Aeronautical Research Institute, University of Tokyo
Report No. 360, December 1960

Statistical Design of Control Systems Containing Digital Devices†

By

Yasushi ISHII

Summary. Sampling and quantization of signals are essential to the digital devices such as digital computers and analog-digital converters. This paper develops statistical optimum design methods of the control systems containing these digital devices under random inputs.

In the first part, neglecting the quantization effects, the pulse transfer function of an ordinary sampled-data control system, which minimizes the mean square error, is presented as a function of the spectral density of the reference input signal, under several constraints which come from the physical realizability of the system, time constants and dead time of the controlled process, etc. Although quantizing operation is nonlinear, in statistical treatments it can be linealized as a superposition of a white noise on the signal, and the corrections to the optimum design due to quantization are discussed theoretically.

In the later part of this paper, a high speed electronic simulator of sampled-data systems, which was designed and constructed for this study, is described. To examine the theory, experiments of the statistical optimum design were performed using this simulator for several kinds of random inputs. These results are compared with those numerically calculated.

1. INTRODUCTION

In the recent field of automatic control, there have been more and more demands for the high accuracy of control and the complex manipulation of information, and to meet these demands the applications of some digital devices operating in real time—such as analog-digital converters (A-D converters), digital data transmission links, digital computers connected directly to control systems—to this field have become important.

One of the distinctive features of the system containing these digital devices is that at one or more points in the system, the signals consist of sequences of numbers which are discrete in the time domain, rather than are conveyed by continuous quantities. That is, sampling operations of the signals are involved in the system and the information is transmitted intermittently. In an A-D converter based on the time modulation principle the iterative counting of the number of pulses necesitates the sampling of the signals to be converted. In a digital computer, the inputs and outputs must be intermittent, because a computing

† This paper was previously published in Japanese on the "Automatic Control" Vol. 6, No. 5 (1959, Tokyo).

operation needs finite time and is not able to be performed continuously.

Another important feature is quantization of signals, that is, in order to convert the analogous signals which are represented by physical quantities such as voltage, current, pressure, position, etc. to numbers or other equivalent cords, the amplitude scale of the signals must be divided into a number of sections or quanta and any signal value which falls somewhere within a section must be represented by a number corresponding to some value within the section, ordinarily the center value, because numbers are discrete in the signal amplitude domain as well as in the time domain. For example, if a digital voltmeter whose quantum has the magnitude of 1V delivers a digital representation of 37V as its output, the input voltage would be somewhere within the range of 36.5V$\sim$37.5V. Thus, in the control systems containing the digital devices, there are two kinds of discontinuities of the signals, one is due to sampling in the time domain and the other is due to quantization in the signal amplitude domain.

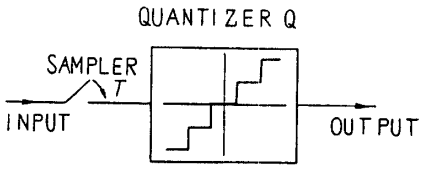

FIGURE 1. Block diagram of an A-D converter.

Although digital signals represented by numbers are substantially different from analogous signals which are represented by physical quantities, in the theoretical treatments of the control systems there is no need to make a distinction between these two forms of signals. For instance, an A-D converter may be considered as a combination of a sampler or switch which closes momentarily every T seconds and a quantizer having a staircase input-output relation as symbolically shown in Fig. 1. The input to the A-D converter is in an analog form and the output is in a digital form, but it is not necessary to make distinction between the forms of the signals so far as the converter is considered as Fig. 1 and treated as an element in the block diagram of the control system. However, it should be noted here that the sampling is a linear operation, while the quantizer has nonlinear characteristic.

In this paper the statistical optimum design methods of the control systems containing the digital devices under random inputs are studied theoretically and experimentally. In the first part, neglecting the quantization of signals, the system is considered as an ordinary sampled-data control system and the pulse transfer function which minimizes the mean square error is derived as a function of the spectral density of the input signal, under several constraints owing to the physical realizability of the system, the time constants and the dead time of the controlled process, etc. Although the quantizing operation is nonlinear as already mentioned, it can be linealized as a superposition of a kind of white noise or quantization noise and the corrections due to this noise to the optimum design are discussed theoretically.

In the later part of the paper, a high speed electronic simulator of sampled-data systems, which was designed and constructed for this study is described. To examine the theory, experiments of the statistical optimum design were performed using this simulator for several kinds of random inputs, and the results are shown with the numerically calculated ones.

2. Optimum Design Theory for Ordinary Sampled-Data Control Systems

Of the two features of digital devices mentioned in the preceding section, only the sampling operation will be considered in this section, and optimum design theory for linear sampled-data control systems such as shown in Fig. 2 [*1*] will be developed. On the generalized form of sampled-data control systems shown in

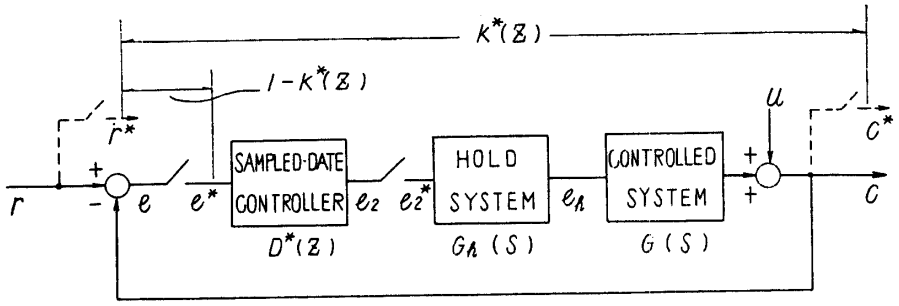

FIGURE 2. Typical sampled-data control system.

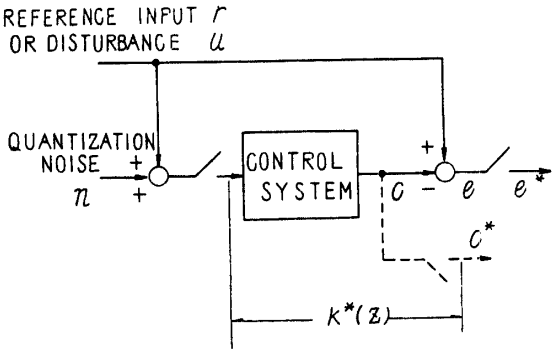

FIGURE 3. A reduced form of the block diagram of Fig. 2.

Fig. 3, the problem to be solved is described as to find the pulse transfer function $K^*(z)$ of the whole control system that minimizes the control error $e(t)$ under stationary random input $r(t)$. The magnitude of the control error is represented by the mean square value of that signal at sampling instants, that is,

$$\overline{\{e(nT)\}^2}=\lim_{N\to\infty}\frac{1}{2N+1}\sum_{n=-N}^{N}\{e(nT)\}^2 \tag{1}$$

where n and N are integers and T is the sampling period. In Fig. 3,

$$e^*(t)=r^*(t)-c^*(t) \tag{2}$$

and

$$E^*(z)=\{1-K^*(z)\}R^*(z) \tag{3}$$

where $E^*(z)$ and $R^*(z)$ are z-transforms of $e(t)$ and $r(t)$ respectively. Therefore, denoting the pulse spectral density of the input as $\Phi_r^*(z)$†, the mean square value of the control error is represented as [2][3]

$$\overline{[e(nT)]^2}=\frac{1}{2\pi jT}\oint_{\substack{\text{unit}\\ \text{circle}}}\{1-K^*(z)\}\{1-K^*(z^{-1})\}z^{-1}\Phi_r^*(z)dz \qquad (4)$$

where $z=e^{j\omega t}$ and ω represents angular frequency. Thus, the problem is reduced to determine $K^*(z)$ which minimizes the value of Eq. (4) under several constraints imposed on $K^*(z)$.

The first of the constraints is

(1) the physical realizability of $K^*(z)$.

