P. V. Norden

# Curve Fitting for a Model of Applied Research and Development Scheduling*

**Abstract: Analysis of the research and development process has suggested improved techniques for estimating the anticipated effort requirements for engineering projects. Logistic growth curves have been fitted to the time distribution of cumulative man-hours, and an effort-distribution array permits the extraction of meaningful information from historical cost records. These provide a consistent basis for adjusting current project schedules and making further estimates.**

## Introduction

In almost all industrial or military research and development activities, we are interested in predicting the probable effort requirements for applied research and development projects. As the pace of competition accelerates, and the technology of many products becomes more complex, an increasing premium is placed on accurate forecasting methods. To date, much of the estimating for R & D projects has been done on an *ad hoc* basis, and even educated guesses in this area have at times turned out to be wrong by as much as 300% to 400%.

Traditionally, R & D estimates have been made by comparing the job at hand with jobs previously completed. When job breakdowns, matrices, and estimating check-lists are available, forecasts are more accurate, but all too often the records are not truly comparable from job to job. Since an experienced estimator largely weighs the performance requirements of a new project against his recollection of relevant prior jobs, he may obtain a highly subjective estimate. The present study investigates broad, stable patterns and relationships in the R & D process, so as to provide a basis for at least limited improvement in forecasting accuracy. Since the outcome of a development project is affected by a large number of factors with wide variances, a prediction accuracy of $\pm 25\%$ of total actual time and effort would be considered a significant improvement for many projects. On major systems projects, which often run from 3 to 5 years, an *a priori* estimate which falls within $\pm 50\%$ of actual would be of value for planning purposes.

In scheduling we are interested not only in the total effort requirement, but also in its distribution in time; in this paper we examine the problem with emphasis on the distinction between the phases of work and the items and problems *worked on*. The distribution of effort within sequential phases of work appears to be more sensitive to task complexity than the traditional measure of total effort spent on subtasks. It should be emphasized at the outset that we are not dealing with phenomena which exhibit stability in a rigorous quality-control sense.

*Problem complexity* is defined as a function of the magnitude and difficulty of the task at hand. *Magnitude* is measured by a count of suitably chosen unit elements comprising the primary task, while *difficulty* is represented as an index based on previous comparable work. With this in mind, we set up an organized set of records (the effort-distribution array) and provide a body of rules (the logistic and phase curves) for developing forecasts and schedules from these records. Starting with the best information obtainable from expert estimators, we refine our knowledge of the process parameters by feeding data back into an accumulating body of records, in a consistent manner, as each project is run. The method is capable of continued increase in accuracy if systematically applied for several successive projects. It has been programmed for the IBM 704 Data-Processing Machine.

This paper will discuss the research and development process, examine the transformation concept and its application to forecasting and monitoring, and describe the method of developing an effort-distribution array for a given project.

## The research and development complex

Applied research and development may be regarded as two segments of a chain which begins with fundamental research and terminates with the manufacture of new

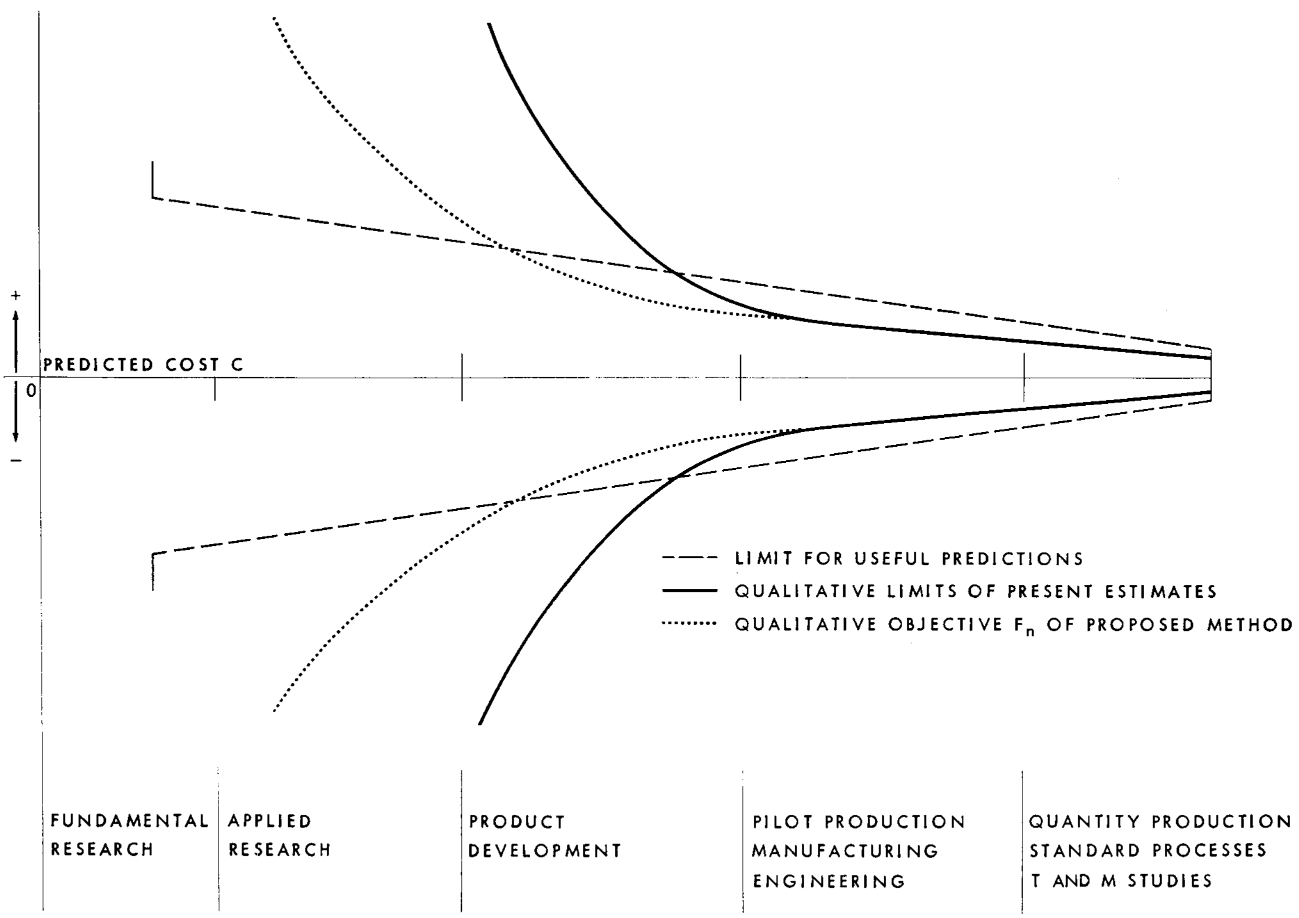
*Portions of this paper were presented at the Tenth National Meeting of the Operations Research Society of America at San Francisco, November 15, 1956.

products. This sequence is shown in the lower portion of Fig. 1. If a prediction of total time and man-hours to complete a given project is made, we may plot the actual outcome, once the job has been run, as a percent deviation from the prediction. We define the confidence limits as the boundaries of the areas within which these deviations may be expected to lie. In Fig. 1 three confidence intervals are shown. The limits tend to converge because, as the production phase is approached, the estimator has increasing knowledge concerning objectives to be achieved, the operating parameters and statistical stability of the process, and the methods and technology to be employed.

Knowledge of the process being measured tends to decrease as we go towards research. This is illustrated by the three pairs of limits of Fig. 1. The dashed lines represent tolerable prediction limits in a given situation. Predictions useful for budgeting and planning should have a probability of accuracy within the indicated limits. The solid lines give a qualitative estimate of what is presently achieved with conventional forecasting methods. And the dotted lines indicate the improvement that is sought with the techniques proposed here, namely, to extend the useful prediction range over the full area of engineering development, and slightly into applied research. Much of the latter and all of fundamental research are considered too open-ended for the method to apply. It should be emphasized, however, that since the study is still in progress, the relationships discussed below represent a working hypothesis, and are subject to verification in the light of additional data.

The present investigation is, therefore, concerned with the time span between the inception of a development project and the point where the first article meets all applicable specifications, and is released for manufacture. This interval has been subdivided into the sequence of activities shown in Fig. 2. In the field of electromechanical systems and equipment design this sequence appears to have considerable stability. In other fields, slightly different activities may be appropriate, although there appears to be widespread agreement in general on the nature and ordering of activities in research and development. This sequence is essentially independent of the content of

*Figure 1* **Probable deviation limits for actual vs estimated costs**



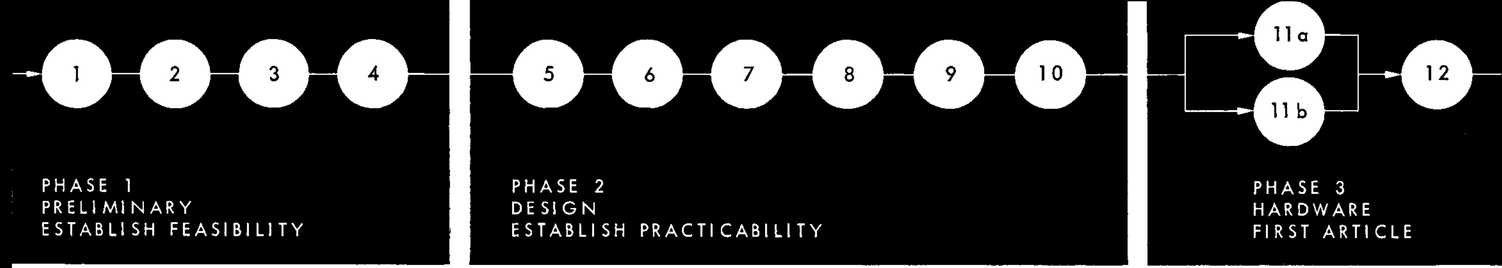



*Figure 2* **Typical work phases and activities in electromechanical development.**

| | |
|---|---|
| **Start of project** | • *Initial statement (or specification) of design and performance objectives* • |
| **1. Basic conceptualization and planning** | • *Formulation of problem* • *Literature search* • *Analysis of problem* • *Fundamental inquiry* • *Formulation of hypotheses* • *Conjecture as to nature or technique of possible solution* • *Gestalt generation: "This looks like a possible solution"* • |
| **2. Crystallization and methodology** | • *Determining approach to experimentation* • *Design of experiments* • *Computation and simulation of anticipated relationships* • *Design and sketching of experimental structures* • *Determination of problem-solving methodology* • |
| **3. Experimentation and reduction to practice** | • *Build breadboard models* • *Construct test equipment for experiment* • *Conduct experiment* • *Take measurements* • *Analyze results* • *Test original hypotheses* • *Evaluate* • |
| **4. Preparing specifications to guide subsequent design** | • *Write down recommendation based on the findings of the foregoing work* • *Specifications* • *Definitions* • *System philosophy and structure* • |
| **5. Design and logic planning** | • *Engineering approach to the whole design* • *Development and evaluation of alternative configurations, structures and the way they will be combined* • |
| **6. Design (for experimental model)** | • *Design and layout of structural and functional features of all parts of the equipment for experimental model* • |
| **7. Drafting (for experimental model)** | • *Preparation of sketches, schematics, detail and assembly drawings, parts lists and bills of materials* • *Checking of the above for experimental model* • |
| **8. Building and testing of experimental model** | • *Fabrication, assembly and test of experimental model, experimental circuits, components and techniques* • *Include construction of test equipment intended for this phase primarily* • *(Charge engineering changes found necessary here, to Design 9)* • |
| **9. Design (for final model)** | • *Same as 6, but intended for engineering model ("First article")* • *Include industrial design here* • |
| **10. Drafting (for final model)** | • *Same as 7, but intended for engineering model ("First article")* • |
| **11a. Engineering model** | • *Fabrication and assembly of engineering model* • *Include such testing as is incidental to assembly, but not performance testing of completed device* • |
| **11b. Design and build test equipment** | • *Design and construction of all test equipment intended for use with the engineering model* • |
| **12. System testing and debugging** | • *Performance testing of completely assembled engineering model* • *Debugging activities* • *All engineering changes short of radical redesign* • *Evaluation activities* • |

the project, i.e., the nature of the items *worked on*. The activities themselves continue while the items to which they are applied may change.

