THE MEASUREMENT OF LINEAR SYSTEM IMPULSE RESPONSES BY MEANS OF CROSSCORRELATION WITH PSEUDORANDOM BINARY TEST PERTURBATIONS

2.1 INTRODUCTION

There has been considerable effort over the past two decades to improve the operation and control of existing industrial processes; and the design of new plant. This has led to an increasing interest in the measurement of system dynamics.

In the initial stages, the measurement of system transfer-functions was carried out by injecting sinusoidal signals at different frequencies, and measuring the amplitude and phase of the response. The method is very time consuming for two reasons : firstly, each frequency must be fed seperately, and secondly in the usual case where noise is present in the system, averages for each frequency must be considered over a long time.

Another conventional method of determining the system dynamics is by means of transient inputs. Indeed, time-domain methods have found increasing use: the system impulse response measurements with step or pulse input are quite familiar. But, a drawback results from the need to use a large perturbation signal which adversely effects the normal operation of the system. If the test signal is made smaller, the experimentation time must be so long that difficulty arises in holding the system steady over a sufficiently long period of time.

So much so, any method which can give good measurement of the dynamics in a reasonably short time in the presence of system disturbances, and which does not interfere with the normal operation of the system would be most attractive. For this reason, correlation methods are currently finding much favour.

It is well known that the dynamic response of a linear system can be obtained by using white noise as the input and crosscorrelating it with the output of the system. However, white noise (characterised by flat power spectrum of infinite bandwidth) is a mathematical fiction - it is awkward to generate a flat power spectrum at low frequencies and difficult to achieve repeatable results. Another drawback results from the need to perform crosscorrelation over a long period of time.

Hence, use of binary random noise as test signal in correlation analysis has been subsequently considered. (Cooper, 1960; Levin, 1960; Heath and Gribble, 1961; Hughes and Noton, 1962; Corran and Cummins, 1963) Poortvliet, 1963). But here too, long integration time was involved to ensure the autocorrelation function of the test signal (e.g., The random telegraph signal, the discrete interval binary noise etc.) to that of white noise. Furthermore, difficulties were experienced in processing and controlling the parameters of the perturbations. And the results have been found to be somewhat unsatisfactory.

More resently use has been made of certain periodic sequences, which possess properties in correlation to those of white noise but which are deterministic. These are known as m- or maximum length linear sequences or loosely pseudorandom sequences.

Over the past few years, attention has been confined to the use of 2-level m-sequences, which are sometimes called chain codes - (Cummins 1964; Korn 1964, Briggs et al 1964-65, Hazlerigg and Noton 1965, Barber and Hammond 1965, Chang 1965, Douce and Ng 1965, Godfrey 1965, 1966, 1967, 1968 and 1969, Gupta 1965, Barker 1967, 1968 and 1969, Davies WDT 1966, Douce and Walker 1966, Briggs et al 1967, Davies and Douce 1967, Wilson 1967, Macleod 1968 and 1969, Davies WDT 1968, Brown 1968 and 1969, Clarke and Briggs 1970, Anca Tomescu 1970, Nikiforuk et. al, 1970, and 1971, Hoffman et. al. 1972; (It may be noted that these references cover applications pertaining to only single input and single output linear systems. References with regard to use of prbs as test signals in the multivariable and nonlinear system identification are stated in the chapters four and five respectively where the subject is taken up for study).

Although the theory of the method is well established, the mathematical relationships involved are not simple to understand. Further, in the practical application of this chain code system testing, a number of errors arise due to the existence of (a) imperfect input transducer characteristics, (b) non-ideal test-input characteristics, (c) and wide-band and band-limited noise in the system. Either to eliminate or to minimize these errors in the measured values of system impulse response, several correlation schemes have been explored, developed, and refined. However, a quite satisfactory method of compensating all the above errors has not yet emerged. Another problem in this connection is that of finding the feedback connections of the shift register sequence generator for providing delayed replicas

(needed in the above correlation experiment) of the generated basic m-sequence.

The content of this chapter is, therefore, presented as follows -

Section 2.2 is devoted to a simplified exposition of the theory of correlation method of system dynamic analysis. In Section 2.3, an attempt is made to collect together and compare the most important features of the presently available correlation schemes for single-input/ single output time-invariant linear system identification using random and pseudorandom techniques. As a result of this comparative study, in Section 2.4, a new correlation scheme is described in which ~~the~~ simultaneous crosscorrelation of system output with the input pseudorandom binary test perturbation and its phase inverse is affected to obtain an accurate estimate of the dynamics without incurring errors due to imperfections in the test signal and corrupting effects of noise and drift in the system. In Section 2.5, a simple and quick method of evaluating the binary feedback shift register connections for providing the desired delayed version of the originally generated pseudorandom binary test sequence (as required in the above crosscorrelation experiment) is broughtforth.

2.2. LINEAR SYSTEM DYNAMICS BY CROSSCORRELATION METHOD

2.2.1 Input-output relations for linear systems

Consider the system shown in Fig.(2.1). The input signal $x(t)$ produces a response signal $y(t)$. In experimental work, the signals $x(t)$ and $y(t)$ represent perturbations added to steady-state input and output levels; here and subsequently, input and output will be taken as meaning deviations from the normal operating levels. It is assumed that there is noise in the process and/or the measuring device, and that this may be represented by an output noise signal $n(t)$. The measurable output signal $z(t)$ is, then, the sum of the system response $y(t)$ and the hoise $n(t)$. Thus:

$$z(t) = y(t) + n(t) \qquad \dots \qquad \dots \qquad \dots \quad (2.1)$$

The system is assumed to be linear, physically realizable and time-invariant. The principle of superposition is thus valid; and the responses to different inputs $x(t)$ can be added to derive the response to an input, which is the sum of these different inputs. When care is exercised to keep the deviations from the operating level, caused by the applied inputs, within reasonable limits, the assumption of linearity is quite valid for practical systems.

The process-response, $y(t)$, is given by a weighted sum of inputs which have occured in the past. The past inputs are

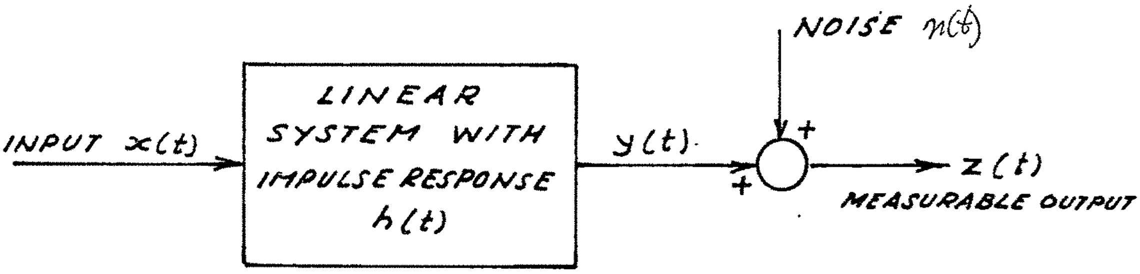

FIG. 2·1 (a) BLOCK DIAGRAM OF SYSTEM

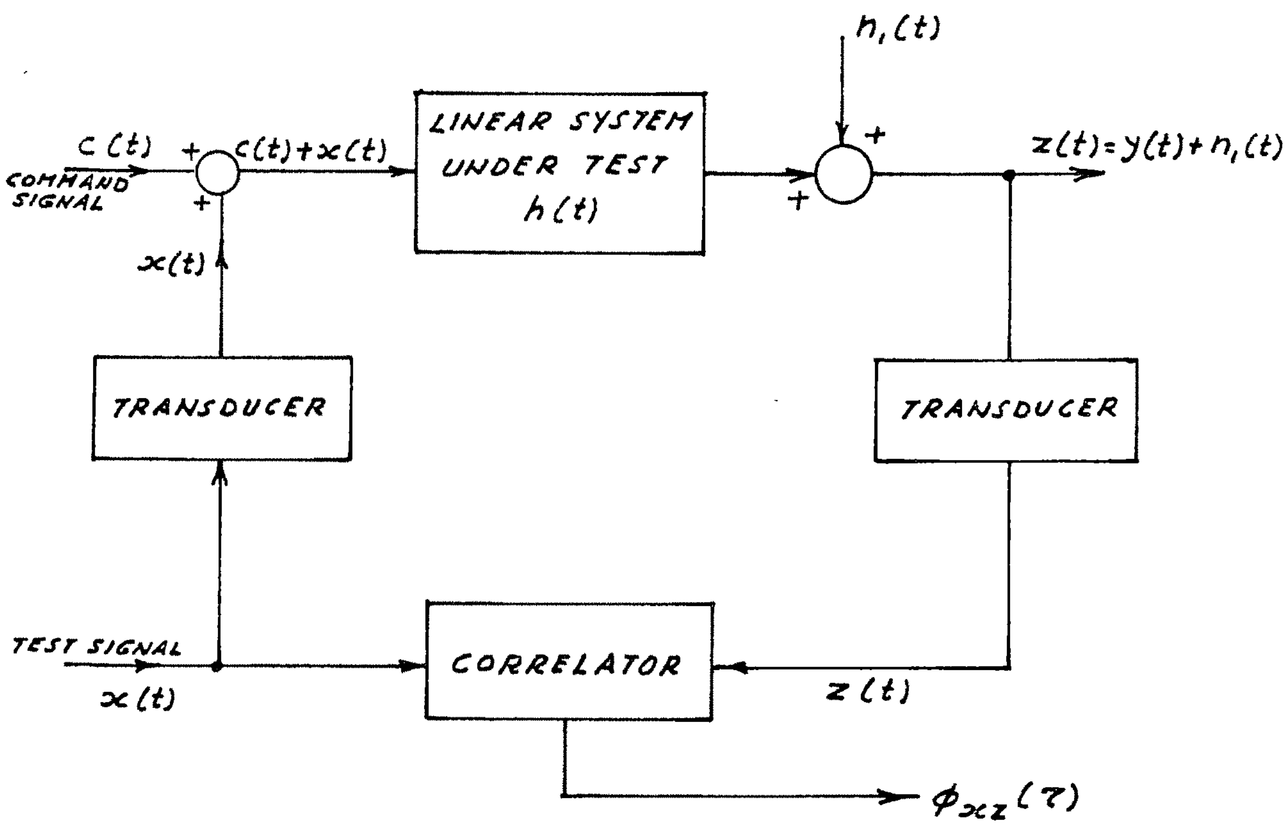

FIG. 2·1 (b) USUAL IMPULSE RESPONSE ESTIMATION SCHEME multiplied by a function h(t), called the system weighting function or the impulse-response of the system. The time-domain relationship between the input, x(t), and the system response y(t) is given by :

$$y(t) = \int_{0}^{\infty} h(u).x(t - u)\,du \quad \ldots \quad \ldots \qquad (2.2)$$

where u is a time variable. Equation (2.2) is called the convolution-integral, and is illustrated in Fig.(2.2). As shown in the figure, Fig.(A) represents a typical weighting function. In Fig. (B) is shown an input signal, arbitrarily chosen. x(-u), depicted in Fig. (C), is the signal x(u) rotated about u = 0 axis. x(t - u) is then the signal x(-u) shifted by an amount t, as shown in Fig. (D). The product signal h(u).x(t - u) is shown in Fig. (E). The system response y(t) is then the integral of this product over values of u from zero to infinity.

In a practical situation, it is normally assumed that inputs occuring at times greater than T_s in the past have no effect on the present output. In such a case eqn. (2.2) modifies to the form :

Nota Correct Statement

$$y(t) = \int_{0}^{T_s} h(u).x(t - u)\,du \quad \ldots \quad \ldots \qquad (2.3)$$

whereby T_s is understood to be the settling time of the

This is why the integral is only between 0 and T_s,

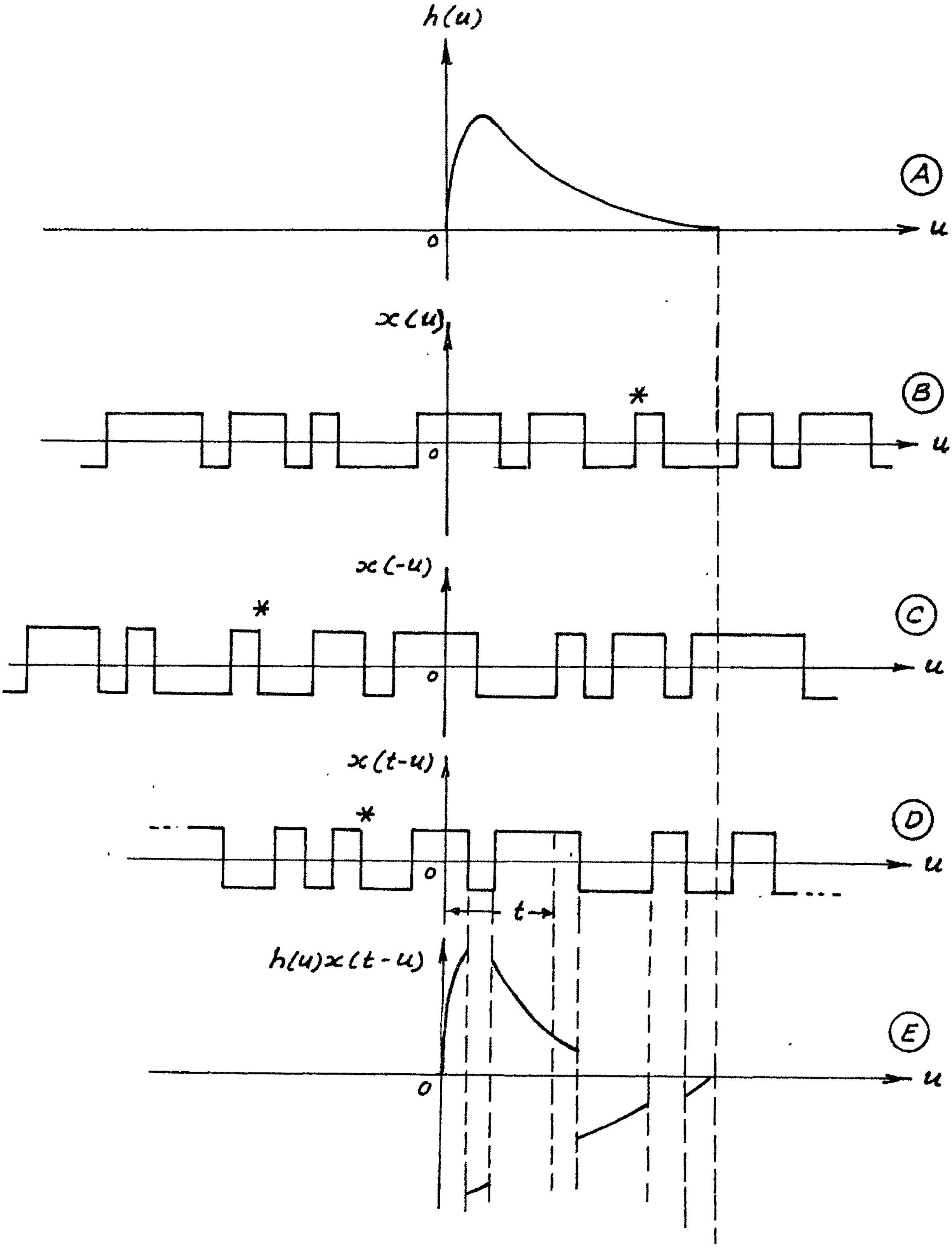

FIG. 2·2 REPRESENTATION OF CONVOLUTION INTEGRAL linear system. (Generally T_s is taken as 5 times the greatest ~~longest~~ system time-constant).

The measurable output z(t) may then be expressed as :

$$z(t) = \int_{0}^{T_s} h(u).x(t - u)du + n(t) \quad \ldots \quad (2.4)$$

Eqn. (2.4) gives the time-response of the linear system to an arbitrary input signal x(t). In principle, the impulse-response h(u) may be obtained by solving this equation, i.e. by deconvolution method. Obviously, this deconvolution is by no means straight forward unless the input signal assumes a simple form. For instance, when x(t) is a short-duration pulse, output, z(t) may be written as -

$$z(t) \simeq A\, h(t) + n(t) \quad \ldots \quad \ldots \quad (2.5)$$

where A is a constant depending on the input signal energy. Needless to say, the accuracy of the result depends on the pulse-duration - the accuracy becomes more as the pulse duration is shorter.

However, as stated in the introduction of this chapter, application of the simple convolution integral under normal operating conditions of the system requires either a large amplitude perturbation or a long experimentation time for measurement of system dynamics to a reasonable accuracy. It will not be seen that correlation method is well-suited under the circumstances.

2.2.2 Correlation functions

Before considering how correlation methods can be used to measure the system dynamics, the meaning of correlation function and some of their properties of pertinent interest are stated in this sub-section.

A function whose value changes in an unpredictable manner with respect to changes in the independent variable is known as a random function. In cases, where the independent variable is the time, the random function is called a stochastic function. Principal concern here is with a class of random functions, the stationary stochastic functions - which means the statistics of the function do not change with change of the time. A set of such random functions (ensemble) constitutes a stationary random process. The ensemble can be described by certain probability distributions. Thus a stationary stochastic function x(t) is said to be known when all its n-dimensional probability distributions are known. However, the experimental determination of these quantities is quite involved and usually the analysis of such a signal is limited to knowing the first - and second probability distributions. It is well known from the probability - theory that the average of the function x(t) is given by

$$\overline{x(t)} = \int_{-\infty}^{\infty} x.p(x)\,dx \quad \ldots \quad \ldots \qquad (2.6)$$

whereby p(x) is the one-dimensional probability density function satisfying the conditions :

$$p(x) \geqslant 0$$
$$\int_{\infty}^{\infty} p(x)\,dx = 1$$

and p(x) dx is approximately the probability that the value of the function x(t) lies between x and x + dx.

