decisions for storage and dispatchable hydro resources are typically made to maximize the benefits from the limited supply of other time-shiftable generation resources.

Prior to PURPA, most long-term purchased power contracts had separate capacity and energy elements. These were mostly for fuel-dependent power plants. This rate form allowed the owner to obtain capital cost recovery in a predictable payment and the receiving utility to control the output as needed to fit varying loads, paying for short-run variable costs as incurred. Today many power purchase contracts are expressed entirely on a volumetric basis, based on an expected pattern of output. This change in how contracts are priced in the wholesale market does not dictate any particular approach to how costs are allocated in the retail rate-setting process.

Generation Capacity Costs

Beyond these energy needs, most regions of the United States also plan around the amount of shared generation capacity needed, and these processes can drive a significant amount of generation costs. The amount of capacity required by a utility system, typically denominated in **megawatts** (MWs) or gigawatts at the time of the system coincident peak, determines whether the utility should retire existing plants, add new resources or delay planned retirements, or keep the system as it is. All those decisions have costs and benefits. This determination may be made by an ISO/RTO, a holding company or other aggregation of interconnected load.

Although the typical planning procedures used to date by utilities and ISOs have often served their original purposes to measure the least-cost resources available at the utility system level, these procedures often oversimplify important aspects of overall capacity and reliability issues. The key principle is that reliability-related costs are not all "caused" by one hour or a few hours of demand during the year. A system must have some form and level of capacity available at all hours. Loss-of-energy expectation[27] studies generally show that adding capacity at any hour to a system, even **off-peak** hours, has a small but discernible beneficial impact on reliability. Many resources can be justified only if all of the attributes are considered, including contribution to meeting peak demand and contribution to meeting other needs such as fuel cost reduction.

The typical vertically integrated utility calculates the installed capacity requirement by determining what amount of existing and new capacity will provide acceptable reliability, measured by such statistical parameters as the mathematical expected value of the number of hours in which it cannot serve load or of the amount of customer energy it will not be able to serve in a year, due to insufficient available generation. Those expected values are computed from models that simulate the scheduling of generation maintenance and the random timing of forced outages for many potential combinations of outages and load levels. In large portions of North America, the capacity requirement is determined regionally by an ISO/RTO and then allocated to the load-serving entities, transmission control areas or utilities.[28]

Required reserves are usually expressed as the percentage **reserve margin**, which is:

$$(\text{capacity} - \text{peak load}) \div \text{peak load; or}$$
$$(\text{capacity} \div \text{peak load}) - 1$$

Capacity may be defined as installed capacity, demonstrated capacity or unforced capacity (installed capacity reduced by the resource's forced outage rate). There may be special provisions to recognize that an installed MW of solar, wind or seasonal hydro capacity is not equivalent to an installed MW of combustion turbine capacity with guaranteed fuel availability or a MW of battery storage capacity located at a distribution substation. Capacity requirements may also be satisfied with curtailable load, energy storage or expected price response to peak pricing. The cost of capacity to meet a very short-term need is very different from the cost of **baseload capacity** that serves customers around the clock

27 Different analysts refer to related measures as loss-of-load hours, loss-of-load expectation, expected unserved energy and loss-of-load probability.

28 Some of the utilities in the ISOs/RTOs are restructured and do not provide generation services, so the cost of service study need not deal with generation costs. However, all the utilities in the SPP and most of those in MISO are vertically integrated, as are some jurisdictions in PJM (West Virginia, Virginia, Kentucky and the PJM pieces of North Carolina, Indiana and Michigan) and ISO-NE (Vermont) and municipal and cooperative utilities in most restructured jurisdictions.

and throughout the year, and the cost analyst must be aware of these differences.

Peak load is generally the utility's maximum hourly output requirement under the worst weather conditions expected in the average year (e.g., the coldest winter day for winter-peaking utilities or the hottest summer day for summer-peaking utilities). In the ISOs/RTOs, the peak load is usually the utility's contribution to the actual or expected ISO/RTO peak load. Although the reserve margin is often stated on the basis of a single peak hour as a matter of measurement convention, the derivation of the reserve margin takes into account far more information than the load in that one hour. The most important parameters in determining the required reserve margin are the following:

- **Load shape**, especially the relationships among the annual and weekly peaks and the number of other hours with loads close to the peaks. The system must have enough reserve capacity to endure generation outages at the high-load hours. The near-peak hours matter because the probability of any given combination of outages coinciding with the peak hour is very low, but if there are hundreds of hours in which that combination of outages would result in a supply shortage, the probability of loss of load would be much larger.
- Maintenance requirements. Utilities attempt to schedule generator maintenance in periods with loads lower than the peak, typically in the autumn and spring, and occasionally in the winter for strongly summer-peaking utilities and in the summer for strongly winter-peaking utilities. Utilities with both modest maintenance requirements and several months with loads reliably well below those in the peak months can schedule all routine maintenance in the off-peak months while leaving enough active capacity to avoid any significant risk of a capacity shortage in those months. But many utilities have large maintenance requirements (especially for coal-fired and nuclear units) and only modest reductions in peak exposure in the shoulder months. After subtracting required maintenance, the effective reserve margin may be very similar throughout the year, increasing the chance that a combination of outages will result in loss of load. As a result, high loads in any month (or perhaps any week) contribute to the need for installed capacity.
- Forced outage rates. All generation units experience some mechanical failures. The higher the frequency of forced outages, the more likely it is that a relatively high-load hour will coincide with outages, eliminating available reserve and resulting in the loss of load.
- Unit sizes. If all of a system's units were very small (say, under 1% of system peak), the random outages could be expected to spread quite evenly through the year. With larger units, outages are much lumpier, and loss of a small number of large units can create operating problems. Hence, systems with larger units tend to need higher reserve margins, all else being equal.
- Other operating constraints. Although hydro resources have the highest overall reliability, they produce power only when water is available to run them. Some hydro resources are required to be operated for flood control, navigation, irrigation, recreation, wildlife or other purposes, and these other constraints may affect the ability of the resource to provide power at full capacity when system peak loads occur.

Some of the factors in this list affect the reliability value of various types of generation, while others highlight the types of load that increase required capacity reserve levels. A large unit with frequent forced outages may contribute little to ongoing system reliability even though it has a significant nameplate capacity. If such a unit has high ongoing costs that could be reduced or eliminated through retirement, continued operation must primarily be justified by its energy benefits. On the demand side, long daily periods of high loads can mean that many weekday hours (and even some weekend hours) in each month will contribute to capacity requirements, proportionately shifting capacity responsibility toward customers with high **load factors**. Table 2 on the next page summarizes cost drivers for power supply capacity.

The value of capacity is partly a function of the type of capacity and the location of that capacity. Although required capacity (measured in MWs) is determined by demand in a subset of hours, along with the characteristics of the power plants, the cost of capacity (measured in dollars per MW-year) is in large part determined by energy requirements.

In the previous millennium, the cheapest form of

Table 2. Cost drivers for power supply

Resource type	Purpose	Investment-related costs	Maintenance costs	Fuel costs
Baseload nuclear, geothermal	Power at all hours	High	High	Low
Coal, intermediate combined cycle	Power at many hours	Medium	Medium	Medium
Peaking	Power in peak hours, plus reserves at all hours	Low	Low	High
Hydro	Power at some or all hours	Very high	Low	Low or none
Wind	Power at some hours	High	Low	None
Solar	Power at some hours	High	Low	None
Storage	Power at peak hours, plus reserves at all hours	High	Low	Low — for purchased kWhs

capacity to serve peak needs was typically considered to be a combustion turbine. These units had low investment costs and low ongoing O&M expenses but were inefficient and typically used more expensive fuels. These characteristics made them perfect to run infrequently during peak times and for other short-term reliability needs. Conversely, it made sense to make major investments in units with high upfront costs but high efficiency and cheap fuel prices and to run these units nearly year-round. These major investments were driven by year-round energy requirements, not peak loads.

Today, in contrast, the least expensive form of capacity to serve extreme peak loads may not be a generating unit at all. For very low-duration loads, demand response, customer response to critical peak pricing or battery storage may be the least-cost resource to serve a very short-duration peak, sometimes described as a needle peak. The ability to curtail an end-use load saves not only the amount of capacity represented by the reduced load but also the marginal line losses and reserves that would be required to reliably sustain that load. Similarly, the ability to dispatch DERs also avoids line losses that would be required to deliver generated capacity to that location.[29]

5.1.2 Transmission

The costs of transmission lines depend on the length of the lines, the terrain they must cover and the amount of power they need to carry at different times, sometimes in either direction. The maximum usage of many transmission lines is not necessarily at system peak hours, and the usage of certain lines can change significantly over time. Carrying more power requires larger conductors, multiple conductors and/or higher voltages, all of which increase costs.

If each load center in a utility's territory had about the amount of generation required to meet its peak load, and the power plants were similar so the utility had no interest in exporting power from one area to another, the transmission system would exist primarily to allow each load center to draw on the others for backup supply when local generation was unavailable. In real utility systems, power plants are often distributed very differently from load, with large centralized plants built to capture economies of scale, often in areas far from major load centers. Generation may be sited remotely away from load for environmental reasons, to facilitate access to fuel and to minimize land costs and land use conflict. Generation plants also tend to vary considerably in fuel cost, efficiency and flexibility; allowing the utility to use the least-cost mix of generation at all load levels may require additional transmission.

By contrast, demand response, energy efficiency and energy storage can be very carefully targeted geographically to provide needed capacity in a specific area without the need for any additional transmission.

Although separating all the causes of the structure of an existing transmission system can be difficult, especially for a

29 The capacity saved can be as high as 1.4 times the load reduced, when marginal line losses and reserves are taken into account. For a detailed discussion of this, see Lazar and Baldwin (2011).

utility whose distribution of load and generation has changed over the decades, decisions about the nature and location of generation facilities can have important effects on the costs of the transmission system.

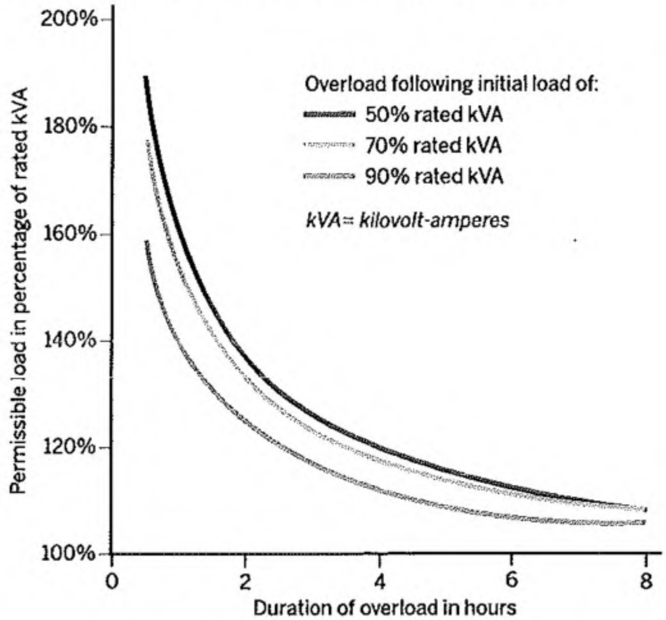
Energy load over the course of many hours also affects the sizing and cost of transmission. Underground transmission is particularly sensitive to the buildup of heat around the lines, so the duration of peak loads and the extent to which loads decline from the peak period to the off-peak period affects the sizing of underground lines. An underground line may be able to carry twice as much load for a 15-minute peak after a day of low loads as for an eight-hour peak with a high daily load factor. To reduce losses and the buildup of heat from frequent high loads, utilities must install larger cables, or more cables, than they would to meet shorter duration loads.

The capacity of overhead lines is often limited by the sagging caused by thermal expansion of the conductors, which also occurs more readily with summer peak conditions of high air temperatures, light winds and strong sunlight. Overheating and sagging also reduce the operating life of the conductors. A transmission facility normally will have a higher capacity rating for winter than for summer because the heat buildup is ameliorated in cooler weather.

The costs of substations, including the power transformers on which they are centered, are determined by both peak loads and energy use. The capacity of a station transformer is limited by the buildup of heat created by electric energy losses in the equipment. Every time a transformer approaches or exceeds its rated capacity (a common occurrence, since transformers can typically operate well above their rated capacity for short periods), its internal insulation deteriorates and it loses a portion of its useful life.