The 13 activities shown in Fig. 2 are grouped into three phases:

Phase 1: Creative (exploratory); establish feasibility.
Phase 2: Design; establish practicability.
Phase 3: Experimental Hardware; physical embodiment.

Every part of a given R & D project must pass through these phases and activities at some time before the project is completed.

For purposes of estimating, it is customary to attack a project in parts. The nature and extent of this partition, of course, depends largely on the field or industry concerned. Figure 3 shows a typical schematic of such a breakdown. While there are no hard and fast rules governing the proper choice of subdivisions or how far this process should be continued, experience has shown that at any given level of project responsibility, second- or third-order breakdowns are preferable. The partitioning process is inherently scaleless. While carrying the breakdown to the "nut and bolt" level is patently absurd, these procedures are as applicable to the development of an engine as to the design of an entire aircraft. In this discussion, first-order breakdowns are termed tasks ($\tau$), second-order breakdowns are elements ($e$) and third-order breakdowns are subelements ($\varepsilon$).

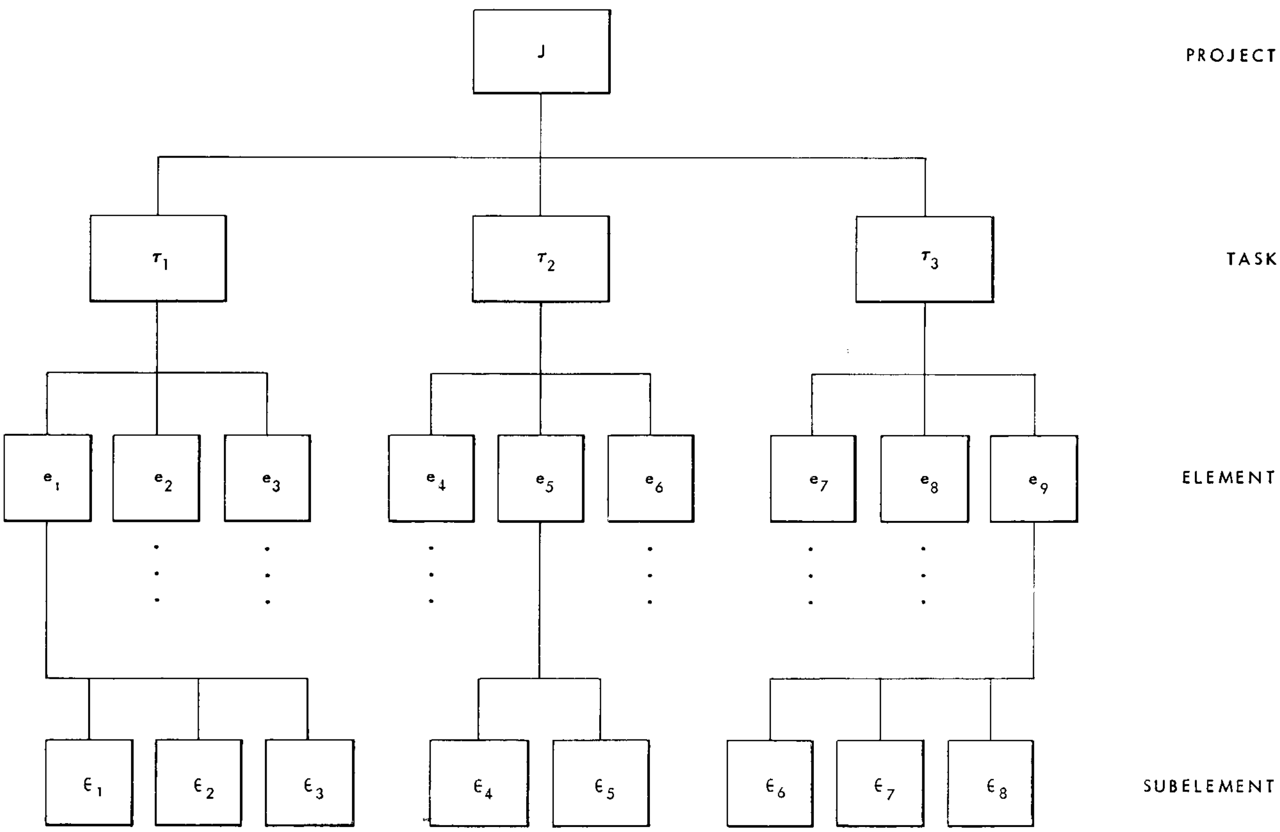
Some considerations in choosing the size of the smallest subdivisions are as follows:

1. They should be of sufficient magnitude to warrant the assignment of at least one man for a significant period, the length of time depending on the operation and its associated scale factor.
2. As far as possible, they should be physically self-contained. In dealing with machines or processes, a good choice appears to be a subassembly or clearly identifiable process stage. However, a definable problem, as distinct from a piece of hardware, is also a suitable element. (Military environmental requirements illustrate this point.)
3. They should be manageable by small numbers of men or teams. If the subelement chosen is too large, prediction accuracy is generally sacrificed.

Initial decisions in this area are best made on the basis of a company's experience. Most businesses can find natural first-order tasks. For example, in camera development we may find such items as body, film magazine, lens assembly, and drive mechanism. Computers may suggest input-output devices, memories, central processing units, controls, and power supplies. Here again, consistency of choice from project to project is essential to the refinement of any associated measures of complexity and prediction accuracy.

The final step in scheduling an actual project is the construction of a Gantt chart or its functional equivalent.

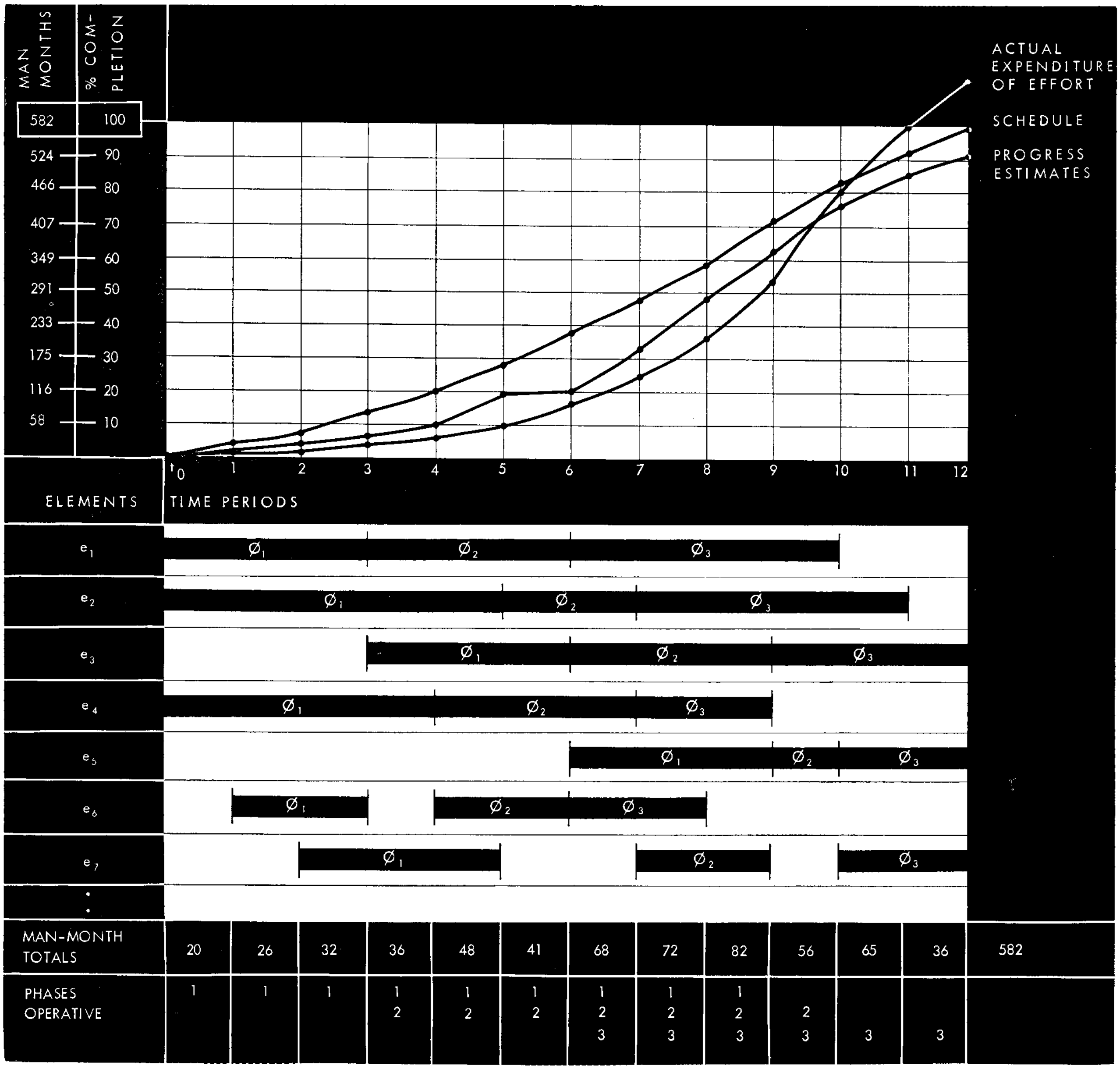
*Figure 3* **Job breakdown.**



Consistent with our concepts of work phases and job elements, the Gantt chart in Fig. 4 is a master project schedule which indicates the time when each task is scheduled to pass through the various activity phases. A limiting condition would be to lay activities end to end in time, but this is patently impractical in all but the simplest of assignments. Knowing which activities can be started simultaneously and which are sequentially dependent is more an art than a science, and the decision is usually left to planners experienced in their field. However, it will be shown how decomposition of cumulative effort curves can assist materially in approximating optimal Gantt charts.