Since $K^*(z)$ is the pulse transfer function of a stable control system, the weighting sequence $k(nT)$ of the control system or inverse z-transform of $K^*(z)$ is given as

$$k(nT)=\frac{1}{2\pi j}\oint_{\substack{\text{unit}\\ \text{circle}}}K^*(z)z^{n-1}dz\dagger\dagger \qquad (5)$$

Assuming the integration contour of Eq. (5) as shown in Fig. 4 and taking into account the condition that $k(nT)=0$ for $n<0$ in a physical realizable system, we get the statement that

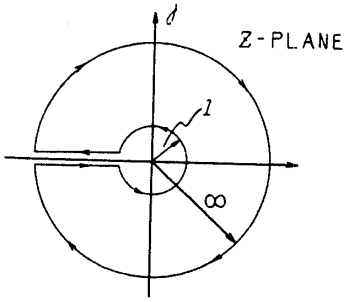

FIGURE 4. The integration contour.

$$\left.\begin{array}{l}|K^*(z)|<\infty \text{ when } z\to\infty \text{ and}\\ K^*(z) \text{ should not have any poles outside the unit circle.}\end{array}\right\} \qquad (6)$$

However, the physical realizability does not impose any significant constraint on the system by itself, because the optimum physical realizable pulse transfer function of the whole control system is always 1 for any random input. Hence, as the second constraint we take that

(2) the maximum order of z in the descending power series expansion of $K^*(z)$ is $-m$.

---

† $\Phi_r^*(z)=T\sum_{k=-\infty}^{\infty}\phi_r^*(k)z^{-k}$, $\phi_r^*(k)=\lim_{N\to\infty}\frac{1}{2N+1}\sum_{n=-N}^{N}r(nT)r\{(n+k)T\}$, k: an integer

†† In this paper, the pulse transfer function $K^*(z)$ is defined as the two-sided z-transform of the weighting sequence of the system, that is

$$K^*(z)=\sum_{n=-\infty}^{\infty}k(nT)z^{-n}$$

The inverse z-transform of Eq. (5) can be derived from the inverse Fourier transform just in the same manner for ordinary z-transforms [4]. The poles of $K^*(z)$ lying outside the unit circle correspond to the portion of $k(nT)$ for $n<0$.

This statement means that the controlled system shown in Fig. 2 has a dead time of the magnitude

$$(m-1)T+\delta T \tag{7}$$

(m: a positive integer, δ: zero or positive fraction less than unity)

and the response of the controlled variable $c(t)$ for a change of $r(t)$ at a sampling instant is delayed by m sampling periods. Even in the case where the controlled system has no dead time, the time delay of one sample is enevitable because of the energy reserving properties of the controlled system. In practice it is impossible to control the system with no time delay, since if it were so, the manipulated variable would contain impulses and infinite power rate would be required. Thus, the general form of $K^*(z)$ should be

$$K^*(z)=\frac{k_m z^{-m}+k_{m+1}z^{-(m+1)}+\cdots}{1+l_1 z^{-1}+l_2 z^{-2}+\cdots} \tag{8}$$

(k's, l's: constant coefficients)

This is the mathematical expression of the constraint (2).

Eq. (4) which gives the mean square value of the control error in the error-sampled system as shown in Fig. 2 is also applicable to the other important types of sampled-data control systems, such as tele-control systems including PCM (Pulse Code Modulation) data transmission links [5] and regulating systems using A-D converters as primary means. Moreover, replacing $\Phi_r^*(z)$ in Eq. (4) by $\Phi_u^*(z)$, the pulse spectral density of the stationary random disturbance $u(t)$ in Fig. 2, the equation gives the mean square value of the control error caused by the disturbance, because

$$E^*(z)=-\{1-K^*(z)\}U^*(z) \tag{9}$$

Thus the disturbance may be considered as an equivalent to the reference input. Hence, in the following parts of this paper the statistical optimum design method based on Eq. (4) will be developed.

In order to determine the form of $K^*(z)$ which minimizes the value of Eq. (4) under the constraints (1) and (2), we eliminate the constraint (2) at a first place by setting

$$K^*(z)=z^{-m}F^*(z) \tag{10}$$

where $F^*(z)$ is the pule transfer function of any physical realizable system, and the optimum form of which is now to be determined. Substituting Eq. (10) in Eq. (4), we get

$$\overline{\{e(nT)\}^2}=\frac{1}{2\pi jT}\oint\{z^m-F^*(z)\}\{z^{-m}-F^*(z^{-1})\}z^{-1}\Phi_r^*(z)dz \tag{11}$$

To determine the optimum form of $F^*(z)$ the method of calculus of variation is employed, that is, if the mean square value of the control error $\overline{e^2}$ is increased to $\overline{e^2}+\Delta\overline{e^2}$ when $F^*(z)$ is replaced by $F^*(z)+\lambda\Delta F^*(z)$, where λ is a parameter and $\Delta F^*(z)$ is an arbitrary function of the form of $F^*(z)$, then

$$\overline{e^2}+\Delta\overline{e^2}=\frac{1}{2\pi jT}\oint[\{z^m-F^*(z)\}-\lambda\Delta F^*(z)][\{z^{-m}-F^*(z^{-1})\}-\lambda\Delta F^*(z^{-1})]z^{-1}\Phi_r^*(z)dz \tag{12}$$

Assuming an optimum $F^*(z)$ that minimizes $\overline{\{e(nT)\}^2}$, the first derivative of $\overline{\Delta e^2}$ with respect to λ should be zero when $\lambda=0$, namely,

$$\left(\frac{\partial \overline{\Delta e^2}}{\partial \lambda}\right)_{\lambda=0}=\frac{-1}{2\pi jT}\oint\{z^m-F^*(z)\}\Delta F^*(z^{-1})z^{-1}\Phi_r^*(z)dz$$
$$+\frac{-1}{2\pi jT}\oint\{z^{-m}-F^*(z^{-1})\}\Delta F^*(z)z^{-1}\Phi_r^*(z)dz=0 \tag{13}$$

The integrals in Eq. (13) must always be zero respectively, since $\Delta F^*(z)$ is assumed as an arbitrary function. Hence, for the first integral in Eq. (13) we have the following conditions to be satisfied;

$$\left.\begin{array}{l}|\{z^m-F^*(z)\}\Phi_r^*(z)|\to 0 \text{ when } z\to 0 \text{ and}\\ \{z^m-F^*(z)\}\Phi_r^*(z) \text{ has no poles inside the unit circle.}\end{array}\right\} \tag{14}$$

The second integral becomes zero automatically, when the above conditions for the first integral is satisfied. Then, factoring the pulse spectral density of the input into two components as

$$\Phi_r^*(z)=\overset{+}{\Phi}{}_r^*(z)\cdot\overset{-}{\Phi}{}_r^*(z) \tag{15}$$

where

$$\left.\begin{array}{l}\overset{+}{\Phi}{}_r^*(z) \text{ includes all poles and zeros in the unit circle}\\ \overset{-}{\Phi}{}_r^*(z) \text{ includes all poles and zeros outside the unit circle}\end{array}\right\} \tag{16}$$

and taking into account the physical realizability of $F^*(z)$ as stated in Eq. (6), the optimum form of $F^*(z)$ is determined as

$$F^*(z)_{\text{opt}}=\frac{1}{\overset{+}{\Phi}{}_r^*(z)}\{f_0+f_1z^{-1}+f_2z^{-2}+\cdots\} \tag{17}$$

where f's are the coefficients in the descending power series expansion of $z^m\overset{+}{\Phi}{}_r^*(z)$ shown as

$$z^m\overset{+}{\Phi}{}_r^*(z)=f_{-m}z^m+f_{-(m-1)}z^{(m-1)}+\cdots+f_0+f_1z^{-1}+f_2z^{-2}+\cdots \tag{18}$$

The optimum pulse transfer function of the whole control system, which minimizes the mean square valne of the control error, is obtained by multiplying $F^*(z)_{\text{opt}}$ by z^{-m}, that is,

$$K^*(z)_{\text{opt}}=z^{-m}F^*(z)_{\text{opt}}$$
$$=\frac{z^{-m}}{\overset{+}{\Phi}{}_r^*(z)}\{f_0+f_1z^{-1}+f_2z^{-2}+\cdots\} \tag{19}$$

Substitution of Eq. (19) in Eq. (4) gives the minimum value of the control error as

$$\overline{\{e(nT)\}^2}_{\min}=\frac{1}{T}\{f_{-m}^2+f_{-(m-1)}^2+\cdots+f_{-1}^2\} \tag{20}$$

Eq. (19) and Eq. (20) are the conclusions of this section.†

It is emphasized here that $F^*(z)_{\text{opt}}$ is nothing other than the pulse transfer function of a optimum predictor which is physically realizable and z^m in Eq. (11) is that of the ideal predictor. These facts are interpreted as a consequences of

† Recently, a similar theory was presented by S. S. L. Chang [6] in different forms.

the time delay of m samples in the response of controlled variable $c(t)$ to the input signal $r(t)$, that is, the best way which is able to be accomplished by the control system under these situations is to predict the future value of the input signal at the sampling instant of m samples later and to force the controlled variable to have that predicted value at that sampling instant.

