1957 IRE National Convention Recor

PART 4

Sessions sponsored by

IRE Professional Groups on

Automatic Control

Electronic Computers

Medical Electronics

at

the IRE National Convention, New York, N. Y.

March 18-21, 1957

The Institute of Radio Engineer

PART 4 - COMPUTERS; AUTOMATIC CONTROL; MEDICAL ELECTRONICS

TABLE OF CONTENTS

Page

Session 1: Nonlinear Control Systems
(Sponsored by the Professional Group on Automatic Control)

Nonlinear Compensating Networks for Feedback Systems . . . . E. Mishkin and J. G. Truxal 3
Direct Synthesis through Block Diagram Substitutions. O. J. M. Smith 8
A Nonlinear Control System for Wide Range Input Signals . . . . J. Tou and Y. H. Ku 17
Switching Discontinuities in Phase Space . . . . J. C. Hung and S. S. L. Chang 22
Nonlinear Techniques Applied to the Analysis of Pilot Induced Oscillations. . . . . . I. H. Van Horn 27

Session 9: Automatic Control -- General
(Sponsored by the Professional Group on Automatic Control)

Digital Computers in Feedback Systems . . . . J. R. Ragazzini 33
Sampling in Linear and Nonlinear Feedback Control Systems . . . . J. Kukel 43
Solution of Statistical Problems by Automatic Control Techniques . . . . R. L. Cosgriff 57
The Design of Optimum Filters and Predictors . . . . C. W. Steeg 62

Session 18: Medical Electronics
(Sponsored by the Professional Group on Medical Electronics)

A Magnetic-Tape Recording System for Teaching Electrocardiography
. . . . G. N. Webb and W. R. Milnor 74
Ultrasonic Dosimetry for Medical Use . . . . V. T. Tomberg 80
An Electrical Network Analogy for Isotope Kinetics . . . . R. L. Schoenfeld and M. Berman 84
Electronic Control of Artificial Respiration . . . . L. H. Montgomery 90
The Incorporation of Picture Storage with the Technique of X-ray Fluoroscopy
(Abstract) . . . . W. Gombash, Jr. 94

Session 37: Symposium on Applications of Computers in Biology and Medicine
(Sponsored by the Professional Groups on Medical Electronics and Electronic Computers)

Abstract of Symposium 95

Session 26: Electronic Computers I -- Digital Computers
(Sponsored by the Professional Group on Electronic Computers)

An RCA High-Performance Tape Transport Equipment . . . . S. Baybick and R. E. Montijo 96
A Magnetic Pulse-Current Regulator . . . . J. D. Lawrence, Jr., and T. H. Bonn 102
Diodeless Magnetic Core Logical Circuits . . . . L. A. Russell 106
Circuit Design Employing A Digital Computer to Attain Longest Mean Time
to Failure . . . . J. Alman, P. Phipps and D. Wilson 115

Page

Session 26 *(continued)*

Considerations in the Design of Character Recognition Devices . . E. C. Greanias and Y. M. Hill 119

Session 41: Electronic Computers II -- Symposium on Computers in Simulation, Data Reduction and Control
(Sponsored by the Professional Group on Electronic Computers)

Digital Computers in Continuous Control Systems . . . . E. L. Braun 127
Computers in the Process Industry . . . . W. F. Gunning 136
Aspects of Real-Time Simulation . . . . W. F. Bauer 142
Digital Information Processing for Machine-Tool Control . . . . A. K. Susskind 145

Session 49: Electronic Computers III -- Mainly Analog
(Sponsored by the Professional Group on Electronic Computers)

Computation with Pulse Analogs . . . . N. Rubenfeld 150
A Cyclic Digital-to-Analog Decoder . . . . G. H. Myers 156
An Automatic Analog Computer Method for Solving Polynomials and Finding Root Loci . . . . L. Levine and H. F. Meissinger 164
Magnetically Controlled Counters . . . . E. A. Sands 173
Systematic Tracing of Discrepancies in Analog Computers . . . . . . E. Grosswald and M. Plotkin 175

NONLINEAR COMPENSATING NETWORKS FOR FEEDBACK SYSTEMS

E. Mishkin and J. G. Truxal
Microwave Research Institute
Polytechnic Institute of Brooklyn
Brooklyn, New York

INTRODUCTION

If we consider nonlinear feedback systems which are unstable as a primary consequence of the nonlinearity (i.e., in which the presence and nature of the nonlinearity plays a key role in the oscillation), stabilization has in the past generally taken either of two forms: (1) the nonlinearity can be cancelled by another nonlinearity which results in a linear, overall system; or (2) the describing-function analysis is the starting point for an attempt to find, by suitable exercise of the designer's ingenuity and imagination, a nonlinear compensation network which will appropriately shape the amplitude and/or frequency loci such that intersection is avoided.

If we restrict our consideration to nonlinearities describable in terms of describing functions (i.e., if we exclude for example multiplicative nonlinearities), the second approach holds considerable promise. An alternate approach, valid at least whenever the nonlinearity can be represented as a frequency-independent block in the system block diagram, has been proposed by several authors in essentially the form shown in Fig. 1 or an equivalent configuration. Here g_1, n, and g_2 represent the given system (which of course in general would be very much more complex), with g_1 and g_2 linear and n nonlinear. If the compensating elements n_1 and g_{2o} are chosen such that n_1 is the complement of n [i.e., (m_2+m_{2o}) is a linear function of m_1] and g_{2o} is identified with g_2, the distortion introduced in the fed-back signal by n is exactly cancelled by the minor-loop feedback. Hence, insofar as stability is concerned, the overall system is linear and no assumption re the validity of the describing function is made, although the actual closed-loop system is certainly still nonlinear.

While such an approach looks promising on paper, two primary difficulties arise in practice:

(1) The g_{2o} and, to a lesser extent, n_1 blocks are ordinarily far more complex than desirable (or necessary). This difficulty is particularly apt to arise if the original nonlinearity requires an appreciable expansion of the original block diagram in order to leave the nonlinearity represented as a frequency-independent block (as in the case of the manipulation required with backlash or velocity saturation).

(2) Certain difficulties may arise in practice because the minor loop may respond so fast and accurately that the g_2 block is never actuated (e.g., if n contains a dead zone, g_2 an integration, and the input is a ramp).

The answer to both difficulties lies in easing the requirements on the specific characteristics of the elements in the minor loop. It is the purpose of the following discussion to indicate the form design might take, then, if we started from the ideal-model philosophy exemplified in Fig. 1, but then modified this idealism to bring both items (1) and (2) above into consideration. The discussion is phrased in terms of the specific examples of hystersis as the n element and of a very simple, two-nonlinearity system. At the present time, the bounds of appreciability and usefulness of this approach to nonlinear compensation are not clear.

The Describing Function

The nonlinear system shown in Fig. 2a contains linear elements g_1 and g_2 and a nonlinear component n that may represent a hysteris phenomenon inherent in the system. The relationship between m_1 and m_2, the input and output respectively of the nonlinear element, is shown in Fig. 2b.

We shall concern ourselves mainly with the problem of the system stability that may be upset by the varying nature and level of the transmitted signal. Under unstable operation conditions, the signals appearing at the various points of the system will have an alternating character with respect to time. In general, they will be nonsinusoidal, but if g_1, g_2 present a transfer function similar to that of a low-pass filter, it will be reasonable to assume that at least m_1, the signal at the input to the nonlinearity, will be nearly sinusoidal.

Further analysis will be greatly facilitated, if n is replaced by a linear element with a complex gain N that is a function of the amplitude M_1 of the signal m_1. The choice of N is to be made so as to minimize the mean square error

$$e_m^2 = \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} \left[m_2(t) - N M_1 \sin \omega t \right]^2 dt , \quad (1)$$

i.e.

$$\frac{\partial e_m^2}{\partial N} = - \frac{2}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} \left[m_2(t) - N M_1 \sin \omega t \right] M_1 \sin \omega t = 0, \quad (2)$$

and

$$N = \frac{2}{T\,M_1} \int_{-\frac{T}{2}}^{+\frac{T}{2}} m_2(t) \sin \omega t \, dt \quad . \qquad (3)$$

N presents in other words the describing function: i.e. the complex ratio between M_2, the amplitude of the first harmonic of the output, and the amplitude M_1 of the sinusoidal-input to the nonlinear element *n*.

