Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

66-11,392

BOBIS, Arthur H., 1936-

A SYSTEMATIC EVALUATION OF ALTERNATE CHEMICAL ROUTES.

Newark College of Engineering, D.Eng.Sc., 1966
Engineering, chemical

University Microfilms, Inc., Ann Arbor, Michigan

A SYSTEMATIC EVALUATION OF ALTERNATE CHEMICAL ROUTES

BY

ARTHUR H. BOBIS

A DISSERTATION

PRESENTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE

OF

DOCTOR OF ENGINEERING SCIENCE IN CHEMICAL ENGINEERING

AT

NEWARK COLLEGE OF ENGINEERING

Newark, New Jersey
1966

APPROVAL OF DISSERTATION

SYSTEMATIC EVALUATION OF ALTERNATE CHEMICAL ROUTES

BY

ARTHUR H. BOBIS

FOR

DEPARTMENT OF CHEMICAL ENGINEERING

NEWARK COLLEGE OF ENGINEERING

BY

FACULTY COMMITTEE

APPROVED:

CHAIRMAN _____

_____

_____

_____

_____

NEWARK, NEW JERSEY

APRIL, 1966

VITA

Arthur H. Bobis received an A. B. from Columbia in 1957 and a B. S. in Chemical Engineering from Columbia in 1958. He received the M. S. degree in Chemical Engineering from Pennsylvania State University in 1960. He began his studies toward the doctorate at Newark College of Engineering in 1961.

He is currently employed in the Systems Analysis Section of the American Cyanamid Company at Bound Brook, New Jersey.

ACKNOWLEDGMENTS

The author wishes to thank his advisor, Dr. L. Bryce Andersen, for his direction during the course of this research. Dr. Andersen's suggestions were an invaluable aid in bringing the dissertation to a successful conclusion.

The author is indebted to the American Cyanamid Company where he is employed, and Mr. T. L. Koehler, his supervisor at American Cyanamid, for allowing him to devote fifty percent of his time during the past two years to this research.

The author also wishes to thank the scientists at American Cyanamid Company who provided the information and data that were used to test the concepts developed in this research. The interest of Dr. H. Merlin, who altered his own experimental program to suit the author's needs, is especially appreciated.

ABSTRACT

A systematic method for selecting from among alternative routes for chemical synthesis is developed and presented. Consideration is given not only to making the best decision, but of reaching that decision with a minimum of experiments.

A method is developed whereby the calculated differences in process cost due to variations in the levels of such responses as chemical yield, usage and recovery are displayed graphically. Such a representation by indicating the potential worth of any experiment focuses attention on those parts of the synthesis which have an economic bearing on the ultimate process cost.

The probability of achieving a particular cost is considered next. Cost functions for each of the proposed routes are developed which relate the effects of the process responses to the process costs. The process responses are in turn related to the settings of the process variables through regression equations. A second order experimental design is generated in the important variables. The data from these design points serve as the basis for the regression equation. The responses to the indicated experiments are estimated listing both a best guess and the range around the best guess. As an experiment is run, the experimentally obtained

value replaces the estimated value. The experimentally obtained result and the remaining estimated results are treated identically. The variance about the experimentally obtained value is calculated from experimental error. The variance about the estimated value is calculated from the range about the guessed response. The variance for each of the design points is used to weight the contribution of that point in the analysis.

The results of the analysis are probability distributions for the costs of each process. It is only when the overlap between the probability-cost distributions for competing processes is satisfactorily small that discrimination between routes is achieved. Only by running experiments and replacing estimated results with actual results can the variance about the calculated optimum value be reduced with a concurrent improvement in process discrimination. However, experiments are run only if improved discrimination is required. This feature reduces the total number of experiments required to select preferred routes.

An actual problem is presented in which it is shown how these techniques lead to a process selection on the basis of very few experimental runs.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

The Problem

Since 1957 Research and Development expenditure in the chemical industry has increased by over 60 percent. Although the increase in sales lagged only slightly behind these increases in R&D costs, the profits have risen only half as fast (1). Traditional guide marks, such as the ratio of R&D expenditure to gross national profit, have shown an undesirable downward trend. Charles Allen Thomas, Monsanto's Chairman of the Board, declared two years ago that "R&D is now stumbling in a morass of projects, sinking in a sea of money, and is being built on a quicksand of changing objectives" (2).

In 1966 it is estimated that 1.57 billion dollars (3) will be spent by the chemical and allied industries on research and development. This represents another 10 percent increase over the previous year. There is growing concern as to whether this money will be spent wisely and efficiently. The problems exist at two levels -- the management level where the problems of budget size, budget allocation, and project selection must be properly resolved to insure that the researcher deals with meaningful assignments; and again at the level of the researcher where the problems of discovery, process selection and process optimization must be resolved with a minimum of lost effort.

The research manager is turning more and more to the utilization of quantitative methods for guidance as to where he can best spend his research dollar.

Dean and Sengupta (4) describe their activities on behalf of three major chemical companies which remain unidentified except for the fact that they all share a long experience in R&D and have grown principally by the introduction of new and improved projects. The authors tabulate such parameters as cumulative product research dollars, cumulative process research dollars, total production costs and the ratios of total sales to old sales and the gross value of the plant to total sales as functions of time. They then show that the company's market share could be related to the ratio of R&D expenditure to that of the sales, administrative and general expenditures. Using these data, the authors were able to work out an appropriate budget and an optimum distribution of that budget between product and process research. It is reported that the results have been more than gratifying to the sponsors. Freeman (5) developed a stochastic model which can be used to assess the expected return from research projects, Hess (6) presents a set of recursive equations which are useful for both project selection and budget distribution. Hess' model reappears in a paper by Saunders (7) who reports its modification and utilization by Monsanto. Asher (8) reports the use by Abbott Laboratories of his linear programming model for the allocation of R&D efforts. Bobis and Sprague (9) report an effort to combine the advantages of management experience with that of mathematical models by having the Research Manager work directly with the model via a computer programmed in "conversational mode." This effort is currently being pursued at American Cyanamid. Horowitz (10) presents a regression equation by which he tries to understand the influences of current sales levels, profits and the

accumulated previous research expenditure on the size of the annual research budget and the influences of the size of that budget on future profits. Approaches such as these hold out the promise of allowing the manager to make more meaningful decisions as to the appropriate expenditure of his resources.

A quantitative approach should also be of use to the research worker in deciding how he can experiment efficiently. When a choice between alternatives exists it is important to find the best choice as soon as possible and to experiment in such a way that the inadequacy of any of the alternatives is quickly determined. This work is concerned with the development of a quantitative approach to be used for early discrimination between alternative routes for chemical synthesis.

A typical problem of selecting between alternatives is shown in Figure 1-1. Four routes: A, B, C and D are shown as horizontal lines. The length of any one line is a measure of the dollars spent to reach that particular stage of the investigation. In each case the investigation is continued until it either becomes clear that the route is not feasible at which point it is compared with all other surviving routes. Since only one route will ultimately survive, all of the research effort expended on its competitors is of limited value. The problem is not one of optimization but rather one of discrimination and it should be possible to discriminate at suboptimal levels.

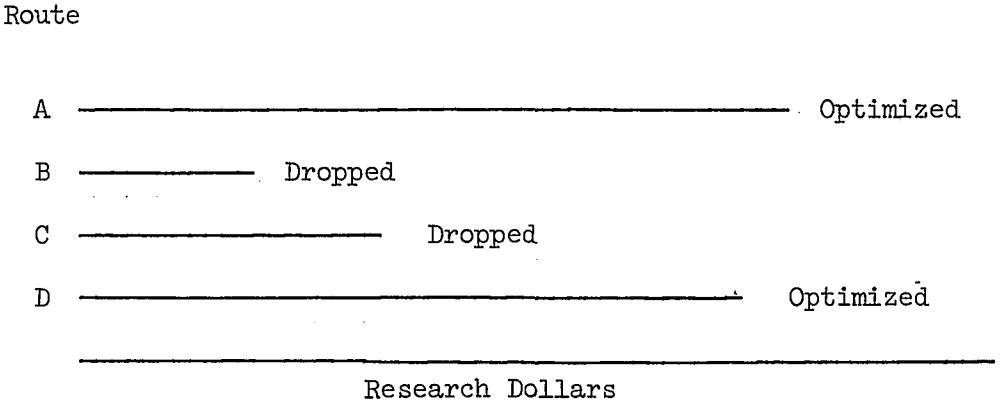

Fig. 1-1
Hypothetical Research on Four Alternates

Continuing experimentation up to the point of optimization for routes which do not succeed is only part of the problem. The other is associated with the failure to terminate experimentation which even if totally successful could no longer have had any bearing on the economic decisions of continuing with the program. In such cases it would appear that the experimenter while pursuing his particular interest often loses sight of the original economic objective -- even when that objective is merely the reduction of costs below some stationary bench mark -- the cost of the present process. His accomplishment, which always could have been measured against that bench mark, was in fact so measured only at periodic reviews and not on an experiment-to-experiment basis. The problems of this experimenter are much greater in a dynamic situation where his accomplishment must be compared against the moving bench mark -- the case where several other experimenters are simultaneously pursuing competitive routes.

What is needed is some technique that will keep the experimenter continually informed as to the economic aspects of his problem and the status of all related approaches, while at the same time highlighting those areas of the problem where additional research would be beneficial. In Chapter 2 a graphical representation of the effect of changes in the process responses on the process costs is developed. The graph which can be drawn prior to any experimentation can be used to assess the economic potential of each experiment. The rest of the thesis is devoted to resolving the companion problem of determining the probability that the economic potential will be realized.

The techniques to be discussed represent a formalized approach to the problem of selecting between alternative routes. The methodology employed in the resolution of these problems was designed to be used with a minimum -- and in fact in the absence -- of experimental evidence. For this reason, decisions concerning the continuance or discontinuance of work on an alternative route can be made at the earliest possible moment.

Background

The basis for this work derived primarily from the science of experimental design, founded by Sir Ronald Fisher, which deals mainly with problems of estimating constants and estimating the differences resulting from treatments in comparative experiments. Finney (11) indicates that what is meant by design of experiments is:

(1) The specification of the response or the measurements to be made on each experiment.

(2) The selection of variables whose effect is of concern.

(3) The settings of the levels of those variables.

(4) Determining the combinations of the levels of the variables at which the experiments will be run.

Prior to 1923, the main concern of the statisticians was that of analyzing data. At that time, Fisher started to publish a series of papers which led to his book The Design of Experiments (12). Two important principles to emerge from this treatment are that

> ... the design of an experiment in great measure determines the form of statistical analysis appropriate to the results ... the success of an experiment in answering the questions that interest the experimenter without excessive expenditure of time and resources depends largely on the choice of design.

Fisher at this time was interested in the problems of agronomy. The earliest designs, therefore, catered to the specific needs of the agricultural experimenter. Primary among these were problems of soil variation and the elapsed time between planting and harvesting. For this reason, these designs placed heavy emphasis on redundancy, randomization and the simultaneous inclusion of a large number of treatments. The Latin square design first described by Euler in 1782 was recognized by Fisher as having properties that would satisfy the agriculturists' demands. In these designs, the experimental laboratory which was an open field was treated as a 2-way table of rows and columns. Protection against confounding the effects of treatments with that of soil heterogeneity was accomplished by insuring that each

treatment appeared in a random position in each row and column. Discussion and analysis of these designs are given in several references (13, 14).

A class of designs of even greater utility is the factorial design. In these the treatments consist of the combinations of levels of two or more different variables. If m levels of n variables are to be included, the design will be described as an m^n factorial. Fisher states the advantages of these designs are:

> Every trial supplies information about each of the main questions that the experiment is designed to examine ... In addition to being able to measure the effects of each of the variables with the same precision, as though the whole of the experiment is devoted to each of them, it measures all possible interactions between these ingredients with the same precision.

Despite the power of these designs, the large number of experiments that result when either "m" or "n" is large, militated against its use. Fisher recognized this drawback and discussed the concept of confounding or fractional replication. He points out that if there is sufficient reason to doubt the existence of the highest order interactions, other variables can be run at the levels which would have been used to calculate those interactions. A treatment of confounding is given by Davies (15).

Often an experimental program can be subdivided into a series of blocks by setting the block contrast equal to higher order interactions, thus eliminating the necessity of performing all the experiments in an

identical environment, which adds to the flexibility of the factorial designs. Finney (16) provides the necessary mathematics.

Despite the body of knowledge generated by Fisher and his peers, workers in fields other than agriculture, and particularly those in the chemical industry, did not take advantage of this technology.

Failure on the part of the chemical researcher to exploit these accomplishments can be attributed to the nature of the two experimental problems. The agronomist runs his experiments simultaneously; the chemical researcher runs his sequentially. The agronomist is concerned most often with qualitative variables and is thus looking for optimum combinations of variables. The chemical researcher is most often concerned with quantitative variables and thus is looking for optimum settings of these variables. The agronomist operates in the face of nature and his ability to reduce his error is limited. He invests his resources effectively by running redundant experiments. The chemical researcher creates his own environment which is insulated against nature, so that he spends his resources effectively if he invests in control.

In 1951, Box and Wilson (17) published a paper in which they demonstrated that Fisher's designs could be effectively utilized and serve the specific needs of the chemical worker. They proposed a sequential approach to the problem of optimization. Using the method of steepest ascent, they described the appropriate design strategies for an experimental region which was remote from the

optimum area and then presented another strategy to be used when the experimental region was in proximity to the optimum. Their basic strategy was to fit a regression equation to the results of a small experimental design and then to differentiate the resulting equations with respect to the variables and thereby find the direction in which the optimum lay. The experimenter continues to experiment along a line in the indicated direction until the response falls off. Then a new design is constructed and the process is repeated until a maximum is found.

The criterion for the best design for fitting a regression equation is that the coefficients in the model be separately estimated with minimum variance. In 1952, Box (18) showed that for models linear in the variables this criterion is satisfied when the design is so chosen that each effect may be independently estimated. Designs with this property are said to be orthogonal. Two level factorials and their fractional replicates are examples of this class of designs.

In selecting designs to fit polynomials of a higher degree, Box and Hunter (19) introduced the concept of rotatability. Designs of this class have the property that the estimated response has a constant variance at all points which are the same distance from the center of the design. The designs for fitting second order polynomials consist of a two level factorial augmented with runs at the center and with two equiradial points outside the factorial for each of the variables.

Although the form of this design is fixed the scale of the design in the space of the variables and the number of center points can be varied. Box and Draper (20, 21) have shown how to choose the values for these two parameters of the design so that the polynomial best describes the response over a given region of interest, even if a polynomial of higher degree would describe the surface more accurately.

In a development which was directed at the chemical experimenter's needs to experiment sequentially statisticians suggested designs utilizing small blocks of experiments. Box (22) offered a paper on the integration of techniques in which he stated his philosophy concerning the growth of knowledge through a continuous iteration between hypothesis, design, analysis and hypothesis. The concept of running experiments in small blocks which could be augmented with additional runs, or supplanted by an entirely new strategy, without having sacrificed too many experiments, was stressed.

In two papers in 1961, Box and Hunter (23, 24) presented the 2^{k-p} designs in which they emphasized designs of Resolution V. These are fractionated designs which permitted independent estimates of main effects and two factor interactions. For example, five variables could be studied with a Resolution V design with 16 runs. DeBaun (25) also discussed blocking strategies for rotatable designs. The saturated designs of Plackett and Berman (26) utilized a basic block size of four. Such designs as the 2^{3-1} were a great attraction to chemists

since the confounded interactions could often be interpreted from either physical relationships or merely by the size of the main effects. Daniel (27) offered a method for clarifying confounded interactions by running two experiments in which the effect appeared at opposite levels of the contrast. Hunter (28) presented a mathematical formula which allowed clarification of such confounding with the addition of only a single experiment. This strategy allows the experimenter to run his basic block and then continue his experimentation by adding one experiment at a time.

An attempt is made in this dissertation to provide the experimenter with a design which minimizes his commitment to the number of experiments. The basic structure of the designs that will be discussed will either be central composite or rotatable designs, yet the experimenter will be free to revise his strategy as the results of each experiment become known. The experimenter will be asked to provide estimates of the results of indicated experiments which will be used in the analysis along with experimentally obtained values. Second order response surfaces will then be fitted.

The experimenter is called on to augment the experimental results with subjective judgment. In a review of subjective judgment reported by Cohen (29), he illustrates several studies which show that the subjective judgment of an individual increases as he becomes familiar with the situation that he is judging.

Savage (30) states that the analysis of most problems can be aided by subjective judgment, and that when faced with problems of uncertainty it is often his only recourse. Dean (31) states that

> Adjustments to allow for uncertainty may be challenged as nothing more than guesses. Perhaps they are, but even so they are guesses that must be made and will be made either explicitly or implicitly.

Hertz (31) in a discussion on risk analysis shows the necessity for employing subjective probability by pointing to the serious shortcomings that occur in any economic analysis in which such estimates are omitted. In discussing estimates of unknown factors, Hertz states that

> ... the range is relatively easy to estimate; if a guess has to be made -- as it often does -- it is easier to guess with some accuracy a range rather than a specific value.... the ranges are directly related to the degree of confidence that the estimator has in his estimate.

It is the author's contention that subjective probability can be used in areas of experimentation. The experimenter gains more knowledge by running experiments and as he does so, his ability to make accurate estimates will increase and his stated range about those estimates will decrease. Early discrimination between alternate chemical routes will rely heavily on both the concepts of experimental design and of subjective probability.

The background presented here has been general in nature. Specific detail about techniques utilized in the work are presented within the body of the text.

Chapter 2

THE VALUE OF THE EXPERIMENT

In any research, process development or process improvement project, there is a period between the time that the decision to invest in technical effort has been made, and the time that actual experimentation begins, when the experimenter attempts to decide which of several alternate routes should be investigated and in which order they should be investigated. The experimenter would hopefully choose that alternative that would allow him to fill his objective at maximum gain and to order his work so that the most critical areas are studied first.

The technique is here developed to aid the experimenter in reaching these decisions. The key to the technique is the recognition that while it is impossible to compute the process cost accurately early in a research project, it is usually possible to compute accurate differential costs. As a result, the possible effect of uncertainties about raw material costs and utilization, yields, recoveries, labor utilization and capital requirements on the overall project economics can be assessed and most important, the assessment can be made before any experimentation has started.

Several examples can be cited. An experimenter looking at a particular reaction step in one route believes that the results of the literature search shows that his yield might be as low as 70% or as high as 90%. Although he does not know what his ultimate yield will be, he certainly can calculate the value of the increase in yield.

He may know from experience that he will recover somewhere between 95 and 99% of the solvent, and although he does not know how much he will actually recover, he can compute the worth of the additional 4%.

Similar questions arise as to labor utilization and capital cost, but in the present treatment, attention is focused only on raw material costs which do in fact comprise a large portion of the total manufacturing cost.

If the differential cost is calculated for all the areas of uncertainty within each route, the experimenter will know how many significant economic levers there are inside each route. He can then use his judgment to decide upon the amount of time that should be devoted to finding optimum settings of these levers. He can, of course, introduce his notions of the probability of success into his general assessment.

Development of Cost Equation

That portion of the raw material cost contributed by the chemicals which participate stoichiometrically is given by:

$$\sum_{i=1}^{n} \left(\frac{E_i M_i}{Y M_p} - \left(\frac{E_i M_i}{Y M_p} - \frac{M_i}{M_p} \right) R_i \right) C_i \qquad 2\text{-}1$$

Where: M_i is the equivalent weight of the participating chemical

M_p is the molecular weight of the product

Y is the chemical yield of this reaction step

R_i is the recovery of the particular chemical

C_i is the cost of the particular chemical

E_i is the ratio of the input level of the participating chemical to its equivalent weight

Only the values M_i and M_p are constant, whereas the other parameters, R_i, Y, C_i and E_i can take on a range of reasonable levels. The term in the brackets $\left(\frac{E_i M_i}{Y M_p} - \frac{M_i}{M_p} \right)$ is the difference between the quantity of material charged per pound of product formed and the amount of that material that ends up in a pound of product. In an ideal case, this represents the amount of material available for recovery. However, most often owing to byproduct formation, something less than this quantity is available for recovery. Since this is only an estimate, the uncertainty as to the amount of reactant consumed in byproduct formation can be expressed by a compensatory adjustment of R_i. If this is not satisfactory, a more exact equation of the following form can be used in which the concept of conversion is introduced.

$$\sum_{i=1}^{n} \left(\frac{E_i M_i}{X_i Y M_p} - \left(\frac{E_i M_i}{X_i Y M_p} - \frac{M_i X_i}{X_i Y M_p} \right) R_i \right) C_i \qquad 2\text{-}2$$

Where: X_i is the fraction of reactant which is converted

Y is the fraction of converted material that goes to product

Although expression 2-2 is a more exact representation of the actual situation, it introduces an additional variable, conversion

which must also be estimated. It has been found that generally the simpler expression is preferred.

The cost contributed by solvents, surfactants, catalysts and buffers can be calculated from the estimates of the ratio of the charge of these chemicals to that of one of the chemicals participating in the stoichiometry. Relationship 2-3 is used.

$$\sum_{j=1}^{n} E_j U_1 (1-R_j) C_j \qquad 2\text{-}3$$

Where: C_j is the cost of the particular solvent catalyst or buffer

E_j is the ratio of the weight of the particular solvent, catalyst or buffer to the weight of Reactant 1

U_1 is the usage of Reactant 1 calculated from expression 1-1

R_j is the recovery of the particular chemical

Credit for byproduct formation can be calculated from the expression:

$$\sum_{k=1}^{p} \left(\frac{E_k M_k}{M_p} \right) C_k \qquad 2\text{-}4$$

Where: E_k is a factor which explains the stoichiometric excess or deficiency of the byproduct

C_k is the credit for the byproduct

The cost of any intermediate chemical or product can now be calculated from the summation of expressions 2-1, 2-3, and 2-4 and this is given in expression 2-5.

$$\text{Cost} \sum_{i=1}^{n} \left(\frac{E_i M_i}{Y M_p} - \left(\frac{E_i M_i}{Y M_p} - \frac{M_i}{M_p} \right) R_i \right) C_i$$

$$\sum_{j=1}^{n} E_j U_1 (1-R_j) C_j - \sum_{k=1}^{n} \left(\frac{E_k M_k}{M_p} \right) C_k \qquad 2\text{-}5$$

These equations can readily be used to calculate the anticipated cost of any particular settings of the variables Y, C, R, and E. There are so many areas of the research in which the estimates are not firm, the effects of changes in the levels of the estimated parameters often must be considered in combination, and many of the effects are not linear; so that a graphical output becomes a practical means of assessing the results. The complexity and tediousness of plotting the cost as functions of the many parameters particularly in multistep processes leads to the necessity of developing a computer routine which will serve the dual purpose of performing the calculations and plotting the results. A program that utilizes the equations that have just been presented is shown in Appendix A. It is a 1620 program written in Fortran II. The curve plotting routine is shown in Appendix B. This, too, is a Fortran II program written for the 1620; the plots are completed on an IBM 870 autoplotter that uses the cards generated from the 1620 program.

A One-step Process

For a given set of C_i, R_i and E_i, expressions 2-1 and 2-3 can be reduced to the form:

$$\text{Cost} = \frac{K_1}{Y} + K_2 \qquad 2\text{-}6$$

Where: K_1 and K_2 are arbitrary constants associated with those chemicals that participate stoichiometrically.