3. Optimum Design Theory for the System Having Finite Settling Time Response

In a system employing a sampled-data controller, it is possible to cause the system to have a finite settling time so far as the values at sampling instants are concerned [*1*]. Such a response is not seen in ordinary continuously controlled systems and is one of the distinguishing marks of sampled-data control systems. The pulse transfer function of the system having finite settling time response is expressed as

$$K^*(z)=k_m z^{-m}+k_{m+1}z^{-(m+1)}+\cdots+k_{m+p}z^{-(m+p)} \tag{21}$$

which shows the response of the controlled variable to the input signal is delayed by m samples and settled in $(p+1)$ sampling periods. If it is also assumed that the system has no off-set for step input, the coefficients in Eq. (21) must take the following relation;

$$k_m+k_{m+1}+\cdots+k_{m+p}=1 \tag{22}$$

In general, the requirement that the system follows a test function of the form of t^μ without steady state error places $(\mu+1)$ equations to be satified by the coefficients of $K^*(z)$, thus decreasing the degrees of freedom for statistical optimum design of $K^*(z)$. In addition to the constraints (1) and (2) mentioned in the preceding section, the two additional restrictions as shown by Eq. (21) and (22), that is,

(3) the finite settling time response of the control system and

(4) the requirement for zero off-set for step input,

will be considered in this section and the statistical optimum design method under these four constraints will be derived.

The integrand

$$\{1-K^*(z)\}\{1-K^*(z^{-1})\}\Phi_r^*(z)=\Phi_e^*(z) \tag{23}$$

in Eq. (4) which gives the mean square error of the basic sampled-data control system of Fig. 3, is no more than the pulse spectral density of the control error $e(t)$. Therefore, taking the expansion forms of $\Phi_r^*(z)$, $\Phi_e^*(z)$ by their pulse correlation functions $\phi_r^*(k)$, $\phi_e^*(k)$ [*2*][*3*] (k: an integer, cf. the footnote on p. 218)

$$\Phi_r^*(z)=T\{\cdots+\phi_r^*(2)z^2+\phi_r^*(1)z+\phi_r^*(0)+\phi_r^*(1)z^{-1}+\phi_r^*(2)z^{-2}+\cdots\} \tag{24}$$

$$\Phi_e^*(z)=T\{\cdots+\phi_e^*(2)z^2+\phi_e^*(1)z+\phi_e^*(0)+\phi_e^*(1)z^{-1}+\phi_e^*(2)z^{-2}+\cdots\} \tag{25}$$

and substituting Eq. (21), (24) and (25) into Eq. (23) to compare the z^0 terms in both sides of the equation; we get the relation

$$\overline{\{e(nT)\}^2}=\phi_e^*(0)$$
$$=\left[1+\sum_{i=m}^{m+p}k_i^2\right]\phi_r^*(0)+2\left[\sum_{j=1}^{m+p}\phi_r^*(j)\left\{\sum_{i=m}^{m+p-j}k_ik_{i+j}-k_j\right\}\right] \tag{26}$$

(i, j: integers, k_i, k_{i+j}: coefficients in Eq. (21))

Thus, the mean square error $\overline{\{e(nT)\}^2}$ is represented by the coefficients $k_m, k_{m+1}, \cdots, k_{m+p}$. If Eq. (26) is differentiated with respect to k's and equated to zero, we have a set of equations

$$\frac{\partial\phi_e^*(0)}{\partial k_i}=0 \tag{27}$$
$$(i=m, m+1, \cdots, m+p)$$

from which we can determine the values of k's that minimize $\overline{\{e(nT)\}^2}$ under the constraints (1), (2) and (3). Similarly, if the values of k's are determined using Eq. (21) and (22) simultaneously, the optimum values of k's under the constraints (1), (2), (3) and (4) can be obtained.

For examples, a sampled-data control system which has the indicial response of the form as shown in Fig. 5 will be considered. The pulse transfer function of the system is

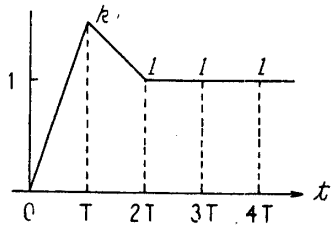

FIGURE 5. Indicial response of the system given by Eq. (28).

$$\left.\begin{aligned}&K^*(z)=k_1z^{-1}+k_2z^{-2}\\&k_1+k_2=1\end{aligned}\right\} \tag{28}$$

From Eq. (26)

$$\overline{\{e(nT)\}^2}=(1+k_1^2+k_2^2)\phi_0+2k_1(k_2-1)\phi_1-2k_2\phi_2 \tag{29}$$

where $\phi_i=\phi_r^*(i)$. Substituting $k_2=1-k_1$ in Eq. (29) and setting $d\overline{\{e(nT)\}^2}/dk_1=0$, we have the following expressions for the optimum values of the coefficients.

$$\left.\begin{aligned}&(k_1)_{\text{opt}}=\frac{1}{2}\left(1+\frac{\phi_1-\phi_2}{\phi_0-\phi_1}\right)\\&(k_2)_{\text{opt}}=1-(k_1)_{\text{opt}}=\frac{1}{2}\left(1-\frac{\phi_1-\phi_2}{\phi_0-\phi_1}\right)\end{aligned}\right\} \tag{30}$$

Substituting Eq. (30) in Eq. (29) the mean square value of the control error, when the coefficients are set optimum, is readily calculated as

$$\overline{\{e(nT)\}^2}_{\min}=\frac{1}{2}\frac{1}{\phi_0-\phi_1}\{(\phi_0-\phi_1)+(\phi_1-\phi_2)\}\{3(\phi_0-\phi_1)-(\phi_1-\phi_2)\} \tag{31}$$

For the sampled-data control system which has a indicial response of the form as shown in Fig. 6,

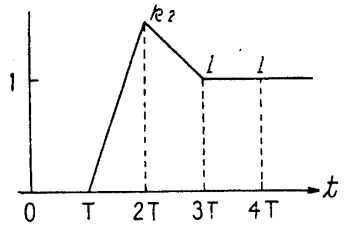

FIGURE 6. Indicial response of the system given by Eq. (32).

$$\left.\begin{aligned} K^*(z)&=k_2z^{-2}+k_3z^{-2} \\ k_2+k_3&=1 \end{aligned}\right\} \tag{32}$$

we get the following results through the similar procedure.

$$\left.\begin{aligned} (k_2)_{\text{opt}}&=\frac{1}{2}\left(1+\frac{\phi_2-\phi_3}{\phi_0-\phi_1}\right) \\ (k_3)_{\text{opt}}&=1-(k_2)_{\text{opt}}=\frac{1}{2}\left(1-\frac{\phi_2-\phi_3}{\phi_0-\phi_1}\right) \end{aligned}\right\} \tag{33}$$

$$\begin{aligned} &\overline{\{e(nT)\}^2}_{\min} \\ &=\frac{1}{2}\frac{1}{\phi_0-\phi_1}\{(\phi_0-\phi_1)+(\phi_2-\phi_3)\}\{3(\phi_0-\phi_1)-(\phi_2-\phi_3)\}+2(\phi_1-\phi_2) \end{aligned} \tag{34}$$

It is worth while to compare the optimum pulse transfer function given by Eq. (19) in Sec. 2, which is designated here as $K_1^*(z)_{\text{opt}}$, with that derived in this section which is designated as $K_2^*(z)_{\text{opt}}$. $K_1^*(z)_{\text{opt}}$ is optimum for a random input when the constraints (1) and (2) are imposed and $K_2^*(z)_{\text{opt}}$ is the optimum pulse

