Determination of Optimum Production Tolerances by Analog Simulation

R. B. McGHEE† AND A. LEVINE†

Introduction

THE ELECTRONIC analog computer has been widely utilized in recent years to obtain optimum parameter values for various types of control systems.[1,2] The procedure usually followed is to determine a control-system configuration by simplified analysis and then to program a computer in a realistic way, including all significant nonlinearities and appropriate random input disturbances. Next, a performance criterion of some sort is selected for the control system under study, and system parameters (gains, time constants, etc.) are varied until an optimum system is obtained. At this point, simulation activity usually ceases and is not resumed again unless the actual control system either cannot be constructed to obtain the desired parameter values or else fails to attain the level of performance indicated by the earlier simulation.

While this approach has produced many successful systems, it is the contention of the authors that it is seriously deficient in one respect: the essentially statistical nature of an actual system resulting from a manufacturing process is completely ignored. The representation of a control system as a differential equation with known coefficients is not realistic. An accurate simulation must effectively consider an ensemble of control systems, (*i.e.*, a random differential equation) characterized by parameters with common statistical properties. These statistical properties ordinarily take the form of production tolerances. Such tolerances must be considered concurrently with the search for optimum parameters if the simulator is to provide a reliable evaluation of the system performance.

The addition of random parameter values to a control-system simulation not only allows a more realistic determination of performance, but it also permits a more systematic and organized assignment of production tolerances. Conventionally, while the mean values for parameters are selected primarily on the basis of performance, the tolerances are selected primarily on the basis of cost. These two indices are somewhat at odds. For example, if cost were no object, but the desire to maximize performance dominated, then the ideal choice for each tolerance would be zero. On the other hand, if one desired simply to minimize cost per unit, infinite tolerances (one hundred percent acceptance of production) would be the proper policy. In the practical situation, a compromise between performance and cost must be effected. This compromise is too often accomplished on the sometimes dubious basis of intuition and "engineering judgment." It is unfortunate that such methods are used at a decision level so vital to the ultimate performance and cost of a product; more so, since the trade-off can be made more objective through analog simulation. By using a simulator to obtain performance as a function of the tolerance assignment, one can choose tolerances so as to maximize performance for a fixed cost or minimize cost for a fixed performance. In statistical terms, one can use the simulator to obtain optimum values for both the mean (nominal value) and standard deviation (production tolerances) for each important parameter.

A first step in this direction has been taken by R. C. Davis.[3] He postulated a linear dependence of a system performance index on parameter variances and obtained the necessary degradation coefficients by regression analysis of an analog computer experiment in which parameters were varied randomly. Although Davis did not explicitly mention cost, the coefficients he derived were utilized in juggling tolerances in such a way as to ease manufacturing difficulties without degrading performance. In principle, it would seem that these coefficients could be used to derive the original tolerances in an optimum manner. While Davis' method is probably the only sensible one if one is restricted to real-time simulation, it will be demonstrated with a simple practical example, that there are systems for which such a linearization can lead to wholly erroneous and misleading conclusions. The following paragraphs describe an analog simulation technique leading to an optimum tolerance-assignment which avoids the difficulties associated with an *a priori* assumption concerning the form of the functional dependence of performance on tolerances.

† Hughes Aircraft Co., Culver City, Calif.

[1] C. L. Johnson, *Analog Computer Techniques*, pp. 45–64, McGraw-Hill, New York, 1956.

[2] F. E. Nixon, *Principles of Automatic Controls*, pp. 287 to 305, Prentice-Hall, Englewood Cliffs, N. J., 1953.

[3] R. C. Davis, "A Statistical Method for Analyzing the Effects of Missile Guidance System Tolerance on Hit Probability," *Proc. Third Exploratory Conference on Missile Model Design for Reliability Prediction, White Sands Missile Range*, pp. 85–96, April 1959.