Figure 16 illustrates the effect of the length of the peak load, and the load in preceding hours, on the load that a transformer can carry without losing operating life (Bureau of Reclamation, 1991, p. 14). The initial load in Figure 16 is defined as the maximum of the average load in the preceding

Figure 16. Permissible overload for varying periods



Source: Bureau of Reclamation. (1991). *Permissible Loading of Oil-Immersed Transformers and Regulators*

two hours or 24 hours.[30] A transformer that was loaded to 50% of its rating in the afternoon can endure an overload of 190% for 30 minutes or 160% for an hour. If the afternoon load was 90% of the transformer rating, it could carry only 160% of its rated load for 30 minutes or 140% for an hour.[31]

Similarly, if the transformer's high-load period is currently eight hours in the afternoon and evening, and the preceding load is 50% of rated capacity, afternoon load reductions that cut the high-load period to three hours would increase the permissible load from about 108% of rated capacity to about 127%. Under these circumstances, the transformer can meet higher load without replacement or addition of new transformers.

Short peaks and low off-peak loads allow the transformer to cool between peaks, so it can tolerate a higher peak current. Long overloads and higher load levels increase the rate of aging per overload, and frequent overloads lead to rapid failure of the transformer.

30 This specific example is for self-cooled and water-cooled transformers designed for a 55 degrees Celsius temperature rise; other designs show similar patterns.

31 Utilities recognize that the length of overloads is critical to determining whether a transformer needs to be replaced. For example, Potomac Electric Power Co. (Pepco) in Maryland has established standards for replacing line transformers when the estimated average load over a five-hour period exceeds 160% of the rating of overhead transformers or 100% for pad-mounted transformers (Lefkowitz, 2016, p. 41). The company has not found it necessary to establish comparable policies for shorter periods.

Table 3. Cost drivers for transmission

Connection to (or between)	Purpose	Typical length of line	Investment-related costs	Maintenance costs
Remote baseload generation	Power at all hours	Long	High	Low
Remote wind or solar	Power at some hours	Long	High	Low
Peaking resources	Power in peak hours, plus reserves at all hours	Short	Low	Low
Hydro	Power at some or all hours	Long	High	Low
Neighbor utilities	Reserve sharing; energy trading	Short to long	Vary	Low
Substations networked for reliability	Power at some hours	Short	Medium	Low
Storage and substations	Power at peak hours, plus reserves at all hours	Very short	Very low	Low

In a low load factor system, these high loads will occur less frequently, and the heavy loading will not last as long. If the only high-demand hours were the 12 monthly peak hours, for example, most transformers would be retired for other reasons before they experienced significant damage from overloads. In this situation, larger losses of service life per overload would be acceptable, and the short peak would allow greater overloads for the same loss of service life.

With high load factors, there are many hours of the year when the transformers are at or near full loads. In this case, the transformer must be sized to limit overloads to acceptable levels and frequency of occurrence commensurate with a reasonable projected lifespan for the asset. If the transformer is often near full capacity with frequent overloads, it will fail more rapidly.

Transmission lines serve many purposes, including connecting remote generating plant to urban centers and enabling the optimal economic interchange of power between regions with different load patterns and generation options. Each transmission segment can be separately examined and allocated on a cost-reflective basis. Table 3 provides examples of this.

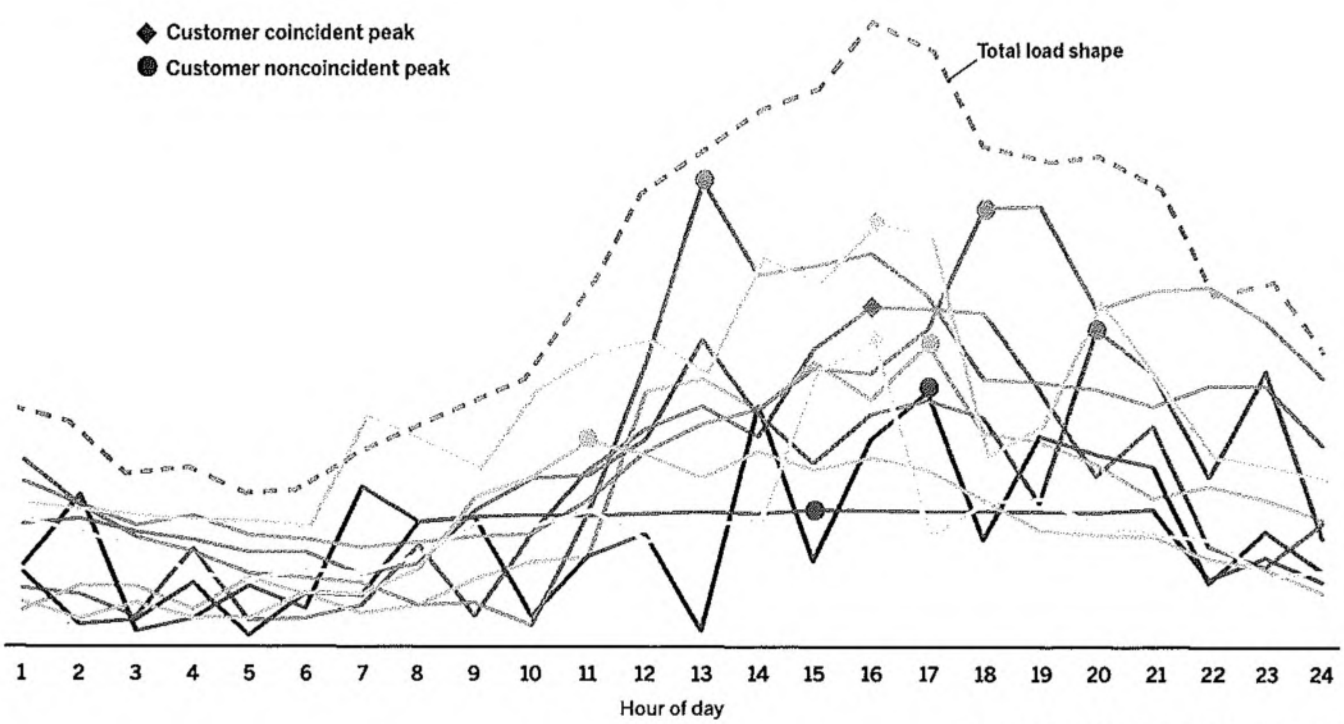
5.1.3 Distribution

The factors driving load-related distribution costs are similar to those for transmission. Different components are built and sized for different reasons; some serve the shared needs of hundreds or thousands of customers, while other components are designed to serve a single customer. Substations and line transformers must be larger — or will wear out more rapidly — if they experience many high-load hours in the year and if daily load factors are high. Underground and overhead feeders are also subject to the effects of heat buildup from long hours of relatively high use.

The allowable load on distribution lines is determined by both thermal limits and allowable voltage drop. Higher loads on a primary feeder may require upgrades (raising the feeder voltage, adding a new feeder, reconductoring to a larger wire size, increasing supply from single-phase to three-phase) to maintain acceptable voltage at the end of the feeder. Small secondary customers can be farther from the line transformers than large customers (allowing the utility to use fewer transformers to serve the same load) and can be served with smaller conductors.

As with station transformers, line transformers can handle moderate overloads for relatively short periods of a few hours but will deteriorate quickly if subjected to extended overload conditions. Therefore, the sizing of transformers takes into consideration not only the maximum capacity required but also the underlying load shape. Figure 17 on the next page shows actual data from a confidential load research sample on a summer peak day for 10 residential customers who share a line transformer. Although no group of 10 customers is identical to any other group of 10 customers, this demonstrates how diversity determines the need for the sizing of system elements. Only three of the 10 customers peak at the

Figure 17. Summer peak day load from 10 residential customers on one line transformer



Source: Confidential load research sample

same time as the 4 p.m. coincident peak for the group, and the coincident peak is only 86% of the sum of the individual peaks on this day. Furthermore, although not shown in this figure, this coincident peak is only 64% of the sum of the annual noncoincident peaks for the individual customers. It is important to note that a group of 10 residential customers is often less diverse than the combined loads from multiple customer classes, which determine the need for substation and generation capacity upstream of the final line transformer.

It is important to note that the load exceeds 50 kVA for only three hours and is below 40 kVA for 18 hours of this summer peak day. Referring back to Figure 16, under these circumstances, a 50-kVA transformer would likely be adequate to serve this load, because the overload is for only a short period. By contrast, the sum of the maximum noncoincident peak loads of the 10 customers is more than 90 kVA.

A large portion of the distribution investment is driven primarily by the need to serve a geographical region. Once a decision is made to build a circuit, the **incremental cost** of connecting additional customers consists mostly of additional line transformers (if the new customer is isolated from others) and secondary distribution lines. This is true even if those investments may serve multiple customers, particularly in urban and suburban areas. These shared facilities are largely justified by the total revenues of the customers served, not the peak load or number of customers. A particular transmission line, substation or feeder to serve an area could be justified by a single very large load, a small number of large customers or a large number of very small customers.

Nearly every electric utility has a line extension policy that sets forth the division of costs incurred to extend service to new customers. Typically, this policy provides for a certain amount of investment by the utility, with any additional investment paid for by the new customers. These provisions are intended to ensure that new customers pay the incremental cost of connecting them to the system without raising rates to other customers. For most utilities, there is no corresponding credit where new service has a cost that is lower than the

Table 4. Cost drivers for distribution

Type	Purpose	Investment-related costs	Maintenance costs
Substations	Power at all hours; capacity for high-load hours	High	Low
Primary circuits	Power at all hours; capacity for high-load hours	High	Low
Line transformers	Power at all hours; capacity for localized high-load hours	Medium	Low
Secondary service lines	Power at all hours; capacity for localized high-load hours	Medium	Low
Meters: Traditional	Measuring usage	Low	Low
Meters: Advanced	Multiple functions	Medium	Low

average embedded cost of service, a circumstance that results in benefits to the utility and other ratepayers.

The final components in the distribution system are meters, typically installed for all residential and general service customers but not for very predictable loads like traffic signals or streetlights. How to classify the cost is a matter of debate. On one hand, a meter is needed because usage levels vary from customer to customer and month to month, a theoretically usage-related cost. But on the other hand, one meter is needed for every metered customer, and meter costs do not typically vary from customer to customer within a class. In addition, **smart meters** entail both higher direct investment costs and back office investments but provide generation, transmission and distribution system benefits by allowing more precise measurement and control of local loads and more accurate assignment of peaking capacity requirements. Lastly, the cost of current transformers and potential transformers necessary to meter large customers should be included as part of their metering costs — an issue common between embedded and marginal cost methods.[32] Table 4 summarizes cost drivers in the distribution system.

5.1.4 Incremental and Complementary Investments

Good economic analysis should distinguish properly between complementary or alternative investments, which substitute for one another, and incremental investments, which add costs to the system.

Customers receive service at different voltages and with different types of equipment. Most of the distinctions among types of equipment represent alternative or complementary methods for providing the same service. For example, various primary distribution feeders operate at 4 kV, 13 kV or 25 kV and may be overhead or underground construction, depending on load density, age of the equipment, local governmental requirements and other considerations. Although the power flowing from generation to a customer served at 25 kV may not flow over any 4-kV feeder, the 4-kV feeders serve the same function as the 25-kV feeders and (in places in which they are adequate) at lower cost.[33] Serving some customers at 4 kV and spreading the feeder costs among all distribution customers does not increase costs allocated to the customers served directly from the 25-kV feeders; converting the 4-kV feeders to a higher voltage would likely increase costs to all distribution customers, including those now served at 25 kV. In this situation, all the feeders should be treated as serving a single function, and all their costs should be allocated in the same manner.

Similarly, most customers served by single-phase primary distribution are served with that configuration because it is cheaper than extending three-phase primary distribution, which they do not require because of the nature of their loads.

32 Current transformers reduce the amperage so a meter can read it. Potential transformers reduce the voltage for meter reading (Flex-Core, n.d.).

33 Conversely, the 4-kV supply to some customers is from transformers fed directly from transmission without using the 25-kV system.

On the other hand, some distinctions in voltage level represent incremental investment:

- Most customers served at distribution voltages cannot take service directly from the transmission system. Even if a transmission line runs right past a supermarket or housing development, the utility must run a feeder from a distribution substation to serve those customers. Distribution in its broadest sense is thus principally an incremental service, rather than an alternative to transmission, needed by and provided to some customers but not all.[34]
- Similarly, most customers who take service at secondary voltage have a primary line running by or to their premises yet cannot take service directly at primary voltage.[35] The line transformers are incremental equipment that would not be necessary if the customers could take service at primary voltage.[36]

These incremental costs should be functionalized so that they are allocated to the loads that cause them to be incurred, while each group of complementary costs (such as various distribution voltages) generally should be treated as a single function and recovered from all customers who use any of the alternative facilities.

