

Cross-wavelet Transform as a New Paradigm for Feature Extraction from Noisy Partial Discharge Pulses

D. Dey, B. Chatterjee, S. Chakravorti and S. Munshi

Jadavpur University
Department of Electrical Engineering
Kolkata 700032, India

ABSTRACT

In this work a new approach based on cross-wavelet transform towards identification of noisy Partial Discharge (PD) patterns has been proposed. Different partial discharge patterns are recorded from the various samples prepared with known defects. A novel cross-wavelet transform based technique is used for feature extraction from raw noisy partial discharge signals. Noise is a significant problem in PD detection. The proposed method eliminates the requirement of denoising prior to processing and therefore it can be used to develop an automated and intelligent PD detector that requires minimal human expertise during its operation and analysis. A rough-set theory (RST) based classifier is used to classify the extracted features. Results show that the partial discharge patterns can be classified properly from the noisy waveforms. The effectiveness of the feature extraction methodology has also been verified with two other commonly used classification techniques: Artificial Neural Network (ANN) based classifier and Fuzzy classifier. It is found that the type of defect within insulation can be classified efficiently with the features extracted from cross-wavelet spectra of PD waveforms by all of these methods with a reasonable degree of accuracy.

Index Terms — Artificial neural network (ANN), cross-wavelet transform, cross-wavelet spectrum, fuzzy classifiers, partial discharge (PD), pattern classification, rough set theory.

1 INTRODUCTION

PARTIAL Discharge (PD) is an electrical discharge that occurs across a portion of the insulation between two electrodes, without completely bridging the gap. PD occurs within an electrical insulation without any significant manifestation outside and leads to unwanted failure. After initiation, PD can propagate and develop into electrical trees until the insulation is so weakened that it fails completely with permanent breakdown. It is necessary therefore, to ascertain the type of discharge present to judge its severity. Noise contamination is one of the significant problems of PD detection. PD signals have high frequency components making it difficult to distinguish them from noises. Therefore, denoising is an important issue in PD pattern classification. Researchers have reported several efficient methodologies for denoising of PD signals with various types of noise contamination [1-4]. Different algorithms are also reported for PD pattern classification in [5-9]. Most of the classification methods require denoised PD pulses and almost all denoising schemes need expert operator. This work is aimed at proper classification of PD patterns from noisy waveforms recorded

in the laboratory environment without denoising them so that, the method can be used to develop an automated and intelligent PD detector that requires minimal human expertise during its operation and analysis.

It is known that PD patterns can be classified on the basis of the phase locations and frequency contents of the pulses [2]. Cross-wavelet transform method, which is an extension of wavelet analysis, gives a measure of correlation between two waveforms in time-frequency domain. So, cross-wavelet spectrum finds the regions in time-frequency plane where two waveforms possess high common power. Keeping this in mind, feature extraction based on cross-wavelet transform is proposed as a new methodology for PD pulse pattern classification in the present work. The features obtained from cross-wavelet spectra between noisy PD waveforms and that superposed on power frequency reference signal are used to classify different defects in the insulation.

Cross-wavelet transform has already been used successfully for impulse fault classification of transformers by the authors [10]. Moreover, the cross-wavelet spectrum analysis has also been reported by other researchers in geosciences [11-12], biomedical signal processing [13] and transient analysis [14].

A Rough set theory (RST) based decision support system is used for classification of defect patterns in the present

approach. Moreover, the efficiency of the feature extraction methodology has also been judged with other classification methods like, Artificial Neural Network (ANN) and Fuzzy logic based classifiers.

To justify the proposed approach four common types of defects have been emulated in the samples and the respective PD waveforms are recorded with standard laboratory setup and data acquisition hardware.

Results show that the proposed feature extraction method gives acceptable accuracy with all three classification tools, thus establishing the effectiveness of the proposed scheme.

2 PD DATA ACQUISITION SETUP

The PD pulses are acquired through a real-life laboratory setup. The schematic is shown in Figure 1 and the photograph of the setup is given in Figure 2. When a test voltage, above PD inception, is applied across the sample under test, partial discharge takes place within the void and PD pulses are filtered out from the applied high voltage signal through the coupling capacitor. The PD signal appears across the inductor, where the role of the damping resistor is to prevent sustained oscillatory signal across the inductor in the case of a PD pulse. C1 and C2 constitute the potential divider for the power frequency high voltage and the corresponding low voltage is used for phase-reference of PD pulses as indicated in Figure 1. The complete connection diagram of the data acquisition setup is shown in Figure 3. It is worth mentioning here that the bandwidth of the used detector is 50 MHz.

3 OVERVIEW OF CROSS-WAVELET TRANSFORM

As stated earlier, Cross-wavelet transform may be considered as an extension of wavelet based analysis. Though the detailed mathematical background can be found in [11-14], an overview of cross-wavelet transform is given below. If $x(t)$ and $y(t)$ are two time domain signals, the cross-wavelet transform is defined as:

$$W^{xy}(s, \tau) = \frac{1}{c_{\psi}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W^x(a, b) W^{y*} \left(\frac{a}{s}, \frac{b-\tau}{s} \right) \frac{dadb}{a^2} \quad (1)$$

Here, $W^x(s, \tau)$ and $W^y(s, \tau)$ are the wavelet transform of $x(t)$ and $y(t)$, respectively, with respect to a mother wavelet $\psi(t)$. ‘ s ’ and ‘ τ ’ are usual ‘dilation’ and ‘translation’ parameters. In the present study Morlet mother wavelet is considered. This choice of mother wavelet is purely dependent on the nature of the problem. In this case it has been found that the performance of the scheme using Morlet wavelet is better than other commonly used mother wavelets. However, depending on the problem some other mother wavelet may be chosen. c_{ψ} is a constant, having value, $c_{\psi} = \int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty$.

It is evident from equation (1) that cross-wavelet spectrum shows regions in time-frequency plane where two waveforms are having high common power. The Cross-wavelet spectrum

is plotted using the magnitude of W^{xy} and the phase angle, $\phi = \tan^{-1} \frac{\Im\{W^{xy}\}}{\Re\{W^{xy}\}}$. Here, $\Re\{W^{xy}\}$ and $\Im\{W^{xy}\}$ indicate the real and imaginary part of W^{xy} respectively.

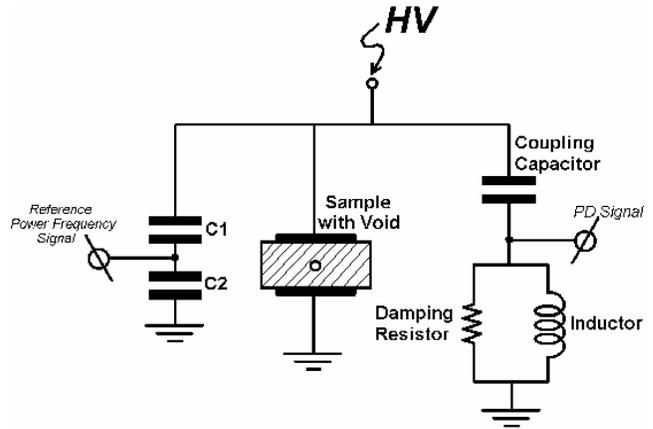


Figure 1. Schematic of PD data acquisition setup.