Proposed Tolerance Optimization Technique

Determination of the Performance Function

In order to arrive at an optimum set of tolerances, it is necessary first to determine the dependence of the control-system performance index, $P(\mathbf{Q})$, on the tolerance vector $\mathbf{Q}$. It is proposed that $P(\mathbf{Q})$ be determined by simulation according to the following procedure:

(1) Optimum nominal parameters are determined in the usual way.

(2) A tolerance vector, $\mathbf{Q}$, whose components are the standard deviations assigned to each parameter, is selected. This vector, together with the mean values determined in Step 1, is used to adjust filters operating on a stationary noise source so as to obtain voltages having the same mean and variance as the parameters under investigation.

(3) These voltages are fed to integrators as initial conditions while the computer is in *reset*. Each integrator in turn provides the input to a multiplier representing the value of the random coefficient in the simulation. When the computer is placed in *operate*, the integrators have no input. Thus, the value of the coefficient is fixed during the solution of the differential equation, but random when the computer is in *reset*. This effectively simulates a new production sample each time the computer is placed in operate.

(4) A large number of runs is made with a fixed tolerance-vector. An average performance-index is obtained for the whole set of runs (*e.g.*, probability of hit for a missile system).

(5) The components of the tolerance vector are varied in a systematic way, with a large number of computer runs being made for each new set of tolerances.

(6) The results are plotted with the aid of regression analysis using a digital computer. An arbitrarily high order of polynomial fit may be used if sufficient data is taken on the analog computer.

The enumerated steps lead to a representation of a system performance-function, $P(\mathbf{Q})$, without an *a priori* assumption regarding the form of the functional dependence of the system performance on parameter values.

It should be noted that step 5 may easily involve hundreds of thousands of runs. For this reason, it appears that the proposed simulation technique is economically feasible only when high-speed analog computers such as the GPS (General Purpose Simulator Instruments Company) or Philbrick Repetitive Computer are employed.[4] These computers are capable of providing solutions at rates up to 50 cps and can compute average performance indices simultaneously with solution of the differential equations of the control system.

Determination of the Cost Function

Once $P(\mathbf{Q})$ is obtained, it is necessary to determine the dependence of the unit cost of the system, $C(\mathbf{Q})$, on the tolerance assignment. This is a difficult problem which has long concerned manufacturing industries. It does not appear that analog simulation can aid in the determination of $C(\mathbf{Q})$. Nevertheless, $C(\mathbf{Q})$ is required in order to optimize $\mathbf{Q}$, so it is therefore necessary to discuss the estimation of costs sufficiently to indicate the difficulties to be expected and the general nature of the cost function.

Among the factors to be taken into account in the determination of $C(\mathbf{Q})$ are the raw cost per unit, salvage value, maintenance costs resulting from tolerances too wide or too narrow, lifetime, etc. Initial estimation of $C(\mathbf{Q})$ based on such considerations can probably best be accomplished by reference to manufacturing experience with previous products similar to the one under consideration. As production proceeds, the original estimate may be refined on the basis of actual costs. In addition to the experience accumulated by various manufacturing organizations, considerable literature exists on the subject of cost estimation.[5,6]

While the general problem is much too extensive to treat here, one can observe from experience that $C(\mathbf{Q})$ is generally hyperbolic in nature. That is, reduction of any element of $\mathbf{Q} = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ to zero will result in infinite cost while minimum cost is achieved by making all tolerances infinite.

Optimization of the Tolerance Vector

After $P(\mathbf{Q})$ has been determined, the optimization of $\mathbf{Q}$ may be considered. The first step required is to select a criterion for optimization. An obvious criterion is to optimize performance for a fixed cost. That is, on a surface of constant cost in the $\mathbf{Q}$-space, one seeks the point of maximum performance. On the other hand, it might be more desirable to minimize cost for a fixed performance. More complicated optimization criteria are possible by specifying the cost one is willing to accept as a function of performance, etc.