Likewise, the average of the function $x(t).x(t + \tau)$ is given by :

$$\overline{x(t).x(t + \tau)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2.p(x_1, x_2, \tau)\,dx_2 dx_1 \quad .. \; (2.7)$$

whereby $p(x_1. \; x_2, \; \tau)$, the second probability density function specifies the probability that the value of the function x(t) at time t_1 lies between x_1 and x_1 + dx, and that τ secs later ~~lies~~ the value lies between x_2 and x_2 + dx.

Further, ergodicity is the property of stationary random processes, which essentially implies that after a sufficient length of time the effect of initial conditions is negligible. Ergodicity, thus, involves the assumption that the time-average over one member function of a stationary random process is equivalent to the average over the ensemble of functions. Hence, the following equations may be written for a stationary stochastic function x(t) :

$$\overline{x(t)} = \int_{-\infty}^{\infty} x.p(x)\,dx = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} .x(t)\,dt = \text{constant} \quad \ldots (2.8)$$

$$\overline{x(t).x(t+\tau)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 . p(x_1, x_2, \tau)\,dx_1 dx_2$$

$$= \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t).x(t+\tau)\,dt = R_{xx}(\tau) \quad \ldots (2.9)$$

The function $R_{xx}(\tau)$ is called the ' ideal autocorrelation function of the signal x(t)'. Thus, the autocorrelation function can be considered either in the light of the shifting, multiplying and averaging process or from the viewpoint of the probability distributions describing the stationary time series, in particular the second probability distribution. For experimental measurement of the autocorrelation function, the former viewpoint is ordinarily strongly preferable. Qualitatively, the autocorrelation function is a measure of the regularity of the function. If the value τ secs from now is closely dependent upon the present value, $R_{xx}(\tau)$ will in general be large.

Some important properties of present interest of the autocorrelation function are listed below :

1. $\phi_{xx}(\tau) = \phi_{xx}(-\tau)$
2. $\phi_{xx}(\tau) \leq \phi_{xx}(0)$
3. $\lim\limits_{\tau \to \infty} \phi_{xx}(\tau) = (\bar{x})^2$
4. $\phi_{xx}(0) = \overline{x^2}$
5. $\phi_{xx}(\tau)$ is periodic if $x(t)$ is periodic.
6. If $x(t)$ is the sum of two uncorrelated component signals $x_1(t)$ and $x_2(t)$, i.e.

   $x(t) = x_1(t) + x_2(t)$, then ,

   $\phi_{xx}(\tau) = \phi_{x_1x_1}(\tau) + \phi_{x_2x_2}(\tau)$

Equation (2.9) (Time-average) is illustrated in Fig. (2.3).

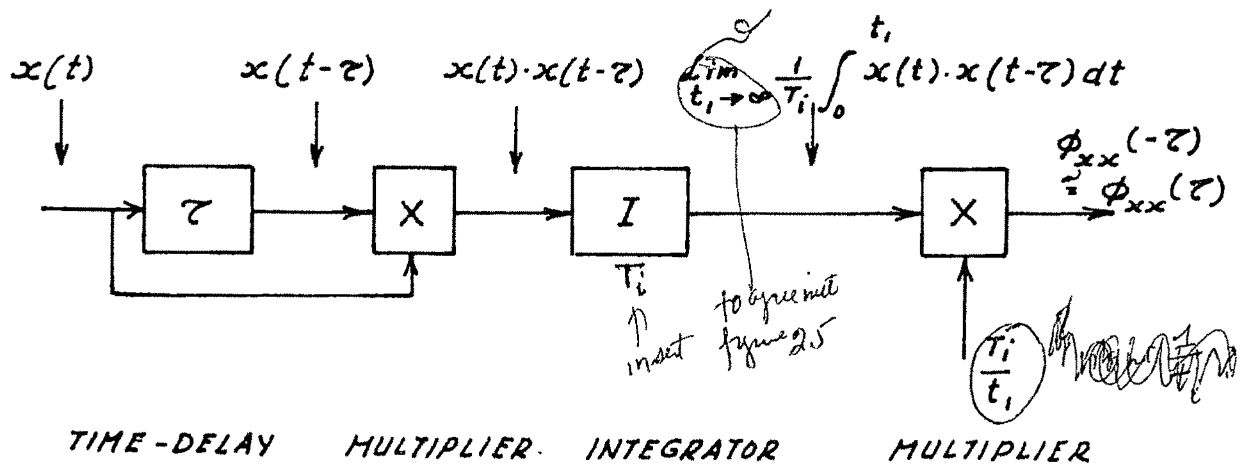

FIG. 2·3 BLOCK-DIAGRAM FOR THE DETERMINATION OF AUTO CORRELATION FUNCTION $\phi_{xx}(\tau)$

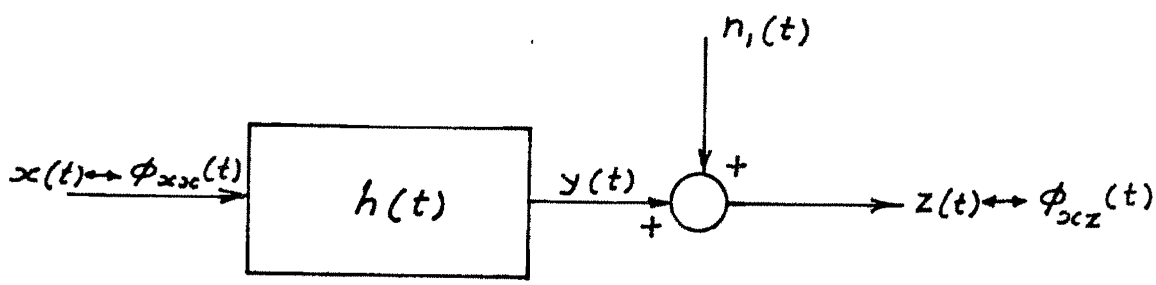

FIG. 2·4 ANALOGY BETWEEN $x(t)$, AND $\phi_{xx}(t)$, AND $z(t)$ AND $\phi_{xz}(t)$.

While the autocorrelation function specifies the statistical dependence of a function and its time-shifted version, the function called as 'crosscorrelation function' serves to characterize the statistical dependence of two different time-functions. Thus, the ideal crosscorrelation function, $R_{xz}(\tau)$, between two different functions x(t) and z(t) which are stationary (i.e. whose statistical properties such as mean level and mean squared amplitude remain constant with time) is defined by the following equation :

$$R_{xz}(\tau) = \overline{x(t).z(t+\tau)} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t).z(t+\tau)\,dt \qquad \ldots (2.10)$$

Since the signal corresponds to a stationary process, the crosscorrelation function $R_{xz}(\tau)$ may also be expressed as :

$$R_{xz}(\tau) = \overline{x(t-\tau).z(t)} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t-\tau).z(t)\,dt \qquad \ldots (2.11)$$

The physical interpretation of the above equation is that very long portions of the signals x(t) and z(t) are to be considered, and z(t) is time shifted with respect to x(t) by an amount of τ. The signals x(t) and $z(t+\tau)$ are then multiplied together and the integral of this multiplication determined. $R_{xz}(\tau)$ is then obtained by dividing the result of integration by the time over which the product $x(t).z(t+\tau)$ is integrated.

Some properties of crosscorrelation function of a stationary signal are stated below :

(i) The crosscorrelation functions $R_{xz}(\tau)$ and $R_{zx}(\tau)$ are related to each other by the simple formula:

$$R_{xz}(\tau) = R_{zx}(-\tau)$$

It may be noted here, that unlike the autocorrelation function, the crosscorrelation function is not necessarily an even function of its argument. In general, it is neither even nor odd. Further, the value of $R_{xz}(\tau = 0)$ need not be the maximum.

(ii) Similar to autocorrelation function, the crosscorrelation function tends to zero as the displacement tends to infinity in either directions, assuming zero d-c components in the signals. i.e.

$$R_{xz}(\tau \rightarrow \pm \infty) = 0$$

(iii) The crosscorrelation function $R_{xz}(\tau)$ is continuous for all values of the argument τ in the range $(-\infty, \infty)$ if $R_{zz}(\tau)$ is continuous at its origin.

(iv) The crosscorrelation function is expressible as a Fourier integral. Thus,

actually the dual relationship

$$R_{xz}(\tau) = \int_{-\infty}^{\infty} R_{xz}(w)\, e^{jw\tau} dw$$

where the function $R_{xz}(w)$ is called the cross-power density spectrum of the functions x(t) and z(t).

$R_{xz}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_{xz}(\tau)\, e^{-j\omega\tau} d\tau$

2.2.3 Correlation method for dynamic analysis

The use of correlation functions in the study of system dynamics can now be examined. For this purpose, the following empirical correlation functions are defined :

Empirical autocorrelation function $\phi_{xx}(\tau)$ is given by :

$$\phi_{xx}(\tau) = \frac{1}{T}\int_0^T x(t).x(t+\tau)\,dt \quad \ldots \quad \ldots \quad (2.12)$$

Empirical crosscorrelation function is given by :

$$\phi_{xz}(\tau) = \frac{1}{T}\int_0^T x(t-\tau).z(t)\,dt \quad \ldots \quad \ldots (2.13)$$

whereby T is now finite, since in experimental work the observation time is always finite.

As T approaches infinity -

$$\phi_{xx}(\tau) \longrightarrow R_{xx}(\tau)$$

$$\phi_{xz}(\tau) \longrightarrow R_{xz}(\tau)$$

From equation (2.1) and (2.2), the crosscorrelation function $\phi_{xz}(\tau)$ may be written :

$$\phi_{xz}(\tau) = \frac{1}{T}\int_0^T x(t-\tau)\,\left[y(t) + n(t)\right]dt$$

$$= \frac{1}{T}\int_0^T x(t-\tau)\,dt \int_0^{T_s} h(u).x(t-u)\,du$$

$$+ \frac{1}{T}\int_0^T x(t-\tau).n(t)\,dt \quad \ldots \quad (2.14)$$

By changing the order of integration -

$$\phi_{xz}(\tau) = \frac{1}{T}\int_{o}^{T_s} h(u).du. \int_{o}^{T} x(t-\tau).x(t-u)dt$$

$$+ \frac{1}{T}\int_{o}^{T} x(t-\tau).n(t)dt \quad \ldots \quad \ldots \quad (2.15)$$

Now, in accordance with the definitions of empirical auto- and crosscorrelation functions given in equations (2.12) and (2.13), ϕ_{xz} may be written as :

$$\phi_{xz}(\tau) = \int_{o}^{T_s} h(u).\phi_{xx}(\tau - u)dt + \phi_{xn}(\tau) \quad \ldots \quad (2.16)$$

where $\phi_{xn}(\tau)$ is the crosscorrelation function between the input signal x(t) and the noise signal n(t) referring to the linear system shown in Fig. (2.1).

In case, the disturbance n(t) in the system is assumed to be unaffected by the test signal, x(t), the average (over all time) of their product becomes zero, so that -

$$\phi_{xn}(\tau) = 0 \quad \ldots \quad \ldots \quad \ldots \quad (2.17)$$

Hence, assuming a very large correlation time, eqn.(2.16) can be written as :

$$\phi_{xz}(\tau) \simeq \int_{o}^{T_s} h(u).\phi_{xx}(\tau - u)du \quad \ldots \quad (2.18)$$

The significant feature of eqn. (2.18) is its resemblance to eqn. (2.3). Comparison of the two equations indicates that $\phi_{xz}(\tau)$ would be the response of the system when excited by the input $\phi_{xx}(\tau)$. This analogy is shown in Fig. (2.4). Further, in many practical systems, it is usually necessary to limit the amplitude of the test signal x(t), both for reasons of plant safety and also in order to preserve the assumed system nonlinearity. In such a case, the noise term n(t) in eqn. (2.4) may become appreciable part of the measured output z(t). Thus, in determining the dynamic characteristics of a system in presence of hoise, the correlation function relationship of equation (2.18) is to be preferred to the direct evaluation of eqn. (2.4). However, the solution of the integral equation (2.18) for h(u) would generally be a complicated procedure, but by using certain test signals, several practical and theoretical simplifications emerge.

In particular, if the autocorrelation function of the input signal x(t) approximates to a delta-function (a large amplitude, very narrow pulse of unit area), that is :

$$\phi_{xx}(\tau) \cong \delta(\tau) \qquad \ldots \qquad \ldots \qquad \ldots \qquad (2.19)$$

then, assuming that $\phi_{xn}(\tau)$ equals zero, eqn.(2.18) reduces to its simplest form:

$$\left.\begin{aligned} \phi_{xz}(\tau) &\simeq h(\tau), && \text{for} \quad \tau > 0 \\ &\simeq h(0)\,/\,2 && \text{for} \quad \tau = 0 \end{aligned}\right\} \quad \ldots \quad (2.20)$$

why?

Equation (2.20) states that the determination of an approximate estimate of the system impulse response is synonymous with the measurement of the crosscorrelation function between a suitably chosen test perturbation and the response it evokes from the system. This is shown in Fig. (2.5).

In brief, the measured values of the crosscorrelation function are proportional to the ordinates of an approximate weighting function of the system. However, as previously stated, the estimated values of the weighting function will be subject to variations due to the existence of imperfect input transducer, imperfections in the input signal characteristics, and wide-band and band-limited noise. Thus, confidence in the estimated results is reduced. In order to eliminate or minimize such errors, several ideas have been explored, and in what follows immediately, a comparative study of the presently available correlation schemes for the linear system identification is given.

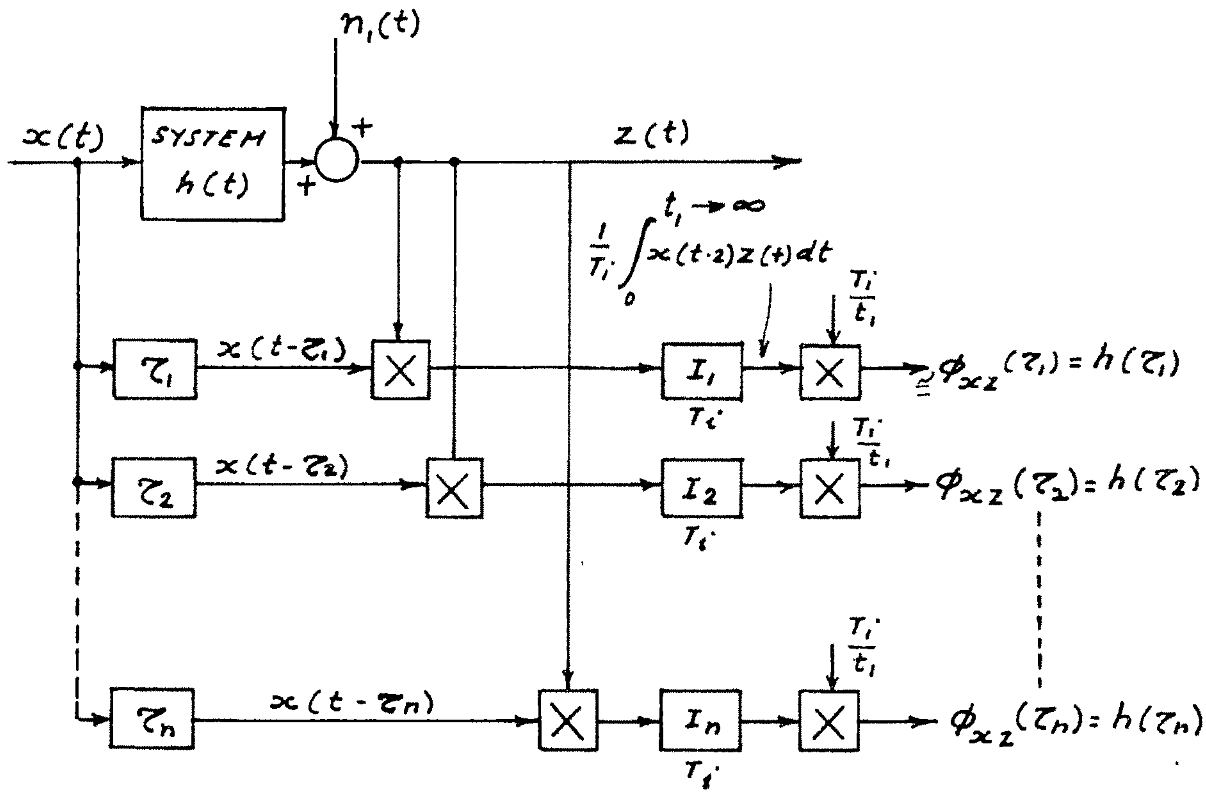

FIG. 2.5 BLOCK-DIAGRAM FOR MEASUREMENT OF IMPULSE RESPONSE $h(t)$

I have not entered from here on detailed amendments like in the earlier pages. Student and adviser ought go thru and amend themselves similarly. — I have indicated such for the Conclusions and References.