Once a Gantt chart has been laid out we have effectively fixed the sequence, duration and manpower requirements of each element-activity combination, and thereby uniquely determined a schedule polygon for the project, as shown in Fig. 4, top. This polygon is the curve developed by cumulatively adding the total number of men scheduled each month. In addition to the man-month polygon, two additional curves are shown on the upper portion of Fig. 4. The data for the "progress estimate" curve are taken from periodic progress reports prepared by project engineers or other R & D administrators. "Actual expenditures of effort" are taken from accounting records. The ordinate is expressed both in percent-com-

*Figure 4* **Relationship of Gantt chart to schedule polygon.**



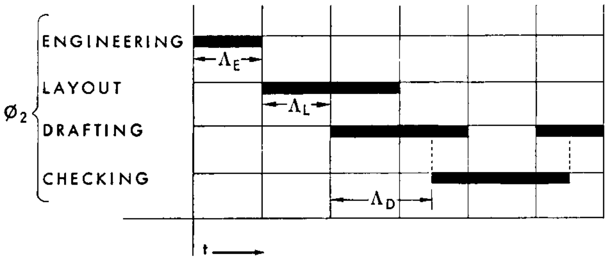



*Figure 5* **Typical activity lags in Phase 2. The $\Lambda$ intervals represent lead times between the start of successive elements.**

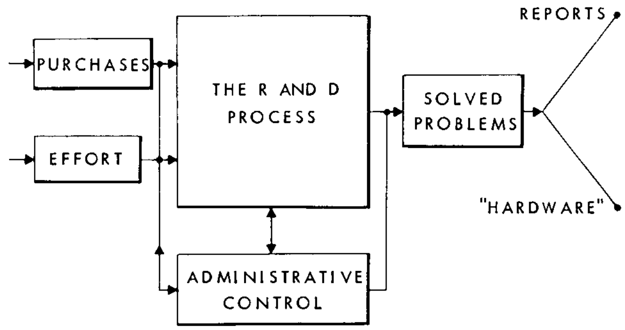



*Figure 6* **The R & D transformation process.**

pletion and in a suitable effort scale, such as man-months. The total appropriation for the job may logically be equated to 100% completion. It then becomes possible to evaluate schedule, progress, and expenditures in terms of a common percentage scale.

To illustrate the inherent scalelessness of the phase concept, Fig. 5 shows the time lags between several successive activities of Phase 2, the design phase. Whether we emphasize lags between successive activities or successive phases depends solely on the coarseness with which we wish to estimate. In general, tasks and projects are best associated with their respective phase lags, while elements and sub-elements can be considered in terms of activity lags. Preliminary studies indicate that the relative difficulty of diverse tasks correlates well with their associated lead times. While the sequence illustrated in Fig. 5 is not atypical, the objection might be raised that these lags are not actually as clear-cut as represented here. These points will be treated more fully below.

### The transformation process

An essential difference between the R & D and the production processes lies in the fact that the production process has for its output a population of discrete objects capable of being uniquely identified and counted, while the R & D process produces solutions to problems. The output of the R & D process is new knowledge embodied in written reports and experimental hardware, such as prototype models, shown in Fig. 6. The input consists of human effort and money distributed in time. While the costs of materials, tools, test equipment, and purchases, et cetera, do not seem to have a functional relationship to the difficulty of the project, the availability of electronic computers, nuclear reactors, and similar high-power computational and test equipment, contributes to the productive capacity of a group. Use of such equipment should be noted explicitly in any study of project effort.

We are concerned here, however, solely with the man-hours of engineering effort. The lower portion of Fig. 4 indicates that creative human effort is applied to the work in successive phases. In each of these, the rate of application of effort builds up to a peak and then diminishes, and the waning of a leading phase is generally associated with the waxing of the following phase. This phenomenon is depicted in Fig. 7, which shows the rate-of-effort curves for each phase of a typical project. The relation of the three phase curves, with respect to magnitude as well as location in time, is by no means fortuitous, but causally related. Preliminary planning by a few men gives rise to a series of questions which must be explored by increasing numbers of other investigators. Eventually, the required answers are found and Phase 1 draws to a close. But while it was operative it generated a sizeable number of specifications and engineering directives, which call for a relatively larger group of designers and development personnel for Phase 2. These in turn produce drawings from which the Phase 3 hardware is ultimately built, and when the latter is "de-bugged" the project is terminated.

In practice, the trailing legs of the curves of Fig. 7 often extend for protracted periods at a low level, since a small liaison, manual writing, or "clean-up" group is often retained to attend to the inevitable odds-and-ends as the project is released to production. However, one can agree that a phase shall be considered terminated as soon as the effort level drops below, say, 1 percent to 5 percent of the phase peak.

Since the phase curves overlap, we obtain points of intersection at points on the time scale at which the *rate* of effort for the leading phase exactly equals that of the lagging phase. We term these time points $\lambda_1, \lambda_2, \ldots$, and propose the time between them as real-life equivalents of the idealized lead-times of Fig. 5. The utility of $\lambda_i$ in project monitoring is treated further in the section entitled, "Establishing the Reference Project."

It is apparent that the rate at which engineering manpower is applied to a project at any time in its life is the sum of the manpower rates for each phase:

$$\Phi(t) = \phi_1(t) + \phi_2(t) + \phi_3(t). \tag{1}$$

The total effort (cumulative man-hours, man-months, etc.) absorbed by a project from its inception to some time $t_i$, is given by the area under the phase curves, so that for projects of sufficient magnitude to treat the functions as continuous

$$F(t_i)=\int_{t_0}^{t_i}\Phi(t)\,dt=\int_{t_0}^{t_i}\phi_1(t)\,dt+\int_{t_0}^{t_i}\phi_2(t)\,dt+\int_{t_0}^{t_i}\phi_3(t)\,dt. \quad (2)$$

This equation defines $F(t_i)$, the cumulative effort expended on the project. Analysis of historical project records appears to indicate that the cumulative effort of many projects may be described by logistic growth curves of the form

$$F=\frac{K}{1+Ae^{B(t)}}. \quad (3)$$

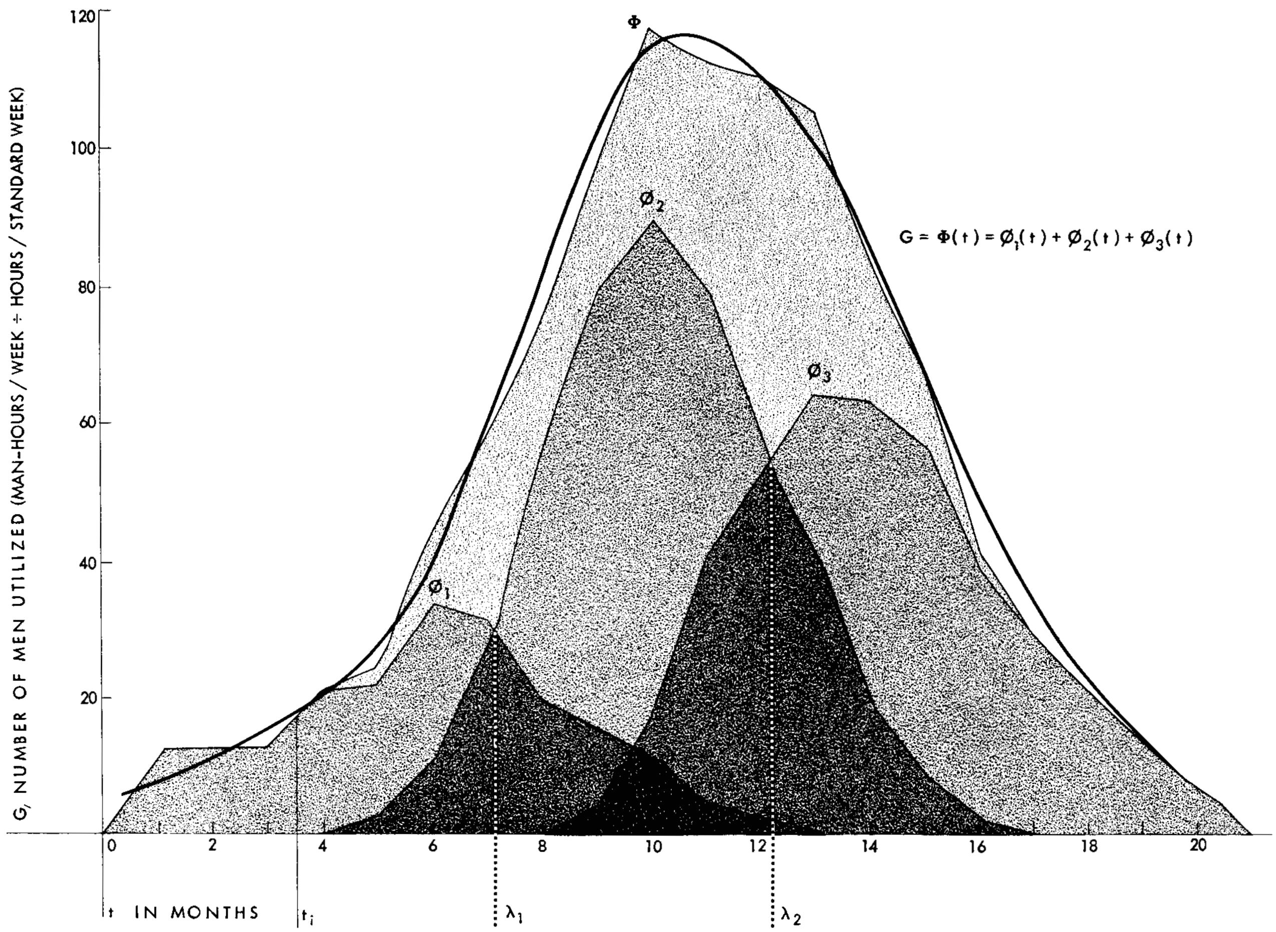
In the present investigation, a curve-fitting program has been written, based on the equation:

$$F(t_i)=\frac{K}{1+\exp(a+bt_i+ct_i^2+dt_i^3)}. \quad (3a)$$

The program first tries the Pearl-Reed curve, in which both $c$ and $d$ equal zero. If this does not fit well it selects the best values, in the least-squares regression sense, of all the parameters $a$, $b$, $c$, and $d$, for given values of $K$ and $t_n$. It also determines the closeness of fit in terms of the standard error of estimate. It is intended to increase the utility of this program by adding routines permitting iterative scanning of incremental and decremental values of $K$ and $t_n$. The dotted line in Fig. 8 shows the logistic curve fitted to the cumulation curve (solid line) of the project of Fig. 7.

