

FEEDBACK CONTROL IN PHYSICAL SCIENCE:  
A LAYMAN'S APPROACH

W. STEPHEN CHEUNG

W. W. HANSEN LABORATORIES OF PHYSICS  
STANFORD UNIVERSITY  
STANFORD, CALIFORNIA 94305 U.S.A.

*in*

*Proceedings of the 1983 International School and Symposium on  
Precision Measurement and Gravity Experiment, Taipei, Republic of  
China, January 24 - February 2, 1983, ed. by W.-T. Ni (Published  
by National Tsing Hua University, Hsinchu, Taiwan, Republic of  
China, June, 1983)*

OUTLINE

I. Introduction .....	51
II. Open and Closed Loop Systems .....	51
III. Positive and Negative Feedback .....	52
IV. Laplace Transform and Transfer Functions .....	52
V. Connections of Transfer Functions .....	56
VI. Properties of a Closed Loop System .....	60
VII. Stability of a Physical System .....	65
VIII. Methods of Stability Study .....	69
Concluding Remarks .....	71
Acknowledgement .....	71
Suggested References .....	72
References .....	72

## FEEDBACK CONTROL IN PHYSICAL SCIENCE: A Layman's Approach

W. Stephen Cheung

W. W. Hansen Laboratories of Physics  
Stanford University, Stanford, CA 94305I. Introduction

The use of feedback and control technology has grown so much that it now touches every field of science. It is therefore impossible within the scope of this article to present a rigorous or detailed treatment of this subject. Rather, we intend to use simple examples to bring out the important aspects of control theory and hope the readers will investigate each aspect more deeply whenever and wherever desired. The author will be more than satisfied if the readers come away with a reasonable exposure to subject and recognize the need for the synthesis of a control system when they encounter certain design problems.

II. Open and Closed Loop Systems

Systems in which the output quantity has no effect on the input quantity are termed open loop control systems. Often, physical measurements are outputs of open loop systems.

In Fig. 1, the ammeter is the simple coil-needle type. Ideally, its presence or absence does not change the current value to be measured. The needle movement, which gives a reading of the current, has no effect on the current itself. Note that a modern ammeter, e.g., a digital one, is not a good example of an open loop system because the electronics design inside is the feedback type.

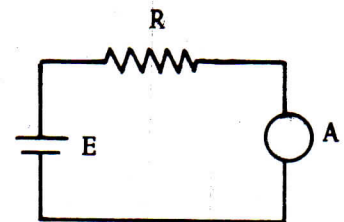


Fig. 1

Unfortunately, there are a large number of open loop systems which do not respond in a desirable manner unless their input quantity can be regulated. A system in which the output and input values are compared and "actions" taken so that a desired output value, or behavior, is maintained is known as a closed loop system. For example, a person may wish to pick up a pencil on a table. His eyes first locate the pencil and guide his hand toward it. If one follows the motion of the hand, one may find, for example, that if the hand moves a little to the left of the desired straight line, the eye/brain system will move the hand back to the right a little. Eventually, it converges on the object.

In the above example, the person's eyes detect, or sense, the deviations, i.e., errors of the hand from the desired track. The instant-



aneous location of the hand relative to the pencil is the input to the eyes/brain system and the subsequent movement of the hand is the output. Clearly, the output is somehow fed back as part of the system's input. This closed loop system becomes open as soon as the person shuts his eyes. It is left as exercises for the readers to think of examples of other open and closed loop systems we find in everyday life and various areas of science.

To pursue the above example further, the amount of deviation from the prescribed course is small if the eyes and the hand are "quick". Technically, this means that the system has a fast response and/or high sensitivity. As soon as a detectable error is sensed, the eyes/brain system put some "emphasis" on correcting it. That is, it processes the error signal with a certain amount of gain. Hence, sensitivity of a system is related to its overall gain.

The readers must not get the impression that a system with infinitely large gain is the most desirable. In this same example, a person with too high gain in his eyes/brain system will over-correct his hand and result in even larger deviation in the opposite direction. As the eyes see this even larger error signal, the over-correction becomes still larger. Consequently, the hand "oscillates" about the desired course with increasing amplitudes. Therefore, too much gain is not good for a system and oscillation may result.

The hand cannot react "instantaneously" to the error signal input, nor can it move at infinite speed. The mechanism of the human body imposes an upper limit to the hand's response as well as a time lag between the sensed error and the corresponding correction. Hence, a physical system is partly characterized by its frequency response and rise time.

### III. Positive and Negative Feedback

In the example discussed above, the closed loop control system tends to make the final error signals vanishingly small. In other words, the output moves in the opposite direction to the sensed error in order to counteract the error. This is known as negative feedback.

A positive feedback system is one in which the output correction goes in the same direction as the error, thereby causing the error amplitude to grow further. This is not meant to imply that all positive feedback systems are undesirable though. Without positive feedback, wave generators and some other types of systems would not work.

### IV. Laplace Transform and Transfer Functions

The simple examples discussed above, one hopes, provided the reader with an intuitive definition of the following terms: open loop system,



closed loop system, positive feedback, negative feedback, gain, error signal, oscillations, frequency response, and rise time. After reflecting upon the general meaning of these terms, we ask the readers to look up their accepted definitions in textbooks for their precise meaning. The readers should remember that different authors may use slightly different definitions of these terms.

Give this background, we shall now prepare a mathematical foundation for use in the design and analysis of feedback systems. Those who have had an introductory course in differential equations know that the "trick" used for solving a linear differential equation of the kind

$$\sum_{i=0}^n a_i \frac{d^i}{dt^i} y(t) = F(t) \quad (1)$$

is to substitute  $y(t) = e^{st}$  and end up solving for the roots of the characteristic equation

$$\sum_i^n a_i s^i = 0. \quad (2)$$

The advantage lies in the reduction of the problem to a relatively simple algebraic manipulation.

For any physical system that can be modelled linearly, a differential equation like Eq. (1) is usually obtained. In general, the left hand side of Eq. (1) describes the natural behavior of the system while the right hand side describes external parameters.

A more systematic approach to solving Eq. (1) is the use of the Laplace transformation. We shall use a familiar example to illustrate what we have discussed in this section so far.

Fig. 2 describes an external force  $F(t)$  acting on a damped harmonic oscillator whose equation of motion can be written as

$$\ddot{\chi} + 2\zeta\omega_n \dot{\chi} + \omega_n^2 \chi = \frac{F(t)}{m} \quad (3)$$

where  $\omega_n^2 = k/m$  and  $\zeta$  is suitably defined as the damping coefficient.

