

Defect classification for steam generator tubes of a nuclear power plant using ultrasonic
nondestructive evaluation

by

Aamir Masood Khan

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE

Major: Electrical Engineering
Major Professor: Satish S Udpa

Iowa State University
Ames, Iowa
2001

Copyright © Aamir Masood Khan, 2001. All rights reserved.

Graduate College
Iowa State University

This is to certify that the Master's degree of

Aamir Masood Khan

has met the thesis requirements of Iowa State University

Signatures have been redacted for privacy

This thesis is dedicated to my sweet family:

Abboo, Ammi, Aeysha Baji, Raana, Guriya, Saad

And

Abba Jee

TABLE OF CONTENTS

LIST OF FIGURES	vi
LI ST OF TABLES	vii
ACKNOWLEDGEMENTS	viii
CHAPTER 1. INTRODUCTION	1
1.1 Non-Destructive Evaluation (NDE) – An Overview	1
1.2 Ultrasonic NDE	2
<i>1.2.1 The Basic Ultrasonic NDE System</i>	2
<i>1.2.2 Data Presentation Modes in Ultrasonic NDE</i>	4
1.3 Pattern Recognition in Ultrasonic NDE	5
1.4 Organization of the thesis	7
CHAPTER 2. THE PROBLEM DEFIFNITION	8
2.1 Ultrasonic NDE in Steam Generator Tubes	8
2.2 Research Objectives	14
CHAPTER 3. FEATURE EXTRACTION AND SELECTION	15
3.1 Overview	15
3.2 Fast Fourier Transform (FFT) Coefficients as Features	15
<i>3.2.1 Introduction to the Fourier Transform (FT)</i>	15
<i>3.2.2 Computation of the DFT – The Fast Fourier Transform Algorithms</i>	17
3.3 Procedure and Results	18
<i>3.3.1 Preprocessing</i>	18
<i>3.3.2 Feature Extraction</i>	18
3.4 Discrete Wavelet Transform (DWT) Coefficients as Features	20

3.4.1 <i>Introduction to Wavelet Theory</i>	20
3.4.2 <i>Mathematical Background</i>	21
3.4.3 <i>Time-Frequency Representation of a Signal</i>	23
3.4.4 <i>The Continuous Wavelet Transform</i>	26
3.4.5 <i>Discretization of the Continuous Wavelet Transform: The Wavelet Series</i>	27
3.4.6 <i>The Discrete Wavelet Transform</i>	29
3.4.7 <i>The Subband Coding and Multiresolution Analysis</i>	29
3.5 Procedure and Results	34
3.5.1 <i>Preprocessing</i>	34
3.5.2 <i>Feature Extraction</i>	34
CHAPTER 4. CLASSIFICATION	37
4.1 An Overview of Pattern Recognition	37
4.2 Introduction to Neural networks	37
4.3 Applications of Neural Networks	38
4.4 The MLP Back propagation Neural Network	40
4.4.1 <i>The Back Propagation Algorithm</i>	41
4.5 Procedures and Implementation	43
CHAPTER 5. RESULTS AND DISCUSSION	45
5.1 The Results Obtained – An Analytical Discussion	45
5.2 Conclusions and Future Work	47
REFERENCES	50

LIST OF FIGURES

Figure 1.1 Basic Elements of an Ultrasonic NDE system	3
Figure 1.2 Various stages in a Pattern recognition system	6
Figure 2.1 The inspection probe shown inside the steam generator tube	9
Figure 2.2. The illustration of signal collection	11
Figure 2.3 A-Scan from a crack file	12
Figure 2.4 B-Scan from a crack file	13
Figure 2.5 A-Scan from a deposit file	13
Figure 2.6 B-Scan from a deposit file	14
Figure 3.1 A 482-point A-Scan from a crack file	19
Figure 3.2 A 482-point A-Scan from a deposit file	19
Figure 3.3 FFT Coefficients from of a deposit and crack	20
Figure 3.4 The Subband Coding Algorithm	31
Figure 3.5 Pre-processed crack signal	35
Figure 3.6 Pre-processed deposit signal	35
Figure 3.7 DWT coefficients of a crack signal	36
Figure 3.8 DWT coefficients of a deposit signal	36
Figure 4.1: Basic neural network model	39
Figure 4.2 The MLP feed forward neural network	40

LIST OF TABLES

Table 5.1 Summary of classification results for B-scans	46
Table 5.2 Summary of classification results for B-scans	48

ACKNOWLEDGEMENTS

I would like to start by saying that All prays is for Allah, the Most Merciful and the Most Beneficent, who made this task accomplishable for me.

I would like to express my utmost gratitude to Dr Satish S Udpa for his continuous help (which was always full of patience despite my ignorant questions), unmatched professional guidance (that always reflected his thorough, in-depth knowledge and perfect professional command), and administrative and moral support for the entire duration of this research project. I would also like to thank Dr Lalita Udpa for her valuable suggestions that let me keep in the right direction. Also, I would like to express my gratitude to Dr Thomas J Rudolphi for his interest in my thesis research.

I feel deeply obliged to National University of Sciences and Technology (NUST), Pakistan. The opportunity of undergoing a split MS program at ISU made a very positive contribution towards enhancing my professional capabilities.

I would like to express my deepest love to my sweet family – my parents, wife, sisters, my son, and my grand father - for their continuous prayers and consistent support during the entire duration of this research (as always). Special thanks my little son Saad-Bin-Aamir, who acted as a *re-energizer* by his sweet actions whenever I went back home really tired. Finally, thanks to all my colleagues at Material Assessment Research Group and my friends here at ISU for their help, advice and friendship. They all really made a healthy and comfortable environment for me while carrying out this challenging task.

CHAPTER 1. INTRODUCTION

1.1 Non-Destructive Evaluation (NDE) – An Overview

The oldest approach to test the integrity of a structure is the proof test, whereby a load is directly applied to the structure in question. However, in today's technologically advanced world, operational parameters such as what load to apply, for how long, and whether damage may have been induced or extended in the process are quantified by material science and the subject of applied mechanics. Unfortunately, the proof test is often an unsatisfactory approach in assuring the integrity of many modern structures because of time and cost constraints or difficulties in defining or realizing the necessary loading conditions. In addition, the consequences of failure of a massive component, such as a large pressure vessel, are too great to accept even in the test environment.

Consequently, the proof test has been augmented by other, less direct, approaches whereby the presence and severity of flaws in the structure are estimated through their interaction with some form of interrogating energy and the suitability of the part for continued service is estimated. This process is known as Nondestructive Evaluation (NDE) or Nondestructive Testing (NDT) [19]. NDE is an interdisciplinary activity, which includes various stages beginning with the definition of measurement procedures to detect flaws and other failure-related conditions, the design and development of instrumentation to perform the measurements, followed by the interpretation of the signals obtained in order to determine whether the structural part should be accepted or rejected. There are numerous advantages of carrying out successful NDE. Generally, these all stem from the more efficient use of materials and parts and depending on the specific engineering tradeoffs made, can

include improved safety, reliability and performance, utilization of a greater fraction of potential part lifetimes, avoidance of added manufacturing costs to defective materials or parts, and reduced weight and fuel costs. These advantages can be derived by applying NDE at various stages in the life history of a component, ranging from the inspection of raw materials to the rectification of a used part for further service.

1.2 Ultrasonic NDE

Nearly all forms of known energy have been used in NDE, including electromagnetic (ranging in frequency from dc resistivity measurements through quasi static and RF impedance measurements to visual and X-ray inspections), ultrasonic, neutron, and thermal energies. The choice of the appropriate form of energy to be used for NDE depends on the specific task under consideration. However, ultrasonic energy is one of the most widely used [19]. Several reasons for this include the fact that it can be inexpensively generated and detected, it can propagate deeply into the interior of many structures without excessive attenuation and the return signals have sufficient information-carrying capacity to determine important failure-related characteristics of the flaws or material.

1.2.1 The Basic Ultrasonic NDE System

Sound generated above the human hearing range (typically 20 KHz) is called ultrasound. Ultrasonic NDE introduces high frequency sound waves into a test object to obtain information about the object without altering or damaging it in any way. Two basic quantities are measured; they are the time of flight (which is the amount of time for the sound to travel through the sample) and amplitude of the received signal. In addition, measurements of relative change in signal amplitude can be used in sizing flaws or measuring attenuation

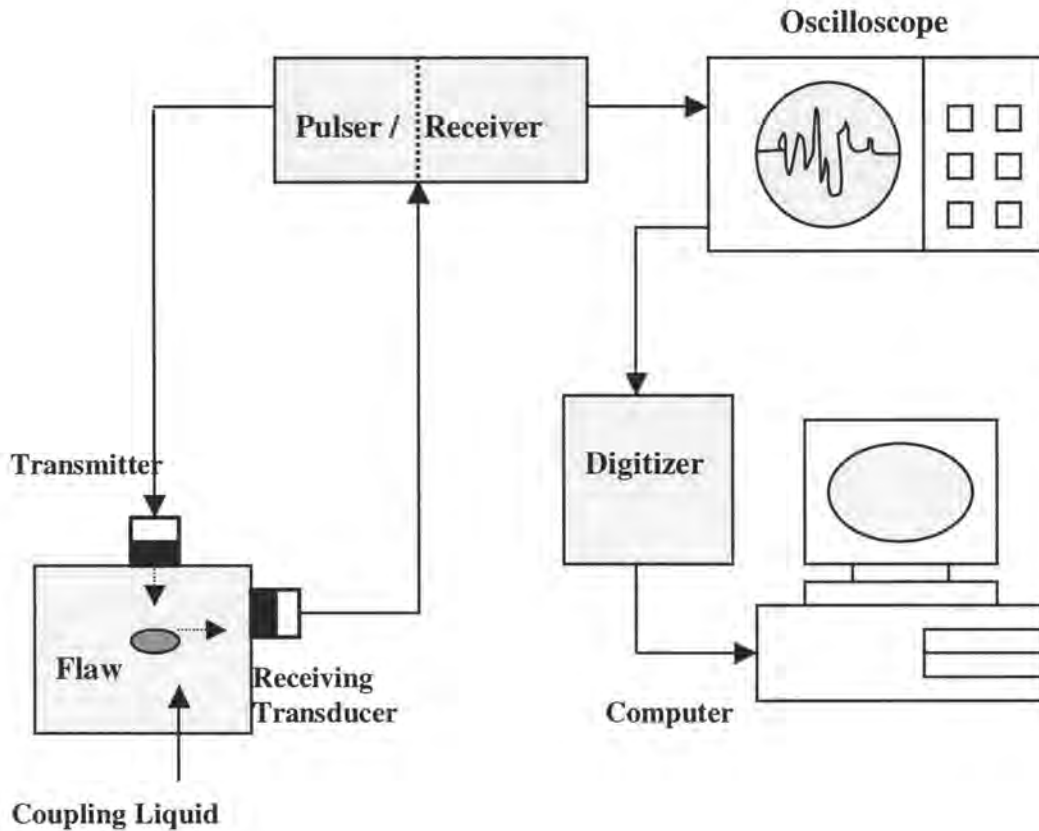


Figure 1.1. Basic Elements of an Ultrasonic NDE system

of a material. The basic elements in an ultrasonic NDE system are shown in figure 1.1:

Short bursts of elastic wave energy are injected into a part by a transducer. The elastic wave pulse varies in length from about $1\frac{1}{2}$ cycles for high-resolution flaw detection or thickness measurements to longer tone bursts for greater sensitivity or in precise attenuation or velocity measurements. The wave either consists of longitudinally polarized elastic wave energy or transversely polarized, shear wave energy. The longitudinal wave is a compressional wave in which the particle motion is in the same direction as the propagation of the wave. The shear wave is a wave motion in which the particle motion is perpendicular to the direction of propagation. Most commercially available instrumentation is tuned to 1 to

10 MHz frequency range. However, special purpose applications range in frequency from 100 kHz or less in highly attenuative materials to 100 MHz or more in low-loss materials where quite small flaws must be found. The waves are excited and detected by a piezoelectric transducer, coupled to the part through a liquid bond (since unlike light waves, ultrasound cannot travel in vacuum and requires an elastic medium such as a liquid or solid for its propagation). The exciting electrical signal is most commonly a quasi-impulse with peak voltage in the range of 100 to 1000 Volts. The received signals are amplified in a receiver and subsequently rectified and displayed in video form [17].

Ultrasonic inspection is an active measurement, in that energy is injected into the part by a transmitting transducer and the same, or a second receiving transducer then detects the energy reflected/scattered by the flaw. It is worth mentioning at this point that besides all its advantages, one of the most difficult aspects of ultrasonic testing is the interpretation of signals from unknown reflectors, e.g., the determination of the size, shape, and the orientation of these reflectors and / or the decision whether the part is acceptable for service. This is an area of active research. However, unknown reflectors are often classified in terms of the strength of their signal with respect to that of a standard reflector.

