The Identification of Nonlinear Reaction Systems with Two-Level Inputs

G. H. HARRIS and LEON LAPIDUS

Princeton University, Princeton, New Jersey

The general problem of characterizing a nonlinear, physical process from its input-output records is examined. A method of identifying the class of nonlinear, time-invariant processes with switched two-level inputs and finite memories is successfully applied to several nonlinear systems. The identification model consists of a finite, orthogonal series expansion for the output as a function of the past of the input. Of particular interest is the characterization and synthesis of a nonlinear, continuous stirred-tank reactor. This system, as well as others that are investigated, is simulated on a digital computer and identified after an initial observation period. No specific knowledge of the unknown process is presumed; however, only moderately fast input switching is permitted. After synthesizing the identification model, the process and model outputs are compared.

The analysis of a dynamic physical process and the intelligent design of its automatic control system can be accomplished only if the dynamic characteristics of the process are known. In its most general interpretation, knowledge of process dynamics implies cognizance of the functional relations between the inputs and outputs of a system. The fundamental problem of identifying an unknown process, or determining its input-output relationships, is termed the *identification problem*. Any technique that promises to characterize these process dynamics must necessarily have an experimental basis if initially the process characteristics are truly unknown.

The identification problem is conveniently subdivided into several classifications, one of the most crucial being whether the process is linear or nonlinear. The identification of linear systems, although still relatively difficult to achieve in practice, has a firm theoretical foundation and has, in certain simplified cases, succumbed to a host of mathematical techniques. Indeed, the identification of linear process dynamics is perhaps the most important single problem in linear system analysis and feedback control theory at the present time. However, the identification of nonlinear systems, a vastly more important and difficult problem, has received scant attention until recently. Because the entire realm of nonlinear mathematics is exceedingly intractable, it comes as no surprise that little success in the identification of nonlinear process dynamics has been achieved.

A second important classification of the identification problem is whether the system is time invariant or time varying. The successful identification of process dynamics, whether the system is linear or not, depends critically upon the degree to which the process parameters vary with time. If the characteristics of the system vary appreciably with respect to the observation time required to identify them, an accurate identification will be exceedingly difficult, if not impossible, to achieve. For this reason the ensuing development pertains generally to time-invariant systems or to those systems whose process parameters vary slowly over the observation interval. In addition, attention is limited to systems with finite memories or, more precisely, to systems in which the influence of the past input on the present output decays with time.

The identification of the class of nonlinear, time-invariant processes with switched two-level inputs and finite memories is of considerable interest for several reasons. Two-position control action (also called on-off, bang-bang, and relay control action), the simplest type in use, is undoubtedly the most widely employed mode of feedback controller action in industry (*9*). Because of the current interest in the design of optimal control systems, systems whose control variables are bounded are frequently optimized or suboptimized by setting the controls at their maximum or minimum values, that is, bang-bang control. For certain systems (*5, 7, 15*), the solution to the time-optimal control problem is known to be bang-bang over the entire control period or at least a portion of it. In other cases bang-bang control may furnish a practical solution to the optimal control problem (*11*).

HISTORICAL BACKGROUND

Since the area of investigation described in this paper is basically new to chemical engineers, the authors feel that a certain amount of historical background is of interest and importance. Nonlinear synthesis may be thought of as evolving from the theoretical development of Wiener (*21*). This theory is the basis for the experimental analysis of a very broad class of nonlinear systems, being applicable to nearly any physically realizable, time-invari-

G. H. Harris is at Arthur D. Little, Inc., Cambridge, Massachusetts.

ant, nonlinear system. Included in this class are linear systems and zero-memory (variously called time-independent, no-storage, or noninertial) systems. It must be stressed that unlike other techniques which are limited to a very narrow class of nonlinear systems, for example, the describing-function approach, Wiener's method is quite general in scope, treating systems from the "black box" viewpoint. Although the theoretical value of such an approach cannot be overstated, it should be anticipated that the practical application of the theory might present severe problems. To date no analysis of any real nonlinear system has been reported, nor have any specific details concerning the implementation of Wiener's theory appeared.

Briefly, Wiener's method consists of exposing an unknown nonlinear system to a random noise input and then measuring certain averages of products of its output with nonlinear functions of the input. These averages are the characterizing coefficients of the nonlinear system. Knowing these coefficients, one can, in theory, synthesize a mathematical or computer model equivalent in some sense to the nonlinear system under test. Both the analysis and synthesis aspects of the method require the availability of rather extensive computer facilities.

Wiener's theory is limited to stable systems with finite memories. These systems possess the necessary property that the present output is influenced to an arbitrarily small extent by that portion of the past of the input beyond some arbitrarily large but finite time. This means, generally, that in such systems the influence of the past input on the present output attenuates with time. In addition, attention is restricted to nonlinear systems operating on continuous, real-valued functions of time that are integrable square on the semi-infinite interval $(-\infty, 0)$, thereby generating continuous, real, time functions as outputs. All physical systems with finite physical inputs, both in amplitude and duration, are included in this classification.

Wiener's fundamental contribution to nonlinear system analysis stems from the introduction of random noise as a general probe for these systems. He has postulated that the response to white Gaussian noise, noise with a Gaussian amplitude distribution and a flat power density spectrum, can serve to characterize or identify nonlinear systems.

As is generally known, nonlinear systems, in contradistinction to linear ones, must be tested with every conceivable input if the response to every input is to be known. With white Gaussian noise as the test signal, there is a finite, nonzero probability that any given time function will be approximated arbitrarily closely over any finite time interval by some sample of this noise. Since white noise possesses a flat frequency spectrum over all frequencies, the system is effectively exposed to every physical input. Two systems which respond identically to white Gaussian noise are said to be equivalent and will have the same response to any given input. It is this property of the white Gaussian input that materially simplifies the problem of introducing every step input into the system and leads to its eventual characterization.

Ideal white noise, with a constant spectral density for all frequencies, cannot be physically realized because the power of the noise signal is infinite. For practical purposes, however, a signal possessing a flat frequency spectrum over a frequency band considerably wider than the system bandwidth can be considered white noise; this is band-limited white noise. White Gaussian noise, alternatively referred to by Wiener as Brownian motion or shot noise, can be approximated by the output of a shot-noise generator. The shot effect, output of a shot-noise generator, is, then, the actual physical input postulated as the universal probe for nonlinear systems.

Under Wiener's guidance these ideas have been extended by several writers such as Bose (*8*), Singleton (*19*), and others. It is Bose's extension of Wiener's basic theory that leads to a practical solution of the characterization of nonlinear systems with two-level inputs. Such a procedure was elucidated by Roy and DeRusso (*17, 18*) and by Harris (*12*). For a detailed discussion of the basic Wiener approach, the book of Mishkin and Braun (*16*) and the work of Harris (*12*) is recommended.

SUMMARY OF CURRENT WORK

In this paper a method for identifying the class of nonlinear, time-invariant processes with switched two-level inputs and finite memories is successfully applied to several nonlinear systems. After an initial observation period during which the system is identified, a model of the real process is synthesized and the model output for an arbitrarily switched input obtained. The nonlinear processes themselves are simulated on a digital computer. Since the real processes were, in fact, known a priori, a measure of the accuracy of the model could be obtained for those systems investigated.

The identification model for such processes consists of a finite series expansion for the output as a function of the past of the input. The terms of the expansion are mutually orthogonal in time, independent of the two-level input time function. The orthogonalization is achieved by partitioning the input space into nonoverlapping or orthogonal cells. At any instant the past of the input resides in one particular cell; the model output is given by the value of the previously determined orthogonal expansion coefficient assigned to that cell. The model coefficients are determined via the minimization of a mean-square-error criterion. As a result of this minimization, each coefficient is evaluated by averaging the actual system output during the time that a particular cell is occupied. In order to reduce the number of coefficients in the model, the switching time is not permitted to be less than half the effective duration of the system's memory.

Whereas the work of Roy and DeRusso was essentially theoretical in postulating the identification model, this current study investigates its practical application to complex systems. In so doing, the scope of the model is significantly expanded due to the availability of large, high-speed digital computers.

This investigation goes beyond prior work in an attempt to determine the influence of the following three critical parameters on the accuracy of model simulations: (1) number of taps (sampling points) along the input memory, (2) distribution of taps along the input memory, and (3) estimate of the effective duration of the system's memory. In addition, the important subjects of (1) choice of an error weighting function, (2) multivariable systems, (3) identification criteria, (4) minimum identification time, and (5) improvement of simulation characteristics are discussed, and, in the latter four cases, investigated through actual system identifications.

