crank arms which can be adjusted from the full stop to any throttled discharge position. Discharge of the on-storage conveyor lines occurs during the first and part of the second shift. Within the limits of practicability, this operation is scheduled so the conveyor system is empty of both stored and first-shift production by the beginning of the second shift.

A simple weighted chute at the end of each on-storage conveyor permits directing the released cartons onto a 3-wide, inclined, powered conveyor. Each of the three inclined conveyors is connected through power or gravity conveyor lines to a 4-high swing belt. Three such swing belts are grouped into one assembly. These belts, in turn, are connected to other conveyors which lead to three rail docks and one combination truck-rail dock. This arrangement permits delivery of cartons from any of the 69 conveyor lines to any loading dock without manual handling. In actual operation, loading may vary from a solid railroad car of one product to a truck having a wide variety of products arranged for store-to-store route deliveries.

Other plans and arrangements were studied before the afore-described system was selected, but each of the alternatives lacked the flexibility and economy required to accomplish the desired result. One of the by-products and significant long-term benefits anticipated in the selected system is the saving in carton breakage. Throughout the planning, emphasis has been laid on reducing material handling costs which are generally estimated as 25–30% of all direct manufacturing costs.[4]

Conclusions

This food product installation indicates that, with the dependable control methods now available, automation will undoubtedly be extended into more process plans than were heretofore considered practical. It shows that by combining commercially developed control devices, a very special tailor-made result can be obtained. A paper punch tape has been used as a memory circuit; automatic elevator control for jib crane operation; and Cypak static control units for coding, decoding, and other operations. The use of Cypak has simplified the design and resulted in a system which should require minimum maintenance over the years. The end result of this installation is a conveyor system which sorts out 31 products and dispatches them to 72 on-storage conveyor lines having an aggregate of 11,300 feet of lineal carton storage space. The system makes possible shipment of three shifts of daily production in an essential single shift without manual handling from production to the shipping facility. Further, it reduces to a minimum the manual warehouse handling of those cartons which are not scheduled for immediate shipment. The tape record is available for production analysis.

References

1. ON THE CONTROL OF MAGNETIC AMPLIFIERS, **R. A. Ramey.** *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 2124–28.
2. CYPAK SYSTEMS, **Editorial Staff.** *Westinghouse Engineer*, East Pittsburgh, Pa., July 1955, p. 4.
3. CYPAK SYSTEMS, **J. P. Baker.** *Ibid.*, July 1956, p. 2.
4. PLANT LAYOUT (book), **R. W. Mallick, A. T. Gaudreau.** John Wiley & Sons, Inc., New York, N. Y., 1951.

General Synthesis Procedure for Computer Control of Single-Loop and Multiloop Linear Systems (An Optimal Sampling System)

R. E. KALMAN
NONMEMBER AIEE

J. E. BERTRAM
ASSOCIATE MEMBER AIEE

Synopsis: This paper is concerned with the problem of designing optimal systems for the control of a plant governed by a linear differential equation with constant coefficients. Control is exerted by means of piecewise constant signals which can change only at the "sampling instants." Optimality means here, as in some nonlinear problems, that the system achieves equilibrium with zero steady-state error from any initial state as quickly as possible.

The sequence of signals required for optimal control in this sense is ideally a linear combination of the state variables of the plant being controlled. In the simplest regulator problem, when all state variables are accessible, the optimal system is realized by a very simple multifeedback arrangement in which the only unconventional component is the sample-and-hold element. When some states are not accessible, when the plant includes time delays, and in some types of follow-up systems, the physical realization of optimal control requires real-time analog or digital computation. The requirements of machine computation for control purposes are derived here in a general way; the theory presented includes as special cases many commonly used methods for control system synthesis.

IT HAS BEEN pointed out in several recent papers[1–4] that the use of sampled-data techniques results in control systems superior in many respects to unsampled systems. Yet until now few practical applications of sampled-data control systems exist, except in those cases where the input or error signal of the system is inherently sampled. Probably the major reasons why sampling techniques have not been more widely used is the complexity and expense of the sampling controller compared to the simple elements available to compensate unsampled systems. Further, in those cases where sampling controllers have been tried, it was found that adjustment of the controller parameters for optimal operation would be very difficult unless the plant transfer function could be measured with high accuracy. An additional difficulty is that a sampling system de-

Paper 58-1161, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted March 13, 1958; made available for printing July 25, 1958.

R. E. KALMAN is with the Research Institute for Advanced Studies (RIAS), Baltimore, Md., and J. E. BERTRAM is with International Business Machines Corporation, Ossining. N. Y.

This research was supported in part by the U.S. Air Force under Contract No. *AF 18(600)-677* at Columbia University, New York, N. Y., and was monitored by the Office of Scientific Research, Air Research and Development Command.

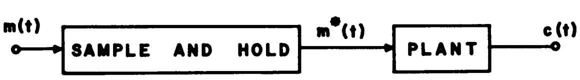


Fig. 1 (left). The system to be controlled

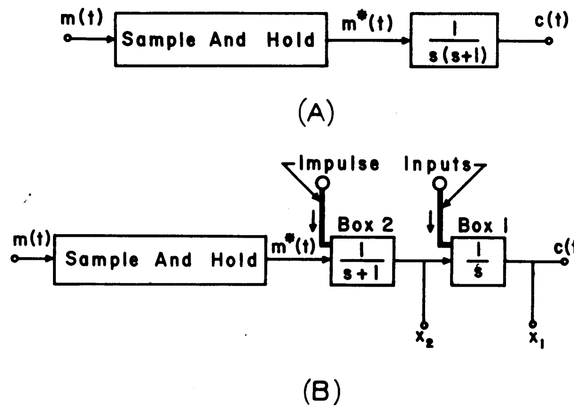


Fig. 2 (right). (A) An example. (B) Selection of state variables

signed by conventional methods to perform in an optimum fashion at sampling instants does not always have satisfactory performance at other times.[5]

In this paper a design technique is developed which results in optimal performance at all times, not just at the sampling instants. The sampling controller for optimal performance can be easily constructed. The experimental adjustment procedure to achieve optimal performance is found to be quite simple.

The significance of the principles introduced here is quite broad; numerous additional matters such as compensation for time delay and saturation, proper use of real-time machine computation, etc., are discussed briefly. This paper is directly mainly to questions of engineering interest; a rigorous and detailed discussion of the mathematical questions will be found in an (as yet) unpublished paper by R. E. Kalman, entitled, "On the General Theory of Control Systems."

Statement of Problem

Consider the system shown in Fig. 1 which consists of the following.

