General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

S&T

Semi-Annual Report

Real Time Flight Simulation Methodology

Submitted to:

National Aeronautics and Space Administration
Langley Research Center
Hampton, Virginia

Submitted by:

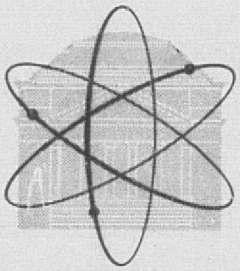
E. A. Parrish
Associate Professor

G. Cook
Professor

E. S. McVey
Professor

SCHOOL OF ENGINEERING AND APPLIED SCIENCE

RESEARCH LABORATORIES FOR THE ENGINEERING SCIENCES



UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VIRGINIA 22901

Report No. UVA/528085/EE76/103

August 1976

(NASA-CR-148728) REAL TIME FLIGHT SIMULATION METHODOLOGY Semiannual Report (Virginia Univ.) 65 p HC $4.50 CSCL 01A

N76-30202

Unclas
G3/05 50389

Semi-Annual Report

Real Time Flight Simulation Methodology

Submitted to:

National Aeronautics and Space Administration
Langley Research Center
Hampton, Virginia

Submitted by:

E. A. Parrish
Associate Professor

G. Cook
Professor

E. S. McVey
Professor

Department of Electrical Engineering
RESEARCH LABORATORIES FOR THE ENGINEERING SCIENCES
SCHOOL OF ENGINEERING AND APPLIED SCIENCE
UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VIRGINIA

Report No. UVA/528085/EE76/103

August 1976

Copy No. 1

Foreword

This report presents the current status and results achieved during the first six months of effort on Real Time Flight Simulation Methodology. The work is a continuation of research performed during the first year on substitutional methods for digitization, input signal dependent integrator approximations and digital autopilot design. Topics reported on herein include Digital Autopilot Design, Digital Simulation Methods for Linear Systems, Interactive Simulator Design Package for The Design of Real Time Simulators and Study of Charge Coupled Devices for Simulation.

1. DIGITAL AUTOPILOT DESIGN

1.0 Introduction

In the final technical report [1] for Grant Number NASA NSG 1151, page 36, it was concluded:

> "The performance of a continuous autopilot which is implemented digitally is affected so much by the zero order hold that the discretization method is a secondary consideration. Tustin is a relatively simple method that is satisfactory. Computer speed (sampling rate) may be established on the basis of the phase shift a designer will allow to be introduced by the zero order hold. Phase lag decreases as sample speed increases. Gain constants should be maintained to keep steady state errors constant. The relative stability of the aircraft will be decreased by the digital implementation so the value of computer sampling time should be based on the decrease in phase margin a designer is willing to allow. This value can be determined through sensitivity studies."

An example sensitivity study is presented here to demonstrate how a digital autopilot designer could make a decision on minimum sampling rate for computer specification. It consists of comparing the simulated step response of an existing analog autopilot and its associated aircraft dynamics to the digital version operating at various sampling frequencies and specifying a sampling frequency that results in an acceptable change in relative stability. In general, the zero order hold introduces phase lag which will increase overshoot and settling time. It should be noted that this solution is for substituting a digital autopilot for a continuous autopilot. A complete redesign could result in results which more closely resemble the continuous results or which conform better to original design goals.

1.1 Sensitivity Study

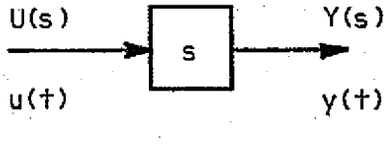
The autopilot is an open-loop system, insofar as the complete aircraft system is concerned; but comparison between the analog and the digital autopilot should be based on their performance in the overall closed-loop system which includes the aircraft dynamics as noted above.

The simplified model of an autopilot system [2] for pitch attitude control shown in Fig. 1.1 was chosen for the study, because it is representative of the kind of problem a designer would typically confront.† The transfer function of the short-period approximation is used to represent the aircraft dynamics. An equivalent block diagram is shown in Fig. 1.2 in which the autopilot portion is separated from the aircraft transfer formation. Figure 1.3 depicts the digital autopilot system which is to replace the analog autopilot of Fig. 1.2. The constants K(z) and H(z) of Fig. 1.3 are to be determined by applying a discretization method to the continuous transfer functions of the analog autopilot of Fig. 1.2. It is noted that, in this example, the continuous transfer functions consist of gain constants and a differentiator. By applying Euler's method [3] to the differentiator*, the following discrete transfer functions for K(z) and H(z) are obtained where S(amp) = 5.6 and S(rg) = 1.19 and K(z) is set equal to S(amp).

---

†The transfer function representing the aircraft dynamics is for a four-engine jet transport flying in straight and level flight at 40,000 feet with a velocity of 600 ft/sec (355 knots) with the compressibility effects neglected.

*Discrete differentiation operator with Euler's method



$y(t) = u'(t)$

Applying Euler's method, yields

$u[(n+1)T] = u[nT] + Ty[nT]$

In z- domain $zU(z) = U(z) + TY(z)$

$$\frac{Y(z)}{U(z)} = z_{dis}[s] = \frac{z-1}{T}$$

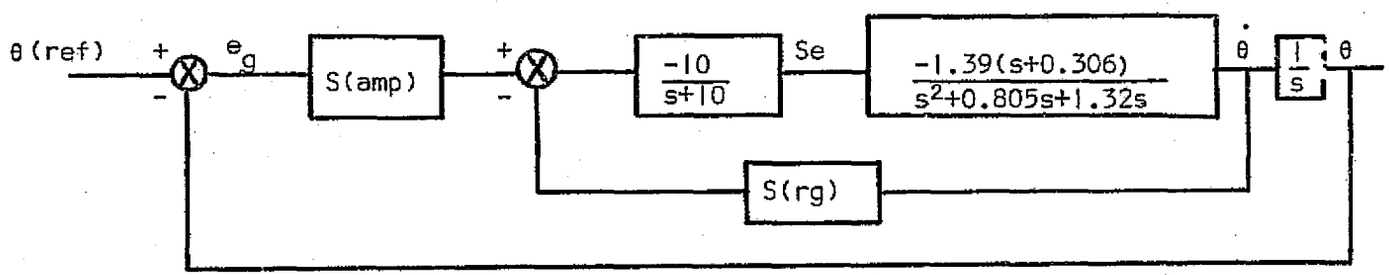


Figure 1.1 Block Diagram for Jet Transport and Autopilot for Pitch Attitude Control

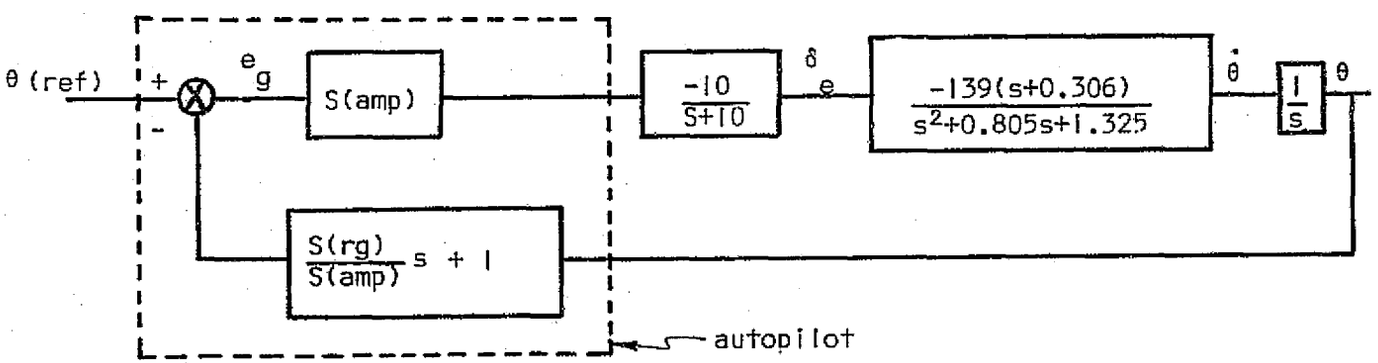


Figure 1.2 System of Figure 1.1 in Reduced Form

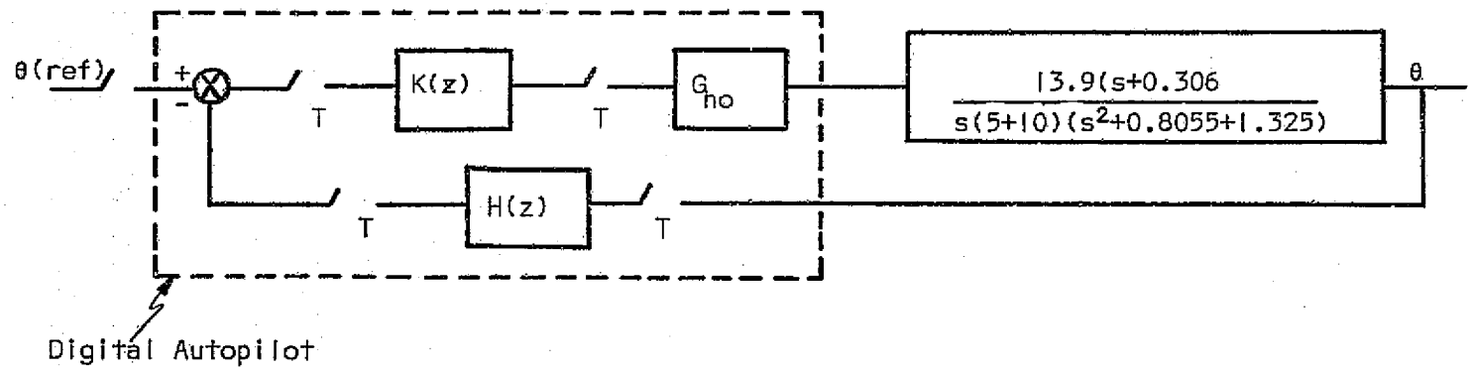


Figure 1.3 Digital Replacement for System of Figure 1.2.

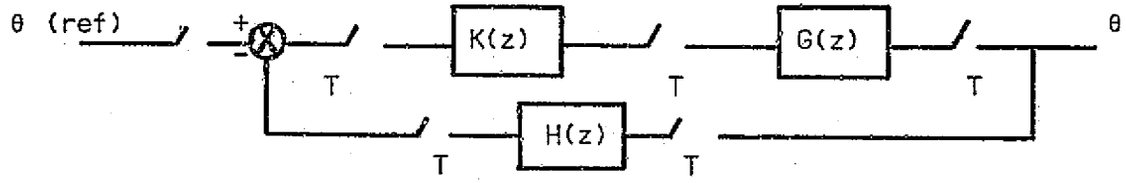


Figure 1.4 System of Figure 1.3 in Reduced Form.

$$H(z) = Z_{dis}\left[\frac{S(rg)}{S(amp)}\, s + 1\right] = Z_{dis}\left[\frac{1.19}{5.6}\, s + 1\right]$$

$$= 0.212 + \frac{z - 1}{T} + 1 \qquad (1.1)$$

The zero-order-hold and the aircraft dynamics are combined to be represented as in Fig. 1.4 where

$$G(z) = Z[G_{ho}(s)\, D(s)]$$

$$= Z\left[\frac{1 - e^{-TS}}{s} \cdot \frac{13.9(s + 0.306)}{s(s + 10)(s^2 + 0.805s + 1.325)}\right]$$

$$= 13.9(1 - z^{-1})Z\left[\frac{s + 0.306}{s^2(s + 10)(s^2 + 0.805s + 1.325)}\right]$$

$$= 13.9(1-z^{-1})Z\left[\frac{0.02309434}{s^2} + \frac{0.059131363}{s} - \frac{0.0010393}{s + 10} + \frac{(-0.058154(s + 0.4025)}{(s + 0.4025)^2+1.078422^2}\right.$$

$$\left. + \frac{(-0.052705)(1.078422)}{(s + 0.4025)^2 + 1.78422^2}\right]$$

$$= G_1(z) + G_2(z) + G_3(z) + G_4(z) + G_5(z) \qquad (1.2)$$

and

$$G_1(z) = 13.9(1 - z^{-1})Z\left[\frac{0.02309434}{s^2}\right] = \frac{0.321011326T}{z - 1}$$

$$G_2(z) = 13.9(1 - z^{-1})Z\left[\frac{0.059131363}{s}\right] = 0.821925946$$

$$G_3(z) = 13.9(1 - z^{-1})Z\left[\frac{-0.0010393}{s + 10}\right] = \frac{-0.014442(z - 1)}{z - e^{-10T}}$$

$$G_4(z) = 13.9(1 - z^{-1})Z\left[\frac{-0.058154(s + 0.4025}{(s + 0.4025)^2 + 1.078422^2}\right]$$

$$= \frac{(0.808341)(z - 1)(z - e^{-0.4025T}\cos 1.078422T)}{z^2 - 2ze^{-0.4025T}\cos 1.078422T + e^{-0.805T}}$$

$$G_5(z) = 13.9(1 - z^{-1})Z\left[\frac{(-0.052705)\cdot(1.078422)}{(s + 0.4025)^2 + 1.078422^2}\right]$$

$$= \frac{(-0.7326)(z - 1)\ e^{-0.425T}\sin(1.078422T)}{z^2 - 2ze^{-0.4025T}\cos 1.078422T + e^{-0.805T}} \qquad (1.3)$$

In order to minimize round-off and truncation errors in the calculations for the system response, G(z) is put in parallel form as shown in Fig. 1.5. By transforming the transfer functions of Fig. 1.5 into time-domain recursive relationships, the output time response can be obtained. Preliminary calculations using Hewlett-Packard time-sharing Basic had 32-bit floating-point arithmetic, and the output responses obtained at different sampling frequencies are shown in Fig. 1.6. It indicates that the word length of the Hewlett-Packard minicomputer is not large enough to make the truncation and the round-off errors in the calculation negligibly small. The indication appears particularly evident in Fig. 1.6 in the case of T = 1 msec (or f = 1000 cps which is the fastest sampling frequency used).

In order to reduce calculation errors a CDC 6400 was used which has a word length of 60 bits, thus having 120-bit floating-point arithmetic in Basic. The results in Fig. 1.7 and in Table 1.1 show that the percent overshoot increases uniformly as the sampling frequency decreases.

