1974

Adaptive Nonlinear Modeling for Ultrasonic Signal Processing

Anthony N. Mucciardi
Adaptronics, Inc.

Adaptive Nonlinear Modeling for Ultrasonic Signal Processing

Abstract
The purpose of this presentation is to introduce to the NDE community an empirical modeling technique which has been under development during the past thirteen years and applied particularly to many problem areas in the last five. The application of interest to today's audience is that of classification of flaw geometry from ultrasonic signals.

Disciplines
Materials Science and Engineering | Structures and Materials

ADAPTIVE NONLINEAR MODELING FOR ULTRASONIC SIGNAL PROCESSING

Anthony N. Mucciardi
Adaptronics, Inc.
McLean, Virginia

The purpose of this presentation is to introduce to the NDE community an empirical modeling technique which has been under development during the past thirteen years and applied particularly to many problem areas in the last five. The application of interest to today's audience is that of classification of flaw geometry from ultrasonic signals.

The project in which we will be applying this methodology will first be described and then I will spend the remaining portion of the talk giving a brief summary of the technique itself. This project is an eighteen-month program with the NDE branch of the Air Force Materials Laboratory. It has just begun six weeks ago.

The main program objectives are to evaluate the efficacy of this particular signal processing methodology for characterization of material flaw descriptors. Our program consists of two tasks. Task 1 is to demonstrate the capability of classifying flatbottom holes in the range of 0 to 8/64ths (in steps of 1/64th) from ultrasonic signatures. The second task is to infer fatigue crack length over a range of 0 to 250 mils.

Other objectives of the program involve assessing the information content of ultrasonic NDT signals. The returned ultrasonic signal will be parameterized using different types of parameter domains that I will describe shortly. We will be interested in identifying those parameters that contain the most discriminatory information relative to distinguishing between different hole sizes or estimating different crack lengths.

Once models have been constructed using this empirical technique, we will determine the model's sensitivity with respect to each of these best found parameters. These results should establish correlations between ultrasonic waveform parameters and parameter combinations and the physical phenomena itself. For example, why might a high frequency component parameter and a particular waveform shape parameter turn out to be highly informative for inferring crack lengths?

As I mentioned earlier, our program consists of two tasks. The first one is to demonstrate the capability of this methodology for characterizing flatbottom holes. We will use two sets of specimens with nine blocks each--one with no hole, and the other eight with holes of diameters from 1/64" to 8/64" in steps of 1/64". The transducer will be oriented axially, straight down the diameter of the hole. A minimum of ten to twenty shots will be taken per test block. Each time the equipment will be disassembled and reassembled.

In the second task, fatigue cracks will be grown from quarter-inch drilled holes as shown in Fig. 1. A five MHz broad-band transducer will be used along with a Biomation 8100 transient recorder which is capable of recording up to 20 MHz signals. The recorder will be interfaced to a Data General Supernova minicomputer from which we will extract a digital record which is the resultant ultrasonic signal for a particular shot. The signals will then be suitably parameterized and analyzed. The material used in both tasks is 7075-T6 aluminum. In this second task, the transducer will be situated at an angle relative to the plate and approximately perpendicular to the crack. The transducer orientation will be varied by plus and minus 10 degrees as shown in Fig. 1 in order to take into account the variability problems in field setup conditions.

The modeling synthesis procedure that will be used in this project is one which Adaptronics has developed over the past thirteen years. It is called a nonlinear adaptive learning network and it represents a very

