

5

Advanced Signal Processing Techniques for Partial Discharge Measurement

5.1 Partial Discharge in Power Transformer

A large number of power transformers in many countries have approached or even exceeded their designed lives. For instance, National Grid, UK operates with approximately 780 transformers over 100 MVA and about half of these transformers were aged 40 years or more in 2010 [1]. In Hydro-Québec, Canada, the average age of the whole transformer fleet was over 30 years in 2012 [2]. In New South Wales, Australia, Endeavour Energy reported that almost half of its power transformers were aged 36 years or more in 2007 [3]. Any failure of a power transformer can be disastrous. Therefore, the condition of transformers needs to be monitored and evaluated continuously.

A transformer's condition is largely dependent on its insulation system. Many surveys have reported that insulation problems are one of the key sources of transformer failure [4–7]. Therefore, various diagnostic methods have been deployed in utilities for assessing transformer insulation conditions over the past decades. Among these methods, Partial discharge (PD) measurement is widely adopted due to its capability for providing continuous online condition monitoring of a transformer without disturbing its normal operation. Through the PD measurement, many types of insulation defect can be revealed. A summary of the insulation defects that can induce PD in a power transformer is given below [8, 9]:

- contamination of liquid and solid insulation due to moisture;
- air trapping in insulation due to poor vacuum of oil;
- gas bubble formation due to overheating and discharges;
- floating metal particles in oil;
- protrusion in HV structures;
- overstressed connections;
- poor earthing.

PD is a rather complicated phenomenon and stochastic in nature. Techniques for properly performing PD measurements of a transformer, effectively analyzing the measured PD signals, and subsequently making an informed condition assessment on a transformer's insulation system are much needed. These techniques are covered in detail in this chapter.

5.2 Overview of PD Analysis

The ultimate goal of PD analysis is to evaluate the characteristics (extensities, types, and locations) of PD events inside a transformer. This can be helpful in evaluating the health condition of a transformer insulation system. The major steps of PD signal analysis are shown in Figure 5.1.

After acquiring a signal from a field transformer by a PD measurement system such as an inductive measurement system using a high-frequency current transformer (HFCT) and a capacitive measurement system (i.e., IEC 60270 [10]), performing signal processing on the acquired signal is essential. This is mainly due to the contamination of an acquired signal by extensive noise. In PD measurements, common types of noise include discrete spectral interference (DSI), stochastic noise, and white noise, as listed in Table 5.1 [11–14].

Among these three types of noise, DSI is a narrow-band noise and the others are wide-band noise [14]. Depending on the source of noise, DSI consists of frequency components up to around 26 MHz for AM radio, 300 MHz for FM radio, and 1880 MHz for mobile communications. In general, this type of noise can have a higher amplitude compared to PD signals [15]. Since DSI is periodic with fixed frequencies, it may be removed by digital filtering, such as a notch filter [16]. However, it is a non-adaptive method since two parameters, forgetting factor and bandwidth, need to be predetermined [16, 17].

Stochastic noise is considerably difficult to remove, due to its similarity with PD signals. One approach to remove stochastic noise is to apply artificial neural networks (ANNs) to recognize different types of PD signal and noise [18]. However, this approach requires constructing a database that consists of fingerprints of different types of PD signal for training algorithms. It may be limited by the availability of fingerprints for proper training the algorithm.

White noise always exists during online PD measurement of transformers. It possesses a flat power spectrum with equal power at all frequencies. This type of noise can be suppressed by gating methods when they have lower amplitude compared to the PD signals.

The noise distorts PD signals, which can significantly influence the performance of the later steps of PD signal analysis. Therefore, the first step of PD signal analysis is de-noising, which aims at removing the noise and extracting the PD signals. De-noising can be achieved by using hardware, but software-based methods provide an adaptive means of removing various types of noise. Some typical software-based methods are signal decomposition, blind processing, and graphics-based techniques. These techniques will be detailed in this chapter.

