# A Model for Selecting Minimum Cost Combinations of Automatic Data Processing Equipment

By Martin E. Abel and Hyman Weingarten

**The practical problem of getting computational work done at the lowest cost subject to certain specifications interests many people. This paper presents a decision model which will enable them to choose, from among the alternatives available, the combination of computing equipment that minimizes computing costs for a given amount and composition of work. This model, together with other considerations, can provide a systematic basis for selecting a machine complex for a data processing installation. The model is useful also in deciding whether to change a machine complex as a result of changes in the workload or the availability of improved equipment.** **The authors are indebted to Lee M. Day and Burton French of Farm Production Economics Division, Economic Research Service, for helpful suggestions in the preparation of this paper.**

THE ADVENT of electronic data-processing equipment has revolutionized our thinking about the number and variety of quantitative problems that can be solved. Problems that once took weeks, months, or years to solve can now be solved in seconds, minutes, or hours. Further, this equipment makes possible a higher degree of accuracy in computations.

There has been a veritable explosion in the sales of electronic data-processing equipment in recent years. Many valid reasons lie behind this enormous increase in sales. Greater efficiency and economies in numerous computational operations have been achieved. But many purchasers of these machines also have found that owning computing equipment is more costly than they expected, because "instant decision making, major cutbacks in personnel and programming for the layman are at present and for the next few years, fanciful pipedreams. . . . Overbuying . . . has become increasingly common. The additional cost for faster, larger memories, added tape drives, and other peripheral equipment may not seem significant at the time of contract signing. . . . However, much of this added potential may, in fact, be outdated and uneconomical when the organization eventually finds adequate use for it and, of course, initial problems and cost of installation are magnified." [1]

## The Cost Problem

The cost of getting a data-processing job done has many dimensions. An obvious one is the dollar price that must be paid for the computational work. However, at times other factors may be more important than the dollar cost. Among these factors are (1) accuracy of computations and (2) computational time. For example, accuracy and speed in data processing are of paramount importance to a space program concerned with orbiting a man about the earth. Voluminous amounts of data must be accurately processed in seconds or minutes so that the results can be used to make critical decisions that may determine the success or failure of the venture. In this instance, technical considerations override cost.

On the other hand, many computations do not have such "critical" requirements. In economic research, for example, many computations can be done on a desk calculator or a small electronic computer within the desired time and accuracy limits. The direct dollar costs of computations are of major importance here.

### A Simple Case

Selecting the minimum cost method of data processing may be a simple task for an organization which is concerned with relatively few types of problems. When the workload is homogeneous, only a few alternative computing methods would have to be considered.

We have illustrated graphically a simple cost minimization problem. A schedule showing how the hypothetical computational cost per problem

[1] *Datamation*, Vol. 9, No. 2, February 1963, pp. 25–26.

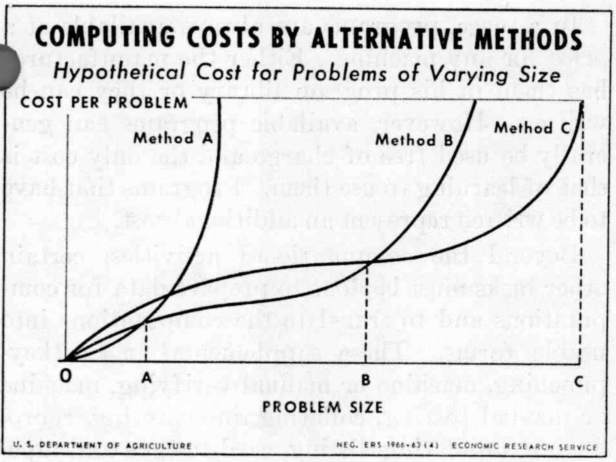


FIGURE 1

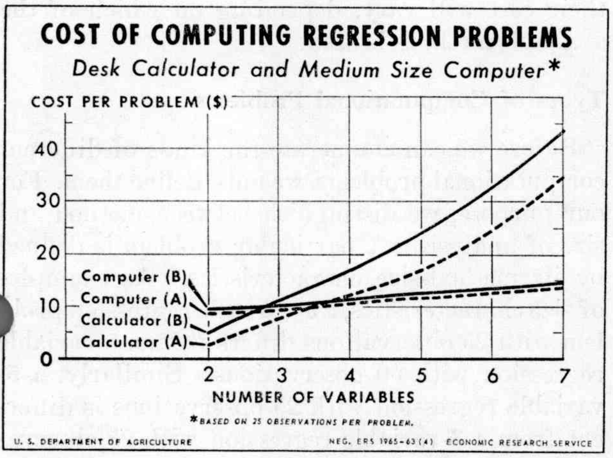


FIGURE 2

varies with the size of the problem is presented in figure 1 for each of three different computing methods. For problems that fall in the size range from 0 to A, method A is the cheapest; for problems whose size is between A and B, we would select method B; and for problems larger than size B, we would select method C.