The error made by this type of presenting a nonlinear element consists of all the higher harmonics that are part of $m_2(t)$. A better approximation can be obtained, if necessary, by considering the higher harmonics of $m_2(t)$ as additional sinusoidal disturbing signals of appropriate frequencies--their amplitudes depend upon the level of the m_1 signal.

Fig. 3 shows the describing function N, its magnitude $|N|$ and phase shift ϕ_N as functions of the relative amplitude M_1/H.

Stability Analysis

For the system shown in Fig. 2a

$$\frac{C}{R} = \frac{G_1\ NG_2}{1 + G_1\ NG_2} \quad . \qquad (4)$$

The stability analysis can be performed with the aid of the Nyquist plot shown in Fig. 4. The system is stable at low amplitudes M . At P the $G_1(j\omega)\ G_2(j\omega)$ curve cuts the wandering - 1/N point, the system becomes unstable and the amplitude M_1 keeps on increasing until a stable mode of oscillations is established at Q at about 2 $^{rad}/_{sec}$.

There are several methods by which this system can be stabilized. The gain K of the linear part when sufficiently decreased will cause the system to be stable. Similarly the linear part G_1, G_2 can be compensated in accordance with well known methods so as to yield a resultant $G(j\omega)$ curve that will neither encircle nor cut the moving - 1/N point. We shall concern ourselves here with compensating methods by means of nonlinear networks which can be inserted in parallel or rather in a minor feedback loop. The advantages of this method will be shown later, but it is clear that in general, if anywhere nearly similar dynamic characteristics are to be achieved over a reasonable range of signal amplitudes, nonlinear systems must be compensated by nonlinear networks.

Nonlinear Compensation

It is apparent directly from the Nyquist plot in Fig. 4 that, in order to stabilize the system shown, one must limit the possible growth of the signal at m_1 or m_2. This can be accomplished by inserting a nonlinearity n_1, in parallel with n, followed by a linear model g_{2o} (see Fig. 1). The transfer function G_{2o} is to equal G_2. The n_1 characteristic is shown in Fig. 5. The deadzone abscissa M_o, in Fig. 5, is to be somewhat smaller than the value M for which the system becomes unstable and the absolute value of the negative slope of n_1 is to equal the slope of N.

Insofar as stability is concerned, and considering $G_{o2} = G_2$, the diagram of the compensated system can be redrawn as in Fig. 6. A new describing function for the composite nonlinearity n + n_1 is to be obtained by adding the ordinates of the gains and phase-angles of the individual describing functions N and N_1 (see Fig. 7). The dashed line on Fig. 4 is the - $1/N+N_1$ curve for amplitudes $M > M_o$.

In general, the slopes of n and n_1 cannot be perfectly matched. Assuming that the slope of n is unity, while that of n_1 is - 0.9, the resultant value of $|N| + |N_1|$ is 0.1, for $M_1 \rightarrow \infty$. In Fig. 4, the end point of the - 1/N + N_1 curve will move from infinity to - 10. Should there be a need for increased relative stability, for a certain range of amplitudes M_1, the absolute value of the slope of N_1 is to be increased, for this range of amplitudes.

The compensation method by means of nonlinear networks has certain advantages over the method of adjusting the linear portion by reducing the gain K. In the latter case, the system overall performance is being slowed down, while the addition of n_1 introduces a form of an automatic gain control at signal levels which the original system cannot handle and remain stable.

The compensating nonlinearity n_1 and the model g_{2o} appear in the minor loop, which is to transfer a signal at as low a power level as possible. It should therefore be easy to realize the stabilizing elements using electrical and mechanical transducers.

Overall Model-Feedback Design

It has been assumed in the previous section that the model g_{2o} has an identical transfer function with that of g_2. G_2 may present a complex block and it might be advantegeous to retain the freedom in designing both, the linear part g_{20} and the nonlinear network n_1 of the stabilizing model feedback loop.

The block diagram of Fig. 1 can be rearranged as shown in Fig. 8. The transfer function of the g(n) block is

$$G(n) = \frac{G_1}{1 + G_1\ NG_2} \qquad (5)$$

When the signal at m_1 is small--at a level lower than that corresponding to point P in Fig. 4-- the "open-loop" $G(n)$ is stable. The model loop $n_1\ g_{2o}$ is not to exercise yet any stabilizing effect upon the system and the nonlinearity n_1 may have a dead zone smaller than m_{1P}-the magnitude of m_1 at P. Between P and Q the $G_1(j\omega)\ G_2(j\omega)$ locus encircles the $-1/N$ point once; hence for values of m_1 between P and Q, the $G(n)$ function has a single pole in the right half P plane. In this interval then, n_1 and g_{2o} have to be chosen in such a way as to cause the overall $G(n)\ G_{2o}(j\omega)$ curve to encircle the new travelling $-1/N + N_1$ point once in the negative mathematical sense.

$G(n)$ depends upon the signal level at m_1. To each value of m_1 corresponds a different $G(n)$ curve. The above outlined procedure will require the drawing of two $G(n)$ curves corresponding to the two extreme values of m_1.

Systems with Two Nonlinearities

Two nonlinearities may follow, or parallel one another in a feedback control system. When the nonlinear elements are frequency independent, a resultant nonlinear relationship, between the input to the first and the input of the following element, can be derived from the characteristics of the individual elements. One is then confronted with the usual problem of a nonlinear element in an otherwise linear system.

A more complicated problem occurs when a frequency dependent linear network separates the two nonlinearities, or when one of the nonlinear elements is frequency dependent. The analysis becomes very cumbersome as the properties of the network are simultaneously amplitude and frequency dependent and are presented in the form of a family of curves instead of a single plot.(1) (2) In practice it involves the aid of a computing device. In case of a single frequency-dependent nonlinearity, frequently, the frequency and amplitude dependent parts can be separated and the analysis of the problem extensively simplified.(3)

Let n_1 and n_2 be two nonlinear elements, amplitude-dependent only, and g a linear network arranged as shown in Fig. 10b. n_1 may present a dead-zone phenomenon and n_2 a saturation effect. The particular characteristics and the describing functions N_1 and N_2 are drawn in Fig. 9. Let g be a single integration $1/p$. When m_1 is a sinusoidal signal of amplitude M_1 and in the stability analysis of the overall system it is reasonable to neglect all the higher harmonics of m_2, n_1 can be best approximated by the describing function N_1. The integration element $1/p$ discriminates against the higher harmonics of m_2 in proportion to the order of the harmonic. The signal at m_3 can be considered as sinusoidal and the describing function N_2 substituted for the nonlinear element n_2. Choosing particular values for the amplitude M_1, amplitude M_2 is obtained from Fig. 9a, and consecutively amplitudes M_3 and M_4 result from Figs. 9b and c. The ratio $|N| = M_4/M_1$, and the -90° phase shift caused by $1/p$, constitute the resultant describing function of the overall nonlinearity n_1 $-g$ $-n_2$. It is shown in form of a family of curves in Fig. 10.

Stability Analysis

Let n, in Fig. 2a, present the n_1 $-g$ $-n_2$ arrangement described in the previous section. Fig. 11a shows the Nyquist plot for

$$G_1(j\omega)\ G_2(j\omega)\ e^{-j90^\circ} = \frac{K}{(j\omega+1)(j\omega+4)}\ e^{-j90^\circ}$$

and the $-1/|N|$ plot. The -90° phase shift of N, caused by the integration $1/p$, is combined here with the $G_1(j\omega)\ G_2(j\omega)$ plot. At about $\omega = 2$ rad/sec the phase angle of $G_1\ G_2\ e^{-j90^\circ}$ is -180°. The $-1/|N|$ point travels along the real axis. When K is large enough, the $G_1(j\omega)\ G_2(j\omega)\ e^{-j90^\circ}$ curve will cut the $-1/|N|$ plot at two points of which the one corresponding to the lower signal level M_1 presents an unstable mode of oscillation. At the other point, M_1 is larger and the system oscillates at a constant amplitude(stable limit cycle.)

