ROUGH SET BASED IDENTIFICATION AND CLASSIFICATION OF SINGLE AND MULTIPLE PARTIAL DISCHARGE SOURCES

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ELECTRICAL ENGINEERING
2016

By

ARNAB BAUG

EXAMINATION ROLL NO.: M4ELE1614

REGISTRATION NO.: 128906 of 2014-15

JADAVPUR UNIVERSITY

Under the Supervision of

Dr. BISWENDU CHATTERJEE

DEPARTMENT OF ELECTRICAL ENGINEERING
FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY
KOLKATA-700032

&

Dr. SOVAN DALAI

DEPARTMENT OF ELECTRICAL ENGINEERING
FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY
KOLKATA-700032

JADAVPUR UNIVERSITY

Faculty of Engineering and Technology

CERTIFICATE OF RECOMMENDATION

*This is to certify that this dissertation entitled "**Rough Set based Identification and Classification of Single and Multiple Partial Discharge Sources**" has been carried out by **ARNAB BAUG** under our supervision and be accepted in partial fulfillment of the requirement for the degree of Master of Electrical Engineering.*

Dr. Biswendu Chatterjee
Assistant Professor
Electrical Engineering Department
Jadavpur University
Kolkata-700032

Dr. Sovan Dalai
Assistant Professor
Electrical Engineering Department
Jadavpur University
Kolkata-700032

Prof. Swapan Kumar Goswami
Head of the Department
Electrical Engineering Department
Jadavpur University
Kolkata-700032

Prof. Sivaji Bandyopadhyay
Dean,
Faculty of Engg. and Technology
Jadavpur University
Kolkata-700032

JADAVPUR UNIVERSITY

Faculty of Engineering and Technology

CERTIFICATE OF APPROVAL

The foregoing thesis is hereby approved as a credible study of an engineering subject, carried out and presented in a manner satisfactory to warrant its acceptance as pre-requisite to the degree for which it has been submitted. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein but approve the thesis only for which it has been submitted.

_____ _____

Signature of the Examiner(s) *Signature of the Supervisor*

Declaration of Originality and Compliance of Academic Ethics

I hereby declare that this thesis contains literature survey and original research work by the undersigned candidate, as part of the Master in Electrical Engineering studies.

All information in this document has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct. I have fully cited and referenced all material and results that are not original to this work.

Name : **ARNAB BAUG**

Examination Roll No. : **M4ELE1614**

Registration No. : **128906 of 2014-15**

Thesis Title : ***ROUGH SET BASED IDENTIFICATION AND CLASSIFICATION OF SINGLE AND MULTIPLE PARTIAL DISCHARGE SOURCES***

.....

Signature with Date

Acknowledgements

I would like to take this opportunity to humbly express my gratitude for the innumerable gesture of help, cooperation and inspiration that I have received from my teachers, friends and well wishers during this course.

I feel honored to express my profound regard and deep sense of gratitude to my guides **Dr. Biswendu Chatterjee** and **Dr. Sovan Dalai**, Electrical Engineering Department, Jadavpur University, Kolkata, for allowing me to do my work in this exciting field. They have been the persons to instill in me a sense of commitment, dedication and optimism. I am highly obliged and grateful for their excellent guidance, endless encouragement, unique cooperation extended to me, right from the time of onset of this colossal task till its successful completion.

I would also wish to express my sincere gratitude to Prof. Abhijit Mukherjee, Prof. Kesab Bhattacharya and Prof. Sivaji Chakravorti, Dept. of Electrical Engineering, Jadavpur University for their keen interest and active support in this work.

I would like to thank Prof. Swapan Kumar Goswami, Head of the Department, Electrical Engineering, and Jadavpur University for providing the necessary facilities for carrying out this thesis.

I offer my sincerest gratitude to Dr. Subrata Biswas, Department of Electrical Engineering, Netaji Subhash Engineering College, Kolkata and Dr. Debangshu Dey, Electrical Engineering Department, Jadavpur University, Kolkata for their continuous support and words of wisdom throughout my thesis.

I would like to give a special thanks to Mr. Riddhi Ghosh, Mr. Nasirul Haque, Mr. Soumya Chatterjee and Mr. Suhas Deb for their spontaneous support and inspiration in carrying out this work.

I would like to thank my colleagues, Niladri Ray Choudhury and Amrendra Kumar, for their help and endless effort to complete this work during M.E. course.

Words would be trite to express the extent of dependency, love and fortitude towards my parents. Without their love and encouragement it would have been impossible to complete this work.

Arnab Baug

CONTENTS

CHAPTER 1

INTRODUCTION

CHAPTER 1.

INTRODUCTION

1.1 INTRODUCTION

The phenomena of Partial Discharge has become the center of attraction of the researchers for last two decades because of its implicit influence upon the lifetime of insulation system of any high voltage equipment. Partial Discharge degrades the quality of insulating media which may ultimately transpire into failure of the equipment and subsequent revenue loss. Therefore, efficient detection of Partial Discharge phenomena inside high voltage equipments is very important for on-line and off-line condition monitoring of such equipments. This chapter provides basic insight into the Partial Discharge phenomena inside insulation systems and different detection methods of Partial Discharge that are commonly employed. At the later part of the chapter, the main objectives of the present work have been discussed alongwith brief outline of this thesis.

1.2 PARTIAL DISCHARGE

Partial Discharge (PD) is a localized electrical discharge within any insulation system, as applied in electrical apparatus, components or systems, that only partially links the insulation between conductors and may or may not occur adjacent to a conductor [1].

In insulation systems, Partial Discharges occur due to various reasons which mainly include local defects and imperfections inside the insulation such as voids, cracks or bubbles and presence of foreign particles or irregularities on the surface of the insulations [2]. The insulation system may comprise of solid, liquid or gaseous dielectrics, or any combination of these materials [3].

Generally, PDs are consequence of local electrical stress concentrations appearing in such local defects or surface irregularities when the insulation is subjected to high voltage. When a void inside an insulating medium undergoes breakdown due to high electric stress, short time bursts of charge flow through the void takes place. Such short time charge bursts give rise to PD pulses. Normally for discharges in solid and liquid media, current pulses of duration of much less than 1 μs appear through the void. In case of gaseous dielectric media, pulseless discharges of more continuous form may occur [3].

In general, PDs remain confined within a part of the dielectric medium, which explains its characterisation as 'localised electrical discharge'. Contrary to the total breakdown, PD partially bridges the electrodes between which the high voltage has been applied. Technically the term 'Partial Discharge' describes a wide variety of discharge phenomena such as [3] -

(i.) Internal discharges occurring within voids or cavities in solid dielectrics and bubbles in liquid dielectrics.

(ii.) Surface discharges occurring at the boundaries and interfaces between different types of dielectric media. Surface discharges can also occur due to presence of contaminations on the surface.

(iii.) Corona discharges occurring in gaseous media which are remotely located from solid and liquid dielectrics.

(iv.) Discharge channels in solid dielectrics due to energy impact of continuous discharges. This type of PD phenomenon is generally called treeing [4] [5].

To comprehend the PD phenomenon in a void within a solid or liquid dielectric media, consider the scheme of an insulation system as given in Figure 1.1. The insulation system comprises of solid or liquid dielectric medium between two electrodes. There is a void inside the dielectric medium which has been assumed to be gas filled. In practical situations, this insulation system represents dielectric test specimen or any HV apparatus which is undergoing PD test.

If high voltage is applied between the electrodes then the dielectric medium alongwith the void experiences very high electric field stress. The dielectric medium and the void constitute capacitances between the electrodes. In Figure 1.1, the capacitance formed by void is represented as C_V. The part of dielectric between the top electrode and top surface of void constitutes capacitance C_S'. Similarly, capacitance C_S'' is formed between bottom surface of void and bottom electrode. Clearly from Figure 1.1 it can be observed that capacitances C_V, C_S' and C_S'' are in series combination. The equivalent capacitance will be,

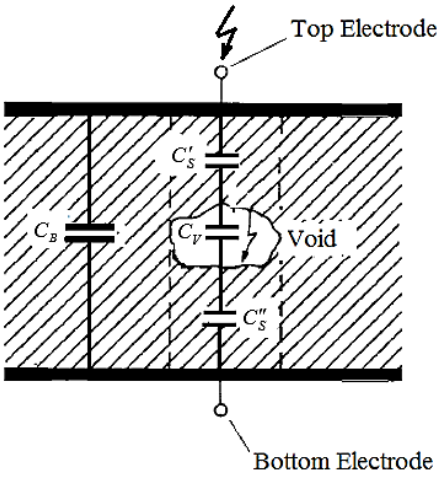


Figure 1.1 Insulation system consisting of a void

$$C_{eq} = C_S' \parallel C_V \parallel C_S'' = C_V \parallel (C_S' \parallel C_S'') = C_V \parallel C_S \tag{1.1}$$

Therefore, capacitances C_S' and C_S'' can be combined together to form capacitance C_S. Capacitance C_B represents the remaining body of the dielectric between the electrodes.

Usually for a practical insulation system, $C_B >> C_V >> C_S$.

When voltage applied between the electrodes is sufficiently high, the gaseous media inside the void undergoes breakdown because of enhanced field gradients by the difference in permittivities. This breakdown is localised within the void only and rest of the dielectric is not affected. Therefore, the void becomes a source of PD. The gaseous discharge inside void creates electrons and ions of both polarity. These charges get deposited at the surfaces of the void which creates additional polarisation of the dielectric material. The breakdown process of the void depends on various important factors such as [6 - 10] –

- Type of gaseous medium inside the void
- Size and shape of the void
- Surface properties and amount of trapped charges on the walls of the void
- Dielectric constant of the surrounding medium and environmental conditions.
- Type of contaminations present in the surrounding medium

If the applied voltage is further increased then the discharge continues. For ac voltage of sufficiently high peak amplitude, the discharge will be repeated in each cycle. Instead, if increasing dc voltage is given input then some discharge events take place during the increase of the voltage. If the dc voltage becomes steady, then partial discharges stop due to accumulated charges on the surface of the void [3].

Due to rapid discharge of void capacitance C_V, voltage across the void is considerably lowered during PD. The voltage drop across the void, due to occurrence of PD, lowers the potential across the test object. Drop in voltage across the terminals of test object will be given by,

$$\Delta V_t = \frac{C_S}{C_S + C_B} \Delta V \tag{1.2}$$

In equation (1.2), $\Delta V = \frac{\Delta q_V}{C_V}$ denotes the voltage drop across the void during discharge and Δq_V is the charge flow through the void during discharge. It seems that by measuring the voltage drop across the test object ΔV, the actual charge flow Δq_V can be measured from equation (1.1). However, values of capacitances C_S and C_B are generally not known. Due to this reason, direct measurement of actual charge flow during Partial Discharge is impossible [3].

1.3 DETECTION OF PARTIAL DISCHARGE

The impact of partial discharges, which may be a sustained one or may be sporadic in nature, on the life of insulation has long been recognized as a crucial issue and investigation of such effects have gained substantial importance [3] [11 - 13] . The number of discharge events occurring in an insulation system during a chosen time interval is directly dependent on the kind of voltage applied and will be largest for ac voltages. Every discharge event causes a short-time avalanche of high energy electrons or accelerated ions which cause chemical transformations of many types by their energy impact. This in turn results in gradual deterioration of the material and if unchecked may ultimately lead to complete failure of the apparatus involved. Hence, presence of Partial Discharge in an apparatus is an indication to transpiring dielectric defects in the insulation of the apparatus. However actual deterioration will be explicitly dependent on the type of dielectric media used. By measuring the intensity of PD inside an equipment, qualitative and quantitative idea about the deterioration of insulation of that equipment can be obtained [14].

Measurement and detection of partial discharge has drawn considerable attention in recent times and most of the research in this field are directed towards establishment of correlation between partial discharge and longevity of the equipments [3]. However such quantitative formulations are very hard to guarantee. Needless to say, early detection and localization of Partial Discharges may prevent the impending apparatus outage and the resulting inconvenience or revenue loss. Therefore, for effective condition monitoring of equipments and systems, efficient detection and measurement of partial discharge is very crucial [13].

As partial discharges cannot be measured directly, usually its energy by-products are measured [15]. Different techniques that are employed for PD measurement are as follows-

- **High Frequency Radio Waves**: Partial Discharges can be measured by detecting and analyzing High Frequency (HF) radio waves arising out of PD sources [16 - 24]. The PD measurement by HF radio waves can be segregated into two classes based on the range of frequency chosen for measurement - Very High Frequency (VHF) measurement and Ultra High Frequency (UHF) measurement.

  For measurement of PDs by VHF radio waves, the frequency range of 30 MHz to 300 MHz is generally used [15]. Measurement of VHF radio waves are usually performed by employing aerial antennas [16] [18]. Less commonly, window-type receiver devices are used to monitor the VHF radio waves. The signal attenuation and noise reduction in this type of measurement is less compared to measurement of UHF radio waves [17].

For UHF measurement, normally the frequency range of 300 MHz to 3 GHz is chosen [14]. UHF measurements are performed using window style radio wave receiver [19] [24]. It is one of the widely used technique for PD measurement because the UHF radio waves are less affected by external noise [20 - 23]. However, UHF radio signals are strongly attenuated when they pass through the bushings or travel longer distances within the oil and noises may be introduced by external sources which also operate in the same UHF band. Radio waves generated due to discharges occurring in oils have an upper frequency limit of only a few hundred kilohertz and cannot be detected by this method.

- **Radio Frequency Current Transformers (RFCTs)**: PD Measurement by RFCTs also involve detection of the radio waves but they are generally designed to measure comparatively low frequency span, up to several hundreds of MHz. They can be employed for both off-line and on-line measurements.
- **Coupling Capacitor (CC)**: It is a type of electrical PD detection method and is most frequently used [3] [25]. This method aims to approximate the effects of charge flow due to PD pulse by measuring a quantity called apparent charge, which is defined as the unipolar charge which, if injected between the terminals of the relevant test object within a very short time span, would give the same reading on the measuring instrument as the actual PD current pulse [3]. Apparent charges are generally measured in picocoulombs. Series assembly of a stable storage capacitor and a low resistance coaxial shunt is connected across the test object [1]. The storage capacitor is called the coupling capacitor. It injects the apparent charge to the test object during the short time span of PD. The measuring instrument is connected across the shunt.
- **Acoustic Emissions:** Measurement of PD by detecting and analyzing acoustic emissions from PD sources is one of the extensively used methods [26 - 32]. The acoustic emissions are captured by one or more ultrasonic transducers, which are sensitive to the frequency range of the emanated acoustic waves from PD sources. Considering the propagation characteristics of the acoustic waves in insulation media and different possible apparatus structures [33] [34], the frequency range 20 kHz to 500 kHz is generally chosen for PD measurement [15].
- **Ultraviolet Radiations:** This method of PD measurement is basically an optical method [35 - 38]. The frequency band of the optical signals radiated by a PD source, spans from ultraviolet to infrared region [35]. Specifically, UV radiations are generated by corona discharges, external surface PDs and arcing due to flashover. Optical

radiations are captured by optical sensors, sensitive to blue-UV region of light spectrum. Clearly the optical method provides isolation and is immune to unwanted noise if suitable light blocking filters are used [15]. However, this method is only applicable for transparent solid dielectrics, gaseous dielectrics and liquid dielectrics to some extent. The opacity of the media should not diminish the intensity of PD signals significantly during transmission. Also the PD generated optical signals must have lines-of -sight from the sensors.

- **Dissolved Gas Analysis (DGA):** PD measurement by DGA monitoring is a common tool for condition monitoring of oil filled equipments such as transformers [39 - 43]. Due to chemical effect of PDs, different types of gases are generated in the oil inside the equipment which remain dissolved in the oil. For DGA analysis, first oil is sampled from inside the equipment and dissolved gases are separated from oil by employing extraction techniques. Then the extracted gases are identified by gas chromatography. Lastly, the amount of PD present inside the equipment is ascertained by observing the levels of gases in the oil. However, this method cannot pinpoint exact location of PD inside the equipment, which is its main disadvantage [44].
- **Power Loss:** Any PD phenomena is associated with power loss which is proportional to the applied voltage. Due to presence of PD power factor changes with respect to applied voltage. PD can be detected and measured by observing the change in the power factor.

Each of the methods described above has its own advantages and disadvantages. Modern trend is to employ combination of these methods for more efficient PD detection [45 - 50]. Recently, different Artificial Intelligence and Machine learning based PD measurement techniques are also finding prominence [50 -54].

1.4 OBJECTIVES OF THE THESIS

The main objective of the present work is to identify and locate single and multiple Partial Discharge sources based on both optical and acoustic signals. This work has been taken up to investigate whether presence of PD phenomena inside emulated metallic enclosure of an equipment (such as a transformer) can be detected from outside by analysing the optical or acoustic signals captured by suitable sensors located at the inside and outside wall respectively.

For that purpose, an optical PD source and an acoustic PD source have been fabricated. In addition to these, a steel-made cubical box has been constructed which emulates the equipment enclosure. For capture of either type of signals from these sources, optical and

acoustic sensors have been mounted on the walls of this box. Each type of PD sources have been separately placed inside the box at different strategic locations and data corresponding to each type of PD signal for all the locations have been recorded. For emulation of single PD phenomena, only one type of PD source was placed inside the box. Whereas, to emulate multiple PDs, two such PD sources have been placed at different locations inside the box. Recording of each type of PD signals for all the locations inside the cubical box has been termed as PD data acquisition. The placement of PD sources at a given location inside the box has been termed as PD events. The recorded optical signals and acoustic signals have been analysed separately.

For optical PD data acquisition, wavelength range corresponding to visible light (300 nm - 500 nm) has been chosen. As the optical signals must have a line of sight, the optical sensors have been placed at the inside walls of the cubical box. In the present work, positions of the optical PD source inside the cubical box has been identified by employing feature classification techniques. It has been observed that the recorded optical signals are mostly sparse in nature. Mathematical Morphology, a time domain based signals transformation tool has been employed to transform and fill up the sparse domains of the captured optical PD signals. After that, suitable statistical features have been extracted from the morphologically transformed optical PD signals. These extracted features have been classified with two separate classification techniques – Rough Set Theory and Sparse Representation Classification. Lastly, the performances of these classification techniques have been compared.

For acoustic PD data acquisition, the frequency range of 20 kHz – 500 kHz has been considered. Moreover the acoustic sensors have been placed at the outside walls of the cubical box. The locations of acoustic PD source inside the cubical box have been identified by adopting a completely different approach. In the present work, the locations have been identified by two parameters, one is sequence of arrival of PD generated acoustic waves at different acoustic sensors mounted on outside walls of the cubical box and another one is levels of peak amplitudes of the captured acoustic signals by those sensors.

1.5 OUTLINE OF THE THESIS

Chapter 1: This chapter provides introduction to the Partial Discharge phenomenon as applicable in high voltage equipments and different detection and measurement methods of Partial Discharge.

Chapter 2: This chapter describes the construction of optical and acoustic PD sources, scheme of placement of sensors and experimental setup for PD data acquisition through optical and acoustic sensors in detail.

Chapter 3: This chapter explains theoretical formulation and working principle of time domain based signal transformation technique, Mathematical Morphology. The feature extraction from morphologically transformed optical PD signals has also been discussed in this chapter.

Chapter 4: This chapter presents theoretical background and working principles of the classification techniques – Rough Set Theory and Sparse Representation Classification.

Chapter 5: This chapter provides the theoretical background behind the identification of acoustic PD sources based on sequence of arrival and levels of peak amplitude.

Chapter 6: This chapter presents the results pertaining to the identification and localisation of both types of single and multiple PD sources using proposed methods.

Chapter 7: The concluding notes and future scopes of the present work have been presented in this chapter.

CHAPTER 2

EXPERIMENTAL SETUP

CHAPTER 2.

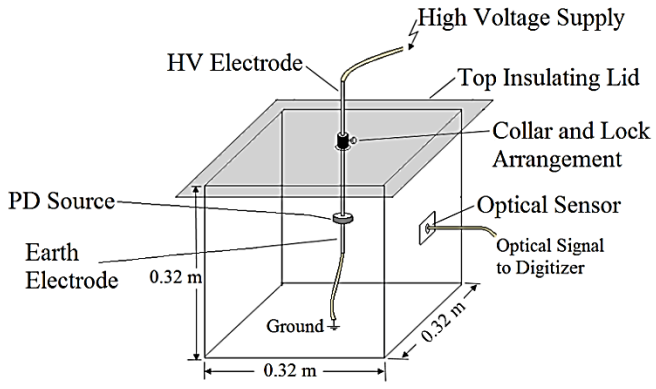
EXPERIMENTAL SETUP & PD DATA ACQUISITION

2.1 INTRODUCTION

Data acquisition through real life experiments is one of the crucial parts of any research work. As already discussed, the aim of the present work is to identify and localize partial discharge sources based on optical and acoustic signals emanated from such sources. It has been also mentioned that this work has been taken up to investigate the PD phenomena inside an emulated metallic enclosure of an equipment and to ascertain whether the presence of such phenomena can be detected from outside by recording and analysing the optical or acoustic signals captured by suitable sensors located at the outside and inside wall respectively. The experimental setup used in the present work emulates such occurrences of PDs in small scale. In the present work, optical signals and acoustic signals that are emitted from the PD sources have been analysed separately. For these two methods, the experimental procedure and apparatus used are same. Different approaches have been taken for Optical PD and Acoustic PD data acquisition techniques. Various aspects of the experimentation have been discussed in the following sections.

2.2 PARTIAL DISCHARGE SOURCE SIMULATOR (PDSS) BOX

For PD data acquisition, a cubical steel box with insulated top lid has been constructed. This box has side length of 0.32 m or 32 cm. The details of this box has been schematically shown Figure 2.1(a). The PD source is placed inside this box and made to discharge continuously for a short time so that the emitted optical or acoustic signals can



(a)

(b)

Figure 2.2 (a) Detailed view of Partial Discharge Source Simulator (PDSS) box (b) Actual photograph of PDSS box

be captured through sensors. The insulated top lid of the box holds the solid conductor of HV supply for the PD source. A fixing collar and a lock arrangement hold this conductor tightly. This box has been labelled as Partial Discharge Source Simulator box or PDSS box from now on.

As two types of PD sources i.e. optical PD and acoustic PD sources have been considered, each of the sources requires different sensors for capturing the respective signals. Depending upon the type of PD signal, optical or acoustic sensors will be mounted on the walls of the PDSS box. To record the optical signals coming out of PD source, the PDSS box has been kept air-filled. Whereas in case of capturing the acoustic signals, the box has been filled with transformer oil.

2.3 PARTIAL DISCHARGE (PD) SOURCE

Generation of partial discharge from breakdown of a void under high electric field stress is a well-known phenomenon. Here this phenomenon has been utilized to create the required PD sources. For two types of PD scheme, two types of PD sources have been constructed i.e. Optical PD source and Acoustic PD source.

2.3.1 OPTICAL PARTIAL DISCHARGE SOURCE

Optically transparent materials are obvious choices for crafting the optical PD source. In the laboratory, an artificial cylindrical void with known dimensions has been created inside a transparent acrylic disc having 10 mm diameter and 3 mm thickness. The cylindrical void has 0.8 mm diameter and a length of 1.5 mm. Figure 2.2 schematically depicts PD source made up of transparent acrylic discs with cylindrical void inside. The breakdown of this void actually generates the optical PD signals. The inception voltage for the void, having dimensions shown in Figure 2.2, is calculated to be 7 kV. The void

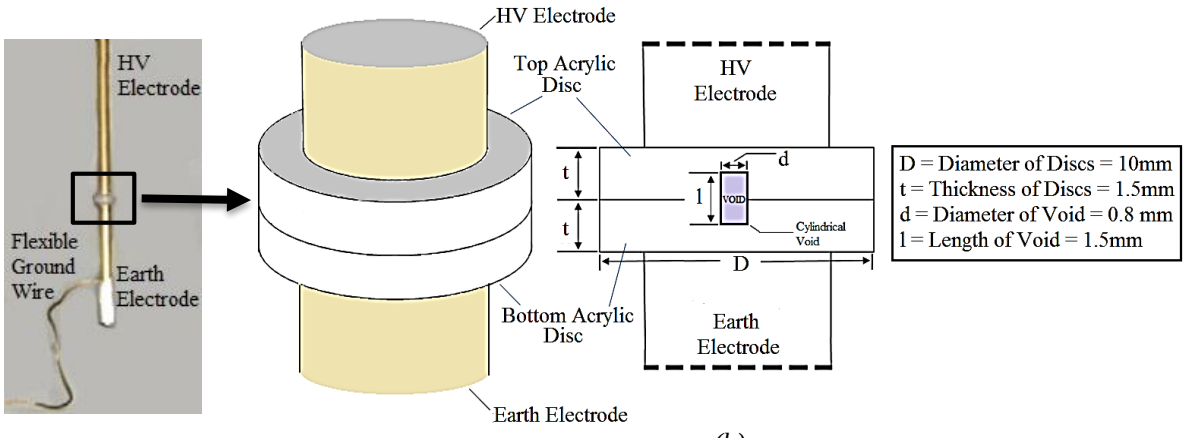


Figure 3.2 (a) Optical Partial Discharge Source with plane-plane electrode system (b) Detailed schematic view showing the dimensions of transparent acrylic discs with cylindrical void inside

should undergo continuous breakdown at least throughout the time required for data acquisition. For that reason, the high voltage applied to the top electrode of Figure 2.2 has been kept higher than 7 kV during the data acquisition. Optical radiations arising out of PD travel through transparent acrylic disc and air to reach the optical sensors mounted on walls of PDSS box. The acrylic discs are sandwiched between two brass electrodes. One of the electrodes is connected to HV supply and other one is kept at earth potential by connecting a flexible copper wire as earth conductor. For optical PD source, both these electrodes have plane cross-section. Hence this PD source has plane-plane electrode system and due to this, the voltage gradient between top and bottom planes of the void is uniform. When the void undergoes breakdown, the resulting short burst of charge flow is volumetrically uniform. The frequency of PD generated optical radiations lies in the upper frequency range of visible light spectrum (i.e. in the blue region) and extends into the UV region. Optical signals, being electromagnetic radiation, will undergo relative change in speed if any other medium was used inside the PDSS box instead of air. Although relative phase displacement between signals captured by different sensors remain same, there will be relative decrease in magnitude due to absorption in the media. For this reason PDSS box has been kept air filled during optical PD data acquisition.

2.3.2 ACOUSTIC PARTIAL DISCHARGE SOURCE

The acoustic PD source has different assembly than its optical counterpart. Instead of using plane-plane electrode system, here one of the plane-faced electrode has been replaced with a point-faced electrode. So the electrode system becomes the point-plane electrode system for acoustic PD source. Figure 2.3 schematically shows the acoustic PD source with point-plane electrode system. An acrylic disc having thickness of 3 mm and

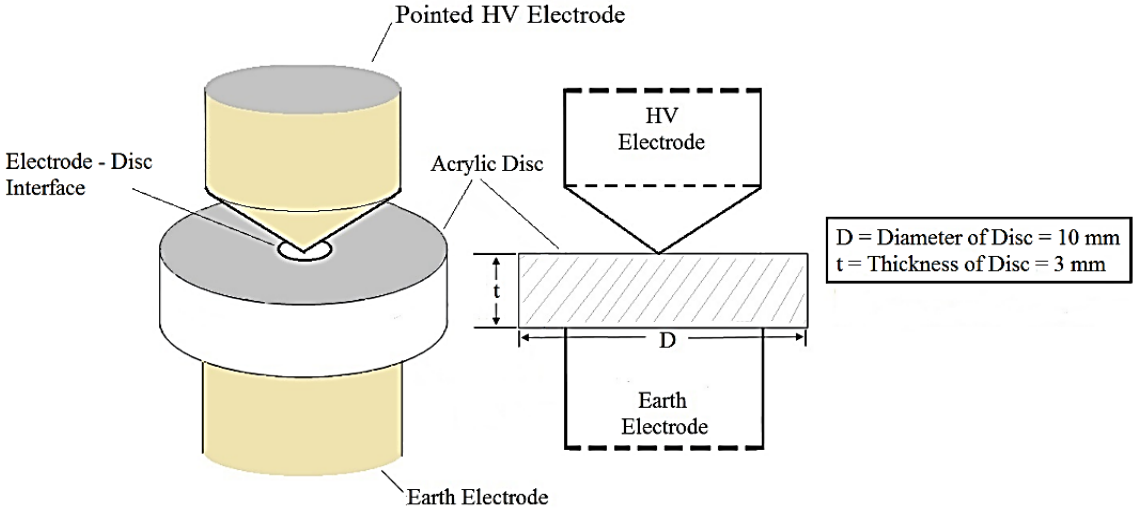


Figure 2.4 Acoustic Partial Discharge Source made up of transparent acrylic discs with cylindrical void inside and point-plane electrode system

diameter of 10 mm has been placed between above mentioned point-plane electrodes. Clearly it can be seen from Figure 2.3 that, due to point electrode, the electric field stress at the interface of point electrode and the acrylic disc wall will be very much high. When the voltage applied to HV electrode is sufficiently large, very high electric stress will cause partial breakdown of oil at the vicinity of that interface. The resulting blast and subsequent energy release at that interface creates the mechanical disturbance that travels through the transformer oil as acoustic waves. So in this case, the PD generated energy is converted into a mechanical wave. These PD generated acoustic waves are captured by the acoustic sensors mounted on walls of the PDSS box. The generated acoustic waves may get significantly attenuated due to energy dissipation, as they reach the wall of the box. In order to have higher amplitudes of the captured acoustic signals, such that these signals are clearly differentiable from the prevailing background noise, the initial energy release from the blast should be high. This initial energy release depends on the amount of non-linear electric stress the oil is subjected to, in vicinity of the interface, which in turn is dependent upon the amount of voltage applied to the HV electrode. Hence to capture acoustic signals of much higher amplitudes, the voltage applied to the HV electrode should be of much greater value than that is applied in case of optical PD source. The frequency of PD generated acoustic waves lie in the range of ultrasonic spectrum (20 kHz – 500 kHz).

2.4 PLACEMENT OF PD SOURCE INSIDE PDSS BOX

Placement of either type of PD source inside the PDSS box is an important task. The PD source is placed inside the box at specific positions which have been strategically chosen. Also a well-defined notation system has been developed to identify them. For acoustic and optical PD, these notations are different. The PD source is placed inside the PDSS box by inserting the HV conductor through a hole on the insulated top lid. The flexible earth conductor has length sufficient to facilitate free movement of PD source inside the box.