The possible meanings of the several growth curves which might describe the R & D process are of some interest. At the beginning of an R & D project, there is a latent body of unsolved problems. The number and complexity of this body of problems may not be known. At the outset it is necessary to analyze the total problem before the subproblems and their sequentially dependent implications can even be identified. This initial analysis is made most efficiently by a very small number of individuals. As approaches to problem solutions are conceived, more personnel must be added to the project to

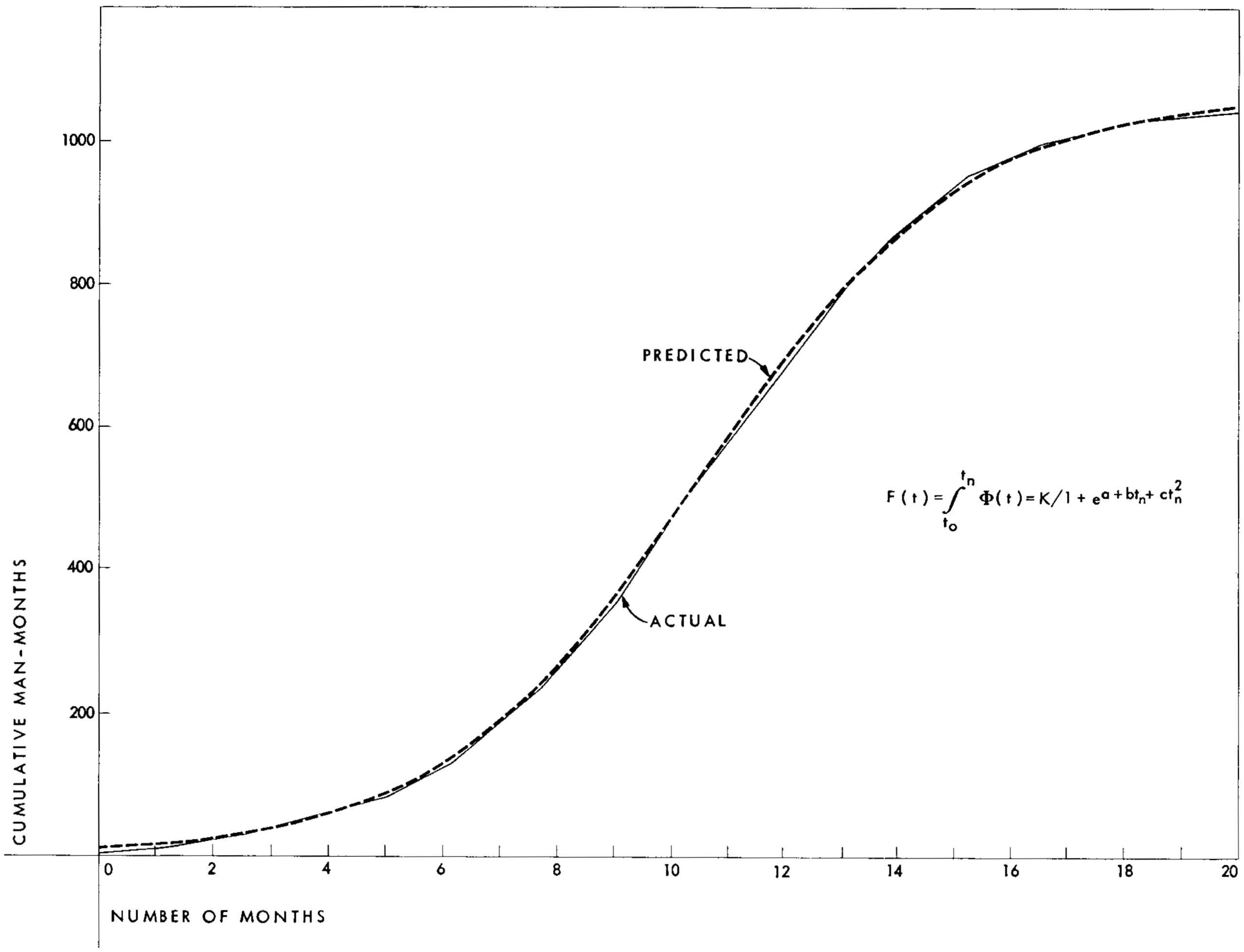
*Figure 7* **Rate of application of effort for sample project, by phases.**
*The number of men assigned to each phase of work per month, is shown by curves $\phi_1$, $\phi_2$, and $\phi_3$. Their algebraic sum produces curve $\Phi$, the manpower distribution for the entire project.*



begin work on the subproblems now formulated. This analytic, planning, and problem-solving process is carried on into the design and drafting stages of the job, until the original body of problems is fully explored and exhausted. From this point on there generally is a decline in the number of men needed on the job. Through the experimental assembly, testing and debugging phases, there is a gradual "phasing out" of engineering and technical personnel as the item under development is perfected. This process of growth and decline and the resulting logistic curve is readily related to population growth phenomena and the growth of living organisms, which have been extensively studied.[1-4] In particular, the Pearl-Reed curve, in which $B(t)$ is linear, has been fitted to the growth of a population of fruit flies in a bottle, and describes the change in population density in a confined space, as a function of time. There is thus an analogy between the exhaustion of the set of unsolved problems in the R & D process and the exhaustion of space in the bottles.

Davis,[1] however, points out the interesting fact that there is an essential difference between the growth of a population which is not subject to a *central* mechanism of control (the flies in the bottle), and the growth of a population which is. The latter situation is illustrated by the growth of the cells of a pumpkin or the increase in weight of an animal from birth to maturity. In this case, $B(t)$ in Eq. (3) has been shown to approximate the cubic function $bt+ct^2+dt^3$, with the coefficients of odd power terms

*Figure 8* **Pearl-Reed logistic curve fitted to cumulative effort for a sample project.**




| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\phi_1$ | 0 | 12 | 13 | 13 | 21 | 22 | 34 | 31 | 19 | 15 | 11 | 4 | 2 | | | | | | | | |
| $\phi_2$ | | | | | | 3 | 11 | 27 | 56 | 79 | 89 | 68 | 56 | 41 | 19 | 8 | 2 | | | | |
| $\phi_3$ | | | | | | | | | | 4 | 17 | 40 | 52 | 64 | 63 | 57 | 38 | 28 | 20 | 12 | 7 |

ACTIVITY PHASES OPERATIVE. MAN-MONTHS UTILIZED

*Fig. 9a*

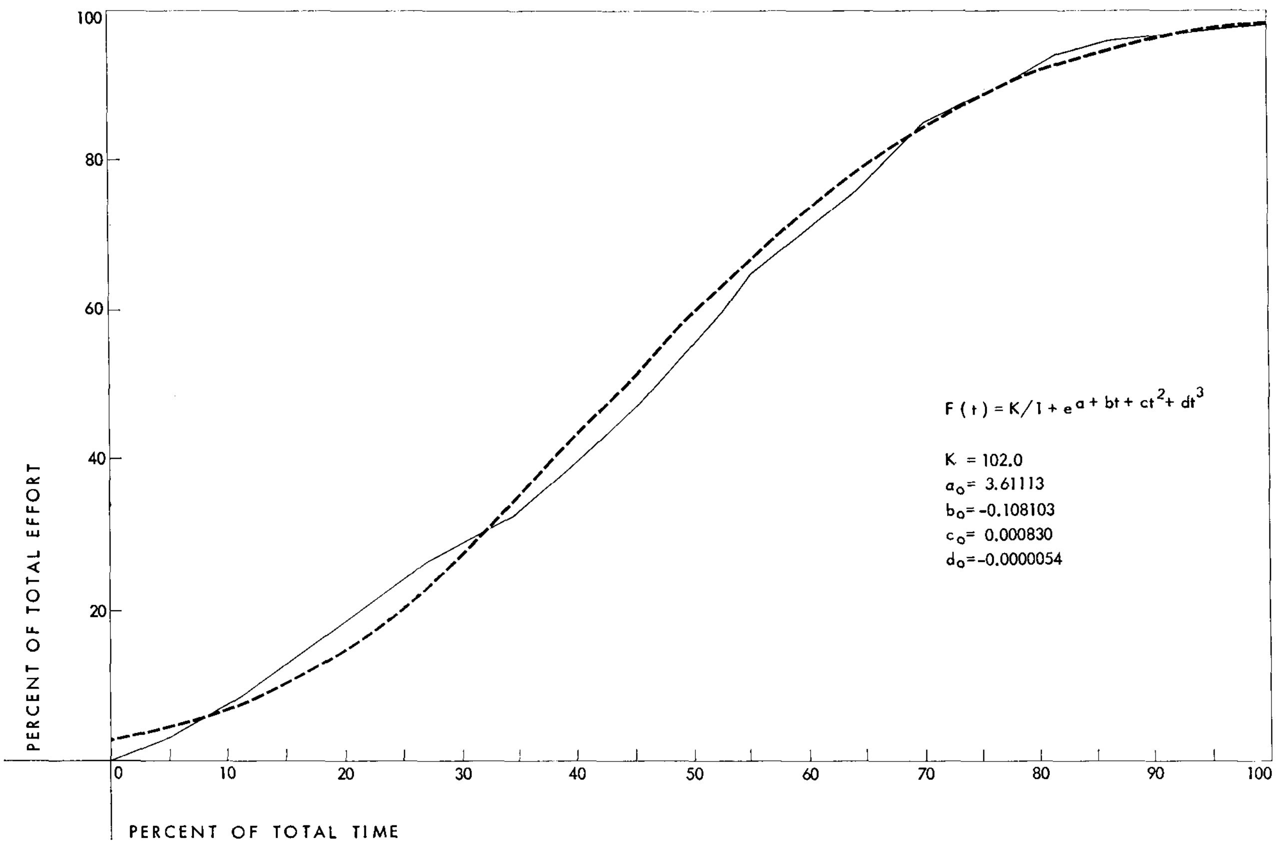



*Figure 9* **Cumulative effort distribution.**
*a) Development of a large computer F(t).*
*b) Development of calculator of moderate size.*
*c) Development of very large computer system.*

negative. If we consider the managerial direction of an R & D project a type of central control function, we may suspect that the cubic function would apply.