The above system will be unstable when the tip of the $-1/|N|$ locus, for $\omega = 2$, is encircled by the $G_1(j\omega)\ G_2(j\omega)\ e^{-j90^\circ}$ plot.

When the n_1 and n_2 nonlinearities contain no energy storages the phase shift of the n element is caused by the g network only. This phase shift is a function of the frequency and can be combined with the $G_1(j\omega)\ G_2(j\omega)$ function. The Nyquist plot will then determine the frequency at which the system will oscillate if unstable. For purposes of stability analysis the describing function $|N|$ of the overall n_1 $-g$ $-n_2$ network is to be found for this frequency only. The system exhibits a stable limit cycle when the $G_1(j\omega)\ G_2(j\omega)\ e^{-j\Phi_g}$ cuts the $-1/|N|$ locus, drawn at the critical frequency ω for which the angle of $G_1(j\omega)\ G_2(j\omega)\ e^{-\Phi g}$ is -180°, at two points. The system will go in a destructive mode of oscillations when the $-1/|N|$ plot is terminated inside the $G_1(j\omega)\ G_2(j\omega)\ e^{-\Phi g}$ locus.

Conclusions

Nonlinear systems can be effectively stabilized by means of the feedback model loop method. Systems with more than one nonlinear element (containing no energy storages), when unstable oscillate at a certain critical frequency. They can be stablized with the aid of narrow band-stop filters. In both cases additional work is required in the resultant response of the system.

References

1. R. J. Kochenburger, "Limiting in Feedback Control Systems", Trans. A.I.E.E., Vol. 72, Part II, Applications and Industry, pp. 180-194, 1953.

2. J. G. Truxal, "Automatic Feedback Control System Synthesis", McGraw-Hill Book Co., N. Y. 1955, pp. 585-599.

3. G. Casserly and J. G. Truxal, "Measurement and Stabilization of Nonlinear Feedback Systems", I.R.E. Convention Record 1955, Part 4, pp. 52-61.

Acknowledgement

The work reported on here has been sponsored in part by the office of Ordnance Research, U.S. Army and the Air Force Office of Scientific Research. The authors sincerely acknowledge the important contributions made by Mr. Kenneth R. Kaplan and Mr. Richard A. Haddad.

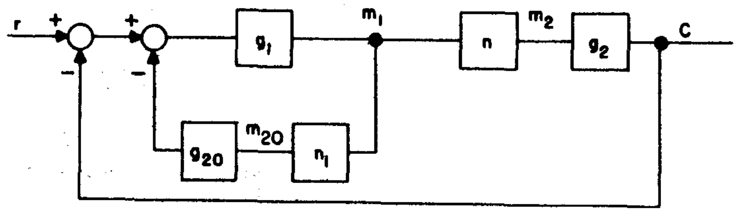


Fig. 1
Model compensation.

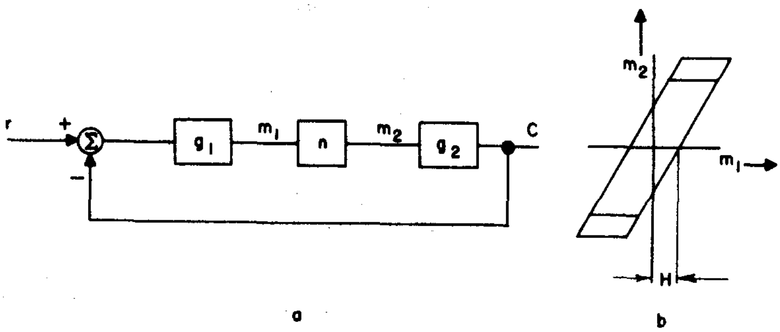


Fig. 2
Nonlinear system.

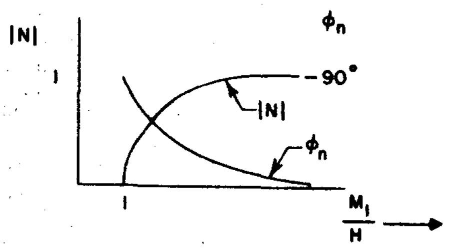


Fig. 3
Describing function of hysteresis element.

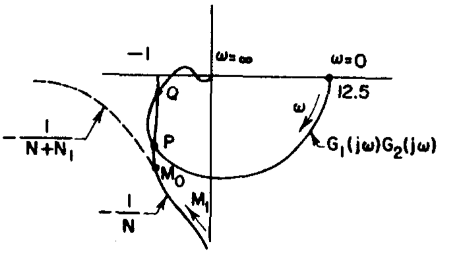


Fig. 4
Nyquist plot for system with a hysteresis element.

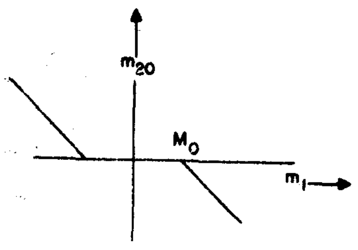


Fig. 5
Characteristic of compensating nonlinear element n_1.

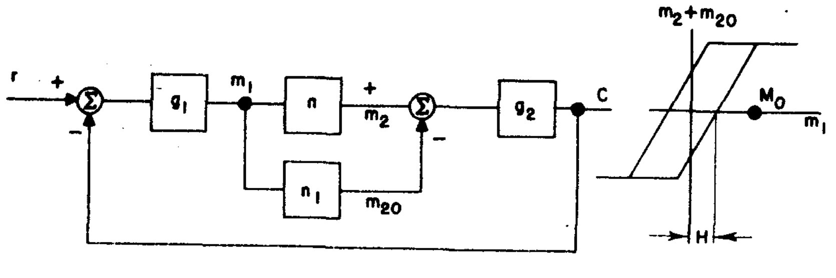


Fig. 6
Compensating system diagram for stability analysis.

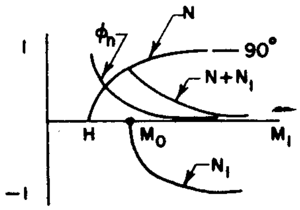


Fig. 7
Resultant describing function.

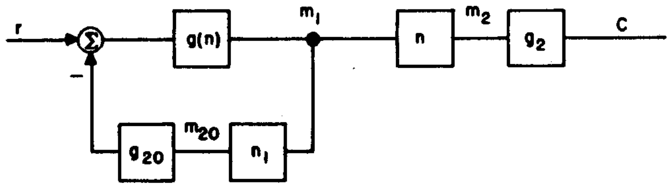


Fig. 8
New block diagram.

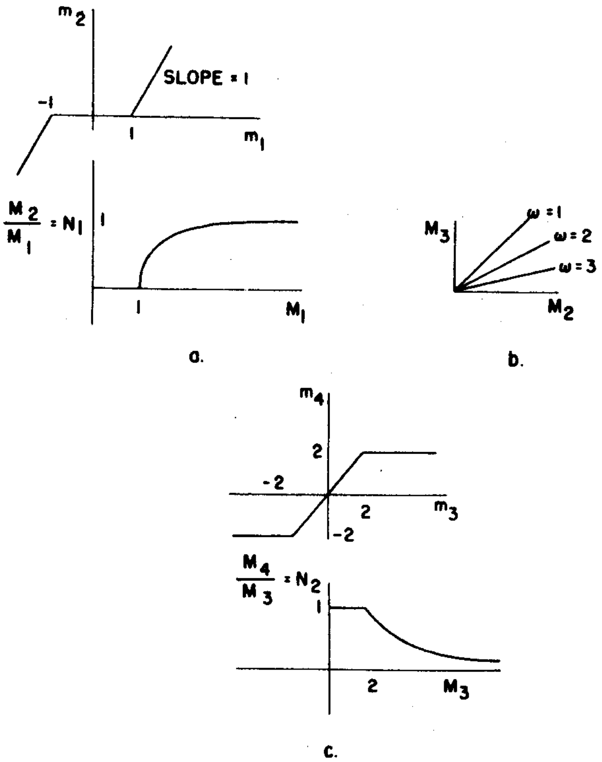


Fig. 9
Characteristics and describing functions of the individual elements.