2.4.1 FOR OPTICAL PD DATA ACQUISITION

Inside the box, several locations have been chosen for placing the optical PD source. These locations are symmetrically distributed with respect to all the walls of the box. It has been observed that at these locations, optimum signals are captured by the optical sensors. The coordinates of these locations can be treated as imaginary grid points of a virtual cubic region having 16 cm side and 8 cm apart from all the walls of PDSS box.

This has been depicted in Figure 2.4. Three parallel cross-sectional surfaces of that virtual cube, that are designated as A, B and C, have been considered. Upon these surfaces grid positions are uniformly distributed and are consecutively 8 cm apart from each other. The grid points are numerically marked, starting from the middle point of the surface A, which has been designated as '1A'. Then corner points of the surface A are marked progressively as '2A', '3A', '4A' and '5A'. Lastly midpoints of the four sides are marked as '6A', '7A', '8A' and '9A'. In this way there will be nine grid points belonging to surface A. Same procedure has been followed for marking the grid points on cross-sectional surfaces B and C. Clearly there will be in total 27 such grid positions in the virtual cubic region. All the marked grid points are shown in Figure 2.4.

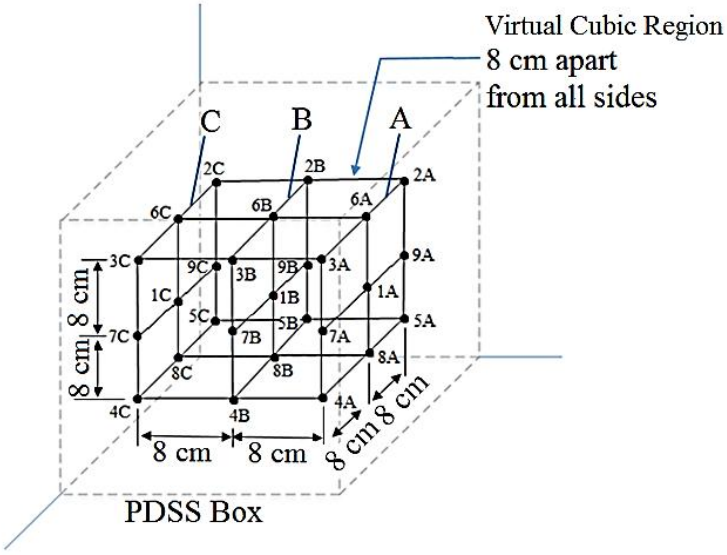


Figure 2.4 Imaginary grid points on three parallel square surfaces A, B and C in the virtual cubic region inside the PDSS box

2.4.2 FOR ACOUSTIC PD DATA ACQUISITION

For marking the positions of acoustic PD source, a different approach has been taken. Figure 2.5 given below shows the marked locations for placement of acoustic PD source.

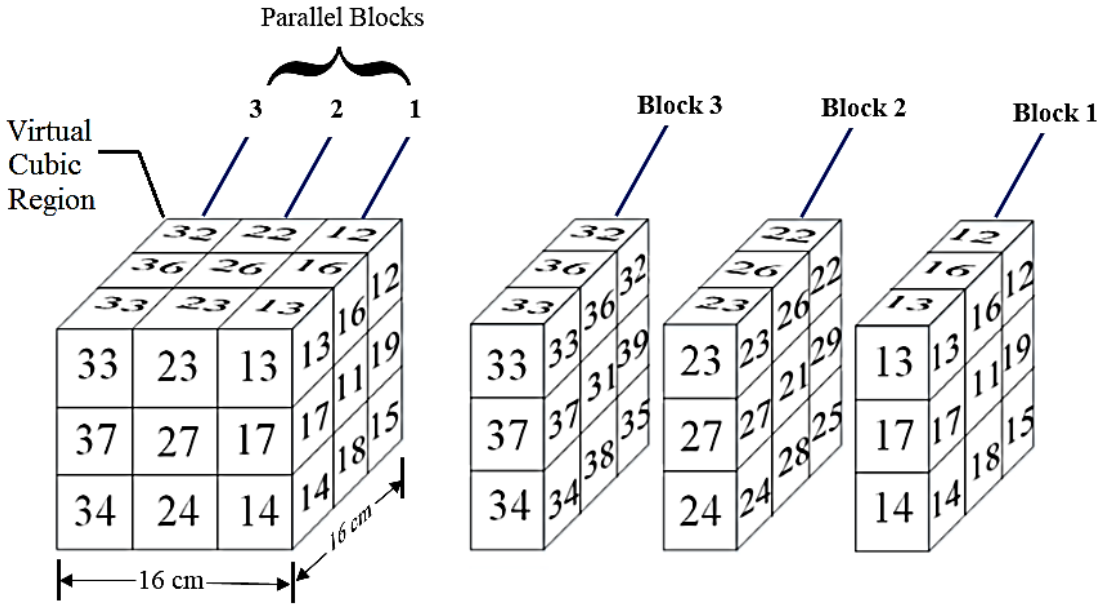


Figure 2.5 Imaginary cubic sub-regions in three parallel square blocks 1, 2 and 3 for placing the acoustic PD source inside the PDSS box

Here the same virtual cubic region of Figure 2.4 has been considered. But instead of placing the acoustic PD source at specific points, several different sub-regions have been chosen for placing the source. The virtual cubic region has been subdivided into 27 small cubic sub-regions having equal volume and uniformly distributed into three parallel blocks, as shown in Figure 2.5.

The small cubic regions have been marked as follows – three parallel blocks have been numerically marked as 1, 2 and 3. Now the cubic sub-region at the middle of block 1 has been marked '11'. Next the corner sub-regions are progressively marked as '12', '13', '14' and '15'. Lastly the regions at the middles of the four sides have been progressively marked as '16', '17', '18' and '19'. Sub-regions in the blocks 2 and 3 have been marked in similar fashion. All the marked cubic sub-regions have been shown in Figure 2.5. In this case also the total number of sub-regions is 27. The acoustic PD source has been placed at these designated regions for data acquisition.

2.4.3 PD EVENTS & NOTATION SCHEME

The PD source is placed at one of the imaginary grid positions (for optical PD data acquisition) or sub-regions (for acoustic PD data acquisition) by maneuvering the HV electrode fixed with it. Once desired vertical and horizontal position is obtained, the HV electrode is fixed with the top insulating lid tightly by collar and lock arrangement. Location of any point or region inside the box can be found out by setting the origin of coordinate system to any one of the base corner points of the PDSS box. The points of Figure 2.4 and the sub-regions of Figure 2.5 have been marked with respect to same coordinate system reference. As in both the cases, the locations or sub-regions are on or inside the same virtual cubic region, the optical PD locations are situated on one of the outer surfaces of the corresponding acoustic PD sub-regions. For example, location 1A of Figure 2.4 is situated on the left-side surface of sub-region 11 of Figure 2.5. Hence by this marking scheme, a parity is maintained between the two types of source placement.

If a single PD source is placed inside PDSS box and all the sensors capture signal from this discharge, then this event is termed as a Single Partial Discharge (SPD) event. On the other hand, when two such PD sources are placed at two locations of the virtual cubic region and PD is made to occur simultaneously in both the locations such that all sensors capture signal from both the discharges, then this event is termed as Double Partial Discharge (DPD) event. In the present work, 27 SPD events have been considered for both the optical PD and acoustic PD. In addition to that, 20 DPD events have been considered for both the cases. For further manipulation of the recorded signals, every PD

event has been assigned with suitable notation. Table 2.1 shown below summarizes the PD events and the corresponding notation scheme.

Table 2.1. SPD and DPD events and their corresponding notations

Type of PD Event	PD locations or sub-regions and corresponding notations
Optical SPD Events	1A, 2A, 3A, 4A, 5A, 6A, 7A, 8A, 9A, 1B, 2B, 3B, 4B, 5B, 6B, 7B, 8B, 9B, 1C, 2C, 3C, 4C, 5C, 6C, 7C, 8C, 9C
Optical DPD Events	1A_1C, 1A_2C, 1A_7C, 1C_2A, 1C_7A, 2A_3A, 2A_4A, 3A_5A, 5A_4A, 9A_7A, 2B_3B, 2B_4B, 3B_5B, 5B_4B, 9B_7B, 2C_3C, 2C_4C, 3C_5C, 5C_4C, 9C_7C
Acoustic SPD Events	11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 31, 32, 33, 34, 35, 36, 37, 38, 39
Acoustic DPD Events	11_31, 11_32, 11_37, 31_12, 31_17, 12_13, 12_14, 13_15, 15_14, 19_17, 22_23, 22_24, 23_25, 25_24, 29_27, 32_33, 32_34, 33-35, 35_34, 39_37

In Table 2.1, the notation for SPD events is same as the position of PD source at the corresponding grid point of Figure 2.4 or sub-region of Figure 2.5. For example, if optical PD source is placed at location ‘1A’ then the notation to denote set of five signals captured by five sensors will be ‘1A’. Similarly notation ‘11’ denotes set of signals captured when the acoustic PD source is placed inside sub-region ‘11’.Whereas for a DPD event, the notation to is of the form ‘1A_1C’ for optical PD events or ‘11_31’ for acoustic PD events. Notation ‘1A_1C’ implies that two optical PD sources have been placed at locations ‘1A’ and ‘1C’ and PD is taking place at both the locations simultaneously. The notation for DPD event ‘11_31’ can be interpreted in similar fashion. For 27 locations or sub-regions, a large number of DPD events are possible but only 20 of such events have been selected. Two locations or sub-regions, which are directly adjacent to each other or are situated along the same vertical line, have been avoided while selecting the DPD events mentioned in Table 2.1.

2.5 PLACEMENT OF SENSORS IN PDSS BOX

In order to detect the optical or acoustic signals from the PD source, suitable type of sensors have been placed on the walls of PDSS box. The placement of sensors should be such that they would be able to capture good quality optical or acoustic signals from all the PD locations successfully.

Let us now take the example of optical PD locations as shown in Figure 2.4. Assuming that the optical PD source has been placed at one of the locations on surface A, the square face of PDSS box right in front of surface A is a logical choice to place an optical sensor. This is because that particular square face has the least distance form all the locations of surface A. Let this optical sensor be marked as OS1. Clearly, optical signals captured by OS1, when the optical PD source has been placed at one of the locations on surface A, would have high amplitudes and intensities. Now when the optical PD source has been placed at distant locations, such as locations on surface C, the amplitude and intensity of the captured signals by OS1 will become very low due to attenuation, owing to greater relative distance of these locations from OS1. Hence in order to capture good quality optical signals from the locations on surface C, another optical sensor should be placed at the square face which is right in front of that surface. Extending this reasoning for all the locations, it can be inferred that for six faces of the virtual cubic region of Figure 2.4, there should be an optical sensor placed in front of every face i.e. optical sensor should be mounted on every square face of PDSS box. But as the top lid of the box holds the HV conductor, placing a sensor on top lid will be unsuitable. Hence to avoid clumsiness, no sensor can be placed on the top lid of the box. The placement of five optical sensors on five walls of the PDSS box has been schematically depicted in Figure 2.6. For simplicity five sensors have been placed right in the middle points of the respective square faces. As the PD generated optical radiations must have line of sights from the walls, the optical sensors have been mounted on the inside walls of the PDSS box for optical PD data acquisition

This contention can be extended for placing the acoustic sensors in similar manner and the scheme of sensor placement, shown in Figure 2.6, is also applicable for acoustic PD data acquisition. In case of PD generated acoustic signals, effect of attenuation can be much significant if sensors are placed much further from the PD source and the acoustic signals may become too noisy. Moreover, sensors which are at different distances from the PD locations will experience different signal arrival times.

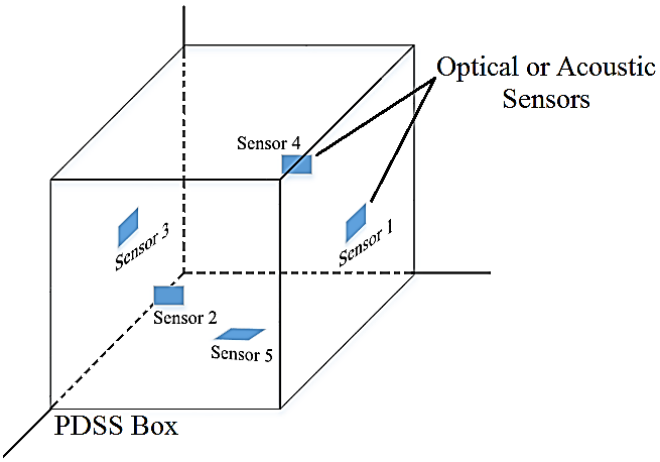


Figure 2.6 Placement of Optical or Acoustic Sensors in PDSS Box

Unlike the placement of optical sensors, the acoustic sensors have been mounted at the outside walls of the PDSS box. The advantages of placing the acoustic sensors outside the walls include easy fixation, better flexibility and scope for reconfiguration. Also the sensors may malfunction and get damaged due to direct contact with transformer oil if placed inside the box. Hence placing five acoustic sensors at the middle points of five outside walls of PDSS box is an optimum arrangement for acoustic PD data acquisition.

If fewer sensors were used then some of the important information from the PD sources might have been lost. Although there is no generalized strategy to choose the optimum number and position of sensors, depending on relevant situation suitable schemes are to be implemented. Optical and acoustic sensors have been marked with suitable notations for identification. Placement schemes, sensitive regions and notations for optical and acoustic sensors have been shown in Table 2.2 given below.

Table 2.2 Placement schemes, sensitive regions and notations for optical and acoustic sensors

Type of the sensor	Placement in the PDSS box	Sensitive to frequency spectrum	Notation
Optical Sensor	At inside walls	Blue-UV spectrum	OS1, OS2, OS3, OS4, OS5
Acoustic Sensor	At outside walls	Ultrasonic spectrum	AS1, AS2, AS3, AS4, AS5

Various salient properties and specifications of the optical and acoustic sensor used in the present work can be found in [56] [57].

2.6 CIRCUIT ARRANGEMENT FOR PD DATA ACQUISITION

Similar circuit arrangements have been used for optical and acoustic PD data acquisition. The schematic of the experimental setup used in the laboratory for optical PD data acquisition has been shown in Figure 2.7. The optical PD source comprises of plane-plane electrode system. The HV electrode of the PD source is connected to a 10 kVA,

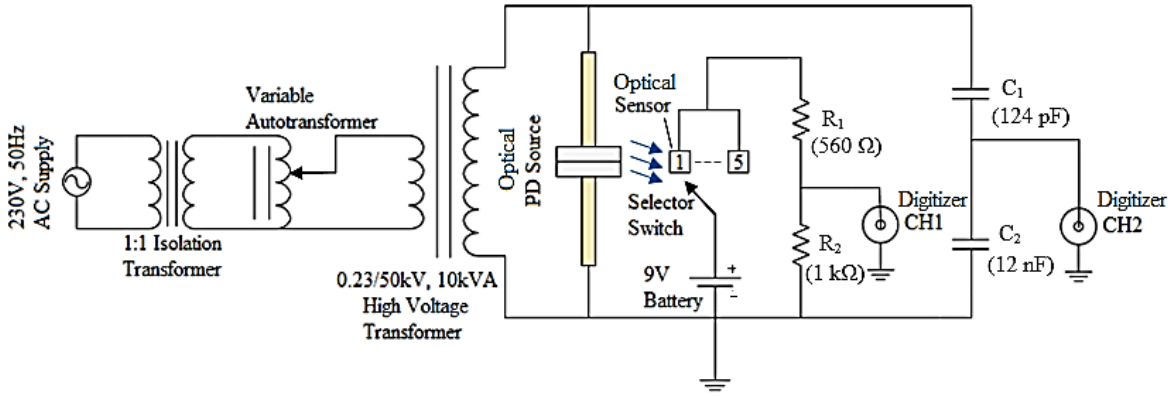


Figure 2.7 Schematic circuit diagram of experimental procedure for optical PD data acquisition

230V/50kV high voltage testing transformer. An assembly of 1:1 isolation transformer

and an autotransformer, as shown in Figure 2.7, is connected to the input of the testing transformer. As mentioned before, the inception voltage of discharge for the void shown in Figure 2.2 is 7 kV. In order to obtain sustained partial discharge, the test voltage has been kept slightly higher than the inception voltage i.e. at 9kV for optical-PD data acquisition.

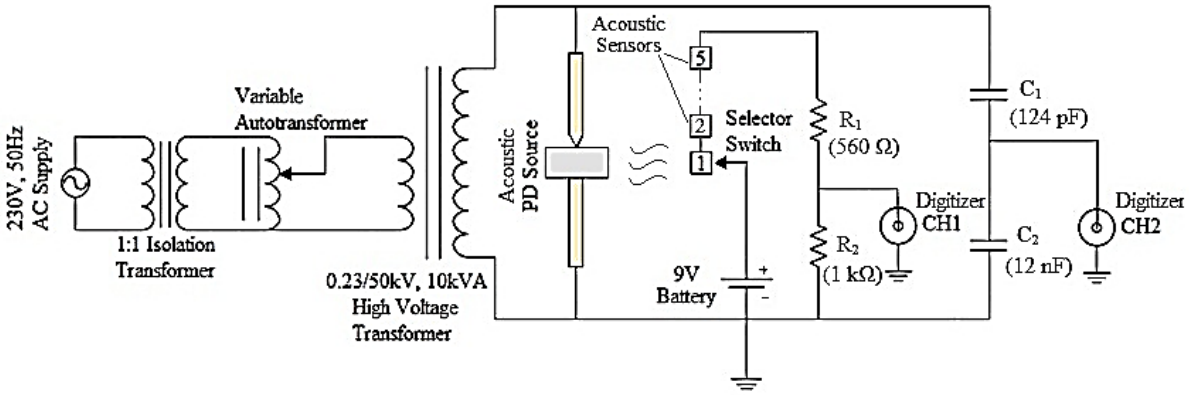


Figure 2.9 Schematic circuit diagram of experimental procedure for acoustic PD data acquisition

The schematic of the experimental setup for acoustic PD data acquisition has been shown in Figure 2.9. The acoustic PD source comprises of an acrylic disc sandwiched in between pointed electrode and plane electrode. The test voltage has been kept at a much higher value of 30kV such that sustained explosion in oil can occur at the site of PD. This higher test voltage yields acoustic signals having higher amplitudes.

Figure 2.10 High Voltage Testing Transformer

The five sensors, placed at walls of the PDSS box, capture the optical or acoustic signals emitted out by the PD sources. The captured signals are digitized by a digitizer which facilitates data acquisition in a computer. As shown in the schematics of Figure 2.8 and 2.9, there is a selector switch to select a particular sensor. When a particular sensor is selected, a current, proportional to amount of signal energy collected by that sensor, flows through the R_1, R_2 resistance bridge. A two channel digitizer from National Instrument™ with sampling rate of 100MS/s has been used in the present work to digitize the analog signals captured by the sensors [58]. The voltage drop across R_2 which is proportional to the current, has been fed to the CH1 (Channel 1) of the digitizer. A capacitor divider reduces the high voltage input to a much smaller value of 5-8 V_p. This capacitor bridge provides the phase

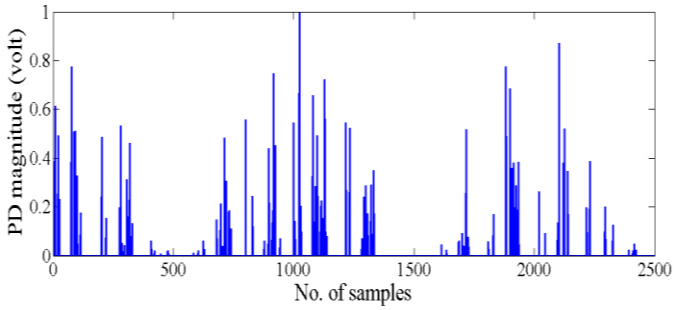
reference of the digitizer through CH2 (Channel 2). The time duration of recorded analog signal has been measured from the zero crossing of positive half cycle of high voltage supply applied to the sample.

- For optical signals, the time duration of each data capture is 0.025s or 25ms.
- For acoustic signals, the time duration of each data capture is 500µs.

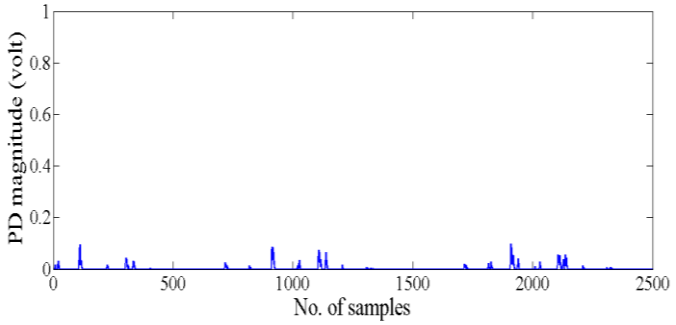
The digitizer is connected to computer through USB. For the data acquisition, a software has been developed in LabVIEW© environment. In order to ensure a fixed starting point with respect to the supply frequency cycles, the data capture through sensors start whenever the power frequency reference potential in CH2 of digitizer becomes greater than a preset threshold value, V_{ref}. For an SPD event, first the PD source is placed at a particular location of Figure 2.4 or Figure 2.5 and five sensors capture optical or acoustic signals. This is a single data set of five data from five sensors. This process is repeated five times with an interval of 1 hrs between two consecutive data capture. By this process total of five data sets, each containing five data corresponding to five signals captured by five sensors, is obtained for a single PD source i.e. a single SPD event. The same procedure is applied to all the DPD events. A complete database of all the data pertaining to signals captured by all the sensors for every optical and acoustic PD event has been created.

One set of optical PD signals, recorded by five optical sensors, corresponding to SPD event '1A' and DPD event '1A_1C' have been respectively shown in Figure 2.10 and Figure 2.11. In Figure 2.10, note that the sensor S1 records maximum intensity because it is closest to PD location 1A, whereas intensities of signals captured by other sensors are lower. Therefore for any optical SPD event, optical sensors which are close to location of optical source record optical signals of higher intensity. Also note from Figure 2.10 and Figure 2.11 that the optical signals are mostly sparse in nature i.e. in general, the domains of optical signals are sparse. This phenomena happens due to sporadic charge flow through the partially discharging cylindrical void, inside the optical PD source of Figure 2.2, during the time span of data capture.

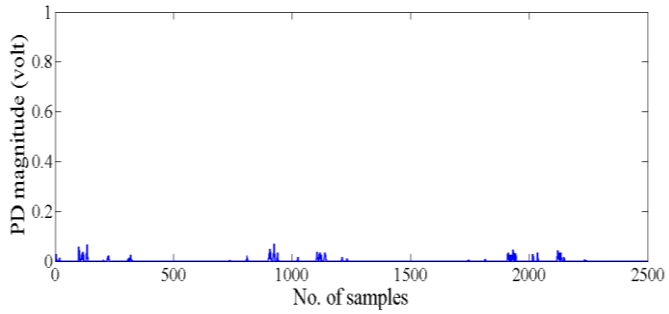
One set of acoustic PD signals, recorded by five acoustic sensors, corresponding to SPD event '31' and DPD event '12_13' have been respectively illustrated in Figure 2.12 and 2.13. It can be seen from these figures that, the acoustic signals have completely different nature compared to optical signals. Also unlike the optical signals, domains of acoustic signals are not sparse. Detailed discussion about the nature of acoustic PD signals have been provided in Chapter 5.



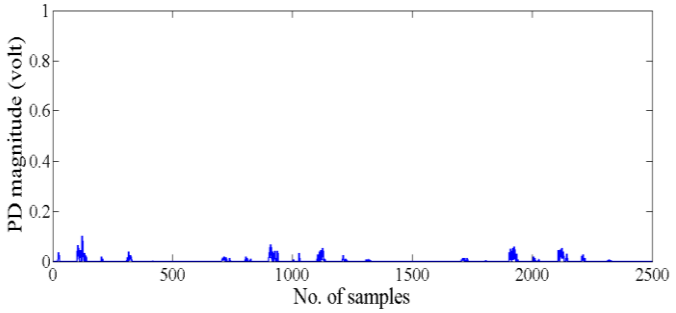
(a)



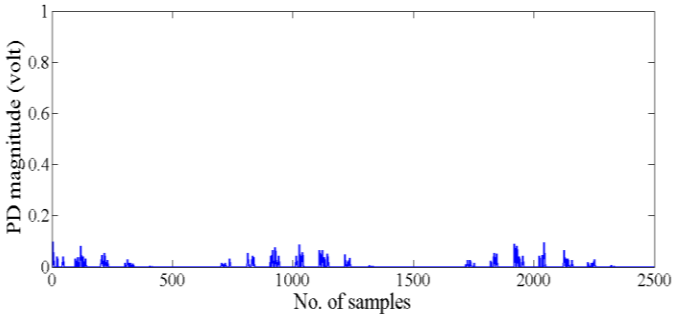
(b)



(c)

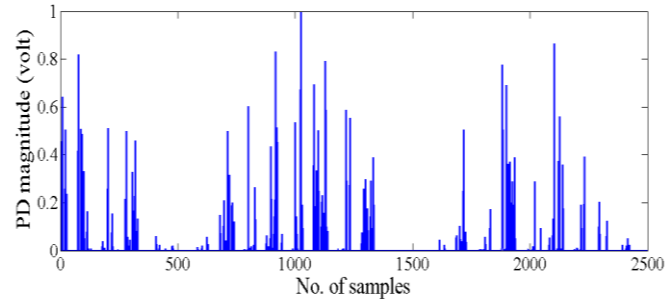


(d)

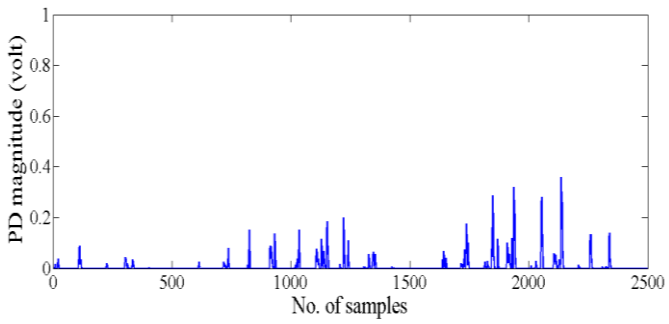


(e)

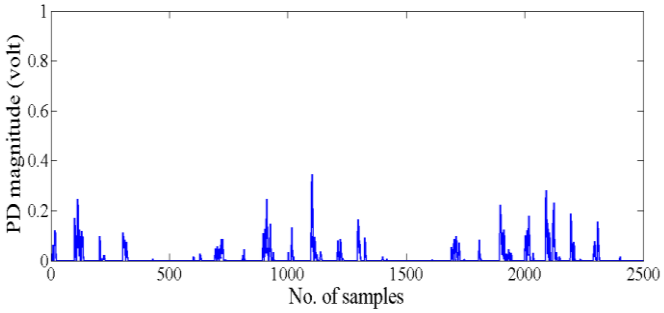
Figure 2.11 Optical signals captured by different sensors when optical PD source is placed at '1A' a) Sensor, OS1 (b) Sensor, OS2 (c) Sensor, OS3 (d) Sensor, OS4 (e) Sensor, OS5.



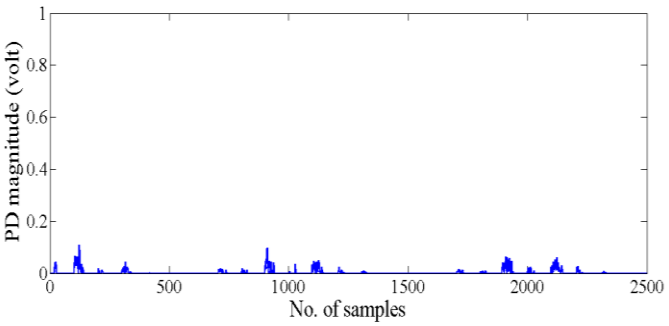
(a)



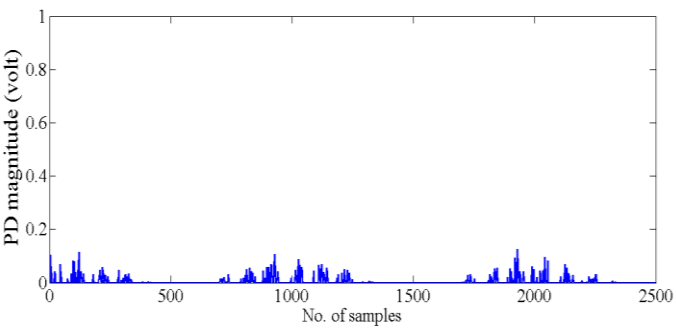
(b)



(c)



(d)



(e)

Figure 2.12 Optical signals captured by different sensors when two optical PD sources are placed at '1A' and '7C' a) Sensor, OS1 (b) Sensor, OS2 (c) Sensor, OS3 (d) Sensor, OS4 (e) Sensor, OS5

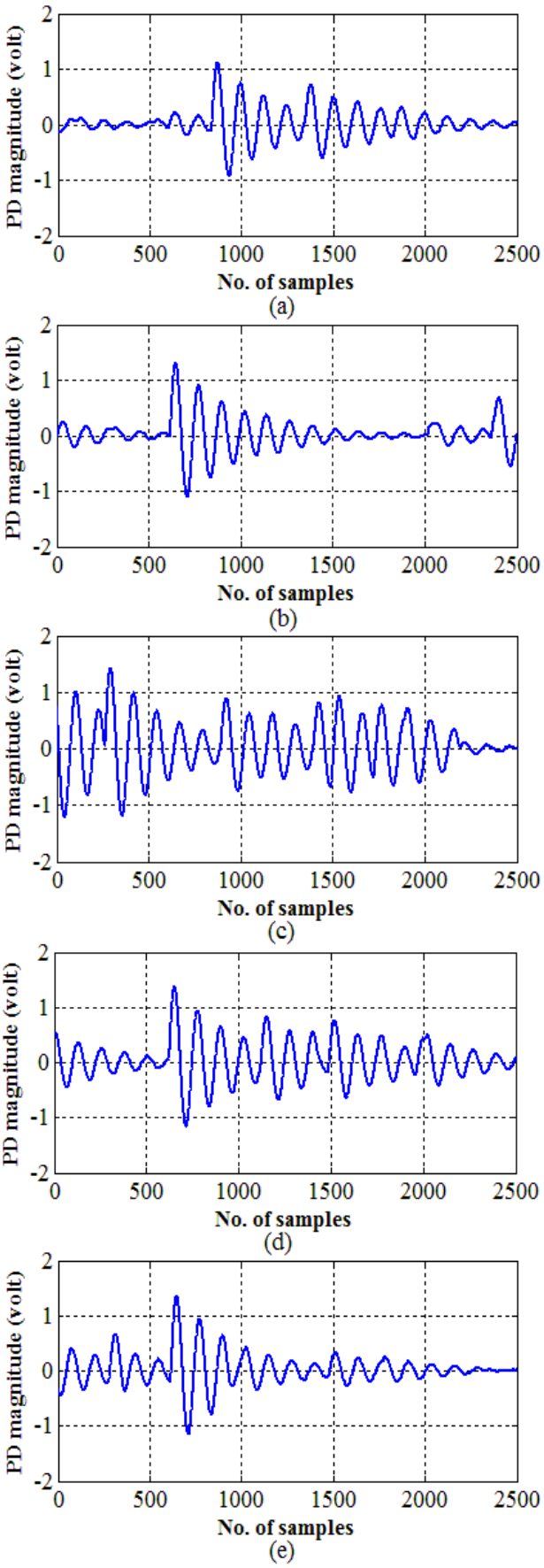


Figure 2.12 Acoustic signals captured by different sensors when acoustic PD source is placed at '31' a) Sensor, AS1 (b) Sensor, AS2 (c) Sensor, AS3 (d) Sensor, AS4 (e) Sensor, AS5.