The byproduct expression at a fixed value of E_k and C_k is reduced to a constant. Therefore, at fixed levels of all the R's, C's and E's, equation 2-5 maintains the general form and can now be represented as

$$\text{Cost} = \frac{CAS}{Y} + CAI \qquad 2\text{-}7$$

Where: CAS and CAI are arbitrary constants associated with the sum total of reactants, solvents and byproducts.

The cost is shown to vary linearly with 1/Y. Alternatively, once the values of the constants are established, the cost can then be calculated for any value of Y and a curve relating cost to yield could be plotted. If the cost of A is known at two different yields of A, the costs may be designated CSTAI and CSTA2 and the yields YA1 and YA2. The constants in equation 2-7 can be calculated from equations 2-8 and 2-9.

$$CAS = (CSTA2 - CSTA1)/(1/YA2 - 1/YA1) \qquad 2\text{-}8$$

$$CAI = CSTA1 - CAS/YA1 \qquad 2\text{-}9$$

With CAS and CAI established, the lines can now be computed and plotted for each combination of the E's, C's and R's. The final cost need merely be known at any two yield levels for each of the combinations involved.

The difference between the resulting lines measures the economic effects of changes in these process parameters.

Multistep Process

For a multistep process, the procedure is more complex since the cost of any successive step depends not only on the conditions of the particular step, but also on the cost of the chemical produced in the previous step. The cost of the particular step (Step B) must be devined not only at some yield level of that step and settings of the E's, C's and R's but must also be defined in terms of the conditions of the prior step (Step A). The cost of B is shown in Equation 2-10.

$$\text{Cost B} = \left(\frac{E_1 M_1}{Y M_p} - \left(\frac{E_1 M_1}{Y M_p} - \frac{M_1}{M_p} \right) R_1 \right) C_1 + \sum_{i=2}^{n} \left(\frac{R_i M_i}{Y M_p} - \left(\frac{R_i M_i}{Y M_p} - \frac{M_i}{M_p} \right) R_i \right) C_i \qquad 2\text{-}10$$

The equation is simplified in that the contribution of solvents and byproducts is not included; however, the general form of the equation still applies.

The term to the left of the plus sign in equation 2-10 refers to the chemical made in the previous step, while the term to the right of the plus sign pertains to all the other chemicals introduced in Step B. By the same logic developed for the one-step process for fixed E_i, R_i and C_i, and in this case at fixed levels of Y, equation 2-11 would result.

$$\text{Cost B} = \left(\frac{BSS}{YB} + BSI \right) \text{Cost A} + \frac{BIS}{YB} + CII \qquad 2\text{-}11$$

Where BSS and BSI are constant for fixed E_1 and R_1, and BIS and BII are constants for fixed E_i, R_i and C_i .

The cost of A is shown as a variable to indicate that it is a function of the conditions of Step A. Therefore, at a fixed value of YB, one obtains the straight line relationship 2-12.

$$\text{Cost B} = S1(\text{COSTA}) + I1 \qquad 2\text{-}12$$

where I1 and S1 are arbitrary constants calculated at a particular value for YB. It is now possible to calculate the values of the constants in 2-12 if the cost of B is known at two different costs of A. Call the costs of A, CSTA1 and CSTA2 and those for B, CSTB1 and CSTB2.

$$S1 = (CSTB2 - CSTB1)/(CSTA2 - CSTA1) \qquad 2\text{-}13$$

$$I1 = [CSTB1 - S1(CSTA1)] \qquad 2\text{-}14$$

Equations 2-13 and 2-14 can be used to solve for the constants in Equation 2-12. If the costs of B are known for the same two costs of A at another level of the yield of B, the constants for another straight line could be calculated.

$$S2 = (CSTB4 - CSTB3)/(CSTA2 - CSTA1) \qquad 2\text{-}15$$

$$I2 = CSTB3 - S2(CSTA1) \qquad 2\text{-}16$$

The purpose is, of course, to express the cost of B as a continuous function of the yield of B. This can now be done by establishing the relationship between the constants S and I and YB.

Since the cost of B for a fixed cost of A can be expressed as a linear function of 1/Y, the two constants S and I can be separately expressed as functions

$$S = BSS/Y + BSI \qquad 2\text{-}17$$

$$I = BIS/Y + BII \qquad 2\text{-}18$$

Where $BSS = (S2 - S1)/(1/YB2 - 1/YB1)$

$BIS = (I2 - I1)/(1/YB2 - 1/YB1)$

$BSI = S1 - BSS/YB1$

$BII = I1 - BIS/YB1$

These values can now be directly substituted in equation 2-11.

Additional steps can be added in similar fashion.

The plot for the cost of a multistep process can be generated from a knowledge of the cost of the last step at two different yield levels of that step and two different costs of the intermediate produced in the previous step. The same information must be available for all of the previous steps. These results depend on the settings of the C_i's, E_i's and R_i's and new constants must be used if changes in those settings are to be evaluated.

The phase of the work can be most useful in determining experimental priorities and the extent of the research commitment. When changes in the parameter levels are plotted against cost, the output resembles a sensitivity analysis, where the approximate slope of the various lines is a measure of the importance of the particular parameter (33). There are, however, two major distinctions between this analysis and a sensitivity analysis.

(1) The sensitivity of the cost to changes of the parameters is used to establish experimental priority rather than merely to indicate the resultant variation in the response.

(2) An attempt is made to force the values of the parameters toward the high side of the range by experimentation rather than to try to reduce the range by improved sampling.

An illustration of how the graphical cost analysis can be used is provided by the following example. The equations and curve plotting routine just described were used to process information supplied by experimenters at the American Cyanamid Company. The decisions which were reached by the experimenters are, of course, their own. The graphical analysis which was developed in this dissertation provided the basis for those decisions.

An Example

The example comes from a cost reduction program. As in the case of all the examples, the data will be coded. Two routes, I and II, were considered. Route I involved a two-step synthesis, both steps of which were relatively uncomplicated and extensive literature references to similar reactions allowed the experimenter to make estimates of the possible range in the yields of the individual steps. In addition to the uncertainty about the yield, there was uncertainty about the ability to recover a byproduct and uncertainty as to whether the company could develop a lower cost process for one of the purchased raw materials. The cost analysis utilizing the equations just developed is shown in Figure 2-1. The yield for the formation of the intermediate product is shown along the abscissa. The yields for the second reaction are shown by the dashed and solid lines. There are three lines in each group -- each one representing a different estimate. The topmost line refers to the use of raw material (T) at its purchased price. The middle line is based on an estimate that this raw material could be manufactured internally at a reduced cost. The lowest line is

based on an assumption that in addition to making T internally, 90 percent of the low cost byproduct could be recovered.

The very earliest experiments resulted in a yield for the first step of about 81 percent and a second-step yield of about 86 percent. This result is located on the figure by the letter a. At this point a decision had to be made as to what was the best way to continue with the research.

Analysis of Figure 2-1 indicated that the preferred strategy was to try to optimize the yield of the second step as this offered a potential savings of about 30 cents per pound (a to c), while optimization of the first step would only have been worth 15 cents per pound (a to b). This work was in fact successful and yields of better than 95 percent were obtained.

At this point the decision between optimizing Step 1 or investigating an alternative synthesis of chemical T had to be faced. Since the potential gain for both choices was comparable, the decision was based on the experimenter's belief that the optimization of Step 1 could be accomplished with a lesser effort. Again he was successful and yields of 93 percent were realized.

Experimentation through the path a c d insured that the critical areas of the work were investigated sequentially. This was particularly important since competitive routes were also under consideration. Failure to increase the yields of the first step would not have been

Differential Cost Analysis of a Route
Containing Four Responses of Interest

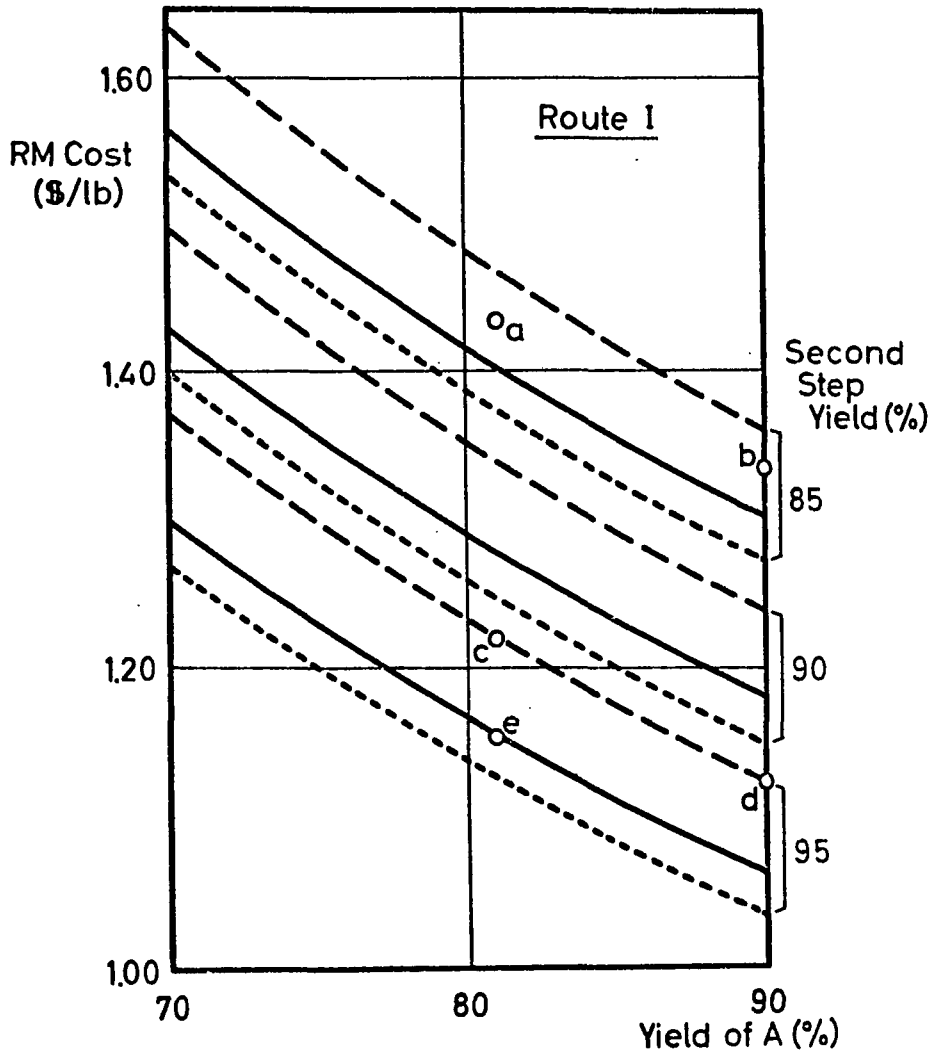

Fig. 2-1

T ($/lb)	By-product Recovery (%)	
.34	0	– – – – –
.17	0	———
.17	90	----------

decisive insofar as terminating work on route I, whereas failure to increase the second step yields up to optimum levels would have been.

Figure 2-2 shows the cost analysis for the competitive route II. This was also a two-step synthesis, requiring the formation of B and subsequent reaction of B in the second step. It was expected that some unreacted B could be recovered. In the figure, the finished raw material cost is plotted on the coordinate axis and the cost of the product made in the first step is shown on the abscissa. It is interesting to note that the cost of B rather than the yield of B was the parameter which was plotted. A separate analysis for the intermediate step had shown that efficient solvent recovery and utilization of by-product material was as important as yield; and thus, the cost was a much more convenient parameter to use.

The experimenter found literature references to two reactions which were similar to that of the second step -- one reporting a yield of 52 percent and the other a yield of 81 percent. These values are plotted in the graph. The 67 percent figure shown in the graph is about in the middle of these two and aids in the interpolation. Similar reasoning permitted the selection of the 30-75 cent range for the manufacturing cost of the intermediate and the 50 percent recovery of unreacted B was the best that the experimenters hoped to obtain.

It is necessary to specify the levels of all the parameters at which the effects of a single parameter is computed as a result of the interactions between the variables. An interaction can be recognized

easily because of the non-parallelism of lines and can, of course, have a strong influence on the experimental strategy. One can read from the graph that the yield increase from 52 to 81 percent would be worth about 70 cents per pound when the cost of B is 30 cents per pound and it would then be worth better than one dollar when the cost of B is 75 cents per pound. The reduction in the cost of B from 75 cents to 30 cents per pound is worth about 40 cents at the 81 percent yield level and about 50 cents at the 52 percent yield level. Also, at the 81 percent yield level the recovery is worth only 3 cents per pound; whereas, at the 52 percent level it could be worth as much as 30 cents per pound. This point will be referred to again later.

In Figure 2-3 routes I and II are shown side by side on a graph in which the abscissa for route I is the yield of A as shown before and that for II is the cost for B as previously shown. The ordinate is common for both and is the finished product raw material cost in dollars per pound. Because this graph was available before any experimentation was done, the experimenters were in a position to choose that route which should receive the major attention and the portions of that route which should be studied first. Route II appeared to have the better overall potential and was assigned a higher priority. The high priority of route II existed only as long as the cost of intermediate B could be maintained below 60 cents per pound. Early experimentation showed this not to be the case and work was immediately abandoned and the emphasis was shifted to route I. Note now, that any effort expended on recovering the unreacted intermediate would only be of value when

Differential Cost Analysis of a Route Containing
Three Responses of Interest

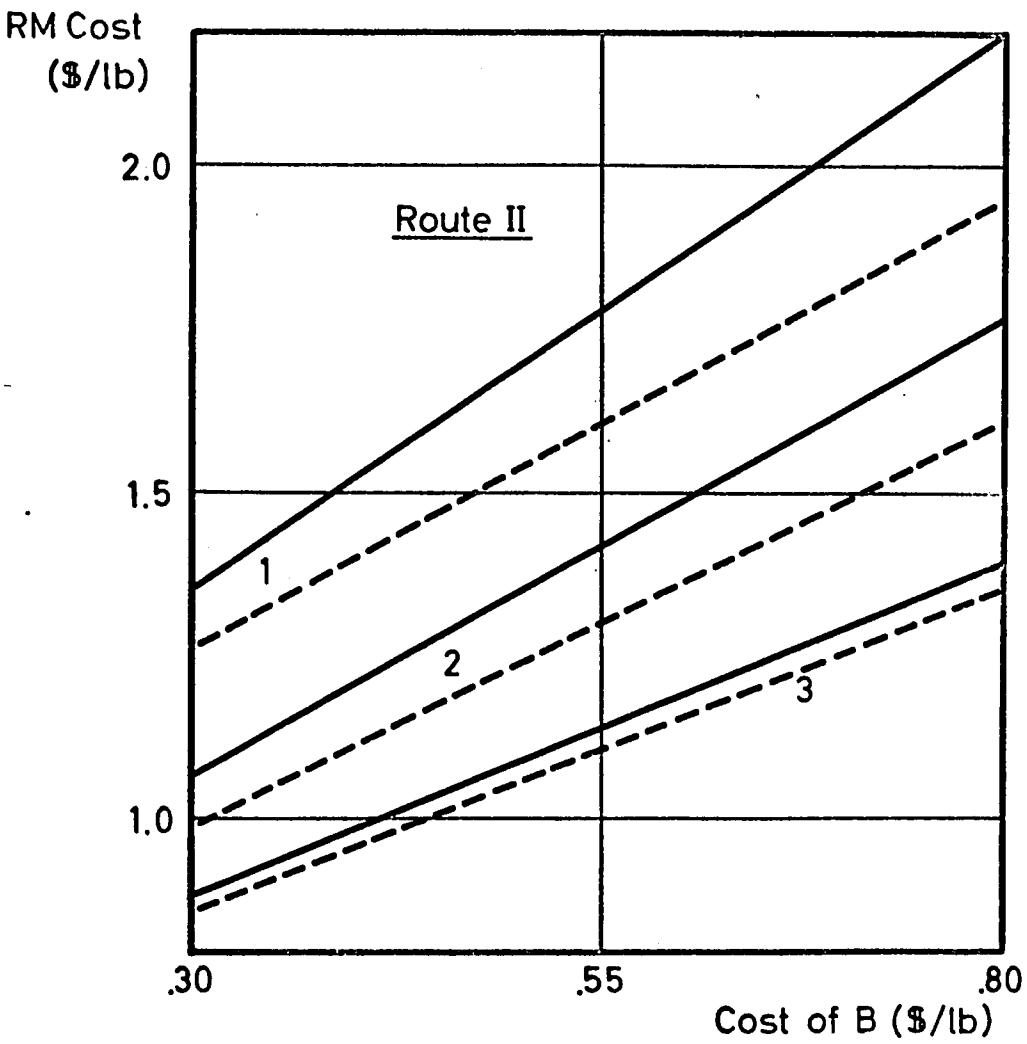

Fig. 2-2

Recovery of B (%)	Second Step Yield (%)
0 = ———	1 = 52
50 = - - - - - - -	2 = 67
	3 = 81

Comparative Analysis of the Two Routes

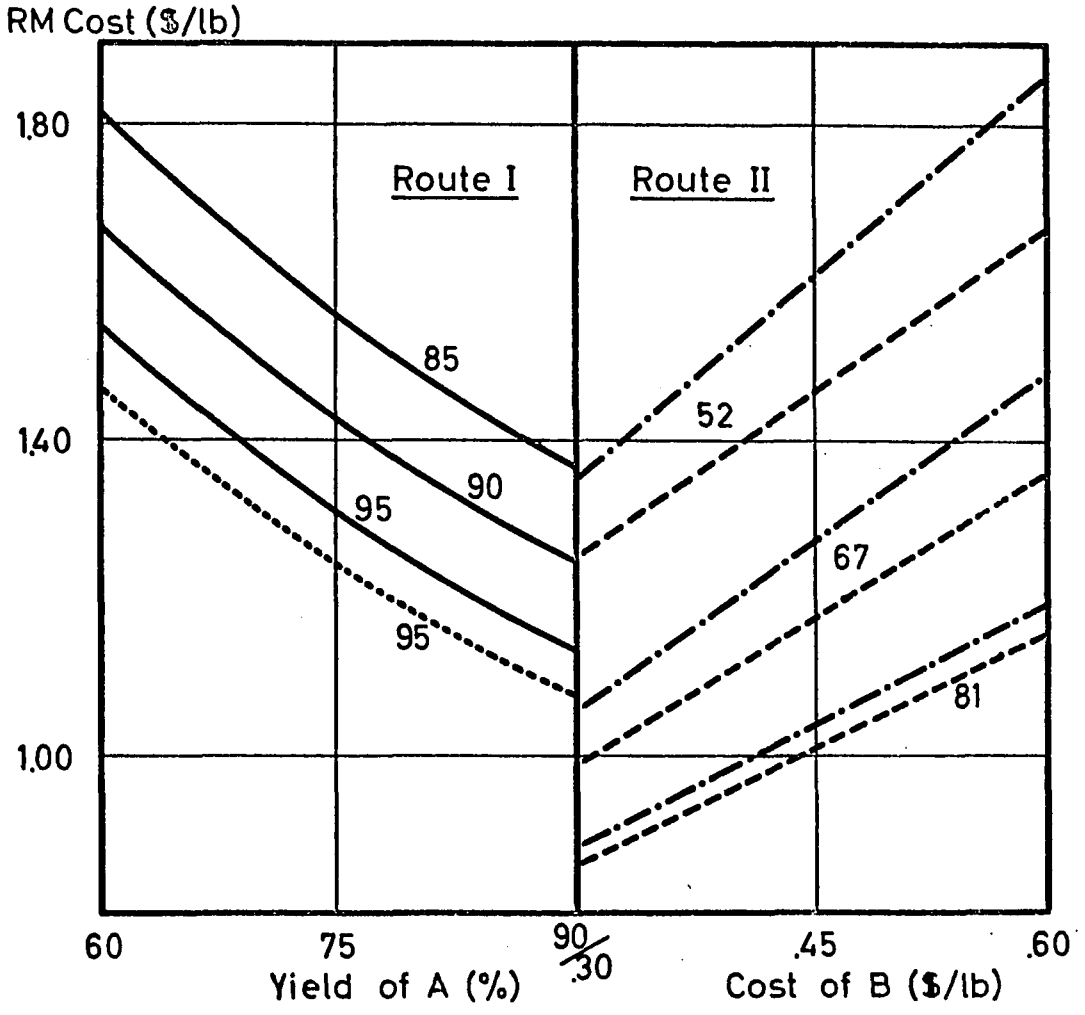

Fig. 2-3

Parameter: Second Step Yield (%)

T ($/lb)	Recovery of B (%)
.34 ———	0 —·—·—·
.17 ·····	50 - - - - - - -

the cost of B was high and the yield low. At these conditions route I would be the route of choice, no matter how successful the work was.

The graphs provide information as to the value of an experiment were it to be successful. The remainder of the dissertation treats the problem of quantitatively defining the probability that the experiment will be a success.

