PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Time-Lag in a Control System

A. Callender, D. R. Hartree and A. Porter

Phil. Trans. R. Soc. Lond. A 1936 **235**, 415-444
doi: 10.1098/rsta.1936.0007

XII—Time-Lag in a Control System

By A. Callender, *Research Department, I.C.I. (Alkali) Limited*, D. R. Hartree, *F.R.S.*, *and* A. Porter, *University of Manchester*

(Received July 15—Revised November 26, 1935—Read February 2, 1936)

1—Introduction

It often happens in physical experiments and in technical processes that there is some physical quantity, subject to random disturbances, which it is required to keep as nearly constant as possible by the operation of some controlling gear. For example, it may be required to keep a room, or a reaction vessel, at a constant temperature by controlling the electric current passing through a heating coil, or steam passing through heating pipes, or the operation of a heating or cooling engine.

The operation of the controlling gear could be by trial and error, but it may often be desirable that it shall be made to depend in some definite way on the behaviour of the physical quantity to be controlled ; the determination of such behaviour and the consequent controlling action may each be carried out either by an operator or by automatic means. It is convenient to make a distinction between the element (such as rheostat or steam valve) whose setting directly affects the physical quantity concerned, and the apparatus employed (if any) to make appropriate adjustments of this setting. We will call the former the " controlling gear " and the latter the " control apparatus " and the two together the " control system ".

Whether the controlling action is automatic, or contains a manual element, it may often happen that there is a time-lag between any particular incident in the behaviour of the quantity being controlled, and the result of the operation of the controlling gear brought about by this incident ; in the example considered, there may be a time-lag between the occurrence of a rise of temperature in the room and the decrease of emission of heat from the heating elements, due to the time required for the thermometer, which is recording the temperature in the room, to acquire this temperature, and to the time required for the resistance to cool down when the current through it is decreased. Time-lags of several minutes may occur in large scale practice ; a value of the order of 1 minute is common.

The primary object of the work summarized in this paper was to make a theoretical study of a general class of such control systems with a time-lag, and to obtain some guidance as to the optimum values of the constants in the relation between the quantity controlled and the operation of the controlling gear.

The argument is general, and is independent of the particular physical quantity controlled (*e.g.*, temperature, pressure, CO_2 content, voltage) and of the particular type of controlling gear (*e.g.*, steam valve or rheostat), though we shall later indicate briefly the application of the results to particular kinds of controlling gear as examples.

A systematic theoretical investigation of the factors involved in the process of control is interesting in itself, and is desirable, if not absolutely necessary, as a basis for detailed design of control apparatus, as its results are much more precise and definite, as well as more general, than any which could be reached by a purely empirical investigation with an actual control system, and also because the field is too extensive to be covered adequately by such an empirical study.

An investigation somewhat similar to that here described has been given recently by HAZEN,* in a theoretical study of the operation of servo-mechanisms ; he studies the effect of time-lag in connexion with two kinds of servo-mechanism, but not in connexion with what he terms "continuous-control servo-mechanism", which corresponds most closely to the type of control with which we are concerned. Our problem of control of a quantity subject also to uncontrolled disturbances is somewhat different from that of control of a power supply by a *given* motion of one part of the mechanism ; also the combination of continuous control and the occurrence of a time-lag is fundamental here, so that this work, though related to HAZEN's, does not overlap it, and was actually carried out independently.

There are two conditions which control must satisfy in order to be of practical value ; it must be stable, that is to say it must not give rise to oscillations of increasing amplitude in the quantity controlled, and it must be quick-acting, that is to say, if the quantity controlled departs from its standard value, the control must operate so as to bring it back to that value as soon as possible.

It is easy to see in a qualitative way that, especially when there is a time-lag in the control system, these conditions are likely to be to some extent incompatible. Consider for definiteness temperature control, and suppose that the temperature increases. The control operates so as to decrease it towards the normal or standard value, but on account of the time-lag the control does not respond to the temperature reaching the normal value until after this value has been passed† ; that is, the temperature overshoots the normal value and becomes too low. The shorter the time required for the return to the normal value, compared with the time-lag, the further will the temperature overshoot before the control responds to its reaching the normal value, and it is clear that there is a danger of setting up an oscillation of increasing amplitude if we try to make the return to the normal value too rapid.

* 'J. Franklin Inst.,' vol. 218, p. 279 (1934).

† For simplicity, the argument is here worded as if the operation of the control depended on the value of the temperature only. Actually the situation will be less simple, as the operation of the control may also depend on the time derivatives and integrals of the temperature. Indeed, it is just by making the operation of the control depend on these latter quantities that we can hope to lessen the degree to which the two conditions are incompatible.

It should be realized that the complete elimination of fluctuations is impossible, because it is only by the occurrence of such fluctuations that appropriate controlling actions are called for.

It should also be mentioned that in practical examples a formulation involving a time-lag is not always an exact way of describing the behaviour of the controlled system and the action of the controlling gear on it, but rather is often an approximate way of describing the behaviour of the solution of some set of differential equations which are too complicated to formulate precisely or to solve. For example, the relation between the increase of flow of steam through heating pipes and the increase of temperature of the contents of a vessel should strictly be described by the equations of heat transfer through the walls of the pipes and the containing vessel and through the contents of the vessel ; but for practical purposes this relation can be approximately described by a time-lag.

2—General Statement of the Problem

We shall write $\theta(t)$ for the departure, at time t, of the quantity to be controlled from its normal or standard value. Variation of $\theta(t)$ may be due to three causes ; firstly through uncontrolled disturbances, such as fluctuations in the temperature of the surroundings of a vessel which we require to keep at constant temperature, or variations of voltage on the mains from which the current for heating coils is taken ; secondly, through the operation of the control gear, and thirdly, apart from changes due to these causes, a departure of θ from zero may in itself give rise to a variation of θ as for a vessel in a constant temperature bath.

We shall suppose that in general*

$$\frac{d\theta(t)}{dt} = \mathrm{D}(t) + \mathrm{C}(t) - m\theta(t), \qquad (1)$$

where $\mathrm{D}(t)$ is the effect of uncontrolled disturbances, which is to be regarded as a given function of t, and will be called the disturbing function, $\mathrm{C}(t)$ is the effect, at time t, of the operation of the control, and will be called the controlling function, and $-m\,\theta(t)$ is the inherent effect of variation of $\theta(t)$ from its zero value.

We shall suppose that the mode of operation of the control is determined by the behaviour of the quantity θ to be controlled, so that the disturbing function $\mathrm{D}(t)$ affects the control, not directly, but only through the variation of θ to which it gives rise. For a system with time-lag, the function $\mathrm{C}(t)$ then depends on the behaviour of θ, not at time t, but at time $t - \mathrm{T}$, where T is the time-lag, which we shall suppose

* [*Note added in proof*, *April* 22, 1936.—The methods of this paper can be applied to a more general case in which the system possesses inertia or its equivalent, the effect of which is to add a term proportional to $d^2\theta(t)/dt^2$ to the left-hand side of (1). Results for this case, when the true time-lag T is zero, can easily be obtained, and, although of little practical value, are quite interesting, especially as a basis of comparison.]

to be constant. Then the effect of the control, at time $t + \mathrm{T}$, is to be determined by the behaviour of $\theta(t)$, at time t. The dependence of $\mathrm{C}(t + \mathrm{T})$ on $\theta(t)$ describes the behaviour of the control, and this dependence we shall call the law of control; the law which we shall consider here is

$$-\dot{\mathrm{C}}(t+\mathrm{T}) = n_1\theta(t) + n_2\dot{\theta}(t) + n_3\ddot{\theta}(t) \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where dots denote time differentiation as usual, and n_1, n_2, n_3 are constants.

To see the physical significance of a law of control of this general type, consider for the moment the example of temperature control by steam heating, the steam supply being controlled by a valve. Then the setting of the valve determines the controlling function, and a law of control of the form (2), with $n_1 \neq 0$, means that not the valve setting itself, but its time rate of change, depends on the values of the temperature and its derivatives, so that the valve setting itself depends not only on the behaviour of the temperature at an instant, but on the time integral of the temperature, and so on its previous history. The dependence of $\mathrm{C}(t + \mathrm{T})$ on $\int \theta \, dt$ when $n_1 \neq 0$, illustrated by this example, has important consequences, as will appear in the further discussion. One consequence may be noted at once, namely, that even if θ has the same behaviour at two times t_1 and t_2, $\mathrm{C}(t + \mathrm{T})$ may not be the same for $t = t_1$ and $t = t_2$, since $\int_{t_1}^{t_2} \theta \, dt$ need not be zero.

It may be possible to obtain better control with control laws of other general types,* but a law of control of the type (2) is easily attained in practice, and, for a certain region of values of n_1, n_2, n_3, m, and T, it gives a stable control, as will appear from § 3. For the control to be satisfactory in practice, it is necessary that the set of values of the constants should lie in the region for which the control is stable, but this is not by any means sufficient, and we shall be concerned to determine the suitable ranges of these constants for practical control, rather than the boundary of the region for which the control is stable.

In the analysis it is usually convenient to take the time-lag T as unit of time. We shall write

$$t/\mathrm{T} = \tau, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and

$$m\mathrm{T} = \mu, \qquad n_1\mathrm{T}^2 = \nu_1, \qquad n_2\mathrm{T} = \nu_2, \qquad n_3 = \nu_3, \quad . \quad . \quad . \quad . \quad (4)$$

and shall call ν_1, ν_2, ν_3 the " control constants " of the system.

* [*Note added in proof, April* 22, 1936.—For example, a control law of the form

$$-\dot{\mathrm{C}}(t+\mathrm{T}) = n_1 z(t) + n_2 \dot{z}(t) + n_3\ddot{z}(t),$$

where $z(t)$ is an auxiliary variable related to $\theta(t)$ by

$$\dot{z}(t) + \mathrm{B}_1 z(t) = \mathrm{B}_2 \dot{\theta}(t) + \mathrm{B}_1 \theta(t),$$

where B_1 and B_2 are constants, appears to have certain advantages and to be easily attainable in practice, and the behaviour of control systems operating according to such a law of control is being investigated on lines parallel to those of the present paper.]

We shall also write

$$\mathrm{T\,D}(t) = \psi(\tau), \qquad \mathrm{T\,C}(t) = c(\tau) \qquad \ldots \ldots \ldots \quad (5)$$

and refer to $\psi(\tau)$ as the disturbing function and write $\theta(\tau)$ for θ regarded as a function of τ rather than of t. Then (1) and (2) become

$$\frac{d\theta(\tau)}{d\tau} = \psi(\tau) + c(\tau) - \mu\theta(\tau), \qquad \ldots \ldots \ldots \quad (6)$$

$$-\frac{dc(\tau+1)}{d\tau} = \nu_1\theta(\tau) + \nu_2\frac{d\theta(\tau)}{d\tau} + \nu_3\frac{d^2\theta(\tau)}{d\tau^2}. \qquad \ldots \ldots \quad (7)$$

If (7) is written in the integrated form

$$-c(\tau+1) = \nu_1\int^{\tau}\theta(\tau)\,d\tau + \nu_2\theta(\tau) + \nu_3\frac{d\theta(\tau)}{d\tau}, \qquad \ldots \ldots \quad (8)$$

it might at first sight appear that, on account of the indefinite lower limit in $\int\theta\,d\tau$, $c(\tau)$ contains an arbitrary additive constant whose value might affect the behaviour of the system. But clearly, if there is no disturbance (*i.e.*, $\psi = 0$), and θ happens to be steady at the value 0, the effect of the control must not be such as to change this behaviour of θ, for any such change produced in these circumstances would be contrary to the whole idea of control. From (6), (7) this means that if at any time τ_0 (either initially or in the course of the operation of the control system) a state is reached in which

$$\psi(\tau_0) = 0 \quad \text{and} \quad \theta(\tau_0) = \frac{d\theta(\tau_0)}{d\tau} = \frac{d^2\theta(\tau_0)}{d\tau^2} = 0, \qquad \ldots \ldots \quad (9)$$

then

$$c(\tau_0+1) = 0. \qquad \ldots \ldots \ldots \ldots \quad (10)$$

This condition, which corresponds to the determination of the zero setting of the control gear (*e.g.*, of the steam valve or rheostat for temperature control), is equivalent to the specification of the constant of integration, or the lower limit of the integral, in (8), and this condition will be satisfied in all the solutions we shall consider. The integral in (8) is to be read in this sense, not as an indefinite integral, though no lower limit is specified as the appropriate lower limit will be different in different cases.