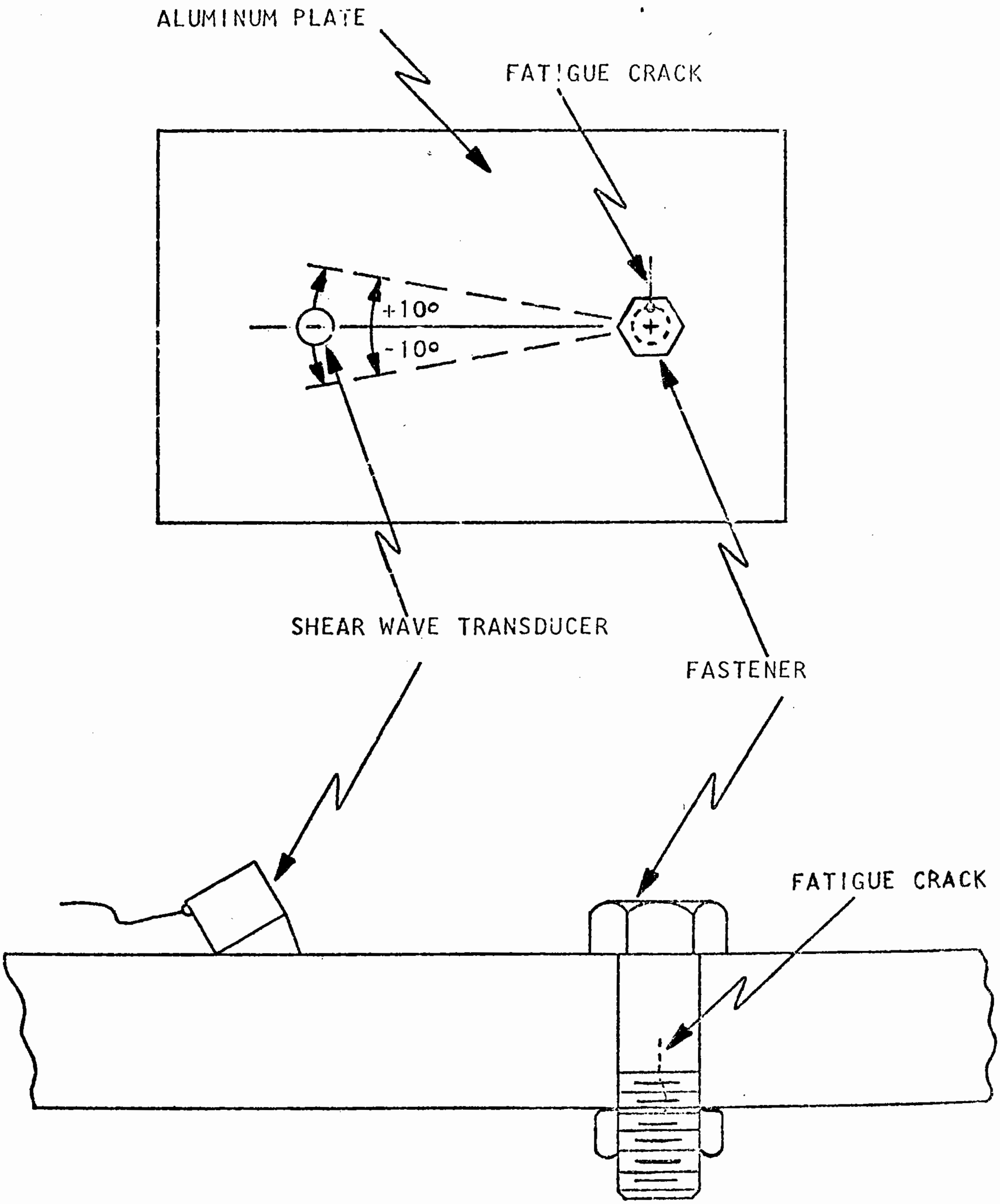


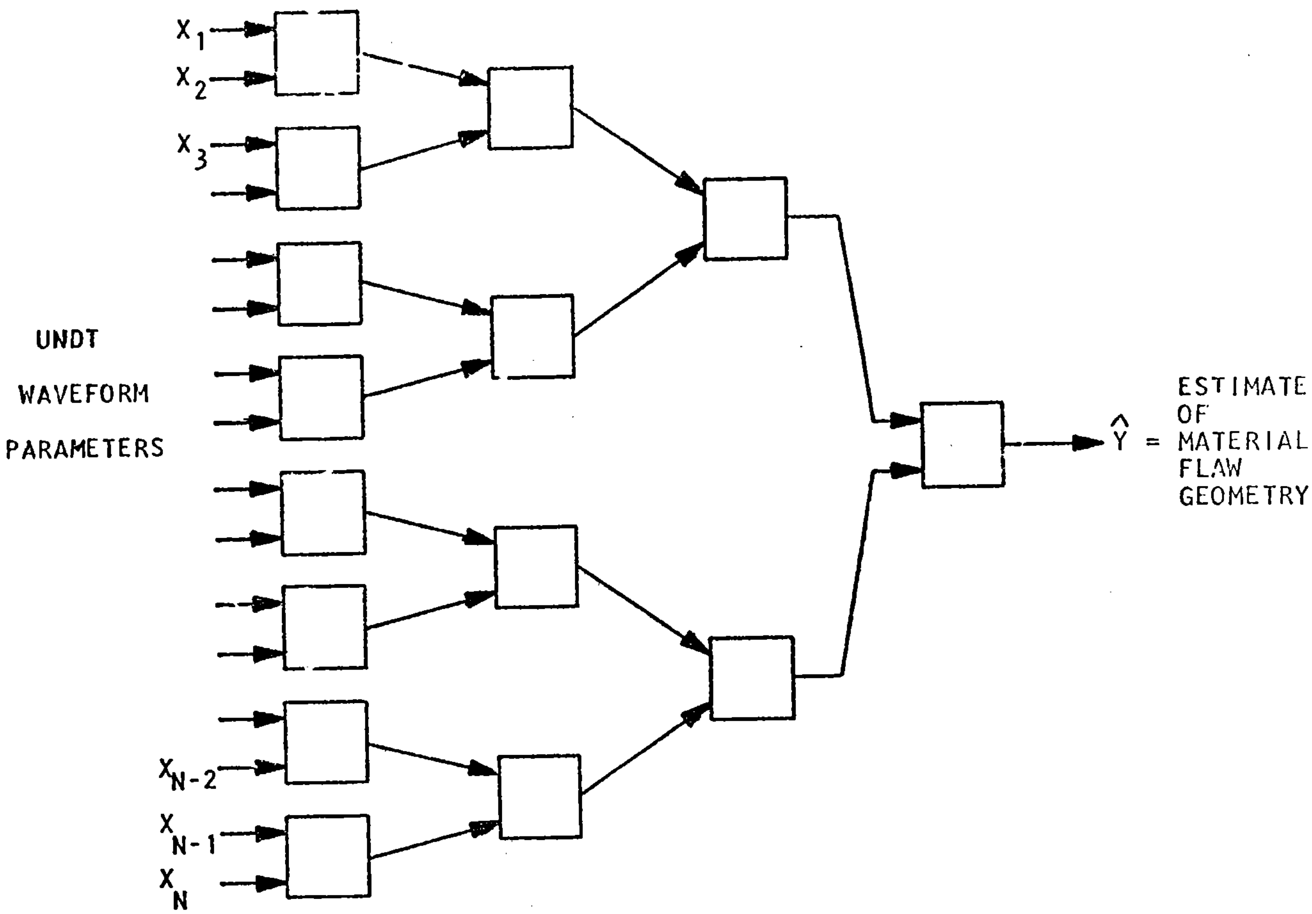
Fig. 1. Fatigue Crack Specimens

interesting modeling methodology. The idea is to construct from a data base an empirical model which considers both linear and nonlinear interactions of ultrasonic waveform parameters in such a way that estimates of material flaw geometry can be inferred with high accuracy.

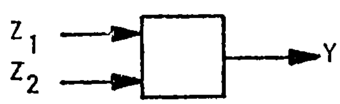
The model, in effect, is a computing network which is adaptively "grown". It consists of the elements shown in Fig. 2. As you can see, each element in this network considers only a pair of the waveform parameters and performs the second degree nonlinear transformation shown at the bottom of Fig. 2. The idea in this particular type of transformation accounting for those linear and nonlinear interactions necessary for accurate estimates of the variable to be modeled. This particular methodology is an empirically-based technique in which, as its synthesis proceeds, the variables which are most informative are automatically identified. A screening procedure takes place in the first layer of computation in which only those parameters which contain flaw geometry discrimination information are retained. The other less discriminating parameters are filtered out.

The elements in the network all have the same basic structure as shown in Fig. 3. Each element considers a pair of inputs, Xi and Xj. These can be the original ultrasonic parameters, such as frequency terms or other waveform parameters, or they could be outputs from a previous layer (which are now <u>functions</u> of the X's).

The way the model is synthesised is as follows. Let us assume we have a data base collected. We have recorded signals of various crack lengths in specimens for which the crack length was known. Assume that we took 20 records for each specimen of a given crack length and that each record was made after the equipment had been disassembled and reassembled. Let us keep ten of the 20 records for synthesizing our model and use the other ten as an independent testing set to test the model's accuracy on data not used in the design phase. Therefore,

