#3

THEORY OF NONLINEAR TRANSDUCERS

H. E. SINGLETON

LOAN COPY only

TECHNICAL REPORT NO. 160

AUGUST 12, 1950

RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

The research reported in this document was made possible through support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the Army Signal Corps, the Navy Department (Office of Naval Research) and the Air Force (Air Materiel Command), under Signal Corps Contract No. W36-039-sc-32037, Project No. 102B; Department of the Army Project No. 3-99-10-022.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report No. 160 August 12, 1950

THEORY OF NONLINEAR TRANSDUCERS

H. E. Singleton

Abstract

Transducers are classified according to the manner in which they make use of storage to control their output. The state of a transducer is specified by the position of a point in multi-dimensional space. The system function of the transducer is then determined by attaching to each state point a set of numbers or probabilities characteristic of the output for the corresponding state. The technique of synthesizing nonlinear transducers is considered, and standard forms for synthesizing the various kinds of transducers are obtained. It is shown that a necessary and sufficient condition for an invariant finite state transducer to be capable of synthesis in terms of a finite number of linear elements and rectifiers is that the state-defining regions be bounded by hyperplane surfaces. It is found that any nonlinear system function is determined by certain higher-order autocorrelation functions of the input and crosscorrelations between the input and output.

The problem of optimum design of nonlinear filters is discussed. It is shown that if the probability distributions of noise and signal are gaussian, the optimum mean-square filter is linear. For finite-state filters, the criterion of minimizing the probability of error is employed. The problem is to specify optimum boundaries for the state-defining regions, and specific design equations are obtained. Several examples of filter design are given, including an application to the radar search problem.

THEORY OF NONLINEAR TRANSDUCERS

Introduction

With the publication of Wiener's work, "The Extrapolation, Interpolation, and Smoothing of Stationary Time Series", network theory gained additional purpose and direction, and achieved a new maturity (1). In this classical treatise Wiener applied statistical methods to the problem of selecting optimum linear filter response characteristics. But the book, originally published in 1942 as an N.D.R.C. report, had a greater significance than is explicit in its treatment of linear filters. In its statement of the concept that communication signals be regarded as stationary time series and studied with the tools of mathematical statistics, it laid the groundwork for the development of the science of information theory, and was itself a first example of the working of this powerful new discipline.

Information theory has given precise meaning to the notion of quantity of information (2, 3, 4), and has made it possible to devise ways of measuring the amount of information carried by a signal. The statistical theory has also provided the foundation for studies of the theoretical limitations on the rate of transmission of information in the presence of noise (5, 6, 7). In the further development of communication systems, information theory may be expected to assume the responsibility of judging how well the systems do their jobs. The networks in the systems will be designed in accordance with the demands of information theory to insure the most effective utilization of noisy or limited media for the transmission of information (8).

The same decade that has seen the development of information theory has also witnessed a more general introduction of communication systems which make use of pulse modulation. The use of pulse techniques tends to draw network design emphasis away from linearity and equalization, since by means of nonlinear devices the signal can be reshaped or regenerated when required. Moreover, multi-channel pulse communication systems make use of time division multiplex, mixing and separating channels in the time domain, and avoiding the use of highly selective filters. It is no accident that the development of information theory and the exploitation of pulse modulation should occur together; for it has been shown that the performance of certain quantized pulse communication systems can be made to approach the limiting ideal performance specified by information theory (5, 9, 10).

The use of pulse regeneration in the repeaters and decoders of pulse communication systems implies that the corresponding network theory should include an additional element, namely the electronic switch. Although electronic switches and other nonlinear circuit elements are in common use in transducers employed in pulse communication systems, there has been very little study of such transducers from a general point of view. There has been no published work at all on the optimum design of such transducers.

A principal purpose of the present paper is to develop methods of selecting optimum response characteristics for a certain class of nonlinear transducers. The function of the transducers may be described as "information processing", and may include noise discrimination, filtering, prediction, decoding, etc. Interest is centered chiefly on transducers which are especially useful in the processing of pulse communication signals. An additional purpose of the paper is to develop some of the properties of such transducers, in particular those properties which are related to the statistical behavior of the transducers. In this connection it should be mentioned that Wiener has recently suggested a method for making use of the statistical response to random noise of an unknown nonlinear transducer to obtain a complete mathematical characterization of the device. This method will be described in a later section.

Wiener's filter theory provided the solution to the problem of selecting optimum response characteristics for linear transducers operating on prescribed ensembles of noise and signal. Beyond the requirement of invariance in time, the statistics of the ensembles were not restricted, but the transducers were required to be linear. In the present paper, on the other hand, the transducers will be allowed to become nonlinear, while the statistics characterizing the signals will often be restricted. In most of the illustrative examples it will be required that the signal statistics conform to those of one of the quantized pulse communication systems. It is natural that such restrictions be employed, since it is likely that pulse modulation systems will be the first to make use of optimum nonlinear information processing devices. Furthermore, as has been indicated, information theory has shown that certain quantized communication systems are able to approach the limiting rate of ideal systems.

For the development of the optimum linear filter theory there was already available a highly developed theory of linear networks themselves, including necessary and sufficient conditions on transfer functions for physical realizability (11, 12), and a number of synthesis procedures (11, 12, 13, 14). On the other hand, no nonlinear transducer theory corresponding to the theory of linear transducers is available to provide a basis for the study of optimum nonlinear systems. A part of the present paper is accordingly devoted to developing some of the properties of nonlinear transducers. In general, the properties investigated are those which are pertinent to the selection of optimum nonlinear systems, or which depend on or characterize the statistical behavior of nonlinear systems. Principal emphasis of this part of the paper is also placed on transducers which are especially suitable for processing quantized signals, but a few properties of more general transducers are also discussed.

I. Classification of Transducers

1.1 General Remarks

By a transducer is meant a device which operates on one time series (the input) to produce another time series (the output). Since we are interested in the way transducers function as information processing devices, we shall restrict the inputs to be considered to those produced by information sources. The information produced by a source can be measured in terms of a set of probability distribution functions (2), and so from the point of view of information processing, a source is completely described by the set of probability distribution functions which govern its operation. Thus any transducer is always discussed with reference to the information source with which it is associated. There is a certain contrast here with the linear theory, where h(t) affords a complete description of the linear transducer. In the nonlinear theory, the system function which specifies the behavior of the transducer may in general be different for different sources. For example, if the input is quantized, it is convenient to make use of a system function which is defined only for the possible inputs, since a more general description might be unnecessarily complex. This is not a real limitation, because any given transducer will usually be designed to work with a particular source, and its response to other sources is not of interest.

The transducers to be studied are restricted to those which have a finite memory, and which operate on a finite portion T of the immediate past of the input. If the input is a continuous function of time, it might at first appear that complete description of a long portion T of the past of the input could not be stored in a memory system having a finite capacity. If our sources were completely general, this would in fact be the case, but the sources considered are always restricted in some way. In particular, if the source is band limited to the bandwidth W, the past of its input over the time interval T may be specified by 2TW numbers (5). These numbers could be, for example, the values of the function at 2TW epochs equally spaced on the time interval T. If instead of a bandwidth limitation, the source is restricted to produce voltage pulses at quantized time intervals, the past of the input is specified by the values of the pulses produced during the interval T. Whatever the method chosen for specifying the past of the input, it will always be done in terms of a finite set of numbers.

Since the transducer operates on a portion of the past of the input covering a time interval T, the only information available to the transducer on which to base its output is the finite set of numbers which characterize the past of its input. Let us assume that there are s of the numbers. We may consider the state of the transducer as defined by the values of these s numbers. If we think of the values over which the s numbers can range as being measured along coordinate axes in an s-dimensional space, called the transducer space, the coordinates of a point in the space can be used to specify the state of the transducer. Since the s numbers which characterize the past of the input to the transducer are functions of time, the state point will move about as time progresses.

1.2 Classification of Transducers

We are now in a position to give a more precise definition of various kinds of transducers:

Invariant transducers. If a definite number can be assigned to each point in transducer space, such that the output of the transducer equals this number when the transducer is in the state defined by the point, the transducer is called time-invariant, or simply invariant.

Stationary transducers. If a fixed probability distribution can be assigned to each point in transducer space, such that the output of the transducer has this probability distribution when the transducer is in the state defined by the point, the transducer will be said to be stationary.

Finite state transducers. If the transducer can produce only a finite set of output values, it is called finite state.

If a transducer is invariant, it is also stationary, but the converse is not generally true. An invariant transducer will be finite state if the input to the transducer is quantized, so that the state point can occupy only a finite number of different positions. On the other hand, even if the input is not quantized and the state point varies over an infinite set of points, an invariant transducer can still be finite state. This will occur if the transducer space is broken up into a finite number of volume elements, and the output of the transducer remains constant as long as the state point is in any one volume element. Stationary transducers will be finite state under the same conditions as invariant transducers, provided in addition that all the defining probability distributions of the output are discrete.

1.3 Other Kinds of Transducers

It is of course possible to define many other classes of transducers. One important class might be transducers whose output depends on certain characteristics of the past of the input which are not restricted to have been determined in any particular time interval. A number of dimensions in transducer space would thus be devoted to storing certain types of information which might depend on portions of the past of the input indefinitely remote, rather than on just the immediately preceding time interval T. In the case of a discrete input, a typical datum of this type might be a record of the total number of pulses received up to the present, or a record of the value of the largest pulse so far received.

As has been pointed out, we are primarily interested in the application of nonlinear transducers to the processing of pulse communication signals. The natural transducer to use for this purpose is the invariant finite state transducer. Accordingly, we next describe a standard form in which any invariant finite state transducer can be represented. We then go on to discuss some of the properties of such transducers.

II. Synthesis of Invariant Transducers

2.1 Synthesis in Terms of Linear Elements and Rectifiers

Various standard forms for representing linear transducers have been known for some time. These are embodied in the Wiener-Lee canonical form, the Darlington insertion loss theory, the Cauer canonical form, and the Gewertz theory. The Wiener-Lee form (13, 14) is especially suitable for modification to allow the introduction of nonlinearity. Ways in which this modification may be accomplished will next be discussed.

We shall prove the following theorem: any invariant transducer driven by a finite quantized source can be synthesized by means of a finite number of resistors, capacitors, inductors, rectifiers*, and direct-current sources.

We prove this first for the case where the input symbols to the transducer consist of quantized voltage pulses of constant duration, uniformly spaced in time and having a finite number of different amplitudes. An example is shown in Fig. 1. It is unimportant whether the duration of the pulses equal the interval between pulses or not.

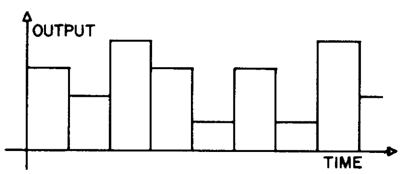


Fig. 1 Typical output of a finite quantized source.

Let the quantized input pulses have N possible different amplitudes, and let the transducer be capable of storing at most s input symbols. Then the transducer can have at most N^s possible states. Let the variable S_i designate the various states, where i ranges from 1 through N^s. Corresponding to each state there will be a specific output symbol. The output of the transducer can be represented by a function of the state:

$$\text{Output} = F(S_i) \quad .$$

Since the function F describes the behavior of the transducer under all conditions which can occur, it may be conveniently regarded as the system function of the transducer. The function F, and therefore the behavior of the transducer, can be represented schematically as shown in Fig. 2. It is clear, since the input is quantized, that the transducer will be finite state.

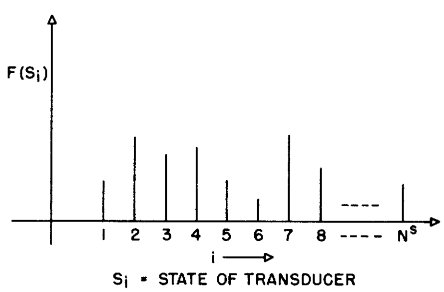


Fig. 2 Graphical representation of the system function of an invariant finite state transducer.

The past input symbols which determine the output at any instant of time can be stored in a delay line consisting of a ladder network of inductors and capacitors, terminated in a resistor. The delay line is tapped at intervals equal to the time interval between input symbols. Although in the present quantized case this implicitly limits the transducer to one output symbol per input symbol, this limitation will be removed when we consider continuous inputs.

---

*By a rectifier is meant an element which is open-circuited for one polarity of current, and short-circuited for the other polarity.

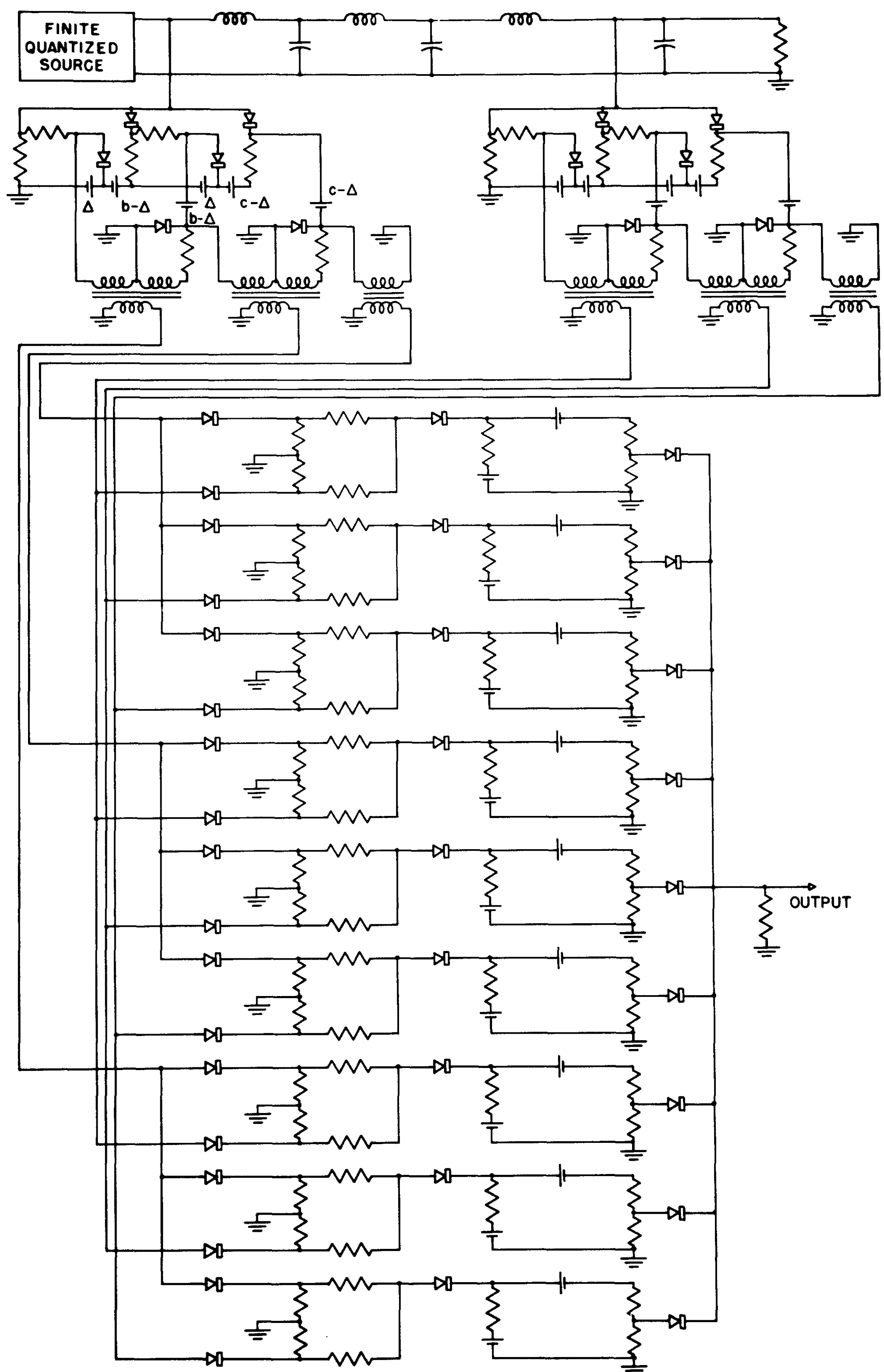


Fig. 3 An invariant finite state transducer driven by a finite quantized source.