1.2.2 Data Presentation Modes in Ultrasonic NDE

The ultrasonic data can be presented for further processing / classification in several formats [23]. Most common formats are the A-Scan, B-Scan and C-Scan. The A-Scan is a display format in which the received pulse amplitude is represented as a displacement along one axis (usually the y-axis) and the travel time of the ultrasonic pulse is represented as a displacement along the other axis (usually the x-axis). The B-Scan is a two dimensional graphical presentation, in rectangular coordinates, in which the travel time of an ultrasonic

pulse is represented as a displacement along one axis, and transducer movement is represented as a displacement along the other axis. In other words, the B-Scan is the combined representation of all the A-Scans placed together. Finally, the C-Scan is a two dimensional graphical presentation, in which the discontinuity echoes are displayed in a top view on the test surface.

There are other formats of data presentation used in the industry like the D-Scan, in which the time-of-flight values are displayed in a top view on the test surface. This is a modified C-scan in which amplitudes are displayed. Also F-Scan and P-Scans can be used for data presentation, but are less commonly used in practice [23]. An F-Scan is also a modified C-scan in which values of certain feature (e.g. center frequency) are recorded and displayed instead of amplitudes. **Projection Imaging Scanning** is a projection of a B-scan result, which is performed in top view of the test piece, to any side view of the test piece. It is used for 3-D visualization of defects or corrosion. Ultrasonic weld inspection using P-scan imaging is mainly used for weld inspection with either a manual or a mechanized scanner. Corrosion mapping technique is applicable to pipes, vessels, storage tanks, ships and many other structures.

1.3 Pattern Recognition in Ultrasonic NDE

Pattern Recognition is the scientific discipline whose goal is the classification of objects (*patterns*) into a number of categories (*classes*) [18]. Depending on the application, these objects can be images or signal waveforms or any type of measurements that needs to be classified. Pattern recognition is an integral part in most *machine intelligent systems* built for decision making. A few examples of pattern recognition systems are *machine vision*, *character recognition*, *speech recognition*, *fingerprint identification*, *signature*

authentication, text retrieval and Computer-aided diagnosis. The medical community employs such techniques for interpreting medical data, such as, X-rays, computed tomographic images, ultrasound images, electrocardiograms (ECGs), and electroencephalograms (EEGs).

As in any one of the above-mentioned applications, a precise pattern recognition system serves as the integral part in ultrasonic NDE. Various stages involved in the design of such a classification system are shown in Fig 1.2.

Feature generation is the most critical stage in ultrasonic NDE. The purpose of feature generation is to transform a given the input sample space into a new “feature space” (which is, in most cases, a lower dimensional space), so as to achieve *information packing*. This means that most of the classification-related information is “squeezed” into a relatively small number of features, resulting in a much better form of input data [18]. The data in the feature space is better than the raw data in terms of its compactness, the ability to carry discriminatory information, and lower overall computational effort for the classification system.

In practice, a larger than necessary number of feature candidates is generated and then the “best” of them are selected in the *feature selection* stage. The most suitable set of features

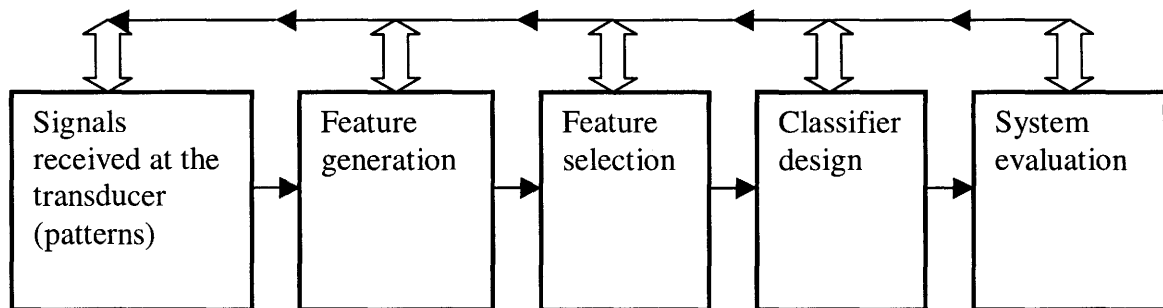


Figure 1.2 Various stages in a Pattern recognition system

are applied to the *classification stage*. There are numerous ways to design classifiers and very often this is a problem specific task. The classification can be *supervised* or *unsupervised* based on whether a set of training data is available or not. Examples of classification algorithms are Bayesian Classification, The Nearest-Neighbor Rule, K-Means Clustering, The Perceptron Algorithm, ID-3 Algorithm and classification using various types of Artificial Neural Networks (ANNs).

Finally, once the classifier has been designed, one needs to assess the performance of the designed classifier. This typically involves the estimation of the *classification error rate*. This task is usually performed at the *system evaluation stage*.

As is apparent from the feedback arrows shown in figure 1.2, the various stages are not independent. On the contrary, they are interrelated and, depending on the results, one may go back to redesign earlier stages in order to improve the overall system performance.

1.4 Organization of the thesis

The remaining four chapters will describe in detail all the stages involved in developing a defect classification system for steam generator tubes in a nuclear power plant using ultrasonic NDE. Chapter 2 explains the specific problem and the aim of this research. Chapter 3 describes how various features were generated and selected. Chapter 4 deals with the classifier design stage and includes a comprehensive coverage of MLP neural networks using back propagation learning algorithm. Finally, Chapter 5 describes the results obtained and includes a discussion of these results. Alternative methods are also proposed to tackle the problem from a different viewpoint that may result in better performance.

CHAPTER 2. PROBLEM DEFINITION

2.1 Ultrasonic NDE in Steam Generator Tubes

Steam generators are large components used in pressurized-water nuclear power plants. They convert water into steam from heat produced in the reactor core. Inside the steam generators, hot radioactive water is pumped through thousands of feet of tubing. Each steam generator can contain anywhere from 3,000 to 16,000 tubes, each under high pressure to prevent it from boiling. That water flowing through the inside of the tubes then heats non-radioactive water on the outside of the tubes. This produces steam that turns the blades of turbines to make electricity. The steam is subsequently condensed into water and returned to the steam generator to be heated once again. These tubes have an important safety role because they constitute one of the primary barriers between the radioactive and non-radioactive sides of the plant. For this reason, the integrity of the tubing is essential in minimizing the leakage of water between the two "sides" of the plant.

There is the potential that if a tube bursts while a plant is operating, radioactivity from the primary coolant system, the system that pumps water through the reactor core, could escape directly to the atmosphere in the form of steam. Therefore, it is a high priority task to ensure that steam generator tube degradation is carefully monitored through routine inspections. These tubes must have an extremely low probability of abnormal leakage and must be periodically inspected and tested. Routine inspection through various nondestructive evaluation techniques is the most appropriate means of detecting defects in these tubes. Eddy current examination techniques, which are used more commonly to inspect steam generator

turbines, provide a good screening for tube integrity but lack the ability to accurately size certain flaw mechanisms. Due to this sizing uncertainty, conservative decisions must be made with regard to remedies resulting in premature plugging and reduced inspection cycle lengths. It has been argued in some cases that ultrasonic examination techniques could provide better sizing and characterization.

The ultrasonic inspection system setup for the steam generator tubes is illustrated in Figure 2.1. As shown in the figure, the ultrasonic inspection probe moves along the axis of the tube with water sealed between the tube wall and the probe. The region of interest is a region where cracks are formed on the tube wall. This region extends axially around 2 inches along the tube within and above the tube-sheet support. Cracks are most likely to occur in this region because the tubes expand against the “tube sheet” (the region where the tube is in contact with a surrounding structure). In addition, “deposits” containing chemical precipitates and dissolved metallic compounds are formed on the tube sheet in this region.

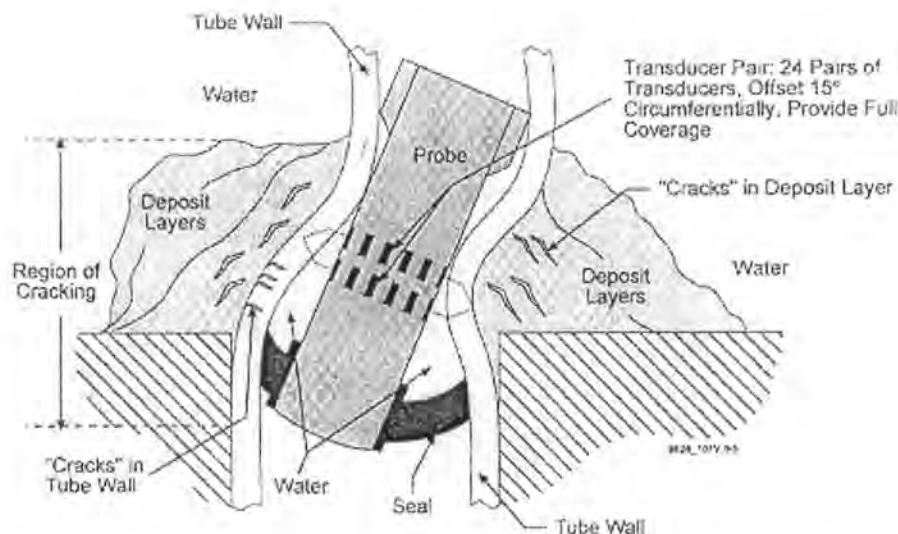


Figure 2.1 The inspection probe shown inside the steam generator tube.

The accumulation and expansion of the deposit material produces localized bowing of the tube immediately above the tube sheet [21].

Inspection probes are commonly employed for ultrasonic testing in steam generator tubes. The probes have ultrasonic transducers mounted along their circumferential surfaces that are used to transmit and /or collect time-of-flight information from the tubes. There are 24 pairs of transducers on the inspection probe used in this project. As shown in Figure 2.2, each pair of transducers is focused at the same point on the outer diameter (OD) of the tube. The transducers have nominal center frequencies of 8-11 MHz (-6dB from ~6 to ~16 MHz). The sampling frequency is 80 Mb/s, and the tube wall thickness is approximately 0.047 inches. The probe uses pairs of oppositely directed transducers to collect reflected shear wave data from two opposing directions in one pass through the tube.

The probe transmits 45 degrees (into the tube wall) refracted shear waves for the detection of OD initiated circumferential cracks. Detection is based on receiving signals reflected from the base of the crack. However, it has been observed in practical testing that signals are also reflected from the OD deposits. As the probe moves along the tube axis, data are collected from each transducer. At every axis position, the 24 pairs of transducers collect information from positions around the tube with an offset of 15° ($360^\circ/24$) between adjacent pairs of transducers.

For each transducer, a 482-point A-scan is collected. The A-scan from the opposite transducer contains information of the same transducer position (as shown in Figure 2.2). The signals from opposite transducers can be concatenated to form a composite A-scan of 964 points.

The data collected in B-scans format are from two types of “classes”, that is, cracks and deposits. The cracks need to be identified during the routine testing since these can result into tube leakage and have dangerous consequences. Deposits, on the other hand are benign and do not pose a serious danger for the proper functioning of the reactor. Therefore, these two “classes” need to be discriminated from each other through an automated testing/evaluation system. An automated system is preferred since a large number of signals need to be evaluated and a manual process can be exceedingly slow.

Figure 2.3 shows a 482-point A-scan that contains information about a crack. Combining all the A-scans along the tube axis results in a 2-Dimensional B-scan image for one pair of transducers. The B-scan is shown in Figure 2.4 where the abscissa is the time axis

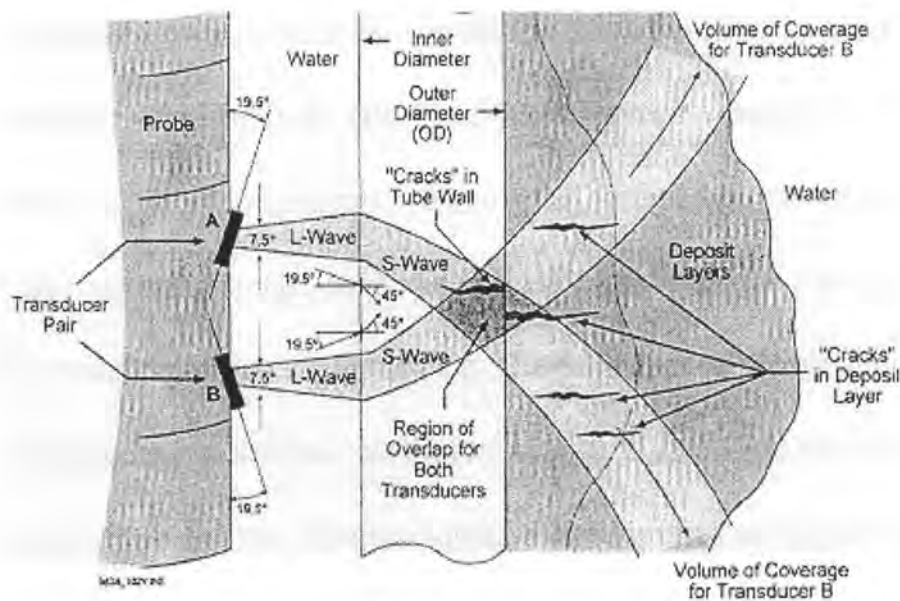


Figure 2.2. The illustration of signal collection

and the ordinate is the axial position along the tube. Different gray levels in the image represent different amplitudes, where “black to white” refers to “low to high” amplitudes respectively. In this image, two football-shaped images indicate crack signals from two opposite transducers. The first and third vertical strips with high amplitudes are reflections from the inner diameter (ID) and the second and fourth are the multi-reflections from the ID. Figure 2.5 and 2.6 represent the A-scan and B-scan displays obtained from a deposit. As obvious from the figures it is not always possible to visually differentiate crack signals from those caused by deposits.