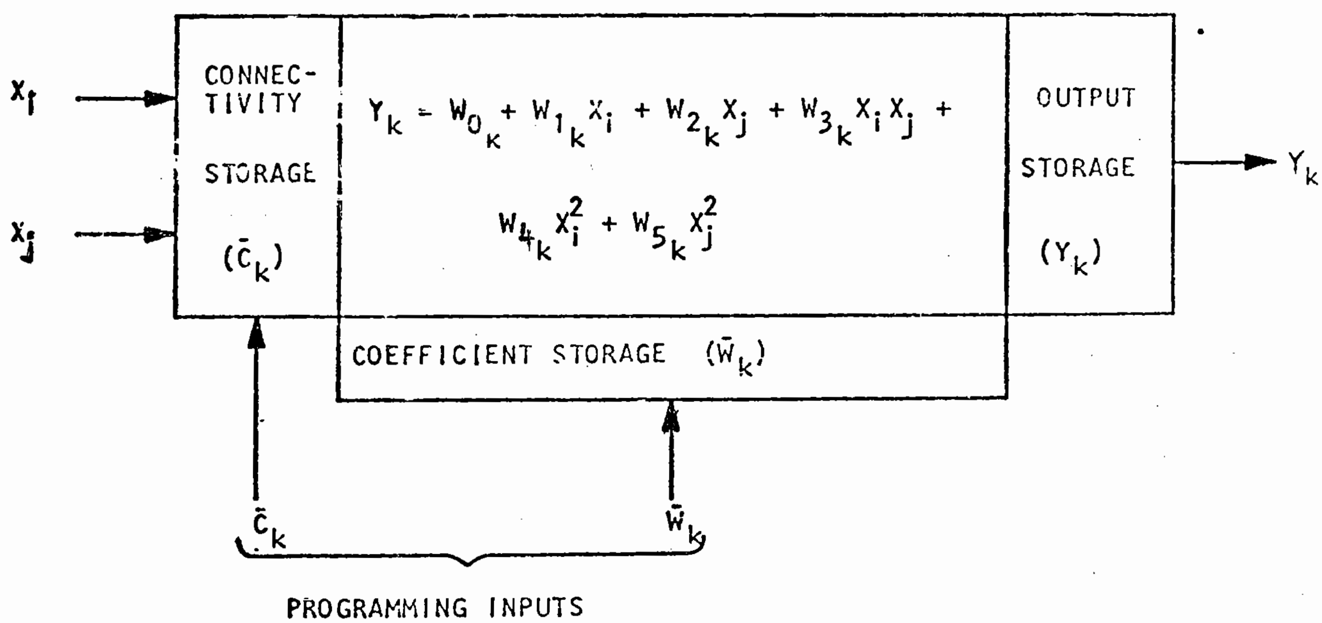


GENERALIZED 2-INPUT ELEMENT USED IN THE ABOVE NETWORK



$$Y = W_o + W_1Z_1 + W_2Z_2 + W_3Z_1Z_2 + W_4Z_1^2 + W_5Z_2^2$$

Fig. 2. Illustration of Nonlinear Adaptive Learning Network that Estimates One Measure of Material Flaw Geometry



- LINEAR TERMS
- CROSS-PRODUCT TERM
- SQUARE TERMS

Fig. 3 Hypercomp™ Polynomial Transformation Programmable Building-Block Element ("Primitive") we have one waveform recorded for each specimen and for each shot of each specimen. A set of parameters is computed from each record. Initially, a reasonably exhaustive list of parameters is computed. Such signal parameters as power spectrum, cepstrum, auto- and cross-correlation, and shape parameters (which are weighted integrals) will be computed. The idea behind this empirical modeling technique is to be as exhaustive as possible initially in the selection of parameters and to let the technique give us information about where the information is and where it is not.

Consider an example. Perhaps we initially generate a candidate list of N parameters. First of all, using the second degree nonlinear transformational element of Fig. 3, we evaluate how informative the first pair of parameters X_1 and X_2 are. We attempt to construct the entire model just based on this pair using the structure of Fig. 3. The modeling error rate obtained from X_1 and X_2 is recorded, and this value is remembered. Next, a similar model is synthesised based on parameters X_1 and X_3. Assume, for example, that X_3 contains information and X_1 and X_2 do not. Consequently, the error rate for the models using X_3 and X_1 and X_2, respectively, will be lower than that for the model based on X_1 and X_2.

We cycle in turn through all possible pairs of variables, asking effectively, "If we had to model using only this pair of parameters, how well would we do?" This procedure is illustrated in Fig. 4. After considering all $N(N-1)/2$ pairs in the first layer of computation, we have identified that waveform parameter that contains the most information towards modeling, at least at the pair-wise level. The remainder of the parameters is discarded.

Now we go to a second layer and continue this model development, as shown in Fig. 4. We now take as a pair of inputs the outputs from the first layer. For example, the first element in the second layer