The arrangement for producing an arbitrary output as in Fig. 2 from an input such as that of Fig. 1 is as shown in Fig. 3. For simplicity the transducer is shown as having storage capacity for only one past symbol, so that its output at any time is a function of the input symbol which is being produced and of the immediately preceding symbol. Also for simplicity the source is assumed to produce only three different symbols, say pulses having amplitudes of a, b, and c volts, where $0 < a < b < c$ and $\Delta > 0$ is chosen such that $a > \Delta$, $b - a > \Delta$, and $c - b > \Delta$. Each stored symbol is applied to a set of rectifiers which are biassed to voltages indicated on the diagram. Current will flow in the load resistors when the input symbol is greater than the biassing voltage. By means of a resistor-rectifier clipper which is biassed Δ volts higher than the corresponding load resistors, the voltage taken from each load resistor is limited to exactly Δ volts. The voltages obtained from adjacent load resistors are applied to opposite ends of the center-tapped primary of an ideal transformer. It is evident that only one of the secondaries of the three output transformers will be energized for each input pulse. The resistor-rectifier clippers at the right-hand terminals of the center-tapped primaries are to prevent coupling from one transformer to the next. The various possible combinations of secondary windings are connected to conventional coincidence circuits, only one of which will be energized at a time. The required amount of voltage is taken from each coincidence circuit by means of a resistive voltage divider, and the outputs of all the voltage dividers are mixed by means of a set of diodes connected to a common load resistor.

It is not necessary that the period and duration of the quantized pulses remain constant, as has been assumed up to now. If the leading and trailing edges of the pulses are quantized to discrete epochs kt_o, $k = 1, 2, \ldots$, the argument remains unchanged, except to note that the taps on the delay line should be placed at intervals equal to t_o. Each possible state of the transducer will therefore still produce a unique result. Although the discussion has been carried through for a particular number of possible symbols and for a specific amount of storage, these restrictions are in no way necessary. This completes the proof of the theorem.

For the case of an unquantized source, we obtain the following result: a necessary and sufficient condition that any finite state transducer be capable of synthesis by means of a finite number of resistors, inductors, capacitors, rectifiers, and direct-current sources, is that the regions defining the states of the transducer be bounded by plane (or hyperplane) surfaces.

It makes little difference whether the source is discrete or continuous. If it is discrete, the taps on the delay line are placed at intervals equal to the separation between symbols. If the source is continuous, we assume that the bandwidth occupied by its output is finite. Let W be the bandwidth of the source. Then the taps on the delay line are placed at intervals equal to $1/2W$. As is well known, the past of the input is then specified in the range $t = 0$ to $t = s/2W$, where s is the number of taps on the delay line.

We first prove the sufficiency of the condition. To that end, we describe a transducer which will indicate in which of a set of sub-ranges a voltage varying over a continuous range lies. In Fig. 4 the ranges are defined by means of the voltages to which

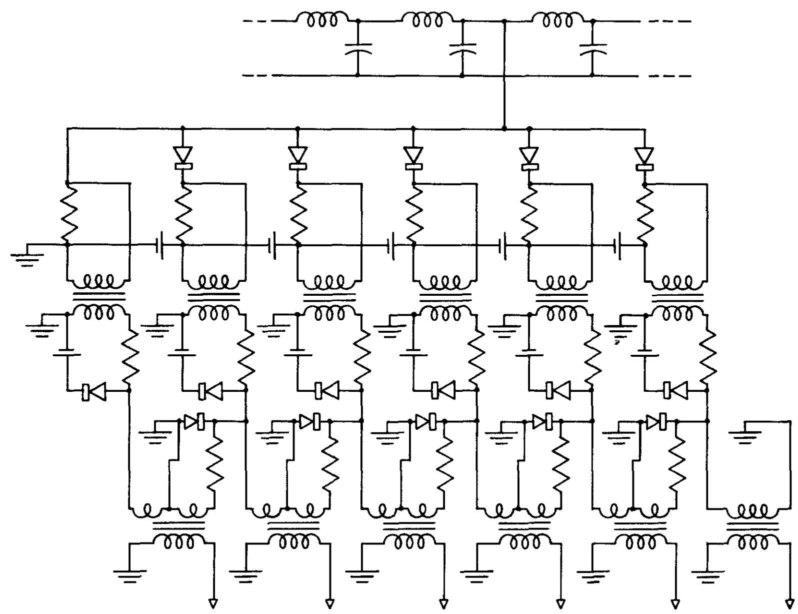


Fig. 4 Transducer to indicate range in which a varying voltage lies.

the upper rectifiers are biassed. The ideal transformer across each of the rectifier load resistors steps up the voltage by a very large factor. The resulting voltage is clipped by a resistor-rectifier clipper biassed to a small positive voltage. By making the voltage step-up of the ideal transformer sufficiently large, a standard voltage is obtained at the output of the clipper, independent of how much greater than the bias voltage the input voltage is. These resulting voltages are then applied in pairs to opposite ends of the center-tapped primaries of ideal transformers, as in the earlier figure. Only one of the transformer secondaries will be energized, in accordance with the range in which the input voltage lies. An exception occurs when the input voltage lies just above and very close to one of the biassing voltages. In this case, the voltage across the secondary of the step-up transformer will not be large enough to be clipped, so that a nonstandard output will be obtained from two range indicators, rather than a standard output from one. The proportion of time during which such failures occur can be made arbitrarily small by increasing the step-up ratio of the transformer and decreasing the clipping bias voltage. It should be noted that this peculiarity is not limited to the representation selected here, but is characteristic, in one form or another, of any finite-state transducer driven by an unquantized continuous source. That is, the peculiarity should be present in the representation.

We next note that any plane surface in the s-dimensional space of the transducer can be represented by an equation of the form

$$a_1 x_1 + a_2 x_2 + \dots + a_s x_s = c \quad . \tag{1}$$

For all points[x] = $(x_1\ x_2, \ldots, x_s)$ on one side of the plane, $\sum a_i x_i > c$, while for all points [x] on the other side of the plane, $\sum a_i x_i < c$. This is true since Eq. 1 is continuous in the x_i, and it is therefore impossible to pass continuously from a point where $\sum a_i x_i > c$ to a point where $\sum a_i x_i < c$ without passing through a point in the plane.

Now assume that the region R_k which defines the state S_k is bounded by a set of n plane surfaces:

$$\sum_{i=1}^{s} a_{ji} x_i = c_j, \quad j = 1, \ldots, n \quad . \tag{2}$$

In general some of the planes represented by Eq. 2 will not only bound, but will also pass through the region R_k, breaking it up into a number of sub-regions. Each of these sub-regions will be bounded by a subset of the planes represented by Eq. 2. The condition that a point [x] lie within a particular one of sub-regions bounded by the planes

$$\sum_{i=1}^{s} a_{ji} x_i = c_j, \quad j = j_1, j_2, \ldots \tag{3}$$

is that [x] satisfy simultaneously a set of inequalities of the form:

$$\sum_{i=1}^{s} a_{ji} x_i > c_j, \quad j = j_1, j_2, \ldots \tag{4}$$

since as long as the point [x] stays within a sub-region it cannot cross any of the bounding planes. The condition that a point lie within the region R_k is then that it satisfy one of the sets of inequalities of the form of Eq. 4, which correspond to the various sub-regions.

For each of the sub-regions of R_k, each of the linear combinations

$$\pm \sum_{i=1}^{s} a_{ji} x_i, \quad j = j_1, j_2, \ldots,$$

can be formed as shown in Fig. 5, by adding together the weighted output of the various taps. By means of a transducer such as the one already described which indicates in which set of ranges a given voltage lies, the comparison

$$\pm \sum_{i=1}^{s} a_{ji} x_i > c_j \tag{5}$$

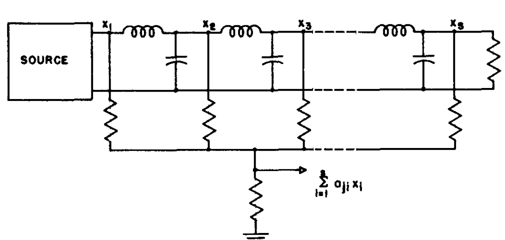


Fig. 5 Formation of weighted sum of past samples.

can be made. By doing this for each of the inequalities for which $j = j_1, j_2, \ldots$, a set of outputs is obtained. A set of coincidence circuits is used to indicate whether all the conditions are met, and if they all are, the coincidence circuit produces the output which corresponds to the region R_k. A similar process is carried out for the other sub-regions of R_k, and the outputs of all the sub-regions are mixed. The other state defining regions of the transducer are treated the same way as R_k.

This completes the proof of the sufficiency of the condition that an invariant finite state transducer be capable of synthesis with linear elements, rectifiers, and direct-current sources. The proof has been obtained by producing a transducer which meets the condition and is composed of the stated elements. A typical two-dimensional example is shown in Fig. 6.

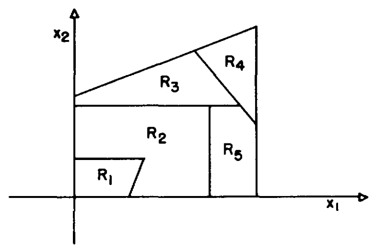


Fig. 6 Schematic representation of a possible two-dimensional finite state transducer.

To show that the condition is also necessary, we note that the systems we have been discussing are essentially switched systems which are governed by linear laws between transitions. If we choose any point $[x]_0$ continuous in $[x]$ on a bounding surface, an expansion of the surface about the point is linear if the bounding surface is planar. Hence if we vary one of the x's, say x_i, in such a way that x passes through the bounding surface in the neighborhood of $[x]_0$, the value of x_i at which a transition of state occurs will be a linear function of the other x's. On the other hand, if we assume that a certain portion of a bounding surface in the neighborhood of $[x]_0$ is not planar, an expansion of the surface about $[x]_0$ must include terms higher than the first. If now we vary x_i so that $[x]$ passes through the bounding surface in the neighborhood of $[x]_0$, the value of x_i at which a transition occurs will be a nonlinear function of the other x's. Even for infinitesimal changes in the x's a continuous nonlinearity is obtained. Such behavior cannot be exhibited by an ideal rectifier whose output is either linear or discontinuous. It follows that if ideal rectifiers and linear elements are to be used, the surfaces must be planar.

Any invariant finite state transducer can of course be approximated by a transducer composed of linear elements, rectifiers, and direct-current sources, since any surface can be approximated by a set of planes. In fact, the standard form which has been obtained is also applicable to time-invariant nonfinite state transducers, in the sense that it will approximate the output of such a transducer arbitrarily closely. Consider the s-dimensional space $x_1, \ldots, x_s$ of the transducer. From the point of view of the transducer, the relevant part of the past of the input to the transducer is described by the position of a point in this space, as we have seen, and the point will move about as time progresses. The transducer is completely specified by attaching a number (equal to the output of the transducer for the corresponding state) to each point in the space. If the input signal is bounded the space of the transducer is finite. If the numbers representing the output of the transducer are continuous in the space coordinates, it is possible to break the space up into a finite number of cells, and to assign a cell-number to each cell such that the output for any point in the cell will be arbitrarily close to this cell-number. A finite state transducer whose states are defined by the cells and whose output for each state is equal to the corresponding cell-number, will thus approximate arbitrarily closely to the output of the original transducer. The above argument may be considered as a proof of the following theorem: any invariant transducer can be approximated arbitrarily closely by means of a transducer composed of resistors, capacitors, inductors, rectifiers, and direct-current sources.

2.2 Synthesis in Terms of Function Generators

We have completed the discussion of the synthesis of transducers whose only nonlinear elements are rectifiers, and now go on to consider the use of other nonlinear elements. As long as we are dealing with finite state transducers, each state S_i may be associated with a region R_i in $[x]$-space, where R_i is bounded by a set of surfaces:

$$f_j([x]) = 0, \; j = 1, \ldots, n \; . \qquad (6)$$

In a later section we shall see that an important practical case arises when none of the surfaces $f_j = 0$ passes through the region R_i. Under this restriction, the condition that a point $[x]$ be interior to R_i is that

$$\pm f_j([x]) > 0, \; j = 1, \ldots, n \; . \qquad (7)$$

If a function generator is available for producing each of the functions $f_j([x])$, a set of comparison circuits and a coincidence circuit is all that is required to find out whether $[x]$ belongs to R_i or not. The transducer can be synthesized in the form shown in Fig. 7. There will be as many coincidence circuits as there are states S_i.