As more data becomes available, the estimate of the cost function will change, thereby shifting the

[4] Johnson, op. cit., pp. 222–232.

[5] D. H. Evans, "Optimum Tolerance Assignment to Yield Minimum Manufacturing Cost," *Bell System Technical Journal*, Vol. 37, pp. 461–484, March 1958.

[6] E. W. Pike and T. R. Silverberg, "Assigning Tolerances for Maximum Economy," *Machine Design*, Vol. 25, p. 139, Sept. 1953.

optimum. However, *the original performance function, P*(**Q**)*, remains valid; so the apportionment of tolerances may be altered to a new optimum without further simulation.*

Example

Description of Simulation

In order to establish the feasibility of the proposed technique, a simulation of a hypothetical radar-homing missile was undertaken. Besides random parameters, this simulation included random radar tracking noise and a random missile-heading error at launch. The performance index selected was probability of hit, P_h as measured against an idealized strip target. To limit the amount of data to be taken, only two parameters were varied randomly. These parameters were the navigation gain, K, and the principal missile filtering-time constant, τ. The missile which was simulated derived steering information from the angular rate of the line of sight from missile to target, so K is given by

$$K = \frac{\text{steady-state missile turning-rate}}{\text{line-of-sight angular rate}} \quad (1)$$

Fig. 1 is a functional block diagram of the missile simulated.

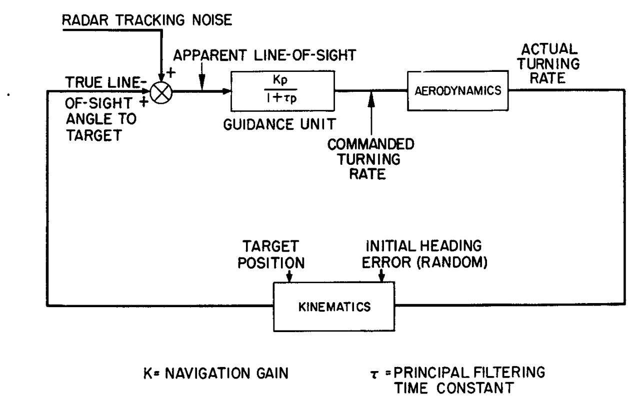

Fig. 1—Radar homing missile navigating in a plane.

To measure P_h with sufficient precision, it was necessary to simulate five hundred attacks for each value of the tolerance vector $\mathbf{Q} = (\sigma_K, \sigma_T)$. By considering four values for each tolerance, a total of sixteen cases was generated, resulting in a grand total of eight thousand simulated attacks. Since a high-speed computer was not available at the time, this simulation was performed on a real-time basis using automatic sequencing equipment. The total running time was about one week. Thus, this example shows quite clearly that a problem of realistic complexity will require high-speed simulation.

In this simulation, it was assumed that the population statistics for both K and τ were Gaussian with mean values K_0 and τ_0 and standard deviation σ_K and σ_T respectively. The sample values were therefore obtained by low-pass filtering of noise generators, using integrators which did not reset. Fig. 2 shows the computer program used to randomly vary the navigation gain, K, from run to run.

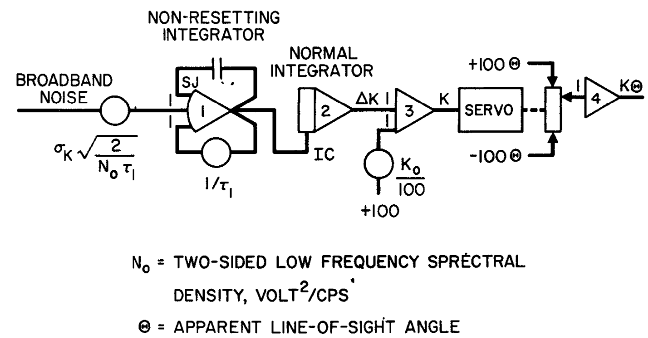

Fig. 2—Random variation of navigation gain.