2.3 COMPARATIVE STUDY OF CURRENTLY AVAILABLE CORRELATION SCHEMES FOR LINEAR SYSTEM IDENTIFICATION

2.3(A) : Scheme 1 - Correlation analysis using binary random noise as test signal.

(Cooper 1960, Levin 1960, Hughes and Noton 1962, Poortvliet 1963)

2.3.(A.1) Estimation of system weighting function

The first comprehensive report to be mentioned in this connection is the work of Hughes and Noton, who have examined a procedure for the emperical determination of system dynamic characteristics in which the time of measurement may be exchanged for test-signal/noise ratio.

In this scheme, Fig. (2.6), the test signal $x(t)$, which is derived from binary noise, corresponded to a sample of length T, from an infinite random time series. Its amplitude is considered to be $\pm$ a, and the repetetion period T is kept significantly large in comparison with the time constants of the system under test. The value of $x(t)$ is constant over any one of the N sub-intervals of length t_o, $(T = Nt_o)$, and the probability of a change of sign at the end of the sub-interval is one half. Thus $x(t)$ here is a repetetive random test-input, and is depicted in Fig. (2.7) together with its autocorrelation function $\phi_{xx}(\tau)$. Further, with the choice of

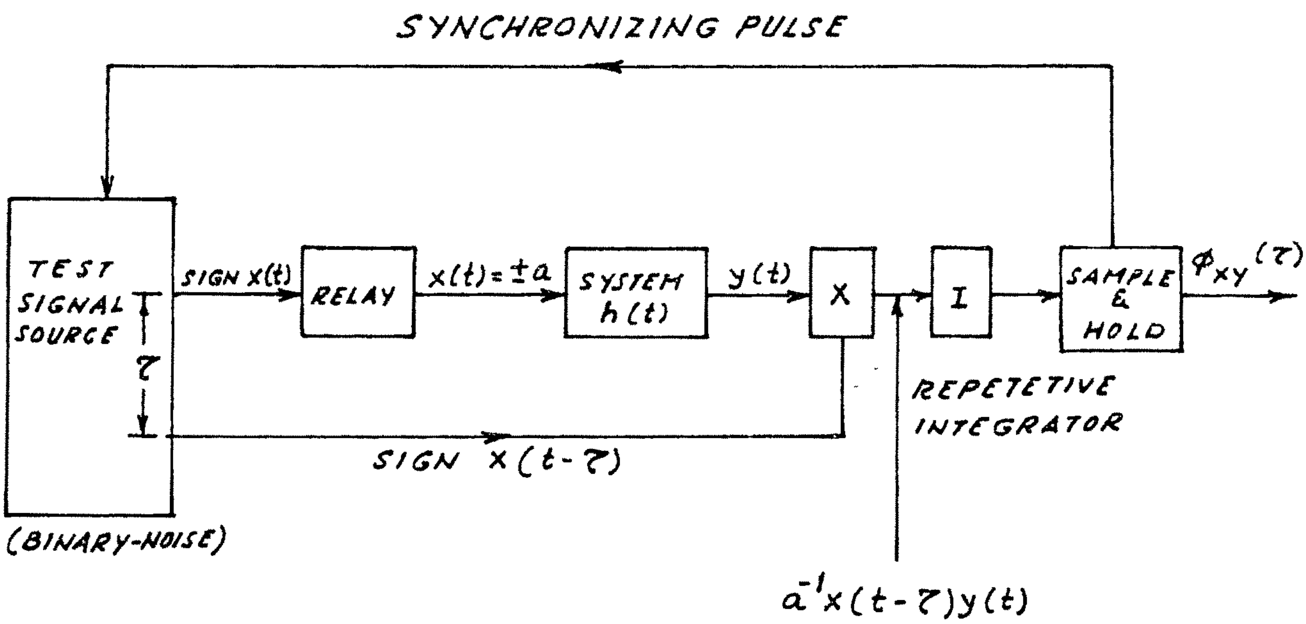

FIG. 2·6 BLOCK-DIAGRAM OF THE CROSS-CORRELATOR USED IN SCHEME 1

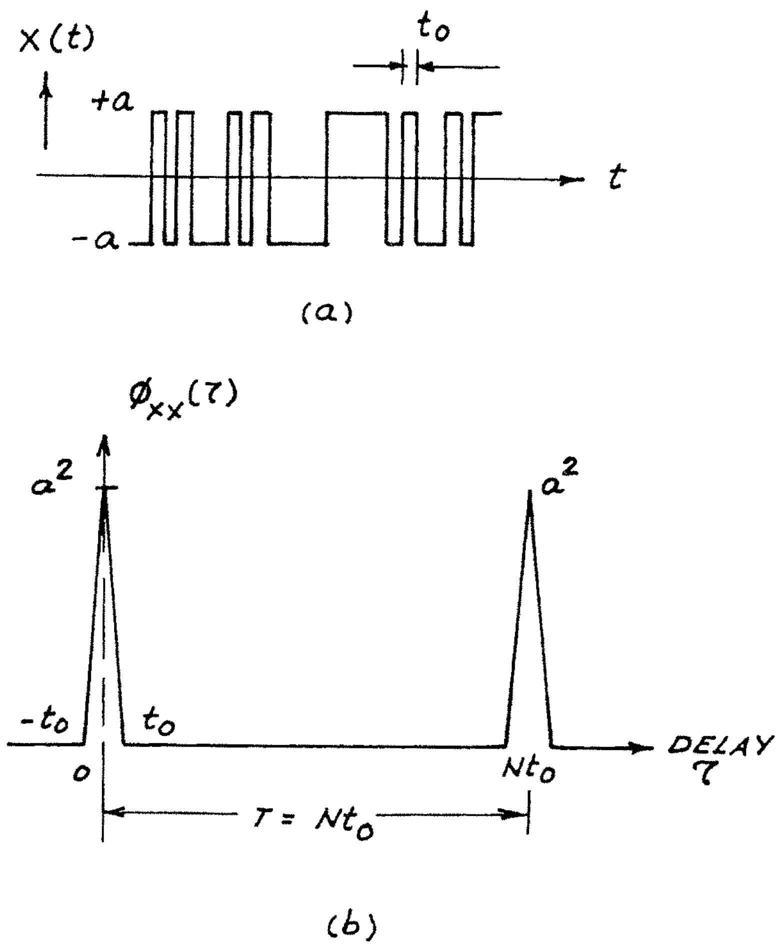

FIG. 2·7 (a) RANDOM BINARY TEST SIGNAL USED IN SCHEME 1

(b) AUTO CORRELATION-FUNCTION OF OF THE TEST SIGNAL the binary form for input, the multiplication operation has been converted into a simple switching process.

Now, assuming that the sub-interval t_o of x(t) to be quite small compared with the time constants of the system, the autocorrelation function of the input signal is expressed as -

$$\phi_{xx}(\tau) = \sum_{n=-\infty}^{n=+\infty} a^2 t_o \, \delta(\tau - nT) \quad \ldots \quad \ldots \quad (2.21)$$

where $\delta(\tau)$ is a symmetrical Dirac delta function. And to a first approximation, $\phi_{xx}(\tau)$ is considered to be an impulse of strength $a^2 t_o$ centered at $\tau = 0$.

Referring to eqn.(2.16) the emperical input-output crosscorrelation function $\phi_{xz}(\tau)$ for the configuration of Fig. (2.6) is :

$$\phi_{xz}(\tau) = \int_o^{T_s} h'u) \cdot \phi_{xx}(\tau - u)du + \phi_{xn_1}(\tau) \quad \ldots \quad (2.22)$$

Substituting for $\phi_{xx}(\tau - a)$,

$$\phi_{xz}(\tau) = a^2 t_o \, h(\tau) + \phi_{xn_1}(\tau) \quad \ldots \quad \ldots \quad (2.23)$$

Assuming the effects of noise in the measurement of $\phi_{xz}(\tau)$ to be small, the weighting function $h(\tau)$ is determined. And, it has been shown that the method works even when the noise and signal power levels are comparable, in which case crosscorrelation over a number of test signal periods is necessary to obtain accurate results.

2.3.(A.2) Errors associated with the scheme

Evidently, the measured impulse response will be subject to variations in environmental disturbances, and assumptions made.

Eqn. (2.21) is derived assuming that :

(i) the input-transducer (Fig.2.1) is ideal,

(ii) the basic interval t_o of input $x(t)$ is zero, and

(iii) its autocorrelation function $\phi_{xx}(\tau)$ can be thought of an impulse of strength ' $a^2 t_o$ ' centered at $\tau = 0$.

In practice, none of these is true and thus errors due to imperfect input transducer, finite band width of the input and imperfections in its autocorrelation function are unavoidable. In such a case, the input-output correlation function in eqn. (2.21) is -

$$\phi_{xz}(\tau) = a^2 t_o \, h(\tau) + \phi_{xn_1}(\tau) + e_t(\tau) + e_x(\tau) \qquad \ldots (2.22)$$

where $e_t(\tau)$ and $e_x(\tau)$ account, respectively, for the errors in the measurement of $\phi_{xz}(\tau)$ due to non-ideal input transducer, and imperfections of the input signal.

Also, the noise-term, $\phi_{xn_1}(\tau)$, refers to the random noise in the system say n(t) ; and to the disturbance caused to the measurement of $\phi_{xz}(\tau)$ by the slowly varying operational inputs, called the low-order output drift, d(t). Thus :

$$n_1(t) = n(t) + d(t) \quad \ldots \quad \ldots \qquad (2.23)$$

so that -

$$\phi_{xn_1}(\tau) = \phi_{xn}(\tau) + \phi_{xd}(\tau) \quad \ldots \qquad (2.24)$$

Further, it is shown that errors due to wide-band noise may be minimised by averaging the input-output correlation function over a large number of test signal periods rather than just one period, i.e.

$$\phi_{xz}(\tau) = \frac{1}{mT}\int_0^{mT} x(t-\tau).z(t)\,dt \quad \ldots \qquad (2.25)$$

Specifically, the rms error in the impulse response h(t) varies inversely as the square root of m, the number of periods over which the integration is carried out.

Concerning the slowly varying operational inputs, the output drift d(t) is approximated over the input signal period T, as -

$$d(t) = k_o + k_1 t \qquad \ldots \qquad \ldots \qquad (2.26)$$

where k_o and k_1 are merely constants. Substituting for d(t), $\phi_{xd}(\tau)$ becomes -

$$\phi_{xd}(\tau) = \frac{k_o}{mT} \int_o^{mT} x(t - \tau)\,dt + \frac{k_1}{mT} \int_o^{mT} t.x(t - \tau)\,dt \qquad \ldots \qquad (2.27)$$

As can be seen, the first term in eqn. (2.27) may be eliminated by making the mean value of the input to be zero over its period T. The second term, although believed to be negligible, cannot be totally discarded.

Thus, the final expression in this scheme for input-output correlation function $\phi_{xz}(\tau)$ is given by -

$$\phi_{xz}(\tau) = a^2 t_o\, h(\tau) + \overset{*}{k} e_t(\tau) + \overset{*}{k} e_x(\tau)$$

$$+ \frac{1}{mT} \int_o^{mT} x(t - \tau) \left[n(t) + k_o + k_1 t \right] dt \qquad \ldots \qquad (2.28)$$

where $\overset{*}{k}$ in the second and third terms is a constant and accounts for the error $e_t(\tau)$ and $e_x(\tau)$ when correlation is performed over m periods of the test signal.

2.3(A.3) Remarks

The main conclusions of the scheme are as follows :

(i) Errors due to finite band-width of the test signal can be reduced by making the sub-interval of the signal t_o to be much smaller than the minimum time constant of the system.

(ii) Errors due to non-ideal autocorrelation function of the test signal can be minimized by making the test signal period sufficiently large.

(iii) Errors due to the wide-band noise can be reduced by carrying out Crosscorrelation over a number of test signal periods (say mT), and

(iv) Errors due to constant output drift can be eliminated by ensuring that the binary test perturbation has zero mean, and that the linearly varying drift error may be minimized with the choice of ideal-binary signal.

In practice, there is a limit to choosing the values of the basic interval t_o, signal repetetion period T, and the number of correlation periods m.

If the basic interval t_o is kept very small, then, the speed at which the input-transducer (Fig.2.1) moves from one binary state to the other increases, and consequently the transition time of the modulator is of increasing significance and requires consideration. So much so, it is desirable that a reasonable basic interval of the test signal be chosen.

The repetetion period T of the binary random signal cannot be also made very large because of the requirement set for rapid determination of any changes in the dynamic characteristics of the system.

Also, the value of the number m cannot be made much greater because of the time involved.

As such, departure from ideal performance of the crosscorrelator should be rather expected.

Besides the above, this scheme has two practical disadvantages :

(i) Theoretically speaking, infinite correlation time is necessary to ensure that the autocorrelation function of the random test signal approximates to a true delta function. Further, the test signal being essentially stochastic, the effect of a finite correlation time can be determined only in a probabilistic sense.

(ii) It is not easy to transmit an amplitude modulated random noise without distortion, especially when electromechanical transducers are involved.

For these reasons, deterministic signals possessing the required autocorrelation function are preferred, and in what now follows, a correlation scheme using such a perturbation is considered.

2.3(B) : Scheme 2 - Correlation analysis using pseudo-random binary test signal. (Single-input/Single-output system)

(Cummins 1964, Korn 1964, Briggs et.al. 1964-65, Hazlerigg and Noton 1965, Barber and Hammond 1965, Chang 1965, Doüce and Ng 1965, Godfrey 1965, 1966, 1967, 1968, 1969; Gupta 1965, Barker 1967, 1969; Davies WDT, 1966, Douce and Walker 1966, Briggs et.al. 1967, Davies and Douce 1967, Wilson 1967, Macleod 1968, 1969; Davies WDT 1968, Brown 1969, Clarke and Briggs 1970, Anca Tomesscu et. al.1970, Nikiforuk et.al.1970,1971; Hoffman et.al.1972.)

2.3(B.1) Pseudorandom binary test perturbations

In recent years, workers have become aware of the benefits of employing pseudorandom binary test signals(prbs) with idealized autocorrelation functions. A major advantage of this cyclic test signal over the random signal is that, since the characteristic of the signal are well defined, explicit expressions for the system dynamics can be obtained when the effects of noise present in the system are negligible. Even when the system output is contaminated with noise, more confidence can be established in the results because, the significant errors in the results are now known to be contributed by noise alone.

There are at least two ways of generating pseudorandom binary signals -

One is based on Gaussian number theory and makes use of the theory of quadratic residue codes, which exist for $N = 4k - 1$, with k an integer and N a prime number (i.e. for N = 3, 7, 11, 19, 23, 31, 43, ...).(Corran and Cummins 1962, Everett 1966). A sequence is found from the rule that the r'th digit, $d_r = +1$ if r is a square modulo N; and is -1 otherwise; r modulo N means $r - sN$, where s is the largest integer such that sN is less than or equal to r. This rule is applied for the example of N = 31 in Table 2.1; where the numbers in the right hand column are squares modulo-31. It follows that the quadratic residue code of length 31 has +1s in positions 1, 2, 3, 4, 5, 7, 9, 10, 14, 16, 18, 19, 20, 25 and 28; and -1s in positions 6, 8, 11, 12, 13, 15, 17, 21, 22, 23, 24, 26, 27, 28, 29 and 30. The Nth digit (the 31st digit in the example) may either be +1 or -1. It may be seen from Table 2.1 that the same numbers are obtained in the fourth column from $r = 1/2.(N + 1)$ onwards, as those from r = 1 to $r = 1/2.(N - 1)$, but for the reverse order. This is true for all quadratic residue codes, and it is therefore, only necessary to complete the squares, modulo N, upto $r = \frac{1}{N}(N-1)$.

Table 2.1 Computation of the quadratic residue sequence of length 31

r	r^2	Multiple of 31 immediately below r^2	r^2 mod 31
1	1	0	1
2	4	0	4
3	9	0	9
4	16	0	16
5	25	0	25
6	36	31	5
7	49	31	18
8	64	62	2
9	81	62	19
10	100	93	7
11	121	93	18
12	144	124	20
13	169	155	14
14	196	186	10
15	225	217	8
16	256	248	8
17	289	279	10
18	324	310	14
19	361	341	20
20	400	372	18
21	441	434	7
22	484	465	19
23	529	527	2
24	576	558	18
25	625	620	5
26	676	651	25
27	729	713	16
28	784	775	9
29	841	837	4
30	900	899	1

The second class of prbs is based on the properties of digital filters and their maximum length null sequences. The generation and properties of this class is discussed in chapter 1, where it is shown that maximum length sequences (m - sequence) have lengths $N = 2^n - 1$, n being an integer, and these may be generated by means of an n-stage shift register circuit with modulo-2 feedback. Denoting the period of the master clock signal (that activates the shift register) by t_o , the maximum repetetion period T of the m-sequence - or pseudorandom binary sequence signal is given by :

$$\left.\begin{aligned} T &= (2^n - 1)t_o \\ &= Nt_o \end{aligned}\right\} \qquad \ldots \qquad \ldots \qquad (2.29)$$

A prbs switches from one level to the other only at time intervals t_o in a known way, and thus any experiment using one can be repeated using exactly the same input, an impossibility with random-signals. The name 'pseudorandom' derives from the manner in which the signal autocorrelation function approximates to that of white noise. A typical prbs generated by a 5-stage shift-register, together with its autocorrelation function is depicted in Fig.(2.8).