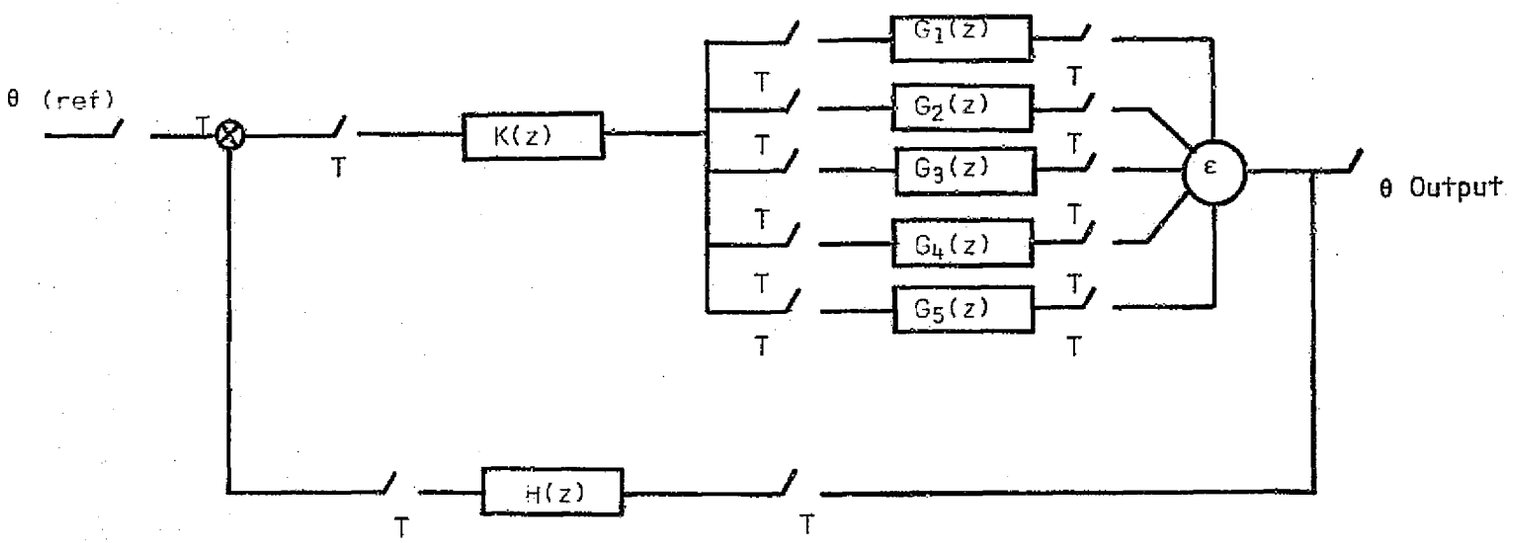


Figure 1.5 Digital Autopilot System of Figure 1.4 in Equivalent Parallel Form

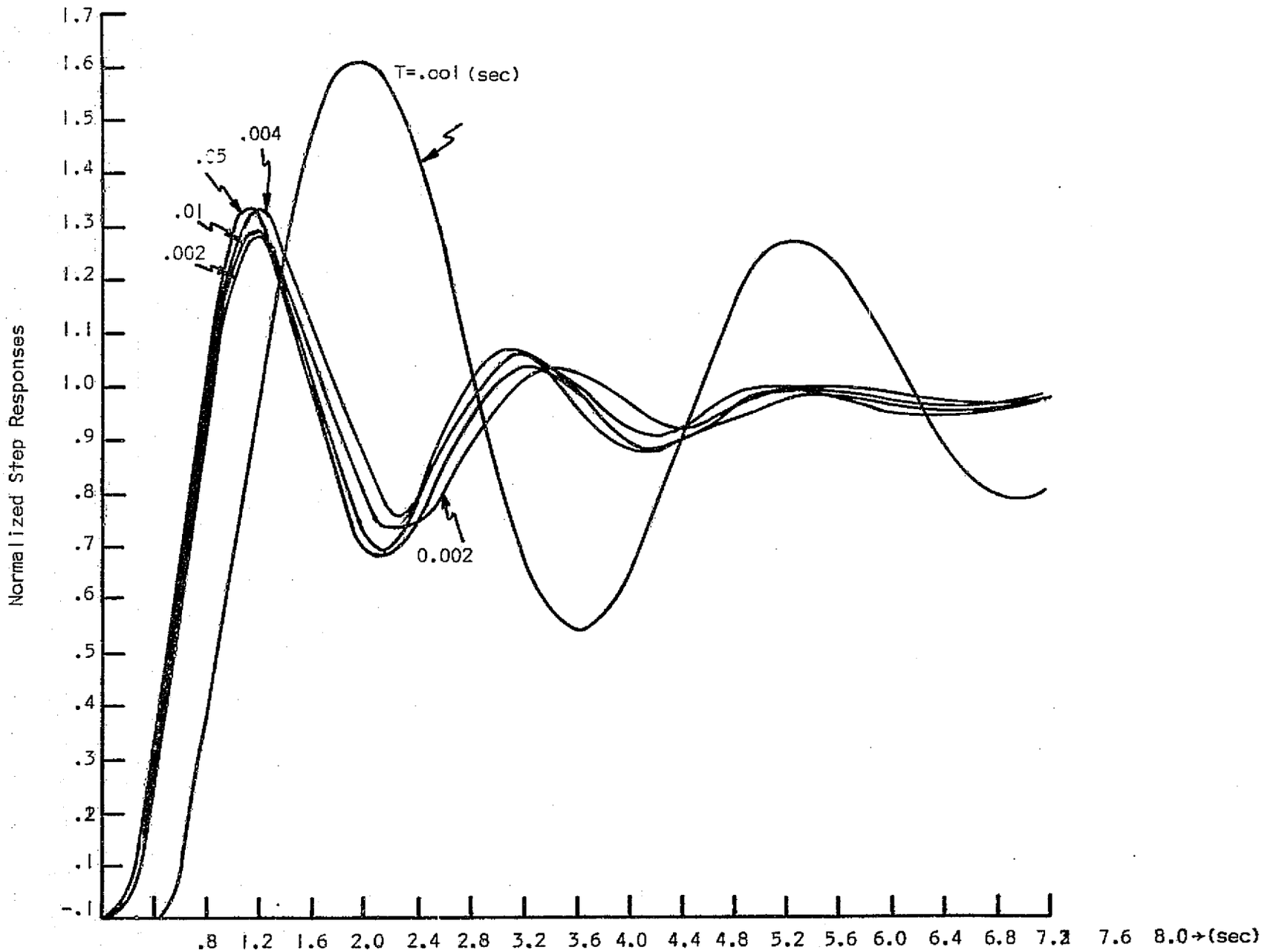


Figure 1.6 Simulated Digital Autopilot System Output Responses Using the Hewlett Packard Computer.

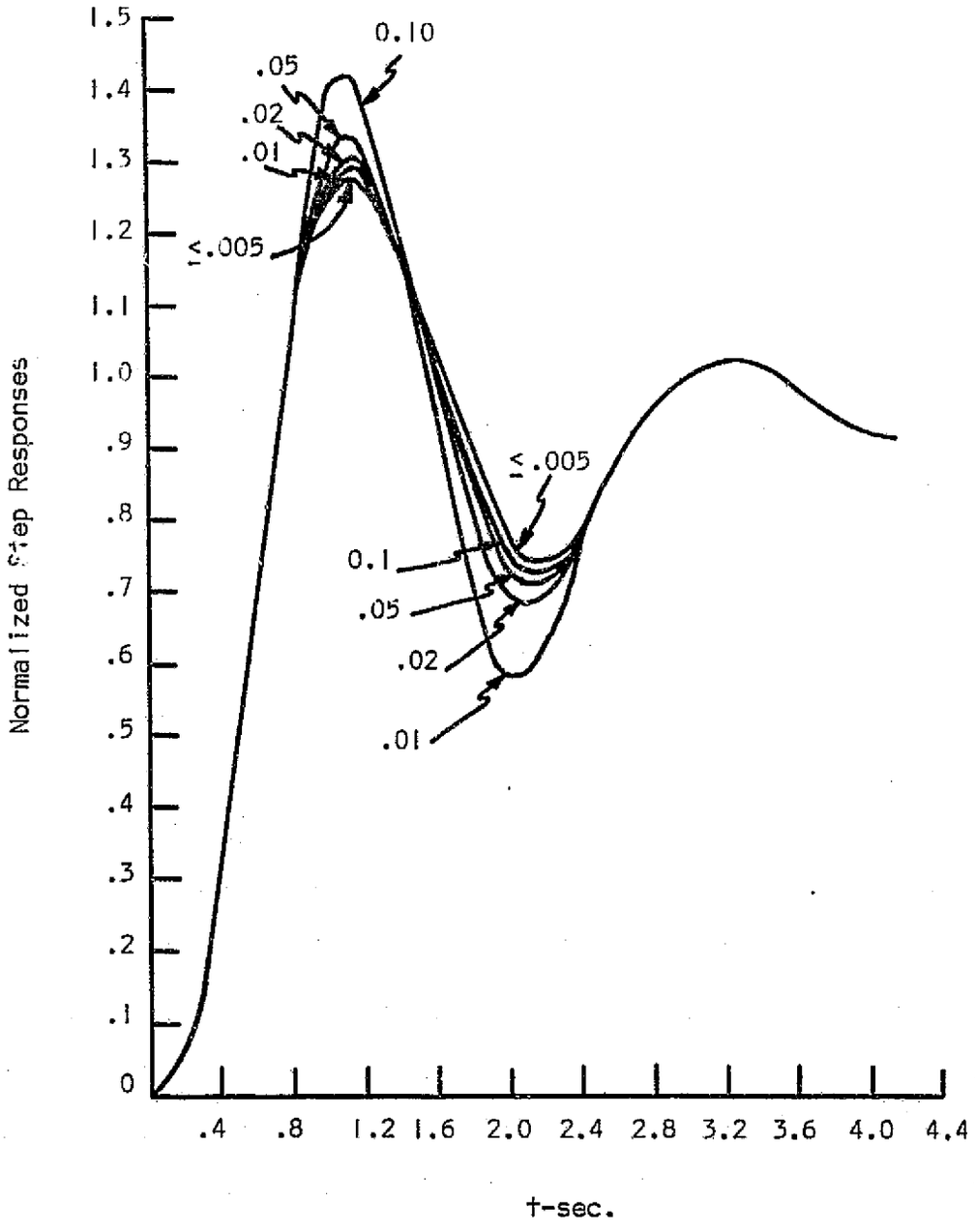


Figure 1.7 Simulated Digital Autopilot System Output Responses Using CDC 6400.

II. DIGITAL SIMULATION METHODS FOR LINEAR SYSTEMS

2.0 Introduction

The methods described in this section are applicable primarily to linear time-invariant systems. Time varying systems whose parameters vary very slowly could also be treated by these techniques, and in some cases weakly nonlinear systems might also yield to this type treatment. However, the approaches are based on the solution of time invariant linear differential equations and, therefore, are most accurate for this class of problems.

2.1 Use of Padé Approximants to the Matrix Exponential for Computer Solutions of State Equations

Work in this area was continued at a very low level. Previous efforts [1] had resulted in Fig. 2.1. This figure can be used to determine the percent error for a given step size and order of approximant. Earlier efforts were concerned with negative real eigenvalues only.

Since our last report, consideration has been given to systems with complex eigenvalues; however, still restricted to the left-half plane. It has now been determined that, for eigenvalues of a given magnitude, the percent error for the various approximants is greatest when the imaginary part of the eigenvalue is zero. Thus, the results presented in Figure 2.1 can be viewed as a worst case for complex eigenvalues. The user need only compute the magnitudes of the various eigenvalues of the A matrix and define λ as the largest of these. This value of λ can then be used in conjunction with Figure 2.1. The user is guaranteed that the percent error actually incurred in the approximation will be no greater than that read off the figure.

Tighter bounds could be obtained by taking into account the angle of the eigenvalues; however, multiple plots would be required, and the results would not be nearly as convenient for the user.

A paper on this topic [4] has been submitted for publication to the IEEE Transactions on Automatic Control.

Next, the phase margin of the digital autopilot system with different sampling periods was calculated and compared with that of the continuous autopilot system. In this calculation for which a hand calculator was used, up to nine digits was retained in each calculation step in order to ensure a degenerate accuracy. The results in Table 1.2 and Fig. 1.8 show that, as the sampling frequency increases, the phase margin of the digital autopilot system converges to that of the continuous autopilot system as it should.

1.2 Conclusions

It was found that increasing the sampling period made the system less stable, i.e. the percent overshoot becomes larger. The phase margin becomes smaller as expected. A designer can observe the step response and decide how much increased overshoot can be allowed. Based on this decision, it is simple to determine the corresponding phase margin from Fig. 1.9 and its associated sampling period.

TABLE I.1

Percent Overshoot vs. Sampling Period (Obtained with CDC 6400)

Sampling Period (sec)	Percent Overshoot
continuous case	27.3%
T = 0.0001	27.3%
T = 0.001	27.3%
T = 0.01	28.2%
T = 0.05	33.2%
T = 0.1	41.0%

TABLE I.2

Phase Margin vs. Sampling Period

Sampling Period (sec)	ϕm
continuous case	27.4°
T = 0.001	27.4°
T = 0.01	26.7°
T = 0.05	24.3°
T = 0.1	21.0°

