

# Chapter 2

## Impulse Fault Analysis

### 2.1 Introduction

Outdoor transformers are often subjected to lightning and switching impulse/surges; these surges propagate along the transmission lines and thereafter into the winding of transformer. Any weakness of insulation may result in failure of the transformer due to such overvoltages. To prevent breakdown of the transformer and thereby interruption of supply, the equipment must be designed to withstand such high-voltage surges, the peak of which may be several times the normal working/rated voltage of the system. So, the impulse withstand test is an acceptance test for transformers.

The impulse test is normally done with impulse voltages, having standard wave shape of  $1.2/50 \mu\text{s}$  as per IEC 60076-4 [1]. These standard wave shapes are impressed on one side of the transformer winding, while keeping other end of the winding and tank to ground through a low resistance (in the case of tank current method). During impulse testing of transformers the oscillographic traces of applied voltage and corresponding current waveforms are investigated for judging the insulation condition of the transformers. This procedure requires significant human expertise and knowledge for proper judgment of insulation condition. Development of high speed processors with machine intelligence techniques makes it possible to analyze the impulse test waveforms by the machine itself without/with minimal requirement of human intervention. Different computer-aided methodologies, such as, artificial neural network (ANN) [2], fuzzy systems [3], and wavelet-based analysis [4] have already been reported for the classification and localization of fault within the transformer winding. Recent studies include information granulation-based approach using rough set analysis for this purpose.

For most of the cases regarding the fault analysis by machine intelligence techniques, faults are simulated either in a software-based model or in an analog model of the transformer. Suitable features are extracted from the response waveforms (fault current waveforms). A rule base is generated on the basis of those features with known fault conditions employing a suitable classification algorithm. This phase may be considered as training phase of a supervised learning. Thereafter the performance of the scheme can be tested upon real-life data.

## 2.2 Impulse Test Methods as per IEC

The impulse test is normally done with impulse voltages, having standard wave shape of  $1.2/50 \mu s$  as per IEC 60076-4, as stated earlier [1].

The routine impulse test normally consists of one reduced and one full-wave impulse, or two full-wave impulses, with the standard wave shape of  $1.2/50 \mu s$  as per IEC. The full-wave impulses are of a crest value equal to the rated BIL of the terminal being tested. These standard wave shapes are impressed on one side of the transformer winding, while keeping other end of the winding and tank to ground through a low resistance (called *tank current method*). The winding currents are recorded by measuring the voltage across the standard or known low resistance shunt. These currents are analyzed in order to find the faults within the transformer winding.

For Impulse testing of a three-phase winding, the impulse voltage is applied at one terminal at a time. The other two terminals are shorted at that time and grounded through a resistance equal to the surge impedance of the line, as shown in Fig. 2.1 [5]. Different test voltage levels depending on the voltage rating of the transformer winding under test is given in Table 2.1.

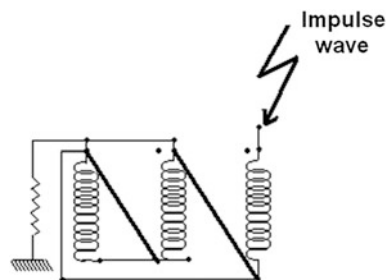
## 2.3 Faults in a Winding and Data Acquisition of Fault Current

Transformer failures during impulse test can be classified into four major classes as listed below:

- Winding-to-ground failure involving major insulation.
- Failure across relatively large portions of the winding but not involving ground.
- Inter winding breakdown, i.e., between high voltage (HV) and low voltage (LV).
- Failure between small portions of windings, such as coils, sections, and turns.

All the above-mentioned kinds of faults may be broadly classified under two heads, namely the *series faults* and the *shunt faults*. Series faults represent faults between turns, layers, and disks, whereas shunt faults stand for winding to ground, winding to tank, and winding to winding faults [1, 5].

**Fig. 2.1** Winding connections for impulse test



**Table 2.1** Rated withstand voltages for transformer windings with highest voltage for equipment  $U_m < 300$  V

Highest voltage for equipment, $U_m$ kV <sub>rms</sub>	Rated lightning impulse withstand voltage <sup>a</sup>	
	kV <sub>peak</sub>	
	List 1	List 2
3.6	20	40
7.2	40	60
12	60	75
17.5	75	95
24	95	125
36	145	170
52	250	
72.5	325	
123	450/550	
145	550/650	
170	550/650/750	
245	750/850/950	

<sup>a</sup> The choice between List 1 and List 2 in this table for  $U_m < 52$  kV and the choice between alternative rated withstand voltages for  $U_m = 123, 145, 170$ , and  $245$  kV depends on the severity of overvoltage conditions to be expected in the system and on the importance of particular installation

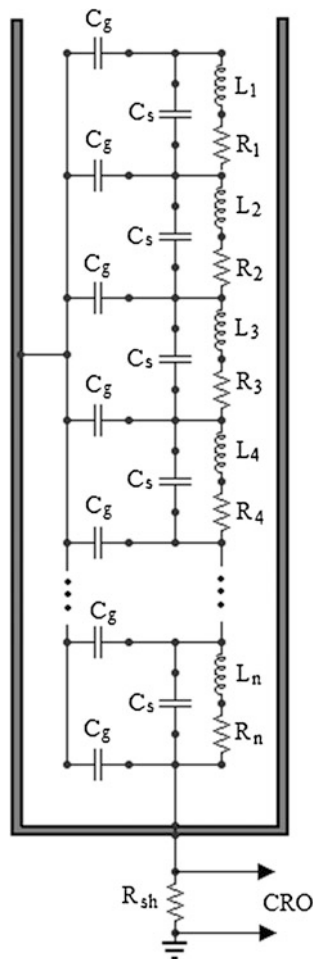
In most of cases the winding responses to faults were investigated by the ‘tank current method’ of fault detection, which is also called the ‘line current method’. In this method impulse is impressed on line end of the winding while the earth end of the winding and the transformer tank is connected to ground through a low resistance. The schematic of the “tank-current” method is given in Fig. 2.2, where the tank and the earth end of the winding are connected to ground through a shunt resistance ( $R_{sh}$ ).  $C_g$ ,  $C_s$ ,  $L_n$ , and  $R_n$  represent the capacitance to ground, series capacitance, equivalent inductance, and resistance of the winding for each disc, respectively. In ‘tank current’ method, the current components shunting out through the ground capacitors are also included in the recording of the response. This makes the ‘tank current’ method more sensitive as compared to the conventional ‘neutral current’ method.

These waveforms can be transferred to a PC for further analysis using suitable interfaces. A typical scheme for data acquisition is shown in Fig. 2.3 [6].

## 2.4 Impulse Fault Current Analysis

Detection of transformer insulation failure during impulse test in laboratory has been an important issue over a long period of time. Advancement in instrumentation and computer based signal processing tools have reformed this issue. Various tools and techniques have been used over the years for impulse fault

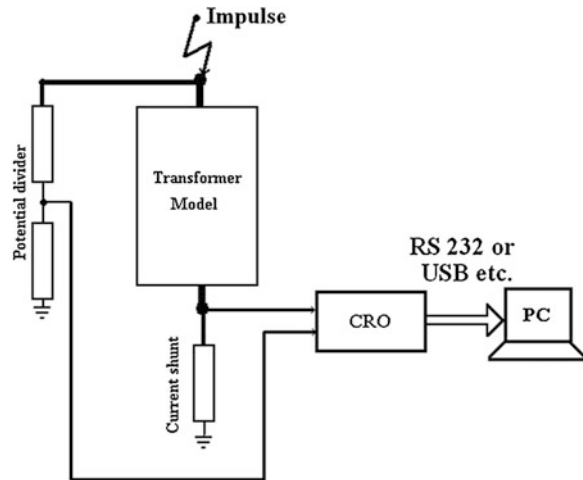
**Fig. 2.2** The schematic of tank-current method



diagnosis in transformers, including manual oscillographic analysis [7], frequency domain analysis [8], and modern signal processing techniques like wavelet [9], fractal [4], artificial intelligent tools like ANN [10], and fuzzy techniques [2].

The initial voltage distribution along the winding length in a transformer when a surge voltage is applied, depends upon the ratio of winding-to-ground capacitance ( $C_g$ ) to interturn capacitance ( $C_s$ ). These capacitances of the winding depend upon the winding geometry, i.e., the physical disposition of the turns and coils relative to each other and to the grounded parts of the structure. Therefore, different winding styles, e.g., layer and disc windings, have different values of the ratio ( $C_g/C_s$ ) and as a result they perform differently under the application of a voltage surge. In the present chapter, the schemes described with experimental results for the analysis of impulse faults of transformers, are based on disc-type

**Fig. 2.3** Schematic of data acquisition setup



windings [11, 12]. However, the schemes can easily be extended or modified for other types of transformer windings.

Abetti [12] found that the response of transformer windings to transient voltages, such as lightning and switching surges, was quite different from that under power frequency voltages. Transformers have complex arrangement of coils around an iron core, and due to close spacing of these coils, they have significant stray capacitances and inductances, which exhibit complex frequency-dependent characteristics. A transformer winding when subjected to a step surge, have been seen to behave initially like a capacitive network. The capacitive voltage distribution is followed by voltage oscillations which subsequently die down to yield finally a voltage distribution determined by winding resistances.

Malewski and Poulin [13] and Vaessen and Hanique [14] have shown experimentally that for higher frequencies essentially above 1 kHz, transformers behave linearly and the iron core does not play a significant role. Similar observations are also reported by the authors in [11].

A number of high-frequency transformer models, suitable for computer-aided transient simulations using electromagnetic transient program (EMTP) have been proposed in literatures [16–21].

Roy and Biswas [11] presented a simplified lumped parameter analog model for impulse studies of transformer windings. The modeling was carried out by considering it as an isolated winding, and the iron core was replaced by air core. The authors of [7, 22–24] discussed several methods for calculation of the self-inductance, mutual inductance, series capacitance, and the shunt capacitance of transformer disc winding.

### 2.4.1 Conventional Methods

Earlier contributions toward fault diagnosis were made by Hagenguth [25, 26]. He introduced the concept of fault analysis by carefully studying the voltage and current waveforms recorded by cathode ray oscilloscope (CRO). According to him, failures can be detected by meticulous comparison of the recorded wave shapes of the reduced voltage wave and the full voltage wave. The neutral current method, as this method is called, still remains one of the popular methods of impulse fault detection in transformers.

Stenkvist [7] in his CIGRE report, pointed out two different connections for the measurement of impulse current—namely the line-current (tank current) connection and the neutral current connection. According to him, the tank-current method is practically more suitable for detection and location of faults. Figure 2.4 shows typical winding current wave and impulse voltage wave [4].

Oscillographic comparison method, though apparently simple and straightforward, has lost its popularity over the years due its dependency on the knowledge and experience of the experts performing the analysis.

