$12.63	-
Summer Mid-Peak	$3.63	$3.12	-
Summer Maximum Demand (NCP)	-	-	$10.88
Transmission			
per kW			
Base, NCP	$4.88	$7.19	$12.05
Summer On-Peak	-	-	$2.13
Winter On-Peak	-	-	$0.66
Distribution			
per kWh component (including UDC costs, such as public purpose programs)	$0.024	$0.021	$0.004
of which, distribution only	$0.002	$0.000	$0.001
per kW component			
Base, NCP	$13.67	$10.37	$12.41
Summer On-Peak	$0.00	$6.01	$8.12
Summer Mid-Peak	$0.00	$2.06	-
Winter Mid-Peak	$0.00	$0.12	-
Winter On-Peak	-		$6.91
Customer charge per meter per month	$634.89	$599.59	$465.74

[28] SCE rates: https://www.sce.com/wps/portal/home/regulatory/tariff-books/rates-pricing-choices/business-rates/!ut/p/b1/tVJNU8IwEP01PYaktPTDWwccbB1UBMa2FyYNSRttk5IGUX-9geGgDogczCnZffuy7-3CHKYwF_iVl1hzKXC9e-feMolHkT12-_HYT0Yoehj5o8WjZ7u3tgFkBoBOnAidq3-COcyJ0K2uYNYRuiRSaCr0kgoLHe4WUrTc1FhL9W4hjRVnDBRSvnQmgzXtQKs44aIEpJKcUBMuNh0XtOvAPr_7oyV8BbP-gCCfhRjYrHCAawcrEBQ-Box5RUBYyBwawuQPovtqMpyUhhbrCnDBJEy_NgbTo43B9Edjhok_r9d5ZEzYiX3TMP03F_ZeG2XhGF3fJPcoHs-nDoqdKbqbRZGDkHcA_DJO401Zy2K_GlkkCicwJijKqKKqt1EmXGnddlcWstB2u-2VUpY17RHZWOhYSSU7I_k7EmZmqfyTAxi6cHbhRM8QehcTts2iCZxBXQY6jEFevDsfc9Y00Sdd9rfQ/dl4/d5/L2dBISEvZ0FBIS9nQSEh/

PGE rates: https://www.pge.com/tariffs/index.page

SDGE rates: http://regarchive.sdge.com/tm2/ssi/inc_elec_rates_comm.html and http://www2.sdge.com/tariff/com-elec/eecc.pdf

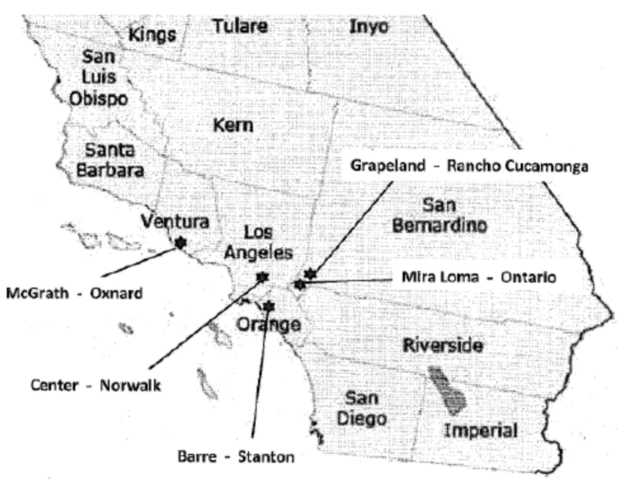
Coincident-Peak-Related Costs Should Be Recovered With Time-Varying and Critical Peak Rates

The vast majority of the power system is needed to provide service at all hours of the year, so while there are some expenses that are specifically peak-demand-related, many expenses are incurred to ensure reliable service over a broader number of hours of the year. Baseload and intermediate generation, reserves, variable renewable generation, transmission and distribution systems, billing and collection systems, and overhead are all required even if power systems have completely uniform loads. But there are some very significant costs that are required to meet peak demands, particularly during extreme peak events that occur for very few hours of the year. These costs include:

- Peaking generation and associated fuel supply
- Additional distribution system capacity needed only for peak hours
- Demand response capability to reduce loads

Many costs that are often classified as demand-related in utility cost studies, however, are typically not associated with meeting extreme peak demand. For example, peaking generators are typically built close to load centers specifically to avoid the need for transmission upgrades. Today's pollution control technology has allowed a limited number of peakers to be built in dense urban areas; units like the Lake (Burbank) and Canyon (Anaheim), which are located in the heart of these municipal service territories, and SCE peaker locations are shown on the map below. [29] Similarly, demand response measures, such as smart thermostats, industrial load rescheduling, or deployment of local storage (thermal or electrical), avoids not only generation, but may also avoid the distribution capacity needed to serve extreme events if deployed geographically where distribution loads are coincident to system loads.

Figure 3. SCE Primary Peaker Locations

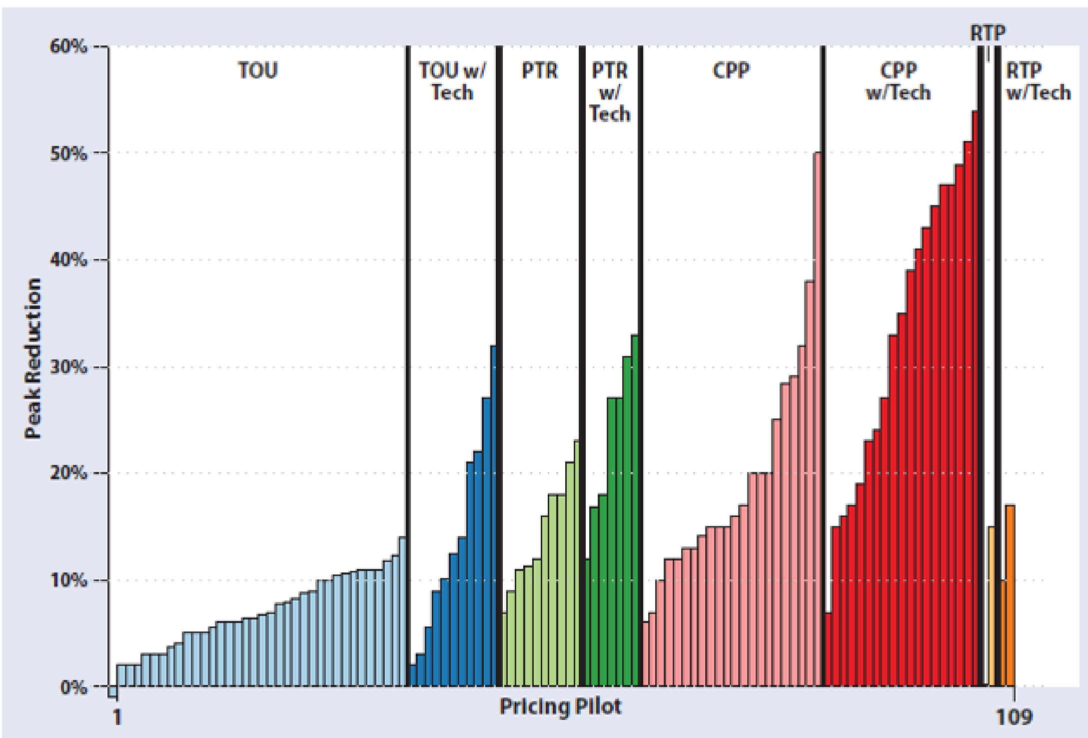


[29] From SCE-02 Vol. 09, 2015 General Rate Case.

The bulk transmission system is normally sized only to bring remote generation, sited for economic reasons, into the load centers, and to facilitate economy energy transfers with out-of-region utilities. Today, transmission is increasingly used to move power from remote wind and solar production locations, so the extent to which transmission expense is caused by peak demand is subject to debate. By contrast, the company's baseload, wind, solar, and geothermal resources are spread across California and beyond California's borders, reaching into five other states.

Extreme system coincident peaks are not predictable far in advance. A combination of extreme weather, reduced output of one or more generating units, and possible transmission and distribution system failures or maintenance activities comes together to create sporadic periods when the system is under great stress. A rate design that simply defines "4-8 p.m., Monday-Friday, June-September" as the "on-peak" period (about 350 hours, in this definition) will likely capture the period when the system is under stress—but will *also* capture hundreds of hours per year that are less stressed, and when rates should not really discourage consumption as aggressively. A coincident peak demand charge, such as that used by SCE, has a similar effect. Conversely, a critical peak pricing alternative, invoked during periodic periods of system stress (that will nearly always, but not universally, fit within the defined time window used for CP demand charges), will more accurately convey to consumers the actual times of system stress when active load management is particularly valuable.

Figure 4: Average Peak Reduction Results from Critical Peak Pricing Pilots[30]



[30] Ahmad Faruqui, Ryan Hledik, & Jennifer Palmer, "Time-Varying and Dynamic Rate Design," Regulatory Assistance Project, 2013.

California has experimented with critical peak pricing, as have other states. The load reductions achieved with CPP have exceeded 40% in some pilots. Figure 4 on the previous page shows the results of more than 100 pricing pilots; as is evident, the *combination* of **critical peak pricing** and **load control technology** deployment to help consumers adapt has proven most effective.

Because the timing of system peaks is difficult to predict, and based on the effectiveness of critical peak pricing and other dynamic pricing mechanisms like real-time pricing, RAP recommends that demand charges generally be eliminated or de-emphasized as tools for signaling the need to constrain usage at times of system stress. The costs currently recovered in most demand charges should be reflected in time-varying energy rates.

Shared Capacity Costs Should Be Recovered With Time-Varying Rates

The recovery of capacity costs (the investment in generation, transmission, and distribution system capacity) is the most significant departure from historical practice that we discuss in this paper, and the genesis of that proposal is important to understand.

Ultimately it is the result reached, not the method employed, that determines the effectiveness of a rate design in achieving regulatory and policy goals. In their seminal text *Public Utility Economics*, Garfield and Lovejoy identified a set of criteria to consider in the recovery of system capacity-related costs.