To be more specific, we estimated the actual cost of computing regression problems by a desk calculator and a medium-size electronic computer. We fixed the number of observations in each problem at 25 and designated problem size by the number of variables. Estimated cost schedules are presented in figure 2. The cost curves labeled "A" for both calculator and electronic computer represent the estimated costs of getting all relevant coefficient estimates. The curves labeled "B" cover costs that include additional calculations for residuals, estimated values, and a test for serial correlation—information obtained from the same basic data. For the A situation it would be cheapest to compute problems of two and three variables on a desk calculator and larger problems on the electronic computer. However, when the additional computations are added to each problem under the B condition, this illustration suggests that it would pay to compute only two-variable problems on a desk calculator.

For the sake of simplicity, we used only the number of variables to indicate problem size. Actually, the number of observations is another dimension that would determine the size of the problem. Our analysis could be extended to a family of curves which represent alternative combinations of numbers of observations and variables.

When the total computational workload, the variety of problems, and the number of alternative computing methods are small, a series of graphic analyses is a useful approach to determining the minimum cost combination of computing methods.

### A Complex Case

The cost minimization problem becomes complex when an organization has many different types of computing problems. This great variety forces us to consider a larger number of alternative methods of computation. In this situation, graphic analysis is a laborious way to get the cost minimization solution. A more systematic approach is desirable. Such an approach is discussed in the next section.

## A Decision Model

We now turn to a more formal decision model that would assist administrators or management personnel to answer such questions as: (1) Given a number of problems, what is the cheapest way to get the computational job done? (2) Is it desirable for an organization to own at least some of the necessary pieces of computing equipment and if so, what pieces should it purchase? (3) How much of a change in the volume or composition of the computational problems is required before it is profitable to change from one machine complex to another? (4) If new equipment becomes available, would its adoption reduce computational costs? A model that provides answers to all of these questions can be used on a continuing basis.

Appropriately modified, such a model can also be used by many organizations.

Four sets of data are needed to determine the minimum cost equipment complex for a given total computing job. It is necessary to know (1) the different computing machines and corresponding machine programs that are available; (2) the number and types of computational problems; (3) the number of problems of each type that can be done on each machine-program combination in some time interval, say 1 hour; and (4) certain costs for each machine, including: costs of purchase, installation, and maintenance; costs of training or retraining personnel in the operation of equipment; costs of writing programs; and costs incurred when the machines are actually operating. With this information a formal model can be constructed that will help to answer the four questions posed above. Before proceeding to the model, however, let us discuss each set of data in more detail.

### Machines and Programs

Computing equipment consists of the machines and the appropriate programs. Programs are sets of coded instructions designed to solve different types of problems. We shall refer to each machine-program combination as a computing *activity*. There are many activities to be considered because the number of both machines and programs available is large.[2]

The total volume of work that can be done on a given machine depends upon its technical (physical) capacity and output as well as the number of programs available for it. If two machines, A and B, have the same technical capacity and the same initial and operating costs, but many more programs are available for A, a larger number and variety of problems can be processed on A than on B. Thus, the capacity of A can be more effectively utilized and the resulting total cost per problem will be lower for A than for B.

In a sense, programs are always available at a price for any machine. Either the manufacturer has them in his program library or they can be written. However, available programs can generally be used free of charge and the only cost is that of learning to use them. Programs that have to be written represent an additional cost.

Beyond the computational activities, certain other tasks must be done to prepare data for computations and to translate the computations into usable forms. These supplemental tasks (keypunching, machine or manual verifying, machine or manual sorting, collating, interpreting, reproducing, tabulating, listing, card-to-tape and tape-to-card conversions, and manual coding and decoding) will, of course, enter into cost computations. They are separate computational activities and their cost will vary, depending on which of the available methods is used.