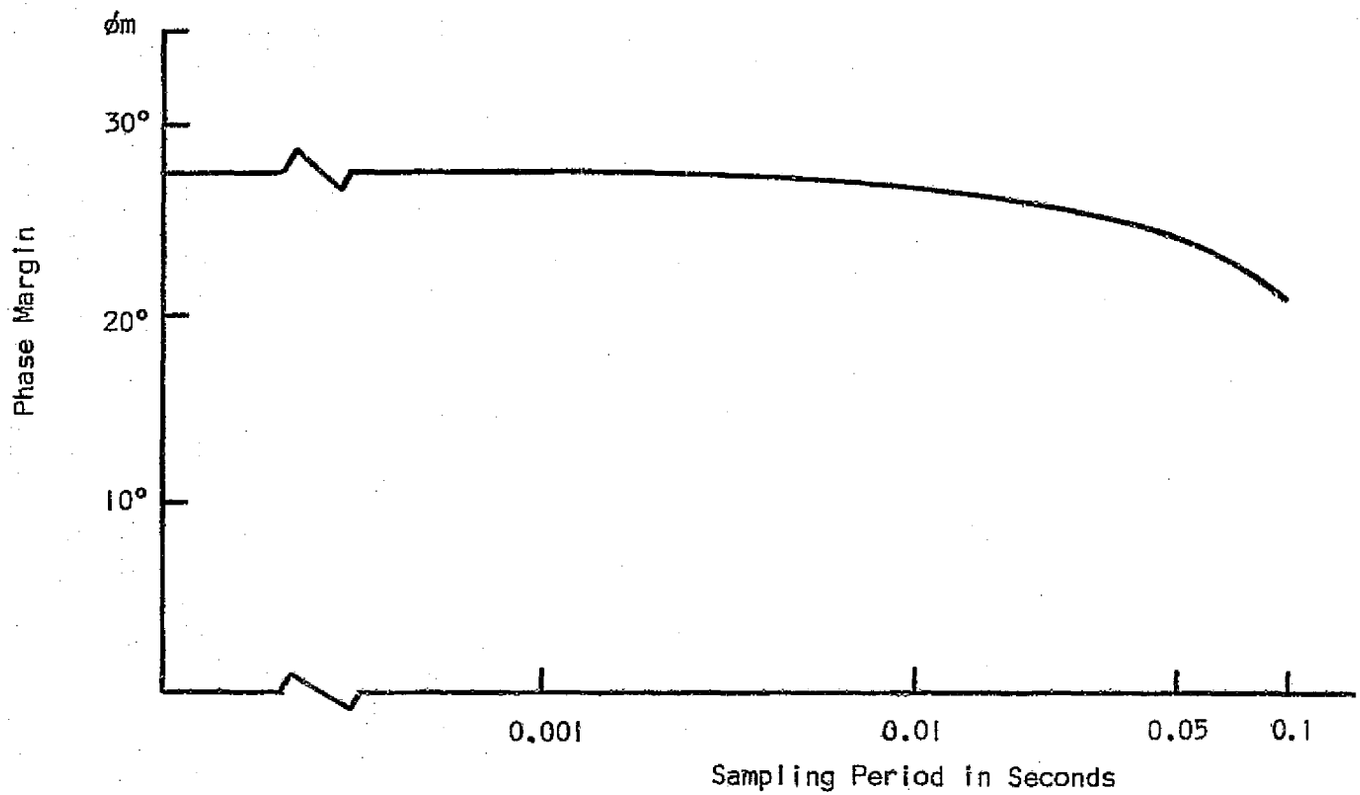


Figure 1.8 Phase Margin vs Sampling Period

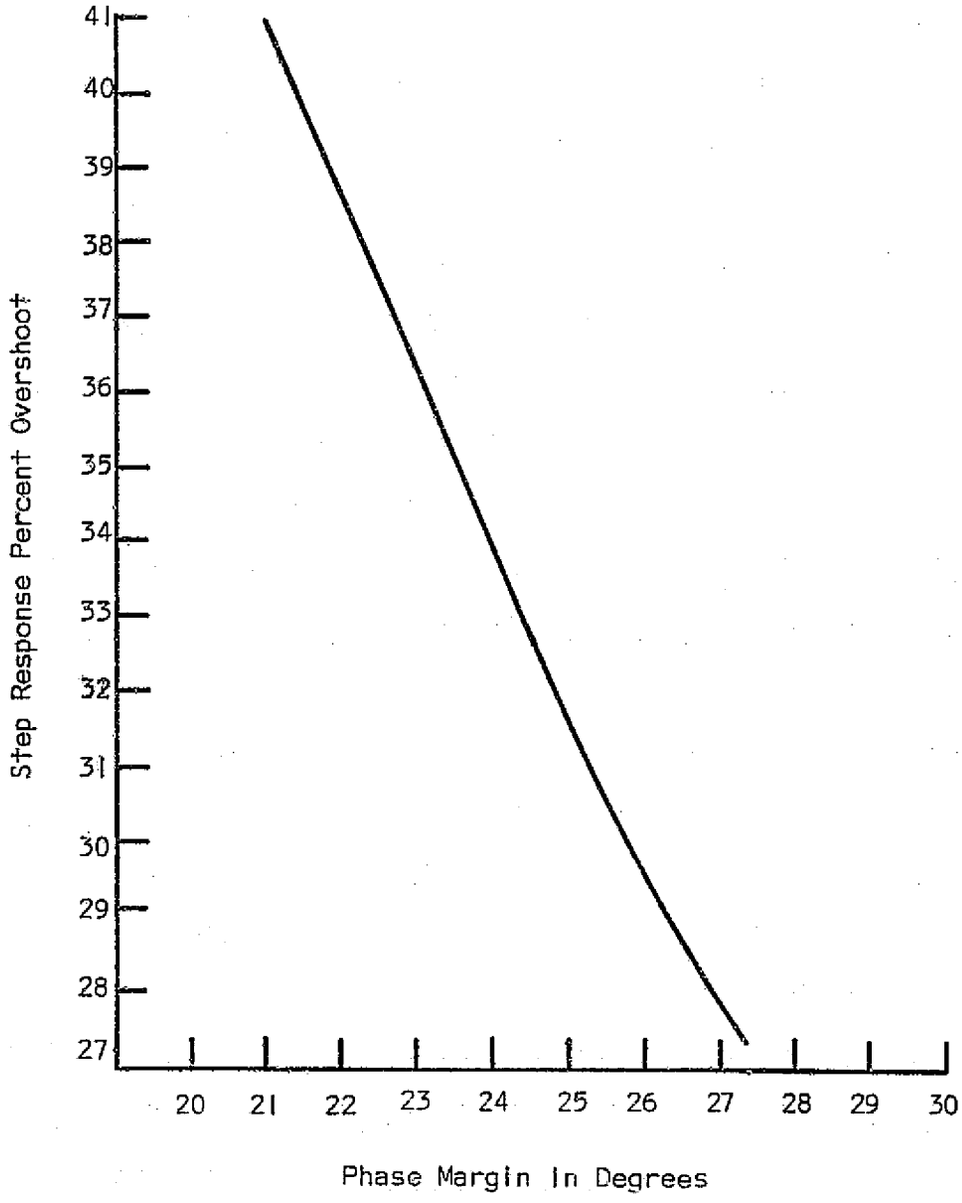


Figure 1.9 Relationship Between the Phase Margin and the Percent Overshoot.

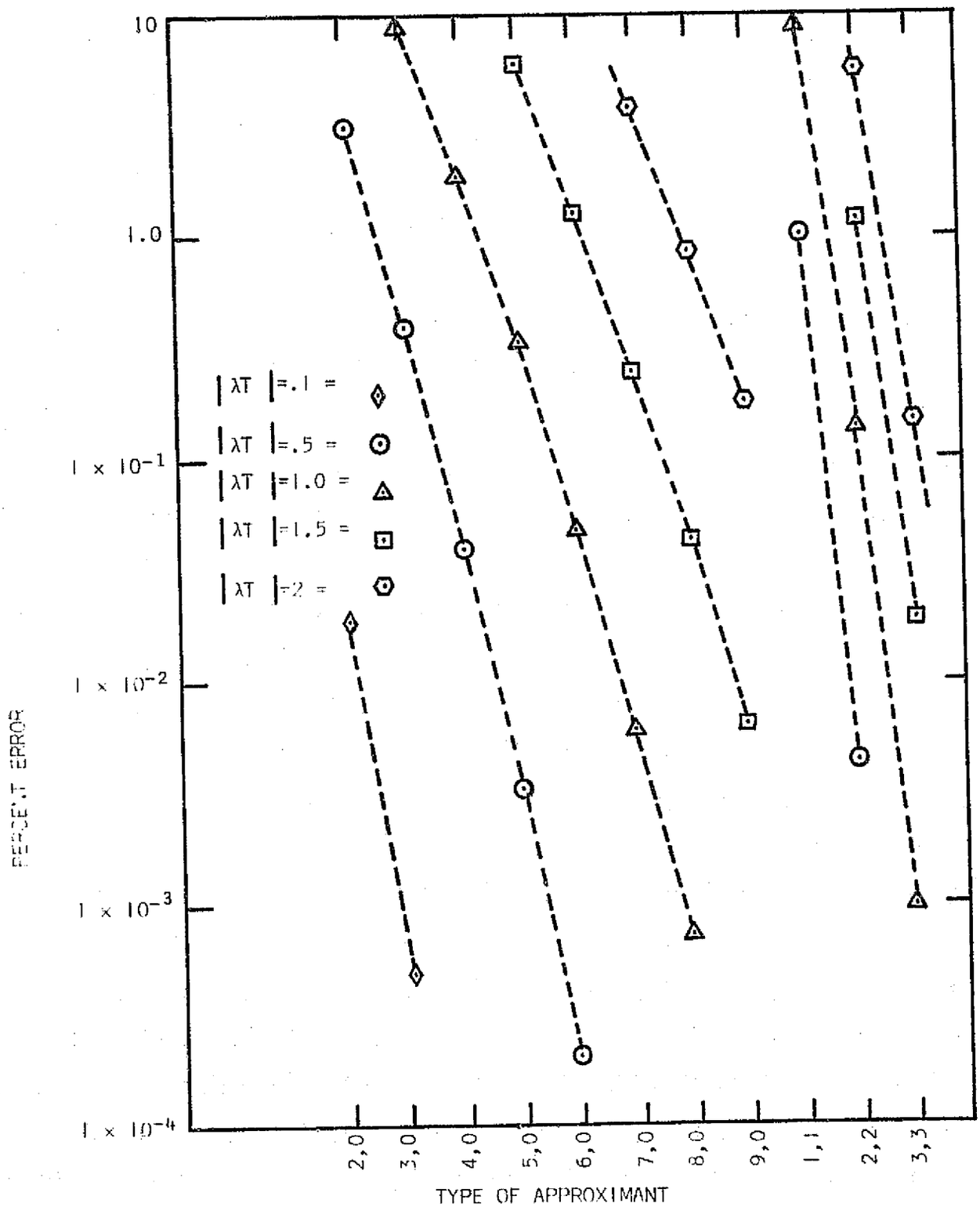

10
1.0
1 x 10⁻¹
1 x 10⁻²
1 x 10⁻³
1 x 10⁻⁴
PERCENT ERROR
|λT|=.1 =
|λT|=.5 =
|λT|=1.0 =
|λT|=1.5 =
|λT|=2 =
2,0
3,0
4,0
5,0
6,0
7,0
8,0
9,0
1,1
2,2
3,3
TYPE OF APPROXIMANT

Figure 2.1

2.2 Analytical Integration of State Equations Using Interpolated Input

2.2.0 Introduction

The basics of this approach were described in last year's final report [1]. The main idea is to use sampled values of the input signal, interpolate between the samples via some type of hold device and then use the analytical solution of the state equations to obtain the output signal. An important feature of this method is that the system is modeled exactly. The approximation is in sampling and interpolating the input.

The solution, when one uses a first-order polynomial to represent the input between samples, is obtained by using the general solution

$$X_N = e^{AT}X_{N-1} + \int_{(N-1)T}^{NT} e^{A(NT-\tau)}BU(\tau)d\tau$$

and approximating $U(\tau)$ by

$$U(\tau) \simeq U_{N-1} + (U_N - U_{N-1})\frac{\tau - (N-1)T}{T}$$

The result is

$$X_N \simeq e^{AT}X_{N-1} + [-A^{-1} + \frac{1}{T}A^{-2}(e^{AT} - 1)]BU_N + [A^{-1}e^{AT} - \frac{1}{T}A^{-2}(e^{AT} - 1)]BU_{N-1}$$

Regardless of the type polynomial used for the interpolation, one can write its transfer function, and it is simply a hold device of one form or another. Figure 2.2 shows a block diagram of the process.

One would like to know the relationships between bandwidth of the input signal, W_I, bandwidth of the plant, W_H, and the sampling frequency, W_S

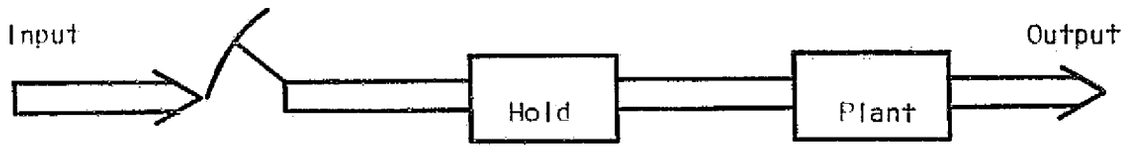


Figure 2.2

as they relate to the accuracy of the discrete representation. Certain minimal requirements on sampling frequency may be obtained by assuming that the hold device is perfect (that is, it has unity gain and no phase shift) over the frequency range $0 - \omega_{s/2}$ but it does not have perfect rejection outside this band, and therefore higher frequencies may be transmitted through. Once these minimal requirements have been established, one can then examine the imperfections of the hold devices over the frequency range $0 - \omega_{s/2}$ and make comparisons between the various ordered polynomial fits.

2.2.1 System Bandwidth Greater Than Input Signal Bandwidth

Consider the case where

$$\omega_H > \omega_I \tag{2.1}$$

and

$$\omega_s > 2\omega_I \tag{2.2}$$

Figure 2.3a shows the spectrum of the input signal, and Fig. 2.3b shows the spectrum of the sampled input signal. In Fig. 3.2c we illustrate the characteristics of a somewhat idealized hold device with unity gain over the frequency range $0 - \omega_{s/2}$ and with non-zero gain at the higher frequencies. Figure 2.3d shows the spectrum of the signal at the output of the hold device. In fig. 2.3e is shown an idealized gain characteristic of the plant. The question we wish to answer is what must be the relationship among ω_s, ω_I, and ω_H so that the final output signal will be the same whether or not sampling occurred. From Fig. 2.3d and 2.3e it is clear that we must have

$$\omega_H < \omega_s - \omega_I \tag{2.3}$$

or

$$\omega_s > \omega_H + \omega_I \tag{2.4}$$

Note that, since

$$\omega_H > \omega_I \tag{2.5}$$

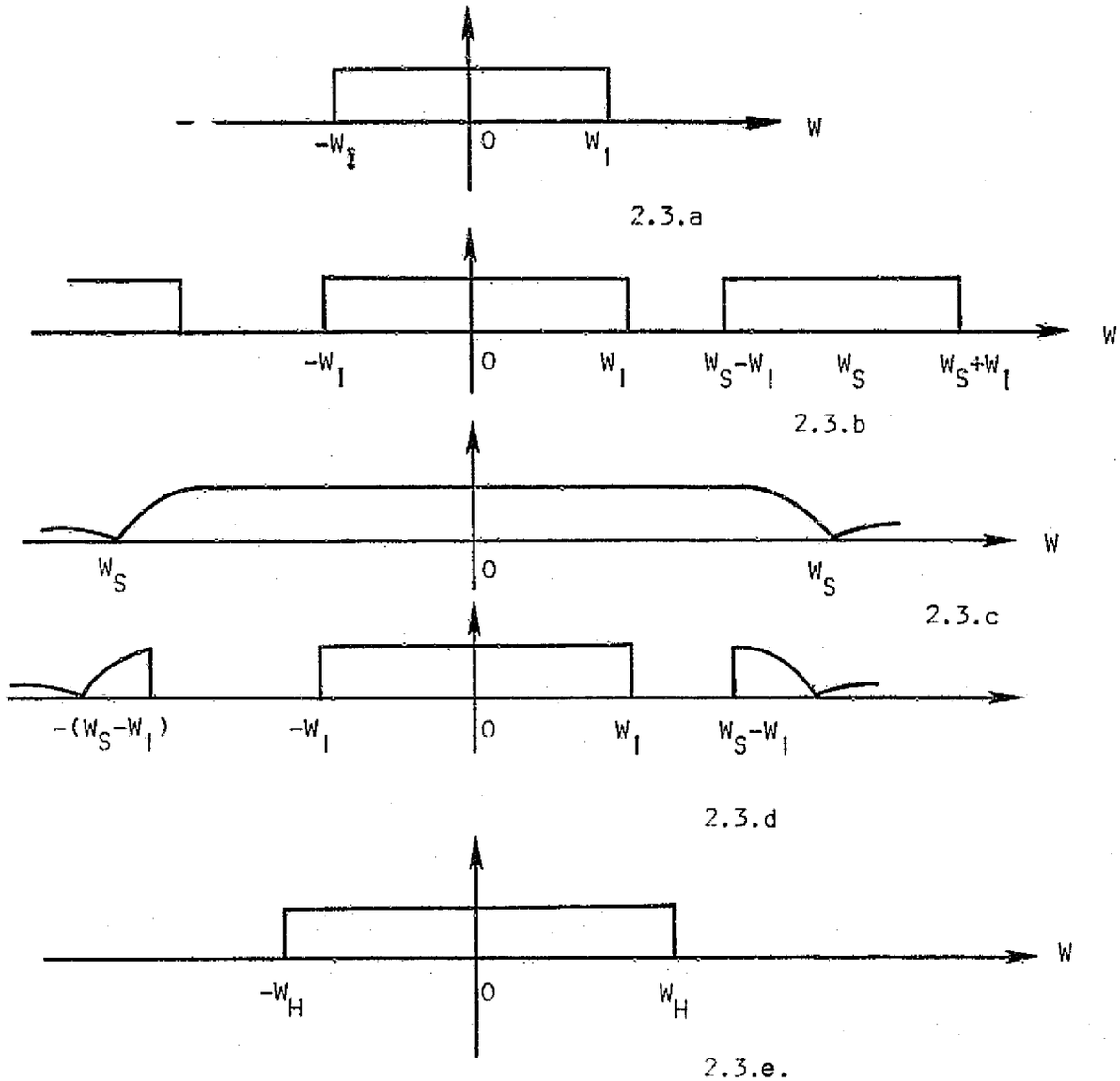

W
$-W_I$
0
W_I
2.3.a
W
$-W_I$
0
W_I
W_S-W_I
W_S
W_S+W_I
2.3.b
W
W_S
0
W_S
2.3.c
$-(W_S-W_I)$
$-W_I$
0
W_I
W_S-W_I
2.3.d
W
$-W_H$
0
W_H
2.3.e.