In other situations, distinguishing between incremental and complementary costs can be more complicated. Examples include the treatment of transmission equipment at different voltages and the treatment of secondary poles. Many embedded cost of service studies treat subtransmission as an incremental cost separate from transmission and charge more for delivery to customer classes served directly from the subtransmission system or from substations fed by the subtransmission system. For the most part, utilities use lower transmission voltage where it is less expensive than higher voltages, either due to the lower cost of construction relative to the total load that needs to be served by the line or the happenstance that the subtransmission line is already in place. If it is less expensive to serve customers with the lower voltage, it would be inequitable to charge them more for being served at that voltage.

Similarly, distribution poles carrying only secondary lines are less expensive than poles carrying primary lines. If a customer served by a secondary-only pole had to be served at primary voltage instead, the primary pole would be more expensive, and that higher cost would almost certainly be allocated to all distribution customers. Secondary poles (unlike line transformers and most secondary lines) are lower-cost alternatives to some primary poles.[37]

5.2 Determining Customer Classes

In addition to administrative simplicity, the purpose of separating customers into broad classes flows from the idea that different types of customers are responsible for different types of costs, and thus it is fairer and more efficient to charge them separate rates. One set of rates for each customer class, based on separate cost characteristics, is the key feature of postage stamp pricing for electric utilities. As a result, it is very important to determine appropriate customer classes with different cost characteristics at the outset of a cost of service study. The number of classes will vary from utility to utility and may vary depending on the costing methodology being used. In addition to equitable cost allocation, different rate structures are often used for different rate classes. For example, residential customer classes generally do not have demand charges today, but most large industrial classes do. This means that decisions regarding the number and type of customer classes can also have rate design implications,

34 In some cases, a distribution substation and feeder can bring service to customers that would otherwise be served by an extension of the transmission system at higher cost. Identifying and accounting for that limited complementary service is probably not warranted in most embedded cost of service study applications.

35 Another way of looking at this relationship is that secondary customers are those for whom providing service at secondary has a lower total cost than providing service at primary. Sharing utility-owned transformer capacity is less expensive than having each customer build its own transformer. See Chapter 11 for a discussion of primary and secondary distribution and their allocation.

36 Although most networked secondary conductors parallel primary lines and are incremental to the primary system, a limited number of secondary conductors extending beyond the primary lines are complementary, because they avoid the need to extend primary lines.

37 Similarly, a portion of the secondary lines replaces primary lines. If the customers that can be served with secondary poles required primary service, the utility would need to extend the primary lines rather than secondary lines. Hence, a portion of the secondary lines is also complementary to the primary system, rather than additive.

although this is not necessarily permanent.

Most utilities distinguish among residential customers, small commercial customers, large commercial customers, industrial customers and street lighting customers. The commercial and industrial classes often are collectively termed general service rate classes. In many cases, general service customers are categorized by voltage levels. Customers served at primary distribution voltage generally do not use, and should not be allocated, costs of secondary distribution facilities, and customers served at transmission voltage generally do not use, and should not be allocated, costs of distribution facilities. Many utilities also separate general service classes with even greater granularity than using simple voltage criteria.

One area where utility practices can vary significantly is whether there is more than one residential class or, alternatively, multiple residential subclasses. Some utilities separate out residential customers based on a measure of size, such as peak demand or energy use. This can be significant in jurisdictions that categorize farms or large master-metered multifamily buildings as residential in a formal sense. Some jurisdictions also create separate classes based on the usage of specific technologies like electric resistance heating. In some jurisdictions, low-income discount customers are treated as a separate rate class.

The creation of multiple residential classes or subclasses is typically justified on cost grounds. There are inarguably many cost distinctions among different types of residential customers, and simple postage stamp cost allocation and rate structures may not capture many of those distinctions. Regulators and utilities have long analyzed the causes of such differences, which vary widely across the country. Some of the distinctions are based on technology (or, more accurately, as a proxy for the load impacts of certain technologies), such as electric space heating, electric water heating, solar or other distributed generation and even electric vehicles. Other distinctions are based on the characteristics of service. Those with relatively large impacts on cost allocation include:

- Single family versus multifamily.
- Urban (multiple customers per transformer) versus rural (one customer per transformer).
- Overhead service versus underground service.

A word of caution is appropriate here. With respect to technology-driven class characteristics such as electric space heat, water heat, vehicles or solar installations, singling out customers based on technology adoption has serious practical and theoretical downsides. Furthermore, addressing one minor cost distinction is likely not fair or efficient if several other major cost distinctions, such as those listed above, are not addressed. It is wiser to consider multiple customer and service characteristics simultaneously to create technology-neutral subclasses for both cost allocation and rate design purposes.

To begin, electric space heating customers are likely to have different load characteristics from the nonheating customers, with significantly more usage and a different daily load shape in the winter. For a winter-peaking system, this could mean that electric heating customers should be allocated proportionately more costs. Conversely, in a summer-peaking system, electric heating customers should be allocated proportionately fewer overall costs. However, this issue, which is essentially a question of a potential intraclass cross-subsidy between types of residential customers, can also be addressed through changes to rate design. Seasonally differentiated rates, if based appropriately on cost causation, can achieve the same distributional impact as separate rate classes for heating and nonheating customers while bringing additional benefits from the improved efficiency of pricing.

The creation of an electric heating rate class can have other implications. In regions where electric heating customers are disproportionately low-income, this decision also has significant equity implications. There can also be environmental repercussions to this choice. Concerns would arise, for example, if electric heating rates promote use of gas and coal in power plants to replace direct burning of gas on-site for heating, which historically was often more efficient on a total energy basis. Recent developments in efficient electric heating, particularly air and ground source heat pumps, may have switched the valence of these questions. In certain areas, higher-income customers may be disproportionately adopting efficient electric heating. And the new electric technologies may now be significantly cleaner and more efficient than on-site combustion of natural gas, particularly if powered by zero emissions electric resources. A seasonal and time-varying cost study and time-varying rates may enable appropriate cost recovery without need for a separate class.

Several states have considered creating a separate rate class for customers with solar PV systems. Because solar customers may have different usage patterns than other customers, this is reasonable to investigate. However, it is not clear that there is a significant cross-subsidy to address, particularly at low levels of PV adoption. Current rate design practices for solar customers in many jurisdictions — such as net metering using flat volumetric rates, monthly netting and crediting at the retail rate — are fairly simple. These rate design practices could be improved significantly over time and integrated with broader rate design reforms. For example, a time-varying cost study would allow the creation of more granular time-varying rates so that solar customers pay an appropriate price for power received during nonsolar hours and are credited with an appropriate price for power delivered to the distribution system during solar hours. This would include changes to netting periods, which would reveal more information about how a solar customer actually uses the electric system.

In terms of rate classes for specific technologies, some utilities separate out customers with electric water heating as a proxy for a flat load shape and the potential for load control. In the future, some utilities may seek to make electric vehicle adoption a separate rate class as a substantially controllable load with distinct usage characteristics. However, these technologies may not need consideration as a separate rate class, particularly given efforts to improve the cost causation basis of rate design more generally. Again, time-varying rates will appropriately charge customers with peak-oriented loads and appropriately benefit customers with loads concentrated in low-cost hours or controlled into those hours.

Some utilities have implemented separate rate classes for single-family and multifamily residential customers. There are many reasons to believe that the cost of serving multifamily buildings is substantially lower than serving single-family homes on average:

- Shared service drops.
- Increased diversity of load for line transformers and secondary distribution lines, enabling more efficient sizing.
- Reduced cost of distribution per customer, since no distribution lines are required between customers in the building.[38]
- Reduced coincidence with both summer and winter peak loads because common walls reduce space conditioning use relative to single-family units of the same square footage, and because lighting and baseload appliances such as refrigerators and water heaters (if electric) are a larger percentage of loads for units with fewer square feet.
- Reduced need for secondary distribution lines in cases where the multifamily building can be served directly from the transformer.
- Reduced summer peak coincidence if space cooling is provided through a separate commercial account for the building, rather than as part of the individual residential accounts.
- Reduced costs of manual meter reading, where still applicable.

There may be countervailing considerations in some service territories, such as if multifamily buildings are served by more expensive underground service and single-family buildings are served with cheaper overhead lines. A similar set of considerations may cause some utilities to disaggregate customers by geography, such as those residing inside and outside city limits.[39] Customers in deeply rural areas tend to be more expensive to serve, since they typically are too far from their neighbors to share transformers, require a long run of primary line along the public way, and generally

38 This distinction is important where some distribution costs are classified as customer-related. In those situations, each multifamily building (rather than each meter) should be treated as one customer, as would a single commercial customer of the same size and load.

39 For example, Seattle City Light, a municipal utility, has two rate schedules for most commercial and industrial classes within the city: one for the highly networked higher-cost underground system in the urban core, and another for the balance of the city, plus separate higher rates for the adjacent cities and towns where it provides service. Compare Schedules MDC, MDD, MDS and MDT at Seattle City Light (n.d.). The city of Austin, Texas, also applies different rates to customers outside the city limits (Austin Energy, 2017). In many places, cities impose franchise fees or municipal taxes that make customer bills inside cities higher than those outside cities, even though the cost data may suggest the opposite is more equitable.

have higher unit costs related to lower load per mile of distribution line.[40]

Analysts may want to employ a simple standard for deciding when to divide a subclass for analytical purposes, based on whether the groups are large enough and distinct enough to form a separate class or subclass. One such guideline might be that, if more than 5% of customers or 5% of sales within a class have distinct cost characteristics, differentiation is worth considering. If fewer than that, although the per-customer cost shifts may be significant, the overall impact on other customers will likely be immaterial. If 2% of the load in a class is paying 20% too much or too little, for example, other customers' bills will change only 0.4%. But if 15% of the load is 20% more or less expensive, the impact on other users rises to 3%. The trajectory of these impacts over time can also be relevant.

Although improved distributional equity from additional rate classes is a laudable goal, and indeed advances the primary goal of cost allocation, there are countervailing considerations that may dictate keeping the number of rate classes on the smaller side. First, there are administrative and substantive concerns around adding rate classes, both in litigation at state regulatory commissions and in real-world implementation. Some potential distinctions among customers may be difficult to implement because they involve subjective and potentially controversial determinations by on-the-ground utility personnel. In creating new distinctions, regulators, utilities and stakeholders must all have confidence that there are true cost differentials between the customer types and that there will be little controversy in the application of the differentials. Some analysts object to customer classes based on adoption of particular end uses, although this may serve as a proxy for significantly different usage profiles. Furthermore, some utilities and parties in a rate case may propose rate classes that effectively allow undue discrimination. If the proper data aren't available to scrutinize such claims, either publicly or for parties in a rate case, then this may allow an end-run around one of the significant motivations for postage stamp pricing: preventing price discrimination.

Lastly, as described above for electric heating and solar PV customers, rate design changes can also address certain cross-subsidies within customer classes in a relatively straightforward manner that also provides additional efficiency benefits. In principle, perfectly designed time- and location-varying pricing for all electric system components and externalities, applied identically to all customers, could eliminate the need for customer classes and cost allocation entirely while providing perfectly efficient price signals. This is unlikely to be the case for the foreseeable future but illustrates the conceptual point that an efficient improvement to rate design may be a strictly preferred option compared with the creation of a new rate class. For example, certain types of customers could be put on technology-neutral time-varying rates on an opt-out or mandatory basis, such as customers with storage, electric vehicles or distributed generation.

5.3 Load Research and Data Collection

Any cost of service study, as well as rate design, load forecasting, system planning and other utility functions, depends heavily on load research data. Cost allocation, in particular, requires reasonably accurate estimates for each class or group distinguished in the analysis, the number of customers, their energy usage (annual, monthly and sometimes more granular time periods), their kW demand at various times and under various conditions, and sometimes more technical measures such as **power factor**. The key principle is that there is diversity among customers in each class, meaning the consumption characteristics for the group are less erratic than those of any individual customer. Load research is the process of estimating that diversity.

At the very least, these data must be available by class across the entire system. For some applications, these data are useful and even essential at a more granular level, such as for each substation, feeder or even customer. Ideally, the cost of service study would be able to draw on information about the hourly energy usage by class, as well as the contribution of each class to the sum of the customer contributions to the maximum loads across the line transformers serving the

40 These factors may be offset by the utility's policy for charging new customers for extending the distribution system, as discussed in Section 11.2

class, the feeders serving the class, the substations serving the class and so on. Modern AMI and advanced distribution monitoring systems, if properly configured, can provide those data. Some utilities now routinely collect interval load data at each level of the system, while others are starting to acquire those capabilities.

The data needed for different cost allocation frameworks and methods can vary greatly, and it is difficult to generalize because of this. But at a high level, embedded cost techniques rely on one year of data or the equivalent forecast for one year. For many inputs, marginal cost techniques often rely on multiple years of data in order to estimate how costs are changing with respect to different factors over time. Different data may be needed for each step of the process, starting from the functionalization of costs down to the creation of **allocation factors**, or allocators, to split up the costs to customer classes.