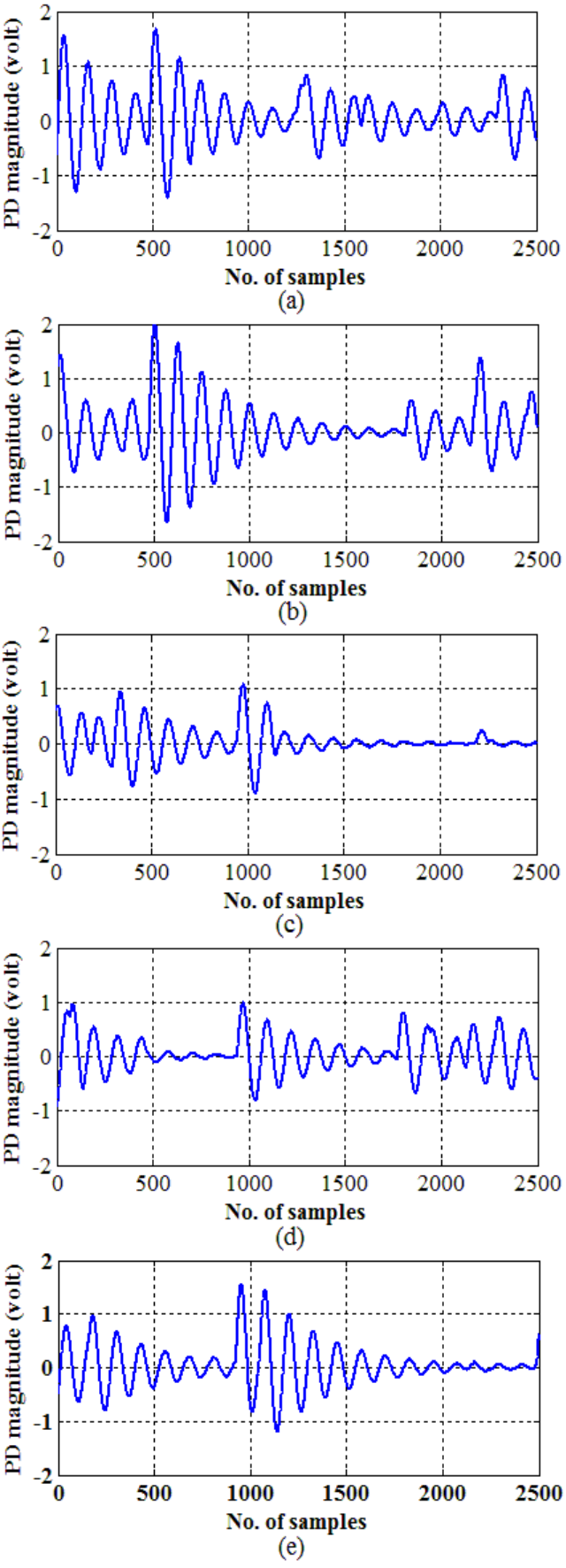


Figure 2.13 Acoustic signals captured by different sensors when two acoustic PD sources are placed at '12' and '13' a) Sensor, AS1 (b) Sensor, AS2 (c) Sensor, AS3 (d) Sensor, AS4 (e) Sensor, AS5.

CHAPTER 3

MATHEMATICAL MORPHOLOGY AIDED FEATURE EXTRACTION FROM OPTICAL PD SIGNALS

CHAPTER 3.

MATHEMATICAL MORPHOLOGY AIDED FEATURE EXTRACTION FROM OPTICAL PD SIGNALS

3.1 INTRODUCTION

In the present work, locations of optical PD source have been identified by extracting suitable features form mathematically transformed optical PD signals and subsequently classifying the extracted features with the help of classification techniques. First, the PD generated optical signals have been transformed by employing Mathematical Morphology, a time domain-based signal transformation tool. Then, different statistical features have been extracted from the transformed PD signals. The extracted features have been classified by two separate classification techniques - Rough Set Theory and Sparse Representation Classification. Mathematical Morphology and subsequent feature extraction from transformed optical PD signals have been discussed in this chapter.

3.1.1 FEATURE EXTRACTION

Artificial Intelligence (AI) based techniques are very popular for data analysis and interpretation due to their robust theoretical background and wide range of applicability. One of the fundamental aims of AI is to achieve rationalization of the machines i.e. to help machines make decisions like rational beings and improve performance through experience. The techniques followed for implementing this rationalization are collectively called the Machine Learning techniques. The concept of Machine Learning evolved from other branches of AI such as pattern recognition and computational learning theory. Machine learning deals with the construction of algorithms that can formulate decisions by analysing a set of input data or can learn to make predictions about that data.

One of the important tasks performed by the Machine learning algorithms is called the dimensionality reduction which is defined as the process of mapping a large set of input data into a lower dimensional space. One of the major problems that stems while processing a large set of data is the large number of associated variables to be handled. Also there could be unknown redundancy present in the data.

Processing that large amount of data and associated variables may require larger memory for storage, more time for processing and unnecessary large computational facility. If that data is given input to a classification algorithm then the algorithm may perform poorly due to overfitting and may not be able to classify new samples. Hence, one of the main objectives while processing a large data will be to analyse out information from the data such that the complete set of associated variables can be represented by a minimal set of variables which is as informative as possible and the desired task can be performed with that reduced representation instead of using the whole data.

For that purpose, some measureable and quantifiable properties are analysed out from the data which are called the explanatory variables or features. Features may be categorical, ordinal, integer valued or real valued. The process of computing and obtaining features is called the feature extraction. Typically, several features are extracted from a set of data to form the feature vector. Feature extraction reduces the variables required to represent a large set of data which facilitates easier and faster classification of that data. Hence, feature extraction prepocesses the input data to make it more explanatory and manipulable. If input data pertaining to a set of signals are to be processed for the purpose of classification, then it will be advantageous to reduce the dimension of the signal space by feature extraction. But, extraction of features from a signal by mere visual inspection will be very difficult. Due to this reason, different specialized feature extraction techniques are employed for extraction of the features.

3.2 MATHEMATICAL MORPHOLOGY

Morphology is defined as the study of shapes. Mathematical Morphology (MM) is a time-domain based non-linear signal transformation technique that deals with the mathematical theory of characterizing and manipulating signals with sets. The theory of Mathematical Morphology was collaboratively developed by Jean Serra and Georges Matheron in 1964 [58 - 61]. Since then it has seen tremendous theoretical development in next four decades.

A signal is a function which describes any quantity that exhibits variation in time or space and provides information about the status of a physical phenomenon or entity. Signal processing is a general technique by which signals are manipulated and transformed in such a way that important information can be extracted from the processed signals [64]. In that sense, Mathematical Morphology is a specialized signal processing technique. MM deals with the shape, structure and form of the signal presented to it and locally modifies geometric features of that signal [62]. The operation of MM can be considered as imparting filtering effect upon the signal and such morphological

transformation operations are collectively termed as the Morphological filters [62] [63]. In general, morphological filters perform set-operations and transforms the graph of the signal to provide quantitative description of signal's geometric structure. Hence, MM is similar to other conventional linear signal processing techniques that employ filtering windows. But there is a fundamental difference between MM and linear transformations. The basic morphological operations are non-linear in nature and hence the algebra involved is much different form the linear algebra. Moreover it exclusively works in time domain, unlike other conventional frequency domain based transform methods [65].

MM is most commonly employed for digital image processing and has become one of the widely used sophisticated mathematical tool for such applications. The areas of application of MM in various image processing and applications (such as biomedical image processing) are wide and numerous. Some of the applications of MM in image processing are – shape recognition, shape smoothing, texture analysis, thinning, enhancement, edge detection, representation and coding etc. In addition to these, MM is a well-established and effective technique for denoising and nonlinear filtering [66 - 69].

3.2.1 REPRESENTATION OF SIGNALS BY SETS AND FUNCTIONS

The theory and technique of MM is based on various important fields of mathematics such as set theory, lattice theory, random functions and topology. Morphological operations are applicable for both continuous and discrete spaces. One of the basic assumptions in the context of MM is that signals can be completely characterized by sets [62]. In general, an n-dimensional signal can be mathematically represented as a function of n-independent variables. As a special case, if that function only takes two distinct values, then it can be shown that the signal can be represented as an n-dimensional set in Euclidean space.

Let us take the example of a bi-level signal, as shown in Figure 3.1. Bi-level signals are characterized by two levels of amplitude i.e. the signal continuously toggles between high-level amplitude and low-level amplitude. Let these amplitudes be denoted in binary form i.e. high-level amplitude is denoted as 1 and low-level amplitude is denoted as 0.

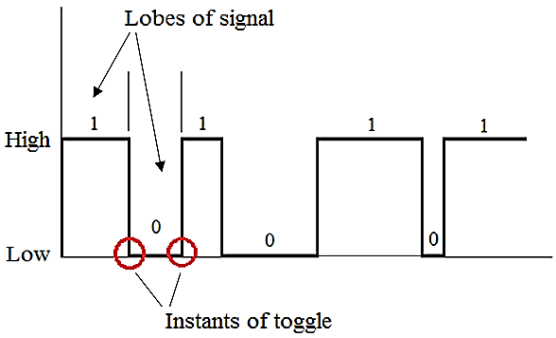


Figure 3.1 Bi-level signal comprising lobes between consecutive instants of toggle

As the amplitude of the signal can assume two distinct values, bi-level signals are also be termed as binary signals. At the instants of toggle, signal changes its amplitude from one level to another and vice-versa. The part of the signal between two consecutive instants of toggle has been termed as lobe. The signal can be considered to be made up of many such lobes. Alternate lobes of the signal have same amplitude levels.

Let parameter ℓ denote the lobes of the signal and a function $f(\ell)$ associated with the bi-level signal be defined as,

$$f(\ell) = amplitude\ \ of\ lobe\ \ell \tag{3.1}$$

Clearly, $f(\ell)$ can assume two possible values. As the amplitude levels have been denoted in binary form, $f(\ell)=1$ for the lobes which have high-level amplitudes and $f(\ell)=0$ for lobes which have low-level amplitudes. All the lobes which are at high level amplitude can be comprehensively described by the set,

$$F = \{\ \ell \mid f(\ell)=1\} \tag{3.2}$$

The lobes which are at low-level amplitude are represented by the complement of set F, as shown below.

$$F^c = \{\ \ell \mid f(\ell)=0\ \} \tag{3.3}$$

Hence, the set F alongwith its complement completely describes the bi-level signal. Function $f(\ell)$ assumes value 1 for all the elements which are members of set F and assumes value 0 for all the elements which are not in F. This type of function is called the characteristics function or indicator function, which maps the elements of a set into the discrete range of {0, 1}. Generalising the foregoing analysis it can be inferred that, bi-level signals are comprehensively represented by sets, whose elements are mapped by binary characteristic functions.

Now consider a more general case of a multilevel signal, shown in Figure 3.2. As depicted in that figure, the amplitude of the signal can assume a finite set of discrete

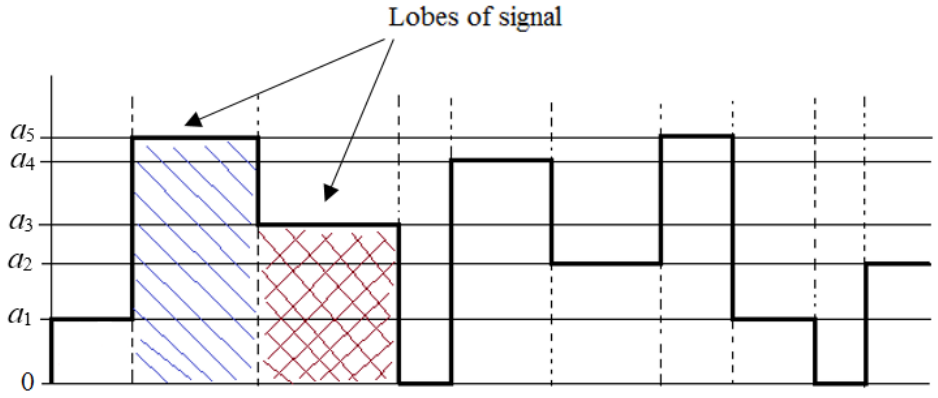


Figure 3.2 Multilevel signal

values i.e. levels of amplitude are discretely distributed. Similar to binary image, here also we can mark the lobes between two consecutive points, at which the amplitude of the signal changes from one value to another.

Let a function $\phi(\ell)$ associated with the multilevel signal be defined as,

$$\phi(\ell) = amplitude\ \ of\ lobe\ \ell \tag{3.4}$$

The lobes of the signal which have the same level of amplitude can be grouped together to form a level set. As levels of amplitudes for all the lobes are discretely distributed, $\phi(\ell)$ is a discrete valued function. If parameter a_j denotes the j^{th} level of amplitude, then the multilevel signal can be represented as a function of a_j. If there are m different levels of amplitude in the signal then,

$$\Phi(a_j) = \{\ \ell \mid \phi(\ell) \geq a_j\},\ j = 1,2,3,...,\ m \tag{3.5}$$

Here, a_j only assumes discrete values. Lobes which have level of amplitude not less than a_j forms the level set, $\Phi(a_j)$. Levels sets are related to each other by the relation, $\Phi(a_{j+1}) \subset \Phi(a_j)$. For j^{th} level of greyness intensity, complement of the level set $\Phi(a_j)$ is given by, $\Phi^{C}(a_j) = \{\ \ell \mid \phi(\ell) < a_j\},\ j = 1,2,3,...,\ m$.

Thus each specific amplitude level in the signal constitutes a level set. If all the discrete amplitude levels are taken into account then the whole signal can be completely represented as a function comprising stack of level sets. Hence contrary to bi-level signal, the multi-level are represented as functions. In general, multilevel signals are represented by functions, comprising stacks of level sets.

Thus, signals can be mathematically represented both as a set and a function. But the primitive representation will be set and functions are to be considered as particular cases of such sets. Sets will be obtained from functions by inducting the level cuts, which is a reverse process of stacking. Basic set operations can be translated to signals if the corresponding set follows certain conditions. A signal can be completely characterized by a set and that set shall be deemed suitable for morphological operations, if the set adheres to following conditions [62] [63] -

(i.) The elements of the set can be partially ordered and all the non-empty subsets can be properly identified alongwith their boundaries.

(ii.) Each non-empty subset of that set has a maximum and minimum.

Clearly every electrical signal, whether continuous or discrete, satisfies these two conditions and such signals can be morphologically processed. Similar set representations

can be also formulated for continuously varying signals, but such formulations will be very much complex in nature. In the context of MM, representation of signals by sets is essential because all morphological operations are applicable for sets only [62].

3.2.2 STRUCTURING ELEMENTS AND FUNCTIONS

The basic concept in MM is to probe a given signal with a simple and predefined element. That probe element interacts with the signal and locally modifies geometric characteristics of the signal. This probe is called the Structuring Element (SE) [58 - 65]. The information that can be extracted from a signal depends upon the way the signal is probed or observed [61]. Interaction of SE with the signal provides the basis for deductions on how the element touches or misses the signal. Depending on the information to be extracted, a given signal can be probed in many ways and accordingly the SE could be of different shapes and dimensions. Hence, shape and size of an SE are the two most important characteristics [65]. Also the signal to be probed can be a bi-level signal or a multilevel signal. Furthermore, multilevel signal can be of discrete or continuous type. Hence, different classes of structuring elements will be required to probe bi-level and multilevel signals

- The structuring element required for probing a bi-level signal will be a set, defined on Euclidean subspace and characterised by a binary indicator function.
- The structuring element intended for probing a multilevel signal is a function instead of a set and is called the Structuring Function. Structuring function will be discrete or continuous depending upon the type of multilevel signal to be processed.

Types of SE required corresponding to different types of signal to be processed have been summarized in Table 3.1.

Table 3.1 Different types of structuring element

Type of Signal		**Type of Structuring Element**
Bi-level or Binary signal		Set defined on Euclidean subspace and characterised by a binary indicator function.
Multilevel signal	Discrete multilevel signal	Discrete–valued structuring function
	Continuous multilevel signal	Continuous structuring function

Some of the commonly used shapes of SE are linear, inclined line, circular, semi-circular, square, disk, polygon, beeline etc. Different types of continuous and discrete structuring elements have been shown in Figure 3.3.

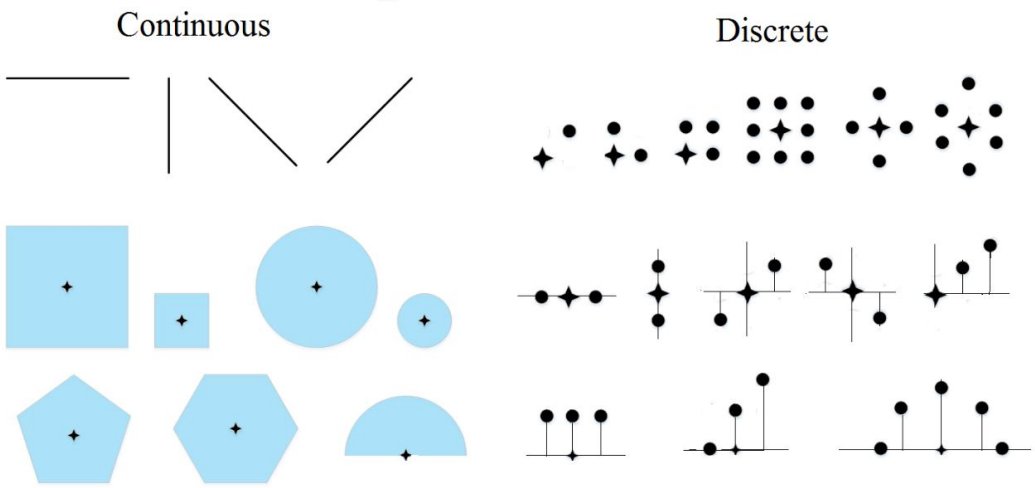


Figure 3.3 Different types of continuous and discrete structuring elements (The symbol ' ✦ ' denotes the reference points or the origins of respective structuring elements)

Structuring elements operate upon a chunk of data obtained from the signal and not on the whole signal. Thus the interaction of SE with the signal is highly localised [62]. To probe the whole signal, the SE needs to be translated through the body and along the boundary of the signal. The effect of SE is more pronounced on the boundary of the signal than on the body and is proportional to the size of SE. In any case rotation of SE is not allowed. It is important to note that SE can be placed anywhere on the signal. The placement of SE is dictated by the reference point or origin of the SE. Choice of the reference point is completely arbitrary. In Figure 3.3, the symbol ' ✦ ' denotes the reference points or the origins of respective structuring elements. It is normal practice to choose structuring elements which are symmetric with respect to reference point. If asymmetric structuring elements are chosen then they need to be transposed with respect to origin before operation.

The structuring element preserves details and reduces noises from the signal. For a given application, a specific structuring element performs optimally. Choice of shape and size of optimal SE are governed by several important factors such as dimensions of signals being transformed, type and frequency of interference present in the signal and sampling rate of digitization [58 - 60] [65]. Ultimately the selection of SE is based on past experience and trial-and-error method. Generally it is noticed that with the increase in size of the data to be analysed the size of SE is also to be increased, but not necessarily proportionately [65].

3.2.3 MORPHOLOGICAL OPERATIONS

Different mathematical operations can be performed in MM by probing a signal with a structuring element [58] [61]. The two most basic operations in MM are Dilation and Erosion of a signal.

Let, F denote a set corresponding to a binary signal, $f(x)$ denote a multilevel continuous signal and $f(n)$ denote a multilevel discrete signal defined in the discrete domain, $D_f = \{f_0, f_1, ..., f_n\}$. Binary signal F is probed with set B defined on Euclidean subspace. Multilevel continuous signal $f(x)$ is probed with continuous structuring function $b(x)$, having same dimensions as $f(x)$. Lastly, multilevel discrete signal $f(n)$ is probed with discrete structuring function $b(m)$ defined in the discrete domain $D_b = \{b_0, b_1, ..., b_m\}$. Here n and m are integers and $n > m$. Note that all the structuring elements have been defined following the conventions provided in Table 3.1. Also it is assumed that all the signals are suitable for morphological operations and all the structuring elements are symmetric to origin.

3.2.3.1 DILATION AND EROSION

In dilation operation, first SE is placed somewhere in the signal and if SE touches the signal then the dilated signal includes the SE alongwith its origin. Then the SE is translated to the whole signal including its boundary to yield the dilated signal. Depending on the type of input signal and corresponding SE, Dilation operation can have three forms. Equations (3.6), (3.7) and (3.8) depict these three forms of Dilation. In these equations, operator ' $\oplus$ ' is the Dilation operator.

(i.) Binary Dilation is closely related to Minkowski set addition. Dilation of F by structuring element B is given as,

$$F \oplus B = \{\varphi + \beta \mid \varphi \in F \wedge \beta \in B\} \tag{3.6}$$

(ii.) Dilation of continuous function $f(x)$ with structuring function $b(x)$ is given by,

$$D(x) = (f \oplus b)(x) = \sup_y \{f(y) + b(x - y)\} \tag{3.7}$$

In the above equation, 'sup' denotes the supremum or least upper bound of a partially ordered set. (Function $f(x)$ can be represented as partially ordered set.)

(iii.) Dilation of discrete function $f(n)$ with structuring function $b(m)$ is similar to its continuous counterpart and is defined as,

$$D(n) = (f \oplus b)(n) = \max\{f(n - m) + b(m)\} \tag{3.8}$$

$$0 \le (n - m) \le n, m \ge 0$$

Dilation thickens the boundary of the signal i.e. this operation 'dilates' any sharp change present in the boundary. Hence any high frequency noise superimposed on the signal will get blunted due to dilation.

Erosion is the dual operation of dilation. In erosion operation, first SE is placed somewhere in the signal and if SE is completely contained by the signal then the dilated signal includes the origin of the SE. The eroded signal is obtained by translating the SE to the whole body of signal including boundary. Erosion can also have of three forms depending on the type of signal probed. These are given in the equations (3.9), (3.10) and (3.11). In these equations, operator ' Θ ' is the Erosion operator.

(i.) Binary Erosion is related to Minkowski set subtraction. Erosion of F by structuring element B is the complement of dilation of F^c by the same structuring element, as shown in equation (3.9).

$$F \Theta B = \{\varphi \mid \forall \beta \in B, \varphi + \beta \in F\} = (F^c \oplus B)^c \tag{3.9}$$

(ii.) Erosion of $f(x)$ with structuring function $b(x)$ is defined as,

$$E(x) = (f \ \Theta \, b)(x) = \inf_y \{f(y) - b(x-y)\} \tag{3.10}$$

In the above equation, 'inf' signifies the infimum or greatest lower bound of a partially ordered set.

(iii.) Erosion of $f(n)$ with structuring function $b(m)$ is similar to continuous counterpart and is given by,

$$E(n) = (f \, \Theta b)(n) = \max\{f(n+m) - b(m)\} \tag{3.11}$$

$$0 \le (n+m) \le n, m \ge 0$$

Erosion performs the reverse operation of dilation and 'erodes' the boundary of the signal. Consequently, the amplitude of the signal alongwith the noise will get attenuated. Hence, both the basic operations, Dilation and Erosion, are capable of denoising the signal. [62]

3.2.3.2 OPENING AND CLOSING

Using the basic operations dilation and erosion, two equally important derivative operations can be formulated. These are called Opening and Closing of a signal [58].

Opening of a signal by a structuring element is defined as Dilation of the eroded signal. This operation opens up narrow valleys, gaps and holes near the boundary but removes small protrusions from the boundary. Three different forms of Opening have been given in equations (3.12), (3.13) and (3.14). In these equations, ' $\circ$ ' is the Opening operator.

(i.) Binary opening of F by structuring element B is defined as,

$$F \circ B = (F \Theta B) \oplus B \tag{3.12}$$

(ii.) Opening of $f(x)$ with structuring function $b(x)$ is given by,

$$O(x) = (f \circ b)(x) = ((f \Theta b) \oplus b)(x) \tag{3.13}$$

(iii.) Opening of $x(n)$ with structuring function $f(m)$ is defined as,

$$O(n) = (f \circ b)(n) = ((f \Theta b) \oplus b)(n) \tag{3.14}$$

Closing of a signal by a structuring element is the dual operation of opening and is defined as Erosion of the dilated signal. Closing performs the reverse operation of opening and fills the narrow valleys and gaps in the proximity of the boundary [65]. Three different forms of Closing have been given in equations (3.15), (3.16) and (3.17). In these equations, '$\bullet$' is the Closing operator.

(i.) Binary Closing of F by structuring element B is given by,

$$F \bullet B = (F \oplus B) \Theta B \tag{3.15}$$

(ii.) Closing of $f(x)$ with structuring function $b(x)$ is defined as,

$$C(x) = (f \bullet b)(x) = ((f \oplus b) \Theta b)(x) \tag{3.16}$$

(iii.) Closing of $x(n)$ with structuring function $f(m)$ is given by,

$$O(n) = (f \bullet b)(n) = ((f \oplus b) \Theta b)(n) \tag{3.17}$$

Opening and Closing operations impart composite effects of dilation and erosion upon the signal [63] [64].

While performing morphological operations, every bi-level signal is considered to be a subset of Euclidean space. Similarly, multilevel signals are considered as functions that map Euclidean subspace into the set of real numbers. Figure 3.4 illustrates the effects of Dilation, Erosion, Opening and Closing by a linear SE upon a noisy sinusoidal signal. This figure clearly shows that, all the basic operations are very effective for denoising.

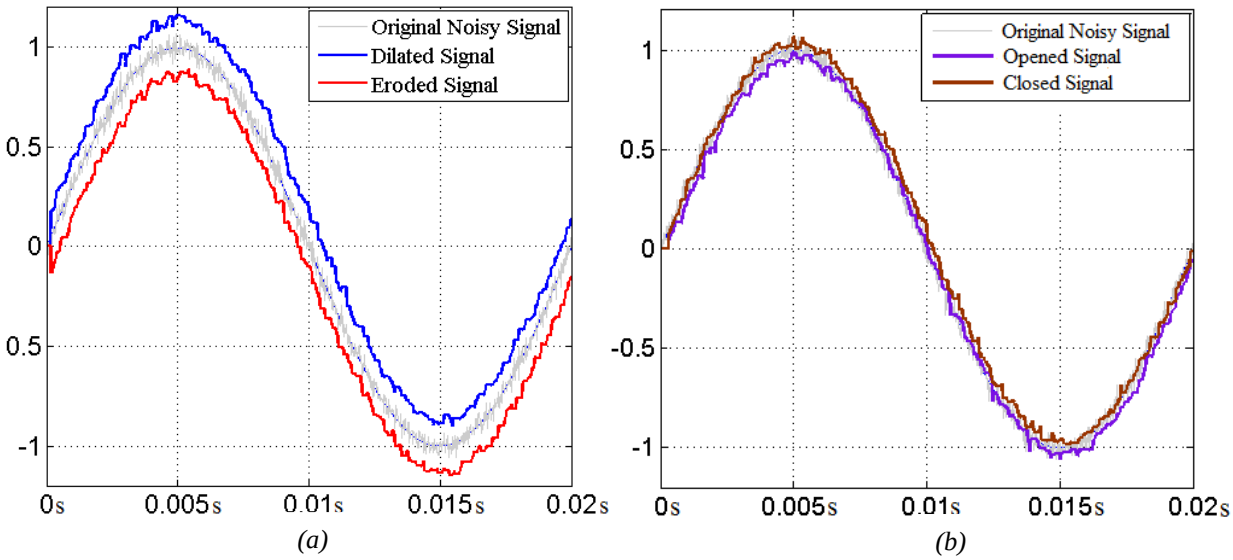


Figure 3.4 Effects of basic morphological operations upon a noisy sinusoidal signal by a linear SE (a) Dilation and Erosion (b) Opening and Closing

Different properties of the four basic morphological operations have been provided in Table 3.2.

Table 3.2 Different Properties of Dilation, Erosion, Opening and Closing

Property	Dilation	Erosion	Opening	Closing
Extensiveness (Resultant signal after operation is a proper subset of original signal)	Extensive	Anti-extensive	Anti-extensive	Extensive
Increasingness (Non-empty subsets of signal remain unchanged after operation)	Increasing	Increasing	Increasing	Increasing
Duality	Dual of Erosion	Dual of Dilation	Dual of Closing	Dual of Opening
Idempotency (Signal remains unchanged if operation is performed more than once)	Non-idempotent	Non-idempotent	Idempotent	Idempotent

Apart from these four basic operations, different hybrid operations can be performed upon the signal by performing the basic operations in different possible sequences. Hence, several different hybrid morphological filters can be implemented. Some of these hybrid operation are: Hit-or-miss transform, Top hat transform, Morphological Median Filters, Open-closing Maximal and Close-opening Minimal, Generalized Multi-resolution Morphological Gradient, Multi Resolution Morphological Opening Closing, Multi-resolution Morphological Gradient (MMG), Series MMG etc. [65]

The MM based filters are known for their very low calculation burden because basic operations involve additions and multiplications. Another advantage of is that morphological operations can fill up missing information in a signal [60]. However, there is an issue with the morphological filters when real time streams of data form a digitized signal are processed. As already mentioned, morphological operations act upon the signals like filters. Generally during real time processing, data from the digitized signal are passed through the filter window continuously. But until the window gets filled up, morphological filters will not be able to yield any output. Therefore some time delay is introduced at the start of the processing during filling up of the window. The filter delay introduced by morphological operations, while processing a real time data, is proportional to the size of the corresponding SE [65]. Hence tuning of length of the SE is very important. A small length of SE decreases the computational burden but will not filter out the noise and the lower order harmonics will become prominent. On the other hand, if too large a length is chosen then noise in the signal will be reduced but the consequent filter delay will introduce lag in the output signal. Thus, an optimal length of SE needs to be obtained for a particular application by trial and error method [65].