In practice, signals acquired by PD measurement systems may consist of signals induced by a number of PD sources, which can co-exist in a transformer and discharge simultaneously. PD source separation is executed to separate the PD sources and reveal individual PD patterns related to each source. This can be performed by utilizing the time and frequency characteristics of PD signals to form distinctive clusters in a time/frequency (TF) map, with each cluster corresponding to a single PD source. This technique will be detailed in this chapter.

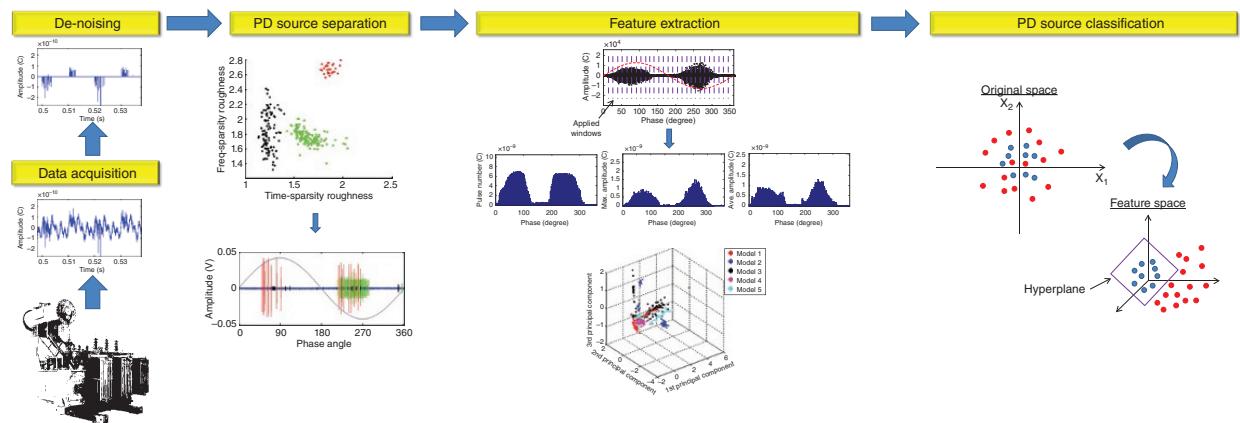


Figure 5.1 Major steps of PD signal analysis.

Table 5.1 Types of noise in PD measurements

Types of noise	Sources of noise
DSI	Communication systems Radio transmissions (e.g., AM and FM)
Stochastic noise	Switching operations (e.g., thyristor) Sparking Automobile spark plugs Electrical noise from detection circuit Arcing between metallic contacts Lightning
White noise	Amplifier Ambience

After PD source separation, individual PD sources can be described by a set of unique features (characteristic attributes), which are normally obtained from phase-resolved PD (PRPD) diagrams. A PRPD diagram consists of the distribution of PD pulse numbers, maximum PD pulse amplitude, and average PD pulse amplitude with reference to the phase angle of AC voltage. Features such as the first four moments (mean, standard deviation, skewness, and kurtosis) can be extracted from these distributions to act as fingerprints to a particular PD source. The features can also be extracted from decomposed signals (using different signal decomposition techniques) in different frequency scales to provide a more accurate representation of PD patterns. The effectiveness of different feature extraction techniques for assisting PD source classification will be examined in this chapter.

5.3 PD Measurement Methods

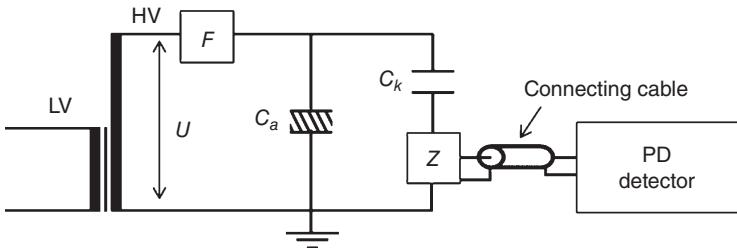
The occurrence of PD events is often accompanied by electromagnetic transients, chemical reactions, emission of sound, pressure waves, light, and heat [19]. These phenomena are utilized to detect the presence of PD events and acquire the PD signals for analysis.