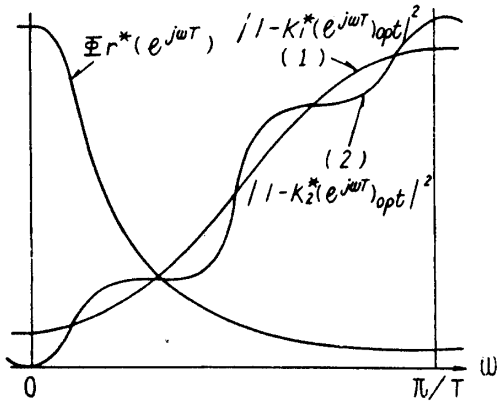

FIGURE 7. Pulse spectral density of the input signal and frequency characteristics of the systems.

transfer function for the same input under the constraints (1), (2), (3) and (4). In the limiting case where the limitation on the settling time of the control system is made to infinitely long, that is where the constraint (3) is removed, $K_2^*(z)_{\text{opt}}$ converges to $K_1^*(z)_{\text{opt}}$ irrespective of the constraint (4). On the rewritten form of Eq. (4)

$$\overline{\{e(nT)\}^2}=\frac{1}{2\pi}\int_{-\pi/T}^{\pi/T}|1-K^*(e^{j\omega T})|^2\Phi_r^*(e^{j\omega T})d\omega \tag{35}$$

the curve of $|1-K_2^*(e^{j\omega t})|^2$, which is shown in Fig. 7, approaches to the curve of $|1-K_1^*(e^{j\omega t})|^2$ closer and closer when the settling time of the system is made longer, and finally the curve (2) coinsides with curve (1) except the points where $\omega=0,\ \pm 2\pi/T,\ \pm 4\pi/T, \cdots (|1-K_2^*(e^{j\omega t})|^2$ should be zero at these points by Eq. (22)). From the preceding discussions we can conclude that the condition (4) can not be a constraint by itself, instead, it can be effective only when combined with condition (3), that is, when the off-set of the control system vanishes within finite settling time.

4. The Effect of Quantization on the Statistical Design

As already mentioned in Sec. 1, an A-D converter may be considered as a combination of a sampler and a quantizer which has a staircase input-output relation. The statistical properties of the quantization error or "round off" error, which comes from the representation of signals by numbers, have been studied as

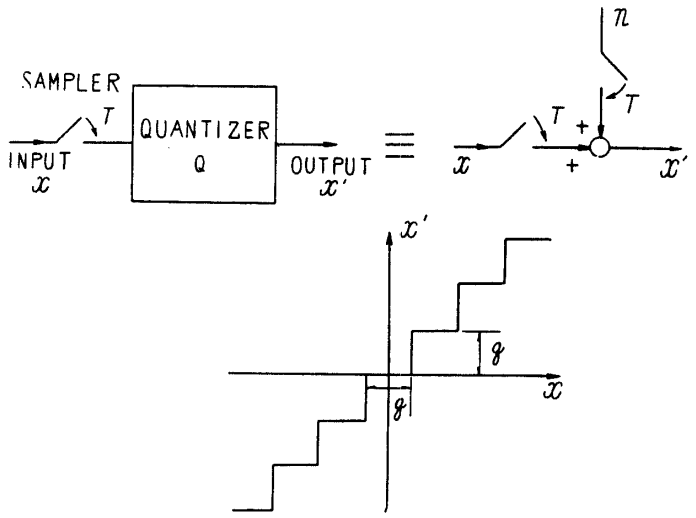

FIGURE 8. Linealization of a quantizer.

problems in PCM communication systems. According to W. R. Bennett [7] and B. Widrow [8], if the width of the quanta is small comparing with the amplitude of the signal to be quantized, the statistical effect of quantizing operation can be approximated as a superposition of a quantization noise, which is statistically independent of the signal to be converted, on that signal. The pulse correlation function $\phi_n^*(k)$ and pulse spectral density $\Phi_n^*(z)$ of this noise are given as

$$\phi_n^*(k)\begin{cases}=0 : k\neq 0\\ =A : k=0\end{cases} \tag{36}$$

and

$$\Phi_n^*(z)=TA \tag{37}$$

where $A=q^2/12$ and q is the width of the quanta. This approximation is especially good when the amplitude distribution of the input signal of the quantizer is expressed by a smooth curve as in Gaussian distribution, and well be applied to an

extremely rough quantization such as for the case where q equals to two standard deviations of the amplitude distribution of the input signal. Thus, the effect of quantization to the statistical design can be treated as addition of fictitious noise to the system input.

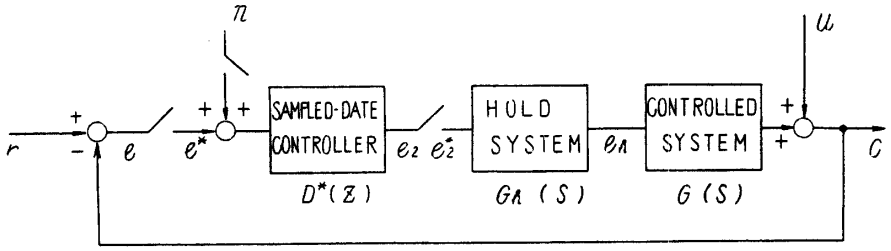

FIGURE 9. The block diagram of a sampled-data control system; in which the A-D converter is replaced by the superposion of the quantization noise.

In the control system of Fig. 2, if the sampled-data controller is a digitally operating one, an A-D converter must precede the controller. Replacement of the A-D converter by a source of quantization noise results the block diagram of Fig. 9. Since the quantization noise $n(t)$ is independent of the reference input $r(t)$ and the pulse transfer function from the noise $n(t)$ to the control error $e(t)$ is $K^*(z)$, the mean square value of the control error is composed of two components, one is due to the reference input and one to the quantization noise, as the following equation.†

$$\overline{\{e(nT)\}^2} = \frac{1}{2\pi jT} \oint_{\substack{\text{unit} \\ \text{circle}}} [\{1-K^*(z)\}\{1-K^*(z^{-1})\}\Phi_r^*(z)+K^*(z)K^*(z^{-1})\Phi_n^*(z)]z^{-1}dz \qquad (38)$$

The above equation indicates that the system of Fig. 9 can be further reduced to Fig. 10.

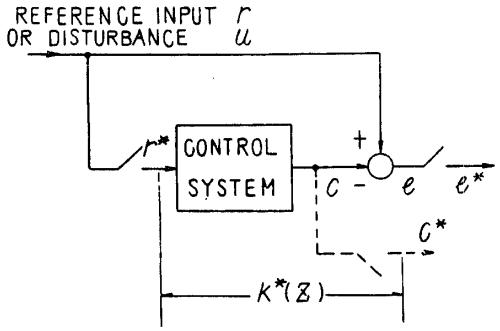

FIGURE 10. A reduced form of the block diagram of Fig. 9.

† In references [7] and [8], the apploximation of a quantizer by a quantization noise source is valid only for the statistical characteristics of the input and output signals of the quantizer. Hence it is not a regorous manner to assume a pulse transfer function from the noise to the control error and evaluate the mean square value of the control error. However, it is possible to derive Eq. (38) using the statistical input-output relationships of the quantizer and the pulse transfer functions of the linear portion of the system, if only one quantizer is included in the control system and the remaining portion is linear.

The pulse transfer function $K^*(z)$ which minimizes $\overline{\{e(nT)\}^2}$ in Eq. (38) under the four constraints stated in the preceding section, can easily be obtained by substiting Eq. (37) in Eq. (38) and following the procedure in Sec. 3, getting the equation

$$\overline{\{e(nT)\}^2}=\left[1+\sum_{i=m}^{m+p}k_i^2\right][\phi_r^*(0)+A]+2\left[\sum_{j=1}^{m+p}\phi_r^*(j)\left\{\sum_{i=m}^{m+p-j}k_ik_{i+j}-k_j\right\}\right]-A \tag{39}$$

corresponding to Eq. (26) which is derived under the condition that the quantization of signals is neglected. Comparison of these two equations shows that they both are in the same form except the constant term $-A$ in Eq. (39). Hence the optimum parameters in $K^*(z)$ when a quantizer is included in the system can be obtained by substituting

$$\phi_0'=\phi_r^*(0)+A \tag{40}$$

for ϕ_0 in Eq. (30) or Eq. (33).†† In practice, the width q of the quantization box is very small comparing with the amplitude of the system input $r(t)$ and the correction is scarcely needed. However, the increment of the control error due to quantization when no correction is made, that is

$$\frac{1}{2\pi jT}\oint K^*(z)K^*(z^{-1})z^{-1}ATdz \tag{41}$$

in Eq. (38) or

$$\left[\sum_{i=m}^{m+p}k_i^2\right]A \tag{42}$$

in Eq. (39) would not be neglected compared with the portion of the control error due to the system input $r(t)$.