Chapter 3

THE PROBABILITY OF SUCCESS

Any attempt to assign a quantitative statement of the probability of success to the objectives within a chemical route must begin with the experimenter or experimental team who, through their combined background reading and experience, become the best source of information regarding the particular synthesis. It is expected that the experimenter will exercise sound judgment in planning his experiments and reporting on the status of the project; such judgments include assessment as to whether the program should continue or should be terminated which is, in fact, a statement as to the probability of success. The conviction of the experimenter develops as he analyzes the accumulated experimental information, but this does not say that he is without conviction in the absence of complete data. Rather, he approaches each experiment with some prior thoughts as to what the results might be. If not, why did he choose that particular experiment at that particular time? He questions the validity of each experimental result often using as his frame of reference his prior thoughts of what the results should have been. Laboratory notebooks universely attest to his proclivity as experiments are repeated, not for the purpose of determining the experimental error, but because the experimental result failed his anticipation, and he is convinced this represents an error in procedure. This situation can be represented graphically by Figure 3-1. Figure 3-1a is a representation of the anticipated value of the response. There is a most likely value

Prior Distribution and Sample

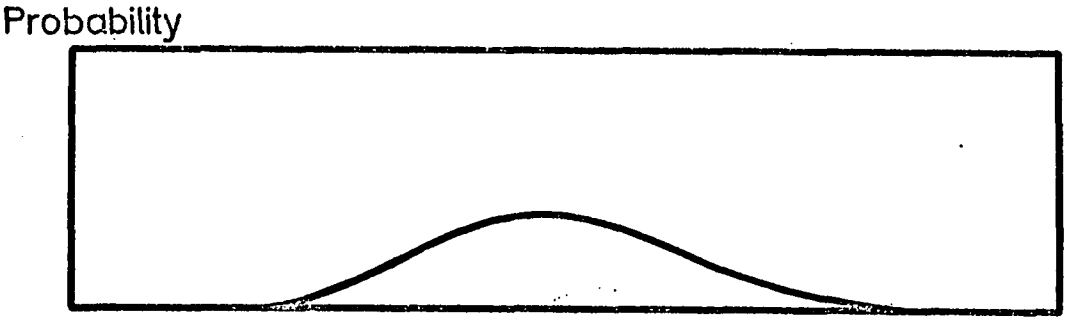

Fig. 3-1a

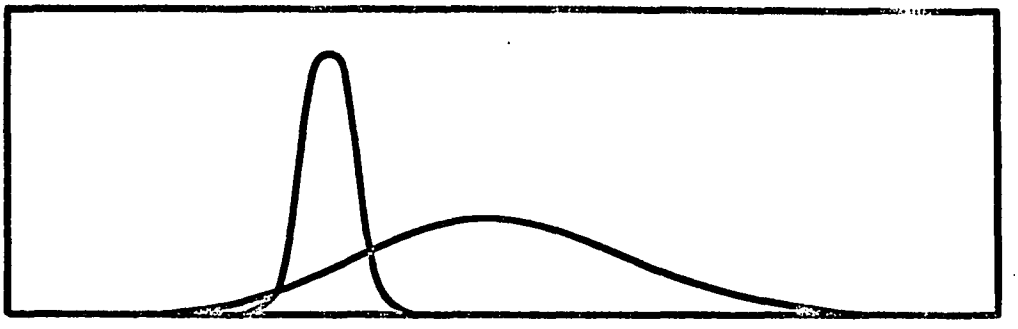

Fig. 3-1b

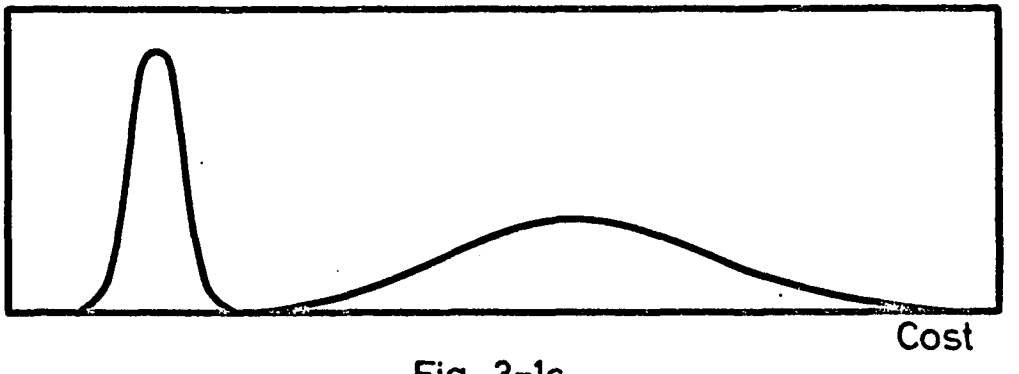

Fig. 3-1c

to which the highest probability has been assigned. Other values are less likely to occur and the probability falls off on either side of the most likely value. Two curves are shown in Figure 3-1b. The broad curve which is repeated from Figure 3-1a and the other much tighter curve which represents the actual experimental result. The value of highest probability is assigned to the actual experimental result, but since both analytical and experimental error are recognized, other values must be admitted and thus the probability distribution results. In Figure 3-1b there is little conflict between the anticipated and experimental results and the experimental value is accepted. In Figure 3-1c, the conflict is shown. The experimenter rationalizes this by either revising his prior estimate of the possible response in light of the experiment or he modifies his estimate of the experimental and analytical error. In the latter case, the experiment is often repeated.

Since the experimenter does have an image of his response, it would be useful if he would enunciate his views in such a form that values can be assigned to his subjective probabilities. Many authors have suggested a series of questions (34, 35) which lead to the elucidation of these distributions. The usual technique involves equating a particular result with betting odds. Although many authors have found this to be satisfactory, the usefulness in this application is rather limited since so many areas of uncertainty exist that such a laborious procedure would become unduly burdensome.

In PERT and CPM analysis (36) estimates of the range and most likely values have provided sufficient data to make meaningful predictions of time distributions. In this case, it has been found that the experimenter could reproducibly state: the most likely result, and the highest and lowest values that he would accept before repeating the experiment. As in PERT no distribution is calculated about the particular estimate; rather the high and low values are assumed to be separated by a fixed number of standard deviations. The value that is assigned to the experimenter's estimate of the upper and lower bound is equivalent to the 95 percent confidence limits and therefore the range is assumed to be four standard deviations apart.

The experimenter will provide information that can be used to establish the variation about the response of a particular experiment; however, this is not the ultimate answer. The experimenter does not think in terms of a single experiment but rather of a series of experiments by which he moves progressively toward his objective. He does not despair if a particular experiment falls on the low side of his expected distribution, but rather this stimulates him to think in terms of conditions for the next experiment. The probability distribution of a single experimental result is of less concern than the probability distribution for the eventual result of the entire series of experiments. To paraphrase, a statement that the yield for a particular experiment should be between 60 and 95 percent is of less interest than a statement placing the eventual result of the total experimental program somewhere between 85 and 95 percent.

The fact that the particular experiment fell outside the range of the eventual expectation is not disturbing, since there are a large number of experiments that still can be run. However, unless some strategy is adopted which takes advantage of each experimental result in the ultimate probability assessment, the desire for "just one more run" will not be eliminated. That is, the failure of a particular run to reach the expectation must somehow be related to the ultimate estimate of the probability of reaching that expectation. The experimenter can manipulate many variables in an effort to improve his results. If he believes, for example, that temperature, concentration and pH are such variables and if his last run was at an adjusted temperature, failure to get improved results says something about the use of temperature as an appropriate variable and also something about his ultimate probability of reaching his expected values. If he in turn finds that neither changes in concentration or pH are effective in raising his yield, he is, of course, forced to modify his expectation of the eventual result. He is free to introduce new variables as they occur to him but the probability of success becomes more remote as the efficacy of each variable is disproved.

What is needed, then, is not a probability statement about the eventual results of a series of experiments, but estimates of the effect on each of the variables to be included in the study on the process response. It is known that the response can be fitted by an expression which is linear in the coefficients of the following form:

$$Y = b_0 + b_1x_1 + b_2x_2 \cdots b_{12}x_1x_2 \cdots b_{11}x_1^2 \cdots + b_{nn}x_n^2 \qquad 3\text{-}1$$

It is required that the experimenter

(1) describe the process variables that are important,

(2) indicate the range of levels for these variables over which he plans to experiment, and

(3) estimate in advance the effects of the variation in these levels.

It is not unusual that the experimenter will provide information about the variables and ranges of operation. This is the usual procedure in setting up an experimental design. Andersen (37) in a review of the subject references several such applications. It is unusual to expect the experimenter to predict the effects in advance of the experimentation but it should be noted that good design strategy requires estimates as to the size of the expected effects, the linearity of the expected effects and indication of those variables which may possibly interact (38).

The experimenter cannot be expected to be able to estimate the most likely value for the coefficients in equation 3-1, yet he is capable of providing estimates of experimental results at various levels of the variables. If the combination of conditions at which the experimenter is asked to make these estimates are well chosen, the coefficients can be calculated from these values. The experimenter is being asked to set up an experimental design and estimate the results of the experiments in advance of their being run. Figure 3-2 shows a central

composite design for three variables. The estimated results are indicated at the design points. Estimates of the range about the most likely value is included in Table 3-1. The three variables: temperature, concentration and mole ratio were varied over the coded ranges. The standard coding of using a plus for the high level of the variable and a minus for the low level of the variable is shown. Hinchon (34) and Harrington (40) use this technique and report correspondence between the estimated and experimentally obtained results.

The estimates of the range about the most likely value is used to determine the variance about each point. The variance is used to appropriately weight the contribution of each of the points in the determination of the coefficients in a regression equation. This is the key to enabling the experimenter to combine estimated results with experimental results.

For each y_i in Table 3-1, an equation in the form of 3-2 can be written.

$$y_i = b_0 + b_1x_1 + b_2x_2 \cdots b_{12}x_1x_2 \cdots b_{11}x_1^2 + b_{nn}x_n^2 \qquad 3\text{-}2$$

This series of equations can be represented in matrix form by

$$Y = XB \qquad 3\text{-}3$$

Where: Y is an (n by 1) vector

X is an (n by m) matrix

B is a (1 by m) vector

)

Central Composite Design for Three Variables

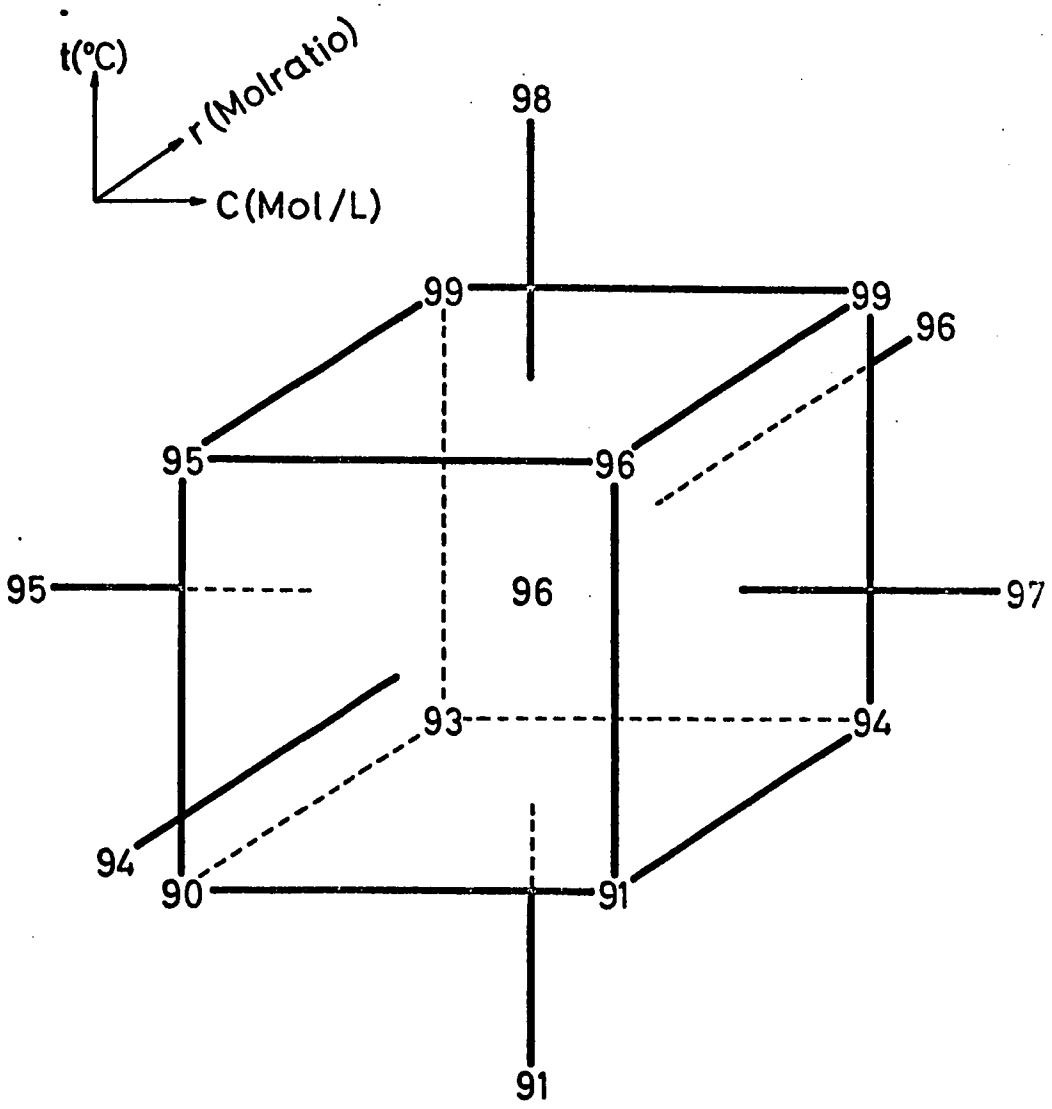

Fig. 3-2

Table 3-1

Matrix Representation of Estimated Yield

No.	Temperature (x_1)	Current Ratio (x_2)	Mole Ratio (x_3)	Estimated Yield Y	Estimated Range of Yield
1	-	-	-	90	80-96
2	+	-	-	95	94-98
3	-	+	-	91	82-96
4	+	+	-	96	95-98
5	-	-	+	93	89-97
6	+	-	+	99	98-99
7	-	+	+	94	91-97
8	+	+	+	99	98-99
9	0	0	0	96	92-96
10	-1.4	0	0	91	83-95
11	1.4	0	0	98	97-98
12	0	-1.4	0	95	90-96
13	0	1.4	0	97	94-97
14	0	0	-1.4	94	85-96
15	0	0	1.4	96	96-98

Example of Coding

Temperature	Code
94	1.4
90	1
80	0
70	-1
66	-1.4

m is the number of coefficients in the model

n is the number of experimental runs

The maximum likelihood estimates (41) of B which is in fact equivalent in this case to the least squares estimates is calculated from equation

$$B = [\ X^T X\ ]^{-1}\ X^T Y \qquad 3\text{-}4$$

In the case when estimated values are being considered, it is unlikely that the variance about each experimental point would be equal. In such case greater reliance should be placed on those values which are known with higher confidence. Weighting the estimates in proportion to the confidence is accomplished by creating the diagonal (n by n) weighting matrix V (42) where the diagonal elements are the variances associated with each of the estimated points.

B can now be calculated from Equation 3-5.

$$B = [\ X^T V^{-1} X\ ]^{-1}\ X^T V^{-1} Y \qquad 3\text{-}5$$

The diagonal elements of the V matrix based on the ranges shown in Table 3-1 are:

16, 1, 12.25, .56, 4, .63, 2.25, .63, 1, 9, 1, 2.25, .56, 14.6, 25.

The greatest variance is associated with the estimate of experiment one. Since the V matrix is inverted, this point carries the least weight in the estimates of the coefficients.

The mechanism has now been developed by which a prediction equation for a particular response can be calculated in advance of experimentation. Either partial differentiation or direct search techniques can now be used to find those x_i's that optimize the response.

The variance estimate about the optimum value can be calculated from Equation 3-6.

$$\text{Var} = X^* \left[X^T V^{-1} X \right]^{-1} X^{*T} \qquad 3\text{-}6$$

where X^* is a (1 by m) column vector each element corresponding sequentially to the linear, interactive and quadratic terms in the prediction model, and locates the optimum condition.

This procedure can, of course, be applied at any stage of the experimental program. As data are accumulated, the ability of the experimenter to assess the response from experiments yet to be run increases, or said another way, his ignorance about the experiment decreases as he looks at more and more data. His ignorance about any of the points that were actually run is reduced to that of the experimental and analytical error.

The design strategy should now be clear. Experiments will be run only when necessary. The estimates of the remaining experiments will be incorporated in the calculation of the regression coefficients. The estimated values and the experimental values are treated identically. If the response of interest is yield, the regression equation can be directly substituted into equation 2-5. The levels of the X_i's that minimize the cost can then be computed. The

variation in Y at these particular x_i's can be calculated from Equation 3-6. A probability distribution about the minimum cost due to a variation in Y can be estimated through a Monte Carlo stimulation employing Equation 2-5. The spread in this distribution is a measure of how well the minimum cost is known. This spread can be reduced by replacing estimates with observations obtained from experimental runs. The best knowledge is derived from an experimentally obtained value but additional runs need only be made when better discriminating power is required.

Figure 3-3 is useful in describing what is meant by discriminating power. In Figure 3-3a two processes are shown in which very little is known about the cost. One might be reluctant to select a route on the basis of this information. The probability cost functions can be described as having poor discriminating power. In Figure 3-3b the same two processes are shown but here, since several experiments have been run, the probability cost distributions are now better known. One process is clearly favored. These probability cost distributions can be said to have strong discriminating power.

The strategy employed in this paper is based on the use of fractional blocks of central composite designs. The following departure from classical use of design theory is made: a response for each of the design point is always included in the determination of the coefficients for a full second-order regression equation. Thus, as the experimentation proceeds through the fractions of the design, the prior information coupled with the observed responses always provides estimates of all the coefficients.

Probability-Cost Distributions for Competing Alternates

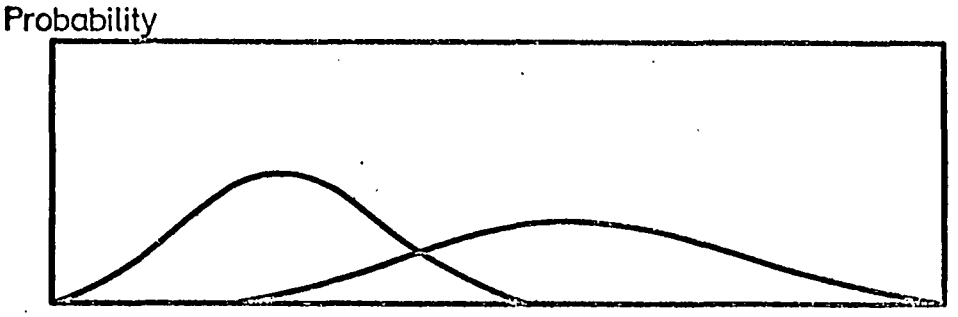

Fig. 3-3a

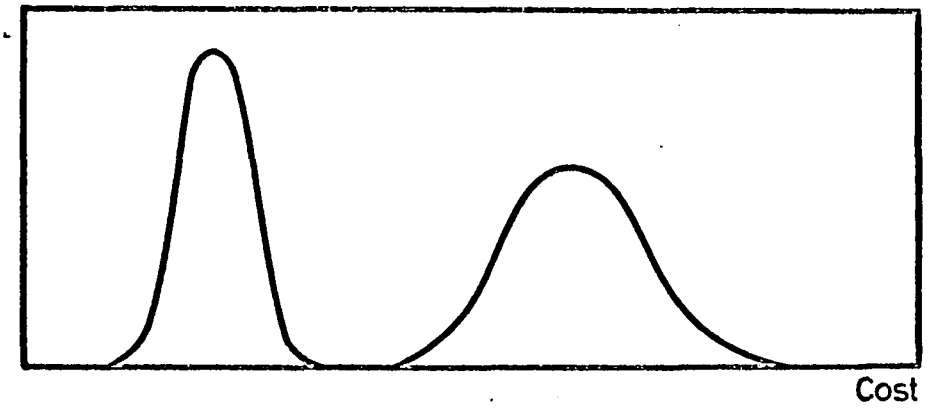

Fig. 3-3b

Such an approach allows the experimenter to employ designed experiments much earlier in the development than heretofore possible. Even though he is trying to fit the surface with a second order polynomial, the commitment on the part of the experimenter to a large number of experiments is minimized. If the sample of experiments that have been run, in conjunction with the modified estimates of the other design points, indicate that a route is unprofitable, the route is abandoned at that point. Further, this approach can allow the experimenter to study wider ranges in the levels of his variables. (The strategy employed here was that of setting the star points of the design at levels close to the natural constraints of the problem.)

The desired result is a continuous decrease in the variance about the optimum point with the addition of each experiment resulting in an increase in the discriminating power of the model. This is not always the case as the variance which is calculated from Equation 3-6 is based on both the location vector X^{*T} and the weighting matrix V. It has been shown that V will continually get smaller as experimental values replace the estimated values and the experimenter's confidence grows. However, should the initial estimate be in grave error, the location vector X^{*T} will move. The most troublesome case occurs when X* moves drastically away from the center of the design. This effect may be sufficient to overwhelm the decrease in V, resulting in a diverging rather than a converging confidence interval about the minimum cost. The discriminating power of the model will therefore decrease.

Under these circumstances it would appear that the experimental design region should be shifted to encompass the new location of the X* vector. The regression equation would then contain correlated estimates of the coefficients. This condition can be redressed by throwing away some of the original postulated results (or even all of them if the shift in space is very large). Even in this situation, the regions of uncertainty will be somewhat better than they would have been in the original location since the observed results from the original design, which may or may not be discarded, continue to give the experimenter greater confidence in guessing responses in the new region. The strategy, therefore, continues to be effective; the original experiments indicated to the experimenter that his original design was poorly placed and that the experimentation could best be performed in some other region of the factor space.

Another difficulty that this problem seems to present is that of discriminating between routes when the discriminating power has been overestimated. This will occur when the experimenter erroneously predicts an optimum near the center of the design, or predicts the experimental results to fall within an unrealistically narrow region. Such predictions would lead to an underestimate of X* and V both of which would contribute to unduly narrow cost distributions. Certainly, whenever the experimenter believes that the information provided by the design is sufficient to enable him to select a route, the route should be checked by running experiments at the anticipated optimum.

Consideration must also be given to the failure of the regression equation to adequately represent the surface given by the predicted and experimental values. The residual sum of squares or lack of fit can be calculated from Equation 3-7.

$$\text{Residual Sum of Squares} = Y^T V^{-1} Y - B^T X^T V^{-1} Y \qquad 3\text{-}7$$

When the predominant portion of the design is filled by estimated responses, these values supply the largest source of variance, which is calculated using the variance covariance matrix $X^T V^{-1} X$. As more experiments are added, the variance calculated from $X^T V^{-1} X$ is reduced while the residual sum of squares may or may not change. Replacing estimated points with experimental values will increase the ability to detect lack of fit. If better models prove to be needed, transformations of the independent variable given by Box and Tidwell (43) or of the dependent variable given by Box and Cox (44) should be considered. The strategy of establishing a new design in the region of the expected optimum should also be useful.

It is, however, anticipated that discrimination between alternate routes can be made without it becoming necessary to reduce the lack of fit. Such was the case in the example which is presented in Chapter 5.

Chapter 4

FINDING THE OPTIMUM VALUE

In the previous chapter, a method for associating an individual response objective with levels of probability was described and an overall experimental strategy was discussed.

The response of ultimate concern is the process cost. This chapter is devoted to the development of a method for utilizing the probability information concerning the many individual areas of experimentation, to compute a measure of the cost potential.

This problem was resolved by taking advantage of existing techniques. The method of direct search was used to determine which experimental conditions would lead to a minimum cost and a Monte Carlo simulation was employed to obtain a probability distribution about that cost.

In the second chapter an equation was presented which could be used to calculate the cost of any particular step. This was Equation 2-5. It was important in Chapter 2 to point out that the values Y, E, R, and C were not fixed and could be varied over wide ranges. In this section, estimating procedures for Y, E, R and C are discussed. Chapter 3 described a method for determining a response. This is of particular value in the determination of the chemical yield, Y, which is usually the variable of greatest concern in the experimental program. It was shown that the objective response, Y, could be related through a quadratic regression equation to the several

process variables. Y could then be replaced in Equation 2-5 with the appropriate regression equation.

The E's were previously described as the ratios of the stoichiometric excesses of one chemical over another in the synthesis. If this ratio was important to the chemical yield, it should appear as one of the process variables, presumably as a concentration term and thus for every E which has a value greater than 1, there should exist an x in the regression equation and the E and the x must be functionally related.

As an example, consider the case where the experimenter wants to study the effect of a stoichiometric ratio of one raw material to the base raw material over the range of 1 to 9. If he chooses to investigate these at the coded levels of -2, -1, 0, 1 and 2, the appropriate coding for x would be $(E-5)/2 = x$, so that E could always be expressed as a function of the midrange and the range. A more general expression is shown in Equation 4-2.

$$E = rx + m \qquad 4\text{-}2$$

Where: m is the midrange

r is the analytical difference in the uncoded levels at the settings -1 and 1.

The maximization of Y does not necessarily correspond to a minimization of cost, since the raw material charges must be considered.

In the previous section it was indicated that the settings of the x_i's that maximize the yields could be found by partial differentiation of the regression equation. The problem of cost minimization cannot be so readily resolved since the regression equation appears in the denominator of the cost equation and some of the x_i's also appear in the numerator. The method for finding the settings of the x_i's that minimize cost will be discussed later in this chapter. It will now be considered that such a method is available and the value of the x_i's that minimize the cost are known.