Where the utility's metering and data collection do not directly provide comprehensive load data for all customers and system components, two options are available. The first and generally preferable option is sampling. Most investor-owned and larger consumer-owned utilities install **interval meters** specifically for load research purposes on a sample of customers in each class that does not have widespread interval metering.[41] The number and distribution of those meters should be determined to provide a representative mix of customer loads within the class (or other subgroups of interest) and to produce estimates of critical values (such as contribution to the monthly system peak load) that reach target levels of statistical significance.[42] These samples are typically a few hundred per class in order to meet the PURPA standard. Second, some smaller utilities borrow "proxy data" from a nearby utility with similar customer characteristics and more robust load research capabilities. Class load data are usually publicly available for regulated utilities. Neither sampled load nor proxy load will provide the precision of comprehensive interval metering, but they can provide reasonable estimates of the contribution of the group to demand at each hour, enabling development of cutting-edge techniques such as time-specific allocation methods.

Different elements of load research data are relevant in the creation of allocation factors for different parts of the system. For example:

- Most residential customers may be served through a transformer shared with other residential, commercial and street lighting customers, so the allocation of transformer costs to each class should ideally be derived from their contribution to the high-load periods of each such transformer.
- Some residential customers are served from feeders that peak in the morning and others from feeders that peak in midday or the evening; some of those feeders may reach their maximum load or stress in the summer and others in the winter. The sum of the class contribution to the various peak hours of the various feeders determines the share of peak-related costs allocated to the class for this portion of the distribution system.
- At the bulk power level, all customers share the generation and transmission system, and the diversity of all usage should be reflected, whether at the highest system hour of the year (a method known as 1 CP, for coincident peak), the highest hour of each month (12 CP) or the highest 200 hours of the year (200 CP), all **on-peak** hours, **midpeak** hours and off-peak hours, or any other criteria relevant for allocation.

Table 5 on the next page shows illustrative load research data for four customer classes. For the purposes of clear examples throughout the manual, we adopt the convention

---

41 Utilities usually have interval meters on customers over some consumption threshold for billing purposes. Smaller customers may have meters that record only total energy consumption over the billing period (typically a month), or both monthly energy and maximum hourly (or 15-minute) demand, neither of which provides any useful data for allocating time-dependent costs.

42 In 1979, FERC issued regulations to implement PURPA § 133 (16 U.S.C. § 2643), which requires the gathering of information on the cost of service. C.F.R. Title 18, Chapter 1, Subchapter K, Part 290.403(b) established the requirement, since repealed, that "the sampling method and procedures for collecting, processing, and analyzing the sample loads, taken together, shall be designed so as to provide reasonably accurate data consistent with available technology and equipment. An accuracy of plus or minus 10 percent at the 90 percent confidence level shall be used as a target for the measurement of group loads at the time of system and customer group peaks." See Federal Energy Regulatory Commission Order 48 (1979).

Table 5. Illustrative load research data

	Residential	Secondary commercial	Primary industrial	Street lighting	Total	Used for
Energy metrics (MWhs)						
Total	1,000,000	1,000,000	1,000,000	100,000	3,100,000	All energy-related costs, including generation, transmission, primary distribution
Total secondary	1,000,000	1,000,000	N/A	100,000	2,100,000	
Energy by time period						
Summer	600,000	650,000	500,000	30,000	1,780,000	
Winter	400,000	350,000	500,000	70,000	1,320,000	
Daytime	600,000	700,000	500,000	0	1,800,000	
Off-peak	400,000	350,000	500,000	90,000	1,340,000	
Midpeak	550,000	600,000	470,000	9,000	1,629,000	
Critical peak	50,000	50,000	30,000	1,000	131,000	
Customer metrics						
Line transformers used	20,000	10,000	N/A	20,000	50,000	Transformers, services
Customers	100,000	20,000	2,000	50,000	172,000	Billing
Demand metrics (MWs)						
Sum of customer NCP	2,000	1,000	N/A	100	3,100	Input to line transformers
Class NCP: circuit	400	400	250	100	1,150	Primary distribution
Class NCP: substation	300	300	225	100	925	Substations
System 1 CP	250	300	200	0	750	Transmission, generation
System monthly 12 CP	225	250	175	10	660	
System 200 CP	200	240	150	10	600	

of a commercial customer class of all general service customers served at secondary voltage, labeled as "Secondary commercial," and an industrial customer class of all general service customers served at primary voltage, labeled as "Primary industrial."

In this illustration, the sum of individual customer noncoincident peak demands is 3,100 MWs, excluding the primary industrial class that is not shown in the table.[43] However, the coincident peak demand served by the utility becomes more diverse as we move up the system, a phenomenon described in more detail in Section 5.1. As a result, the observed coincident peak demands are lower at more broadly shared portions of the system. At the highest level, this illustrative system has a 750-MW coincident peak demand for the highest single hour, labeled as "System 1 CP." In between, the sum of the class NCPs at the circuit level, labeled as "Class NCP: circuit," is 1,150 MWs, and the sum of the class NCPs at the substation level, labeled as "Class NCP: substation," is 925 MWs. Customers served at primary voltage (primary industrial) have no utility-provided line transformers, and the first level at which their demand is typically relevant is the circuit level.

The street lighting class is important to note with respect to the volatility of results. Because this class has zero daytime usage and a very different (typically completely stable overnight) load profile than other classes, it is highly affected by the choice between noncoincident methods and either coincident or hourly methods. In addition, because streetlights represent many points of delivery but are typically located only in places where other customers are nearby, this class almost never "causes" the installation of a transformer or the creation of a secondary delivery point but also does account for a huge number of the individual points of use

43 In Table 5, the sum of customer NCPs for the primary industrial class is shown as "N/A" because these customers do not use line transformers and thus this demand metric is not generally relevant to this class. For more general purposes, we are assuming that the sum of customer NCPs for the primary industrial class in this illustration is 300 MWs, bringing the overall total to 3,400 MWs.

Table 6. Simple allocation factors derived from illustrative load research data

	Residential	Secondary commercial	Primary Industrial	Street lighting	Used for
Energy metrics (MWhs)					
Total	32%	32%	32%	3%	All energy-related costs, including generation, transmission, distribution
Total secondary	48%	48%	N/A	5%	
Energy by time period					
Summer	34%	37%	28%	2%	
Winter	30%	27%	38%	5%	
Daytime	33%	39%	28%	0%	
Off-peak	30%	26%	37%	7%	
Midpeak	34%	37%	29%	1%	
Critical peak	38%	38%	23%	1%	
Customer metrics					
Line transformers used	40%	20%	N/A	40%	Transformers, services
Customers	79%	17%	3%	1%	Billing
Demand metrics (MWs)					
Sum of customer NCP	65%	32%	N/A	3%	Input to line transformers
Class NCP: circuit	35%	35%	22%	9%	Primary distribution (legacy)
Class NCP: substation	32%	32%	24%	11%	Substations
System 1 CP	33%	40%	27%	0%	Transmission, generation
System monthly 12 CP	34%	38%	27%	2%	
System 200 CP	33%	40%	25%	2%	

Note: Class percentages may not add up to 100 because of rounding.

on the system. Put another way, we all like streetlights near our homes and businesses, but nearly all of them go in as a secondary effect of residential or commercial development; a few are along major highways without a nearby residence or business, but these are rare.

The next step is generating allocation factors to be used in the allocation phase of the cost study. For embedded cost studies, these are applied to the total investment and expense by FERC account, while in marginal cost studies they are applied to the calculated unit costs for each type of system component.

Table 6 shows the data above converted to allocation factors. The only implicit assumption is that the circuit-level peak demand for the residential class is one-fourth of the customer NCP demand due to load diversity and that for the commercial class it is one-half, reflecting lower diversity of commercial customer usage across the day compared with residential load. The raw factors are computed simply by dividing each class contribution to each category by the system total, then converting to percentages. For embedded cost of service studies, this manual recommends the use of class hourly energy use as a common allocation factor for all shared system components in generation, transmission and distribution where the system is made up of components essential for service at any hour, but sized for maximum levels of usage, and where the class contribution to that usage varies. The only one of these factors that is not self-explanatory is the midpeak factor, which takes both on-peak and **critical peak** usage into account, reflecting class usage in all higher-cost hours. This is illustrative of the probability-of-dispatch method, in which the likelihood of any resource being dispatched at specified hours is measured. There is no diversity of street lighting usage in this example, but little or no demand imposed at the system peak hours. Customer weighting factors are typically based on the relative cost of meters and billing services for different types of customers, based on complexity.

Table 7. Composite allocation factors derived from illustrative load research data

Method	Components	Residential	Secondary commercial	Primary industrial	Street lighting	Used for
Equivalent peaker	20% system 200 CP/ 80% energy	32%	34%	31%	3%	Generation, transmission
On-peak	50% midpeak/ 50% critical peak	36%	38%	26%	1%	Peaking generation
Average and peak	50% class NCP/ 50% energy	34%	34%	27%	6%	Primary distribution
Minimum system	50% customer/ 50% class NCP: circuit	57%	26%	12%	5%	Circuits (legacy)
Equivalent peaker for transformers	20% delivery points/ 80% customer NCP	60%	30%	0%	11%	Line transformers and secondary service lines

Note: Class percentages may not add up to 100 because of rounding.

In Table 6, we have calculated allocation factors shown as a class percentage of each usage metric. In Part II, we discuss in what circumstances each of these will be appropriate for embedded cost of service studies. In many cases, weighted combinations of these are appropriate. Several commonly used composite allocation factors are shown in Table 7, computed by weighting values in Table 6.

Given the wide diversity of utilities and their load patterns, readers should be careful about overgeneralizing from these illustrative examples. However, some patterns will hold true across the board. For example, the minimum system method will always allocate more costs to classes with large numbers of customers, at least compared with the basic customer method.

6. Basic Frameworks for Cost Allocation

We group cost allocation studies into two primary families. Embedded cost studies look at existing costs making up the existing revenue requirement. Marginal cost studies look at changes in cost that will be driven by changes in customer requirements over a reasonable planning period of perhaps five to 20 years. In the same family as marginal cost studies, total service long-run incremental cost (TSLRIC) studies look at the cost of creating a new system to provide today's needs using today's technologies, optimized to today's needs. Each has a relevant role in determining the optimal allocation of costs, and regulators may want to consider more than one type of study when making allocation decisions for major utilities that affect millions of consumers.

6.1 Embedded Cost of Service Studies

Embedded cost of service studies may be the most common form of utility cost allocation study, often termed "fully allocated cost of service studies." Most state regulators require them, and nearly all self-regulated utilities rely on embedded cost of service studies. The distinctive feature of these studies is that they are focused on the cost of service and usage patterns in a test year, typically either immediately before the filing of the rate case or the future year that begins when new rates are scheduled to take effect. This means there is very little that accounts for changes over time, so it is primarily a static snapshot approach. Embedded cost of service studies are also closely linked to the revenue requirement approved in a rate case, which can be administratively convenient.

Generally speaking, in the traditional model displayed in Figure 18 on the next page, functionalization identifies the purpose served by each cost (or the underlying equipment or activity), classification identifies the general category of factors that drive the need for the cost, and allocation selects the parameter to be used in allocating the cost among classes.[44]

Although they are convenient parts of organizing a cost of service study, functionalization and classification decisions are not necessarily critical to the final class cost allocations. The cost of service study can get to the same final allocation in several ways. For example, consider the reality that a portion of transmission costs is driven by the need to interconnect remote generation to avoid fuel costs. This can be reflected by functionalizing a portion of transmission cost as generation, or by classifying a portion of transmission in the same manner as the remote generation, or it can be recognized by using a systemwide transmission allocator with some energy component. In either case, a portion of costs is allocated based on energy throughput, not solely on design capacity or actual capacity utilization.