Figure 2. Photograph of the developed PD data acquisition setup

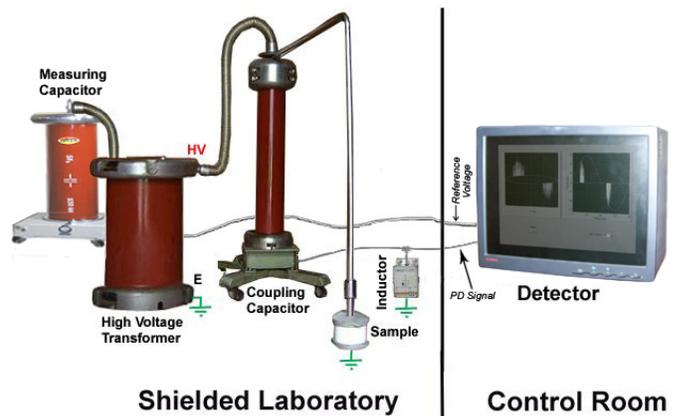


Figure 3. Wiring diagram of the developed PD data acquisition setup.

4 FEATURE EXTRACTION USING CROSS-WAVELET TRANSFORM

4.1 INSULATION DEFECTS UNDER STUDY

The defects that occur within the insulation can commonly be classified as i) narrow void (less than 1 mm² cross-section area), in contact with the electrode; ii) narrow void, away from the electrode; iii) large void (greater than 1mm² cross-section area), in contact with the electrode; iv) large void, away from the electrode. PD pulse patterns depend on the nature (size) and location of voids. In the present study, these four usual patterns of PD pulses are classified by the developed methodology.

The test samples are prepared using acrylic resin. Voids are created in the base material (i.e. acrylic resin) to simulate the defect inside an insulation. A representative diagram alongwith a photograph of a sample are shown in Figure 4. Plane-plane electrode system is used in the present study. A photograph of the electrodes placed across the sample under test is given in Figure 5.

4.2 FEATURE EXTRACTION FROM NOISY PULSES

The present study is aimed at the development of a scheme that is useful for classification of the PD pulses, recorded in laboratory environment. PD pulses get contaminated by noise even in a shielded laboratory environment. But in this case noise level is low compared to that in the outdoor or on-field PD measurements. However, this low noise too, could deform the signal upto such extent that it becomes difficult to detect the defect type and location from the recorded signals even for a human expert. It has been observed that the lowest value of signal to noise ratio is about -10 dB under different measurement conditions. These noises are mostly random in nature. So, denoising schemes are often used to cater to this problem. A typical spectrum of the noise is shown in Figure 6.

In recent times Wavelet analysis based denoising techniques are in use for this purpose. Shim et al [1] proposed use of wavelet transforms to improve their capability for detection and location of PDs in shielded distribution cables. Ma et al [2] applied continuous wavelet transforms for PD pattern recognition. They also implemented wavelet transforms to separate PD from electrical noise and proposed a wavelet transform based threshold value selection method. Satish et al [3] also employed wavelet transforms for PD denoising and comparison of performance with various noise sources. Hang et al [15] described application of the wavelet transform to extract PD signals from narrow-band interference. Zhang et al [4] utilized wavelet transforms to remove narrow band interference and white noise in on-line PD measurement. Ming et al [5] reported wavelet applications for characterization of PD, based on laboratory studies. Wavelet transform based denoising techniques were also applied by Tian et al [6] to improve signal to noise ratio (SNR) for detecting PD within a HV cable joint.

But the significant problem with most of these denoising methods is that they require an expert operator for the

denoising scheme to work. The expertise is required for selection of threshold values of denoising, observation of PD inception voltage and noise level present in that case etc.

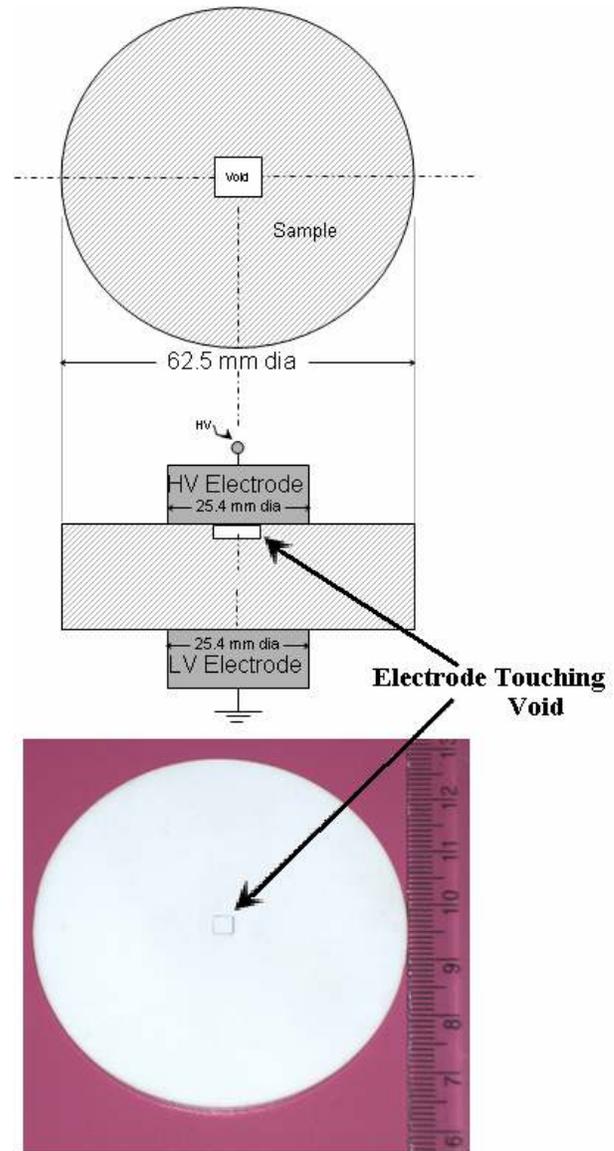


Figure 4. Sample prepared for PD simulation (a) Schematic (b) Actual Photograph.