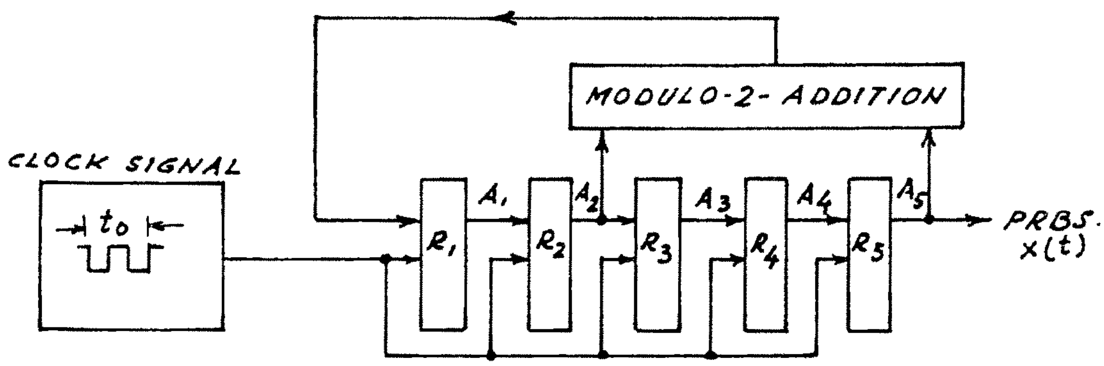

FIG. 2·8 (a) GENERATION OF PRBS USING BINARY F.S.R. AND MODULO-2 LOGIC. ($n=5$)

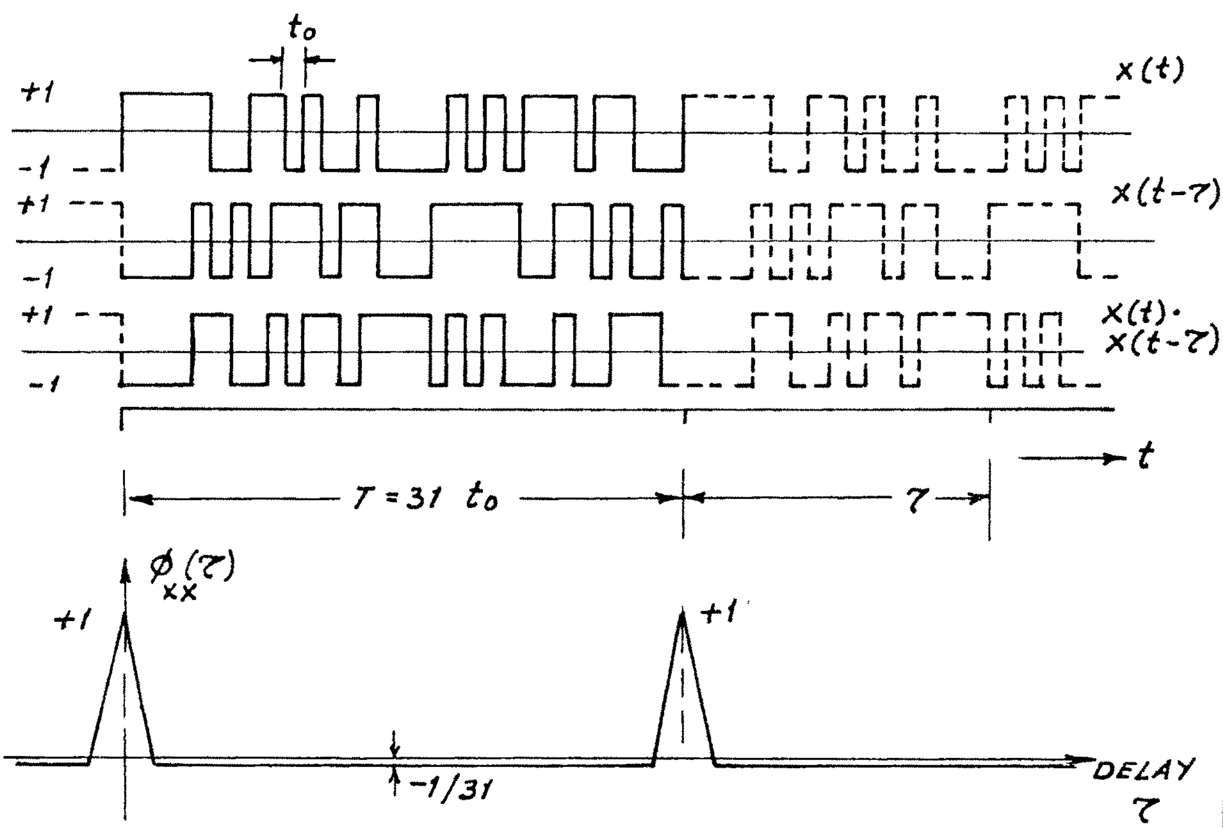

FIG. 2·8 (b) PRBS GENERATED BY F.S.R. OF FIG. (a) AND AUTO CORRELATION FUNCTION OF THE P.R.B.S.

The time autocorrelation of the pseudorandom binary signal, x(t), is given by -

$$\phi_{xx}(\tau) = a^2\left(1 - \frac{N+1}{N}\frac{|\tau|}{t_o}\right) , \text{ for } |\tau| < t_o$$

$$= -\frac{a^2}{N} , \text{ for } t_o \le \tau \le (N-1)t_o \quad \ldots \quad (2.30)$$

where the levels of the prbs are considered to be $\pm$ a, As N is made larger, the autocorrelation function considered over one cycle becomes more like that of white noise, the property that makes prbs so useful in system-identification.

Maximum length sequences have the advantage that they are simple to generate, but the disadvantage that their successive possible lengths are rather widely spaced (by a factor of approximately two). Quadratic residue codes have the advantage that their successive possible lengths are much closer together (since approximately half the numbers of form 4k - 1 are prime) but the disadvantage that they are not easy to generate using logic circuitry, the longer sequences requiring a very large number of logic elements.

The correlation analysis here is concerned with the prbs based on m-sequences only.

2.3.(B.2) : Estimation of weighting function

The correlation scheme employing pseudorandom binary test signal is shown in Fig. (2.9). The conventional experimental configuration of this identification method consist of adding the prbs (defined as the test signal) to the existing input (or set point adjustment), and crosscorrelating the corresponding system output with a delayed version of the reference prbs. The result of this calculation yields an estimate $h(\tau)$ of the system impulse response at a time corresponding to the time delay τ of the delayed prbs with reference to the reference prbs.

To obtain the required points on the impulse response curve, the corresponding delayed prbs must be generated. In section (2.5) of this chapter, the various methodsthat have been proposed for the purpose are characterized and a more simple and quick method of determining the feedback connections of the shift register (m-sequence generator) for providing the delayed replica is presented. For the present, we assume that the delayed replicas can be made available.

The noise signal $n_1(t)$ shown in Fig.(2.9) has the same meaning as referred to that in the scheme 1.

Referring to eqn.(2.16) the input-output correlation function for the system of Fig. (2.9), over m periods of

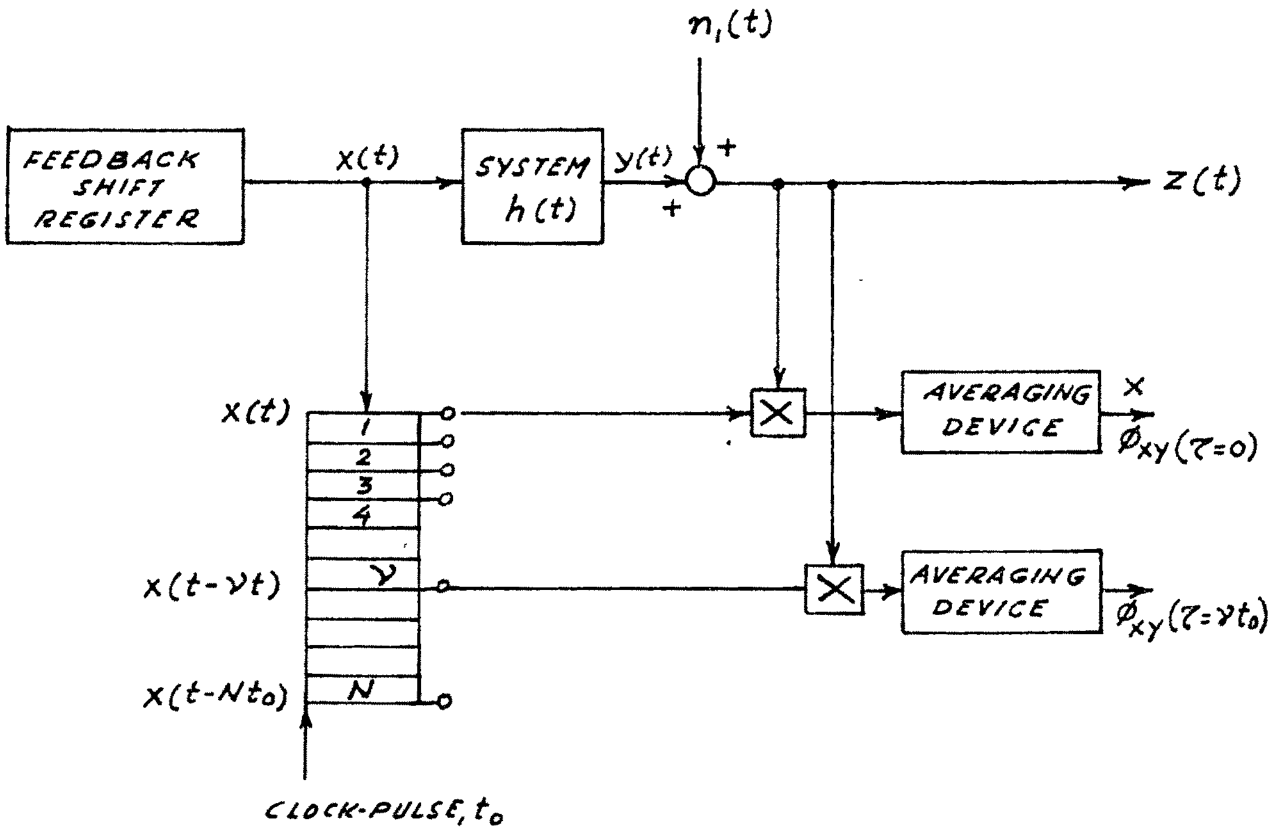

FIG. 2·9 CORRELATION SCHEME 2, ($x(t)$: PRBS).

the test signal from eqn.(2.30) may be expressed as :

$$\phi_{xz}(\tau) = a^2 t_o (1 + \frac{1}{N})\ h(\tau) - \frac{a^2}{N} \int_o^{T_s} h(u)\,du + \phi^*_{xn_1}(\tau) \quad ; \quad ..(2.31)$$

for $\tau > t_o$, provided that the period T of the prbs is greater than T_s, and that h(u) does not vary appreciably over the interval ($-t_o \leq u \leq + t_o$). The integral of the second-term represent the system steady-state gain. The last term $\phi^*_{xn_1}(\tau)$ stands for the corrupting effects of noise when correlation is carried out for m-periods of the test signal.

Now assuming the noise contribution to the measurement to be small, the measured correlator output, corrected for bias, is considered as the direct estimate of the weighting function.

2.3.(B.3) : Errors associated with this scheme

The errors which result from approximating the measured crosscorrelation function to the system impulse response in eqn. (2.31) are primarily due to the existence of :

(i) Imperfect input-transducer,

(ii) bias in prbs and its autocorrelation function,

(iii) finite band-width of prbs,

(iv) wide-band noise, and

(v) low-order output drift.

Accounting for these errors, the correlation function $\phi_{xz}(\tau)$ becomes :

$$\phi_{xz}(\tau) = a^2 t_o(1 + \frac{1}{N})\ h(\tau) - \frac{a^2}{N}\int_o^{T_s} h(u)\,du + \overset{*}{ke}_t(\tau) + \overset{*}{ke}_x(\tau) + \overset{*}{\phi}_{xn}(\tau) + \overset{*}{\phi}_{xd}(\tau) \quad \ldots \quad (2.32)$$

where $\overset{*}{ke}_t(\tau)$ and $\overset{*}{ke}_x(\tau)$ stand for errors due to imperfect transducer, and non-ideal characteristics of test signal respectively, when correlation is considered for a time equal to m periods of the test signal.

(i) Errors due to input transducer has been discussed in a detail manner by Godfrey et. al. (1966) who show that the actual time taken for the production of the signal is not very important provided the transition between the two states are similar. If not, errors are introduced into $\phi_{xz}(\tau)$ at the points corresponding to delays of c = 0, t_o and $k_s t_o$ where k_s depends on the particular prbs under use. Choosing the transition time to be less than one quarter the smallest time constant of the system, the transducer errors are minimized.

(ii) Bias error is shown to be inversely proportional to N (Eqn. 2.30). Unfortunately mean square errors due to wide-band noise are found to be proportional to N, and hence bias error cannot be made suitably small by having N large. Hence correction to $\phi_{xz}(\tau)$ to this effect is inevitable. Usually the estimation of bias is made using several independent values of $\phi_{xz}(\tau)$

(iii) The error due to impulse approximation to $\phi_{xx}(\tau)$, has been shown to be related to the ratio of test-signal bandwidth to system bandwidth. Generally using a ratio of 14 : 1, this error is kept small.

(iv) Effect of wideband noise, as discussed with reference to scheme 1, can be kept small by performing cross-correlation over a large number of test signal periods, although the number seldom exceeds 2 due to the time involved.

(v) Drift, which refers to a slow change of the system output from the desired operating point, introduces notable errors into the estimate of the impulse response especially in chemical and nuclear processes.

Hence elimination of the drift error is an important criterion in the design of an effective crosscorrelator. To this end, several ideas have been explored, and in what now follows, the salient features of existing drift correction schemes are stated.

2.3(B.4) Salient features of existing drift-correction schemes

Douce (1966) has originated the idea that the low-frequency drift signal may be expanded as Taylor-series about time origin, and shown that careful choice of the commencement of input prbs period significantly eliminates the effect of a ramp-formic drift. In mathematical terms, assuming the drift to be :

$$d(t) = kt \quad \ldots \quad \ldots \quad \ldots \quad (2.33)$$

the error in the correlator output has been shown to be equal to the first time moment of the test-signal with zero d.c. level, which becomes a minimum for a particular phase of the input prbs.

Subsequently, Davies (1967) et.al. investigated the problem for polynomial drift expressed as :

$$d(t) = \sum_{i=0}^{N} k_i t^i \quad \ldots \quad \ldots \quad \ldots \quad (2.34)$$

In order to eliminate the drift errors by this technique, (using the notation of Davies) a weighting function $w(T_1)$ is to be chosen, which satisfies the integral equation :

$$\int_0^T w(T_1) \left\{ \int_0^T [d(t + T_1 + s + \tau).x(t + T_1 + s)] \; dt \right\} dt_1 = 0 \quad \ldots \quad (2.35)$$

whereby the observations start at time s and the input signal x(t) of period T (prbs) is considered over the range $(s + T_1)$ to $(s + T_1 + T)$. It has been shown that the above equation is satisfied for polynomial drift signals when $w(T_1)$ is polynomial in T_1, the coefficients of which depend on the order of the drift polynomial, and not on the magnitude of the coefficients of the drift terms. Although quite promising, the method needs considerable off-line computation, besides the required observation time of two test signal periods.

Another noteworthy scheme to minimizing the drift errors is that of Barker (1969), in which a pseudorandom binary signal based on m-sequences has been shown to possess a definite phase that is uncorrelated with constant, and linear signals. Such a phase is called 'reference-phase' of the signal, and it use in crosscorrelation provides the possibility of eliminating constant, linear, quadratic and cubic drift errors by linear weighting of the crosscorrelation over two periods of the input prbs with zero d.c. level. In mathematical terms, the reference phase signal $x_1(t)$ of a prbs $x(t)$ is a translation of $x(t)$ defined by :

$$x_1(t) = x(t + \lambda) \qquad \ldots \qquad \ldots \qquad (2.36)$$

with the properties :

$$\left.\begin{aligned} &\text{(i)} \int_0^T x_1(t)\,dt = 0 \\ &\text{(ii)} \int_0^T t.x_1(t)\,dt = 0 \end{aligned}\right\} \qquad \ldots \qquad (2.37)$$

The particulars regarding the initial setting of the n-stage shift register to realize the reference phase are listed in Table 2.2 for values of n upto 7. For the case of n = 5, the prbs and the corresponding references phase signal are shown in Fig. 2.10

Table 2.2 : $a = (N-1)/2N$ and $b=(N+1)/2N$ are the levels of ref.phase signal.

n	F(D)	N	Initial setting of fsr to obtain the ref.phase							Na	-Nb
			1	2	3	4	5	6	7		
2	$I + D + D^2$	3	1	0	.	.	.	.	.	1	2
3	$I + D + D^3$	7	1	1	0	.	.	.	.	3	4
4	$I + D + D^4$	15	1	0	1	0	.	.	.	7	8
5	$I + D^2 + D^3 + D^4 + D^5$	31	0	1	0	1	0	.	.	15	16
6	$I + D + D^6$	63	0	0	0	1	1	1	.	31	32
7	$I + D^2 + D^3 + D^4 + D^7$	127	1	0	1	1	1	1	0	63	64

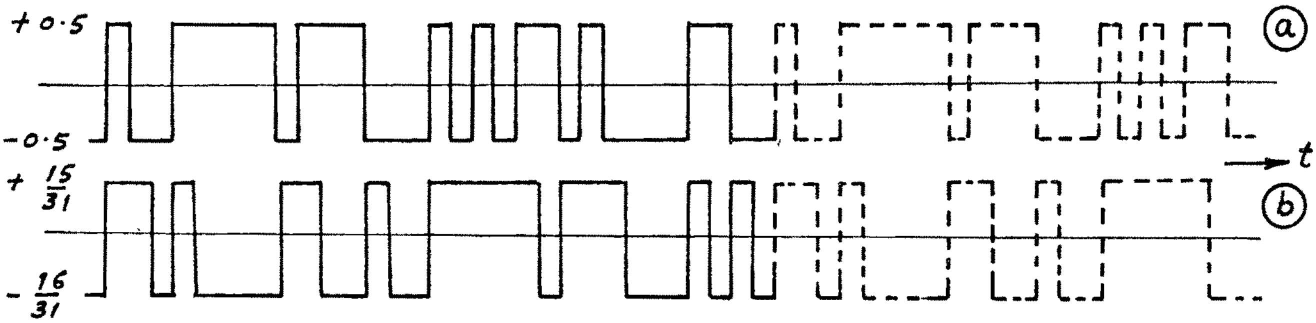

FIG. 2·10 (a) P.R.B.S. GENERATED BY 5-STAGE F.S.R. WITH CHARACTERISTIC POLYNOMIAL $1 \oplus D^2 \oplus D^3 \oplus D^4 \oplus D^5$

(b) CORRESPONDING REFERENCE PHASE P.R.B.S.