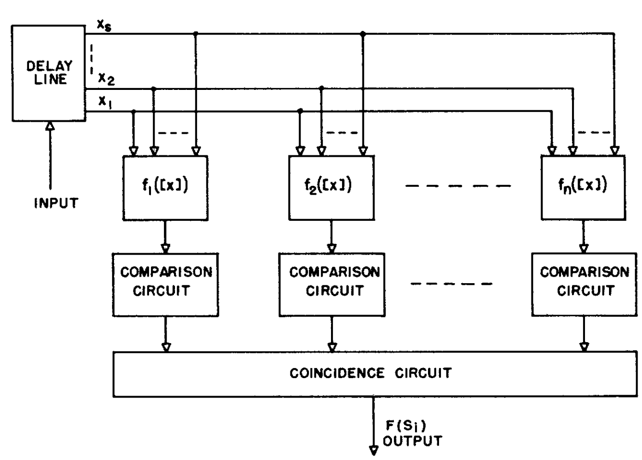


Fig. 7 Finite state transducer with regions having non-planar boundaries.

As an example we write out the equations for the two-dimensional transducer of Fig. 8.

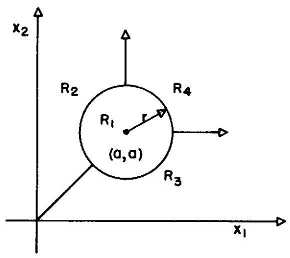


Fig. 8 Schematic representation of a possible two-dimensional finite state transducer.

The condition that the point belong to R_1 is that

$$(x_1 - a)^2 + (x_2 - a)^2 < r \quad .$$

For R_2 the conditions are:

$$(x_1 - a)^2 + (x_2 - a)^2 > r$$
$$0 < x_1 < a$$
$$x_1 - x_2 < 0 \quad .$$

For R_3:

$$(x_1 - a)^2 + (x_2 - a)^2 > r$$
$$0 < x_2 < a$$
$$x_1 - x_2 > 0 \quad .$$

And for R_4:

$$(x_1 - a)^2 + (x_2 - a)^2 > r$$
$$x_1 > a$$
$$x_2 > a \quad .$$

2.3 Expansion in Series

For future reference, we mention one other synthesis technique, a method especially suitable for invariant continuous output transducers. Let the set of numbers $[x] = (x_1, \ldots, x_s)$ specify the state S of the transducer. Then the output of the transducer at time t is given by

$$y(t) = F(S) = F([x(t)]) = F\left[x_1(t), \ldots, x_s(t)\right] , \tag{8}$$

and if y(t) is continuous in the x_i and has continuous derivatives, y may be expanded in a power series:

$$y(t) = \sum_i \cdots \sum_h a_{i, j, \ldots, h} \, x_1^i x_2^j \cdots x_s^h \tag{9}$$

$$= \sum_\alpha A_\alpha X_\alpha (t) \tag{10}$$

where $X_\alpha(t)$ represents one of the products $x_1^i x_2^j \ldots x_s^h$ and A_α is the corresponding $a_{i,j,\ldots,h}$. If we wish to approximate the infinite expansion by a finite expansion containing M terms, we may limit the range of values which the exponents $i, j, \ldots, h$

can assume to 0 to N – 1, and α may be identified as follows:

$$\alpha = i + j\,N + \ldots + h\,N^{s-1}, \tag{11}$$

so that α ranges from 0 to $N^S - 1 = M$. The transducer in this case includes adding and multiplying circuits; the form of the transducer would be as shown in Fig. 9.

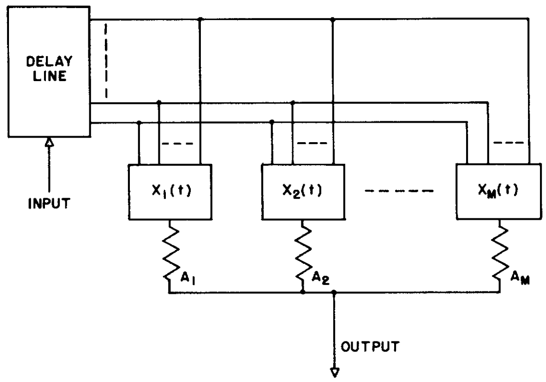


Fig. 9 Series form of nonfinite state transducer.

We shall also find it convenient in a later application to use, instead of a power series representation, a series expansion in orthogonal functions, as employed by Wiener (15).

III. Synthesis of Stationary Transducers

By making use of the finite state transducer synthesis technique which has been developed in the preceding section, we are in a position to describe a standard form for stationary transducers. Instead of a definite output voltage for each state, a voltage which has a fixed probability distribution dependent on the state must be produced. Although the optimum synthesis we later discuss is based on invariant rather than stationary transducers, it is perhaps worth noting that stationary transducers have possible value in the synthesis of secrecy systems, in studying the effects of noise in transmission systems, and in the study of information sources. We shall show, for example, that any information source whose output depends statistically on a fixed finite portion of its past can be represented in terms of an appropriate stationary transducer whose output is its own input. In providing a building block for the synthesis of information sources, it is possible that stationary transducers may also serve as useful laboratory tools.

To provide a basis for the stationary transducer synthesis, we assume that we have a certain invariant finite state transducer called a state evaluator. The state evaluator provides an indication of the state the transducer is in. A schematic representation is shown in Fig. 10.

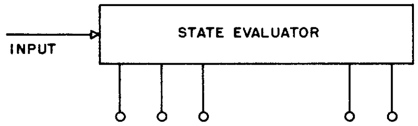


Fig. 10 Block diagram of state evaluator.

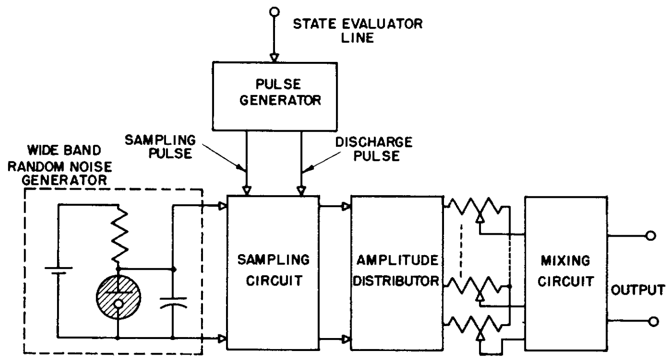
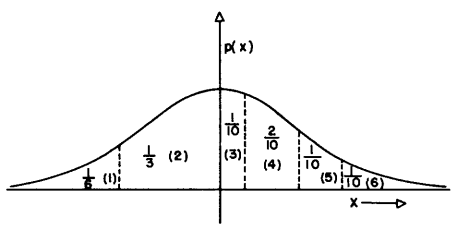


Fig. 11 Distribution control circuit.



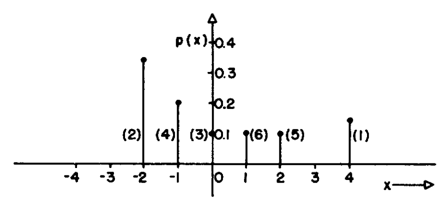


Fig. 12 Separation of gaussian distribution in ranges having assigned probabilities.

Fig. 13 Output probability distribution for separation according to Fig. 12.

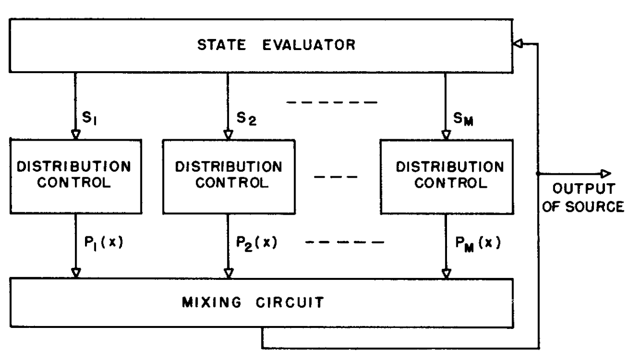


Fig. 14 Stationary transducer used as a source.

A unit voltage appears on the output line corresponding to the state of the transducer. As the state point moves in transducer space from one state-defining region to another, the output line which is energized changes accordingly. The state evaluator is evidently composed of sets of comparison circuits and coincidence circuits, as described in Sect. II, and its design need not be discussed any further here.

In order to convert the state evaluator to a stationary transducer, it is sufficient to connect each of the output lines $S_1, \ldots, S_M$ of Fig. 10 to separate circuits which are each adjusted to produce an output voltage which has an arbitrary prearranged probability distribution.

A method of producing a voltage having an arbitrary discrete probability distribution will now be described. The elements are shown in block diagram form in Fig. 11. To complete the stationary transducer, one of these arrangements is connected to each state indicator line. A wide band noise source is sampled at the time the state indicator line is energized, and the resulting sample is stored on a capacitor in the sampling circuit. The stored voltage is applied to an amplitude distributor which produces a standard voltage on one of a number of output lines, according to the voltage range in which the stored sample lies. By means of the potentiometers shown at the output of the amplitude distributor, any required proportion of the standard voltage is applied to the output circuit. When the transducer again changes state, the capacitor is discharged.

The probability distribution of the noise source is gaussian. The biases in the amplitude distributor are adjusted in such a way as to break the total area under the gaussian curve into sub-areas having predetermined values. Figure 12 shows an example of a possible distribution. The samples of noise are broken up into six amplitude ranges, such that the probabilities of falling in ranges one through six are respectively 1/6, 1/3, 1/10, 2/10, 1/10, and 1/10. Arbitrary output voltages of 4, − 2, 0, − 1, 2, and 1 are then produced for sample voltages falling in these ranges, as shown in Fig. 13.

In Fig. 14 is shown a diagram of a stationary transducer whose input is provided by its own output. The blocks labelled "distribution control" are circuits such as those of Fig. 11. Since the distribution control circuits may be set to have any desired discrete distribution, and since the state evaluator may be connected in such a way as to provide an arbitrary correspondence between distributions and states, it follows that the arrangement is a model for any discrete statistical source whose output depends on a finite portion of its past.

IV. Statistical Analysis of Transducers

4.1 Correlation Analysis of Transducers

It has been shown (16, 17) that for a linear transducer

$$\phi_{io}(\tau) = \int_{-\infty}^{\infty} h(t)\ \phi_{ii}(t-\tau)\, dt \quad . \qquad (12)$$

For the discrete case this may be written

$$\overline{x(n)\,y(n+\tau)} = a_1\,\overline{x(n)\,x(n+\tau)} + a_2\,\overline{x(n)\,x(n+\tau-1)} + \ldots + a_s\,\overline{x(n)\,x(n+\tau-s+1)} \qquad (13)$$

where the output of the transducer is given by

$$y(n) = a_1\,x(n) + a_2\,x(n-1) + \ldots + a_s\,x(n-s+1) \;. \qquad (14)$$

If we know the value of $\phi_{ii}(\tau)$ and $\phi_{io}(\tau)$ for $\tau = 0$ to $\tau = s-1$, the averages in Eq. 13 are known. Letting τ range over the values $\tau = 0$ to $\tau = s-1$ in Eq. 13 provides a set of s simultaneous equations in the s unknowns $a_1, \ldots, a_s$. By solving the equations for the a_i, the system function is determined. Thus the crosscorrelation function of the input with the output, and the autocorrelation of the input are together sufficient to define a linear system. This rather striking result comes about because the form of the equation defining the transducer is known. The problem of completely describing a transducer mathematically is greatly simplified when the form of the equation relating input and output is specified. The remainder of the problem of specifying the transducer is that of finding the coefficients which must be fitted into the known form.

In the nonlinear case it may also happen that the form of the equation describing the transducer is known. This will happen, for example, if the kind of nonlinear elements available is restricted. In this case the coefficients which are required to complete the characterization of the transducer can be found by making use of statistical parameters similar to those employed in the linear analysis.

Let the transducer operate on a finite portion of the past of the input $(t, t-T)$ where $T = s - 1/2W$, and let the input be band limited to the band W. Then the equation of the transducer can be written

$$\begin{aligned} y(t) &= F\left[x(t),\, x\left(t-\tfrac{1}{2W}\right),\, \ldots,\, x\left(t-\tfrac{s-1}{2W}\right)\right] \\ &= \sum_i \cdots \sum_h a_{i,j,\ldots,h}\left[x(t)\right]^i \left[x\left(t-\tfrac{1}{2W}\right)\right]^j \cdots \left[x\left(t-\tfrac{s-1}{2W}\right)\right]^h \\ &= \sum_\alpha A_\alpha\, X_\alpha(t) \end{aligned} \qquad (15)$$

where $X_\alpha(t)$ represents one of the products

$$\left[x(t)\right]^i \left[x\left(t-\tfrac{1}{2W}\right)\right]^j \cdots \left[x\left(t-\tfrac{s-1}{2W}\right)\right]^h$$

and A_α is the corresponding value of $a_{i,j,\ldots,h}$. If 0 to $N-1$ is the range of values which the exponents $i, j, \ldots, h$ can assume, we may identify α as follows:

$$\alpha = i + j\,N + \ldots + h\,N^{s-1} \qquad (16)$$

so that α ranges from 0 to $N^S - 1 = M$.

On the basis of the assumed form for the nonlinear transducer, a complete description is provided by specifying the coefficients A_α. In order to determine the values of A_α in terms of statistical parameters, we multiply Eq. 15 by $X_\beta(t)$, and average over all time:

$$\overline{X_\beta(t)\, y(t)} = \sum_\alpha A_\alpha \, \overline{X_\beta(t)\, X_\alpha(t)} \tag{17}$$

$$\phi_{xy}(\beta) = \sum_\alpha A_\alpha \, \phi_{xx}(\alpha, \beta) \quad , \tag{18}$$

where the higher order correlation functions ϕ_{xy} and ϕ_{xx} are written in place of the corresponding time averages. Eq. 18 must hold for all β, and so provides a set of M linear simultaneous equations in M unknowns which leads to the evaluation of the A_α.

4.2 Experimental Study of Unknown Transducers

An interesting technique for experimentally evaluating the defining coefficients of an unknown transducer can be based on the above result. Let it be required to evaluate the system function of a transducer having a prescribed form of nonlinearity for a signal which is band limited to the band W. The experimental arrangement shown in block diagram form in Fig. 15 could be used. We choose a noise source having a bandwidth large compared to W. Successive samples of the noise are obtained by means

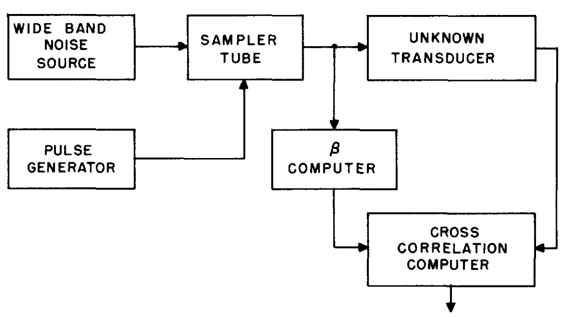


Fig. 15 Experimental study of unknown nonlinear transducer.

of the sampler tube and a pulse generator having a period 1/2W. The amplitudes of the samples are all independent of each other, since the autocorrelation function of the noise becomes negligible in a time small compared to the sampling period. The values of the correlation functions $\phi_{xx}(\alpha, \beta)$ are therefore all easily computed, and are in fact given by

$$\phi_{xx}(\alpha, \beta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} X_\alpha(x_1, \ldots, x_s)\, X_\beta(x_1, \ldots, x_s)\, p(x_1, \ldots, x_s)\, dx_1 \ldots dx_s \, . \tag{19}$$

It remains to evaluate $\phi_{xy}(\beta)$. This is accomplished experimentally, as shown in Fig. 15, by first computing $X_\beta(t)$ and then crosscorrelating the result with the output y(t) of the unknown nonlinear transducer. This amounts to multiplying the two functions and averaging the product. By repeating this measurement for each value of β, the complete set of simultaneous equations represented by Eq. 18 can be set up, and the coefficients A evaluated.