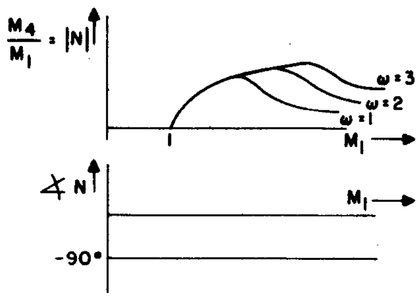


Fig. 10
Overall describing function.

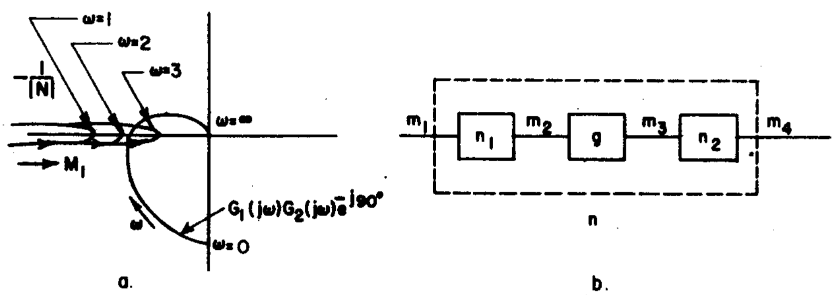


Fig. 11
Nyquist plot for a two-nonlinearity system.

DIRECT SYNTHESIS THROUGH BLOCK DIAGRAM SUBSTITUTIONS.

O. J. M. Smith
University of California
Berkeley, California

Summary

The direct synthesis of a system starts with stating two restrictions: (1) the unalterable components of the system; and (2) the undesired signals. Two engineering judgments must next be introduced: (1) a criteria for best; and (2) a possible, realizable mode of action. These four statements can be combined into a single block diagram which is a statement of the best possible, realizable control. By means of certain block diagram substitutions, this mode of action representation can be changed into the constructional arrangement. Rules for the permissible block diagram substitutions will be given. Examples are given of the application of this method to minimum-phase linear systems with not more than one more pole than the number of zeros, to minimum-phase linear systems with many more poles than zeros, to dead-time systems, statistical predictors, and dead-beat control.

Unalterable Components and Undesired Signals.

Every problem in synthesis has certain restrictions. Because of cost or weight limitations unalterable components may have too narrow a frequency band for the signal spectrum, or too many poles, or be non-linear, or perhaps be unstable. The types of undesired signals that enter a system are load forces or disturbances, noise, and distortion from non-linearities. The more severe these restrictions are, the more difficult it is to find a realizable possible mode of action. However if one eliminates too many restrictions, the synthesis problem becomes trivial.

Minimum Phase Linear System with One More Pole than Number of Zeros.

This design sequence can best be illustrated by reference to Figure 1. Figure 1a shows the unalterable component and the undesired signal in the given system. Examples of systems of this sort are a velocity control where θ_c is the motor torque, θ_L is the load torque, and θ_o is the shaft velocity. Or θ_L may represent the load current in a voltage regulator, and θ_o represent the regulated voltage. The unalterable dynamics,

$$\frac{1}{1 + sT} ,$$

could just as well be written G(s), which must be a minimum phase linear system with a denominator polynomial of degree equal to or one more than the degree of the numerator polynomial. The desired system is an output function θ_o which is exactly equal to an input function θ_i. The desired system should therefore have a transmission or transference of unity as shown in Figure 1b.

We will now hypothesize that the criteria for the best system is absolutely zero error between the input and the output We will further hypothesize that there exists a solution for the system using this criteria and that this is a realizable, possible mode of action. Therefore to convert the given system in Figure 1a into the desired system in Figure 1b one operates on the input signal in Figure 1b with the inverse of the unalterable transference and subtracts a signal exactly equal to the load disturbance. Figure 1c is a statement of the realizable desired mode of operation. It is not the finished design.

To obtain the finished design one uses a series of block diagram substitution rules. The first step in converting this block diagram into the constructional arrangement, is to move the signal compensating for the load disturbance to the input. This results in θ_L passing through the $1/1 + sT$ in Figure 1d, and being subtracted directly from θ_i.

The next rule that we will introduce is that a function may be generated by its inverse in negative feedback around an infinite gain amplifier. From s plane root locus studies it can be shown that this rule is true if the function has an equal number of poles and zeros. It can also be shown that this rule is practically true, and that the closed loop is stable, if the function introduced in the negative feedback branch of the closed loop has not more than one more pole than the number of zeros. In other words, one can generate a function which has one more zero than the number of poles. The application of this rule to the generation of the zero in Figure 1c results in the loop in Figure 1d.

The feedback loop which has been generated will now be expanded by moving the branch point after the amplifier through the load disturbance and into the output lead. The reason for doing this is that in the actual system it is the output variable which can be measured and therefore we should take advantage of this measurement. In the process of this substitution the signal θ_L is introduced into the feedback branch. By arranging the three adders in juxtoposition as shown in Figure 1e, and recognizing that the order of adding signals can be interchanged without affecting the summation, it can be seen that the two load-force correction signals cancel exactly, leaving the final arrangement of the system as in Figure 1f.

This is the constructional arrangement. We have derived the basic philosophy underlying the use of feedback systems instead of communication or filter channels. It can be seen that unity negative feedback around an infinite gain amplifier in series with a dynamic element having not more than one more pole than the number of zeros, will yield the equivalent of perfect flat response with infinite bandwidth. The infinite gain amplifier therefore generates the inverse of the system unalterable dynamics. In addition the infinite gain amplifier generates the negative of the load disturbance. If we had within this loop a nonlinear amplifier, then the infinite gain amplifier would not only compensate for the variation in gain due to the nonlinearity, but would also compensate for all of the harmonic distortion generated by the nonlinear amplifier. The things which it cannot compensate for are unwanted signals or noise signals at the input of the infinite gain amplifier. The actual infinite-gain amplifier is built with a gain of K and a minor positive feedback of $1/K$. This example may have seemed a little too obvious, but it was presented in order to illustrate the individual steps in the synthesis procedure.

Minimum Phase Linear System with Many More Poles than Zeros.

If one were to apply the technique in Figure 1 to the unalterable component shown in Figure 2a, which has a large number of poles but no zeros, he would be unsuccessful in writing a realizable desired mode of operation, because there is no realizable way of generating the inverse of this unalterable component. It is impossible to get a large number of zeros without poles in a realizable transference. Therefore it is impossible to get the desired system shown in Figure 1b if one has the given system shown in Figure 2a. At this point one must ask one's self, What is a possible of action? One possible mode of action is to permit a small amount of error due to a large number of poles located at very high frequency, and attenuating the very high frequencies. With respect to step responses this would mean that one would permit a very rapid rate of rise but not an infinite rate of rise on the output, represented by the cascading of a large number of very small time constants. Without specifying the power spectra of the signal and the noise into this system, it is meaningless to argue about the criteria for the best system or the best response. Let us take for our criteria the fact that Figure 2b is a realizable, possible mode of action, and that it is good enough. T is a very small time-constant, and n is equal to the number of factors in G_1.

Figure 2c shows this realizable mode of operation. It is equivalent to Figure 2b cascaded with the inverse of the unalterable component and the unalterable component. This system is now in a form suitable for block diagram substitutions. Again we will introduce an infinite gain amplifier to generate the inverse of the function

$(1 + sT)^n/G_1$. This is shown in Figure 2d.