Two nonlinear systems selected for a test of the identification model are discussed here in detail. A second-order system with a saturation nonlinearity was initially chosen because it possesses a relatively mild nonlinearity, thereby simplifying the analysis of results. Another system is a highly nonlinear chemical reaction process with multiple outputs; brief mention is also made of a time varying version of the reaction process (*12*). Heretofore, only strictly time-invariant systems had been considered.

For these systems the effect of varying the above three model parameters is shown. Models with as many as 752

coefficients have been employed, which far exceeds the 50-coefficient models of Roy and DeRusso. The increased speed of computation has made possible a hundredfold increase in the available identification time (real system time) and permitted for the first time a system identification to be completed in minimum real time. These improvements in the complexity of the model and in the identification time, not to mention improved tap distributions, lead to practical and highly accurate simulations.

THEORY OF THE IDENTIFICATION MODEL

In this section we summarize the important details of the theory for the identification model. Further analysis can be found in the work of Roy and DeRusso (*17, 18*) and of Harris (*12*).

Characterizing the Two-Level Input

The input $x(t)$ to the system is assumed to switch between two levels denoted by amplitudes of -1 and $+1$. If the input to a real system switches instead between levels A and B, where $A < B$, the linear substitution

$$x_1(t) = \frac{2x(t) - (A+B)}{B-A} = \begin{cases} -1 & \text{if } x = A \\ +1 & \text{if } x = B \end{cases} \quad (1)$$

will transform these levels to -1 and $+1$, respectively.

Since the present value of the output of a general nonlinear system is determined by the past history of its input, it is natural to first consider the characterization of this time function. This characterization is most easily accomplished by sampling the input at n uniform time intervals T and constructing the finite time sequence

$$x(t) \doteq [x(0),\ x(T),\ x(2T), \ldots,\ x((n-1)T)] = [x_0, x_1, x_2, \ldots, x_{n-1}] \quad (2)$$

where $(n-1) = T_m/T$. T_m, an estimate of the effective duration of the system's memory, is frequently chosen to be the system settling time T_s. The settling time is the time required for the system response to a test signal to come within a specified proximity of the system's final value. The two test signals most frequently used are the unit impulse and the unit step function. It is clear from Equation (2) that as $T \to 0$, that is, the sampling frequency approaches infinity and the number of samples $n \to \infty$, the representation of $x(t)$ improves.

Orthogonal Expansion of the Output

The n sampled values of $x(t)$ characterizing the input past are conveniently stored in a digital shift register tapped at n points. At any given time each tap can exist at only one level, so that at each instant one and only one input state (set of n tap levels) can exist in the memory device. The input space consisting of all possible input configurations, of which there are 2^n, has thus been partitioned into nonoverlapping cells. If an expansion is made of the system output as a function of the individual tap levels, a model orthogonal in time will result. This means that each term in the expansion represents the system output for one particular input configuration, that is, there is one term for each input state.

An orthogonal expansion is required so that each term can be evaluated independently of all other terms. The orthogonal expansion is given in Equation (3), where $\hat{y}(t)$ is the model output.

$$\hat{y}(t) = \sum_{\alpha} A_\alpha \Phi(\alpha) \quad (3)$$

A_β is a constant assigned to the β^{th} configuration of tap levels, and α is an index of summation which ranges over the 2^n cells of the input space. The function $\Phi(\beta)$ takes on a value of either 1 or 0, depending upon whether the β^{th} configuration is present or not, respectively.

At any instant in time, the past of the input resides in a particular cell in the input configuration space, and only the coefficient assigned to that cell is nonzero. Suppose this is the β^{th} cell. Then $\Phi(\beta) = 1$ and $\hat{y} = A_\beta$. The remaining $(2^n - 1)$ Φ's are zero.

Since the model output is given by a finite sequence of numbers, $\hat{y}(t)$ is a "staircase" approximation to the actual system output. As the number of taps in the input memory is increased (the input is sampled more frequently), the input characterization becomes more accurate, and as a direct result the orthogonal expansion can be expected to approximate more closely the actual system output.

The coefficients $\{A_\alpha\}$ are determined by minimizing the weighted mean-square error between the actual system output $y(t)$ and the orthogonal model output $\hat{y}(t)$. The weighted mean-square error is given by

$$\overline{e^2} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} w(t)\ [y(t) - \hat{y}(t)]^2 dt \quad (4)$$

where $w(t)$ is a nonnegative weighting function.

Substitution of Equation (3) into Equation (4) gives

$$\overline{e^2} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} w(t)\ [y(t) - \sum_{\alpha} A_\alpha \Phi(\alpha)]^2 dt. \quad (5)$$

This expression is to be minimized with respect to the coefficients $\{A_\alpha\}$. The result may be readily shown to be given by Equation (6):

$$A_\beta = \frac{\overline{y(t)\ w(t)\ \Phi(\beta)}}{\overline{w(t)\Phi(\beta)}} \quad (6)$$

which is the desired expression relating the orthogonal coefficients to certain experimentally measurable time averages of the system's input and output.

Because the weighted mean-square error criterion was chosen, the orthogonal coefficients are to be evaluated by averaging certain quantities over an infinite period of time. In any practical application, of course, this is not possible, since measurements must necessarily be confined to a finite time interval. The coefficients computed on the basis of finite time averaging are such that the finite weighted mean-square error over the elapsed time interval is a minimum. In essence, then, the model coefficients have been determined a posteriori for a finite binary input and are to be used to simulate the system's response to future binary signals. It is therefore desirable that the characteristics of the probing signal during identification be similar to the characteristics of the model input during synthesis, or simulation. In this regard a random binary input might be a logical starting point in the event nothing is known about the future input to the model during simulation. It is also possible that the binary input is random due to conditions beyond the investigator's control. The particular distribution of switching times for a random binary signal will certainly affect the model coefficients evaluated according to Equation (6).

Minimum Mean-Square Error

The minimum mean-square error (MSE) of the identification process is realized when the coefficients $\{A_\alpha\}$ are determined according to Equation (6), and they can be shown to be given by the following expression:

$$\overline{e^2_{\min}} = \overline{y^2(t)w(t)} - \sum_{\alpha} A_\alpha^2\, \overline{w(t)\Phi(\alpha)} \quad (7)$$

Some of the results obtained thus far may be given interesting interpretations when the weighting function $w(t)$ is a constant. Specifically, consider that $w(t) = 1$; the quantity $\overline{\Phi(\beta)}$ may be interpreted as the probability that the β^{th} cell in the input space is occupied at any instant since

$$\overline{\Phi(\beta)} < 1$$

and

$$\sum_{\alpha} \overline{\Phi(\alpha)} = 1$$

Hence, the coefficient A_β in Equation (6) is simply equal to the average value of the system output $y(t)$ over the time that the input configuration β exists, this average taken over an infinitely long observation period. That is to say, A_β is the conditional mean of $y(t)$ assuming that the β^{th} cell in the input space is occupied, or

$$A_\beta = [\overline{y(t)} \mid \Phi(\beta) = 1]$$

The minimum MSE is seen to be the difference between the time average of the square of the system output and the sum of the squares of the orthogonal coefficients, each multiplied by its respective probability of occurrence. Hence, the minimum MSE is simply the difference between the mean-square values of the system output and the model output:

$$\overline{e^2_{\min}} = \overline{y^2(t)} - \overline{\hat{y}^2(t)} \tag{8}$$

for $w(t) = 1$.

Error Due to Finite Time Averaging

The coefficients $\{A_\alpha\}$ are determined according to Equation (6) by forming the ratio of two time-averaged quantities, these averages extending over an infinitely long period of time. Since the duration of any experiments must necessarily be finite, the computed coefficients will be in error; consequently, the minimum MSE, given by Equation (7), cannot be realized. The MSE of the model using the inaccurate coefficients $\overline{e^2}$ will be greater than the minimum MSE, $\overline{e^2_{\min}}$. Roy, Miller, and DeRusso (*18*) have obtained an expression for the average additional MSE, $<\overline{e^2} - \overline{e^2_{\min}}>$, resulting from the use of inaccurate coefficients. This average additional MSE is less than or equal to $\overline{e^2_{\min}}$. Furthermore, if each input configuration occurs at least M times in the input record, the average additional MSE is at most $\overline{e^2_{\min}}/M$.