Table 3. Garfield and Lovejoy Criteria and Alternative Rate Forms

Garfield and Lovejoy Criteria	CP Demand Charge	NCP Demand Charge	TOU Energy Charge
All customers should contribute to the recovery of capacity costs.	No	Yes	Yes
The longer the period of time that customers pre-empt the use of capacity, the more they should pay for the use of that capacity.	No	No	Yes
Any service making exclusive use of capacity should be assigned 100% of the relevant cost.	Yes	No	Yes
The allocation of capacity costs should change gradually with changes in the pattern of usage.	No	No	Yes
Allocation of costs to one class should not be affected by how remaining costs are allocated to other classes.	No	No	Yes
More demand costs should be allocated to usage on-peak than off-peak.	Yes	No	Yes
Interruptible service should be allocated less capacity costs, but still contribute something.	Yes	No	Yes

Table 3[31] lists the criteria identified by the authors, and roughly compares the use of coincident-

[31] Source: Jim Lazar, "Use Great Caution in the Design of Residential Demand Charges," *Natural Gas and Electricity Journal*, February 2016.

peak demand charges, non-coincident-peak demand charges, and TOU and/or dynamic energy charges with respect to whether they serve the above criteria. While the black/white nature of this table may oversimplify, these criteria collectively recognize the importance of load diversity. Most demand charges, by applying a charge for the entire month based on use in 15 minutes to one hour of load, do not recognize diversity. The only rate forms that satisfies all the criteria are the TOU and dynamic energy rates to recover shared system costs.

This clearly shows that the TOU rate design is a more equitable and efficient way to recover capacity costs than either CP or NCP demand charges. The TOU rate recognizes that multiple customers can share the same capacity if their loads have diversity, and do require customers that utilize capacity continuously during the on-peak period to pay for the full cost of the capacity they require. Demand charges by their nature (dividing a pool of costs by a sum of demand billing determinants) inevitably shift costs from higher-load-factor customers to lower-load-factor customers, without justification by cost causation.

Simply put, if a designated "capacity-related cost" is spread over all of the hours of the period for which these costs are incurred (as a time-varying energy rate does), then any user at any window will contribute equally to recovery of the capacity costs. Five customers with intermittent usage, whose combined demands and usage are equivalent to the demand and total usage of a single larger customer, will collectively pay exactly the same amount as the single customer who utilizes the same power and energy at separate intervals of time. The shared capacity customers will likely pay much more than the continuous-demand customer.

We will present a couple of examples of how different customers with complimentary demand may be overbilled if demand charges are the basis of capacity cost recovery.

Table 4: Illustrative Example of Three Commercial Customers On A Shared Circuit

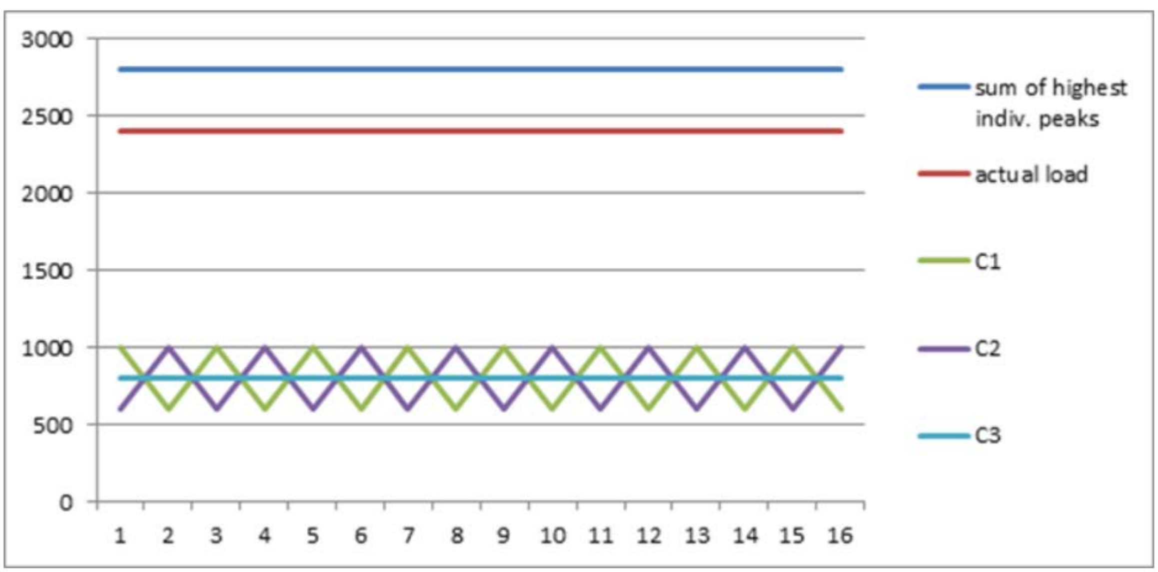
Hours	System Peak	Church	School	Mini-Mart	Total
Weekday 9-4	Mid-Peak	5	45	50	100
Weekday 4-8	On-Peak	5	15	50	70
Nights	Off-Peak	5	5	50	60
Weekend	Off-Peak	45	5	50	100
NCP		45	45	50	140
%		32%	32%	36%	
CP		5	15	50	70
%		7%	21%	71%	

One can illustrate this by looking at Table 4, which outlines a very simple node of a distribution system with three closely located customers: a school, a church, and a 24-hour mini-mart. The school has demand primarily during weekdays, with some after-school loads, and much lower loads during non-school hours. The church has demand primarily on weekends, with much lower loads during all other hours. And the mini-mart has continuous loads at all hours. Table 4 shows the loads in three broad periods, and computes the nodal demand that drives local distribution system capacity needs, as well as the contribution of each customer and the group to the system peak demand. The point is that the school and church are very complementary low-load-factor loads, that can share generation, transmission, and distribution capacity, while the mini-mart requires capacity at all times and cannot share.

Several things are evident from this simple example:

- The school and church have very complementary loads, and can easily share 50 kW of capacity.
- The mini-mart has continuous demand, and requires that capacity at all hours.
- The group of three customers has their individual peak of 100 kW at hours other than the system peak; while they require 100 kw of distribution capacity, they require only 70 kW of generation capacity at the time of the system peak.
- If billed on an NCP basis, the school and church will pay for 64% of of the billed demand, even though they never use more than 50% of the combined capacity requirement. The mini-mart will pay for only 36% of the billed demand, even though it uses 50% of the nodal demand and 71% of the group contribution to system coincident peak demand.
- The group, as a whole, does not peak at the time of the system peak.

Figure 5. Illustrative Example of Three Commercial Customer Loads



Another way of looking at this issue graphically is to compare three hypothetical utility system customers (see Figure 5 above). Customers 1 and 2 each have varying loads, with individual peak demands of 1,000 kW, but at exactly complementary times. Their combined usage never exceeds 1,600 kW. Customer 3 has a 100% load factor, with constant usage at all times of 800 kW, and exactly the same annual energy use as Customer 1 and Customer 2. If demand charges are used to collect system capacity costs, each customer would be billed based on their highest peak demand. Customers 1 and 2 would each be billed for 1,000 kW, (even though their combined load is always 1,600 kW, or 800 kW each) and Customer 3 for 800 kW, a total billing determinant of 2,800 kW. However, since only 2,400 kW of system capacity is required and used, the demand charges would be set at about 86% of the system cost per kW (2400 / (1,000 + 800 + 800)) to produce the system revenue requirement. The result would be that Customer 3, with the 100% load factor, would pay only 86% of their cost of service, while Customer 1 and Customer 2 would each pay 107% of the cost of providing their service, subsidizing the service for Customer 3.

This comes back to the basic premise of this paper that we discuss in Section II: In today's energy environment, load **shape** matters much more than load **factor**. If customer usage is at low-cost times, lower prices should apply. A customer that consumes power continuously is always going to be taking power during peak periods. Customers with varying usage may take power primarily during high-cost or during low-cost periods. Customers with diverse usage patterns can share generation, transmission, and network distribution capacity. Only the site-specific distribution system components are sized to (and thus cost-correlated to) the customer's individual maximum demands.

Street lights are approximately a 50% load factor load, but concentrated in off-peak hours. This is inexpensive to serve. By contrast, office buildings are also approximately a 50% load factor load, concentrated in high-cost hours. While the load *factor* is similar, the load *shape* is not.

Diversity is highest at the bulk power (generation and transmission) level, where hundreds of thousands (or even millions) of customers are served by the same resources. Any customer anywhere on the system can complement the usage of any other customer. At the distribution substation level, there are still hundreds or thousands of customers whose diverse usage patterns create the substation capacity requirement. At the point of connection to the distribution circuit, however, for larger commercial customers, it is most common to find a dedicated transformer or transformer bank for each customer, meaning the components must be sized for the individual customer's highest demand whenever it occurs. However, if distribution costs are recovered in time-varying charges, the continuous demand customer, who uses the distribution circuit capacity more uniformly, will benefit from the off-peak rates that will apply to much of their usage. Shared-demand customers will bear the same costs—but those with on-peak usage will bear more, and those with off-peak usage less.