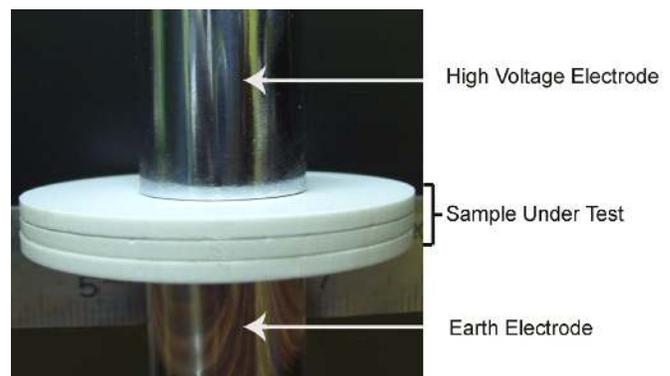


Figure 5. Actual Photograph of the electrode placement across the sample.

The aim of the present work is not only developing a scheme to eliminate the effect of noises from PD pulses but to do it in such a way that requires minimal human expertise, so that the scheme can be used in an intelligent, integrated and automated PD detection unit. Therefore, cross-wavelet transform method is used in the present study.

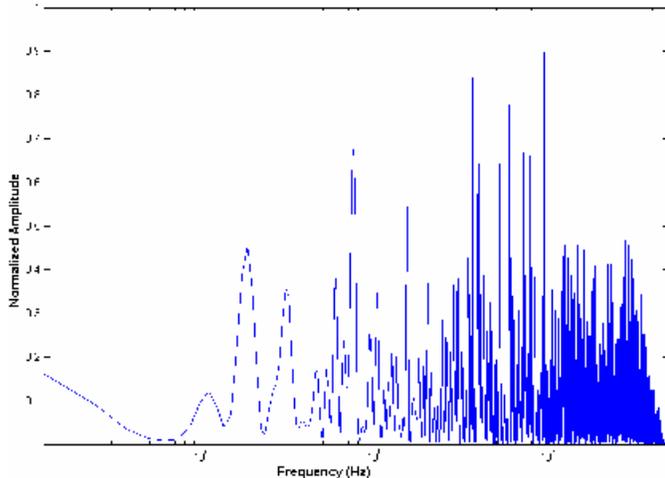


Figure 6. Typical frequency spectrum of recorded noise.

In this work, noisy PD signals are recorded from various samples with different defect (void) types. These waveforms are then superposed on the reference power frequency signal. This superposition is performed by the algebraic addition of the digitally stored data samples. One of such recorded data is shown in Figure 7a. The recorded PD waveforms were corrupted with noises which is evident from Figure 7a. Figure 7b shows the waveform when this noisy data is superposed on the power frequency reference signal. The random nature of the noise is responsible for low value of correlation of noise with itself, because random signals have very low auto-correlation sequence. Cross-wavelet transform gives the correlation of two signals in time-frequency domain. It basically uses continuous wavelet transform of two signals. Due to the random nature of the noise, its contribution in the wavelet coefficient values in the time-frequency domain is randomly distributed, which is not the case for PD pulses. So, when the correlation of such wavelet coefficients are computed for Cross-wavelet transform, the contribution of noise becomes low in terms of coefficient values in cross-wavelet spectrum. Therefore, prior denoising of the recorded waveforms could be avoided in the present analysis. Raw noisy waveforms are therefore used for processing. Cross-wavelet transforms between the superposed waveform and the corresponding original noisy waveform are obtained for all the recorded signals. Seven features, explained later, are extracted from these cross-wavelet spectra to classify the defect type [10].

The reason behind the choice of these two waveforms (i.e. PD pulses and the superposed waveform) to obtain the cross-wavelet spectra is explained below:

It is known that the nature of PD patterns depend on the phase locations and frequency contents of the pulses. So, the defect may be classified on the basis of this information, if they can be extracted from PD pulses [2]. As cross-wavelet transform gives a measure of correlation between two

waveforms in time-frequency domain, it is a possible choice for extracting the hidden features from the PD pulse patterns to classify the type of defects. In the present case of application two types of signals are recorded: a) the noisy PD pulses; b) the power frequency reference signal. So, the possible cross-wavelet spectrum which can be used for feature extraction to classify PD patterns may be one of the following:

- i) Cross-wavelet spectrum obtained from the cross-wavelet transform of noisy PD pulse and the power frequency reference signal
- ii) Cross-wavelet spectrum obtained from the cross-wavelet transform of noisy PD pulse with itself
- iii) Cross-wavelet spectrum obtained from the cross-wavelet transform of PD pulse and power frequency reference signal superposed on noisy PD pulse

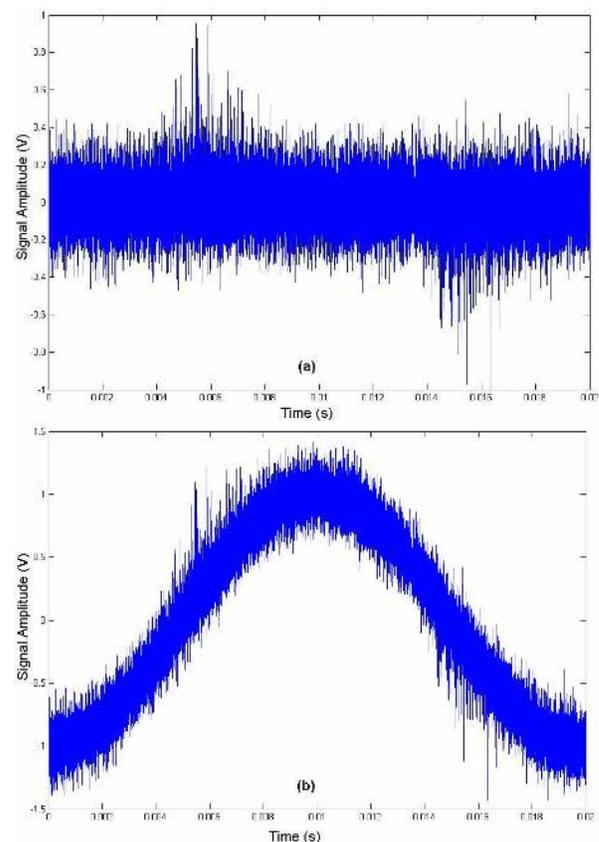


Figure 7. Signals considered for Cross-wavelet Transform: (a) Noisy PD signal (b) PD signal superposed on power frequency reference.

In the first case, as the frequency content of the noisy PD signal is very high with respect to the power frequency signals, the coefficient values for cross-wavelet spectrum are low and close to each other, thus giving poor performance in classification.

In the second case, the phase information has been lost because two waveforms used in cross-wavelet transform are same. So, there occurs a loss of information, causing poor performance of the classification algorithm.

It has been found in the present case that the third one gives acceptable performance for classification of the PD pulses using the scheme studied in this work. So, this method is applied in the

present study. A pictorial representation of the scheme is shown in Figure 8. However, there exist other possibilities of selecting different combinations of waveforms for some other problem.