6.1.1 Functionalization

In this first step, cost of service studies divide the utility's accounting costs into a handful of top-level functions that mirror the elements of the electric system. At a minimum, this includes three functions:[45]

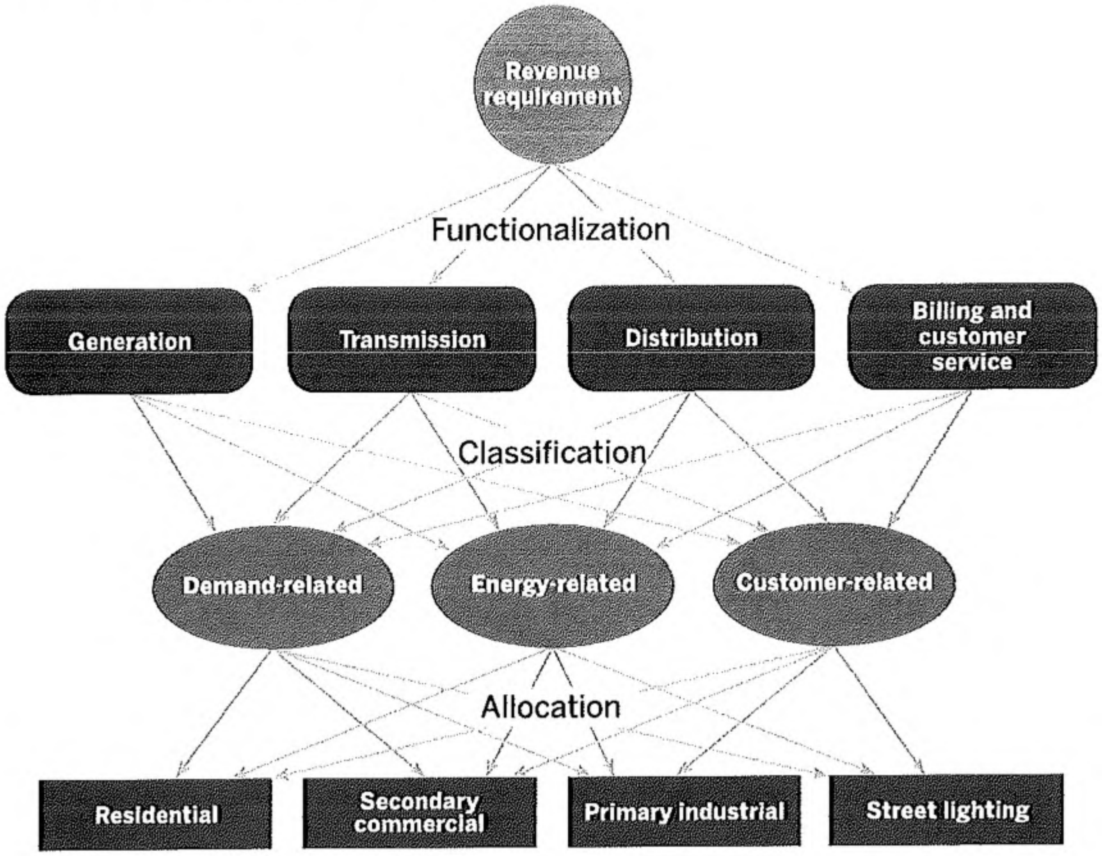
- Generation:[46] the power plants and supporting equipment, such as fuel supply and interconnections, as well as purchased power.
- Transmission: high-voltage lines (which may range from 50 kV to over 300 kV) and the substations connecting

44 The third step is usually called allocation, which is the same as the name of the entire process. This step involves the selection or development of allocation factors. Some analysts refer to this third step as factor allocation to prevent confusion.

45 Some of the costs, such as for energy efficiency programs and advanced meters, may serve multiple functions and must be assigned among those functions or treated as special functional categories.

46 Some sources use the term "production" instead. This manual uses the term "generation" and generally includes exports from storage facilities under this category.

Figure 18. Traditional embedded cost of service study flowchart



those lines, moving bulk power from generation to the distribution system.

- Distribution: lower-voltage primary feeders (in older systems, 4 kV and 8 kV; in newer areas, typically 13 kV to 34 kV) that run for many miles, mostly along roadways, and the distribution substations that step power down to distribution voltages; line transformers that step the primary voltages down to secondary voltages (mostly 120 V and 240 V); and the secondary lines that connect the transformers to some customers' service drops.

Although some utility analysts combine all costs into these three functions, the better practice is to include other functions as well at this stage:

- Billing and customer service: Also known as retail service or erroneously labeled entirely as customer-related costs, these are directly related to connecting customers (service drops, traditional meters) and interacting with them (meter reading, billing, communicating).
- General plant and administrative and general expenses: Overhead investments and expenses that jointly serve multiple functions (e.g., administration, financial, legal services, procurement, public relations, human resources, regulatory, information technology, and office buildings and equipment) can be kept separate at this stage. In some circumstances, these costs could be attributed to certain functions but are not tracked that way in a utility's system of accounts.
- Public policy program costs: In many jurisdictions, these costs are administered and allocated through another process; but if handled in a rate case, energy efficiency and other public policy programs should be tracked separately.

Historically, in most cases functionalization decisions can follow the utility's accounting and are noncontroversial.

The investment that is booked as generation units is usually part of the generation function. But there are exceptions. In some situations, the function of an investment may not match the accounting category. Examples include the following:

- Transmission lines and substations that are dedicated to connecting specific generating plants to the bulk transmission network. These assets are often in the accounting records as transmission but are more properly functionalized as generation.
- Substations that contain switching equipment to connect transmission lines of the same voltage to one another, high-voltage transformers that connect transmission lines of different voltages, and lower-voltage transformers that connect transmission to distribution. These facilities may be carried in the accounting records as entirely transmission or entirely distribution but are properly split between transmission and distribution in the functionalization process.
- Equipment within transmission substations that look like distribution equipment (e.g., poles, line transformers, secondary conductors, lighting). These might be booked in distribution accounts but are functionally part of the transmission substation.

In addition, many cost of service studies subfunctionalize some costs within a function, such as the following:

Generation

- Differentiating baseload generation (which runs whenever it is available or nearly so), intermediate generation (which typically runs several hours daily) and **peaking generation** (which runs only in a few high-load hours and when other generation is unavailable).
- Separating generators by technology to recognize such factors as renewable resources procured to meet energy-based environmental goals, the differing reliability contributions per installed kW of various technologies (e.g., wind, solar, thermal) and the differences in cost structure and output pattern between thermal, wind, solar and hydro resources.

Transmission

- Categorizing lines (and associated substations) by their role in operations, such as networking together the utility's service territory, providing radial supply to scattered distribution substations or importing low-cost baseload energy from distant suppliers.
- Segregating lower-voltage subtransmission facilities (typically under 100 kV) from higher-voltage facilities.
- Treating interconnections differently from the internal transmission network.
- Separating substations from lines.

Distribution

- Separating substations, lines (comprising overhead poles, underground conduit and the wires) and line transformers.
- Segregating costs of system monitoring, control and optimization related to reducing losses, improving **power quality** and integrating distributed renewables and storage.
- Dividing lines into primary and secondary components.
- In some cases, separating underground from overhead lines.

Billing and customer service

- Subfunctionalizing meters, services, meter reading, billing, customer service and other components, each of which may be allocated separately.
- Separating meters by technology — traditional kWh meters, **demand meters**, remotely read meters and advanced meters with hourly load recording and other capabilities — with different costs and different functions (including, for the advanced meters, services to the entire system).

General plant and administrative and general expenses

- Subfunctionalizing by type of cost: pensions and benefits, property insurance, legal, regulatory, administration, buildings, office equipment and so on.

In the future, organizing costs by function probably will still be helpful in organizing thinking about cost causation, but the cost of service study may need to differentiate functions in new ways. For example, distributed generation, storage, energy efficiency, demand response and smart grid technologies can provide services that span generation, transmission and distribution.

6.1.2 Classification

The second step of the process classifies each function or subfunction (i.e., each type of plant and expense) as being caused by one or more categories of factors. In particular, most cost of service studies use the classification categories of demand (meaning some measure of loads in peak hours or other hours that contribute to stressing system reliability or increasing capacity requirements on the generation, transmission or distribution systems), energy and customer number, and some use other categories (e.g., direct assignment, such as of street lighting).

The classification of most costs as demand-, energy- or customer-related dates back many decades. These categories can still be used but need to be interpreted more carefully as the utility system has changed in many ways:

- Utility planning has become more sophisticated.
- Utilities have access to more granular and comprehensive data on load and equipment condition.
- The variety of generation resources has increased to include wind, solar and other renewables with performance characteristics very different from legacy thermal and hydro resources.
- Multiple storage technologies are affecting generation, transmission and distribution costs.
- Legacy hydro, nuclear and fossil resources continue to operate and provide benefits to the utility system, but new similar resources and even continued operation of some existing units may no longer be cost-effective. Until they are retired, all or a portion of costs will remain in the allocation study.
- Demand response programs have increased in scale, role and variety.
- Utility spending on energy efficiency programs has increased.
- Advanced metering technology has added system benefits to a traditionally customer-related asset.

The demand and energy classifications are often treated as totally separate but, as discussed in Chapter 5, the load in many hours contributes to needs that have traditionally been classified to demand, and some hours are more important than others in driving energy costs. With improved information about class loads, and with a range of new technologies, it may be appropriate to move past the traditional energy and demand classifications and create new more granular distinctions, as discussed further in Chapter 17.

Table 8. 1992 NARUC cost allocation manual classification

Cost function	Typical cost classification
Production	Demand-related Energy-related
Transmission	Demand-related Energy-related
Distribution	Demand-related Energy-related Customer-related
Customer service	Customer-related Demand-related

Source: National Association of Regulatory Utility Commissioners. (1992). *Electric Utility Cost Allocation Manual*

Table 8 reproduces a table from the 1992 NARUC *Electric Utility Cost Allocation Manual*, showing how the classification step worked in that period (p. 21).

This was a simplification even at the time, and changes to the industry and in the available data and analytical techniques merit reevaluation and reform. For example, a legacy framework for variable renewable capacity, particularly wind and solar, could treat the investment for utility-owned resources as 100% demand-related, since there are no variable fuel costs. However, power purchase agreements for these same resources are typically priced on a per-kWh basis from independent power producers. This could lead to two different approaches for the same asset depending on the ownership model, an obvious error in analysis that should be avoided by considering the actual products and services being provided. In addition, most of the benefits of wind and solar do not necessarily accrue at peak hours — the underlying justification of a demand-related classification. Similarly, analog meters were only useful for measuring customer usage and billing, but new AMI provides data that can be used for system planning and provides new opportunities for energy management and peak load reduction.

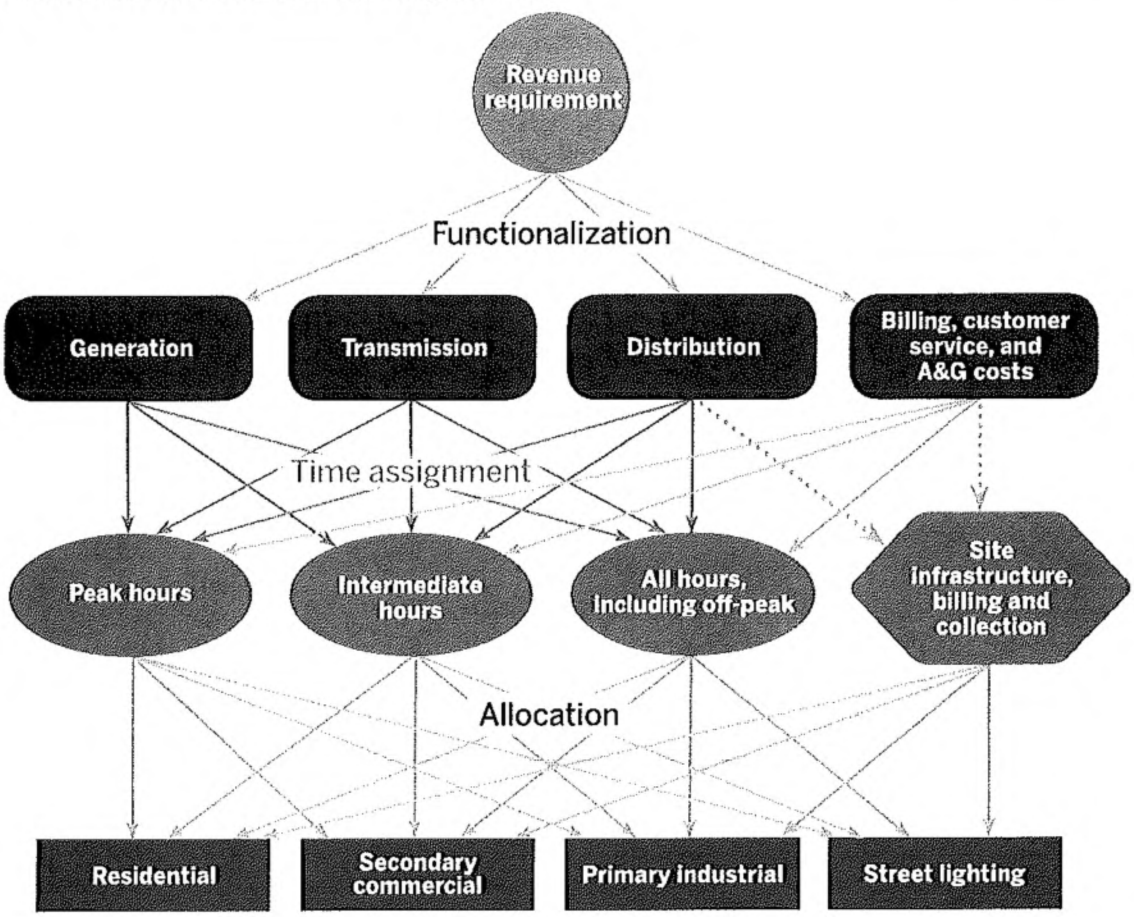
6.1.3 Allocation

The final step of the standard allocation process is the application of an allocation factor, or allocator, to each cost category.[47] An allocator is a percentage breakdown of the selected cost driver among classes. Within each broad type of classification, utilities use multiple allocators for various cost categories. For example, many different measures of "demand" are used to allocate demand-related costs, including various measures of contribution to coincident peaks (a single annual system coincident peak, or 1 CP); the average of several high-load monthly coincident peaks (e.g., 3 CP or 4 CP); the average of all 12 monthly coincident peak contributions (12 CP); the average of class contribution to some number of high-load hours (e.g., 200 CP); or different measurements of class maximum load (class noncoincident peak) at any time during the year. Usage of these peak-based demand allocators is often referred to as the **peak responsibility method.**

Generation allocators are sometimes differentiated among resources, to reflect the usage of different types of capacity and to retain the benefit of legacy resources for historic loads. Customer allocators are often weighted by the average cost of providing the service to customers in the various classes so that the cost of customer relations, for example, may be allocated with a weight of 1 for residential customers, 2 for small commercial, 5 for medium commercial and 20 for industrial.

