MIT EL 99-005 WP

Energy Laboratory

Massachusetts Institute
of Technology

The Value of Reliability in Power Systems - Pricing Operating Reserves -

June 1999

THE VALUE OF RELIABILITY IN POWER SYSTEMS
- PRICING OPERATING RESERVES -

José Fernando Prada

EL99-005 WP July 1999

THE VALUE OF RELIABILITY IN POWER SYSTEMS
- PRICING OPERATING RESERVES -

by
José Fernando Prada

ABSTRACT

The provision of operating reserve in power systems is revisited in the context of the deregulated power industry and of competitive power markets. The operating reserve is stand-by capacity that must be kept ready to generate energy to provide for unplanned outages of generating units. The former problem of establishing reserve requirements in the regulated industry has turned into the question of what kind of mechanisms should be devised to allocate and price this service in competitive power markets.

This report analyzes the allocation of operating reserves at the system operator level. In the past, reserve requirements have been defined using deterministic criteria such as "peak load percentage" or "loss of largest unit", which fail to consistently define the risk of supply shortages in the system. Furthermore, these quantity-constrained methods do not explicitly address economic criteria and, when employed in competitive markets, they do not capture the worth of added reliability provided by capacity reserves.

This work proposes the allocation of operating reserves through capacity markets using a stochastic demand model. The insurance-like characteristics of the operating reserve are used to derive a valuation model which is analytically consistent and reflects the economic value to customers of added reliability. The model can be expressed in the form of a demand curve for operating reserve. This curve can be used in auction-type capacity markets to determine the amount of reserve to be provided and its trading price.

FOREWORD

This report is based on a related Master of Science thesis submitted to the Department of Electrical Engineering and Computer Science and the Technology and Policy Program at MIT. The report is deliberately shorter than the original thesis and the presentation of the material has been reorganized for conciseness and clarity.

The thesis supervisor was Dr. Marija D. Ilic, EECS Senior Research Scientist and associated researcher at MIT Energy Lab, who made fundamental contributions to the elaboration of this work. The author also appreciate discussions with Dr. Richard Tabors and Mr. Frank Graves.

Financial support provided by the MIT E-Lab Consortium "New Concepts and Software for the Deregulated Industry" and by the Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología, COLCIENCIAS, is greatly appreciated.

Table of Contents

List of Figures

Introduction

During the last decade the United States and other countries have been restructuring their power sectors, abandoning the former regulated monopolistic model who ruled the provision of electric energy during most part of this century. The new 'deregulated' structures are based on free market principles, favoring competition among private participants and consumer choice. However, the theoretical bases of *deregulation* in the electricity industry are not completely developed yet, and the practical experience with electricity markets is still limited. In effect, the restructuring processes have brought about new problems and many open questions, especially regarding the introduction of competitive or market-based mechanisms and their effect on the reliability of power supply.

Reliability standards in power systems are traditionally established as a series of technical requirements to be fulfilled during planning and operation. In general, reliability requirements are met providing a group of services, known as *ancillary services*, which are necessary to protect the integrity of the system and to guarantee the production and delivery of electric power throughout the electric grid. These ancillary services include coordinated system operation, frequency regulation, energy balance, voltage support and generation reserves, among others. This report studies the valuation of short-term capacity reserves in deregulated power systems, proposing a market-based method that can be easily implemented in actual competitive power markets.

Background

Electric energy is produced and delivered practically on real time and there is no convenient method to readily store it. This makes necessary to maintain a continuous and almost instantaneous balance between production and consumption of electricity in power systems. A way to ensure energy balance is by keeping some margin of generation above the expected demand load, so the system can deal with unexpected mismatches between supply and demand leading to power shortages. Generation margins are attained by providing stand-by plant capacity and they represent *reserves* of generation capacity that can be rapidly utilized in case of a supply shortage.

Utilities have traditionally determined reserve requirements using working rules and more recently probabilistic techniques. They estimate a reasonable amount of capacity to be reserved and kept available, so that credible contingencies will not cause a failure of supply. Nevertheless, even when analytical methods are used, a final decision regarding reserve levels depends on the operator's judgment of what is the acceptable risk of system failure. In fact, although it is not always made explicit, this decision is a trade-off between the additional reliability offered to customers and the cost of keeping the reserves available.

The risk of shortages in generation can be reduced by increasing the investment in generation and the operating cost of keeping installed capacity available. However, overinvestment and high operating costs would be ultimately reflected in the bill paid by the customer. On the other hand, underinvestment and tight generation margins would lead to a low reliability offered to customers. In general, economic efficiency requires that the benefits of improvements in reliability be weighed against the costs of providing additional reliability. Accordingly, the main shortcoming of using quantity-constrained methods to estimate reserve requirements is that economic criteria are not explicitly included in the decision-making process.

Proposed Approach

In theory, capacity *markets* can allocate system reserve efficiently. In such a market the incremental benefits of added reliability are compared to the incremental costs of supplying capacity reserves. At equilibrium, marginal benefit equals marginal cost. The market-clearing process defines both the amount of capacity to be reserved and the corresponding trading price. In the deregulated industry, therefore, a meaningful mechanism to allocate reserves should be market-based. The market matches supply and demand, defines an efficient price for reserves and supports competition on the supply side, being consistent with the principles of economic deregulation.

Some current systems offer practical approaches to capacity markets (e.g. California). However, they use quantity-constrained methods to determine the reserve requirement and employ this ex-ante figure as demand for reserves. This approach provides little information about the value of reserves. Moreover, in markets where generators bid for making available standby capacity, opportunities arise for suppliers' strategic behavior in the reserves market and between the energy and reserves markets. By and large, the main obstacle found in establishing a market for reserves is how to determine the value of added reliability benefits derived from additional capacity available in the system.

The valuation of capacity reserves is less straightforward than the valuation of energy. In effect, spare capacity is not a consumable good as is electric energy. Instead, what capacity reserves provide is a hedge against the contingency of not having enough generation available to meet demand. Essentially, a purchaser of reserves holds the *option* to buy the amount of energy implicit in the 'locked' capacity, and he will do so according to the actual energy deficit confronted. A pricing method that did not consider these insurance- and option-like features would miss the real value of capacity reserves. A suitable valuation model should then associate the price paid for reserved capacity with a premium paid for holding the related option or insurance policy. Alternative mechanisms based on regulated rates or operating cost minimization would lead to less efficient outcomes.

Objectives

This work studies the added-value features of generation reserves and proposes to use these characteristics to value reserves in capacity markets. The purpose of it is twofold. First, to create a suitable framework for markets of operating reserve, and secondly, to introduce a pricing model to value reserves. The objective is to offer a more efficient approach to the provision of operating reserve by taking into account its economic value. The gained insight is helpful to regulators setting up market rules for capacity markets and to *system operators* (ISOs) or *load aggregators* who reserve capacity in behalf of electricity consumers. In this way, they can make better-informed decisions in establishing market rules or purchasing reserves.

Work Organization

The validity of a market-based framework for operating reserves should be based on its consistency with engineering requirements and with accepted criteria for economic efficiency. In addition, it should be implementable in real systems. Accordingly, this report is organized as follows:

- Chapter 1 presents basic concepts of generation reliability and capacity reserves, and Chapter 2 reviews analytical methods and indices used to evaluate generation reliability.
- Chapter 3 discusses policy alternatives for the provision of operating reserve. The trade-off reliability vs. cost and the rationale for efficient allocation of reserves are considered.
- Chapter 4 analyzes the supply and demand sides of reserves markets. We examine the costs of supplying reserves and the benefits of having generation reserves in the system.
- Chapter 5 studies the value of operating reserve. We introduce a pricing model to assess the worth of reserves in capacity markets. Mathematical background is included in Annex 1 and a glossary in Annex 2.

Chapter 1

Generation Reserves

Providing additional generation capacity increases the reliability of power supply and adds value to the service, but it also costs money. Indeed, beyond certain point the added benefits do not justify the additional expenses incurred. In light of this consideration we would like to know how reliable system-wide generation should be and how to price generation reliability in competitive power markets. Before addressing these fundamental issues, we present in this chapter the basic concepts that provide the context to the arguments developed in subsequent chapters.

We begin by reviewing general concepts of power systems reliability, next, we consider ancillary services and the role they play in power systems, finally, we end the chapter discussing generation reserves. Power systems reliability considers the performance of the system as a whole, considering generation facilities as well as the transmission network and the distribution grid. For the purposes of this work only the reliability of generation capacity is analyzed.

1.1 Reliability

The term reliability is broad in meaning. In general, reliability designates the ability of a system to perform its assigned function, where past experience helps to form advance estimates of future performance. A useful definition that illustrates the different dimensions of the concept is the following [1]:

> *Reliability* is the probability of a device or system performing its function adequately, for the period of time intended, under the operating conditions intended.

Reliability can be measured through the mathematical concept of probability by identifying the probability of successful performance with the degree of reliability. Generally, a device or system is said to perform satisfactorily if it does not fail during the time of service. On the other hand, a broad range of devices are expected to undergo failures, be repaired and then returned to service during their entire useful life. In this case a more appropriate measure of reliability is the availability of the device, which is defined as follows:

> The *availability* of a repairable device is the proportion of time, during the intended time of service, that the device is in, or ready for service.

The indices used in reliability evaluation are probabilistic and, consequently, they do not provide exact predictions. They state averages of past events and chances of future ones by means of most frequent values and long-run averages. This information should be complemented with other economic and policy considerations for decision-making in planning, design and operation.

1.2 Power Systems Reliability

The function of an electric power system is to provide electricity to its customers efficiently and with a reasonable assurance of continuity and quality [2]. The task of achieving economic efficiency is assigned to system operators or competitive markets, depending on the type of industry structure adopted. On the other hand, the quality of the service is evaluated by the extent to which the supply of electricity is available to customers at a usable voltage and frequency. The reliability of power supply is, therefore, related to the probability of providing customers with continuous service and with a voltage and frequency within prescribed ranges around the nominal values.

A modern power system is complex, highly integrated and very large. Fortunately, the system can be divided into appropriate subsystems or functional areas that can be analyzed separately [3]. These functional areas are generation, transmission and distribution. Reliability studies are carried out individually and in combinations of the three areas. The evaluation of transmission and distribution reliability is beyond the scope of this work. Nevertheless, the following remarks are important when assessing the reliability of the entire power system:

- The actual degree of reliability experienced by a customer will vary from location to location. Different functional areas may offer different degrees of reliability.
- There should be uniformity between the reliability of various parts of the system. It is useless to strongly reinforce a part if weaker areas exist on the supply chain.
- In deregulated systems, efficient pricing mechanisms for transmission and distribution must consider a reliability component.

Reliability Evaluation

Electric power networks are good examples of reliable systems. In many power systems the average duration of interruptions experienced by a customer is just a few hours per year, which translates into high availability of power supply. In general, to ensure these levels of supply availability, the probability of load being disconnected for any reason is reduced by introducing redundancy in the system. For decades, satisfactory levels of reliability were achieved through empirical methods and policies. However, as systems grew larger and more complex, formal, more rigorous analytical techniques have been applied.

The first techniques used were all deterministic. Typical criteria included planning generation margins equal to a fixed percentage of the forecast peak demand and operating generation margins sufficient to cover the most likely contingencies. Additionally, network capacity is usually installed to meet the N-1 or N-2 criteria, which requires the system to operate with one or maximum two elements out of service. An important shortcoming of these methods is that they do not account for the stochastic nature of system behavior. Indeed, randomly occurring or probabilistic events in the system are easy to recognize: forced outages of generating units, failure of overhead lines, uncertainty in customer demand.

Probabilistic methods can provide more meaningful information to be used in design and resource in planning and allocation. There are two main approaches for probabilistic evaluation of power systems reliability: analytical methods and Monte Carlo simulation. Analytical techniques represent the system by mathematical models and use direct analytical solutions to evaluate a priori reliability indices from the model. Monte Carlo simulation estimates a posteriori reliability indices by simulating the actual random behavior of the system. Whichever approach is used, the predicted indices are as good as the derived models, the relevance of each technique and the quality of the data.

Reliability studies are conducted for two purposes. First, long-term evaluations are performed to assist in system planning. Secondly, short-term evaluations assist in day to day operating decisions. Typical reliability indices used in power systems evaluation are the following:

- Load interruption indices: Average load interrupted per period of time.
- Loss of load probability: Probability of load exceeding available generation.
- Frequency and duration indices: Average number of occurrences and duration of interruptions per time period.

1.3 Ancillary Services and Reliability

The term ancillary services was coined during the first restructuring and liberalization processes. It designates the whole range of services necessary for power systems' successful performance and different from the basic functions of generation, transmission and distribution of energy. The ancillary services support basic energy supply and delivery functions that are essential for bulk power reliability. In effect, the former problem of determining how reliable the service should be has turned into the problem of how to provide and price the different (ancillary) services. They are also commercially important. These services are estimated to cost about $12 billion a year to U.S. electricity consumers, compared to the $15 billion a year that consumers pay for transmission services [4].

The Federal Energy Regulatory Commission FERC, in its benchmark Open Access Order No. 888 [5], defined ancillary services as those "necessary to support the transmission of electric power from seller to purchaser given the obligations… to maintain reliable operations of the interconnected transmission system". Order 888 recognized six types of ancillary services: Scheduling, System Control and Dispatch Service; Reactive Supply and Voltage Control; Regulation and Frequency Response; Energy Imbalance Service; Operating Spinning Reserve and Operating Supplemental Reserve. This list is by no means exhaustive, some other services are not necessarily associated with energy transmission (e.g. black start capability) and sometimes it is difficult to establish a clear-cut division between services.

FERC's Order 888 included a pro-forma tariff for the six ancillary services. This tariff prices the services on the basis of embedded costs. However, because most of these services are provided by generating units, it should be possible to create competitive markets for the procurement and pricing of ancillary services. Creating viable markets for these services is not a trivial task, though, given their complexity and the lack of experience with their unbundled provision and trading. In general, the pricing of ancillary services will depend on how the market is structured. Alternative mechanisms will produce different prices to consumers, producer's profits and system reliability. In addition, since energy and many ancillary services are provided by generating units, their markets will be correlated, with the prices of ancillary services depending on current energy prices.

1.4 Generation Reserves

Improvements in system reliability can be achieved by using better components or incorporating redundancy. Generation redundancy is attained by providing generating capacity above that needed for maximum load demand and transfers. This spare capacity represents the reserve of generation necessary to keep the risk of power shortages below an acceptable level.

The determination of the required amount of generation reserves is an important aspect of both power system planning and operation. The problem can be conceptually divided into the installed capacity requirement and the operating capacity requirement [3]. The *installed capacity reserve* relates to the long-term ability of the system to meet the expected demand requirements while the *operating reserve* relates to the short-term ability to meet a given load.

> The installed capacity considers the capacity that must be planned and constructed in advance to provide for uncertainties in the forecast of demand growth, overhaul of generating equipment and plant maintenance, and generation outages that are not planned or scheduled.

The basic difference between installed and operating capacity is in the time period considered. In the short term there is less uncertainty about load forecast. Moreover, equipment overhaul and maintenance can be scheduled during off-peak load periods. Likewise, real-time balance of energy supply and demand, which is necessary to deal with load fluctuations, is achieved by automatic generation control. Consequently,

> The operating reserve represents the capacity that must be available to replace loss of generation due to forced outages.

The provision of generation reserves in regulated and deregulated systems are conceptually different problems. A very general description for each system is as follows:

- *Centralized Provision*. Necessary standby capacity is estimated in the planning phase; unit additions are scheduled and constructed accordingly. During the operating phase (weeks to days) the installed capacity is scheduled to provide for planned outages and short-term load forecasting. Then, operating reserve is scheduled day ahead to provide for uncertainty in load forecast and for unplanned outages of dispatched generation.

- *Decentralized Provision*. In the long run, installed capacity should be provided by private investors who forecast profits in both energy and reserves markets. In the medium run, a system operator coordinates the maintenance of units, otherwise average market prices of energy and capacity can serve as coordinating signal. That is, power producers will schedule unit maintenance during low season, off-peak periods, where prices for energy and capacity are lower on average. Finally, in the short run (day ahead), operating reserve is allocated by capacity markets.