3.3 FEATURE EXTRACTION FROM MORPHOLOGICALLY TRANSFORMED OPTICAL PD SIGNALS

In the present work, MM has been employed to transform and denoise the optical PD signals obtained through optical sensors and subsequently significant statistical features have been extracted from the transformed optical PD signals. During the PD data acquisition, the analog optical signals emanated from optical PD source were captured after digitizing with the help of a digitizer. Hence, the recorded optical PD signals are to be considered as multilevel discrete signals.

It has been already mentioned in Chapter 2 that most of the captured optical PD signals are sparse in nature due to sporadic charge flow through cylindrical void during the time span of data capture. Features extracted from such sparse signals may not be consistent and may yield very low numeric values. As morphological operations can fill up missing data in a signal, it will be advantageous to employ such operations for filling up of sparse domain of the recorded optical PD signals. Due to this reason, the recorded data pertaining to optical PD signals have been transformed with morphological operations, Dilation and Erosion. As the captured PD signals are multilevel discrete signals, Dilation and Erosion operations as given in equations (3.8) and (3.11) are applicable in this case. The choice of Structuring Element (SE) is crucial for probing the features from the signals. In the present work, a discrete structuring element, $b(m) = [1, 1, 1]$ was chosen. This SE gives satisfactory results as far as the saliency of the extracted features is concerned.

Nine important statistical features have been extracted from the morphologically transformed optical PD signals. The nine features are **Skewness of dilated signal**, **Skewness of eroded signal**, **Kurtosis of dilated signal**, **Kurtosis of eroded signal**, **Equivalent width of dilated signal**, **Centroid of dilated signal**, **Mean Square Width of dilated signal**, **Average or Mean of dilated signal** and **Average of eroded signal** respectively. All the features and their corresponding mathematical expressions have been listed in Table 3.3. In these nine features, $D(n)$ and $E(n)$ are dilation and erosion of the obtained optical PD signals respectively, while μ and σ are the Mean and Standard Deviation of the same signal respectively. Note that apart from these features, some other features could also have been extracted by performing more complex operations on the signals, such as Opening and Closing. But it has been seen that these nine features are sufficient to classify the optical PD type and location with excellent accuracy.

Table 3.3 Different statistical features extracted from morphologically transformed optical PD signals

Sl. No.	Name of the feature	Mathematical expression
1.	Skewness of dilated signal	$F_1 = D_S = \frac{\sum_{n=0}^{N}(D(n)-\mu)^3}{(n+1)\sigma^3}$
2.	Skewness of eroded signal	$F_2 = E_s = \frac{\sum_{n=0}^{N}(E(n)-\mu)^3}{(N+1)\sigma^3}$
3.	Kurtosis of dilated signal	$F_3 = D_k = \frac{\sum_{n=0}^{N}(D(n)-\mu)^4}{(N+1)\sigma^4}$
4.	Kurtosis of eroded signal	$F_4 = E_k = \frac{\sum_{n=0}^{N}(E(n)-\mu)^4}{(N+1)\sigma^4}$
5.	Equivalent width of dilated signal	$F_5 = Ew_D = \frac{1}{D_{\max}}\sum_{n=0}^{N} nD(n)$
6.	Centroid of dilated signal	$F_6 = C_D = \frac{\sum_{n=0}^{N} nD(n)}{\sum_{n=0}^{N} D(n)}$
7.	Mean Square Width of dilated signal	$F_7 = Msw_D = \sqrt{\frac{\sum_{n=0}^{N} n^2 D(n)}{\sum_{n=0}^{N} D(n)}}$
8.	Mean of dilated signal	$F_8 = Avg_D = \frac{\sum_{n=0}^{N} D(n)}{N+1}$
9.	Mean of eroded signal	$F_9 = Avg_E = \frac{\sum_{n=0}^{N} E(n)}{N+1}$

Figure 3.5 depicts the effects of Dilation and Erosion by the chosen SE upon a sample optical PD signal. Clearly it can be seen from this figure that Dilation has filled up the domain of the signal whereas Erosion has diminished variations from amplitude.

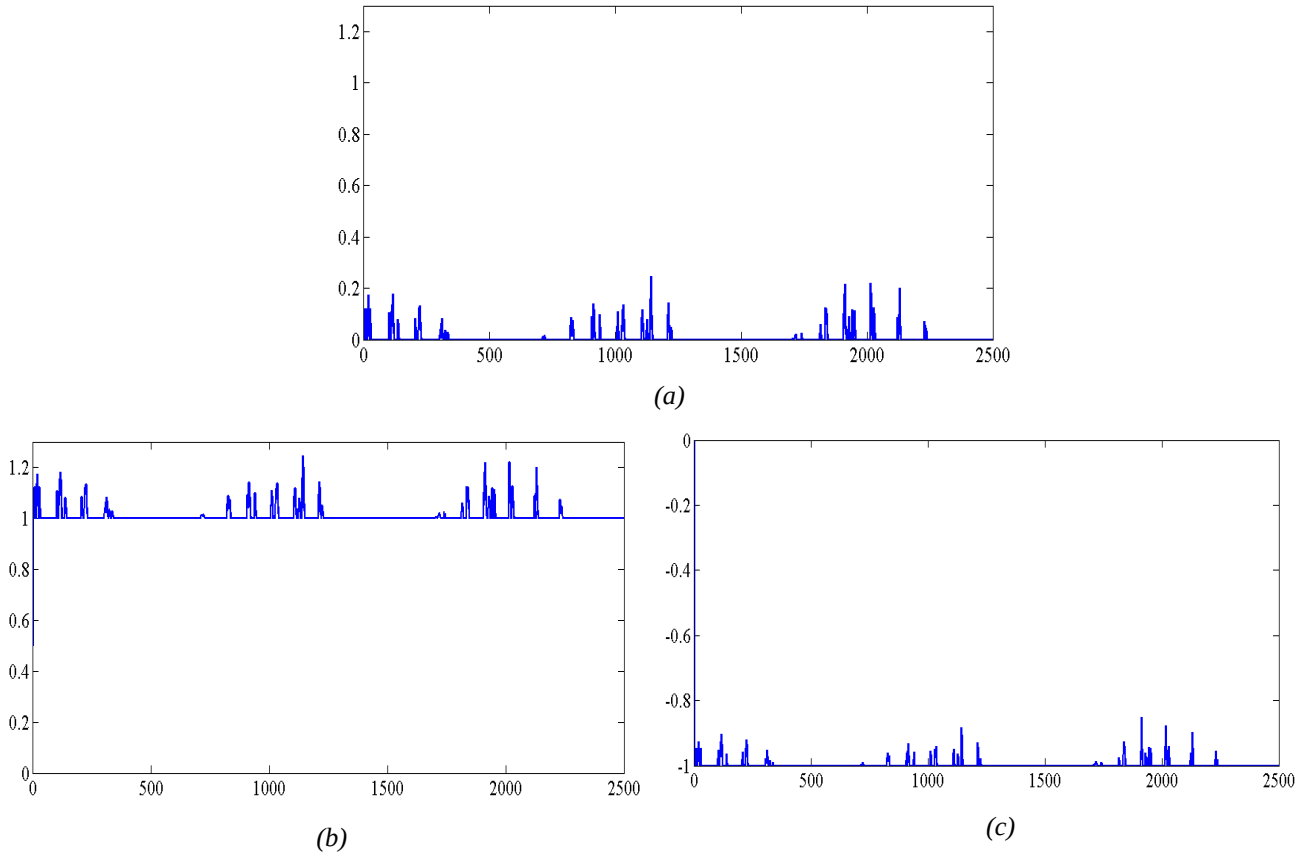


Figure 3.5 Effect of Dilation and Erosion upon a sample PD signal by the chosen SE (a) original signal (b) dilated signal (c) eroded signal

There are five data sets, each set consisting of five data from five sensors, corresponding to each SPD and DPD event. These data sets were captured with an interval of 1 hour. Some of the extracted features from morphologically transformed optical PD data have been presented in Table 3.4. In that table, the five sensors have been represented as S1, S2, S3, S4 and S5. Now consider, for example, Single Optical PD event 1A i.e. when optical PD source has been placed at 1A location of virtual cubic region and five sensors capture optical signals simultaneously. The nine extracted features for the signals captured by each of the five optical sensors have been put in first five rows respectively. This procedure is followed for rest of the SPD and DPD events. Hence there will be five set of nine features for each event pertaining to signals captured by five sensors. In Table 3.4, notations of the optical events have been provided based on the scheme shown in Table 2.1. These extracted features will be further classified by classification techniques to ascertain the type and location of optical PD source. The classification of features has been discussed in the subsequent chapters.

Table 3.4 Extracted features from morphologically transformed optical PD signals

Sensor no.	SPD Event	Extracted features for single optical PD events								
		F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9
OS1	1A	3.55	35.31	25.82	1248.00	1608800	1246.49	1439.97	1.03	-0.99
OS2	1A	-28.29	33.79	940.33	1175.33	2856226	1251.49	1444.39	1.00	-0.99
OS3	1A	-31.91	35.08	1094.09	1237.28	2931388	1251.19	1444.16	1.00	-0.99
OS4	1A	-29.30	34.58	981.28	1213.35	2845782	1251.19	1444.17	1.00	-0.99
OS5	1A	-25.58	33.99	825.69	1185.37	2860528	1250.95	1443.93	1.00	-0.99
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
OS1	9C	-20.09	35.31	627.39	1248.00	2624835	1250.89	1444.00	1.00	-0.99
OS2	9C	-16.20	32.19	481.66	1099.05	2627452	1250.18	1443.32	1.00	-0.99
OS3	9C	3.15	8.45	24.65	112.45	1662812	1251.89	1443.86	1.04	-0.98
OS4	9C	2.83	12.62	19.32	247.91	1636869	1249.55	1442.78	1.04	-0.99
OS5	9C	-9.97	28.75	288.09	936.47	2541650	1250.24	1443.52	1.00	-0.99
Sensor no.	**DPD Event**	**Extracted features for double optical PD events**								
		F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9
OS1	1A_1C	-20.77	35.31	650.98	1248.00	1619121	1246.20	1439.78	1.00	-0.99
OS2	1A_1C	-16.79	32.36	501.41	1106.84	2664961	1251.35	1444.24	1.00	-0.99
OS3	1A_1C	3.00	8.78	24.31	122.73	1632155	1252.31	1445.84	1.03	-0.98
OS4	1A_1C	-25.02	32.98	802.64	1137.75	2852459	1250.93	1443.94	1.00	-0.99
OS5	1A_1C	-18.53	30.75	553.44	1032.56	2801675	1250.53	1443.64	1.00	-0.99
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
OS1	9C_7C	-21.46	35.24	668.44	1244.34	2771578	1250.62	1443.75	1.00	-0.99
OS2	9C_7C	2.86	16.24	54.26	365.25	1864573	1257.48	1449.03	1.01	-0.99
OS3	9C_7C	2.55	6.92	18.65	79.31	1642037	1253.31	1445.60	1.05	-0.98
OS4	9C_7C	1.83	16.96	24.96	420.97	1910036	1249.98	1443.14	1.03	-0.99
OS5	9C_7C	-9.94	30.77	274.50	1039.94	2617502	1249.46	1442.93	1.01	-0.99

CHAPTER 4

FEATURE CLASSIFICATION BY ROUGH SET THEORY AND SPARSE REPRESENTATION CLASSIFIER

CHAPTER 4.

FEATURE CLASSIFICATION BY ROUGH SET THOERY AND SPARSE REPRESENTATION CLASSIFIER

4.1 INTRODUCTION

The statistical features extracted from morphologically transformed optical PD signals have been separately classified employing two different classification techniques – Rough Set Theory and Sparse Representation Classification. The working principles of these classification techniques have been discussed in this chapter.

4.1.1 CLASSIFICATION PROBLEM

Classification is a part of Supervised Machine Learning technique and is defined as the process of categorizing different observations or instances into different classes. Here the term 'observation' is a generic term that can signify a large number of phenomena such as outcome of any experiment, property of any object, occurrence of any event or incident etc. Observations can be qualitative (i.e. in terms of shape, colour, texture, taste, existence, perceptibility etc.) as well as quantitative (comprising numerical values). They are the input arguments to the classification problem.

In a classification problem, correctly identified observations or instances are available in priori alongwith their category memberships (i.e. which observations belong to which classes) and based on the currently available knowledge about these observations, some decisions are to be taken or forecasts are to be made. Category memberships are implemented by membership functions which perform a predictive mapping from the available observations to the set of classes i.e. they assign the classes to the observations.

Classes can be easily assigned to a large set of observations by extracting suitable features from these observations. Hence, preprocessing a set of observations through feature extraction is desirable before actual classification is performed. Same set of observations can yield different features for different membership functions.

Membership functions are most commonly implemented by mathematical or multi-level logical functions. Choice of membership function is largely dependent on type of observations available and the relevant problem in hand.

The decision rule for class assignment can be simply put in the form of lower given proposition.

For an observation, **IF** Feature1 = ‘a’ and Feature2 = ‘b’ and so on … **THEN** this observation belongs to class ‘X’.

Hence in the context of classification, features are the salient characteristics of the observations that make a class of observations different from the other one. A subset of features would have similar set of values for different observations if these observations belong to same class. In case of objects belonging to a different class, that subset of features would have different set of values.

The set of available observations alongwith their feature vectors and assigned classes is known as the training data set. The classes are predictively assigned by the chosen membership functions. Training data set enables the formulation of a set of ‘IF-THEN’ decision rules pertaining to different classes, like the one indicated in the previously shown proposition.

Given an observation of unknown class, the aim is to classify it by extracting the same set of features as the training data set and finding out the decision rule that the feature vector complies with. This decision rule will further help in identifying the class, which the observation belongs to.

Unclassified observations are said to form the testing data set, which is used to validate the predictions warranted by the training data set. Not all observations of the testing data set can be suitably classified. Classification accuracy is defined as the number of observations or instances successfully categorized, expressed as percentage of total number of observations or instances. Thus,

$$Percentage\ Classification\ Accuracy = \frac{No.\ of\ observations\ or\ instances\ successfully\ classified}{Total\ no.\ of\ observations\ or\ instances} \times 100$$

An algorithm that implements the classification is known as the classifier. Classifier algorithms formulate the decision rules from already available information and can be repeatedly employed for making decisions or predictions based on new information available. Percentage classification accuracy is the main index for performance evaluation of such classifier algorithms.

4.2 ROUGH SET THEORY

Rough Set Theory (RST) is one of the most important and powerful mathematical tools for classification. It was first proposed by computer scientist Zdzislaw Pawlak in 1982 [70]. RST is specially suited for applications where the objects to be classified are imprecise and superfluous. Decision rules for class assignment cannot be formulated if input data contains imprecision. RST enables the formulation of decision rules by removing the imprecisions present in the input data [70 - 75].

RST has been formulated from conventional set theory considering the notion of vagueness. In conventional set theory, an element can fully belong to a set (or have total participation in a relation), or it does not (i.e. does not have any participation in any relation). Thus every element precisely belongs to a set and every set is uniquely described by its elements. Hence, all the set boundaries are considered to be infinitely thin. Such a notion of a set is termed as 'crisp' or precise. All the elements in crisp sets and the relevant set universe are characterised by unambiguous and precise information.

But the concept of crispness led to several antinomies or contradictions which could not be eradicated by theoretical knowledge of classical set theory. One way of avoiding these contradictions is to impose restrictions upon the elements which can or cannot form a set. The restrictions are warranted by properly selected axioms which determine how elements can belong to a set. This notion led to the formulation of Axiomatic set theory.

The other way of avoiding contradictions is to introduce the concept of vagueness [70]. Vagueness arises in case of elements which are imprecise, i.e. they cannot be fully classified into a set or cannot have full participation in a relation. Several different modifications to the classical set theory were suggested to incorporate vagueness present in a set of data. For example, Fuzzy Set Theory is an ingenious approach to handle imprecision present in input data by hypothesising that a set of data can belong to a class in a certain level or percentage [71].

Rough Set Theory handles vagueness in a different way. In RST, the imprecision is represented as the boundary region of a set. As the knowledge about the universe and its elements are imprecise, with respect to a relation defined on the universe, if any element does not belong to a set or its complement then it belongs to the boundary region. Hence, boundary region of a set comprises of elements which neither belong to the set nor to its complement. When this boundary is non-empty, the set is 'rough'. If the boundary is empty, then the set is 'crisp'. Hence, RST generalises the classical set theory.

4.2.1 INFORMATION SYSTEM

RST explicitly works upon input data presented in tabular form. This table is called the data table or decision table. In a decision table, each row represents an object or a case, each column represents an attribute and entries of the table represent attribute values. The attributes normally include condition attributes and decision attributes. Condition attributes denote some property or observation of the objects. Decision attributes denote the class of the objects which are determined based on those properties or observations.

The decision table is called an Information System [35] [74]. It can be mathematically represented as,

$$T = \langle U, Q, V, f \rangle \tag{4.1}$$

Here, U is the finite set of all objects i.e. it represents the universe of objects, Q is the set of attributes. If q is an attribute variable then $V = \bigcup_{q \in Q} V_q$ denotes set of values accepted by q i.e. the domain of q. Lastly, f is the decision function that maps from $(U \times Q)$ matrix to V and classifies every object to form the set of decision attributes.

An example of a decision table has been presented in Table 4.1. This table shows information about a set of transformers which have been accepted or rejected based on two condition attributes - whether distortion in current and voltage waveforms during impulse test is within acceptable limit and whether partial discharge present in the transformer is within acceptable limit. Transformers are the objects here and whether the transformers are accepted or rejected is the decision attribute.

Table 4.1 Example of a decision table containing imprecision

Objects (Transformers)	**Condition attributes**		**Decision attribute: Accept or reject the transformer**
	Distortion in current and voltage waveforms within acceptable limit	**Partial discharge within acceptable limit**	
T1	Yes	Yes	Accept
T2	No	No	Reject
T3	Yes	No	Accept
T4	No	No	Reject
T5	Yes	Yes	Accept
T6	Yes	No	Reject

From Table 4.1, simple decisions can be formulated such as - “If distortion in current and voltage waveforms during impulse test is within acceptable limit and partial discharge is within acceptable limit then transformer is accepted”. Decision rules can be formulated from each row of the table.

It can be observed form the information system of Table 4.1 that, objects T3 and T6 have exactly same values for the condition attributes but their corresponding decision attributes are different. Transformer T3 has been accepted and but the other transformer T6 has been rejected. Hence, this decision table contains imprecise information.

The decision rules formulated from any information system must conform two important properties [70][74]-

- **Consistency** – Objects comprising exactly same values for all the attributes must belong same decision class. contain
- **Minimality** – Decision rules are constructed from minimal set of attributes.

For imprecise decision tables, it is very hard to ensure consistency and minimality of the formulated decision rules. In addition to that, if decision table comprises numerical values for any of the condition attributes, it is a normal practice to normalise those numerical values with respect to a well-defined base, such as with respect to maximum and minimum value in the range or mean and standard deviation of those numerical values.

4.2.2 INDISCERNIBILITY RELATION

In a data table, some objects may be characterised by exactly same set of values with respect to some set of attributes. These objects are called 'indiscernible' or indistinguishable with respect to those attributes [35]. Mathematically, if P is a subset of attribute i.e. $P \subseteq Q$ then two objects α_i and α_j belonging to universal set U are indiscernible with respect to set of attributes P if,

$$f(\alpha_i, q) = f(\alpha_j, q), \forall q \in P \text{ and } \alpha_i, \alpha_j \in U \tag{4.2}$$

The set of such indiscernible objects is called an elementary set. Elementary sets represent granules of knowledge about the universe. Based on different subset of attributes, different elementary sets can be formed.

For a subset of attributes P, an equivalence relation defined on U is given by,

$$I_P = \{(\alpha_i, \alpha_j) \in U \times U \mid \forall q \in P, f(\alpha_i, q) = f(\alpha_j, q)\} \tag{4.3}$$

I_P is called the P-indiscernibility relation and represents the knowledge that objects (α_i, α_j) are indiscernible or equivalent with respect to subset of attributes P. Equivalence relation I_P introduces partitions in the set of objects U with respect to P, by forming all possible elementary sets corresponding to each of the different combination of values assumed by P. For each of the possible combinations, objects are grouped to form subsets and objects in same subset have exactly same combination of

attribute values. These elementary subset of objects are termed as 'Information Granules' and all the objects in a granule are considered indistinguishable [74].

Concept of information granules can be simply illustrated by considering the information system of Table 4.1. Let, P = {Distortion in current and voltage waveforms within acceptable limit} and the value assumed by P can be either 'Yes' or 'No'. Clearly, objects T1, T3, T5 and T6 are indiscernible with respect to P and can be grouped together to form an elementary set, because all these objects have attribute value 'Yes'. In similar way, objects T2 and T4 are indiscernible with respect to P and can be grouped together to form another elementary set, as both these objects have attribute value 'No'. Hence, considering all possible combination of values for P, two elementary sets can be formed. Therefore, the equivalence relation I_P introduces two partitions in the set of all objects by distributing the objects into two elementary sets.

$$I_P = \{\underbrace{\{T1, T3, T5, T6\}}_{P = \text{'Yes'}}, \underbrace{\{T2, T4\}}_{P = \text{'No'}}\}$$

Similarly if we assume, P = {Partial discharge within acceptable limit} then,

$$I_P = \{\{T1, T5\}, \{T2, T3, T4, T6\}\}$$

Any number of attributes can be selected to form the granules. Identification of information granules from an information system is the first step of towards reducing imprecision present in the decision table.

4.2.3 SET APPROXIMATIONS

After information granules are identified from an information system with respect to all the attributes, decision rules can be formulated to classify the objects. But due to lack of knowledge about the universe and imprecision in the data, not every granule can be 'crisply' classified. Objects belonging to problematic granules are generally characterised by different values for decision attributes. Set approximation provides a way to delineate such problematic granules.

Every rough set is characterised by two crisp sets, which are called the lower and upper approximations [70]. For any set A, $\underline{P}A$ and $\overline{P}A$ are called the P-lower and P-upper approximation of A with respect to set of attributes P respectively. These approximations are defined as,

$$\text{P-lower approximation} = \underline{P}A = \{\alpha \in A \mid I_P(\alpha) \subseteq A\} \quad (4.4)$$

$$\text{P-upper approximation} = \overline{P}A = \{\alpha \in A \mid I_P(\alpha) \cap A \neq \phi\} \quad (4.5)$$

The lower approximation of set A with respect to P comprises of all objects which are certainly classified as members of A, based on knowledge available in P. On the other hand, upper approximation of set A is the set of objects which 'possibly' belong to A.

The difference between upper approximation and lower approximation represents the P-boundary region of set A and is given by,

$$B_P(A) = (\overline{P}A - \underline{P}A) \tag{4.6}$$

Boundary region consists of those objects which cannot be decisively classified as members of A based on the information available in P. The rest of the universe is P-outside region of A and is given as, $(U - \overline{P}A)$. The set A will be 'Rough' if and only if $B_P(A) \neq \phi$ i.e. the boundary is non-empty, otherwise it is 'Crisp' [74] [75]. All the set approximations have been depicted in Figure 4.1.

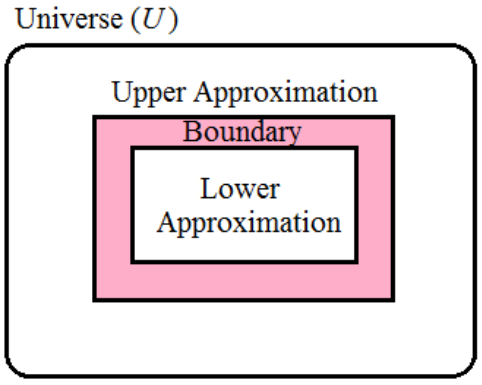


Figure 4.1 Different set approximations

Again considering the information system of Table 4.1, let P = {Distortion in current and voltage waveforms within acceptable limit, Partial discharge within acceptable limit}.

For this set of attributes, the information granules are,

$$I_P = \{\{T1, T5\}, \{T2, T4\}, \{T3, T6\}\}$$

Let all the objects which belong to decision class 'Accept' are to be determined. It is evident that elementary set {T1, T5} unambiguously belongs to this decision class. Therefore, these objects construct the P-lower approximation. Among rest of the granules, elementary set {T3, T6} is a problematic granule because both the objects have same values for set of attributes P but have different decision attributes. Hence these objects cannot be classified crisply and belongs to boundary with respect to set of attributes P. The objects {{T1, T5}, {T3, T6}} together construct the P-upper approximation. Lastly, both the objects in the elementary set {T3, T6} have decision 'Reject'. Hence these objects precisely belong to the P-outside region. All the formulated set approximations have been depicted in Figure 4.2.

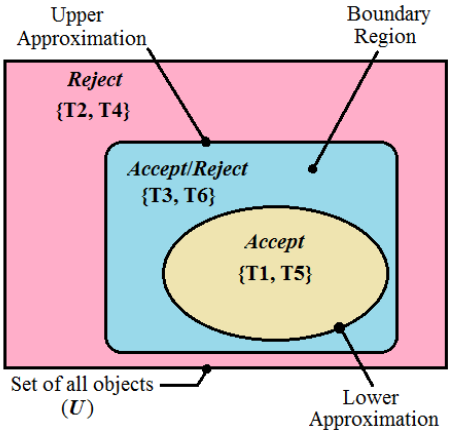


Figure 4.2 Set approximations for information system of Table 4.1

Hence, the Rough Set Theory provides a way to diminish imprecision by identifying objects which certainly belong to or do not belong to a decision class and delineating the objects which belong to the boundary region.

4.2.4 DISCRETIZATION

After information granules are constructed from a decision table, individual objects can be classified into representative classes and objects which cannot be properly classified can be demarcated with the help of set approximations. But for a table of very large size, this is too crude a method because chance of classifying a new object (with set of its attributes matching one of the rows of decision table) will be very low. This is particularly true for information systems which have numerical attributes [35] [74].

The largeness of the decision table is comprehensively given by the Cardinality operator applied to any set A, $card(A)$. Hence for a decision table with very large number of attributes, $card(V_q)$ is very high for some $q \in Q$. Cardinality simply means 'number of elements of a set or participating in a relation'. High cardinality of decision table can be reduced employing a special technique called the 'Discretization'.

Discretization is the process of inducing some partition in the decision table for obtaining higher classification efficiency. These partitions are inducted by generating cuts in the domain of q and individual values for attributes are replaced with class-values i.e. the data within a range are given a particular class which decreases the cardinality [35].

Discretization can be illustrated using a simple information system as shown in Table 4.2. This decision table comprises of six objects and one condition attribute which assumes numerical values.

Table 4.2 A simple decision table

Object	Condition attribute
X1	0.18
X2	0.04
X3	0.36
X4	0.34
X5	0.09
X6	0.26

Suppose that we have additional information about a new object that it belongs to one of the object classes in Table 4.2. If that new object is to be classified using the decision rules formulated from that table, then numerical value pertaining to its condition attribute should match with numerical value of one of the condition attributes. But due to imprecision, it may so happen that the numerical value pertaining to condition attribute of that new object does not completely match with any of the values in Table 4.2. In order to allow for that tolerance, each attribute is assumed to belong within a range and each range is assigned with a class. Any new object which has condition attribute within a range belongs to the corresponding class.

As shown in Table 4.3(a), different ranges of the numeric values have been assigned with different class values, e.g. all numeric values within range (0 – 0.05) are assigned class 0, values within range (0.05 – 0.15) are assigned class 1 and so on. Information system of Table 4.2 is discretized by replacing all the numerical values corresponding to condition attributes with respective class values. The discretized decision table has been shown in Table 4.3(b). Suppose that the new object has condition attribute 0.1. It should not belong to any of the object class if exact matching was mandated. But now the domain of condition attribute has been differentiated into five classes and this object belongs to class 1. Hence, new object is in the same class as object X2.

Range of numeric value	Class value
0 - 0.05	0
0.05 - 0.15	1
0.15 - 0.25	2
0.25 - 0.35	3
0.35 - 0.45	4

(a)

Object	Condition attribute
X1	2
X2	0
X3	4
X4	3
X5	1
X6	3

(b)

Table 4.3 (a) Assigned classes to different ranges (b) Discretized form of Table 4.2

Clearly, this process decreases cardinality because instead of infinite number of possible values between maximum and minimum value in the range, the range is now reduced to only 5 possible values. In this way, numeric values pertaining to all the attributes can be discretized. Generation of optimum cuts in the domain of a numerical condition attribute and identification of ranges is a difficult task. Several popular discretization algorithms are available for this purpose.

4.2.5 REDUCTS AND CORES

As already discussed, identification of equivalence classes reduces imprecision present in the data table because one element of equivalence class represents the entire class. But there may be superfluous information present in the decision table which cannot be removed by identification of indiscernibility relations. The redundancy in the data table can be removed by identifying the attributes which are minimally sufficient to classify each object of the data table [74]. Other redundant attributes are to be dispensed with, as their removal does not hamper the indiscernibility of the objects.

For a subset of attributes P, any attribute $q \in P$, is dispensable if and only if its removal does not disturb the corresponding indiscernibility relation I_P. Mathematically it can be denoted as, $I_P = I_{P-\{q\}}$. Otherwise the attribute will be indispensable.

If all the elements of P are indispensable, then P is called independent set of attributes. This minimally sufficient subset of attributes is called the 'Reduct' and is denoted as $RED(P)$.