Several techniques for representing the values R's (the recovery of the unreacted chemicals or the recovery of solvents) are possible. They, too, can be described by regression equations if the experimenter feels that the recovery will vary with the settings of the process variables. Development would be similar to that shown for the yield, and it should be noted that an entirely different set of variables from those that affect the yield may be involved.

In the case of solvents, the recovery is equal to the difference between the input and the loss and can often be related to the physical laws which involve such variables as temperature, pressure and concentration. Explicit expressions for the loss in these variables can be substituted for the value of R.

Finally, the recovery of solvents is often known from experience to vary within narrow limits and a normal probability distribution can be fitted between the outer extremes of the most likely value of the

distribution. This is the approach that will be used in the examples included at the end of this dissertation.

The C's are the response for a one-step process and become the input for subsequent steps in a multistep process. In the latter event, the form of the input of the C's is identical to the output for the preceding step.

The basis for the calculation of the cost distribution is Equation 2-5. It is here assumed that the following information is available. The setting of the x's which provide the optimum values for Y, Y is distributed normally with an optimum value Y_0 and variance about Y_0 calculated from Equation 3-6. The E's are replaced by a function containing x_i's at known levels. The R's are distributed normally with an expected value R_E and a variance calculated from estimates of the extreme values. The costs for all the input raw materials are known exactly.

The calculation of the cost distribution for such a situation is given by Mood (45). This method involves partial differentiation of the cost equation with respect to Y and R, which assumes that Y and R are distributed normally. In a one-step process this is the case. However, a method is required that will handle all cases and once the number of steps in a process is greater than 1, a new variable of cost, C, is introduced. There is no guarantee that this variable will also be normally distributed.

Monte Carlo Simulation

Quigley and Hess (46) describe an approach for calculating the distribution of variables which are dependent on a complicated function of non-normally distributed independent variables. Their approach utilizes the Monte Carlo technique and is applicable here.

The Monte Carlo technique is a sampling procedure whereby complicated expressions involving combinations of probability distributions may be evaluated. The only requirement is that the probability distribution is known and can be represented numerically. A bibliography of the Monte Carlo methods is given by Meyer (47). The distribution about the estimated parameter is developed by taking repeated samples at random from the known populations of the independent variables. The distribution of results comes closer to approximating the true distribution as the sample size is increased.

Since a large number of samples is required, a computer is suited for this application. Random sampling from a distribution is best done by either selecting numbers from a random number table or by internally generating such a series of random digits. The use of random numbers is so extensive that most computers have random number generating routines, which become available by calling for a random number subroutine.

The method used in this dissertation is an IBM library program (48) which uses the power residue method. A three digit number is entered to start the sequence. The three digit number multiplies a ten digit

prime number. The ten low order digits of this product make up the random number. This random number is in turn multiplied by the prime number to generate the next random number. The process is repeated until the required numbers are obtained. The random numbers so chosen can now be assigned to fit the characteristics of any particular distribution. For example, if 10 percent of a distribution has a particular value, the digits 00000000-09999999 can be assigned to that value.

The distribution of interest can always be expressed in terms of a cumulative probability distribution. Such a distribution is shown in Figure 4-1. The ordinate runs from 0 to 1. The random number is used to enter the cumulative probability distribution from the ordinant. In the figure, the selection of the random digit is 55000000 which is located at .55 on the ordinate axis results in a value of $.0965 per pound being selected from the cost distribution.

In cases when the selection is from a normal distribution, the approach is facilitated by recognizing that for a rectangular distribution with limits of 0 to 1, the mean and variance are .5 and 1/12. The sum of 12 such random numbers would have an expected value of 6 and a standard deviation of 1. The statistic shown in Equation 4-1 is approximately normally distributed with mean 0 and variance 1.

$$S = \frac{\sum_{}^{12} RN_i - 12(.5)}{\sqrt{12}\ \ (1/\sqrt{12})} \qquad 4\text{-}1$$

Use of Cumulative Distribution in Monte Carlo Simulation

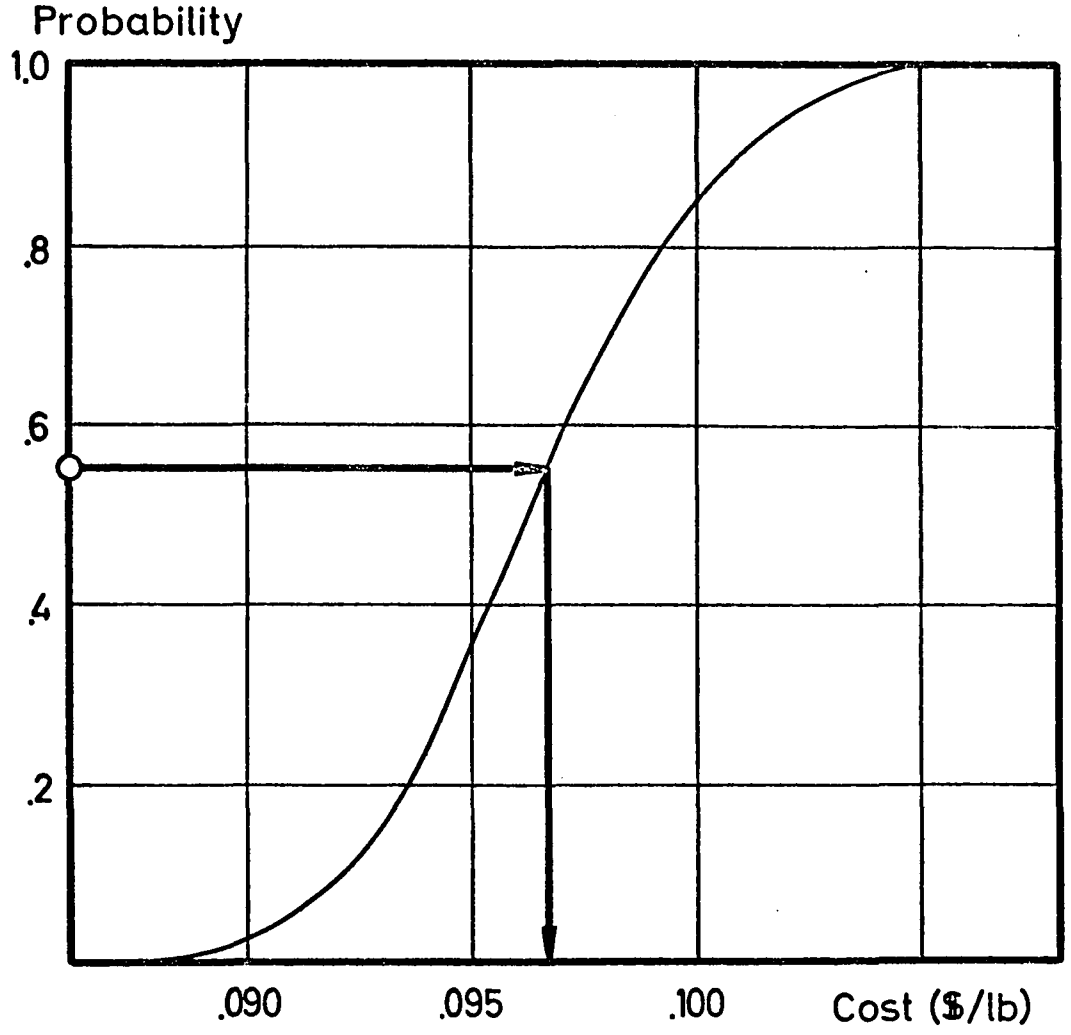

Fig. 4-1

Where: RN_i is a random number from a rectangular distribution.

Normally distributed random numbers are calculated from 4-2:

$$R_i = \sum_{i=1}^{12} RN_i - 6 \qquad 4\text{-}2$$

These normally distributed random numbers can be transformed to the range of interest by Equation 4-3.

$$RT_i = M + sR_i \qquad 4\text{-}3$$

Where: M is the mean of the parent distribution

s is the standard deviation of the parent distribution

The program used here is written for the IBM 1620 and is shown in Appendix C. The rapid convergence toward the actual shape of the distribution is shown in Figure 4-3 where a cumulative cost distribution is drawn after 100, 500 and 5,000 random selections. The curves for the latter two samples are superimposed. In all other cases calculated in this dissertation, a selection of 500 was used. Figure 4-3 is the probability cost distribution calculated for one of the examples in Chapter 5.

The skewed characteristic results from the necessity of truncating the variable Y against the upper bound of 100%. An optimum value for Y of 96.5 was calculated with a confidence interval of $\pm$ 5.7. All values calculated from the random sample that led to values of Y

'Effect of Increasing the Sample Size in Monte Carlo Simulation

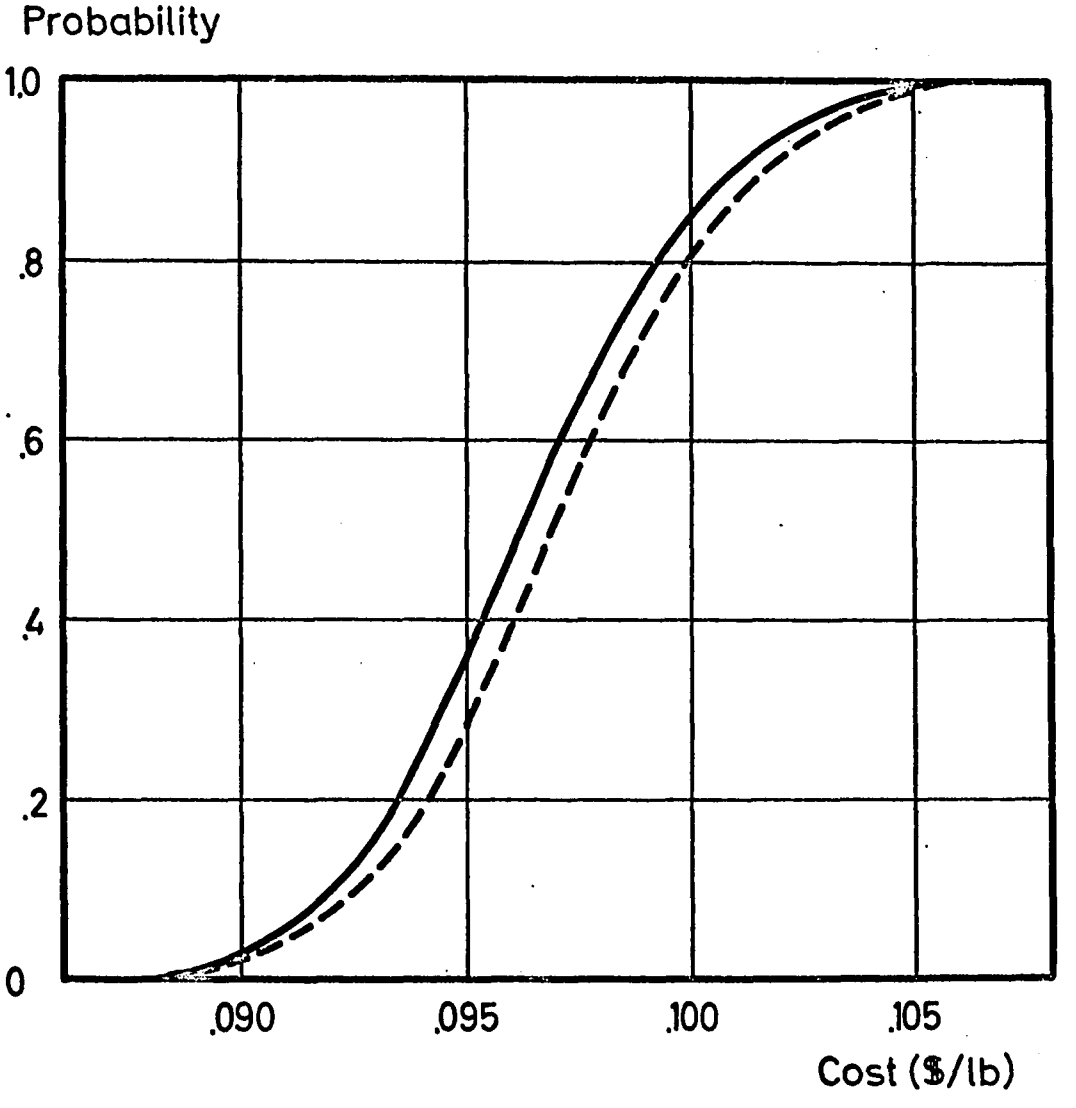

Fig. 4-2

Route II : 100 Trials
" : 500 & 5000 Trials

greater than 100 are set equal to 100. The shape of this cost distribution now becomes the input for the next step. The non-normality of this distribution necessitated the use of the Monte Carlo approach in calculating distributions about all subsequent costs.

Direct Search

Direct search is a sophisticated trial and error technique that has been applied successfully to many problems that have defied solution by classical methods. It often provides faster solutions for many problems that are solvable by classical methods. The technique credited to Hooke and Jeeves (49) has been extensively modified by Wood (50, 51). It has now reached the point where it is the method of choice for finding the optimum solution for complicated functions subjected to several constraints. Wilde (52) offers a critique of the method.

The problem of finding the levels for the variables that minimize the cost is ideally suited for this technique. The strategy is described in the simplified flow diagram shown in Figure 4-3. The operations that are marked in the boxes on the flow diagram are:

Initialization - Levels of the variables X_1 are selected randomly and tested to see if they fall within their allowable range. If so, the value of the function is then computed.

Type I search - A search is performed by moving one variable at a time by a fixed step size in the direction which produces a reduction in the values of the function. The variables are moved sequentially. If, after moving all of the variables, a reduction in the value of the function has not been obtained, the step size used is reduced

Flow Diagram for Direct Search

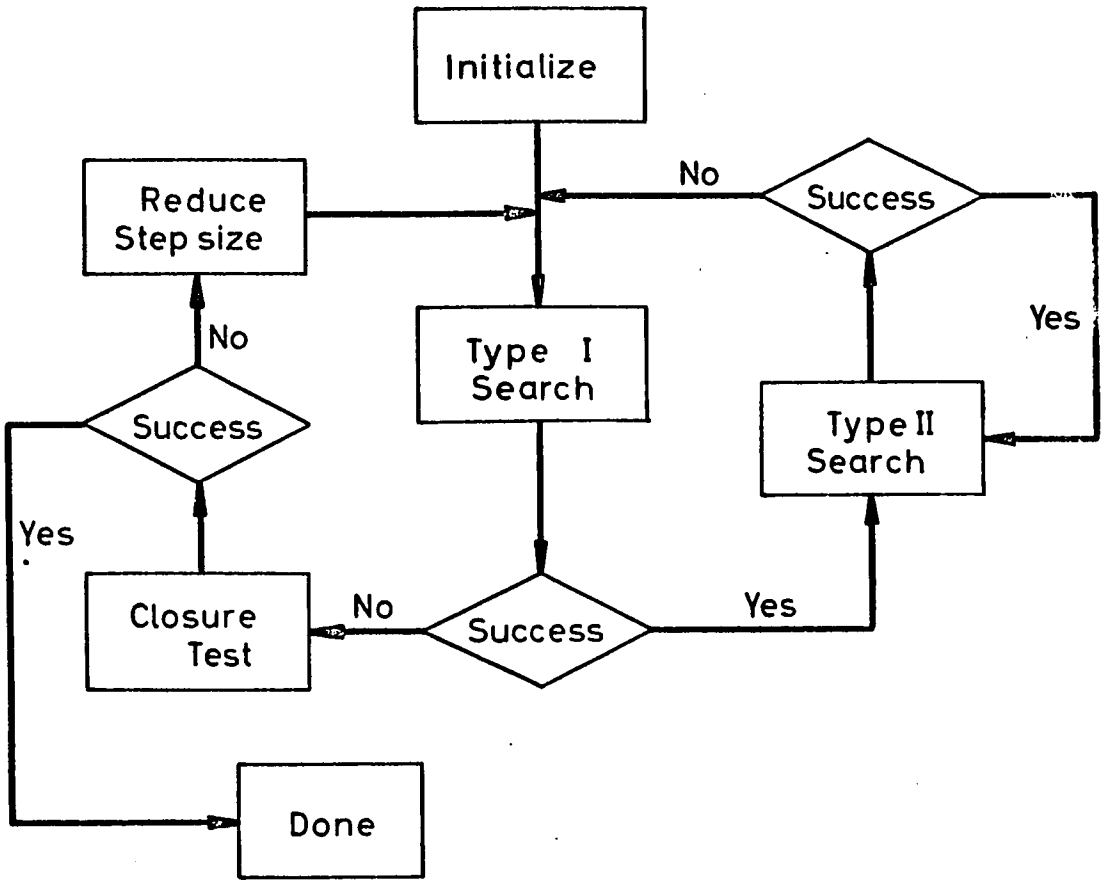

Fig. 4-3

to the next lowest level and the procedure is repeated. One continues this way until the smallest acceptable step size producing no further reduction is found. Successful reduction in the value of the function resulting from a Type I search initiates the start of a pattern move.

Pattern move - All of the variables are moved in the direction and by the same incremental amount that produced reduction in the function when they were individually manipulated during the Type I search. A Type II search immediately follows the Type I search.

Type II search - The Type II search is essentially identical to the Type I search, the only difference being that the value of the function after the pattern move is not computed. Comparison is always made to the best point that was obtained by the Type I search which led to the pattern move.

Closure test - The size of the charge in the variables is compared to a minimum step size. The procedure is successfully terminated when after failing to achieve an improvement by a Type I search it is found that the step size used was at the minimum.

In the problem of interest, the calculated values such as yield and recovery must be constrained to values below 100%. If Y is given as a function

$$Y = F(b_i, x_i) \qquad 4\text{-}4$$

The constraint can be handled by creating a slack variable X_{n+1} and we can then write Equation 4-5

$$\text{CONST} = Y - X_{n+1}$$

$$0 \leq X_{n+1} \leq 1 \qquad 4\text{-}5$$

Now agreement with the constraint is insured by trying to minimize not the cost but a value -- call it SN, which is given by Equation 4-6

$$\text{SN} = \text{COST} + P(\text{CONST})^2 \qquad 4\text{-}6$$

Here P is the number called by Wood a penalty function which multiplies the constraint squared. Thus a mechanism is provided which drives the solution away from an area in which the constraint is violated. The process is continually repeated until the constraint vanishes within some preset limit.

The routines upon which the optimization program operates are shown in Appendix D. The program when run on an IBM 7070 converges within 3 minutes.

Chapter 5

APPLICATION TO AN ACTUAL PROBLEM

An actual situation was selected to test the theories and procedures developed in the earlier chapters. It was agreed to by the American Cyanamid Company that they would use the techniques on a problem of route discrimination. The experimental work and the estimates of the anticipated results from that work was provided by scientists at the American Cyanamid Company. The experimental strategy employed was that described in Chapter 3. The analysis was based on the work presented in Chapters 2-4.

Since the experimentation and program development occurred simultaneously, it was not possible to perform all of the necessary mathematical calculations prior to the initiation of each block of experiments. Rather an experimental design was selected and the experiments were run sequentially with the estimates provided prior to the inception of each new block of runs. A decision to modify the design if it had to be made would have been based primarily on the observed differences between the anticipated and the actual results. In the two reactions which were studied, a change in strategy was not dictated since the results reinforced rather than contradicted the experimenter's prior knowledge.

When the programs were completed it was then possible to treat the data appropriately and to determine just how much experimentation was truly required to discriminate between the alternative routes involved.

It was agreed that in the absence of experience with the technique the test cases should be uncomplicated and the effects of only important variables were to be investigated. The variable time, usually affects the process economics by influencing the size and cost of equipment. Only raw material costs were considered in this example. Time was therefore not included as a variable in the central composite design but rather provisions were made either to sample in time until the maximum yield was obtained or to sample at the appropriate time when this could be determined by visual observation.

Background

The test problem involved the selection of a new synthesis route for a high volume organic chemical. The standard process involved a two-step synthesis with a purification step intermediate between the two reactions. The purification was not specific for the elimination of the impurity and as a result some product was lost. A yield was therefore associated with the purification step. It was discovered that a change in one of the chemicals involved in the purification would not only inhibit product loss but would also enhance the yield of the subsequent step.

Before undertaking this work, an intensive study was made of other alternatives. One which seemed particularly attractive involved a catalytic reaction of the unpurified product resulting from the first step. With this single step both reaction and purification could be completed.

The three alternatives are given below

Present Route

$$A + B + C \longrightarrow ID$$

$$ID + N \longrightarrow PD$$

$$PD + Z \longrightarrow E$$

Where: A, B, C are the reactants in Step 1

ID is the impure product from Step 1

N is the chemical used for purifying ID

PD is the pure product from Step 1

Z is the Step 2 reactant

E is the final product

Alternate Route I

$$A + B + C \longrightarrow ID$$

$$ID + Q \longrightarrow PD$$

$$PD + Z \longrightarrow E$$

Where: Q is the new chemical used in the purification.

Alternate Route II

$$A + B + C \longrightarrow ID$$

$$ID + R \longrightarrow E$$

Where: R is a chemical which selectively reacts with pure D.

Since the first step is common to all three routes, the resulting product from this step can be treated as a purchased chemical of known cost. Alternate I then becomes a two-step synthesis and Alternate II a one-step synthesis. The cost for the present route is, of course, known and provides a stationary bench mark against which the other results must be measured.

Alternate Route I

The graphical analysis discussed in Section I was performed and is shown in Figure 5-1. Three areas of uncertainty were noted. They were the yield of the purification step; the yield of the subsequent reaction and the amount of chemical Z. The effect of increasing the purification yield from 94 to 100 percent is determined from the differences between lines one and two and also the difference between lines three and four. Each one percent increase in yield is worth between .2 and .3 of a cent per pound. An increase in the yield of the second step which is shown along the abscissa is worth about .15 cent per pound. The effect of doubling the usage of Z from the minimum ratio of 1 is established from the differences between lines one and three and two and four. The twofold increase in the usage costs about .7 of a cent per pound. It was realized at this time that the nature of the experiment was such that the amount of Z could never be accurately determined and would vary with the type of equipment that was used. The usage could, however, be estimated. This left two areas for study; the yield of the extraction and the yield of the subsequent step.