It may be mentioned that any departure from the condition expressed by (9) and (10) can be treated as a constant disturbance, so that this condition forms no essential restriction on the solution of (6) and (7).

From the form of these equations (6) and (7) it can be seen that, apart from a change of time scale, the general behaviour of the system depends on μ and the control constants only, and (4) gives the way in which they change when the time-lag T changes.

The equations (1) and (2), or (6) and (7), are linear, as is desirable both for practical and for analytical reasons in order that the superposition principle should apply to their solutions, so that the effects of disturbances occurring at different times should be additive. In practice, the variations of θ are likely to be small, so that linear equations should provide an adequate representation of the behaviour of a real system.

Concerning the form of the law of control, the following general considerations show that the inclusion of at least the terms with coefficients ν_1 and ν_2 is desirable.

If $\nu_1 = 0$, then for $\psi(\tau) =$ constant, equations (6), (7) have a particular solution $\theta = \psi/(\mu + \nu_2)$, whereas $\theta = 0$ is not a solution, showing that if the disturbing function tends to a non-zero value as $\tau \to \infty$, θ may also tend to a non-zero value, and in any case cannot tend to 0. One of the objects of a control system should be to avoid this if possible, and to ensure that even for a constant disturbance, $\theta \to 0$ as $\tau \to \infty$.

If $\nu_1 \neq 0$, on the other hand, $\theta = 0$ is a solution of equations (6), (7) with $\psi(\tau) =$ const., whereas $\theta =$ const. $\neq 0$ is not; and this suggests that taking $\nu_1 \neq 0$ may have the desired effect of ensuring that for a constant disturbance $\theta \to 0$ as $\tau \to \infty$. It will appear later* that, provided ν_1, ν_2, ν_3, μ lie in the region for which the control system described by the equations is stable, this is the case. Speaking descriptively, what happens in such a case is that θ varies until ultimately $\int \theta\, d\tau$ builds up to such a value that the term $-\nu_1 \int \theta\, d\tau$ in $c(\tau + 1)$ (*see* 8), cancels the term ψ in (6) representing the constant disturbance.†

Thus for satisfactory control it appears desirable to include the term in (7) with coefficient ν_1. On the other hand, this term alone would lead to a very sluggish control, as its effect on the controlling function $c(\tau)$ depends on the time integral of θ, and thus takes some time to build up, especially when using for ν_1 a value small enough to give stability. The ν_2 term is therefore necessary in addition, in order to quicken the response of the control to variations of $\theta(\tau)$, and it will appear that a law of control involving ν_1 and ν_2 alone will give satisfactory control with suitable values of the control constants. The incorporation of the ν_3 term, by giving, effectively, still earlier indication of the incipient deviations of $\theta(\tau)$, greatly hastens the checking of these deviations, and, provided ν_3 is not made too large, values of ν_1 and ν_2 can be found so that the control is still satisfactory from the point of view

* The proof depends on the results of § 4; we are concerned here to show in a general way the reasons for adopting a law of control of the form (2) as one whose consequences it is desirable to investigate, and to indicate the significance of the various terms involved, rather than to establish these consequences.

† There is no inconsistency between (9), (10), according to which, if θ is zero and steady, and $\psi = 0$, then $c(\tau) = 0$, on the one hand, and the possibility when $\nu_1 \neq 0$, that if ψ tends to a constant non-zero value, θ may ultimately become zero and steady, and $c(\tau) \to -\psi \neq 0$, on the other hand, since, as we have already pointed out more generally, one consequence of taking $\nu_1 \neq 0$ is that $c(\tau_2 + 1)$ may be different from $c(\tau_1 + 1)$, although the behaviour of θ at τ_1 is the same as at τ_2.

of stability. The practical design of control gear giving a law of control including a ν_3 term may involve some difficulties, but is possible (perhaps with some restrictions on the values of the constants obtainable); the practical difficulties of including terms depending on still higher derivatives of $\theta(\tau)$ appear severe, and hence we shall restrict discussion to a law of control involving the three control constants ν_1, ν_2, ν_3.

For general discussion, and for analytical treatment, it is convenient to consider a particular class of disturbances, namely, those for which

$$\psi(\tau) = 0 \quad \text{for} \quad \tau < 0, \qquad \psi(\tau) = 1 \quad \text{for} \quad \tau > 1,$$

the variation of $\psi(\tau)$ for $0 < \tau < 1$ being specified; one question which arises is that of the comparative effects of disturbances whose onset is represented by different functions $\psi(\tau)$ in the range $0 < \tau < 1$. By the superposition principle, the effect of a general disturbance can be analysed into the effects of disturbances of this particular class, multiplied by suitable factors and beginning at different times, so that by studying the effects of typical disturbances of this kind we shall get a survey of the general behaviour of the system. For a similar reason we may also limit ourselves to solutions for which $\theta(\tau) = 0$ for $\tau < 0$; that is to say, effects of previous disturbances are omitted, and we may consider the effect of each disturbance on the understanding that the quantity to be controlled is steady at its normal value until that disturbance occurs; §§ 4–6 are concerned with solutions of this kind.

The advantage of concentrating attention on disturbances of this kind is that then the behaviour of $\theta(\tau)$ from $\tau = 1$ onwards is determined only by the operation of the control and by $\theta(\tau)$ itself. If $\theta(\tau) = 0$ for $\tau < 0$, then $c(\tau) = 0$ for $\tau < 1$, and the variation of $\theta(\tau)$ in the range $0 < \tau < 1$ is determined by the form of $\psi(\tau)$; this variation of $\theta(\tau)$ in the first interval $(0 < \tau < 1)$ determines the variation of the controlling function $c(\tau)$ in the second interval $(1 < \tau < 2)$ by equation (7), and since $\psi(\tau)$ is now constant, this determines the variation of $\theta(\tau)$ in this interval, by (6), and so on. Thus, subsequent to $\tau = 1$, the variation of $\theta(\tau)$ in any interval $(n < \tau < n + 1)$ is determined by its variation in the previous interval, and determines its variation in the next interval. As already pointed out, in order to be of practical use the control must be stable, that is to say, the range of variation of θ for such a solution must decrease as time goes on, and the more rapidly it decreases, the sooner effects of disturbances disappear.

Although it is convenient to introduce the disturbing function $\psi(\tau)$, both to represent what does occur physically, and from the point of view of subsequent analysis of the effects of different disturbances, these particular solutions can be dealt with without reference to a disturbing function, as follows. The equations with $\psi(\tau) = 0$ have a number of solutions which may be called "normal modes". Just as in a dynamical system we may study the motion resulting from given initial conditions, without enquiring by what disturbance these initial conditions were set up, so here we may study the variation of θ arising from the operation of the control according to a certain initial behaviour of θ, without explicitly introducing any disturbance ψ to give the initial variation of θ. In the dynamical case the initial conditions required are the positions and velocities of the bodies concerned at a single instant, but in the present case the initial conditions consist of the behaviour of $\theta(\tau)$ throughout the time interval $0 < \tau < 1$, as well as the value of $c(1)$ (in which a constant $\psi(\tau)$ for $\tau > 1$ can be included). Also, just as in the dynamical case of a vibrating system, we shall find that the solution from given initial conditions can be expressed as the sum of a number of normal modes with appropriate amplitudes and phases.

It should be noted that the value of $\theta(1)$ (or, alternatively, the maximum value of $\theta(\tau)$ for $0 < \tau < 1$) is inevitable, in the sense that it cannot be affected by any operation of the control, for no action of the control based on the behaviour of $\theta(\tau)$ for $0 < \tau < 1$ can influence $\theta(\tau)$ for $\tau < 1$.

Three methods of investigation have been used for the study of the control equations (6), (7), namely :—

(*a*) Determination of the special solutions, which we have called "normal modes", of the equations, with $\psi(\tau) = 0$. This will be considered in § 3. The behaviour of $\theta(\tau)$ in each of these is found to be exponential or damped harmonic, and their frequencies and damping constants can be found. The primary question of stability can then be dealt with by the condition that the damping constants must all be positive, but this method gives no easy way of finding the amplitudes and phases of the normal modes required to fit any given starting conditions (*i.e.*, given behaviour of $\theta(\tau)$ or $\psi(\tau)$ in the range $0 < \tau < 1$), and hence details of the way in which the initial deviation of $\theta(\tau)$ from zero is checked by the operation of the control cannot be studied.

(*b*) Use of Heaviside operators. Equations (6) and (7) form a pair of linear difference-differential equations with constant coefficients, and invite treatment by Heaviside's operational method* ; this will be considered in § 4, where it will be seen that this method provides analytical solutions for any initial conditions [or any disturbing function $\psi(\tau)$], and so enables the way in which the operation of the control depends on the disturbance to be studied.

(*c*) Numerical investigation of particular cases, either by arithmetical or graphical methods or by use of the differential analyser. This is considered in § 5.

3—The Normal Modes

It may be anticipated that, for $\psi(\tau) = 0$, equations (6), (7), being linear, will have special solutions which are exponential, oscillatory, or damped oscillatory, and we therefore first consider solutions of this form, namely

$$\theta(\tau) = Q e^{\gamma\tau}, \quad \ldots \ldots \quad (11)$$

where Q, γ are constants which may be real or complex.

* *See* Jeffreys, " Operational Methods in Mathematical Physics " (Cambridge, 1927).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Substituting in (6), (7) we find that Q is arbitrary, and that γ must satisfy the equation

$$F(\gamma) \equiv \gamma(\gamma+\mu) + e^{-\gamma}(\nu_1+\nu_2\gamma+\nu_3\gamma^2) = 0. \quad . \quad . \quad . \quad . \quad . \quad (12)$$

This equation has an infinite number of roots, complex roots occurring in conjugate pairs; if the control is to be stable, the real part of γ must be negative or zero for every root, and for the control to be satisfactory in practice, the real part of γ must be definitely negative for each complex root,* so we shall write

$$\gamma = -\alpha \pm i\beta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

and require that α shall be positive; it will be convenient to take β as essentially positive. We shall refer to the normal mode with the smallest value of β as the "fundamental" or "first harmonic" and to those with larger β as "harmonics". If there is a real root $\gamma = -\alpha_0$, we shall refer to the corresponding normal mode as the "simple exponential" (for some values of the control constants, there may be more than one such normal mode).

It should be noted that, for $\nu_1 \neq 0$, $\gamma = 0$ is *not* a solution of (12), so that $\theta =$ constant is not a normal mode, but when $\nu_1 = 0$, $\gamma = 0$ is a solution of (12); and also that in general (12) has no purely imaginary root,† so that there is generally no undamped normal mode.

We shall consider the solutions of (12) under three heads:

(i) $\mu = 0$, $\nu_3 = 0$, ν_1 and ν_2 non-zero;

(ii) $\mu = 0$, ν_1, ν_2, and ν_3 non-zero;

(iii) $\mu > 0$.