Let R be another subset of attributes such that $R \subseteq Q$ and has equivalence relation I_R in U. Now if, utilizing the knowledge described by I_P, some of the objects can be classified into R-elementary set obtained from I_R then that set of classifiable objects are called the P-positive region of R and is denoted as,

$$POS_P(R) = \bigcup_{A \in I_R} \underline{P}A. \tag{4.7}$$

Note that right hand side of equation (4.6) signifies union of all objects which are characterized by I_R and belongs to P-lower approximation of set A (hence they are certainly classified). Any attribute, $q \in P$ will be called R-dispensable in P if $POS_P(R) = POS_{P-\{q\}}(R)$ i.e. its removal does not perturb the classification into R-elementary set. Otherwise it will be termed as R-indispensable in P. Similar arguments follow for independent attributes. Finally, for a subset of attributes S which is R-independent in P and $POS_P(R) = POS_S(R)$, subset S will be called the Reduct of P.

In order to construct the Reduct from a decision table, first equivalence classes are identified by defining the indiscernibility relations with respect to set of attributes. Now if it is seen that, some condition attributes assume exactly same values for a particular decision class, then these condition attributes can be considered as superfluous. Therefore, these redundant attributes can be dispensed with and considered as 'don't cares'. Construction of Reducts is dependent on the subset of attributes chosen and for different subsets, Reducts will be different [35].

Another important quantity is 'Core' which removes imprecision present in a single decision class. Core signifies the set of relations occurring in every Reduct. Cores are obtained from Reducts by intersecting the Reducts corresponding to a single decision class i.e. $CORE(P) = \cap RED(P)$. Condition attributes within a Reduct, which assume different values corresponding to a single decision class, provide vague information about that decision class. Such attribute values are removed by set intersection while forming the Cores. Construction of Reducts and Cores after identification of equivalence classes effectively removes imprecision and redundancy from the information system. Once Cores are formed, decision rules can be easily formulated.

4.2.6 FORMULATION OF DECISION RULES AND CLASSIFICATION OF NEW OBJECTS

Formulation of decision rules implies identification of relations that exist within attributes after removal of imprecision and representing them in 'IF…THEN' format [74]. It can be simply depicted as,

If for an object, ((Attribute1 = a) and (Attribute2 = b) and …) Then the object belongs to decision class X.

Any new object belongs to a certain decision class if it is characterised by exactly same set of attribute values for corresponding to that decision class.

It has been already discussed that, in a classification problem, different observations and their class memberships are available in the form of training data set. Classes are generally assigned to observations by extracting suitable features from these observations. Hence in the context of classification by RST, the training data set is considered as an information system, the observations are considered as objects and different extracted features are considered as attributes. The training data set is made free from imprecisions by identification of equivalence classes, formation of set approximations, discretization and lastly construction of Reducts and Cores. Decision rules are formulated from the Cores. Based on these decision rules, the observations in the testing data set will be classified.

Different steps employed for construction of decision rules from an initial raw data through dimensionality reduction and RST have been shown in Figure 4.3.

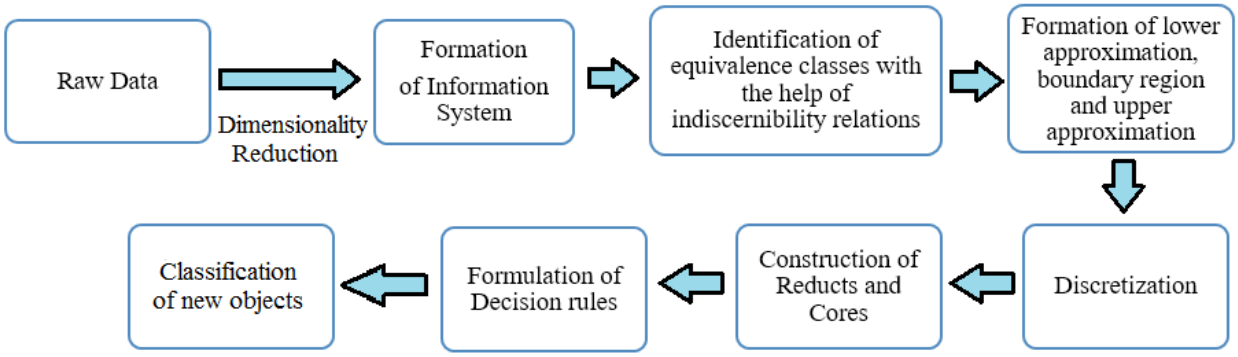


Figure 4.3 Flowchart showing different steps for construction of decision rules from an initial raw data through RST and dimensionality reduction

4.3 SPARSE REPRESENTATION CLASSIFICATION

Sparse representation classification is a relatively new classification technique. It is a part of specialised mathematical methodology called the Parsimony which is defined as the principle of solving a problem with least amount of resources [76] [77].

Any matrix or vector will be called sparse if most of its elements are zero. Sparse representation aims to characterise a system employing such sparse matrices or vectors. Fundamentally, sparse representation deals with the methodology of finding approximate sparse solutions to a linear system of equations.

A system of linear equations is a collection of several linear algebraic equations involving same set of variables. Unique solution to such a system exists if number of equation is exactly equal to number of unknown. But in many situations, it is often required to find solution of a system where number of equations is less than the number of unknown variables. Such systems have infinite number of solutions possible and are generally termed as underdetermined systems.

Among the infinite number of possible solutions, the optimum choice would be to obtain the sparse solutions i.e. solutions with few non-zero elements [78]. Sparse solutions facilitate representation of the underlying physical system with few number of candidate variables. Finding sparse solution of a system provides several practical benefits because such representation will require very low resources to describe the physical system. However finding the sparse solutions to an underdetermined system is generally very difficult and is performed by employing specialised optimization techniques [81] [88].

Sparse representation has wide variety of applications is many fields. It is most commonly employed for recovery and reconstruction of signals [87]. In such applications, the signals to be reconstructed are represented as sparse linear combinations of elementary signals called the atoms. Another class of reconstruction problem is characterised by imperfect and noisy signals which are not exactly sparse. Candes et al. showed that such systems can also be determined by sparse representation with suitable modifications allowing for noise [79] [80].

The concept of sparse representation of signals by linear combination of atoms can be applied to classification problem where the test samples are expressed as sparse linear combination of training samples [88]. With the help of the linear combinations, the test samples will be classified.

4.3.1 MATHEMATICAL NORM

Concept of the mathematical norm of a vector or a matrix is very important in the context of sparse representation. Mathematically, Norm is the representative of size of the relevant vector or matrix. The L_P-norm of a vector $\mathbf{x}$ is defined as the p^{th} root of the summation of all the elements of $\mathbf{x}$, with each of the element raised to p^{th} power. Thus, L_P-norm is mathematically given as,

$$\| \mathbf{x} \|_p = \left(\sum_{i=1}^{n} | x_i |^p \right)^{1/p} \tag{4.8}$$

where, x_i is the i^{th} element of the vector $\mathbf{x}$ and total number of elements in $\mathbf{x}$ is n. In general, $\mathbf{x}$ is a complex vector having all the elements as complex numbers. If all ordered n-tuples of complex numbers form a complex n-dimensional vector space C^n then, $\mathbf{x} \in C^n$. Note that in equation (4.8) the modulus of each complex element gives their respective absolute values. In case the elements are real numbers, the modulus of each element will give their respective numeric values. Hence, the norm is a function that assigns strictly positive size or length to a vector in a vector space.

Given a non-empty complex vector space V, a norm on V is a function or mapping from V to the set of real numbers, $\Re$, i.e. $\ell : V \to \Re$. Here $V \subset C_n$ and if V is a finite dimensional set then it is termed as Hilbert Space. The n-dimensional real vector space R^n, formed by all vectors with n real numbers as components, is also a proper subset of C^n. Hence the norm function is inclusive of R^n. For two vectors $\mathbf{f}, \mathbf{g} \in V$ and $c \in C^1$, the norm function strictly satisfies following properties,

(i.) $\ell(\mathbf{f}) = \| \mathbf{f} \| > 0$ for $\mathbf{f} \neq 0$ and $\| 0 \| = 0$

(ii.) Absolute Scalability : $\| c \cdot \mathbf{f} \| = | c | \cdot \| \mathbf{f} \|$

(iii.) Triangle Inequality or Subadditivity : $\| \mathbf{f} + \mathbf{g} \| \leq \| \mathbf{f} \| + \| \mathbf{g} \|$

Norm of a matrix is a natural extension of the concepts of a vector norm. For an $m \times n$ matrix $\mathbf{A}$, the L_P-norm is defined as,

$$\| \mathbf{A} \|_p = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} | a_{ij} |^p \right)^{1/p} \tag{4.9}$$

Clearly, like the vector norm, the L_P-norm of matrix $\mathbf{A}$ is the p^{th} root of the summation of all the elements of $\mathbf{A}$, with each of the element raised to p^{th} power.

In equation 4.9, $\mathbf{A}$ is a complex matrix and the corresponding complex matrix space is $C^{m\times n}$. For a matrix having all real elements, the complex matrix space reduces to real matrix space $R^{m\times n}$. Lastly, due to presence of modulus in the expression of norm in equation (4.9), the norm function is again given as, $\ell : V' \to \Re$ where $V' \subset C^{m\times n}$. Matrix norm satisfies all the properties of vector norm including scalability and subadditivity.

4.3.1.1 L_1, L_2 AND L_0-NORM

For different integer values of p different norms are obtained. Among these norms L_1, L_2 and L_0-norm are very important and are most extensively used.

- the L_1-norm is obtained by putting $p = 1$ in equation (4.10) and is given by,

$$\| \mathbf{x} \|_1 = \left(\sum_{i=1}^{n} | x_i | \right) \tag{4.10}$$

  Hence the L_1-norm is basically the summation of absolute values of all the elements of vector $\mathbf{x}$. The L_1-norm of a matrix has similar definition. This norm is called the Manhattan norm or Taxicab norm and is widely used in various fields.

- The L_2-norm is obtained by putting $p = 2$ in the equation (4.11). This norm is the most popular of all norms and is used in almost every field. Mathematically this norm is given by,

$$\| \mathbf{x} \|_2 = \left(\sum_{i=1}^{n} | x_i |^2 \right)^{1/2} = \sqrt{\left(\sum_{i=1}^{n} | x_i |^2 \right)} \tag{4.11}$$

  This norm denotes the square root of summation of each element squared and is generally termed as the Euclidean norm. This norm is a standard quantity for measuring the difference between two vectors. The Euclidean distance between two points in space is basically the L_2-norm computed for the difference of the position vectors of these two points.

- The value $p = 0$ gives a special type of norm called the L_0-norm and is of great importance in the context of sparse representation and classification. The meaning of equation (4.12) becomes vague if $p = 0$. Instead, the L_0-norm of vector $\mathbf{x}$ is defined as,

$$\| \mathbf{x} \|_0 = \#\left(i \mid x_i \neq 0 \right) \tag{4.12}$$

  Clearly it denotes the total number of non-zero elements in a vector and is representative of the cardinality. The L_0-norm of a matrix can be defined similarly.

4.3.2 SOLUTION TO UNDERDETERMINED SYSTEMS - THE MINIMISATION PROBLEMS

A linear system of m equations in n unknown variables $x_1, x_2, \ldots, x_n$ is given by,

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = y_1 \\ &a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = y_2 \\ &a_{31}x_1 + a_{32}x_2 + \cdots + a_{3n}x_n = y_3 \\ &\vdots \\ &a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = y_m \end{aligned} \tag{4.13a}$$

These equations can be represented in the form,

$$\sum_{j=1}^{m} a_{ij}x_i = y_j,\ i = 1, 2, \ldots, n \tag{4.13b}$$

or can be represented in more compact matrix form,

$$\mathbf{Ax} = \mathbf{y} \tag{4.13c}$$

where, $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$ is the solution vector in n-dimensional real space R^n, $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_m \end{bmatrix}^T$ is a given vector in m-dimensional real space R^m and $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$ is the coefficient matrix. The system of linear equations (4.13) will be underdetermined if $m < n$ and overdetermined if $m > n$. We consider only underdetermined case where the solutions are not unique and infinite number of possible solutions exist. In many practical situations, it is often required that a unique solution for such an underdetermined system is to be found out. Generally in those cases, the best possible solution or optimum solution is sought [78].

The problem of finding best possible solution to an underdetermined system should be considered as an optimisation problem. Optimisation is referred to as the method of finding best possible or optimum solution of a problem. The main objective of an optimisation problem is to maximize or minimize a specific function, called the objective function, with respect to a set of constraints. The minimisation problem is the class of optimisation problem in which the objective function is exclusively minimised.

Conventionally, the difficulty of finding an optimum solution is resolved by selecting the solution which has the lowest L_2-norm. Thus, the optimum solution $\hat{\mathbf{x}}$ is given by,

$$\hat{\mathbf{x}} = \min \|\mathbf{x}\|_2 \text{ subject to } \mathbf{Ax} = \mathbf{y} \tag{4.14}$$

The optimisation problem of equation (4.14) is known as the L_2-minimisation problem and can be easily solved by finding pseudoinverse of $\mathbf{A}$ or by Lagrangian multiplier method [88]. As L_2-minimisation finds the solution which has least summation of square, it is also known as Least Square Optimisation. Although it is easy to compute the solution for this optimisation problem, the solution is not unique one. Moreover, the solution is not informative and is generally dense with large non-zero elements. Hence memory requirements will be high and convergence of algorithms will be slow [88].

Instead of finding the solution which has the least summation of square, a more realistic and practical approach will be to find the sparsest solution i.e. the solution which involves least number of non-zero elements [78]. As, L_0-norm of any vector denotes the total number of non-zero elements in that vector, finding the sparsest solution of underdetermined linear system of equation (4.13) implies finding the solution which has lowest L_0-norm. Hence, the problem formulation now changes to as follows,

$$\hat{\mathbf{x}} = \min \|\mathbf{x}\|_0 \text{ subject to } \mathbf{Ax} = \mathbf{y} \tag{4.15}$$

The problem of finding the optimum solution by minimising the L_0-norm is termed as the L_0-minimisation problem. The sparsest solution will represent the corresponding underdetermined system with fewest candidate variables which are not known in priori. Hence, representation of any linear system by the sparsest solution provides very crucial practical advantages such as low memory requirement and faster execution of algorithm.

But unfortunately, finding the sparsest solution to the combinational optimisation problem of equation (4.15) is very complex and almost impossible to solve. In terms of computational complexity, L_0-minimisation problem is generally categorised as Non-deterministic Polynomial-time hard or NP hard problem [84]. This problem can be solved sub-optimally by employing iterative methods such as matching pursuits. Alternatively, the L_0-minimisation problem can be approximated as an L_1-minimisation problem, under the condition that optimum solution $\hat{\mathbf{x}}$ is sufficiently sparse. This alternative approach was proposed by David Donoho in 2006 [78].

In L_1-minimisation problem, the L_0-norm of equation (4.15) is replaced with L_1-norm. Therefore, the formulation of L_1-minimisation problem is given by,

$$\hat{\mathbf{x}} = \min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{Ax} = \mathbf{y} \tag{4.16}$$

Hence, now the solution which has the lowest L_1-norm (i.e. least absolute sum) is

computed. Solution to the L_1-minimisation problem of equation (4.16) can be found out by employing the Convex optimization algorithms such as linear programming or non-linear programming. Convex optimisation is a special type of minimisation technique in which convex functions are minimized over convex sets and if local minimum exists within a convex set then it is the global minimum and is unique. It is similar to the LASSO method in machine learning and statistics [82] [83].

4.3.3 SPARSE LINEAR REPRESENTATION OF SIGNALS

Sparse linear representation of a signal is a relatively new technique in signal processing and classification. Sparse representation provides a way to represent or reconstruct a signal using fewest possible samples. The Nyquist-Shannon sampling theorem in digital signal processing provides a sufficient condition for sampling rate that ensures perfect reconstruction of a bandlimited signal by discrete sequence of equidistant samples without any loss of information due to sampling. This sampling rate is known as Nyquist rate. Any continuous bandlimited signal should be sampled at a rate higher than the Nyquist rate, otherwise there will be imperfections in the reconstructed signal known as aliasing. In sparse representation, the aim is to reconstruct and compress a signal using lower sampling rates than that mandated by Nyquist-Shannon theorem [86].

In sparse representation, a signal is represented as the linear combination of elementary signals known as the atoms, which are the primary information available for reconstruction. The dictionary is said to be made up of atoms such that the total number of atoms available for reconstruction is more than sufficient and hence the dictionary is 'overcomplete' [88].

Let us define the problem mathematically. Suppose that an M-dimensional signal $\boldsymbol{\psi}$, having components $\psi_1, \psi_2, \ldots, \psi_M$ along M different dimensions, is to be reconstructed using elementary signals or atoms, having same dimensions as the signal. Also, let us assume that there are N number of atoms. If $\breve{\boldsymbol{\alpha}}_{\mathbf{j}}$ denotes j^{th} atom and α_{ij} denotes i^{th} component of j^{th} atom, then the components of the atoms can be arranged in the columns of a matrix $\mathbf{D}$, as shown below.

$$\mathbf{D} = \begin{bmatrix} \breve{\boldsymbol{\alpha}}_{\mathbf{1}} & \breve{\boldsymbol{\alpha}}_{2} & \cdots & \breve{\boldsymbol{\alpha}}_{\mathbf{N}} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2N} \\ \vdots & & & \\ \alpha_{M1} & a_{m2} & \cdots & \alpha_{MN} \end{bmatrix} \tag{4.17}$$

where, $\breve{\boldsymbol{\alpha}}_{\mathbf{j}} = [\alpha_{1j} \quad \alpha_{2j} \quad \cdots \quad \alpha_{Mj}]^T$, $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, N$.

The matrix $\mathbf{D}$ is called the dictionary. Every component of signal $\boldsymbol{\psi}$ should be represented as linear combination of respective components of all the atoms. Therefore, i^{th} component of signal $\boldsymbol{\psi}$ would be represented as,

$$\psi_i = \beta_1\alpha_{i1} + \beta_2\alpha_{i2} + \cdots + \beta_N\alpha_{iN} \tag{4.18}$$

where, $\beta_1, \beta_2, \ldots, \beta_N$ are the scalar coefficients to be determined and $i = 1, 2, \ldots, M$. Putting all possible values of i, every component of $\boldsymbol{\psi}$ can be represented as linear combination of respective components of atoms. If representations of all the components are gathered together, then a system of linear equations is obtained, as shown below.

$$\begin{aligned}
\psi_1 &= \beta_1\alpha_{11} + \beta_2\alpha_{12} + \cdots + \beta_N\alpha_{1N} \\
\psi_2 &= \beta_1\alpha_{21} + \beta_2\alpha_{22} + \cdots + \beta_N\alpha_{2N} \\
\psi_3 &= \beta_1\alpha_{31} + \beta_2\alpha_{32} + \cdots + \beta_N\alpha_{3N} \\
&\vdots \\
\psi_M &= \beta_1\alpha_{M1} + \beta_2\alpha_{M2} + \cdots + \beta_N\alpha_{MN}
\end{aligned} \tag{4.19a}$$

Or vectorially, $\boldsymbol{\psi} = \beta_1\breve{\boldsymbol{\alpha}}_{\mathbf{1}} + \beta_2\breve{\boldsymbol{\alpha}}_{\mathbf{2}} + \cdots + \beta_N\breve{\boldsymbol{\alpha}}_{\mathbf{N}}$ (4.19b)

The linear representation of signal $\boldsymbol{\psi}$ by system of linear equations (4.19a) can be alternatively denoted by the equivalent matrix equation,

$$\boldsymbol{\psi} = \mathbf{D}\boldsymbol{\beta} \tag{4.20}$$

$$\text{where, } \boldsymbol{\beta} = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_N \end{bmatrix}^T$$

Vector $\boldsymbol{\beta}$ is made up of scalar coefficients that linearly encode the components of atoms into reconstruction of respective components of signal $\boldsymbol{\psi}$ and is called the coding vector. All the components of vector $\boldsymbol{\beta}$ simultaneously satisfy the system of equations (4.19a). Therefore, the unknown scalar coefficients can be determined by solving equation (4.19) or (4.20).

Now, suppose that there are P number of such M-dimensional signals $\breve{\boldsymbol{\psi}}_{\mathbf{1}}, \breve{\boldsymbol{\psi}}_{\mathbf{2}}, \ldots, \breve{\boldsymbol{\psi}}_{\mathbf{P}}$ which are to be reconstructed from the same N number of M-dimensional atoms. Let, $\breve{\boldsymbol{\psi}}_{\mathbf{k}}$ denote k^{th} signal and ψ_{ik} denote i^{th} component of k^{th} signal. The components of the signals can be arranged columnwise to form the matrix $\boldsymbol{\Psi}$.

$$\therefore \boldsymbol{\Psi} = \begin{bmatrix} \breve{\boldsymbol{\psi}}_{\mathbf{1}} & \breve{\boldsymbol{\psi}}_{\mathbf{2}} & \cdots & \breve{\boldsymbol{\psi}}_{\mathbf{P}} \end{bmatrix} = \begin{bmatrix} \psi_{11} & \psi_{12} & \cdots & \psi_{1P} \\ \psi_{21} & \psi_{22} & \cdots & \psi_{2P} \\ \vdots & & & \\ \psi_{M1} & \psi_{M2} & \cdots & \psi_{MP} \end{bmatrix} \tag{4.21}$$

$$\text{where, } \breve{\boldsymbol{\psi}}_{\mathbf{k}} = \begin{bmatrix} \psi_{1k} & \psi_{2k} & \cdots & \psi_{Mk} \end{bmatrix}^T \text{ and } k = 1, 2, \ldots, P.$$

If alike components of the signals are represented as linear combinations of alike components of atoms, then like the previous case, a system of linear equations will be obtained. Therefore, similar to vectorial representation of equation (4.19b), the k^{th} signal will be given by,

$$\breve{\boldsymbol{\psi}}_{\mathbf{k}} = \beta_{1k}\breve{\boldsymbol{\alpha}}_{\mathbf{1}} + \beta_{2k}\breve{\boldsymbol{\alpha}}_{\mathbf{2}} + \cdots + \beta_{Nk}\breve{\boldsymbol{\alpha}}_{\mathbf{N}} \tag{4.22}$$

Considering all possible values of k, system of linear equations corresponding to linear representation of all the signals can be compactly denoted by the matrix equation,

$$\boldsymbol{\Psi} = \mathbf{DB} \tag{4.23}$$

Here, matrix $\mathbf{D}$ is the dictionary as before. The matrix $\mathbf{B} = [\breve{\boldsymbol{\beta}}_1 \quad \breve{\boldsymbol{\beta}}_2 \quad \cdots \quad \breve{\boldsymbol{\beta}}_{\mathbf{P}}]$ represents the coding matrix whose columns are individual coding vectors. The k^{th} coding vector is given by, $\breve{\boldsymbol{\beta}}_{\mathbf{k}} = [\beta_{1k} \quad \beta_{2k} \quad \cdots \quad \beta_{Nk}]^T$. Clearly, for reconstruction of P number of signals, equal number of coding vectors will be required. Each coding vector is uniquely associated with the respective signal. For instance, coding vector $\breve{\boldsymbol{\beta}}_{\mathbf{1}}$ is uniquely associated with signal $\breve{\boldsymbol{\psi}}_{\mathbf{1}}$ and so on. Note that for the compatibility, the dimensions of the signals to be reconstructed and the atoms must be same. As matrix equation (4.23) essentially denotes a set of linear equations, similar to equation (4.19b), the solution to this equation will be the matrix $\mathbf{B}$.

Clearly, if the number of atoms available is much greater than dimension of signals to be reconstructed i.e. $N >> M$, then systems corresponding to equations (4.19) or (4.23) will be highly underdetermined. Hence, infinite number of solutions to these equations are possible and accordingly infinite number of linear combinations are possible. But among infinite possibilities, the sparsest solution is selected, such that the signals are represented by most minimum subset of atoms. Hence for equation (4.19), the optimum solution $\hat{\boldsymbol{\beta}}$ will be the solution which has the lowest L_0-norm. The associated L_0-minimisation problem can be formulated as,

$$\hat{\boldsymbol{\beta}} = \min \|\boldsymbol{\beta}\|_0 \text{ subject to } \mathbf{D}\boldsymbol{\beta} = \boldsymbol{\psi} \tag{4.24}$$

But as solving this optimisation problem is NP hard, the problem is approximated as an L_1-minimisation problem as given below.

$$\hat{\boldsymbol{\beta}} = \min \|\boldsymbol{\beta}\|_1 \text{ subject to } \mathbf{D}\boldsymbol{\beta} = \boldsymbol{\psi} \tag{4.25}$$

Optimum solution to equation (4.23) can be similarly obtained. The technique of acquiring and reconstructing a signal by solving an underdetermined system of linear equations is commonly known as Compressive Sampling or Compressed Sensing [86].

In the context of signal processing, signal reconstruction and discrimination both are equally important. Difference between these two techniques have been widely reported and investigated in literature. Signal reconstruction methods aim to formulate a representation that can get rid of signal corruptions, such as noise, missing data etc., and effectively tries to recover every possible information from the corrupted signal through reconstruction. On the other hand, signal discrimination methods generally classify different signals into different classes assuming that the signals are ideal i.e. devoid of noise or missing data. Such assumptions seems unrealistic because in reality signals can be corrupted i.e. there can be noise in the signal or signal may contain missing data. This calls for formulation of new methodology that can combine advantages of both reconstruction and discrimination methods. The classification techniques that can reconstruct and discriminate corrupted signals are generally known as Robust Classification [87].

Sparse representation is very well suited for robust classification because of its inherent theoretical structure. A small dense noise can be incorporated in the problem formulation by changing the basic equation (4.19) into following form,

$$\boldsymbol{\psi} = \mathbf{D}\boldsymbol{\beta} + \varepsilon \tag{4.26}$$

Here ε denotes the white Gaussian noise. However, sparse representation may not always perform optimally to accomplish all three objectives - sparsity, reconstruction and discrimination [87]. Hence, tradeoff exists among these objectives, which can be adjusted by introducing a suitable weighing factor in the problem formulation. If parameter λ denotes the tradeoff factor, then the L_1-minimisation problem for reconstruction and discrimination of corrupted signals can be given as,

$$\hat{\boldsymbol{\beta}} = \min \| \boldsymbol{\beta} \|_1 \text{ subject to } \| \mathbf{D}\boldsymbol{\beta} - \boldsymbol{\psi} \|_2^2 + \lambda \| \boldsymbol{\beta} \|_1 \leq \varepsilon \tag{4.27}$$

4.3.4 CLASSIFICATION BASED ON SPARSE REPRESENTATION

The concept of reconstructing a signal from sparsely populated vectors made up of atoms can be extended to classification problem. As already discussed, in a classification problem, observations and class memberships are available as the training data set. Observations in the training data set are termed as the training samples. Now in the classification through sparse representation, the basic assumption is that that the test samples can be represented as linear combination of training samples i.e. test samples lie in the linear subspace of training samples [88]. The aim will be to find out the sparsest linear combination.

4.3.4.1 TEST SAMPLES REPRESENTED AS SPARSE LINEAR COMBINATION OF TRAINING SAMPLES

Let us assume that in a training data set there are N numbers of classes and K number of M -dimensional training samples in each class. Also assume that there are P number of M -dimensional test samples in the corresponding testing data set.

The training samples are arranged in the columns of matrix **Train**. This matrix is analogous to the dictionary for signal reconstruction. If, $\breve{\mathbf{Tr}}_{\mathbf{j}}$ denotes training samples corresponding to j^{th} class and Tr_{ip_j} denotes i^{th} component of p^{th} training sample belonging to j^{th} class, then the matrix **Train** will be given by,

$$\mathbf{Train} = [\breve{\mathbf{Tr}}_{\mathbf{1}} \quad \breve{\mathbf{Tr}}_{\mathbf{2}} \quad \cdots \quad \breve{\mathbf{Tr}}_{\mathbf{N}}] \text{ where, } \breve{\mathbf{Tr}}_{\mathbf{j}} = \begin{bmatrix} Tr_{11j} & Tr_{12j} & \cdots & Tr_{1Kj} \\ Tr_{21j} & Tr_{22j} & \cdots & Tr_{2Kj} \\ \vdots & & & \\ Tr_{M1j} & Tr_{M2j} & \cdots & Tr_{MKj} \end{bmatrix} \tag{4.28}$$

$$i = 1, 2, \ldots, M \ , \ j = 1, 2, \ldots, N \text{ and } p = 1, 2, \ldots, K$$

Similarly, the test samples are arranged in the columns of matrix **Test**. If, $\breve{\mathbf{Te}}_{\mathbf{k}}$ denotes k^{th} test sample and Te_{ik} denotes i^{th} component of k^{th} test sample then,

$$\mathbf{Test} = [\breve{\mathbf{Te}}_{\mathbf{1}} \quad \breve{\mathbf{Te}}_{\mathbf{2}} \quad \cdots \quad \breve{\mathbf{Te}}_{\mathbf{P}}] = \begin{bmatrix} Te_{11} & Te_{12} & \cdots & Te_{1P} \\ Te_{21} & Te_{22} & \cdots & Te_{2P} \\ \vdots & & & \\ Te_{M1} & Te_{M2} & \cdots & Te_{MP} \end{bmatrix} \tag{4.29}$$

$$\text{where, } \breve{\mathbf{Te}}_{\mathbf{k}} = [Te_{1k} \quad Te_{2k} \quad \cdots \quad Te_{Mk}]^T \text{ and } k = 1, 2, \ldots, P \, .$$

Now the aim in the classification through sparse representation is to represent a test sample as the linear combination of training samples. Therefore, p^{th} test sample will belong to j^{th} class if its i^{th} component is characterised by linear combination of i^{th} components of all the training samples which belong to the same class.

$$\therefore \ Te_{ik} = Tr_{i1_j}\sigma_{1i_j} + Tr_{i2_j}\sigma_{2i_j} + \cdots + Tr_{iK_j}\sigma_{Ki_j} \tag{4.30}$$

The scalar coefficients $\sigma_{1i_j}, \sigma_{2i_j}, \ldots, \sigma_{Ki_j}$ are associated with training samples belonging to the j^{th} class only. All the components of all the test samples in **Test** can be similarly represented. But class memberships of the test samples are initially unknown. Hence, test samples should be characterised by all the training samples from all the classes in the training data set.