The R & D effort distributions for widely different classes of products can be approximated by the family of logistic curves. Figure 9a shows the curve for the development of a large computer, on a percentage (dimensionless) plot. This project was completed several years before this study was begun, and the accounting information was not segregated on as rigorous a classification scheme as the one to be outlined below. Nevertheless, the logistic curve gives an adequate fit for planning purposes. Data for a calculator and for a very large computer system are shown in Figs. 9b and 9c. Figure 10 shows the effort cumulation for a small electromechanical device manufactured by a small company, also prior to the formulation of this method. This illustration includes a tolerance interval of $\pm 2.5\%$ of the nominal values of the parameters $a$ and $b$. While their numerical values differ for the two companies, the functional relationship holds.

There is a temptation to speculate whether given organizations or product types give rise to characteristic distributions. Johnson and Turner,[5] in a study of multiproduct engineering-sales relationships at the Minneapolis-Honeywell Regulator Co., show curves for engineering effort which were strikingly similar to our $\Phi(t)$, and were led to state: "Most devices exhibited a similar pattern in the way that the engineering effort increased to a peak and then decreased rapidly to a low level. The low level continued during the sales life of the product. A remarkable similarity appeared in the sales pattern."

This consistency of pattern in the effort curves is noteworthy in the light of the wide ranges of project magnitude, nature of devices developed, size of company, and time span of data collection.

While it is possible to achieve good fits to project data with other growth curves, the Pearl-Reed has the considerable advantage that the parameter $K$ estimates the upper asymptote of the curve explicitly, providing a direct value for the total effort predicted for a given project.

The upper terminus of the cumulation function represents two highly important quantities: total effort and total time. The shape of the function describes the time path along which the end of the project is approached. For forecasting, it remains to relate these factors to task complexity in operational situations.

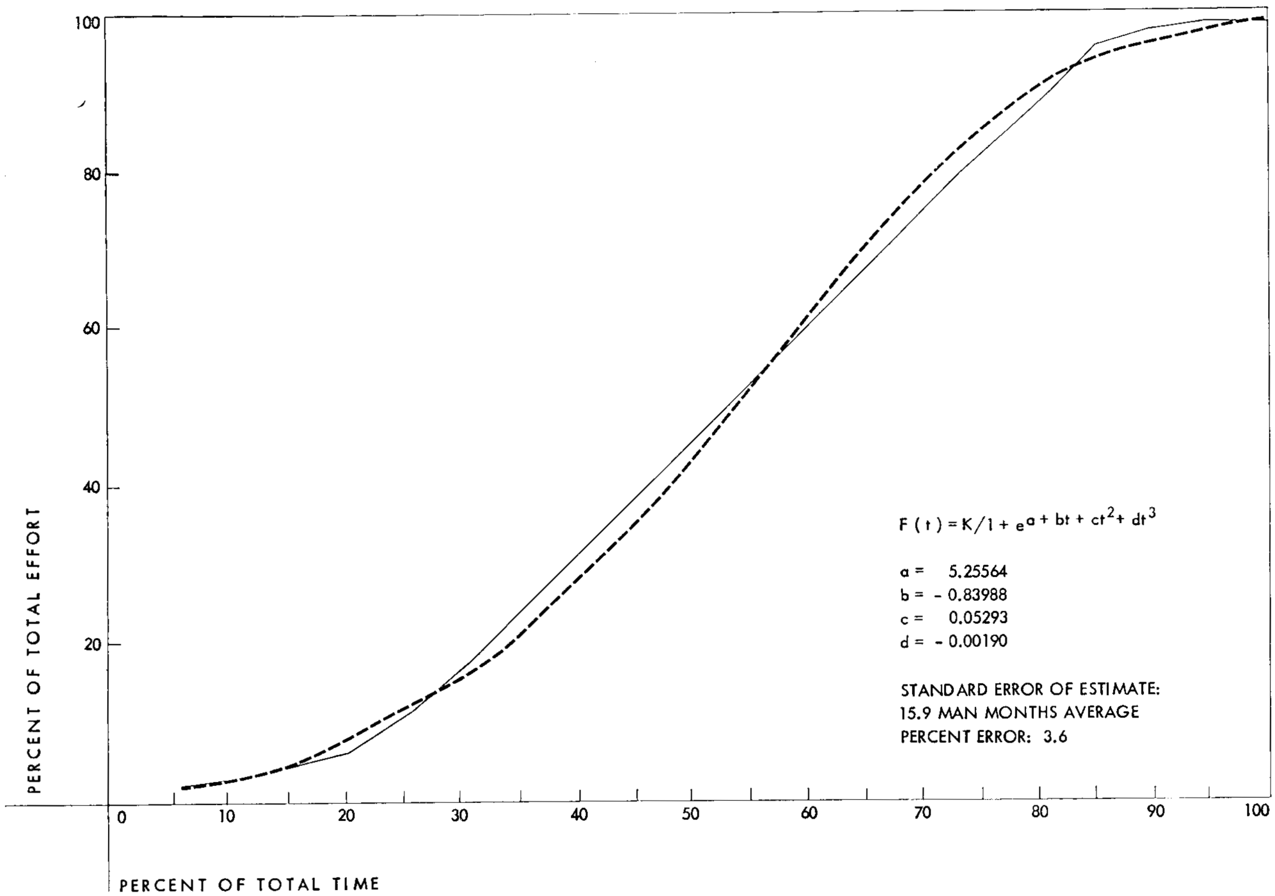

PERCENT OF TOTAL EFFORT
100
80
60
40
20
F ( t ) = K/1 + e^a + bt + ct^2 + dt^3
a = 5.25564
b = - 0.83988
c = 0.05293
d = - 0.00190
STANDARD ERROR OF ESTIMATE:
15.9 MAN MONTHS AVERAGE
PERCENT ERROR: 3.6
0
10
20
30
40
50
60
70
80
90
100
PERCENT OF TOTAL TIME

*Fig. 9b*

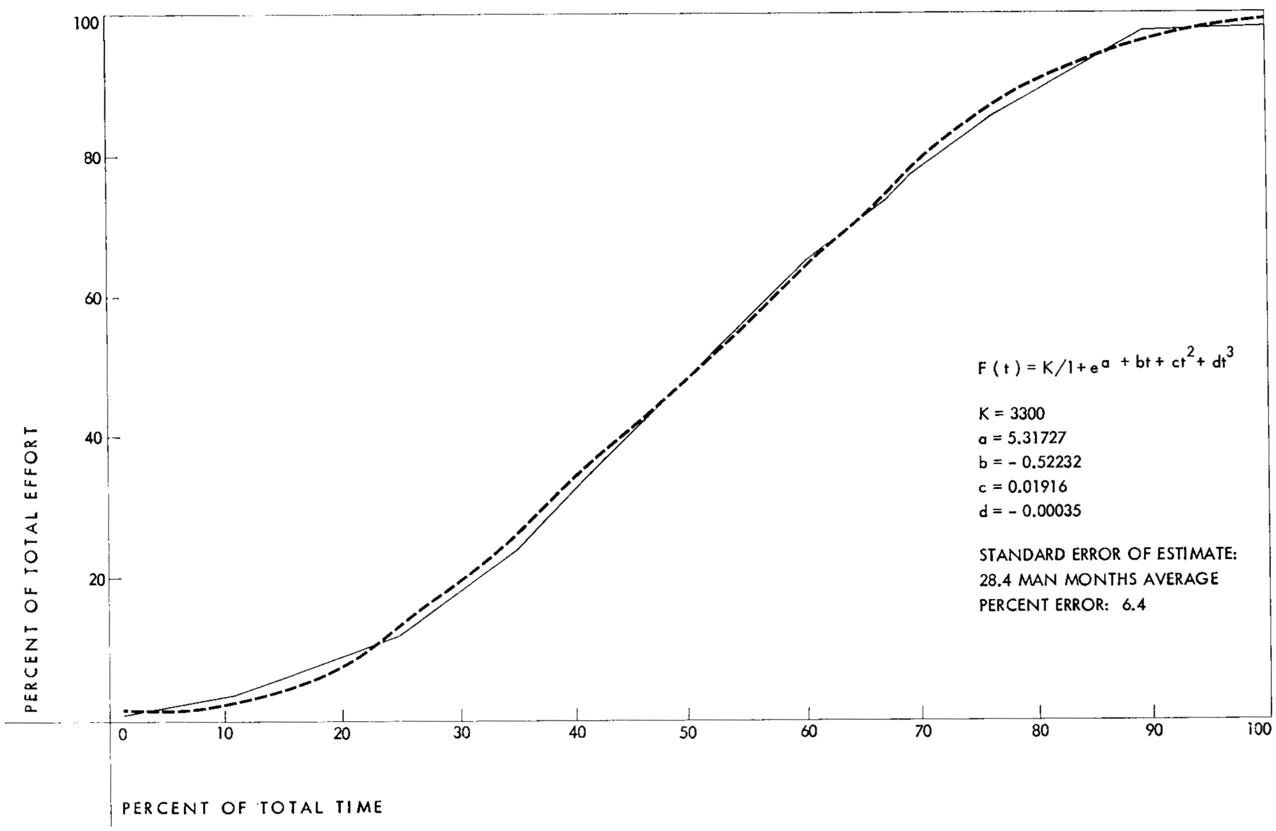

PERCENT OF TOTAL EFFORT
100
80
60
40
20
F ( t ) = K/1+e^a + bt + ct^2 + dt^3
K = 3300
a = 5.31727
b = - 0.52232
c = 0.01916
d = - 0.00035
STANDARD ERROR OF ESTIMATE:
28.4 MAN MONTHS AVERAGE
PERCENT ERROR: 6.4
0
10
20
30
40
50
60
70
80
90
100
PERCENT OF TOTAL TIME

*Fig. 9c*

## Establishing the reference project

Our estimating procedure is based on the careful analysis of a project typical of a company's class of work, called the *reference project*. It is possible to select such a project from previous jobs, but unless the available historical records are unusually detailed and accurate, it is preferable to establish a current job as the reference for determining the parameters $K$ and $t_n$ of later projects.