1.5 Operating Reserve

Assuming there is sufficient installed capacity in the system, the allocation of operating reserves consists in the decision concerning the capacity and units to commit to replace failed generating units. The risk of load interruption upon the failure of a generating unit can be minimized keeping part of the reserve 'spinning'; that is, as units connected to the grid, synchronized and ready to take load, or keeping available a group of units with quick-start capability. These units can be rapidly brought on-line and pick up load.

Both the spinning and non-spinning reserve form the *operating reserve* of the system. Non-spinning reserve can only be provided by hydraulic or gas turbine units which have start-up times in the order of minutes, whereas spinning reserve can be provided by a broader range of units. Actually, the division between spinning and non-spinning reserve can be actually one of definition. Fast-start units can be considered spinning reserve; interruptible loads and assistance from interconnected systems can be included in both categories. Accordingly, some systems may or may not include non-spinning reserve when assessing generation reliability.

Besides the immediate response group, conventionally able to be brought on-line in less than 10 minutes, a slower contingent of reserves, or "hot reserve", can be kept available. The hot reserve is capacity generally provided by thermal generation where the turbo-alternator is shut down but the boiler is left in a hot state. Thus, some regions like New York and New England require additional reserve that must be fully available within 30 minutes. California ISO requires a replacement reserve to be fully available within 60 minutes. This additional reserve (replacement, secondary) is used to redispatch after contingencies and to restore operating reserve requirements.

Requirements

NERC's Operating Manual [6] recommends keeping half of the operating reserve spinning. NERC also specifies that following a loss of resources, a Control Area shall take appropriates steps to reduce its Area Control Error (ACE) to meet the Disturbance Control Standard. The ACE is a measure of instantaneous unbalance of actual and scheduled generation and demand, taking into account energy exchanges with other control areas. NERC's traditional operating criteria for disturbance conditions are the B1 and B2 standards. The B1 standard requires the ACE to return to zero within 10 minutes following the start of the disturbance. B2 requires the ACE to start to return to zero within 1 minute following the start of the disturbance. The immediate response and the 10-minute full response associated with the B1 and B2 standards are cited as the origin of spinning and non-spinning requirements [7].

However, there is no agreement in minimum operating reserve standards. In the East Central Area Reliability Council (ECAR), the requirements for spinning and non-spinning reserve are both 3% of daily peak load. In the mid-Atlantic region, spinning reserve must be greater of 700 MW or the largest unit on line. In Florida, spinning reserve must equal 25% of the largest unit on-line. The Western Systems Coordinating Council requires reserves equal to 5% of load supplied by hydroelectric resources plus 7% of the load supplied by thermal generation, with spinning reserves not less than one half of the total operating reserve. In effect, beyond NERC's operating standards and each region's operating experience, there are no other bases for the adopted reserve requirements.

Chapter 2

Reliability Evaluation

The reliability of supply in a power system can be improved by increasing the investment in installed capacity and by incurring the operating costs of keeping reserves in service. For this reason, it is necessary to measure the reliability that these expenditures are buying. This chapter reviews different methods employed to evaluate the reliability of system-wide generation. We first introduce a stochastic model to evaluate the risk of supply shortages, next, we review common reliability measures used to assess said risk, and finally, we compare different reliability indices, looking at their physical relevance and usefulness. Most of the mathematical background used in this chapter is presented in Annex 1.

2.1 Generation Reliability

A power system, as any other system, consists of a set of components interconnected in some purposeful way. The object of a reliability study is to derive suitable measures of successful performance on the basis of component failure information and system configuration. For generation reliability studies the components of interest are the generating units and system configuration refers to the specific units scheduled to serve the load.

The indices used to measure generation reliability are probabilistic estimates of the ability of a particular generation configuration to supply the load demand. These indices are better understood as estimates of system-wide generation adequacy and not as absolute measures of system reliability. The indices are sensitive to basic factors like unit size and unit availability, and they are most useful when comparing the relative reliability of different generation configurations.

The basic elements used to evaluate generation adequacy are shown in Fig. 2.1. The system is deemed to operate successfully as long as there is sufficient generation capacity to supply the load. First, mathematical representations of generation and load are combined to model the risk of supply shortages in the system. Secondly, probabilistic estimates of shortage risk are used as indices of bulk power reliability for the considered configuration.

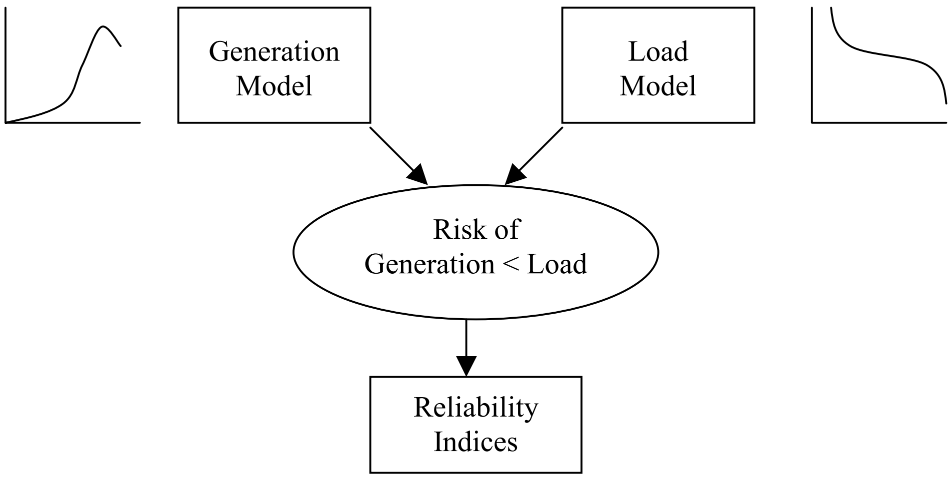


Fig. 2.1: Elements of generation reliability evaluation.

This approach only considers bulk generation and the aggregate load in the system. Evidently, the transmission and distribution grids are very important to evaluate the reliability offered to single customers. However, the model is sufficient for the purpose of comparing the adequacy of different generation configurations. Accordingly, the derived indices do not reflect generation deficiencies at any particular customer load point but they measure overall adequacy of generation capacity.

2.2 Risk of Supply Shortages

A model of bulk generation must consider the size of generation units and the two main processes involved in their operation, namely the failure and the restoration processes. A failure in a generating unit results in the unit being removed from service in order to be repaired or replaced, this event is known as an *outage*. Such outages can compromise the ability of the system to supply the load and affect system reliability.

An outage may or may not cause an interruption of service depending on the margins of generation provided. Outages also occur when the unit undergoes maintenance or other scheduled work necessary to keep it operating in good condition.

- A *forced outage* is an outage that results from emergency conditions, requiring that the component be taken out of service immediately.
- A *scheduled outage* is an outage that results when a component is deliberately taken out of service, usually for purposes of preventive maintenance or repair.

The status of a generating unit is conveniently described as residing in one of several possible states [8]. A hierarchical representation of said states is shown in Fig. 2.2.

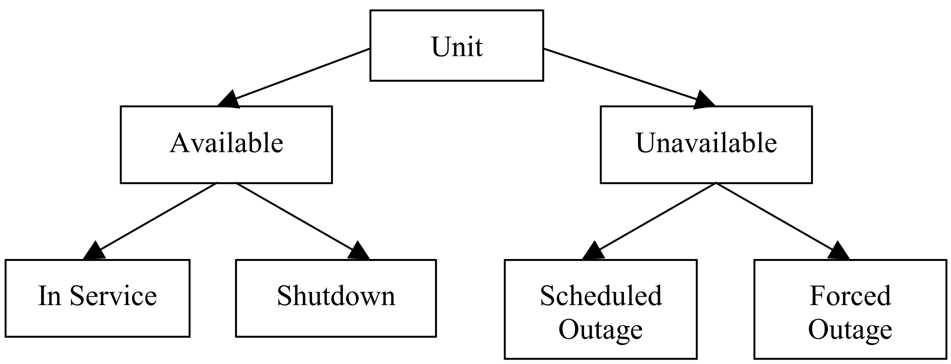


Fig. 2.2: Generating unit states

To investigate the effect of a unit on system generation reliability, it is sufficient to know its capacity and the probability of residing in each state. The state space representation of generation units is presented in the next section and the state space approach to reliability evaluation in section A1.2.

State Space Representation

The operating life of a generation unit can be represented by a simple two-state model in a "service-repair" process as shown in Fig. 2.3, where λ and μ are the unit failure and repair rate respectively (section A1.1). The most important quantity for generation reliability analysis is the probability of unit failure. The long-run failure probability, known as the unavailability of a unit, U, is defined by eq. (2.1).

$$U = \frac{\sum [\text{down time}]}{\sum [\text{down time}] + \sum [\text{up time}]} \tag{2.1}$$

The unit unavailability can be expressed in terms of unit's failure and repair rates (section A1.2), as indicated in eq. (2.2).

$$U = \frac{\lambda}{\lambda + \mu} = \frac{r}{m + r} \tag{2.2}$$

Where λ = unit failure rate, μ = unit repair rate.

m = mean time to failure = $1/\lambda$, r = mean time to repair = $1/\mu$

$T = m+r$ = mean cycle time, $f = 1/T$ = cycle frequency = $\mu.U$

The parameter U is a good approximation of a unit failure probability even when preventive maintenance is considered, provided that maintenance is scheduled during low demand periods. The unavailability is then an adequate estimator of the probability of finding a unit out of service at some point in the future.

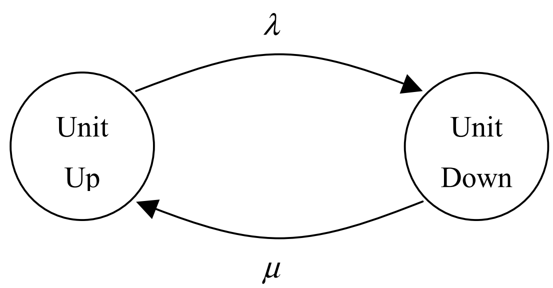


Fig. 2.3: Two-state model

The unit unavailability is commonly referred to as the 'forced outage rate', *FOR*, which in fact is not a rate but the ratio of eq. (2.3). If computed over a long period of time, the *FOR* is equivalent to unit unavailability. Models with multiple states can be used to represent partial outages as derated states. Multistate models are also useful to accommodate intermittent operation and start-up failure rates. Of course, the level of detail of the model depends on the degree of accuracy sought. In most reserve studies the two-state representation is sufficient.

$$\text{FOR} = \frac{\text{forced outage hours}}{\text{in service hours} + \text{forced outage hours}} \tag{2.3}$$

Capacity Outage Distribution

The final step in building a generation model is to combine the capacity and availability of the individual units to estimate available generation in the system. The result is a capacity model, in which each generating unit is represented by its nominal capacity g_i and its unavailability index u_i (or forced outage rate). For each of the N generators in the system, the available capacity $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}$, i =1…N, is a random variable that can take the value 0 with probability u_i and the value g_i with probability a_i =1-u_i [9], as shown in Fig. 2.4.

$$\tilde{\boldsymbol{g}}_{\boldsymbol{i}}\ (g_i, p_i) \begin{cases} \nearrow\ (g_i, a_i = 1 - u_i) & \rightarrow \text{ (unit available)} \\ \searrow\ (0, u_i) & \rightarrow \text{ (unit on outage)} \end{cases}$$

Fig. 2.4: Unit capacity

The total generating capacity available in the system is: $\tilde{\boldsymbol{G}}_{\boldsymbol{A}} = \sum_{i=1}^{N} \tilde{\boldsymbol{g}}_{\boldsymbol{i}}$ (2.4)

$\tilde{\boldsymbol{G}}_{\boldsymbol{A}}$ is a random variable itself. We assume that all units can fail and be repaired independently of failures and repairs of other units. Under these conditions, the probability distribution of $\tilde{\boldsymbol{G}}_{\boldsymbol{A}}$ can be obtained combining the single probabilities of the different $\tilde{\boldsymbol{g}}_{\boldsymbol{i}}$. The result is a discrete capacity distribution $\tilde{\boldsymbol{G}}_{\boldsymbol{A}} = \{G_j, p_j\}$, j = 1…2^N, with a sample space of 2^N capacity states. Each capacity state represents an outage event with one or several units out of service.

The capacity of the *j-th* state, G_j, with k available units and N-k failed units is the sum of the capacities of the k available units, or

$$G_j = g_1 + \ldots + g_k \tag{2.5}$$

The probability of finding the *j-th* state is equal to the product of the probabilities a_i of the k available units and the probabilities u_i of the N-k out-of-service units, that is:

$$P_i = a_1 a_2 \ldots a_k \,.\, u_1 u_2 \ldots u_{N-k} \tag{2.6}$$

In general, the probability distribution of $\tilde{\boldsymbol{G}}_{\boldsymbol{A}}$ is given by the individual terms of the following binomial expansion [10]:

$$\prod_{i=1}^{N} (a_i + u_i) = a_1 a_2 ... a_N + a_1 a_2 ... a_{N-1} u_N + ... + a_1 a_2 ... a_{N-2} u_{N-1} u_N + ... + u_1 u_2 ... u_N \tag{2.7}$$

There are 2^N possible different capacity states. In practice, several states have the same capacity so they can be grouped in a single state with the same capacity and probability equal to the sum of the single probabilities. Finally, the model is reduced to a series of capacity states and probabilities defined as follows:

$$\widetilde{\boldsymbol{C}}_A = \{C_j, p_j\},\ j = 1 \ldots l,\ l \leq 2^{\mathrm{N}}.\ \text{ Where } C_j, p_j = \begin{cases} G_m, \sum P_m & \text{for } C_j = G_m \\ G_i\, , P_i & \text{otherwise} \end{cases} \tag{2.8}$$

This capacity probability distribution is usually tabulated and referred to as the *capacity outage probability table*. The construction of a capacity outage table for a simple three-unit system is illustrated in section 2.4. The following example shows the capacity outage distribution of a six-unit system.

Example 2.1: Consider a power system consisting of six generating units. The capacity of unit 1 is 300 MW, units 2 and 3 are 200 MW each, and units 4, 5, and 6 are 100 MW each. The forced outage rate of each unit is 0.05.

- The nominal installed capacity of the system is 1,000 MW. The capacity outage distribution of the system is shown in Table 2.1.
- Table 2.1 indicates the amount of capacity out of service (col. 2) and available (col. 3) for each state, the probability of each state (col. 4) and the cumulative probabilities (col. 5). The probability distribution is shown in Fig. 2.5.
- The average capacity available is 950 MW with a standard deviation of 97.5 MW. The distribution of available capacity is typically skewed as in Fig. 2.5.

Generation Shortages

The applicable capacity outage distribution needs to be combined with an appropriate system load representation to derive a measure of generation shortage risk. However, realistic load modeling is one of the more difficult problems in power systems [11]. A simple static, constant power, approach represents the aggregate load in the system using either demand duration histograms, in which the number of hours the load exceeds any given level is plotted, or historical load curves for typical days, weeks and seasons. For reliability evaluation the load duration models are most helpful.