Cross-wavelet spectrum of the waveforms shown in Figure 7 is given in Figure 8.

In Figure 8, $|W^{xy}|$ values at different ‘time’ and ‘scale’ are plotted. The x-axis is considered as ‘time’ axis and y-axis shows the ‘scale’, which can be considered as the inverse of frequency. The color of the figure at a point shows the value of $|W^{xy}|$ at that location time-frequency plane. The color-bar given on the right side of Figure 8 indicates the value corresponding to a color. Higher the value, higher is the common power at that time-frequency point.

Black arrows show the phase angle. Arrows shown in Figure 8, pointing towards right indicate “in-phase” (i.e. phase difference in zero) and arrows pointing left indicate “anti-phase” (i.e. phase difference is 180 degrees) conditions. In general, the angles made by the arrows with the time axis indicate their respective phase angles.

The ‘U’ shaped and black colored line shows the “cone of influence” (COI) which indicates the effective region for the analysis. COI separates the region where edge effects due to zero padding are significant. Similar to spectral analysis, errors will occur at the edges (i.e. beginning or end) of the waveform in the case of cross-wavelet because of the finite length of time series. Padding with zeros introduces discontinuities and the amplitude of the coefficients near the edge decreases at larger scales as more zeroes are added. The cone of influence is the region of the cross-wavelet spectrum in which these effects become important. This is defined as the e-folding time for the autocorrelation of wavelet power at each scale. It ensures that the power for a discontinuity drops by a factor of e^{-2} (where, $e= 2.7182\dots$) and the edge effects are negligible beyond this point.

Feature extraction is the most important aspect of any classification or clustering problem. Choice of the features for a particular problem may be considered as a combination of trial-and-error method and application of experience. For identification of PD patterns seven features are extracted from each of the cross-wavelet spectra, in a way similar to that followed in the case of impulse fault identification in [10] by the authors. These parameters represent the significant features of the $|W^{xy}|$ surface and produced good results in impulse fault identification [10].

The features (from A_1 to A_7) are described below:

$$1. A_1 = \frac{\sum_s \sum_\tau s \tau |W^{xy}(s, \tau)|}{\sum_s \sum_\tau |W^{xy}(s, \tau)|} \quad (2)$$

$$2. A_2 = \sqrt{\frac{\sum_s \sum_\tau s^2 \tau^2 |W^{xy}(s, \tau)|}{\sum_s \sum_\tau |W^{xy}(s, \tau)|}} \quad (3)$$

$$3. A_3 = \frac{\sum_s \sum_\tau |W^{xy}(s, \tau)|}{|W^{xy}(s, \tau)|_{peak}} \quad (4)$$

$$4. A_4 = \frac{\sum_s \sum_\tau |W^{xy}(s, \tau)|}{(s_{max} - s_{min})(\tau_{max} - \tau_{min})} \quad (5)$$

$$5. A_5 = \sqrt{\frac{\sum_s \sum_\tau (A_4 - |W^{xy}(s, \tau)|)^2}{(s_{max} - s_{min})(\tau_{max} - \tau_{min})}} \quad (6)$$

$$6. A_6 = \text{“s” at peak of } |W^{xy}(s, \tau)| \text{ i.e. } |W^{xy}(s, \tau)|_{peak} \quad (7)$$

$$7. A_7 = \text{“}\tau\text{” at peak of } |W^{xy}(s, \tau)| \text{ i.e. } |W^{xy}(s, \tau)|_{peak} \quad (8)$$

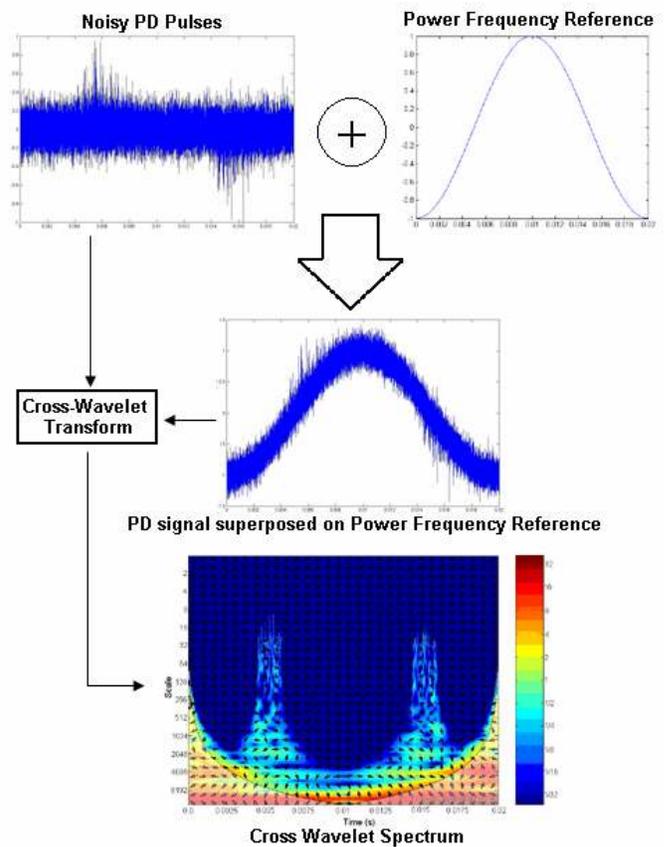


Figure 8. Scheme to obtain Cross-wavelet spectrum for feature extraction

However, one may choose some other features like, location of local peaks of $|W^{xy}|$ surfaces, if any, or some features from phase angle information $\phi(s, \tau)$ depending upon the nature of the problem. In the present case, the seven features mentioned above are found to be sufficient for classification of PD patterns for four common types of defects. The computation time for this feature extraction method on a Pentium IV processor (2.68 GHz, with 1 GB RAM) based Personal Computer (PC), is found as 30 s approximately.

5 CLASSIFICATION OF EXTRACTED FEATURES

To classify the extracted features, a classification technique based on Rough set theory (RST) is used. It is evident that there is no a priori knowledge regarding which features of the cross-wavelet spectrum will be suitable and should be taken for classification of PD patterns. So, the data table obtained after the extraction of the features may contain imprecise or superfluous information. As Rough sets are well-suited for this kind of problems [16-21], a classifier based on Rough sets is preferred here. However, other classification techniques can also be used for the same purpose. So, the classification performances have also been studied with two other commonly used classification tools: ANN based classifier and Fuzzy c-means clustering. As these two techniques are well known, an overview of RST based classification is given here only.

In RST, data is presented in a decision table in which each row represents an object (e.g. information regarding the occurrence of PD for different defect types in different test samples) and each column represents an attribute. For example, in the present problem decision table contains extracted features (A₁-A₇) as the 7 condition attributes and the defect type as the decision attribute. For decision attribute the following notation is followed in the present problem: D₁= narrow void (less than 1 mm² cross-section area), in contact with the electrode; D₂= narrow void, away from the electrode; D₃= large void (greater than 1mm² cross-section area), in contact with the electrode; D₄= large void, away from the electrode. The normalized decision table used for training or rule generation is shown in Table 1.