Other costs, such as A&G expenses, are sometimes allocated on the basis of a labor allocator where the classification and allocation of underlying labor costs for the

Figure 19. Modern embedded cost of service study flowchart



47 Note that "allocation" is the term normally used for the entire process of assigning revenue requirements to classes and is also the term used for the last step of that process.

system is used for a set of other purposes. This is sometimes referred to as an internal allocator because it comes internally from previous calculations in the process. This is in contrast with "external allocators" based on facts and calculations outside of the cost allocation process, such as system peak and energy usage. Lastly, a variety of costs may be allocated based on a revenue allocator, which is based on the division of costs across all the classes.

6.1.4 Potential for Reform

As hourly data become available for all parts of the system, from transmission lines and substations through distribution feeders and line transformers to individual customers, an additional approach to classification and allocation becomes feasible: assigning costs directly to the time periods or operating conditions in which they are **used and useful**. This approach may entirely bypass the traditional classification step, at least between energy and demand.[48] Some relatively recent approaches recognize the complexity of cost drivers and combine classification and allocation into time-varying direct assignment of costs, as explained in Part II.

These time-varying allocation methods are discussed in Chapter 17 and Section 9.2; Figure 19 shows a simplified version.

Table 9 shows a simplified allocation study (very few cost categories and only two customer classes) and a caricature of the effect of using very different approaches. Both are embedded cost studies, but they produce dramatically different results.

The first study uses what might have passed for a reasonable cost allocation method a few decades ago, with all generation capacity and transmission costs allocated

Table 9. Results of two illustrative embedded cost of service study approaches

		Legacy study: Peak responsibility/minimum system			Modern study: Base-peak/basic customer		
Cost category	Revenue requirement	Allocation method	Residential	Commercial and industrial	Allocation method	Residential	Commercial and industrial
Generation							
Baseload	$100,000,000	Peak demand (1 CP)	$60,000,000	$40,000,000	All energy	$50,000,000	$50,000,000
Peaking	$50,000,000	Peak demand (1 CP)	$30,000,000	$20,000,000	On-peak energy	$27,500,000	$22,500,000
Fuel	$100,000,000	All energy	$50,000,000	$50,000,000	All energy	$50,000,000	$50,000,000
Subtotal			$140,000,000	$110,000,000		$127,500,000	$122,500,000
Transmission	$20,000,000	Peak demand (1 CP)	$12,000,000	$8,000,000	75% all energy/ 25% on-peak energy	$10,300,000	$9,800,000
Distribution							
Circuits	$50,000,000	50% peak demand/ 50% customer	$37,500,000	$12,500,000	75% all energy/ 25% on-peak energy	$25,600,000	$24,400,000
Transformers	$20,000,000	Customer	$18,000,000	$2,000,000	75% all energy/ 25% on-peak energy	$10,300,000	$9,800,000
Advanced meters	$10,000,000	Customer	$9,000,000	$1,000,000	50% customer/ 25% all energy/ 25% on-peak energy	$7,100,000	$2,900,000
Subtotal			$64,500,000	$15,500,000		$43,000,000	$37,000,000
Billing and collection	$20,000,000	Customer	$18,000,000	$2,000,000	Customer	$18,000,000	$2,000,000
Total	$370,000,000		$234,500,000	$135,500,000		$198,750,000	$171,250,000
Average per kWh	$0.123		$0.156	$0.09		$0.133	$0.114
Difference						-15%	+26%

Note: Numbers may not add up to total because of rounding.

48 Some costs associated with providing service under rare combinations of load and operating contingencies may not fit well into this framework.

Table 10. Illustrative allocation factors

Method	Residential	Commercial and industrial
Peak demand (1 CP)	60%	40%
All energy	50%	50%
On-peak energy	55%	45%
Customer	90%	10%
50% peak demand (1 CP)/ 50% customer	75%	25%
75% all energy/ 25% on-peak energy	51.3%	48.8%
50% customer/ 25% all energy/ 25% on-peak energy	71.3%	28.8%

on the highest-hour peak demand and most distribution costs allocated based on customer count. The second uses a simple time-based assignment method, in which all costs are allocated to usage in the hours for which the costs are incurred. This method recognizes that costs have a base level needed to provide service at all hours and incremental costs to provide service at peak hours. It also recognizes the multiple purposes for which advanced meter investments are made. The results are quite striking, with the second study showing a residential class revenue requirement 15% lower than the first. This set of assumptions probably forms the bookends between which most well-developed embedded cost studies would fall.

The first approach presents a legacy method that some industrial and large commercial customer representatives still sometimes propose. The second is a method that residential consumer advocates often champion. This change in method drives a significant change in the result. Both of these are "cost of service" results.

The point of these illustrative examples is not to suggest a specific approach, nor to defend any of the individual allocation methods shown, but to illustrate how different classification and allocation assumptions affect study results. Simply stating that a proposed cost assignment between classes is "based on the cost of service" may ignore the very important judgments that goes into the assumptions of the study. Table 10 shows the illustrative allocators that drive the results in Table 9.

Figure 20 on the next page shows a Sankey diagram for the legacy embedded cost of service study shown in Table 9. In that legacy study, most costs are classified as demand-related, and 60% of demand-related costs get allocated to the residential class. Similarly, a significant amount of costs are classified as customer-related, which are then overwhelmingly allocated to the residential class. This is because the **minimum system method** classifies all metering, billing and line transformers as customer-related, along with a portion of the distribution system.

In contrast, Figure 21 on Page 77 shows a Sankey diagram for the modern study in Table 9. More than half of peak hours costs are allocated to the residential class, but the peak hours classification is much less significant than the demand-related classification in the legacy study. Similarly, the basic customer method classifies only billing and a portion of advanced metering costs as customer-related. These costs are still primarily allocated to the residential class, but the aggregated differential nevertheless comes out significantly lower than in the legacy study. The remainder of advanced metering costs is split between all energy and on-peak energy because the purpose of these investments is to reduce energy costs and peak capacity requirements.

Figure 20. Sankey diagram for legacy embedded cost of service study

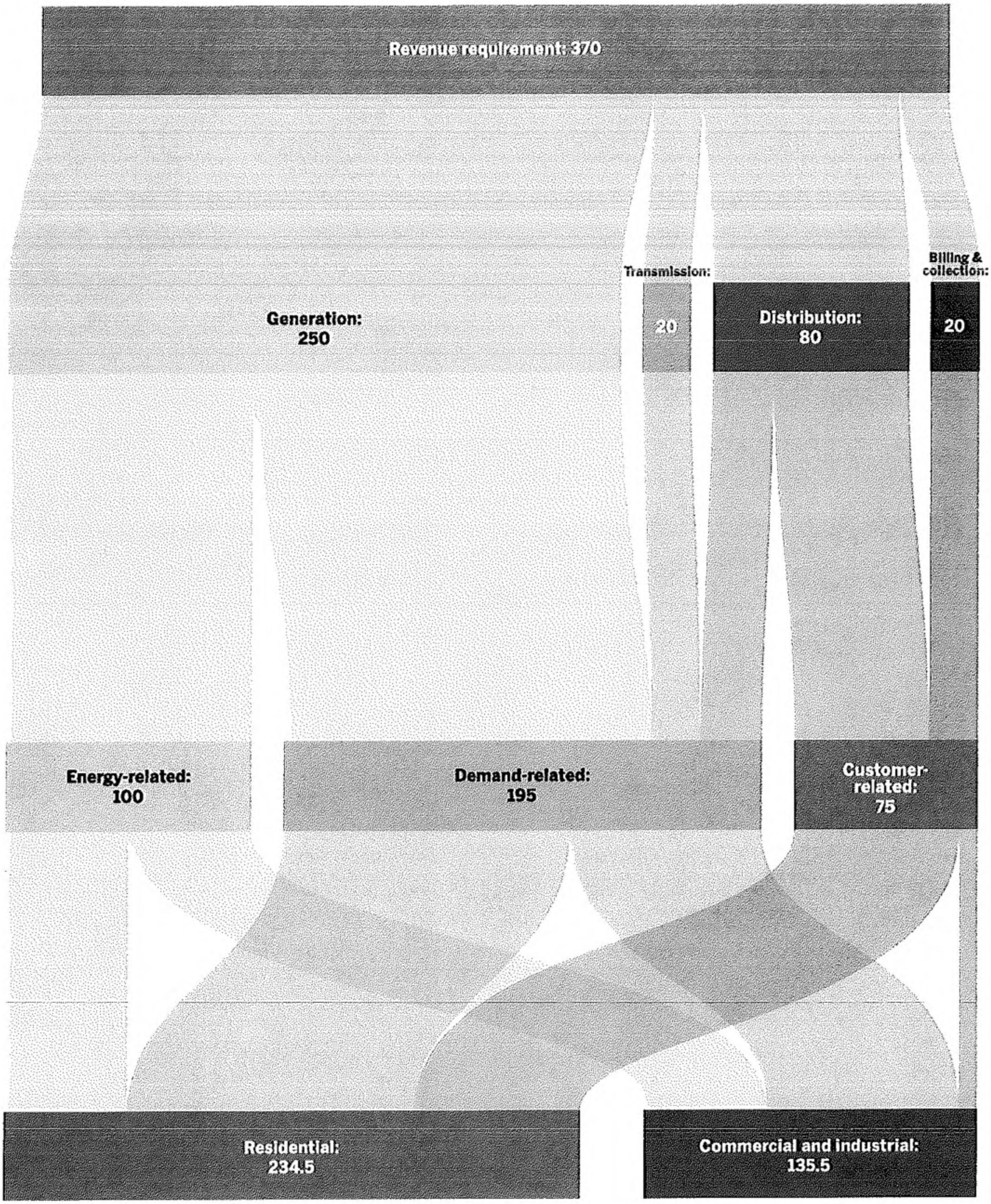
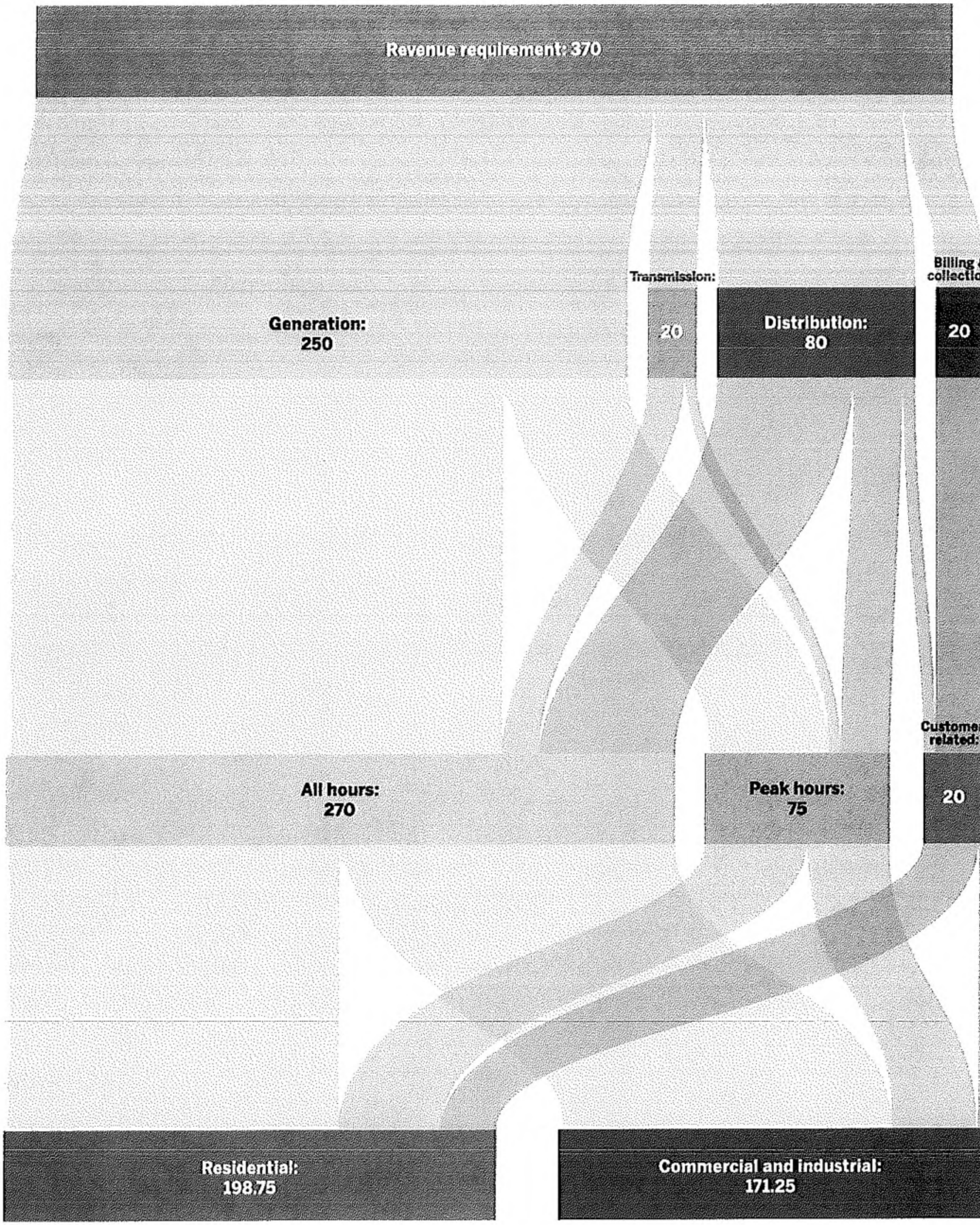