Table 2.1 – Capacity Outage Table

STATE J	CAPACITY OUT (MW)	CAPACITY IN, C_J (MW)	PROBABILITY $P[C_A = C_J]$	CUM. PROBABILITY $P[C_A \leq C_J]$
0	0	1000	0.735092	1.000000
1	100	900	0.116067	0.264908
2	200	800	0.083487	0.148841
3	300	700	0.051014	0.065354
4	400	600	0.008788	0.014340
5	500	500	0.004727	0.005552
6	600	400	0.000666	0.000825
7	700	300	0.000141	0.000159
8	800	200	0.000018	0.000018
9	900	100	0.000000	0.000000
10	1000	0	0.000000	0.000000

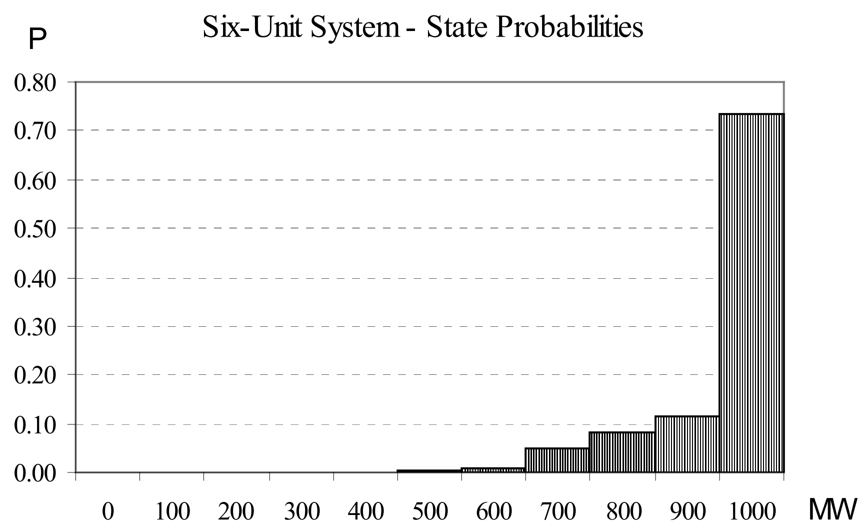


Fig. 2.5: Capacity probability distribution

The simplest load duration model is one in which each day is represented by its daily peak load. The individual peak loads are arranged in descending order to form a cumulative load model known as the *daily peak load variation* curve. Another method uses hourly load values in a given period and organize them in descending order to produce the *load duration* curve.

The advantage of this representation is that the area under the duration curve is the energy required in the period considered. Fig. 2.6 shows the typical shape of a load duration curve.

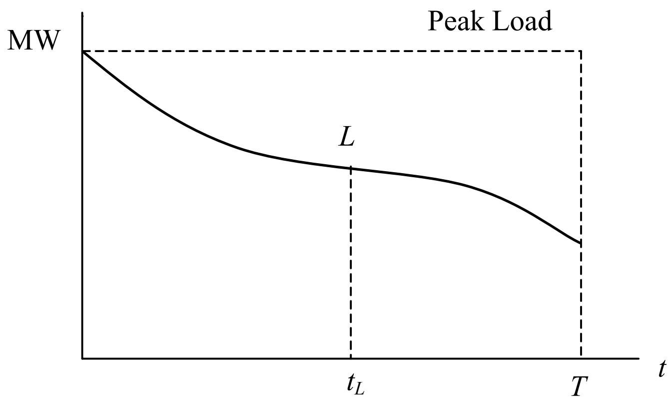


Fig. 2.6: Cumulative load curve

A supply shortage will occur whenever the system load exceeds the generating capacity remaining in service. If L is the system load, the probability of having power shortages will be the probability of all the outage events for which C_A is less than L, or $P[C_A \leq L]$.

2.3 Generation Reliability Indices

The application of probability models to the evaluation of generation reliability allows the integration of different unit sizes and types, the effects of maintenance, the capacity of interconnections and other factors. In addition, economic aspects can be better accommodated. The analytical methods commonly employed are the "loss of load" and the "frequency and duration" approaches.

Loss of Load

A loss of load will occur whenever the system load exceeds the generating capacity in service. The overall probability that the load demand will not be met is called the Loss-of-Load Probability or LOLP. For an expected load L and available generation capacity C_A, the LOLP is:

$$LOLP = \sum_j P[\widetilde{C}_A = C_j].P[L > C_j] \tag{2.9}$$

The simplest case is when the load is constant and known. If *Lo* is the expected load, the loss-of-load probability will be the probability of all the outage events leaving the system with an available capacity lower than *Lo*:

$$LOLP = \sum_j P[Lo > C_j] \tag{2.10}$$

Equation (2.10) is equal to the cumulative probability of $C_j < Lo$. Therefore, the LOLP can be read directly from the capacity outage table for a given dispatch. The assumption of a constant load is sufficient for evaluating short-run generation adequacy, for instance in systems where the dispatch is determined hourly. In this case load uncertainty is small and load fluctuations are taken care of by load following services.

The LOLP can be used to measure loss-of-load risk hour by hour or just consider the expected peak load during the dispatch period. For long-run and installed capacity evaluation, a cumulative load curve is used. The LOLP calculation is illustrated in Fig. 2.7 with a daily peak load curve. O_k is the magnitude of the k-th outage in the system, p_k is the probability of a capacity outage of magnitude O_k, and t_k is the number of days that an outage of magnitude O_k would cause a loss of load in the system.

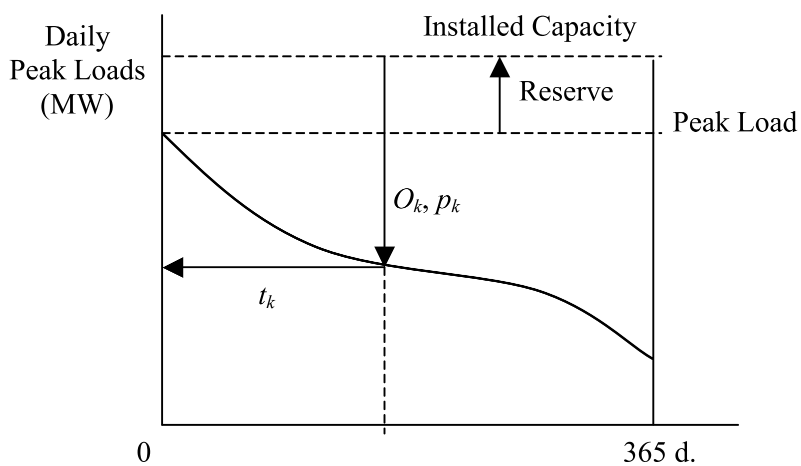


Fig. 2.7: LOLP calculation

Capacity outages less than the reserve will not contribute to loss-of-load risk. A particular capacity outage greater than the reserve will contribute to the overall risk by the amount p_k x t_k.

The system LOLP for the period is:

$$LOLP = \sum_{k} p_k \, . \, t_k \tag{2.11}$$

Equation (2.11) is an expected value instead of a probability, and it is also known as the loss of load expectation LOLE. When the daily peak load curve is used, the value of LOLE is in days for the period of study, usually days per year. A widely accepted LOLE risk criterion is the “one day in ten years” or 0.1 days/year standard.

Loss of Energy

The loss-of-energy method is a variation of the loss-of-load method. Here the measure of interest is the ratio of expected non-served energy to total energy demand over a period of time. If E_k is the energy not supplied due to a capacity outage O_k, and E is the total energy demand during the period of study, the Loss-of-Energy Probability LOEP is given by the following ratio:

$$LOEP = \sum_{k} \frac{E_k \, . \, p_k}{E} \tag{2.12}$$

Equation (2.12) is also known as the Loss-of-Energy Expectation LOEE. Again, the simplest case is when the load is constant and known. If Lo is the expected load during say 1h, the energy demanded is $Eo = Lo$ x 1h (MWh), and the system loss-of-energy probability during the hour will be:

$$LOEP = \sum_{j} \frac{E_j \, . \, p_j}{E_o} \approx \sum_{j} \frac{(Lo - C_j) \, . \, p(C_j)}{Lo} \quad , \quad \text{for } Lo > C_j \tag{2.13}$$

For longer periods and installed capacity evaluation the load duration curve is used. Any capacity outage exceeding the reserve will result in load interruption and energy curtailment. The non-served energy is the shaded area in Fig. 2.8. The system LOEP is given by eq. (2.12), where:

$$E_k = \int_0^{t_k} (L - C_k) \, dt \quad \text{and} \quad E = \int_0^{8760} L \, dt \tag{2.14}$$

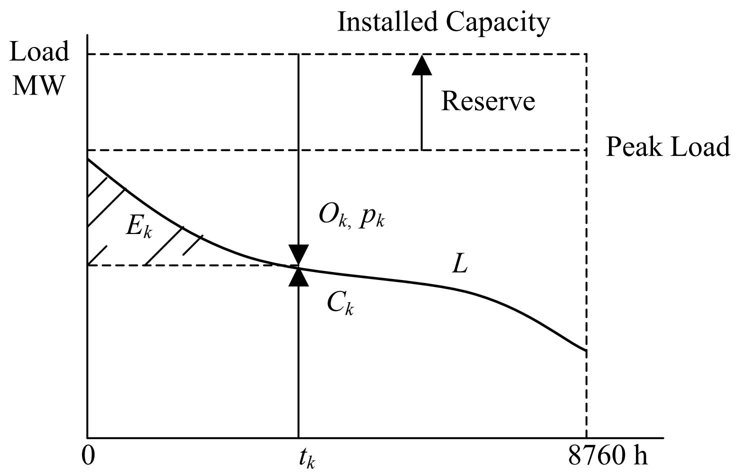


Fig. 2.8: LOEP calculation

Frequency and Duration

The LOLP and LOEP indices do not give indications about the frequency of occurrence or likely duration of a generation deficit. The frequency and duration FAD method measures these figures and is helpful to evaluate customer point reliability. The FAD method utilizes the transition rate parameters λ and μ of generation units presented in section 2.2. This technique applies the state-space approach (section A1.2) to the set of units present in the system. In short, each possible combination of units in up (in service) and down (forced outage) states defines a capacity state of the system. The resulting states are characterized by their available capacity, the associated state probabilities and the (intrastate) transition rates. The steps of a frequency and duration analysis are as follows:

- The capacities C_j and the probabilities p_j of each state are calculated as in section 2.2 for the system capacity outage distribution.
- The frequency of encountering a state j, f_j , is the expected number of stays in (or arrivals into, or departures from) j per unit time, computed over a long period.
- The frequency of state j is $f_j = p_j.(\lambda_{j+} + \lambda_{j-})$, where λ_{j+} is the transition rate from state j to higher capacity states and λ_{j-} the transition rate to lower capacity states.
- The average state duration T_j is defined by the relation $p_j = f_j \,.\, T_j$.

In the more general case, this representation is combined with a load model to identify marginal states, that is, states where a transition to a lower capacity state results in a generation deficit ($C_j < L$). Next, cumulative probabilities and frequencies are computed for the marginal states and suitable indices are derived.

Deterministic Criteria

A common practice developed over many years has been to evaluate system reliability in terms of working rules, which we will call deterministic criteria to differentiate them from the probabilistic techniques previously presented. The first rule, or 'percentage reserve' criterion, defines a target generation margin. The percentage reserve is established ad-hoc for each system. Representative ranges are 10-30% of peak demand in installed capacity and 2-10% in operation. This criterion compares the adequacy of reserve requirements in totally different systems on the sole basis of their peak load. Indeed, appreciable differences in capacity reserve may be needed to provide similar levels of reliability in systems with comparable peak loads.

Another widely used criterion calls for a reserve equivalent to the capacity of the largest unit on the system plus a fixed percentage of the dispatched capacity. This is somehow a sounder requirement because it is based on the criterion of providing for the largest credible contingency. In fact the probability of having a double outage event is much lower than losing a single unit. Anyway, these criteria do not consistently define the true risk of generation shortages. They clearly cannot account for the stochastic behavior of the system and they are hard to combine with economic criteria. Indeed, they may provide a misguiding sense of confidence in the adequacy of system generation on the basis of rule of thumbs.

2.4 Reliability Calculations

We present some simple numerical examples to illustrate the kind of calculations involved in generation reliability evaluation.

Capacity Outage Table Calculation

The following example illustrates the construction of a capacity outage table.

Consider a system consisting of three 25 MW units (#1, 2 and 3), each one having forced outage rates of 0.02. Table 2.2 shows the outage distribution for the system. Column 4 shows the probability of finding available the exact amount of capacity indicated in column 3, while column 5 shows the probability of finding available an amount of capacity equal or less than column 3.

Table 2.2 – Three-Unit System Capacity Outage Table

UNITS OUT #	CAPACITY OUT	CAP. IN (C_A)	PROBABILITY $P[C = C_A]$	CUM. PROB. $P[C \leq C_A]$
None	0 MW	75 MW	(0.98) (0.98) (0.98) = 0.9412	1.000
1, or 2, or 3	25 MW	50 MW	3 x (0.02) (0.98) (0.98) = 0.0576	0.0588
1,2 or 1,3 or 2,3	50 MW	25 MW	3 x (0.98) (0.02) (0.02) = 0.0012	0.0012
#1,2,3	75 MW	0 MW	(0.02) (0.02) (0.02) = 0.0000	0.0000

The cumulative probability of outages decreases as capacity on outage increases. In practice, the probability of having large amounts of capacity out due to the outage of several units is very low, and the capacity outage table can be truncated for probabilities below a specified amount (e.g. 10^{-7}).

Units can be easily added using conditional probabilities. For instance, adding a 25 MW unit with FOR = 0.02 will modify the probabilities of having 100 MW available (no outages) to 0.9412 x 0.98 = 0.9224, and the probabilities of 75 MW available to 0.9412 x 0.02 + 0.0576 x 0.98 = 0.0753.

Reliability Indices Calculation

To illustrate the calculation of reliability indices we will use the generation model of Table 2.2. We will combine that model with the load shown in fig. 2.9 representing a simplified load duration curve.

An on-peak load of 70 MW is assumed to last 40% of the time (3500h) and an off-peak load of 40 MW to be present the rest of the year. The capacity of the system above the peak load or system reserve is 5 MW, or approximately 7% of the peak load.

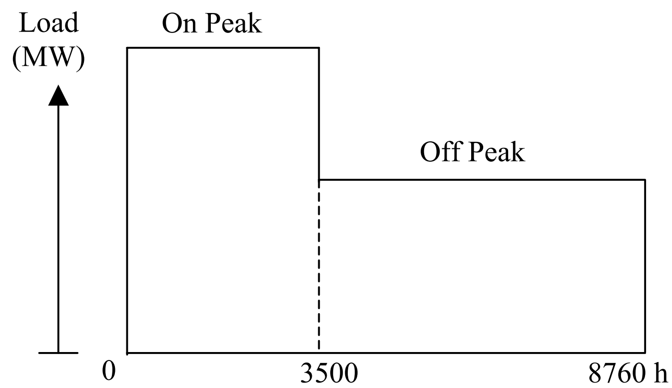


Fig. 2.9: Load model for calculations

- **LOLP**. The lowest outage level of 25 MW will cause loss of load only during 3500h, with associated probability 0.0576. Carrying out the calculations for the other capacity outages we obtain the loss-of-load probability and expectation:

  LOLP = 3500h x 0.0576 + 8760h x 0.0012 + 8760h x 0.0000 = 212 h/yr.

  LOLE = [20MW x 0.0576 + 45MW x 0.0012] x 0.4 + [15MW x 0.0012] x 0.6
  = 0.49MW

- **LOEP**. The expected energy curtailed in the three-unit system is:

  LOEE = 20MW x 3500h x 0.0576 + 45MW x 3500h x 0.0012
  + 15MW x 5260h x 0.0012 = 4,316 MWh

  The total energy demanded by the system is:

  E_T = 70MW x 3500h + 40MW x (8760h - 3500h) = 455,400 MWh

  Finally, LOEP = 4,316 / 455,400 = 0.00948

- **FAD**. The FAD requires a state space model of the system and the knowledge of the failure and repair rates of each unit. An example for a two-unit system is shown next.

Two-Unit System

The following example is taken from [9]. Consider a system with two generating units and the parameters of Table 2.3. The state space model for the two-unit system is shown in Fig. 2.10.