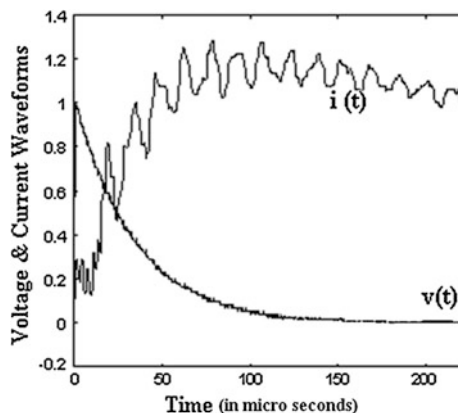
### 2.4.2 Modern Analysis Tools

The problem of impulse fault current analysis using modern machine intelligence tools may be addressed in two stages:

- (1) extraction of suitable features from the fault current waveforms;
- (2) classification of the features for proper fault identification with suitable classification algorithm.

These two aspects are discussed in details in the following sections.

**Fig. 2.4** Applied impulse voltage  $v(t)$  and winding current  $i(t)$  of a transformer under test



### 2.4.2.1 Feature Extraction Methodologies

With the advent of modern digital computers, high speed data acquisition devices, processors and analysis softwares, traditional techniques of fault current analysis have been over-shadowed by different signal/image processing tools/techniques.

Wavelet analysis has been applied successfully for transformer condition assessment and fault diagnosis in recent times by many researchers [27–30]. The inherent non-stationary pattern of transformer current waveforms during different fault conditions can be effectively classified using the frequency-selective feature of wavelet transform.

Purkait and Chakravorti [27] described the Multi-resolution analysis technique for PD detection during impulse testing of transformers. Koley [5, 30] described a technique based on wavelets for detection of impulse faults in transformers. For this purpose, they divided the HV winding of the transformer into a number of sections and they have extracted features of the faults for each of the sections from wavelet packet-based analysis. They have applied support vector machine (SVM)-based classifier to classify the extracted features for identification of faults. Their method is good enough to identify the faults within 10 % of the winding length.

Investigation results have also been reported in [6, 28, 31] on the use of wavelet analysis for the pattern recognition of winding current waveforms and classification of impulse faults. With the help of EMTP, series as well as shunt faults have been simulated at different locations along the entire length of disk-type HV winding of three different transformers. The authors studied the current waves employing continuous wavelet transform (CWT). Certain parameters were then selected for pattern classification of these fault current waves.

The authors of [32, 33] reported investigation results on the use of fractal geometry for the feature extraction and pattern recognition of winding current waveforms and classification of impulse faults therein. Fractal dimension (FD) is an intrinsic parameter of the image describing its fractal nature and lacunarity is a parameter describing the denseness (rather lack of denseness or lacuna) in the image.

#### Time–Frequency Domain Methods: Wavelet Transform

Any transform is basically a mathematical operation that maps a function or sequence from one domain to another. Usually mathematical transformations of different signals are performed to extract further information from that signal which are not evident from the raw (e.g. time-domain) signal. So, the advantages of using a transform are summarized below:

- Transformations, if suitably chosen, make an equation easier to solve than the original equation (e.g. Laplace transforms converts differential equations in a form that is easier to solve).

- The transform of a function gives additional or hidden information, which may not be observable in its original domain.
- The transform of a function or sequence may require less storage; hence, provide data compression or, reduction.
- An operation may be easier to apply on the transformed function, rather than the original function (like convolution)

The choice of transformation technique depends on the application.

In most of the signal processing applications, the knowledge of the frequency content of the signal plays an important role. The Fourier transform is probably the most popular transform which is used to obtain the frequency spectrum of a signal. However, Fourier transform is only suitable for stationary signals, i.e., signals whose frequency content does not change with time. The Fourier Transform, while it tells how much of each frequency exists in the signal, it does not tell at which time these frequency components occur. Signals such as image and speech have different spectral characteristics at different space and time, i.e., they are non-stationary. To analyze these signals, both frequency and time information are needed simultaneously, i.e., a time–frequency representation of the signal is needed.

To solve this problem, the short-time Fourier transform (STFT) [34, 35] was introduced. The major drawback of the STFT is that it uses a fixed window width. The Wavelet Transform, which has evolved considerably in the past few decades, provides a better time–frequency representation of the signal [35–38] than STFT.

*Continuous wavelet transform.* The CWT was developed as an alternative approach to the STFT to overcome the window width or, resolution problem. In the wavelet transform a signal is multiplied with a function (i.e. the *wavelet*), similar to the *window function* in the STFT, and the transform is computed separately for different segments of the time-domain signal.

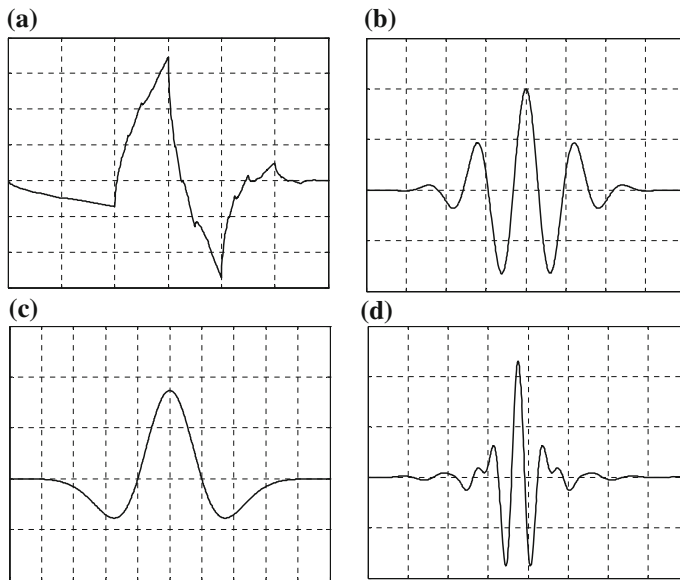
The CWT is defined as follows [5, 6, 34–37],

$$\text{CWT}(\tau, s) = \Psi(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} f(t) \Psi^* \left( \frac{t - \tau}{s} \right) dt \quad (2.1)$$

As in the above equation, the transformed signal is a function of two variables,  $\tau$  and  $s$ , the *translation* and *scale* parameters, respectively.  $\psi(t)$  is the transforming function, and is called the *mother wavelet*. The term *mother wavelet* gets its name due to the important property of the wavelet analysis as explained below:

The term *wavelet* means a small wave. The smallness refers to the condition that this (window) function is of finite length. The *wave* refers to the condition that this function is oscillatory. The term *mother* implies that the functions with different region of support that are used in the transformation process are derived from one main function, or the *mother wavelet*. In other words, the *mother wavelet* is a prototype for generating the other window functions. There are various types of mother wavelets like Daubechies, Morlet, Mexican Hat, Meyer, etc., as shown in Fig. 2.5 [5, 6, 34].





**Fig. 2.5** Mother-wavelet families: **a** Daubechies2, **b** Morlet, **c** Mexican Hat, **d** Meyer

The term *translation* is used in the same sense as it is used in the STFT, i.e., it is related to the location of the window, as the window is shifted through the signal. This term, obviously, corresponds to time information in the transform domain.

The parameter *scale* in the wavelet analysis is similar to the scale used in geographical maps. As in the case of maps, high *scales* correspond to a non-detailed global view (of the signal), and low *scales* correspond to a detailed view. Similarly, in terms of frequency, low frequencies (high scales) correspond to a global information of a signal (that usually spans the entire signal), whereas high frequencies (low scales) correspond to a detailed information of a hidden pattern in the signal (that usually lasts a relatively short time). So, it can be noted that *scale* is related to the inverse of frequency. However, the inverse of scale does not give the value of frequency directly. Scale ( $s$ ) and frequency ( $f$ ) are related according to the following relation [34–38]:

$$f = \frac{f_c}{s \cdot \Delta} \quad (2.2)$$

Here,  $f_c$  = center frequency of the mother wavelet and  $\Delta$  = sampling period of the signal.

So, *scaling* as a mathematical operation either dilates or compresses a signal. Larger *scales* correspond to dilated (or stretched out) signals and smaller scales correspond to compressed signals.

*Computation of the CWT.* The CWT is provided by Eq. (2.1), where  $f(t)$  is the signal to be analyzed.  $\psi(t)$  is the *mother wavelet* or the *basis function*. All the wavelet functions used in the transformation are derived from the mother wavelet through translation (shifting) and scaling (dilation or compression).

The mother wavelet used to generate all the basis functions is chosen depending on some desired characteristics associated with that function. The translation parameter  $\tau$  relates to the location of the wavelet function as it is shifted through the signal. Thus, it corresponds to the time information in the wavelet transform.

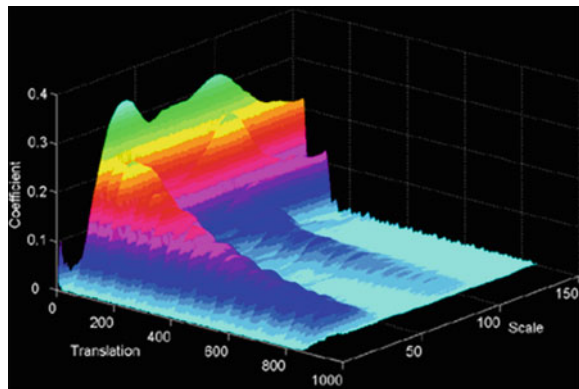
The scale parameter  $s$  is related to  $\left| \frac{1}{\text{frequency}} \right|$  and corresponds to frequency information, as stated earlier.

The wavelet series is obtained by discretizing CWT. This aids in computation of CWT using computers and is obtained by sampling the time-scale plane. The sampling rate can be changed according to *scale* change without violating the *Nyquist criterion*. For example, as the scale goes higher (i.e. at lower frequencies), the sampling rate can be decreased; thus reducing the number of computations [5, 34–38]. Usually the absolute coefficient values of CWT are plotted as a 3D surface in the time–frequency domain. A typical representation of such a plot of the CWT coefficients of an impulse fault current waveform of a transformer using Daubechies4 mother wavelet is shown in Fig. 2.6.

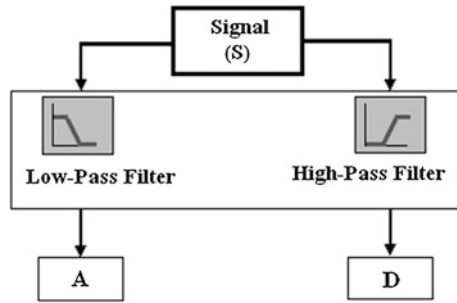
*Discrete wavelet transform (DWT).* In discrete wavelet analysis, the terms that are mostly used are *approximations* and *details*. The *approximations* are the high scale, i.e., low-frequency components of the signal. The *details* are the low scale, i.e., high-frequency components. Figure 2.7 shows the basic filtering process of a signal  $S$  in order to obtain low frequency and high-frequency components as done in DWT [5, 6, 35].