### Types of Computational Problems

Before we can enumerate the kinds of different computational problems, we must define them. For our purpose, we distinguish between method and size of analysis. A particular problem is defined by its method-size characteristics. As examples of size characteristics, a 5-variable regression problem with 25 observations differs from a 5-variable regression with 50 observations. Similarly, a 5-variable regression with 25 observations is different from a 3-variable regression with 25 observations.

Some of the methods of analysis used in the U.S. Department of Agriculture are:

1. Regression analysis
2. Linear programming analysis
3. Simultaneous equation analysis
4. Input-output analysis
5. Seasonal variation analysis
6. Analysis of variance

In addition, there may be data reduction problems involving simple arithmetic operations and the tabulation and summarization of large volumes of data.

The methods of analysis given above generally involve more than one size dimension. For example, in a limited-information, simultaneous equation analysis there are three dimensions—the number of equations, the number of variables, and the number of observations. However, data re-

[2] Electronic computers can be grouped by their capacity. For simplicity, this paper is concerned only with 10K and 2K capacity machines, where K represents 1,000 words of addressable internal storage available. A listing of manufacturers of various-sized computers is customarily found in each November issue of *Datamation*. The most recent listing is in Vol. 8, No. 11, November 1962, pp. 32–40.

duction problems have only one size dimension; i.e., the volume of data to be processed.

Considerable reduction in the number of different types of problems used in the model can be obtained by defining a problem over a size interval. Setting up the intervals is a subjective matter. They should be chosen so that average values can be used for them, with a probable error not large enough to greatly affect the minimum cost solution. Of course, the smaller the interval, the more sensitive will be the allocation of problems among computational methods.

### Computational Efficiency

To determine computational costs it is important to know the number of problems of each type that can be computed per hour of machine time for each computing activity. Similar data are required also for supporting activities. In this way we relate the technical computing capacity of each machine to the amount of computing work to be done. Stated another way, given the number of problems, it is possible to estimate the time required under each activity to get the computing job done.

The required information on computational speed is often provided by machine manufacturers. Where it is not provided, the persons constructing the decision model would have to estimate computational speed. Further, estimates of manual time requirements for many of the supplemental activities would have to be made. It is our opinion that the collection of accurate performance data should be an integral function of any data-processing center.

### Cost of Machine Operation

Two sets of cost calculations are relevant. The first applies when computational services are purchased; the second when an organization owns a complex of computing equipment.

When computational services are purchased, the lowest obtainable price per unit of time used for each activity should be entered in the model. Since each activity is associated with a particular type of machine, the lowest price available could be chosen for each machine needed to complete a computation. It may not be necessary to obtain prices for each supplemental activity, since these may be included in the price charged for the computations.

The cost computations are considerably more complicated when a complex of machines is owned than when computational services are purchased. We have to distinguish between two types of cost—*fixed* and *variable* costs.

Fixed costs are those that are incurred in owning a machine. They do not vary with the level of machine use. They include (*a*) site preparation and installation costs, (*b*) depreciation charges based on the purchase price and estimated life of the machine, (*c*) costs of training or retraining personnel in the operation of the machine, and (*d*) costs of writing programs for the machine if they are not in the manufacturer's program library.

Variable costs are those incurred in the use of a machine. They include the costs of power, maintenance, supplies such as cards, tapes, and listing paper, as well as any portion of the salaries of machine operators, clerical workers, and supervisors that would vary with the level of machine use. They include also any assistance given by systems analysts and programmers in reviewing the preparation of input data for machine processing.

We are interested primarily in comparing machines on the basis of variable costs. However, fixed costs cannot be ignored. Total fixed costs may differ significantly among various complexes. They are an important function of the extent to which the computing capacity of a machine is utilized. The greater the utilization of the machine, the lower the fixed cost element in each problem. This is because total fixed cost is divided among more problems.

We want to minimize the total cost per problem (the combined fixed and variable cost components). And we seek the equipment complex that does this. For example, given a total computational workload, machine complex A may have lower variable costs than complex B. However, if complex A operates at a lower level of capacity than B, its fixed cost factor may be sufficiently large to cause the total cost per problem to be greater for A than for B. For this reason, fixed costs cannot be ignored in the decision model and utilization of machine capacity is an important consideration in many cost computations.