The foregoing discussion is an example of the quantization effect on a linear sampled-data control system in which only one quantizer is included. For the general case where n quantizers are included in the control system, it is not always possible to analyse the effect of the quantization by assuming the n noise sources, which are statistically independent of each other, in places of the quantizers. In practice, however, many important types of sampled-data control systems can be reduced to the form of Fig. 10 when the quantization effect is taken into consideration, and therefore the method stated in this section may be applied.

5. Calculation of the Optimum Design and its Comparison with Experimental Results

An example of the statistical design method stated in Secs. 2 and 3 will be illustrated in this section. The assumed auto-correlation function and spectral density of the random input are as follows:

$$\phi_r(\tau)=P_0e^{-a|\tau|}\cos b\tau \tag{43}$$

$$\begin{aligned}\Phi_r(\omega)&=\int_{-\infty}^{\infty}\phi_r(\tau)e^{-j\omega\tau}d\tau\\&=\frac{2P_0}{a}\cdot\frac{1+(b/a)^2+(\omega/a)^2}{\left\{1+\left(\frac{b+\omega}{a}\right)^2\right\}\left\{1+\left(\frac{b-\omega}{a}\right)^2\right\}}\end{aligned} \tag{44}$$

†† For Eq. (19) in Sec. 2, we can also get the correction due to the quantization of the signals.

Especially when $b=0$ in Eq. (43) and (44), they are reduced to

$$\phi_r(\tau)=P_0e^{-a|\tau|} \tag{45}$$

$$\Phi_r(\omega)=\frac{2P_0}{a}\cdot\frac{1}{1+\left(\frac{\omega}{a}\right)^2} \tag{46}$$

These are shown graphycally in Fig. 11. A reason for assuming the auto-correlation functions as Eq. (43) or Eq. (45) is that such kinds of random signals may be obtained by passing the random signal having a flat spectral density i.e. white noise, through first or second order resistance-capacitance combination systems, and many kinds of auto-correlation function of random signals encountered in practice may well be approximated by Eq. (43) or Eq. (45).

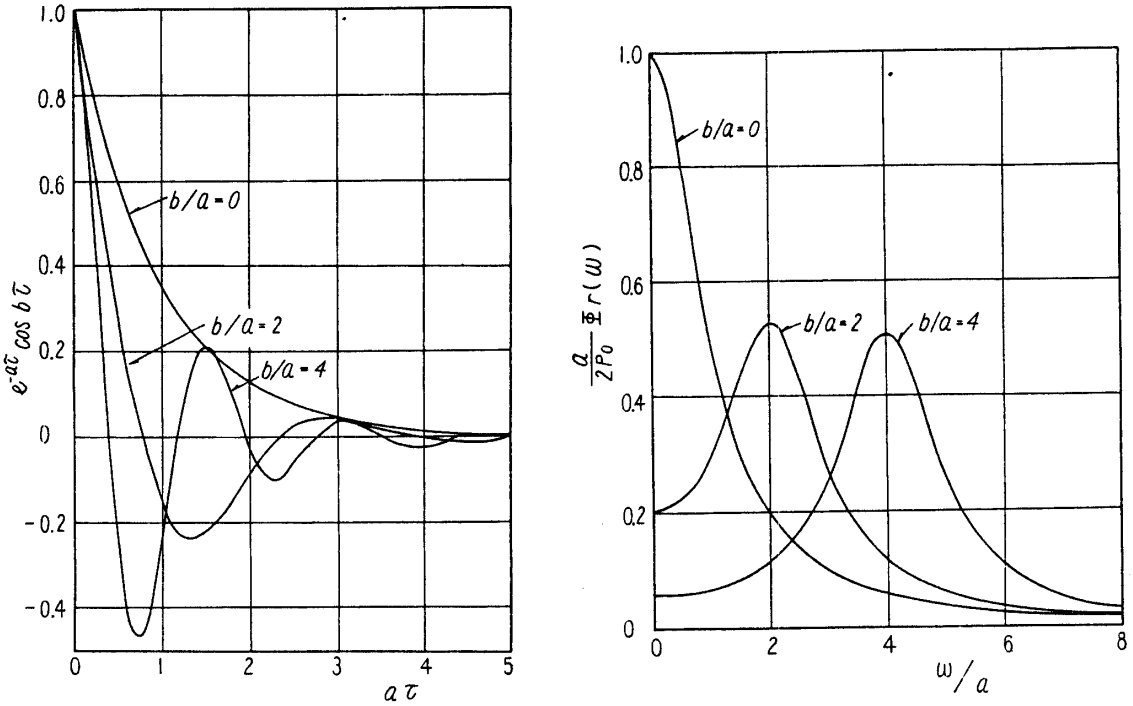

FIGURE 11. Pulse correlation function and spectral density of the input signal.

The pulse correlation function corresponding to Eq. (43) is readily obtained as

$$\phi_r^*(k)=P_0e^{-\alpha|k|}\cos\beta k \tag{47}$$

where $\alpha=aT$, $\beta=bT$, T is the sampling period and k is an integer. The pulse spectral density $\Phi_r^*(z)$ can be calculated by taking the two-sided z-transform of Eq. (47) and multiplying it by T. Hence, using the tables of ordinary z-transforms [9], we get the following tranform for the positive half part of Eq. (47).

$$\text{for}\quad k\geq 0: P_0T\frac{z^2-dz\cos\beta}{z^2-2dz\cos\beta+d^2} \tag{48}$$

$$(d=e^{-\alpha})$$

Substituting z^{-1} for z in Eq. (48), we get the transform for the negative half, thus

$$\text{for}\quad k\leq 0: P_0T\frac{z^{-2}-dz^{-1}\cos\beta}{z^{-2}-2dz^{-1}\cos\beta+d^2} \tag{49}$$

It must be taken into account that the value of $\phi_r^*(k)$ for $k=0$ is doubly transformed, so

for $$k=0: -P_0T \tag{50}$$

Taking the sum of these three parts the complete transformation of Eq. (47) is achieved,

$$\Phi_r^*(z)=P_0Td(1-d^2)r_2\cos\beta\frac{z(z-r_1)z^{-1}(z^{-1}-r_1)}{(z^2-2dz\cos\beta+d^2)(z^{-2}-2dz^{-1}\cos\beta+d^2)} \tag{51}$$

where r_1 and r_2 are two roots of a quadratic equation and calculated from the following set of equations.

$$\left.\begin{aligned} &r_1r_2=1\\ &r_1+r_2=\frac{d^2+1}{d\cos\beta}\\ &|r_1|\leq|r_2| \end{aligned}\right\} \tag{52}$$

Eq. (51) must be factored into two components according to Eqs. (15) and (16), then

$$\overset{+}{\Phi}_r^*(z)=\sqrt{P_0Td(1-d^2)r_2\cos\beta}\,\frac{z(z-r_1)}{z^2-2dz\cos\beta+d^2} \tag{53}$$

$$\overset{-}{\Phi}_r^*(z)=\sqrt{P_0Td(1-d^2)r_2\cos\beta}\,\frac{z^{-2}(z^{-1}-r_1)}{z^{-2}-2dz\cos\beta+d^2} \tag{54}$$

The pulse spectral density for Eq. (45) is obtained similarly, thus

$$\phi_r^*(k)=P_0e^{-\alpha|k|} \tag{55}$$

$$\Phi_r^*(z)=P_0T(1-d^2)\frac{zz^{-1}}{(z-d)(z^{-1}-d)} \tag{56}$$

$$\overset{+}{\Phi}_r^*(z)=\sqrt{P_0T(1-d^2)}\,\frac{z}{z-d} \tag{57}$$

$$\overset{-}{\Phi}_r^*(z)=\sqrt{P_0T(1-d^2)}\,\frac{z^{-1}}{z^{-1}-d} \tag{58}$$