The original signal,  $S$ , is passed through two complementary filters and two output signals are obtained (“A” and “D” as shown in Fig. 2.7). Here, “A” is the low-frequency or *approximate* and “D” is the high-frequency or *detailed* part of the signal  $S$ . Unfortunately, this operation on a real digital signal ends up with twice as much data as when started. Let, for instance, the original signal  $S$  consists

**Fig. 2.6** Typical CWT surface for a PD signal



**Fig. 2.7** The filtering process, at its most basic level

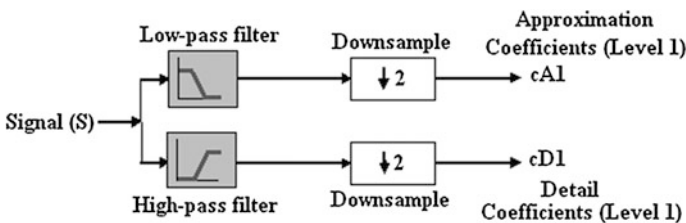


of 2,000 samples of data. Then the resulting signals will each have 2,000 samples, for a total of 4,000. There exists a more subtle way to perform the decomposition using wavelets, by keeping only one point out of every successive two in each of the two 4,000-length samples to get the complete information. This is the notion of *down-sampling*. It produces two sequences called *cA* and *cD* each having 1000 data points. Figure 2.8 describes the process of single step DWT. Starting from *S*, the first step produces two sets of coefficients: approximation coefficients *cA1*, and detail coefficients *cD1*. These vectors are obtained by convolving *S* with the impulse response of the low-pass filter for *approximation*, and with that of the high-pass filter for *detail*. A one-stage DWT of impulse fault current signal has been performed and shown in Fig. 2.9. In this case the detail coefficients *cD* are small and consist mainly of a high-frequency noise, while the approximation coefficients *cA* contain much less noise than the original signal.

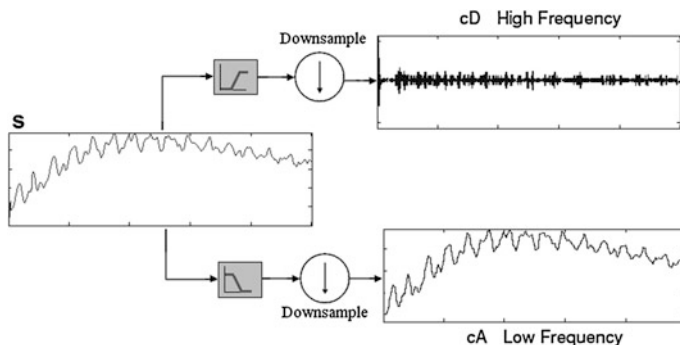
In the case of multi-level DWT, the next step splits the approximation coefficients *cA1* in two parts using the same scheme, replacing *S* by *cA1* and producing *cA2* and *cD2*, and so on. This is called the wavelet decomposition tree and shown up to level 3 in Fig. 2.10.

Though, in the case of DWT the approximation coefficients (*cA*) are usually decomposed further so that one signal is resolved down into many lower resolution components, one can also resolve the detailed part (*cD*) depending upon the information that is required from the signal in the respective problem.

In reality, the decomposition can proceed only until the individual details consist of a single sample or pixel. It can be proved easily that given a signal *S* of

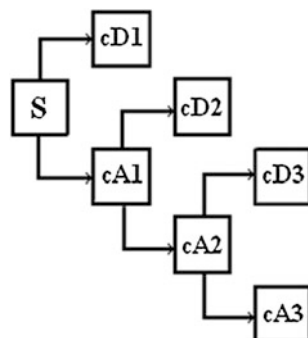


**Fig. 2.8** Single step DWT



**Fig. 2.9** Single step DWT of impulse fault current

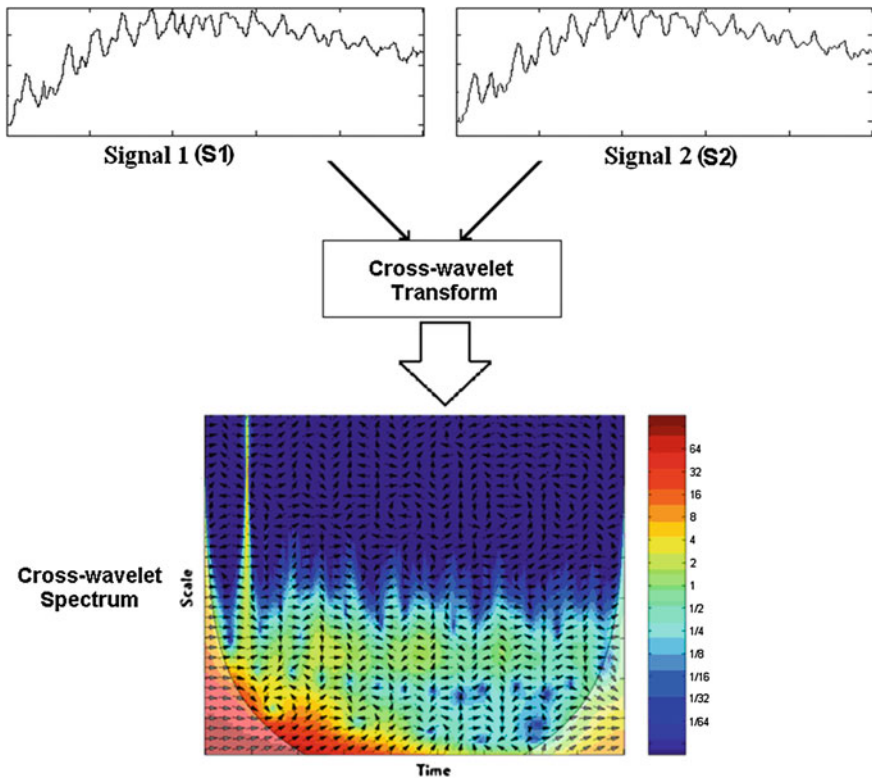
**Fig. 2.10** A typical Wavelet decomposition tree up to level 3



length  $N$ , the DWT consists of  $\log_2 N$  levels of decomposition at most. In practice, a suitable number of levels is selected based on the nature of the signal, or on the basis of a suitable criterion.

### Cross-Spectrum Analysis: Cross-Wavelet Transform

*Physical significance.* Cross-wavelet transform may be considered as an extension of wavelet-based analysis [39, 40]. The mathematical operation for cross-wavelet transform is similar to the computation of CWT. As the name suggests, cross-wavelet transform is performed between two signals. As it is discussed that the CWT represents a signal in time–frequency space, cross-wavelet transform of two time series gives cross-spectrum or cross-wavelet spectrum in time–frequency domain. The cross-wavelet transform gives a measure of correlation between two waveforms in time–frequency domain. The cross-wavelet spectrum shows regions in time–frequency space where two waveforms are having high common power. To demonstrate the operation, cross-wavelet spectrum of two impulse fault current waveforms of a transformer is shown in Fig. 2.11. As it is just a representative



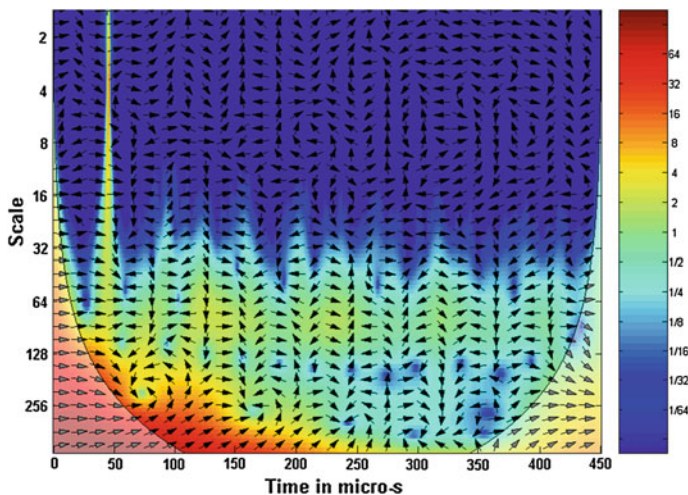
**Fig. 2.11** A representation of cross-wavelet spectrum

diagram of the procedure, magnitude values are not shown in the axes and how the cross-wavelet spectrum is plotted in the colored picture, is explained in below section

*Mathematical expressions.* The mathematical operation for cross-wavelet transform is similar to the computation of CWT, as stated earlier. Mathematically, the cross-wavelet transform of two signals,  $x(t)$  and  $y(t)$  is defined as [31, 39–43]:

$$W^{xy}(s, \tau) = \frac{1}{k_\psi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W^x(a, b) W^{y*}\left(\frac{a}{s}, \frac{b - \tau}{s}\right) \frac{da db}{a^2} \quad (2.3)$$

Here,  $W^*(s, \tau)$  and  $W^y(s, \tau)$  are the CWT of two signals  $x(t)$  and  $y(t)$  respectively with respect to a mother wavelet  $\psi(t)$ . The choice of mother wavelet depends on the nature of the waveform being processed.  $K_\psi$  is a constant, defined as,  $k_\psi = \int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty$ . 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



**Fig. 2.12** A typical Cross-wavelet spectrum of two current waveforms obtained from impulse testing of a transformer

$\Re\{W^{xy}\}$  indicate the real and imaginary part of  $W^{xy}$  respectively. For clarity arguments are omitted.

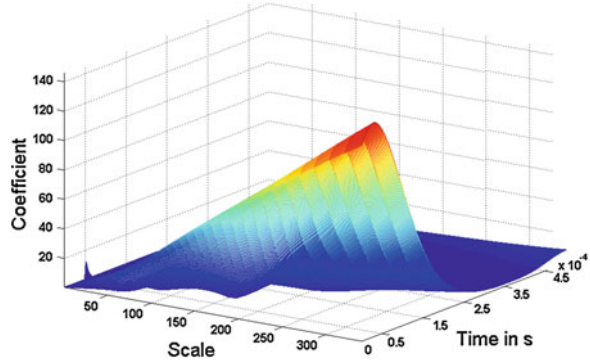
For further explanation, the cross-wavelet spectrum shown in Fig. 2.11 has been magnified and given in Fig. 2.12. The ‘U’ shaped and black colored line shows the “cone of influence” (COI). COI indicates the region where edge effects due to zero padding are significant. Similar to spectral analysis, errors 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.

In Fig. 2.12,  $|W^{xy}|$  values at different ‘time’ and ‘scale’ are plotted. The  $x$ -axis is considered as ‘time’ axis and  $y$ -axis shows the ‘scale’, which is related to the inverse of frequency. The color of the figure at a point shows the value of  $|W^{xy}|$  at that time–frequency space. The color-bar given in the right side of Fig. 2.12 indicates the value corresponding to a color. Higher the value, higher is the common power at that time–frequency point. The  $|W^{xy}|$  surface for the above cross-wavelet spectrum is shown in Fig. 2.13 [31].