The experimental procedure which has been described is based on a method originally suggested by Wiener (15). We now describe Wiener's method, which has the virtue of not requiring the solution of any simultaneous equations. The idea is to use normal and orthogonal functions in place of the products $X_\alpha(t)$ in the expansion of the system function in Eq. 15. We first rewrite Eq. 15, replacing $x(t), \ldots, x(t-s-1/2W)$ by a set of variables $u_1(t), \ldots, u_s(t)$ which characterize the past of the input. Thus

$$y(t) = F\left[u_1(t), u_2(t), \ldots, u_s(t)\right] . \quad (20)$$

The system function F is next expanded in a series of Hermite functions (19), giving

$$y(t) = F(u_1, \ldots, u_s) = \sum_i \cdots \sum_h a_{i,j,\ldots r} H_i(u_1) \ldots H_h(u_s) \exp\left[-\frac{u_1^2 + \ldots + u_s^2}{2}\right]$$

$$= \sum_\alpha A_\alpha H(\alpha) \exp\left[-\frac{u_1^2 + \ldots + u_s^2}{2}\right] \quad (21)$$

where $H(\alpha)$ represents the polynomial $H_i(u_1) \ldots H_h(u_s)$, and α is defined as in Eq. 16.

If we now multiply by $H(\beta)$ and average over all time we find

$$\overline{y(t)\,H(\beta)} = \sum_\alpha A_\alpha \overline{H(\alpha)\,H(\beta) \exp\left[-\frac{u_1^2 + \ldots + u_s^2}{2}\right]} \quad (22)$$

In order to calculate the time averages in the summation, we make use of the ergodic theorem to replace the time averages by phase averages. The phase averages can be conveniently calculated if we choose an input whose parameters $u_1, \ldots, u_s$ have independent gaussian distributions, and which have been normalized to have a standard deviation of unity. Then the joint probability distribution of $u_1, \ldots, u_s$ is

$$(2\pi)^{-s/2} \exp\left[-\frac{u_1^2 + \ldots + u_s^2}{2}\right]$$

and the sum of the phase averages becomes:

$$(2\pi)^{-s/2} \sum_\alpha A_\alpha \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H(\alpha)\,H(\beta) \exp\left[-(u_1^2 + \ldots + u_s^2)\right] du_1 \ldots du_s \quad . \quad (23)$$

Since the Hermite functions are orthogonal over $(-\infty, \infty)$, expression 23 reduces to $(2\pi)^{-s/2} K^2 A_\beta$ and we obtain:

$$A_\beta = \frac{(2\pi)^{s/2}}{K_\beta^2} \overline{y(t)\,H(\beta)} \quad . \quad (24)$$

Here the constant K_β is the product of the normalizing factors associated with the Hermite polynomials whose product equals $H(\beta)$. Thus (19)

$$K_\alpha = \left[\pi^{s/2} 2^{i+j+\ldots+h} (i!)(j!)\ldots(h!)\right]^{1/2} . \tag{25}$$

Wiener's method lends itself readily to the experimental evaluation of unknown system functions, but even if the representation requires the use of only a few Hermite functions, the complexities involved in automatically computing $H(\beta)$ become very great. Nevertheless, as electronic methods for multiplication continue to be improved and become more available, Wiener's method of studying unknown nonlinear transducers may receive considerable application.

In the equations which have been derived, we related the characteristic parameters of a nonlinear transducer to the higher order correlation functions

$$\phi_{xy}(u_1, \ldots, u_s) \text{ and } \phi_{xx}(u_1, \ldots, u_s)$$

of the input and output signals. These correlation functions represent certain averages, which can of course be computed from a knowledge of the joint probability distributions

$$P(u_1, \ldots, u_s) \text{ and } P(y : u_1, \ldots, u_s) .$$

It is perhaps worth noting that since we are dealing with invariant transducers, the distribution $P(y : u_1, \ldots, u_s)$ is alone sufficient to specify the transducer; because for each set of values $u_1, \ldots, u_s$ there is only one possible output, so that

$$\begin{aligned} P(y : u_1, \ldots, u_s) &= 1 \text{ for } y = F(u_1, \ldots, u_s) \\ &= 0 \text{ otherwise} . \end{aligned} \tag{26}$$

Also, if the variables $u_1, \ldots, u_s$ are quantized so that the transducer can have only a finite number of states, and if a unique value of y corresponds to each state, then the transducer is completely characterized by the set of distributions

$$P\left[y_j : u_{ki}\right], \ k = 1, \ldots, s .$$

where $P\left[y_j : u_{ki}\right]$ is the probability that y has its jth value when u_k has its ith value. For we may write

$$P(u_{ki} : y_j) = \frac{P(u_{kj})\, P(y_j : u_{ki})}{\sum_i P(u_{ki})\, P(y_j : u_{ki})} \tag{27}$$

and since there is only one state for each y_j, it follows that $P(u_{ki} : y_j)$ will be unity for one value of i and zero for all others. Thus by holding j constant and ranging over all k we can find the state S_j which corresponds to y_j. By carrying out the same procedure for each value of j we completely specify the transducer. We need not actually know $P(u_{ki})$ in order to carry out this process, since $P(y_j : u_{ki})$ will differ from $P(u_{ki} : y_j)$ only by the factor $P(u_{ki})/P(y_j)$, which is a positive number for all i, j. Hence $P(y_j : u_{ki})$ will differ from zero for each combination of k and y_j only for one u_{ki}.

4.3 Influence of Transducers on Probability Distributions

This section on the statistical analysis of transducers will be concluded with a few remarks regarding the influence of a transducer on the statistics of the signal on which it operates. The transducer and the statistics of the input are assumed to be known, and it is required to calculate the statistics of output. In the corresponding linear case it is usual to consider only the spectra, or correlation functions. Thus if we have a linear transducer with a certain frequency response characteristic and the input is a white noise, we say that the power spectrum of the output has the same form as the frequency response. Such statistical assertions are of great value in judging the performance of transducers, and have in fact provided the basis for much linear filter design.

We shall consider a fairly general case, in which it is desired to calculate the conditional probability distribution functions of the output when the transducer is specified and the set of input probability distribution functions is known. Let the input time series consist of a sequence of uniformly spaced quantized pulses x(n), where n is a discrete time parameter, and x(n) is the nth pulse applied to the input of the transducer. The sequence x(n) is assumed to be a Markoff process (20) of order r, so that it is governed by the joint probability distribution function

$$p\left[x(n),\ x(n-1),\ \ldots,\ x(n-r)\right] \ . \tag{28}$$

Let each pulse have one of N different amplitudes. Since the transducer operates on s input pulses, it can have a maximum of $N^s = M$ possible states S_m, where $m = 1, \ldots, M$. If we designate by $x_{u(k)}$ the value of x stored in the kth position in the transducer, where u(k) is an index which ranges from 1 to N, the possible states S_m can be ordered according to the equation

$$m = u(1) + \left[u(2) - 1\right] N + \ldots + \left[u(s) - 1\right] N^{s-1} \ . \tag{29}$$

The probability of occurrence of the state S_m is then

$$p\left[S_m\right] = p\left[x_{u(1)},\ x_{u(2)},\ \ldots,\ x_{u(s)}\right] \ . \tag{30}$$

If $s \leqslant r$, $p\left[S_m\right]$ is found by summing over the extra variables in expression 28. If $s > r$, so that $s - r = a$, $p\left[S_m\right]$ is found from

$$p\left[S_m\right] = p\left[x_{u(a)},\ x_{u(a+1)},\ \ldots,\ x_{u(s)}\right] \\ p\left[x_{u(1)},\ x_{u(2)},\ \ldots,\ x_{u(0-1)} : x_{u(a)},\ \ldots,\ x_{u(s)}\right] \ . \tag{31}$$

But

$$p\left[x_{u(1)},\ x_{u(2)},\ \ldots,\ x_{u(a-1)} : x_{u(a)},\ \ldots,\ x_{u(s)}\right] \\ = \prod_{k=1}^{a-1} \frac{p\left[x_{u(a-k)},\ x_{u(a-k-1)},\ \ldots,\ x_{u}(r+a-k)\right]}{p\left[x_{u(a-k-1)},\ x_{u(a-k-2)},\ \ldots,\ x_{u(r+a-k)}\right]} \ . \tag{32}$$

Similarly the conditional probability that state j follow state i, where

$$i = v(1) + \left[v(2) - 1\right] N + \ldots + \left[v(s) - 1\right] N^{s-1}$$

and

$$j = w(1) + \left[w(2) - 1\right] N + \ldots + \left[w(s) - 1\right] N^{s-1}$$

is given by

$$p\left[S_j(n+1) : S_i(n)\right] = p\left[x_{w(1)} : x_{v(1)}, x_{v(2)}, \ldots, x_{v(s)}\right],$$

$$p\left[S_j(n+2) : S_i(n)\right] = p\left[x_{w(2)} : x_{w(1)}, x_{v(1)}, x_{v(2)}, \ldots, x_{v(s)}\right],$$

etc. (33)

On the basis of the probability distributions determined for the states, we may now find the probability distributions of the output. Let the equation of the transducer be

$$A_m = F\left[S_m\right], \tag{34}$$

where A_m is the output when the transducer is in state S_m. If the output time series is represented by y(n), each value of y will equal one of the numbers A_m. It follows that the output probability distribution is

$$P(y_t) = \sum_j p\left[S_j\right] \tag{35}$$

where the summation is over all j for which $A_j = y_t$.

Similarly

$$P\left[y_q(n+1) : y_t(n)\right] = \sum_k \sum_j p\left[S_k(n+1) : S_j(n)\right], \tag{36}$$

where the summation is over all j and k such that $A_k = y_q$ and $A_j = y_t$.

If a transducer has a unique output corresponding to each state and $s \geqslant r$, the probability distribution of the next output symbol is completely determined by the present output symbol. It follows that the output in this case is a first order Markoff process, i.e., $r_o = 1$, where r_o is the order of the output Markoff process.

On the other hand, if $s < r$, the probability distribution of the next input symbol, and hence of the next output symbol, is dependent not only on the s stored symbols but also on the $r - s$ additional symbols which are no longer stored. These $r - s$ input symbols may be determined, however, if the preceding $r - s$ output symbols are known. The probability distribution of the next input symbol, and hence of the next output symbol, is determined by these $r - s$ output symbols plus the present output symbol. That is, the order of the output Markoff process is $r_o = r - s + 1$.

V. Synthesis of Optimum Nonlinear Filters

5.1 General Remarks

The problem to be considered in this section is that of separating two interfering signals, one of which may be a noise. If the distribution of the energy in the signal as a function of frequency is different from that of the noise, a partial separation can be affected by means of a linear filter. It has been shown by Wiener (1) that if the noise and signal are independent, the power density spectra of the signal and noise are sufficient to prescribe the best possible linear filter in the mean square error sense.*

In this incoherent case, the treatment given the filter problem by Lee (18) makes it clear that essentially none of the noise energy which lies in a band where the signal energy is relatively large will be effectively removed by the optimum linear filter. In particular, if the noise and signal have the same spectrum, no separation can be accomplished, in spite of the fact that other statistical parameters of the noise and signal may differ considerably. The ability of a linear device to separate noise and signal is thus limited because direct use is not made of parameters which provide a more complete statistical characterization.

The foregoing considerations indicate that in the general case, a nonlinear device can be expected to do a better job of separating noise and signal than a linear filter (Cf. Sect. 5, 6), and that the operation of the device should be governed by statistical parameters of the noise and signal. It is possible to consider a noise separating device as having for its function to make the information in a corrupted signal more readily understandable to a user (receiver, decoder). Thus the noise discriminator is an information processing device. The information carried by the signal is defined by sets of amplitude probability distribution functions (2). It follows that the behavior of the device should take into account the same sets of probability distributions. This is done, of course, in the synthesis of optimum linear filters, since the power spectra may be obtained from the probability functions. It may be noted that as many different time series governed by different sets of probability distributions as desired can be constructed which lead to the same power spectrum. All these time series would be processed by the same linear filter: on the other hand, the optimum nonlinear filter would in general be different for each of the series.

5.2 An Example

Before going on to formulate the filter problem on a general basis, it may be of interest to illustrate some of the ideas involved by working out a simple example. The example chosen is that of a signal and noise having identical spectra.

The signal consists of a sequence of uniformly spaced independent pulses which may have amplitudes of +1, 0, and – 1, with the following probabilities:

*If the noise and signal are not independent, the cross-power spectrum must also be known.

$$p(+1) = p(0) = p(-1) = \frac{1}{3} \quad . \tag{37}$$

The noise consists of a similar sequence of uniformly spaced independent pulses which may have amplitudes of + 2, 0, and − 2, with probabilities:

$$p(+2) = p(-2) = \frac{1}{12}, \; p(0) = \frac{5}{6} \quad . \tag{38}$$

This situation corresponds to interference between two signals on the same channel, with the signals having the same general form but different statistical characteristics. The signal and noise have identical power spectra and equal power, so that no separation can be carried out by means of a linear filter. A nonlinear filter is easily designed, however, by taking into account the probability distributions.

The noise and signal are combined additively, so that the input to the transducer consists of a sequence of pulses with amplitudes ranging from + 3 to − 3. The possible combinations and the corresponding probabilities are shown in Table I.