Differential Cost Analysis - Route I

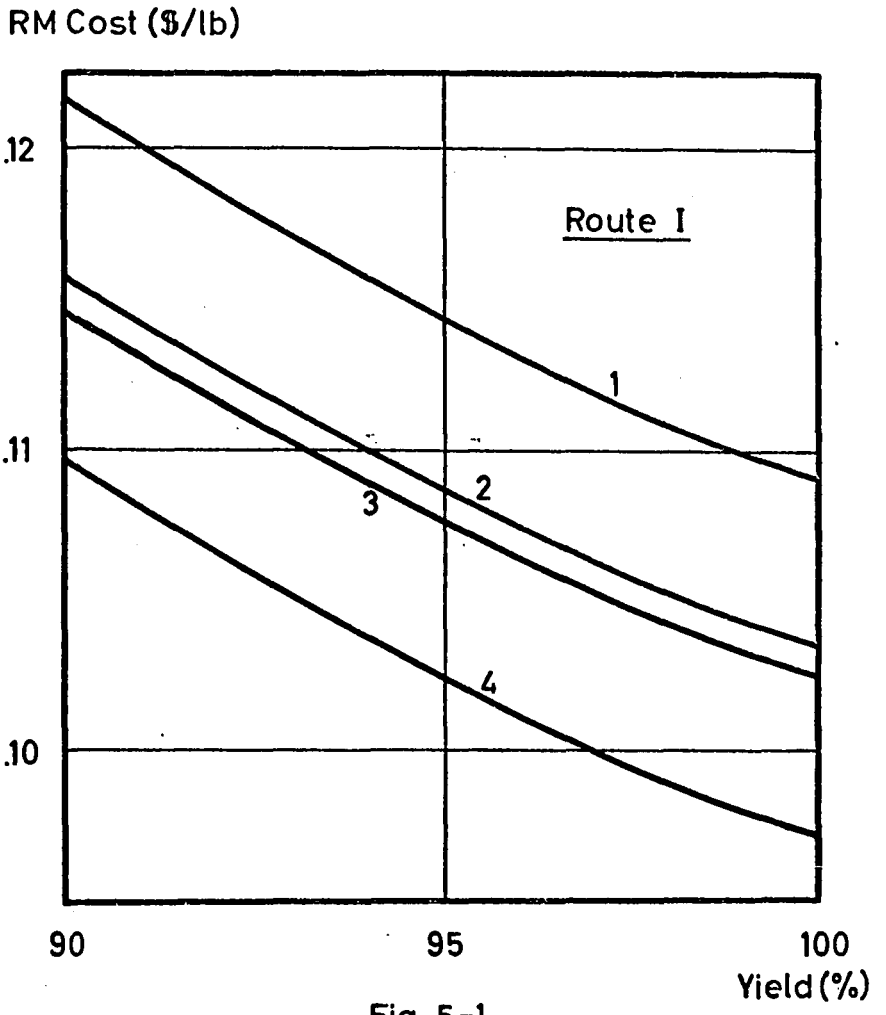

Fig. 5-1

	Step 1 Yield (%)	Z/PD Usage
1	94	2
2	100	2
3	94	1
4	100	1

Since this process was a modification of an existing plant process, experience existed concerning the possible effects of the variables. Three variables were chosen for the investigation: temperature (X_1), concentration or amount of water (X_2) and the ratio of Q to ID (X_3). The method of purification was essentially an extraction of the pure material from a residual tar cake. The longer the extraction period the greater would be the total amount of product extracted; however, at these extended periods, the dissolution of some of the impurity would also take place. Since this could be visually observed by a change in the color of solution from clear to yellow, time was not included as a variable but the experiments were run until the appearance of the slight yellow color. Dr. H. Grethlein who ran the experiments also made the prior estimates. Dr. G. Goulandris, who was familiar with the work, provided another independent set of estimates. A central composite design was selected to study the response surface. The design is shown in Table 5-1 together with the estimates of both Grethlein and Goulandris. Grethlein, who was more familiar with the problem, indicated this better understanding by selecting relatively narrow ranges about the estimates. Although Goulandris could not indicate such conviction and was in general more pessimistic about the results, both experimenters believed that about a 99% yield could be obtained and that this would most likely result when both X_1 and X_3 were at their high levels (lines 6 and 8). In fact, both experimenters believed that X_1 and X_3 would exert strong positive effects and discounted X_2 as being important in the range studied. This thinking was corroborated by the experimental results which were eventually

TABLE 5-1

Original Estimates
Alternate Route I, Step 1

Line	Conditions X_1	X_2	X_3	Grethlein Most Likely	Grethlein Range	Goulandris Most Likely	Goulandris Range	Actual
1	-	-	-	90	80-96	60	40- 70	84.29
2	+	-	-	95	94-98	75	65- 85	97.50
3	-	+	-	91	82-96	60	45- 70	85.53
4	+	+	-	96	95-98	80	65- 85	99.40
5	-	-	+	93	89-97	80	60- 85	97.49
6	+	-	+	99	98-99	99	95-100	98.53
7	-	+	+	94	91-97	80	60- 90	94.24
8	+	+	+	99	89-99	99	95-100	99.15
9	0	0	0	96	92-96	95	85-100	98.32
10	-1.4	0	0	91	83-95	40	20- 70	
11	1.4	0	0	98	97-98	98	80-100	
12	0	-1.4	0	95	90-96	95	70-100	
13	0	1.4	0	97	94-97	95	70-100	
14	0	0	-1.4	94	85-96	95	70-100	
15	0	0	1.4	96	96-98	98	80-100	

obtained. Grethlein did a much better job of estimating the true values but it should be recalled that he had had more experience with the particular route, and further, although Goulandris did a poorer job of estimating, he so indicated this by providing estimates with wide ranges.

Some of the experimental values fell outside of the ranges given by the experimenters but in the region where the yield was maximized, concurrence between the expected and the experimental results was found.

In Table 5-2, three columns of estimates are given; those prior to any experimentation, those after running the center point and a half replicate of the design and again after running the full factorial. The actual experimental results are included in the table and these are always indicated by the dash given under the range.

In general, these estimates which were given by Grethlein showed remarkable stability. The only major modification to the estimates was made for lines 14 and 15 where the strong positive effect of X_3 was initially underrated.

It is interesting to note how the experimenter's confidence in his knowledge increased as more experiments were added. The variances calculated from the ranges are given in lines 16 and 17. Three variances are shown: σ_I^2, which is a pooled average of the center points and all the points of the face of the cube, σ_{II}^2 which is a pooled average of all the points which extend beyond the face of the cube, called star points, and σ_E^2, the estimated experimental error.

TABLE 5-2a

Modification of Estimates with Experimentation

				No Expt.		5 Expts.		9 Expts.	
Line	X_1	X_2	X_3	Most Likely	Range	Most Likely	Range	Most Likely	Range
1	-	-	-	90	80-96	91	83- 92	84.29	-
2	+	-	-	95	94-98	97.50	-	97.50	-
3	-	+	-	91	82-96	85.53	-	85.53	
4	+	+	-	96	95-98	95	93- 98	99.40	-
5	-	-	+	93	89-97	97.49	-	97.49	-
6	+	-	+	99	98-99	99.5	98-100	98.53	-
7	-	+	-	94	91-97	95	93- 97	94.24	-
8	+	+	+	99	98-99	99.15	-	99.15	-
9	0	0	0	96	92-96	98.32	-	98.32	-
10	-1.4	0	0	91	83-95	88	85- 95	93	90- 97
11	1.4	0	0	98	97-98	99.5	98-100	99.5	98-100
12	0	-1.4	0	95	90-96	96	93- 98	98	97- 99
13	0	1.4	0	97	94-97	97	90- 99	98	97- 99
14	0	0	-1.4	94	85-96	91	85- 95	90	88- 94
15	0	0	1.4	96	96-98	99.5	98-100	99	98-100

TABLE 5-2b

Variances

Line	Source	No Expts.	5 Expts.	9 Expts.
16	σ^2_I	4.1	1.7	-
17	σ^2_{II}	3.3	3.3	.7
18	σ^2_E		.25	.25

In other work, it was found that σ_I^2 is generally lower than σ_{II}^2. In this case the experimenter felt less certain about the results from experiments in which both X_1 and X_3 were simultaneously set at the low level (lines 1 and 3) than at the star points where the effect of only a single variable had to be assessed.

After five runs the new value for σ_I^2 was 1.7 and that for σ_{II}^2 was 3.3 which is a change in the expected direction. The four runs at the corners of the cube provided sufficient information about the corner points to allow some contraction in the range of the estimates. After 9 runs, σ_{II}^2 is also reduced in this case to a value of 0.7. Table 5-3 indicates the excellent correspondence between the initial and final estimates. The direction and magnitude of the linear coefficients were estimates extremely well. The slight changes in the optimum settings of the variables and the imperceptible changes in cost reflect the stability of the system.

Figure 5-2 shows the probability cost distributions for the first step of this route. These frequency distributions will serve as input for the next step. The improvement in knowledge resulting from additional experimentation is reflected by the reduction in the range of these cost distributions. The figures are based on 500 Monte Carlo selections. The data and all other similar data are presented in Appendix E; the curves shown are smoothed by eye.

TABLE 5-3

Coefficients

	No. Expt.	6 Expts.	9 Expts.
B_0	95.97	98.25	98.31
B_1	2.51	2.98	4.10
B_2	0.58	-0.01	0.06
B_3	1.39	2.82	2.84
B_{12}	-0.02	0.62	0.57
B_{13}	0.34	-1.40	-2.64
B_{23}	-0.05	0.41	-0.72
B_{11}	-0.02	-1.40	-1.93
B_{22}	0.08	-0.74	-0.15
B_{33}	-0.08	-1.14	-1.72
	Optimum Variable Setting		
X_1	1.38	1.40	1.40
X_2	1.06	0.97	0.67
X_3	-0.01	0.12	0.40
Yield	100%	100%	100%
Cost	0.08845	0.08852	0.08854

Frequency-Cost Distribution - Route I, Step 1

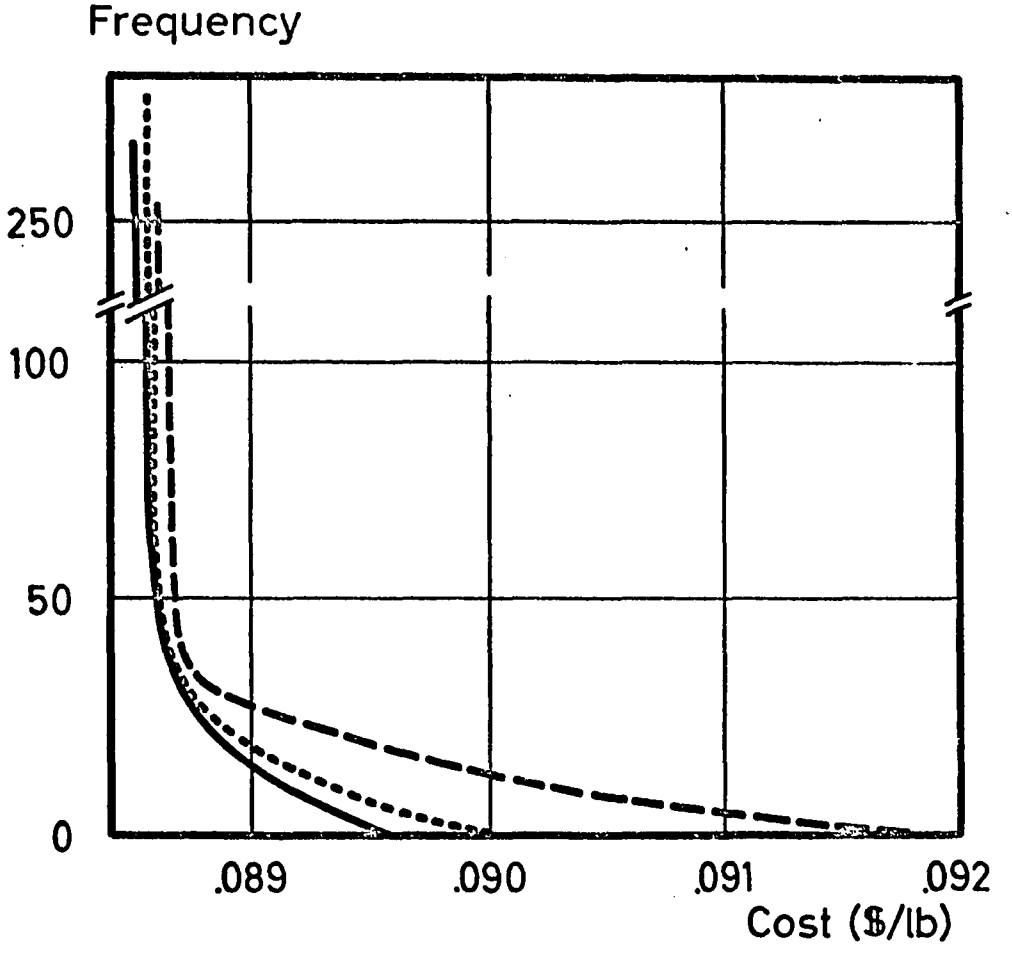

Fig. 5-2

Route I - Step 1 - 0EXPT
" - " - 5EXPT
" - " - 9EXPT

The Second Step - Alternate I

The second process step involved the reaction of the purified intermediate with the chemical reagent Z to produce the final product. The medium in which the reaction takes place now contained Q rather than N. A single experiment was run in the new medium at what were the optimum variable settings for the standard process. A yield of 96% was obtained. Since a good understanding of the factor space was available from the experience with the standard process, fairly tight estimates of the points in a designated experiment were obtained. Experience had shown that the only two variables of importance were pH and temperature. Table 5-4 lists the estimated guesses based on a single run at the center of that design.

As neither of the variables contributed to the cost in any way other than through their influence on yield, the regression model calculated from these estimates could in this case be differentiated to establish the minimum step cost.

One other area of uncertainty involved the utilization of Z, a gas. It was anticipated that the usage of this chemical could be reduced by better design. The vagaries of scale-up were such that there was reluctance to state the usage with any more certainty than that it should be no greater than the present stoichiometric excess of 1.5 and certainly not much lower than a ratio of 1.1.

TABLE 5-4

Estimates of Alternate Route I, Step 2

(After a Run at the Anticipated Optimum)

Line	X_1	X_2	Most Likely	Range
1	-	-	90	87-95
2	+	-	93	91-97
3	-	+	80	76-84
4	+	+	85	81-88
5	0	0	96	-
6	-1.4	0	70	66-75
7	1.4	0	75	70-81
8	0	-1.4	94	90-97
9	0	1.4	88	85-93

The three areas of uncertainty that existed in this route are defined below:

(1) The yield of the purification step for which the estimates in Table 5-1 were made.

(2) The yield of the final step for which a single run had been made, which is included with the estimates in Table 5-4.

(3) Variation in the utilization of Z which was expected not to be lower than 1.1 nor greater than 1.5.

The cost distribution calculated from these data and those for the purification step are given in Figure 5-3. Three curves are shown; they represent the case where only a single run was performed on Step 2, the case when five runs of Step 1 were added and the case where nine runs of Step 1 were added. The gain in information is seen to be small. The data for these distributions were also obtained from a Monte Carlo sample of 500.

The distributions are graphic evidence that experimentation added little to the existing knowledge and that one or two experiments were all that were necessary to define the situation. This should not have been entirely unexpected. The route was similar to the present route and the excellent information available concerning the present route could and was translated through the experimenter's estimates of the experimental results.

Frequency-Cost Distribution - Route I

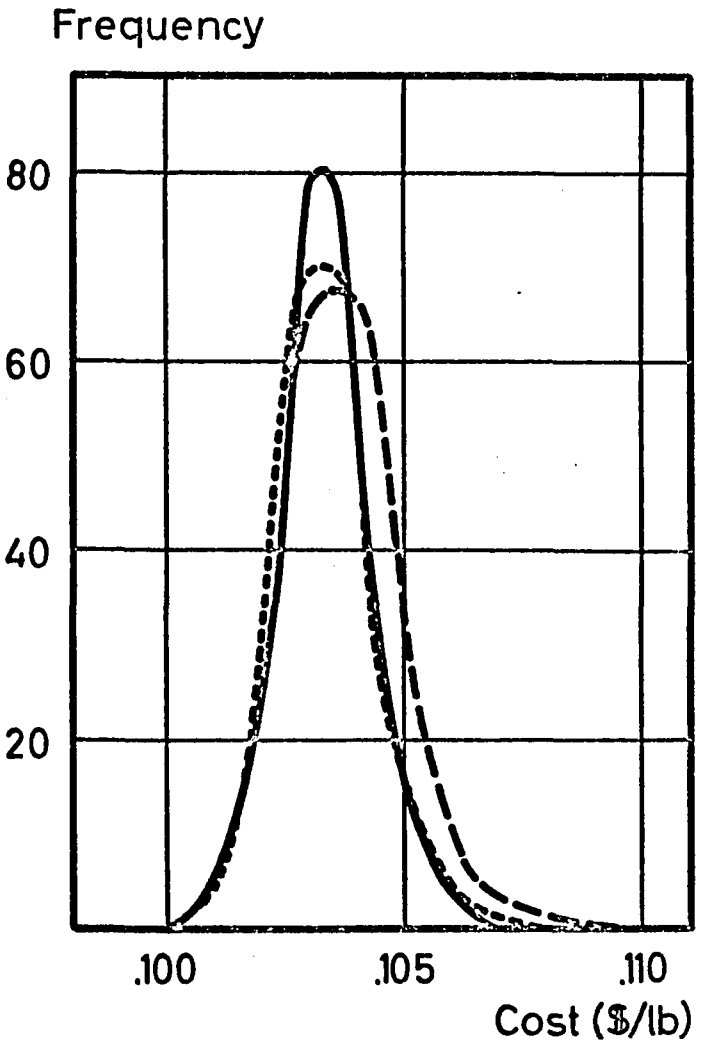

Fig. 5-3

Route I - 1EXPT
" -6EXPT
" -10EXPT

Alternate II

This route represented a completely new approach to the synthesis of E. ID was taken up in a solvent where in the presence of a catalyst a selective reaction took place between D and Z leaving the impurities behind in the mother liquor. Several experiments were run in which different classes of solvents and catalysts were tried. This work which involved solvents and catalysts that were too expensive to be considered in a commercialization of the process provided the basic information which led to the selection of a low cost solvent catalyst system and provided a basis for the estimates of the response surface when the lower priced catalyst solvent system was used. The experimental work was carried out by Dr. H. Merlin and the author. Both men participated in the planning and running of the experiment, but so as not to bias the purpose of this paper, Dr. Merlin alone provided the estimates of the results.

The graphical analysis shown in Figure 5-4 indicated three areas of concern; the yield, solvent usage, and catalyst usage.

The large contribution that the solvent adds to the cost is apparent from the figure. The usage is affected by both the initial charge and eventual recovery. The effect on cost of a reduction in the ratio of solvent to ID of from ten to five can be computed from the difference between lines one and two. Complete elimination of the solvent cost is shown on line three where 100 percent recovery was assumed.

Conversely, the contribution of the catalyst to the cost is small as indicated by the difference between lines three and four.

A one percent increase in yield is worth from .05 to .2 cent per pound, depending upon the yield level and solvent requirements.

As the solvent and catalyst usages were in part affected by the level required to provide adequate yields, an experimental program, investigating the effect of the important variables on yield, was in order. It was expected that for a particular solvent catalyst system, four variables would be important: temperature (X_1), ratio of solvent to reactant (X_2), ratio of catalyst to reactant (X_3) and time which was studied by withdrawing samples from the reactor at various intervals. The response of interest was the maximum yield.

A central composite design was chosen as the basis for the experimental program. The limits of temperature were set by natural constraints as was the lower level of the solvent. The upper level for solvent and catalyst were, of course, limited by economics. An upper bound was selected which set reasonable limits on the concentrations.

Merlin's estimates are shown in Table 5-5. Despite the initially wide ranges suggested, as shown by a comparison of σ_I^2 and σ_{II}^2 for the two routes, his only poor estimates were for Experiments 2, 3 and 9. As the experimentation progressed, the expected reduction in σ_I^2 and σ_{II}^2 occurred. After six runs, good estimates for Runs 2 and 3 were obtained. Since Experiment 9 was the center point and was run first, this estimate could never be corrected.

Differential-Cost Analysis - Route II

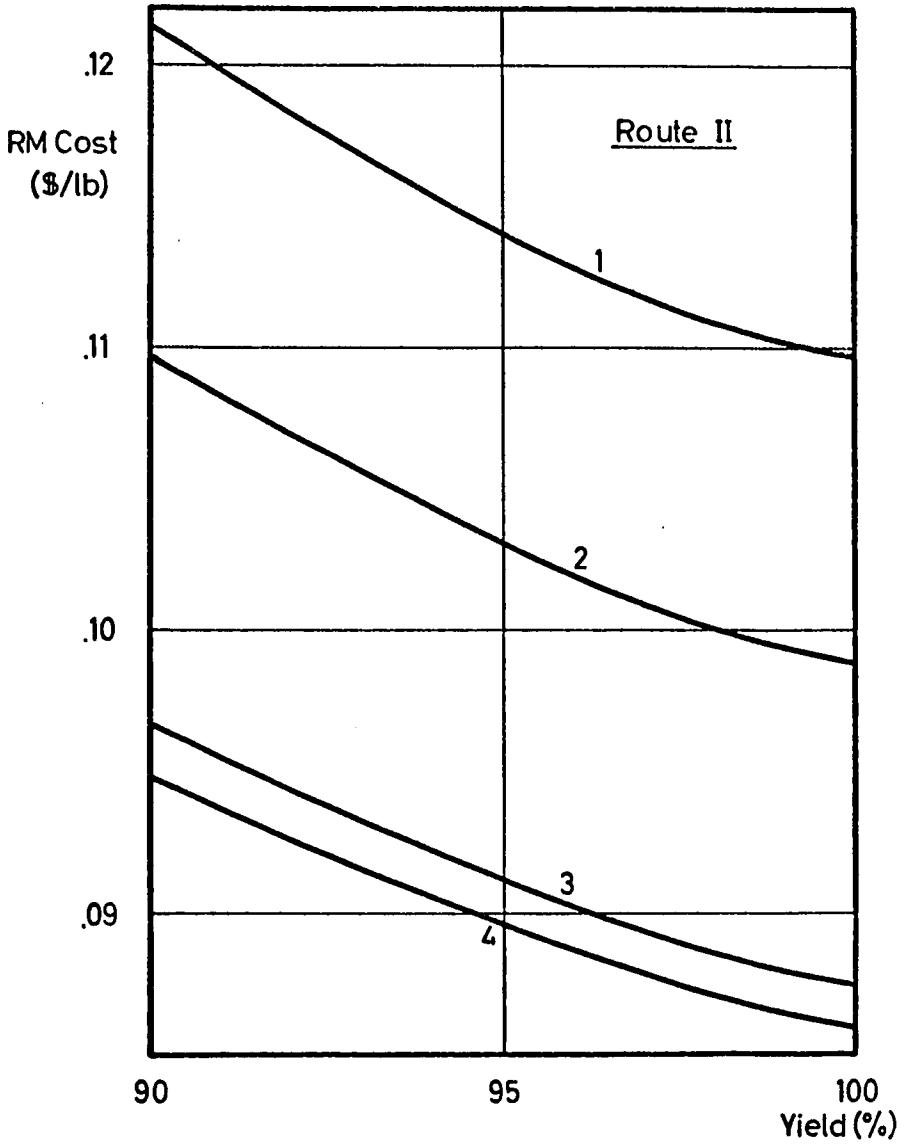

Fig. 5-4

	Solv./ID Ratio	Cat./ID Ratio	Solv. Recov.(%)	Cat. Recov.(%)
1	10	.1	95	50
2	5	.1	95	50
3	5	.1	100	50
4	5	.1	100	100

The coefficients which are given in Table 5-6 show that the original estimates of large positive temperature and solvent effects prevailed as runs were added, but the estimate of the catalyst effect was revised downward. These coefficients will allow estimates of the optimum yield and indicate the appropriate settings for the initial catalyst and solvent charge. This, however, will not completely define the usages of these two chemicals since both can be recovered. It is expected that the solvent can be recovered through distillation. Engineers, experienced in building recovery systems for this particular solvent, suggested that a recovery between 94 and 98 percent was possible. The catalyst could, in a sense, be recovered also as the solution could be recycled with only limited replacement. The graphical analysis had indicated that with the low levels of usage required, accurate determination of the recovery of the catalyst was unnecessary and rather than run additional experiments, Merlin merely estimated that recoveries between 40 and 80 percent were possible. Figure 5-4 is the probability cost distribution for this route. The range is considerably wider than that for Route I and it is only through experimentation that clear understanding of the process costs can be had. The four curves shown correspond to the cases of no experimentation, two experiments, six experiments and nine experiments. The additional increase in knowledge between six and nine experiments leads one to wonder whether the additional experiments were really necessary. Once again, satisfactory understanding of the process route was available with limited experimentation.