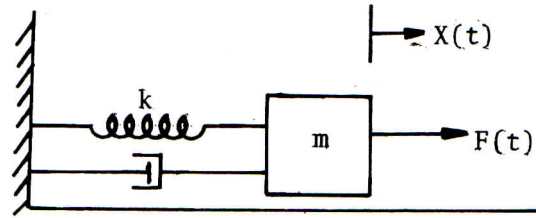


Fig. 2

When  $x(0) = 0$  and  $\dot{x}(0) = 0$ , Laplace transform of Eq. (3) gives

$$x(s)[s^2 + 2\zeta\omega_n s + \omega_n^2] = \frac{F(s)}{m} \quad (4)$$

This can be rewritten as

$$\frac{x(s)}{F(s)} = \frac{1}{m(s^2 + 2\zeta\omega_n s + \omega_n^2)} = \frac{1}{m(s-a)(s-b)} \quad (5)$$

Here  $a$  and  $b$  are roots of the characteristic equation

$$s^2 + 2\zeta\omega_n s + \omega_n^2 = 0. \quad (6)$$

The roots  $a$  and  $b$  can be real or complex depending upon the magnitude of  $\zeta$ .

The ratio  $x(s)/F(s)$  is known as the transfer function. Fig. (3) gives a good idea of what a transfer function means. In Fig. 3, the box represents the spring system. When an external force  $F(s)$  is applied to this spring system,

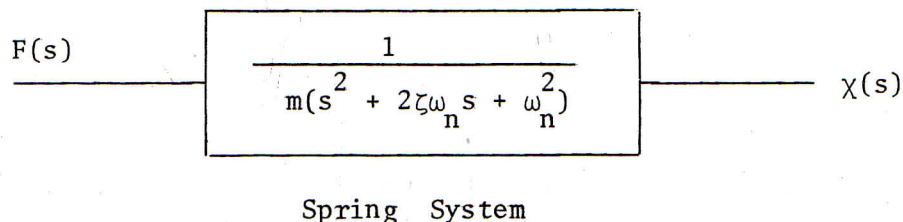


Fig. 3

the displacement behavior,  $\chi(s)$ , of the object  $m$  can be obtained. Note that the independent variable is  $s$  (the Laplace variable), whose dimension is frequency, instead of time  $t$ .

If a  $s$ -plane is constructed with  $\text{Re}(s)$  as the horizontal axis and  $\text{Im}(s)$  as the vertical axis, we can plot the roots  $a$  and  $b$  graphically. The location of  $a$  and  $b$  are called poles of the transfer function  $1/(s-a)(s-b)$ . The poles are conventionally marked as crosses "x" on the  $s$ -plane. In the case of the damped oscillator, the poles  $a$  and  $b$  can be both real ( $a_1$  and  $b_1$ ) or complex ( $a_2$  and  $b_2$ ) as illustrated in Fig. 4. Note that  $a_2$  and  $b_2$  form a conjugate pair.

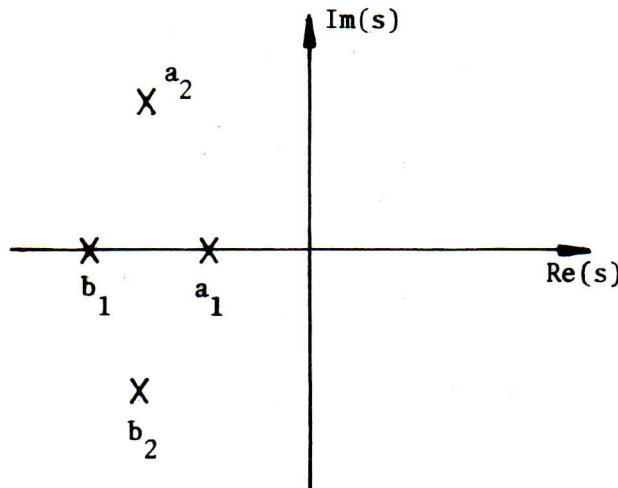


Fig. 4

Let us write

$$a = \alpha + j\beta,$$

and

$$b = \alpha - j\beta$$

where

$$j = \sqrt{-1}.$$

Mathematically, the meaning of the poles is that when we inverse transform  $(s)$  to  $(t)$ , we get

$$\chi(t) = Ae^{(\alpha+j\beta)t} + Be^{(\alpha-j\beta)t} + \text{other terms.} \quad (7)$$

One should not think only in terms of infinities for the points  $s=a$  and  $s=b$ , but rather that the location of the poles, i.e., the magnitudes of  $\alpha$  and  $\beta$ , in general, determine the frequency response of the system. If  $\alpha=0$ , the system is purely oscillatory. If  $\alpha < 0$  and  $\beta \neq 0$ , the oscillations will decay. If  $\alpha < 0$  and  $\beta$  is imaginary, there will be no oscillations.

A transfer function is not always composed of poles. If  $\chi(0) \neq 0$  and



$\dot{\chi}(0) \neq 0$  in our oscillator example, Eq. (4) becomes

$$\chi(s)[s^2 + 2\xi\omega_n s + \omega_n^2] - s\dot{\chi}(0) - \chi(0) = \frac{F(s)}{m},$$

and

$$\chi(s) = \frac{F(s) + s\dot{\chi}(0) + \chi(0)}{m(s^2 + 2\xi\omega_n s + \omega_n^2)} \quad (8)$$

Without going further into this example, we shall state that, in general, a transfer function is given by

$$G(s) = \frac{\prod_{i=0}^n (s + z_i)}{\prod_{j=0}^m (s + p_j)} \quad (9)$$

where  $p_j$ 's are the poles and  $z_i$ 's are known as zeroes of the transfer function  $G(s)$ . On the  $s$ -plane, zeroes are labelled as circles "o" as shown in Fig. 5.

We shall also state that for most systems,  $n \leq m$ . It is also sufficient, but important, to point out to the readers that the zeroes are typically due to the boundary conditions so they are studied for the transient behavior of the system.

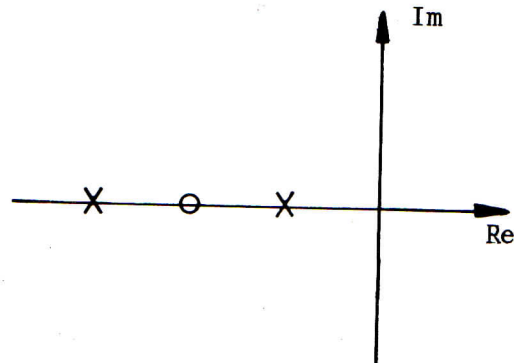


Fig. 5

## V. Connections of Transfer Functions

A useful feature of working in the  $s$ -plane is that we can draw and connect block diagrams of the transfer functions for the system and more easily understand the signal flow.