The average additional MSE, $<\overline{e^2} - \overline{e^2_{\min}}>$, is reduced if the computation of the $\{A_\alpha\}$ is based on an input record containing several distinct occurrences of each input configuration. Successive occurrences of the configuration α are called distinct provided one or more different configurations intervene in the input record. If the input record contains the configuration α M distinct times, then M coefficients $\{A_{\alpha,i}\}$ can be computed, and the sample mean $\overline{A}_\alpha$ evaluated according to

$$\overline{A}_\alpha = \frac{1}{M} \sum_{i=1}^{M} A_{\alpha,i} \tag{9}$$

In calculating a particular coefficient $\overline{A}_\alpha$ according to Equation (9), one would like to have some criterion to establish whether sufficient averaging had occurred to ensure a relatively constant value of $\overline{A}_\alpha$. The central limit theorem of probability theory asserts that the probability distribution of the sample mean tends to become Gaussian as the number of statistically independent samples is increased without limit. For a finite sample size, the use of this theorem will allow rough confidence limits to be found for the coefficients $\{A_\alpha\}$. The Student t distribution provides $100(1 - a)\%$ confidence limits for the unknown mean A_α:

$$\overline{A}_\alpha \pm \frac{t_{M-1,\,a}\, s}{M^{1/2}}$$

Here s is the computed sample standard deviation based on M observations and $t_{M-1,\,a}$ is the value of the t distribution at the significance level a for $M - 1$ degrees of freedom.

Choice of Error Weighting Function

It will sometimes be advantageous to choose a time-dependent error weighting function $w(t)$ rather than one which is constant. This decision is influenced to a considerable extent by the a priori knowledge of the system under investigation. For example, if it were known that the system was time invariant, then assuming no additional a priori information, it would be reasonable for one to choose $w(t) = 1$. If, on the other hand, the system were known to be time varying, the choice $w(t) = 1$ would not be particularly advantageous, since data obtained in the distant past and that of more recent origin would contribute equally to the MSE.

This is not desirable for it would mean that current data, reflecting the changing system parameters, would be corrupted by data obtained from previous epochs. The deleterious effect on the model coefficients of choosing $w(t) = 1$ for a time varying system is easily discerned from Equation (6). For a time varying system, then, a nonnegative, monotonic nondecreasing weighting function $w(t)$ is in order.

As another example of incorporating a priori knowledge of the system into the model, as well as achieving a particular type of model approximation to the actual system output, consider the following situation. Suppose it is known that the system has two steady state points, y_* and y^*, such that when the input $x(t) = -1$, $y(t) \to y_*$ and when $x(t) = +1$, $y(t) \to y^*$ Since the steady state points are accurately known, it is desired to weight significantly the error expression immediately following a switch in the input but to attach little weight to the error as the steady state value is approached. Thus, the model output $\hat{y}(t)$ will closely approximate the actual system output $y(t)$ during the transient operation of the system at the expense of a more accurate approximation as the steady state is approached. This end could be achieved by choosing $w(t)$ to be a function of the magnitude of the difference $[y(t) - y^*_*]$ for $x = \pm 1$. A possible relationship is

$$w(t) = |y(t) - y^*_*|^n \quad \text{for} \quad x = \pm 1, \quad n > 0$$

Input Switching Constraint

Although only single-input, single-output systems have been considered thus far, the identification procedure can, in principle, be extended to multivariable systems; the single-input, multioutput system can be treated rather easily. In this case there are 2^n coefficients $\{A_\beta\}$ corresponding to the 2^n cells of the input space, associated with each output variable. For q outputs, then, there is a total of $q \times 2^n$ coefficients to be evaluated. This number does not represent a significant increase over the first case in the number of coefficients to be handled. The β^{th} coefficient for the i^{th} output variable is simply

$$A_{i,\beta} = \frac{\overline{y_i(t)w(t)\Phi(\beta)}}{\overline{w(t)\Phi(\beta)}} \tag{10}$$

where $y_i(t)$ is the i^{th} output of the system. Each coefficient $A_{i,\beta}$ depends only on the past history of the i^{th} output variable and the past of the input.

For multi-input systems the dimensions of the input space are so large that identification by this method is all but impossible. Since the state of each input variable exists independently of all other variables, the number of cells in the input space is 2^{np} for p inputs with n taps on each input memory. If there are q outputs, the number of coefficients to be evaluated is $q \times 2^{np}$, a number which can easily assume astronomical proportions even for small values of q and p, depending upon the value of n.

Since the number of coefficients to be evaluated increases as $q \times 2^{np}$, the restriction $p = 1$ (only one input) is necessary to reduce the complexity of the identification model to the realm of practicality. The evaluation of $q \times 2^n$ coefficients, where n is of the order of 20 or more (more than a million coefficients), is still not however a practical matter; and it becomes necessary to restrict the number of model coefficients further. It is assumed henceforth that the time between consecutive input switches, or the switching time T_{sw}, is not less than half the effective duration of the system's memory T_m, that is

$$T_{sw} \geq \frac{1}{2} T_m \tag{11}$$

Therefore, no more than two switches can occur during a time interval T_m, thereby drastically reducing the number of allowable input configurations. For this reason, the models proposed here are limited to systems whose inputs satisfy the above constraint.

The initial assumption that 2^n distinct input states exist implies that the input can switch at least as fast as T_m/n, since 2^n waveforms can be distinguished in the input memory. This assumption is unreasonable in many control system applications (when n is of the order of 20 or more). In most chemical process control systems, for example, such rapid switching is impractical due to the general sluggishness of these processes. It is expected, however, that a minimum switching time does exist, although admittedly it may not be as large as $\frac{1}{2} T_m$. The number of allowable input tap configurations, hence the number of coefficients, satisfying the minimum switching time constraint is easily determined for an input memory with uniformly spaced taps to be

$$N = 2n + \left[\frac{n}{2}\right]\left[\frac{n+2}{2}\right] \tag{12}$$

where the brackets denote truncation to the nearest integer less than or equal to the expression within the brackets. Table 1 presents the number of allowable input configurations for selected values of n.

TABLE 1. NUMBER OF ALLOWABLE INPUT CONFIGURATIONS FOR MEMORIES WITH UNIFORMLY SPACED TAPS

Taps	Configurations
10	50
16	104
21	152
31	302
41	502
51	752
61	1,052
71	1,402

Although only memories with uniformly spaced taps have been considered thus far, a tapered-tap arrangement (nonuniform distribution) can be expected to produce a smaller MSE term than will uniform spacing. This statement is valid for systems in which the influence of the past input on the present output attenuates with time. An input memory whose distribution of taps generally diminishes with distance down the memory (in the direction of the more distant past) will be more efficient than one with uniform-tap spacing. This is not to say that it is always desirable to have the closest tap spacing at the very beginning of the memory. Systems with appreciable time delays could best be identified by using memories with the closest tap spacing in the vicinity of the delay time.

The optimum distribution, in the sense of realizing the minimum MSE, of n taps along the memory is not readily determined. Consequently, there is considerable latitude for the investigator to apply what intuitive or a priori knowledge he has of the system under study. In fact, it is very difficult even to determine the number of allowable input configurations for n taps with a given nonuniform distribution; however, Equation (12) for uniform-tap spacing usually provides a good estimate.

Identification Time

An important aspect of any identification procedure is the time required for a complete identification of a given unknown system. Ideally, Equation (6) for the evaluation of coefficient A_β requires an infinite identification time. Since input records must necessarily be finite, the problem arises as to the selection of reliability criteria for the coefficients based on the assumption of finite input-output data. Kerr and Surber (*14*) refer to this as the *sufficient test signal problem.*

The identification of time-invariant systems is enhanced by the use of long operating records, not only as a result of prolonged coefficient averaging but also due to the effect of input-output noise smoothing. Such is not the case, however, for time varying systems. Although the orthogonal model can be used to describe systems with slowly varying characteristics, the coefficients themselves are only implicitly aware of changing system parameters. The coefficients are not explicit functions of time; they are suitable time averages of output conditions observed over a prior identification period. The identification model is thus time varying in the sense that it consists of a time sequence of time-invariant models, each model persisting until the next system identification is completed.

Since a sequence of time-invariant models is used to approximate a time varying system, it is desirable to base the system identifications on observation periods as short as possible, consistent with the decreased effect of noise smoothing. As Kerr and Surber (*14*) have observed, "for a given rate of parameter variation, a given noise level, and a given type of control signal variation, there should exist an optimum record length if an approximate time-invariant model is used to represent the system for a definite time into the future, that is, until the results of the next computation of parameters are available." For a time-invariant system there will be no optimum identification time since there is no conflict of requirements. However, there will exist a minimum identification time to yield steady state values of the coefficients, constant within acceptable tolerances.