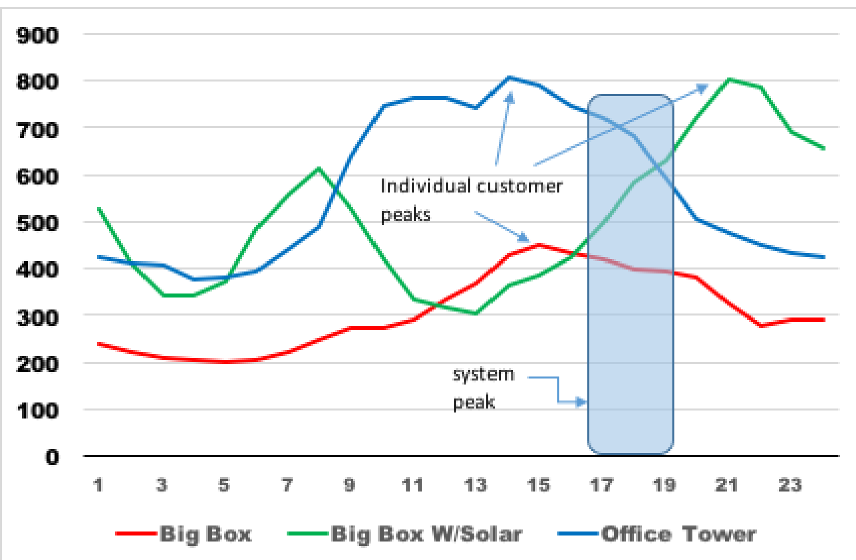
An Example of Diversity: The 2014 Commercial Solar Case

The CPUC recognized many of the issues this paper is addressing in a 2014 rate case (Application 12-12-002). In that docket, a group of large commercial customers with rooftop solar systems petitioned the Commission for relief from the utility rate design, which included both NCP and CP demand charges. The argument was that these facilities posed their highest peak demands to the system on cloudy days, and those were not system peak days. That decision explicitly discusses the arguments against both NCP demand charges (misaligned timing of charges with system peaks) and CP demand charges (failure to recognize load diversity *within* the peak period). The Commission ruled that 75% of generation-related demand charges should be shifted to the TOU energy blocks. They found that this would result in an equitable sharing of capacity costs. In a separate docket (D. 12-12-080), the Commission extended this finding to PG&E.

The evidence in that docket showed that even at the large commercial level, individual customers have diversified peaks that may differ from the system peak. This means that any demand charge measured on a short interval of one hour or less will shift costs to those customers with lower load factors who will have diversity with other customers and can share generation, transmission, and to a lesser extent, distribution capacity.

Simply looking at three specific Southern California customers illustrates this. The curves show the hourly loads on a peak day for a big-box store with solar, one without solar, and an office tower. The two non-solar customers peak around mid-day; the solar customer's load on the utility peaks in the evening. The combined peak of these three customers occurs at a time different from the peak use of any of the three. With hundreds of thousands of commercial customers in California, the diversity effect is very large. Only time-varying volumetric rates, whether TOU or dynamic, equitably share system capacity costs among multiple customers with different load patterns. Any rate design that bases capacity cost recovery on short intervals, whether concurrent with the system peak or not, is ineffective at encouraging each customer to shape load based on true system needs.

Figure 6: Hourly Loads for Three Southern California Customers[32]



Three things are important to note in this depiction of the diversity effect of multiple customers:

1. All three customers peak at different times of the day;
2. None of the three customers has their individual peak at the time of the maximum combined system peak; and
3. A customer assigned a demand charge based on their highest usage within a multi-hour

[32] Source: Sean Swe, Rate Analyst, Burbank Water and Power.

coincident peak period will bear a higher share of capacity cost than their actual share of usage at the system peak period.

A TOU rate recovering capacity costs will provide an incentive for customers to constrain their usage during each hour of the on-peak period. A CPP rate will provide an incentive for customers to constrain their usage during specific hours of system stress. Both of these are superior price signals compared with a demand charge levied on the highest use within a pre-defined multi-hour period.

What Can We Learn from SMUD?

C&I customers have the potential to invest in cost-effective DERs or DER portfolios that benefit the enterprises themselves and the management of the electric system, while contributing to achievement of California's clean energy goals. For example, combined solar and storage systems can make it possible for a customer to meet part of its own energy needs while providing capacity benefits to the system if the storage unit can be called upon during peak hours that may differ from the customer's own peak hours. Ice storage systems that "charge up" during times of system excess can help reduce renewable curtailment, help eliminate periods of negative pricing, and provide customers with a way to more cost-effectively cool spaces. EV charging, either for fleets, a company's own workforce, or the patrons visiting a commercial entity, can be managed to occur when solar is plentiful and power is cheap, thus enhancing the economics of transportation electrification, and curtailed when energy costs are high. These are but a few examples of the potential contributions that C&I customers can make toward integrating renewable energy, managing demand and reducing the need for expensive, higher-carbon peak resources.

The examples discussed above can be encouraged with good rate design—or hindered by the opposite. For example, non-coincident peak demand charges can be a significant barrier to commercial and industrial customers' willingness to provide EV charging on site or to meet its own EV fleet charging needs. Though the individual customer's peak may not coincide with the system's most stressed time, this rate structure would give the customer a strong disincentive to allow EVs to charge, if doing so would cause them to trigger a new spike in their local monthly demand. Reducing or eliminating NCP demand charges, with more costs recovered through a CP demand charge or TOU energy charges, will greatly enhance the economics for C&I customers debating whether to host EV charging at low-cost times for the grid, as well as for the drivers of EVs (assuming some amount of the charging cost is passed through to them).

Table 5 on the following page compares the rates for a 300 kW secondary voltage electricity customer (supermarket; medium office building; big box retail) on the Southern California Edison system and the Sacramento Municipal Utility District system. We pick one utility to compare with SMUD for simplicity, but any one of the three utilities could have been used for this comparison. We do not intend to call out SCE as having worse rate design than the other IOUs. The notes discuss the effect of each rate element.

Table 5. Comparison of SCE and SMUD rates for 300 kW Secondary Voltage Customer

SCE TOU-3					
	Rate	**Unit**	**Metric**	**Costs Covered**	**Comment**
Customer Charge	$446.13	Month		Customer-specific and transformer	Greatly exceeds metering and billing costs; transformer cost varies with kW. Lower to a cost-based level consistent with customer-specific metering and billing costs.
Distribution Demand Charge	$13.17	kW	NCP	~60% of distribution cost	NCP application does not reflect cost causation; concentrates costs on lower load factor customers who can share capacity. Isolate customer-specific capacity costs such as final transformer in demand charge, perhaps in contract facilities charge. Shift balance to a TOU energy rate to reflect sharing of capacity.
Distribution Energy[33]	$0.02570	kWh	All	~40% of distribution cost	Lack of TOU differentiation does not reflect cost causation; shift to a TOU basis, with off-peak at current level, and demand charges shifted into mid-peak and on-peak periods.
Transmission Demand	$4.64	kW	NCP	All transmission cost	NCP application does not reflect cost causation; concentrates costs on lower load factor customers who can share capacity. Shift to a TOU basis with baseload transmission costs reflected in all hours; peaking transmission reflected in peak hours.
Generation Demand					
Summer On-Peak	$17.42	kW	15-minute	Most generation capacity cost	Exceeds peaking capacity cost for short-duration capacity such as DR appropriate to 15-minute metric. Shift to TOU or critical peak volumetric prices.
Summer Mid-Peak	$3.43	kW	15-minute	Limited generation capacity cost	Shift to TOU.
Winter Mid-Peak	-				.
Generation Energy					Shift to a broad TOU rate
Summer On-Peak	$0.08819	kWh	12-6 pm	Marginal variable energy cost	Include relevant capacity costs in TOU rates.
Summer Mid-Peak	$0.05095	kWh	Other	Marginal variable energy cost	Include relevant capacity costs in TOU rates.
Summer Off-Peak	$0.03226	kWh	Night/ weekend	Marginal variable energy cost	Include relevant capacity costs in TOU rates.
Winter Mid-Peak	$0.04662	kWh	8 am-9 pm	Marginal variable energy cost	Include relevant capacity costs in TOU rates.
Winter Off-Peak	$0.03712	kWh	Night/ weekend	Marginal variable energy cost	Include relevant capacity costs in TOU rates.

SMUD Equivalent Rate: GS-TOU-3 300-499 kW					
	Rate	**Unit**	**Metric**	**Costs Covered**	**Comment**
Customer Charge	$106.85	Month		Metering, billing, collection, customer service	Reflects customer-specific costs for metering and billing.
Site Infrastructure	$3.76	kW	Contract demand	At-site equipment, including transformer	Consistent with site-specific costs; fixed from month to month.
Summer Super Peak	$7.57	kW	15-minute	Needle-peaking generation, transmission, distribution?	Only applicable five hours a day, weekdays, summer. Use of longer metric would improve equity to shared demand customers.
Energy Charge					
Summer Super Peak	$0.1986	kWh		Baseload, intermediate, peaking generation, transmission, distribution	Incorporation of most capacity costs into TOU energy rates adds simplicity, equity.
Summer On-Peak	$0.1357	kWh		Baseload, intermediate generation, transmission, distribution	
Summer Off-Peak	$0.1079	kWh		Baseload generation, transmission, distribution	
Winter On-Peak	$0.1032	kWh		Baseload, intermediate generation, transmission, distribution	
Winter Off-Peak	$0.0820	kWh		Baseload generation, transmission, distribution	Off-peak energy rates recover baseload resource costs.

[33] Calling this "distribution energy" may be misleading or confusing. Almost all of the charge is Public Purpose Programs Charge, New System Generation Charge, etc. Only $0.0021 is "distribution."

The SCE on-peak energy charge in summer is about one-third that of SMUD, while the demand charges are three-times those of SMUD. The SCE rate creates powerful incentives, in the form of a combined generation, transmission, and distribution demand charge of about $30/kW, for customers to focus on limiting their individual 15-minute peak demand within the on-peak period. By contrast, the SMUD rate, by recovering most of these costs in the TOU energy rate, provides a limited incentive for focus on the 15-minute demand, and a much larger focus on controlling usage across the entire super-peak period. Since generation, transmission, and network distribution capacity is shared, an individual customer's 15-minute demand is largely irrelevant to system capacity planning; the collective usage of all customers in the various periods drives the resource development needed to provide reliable service.

The SMUD rate is not ideal, in that the super-peak demand charge is still based on the customer's highest 15-minute usage within a 132-hour (six hours a day, weekdays, per month) period. A lot of diversity is possible within that period. The challenge, of course, is to identify in advance when the system peak will occur, to set prices that coincide with that peak. A better approach, in our opinion, is to use TOU energy charges to recover predictable capacity costs across the broad system-peak period, and then to use critical peak pricing or other demand-response measures for the extreme system peak periods, which occur only sporadically on days that cannot be predicted in advance. California's extensive experience, and global experience, with critical peak pricing shows that it is effective.