Table I

Received Signal	Possible Combinations Signal	Noise	Probability
+ 3	+ 1	+ 2	$\frac{1}{36}$
+ 2	0	+ 2	$\frac{1}{36}$
+ 1	+ 1	0	$\frac{10}{36}$
+ 1	− 1	+ 2	$\frac{1}{36}$
0	0	0	$\frac{10}{36}$
− 1	− 1	0	$\frac{10}{36}$
− 1	+ 1	− 2	$\frac{1}{36}$
− 2	0	− 2	$\frac{1}{36}$
− 3	−1	− 2	$\frac{1}{36}$

From the table the values which the signal and noise must have are evident for all received pulses other than + 1 or − 1. One or the other of these two ambiguous pulses is received 22/36 of the time. An ambiguous pulse can be caused by either of two possible combinations, one of which is of high probability (10/36), and the other of low probability (1/36). Since the magnitude of the error caused by a wrong selection is independent of which of the two possible combinations is chosen, it follows that the combination of high probability should always be selected to represent an ambiguous pulse. The optimum transducer can thus be specified by the input-output relation given in

Table II. Any number of graphs can be constructed which will pass through the points given in Table II, and a suitable graph which can be conveniently synthesized is shown in Fig. 16.

Table II

Input	Output
+ 3	+ 1
+ 2	0
+ 1	+ 1
0	0
− 1	− 1
− 2	0
− 3	− 1

A block diagram which will produce the input-output relation of Fig. 16 is shown in Fig. 17. The upper flipflop is arranged to trigger when the input exceeds + 1.5, and when it triggers it adds − 2 to the original input. The lower flipflop triggers when the input is less than − 1.5, and when it triggers it adds + 2 to the original input.

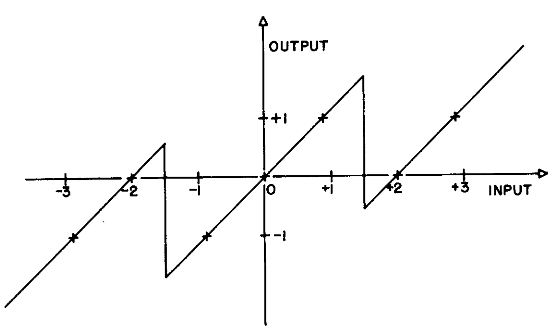


Fig. 16 Response for minimum probability of error.

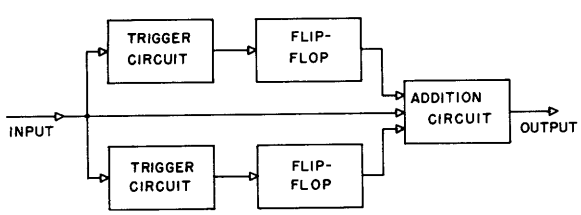


Fig. 17 Nonlinear transducer with response of Fig. 16.

The result is the input-output curve of Fig. 16. A circuit realization of the block diagram would require two double triodes and a number of resistors, capacitors, and crystal diodes.

The effectiveness of the transducer in this case is measured by its reduction of the mean square error from 6/9 to 2/9, and by its reduction of the probability of error from 3/18 to 1/18.

5.3 Formulation of the Filter Problem

We consider a message $f_m(t)$ whose past over the time interval (t, t − T) can be described by the set of parameters $u_1(t)$, $u_2(t)$, ..., $u_s(t)$. Such a description will be possible, as we have seen, if for example the message is band limited to the band W, and the values u_1, ..., u_s represent samples of the message at the points t, t − 1/2W, ..., t − s − 1/2W = t − T, where s = 2TW + 1. The description in terms of u_1, ..., u_s is of course especially applicable to pulse communication signals.

We assume that the message as received at some point is additively perturbed by a noise $f_n(t)$ whose past over the same time interval $(t, t - T)$ can be described by the set of parameters $v_1(t), v_2(t), \ldots, v_s(t)$. The received combination of signal and noise will then be represented by

$$u_1(t) + v_1(t), \ldots, u_s(t) + v_s(t) = w_1(t), \ldots, w_s(t) \quad .$$

We denote the set of functions $u_1(t), \ldots, u_s(t)$ by $\left[u(t)\right]$, the set of functions $v_1(t), \ldots, v_s(t)$ by $\left[v(t)\right]$, and the set of functions $w_1(t), \ldots, w_s(t)$ by $\left[w(t)\right]$. We assume that $\left[u\right]$ and $\left[v\right]$ are stationary random functions with given characteristic probability distributions, so that the probability distributions $p(\left[u\right]) = p(u_1, \ldots, u_s)$, $p(\left[v\right]) = p(v_1, \ldots, v_s)$, and $p(\left[w\right]) = p(w_1, \ldots, w_s)$ can be calculated.

The general filter problem may now be stated as follows: what operator $F(\left[w(t)\right])$ best approximates, in a prescribed statistical sense, a desired function $f\left[f_m(t)\right]$

The problem as formulated includes prediction, detection, decoding, and ordinary filtering as special cases. If, for example, the function $f\left[f_m(t)\right] = f(\left[u(t)\right])$, we might have a filter or a decoder. The transducer does not know the values of $u_1, \ldots, u_s$, but is nevertheless required to produce the best approximation to some function of them. As another example, if $f\left[f_m(t)\right] = f_m(t + \alpha)$, $\alpha > 0$, we would have a problem in prediction.

The answer to the problem as formulated will evidently have to be expressed in terms of the probability distributions of $\left[u\right]$, $\left[v\right]$, and $\left[w\right]$. As we have noted, Wiener obtained the solution to the problem for the case in which f and F are linear operators on the message and on the past of the message plus noise respectively. The result is expressed in the Wiener-Hopf equation:

$$\phi_{id}(\tau \pm \alpha) = \int_{-\infty}^{\infty} h(t)\, \phi_{ii}(t - \tau)\, dt, \quad \tau > 0 \quad , \tag{39}$$

where h(t) is the linear operator on the past of the message plus noise which is required to be found, and the correlation functions are determined by the probability distributions and the desired operation on the message. The statistical sense in which h(t) is the best linear operator is that it gives minimum average error power.

We now proceed to investigate some of the techniques which can be used to solve the problem when the functions f and F are not restricted to be linear. We shall usually require that F assumes a prescribed mathematical form. The form may of course be very inclusive, as in the case of a power series or series of Hermite functions. Or the form of F may be such that the space of $\left[w\right]$ is broken up into a large number of small cells, and $F(\left[w\right])$ required to be constant in each cell. By making the number of cells large enough, F can be made to approximate arbitrarily closely to the most general invariant transducer (Cf. Sect. II).

5.4 Determination of Series Representation for Minimum Mean Square Error

As we saw in Sect. II and Sect. IV, we may represent F ([w]) by a series:

$$y(t) = F([w]) = \sum_{a=1}^{M} A_a W_a(t) \tag{40}$$

where the W(t) are known functions, such as products of the functions $w_1, \ldots, w_s$, or products of Hermite functions of the w_i. The filter problem becomes that of finding the best values of A_a for approximating to $f\left[f_m(t)\right] = f_d(t)$. The values of the A_a for minimum mean square error can be obtained as follows:

Let the error at time t be given by

$$\epsilon(t) = y(t) - f_d(t) \quad . \tag{41}$$

The mean square error is

$$E = \overline{\left[\epsilon(t)\right]^2} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left[y(t) - f_d(t)\right]^2 dt$$

$$= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left[\sum_{a} A_a W_a(t) - f_d(t)\right]^2 dt$$

$$= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left[f_d(t)\right]^2 dt - 2\sum_{a} A_a \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} W_a(t)\, f_d(t)\, dt$$

$$+ \sum_{a,\beta} A_a A_\beta \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} W_a(t)\, W_\beta(t)\, dt \quad . \tag{43}$$

The limits in this equation are higher order correlation functions, and they may be evaluated in terms of the known probability distributions. Thus we have

$$\phi_{dd}(0) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left[f_d(t)\right]^2 dt$$

$$= \int_{-\infty}^{\infty} \left[f_d\right]^2 p\left[f_d\right] df_d, \tag{44}$$

$$\phi_{wd}(a) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} W_a(t)\, f_d(t)\, dt$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_\alpha f_d \, p\left[W_\alpha, f_d\right] d\left[w\right] df_d \tag{45}$$

$$\phi_{ww}(\alpha, \beta) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} W_\alpha(t) \, W_\beta(t) \, dt$$

$$= \int_{-\infty}^{\infty} W_\alpha W_\beta \, p\left[W_\alpha, W_\beta\right] d\left[w\right] \quad . \tag{46}$$

Substituting,

$$E = \phi_{dd}(0) - 2 \sum_{\alpha} A_\alpha \, \phi_{wd}(\alpha) + \sum_{\alpha, \beta} A_\alpha A_\beta \, \phi_{ww}(\alpha, \beta) \quad . \tag{47}$$

In order to make E a minimum, we must have

$$\frac{\partial E}{\partial A_\gamma} = 0 \text{ for all} \quad . \tag{48}$$

Thus

$$\frac{\partial E}{\partial A_\gamma} = -2\phi_{wd}(\gamma) + 2 \sum_{\alpha} A_\alpha \, \phi_{ww}(\alpha, \gamma) = 0 \tag{49}$$

or

$$\phi_{wd}(\gamma) = \sum_{\alpha} A_\alpha \, \phi_{ww}(\alpha, \gamma), \ \gamma = 1, \ldots, M \quad . \tag{50}$$

This is a set of M simultaneous equations in M unknowns, and hence leads to an evaluation of the A_α and of the system function F.

We now show that this necessary condition specified by Eq. 50 that E be a minimum is also sufficient. That the values of A actually specify a minimum and not a maximum, may be verified by observing that $\partial E/\partial A_\gamma > 0$ for A_γ greater than its optimum value, and $\partial E/\partial A_\gamma < 0$ for A_γ less than its optimum value, since $\phi_{ww}(\gamma, \gamma)$ is positive. Hence E is increased by any departure of A_γ from optimum. Furthermore, the error is represented by a polynomial of the second degree in the A's so that there cannot be more than one minimum. Hence condition 50 is sufficient.

If now the A's are obtained by solving Eq. 50, the mean square error becomes, by substituting Eq. 50 in Eq. 47,

$$E = \phi_{dd}(0) - 2 \sum_{\alpha} A_\alpha \, \phi_{wd}(\alpha) + \sum_{\alpha} A_\alpha \, \phi_{wd}(\alpha)$$

$$= \phi_{dd}(0) - \sum_{\alpha} A_\alpha \, \phi_{wd}(\alpha)$$

$$= \phi_{dd}(0) - \sum_{\alpha,\beta} A_\alpha A_\beta \phi_{ww}(\alpha,\beta) \quad . \tag{51}$$

5.5 Example of Power Series Synthesis

As an illustration of the use of the results which have been developed, we apply them to the design of a minimum mean square error transducer for the noisy transmission system used in the previous example. This will enable us to compare the system functions for minimum mean square error and for minimum probability of error.

Since all the pulses are independent of each other, the transducer will not make use of storage. The output can accordingly be represented by

$$y(t) = \sum_{i=0}^{M} A_i \left[w(t)\right]^i \quad . \tag{52}$$

The input can take on only seven different values: +3, +2, ..., −3, and the output must take on a unique value for each input. Because the input-output relation must be skew-symmetrical, and go through three prescribed points on each side of the origin, it is clear that Eq. 52 must contain only odd powers of w, and need have no terms of degree higher than the fifth; i.e., M = 5. The design equations are

$$\phi_{wd}(\gamma) = \sum_{\alpha=1}^{M} A_\alpha \phi_{ww}(\alpha,\gamma), \quad \gamma = 1, 2, \ldots, M \quad . \tag{53}$$

Making use of Eqs. 45 and 46 we compute

$$\phi_{wd}(\gamma) = \overline{w^\gamma y_d}; \quad \phi_{wd}(0) = \phi_{wd}(2) = \phi_{wd}(4) = \cdots = 0,$$

$$\phi_{wd}(1) = \frac{2}{3}, \quad \phi_{wd}(3) = 2, \quad \phi_{wd}(5) = 14 \quad .$$

Also

$$\phi_{ww}(\alpha,\gamma) = \overline{w^\alpha \; w^\gamma}, \quad \overline{w} = \overline{w^3} = \overline{w^5} = \cdots = 0,$$

$$\overline{w^2} = \frac{4}{3}, \quad \overline{w^4} = 6, \quad \overline{w^6} = \frac{134}{3}, \quad \overline{w^8} = \frac{1138}{3}, \quad \overline{w^{10}} = 3338 \quad .$$

Substituting in Eq. 53, we obtain:

$$A_0 = A_2 = A_4 = 0$$

$$A_1 \frac{4}{3} + A_3 6 + A_5 \frac{134}{5} = \frac{2}{3}$$

$$A_1 6 + A_3 \frac{134}{3} + A_5 \frac{1138}{3} = 2$$

$$A_1 \frac{134}{3} + A_3 \frac{1138}{3} + A_5 3338 = 14 \quad . \tag{54}$$

The solution is:

$$A_1 = 1.20$$
$$A_3 = 0.485$$
$$A_5 = 0.0424 \quad , \qquad (55)$$

so that the equation of the optimum mean square transducer is

$$y = 1.26\ w - 0.495\ w^3 + 0.0424\ w^5 \quad . \qquad (56)$$

A plot of Eq. 56 is given in Fig. 18. It is clear that the probability of error is increased as compared with the earlier design, since for the ambiguous inputs +1 and −1, the correct output is never obtained. The probability of error is in fact raised from 1/18 for the earlier design to 11/18 for the present design. On the other hand, the mean square error is reduced from 2/9 = 22/99 for the earlier design to 20/99 for the present design.

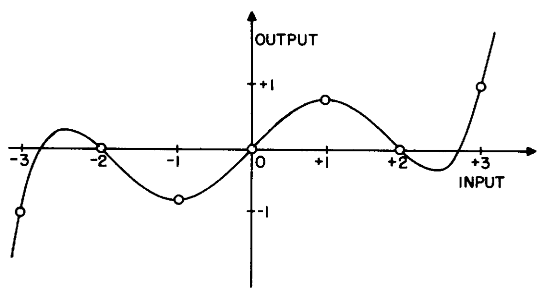


Fig. 18 Response for minimum mean square error.

A transducer having the system function of Eq. 56 is easily constructed if circuits for raising voltages to the third and fifth powers are available. The result would be as shown in Fig. 19. But inasmuch as the transducer is required to function only for integer inputs ranging over (+3, −3), the form shown in Fig. 20 can be used. The pulse amplitude distributor can be of the crystal diode type shown in Fig. 4, or a modification of that circuit. Because amplitude distributors are usually easier to construct than circuits for raising voltages to powers, the arrangement of Fig. 20 may be preferable.