Black arrows in Fig. 2.12 show the phase angle at that time–frequency space. Arrows, pointing towards right indicate “in-phase” (i.e. phase difference is zero) and arrows pointing left indicate “opposite-phase” (i.e. phase difference is  $180^\circ$ ). In general, the angles made by the arrows with the time axis indicate their respective phase angles.

Cross-wavelet transform has already been applied successfully in geosciences [40, 41] and biomedical signal processing [39]. Very recently it has also been proposed for transient analysis [31, 42, 44]. However, the cross-wavelet spectrum analysis can be used as an efficient tool in the processing of several signals obtained from different HV and power system measurements. Fault analysis in

**Fig. 2.13** The surface of  $|W^{xy}|$  in time–frequency space



different power equipment may be considered as one of such applications. More specifically for impulse fault analysis, cross-wavelet transform is used as a method for feature extraction. It is known that location and type of fault has significant effect on the time–frequency spectrum of the impulse test waveforms. So, the cross-wavelet transform of these waveforms with one another may give some important information for localization and identification of fault in the transformer winding. Therefore, analysis of cross-wavelet spectrum is a promising technique for transformer impulse fault identification. For similar reasons, cross-spectrum analysis has been used for pattern classification of partial discharge (PD) pulses by the researchers [45]. It can be easily understood that there exists several such domains where cross-spectrum analysis can be applied.

Different types of features can be extracted from cross-wavelet spectrum. For example, in [31] many features are extracted from  $|W^{xy}|$  as:

1.

$$F_1 = \frac{\sum_s \sum_\tau s\tau |W^{xy}(s, \tau)|}{\sum_s \sum_\tau |W^{xy}(s, \tau)|};$$

2.

$$F_2 = \sqrt{\frac{\sum_s \sum_\tau s^2 \tau^2 |W^{xy}(s, \tau)|}{\sum_s \sum_\tau |W^{xy}(s, \tau)|}};$$

3.

$$F_3 = \frac{\sum_s \sum_\tau |W^{xy}(s, \tau)|}{|W^{xy}(s, \tau)|_{\text{peak}}};$$

4.

$$F_4 = \frac{\sum_s \sum_\tau |W^{xy}(s, \tau)|}{(s_{\max} - s_{\min})(\tau_{\max} - \tau_{\min})}$$

5.

$$F_5 = \sqrt{\frac{\sum_s \sum_\tau (F_4 - |W^{xy}(s, \tau)|)^2}{(s_{\max} - s_{\min})(\tau_{\max} - \tau_{\min})}}$$

6.  $F_6 = "s"$  at peak of  $|W^{xy}(s, \tau)|$ , i.e.,  $|W^{xy}(s, \tau)|_{\text{peak}}$ 7.  $F_7 = "\tau"$  at peak of  $|W^{xy}(s, \tau)|$ , i.e.,  $|W^{xy}(s, \tau)|_{\text{peak}}$ 

Five more features, similar to  $F_1$  to  $F_5$ , may also be extracted from the phase angle  $\phi(s, \tau)$  data. They are:

8.

$$F_8 = \frac{\sum_s \sum_\tau s\tau |\phi(s, \tau)|}{\sum_s \sum_\tau |\phi(s, \tau)|};$$

9.

$$F_9 = \sqrt{\frac{\sum_s \sum_\tau s^2 \tau^2 |\phi(s, \tau)|}{\sum_s \sum_\tau |\phi(s, \tau)|}}$$

10.

$$F_{10} = \frac{\sum_s \sum_\tau |\phi(s, \tau)|}{|\phi(s, \tau)|_{\text{peak}}};$$

11.

$$F_{11} = \frac{\sum_s \sum_\tau |\phi(s, \tau)|}{(s_{\max} - s_{\min})(\tau_{\max} - \tau_{\min})}$$

12.

$$F_{12} = \sqrt{\frac{\sum_s \sum_\tau (F_{11} - |\phi(s, \tau)|)^2}{(s_{\max} - s_{\min})(\tau_{\max} - \tau_{\min})}}$$

However, one may choose some other features like, location of local peaks of  $|W^{xy}|$  and  $\phi(s, \tau)$  surfaces, if any, or some more features from  $|W^{xy}|$  and  $\phi(s, \tau)$  depending upon the nature of the problem.



### Correlation-Based Method

Cross-correlation is a mathematical operation similar to convolution technique and measures the degree of similarity between the two signals correlated. Correlation plays an important role in the digital signal processing, especially in applications dealing with radar and sonar. In practical applications in which correlation is used, one of the sequences is contaminated by the other forms of interference. Then, the correlation sequence is a function of the statistical characteristics of that interference. In case of feature extraction for impulse fault analysis, let, the impulse fault current of healthy (no-fault) transformer winding as  $A(n)$ , and that in case of a particular fault condition as  $B(n)$ . The cross-correlation sequence of  $A(n)$  and  $B(n)$  is a function bearing the characteristics of that particular fault condition. The statistical characteristics of correlation sequence indicate the nature of the insulation failure like type, nature and location of faults. The cross-correlation sequence of these impulse currents is given by [46–49]

$$E_{AB}(m) = \begin{cases} \sum_{n=0}^{W-m-1} A_{n+m} B_n & m \geq 0 \\ E_{BA}(-m) & m < 0 \end{cases} \quad (2.4)$$

where  $m = -W, \dots, -1, 0, 1, \dots, +W$ , represents time shift parameter and the subscript  $AB$  represents the signals correlated. Let,  $W$  be the finite number of samples of each of the signals then the resultant cross-correlation sequence will have  $2W - 1$  number of samples. For simplicity  $E_n$  has taken as the cross-correlation sequence of the two impulse currents, then sample features which may be extracted from the cross-correlation sequence are given from Eqs. (2.5a) to (2.5c).

$F1$  = Maximum value of the sequence ( $E_{\max}$ )

$$F2 = \frac{\sum_{n=-W}^W n E_n}{E_{\max}} \quad (2.5a)$$

$$F3 = \frac{\sum_{n=-W}^W n E_n}{\sum_{n=-W}^W E_n} \quad (2.5b)$$

$$F4 = \frac{\sum_{n=-W}^W |n| E_n}{\sum_{n=-W}^W E_n} \quad (2.5c)$$

However, one may choose other features like index of the maximum value, kurtosis, skewness and root mean abscissa depending upon the nature of the pattern identification problem in hand.

## Fractals Based Method

*Fractals* have been found to be successful to provide a description of naturally occurring phenomena and shapes, where conventional and existing mathematical models were found to be inadequate. In recent years, this technique has attracted increased attention for classification of textures and objects present in images and natural scenes [50–53].

The applications of fractals in the field of electrical insulation commenced in early 1980s. Niemeyer et al. [54] introduced a new stochastic model to describe the discharge pattern of dielectric breakdown in terms of fractal properties.

Fuji et al. [55] presented results to show that the ‘electrical tree’ shapes of the discharge process within the insulation material have fractal properties.

Satish and Zaengl [56] discussed the results on the use of fractal features for recognition of 3D PD patterns. They computed two fractal features, like the FD and lacunarity from 3D discharge patterns and analyzed their ability to discriminate among various discharge patterns.

Candela et al. [57] described a novel PD defect identification method by means of statistical and fractal parameters and a neural network.

Several researchers [58–61] have also reported suitable and effective use of fractal geometry for pattern classification of waveforms and profile-like curves. A variety of algorithms are available for the computation of FD.

Presently there is an increasing trend on the application of fractal techniques in the field of signal processing and pattern recognition, as it has solved many pattern recognition problems efficiently [62, 63]. That is why, fractal techniques have been applied for PD pattern recognition [57] and in the field of classifying impulse faults of transformers [32, 64]. Researchers in [64] have successfully classified type and location of fault within 33 % of transformer winding with the help of two fractal features—FD using box count algorithm, and Lacunarity.

In this section, the description of different Fractal features effective for impulse fault analysis of transformer winding is given. These fractal features include calculation of FD using box count and Higuchi algorithm, Lacunarity and Approximate Entropy (ApEn), as they were found to be efficient for this purpose by the researchers [5, 64].

*Basic concept of fractals.* The word fractal was first coined by the polish scientist Benoit Mandelbrot [50]. A *fractal* is a rough fragmented geometric shape that can be subdivided in parts, each of which (at least approximately) is a reduced size or copy of whole (alternatively called *self sufficient* functions or patterns). Fractal is any pattern that reveals greater complexity as it is enlarged. Figure 2.14 shows one such fractal shape.

This quality is usually called ‘self-similarity’ and can be also looked at as the object being composed of many copies of itself, each of which is scaled down with possible translations and rotations. However, many fractal objects found in nature are not self-similar, but similar only in a statistical sense, and hence referred to as statistically self-similar, i.e., these are only statistically scale invariant when magnified [50].

**Fig. 2.14** Typical fractal shape



*Box count method.* FD is an important characteristic of fractals because it contains information about their geometric structure. The term FD was introduced by Mandelbrot [50]. According to him, a set  $A$  in a Euclidean  $n$  dimensional-space, is said to be self-similar when  $A$  is the union of  $N$  distinct (non-overlapping) copies of itself, each of which has been scaled down by a ratio  $r$  in all coordinates. The fractal dimension  $D$  of the set  $A$  is given by the relation [5, 50, 64]:

$$Nr^D = 1$$

or,

$$D = \frac{\log N}{\log\left(\frac{1}{r}\right)} \quad (2.6)$$

Let, the entire set  $A$  is covered by a single  $n$ -dimensional box of size  $L_{\max}$  which can be a multiple of the side length  $L$  in the  $x$ - $y$  plane. If the set  $A$  is now scaled down by the ratio  $r$ , then the number of boxes of size  $L = rL_{\max}$  needed to cover the whole can be derived from (2.6) as:

$$N(L) = \frac{1}{r^D} = \left[ \frac{L_{\max}}{L} \right]^D \quad (2.7)$$

The simplest way to estimate  $D$  from (2.7) is to divide the  $n$ -dimensional space into a grid of boxes with side length  $L$  (e.g. square of side  $L$  for the case of two-dimensional space) and to count the number of non-empty boxes. If such a box contains at least one image point, it is called a non-empty box. For a fixed  $L$ , let the total number of non-empty boxes be  $N(L)$ . Several values of  $L$  are chosen and the corresponding  $N(L)$  are determined. Fractal dimension  $D$  is then obtained from the slope of a least square linear fit on the plot  $\log(L)$  versus  $\log(N(L))$  [5, 64].

*Higuchi method.* Higuchi's algorithm for the calculation of FD regarded as one of the most effective and stable estimator of FD [5, 60, 64].