NCP demand charges incentivize customers to use storage in a way that benefits their own load shape, not provide grid benefits (i.e., at the system's peak times).

Optional Real-Time Pricing for Sophisticated Energy Managers

Some analysts argue that real-time pricing—where a significant amount of shared system costs are recovered through an energy charge that fluctuates freely every hour—gives customers a much stronger incentive to curtail loads at high-cost times in a way that reduces costs and carbon emissions. Others argue that real-time pricing signals in California today are not strong enough nor predictable enough to induce investment in energy management technologies. These analysts would argue that a TOU pricing targeted on a narrow period, say a two-hour period, or a critical peak price is a more concentrated price signal and more certain, and thus would more readily elicit active load management and possibly even an incremental DER investment response. Real-time prices more accurately convey short-run market price information and they can be readily extended to more granular locational pricing on the distribution system over time, so there is reason to proceed with RTP pilots that give active energy managers the opportunity to build value propositions around these prices. However, there is also a case to be made for continuing to offer TOU and critical peak pricing options, as they may better reflect long-run costs and facilitate greater customer understanding and response during occasional periods of system stress.

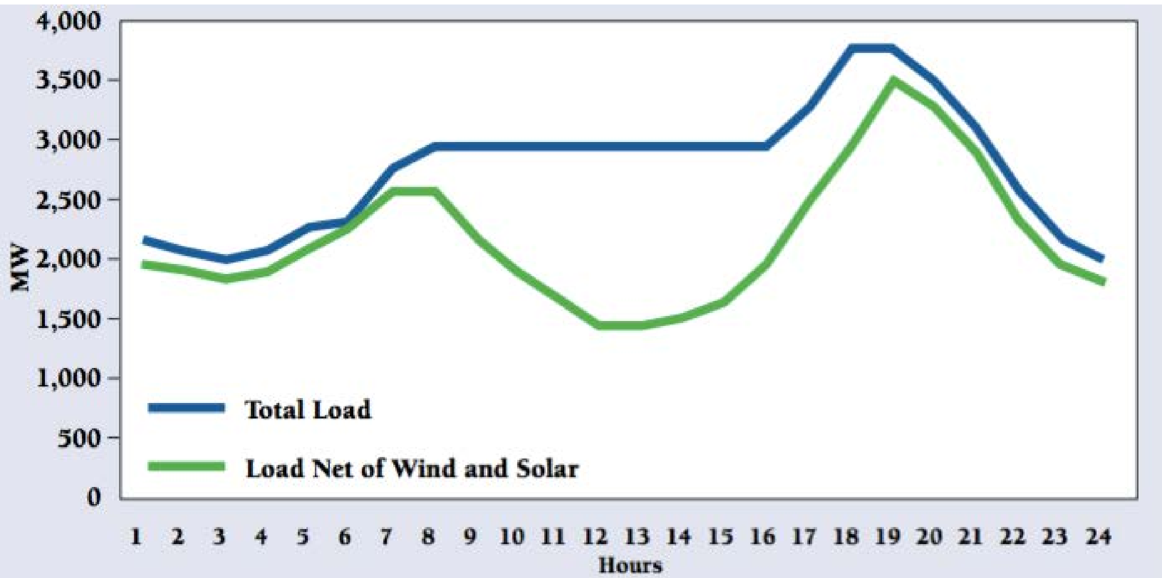
We will talk about experiences with RTP and CPP more in Section V, as we examine what we can learn from others outside California.

Some Examples of How Rate Design Affects Beneficial DER Adoption and Operation

California faces the challenge of increasing reliance on variable renewable resources, and many opportunities to utilize emerging technologies to meet that challenge. Beneficial adoption and operation of DER resources can help with these integration challenges, but rate design needs to support and not deter such beneficial use.

The prospect of managing larger ramping requirements, along with attendant grid stability challenges, is raising some concerns among utility engineers and managers. See Figure 7 below.[34]

Figure 7. Illustration of Daily Load Pattern Faced by Utility



These issues are being exacerbated by the problems with California's rate design—particularly the still insufficient emphasis (discussed above) on time-varying rate designs. We identify below several separate elements of these challenges and opportunities.

1. High-impact EE providing savings in key hours

There is need for increased energy efficiency programs targeted at specific customer uses in problem hours. For example, LED lighting retrofits can help reduce the upward movement seen in both the blue and green lines in Figure 7 during the hours around 5 p.m. (i.e., the "shoulder period").

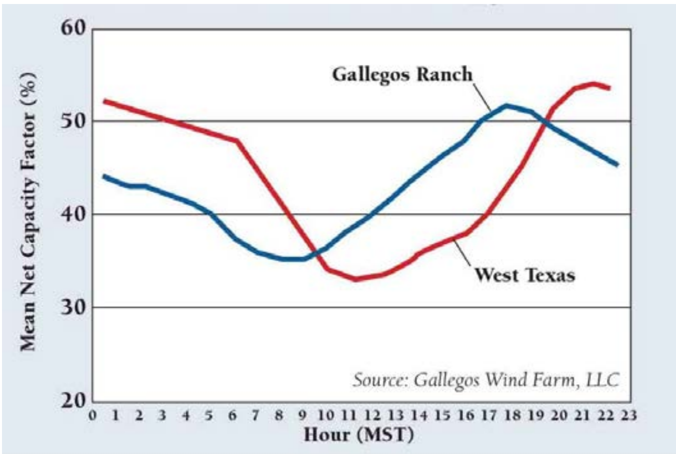
In addition, air conditioning is a major driver of peak demand, and could be a profitable target for EE and thermal storage investments. A time-varying rate directed toward A/C would pay a double dividend, as A/C has great potential for both demand response and thermal energy storage).

[34] Jim Lazar, "Teaching the 'Duck' to Fly (Second Edition)," Regulatory Assistance Project, https://www.raponline.org/knowledge-center/teaching-the-duck-to-fly-second-edition/.

2. DG that performs in key hours

Except in the spring months, solar energy is a generally load-matched resource, serving loads during the business day, when loads historically have increased. But at some point, other times of day become a challenge, and solar ceases to be well matched to loads. For example, one wind developer in New Mexico published a very favorable wind profile:

Figure 8. Annual 80-m Diurnal Wind Energy Patterns



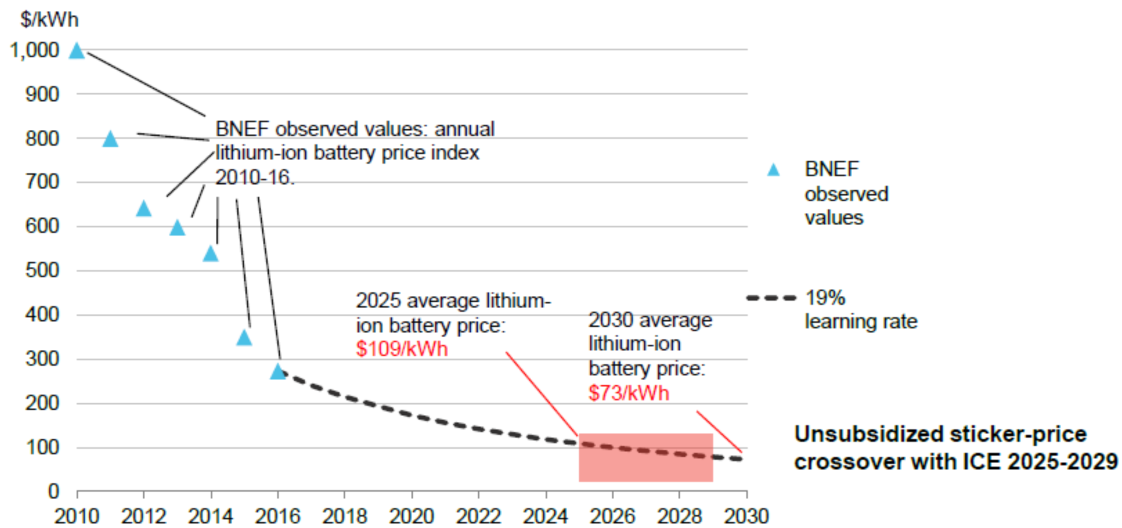
Using biomass resources, which are generally fully dispatchable, becomes an important option. Redispatch of California's extensive hydro resources is an important opportunity; these have historically been dispatched during the mid-day peak, but as the peak to be served with controllable resources shifts, so can the dispatch of these resources.

3. Storage: both thermal and electrical storage

The cost of electrical energy storage has been coming down rapidly, although it is still quite expensive. There are growing expectations of further price declines.

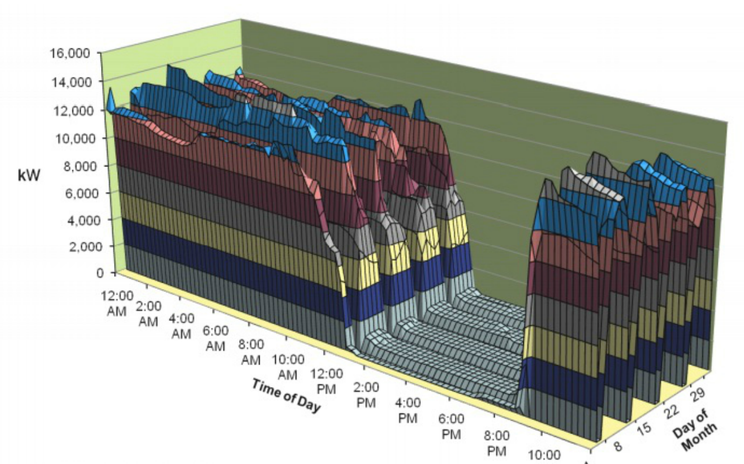
Storage can provide peak power anytime (and almost anywhere); renewable energy produced during low-demand periods of the day can effectively be stored until a later, more valuable time period. An extremely important potential resource is ice and chilled water storage for air conditioning. These technologies can be charged during low-cost periods of wind and solar production, and then utilized for cooling later in the day. Austin Energy has deployed several central cooling plants in their commercial core. Advanced non-residential rate design will encourage cost-effective deployment of customer storage resources.