Table 2.3 – Unit Parameters

Unit	Capacity	Availability	λ (days^{-1})	μ (days^{-1})
1	20 MW	0.98	0.01	0.49
2	30 MW	0.98	0.01	0.49

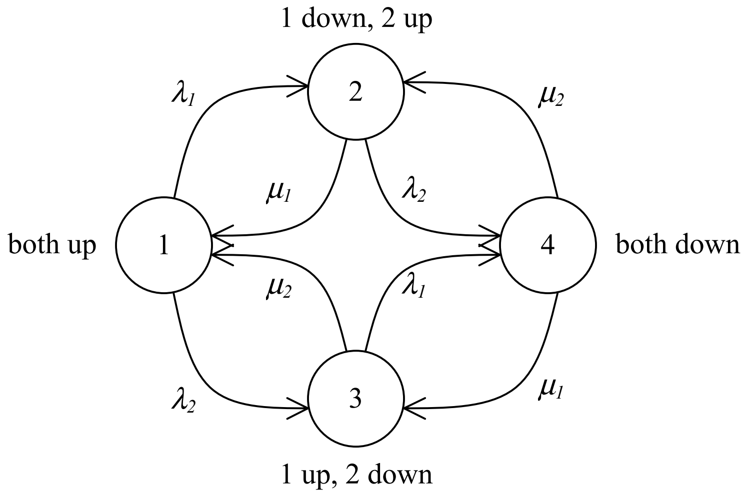


Fig. 2.10: Two-unit model

The capacity table of the system, including rates of departure, is shown in Table 2.4 and the frequency and duration of each particular state in Table 2.5.

Table 2.4 – Two-Unit Capacity Table

State #	Capacity	Availability	Rate of departure, λ_d
1	50 MW	(0.98)(0.98) = 0.9604	$\lambda_1 + \lambda_2 = 0.02$
2	30 MW	(0.02)(0.98) = 0.0196	$\mu_1 + \lambda_2 = 0.50$
3	20 MW	(0.98)(0.02) = 0.0196	$\lambda_1 + \mu_2 = 0.50$
4	0 MW	(0.02)(0.02) = 0.0004	$\mu_1 + \mu_2 = 0.98$

Table 2.5 – Capacity States Frequency and Duration

State #	Capacity	frequency (days^{-1})	duration (days)
1	50 MW	0.9604 x 0.02 = 0.0192	1 / 0.02 = 50
2	30 MW	0.0196 x 0.50 = 0.0098	1 / 0.50 = 2
3	20 MW	0.0196 x 0.50 = 0.0098	1 / 0.50 = 2
4	0 MW	0.0004 x 0.98 = 0.0004	1 / 0.98 = 1

2.5 Comparison of Reliability Indices

We have reviewed basic concepts and techniques used to evaluate generation reliability in power systems. The results are expressed as a series of reliability indices that measure the risk of supply shortages for a given configuration. However, in practice there is no uniformity in the interpretation and application of the different risk measures.

- **Deterministic Criteria**: The probabilistic methods are far superior to the percentage reserve and other rules of thumb often used. They provide analytical basis to consistently define system risk for different configurations. Deterministic criteria are insensitive to factors that significantly influence system reliability, such as unit size, failure rates or load characteristics. In fact, the reliability of two systems with same percentage reserve but different unit composition may be quite different. Moreover, the percentage reserve conveys the misguiding idea that all the risk can be removed keeping a fixed amount of reserves.

- **LOLP/LOLE:** This is the probability of system failure (to serve the load) based on a load duration curve or daily peak load curve. Depending on which load model is used the LOLP have different meanings. This index is often expressed as the expected fraction of time, LOLE, on which the system will be observed undergoing an outage event that leads to load of loss. All loss-of-load events count for its time contribution and not for the magnitude of the loss. LOLP/LOLE is easy to calculate and understand but it does not differentiate small capacity outages from large ones.

- **FAD:** The frequency of system failure measures the average number of failure occurrences per unit time. The corresponding duration indicates the average residence time on the failure states. This information is not provided by LOLP, but FAD does not either give information about the size of the outages when they occur. The frequency and duration of capacity outages have a greater physical significance than LOLP, but the FAD models require more detailed information about each generating unit and more computational effort.

- **LOEP:** It measures the expected fraction of system energy not served due to capacity outage events. The loss-of-energy approach has much greater physical relevance than the other approaches and takes into account the magnitude of the different outage events.

Reliability Evaluation and Capacity Reserves

Methods for evaluating generation adequacy can be very sophisticated, with generation and load models more elaborated than the ones presented here. In any case, the selection of a method and a risk index depends on the specific application. With regard to capacity reserves, different techniques of reliability assessment are used to determine the reserves necessary to keep the risk of supply shortages below a predetermined level. However, the methods do not provide any indication about the adequate level of that risk. In fact, none of them answers the simple question of how reliable the system should be. Indeed, any risk level could be selected to estimate the necessary reserves. In the next chapter we will explain how the selection of the 'risk level' must take into account the value of reliability to customers.

Our goal is to single out a measure of generation reliability that is also useful to measure the economic benefits to customers due to the presence of reserves in the system. For this purpose the LOEP method is the best approach. The loss-of-energy index measures the impact of capacity outages on customers by means of the expected energy curtailed. Other indices like LOLP and FAD focus on time recurrence and frequency of generation deficits without considering the magnitude of the shortages.

Evidently, from the point of view of customers, the effects of outages resulting on an average 1 MWh curtailment will be very different from others resulting on an average 1,000 MWh curtailment, even if they have the same recurrence. In effect, as it will be clearer in subsequent chapters, we will keep the basic generation model expressed in the capacity outage distribution, and instead of a single point estimate as LOEP, we will use the incremental probability of energy curtailments as the appropriate index of risk.

Chapter 3

Reserves Allocation

A decision concerning adequate levels of generation reliability and efficient provision determines the required reserve and the means to allocate it on a power system. In this chapter we discuss the main elements of this decision. We analyze the factors to be considered, the trade-offs involved, different alternatives and the rationale for efficient allocation of reserves in deregulated systems.

3.1 Reliability vs. Cost

The function of a power system is to supply electricity economically and with a reasonable assurance of continuity and quality. However, due to the integrated nature of power systems, failures in any part of the system can cause service disruptions. From the customer standpoint, power disruptions may be experienced as frequency and voltage reductions, as unstable supply with erratic frequency and power fluctuations or as a total interruption of supply. Although all these events impose costs on customers, in practice the effects of supply interruptions are the most severe.

Ideally, supply should be made continuously available to customers, but that is costly and arguably not feasible. In fact, interruptions of supply are caused by power outages, which are predominantly events of stochastic nature involving the failure of one or several components in the system. Due to the random aspect of system failures, it is accepted that any system will present a definite risk of suffering a number of future power shortages. That is, unbalances between power supply and demand leading to load interruptions. The risk can be reduced by installing better equipment or by providing additional capacity as generation reserves. The reserve can be dispatched to replace lost generation, effectively reducing the probability of load curtailments.

Consequently, in order to lessen the effects of power shortages on customers, it is necessary to invest in installed capacity and to incur the operating costs of keeping reserves available. As generation reliability is improved, a trade-off occurs between the increased costs of capacity reserves and the increased benefits to customers, as avoided costs, from fewer power shortages. Therefore, when making decisions concerning adequate levels of generation reliability, the factors to consider are the incremental costs, the benefits expected and the allocation of capital and operating resources among the different parts of the system. The objective is to determine an optimal balance between the economic benefits of higher reliability and the corresponding costs.

3.2 Reserves Provision

What the optimal level of reliability is? How much should be spent? Who should decide: power producers, regulators or customers? On what basis should the decision be made? The answers to these questions represent the policy a system follows with regard to its reserves of capacity. The central issue is how much reserve should be provided and by whom. In order to gain insight into system reserve requirements we look first at the regulated model, then we will turn back to deregulated systems.

Regulated Systems

The power supply industry in a regulated system is formed by electrical utilities, which are granted regional monopolies to provide the service at a regulated cost-based price. The concession of power supply monopolies allows utilities to exclusively serve a group of "captive" customers within a certain geographic area. The power franchise also makes them responsible for the planning and operation of the system and the reliability of supply. The development of reserve requirements in the monopolistic industry helps us understand important points.

First, it is clear that keeping reserves of generation adds value to the service, and all utilities have included reserve provision in their planning and operating activities associated with system reliability. Second, utilities have interconnected their systems because, among other reasons, capacity reserves can be shared, lowering the individual reserve margins and affording considerable capital economies. Third, the trade-off reliability-cost has always been recognized. For instance, generation reliability could be greatly improved if all the installed capacity were kept spinning, but this is recognized as not economical.

Utilities determine reserve requirements on the basis of operating experience and sound judgment, without including explicit economic criteria. In fact, the goal is to offer high reliability provided that cost is not excessive. Two reasons may explain this approach, which by and large have resulted in satisfactory levels of reliability, but uncertainty about the efficiency on the allocation of capital resources.

- First, utilities do not have incentives to reduce costs since cost-based pricing allows them to recover all their expenditures. Costs are passed to consumers, who do not realize which portion of their bills goes to cover reliability services.
- Second, tremendous difficulties are found in establishing the worth of reliability to consumers, especially considering that cost-based pricing does not provide information about the economic value of the service. In fact, acceptable methods to evaluate the worth of providing reliable service are not well established yet.

Deregulated Systems

The provision of generation reserves in deregulated systems depends on the institutional arrangements and competitive mechanisms selected, the degree of coordination among system participants and the extent to which decision-making is decentralized. In fact, for any arbitrary market structure, the provision of capacity reserve is not necessarily the responsibility of a system operator [12]. Furthermore, the conceptual base driving deregulation of the power industry favors competitive markets and decentralized decision-making as preferred mechanisms to allocate physical resources and foster individual choice. The provision of reserves should be in line with these concepts.

All practical designs for competitive systems establish the need of physically distinct enterprises to coordinate system operation and to coordinate the trading of products and services. These functions are assigned to independent organizations, known as the Independent System Operator (ISO) and the Power Exchange (PX). For convenience, we will refer hereafter to both as the PX/ISO.

This work seeks a mechanism to price and allocate operating capacity reserves. The mechanism should be consistent with the principles of deregulation, based on economic criteria and implementable in practical market structures, in order to determine a balance between reliability and cost that is economically efficient. Due to economies of coordination, we propose to assign to the PX/ISO the allocation of system reserves. What we need is, therefore, a decision criterion for reserve allocation to be used at the PX/ISO level. In the next section we look at some alternatives.

3.3 Decision Criteria

Cost-Effectiveness

The traditional criterion in systems with centralized decision making has been to use least-cost resources in order to meet arbitrary levels of generation reliability. This sort of *cost-effectiveness* criterion implies an a priori selection of a reliability level, usually based on experience and judgment. Gains realized from higher reliability are not considered. However, an increase in reliability may be advisable even if it results in a slight increase in cost, and a slight reduction in reliability may be acceptable if it results in significant savings. Accordingly, to reach an economically efficient outcome, the benefits gained by reliability improvements should be assessed against the costs of additional capacity.

Cost-Benefit Analysis

A better approach compares the incremental cost of reserves with the corresponding decline in outage costs, that is, the economic costs incurred by consumers because of supply interruptions. The objective is to minimize investment and operating plus outage costs over the period considered. The point of minimum cost marks the optimal level of reliability to be used as a benchmark in the system. The method is illustrated in Fig. 3.1. The investment and operating costs can be represented by curve RC, function of any suitable reliability index. Outage costs, represented by curve OC, decrease as reliability increases. The total cost curve TC is the sum of the individual cost curves RC and OC. Total cost presents a minimum at R*, which determines the optimal level of reliability.

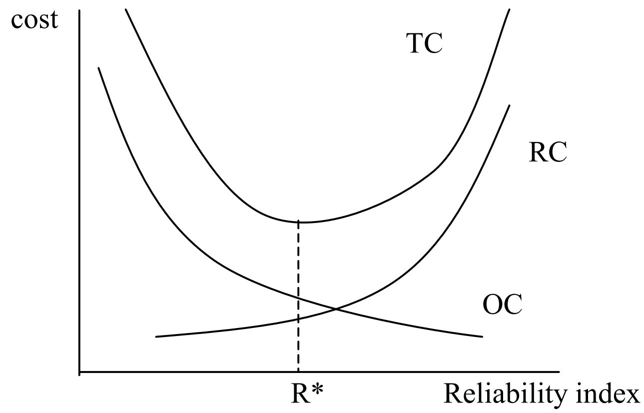


Fig. 3.1: Total reliability costs

The reliability level R* is treated as a variable and total social cost is minimized. This is equivalent to a *cost-benefit analysis*, which maximizes net social benefit. There are two difficulties in applying cost-benefit analysis. First, significant problems are found in assessing customer outage costs, and second, outage costs need to be related to an appropriate risk index used as measure of system reliability. Despite these difficulties, the cost-benefit analysis is a valid economic approach, but it is based on centralized decision-making. The cost-benefit approach does not incorporate individual choice being hardly compatible with competitive electricity markets, where suppliers decide individually the amount of capacity to commit.

Market-based Allocation

The basic criterion of economics relative to whether people are better off is the maximization of individual preferences, which translates into maximizing the utility consumers derive from a resource. It is known that this value is maximized when the allocation of resources is Pareto efficient, that is, when none can be made better off without making someone else worse off. A Pareto optimal resource allocation defines the economic benchmark for efficiency [13]. If markets are fairly competitive they lead to Pareto efficient outcomes and price reflects marginal social costs. Thus, competitive markets 'alone' allocate resources efficiently, without need of centralized direction. The pricing mechanism transmits the relevant information among market participants allowing individuals to decide what is best for themselves.

Market allocation is economically efficient. It allows decentralized decision-making and fosters individual choice. A meaningful mechanism to allocate capacity reserve in deregulated systems should be market-based. The market mechanism is shown in fig. 3.2. The supply curve S represents the price at which suppliers are willing to sell, equal to their marginal costs. The demand curve D indicates how much consumers are willing to pay, equal to the marginal value of the good. At equilibrium, supply equals demand and the market settles at the clearing price P* and the efficient level of production R*, which maximizes net social benefit [14].

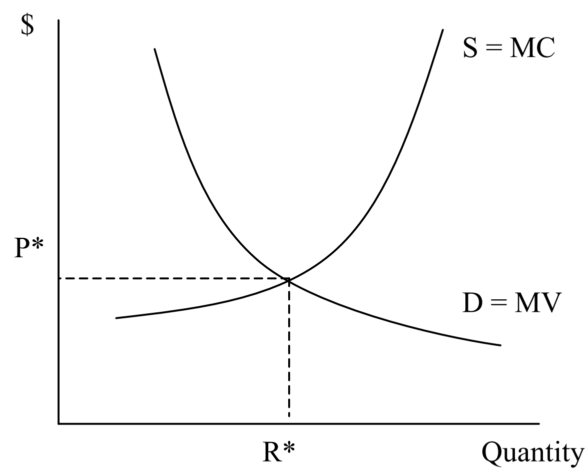


Fig. 3.2: The market mechanism

In theory, capacity markets can allocate reserves efficiently. In such a market the marginal benefit from increased reliability is equal to the marginal cost of supplying capacity reserves. When the market clears, it determines both the amount of capacity to be reserved R* and the trading price for reserves. R* defines the adequate level of reliability in the system, which is the one maximizing net benefit. A capacity market supports competition among reserve providers and sets the efficient price for reserves equal to the marginal cost of supply.

3.4 Markets for Capacity Reserves

We have argued in favor of market-based allocation of operating reserves and of allocation at the ISO level on the grounds of economic efficiency. In this section we propose a workable framework for a competitive market of reserves.

Proposed Market Framework

The proposed market structure follows usual auction-type electricity spot markets where hourly supply and demand bids for energy are submitted day ahead. The PX/ISO collects the bids and establishes the price and traded amount of energy hour by hour for the next day. In a market for reserves, the PX/ISO will collect hourly bids of capacity and will procure the service on behalf of the consumers. Afterwards, a schedule of operating reserve and capacity prices for the 24 hours of the next day is produced. The PX/ISO intervenes only in its role of market maker and allocates the cost of purchased reserve among customers. This can be done on a simple pro-quota basis or any other appropriate method.

Supply Bidding

The bidding to supply reserves is a transparent process. The tenders will contain amounts and prices of capacity at which spinning and non-spinning reserves will be made available during the next 24 hours. The bids will contain only a component for capacity. In case a reserve unit is effectively utilized, the energy consumed will be paid at the relevant spot price, either the hour-ahead or the real-time market, for the period during which the unit is kept in service. In theory, the energy price will be set by the energy bid of the marginal reserve unit of the system. Energy and capacity reserves bids should be submitted at the same time to avoid strategic gaming between those markets. Assuming competition among power producers and no market power exercise, the bids should reflect the marginal cost of making reserves available.