## Development of the Decision Model

Once the four sets of data just described have been obtained, the minimum cost combination of computing equipment can be determined. In this paper the formal scheme used is the cost minimization method in linear programming. The mathematical details of the problem are not presented. The general approach is sketched in narrative form. Those familiar with linear programming will be able to translate the description of the solution into a formal linear programming problem. Those not familiar with linear programming should be able to follow the general procedure outlined and see the nature of the decision model.[3]

Having obtained data on the number of machines and programs available (computing activities), the number and types of problems that can be computed on each machine in a given time, and the cost of operating each machine for the same period of time, one would organize these data in such a way as to provide the desired information. The problem can be formulated in two stages. The first stage assumes that no computing equipment is owned and all computational services are purchased. A minimum cost solution is obtained under this assumption. The second stage assumes that a combination of computing equipment is to be owned. The least cost combination of all equipment available is determined.[4]

The first stage of the minimum cost allocation of problems among the computing methods is relatively simple. Associated with each computing activity are the hourly cost of the machine and the number of problems of a specific type that can be computed in 1 hour. Given the total number of such problems that are to be solved, the amount of time each machine would take can be determined. This time multiplied by the appropriate price gives the cost for each machine. This procedure is then carried out for each specific type of problem.

An organization might want to own a computer complex even though it could get its work done more cheaply by hiring computational services. The convenience of having its own machines may be worth something to the organization, though it may not be able to attach a price to the factors that determine convenience. The model is helpful in determining the total cost of this convenience. For the difference between the cost of getting a given job done by renting computational services and the cost of owning a given machine complex represents the total convenience cost. Although convenience factors may be subjective, knowing their total cost can be helpful in deciding the degree of convenience that is worth while.

Based on the first stage of the cost problem, some machines can be ruled out as being too costly to own. There may be several reasons: (1) Some machines may have so few programs available for them that only a very small part of the total computing job could be done on them. When the costs of writing additional programs for these machines are considered, it may be cheaper to use other machines for which programs exist. (2) Some machines would be the lowest cost method for only a limited number of problems and, therefore, it would be cheaper to rent the services of the machines for those problems and own other machines to do the rest of the computing job. These are decisions made by the people who construct and solve the model.

Having eliminated some machines as being too costly to own, a machine complex that will get the computing job done at minimum cost has to be selected from those remaining. The possibility of purchasing machine time for certain types of problems is still open. And, if a limited number of these problems are the only ones for which a certain machine provides the lowest cost computations, these problems can be omitted and the minimum cost solution is determined on the basis of the remaining problems.

[3] There are many books on the subject of linear programming. Two references are: Earl O. Heady and Wilfred Candler, *Linear Programming Methods* (Ames: The Iowa State College Press, 1958), and Robert Dorfman, Paul A. Samuelson, and Robert M. Solow, *Linear Programing and Economic Analysis* (New York: McGraw-Hill Book Co., Inc., 1958).

[4] For those familiar with the cost minimization problem in linear programming, there are several ways in which the minimum cost combinations of machines can be obtained. One way is by use of an integer program which restricts the solution to multiples of whole machines. Another way is to force different combinations of machines into the solution by placing artificial prices on these machines to get solutions in terms of whole machines. We then compare the total cost of each combination forced into the solution and select the minimum cost combination.

The construction of the model would be a large undertaking. If it had no use except the selection of a computer complex for one organization at one point in time, it might be doubtful whether all the work involved could be justified. Fortunately, the model has many additional uses that might justify its construction.

### Continued Use of the Model

The model has many uses beyond that of selecting a minimum cost combination of computing machines for a single organization at one point in time. With minor modifications, it can be used by other organizations. Also, it can serve as a model for deciding when it is profitable to change a given computer complex. Such changes would stem from (*a*) the availability of new machines that have a cost advantage over existing machines for the type of work done, and (*b*) a change in the total level and composition of the computational work.

The electronic computer industry is a technologically dynamic one. New machines and programs are being developed at a very rapid rate. To keep abreast of these developments and to insure that the computing job is done at minimum cost, new machines or machine-program combinations that become available are added to the decision model as new computing activities. The number of problems of each type that can be computed on each new machine and the cost of operation would have to be known before purchase of a new machine could be considered. Given these data one can compare each new machine with existing equipment and decide whether or not it would be advantageous to buy new equipment.[5]

The total volume and composition of computational problems may also change sufficiently to warrant a change in the computer complex. An increase in volume of work might require merely expanding the number of machines of the type presently used. Or a change in composition of problems might involve a change in the type of machines used. Further, *projected* levels and composition of the workload can be analyzed within the framework of the model. Because of expected growth in the volume of work, an organization might want to consider investing in a computer complex with a capacity beyond current needs. It could project the total workload to levels and composition that would require a larger computing capacity, and then consult the model to decide whether or not the investment is justified.