σ_T \ σ_k	0%	10%	20%	30%
0%	0.34	0.30	0.31	0.29
10%	0.30	0.31	0.29	0.29
20%	0 30	0.26	0.30	0.30
30%	0.24	0.22	0.28	0.27

Fig. 3—Measured probability of hit as a function of parameter standard deviations.

Fig. 3 is a table of the results obtained in the simulation. This table presents the measured P_h as a function of the percentage standard deviation assigned to the randomized parameters, K and τ. The raw data of this table was subjected to quadratic regression analysis on σ_K and σ_T using a digital computer; *i.e.*, the data points were fitted in a least-square sense by

$$P_h = P_0 + a_1\sigma_K + a_2\sigma_T + a_{11}\sigma_K{}^2 + a_{12}\sigma_K\sigma_T + a_{22}\sigma_T{}^2 \quad (2)$$

The results are plotted in Fig. 4. Fig. 5 is a contour map of Fig. 4 with hypothetical cost curves included. Examples of minimum-cost and maximum-performance optima are indicated.

Significance of the Experimental Results

The first and most obvious observation that can be made from Fig. 4 is that linearization of the dependence of P_h on σ_K and σ_T is impossible. Linear dependence would result in equally spaced, parallel,

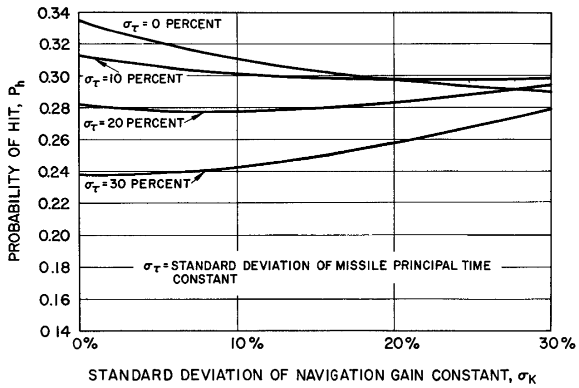

Fig. 4—Dependence of probability of hit on tolerance assignment for a homing missile.

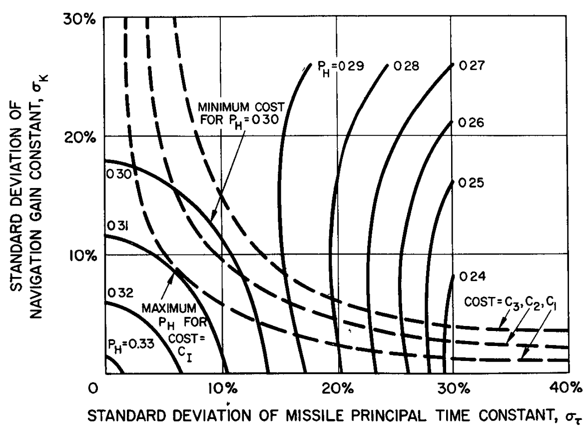

Fig. 5—Optimization of tolerances for minimum cost or maximum performance.

straight lines on Fig. 4. Replotting with $\sigma_K{}^2$ as the abscissa would show that linearization of the dependence of P_h on $\sigma_K{}^2$ and $\sigma_T{}^2$ is also impossible. Further examination of the figure reveals the remarkable fact that the performance of this system can actually be degraded by tightening tolerances. For example, when $\sigma_T = 30$ percent, the probability of hit is decreased from .28 to .24 as the tolerance on σ_K is reduced from 30 percent to 0 percent. This is quite at odds with the widespread opinion that when tolerances are relaxed in one part of a system, they must be tightened elsewhere to maintain the system performance. In the present example, this apparently anamolous behavior results from a cancellation of effects when K and τ both vary widely from their nominal values.