With four independent parameters, the field to be investigated is rather extensive; cases (i) and (ii) have been examined rather thoroughly, case (iii) less so. Control is clearly easier to achieve when $\mu > 0$ than when $\mu = 0$, hence the determination of suitable control constants in the latter case demanded thorough investigation in the first instance. Also in practical examples of control problems μ is often so small compared to ν_2 that $\mu = 0$ is a sufficiently accurate simplifying approximation.

Case (i) $\mu = 0$, $\nu_3 = 0$—As already mentioned, it may not always be easy to design a practical control for which $\nu_3 \neq 0$, so in the first instance we shall consider the case $\nu_3 = 0$; this will also be found to give a basis from which to start the investigation of the more general case $\nu_3 \neq 0$, and will provide a comparison to show the effect of ν_3 on the behaviour of the control. In this case, (12) becomes

$$\gamma^2 e^{\gamma} + \nu_1 + \nu_2\gamma = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

* If one or more values of γ were pure imaginary, it would be possible for θ, once disturbed from the value zero, to continue to oscillate with undiminished amplitude, although the disturbance ceased; and this would not be satisfactory from the point of view of control.

† Substitution of $\gamma = \pm i\beta$ in (12), and separation of real and imaginary parts, gives two equations for β which are inconsistent unless there is a particular relation between ν_1, ν_2, ν_3, and μ. This relation specifies the boundary of the region of values of ν_1, ν_2, ν_3, μ for which the system described by equations (6) and (7) is stable.

The real roots $\gamma = -\alpha_0$ if any, are given by the intersection of the curves of $-\alpha^2 e^{-\alpha}$ and $\nu_1 - \nu_2\alpha$ (*see* fig. 1). It is clear that for any negative value of ν_1, there is an intersection at a negative value of α, giving an increasing exponential behaviour for θ; hence ν_1 must be positive, as would be expected from the general point of view of the object of the control. It may similarly be expected that ν_2 must be

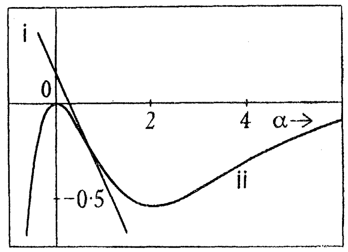


FIG. 1—Curve i, $\nu_1 - \nu_2\alpha$; curve ii, $-\alpha^2 e^{-\alpha}$.

positive (and study of the complex roots of (14) confirms this); and for positive values of ν_1, ν_2 it is clear from fig. 1 that there is certainly one, and usually only one, real root $\gamma = -\alpha_0$; and in order that it should not be too small, ν_1 must not be too small or ν_2 too large.

To obtain a survey of the complex roots $\gamma = -\alpha \pm i\beta$ it is convenient to use a diagram such as fig. 2 showing contours of constant ν_1 and of constant ν_2 in the

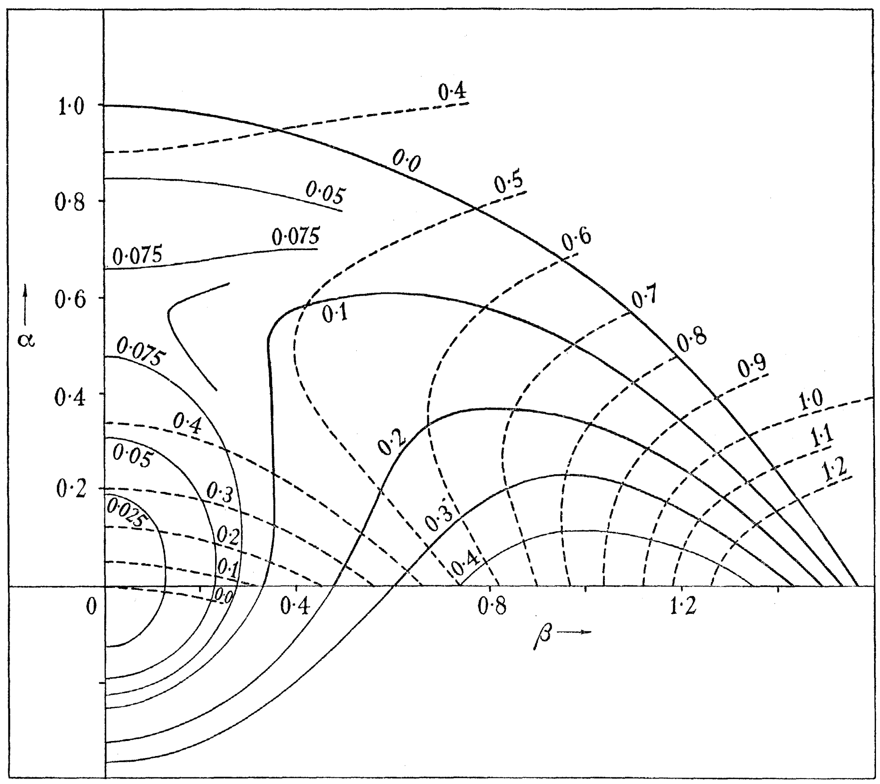


FIG. 2—ν_1 and ν_2 contours on an (α, β) diagram for $\nu_3 = 0$, $\mu = 0$. Contours of constant ν_1 ——; contours of constant ν_2 - - -.

(α, β) plane; the values of α, β for any ν_1, ν_2 are then given by the intersections of the corresponding ν contours; fig. 2 includes the neighbourhood in which the fundamental lies in systems of practical importance. It will be shown later, without recourse to any approximation, that, when $\nu_3 < 1$, then if α is positive for the fundamental, it is positive for all the harmonics, so that to ensure positive damping for the oscillatory normal modes it is sufficient to consider the fundamental alone. Nevertheless, since high damping of the fundamental and high damping of the simple exponential are of limited compatibility, the latter also must be kept in view and a suitable compromise arrived at.

Usually for the high harmonics, and often for the fundamental also, α and ν_1/ν_2 are fairly small compared to β, and then an approximate solution of (14) is given by

$$\beta = (2n + \tfrac{1}{2})\pi - [(\nu_1/\nu_2) + \log(\beta/\nu_2)]/\beta$$

$$\alpha = \log(\beta/\nu_2).$$

Hence, as far as this approximation goes, α increases with β, so that if the fundamental is positively damped, the higher harmonics are more so; and also α/β ultimately decreases with increasing β (though it may increase at first), so that the approximation becomes better for the higher harmonics. Table I shows a test of the approximate solution, in which it will be seen to be surprisingly good for the case considered.

TABLE I—FREQUENCIES AND DAMPING CONSTANTS, AND VALUES OF $F'(\gamma)$, FOR THE NORMAL MODES FOR $\nu_1 = 0 \cdot 3$, $\nu_2 = 1$, $\nu_3 = \mu = 0$.

Order of harmonic	Correct values		Values from approx. formulae		$F'(\gamma)$
$n+1$	α	$\beta - (2n+\frac{1}{2})\pi$	α	$\beta - (2n+\frac{1}{2})\pi$	
Simple exponential	0·4128	—	—	—	0·856
1 (fundamental).	0·1719	−0·4010	0·1633	−0·3935	$-1\cdot483\ e^{(-0\cdot6061\,i)}$
2.	2·0671	−0·3044	2·021	−0·3077	$-59\cdot39\ e^{(+0\cdot4095\,i)}$
3.	2·6555	−0·2092	2·634	−0·211	$-198\cdot6\ e^{(+0\cdot3073\,i)}$
4.	3·0216	−0·1622	3·008	−0·163	$-417\ e^{(+0\cdot2482\,i)}$
5.	3·2885	−0·1343	3·283	−0·135	$-719\ e^{(+0\cdot2076\,i)}$

Case (ii) $\mu = 0$, $\nu_3 \neq 0$—When we extend the investigation to three independent control constants, ν_1, ν_2, ν_3, but still $\mu = 0$, the equation for the complex frequencies γ of the normal modes becomes

$$\gamma^2 e^{\gamma} + \nu_1 + \nu_2\gamma + \nu_3\gamma^2 = 0. \quad \dots\dots\dots\dots \quad (15)$$

The real roots $\gamma = -\alpha_0$ are given by the intersections of the parabola $\nu_1 - \nu_2\alpha + \nu_3\alpha^2$ with the curve $-\alpha^2 e^{-\alpha}$; if $\nu_2{}^2 < 4\,\nu_1\nu_3$ the parabola will not cross the α axis, so there will certainly be no intersection and no real roots, and there may be no real root even when this condition is not satisfied. This is the first

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

distinction from the case when $\nu_3 = 0$, when there is always a real root giving a "simple exponential" normal mode.

It is convenient to refer a case with $\nu_3 \neq 0$ to the case with $\nu_3 = 0$ which has the same value of γ for the fundamental; we shall call the values of ν_1, ν_2 for the latter case the "basic values", and write them ν_1', ν_2'.

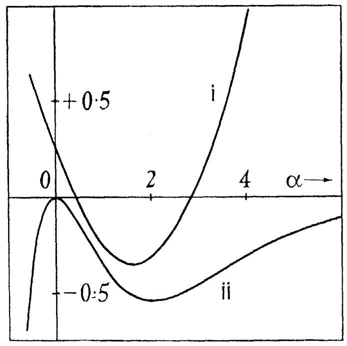


FIG. 3—Curve i, $\nu_1 - \nu_2\alpha + \nu_3\alpha^2$; curve ii, $-\alpha^2 e^{-\alpha}$.

Then

$$\nu_1' + \nu_2'\gamma = \nu_1 + \nu_2\gamma + \nu_3\gamma^2$$

and writing $\gamma = -\alpha + i\beta$ and separating real and imaginary parts

$$\left.\begin{aligned} \nu_1 &= \nu_1' + (\alpha^2 + \beta^2)\,\nu_3 \\ \nu_2 &= \nu_2' + 2\alpha\,\nu_3 \end{aligned}\right\} \quad \ldots \ldots \ldots \quad (16)$$

Thus a suitable linear increase of ν_1 and ν_2 with ν_3 leaves the fundamental normal mode unchanged. Also if, for given basic values ν_1', ν_2', the control constants ν_1 and ν_2 for $\nu_3 > 0$ are given by (16) we have

$$(\nu_1 + \nu_2\gamma + \nu_3\gamma^2) - (\nu_1' + \nu_2'\gamma) = \nu_3\,[\beta^2 + (\alpha + \gamma)^2],$$

which is positive for all real γ, so the parabola in fig. 3 lies always above the straight line in fig. 1 for the same basic values; hence the lowest value of α_0 for the simple exponential (if any) with $\nu_3 > 0$ is necessarily greater than that for $\nu_3 = 0$ and the same basic values.

Thus, considering always the same pair of basic values, the introduction of $\nu_3 > 0$ does not affect the fundamental, and either eliminates the simple exponential or increases its damping, so that, provided the damping of the harmonics remains satisfactory, the stability of the control is not impaired and its rapidity of response is probably increased, as explained in § 2.

To examine the behaviour of the harmonics, write (15)

$$e^{-\alpha+i\beta} = -\,[\nu_3 + (\nu_2/\gamma) + (\nu_1/\gamma^2)]. \quad \ldots \ldots \ldots \quad (17)$$

For the high harmonics $|\gamma|$ is large, so that when $\nu_3 \neq 0$,

$$\alpha \to \log\,(1/\nu_3),$$

instead of increasing indefinitely with the order of the harmonic, as occurs when $\nu_3 = 0$, when $\alpha \sim \log(\beta/\nu_2)$. Further, it is clear that ν_3 must be less than 1, and that if it is too nearly 1, the high harmonics will only be slightly damped. Hence, apart from any question of mechanical design, there is a limit to the values of ν_3 for which satisfactory control is possible.