TABLE 5-5a

Estimates for Route II After Each Block of Experiments

Line	Conditions X_1	X_2	X_3	0 Expt. Most Likely	0 Expt. Range	2 Expts. Most Likely	2 Expts. Range	6 Expts. Most Likely	6 Expts. Range	9 Expts. Most Likely	9 Expts. Range
1	-	-	-	60	25- 75	60	25- 75	64	-	64	-
2	+	-	-	75	60- 85	77	65- 87	85	83- 87	87.5	-
3	-	+	-	85	79- 98	85	79- 97	98	96- 99	98	-
4	+	+	-	94	90- 99	97	94- 99	96	-	96	-
5	-	-	+	65	55- 78	65	55- 78	59	58- 62	64	-
6	+	-	+	90	85- 95	87.5	-	87.5	-	87.5	-
7	-	+	+	92	88- 98	94	90- 98	98	-	98	-
8	+	+	+	98	96- 99	98	96- 99	96	95- 99	96	-
9	0	0	0	80	92- 95	95	-	95	-	95	-
10	-1.4	0	0	75	68- 88	75	68- 88	75	70- 80	75	70-80
11	1.4	0	0	93	90- 98	93	90- 98	95	92- 98	95	92-98
12	0	-1.4	0	60	20- 68	60	20- 68	30	20- 45	35	20-45
13	0	1.4	0	100	95-100	100	95-100	98	96-100	98	95-99
14	0	0	-1.4	76	70- 92	76	70- 92	95	92- 98	98	93-99
15	0	0	1.4	96	92- 99	96	92- 99	95	92- 98	98	93-99

TABLE 5-5b

Variance

Line	Source	No Expts.	2 Expts.	6 Expts.	9 Expts.
16	σ^2_I	32.7	31.5	.8	-
17	σ^2_{II}	38.7	34.5	8.8	8.8
18	σ^2_E		.25	.25	.25

TABLE 5-6

Estimates of Coefficients

Coefficient	0 Expt.	2 Expts.	6 Expts.	9 Expts.
B_0	96.47	94.90	94.69	94.01
B_1	7.61	7.43	5.63	5.35
B_2	11.17	11.22	11.40	11.09
B_3	5.60	4.98	0.37	0.19
B_{12}	-4.93	-3.44	-6.77	-6.38
B_{13}	-0.54	-1.72	0.71	0.0
B_{23}	-2.60	-2.54	0.23	0.0
B_{11}	-2.18	-5.24	-3.3	-0.65
B_{22}	-1.13	-5.02	-6.4	-7.45
B_{37}	-1.35	-2.63	0.8	0.22
	Optimum Variable Settings			
X_1	0.627	0.605	-0.512	-1.400
X_2	1.400	1.388	1.398	1.333
X_3	-1.14	-1.145	-1.398	-1.400
Yield	0.965	0.965	1.00	0.989
Cost	0.0974	0.0974	0.0929	0.0939

Figure 5-5 is a comparison of the probability cost distribution for both routes. Here Figures 5-3 and 5-4 are combined at the appropriate level of cost. The coding for the lines used in Figures 5-3 and 5-4 applies.

Alternate Route II is shown here as being the preferred route, the mean cost is about $0.093 per pound where that of Alternate Route I is $0.102 per pound. Yet there is a definite probability that Alternate Route II may in fact result in a more costly process. This is a result of the uncertainty associated with the cost estimate and is shown by the overlap of the probability distribution of Alternate II, with that of Alternate Route I. The reduction in the overlap that results from running two experiments can be noted. After five experiments it is virtually certain that Alternate Route II will result in a process with the lower raw material cost. The additional experiments did nothing to improve the discrimination between the routes. These runs merely serve to reduce the variation about the suspected optimum. One should consider whether this experimentation is well invested. Further, it is also clear that any experimentation on Alternate Route I other than the single experiment which was the only experiment used to develop the distribution for the route was unnecessary.

If one randomly selects from the distribution of each route and subtracts the value obtained for Route I from that of Route II, a distribution of the expected gain due to selecting Route II results. This distribution is shown in Figure 5-7. The area to the left of zero represents the potential loss that could result from this decision.

Frequency-Cost Distribution - Route II

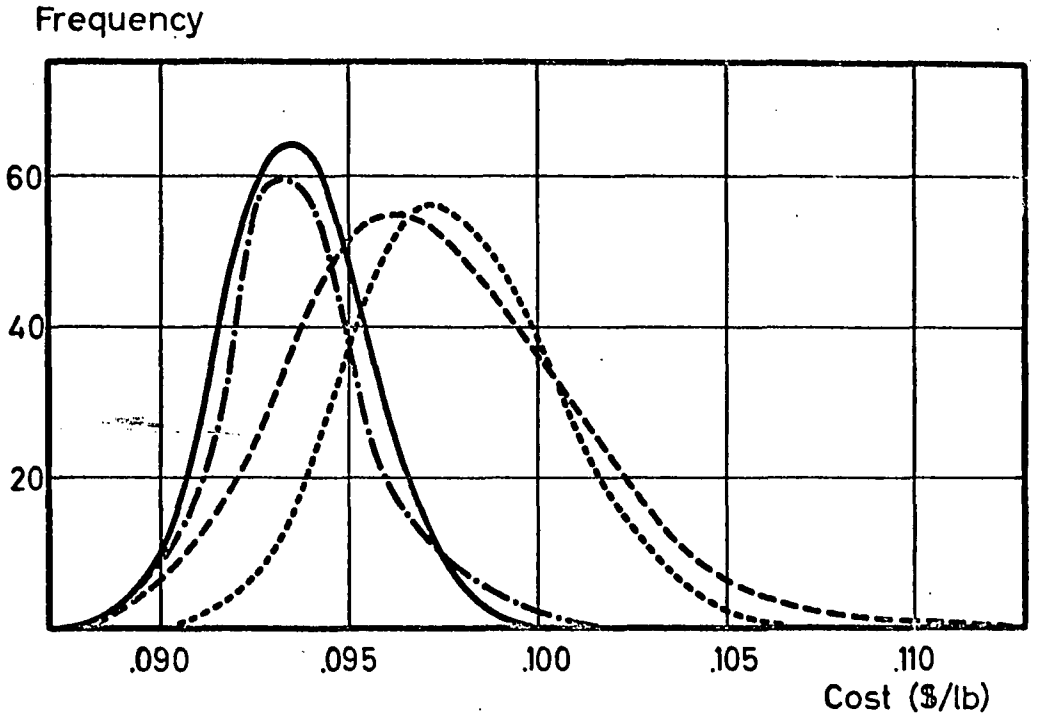

Fig. 5-5

Route II - 0EXPT
" - 2EXPT
" - 6EXPT
" - 9EXPT

A Comparison of the Frequency-Cost Distributions
For Routes I and II

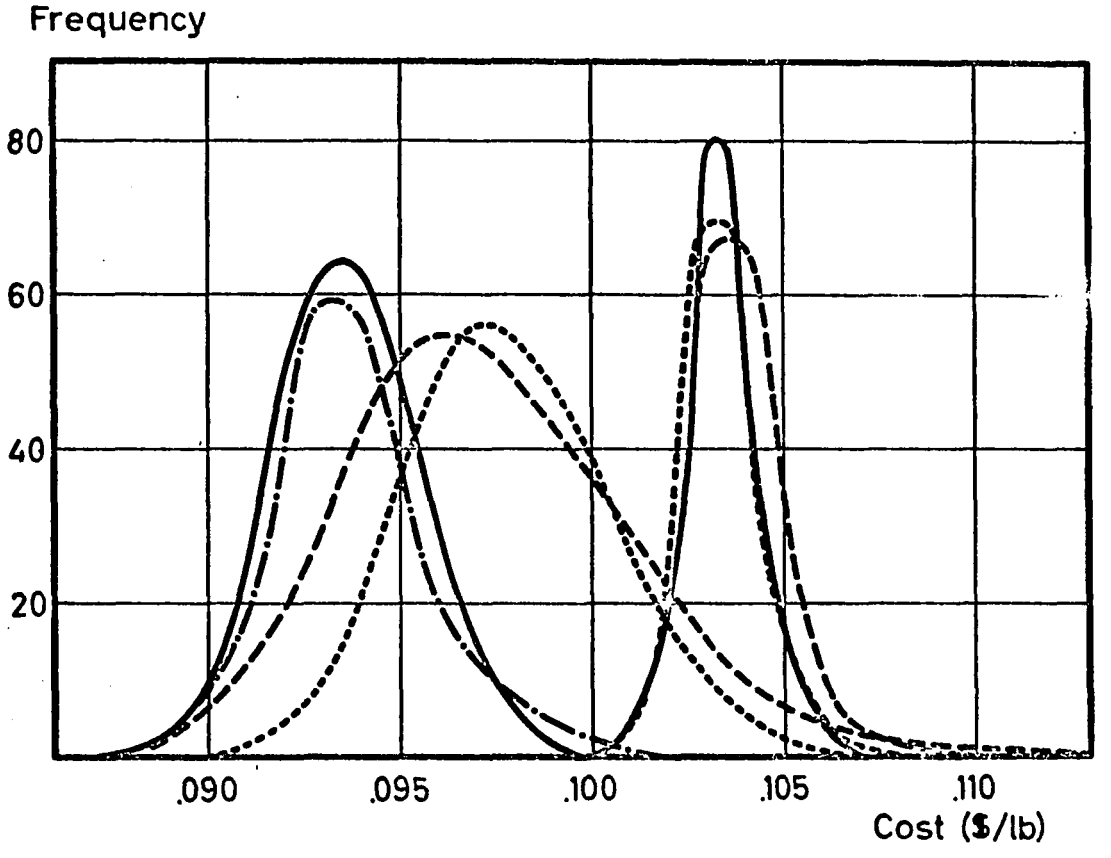

Fig. 5-6

As the definition between the two processes is increased, the area representing potential loss is reduced. Figure 5-7 is calculated from the curves shown in Figure 5-6 in which no experiments were run on Route II and only one was run for Route I. The addition of experiments increases the discriminating power of the model. This can be seen in Figure 5-8 which is calculated for the case when two experiments were run to investigate Route II. The area for a potential loss is considerably reduced. Figure 5-9 which was calculated after six experiments were performed on Route I, already indicates a certain gain.

These curves serve as the basis for decisions concerning the continuation of work on a particular route. Some decision rules could have been established but since each case is subject to different restrictions (size of investment and the effect on profit of delay in entering the market) only the basic information is provided and this can be used as the experimenter sees fit.

In the example shown, a single experiment on Alternate Route I and the exploratory work on Alternate Route II was all that was required to develop Figure 5-6. The experimenter might well have chosen to eliminate Route I from consideration at this point.

Expected Gain From Choice of Route II
(After One Experiment on Route I and One on Route II)

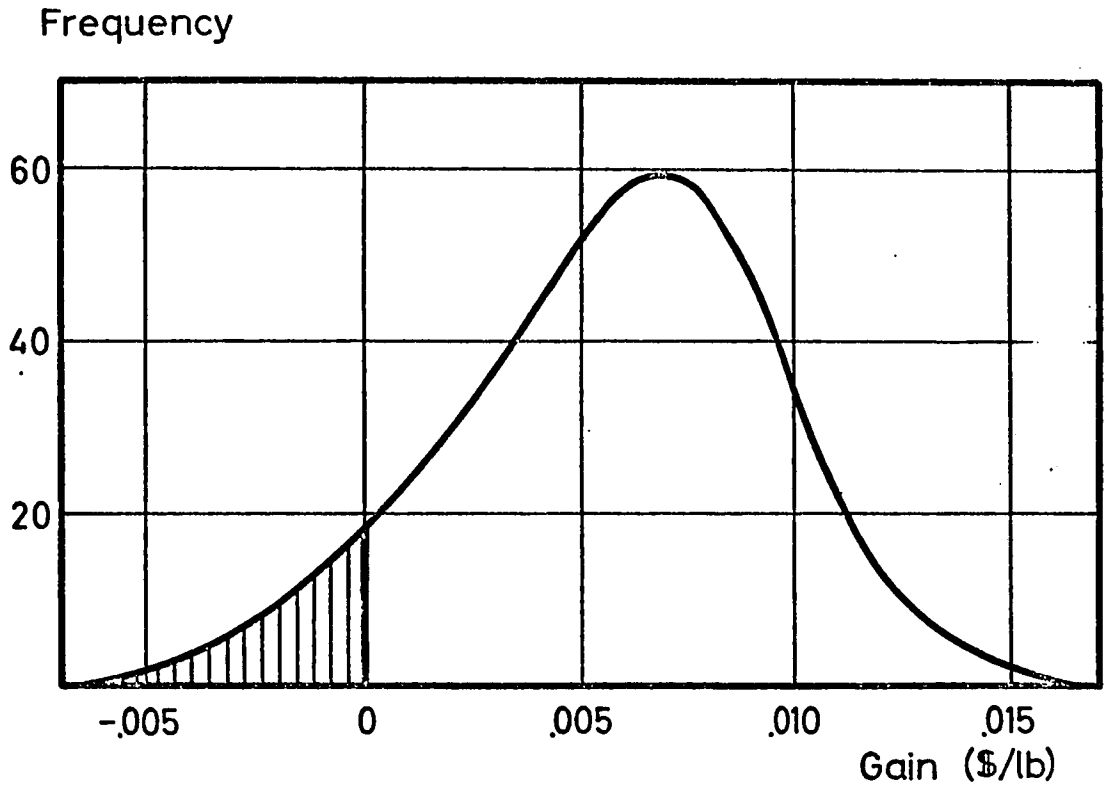

Fig. 5-7

Expected Gain from Choice of Route II
(After One Experiment on Route I and Two on Route II)

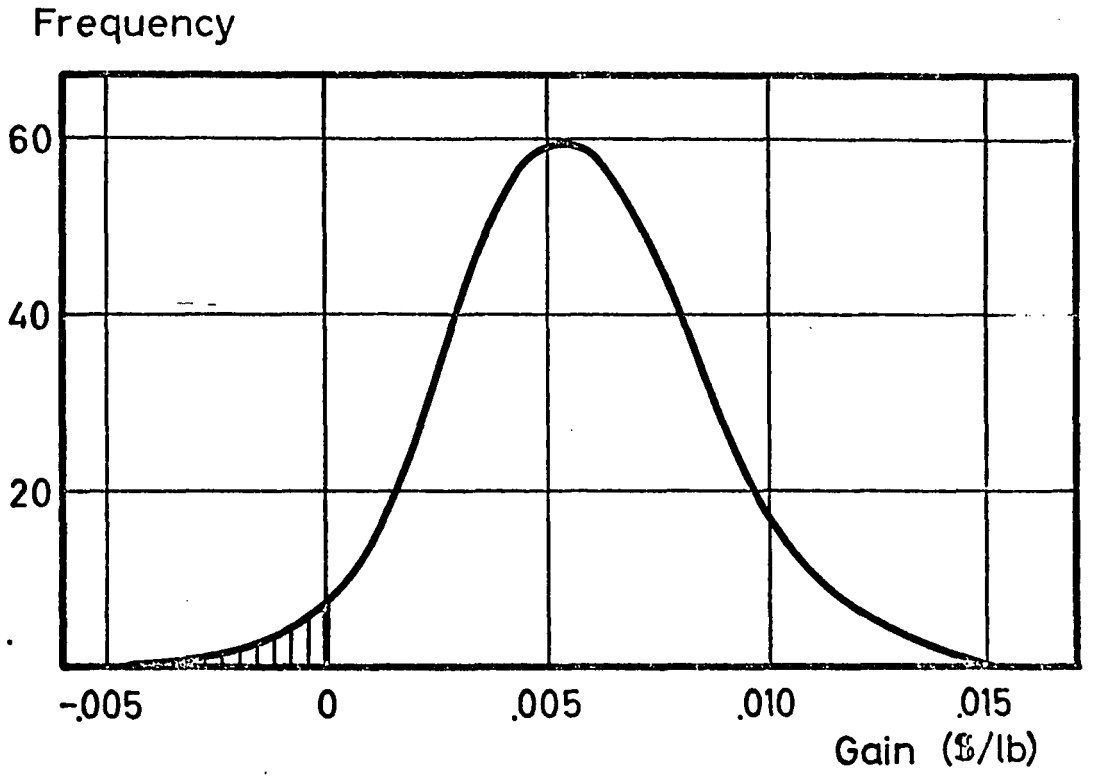

Fig. 5-8

Expected Gain From Choice of Route II
(After One Experiment on Route I and Six on Route II)

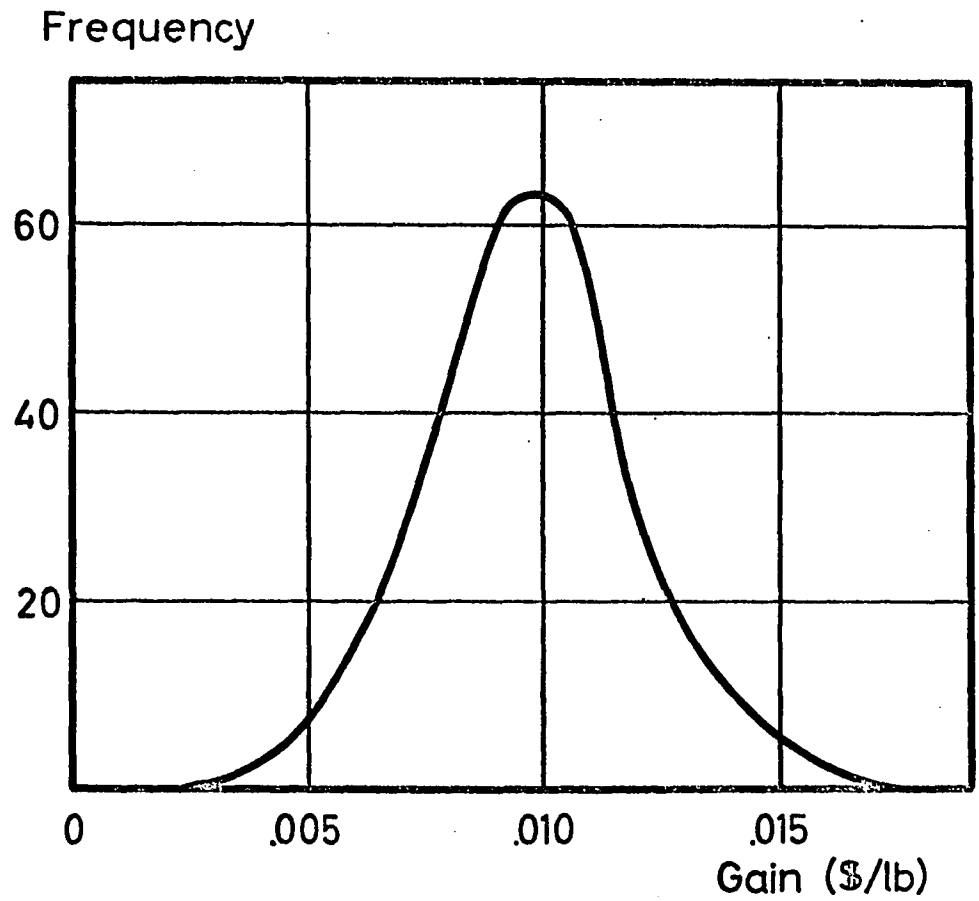

Fig. 5-9

Chapter 6

CONCLUSION AND SUMMARY

A solution to the problem of early discrimination between competing alternate routes is offered in this dissertation.

In Chapter 2 a graphical method was developed which was used to assess the effects of changes in the level of the process responses (yields, recoveries) on the ultimate process cost. This technique establishes the potential value of each experiment and focuses attention on the critical areas within a given route. It was demonstrated through examples that:

(1) The sequence by which the various areas open for investigation are pursued can have a direct influence on the length of the experimental program.

(2) Areas of investigation offering wide variations in the process response may in fact have little bearing on the ultimate problem of route selection.

The technique, therefore, serves to provide information essential to the establishment of an efficient research program.

In Chapters 3 and 4 the relationship between the process cost and the several process responses is extended to a relationship between the process costs and the process variables. These relationships are used to determine the probability of obtaining the indicated cost reductions. The inclusion in the analysis of the experimenters'

estimates of the results of various combinations of the process variables, play an intrinsic role in reducing the number of actual experiments required to define the cost surface.

In the treatment of an actual problem of route discrimination it was shown that:

(1) By incorporating estimated values with experimental results, useful second order regression equations can be developed from runs at only a small fraction of the design points.

(2) The variance about the calculated optimum is reduced with the addition of each new experimental run, but the rate of reduction after completing the first half-replicate is small.

(3) Discrimination between routes can be based on experimental results from a small fraction of the design points.

A summary of the method presented in Chapter 3 is outlined by the following steps:

(1) A second order design is generated with the most important variables.

(2) The responses to the indicated experiments are estimated, listing both a best guess and a range around this best guess.

(3) The usual regression equations are fitted but in this case a weighted variance matrix V is included so that

the precision of the final estimates reflects the uncertainty in the guessed response.

(4) Using this regression equation as part of a function describing the process cost, a vector X* is established representing the design point of minimum cost.

(5) The variance around this design point is calculated.

(6) A few experiments in the design are run. The progression through experiments might be from the original none to a half-replicate, to a full replicate, to a second order design. Re-estimation of the remaining experiments, as well as a thorough analysis, are completed after each stage.

The procedure provides a structural framework within which the experimenter can operate with a minimum of constraints. The structure insures that the data can be sensibly and readily analyzed, while the emphasis on reanalysis after small groups of experiments (even one) satisfies the experimenter's need for flexibility. The approach was designed so that the experimenter's knowledge would be made an integral part of the analysis.

The approach must, of course, withstand the natural reluctance to reach decisions on any information other than that which was won from controlled experiments. There is nowhere in this dissertation an advocation to behave otherwise. Certainly no route should be abandoned until the predicted results are tested at the indicated optimum conditions. What has been suggested is that through the

incorporation of prior knowledge a satisfactory representation of the response surface can be obtained with a reduced number of experiments. The techniques of experimental design have progressed to the point where further reductions in experiments can only come from the inclusion of prior knowledge.

In an interesting exercise, McArthur and Heigl (53) describe a "black box" process which is affected by five hypothetical variables. They ask that the optimum settings for the variables be found with as few experiments as possible. Experience with the black box has shown that no experimental strategy is particularly effective and the required number of experiments is inordinately large. The problem, of course, is that the perverse surface described by McArthur exists only in a black box. There are no fundamental relationships that can be used to explain the results and as such the experimenter's only recourse is to rely on the dictates of the mathematics.

In real situations fundamental relationships exist. The problems are pursued by knowledgeable scientists whose professional life is devoted to trying to understand and apply these relationships. It is expected that such men can successfully use their knowledge to anticipate conditions that will provide favorable experimental results. It is also realistic to expect that their prior assessment as to the consequences of these experiments should be most useful in arriving at a quantitative description of the response surface.

An experimental strategy has been proposed in which central composite designs are employed at an early stage in the experimental program. Available mathematical techniques were utilized to relate the settings of the process variables to the ultimate process costs and to the confidence interval about the calculated optimum cost. Probability cost distributions -- the variance of which is reduced by replacing estimated values with experimental data -- serve as the basis for discrimination.