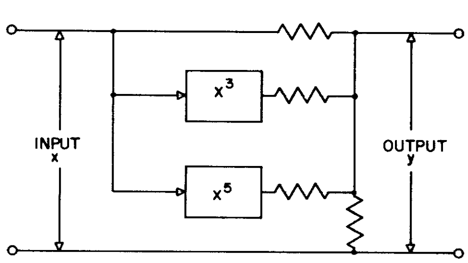


Fig. 19 Nonlinear transducer with response of Fig. 18.

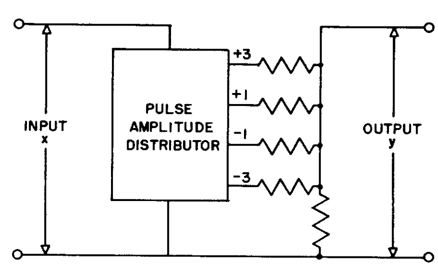


Fig. 20 Nonlinear transducer with response of circled points of Fig. 18.

5.6 Optimum Mean Square Filter for Gaussian Signals

It seems appropriate at this point to examine the problem of optimum mean square filter design from a slightly different point of view. We shall obtain equations which, while not always convenient to apply physically, nevertheless throw additional light on the relation of the filter to the probability distributions of the input time series. In particular we shall show that if the given signal and noise have gaussian probability distributions, the best mean square filter is linear, as has been pointed out by Wiener (3).

The general multi-dimensional gaussian distribution of the set of variables $w_1, \ldots, w_s$ may be written (21)

$$p([w]) = \frac{1}{(2\pi)^{s/2}\sqrt{A}} \exp\left[-\frac{1}{2A} \sum_{j,k} A_{jk} w_j w_k\right] \tag{57}$$

where A is the determinant of the moment matrix $[A]$:

$$[A] = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1s} \\ \cdots & \cdots & \cdots \\ \lambda_{s1} & \cdots & \lambda_{ss} \end{bmatrix} \quad . \tag{58}$$

A_{jk} is the cofactor of the element λ_{jk} of A, and

$$\lambda_{ii} = \sigma_i^2, \quad \lambda_{ik} = \rho_{ik} \sigma_i \sigma_k \quad . \tag{59}$$

In Eq. 59 σ_i is the standard deviation of the random variable w_i, and ρ_{ik} is the correlation coefficient between w_i and w_k. If all the variables $w_1, \ldots, w_s$ belong to the same time series f(t), all the σ_i are equal; and if $w_i(t)$ is separated from $w_j(t)$ by the time interval $(i-j)\tau$, where τ is a constant, then $[A]$ becomes the correlation matrix, i.e.,

$$[A] = \begin{bmatrix} \phi(0) & \phi(\tau) & \cdots & \phi[(s-1)\tau] \\ \phi(\tau) & \phi(0) & \cdots & \phi[(s-2)\tau] \\ \cdots & \cdots & \cdots & \cdots \\ \phi[(s-1)\tau] & \phi[(s-2)\tau] & \cdots & \phi(0) \end{bmatrix} \quad . \tag{60}$$

Eq. 57 shows the general gaussian distribution to be defined completely in terms of a matrix $[A]$, which, in case Eq. 57 is the distribution of a single time series f(t), is in turn completely specified by the autocorrelation function $\phi(k\tau)$, $k = 0, 1, \ldots, s-1$. It is thus clear that a knowledge of the autocorrelation function will permit the calculation of any other statistical parameter of f(t).

In case two functions are involved, the only significant change in the character of $[A]$ is that some of the elements become crosscorrelations. Since the only pertinent statistical parameters in such cases are the correlation functions, and since the correlation functions are sufficient to specify the optimum mean square linear filter, it might be expected that for gaussian signal and noise the optimum mean square filter would not contain any nonlinearity. That this is in fact true is verified below.

We first note that if the probability distribution p(x) of a random variable x is given, the best mean square estimate of x is $\overline{x}$; for the mean square error due to an estimate y is

$$E = \overline{e^2} = \int_{-\infty}^{\infty} (y - x)^2\, p(x)\, dx \tag{61}$$

and minimizing with respect to y we find

$$\frac{dE}{dy} = 2\int_{-\infty}^{\infty} (y - x)\, p(x)\, dx = 0 \quad . \tag{62}$$

Thus

$$y \int_{-\infty}^{\infty} p(x)\, dx = \int_{-\infty}^{\infty} x\, p(x)\, dx \tag{63}$$

$$y = \overline{x} \quad . \tag{64}$$

We now consider the design of an optimum mean square filter for which the desired output $f_d(t)$ is equal to the value of one of the members of [u], say u_d. Let the joint distribution of [w] and u_d be gaussian, and let a new variable [z] be defined by

$$z_i = w_i, \quad i = 1, \ldots, s$$

$$z_{s+1} = u_d \quad . \tag{65}$$

Let [z] have a moment matrix [A]. Then

$$p([z]) = \frac{1}{(2\pi)^{\frac{s+1}{2}} \sqrt{A}} \exp\left[-\frac{1}{2A} \sum_{j,k=1}^{s+1} A_{jk}\, z_j\, z_k \right] . \tag{66}$$

According to Eq. 64, the best mean square estimate of $u_d = z_{s+1}$ at any time when [w] $= z_1, \ldots, z_s$ takes on a particular value is given by the conditional mean of $u_d = \overline{z_{s+1}}$ under this condition. But the conditional mean of z_{s+1} for a particular [w] is determined from the conditional probability distribution $p(z_{s+1} : z_1, \ldots, z_s)$. Thus

$$p(z_{s+1} : z_1, \ldots, z_s) = \frac{\exp\left[-\frac{1}{2A} \sum_{j,k=1}^{s+1} A_{jk}\, z_j\, z_k \right]}{\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2A} \sum_{j,k=1}^{s+1} A_{jk}\, z_j\, z_k \right] dz_{s+1}} \tag{67}$$

$$= \frac{\exp\left[-\frac{1}{2A}\left(A_{s+1,\,s+1}\; z_{s+1}^2 + 2\,z_{s+1}\sum_{j=1}^{s} A_{j,\,s+1}\; z_j\right)\right]}{\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2A}\left(A_{s+1,\,s+1}\; z_{s+1}^2 + 2\,z_{s+1}\sum_{j=1}^{s} A_{j,\,s+1}\; z_j\right)\right] dz_{s+1}} \quad . \tag{68}$$

Eq. 68 also represents a gaussian distribution. It may be put in standard form by multiplying both numerator and denominator by a constant which will complete the square of the expression in the exponential. The numerator becomes

$$\exp\left[-\frac{A_{s+1,\,s+1}}{2A}\left(z_{s+1} + \sum_{j=1}^{s}\frac{A_{j,\,s+1}}{A_{s+1,\,s+1}}\; z_j\right)^2\right] \quad . \tag{69}$$

The conditional mean of z_{s+1} is therefore given by

$$\overline{z_{s+1}} = -\sum_{j=1}^{s}\frac{A_{j,\,s+1}}{A_{s+1,\,s+1}}\; z_j \quad . \tag{70}$$

Or, writing $f_o(t)$ for the output of the filter

$$f_o(t) = \frac{-1}{A_{s+1,\,s+1}}\sum_{j=1}^{s} A_{j,\,s+1}\; w_j(t) \quad . \tag{71}$$

This shows that in this case the optimum mean square filter is linear.

Although Eq. 71 was derived for gaussian distributions of signal and noise, it also specifies the optimum linear filter for any distributions. This is because the optimum linear filter depends directly on the correlation functions, so that combinations of signal and noise which have different sets of probability distributions, but the same set of correlation functions, result in the same optimum linear filter. Eq. 71 is in fact the same result which is obtained by solving the Wiener-Hopf equation for the case we are considering.

In the general case we have from Eq. 65 for the equation of the optimum mean square filter

$$f_o = \frac{\int_{-\infty}^{\infty} u_d\, p(u_d,\, w_1,\, \ldots,\, w_s)\, du_d}{\int_{-\infty}^{\infty} p(u_d,\, w_1,\, \ldots,\, w_s)\, du_d} \quad . \tag{72}$$

5.7 Finite State Filters

If one assumes at the outset that a finite state transducer is to be used, as usually seems best for pulse communication systems, two convenient methods for handling the problem suggest themselves. One method is especially convenient for unquantized signals, and the other appears to be more natural for quantized signals. In the first method the space of the transducer is assumed to be broken up into regions in a predetermined way, for example into cells whose boundaries are perpendicular to coordinate axes. The problem is then to determine a number to assign to each cell, such that if the output of the transducer equals this number when the input lies in the corresponding cell, the performance of the transducer is optimum in some prescribed statistical sense. The advantage of this method is that the transducer is easily synthesized in the standard form of Sect. II. The disadvantage is, of course, that while the transducer is the best of all transducers having the prescribed call arrangement, some other transducer with a different cell arrangement might be still better. Improvement of this kind will not be important if the number of cells are sufficiently large.

In the second method the arrangement and location of the regions are not specified in advance, but the numbers which are available for assignment to the various regions are assumed to be known, since the signal is quantized. The problem is then to specify the boundaries for each region in such a way as to give optimum performance in a prescribed statistical sense. This method does not suffer from the disadvantage mentioned for the first method.

We now outline briefly two procedures suitable for designing according to the first method. Because our principal interest is in quantized signals, we will then go on to discuss in detail a procedure for handling the second method.

Let the condition that a point $[w] = (w_1, \ldots, w_s)$ lie within the ith cell in w space be that $w_{11} < w_1 < w_{12}$; $w_{21} < w_2 < w_{22}$; ...; $w_{s1} < w_s < w_{s2}$. Let $f\left[f_m(t)\right] = f_d(\, u(t)\,)$, and let y_i be the number assigned to the ith cell. A sketch of the arrangement is shown in Fig. 21. The mean square error due to the labelling of the ith cell is

$$E_i = \int_{\text{cell } i} \int_{\text{all } u} \left[y_i - f_d([u])\right]^2 p([u], [w])\, d[u]\, d[w] \quad . \tag{73}$$

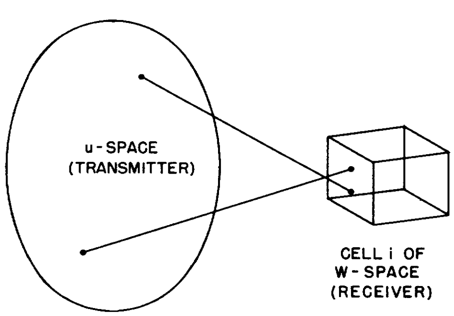


Fig. 21 Schematic representation of two different transmitted signals which are received in the same cell of w space.

Since the cells are fixed we may minimize the error due to each cell separately, and the total error

$$E = \sum_{i} E_i \tag{74}$$

will also be a minimum. In order for E_i to be a minimum we must have

$$\frac{\partial E_i}{\partial y_i} = 0 \quad .$$

Thus

$$\frac{\partial E_i}{\partial y_i} = \int_{\text{cell } i} \int_{\text{all } u} 2 \left[yi - f_d(\left[u\right]) \right] p(\left[u\right], \left[w\right]) \, d\left[u\right] \, d\left[w\right] \tag{75}$$

so that

$$\int_{\text{cell } i} \int_{\text{all } u} y_i p(\left[u\right], \left[w\right]) \, d\left[u\right] d\left[w\right] = \int_{\text{cell } i} \int_{\text{all } u} f_d(\left[u\right]) \, p(\left[u\right], \left[w\right]) \, d\left[u\right] d\left[w\right] \quad . \tag{76}$$

Hence

$$y_i \, p(w_i) = \int_{\text{cell } i} \int_{\text{all } u} f_c(\left[u\right]) \, p(\left[u\right]) \, p(\left[w\right] : \left[u\right]) \, d\left[u\right] d\left[w\right]$$

$$= \int_{\text{all } u} f_d(\left[u\right]) \, p(w_i : \left[u\right]) \, p(\left[u\right]) \, d\left[u\right]$$

$$= \overline{f_d(\left[u\right]) \, p(w_i : \left[u\right])}, \tag{77}$$

where $p(w_i)$ is the probability that a received point lie in the ith cell, and $p(w_i : [u])$ is the probability that a transmitted point $[u]$ be received in the ith cell. The result is

$$y_i = \frac{\overline{f_d(\left[u\right]) \, p(w_i : \left[u\right])}}{p(w_i)}, \quad i = 1, \ldots, M \quad . \tag{78}$$

The other procedure mentioned for finding the optimum y_i for predetermined cells makes use of a general series expansion. Setting

$$y(t) = F(\left[w\right]) = \sum_{\alpha=1}^{M} A_\alpha \, W_\alpha(t) \quad , \tag{79}$$

we find the optimum A's by the method already developed. By averaging y(t) over each cell i, the optimum average output for each cell is obtained. Thus

$$y_i = \frac{1}{p(w_i)} \int_{w_{11}}^{w_{12}} \cdots \int_{w_{s1}}^{w_{s2}} F([w])\, p([w])\, dw_1 \ldots dw_s \quad . \tag{80}$$

5.8 Selection of Optimum Boundaries

We next consider the problem of specifying the state defining regions in such a way as to give an optimum result in some prescribed statistical sense, when the signal is quantized. Let $[u]_i = [u_1, u_2, \ldots, u_s]_i$ be the transmitted signal, where M is the number of possible transmitted signals and $i = 1, \ldots, M$. The noise will in general not be quantized, so the received noise plus signal will also be unquantized. Let the desired output of the transducer for each transmitted $[u]_i = f([u]_i)$. From the known probability distribution for $[u]_i$, $[v]$, and $[w]$, we can find the joint distribution $p([w] : [u]_i)\, d[w]$ is the probability that a transmitted signal $[u]_i$ is received in the volume element $d[w]$ of w space centered at $[w]$.

The problem is now to select M non-overlapping regions R_i, $i = 1, \ldots, M$, in w space under the condition that the output $f([u]_i)$ is produced for any point $[w]$ received in the region R_i, in such a way as to optimize some statistical criterion of performance, under the constraint that $\sum_i R_i$ includes all of w space. We shall select the regions in such a way as to minimize the probability of error.