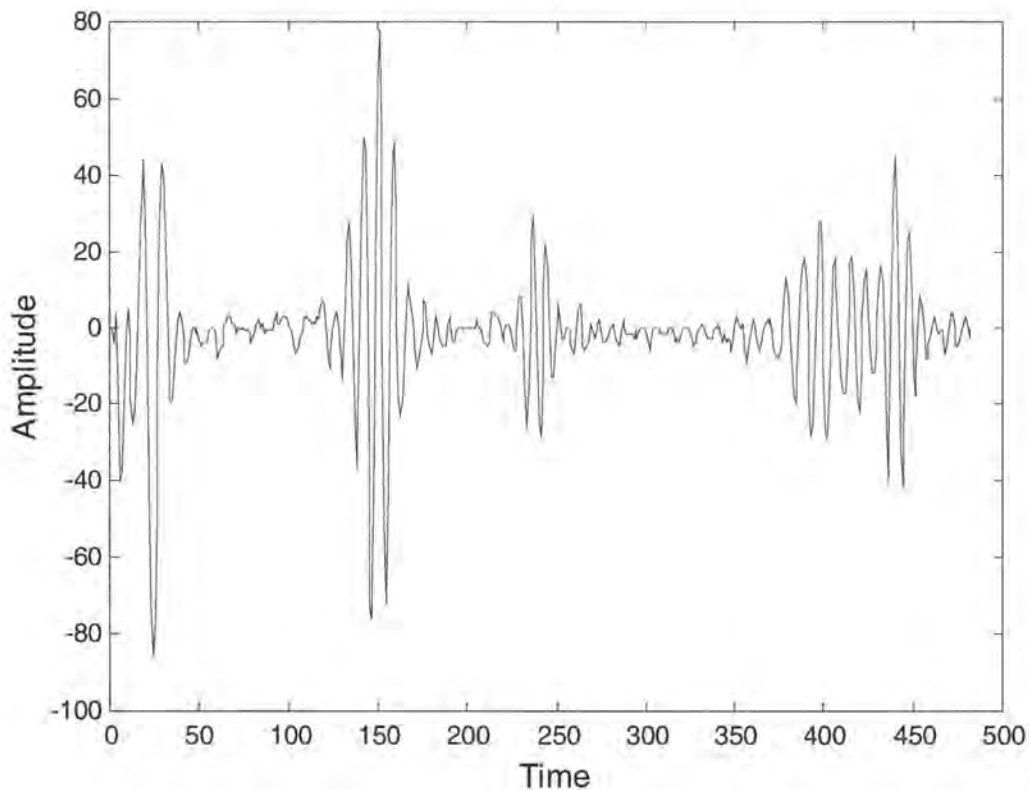


Figure 2.3 A-Scan from a Crack file

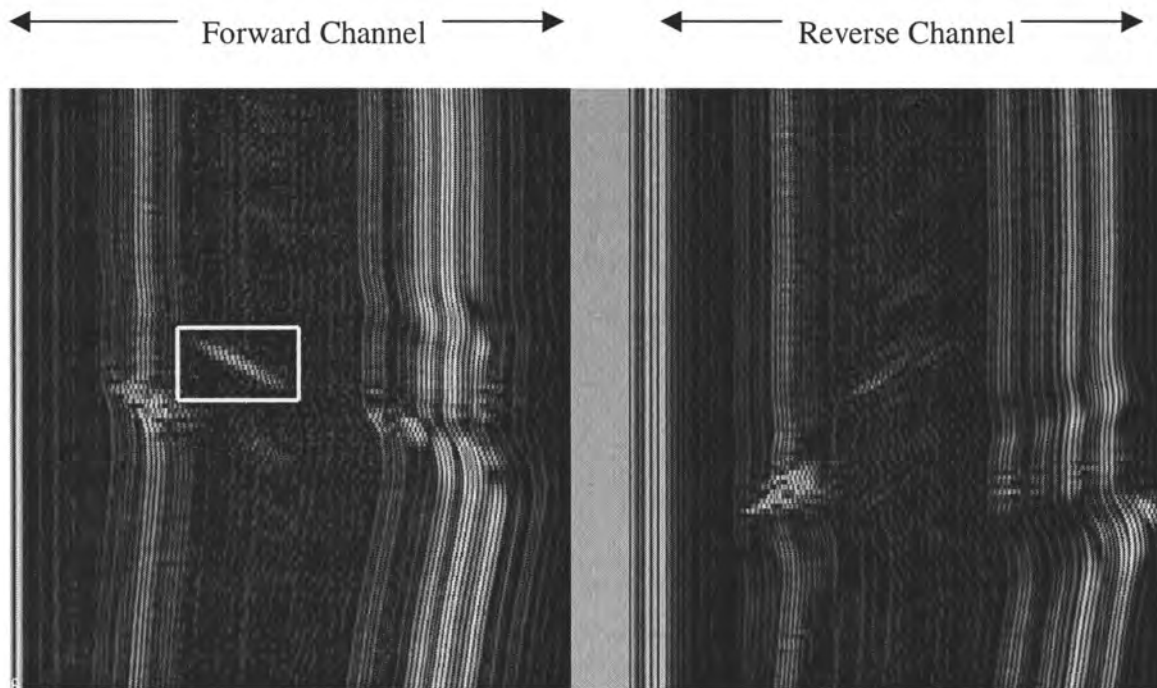


Figure 2.4 B-Scan from a Crack file

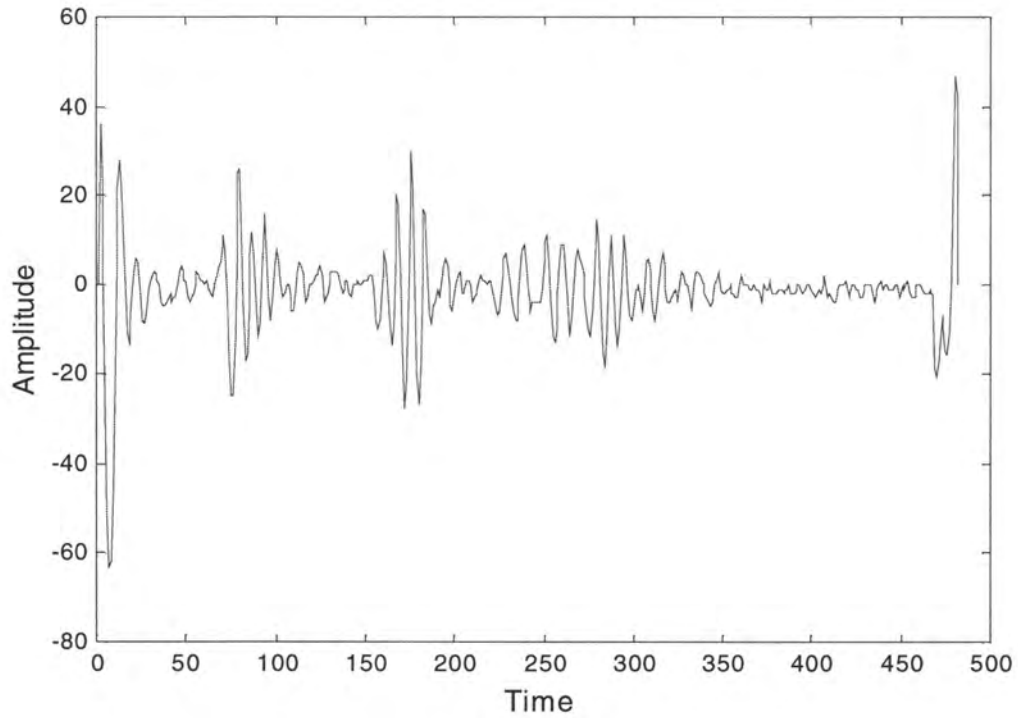


Figure 2.5 A-Scan from a Deposit file

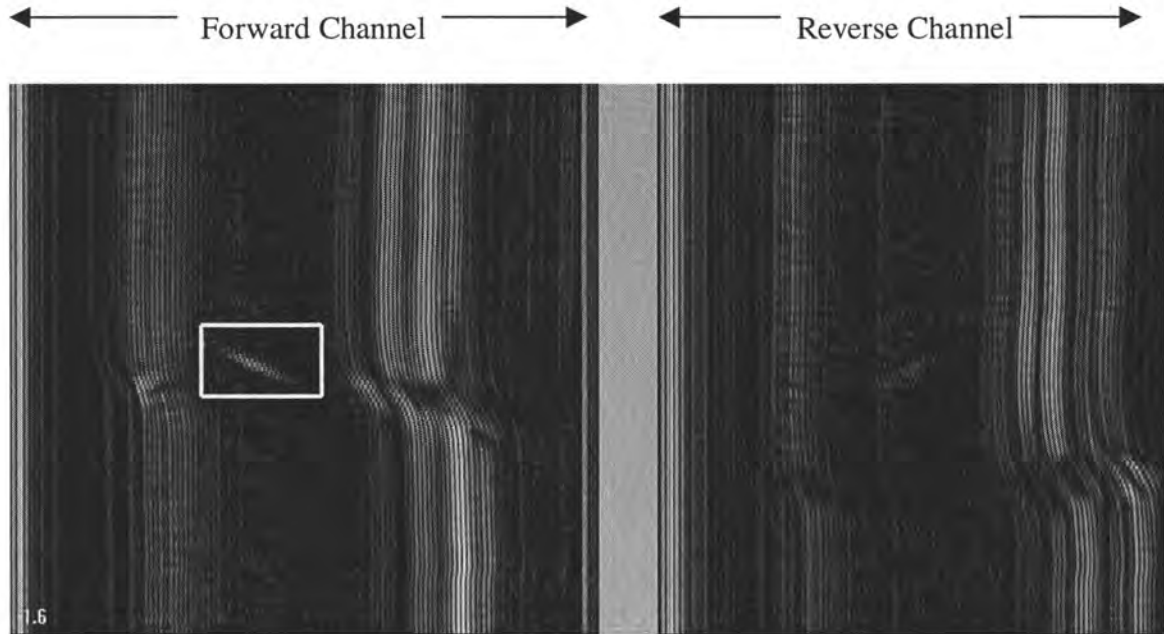


Figure 2.6 B-Scan from a Deposit file

2.2 Research Objectives

The objectives of this thesis research are:

- To carefully investigate the nature of the ultrasonic signals collected from “cracks” and “deposits”.
- To apply various signal processing techniques for extracting features that carry the relevant discriminatory information that allows differentiation between the two “classes”, namely the *cracks* and *deposits*.
- To apply pattern recognition method(s) for developing a classification system that can classify *cracks* and *deposits* with minimum misclassification error.

CHAPTER 3. FEATURE EXTRACTION AND SELECTION

3.1 Overview

As described in chapter 1, feature generation and feature selection are the most critical tasks in the ultrasonic NDE signal analysis system. The aim of feature generation is to squeeze “relevant” information into a relatively very small number of features as compared to the length of the input sample vector. The feature generation task is very often problem dependent. For ultrasonic NDE, transform-based features are one of the appropriate choices. The basic reason behind transform-based features is that an appropriately chosen transform can exploit and remove information redundancies, which usually exist in the set of signals obtained by the measuring devices, i.e., the transducers. Various examples of these transforms are the Karhunen-Loeve (K-L) Transform, Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), Discrete Sine Transform (DST), Hadamard Transform, Haar Transform and the Discrete Wavelet Transform (DWT). A comprehensive treatment on each of these transforms can be found in [18].

3.2 Fast Fourier Transform (FFT) Coefficients as Features

3.2.1 Introduction to the Fourier Transform (FT)

Like most of the *raw* signals in practice, the A-scan is a *time domain* signal representation. In other words, it is a *time-amplitude* representation of the signal. This representation is not always the best representation of the signal for most signal processing related applications. In many cases, it may be possible to obtain discriminatory information

by transforming the signal into an appropriate domain. The Fourier transform of a time-domain signal results in the frequency spectrum of the signal, which represents the spectral components present in the signal.