Demand Side

The PX/ISO represents the demand side in reserves markets and procures the service on behalf of customers, that is, the aggregate load demand. Individual choice is favored by allowing consumers to participate in the market as interruptible loads. In this way consumers can choose to self-provide reserves or sell them back. Interruptible loads are equivalent to generation capacity from the point of view of reserves. Thus, in case of a power shortage some load is voluntarily curtailed at a certain price, reducing the magnitude of the shortage. Interruptible loads can compete on equal terms with generation reserves, with the beneficial effect of enlarging the competitive base of the supply side. We have not discussed how the PX/ISO can establish the worth of reliability and establish the demand for reserves to confront with suppliers' bids. That is the subject of chapter 5 and the analysis will be postponed until then.

Sequential Markets

Alternative structures for the operation of competitive energy and capacity reserves markets, which may result in equally valid resource allocation, can have significantly diverse implications from the perspective of decision-making decentralization and individual profitability. One way to organize these markets is through the simultaneous determination of quantities and prices for each market. An alternative approach involves sequential market calculations. In *sequential markets* the energy market is cleared first and the results represent the starting point for the reserves market [4].

In the simultaneous approach, the system operator collects all the information concerning costs and characteristics of generating units. Then, based on some a priori determination of reserve requirements, it runs a cost-minimization, unit-commitment type program in order to determine a cost-effective allocation of generation resources. This approach achieves lowest system cost for a set of operating constraints, but it presents all the shortcomings of cost-effective and centralized models. In particular, the arbitrariness of the selected reserve requirements and the limitations imposed on individual choice. The owners of generation may not be willing to disclose all required cost information and be subject to the results of a mysterious program. They will want to be responsible for their price and output decisions in different markets, and they will want to take the chances of profitable opportunities and the risks of poor decisions.

In sequential markets competitive auctions are conducted for energy and operating reserve services [15]. Auctions have the advantage of being simple, easy to understand and perceived as fair contests. In day-ahead auctions, individual generators are allowed to bid different hourly prices for energy and capacity reserves. Energy and reserve bids are collected simultaneously, and suppliers may offer the same capacity in all markets. Units with the lowest bids are selected to balance supply and demand in the different markets. The energy market is cleared first, defining energy output, price and dispatched units. The capacity committed in generation and regulation services is withdrawn from the operating reserve supply and the market for reserves is cleared next. The result is a set of market-clearing prices and quantities for energy and reserves.

With sequential-market clearing, the markets for energy and operating reserve are closely coupled in supply. Units left to supply reserve are not dispatched in the energy market so no energy opportunity cost is incurred. In other words, energy production is not an alternative for units providing reserves. Furthermore, base load units are the first to be dispatched in the energy market and do not really belong to the reserves market.

"Real World" Examples

A few current systems apply market concepts to the determination of their operating reserve. However, a satisfactory method has not been developed yet. In general, the results obtained point out to low capacity prices during most of the time and extremely high prices when capacity is scarce. The first behavior is what is expected if there is enough installed capacity in the system, because reserve supply costs are low. The price peaks can be due to defects on market design and due to exercise of market power by power producers. For illustration, we outline below how reserve is allocated in a couple of real systems, UK and California.

United Kingdom

In UK a capacity payment is made to every MW declared available every half an hour. For no dispatched units, the payment is determined as the LOLP multiplied by the difference between a value of lost load (VLL) in £/kWh and the cost of generation determined by the unit bid price also in £/kWh. The VLL was established by regulators as £2/kWh in 1990 and is updated annually with the inflation index. The payment is included in the system price as uplift.

Capacity Payment (£/kWh) = LOPL x (VLL – Bid)

The capacity payments have been heavily criticized, raising to hundreds of pounds per MWh when capacity is scarce, with some consumers being forced to reschedule production when facing those prices. Besides market power issues, one concern is that the LOLP calculated may overstate the risk of outages in the system [16].

California

In California an auction-type reserve market has been established with supply side bidding. The required amount of operating reserve is established as a percentage of the expected hourly load [17]. This is equivalent to use a complete inelastic demand, as shown in figure 3.3, which do not convey any information about the economic value or reserve.

In addition, this favors strategic behavior of suppliers, who can raise their bids or withhold capacity and get higher prices. Actually, that seems to have happened, prices for replacement reserves reached \$750, \$2500 and \$5000/MW during few hours on July 9, 1998. California ISO decided not to buy replacement reserves on July 10, when prices could have reached \$9,999/MW, equal to the bidding software limit [18]. As a fix, California ISO capped prices at \$500/MW, decision approved by FERC on July 1998.

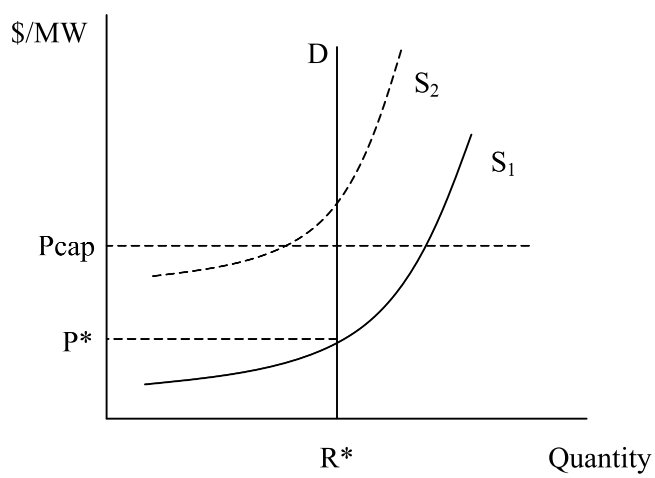


Fig. 3.3: Inelastic demand

Chapter 4

Supply and Demand

The supply in a competitive market of reserves reflects the cost to producer of making their generating capacity available. On the other hand, the demand reflects the benefits to consumers from having generation reserves in the system. In the latter case, consumer benefits are accrued as avoided interruption costs. In this chapter we analyze the components of supply and interruption costs.

4.1 Supply Costs

The aggregate supply in the proposed market framework is composed of separate bids that represent how much capacity producers will make available at different prices. The market supply curve is the horizontal sum of those bids. Producers will commit capacity as long as they get a price greater or equal to their marginal production costs. In addition, they will shut down operation if price stays consistently below average variable cost. Consequently, the aggregate supply curve represents the marginal cost schedule - above average variable cost - of providing capacity reserves.

Capacity reserves are provided mostly by generation units and to a less extent by interruptible loads. The costs associated with generation units are conventionally divided into capital, fuel and operation and maintenance (O&M) costs.

- The *capital costs* are investments in plant and equipment to provide physical infrastructure, industrial facilities, generation hardware and plant auxiliary services. Capital costs are associated with the construction of the generation plant and are incurred at the beginning of plant's operating lifetime.
- The *fuel costs* include purchase, transportation and storage of the source of energy. Fuel costs are directly related to generation of electric energy. They can be signifi-

cant for fossil-fueled units, lower for nuclear plants and can be positive for hydroelectric plants, considering the shadow price of water in reservoir.

- *Operation and Maintenance costs* are associated with the day to day operation of the plant. O&M costs include overhead costs - administration, rents, insurance, etc. - of running the plant and the cost of consumables and labor involved in operation and maintenance of the plant.

Cost Analysis

Capital costs are variable costs in the long run because they depend on the size of the plant. Once the nominal capacity is defined, capital costs become fixed costs, not depending on the actual output of the power plant. After the plant is built, capital costs are irrecoverable except for some salvage value if the facilities are dismantled. In effect, capital costs are sunk costs in the short run, irrelevant for future economic decisions. On the other hand, fuel costs are pure variable costs, accounting for most of the total variable cost. Overhead costs are fixed O&M costs, while variable O&M costs are related to equipment wear and tear.

The operating reserve is installed capacity committed to provide for unexpected plant outages. Since the commitment decision is part of the operating phase, capital costs are irrelevant. The costs of interest are, consequently, fuel and O&M costs. However, reserves are provided as pure capacity, that is, no energy generation occurs and no fuel costs are involved. Actually, there is some amount of energy consumed to provide for mechanical losses and plant auxiliary services. Next, we study differences in supply costs among types of reserves.

Non-Spinning Reserve

The non-spinning reserve contingent consists of fast-start units like gas-fired turbine and hydroelectric stations. As provider of non-spinning reserve, a generation unit is at a standstill. There are no mechanical losses involved and no equipment wear and tear. Operation is reduced to supervisory functions and only routine maintenance is carried out. Consequently, the marginal cost of providing non-spinning reserve is practically zero. Thus, there are no cost differential between committing, for instance, half or maximum unit output as non-spinning reserve.

Some O&M costs are indeed incurred; otherwise the plant would be shut down and closed. In this case, the capacity bid can be based on average cost, that is, O&M costs divided by the committed capacity. There is an incentive to bid full plant capacity, since costs can be spread out to yield a lower $ per MW bid. The costs incurred are low relative to fuel and full O&M costs. Accordingly, the bid price for non-spinning reserve should be significantly lower than same unit bid price for energy.

Spinning Reserve

Intuitively, spinning reserves have to cost more than non-spinning reserves. Otherwise there would be no need to employ non-spinning reserve. This is generally true, with a subtle distinction. There can be spinning reserve available either from the idle capacity of no fully dispatched units or from units that are on-line but not supplying any load. The first type of spinning reserve has zero marginal cost, because all costs are charged to energy sales. However, the total amount of this capacity in the system is limited, and it would be better used if committed to provide other “faster” ancillary services like frequency regulation or energy imbalance.

The second type of spinning reserve is capacity kept on-line, spinning at nominal frequency, ready to take load but without actually supplying energy. The spinning reserve group can, in theory, consist of any station in the system. In practice, base load units belong to the energy market. When providing spinning reserve, a generation unit is subject to mechanical losses, to equipment wear and tear and it drives some auxiliary systems (lubrication, cooling, etc.). Some fuel is used to generate enough energy to cover the mechanical losses and to feed auxiliary systems. Operation includes control and monitoring.

Evidently, the incremental cost (on the very first MW) of providing spinning reserves is higher than for non-spinning reserves. This is essentially a start-up cost that can be spread out over the whole MW output. The capacity bid price should reflect this start-up cost. Start-up costs are appreciably lower than fuel and O&M costs incurred when generating energy. Accordingly, the bid price to provide spinning reserves should be lower than same unit bid price for energy.

Reserves Price

In a competitive market, the price paid for capacity reserves should converge to supply cost. This cost is much lower than the cost per MWh generated, so one can expect average market prices for reserves to be lower than average prices for energy. Appreciable deviations from this rule can be due to monopolistic practices. Reference [4] suggests that the total amount paid for ancillary services prices may come to about 10% of the energy price.

It has been suggested that the real price of reserves is their energy opportunity cost, that is, the foregone opportunity of employing the committed capacity to generate electricity and sell it in the energy market. The assumption is that all capacity committed as operating reserves would have been dispatched otherwise. However, in reasonably competitive markets, a supplier would not take capacity out of the energy market to put it in the reserve market where the price that he can expect to be paid is appreciably lower. In effect, well-functioning power markets will allocate first the most efficient resources to energy generation and next to reserve provision or other ancillary services.

4.2 Outage Costs

In power systems, operating reserves are necessary to provide for unplanned plant outages that could result in power shortages. In general, when load demand exceeds available supply and no corrective action is taken, the system becomes unstable. A dynamic process is triggered, which starts with frequency fluctuations and power surges, followed by unpredictable tripping of and/or damage to generators and lines, finally culminating in widespread blackouts. Therefore, in order to protect the system and the customers against a cascading blackout, it is necessary to establish an ordered reduction of load. This process, called *load shedding*, helps to limit shortage costs by protecting the integrity of the system and especially by decreasing the number of customers who would suffer interruptions of supply.

Load Shedding

Every time load demand begins to exceed available supply, including generators' overload limits and maximum transfers from interconnected systems, a series of measures must be taken to reduce load. First, voluntary decreases or interruptible contracts would be called, followed by voltage and frequency reductions and disconnection of loads. The decrease of system voltage and frequency can reduce modest amounts of load, so at the end a power shortage will result in some load being cut off the system.

A rational load shedding scheme, that is, a satisfactory procedure to disconnect load and apportion energy curtailment among customers is one of the more difficult problems in power systems operation. In general, at certain aggregate levels, it would be possible to carry out an across the board curtailment, spreading the loss evenly among consumers. Another approach is to curtail always a group of customers. This may favor groups of customers over others, and involves policy decisions rather than technical problems.

An optimal procedure of load shedding should minimize total social cost. This implies to curtail first those customers that stand to lose the less from supply interruptions. Thus, for instance, residential consumers could be curtailed first, not having production losses to show. For the purposes of reserve allocation we will consider that load shedding would take place in an orderly way, in which load curtailment is apportioned to minimize social cost.

Cost Estimation

Power shortages result in interruptions of supply, which in turn translate into curtailments of energy to some customers. The effect of having capacity reserves in the system is to reduce the frequency and severity of said curtailments. Accordingly, the reliability benefits of generation reserves will appear as reduction in costs associated with interruptions of supply.

All the economic costs suffered by consumers when electricity is not available are known as *outage costs*. Outage costs are directly related to unserved energy. In effect, consumers are indifferent about where they get electricity from, and no outage costs are incurred when another source of power is available (e.g. a backup system), except those necessary to have access to the alternative source. In conclusion, the benefits of operating reserves are accrued to consumers as avoided outage costs due to reductions of unserved energy.

Direct outage costs result from the interruption of supply while indirect costs result from responses to that interruption. Thus, direct costs include lost production, opportunity costs of idle production factors (materials, labor, equipment), process restart costs, spoilage of materials and food, equipment damages, lack of transportation, foregone leisure activities, uncomfortable building temperatures and costs associated with increased risk to human health and safety. Indirect costs include operation of backup systems, overtime production to make up lost output, evacuation of buildings and civil disobedience during extended blackouts or no operation of safety devices.

There are two approaches to the evaluation of outage costs [3, 19]. One approach attempts to directly estimate the effects of outages on consumers' productive activities and behavior. The other is based on observed or estimated willingness-to-pay for increased reliability. The method proposed in this work is a variant of the willingness-to-pay approach.

Damage Function

In general, outage costs depend on the nature of activities of each customer and how they depend on electrical supply. Hence, methods of estimating outage costs are consumer specific. Customer characteristics include type, customer's activities, and size of operation, demand and energy requirements and demographic data. Moreover, outage costs also depend on the attitude and preparedness of customers, which in turn is related to current reliability levels, perceived as expected frequency and duration of interruptions. In fact, consumers progressively adapt their productive activities and way of life to given levels of reliability.

The direct estimation of outage costs requires detailed information concerning economic activity and outage effects. Usually, consumers are divided into groups or categories, which are considered to face similar outage costs, e.g. residential, industrial.

The methodologies commonly used infer costs from proxy variables or recur to customer surveys. Some measures used are lost revenue, foregone production, electricity value added and foregone leisure time. In surveys, customers are asked to estimate their losses due to supply interruptions of different duration and occurring at different times of the day and year.

The information retrieved is usually compiled in a customer damage function (CDF). The CDF displays outage costs expressed in kilowatts of annual peak demand ($/kW) for different customer categories and interruption durations. The CDF is aggregated for a particular load configuration and peak demand to produce an average estimation of dollars per kWh of unserved energy. The CDF estimation assumes that curtailments will be distributed proportionally across customer sectors. This method is straightforward to apply but based on severely limiting assumptions, and it focuses only on average, not incremental outage costs.

Willingness to Pay

This approach measures outage costs by the expected reduction in net social benefit due to interruptions of supply. Net social benefit from electricity consumption is defined as the difference between what a consumer is willing to pay for electricity and what the consumer actually pays for it. The net benefit from electricity consumption or *consumer surplus* can be easily calculated if the demand curve is known; it is the area below the demand curve and above the price line, as illustrated in figure 4.1.