Mathematically, information system $T = \langle U, Q, V, f \rangle$. Here, U is the finite set of objects and Q is the set of attributes. $V = \bigcup_{q \in Q} V_q$, where V_q is the domain of the values of q and f denotes decision function as, $f: U \times Q \rightarrow V$. If the table is having a large number of attribute values i.e. $card(V_q)$ is very high for some $q \in Q$, then there is a very low chance of classifying a new object by the rules generated directly from the table. Here, $card()$ means cardinality operator, which means “number of elements of a set”. Therefore, discretization of the decision table is required for large real-valued decision table. Discretization of a data table indicates some partitioning of the attribute values. In the present problem Maximal Discernible (MD) heuristic is followed which is discussed in details in [22]. The discretized decision table is shown in Table 2.

In RST, for different attributes, objects are called indiscernible, i.e. similar, if they are characterized by the same information. If $P \subseteq Q$ and $x_i, x_j \in U$, then x_i and x_j are indiscernible wrt the set of attributes P , if $f(x_i, q) = f(x_j, q), \forall q \in P$. For example, let $P = \{A_1\}$. The part of the discretized decision table given in Table 2 shows

that, at least objects 2, 3, 4, 6 are indiscernible with attribute A₁, as all these objects are having same value (i.e. value=1) for this feature. Similarly objects 5,..., 67 and 70 are also indiscernible using A₁ as they possesses same values (i.e. value =3) for these objects. An elementary set is the set of all indiscernible objects. So, for $P \subseteq Q$, an equivalence relation on U , called P -indiscernibility relation is given by I_P . Considering the previous example of $P = \{A_1\}$, it can be found that the P -elementary set includes sets like $\{2,3,4,6,\dots\}$, $\{5,\dots,67,70\}$, $\{\dots,68,69,71,72\}$ and $\{1,\dots\}$ etc., because all the elements of each of these sets are having same attribute values for attribute A₁.

Table 1. Decision table considered for Rule generation or Training

Objects	Condition Attributes							Decision Attribute
	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇	
1	0.52	0.35	0.86	0.79	0.48	0.29	0.86	D ₁
2	0.48	0.29	0.86	0.87	0.48	0.25	0.86	D ₂
3	0.48	0.25	0.86	0.90	0.48	0.27	0.86	D ₂
4	0.48	0.27	0.86	0.90	0.55	0.38	0.79	D ₄
5	0.60	0.49	0.70	0.62	0.56	0.38	0.79	D ₁
6	0.44	0.29	0.86	0.88	0.46	0.27	0.89	D ₃
.
.
.
67	0.66	0.52	0.52	0.39	0.66	0.55	0.52	D ₄
68	0.71	0.62	0.56	0.42	0.66	0.55	0.56	D ₁
69	0.71	0.55	0.53	0.40	0.66	0.52	0.50	D ₁
70	0.66	0.52	0.50	0.38	0.71	0.55	0.50	D ₃
71	0.71	0.59	0.59	0.47	0.74	0.66	0.67	D ₄
72	0.7	0.66	0.61	0.47	0.71	0.59	0.56	D ₄

Table 2. Discretized Decision table for Rule generation

Objects	Condition Attributes							Decision Attribute
	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇	
1	2	2	4	3	1	0	5	D ₁
2	1	2	4	4	1	0	5	D ₂
3	1	1	4	5	1	0	5	D ₂
4	1	2	4	5	2	1	4	D ₄
5	3	3	3	2	2	1	4	D ₁
6	1	2	4	4	1	0	6	D ₃
.
.
.
67	3	3	1	0	2	3	0	D ₄
68	4	4	1	1	2	3	1	D ₁
69	4	3	1	1	2	3	0	D ₁
70	3	3	1	0	3	3	0	D ₃
71	4	3	1	1	3	4	2	D ₄
72	4	4	2	1	3	4	1	D ₄

For any rough set Y , \underline{PY} and \overline{PY} are called P -lower and P -upper approximation of Y and defined as, $\underline{PY} = \{x \in Y \mid I_P(x) \subseteq Y\}$ and $\overline{PY} = \{x \in Y \mid I_P(x) \cap Y \neq \emptyset\}$ respectively. This indiscernibility relation can reduce a decision table. This can be done by keeping only one element

of an equivalence class and also keeping those attributes which preserve the indiscernibility relation. In other words, keeping all the information intact and removing the superfluous attributes. Thus obtained *minimal* sets of attributes are called *Reduct*. The *CORE* is the set of relations occurring in every *Reduct*, i.e. $CORE(P) = \bigcap RED(P)$.

From the *CORE* and *Reducts* one can generate the decision rules. Usually these rules are considered in “IF...THEN” formats. This is illustrated in the following paragraphs. For better understanding of the method a small decision table having six objects with three condition attributes and one decision attribute is considered and is shown as Table 3.

For a given subset $P \subseteq Q$, an attribute $q \in P$ is dispensable in P if and only if, $I_P = I_{(P-\{q\})}$; otherwise q is indispensable.

If every element in P is indispensable then P is called *independent* otherwise *dependent*. Let $P \subseteq Q$ and $D \subseteq Q$ have equivalence relations in U . The P -positive region of D is indicated as, $POS_P(D) = \bigcup_{Y \in I_D} PY$. In other words, it denotes

the set of elements that can correctly be classified into D -elementary sets obtained from I_D using the knowledge described by I_P . If $q \in P$ and $POS_P(D) = POS_{(P-\{q\})}(D)$ then q is D -dispensable in P , otherwise q is D -indispensable in P . If the set of attributes G ($G \subseteq P$) is a D -independent in P and $POS_G(D) = POS_P(D)$, then G is called D -reduct of P or in general *Reduct* of P .