Consider the following example which uses the oscillator discussed above (Fig. 6). Suppose the position of the oscillating mass is monitored with a device\* whose output is a voltage which is some function of  $\chi(t)$ . This can easily be done capacitatively or optically. Fig. 7A gives the

\* A position monitor device with a voltage output is known as a transducer.

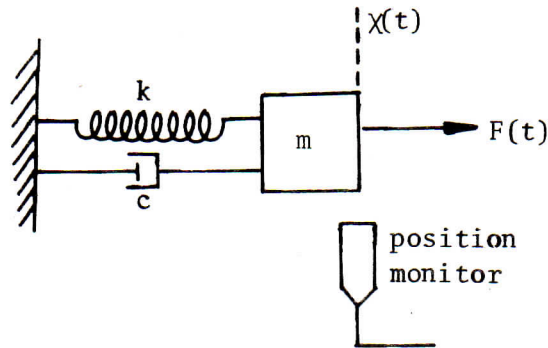
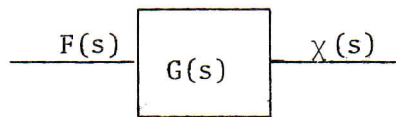


Fig. 6

transfer function of the oscillator without the transducer and Fig. 7B describes how the transfer function of the transducer is connected to that of the oscillator.



$$G(s) = \frac{1}{m(s-a)(s-b)}$$

Fig. 7A

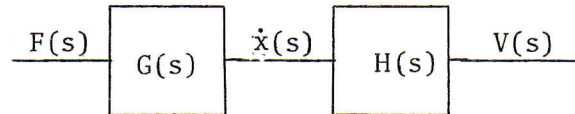


Fig. 7B

We shall not bother with the exact form of  $H(s)$  because it depends upon the physics of the device. However, we do get the following information from the block diagram at this point. A force function  $F(s)$  acts on the mass  $m$  whose subsequent motion,  $x(s)$ , is monitored to give rise to some voltage output,  $V(s)$ .

Even though we shall say nothing about the poles and zeroes of  $H(s)$ , the transfer function of the transducer, we can assume, in general, that  $H(s)$  may contain an amplification factor and other terms. The amplification factor is linear, i.e., a proportionality constant converting position into voltage, while the other factors may impose a frequency range within which the transducer operates.

It should be noted that the system shown in Fig. 6 is still an open loop system, though. Our next task is to make a closed loop system out of the same system.

Consider a mass  $m$  connected to one end of a spring, the other end of which is connected to the floor. There is a damping mechanism which we call  $c$  (Fig. 8). Ideally, the mass  $m$  would be inertially at rest at its equilibrium position. However, floor vibrations and external perturbations will cause unwanted motion on  $m$ .

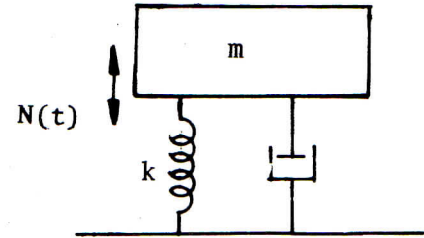


Fig. 8

Let  $N(t)$  be the force due to all the disturbances and  $\chi$  be the position deviation from the equilibrium point. The equation of motion of  $m$  is given by

$$m\ddot{\chi} + c\dot{\chi} + k\chi = N(t). \quad (10)$$

We can define the natural frequency and the natural angular frequency of the oscillator as

$$f_0 = \frac{1}{2\pi} \sqrt{k/m}, \quad (11)$$

and

$$\omega = \sqrt{k/m}. \quad (12)$$

Suppose the position of  $m$  can be monitored and converted into an electrical signal (voltage)  $V$ . We can then employ an electromechanical device (an actuator) that can convert  $V$  into a force  $\mathcal{F}$ . An example of such an actuator is a piezo electrical crystal. By making  $\mathcal{F}$  a special function of  $V$  and then applying it back to the mass  $m$  with suitable phase relationship and magnitude, we can naively expect that the effects of  $N(t)$  on  $m$  could be exactly cancelled by  $\mathcal{F}$ .

From the practical standpoint, it is difficult, if not impossible, to sense  $N(t)$  and construct  $\mathcal{F}(t) = -N(t)$ . It is, however, easy to construct  $\mathcal{F}(t) \propto \chi$ . Let us investigate this possibility. Fig. 9 shows the block diagram connection of this closed loop system.

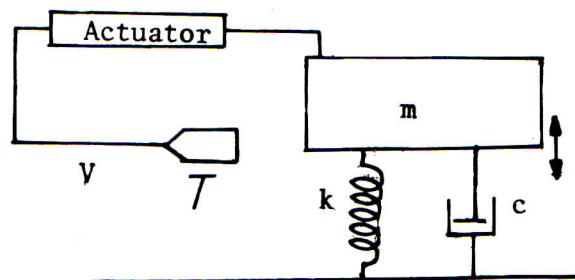


Fig. 9



The equation of motion of the oscillator in the presence of  $N(t)$  and  $\mathcal{J}(t)$  is now

$$m\ddot{\chi} + c\dot{\chi} + k\chi = N(t) + \mathcal{J}(t). \quad (13)$$

Let

$$\mathcal{J}(t) = K\chi. \quad (14)$$

Equation (13) can be rewritten as

$$m\ddot{\chi} + c\dot{\chi} + (k-K)\chi = N(t). \quad (15)$$

The effective spring constant is now  $(k-K)$  and the corresponding natural frequency  $f'_0$  is

$$f'_0 = \frac{1}{2\pi} \sqrt{\frac{(k-K)}{m}} \quad (16)$$

If  $k$  is approximately equal to  $K$ , the effective natural frequency  $f'_0$  will be approximately zero and the effective natural period ( $T'_0 = 1/f'_0$ ) will be very large. What one has achieved in this approximation is an extremely soft spring which has very little response to external disturbances of frequency higher than  $f'_0$ .

We have just described the basic principles of an "active" vibration isolation mechanism. The system is active because feedback is involved and external energy is "pumped" into the system through the actuator. It needs little emphasis on how valuable vibration isolation is to most experimentation. Typical shock absorbers (passive) reduces vibrations of frequencies larger than 30 Hz but not the low frequencies. Not that the vibrations can only be reduced but not totally eliminated. The explanation for this is the same as that for high or low pass electronic filters. The reduction is typically given in percentage, %, or decibel, dB.