If all the components of all the test samples are linearly represented then, similar to formulation of equation (4.19a), a system of linear equations will be obtained which can be comprehensively denoted by the lower given matrix equation.

$$\mathbf{Test} = \mathbf{Train} \times \mathbf{S} \tag{4.31}$$

The matrix $\mathbf{S}$ is called the Sparsity matrix and each row of this matrix is a coefficient vector made up of scalar coefficients. Sparsity matrix and the scalar coefficients can be obtained by solving equation (4.31). If number of training samples available is much more than the number of test samples to be classified, then the system of linear equations denoted by equation (4.31) will be highly underdetermined. In that case, the sparsest solution $\hat{\mathbf{S}}$ will be obtained by L_0-minimisation or alternatively L_1-minimisation if the solution is sparse enough [81]. Recall that the sparsest solution comprises fewest non-zero elements. Ideally, the sparsest solution should be given by,

$$\hat{\mathbf{S}} = \begin{bmatrix} \breve{\boldsymbol{\sigma}}_{\mathbf{1}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \breve{\boldsymbol{\sigma}}_{\mathbf{2}} & \cdots & \mathbf{0} \\ \vdots & & & \\ \mathbf{0} & \mathbf{0} & \cdots & \breve{\boldsymbol{\sigma}}_{\mathbf{N}} \end{bmatrix} \text{ where, } \breve{\boldsymbol{\sigma}}_{\mathbf{i}} = \begin{bmatrix} \sigma_{11i} & \sigma_{12i} & \cdots & \sigma_{1Mi} \\ \sigma_{21i} & \sigma_{22i} & \cdots & \sigma_{2Mi} \\ \vdots & & & \\ \sigma_{K1i} & \sigma_{K2i} & & \sigma_{KMi} \end{bmatrix} \tag{4.32}$$

Thus in $\hat{\mathbf{S}}$, entries of coefficient vector should be zero except those associated with a certain object class [88]. This is because test samples of a certain class lie in the linear subspace of only those training samples which belong to same class.

4.3.4.2 RESIDUAL BASED CLASSIFICATION OF TEST SAMPLES

The fact that the test samples of a certain class are exclusively characterised by the training samples of same class, can be utilised to classify new test samples. Given a new test sample $\mathbf{Te}$ belonging to a certain class, first the sparsest sparsity matrix $\hat{\mathbf{S}}$ is computed from equation (4.31) by L_1-minimisation. Ideally, the non-zero entries in the $\hat{\mathbf{S}}$ will be associated with a single object class and test sample $\mathbf{Te}$ can be easily assigned to that class. However, as the L_0-minimisation problem has been approximated as an L_1-minimisation problem, there can be small non-zero elements present in the rows of $\hat{\mathbf{S}}$ which will jeopardise this method [88].

To get rid of this problem, let us define a function φ_j which identifies the coefficients in $\hat{\mathbf{S}}$ associated with the j^{th} class. Utilizing only the coefficients associated with j^{th} class the given test sample $\mathbf{Te}$ can be approximated as,

$$\tilde{\mathbf{Te}}_{\mathbf{j}} = \mathbf{Train} \times \varphi_j(\hat{\mathbf{S}}) \tag{4.33}$$

Corresponding to every class, the estimate $\tilde{\mathbf{Te}}_{\mathbf{j}}$ is computed. Next the differences between the test sample $\mathbf{Te}$ and the estimates $\tilde{\mathbf{Te}}_{\mathbf{j}}$ are obtained by calculating the residuals,

$$res_j(\mathbf{Te}) = \left\| \mathbf{Te} - \tilde{\mathbf{Te}}_{\mathbf{j}} \right\|_2 = \left\| \mathbf{Te} - \mathbf{Train} \times \varphi_j(\hat{\mathbf{S}}) \right\|_2 \tag{4.34}$$

The test sample belongs to j^{th} class if residual corresponding to that class given by equation (4.34) is minimum. The complete classification procedure by sparse representation has been summarised in the following algorithm.

Algorithm: Sparse Representation-based Classification

1. Input: Matrix pertaining to training data set, $\mathbf{Train} = [\breve{\mathbf{Tr}}_{\mathbf{1}} \quad \breve{\mathbf{Tr}}_{\mathbf{2}} \quad \cdots \quad \breve{\mathbf{Tr}}_{\mathbf{N}}]$, testing data set $\mathbf{Test} = [\breve{\mathbf{Te}}_{\mathbf{1}} \quad \breve{\mathbf{Te}}_{\mathbf{2}} \quad \cdots \quad \breve{\mathbf{Te}}_{\mathbf{P}}]$ (optional error tolerance $\varepsilon > 0$ and trade off factor λ). Mark all training samples with suitable labels.
2. Normalize columns of $\mathbf{Train}$ to obtain unit L_2-norm.
3. Solve L_1-minimisation problem:

   $\hat{\mathbf{S}} = \min \| \mathbf{S} \|_1$ subject to $\mathbf{Train} \times \mathbf{S} = \mathbf{Test}$

   (Or alternatively,

   $\hat{\mathbf{S}} = \min \| \mathbf{S} \|_1$ subject to $\| \mathbf{Train} \times \mathbf{S} - \mathbf{Test} \|_2^2 + \lambda \| \mathbf{Test} \|_1 \leq \varepsilon$)
4. Compute the residual for every test sample:

   $res_j(\breve{\mathbf{Te}}_{\mathbf{k}}) = \left\| \breve{\mathbf{Te}}_{\mathbf{k}} - \mathbf{Train} \times \varphi_j(\hat{\mathbf{S}}) \right\|_2$ for $j = 1, 2, \ldots, N$ and $k = 1, 2, \ldots, P$.
5. Identify and assign the classes to test samples by finding minimum value of residuals, $res_j(\breve{\mathbf{Te}}_{\mathbf{k}})$ corresponding to all the classes.

CHAPTER 5

TIME OF ARRIVAL AND LEVEL OF AMPLITUDE BASED CLASSIFICATION OF ACOUSTIC PD SIGNALS

CHAPTER 5.

TIME OF ARRIVAL AND LEVEL OF AMPLITUDE BASED CLASSIFICATION OF ACOUSTIC PD SIGNALS

5.1 INTRODUCTION

In the present work, the locations of the acoustic PD source inside PDSS box have been identified based on two parameters, one is sequence of arrival of PD generated acoustic waves at different acoustic sensors mounted on the outside walls of the PDSS box and another is levels of peak amplitudes of the captured acoustic signals by those sensors. The theoretical background behind the identification method based on these parameters has been discussed in this chapter.

5.2 TIME OF ARRIVAL AND PROPAGATION DELAY

It has been discussed in Chapter 2 that the acoustic PD source, employed in the present work, has different assembly than its optical counterpart. The acoustic PD source comprises point-plane electrode system and an acrylic disc sandwiched between the electrodes, as shown in Figure 2.3 of Chapter 2. Moreover, for acoustic PD data acquisition, the PDSS box has been filled with transformer oil. Therefore, when potential applied to the HV electrode of acoustic PD source is sufficiently high, the electric stress at the vicinity of interface between top electrode and acrylic disc becomes very high. The oil present at the vicinity of that interface partially breaks down due to this high electric stress and subsequently rapid energy release takes place from that interface. This energy travels in the form of elastic wave through the oil in all directions. These transient elastic waves are generally called the acoustic emissions. The group of oscillations constituting an acoustic signal is called an acoustic emission burst.

As per the IEEE Standard C57.127-2007 [15], the frequency of acoustic waves generated due to Partial Discharge lies in the range 20 kHz to 500 kHz. In addition to that, the generated acoustic waves normally have large bandwidth (Typically more than 1 MHz). The acoustic sensors employed for detection of acoustic signals from PD source should be sensitive to that aforementioned range of frequency and preferably should be an ultrasonic piezoelectric transducer. In the present work, such type of acoustic sensors have been used.

5.2.1 TYPES OF ACOUSTIC EMISSIONS AND CRITICAL ANGLE

For the given assembly of acoustic PD source and the steel made PDSS box, the acoustic emissions from the PD source give rise to two types of elastic waves:

- **Longitudinal Wave or Pressure Wave** – This type of elastic wave travels through the medium by creating alternate compression and rarefaction in the direction of propagation. Pressure waves exist both in transformer oil and wall of the PDSS box.
- **Transverse Wave or Shear Wave** – It is a kind of shock wave which is generated when the acoustic wave emanated from PD source directly impinges at normal incidence upon the box wall and creates vibrations inside the wall perpendicular to the propagation direction of the direct wave. Shear waves only exist in the box wall and not inside the oil because propagation of this type of wave requires acoustically solid material.

Another important quantity related to wave propagation and refraction is the critical angle. While travelling, if a wave reaches a boundary between two media having different refractive indices then the wave is partially refracted at the boundary surface and partially reflected. Due to refraction, the direction of wave travel inside the second medium changes depending on the refractive indices of two media. If the velocity of the wave is more (or refractive index is more) in the second medium than that was in the first medium then path of the wave bends towards the boundary. Critical angle is defined as the largest incidence angle that a wave can make with the normal to the boundary at the point of incidence while travelling from one medium to another medium such that the refracted wave in the other medium travels exactly parallel to the boundary. If the angle of incidence is more than the critical angle then the wave undergoes total internal reflection instead of refraction.

This critical angle is very important in the context of acoustic wave propagation inside the PDSS box and subsequent capture by acoustic sensors. The acoustic emissions that are generated from the acoustic PD source placed inside the box travel in all directions through oil to reach the walls of the box. Whether the acoustic waves get refracted into the walls or get

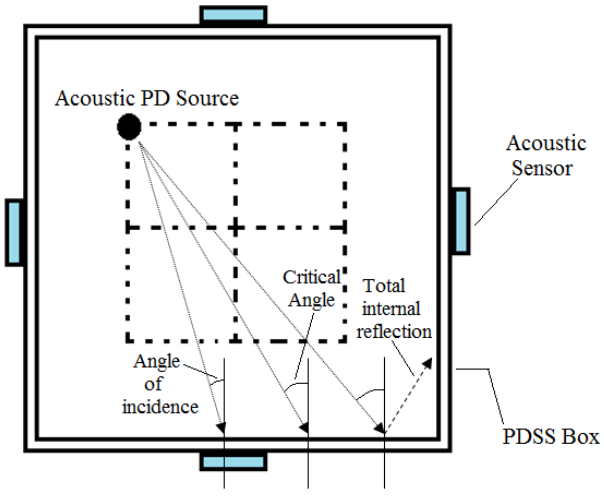


Figure 5.1 Critical Angle and Total Internal Reflection inside the PDSS box

reflected from the walls will depend upon the angle of incidence of the acoustic wave upon the inside wall. This angle of incidence primarily depends upon the position of acoustic PD source inside the PDSS box.

In the present scheme, the acoustic sensors have been placed at outside walls of the PDSS box. These sensors will be able to capture the acoustic signals only when the acoustic emissions from the PD source alight upon the inside wall of the PDSS box and penetrate the wall by refraction. Clearly, this can only happen if angle of incidence of the acoustic emissions is less than the critical angle. Otherwise the generated acoustic waves will unnecessarily get reflected multiple times inside the box which may diminish the amplitudes and increase the time of travel. This phenomenon has been schematically depicted in Figure 5.1. Therefore, the acoustic PD source should be placed such that the angle of incidence on the inside walls for every possible direction of acoustic emissions do not exceed the critical angle. As both the parameters, amplitude and time of travel, are important for identification of the location of the acoustic PD source inside the PDSS box, the placement of acoustic PD source inside the box becomes a crucial issue.

5.2.2 CHOICE FOR PLACEMENT OF ACOUSTIC SENSORS

Whether the acoustic sensors will be mounted on the inside or outside of the wall is a judicious issue. Sensors may be mounted on any side of the wall with advantages and disadvantages corresponding to each type of placement.

- **Outside Mounting** - Mounting the sensors outside the walls has various advantages. The positions of sensors can be reconfigured as necessary to get clearer acoustic signals. Moreover, the assembly of sensors can be moved from one equipment to another equipment or can be retrofitted to an existing equipment. However main disadvantage of externally mounted sensors is that they are more susceptible to noise.
- **Inside Mounting** - Inside mounting of sensor has advantages that the acoustic signals captured by the sensors will have higher amplitudes facilitating clearer measurement. The acoustic signals will be less noisy and signal-to-noise ratio will be high. Also problems associated with critical angle will no longer be present. However once installed, the sensors cannot be moved to get clearer signals. Also the system is not easy to replace and cannot be fitted to another equipment.

Clearly, external mounting of sensors has more advantages than internal mounting and it is general practice to mount the sensors outside the wall. Due to this reason in the present scheme, the acoustic sensors have been placed at outside walls of the PDSS box.

5.2.3 ACOUSTIC WAVE PROPAGATION

Propagation of acoustic waves inside the PDSS box and through the walls of the box is a complex phenomenon. It has been assumed in the present analysis that the acoustic wave travel at constant speeds inside the oil and inside the wall. Also the pressure and shear waves generated due to acoustic emissions have been assumed to be ideal.

The phenomenon of wave propagation inside and in the walls the PDSS box has been schematically illustrated in Figure 5.2. When the acoustic PD source is placed at the location shown in Figure 5.2 and PD is initiated, the acoustic waves are generated which travel in all directions through oil.

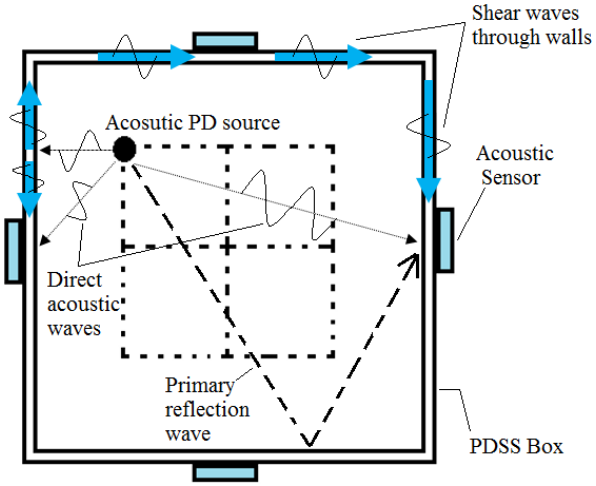


Figure 5.2 Wave propagation in PDSS box due to acoustic PD source

First, let us consider the acoustic waves from the PD location which reach the walls directly through oil. If angles of incidence of the direct waves upon a particular inside wall is less than critical angle, then the direct waves penetrate into that walls by refraction and reach the sensors mounted at other side of that wall. As acoustic emissions travel relatively slow, all the sensors do not experience simultaneous arrival of direct acoustic waves from the PD location. The distance travelled by the acoustic wave through the oil is dependent upon the speed of acoustic wave in the oil. The distance travelled can be simply given by,

Distance travelled by acoustic wave = Speed of the acoustic wave × Time of travel

The sensors which are at same distance from PD location, the direct acoustic wave travels to them in equal time. On the other hand, for the sensors which are at different distances the arrival time will be different. In Figure 5.2, the walls which are closest to the PD location, experiences earliest arrival of acoustic emissions. The acoustic sensors mounted on the closest walls will pick up the signals first. For other distant walls the arrival times are inversely proportional to their distance from the PD location.

Therefore depending upon the distance from the PD location, every sensor is associated with unique time of arrival of direct acoustic waves from a particular location. If location of acoustic PD source is changed, the arrival times are changed accordingly. The propagation delay of direct acoustic waves experienced by different sensors is an indication about the distance of these sensors from the PD location. Speed of acoustic waves in transformer oil is assumed to be 1413 m/s when measured at 20 ^{O}C [15].

Now let us turn attention to another type of propagation phenomenon. Upon reaching the closest wall, the acoustic emissions give rise to shear waves that travel through the walls only. Alongwith the direct acoustic waves travelling through oil, these shear waves also reach other sensors. Therefore, there are two paths of the acoustic waves through which they can reach the sensors from the PD locations. One is directly through the oil and one is through the walls. Speed of acoustic waves is more in the steel than in the transformer oil. Therefore the shear waves through the walls travel faster than the direct waves travelling through oil and may arrive at sensors before the direct wave do.

Typical nature of one acoustic emission burst from a PD source and recorded by an acoustic sensor has been shown in Figure 5.3. The sharp rise in amplitude up to positive peak marks the time instant when the direct acoustic wave arrives and impinges on the wall where the sensor is placed within the critical angle for pressure waves. This time instant associated with the positive peak is very important parameter to ascertain the distance of the sensor from the PD source.

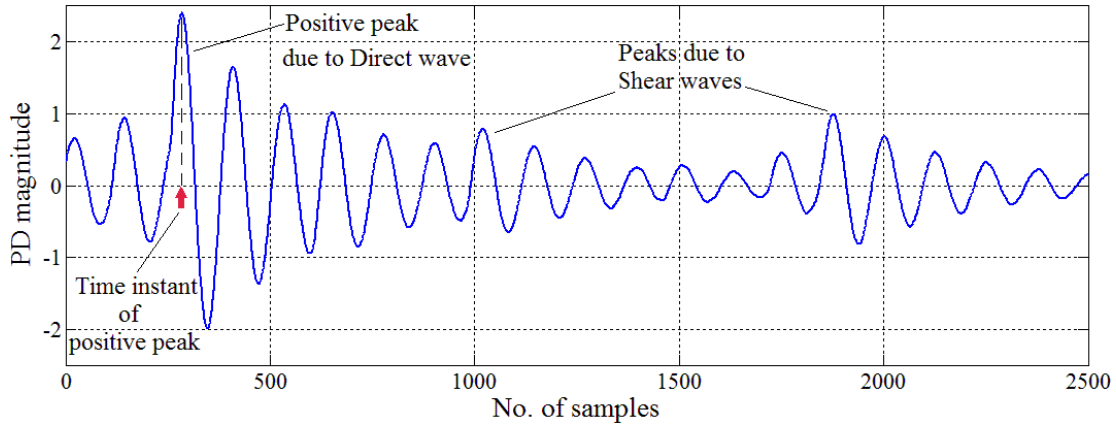


Figure 5.3 Example of one acoustic emission burst

After the occurrence of the positive peak, the wave undergoes damping due to energy loss and the amplitude gradually decreases through several cycles of oscillation. Ideally the amplitude should decay to very low value after several such cycles. But due to arrival of shear waves through the walls, the amplitude rises up again to reach another peak and subsequently decays down from that peak through damped oscillations.

Several such secondary peaks can occur due to a number shear wave arriving at different times. If sufficiently strong shear wave arrives a sensor before the direct wave, then two distinct positive peaks having comparable amplitudes may occur. In those cases it will be very hard to ascertain which peak has been created by which type of wave.

As the direct waves travel through oil, they undergo energy dissipation due to absorption in the oil. The peak value of direct wave, recorded by a sensor, is inversely proportional to average energy dissipation in oil during the travel of direct wave through the oil. Therefore, the less time direct wave spends within oil, the less will be energy dissipation and more will be the peak amplitude. In similar way, shear waves also undergo energy dissipation during their travel through steel walls. However, shear waves travel more path than the direct waves and consequently the shear waves experience more energy dissipation. Therefore, the peaks generated by shear waves are generally of low amplitude than those generated by direct waves.

In addition to these two types of waves, there may be additional acoustic waves due to primary and secondary reflection within the PDSS box. But amplitudes of these reflected waves are of very low value and can be considered negligible.

In practical situations, the oil may contain contaminations. Moreover, effects of temperature and humidity may be present. In such situations, transformer oil generally behaves as an anisotropic medium. Acoustic wave propagation through such anisotropic medium are not ideal because velocity of the wave in different directions may be different. Also roughness present in wall surface may induce microlevel reflection and refractions. The sum effect of these phenomena will be to create high frequency noise superimposed in the waveform of acoustic emission. These noises distort the acoustic signal.

5.2.4 IDENTIFICATION OF ACOUSTIC PD EVENTS BASED ON SEQUENCE OF ARRIVAL OF DIRECT WAVES

The fact that every sensor is associated with a unique arrival time can be utilised to identify the locations of the acoustic PD source. The time instant of occurrence of peak in the signal captured by a sensor denotes the time of arrival of direct wave at that sensor.

For a given location of acoustic PD source inside the PDSS box, five acoustic sensors placed on five walls of the box experience different arrival times. If sequences of arrival of direct waves are obtained for each location then this information can be utilised to identify any unknown location of the PD source.

The acoustic PD source was placed at one of the cubic sub-regions inside the PDSS box. For convenience the sub-regions and the placement of the sensors have been shown in Figure 5.4 and Figure 5.5 respectively. Now let us consider the sub-region '11' in the middle of block 1.

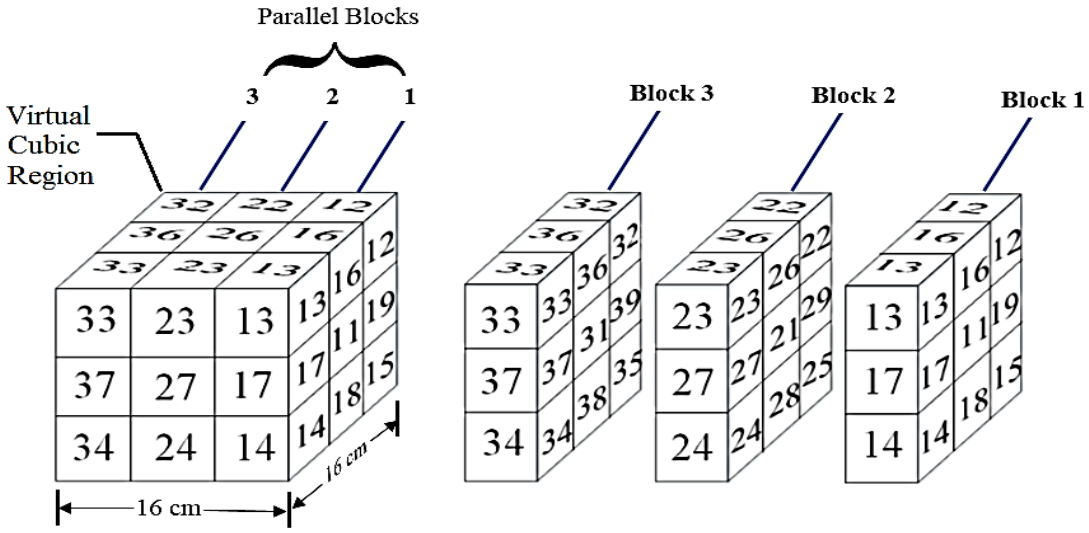


Figure 5.4 Imaginary cubic sub-regions in three parallel square blocks 1, 2 and 3 for placing the acoustic PD source inside the PDSS box

The waveforms of captured acoustic waves by five sensors when acoustic PD source is placed at '11' have been shown in Figure 5.6. From Figure 5.4 and Figure 5.5, it can be seen that this sub-region is closest to sensor AS1 which is placed outside the wall just in front of it. The sensor AS3 is farthest from this sub-region. The rest of the sensors AS2, AS4 and AS5 are equidistant from this sub-region. These sensors are farther from AS1 but are closer than AS3. Hence direct acoustic waves emanated from acoustic PD source will reach sensor AS1 first. Then the wave reaches simultaneously AS2, AS4 and AS5. Lastly the wave reaches sensor AS3. From Figure 5.6, this sequence of arrival can be readily ascertained by observing the time instants of occurrence of peaks. As direct wave reaches sensor AS1 earliest, the waveform of Figure 5.6(a) pertaining to signal captured by sensor AS1, shows earliest peak. In the waveforms of Figure 5.6(b), (d) and (e), pertaining to signals captured by sensors AS2, AS4 and AS5 respectively, the peaks occur at approximately same time but later than the waveform of Figure 5.6(a).

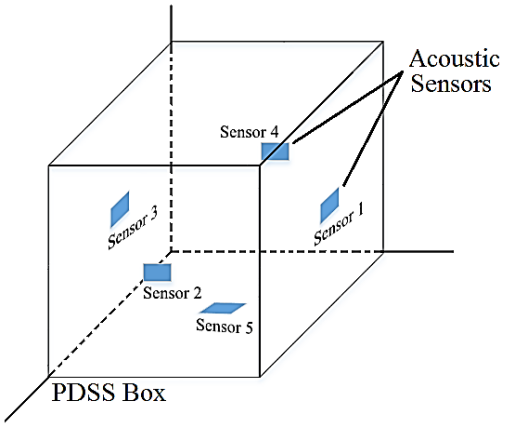


Figure 5.5 Placement of Acoustic Sensors in outside walls of PDSS Box

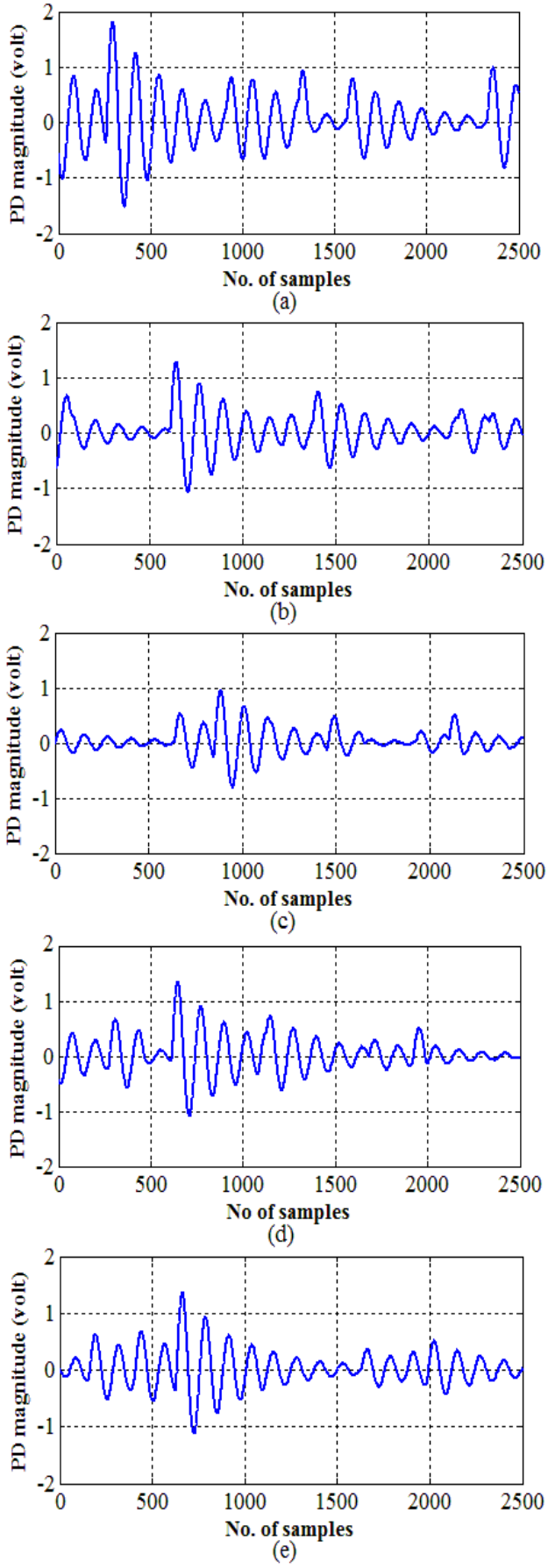


Figure 5.6 Acoustic signals captured by different sensors when acoustic PD source is placed at '11' a) Sensor, AS1 (b) Sensor, AS2 (c) Sensor, AS3 (d) Sensor, AS4 (e) Sensor, AS5.

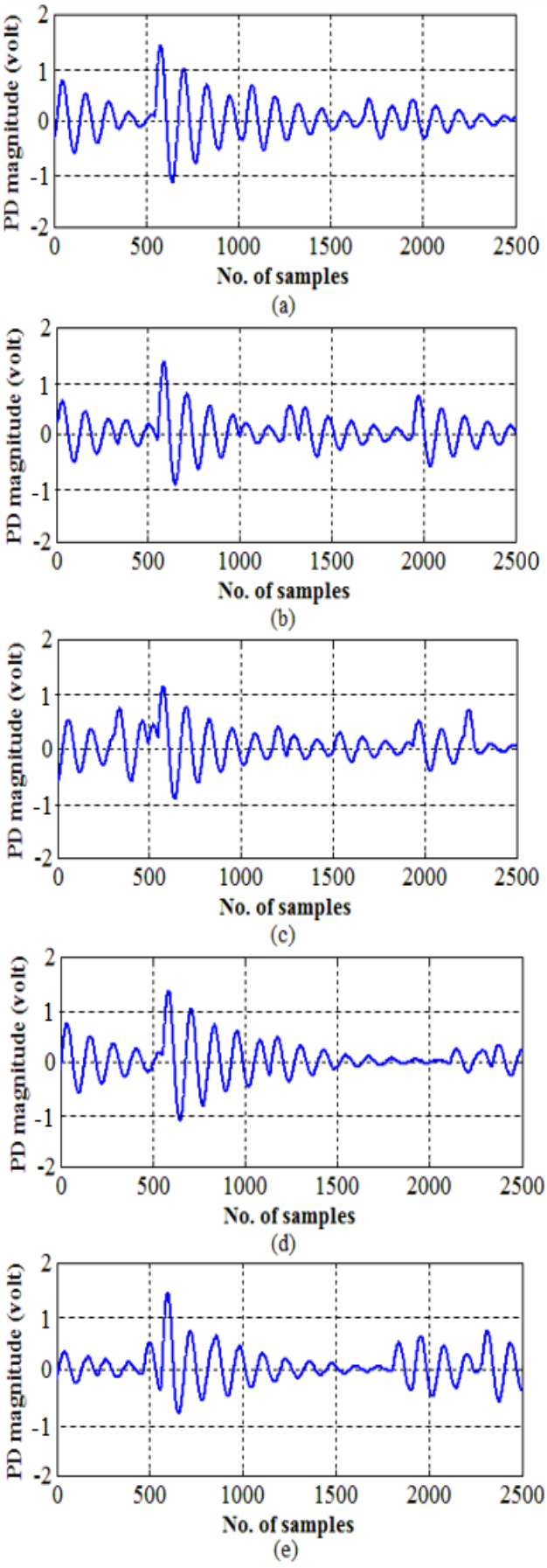


Figure 5.7 Acoustic signals captured by different sensors when acoustic PD source is placed at '21' a) Sensor, AS1 (b) Sensor, AS2 (c) Sensor, AS3 (d) Sensor, AS4 (e) Sensor, AS5.

This happens because sensors AS2, AS4 and AS6 are equidistant from location '11' but are farther than the sensor AS1. In the waveform of Figure 5.6(c), pertaining to signal captured by sensor AS3, the peak occurs later than all the other waveforms. This is because direct wave arrives that sensor at last.