Further, taking the modulus of both sides of (17)

$$\alpha = -\log|\nu_3 + (\nu_2/\gamma) + (\nu_1/\gamma^2)|, \quad \ldots \ldots \ldots \quad (18)$$

and this is one relation which must be satisfied by the roots $-\alpha \pm i\beta$ of (15). For this relation, $\alpha = 0$ when

$$(1 - \nu_3{}^2)\,\beta^4 + (2\nu_1\nu_3 - \nu_2{}^2)\,\beta^2 - \nu_1{}^2 = 0\,;$$

when $\nu_3 < 1$, this has one and only one real positive root, say β_0, and also for large β, α given by (18) is certainly positive. Hence α must be positive for all roots of (15) for which $\beta > \beta_0$; and so, if the fundamental is positively damped, the harmonics must be also; and a positively damped fundamental is ensured by taking ν_1, ν_2 related to ν_3 by (16) with basic values known from case (i) to give a positively damped fundamental.

Case (iii) $\mu \neq 0$—When $\mu \neq 0$, so that the behaviour of the system is described by four independent parameters, μ, ν_1, ν_2, ν_3, the field to be covered is still more extensive, but the relations (16) between the values of ν_1, ν_2, ν_3, when $\nu_3 \neq 0$ and the "basic values" $\nu_1{}'$, $\nu_2{}'$ giving the same fundamental when $\nu_3 = 0$, still apply.

The equation for the complex frequencies γ is now equation (12) and the real roots $\gamma = -\alpha_0$, if any, are given by the intersections of the curve of $\alpha\,(\mu - \alpha)\,e^{-\alpha}$

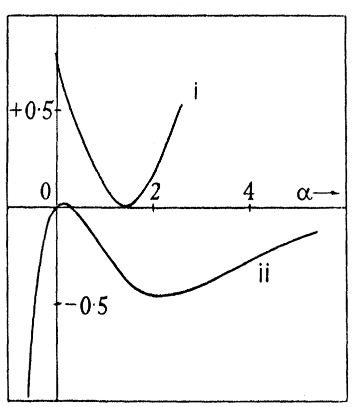


FIG. 4.—Curve i, $\nu_1 - \nu_2\alpha + \nu_3\alpha^2$; curve ii, $\alpha\,(\mu - \alpha)\,e^{-\alpha}$, $\mu = 0{\cdot}3$.

with the parabola $\nu_1 - \nu_2\alpha + \nu_3\alpha^2$ (*see* fig. 4); as before, if $\nu_3 = 0$ and ν_1, ν_2 are positive there is always an intersection for $\alpha > 0$, and there may be an intersection if $\nu_3 \neq 0$, and, if this exists at all, the lowest value of α_0 for it is greater than for $\nu_3 = 0$ with the same basic values of ν_1', ν_2'.

If $\nu_3 = 0$ and $\nu_1 = \mu\nu_2$, then (12) factorizes and gives

$$\gamma + \mu = 0, \quad \text{or} \quad \gamma e^{-\gamma} + \nu_2 = 0.$$

The first root gives the normal mode $\theta(\tau) = Qe^{-\mu\tau}$, which is the same behaviour as θ would have if it were disturbed from its zero value and left to return to it without further disturbance or operation of the control* ; thus, if θ happens to be varying in this way the control does nothing to alter the situation. If the value of μ is such that this condition is satisfied with values of the control constants which also give reasonable fundamental values for α and β, then these should be chosen because, in the long run, the total movement of the controlling gear would thereby be kept down. It should be noted that the inclusion of a small ν_3 term does not immediately destroy the possibility of this situation. A large ν_3 term will do so, but this particular loss is then off-set by the better control thereby obtained.

For investigating the available field for $\mu \neq 0$, $\nu_3 = 0$, consider the general form to be expected of a diagram similar to fig. 2, giving ν_1 and ν_2 contours in the (α, β) plane. The $\nu_1 = 0$ contour is given by

$$\beta \cos \beta = (\alpha - \mu) \sin \beta,$$

so that as $\beta \to 0$, $\alpha \to 1 + \mu$ and when $\alpha = 0$, $\tan \beta = -\beta/\mu$.

Hence the total useful field available when $\mu \neq 0$, which is certainly bounded by $\beta = 0$, $\alpha = 0$, and $\nu_1 = 0$, is somewhat larger than when $\mu = 0$; it will later be found important that the same order of values for α and β for the fundamental will now be obtained by using larger values for ν_1, or, alternatively, that larger values of α for the fundamental can now be obtained without demanding values of ν_1 dangerously small from the point of view of the damping of the simple exponential.†

4—Use of Heaviside's Operational Method

We shall now consider the particular kind of solution for which $\theta(\tau) = 0$ when $\tau < 0$, and, as a result of some disturbance in the interval $0 < \tau < 1$, $\theta(\tau)$ varies from zero for $\tau > 0$. As we have seen, the effect of any disturbance can be found by superposition of such special solutions.

* This type of behaviour has been termed "working to a curve of return" by Hodgson and Robinson, 'Proc. Inst. Mech. Eng.,' vol. 126, p. 66 (1934).

† [*Note added in proof*, *April* 22, 1936.—When $\mu > 0$, control which, although somewhat sluggish, is nevertheless stable can be obtained with $\nu_2 = \nu_3 = 0$ and only $\nu_1 \neq 0$; whereas when $\mu = 0$, it is evident from fig. 2 that for $\nu_2 = 0$, the value of α for the fundamental is negative, which shows that when $\mu = 0$, "hunting" must occur if only the term in ν_1 in the law of control is used.]

We shall first consider the operational representation of some convenient types of disturbing function. If $p \equiv \frac{d}{d\tau}$, and p^{-1} is interpreted in HEAVISIDE's sense as $\int_{\tau=0} \ldots d\tau$, and $\mathrm{H}(\tau)$ is HEAVISIDE's unit function

$$\mathrm{H}(\tau) = 0 \quad \tau < 0$$
$$\mathrm{H}(\tau) = 1 \quad \tau > 0,$$

then if

$$\psi(\tau) = \phi(p)\,\mathrm{H}(\tau), \qquad (19)$$

$\phi(p)$ is called the " operational representation " of $\psi(\tau)$.

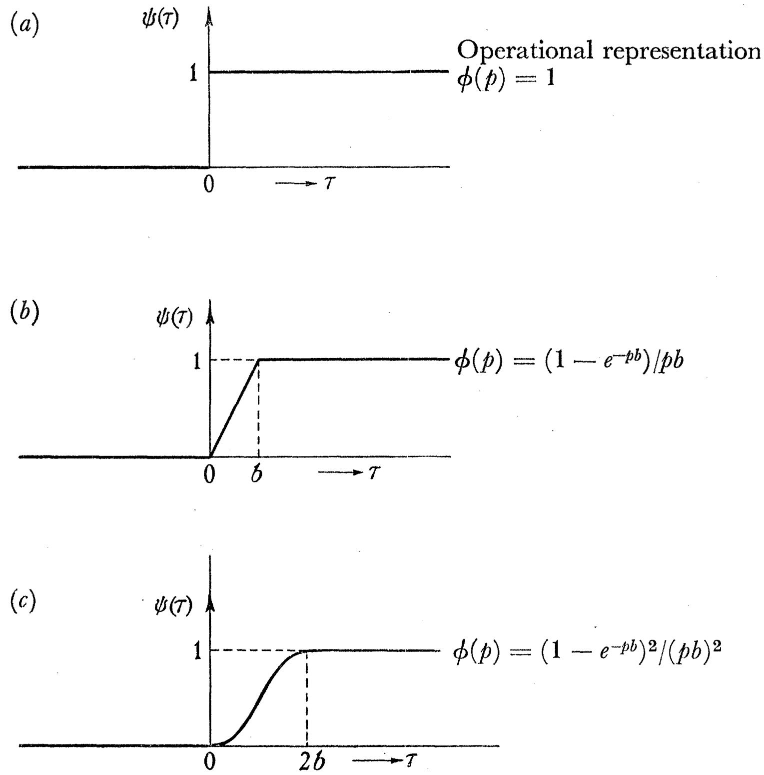


FIG. 5.

Three convenient disturbing functions and their operational representations are shown in fig. 5. The disturbance shown in (*a*) is the most drastic, and is just $\psi(\tau) = \mathrm{H}(\tau)$, so that its representation is just $\phi(p) = 1$. For some purposes, such a disturbance is too sudden, and we need a function $\psi(\tau)$ which is continuous. The simplest is that shown in (*b*), which is

$$\begin{aligned} \psi(\tau) &= 0 & \tau &< 0 \\ &= \tau/b, & 0 < \tau &< b \\ &= 1 & \tau &> b\,; \end{aligned}$$

it can easily be verified that its representation is

$$\phi(p) = (pb)^{-1}(1 - e^{-pb}).$$

A still less sudden disturbance, for which $\psi(\tau)$ and $d\psi(\tau)/d\tau$ are continuous, is given by

$$\begin{aligned} \psi(\tau) &= 0 && \tau < 0 \\ &= \tau^2/2b^2 && 0 < \tau < b \\ &= 1 - (\tau - 2b)^2/2b^2 && b < \tau < 2b \\ &= 1 && \tau > 2b \end{aligned}$$

shown in (*c*); its representation can be verified to be

$$\phi(p) = (pb)^{-2}(1 - e^{-pb})^2.$$

We have also

$$f(\tau - 1) = e^{-p} f(\tau)$$

and e^{-p} commutes with p^{-1},* so if $\phi(p)$ is the operational representation of the disturbing function, the operational form of (6), (7) is

$$p\theta = \phi(p) + c - \mu\theta$$

$$pc = -e^{-p}(\nu_1 + \nu_2 p + \nu_3 p^2)\,\theta$$

and that of their solution is

$$\theta = \frac{p\,\phi(p)}{F(p)} \qquad \text{. . . . . (20)}$$

where (*cf.* (12))

$$F(p) \equiv p(p + \mu) + e^{-p}(\nu_1 + \nu_2 p + \nu_3 p^2).$$

This operational result can be interpreted in two ways, either by the "partial fraction rule"† or by expansion in negative exponentials. The situation is very similar to that arising in the application of HEAVISIDE's operators to problems of wave motion on strings of limited length,‡ in which the interpretation by the partial fraction rule gives the solution, from given initial conditions, as an expansion in normal modes, and the expansion in negative exponentials gives a solution describing the successive arrival of waves reflected from the two ends, at intervals corresponding to the time-lag in the present case.

The interpretation of (20) by the partial fraction rule gives

$$\theta = \Sigma_\gamma \frac{\phi(\gamma)}{F'(\gamma)}\, e^{\gamma\tau} \qquad \text{. . . . . (21)}$$

* *See* JEFFREYS, *op. cit.*, p. 18.

† *See* JEFFREYS, *op. cit.*, p. 11.

‡ *Cf.* in particular the example considered in JEFFREYS's book, *op. cit.*, § 4.4.

where the summation is over the roots of $F(\gamma) = 0$. Now $F(\gamma) = 0$ is just equation (12) for the values of γ for the normal modes, so that in (21) we have θ expressed as a superposition of such normal modes, with amplitudes and frequencies now determined in terms of the disturbance, specified by its operational representation $\phi(p)$.