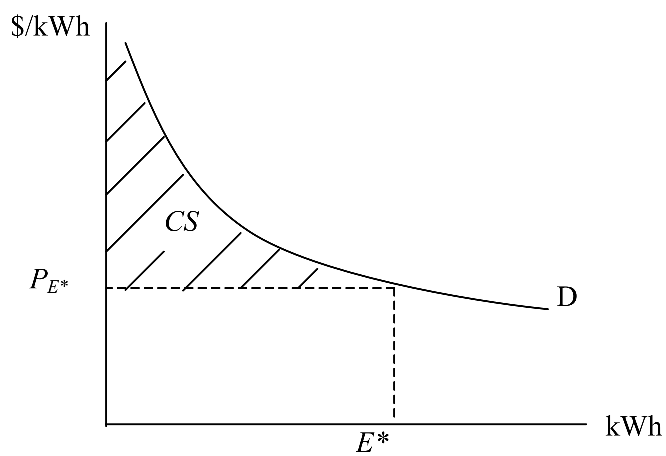


Fig. 4.1: Consumer surplus

The demand curve D expresses the willingness to pay WTP of consumers for incremental units of electricity. WTP measures benefits of incremental units of energy. A related concept, willingness to accept WTA, expresses how much consumers would accept for giving up a unit of electricity. WTA measures the cost to consumers of decreasing units of energy, and represents how much they should be compensated to accept reductions in consumption. Economic theory explains that the values of WTP and WTA should not differ.

The demand curve for energy represents, therefore, the WTA for decreasing units of electricity, beginning from the market equilibrium quantity E^*. This is equivalent to consumer costs associated with unserved energy. As explained before, we assume energy curtailments are carried out according to a least-cost criterion. That is, beginning with the marginal consumption, indicated as the portion of the demand curve between points A and B in Fig. 4.2. Consequently, the value of unserved energy is the area under the demand curve net of the savings realized by not paying the energy price. This area represents the lost surplus of marginal consumers, or area ABC in Fig. 4.2.

In conclusion, outage costs due to interruptions of supply are measured in terms of lost surplus of marginal consumption. The main advantage of the lost surplus approach is that outage costs are directly derived from consumers' preferences as revealed in the energy market, without recurring to arbitrary categorization and aggregation of customers and indirect calculations of their outage costs.

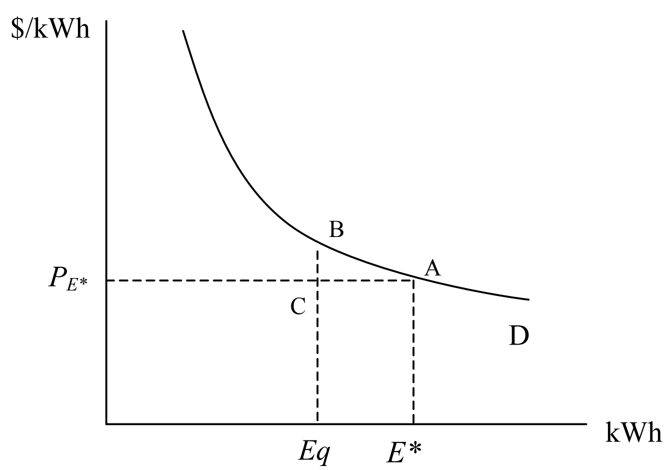


Fig. 4.2: Outage costs

The WTP approach is simple to apply and requires information readily available to the ISO/PX. In systems with demand-side bidding the ISO/PX can directly calculate hourly outage costs to customers due to unserved energy, beginning from the market-clearing point. The incremental outage cost for different levels of energy curtailment is the incremental lost surplus.

In systems with no demand-side bidding, a good approximation can be achieved using the market equilibrium information and an estimated price elasticity of demand, ε. For an isoelastic demand we can write the equation of the demand curve as:

$$D = \alpha . P^{\varepsilon} \text{, where } \alpha = E^{*}/(P_{E^{*}})^{\varepsilon} \text{ and } \varepsilon < 0 \tag{4.1}$$

With reference to Fig. 4.2, we can find area ABC using equation (4.1). The area is given by the following expression:

$$\Delta ABC = \int_{E_q}^{E^*} \left(\frac{\alpha^{1/|\varepsilon|}}{D^{1/|\varepsilon|}} - P_{E^*} \right) dD$$

$$\Delta ABC = \frac{\alpha^{1/|\varepsilon|}}{1/|\varepsilon| - 1} \cdot \left[\frac{1}{E_q^{(1/|\varepsilon|-1)}} - \frac{1}{E^{*(1/|\varepsilon|-1)}} \right] - P_{E^*} . (E^* - Eq) \tag{4.2}$$

Equation (4.2) is a good estimate for curtailments in the neighborhood of E^*, which is generally true because actual outages are expected to be small compared with total consumption.

Chapter 5

Reserves Value

In order to procure generation reserves on behalf of customers it is necessary to establish the reliability benefits of this service. In this chapter we introduce a valuation model to establish the worth of said benefits and to translate this worth into a demand for reserves.

First, we examine the sources of value of retaining additional capacity, and next, we present a model to establish the value of different levels of reserve in the system. Finally, we use the model to establish the incremental value of reserves and their demand.

5.1 Sources of Value

The value of capacity reserves is less intuitive than the value of energy. In effect, capacity is not a consumable good as is electrical energy, but the ability to produce this energy at a determined rate and with certain defined characteristics as cost, availability and time of response. Actually, in power markets scheduled ahead of time, energy is traded through a series of *forward contracts*, while capacity payments correspond to *option contracts* on energy [20].

Thus, in electricity 'spot' markets settled ahead for next-day hourly trades, the scheduled dispatch corresponds to a group of forward contracts to deliver energy at the specified market-clearing prices for each of the next twenty-four hours. The 'spot' price represents the forward price of the market for delivery at each hour. Likewise, payments made to available capacity represent premiums paid for energy option contracts, which can be called (dispatched) paying a specified amount.

Option Contracts

A *call option* contract in financial markets gives its holder the right to buy an underlying asset (e.g. stock) for a specified price, called the strike price, on or before the maturity of the contract. A trader 'writes' such a contract in return for a fixed fee or premium [21]. The premium reflects the expected value to the holder of being able to call the contract. A call option is exercised when the market value of the underlying asset is greater than the strike price. The difference between the underlying asset value and the strike price is the exercise value of the option.

Capacity contracts are call-like options on the energy (underlying asset) that can be generated during the relevant dispatch period if the capacity is used. The value of the underlying asset is the spot price for energy. The strike price of this call option can be a contractual energy price or generator's own energy bid price. In the latter case, the exercise value would be the difference between the spot price and the bid price, and payments for capacity could be calculated as premiums for financial options.

Option Value

Capacity reserves are called on because of generation outages and do not make part of a price hedging strategy [22]. As a result, the decision to call the reserve is not based on the difference between spot and strike prices, and there is no well-defined exercise value on 'financial' terms. Consequently, pricing methods for financial options (e.g. the Black-Scholes formula) are not applicable to assess reserve payments.

The option-like characteristics of capacity contracts provide useful insight into the value of operating reserves. In fact, a purchaser of reserves holds the option to buy the energy that can be generated using the 'locked' capacity. The decision to call the reserved capacity will depend on the occurrence of outages and on the actual deficit of energy confronted. Consequently, reserving capacity is rather a strategic option whose 'exercise' value is the worth of outage costs that are not incurred because the reserve is called on.

The value of reserves to customers is equivalent to the premium of holding the associated 'call' option. In general, reserves can be used totally, partially or not used at all. Since the 'exercise' value of calling reserves on is uncertain, the premium reflects the expected worth of the avoided outage costs.

Insurance Value

To understand the value of operating reserves is necessary to analyze the protection they provide against the risk of energy curtailments due to generation shortages. In theory, this risk could be removed through an insurance contract. With such a contract customers would be entitled to a financial compensation each time their load is interrupted. If customers are to be totally compensated for the economic losses incurred when their energy consumption is curtailed, the fair price of this insurance policy will be the expected value of those losses [23].

In practice, there are no insurance markets for power outages. The other way to hedge against the risk of curtailment is to procure operating reserves. In this case, instead of financial compensation of load interruptions, the physical risk of shortages is reduced by increasing the capacity available in the system. The stochastic nature of this risk and its evaluation were studied in chapter 2.

For this self-provided insurance, the customer will be willing to pay a premium (the fair price of risk) equivalent to the reduction in expected outage costs. The premium represents the worth of reserves for customers, which is different from purchase costs. The optimal level of reserves is determined when incremental reductions in expected outage costs are equal to incremental costs of acquiring reserves.

5.2 Valuation of Reserves

Payments for generation reserves should reflect the value of available fast-start capacity in terms of avoided outage costs. Accordingly, as a single purchaser in the proposed market allocation framework and in order to establish the worth of reserves in the system, the PX/ISO must simply evaluate the reduction on expected outage costs due to additions of reserve units. This is equivalent to calculate the reduction in expected lost surplus of marginal consumption. The mechanism is illustrated with the following example.

Example 5.1: Consider the following single customer – single producer system.

- A power producer owns enough installed capacity to meet the hourly demand of its only customer, which happens to be price elastic with elasticity ε = -0.5. Customer's hourly demand is given by the expression $D = 5000.P^{-0.5}$, where D is energy consumption in MWh and P the energy price in \$/MWh.

- Assume this single buyer – single seller market is in equilibrium at P = \$ 25/MWh, at the point where demand and generation (neglecting losses) are equal to 1,000 MWh. The market equilibrium and consumer surplus are shown in Fig. 5.1.

- To supply the hourly load demand, the 'utility' dispatches 800 MW using modern generating units which are available practically 100% of the time. For the remaining 200 MW an old, low-cost unit is used, with forced outage rate of 0.25.

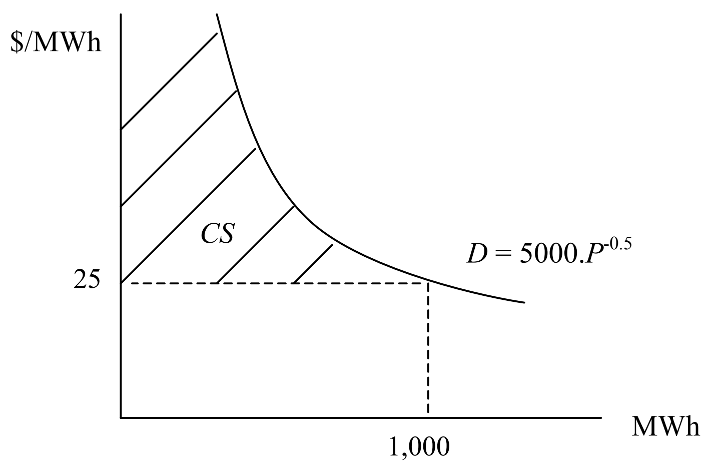


Fig. 5.1: Equilibrium and consumer surplus

- The customer realizes that on average, even though she contracted 1,000 MW each hour, she only gets 800 MW during 25% of the time. Although she does not pay for the unserved energy, she still loses the consumer surplus associated with the last 200 MWh of consumption.

- The 'utility' explains that outages in the older unit are beyond its control, and are due to unavoidable random failures in electromechanical systems. He proposes instead to increase the reliability of supply by keeping some higher-cost capacity in standby. This reserve can pick load when the old unit is out.

- The customer has to figure out how much she would be willing to pay for this service. She recognizes that, if the utility can remove the risk in the supply of the marginal 200 MWh, the maximum she would pay is the avoided lost surplus. The surplus loss, corresponding to the shaded area in Fig 5.2, can be evaluated using eq. (4.2) as follows:

$$SL = \int_{800}^{1000} \left(\frac{5000^2}{D^2} - 25 \right) dD = \frac{5000^2}{D} \Bigg|_{1000}^{800} - 25 \text{ x } 200 \approx \$\,6{,}250 - 5{,}000 = \$\,1{,}250$$

- Moreover, since the probability of using the reserve is 0.25, the expected value of lost surplus is *E[SL]* = 0.75 x \$0 + 0.25 x \$1,250 ≈ \$312. In conclusion, the customer would pay up to \$312 / 200 MW = \$1.56 per MW of reserve.

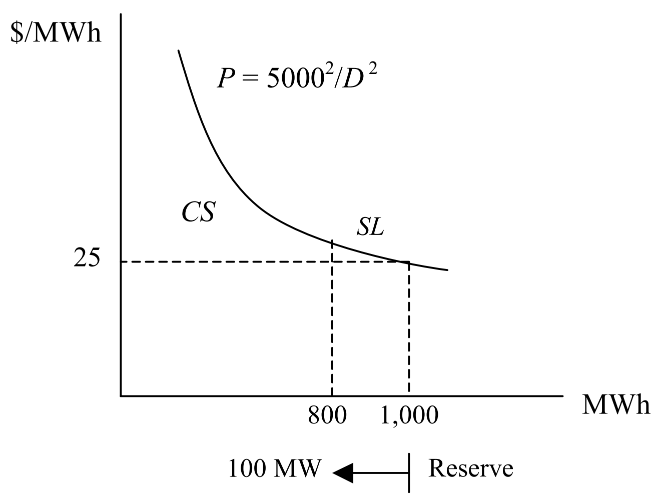


Fig. 5.2: Reserves evaluation

- Assume now that the cost of providing the 200 MW reserve in the system is about 5% the cost of generation, that is \$25 x 0.05 = \$1.25/MW-h. Therefore, the customer and the power producer will agree in some price between \$1.56 and \$1.25 per MW of reserve.

Valuation Model

The previous example contains all the basic elements of a valuation model for operating reserves. The worth to customers of additional capacity is the reduction in expected lost surplus of marginal consumption. All the information required is a risk model of generation shortages and the demand for energy, which is known in systems with demand-side bidding. Otherwise, the demand curve can be inferred using estimates of consumption price elasticity. This approximation should be good enough for the size of expected generation outages, which are small compared with the total dispatched generation.

Generation Risk Model

The generation risk model is the capacity outage probability distribution introduced in chapter 2. The parameter of interest is the long run probability of finding the unit in operation, or *unit availability*. Based on units' capacity and availability, a system operator can build the outage probability distribution for a given dispatch. The process can be easily implemented using a software routine.

The resulting outage distribution is a discrete probability distribution that shows the probability of finding different levels of capacity in the system. The first level is the nominal capacity, when all dispatched units are in, and the last is zero capacity, when all units are out. The risk model of discrete capacity states and associated probabilities can be expressed as follows:

$$\tilde{\boldsymbol{C}}_{\boldsymbol{A}} = \left\{ C_j , p_j \right\}, \ \ j = 0, 1, 2, \ldots l \quad \text{and} \quad l \leq 2^N - 1 \tag{5.1}$$

$\tilde{\boldsymbol{C}}_{\boldsymbol{A}}$ is a random variable corresponding to available generation in the system with N dispatched units. C_j represents the available capacity of the j-th level and p_j the associated probability. There are l+1 different capacity levels, but in practice p_j decreases rapidly and only the first levels are relevant.

Expected Surplus Loss

The price of electricity P_E and the expected demand load L_D for a dispatch period, say 1 hour, are known once the energy market is cleared. This information can be combined with the available capacity to establish the expected surplus loss.

With reference to Fig. 5.3 and for the initial dispatch without reserves, the capacity committed C_0 is equal to the expected load L_D. For an outage reducing available capacity to C_j, the corresponding generation shortage is L_D -C_j, with probability p_j.

The energy not supplied is (L_D - C_j) x 1h and the associated surplus SL_j is the area UVW below the demand curve D. Evidently SL_0 = \$0.