The Fourier transform is defined as follows [12].

$$X(f) = \int_{-\infty}^{\infty} x(t) \cdot \exp(-j2\pi ft) dt, \quad (3.1)$$

where $x(t)$ is a continuous time-domain signal and $X(f)$ is its Fourier Transform. The discrete version of the Fourier Transform, known as The Discrete Fourier Transform (DFT), is defined as follows.

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp(-j2\pi kn / N), \quad k=0,1,2, \dots, N-1 \quad (3.2)$$

where $x(n)$ is a discrete-time signal and $X(k)$ is the DFT of $x(n)$.

The Fourier transform basically expands the signal in terms of basis functions that are complex exponentials. If the result of the integral (3.1) is a large value, it means that the signal $x(t)$ has a dominant spectral component at frequency ' f '. The information provided by the integral in (3.1), corresponds to all time instances. Consequently, the Fourier transform is not suitable if the signal spectrum varies with time. Signals whose spectra vary with time have traditionally been analyzed using time-frequency distributions such as the Winger transform [5,6]. Alternately, they have been analyzed using the Short Time Fourier Transform (STFT) [3,8,15] where the spectrum of moving windowed segment of the signal is computed. Additional details of the STFT are provided in Section 3.4.3.

3.2.2 Computation of the DFT – The Fast Fourier Transform Algorithms

There are two approaches employed for computing the DFTs [13]. One is a *divide-and-conquer* approach in which a DFT of size N , where N is a composite number, is reduced to the computation of smaller DFTs from which the larger DFT is calculated. In particular, most important computational algorithms are the ones where the size N is a power of 2 and when it is a power of 4. These algorithms are known as radix-2 and radix-4 *Fast Fourier Transform (FFT) algorithms*, respectively.

It can be observed from (3.2) that for each value of k , direct computation of $X(k)$ involves “ N ” complex multiplications ($4N$ real multiplications) and “ $N-1$ ” complex additions (“ $4N-2$ ” real additions). Consequently, to compute all “ N ” values of the DFT requires N^2 complex multiplications and $N^2 - N$ complex additions. Direct computation of the DFT is basically inefficient primarily because it does not exploit the symmetry and periodicity properties of the phase factor $\exp(-j2\pi kn/N)$. The FFT algorithms exploit these two basic properties of the phase factor to reduce the number of complex multiplications and thus, result in much more computationally efficient method of determining the DFT as compared to the direct computation. There are two approaches used in these FFT algorithms, namely the *decimation-in-time* and *decimation-in-frequency* algorithms that are based on whether the decimation is being done in the time domain or in the frequency domain.

The second approach is based on formulation of the DFT as a linear filtering operation on the data. This approach leads to two algorithms, the *Goertzel* algorithm and the *Chirp-z Transform* algorithm for computing the DFT via linear filtering of the data sequence [13].

3.3 Procedure and Results

3.3.1 Preprocessing

As mentioned earlier, the flaw signals are available in the B-scan format. The B-scans for one of the transducer pairs were shown in Figure 2.5. In this figure, the football-shaped images represent a crack. The B-scans for a deposit file were shown in Figure 2.6

It is obvious from these figures that there is no visual difference between the two. It is anticipated that transforming the signal into an appropriate domain may result in features that are spaced apart in the feature space if they are derived from signals of different classes and are close if they belong to the same class. For the problem in hand, A-scans were analyzed to obtain features. Each A-scan is 482 points in length. The A-scans from a crack file and a deposit file are shown in Figure 3.1 and Figure 3.2 respectively. The region of interest in the A-scan is only the *flaw region*, that is, where the crack (or deposit) information is available. Thus, 101 points were chosen from the A-scan for further processing. After selecting these 101 points, the data was normalized before computing its FFT.

3.3.2 Feature Extraction

After normalization of the A-scan data, FFT was taken and the FFT coefficients were stored as features. The amplitudes of FFT coefficients from a crack and deposit signal are shown in Figure 3.3. It is clear from both these figures that the plot is symmetric about the middle point in each case. Furthermore, negligible discriminatory information is contained in points in the range from 30 to 51. (The amplitudes of FFT coefficients are almost zero in this range). Therefore, the first 30 points in the FFT were chosen as features to be tested by the classification algorithms.

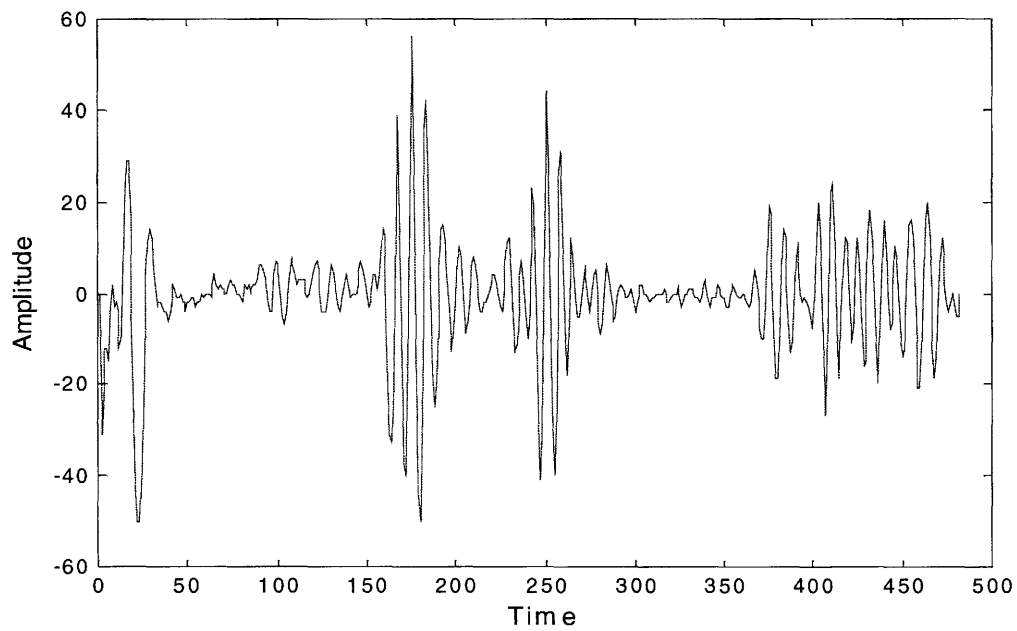


Figure 3.1 A 482-point A-Scan from a Crack file.

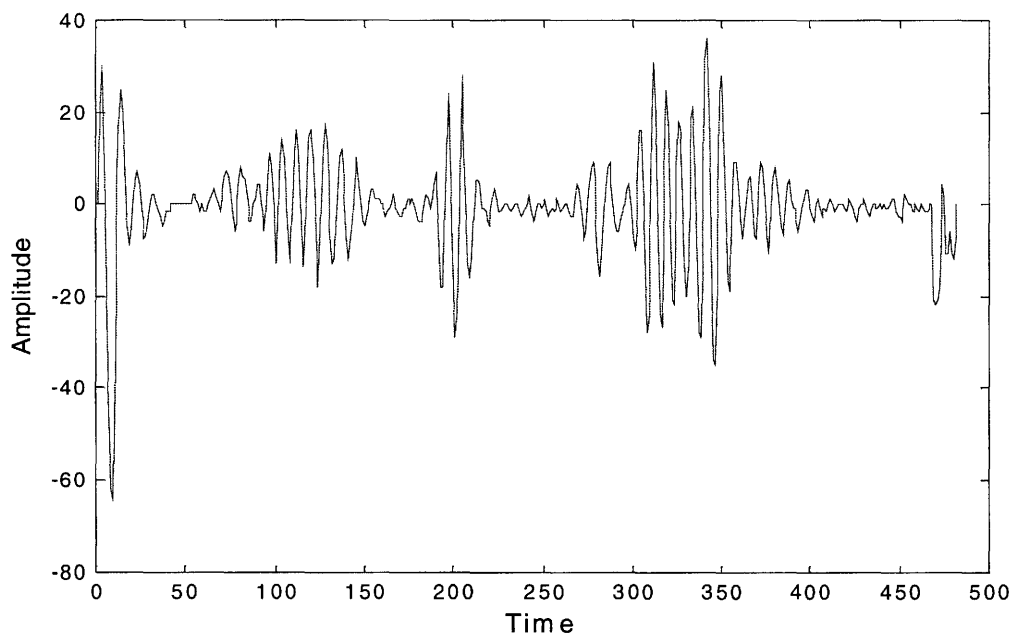


Figure 3.2 A 482-point A-Scan from a Deposit file.

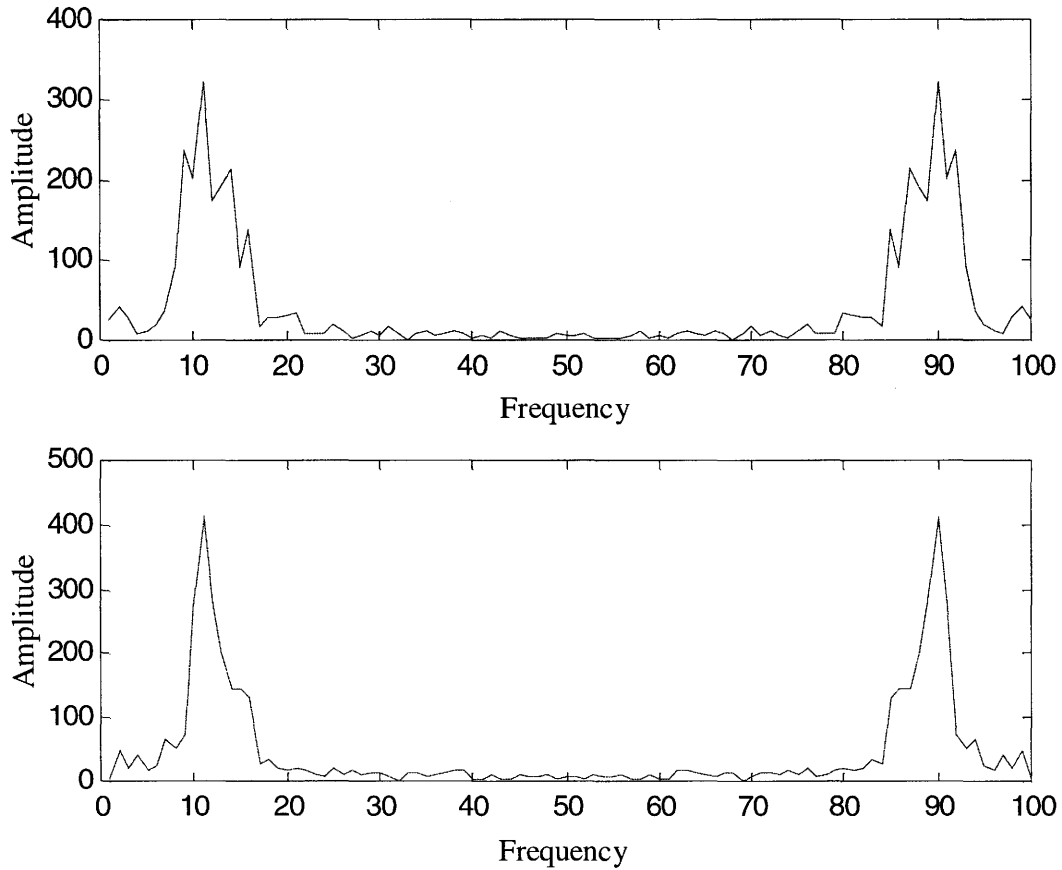


Figure 3.3 FFT Coefficients from of a deposit (up) and crack (below).

3.4 Discrete Wavelet Transform (DWT) Coefficients as Features

3.4.1 Introduction to Wavelet Theory

The Wavelet theory is used as an analytical tool for signal processing (including speech and image compression), numerical analysis, and mathematical modeling. The history of modern wavelet theory is not very long and the early work was carried out in the 1980's by Morlet , Grossmann , Meyer[3], Mallat [10,11], and Ingrid Daubechies [7,8].

The wavelet transform is of particular interest for the analysis of non-stationary signals, that is, signals whose frequency response varies with time. The Wavelet transform provides an alternative to the Short Time Fourier Transform (STFT) by using short windows at high frequencies and long windows at low frequencies, instead of using a single analysis window as in the case of the STFT. The functions to be analyzed can be solutions of a differential equation with shock [8], or integral kernels of singular integral operators, or 1 or 2-dimensional signals, as in sound (speech , music), ultrasound , time series or images.

The wavelet transform can be seen as signal decomposition onto a set of basis functions called *wavelets*. The *wavelets* are obtained from a single prototype *mother* wavelet by dilations and contractions (scalings) as well as shifts [7]. Wavelets may be viewed from two aspects, that is, “mathematical” and “algorithmic ”. The mathematical aspect is covered first followed by the algorithmic implementation that describes an efficient coding scheme for obtaining the Discrete Wavelet Transform.

3.4.2 Mathematical Background

The following terms need to be described before explanation and interpretation of the Wavelet transform.