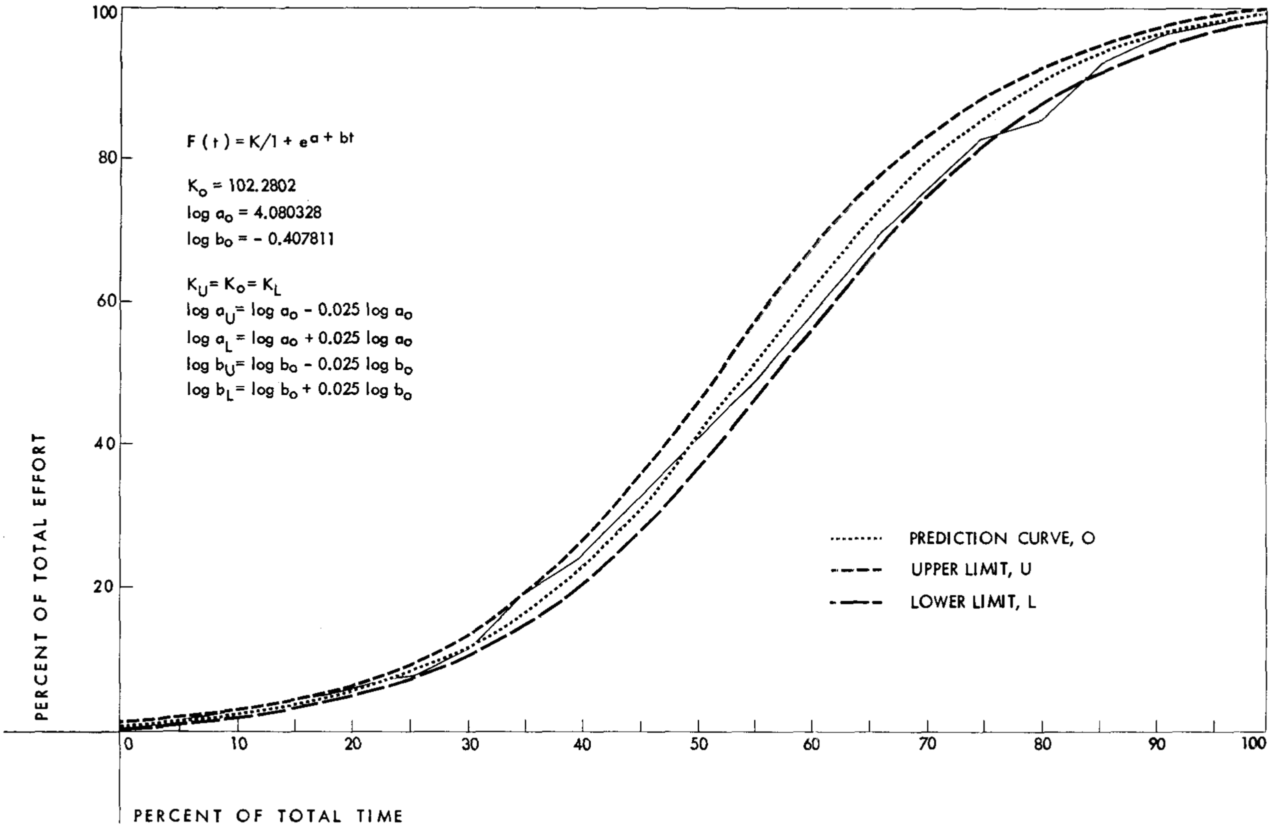
Thus, if $K_R$ is the total number of man-hours (weeks-months) of the reference job, $K=K_R\Delta$, where $\Delta$ is a proportionality coefficient which is taken to be a measure of task complexity.

Complexity, as a function of the magnitude and difficulty of a job, is not easily measured. However, in any breakdown of a job into tasks, tasks into elements, and so on, the magnitude can be taken as the unweighted count of the smallest subelements. In the early stages of the forecasting procedure the weighting function may have to be supplied by an experienced estimator, but after repeated analyses of successive projects this information should emerge from the data. The estimation of difficulty is accomplished by suitably developed performance indices $\psi$, which are discussed in the section entitled "Estimating and Monitoring."

An effort-distribution array of the type shown in Fig. 11 is a useful device for accumulating data while establishing the reference project. If we list the subelement set $\varepsilon_j$ across the top of the array, as column headings, and list the activities $I_i$ down the left-hand side for row headings, each cell of the array will correspond to a unique element-activity combination. Thus $T_{ij}$ represents the total number of man-hours generated by applying the $i$-th activity to the $j$-th job element. The column totals define the total time spent in all activities on a given job element, while the row totals show the total effort spent in each of the several activities across the entire job. The cross-footed total in the lower right-hand corner gives the grand total effort. By proper grouping of rows or columns we can determine the effort attributable to phases and tasks. The effort-distribution array used here is a refinement of job matrices variously used in industry. It differs from these primarily in the strict adherence to the time-sequential arrangement of the work phases, and a rigid distinction between the items and problems *worked on* ($\varepsilon$), and the nature of the activity ($I$) applied to them. Furthermore, the array preserves the input-output concepts discussed earlier and generates all the information required for estimating and monitoring purposes.

We designate by $T$ the concept of time as a commodity (in man-hours, etc.), and by $t$ the notion of elapsed or calendar time. The array presents an instantaneous picture in calendar time. In practice, we make use of two

*Figure 10* **Cumulative effort distribution.**
*Pulse mechanism for aerial camera control.*



types of arrays, one showing effort during a single accounting interval and the other showing cumulative effort to date. If the accounting intervals are small relative to the duration of the project, then in the limit, the row totals of the first type of array generate points of the rate-of-effort, or phase, curves; while the grand-total term generates points on the over-all project curve:

$$\sum_{i=1}^{3} \sum_{j=1}^{n} T_{ij} \rightarrow \phi_1(t)$$

and

$$\sum_{i=1}^{m} \sum_{j=1}^{n} T_{ij} \rightarrow \Phi(t).$$

Similarly, the cumulative arrays generate the integral functions

$$\sum_{t} \sum_{i=1}^{3} \sum_{j=1}^{n} T_{tij} \rightarrow \int_0^t \phi_1(t)\,\mathrm{dt},$$

and

$$\sum_{t} \sum_{i=1}^{m} \sum_{j=1}^{n} T_{tij} \rightarrow \int_0^t \Phi(t)\,\mathrm{dt}.$$

Thus the row totals describe effort input to the R & D process, while the column totals provide level-of-complexity information for the elements of the job.

It is worth noting that each column vector and each row vector of the array is composed of independent functions of time. This property enables us to use this device for projects differing widely in magnitude, and for such activities as military product development, which frequently starts with a detailed specification that truncates Phase 1.

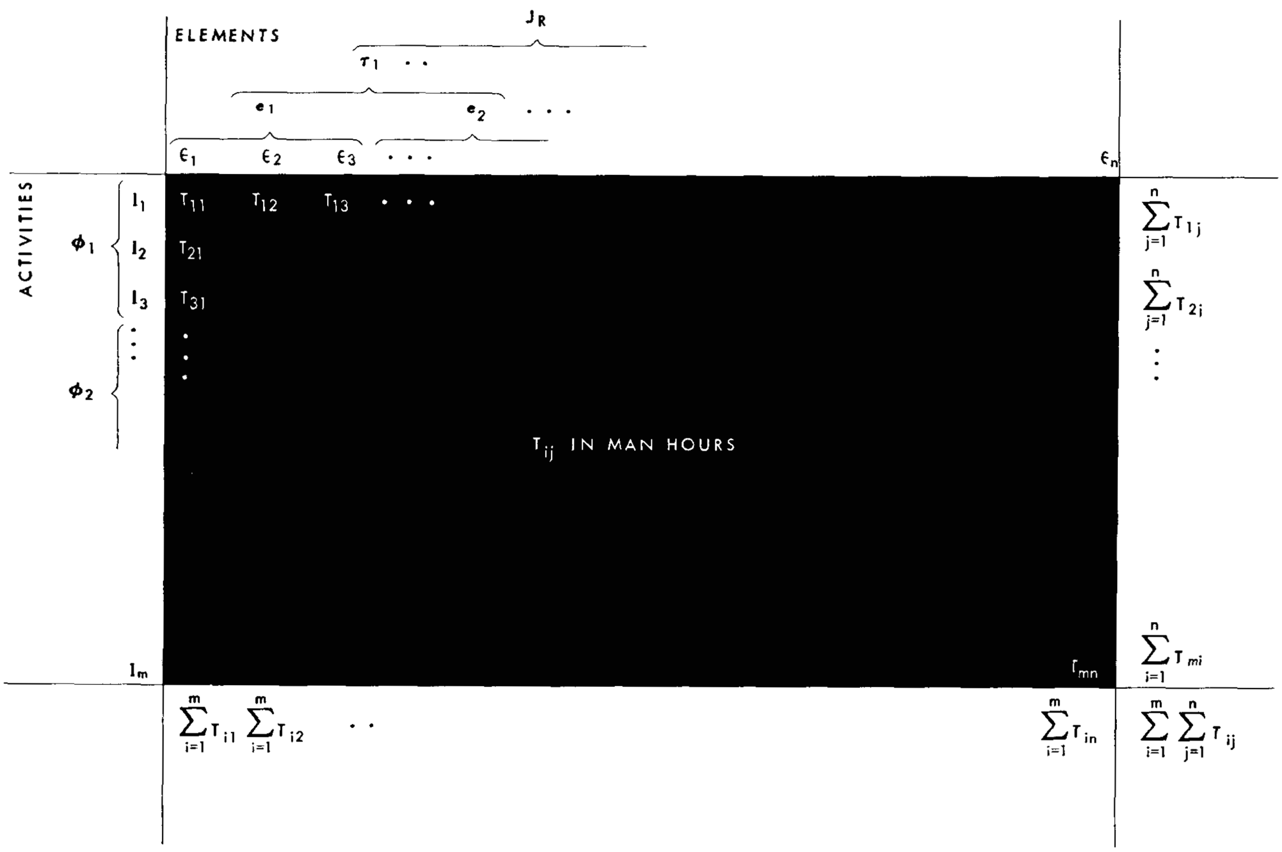
Data-handling procedures have been developed to transfer and process the information from the work place to the array. Effort data are generated as soon as work is performed. Once each day this information is entered at the source on a time sheet which is a sub-matrix of the array and includes only that part on which a given individual is likely to be working.

The research worker or engineer is thus called upon to make a daily judgment as to

a) Which item he worked on;
b) Which activity he pursued;
c) How many hours he spent on a particular combination.

He enters the quantity ($c$) in the appropriate cell of the matrix on his time sheet. The information is validated, suitably coded, and punched into IBM cards. An IBM 704 program has been written which generates the array and produces a print-out of the format shown in Fig. 11.

*Figure 11* **Effort distribution array.**
*Column totals: time spent on each job element; row totals: time spent in each type of activity.*



The accumulated data are stored on tape for re-use the following week. The phase curves and cumulation distributions are then plotted from the row totals and smoothed for monitoring purposes. A separate program has been written for normalizing the recorded man-hours, taking into account differences in productivity of different man-assignment combinations. It may be noted that the concept of individual differences in job performance is important not only to cost estimates but to any prediction system based on the analysis of human effort, and further study in this area is needed.