A more general and far reaching conclusion arising from this example is concerned with the form of the mathematical model selected to represent the dependence of the system-performance index on parameter *values*. In the more conventional approach to the study of tolerances, one begins by assuming a specific model for this dependence.[7] For example, on the basis of experience gained in the optimization of nominal values, the system designer might decide that an adequate model for the relationship between P_h and the values of the parameters, K and τ, is given by the quadratic expression:

$$P_h = \gamma_0 + \gamma_1(K - K_0) + \gamma_2(\tau - \tau_0) + \gamma_{11}(K - K_0)^2 + \gamma_{12}(\tau - \tau_0)(K - K_0) + \gamma_{22}(\tau - \tau_0)^2 \quad (3)$$

Now since the values of K and τ were independent in this experiment (and very often in practice) the average or expected value of P_h measured over the ensemble of missiles is given by

$$P_h^{ave} = E(P_h) = \gamma_0 + \gamma_{11}E(K - K_0)^2 + \gamma_{22}E(\tau - \tau_0)^2 \quad (4)$$

where the symbol E denotes the expected value or theoretical average resulting from repeated estimation of P_h. The other terms vanish because they have zero expected value. Now, by definition, the ensemble averages in the above expression are just the variances; *i.e.*,

$$E(K - K_0)^2 = \sigma_K{}^2 = \text{variance of } K \quad (5)$$

$$E(\tau - \tau_0)^2 = \sigma_T{}^2 = \text{variance of } \tau \quad (6)$$

so

$$P_h^{ave} = \gamma_0 + \gamma_{11}\sigma_K{}^2 + \gamma_{22}\sigma^{k2} \quad (7)$$

This is the analytic origin of the common assumption that performance depends linearly on the variances. If one accepts this assumption, then all that is required is to perform an experiment, selecting many values of the variables and recording the performance index along with the values of the variables. Regression analysis is then accomplished on equation (3) yielding the desired coefficients, γ_0, γ_{11}, and γ_{22}.

While the above approach will succeed if the assumed model (3) is in fact correct, it has already been shown that the resulting conclusion of linear dependence of performance on parameter variances cannot possibly describe the behavior of the example at hand. The reason is that the effect of the terms σ_1, σ_2, and $\sigma_1\sigma_2$ are absent from (7) whereas a regression analysis of the experimental results based on (2) shows that $\sigma_1\sigma_2$ is in fact the most significant term. This term represents interaction effects which are of prime importance in this system. The actual financial cost of ignoring this term could be considerable since utilization of (7) would lead both to unnecessarily tight tolerances on K and to degraded performance when σ_T is large.

The process of deriving the effects of tolerances from a model for performance-dependence on parameter values can be turned around. That is, if one knows the coefficients in an equation like (2),

$$P_h = a_0 + a_1\sigma_K + a_2\sigma_T + a_{11}\sigma_K{}^2 + a_{12}\sigma_K\sigma_T + a_{22}\sigma_T{}^2 \quad (2)$$

then the effect of parameter values can be inferred (though not uniquely). For example, if it were true

[7] Evans, op. cit.

that

$$P_h = a_0 + a'_1 \mid K - K_0 \mid + a'_2 \mid \tau - \tau_0 \mid + a_{11}(K - K_0)^2 + a'_{12} \mid K - K_0 \mid \mid \tau - \tau_0 \mid + a_{22}(\tau - \tau_0)^2 \quad (8)$$

then (2) would follow and the coefficients in (8) could be obtained from a regression on tolerances such as was performed in the present example.

To summarize, in the light of the preceding discussion, it is maintained that one should in general reverse the usual approach to tolerance analysis by:

1. Performing a simulation with randomized coefficients as described in the section above on Tolerance Optimization.
2. Derive the mathematical model for the system from a regression analysis of the dependence of system performance on the tolerance vector.

In this way, new insight can be obtained into the relationship between performance and parameter values, and serious errors in tolerance assignment can be avoided.