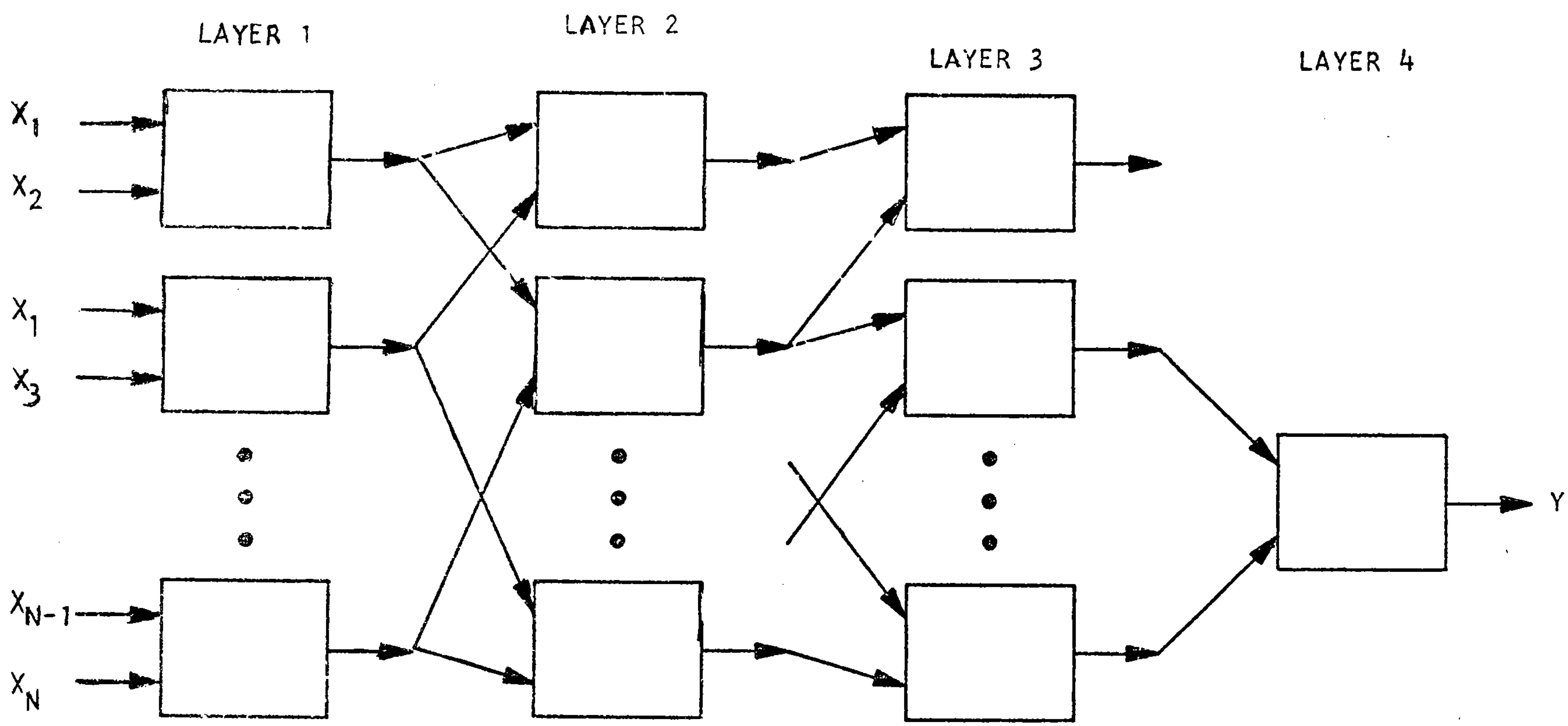


Fig. 4 Evolution of Hypercomp[TM] Network

evaluates the effectiveness of its two input variables. Notice that one is a function of X_1 and X_2 and the other is a function of X_1 and X_3. Consequently, when we multiply these two inputs, we are really treating implicitly up to four-way interactions of the original waveform parameters X_j. We have terms like $X_1^2 \cdot X_2, X_3$, and so on.

In this second layer of computations, even though we are asking the same question as in the first layer--"How well does this pair of variables perform?"--we are now considering variables which are really each functions of two of the x-variables. Therefore, we are examing four-way interactions in this second layer and asking the question, "How efficacious is this particular four-way interaction?"

What we do is begin to refine the model even further. We start looking at nonlinearities of nonlinearities, retaining those which are important and discarding those which are not. The error rate decreases further as the model becomes richer. It can be seen that this synthesis procedure is, in effect, self-programming from the data. No *a priori* considerations are required to generate this nonlinear transformation except perhaps in the selection of the original inputs.

Now, at the third layer, we consider again only a pair of variables. Notice that each one of these variables, which are outputs from the second layer, is a function of up to four X_i. Consequently, the third layer treats up to eight-way interactions of the X_i. Thus, the complexity of the model is increasing geometrically layer by layer, yet our work load is being maintained constant. We are always solving a small search problem for the element coefficients for a pair of variables. In this particular iterative fashion we adaptively construct a model from a data base which satisfies the criterion of interest directly, namely, how accurately can we discriminate flaws.

As I mentioned earlier, if we continue this process, we would eventually drive the error rate to zero because we would have enough degrees of freedom in this nonlinear model to finally overwhelm the problem. This is illustrated in Fig. 5. The solid line is the error versus the number of layers in this network. It asymtotically approaches zero. However, if during the model synthesis, before proceeding to the next layer in the network, we monitor the error rate on the testing set, we will obtain the dashed curve. It tends to track the solid curve, which is the design set error rate, until a certain point layer is reached at which the test set error rate begins to diverge. So, if one were using only one set of data to design a model, overfitting would be nearly impossible to avoid. For example, if a two per cent error rate was achieved by the model based on the design set, it is not uncommon to find that the error rate jumps to 45 per cent on a new set of data. This is, of course, due to overfitting the data base.

One avoids overfitting by using this training-testing program before going to the next layer. The error rate is checked on an independent set of testing data in order to prevent overfitting. As new data is developed in the future, the model can then perform as well on it as on the data for which the system was designed.

In summary, there are six major advantages of the adaptive nonlinear signal processing approach compared with conventional methods.

1) It eliminates guesses regarding which waveform parameters are informative. If you recall, a priori information is not needed. To the extent that it is available, it can be included in what we called the first layer, namely, you may have intutive knowledge that certain high frequency terms are relatively important, and these specific parameters can be incorporated as part of the inputs, X_i. In general, we can be rather exhaustive in the list of candidate parameters in the first layer and let the algorithm filter out those parameters which are relatively uninformative.

2) There is no limit in principle to the number of waveform parameters that can be considered. We have treated many hundreds in previous projects. In general, we find just a small fraction to be relevant.