The probability that a transmitted signal $[u]_i$ is received in R_i is given by

$$\int_{R_i} p([w] : [u]_i)\, d[w] \quad , \tag{81}$$

so that the probability that $[u]_i$ is received in some other region is

$$1 - \int_{R_i} p([w] : [u]_i)\, d[w] \quad . \tag{82}$$

This last expression is therefore the relative frequency with which the transmitted $[u]_i$ are incorrectly evaluated at the receiver. Multiplying this by p_i, the relative rate of transmission of $[u]_i$, gives the relative frequency of errors due to transmitting $[u]_i$. By summing over all i we find the total probability of error,

$$p(E) = \sum_i p_i \left[1 - \int_{R_i} p([w] : [u]_i)\, d[w] \right]$$

$$= 1 - \sum_i \int_{R_i} p_i\, p([w] : [u]_i)\, d[w] \quad . \tag{83}$$

In this formulation, a region R_i need not be continuous; that is, R_i may consist of several separate sub-regions. Our objective is not to obtain equations defining the boundaries of each R_i explicitly, but instead to obtain design equations which will enable us to determine in which region a received signal plus noise falls. The design equations will then permit the transducer to be constructed. With this end in view, we now show that for p(E) to take on its minimum value, it is necessary and sufficient that a point $[w]$ belong to R_i, $i = 1, \ldots, M$, if

$$p_i\, p([w] : [u]_i) > p_j\, p([w] : [u]_j) \text{ for all } j \neq i \quad . \tag{84}$$

To show the necessity of the condition, we assume that the probabilities are continuous in $[w]$, that p(E) is a minimum, and that the neighborhood of a certain point $[w]$ belongs to R_i. We now ask whether there is a j such that

$$p_i\, p([w] : [u]_i) < p_j\, p([w] : [u]_j) \quad . \tag{85}$$

That there is no such j follows from the fact that $p_i\, p([w] : [u]_i)$ is the integrand of the integral in Eq. 83; so that if such a j existed the value of the summation would be increased by reassigning some neighborhood of $[w]$ to R_j. This would contradict the initial assumption, and shows the necessity of the condition.

To show the sufficiency of the condition we assume that each received point $[w]$ belongs to R_i, $i = 1, \ldots, M$, if $p_i\, p([w] : [u]_i) \geqslant p_j\, p([w] : [u]_j)$ for all $j \neq i$. We now divide the inequality by $p([w])$, obtaining

$$\frac{p_i\, p([w] : [u]_i)}{p([w])} > \frac{p_j\, p([w] : [u]_j)}{p([w])} \text{ for all } j \neq i \tag{86}$$

or

$$p([u]_i : [w]) > p([u]_j : [w]) \text{ for all } j \neq i \quad . \tag{87}$$

That is, the probability that $[u]_i$ was transmitted if $[w]$ was received is greater than or equal to the probability that any other point $[u]_j$ was transmitted. Therefore, by assigning $[w]$ to R_i, we minimize the probability of error as far as this particular $[w]$, is concerned. The same procedure is followed for each received point $[w]$, so that the total error p(E) is also minimized.

The set of equations

$$p_i\, p([w] : [u]_i) - p_j\, p([w] : [u]_j) > 0 \text{ for all } j \neq i \tag{88}$$

and $i = 1, \ldots, M$ are the design equations of the transducer. The process of designing the optimum filter consists primarily of designing equipment which evaluates the expressions

$$p_i\, p([w] : [u]_i) - p_j\, p([w] : [u]_j) \tag{89}$$

for all i and $j(j \neq i)$ each time a point $[w]$ is received. This equipment can usually be put in the form of amplitude discriminators, as will be shown. The results of the evaluations are led to M coincidence circuits, one of which corresponds to each of the regions R_i. Any received point $[w]$ will cause one of the coincidence circuits to produce an output, according to the region in which $[w]$ lies. The output of the ith coincidence circuit, corresponding to the region R_i, is adjusted to equal $f([u]_i)$, $i = 1, \ldots, M$.

Under suitably restrictive conditions on the noise, it may be shown that the probability of obtaining more than one output for an arbitrary value of $[w]$ is zero. For example, if the noise has a probability distribution

$$\phi([v]) = \phi(v_1, \ldots, v_s) \tag{90}$$

which is continuous in the v_i, $i = 1, \ldots, s$, we have that the boundary of any region R_i consists of sections from a set of surfaces of the form

$$p_i \, p([w] : [u]_i) - p_j \, p([w] : [u]_j) = 0 \tag{91}$$

or

$$p_i \, \phi([w] - [u]_i) - p_j \, \phi([w] - [u]_j) = 0 \quad . \tag{92}$$

For each pair of values of i, j, Eq. 92 represents a surface in w space, and the ratio of the measure of the set of points on a section of the surface to the measure of the set of points in any finite volume of the same space is zero (21). The ratio of the probability that an arbitrary point $[w]$ falls in the volume to the probability that it falls on the surface is just the ratio of the integrals of the probability distribution over the two corresponding sets of points. Because of the assumed continuity of the distributions, the probability function cannot be concentrated on the surface, so that the ratio must be zero. But more than one output can be obtained only if $[w]$ falls on a surface. Hence the probability of more than one output is zero.

5.9 Example: Optimum P.C.M. Decoder

As a first example we choose the case of a two-dimensional transducer, since the result may conveniently be shown on a two-dimensional graph. A two-dimensional transducer will generally be necessary for detecting or filtering code groups which consist of two pulses per group. We assume therefore that our signal consists of a sequence of pulses uniformly spaced by a time interval 1/2W which can take the values +E or − E at the receiver, and that the signal is transmitted in code groups consisting of pairs of pulses over a transmission channel having a bandwidth W. This would correspond to a simple P.C.M. system. Let the code groups be governed by the following set of probabilities:

$$p(+E, +E) = a \qquad p(+E, -E) = b$$
$$p(-E, -E) = a \qquad p(-E, +E) = b \tag{93}$$

where $2(a + b) = 1$, and let each code group be independent of all the others.

The transmitted pulses are assumed to be additively perturbed by a white noise which is also band-limited to the bandwidth W, so that samples of the noise which are separated by the interval 1/2W are independent and have a gaussian distribution. Let the power in the received noise equal N. Then the joint probability distribution of the two noise voltages v_1 and v_2 which add to the two signal pulses in a code group is given by

$$p(v_1, v_2) = \frac{1}{2\pi N} \exp\left[-\left(\frac{v_1^2}{2N} + \frac{v_2^2}{2N}\right)\right] . \tag{94}$$

According to Eq. 84, the optimum design for minimizing the probability of error in the ith code group is found from the condition that

$$p_i\, p([w] : [x]_i) > p_j\, p([w] : [x]_j) \text{ for all } j \neq i \quad . \tag{95}$$

Let us set

$$\begin{aligned} [x]_1 &= (+E, +E) \\ [x]_2 &= (+E, -E) \\ [x]_3 &= (-E, +E) \\ [x]_4 &= (-E, -E) \quad . \end{aligned} \tag{96}$$

Then

$$\begin{aligned} p([w] : [x]_1) &= \frac{1}{2\pi N} \exp\left[-\frac{(w_1 - E)^2 + (w_2 - E)^2}{2N}\right] \\ p([w] : [x]_2) &= \frac{1}{2\pi N} \exp\left[-\frac{(w_1 - E)^2 + (w_2 + E)^2}{2N}\right] \\ p([w] : [x]_3) &= \frac{1}{2\pi N} \exp\left[-\frac{(w_1 + E)^2 + (w_2 - E)^2}{2N}\right] \\ p([w] : [x]_4) &= \frac{1}{2\pi N} \exp\left[-\frac{(w_1 + E)^2 + (w_2 + E)^2}{2N}\right] \quad . \end{aligned} \tag{97}$$

From Eq. 95 the set of conditions under which a received point $[w] = (w_1, w_2)$ belongs to R_1 is as follows:

$$\begin{aligned} \frac{a}{2\pi N} \exp\left[-\frac{(w_1 - E)^2 + (w_2 - E)^2}{2N}\right] &> \frac{b}{2\pi N} \exp\left[-\frac{(w_1 + E)^2 + (w_2 - E)^2}{2N}\right] \\ &> \frac{b}{2\pi N} \exp\left[-\frac{(w_1 - E)^2 + (w_2 + E)^2}{2N}\right] \\ &> \frac{a}{2\pi N} \exp\left[-\frac{(w_1 + E)^2 + (w_2 + E)^2}{2N}\right] \quad . \end{aligned} \tag{98}$$

Simplifying, the conditions become:

$$\log a + E\frac{w_1}{N} + E\frac{w_2}{N} > -E\frac{w_1}{N} + E\frac{w_2}{N} + \log b$$

$$> E\frac{w_1}{N} - E\frac{w_2}{N} + \log b$$

$$> -E\frac{w_1}{N} - E\frac{w_2}{N} + \log a \qquad (99)$$

or

$$w_1 > -\frac{1}{2}\left(\frac{N}{E}\right) \log \frac{a}{b}$$

$$w_2 > -\frac{1}{2}\left(\frac{N}{E}\right) \log \frac{a}{b}$$

$$w_1 + w_2 > 0 \quad . \qquad (100)$$

In a similar way it is found that the conditions that $[w]$ belong to R_2 are:

$$w_1 > \frac{1}{2}\left(\frac{N}{E}\right) \log \frac{a}{b}$$

$$w_2 < -\frac{1}{2}\left(\frac{N}{E}\right) \log \frac{a}{b}$$

$$w_1 - w_2 > 0 \quad . \qquad (101)$$

For R_3 the conditions are:

$$w_1 < -\frac{1}{2}\left(\frac{N}{E}\right) \log \frac{a}{b}$$

$$w_2 > \frac{1}{2}\left(\frac{N}{E}\right) \log \frac{a}{b}$$

$$w_1 - w_2 < 0 \quad . \qquad (102)$$

And for R_4 the conditions are

$$w_1 < \frac{1}{2}\left(\frac{N}{E}\right) \log \frac{a}{b}$$

$$w_2 < \frac{1}{2}\left(\frac{N}{E}\right) \log \frac{a}{b}$$

$$w_1 + w_2 < 0 \quad . \qquad (103)$$

All of the conditions 100-103 can be conveniently presented in graphical form, as shown in Fig. 22. The dashed lines divide w space into four regions, corresponding to the case $a/b > 1$. In this case the conditions $w_1 - w_2 > 0$ for R_2 and $w_1 - w_2 < 0$ for R_3 are included in the other two conditions on R_2 and R_3. With reference to the transducer diagram of Fig. 22, we may say that the function of the transducer is to recognize in which of the four regions a received signal plus noise falls, and to produce the correct output corresponding to that region.

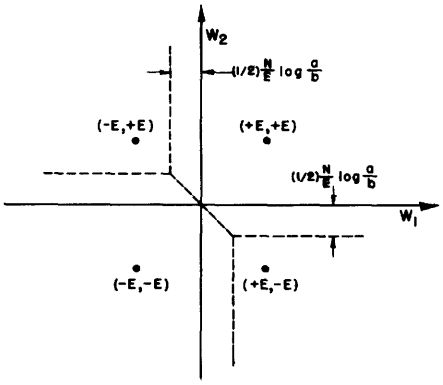


Fig. 22 Two-dimensional transducer diagram.

The transducer diagram corresponds to what one would intuitively expect to be best. If $a/b > 1$ is held fixed and N/E becomes very large, best results will be obtained by simply taking (+E, +E) when $w_1 + w_2 > 0$ and (−E, −E) when $w_1 + w_2 > 0$. On the other hand if N/E becomes very small, the diagram reduces to the four quadrants. This result would be expected in the case of a small noise, since the signal is then easily decoded without reference to the probabilities.

A transducer for carrying out the operation indicated graphically in Fig. 22 would comprise a number of comparison circuits for comparing the input signals with the prescribed reference voltages, and a set of coincidence circuits for indicating the region in which the voltage lies. There would be four such coincidence circuits, one for each region. A block diagram of the arrangement is shown in Fig. 23. Successive pairs of samples

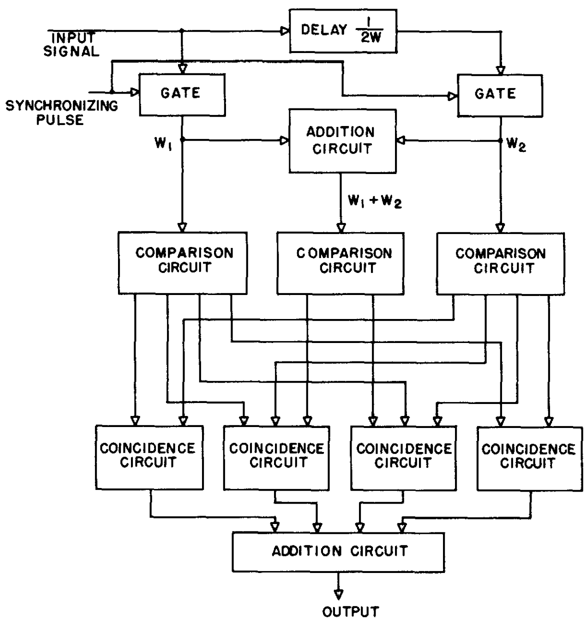


Fig. 23 Block diagram of two-dimensional transducer.

of signal plus noise are obtained by keying the gate tubes with synchronized timing pulses which may be obtained from the input signal. The voltages w_1, w_2, and $w_1 + w_2$ are then sent to separate comparison circuits. The comparison circuits produce output voltages on appropriate lines, as shown, in accordance with the magnitude of the sample voltages. Each coincidence circuit functions in such a way as to produce an output only when all the lines leading to it are keyed, and this will happen in accordance with the region in which a received signal plus noise lies. If the transducer is to function as a decoder, each coincidence circuit can be adjusted to have an output representing the particular code group to which it corresponds. If on the other hand it is desired to reproduce the original signal, as at a repeater station, this can be accomplished by keying suitable circuits with a trigger pulse derived from the coincidence circuits.

5.10 Example: Radar Search Problem

As a final example we discuss the application of the methods which have been developed to an important practical problem, namely that of detecting a target by radar in the presence of interference. The interference may be due to ground clutter or sea return, to jamming, or in the case of small or very distant targets, to random thermal noise originating in the first stages of the receiver. In practice targets are usually detected by an operator observing the pattern of received echos on a cathode ray tube screen. Experiments have shown that many targets are missed by the operator, either because he is observing some other part of the cathode ray tube screen when the target appears, or because of fatigue (targets are missed during the latter part of a long watch more often than during the early part).

In radar systems of the type we are considering, the behavior of the electrical part of the system is essentially linear: small or distant targets give very weak indications on the screen; larger targets give stronger indications. A nonlinearity is usually introduced by the human operator: if the target echo is extremely weak he does not observe the target at all, and takes no action. On the other hand, at least in the ideal case, if the target response passes a threshold value, the operator does observe it; and takes a decisive action. He may, for example, if the target is a submarine schnorkel, approach the target and drop a depth bomb. From this point of view the human operator functions as a nonlinear filter, with an all or nothing response, according as a target is or is not present. The response of the eye is apparently fairly linear, so the nonlinear response of the human observer is due to operations on the input signal which are carried on in the brain. Furthermore, the response of the human operator takes into account the statistics of the target return and of the interference. For even though the signal level is very weak, the statistical character of the target echo may be very different from that of the surrounding interference, and so permit the target to be identified by an experienced observer.