1. The plant, which is governed by an ordinary, linear, time-invariant differential equation. As is customary in the control systems literature, the plant is described by a rational transfer function.
2. The sample-and-hold element, which replaces the continuous control signal $m(t)$ by a piecewise constant or sampled signal $m^*(t)$, described by

$$m^*(t)=m(kT),\ kT\leq t<(k+1)T, \quad k=0, 1, 2, \ldots \tag{1}$$

where the sampling period T is a positive constant, the point $t=kT$ in time is called the k^{th} sampling instant. Frequently, it will be convenient to write $f(k)$ instead of $f(kT)$, where $f(t)$ is some function of time.

The design of control system can be formulated in general as follows: Given an arbitrary input signal $r(t)$, select the control signal $m(t)$ in such a way that the error $e(t)$

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

is as small as possible in some precisely defined sense.

This paper is concerned primarily with the following problem:

Find the function of time $m(t)$, that is, the sequence of numbers $m(0)$, $m(1)$, ... which will bring the plant from any condition existing at $t=0$ to equilibrium with zero error in a minimum number of sampling periods.

In the following sections a complete solution of this problem will be given in the special case when $r(t)=$constant. Such a control system is called a regulator; the problem stated is then that of designing a regulator which eliminates the effect of any suddenly applied disturbance to the plant in the minimum length of time consistent with a fixed choice of the sampling period T.

If $r(t)$ is an arbitrary function of time, the problem cannot be solved in general. However, when $r(t)$ is of the form

$$r(t)=\sum_{i=1}^{m} b_i t^{i-1},\ t\geq 0 \tag{3}$$
$$=0,\ t<0$$

(where the b_i are arbitrary real numbers) the solution can be readily obtained by modifying the solution of the regulator problem.

Conventional methods[2] for designing sampled-data systems are frequently derived in such a way as to make $e(kT)=0$ at every sampling instant after, say N sampling periods following the application of a prototype input (such as a step or a ramp) to a system at rest. The disadvantage of such methods is that the plant may not have reached equilibrium in N steps, so that even though the error will be zero at sampling instants it will not be zero between the sampling instants at any time $t\geq NT$. This phenomenon is well known in the theory of sampled-data systems and is referred to as intersample "ripple." In the present formulation of the problem, the plant is required to reach equilibrium when the error becomes zero. As a result, the ripple problem is completely eliminated.

It is interesting to note, as a matter of historical interest, that the present formulation of the problem, based on the ripple-free requirement, was announced by one of the present authors in 1954, together with the solution in a simple special case.[6] At that time, the basic idea was derived from certain well-known methods in the theory of nonlinear (saturating) control systems; here the earlier idea is generalized using techniques for the theory of nonlinear dynamic systems.

System Description

At the sampling instants the behavior of the system shown in Fig. 1 may be represented by the equation

$$\mathbf{x}(k+1)=G(T)\mathbf{x}(k)+m(k)\mathbf{h}(T) \tag{4}$$

where $\mathbf{x}(k)$ is a vector representing the state of the plant at the k^{th} sampling instant; $G(T)$ is an $n\times n$ matrix whose elements are functions of T, called the transition matrix[7] of the differential equation governing the behavior of the plant; $\mathbf{h}(T)$ is a vector which represents the effect on the state of the plant at the $(k+1)^{\text{th}}$ sampling instant of a unit-step control signal applied to the plant at the k^{th} sampling instant.

To illustrate how equation 4 is obtained explicitly, consider the example of Fig. 2(A). This example will be used throughout the paper. The first step is to redraw the block diagram of the plant in such a way as to make the state variables (i.e., the components of the state vector) accessible. This can be done by decomposing the plant into a larger number of block in such a way that each block contains a transfer function with one pole and no finite zero. There are many possible ways of doing this; however, the choice is considerably narrowed by requiring that as many of the state variables as possible must be physically measurable quantities. A possible choice is shown in Fig. 2(B).

To define and motivate the preceding terminology, consider now the differential equations corresponding to Fig. 2(B).

$$dx_1/dt=x_2$$

$$dx_2/dt=-x_2+m^*(t) \tag{5}$$

Any two values of $x_1(kT)$, $x_2(kT)$ at the k^{th} sampling instant define a point in the x_1,x_2 plane which is called the state space (or phase space) of the system, equation 5. Analogously, any point in the state space is called a state. When $m^*(t)=0$, the evolution of the system, equation 5, from an arbitrary present state $(x_1(kT), x_2(kT))$ can be represented by a curve in the state space, called the trajectory, along which time is a parameter; see Fig. 3 where the arrow indicates the positive direction of time. Since the solution of a linear differential equation is uniquely determined for all finite values (positive or negative) of time by the initial state, there is one and only one trajectory passing through any given point in the x_1, x_2 plane. It is convenient to think of the behavior of the system, equations 5, in the absence of an input as the motion of a point representing the state along a trajectory. If the point does not move at all, i.e., if the trajectory through the point is the point itself, the point is called an equilibrium state. The equilibrium states of equation 5 are the states for which $dx_1/dt=dx_2/dt=0$.

The fundamental matrix expresses the dependence of any future (or past) state on the present state. To find the elements of this matrix, it is only necessary to express the state of the plant at any time explicitly as a function of the present state. Now, the present value $x_1(kT)$ can be brought about by applying an impulse at $t=kT$ of area $x_1(kT)$ to the input of box 1, if the system in Fig. 2(B) is initially in the zero state. A typical germ $g_{ij}(T)$ of $G(T)$ is therefore obtained by calculating the response of the plant observed at state variable x_i at time $t=(k+1)T$ when a unit impulse is applied at $t=kT$ to the input of the box containing state variable x_j. In case of Fig. 2(B),

$$G(T)=\begin{Vmatrix} g_{11}(T) & g_{12}(T) \\ g_{21}(T) & g_{22}(T) \end{Vmatrix}=\begin{Vmatrix} 1 & 1-e^{-T} \\ 0 & e^{-T} \end{Vmatrix} \quad (6)$$

To calculate the effect of the control signal $m^*(t)$, note first that it is constant between sampling instants by virtue of its definition. In other words, the control signal is always a step function. To obtain the response observed at state variable x_i at time $t=(k+1)T$ as a result of applying a step at $t=kT$ to the input of the box containing the state variable x_j, it is only necessary to integrate the corresponding impulse responses, which have already been determined. Thus:

$$\int_0^T G(\tau)d\tau=H(T)$$
$$=\begin{Vmatrix} \int_0^T g_{11}(\tau)d\tau & \int_0^T g_{12}(\tau)d\tau \\ \int_0^T g_{21}(\tau)d\tau & \int_0^T g_{22}(\tau)d\tau \end{Vmatrix}$$
$$=\begin{Vmatrix} T & T-1+e^{-T} \\ 0 & 1-e^{-T} \end{Vmatrix} \quad (7)$$

In the present example, however, the control signal is applied only to the box containing x_2. Therefore only the second column of $H(T)$ is of interest, so that

$$\mathbf{h}(T)=\begin{Vmatrix} T-1+e^{-T} \\ 1-e^{-T} \end{Vmatrix} \quad (8)$$

More generally, one can write

$$\mathbf{h}(T)=H(T)\mathbf{b} \quad (9)$$

where the vector $\mathbf{b}$ is the state resulting from the application of an impulse to the input of the plant which is initially in the state $\mathbf{x}=0$.