Figure 9. Declining Costs of Battery Storage[35]



A less effective, but still valuable form of thermal storage can be achieved without any capital investment, by simply pre-cooling a structure before the on-peak period, and allowing it to "coast through" the high-cost period. In buildings with substantial structural, furniture, paper, and equipment thermal mass, the temperature will rise slowly during the coast-through period.

Figure 10. Central Cooling Energy Management at Austin Energy



Source: Austin Energy

[35] Bloomberg New Energy Finance.

4. EV fleets

EV batteries present the opportunity to control the timing of when they are charged, and perhaps the future option of reversing flow from batteries during periods of severe grid stress.

Ideally, EV owners and operators will choose to charge EVs during the times of day when power is plentiful and lower-cost, and impose little or no incremental peak demand to the system.

A key emerging opportunity is the charging of EV fleets. Providing the right incentives for commercial managers of these fleets is a promising area, because these fleets many ramp up quickly in size and because these managers should be sensitive to charging costs and have the flexibility to manage charging.

However, demand charges (especially NCD charges) are a major barrier to economic EV fleet charging. A recent study from the Rocky Mountain Institute identified demand charges as a major barrier to EV deployment.[36]

5. Workplace charging of employee EVs

Another significant opportunity is in the area of workplace charging of employee EVs. Under current typical NCP demand charge rate designs, employers would likely spike their monthly demands by allowing workplace charging. But, if it can be limited to low-cost hours, workplace charging can help adapt the system to higher levels of variable renewables. Employees who own EVs sometimes need to charge during the day, while at work, and may take advantage of such opportunities at the workplace. From a power system point of view, this may make sense as it can help align power usage with, say, low-energy-cost late mornings. Good rate design—particularly time-dependent design principles discussed in this paper—can help encourage more workplaces to offer employee and visitor charging.

The Electric School Bus: A Special Opportunity

Many transit agencies are implementing electrification programs, and SCE recently filed for a five-year waiver of demand charges to help enable transit electrification. However, electric school buses may be an even more attractive option:

a) Transit buses operate for up to 300 miles per day; carrying adequate battery capacity is difficult and expensive, and deploying high-speed charging systems is expensive and challenging for the grid.

b) School buses typically travel about 50 miles per day, making the battery capacity lower, and adding flexibility to charge at mid-day (solar) or overnight (wind).

c) School buses are typically sedentary from 10–2, peak hours to absorb excess solar generation.

d) School buses could be an ideal vehicle-to-grid resource. One study by the Vermont Energy Investment Corporation showed that the school bus fleet of New England could supply about 1,000 MW of peaking capacity to the grid in a V2G configuration.

e) BUT: Charging on this type of schedule will produce extremely poor load factors, but can produce very good load shape.

Rate design will have a significant influence on the viability of this technology.

[36] Rocky Mountain Institute, "From Gas to Grid," 2017, https://www.rmi.org/insights/reports/from_gas_to_grid/

6. Suboptimal operation of existing "flexibility"

California's power system has existing flexibility in several forms. There are existing hydro resources (discussed above) that provide great flexibility. The Helms pumped storage facility, originally constructed to provide flexibility to the output of the Diablo Canyon nuclear plant, can be used for any purpose. System flexibility resources can be directly controlled by the grid operator, and easily redeployed.

But customer flexibility is a different matter. Customers will respond to economic incentives and programmatic offerings. Existing air conditioner cycling programs, for example, may require programmatic changes to align the ability to control these loads to current needs.

Current rate designs misalign customer benefits (opportunities for bill reduction) with grid benefits, causing misuse of storage resources (from a grid perspective) and neglected opportunities to provide grid benefits (e.g., by increasing energy uptake during renewable curtailment hours or by pre-cooling during hot summer days).

Customer investment in new storage technology, whether electrical or thermal, depends on the rate design the customer faces. And proper utilization of that technology is important. High demand charges, as now exist, will indeed encourage customer investment in storage; Calmac, a leading provider of ice storage for commercial air conditioning, has indicated to the authors that a demand charge of $14/kW (or a $.06/kWh TOU differential) will produce an acceptable return on their systems for most commercial building operators to invest in storage. BUT, with a demand charge in place, the incentive then will be to use that storage to levelize load at the lowest level achievable, not to optimize load across all hours of the day. Only a time-varying rate will encourage the building operator to optimize usage in the context of overall grid costs. For example, the office tower in Figure 6 on p. 31 would likely use storage between 10 a.m. and 4 p.m., to reduce their demand charge (and thereby their demand charge); however, the system would be better off if that same storage were used to reduce load between 5 p.m. and 8 p.m., when the system peak occurs. The shift of demand charge revenues into TOU rate periods will enhance customer deployment and utilization of cost-effective on-site storage technologies.

7. Solar plus storage: Bad rate design results in uneconomic deployment

NCP demand charges can influence C&I customers with solar PV generation on-site to invest in otherwise uneconomic storage in order to reduce demand charges. In particular, customers with on-site solar may serve their daytime peak demand with solar, leaving a short spike at the end of the solar day before operations wind down. This may be uneconomic when the customer's on-peak load does not match the system load—the storage may be deployed to shave a peak that has little or no consequence for the utility system, and then the storage may be unavailable to reduce system peak demand. RAP addressed this in a posting earlier this year[37], and a recent paper from NREL and LBNL studied this phenomenon in greater detail.[38]

[37] Jim Lazar, "With Sinking Storage Costs, Big Box Solar Could Really Take Off," Regulatory Assistance Project, 2017, http://www.raponline.org/blog/with-sinking-storage-costs-big-box-solar-could-really-take-off/

[38] Gagnon et. al., "Solar + Storage Synergies for Managing Commercial-Customer Demand Charges," Lawrence Berkeley National Laboratory, https://emp.lbl.gov/publications/solar-storage-synergies-managing

V. What Can We Learn from Others?

Comparison of Large Commercial Rates of Major US Utilities

As a part of this project, RAP reviewed the commercial rates of general application for many of the largest electric utilities in the US, for the large utilities in California, and a few selected international examples. Table 6 on p. 41 shows a partial summary of that review.

In keeping with the principles of smart rate design we outlined above and the other opportunities and challenges we have explored thus far, key features we looked for in these rate designs include:

- Reasonable customer charges reflecting the recovery of customer-specific costs such as the service drop, metering, billing, and basic customer service expenses;
- Small NCP demand charges focused on the most local distribution cost recovery, such as the final line transformer, where component sizing must track individual customer load;
- Coincident peak demand charges for generation, transmission, and shared distribution system costs, focused on no more than a five-hour peak period;
- Seasonal energy charges;
- TOU energy charges; and,
- The emphasis of cost recovery in energy charges, rather than demand charges.

We note that few of the national examples are as sophisticated as the existing California IOU rates. In fact, no other domestic utility that we surveyed offers a tariff with both seasonal and time-varying energy charges (some have one but not the other). We also found it notable that very few other utilities outside of California from non-restructured states offer separate distribution and generation demand charges. This additional level of information granularity available in California may become useful as more customer-sited generation resources are brought onto the system, though we note that both LADWP and PG&E do not yet offer that type of rate. We were surprised to find a number of Wright-Hopkinson rate forms (load factor blocks), as we generally view this rate design as quite antiquated. We found very few utilities with coincident peak demand charges, though several do have time-varying energy charges. Outside of California, most utilities offer a low customer charge (with the exception of Portland General Electric). Georgia Power's tariff is notable because it is one of the few with a CP demand charge and time-varying energy charges.

Pragmatically, the only example that met all of our general criterial is the current rate from SMUD, set forth earlier in Table 5 on p. 33, which we consider a good illustrative rate for discussion for the other California utilities. SCE offers a generation demand charge that is time-varying, but it is not coincident with system peak and it is combined with an NCP distribution demand charge. SDG&E is quite similar in that they offer time-varying generation and distribution demand charges, but neither are coincident with system peak. The SMUD rate design is not perfect, but it aligns with most principles.

Table 6. Overview of Non-Residential Rates Applicable to 300kW Commercial Customer

		Customer Charge	Combined or Distribution Demand Charges				Separate Generation Demand Charges			Hopkinson Rate	Energy Charges	
Utility	**Schedule**	**$/Month**	**Flat?**	**Seasonal?**	**TOU?**	**Coincident ≤ 5 Hours**	**Seasonal?**	**TOU?**	**Coincident ≤ 5 Hours**	**Load Factor Blocks**	**Seasonal?**	**TOU?**
LADWP	A2B	28	No	No	Yes	Yes	No	No	No	No	Yes	Yes
PG&E	A10	138	No	Yes	No	No	No	No	No	No	Yes	Yes
SDG&E	AL-TOU	116	No	Yes	Yes	No	Yes	Yes	No	No	Yes	Yes
SCE	TOUGS3	446	Yes	No	No	No	Yes	Yes	No	No	Yes	Yes
SMUD	GSTOU3	107	Yes	No	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Florida Power	GSDT1	25	Yes	No	No	No	Yes	Yes	No	No	No	Yes
Virginia Electric	GS2t	26	Yes	No	No	No	Yes	Yes	No	No	No	Yes
Georgia Power	GSD10	209	No	No	Yes	Yes	-	-	-	No	No	Yes
Duke SC*	LGS	17	Yes	No	No	No	-	-	-	Yes	No	No
Detroit Edison	D4	14	Yes	No	No	No	-	-	-	Yes	No	No
Duke Florida	GSDT1	12	No	No	Yes	No	-	-	-	No	No	Yes
Ameren Missouri*	LGS	94	No	Yes	No	No	-	-	-	Yes	Yes	No
Alabama Power*	LPM	50	Yes	No	No	No	-	-	-	Yes	No	No
Duke NC	SGSTOU42A	29	No	No	Yes	No	-	-	-	No	No	Yes
NoStates Power*	A14	26	No	Yes	No	No	-	-	-	Yes	No	No
PEPCO MD	MGT3A	40	No	Yes	Yes	No	-	-	-	No	Yes	No
Portland GE	NEDNR	420	No	No	Yes	No	-	-	-	No	No	Yes
Eskom South Africa	Megaflex		Yes	No	No	No	-	-	-	No	Yes	Yes
Shanghai China			Yes	No	No	No	No	Yes	No	No	Yes	Yes

Note: "Flat" demand charges are those that do not have any time or seasonal variation to them. Utilities marked with * also offer customers of this size optional tariffs with time-varying energy charges.