Figure 2.3

the requirement of inequality (2.4) is more stringent than that of inequality (2.2) which would be the only requirement if we assumed ideal low-pass filtering with cut-off frequency near ω_l.

Consider now the case where

$$\omega_H > \omega_l \tag{2.6}$$

and

$$\omega_s < 2\omega_l \tag{2.7}$$

Figures 2.4a through 2.4e depict this situation. Looking at Figs. 2.4d and 2.4e, one sees that the output signals will be the same with the sampling as it would without the sampling if, and only if,

$$\omega_H < \omega_s - \omega_l \tag{2.8}$$

or

$$\omega_s > \omega_H + \omega_l \tag{2.9}$$

However, this contradicts inequalities (2.6) and (2.7). Therefore, we conclude that one cannot reproduce the output properly if

$$\omega_H > \omega_l \tag{2.10}$$

and

$$\omega_s < 2\omega_l \tag{2.11}$$

Thus, for the case where

$$\omega_H > \omega_l \tag{2.12}$$

the conditions for good output reproduction are

$$\omega_s > 2\omega_l \tag{2.13}$$

and

$$\omega_s > \omega_H + \omega_l \tag{2.14}$$

Since inequality (2.14) is the stronger, its fulfillment guarantees that inequality (2.13) will be fulfilled.

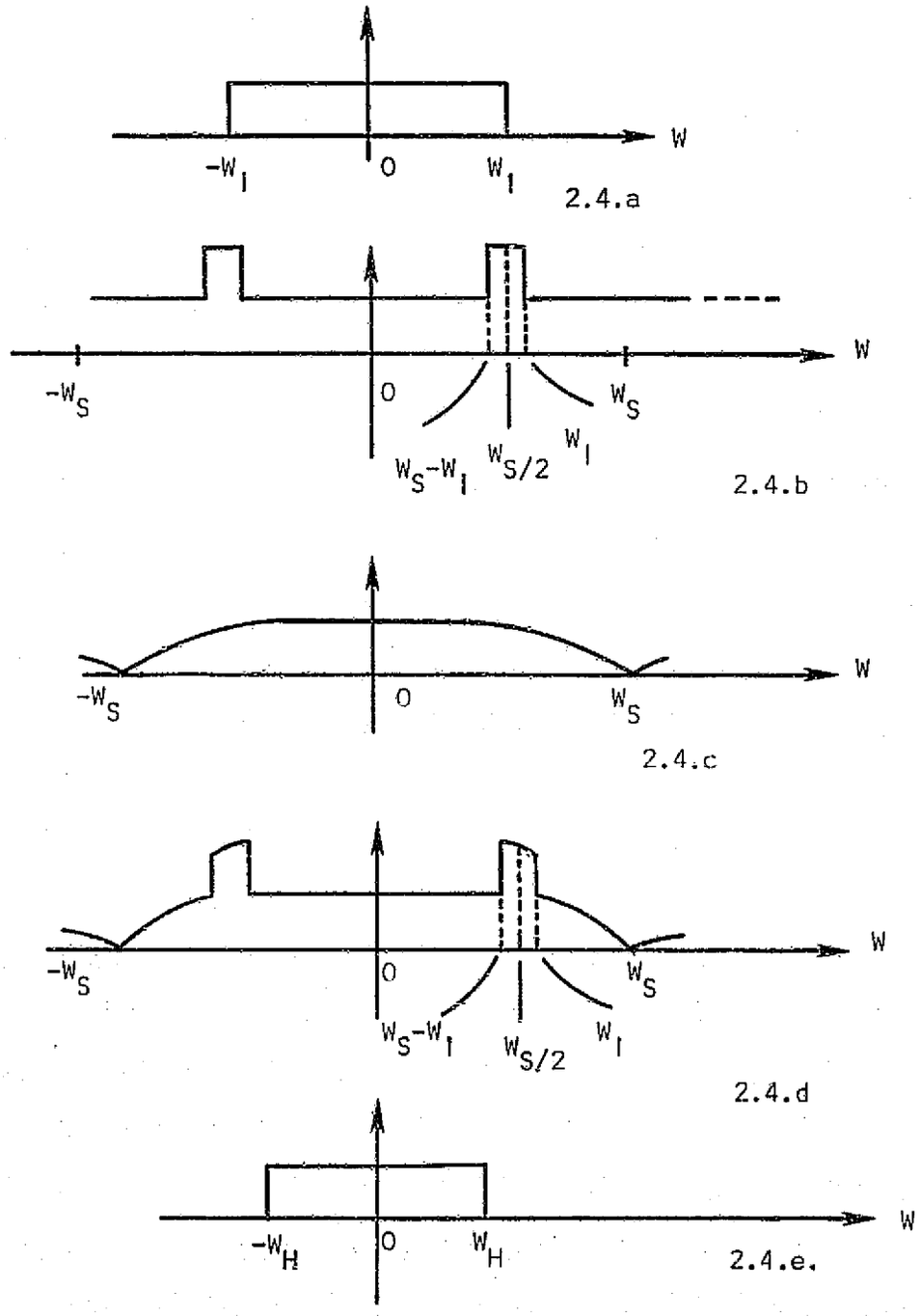

W
$-W_I$
0
W_I
2.4.a
W
$-W_S$
0
W_S
W_I
W_S-W_I
$W_S/2$
2.4.b
W
$-W_S$
0
W_S
2.4.c
W
$-W_S$
0
W_S
W_S-W_I
$W_S/2$
W_I
2.4.d
W
$-W_H$
0
W_H
2.4.e.

Figure 2.4

2.2.2 System Bandwidth Less Than Input Signal Bandwidth

Now, consider the case

$$\omega_H < \omega_I \tag{2.15}$$

and

$$\omega_S > 2\omega_I \tag{2.16}$$

Looking again at Figs. 2.3d and 2.3e, we see that the output will be reproduced properly for this case.

Finally, consider the case

$$\omega_H < \omega_I \tag{2.17}$$

and

$$\omega_S < 2\omega_I \tag{2.18}$$

Looking at Figs. 2.4d and 2.4e, we see that the output may still be reproduced properly if

$$\omega_H < \omega_S - \omega_I \tag{2.19}$$

or

$$\omega_S > \omega_H + \omega_I \tag{2.20}$$

This is a less stringent condition than inequality (2.16), but it still guarantees that the output will be the same with sampling as it would without the sampling.

The final conclusion is that the same minimal conditions on sampling frequency

$$\omega_S > \omega_I + \omega_H \tag{2.21}$$

apply whether

$$\omega_H > \omega_I \tag{2.22}$$

or

$$\omega_I > \omega_H \tag{2.23}$$

These are minimal conditions and guarantee no aliasing or high frequency distortion for the case of the somewhat idealized hold and the ideal cut-off characteristics of the plant which we have assumed. The next step is to examine the actual low frequency ($0 - \omega_{s/2}$) characteristics of the holds represented by the various polynomial interpolation schemes and compare the accuracy taking these imperfections into account.

2.3 Discrete Representation Obtained via the Frequency Domain

The design procedure developed using this approach was discussed, along with an application in last year's final report [1]. Since then, a paper [5] has been written and accepted for publication in the IEEE Transactions on Industrial Electronics and Control Instrumentation.

III. INTERACTIVE SIMULATOR DESIGN PACKAGE FOR THE DESIGN OF REAL-TIME SIMULATORS

3.0 Introduction

This section describes the status of research on an interactive software support system which will aid the design of optimum simulation models. The generic type of system under study is shown in Fig. 3.1. When programming is completed, it is envisioned that the Simulator Design Package (SDP) can be used to evaluate a number of different standard integrator models (for example, Tustin, Optimum Discrete Approximation) on the basis of selectable error criteria or design an entirely new model suitable for a particular problem. In the latter case the model would be designed on an interactive basis using selectable algorithms to find an optimal form.

In previous work [1] we examined a number of different substitution methods to determine which was most suitable under vairous error criteria. Based on these results, a number of substitution formulas have been chosen for inclusion in SDP. Consequently, most of the work during the past six months has concentrated on the evaluation of optimization algorithms and discrete representations. At this stage, very little programming has been accomplished, but the elements of SDP have been defined, and programming is now underway. Thus, the emphasis in this report lies in Mode 2 operation of SDP; that is, that mode of operation in which the user interacts with the system as it attempts to converge iteratively to an optimum model.

This section contains two sub-sections which report on different facets of the development of SDP. In the first sub-section we are concerned with assuming a discrete representation for a given continuous transfer function and then iterating to solve for optimal values of the parameters concerned. In the second sub-section a similar effort is reported on in which a form is assumed for an integration operator, and then a random search method is used to determine the optimum parameters.

3.1 Digital Simulation and Optimization via Gradient Techniques

In previous reports we have examined the IBM, Tustin, Sage, etc., methods for digital simulation. They all share a common characteristic:

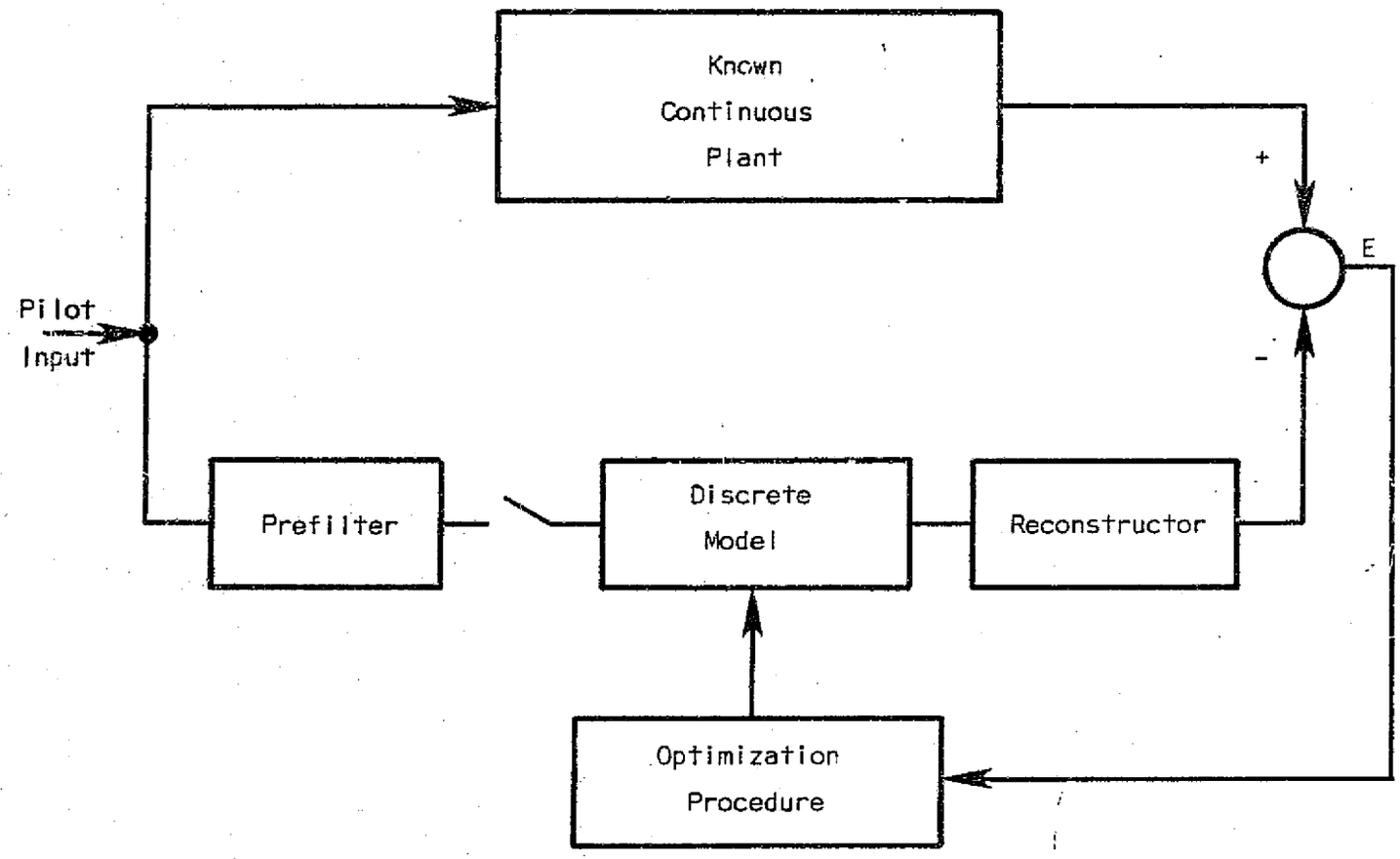


Figure 3.1 System Configuration for Model Development

once the parameters for the digital system from the continuous system are obtained, they cannot be altered to adjust to changes in input. These methods, along with z-transforms, can work satisfactorily if the input can be categorized. However, in many physical situations, the inputs cannot be categorized in this fashion; therefore, some method in which the digital system coefficients can be adjusted for a certain input will indeed be a more accurate representation of a continuous system. In other words, a simulation model for a given continuous system may not be unique but may be a function of the input.