In explicit notation, the vector equation 4, for the example is:

$$x_1(k+1)=x_1(k)+(1-e^{-T})x_2(k)+(T-1+e^{-T})m(k)$$
$$x_2(k+1)=e^{-T}x_2(k)+(1-e^{-T})m(k) \quad (4A)$$

Two important properties of the transition matrix are needed in the sequel. First, if $m(k)=0$ for all k, then using equation 4 several times shows that

$$\mathbf{x}(k+l)=G(lT)\mathbf{x}(k) \quad (10)$$

for any integers (in fact, any real numbers) k, l. But, since by the afore-mentioned observation $\mathbf{x}(k)=G(kT)\mathbf{x}(0)$ and $\mathbf{x}(k+l)=G((k+l)T)\mathbf{x}(0)$, it follows that $\mathbf{x}(k+l)=G(lT)\mathbf{x}(k)=G(lT)G(kT)\mathbf{x}(0)=G((k+l)T)x(0)$ for any state $x(0)$. Therefore

$$G((k+l)T)=G(kT)G(lT) \text{ (any } k,\, l) \quad (11)$$

In particular, letting $l=-k$, it follows:

$$G^{-1}(kT)=G(-kT) \quad (12)$$

Equations 11 and 12 make it very easy to calculate powers and inverse powers of the transition matrix.

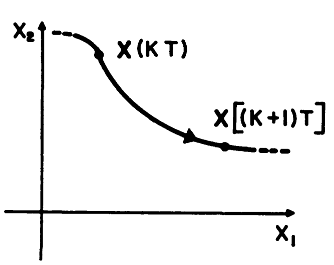


Fig. 3. Trajectory in state space

The Canonic Co-ordinate System

Consider now the solution of the basic problem, with $r(t)=0$. In other words, it is desired to find a sequence $m(0)$, $m(1)$, ... which will take any initial state $\mathbf{x}(0)$ to equilibrium with zero error, i.e., to the state $\mathbf{x}=0$, in a minimum number of sampling periods.

If $\mathbf{u}_1$, $\mathbf{u}_2$ are the unit vectors along the co-ordinate axes in the state space of the running example, then the state vector at any time kT can be written in the form (see Fig. 4)

$$\mathbf{x}(k)=x_1(k)\mathbf{u}_1+x_2(k)\mathbf{u}_2$$

Now consider the question: From what initial states $\mathbf{x}(0)$, if any, is it possible to reach the origin of the state space in one sampling interval? Let $\mathbf{x}^1(0)$ be any initial state with this property; setting $\mathbf{x}(1)$ equal to zero in equation 4 leads to

$$\mathbf{x}(1)=0=G(T)\mathbf{x}^1(0)+m(0)\mathbf{h}(T)$$

Solving for $\mathbf{x}^1(0)$ and using equation 12 shows that

$$\mathbf{x}^1(0)=-m(0)G^{-1}(T)\mathbf{h}(T)$$
$$=-m(0)G(-T)\mathbf{h}(T) \quad (13)$$

Letting

$$\mathbf{v}_1=G(-T)\mathbf{h}(T) \quad (14)$$

equation 13 shows that every state $\mathbf{x}^1(0)$ can be written in the form

$$\mathbf{x}^1(0)=y_1\mathbf{v}_1 \quad (15)$$

where y_1 is some real constant. The set of all states satisfying equation 15 clearly constitute a straight line in the state space.

The set of all states from which the origin can be reached in at most two steps is clearly the same set from which a state of the form given by equation 15 can be reached in at most one step. Let $\mathbf{x}^2(0)$ be a state from which the origin can be reached in two steps; then, again using equation 4,

$$\mathbf{x}(1)=y_1\mathbf{v}_1=G(T)\mathbf{x}^2(0)+m(0)\mathbf{h}(T)$$

Solving for $\mathbf{x}^2(0)$ and again using equation 12 as well as equation 14 leads to

$$\mathbf{x}^2(0)=G^{-1}(T)(y_1\mathbf{v}_1-m(0)\mathbf{h}(T))=$$
$$y_1G(-2T)\mathbf{h}(T)-\mathrm{m}(0)G(-T)\mathbf{h}(T)$$

Letting $\mathbf{v}_2=G(-2T)h(T)$, it follows from equation 14 that every $\mathbf{x}^2(0)$ can be written in the form

$$\mathbf{x}^2(0)=y_1\mathbf{v}_1+y_2\mathbf{v}_2 \quad (16)$$

where y_1, y_2 are real numbers. Using the

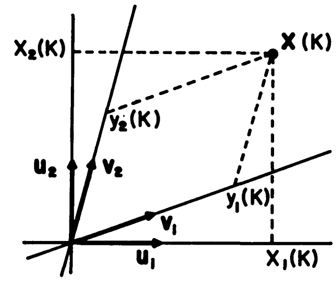


Fig. 4. Co-ordinate systems

explicit forms of $G(T)$ and $\mathbf{h}(T)$ given by equations 7 and 8, $\mathbf{v}_1$, $\mathbf{v}_2$ can be written as linear combinations of the unit vectors $\mathbf{u}_1$, $\mathbf{u}_2$

$$\mathbf{v}_1=(T+1-e^T)\mathbf{u}_1+(e^T-1)\mathbf{u}_2$$
$$\mathbf{v}_2=(T-e^{2T}+e^T)\mathbf{u}_1+(e^{2T}-e^T)\mathbf{u}_2 \qquad (17)$$

It is easily verified from equation 17 that the vectors $\mathbf{v}_1$, $\mathbf{v}_2$ are linearly independent for any value of T, i.e., they do not lie along the same line through the origin of the state space. Therefore $\mathbf{v}_1$, $\mathbf{v}_2$ may be regarded as unit vectors along the axes of a new co-ordinate system, $\mathbf{x}(0)$ having the co-ordinates y_1, y_2 in this new system, as given by equation 16. Since any vector in the two-dimensional state space of the running example can be represented as the linear combination of at most two linearly independent vectors, it follows also from equation 16 that the origin can be reached always in at most two steps.