### An Illustration of Model Construction

Sufficient data to construct and solve a complete model of the type described were not available to the authors. In what follows the limited data available are used to illustrate the construction of the model. The data are estimates based largely on experience in the Economic Research Service. The reader is cautioned that much of the data are necessarily crude estimates because of limited experience. We cannot emphasize too strongly the need to collect accurate data on a continuing basis because the answers obtained from the decision model are only as good as the data employed in the model.

The basic data used in our illustration are summarized in table 1. These data refer to three computing methods (a 10K machine, a 2K machine, and a desk calculator) and three selected types of problems (regression, linear programming, and seasonal analysis). We use this information to obtain the minimum cost solution for a specified total workload. The data in table 1 plus the assumed workload are presented in table 2, a format that facilitates cost calculations by the method of linear programming.

At the top of table 2 we have listed the three computing machines and the estimated hourly cost of each. We assume that all computational services are purchased. In addition, we list a data preparation activity and its hourly cost. For the 10K and 2K machines, estimates of computational time are based on specific programs. On the right-hand side of the table we list an assumed workload for which we want to obtain the minimum cost combination of computing activities. This workload consists of 200 five-variable regression problems, 4 size groups of linear program-

[5] For discussion of a method for deciding whether it pays to retain or replace a piece of equipment, see Glenn L. Johnson, "Supply Function—Some Facts and Notions," *Agricultural Adjustment Problems in a Growing Economy*, ed. by E. O. Heady et al. (Iowa State College Press, Ames, 1958), pp. 74–93; and Clark Edwards, "Resource Fixity and Farm Organization," *Jour. Farm Econ.*, Vol. 41, No. 4, November 1959, pp. 747–759.

**TABLE 1.**—*Estimated hourly machine costs, computing rates, and data preparation requirements for 3 kinds of equipment and 3 selected types of problems*

| Types of problems | Number of problems that can be computed per hour | | | Data preparation time ($5 per hour) for each hour of machine time | |
|---|---|---|---|---|---|
| | 10K machine [1] ($225 per hour) | 2K machine [1] ($30 per hour) | Desk calculator ($2.50 per hour) | 10K machine [1] | 2K machine [1] |
| | *Problems* | *Problems* | *Problems* | *Hours* | *Hours* |
| Regression problems (5 variables, 25 observations): | | | | | |
| Several problems computed from 1 basic set of data | 200 | 8 | ---- | 50 | 3 |
| Each problem prepared and processed separately | ---- | 5 | 0.105 | ---- | 8 |
| Linear programming problems: | | | | | |
| 5 rows, 10 columns | 120 | 12 | .125 | 120 | 18 |
| 25 rows, 45 columns | 20 | 1.333 | ---- | 80 | 16.024 |
| 100 rows, 500 columns | 4 | .025 | ---- | 60 | 25 |
| 200 rows, 1,000 columns | 1.5 | ---- | ---- | 90 | ---- |
| Seasonal analyses (12 years, 144 observations) | 30 | 3 | .125 | 30 | 3 |

[1] $K$=1,000 words of addressable internal storage capacity.

ming problems containing 120, 20, 4, and 3 problems, respectively, and 30 seasonal analyses.

The data in table 2 are of two types. One set gives the number of problems of a specific type that can be computed per hour on each machine. These data appear in lines 1, 2, . . ., 6. The other set of data gives the amount of data preparation time required for 1 hour of computation under each computing activity. These data are given in lines 1a, 2a, 2b, etc.

Line 1 is interpreted as follows: 200 five-variable, 25-observation regression problems can be run in 1 hour on the 10K machine; 5 regression problems of type I can be run in 1 hour on the 2K machine; 8 regression problems of type II can be run on the 2K machine in 1 hour and 0.105 problem can be run on a desk calculator in 1 hour. Thus, we could compute our total regression workload, using 1 hour of 10K machine time, 40 hours of 2K machine time if the problems are prepared singly, 25 hours of 2K machine time if several problems are obtained from a set of basic data, or 1,905 hours of desk calculator time.