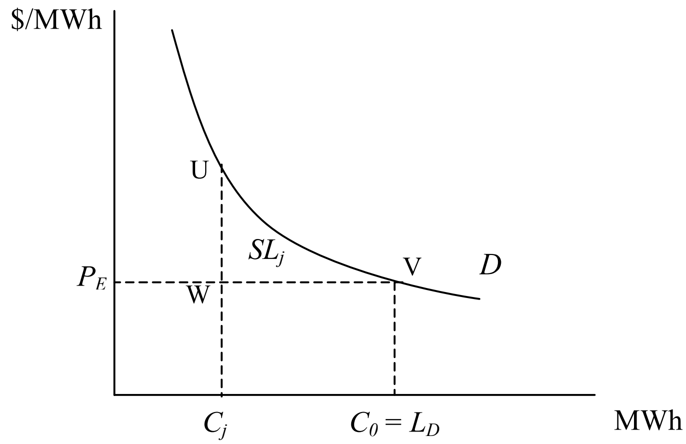


Fig. 5.3: Surplus loss

Reserves Worth

The total monetary exposure of the aggregate consumption, for a dispatch without generation reserves, is the expected surplus loss due to load interruptions. The expected loss is given by eq. (5.2) and represents the fair value or premium for an insurance contract that would remove all the economic risk due to energy curtailments.

$$E[SL] = \sum_i p_i \,.\, SL_i \tag{5.2}$$

Such a financial hedging strategy, even if available, would be extremely complicated to apportion among individual customers. We are interested, instead, in reductions of curtailment risk provided by additions of fast-response capacity. In effect, every time a unit of reserve is added to the system, the probability of supply shortages decreases and the expected surplus loss is reduced. This reduction on expected loss determines the worth of added capacity and, consequently, the value of reserves.

To assess the value of reserves, a market maker simply adds units of reserve to the dispatch following the 'merit' cost order derived from supplier bids. He finds the new expected surplus loss for each unit addition and takes the reduction ΔSL as the worth of the added capacity. The successive reductions on ΔSL, due to increments on capacity, can be added up to build a curve representing the value of different levels of reserve.

In order to simplify calculations, the 'merit' order of reserve units can be modified to reflect not the bid capacity, which is itself subject to unit availability, but to reflect the expected capacity available from each unit. The available capacity is the bid capacity times its availability. In this way, the added reserve is not subject to outages and will remove all curtailment risk for generation shortages less or equal to the available capacity.

Now the added value ΔV_j of additional capacity ΔR_j, covering shortages from C_{j-1} to C_j, is simply the expected reduction of surplus loss. The value of total reserves R_j is the sum of all ΔV_j. See Fig 5.4.

$$\Delta V_j = p_j \,.\, \Delta SL_j \tag{5.3}$$

$$V(R_j) = \sum_{i<j} p_i \,.\, \Delta LS_i \tag{5.4}$$

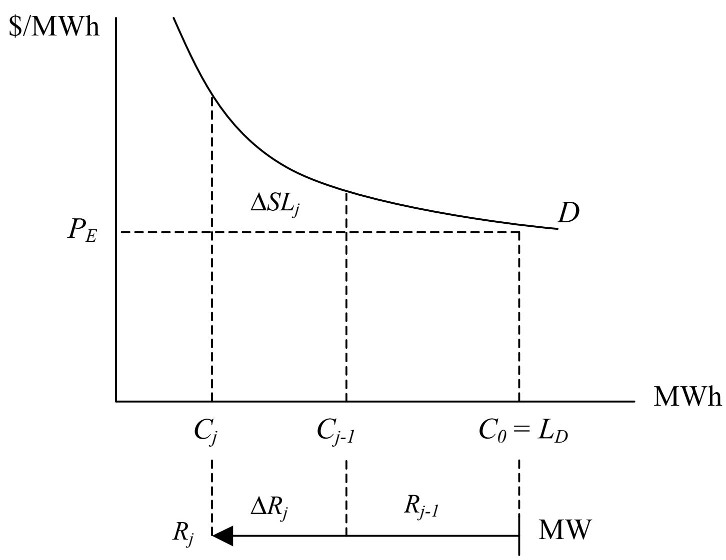


Fig. 5.4: Added value of reserves

Example 5.2: Consider the six-unit system of example 2.1 and an hourly, price-sensitive load described by the curve D=5000.$P^{-0.5}$, where P is the price of electricity in \$/MWh. We want to know the value of reserves for this system.

- The market equilibrium, where supply equals demand, is assumed to be at P_E = \$25/MWh and L_D=1,000 MWx1h.
- With reference to Fig. 5.4 and using eq. (4.2), the value of additional capacity is calculated as follows.

$$\Delta V_j = p_j \cdot \int_{C_j}^{C_{j-1}} \left(\frac{5000^2}{D^2} - 25 \right) dD = p_j \cdot \left\{ 5000^2 \left[\frac{1}{C_j} - \frac{1}{C_{j-1}} \right] - 25 \text{ x } (C_j - C_{j-1}) \right\} \quad (5.5)$$

- The supply shortage risk model is the same of example 2.1. The different capacity states, probabilities and reserve calculations are tabulated in Table 5.1.

Table 5.1 – Reserve Evaluation

J	Available Capacity C_j (MW)	State probability $p_j = P[C_j]$	Shortage & Reserves L_D-C_j (MW)	Incremental Surplus Loss ΔSL_j (\$)	Added Value ΔV_j (\$)	Reserves Value $V(R_j)$
0	1000	0.73509	0	0	0	0
1	900	0.11607	100	278	32.3	32.3
2	800	0.08349	200	972	81.2	113.5
3	700	0.05101	300	1,964	100.2	213.5
4	600	0.00879	400	3,452	30.3	243.8
5	500	0.00473	500	5,833	27.6	271.4
6	400	0.00067	600	10,000	6.7	278.1
7	300	0.00014	700	18,333	2.6	280.7
8	200	0.00002	800	39,167	0.8	281.5

- The value of operating reserves in this system is shown in Fig. 5.5. The worth of reserves to customers rises rapidly for the first MW of additional capacity, up to around 300 MW. After that level there are diminishing benefits from additions of generation reserves.

5.3 Demand for Reserves

To establish a market of operating reserves is necessary to determine the demand of customers for fast-response capacity. The demand is confronted with the supply curve derived from generators' bids and the market is cleared where both curves intersect. The market equilibrium defines the optimal level of reserves and their price.

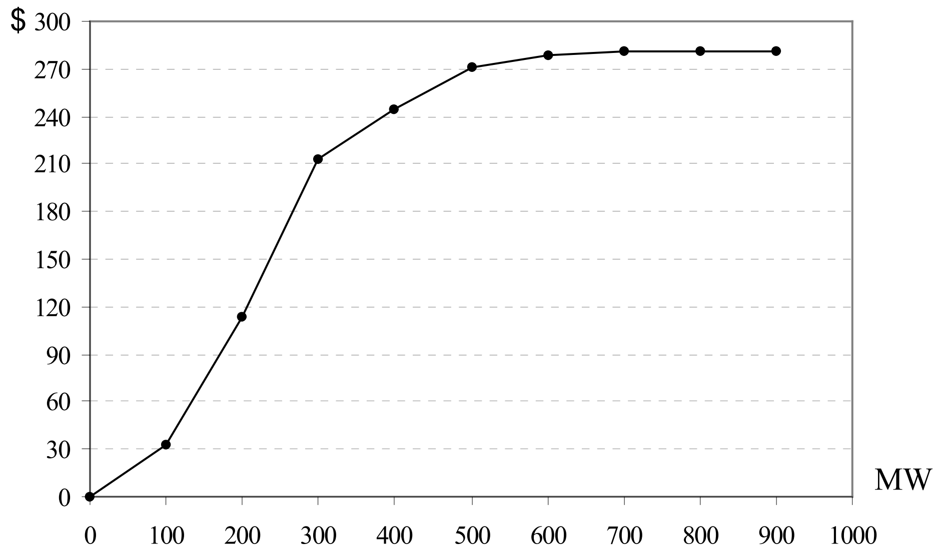


Fig. 5.5: The worth of reserves

The demand tells how much customers are willing to pay for generation reserves and reflects the reliability benefits of said reserves. The benefits are accrued as reductions in expected surplus loss due to load interruptions, and by definition, the demand represents marginal benefits to consumers. Accordingly, the curve of demand indicates the incremental value of incremental units of available capacity. Using eq. (5.3) we obtain the following expression for the marginal value (demand) of reserves:

$$D_j = MVj = \Delta V_j \,/\, \Delta R_j \;\rightarrow\; D_j = p_j \,.\, \Delta V_j \,/\, \Delta R_j \qquad (5.6)$$

Example 5.3: We want to find out the demand for operating reserves in the system of example 5.2. The demand can be determined directly from Table 5.1 or using eq. (5.6), and from the fact that the incremental additions of capacity reserves ΔR_j are constant and equal to 100 MW.

- Using Table 5.2, the incremental value of reserves is equal to column 6 divided by 100 MW.
- Using eq. (5.5), (5.6) and the relation $R_j = L_D - C_j$, we obtain the following expression for reserves demand:

$$D_j = \Delta V_j / \Delta R_j = p_j \cdot \left[\frac{5000^2}{(L_D - R_j)(L_D - R_{j-1})} - 25 \right] \quad (\$/\text{MW}) \qquad (5.7)$$

- Either way we get the following demand schedule:

Table 5.2 – Reserves Demand

R_J (MW)	100	200	300	400	500	600	700	800	900
D_J (\$/MW)	0.323	0.812	1.002	0.303	0.276	0.067	0.026	0.008	0.000

- The curve of demand is shown in Figure 5.6. The curve initially increases because the first MWs of reserve are highly valued. The increment in avoided surplus loss is the predominant factor. Past certain point the curve becomes strictly decreasing, additional reserves are less valuable because the probability of using them becomes very small.

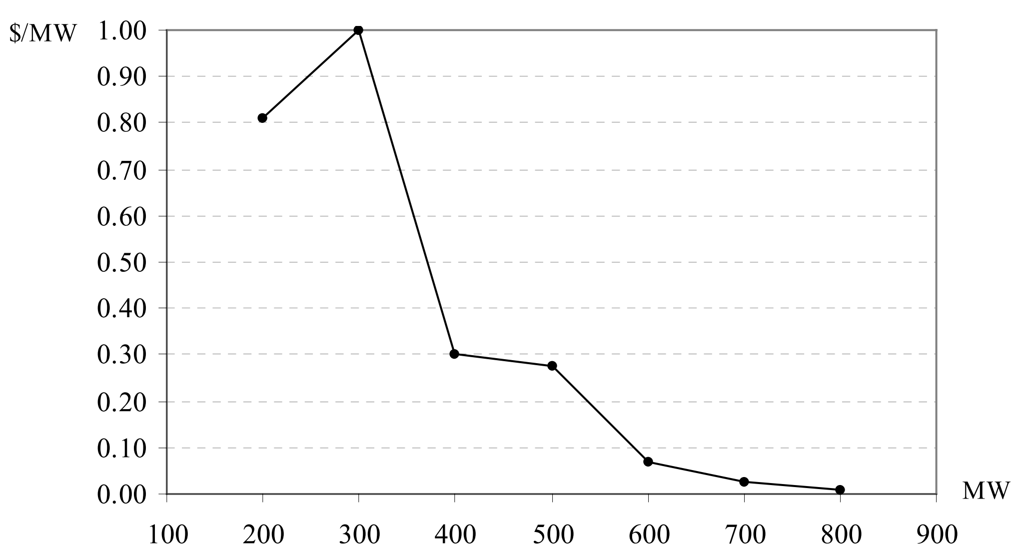


Fig. 5.6: Demand for reserves
Fig. 5.6: Demand for reserves

5.4 Market of Reserves

A well-functioning market of reserves is essential for the successful operation of modern power markets. It ensures not only the efficient provision of a basic reliability service but brings other beneficial effects by providing another source of revenue and by reducing speculation among generators.

Cost Recovery

A main issues in competitive markets for electricity is the uncertainty in generation cost recovery. Marginal cost pricing ensures allocation efficiency and recovery of generation variable costs, but not necessarily recovery of fixed costs. In theory, producer surplus in energy sales would yield enough revenue to recover all costs.

However, base load units still may present some economies of scale. In addition, the intermittence of dispatch is also important. Peaking units that are dispatched few hours each year need high prices to recover fixed costs in a short period of time. Similar situation exists regarding start-up costs.

The stream of payments received for capacity reservation helps to recover fixed and start-up costs. For marginal units they represent an additional source of revenue besides energy sales. The effect is that marginal cost bidding becomes a more acceptable strategy for generators.

Strategic Behavior

When capacity is scarce, poor market designs allow generators to speculate in energy and capacity markets, withdrawing capacity and manipulating bids to get higher prices. Single-auction bidding for energy and reserves and sequential market clearing reduce strategic opportunities.

A single energy-capacity bid price limits speculation. First, a unit should not be allowed to bid only reserves, since reserves that could not be dispatched in a contingency are worthless. In addition, from the point of view of costs, there is no difference between a unit scheduled to dispatch in advance or called upon a contingency.

Price Caps

In absence of markets for capacity, reserves could be created paying the full market price for energy without actually consuming that energy. Generators would be indifferent because they would receive at least its bid price with a high probability of saving fuel costs. Consequently, for arbitrage reasons, the maximum price that should be paid for reserves is the energy bid price of the marginal unit reserved.

The rationale is that, in case the reserve were purchased as energy in the primary market, the marginal reserve unit would be the marginal unit in the dispatch, establishing a market clearing price equal to its bid. This 'natural' price cap will adjust itself, hour by hour, to the conditions of the system, and can be used to avoid speculation in reserves prices.

Conclusion

This work challenges the widespread practice of defining reserve requirements in power systems operation based on deterministic criteria. These quantity-constrained methods rely on an arbitrary selection of adequate levels of reserve and do not address economic efficiency beyond some cost minimization. Moreover, when these methods are applied in competitive power markets, they could provide perverse incentives to power producers.

The proposed market-based framework for allocation and pricing of operating reserves is economically efficient, and fosters individual choice and competition in power markets. In addition, it reduces opportunities for strategic behavior and provides a simple mechanism to determine reserves and their trading price.

Integral to the proposed market framework, we have introduced a model to asses the economic value of reserves, which allows a market maker or system operator to organize markets of reserves in coordination with energy markets. The valuation model is based on the probabilistic assessment of the economic risk faced by customers due to power outages.

Similar value-based methods could be developed for installed capacity reserves and other generator-provided ancillary services. Further research in this and related fields is encouraged.

Annex 1

Reliability Models

A1.1 Exponential Distribution

A random variable $\tilde{X}$ is said to be exponentially distributed when it has the following probability density function, where λ is a positive constant:

$$f(x) = \lambda . e^{-\lambda x} \tag{A1.1}$$

The corresponding cumulative probability distribution and mean value of $\tilde{X}$ are:

$$P[\tilde{X} \leq x] = F(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} \tag{A1.2}$$

$$E(\tilde{X}) = \int_0^\infty x e^{-\lambda x} dx = {}^1\!/_\lambda \tag{A1.3}$$

If X denotes the up time of a component, then λ is the reciprocal of its mean up time and is known as the component failure rate. $F(t)$ represents the failure distribution of the component and its reliability is defined as $R(t) = 1 - F(t)$. The key property of the exponential distribution is that, at any time t, the probability of failure in the next instant is always equal during the entire operating period, that is, the failure rate is constant. This means the component does not age.

$$P \text{ [failure between } t \text{ and } t + \Delta t \text{ / component operating at } t] = \lambda . \Delta t \tag{A1.4}$$

This result is the usual assumption of constant failure rate in the flat portion of a "bathtub" hazard function. The exponential distribution is a good description for the up time of an electromechanical component over their useful life. When the down time of a repairable component is also assumed to be exponentially distributed, μ designates its repair rate and it is the reciprocal of the mean down time.

A1.2 State Space Approach for Reliability Calculations

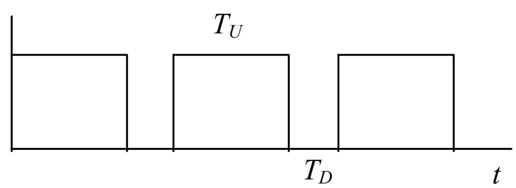
The state space approach is a powerful method for reliability calculations. It is based on Markovian stochastic processes and can be applied to independent components as well as the entire system. The steps of the approach are as follows:

- Enumerate system states and identify states associated with system failure.
- Determine interstate transition rates. The transition rate from state i to j is the rate of the system passing from state i to j.
- Calculate state probabilities and overall system failure probability.