Instead of expanding the feedback, two additional feedback loops will be introduced. In Figure 2e a major negative feedback loop has been introduced from the output to the input and in order not to change the input-output relationship an equivalent minor positive feedback loop has been added to the input. The outer feedback loop is the one which we wish to retain and the two inner feedback loops we wish to combine. It can be seen that they will combine in such a manner that the numerator polynomial of the feedback channel is equal to

$(1 + sT)^{n-1}$. The zeros represented by the function $Z_\bullet$ in Figure 2f are minimum phase, and quite easily constructed. They lie on a circle whose center is at $s = -1/T$, and whose radius is equal to $1/T$. There is one zero at the origin and all the others are equa-angularly spaced. The entire function Z/G_1 is therefore minimum phase and realizable. The complete system is not perfect, and the error is exactly equal to the output of the block Z/G_1. One can consider that this block is in parallel with the unalterable dynamic element $1/G_1$ and that this parallel combination generates as many zeros as poles, and therefore stabilizes the infinite gain loop. Z/G_1 can be thought of as a high frequency by-pass which takes the high frequencies which are important to stability, but unimportant to the output, and by-passes them around the unalterable element. Again, the infinite gain amplifier may be replaced by an amplifier with finite gain K, and $1/K$ in positive minor loop feedback. Combining the two minor loops results in an equivalent of $\left(\frac{Z}{G_1} - \frac{1}{K}\right)$. This shifts the zeros slightly from their circular locus.

Finite Gain Control

In cases where the input signal to a system is mixed with noise, it is not desirable for the system to faithfully reproduce both the signal and the noise at the output. If the attenuation rate of the signal is one decade-per-decade greater than the attenuation rate of the noise, then a single time-constant is the appropriate filter to separate the two, and the turnover frequency of this filter should equal the frequency at which the signal power density spectrum equals the noise power density spectrum. The design in this case is illustrated in Figure 3. Figure 3a is the given system with an unalterable component containing a large time-constant. Figure 3b is the desired filter in which a very small time-constant T_2 is chosen to provide the necessary filtering of the signal from the noise. The mode of operation of the system as a whole is derived from Figure 3b by multiplying by the inverse of the unalterable component, and then by the unalterable component. This is the first step toward finding the constructional arrangement. A major negative feedback loop is introduced from output to input, and its twin the corresponding positive feedback loop is also shown in Figure 3d. The inside loop is shrunk so that it encompasses only the control amplifier and its compensating networks in Figure 3e. Resolving this feedback system into its equivalent yields the final system shown in Figure 3f. Here the gain is not infinite at all frequencies, but has the constant component of T_1/T_2 which is a very large amount equal to the ratio of the unalterable time-constant to the desired time-constant. In addition there is an infinite gain component at zero frequency, produced by the integral action of $1/sT_2$. This control function looks surprisingly familiar. It is the type one has on a voltage regulator using an amplidyne in a high speed proportional loop and a motor driven field rheostat for the slow speed integral. A combination of the methods shown in Figures 2 and 3 can be applied to the most complex minimum phase linear systems. Now let us consider other types of systems.

Dead-Time Systems

The output-over-input transference of a conveyor belt is commonly known as dead time. This type of transference is also present in process controls with transportation lag, with flow delay between input and output, in flow reactors, and in fractionating towers. It is the transference of a matched reflectionless distortionless transmission line. Feedback systems in which the major component is a dead time are very difficult to stabilize if the gain is sufficient to make the system useful. If there is a single time constant cascaded with the dead time, the maximum permissible loop gain is equal to the ratio of the time constant to the dead time. If this ratio is low then the loop gain is not sufficient to override load disturbances, changes in the system transference due to nonlinearities, and noise generated in the system. These stability problems arise in process controls because the system design attempts to achieve a mode of operation which is unrealizable. The feedback loop from output to input is trying to make the output follow the input perfectly. But this is an impossible mode of operation, because if the input should change suddenly, absolutely nothing will happen at the output until after the elapse of a time equal to the dead time of the system. If we are to build the best dead-time control system, then we must invent a new statement of the mode of operation. This statement is: The output should exactly equal the input delayed by the amount of the dead time of the system. This is quite possible. If a conveyor belt loses none of its material, then whatever is dumped on at time t will come off at time $t + T_d$ where T_d is the system dead-time.

Let us apply this philosophy to the process control shown in Figure 4a. The given system has a dynamic minimum-phase member G_1, followed by a dead time T_1. An undesirable load disturbance is added to the system and the sum of these two components passes through the dead time T_2 to the output. There is also a feedback function F which is given as unalterable. A realizable mode of operation for this system is shown in Figure 4b A control system is built such that the input is reproduced faithfully at Θ_g. This then passes through the dead-time T_1 and T_2 producing at Θ_o a faithful reproduction of the input delayed by the time $T_1 + T_2$. In addition there is at the output an extra signal produced by the load disturbance Θ_L. We must find out what is the best realizable compensation for this undesired signal. In our actual system it is quite impossible for us to know whether or not the undesirable signal Θ_L exists until after the time T_2 when a measurement can be effected at the output. In addition if we immediately try to correct for the presence of Θ_L this correction will not reach the output until after an additional time $T_1 + T_2$. Therefore the best possible realizable correction for the application of a load disturbance would completely remove the load disturbance at the output of the system only after the time $T_1 + 2T_2$ had elapsed. Going back to our statement of the realizable mode of operation in Figure 4b, we will inject the load disturbance Θ_L delayed by the dead-time T_2 into the feed-back loop so that it will get to work and generate the negative of this signal at Θ_g as soon as possible to compensate for the actual load disturbance.

In Figure 4b it is assumed that the feedback loop with G_2 and F is "perfect". In actual practice we shall design a loop "satisfactory". This would be a design like Figure 3, where G_2

contains G_1 , and other components sufficient so so that the closed loop transfer function has high precision and speed of action. It is essentially the closed loop design that one would have if the dead times were removed from the system. In Figure 4c a major negative feedback loop from output to input has been introduced and its twin a minor positive feedback loop has also been added to keep the overall system transference equal to that of Figure 4b. The two minor loops are then combined to yield the final system shown in Figure 4d.

The minor feedback loop acts like a short term predictor. When the input is changed the minor feedback loop predicts immediately what the output will eventually do and introduces this in negative feedback. After the time $T_1 + T_2$ the minor feedback loop suddenly turns off and the major feedback loop suddenly comes on, so that the system switches automatically from predictor control to feedback control after the time $T_1 + T_2$ following each component of a transient. This system is linear, and the principle of superposition holds. It operates like a linear switch which switches from open loop control during the time $T_1 + T_2$ after each transient component to closed loop control after this time.

The method shown in Figure 4 is applicable to any non-minimum-phase system. All right-side s plane zeros have the characteristics of dead time. They cannot be removed by feedback. They can only be removed by feed-forward elements somewhat similar to the Z/G_1 component in Figure 2f. But even this is evading the basic characteristic of the non-minimum-phase elements, which is that they act like dead times. To handle non-minimum-phase systems, factor out the right-side zeros, and group with them a set of left-side s plane poles located at the mirror images of the zeros in the $j\omega$ axis. This product will be an all pass network with constant gain at all frequencies, but with a lagging phase shift which increases with frequency. Treat this combination in the same manner as the dead times shown in Figure 4.

Statistical Predictor

The formulation of the synthesis problem on a statistical basis is shown in Figure 5a. The feedback system has an input which is the sum of a desired signal θ_s and an undesired noise component θ_n (this may be the equivalent of load disturbances and other undesired signals in the system). The feedback system contains an unalterable output transducer with dynamic elements H and a variable control amplifier with compensating network G. One wishes to find the design for G, with H unalterable. The design for G should yield the minimum possible difference between the actual system output θ_o and the desired signal θ_s. This difference is defined as $\theta_e = \theta_s - \theta_o$. In Fig. 5a the error signal is actually shown. This is not usually available, either in an actual system, or in a laboratory set-up, since the undesired noise is inextricably mixed with the signal to form θ_i and cannot be separated. The arrangement shown in Fig. 5a is intended only to illustrate the equation for θ_e, and not the way in which it would actually be found.