Active vibration isolation has been studied and realized by many experimenters. Sakuma<sup>1</sup> of BIPM and Ritter<sup>2</sup> of Virginia employ piezoelectric crystals. The superspring of Faller<sup>3</sup> works on the same principle but employs a different actuator.

Our interest here, though, is not in the actual construction of any of the schemes mentioned above, but rather in the development of a sufficient block diagram model for the closed loop system. One way of connecting the transfer functions in this closed loop system is shown in Fig. 10. In this figure,  $Y(s)$  is the transfer function converting voltage  $V$  into force  $\mathcal{J}$ ; and it has a certain pole/zero distribution characterizing it.

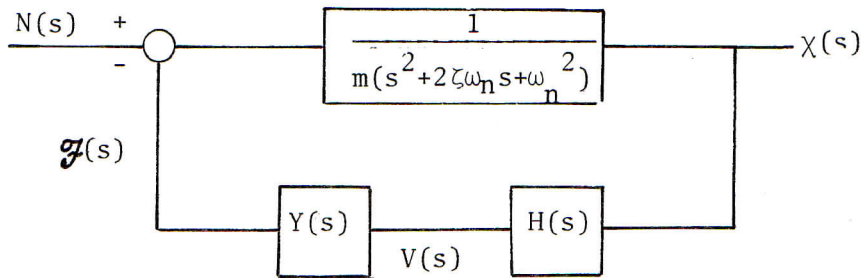


Fig. 10

The loop is closed at the point where the summation takes place. Note, however, that  $\mathcal{Z}$  is summed negatively to  $N(s)$ , i.e., the resulting  $\epsilon$  is given by

$$\epsilon = N(s) - \mathcal{Z}(s). \quad (17)$$

One can regard  $\epsilon$  as an error value which the closed loop system is designed to minimize. Note that the error signal of a control system can never be truly zero; it is the source of information to the control system.

## VI. Properties of a Closed Loop System

The motivation for constructing a closed loop system is one or both of the following reasons:

1. The open loop system has undesired features, e.g., poor frequency response,
2. The open loop system is intrinsically unstable.

We shall study an example for each case. To help us understand these examples better, we shall first develop some mathematical tools.

Consider an open loop transfer function  $G(s)$  whose poles and zeroes have been or can be determined. If the output  $Y$  is taken and summed back to the input  $C$  after operating on it with another transfer function  $H(s)$ , a typical closed loop system is formed (Fig. 11). The function  $H(s)$  can be a simple numerical value or can be complex (a closed loop system is called a unit feedback system when  $H(s)=1$ ). Simple algebra yields

$$\frac{Y(s)}{C(s)} = \frac{G(s)}{1 \pm G(s)H(s)} \quad (18)$$

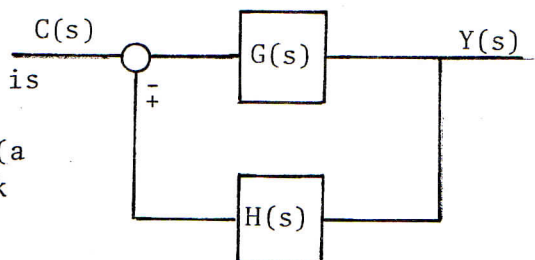


Fig. 11

For a negative feedback system, the sign in the denominator is (+) while for a positive feedback system, the sign is (-). We shall assume negative feed-

back for our present discussion unless it is stated otherwise. Therefore, we shall study the case

$$\frac{Y(s)}{C(s)} = \frac{G(s)}{1 + G(s)H(s)} \quad (19)$$

If  $|G(s)H(s)| \gg 1$ , we have

$$\frac{Y(s)}{C(s)} \approx \frac{1}{H(s)} \quad (20)$$

In particular, if  $H(s) = 1$ , i.e.,  $G(s)$  is very large, we have

$$\frac{Y(s)}{C(s)} \approx 1. \quad (21)$$

Note that the effect of  $G(s)$  in a unity (negative) feedback system, with large  $G(s)$ , is almost nil. It is true that the high gain has been lost but we have also eliminated the frequency effect. Hence, over a frequency range for which  $|G(s)| \gg 1$ , we have a "flat" response.

If  $H(s) = \beta$  where  $\beta$  is just a numerical value, we have

$$\frac{Y(s)}{C(s)} \approx \frac{1}{\beta} \quad (22)$$

Hence, by appropriately choosing the magnitude of  $\beta$ , we can construct an amplifier or attenuator with a flat response over a certain range of frequencies.

We shall study an example that the simple mathematics we have developed can explain. This example is very important in electronics.

Fig. 12 shows the symbol of an operational amplifier (op-amp). There are two inputs, positive and negative, and one output. When powered, the input voltages  $V_+$  and  $V_-$ , and the output voltage  $V_0$  are given by

$$V_0 = A_0(V_+ - V_-) \quad (23)$$

where  $A_0$  is called open loop gain. In simple terms, the op-amp compares the (+) input and the (-) input, then amplifies the difference  $(V_+ - V_-)$  by a factor  $A_0$ , and delivers to the output. Unfortunately, the open loop gain, its magnitude and phase, are functions of frequency. Fig. 13 shows the magnitude of  $A_0$  as a function of frequency for a 741 op-amp.

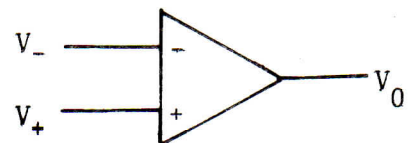


Fig. 12

Because of this actual frequency dependent gain, the open loop op-amp is not a good candidate for use as an amplifier. We would rather specify the gain we need, and have it fixed over the operating frequency range.