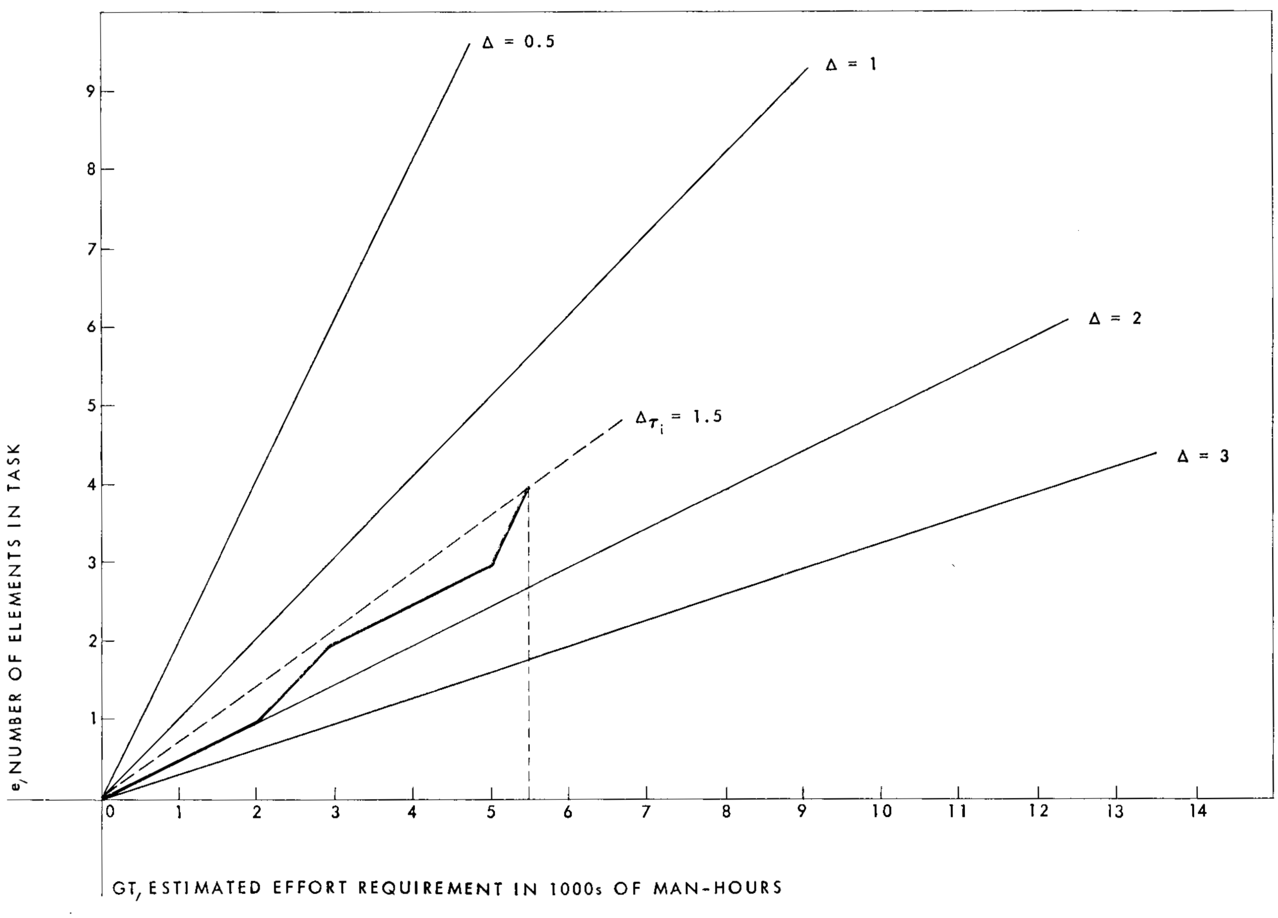
### Estimating and monitoring

Once a reference project has been analyzed as described, the estimation of a new project reduces to specifying a cumulation function $F(t)$ as in Eqs. (2) and (3a). This is equivalent to predicting values for $K$ and for the parameters $a$, $b$, $c$, and $d$. If the method is being applied for the first time, the values of the reference project parameters must be used for the forecast. If a number of projects have been administered in a manner consistent with these procedures, mean values may be superior, provided that their associated variances are sufficiently small. Otherwise the last previous comparable job may furnish more appropriate values.

The estimate for $K$ is based on the terminal cumulative array for the reference project. The effort distribution of the reference project is taken as unity for every element and activity combination of which it consists. This is consistent with the concept of scalelessness and applies equally to elements, tasks, or the entire job. We proceed to build up the estimate of $K$ from estimates for the smallest functionally stable modules, or sub-elements. As previously stated, intelligent choices for these must be based on each company's experience in its own field. In general, estimates synthesized from predictions for each part are considerably more accurate than forecasts of the project *in toto*. The effort-distribution array enables us to perform this synthesis systematically.

Essentially, we compare the new job feature-by-feature with the reference job. The column totals of the reference array show a total effort figure, taken as unit complexity for each element. We can now either estimate the total effort requirements of the elements directly in man-hours, or assign an appropriate measure of complexity $\delta_j$ to each. Sundry companies have developed techniques for this. Epstein[6] treats the problem in terms of a "value function." Another method calls for development of a performance index. Such an index takes into consideration increased levels of performance or output; unusual power

*Figure 12* **Task complexity determined by graphic addition of relative complexity of components.**



consumption constraints; severe size, weight, and cost restrictions; unusual environmental requirements, et cetera. The following may serve as an example in the area of memories for electronic computers. Specifications for these devices often include these considerations:

| *Feature* | | *Direction of Improvement* |
|---|---|---|
| A | Time/memory cycle (secs) | Decrease |
| B | Word length (bits/word) | Increase |
| C | Capacity of unit array (bits) | Increase |
| D | Volume of unit array (in$^3$) | Decrease |
| E | Supply voltage variation tolerance (%) | Increase |

A performance index $\psi$ may be constructed from the above:

$$\psi = \frac{\text{BCE}}{\text{AD}}.$$

The dimensionality of all factors, as well as their functional characteristics, must be statistically determined in each case. However, if a measure such as this is used consistently over an extended period of time, it can become quite valuable for estimating $\delta_j$, since $\delta_j$ is a proportionality coefficient which measures the effort requirements to achieve a given value of $\psi$.

The element estimates may be added either graphically, as in Fig. 12, or computationally. The radiating delta lines refer to the complexity level of the next higher breakdown echelon. That is, if we are synthesizing effort by elements, the result is the net complexity of the composite task. Overall project complexity is analogously computed. For new project features which have no functional counterpart in the reference array we can use either the average value for $\delta_R$ as base, or estimate on practical experience alone. Having arrived at a value for $\Delta_J$, we must select a feasible total time for the project's duration. A family of curves as illustrated in Fig. 13 is useful in this regard. The curves plotted are equilateral hyperbolas defined by the equation $Gt = K_R\Delta$, relating the average effective number of men $G$ to the calendar time $t$ of project duration. The delta curves are thus constant-complexity contours. The dotted lines indicate an example: a hypothetical project of difficulty level 3 is seen to require perhaps 300 men for a four-year period. The curves describe all possible ranges of combinations of men and time to complete a job with a specified amount of total effort. Which one of these is optimal in a given situation depends on practical considerations such as contractual commitments, availability of personnel, competitive release dates, et cetera. Furthermore, the relation is obviously not defined over the entire $Gt$ plane.

$G$ only gives the net rate of application of manpower. Since our rate of effort utilization was shown to follow $\Phi(t)$, we must also examine the peak number of men called for during the project. This is determined by solving $\Phi'(t) = 0$, and may influence our choice for $t_n$. Ultimately a feasible combination of $K$ and $t_n$ is chosen and the task of generating a detailed project schedule remains.

For this purpose, an adaptation of the method developed by Johnson and Turner[5] may be utilized. In examining the relationship between engineering effort and sales income for a given product, both as functions of time, Johnson and Turner found sufficient regularities to postulate the one as stimulus and the other as response, and to apply the Laplace transform to generate a transfer function relating the two. As seen in Fig. 7, we are faced with an analogous situation. We may therefore consider the use of a similar transfer function relating the analytic expressions of the phase curves with the overall function $\Phi(t)$:

$$\theta_i(s) = \int_0^\infty e^{-st}\phi_i(t)\,dt \Big/ \int_0^\infty e^{-st}\Phi(t)\,dt.$$

Since this transformation is reversible, we can obtain the phase curves uniquely by making use of the inverse Laplace transform

$$\phi_i = \mathcal{L}^{-1}\{\theta_i(s)\cdot\mathcal{L}[\Phi(t)]\}.$$

If we had confidence that all projects have closely similar time distribution of effort in time, a simple proportional scaling technique would suffice to determine the new value of $\phi_i$. However, this is not generally the case, and the Laplace transform is a convenient tool which enables us to handle the variations in the shape of $\Phi$ (e.g., skewing) with a single computer program. We proceed on the assumption that, for practical purposes, this transfer function is sufficiently stable between two successive projects for useful predictions to be made. Since the activity content of the several phases is defined (Fig. 2), it is then comparatively easy to develop a Gantt chart compatible with the predicted value for $\phi_i$.