Figure 21. Sankey diagram for modern embedded cost of service study



"Fixed" versus "variable" costs

In the past, some cost allocation studies have relied on a simplified model of cost causation, in which certain costs are labeled as variable and then classified as energy-related and apportioned among classes based on class kWh usage. The remaining costs, labeled as fixed, are classified as demand-related or customer-related and allocated on some measure of peak demand or customer number, respectively.[49] This antiquated approach is based on fundamental misconceptions regarding cost causation. But it still underlies many arguments about cost allocation, perhaps because it typically works to the benefit of customer classes with high load factors and small numbers of customers — which describes most utilities' large industrial classes, data centers and even supermarkets.[50] This technique ignores the reality that modern electric systems trade off capital, labor, contractual obligations, fuel and other expenditures to minimize costs.

One of the problems with using the fixed/variable dichotomy to classify costs is the ambiguity of the concept of a cost being "fixed." Nearly all observers agree that certain generation costs are variable because they are short-term marginal costs that vary directly with usage patterns. These costs include:

- Fuel purchasing and disposal costs.[51]
- Variable operating costs related to consumables (e.g., water, limestone, activated carbon, ammonia) injected to increase output, reduce emissions or provide cooling to the power plant as it produces energy.
- Allowances or offsets that must be purchased to emit various pollutants.
- Purchased power charges that depend on the amount of energy taken by the utility.[52]

Over the decades, nearly every other utility cost has been described as fixed in one context or another: capital, labor, materials and contract services. Most of these costs are fixed for the coming year, in the sense that they are committed (investments made, contracts signed, employees hired) and will not be immediately changed by usage levels (energy, demand or number of customers). However, almost all of these cost accounts are variable over a period of several years, and energy consumption may affect:

- Whether excess generation capacity or other redundant facilities can be retired or mothballed in order to reduce operating and capital expenditures or repurposed to increase the net benefits of the facility.
- Whether additional facilities are needed (increasing capital and operating costs).
- Whether contracts are extended.
- The cost of capacity that is built (e.g., combined cycle versus combustion turbine plants, larger T&D equipment to reduce losses).

As a result, these costs are not fixed over the planning horizon. From an economic perspective more generally, all costs vary in the long run.

Relatedly, nearly all competitive businesses and fee-charging public services recover their fixed costs based on units sold. Customers do not pay an access fee to enter a supermarket.

49 In rate design, this approach has been extended to argue that all "fixed" costs must be recovered through **fixed charges**, often meaning customer and demand charges. These approaches promote neither equity nor efficiency.

50 Similarly, the fixed/variable approach is attractive to those who would justify rate designs with lower energy charges and higher customer and demand charges.

51 In previous decades, utilities would even argue that some fuel costs are fixed, on the grounds that having fuel on hand was necessary to allow the plant to function when required, or that a certain amount of fuel was required for startup, before any energy could be generated. These arguments appear to have largely disappeared, although similar issues are raised by the fuel security debate at FERC.

52 Many observers would add another category — expenses whose amount and timing vary with hours of operation, output or unit starts — even though not all cost of service studies separate those costs from other O&M expenses.

Restaurants, theaters and airlines have many costs that can be characterized as fixed (land, buildings, equipment, a large share of labor) and vary their unit prices by time of use but ultimately recover their capital investments and long-term costs from sales of output. RAP has done extensive analysis of utility distribution system investment and the relationship of that investment to the number of customers, peak demands and total kWhs. We found that these costs are roughly linear with respect to each of these metrics (Shirley, 2001).

Some version of the fixed/variable distinction may have been close to reality in the middle of the last century. Most utilities relied primarily on fossil steam plants, using newer, more efficient plants to serve baseloads and older plants to serve intermediate and peak loads. The capital costs of each were not very different. Fuel costs for oil, coal and natural gas were not very different. And because little was required in terms of emissions controls, coal plants were not much more expensive than other fossil-fueled plants.[53] By the 1970s, however, conditions had changed radically. Oil prices rose dramatically, new coal plants were required to reduce air emissions, and new generation technologies arose: nuclear, with high capital and O&M cost but low fuel prices; and combustion turbines, with low capital and O&M costs but high fuel costs. Utilities suddenly had a menu of options among generation technologies, including the potential for trading off short-term fuel costs for long-term capital investments. Today that menu has expanded even more and includes storage, demand response, price-responsive customer load and distributed generation.

As a result, the fixed/variable distinction has lost relevance and adherents over the last several decades. For example, many regulators classify capital investments using methods that recognize the contribution of energy requirements to the need for a wide variety of "fixed" costs for generation, transmission and distribution.[54]

53 In some areas, such as the U.S. Northwest, Manitoba and Québec, utilities had access to ample low-cost hydro facilities and mostly avoided construction of thermal generation.

54 These methods are discussed in chapters 9, 10 and 11.

6.2 Marginal Cost of Service Studies

The fundamental principle of marginal cost pricing is that economic efficiency is served when prices reflect current or future costs — that is, the true value today of the resources that are being used to serve demand — rather than historical embedded costs. Advocates for a marginal cost of service study approach work backward from this pricing concept to suggest that cost allocation should be based around marginal costs as well. Critics of marginal cost methods often point out that this economic theory is appropriate only when other conditions are present, including that all other goods are priced based on marginal costs, that there are no barriers to entry or exit from the market and that capital is fungible.

This is a very broad concept because it abstracts from and does not consider both theoretical and computational issues associated with the development of marginal costs. In contrast to the static snapshot that is typical of embedded cost approaches, marginal cost of service studies account for how costs change over time and which rate class characteristics are responsible for driving changes in cost. Importantly, marginal costs can be measured in the short run or long run. At one extreme, a true short-run marginal cost study will measure only a fraction of the cost of service, the portion that varies from hour to hour with usage assuming no changes in the capital stock. At the other, a total service long-run incremental cost study measures the cost of replacing today's power system with a new, optimally designed and sized system that uses the newest technology. In between is a range of alternatives, many of which have been used in states like Maine, New York, Montana, Oregon and California in determining revenue allocation among classes.

There is a strong theoretical link between optimal rate design and long-run marginal costs. Allocation based on marginal costs works backward from this premise; because pricing should be determined on this basis, cost allocation should as well. In its simplest form, a marginal cost study computes marginal costs for different elements of service, which can be estimated using a number of techniques, including proxies, regressions and other cost data. Table 11 shows illustrative marginal costs for different elements of the electric system.

Different marginal cost of service studies may base their costing on different elements of the system or different combinations. The categories of costs included in each element can also be more or less expansive. The estimated marginal costs are then multiplied by the billing determinants for each class. This produces a class marginal cost revenue requirement and, when combined with other classes, a system MCRR. However, revenue determination solely on this marginal cost basis will typically be greater or less than the allowed revenue requirement, which is normally computed on an embedded cost basis. It is only happenstance if marginal costs and embedded costs produce the same revenue or even similar levels of revenue. As a result, a marginal cost of service study must be adjusted to recover the correct annual amount from the revenue requirement.

Two notable long-run methods are discussed in this section: the long-run marginal cost approaches advocated by Lewis Perl and his colleagues at the consulting firm National Economic Research Associates (NERA) — now NERA Economic Consulting — and the total service long-run incremental cost approach.[55] In the 1980s, during the PURPA hearing era, many states considered and a few adopted the **NERA method** to measuring long-run marginal costs. California, Oregon, Montana and New York are examples of states that began relying on this approach to measuring marginal costs. This methodology generally looked at a 10-year or longer time horizon to measure what costs would change in response to changes in peak demand and energy requirements during different time periods and the number of customers served (National Economic Research Associates, 1977). One essential element of this was to define the cost of generation to meet peak period load growth (peaker units and associated T&D capacity) as much higher than the cost to meet off-peak load growth (increased utilization of existing assets). This approach was influenced by Alfred Kahn's theoretical focus on peak load costs and management (Kahn, 1970), and he himself was associated with NERA for many years.

For generation, one of the theoretical advances that made marginal cost of service studies attractive when they were first developed in the late 1970s was that generation costs were made up of capacity and energy costs, but the embedded plant was not classified to obtain these costs. Marginal energy costs were based on the incremental operating costs of the system (discussed in Chapter 18 in more detail), while capacity costs were the least cost of new capacity (at the time, typically a combustion turbine). The annualization for the capacity costs of all types is not based on the embedded rate of return but on a **real economic carrying charge** (RECC) rate that yields the same present value of revenue requirements when adjusted for inflation.

Table 11. Illustrative marginal cost results by element

	Units	Cost per unit
Customer connection	Dollars per year	$80
Secondary distribution	Dollars per kW	$40
Primary distribution	Dollars per kW	$80
Transmission	Dollars per kW	$50
Generation capacity	Dollars per kW	$100
Energy by time period		
On-peak	Dollars per kWh	$0.10
Midpeak	Dollars per kWh	$0.07
Off-peak	Dollars per kWh	$0.05

For transmission and distribution costs in the NERA method, the marginal costs have typically been estimated by determining marginal investment for new capacity over a number of historical and projected years and relating that investment to changes in some type of load or capacity measure in kWs. This relationship can be found either using regression equations (cumulative investment versus cumulative increase in load over the time period) or by simply dividing the number of dollars of investment by the total increase in load over the time period. O&M costs are generally based on some type of average over a number of historical and projected years, although obvious trends or anomalies can be taken into account.

55 Short-run marginal cost approaches are actually much simpler, primarily varying fuel consumption and purchased power costs, but are applicable only in a limited number of circumstances.

For customer costs, the same type of arguments over classification between distribution demand and customer costs occur as in embedded cost studies. The marginal cost study needs data on the current costs of hooking up new customers by class. The method for annualizing the costs is in dispute (RECC versus a **new-customer-only method** that assigns the costs by new and replacement customers). O&M costs are again typically based on some type of average over historical and projected years.

The time horizon used for the NERA approach has proven controversial because it assumed the utility would install exactly the number of new customer connections and distribution lines required by new customers (i.e., all customer costs are "marginal") but would consider the adequacy of existing generation and transmission (which may be oversized to meet current needs) in determining the need for additional generation and transmission (meaning only some G&T costs are "marginal"). Many utilities have used a 10-year time horizon in this analysis, a period in which many found substantial excess capacity and, therefore, relatively low costs to meet increasing power supply needs. In addition, this methodology, as most often used, treats the cost of increased off-peak usage as only the fuel and variable power costs and losses associated with operating existing resources for additional hours, with no associated investment-related or maintenance-related cost, despite the reliance on expensive investments to produce that power.