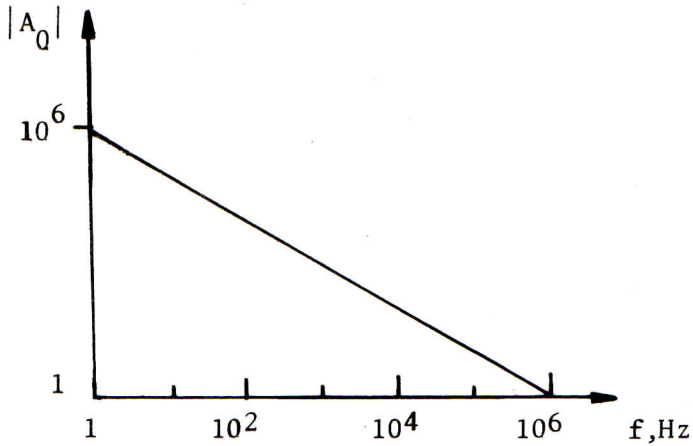


Fig. 13

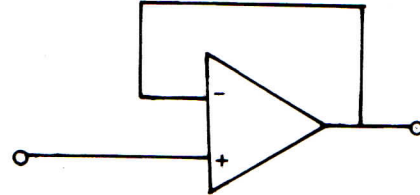


Fig. 14

Suppose we wish to design a unity gain amplifier using an op-amp, we can connect the op-amp as shown in Fig. 14. Readers who are not familiar with this type of circuitry may wish to compare Fig. 14 with Fig. 11 and review Eq. (19) through Eq. (21).

The output/input transfer function is now given by

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{A_0(s)}{1 + A_0(s) \cdot 1} \quad (24)$$

since  $H(s) = 1$ . At frequency values where  $|A(s)| \gg 1$ , we have

$$\frac{V_{\text{out}}}{V_{\text{in}}} \approx 1$$

and the amplification factor is unity over this frequency range. This unity gain amplifier fails at, for example,  $f = 1 \text{ MHz}$  where  $|A_0| = 1$  and the approximation  $|A_0| \gg 1$  no longer holds.

Suppose an amplifier with a gain of 10 is designed; we shall require a value of  $\beta$  for which

$$\frac{V_{\text{out}}}{V_{\text{in}}} = 10 = \frac{A(s)}{1 + A(s)\beta} \quad (25)$$

Over the frequency range for which  $|A_0(s)| \gg 1$  and  $|\beta A_0(s)| \gg 1$ , we find that

$$10 \approx \frac{1}{\beta}$$

so

$$\beta \approx 0.1.$$

(26)

Eq. (26) suggests that 10% of the output is fed back (negatively) to the input. Recall, from simple electronics, that two resistors can be arranged to form a voltage divider as shown in Fig. 15. We must be careful in employing the circuit shown in Fig. 15 because  $V_1$  should be the output of the op-amp while  $V_2$  should be the input negatively fed back to the op-amp.

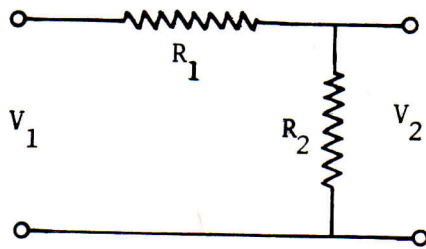


Fig. 15

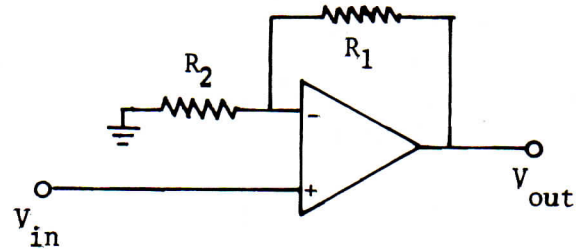


Fig. 16

The actual connection is shown in Fig. 16. It can be shown that the voltage gain of this connection is given by

$$\frac{V_{out}}{V_{in}} = \frac{1}{\beta} = \frac{R_1 + R_2}{R_2} = 1 + \frac{R_1}{R_2} \quad (27)$$

For  $\beta = 0.1$ , we have  $R_1/R_2 = 9$  and one can, for example, choose  $R_1 = 9\text{k}\Omega$  and  $R_2 = 1\text{k}\Omega$  or choose  $R_1 = 90\text{k}\Omega$  and  $R_2 = 10\text{k}\Omega$ . The result is that the gain of this feedback network is now governed by the designer's choice of resistors  $R_1$  and  $R_2$ .

Some readers may wonder whether the op-amp itself has any resistance (impedance) when used as a voltage divider as shown in Fig. 16. It turns out that the input impedance of an op-amp is ideally infinite, about  $10^6\Omega$  for a 741 op-amp, and some much higher. Therefore, as long as  $R_1$  and  $R_2$  are chosen to be less than the input impedance value, the given circuit should work. The practical considerations of optimizing the values of  $R_1$  and  $R_2$  then change to solving the problems of offset and other effects that are not necessarily related to feedback as presently discussed.

Note that, according to Eq. (25), as the closed loop gain  $V_{out}/V_{in}$  increase,  $\beta$  decreases. Consequently, the frequency range in which  $|\beta A_0(s)| \gg 1$  becomes smaller compared to, say, that of a unity gain system. One way of looking at this point is shown in Fig. 17.

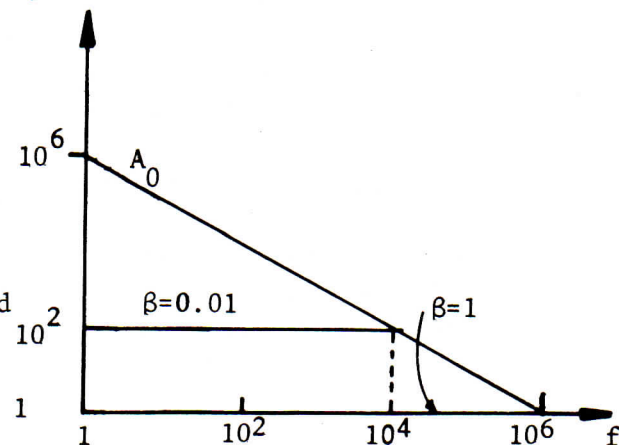


Fig. 17

when  $\beta=1$ , the approximation  $\beta A_0 \gg 1$  is no longer valid at  $f = 1$  MHz. Similarly, when  $\beta=0.01$ , the approximation fails at  $f=10$  kHz. This conveniently brings up another very useful item of technical vocabulary, namely, bandwidth. We can say that the bandwidth of the unit gain amplifier is approximately 1 MHz because its response is flat up to that frequency and then falls off. Similarly, the bandwidth of the amplifier of gain 100 is 10 kHz using the same op-amp.

Before we go on to develop more mathematical expressions on feedback, the author wishes to comment on the popularity (and simplicity) of using op-amps. Most mathematical operations in electronics such as addition, subtraction, multiplication, division, differentiation, integration, and even logarithm can be performed with relatively simple connections.