TABLE OF NOMENCLATURE

English Alphabet:

C_i	Cost of chemical reactant
C_j	Cost of solvent, catalyst, buffer, surfactant
C_k	Credit for byproduct
E_i	Ratio of the charge of the reactant to its equivalent weight
E_j	Ratio of charge of solvent, catalyst, etc. to equivalent weight of reactant 1
E_k	Factor explaining the stoichiometric excess or deficiency of the byproduct
M_i	Equivalent weight of reactant
M_p	Equivalent weight of product
R_i	Recovery of reactant
R_j	Recovery of solvent, catalyst, etc.
U_1	Usage of reactant 1
X_i	Fraction of chemical convertor
Y	Chemical yield

Chemical codes in Chapter 5:

A,B,C	Reactants in present route
E	Final product
ID	Impure product of Step 1 in present route and Alternate Route I
N	Chemical used to purify ID in present route
PD	Pure ID
Q	Chemical used to purify ID in Alternate Route I

R — Reactant in Alternate Route II

Z — Reactant in Step 2 Present Route and Alternate Route I

Matrix Notation:

B — A vector of coefficients in the prediction equation

V — A weighting matrix

X — A matrix describing the settings for the effects for which the coefficients are to be determined

X* — A column vector locating the variable settings at minimum cost

Y — A vector indicating the responses at the different experimental conditions

Greek Letters:

σ_E^2 — Variance due to experimental error

σ_I^2 — Pooled variance calculated from estimates of the range about the factorial and center points

σ_{II}^2 — Pooled variance calculated from estimates of the range about the star points

Appendix A

COST ANALYSIS

Program 422 is written in Fortran II for an IBM 1620 computer. Its purpose is to calculate the cost of a chemical at various levels of yield and usages of all participating chemical reagents. The program involves separate calculations indicating the contribution to the cost of:

(1) All chemicals participating in the stoichiometric equation.

(2) All solvents, catalysts, surfactants and buffers.

(3) All byproducts.

The total cost is the sum of the three equations.

The input data include:

NY The number of yield levels at which the calculations will be performed.

NR The number of chemicals participating in the stoichiometric relations.

NS The number of solvents, catalysts, etc.

NB The number of byproducts.

Y(I) The values of the yield at which the calculations will be performed.

Each chemical that is entered requires the following information:

XMW The molecular weight

STR The stoichiometric coefficient

XCOSR The cost of the chemical

XE The ratio of the amount charged to the stoichiometric requirement

RR The recovery

PR The purity of the chemical (needed if cost is given on an as is basis)

The entries are similar for the other two classes of chemicals with the exception that SF replaces XE where:

SF is the ratio of the pounds of solvent or catalyst, etc. to those of the first chemical read in the previous section.

Other input includes:

XMWP The molecular weight of the product.

STP The stoichiometric coefficient for the product.

All this information is utilized in the calculation section which begins with statement 39. The calculations are based on Equations 2-1, 2-3, and 2-4.

The program calculates the cost at fixed values for XCOSR, XE, RR at each level of Y. The program then accepts a change if indicated in the values XCOSR, XE, RR and runs through that particular calculation for the indicated yield levels.

```
C       AMERICAN CYANAMID CO., BOUND BROOK, N.J. - P0422
C       SEARCH I MODIFIED-RAW MATERIALS COST ANALYSIS
C       A.H. BOBIS
C       MARCH 10 1964
        DIMENSION DUMMY(16),Y(5),H1(3,8),H2(3,8),H3(3,8)
        DIMENSION XMWR(8),STR(8),XCOSR(8),XE(8),RR(8),PR(8)
        DIMENSION XMWS(5),STS(5),XCOSS(5),SF(5),SR(5),PS(5)
        DIMENSION XMWB(5),STB(5),XCOSB(5),XB(5),BR(5),PB(5)
C       FORMAT STATEMENTS
     92 FORMAT(6HSEARCH//)
     93 FORMAT(60H------------------------------------------------
       1----------)
     94 FORMAT(7HYIELD = F6.1,8H PERCENT/)
     95 FORMAT(1X)
     96 FORMAT(9H I J ITEM9X7HCOST/LB5X5HRATIO5X5HRECOV5X5HUSAGE
       15X4HCOST)
     97 FORMAT(/7HPRODUCT11X7HCOST/LB)
     98 FORMAT(15H--------------,5F10.4)
     99 FORMAT(5I5)
    301 FORMAT(I2,I2,2A4,A3,6F10.4)
    311 FORMAT(79X,1H-)
C
C       READ ORIGINAL DATA
     40 READ 93
        READ 99,NR,NS,NB,NY
        NT=NR+NS+NB
        READ 98,Y(1),Y(2),Y(3),Y(4),Y(5)
C       INPUT ROUTINE
        DO 217 N=1,NT
        NTR=1
    500 READ 301,I,J,H1(I,J),H2(I,J),H3(I,J),Z1,Z2,Z3,Z4,Z5,Z6
        GO TO (401,402,403),I
    401 XMWR(J)=Z1
        STR(J)=Z2
        XCOSR(J)=Z3
        XE(J)=Z4
        RR(J)=Z5
        PR(J)=Z6
        GO TO (217,700),NTR
    402 XMWS(J)=Z1
        STS(J)=Z2
        XCOSS(J)=Z3
        SF(J)=Z4
        SR(J)=Z5
        PS(J)=Z6
        GO TO (217,700),NTR
    403 XMWB(J)=Z1
        STB(J)=Z2
```

```
      XCOSB(J)=Z3
      XB(J)=Z4
      BR(J)=Z5
      PB(J)=Z6
      GO TO (217,700),NTR
  217 CONTINUE
      READ 98,XMWP,STP
C
C     PUNCH TITLE AND BEGIN CALCULATIONS
   39 DO 1000 KY=1,NY
      PUNCH 311
      PUNCH 92
      PUNCH 93
      PUNCH 95
      Z6=100.*Y(KY)+.05
      PUNCH 94,Z6
      PUNCH 95
      A=XMWR(1)*XE(1)*STR(1)/PR(1)
      TCOSR=0.
      TCOSS=0.
      TCOSB=0.
C
C      REACTANT SECTION
      PUNCH 96
      PUNCH 95
      I=1
      DO 601 N=1,NR
      C=Y(KY)+(XE(N)-Y(KY))*(1.-RR(N))
      XUSER=STR(N)*XMWR(N)/(PR(N)*Y(KY)*XMWP*STP)*C
      COSR=XUSER*XCOSR(N)
      IF(N-1)52,1050,1051
 1050 G=XUSER
 1051 J=N
      Z1=XCOSR(N)
      Z2=XE(N)
      Z3=RR(N)
      Z4=XUSER+.00005
      Z5=COSR+.00005
      GO TO 600
C     STATEMENT 600 IS IN THE OUTPUT ROUTINE
  601 TCOSR=TCOSR+COSR
C
C     SOLVENT AND OTHER CHEMICALS SECTION
      IF(NS)52,151,914
  914 I=2
      N1=NR+1
      NF=NR+NS
      DO 602 N=N1,NF
```

```
      J=N-NR
      XUSES=SF(J)*G*(1.-SR(J))
      COSS=XUSES*XCOSS(J)
      Z1=XCOSS(J)
      Z2=SF(J)
      Z3=SR(J)
      Z4=XUSES+.00005
      Z5=COSS+.00005
      GO TO 600
C     STATEMENT 600 IS THE OUTPUT SECTION
  602 TCOSS=TCOSS+COSS
C
C     BYPRODUCT SECTION
  151 IF(NB)52,800,920
  920 I=3
      N1=NR+NS+1
      NF=NR+NS+NB
      DO 603 N=N1,NF
      J=N-NR-NS
      BPM=(XB(J)*STB(J)*XMWB(J)*BR(J))/(STP*XMWP)
      COBP=BPM*XCOSB(J)
      Z1=XCOSB(J)
      Z2=XB(J)
      Z3=BR(J)
      Z4=-(BPM+.00005)
      Z5=-(COBP+.00005)
      GO TO 600
  603 TCOSB=TCOSB+COBP
  800 TCOSC=TCOSR+TCOSS-TCOSB+.00005
      PUNCH 97
 1000 PUNCH 98,TCOSC
C     CALCULATIONS AND OUTPUT COMPLETED
      GO TO 50
C     STATEMENT 50 IS START OF TRANSFER SECTION
C     OUTPUT ROUTINE
  600 PUNCH 301,I,J,H1(I,J),H2(I,J),H3(I,J),Z1,Z2,Z3,Z4,Z5
      GO TO (601,602,603),I
C     TRANSFER SECTION(TO MODIFY
C
   50 READ 99,NTR,NT
      GO TO (40,51,52),NTR
   51 DO 700 L=1,NT
      GO TO 500
C     STATEMENT 500 IS AN ENTRANCE TO INPUT ROUTINE
  700 CONTINUE
      GO TO 39
C     STATEMENT 39 IS THE FIRST IN THE CALCULATION SECTION
   52 PAUSE
```

```
GO TO 50
END
```

Appendix B

PLOTTING THE DIFFERENTIAL COST ANALYSIS

Program Number 431B is a 1620 computer program written in Fortran II. The card output is then fed to an IBM 870 (autoplotter) which prints graphs of the type shown in Figures 2-1, 2-2, 5-1 and 5-2.

The program is written for a three step process and is designed to develop the following curves:

(1) Cost of any step versus the yield of any step in which the yield of the other steps are treated as parameters.

(2) Cost of the first step versus the overall yield.

(3) The final cost versus the cost of step 1, with the yields of steps 2 and 3 treated as parameters.

(4) The final cost versus the cost of step 2, with the yield of step 3 treated as a parameter.

For any given line the raw material cost, excess and recoveries are fixed. Equations 2-7 to 2-18 provide the necessary mathematics.

The input, besides that described in Chapter 2, includes information about the scaling of the two axis and the line density. Each line is indicated with a black or red dot or cross. The desired symbol is under the users control. The cost is calculated at increments fixed by the desired line density. A card is punched which contains the values of the coordinates. These cards are fed to the 870 and are the basis for the graphs.

Appendix C

MONTE CARLO SIMULATION

Program 598A is written in Fortran II for an IBM 1620 computer. The program calculates the raw material costs for a particular chemical step based on Equation 2-5. The values used in the calculation for the chemical yield, raw material cost, stoichiometric excess and chemical recoveries are randomly selected from probability distributions defined by a mean and standard deviation. The calculation is repeated until a representative probability for the raw material cost of the particular process step is obtained.

The input to the program include:

N The number of chemicals that participate in the stoichiometric equation.

M The number of solvents, catalysts, etc.

LOOP The total number of costs that are to be calculated.

NC The number of classes used to define the probability cost distribution.

YM The mean yield.

YSD The standard deviation for the yield distribution

P The molecular weight of the product.

ARG The input to start the random number series.

CM(I) The mean cost for chemical I.

CSD(I) The standard deviation for the cost distribution.

RM(I) The mean recovery for chemical I.

RSD(I) The standard deviation for the recovery distribution.

E(I) The ratio of the initial charge to the stoichiometric requirement for chemical I.

WM(I) The molecular weight for chemical I.

As the problem of interest did not involve byproducts, the program was not generalized to handle that situation.

Since the random number routine is time consuming, the program checks each standard deviation to determine whether or not it is zero and calculates the random number for only the non-zero entries. The normalized random numbers for yield and recovery were truncated at the upper bound of 100 percent.

500 cost calculations were completed for each case. The highest and lowest values obtained are subtracted and divided by the number of classes to establish the class size. The difference between any value and the minimum value is then divided by the class size for the purposes of assigning that cost to a particular class.

Program 0598B uses as input the accumulated frequency-cost distribution generated by program 0598A. By dividing each entry in the accumulated distribution by the number of values used in its generation, the distribution is scaled from zero to one. Numbers randomly selected from a uniform distribution are then used to enter

the accumulated distribution as shown in Figure 4-2. This entry determines the cost of the raw material to be used in a particular calculation.

The rest of the program is essentially the same as 0598A.

The additional entries are:

CMIN The minimum value of the frequency distribution generated through the use of 0598A.

CLSZE The class size of that distribution.

```
C       AMERICAN CYANAMID CO., BOUND BROOK, N.J. - PROGRAM NO. 0598A
C       MONTE-CARLO COST ANALYSIS PROGRAM STEP 1
C       INPUT SUMMATION PARAMETER, NUMBER OF COST VALUES, AND
C       NUMBER OF CLASSES
C       A.H. BOBIS
C       NOVEMBER 8 1965
        DIMENSION CM(3),CSD(3),RM(3),RSD(3),E(3),WM(3),C(3),R(3),
       1COST(500) KF(20)
   99 READ 100, N,M,LOOP,NC
        READ 101,YM,YSD,P,ARG
        PUNCH 200
        DO 10 I=1,N
   10 READ 101,CM(I),CSD(I),RM(I),RSD(I),E(I),WM(I)
        DO 30 NT=1,LOOP
        DO 20 I=1,N
C       COMPUTE INPUT COSTS
        IF(CSD(I))1,2,1
    2 C(I)=CM(I)
        GO TO 3
    1 C(I)=RAND(-ARG )*CSD(I)+CM(I)
C       COMPUTE RECOVERIES
    3 IF(RSD(I))4,5,4
    5 R(I)=RM(I)
        GO TO 6
    4 R(I)=RAND(-ARG )*RSD(I)+RM(I)
    6 IF(R(I)-1.00)20,20,15
   15 R(I)=1.00
   20 CONTINUE
C       COMPUTE YIELD
        IF(YSD)7,8,7
    8 Y=YM
        GO TO 9
    7 Y=RAND(-ARG )*YSD+YM
    9 IF(Y)16,16,17
   16 Y=.00001
   17 IF(Y-1.00)27,27,25
   25 Y=1.00
C       COMPUTE COSTS OF INTERMEDIATE PRODUCT
   27 TEMP=0.
        J=N-M
        DO 28 I=1,J
   28 TEMP= (E(I)*(1.-R(I))/Y+R(I))*C(I)*WM(I)+TEMP
        IF(M)31,30,31
   31 J=N-M+1
        DO 29 I=J,N
   29 TEMP=C(I)*E(I)*(E(1)*WM(1)/Y-(E(1)*WM(1)/Y-WM(1))*R(1))*
       1(1.-R(I))+ TEMP
   30 COST(NT)=TEMP/P
```

```
C       COMPUTE MIN AND MAX COST OF INTERMEDIATE PRODUCT
        CMIN=COST(1)
        CMAX=CMIN
        DO 70 NT=1,LOOP
        IF(COST(NT)-CMIN)40,50,50
   40   CMIN=COST(NT)
   50   IF(COST(NT)-CMAX)70,70,60
   60   CMAX=COST(NT)
   70   CONTINUE
C       CLEAR KF ARRAY FOR FREQUENCY COUNT
        DO 80 I=1,NC
   80   KF(I)=0.
        FNC=NC
C       COMPUTE SIZE OF CLASSES
        CLSZE=(CMAX-CMIN)/FNC
C       COUNT NUMBER IN EACH CLASS
        DO 90 NT=1,LOOP
        A=(COST(NT)-CMIN)/CLSZE
        IF(COST(NT)-CMAX)87,85,95
   85   I=A
        GO TO 90
   87   I=A+1.
   90   KF(I)=KF(I)+1
C       OUTPUT MIN AND MAX COSTS, CLASS SIZE AND FREQUENCY
C       DISTRIBUTION
        PUNCH 102, CMIN,CMAX,CLSZE
        PUNCH 103
        PUNCH 104,(I,KF(I),I=1,NC)
        PUNCH 105
        GO TO 99
   95   CALL EXIT
  200   FORMAT(33HMONTE-CARLO COST ANALYSIS STEP 1$)
  100   FORMAT(4I5)
  101   FORMAT(7F10.4)
  102   FORMAT(6HCMIN =,F10.6,5X,6HCMAX =,F10.6,5X,7HCLSZE =,
       1F10.7,1H)
  103   FORMAT(1H$/5HCLASS,5X,10HFREQUENCY$/1H$)
  104   FORMAT(I4,7X,I5,1H$)
  105   FORMAT(1H*)
        END
```

```
C       AMERICAN CYANAMID CO., BOUND BROOK, N.J. - PROGRAM NO. 0598B
C       MONTE-CARLO COST ANALYSIS - STEP 2
C
C       THIS PROGRAM USES FUNCTION RN FOR GETTING RANDOM NUMBERS
C       FROM A NORMAL DISTRIBUTION
C       A.H. BOBIS
C       NOVEMBER 22 1965
        DIMENSION CM(3),CSD(3),RM(3),RSD(3),E(3),WM(3),C(3),R(3),
       2COST(500) KF(20),ESD(3)
   99 READ 108, N,M,LOOP,NC
        READ 101,CMIN,CLSZE
        READ 101,YM,YSD,P,ARG
        DO 1 I=1,N
    1 READ 101,RM(I),RSD(I),E(I),WM(I)
        DO 2 I=2,N
    2 READ 101,CM(I),CSD(I)
C       READ IN FREQ DISTRIBUTION FROM STEP 1
        READ 100,(KF(I),I=1,NC)
        DO 10 I=2,NC
   10 KF(I)=KF(I-1)+KF(I)
        CLOOP=LOOP
        DO 30 NT=1,LOOP
C       COMPUTE STOICHIOMETRIC RATIOS
        DO 13 I=1,N
        IF(ESD(I))7,13,7
    7 E(I)=RN(ARG)*ESD(I)+E(I)
   13 CONTINUE
C       COMPUTE RECOVERIES
        DO 3 I=1,N
        IF(RSD(I))4,5,4
    5 R(I)=RM(I)
        GO TO 3
    4 R(I)=RN(ARG)*RSD(I)+RM(I)
    8 IF(R(I)-1.00)3,3,9
    9 R(I)=1.00
    3 CONTINUE
C       COMPUTE INPUT COSTS OF C(2),...,C(N)
        DO 19 I=2,N
        IF(CSD(I))11,12,11
   12 C(I)=CM(I)
        GO TO 19
   11 C(I)=RN(ARG)*CSD(I)+CM(I)
   19 CONTINUE
C       COMPUTE YIELD
        IF(YSD)24,25,24
   25 Y=YM
        GO TO 26
   24 Y=RN(ARG)*YSD+YM
```

```
   26 IF(Y)15,15,16
   15 Y=.00001
   16 IF(Y-1.00)17,17,18
   18 Y=1.00
C     COMPUTE RANDOM ORDINATE FROM FREQ DISTRIBUTION FROM STEP 1
C     FOR INPUT COST C(1)
   17 KOFC=RAND(ARG)*CLOOP
      DO 20 J=1,NC
      IF(KF(J)-KOFC)20,35,35
   35 CK=J
      IF(J-1)32,34,32
   34 DEN=KF(1)
      GO TO 33
   32 DEN=KF(J)-KF(J-1)
   33 DIF=KF(J)-KOFC
      DIF=CLSZE*DIF/DEN
      GO TO 21
   20 CONTINUE
   21 C(1)=CMIN+CK*CLSZE-DIF
C     COMPUTE COST OF FINAL PRODUCT
      TEMP=0.
      J=N-M
      DO 28 I=1,J
   28 TEMP= (E(I)*(1.-R(I))/Y+R(I))*C(I)*WM(I)+TEMP
   30 COST(NT)=TEMP/P
C     COMPUTE MIN AND MAX FINAL COSTS
      CMIN=COST(1)
      CMAX=CMIN
      DO 70 NT=2,LOOP
      IF(COST(NT)-CMIN)40,50,50
   40 CMIN=COST(NT)
   50 IF(COST(NT)-CMAX)70,70,60
   60 CMAX=COST(NT)
   70 CONTINUE
C     CLEAR KF ARRAY FOR FREQUENCY COUNT
      DO 80 I=1,NC
   80 KF(I)=0
C     COMPUTE CLASS SIZE
      FNC=NC
      CLSZE=(CMAX-CMIN)/FNC
C     COUNT NUMBER IN EACH CLASS
      DO 90 NT=1,LOOP
      A=(COST(NT)-CMIN)/CLSZE
      IF(COST(NT)-CMAX)87,85,85
   85 I=A
      GO TO 90
   87 I=A+1.
   90 KF(I)=KF(I)+1
```

```
C       OUTPUT FREQUENCY DISTR
        PUNCH 104,(I,NF(I),I=
C       FIND MEAN OF EACH CLASS
        COST(1)=CMI
        DO 96 I=2,
   96   COST(I)=COST(I-
C       OUTPUT MEAN OF
        PUNCH 101,(COST(I),
        GO TO 99
  100   FORMAT(11X,I
  101   FORMAT(5F10.4)
  104   FORMAT(I4,7X,I
  108   FORMAT(4I5)
        END
```

```
C       OUTPUT FREQUENCY DISTRIBUTION
        PUNCH 104,(I,KF(I),I=1,NC)
C       FIND MEAN OF EACH CLASS, USE 1ST NC WORDS OF COST ARRAY
        COST(1)=CMIN+CLSZE/2.
        DO 96 I=2,NC
   96   COST(I)=COST(I-1)+CLSZE
C       OUTPUT MEAN OF EACH CLASS
        PUNCH 101,(COST(I),I=1,NC)
        GO TO 99
  100   FORMAT(11X,I4)
  101   FORMAT(5F10.4)
  104   FORMAT(I4,7X,I4)
  108   FORMAT(4I5)
        END
```

Appendix D

DIRECT SEARCH MAIN ROUTINE

The calculation of the minimum cost is based on modifications by Wood of the Hooke and Jeeves method of Direct Search. Wood provides the requisite optimization procedure which can be regarded as a huge subroutine which operates on a particular problem, or set of equations. A specific program for each problem is required. This should not cause concern as the basic equations are easily programmed. To illustrate this point the main routine written for calculating the cost of Alternate Route I step 1 is provided. It consists of any necessary regression equations and the cost Equation 2-5.

The second order regression equation for the yield of the step is converted to a constraint equation by subtracting a slack variable X_4. The cost is calculated from the following input data:

XC(I) The cost of the first raw material.

XM(I) Its molecular weight.

E(I) Its excess ratio.

XMP The molecular weight of the product.

XC(2) The cost of the second raw material.

The excess of the second raw material is given by: .025(x(3)) + 1.05, an expression similar to Equation 4-2 in the text. X_3 is one of the variables whose setting the technique attempts to optimize.

```
C       A.H. BOBIS
C       OCTOBER 18 1965
   80 CONST(1) = B0 + B1*X(1) + B2*X(2) +B3*X(3) +B12*X(1)*X(2)
     1+ B13*X(1)*X(3) + B23*X(2)*X(3) + B11*X(1)**2 + B22*X(2)
     2**2 + B33*X(3)**2 - X(4)
      Y=CONST(1)+X(4)
      YAXMP = Y*XMP
      COST = XC(1)*XM(1)*E(1)/YAXMP + XC(2)*XM(2)*(.025*X(3)+1.05)
     1/YAXMP
      SN=COST+PEN(1)*CONST(1)**2
```

Appendix E

OUTPUT FROM MONTE CARLO SIMULATION

The results of the Monte Carlo simulations are given in Chapter 5 as smoothed curves. The raw data are presented in this Appendix. In all cases the 500 trials are partitioned among 20 classes. The form of the output for the several cases differ as the programs were continually modified.