In the n^{th} order case, it can be rigorously shown (in the previously mentioned unpublished paper) that the set of n vectors

$$\mathbf{v}_1,\ldots,\mathbf{v}_n \text{ where } \mathbf{v}_i=G(-iT)\mathbf{h}(T) \qquad (18)$$

are linearly independent if and only if

$$T\omega\neq q\pi \qquad (19)$$

where q is any integer and ω is the imaginary part of any of the complex poles of the transfer function of the plant. The co-ordinate system for which the $\mathbf{v}_i$ in equation 18 are the unit vectors will be called the canonical co-ordinate system.

The foregoing may be summarized as follows:

If the sampling period T satisfied condition 19, then any state $\mathbf{x}$ may be written in the form

$$\mathbf{x}=\sum_{i=1}^{n}y_i\mathbf{v}_i \qquad (20)$$

where $\mathbf{v}_i=G(-iT)\mathbf{h}(T)$ and n is the order of the system. Moreover, if j is the largest integer for which $y_j\neq0$, then the origin may be reached from $\mathbf{x}$ in exactly j steps.

It should be noted also that the maximum response time of an n^{th} order optimal system is nT seconds. Thus the response time of a sampled system can be made arbitrarily small by selecting a short sampling period. However, by doing so, the magnitude of the control signals required will increase rapidly and the fast response may not be obtainable because plant saturation will ultimately limit the effect of the control signal.[6] As a rough rule, it probably does not pay to decrease T below about 10% of the dominant time constants of the plant.

The Control Signal Sequence

The next problem is to determine the actual values of $m(0)$, $\ldots$, $m(n)$ which will take an arbitrary initial state $\mathbf{x}(0)$ into the origin. Considering again the running example, it is known that every initial state can be written in the form:

$$\mathbf{x}(0)=y_1(0)\mathbf{v}_1+y_2(0)\mathbf{v}_2 \qquad (21)$$

If $m(0)$ is chosen in an optimal way, the state at the next sampling instant must be given by

$$\mathbf{x}(1)=y_1(1)\mathbf{v}_1 \qquad (22)$$

Equation 4 for $k=0$ is

$$\mathbf{x}(1)=G(T)\mathbf{x}(0)+m(0)\mathbf{h}(T)$$

Substituting equation 21 for $\mathbf{x}(0)$

$$=G(T)(y_1(0)\mathbf{v}_1+y_2(0)\mathbf{v}_2)+m(0)\mathbf{h}(T)$$

Using equation 18, the definition of the $\mathbf{v}_i$, and also equation 11, it further follows

$$\mathbf{x}(1)=y_2(0)\mathbf{v}_1+(y_1(0)+m(0))\mathbf{h}(T)$$

Hence $\mathbf{x}(1)$ will have the form required by equation 22 if and only if

$$m(0)=-y_1(0) \qquad (23)$$

in which case

$$y_1(1)=y_2(0) \qquad (24)$$

If $\mathbf{x}(1)$ is to be reduced to zero in one step

$$\mathbf{x}(2)=0=y_1(1)G(T)\mathbf{v}_1+m(1)\mathbf{h}(T)$$
$$=(y_1(1)+m(1))\mathbf{h}(T)$$

Therefore

$$m(1)=-y_1(1)=-y_2(0) \qquad (25)$$

These results can be easily generalized as follows:

At every sampling instant, the optimal control signal $m(k)$ is given by:

$$m(k)=-y_1(k) \qquad (26)$$

where

$$\mathbf{x}(k)=\sum_{i=1}^{n}y_i(k)\mathbf{v}_i \qquad (20)$$

The sequence of control signals $m(k)$ which takes an arbitrary initial state $\mathbf{x}(0)$ to the origin in the absence of any disturbances is given by

$$m(0)=-y_1(0),\ m(1)=-y_2(0),\ \ldots,\ m(n-1)=-y_n(0) \qquad (27)$$

If $m(k)$ satisfies equation 26, then

$$\mathbf{x}(k+1)=\sum_{i=1}^{n-1}y_{i+1}(k)\mathbf{v}_i \qquad (28)$$

and therefore

$$y_i(k+1)=y_{i+1}(k) \qquad (29)$$

A simple and interesting geometrical interpretation of equations 28 and 29 is this. The initial state of the plant may lie anywhere in the n-dimensional state space. After one sampling period and the application of a control signal as given by equation 26, the state can be expressed as a linear combination of only $n-1$ linearly independent vectors $\mathbf{v}_1$, $\ldots$, $\mathbf{v}_{n-1}$; in other words, after one sampling period, the initial state always lies in an $(n-1)$-dimensional space, provided there are no external disturbances. After two sampling periods, the initial state is taken into a space of one less dimension during each sampling period. Ultimately, the state must be in a space of dimension zero which is an equilibrium point in the state space.

Physical Realization of the Control Signal

According to equation 26, the optimal control signal is obtained by computing the first component in the canonical co-ordinate system of the state of the plant at each sampling instant. In general, the state variables cannot be directly measured in the canonical co-ordinate system; on the other hand, any component of a vector in one co-ordinate system can be expressed as a linear combination of the components of the same vector in a different co-ordinate system. Thus it is always possible to write:

$$m(t)=-y_1(t)=-\alpha_1x_1(t)=\ldots-\alpha_nx_n(t) \quad \text{(all } t\text{)} \qquad (30)$$

where the α_i are real constants and the x_i are directly measurable state variables. To calculate these constants, observe that

$$\sum_{i=1}^{n}x_i\mathbf{u}_i=\sum_{i=1}^{n}y_i\mathbf{v}_i \qquad (31)$$

by definition. Moreover, since $\mathbf{v}_j=G(-jT)\mathbf{h}(T)$

$$\mathbf{v}_j=\sum_{i=1}^{n}p_{ij}\mathbf{u}_i$$

where

$$p_{ij}=\sum_{l=1}^{n}g_{il}(-jT)h_l(T) \qquad (32)$$

The $g_{il}(-jT)$ are the elements of the transition matrix in the physical co-ordinate system as given by equation 6. Likewise, the $h_j(T)$ are components of $\mathbf{h}(T)$ in the physical co-ordinate system as given by equation 8. Substituting into equation 31

$$\sum_{i=1}^{n}y_i\mathbf{v}_i=\sum_{i=1}^{n}\left(\sum_{j=1}^{n}p_{ij}y_j\right)\mathbf{u}_i=\sum_{=1}^{n}x_i\mathbf{u}_i \qquad (33)$$