This reasoning can be extended for every sub-region of Figure 5.4. For example, if we consider the sub-region '21', it can be seen from Figure 5.4 that this sub-region is equidistant from all the sensors. Therefore acoustic emissions from this PD location reach each of the sensors simultaneously. This can be clearly comprehended from the waveforms of Figure 5.7 by looking at the time instants of occurrences of peaks in captured acoustic signals.

The sequence of arrival of direct acoustic wave at five sensors corresponding to each PD location of Figure 5.4 can be considered as a qualitative feature uniquely associated with each location. The sequences of arrival associated with each location can be more systematically presented by utilising the notation scheme of Table 5.1.

Table 5.1 Notations to identify sequence of arrival of direct waves

Notation	Interpretation
I	Sensor is closest to the relevant PD location
II	Sensor is at moderate distance from the PD location
III	Sensor is farthest from PD the location

The sensors which are closest to PD location, direct wave arrives earliest at those sensors. Those sensors have been marked as 'I'. The sensors which are at moderate distances from the PD location experience the next arrival. Therefore those sensors have been marked as 'II'. The sensors which are farthest and experience last arrival of direct wave are marked as 'III'. Utilising this notation, sequences of arrival of direct wave at five sensors for all the PD events can be marked. The sequences of arrival for PD events '11' and '21' has been presented in Table 5.2

Table 5.2 Sequence of arrival for PD events '11' and '21'

PD Event	AS1	AS2	AS3	AS4	AS5
11	I	II	III	II	II
21	I	I	I	I	I

Identification of PD sub-regions based on unique time of arrival works excellently for single PD events i.e. when only one PD source is placed inside the PDSS box. But when double PD events are considered, the situation becomes quite complex.

When two PD sources are placed inside the PDSS box and made to discharge simultaneously, acoustic waves are generated from both the sources. The direct acoustic waves reach the sensors at different times which are at different distances from these sources. Therefore, the waves get unevenly mixed up by superposition during capture. In that mixed wave, the peaks cannot be marked clearly and there can be multiple peaks.

But if it happens that a sensor is equidistant from both the locations, then it can be fairly assumed that acoustic waves reach that sensor simultaneously. The acoustic waves from two PD sources get superimposed in phase with each other. Therefore peaks of both the waves become concurrent so that the peak in the mixed wave can be identified.

Let us take the example of double PD event '12_14' i.e. two acoustic PD sources have been placed at '12'and '14' and partial discharge is taking place simultaneously. The waveforms captured by five sensors have been depicted in Figure 5.8. It is evident from Figure 5.4 that both these sub-regions are individually equidistant from sensors AS1 and AS3. AS1 is the closest sensor and AS3 is the farthest sensor.

It can be seen in the waveforms of Figure 5.8 (a) and (c), pertaining to signals captured by sensors AS1 and AS3, that the peaks can be easily marked. This happens because for each of these sensors, the direct acoustic waves from the PD locations reach simultaneously and get superimposed in phase. Therefore the peaks of individual direct waves are approximately concurrent. But for other sensors, the waveforms are of poor quality and peaks cannot be easily marked which is because the rest of the sensors are at different distances from these PD locations.

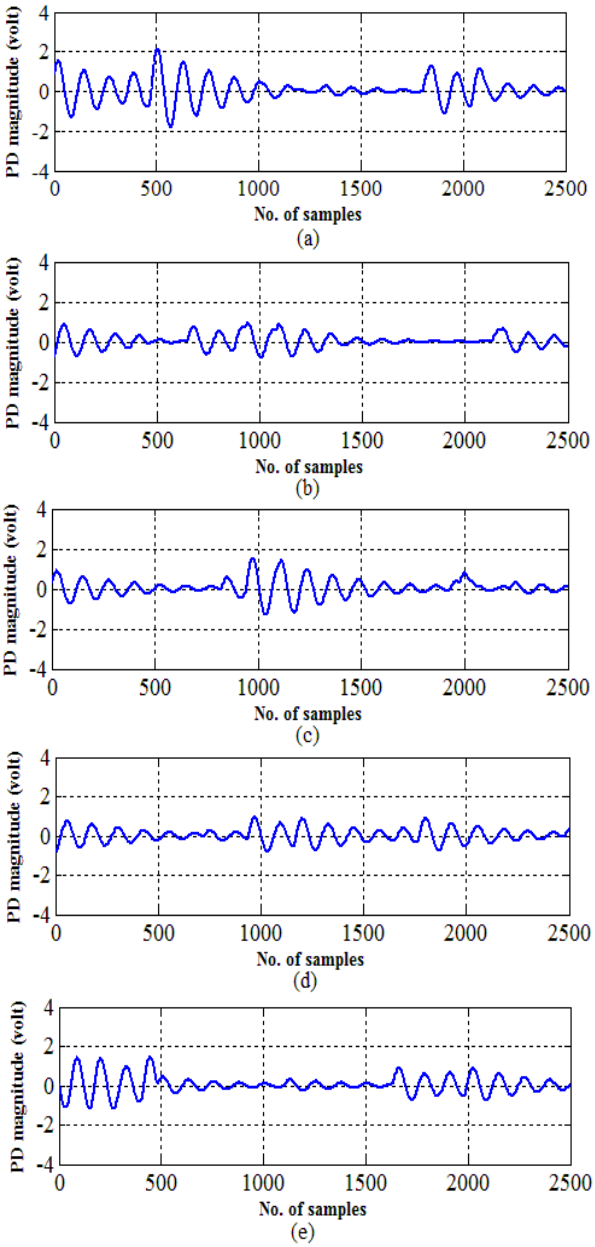


Figure 5.8 Acoustic signals captured by different sensors when two acoustic PD sources are placed at '12' and '14' a) Sensor, AS1 (b) Sensor, AS2 (c) Sensor, AS3 (d) Sensor, AS4 (e) Sensor, AS5.

Hence for DPD event ‘12_14’, the sequences of arrival of direct wave can only be ascertained for sensors AS1 and AS3. For rest of the sensors sequence cannot be obtained. It may so happen that sequence pertaining to ‘12_14’ matches with any other DPD event. For example consider the DPD event ‘13_15’. Referring to Figure 5.4 and Figure 5.5, it can be seen that for this event also the sensors AS1 and AS3 are individually equidistant from locations of PD sources while rest of the sensors are at different distances. In terms of notation introduced in Table 5.1, the sequence of arrival for DPD events ‘12_14’ and ‘13_15’ will be as given in Table 5.3. Note that for AS2, AS4 and AS5, the time of arrival cannot be evaluated and hence no sequence can be assigned.

Table 5.3 Sequence arrival for PD events ‘12_14' and ‘13_15’

PD Event	AS1	AS2	AS3	AS4	AS5
12_14	I	-	III	-	-
13_15	I	-	III	-	-

Therefore DPD events ‘12_14’ and ‘13_15’ are indistinguishable with respect to sequence of arrival. Hence, information of the sequences of arrival is sufficient for identification of SPD events but is not sufficient for identification of DPD events.

5.3 IDENTIFICATION OF ACOUSTIC PD EVENTS BASED ON PEAK AMPLITUDE LEVELS OF ACOUSTIC WAVES

The other parameter that is associated with the acoustic PD waves is the peak amplitude of acoustic signals. It has been already discussed that the peak amplitude of direct acoustic wave vary inversely with respect to the distance between PD location and the sensor. If sensor is more distant from the PD location then, due to energy dissipation in transit, the peak amplitude of the direct wave decreases. Hence the peak amplitude in the signal captured by a sensor is also an indication of the distance of this sensor from the PD location. The sensors which are closest to PD location capture signals comprising highest peaks. For the sensors which are farthest the reverse happen. Rest of the intermediated sensors capture signals whose peak amplitude varies within these maximum and minimum value.

For example, refer Figure 5.6. As sensor AS1 is closest to location ‘11’, the signal captured by AS1 has highest peak. Whereas, for signals captured by sensors S2, S4 and S5, the peak amplitudes are approximately same because these sensors are equidistant. But these peak amplitude are lower than that of the signal captured by AS1 because AS2, AS4 and AS5 are more distant than AS1. Lastly, the signal captured by AS3 has least

peak amplitude as it is farthest from the PD location ‘11’. Similarly, peak amplitudes in the signals of figure 5.7 are approximately of same value because all the sensors are equidistant from PD location ‘21’. In this way, by observing the relative differences in the peak values of signals captured by five sensors, every single PD location of Figure 5.4 can be uniquely characterised.

This reasoning can also be extended for DPD events. However, in DPD events there may be several local peaks of comparable amplitudes in the signals captured by some of the sensors. In those cases amplitude of the peak which has occurred earliest will only be considered. This is because that peak occurs due to arrival of direct wave from the PD source which is closest to that sensor. For example in Figure 5.8(d), pertaining to signal captured by sensor AS4, there are several local peaks which have comparable amplitudes. But the first peak corresponds to PD source placed at ‘12’ and amplitude of that peak will only be considered.

Therefore, peak amplitudes of acoustic signals, captured by five sensors for a SPD and DPD events is another unique numerical feature. Recall that for each PD event there are five data sets, each containing five data corresponding to five signals captured by five sensors. Peak values of all the signals captured by five sensors for all the PD events have been obtained.

As we are more concerned with the relative difference between peak values of captured signals pertaining to each single event, the peak values of all the signals for each acoustic sensor have been normalised with respect to maximum and minimum values in the range. Normalised peak values of signals captured by acoustic sensors for some of the SPD and DPD events have been shown in Table 5.4.

Table 5.4 Normalised peak values of signals captured by different acoustic sensors for SPD and DPD events

SPD Event	Normalised peak values of signals captured by different acoustic sensors				
	AS1	**AS2**	**AS3**	**AS4**	**AS5**
11	0.593	0.411	0.287	0.458	0.573
11	0.991	0.502	0.364	0.373	0.540
11	0.884	0.473	0.348	0.747	0.456
11	0.837	0.551	0.360	0.525	0.474
11	0.755	0.517	0.464	0.643	0.656
. . .	. . .	. . .	. . .	. . .	. . .
39	0.347	0.292	0.410	0.663	0.484
39	0.471	0.389	0.735	0.439	0.495
39	0.432	0.401	0.837	0.993	0.492
39	0.307	0.412	0.633	0.650	0.581
39	0.288	0.369	0.755	0.690	0.493
DPD Event	**Normalised peak values of signals captured by different acoustic sensors**				
	AS1	**AS2**	**AS3**	**AS4**	**AS5**
11_31	0.311	0.393	1.000	0.530	0.568
11_31	0.418	0.489	0.672	0.630	0.759
11_31	0.455	1.000	0.703	0.516	0.781
11_31	0.464	0.820	0.448	0.596	0.566
11_31	0.386	0.546	0.339	0.621	0.580
. . .	. . .	. . .	. . .	. . .	. . .
39_37	0.370	0.288	0.596	0.680	0.697
39_37	0.381	0.593	0.855	0.372	0.670
39_37	0.543	0.352	0.664	0.870	0.596
39_37	0.522	0.796	0.779	1.000	0.472
39_37	0.357	0.773	0.665	0.490	0.615

CHAPTER 6
RESULTS AND DISCUSSIONS

CHAPTER 6.

RESULTS AND DISCUSSIONS

6.1 INTRODUCTION

The main objective of the present work is to identify the type and location of single and multiple PD sources inside PDSS box based on the optical and acoustic PD signals emanated from these sources. For that purpose an optical PD source and an acoustic PD source was constructed. These sources were separately placed inside the PDSS box. The signals pertaining to each type of source were captured through the optical or acoustic sensors placed at the walls of the box. The placement of source inside the box has been as termed as events. The optical PD events have been identified by employing classification techniques, whereas the acoustic PD events have been identified based on sequence of arrival and levels of peak amplitudes. Results pertaining to identification of both types of single and multiple PD sources using these proposed methods have been discussed in this chapter.

6.2 IDENIFICATION OF OPTICAL PD EVENTS

The optical PD signals obtained through optical sensors and have been transformed by applying morphological operations. After that, suitable statistical features have been extracted from transformed optical PD signals. These features have been classified by two different techniques – Rough Set Theory and Sparse Representation Classification.

6.2.1 FEATURE CLASSIFICATION BY ROUGH SET THEORY

To classify the extracted features with the help of Rough Set Theory, first the information system has been constructed. This information system comprises of statistical features which have been obtained after applying morphological operations to the captured PD signals pertaining to each of the optical PD events. The main objective is to formulate decision rules utilizing these extracted features. Suppose optical PD source has been placed at one of the imaginary grid positions shown in Figure 2.4 of Chapter 2, which is unknown, and optical signals radiated from the optical PD source have been captured by five optical sensors. Now these formulated decision rules will identify that unknown location by analysing the features of all the captured optical signals. The constructed information system has been presented in Table 6.1.

Table 6.1 Normalised decision table pertaining to optical SPD and DPD events

Objects	Condition attributes										Decision attributes: Type and location of optical PD source
	F_{11}	F_{12}	F_{13}	F_{14}	F_{15}	F_{16}	F_{17}	F_{18}	F_{19}	… F_{59}	
1	0.988	1.000	0.023	1.000	0.549	0.993	0.995	0.974	0.000	… 0.055	**1A**
2	0.988	1.000	0.024	1.000	0.548	0.993	0.995	0.974	0.000	… 0.054	**1A**
3	0.989	1.000	0.024	1.000	0.549	0.993	0.994	0.974	0.000	… 0.055	**1A**
4	0.959	0.366	0.027	0.209	0.565	0.995	0.997	0.973	0.419	… 0.000	**2A**
5	0.959	0.367	0.028	0.210	0.560	0.995	0.997	0.973	0.418	… 0.000	**2A**
6	0.957	0.381	0.029	0.225	0.565	0.995	0.997	0.972	0.404	… 0.000	**2A**
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	… . . .	. . .
79	0.329	1.000	0.573	1.000	0.896	0.996	0.997	0.947	0.000	… 0.127	**9C**
80	0.316	1.000	0.588	1.000	0.904	0.996	0.997	0.947	0.000	… 0.124	**9C**
81	0.331	1.000	0.571	1.000	0.903	0.996	0.997	0.947	0.000	… 0.126	**9C**
82	0.989	0.981	0.024	0.974	0.552	0.993	0.994	0.976	0.037	… 0.136	**1A_1C**
83	0.988	0.980	0.023	0.973	0.550	0.993	0.994	0.976	0.038	… 0.138	**1A_1C**
84	0.987	0.982	0.024	0.976	0.549	0.993	0.995	0.975	0.036	… 0.131	**1A_1C**
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	… . . .	. . .
139	0.292	0.998	0.610	0.997	0.945	0.996	0.997	0.948	0.002	… 0.103	**9C_7C**
140	0.283	0.998	0.621	0.997	0.949	0.996	0.997	0.948	0.002	… 0.100	**9C_7C**
141	0.296	0.998	0.605	0.997	0.948	0.996	0.997	0.948	0.002	… 0.103	**9C_7C**

It is evident from the constructed information system of Table 6.1 that, the objects are information related to optical PD sources at different positions i.e. the objects correspond with different types of optical PD events. The condition attributes are features which were extracted from morphologically transformed optical PD signals and decision attributes are type and location of the optical PD source.

For each optical PD event, five data sets are available. Each data set contains five data pertaining to five signals captured by five sensors mounted on the walls of PDSS box. All the captured signals have been transformed by applying morphological operations,

Dilation and Erosion, and nine features have been extracted from each of the transformed signals. As there are five signals in each data set corresponding to five sensors, there will be in total 45 features in a single data set. For better manipulation, every feature has been assigned with suitable notation. The extracted features have been represented by notation F_{ij} which denotes ' j^{th} feature of the transformed optical signal which has been captured by i^{th} optical sensor', where $i = 1, 2, \ldots, 5$ and $j = 1, 2, \ldots, 9$. For example, F_{34} denotes 4^{th} feature (i.e. Kurtosis) of the transformed optical signal which has been captured by 3^{rd} optical sensor OS3 and so on. (Refer Table 3.3 of Chapter 3)

There are 27 single and 20 double optical PD events i.e. in total there are 47 optical PD events. Out of the five data sets corresponding to a single event, three data sets have been taken (60% of five data set) for construction of decision rules. As there are in total 47 optical PD events and there are five data sets corresponding to each event, there must be 235 (=47×5) objects. Out of these, 60% of the objects, i.e. 141 objects were used to develop the decision rules. Therefore, these 141 objects collectively construct the training data set. The rest 94 objects construct the testing data set.

In Table 6.1 for each event, normalised values of extracted features of signals captured by five sensors have been put in a single row. For example, normalised features F_{11} to F_{19} correspond to signal captured by Sensor OS1, normalised features F_{21} to F_{29} pertain to signal captured by Sensor OS2 and so on. Three data sets for each SPD and DPD event have been put in three consecutive rows (For example, first three rows of Table 6.1 are three sets of extracted and normalised features pertaining to SPD event '1A' etc.). As there are, in total, 47 SLPD & DLPD sources and 45 attributes, in our case the size of corresponding $(U \times Q)$ matrix will be (141×45). As all the conditional attributes are numerical type, each of the attributes has been individually normalized with respect to maximum value and minimum value in the range. Due to normalization, every attribute has numeric values confined within range 0 to 1. This is the benefit of normalization because all the numeric values are now properly referenced.

As discussed in Chapter 4, different information granules can be formed by choosing different subset of attributes P. For example in Table 6.1, if P = {Type and Location of Source} then $I_P = \{(1,2,3),(4,5,6),...,(139,140,141)\}$.

As all the attributes of the decision table are numerical, domain of each attribute has been individually discretized for increased flexibility.

In the present work, the discretization has been performed by employing Maximum Discernible (MD) Heuristics. More details of MD heuristics can be found in [73]

In addition to identification to equivalence classes and discretization, the decision table has been further processed by removing dispensable attributes and forming the Reduct comprising of indispensable attributes. The discretized decision table with Reducts have been presented in Table 6.2.

Table 6.2 Discritized decision table showing reducts

Object	**Condition attributes**											**Decision Attributes: Type and location of source**
	F_{11}	**F_{14}**	**F_{19}**	**F_{21}**	**F_{22}**	**F_{24}**	**F_{34}**	**F_{39}**	**F_{42}**	**F_{49}**	**F_{51}**	
1	9	6	0	1	8	9	8	0	7	0	1	**1A**
2	9	6	0	1	8	9	8	0	7	0	1	
3	9	6	0	1	8	9	8	0	7	0	1	
4	9	2	2	7	6	7	**5**	1	0	9	**2**	**2A**
5	9	2	2	7	6	7	**5**	1	0	9	**3**	
6	9	2	2	7	6	7	**6**	1	0	9	**2**	
7	9	1	3	**10**	4	**2**	7	0	5	1	**5**	**3A**
8	9	1	3	**9**	4	**3**	7	0	5	1	**4**	
9	9	1	3	**9**	4	**2**	7	0	5	1	**5**	
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
79	3	6	0	4	7	9	1	5	2	6	6	**9C**
80	3	6	0	4	7	9	1	5	2	6	6	
81	3	6	0	4	7	9	1	5	2	6	6	
82	9	6	0	3	7	**8**	1	5	7	1	3	**1A_1C**
83	9	6	0	3	7	**8**	1	5	7	1	3	
84	9	6	0	3	7	**9**	1	5	7	1	3	
85	9	6	0	1	8	9	7	1	**6**	2	2	**1A_2C**
86	9	6	0	1	8	9	7	1	**5**	2	2	
87	9	6	0	1	8	9	7	1	**6**	2	2	
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
139	3	6	0	**10**	3	3	1	7	3	5	6	**9C_7C**
140	3	6	0	**9**	3	3	1	7	3	5	6	
141	3	6	0	**10**	3	3	1	7	3	5	6	

In Table 6.2, 11 indispensable attributes have been shown to from the Reduct while other superfluous attributes have been discarded. These attributes are F_{11}, F_{14}, F_{19}, F_{21}, F_{22}, F_{24}, F_{34}, F_{39}, F_{42}, F_{49} and F_{51}.

From the Reducts, the Cores are to be formed. For illustration of construction of Cores, consider a portion of first six rows of discretized decision table, as shown in Table 6.3. If first three attributes constituting the subset $P = \{F_{11}, F_{14}, F_{19}\}$ is considered, it is seen that attribute F_{11} has the same value (= 9) corresponding to SPD events 1A, 2A, 3A and hence it is superfluous due to its omnipresence in all the objects. So this attribute can be dispensed with and considered as 'don't care'. Whereas attributes $\{F_{14}, F_{19}\}$ have different set of values for each of six objects. So these features can classify the objects in terms of type and location of source.

Table 6.3 Truncated discretized data table for illustration

Object	**Condition attribute**			**Decision attribute: Type and location of source**
	F_{11}	**F_{14}**	**F_{19}**	
1	9	6	0	**1A**
2	9	6	0	**1A**
3	9	6	0	**1A**
4	9	2	2	**2A**
5	9	2	2	**2A**
6	9	2	2	**2A**

Again referring to Table 6.3, if R = {Type and Location of Source} then, $I_R = \{(1,2,3),(4,5,6)\}$ and for $P = \{F_{11}, F_{14}, F_{19}\}$, $I_P = \{(1),(2),(3),...,(6)\}$. Now, $POS_P(R) = \{(1),(2),(3),...,(6)\}$ and, $POS_{\{P-\{F_{11}\}\}}(R) = \{(1),(2),(3),...,(6)\} = POS_P(R)$. Therefore, attribute F_{11} is R-dispensable in P. Similarly it can be shown that attributes F_{14} and F_{19} are R-indispensable in P. The reduced decision table, obtained from Table 6.3 after feature F_{11} is eliminated, has been shown in Table 6.4.

Table 6.4 Reduced data table obtained from Table 6.3

Object	**Condition attribute**			**Decision attribute: Type and location of source**
	F_{11}	**F_{14}**	**F_{19}**	
1	-	6	0	**1A**
2	-	6	0	**1A**
3	-	6	0	**1A**
4	-	2	2	**2A**
5	-	2	2	**2A**
6	-	2	2	**2A**

From this reduced data table, one can easily define the decision rules, such as-

If **F_{14} = 6** and **F_{19} = 0**, then type of PD source is **1A**.

If **F_{14} = 2** and **F_{19} = 2**, then type of PD source is **2A** and so on.

Of course these decision rules are only demonstrative and for the complete decision table the rules will be different. In this case, intersection of Reducts for each decision class will lead to null set and Reducts will be identical to Cores. But this may not the case for the entire table. The Reducts for each decision class were already obtained in Table 6.2. The Core values are obtained by intersecting the Reducts corresponding to each decision class, i.e. each optical PD event. The finalised form of complete data table showing the Cores for all the optical SPD and DPD Events has been shown in Table 6.5.

Table 6.5. Finalised form of complete data table showing the Cores for SLPD and DLPD Events

Object	**Condition attributes**											**Decision attributes: Type and location of source**
	F_{11}	**F_{14}**	**F_{19}**	**F_{21}**	**F_{22}**	**F_{24}**	**F_{34}**	**F_{39}**	**F_{42}**	**F_{49}**	**F_{51}**	
1	9	6	0	1	8	9	8	0	7	0	1	**1A**
2	9	6	0	1	8	9	8	0	7	0	1	
3	9	6	0	1	8	9	8	0	7	0	1	
4	9	2	2	7	6	7	-	1	0	9	-	**2A**
5	9	2	2	7	6	7	-	1	0	9	-	
6	9	2	2	7	6	7	-	1	0	9	-	
7	9	1	3	-	4	-	7	0	5	1	-	**3A**
8	9	1	3	-	4	-	7	0	5	1	-	
9	9	1	3	-	4	-	7	0	5	1	-	
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
79	3	6	0	4	7	9	1	5	2	6	6	**9C**
80	3	6	0	4	7	9	1	5	2	6	6	
81	3	6	0	4	7	9	1	5	2	6	6	
82	9	6	0	3	7	-	1	5	7	1	3	**1A_1C**
83	9	6	0	3	7	-	1	5	7	1	3	
84	9	6	0	3	7	-	1	5	7	1	3	
85	9	6	0	1	8	9	7	1	-	2	2	**1A_2C**
86	9	6	0	1	8	9	7	1	-	2	2	
87	9	6	0	1	8	9	7	1	-	2	2	
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
139	3	6	0	-	3	3	1	7	3	5	6	**9C_7C**
140	3	6	0	-	3	3	1	7	3	5	6	
141	3	6	0	-	3	3	1	7	3	5	6	

For any particular decision class, the attributes which does not help in classification and are superfluous, those have been considered 'don't care' and have been indicated as '-'. Take for example the three Reducts of decision class '2A' in Table 6.2. The attributes F_{34} and F_{51} have different values for this decision class and gives vague information about the optical PD event '2A'. The intersection of these Reducts results into dispensation of these attributes because these do not help in classifying PD optical PD event '2A'. So exempting these features the Core for decision class '2A' is obtained. In this way, by further elimination of other dispensable attributes within Reducts of each decision class, all the Cores are obtained from which decision rules have been formulated.

The formulated decision rules are in 'IF…THEN' format, as discussed before. The formulated decision rules have been shown in Table 6.6. In these decision rules, the operator '∧' denotes the logical AND operator. There are 94 data sets (40% of 235 data sets) used for testing the decision rules. The testing data sets have been normalised and discretized with respect to same references used in Table 6.2.

Table 6.6 Decision rules obtained from Table 6.5 to identify the type and location of PD source

Decision Rule No.	**Statement of the rule to identify the PD source**	
	IF	**THEN**
1	(F_{11} = 9 ∧ F_{14} = 6 ∧ F_{19} = 0 ∧ F_{21} = 1 ∧ F_{22} = 8 ∧ F_{24} = 9 ∧ F_{34} = 8 ∧ F_{39} = 0 ∧ F_{42} = 7 ∧ F_{49} = 0 ∧ F_{51} = 1)	**Type of source is 1A**
2	(F_{11} = 9 ∧ F_{14} = 2 ∧ F_{19} = 2 ∧ F_{21} = 7 ∧ F_{22} = 6 ∧ F_{24} = 7 ∧ F_{39} = 1 ∧ F_{42} = 0 ∧ F_{49} = 9)	**Type of source is 2A**
. . .	. . .	. . .
27	(F_{11} = 3 ∧ F_{14} = 6 ∧ F_{19} = 0 ∧ F_{21} = 4 ∧ F_{22} = 7 ∧ F_{24} = 9 ∧ F_{34} = 1 ∧ F_{39} = 5 ∧ F_{42} = 2 ∧ F_{49} = 6 ∧ F_{51} = 6)	**Type of source is 9A**
28	(F_{11} = 9 ∧ F_{14} = 6 ∧ F_{19} = 0 ∧ F_{21} = 3∧ F_{22} = 7 ∧ F_{34} = 1 ∧ F_{39} = 5 ∧ F_{42} = 7 ∧ F_{49} = 1 ∧ F_{51} = 3)	**Type of source is 1A_1C**
. . .	. . .	. . .
47	(F_{11} = 3 ∧ F_{14} = 6 ∧ F_{19} = 0 ∧ F_{22} = 3 ∧ F_{24} = 3 ∧ F_{34} = 1 ∧ F_{39} = 7 ∧ F_{42} = 3 ∧ F_{49} = 5 ∧ F_{51} = 6)	**Type of source is 9A_7C**

It has been observed that, the formulated decision rules have successfully identified 52 out of 54 different SPD events and 36 out of 40 different DPD events. So the classification accuracies for identifying single location and double location PD sources are approximately 96% and 90% respectively. All these information have been summarized in Table 6.7. It should be noted that the classification accuracy explicitly depends on the

Structuring Element chosen while extracting features by Mathematical Morphology and the discretization algorithm to assign class values on attributes. If these are changed or recalibrated then the accuracy can be improved further.

Table 6.7 Classification accuracies for identification of optical PD source by Rough Set Theory

Total No. of single PD source	**No. of successful classification**	**% Accuracy**
54	52	96
Total No. of double PD Source	**No. of successful classification**	**% Accuracy**
40	36	90

6.2.2 FEATURE CLASSIFICATION BY SPARSE REPRESENTATION CLASSIFIER

Classification of features by sparse representation classifier algorithm requires a different approach. It has been discussed in Chapter 4 that, for classification through sparse representation, the test samples in the testing data set are to be represented as sparse linear combination of training samples. The training samples of all the classes are arranged in the columns of matrix **Train**. Similarly, the test samples are arranged in the columns of another matrix **Test**. If number of training samples is more than number of testing samples, then the sparse linear combination of training samples are obtained by finding the sparsest solution to the matrix equation (6.1).

$$\mathbf{Test} = \mathbf{Train} \times \mathbf{S} \tag{6.1}$$

The sparsest solution is approximately obtained by employing L_1-minimisation technique.

Similar to the information system processed by RST, the constructed training data set comprises 60% of total optical PD events. There are five data sets for each of the optical PD event and each data set consists of five data pertaining to five signals captured by five sensors. Out of the five data sets, three data sets have been used for training. Additionally, there are 45 features related to a single data set (Nine extracted features from each of the five morphologically transformed optical PD signals). Three data sets, pertaining to a single optical PD event and comprising 45 features, have been arranged in three consecutive columns. In this way all the data sets have been arranged to form the training data set. The classes are the different optical PD events corresponding to the positions of optical source inside the PDSS box. Therefore, the training data set comprises of 47 classes (corresponding to 27 optical SPD events and 20 optical DPD events) and 3 training samples in each class. The training data set has been presented in the Table 6.8.