Considering  $x(1), x(2), x(3), \dots, x(N)$  as the time sequence to be analyzed, let us construct  $k$  new time series  $x_m^k$  as

$$x_m^k = \left\{ x(m), x(m+k), x(m+2k), \dots, x\left(m + \left\lfloor \frac{N-m}{k} \right\rfloor k\right) \right\},$$

for  $m = 1, 2, \dots, k$

where  $m$  indicates the initial time value,  $k$  indicates the discrete time interval between the points and  $\lfloor a \rfloor$  means integer part of  $a$ . For each of the time series  $x_m^k$ , the average length  $L_m(k)$  is computed as

$$L_m(k) = \frac{\sum_{i=1}^{\lfloor \frac{N-m}{k} \rfloor} |x(m+ik) - x(m+(i-1)k)| (n-1)}{\lfloor \frac{N-m}{k} \rfloor k} \quad (2.8)$$

where  $N$  is the total length of the data sequence  $x$  and  $(N-1)/\lfloor (N-m)/k \rfloor k$  is a normalization factor. An average length is computed for all time series having the same delay (or scale)  $k$ , as the mean of the  $k$  lengths  $L_m(k)$  for  $m = 1, \dots, k$ . This procedure is repeated for each  $k$  ranging from 1 to  $k_{\max}$ , yielding a sum of average lengths  $L(k)$  for each  $k$  as indicated in (2.8)

$$L(k) = \sum_{m=1}^k L_m(k) \quad (2.9)$$

The total average length for scale  $k$ ,  $L(k)$ , is proportional to  $k^{-D}$ , where  $D$  is the FD by Higuchi's method. In the curve of  $\ln(L(k))$  versus  $\ln(1/k)$ , the slope of the least squares linear best fit is the estimate of the FD [65].

*Lacunarity*. It has been observed by the researchers [5, 50, 64] that FD alone is insufficient for the purpose of discrimination, since two differently appearing surfaces can have the same value of  $D$ . To overcome this, Mandelbrot [50] introduced the term called *lacunarity*  $\chi$ , which quantifies the *denseness* or *compactness* of an image surface. Many definitions of this term have been proposed and the basic idea is to quantify the gaps or lacunae present in the given pattern.

A procedure for calculating Lacunarity is given in [52]. Let  $P(m, L)$  be the possibility that there are  $m$  points within a box of side  $L$  (i.e. square of side  $L$ ), which is centered about an arbitrary point on the waveform. Then  $P(m, L)$  is normalized, as below for all  $L$

$$\sum_{m=1}^n P(m, L) = 1 \quad (2.10)$$

where  $n$  is the number of possible points within the box of side length  $L$ . Let, the image is of size  $L_{\max} \times L_{\max}$ . If the entire waveform is overlayed with boxes of size  $L$ , then the number of boxes with  $m$  points inside the box is  $(L_{\max}/m)P(m, L)$ .

Defining  $G(L)$  and  $G^2(L)$ , Voss [52] defined lacunarity as:

$$G(L) = \sum_{m=1}^n mP(m, L) \quad (2.11)$$

$$G^2(L) = \sum_{m=1}^n m^2P(m, L) \quad (2.12)$$

$$\chi(L) = \frac{G^2(L) - [G(L)]^2}{[G(L)]^2} \quad (2.13)$$

*Approximate entropy.* ApEn is a “regularity statistic” that quantifies the unpredictability of fluctuations in a time series [66]. ApEn reflects the likelihood that “similar” patterns of observations will not be followed by additional “similar” observations. A time series containing many repetitive patterns has a relatively small ApEn; a less predictable (i.e. more complex) process has a higher ApEn [5, 67].

The algorithm for computing ApEn has been published [66]. Here a brief summary of the calculations, as applied to a time series of impulse current responses  $i(t)$  of transformer under test has been provided. Given a sequence  $S_N$ , consisting of  $N$  samples of impulse current response  $i(1), i(2), \dots, i(N)$  the values of two input parameters,  $m$  and  $r$  has been chosen, to compute the approximate entropy,  $\text{ApEn}(S_N, m, r)$ , of the sequence. The second of these parameters,  $m$ , specifies the pattern length, and the third,  $r$ , defines the criterion of similarity. Denoting a subsequence or *pattern* of  $m$  samples, beginning at measurement  $t_0$  within  $S_N$ , by the vector  $p_m(t_0)$ , two patterns,  $P_m(t_0)$  and  $P(j)$ , are *similar* if the difference between any pair of corresponding measurements in the patterns is less than  $r$ , i.e.,  $|i(t_0 + k) - i(j + k)| < r$  for  $0 \leq k < m$

Now considering the set  $P_m$  of all patterns of length  $m$ , i.e.,  $p_m(1), p_m(2), \dots, p_m(N - m + 1)$ , within  $S_N$ , let us define

$$C_m(r) = \frac{n_m(r)}{N - m + 1} \quad (2.14)$$

where  $n_m(r)$  is the number of patterns in  $P_m$  that are similar to  $P_m(t_0)$  (given the similarity criterion  $r$ ). The quantity  $C_m(r)$  is the fraction of patterns of length  $m$  that resemble the pattern of the same length that begins at interval  $t_0$ . Now after calculating  $C_m(r)$  for each pattern in  $P_m$ , and defining  $C_m(r)$  as the mean of these  $C_m(r)$  values. The quantity  $C_m(r)$  expresses the prevalence of repetitive patterns of length  $m$  in  $S_N$ . Finally, the ApEn of  $S_N$ , for patterns of length  $m$  and similarity criterion  $r$ , is defined as:

$$\text{ApEn}(S_N, m, r) = \ln \left[ \frac{C_m(r)}{C_{m+1}(r)} \right] \quad (2.15)$$

### 2.4.2.2 Classification Methodologies

Faults in different parts of a transformer winding during impulse tests generate distinct signatures in the time–frequency spectrum. Although discrimination of a large set of such signatures is difficult and some time impossible for the human eye, this complex task can be successfully handled by modern classification algorithm like ANN, fuzzy logic-based classifiers, etc., provided proper features are extracted from the response signals.

De and Chatterjee [33, 68] proposed an ANN approach for detection and diagnosis of fault nature and fault location in oil-filled power transformers during

impulse testing. In their work, the “self-organizing feature map” (SOFM) algorithm with Kohonen’s learning and learning vector quantization-based classification have been successfully applied to the problem of fault current waveform pattern classification. The studies on impulse fault diagnosis using the above mentioned methods showed that the localization of the faults could be done within 25 % of the winding length.

Purkait et al. [2] described development of an adaptive fuzzy classifier, which can effectively determine various classes or categories of series and shunt faults in a wide range of power transformers. They performed the classification of faults successfully for 33 % of the winding length.

Authors in [69] have applied Fuzzy ARTMAP on frequency response data to localize impulse faults in power transformers. They found that localization of faults can be successfully done within 20 % of the winding length.

In recent years SVM and rough set theory (RST)-based classifiers are also used successfully by the researchers [5, 6, 31].

## Overview of Different Classification Methods

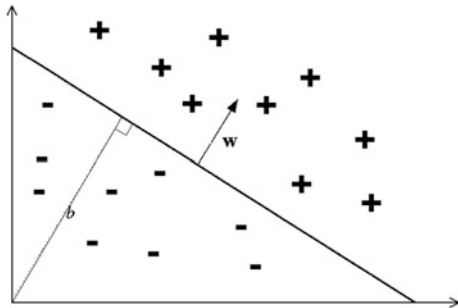
Construction of a machine capable of learning from experience has been the object of both philosophical and technical debate for a long time. The technical aspect of the debate has received enormous impetus from the advent of electronic computers. It has been demonstrated that machines can display a significant level of learning ability, though the boundaries of this ability are not clearly defined yet [5, 70]. However, a machine, capable of learning from experience through systematic training, is called ‘learning machine’.

It is worth mentioning here that, as ANN and fuzzy logic-based classifiers are being used by the researchers successfully for quite a long time, they are not discussed in this chapter. However, there should not be any doubt that they are indeed very powerful tools for such fault classifications.

In *supervised* learning, the machine is given a *training set* or, *examples* (called *inputs*) with associated labels (called *output/target* values). The inputs are in the form of feature vectors, so that the input space is a subset of  $\mathbb{R}^n$ . Once the attribute vectors are available, a number of sets of hypotheses could be chosen for the problem which relates the input and output vectors. Among these, linear functions are the best understood and simplest to apply. Traditional statistics and the classical neural network literature have developed many methods for discriminating between two classes of instances using linear functions, as well as methods for interpolation using linear functions [71].

*Linear classification.* Linear classification is frequently performed by using a real-valued function  $f: X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  in the following way: the input  $x = (x_1, \dots, x_n)$  is assigned to the positive class, if  $f(x) \geq 0$ , and otherwise to the negative class. Consider the case where  $f(x)$  is a linear function of  $x \subseteq X$ , so that it can be written as

**Fig. 2.15** A separating hyperplane ( $w, b$ ) for a two-dimensional training set



$$f(x) = \langle w \cdot x \rangle + b = \sum_{i=1}^n w_i x_i + b$$

where  $(w, b) \in R^n \times R$  are the parameters that control the function.

A geometric interpretation of this kind of hypothesis is that the input space  $X$  is split into two parts by the hyperplane defined by the equation  $\langle w \cdot x \rangle + b = 0$ , as shown in Fig. 2.15. The vector  $w$  defines a direction perpendicular to the hyperplane, while varying the value of  $b$  moves the hyperplane parallel to itself.

The theory of linear discriminants was developed by Fisher in 1936, while neural network researchers studied perceptrons in the early 1960s, mainly due to the work of Rosenblatt [72]. The quantities  $w$  and  $b$  are the weight vector and bias, terms borrowed from the neural network literature.

*Multi-class discrimination.* The problem of two-class discrimination studied so far can also be solved by defining a weight vector  $w_i$  and a bias  $b_i$  for each class. Each time a new instance has to be classified, both functions are evaluated, and the point  $x$  is assigned to class 1 if  $\langle w_1 \cdot x \rangle + b_1 \geq \langle w_{-1} \cdot x \rangle + b_{-1}$ , or to class  $-1$  otherwise. This approach is equivalent to discrimination using the single hyperplane  $(w, b)$ , with the substitutions  $w = w_1 - w_{-1}$ ,  $b = b_1 - b_{-1}$ .

For a multi-class classification problem the output domain is  $Y = \{1, 2, \dots, m\}$ . The generalization of linear learning machines to the  $m$ -class case is straightforward: to each of the  $m$ -classes are associated a weight vector and a bias,  $(w_i, b_i)$ ,  $i \in \{1, \dots, m\}$ , and the decision function is given by

$$c(x) = \arg \max_{1 \leq i \leq m} (\langle w_i \cdot x \rangle + b_i)$$

Geometrically this is equivalent to associating a hyperplane to each class, and to assigning a new point  $x$  to the class whose hyperplane is furthest from it. The input space is split into  $m$  simply connected and convex regions. Algorithms for learning the  $m$  hyperplanes simultaneously from data exist, and are extensions of the basic procedures outlined above [5].