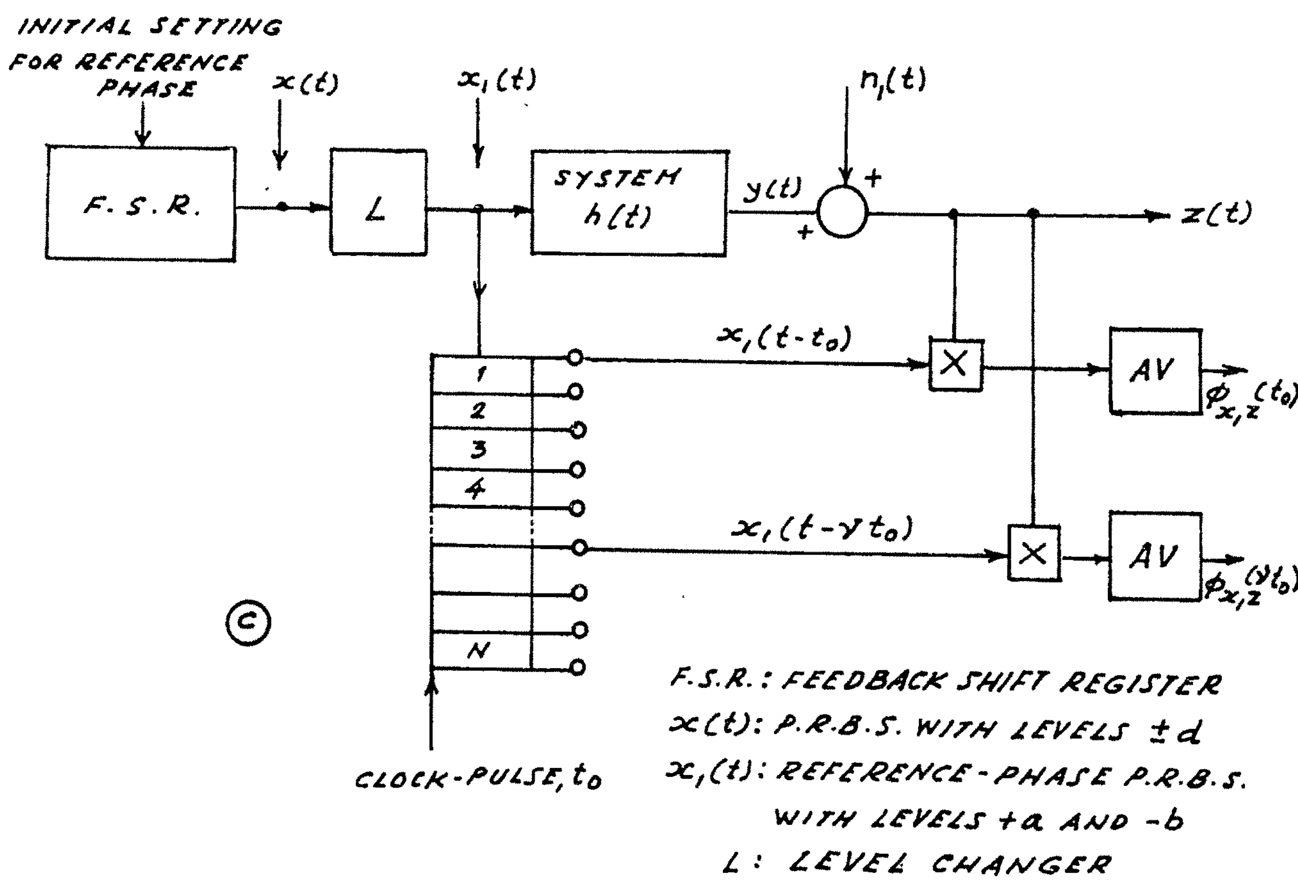

FIG. 2·10 (c) IMPROVED-VERSION OF CORRELATION-SCHEME 2 WITH REFERENCE PHASE P.R.B.S. AS TEST-SIGNAL

However the scope of this reference phase approach in presence of higher order drifts needs consideration. Further, since the signal levels are made unequal, real multiplication is to be effected into the correlator.

Later, Brown (1969) proposed a drift correction scheme wherein a modified m-sequence having 'inverse repeat' property is utilized, as the test signal. By this scheme, the effect of Mth order drift can be eliminated by crosscorrelating over (M + 2) half periods of the modified m-sequence with a binomial weighting given to each half period. Subsequently, Macleod (1969) developed an algorithm, by which it is possible in principle, to eliminate the drift effects upto (N - 2)th power (Nt_o is the period of the signal) by using two periods only of the output using conventional prbs as the input signal. But, when the binomial coefficients become prohibitively large, the feasibility of such a scheme is yet to be examined.

In another drift correction scheme, Brown (1969) has brought-forth a remarkable concept of the drift signals. By considering the drift over q + 1 periods of the input prbs as a truncated Fourier series, it has been shown that a.c. components of drift may be exactly compensated upto the qth harmonic, with uniform distribution of weights. A broader treatment of the drift by Fourier-series expansion thus seems to merit additional effort.

Later, Nikiforuk (1970,1971) et.al. proposed some methods for eliminating the effect of polynomial type drift. The methods are similar to the one suggested by Macleod in that a convolution summation using alternating sign binomial coefficients is performed on system output. However, in these methods the need for the deconvolution process is eliminated. The weakness of these methods is that they tend to amplify the effect of high frequency disturbances. Better results can be arrived at, if these ideas are used together with the reference phase concept of Barker.

2.3.(B.5) : Remarks

Referring back to eqn. (2.32), in view of the above discussion, the input-output correlation may now be written as :

$$\phi_{xz}(\tau) \approx a^2 t_o (1 + \frac{1}{N})\, h(\tau) \quad \ldots \quad \ldots \quad (2.38)$$

provided that -

(i) the bias is corrected ,
(ii) the ratio of signal bandwidth to system bandwidth is of the order of 14 : 1,
(iii) the inter-state transition band of the input transducer is made one quarter the smallest system time constant,
(iv) crosscorrelation is performed for at least 2 periods of the test signal and
(v) a suitable drift correction is employed.

2.3.(C) : Correlation method for linear system identification using matched filters - Scheme 3

(Anca Tomeseu 1970; Korn 1964, Truxal 1958)

2.3.(C.1) : Use of matched-filter in crosscorrelation

Let x(t) be a pseudorandom binary signal acting as a control signal to a linear system, the impulse response of which is denoted by h(t). Then, the filter matched to the input signal x(t) is the filter with the impulse response proportional to the delayed "image" of the signal (Fig.2.11) - i.e.

$$g(t) = A.x\left[-(t-\tau_0)\right] \quad \ldots \quad \ldots \quad (2.39)$$

where A is a constant, and the finite time-delay τ_0 ensures the physical realizability of the filter.

Now, consider the scheme of Fig(2.12), where in the response y(t) of a linear system to the test signal x(t) serves as the input to the matched filter with impulse response g(t), as given by eqn.(2. 39).

The output of the matched-filter $y_1(t)$ is given by :

$$y_1(t) = A \int_{-\infty}^{\infty} y(u).x(u+\tau_0 - t)\,du$$

$$= A.R_{yx}(\tau_0 - t) \quad \ldots \quad \ldots \quad (2.40)$$

Equation (2.40) states that the output of the match filter is a measure of the input-output correlation of the

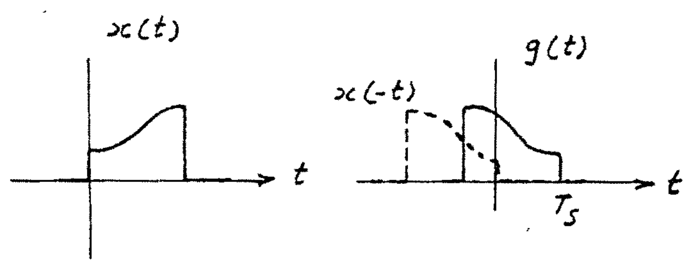

FIG. 2·11 IMPULSE RESPONSE OF A FILTER MATCHED TO THE SIGNAL $x(t)$

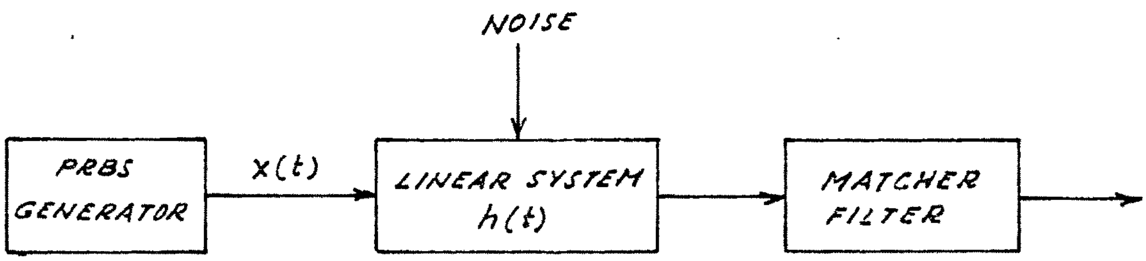

FIG. 2·12 IMPULSE RESPONSE MEASUREMENTS USING MATCHED FILTERS

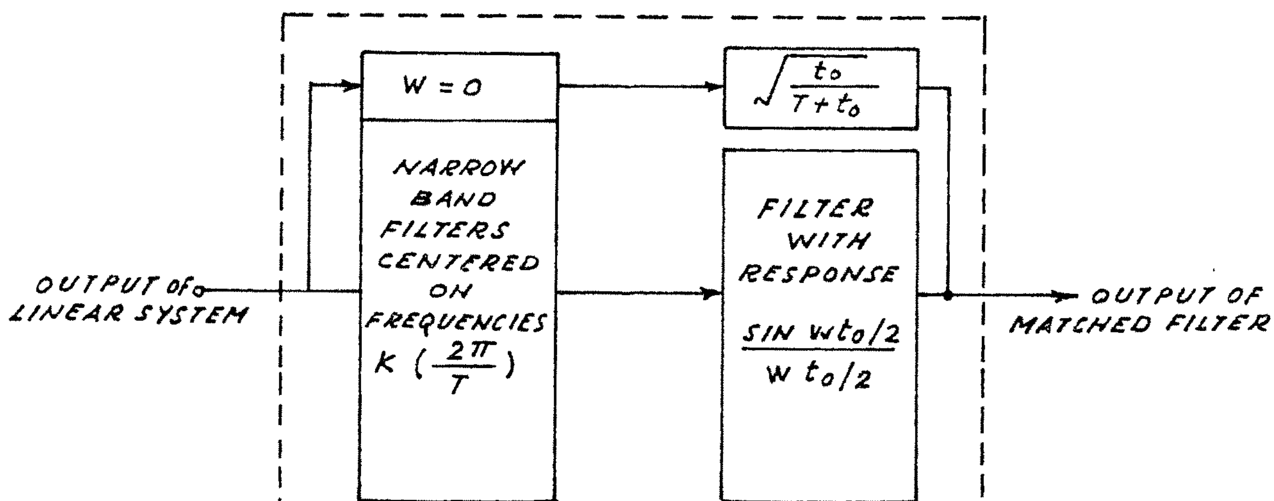

FIG. 2·13 REALIZATION OF MATCHED-FILTER OF FIG. 2·12 of the linear system. Thus, in this way too, the impulse response of a linear system may be estimated. The advantages of this scheme are its high simplicity and possibility to realize the continuous version of the system weighting function, $h(t)$.

2.3.(C.2) : Synthesis of the matched-filter

From eqn. (2.39), the transfer function of the matched filter may be expressed as -

$$G(w) = e^{-jw\tau_o} \cdot X(w) \qquad \ldots \qquad \ldots \qquad (2.41)$$

where $X(w)$ is the Fourier-transform of $x(t)$.

Using Wiener-Khinchine relations, it is shown that the power-spectral density of the pseudorandom signal is -

$$\phi_{xx}(w) = 2\pi \frac{(at_o)^2}{T^2} \left[- \frac{T}{t_o} \delta(w) + \sum_{k=-\infty}^{\infty} \frac{(T + t_o)}{t_o} \cdot \left(\frac{\sin k\pi (t_o / T)}{k\pi (t_o / T)} \right)^2 \delta\left(w - \frac{2k\pi}{T} \right) \right] \quad .. \quad (2.42)$$

where k is an integer.

Eqn. (2. 42) essentially implies that prbs acts as a band-limited white-noise, and if the band width B, of the system under test is such that -

$$B < \frac{1}{4\pi t_o} \qquad \ldots \qquad \ldots \qquad \ldots \qquad (2.43)$$

then, for all practical purposes, the periodic prbs is a good approximation to white noise.

Now, using the relation between the power spectral density $\phi_{xx}(w)$ and the spectral density $X(w)$, the square of the modulus of the filter transfer function becomes :

$$|G(w)|^2 = 4\pi^2 \left(\frac{at_o}{T} \right)^2 .$$

$$\left[-\frac{T}{t_o} \delta(w) + \left(\frac{T + t_o}{t_o} \right) \sum_{k=-\infty}^{\infty} \left(\frac{\sin \pi k (t_o / T)}{k\pi (t_o / T)} \right)^2 \delta\left(W - \frac{2k\pi}{T} \right) \right] \quad \ldots \quad (2.44)$$

It has been shown that it is possible to realize the transfer function, essentially, by means of required number of parallel connected narrow-band filters with centre frequencies $K(\frac{2\pi}{T})$, followed by a filter with the amplitude response :

$$\frac{\sin(w\, t_o / 2)}{(wt_o / 2)}$$

and a low pass-filter with a fixed attenuation. This situation is shown in Fig. (2.13).

From eqn. (2.44), it is clear that it is enough to approximate a bandwidth equal with a part of the order of $10^{-1}(2\pi / t_o)$ of the function :

$$\frac{\sin(wt_o / 2)}{(wt_o / 2)} ,$$

which may contain a number of $(1, \ldots, 5).10^1$ narrow band filters with centre frequencies $K(2\pi / T)$.

2.3.(C.3) : Remarks

This technique essentially involves the design of a matched-filter to the image of the test-signal to arrive at the input-output crosscorrelation of the linear system.

The method is remarkably simple and yields a continuous variation of the system impulse response as a function of time. Hence, the method is well-suited for linear systems which satisfy the following requirements :

T_{min} of the system $\gg$ the basic interval of the input prbs,

T_{max} of the system $\ll$ the period of the input prbs,

which permit the use of impulse approximation to the autocorrelation of the input signal (prbs).

However, this scheme has the following disadvantages :

(i) Proper adjustment of centre frequencies of the narrow-band filters is essential,

(ii) use of too many narrow-band filters (to obtain accurate results) may prove the scheme to be uneconomical, and

(iii) the effect of output drift on the matched filter response need to be considered. Further, the sensitivity of the filter to wide-band noise is also of some importance.

2.3.(D) : Conclusions of the comparative study of the existing correlation schemes.

Remarks on the correlation schemes 1, 2 and 3 are already stated at the end of sub-sections 2.3 (A, B and C). In view of this, the following may now be summarised:

The test-signal in the scheme 1, namely a simple amplitude modulated random-noise, has the great merit that its autocorrelation function is a delta function. And, the measured crosscorrelation function need not be corrected for bias. But, the variance of the correlator output using a random signal will be certainly more than that using a deterministic test perturbation.

Test signal in the scheme 2, namely the pseudorandom binary signal is easy to generate as well as to transmit, besides giving reasonably accurate results even when the system output is highly contaminated with noise, by using a proper drift correction scheme. However, at times, there is need for double signal generators or special command equipment in such a scheme. Moreover, the impulse response ordinates for discrete values of τ only are realized, but not a continuous variation.