Line 1a gives us the amount of data preparation time required per hour of computing time for each machine. For simplicity we have grouped all data preparation activities into one group. In practice we would want to carry many of these activities separately. We can read line 1a as follows: The number of regression problems that could be done in 1 hour on the 10K machine would require 50 hours of data preparation time; the number of single problems that could be done on the 2K machine would require 8 hours of data preparation time; etc. The −1 inserted in the data preparation column in combination with the equality relationship insures that the amount of data preparation time *needed* to support each activity is equal to the amount of data preparation time actually *used*. Stated another way, the −1 in the data preparation column indicates that the total amount of data preparation time used subtracted from the total amount of data preparation time needed on the computing activities equals zero. In this way data preparation costs as well as computational costs enter the model to give a complete cost analysis.

Assume we have 200 regressions to compute, of the type listed in table 2. The total cost on the 10K machine would be 1 hour of machine time at $225 per hour plus 50 hours of data preparation time at $5 per hour, for a total cost of $475. On the 2K machine, processing each problem separately would require 40 hours of machine time (200÷5) at $30 per hour plus 320 hours of data preparation time at $5 per hour, for a total cost of $2,800. If, however, we were to compute several problems from one basic set of data, only 25 hours of 2K machine time and 75 hours of data preparation time would be required, for a total cost of $1,125. To do the same job on the desk calculator would require 1,905 hours at $2.50 per hour for a total cost

TABLE 2.—*An example of a simplified decision model*

| Price per hour | $225 | | | $30 | | | | $2.50 | | | $5 | Number and type of computational problems | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 10K machine | | | 2K machine | | | | Desk calculator | | | | | |
| Computing activity | Regression [2] | Linear programming | Seasonal analysis | Regression I [1] | Regression II [2] | Linear programming | Seasonal analysis | Regression [1] | Linear programming | Seasonal analysis | Data preparation | Number [3] | Type |
| 1 | 200.000 | | | 5.000 | 8.000 | | | 0.105 | | | | 200 | 5-variable, 25-observation regression problems. |
| 1a | 50.000 | | | 8.000 | 3.000 | | | | | | -1. | 0 | |
| 2 | | 120.000 | | | | 12.000 | | | 0.125 | | | 120 | 5-row, 10-column linear programming problems. |
| 2a | | 120.000 | | | | | | | | | -1. | 0 | |
| 2b | | | | | | 18.000 | | | | | -1. | 0 | |
| 3 | | 20.000 | | | | 1.333 | | | | | | 20 | 25-row, 45-column linear programming problems. |
| 3a | | 80.000 | | | | | | | | | -1. | 0 | |
| 3b | | | | | | 16.024 | | | | | -1. | 0 | |
| 4 | | 4.000 | | | | .025 | | | | | | 4 | 100-row, 500-column linear programming problems. |
| 4a | | 60.000 | | | | | | | | | -1. | 0 | |
| 4b | | | | | | 25.000 | | | | | -1. | 0 | |
| 5 | | 1.500 | | | | | | | | | | 3 | 200-row, 1,000-column linear programming problems. |
| 5a | | 90.000 | | | | | | | | | -1. | 0 | |
| 6 | | | 30.000 | | | | 3.000 | | | 0.125 | | 30 | 12-year, 144-observation seasonal analyses. |
| 6a | | | 30.000 | | | | 3.000 | | | | -1. | 0 | |

[1] Data in this column assume that each problem is prepared and processed separately.
[2] Data in this column assume that several problems are computed from one basic set of data.
[3] All relationships are equalities.

of $4,762.50. Thus, the 10K machine is the cheapest for the type of regression problem specified here.

Similar cost comparisons can be made for each of the linear programming and seasonal analysis problems, and for any alternative number of problems. If no number appears in a computing activity for a particular size problem, that activity is not a feasible method of computation.

## Conclusion

The use of electronic data-processing equipment is rapidly expanding. Although this equipment greatly facilitates computational work, it is expensive. It is desirable to know the minimum cost combination of equipment for a given level of computational work. A wrong decision can be very costly.

We have presented a model which helps us to formally determine the cheapest way to get the computational job done. Further, the model is useful on a continuing basis to determine when it would pay to change an existing machine complex. This may be desirable as new equipment becomes available that is more economical than the old, or as the size and composition of the computing job change. Both types of changes can occur at rather rapid rates.