All these definitions can be explained using Table 3. For example, if P is taken as, $P = \{A_1, A_2, A_3\}$, and $D = \{‘type\ of\ defect’\}$ (i.e. decision attribute), then $I_P = \{1\}, \{2\}, \{3\}, \{4\}, \{5\}$ and $\{6\}$; $I_D = \{1,2\}, \{3,4\}$ and $\{5,6\}$. Also, $POS_P(D) = \{1,2,3,4,5,6\}$. If the attribute A_1 is removed from P then, $POS_{(P-\{A_1\})}(D) = \{2,3,4,5\}$. Clearly, $POS_{(P-\{A_1\})}(D) \neq POS_P(D)$. Therefore the attribute A_1 is D -indispensable in P . Similarly, removing attribute A_2 gives, $POS_{(P-\{A_2\})}(D) = \{2,3,4,6\} \neq POS_P(D)$. Therefore attribute A_2 is also D -indispensable in P . Again, for attribute A_3 it is easy to observe that, $POS_{(P-\{A_3\})}(D) = \{1,2,3,4,5,6\} = POS_P(D)$. So, A_3 is D -dispensable in P . Thus, the set $\{A_1, A_2\}$ is the D -reduct of P . Therefore, the simplified or reduced form of Table 3 is given in Table 4. ‘-’ indicates “don’t care” (i.e. dispensable) condition. It can be said that, attribute values, $(A_1=4 \wedge A_2=2)$ is the characteristic for decision class ‘D₁’. Similarly, $(A_1=3 \wedge A_2=3)$ is the characteristic of decision class ‘D₂’ and $(A_1=4 \wedge A_2=1) \vee (A_1=0 \wedge A_2=2)$ is the characteristic of ‘D₃’. ‘ \wedge ’ and ‘ \vee ’ are logical “AND” and “OR” operators respectively. Intersections of these *Reduct* values for each of the decision class (i.e. D₁, D₂ and D₃) will give the *CORE* for the respective class. For decision class ‘D₁’ the intersection of $(A_1=4 \wedge A_2=2)$ and $(A_1=4 \wedge A_2=2)$ gives *CORE* values $A_1=4$ and $A_2=2$. Similarly, for ‘D₂’ the *CORE* values are $A_1=3$ and $A_2=3$. Again, for the decision class ‘D₃’ no such *CORE*

value is obtained from the Table 4, as the intersection of $(A_1=4 \wedge A_2=1)$ and $(A_1=0 \wedge A_2=2)$ is null. Furthermore, this reduced Table 4 can be used to generate decision rules. The decision rules obtained from the *Reduct* and *CORE* values are given in Table 5.

Table 3. A typical Decision table for illustration

Object	Condition Attributes			Decision Attribute
	A ₁	A ₂	A ₃	
1	4	2	2	D ₁
2	4	2	1	D ₁
3	3	3	0	D ₂
4	3	3	2	D ₂
5	4	1	2	D ₃
6	0	2	2	D ₃

Table 4. Reduced form of Table 3

Object	Condition Attributes			Decision Attribute
	A ₁	A ₂	A ₃	
1	4	2	-	D ₁
2	4	2	-	D ₁
3	3	3	-	D ₂
4	3	3	-	D ₂
5	4	1	-	D ₃
6	0	2	-	D ₃

Table 5. Decision Rules obtained from *CORE* and *Reducts*

Decision Rule No.	Statement of the Rule	
	IF	THEN
1	$(A_1=4 \wedge A_2=2)$	The defect type is D ₁
2	$(A_1=3 \wedge A_2=3)$	The defect type is D ₂
3	$(A_1=4 \wedge A_2=1) \vee (A_1=0 \wedge A_2=2)$	The defect type is D ₃

6 RESULTS AND DISCUSSIONS

Using the data acquisition setup, detailed previously, PD signals are recorded in the laboratory over one full cycle of power frequency wave (i.e. 20 ms) for four common types of defects: D₁ to D₄. These data are contaminated with real-life noises (i.e. these noises are not simulated or synthetically added) present in the laboratory environment. The total number of acquired PD pulse data sets is 120. Each of the four classes has 30 data sets and the total number of cross-wavelet spectra is 120. The complete data table obtained after feature extraction, has 120 rows (objects), and 8 columns with 7 condition attributes (i.e. extracted feature vector) and one decision attribute (defect type). 60% of the data (i.e. 72 data sets), randomly chosen, are considered to obtain the decision rules (or during training), as evident from Table 1. The remaining 40% of the data (i.e. 48 data sets) are used for testing.

To obtain the decision rules for classification of defect patterns from the extracted features using Rough set theory,

following steps are followed:

Step 1: The data table is discretized. A part of the discretized decision table is shown in Table 2.

Step 2: Identical Attributes and cases are eliminated

Step 3: Dispensable attributes are removed

As in the present case no identical attributes or cases are observed and all the attributes have been found to be ‘indispensable’, the discretized table remains unchanged after these steps for the present problem.

Step 4: CORE values are obtained from Reducts.

The sample computation of Reduct and CORE is described earlier. Following the same procedure the final form of the decision table is obtained and is shown in Table 6.

Step 5: Decision rules are generated from the final table of CORE and Reducts. Finally 4 IF...THEN rules are obtained. The rule-set obtained finally is shown in Table 7.

Table 6. Simplified form of the Decision table for Rule generation

Objects	Condition Attributes							Decision Attribute
	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇	
1	2	-	4	-	-	0	-	D ₁
2	1	-	4	-	1	0	-	D ₂
3	1	-	4	-	1	0	-	D ₂
4	1	-	-	5	2	-	4	D ₄
5	3	-	3	-	-	1	-	D ₁
6	-	2	4	4	-	-	-	D ₃
.
.
67	3	-	-	0	2	-	0	D ₄
68	4	-	1	-	-	3	-	D ₁
69	4	-	1	-	-	3	-	D ₁
70	-	3	1	0	-	-	-	D ₃
71	4	-	-	1	3	-	2	D ₄
72	4	-	-	1	3	-	1	D ₄

Table 7. Decision Rules obtained from CORE and Reducts

Decision Rule No.	Statement of the Rule	
	IF	THEN
1	$(A_1=4 \wedge A_3=1 \wedge A_6=3)$	Type of Defect is D ₁
2	$(A_1=1 \wedge A_3=4 \wedge A_5=1 \wedge A_6=0)$	Type of Defect is D ₂
3	$(A_2=2 \wedge A_3=4 \wedge A_4=0)$	Type of Defect is D ₃
4	$(A_1=4 \wedge A_4=1 \wedge A_5=3 \wedge A_7=2)$	Type of Defect is D ₄

It is worth mentioning here that the rules presented in Table 7 do not satisfy some of the objects used for rule generation in Table 2. The objects marked by grey color in Table 6 are those cases. The corresponding values of the attributes which are responsible for the mismatch are also indicated in grey color. Among the 72 objects considered for rule generation, only 7 such cases were found. As only 12 objects among the total 72 cases are shown in the Tables 2 and 6, it is unlikely that these cases appear within those 12 objects when all of 72 cases are arranged at random. Such a decision table may create an impression that CORE and Reducts are obtained by using all the objects during rule

generation. To avoid this misconception, all of the 7 odd cases are shown in the tables intentionally. The rest of the cases are chosen randomly.

It is observed that using all the 72 objects, the number of CORE values obtained were very low. While a closer look revealed that removal of some objects improves the performance significantly. Those cases were object number 1, 4, 5, 6, 67, 70 and 72 as shown in Table 6. These objects are considered as garbage or ‘improper’ data in the information system. So, these cases are not considered during the computation of decision rules.