This model can be achieved by a variety of constraints on the response of the digital system: The mean square error in time domain should be minimum; the error in frequency domain is minimum, etc. Therefore, we are confronted with an optimization procedure; and the digital system derived by this procedure should be optimal in some way. Since most performance criteria are nonlinear, an iterative algorithm is required. This algorithm is central to the process and must be modified so that interactive actions between the user and the computer are possible at all points during the convergence of the algorithm.

In order to achieve an effective optimization technique for a digital simulation system, we must first define a performance criterion to be optimized, the possible constraints on the variables, the form of the digital system (so that the performance criterion can be obtained as a function of system parameters) and, finally, an optimization algorithm.

3.1.1 Performance Criteria

The natural choice here is the simulation system error, since our ultimate goal is to match the response of the continuous system and that of the digital system, i.e., the response $\{y_n\}$ produced by an input sequence $\{x_n\}$ is exactly matched to the $\{y(nT)\}$ sequence obtained from sampling the continuous response. This match should occur at all points $\{x_n\}$. This exact matching is, of course, impossible; therefore, we must try to come as close as possible to the ideal situation. This leads to the definition of various errors.

Many error criteria have been proposed in literature. The two most common are the time domain error and the frequency domain error. Frequency domain error is much easier to define than the time domain error, since we only have to compare the frequency responses of the two systems in the range 0-T/2. Moreover, frequency responses of most systems are smooth; therefore, only a few comparisons are needed over the frequency range to obtain good results.

During the first few months of this research, efforts have been concentrated on the frequency domain error. We will repeat the same procedure here for the time domain error in the hope that we can obtain two optimization programs, one based on frequency domain error and the other based on time domain error. The user can try both methods for a given continuous system and choose the better one of the two.

The definition of error is defined as

$$E = \sum_{m=1}^{M} \left| H(j\omega_m) - \bar{H}(j\omega_m) \right|^2, \quad 0 \leq \omega_1 < \omega_2 \cdots < \omega_m < \frac{\pi}{T} \qquad (3.1)$$

where $H(j\omega_i)$ is the desired characteristic at $\omega = \omega_i$, and $\bar{H}(j\omega_i)$ is the simulation characteristic at the same frequency. Unlike time domain error definition, the frequency domain error poses some problems. First, by itself, this error measure presents no constraints on poles locations. Using this error definition, a simulation digital system may have its poles outside the unit circle in z-plane. Second, even if the poles are inside the unit circle, the transient responses of the simulated system and the digital system may be grossly different, resulting in an inaccurate simulation. Therefore, along with this error criterion, some constraints must be placed on the pole locations of the digital transfer function to limit time domain error, as we will see later.

3.1.2 Forms of Digital System

The form of a digital system should be either a cascade or parallel combination of first-or second-order systems, because they are least

sensitive to quantization and round-off errors. Steiglitz [6] proposed the following form:

$$H(z) = \frac{\sum_{k=1}^{K} a_k z^{-(k-1)}}{1 + \sum_{k=1}^{N} b_k z^{-k}} \tag{3.2}$$

This form has some difficulties: first, if control is to be exercised over the pole locations, the denominator must be factored at certain stages in the minimization process; second, the pole locations may be extremely sensitive functions of the coefficients b_k. Therefore, another form is proposed:

$$H(z) = A \prod_{k=1}^{K} \frac{1 + a_k z^{-1} + b_k z^{-2}}{1 + c_k z^{-1} + d_k z^{-2}} \tag{3.3}$$

In this form there are 4k + 1 unknowns (A, a_k, b_k, c_k, d_k) to adjust for minimum error. The optimization procedure used by Steiglitz is based on the algorithm developed by Fletcher and Powell [7]. In order to reduce the dimensionality, the error is minimized with respect to A, leaving 4k unknowns. By supplying various starting points, the least of the minimum values of the error can be obtained, and, also the coefficients a_k, b_k, c_k, d_k. Since the procedure does not place any constraints on pole locations, after the coefficients for H(z) are obtained, we must convert all poles outside the unit circle into the unit circle and start the process again to obtain the final digital form. The conversion of poles (or zeroes) has no effect on the error. Suppose H(z) has a real pole at $z = \alpha$. Replacing $z = \frac{1}{\alpha}$ is equivalent to multiplying by

$$\frac{z - \alpha}{z - \frac{1}{\alpha}} \left| \frac{z - \alpha}{z - \frac{1}{\alpha}} \right| = |\alpha| \tag{3.4}$$

when α is on the unit circle. The same properties hold for complex poles. Therefore, with an adjustment in A, the definition of error is unaffected by this conversion.

A severe disadvantage of this method is that transient response is not being taken into consideration. A pole may fall near, or on, the unit circle which may prevent the response from decaying as fast as that of the continuous response.

Another form which was proposed by D. Schroeder [8] retains the cascade characteristic and also allows the pole magnitudes to be constrained. The form is given by

$$H(z) = A \frac{\prod_{k=1}^{K} (1 + a_{2k-1}z^{-1} + a_{2k}z^{-2})(1 + a'_{Nz}z^{-1})}{\prod_{k=1}^{NCP/2} (1 + 2b_{2k}\cos b_{2k-1}z^{-1} + b^2_{2k}z^{-2}) \prod_{k=NCP+1}^{NP} (1 + b_k z^{-1})} \qquad (3.5)$$

where

NZ = number of zeroes

NP = number of poles

NCP = number of complex poles

K = largest integer $\leq \frac{NZ}{2}$

$a'_{NZ} = 0$ if $NZ = 2K$ (even number of zeroes)

$a'_{NZ} \neq 0$ if $NZ = 2K + 1$ (odd number of zeroes)

This form will be discussed in more detail later, as it proves to be the best form available under the frequency domain error criterin.

3.1.3 Optimization Algorithm

The method of Fletcher and Powell [7] is chosen here for the following reasons:

(1) It has been widely used in many similar problems reported in literature with a high degree of success.

(2) It always converges, and it converges rapidly (it converges quadratically for quadratic functions).

A number of other methods can be used, and the Fletcher-Powell algorithm is, by no means, the best (or final) choice. Its availability as a system supported subroutine in some systems makes it an attractive choice for the initial investigation. In the future other optimization algorithms will be studied in order to evaluate their performances so that the best possible method will be used in the proposed software package. A brief outline of the Fletcher-Powell algorithm is presented here.

The algorithm performs a one-dimensional minimization at each cycle, along a direction determined by the gradient and an updated estimate of the Hessian matrix. The generalized Taylor expansion:

$$f(x + u) = f(x) + g(x)u + \frac{1}{2} u^{+}G(x)u + \ldots \tag{3.6}$$

where x is an n-dimension vector, and u is an n-dimension incremental vector, g(x) is the gradient vector, and G(x) is the n x n matrix of second-order partial derivatives. The displacement between the point x and the minimum x_0 is given by:

$$(x_{min} - x) = -G^{-1}(x_{min})g(x) \tag{3.7}$$

A method of successive linear search is used to find G^{-1}. Fletcher and Powell suggested that initially G° be chosen as the identify matrix I; from this, a sequence of matrices $G^{(i)}$ is generated which will converge to G^{-1}. The search is terminated when the function value $f(x^{\dagger})$ has not decreased in the last step and the gradient vector is small.

By itself, this algorithm imposes no constraints on the variable. Therefore, we must place artificial limits on the pole locations before applying the algorithm.

Let us now return to the minimization of error to obtain the "optimal" coefficients of a digital model. Recall:

$$E = \sum_{m=1}^{M} \left| H(j\omega_m) - \tilde{H}(j\omega_m) \right|^2, \quad 0 < \omega_1 < \omega_2 < \cdots < \omega_m < \frac{\pi}{T} \tag{3.8}$$

and

$$\tilde{H}(z) = A \frac{\prod_{k=1}^{K} (1 + a_{2k-1}z^{-1} + a_{2k}z^{-2})(1 + a'_{NZ}z^{-1})}{\prod_{k=1}^{NCP/2} (1 + 2b_{2k}\cos b_{2k-1}z^{-1} + b_{2k}z^{-1}) \prod_{k=NCP+1}^{NP} (1 + b_k z^{-1})} \tag{3.9}$$

The constraints on the coefficients are simple: the poles, both real and complex, are required to lie within a circle of predetermined radius. This predetermined radius is obtained from the pole locations in the s-plane. This restriction will ensure that the transient responses of the two systems will decay at the same rate

$$z\text{-pole} = e^{T(s\text{-plane pole})} \tag{3.10}$$

A less restrictive form is:

$$|z| \leq |e^{sT}| = e^{T(\text{Real } s)} \tag{3.11}$$

Since the Fletcher-Powell algorithm requires the function value and the gradient vector at each iteration, the error function must be manipulated to provide these requirements. First, replace $z = \alpha j^{\omega T}$ to get

$$E = \sum_{m=1}^{M} \left\{ \left| H(j\omega_m) - A \frac{\prod_{k=1}^{K} (1+a_{2k-1}e^{-j\omega_m T}+a_{2k}e^{-j2\omega_m T})(1+a'_{NZ}e^{-j\omega_m T})}{\prod_{k=1}^{NCP/2} (1+2b_{2k}\cos b_{2k-1}e^{-j\omega_m T}+b_{2k}^2 e^{-j2\omega_m T}) \prod_{k=NCP+1}^{NP} (1+b_k e^{-j\omega_m T})} \right|^2 \right\} \tag{3.12}$$

A can be obtained by matching the steady-state gains of the two systems:

$$A = H(0)\,\frac{\prod_{k=1}^{NCP/2}(1+2b_{2k}\cos b_{2k-1}+b^2)\;\prod_{k=NCP+1}^{NP}(1+b_k)}{\prod_{k=1}^{K}(1+a_{2k-1}+a_{2k})(1+a_{NZ}')} \tag{3.13}$$

With A eliminated, we now have NZ numerator coefficients and NP denominator coefficients by which to minimize the error. The error can be written:

$$E = \sum_{m=1}^{M}\left[(HR_m - HBR_m)^2 + (HI_m - HBI_m)^2\right] \tag{3.14}$$

where

$$\left.\begin{aligned} HR_m &= \text{Real}\,(H(j\omega_m)) \\ HI_m &= \text{Imag}\,(H(j\omega_m)) \end{aligned}\right\} \text{ known constants}$$

$$\left.\begin{aligned} HBR_m &= \text{Real}\,(\bar{H}(j\omega_m)) \\ HBI_m &= \text{Imag}\,(\bar{H}(j\omega_m)) \end{aligned}\right\} \text{ contain } a_i\text{'s},\ b_i\text{'s}$$

The gradient vector is given by

$$g = \left(\frac{\partial E}{\partial a_1}, \frac{\partial E}{\partial a_2}, \ldots, \frac{\partial E}{\partial a_{NZ}}, \frac{\partial E}{\partial b_1}, \frac{\partial E}{\partial b_2}, \ldots, \frac{\partial E}{\partial B_{NP}}\right) \tag{3.15}$$

when the i th element is

$$\frac{\partial E}{\partial a_i} = -2\sum_{m=1}^{M}(HR_m - HBR_m)\frac{\partial HBR_m}{\partial a_i} + (HI_m - HBI_m)\frac{\partial HBI_m}{\partial a_i} \tag{3.16}$$

with the $\frac{\partial E}{\partial b_i}$ given by a similar expression. In order to obtain an explicit expression for the gradient vector $\frac{\partial HBR_m}{\partial a_i}$, $\frac{\partial HBI_m}{\partial a_i}$, $\frac{\partial HBR_m}{\partial b_i}$ and $\frac{\partial HBI_m}{\partial b_i}$ must be obtained. The concept is simple, but the algebraic manipulation is extremely tedious. We only summarize the procedure and list the final results here.

First, we use the definition of $\tilde{H}(z)$ in the digital form, replacing A with its equivalent expression obtained from matching the d.c. gains. Second, we separate

$$\tilde{H}(j\omega_\ell) = HBR_\ell + jHBI_\ell \qquad (\ell = 1,2,\ldots M) \tag{3.17}$$

Third, we isolate a given variable and obtain the partial derivatives

$$\frac{\partial}{\partial a_i}(HBR_\ell) = Real\left[\frac{\partial}{\partial a_i}(HBR_\ell + jHBI_\ell)\right] \tag{3.18}$$

and

$$\frac{\partial}{\partial a_i}(HBI_\ell) = Imag\left[\frac{\partial}{\partial a_i}(HBR_\ell + jHBI_\ell)\right] \tag{3.19}$$

For i even:

$$\frac{\partial}{\partial a_i}[HBR_\ell + JHBI_\ell]$$

$$= \left[\frac{HBR_\ell + jHBI_\ell}{1+a_{i-1}+a_i}\right] \left\{\frac{[(1+a_{i-1})\cos\omega_\ell T - 1 - a_{i-1}\cos\omega_\ell T] + j(I)}{1+a_{i-1}\cos\omega_\ell T + a_i\cos 2\omega_\ell T) - j(a_{i-1}\sin\omega_\ell T + a_i \sin 2\omega_\ell T)}\right\} \tag{3.20}$$

where

$$I = [a_{i-1}\sin\omega_\ell T - (1+a_{i-1})\sin 2\omega_\ell T) \tag{3.21}$$

For i odd:

$$\frac{\partial}{\partial a_i}[HBR_\ell + jHBI_\ell]$$

$$= \frac{[HBR_\ell+jHBI_\ell][\cos\omega_\ell T(1+a_{i+1})-1-a_{i+1}\cos 2\omega_\ell T)+j(a_{i+1}\sin 2\omega_\ell T-(1+a_{i+1})\sin\omega_\ell T)]}{[1+a_i+a_{i+1}][(1+a_i\cos\omega_\ell T+a_{i+1}\cos 2\omega_\ell T)-j(a_{i-1}\sin\omega_\ell T+a_i\sin 2\omega_\ell T)} \tag{3.22}$$

For NZ odd:

$$\frac{\partial}{\partial a_{NZ}}[HBR_\ell + jHBI_\ell] = \frac{[HBR_\ell + jHBI_\ell]\,[(\cos\omega_\ell T - 1) - j\sin\omega_\ell T]}{(1 + a_{NZ})[(1 + a_{NZ}\cos\omega_\ell T) - j\,a_{NZ}\sin\omega_\ell T]} \tag{3.23}$$

The derivatives with respect to the denominator coefficients are:

For i even:

$$\frac{\partial}{\partial b_i}[HBR_\ell + jHBI_\ell] = \frac{2[HBR_\ell + jHBI_\ell]}{[1 + 2b_i\cos b_{i-1} + b_i^2]} \left\{\frac{R + jI}{R' + jI'}\right\} \tag{3.24}$$

where

$$R=[(\cos b_{i-1}+b_i)(1+2b_i\cos\omega_\ell T\cos b_{i-1}+b_i^2\cos 2\omega_\ell T$$
$$-(1+2b_i\cos b_{i-1}+b_i^2)(\cos b_{i-1}\cos\omega_\ell T+b_i\cos 2\omega_\ell T)] \tag{3.25}$$

$$I=[(1+2b_i\cos b_{i-1}+b_i^2)(\cos b_{i-1}\sin\omega_\ell T+b_i\sin 2\omega_\ell T) - (\cos b_{i-1}+b_i)(2b_i\cos b_{i-1}\sin\omega_\ell T+b_i^2\sin 2\omega_\ell T)] \quad (3.26)$$

$$R' = [1 + 2b_i\cos b_{i-1}\cos\omega_\ell T + b_i^2\cos 2\omega_\ell T] \quad (3.27)$$

$$I' = -[2b_i\cos b_{i-1}\sin\omega_\ell T + b_i^2\sin 2\omega_\ell T] \quad (3.28)$$

For i odd:

$$\frac{\partial}{\partial b_i}[HBR_\ell + jHBI_\ell] = \frac{2[HBR_\ell + jHBI_\ell]}{(1 + 2b_{i+1}\cos b_i + b_{i+1}^2} \left\{\frac{R_1 + jI_1}{R_1' + jI_1'}\right\} \quad (3.29)$$

where

$$R_1=(1+2b_{i+1}\cos b_i+b_{i+1}^2)b_{i+1}\cos\omega_\ell T\sin b_i-b_{i+1}\sin b_i(1+2b_{i+1}\cos b_i\cos\omega_\ell T+b_{i+1}^2\cos 2\omega_\ell T) \quad (3.30)$$

$$I_1=b_{i+1}\sin b_i(2b_{i+1}\cos b_{i+1}\sin\omega_\ell T+b_{i+1}^2\sin 2\omega_\ell T)-(1+2b_{i+1}\cos b_i+b_{i+1}^2)b_{i+1}\sin b_i\sin\omega_\ell T \quad (3.31)$$

$$R_1' =1+2b_{i+1}\cos b_i\cos\omega_\ell T+b_{i+1}^2\cos 2\omega_\ell T \quad (3.32)$$

and

$$I_1' = -[2b_{i+1}\cos b_i\sin\omega_\ell T+b_i^2\sin 2\omega_\ell T] \quad (3.33)$$

Finally, for i = NCP + 1, NCP + 2,..., NP

$$\frac{\partial}{\partial b_i}[HBR_\ell + jHBI_\ell] = \frac{[HBR_\ell + jHBI_\ell][(1 - \cos\omega_\ell T) + j\sin\omega_\ell T]}{(1 + b_i)[(1 + b_i\cos\omega_\ell T) - jb_i\sin\omega_\ell T]} \quad (3.34)$$

We now have explicit functions for the gradient vector, and the Fletcher-Powell algorithm can be applied after the constraints have been taken into account.

Computer programs have been written for this method by D. Schroeder[8]; and at the time of this report similar local programs are being developed. Results will be included in the next report.

In summary, the following data are needed:

(1) The form of the desired continuous system to be simulated

$$\overline{H}(j\omega) = HR(j\omega) + JHI(j\omega) \text{ for } 0 \leq \omega \leq \frac{\Pi}{T} \tag{3.35}$$

(2) Sampling interval T

(3) Frequencies at which the error measure is evaluated, i.e., ω_m for $0 < \omega_1 < \omega_2 \cdots < \omega_m < \frac{\Pi}{T}$

(4) The maximum radius of poles (to be used for constraints)

(5) Some expected minimum error

(6) Number of zeroes, poles, and complex poles of the digital system

(7) Initial guesses for the numerator and denominator coefficients

Further studies on this method are underway. It is envisioned that a similar approach can be applied for a time domain error criterion, even though an accurate time domain measure, suitable for minimization techniques, appears to be complicated mathematically.

3.1.4 Future efforts

(1) Further investigations are to be directed in the method presented above.

(2) Develop similar techniques, using time domain error criteria.

(3) Develop general programs for IBM (which seem to be the most difficult from the programming point of view), Tustin and Optimum Discrete Approximation approach. The programming difficulties for the optimum

discrete approximation are at least circumvented by a study carried out during earlier work on this grant. It appears that in the Optimum Discrete Approximation method, a system can be broken down into a cascade or parallel combination of first-and second-order systems for which the discrete approximators are readily available. Therefore, as far as programming is concerned, the Optimum Discrete Approximation method can be considered just another substitutional method.

3.2 Digital Simulation and Optimization via Random Search Techniques

3.2.0 Introduction

Real-time digital simulation of physical systems has received attention for a number of years [9], [10]. This paper discusses a technique for the development of a discrete time integration operator to be used in the simulation process. The integration operator can be optimized for a particular system subjected to a set of specified inputs. The class of systems being investigated are those which can be represented by the set of state equations

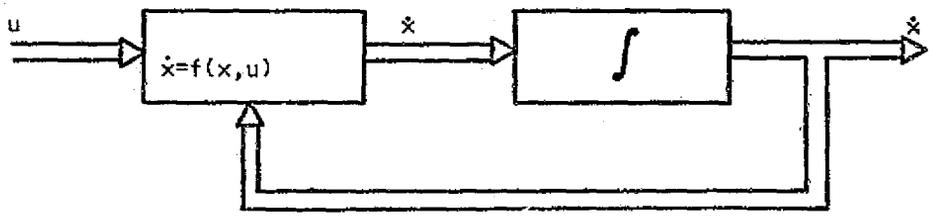
$$\dot{x} = f(x, u) \tag{3.36}$$

where x is the n x 1 state vector, u is for the r x 1 control vector, and f is the set of n functions, typically nonlinear.

3.2.1 Integration Operator

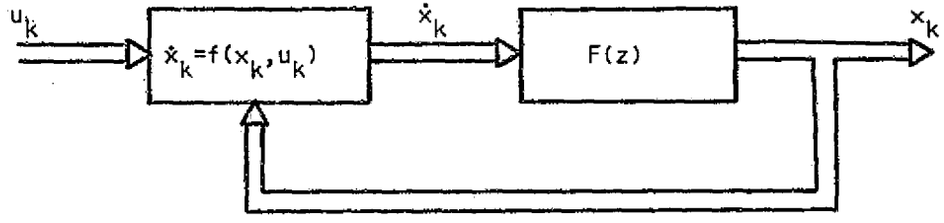
Figure 3.2a is a block diagram of the mathematical relations in Eqn. (3.36). The vectors x and u are acted upon by the functional relations f(x, u), producing the vector $\dot{x}$. This is then integrated to produce the state vector x. Figure 3.2b is a block diagram of a discrete approximation to the continuous time system. The control vector u is assumed to be sampled at a uniform rate, producing the input samples u(k). The equations

$$x(k) = f[x(k), u(k)] \tag{3.37}$$



(a)

CONTINUOUS TIME SYSTEM



(b)

DISCRETE TIME SIMULATION

$$F(z) = \frac{\sum_{i=o}^{M} a_i z^i}{\sum_{j=o}^{N} b_j z^j}$$

N>M FOR REALIZABLE SIMULATION

Figure 3.2 Digital Simulation of Nonlinear System.

are in the same form as those representing the continuous time system. For example, if

$$\dot{x}_1 = \cos x_2 + \sin x_3 + u$$

then

$$\dot{x}_1(k) = \cos x_1(k) + \sin x_3(k) + u(k)$$

The function F(z) represents the discrete integration operator. For the simulation to be realizable, the denominator of F(z) must be of a higher power in z than the numerator [9]. This can be shown to be correct as follows. Let F(z) be

$$F(z) = \frac{T}{2}\frac{z+1}{z-1}$$

which is the familiar Tustin substitution operator for continuous integrator $\frac{1}{s}$.

$$\frac{x(z)}{\dot{x}(z)} = \frac{T}{2}\,\frac{z+1}{z-1}\,\frac{T}{2}\frac{1+z^{-1}}{1-z^{-1}}$$

$$(1-z^{-1})x(z) = \frac{T}{2}(1+z^{-1})\dot{x}(z) \qquad (3.38)$$

$$x(z) = z^{-1}x(z) + \frac{T}{2}(1+2^{-1}(\dot{x}(z)$$

$$x(k) = x(k-1) + \frac{T}{2}[\dot{x}(k) + \dot{x}(k-1)]$$

The calculation of the state at time k is seen to depend on its derivative at time k. However, from the state equations, the derivative at time k is a function of the state at time k, (3.37). This leads to an equation of the form

$$x(k) = x(k-1) + \frac{T}{2}\{f[x(k), u(k)] + f[x(k-1), u(k-1)]\}$$

Since the functions f(x, u) are typically nonlinear, the equation cannot be solved by factoring x(k) out of the right-hand side of the equation. Therefore, (3.38) represents an unrealizable simulation. If the integration operator had been chosen as

$$F(z) = \frac{T}{2}\left(\frac{z+1}{z-1}\right) z \quad = \frac{T}{2}\frac{z^{-1} + z^{-1}}{1 - z^{-1}}$$

then

$$x(k) = x(k-1) + \frac{T}{2}[\dot{x}(k-1) + \dot{x}(k-2)] \tag{3.39}$$

which is a realizable simulation. Adding the z^{-1} term to the Tustin operator, however, degrades its performance. It has been shown, though, that to be closed-loop realizable, the power of the denominator must exceed that of the numerator. What is desired is a closed-loop realizable operator which can be optimized for a particular system being driven by a set of known inputs.

To this end, a discrete-time integration operator of the following form is chosen.

$$F(z) = \frac{T\sum_{j=0}^{N} \lambda_j z^j}{z^N(z-1)} \tag{3.40}$$

where T is the sampling period, and the λ's are a set of free parameters, the values of which are to be optimized. This operator yields a realizable simulation, since the power of the denominator is always one greater than that of the numerator. The pole at z = 1 corresponds to a pole at the origin in the complex s-plane, and the N^{th} order pole at the origin corresponds to an N^{th} order pole at infinity in the s-plane [11]

The free parameters in F(z) are optimized using an idealized model form of a model reference adaptive control system [12]. Figure 3.3 is a block diagram of this form. An exact, or nearly exact, solution to the system's differential equations represents the ideal model of the system. The discrete simulation, including the parameters to be optimized, represents the controlled process. A set of inputs is applied to the model and the process, and the error at each sample time is squared and summed. At the end of one run, the free parameters are perturbed under the control of an optimization technique; and the sequence is repeated. This continues until the error of the digital simulation is sufficiently small.

3.2.2 Optimization Technique

The perturbation of parameters in F(z) is controlled by an Adaptive Random Search Optimization technique (ARSO). Random perturbation methods have been shown to solve a large class of optimization problems faster than gradient techniques when the number of unknown parameters exceeds four [13], [14], [15]. In addition, the convergence time has empirically been shown to increase linearly with the number of unknown parameters rather than exponentially or quadratically [15].

In the ARSO technique, the mean and variance of a uniformly distributed random variable are adaptively selected for each unknown parameter based on recent successful experiments. In this way the step size and direction are controlled by past successful perturbations. Adaptive step size has been shown by Schumer and Steiglitz [15] to be a powerful technique in multidimensional problems without ridges or valleys. Adding adaptive step direction should tend to broaden the range of applicability.

The performance index used with ARSO is a vector valued one; that is, one which requires simultaneous minimization of all components [13]. This allows consideration of several cost functions, such as integral square error, minimum energy, etc., at the same time. In the nonlinear problem to be studied, air-craft dynamics, the mean square error for each of the state variables is used as a component in the performance index.

OPTIMAL DIGITAL SIMULATOR DEVELOPMENT
VIEWED AS A MODEL REFERENCE ADAPTIVE CONTROL
SYSTEM USING ADAPTIVE RANDOM SEARCH OPTIMIZATION TECHNIQUES

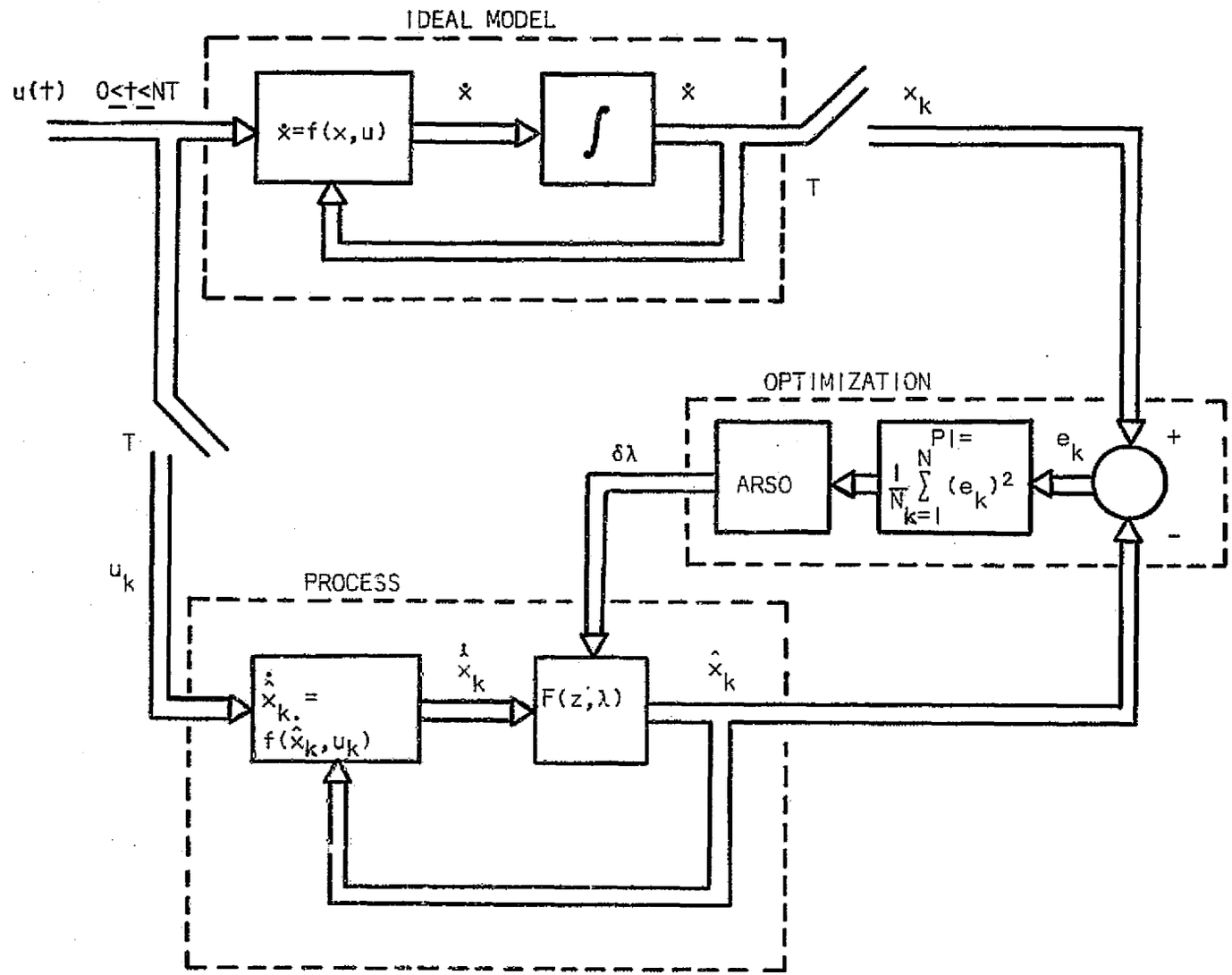


Figure 3.3

That is,

$$J_i = \frac{1}{N} \sum_{j=1}^{N} (x_{ij} - x^*_{ij})^2 \tag{3.41}$$

and

$$J = [J_1 \; J_2 \; \ldots \; J_n]^T \tag{3.42}$$

where N is the number of sample periods per trial, x_{ij} is the calculated value or the i^{th} state variable at the j^{th} sample time, and x^*_{ij} is the exact value. For a trial to be considered a success, at least one component of J must be reduced, and no component may increase in value.