Design of Experiments

The Need for Experimental Design

The experiment just described involved only two parameters. In a more realistic problem, it is likely that a much larger number of system parameters would be considered in the assignment of optimum tolerances. Utilization of the method presented in this paper leads to an alarming increase in the total number of runs as the number of parameters increases unless the experiment is carefully designed. The problem of determining just how to take data so as to obtain the most information with a minimum of effort falls within the realm of statistical design of experiment. This field is so extensive that it is neither possible nor desirable to consider it in detail in this paper. Fortunately, there are several excellent texts on the subject which treat a variety of types of experiment.[8]

While a general discussion of design of experiment is not appropriate, there are two aspects of this field which are particularily useful in experiments of the type discussed in this paper. In the first place, it is necessary to be able to estimate the total amount of data required to define $P(\mathbf{Q})$ before beginning the experiment. Secondly, in the event that this number is too large, it is very desirable to have a means for reducing this total by a judicious choice of data points in the tolerance-vector space. Both of these problems are considered below in connection with the present experiment.

Estimation of the Amount of Data

In the missile simulated, the measured values of the performance index, P_h, possess a binomial distribution, because they result from repeated trials of the same experiment. One can therefore estimate the variability of successive measurements of P_h.[9] It turns out that if P_h is the true or theoretical probability of hit and $\hat{P}_h$ is the measured ratio of hits to total firings, then

$$P_h = \hat{P}_h \pm 1/\sqrt{n} \quad (9)$$

n = total number of simulated firings

with probability .95 or better. Consequently, since the data in Fig. 3 was obtained by making five hundred simulated firings per data point,

$$P_h = \hat{P}_h \pm 1/\sqrt{500} \cong \hat{P}_h \pm .045 \quad (10)$$

That is, we can have at least ninety-five percent confidence that the true P_h lies within $\pm.045$ of the values listed in Fig. 3.[10] Utilization of quadratic regression (2) considerably improves the confidence we may have in the final result of Fig. 3.

If one were to decide that a precision of $\pm.045$ with ninety-five percent confidence was required of every point, and furthermore, that m values of each of k parameters must be considered, then for this example the total number of runs would be given by

$$n = 500\, m^k \quad (11)$$

This number may be so large as to make a "brute force" determination of $P(\mathbf{Q})$ impractical, even with a high speed computer. In such an event, some means such as the experimental design described below must be found for reducing this total.

Factorial Design

In the example treated in this paper, the data points were taken at uniformly spaced points in a σ_K, σ_T space (Fig. 3). Fortunately, it is not necessary to take data at all of these points to determine a quadratic relationship between $P(\mathbf{Q})$ and $\mathbf{Q}$. One type of design of experiment which allows fewer data points to be considered is called "fractional factorial design".[11] This design method involves a sequential determination of the points at which observations should be made, and is particularly well suited to the determination of $P(\mathbf{Q})$. The total number of computer runs may be drastically reduced, amounting possibly to several orders of magnitude in the total effort required to define $P(\mathbf{Q})$. While the complete

[8] W. G. Cochran and G. M. Cox, *Experimental Designs*, John Wiley and Sons, New York, 1950.

[9] P. G. Hoel, *Introduction to Mathematical Statistics*, John Wiley and Sons, New York, Chapter 10, 1947.

[10] Methods also exist for the estimation of confidence intervals in the event that the variables do not possess a binomial distribution. *Ibid.*, Chapter 10.

[11] J. S. Hunter, "Designing and Interpreting Tests," *Control Engineering*, pp. 137–141, Sept. 1959.

σ_T \ σ_k	0%	30%
0%	0 34	0 29
30%	0 24	0.27

Fig. 6—Measured probability of hit as a function of parameter standard deviations.

theory is somewhat involved, Fig. 3 can be used to illustrate the basic idea of fractional factorial design.