TABLE E-1

12-9-65

MONTE CARLO COST ANALYSIS ROUTE 1 STEP 1

0 EXPTS

CMIN .0885 CMAX .0918
CLSZE .000166

CLASS	FREQUENCY
1	254
2	35
3	26
4	24
5	22
6	24
7	20
8	22
9	20
10	12
11	7
12	7
13	10
14	3
15	4
16	3
17	3
18	0
19	3
20	1

12-9-65

MONTE CARLO COST ANALYSIS ROUTE 1 STEP 1

5 EXPTS

CMIN .0885 CMAX .0901
CLSZE .000082

CLASS	FREQUENCY
1	277
2	35
3	21
4	33
5	20
6	20
7	21
8	11
9	10
10	14
11	18
12	11
13	3
14	2
15	1
16	2
17	0
18	0
19	0
20	1

12-9-65

MONTE CARLO COST ANALYSIS ROUTE 1 STEP 1

9 EXPTS

CMIN .0885 CMAX .0896
CLSZE .000052

CLASS	FREQUENCY
1	266
2	19
3	32
4	23
5	30
6	13
7	19
8	17
9	14
10	16
11	9
12	7
13	5
14	11
15	2
16	6
17	4
18	2
19	3
20	2

TABLE E-2

12-20-65

MONTE CARLO COST ANALYSIS ROUTE 1 STEP 2

0 EXPT STEP 1 1 EXPT STEP 2

CMIN .0997 CMAX .1095
CLSZE .00199

CLASS	FREQUENCY
1	2
2	7
3	15
4	19
5	55
6	45
7	63
8	70
9	64
10	57
11	40
12	22
13	18
14	10
15	4
16	3
17	3
18	2
19	0
20	1

12-20-65

MONTE CARLO COST ANALYSIS ROUTE 1 STEP 2

5 EXPT STEP 1 1 EXPT STEP 2

CMIN .0998 CMAX .1078
CLSZE .001785

CLASS	FREQUENCY
1	2
2	4
3	8
4	11
5	35
6	43
7	66
8	68
9	48
10	73
11	41
12	40
13	20
14	17
15	13
16	2
17	4
18	2
19	2
20	1

12-20-65

MONTE CARLO COST ANALYSIS ROUTE 1 STEP 2

9 EXPT STEP 1 1 EXPT STEP 2

CMIN .0996 CMAX .1074
CLSZE .00165

CLASS	FREQUENCY
1	1
2	3
3	7
4	11
5	8
6	33
7	37
8	48
9	82
10	70
11	51
12	46
13	43
14	26
15	14
16	8
17	7
18	3
19	0
20	2

TABLE E-3

1-5-66

MONTE CARLO COST ANALYSIS ROUTE 2

0 EXPTS

CMIN .0892 CMAX .1095
CLSZE .00102

CLASS	FREQUENCY
1	3
2	8
3	13
4	22
5	35
6	49
7	61
8	50
9	53
10	50
11	47
12	33
13	27
14	12
15	18
16	7
17	4
18	4
19	3
20	1

1-5-66

MONTE CARLO COST ANALYSIS ROUTE 2

2 EXPTS

CMIN .0912 CMAX .1062
CLSZE .00075

CLASS	FREQUENCY
1	10
2	5
3	14
4	28
5	33
6	45
7	51
8	46
9	53
10	48
11	56
12	36
13	26
14	20
15	13
16	10
17	3
18	0
19	2
20	1

1-5-66

MONTE CARLO COST ANALYSIS ROUTE 2

6 EXPTS

CMIN .0884 CMAX .0987
CLSZE .00051

CLASS	FREQUENCY
1	3
2	6
3	9
4	19
5	15
6	27
7	45
8	55
9	47
10	43
11	47
12	46
13	36
14	24
15	36
16	14
17	11
18	9
19	3
20	5

1-5-66

MONTE CARLO COST ANALYSIS ROUTE 2

9 EXPTS

CMIN .0891 CMAX .100798
CLSZE .00058

CLASS	FREQUENCY
1	5
2	7
3	14
4	25
5	35
6	56
7	45
8	50
9	46
10	58
11	56
12	34
13	24
14	23
15	13
16	3
17	2
18	3
19	0
20	1

TABLE E-4

MONTE CARLO ANALYSIS EXPECTED GAIN

CHOICE OF ROUTE 2 OVER ROUTE 1

ROUTE 1		ROUTE 2
STEP 1	0-EXPTS	0-EXPTS
STEP 2	1-EXPTS	

CLASS	FREQUENCY	MEAN
1	3	-.0047
2	2	-.0036
3	11	-.0025
4	8	-.0014
5	14	-.0004
6	24	.0006
7	30	.0017
8	35	.0028
9	40	.0039
10	54	.0050
11	61	.0061
12	52	.0071
13	62	.0082
14	42	.0093
15	29	.0104
16	15	.0113
17	9	.0126
18	6	.0136
19	1	.0147
20	2	.0158

MONTE CARLO ANALYSIS EXPECTED GAIN

CHOICE OF ROUTE 2 OVER ROUTE 1

ROUTE 1		ROUTE 2
STEP 1	0-EXPT	2-EXPT
STEP 2	1-EXPT	

CLASS	FREQUENCY	MEAN
1	1	-.0029
2	2	-.0020
3	6	-.0012
4	6	-.0003
5	6	.0004
6	23	.0013
7	29	.0022
8	49	.0030
9	49	.0039
10	54	.0047
11	49	.0056
12	54	.0064
13	42	.0073
14	44	.0081
15	31	.0090
16	18	.0098
17	10	.0107
18	8	.0116
19	6	.0124
20	3	.0133

MONTE CARLO ANALYSIS EXPECTED GAIN

CHOICE OF ROUTE 2 OVER ROUTE 1

ROUTE 1 ROUTE 2
STEP 1 0-EXPT 6-EXPT
STEP 2 1-EXPT

CLASS	FREQUENCY	MEAN
1	2	.0034
2	3	.0041
3	5	.0047
4	11	.0054
5	15	.0060
6	26	.0066
7	28	.0073
8	41	.0079
9	56	.0086
10	39	.0092
11	63	.0099
12	57	.0105
13	57	.0111
14	34	.0118
15	19	.0124
16	17	.0131
17	11	.0137
18	10	.0143
19	4	.0150
20	3	.0156

Appendix F

LACK OF FIT

The effectiveness of the prediction equations to represent the surface of experimental and estimated results is given by presenting the difference between the given and predicted values.

A good fit is obtained. The single exception is for run 13 Route II where the second order model fails to predict the expected low result given by the estimated value 30. The resulting weighted residual sum of squares would indicate a serious lack of fit. However, the variable setting for the particular case is so far removed from the region of interest that it does not influence the optimization procedure.

TABLE F-1

2-17-66

LACK OF FIT ROUTE 1 STEP 1

0 EXPTS

RUN	DATA	PRED	DIFF
1	.9000	.9172	-.0172
2	.9500	.9611	-.0111
3	.9100	.9302	-.0202
4	.9600	.9734	-.0134
5	.9300	.9392	-.0092
6	.9900	.9966	-.0066
7	.9400	.9502	-.0102
8	.9900	1.0069	-.0169
9	.9600	.9596	.0004
10	.9100	.9240	-.0140
11	.9800	.9944	-.0144
12	.9500	.9530	-.0030
13	.9700	.9693	.0007
14	.9400	.9385	.0015
15	.9600	.9774	-.0174

2-17-66

LACK OF FIT ROUTE 1 STEP 1

5 EXPTS

RUN	DATA	PRED	DIFF
1	.9100	.8887	.0212
2	.9750	.9639	.0110
3	.8553	.8659	-.0106
4	.9500	.9659	-.0159
5	.9749	.9637	.0111
6	.9950	.9829	.0120
7	.9500	.9614	-.0114
8	.9915	1.0054	-.0139
9	.9832	.9824	.0006
10	.8800	.9135	-.0335
11	.9950	.9970	-.0020
12	.9600	.9680	-.0080
13	.9700	.9678	.0021
14	.9100	.9200	-.0100
15	.9950	1.0002	-.0052

2-17-66

LACK OF FIT ROUTE 1 STEP 1
9 EXPTS

RUN	DATA	PRED	DIFF
1	.8429	.8471	-.0042
2	.9750	.9706	.0043
3	.8553	.8514	.0038
4	.9940	.9976	-.0036
5	.9749	.9712	.0036
6	.9853	.9891	-.0038
7	.9424	.9467	-.0043
8	.9915	.9872	.0042
9	.9832	.9831	.0001
10	.9300	.8879	.0420
11	.9950	1.0027	-.0077
12	.9800	.9794	.0005
13	.9800	.9810	-.0010
14	.9000	.9095	-.0095
15	.9900	.9891	.0008

TABLE F-2

2-17-66

LACK OF FIT ROUTE 2

0 EXPTS

RUN	DATA	PRED	DIFF
1	.6000	.5024	.0975
2	.7500	.7440	.0059
3	.8500	.8563	-.0063
4	.9400	.9408	-.0008
5	.6500	.6770	-.0270
6	.9000	.8971	.0028
7	.9200	.9270	-.0070
8	.9800	.9900	-.0100
9	.8000	.8647	-.0647
10	.7500	.7154	.0345
11	.9300	.9286	.0013
12	.6000	.6837	-.0837
13	1.0000	.9965	.0034
14	.7600	.7597	.0002
15	.9600	.9165	.0434

2-17-66

LACK OF FIT ROUTE 2

2 EXPTS

RUN	DATA	PRED	DIFF
1	.6000	.5067	.0932
2	.7700	.7586	.0113
3	.8500	.8506	-.0006
4	.9700	.9648	.0051
5	.6500	.6915	-.0415
6	.8750	.8744	.0005
7	.9400	.9339	.0060
8	.9800	.9791	.0008
9	.9500	.9489	.0010
10	.7500	.7422	.0077
11	.9300	.9502	-.0202
12	.6000	.6934	-.0934
13	1.0000	1.0074	-.0074
14	.7600	.8276	-.0676
15	.9600	.9670	-.0070

2-17-66

LACK OF FIT ROUTE 2

6 EXPTS

RUN	DATA	PRED	DIFF
1	.6400	.6326	.0073
2	.8500	.8661	
3	.9800	.9914	
4	.9600	.9543	
5	.5900	.6063	
6	.8750	.8683	
7	.9800	.9744	
8	.9600	.9658	
9	.9500	.9469	
10	.7500	.8027	
11	.9500	.9601	
12	.3000	.6609	
13	.9800	.9803	
14	.9500	.9685	
15	.9500	.9581	

2-17-66

LACK OF FIT ROUTE 2

9 EXPTS

RUN	DATA	PRED	DIFF
1	.6400	.6313	.0086
2	.8700	.8657	.0042
3	.9800	.9806	-.0006
4	.9600	.9600	0.0000
5	.6400	.6349	.0050
6	.8750	.8694	.0055
7	.9800	.9843	-.0043
8	.9600	.9637	-.0037
9	.9500	.9401	.0098
10	.7500	.8526	-.1026
11	.9500	1.0022	-.0522
12	.3500	.6387	-.2887
13	.9800	.9493	.0306
14	.8500	.9417	-.0917
15	.9500	.9469	.0030

APPENDIX G

SAMPLE CALCULATION

The frequency-cost distribution can be calculated at any point in the analysis. As an example a sample calculation is shown in which the information consists entirely of estimated results. A fictitious example is presented, which utilizes the estimates given in Chapter 5 for Route II.

The following reaction is considered:

$$A \xrightarrow[S]{C} B + \text{byproducts}$$

The pertinent data for the chemicals are summarized in Table G-1.

TABLE G-1

DATA FOR SAMPLE CALCULATION

Chemical	Purpose	Molecular Weight M_i	Cost C_i	Stoichiometric Excess E_i	Recovery R_i
A	Reactant	60	0.17	1	0.0
C	Catalyst	1	0.066	Variable	0.8
S	Solvent	1	0.090	Variable	0.96
B	Product	40		1	1.0

The only process response in this one-step synthesis is the yield. It is believed that the only three variables of interest are the temperature (x_1), the amount of catalyst (x_2) and the amount of solvent (x_3).

A central composite design is developed and estimates of the most likely and the best and worst possible results are made. These estimates are shown in Table G-2.

TABLE G-2

ESTIMATES FOR SAMPLE CALCULATION

Run	Conditions			Most Likely	Range
	x_1	x_2	x_3		
1	-	-	-	60	25-75
2	+	-	-	77	60-85
3	-	+	-	85	79-98
4	+	+	-	97	90-99
5	-	-	+	65	55-78
6	+	-	+	87	85-95
7	-	+	+	94	88-98
8	+	+	+	98	96-99
9	0	0	0	95	92-95
10	-1.4	0	0	75	68-88
11	1.4	0	0	93	90-98
12	0	-1.4	0	60	20-68
13	0	1.4	0	100	95-100
14	0	0	-1.4	76	70-92
15	0	0	1.4	96	92-99

where: x_1 at + = 100°C.; x_1 at - = 80°C.

x_2 at + = 0.01 lb./lb. A; x_2 at - = 0.001 lb./lb. A

x_3 at + = 10 lb./lb. A; x_3 at - = 5 lb./lb. A

The diagonal elements of the weighting matrix V can be calculated from the estimates of the range. Assuming the range to be four standard deviations apart, the variance for each of the points is computed. The diagonal elements for the sample problem are

156, 39.2, 22.6, 5.1, 33, 6.3, 6.3, 0.6, 0.6, 25, 4, 144, 1.6, 30.2, 3.1

Y is next set equal to the quadratic expression G-1.

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 \qquad \text{G-1}$$

The values for the coefficients are calculated with the use of equation G-2.

$$B = \left[X^T V^{-1} X\right]^{-1} \quad X^T V^{-1} Y \qquad \text{G-2}$$

The solution to G-2 for the sample problem is given in Table G-3.

TABLE G-3

COEFFICIENTS FOR SAMPLE CALCULATION

b_0	0.865
b_1	0.076
b_2	0.112
b_3	0.056
b_{12}	-0.039
b_{13}	-0.005
b_{23}	-0.026
b_{11}	-0.021
b_{22}	-0.012
b_{33}	-0.013

Equation 2-5 is then expanded to the form given in Equation G-3.

$$\text{Cost} = 0.17\ U_1 + 0.066\ E_2U_1\ (1-0.8) + 0.09\ E_3U_1\ (1-0.96) \qquad \text{G-3}$$

$$U_1 = \frac{60}{40\ Y} \qquad \text{G-3b}$$

$$E_2 = \left[10^{(0.5\left[x_2\right] - 0.5)}\right] 10^{-2} \qquad \text{G-3c}$$

$$E_3 = 2.5\ x_3 + 7.5 \qquad \text{G-3d}$$

Equation G-2 replaces Y and the coefficients given in Table G-3 are employed.

The direct search program is then used to compute the settings for the variables x_i that minimize the cost.

The result of this computation is:

$x_1 = 0.648$

$x_2 = 1.4$

$x_3 = -1.143$

Minimum Cost = \$0.2904

The yield at the minimum cost = 96.5%

The variance about the minimum cost is then calculated with the use of Equation G-4.

$$X^{*T} \left[X^T V^{-1} X \right]^{-1} X^* \epsilon^2 \qquad \text{G-4}$$

where X* is

1.000

0.648

1.400

-1.143

0.907

-0.740

-1.600

0.420

2.000

1.307

and ϵ^2 is 1×10^{-4}.

The variance calculated from this expression is 8.52×10^{-4}.

The Monte Carlo cost simulation program for a one-step process (program 0598A) is then used to determine the frequency-cost distribution about the indicated optimum.

The data in Table G-1 apply. The standard deviation about the optimum yield is 2.92×10^{-2}. The standard deviation about the estimated catalyst and solvent recoveries are 0.1 and 0.01, respectively.

The frequency-cost distribution developed from these data is shown in Table G-4.

TABLE G-4

FREQUENCY-COST DISTRIBUTION FOR SAMPLE PROBLEM

Class Mean $/Lb.	Frequency
0.2675	3
0.2705	3
0.2735	11
0.2765	31
0.2795	38
0.2825	46
0.2855	62
0.2885	54
0.2915	46
0.2945	57
0.2975	34
0.3005	39
0.3035	20
0.3065	23
0.3095	9
0.3125	15
0.3155	4
0.3185	2
0.3215	0
0.3245	3

REFERENCES

1. Kiefer, D. M., "Winds of Change in Industrial Chemical Research," C&E News, March 23, 1964, pp. 88-104.

2. Lillieholm, W. D., "For Research A Mighty Challenge," Chemical Week, August 14, 1965, pp. 49-76.

3. Danilov, V. J., "I.R. Annual Forecast 23-Billion for Research", Industrial Research, January 1966, pp. 31-44.

4. Dean, B. V. and Sengupta, S. S., "Research Budgeting and Project Selection", I.R.I. Transactions on Engineering Management, December 1962, pp. 158-169.

5. Freeman, R. J., "An Operational Analysis of Industrial Research", Ph.D. Dissertation, M.I.T., Cambridge, Massachusetts, June 1957.

6. Hess, S. W., "A Dynamic Programming Approach to R&D Budgeting and Project Selection," I.R.I. Transactions on Engineering Management, December 1962, pp. 170-179.

7. Saunder, W. E., "A Method for Allocating Research and Development Expenditures," Commercial Chemical Development Association, May 1965, pp. 19-38.

8. Asher, D. T., "A Linear Programming Model for the Allocation of R&D Efforts," I.R.I. Transactions on Engineering Management, December 1959, pp. 154-157.

9. Bobis, A. H. and Sprague, C. R., "An Exercise in Long Range Planning," Presentation I.R.I. National Meeting, November 13, 1965.

10. Horowitz, I., "Regression Models for Company Expenditures on and Returns from Research and Development," I.R.I. Transactions on Engineering Management, March 1957.

11. Finney, D. J., "An Introduction to the Theory of Experimental Design,"The University of Chicago Press, 1960.

12. Fisher, R. A., The Design of Experiments, Oliver & Boyd, London, 1935.

13. Fisher, R. A. and Yates, F., "The 6x6 Latin Squares," Proceedings of the Cambridge Philosophical Society, 30, pp. 492-502.

14. Finney, D. J., "Some Orthogonal Properties of the 4x4, 6x6 Latin Squares," Annals of Eugenics, 1944.

15. Davies, O. L., Editor, Statistical Methods in Research and Production, Oliver & Boyd, London, 1958.

16. Finney, D. J., "The Fractional Replication of Factorial Arrangements," Annals of Eugenics, 1945, pp. 291-301.

17. Box, G. E.P. and Wilson, K. B., "On the Experimental Attainment of Optimum Condition," Journal Royal Statistical Society, V. 13, 1951, pp. 1-45.

18. Box, G. E. P., "Multi-Factor Designs of First Order," Biometrika, Vol. 39, 1952, pp. 49-57.

19. Box, G. E. P. and Hunter, J. S., "Multi-Factor Experimental Designs for Exploring Response Surfaces," Annals of Math. Stat., Vol. 28, 1957, pp. 195-241.

20. Box, G. E. P. and Draper, N. R., "The Choice of a Second Order Rotatable Design," Biometrika, 50, 1936, pp. 335-352.

21. Box, G. E. P. and Draper, N. R., "A Basis for the Selection of Response Surface Design," J. American Statistical Association, 54, pp. 622-654.

22. Box, G. E. P., Presentation Gordon Research Conference, July, 1962.

23. Box, G. E. P. and Hunter, J. S., "The 2^{k-p} Fractional Factorial Designs," Part I, Technometrics, August 1966, pp. 311-351.

24. Box, G. E. P. and Hunter, J. S., "The 2^{k-p} Factorial Designs," Part II, Technometrics, Vol. 3, November 1961, pp. 449-477.

25. DeBaun, R. M., "Block Effects in the Determination of Optimum Conditions," Biometrics, Vol. 12, 1956, pp. 20-25.

26. Plackett, R. L. and Burman, J. P., "Design of Optimum Multi-factorial Experiments," Biometrika, Vol. 35, pp. 305-331.

27. Daniel, C., "Fractional Replication in Industrial Experimentation," 11th transactions ASQC.

28. Hunter, J. S., "Sequential Factorial Estimation," Technometrics, Vol. 6, 1964, pp. 41-55.

29. Cohen, J., "Scientific American," 197 N.O.S., pp. 128-135, Nov. 1951.

30. Savage, L. J., The Foundations of Statistical Inference, Nethuen & Co., Ltd., London, 1962.

31. Dean, J., Managerial Economics, Prentice-Hall, Englewood Cliffs, N. J., 1951, p. 568.

32. Hertz, D. B., "Risk Analysis in Capital Investment," Harvard Business Review, January 1964, pp. 95-106.

33. Bowman, E. H. and Fetter, R. B., Analysis for Production Management," McGraw-Hill, New York, 1959, pp. 387.

34. Schlaiffer, R., Probability and Statistics for Business Decisions, McGraw-Hill, New York, 1959.

35. Kahr, L. T., "Estimates of Subjective Probability," Presentation at the A.I.C.H.E. Meeting, December 9, 1965.

36. General Information Manual, PERT, IBM.

37. Anderson, L. B., "Factorial Design of Experiments," Chem. Eng., September 12, 1963, pp. 99-104.

38. Hunter, J. S., "Notes on the Methods for Fitting Response Surfaces," American Cyanamid Publication, 1960.

39. Hinchon, J., "Experimentation Without Data," Presentation at ASQC National Convention, May 1965.

40. Harrington, E. C., Jr., "A Procedure for Sequential Experimentation," CEP Symposium Series 42, Vol. 59, 1963, pp. 1-7.

41. Fisher, R. A., "On the Mathematical Foundations of Theoretical Statistics," Philosophical Transactions of the Royal Society, A, Vol. 2222, 1922.

42. Graybill, F. A., "An Introduction to Linear Statistical Models," McGraw-Hill, New York, 1961, pp. 117.

43. Box, G. E. P. and Tidwell, P. W., "Transformation of the Independent Variables," Technometrics, 4, November 1962, pp. 531-550.

44. Box, G. E. P. and Cox, D. R., "An Analysis of Transformations," JRSS Series B, 26, 1964, pp. 211-252.

45. Mood, A. M., Introduction to the Theory of Statistics, McGraw-Hill, New York, 1950, p. 192.

46. Hess, S. W. and Quigley, H. A., "Analysis of Risk Investments Using Monte Carlo Techniques," CEP Symposium Series, 42, Vol. 59, 1963,pp. 55-63.

47. Meyer, H. A., Editor, Symposium on Monte Carlo Methods, Wiley New York, 1956.

48. Larrivee, J. A., Mathematics and Computers, McGraw-Hill, New York, 1947, pp. 175-191.

49. Hooke, R. and Jeeves, T. A., "Direct Search," Solution of Numerical and Statistical Problems, Journal of the ACM, Vol. 8, No. 2, April 1961, pp.212-229.

50. Wood, C. F., "Optimization Subroutine for IBM 704 Computer Based on "Direct Search" Technique", Westinghouse Research Report 6-4, 1210 4-R2.

51. Wood, C. F., "Recent Developments in "Direct Search" Techniques," Westinghouse Research Report 62-159-522-R1, 1962.

52. Wilde, D. J., "An Experimental Search for an Optimum," Preprint. Prentice-Hall, Englewood Cliffs, N. J.

53. McArthur, D. W. and Heigl, J. J., "Strategy in Research," Transactions of the American Society for Quality Control, 1957, pp. 1-18.