The test dataset is discretized and tested by the generated decision rules to judge the validity of the rules. Among the 48 test data, the decision rules have classified the PD pattern correctly 42 times.

To judge the efficiency of the features for classification of PD patterns, two common classification techniques- Artificial Neural Network (ANN) based classification and Fuzzy C-Means (FCM) clustering- have also been used.

A feed-forward neural network having a network structure as (7-30-4) i.e. seven neurons are in the input layer, thirty neurons are in the hidden layer and four neurons in the output layer, is used in this work. Sigmoidal activation function is used in the network. Gradient based back propagation algorithm is used for training of the network. The seven dimensional feature vector is given as the input to this ANN and the output from the network is the type of PD pattern, i.e. one of the four common defect types as stated earlier. 1000 iterations are considered during training.

60% of the total data (i.e. 72 data sets), randomly chosen, are considered for training of ANN and the remaining 40% of the data are used for testing, as before. Among the 48 test data the ANN has correctly identified the defect type 40 times. So, the percentage of success is about 83%, which is reasonable for classification of noisy PD data.

Similarly, Fuzzy C-Means (FCM) clustering [23] is also applied on the extracted features for classification with the same data. The scatter-plot of the extracted features with respect to the first feature (A₁) is shown in Figure 9. In this case the number of correct identification is 41 out of 48, which is quite acceptable, too.

To judge the efficiency of the scheme against noise contamination, the same methodology is applied on the recorded data but in this case the recorded signals are denoised by wavelet decomposition technique, detailed in [4], prior to processing. In a similar way features are extracted and classified with different techniques. Moreover, the sensitivity of the proposed scheme on the number of data is also given in Table 8.

It is evident from the results that there is not much difference in terms of performance of the three classification techniques. All the three methods give acceptable range of accuracy in detection of the type of defect with noisy waveforms. Though the features extracted from denoised waveforms give better results than those of the noisy pulses, the performance with the noisy waveforms

too, have acceptable level of accuracy. It is a significant aspect of the present study because it can be used to develop an automated and intelligent PD detector that requires minimal human expertise during operation and analysis.

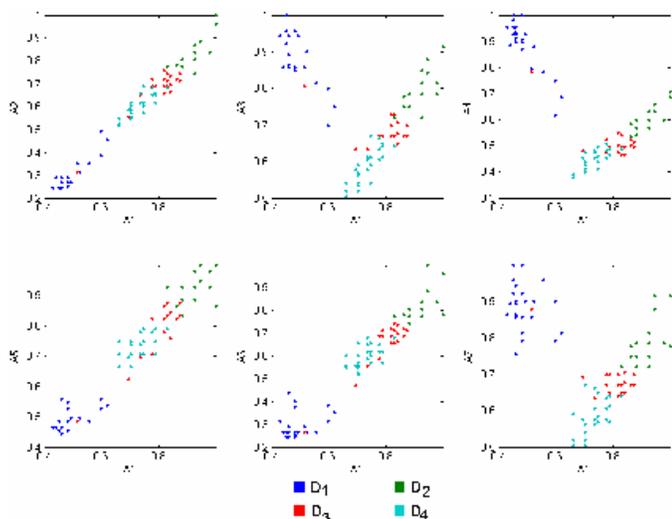


Figure 9. Scatter-plot of the extracted features with respect to the first feature (A₁).

Table 8. Effect of no. of data sets and noise on performance of the proposed scheme

No. of Training Datasets	Noisy or Denoised	No. of Testing Datasets	No. of Successful identifications (% success)		
			RST	ANN	FCM
72	Noisy	48	42 (87.5%)	40 (83.3%)	41 (85.4%)
60	Noisy	60	50 (83.3%)	50 (83.3%)	49 (81.6%)
30	Noisy	30	21 (70%)	18 (60%)	21 (70%)
72	Denoised	48	44 (91.6%)	45 (93.7%)	44 (91.6%)
60	Denoised	60	54 (90%)	52 (86.6%)	51 (85%)
30	Denoised	30	22 (73.3%)	20 (66.6%)	23 (76.6%)

The RST has the capability to extract information from imprecise or superfluous (noisy) data as explained in [16-22]. So, RST has a slight edge over the other two classification tools in terms of the performance in the present case, which is also evident from the relative comparisons shown in Table 8.

7 CONCLUSIONS AND FUTURE WORK

It has been shown in this work that cross-wavelet spectrum may be considered as an efficient feature extraction tool for PD pattern classification from noisy PD pulses. Successful classifications with real-life experimental data provide the justification of this approach. However, the methodology, presented here, is intended for PD detection in the laboratory environment. Performance of the scheme for outdoor testing

(e.g. on-field PD detection for cables etc.) has not been investigated, which lie well within the scope of future work.

Nevertheless, it does not reduce the importance of the findings in the present work. Because, majority of the PD testing is performed in the laboratory environment and noise contamination is also a significant problem there. The proposed methodology is proved to be very efficient for classification of those noisy PD pulses. So, the developed scheme can be used in an automated and intelligent PD detector that requires minimal human expertise during operation and analysis, which is a significant feat.

ACKNOWLEDGEMENT

The authors are thankful to DST, GOI, for extending financial help through the SERC project grant no.SR/S3/EECE/38/2006 to carry out this work.