Let us write down the expression for a unity gain negative feedback system (Fig. 18).

$$\frac{Y(s)}{C(s)} = \frac{G(s)}{1 + G(s)} \quad (28)$$

Assume the poles and zeroes of  $G(s)$  are known, i.e., we may write

$$G(s) = \frac{\prod_{i=0}^n (s+z_i)}{\prod_{j=0}^m (s+p_j)}$$

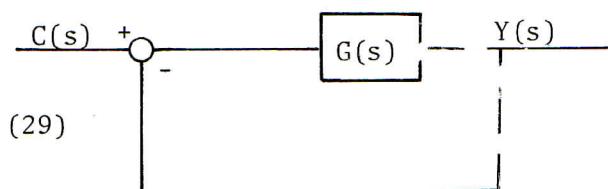


Fig. 18

The closed loop transfer function can then be rewritten as

$$\frac{Y(s)}{C(s)} = \frac{\prod_{i=0}^n (s+z_i)}{\prod_{j=0}^m (s+p_j) + \prod_{i=0}^n (s+z_i)} \quad (30)$$

Next, suppose we can write the denominator of Eq. (30) as  $\prod_{i=1}^{\ell} (s+p_i^c)$  where the superscript, c, denotes a closed loop system and  $\ell=m$  if  $m>n$  and  $\ell=n$  if  $n>m$ . Clearly, the locations of the closed loop poles,  $p_i^c$ 's are different from the locations of the open loop poles  $p_j$ 's.

In general, the closed loop system shown in Fig. 11 may have closed loop poles located at very different places from their open loop locations. Not only may the number of (closed loop) poles increase but also the zeroes of each open loop transfer function have their parts in determining the closed loop poles. Hence, the poles "move" when a loop is closed.



It was mentioned early in this section that one motivation of having a closed loop system is that the system may be open loop unstable. The following section will discuss stability problems from the locations of poles in the s-plane and an example will be presented.

## VII. Stability of a Physical System

The concept of stability and instability can be understood by graphically inspecting the poles of a transfer function. Let us assume a transfer function with the form

$$G(s) = \frac{1}{(s+a)(s+b)(s+c)} \quad (31)$$

where

$$a = \alpha + j\beta \quad (32)$$

$$b = \alpha - j\beta \quad (33)$$

and  $c$  is purely real. The inverse transform of  $G(s)$  yields

$$G(t) = Ae^{-(\alpha+j\beta)t} + Be^{-(\alpha-j\beta)t} + Ce^{-ct} \quad (34)$$

If  $\alpha$  and  $c$  are positive, then  $G(t)$  will be finite at  $t=\infty$  and any oscillations due to the  $j\beta$  term will die down. How fast this decaying oscillation takes place is governed by  $\alpha$ . Therefore, as shown in Fig. 19, if a pole lies on the negative real axis, it corresponds to a simple decay terms. If a pole lies off the real axis of the Left Half Plane (LHP), it corresponds to a decaying oscillatory term.

On the other hand, if a pole lies on the Imaginary axis, it corresponds to a purely oscillatory term with no decay. For example, imagine that a person's hand tries to pick up a pencil, but that his hand just swings left and right about the pencil and never converges.

In this case, there is no decay and a purely oscillating system is regarded as unstable.

If a pole lies on the Right Half Plane (RHP), it means that the model contains an exponentially growing term, which may or may not oscillate. Such a term "blows up" at  $t=\infty$ . A physical system usually does not end up this way. This is because we begin by modelling the system with a linear differential equation and we get a positive pole. As time,  $t$ , increases, the system will oscillate so much that it is no longer possible to describe

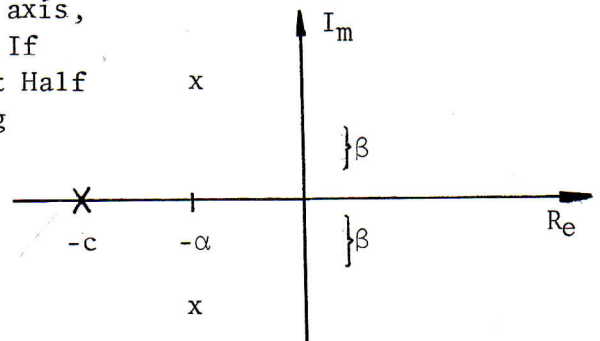


Fig. 19

it by a linear model.

One may ask whether or not a pole at  $s=0$  is stable. A pole at  $s=0$  corresponds to a constant term. This point is mathematically stable but any infinitesimal and physically real value of  $\alpha$  or  $\beta$  will result in a system with either a decaying or growing oscillatory term. Again, since an exact model of a physical system is seldom possible, a pole at the origin is generally regarded as unstable, or at least a source of potential trouble.

Suppose a system is open loop unstable, i.e., one or more of its poles lie in the RHP; it may be made stable by means of feedback. We have seen in the last section that closing a loop moves the poles around. It is therefore conceivable that the pole(s) on the RHP can be "moved" to the LHP by suitably choosing  $H(s)$ , the feedback transfer function. We shall use magnetic suspension to show how an open loop unstable system can be made stable by means of feedback.

Owing to the static nature of the earth's gravity field, it is impossible to produce a naturally force-free object in the laboratory without introducing another force. If we wish to have a contactless suspension of a mass  $m$ , we can take advantage of magnetic or electric force in order to balance the gravity force.

If a piece of soft iron is put near a permanent or electric magnet, an attractive force is felt. But the point where the magnetic force balances gravity is unstable. In other words, the soft iron will either move towards the magnet or fall away (assuming the magnet is held vertically). There is a general theorem that prohibits passive levitation along any axis of an object by electric or ferromagnetic means. Such theorem is known as Earnshaw's theorem<sup>4</sup>. Note, however, that diamagnetic and superconductive levitation can be done passively.