The matched filter method of scheme 3 is simple and gives a continuous variation of system impulse response, but needs too many precise narrow band filters which may react adversely to wide-band and band-limited system noise.

In view of the merits and demerits of the existing correlation schemes as stated above, it is clear that a satisfactory method of dealing with bias, wide-band and band limited noise effects has not yet been advanced. And in what follows, a new correlation scheme for impulse response measurements is described, which is superior in some respects to those already in use. The proposed scheme gives good measurements of the dynamics, whether the output drift is represented by (i) Truncated Fourier series or (ii) a polynomial of degree P(say) over the measurement interval.

2.4 DYNAMIC ANALYSIS BY MEANS OF SIMULTANEOUS CROSS-CORRELATION WITH A PSEUDORANDOM BINARY TEST SIGNAL AND ITS PHASE-INVERSE

A new correlation scheme for linear system impulse response measurements is depicted in Fig. (2.14). The input $x(t)$ is a pseudorandom binary perturbation, with levels $\pm$ 1, basic interval t_o, and period $T = Nt_o$ where $N = 2^n - 1$, n being a chosen integer. The signal $x^*(t)$ is the phase-inverse of the signal $x(t)$. And, the signals $y(t)$, $n_1(t)$ and $z(t)$ represent, as before, the system response due to test signal $x(t)$, the total disturbance due to wide-band and band-limited noise, and the measurable system output respectively. In this scheme, the

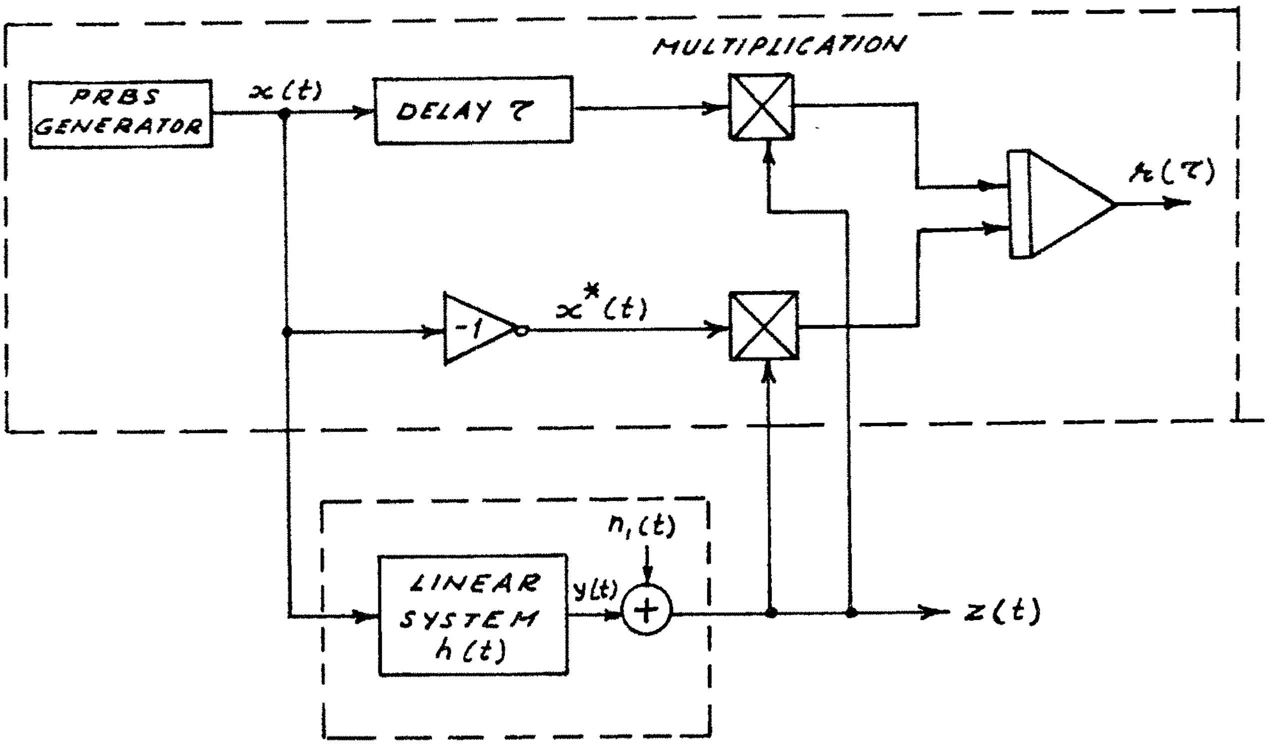

FIG. 2.14 NEW SCHEME WITH SIMULTANEOUS CROSS-CORRELATION WITH A PRBS AND ITS PHASE-INVERSE

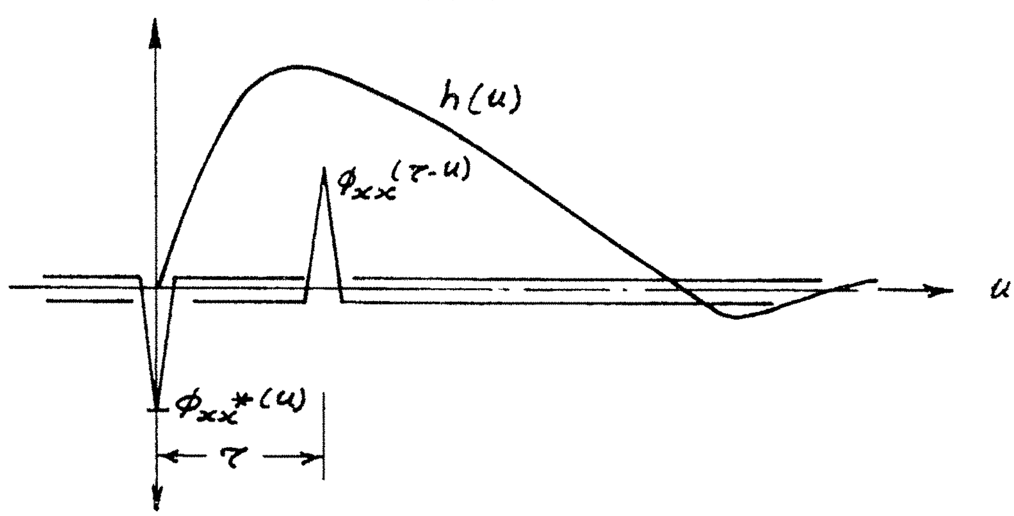

FIG. 2.15 INTERPRETATION OF CONVOLUTION INTEGRALS IN EQN. (2.49)

measurable output $z(t)$ is correlated simultaneously with the pseudorandom binary input signal $x(t - \tau)$ and its phase inverse $x^*(t)$ (for $\tau = 0$). The scheme hence derives the name 'Simultaneous Crosscorrelation method.'

From Fig.(2.14), performing crosscorrelation over m periods of $x(t)$, the integrator output is given by :

$$r(\tau) = \frac{1}{mT}\int_0^{mT} x(t - \tau).z(t)\, dt + \frac{1}{mT}\int_0^{mT} x(t).z(t)dt \quad \ldots \quad (2.45)$$

Substituting from $z(t)$ from eqn.(2.1) and using the convolution integral in eqn. (2.2) for $y(t)$, the following equation may be obtained for $r(\tau)$:

$$r(\tau) = \int_0^{T_s} h(u).\phi_{xx}(\tau - u)du + \phi^*_{xn_1}(\tau) + \int_0^{T_s} h(u)\ \phi_{xx^*}(u)du + \phi^*_{xn_1}(0) \ldots (2.46)$$

Since the signal $x^*(t)$ is the inverse of the signal $x(t)$,

i.e. $\quad x^*(t) = - x(t) \quad \ldots \quad \ldots \quad (2.47)$

the equation (2.46) may be rewritten in the form :

$$r(\tau) = \int_0^{T_s} h(u).\phi_{xx}(\tau - u)\ du - \int_0^{T_s} h(u).\phi_{xx}(u)\ du + \phi^*_{xn_1}(\tau) - \phi^*_{xn_1}(0) \quad \ldots \quad (2.48)$$

And it is of present interest to examine this correlation scheme, described by the above eqn. (2.48) when :

Case (1) : drift signal d(t) is negligible,

Case (2) : drift signal may be represented as a truncated 'Fourier Series' over the measurement interval mT, and

Case (3) : drift signal may be represented as a polynomial of degree P (say) as : $d(t) = \sum_{i=0}^{P} k_i t^i$ over the measurement interval.

2.4.1 Case (1) : Drift signal d(t) is negligible : In case the drift-effects can be ignored, eqn. (2.48) may be expressed as :

$$r(\tau) = \int_0^{T_s} h(u).\phi_{xx}(\tau - u)du - \int_0^{T_s} h(u).\phi_{xx}(u)\ du \quad ..(2.49)$$

whereby the effects of wide-band noise n(t) in the noise-term $n_1(t)$ is assumed to be compensated by performing correlation over the time-interval mT.

The interpretation of the convolutional integrals in the eqn. (2.49), with x(t) referring to input prbs, is shown graphically in Fig. (2.15). Here, as the autocorrelation function $\phi_{xx}(\tau)$ is an even function of its argument τ, the rotation about $\tau = 0$ axis in the first term makes no difference.

Substituting for the values of the autocorrelation function from eqn. (2.30) in the above equation (2.49), the correlator output $r(\tau)$ becomes :

$$r(\tau) = \left[\left(\frac{N+1}{N} \right) t_o \, h(\tau) - \frac{1}{N} \int_o^{T_s} h(u) \, du - \left(\frac{N+1}{N} \right) t_o \, h(0) + \frac{1}{N} \int_o^{T_s} h(u) \, du \right]$$

$$= \left(\frac{N+1}{N} \right) t_o \left[h(\tau) - h(0) \right] \quad \ldots \quad (2.50)$$

In practical systems, $h(0)$ is usually zero. In such a case,

$$r(\tau) = \left(\frac{N+1}{N} \right) t_o \, h(\tau) \quad \ldots \quad \ldots \quad (2.51)$$

The correlator output thus directly yields accurate estimate of the system impulse response.

2.4.2 Case 2 : Truncated 'Fourier - Series' representation of the drift signal : Consider the output drift signal $d(t)$ over m periods of the test signal to be expanded into a truncated Fourier Series :

$$d(t) = \frac{\alpha_0}{2} + \sum_{i=1}^{m} \left(\alpha_i \cos i w_m t + \beta_i \sin i w_m t \right) + \epsilon(t) \, , \; 0 < t < mT \quad \ldots \quad (2.52)$$

where $w_m = \frac{2\pi}{mT}$ is the fundamental frequency and $\epsilon(t)$ is the residual. The expansion is valid over one fundamental period $0 < t < mT$.

Now, consider the correlator output $r(\tau)$ given in eqn. (2.48), which is rewritten below for convenience :

$$r(\tau) = \int_0^{T_s} h(u).\phi_{xx}(\tau - u)\, du - \int_0^{T_s} h(u).\phi_{xx}(u)\, du$$

$$+ \phi^*_{xn_1}(\tau) - \phi^*_{xn_1}(0)$$

Assuming that the effects of wide-band noise can be made almost negligible by integration over the time-interval mT, $r(\tau)$ may be simplified with the help of the equations (2.23) and (2.51) to the following form :

$$r(\tau) = Kh(\tau) + \frac{1}{mT}\int_0^{mT} x(t - \tau)\, d(t)\, dt - \frac{1}{mT}\int_0^{mT} x(t)\, d(t)\, dt \qquad \ldots \quad (2.53)$$

$$= Kh(\tau) + \frac{1}{mT}\int_0^{T} x(t - \tau)\left\{\sum_{j=0}^{m-1} d(t + iT)\right\} dt$$

$$- \frac{1}{mT}\int_0^{T} x(t)\left\{\sum_{j=0}^{m-1} d(t + iT)\right\} dt \qquad \ldots \quad (2.54)$$

By definition of $d(t)$ stated in eqn. (2.52), the curly-bracketed part of the second - and third - terms in eqn. (2.54) can easily be shown to be identically zero for $n \neq 0$, making use of the following trigonometric identities -

$$\sum_{i=0}^{m-1} \cos iw_m(t + jT) \equiv 0, \text{ for all } t \text{ and } i \text{ except for } i = 0 \qquad \ldots \quad (2.55)$$

$$\sum_{j=0}^{m-1} \sin iw_m(t + jT) \equiv 0, \text{ for all } t \text{ and } i \qquad \ldots \quad (2.56)$$

Furthermore, the constant drift component $a_o/2$ is automatically eliminated due to the presence of negative sign before the third term in eqn. (2.54).

The correlator output may then be written as :

$$r(\tau) = Kh(\tau) , \quad K = \left(\frac{N+1}{N}\right) t_o = \text{constant} \qquad \ldots \quad (2.57)$$

Thus, the effects of constant and alternating components of the output drift signal can be exactly compensated upto the (m - 1)th harmonic making use of simultaneous crosscorrelation over m periods of the test signal.

2.4.3 Case (3.A) : Polynomial representation of the drift signal (assumption is that drift coefficients can be measured)

Now, consider the case when the drift signal d(t) over m periods of the test signal w may be approximated by a polynomial given by -

$$d(t) = k_o + k_1 t + k_2 t^2 + \ldots\ldots + k_i t^i + \ldots + k_p t^p$$

$$= \sum_{i=0}^{p} k_i t^i \qquad \ldots \qquad \ldots \qquad \ldots \qquad (2.58)$$

As before, assuming that wide-band noise effects to be compensated by crosscorrelation over time mT, (although the value of m seldom exceeds 2 due to time involved), and substituting for d(t) in eqn. (2.48) , the following equation can be written for the correlator output :

$$r(\tau) = K\,h(\tau) + \frac{1}{mT}\int_{0}^{mT} x(t-\tau)\left[\sum_{i=0}^{P} k_i t^i\right] dt$$

$$-\frac{1}{mT}\int_{0}^{mT} x(t)\left[\sum_{i=0}^{P} k_i t^i\right] dt \quad \ldots \quad (2.59)$$

From the above equation, it is obvious that the drift-error will be different for each value of τ. To avoid this situation, Davies (1968) suggested that we may consider the time-advanced-output instead of time-delayed input. Making use of this concept, the new correlation scheme of Fig. (2.14) is modified and is shown in Fig. (2.16).

The output of the correlator of Fig. (2.16) will be :

$$r(\tau) = \frac{1}{mT}\int_{0}^{mT} x(t).z(t+\tau)\,dt - \frac{1}{mT}\int_{0}^{mT} x(t).z(t)\,dt \quad \ldots \quad (2.60)$$

Making use of eqn. (2.59), the above equation becomes :

$$r(\tau) = K\,h(\tau) + \frac{1}{mT}\int_{0}^{mT} x(t).\sum_{i=0}^{P} k_i (t+\tau)^i\,dt$$

$$-\frac{1}{mT}\int_{0}^{mT} x(t)\cdot\sum_{i=0}^{P} k_i t^i \quad dt \quad \ldots \quad (2.61)$$