It is convenient to express θ in real form for practical use. The complex roots of $F(\gamma) = 0$ occur in conjugate pairs $\gamma = -\alpha \pm i\beta$, and since ϕ and F' are sums of products of exponentials and polynomials, it follows that $\phi(-\alpha + i\beta)$ and $\phi(-\alpha - i\beta)$ are conjugate, and similarly for F'. So, taking each pair of terms $\gamma = -\alpha \pm i\beta$ together, we have altogether

$$\theta = \sum_{\gamma\,\mathrm{real}} \frac{\phi(\gamma)}{F'(\gamma)} e^{\gamma\tau} + \sum_{\substack{\gamma\,\mathrm{complex} \\ \beta > 0}} 2 \frac{|\phi(\gamma)|}{|F'(\gamma)|} e^{-\alpha\tau} \cos\left[\beta\tau + \arg\left(\frac{\phi(\gamma)}{F'(\gamma)}\right)\right]. \quad . \quad (22)$$

Now for a stable system, $\alpha \geq 0$ for every normal mode, and we have seen that in general $\alpha \neq 0$ for each complex root of (12), and also that, provided $\nu_1 \neq 0$, $\gamma = 0$ is not a root, so is not included in the sum over γ in (21) or (22). Hence the effect of the disturbance $D(\tau) = H(\tau)$ [(*a*), fig. 5; $\phi(p) = 1$] on a stable system with $\nu_1 \neq 0$ is in general* to give $\theta \to 0$ as $\tau \to \infty$. Whereas if $\nu_1 = 0$, $\gamma = 0$ is a root of (12), and $F'(0) = \mu + \nu_2$, hence the sum in (21) over the roots γ of (12) includes the constant term $\psi(0)/(\mu + \nu_2)$, so that the effect of the same disturbance on a stable system with $\nu_1 = 0$ is in general* to give $\theta \to 1/(\mu + \nu_2)$ as $\tau \to \infty$. These analytical results confirm the general descriptive statements made on p. 6 concerning the difference between the behaviour of stable systems with $\nu_1 \neq 0$ and $\nu_1 = 0$ under a constant disturbance, and show the importance of the ν_1 term in the law of control.

Taking the effect of the disturbance $D(\tau) = H(\tau)$ as standard, we see that, for any other disturbance represented by $\phi(p)$, the amplitudes of the normal modes are multiplied by $|\phi(\gamma)|$; for example, for the disturbances (*b*) and (*c*) of fig. 5, the amplitudes are multiplied by factors

$$|(1 - e^{-\gamma b})/\gamma b| \qquad \text{and} \qquad |(1 - e^{-\gamma b})/\gamma b|^2,$$

respectively.

As an example of the effects of different disturbances, Table II gives the amplitudes and phases, calculated from (22), of the first few normal modes of a system with $\nu_1 = 0\cdot3$, $\nu_2 = 1$, $\nu_3 = \mu = 0$ (for which the values of γ and $F'(\gamma)$ are given in Table I), under five different disturbances, namely, for that shown in fig. 5*a*, and for two values of b ($0\cdot1$ and $0\cdot5$) for each of those shown in fig. 5*b* and *c*. The amplitudes of the first and second harmonics are much the same for different disturbances, but the phase shifts are considerable, and are mainly responsible for the different behaviours of θ in the various cases; since the first harmonic is

* In the exceptional case in which $\alpha = 0$ for one (or more) of the complex roots of (12) and $\alpha > 0$ for the other complex roots, the result of the same disturbance is that θ ultimately oscillates about the value $\theta = 0$ if $\nu_1 \neq 0$, and about $\theta = 1/(\mu + \nu_2)$ if $\nu_1 = 0$.

Table II—Amplitudes and Phases of Normal Modes Excited by Different Disturbances for System with $\nu_1 = 0\cdot3$, $\nu_2 = 1$, $\nu_3 = \mu = 0$.

Disturbing function $\psi(\tau)$	H (τ) Fig. 5 (a)		$(pb)^{-1}(1-e^{-pb})$ H (τ) Fig. 5 (b), $b = 0\cdot1$		$(pb)^{-1}(1-e^{-pb})$ H (τ) Fig. 5 (b), $b = 0\cdot5$		$(pb)^{-2}(1-e^{-pb})^2$ H (τ) Fig. 5 (c), $b = 0\cdot1$		$(pb)^{-2}(1-e^{-pb})^2$ H (τ) Fig. 5 (c), $b = 0\cdot5$	
	Amplitude	Phase at $\tau = 0$ (radians)	Amplitude	Phase at $\tau = 0$ (radians)	Amplitude	Phase at $\tau = 0$ (radians)	Amplitude	Phase at $\tau = 0$ (radians)	Amplitude	Phase at $\tau = 0$ (radians)
Simple exponential .	1·168	—	1·192	—	1·297	—	1·217	—	1·441	—
1st harmonic (fundamental)	1·348	$\pi+0\cdot6061$	1·358	$\pi+0\cdot547$	1·388	$\pi+0\cdot309_5$	1·370	$\pi+0\cdot488$	1·430	$\pi+0\cdot013$
2nd harmonic . . .	0·034	$\pi-0\cdot4095$	0·036	$\pi-0\cdot800$	0·032	0·420	0·040	$\pi-1\cdot191$	0·030	$\pi+1\cdot249$
3rd harmonic . . .	0·010	$\pi-0\cdot3073$	0·010	$\pi-1\cdot035_5$	0·004	$\pi+0\cdot187$	0·010	1·378	0·002	$\pi+0\cdot681$
4th harmonic . . .	0·005	$\pi-0\cdot2482$	0·005	$\pi-1\cdot316$	$0\cdot002_5$	0·575	0·005	0·758	0·001	$\pi+1\cdot403$
5th harmonic . . .	0·003	$\pi-0\cdot2076$	0·003	$\pi-1\cdot619$	$0\cdot000_5$	$\pi-0\cdot214$	0·002	0·112	0·000	—

very much the least damped normal mode in this case, it follows that after a comparatively short time the effects of different disturbances, of the same final magnitude, will be mainly phase changes of the subsequent variation of θ.

We have so far been concerned with the interpretation of the operational result (20) in terms of the expansion of θ in normal modes by the partial fraction rule. The other interpretation of (20), by expansion in negative exponentials, gives results very similar to those for wave problems, but is too complicated to be of use in this case if the variation of θ after the first few units of τ is required. It gives directly the results which could also be obtained by an iterative process, by integrating (6), (7) through successive periods of length unity (in τ).

5—Use of Numerical and Mechanical Methods of Integration

A general idea of the range of values of the control constants which give satisfactory control can be obtained by the method of § 3, and further investigation requires actual solutions of the equations, with different disturbing functions $\psi(\tau)$, to show the behaviour of the control under different circumstances. This can be done by the methods of § 4, but the evaluation of any large number of solutions by those methods would be very tedious.

Some solutions have been obtained by numerical and graphical methods of integration, and a fairly extensive set of rough solutions has been obtained by the use of a model of the machine designed and built by Bush* for the solution of differential equations, known as the differential analyser. The model, of which an account has been published elsewhere,† was constructed in the Physical Laboratories of the University of Manchester, and, despite the roughness of its mechanical construction as compared to Bush's machine, it has given valuable results in this work, and this work has given valuable experience in handling such a machine, previous to the erection of the full-size differential analyser lately installed there.‡ This machine, and the model, incorporate a development suggested by the present problem and likely to be of use in others, namely, a special input table for handling just such equations as (1) and (2), in which $d\theta/dt$ at time t depends on the values of $\theta(t)$ and other quantities, not only at time t but at time $t - \mathrm{T}$. In the present problem, a pencil draws, at the abscissa $(t + \mathrm{T})$, a curve whose ordinate is

$$\text{constant} - [n_1 \int \theta(t)\, dt + n_2\, \theta(t) + n_3\, \dot{\theta}(t)]$$

and a pointer at the abscissa t is made to follow the curve so drawn. For $0 < t < \mathrm{T}$ the pointer is made to follow $\mathrm{D}(t)$ which will have previously been plotted, as shown dotted in fig. 6, which shows the set-up of the machine for equations (1) and (2) in

* 'J. Franklin Inst.,' vol. 212, p. 447 (1931).

† Hartree and Porter, 'Mem. Proc. Manch. Lit. Phil. Soc.,' vol. 79, p. 51 (1935).

‡ Hartree, 'Nature,' vol. 135, p. 940 (1935).

their general form ($\nu_3 \neq 0$, $\mu \neq 0$) and follows Bush's standard diagrammatic representation of the various units of the machine, with an obvious modification for the special input table.

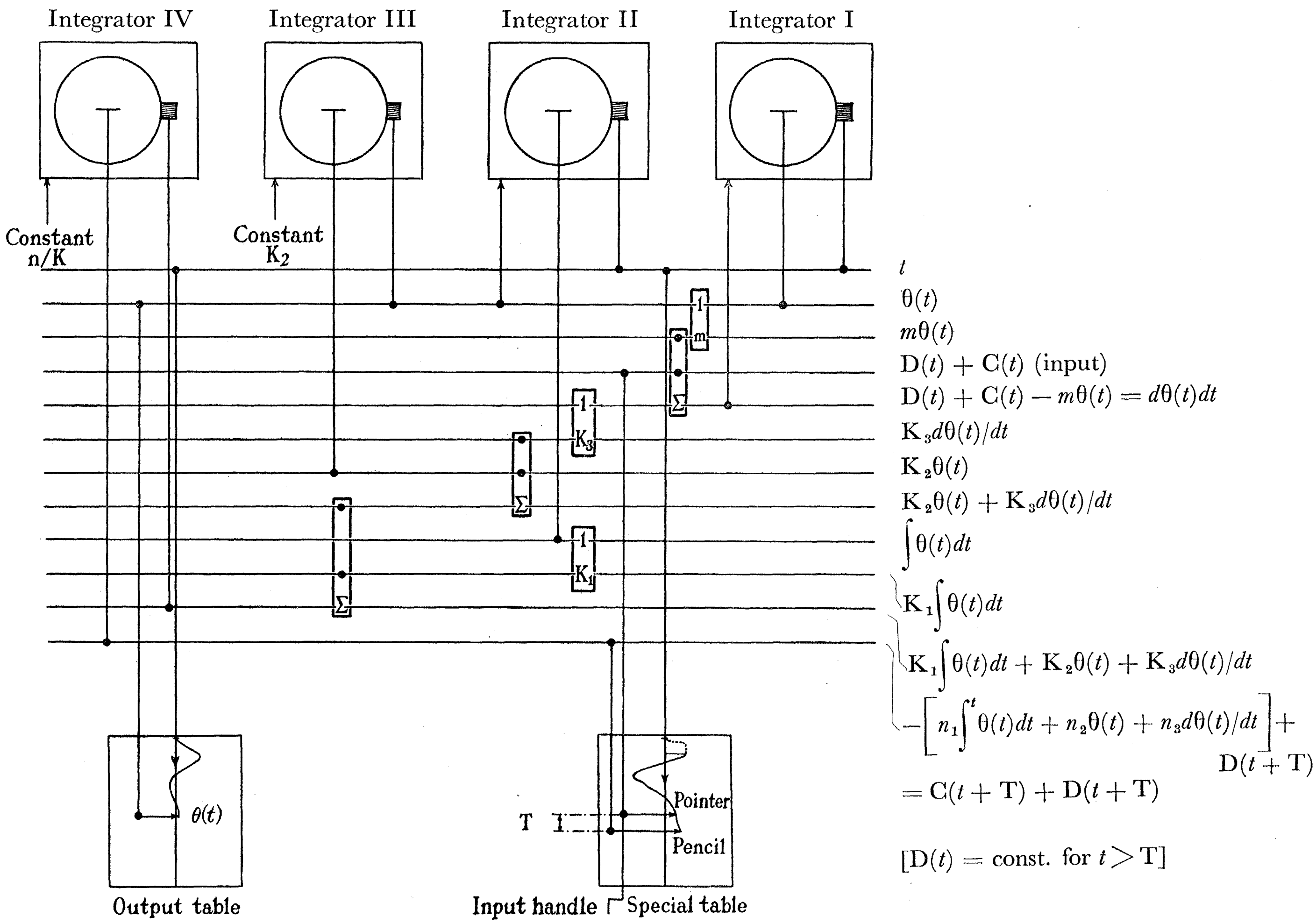


Fig. 6—Schematic set-up of Model Differential Analyser for

$$\begin{cases} d\theta(t)/dt = D(t) + C(t) - m\,\theta(t) \\ -C(t+T) = n_1 \int^t \theta(t)\,dt + n_2\theta(t) + n_3 d\theta(t)/dt \end{cases}$$

The values of K_1, K_2, and K_3 are chosen so that $n_1/K_1 = n_2/K_2 = n_3/K_3$, and so that K_1 and K_3 can be obtained by combination of available gear wheels; the multiplication by the ratio n/K is carried out on Integrator IV. The 't' shaft is driven by the independent-variable motor.