We shall use an electromagnet to demonstrate the nature of this instability. Fig. 20 shows an electromagnet whose current is described by the variable  $i$ . We shall use an inverse square law to describe the magnetic force which can be written in the form

$$F(i, \chi) = -\frac{Ci^2}{\chi^2}, \quad (35)$$

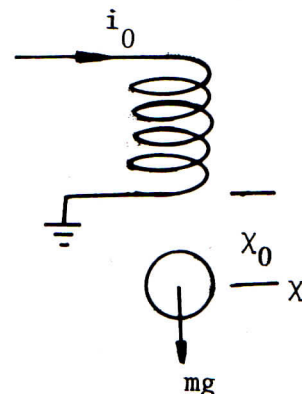


Fig. 20

where  $C$  is a constant depending on the geometry, and the negative sign means that it is an attractive force. The boundary condition is such that

$$mg = \frac{Ci_0^2}{\chi_0^2} \quad (36)$$

In other words, at a given current  $i_0$ , the magnetic force acting on the mass  $m$  at  $\chi_0$  is equal to its weight. The readers are reminded that  $\chi_0$  is not a stable point. The current is now varied by  $\Delta i$  and we are interested in the subsequent variation of  $\chi$ . We can expand Eq. (35) about  $i_0$  and  $\chi_0$  to obtain

$$F(i_0 + \Delta i, \chi_0 + \Delta \chi) = -\frac{Ci_0^2}{\chi_0^2} - \frac{2Ci_0\Delta i}{\chi_0^2} + \frac{2Ci_0^2\Delta \chi}{\chi_0^3} + \text{higher order terms.} \quad (37)$$

If we work out the boundary conditions and let  $\chi = \chi_0 + \Delta \chi$ , the equation of motion can be simplified as

$$\Delta \ddot{\chi} - \frac{g}{\chi_0} \Delta \chi = -\frac{g\Delta i}{i_0} \quad (38)$$

If we take the Laplace transform of Eq. (38) with initial conditions  $\Delta \chi(0)=0$  and  $\Delta \dot{\chi}(0)=0$ , we have

$$\Delta \chi(s) \left[ s^2 - \frac{g}{\chi_0} \right] = \frac{-g\Delta i(s)}{i_0} \quad (39)$$

The transfer function is

$$\frac{\Delta \chi(s)}{\Delta i(s)} = \frac{-g/i_0}{s^2 - g/\chi_0} = \frac{-g/i_0}{(s - \sqrt{g/\chi_0})(s + \sqrt{g/\chi_0})} \quad (40)$$

Hence, for a positive change of current, there results a negative change of position. Moreover, we see that this open loop transfer function has a pair of poles, one positive and one negative on the real axis as shown in Fig. 21. These poles correspond to the solution of an equation of the form

$$\Delta \chi(t) = Ae^{+t/\tau} + Be^{-t/\tau}, \quad (41)$$

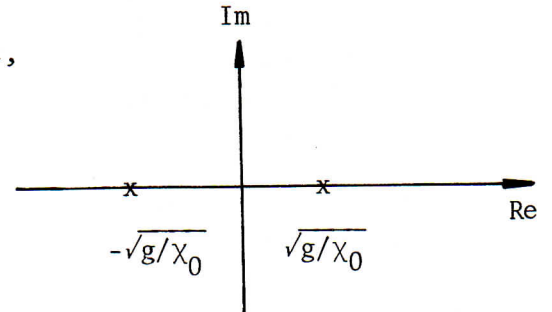


Fig. 21

where  $\tau = \sqrt{\chi_0/g}$ . Clearly, the first term goes to infinity as time  $t \rightarrow \infty$  and therefore explains why the point  $\chi_0$  is not a point of stable equilibrium.



If we establish a closed loop situation, we may "move" the positive pole leftward. In other words, the closed loop system may be stable despite its open loop instability. Therefore, we need to devise an appropriate  $H(s)$  of Fig. 11. When we use Fig. 11 to study our case of magnetic suspension,  $G(s)$  represents the dynamics of the suspended object which are fixed. On the other hand,  $H(s)$  is a transfer function that converts position to force. If we monitor the object's position with, for instance, an optical sensor or by a capacitive bridge, we can obtain a voltage proportional to the displacement of the ball relative to a fixed point, say,  $\chi_0$ . This displacement from the fixed point is our error signal and it is on this signal (and the reference) that the control system operates. The voltage can be amplified and converted into a current which can then be added to the idling current flowing through the solenoid. The polarity of this correction current is, of course, such as to yield a restoring force when the object falls and vice versa.

Suppose  $H(s)$  is  $s$ -independent, i.e.,  $H(s) = H$ , the closed loop transfer function is given by

$$\frac{\Delta\chi(s)}{\Delta i(s)} = \frac{\frac{g/i_0}{s^2 - g/\chi_0}}{\frac{Hg/i_0}{s^2 - g/\chi_0}} \quad (42)$$

The denominator of this closed loop transfer function is

$$s^2 - \frac{g}{\chi_0} + \frac{Hg}{i_0} = 0. \quad (43)$$

Clearly, we will either wind up with  $s=0$  or a pair of purely imaginary poles. The latter indicates that the object will oscillate about the equilibrium point. In order to have the object stably levitated at some designated position, say  $\chi_0$ , we must obviously require some damping. Surely, if  $H(s)$  contains an  $s$ -dependent term, the algebra may yield two real poles on the real axis. Since we are trying to construct  $H(s)$  electronically, the  $s$ -term (or damping term) must be electrical also. In fact, the levitated object is usually in a vacuum, so environmental or viscous drag is minimized. The electronics configuration that yields an appropriate  $s$ -term is a differentiation circuit, whose behavior is analogous to velocity dependent viscous damping. Hence, the overall circuit needed is one in which we have a proportional and a derivative circuit connected in parallel. These circuits are shown in detail in the several magnetic suspension papers submitted to the conference.

Most active magnetic suspension systems used today are designed to

meet different types of specifications. Because control handbooks now contain many standardized rules for designing control systems, a magnetic or electric suspension is normally designed around these rules rather than around a basic understanding of the physics involved. Indeed, having the proportional and derivative networks in parallel, as we proposed, may not be the only functional requirement we need; but it is the most essential part of this particular control problem. Additional features simply serve to make it work better in some detail.

The magnetic suspension that has just been discussed is a vivid example of a physical system that is open loop unstable but closed loop stable. However, the readers are cautioned that a closed loop system is not free of problems. The rest of this section will be devoted to stability of a closed loop system.

It was said in the Introduction that too high a gain setting may make a closed loop system oscillate, therefore producing an undesirable result. Another way for a feedback system to fail involves each function. Since each may change the phase of the processed signal somewhat, the accumulated phase shift may cause an adverse effect. For example, a total phase shift of  $180^\circ$  will make (unexpectedly) a negative feedback system into a positive feedback system. Note, though that phase shift is frequency, or  $s$ , dependent. For example, neither the magnitude nor the phase on the low frequency component of a certain signal changes when it is passed through a low phase electronic filter but both the magnitude and the phase of the high frequency component will change. It is therefore conceivable that a feedback system would work very well in a certain frequency range but oscillate in another range. This is, in fact, a familiar phenomenon in electronics.