For these random inputs, the optimum pulse transfer function of the control system under the constraints (1) and (2) in Sec. 2 is now readily obtainable by substituting the power spectral densities formulated above in Eq. (18) and (19). For instance, when $m=1$, that is the response of the controlled variable to the reference input is delayed by one sample, Eq. (18) becomes as

$$\begin{aligned} z^m\overset{+}{\Phi}_r^*(z)&=\sqrt{P_0Td(1-d^2)r_2\cos\beta}\,\frac{z^3-r_1z^2}{z^2-2dz\cos\beta+d^2}\\ &=\sqrt{P_0Td(1-d^2)r_2\cos\beta}\left\{z+\underbrace{\frac{(2d\cos\beta-r_1)z^2-d^2z}{z^2-2dz\cos\beta+d^2}}\right\} \end{aligned} \tag{59}$$

the part of zero and negative powers of z

Substitution of (59) in Eq. (19) results in a rational fractional function of z as the optimum pulse transfer function, that is

$$K^*(z)_{\text{opt}}=\frac{(2d\cos\beta-r_1)z^{-1}-d^2z^{-2}}{1-r_1z^{-1}}$$
$$=\frac{(k_1)_{\text{opt}}z^{-1}+(k_2)_{\text{opt}}z^{-2}}{1+(l_1)_{\text{opt}}z^{-1}} \tag{60}$$

Means square value of the control error for this case is calculated from Eq. (20)

$$\{\overline{e(nT)}\}^2_{\min}=P_0d(1-d^2)r_2\cos\beta \tag{61}$$

In Fig. 12, numerically calculated values of the parameters in Eq. (60) and corresponding values of Eq. (61) for the spectral density of $b/a(=\beta/\alpha)=2$ are shown as functions of α, where $\alpha=aT$ or

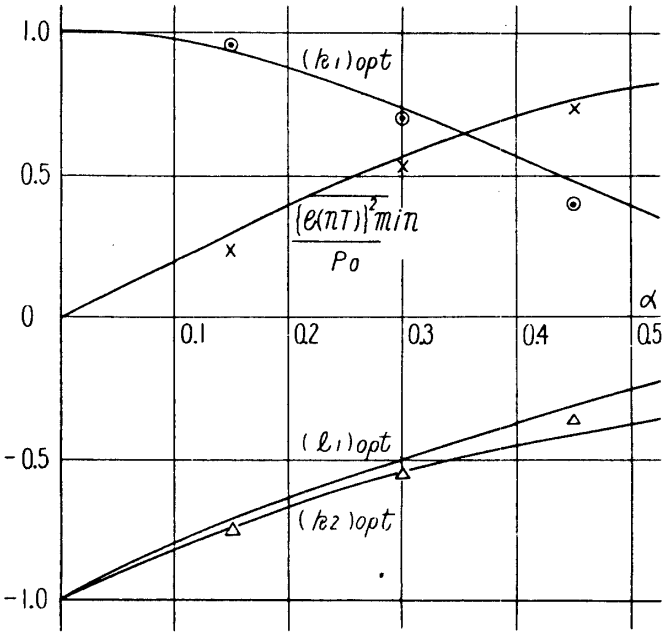

FIGURE 12. Calculated and experimental values of the optimum parameters in Eq. (60) and the minimum mean square error (for the random input of $a/b=2$, solid lines: calculated values).

$$f_s=\frac{1}{T}=\frac{a}{\alpha} \tag{62}$$

(f_s : sampling frequency)

that is, for constant a, α is decreased with the increase of f_s. Experimentally obtained values of the optimum parameters and minimum mean square value of the control error are also shown in Fig. 12. These values were obtained by setting the pulse transfer function of the simulator, which will be illustrated in the later section, as

$$K^*(z)=\frac{k_1z^{-1}+k_2z^{-2}}{1+l_1z^{-1}} \tag{63}$$

and seeking the optimum parameters that minimize the value of $\{\overline{e(nT)}\}^2$ on a vacuum tube voltmeter. Thc experiments were carried out for $\alpha=0.15$, 0.30 and

0.45, but the experimental values of l_1 were not shown, because the effect of parameter l_1 on the magnitude of the control error is insignificant except the vicinity of $+1$ or -1 (if $|l_1|\geq 1$, the control system becomes unstable) and it is difficult to determine the minimum point of $\overline{\{e(nT)\}^2}$ by changing the value of l_1. This fact may be illustrated by expanding Eq. (63) as

$$K^*(z)=\frac{k_1z^{-1}+k_2z^{-2}}{1+l_1z^{-1}}=k_1z^{-1}+(k_2-k_1l_1)z^{-2}+\cdots \tag{64}$$

and expressing the sampling value of the controlled variable $c(nT)$ as the weighted sum of the values of reference input $r(nT)$, thus

$$c(nT)=k_1\cdot r(nT-T)+(k_2-k_1l_1)\cdot r(nT-2T)+\cdots \tag{65}$$

Eq. (65) shows that l_1 is not included in the weight of the latest value of the input $r(nT-T)$, which is most effective to the present value of the controlled variable $c(nT)$. The parameter l_1 is contained in the second term as the form of k_1l_1 where k_1 takes the value less than unity, hence, l_1 is less effective on the control error than the other parameters.

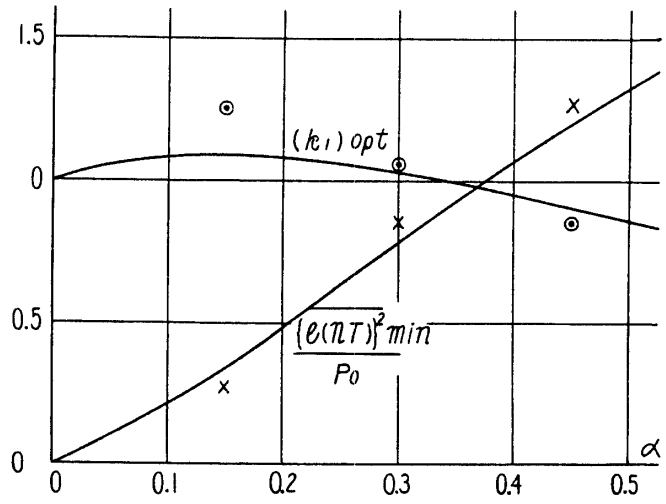

FIGURE 13. Calculated and experimental values of the optimum parameters in Eq. (30) and the minimum mean square error (for the random input of $b/a=2$, solid lines: calculated values).

In Fig. 13 a numerical example of the results of Sec. 3, that is, the optimum setting when the conditions of finite settling time response of the control system and zero off-set for a step input are taken into accounts, is shown. The spectral density of the reference input and other constraints of the control system are assumed to be same as in the foregoing example. That is, the pulse spectral density

$$\phi_r^*(i)=\phi_i=Pe^{-\alpha|i|}\cos\beta i \qquad (\beta/\alpha=2) \tag{66}$$

is substituted in Eq. (30) and Eq. (31), then $(k_1)_{\text{opt}}$, $(k_2)_{\text{opt}}$ and $\overline{\{e(nT)\}^2}_{\text{min}}$ are calculated as functions of α. The experimental results is also shown in the figure.

Comparison of Fig. 13 and Fig. 12 shows that the value of $\overline{\{e(nT)\}^2}_{\min}$ in Fig. 13 is always larger than that of Fig. 12, that is a consequence of additional constraints in the latter case. Furthermore, the value of $\overline{\{e(nT)\}^2}_{\min}$ in Fig. 13 exceeds unity while, in Fig. 12, the curve of $\overline{\{e(nT)\}^2}_{\min}$ never exceeds unity, because even in the worst condition $\overline{\{e(nT)\}^2}$ is unity, if the controlled variable is held at the mean value of the reference input, or no control action is made.

To examine the theory, the numerical calculation and experiments of the optimum design were executed for several cases other than the example stated in this section, and satisfactory accord of the experimental and calculated results was achieved.