In view of experiments showing a large number of targets missed due to human failure, it is natural to inquire whether further electrical operation on the radar return might not make the human observer's job easier. In particular, it has been suggested that the necessary nonlinearity (which must exist in certain complete systems) be introduced ahead of the indicating device by means of an optimum nonlinear filter, in such a way that either an unmistakable signal is presented to the human operator, or no signal whatsoever appears on the indicator screen. If the highly complex human link is to be replaced in this way, it is to be expected that the nonlinear filter which takes its place should also be complex, at least relative to more conventional filters. The advantage to be gained would be freedom from oversight and fatigue.

We consider the particular case of an airborne radar searching for a submarine schnorkel. In order to carry through a design with the aid of relations 84, we shall assume a certain probability distribution for the return in the absence of a target (the so-called sea return), and another distribution for the return when a target is present. The distribution assumed for a single echo in the absence of a target is of the form

$$p(w : 0) = \frac{1 - c}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{(w - m_n)^2}{2\sigma_n^2}\right] + \frac{c}{\sqrt{2\pi}\sigma_o} \exp\left[-\frac{(w - m_o)^2}{2\sigma_o^2}\right] . \quad (104)$$

The first of the two normal distributions is due to thermal noise and has a relatively small σ_n and m_n, and takes account of the fact that waves sometimes mask sizeable areas behind them. The second of the two normal distributions is due to reflections from waves. If the sea were smooth, the sea return would be absent (c = o), and the limitation on range would be due to thermal noise in the receiver. The parameters c, σ_o, and m_o are functions of the coordinates with respect to the aircraft, of the elementary area under consideration, and will moreover vary with flight conditions and weather. These influences on the statistics of the sea return are taken into consideration by the human observer, and must be taken into consideration by the nonlinear filter as well. It is probable that the parameters can be continuously evaluated, in an approximate fashion, by an auxiliary computer associated with the filter, and the results can be fed continuously to the filter, varying as flight altitude, weather conditions, etc., change. This is, of course, the way a good human observer works. The auxiliary computer need not be located in the searching aircraft, but can be placed on land or on a ship, according to requirements.

In the multi-dimensional case, assuming the echos on successive scans to be independent of one another, which they clearly are for sea return, the joint distribution for s echos becomes:

$$p(w_1, \ldots, w_s : 0) = \prod_1^s \left[\frac{1 - c}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{(w_i - m_n)^2}{2\sigma_n^2}\right) + \frac{c}{\sqrt{2\pi}\sigma_o} \exp\left(-\frac{(w_i - m_o)^2}{2\sigma_o^2}\right)\right] . \quad (105)$$

The distribution of a single echo when a target is present will be assumed to be gaussian, with mean m_1, and standard deviation σ_1:

$$p(w : 1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{(w - m_1)^2}{2\sigma_1^2}\right] . \quad (106)$$

The parameters m_1 and σ_1 will also depend on the coordinates of the schnorkel with respect to the aircraft, and on the flight conditions and weather. They cannot be continuously evaluated during flight, and would therefore have to be evaluated experimentally for a wide range of conditions as part of the program of developing the nonlinear filter.

The joint distribution for s echos when a target is present becomes:

$$p(w_1, \ldots, w_s : 1) = \frac{1}{(2\pi)^{s/2} \sigma_1^s} \exp\left[-\sum_1^s \frac{(w_i - m_1)^2}{2\sigma_1^2}\right] . \tag{107}$$

In order to determine the optimum filter it is now necessary to adopt a criterion of performance. It might at first appear that a suitable criterion would be that of minimizing the probability of error. Equation 84 would be then be directly applicable, and would give, on the basis of s scans of an elementary area,

$$p_1 \, p(w_1, \ldots, w_s : 1) > p_o \, p(w_1, \ldots, w_s : 0) \tag{108}$$

as the design equation of the filter, where p_1 is the relative frequency with which targets appear in the elementary area under consideration, and $p_o = 1 - p_1$. That is, when the inequality 108 is satisfied, a target is judged to be present. Since $p_o >> p_1$, it is clear that 108 would almost never be satisfied, and the filter would indicate no target at almost all times. This result is in keeping with the criterion mentioned, and shows that another criterion should be chosen. The human observer has as his criterion to miss as few submarines as possible while searching as much area as possible. The reduction in area searched can be considered as coming about due to false alarms, which take time to investigate. If the cost of missing a submarine which is present is taken as a, and the cost of investigating a false alarm is taken as b, the criterion adopted by the human observer would be to minimize $e_1 a + e_o b$, where e_1 is the relative frequency at which targets are missed and e_o is the relative frequency at which false alarms are called. From Eq. 83 we find

$$e_1 a + e_o b = a p_1 \left[1 - \int_{R_1} p([w] : 1) \, d[w]\right] + b p_o \left[1 - \int_{R_o} p([w] : 0) \, d[w]\right] . \tag{109}$$

Since a and b are independent of R_o and R_1, the condition that a point belong to R_1 will be found in the same way as before, from Eq. 84, giving

$$a p_1 \, p(w_1, \ldots, w_s : 1) > b p_o \, p(w_1, \ldots, w_s : 0) . \tag{110}$$

Condition 110 can be met, as contrasted to condition 108, whenever the cost of missing a submarine is sufficiently greater than the cost of investigating a false alarm.

In the same way that $p([w] : 1)$ and $p([w] : 0)$ depend on the coordinates of the area under consideration, the values of a and b can vary with these coordinates. For if a target is out on the fringe of a search path, and we incorrectly say that no target is there, we have very little chance of rectifying the error. On the other hand, if the target is dead ahead and we miss it, it may be picked up later.

Substituting Eqs. 105 and 107 in Eq. 110 gives

$$\frac{ap_i}{\sigma_1^s} \exp\left[-\sum_1^s \frac{(w_i - m_1)^2}{2\sigma_1^2}\right] > bp_o \prod_1^s \left[\frac{1-c}{\sigma_n} \exp\left(-\frac{(w_i - m_n)^2}{2\sigma_n^2}\right) + \frac{c}{\sigma_o} \exp\left(-\frac{(w_i - m_o)^2}{2\sigma_o^2}\right)\right] . \tag{111}$$

The optimum nonlinear filter in this case is a computer. Its function is to evaluate the two sides of the inequality 111, and compare the results to determine which side is larger. If the left hand side is larger, a warning is given.

Condition 111 is based on statistics which define the coherence from pulse to pulse. In the same way it would be possible to take into consideration the statistics relating the returns produced by reflections of the same pulse when incident on different areas. It seems most convenient to do this in the design of the linear filters associated with the i-f and video amplifiers.

It may be expected that σ_o and σ_n will both be small compared to $m_n - m_o$. In this case, one of the two terms on the right side of 111 will usually be much larger than the other and the inequality may be written:

$$-\sum_1^s \frac{(w_i - m)^2}{2\sigma_1^2} > s \log \frac{bp_o \sigma_1}{ap_1} + k \log \frac{1-c}{n} + (s-k) \log \frac{c}{\sigma_o} - \sum_{(k)} \frac{(w_i - m_n)^2}{2\sigma_n^2} - \sum_{(s-k)} \frac{(w_i - m_o)^2}{2\sigma_o^2} \tag{112}$$

where k is the number of returns for which

$$\frac{1-c}{\sigma_n} \exp\left[-\frac{(w_i - m_n)^2}{2\sigma_n^2}\right] \quad \frac{c}{\sigma_o} \exp\left[-\frac{(w_i - m_o)^2}{2\sigma_o^2}\right] \tag{113}$$

and $\sum_{(k)}$ indicates summation over those terms for which 113 holds. Simplifying 112 gives

$$-\sum_{1}^{s}\frac{(w_i - m_1)^2}{2\sigma_1^2} > s\log\frac{bp_o\sigma_1}{ap_1\sigma_o} + k\log\frac{1-c}{c}\frac{\sigma_o}{\sigma_n}$$

$$-\sum_{(k)}\frac{(w_i - m_n)^2}{2\sigma_n^2} - \sum_{(s-k)}\frac{(w_i - m_o)^2}{2\sigma_o^2} \quad . \tag{114}$$

The computer which carries out the computation 114 may be of any convenient design, such as a modern high speed digital computer. It could be carried on a ship, for example, and operate on signals [w] received from an aircraft.

It is worth noting that in case c = 1, which might correspond to the case of a fairly smooth sea, 114 would become:

$$-\sum_{1}^{s}\frac{(w_i - m_1)^2}{2\sigma_1^2} > -\sum_{1}^{s}\frac{(w_i - m_o)^2}{2\sigma_1^2} + s\log\frac{bp_o\sigma_1}{ap_1\sigma_o} \quad . \tag{115}$$

If now in addition $\sigma_1 = \sigma_o$, we obtain

$$\sum_{1}^{s}\frac{w_i m_1}{\sigma_1^2} - \frac{sm_1^2}{2\sigma_1^2} > \sum_{1}^{s}\frac{w_i m_o}{\sigma_o^2} - \frac{sm_o^2}{2\sigma_1^2} + s\log\frac{bp_o}{ap_1} \quad . \tag{116}$$

Or rewriting:

$$\frac{1}{s}\sum_{1}^{s} w_i > \frac{m_1 + m_o}{2} + \frac{\sigma_1^2}{m_i - m_o}\log\frac{bp_o}{ap_1} \tag{117}$$

As would be expected in this simple case, relation 117 simply requires (when $m_1 > m_o$) that the average value of the return exceed the average of the two means plus an additional term which takes into account the probabilities p_o and p_1 and the costs a and b.

Conclusion

Principal emphasis in this study of nonlinear transducers has been placed on information processing devices intended for use in receivers, or in equipment similar to receivers. Thus the field of study parallels roughly the general area studied by Wiener in his book on optimum linear filters, with the difference that the devices which are optimized are allowed to be nonlinear. Within the limits of this field it has been found possible to handle the problem of determining optimum invariant devices for processing various kinds of signals of importance. Specific design equations are given in these cases. It has not been found possible to summarize the results in a closed form, as was done by Wiener for the linear case, but the methods which have been developed should prove applicable, with minor modifications, to new problems which may arise in the general area selected for study.

The results which have been obtained lead to equipment which is physically realizable in one or another of the standard forms which are proposed. It is possible that in many cases simpler equipment which produces the same results can be designed, and this possibility needs to be investigated. The problem involved is related to that of the logical design of computers.

The examples which have been discussed indicate that optimum nonlinear equipment tends to be relatively complex. Only in the case of gaussian signal and noise is the optimum filter found to reduce to a linear device. The examples also emphasize that a considerable body of statistical data must be obtained before optimum design can be carried out. Because of the tendency toward complexity, and also because of the requirement for large amounts of statistical data, it is not to be expected that optimum nonlinear devices will generally replace linear networks in jobs which are normally performed by linear networks. We may instead conclude that optimum nonlinear devices will probably first receive extensive use in processing information which is normally processed by a nonlinear device anyway. By making use of the nonlinear theory, an optimum system replaces what might otherwise have to be designed on an intuitive or a cut-and-try basis. Again, as has been discussed by Wiener in "Cybernetics", optimum transducer theory may make possible the design of transducers which will be good enough to release human beings from certain routine tasks. As illustrated by our example of the radar search problem, some tasks are now performed by human beings, not because the human being is especially suited to the job, but because other equipment which would do it has not been available.

Closely related to the problem which has been studied is that of the design of optimum coding equipment. The problem of optimum coding may in a sense be considered to include the problem of optimum reception; and in fact to constitute the problem of optimum communication system design. It is hoped that some of the techniques which have been employed in this thesis may prove useful in the study of this more general problem.

Acknowledgement

The author wishes to express his deep appreciation to Professor J. B. Wiesner, who directed this research and led it to its present form; to Professor R. M. Fano for much constructive criticism; to Professor Y. W. Lee, who introduced the author to this field, and whose work provides a basis for this research; and to Messrs. A. J. Lephakis and C. A. Stutt for many helpful suggestions and comments. The author is also greatly indebted to Professor Norbert Wiener, who by generously discussing his own research on nonlinear networks with the author, has been a source of inspiration and encouragement.

References

1. N. Wiener: The Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications, (Wiley and Sons, N.Y. 1949).
2. C. E. Shannon: B.S.T.J. 27, 3, 379-423 (1948); 27, 4, 623-656 (1948).
3. N. Wiener: Cybernetics (Wiley and Sons, N.Y. 1948).
4. R. M. Fano: R.L.E. Technical Report, No. 65, M.I.T. (1949).
5. C. E. Shannon: Proc. I.R.E. 37, 1, 10-21 (1949).
6. W. G. Tuller: Proc. I.R.E. 37, 5, 468-478 (1949).
7. R. M. Fano: R.L.E. Technical Report, No. 149, M.I.T. (1950).
8. R. L. Dietzhold: Elec. Eng., 67, 9, 895-899 (1948).
9. B. M. Oliver, J. R. Pierce, C. E. Shannon: Proc. I.R.E. 36, 11, 1324-1331 (1948).
10. M. J. E. Golay: Proc. I.R.E. 37, 9, 1031 (1949).
11. S. Darlington: J. Math. Phys. 18, 257-353 (1939).
12. R. M. Fano, A. W. Lawson, Jr. : The Theory of Microwave Filters, 9, 9, M.I.T. Radiation Laboratory Series, (McGraw Hill, N.Y. 1946).
13. Y. W. Lee: J. Math. Phys. 11, 83-113 (1932).
14. N. Wiener, Y. W. Lee: U.S. Patents No. 2,128,257, 2,024,900, and 2,124,599.
15. N. Wiener: Seminar in Nonlinear Networks, Research Laboratory of Electronics, February, 1949, Unpublished.
16. Y. W. Lee: Quarterly Progress Report, Research Laboratory of Electronics, M.I.T. p. 60, (January 15, 1950).
17. Y. W. Lee, J. B. Wiesner: Electronics 23, 6, 86-92 (1950).
18. Y. W. Lee: Course 6.563 Class Notes, Unpublished.
19. E. C. Titchmarsh: Introduction to the Theory of Fourier Integrals (Oxford University Press, London, 1937).
20. M. Frechet: Theorie des evenements en chaine, (Gauthier-Villars, Paris, 1938).
21. H. Cramer: Mathematical Methods of Statistics, (Princeton University Press, Princeton, N.J. 1946).