There are many methods of analyzing the stability of a physical system. Normally, the stability of an open loop system is readily understood by considering the poles and zero locations, i.e., finding out if the pole(s) is (are) in the RHP. What one is often interested in is whether or not a certain closed loop system is stable; and if stable, how much margin does it have against disturbances; how fast does it respond to an input; and will its output be exactly at where it is expected? These questions are all common design criteria used or specified by control engineers. They are only mentioned here but will not be extensively discussed since such discussions are readily and thoroughly available in the standard textbook.

#### VIII. Methods of Stability Study

Consider again the closed loop transfer function with negative feedback in Fig. 11. According to Eq. (19), we have



$$\frac{Y(s)}{C(s)} = \frac{G(s)}{1 + G(s)H(s)} \quad (44)$$

The denominator  $1+G(s)H(s)$  is sensitive to either  $|G(s)H(s)| = 1$  or  $\angle G(s)H(s) = 180^\circ$ , or both.

It is generally easy to determine (empirically and theoretically)  $G(s)$  and  $H(s)$ . Empirically, if  $G(s)$  and  $H(s)$ , individually or together, produce stability, one then breaks the loop as shown in Fig. 22 and obtain  $G(s)H(s)$  from open loop measurements. From the results for  $G(s)$   $H(s)$ , the closed loop behavior can be predicted. This approach is known as predicting the closed loop behavior by studying that of the open loop.

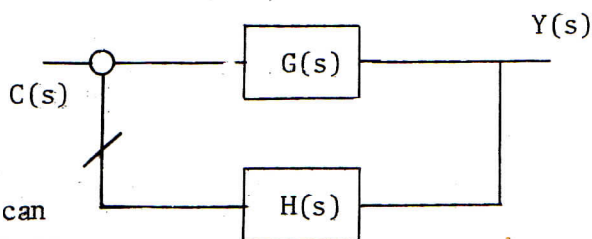


Fig. 22

One method, known as Bode plot, is to plot the magnitude and phase of  $G(s)H(s)$  versus frequency. Fig. 23 illustrates this idea. Bode plot is a log-log plot. As found in the open loop Bode plot, the locations where  $|A(s)| = 1$  (0dB) and  $\angle A(s) = \pi/2$  are of interest [ $A(s)$  is the overall open loop transfer function, i.e.,  $A(s) = G(s)H(s)$ ]. After finding them, we compare them with the desired response and see if  $H(s)$  needs to be modified.

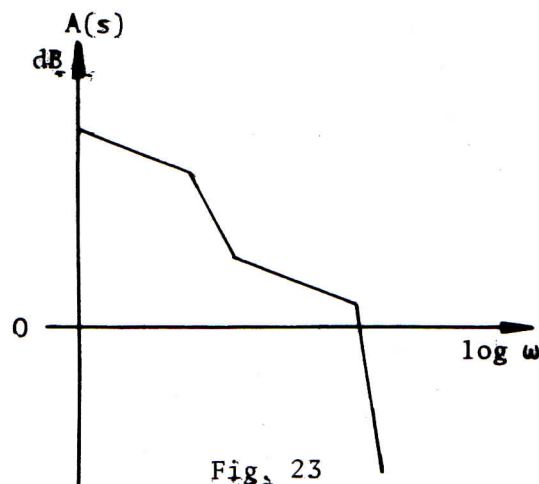


Fig. 23

The second method is to use the system's root locus. The experimenters usually have control over the system's gain, i.e., they can increase the gain (or amplification factor) at will. Therefore, there is a gain term, call it  $K$ , built in  $G(s)$ ,  $H(s)$ , or both. Since the closed loop poles are determined by the sum  $1+G(s)H(s)$ , they are, therefore, dependent upon  $K$ . The root locus is a graphical trace of the location of the closed loop

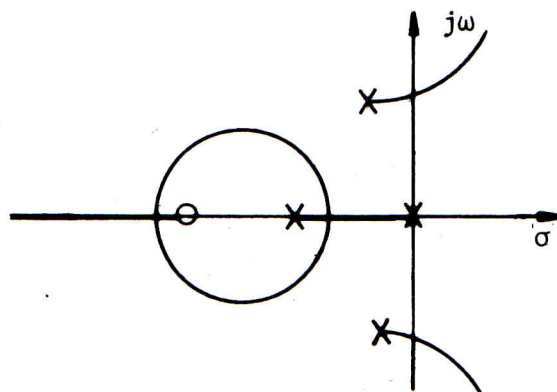


Fig. 24



poles and zeroes of the transfer function  $1+G(s)H(s)$  as  $K$  is varied from zero to infinity. The trace shows when and how the closed loop poles produce instability or oscillation. Fig. 24 shows how a root locus looks.

Other methods such as the Nyquist plot is also used but we shall not cover them. Bode plots are widely used in electronic documents because the log axis covers a wide spectrum of frequency. The poles and zeroes of an open loop transfer function show up as the "corner frequencies" on the Bode plot. Unfortunately, RHP poles cannot be represented by such a plot. Hence, our magnetic suspension example is not suitably analyzed solely by the Bode plot because the plot does not reveal these poles. In this case, the root locus method is more exact.

#### Concluding Remarks

The subject of feedback control was treated here in an overly simplified manner. The concepts of open and closed loop systems, positive and negative feedback, and stability were brought out by simple examples. The use of feedback in operational amplifiers and in magnetic suspension are covered extensively. Finally, two graphical methods of depicting the behavior of a physical system were briefly discussed.

It is the author's personal feeling that a closed loop system is rarely studied or assembled by starting from its physical fundamentals and then proceed to achieve its best engineering performance. Often, a physicist perceives what turns out to be a control system but he is not quite aware of the rules and ways to make this system perform its best. Similarly, a control engineer often builds his system around the rules but does not look deeper into the fundamental nature of this system. Therefore, it is the author's hope that this paper will initiate some awareness of feedback among students who will go into experimental science.

#### Acknowledgement:

The author wishes to thank Dr. George T. Gillies of BIPM, France, for his comments on this article during early preparation, and Professor Rogers C. Ritter and Carl H. Leyh of University of Virginia for their helpful discussions in this subject.

## Suggested References

1. Ogata, K., Modern Control Engineering, Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1970.
2. Stante, D., Automatic Control System Technology, Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1980.

## References

1. To the knowledge of the author, this work has not been published.
2. Unpublished, private communication.
3. J.E. Faller, Introduction To How Things Fall, PMGE conference proceeding.
4. R.C. Ritter, et.al., Analysis of Vertical Magnetic Suspensions, PMGE conference proceeding.