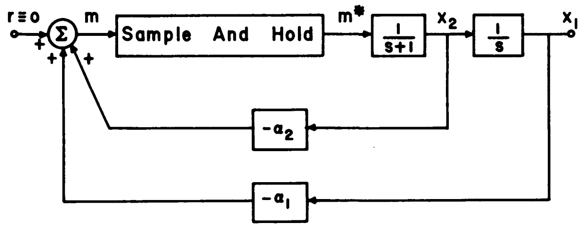
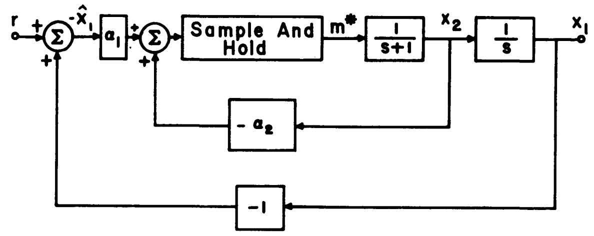


Fig. 5. (left). Physical realization of regulator system

Fig. 6 (right). Physical realization of follow-up system



Since the $\mathbf{u}_i$ are linearly independent, equation 33 implies

$$\sum_{j=1}^{n} p_{ij} y_j = x_i \text{ (all } i) \tag{34}$$

Thus the α_i can be obtained by solving the set of linear equations 34 for y_1. (Because the $\mathbf{v}_i$ were chosen so as to form a co-ordinate system, equation 34 always has a unique solution.)

In the running example, see equation 17,

$$P = \begin{Vmatrix} T+1-e^T & T-e^{2T}+e^T \\ e^T-1 & e^{2T}-e^T \end{Vmatrix}$$

Therefore

$$y_1 = \frac{p_{22}}{\det P} x_1 + \frac{(-p_{12})}{\det P} x_2$$

Comparing with equation 30, the final result is seen to be

$$\alpha = e^T/T(e^T-1),$$
$$\alpha_2 = (e^{2T}-e^T-T)/T(e^T-1)^2$$

Fig. 5 shows the block diagram of the optimal regulator system when $r(t)=0$.

Nonzero Inputs

There are several interesting problems which must be discussed if the solution for the optimal regulator system is to be extended to follow-up systems.

Arbitrary Step Inputs

This problem is equivalent to making the output c equal to some arbitrary reference input r in such a way that the plant is in equilibrium with zero system error.

In the regulator problem previously treated, this equilibrium point with zero error is the origin of the state space.

Provided the state x_1 is the system output, zero error at equilibrium is achieved when $x_1 = r$ and the other components of the state vector are determined by the equilibrium conditions.

If, as in the running example, the plant contains at least one integration, then the differential equations of the plant in terms of the state variables may be written as follows:

$$dx_1/dt = x_2 \tag{35}$$

plus other equations which do not involve the state variable x_1. Form the modified state variables

$$\hat{x}_1 = x_1 - r$$
$$\hat{x}_k = x_k \qquad 1<k\leq n \tag{36}$$

Since $dr/dt=0$, substituting equation 36 into equation 35 leads to an equation of the same form as equation 35 in $\hat{x}_1$. The rest of the equations of the plant are unaffected. In equilibrium,

$$\hat{x}_1 = r$$
$$\hat{x}_k = 0 \qquad 1<k\leq n$$

which shows that the follow-up action is achieved with the regulator design by utilizing the modified state $x_1 - r$ in place of x_1.

Thus the regulator system of Fig. 5 which contains an integration is converted to a follow-up system simply by taking the difference between state x_1 and the input r to create the modified state $\hat{x}_1$ prior to multiplying by the feedback coefficient α_1. This system is shown in Fig. 6.

In those cases where the plant does not contain an integration proceed as follows. From Laplace transform considerations note that in the steady state the control signal is given by

$$m_{ss} = r/g(0) \tag{37}$$

where $g(s)$ is the transfer function of the plant. Using equation 4, it is seen that the equilibrium state $\mathbf{z}$ is given by

$$\mathbf{z} = G(T)\mathbf{z} + m_{ss}\mathbf{h}(T) \tag{38}$$

By a shift in the origin of the canonic co-ordinate system, it follows that any state may be written in the form

$$\mathbf{x}(k) = G(-nT)\mathbf{z} + \sum_{i=1}^{n} y_i(k)\mathbf{v}_i \tag{39}$$

where the $y_i(k)$ are unique real constants. If, as before, the control signal is given by

$$m(k) = -y_1(k) \tag{26}$$

$y_1(k)$ to be determined by equation 39, then after one sampling period

$$\mathbf{x}(k+1) = G(-(n-1)T)\mathbf{z} + \sum_{i=1}^{n} y_i(k+1)\mathbf{v}_i \tag{40}$$

where $y_i(k+1) = y_{i+1}(k)$. Multiplying both sides of equation 38 by $G(-nT)$ and rearranging shows

$$G(-nT)\mathbf{z} = G(-(n-1)T)\mathbf{z} + m_{ss}\mathbf{v}_n$$

Substituting this into equation 40 results in

$$\mathbf{x}(k+1) = G(-nT)\mathbf{z} + \sum_{i=1}^{n} y_i(k+1)\mathbf{v}_i$$

where

$$y_i(k+1) = y_{i+1}(k) \quad i=1, \ldots, n-1 \tag{41}$$
$$y_n(k+1) = -m_{ss}$$

This shows that expression 39 can be used to compute $y_1(k)$ for every value of k. The foregoing may be summarized as follows.

For step inputs of arbitrary magnitude, the optimal control function is given by

$$m(k) = -y_1(k)$$
$$= \sum_{i=1}^{n} \alpha_i \left(\sum_{j=1}^{n} g_{ij}(-nT)z_j - x_i \right) \tag{42}$$

$\mathbf{z}$ being determined by equation 37, and the αi by equation 34.

Moreover, if j is the largest integer in equation 39, such that $y_j \neq -m_{ss}$, then equilibrium is reached in exactly j steps. This shows that the modified state variables are given in general by $\hat{\mathbf{x}} = \mathbf{x} - G(-nT)\mathbf{z}$.