Thus,

$$r(\tau) = K\,h(\tau) + \sum_{i=0}^{P-1} M_i F_i(k_j\,,\ \tau), \quad \ldots \quad (2.62)$$

$$0 < j < P.$$

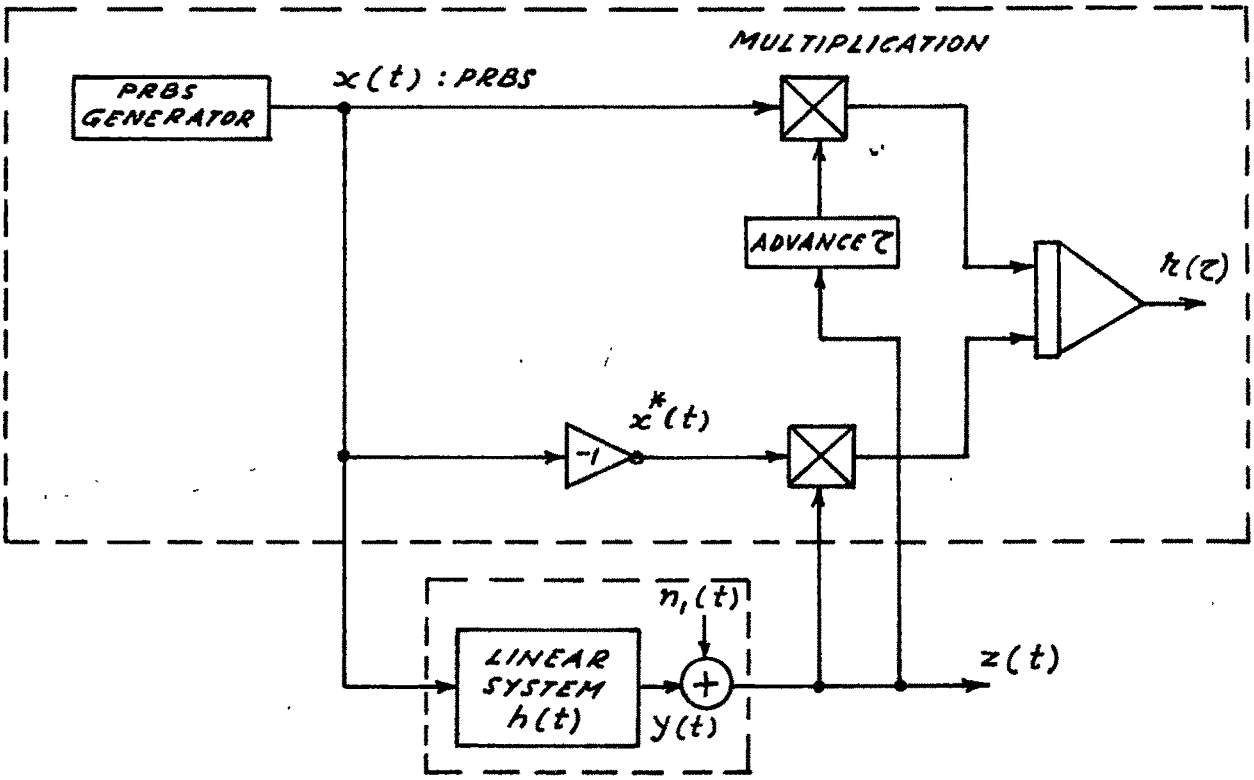

FIG. 2·16 SIMULTANEOUS CROSS-CORRELATION SCHEME EMPLOYING DAVICS'S DRIFT CORRECTION PRINCIPLE

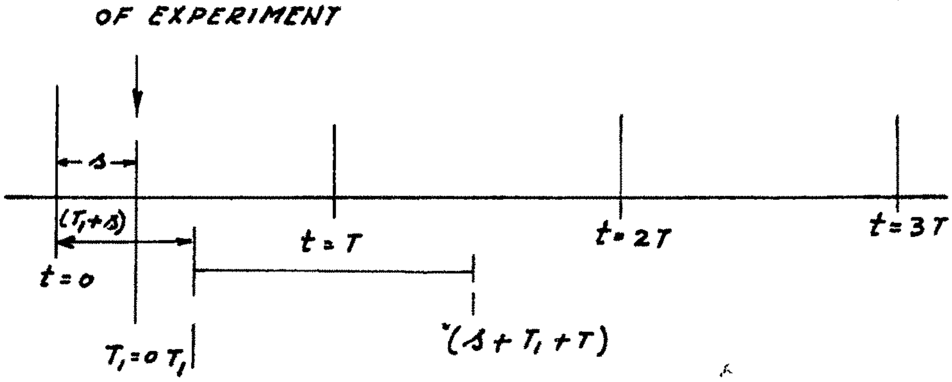

FIG. 2·17 DEFINITION OF TIME-ORIGIN AND OBSERVATION PERIOD IN DAVICS'S DRIFT-CORRECTION METHOD.

where M_1, M_2, ... M_{P-1} are the normalized time moments of the input prbs with the zeroth time moment $M_o = 1/N$ and where $F_i(k_j, \tau)$ is a function of the drift parameters and τ.

For instance, in case d(t) is approximated as :

$$d(t) = \sum_{i=0}^{4} k_i t^i \qquad \ldots \qquad \ldots \qquad (2.63)$$

then,

$$F_o(k_j, \tau) = (k_1 + k_2\tau^2 + k_3\tau^3 + k_4\tau^4)$$

$$F_1(k_j, \tau) = (2k_2 + 3k_3\tau^2 + 4k_4\tau^3)$$

$$F_2(k_j, \tau) = (3k_3 + 6k_4\tau^2)$$

and $$F_3(k_j, \tau) = (4k_4) \qquad \ldots. \qquad \ldots. \qquad (2.64)$$

Since the normalized time moments M_o, M_1, ..., M_{P-1} can be obtained by off-line computation for a fixed m and input prbs x(t), the drift coefficients k_o, k_1, ..., k_P can be calculated by measuring the correlator output at $P + 1$ values of $\tau > T_s$, i.e. after the system has settled. Once the drift coefficients are known, due corrections can then be applied to arrive at accurate estimates of the system impulse response.

However, it may not always be feasiable to measure the drift parameters. In such a case, effects of the polynomial drift can be reduced to zero, by generating a modified form of the test signal to be applied to the crosscorrelator has originated by Davies and Douce (1967). This idea, when applied to the present 'simultaneous crosscorrelation scheme' yields favourable results than those given by Davies and Douce, as shown below -

2.4.3. Case (3.B) : Polynomial representation of the drift signal - Application of the drift correction principle of Davies and Douce (1967) to the present 'Simultaneous Crosscorrelation Scheme.'

It is seen that the impulse response can be found by performing the crosscorrelation over any time interval of duration T. If many observations are made, their average will be a more accurate estimate of the impulse response. In particular, different weights may be given to different determinations.

Let it be assumed that the output measurements in the correlation scheme of Fig. (2.14) start at time s. Then, from Fig. (2.17), using the input prbs for $(s + T_1) \leq t \leq (s + T_1 + \tau)$ the measured impulse response can be obtained from -

$$r(\tau) = \frac{1}{T}\int_0^T x(t + T_1 + s).z(t + T_1 + s + \tau)dt - \frac{1}{T}\int_0^T x(t + T_1 + s).z(t + T_1 + s)dt \qquad \ldots (2.65)$$

In presence of drift, the correlator output depends on the starting time and to emphasize this, the impulse response obtained from eqn. (2.65) may be denoted by $h(T_1 + s)$.

Now, consider the effect of adding determinations together with different weights $w(T_1)$ when T_1 varies from 0 to T.

The new estimate $h'(\tau)$ is given by :

$$h'(\tau) = \frac{\int_0^T w(T_1).\ h(T_1 + s)\,dT_1}{\int_0^T w(T_1)\ dT_1} \qquad \ldots \qquad \ldots \qquad (2.66)$$

The weighting function $w(T_1)$ can be chosen to minimize the drift effects on the new estimate $h(\tau)$.

In view of the statements implied by eqns. (2.65) and (2.66), the correlator output $r(\tau)$ is free from the influence of drift provided that

$$\int_0^T w(T_1) \left[\int_0^T d(t + T_1 + s + c)\ .\ x(t + T_1 + s)\ dt - \int_0^T d(t + T_1 + s).\ x(t + T_1 + s)\ dt \right] dT_1 = 0$$

$$\ldots \qquad (2.67)$$

This can be satisfied for polynomial drift signals by assuming that $w(T_1)$ is a polynomial in T_1 , the coefficients of this polynomial being dependent on the order of the drift polynomial, but independent of the magnitude of the coefficients of the drift terms as discussed below -

If the drift is given by -

$$d(t) = k_o + k_1 t \quad \ldots \quad \ldots \quad (2.68)$$

then, eqn. (2.67) becomes :

$$\int_0^T w(T_1) . \left[\int_0^T k_o + k_1(t + T_1 + s + \tau) \, . \, x(t + T_1 + s) \, dt - \int_0^T k_o + k_1(t + T_1 + s) \, . \, x(t + T_1 + s) \, dt \right] dT_1 = 0$$

$$\ldots \quad (2.69)$$

which is satisfied for all $w(T_1)$ if $x(t)$ has zero mean value. As there are $(N + 1)/2$ number of +1s and $(N - 1)/2$ number of -1s in a prbs of period N digits, zero mean value can be realized by adjusting the levels of the sequence such that -

+1 corresponds to the level $(N - 1)/N$, and

-1 corresponds to the level $-(N + 1)/N$.

For the drift d(t) given by -

$$d(t) = k_o + k_1 t + k_2 t^2, \ldots \quad \ldots \quad \ldots \quad (2.70)$$

Further to eqn. (2.69) we must have :

$$\int_0^T w(T_1) \left[\int_0^T (t + s + T_1 + \frac{\tau^2}{2}) . \, x(t + T_1 + s) \, dt \right] dT_1 = 0$$

which reduces to :

$$\int_0^T w(T_1) \left[\int_0^T t.\ x(t + T_1 + s)\ dt \right] dT_1 = 0$$

or

$$\int_0^T w(T_1).\ m_1(T_1 + s)\ dT_1 = 0 \qquad \ldots \qquad \ldots \qquad (2.71)$$

where $m_1(T_1 + s)$ is the moment of the input prbs defined as :

$$m_j(T_1 + s) = \frac{1}{T} \int_0^T t^j.\ x(t + T_1 + s)\ dt \qquad \ldots \qquad (2.72)$$

Eqation (2.71) can be satisfied if the weighting function for $w(T_1)$ = constant, because the average value of any moment of the input prbs over all T_1 equals zero, following the requirement that it has zero mean value.

Similarly, for $d(t) = (k_o + k_1 t + k_2 t^2 + k_3 t^3)$, the requirement in addition to those given by (2.67) and (2.71) is :

$$\int_0^T w(T_1) . \left[m_2(T_1 + s) + 2T_1 m_1(T_1 + s) \right] dT_1 = 0 \ldots (2.73)$$

Now suppose that -

$$w(T_1) = 1 + AT_1 + BT_1^2 \qquad \ldots \qquad \ldots \qquad (2.74)$$

Then, eqns. (2.71) and (2.73) become :

$$\int_0^T (AT_1 + BT_1^2).\ m_1(T_1 + s)\ dT_1 = 0 \qquad \ldots \qquad (2.75)$$

and

$$\int_0^T \left[(AT_1 + BT_1^2) . \, m_2(T_1 + s) \right.$$

$$\left. + \, (2T_1 + 2AT_1^2 + 2BT_1^3) . \, m_1(T_1 + s) \right] dT_1 = 0$$

... (2.76B)

With the help of the definition for moment of a moment function given by :

$$M_{ij}(s) \; : \; \frac{1}{T} \int_0^T T_1^i . m_j(T_1 + s) \, dT_1 \quad \ldots \quad (2.77)$$

equations (2.76) become :

$$A \, M_{11}(s) + B \, M_{21}(s) = 0 \quad \ldots \quad (2.78A)$$

$$A \, M_{12}(s) + 2M_{21}(s) + B \, M_{22}(s) + 2 \, M_{31}(s) = -2 \, M_{11}(s)$$

... (2.78B)

Thus, the coefficients A and B in the weighting function $w(T_1)$ depend only on the starting time s of the experiment.

Solving the two sets of simultaneous equations (eqns. 2.78A and 2.78B) give A and B to eliminate the drift given by :

$$d(t) = (k_o + k_1 t + k_2 t^2 + k_3 t^3) \quad \ldots \quad (2.79)$$

It may be noted here that in the conventional scheme of Davies and Douce, to eliminate the drift described by eqn.(2.79) they formulated simultaneous eqns. (like those of eqns. 278A and 2.78B) which involve three unknowns, (A, B, and C) as against two (A and B only) in the present simultaneous crosscorrelation scheme. In other words, by the present scheme, with quadratic weighting function $w(T_1)$, it is possible to eliminate effects of the drift signal d(t) given by eqn. (2.82); whereas Davies's scheme needs a weighting function $w(T_1)$ of the form :

$$w(T_1) = 1 + AT_1 + BT_1^{\,2} + CT_1^{\,3}$$

to bring out the same results.

Thus, by means of the simultaneous crosscorrelation of a linear time-invariant system output with the pseudorandom binary test signal and its phase inverse (Figs. 2.14 and 2.16), the impulse response can be accurately estimated.

2.4.4 SOME ADVANTAGES OF THE PROPOSED SIMULTANEOUS CROSS-CORRELATION SCHEME

The proposed scheme of measurement of the impulse response of a linear time invariant system performs simultaneous cross-correlation of the system output with the pseudorandom binary test perturbation and its phase inverse. The scheme has the following advantages :

(i) The scheme eliminates errors in the correlator output due to bias in the autocorrelation function of the input prbs, wide-band noise, and the constant-term in the drift signal.

(ii) With proper choice of the observation interval (mT), the scheme of Fig. (2.14), suited for Fourier-Series expansion representation of the drift signal, exactly compensates a.c. components of the drift signal upto (m - 1)th harmonic.

(iii) The modified scheme shown in Fig. (2.16), suited for polynomial drift, yields the drift constants from the measurements of the correlator output during the normal experimental run.

(iv) The application of Davies et.al.'s method (1967) to the scheme of Fig. (2.16) gives favourable results to determine a weighting function $w(T_1)$ that eliminates the polynomial drift-effects. Davies's drift correction method may, therefore, be used when it is not possible to measure the drift parameters.

(v) It is economical in the sense that no extra equipment is required. And, the additional calculations involved are extremely simple.

So far, the content of the present chapter pertains to the identification problem of a linear time-invariant system. An instructive question is: is the crosscorrelation technique using the pseudorandom input applicable to the identification of a linear time-varying system ? Recent work of Hoffman (1972) says 'Yes' to this question, in principle. Hoffman has shown h that it is possible to track the weighting sequence of a time varying process provided that the rate of variation is sufficiently smooth. Hoffman's method is based on the assumption that the time variation of each term of the weighting sequence can be described over a period of at least $i_m + 1$ times the system settling time by the i_m terms of Taylor - Series expansion. The periodicity of the input is essential to the method. To the author's knowledge, but for this communication of Hoffman, no regorous analytical studies have been made with reference to the time varying systems.

Thus, the deterministic pseudorandom binary signals are well-suited as test perturbations in system identification by means of the crosscorrelation method. The main reasons for the increasing interest in this chain-code testing of an unknown system may be summarized as under :

(i) The chain-code or pseudorandom binary signal has its time autocorrelation function very similar to that of white noise. The unwanted bias in its autocorrelation poses no problem as the d.c. level for the purpose can either be determined or the bias can be eliminated by the method discussed in the present chapter.

(ii) Prbs can be easily generated with the help of a linear feedback shift-register (fsr).

(iii) With proper combination of shift pulse interval and the number of stages of the shift register, it is possible to confine the power of the almost white noise generated by the m-sequence to the interested range of frequencies.

(iv) The statistical properties of the so generated prbs are not altered with environmental changes.

(v) The processing of a prbs is easy and provides simple calculus of the input-output correlation functions in the system identification.

(vi) The use of a prbs leads to an appreciable reduction of the smoothing time (i.e. smaller variance of the resulting information after the chosen interval of correlation).

(vii) Results obtained with a prbs are reasonably precise even in presence of noise.

(viii) A prbs possesses a characteristic phase, called as reference phase, the use of which eliminates estimation errors due to constant, linear, quadratic and cubic drift signals by linear weighting over two periods of the sequence.

(ix) The chain-code method of system testing is well-suited to on-line working, which is an economical proposition to many industrial processes.

(x) By virtue of its periodicity and well-defined properties, any experiment using a prbs can be repeated using exactly the same sequence which is impossible with true white noise.

Thus, the use of a prbs as test input greatly overcomes difficulties associated with the correlation experiments using conventional test perturbations.

2.5 DETERMINATION OF BINARY SHIFT REGISTER CONNECTIONS FOR THE REALIZATION OF DELAYED COPIES OF THE BASIC m-SEQUENCE

Referring to the practical considerations of the correlation method of identifying an unknown system using m-sequence test signals, it is seen that this method of system identification consists of adding the m-sequence (sometimes called as the basic sequence or reference sequence) to the existing input (or set-point adjustment), and crosscorrelating the corresponding system output with a delayed version of the reference m-sequence. An estimate of the system impulse response $h(\tau)$ at time $t = \tau$ is then obtained by measuring the correlator output corresponding to the delay τ of the delayed m-sequence with respect to the reference sequence.

Thus, to obtain the required ordinates of the impulse response $h(\tau)$, the corresponding time-delayed replicas of the reference m-sequence must be generated. To this end, some interesting procedures have been documented. (Tsao 1964, Davies 1965, 1968, Ireland and Marshall 1968, Stepleton 1971).

Tsao's method of obtaining the delayed versions uses the shift and add property of the linear m-sequences (this property is discussed in the previous chapter). The technique is very noteworthy, but involves considerable computation. Another method by Davies uses polynomial division which is quite straight forward but weakness of the method is that the whole division is to be performed for each delay to obtain the feedback connections of the sequence generator. The method of Ireland and Marshall involves matrix albegra and related algorithm. Finally the recent method of Stepleton using extended shift register is nothing but an inversion of Davies method. So much so, a satisfactory method of dealing with the delay generation has not yet been developed.

An attempt is, therefore, made in this section to present a simple method to determine the binary shift register connections for the realization of the required delayed copies of the basically generated m-sequence. The method makes use of the familiar concept of generating function associated with the periodic m-sequence (Golomb 1967). As the only calculation involved in the present technique is simple modulo-2 addition, it ~~which~~ can provide quick results.

Before considering how the generating function concept can be used to find the shift register connections for the delay generation, it is necessary to recall the salient features of m-sequence generation using the linear binary shift register and a few of its properties of present interest.

2.5.1 Salient features of m-sequence generation using feedback shift register

As discussed in Chapter 1, binary sequences may be generated by a system which is in its simplest form has the logical arrangement shown in Fig. (2.18). This represents a basic n-stage shift register to which modulo-2 gate has been added. These gates may be connected to various stages of the register, the outputs of the stages are fed to the gates and the outputs of the gates are fed back to some other stage of the register, so that a single (or multiple) closed-loop is formed. On the application of a shift pulse the digits in the shift register move one stage to the right, and the modulo-2 sum-digit enters the first stage. If a train of shift pulses is applied, the process continues and the contents of the register progress in some cyclic manner. The output which may be taken from any of the stages is a linear binary sequence.