*Linear regression.* The problem of linear regression consists of finding a linear function  $f(x) = \langle w \cdot x \rangle + b$  that best interpolates a given set  $S$  of training points labeled from  $Y \subseteq R$ . Geometrically this corresponds to a hyperplane fitting the given points.

This problem has been studied since the eighteenth century, and the best-known solution is that proposed independently by Gauss and Legendre of choosing the line that minimizes the sum of the squares of the distances from the training points. This technique is known as method of least squares, and is known to be optimal in the case of linear targets corrupted by Gaussian noise.

*Support vector machine.* SVMs are a system for efficiently training the linear learning machines in the kernel-induced feature spaces. An important feature of these systems is that, while enforcing the learning biases necessary for generalization, they also produce ‘sparse’ dual representations of the hypothesis, resulting in extremely efficient algorithms [5, 73]. Another important feature of the support vector approach is that due to Mercer’s [74] conditions on the kernels the corresponding optimization problems are convex and hence have no local minima.

The aim of support vector classification is to devise a computationally efficient way of learning ‘optimum’ separating hyperplanes in a high dimensional feature space, where ‘optimum’ hyperplanes are the ones optimizing the generalization bounds and by ‘computationally efficient’ means algorithms able to deal with sample sizes of the order of say, 100,000 instances.

*Maximal margin classifier:* The simplest model of SVM, which was also the first to be introduced, is the so-called maximal margin classifier. It works only for linearly separable data in the  $n$ -dimensional feature space, and hence cannot be used in many real-world situations. Nonetheless it is the easiest algorithm to understand, and it forms the main building block for the more complex SVMs.

The maximal margin classifier optimizes the bound of error by separating the data with the maximal margin hyperplane, given that the bound does not depend on the dimension of the space and this separation can be sought in any kernel-induced feature space. The maximal margin classifier finds the maximal margin hyperplane in an appropriately chosen kernel-induced feature space.

This strategy can be implemented by reducing it to a convex optimization problem: minimizing a quadratic function under linear inequality constraints. In case of linear classifiers there is an inherent degree of freedom, due to the fact that the function associated with the hyperplane  $(w, b)$  does not change in rescaling the hyperplane to  $(\lambda w, \lambda b)$ , for  $\lambda \in R^+$ . However, there will be a change in the margin as measured by the function output as opposed to the geometric margin. The margin of the function output is referred as the functional margin. It is possible to optimize the geometric margin by fixing the functional margin to be equal to 1 and minimizing the norm of the weight vector [73]. It is worth mentioning here that, hyperplanes with functional margin 1 are sometimes known as *canonical hyperplanes*. If  $w$  is the weight vector realizing a functional margin of 1 on the positive point  $x^+$  and the negative point  $x^-$ , its geometric margin can be computed as follows. Now, a functional margin of 1 implies

$$\langle w \cdot x \rangle + b = +1$$

$$\langle w \cdot x \rangle + b = +1$$



and to compute the geometric margin  $w$  must be normalized. The geometric margin  $\gamma$  is then the functional margin of the resulting classifier

$$\begin{aligned}\gamma &= \frac{1}{2} \left( \left\langle \frac{w}{\|w\|_2} \cdot x^+ \right\rangle - \left\langle \frac{w}{\|w\|_2} \cdot x^- \right\rangle \right) \\ &= \frac{1}{2\|w\|_2} (\langle w \cdot x^+ \rangle - \langle w \cdot x^- \rangle) \\ &= \frac{1}{\|w\|_2}\end{aligned}$$

Hence, the resulting geometric margin will be equal to  $1/\|w\|_2$  [5, 73, 74].

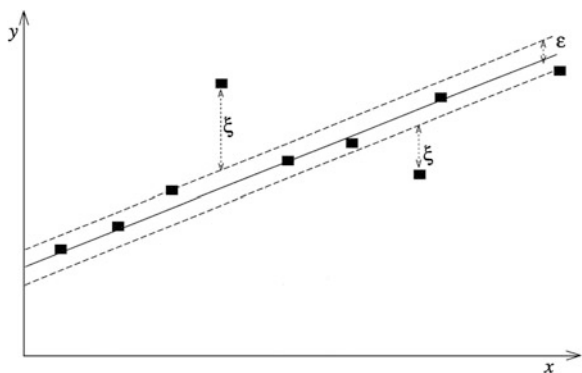
*Support vector regression:* The support vector method can also be applied to the case of regression, maintaining all the main features that characterize the maximal margin algorithm: a nonlinear function is learned by a linear learning machine in a kernel-induced feature space while the capacity of the system is controlled by a parameter that does not depend on the dimensionality of the space.

The classification approach is motivated by seeking to optimize the generalization bounds given for regression. Figure 2.16 shows an example of a one dimensional linear regression function with an  $\varepsilon$ -insensitive band. The variable  $\zeta$  measure the cost of the errors on the training points. These are zero for all points inside the band. Figure 2.17 shows a nonlinear regression function.

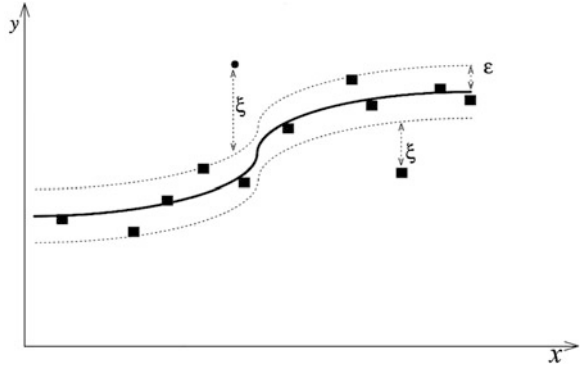
*RST-based classification.* RST theory was developed by Zdzislaw Pawlak [42, 43, 75–79] in the early 1980s. It deals with the analysis of data tables. The data can be acquired from measurements or from human experts. The main goal of the rough set analysis is to synthesize approximation of concepts from the acquired data. The term *classification* concerns any context in which some decision is taken or a forecast is made on the basis of currently available knowledge or information, as stated earlier.

RST can be applied for classification of data by simplifying a decision table. RST-based rule generation shows that RST is an effective methodology for classification of data patterns where the information system contains imprecise,

**Fig. 2.16** Example of a one dimensional linear regression function with an  $\varepsilon$ -insensitive band



**Fig. 2.17** A nonlinear regression function



superfluous and inconsistent data. It is known that prior to classification of data, some features are extracted from the data so that a classification algorithm can classify the data patterns efficiently. Selection of proper features for efficient classification is not an easy task. For example, in the present problem of impulse fault classification significant features may be extracted by using any method explained in the previous sections; but the difficulty of choosing appropriate features is an issue. Improper selection of feature vector may make the data table inconsistent or, it may contain superfluous data requiring unnecessarily higher time of processing. In this context RST-based classification is very effective. That is why RST has successfully been used for these types of problems [42, 43, 75–79].

*Overview of RST:* Even though details of the RST can be found in [42, 43, 75–79], a brief overview of RST is given in the following paragraphs to understand the classification procedure. In RST, data is presented in a decision table in which each row represents an object (e.g. impulse fault information at different disc position) and each column represents an attribute. For example, in cross-wavelet transform-based feature extraction for impulse fault analysis as explained in [31] the decision table contains extracted features ( $F_1$ – $F_{12}$ ) as the 12 condition attributes and the fault location and type as the decision attribute.

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.,  $\text{card}(V_q)$  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,  $\text{card}()$  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 many cases maximal discernible (MD) heuristic may be followed which is discussed in details in [79].

In short, to obtain the decision rules for classification of faults using RST, the following steps are followed:

- Step 1: *The data table is discretized.* This means that each condition attribute values are divided into ranges. MD heuristic may be followed which is discussed in details in [79], as stated in the previous section
- Step 2: Identical Attributes and cases are eliminated to simplify the table by removing superfluous data
- Step 3: Dispensable attributes are removed
- Step 4: *Reducts and CORE are obtained.* The sample computation of *Reduct* and *CORE* is available in [42]
- Step 5: Decision rules are generated in “IF\_THEN” form using the final data table of *CORE* and *Reducts*

*An example of RST-based Rule generation for classification* [6, 31]: Let us consider Table 2.2 as a data table with three features (condition attributes) and one decision attribute (fault type). Here, ‘SE1’ indicates the series fault in the winding location L1. Similarly ‘SH2’ implies shunt fault in location L2 and so on.

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 with respect to the set of attributes  $P$ , if  $f(x_i, q) = f(x_j, q), \forall q \in P$ . For example, let  $P = \{F_1\}$ . The part of the discretized decision table given in Table 2.2 shows that, at objects 1, 3, 4 are indiscernible with attribute  $F_1$ , as all these objects are having same value (i.e. value = 7) for feature  $F_1$ . Similarly objects 5 and 6 are also indiscernible using  $F_1$  as they possesses same values (i.e. value = 5) 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  $P = \{F_1\}$ , it can be found that the  $P$ -elementary set includes elementary sets like  $\{1, 3, 4\}$ ,  $\{2\}$ , and  $\{5, 6\}$ , because all the elements of each of these sets are having same attribute values for attribute  $F_1$ . The equivalence classes of the partition induced by the  $P$ -indiscernibility relation are called *information granules*. This is explained later.

For any rough set  $Y$ ,  $\underline{P}Y$  and  $\overline{P}Y$  are called  $P$ -lower and  $P$ -upper approximation of  $Y$  and defined as,  $\underline{P}Y = \{x \in Y | I_P(x) \subseteq Y\}$  and  $\overline{P}Y = \{x \in Y | I_P(x) \cap Y \neq \emptyset\}$  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

**Table 2.2** Truncated decision table for illustration

Object	Condition attributes			Decision attribute
	$F_1$	$F_2$	$F_3$	
1	7	6	1	SE1
2	3	6	0	SE1
3	7	6	0	SH1
4	7	2	0	SH1
5	5	0	5	SE2
6	5	6	5	SE2

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.,  $\text{CORE}(P) = \cap \text{RED}(P)$ . Therefore, CORE represents the *most important* part of the knowledge. 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 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,  $\text{POS}_P(D) = \bigcup_{Y \in I_D} PY$ . 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  $\text{POS}_P(D) = \text{POS}_{(P-\{q\})}(D)$  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  $\text{POS}_G(D) = \text{POS}_P(D)$ , then  $G$  is called  $D$ -reduct of  $P$  or in general *Reduct* of  $P$ .