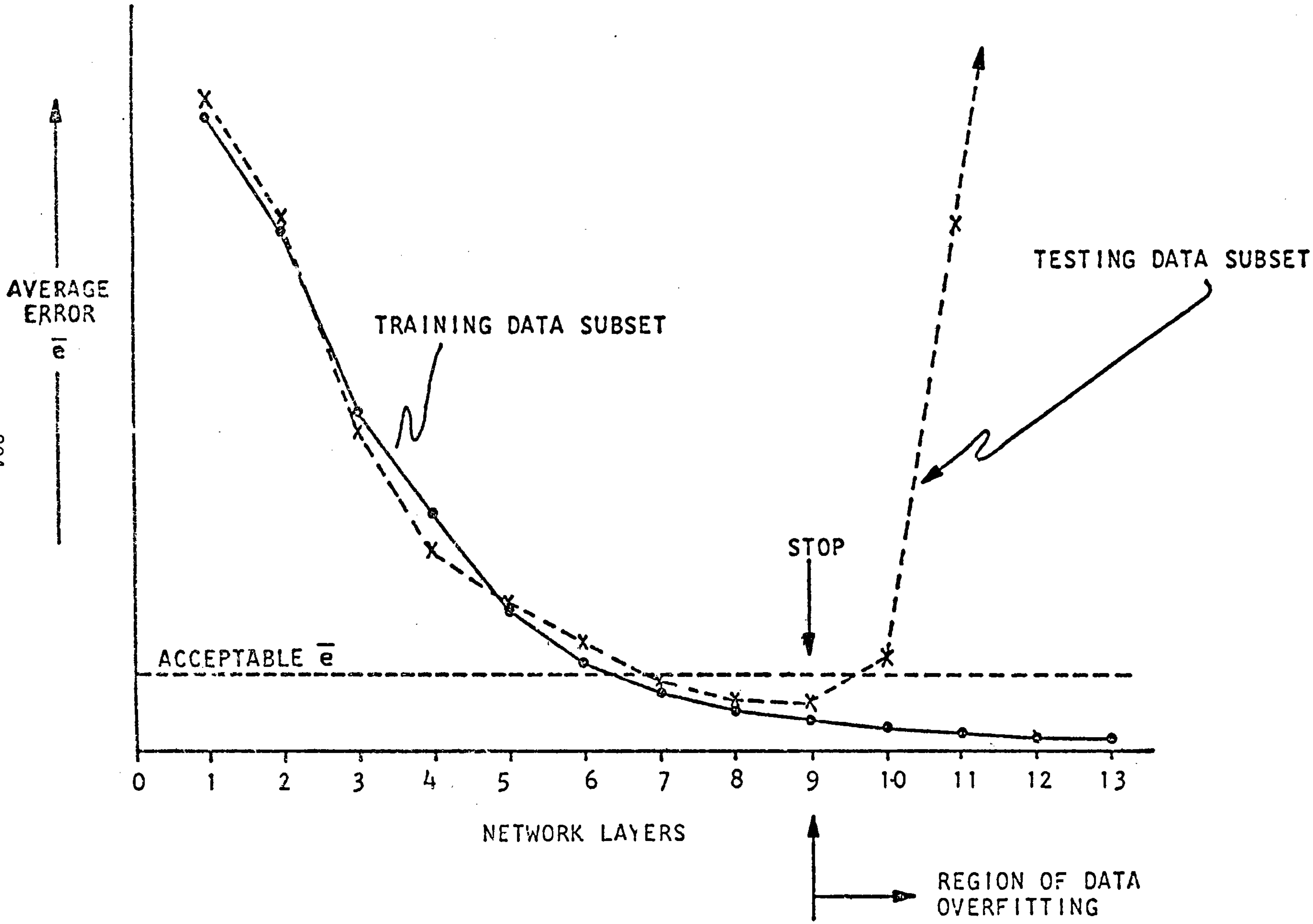


Fig. 5 Illustration of Detection and Avoidance of Overfitting in Synthesis of Hypercomp™ Trainable Networks

3) The nonlinear and higher-order interactions between parameters are automatically handled and discovered by this model synthesis procedure. That is, from the data base itself, we can discover those higher-order interactions that, in fact, are most discriminative.

4) Overfitting a sparse data base is avoided. A very small data set does not present a prohibitive problem. We can work with small data bases using the model synthesis training-testing procedure described above.

5) Any performance criterion can be used to train the adaptive processor. That is, we may use error metrics other than the mean square agreement between the predicted variable and the true variable. For example, we could additionally incorporate the cost of obtaining a particular error. It may be that it is much more costly to err on those crack lengths below a certain value than on those above this value. Therefore, the criterion which is used to guide this training synthesis can reflect the asymmetrical cost function.

6) The adaptive processor can operate in real-time and is simple in its mechanization.

This technique, by the way, is not only relevant for ultrasonic work, but also for acoustic emission and other types of NDE waveform processing.

I would like to finish this talk by giving you an idea of what these nonlinear networks look like for a problem in language classification that we recently worked on. This project is similar to classifying the diameter of flatbottom holes into one of 9 categories. The language problem consists of examining parameters recorded after 30 seconds of speech and determining which of five foreign languages the speaker was uttering. The classifier must be insensitive to both speaker and text.

The network that discriminated between languages 1 and 2 (Fig. 6) used only four of the 39 parameters that were discovered in the synthesis phase to be the relevant parameters for this parwise discrimination. The network that was automatically synthesized to discriminate between languages 1 and 3 again used four parameters, but a different group of four from the first network (Fig. 7). Once again, a fairly simple structure was found for discrimination between languages 1 and 4 (Fig. 8). On the other hand, notice the much more comples structure required to discriminate between languages 1 and 5 (Fig. 9).

Another interesting item in Fig. 9 is that parameter X_{12} is used over and over again. It turns out that this parameter is not a very good discriminator by itself; in fact, it is a very weak parameter. However, in combination with other more singly informative parameters it seems to provide enough information so that terms using X_{12} are quite discriminatory.