1. Any function $f(t)$ can be written as a linear combination of the *basis functions* $\phi_k(t)$ and the corresponding coefficients μ_k as follows:

$$f(t) = \sum_k \mu_k \phi_k(t) \quad (3.3)$$

For example, the complex exponential (sines and cosines) functions are the basis functions for the Fourier transform.

2. The *inner product* of two functions is defined as follows:

$$\langle f(t), g(t) \rangle = \int_a^b f(t) \cdot g^*(t) dt, \quad (3.4)$$

where $f(t)$ and $g(t)$ are two functions in $L^2 [a,b]$ (that is, they are set of square integrable in the interval $[a,b]$).

3. Two functions f and g are said to be *orthogonal* to each other if their inner product is zero.

That is

$$\langle f(t), g(t) \rangle = \int_a^b f(t) \cdot g^*(t) dt = 0 \quad (3.5)$$

4. A set of functions $\{\phi_k(t)\}$, $k=1,2,3,\dots$, is said to be *orthonormal* if

$$\int_a^b \phi_k(t) \phi_l^*(t) dt = 0, k \neq l \quad (\text{orthogonality condition}) \quad (3.6)$$

and

$$\int_a^b \|\phi_k(t)\|^2 dt = 1, \text{ or equivalently} \quad (3.7)$$

$$\int_a^b \phi_k(t) \phi_l^*(t) dt = \delta_{kl}, \quad (3.8)$$

where δ_{kl} is the *Kronecker delta* function, defined as:

$$\delta_{k,l} = \begin{cases} 1, & \text{if } k = l \\ 0, & \text{if } k \neq l \end{cases} \quad (3.9)$$

The orthonormal basis functions are of particular importance because the orthonormal bases

allow computation of the analysis coefficients in a straightforward way using the property of orthonormality. For orthonormal bases, the coefficients μ_k can be calculated as

$$\mu_k = \langle f, \phi_k \rangle = \int f(t) \cdot \phi_k^*(t) dt, \quad (3.10)$$

and the function $f(t)$ can then be reconstructed by using (3.3) by substituting the μ_k coefficients. This yields

$$f(t) = \sum_k \mu_k \phi_k(t) = \sum_k \langle f, \phi_k \rangle \phi_k(t) \quad (3.11)$$

3.4.3 Time-Frequency Representation of a Signal

The frequency *spectrum* of a function $x(t)$ is given by the Fourier transform, which was defined in (3.1) as follows:

$$X(f) = \int_{-\infty}^{\infty} x(t) \cdot \exp(-j2\pi ft) dt \quad (3.12)$$

The Discrete Fourier Transform (DFT), was also defined in (3.2) as follows.

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp(-j2\pi kn / N), \quad k=0,1,2, \dots, N-1 \quad (3.13)$$

where $x(n)$ is a discrete-time signal and $X(k)$ is the DFT of $x(n)$.

As mentioned earlier, the Fourier transform basically expands the signal in terms of complex exponential basis functions. It describes what all frequency components are present in $x(t)$, but it's not possible to get the timing information of these components (this information is all hidden in the phase of $X(f)$). This is because the information provided by the integral in (3.12), corresponds to all time instances.

For the analysis of non-stationary signals (that is, the signals whose frequency response varies with time), it is desirable to have a frequency decomposition that is ‘local’ in time. This is what is achieved by so called *time-frequency representations*. The Short Time Fourier Transform (STFT), the Continuous Wavelet Transform (CWT), and the Discrete Wavelet Transform (DWT) come under this category of time-frequency representations.

In case of STFT, the function x is first ‘windowed’ by multiplying it with a fixed $g(t)$, that is the window function. This effectively restricts x to an interval (with smooth edges). Then the Fourier coefficients of this product are computed. This process is repeated with shifted versions of g , i.e. $g(t-nt_o)$, leading to a family of windowed Fourier coefficients[8],

$$S_{m,n}(x) = \int f(s)g(s-nt_o)e^{jmw_o s} ds, \quad (3.14)$$

where m,n are integers. These can also be viewed as the inner products (in $L^2(\mathbf{R})$) of x with the

$$g_{m,n}(t) = e^{-jmw_o s} g(t-nt_o) \quad (3.15)$$

Each $g_{m,n}$ consists of an envelope function, shifted by nt_o . The index n gives the time localization of $g_{m,n}$, the index m being its frequency.

The Wavelet transform is similar to the STFT in that it also computes the inner products of x with a sequence of functions $\psi_{m,n}$, with m indicating frequency localization, and n time localization as shown by the following equation [6]:

$$W_{m,n}(x) = \int x(s)\overline{\psi_{m,n}(s)}ds \quad (3.16)$$

But, in this case the $\psi_{m,n}$ are generated in a different way. The basic wavelet ψ (commonly known as the “mother wavelet”), is typically well concentrated in time and in frequency,

and has integral zero, that is

$$\int \psi(t) dt = 0, \quad (3.17)$$

which means that it has at least some oscillations. The $\psi_{m,n}$ are then generated by dilations and translations as follows:

$$\psi_{m,n}(t) = a_o^{-m/2} \psi(a_o^{-m} t - n b_o), \quad (3.18)$$

where $a_o > 1$ and $b_o > 0$ are fixed parameters (similar to ω_o, t_o in (3.14)). Changing m in (3.18) amounts to packing the oscillations of ψ into a smaller or larger width, i.e. to wavelets with higher or lower frequency ranges. For fixed m , the $\psi_{m,n}$ are the translates of $\psi_{m,o}$ by $n a_o^m b_o$, i.e. wavelets are translated by amounts proportional to their width. There are various types of mother wavelets - the high frequency wavelets are narrow, whereas low frequency wavelets are wide. This is the main difference between STFT and Wavelet transform: the $g_{m,n}$ of the STFT all have the same width. *The wavelet transform is particularly well adapted to functions, signals or operators with highly concentrated high frequency components superposed on longer lived low frequency components.*

There are several types of wavelet transforms [15], and, depending on the application, one may be preferred to the others. For a continuous input signal, the time and scale parameters can be continuous, leading to the Continuous Wavelet Transform (CWT). They may as well be discrete leading to a Wavelet Series expansion. Finally, the wavelet transform can be defined for discrete-time signals, leading to a Discrete Wavelet Transform (DWT).

3.4.4 The Continuous Wavelet Transform (CWT)

The continuous wavelet transform is defined as follows:

$$CWT \ \Psi_x(\tau, s) = \Psi_x(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t) \psi^* \left(\frac{t - \tau}{s} \right) dt \quad (3.19)$$

According to the definition of the inner product described earlier, the CWT can be thought of as the inner product of the test signal with the basis functions $\psi_{\tau,s}(t)$:

$$CWT \ \Psi_x(\tau, s) = \Psi_x(\tau, s) = \int x(t) \psi_{\tau,s}^*(t) dt \ , \quad (3.20)$$

where

$$\psi_{\tau,s} = \frac{1}{\sqrt{|s|}} \psi \left(\frac{t - \tau}{s} \right) \quad (3.21)$$

The transformed signal is a function of two variables, τ and s , the *translation* and *scale* parameters, respectively. ψ is the transforming function (the mother wavelet) . The term *translation* is related to the location of the window, as the window is shifted through the signal. The scale s is defined as the reciprocal of frequency and as a mathematical operation, it either dilates or compresses the signal. Low frequencies (high scales) correspond to a global information of a signal, whereas high frequencies (low scales) correspond to a detailed information of a hidden pattern in the signal. Also, larger scales correspond to dilated signals and small scales correspond to compressed signals. The mother wavelet is chosen to serve as a prototype for all windows in the process of computing the CWT. All the windows that are used are the dilated (or compressed) and shifted versions of the mother wavelet. There are a number of functions that are used for this purpose. The Morlet wavelet and the Mexican hat function are two common examples.

The above definition of the CWT (3.21) shows that the wavelet analysis is a measure of similarity between the basis functions (wavelets) and the signal itself. The calculated CWT coefficients refer to the closeness of the signal to the wavelet at the current scale.

An interesting fact is that there do exist well-behaved functions $\psi(t)$ that can be used as prototype wavelets. This is in sharp contrast with the STFT, where it is impossible to have orthonormal bases with functions well localized in time and frequency (that is, for which the time-bandwidth product is a finite number).

3.4.5 Discretization of the Continuous Wavelet Transform: The Wavelet Series

It is apparent that neither the FT, nor the STFT, nor the CWT can be practically computed on computers by using analytical equations, integrals, etc. It is, therefore, necessary to discretize the transforms by sampling the time-frequency (scale) plane. In the case of WT, the scale change can be used to reduce the sampling rate (according to Nyquist's rule), which will save a considerable amount of computation time. Therefore, if the time-scale plane needs to be sampled with a sampling rate of N_1 at scale s_1 , the same plane can be sampled with a sampling rate of N_2 , at scale s_2 , where, $s_1 < s_2$ (corresponding to frequencies $f_1 > f_2$) and $N_2 < N_1$. The actual relationship between N_1 and N_2 is

$$N_2 = \frac{s_1}{s_2} N_1, \text{ or}$$

$$N_2 = \frac{f_2}{f_1} N_1 \quad (3.22)$$

The discretization can be done without any restriction as far as the analysis of the signal is concerned. If synthesis is not required, even the Nyquist criteria does not need to be satisfied. The restrictions on the discretization and the sampling rate become important if, the

signal reconstruction is desired. The signal can be reconstructed if the time and scale parameters are discretized under certain conditions. The scale parameter s is discretized first on a logarithmic grid. The time parameter is then discretized with respect to the scale parameter, i.e., a different sampling rate is used for every scale. In other words, the sampling is done on the *dyadic* sampling grid. Details about this sampling process can be found in [14]

The mathematical interpretation of the discretization procedure is as follows: The scale discretization is $s = s_o^j$, and translation discretization is $\tau = ks_o^j \tau_o$, where $s_o > 1$ and $\tau_o > 0$. It is important to note that the translation discretization is dependent on scale discretization with s_o .

The continuous wavelet function was defined as:

$$\psi_{\tau,s} = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right)$$

by inserting $s = s_o^j$, and $\tau = ks_o^j \tau_o$, we get

$$\psi_{j,k}(t) = s_o^{-j/2} \psi(s_o^{-j} t - k \tau_o) \quad (3.23)$$

If $\psi_{j,k}$ constitutes an orthonormal basis, the wavelet series transform becomes

$$\Psi_x^{\psi_{j,k}} = \int x(t) \psi_{j,k}^*(t) dt, \text{ or} \quad (3.24)$$

$$x(t) = c_\psi \sum_j \sum_k \Psi_x^{\psi_{j,k}} \psi_{j,k}(t) \quad (3.25)$$

A wavelet series requires that $\psi_{j,k}$ are either orthonormal, biorthogonal, or frame [14].

If $\psi_{j,k}$ are not orthonormal, Equation 3.24 becomes

$$\Psi_x^{\psi_{j,k}} = \int x(t) \psi_{j,k}^*(t) dt \quad (3.26)$$

where $\psi_{j,k}^*(t)$, is either the dual biorthogonal basis or dual frame.

If $\psi_{j,k}$ are orthonormal or biorthogonal, the transform will be non-redundant, where as if they form a frame, the transform will be redundant. On the other hand, it is much easier to find frames than it is to find orthonormal or biorthogonal bases [14].

3.4.6 The Discrete Wavelet Transform

The wavelet series is simply a sampled version of the CWT, and the information it provides is highly redundant as far as the reconstruction of the signal is concerned. This redundancy requires a significant amount of computation time and resources. The discrete wavelet transform (DWT), on the other hand, provides sufficient information both for analysis and synthesis of the original signal, with a significant reduction in the computation time. The DWT is considerably easier to implement when compared to the CWT and various algorithms exist that are used for its computation. The *Subband Coding* scheme for implementation of DWT is explained next.

3.4.7 The Subband Coding and The Multiresolution Analysis

In case of DWT, a time-scale representation of a digital signal is obtained using digital filtering techniques in DWT. The DWT analyzes the signal at different frequency bands with different resolutions by decomposing the signal into a coarse approximation and detail information. DWT employs two sets of functions, called *scaling* functions and *wavelet* functions, which are associated with low pass and high pass filters, respectively. The decomposition of the signal into different frequency bands is simply obtained by successive

high pass and low pass filtering of the time domain signal. The original signal (say $x[n]$) is first passed through a half-band high pass filter $g[n]$ and a low pass filter $h[n]$. After the filtering, half of the samples can be eliminated according to the Nyquist's rule, since the signal now has a highest frequency of $\pi/2$ radians instead of π . The signal can, therefore be subsampled by 2. This constitutes one level of decomposition and can mathematically be expressed as follows:

$$y_{high}[k] = \sum_n x[n].g[2k - n] \quad (3.27)$$

$$y_{low}[k] = \sum_n x[n].h[2k - n] \quad (3.28)$$

where $y_{high}[k]$ and $y_{low}[k]$ are the outputs of the high pass and low pass filters, respectively, after subsampling by 2.