IDENTIFICATION STUDIES OF NONLINEAR SYSTEMS

Preliminary Details and Analysis

In order to make a realistic appraisal of the value of the identification model, as well as to determine the

characteristics of its performance, several nonlinear systems have been investigated. The identification and synthesis procedures have been performed entirely on an IBM 7090/7094 digital computer. Rather than construct real nonlinear systems to be investigated, the process dynamics of these systems have been simulated on the digital computer. The computer simulated a given nonlinear system and, at the same time, identified it by evaluating the model coefficients $\{A_\alpha\}$. Having these coefficients, the computer next constructed a model of the system and proceeded to synthesize the model output $\hat{y}(t)$. The model output was then compared with the actual computed process output $y(t)$.

The system is to be identified by probing it with a switched two-level signal subject to the constraint $T_{sw} \geqq T_m/2$. It is often very desirable that the process identification be made under normal operating conditions, conditions which the investigator cannot control. In order to approximate normal operating conditions, each system investigated in this study is subjected to a randomly switched two-level signal with a uniform distribution of switching times in the interval $T_m/2 \leqq T_{sw} \leqq T_m$.

At other times the investigator is free to conduct independent experiments on the system, adjusting the control variable at will. Generally, the deterministic input signal will be selected so as to reduce the identification time to a minimum. This is accomplished by choosing a sequence of switching times which will, first, generate all allowable input configurations and, second, generate those input configurations whose associated coefficients have not yet met the stipulated criteria of measurement. There is no way to determine that unique sequence of switching times which will lead to the identification of the system in minimum time. However, any procedure whereby the switching times are sequentially chosen as described above should yield a near-minimum identification time.

The effect of varying the following model parameters on the accuracy of the model simulations was investigated: (1) number of taps along the input memory; (2) spacing of taps along the input memory; and (3) estimate T_m of the effective duration of the system's memory. In addition, some of the other features examined were: (4) degree to which the computed coefficients $\{A_\alpha\}$ approached constancy; (5) identification time of the unknown systems; and (6) over-all accuracy of the identification and synthesis procedures. Each of these points is discussed briefly in this section and in more detail later, when the results of the simulations are presented.

The identification process will certainly improve in accuracy as the number of taps along the input memory is increased, since this corresponds to increasing the sampling frequency of the input. Unfortunately, there are two adverse effects resulting from an increase in the number of taps: (1) the theoretical minimum identification time for a random binary input increases almost linearly with the number of taps (*12*), and (2) the number of allowable input configurations increases as a quadratic function of the number of taps. For a single-input, single-output system, the fast access digital storage requirement is approximately twice the number of allowable input configurations, so that some point is ultimately reached beyond which it becomes impractical to increase the number of taps.

For a fixed number of taps and a given unknown system, there is some optimum spacing of taps to give the least MSE in the model approximation to the actual system output. Since this optimum spacing cannot be determined analytically, a number of different tap spacings were investigated and then compared with the uniform tap distribution.

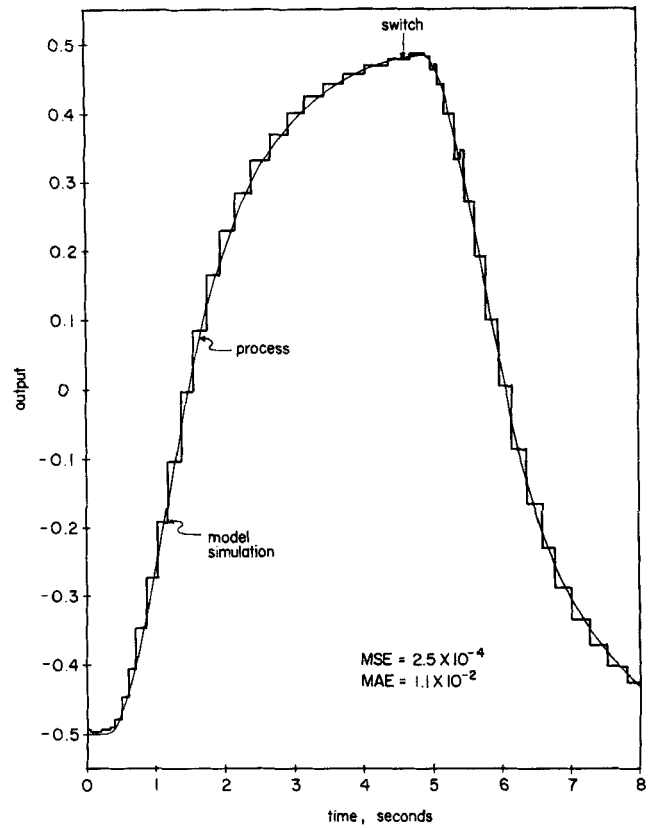

Fig. 1. Model simulation of system with saturation nonlinearity (thirty-one tapered taps, $T_m = 5.4$ sec., nonrandom input).

The estimate of the effective duration of the system's memory used by the model T_m will greatly affect the accuracy of the identification process. For a given number of taps there will be an optimum value of T_m. If T_m exceeds this optimum value, those taps beyond the effective length of the system's memory will be mostly ineffectual. The situation is more serious if T_m is considerably less than this optimum value. In this case the effect on the output of the neglected portion of the past input within the system's memory cannot be ascertained, and a large MSE will result. A good approximation to the optimum value of T_m is the settling time of the system T_s. This is defined to be the time required for the system response to a unit step function to come within 1% of the system's final value. The system will usually have two steady states, an upper state y^*, which the system will approach when the input is at its upper level, $x = +1$, and a lower steady state y_*, corresponding to an input $x = -1$. The settling time resulting from a step increase in the input and that for a step decrease will generally be different. The two settling times are given implicitly by

$$\left| \frac{y(T_s) - y_*^*}{y^* - y_*} \right| = 0.01 \qquad (13)$$

For purposes of identifying the system, the duration of the memory is usually taken as the larger of the two settling times.

A measure of the constancy of the model coefficient $\bar{A}_\alpha$ is furnished by its computed variance. As the number of samples increases, the variance can be expected to decrease, thereby reducing the confidence interval about A_α. When the confidence intervals for all the unknown means

$\{A_\alpha\}$ have been reduced to prescribed tolerances, the identification of the system is considered to be complete. In this investigation 95% confidence limits are employed for each coefficient $\{A_\alpha\}$.

The accuracy of the model simulation over an extended period of time subsequent to identification was estimated by evaluating both the MSE and the mean-absolute error (MAE) of the simulation. The MAE provides a second measure of simulation accuracy.

Identification of a Second-Order System with a Saturation Nonlinearity

The first nonlinear system to be investigated by the identification procedure was a second-order process with a saturation nonlinearity. This relatively simple system, also studied by Roy and DeRusso (*17*), consists of two first-order transfer stages separated by a saturating element. This system satisfies the following set of ordinary, nonlinear differential equations:

$$dy_1/dt + y_1 = x(t) \tag{14}$$

$$dy_2/dt + y_2 = 1/2 \text{ sat } (2y_1) \tag{15}$$

where $y_1(t)$ and $y_2(t)$ are the outputs from the first and last stages, respectively; $x(t)$ is the two-level input to the system; and

$$\text{sat } (z) = \begin{cases} -1 & \text{for } z \leq -1 \\ z & \text{for } -1 \leq z \leq 1 \\ +1 & \text{for } z \geq 1 \end{cases}$$

For the input $x(t)$ switching between the levels -1 and $+1$, the steady state operating values are $-1/2$ and $+1/2$, respectively. The settling time is 5.39 sec. regardless of the direction of the step change in the input.

Figure 1 shows the response of a model with thirty-one taps in a tapered arrangement on its input memory of length $T_m = T_s = 5.4$ sec. All simulations, of which only one is shown here, were performed over a period of at least fifty settling times. Figure 1 depicts the responses of the system and model, initially at their steady states, to a step increase in the input. The latter portion shows the response to a subsequent step decrease in the input. It is evident that the simulation improves as the steady state is approached. The model output, or set of coefficients $\{\overline{A}_\alpha\}$, is seen to be quantized into levels, each of which persists until the corresponding input configuration changes. The vertical lines are not part of the model output; they are drawn solely to aid in visualizing the simulation.

The effect of varying the number of taps on the accuracy of the model simulation is apparent from Figure 2, in which the MSE in the simulation over an extended period of time subsequent to identification is plotted against the number of taps. All three curves in the figure are for models with uniformly spaced taps. The nonrandom input was selected to identify the system in minimum time. It is evident that a further reduction in the MSE can be achieved by using more than fifty-one taps along the input memory.

For a memory with a given number of taps, it was always possible to find a tapered-tap distribution which was superior to the uniform arrangement. If the distribution of taps was too highly concentrated at the beginning of the memory, so that the taps were widely separated at the end, the performance of the model was inferior to that resulting from uniform spacing. In all cases studied those tapered-tap distributions leading to an increase in the number of allowable input configurations, consistent with the general rules of efficient tap spacing, produced an improved simulation over the uniform spacing case. This latter distribution can be considered a standard of comparison and is one that can always be improved upon.