The unknown parameters are perturbed in the following manner.

$$\lambda_i(j + 1) = \lambda^*_i + \delta\lambda_i(j) \tag{3.43}$$

where $\lambda(j + 1)$ is the new value of the i^{th} parameter, λ^*_i is the "best-to-date" value of the i^{th} parameter; that is, the value of λ_i when the minimum-to-date value of the J vector was calculated, and $\delta\lambda_i(j)$ is the random perturbation for the i^{th} parameter. The perturbation is calculated as follows:

$$\delta\lambda_i(j) = \mu_i(j) + \sqrt{3\sigma^2(j)}\,[2\,\mathrm{RND}(o) - 1] \tag{3.44}$$

where $\mu_i(j)$ is the current value for the mean of the i^{th} random variable, $\sigma^2(j)$ is the current value for the variance, and RND(o) is a uniformly distributed random variable on the interval [0,1]. Equation (3.44) produces a random number from a uniformly distributed random variable with mean μ and variance σ^2. At the present time, the variance is the same for each of the parameters, while a mean is calculated for each. The use

of individual variances, as well, is an area to be studied. Stability considerations may place constraints on the parameter calculated in (3.43). If a particular perturbation places a value outside its limit for stability, the value may be moved deterministically inside the limit. The particular system will determine what the stability limits for the parameters are.

When a particular trial is successful; that is,

$$J_i \leq J_i^* \qquad 1 \leq i \leq n$$

where J_i^* is the minimum value of the i^{th} element of J, the means of the random variables are updated.

$$\mu_i(j) = \lambda_i^* - m_i \tag{3.45}$$

where

$$m_i = \frac{c_i(1)}{2} + \frac{c_i(2)}{4} + \frac{c_i(3)}{8} + \frac{c_i(4)}{16} + \frac{c_i(5)}{16} \tag{3.46}$$

The c_i's are past values of λ_i^*; that is, previous values of the "best-to-date" parameters. $c_i(1)$ is the most recent value of λ_i^*. Since the most recent best value corresponds to a smaller performance index than a previous best value, (3.45) tends to select the most favorable direction for the next perturbation. For an unsuccessful trial, the mean is not changed until the number of consecutive failures becomes large. The strategy for the mean and variance is as follows:

Initially, the mean is zero; and the variance is equal to two. This permits large perturbations in the values of the parameters. As the trials become successful, the means are computed by (3.45); and the variance is greatly reduced. This permits the parameter values to follow favorable terrain. The mean is updated at each success as previously mentioned. When the number of consecutive failures reaches 100, the variance is increased to allow larger step sizes. At this point, the means remain

unchanged. If the number of consecutive failures reaches 1000, the variance is greatly increased; and the means are set to zero. This allows the possibility of jumping out of a local minimum to an area with a smaller performance index. When the number of consecutive failures reaches 10,000, the program is terminated. If, at any time, a successful trial occurs, the variance is reduced; and the means are updated as before.

3.2.3 Preliminary Results

Thus far, the ARSO technique has only been used for the devopment of general-purpose integration operators for linear systems. A three-element performance index was used to measure deviation of the integrator operators from ideal integration. Frequency domain magnitude error from zero to one-half the sampling frequency, phase error over the same frequency range, and phase error from zero to eight-tenths of one-half the sampling frequency made up the three components of J. One and four parameters were used in F(z) on different experiments. The two operators obtained are shown below.

$$F(z) = \frac{.8599T}{z - 1}$$

$$F(z) = \frac{T(.9718z^3 - .1768z^2 - .0386z + .1613)}{z^3(z - 1)}$$

The performance index for these two operators, as well as that for the forward difference operator, which is closed-loop realizable, are shown in Table 3.1. The operators were also evaluated in the time domain by using them to solve a system of seven linear differential equations using a fourth-order Runge-Kutta solution as the ideal model. Table 3.2 summarizes the results. The performance index was of the form in (3.41) and (3.42). Reasonable agreement between frequency domain optimization and time domain evaluation is realized for the linear system. This is expected from application of Parseval's Theorem [16].

TABLE 3.1

Frequency Domain Optimization

RMS ERRORS

Operator	J_1	J_2	J_3
ARSO - 1	1.18161 dB	52.9347 degrees	42.5422 degrees
ARSO - 4	2.61323 dB	50.5897 degrees	36.5964 degrees
Forward Diff.	1.76163 dB	52.9347 degrees	42.5422 degrees

TABLE 3.2

Time Domain Evaluation

RMS ERRORS

	ARSO - 1	ARSO - 4	Fwd. Diff.
J1	.217174	.53871	.531912
J2	4.99025	11.4281	11.5529
J3	53.9662	109.322	112.056
J4	319.485	574.329	583.429
J5	1004.97	1550.95	1498.52
J6	1683.85	1872.31	1482.75
J7	1034.9	1014.4	661.88

$$\int_{-\infty} f^2(t)dt = \frac{1}{2\pi j} \int_{-\infty} F(s)F(-s)ds \qquad (3.49)$$

which relates the frequency and time domains. Of course the user should formulate his performance index to reflect what the major concern is. The following transfer function was put in state variable form and used for the evaluation.

$$F(s) = \frac{36}{s^2 + 8.4s + 36} \cdot \frac{s^2 + 2.31 + 2.72}{s^2 + 5.62s + 3.1} \cdot \frac{s + 1.65}{s + .62} \cdot \frac{s^2 + 7.25s + 81}{1.125s^2 + 13.33s + 81} \qquad (3.50)$$

This is the transfer function for an autopilot and therefore is of interest to us.

3.2.4 Further Study

The optimization technique described here will be applied to simulating the dynamic behavior of an aircraft. Initially, motion in a vertical plane will be considered, with more degrees of freedom added as familiarity with the optimization technique is gained. Modifications in the strategy of ARSO, such as individual variances for the random variables, will also be studied.

IV. STUDY OF CHARGE COUPLED DEVICES FOR SIMULATION

4.0 Introduction

The objective of this work was to investigate the feasibility of using CCDs (charge coupled devices) in real-time aircraft simulation. When a single digital computer is used for simulation, the number of calculations made serially is so large that even with relatively large computers the sampling rate cannot be made fast enough to make discretization error negligibly small. For this reason CCD devices were primarily examined to determine if the sampling rate could be made large.

Charge coupled devices are relatively new. The Bell Telephone Laboratories announced the initial discovery of CCDs in May of 1970 [17-19]. A CCD is a semiconductor component [20] which is capable of storing quantities of charge and moving them practically intact from one storage location to another. The quantities of charge represent the magnitude of signal samples. If the movement takes place at regular time intervals, a discrete delay line action is achieved. The addition of signal extraction, weighting, and other manipulations such as summing yield analog data processing of unique potential. Obviously, CCDs can also be used to process digital data.

The CCDs belong to the class of components called charge transfer devices of which the "bucket brigade device" (BBD) [21] is a member. The BBD is older but of less potential value for signal processing than the CCD [22].

For an in-depth review there are several papers available which describe the physical principles of CCDs, how they are constructed, and details about their operational performance [23]. Briefly, they consist of a substrate of doped semiconductor material (such as n-type) on which a very thin layer of insulation is placed. Cells are then formed by depositing metal electrodes on the insulating surface. When the electrodes are biased with respect to the substrate, a potential well is formed in the semiconductor materials. Charge is injected into the device with a PN junction (in the case of n-type substrates). The charge may then be moved

between adjacent cells by appropriately pulsing the associated electrodes. After transfer of charge from a cell, the electrode voltage may be returned to a quiescent value. The transfer between cells is not perfect but is quite adequate for many applications, having an efficiency on the order of 0.9999 to .99999, i.e. the transfer inefficiency ε is on the order of 10^{-4} to 10^{-5} [24]. Thus, a single packet of charge can be moved through many cell locations without obtaining significant signal distortion. The ε depends on clock frequency as one would expect. At the present state of development, operation is usually restricted to less than 10 MHz. Transfer efficiency is important to analog data processing because, unlike digital processing, signal levels may not be re-established. The thermal generation of electron-hole pairs is a serious source of distortion, especially at elevated temperatures. This restricts storage in a cell to times less than one second (usually, one or two orders of magnitude less than one second) at room temperature. Storage time is reduced by a factor of two for every 10°C rise in temperature. This is illustrated later with calculations.

The types of operations performed include all kinds of linear transformations such as correlations, matched filtering, discrete Fourier and Hadamard transforms, Karhounen-Loeve expansions, etc. In addition, fast time simulation can make use of discrete convolution. With some peripheral circuitry, these operations may be implemented by various combinations of the basic building blocks [25] consisting of:

1. Serial In/Serial Out (SI/SO) Registers
2. Serial In/Parallel Out (SI/PO) Registers
3. Parallel In/Serial Out (PI/SO) Registers

The SI/SO Register is the simplest form of CCD processor. It consists of a large linear array of cells forming a discrete analog shift register or delay line. This device is shown in Figure 4.1a.

As shown in Figure 4.1b, the basic SI/PO structure can be used to extract desired data points after a given delay time. If tap weights are

Figure 4.1a Serial In/Serial Out Discrete Analog Delay Line (N Storage Cells)

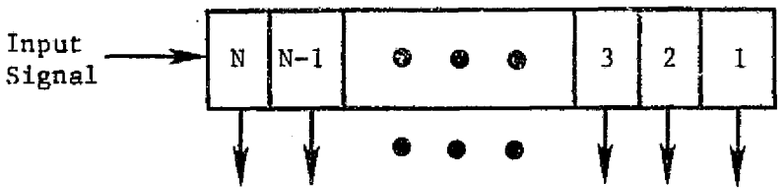


Figure 4.1b Serial In/Parallel Out Structure

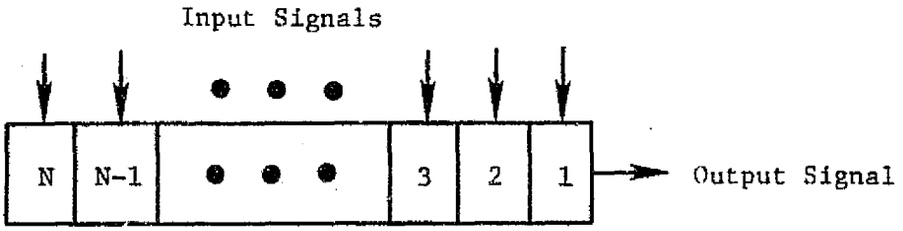


Figure 4.1c Parallel In/Serial Out Structure

used with current summing on the outputs, then it is possible to implement various linear transformations. For example, Figure 4.2 depicts a non-recursive or transversal filter whose output at the nth clock period is the weighted sum of the previous k samples,

$$g(n) = \sum_{i=o}^{k} h_i x_{n-i}, \quad n \geq k \tag{4.1}$$

This is, of course, discrete convolution. If the tap weights h_i in (4.1) define samples of the time inverse of a desired signal, then the CCD processor of Figure 4.2 is a matched filter.

The PI/SO building block is shown in Figure 4.1c. Here, the inputs are synchronously sampled and stored in particular cells of the linear array.

4.1 Simulation Considerations

The most appropriate approach for the simulation being considered is the discrete convolution described by (4.1) which applies to linear stationary systems. Unfortunately, the equations governing aircraft systems are both non-stationary and nonlinear; but for operating conditions where the assumption of small perturbations is valid, the equations can be linearized at each equilibrium condition of interest. This removes the linearity and time-varying restrictions but at the price of either having variable tap weight devices or a large number of devices which can be switched in as required. Practical variable tap weight devices are not available, and development progress appears uncertain so the second approach was assumed.

The idea here is to use a bank of CCD transversal filters, as shown in Figure 4.3, each of which covers one specific equilibrium condition of the system to simulated in order to cover the whole operating range desired in the aircraft simulation.

For small perturbations the motion of an aircraft can be considered to consist of modes with relatively long periods (such as phugoid and spiral divergence) and modes with relatively short periods (such as short-

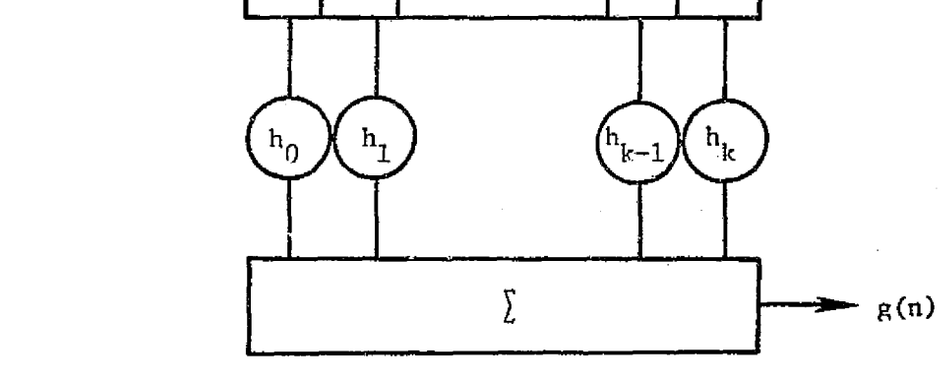


Figure 4.2 A CCD Transversal Filter

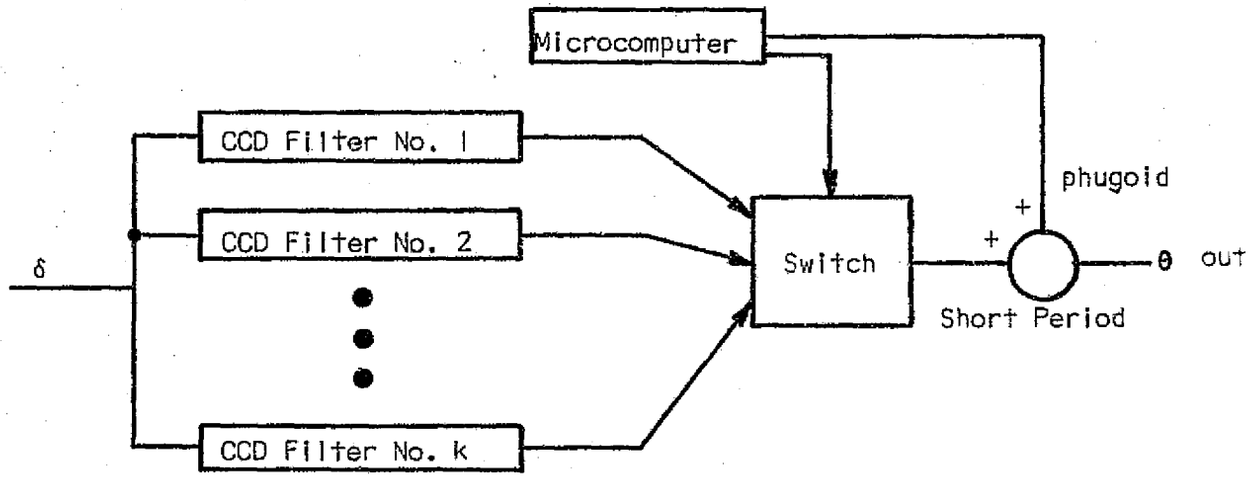


Figure 4.3 Simulation of a Typical Longitudinal Mode of Aircraft with a bank of CCD Transversal Filters.

period and roll subsidence). As will be discussed later, there is a practical limit on the total delay of CCD transversal filters, e.g., 100 m sec at room temperature. Even though much longer delays can be achieved by cooling the devices total delay of more than 10 to 20 sec is not desirable because discretization effect becomes significant due to the required decrease in sampling frequency. This is due to the limited number of stages of a CCD transversal filter. Therefore, the first compromise will be to use a small computer to digitally simulate modes with long periods while using CCD transversal filters for modes with short periods.