Although it is not the only possible choice, DWT coefficients are usually sampled from the CWT on a dyadic grid, i.e., $s_0 = 2$ and $\tau_0 = 1$, yielding $s=2^j$ and $\tau=k.2^j$, as described earlier. Subband coding starts with passing this signal through a half band digital lowpass filter with impulse response $h[n]$. This filtering operation can be written in the form of the convolution sum as follows:

$$x[n] * h[n] = \sum_{k=-\infty}^{\infty} x[k].h[n - k] \quad (3.29)$$

This decomposition halves the time resolution and doubles the frequency resolution. The above procedure can be repeated for further decomposition. At every level, the filtering and

subsampling will result in half the time resolution and double the frequency resolution. Figure 3.4 illustrates this procedure, where $x[n]$ is the original signal to be decomposed, and $h[n]$ and $g[n]$ are lowpass and highpass filters, respectively. The bandwidth of the signal at every level is marked on the figure as "f".

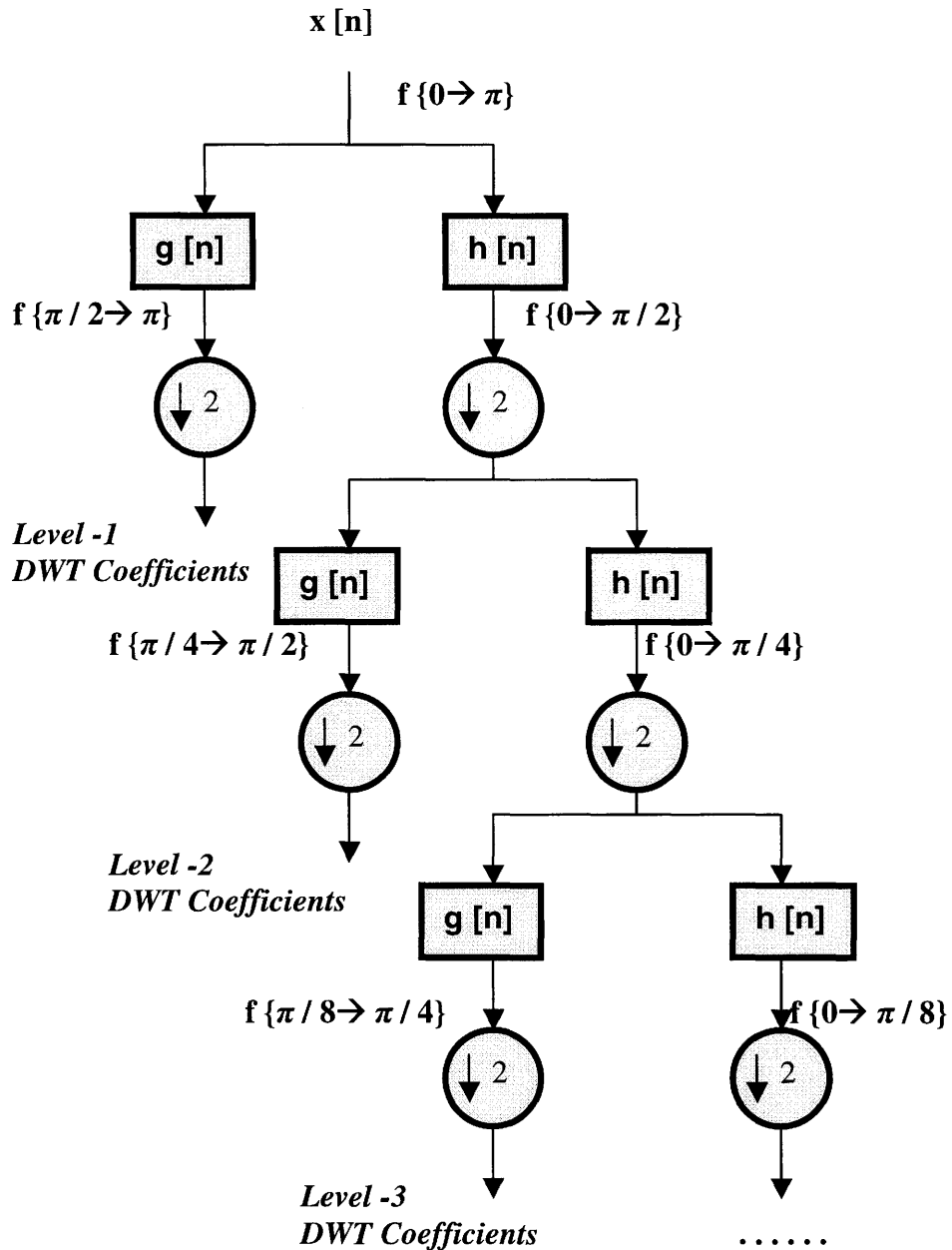


Figure 3.4 The Subband Coding Algorithm

The frequencies that are most prominent in the original signal will appear as high amplitudes in that region of the DWT signal that includes those particular frequencies. Also, the time localization will have a resolution that depends on which level they appear. If the main information of the signal lies in the high frequencies, the time localization of these frequencies will be more precise, since they are characterized by more number of samples. If the main information lies only at very low frequencies, the time localization will not be very precise.

This procedure in effect offers a good time resolution at high frequencies, and good frequency resolution at low frequencies. Due to successive subsampling by 2, the signal length must be a power of 2, or at least a multiple of power of 2, in order for this scheme to be efficient. The length of the signal determines the number of levels that the signal can be decomposed.

One important property of the discrete wavelet transform is the relationship between the impulse responses of the highpass and lowpass filters [10]. The highpass and lowpass filters are related by

$$g[L - 1 - n] = (-1)^n . h[n] \quad , \quad (3.30)$$

where $g[n]$ is the highpass, $h[n]$ is the lowpass filter, and L is the filter length. The two filters are odd index alternated reversed versions of each other. Lowpass to highpass conversion is provided by the $(-1)^n$ term. Filters satisfying this condition are known as the *Quadrature Mirror Filters (QMF)*[10].

The two filtering and subsampling operations can be expressed by the following two equations:

$$y_{high}[k] = \sum_n x[n].g[-n + 2k] \quad (3.31)$$

$$y_{low}[k] = \sum_n x[n].h[-n + 2k] \quad (3.32)$$

The reconstruction in this case is very easy, since halfband filters form orthonormal bases. The above procedure is followed in reverse order for signal reconstruction. The signals at every level are upsampled by two, passed through the synthesis filters $g'[n]$, and $h'[n]$ (highpass and lowpass, respectively), and then added. The analysis and synthesis filters are identical to each other, except for a time reversal. Therefore, the reconstruction formula becomes (for each layer)

$$x[n] = \sum_{k=-\infty}^{\infty} (y_{high}[k].g[-n + 2k] + (y_{low}[k].h[-n + 2k])) \quad (3.33)$$

However, if the filters are not ideal halfband, then perfect reconstruction cannot be achieved. Although it is not possible to realize ideal filters, under certain conditions it is possible to find filters that provide perfect reconstruction. The most well known wavelets are the ones developed by Ingrid Daubechies, and they are known as Daubechies' wavelets [14].

The flexibility of choosing the appropriate mother wavelet is one of the strongest advantages of using Wavelet transforms, since the choice of mother wavelet for a particular

problem improves the signal processing capability of the technique [1]. Tailoring the mother wavelet to the actual problem is mostly possible and should be done.

3.5 Procedure and Results

3.5.1 Preprocessing

As described earlier, the flaw data is available in the B-scan format. The number of A-scans in a flaw B-scan varies from 8 to around 15. The selected A-scans were chosen to be 101 points in length. The A-scans were time-aligned since the features to be extracted in this case were the DWT coefficients, and the discrete wavelet transform is not independent of timing of the input signal as is the case with the FFT. The length of the time-aligned signal was 256 points, since zeros were to be appended to the signal. After time-alignment, the A-scans were processed to be zero-mean signals and finally normalized before extracting the features. Figure 3.5 and 3.6 show the crack and deposit A-scans after pre-processing.

3.5.2 Feature Extraction

After the preprocessing stage, the time-aligned, zero-mean, and normalized A-scans are applied to the subband coding algorithm for obtaining DWT coefficients as features. Since the data length is 256, eight levels of decomposition were taken, which included 8 detailed level coefficients and one, level 8 approximate coefficients. The length of the DWT coefficients was 306 points as shown in Figure 3.7 and 3.8 for a crack and a deposit signal respectively. Daubechies wavelet was used as the mother wavelet for the decomposition. As the frequency of the transducers is 8-11 MHz, the relevant frequency band was selected amongst the eight detailed and one approximate coefficient. After careful consideration of the output signal and the various frequency bands at each level of decomposition, the coefficients in the range 75 to 112 were finally selected as features. This range corresponds

to the frequency band of interest (since the transducers have a central frequency of 8-11MHz). Before applying these coefficients to a classification algorithm, the data was again normalized to improve the classification performance. Finally, the 38-point DWT coefficients were applied as the input to the classifier.