6. The High Speed Simulator of Sample-Data Systems

The experiments of statistical design mentioned previously were carried out electrically according to the block diagram shown in Fig. 14. The random signal which has the spectral density of the form shown in Fig. 11 is achieved by passing the noise having flat spectral density or "white noise" from the noise generator through a filter which consists of C-R-L passive elements. This random signal is

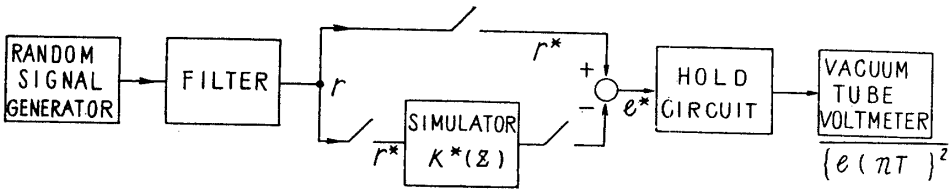

FIGURE 14. Block diagram of the experiments.

the input to the high speed electronic simulator which simulates the behaviour of a sampled-data control system as a whole and the output of the simulator is subtracted from the input giving the control error. A zero-order hold circuit clamps the values of the error signal at sampling instants for the intervals between samples. The waveform of the output of the hold circuit is a train of rectangles, which changes stepwise its value at each sampling instant. The mean square value of the control error at sampling instants $\overline{\{e(nT)\}^2}$ is obtained by measuring the effective voltage of this signal by a vacuum tube voltmeter. The optimum condition of the control system is searched for by changing the parameters of the simulator step by step to get the minimum reading of the voltmeter. It is a matter of course that if there is a constraint such as Eq. (22) the parameters of the simulator must be changed while holding always these relations. The simulator is able to simulate the part of sampled-data controller shown in Fig. 2 not the whole control system as in this experiment, and is also able to be used to investigate the indicial response of a sampled-data system on a cathode-ray oscilloscope.

The principle of the simulator is that suggested by A. R. Bergen and J. R. Ragazzini [*1*] which uses sampled-data delay lines and a summing amplifier with various weights. The simulator constructed for this experiment is an all-electronic high speed type in which sampled-data delay lines by analog method are adopted.

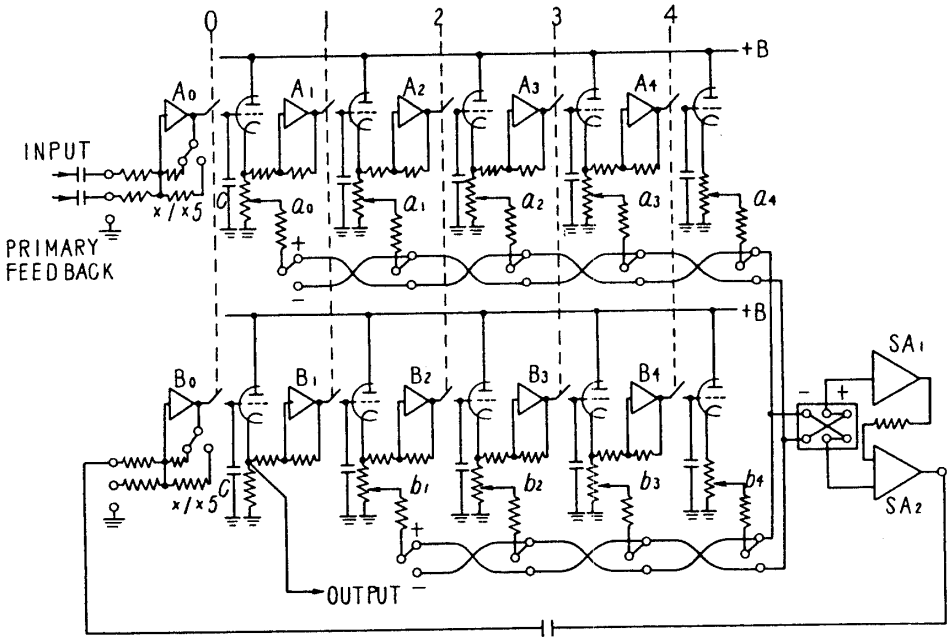

FIGURE 15. Block diagram of the simulator.

In Fig. 15, the block diagram of the simulator is shown, where $A_0, A_1, \cdots B_0, B_1, \cdots$ are amplifiers with gain of -1 (including each cathode follower) and $0, 1, 2, \cdots$ represent electronic switches. When the electronic switches are switched on for a short time in order of $4, 3, \cdots 0$, the past values of the input and output signals of the simulator at each sampling instants, which have been stored in the condenser in voltage forms, are shifted respectively to the next stages toward right and new samples are taken in by the simulator. The sampling instant of the simulator is the moment when the switch 0 is made on, hence the sampling operation takes a finite time duration equal to the width of the switching pulse and five switching pulses which are generated in order $4, 3, \cdots 0$ are needed to accomplish the complete cycle of a sampling period.

The signal values stored in the delay lines are summed up by the amplifiers SA_1 and SA_2 with the weight of $a_0, a_1, \cdots b_0, b, \cdots$ respectively, getting the linear pulse transfer function of the form

$$D^*(z)=\frac{a_0+a_1z^{-1}+a_2z^{-2}+\cdots+a_pz^{-p}}{1+b_1z^{-1}+b_2z^{-2}+\cdots+b_qz^{-q}} \tag{67}$$

In Eq. (67), $a_0, a_1, \cdots b_0, b_1, \cdots$ are parameters and when they take the values larger than unity, the amplifiers A_0 and B_0 are switched to have gain of $\times 5$. If a parameter takes negative value, the connection of that stage to the summing amplifiers must be changed to the side of negative sign. In this instance, however, it must be taken into account that the polarity of the signal values in the delay lines are changed from stage to stage because the gain of A's and B's are -1. The output of the simulator is taken from the cathode follower of the B_0 stage. Therefore, the waveform of output is a train of steps, not samples discrete in time, and hence, the similator is regarded as including the hold circuit in the block diagram of Fig. 14.

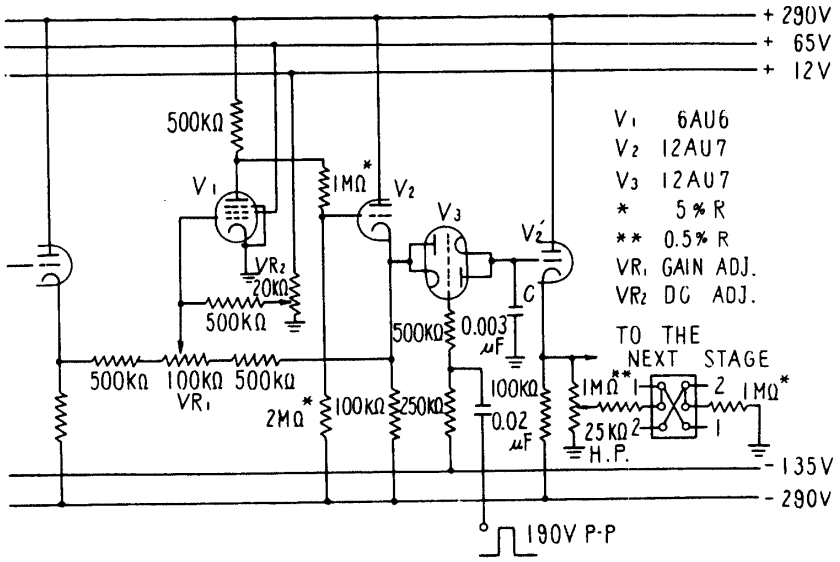

FIGURE 16. Circuit diagram of a stage of the delay lines in the simulator.