If the plant does not contain an integration, then its transition matrix has no characteristic root equal to one. Consequently the matrix $1-G(T)$ can be inverted and $\mathbf{z}$ can be found from

$$\mathbf{z} = m_{ss}(1-G(T))^{-1}\mathbf{h}(T) \tag{43}$$

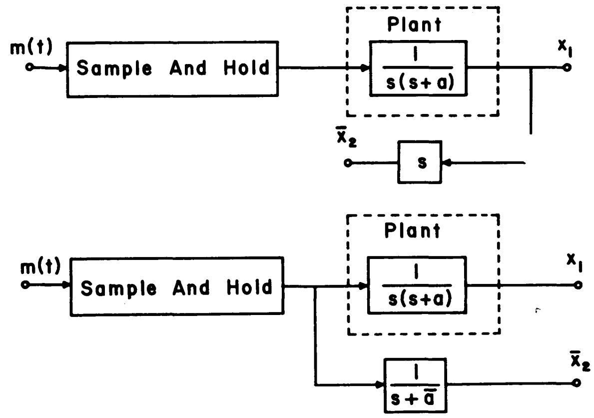
In this case, however, the accuracy of the output in the steady state depends on the accuracy with which the transition matrix is known. Such systems are called calibrated and are frequently undesirable in practice.

If the plant does contain an integration, then $m_{ss}=0$. In this case, the transition matrix has a characteristic root equal to one and therefore there is a vector $\mathbf{z}$ satisfying

$$G(T)\mathbf{z} = \mathbf{z} \text{ and hence also } G(-nT)\mathbf{z} = \mathbf{z} \tag{44}$$

By substituting these relations into equation 42, it follows that the modified state

Fig. 7. Methods for simulating state variables



variables are now given by $\hat{\mathbf{x}}=\mathbf{x}-\mathbf{z}$. Since it is desired to have $x_1=c=r$ in the steady state, when $\hat{\mathbf{x}}=0$, $z_1=r$ and the rest of the components of the vector $\mathbf{z}$ are then determined uniquely by equation 44. Hence if there is an integration in the plant, the steady-state accuracy does not depend on precise knowledge of the transition matrix.

Arbitrary Polynomial Inputs

Suppose the permissible types of inputs are given by equation 3. Assume, for simplicity, that the system contains at least m integrators. Then the differential equations of the plant in terms of the state variables are

$$\begin{aligned} &dx_1/dt=x_2\\ &dx_2/dt=x_3\\ &\quad\cdots\\ &dx_m/dt=x_{m+1} \end{aligned} \tag{45}$$

plus other equations which do not involve the state variables $x_1, \ldots, x_m$. If the modified state variables are defined as

$$\begin{aligned} &\hat{x}_1=x_1-r\\ &\hat{x}_2=x_2-dr/dt\\ &\quad\cdots\\ &\hat{x}_m=x_m-d^mr/dt^m\\ &\hat{x}_k=x_k \qquad m<k\leq n \end{aligned} \tag{46}$$

then substituting into equation 45 shows that the modified state variables obey the same differential equations. Hence in equilibrium when $\hat{\mathbf{x}}=0$,

$$\begin{aligned} &x_1=r\\ &x_2=dr/dt\\ &\quad\cdots\\ &x_m=d^mr/dt \end{aligned}$$

which shows that the output will be identical with the input after, at most, n sampling periods. Note, however, that equations 46 require the knowledge of the derivatives of the input. This may be difficult to obtain in some cases.

Some Practical Aspects

Simulation of the State Variables

In some practical cases it is not possible to measure directly all of the state variables. With such plants other elements must be added in order to obtain approximations, $\bar{x}_i$, to those state variables which are inaccessible. Two common methods of obtaining the approximate state variables are the following.

1. If $g_{ij}(s)$ is the transfer function relating state variable x_i, which is measurable, to state variable x_j, which is not measurable, then $\bar{x}_j$ may be obtained by passing the measurable signal x_i through the transfer function $1/g_{ij}(s)$.

2. Construct a model or analog of the plant in which all of the desired state variables can be measured, and subject it to the same input sequence as that applied to the actual plant. Because the output of the plant model is required only at sampling instants, the model can be simulated also by a digital computer.

Both of these methods have been quite well known in the control systems art, although their full significance was perhaps not understood in the past. In fact, method 1 includes the very popular tachometer-feedback technique of stabilization; see also the experimental results which follow. The use of models to simulate the behavior of the plant has also received much attention in recent years; see, for instance, a recent paper by Reswick.[8] Finally, common methods of compensation of sampled-data systems may also be regarded as special cases of method 2.

To illustrate the application of these methods, assume that in the running example it is not possible to directly measure the state variable x_2. Fig. 7 shows how the variable $\bar{x}_2$ may be obtained by either method 1 or 2. Using method 1, the state variable x_1, which can be measured, is differentiated to obtain an approximation to x_2 which, by equation 5, is dx_1/dt. Using method 2, the sampled control signal $m^*(t)$ is applied to the element whose Laplace transform is $1(s+\bar{a})$.

In method 1 the transfer functions required to obtain the approximations have at least a simple pole at ∞. In practice, such elements are not realizable although tachometers and accelerometers are reasonable approximations over a certain range of frequencies. Because of this difficulty of physical realization, method 2 is frequently more satisfactory in practice.

In using method 2, note that $\bar{x}_2=x_2$ only if the model is identical with the plant element it is simulating $(\bar{a}=a)$ and only if there are no (disturbance) inputs applied to the plant which do not also affect the model in the same way. Clearly, the best results are achieved by measuring as many state variables as possible. In some practical cases, it may be possible to ignore some of the state variables, since they may correspond to time constants very much smaller than T and therefore differ negligibly from their steady-state values at the sampling instants. A detailed study of the approximation problem will be made at a later time.

Compensation of Plant Time Delay

Suppose the control signal affects the plant only after a definite time delay τ. This may happen particularly often in process control problems where the time delays are caused by transportation of liquids and gases through long pipes. As is well known, such time delays may cause serious problems in the stabilization of the closed-loop system. A simple way of canceling the detrimental effect of time delay as far as over-all system stability is concerned is the following.