4.2 Error Sources

The transfer inefficiency ε per unit cell represents the fractional portion of the signal charge which is left behind when a transfer takes place. At the present, CCDs are constructed with typical values of $\varepsilon \sim 10^{-4}$ to 10^{-5}. The ε remains approximately constant below clock frequencies of $f_c \approx 1$ MHz. Above this ε starts to increase due to the finite transit time of the signal charge from one cell to another. The ε also depends on the quantity of charge in a cell. A d-c bias change called a "fat-zero" is required to make ε less dependent of signal amplitude. The transfer inefficiency affects the CCD operation such that the input signal is dispersed or spread out in time, thus limiting the filter length, i.e. the number of delay stages.

The mechanism for dark current (or leakage current) is the thermal generation of electron-hole pairs. It causes the stroage wells to be slowly filled with minority carriers which gradually mask the stored information. Since the amount of charge is proportional to time, the effect is to limit the low frequency operation of CCD devices. It is found experimentally that total delays up to 100 m sec at room temperature can be achieved before dark current effects become significant [26]. This is true for all devices, irrespective of size or number of elements. However, the dark current is temperature dependence and decreases by a factor of approximately 2 for every 10°C drop in temperature. Thus, dark current can be minimized by sufficient cooling of the CCD. This would be practical for the simulation of large systems.

Tap weight characteristics are a source of simulation error. There is a practical limit on the accuracy of the weighting coefficients in CCD transversal filters. There is also a limitation of the range of maximum to minimum values imposed by practical consideration. This requires that tap weights smaller than the minimum realizable value either be set to the minimum realizable value or zero.

Because the CCD transversal filter has a finite number of elements and the sampling frequency is limited, the total delay that can be achieved with CCD transversal filters is also limited. Thus, non-zero values in a weighting function at instants of time exceeding the total delay have to be truncated. Maximum charge holding capacity, i.e. saturation, can introduce simulation errors. Errors due to linearization have not been studied

4.3 Rough Estimate of Required Number of Filters

The number of transversal filters required is a very important parameter and is considered in the following. Generally, the coefficients of linearized dynamic equations of motion for aircraft systems are determined by the following flight conditions:

(1) Mach number

(2) Angle of attack

(3) Altitude

(4) c.g. location

Suppose the six-degree-of-freedom equations of motion are uncoupled. Then, a typical longitudinal set of equations would have four different excitation inputs: (a) elevator deflection, (b) flap deflection, (c) throttle change, and (d) spoiler deflection. Since longitudinal motion has three-degree-of-freedom and each degree-of-freedom is implemented with a separate set of CCD devices, a total of twelve CCD transvers l filter is necessary for each different equilibrium condition of the longitudinal motion. That means that, in case of coupled motion, approximately eighty CCD devices would be necessary for each different

equilibrium condition for six-degree-freedom simulations. Considering that the four flight conditions can vary independently to cause changes in the coefficients of the equations of motion and considering the wide range in stability derivatives, particularly due to Mach number, altitude and angle of attack, it is not difficult to imagine that thousands of CCD transversal filters would be needed to adequately cover the whole range of interest in ordinary aircraft simulation.

4.3 Future Work

The study was not persued in depth because the first rough estimates of performance for modes with large settling time and the number of CCD section required for realistic flight operation indicated that present technology favors digital technology. Use of microprocessors in parallel operation seems far more promising now than CCDs, so the research has been moved in that direction. Fundamental architecture and software problems are being considered to try to achieve real time simulation that is superior to present large computer systems.

The use of CCDs for a single set of flight conditions may be practical but this question was not addressed because microcomputers offer a possible solution to the more general problem.

Appendix 4.1

Dark Current Considerations

As stated in the text, the achievable total delay of CCD transversal filters is limited, due to dark current. At room temperatue, total delay up to 100 msec can be achieved before dark current effects become significant in many applications.

In order to increase the useful storage time, cooling of CCD devices have been considered. In the following it is shown that, by cooling a CCD device from room temperature (300°K) to 240°K, the dark current is reduced approximately 500 times below than that of room temperature. That means that, theoretically, total delay as large as 50 seconds can be achieved at 240°K before dark current effects become significant.

Dark current has six possible sources [27]:

(1) Generation of carriers via generation-recombination centers in the depletion region.

(2) Generation of carriers via surface states at the $Si-SiO_2$ interface.

(3) Avalanching at the channel-stop boundary.

(4) Diffusion of minority carriers out of the neutral region of the bulk.

(5) Processing errors.

(6) Tunneling between bands.

Among the above six possible sources of dark current, carrier generation in the depletion region (Source 1) and carrier generation via surface states at the silicon-silicon-dioxide interface (Source 2) are the dominant sources [27]. Thus, the dark current, J_D, can be written as

$$J_D = \frac{qn_iW}{2\tau} + \frac{qn_iS_o}{2} = n_i\left(\frac{qW}{2\tau} + \frac{qS_o}{2}\right) \tag{4.2}$$

where

n_i = intrinsic carrier concentration

W = depletion region width

τ = bulk lifetime

S_o = surface recombination velocity when the surface is depleted.

q = charge of an electron

Using the formula

$$S_0 = N_{ss}\pi kT\, V_{th}\, \sigma \tag{4.3}$$

where

N_{ss} = the number of surface states per cm^2 per eV

T = temperature (°k)

σ = capture cross-section for electrons or holes, whichever is smaller

$V_{th} = \sqrt{3kT/m}$ = thermal velocity of the electrons (4.4)

$\tau = C/V_{th}$ (4.5)

with

C = temperature independent constant

yields

$$\frac{qW}{2\tau} \alpha\, V_{th} \alpha \sqrt{T} \tag{4.6}$$

and

$$\frac{qS_o}{2} \alpha\, T \cdot V_{th} \alpha\, T^{3/2} \tag{4.7}$$

The ratios of these values at two different temperatures are

$$\frac{(\frac{qW}{2\tau})_{T = 240°k}}{(\frac{qW}{2\tau})_{T = 300°k}} = \frac{\sqrt{240}}{\sqrt{300}} = 0.9 \qquad (4.8)$$

$$\frac{(\frac{qS_o}{2})_{T = 240°k}}{(\frac{qS_o}{2})_{T = 300°k}} = (\frac{240}{300})^{3/2} = 0.72 \qquad (4.9)$$

Thus,

$$0.72 < \frac{(\frac{qW}{2\tau} + \frac{qS_o}{2})_{T = 240°k}}{(\frac{qW}{2\tau} + \frac{qS_o}{2})_{T = 300°k}} < 0.9 \qquad (4.10)$$

From (4.10) it is observed that the term $(\frac{qW}{2\tau} + \frac{qS_o}{2})$ varies little with temperature.

Now, calculate the change of n_i with temperature change. It is

$$n_i^2 = A_o T^3 \epsilon^{-Eg/kT} \qquad (4.11)$$

where A_o = constant

eg = Energy Gap

k = Boltzman's constant (=8.62 x 10^{-5}eV/°K)

Thus,

$$\frac{(n_i^2)_{T = 240°K}}{(n_i^2)_{T = 300°K}} = \frac{(T^3)_{T = 240°K}}{(T^3)_{T = 300°K}} \cdot \frac{(\epsilon^{-Eg/kT})_{T = 240°K}}{(\epsilon^{-Eg/kT})_{T = 300°K}} \qquad (4.12)$$

so

$$\frac{(T^3)_{T = 240°K}}{(T^3)_{T = 300°K}} = \left(\frac{240}{300}\right)^3 = 0.512 \tag{4.13}$$

and

$$(Eg/kT)_{T = 240°K} = \frac{1.12\ eV}{(8.62 \times 10^{-5} eV/°K)(300°K)} = 55.1$$

$$(Eg/kT)_{T = 300°K} = \frac{1.12\ eV}{(8.62 \times 10^{-5} eV/°K)(300°K)} = 43.31 \tag{4.14}$$

and, it follows that

$$(\varepsilon^{-Eg/kT})_{T = 240°K} = \varepsilon^{-55.1} = 1.176 \times 10^{-24}$$

$$(\varepsilon^{-Eg/kT})_{T = 300°K} = \varepsilon^{-43.31} = 1.551 \times 10^{-19} \tag{4.15}$$

Therefore,

$$\frac{(n_i^2)_{T = 240°K}}{(n_i^2)_{T = 300°K}} = 0.512 \cdot \frac{1.176 \times 10^{24}}{1.551 \times 10^{19}} = 3.88 \times 10^{6} \tag{4.16}$$

and

$$\frac{(n_i)_{T = 240°K}}{(n_i)_{T = 300°K}} = 1.97 \times 10^{-3} \cong (500)^{-1} \tag{4.17}$$

From this calculation it is found theoretically that dark current becomes 500 times smaller at T = 240°K than that at T = 300°K.

REFERENCES

1. Parrish, E. A., McVey, E. S., Cook, G., and Henderson, K., "The Investigations of Optimal Discrete Approximations For Real Time Flight Simulations," Final Technical Report No. EE-4041-102-76, University of Virginia, Charlottesville, Virginia, March 1976.

2. Blakelock, J. H., "Automatic Control of Aircraft and Missles," 1965, pp. 32 and 59.

3. Conte, S. D., and de Boor, Carl, "Elementary Numerical Analysis," Second Edition, 1972.

4. Price, M. G., and Cook, Gerald, "Error-Step Size Trade-Offs for Padé Approximants to Exponentials," submitted to IEEE Trans. on Automatic Control.

5. Beale, G. O., and Cook, Gerald, "Frequency Domain Synthesis of Discrete Representations," accepted, IEEE Trans. on IECI.

6. Steiglitz, K., "Computer Aided Design of Recursive Digital Filters," IEEE Trans. Audio Electroacoustic, Vol. AU-18, No. 2 (June 1970), pp. 123-129.

7. Fletcher, R., and Powell, M. J. D., "A Rapidly Convergent Descent Method for Minimization," Computer Jnl., Vol. 6, No. 2 (1963), pp. 163-68.

8. Schroeder, D., "A New Optimization Procedure for Digital Simulation," Ph.D. Dissertation, University of New Mexico, 1972.

9. Sage, A. P., "A Technique for the Real Time Digital Simulation of Non-Linear Control Processes," Proc. of Region III IEEE Conf., April 1966.

10. Sage, A. P., and Smith, S. L., "Real Time Digital Simulation for Systems Control," Proc. of the IEEE, Vol. 54, No. 12 (December 1966).

11. Kuo, B. C., "Analysis and Synthesis of Sampled Data Control Systems," Prentice-Hall, 1963.

12. Beck, M. S., "Adaptive Control-- Fundamental Aspects and Their Application," Proc. of 1st Anl. Advanced Control Conf. (Dun-Donnelly Publishing, Purdue University, April 29-May 1, 1974).

13. Andrews, M., and Korn, G. A., "Hybrid Computer Methods for Direct Functional Optimization," IEEE Trans. on Computers, Vol. c-24, No. 10 (October 1975).

14. Korn, G. A., and Kosako, H., "A Proposed Hybrid Computer Method for Functional Optimization," IEEE Trans. on Computers (February 1970).

15. Schumer, M. A., and Steiglitz, K., "Adaptive Step Size Random Search," IEEE Trans. on Automatic Controls, Vol. AC-13, No. 3 (June 1968).

16. Eveleigh, V. W., Adaptive Control and Optimization Techniques, McGraw-Hill, 1967.

17. Electronics (May 11, 1970), p. 112.

18. Boyle, W. J., and Smith, G. E., "Charge-Coupled Semiconductor Devices," Bell Syst. Tech. J., Vol. 49 (1970), pp. 487-93.

19. Amilio, G. F., et al., "Experimental Verification of the Charge-Coupled Concept," Bell Syst. Tech. J., Vol. 49 (1970), p. 593.

20. Kosonocky, W. F., "Charge-Coupled Devices -- An Overview," 1974 WESCON Professional Program, Session 2, Los Angeles, California, September 1974.

21. Butler, W. J., et al., "Practical Considerations for Analog Operation of Bucket Brigade," IEEE J. Solid State Circuits, SC-6 (December 1971), p. 391-95.

22. Buss, D. D., et al., "Charge Transfer Devide Transversal Filters for Communication Systems," submitted to IEEE Trans. on Comm.

23. Boyle, W. J., and Smith, G. E., "Charge-Coupled Devices -- A New Approach to MIS Device Structures," IEEE Spec. (July 1971), pp. 18-27.

24. White, M. H., and Webb, W. R., "Study of the Use of Charge-Coupled Devices in Analog Signal Processing Systems," Final report, Contract No. N00014-74-C-0069, Westinghouse Defense and Electronic System Center, May 1974.

25. Lampe, D. R., "CCD's for Discrete Analog Signal Processing (DASP)," IEEE Intercon, New York, March 1974.

26. Burt, D. J., "Charge-Coupled Devices and Their Applications," Electonics and Power (February 1975), pp. 93-97.

27. Tasch, A. F., Jr., et al., "Dark-Current and Storage-Time Considerations in Charge-Coupled-Devices," Proc. CCD Applications Conference, Naval Elec. Lab. Center, San Diego, California, September 18-20, 1973, pp 179-87.