Texas: Energy-Only Market

Texas has some features worth paying attention to in California, related to the design of the ERCOT market. First, because the ERCOT energy market—which has a footprint that covers most of the state—does not include any generation capacity cost mechanism, there is no per-kW cost that needs to flow through to generation cost allocation and rate design. In other words, ERCOT is an "energy-only" market in which generators (and other resources) in the ERCOT market recover their captial costs during peak periods when energy prices are high.[39] All generation costs (including fuel costs and capital costs, etc.) are allocated down to the LSE level in the form of energy (per/kWh) prices. As a result, the scope for efficient time-varying allocation of generation costs to end users is, at least in principle, greater than in jurisdictions where generator fixed costs are recovered through fixed charges.[40]

Second, in ERCOT, embedded transmission costs are allocated to some medium/large commercial end-users based on the individual customer's coincident peak, with usage in four ERCOT peak hours (one in each of four summer months) used to determine charges for the subsequent year. In the ERCOT footprint, service providers help commercial and industrial consumers to predict the ERCOT system peak hours, so that usage can be adjusted and costs reduced.[41] These customers typically receive forecasted warnings of upcoming grid peak periods that will determine the customer's transmission charge for the following year.

Illinois Hourly Pricing Tariffs

In Illinois, the two largest utilities offer real-time pricing options known as "hourly pricing."[42] These options are for residential customers, but we discuss them here because the general approach could easily be applied to non-residential customers in California. The tariff includes a generation (production) energy cost component—a per-kWh price that fluctuates hourly, in line with PJM LMP wholesale market prices. Other per kWh and per kW components collect transmission and distribution costs, as well as generation capacity costs not included in the PJM LMP wholesale prices. The results reported to date indicate that residential customers would have paid on average $86 less per year had they been on the real-time pricing tariff, and the benefits were broadly evident with more than 90% of customers projected to have lower bills under the RTP tariff**.**

Georgia Power Real-Time Pricing Tariffs

Georgia Power is a large investor-owned utility with a significant industrial base. For more than ten years, it has offered these customers a real-time pricing option. This is characterized as "baseline-

---

[39] Note that, as in California, congestion costs are also reflected in ERCOT wholesale energy market prices (and these congestion costs reflect the marginal cost of grid usage at a given location).

[40] It is important not to try to make a direct analogy between California and Texas. Texas has a much greater degree of retail competition. Retailers offer a range of different rate options, some of which are time-varying in nature.

[41] For detailed discussion, see http://energytariffexperts.com/blog/2013/7/17/ercot-4cp-june-2013-review and https://www.electricchoice.com/blog/4-coincident-peak-program/

[42] See Jeff Zehtmayr and David Kolata, "The Costs and Benefits of Real-Time Pricing," Illinois Citizens Utility Board, 2017, https://citizensutilityboard.org/wp-content/uploads/2017/11/20171114_FinalRealTimePricingWhitepaper.pdf

referenced," because customers only experience real-time prices for deviations in their usage (up or down), not for the total usage. The mechanism has the following characteristics:

- A customer baseline is established for each participating customer;
- Usage at the baseline is priced at a price determined through regulation, based on the utility cost of service;
- The customer is given notice of day-ahead prices; and
- Deviations from the baseline usage are charged or credited at the real-time price.

In essence, the customer "subscribes" to power at a regulated price, and then can consume greater or lesser amounts at a real-time price.[43] These tariffs have proven acceptable, and in 2011 became the standard tariff for large-use customers. An option to choose a fixed-price tariff is available after three years on a real-time rate.

Washington Real-Time Pricing Experiment

In 1996, industrial customers of Puget Sound Energy, the largest electric utility in Washington State, requested access to wholesale market pricing for electricity. The approach that was approved had three key elements:

a) a transition charge for three years, during which time they paid a portion of the cost for stranded utility generating capacity until it could be absorbed by growth in usage by other customers;

b) a delivery charge based on the cost of transmission services; and

c) a daily price for on-peak and for off-peak power, based on day-ahead wholesale prices at the largest regional trading hub for electricity.

For the first three years, wholesale market prices were significantly lower than the costs embedded in retail rates, and the customers saved millions of dollars. In the fourth year, the western United States suffered a drought that reduced hydropower availability and put extreme pressure on natural gas supplies to provide relief generation, generally known as the California Energy Crisis of 2000-2001. Wholesale market prices soared to previously unknown levels. The customers, fully exposed to market prices, took drastic steps to adapt, including renting onsite diesel generators and curtailing operations. One major industrial facility, the Georgia-Pacific pulp and paper mill in Bellingham, WA, did not survive the economic impact of the power crisis, and closed permanently. Eventually, in October 2000, the customers approached the Washington Utilities and Transportation Commission for regulatory relief, which was granted in the form of permission to enter into long-term contracts for power with non-utility suppliers, a form of open access not available to other retail customers.

Experience in regions providing open access to industrial customers suggests that some large users will choose a fixed-price plan over a dynamic rate, because the stability of cost allows them to make

[43] Georgia Real Time Pricing Day-Ahead With Adjustable CBL Schedule "RTP-DAA-4."

reasoned business decisions. An industry making sales commitments at contract prices for delivery months or even years ahead may prefer to "lock in" as many cost drivers as possible, including power supply costs.

We hope the experience of the energy crisis period is never repeated, but this is a cautionary tale worth remembering as customers assume greater risk for self-provision.

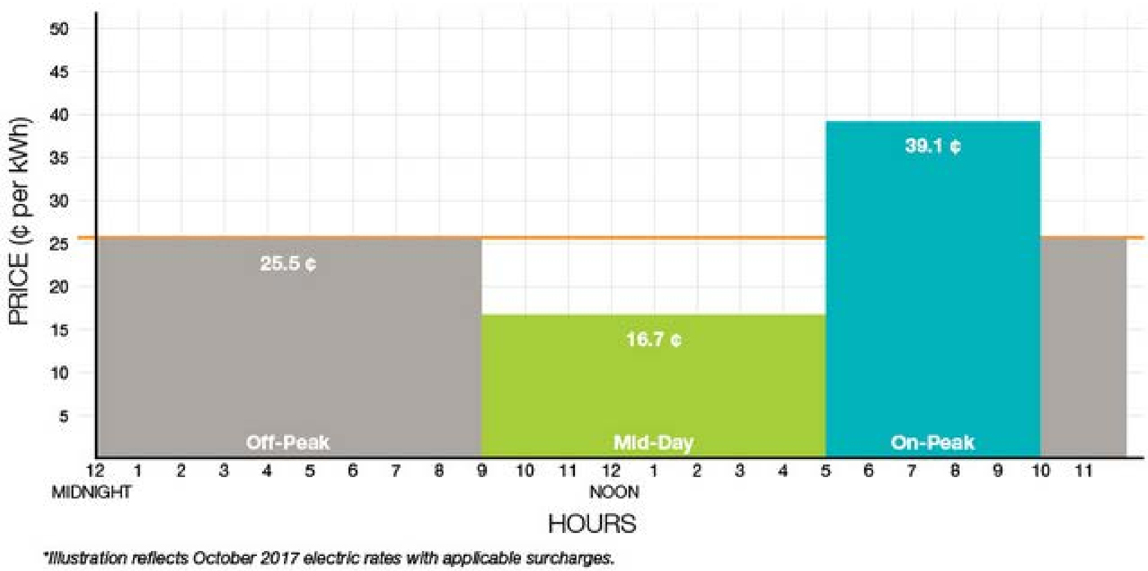
Hawaii TOU Rates

Hawaii has the highest level of installed solar as a percentage of system load of any US state, with installed solar exceeding 50% of peak demand on some islands. Renewables are being regularly curtailed, and new solar connections are being strictly limited to enable adequate grid management capability. The Hawaii PUC recently allowed additional solar installations if they are controllable by the grid operator, or have attached storage that can be programmed to take any excess generation that might be exported.

While this is a residential example, it is relevant as the only example we have found of a system where solar energy deployment has resulted in the mid-day being the lowest-cost pricing period. The CAISO has suggested that this is appropriate for California, at least during the spring months and weekends.