Some tap distributions, however, leading to a decrease in the number of allowable input configurations produced an improvement in the simulation, although in no case was it superior to the improved simulation for that tap distribution which provided the greatest number of allowable configurations. It would seem that as a general rule those tap distributions, consistent with the investigator's intuition regarding efficient tap spacing, which lead to the greatest number of allowable input configurations, will provide the best simulations.

Several different estimates of the effective duration of the systems memory T_m were employed in order to determine their effect on the simulation. Figure 2 shows the results of using $T_m = T_s = 5.4$ sec. and $T_m = 7.2$ sec. The superiority of $T_m = 5.4$ sec. over $T_m = 7.2$ sec. for a given number of taps was anticipated from remarks made earlier. A 21-tap memory with $T_m = 2.7$ sec., or $1/2\ T_s$, produced a MSE of 0.3×10^{-2}, nearly seven times that for a 21-tap memory with $T_m = 5.4$ sec. and greater than the error for a 10-tap memory with $T_m = 7.2$ sec. T_s, it would seem, is a good approximation to the optimum value of T_m.

Two aspects of the experimental identification procedure, the constancy of the coefficients $\{A_\alpha\}$ and the identification time, merit discussion. Although the present concept of a system identification time, in the sense of achieving acceptable confidence limits about the $\{\overline{A}_\alpha\}$, is clear, a priori knowledge of what constitutes acceptable confidence limits for an unknown system is not clear. Once confidence limits for the $\{\overline{A}_\alpha\}$ have been evaluated, it is still not possible to determine in advance their effect on the accuracy of the model simulation for future inputs. The identification time is, of course, drastically influenced by the type of signal used, probabilistic or deterministic.

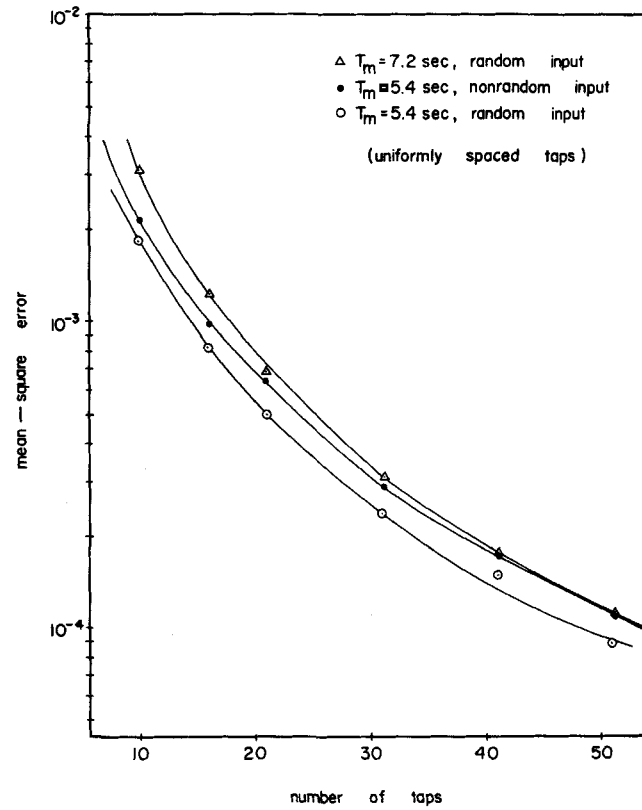

Fig. 2. Accuracy of model simulations of system with saturation nonlinearity (MSE).

TABLE 2. OCCURRENCE TIMES FOR ALLOWABLE INPUT CONFIGURATIONS AND IDENTIFICATION TIMES OF SECOND-ORDER SYSTEM WITH SATURATION NONLINEARITY

No. of taps (uniform spacing)	Random switching		Switching to achieve minimum ID time	
	Occurrence time*	ID Time* con. lim. = ±1.0	Occurrence time*	ID Time con. lim. = ±0.1
10	56.6	73.8	9.38	18.2
16	49.5	83.9	11.2	23.2
21	79.6	90.5	16.9	26.2
31	136.0	150.0	20.2	31.6
41	111.0	169.0	30.0	44.1
51	175.0	247.0†	31.8	59.3

* Times are normalized by T_m.
† Computation terminated at this time; system not identified.

For the randomly switched input, the prescribed confidence limits were initially set rather large, at ± 1.0, in order to determine the effect of coarse tolerances. Table 2 presents typical identification times, normalized by T_m, for memories with uniformly spaced taps. Because of the random nature of the input, it was found that the identification times varied by as much as 100% on different occasions. With these large confidence limits, identification was achieved as soon as all input configurations had occurred at least twice. Computed confidence limits about the model coefficients for a 16-tap memory did not exceed ± 0.5 nor ± 0.2 for a 41-tap memory. No noticeable improvement in the MSE resulted from a reduction of the prescribed confidence limits to ± 0.1.

Table 2 also displays the occurrence times for the allowable input configurations. This is the time required for all allowable input configurations, for any particular memory, to occur at least once. The occurrence time is only dependent upon the characteristics of the memory and the type of input used. It is in no way influenced by the unknown system. Each occurrence time in the table is, therefore, representative of any model with an equal number of uniformly spaced taps.

The minimum occurrence time and minimum identification time for various memories, using an input specifically selected to minimize these quantities, are also given in Table 2. The values in the third and fourth columns represent, in several cases, fourfold reductions in comparable quantities in the first and second columns. Once again, the minimum occurrence time is representative of any model with an equal number of uniformly spaced taps. The identification times are reported for models having coefficients with prescribed confidence limits of ± 0.1.

Mention should be made of the observed effect of the identification criteria on the accuracy of the model simulation. With various prescribed confidence limits on the model coefficients, limits from ± 0.01 to ± 1.0 to none at all, simulations were performed with several different memories. The MSE in the simulation was found to be little influenced by the criteria chosen. In fact, the accuracy of the simulation with no criteria (wherein the occurrence time is also the identification time) seldom differed by more than 10% from simulations employing relatively small confidence limits. This invariance of the model accuracy to the prescribed identification criteria became more pronounced as the number of taps increased. In the theoretical development it was noted that the average effect of using finite time averages in evaluating the model coefficient is, at worst, to double the MSE. It appears therefore, on the basis of the present second-order, nonlinear system that a practical criterion of identification is simply the occurrence of all allowable input configurations. The additional accuracy in the simulation gained by specifying confidence limits on the coefficients does not seem to justify the greatly increased identification time.

TABLE 3. VALUES OF INPUTS AND PHYSICAL PARAMETERS FOR TIME-INVARIANT CSTR SYSTEM

A_0	= 6.50×10^{-3} mole/liter	E	= 28,000 cal./mole
T_{0A}	= 300°K.	R	= 1.987 cal./(mole)(°K.)
T_{0B}	= 370°K.	$(-\Delta H)$	= 27,000 cal./mole
T_c	= 300°K.	$\rho = \rho_c$	= 1.0 g./cc.
F	= 10.0 cc./sec.	$c = c_c$	= 1.0 cal./(g.)(°K.)
F_c	= 5.0 cc./sec.	U	= 10.0 cal./(sec.)(°K.)
V	= 1,000 cc.	K	= 5.0 cc./sec.
k_0	= 7.86×10^{12} sec.$^{-1}$		

TABLE 4. STEADY STATE POINTS FOR TIME INVARIANT CSTR SYSTEM

Input Level $x = +1$			Input Level $x = -1$		
Steady state point	T^s, °K.	A^s, mole/liter	Steady state point	T_s, °K.	A_s, mole/liter
A	460.91	1.5310×10^{-4}			
B	414.94	2.7072×10^{-3}			
C	346.88	6.4884×10^{-3}	C'	300.00	6.5000×10^{-3}

Identification of Nonlinear CSTR System

Another nonlinear system to be investigated by the identification procedure was the well-known model of a continuous stirred tank reactor (CSTR) in which a single exothermic reaction occurs. This particular system has been the object of considerable attention, and recently several studies have been reported of efforts to determine its dynamic characteristics from process operating records (*1, 4, 10, 13*). In these investigations there has been no need to restrict the control variable to operate at two levels as in this study. However, the previous techniques are applicable only to linear systems, giving erroneous results for nonlinear processes. Since the process model

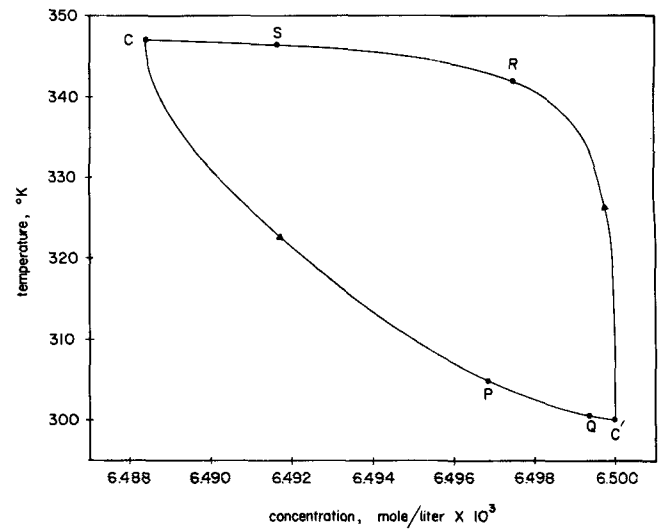

Fig. 3. Steady state trajectories for CSTR system.

has been described in detail elsewhere (2), only a brief description will be presented here.