Conventional PD measurement methods are described in IEC 60270 [10]. Some non-conventional methods are associated with electromagnetic and acoustic PD detection [19]. Table 5.2 summarizes various PD measurement methods.

The circuitry diagram of the conventional PD measurement system is depicted in Figure 5.2. In this figure, a HV power supply is required which should exhibit a low background noise level. An analog filter F can also be introduced to further reduce the noise. In the figure, C_a represents a test object. C_k is a coupling capacitor, which acts as a high impedance for power frequency of the voltage supply and a low impedance for high frequencies of PD signals generated from the test object [20]. The measuring impedance Z is used to convert PD current to voltage for signal acquisition by a PD detector [21]. The measuring impedance is either an RC or RLC circuit, in which the measured voltage pulses appear to be unidirectional or attenuated oscillatory waveforms, respectively. This conventional PD

Table 5.2 PD measurement methods

Conventional methods	Non-conventional methods
Narrow-band detection	Electromagnetic transients detection
Wide-band detection	High frequency (HF)/very HF (VHF) (3 ~ 300 MHz) Ultra HF (UHF) (300 MHz ~ 3 GHz) AE detection (10 ~ 300 kHz) Optical occurrences detection Chemical compounds detection

**Figure 5.2** Conventional PD measurement method (defined in IEC 60270).

measurement method is also called “direct coupling method” or “capacitive method.”

During PD measurements, the acquired PD signals may be contaminated by numerous noise sources. Therefore, PD measurements should be carried out in a frequency range where the noise level is as low as possible. According to IEC 60270 [10], the PD measurements can be performed in a narrow band or wide band. In a narrow-band system, the bandwidth Δf is recommended to satisfy

$$9 \text{ kHz} \leq \Delta f \leq 30 \text{ kHz} \quad (5.1)$$

In a wide-band PD system, transient PD pulses can be recorded with higher time resolution [11]. Based on IEC 60270, the bandwidth Δf of a wide-band system is recommended to satisfy

$$100 \text{ kHz} \leq \Delta f \leq 400 \text{ kHz} \quad (5.2)$$

Figure 5.3 shows a PRPD diagram acquired using a commercial product (Omicron MPD 600), which complies with IEC 60270.

Though the above capacitive system is capable of acquiring PD signals with relatively low noise, in practice the non-conventional PD measurements can provide a more flexible means for online PD measurements. In non-conventional PD measurements, a UHF antenna and inductive sensors such as a Rogowski coil and HFCT are adopted to acquire PD signals [22–25]. The acquired PD signals have various frequency components, depending on the types of PD source and insulation material [26]. A UHF PD measurement system is normally applied for the frequency range 300 MHz ~ 3 GHz (refer to Section 5.7 for a brief review of UHF).

Inductive sensors (HFCT, Rogowski coil) are clamped on a grounding wire or neutral of a transformer to measure high-frequency currents induced from PD.