6—Discussion of Results

The merits of any selected control constants will be judged by the behaviour of θ subsequent to any disturbance. The behaviour for the duration of one time-lag after the initiation of the disturbance is inevitable, but after this a quick return

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

towards steadiness at $\theta = 0$ will always be desired. The decision as to whether this should occur with as little over-shooting as possible, or whether θ should preferably be caused to decrease rather more quickly, but with repeated slight over-shooting to the opposite side before finally returning to zero, can be made only with reference to the particular control problem envisaged. Quite apart from the quicker initial decrease, there certainly will be cases in which it will be an advantage if an initial deviation of θ is followed by a temporary deviation of the opposite sign which might partly correct any undesired effects of the original deviation.

In practical control mechanisms, the time-lag and the values of the control parameters n_1, n_2, n_3 may themselves be subject to variations in the course of time, and it is important that such variation shall not impair the behaviour of the controlling gear. We shall call the control " flexible " when its behaviour is satisfactory over an appreciable range of values of n_1, n_2, n_3, and T. Using the differential analyser, it is easy to study the effect of such changes, and so to examine the flexibility of the system.

In discussing the results it is convenient to give diagrams showing the behaviour of the solution in some typical examples ; a large number of such solutions have been obtained by the use of the differential analyser or otherwise, and only a small selection will be given here to illustrate the main points which arise.

For the case of temperature control $\theta(\tau)$ is the temperature deviation and $\psi(\tau) + c(\tau)$ on the right-hand side of (6) is the rate of heat supply. In the diagrams we shall usually show curves both of $\theta(\tau)$ and of $\psi(\tau) + c(\tau)$; the beginning of the latter curve shows the form of the disturbance considered, which is always one of those shown in fig. 5, reaching a maximum value 1. In order that the curves should be easily comparable, to show the effects of the different types of control, the scales are taken the same for all.

Case (*i*). $\mu = 0$, $\nu_3 = 0$, $\nu_1 \neq 0$, $\nu_2 \neq 0$—The total field available for the fundamental is the region of fig. 2 bounded by $\beta_1 = 0$, $\alpha_1 = 0$, and $\nu_1 = 0$ (α_1, β_1 being the values of α, β for the fundamental). The desirability of a quick return of θ to the neighbourhood of $\theta = 0$ demands a high value of β, and this should be coupled with adequate damping indices both for the fundamental itself and for the simple exponential. (As already stated, the two latter aims are of limited compatibility.)

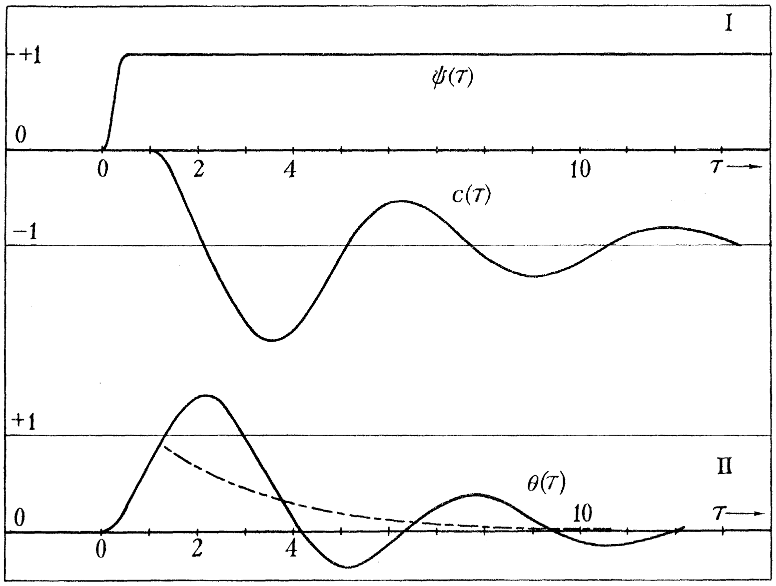
Such considerations as these show that the workable field covers a region surrounding the line joining the points, Table III.

Table III

ν_1	ν_2	α_1	β_1	
0·156	0·568	0·40	0·60	(*a*)
0·29	0·95	0·20	1·13	(*b*)

Any final selection of ranges of values for obtaining good types of return to steadiness must be based on examples showing the actual behaviour of θ in various cases.

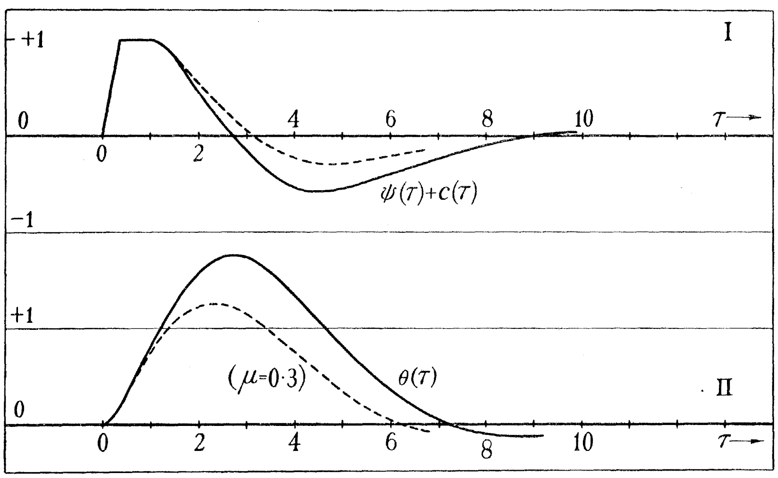
Figs. 7, I and II, show results for the above case (*b*) $\nu_1 = 0\cdot29$, $\nu_2 = 0\cdot95$. The example shows a typical choice of disturbing function; that used attains its final unit value in the first half time-lag. It is found that small changes in the actual



FIGS. 7—I. Curves of $\psi(\tau)$ and $c(\tau)$; II. curve of $\theta(\tau)$. Example (*b*) of Table III.

Full curves $\mu = \nu_3 = 0$, and $\begin{cases} \nu_1 = 0\cdot29 & \alpha_1 = 0\cdot20 \\ \nu_2 = 0\cdot95 & \beta_1 = 1\cdot13 \end{cases}$

Broken curve, contribution to θ from simple exponential.



FIGS. 8—Example (*a*) of Table III.

Full curves $\mu = \nu_3 = 0$, and $\begin{cases} \nu_1 = 0\cdot156 & \alpha_1 = 0\cdot40 \\ \nu_2 = 0\cdot568 & \beta_1 = 0\cdot60 \end{cases}$

Broken curves, $\mu = 0\cdot3$, $\nu_3 = 0$, same α_1 and β_1.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

form used have no appreciable effect on the behaviour of $\theta(\tau)$; this agrees with the results in Table II for which the values of ν_1 and ν_2 are very similar. The way in which $c(\tau)$ ultimately will attain the value -1, thus producing $c(\tau)+\psi(\tau)=0$, is clear; in future examples this sum alone will be plotted in order to save space.

In fig. 7, β_1 is high and the initial return of $\theta(\tau)$ is therefore good, but the damping is low. Moreover, the high ν_2 combined with the low ν_1 gives a low damping of the simple exponential as shown approximately by the broken line.

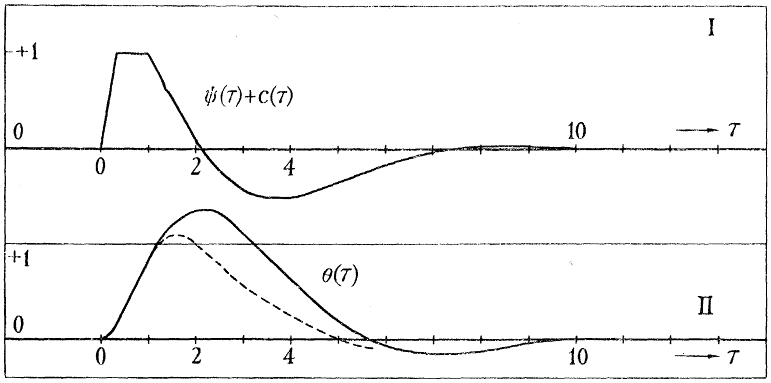
Fig. 8 shows corresponding results for the other extreme (case (*a*) of Table III). Here $\psi(\tau)$ is taken as linear in τ as far as $\tau=\frac{1}{3}$. $\theta(\tau)$ is almost dead-beat, but the maximum value of θ is high because β_1 is low. In itself this control is unnecessarily sluggish, but it is found to have the practical advantage of flexibility, which remains when these values of ν_1, ν_2 are taken as the basic values for a control system with $\nu_3 \neq 0$. Two-term control, that is to say control based on equation (2) in which only ν_1 and ν_2 are non-zero, is adequate for many purposes; apart from any possible need for flexibility, one would choose constants giving a result probably somewhat nearer fig. 7 than fig. 8, unless a dead-beat tendency for $\theta(\tau)$ were specially desired. Figs. 7 (II) and 8 (II) well illustrate limits between which $\theta(\tau)$ may be made to behave in any intermediate manner.

Case (*ii*). $\mu=0$, $\nu_3\neq 0$, $\nu_2\neq 0$, $\nu_1\neq 0$—We have seen already that for stable control $\nu_3<1$.

An effect of ν_3, actually found by examination of curves, will now be explained in order to obviate the need for a large number of figures. Examination of two-term control shows that high damping for the fundamental can be obtained only in conjunction with somewhat sluggish return of $\theta(\tau)$, *i.e.*, a high α_1 cannot be accompanied by a high β_1 for, as seen in fig. 2, this requirement would demand a low value of ν_1 and it would then be found that the damping of the simple exponential was too small. The introduction of ν_3 is very useful in that it enables the basic constants ν_1' and ν_2' to be chosen as if for a reasonably high α_1, and, therefore, a low β_1, but without, at the same time, causing sluggish initial return of $\theta(\tau)$.

Taking, then, the values of ν_1, ν_2 for the case of fig. 8, which tends this way, as basic values, fig. 9 shows the effect of two values of ν_3. Full and broken lines show respectively $\nu_3=0\cdot227$ and $\nu_3=0\cdot5$. The advantage of even small values of ν_3 is clearly seen by comparing the three cases. By bringing the controlling function into operation early the maximum error of $\theta(\tau)$ is reduced, and moreover in the example with $\nu_3=0\cdot227$, the maximum value which has to be attained by $c(\tau)$ is actually slightly less than for $\nu_3=0$. That both these results are already obtained when using this fairly small value of ν_3 is sometimes important, because although larger values would give better control they might lead to excessive operation of the controlling gear, which would not only cause wear but might also disturb other apparatus or plant supplied from the same source of energy (for example, the pressure of a common steam-main may itself be unduly disturbed in this way).