An irreversible, exothermic chemical reaction $A \rightarrow B$ is assumed to take place in a perfectly stirred-tank reactor, such that the concentration and temperature of the components within the reactor are uniform and the same as those in the effluent. Cooling water flowing through a jacket surrounding the reactor provides for the removal of heat. Taking the reaction as first order, one can write the following dynamic material and energy balances for the reactor:

$$\frac{dA}{dt} = \frac{F}{V}(A_0 - A) - kA \tag{16}$$

$$\frac{dT}{dt} = \frac{F}{V}(T_0 - T) + \frac{(-\Delta H)}{\rho c} kA - \frac{UF_c}{V\rho c(K + F_c)}(T - T_c) \tag{17}$$

where the reaction rate constant k is given by the Arrhenius expression

$$k = k_0 \exp(-E/RT) \tag{18}$$

Equations (16) and (17) taken together form a set of simultaneous, ordinary nonlinear differential equations which describe the dynamic behavior of the continuous stirred-tank reactor. Clearly, the assumption of a first-order reaction does not detract from the inherent difficulty in obtaining an analytical solution to this problem. Equations (16) and (17) are easily and rapidly integrated on the IBM 7090/7094 digital computer.

In this reactor system it is desired to control the dependent variables $A(t)$ and $T(t)$ according to some control scheme, as yet unspecified, through the on-off manipulation of a suitable input or control variable (manipulated variable). The application and effect of two-level, or on-off, control action in regard to the CSTR has been discussed to some extent in the literature (3, 6, 11).

There are several suitable control variables which might be used singly or in combination: (1) the inlet temperature T_0; (2) the inlet concentration A_0; (3) the inlet coolant temperature T_c; and (4) the coolant flow rate F_c. First, the inlet temperature T_0 was selected as the control variable and the system identified. Later, with the coolant flow rate F_c as the control variable, the system was again identified. It is anticipated that there will be inherent differences in the results of these two identifications, for the first CSTR system, with T_0 as input, is time invariant whereas the second, with F_c as input, is time varying. Only the results of the former system simulations are discussed here even though the time varying system simulations turned out to be excellent. The inlet temperature T_0 is assumed to switch between two levels T_{0A} and T_{0B} where $T_{0A} < T_{0B}$. The switching variable $x(t)$ is introduced

$$T_0(t) = \begin{cases} T_{0A} & \text{if } x(t) = -1 \\ T_{0B} & \text{if } x(t) = +1 \end{cases} \tag{19}$$

The values of the physical parameters and constant input conditions used in this system appear in Table 3.

The steady state, or equilibrium, points for this system are given in Table 4. There are two sets of steady state points, one each for the inputs $x = -1$ and $x = +1$, denoted (A_s, T_s) and (A^*, T^*), respectively. The steady state points labeled A, C, and C' can be shown to be locally asymptotically stable, while point B is locally unstable (20). Although there exist three steady state points for the system when $x = +1$ and one when $x = -1$, only points C and C' are of interest. For the on-off control action specified previously, the system will have all its trajectories in the phase plane of A and T and approach one of the two steady state points C or C', provided the process is initially located at or near either of these points. Figure 3 shows two such trajectories whose origins are C and C'. Points R and P represent the state of the sys-

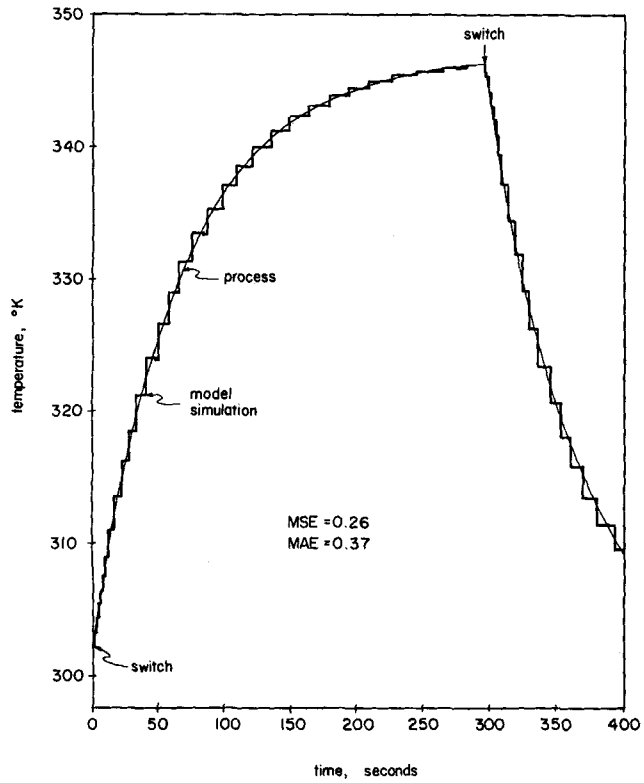

Fig. 4. Temperature simulation of CSTR system (thirty-one tapered taps, T_m = 300 sec., random input).

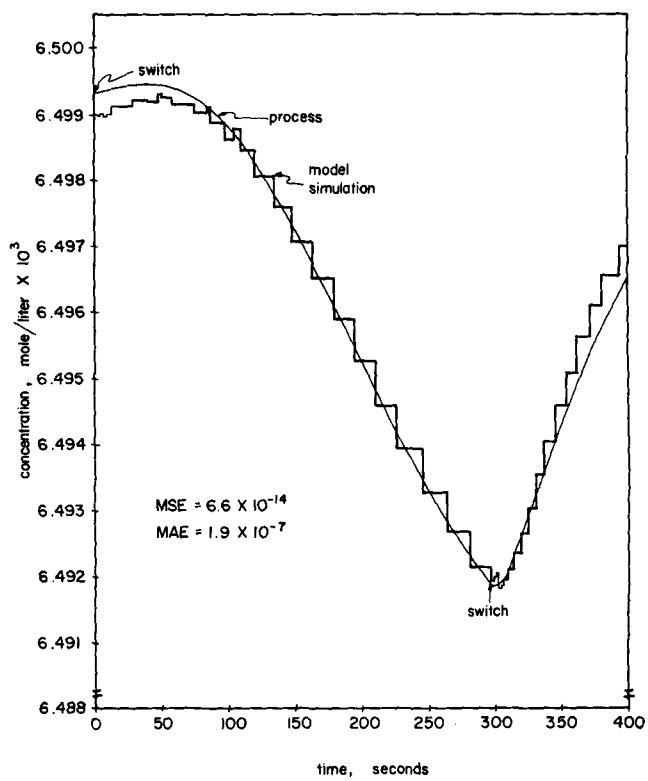

Fig. 5. Concentration simulation of CSTR system (thirty-one tapered taps, T_m = 300 sec., random input).

tem after an elapsed time $T_s/2$; points S and Q show its location after time T_s.

The settling time of the system will depend upon which variable, A or T, is selected to measure the system's proximity to a steady state. Since the temperature undergoes a much greater percentage variation than does the concentration (for this reason the former is of greater interest), the temperature was chosen to measure the system's approach to the steady state. By numerical solution of the system differential equations, the two settling times, one for each steady state, were found to differ by only 2%, the average value being 310 sec. The estimate of the system's memory used by the model T_m was taken as 300 sec.

The nature of the system's trajectories resulting from the specified two-level input is of interest. The switching time of the input, it will be recalled, is constrained to be not less than $T_s/2$. The system, originating at C' with $x = +1$, will reverse trajectory $C'RSC$ until the input switches to $x = -1$; this switch cannot occur before the system reaches point R. When the input does switch to $x = -1$, the system will leave this trajectory and proceed toward steady state C' via some path resembling $CPQC'$. Similarly, the system, originating at C with $x = -1$, will follow trajectory $CPQC'$ until the input switches to $x = +1$. Following the switch, the system, located on arc PQC', will leave this trajectory and proceed toward steady state C via a path similar to $C'RSC$. Regardless of the precise nature of subsequent control switching, subject, of course, to the above constraint, all of the system's trajectories will lie within the area enclosed by the two trajectories shown in Figure 3.