Select the sampling period T so as to have $\tau=qT$, where q is a positive integer. Since the control signal affects the plant only after qT seconds, it should be computed not according to the present state of the plant but according to the anticipated state of the plant qT seconds later. Thus instead of equation 26 the control signal should be given by:

$$m(k)=-y_1(k+q) \tag{47}$$

where $y_1(k+q)$ is the first component of the approximate state vector $\mathbf{x}(k+q)$ in the canonical co-ordinate system.

$$\bar{\mathbf{x}}(k+q)=G(qT)\mathbf{x}(k)+\sum_{i=1}^{q}m(k-i)G((i-1)T)\mathbf{h}(T) \tag{48}$$

Equations 48 characterize the anticipator in Fig. 8, which gives the value of the state q sampling periods after the present sampling instant, provided that no disturbances act on the plant during the interval $kT\leq t\leq(k+q)T$. The signifi-

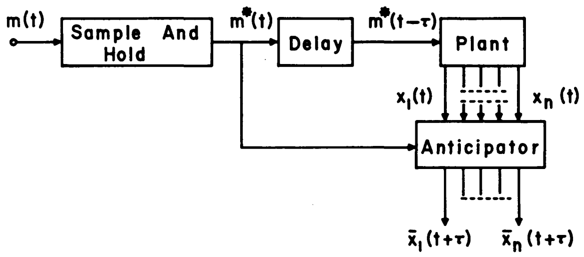


Fig. 8. Compensation for delay

Table I

	K_1	K_2
Calculated (based on measurement of motor transfer function)	14.5	0.0712
Experimental	13.9	0.074

cance of equations 47 and 48 is simply that the effect of the time delay is canceled out as far as the autonomous (i.e., in the absence of inputs) operation of the control system is concerned. Because the plant is assumed linear, the effect of disturbances during the anticipation interval is superposed on the autonomous behavior. If these disturbances are completely random, that is, completely unpredictable, then a system containing a plant with time delay clearly cannot have optimally fast disturbance response. If the disturbances can be predicted in some statistical sense, then the predictor equations for the disturbance are to be incorporated in the equations of the anticipator, thereby improving the disturbance response of the system.

The anticipator is physically realizable by either analog or digital computation since equation 48 requires only knowledge of the present state, a sequence of q past control signals, and knowledge of the elements of the transition matrix. The latter requirement shows that this compensation scheme is of the "calibrated" type. However, as long as the plant contains an integrator, the output will be fed back with coefficient -1, so that the steady-state accuracy does not depend on the precise knowledge of the transition matrix.

This method of compensation of the effects of time delay was first suggested in a general form by Bass.[9]

Experimental Results

An instrument servomechanism shown in Fig. 9 was constructed to perform in the optimal fashion. The components of this system are conventional instrument servo hardware except for the sample-and-hold element which employs a slight modification of the 6-diode gate of reference 10. The error signal was d-c and was sampled and held at a rate of 10 cycles per second. This signal was then used to modulate a 60-cycle-per-second carrier in order to drive a 2-phase servomotor. As is well known, the signal-frequency transfer function of such motors is approximately of the same form as that of the running example. One state variable was the output; an approximation to the other was obtained by attaching a tachometer to the output shaft of the motor.

The adjustment of K_1 and K_2 was found to be extremely simple. In fact, optimal settings could be made in a few minutes by persons unfamiliar with the operation of the system, once the criterion of "dead-beat response" was understood. The calculated and experimentally derived settings for optimal performance were found to differ by only a few per cent. The discrepancy is due mainly to the neglect of additional state variables in the system. See Table I.

Conclusions

This paper presents a method for designing a sampled-data control system having the following properties:

1. The design procedure is applicable to any linear plant with a rational transfer function.

2. The response of the system to any disturbance in state or step input requires at most n sampling periods, n being the order of the plant.

3. After at most n sampling periods following the application of a disturbance or step, the system reaches equilibrium with zero error. Hence the system is not subject to intersample ripple.

4. The design of the system is based directly on the concept of the state and can be readily implemented if all states are accessible.

5. If not all states are accessible the design should incorporate some method of simulating the state which usually requires analog or digital computation.

6. Additional computation is necessary in order to compensate the effect of time delay.

7. The design results in a linear system. It can be shown, however, that even if the control signal is limited by saturation, i.e.,

$$m^*(t)| \leq M$$

the system is still monostable.[11] In other words, any initial state is reduced to equilibrium (in a finite number of steps) if the system is designed on a linear basis.

8. The optimal design may be extended to include the case of saturation. In this case it can be shown[12] that the optimal control signal $m(k)$ is in general a nonlinear function of the state vector $\mathbf{x}(kT)$.

It is hoped that the present discussion will help in simplifying the design of computer-control systems and lead to a better understanding of the reasons why real-time computation can contribute to better control. Using many other types of criteria for optimality, one can arrive at other optimal designs such that the control signal is a linear function of the state.[13] Reference 13 also gives an alternate method for the derivation of the feedback constants α_i in equation 30.

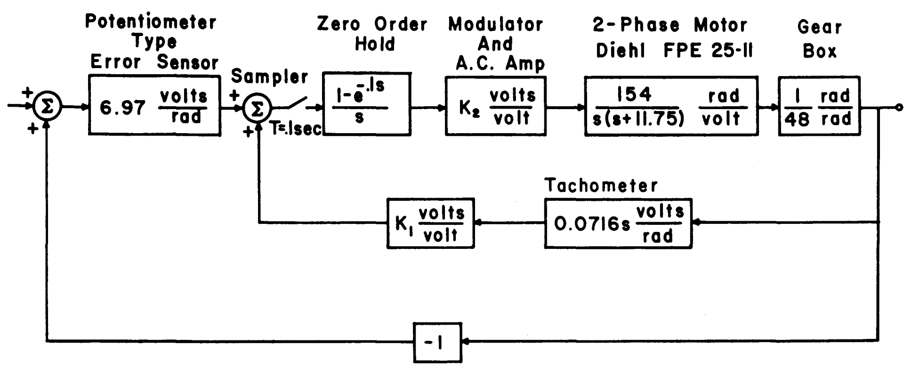


Fig. 9. Block diagram of practical system

References

1. The Pulse Transfer Function and Its Application to Sampling Servo Systems, R. H. Barker. *Proceedings*, Institution of Electrical Engineers, London, England, *Monograph no. 43*, pt. 4, 1952.

2. Sampled-Data Processing Techniques for Feedback Control Systems, A. R. Bergen, J. R. Ragazzini. *AIEE Transactions*, vol. 73, pt. II, Nov. 1954, pp. 236–47.

3. Factors in the Design of Digital Controllers for Sampled-Data Feedback Control Systems, J. E. Bertram. *Ibid.*, vol. 75, pt. II, July 1956, pp. 151–59.

4. Compensation of an Error-Sampled System

BY A MULTIRATE CONTROLLER, G. M. Kranc. *Ibid.*, vol. 76, pt. II, July 1957, pp. 149–59.