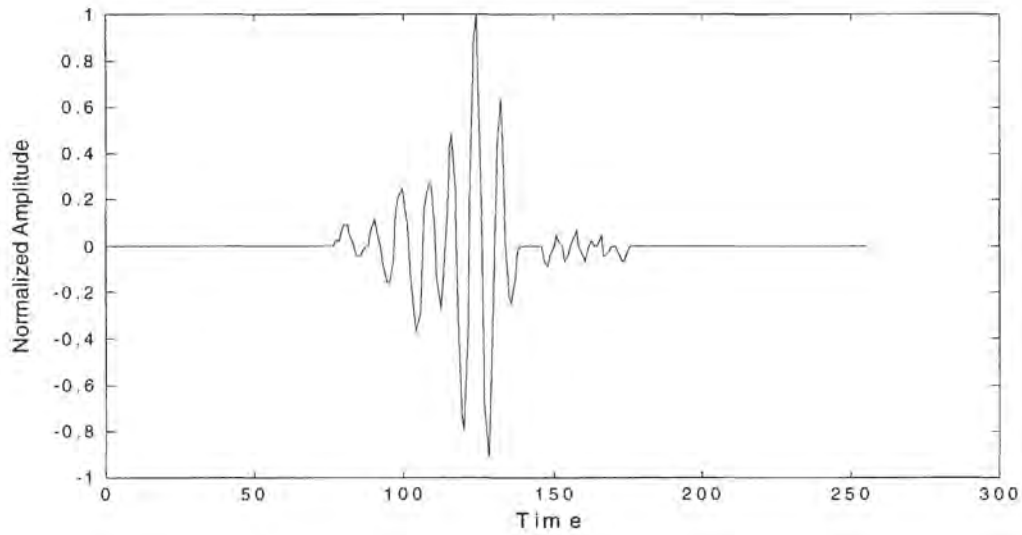


Figure 3.5 Pre-processed crack signal

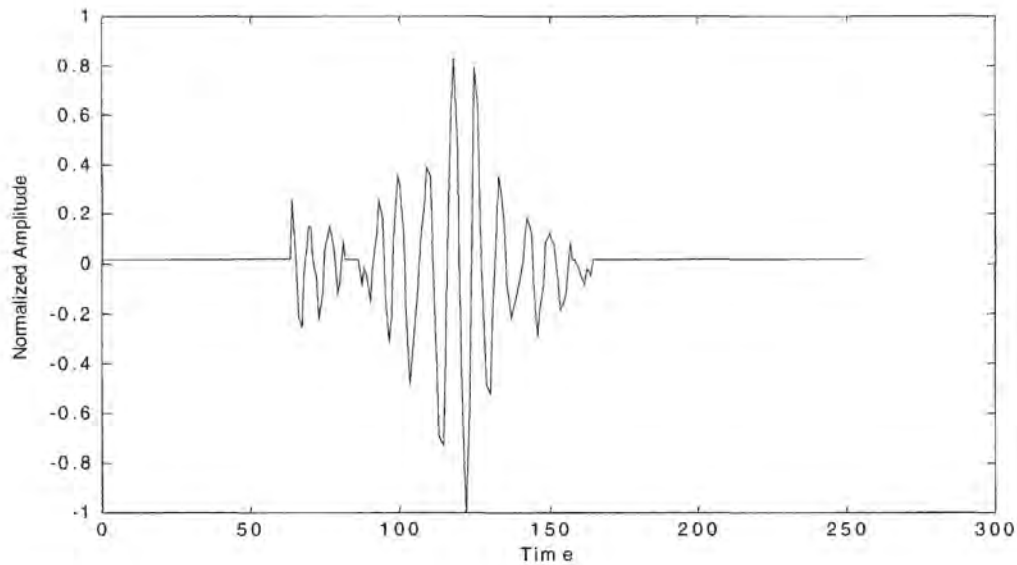


Figure 3.6 Pre-processed deposit signal

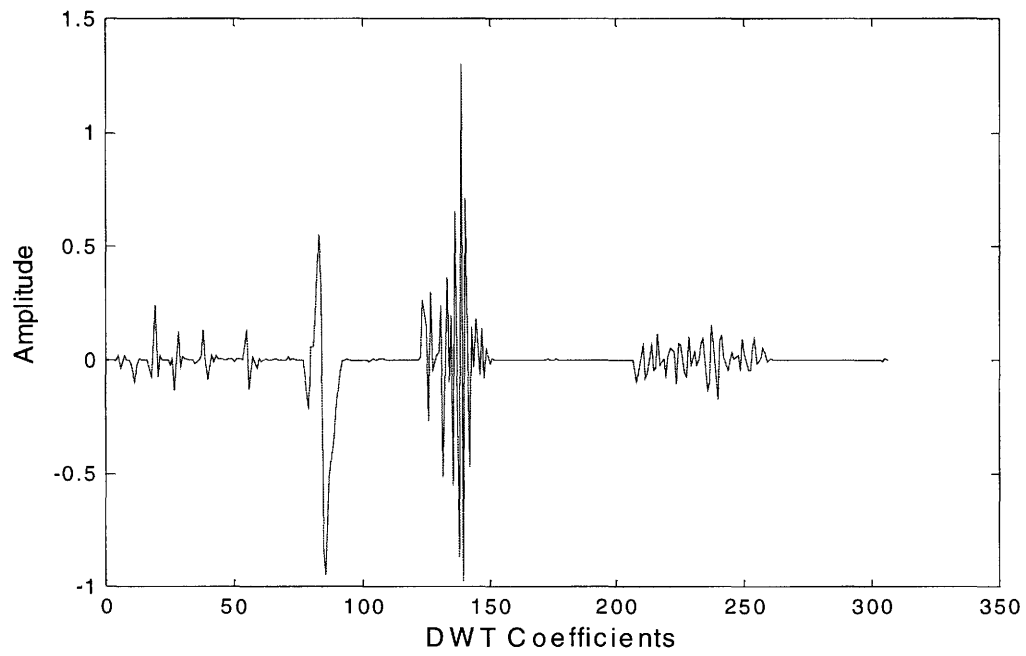


Figure 3.7 DWT coefficients of a crack signal

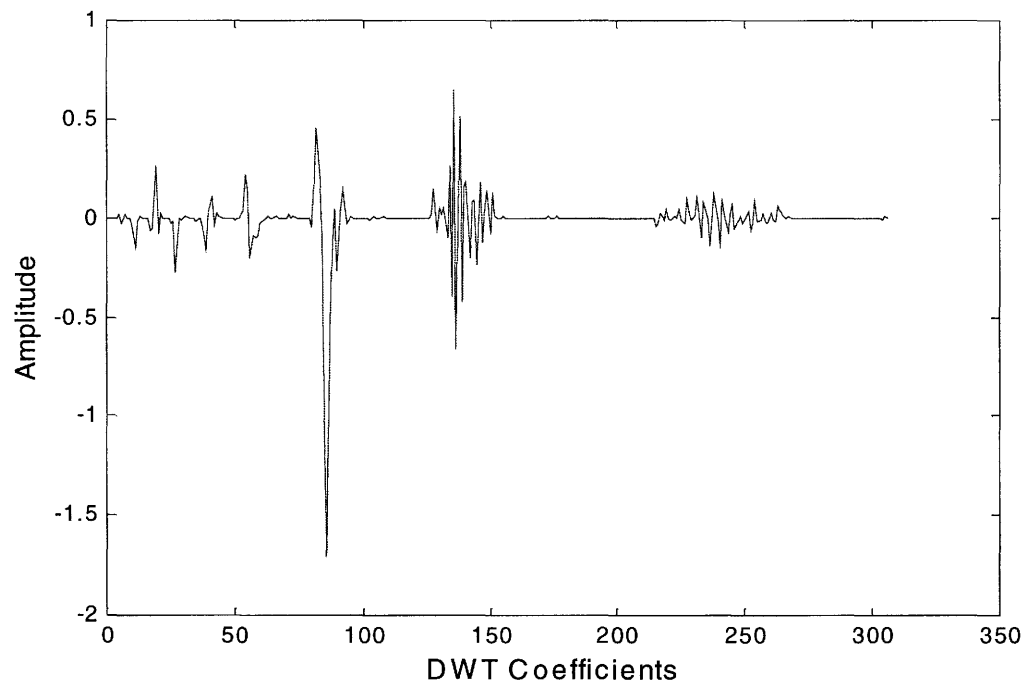


Figure 3.8 DWT coefficients of a deposit signal

CHAPTER 4. CLASSIFICATION

4.1 An Overview of Pattern Recognition

The objective of a pattern recognition or classification system is to determine, on the basis of the observed information, the pattern class responsible for generating a set of measurements similar to the observed data. Correct classification depends on the amount of discriminatory information contained in the measurements and the effective utilization of this information during the preprocessing and feature extraction stages.

The key to defect classification in ultrasonic NDE is to acquire the unique capability to learn the patterns and their inter-relationships. In defect classification for material testing, there are no “standard” defects, and thus learning from a class of defect specimen is much needed to perform effective classification of a *new* specimen. Also, the number of training samples is always limited. *Neural networks* are particularly suited for pattern classification in these situations [4].

4.2 Introduction to Neural networks

A *Neural Network* is a system inspired by biological learning systems and is loosely modeled on the basis of human brain’s learning process. The field of Neural Networks goes by many names, such as parallel distributed processing, neuro-computing, natural intelligent systems, connectionism, machine learning algorithms, and artificial neural networks [18]. It is an attempt to simulate within specialized hardware or sophisticated software, the multiple layers of simple processing elements called *neurons*. Each neuron is linked to certain of its neighbors with varying coefficients of connectivity that represent the strengths or *weights* of

these connections. Learning is accomplished by adjusting these weights to cause the overall network to output appropriate results. In other words, we can train a neural network to perform a particular function by adjusting the values of the weights between neurons.

Commonly neural networks are adjusted, or *trained*, so that a particular input leads to a specific target output. Such a situation is shown in Figure 4.1. Note that the network is adjusted, based on a comparison of the output and the target, until the network output matches the target [9]. Once this is accomplished, the network is said to have *converged*. Typically many such input/target pairs are used to train a network. This type of training is called *supervised training*. The supervised training methods are commonly used in practice, but other networks can be obtained from *unsupervised training* techniques or from *direct design* methods. Unsupervised networks can be used, for instance, to identify groups of data.

4.3 Applications of Neural Networks

Neural networks have been used to perform complex functions in various fields of application including pattern recognition, identification, classification, speech, vision, and control systems [20]. The field of neural networks has a history of some five decades but has found solid applications only in the past fifteen years, and the field is still developing rapidly. Modern applications include high performance aircraft autopilot, aircraft control systems, aircraft component fault detection, automobile automatic guidance systems, credit application evaluation in banking field, target tracking and object discrimination in various weapon systems, integrated circuit chip layout and nonlinear modeling in electronics industry, trajectory control, forklift robot, and manipulator controllers in robotics, and animation and special effects in the entertainment industry [9].

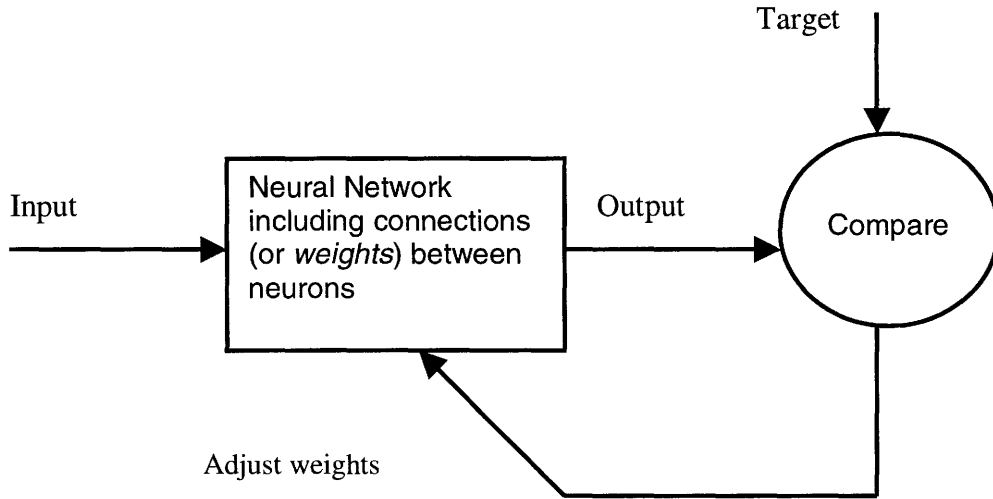


Figure 4.1: Basic neural network model

A number of neural network models have been developed in the recent years for various applications. A taxonomy of six neural networks that can be used as classifiers [4] includes the Hopfield net, Hamming net, Carpenter/ Grossberg net, single and multi-layer perceptrons, and the Kohonen self-organizing feature maps. In general, the recognition capability of a network model is mainly dependent on the available learning algorithms that also determine the convergent properties. It is very difficult to rank these network models even for a specific application. However, the *multi layer perceptron model with back propagation learning* algorithm is one of the most commonly used. Besides the six models listed above, there are several other models suitable for pattern recognition. These include the Nestor Learning Systems (NLS) and its component RCE networks, cellular neural networks, and the Cerebellar Model Articulation Controller (CMAC) neural networks [4].

4.4 The MLP Feedforward Back propagation Neural Network

The multi layer perceptron, having on or more hidden layers, and using a back propagation learning algorithm has the ability to construct arbitrarily complex nonlinear decision boundaries to separate pattern classes. Properly trained back propagation networks tend to give reasonable answers when presented with inputs they have never seen. Typically, a new input will lead to an output similar to the correct output for input vectors used in training that are similar to the input being presented. This generalization property makes it possible to train a network on a representative set of input/output pairs and get good results without training the network on all possible input/output pairs [9]. Figure 4.2 shows the basic architecture of such a network.

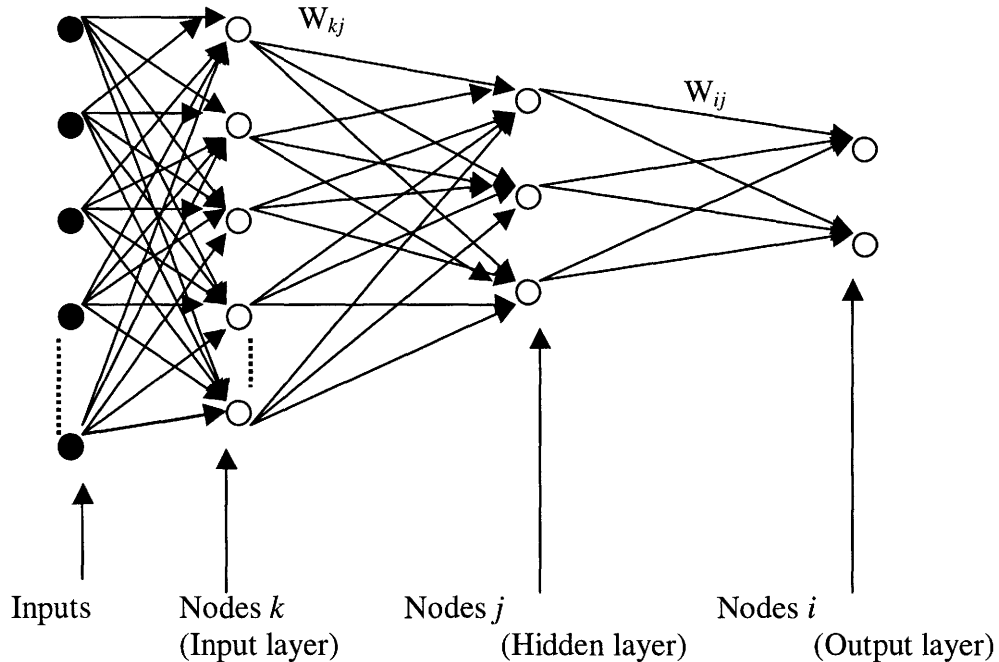


Figure 4.2 The MLP feed forward neural network

4.4.1 The Back Propagation Algorithm

The back propagation algorithm derives its name from the fact that information about errors is also propagated backwards through the network layers and is used to adjust the connections between the layers, thus improving performance.

The algorithm is based on the selection of a suitable error function or cost function, whose values are determined by the actual and desired outputs of the network and the network parameters such as the weights and thresholds. The basic idea is that the cost function has a particular surface over the weight space and therefore, an iterative process such as the gradient descent method can be used for its minimization. The gradient descent method is based on the fact that the gradient of a function always points in the direction of maximum increase of the function. Then, moving to the direction of negative gradient induces a maximal “downhill” movement that will eventually reach the minimum of the function surface over its parameter space. This is a well-established technique for minimizing functions, but does not guarantee that one finds a global minima every time.