The binary feedback shift register can be described by the following equation :

$$[F(D)]\, C = \left(\sum_{i=0}^{n} a_i D^i\right) C = 0 , \quad a_0 = a_n = 1 . \qquad (2.80)$$

whereby the coefficients a_i are all 1's and 0's, and D stands for a time delay equal to one digit period or shift pulse period and C represents the generated cyclic sequence.

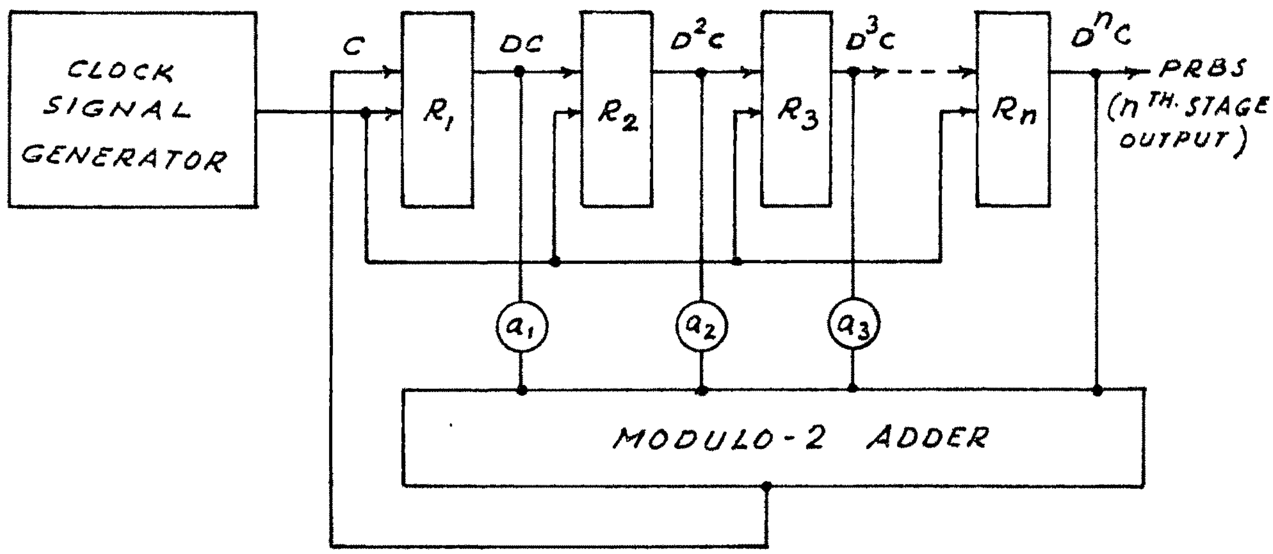

FIG. 2·18 AUTONOMOUS LINEAR n-STAGE BINARY FEED BACK SHIFT REGISTER GENERATING P.R.B.S.

The expression within the brackets of the above equation, which is composed of characteristic delays acting on the variable C (logical output sequence of the modulo-2 adder of Fig. (2.18))is called 'Characteristic Delay Polynomial' and is denoted by F(D). Thus F(D) is given by -

$$F(D) = \sum_{i=0}^{n} a_i D^i \ , \quad \text{with } a_o = a_n = 1 \ ,$$

since the shift register contains n stages and is binary in nature. In case, this polynomial is a primitive modulo-2, the feedback shift register generates m - or maximum length sequence with period equal to $2^m - 1$ digits. The current section is concerned with only binary m-sequences and their delay generation.

2.5.2 Four properties of binary m-sequences

Many useful properties of the binary delay polynomials and their associated cyclic sequences have been stated in Chapter 1. However, four properties of m-sequences relevant to the present aspect of delay generation are given below -

(i) The m-sequence generated by an n-stage 2-level feedback shift register is cyclic with a period T equal to 2^n-1 shift pulse intervals.

(ii) There are $2^n - 1$ phase-distinct m-sequences corresponding to each of the $2^n - 1$ possible nonzero initial states of the feedback shift register.

(iii) If an m-sequence C is operated by any polynomial Q(D) in D (with its coefficients over the modular field of integers 0 and 1), a phase-shifted version of the sequence C results.

(iv) Every m-sequence C represented by the terms $(c_o, c_1, \ldots, c_r, \ldots)$ satisfies a linear recurrence relation of the form :

$$c_r = \sum_{i=1}^{n} a_i\, c_{r-i} \quad \ldots \text{(mod. 2 addition)} \ldots \qquad (2.81)$$

where a_i is 1 or 0 depending on whether the ith stage of the corresponding feedback shift register is involved in the feedback path or otherwise.

2.5.3 Use of 'Generating-Function' to determine feedback shift register connections for delayed m-sequences

A close look at the methods of studying linear recurring sequences reveals that a generating function G(x) may be associated with the 2-level m-sequence $C = \{c_r\}$ as under :

$$G(x) = \sum_{r=0}^{\infty} c_r\, x^r \quad \ldots \quad \ldots \qquad (2.82)$$

Assuming the initial state of the corresponding shift register as :

$c_{-1}, c_{-2}, \ldots c_{-n}$,

and making use of the linear recurrence stated in eqn.(2.81), the generating function G(x) can be shown to be :

$$G(x) = \left(\frac{\sum_{i=0}^{n-1} b_i D^i}{F(D)} \right) x \quad \ldots \quad \ldots \quad \ldots \qquad (2.83)$$

where $F(D) = \sum_{i=0}^{n} a_i D^i$, $a_o = 1$ and $b_j = \sum_{i=1}^{n-j} a_{i+j} c_{-i}$, $0 \leq j \leq n-1$.
with $I = D^o$ as the identift operator.

Eqn.(2.83) may perhaps be better understood, if it is in the form :

$$G(x) = \left[\left\{ \sum_{i=0}^{n-1} b_i \, D^i \right\} . \quad \frac{1}{F(D)} \right] x \quad \ldots \qquad \ldots . \qquad (2.84)$$

It is apparent from eqn. (2.84) that there exists a duality between the characteristic delay polynomial F(D) and the generating function G(x) (i.e. the m-sequence under consideration). Under the circumstances, the term within the curly-brackets of the above equation may be made use of to determine the two-level shift register connections necessary for the realization of delayed versions of the m-sequence as discussed below :

With the initial state of the shift register given by :

$c_{-1} = c_{-2} = \ldots\ldots\ldots = c_{-n+1} = 0$ and $c_{-n} = 1$,

the generating function G(x) in eqn. (2.84) becomes :

$$G(x) = \left[\frac{a_n = 1}{F(D)} \right] x \qquad \text{(since the system is binary and of degree n)} \quad .. \ (2.85)$$

Now, if the m-sequence described by G(x) in eqn. (2.85) is the reference sequence C, then by property (iii) of Sec.2.5.2, the m-sequences described by G(x) in eqn. (2.84) are nearly the delayed version of the reference sequence C. This means that the term within the curly brackets of eqn. (2.84) acts as a ' Delay-Operator' on the reference m-sequence C.

This 'Delay Operator' evidently depends on the initial state of the fsr and thus yields a distinct delayed version of the m-sequence C with each of the $(2^n - 1)$ distinct nonzero initial states (Property ii of Sec. 2.5.2).

Therefore, if a relationship between the m-sequence described by the generating function G(x) and the initial state of the fsr (feedback shift register) is established, then, by merely substituting the initial conditions corresponding to a specified delayed sequence, the required 'Delay Operator' or the 'Shift Register Connections' for the generation of the specified delayed version can be readily obtained. The relationship is derived below :

Relationship between the m-sequence described by the generating function G(x) of eqn. (2.84) and the corresponding initial state of the shift register (Fig. 2.18)

The generating function G(x) in eqn. (2.84) may be written as a ratio of two polynomials as under :

$$G(x) = \sum_{i=0}^{n-1} b_i D^i \cdot \frac{1}{F(D)} \quad x \ldots$$

$$= \frac{F_1(x)}{F(x)} \, , \quad \text{(say)}$$

Here, $F_1(x)$ is of degree $n - 1$ and $F(x)$ is of degree n. Hence, we may write :

$$G(x) = d_o + d_1 x + d_2 x^2 + \ldots + d_r x^r + \ldots \quad \ldots \quad (2.86)$$

Considering only one cycle of the m-sequence of period N digits, we get for the generating function G(x) of eqn. (2.86) as :

$$G_1(x) = (d_o + d_1x + d_2x^2 + \ldots + d_{N-1}x^{N-1}) \qquad \ldots \quad (2.87)$$

Expanding eqn. (2.84), we see that -

$d_o = b_o$

$d_1 = b_1 + a_1d_o$

$d_2 = b_2 + a_1d_1 + a_2d_o$

....

....

and $d_r = b_r + \sum_{i=0}^{r-1} a_{r-i}\, d_i$ where $d_o = b_o$ ... (2.88)

From the above equation -

$d_o = b_o$

But, from eqns. (2.83) and (2.81), we obtain -

$$b_o = \sum_{i=1}^{n} a_i c_{-i}$$

$$= c_o$$

Hence, $d_o = c_o$ ... ... (2.89)

Further,

$d_1 = b_1 + a_1d_o$

Substituting for b_1 and d_o from eqns. (2.83) and (2.89) and using eqn. (2.81) -

$$d_1 = \sum_{i=1}^{n-1} a_{i+1}\, c_{-i} + a_1c_o = \sum_{i=1}^{n} k_i c_{1-i} = c_1$$

Proceeding in the similar manner, it can be proved that -

$$d_r = c_r \qquad \ldots \qquad \ldots \qquad (2.90)$$

Now, since the period of m-sequence $T = N(t_o)$, t_o being the digit period,

$$d_{N-i} = d_{-i} \qquad \ldots \qquad \ldots \qquad (2.91)$$

Making use of eqn. (2.90),

$$d_{N-i} = d_{-i} = c_{-i} \qquad \ldots \qquad (2.92)$$

Thus,

$$d_{N-1} = c_{-1},\ d_{N-2} = c_{-2},\ \ldots.\ d_{N-n} = c_{-n}\ .$$

In view of eqns. (2.87) and (2.92), it is clear that the initial state of the feedback shift register corresponding to an m-sequence described by the generating function G(x) of eqn. (2.84) appears at the end of the sequence in its reverse order.

The above results may also be obtained by considering the succession of states in the shift register of Fig.(2.18) as follows.

Succession of states in the shift register of Fig. (2.18)

Description of State	Output sequence				
	C	DC	D^2C	....	D^nC
Initial state	c_o	c_{-1}	c_{-2}	....	c_{-n}
Second state	c_1	c_o	c_{-1}	....	c_{-n+1}
Third state	c_2	c_1	c_o	....	c_{-n+2}
....	....	....	....	....	....
....	....	....	....	....	....
(N - 1)th state	c_{N-1}	c_{N-2}	c_{N-3}	....	c_{N-n-1}
Nth state	c_N	c_{N-1}	c_{N-2}	....	c_{N-n}

Since the period of the m-sequence equals $T = Nt_o$, it may be written that :

$$
\begin{aligned}
c_N &= c_o \\
c_{N-1} &= c_{-1} \\
c_{N-2} &= c_{-2} \\
&\cdots\cdots \\
&\cdots\cdots \\
c_{N-n} &= c_{-n} \qquad \ldots \qquad \ldots. \qquad (2.93)
\end{aligned}
$$

It is clear from the above equation that the initial state (c_{-1} , c_{-2} , ..., c_{-n}) of the shift register appears at the end of the reference m-sequence C (logical output of mod-2 adder of Fig.2.18) in its reverse order.

Hence, by mere observation of the required delayed sequence, the corresponding initial state of the shift register can be noted, substitution of which in the curly-bracketed term of eqn. (2.84) directly yields the shift register connections for the generation of the specified delayed version of the reference m-sequence.

In mathematical terms, the shift register connections for any given delayed version D^qC ($n < q < 2^n - 1$) of the reference m-sequence C are given by :

$$D^qC = \left(\sum_{i=0}^{n-1} b_i D^i \right) C \quad \ldots \quad \ldots \quad (2.94)$$

(curly bracketed terms of eqn. 2.84)

where the coefficients b_i are evaluated with the initial conditions of the shift register corresponding to the delayed version D^qC.

To clear the above ideas, an example is illustrated below:

<u>Example</u> :

The characteristic equation of a feedback binary shift registef is given by :

$$[F(D)]\, C = (I \oplus D^3 \oplus D^4)\, C = 0 \quad \ldots \quad \ldots \quad (2.95)$$

(Modulo-2 addition)

Determine the shift register connections for the delayed version D^9C of the basic m-sequence C.

Comparing the present characteristic eqn. (2.95) with the general eqn. (2.80), we see that :

$n = 4$, and

$a_0 = a_3 = a_4 = 1$, and $a_1 = a_2 = 0$ ... ... (2.96)

Starting with the initial state '0001', the reference m-sequence C (logical output of the mod-2 adder) as described by the generating function G(x) of eqn. (2.85) may be written as :

$$\begin{array}{cccccccccccccccc} C: & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & \underline{1} & \underline{0} & \underline{0} & 0, .. \\ & c_o & c_1 & c_2 & . & . & . & . & . & . & . & . & . & . & . & c_{14} \end{array}$$

Hence,

$$D^9C: 1 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad \underline{0 \quad 1 \quad 1} \quad 0, ..$$

Now, the initial conditions corresponding to the delayed version D^9C are the last four digits of the above sequence taken in the reverse order (From eqn. (2.93)). i.e.

$c_{-1} = 0,\ c_{-2} = 1,\ c_{-3} = 1$ and $c_{-4} = 0$ ... (2.97)

Substituting these initial conditions in the Delay-Operator (i.e. the curly bracketed term of eqn.(2.84) written separately in eqn. (2.94), we obtain -

$$D^9C = (b_o \oplus b_1D \oplus b_2D^2 \oplus b_3D^3)\ C \quad ... \quad (2.98)$$

where

$$b_j = \sum_{i=1}^{3} a_{i+j}\ c_{-i} \quad \text{(mod-2 addition)} \quad ...$$

Evaluating the coefficients 'b_j' with the help of equations (2.96) and (2.97) :

$$b_o = (c_{-3} \oplus c_{-4}) = 1$$

$$b_1 = (c_{-2} \oplus c_{-3}) = 0$$

$$b_2 = (c_{-1} \oplus c_{-2}) = 1 \text{ , and}$$

$$b_3 = (c_{-1}) = 0$$

Thus,

$$D^9C = I \oplus D^2$$

Since

$$I = D^3 \oplus D^4 ,$$

$$D^9C = (D^2 \oplus D^3 \oplus D^4)\ C.$$

<u>Verification</u> :

D^2C :	0	0	1	0	0	1	1	0	1	0	1	1	1	1	0,	....
D^3C :	0	0	0	1	0	0	1	1	0	1	0	1	1	1	1,	....
D^4C :	1	0	0	0	1	0	0	1	1	0	1	0	1	1	1,	....
$D^2 \oplus D^3 \oplus D^4C$:	1	0	1	1	1	1	0	0	0	1	0	0	1	1	0,	....
:	D^9C															

Thus, making use of the familiar generating function concept, it is possible to determine the shift register connectio for providing delayed m-sequences. The method provides quick results as the only calculation involved is simple mod-2 addition.

2.6 SUMMARY

The main conclusions of the present chapter are as follows :

(i) The theory of the crosscorrelation method of linear time invariant system dynamic analysis is explained and the mathematical relationships involved therein are illustrated by graphical means.

(ii) A comparative study of the presently available correlation schemes for the measurement of linear system impulse responses is made using random and pseudorandom test-signals. The results of this comparative study (stated in section 2.3(D)) have indicated that the correlation scheme 2 which employs the pseudorandom test signals is reliable as well as economical. It is also pointed out that here, more confidence in the estimated system dynamics can be established if a simple and efficient drift correction scheme can be advanced.

(iii) To meet with the above purpose, a new crosscorrelation scheme called as 'Simultaneous Crosscorrelation Scheme' is described in which simultaneous crosscorrelation of the system output with the input prbs and its phase inverse is effected to obtain an accurate estimate of the dynamics without incurring errors due to the test signal imperfections and corrupting effects of noise and drift in the system. It is shown that this technique is applicable to situations wherein the drift is represented either as a (i) truncated Fourier-series or (ii) polynomial of required order, over the measurement interval.

(iv) Further, in situations where the drift coefficients cannot be measured, it has been shown that the drift correction principle suggested by Davies and Douce (1967) can be applied to the present simultaneous crosscorrelation scheme for eliminating the effects of polynomial drift signal. In fact, the results of this approach have indicated that the simultaneous cross-correlation scheme is well-suited for this type of drift correction.

(v) Finally, making use of the familiar generating function concept, a simple method is presented to determine the feedback shift register connections for providing the delayed replicas of the basically generated m-sequence (reference m-sequence). It is seen that this method provides quick results as the only calculation involved is simple modulo-2 addition.

The content of the present chapter is concerned with the use of 2-level shift register m-sequences (pseudorandom binary sequences) in the measurement of single input / single output linear system dynamics. However, most systems in practice are nonlinear and multivariable. And in the identification of these complex systems, there have been indications that 'multilevel' pseudorandom sequences are superior in some respects to the orthodox '2-state' sequences.

And in the next chapter, the theory of multi-level autonomous feedback shift register sequences is developed and their generations using binary logical elements is described which is also successfully experimented. The use of multilevel sequences in multivariable and non-linear system identification is discussed in chapters 4 and 5.