All these definitions can be explained using Table 2.2. For example, if  $P$  is taken as,  $P = \{F_1, F_2, F_3\}$ , and  $D = \{\text{'Location and type of fault'}\}$  (i.e. decision attribute), then  $I_P = \{1\}, \{2\}, \{3\}, \{4\}, \{5\}$  and  $\{6\}$ ;  $I_D = \{1, 2\}, \{3, 4\}$ , and  $\{5, 6\}$ . Also,  $\text{POS}_P(D) = \{1, 2, 3, 4, 5, 6\}$ . If the attribute  $F_1$  is removed from  $P$  then,  $\text{POS}_{(P-\{F_1\})}(D) = \{1, 4, 5, 6\}$ . Clearly,  $\text{POS}_{(P-\{F_1\})}(D) \neq \text{POS}_P(D)$ . Therefore the attribute  $F_1$  is  $D$ -indispensable in  $P$ . Similarly, removing attribute  $F_2$  gives,  $\text{POS}_{(P-\{F_2\})}(D) = \{1, 2, 3, 4, 5, 6\} = \text{POS}_P(D)$ . Therefore attribute  $F_2$  is  $D$ -dispensable in  $P$ . Again, for attribute  $F_3$  it is easy to observe that,  $\text{POS}_{(P-\{F_3\})}(D) = \{2, 4, 5, 6\} \neq \text{POS}_P(D)$ . Similarly,  $F_3$  is also  $D$ -indispensable in  $P$ . Thus, the set  $\{F_1, F_3\}$  is the  $D$ -reduct of  $P$ . Therefore, the simplified or reduced form of Table 2.2 is given in Table 2.3. ‘—’ indicates “don’t care” (i.e. dispensable) condition. It can be said that, attribute values  $(F_1 = 7 \wedge F_3 = 1) \vee (F_1 = 3 \wedge F_3 = 0)$  are the characteristic for decision class ‘SE1’. ‘ $\wedge$ ’ and ‘ $\vee$ ’ are logical “AND” and “OR” operators respectively. Similarly  $(F_1 = 7 \wedge F_3 = 0)$  is the characteristic of decision class ‘SH1’ and  $(F_1 = 5 \wedge F_3 = 5)$  is the characteristic of ‘SE2’. These are called *information granules*. Intersections of these

**Table 2.3** Reduced form of Table 2.2

Object	Condition attributes			Decision attribute
	$F_1$	$F_2$	$F_3$	
1	7	—	1	SE1
2	3	—	0	SE1
3	7	—	0	SH1
4	7	—	0	SH1
5	5	—	5	SE2
6	5	—	5	SE2

**Table 2.4** Decision rules obtained from CORE and Reducts

Decision rule no.	Statement of the rule	
	If	Then
1	$(F_1 = 7 \wedge F_3 = 1) \vee (F_1 = 3 \wedge F_3 = 0)$	Location and type of fault is 'SE1'
2	$(F_1 = 7 \wedge F_3 = 0)$	Location and type of fault is 'SH1'
3	$(F_1 = 5 \wedge F_3 = 5)$	Location and type of fault is 'SE2'

reduct values for each of the decision class (i.e. SE1, SH1, and SE2) will give the CORE for the respective class. For the decision class 'SE1' no such CORE value is obtained from Table 2.3, as the intersection of  $(F_1 = 7 \wedge F_3 = 1)$  and  $(F_1 = 3 \wedge F_3 = 0)$  is null. Similarly, for decision class 'SH1' the intersection of  $(F_1 = 7 \wedge F_3 = 0)$  and  $(F_1 = 7 \wedge F_3 = 0)$  gives CORE values  $F_1 = 7$  and  $F_3 = 0$ . Again, for 'SE2' the CORE values are  $F_1 = 5$  and  $F_3 = 5$ . Furthermore, this reduced Table 2.3 can be used to generate decision rules. The decision rules obtained from the Reduct and CORE values are given in Table 2.4 [31].

### 2.4.3 How to Build an Expert System for Impulse Fault Analysis

The structure of the designed expert system (ES) has the following functional modules: (a) user interface (UI), (b) knowledge base: consisting of the data-base and the rule-bases, (c) inference engine, (d) explanation module, and (e) updating module.

The different functional modules of the ES are interrelated with each other [80].

The necessary input data required for the data-base may be the voltage and current wave records at reduced impulse voltage level and at full impulse voltage level (BIL), respectively. The ES stores and retrieves data from the data-base before the actual fault detection analysis begins. After the analysis, the ES comes out with a decision regarding possible fault and its location in the winding.

With this UI, the user can ask the ES to identify and locate the fault following the rules of fault diagnosis using the combined time and frequency domain knowledge bases. The ES may also have an explanation module, which displays the decision tree to the user to provide transparency in decision making to the user.

For fault diagnosis, the waves to be analyzed are added to the database of the ES via the UI. Then the UI interacts with the user to read the transformer rating and invokes the inference engine which accordingly determines the necessary diagnostic indices. The fault diagnosis algorithm follows an IF\_THEN\_ELSE structure.

It combines the heuristic rules framed on the basis of the acquired knowledge and conventional mathematical pattern recognition parameters to identify the fault current waveform pattern to arrive at the conclusion regarding the type and

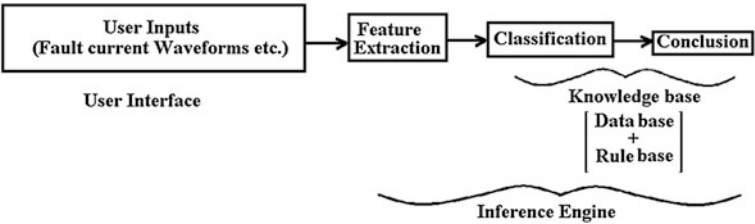


Fig. 2.18 Schematic of ES

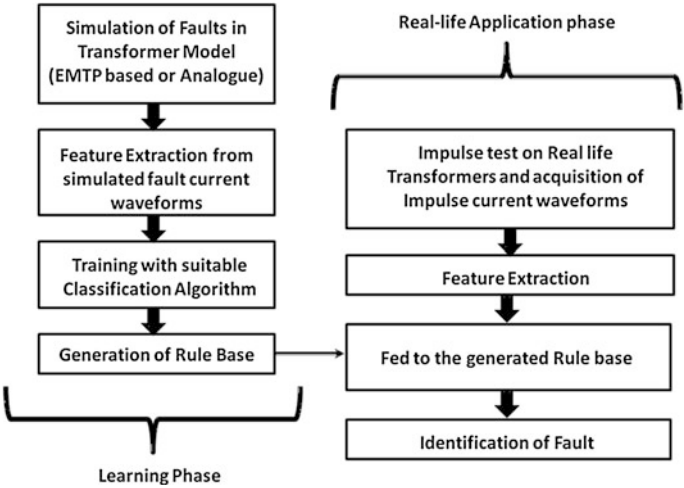


Fig. 2.19 Logical Steps for ES operation

location of the fault. A typical schematic for ES operation is given in Fig. 2.18. The basic logical steps of operation of the ES are given in Fig. 2.19. The ES can be developed based on these figures in any high level languages [80, 81].

2.5 Recent Challenges and Research Directions

In most of the cases referred previously, the fault within the transformer winding is considered as a *static* fault. The faults created due to the defects that are present in insulation prior to the application of impulse sequences and do not vary with the propagation of the impulse wave through the winding, are named as static faults. They can be assumed as the permanent faults. However, in many practical cases there exist faults which are not static. The fault that is developed and varying during the propagation of impulse voltage wave along the winding is termed as

*dynamic* fault. Moreover, there may be more than one fault in the winding. Identification of such faults are difficult than that for the static faults, i.e., simultaneous occurrence of faults of different topologies (static and dynamic). The dynamics of such fault conditions are complex. Therefore, there are growing interests among the researchers for identification of such faults. Authors in [82] have reported a scheme using cross-wavelet transform-based feature extraction and wavelet network-based classification for such problems. However, there is ample scope for the researchers to contribute in this direction.

**Acknowledgments** The authors acknowledge with thanks the immense contribution of Dr. Saibal Chatterjee, Prof. Prithwiraj Purkait, Dr. Chiranjib Koley and Mr. P. Rajamani towards the completion of this chapter. Most of the results discussed in this chapter are the outcome of their experimentation and research.

## References

1. Guide to the lightning and switching impulse testing of power transformers and reactors. IEC Standard, Publication 60076-4, 2002
2. Purkait P, Chatterjee A, Chakravorti S, Bhattacharya K (2003) Translationally adaptive fuzzy classifier for transformer impulse fault identification. IEE Proc Gen Transm Distrib 150(1):33–40
3. Hagenguth JH (1944) Progress in impulse testing of transformers. AIEE Trans 63:999–1005
4. Mamishev AV (1996) Analysis of high impedance faults using fractal technique. IEEE Trans Power Syst 11(1):435–440
5. Koley C (2007) Advanced techniques for identification of faults in transformers due to impulse stresses. Ph. D Thesis, Faculty of Engineering and Technology, Jadavpur University
6. Dey D (2009) Studies on intelligent digital signal processing tools for measurements in high voltage systems. Ph. D Thesis, Faculty of Engineering and Technology, Jadavpur University
7. Stenkvist E (1952) Study of fault detection and failure location during surge testing of transformers. CIGRE 1952, Supplement to paper no. 12-125
8. Satish L (1998) Short-time Fourier and wavelet transforms for fault detection in power transformers during impulse tests. IEE Proc Sci Meas Technol 145(2):77–84
9. Purkait P, Chakravorti S (2002) Time and frequency domain analyses based expert system for impulse fault diagnosis in transformers. IEEE Trans Electr Insul 9(3):433–445
10. Zhang Y, Ding X (1996) An artificial neural network approach to transformer fault diagnosis. IEEE Trans Power Deliv 11(5):1836–1841
11. Roy CK, Biswas JR (1994) Studies on impulse behavior of a transformer winding with simulated faults by analogue modeling. IEE Proc C 141:401–412
12. Abetti PA (1958) Bibliography on the surge performance of transformers and rotating machines. Trans AIEE 77:1150–1168
13. Malewski R, Poulin B (1988) Impulse testing of power transformers using the transfer function method. IEEE Trans Power Deliv 3:476–490
14. Vaessen PTM, Hanique E (1992) A new frequency response analysis method for power transformers. IEEE Trans Power Deliv 7:384–390
15. Abetti PA (1953) Transformer models for the determination of transient voltages. Trans. AIEE, pp. 468–480
16. Vaessen PTM (1988) Transformer model for high frequencies. IEEE Trans Power Deliv 3:1761–1768