REFERENCES

- [1] I. Shim, J.J. Soraghan, and W.H. Siew, "Detection of PD utilizing digital signal processing methods, part 3: open-loop noise reduction", IEEE Electr. Insul. Mag., Vol. 17, No.1, pp.6-13, 2001.
- [2] X. Ma, C. Zhou and I.J. Kemp, "Interpretation of wavelet analysis and its application in partial discharge detection", IEEE Trans. Dielectr. Electr. Insul., Vol.9, pp.446-457, 2002.
- [3] L.Satish and B.Nazneen, "Wavelet-based de-noising of partial discharge signals buried in excessive noise and interference", IEEE Trans Dielectr. Electr. Insul., Vol.10, pp.354-367, 2003.
- [4] H. Zhang, T.R. Blackburn, B.T. Phung and D. Sen, "A Novel Wavelet Transform Technique for On-line Partial Discharge Measurements, Part 1: WT De-noising Algorithm", IEEE Trans. Dielectr. Electr. Insul., Vol. 14, pp. 3-14, 2007.
- [5] Y. Ming and S. Birlasekaran, "Characterization of Partial Discharge Signals Using wavelet transform and statistical techniques", IEEE Intern. Sympos. Electr. Insul., Boston, MA, USA, pp. 9-13, 2002.
- [6] Y.Tian, P.L. Lewin and A.E. Davies, "Comparison of on-line partial discharge detection methods for HV cable joints", IEEE Trans. on Dielectr. and Electr. Insul., Vol.9, pp.604-615, 2002.
- [7] M. M. A. Salama and R. Bartnikas, "Determination of Neural Network Topology for Partial Discharge Pulse Pattern Recognition", IEEE Trans. Neural Networks, Vol. 13, pp. 446-456, 2002.
- [8] N.C. Sahoo, M. M. A. Salama and R. Bartnikas, "Trends in Partial Discharge Pattern Classification: A Survey", IEEE Trans. Dielectr. Electr. Insul., Vol. 12, pp. 248-264, 2005.
- [9] A. Contin, A. Cavallini, G. C. Montanari and G. Pasini, "Digital Detection and Fuzzy Classification of Partial Discharge Signals", IEEE Trans. Dielectr. Electr. Insul., Vol. 9, pp. 335-348, 2002.
- [10] D.Dey, B. Chatterjee, S. Chakravorti and S.Munshi, "Rough-Granular Approach for Impulse Fault Classification of Transformers using Cross-Wavelet Transform", IEEE Trans. Dielectr. Electr. Insul., Vol. 15, no. 5, pp. 1297-1304, Oct 2008.
- [11] C. Torrence and G. P. Compo, "A Practical Guide to Wavelet Analysis," J. American Meteorological Society, Vol. 79, No. 1, pp. 61-79, 1998.
- [12] B. G. Ruessink, G. Coco, R. Ranasinghe, and I. L. Turner, "A cross-wavelet study of alongshore non-uniform near shore sandbar behavior", Proc. Joint Conf. Neural Networks, Vancouver, Canada, pp. 4310-4317, 2006.
- [13] X. Li, X. Yao, J. R. G. Jefferys and J. Fox, "Computational Neuronal Oscillations using Morlet Wavelet Transform," Proc. IEEE Annual Conf. Engg. in Medicine and Biology, Shanghai, China, pp. 2009-2012, 2005.
- [14] M. I. Plett, "Transient Detection With Cross Wavelet Transforms and Wavelet Coherence," IEEE Trans. Signal Processing, Vol. 55, pp. 1605-1611, May 2007.
- [15] W. Hang, T. Kexiong and Z. Deheng, "Extraction of partial discharge signals using wavelet transform", IEEE 5th Intern. Conf. Properties and Applications of Dielectric Materials (ICPADM), Seoul, Korea, Vol. 1, pp. 322-325, 1997.

- [16] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer, Boston, USA, 1991.
- [17] J.T. Peng, C.F. Chien and T.L.B. Tseng, "Rough set theory for data mining for fault diagnosis on distribution feeder", *IEE Proc. Gener. Transmission. Distribution*, Vol. 151, pp. 689-697, 2004.
- [18] J.E. Cabral, J.P. Pinto, E.M Gontijo and J.R. Filho, "Fraud Detection in Electrical Energy Consumers Using Rough Sets", *IEEE Conf. Systems, Man and Cybernetics*, pp. 3625-3629, 2004.
- [19] A. Kusiak, "Rough Set Theory: A Data Mining Tool for Semiconductor Manufacturing", *IEEE Trans. Electronics Packaging Manufacturing*, Vol. 24, pp. 44-50, 2001.
- [20] S. K. Pal and P. Mitra, "Case Generation Using Rough Sets with Fuzzy Representation", *IEEE Trans. Knowledge and Data Engg.*, Vol. 16, pp. 292-300, 2004.
- [21] Y.Y. Yao, "Rough Sets, Neighborhood systems and Granular computing", *IEEE Conf. Electr. Computer Engg.*, pp. 1553-1558, Canada, 1999.
- [22] H. S. Nguyen, *Discretization of Real Value Attributes: A Boolean Reasoning Approach*, Ph.D. Thesis, Dept. of Mathematics, Warsaw Univ. Poland, 1997.
- [23] T. Ross, *Fuzzy Logic with Engineering Applications*, Wiley Pub., NJ, USA, 2004.



Debangshu Dey (M'09) received the B.E.E, M.E.E. and Ph.D. degrees from Jadavpur University, Kolkata, India in 2003, 2005 and 2009, respectively. Presently he is working as a Lecturer in the Electrical Engineering Department, Jadavpur University, Kolkata, India. He has published more than 20 research papers in the fields of his research. His areas of interest are condition monitoring of electrical equipment, intelligent instrumentation, signal conditioning and application of optimization and computational intelligence in electrical measurements.



Biswendu Chatterjee was born in Kolkata, West Bengal in 1973. He did his M.E.E. from Jadavpur University, Kolkata, India in 2004. Presently he is working as a Lecturer in the Electrical Engineering Department, Jadavpur University, Kolkata, India. His areas of interest are measurement, data acquisition and condition monitoring related to high voltage equipment.



Sivaji Chakravorti (M'90-SM'00) received the B.E.E., M.E.E. and Ph.D. degrees from Jadavpur University, Kolkata, India in 1983, 1985 and 1993, respectively. Since 1985 he has been a full-time faculty member of the Electrical Engineering Department of Jadavpur University, where he is currently a Professor in Electrical Engineering. In 1984 he worked at the Indian Institute of Science Bangalore as Indian National Science Academy Visiting Fellow. He worked at the Technical University Munich as Humboldt Research Fellow in 1995-96, 1999 and 2007 respectively. He served as Development Engineer in Siemens AG in Berlin in 1998. He has also worked as Humboldt Research Fellow in ABB Corporate Research at Ladenburg, Germany in 2002. In 2003 he worked as US-NSF guest scientist at the Virginia Tech, USA. He has published more than 100 research papers and has authored a book. He is the recipient of Technology Day Award of AICTE for best project work in 2003. He is a Fellow of the Indian National Academy of Engineering and an IEEE Power Engineering Society Distinguished Lecturer. His current fields of interest are numerical field computation, computer aided design and optimization of insulation system, condition monitoring of large electrical equipment and signal conditioning in high voltage systems.



Sugata Munshi obtained his B.E.E and M.E.E. degrees from Jadavpur University, Kolkata, India in 1980 and 1985, respectively. He worked as an Engineer in the Plasma Physics Division of Saha Institute of Nuclear Physics, India from 1985 to 1990. In 1986, he had training on 'Tokamak' machine in the 'Heavy Engineering Works' of Toshiba Corporation in Japan. In 1990 he joined the Electrical Engineering Department of Jadavpur University as a faculty member. At present, he is a Reader in this department. He has published more than 30 research papers in refereed journals. He was the joint recipient of The President of India's Prize (English) in 1989-90, The Pandit Madan Mohan Malaviya Memorial Prize in 1989-90, The Sir Thomas Ward Memorial Prize in 1994-95, The Tata Rao Medal, awarded by the Institution of Engineers (India) in 1996-97 and Certificate of Merit from IE (India) in 1996-97. His current fields of interest are signal processing, surge phenomena in power equipment and sensor systems.