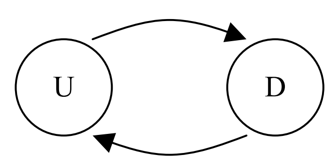
The state space representation is applied next to a repairable component.

Two-State Model

The useful lifetime of a repairable component is described as cycles of service and repair periods whose duration are considered random variables. The process consists of alternating "up" and "down" periods, T_U and T_D. The state space diagram in Fig. A1.1 shows the up (U) and down (D) states and the transitions between them. Perfect repair is assumed so the cycles are repeated during the whole useful lifetime.



(a) Life history



(b) State diagram

Fig. A1.1: Repairable component cycle

The component's life history is determined by the probability distributions $f_U(t)$ and f_D (t). Where $f_U(t)$ is the density function of up times T_U, and $f_D(t)$ is the density function of down times T_D. If $\boldsymbol{X_t}$ is the state of the component at time t, then the following definitions apply:

- Probability of being in the up state: $P_U(t) = P\,[\text{up at } t] = P\,[\boldsymbol{X_t} = U]$

- Probability of being in the down state: $P_D(t) = P\,[\text{down at } t] = P\,[\boldsymbol{X_t} = D]$

- Mean up time or mean time to failure MTTF, m, and mean downtime or mean time to repair MTTR, r:

$$m = \int_0^\infty t.f_U(t)\,dt \qquad \text{and} \qquad r = \int_0^\infty t.f_D(t)\,dt \tag{A1.5}$$

- The mean time between failures MTBF or mean cycle time is $T = m + r$, the cycle frequency is $f = 1/T$.
- Component Availability, A, and Component Unavailability, U:

$$A = \frac{m}{m+r} \qquad \text{and} \qquad U = \frac{r}{m+r} \tag{A1.6}$$

In the general case both $f_U(t)$ and f_D (t) can have any form and calculations require numerical methods. When both up and down times are exponentially distributed, with failure rate λ and repair rate μ, solutions can be obtained for the state probabilities $P_U(t)$ and $P_D(t)$ by solving a two-state Markov stochastic process.

$$P_U(t) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} \tag{A1.7}$$

$$P_D(t) = \frac{\lambda}{\lambda+\mu} - \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} \tag{A1.8}$$

These solutions assume the component is working at $t = 0$. Using eq. (A1.3) is easy to write availability and unavailability indices in terms of λ and μ:

$$A = \frac{\mu}{\lambda+\mu} \qquad \text{and} \qquad U = \frac{\lambda}{\lambda+\mu} \tag{A1.9}$$

Comparing eq. (A1.9) with (A1.7) and (A1.8) we see that $P_U(t) \to A$ and $P_D(t) \to U$ as $t \to \infty$. Consequently, A and U represent the long-run probabilities of finding a unit in the up and down state respectively. This is true for any distribution of T_U and T_D.

A1.3 Two-State Markov Process

A Markov process is a special kind of stochastic process in which the state probabilities at a future instant, given the present state of the process, do not depend on the states occupied in the past. A Markov process is memoryless, that is, the probabilities of a random variable at t_{n+1} depend on the value of the random variable at t_n but not on previous values. The process is illustrated with a two-state model.

Consider a component that can be in two states, either working (up, $X=0$) or failed (down, $X=1$). After failures, the component is repaired and returned to service. The states and the possible transitions are shown in Fig. A1.2. The probability of being in one particular state at $t+\Delta t$ depends on the state the component is at time t, not on the states previously assumed. Accordingly, the process is Markovian.

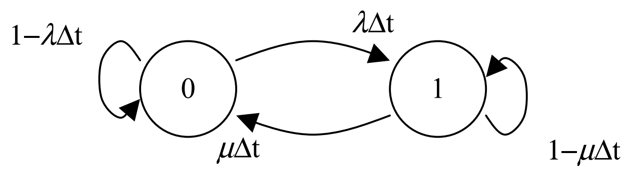


Fig A1.2: Two state Markov process

Assuming the component up and down times, T_0 and T_1, are exponentially distributed, with failure and repair rates λ and μ respectively, we can use eq. (A1.4) to find the transition probabilities p_{ij} from state i to j. Accordingly, $p_{01} = \lambda.\Delta t$ and $p_{10} = \mu.\Delta t$. Also p_{00}=1- $\lambda.\Delta t$ and p_{11}=1-$\mu.\Delta t$, as indicated in Fig. A1.2.

- The state probabilities at $t+\Delta t$ are calculated as follows:

  $p_0(t+\Delta t)= p_0(t).(1-\lambda\Delta t) + p_1(t).\mu\Delta t$, so $p_0(t+\Delta t) - p_0(t)= [-p_0(t).\lambda + p_1(t).\mu].\Delta t$

  $p_1(t+\Delta t)= p_0(t).\lambda\Delta t + p_1(t).(1-\mu\Delta t)$, so $p_1(t+\Delta t) - p_1(t)= [p_0(t).\lambda - p_1(t).\mu].\Delta t$

- Dividing by Δt and letting $\Delta t \rightarrow \infty$ we obtain:

  $p_0'(t)= - \lambda p_0(t) + \mu p_1(t)$ and $p_1'(t)= \lambda p_0(t) - \mu p_1(t)$

- Finally, defining $\mathbf{p}(t)$=[$p_0(t)$ $p_1(t)$], we can write the equations in matrix form:

$$\mathbf{p}'(t) = \mathbf{p}(t) \bullet \mathbf{A} \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} \qquad \text{(A1.10)}$$

A is called the transition intensity matrix. Solving eq. (A1.10) with the initial conditions $\mathbf{p}(0)$ = [1 0] we obtain equations (A1.7) and (A1.8) for $p_0(t)$ and $p_1(t)$ respectively, which are the solutions for the two-state model.

Annex 2

Glossary

Ancillary Services: Services necessary to support the generation and transmission of energy from generating resources to customers while maintaining reliable operation of the system.

Capacity: A measure used to define the maximum rate of electrical energy output of a generating unit, utility or system. Capacity is expressed in units of electrical power, usually megawatts (MW).

Competitive Pricing: Pricing based on competitive markets. The price reflects the actual cost of producing the incremental (or marginal) unit of power rather than the average cost.

Consumer Surplus: Difference between the total value consumers receive from the purchase of electric energy, as described by their demand curve, and the actual amount they pay for it.

Cost of Service Regulation: A pricing method where electricity rates are set so that revenues from electricity sales cover supply costs plus a fair return on invested capital.

Contingency: An unexpected event, usually the loss of one or more elements in the system. Applied to unexpected failure or outage of a system component (generator, transmission line, etc.).

Demand: Rate at which electric energy is delivered to or by a system or piece of equipment, at a given instant or averaged over a designated period.

Deregulation: Reduction or elimination of a regulation from a previously regulated industry.

Distribution: Process of transforming high-voltage electricity to lower voltages and then physically delivering it to consumers.

Elasticity (Price): Measure of consumer's demand variation due to changes in price. Expressed as the ratio of percent change in demand to percent change in price.

Electric energy: Generation or consumption of electric power over a period of time, usually expressed in megawatt-hours (MWh).

Forced Outage: Removal of a unit from service due to component failure, improper operation or human error. Forced outages require immediate removal of the unit.

Forced Outage Rate: Fraction of time for which a generating unit is required but cannot be in service due to an unplanned event.

Forward Contracts: Contractual agreement to exchange a financial asset at the maturity date and at a prespecified forward price.

Generation: The process of producing electric energy from other forms of energy.

Installed Capacity Reserve: Additional generating capacity that must be planned and built in to provide for load growth uncertainty and foreseeable outages.

Interruptible Load: Amount of customer demand that, according to previous arrangements, can be interrupted by direct control or request of the system operator.

ISO- Independent System Operator: An independent operator responsible for the coordination of the physical supply of electricity and for maintaining reliability throughout a system. The "independence" corresponds to the separation of physical operation from merchant functions in competitive models.

Load (Electric): Amount of electric power delivered or required at any specific point or summed up over the entire system.

Load Shedding: Deliberate removal of specific loads from the system in order to protect it during power supply shortages.

Marginal Cost: Incremental cost required to produce one additional unit of output (or reduction in cost from producing one unit less). In power markets is the cost to generators of providing the next MWh of electricity (or the next MW of capacity).

Market: Mechanism by which sellers and buyers interact to buy and sell goods and services, e.g. markets for electricity, capacity, reserves, etc.

Non-Spinning Reserve: Quick-start operating reserve.

Operating Reserve: Consists of on-line (spinning) and quick-start (non-spinning) reserves of generating capacity, which can rapidly pick up large blocks of load if needed.

Option Contracts: Contractual agreement that gives the holder the right to buy (call) or sell (put) a fixed quantity of a financial asset or commodity at a fixed price, within a specified period of time.

Outages: Non availability of electric power to the customer.

Outage Costs: Economic costs resulting from interruptions of power supply. They include, besides direct costs, the indirect costs of actions taken to avoid anticipated outages.

Peak Load: Electric load that corresponds to a maximum level of electric demand in a specified time period, often expressed in MW.

Scheduled Outage: Removal of a unit from service to perform work on specific components that is scheduled in advance and has a predetermined duration (e.g. nuclear refueling, annual overhaul, inspection, etc.).

Performance-based Regulation: Any pricing mechanism that links profits to desired results or targets.

PX - Power Exchange: A market place for the trading of electric energy and other power-related services.

Reliability: Extent to which electric supply is available to customers within accepted standards of voltage and frequency. Reliability is measured by the continuity of the service and voltage and frequency stability about nominal values.

Reserve (Capacity): Available generating capacity above the expected peak load.

Spinning Reserve: On-line operating reserve.

Spot Market: A market in which goods are traded for immediate delivery.

Transmission: Process of conducting the flow of electricity at high voltages from the points of generation to points of distribution.

Utility (Electric): Company or government agency with a monopoly franchise to sell electric energy to end-use customers.

References

[1] Endrenyi, J. 1978. *Reliability Modeling in Electric Power Systems*. John Wiley & Sons, New York, NY.

[2] Billinton, R., Allan, R., and Salvaderi, L. (editors). 1991. *Applied Reliability Assessment in Electric Power Systems*. IEEE Press, New York, NY.

[3] Billinton, R. and Allan, R. 1996. *Reliability Evaluation of Power Systems*. Plenum Press, New York, NY.

[4] Hirst, E. and Kirby, B. “Ancillary Services: The Neglected Feature of Bulk Power Markets”. *Electricity Journal*, Vol. 11, No. 3, Elsevier Science, April 1998, pp. 50-57.

[5] Federal Energy Regulatory Commission. *Promoting Wholesale Competition Through Open Access Non Discriminatory Transmission Services*. Order No. 888, April 1996.

[6] North American Electric Reliability Council. *NERC Operating Manual*. Princeton, NJ, Dec. 1997. Cited in [7].

[7] Hirst, E. and Kirby, B. “Technical and Market Issues for Operating Reserve”. *Electricity Journal*, Vol. 12, No.2, Elsevier Science, March 1999, pp. 36-48.

[8] Allan, R. 1989. “Concepts of Data for Assessing the Reliability of Composite Systems”. IEEE Tutorial Course - *Reliability Assessment of Composite Generation and Transmission Systems*, IEEE Power Engineering Society. IEEE Publishing Services, New York, NY, pp. 14-20.

[9] Telson, M. 1973. *The Economics of Reliability for Electric Generation Systems*. MIT Energy Laboratory, Report No. MIT-EL 73-016, Cambridge, MA.

[10] Knight, U. 1972. *Power Systems Engineering and Mathematics*. Pergamon Press, Oxford.

[11] Ilic, M. and Zaborsky, J. 1999. *Dynamics and Control of the Large Electric Power Systems*. Lecture notes of Course 6.686, Advanced Power Systems, MIT, Cambridge, MA.

[12] Ilic, M. and Galiana, F. 1998. "Power Systems Operation: Old vs. New". *Power Systems Restructuring, Engineering and Economics*. Ed. by M. Ilic, F. Galiana, L. Fink. Kluwer Academic Publishers, Boston, MA, pp. 15-107.

[13] Rosen, H. 1995. *Public Finance*. Irwin/McGraw-Hill, Boston, MA.

[14] Pindyck, R. and Rubinfeld, D. 1995. *Microeconomics*. Prentice Hall, Upper Saddle River, NJ

[15] Kirsch, L. and Rajaraman, R. "Profiting from Operating Reserves". *Electricity Journal*, Vol. 11, No.2, Elsevier Science, March 1998, pp. 40-49.

[16] Green, R. 1998. "The Political Economy of the Pool". *Power Systems Restructuring, Engineering and Economics*. Ed. by M. Ilic, F. Galiana, L. Fink. Kluwer Academic Publishers, Boston, MA, pp. 131-166.

[17] California Independent System Operator. *Ancillary Services Requirements Protocol.* Posted at http://www.caiso.com/ , April 1998.

[18] *The Electricity Daily*. "FERC Gives California ISO Authority to Cap Prices". Vol. 11, No. 14, Elsevier Science Inc., July 21, 1998.

[19] Munasinghe, M. 1979. *The Economics of Power System Reliability and Planning*. The World Bank, John Hopkins University Press, Baltimore, MD.

[20] Hunt, S. and Shuttleworth, G. 1996. *Competition and Choice in Electricity*. John Wiley & Sons, Chichester, England.

[21] Brealey, R. and Myers, S. 1996. Principles of Corporate Finance. McGraw-Hill, New York, NY.

[22] Singh, H. and Papalexopoulos, A. 1998. *Competitive Procurement of Ancillary Services by an Independent System Operator*. IEEE Transactions on Power Systems. PE-427-PWRS-0-06-1998.

[23] Borch, K. 1990. *Economics of Insurance*. North-Holland, Amsterdam.

[24] Singh, C. 1989. "Basic Probability and Reliability Concepts". IEEE Tutorial Course - *Reliability Assessment of Composite Generation and Transmission Systems*, IEEE Power Engineering Society. IEEE Publishing Services, New York, NY, pp. 9-13.

[25] Apostolakis, G. 1999. *Elements of Reliability Theory and Probabilistic Risk Assessment*. Lecture notes of Course 6.938, Engineering Risk-Benefit Analysis, MIT, Cambridge, MA.

[26] Ringlee, R. 1989. "Thoughts on the Applications of Composite System Reliability Methods – Cost Benefit Analysis". IEEE Tutorial Course - *Reliability Assessment of Composite Generation and Transmission Systems*, IEEE Power Engineering Society. IEEE Publishing Services, New York, NY, pp. 56-61.

[27] Gromb, D. 1998. *Finance Theory*. Lecture notes of Course 15.415, Sloan School of Management, MIT, Cambridge, MA.

[28] Graves, F., Grant, E. et al. 1998. "One-Part Markets for Electric Power: Ensuring the Benefits of Competition". *Power Systems Restructuring, Engineering and Economics*. Ed. by M. Ilic, F. Galiana, L. Fink. Kluwer Academic Publishers, Boston, MA, pp. 243-280.

[29] Gibescu, M. and Liu, C. 1998. *Optimization of Ancillary Services for System Security*. Proceedings of the Bulk Power System Dynamics and Control IV - Restructuring, August 24-28, Santorini, Greece.

[30] Hogan, W., Cadwalader, M. et al. 1998. *Reliability, Scheduling Markets, and Electricity Pricing*. Harvard Electricity Policy Group, Cambridge, MA.

[31] Energy Information Administration. *Electricity Prices in a Competitive Environment*. DOE/EIA 0614, Washington D.C., August 1997.

[32] Zobian, Assef. 1995. *A Framework for Pricing Transmission and Ancillary Services in Competitive Electric Power Markets*. S.M. Thesis, Technology and Policy Program, MIT, Cambridge, MA.