Referring to the experimental example, if one knew the relative importance of the terms σ_K, σ_T, and $\sigma_K\sigma_T$ in determining $P(\mathbf{Q})$, then a more intelligent selection of sampling points could be made. Now an *estimate* of the relative effects can be obtained by evaluting $P(\mathbf{Q})$ only at the vertices of the table of Fig. 3. Fig. 6 is an abbreviated table, obtained by using only the four corners of the other table. The relative effect of each of the three terms above is obtained as follows:

(1) Estimated effect of σ_K = average of P_h over column (2) − average P_h over column (1) = $(.29 + .27)/2 - (.34 + .24)/ = -.01$

(2) Estimated effect of σ_T = average of P_h over row (2) − average P_h over row (1) = $(.24 + .27)/2 - (.34 - .29)/2 = -.06$

(3) Estimated effect of $\sigma_K\sigma_T$ = average difference of row (2) − average difference of row (1) = $(.27 - .24)/2 - (.29 - .34)/2 = .04$

Thus it can be seen that the "interaction effect" of $\sigma_K\sigma_T$ is greater than the "main effect" due to σ_K and nearly as great as the "main effect" due to σ_T. While this result is obtained from only four data points, it is (qualitatively) quite in accord with the final result obtained by regression analysis (Fig. 4). The significance of this test is simply that the $\sigma_K\sigma_T$ term cannot be dropped from the regression equation (2). There is, however, a possibility of dropping the linear term σ_K. If this were done, then less data would be required to estimate the remaining regression coefficients. A more sophisticated approach to fractional factorial design must consider confidence intervals and other effects to determine whether or not a given term can be dropped from (2).

When a large number of parameter tolerances are involved in $\mathbf{Q}$, the dropping of even one term from the regression equation may provide a considerable saving in the simulation effort required to obtain $P(\mathbf{Q})$. Even if no terms can be dropped from the regression equation, uniformly-spaced data points do not provide the best grid for measurement of $P(\mathbf{Q})$. Further development of fractional factorial design provides a solution to the problem of determining the optimum location for data points in the parameter-tolerance space.

Special Simulation Considerations

Correlation Between Successive Parameter Values

In order to simplify the analysis of data, it is desirable that successive values of parameters selected from noise generators be uncorrelated. For noise sources shaped by first-order filtering, the correlation of output samples separated by time T is given by

$$\rho(T) = e^{-|T|/\tau_1} \qquad (12)$$

where τ_1 is the filter time constant.[12] Consequently,

Multiplexing of Noise Generators

In a problem of realistic complexity, a sizable number of uncorrelated noise voltages are required to provide parameter values. These voltages can be obtained from a single noise generator by frequency-multiplexing techniques. Rice has shown[13] that an ensemble of samples drawn from a Gaussian noise generator can be described by a Fourier series with uncorrelated random coefficients. Consequently, if such a signal is passed through a bank of bandpass filters, the correlation between filter outputs can be made as small as desired by reducing the spectral overlap of the filters sufficiently. In this process, the correlation between successive samples from a single filter must also be considered. This factor sets a lower limit on the bandwidth of the individual filters.

High-Speed Data Processing

When solution rates are as high as the 50-cps capability of repetitive computers, data reduction must be at least partially accomplished by the computer. One such computer (GPS) has a probability-distribution analyser which automatically provides probability of hit. This means that there is no need to process individual solutions off the computer. It appears that such an instrument or a related one is essential in simulations where hundreds of thousands of runs are required.

Conclusions

The role of simulation in the preliminary design of control systems can be expanded to include selection of optimum manufacturing tolerances. If a high-speed computer is available, one may abandon linear models for the effect of varying tolerances and thereby obtain a realistic method for performing the crucial trade-off between cost and performance. A lesser effort is apt to lead to false conclusions concerning system performance.

[12] J. L. Lawson and G. E. Uhlenbeck, *Threshold Signals*, p. 42' McGraw-Hill, New York, 1950.

[13] S. O. Rice, "Mathematical Analysis of Random Noise," *Selected Papers on Noise and Stochastic Processes*, N. Wax, ed., pp. 157–161, Dover Publications, New York, 1954.