17. de Leon F, Semlyen A (1994) Complete transformer model for electromagnetic transients. *IEEE Trans Power Deliv* 9:231–239
18. Ghahrepetian GB, Mohseni H, Moller K (1998) Hybrid modeling of inhomogeneous transformer winding for very fast transient over-voltage studies. *IEEE Trans Power Deliv* 13:157–163
19. Degeneff RC (1978) A method for constructing terminal models for single-phase n-winding transformers. In: *Proceedings of IEEE PES Summer Meeting*, paper no. A 78 539-9
20. Morched A, Marti L, Ottevangers J (1993) High frequency transformer model for the EMTP. *IEEE Trans Power Deliv* 8:1615–1626
21. Fergestad PI, Henriksen T (1974) Inductance for the calculation of transient oscillation in transformers. *IEEE Trans PAS*, vol. PAS-93, pp. 510–517
22. Chatterjee S (1994) Analog studies on impulse performance of a H.V. transformer winding with tapping coils. M.E.E Thesis, Jadavpur University, Kolkata
23. Grover FW (1973) Inductance calculations. ISA Publication, New York
24. Maxwell JC (1904) A treatise on electricity and magnetism, 3rd edn, vol. II. Clarendon Press, Oxford
25. Hagenguth JH, Meador JR (1952) Impulse testing of power transformers. *AIEE Trans* 71:697–704
26. Wang M, Vandermaar AJ, Srivastava KD (1999) Condition monitoring of transformers in service by the low voltage impulse test method. In: *Eleventh international symposium on high voltage engineering (Conf. Publ. No. 467)*, vol 1, pp 45–48
27. Purkait P, Chakravorti S (2001) Pattern classification of impulse faults in transformers by wavelet analysis. *IEEE Trans Electr Insul* 9(4):555–561
28. Purkait P, Chakravorti S (2003) Wavelet transform based impulse fault pattern recognition in distribution transformers. *IEEE Trans Power Deliv* 18(4):1588–1589
29. Koley C, Purkait P, Chakravorti S (2006) Wavelet aided SVM tool for impulse fault identification in transformers. *IEEE Trans Power Deliv* 21(3):1283–1290
30. Purkait P, Chakravorti S (2003) Impulse fault classification in transformers by fractal analysis. *IEEE Trans Dielectr Electr Insul* 10(12):109–116
31. Dey D, Chatterjee B, Chakravorti S, Munshi S (2008) Rough-granular approach for impulse fault classification of transformers using cross-wavelet transform. *IEEE Trans Dielectr Electr Insul* 15(5):1297–1304
32. Purkait P, Chakravorti S (2004) Can fractal techniques be used for analysis of impulse fault patterns in distribution transformers. *Int J Electr Power Syst Res (Elsevier)* 68:258–267
33. De A, Chatterjee N (2002) Recognition of impulse fault patterns in transformers using Kohonen's self-organizing feature map. *IEEE Trans Power Deliv* 17(2):489–494
34. Mathworks, Inc. Matlab Toolbox, <http://www.mathworks.com/access/helpdesk/help/toolbox/wavelet/wavelet.html>
35. Stark HG (2005) Wavelets and signal processing, 1st edn. Springer, Berlin
36. Daubechies I (1990) The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans Inf Theory* 36:961–1005
37. Ruessink BG, Coco G, Ranasinghe R, Turner IL (2006) A cross-wavelet study of alongshore non-uniform near shore sandbar behavior. In: *Proceedings of joint conference on neural networks*, pp. 4310–4317, Vancouver, Canada
38. Mallat S (1999) A wavelet tour of signal processing, 2nd edn. Elsevier, Amsterdam
39. Plett MI (2007) Transient detection with cross wavelet transforms and wavelet coherence. *IEEE Trans. Signal Processing* 55(5):1605–1611
40. Torrence C, Compo GP (1998) A practical guide to wavelet analysis. *J Am Meteorol Soc* 79(1):61–79
41. Li X, Yao X, Jefferys JRG, Fox J (2005) Computational neuronal oscillations using Morlet wavelet transform. In: *Proceedings of IEEE Annual Conference on Engineering in Medicine and Biology*, pp. 2009–2012, Shanghai, China
42. Pawlak Z (1991) Rough sets: theoretical aspects of reasoning about data. Kluwer, Boston



43. Peng JT, Chien CF, Tseng TLB (2004) Rough set theory for data mining for fault diagnosis on distribution feeder. *IEE Proc Gen Transm Distrib* 151:689–697
44. Rajamani P, Chakravorti S (2012) Identification of simultaneously occurring dynamic disc-to-disc insulation failures in transformer winding under impulse excitation. *IEEE Trans Dielectr Electr Insul* 19(2):443–453
45. Dey D, Chatterjee B, Chakravorti S, Munshi S (2010) Cross-wavelet transform as a new paradigm for feature extraction from noisy partial discharge pulses. *IEEE Trans Dielectr Electr Insul* 17(1):157–166
46. Chandaka S, Chatterjee A, Munshi S (2009) Cross-correlation aided support vector machine classifier for classification of EEG signals. *Expert Syst Appl* 36:1329–1336
47. Chandaka S, Chatterjee A, Munshi S (2009) Support vector machine employing cross-correlation for emotional speech recognition. *Measurements* 42:611–618
48. Rajamani P, Dey D, Chakravorti S (2011) Cross-correlation aided wavelet network for classification of dynamic insulation failures in transformer winding during impulse testing. *IEEE Trans Dielectr Electr Insul* 18:521–532
49. Rajamani P, Dey D, Chakravorti S (2010) Cross-correlation aided fuzzy c-means for classification of dynamic fault in transformer winding during impulse testing. *Electric Power Compon Syst* 38(13):1513–1530
50. Mandelbrot B (1983) *The fractal geometry of nature*. Freeman, New York
51. Lovejoy S, Schertzer D (1986) Scale invariance, symmetries, fractals and stochastic simulations of atmospheric phenomena. *Bull Am Meteorol Soc* 67(1):21–32
52. Voss RF (1985) Random fractals: characterisation and measurement. In: Pynn R, Skjeltrop A (eds) *Scaling phenomena in disordered systems*. Plenum press, New York, pp. 1–11
53. Keller JM, Chen S, Crownover RM (1989) Texture description and segmentation through fractal geometry. *Comput Vis Graph Image Process* 45:150–166
54. Niemeyer L, Pietronero L, Wisemann HJ (1984) Fractal dimension of dielectric breakdown. *Phys Rev Lett* 6(6):661–674
55. Fuji M, Watanabe M, Kitani I, Arii K, Yoshino K (1991) Fractal character of dc trees in polymethylmethacrylate. *IEEE Trans Electr Insul* 26:159–162
56. Satish L, Zaengl WS (1995) Can fractal features be used for recognizing 3-D partial discharge patterns? *IEEE Trans Dielectr Electr Insul* 2:352–359
57. Candela R, Mirelli G, Schifani R (2000) PD recognition by means of statistical and fractal parameters and a neural network. *IEEE Trans. Dielectr Electr Insul* 7(1):87–92
58. Sevcik C (1998) A procedure to estimate the fractal dimension of waveforms. *Complex Int J* 5. [http://www.complexity.org.au/ci\\_louise/vol05/sevcik/sevcik.html](http://www.complexity.org.au/ci_louise/vol05/sevcik/sevcik.html)
59. Esteller R, Vachtsevanos G, Echaz J, Litt B (2001) A comparison of waveform fractal dimension algorithms. *IEEE Trans Circuits Syst I* 48(2):177–183
60. Katz MJ (1988) Fractals and the analysis of waveforms. *Comput Biol Med* 18:145
61. Vanouplines P. Rescaled range analysis and the fractal dimension of pi. [Online]. Available: <http://homepages.vub.ac.be/~pvouplin/pi/compfdim.htm>
62. Krivda A, Galski E, Satish L, Zaengl WS (1995) The Use of Fractal Features for Recognition of 3-D Discharge Patterns. *IEEE Trans Electr Insul* 2(5):889–892
63. Ganapathiraju A, Hamaker JH, Picone J (2004) Applications of support vector machines to speech recognition. *IEEE Trans Signal Process* 52(8):2348–2355
64. Koley C, Purkait P, Chakravorti S (2007) SVM classifier for impulse fault identification in transformers using fractal features. *IEEE Trans Dielectr Electr Insul* 14(6):1538–1547
65. Wong KL (2004) Application of very-high-frequency (VHF) method to ceramic insulators. *IEEE Trans Electr Insul* 11(6):1057–1064
66. Ho KKL, Moody GB, Peng CK, Mietus JE, Larson MG, Levy D, Goldberger AL (1997) Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics. *Circulation* 96(3):842–848
67. Pincus SM (1991) Approximate entropy as a measure of system complexity. *Proc Natl Acad Sci USA* 88:2297–2301

68. De A, Chatterjee N (2001) Impulse fault diagnosis in power transformers using self-organising map and learning vector quantisation. *IEE Proc Gen Transm Distrib* 148(5):397–405
69. De A, Chatterjee N (2004) A fuzzy ARTMAP fault classifier for impulse testing of power transformer. *IEEE Trans Electr Insul* 11(6):1026–1036
70. Penrose R (2002) *The emperor's new mind*. Oxford University Press, Oxford
71. Anthony M, Biggs N (1992) *Computational Learning Theory*, vol 30. Cambridge tracts in theoretical computer science. Cambridge University Press, Cambridge
72. Rosenblatt F (1959) The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev* 65:386–408
73. Cristianini N, Shawe-Taylor J (2000) *Support vector machines*. Cambridge University Press, Cambridge
74. Mercer J (1909) Functions of positive and negative type and their connection with the theory of integral equations. *Phil Trans R Soc London, A* 209:415–446
75. Cabral JE, Pinto JP, Gontijo EM, Filho JR (2004) Fraud detection in electrical energy consumers using rough sets. In: *Proceedings of IEEE Conference on Systems, Man and Cybernetics*, pp 3625–3629
76. Kusiak A (2001) Rough set theory: a data mining tool for semiconductor manufacturing. *IEEE Trans Electron Packag Manuf* 24:44–50
77. Pal SK, Mitra P (2004) Case generation using rough sets with fuzzy representation. *IEEE Trans Knowl Data Eng* 16:292–300
78. Yao YY (1999) Rough sets, neighborhood systems and granular computing. In: *Proceedings of IEEE conference on electrical and computer engineering*, pp. 1553–1558, Canada
79. Nguyen HS (1997) Discretization of real value attributes: a boolean reasoning approach. PhD thesis, Dept. of Mathematics, Warsaw Univ., Poland
80. Purkait P (2002) Artificial intelligence technique based impulse fault diagnosis in transformers. Ph.D thesis, Faculty of Engineering and Technology, Jadavpur University
81. Purkait P, Chakravorti S (2003) Investigations on the usefulness of an expert system for impulse fault analysis in distribution transformers. *Electric Power Syst Res*, Elsevier 65(2):149–157
82. Rajamani P, Chakravorti S (2012) Identification of simultaneously occurring dynamic disc-to-disc insulation failures in transformer winding under impulse excitation. *IEEE Trans Dielectr Electr Insul* 19(2):443–453

Recent Trends in the Condition Monitoring of  
Transformers

Theory, Implementation and Analysis

Chakravorti, S.; Dey, D.; Chatterjee, B.

2013, XVI, 280 p. 205 illus., 105 illus. in color.,

Hardcover

ISBN: 978-1-4471-5549-2