The combination of these assumptions meant that many marginal cost of service studies over the last several decades would come to three basic conclusions:

- Power supply and transmission costs to meet off-peak loads were relatively low, due to available excess capacity.
- Power supply and transmission costs to meet peak load growth were higher.
- Distribution costs always grew in lockstep with the number of customers and distribution demands.

The most serious shortcoming of the NERA methodology is that if power supply is surplus due to imperfect forecasting, it assigns a very low cost to power; if it is scarce, the method assigns a very high cost. Neither of those circumstances is *caused* by the action of consumers in any class, but the presence of either can shift costs sharply among consumer classes. Because of this imbalanced result, regulators have adopted modifications to this methodology to equalize the time horizon for different elements of the cost of service. For example, not all customers will require new service drops and meters over a 10-year period — only new customers and those whose existing facilities fail. Some states apportion costs within functional categories, avoiding this problem and addressing markets with partial retail choice.

In contrast to the NERA approach and other marginal cost approaches, which start from the parameters and investments found in the existing system, the total service long-run incremental cost approach looks at a period long enough so that all costs truly are variable. This allows for an estimate of what the system would look like if it were completely constructed using today's technologies and today's costs. Today, new generation is often cheaper than existing resources, while the cost of transmission and distribution continues to rise.

The TSLRIC approach was developed in the context of regulatory reform for telecommunications (International Telecommunication Union, 2009). In the 1990s, as telecommunication technology advanced rapidly, incumbent local exchange companies (better known as phone companies) faced competition from new market entrants that did not have legacy system costs. These new competitors were able to offer service at lower cost than the local phone companies. Regulators did not want to discourage innovation but also did not want existing customers served by the local phone companies to suffer rate increases if select customers left the system.

The TSLRIC approach constructs a hypothetical system with optimal sizing of components, with neither excess capacity nor deficient capacity. It would use the most modern technology. In the context of an electric utility, it would likely rely on wind, solar and storage to a greater extent than most systems today, which would likely lead to lower costs. But it would also incur the cost of today's environmental and land use restrictions, such as the requirement for lower emissions from generation and undergrounding of transmission and distribution lines. These requirements have substantial societal benefits but can also drive up electric system costs.

One advantage of a TSLRIC study over a NERA-style study is that no class is advantaged or disadvantaged by a current surplus or deficiency of power supply or distribution network capacity, since costs for all classes would be based on an optimal mix of resources to serve today's needs. This is one of the most common critiques of the NERA methodology — that it favors any class that is served dominantly by the elements of a system that are in surplus.

6.3 Combining Frameworks

Several jurisdictions require both an embedded and a marginal cost of service study to support cost allocation and rate design. As a result, utilities and other parties may file several studies in the course of a rate proceeding. A regulator may reasonably use multiple cost studies in reaching decisions, using multiple results to define a range of reasonableness. Within that range, the regulator can apply judgment and all of the relevant non-cost concerns to determine the allocation of the revenue requirements among classes. Furthermore, the different types of studies provide different information that can be used at other stages in the rate-making process.

One approach is to use embedded cost methods to determine the allocation of the revenue requirement among customer classes and then a forward-looking cost method of some kind to design rates within classes. This applies the focus of embedded cost studies on equitably sharing the costs among classes while maximizing the efficiency of price signals in the actual rates that individual customers face in making consumption decisions that will affect future costs. The appropriate form of price signals can also be influenced by externalities that are not part of the embedded costs for a regulated utility. For example, many regulatory agencies that allocate costs among classes on embedded costs have reflected higher long-run marginal costs in adopting inclining block or time-of-use rates for customers with high levels of usage (either because large customers are better able to respond to price signals or because the larger customers have more expensive load shapes, such as for space conditioning).

In some situations, regulators will use one costing method to set rates for existing load while using a different method to set rates for new customers or incremental usage. Some jurisdictions have applied this technique for rate design within classes — as the foundation for most "economic development" rate discounts where marginal costs are lower than embedded costs, as well as for inclining block rates where marginal costs are higher than embedded costs. In addition, some jurisdictions have applied this technique across rate classes, allocating new incremental resources to specific rate classes. Depending on the trajectory of costs, this can have two different intended purposes:

- To provide a foundation upon which to impose on fast-growing classes the high costs of growth and to shelter slower-growing classes from these new costs.
- To provide a foundation to give the benefit of low-cost new resources to the growing class.

This approach to differential treatment of incremental resources may be applicable to situations where costs are being driven by disparate growth among customer classes. In the 1980s, for example, commercial loads in the U.S. grew much faster than residential loads, and this technique could be used to assign the cost of expensive new resources to the classes causing those new costs to be incurred.

6.4 Using Cost of Service Study Results

Quantitative cost of service study results should serve only as a guide to the allocation of revenue responsibility among classes, not as the sole determinant. Even the best cost of service study reflects many judgments, assumptions and inputs. Other reasonable judgments, assumptions and inputs would result in different cost allocations. Additionally, loads may be unstable, significantly changing class revenue responsibility between cost studies, particularly for traditional studies that base costs on single peak hours in one or several months. More globally, concepts of equity extend beyond the cost of service study's assignment of responsibility for causing costs or using the services provided by those costs to include relative ability to pay, gradualism in rate changes, differential risks by function and class and other policy considerations.

Chapter 27 addresses the many ways in which the results of cost of service studies can be used to guide regulators.

7. Key Issues for 21st Century Cost Allocation

Many important cost allocation issues for the current era are fundamentally different from those that existed when NARUC published its 1992 *Electric Utility Cost Allocation Manual*. This chapter sets forth the changes the industry has experienced and describes the approaches that may be needed to address those changes in cost allocation studies.

Inevitably, additional costing issues will emerge and require recognition in future cost of service studies. The fundamental considerations are why the costs were incurred and who currently benefits from the costs. Costs are often categorized using engineering and accounting perspectives that are useful for many applications but must not be allowed to obscure the fundamental questions of causation and benefits.

7.1 Changes to Technology and the Electric System

Technological change has affected every element of the electric system since the studies and decisions that informed the 1992 NARUC cost allocation manual. These changes include:

- Improved distribution system monitoring and advanced metering infrastructure, leading to new comprehensive data on the system and customers.
- Evolution of resource options to include significant amounts of variable renewables, new types of storage, energy efficiency and demand response.
- Significant commitments to DERs behind customer meters, including rooftop solar and storage.
- Beneficial electrification of transportation.
- Changes in fuel prices and the resource supply mix that have dramatically changed the operating pattern of various generation resources (addressed in more detail in Section 7.2).

These changes both enable and require new approaches in order to efficiently and equitably allocate costs across customer classes.

7.1.1 Distribution System Monitoring and Advanced Metering Infrastructure

In the past, customer meters were used solely to measure usage and render bills. Today, so-called smart meters are part of a complex web of assets that enable energy efficiency, peak load management and improved system reliability, in addition to the traditional measuring of usage and rendering of bills.

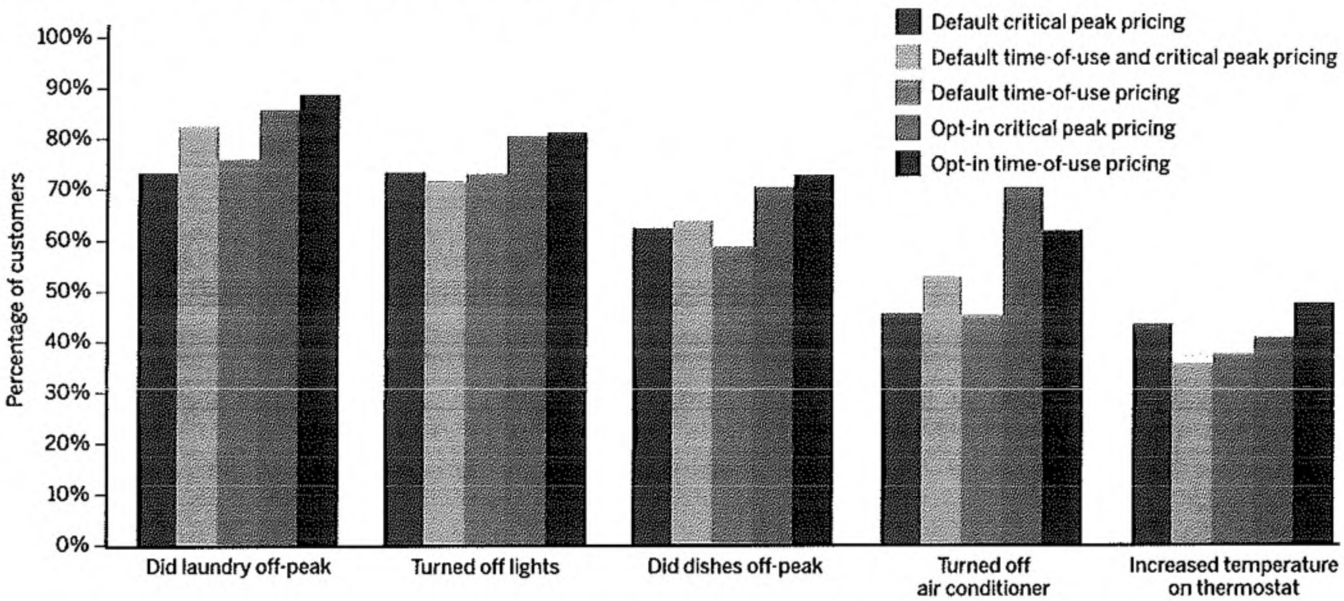
More recently, a number of utilities have used advanced meters to support demand response and other programs. Sacramento Municipal Utility District, for example, ran a pilot program to test the impacts of **dynamic pricing** and smart technology on peak load shaving and energy conservation. Figure 22 on the next page shows how customers in the program took steps to lower their electricity usage during high-load, higher-cost hours (Potter, George and Jimenez, 2014).

Smart meters (along with supporting data acquisition and data management hardware and software) can provide a number of services that improve reliability and reduce costs of generation, transmission and distribution.[56] Analysts have identified a wide range of expected and potential benefits. These include:

- Reduced line losses.
- Voltage control.
- Improved system planning and transformer sizing.
- The ability to implement rate designs that encourage energy efficiency.
- Reduced peak loads.
- Integration of EVs and renewables.

56 The broader concept of "smart grid" includes distribution (and sometimes transmission) automation devices such as automatic reclosers, voltage controls, switchable capacitors and sensors.

Figure 22. Customer behavior in Sacramento Municipal Utility District pricing pilot



Source: Potter, J., George, S., and Jimenez, L. (2014). *SmartPricing Options Final Evaluation*

- Operating savings from, among other things, reduced labor needs and improved outage management.

Lastly, smart meters, distribution sensors and modern computing power provide utilities with large amounts of data that can be used to determine the usage patterns of distribution and transmission equipment in great detail and support direct hourly allocation of costs.

7.1.2 Variable Renewables, Storage, Energy Efficiency and Demand Response

New variable renewable resources, such as wind and solar, are highly capital-intensive, and their contribution to system reliability varies greatly from region to region depending on when their generation occurs relative to peak demand.[57] The emergence of demand response as a service provides an opportunity to meet narrow periods of peak demand with relatively little capital investment by rewarding customers who curtail usage on request.

Investments in renewable resources, driven by policy and economic trends, can greatly change patterns in supply and demand that had been roughly constant for decades. Due to significant solar capacity in some regions, such as California and Hawaii, costs (e.g., extra **spinning reserves,** out-of-merit dispatch or quick-start generation) may also be incurred to rapidly ramp up other generation as solar output falls in the late afternoon, particularly if customer load does not drop dramatically from afternoon to evening.[58] Excess solar generation may create ramping costs, while storage resources may reduce ramping costs by both raising load at the beginning of the ramp period and trimming the peak toward the end of the ramp period.

In Hawaii, June load shapes changed as increased levels of distributed solar were added to the system. Figure 23 on the next page illustrates this, using data from the Federal Energy Regulatory Commission (n.d.). In 2006, the **system peak demand** was approximately 1,200 MWs at 1 to 3 p.m. By 2017, with extensive deployment of customer-sited solar, the peak demand was 1,068 MWs at 9 p.m. A cost allocation scheme must be adaptable enough to be relevant as significant changes in the shape and character of utility-served load take place.

57 Growth in solar resources, whether central or distributed, gradually reduces the reliability value of incremental solar capacity in many respects; the same is true for wind resources with respect to the reliability value of incremental wind and the equivalent for (if they become economically competitive) tidal and wave energy. In contrast, these different resources may be complementary to one another in certain respects.

58 The resulting load shape, first identified by Denholm, Margolis and Milford in 2008, is commonly known as a duck curve. See also Lazar (2016).