The circuit diagram of a stage of the delay lines is shown in Fig. 16. The tube V_3 is an electronic switch which is driven to conducting state by the switching

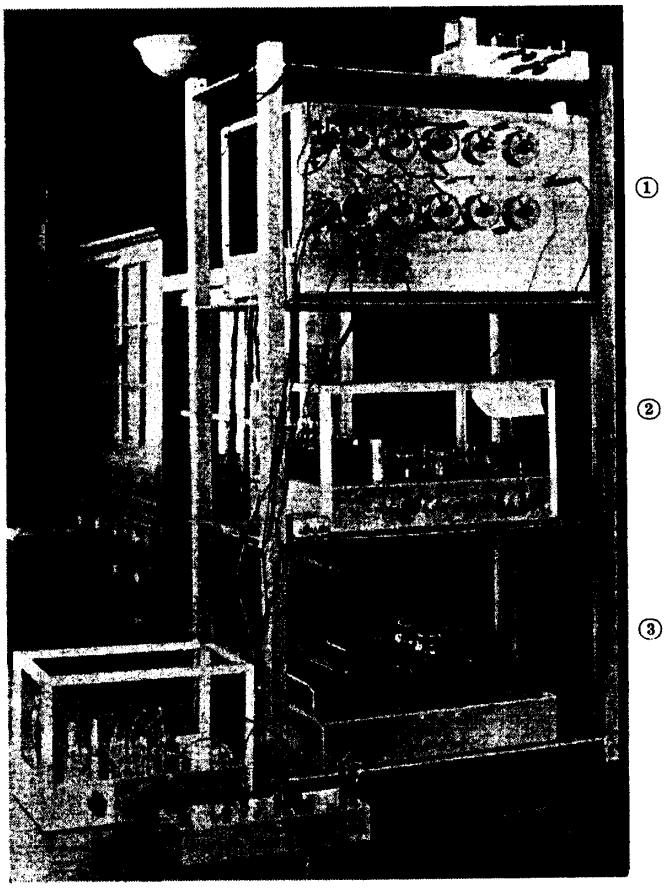

PHOTO. Arrangement of the device.
① simulator. ② switching pulse generator. ③ source.
④ random signal generator and filters.

pulse, thus charging the capacitor to a new voltage. V_3 is operated with the 80% heater voltage of the rating in order to decrease the leakage current between the hearter and cathode. The capacitor C is the memory element in the delay lines and a polystyrene condenser is used to avoid the error due to the absorption of change or "soak in" effect. The drift of DC level in the amplifier which is consisted of V_1 and V_2, is not so serious, because the simulator is AC-coupled by capacitors to the other devices.

The sampling frequency of this simulator is variable between 250 to 1000 c/s and the maximum voltage range of the input and output signal is ± 30V (when the gain of A_0 and B_0 is $\times 1$). The main source of the error of the simulator is that the sampling of signals is not operated instantaneously, instead it needs the time about $T/10$, where T is the sampling period. However, the simulator has sufficient accuracy for this experiment.

It should be noted here that the amplifiers $A_0, A_1, \cdots A_4$ or $B_0, B_1, \cdots B_4$ are never operated simultaneously, but they are operated one after another, so it is sufficient to switch over one amplifier from stage to stage. In the case where a device of this sort is used as a sampled-data controller operating in real time and therefore, the sampling period is sufficiently long, it is advantaneous to adopt the method of construction that mechanical relays switch over an operational amplifier such as those using chopper-stabilized circuits.

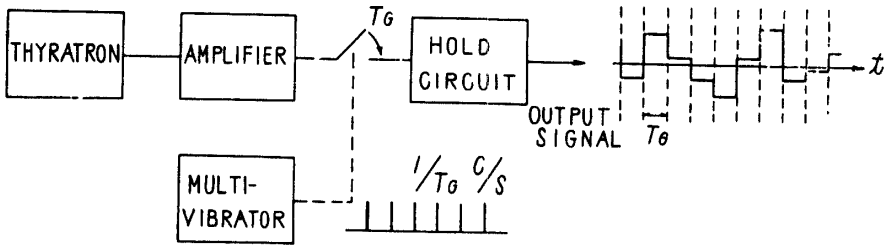

FIGURE 17. Block diagram of the low-frequency random signal generator.

Fig. 17 is the block diagram of the low-frequency random signal generator constructed for the experiment, in which a thyratron used as the noise source. The thyratron noise, after amplified, is sampled at every 1 ms by an electronic switch and is clamped by the hold circuit similar to those of the simulator, giving the output voltage which changes its value irregularly at every sampling instant as shown in the figure. The flicker effect and ham in the amplifier of the generator are cut out by R-C coupling between tue stages of the amplifier and the stable low-frequency random signal is obtainable as the output. This output voltage is filtered to have a specified spectral density as already shown in Fig. 14. For the random signals used in the experiment stated in Sec. 5, the parameter a which specifies the form of spectral density shown in Fig. 11 is set equal to 125 rad/sec or about 20 c/s. A low-frequency random signal generator based on these principles was reported preuiously by D. F. Winter [*10*], for which a mechanical commutator was used as a sampling device. However, the generator illustrated here was consisted of electronic tubes, including sampling and holding circuits.

The vacuum tube voltmeter to read the mean square value of the random signal is one of the types which detect the input voltages by the square-law characteristics of triodes and it was reconstructed so as to be able to measure the frequency component down to 5 c/s with sufficient averaging time constant.

7. Conclusive Remarks

The two essential features of digital devices, that is sampling and quantization of signals, were dealt with in this paper and the statistical design methods of control systems containing these devices were developed. In Secs. 2 and 3, only sampling operation was taken into consideration and the optimum pulse transfer functions of the control systems subjected to random inputs were derived under the constraints such as time delay in the response of the controlled variables, finite settling time of the systems, etc. The effects of the quantization of signals on optimum design are usually trivial as discussed in Sec. 4. To examine the theory of optimum design, the experiments were carried out using a simulator of sampled-data systems and satisfactory accord of the experimental and theoretical results was achieved. In the last part, the simulator constructed for this experiment was illustrated and a method to construct a sampled-data controller operating in real time was suggested.

Throughout this study, the values of the control error only at sampling instants were considered. In practice, however, it is the averaged magnitude of the control error as a continuous signal that must be kept to be minimum. According to the sampling theorem [*11*], the representation of the control error $e(t)$ by its values at sampling instants $e(nT)$ is based on the assumption that the frequency components of $e(t)$ higher than one half of the sampling frequency is not so excessive. Therefore, if the system exhibit very oscillatory response between sampling instants, the method stated in this paper should not be applied carelessly. The similar situation will occur when the optimum pulse transfer function of the control system is decided and the parameters of sampled-data controller such as shown in Fig. 2 are set to cause the system to have the specified pulse transfer function. In such a case, hidden instability—a sampled-data control system is unstable concerning the values of the controlled variables between sampling instants although stable at the sampling time [*12*]—may occur according to the characteristics of the controlled system, such as the ratio of the dead time of the controlled system to the sampling period. To sum up the foregoing comments, it may be said that the statistical design methods developed in this paper should always be used in connection with other design methods and the results of the statistical design must be examined from other points of view such as the indicial response of the system.

6. Acknowledgements

The author wishes to thank Prof. Takashi Isobe for his instructive advices and encouragement. The author also wishes to acknowledge Prof. Juichi Igarashi

who has given many conveniences to him for the experiments in this study.

Aeronautical Research Institute
University of Tokyo, Tokyo
December 14, 1960

References

[1] A. R. Bergen and J. R. Ragazzini: Sampled-Data Processing Techniques for Feedback Control Systems, AIEE Trans., No. 15, p. 236 (1954).
[2] M. Mori: Statistical Treatment of Sampled-Data Control Systems for Actual Random Inputs, ASME Trans., Vol. 80, p. 444 (1958).
[3] M. Mori: Statistical and Transient Approaches to Sampled-Data Control Systems, Rep. Institute of Industrial Science, University of Tokyo, Vol. 7, No. 3 (1958).
[4] G. W. Johnson, D. P. Lindorff and C. G. A. Nordling: Extension of Continuous-Data System Design Techniques to Sampled-Data Control Systems, AIEE App. & Ind., No. 20, p. 252 (1955).
[5] R. H. Barker: A Servo System for Digital Data Transmission, Proc. IEE, Pt. B, Vol. 103, No. 7, p. 52 (1956).
[6] S. S. L. Chang: Statistical Design Theory for Strictly Digital Sampled-Data Systems, AIEE Com. & Elect., No. 34, p. 702 (1958).
[7] W. R. Bennett: Spectra of Quantized Signals, Bell Sys. Tech. Jour., Vol. 27, p. 446 (1948).
[8] B. Widrow: A Study of Rough Amplitude Quantization by Means of Niquist Sampling Theory, IRE Trans., Vol. CT-3, p. 266 (1956).
[9] R. H. Barker: The Pulse Transfer Function and Its Application to Sampling Servo Syetems, Proc. IEE, Pt. 4, Vol. 99, No. 4, p. 302 (1952).
[10] D. F. Winter: A Gaussian Noise Generator for Frequencies Down to 0.001 c/s, Conv. Record IRE, Vol. 2, Pt. 4, p. 23 (1954).
[11] S. Goldman: Information Thory (book), Chap. 2, Prentice-Hall Inc. (1953).
[12] E. I. Jury: Hidden Oscillations in Sampled-Data Control Systems, AIEE App. & Ind., No. 28, p. 391 (1957).