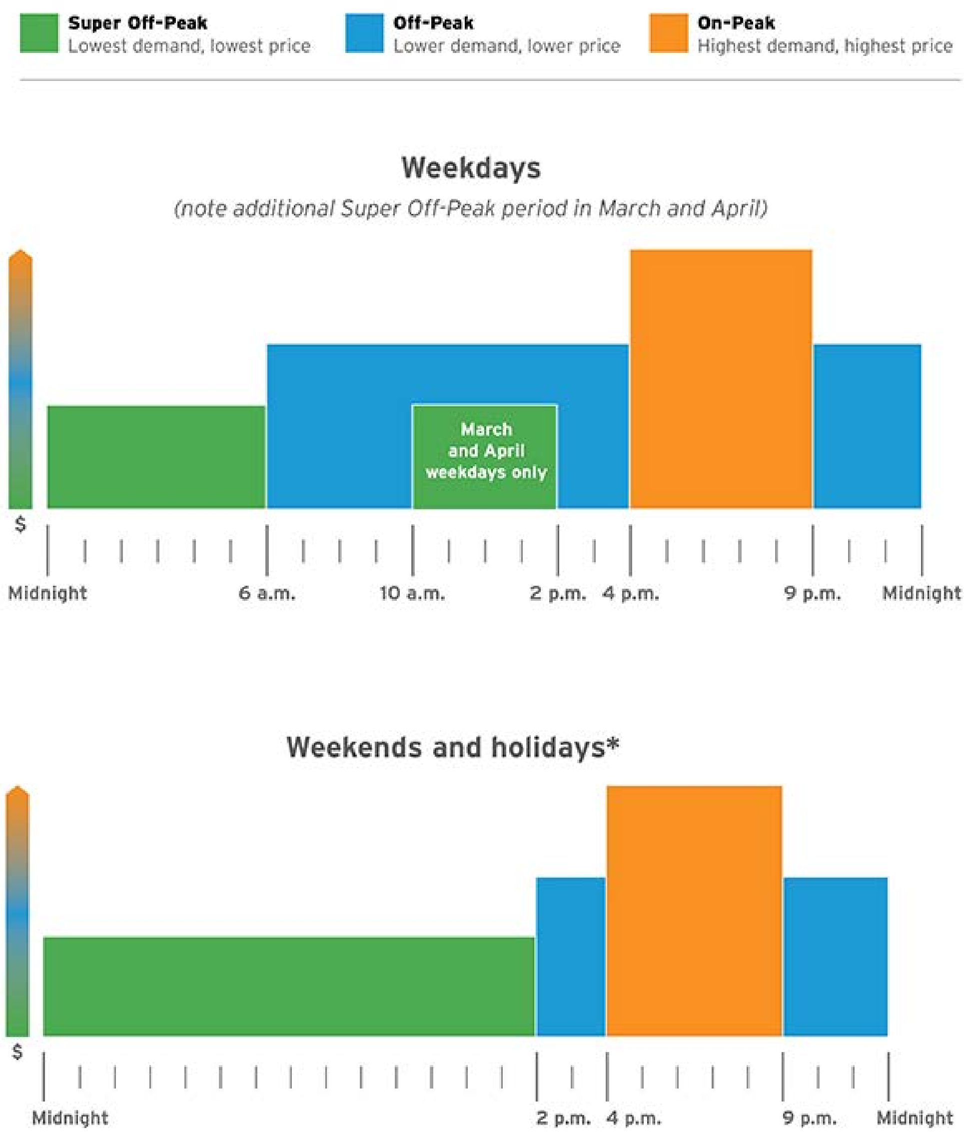
The pilot TOU residential rate in Hawaii may be an indicator of the type of rate form that will become applicable in other sunny regions in the future:

Figure 11. Interim Time-of-Use Rates, Hawaii Pilot



San Diego Gas and Electric has recently filed rates that move in this direction, but do not yet reflect the sharp mid-day depression in costs being experienced in Hawaii. Effective December 1, 2017, SDG&E large commercial TOU rates will have the following form, with a super-off peak rate at night, and during the daytime in the spring "Duck Curve" months. We expect this transition in rates to continue to evolve into the shape of the Hawaii TOU pilot rate design.

Figure 12: San Diego Gas and Electric Commercial Rates, Effective December 1, 2017



Maryland: Exploring Time-Varying Distribution Rates in a Restructured State

In a November 2017 order, the Maryland Public Service Commission took a next step toward time-of-use pricing for residential customers, augmenting their critical peak rewards program that applies to all customers. The Commission directed a workgroup of stakeholders and experts to develop two pilots for each of the three largest utilities. The pilots are worth mentioning because they include both a time-varying rate for distribution service and a time-varying rate for supply, both under default service (called "standard offer service") and through retail supply. For all six pilots, there is to be a five-hour afternoon peak for distribution rates from June to September and a three-hour morning peak from October to May. Retail supplier pilots must offer a three-to-five-hour afternoon peak for supply from June to September, with a winter morning peak being optional. The Commission also allows suppliers to offer additional innovations, such as “free Saturdays,” subject to review. The Commission hopes that this will better incent real-time peak-shaving behavior that advanced meters are now enabling, and is clearly thinking about the innovations needed in rate design to enable the ongoing transformation of the electric sector. These pilots, as described in the Commission’s order, will result in data on customers’ response to time-varying elements of distribution and supply rates that can enable future opportunities for more innovative tariffs. The disaggregated nature of the rate structures to be tested (e.g., separate distribution and supply time-varying elements) should provide insights into how rate design can make things like smart thermostats, electric vehicles, distributed generation, and energy storage more attractive to ratepayers and beneficial to the system.

SMUD: Most Costs in TOU; Coincident Peak Demand Charges

The non-residential rate design we found that best comports with the principles and elements we have described earlier in this paper is that for the Sacramento Municipal Utility District. SMUD’s non-commercial rate has the following characteristics:

- Fixed charge to recovery customer-specific costs of billing, collection, and customer service;
- Site infrastructure cost ($/kW) to recover location-specific capacity costs;
- Super-peak demand charge ($/kW) to recover marginal T&D capacity costs associated with oversizing the system for extreme hours; and
- TOU energy cost to recover all generation costs and remaining T&D costs.

Table 7: SMUD Large Commercial Rate Design

Customer Charge	$108/month	
Site Infrastructure Charge	$3.80/kW/month	
Super Peak Demand Charge	$7.65/kW	
Energy Charge	Summer	Winter
Super Peak	$0.20	N/A
On-Peak	$0.137	$0.104
Off-Peak	$0.109	$0.083

Takeaway Lessons

From our examination of commercial rates globally for this project, we found many antiquated rates, and a few examples of modern rate principle application.

- Most utilities retain NCP demand charges to recover shared capacity costs, which we consider to be poor guidance for cost control, and inequitable to lower load-factor classes;
- A few utilities have migrated to CP demand charges, which are better, but still inferior to time-varying rates;
- Many utilities have implemented TOU rates for large commercial customers, but in many cases only for variable energy-related costs;
- Texas, Illinois, and Georgia have implemented real time pricing programs that have produced benefits, but the Washington experience with market pricing is a cautionary tale indicating that while there are benefits, there also potential risks;
- Hawaii is experimenting with aggressive changes in TOU structure in response to very high solar DG penetration;
- Maryland is implementing an interesting example of a distribution rate with time-varying pricing components; and,
- SMUD has set itself apart as an industry pace-setter with a rate design that reflects most modern rate principles, but we believe their rate design can be improved further.

VI. Concluding Recommendations

RAP's *Smart Rate Design for a Smart Future*[44] undertook an extensive discussion of residential and small commercial rate design, and identified three principles that apply to all customer classes:

- Principle 1: A customer should be able to connect to the grid for no more than the cost of connecting to the grid.
- Principle 2: Customers should pay for grid services and power supply in proportion to how much they use and when they use it.
- Principle 3: Customers who provide services to the grid should be fairly compensated for the value of what they supply.

In this paper, we propose smart non-residential rate principles that build off of the first two of these three. We propose:

- Non-Residential (NR) Principle 1: The service drop, metering, and billing costs should be recovered in a customer fixed charge, but the cost of the proximate transformer most directly affected by the non-coincident usage of the customer, along with any dedicated facilities installed specifically to accommodate the customer, should be recovered in a NCP demand charge.
- NR Principle 2.1: De-emphasize NCP demand charges except as noted in NR Principle 1. All shared generation and transmission capacity costs should be reflected in system-wide time-varying rates so that diversity benefits are equitably rewarded.
- NR Principle 2.2: Shift shared distribution network revenue requirements into regional or nodal time-varying rates. This recognizes that some costs are required to provide service at all hours, and that higher costs are incurred to size the system for peak demands.[45]
- NR Principle 2.3: Consider short-run marginal cost pricing signals and long-run marginal cost pricing signals together in establishing time-varying rates for system resources.
- NR Principle 2.4: Time-varying rates should provide pricing signals that are helpful in aligning controllable load, customer generation, and storage dispatch with electric system needs.
- NR Principle 2.5: Non-residential rate design options should exist that provide all customers with an easy-to-understand default tariff that does not require sophisticated energy management, along with more complex optional tariffs that present more refined

[44] Jim Lazar and Wilson Gonzalez, "Smart Rate Design for a Smart Future," Regulatory Assistance Project, 2015, http://www.raponline.org/knowledge-center/smart-rate-design-for-a-smart-future/

[45] One California municipal utility, for example, has TOU rates for commercial customers that include weekends as off-peak, but for residential customers, summer afternoons remain on-peak, due to distribution system capacity constraints on residential circuits. The same concept could apply in different regions or nodes of a distribution system serving non-residential customers, where capacity constraints are reached at different times of the day or year.

price signals but require active management by the customer or the customer's aggregator.

- NR Principle 2.6: Optimal non-residential rate design will evolve as technology and system operations matures, so opportunities to revisit rate design should occur regularly.

RAP applied these principles to evaluate existing commercial rate designs at each of California's investo- owned utilities. We concluded that if rate design is not changed to better align with these principles, California will continue to see underinvestment in DER resources and under-utilization of DER resources toward meeting the state's policy goals.

RAP searched for rate design examples that better comport with the principles in California and elsewhere. As mentioned, the non-residential rate design we found that best does so is that of SMUD. SMUD's non-commercial rate has a fixed charge to recovery customer-specific costs of billing, collection, and customer service; a site infrastructure cost ($/kW) to recover location-specific capacity costs; a super-peak demand charge ($/kW) to recover marginal T&D capacity costs associated with oversizing the system for extreme hours; and a TOU energy cost to recover all generation costs and remaining T&D costs. SMUD's rate sets it apart as an industry pace-setter, but we believe their rate design can be improved further.