The criteria for minimizing θ_e is dependent upon engineering judgment. We will not belabor the "best" criteria for the "best" system, but will state a single example. If the signal θ_s and the noise θ_n are continuous random functions with Gaussian probability distributions, (stationary random variables), then one may efficiently use the minimum error power as a criteria for design. Setting up the equations for average error power, and minimizing them, will yield the following design:

$$\frac{\theta_o}{\theta_i} = G_c(s) = \frac{1}{\phi^-_{ii}(s)} \mathcal{L}\mathcal{F}^{-1} \frac{\phi_{is}(\lambda)}{\phi^+_{ii}(\lambda)} \qquad (1)$$

$$s = j\lambda \qquad (2)$$

$$\phi_{ii} = \phi^+_{ii} \cdot \phi^-_{ii} \qquad (3)$$

$\phi_{is}(\lambda)$ is the cross power spectrum between input and signal. $\phi_{ii}(\lambda)$ is the self power spectrum of the input. $\phi^+_{ii}(s)$ has the left half s plane poles and zeros of the input self power spectrum. $\phi^-_{ii}(\lambda)$ had the lower half λ plane poles and zeros of the input power spectrum. The complex frequency variable s is equal to $j\lambda$. The transfer function G_c is the optimum transference for the complete feedback loop from θ_i to θ_o. The design of the component G within the loop is given by:

$$G = \frac{1}{H} \frac{G_c}{(1-G_c)} \qquad (4)$$

The design for G is always realizable if H is a minimum phase stable linear transference with an equal number of poles and zeros. The design for G is not always realizable from the above problem formulation when H does not fulfill these conditions. In these cases, our mathematics tells us (1) that one should not use a feedback system such as shown in Fig. 5a, and (2) that one should not minimize the error between the output and the desired signal.

What then should be the pattern for the realizable system, and what signal can be minimized for the best possible mode of action? These are answered in Figure 5b, where H is that part of the unalterable component which contains the excess poles over the number of zeros, and which contains the dead time or non-minimum-phase portions of the unalterable component. Since it is impossible to build a realizable negative dead time or a realizable inverse of the product of a large number of poles, the best that can be done is to predict the best possible signal to present to the input of this "terrible" unalterable component. The mathematically perfect predicted signal is shown as θ_p in Figure 5b. It is the mathematical signal passing through the transference 1/H. The actually available signal which must go through H is θ_a. We would like for the error between these two to be a minimum. This is a realizable form of control and this is a legitimate mode of action to ask for. Solving for the error power as defined in this manner and minimizing it yields

$$\frac{\theta_a}{\theta_i} = G_p(s) = \frac{1}{\phi^+_{ii}(s)} \mathcal{L}\mathcal{F}^{-1} \frac{\phi_{is}(\lambda)}{H(\lambda)\phi^-_{ii}(\lambda)} \qquad (5)$$

This is the design for the entire system preceding H.

This is still but a statement of the mode of control desired. It is not the actual constructional form. One may solve for G inside the feedback loop, from $G = G_p/(1-G_p)$. Following this step, the excess poles of H can be introduced as in Figure 3, and the non-minimum-phase part of H can be introduced as in Figure 4. The result is a feedback system considerably more complex than that shown in Figure 5a, and having a mode of action which it is impossible to describe by the simplified criteria in Figure 5a.

Dead-beat Control

One may be presented with a completely designed feedback system, which is very lightly damped, and be asked to make the transient responses of the system dead-beat by the addition of extra components, without eliminating any of the existing system. This is possible. The system may have a pair of dominant poles with transference

$$\frac{1}{1 + 2\zeta \frac{s}{\omega_o} + \left(\frac{s}{\omega_o}\right)^2} \qquad (6)$$

$$\omega_n = \omega_o \sqrt{1-\zeta^2} \qquad (7)$$

Complex zeros to exactly cancel these poles can be provided by the transference (1+P) in cascade with the input, where P is given by

$$P = K_o + K_1 \epsilon^{-sT_r/2} + K_2 \epsilon^{-sT_r} \qquad (8)$$

$$K_2 = \frac{1}{1 + k - 2k^{1/2} \cos \frac{\omega_n T_r}{2}} \qquad (9)$$

$$K_1 = K_2 \left[-2k^{1/2} \cos \frac{\omega_n T_r}{2}\right] \qquad (10)$$

$$K_o = K_2 \left[2k^{1/2} (\cos \frac{\omega_n T_r}{2}) - 1\right] \qquad (11)$$

$$K_o + K_1 + K_2 = 0 \qquad (12)$$

$$k^{-1} = 1 + \exp\left(\frac{-\zeta \omega_n T_r}{\sqrt{1-\zeta^2}}\right) \qquad (13)$$

T_r is the total step response time or total transient time. At the time T_r after an input step change, all of the transient components will have cancelled out. Another way to state it, is that the input compensator breaks the input transient into several individual transients at different times, and each of these excites its own oscillation which can be represented by a vector. The sum of these vectors is equal to zero at the end of time T_n. The response of this system does not approach the final value asymptotically but equals the final value exactly at the instant at which all of the derivatives at the output are equal to zero. The control function P can be either a single pulse or a double pulse generator. When the control function is a single pulse generator, one-half period response can be achieved. When the control function is a double pulse generator, dead-beat response in any small fraction of a cycle can be realized.

The application of this type of control to an existing system is shown in Figure 6. Figure 6a is the given unalterable unsatisfactory feedback system. It is characterized by Equation 6, having a dominant pair of complex poles. The oscillations in this system can be made negligible by introducing the function P in feed-forward from the input. The system will respond in a dead-beat manner to every input transient. If a similar pulse generator could be provided as feed-forward from the load disturbance as shown in Figure 6d then all load disturbances would be characterized by dead-beat motion of the output from an initial to a final value. This mode of operation is realizable.

To simplify the steps required we will combine the input and the load disturbances as an equivalent input signal as shown in Figure 6c. This also shows a single pulse generator P designed in accordance with Equation 8. The pulse generator P is shown as having an input equal to the system input plus a function of the system load. But the system load disturbance cannot be measured directly. It can only be derived from measurements at the feedback position or at the error position within the existing system. For this reason, the equivalent signal input to P is derived from system measurements as shown in Figure 6d. The branch point ahead of $G_1 G_2$ is moved through these dynamic elements and through the feedback block until it comes from the actual feedback as shown in Figure 6e. The feedback loop around P is isolated as a single local feedback loop and then the two adders are interchanged in position to form the resultant constructional arrangement in Figure 6f. The block $(1+1/FG_1G_2)$ can be approximated by a phase lead network with a gain between one-half and five and phase shift of less than $+180^\circ$ in the frequency bandof interest. The block P has no d-c gain. It is a pulse generator, and with the feedback branch around it, becomes a continuous pulse generator with s-plane poles. It can be built either as a single-reflection transmission line with feedback, or as a continuously reflecting transmission line which is not terminated in the characteristic impedance at either end.

Summary

Direct synthesis depends upon a conception of a mode of operation which is realizable, and includes the unalterable system components. The final system must have an excess of poles over the number of zeros equal to the excess of the unalterable components. The statement of this mode of operation is a block diagram. The values within the block diagram are derived from the criteria (based on engineering judgment) for "best realizable".

Additional block diagram parts are introduced by a series of substitutions, with steps for each of the following:

(1) The unalterable component must be in the system.
(2) Compensation for load disturbances and variations in component gains must be obtained from output feedback.
(3) Corrective forces can be applied only at the input to the unalterable element.
(4) Infinite gain amplifiers can be used to generate inverse functions only if they have not more than one more pole than the number of zeros.
(5) Resultant minor feedback loops should have very low or zero d-c gain.