By generating the complete project forecast as phase curves over time, we automatically produce predicted values for the cross-over points of the several phases. Specifically (see Fig. 7) we predict $\lambda_1$ and the accumulated effort

$$\int_{t_0}^{\lambda_1} \phi_1(t)\,dt.$$

At this cross-over point the project has progressed sufficiently for the effort utilization of the second phase to equal that of the first, and similarly for the third and second. In this regard, the point $\lambda_i$ is superior to straightforward monitoring of $F(t)$. The project must pass through these points, and an examination of the actual and predicted points in time, coupled with the associated cumulative effort, provides a check on our initial estimate of $\Delta$. In practice it is desirable to normalize the actual manhour data first by applying a productivity factor. Then a comparison can be made whether, say, a delayed occurrence of $\lambda_i$ -actual was coupled with a greater cumulative manpower figure than predicted. If not, the time schedule can be reviewed to determine whether the project can still be completed as planned, using the same value of $\Delta$, and to see whether the increased level of exertion is feasible. If a significantly greater use of manpower had been logged at $\lambda_i$, we would have to conclude that we had underesti-

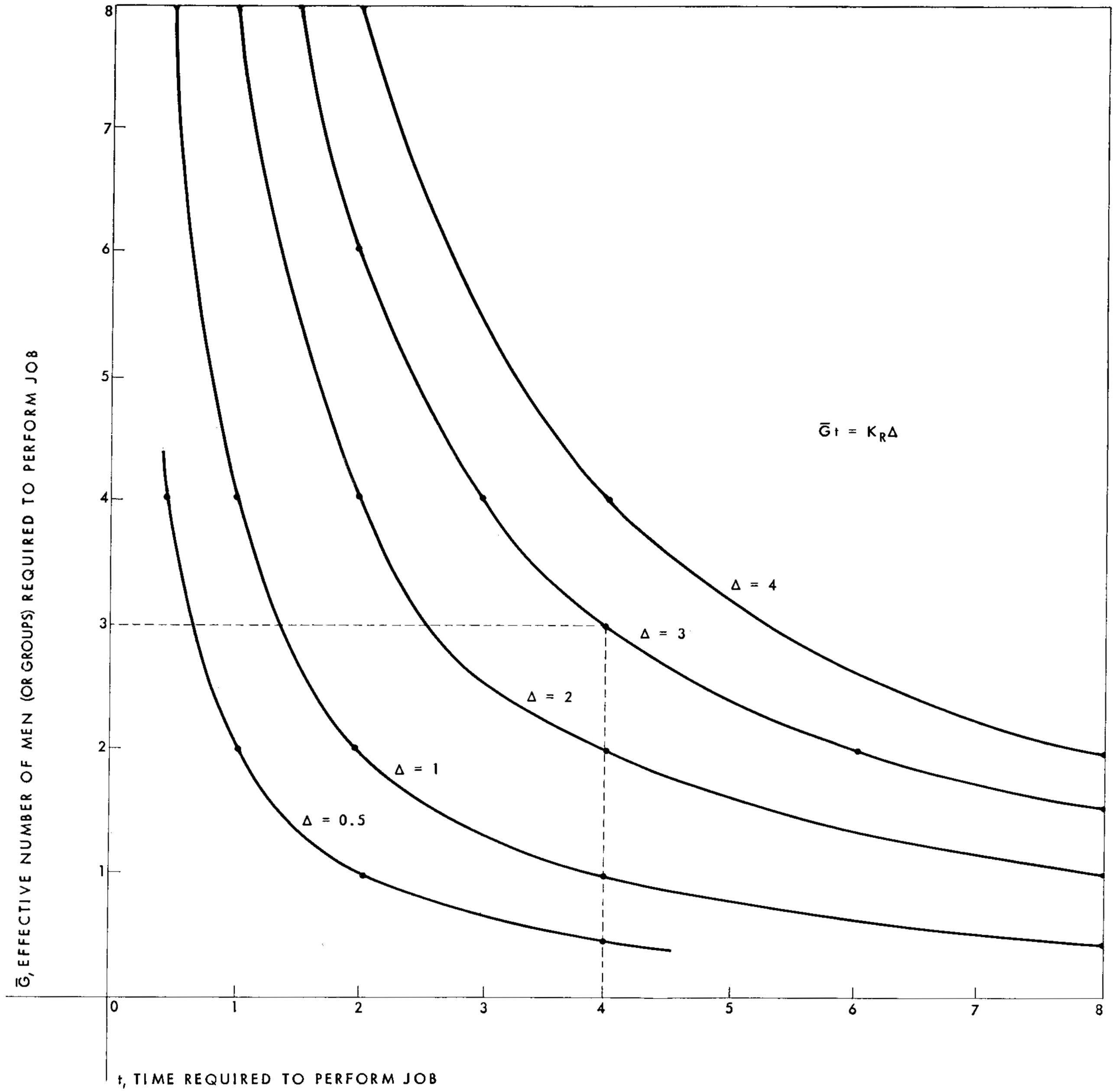



*Figure 13* **Graphical method for selecting time required for project duration.**
*For most companies, this relation is defined only for a limited area near, but not necessarily centered on, the knees of the curves.*

mated Δ, and would then re-schedule with the revised Δ, advising all those concerned of the implications. The required algorithms can be presented graphically or programmed. Since $\lambda_1$ generally occurs when about one-third of either total forecast effort or promised delivery time has been spent, it is a very useful early-warning device. Also, since it is possible to monitor individual tasks and phases independently, we have a flexible tool for changing the basic process as far as feasible, by combining and/or truncating certain activities. Increased familiarity with the behavior of these variables should ultimately lead to increased insight and control over the R & D process.

## Conclusions

The forecasting method described, if conscientiously applied over two or more projects, enables us to compare jobs systematically phase by phase and element by element. It is estimated that accuracy of ±15% of total effort for familiar classes of products under development should be attainable. In addition, generating the effort-distribution array on a computer enables us to monitor project schedules in a matter of days or hours, whereas previously such information was not available for weeks or months, if at all.

Projects to which these techniques are applicable should be sufficiently large for random fluctuations to be expected to cancel. In general, a project should have at least four or five first-echelon tasks and require the services of a minimum of five men for approximately one year or more.

Certain limitations are inherent in the present study. The data were collected in the area of complex electromechanical devices, and the exact choice of effort phases may not necessarily apply to other industries. In addition, the length of time and cost required for analyzing a reference project in this field make the conducting of controlled experiments nearly impossible.

Further study is needed in the following areas:

1. The general problem of personnel interchangeability.
2. Determination of the best composition of a working group in terms of different scientific disciplines.
3. Determination of a unit module of effort, and an optimal economic assignment size for groups of varying magnitude and composition.
4. Collection and analysis of considerably more data to explore the variances of R & D process parameters.

The implications that the natural growth of a development project may be logistic in character are provocative. It means for instance that a rectangular distribution of manpower assigned to a project (one in which we assign a given number of men to the job from the outset, and leave that number unchanged throughout the life of the project) is decidedly uneconomic. Of course this is an extreme example, and not commonly experienced on large projects. But we may speculate that projects which are put on crash programs, or truncated by administrative decree, could also be viewed as arbitrary and presumably uneconomic distortions of a natural pattern of project behavior.

Finally, an underlying assumption of many studies of human effort is the existence of some normal level of exertion. Variations from such a level are certainly conceivable. For instance, there appears to be a tendency for an operating group to converge upon a known completion deadline by consciously or subconsciously adjusting their work speed by a normative mechanism. This ability to attenuate or concentrate work effort to meet an established goal is not infinitely great, of course, and we must be within a certain sphere of forecasting effectiveness for this type of mechanism to operate. Improved forecasting methods and careful monitoring and adjustment can help in bringing applied research and development programs into this sphere on a profitable basis.


## Acknowledgment

Grateful acknowledgment is made to Dr. David B. Hertz, whose encouragement originally prompted the undertaking of this study, and to the Managements of the IBM Research and Product Development Laboratories for their interest and support.


## List of symbols

| | |
|---|---|
| $\delta$ | Complexity coefficient for $\varepsilon_j$ |
| $\Delta$ | Complexity coefficient for $J$ or $\tau$ |
| $e$ | Element (2nd level of job breakdown) |
| $\varepsilon$ | Subelement (3rd level of job breakdown) |
| $G$ | Men, or group, assigned to work |
| $I$ | Activity |
| $J$ | Job or project |
| $K$ | Total effort requirement for $J$ |
| $\lambda$ | Phase lag (crossover) point |
| $\lambda_i$ | All phase lags |
| $\Lambda$ | Lead-time |
| $\mathcal{L}$ | Laplace Transform operator |
| $\phi$ | Activity phase |
| $\phi_i$ | All activity phases |
| $\Phi$ | Rate of effort application for $J$ |
| $\psi$ | Performance index of a piece of equipment |
| $R$ | (Subscript) Reference job |
| $t$ | Calendar time (elapsed time) |
| $T$ | Hours worked (time as a commodity) |
| $\tau$ | Task (1st level of job breakdown) |
| $\theta$ | Transfer function |

## References


1. H. T. Davis, *The Analysis of Economic Time Series,* The Principia Press, Bloomington, Indiana, pp. 247–264.
2. R. Pearl, *Studies in Human Biology,* Williams and Wilkins Company, Baltimore, pp. 373–379 and Chapters XXIV-XXV (1924).
3. R. Pearl, *The Biology of Population Growth,* Alfred A. Knopf, New York, pp. 34–41 (1925).
4. R. Pearl, *Medical Biometry and Statistics,* W. B. Saunders Co., New York and London, Chap. VIII (1940).
5. G. K. Johnson, and I. M. Turner, "Use of Transfer Functions for Company Planning", *Operations Research,* **4,** No. 6, 705–710 (December, 1956).
6. L. I. Epstein, "A Proposed Measure for Determining the the Value of a Design," *Operations Research,* **2,** No. 2, 297–299 (April, 1957).


## Bibliography


A. Abruzzi, *Work Measurement,* Columbia University Press, New York (1952).

R. N. Anthony, *Management Controls in Industrial Research Organizations,* Harvard University Printing Office, Cambridge (1952).

American Management Association, *Rating Employee and Supervisory Performance,* AMA, New York (1954).

E. B. Bowerman, and S. B. Littauer, "Operations Engineering," *Management Science,* **2,** No. 4 (July, 1956).

R. W. Chaffee, "Evaluating Engineers," *Machine Design,* (June, 1951).

J. P. Guilford, *Psychometric Methods,* McGraw-Hill, New York (1954).

R. E. Heiland, and W. J. Richardson, *Work Sampling,* McGraw-Hill, New York (1957).

D. B. Hertz, and G. Feeney, "Operational Organization," *Operations Research for Management,* ed. by McCloskey and Coppinger, the Johns Hopkins Press, Baltimore (1956).

D. B. Hertz, *The Theory and Practice of Industrial Research,* McGraw-Hill, New York, p. 151 (1950).

G. W. Howard, *Common Sense in Research and Development Management,* Vantage Press, New York (1955).

N. W. Kettner, *Linear Programming and Personnel Classification.* Paper presented at the Tenth National Meeting, Operations Research Society of America, San Francisco, California (November 16, 1956).

W. I. McNeill, "Evaluating the Cost and Utility of Industrial Research and Development," *Coordination, Control, and Financing of Industrial Research.* Kings Crown Press, Columbia University, New York (1955).

G. Nadler, *Motion and Time Study,* McGraw-Hill, New York (1955).

J. L. Otis, and R. H. Leukart, *Job Evaluation,* Prentice-Hall, Inc., New York (1949).

A. H. Rubenstein, "Setting Criteria for R & D," *Harvard Business Reviews,* **35,** No. 1, 98 (January-February, 1957).

J. Sherman, and W. J. Morrison, "Simplified Procedures for Fitting a Gompertz Curve and a Modified Exponential Curve," *Journal of the American Statistical Association,* **45,** 87–97 (March, 1950).

W. A. Spurr, and D. R. Arnold, "A Short-Cut Method of Fitting a Logistic Curve," *Journal of the American Statistical Association,* **43,** 127–134 (March, 1948).

W. W. Waite, *Personnel Administration,* The Ronald Press Company, New York (1952).

*Revised manuscript received January 20, 1958.*