Figures 4 and 5 show a typical model simulation of the CSTR system. It should be stressed that the two outputs of the model, temperature and concentration, are independent of one another and dependent only upon the past history of the input as stored in the model's memory. The particular model employed in this simulation had thirty-one taps in a tapered arrangement along its input memory of length $T_m = 300$ sec. The placement of taps along the memory can be ascertained from the pattern of steps which forms the model output. Although it is not perfectly clear, the model output steps fourteen times in the first 50 sec. following an input switch. This simulation and all others were conducted over a period of at least 100 T_s At time $t = 0$ the system, near its steady state, is subjected to a step change in the input from $x = -1$ to $x = +1$, and at $t = 295.6$ sec. the input reverts to $x = -1$. The temperature simulation appears to be superior to the concentration simulation. The MAE in the former is 0.79% of $(T^* - T_s)$,the maximum range over which the temperature can vary, whereas the MAE in the latter is 1.7% of $(A_s - A^*)$, the maximum possible concentration variation.

The effect of varying the number of taps on the accuracy of the model simulation can be seen from Figures 6 and 7. The MSE in the simulation over an extended period of time subsequent to identification is the relevant quantity plotted. There is little tendency for the curves in Figure 6 to become horizontal, and it is expected that a significant reduction in the MSE in the temperature simulation could be achieved by increasing the number of taps beyond 51, the maximum number used. In Figure 7, however, there are definite breaks in the curves. The MSE in the concentration simulation is less than halved in going from twenty-one to fifty-one taps. This behavior, as well as the generally mediocre accuracy achieved in the

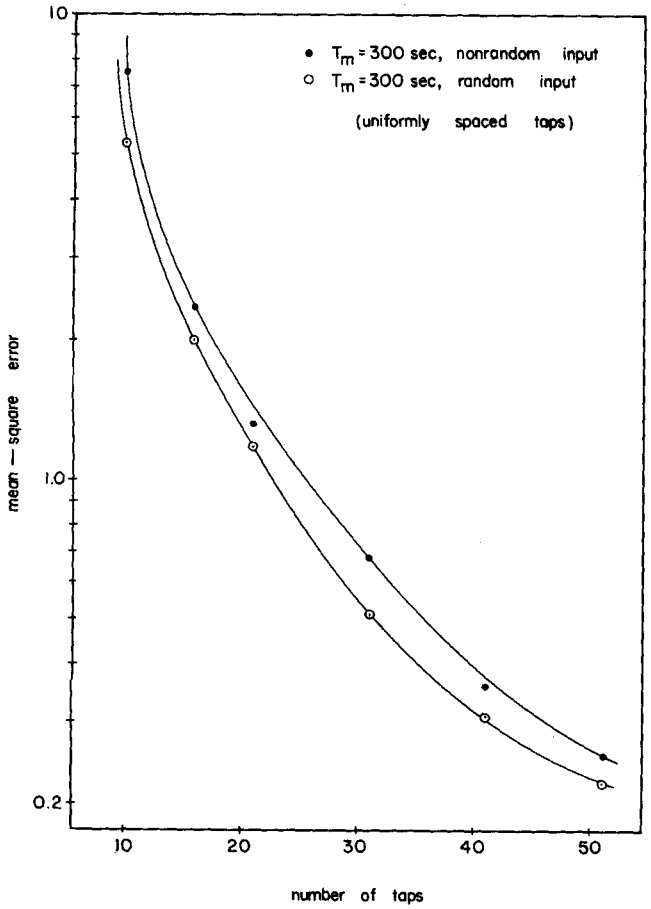

Fig. 6. Accuracy of temperature simulations of CSTR system (MSE).

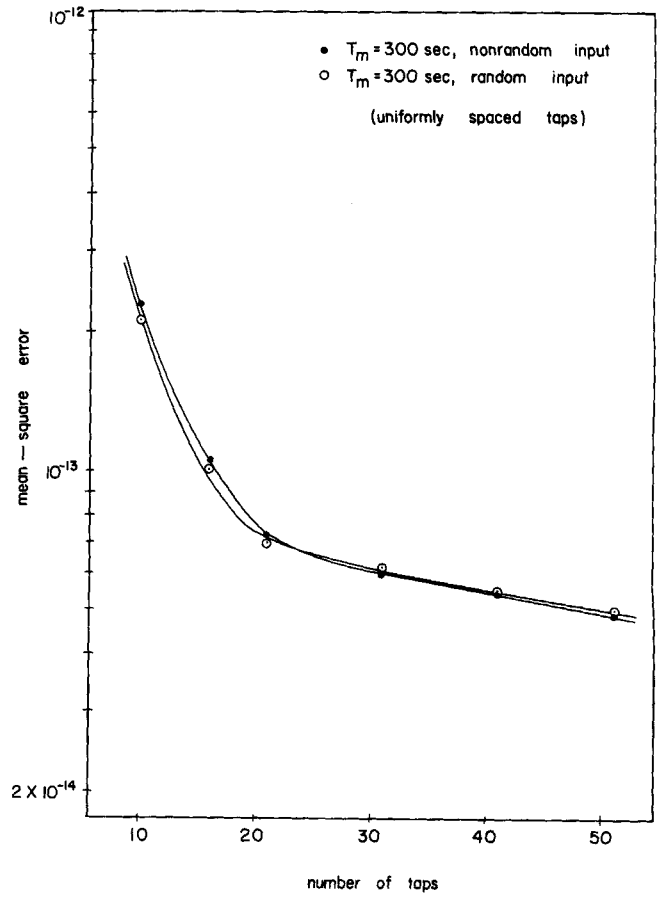

Fig. 7. Accuracy of concentration simulations of CSTR system (MSE).

concentration simulations, is probably a direct result of T_m being too small. Based on concentration measurements rather than temperature measurements, the settling times of the system are 480 sec. when $x = -1$ and 680 sec. when $x = +1$.

Several different tap distributions along the input memory were tried. The results obtained were basically the same as those for the first nonlinear system investigated. Note, however, that the concentration response, Figure 5, has an apparent delay time of 50 sec. when the input switches to $x = +1$; the tapered-tap arrangement depicted is not efficient. The temperature, Figure 4, immediately responds following a step change in the input; the tap arrangement in this case is seen to be highly efficient.

Only one value of T_m, 300 sec., was employed in the identification models of the CSTR system. The difference in the quality of temperature versus concentration simulations has been partially ascribed to the use of an inadequate value of T_m. If T_m is taken as 680 sec., the settling time of the system based on concentration measurements, both the temperature and concentration simulations will not necessarily improve. It was stated earlier that for a single-output system and a given number of taps there is an optimum value of T_m. If this value is exceeded, the taps will be spaced far apart; hence, a poor characterization of the input results. If T_m is less than its optimum value, the effect of the past input on the system's output cannot be adequately incorporated into the model. For the CSTR system there are two outputs, and correspondingly there is an optimum value of T_m for each output. Thus, while $T_m = 300$ sec. leads to an accurate simulation of the system's temperature response, it leads to a mediocre concentration simulation. There appears to be no straightforward method to determine an optimum T_m for an unknown system.

An important aspect of the model simulation has just been revealed: the model parameters (number of taps, tap spacing, T_m) can be chosen to yield a desired end (usually small MSE) with respect to a particular output without regard for the other model outputs. In the present case, for example, the tap spacing can be chosen to minimize the MSE in the temperature response at the expense of an accurate concentration simulation.

The accuracy of the simulations was not significantly improved by the use of small confidence limits on the model coefficients. The occurrence times for a randomly switched input, which are independent of the system being identified, were comparable to those in Table 2 as were the minimum occurrence times. The minimum identification time of models with prescribed confidence limits of $\pm$ 1.0 varied from 61 T_s for a 10-tap model to 72 T_s for a 16-tap model. In no instance could identification of the system be achieved with a randomly switched input; a maximum identification time of 100 T_s was permitted. To reiterate what was stated previously, one observes that the occurrence of all allowable input configurations seems to be the criterion of identification; the criterion is not the satisfaction of imposed confidence limits.

CONCLUSIONS AND EXTENSIONS

The MSE in the model simulations decreased rapidly as the number of input samples in the memory, or taps, increased. The greatest number of taps employed was fifty-one; however, a further reduction in the MSE is possible if more than fifty-one taps are used. Efficient spacing of taps along the input memory can at least halve the MSE for uniformly spaced taps. Several rules for efficient tap spacing have been given.