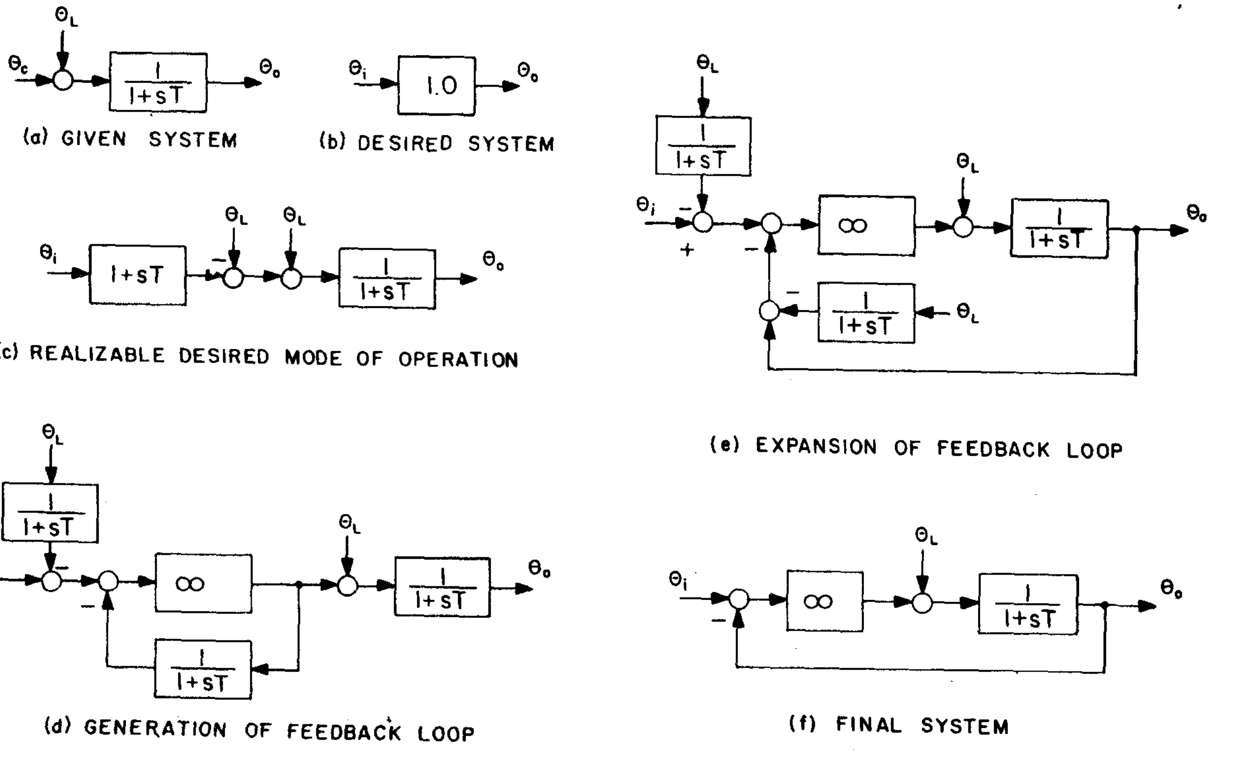


Fig. 1

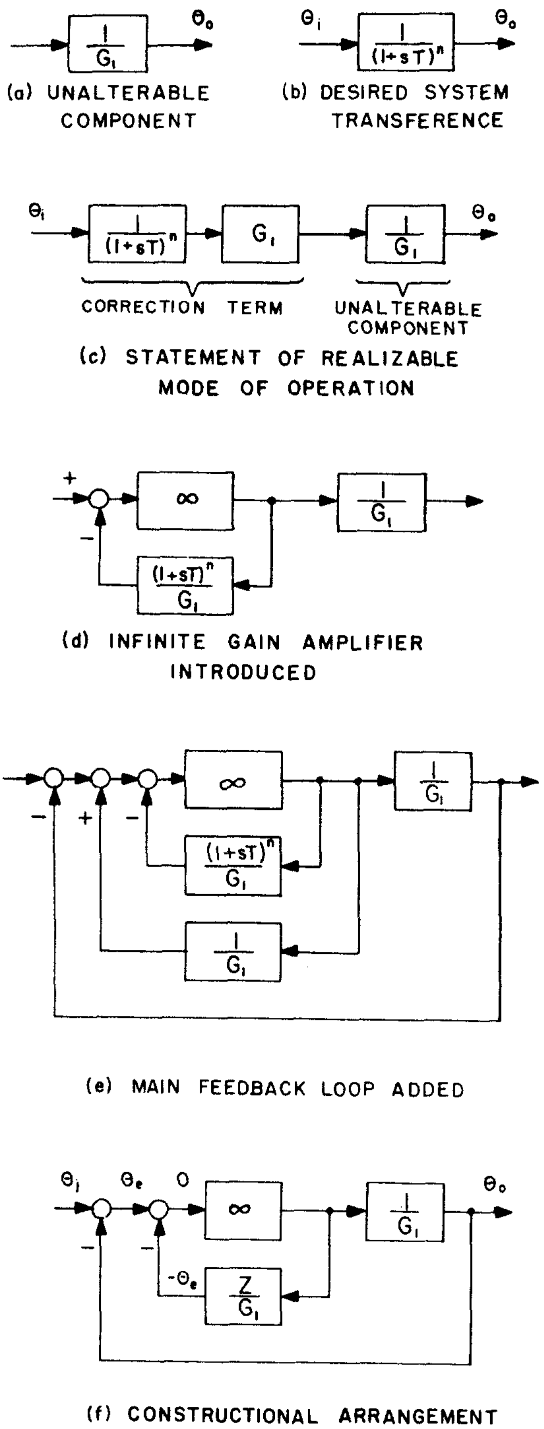

Θo
Θi
Θo
(a) UNALTERABLE COMPONENT
(b) DESIRED SYSTEM TRANSFERENCE
Θi
Θo
G1
CORRECTION TERM
UNALTERABLE COMPONENT
(c) STATEMENT OF REALIZABLE MODE OF OPERATION
∞
(d) INFINITE GAIN AMPLIFIER INTRODUCED
∞
(e) MAIN FEEDBACK LOOP ADDED
Θi
Θe
0
Θo
∞
-Θe
(f) CONSTRUCTIONAL ARRANGEMENT

Fig. 2

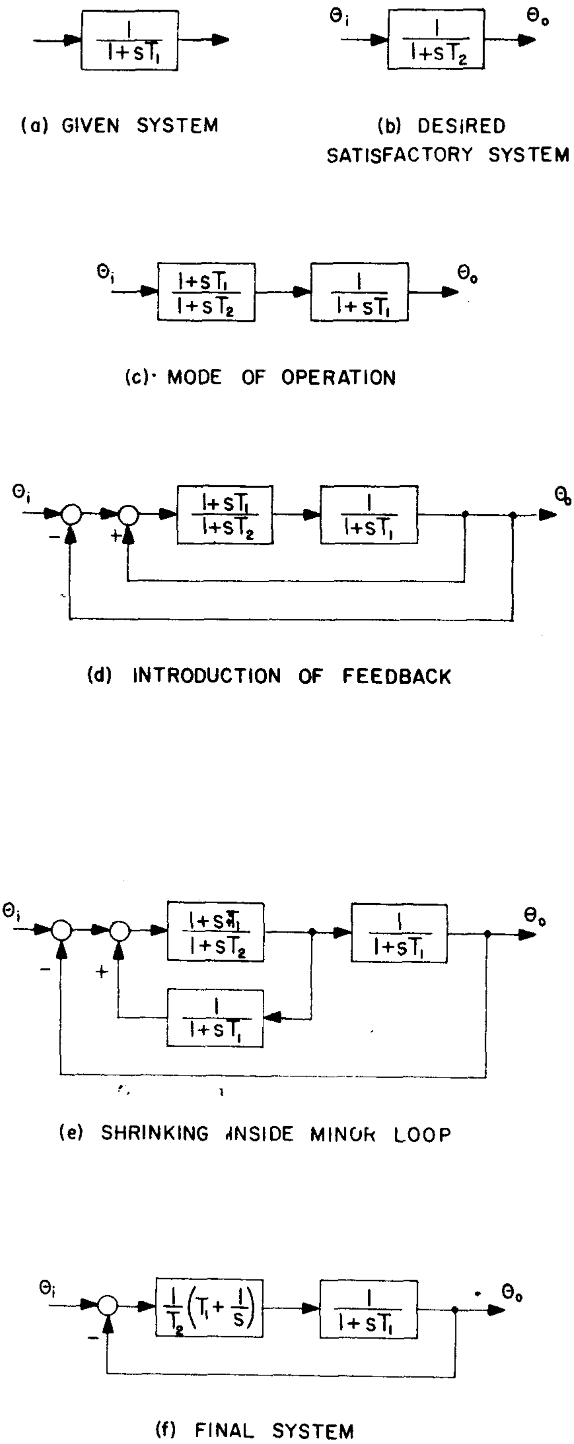
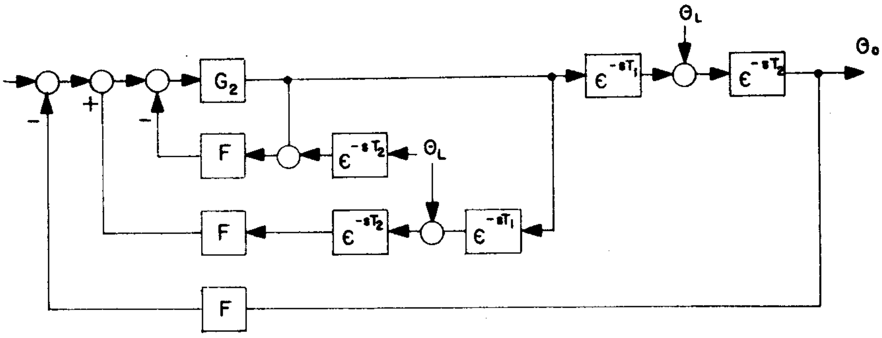
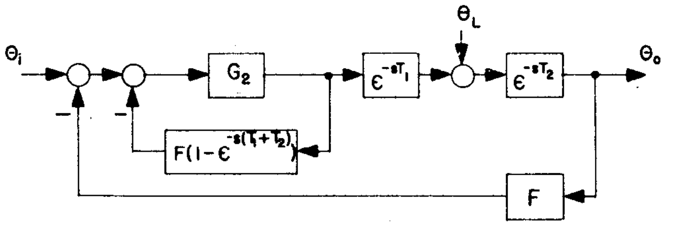