The back-propagation training consists of two computational passes: a *forward* and a *backward* pass. In the forward pass computation, as shown in Figure 4.2, given the selected input vectors $Y_k^{(p)}$ and weight w_{jk} between hidden nodes and input nodes, each hidden node j will produce the output

$$y_j^{(p)} = f\left(\sum_k w_{jk} y_k^{(p)}\right), \quad (4.1)$$

where $f(\cdot)$ is the transfer function. Then, each output node will produce the final output

$$y_i^{(p)} = f\left(\sum_j w_{ij} f\left(\sum_k w_{jk} y_k^{(p)}\right)\right) \quad (4.2)$$

In the backward pass, the error or cost function calculated as the mean square error (MSE) over all output units and over all patterns p is given by

$$E = \frac{1}{2} \sum_p \sum_i (d_i^{(p)} - y_i^{(p)})^2, \text{ or}$$

$$E = \frac{1}{2} \sum_p \sum_i (d_i^{(p)} - f(\sum_j w_{ij} f(\sum_k w_{jk} y_k^{(p)})))^2 \quad (4.3)$$

Now, the gradient descent method is applied to find the gradient descent rule for the hidden layer to output layer connections as follows

$$\Delta w_{ij} = -\eta (\partial E / \partial w_{ij}), \quad (4.4)$$

where η is a constant that determines the rate of learning. Buy using the chain rule, we find

$$\Delta w_{ij} = \eta \sum_p (d_i^{(p)} - y_i^{(p)}) f'(x_i^{(p)}) y_i^{(p)} = \eta \sum_p \delta_i^{(p)} y_i^{(p)} \quad (4.5)$$

In the same manner the gradient descent rule for the input layer to hidden layer connections can be found as follows.

$$\Delta w_{jk} = -\eta (\partial E / \partial w_{jk}) \quad (4.6)$$

By using chain rule, we get

$$\Delta w_{jk} = \eta \sum_p \sum_i (d_i^{(p)} - y_i^{(p)}) f'(x_i^{(p)}) w_{ij} f'(x_i^{(p)}) y_k^{(p)}, \text{ or}$$

$$\Delta w_{jk} = \eta \sum_p \delta_j^{(p)} y_k^{(p)} \quad (4.7)$$

Where

$$\delta_j^{(p)} = f'(x_j^{(p)}) \sum_i \delta_i^{(p)} w_{ij} \quad (4.8)$$

From the equations above, it can be seen that the transfer function must be differentiable. If a *sigmoid* function is chosen as the transfer function, that is

$$f(x) = \frac{1}{1+\exp(x)} \quad (4.9)$$

It can be found that

$$f'(u) = f(u) [1 - f(u)] \quad (4.10)$$

4.5 Procedures and Implementation

The features used for training the neural network were the DWT coefficients that were explained in chapter 3. Various architectures of neural network were tested for optimum performance on the available data. The output layer always has 2 nodes in this problem since it is a *two-class* problem, i.e., the output is either a *crack* or a *deposit*. The number of nodes in the input layer and the number of layers of the neural network were changed and tested for best results. The best results were obtained with a neural network consisting of one input layer with 19 nodes and one output layer with 2 nodes. The length of the input feature vector was 38.

The available database had limited number of crack and deposit files. The complete database was divided into three parts, namely *training*, *testing*, and *validation* databases. The neural network was trained with the randomized training database using the back propagation algorithm and then the performance was tested with the testing database. The training and testing databases were again mixed, randomized, and divided into a new set of training-testing database. Again the neural network was trained with the *new* training database and its performance tested with the *new* testing database. This procedure was repeated many times so as to achieve optimum results. This is an important implementation consideration because

of mainly two reasons. Firstly, the neural network needs to be *trained* on the *best* available data for optimum performance. Secondly, there is randomization involved during the training phase since the algorithm begins with a random selection of the training data itself. Therefore, this process of data shuffling and testing results in an optimum set of “neural network architecture and available training data set”. After obtaining the best possible results with the training-testing databases, the neural network’s performance was finally *validated* with the validation database. It is important to mention that the validation database was not a part of the data set employed for training the neural network. This is important for validating the performance of a neural network since this stage of validation really shows how the network will behave when it encounters new input data.

CHAPTER 5. RESULTS AND DISCUSSION

5.1 The Results Obtained- An Analytical Discussion

As described earlier, the most critical stage in the entire process of designing a classification is the extraction and selection of appropriate features that carry significant discriminatory information between the classes. Having obtained such features, the classification problem is much simplified since there are many known useful algorithms that can classify both linearly and non-linearly separable classes.

Fast Fourier Transform (FFT) coefficients were first tried as features to be tested by the classification stage. The results obtained using the MLP back propagation neural network were not satisfactory at all. The reason for this bad performance is that the DFT is not useful for analyzing time-varying signals (like the A-scans). The time-localization of various frequency components present in the signal is lost and thus the DFT coefficients did not contain enough discriminatory information to distinguish the two classes from each other.

Discrete Wavelet Transform coefficients were obtained using the subband coding algorithm. The DWT, as mentioned earlier, is a very powerful signal analysis tool that works like a magnifying glass for analyzing time-varying signals. Using this algorithm, very good time resolution of the signal is obtained at higher frequencies (with a relatively poor frequency resolution) and a good frequency resolution is achieved at lower frequencies (with of course, a poor time resolution). The relevant DWT coefficients were chosen as features after carefully determining the frequency band of interest in this particular case. Then, the selected coefficients were applied to the MLP back propagation neural network.

Table 5.1. Summary of classification results for A-scans.

	CRACKS	DEPOSITS
Total A-Scans (approximately)	350	480
A-scans used for Training/Testing	200	200
A-scans used for Validation	150	280
Misclassifications in Validation	21	39
Percentage Performance	86	86.07
Overall Performance(%)	86	

The results obtained in this case were quite reasonable and are summarized in the table 5.1. These results are the best ones obtained with the available data set. This means that the overall classification performance was tested for various combinations of the training/testing data and different neural network architectures as explained in Chapter 4. The

performance varied slightly (approximately 3% to 5 %) with the different sets of training/testing data. Also, neural networks with one or two hidden layers were performing well for the 38-point feature vector and introducing a third hidden layer did not improve the performance; instead it resulted in an increased training period. It was found that the performance of the neural network with 19 nodes in one hidden layer produced the best results with a reasonable training time.

The results shown in Table 5.1 were obtained using the A-scans from the crack and deposit B-scan files. Since each B-scan of a flaw signal consisted of 8 to around 18 A-scans, and the classification stage is required to classify the B-scans, these results were improved when computed for the B-scans as compared to those obtained from A-scans. These results with B-scans are listed in table 5.2.

5.2 Conclusions and Future Work

Steam generator tubes form an essential part of a nuclear power plant. Routine testing of these tubes is of considerable importance in order to detect tube degradation and other types of defects that can cause unsafe operation of the system. In this research project, a classification system was developed that distinguishes the crack signals from those of deposits in the steam generator tubes.

Reasonably good results were obtained using the DWT coefficients as features and the MLP back propagation neural network as the classifier. The DWT was proven to be a strong analysis tool and it produced significantly better results as compared to the FFT. Also, the back propagation neural network worked very well as a classifier for the two non-linearly separable classes.

Table 5.2. Summary of classification results for B-scans.

	CRACKS	DEPOSITS
Total B-Scans Available	62	88
B-scans used for Training/Testing	35	40
B-scans used for Validation	27	48
Misclassifications in Validation	2	2
Percentage Performance	92.6	95.8
Overall Performance (%)	94.2	

It is worth mentioning that the database available to us in this project was very limited, especially in case of crack files. The classification results can be further improved if the neural network can be trained with more crack and deposit data. Also, in this research the emphasis was only on obtaining frequency-domain features. This approach is based on the assumption that the cracks and deposits are acting as filters with different frequency

responses for the incoming transducer signals. In other words, both cracks and deposits attenuate different frequencies differently. Other techniques may be used for obtaining different set of features. For example, if the kinematics of the problem is carefully studied and the exact locations of the transducers can be estimated, the inner diameter (ID) and outer diameter (OD) can be precisely computed with the available B-scan data, then the physical location of the flaw can directly give the information about a crack or a deposit. The Synthetic Aperture Focusing Technique (SAFT) can be applied to eliminate or at least minimize the effects of the transducer aperture beam width over the B-scan image [16]. SAFT has been used to restore ultrasonic images obtained either from B or C scans with focusing distortion. With the use of this technique, the image resolution can be improved. The images of the B-scan have a lateral resolution limited by the focus of the ultrasonic transducer, which depends on the aperture area of the transducer.

SAFT provides an accurate measurement of the spatial location and extent of flaws contained in objects such as structural components and welds and nuclear power reactor systems. The transit-time for the ultrasonic beam to travel to and from a point is a hyperbolic function of the probe position and target depth. When the equation of this hyperbola is known, A-scan signals can be shifted in time and added together. When a defect is present constructive interference of the waveforms form a large signal. When no defect is present the interference is destructive and the signal is small. This SAFT processing may be performed in either two dimensions or three dimensions; however, three dimensional SAFT requires considerable processing time [22].

REFERENCES

1. A. Abbate, J. Koay, J. Frankel, S. C. Schroeder, P. Das, "Signal detection and noise suppression using a wavelet transform signal processor: Application to ultrasonic flaw detection" IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control Vol.44, pp.14-26, 1997.
2. A. Abbate, J. Koay, J. Frankel, S. C. Schroeder, P. Das, "Application of Wavelet transform signal processor to ultrasound" IEEE Ultrasonics Symposium Proceedings Vol. 2, pp. 1147 - 1152, 1994.
3. C. S. Burrus, R. A. Gopinath, H. Guo, Introduction to Wavelets and Wavelet transform: A primer, Prentice-Hall Inc., 1998.
4. C. H. Chen, "A comparison of neural network models for pattern recognition" IEEE pp. 45-46, 1990
5. T.A.C.M. Classen, W.F.G. Mecklenbrauker, "The Winger distribution – A tool for time-frequency analysis: Part I: Continuous time signals" Philips Journal of Research Vol.35,pp. 217-250, 1980.
6. T.A.C.M. Classen, W.F.G. Mecklenbrauker, "The Winger distribution – A tool for time-frequency analysis: Part II: Discrete time signals" Philips Journal of Research Vol.35,pp. 276-300, 1980.
7. I. Daubechies, "The Wavelet transform, time-frequency localization and signal analysis" IEEE Transactions on Information theory, Vol. 36, pp. 961-1005, 1990.
8. I. Daubechies, "Different perspectives on Wavelets" Proceedings of symposia in applied mathematics, American Mathematical Society short course, Vol. 47, pp 1-117, January 1993.

9. C. G. Looney, Pattern Recognition using neural networks: Theory and algorithms for engineers and scientists, Oxford University Press, 1997.
10. S. G. Mallat, "A theory for multiresolution signal decomposition: The Wavelet representation" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.11, pp. 674-693, July 1989.
11. S. G. Mallat, "Multiresolution approximation and Wavelet orthonormal bases of L^2 " Transactions of American Mathematical Society, Vol.315, pp. 69-87, September 1989.
12. A. V. Oppenheim, R. W. Schaffer, Discrete time signal processing, Fifth Edition, Prentice-Hall, 1998.
13. J. G. Proakis, D. G. Manolakis, Digital signal processing principles, algorithms and applications, Third Edition, Prentice-Hall, 1996.
14. R. Polikar, "The Wavelet tutorial", last viewed June 2001. Available at
<<http://www.engineering.rowan.edu/~polikar/WAVELETS/WTtutorial.html>>
15. O. Rioul, M. Vetterli, "Wavelets and signal processing" IEEE Signal Processing Magazine, pp. 14-38, October 1991
16. K. Schroeder, C. Haas, M. Kruger, H. Ermert, T. Rastello, D. Vray, G. Gimenez, "Synthetic aperture-based reconstruction of intravascular ultrasound images in the time domain and frequency domain" Proceedings of the IEEE International Symposium on Industrial Electronics, Vol. 3, pp. 1502 – 1507, 1999.
17. L. W. Schmerr, Jr., Fundamentals of ultrasonic nondestructive evaluation, Plenum Press, New York, 1998.
18. S. Theodoridis, K. Koutroumbas, Pattern Recognition, Academic Press, 1999.

19. R.B. Thompson, D.O. Thompson, "Ultrasonics in nondestructive evaluation" IEEE Proceedings Vol. 73, pp. 4-55, December 1985.
20. L. Udpa, S. S. Udpa, "Neural networks for the classification of nondestructive evaluation signals" IEE Proceedings on Radar and Signal Processing Vol. 138, pp. 41-45, February 1991
21. R. Xing, Ultrasonic NDE signal classification on steam generator tubes, M.S. Thesis, Iowa State university, 2000.
22. Yamani, "Three-dimensional imaging using a new synthetic aperture focusing technique" IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, Vol. 44, pp. 943-947, 1997.
23. "Nondestructive testing encyclopedia", The e-Journal of Nondestructive Testing, last viewed June 2001. Available at <http://www.ndt.net/article/az/ut_idx.htm>