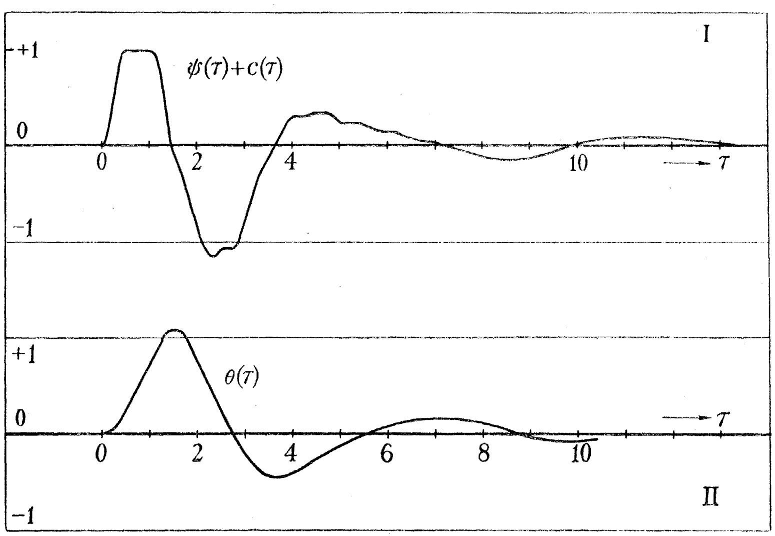
Further investigation, as illustrated by fig. 10, shows that in examples where flexibility is not important, basic values of ν_1' and ν_2' somewhat greater than those of Table III (*a*), may advantageously be used, even when $\nu_3 \neq 0$. Flexibility will, however, be less good. It may be mentioned that in the case of fig. 9, the control



FIGS. 9—Full curves $\mu = 0$ and $\begin{cases} \nu_1 = 0\cdot274 \\ \nu_2 = 0\cdot749 \\ \nu_3 = 0\cdot227 \end{cases}$ $\nu_1' = 0\cdot156$; $\alpha_1 = 0\cdot40$ $\nu_2' = 0\cdot568$; $\beta_1 = 0\cdot60$

Broken curve, $\mu = 0$, $\nu_3 = 0\cdot5$, same ν_1', ν_2'.

constants may be modified so as to correspond to a simultaneous increase of 20% both in time-lag and in controlling gear sensitivity, and the control will still be excellent. Fig. 10 completes the diagrams for case (ii) with a selected example of

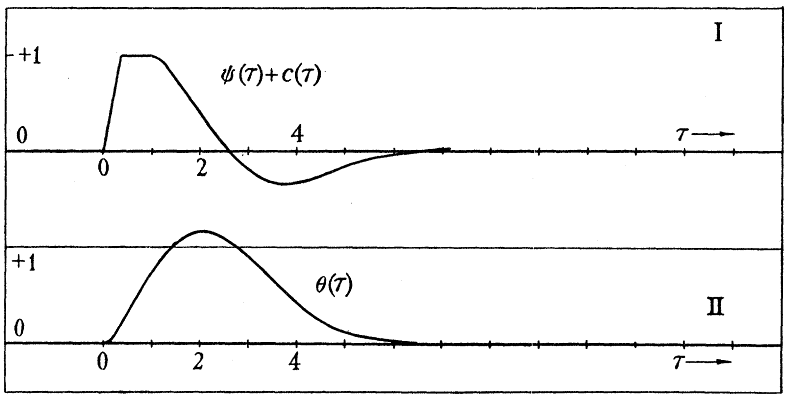


FIGS. 10—$\mu = 0$ and $\begin{cases} \nu_1 = 0\cdot95 \\ \nu_2 = 1\cdot24 \\ \nu_3 = 0\cdot67 \end{cases}$ $\nu_1' = 0\cdot20$; $\alpha_1 = 0\cdot34$ $\nu_2' = 0\cdot79$; $\beta_1 = 1\cdot00$

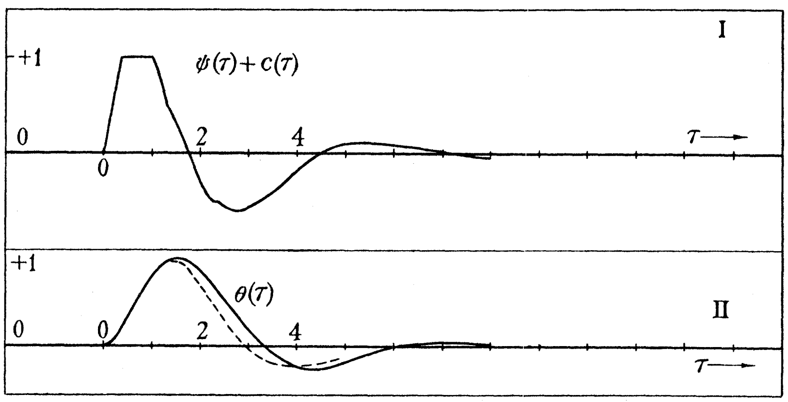
good control; ν_3 is fairly large. In this case the control function builds up more rapidly and further than in previous examples, consequently the rapidity of initial return to normality is higher. There is some "over-shooting"; in fact, comparison of figs. 9 (II) (broken) and 10 (II) well illustrates the choice to be decided upon as mentioned at the beginning of this section. Neither of these cases, however, will be so flexible as that for which the results are shown by the full curves in fig. 9.

In examples where $\nu_3 \neq 0$, it may be noticed that any discontinuity in $\psi(\tau)$ is repeated (to a smaller degree since $\nu_3 < 1$) at regular intervals of one time-lag after its first occurrence.

Case (iii). $\mu \neq 0$, ν_1, ν_2, $\nu_3 \neq 0$—No general examination of this case has been made; a few cases have been tried for $\mu = 0\cdot3$ merely because this value corresponded to a definite practical problem.



FIGS. 11—$\mu = 0\cdot3$ and $\begin{cases} \nu_1 = 0\cdot282 \\ \nu_2 = 0\cdot690 \\ \nu_3 = 0 \end{cases}$ $\alpha_1 = 0\cdot50$ $\beta_1 = 1\cdot00$



FIGS. 12—Full curves $\mu = 0\cdot3$ and $\begin{cases} \nu_1 = 0\cdot752 \\ \nu_2 = 1\cdot066 \\ \nu_3 = 0\cdot376 \end{cases}$ $\nu_1' = 0\cdot282$ $\nu_2' = 0\cdot690$ $\alpha_1 = 0\cdot50$ $\beta_1 = 1\cdot00$

Broken curve, $\nu_3 = 0\cdot5$, same μ, α_1, β_1

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

It appears that for $\mu = 0 \cdot 3$, the fundamental normal mode with $\beta_1 = 0 \cdot 60$, $\alpha_1 = 0 \cdot 40$ gives an unnecessarily sluggish control, even making due allowance for flexibility ; the broken curves in fig. 8 illustrate the point. Figs. 11 and 12 are on the same basis and show a case which uses freedom to combine higher ν_1 with higher α_1. The benefit of ν_3 remains, and the control is good ; further it can be shown that the control constants used in fig. 12 give good flexibility. (The broken curve is for $\nu_3 = 0 \cdot 5$ using same values for α_1 and β_1.)

7—Automatic Control

A general method by which control in accordance with equation (2) may easily be obtained automatically will now be indicated, and some practical examples of this method outlined.

In fig. 13 the controlling gear is represented for definiteness by a valve S governing the supply of steam to a vessel whose temperature deviation θ from standard is indicated by the position of the arm A, and is to be kept as small as possible ; but the argument could easily be extended to other methods of controlling and indicating temperature or other physical quantities.

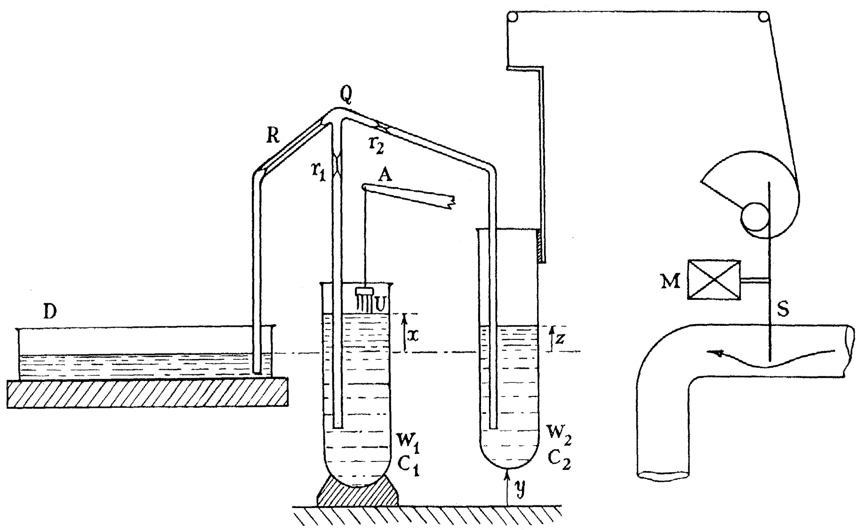


Fig. 13.

If there is a time-lag T between the valve S taking up any setting, and the effect of this setting on the behaviour of the indicated value* of θ, then we may write

$$C(t + T) = S(t),$$

* If some of the time-lag is in the indicating instrument, $\theta(t)$ in (1), (2) must be taken as the *indicated* value of θ at time t.

where $S(t)$ depends on the setting of the valve S at time t, and $C(t)$ is the controlling function defined in § 2. Then (2) becomes

$$-\dot{S}(t) = n_1\,\theta(t) + n_2\,\dot{\theta}(t) + n_3\,\ddot{\theta}(t) \qquad (23)$$

and the valve S is to be moved automatically so that $S(t)$ satisfies this equation. There are several practical methods of achieving this; we shall outline one and indicate two analogous forms.

In fig. 13, let x be the height of the end of the arm A above its position when $\theta = 0$, and suppose that x is proportional to θ. Also let y be the height of a cylindrical vessel W_2 above some arbitrary level, this height being made proportional to the value of $-S(t)$, for example by means of a connexion to the valve S through a cam as shown. A similar vessel W_1 is placed so that the lower end U of a rod hanging from the arm A can touch the surface of liquid in W_1. Liquid in W_2 communicates with that in W_1 through a siphon pipe, from an intermediate point Q of which a branch pipe is led to a dish D of cross-section large compared to those of W_1 and W_2, so that changes of level of the liquid in D can be neglected. This level is adjusted to be that of U when $\theta = 0$.

Now suppose some mechanism M, actuated by the contact at U, causes S (and so W_2) to be moved in such a way that the liquid surface in W_1 is kept just in contact with U. Then if the cross-sections of W_1, W_2 are C_1, C_2 and the resistances* of the pipes from Q to D, W_1, W_2 are R, r_1, r_2, respectively, then it follows from the equations of motion and continuity of the liquid† that

$$\dot{y} = \frac{1}{C_2 R}\left[x + \{(R + r_1)\,C_1 + (R + r_2)\,C_2\}\,\dot{x} + \{R\,(r_1 + r_2) + r_1\,r_2\}\,C_1\,C_2\,\ddot{x}\right], \qquad (24)$$

which is of the form (23), since x, y are respectively proportional to $\theta(t)$ and $-S(t)$.

Thus if contact of the surface of the liquid in W_1 with U is maintained as θ varies, a law of control of the desired form will be obtained.

It is of some interest to note that though any desired values of n_1 and n_2 can be obtained by suitable choice of the constants of proportionality between x and θ, and between y and $-S(t)$, and of C_1, C_2, R, r_1, r_2, the possible values of n_3, for given n_1, n_2 are restricted, for, comparing (24) with (2), it is found that

$$4\,(n_3/n_1) = (n_2/n_1)^2 - \{(R + r_2)\,C_2 - (R + r_1)\,C_1\}^2 - 4R^2\,C_1\,C_2,$$

so that

$$n_3/n_1 \leq \tfrac{1}{4}\,(n_2/n_1)^2$$

and consequently, from (4)

$$\nu_3/\nu_1 \leq \tfrac{1}{4}\,(\nu_2/\nu_1)^2 \qquad (25)$$

* For the velocities of flow here concerned, the resistance of a pipe [(pressure difference)/(rate of flow)] is constant. In deriving (24), the pressure difference is supposed measured in terms of head of liquid.