The settling time, as defined in this study, is perhaps an overly conservative estimate of the effective duration of the system's memory. It might be advantageous to redefine the settling time to be the time required for the system response to a unit step function to come within 5% of the system's final value. There are two settling times, which are in general different, for each output of the system. If the largest of these settling times is taken as the estimate of the length of the system's memory, the model simulation for a given number of taps may not be superior to the case in which the smallest settling time is used. There exists for each system an optimum estimate of the duration of its memory, but this optimum value, like the optimum tap spacing, is not readily determined.

A satisfactory criterion of system identification is merely the occurrence of all allowable input configurations. An alternative criterion, the satisfaction of prescribed confidence limits about the model coefficients, was rejected. The minimum identification time, which is independent of the system investigated and dependent only on the number of taps and the tap spacing, was evaluated for many models. In no case did the minimum identification (occurrence) time exceed one-third of the (occurrence) time required to identify systems with randomly switched two-level inputs, and it frequently was less than one-fifth of the latter time interval.

In addition, for the different systems investigated, the computed MAE in the model simulations decreased from about 3.1% of the total output variation for 10-tap models to about 0.7% for 51-tap models. The results of simulations in which prior identification was achieved through the use of (1) randomly switched input and (2) an input chosen to yield a maximum identification time were in close agreement. The former input usually gave a more accurate simulation than the latter because of the disparity in identification times.

There are three aspects of the identification process which are especially attractive. First, model parameters can be adjusted to give a small error in the simulation of a specified system output at the expense of correspondingly larger errors in the other model outputs. Second, errors in the model performance at discrete times are not propagated; the model output at any time depends solely upon the state of the input and not at all upon its previous value. Third, the model is expected to be relatively insensitive to noise in the measured system outputs since the characterizing coefficients are the result of an averaging process.

Further avenues of investigation incorporating the identification model have been previously mentioned. The use of a weighted-MSE criterion between the actual output of the system and the output of the model will, in certain instances, lead to a greatly improved simulation. The model can be used to track systems with slowly varying process dynamics provided the system can be considered to be time invariant over a period roughly equal to the minimum identification time. Multiple-input systems can, in theory, be identified using the model; however, the number of coefficients to be evaluated will be legion.

Application of the present model is limited to systems whose inputs do not switch rapidly ($T_{sw} \geq T_m/2$). Unquestionably, the first extension should be removal of this constraint on the switching time. A reasonable step in this direction is to restrict the two-level input to at most two switches during a period equal to the duration of the system's memory. Thus, the assumption of a minimum switching time is removed. Regardless of the tap spacing, the number of input configurations N is given by

$$N = n^2 - n + 2$$

where n is the number of taps. For large n there is nearly a fourfold increase in the number of input configurations, or cells, over that resulting from acceptance of the constraint on the switching time. A single-input, single-output model with fifty-one taps, for example, would have 2,552 characterizing coefficients and require at least twice as many words of fast-access digital storage. This presents no storage problem for most digital computers; however, if more than two switches are permitted in the model's memory, an acute shortage soon develops. In general, if s switches are allowed in the input memory, then the number of coefficients is

$$N = 2 \sum_{r=0}^{s} \binom{n-1}{r}$$

where the parentheses denote the binomial coefficients. A 51-tap model with at most three switches in its memory would have over 40,000 coefficients and require more than 80,000 words of fast-access storage. The identification time for such a model would be prohibitive.

ACKNOWLEDGMENT

The authors wish to thank the International Business Machines Corporation, Food Machinery Corporation, and Levers Brothers Company for financial support in the form of research grants and fellowships. This work was supported in part by National Science Foundation Grant NSF-GK-460 and made use of computer facilities supported in part by National Science Foundation Grant NSF-GP-579.

NOTATION

A = concentration of A in reactor and effluent
A_0 = concentration of A in reactor influent
A_α = α^{th} orthogonal expansion coefficient of model output
$\overline{A_\alpha}$ = sample mean of $\{A_{\alpha,i}\}$
$A_{\alpha,i}$ = model coefficient for i^{th} occurrence of α^{th} input configuration
$A_{i,\beta}$ = β^{th} model coefficient for the i^{th} output variable
c = heat capacity of reactor contents
c_c = heat capacity of coolant
E = activation energy of reaction
$\overline{e^2}$ = weighted mean-square error
$\overline{e^2_{\min}}$ = minimum weighted mean-square error
F = reactor influent and effluent flow rate
F_c = coolant flow rate
$(-\Delta H)$ = heat of reaction (>0 if exothermic)
K = $U/2\,\rho_c c_c$
k = reaction rate constant
k_0 = frequency factor in reaction rate expression
M = number of distinct occurrences of a particular input configuration
N = total number of allowable input configurations
n = number of taps along input memory
p = number of system inputs
q = number of system outputs
R = gas constant
s = sample standard deviation
T = temperature within reactor and of effluent; sampling interval; time averaging variable
T_c = inlet coolant temperature
T_m = estimate of effective duration of system's memory
T_s = system settling time
T_0 = temperature of reactor influent
T_{0A}, T_{0B} = switching levels of reactor influent temperature
t = time
T_{sw} = switching time interval
$t_{M-1,a}$ = value of t distribution at significance level a for $M-1$ degrees of freedom
U = product of heat transfer coefficient and heat transfer area
V = reactor volume
$w(t)$ = nonnegative weighting function
$x(t)$ = two-level input
$y(t)$ = system output
$\hat{y}(t)$ = model output
$y_i(t)$ = i^{th} system output
y_1, y_2 = outputs from first and last stages of system with saturation nonlinearity

Greek Letters

α, β = model expansion indices
ρ = density of reactor contents
ρ_c = density of coolant
$\Phi(\alpha)$ = orthogonal function in model output expansion of Equation (3)

Superscripts

* = upper steady state corresponding to $x = +1$
$\overline{\quad}$ = time average

Subscript

$_*$ = lower steady state corresponding to $x = -1$

LITERATURE CITED

1. Angus, R. M., and Leon Lapidus, *A.I.Ch.E. J.*, **9**, 810 (1963).
2. Aris, Rutherford, and N. R. Amundson, *Chem. Eng. Sci.*, **7**, 121 (1958).
3. *Ibid.*, 148.
4. *Ibid.*, **9**, 250 (1958).
5. Aris, Rutherford, "The Optimal Design of Chemical Reactors," Academic Press, New York (1961).
6. Berger, J. S., and D. D. Perlmutter, *A.I.Ch.E. J.*, **10**, 238 (1964).
7. Blakemore, N., and Rutherford Aris, *Chem. Eng. Sci.*, **17**, 591 (1962).
8. Bose, A. G., *Tech. Rept. 309*, Res. Lab. Electronics, Massachusetts Inst. Technol., Cambridge (1956).
9. Eckman, D. P., "Automatic Process Control," Wiley, New York (1958).
10. Goodman, N. R., S. Katz, B. H. Kramer, and M. T. Kuo, *Technometrics*, **3**, 245 (1961).
11. Grethlein, H. E., and Leon Lapidus, *A.I.Ch.E. J.*, **9**, 230 (1963).
12. Harris, G. H., Ph.D. dissertation, Princeton Univ., N. J. (1965).
13. Homan, C. J., and J. W. Tierney, *Chem. Eng. Sci.*, **12**, 153 (1960).
14. Kerr, R. B., and W. H. Surber, Jr., *IRE Trans. Automatic Control, AC-6*, 173 (1961).
15. LaSalle, J. P., in "Contributions to the Theory of Nonlinear Oscillations," S. Lefschetz, ed., Vol. 5, pp. 1-24, Princeton Univ., N. J. (1960).
16. Mishkin, Eli, and Ludwig Braun, ed., "Adaptive Control Systems," McGraw-Hill, New York (1961).
17. Roy, R. J., and P. M. DeRusso, *IRE Trans. Automatic Control, AC-7*, 93 (1962).
18. Roy, R. J., R. W. Miller, and P. M. DeRusso, *IEEE Trans. Appl. Ind.*, **83**, 173 (1964).
19. Singleton, H. E., *Tech. Rept. 160*, Res. Lab. Electronics, Massachusetts Inst. Technol., Cambridge (1960).
20. Warden, R. B., Rutherford Aris, and N. R. Amundson, *Chem. Eng. Sci.*, **19**, 149 (1964).
21. Weiner, Norbert, "Nonlinear Problems in Random Theory," Technology Press, Massachusetts Inst. Technol., Cambridge (1958).

Manuscript received May 18, 1966; revision received August 22, 1966; paper accepted August 22, 1966.