5. SYNTHESIS AND CRITICAL STUDY OF SAMPLED-DATA CONTROL SYSTEMS, E. I. Jury. *Ibid.*, vol. 75, pt. II, July 1956, pp. 141–51.

6. *Discussion* by R. E. Kalman of reference 2, pp. 245–46.

7. THEORY OF ORDINARY DIFFERENTIAL EQUATIONS (book), E. A. Coddington, N. Levinson. McGraw-Hill Book Company, Inc., New York, N. Y., 1955, chap. 3.

8. DISTURBANCE RESPONSE FEEDBACK—A NEW CONCEPT, J. B. Reswick. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 78, 1956, pp. 153–62.

9. A GENERALIZATION OF THE FUNCTIONAL RELATION $Y(t+s)=Y(t)\cdot Y(s)$ TO PIECEWISE-LINEAR DIFFERENCE-DIFFERENTIAL EQUATIONS, W. R. Bass. *Quarterly of Applied Mathematics*, Providence, R. I., vol. 14, 1957, pp. 415–17.

10. PULSE AND DIGITAL CIRCUITS (book), J. Millman, H. Taub. McGraw-Hill Book Company, Inc., New York, N. Y., 1956, p. 445.

11. PHYSICAL AND MATHEMATICAL MECHANISMS OF INSTABILITY IN NONLINEAR AUTOMATIC CONTROL SYSTEMS, R. E. Kalman. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 79, 1957, pp. 553–56.

12. OPTIMAL NONLINEAR COMPENSATION OF SATURATING SYSTEMS BY INTERMITTENT ACTION, R. E. Kalman. *WESCON Convention Record*, Institute of Radio Engineers, New York, N. Y., 1957.

13. OPTIMAL SYNTHESIS OF LINEAR SAMPLING CONTROL SYSTEMS USING GENERALIZED PERFORMANCE INDEXES, R. E. Kalman, R. W. Koepcke. *Transactions*, American Society of Mechanical Engineers, New York, N. Y. (To be published in vol. 81, 1959).

Trends in Automatic Hoist Controls

G. L. TILEY
ASSOCIATE MEMBER AIEE

E. ZUCKER
ASSOCIATE MEMBER AIEE

Synopsis: The paper outlines the problems facing mine hoist engineers in the design and application of automatic hoist equipment. Duty cycle requirements, types of hoists, and the advantages of mine hoist automation are discussed. Particular mention is made of available a-c and d-c control systems. Overriding speed reference equipment and the safety features of modern automatic controls are described. A discussion on hoist brakes is also included, as these are an integral part of each hoist control design. It is the purpose of this paper, and its bibliography, to provide the designer with many references and with a detailed outline of existing thinking on the subject.

MOST of the mineral requirements of North America have been obtained, up to the present, from ore bodies that can be worked by open cast methods. Supplies of readily available ore are running out, however, and it is becoming necessary to go deeper into the earth for the required materials. As a result, increasing attention is being paid to the design of mine hoists, their drives, and associated control systems.

Control Requirements

A thorough investigation of the hoisting requirements of any particular installation should be made before hoisting or control equipment is selected. References 1 to 5 should be consulted on the general subject of mine hoists and mine-hoist duty cycles.

The basic requirements of the hoisting cycle from the control point of view are simple. The hoist must be accelerated, run at full speed, decelerated, and finally stopped. On automatic control, the difficulties arise chiefly in the required degree of accuracy with which the hoist conveyances must be stopped. A multilevel cage hoist is often required to have a landing accuracy of about one inch in a 4,000-foot shaft under all conditions of speed and loading. The difficulty of accomplishing this is further aggravated by the fact that the connection between the hoist and the conveyances is by relatively elastic ropes which are also subject to permanent deformation (rope stretch) creeping around the circumference of a Koepe wheel, etc.

The controls must also fail to safety under all imaginable circumstances. The effectiveness with which safety requirements have been met on present-day mine hoists is illustrated by the fact that these have a better safety record than any other known form of transportation.

Types of Hoists

Figs. 1(A), 1(B), and 1(C) show the three most common types of hoists. The conventional drum hoist [Fig. 1(A)] is admirably suited for service in shafts less than 2,500 feet deep. Two shaft conveyances are attached to ropes that coil on separate drums so arranged that when one conveyance is being raised, the other is being lowered.

At depths of over 2,500 feet, the Koepe or friction hoists become attractive. These hoists, shown in Figs. 1(B) and 1(C), are relatively new to North America. The head rope or ropes are not attached to the hoist wheel but are only in contact with it over an arc of 180 to 210 degrees. Forces are transmitted from the wheel to the rope by friction only. Provided the ratio of rope tensions on either side of the wheel is carefully controlled, there is no danger of the rope slipping. A tail rope is connected between the bottoms of the conveyances in order to keep these tension ratios within safe limits. As a result, the only out-of-balance is the useful load being raised. On a conventional drum hoist, the unbalanced rope in the shaft imposes additional torques on the hoist motor.

When large tonnages have to be handled, the size of the single head rope required by either a drum hoist or a ground-mounted Koepe hoist becomes impractical. A Koepe hoist can, however, be mounted in a tower directly over the shaft, as shown in Fig. 1(C) and then two or four head ropes can be used. This system is similar to a multirope elevator.

As has already been noted, the unbalanced rope weight imposes additional loads on the conventional drum hoist drive. The inertia of the Koepe hoist is also considerably less than that of a double drum hoist, so the inertia loads during acceleration and deceleration will also be reduced.

Duty cycles of a Koepe and a drum hoist are compared in Fig. 2. Both hoists are designed for hoisting 400 tons per hour from a depth of 3,500 feet, using 13-ton-capacity skips. The drum hoist requires a single 2-inch-diameter rope and 14-foot-diameter drums. The friction hoist employs four ropes, each of $1^3/_8$-inch diameter, and has a drum diameter of only 10 feet 6 inches. The former hoist requires a driving motor of 3,500 horsepower (6,500-horsepower peaks) while the latter only needs 2,500 horsepower (3,500-horsepower peaks).

It is obvious that the Koepe hoist is very attractive at these depths. From the control engineer's point of view, however, the system introduces additional complexities. Torques, in particular braking torques, must be carefully controlled and periodical recalibration of the control equipment is necessary to compensate for rope creep.

Paper 58-1210, recommended by the AIEE Mining Industry Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Pittsburgh, Pa., October 26–31, 1958. Manuscript submitted June 26, 1958; made available for printing August 21, 1958.

G. L. TILEY and E. ZUCKER are with the Canadian Westinghouse Company, Ltd., Hamilton, Ont., Canada.