† Inertia terms in the equations of motion have been neglected, as they are negligible in all practical cases yet encountered.

However, in theory, as shown in § 3, and in practice there are reasons why large values of ν_3 are undesirable, so that the restriction (25) has not been found a serious disadvantage.

There are several ways of ensuring that S (and so W_2) should move in such a way that the liquid surface in W_1 follows the motion of U, as closely as is required in order to obtain a good approximation to the desired law of control. One method, which has the advantage of imposing practically no forces on the indicating arm A, is to use mercury as the liquid in the vessels, and, in principle, to use an electrical contact* between U and the liquid surface to operate a motor (M, fig. 13) driving W_2 and S.

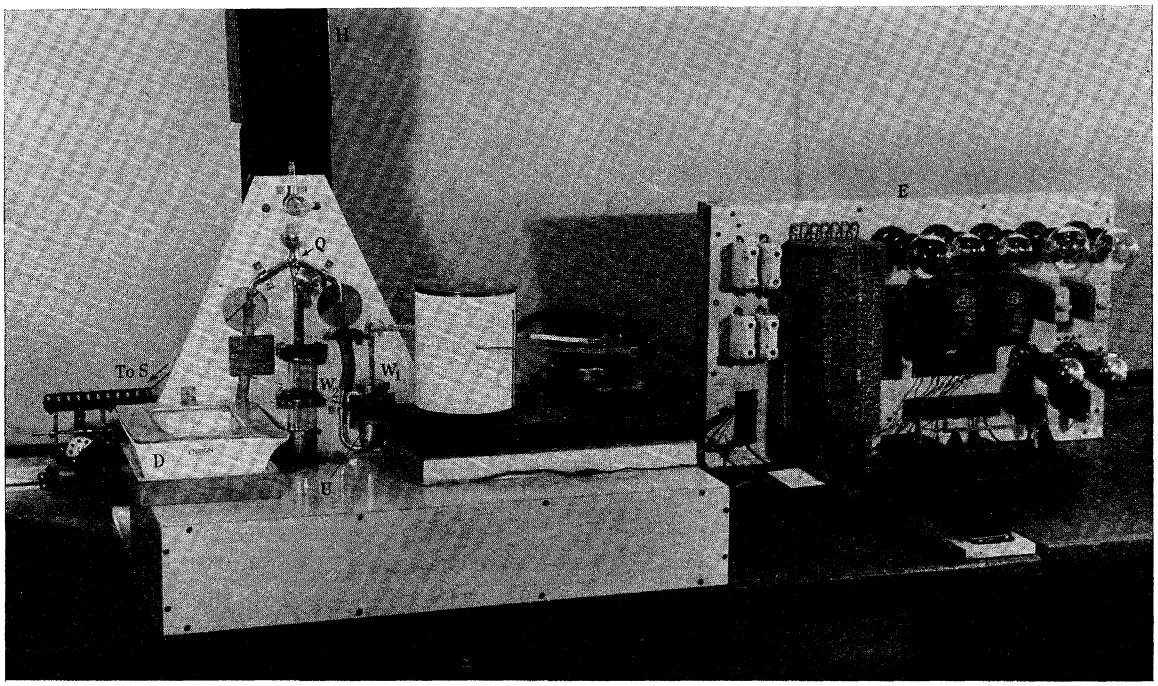


FIG. 14.

An apparatus of the type of fig. 13 was exhibited by I.C.I. (Alkali) Limited at the Royal Society Conversazione, May 3, 1935. In fig. 14 (photograph), D, W_2, and W_1 are the various vessels, E is an electrical relay system by which the contacts at U govern the operation of motor M (not visible in the photograph) so as to give the "following" motion to the liquid surface. For demonstration the steam valve had been replaced by a rheostat governing the electrical energy supply to H, which is a miniature body the temperature of which is being controlled. A "delay", brought about by suitable insertion of lagging material, and approximately constituting a $1\frac{1}{2}$-minute time-lag, is incorporated in the system. A non-electrical form of this

* To avoid sparking, and for other reasons, a group of four contacts operating the motor through a relay system is used ; it is beyond the scope of this paper to describe this arrangement in detail.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

type in which a float on the liquid in W_1 brings about the operation of a pilot valve arrangement has also been designed.

Two analogous forms of this apparatus will now be briefly mentioned.

The electrical circuit of fig. 15 is exactly analogous—the same mathematics applies, using the obvious interpretations. The setting of S determines, on a suitable scale, the potential of one side of condenser C_2, instead of the height of vessel W_2. S is operated so that the potential indicated by the arm B of the electrical instrument G follows the movements of A.

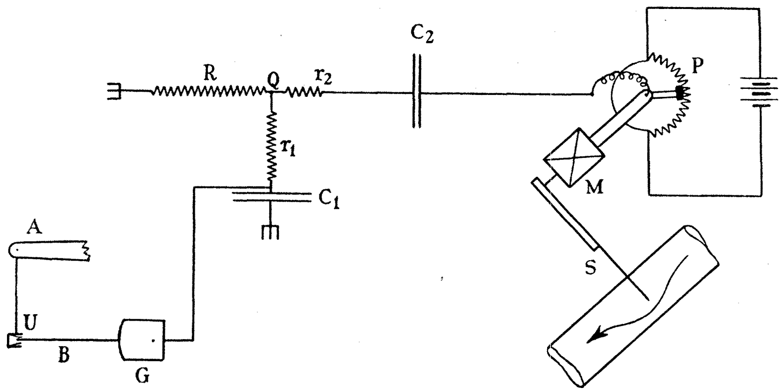


Fig. 15.

In another form, liquid pressures are used (instead of actual heights as in fig. 13), a pressure indicator taking the place of the electrical instrument G in fig. 15, spring loaded pistons or diaphragms being the analogues of the condensers of fig. 15.

This analogue can easily be given still further variations in which, by duplicating some of the vessels and arranging them symmetrically, the equivalent of D can be eliminated.

Practical examples of each of these systems have been used, and give satisfactory operation.

One of the authors (A. C.) wishes to thank the Directors of Imperial Chemical Industries Limited for permission to publish this work, developed, in part, in the Research Department of one of the subsidiary companies, I.C.I. (Alkali) Limited.

Summary

Control gear of some kind is often used to keep the value of a physical quantity, subject to random disturbances, as nearly constant as possible. This paper is concerned, firstly, with a general theoretical study of the operation of such control gear when this operation is determined solely by the behaviour of the quantity to be controlled itself, and when there is a time-lag between this behaviour and the effect

of the consequent control operation ; and secondly, with means of putting the theoretical conclusions into practice.

The "law of control", or relation between the behaviour of the quantity to be controlled and the effect (after the expiration of the time-lag) of the consequent control operation on it, is taken to be linear in the quantity controlled, its time derivative and time integral.

In the absence of a disturbance, the control equations have certain simple solutions, exponential or damped harmonic, which may be called "normal modes", and the behaviour of the controlled quantity when disturbed can be expressed as the sum of normal modes, with amplitudes and phases which are best determined by HEAVISIDE's operational method. For stable control the damping constants of all the normal modes must be positive, and this, coupled with the requirement of quick return of the quantity controlled to its normal value after a disturbance, limits the practically useful range of values of the parameters in the law of control. Suitable ranges of values for these parameters are found and typical examples of the behaviour of the quantity controlled, for different values of the parameters, are given.

A practical method, feasible on the industrial scale, for obtaining automatic control approximately in accordance with a law of control of the form here studied, is outlined, and two analogous methods are indicated.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

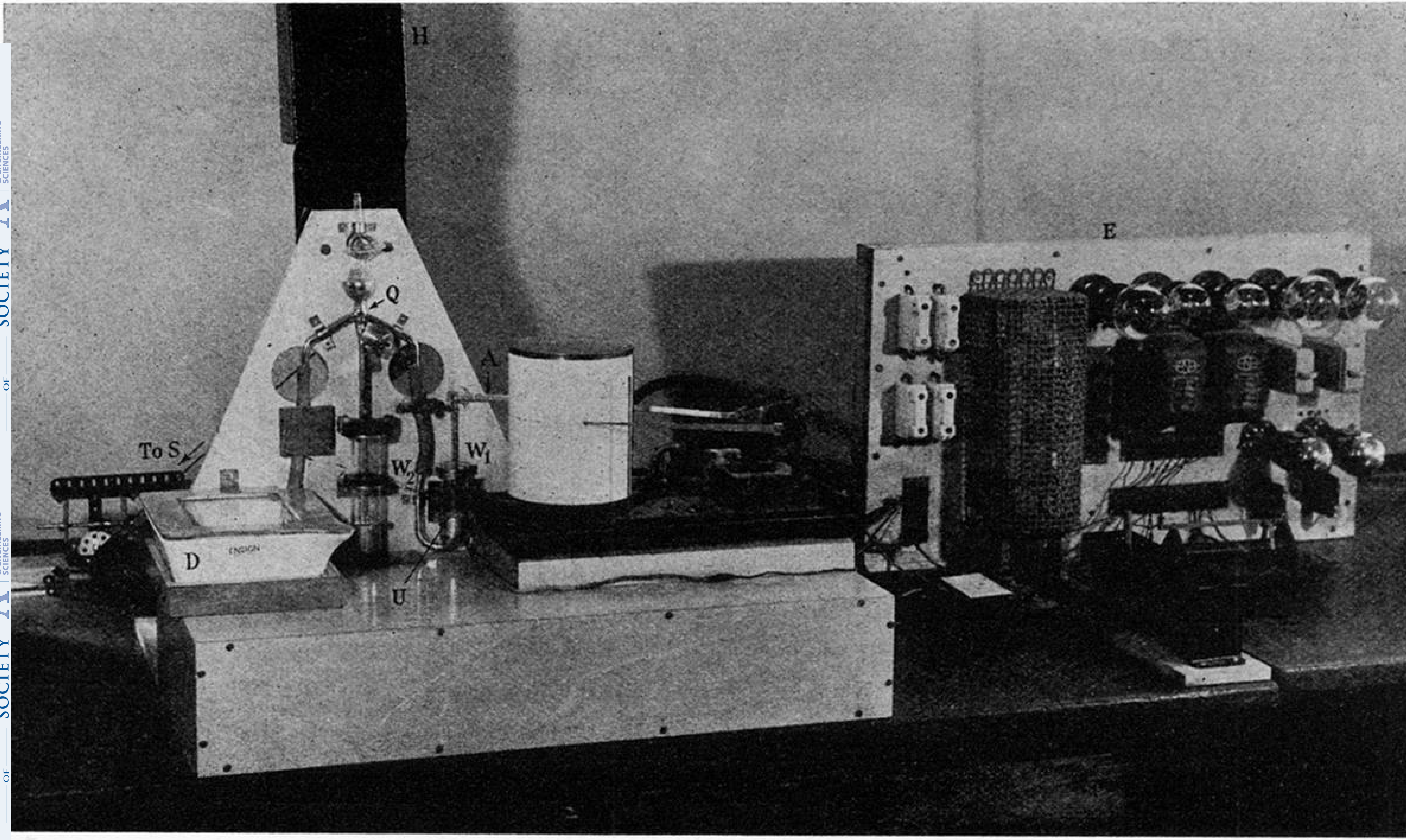


FIG. 14.