One important goal for revision of non-residential rate design should be to better adapt to the incorporation of customer resources, such as thermal or electrical storage, customer provision of ancillary services through smart inverters, and customer load control for peak load management. The general framework of the rate design we propose directly compensates many of these through simple, clear, and compensatory TOU rate elements:

Table 8. Proposed Illustrative Rate Design for Non-Residential Consumers

	Production	Transmission	Distribution	Total	Unit
Metering, Billing			$100.00	$100.00	Month
Site Infrastructure Charge			$2/kW	$2/kW	kW
Summer On-Peak	$0.140	$0.020	$0.040	$0.20	kWh
Summer/Winter Mid-Peak	$0.100	$0.015	$0.035	$0.15	kWh
Summer/Winter Off-Peak	$0.070	$0.010	$0.020	$0.10	kWh
Super Off-Peak	$0.030	$0.010	$0.010	$0.05	kWh
Critical Peak	Maximum 50 hours per year			$0.75	kWh

This design is generally similar to SMUD's, with three important differences. First, it is unbundled between generation, transmission, and distribution to enable more granular application. Second, rather than have a super-on-peak demand charge, those costs are reflected in a critical peak price for up to 50 hours per year. The amount recovered is similar to that for SMUD's super-peak demand charge, but converted to an hourly rate to directly track high-cost hours, and to enable better customer response as system conditions change. Third, we have introduced a super off-peak rate, consistent with the recommendation of CAISO. We have intentionally left the definition of time periods unstated, as these will be specific to particular utilities, to particular nodes within each service territory, and will change over time as loads and resources evolve.

RAP also reviewed a number of real-time pricing tariffs and, while we did not identify one in particular that we would classify as best practice, we have identified lessons learned from Texas, Illinois, Georgia and Maryland that will be useful to the CPUC as it considers RTP optional tariffs. We suggest designing an RTP option that builds from our TOU plus CPP recommendation, and propose the following simple initial design:

- A wholesale energy cost component, charged on a per kWh basis, that fluctuates hourly. This would be based on the relevant CAISO zonal locational marginal price and would replace the "production cost" component of our recommendation above.
- Transmission costs and distribution costs would be collected in the same way that they are collected under our recommendation above, as would any generation capacity costs that aren't accounted for in wholesale rates.

Note that this design would not achieve the full benefits of an ideal RTP approach. In particular, this would not include comprehensive price signals reflecting conditions on the local distribution network. Instead, the hourly pricing innovation here is increased exposure of end users to existing CAISO wholesale prices. Over time, as California introduces new approaches that animate the value stack for resources at the distribution level, new rate designs will be able to incorporate more complex and comprehensive RTP components.

Appendix A: Some Important Rate History

Early Foundations

The best recognized text on utility ratemaking is Bonbright's 1961 *Principles of Public Utility Rates*. Bonbright set forth some principles for a fair rate design that include:

- The related, "practical" attributes of simplicity, understandability, public acceptability, and feasibility of application.
- Freedom from controversies as to proper interpretation.
- Effectiveness in yielding total revenue requirements under the fair-return standard.
- Revenue stability from year to year.
- Stability of the rates themselves, with a minimum of unexpected changes seriously adverse to existing customers. (Compare "The best tax is an old tax.")
- Fairness of the specific rates in the apportionment of total costs of service among the different customers.
- Avoidance of "undue discrimination" in rate relationships.
- Efficiency of the rate classes and rate blocks in discouraging wasteful use of service while promoting all justified types and amounts of use:
    - in the control of the total amounts of service supplied by the company; and
    - in the control of the relative uses of alternative types of service (on peak versus off peak electricity, Pullman travel versus coach travel, single party telephone service versus service from a multi party line, etc.).

Bonbright also provides detailed guidance on fully allocated and marginal cost, on elements of rate design, and on how to approach the "perceptions of equity and fairness" issue.

Another landmark text in utility rate making is Garfield and Lovejoy's *Public Utility Economics* (1964). A portion of this focuses specifically on the recovery of utility system capacity costs—the costs of constructing and maintaining generation, transmission, and distribution facilities. Garfield and Lovejoy cite extensively from a 1949 NARUC rate manual (which we do not have), identifying multiple criteria to set equitable rates to recover these costs. This work is particularly on-point to the issue of commercial rate design that is the focus of this project. We have addressed these in Section II.

Alfred Kahn, an architect of airline deregulation, published *The Economics of Regulation* in 1970, advocating that pricing should reflect marginal costs. After the Public Utility Regulatory Policies Act required states to "consider and determine" whether electricity rates should be based on the cost of service, several states, including California, Oregon, and New York, adopted costing principles based on marginal cost. Most states have retained cost allocation based on accounting (embedded) costs, but many of those use marginal cost principles for rate design, reflecting an accounting cost approach to equity in cost allocation between classes, but accepting the Kahn principles for rate design within class revenue requirements.

Demand Charges: Wright, Hopkinson, and TOU

Demand charges began over a century ago. A significant debate ensued over the better rate form, and demand charges emerged as the preferred alternative, due entirely to the simplicity of demand metering. At that time, the only way to measure TOU energy consumption was with chart recorders that were manually interpreted. Some very large (multi-megawatt) industrial customers were fitted with such systems, but until the emergence of electronic metering, TOU measurement remained relatively difficult.

The early commercial rate forms were the Wright rate (demand charge plus energy charge), and the Hopkinson rate (multiple load factor blocks). Some incorporated both features into rates. An example of a Wright-Hopkinson rate is that for DTE (Detroit Edison):

Figure A-1. Sample Wright-Hopkinson Rate

Power Supply Charges:	
Demand Charge:	$***13.88*** per kW applied to the Monthly Billing Demand
Energy Charges:	***4.7806¢*** per kWh for the first 200 kWh per kW of billing demand
	3.7806¢ per kWh for the excess

First, an NCP demand charge recovers most capacity costs. Second, the balance of capacity costs is embedded in the first 200 kWh/kW of energy charges. Higher-load-factor customers enjoy the end-block rate after capacity costs are recovered.

Decades ago, when all forms of capacity had similar costs, and detailed metering was expensive, this rate form may have been reasonable. In an evolved industry, where advanced metering is available and “capacity” needs are met with a mix of storage, demand response, dispatchable generation, inflexible baseload generation, and intermittent renewable resources, it no longer makes sense.

Rate Design References

- James Bonbright, *Principles of Public Utility Rates*, Columbia University Press, 1961.
- Paul Garfield and Wallace Lovejoy, *Public Utility Economics*, Prentice Hall, 1963.
- Alfred Kahn, *The Economics of Regulation*, MIT Press, 1988.
- Frederick Weston, “Charging for Distribution Utility Services: Issues in Rate Design,” Regulatory Assistance Project, 2000, http://www.raponline.org/knowledge-center/charging-for-distribution-utility-services-issues-in-rate-design/.

- Carl Linvill, John Shenot, and Jim Lazar, "Designing Distributed Generation Tariffs Well," Regulatory Assistance Project, 2014, http://www.raponline.org/knowledge-center/designing-distributed-generation-tariffs-well.
- Jim Lazar and Wilson Gonzalez, "Smart Rate Design for a Smart Future," Regulatory Assistance Project, July 2015, http://www.raponline.org/knowledge-center/smart-rate-design-for-a-smart-future.
- Jim Lazar and Ryan Hledik, "Distribution System Pricing With Distributed Energy Resources," Future of Electric Utility Regulation series, National Renewable Energy Laboratory, May 2016, http://www.raponline.org/knowledge-center/distribution-system-pricing-with-distributed-energy-resources.
- "Distributed Energy Resources Rate Design and Compensation," NARUC Staff Subcommittee on Rate Design, November 2016, https://pubs.naruc.org/pub/19FDF48B-AA57-5160-DBA1-BE2E9C2F7EA0.

Appendix B: Traditional Cost of Service Methods and Their Application to Rate Design

Different states use different methods to apportion costs between classes, and to design rates within classes. This is a short summary of the types of approaches RAP experts have seen in our long experience.

Embedded Cost Approaches

Most states use fully allocated cost of service studies for apportioning rates between classes. Some use the results of these studies to design rates within classes as well. But embedded cost studies come in an infinite variety. All of these divide accounting costs, and the results exactly equal the utility revenue requirements, but produce very different results by class and by the cost drivers to retail rate elements. We have grouped them into just a few categories.

Peak-Responsibility Methods: Fixed production and transmission costs are classified as demand-related, and allocated on some measure of peak demand: 1 CP, 4 CP, and 12 monthly CP are common methods.

Energy-Weighted Methods: Fixed production and transmission costs are classified as partly demand-related, based on the cost of peaking resources, and the balance as energy-related. The demand-related costs are allocated on some measure of peak demand.

Minimum-System Methods: Distribution plant is classified as customer related based on the hypothetical cost of a minimum-sized distribution (or zero intercept) calculation, with the balance classified as demand-related.

Basic-Customer Methods: Distribution plant is divided between customer-specific costs (service drops and meters, typically), and joint costs (poles, wires, and transformers). The joint costs are apportioned on some measure of usage: CP, NCP, and energy allocators are variously used.

Marginal Cost Approaches

A few states use marginal cost studies for electric and gas cost allocation. As with embedded cost studies, however, there is a wide variety of methodologies that have evolved.

Short-Run Marginal Cost: The cost of supplying additional customer requirements using existing facilities. Only costs that vary in the short run (fuel, purchased power, and line losses) are considered. The result is typically much lower than the revenue requirement. This approach is most often used to set economic development and other incentive rates.

NERA-Methodology: Costs that vary within an intermediate planning horizon, such as 10-years, are considered. This will typically include some peaking generation, some transmission, and some distribution capital costs, plus variable costs, but not new baseload generation or remote long-distance transmission. If the utility system is in equilibrium (no excess or deficiency of generation, transmission, or distribution capacity), this method produces a similar result to TSLRIC, below. But, where utilities have temporary excess generation capacity, as is common, this method typically produces a marginal cost for production that is somewhat lower than the production revenue requirement, and is therefore favorable to large-user classes (for whom production costs are a larger share of the cost responsibility). California has used a variation of this approach for many years.

Total-System Long-Run Incremental Cost (TSLRIC): The cost of building an optimized system for the current complement of customers and loads is measured. This has been widely used in telecom, but less often for electricity and gas. It is the theoretically appropriate metric for determining if competitive suppliers are viable. Because existing facilities are typically on utility books at far less than replacement cost, this approach generally produces a marginal cost that is somewhat greater than the revenue requirement.

RAP® Energy Solutions for a Changing World

The Regulatory Assistance Project (RAP)®
Belgium · China · Germany · India · United States

50 State Street, Suite 3
Montpelier, Vermont 05602
USA

1 802-223-8199
info@raponline.org
raponline.org