In summary, an empirical modeling methodology is available which has been applied in many problem areas over the past seven years. It is now being newly introduced to the NDE community via this Air Force-sponsored project. This methodology that can automatically identify a relevant subset of a large number of variables and it can construct a model which considers linear as well as nonlinear high-order interactions. Its utility in ultrasonic material flaw geometry will be revealed throughout the course of this current project.

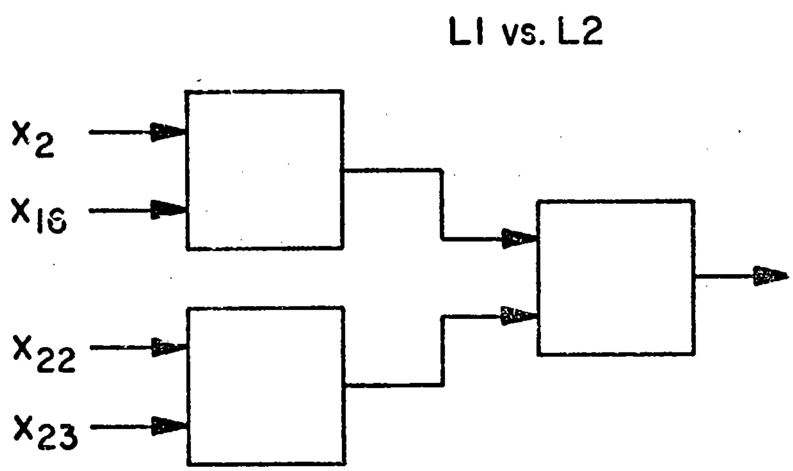


Fig. 6 Hypercomp^{TM} Classifier that Discriminates Languages L1 and L2

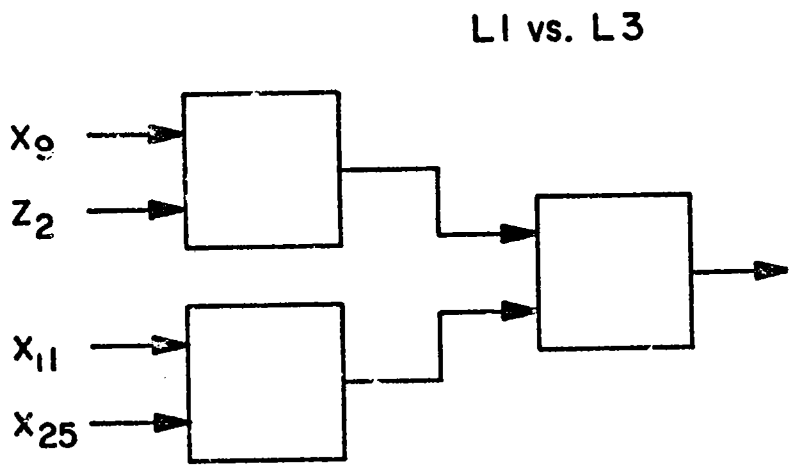


Fig. 7 Hypercomp™ Classifier that Discriminates Between Languages L1 and L3

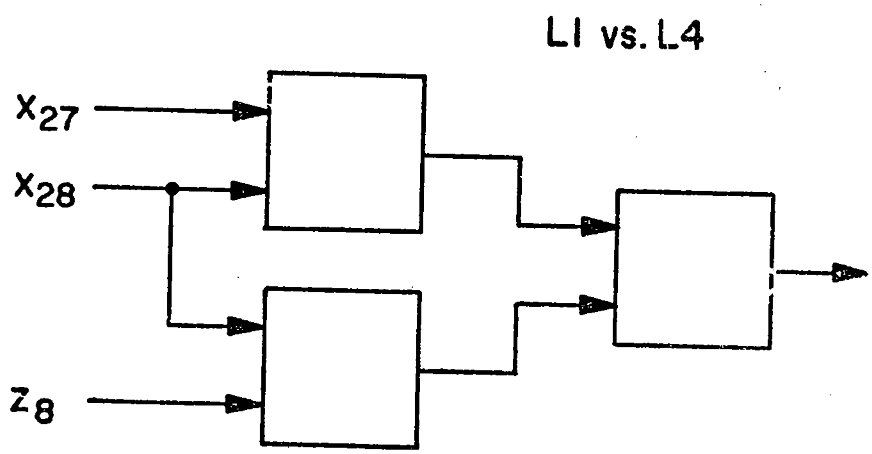


Fig. 8 Hypercomp™ Classifier that Discriminates Between Languages L1 and L4

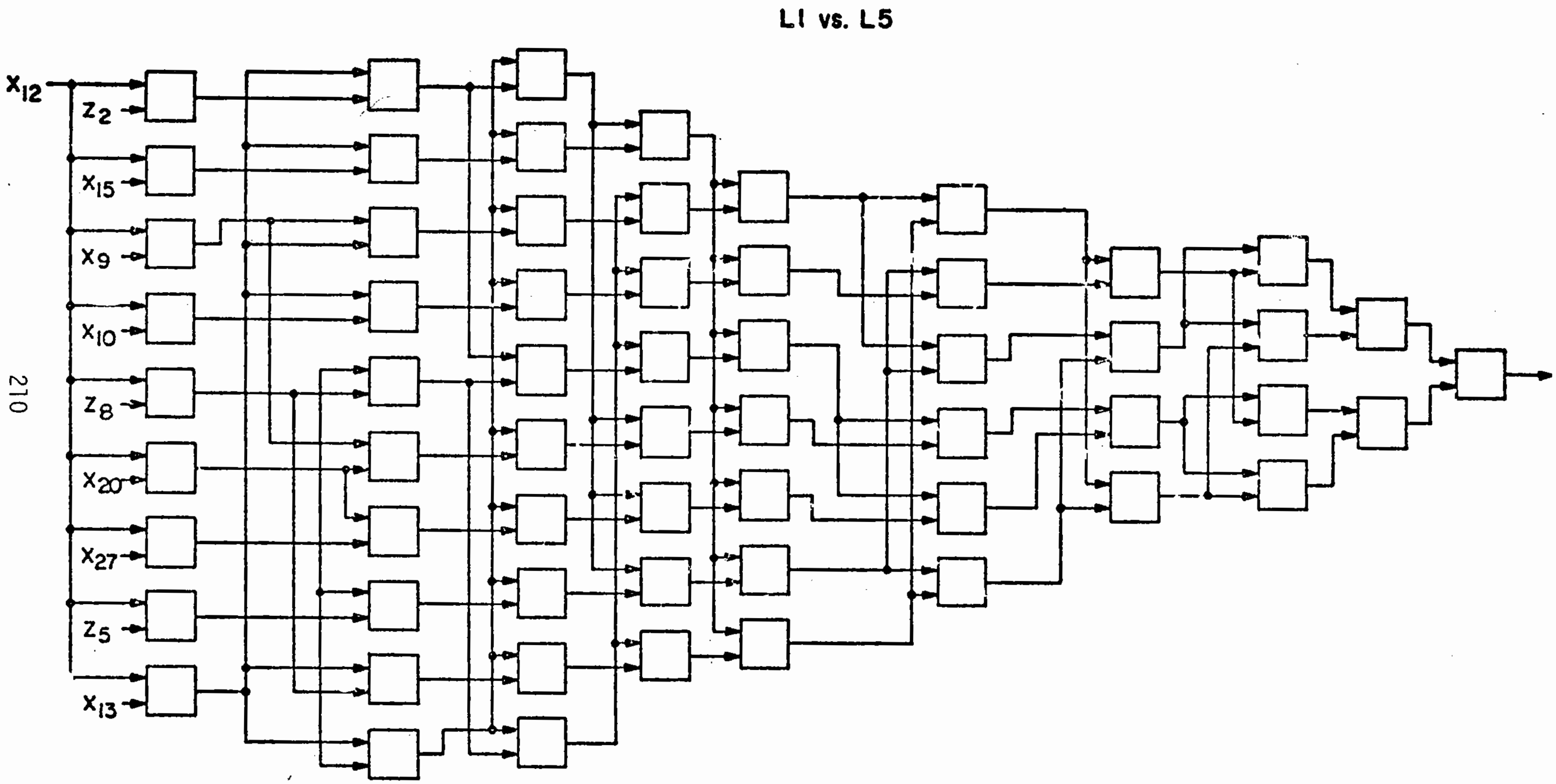


Fig. 9 Hypercomp[TM] Classifier that Discriminates Between Languages L1 and L5

DISCUSSION

DR. BRUCE THOMPSON (Science Center, Rockwell International): I didn't quite understand the network building block element. Are you considering just two parameters per element?

DR. MUCCIARDI: Yes.

DR. BRUCE THOMPSON: From this you construct linear terms, cross-product terms, and so forth?

DR. MUCCIARDI: Yes.

DR. BRUCE THOMPSON: Then you vary the coefficients of these terms to obtain the best correlation with your data set or whatever criteria you are looking for? How do you get those coefficients in that sum?

DR. MUCCIARDI: I didn't describe how the coefficients are found; there just isn't time. I would be happy to discuss it later.

Briefly, the coefficients are found by an optimization procedure. The values that give the best agreement between the predicted value and the true value are found. The criterion can be mean square error or any other that is desired.

DR. BRUCE THOMPSON: That is the essence of the synthesis procedure, i.e., this feed-forward training?

DR. MUCCIARDI: That's right; and then as one iterates this over and over again you are, in effect, searching higher and higher dimensional spaces.

PROF. TIERSTEN (Renseller Polytechnic Institute): This is probably an unfair question to ask, because it is directed at about three talks. I saw three times a fastener going through one plate bolted at the top and bolted at the bottom. What does it do?

DR. MUCCIARDI: The project called for the crack specimens to be prepared in the following way. A quarter-inch hole is drilled; a crack is artifically grown using a crack starter; and, then a nut is attached to the quarter inch bolt and torqued to a desired level. This procedure simulates cracks induced by fasteners.