Table 7.8 Normalised training data set pertaining to sparse representation classification

Dimensions of training samples	**Classes of training samples corresponding to optical SPD and optical DPD events**																
	1A			**2A**			**…**	**9C**			**1A_1C**			**…**	**9C_7C**		
F_{11}	0.988	0.988	0.989	0.959	0.959	0.957	…	0.329	0.316	0.331	0.989	0.988	0.987	…	0.292	0.283	0.296
F_{12}	1	1	1	0.366	0.367	0.381	…	1	1	1	0.981	0.98	0.982	…	0.998	0.998	0.998
F_{13}	0.023	0.024	0.024	0.027	0.028	0.029	…	0.573	0.588	0.571	0.024	0.023	0.024	…	0.61	0.621	0.605
F_{14}	1	1	1	0.210	0.211	0.225	…	1	1	1	0.974	0.973	0.976	…	0.997	0.997	0.997
F_{15}	0.549	0.548	0.549	0.565	0.560	0.565	…	0.896	0.904	0.903	0.552	0.55	0.549	…	0.945	0.949	0.948
F_{16}	0.993	0.993	0.993	0.995	0.995	0.995	…	0.996	0.996	0.996	0.993	0.993	0.993	…	0.996	0.996	0.996
F_{17}	0.995	0.995	0.994	0.997	0.997	0.997	…	0.997	0.997	0.997	0.994	0.994	0.995	…	0.997	0.997	0.997
F_{18}	0.974	0.974	0.974	0.973	0.973	0.972	…	0.947	0.947	0.947	0.976	0.976	0.975	…	0.948	0.948	0.948
F_{19}	0	0	0	0.415	0.413	0.400	…	0	0	0	0.037	0.038	0.036	…	0.002	0.002	0.002
. . .	. . .	. . .	. . .	. . .	. . .	. . .	…	. . .	. . .	. . .	. . .	. . .	. . .	…	. . .	. . .	. . .
F_{55}	0.997	0.993	0.995	0.848	0.836	0.850	…	0.885	0.890	0.889	0.976	0.972	0.974	…	0.912	0.904	0.909
F_{56}	0.997	0.997	0.997	0.997	0.997	0.997	…	0.996	0.996	0.996	0.996	0.996	0.996	…	0.995	0.995	0.995
F_{57}	0.998	0.998	0.998	0.998	0.998	0.998	…	0.997	0.997	0.997	0.997	0.997	0.997	…	0.997	0.997	0.997
F_{58}	0.943	0.943	0.943	0.942	0.942	0.942	…	0.948	0.947	0.947	0.946	0.947	0.946	…	0.950	0.950	0.950
F_{59}	0.055	0.054	0.055	0.000	0.000	0.000	…	0.127	0.124	0.126	0.136	0.138	0.131	…	0.103	0.1	0.103

In Table 6.8, each column comprises 45 features related to one optical PD event. Total number of training samples are 141 and dimensionality of each training sample is 45. In Table 6.8, the notation F_{ij} denotes ' j^{th} feature of the transformed optical signal which has been captured by i^{th} optical sensor'. As there are in total 235 optical PD events, there will be 94 test samples in the testing data set. The sparsest solution to Equation (6.1) has been obtained by solving the L_1-minimisation problem of equation (6.2).

$$\hat{\mathbf{S}} = \min \| \mathbf{S} \|_1 \text{ subject to } \mathbf{Train} \times \mathbf{S} = \mathbf{Test} \tag{6.2}$$

In the present work, the L_1-norm has been minimised by employing Primal-dual algorithm for linear programming.

Unlike RST, the sparse representation classifier does not provide any decision rules. Instead it provides the sparsest solution $\hat{\mathbf{S}}$ to the Equation (6.1). Out of 94 test samples, any number of test samples can be arranged to form matrix **Test**. Sparsest solution $\hat{\mathbf{S}}$ will be different for different set of test samples. However, for calculation of classification accuracy, all the test samples were used.

Classes of the test samples have been ascertained by identifying the non-zero elements in the columns of matrix $\hat{\mathbf{S}}$ and then evaluating the estimates by multiplying such columns with matrix **Train**. The difference between the estimates and test samples were evaluated by calculating the residuals. The test samples have been assigned classes if for a particular class the residual is minimum.

It has been observed that, the sparse representation classifier algorithm has successfully identified '53' out of '54' different SPD events and '38' out of '40' different DPD events. So the classification accuracy for identifying single location and double location PD sources are approximately 98% and 95% respectively. All these information have been summarized in Table 6.9. Comparing with the classification accuracies provided by RST, as shown in Table 6.7, it can be inferred that sparse representation classifier provides better performance.

Table 6.9 Classification accuracies for identification of optical PD source by sparse representation classification

Total No. of single PD source	**No. of successful classification**	**% Accuracy**
54	53	98
Total No. of double PD Source	**No. of successful classification**	**% Accuracy**
40	38	95

6.3 IDENTIFICATION OF ACOUSTIC PD EVENTS

The acoustic PD events have been classified based on two parameters – sequence of arrival of direct waves at different sensors and levels of amplitudes of the acoustic signals captured by those sensors. A suitable notation scheme for identification of sequences of arrival has been introduced in Table 5.1. It has been again provided in Table 6.10 for convenience.

Table 6.10 Notations to identify sequence of arrival of direct waves

Notation	Interpretation
I	Sensor is closest to the relevant PD location
II	Sensor is at moderate distance from the PD location
III	Sensor is farthest from PD the location

Utilising this notation scheme, sequences of arrival for all the acoustic SPD and DPD events have been obtained.

In addition to sequences of arrival, the peak amplitudes of the captured acoustic signals have also been obtained. It has been already discussed in chapter 6 that peak amplitudes of signals captured by five acoustic sensors are uniquely associated with each acoustic PD event. The values of peak amplitudes can be though as numerical features related to each acoustic PD event. The normalised peak amplitudes for all the acoustic PD events have been already shown in Table 5.4 of Chapter 4.

Employing the concepts of Rough Set, all the numerical values pertaining to normalised peak amplitudes of all the captured acoustic signals have been discretized for increased flexibility. The ranges of numerical values and corresponding class values have been provided in Table 6.11.

Table 6.11 Different ranges of numerical values pertaining to peak amplitudes and corresponding assigned classes

Range of Numerical Value	**0 to 0.25**	**0.25 to 0.35**	**0.35 to 0.45**	**0.45 to 0.55**	**0.55 to 0.65**	**0.65 to 0.75**	**0.75 to 0.85**	**0.85 to 0.95**	**0.95 to 1.00**
Assigned Class	0	1	2	3	4	5	6	7	8

All the numerical values of Table 5.4 have been replaced with corresponding class values. The complete information about the sequences of arrival and levels of peak amplitudes corresponding to acoustic PD events have been provided in Table 6.12.

Table 6.12 Sequences of arrival and levels of peak amplitudes pertaining to different acoustic SPD and acoustic DPD events

Acoustic SPD Events	Sequences of arrival					Levels of peak amplitudes				
	AS1	AS2	AS3	AS4	AS5	AS1	AS2	AS3	AS4	AS5
11	I	II	III	II	II	4	2	1	3	4
12	I	II	II	I	II	3	1	1	4	2
13	I	I	II	II	II	3	4	1	1	2
14	I	I	II	II	I	3	3	1	2	5
15	I	II	II	I	I	3	1	1	5	4
16	I	II	III	II	III	3	4	1	2	2
17	I	I	III	III	II	4	4	1	2	3
18	I	II	III	II	I	4	2	1	3	6
19	I	III	III	I	II	6	1	6	3	6
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
37	III	I	I	III	II	1	4	2	0	4
38	III	II	I	II	I	1	2	5	2	4
39	III	III	I	I	I	1	1	8	5	2
Acoustic DPD Events	**Sequence of arrival**					**Level of peak amplitude**				
	AS1	**AS2**	**AS3**	**AS4**	**AS5**	**AS1**	**AS2**	**AS3**	**AS4**	**AS5**
11_31	-	I	-	I	I	1	2	8	3	4
11_32	-	-	-	-	-	1	4	1	4	6
11_37	-	-	-	-	-	4	8	1	3	4
31_12	-	-	-	-	-	2	1	4	4	2
31_17	-	-	-	-	-	8	3	1	1	3
12_13	I	-	II	-	II	3	4	1	3	2
12_14	I	-	II	-	-	5	4	2	1	4
13_15	I	-	II	-	-	5	3	2	1	6
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
33_35	II	-	I	-	-	1	3	3	5	3
35_34	II	-	I	-	-	1	1	4	4	5
39_37	II	-	I	-	II	3	1	4	4	5

Similar to optical PD events, there are five data sets corresponding to each acoustic PD event. Out of those five data sets, three data sets have been utilised for construction of Table 6.12. However it has been observed that the sequences of arrival and levels of peak amplitudes are identical for all three data sets. Due to this reason, sequences of arrival and levels of peak amplitudes for only one set of data has been shown in Table 6.12.

In order to identify an unknown acoustic PD event, first sequences of arrival of direct wave at different sensors and levels of amplitudes of the acoustic signals captured by those sensors are found out. If peaks in all the captured signals are easily distinguishable and are unique, then clearly the unknown event must be a single acoustic PD event. If peaks cannot be distinguished in some of the signals then PD event is most likely a DPD event. Then the numerical values are pertaining to peak amplitudes are assigned classes using the information given in Table 6.11. The unknown acoustic PD event can be easily identified by matching the sequences and levels with the rows of Table 6.12.

Out of the five data sets, the rest two data set has be used to validate identification accuracy. It has been observed that using information given in Table 6.12, 25 out of 27 acoustic SPD events and 17 out of 20 acoustic DPD events can be successfully identified in each of these sets.

CHAPTER 7

CONCLUSIONS

CHAPTER 7.

CONCLUSIONS

7.1 CONCLUSIONS

The present work is aimed at identification and localisation of single and multiple Partial Discharge sources based on optical and acoustic signals under simulated conditions. This work investigated the possibility of PD detection and measurement from outside of an emulated equipment enclosure by analysing the optical or acoustic signals captured by suitable sensors located at the inside and outside wall respectively.

For that purpose, an optical PD source and an acoustic PD source were fabricated. Additionally, a steel-made cubical box, called the Partial Discharge Source Simulator or PDSS box, was constructed which emulates the equipment enclosure. For capture of either type of signals from these sources, optical and acoustic sensors were mounted on the walls of this box. Each type of PD sources were separately placed inside the box at different strategic locations and data corresponding to each type of PD signal for all the locations were recorded by suitable sensors. Recording of PD signals for all the locations inside the PDSS box was termed as PD data acquisition. The placement of PD sources at a given location inside the PDSS box were termed as PD events. For emulation of single PD phenomena, only one type of PD source was placed inside the PDSS box and this event was called Single PD (SPD) event. Similarly to emulate multiple PDs, two such same type of PD sources were placed at different locations inside the box and this event was named as Double PD (DPD) Event.. The recorded optical signals and acoustic signals were analysed separately.

By analysing the recorded optical PD signals, it was observed that these signals are mostly sparse in nature. To fill up the sparse domains of the captured optical PD signals, these signals were transformed by employing Mathematical Morphology, a time domain based signals transformation tool. After that, suitable statistical features were extracted from the morphologically transformed optical PD signals. These extracted features were classified with two separate classification techniques – Rough Set Theory (RST) and Sparse Representation Classification. It has been observed that RST provides very good classification accuracy and can be considered as an effective tool for PD localisation.

The Sparse Representation performed better than the RST in terms of classification accuracy. Therefore Sparse Representation Classification can be considered as a better tool than RST.

The locations of acoustic PD source inside the cubical box were identified by adopting a completely different approach. The locations were recognised based on two parameters, sequence of arrival of PD generated acoustic waves at different acoustic sensors mounted on outside walls of the cubical box and levels of peak amplitudes of the captured acoustic signals by those sensors. The results show that this combined methodology can effectively identify the locations of single as well as double PD sources with very good accuracy.

7.2 FUTURE SCOPES

There are many aspects of the present work which can be improved further. In the present scheme, for emulation of multiple PDs, only two PD sources were placed inside PDSS box. A more number of PD sources can be used simultaneously for obtaining better results. Moreover, retuning of corresponding Structuring Element of MM may further improve the classification accuracy.

The optical method discussed in the present work is particularly useful for PD detection of power transformers, in which case the sensors will be mounted on the inside walls of the tank and they capture signals form PD sources occurring at the outskirts of the coils or assembly. For practical cases high sensitivity sensors are recommended because transformer oil may significantly decrease the intensity of optical signals emitted out by potential PD sources. The application of this method can be justified considering the higher classification accuracy and very low cost of the equipment involved.

It has been reported by various researches that electrical PD signals are more efficient in detecting PD discharges as compared to both optical and acoustic signals. It remains to be seen, whether the developed classifier would work with desired accuracy in case of electrical PD signals or a new algorithm is needed to be developed.

REFERENCES

REFERENCES

[1.] IEC Standard 60270 (Third edition, 2000), Partial Discharge Measurements. International Electrotechnical Commission (IEC), Geneva, Switzerland.

[2.] T. Seghier, D. Mahi, T. Lebey and D. Malec, "The effect of space charge on partial discharges inception voltage in air gaps within high density polyethylene".

[3.] E. Kuffel, W. S. Zaengl and J. Kuffel, High Voltage Engineering: Fundamentals, NEWNES, 2nd Edition, 2000.

[4.] M.T. Suwarno, H. Ichikawa, Y. Suzuoki, T. Mizutani and K. Uchida, "A Model for Electrical Treeing Discharge in Polyethelene", International Symposium on Electrical Insulating Materials, Tokyo, Japan, pp. 443 – 446, 1995.

[5.] K. Wu, H. K. Xie, Y. Suzuoki and T. Mizutani, "The Effects of Conductivity Distribution in Tree Channel on the Behavior of PD in Electrical Tree", International Symposium on Electrical Insulating Materials, pp. 701 – 704, 1998.

[6.] T. Seghier, D. Mahi, A. Nouar and K. Lefkaier, "The Effect of Temperature and the Mutual influence Between two Cavities on the Appearance of Partial Discharges in Gaseous Cavities Contained in the Insulator of High Voltage", Proceedings of IEEE International Conference on Solid Dielectrics, Toulouse, France, July 5 - 9, pp. 598 - 602, 2004.

[7.] T. Seghier, D. Mahi, T. Lebey and D. Malec "Analysis of the electric field and the potential distribution in cavities inside solid insulating electrical materials", International Comsol Conference, Paris, November 5 - 9, 2006.

[8.] T.Seghier, D. Mahi and F. M. Frigura "Effect of temperature and relative humidity on partial discharges activity in artificial air gap embedded in high density polyethylene", Conférence internationale des systèmes d'énergie, Romanie, 21 - 23 November, 2007.

[9.] P. Das & S. Chakravorti, "Studies on Partial Discharge Simulation Based on a Stochastic Model Considering the Variation of Discharge Area and Temperature of the Void Surface", International Journal for Computational Methods in Engineering Science and Mechanics, 10:5, pp. 393 - 405, 2009.

[10.] P. Das & S. Chakravorti, "Simulation of PD patterns due to a narrow void in different E-field distribution", Journal of Electrostatics.

[11.] R. Bartnikas, "Partial Discharges. Their Mechanism, Detection, and Measurement", IEEE Transactions on Dielectrics & Electrical Insulation, Volume: 9, pp. 763 - 808, 2002.

[12.] G. C. Stone, "Partial Discharge Diagnostics and Electrical Equipment Insulation Condition Assessment", IEEE Transactions on Dielectrics & Electrical Insulation, Volume: 12, pp. 891-904, 2005.

[13.] M. Muhr & R. Schwarz, "Partial discharge measurement as a Diagnostic Tool for HV-Equipments", IEEE International Conference on Properties & Applications of Dielectric Materials, pp. 195 – 198, 2006

[14.] M. S. Naidu & V. Kamaraju, High Voltage Engineering, McGraw-Hill, 4th Edition, 2009.

[15.] IEEE standard C57.127 – 2007, IEEE Guide for the Detection and Location of Acoustic Emissions from Partial Discharges in Oil-Immersed Power Transformers and Reactors.

[16.] B. Sarkar, C. Koley, N. K. Roy & P. Kumbhakar, "Low cost RF sensor for partial discharge detection of high voltage apparatus", IEEE International Conference on Condition Assessment Techniques in Electrical Systems (CATCON), pp. 259 - 264, 2013.

[17.] S. Meijer, R. A. Jongen, E. Gulski, P. P. Seitz, T. J. W. H. Hermans & L. Lamballais, "VHF Partial Discharge Detection during After-Laying testing of Power Cables", IEEE International Conference on Solid Dielectrics, 2007.

[18.] R. Giussani, I. Cotton & R. Sloan, "Detection of corona with RF methods and spectra analysis", IEEE International Symposium on Electrical Insulation (ISEI), pp. 132 – 136, 2012.

[19.] M. Kawada, "Ultra wide band VHF/UHF radio interferometer system for detecting partial discharge source", IEEE Power Engineering Society Winter Meeting, Volume: 2, pp. 1482 – 1487, 2002.

[20.] H. R. Mirzaei, A. Akbari, E. Gockenbach, M. Zanjani & K. Miralikhani, "A novel method for ultra-high-frequency partial discharge localization in power transformers using the particle swarm optimization algorithm", IEEE Electrical Insulation Magazine, Volume: 29, Issue: 2, pp. 26 – 39, 2013.

[21.] S. Zheng, C. Li, Z. Tang, W. Chang & M. He, "Location of PDs inside transformer windings using UHF methods", IEEE Transactions on Dielectrics and Electrical Insulation, Volume: 21, Issue: 1, pp. 386 – 393, 2014.

[22.] Z. Tang, C. Li, X. Cheng, W. Wang, Jinzhong Li & Jun Li, "Partial discharge location in power transformers using wideband RF detection", IEEE Transactions on Dielectrics and Electrical Insulation, Volume: 13, Issue: 6, pp. 1193 – 1199, 2006.

[23.] M. Judd, B. Hampton & W. Brown, "UHF partial discharge monitoring for 132 kV GIS", International Symposium on High Voltage Engineering, Montreal, August 1997.

[24.] P. G. Reddy & P. Kundu, "Detection and analysis of partial discharge using ultra high frequency sensor", Annual International Conference on Emerging Research Areas: Magnetics, Machines and Drives (AICERA/iCMMD), pp. 1 – 6, 2014.

[25.] W. Pfeiffer & R. Plessow, "The influence of the coupling capacitance and the frequency of the voltage stress with respect to a partial discharge measurement", IEEE Conference on Electrical Insulation and Dielectric Phenomena, IEEE 1997 Annual Report, Volume: 2, pp. 558 - 561, 1997.

[26.] M. Harbaji, K. Shaban & A. El-Hag, "Classification of common partial discharge types in oil-paper insulation system using acoustic signals", IEEE Transactions on Dielectrics and Electrical Insulation, Volume: 22, Issue: 3, pp. 1674 – 1683, 2015.

[27.] T. Czaszejko & J. Sookun, "Acoustic emission from partial discharges in solid dielectrics", IEEE Electrical Insulation Conference (EIC), pp. 119 – 123, 2014.

[28.] P. Kundu, N. K. Kishore & A. K. Sinha, "Frequency domain analysis of acoustic emission signals for classification of partial discharges", 2007 Annual Report - Conference on Electrical Insulation and Dielectric Phenomena, pp. 146 – 149, 2007.

[29.] T. Boczar & D. Zmarzly, "Application of wavelet analysis to acoustic emission pulses generated by partial discharges", IEEE Transactions on Dielectrics and Electrical Insulation, Volume: 11, Issue: 3, pp. 433 – 449, 2004.

[30.] T. Boczar, "Identification of a specific type of PD from acoustic emission frequency spectra", IEEE Transactions on Dielectrics and Electrical Insulation, Volume: 8, Issue: 4, pp. 598 – 606, 2001.

[31.] S. L. Jones, "The detection of partial discharges in power transformers using computer aided acoustic emission techniques", Conference Record of the 1990 IEEE International Symposium on Electrical Insulation, pp. 106 – 110, 1990.

[32.] S. L. Jones, "The detection of partial discharges in power transformers using computer aided acoustic emission techniques", Conference Record of the 1990 IEEE International Symposium on Electrical Insulation, pp. 106 – 110, 1990.

[33.] R. Meunier & G. H. Vaillancourt, "Propagation behaviour of acoustic partial discharge signals in oil-filled transformers", 12th International Conference on Conduction and Breakdown in Dielectric Liquids (ICDL), pp. 401 – 404, 1996.

[34.] A. O. Akumu, N. Kawaguchi, R. Ozaki, H. Ihori, M. Fujii & K. Arii , "A study of partial discharge acoustic signal propagation in a model transformer", Proceedings of International Symposium on Electrical Insulating Materials (ISEIM 2001), pp. 583 – 586, 2001.

[35.] S. Biswas, D. Dey, B. Chatterjee & S. Chakravorti, "An approach based on rough set theory for identification of single and multiple partial discharge source.", Elsevier International Journal of Electrical Power and Energy Systems, Volume: 46, pp. 163–174, March 2013.

[36.] S. Biswas, C. Koley, B. Chatterjee & S. Chakravorti, "A Methodology for Identification and Localization of Partial Discharge Sources using Optical Sensors", IEEE Transactions on Dielectrics & Electrical Insulation, Volume: 19, Issue: 1, February 2012.

[37.] N. A. Algeelani, M. Afendi, M. Piah, & S. M. Zafar Iqbal, "Optical detection and evaluation of partial discharge on glass insulator", IEEE International Conference on Power Engineering and Optimization (PEOCO), pp. 87 – 91, 2013.

[38.] K. Fujii, M. Yamada, A. Tanaka & K. Kurosawa, "Emission spectrum of partial discharge light in SF6 gas", Conference Record of the IEEE International Symposium on Electrical Insulation, pp. 332 - 335, 1992.

[39.] M. Duval, "A review of faults detectable by gas-in-oil analysis in transformers", IEEE Electrical Insulation Magazine, Volume: 18, No. 3, pp. 8-17, 2002.

[40.] M. R. Samsudin, A. Q. Ramli, A. Berhanuddin & Y. Zaidey, "Incipient fault detection in 33/11kV power transformers by using combined Dissolved Gas Analysis technique and acoustic partial discharge measurement and validated through untanking", Proceedings of the International Symposium on Modern Electric Power Systems (MEPS), pp. 1 – 6, 2010.

[41.] N. A. Muhamad, B. T. Phung, T. R. Blackburn & K. X. Lai , "Dissolved gas analysis of faults in biodegradable oil transformer insulating systems", International Conference on Condition Monitoring and Diagnosis (CMD 2008), pp. 663 – 666, 2008.

[42.] B. Nemeth, S. Laboncz & I. Kiss, "Condition monitoring of power transformers using DGA and Fuzzy logic", IEEE Electrical Insulation Conference, pp. 373 – 376, 2009.

[43.] A. D. Ashkezari, T. K. Saha & C. Ekanayake & H. Ma, "Evaluating the accuracy of different DGA techniques for improving the transformer oil quality interpretation", Australasian Universities Power Engineering Conference (AUPEC), pp. 1 – 6, 2011.

[44.] S. Chakravorti, D. Dey & B. Chatterjee, Recent Trends in the Condition Monitoring of Transformers, Springer, 2013.

[45.] E. Lemke, "Progress in PD Probe Measuring Technique", International Symposium on High Voltage Engineering (ISH), Dresden, Germany, 1991.

[46.] L. E. Lundgaard, "Partial discharge–Part XIV: Acoustic Partial Discharge Detection Practical Application", IEEE Electrical Insulation Magazine, Volume: 8, No. 1, pp. 34 - 43, 1992

[47.] L. Hao, P. L. Lewin & S. J. Dodd, "Extraction of PD signals from an electro-optic modulator based PD measurement system", Conference on Electrical Insulation and Dielectric Phenomena, Southampton, UK, 2006.

[48.] X. Wang, B. Li, H.T. Roman, O.L. Russo, K. Chin & K.R. Farmer, "Acousto-optical PD detection for transformers", IEEE Transactions on Power Delivery, Volume: 21, pp. 1068 – 1073, 2006.

[49.] F. Li, Y. Liu, L. Wang & Yu Chen, "Analysis of the coupling optical fiber ultrasonic sensor for partial discharges detection", IEEE International Conference on the Properties and Applications of Dielectric Materials (ICPADM), pp. 991 – 994, 2015.

[50.] S. Coenen & S. Tenbohlen, "Location of PD sources in power transformers by UHF and acoustic measurements", IEEE Transactions on Dielectrics and Electrical Insulation, Volume: 19, Issue: 6, pp. 1934 – 1940, 2012.

[51.] D. Dey, B. Chatterjee, S. Chakravorti & S. Munshi, "Cross-wavelet Transform as a new paradigm for noisy partial discharge pulses", IEEE Transactions on Dielectrics & Electrical Insulation, Volume: 17, pp. 157-166, 2001.

[52.] R. Badent, K. Kist, N. Lewald & A. J. Schwab, "Partial-discharge diagnosis with artificial neural networks", Proceedings of the International Conference on Properties and Applications of Dielectric Materials, Volume: 2, Pages: 638 - 641, 1994.

[53.] M. B. Ashtiani and S. M. Shahrtash, "Feature-oriented De-noising of Partial Discharge Signals Employing Mathematical Morphology Filter", IEEE Transactions on Dielectrics and Electrical Insulation, Volume: 19, No. 6, pp 2128 - 2136, December 2012.

[54.] M. Hikita, K. Yamanda, A. Nakamura & T. Mizutani, "Measurement of Partial Discharges by Computer and Analysis of Partial Discharge Distribution by Monte Carlo Method", IEEE Transactions on Electrical Insulation, Volume: 25, pp. 453 – 468, 1990.

[55.] Datasheet available at http://www.hamamatsu.com.

[56.] Ultrasonic Transducers

[57.] http://sine.ni.com/nips/cds/view/p/lang/en/nid/203720.

[58.] J. Serra, Image Analysis and Mathematical Morphology, New York: Academic, 1982.

[59.] Image Analysis and Mathematical Morphology, Volume 2: Theoretical Advances by Jean Serra

[60.] J. Serra & Ph. Salembier (Eds.), "Mathematical Morphology and its Application to Signal Processing", Proceedings of the International workshop on Mathematical Morphology and Its Applications to Signal Processing (ISMM), 1993.

[61.] G. Matheron, Random Sets and Integral Geometry, New York, Wiley, 1975.

[62.] P. P. Maragos & R. W. Schafer, "Morphological filters--Part I- Their set-theoretic analysis and relations to linear shift-invariant filters", IEEE Transactions on Acoustics, Speech and Signal Processing, Volume: 35, No. 8, pp. 1153 - 1169, August 1987.

[63.] P. Maragos & R. W. Schafer, "Morphological filters--Part II- Their Relation to Median, Order Statistic, and Stack Filters", IEEE Transactions on Acoustics, Speech and Signal Processing, Volume: 35, No. 8, pp. 1170 - 1184, August 1987.

[64.] R.C. Gonzalez & R. E. Woods, Digital Image Processing, 2nd Edition, Prentice Hall, 2002.

[65.] S. Gautam & S. K. Brahma, "Overview of Mathematical Morphology in Power Systems – A Tutorial Approach", IEEE Power & Energy Society General Meeting, 2009.

[66.] M. Jing, W. Zengping, X. Yan & M. Lei, "Single-ended transient positional protection of transmission lines using mathematical morphology", Proceedings of the International Power Engineering Conference (IPEC), Singapore, November – Decemeber, 2005.

[67.] Y. Tingfang, L. Pei, Z. Xiangjun, & K. K. Li, "Application of Adaptive Generalized Morphological Filter in Disturbance Identification for Power System Signatures", in Proceedings of the Power System Technology International Conference, October 2006.

[68.] Z. Lu, Q. H. Wu & J. Fitch, "A Morphological Filter for Estimation of Power System Harmonics", Proceedings of the Power System Technology International Conference, October 2006.

[69.] O. Sen & R. Zhen, "Application of Improved Mathematical Morphology Method in the Power Quality Monitoring", Proceedings of the Power System Technology International Conference, October 2006.

[70.] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer, Boston, USA, 1991.

[71.] L. Zadeh, Fuzzy Sets, Information and Control, 1965

[72.] Y.Y. Yao, "Rough Sets, Neighborhood systems and Granular computing", Proceedings of the IEEE Conference on Electrical and Computer Engineering, pp. 1553 - 1558, Canada, May 1999.

[73.] H. S. Nguyen, "Discretization of Real Value Attributes: A Boolean Reasoning Approach", PhD Thesis, Dept. of Mathematics, Warsaw University, Poland, 1997.

[74.] D. Dey, B. Chatterjee, S. Chakravorti & S. Munshi, "Rough-granular Approach for Impulse Fault Classification of Transformers using Cross-wavelet Transform", IEEE Transactions on Dielectrics & Electrical Insulation, Volume: 15, No. 5, October 2008.

[75.] M. Bal, "Rough Sets Theory as Symbolic Data Mining Method: An Application on Complete Decision Table", Information Science Letters: An International Journal, Volume: 2, pp. 35 - 47, 2013.

[76.] J. Rissanen, "Modeling by Shortest Data Description", Automatica, Volume: 14, pp. 465 - 471, 1978.

[77.] M. Hansen and B. Yu, "Model Selection and the Minimum Description Length Principle", Journal of American Statistical Association, Volume: 96, pp. 746 - 774, 2001.

[78.] D. Donoho, “For Most Large Underdetermined Systems of Linear Equations the Minimal l1-Norm Solution Is Also the Sparsest Solution”, Communications on Pure and Applied Mathematics, Volume: 59, No. 6, pp. 797 - 829, 2006.

[79.] E. Cande`s, J. Romberg & T. Tao, “Stable Signal Recovery from Incomplete and Inaccurate Measurements”, Communications on Pure and Applied Mathematics, Volume: 59, No. 8, pp. 1207 - 1223, 2006.

[80.] E. Cande`s and T. Tao, “Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies?”, IEEE Transactions on Information Theory, Volume: 52, No. 12, pp. 5406 - 5425, 2006.

[81.] M. Schmidt, “Least Squares Optimization with L1-Norm Regularization”, Notes on LASSO, Department of Computer Science, University of British Columbia, 2005.

[82.] P. Zhao & B. Yu, “On Model Selection Consistency of Lasso”, Journal of Machine Learning Research, No. 7, pp. 2541 - 2567, 2006.

[83.] R. Tibshirani, “Regression Shrinkage and Selection via the LASSO”, Journal of Royal Statistical Society, Volume 58, No. 1, pp. 267 - 288, 1996.

[84.] E. Amaldi and V. Kann, “On the Approximability of Minimizing Nonzero Variables or Unsatisfied Relations in Linear Systems”, Theoretical Computer Science, Volume: 209, pp. 237 - 260, 1998.

[85.] E. Cande`s, “Compressive Sampling”, Proceedings of the International Congress of Mathematicians, 2006.

[86.] R. Chartrand & W. Yi, “Iteratively reweighted algorithms for compressive sensing”, *International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2008.*

[87.] K. Huang and S. Aviyente, “Sparse Representation for signal classification”, Proceedings of the Neural Information Processing Systems Conference, 2006.

[88.] J. Wright, A. Y. Yang, A. Ganesh, S. Shankar Sastry & Y. Ma, “Robust Face Recognition via Sparse Representation”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 31, No. 2, pp 210 - 227, February 2009.