Θi
Θo
(a) GIVEN SYSTEM
(b) DESIRED SATISFACTORY SYSTEM
Θi
Θo
(c) MODE OF OPERATION
Θi
Θo
(d) INTRODUCTION OF FEEDBACK
Θi
Θo
(e) SHRINKING INSIDE MINOR LOOP
Θi
Θo
(f) FINAL SYSTEM

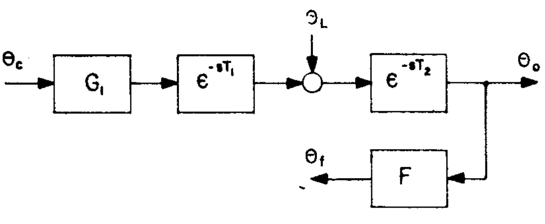
Fig. 3



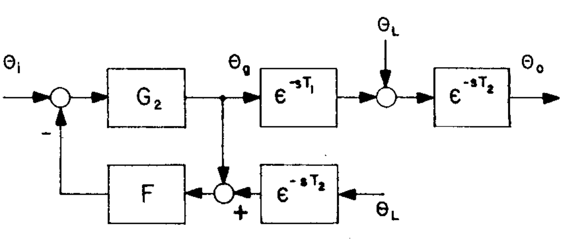
(c) INTRODUCTION OF MAJOR FEEDBACK LOOP AND SHRINKING OF INSIDE MINOR LOOP



(d) FINAL SYSTEM

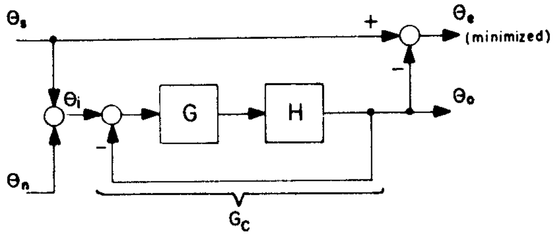


(a) GIVEN SYSTEM

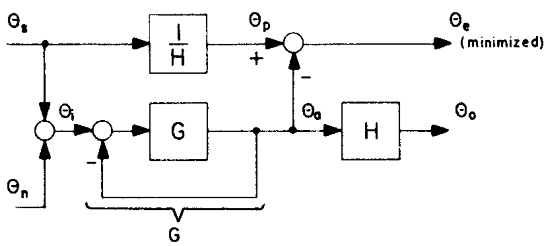


(b) REALIZABLE DESIRED MODE OF OPERATION

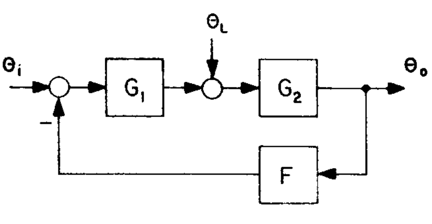
Fig. 4



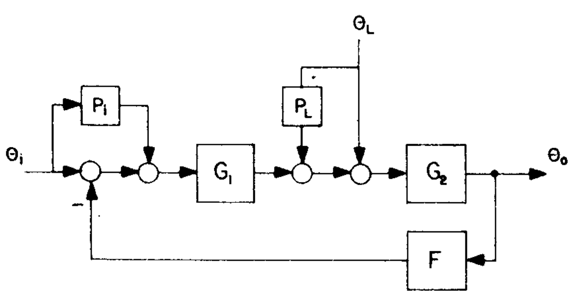
(a) FORM FOR MINIMUM-PHASE H WITH EQUAL NUMBER OF POLES AND ZEROS



(b) FORM FOR NON-MINIMUM-PHASE H WITH MORE POLES THAN ZEROS

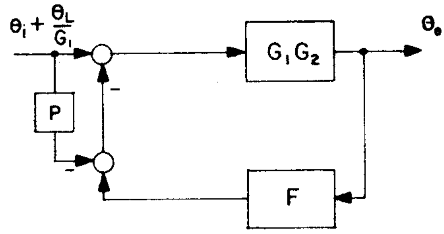


(a) GIVEN SYSTEM

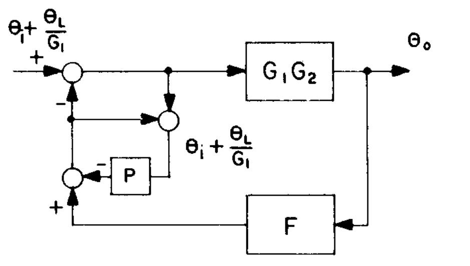


(b) REALIZABLE DESIRED MODE OF OPERATION

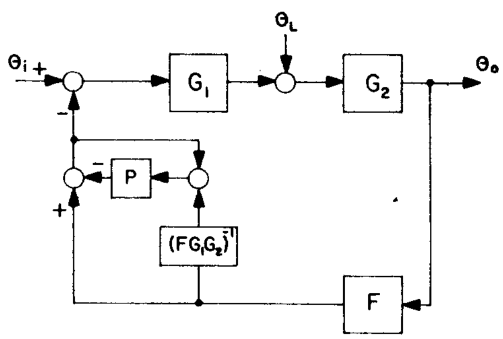
Fig. 5



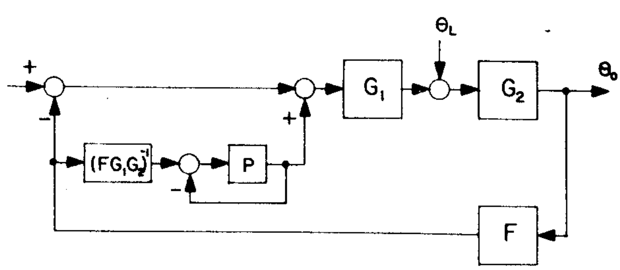
(c)



(d) DETECTION OF LOAD DISTURBANCE THROUGH SYSTEM MEASUREMENTS



(e)



(f) CONSTRUCTIONAL ARRANGEMENT

Fig. 6