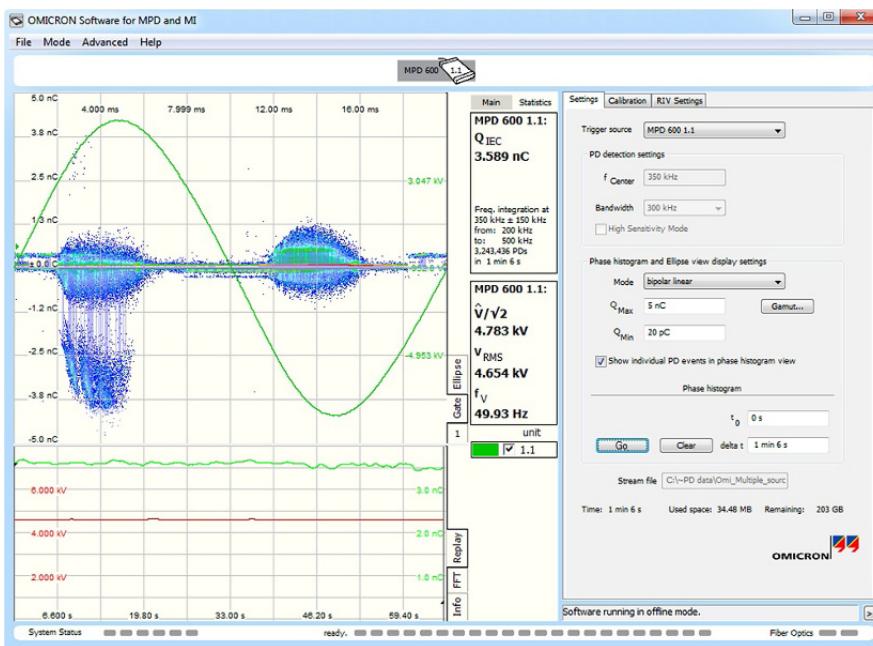


Figure 5.3 An example of an acquired PD signal using Omicron MPD 600, which complies with IEC 60270.

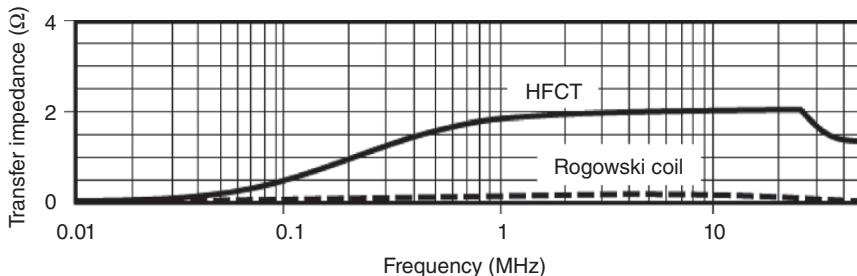


Figure 5.4 Frequency responses of Rogowski coil and HFCT (model: 140/100HC from HVPD).

They can also be clamped on HV leads, but their installations require a temporary disconnection of transformers from the power grid [27]. Rogowski coils are constructed with inductive coils wound on a non-magnetic core, such as plastic for flexibility on clamping. The required number of turns is decided largely by the expected inductance. Since the amplitudes of voltage outputs for PD signals are inductance dependent, changing the number of turns can adjust the amplitudes [22–24]. However, compared to HFCTs, Rogowski coils are considerably less sensitive and have a lower bandwidth. This is because HFCTs use a ferrite core [27].

Figure 5.4 shows a comparison of frequency responses (output voltages to input currents) between a Rogowski coil and a HFCT. From the figure, it can be seen that the sensitivity of the HFCT is much higher than that of the Rogowski coil. Though HFCTs have higher sensitivity, the use of a ferrite core implies that they

are nonlinear, and this may introduce some difficulties for calibration [28]. Nevertheless, using a split-core type of HFCT can ease installation in practice.

An acoustic method is often used for PD source localization. This is achieved by attaching a number of acoustic sensors on a transformer tank (refer to Section 5.7 for PD source localization by acoustic sensors) [29].

Except for the conventional PD measurement method (IEC 60270), there is a lack of standardized guidelines for the non-conventional methods. Therefore, a number of working groups from different organizations are now working on standardizing the PD measurements and the related interpretation techniques. Those working groups include CIGRE WG D1.37, IEC WG14, and CIGRE WG D1.29. For instance, the remit of CIGRE WG D1.29 includes a survey of PD measurement methods on their practical applications, suggestions of benchmarks for evaluating different PD measurement methods, collation of signatures from different PD sources, and determination of severity of PD sources [30].

With consideration of the above issues, this chapter is aimed at providing comprehensive PD signal analysis techniques for utilities to conduct PD measurements of transformers.

5.4 Advanced Signal Processing Techniques

This section presents a number of recently developed techniques for PD analysis, summarized as:

- signal decomposition;
- blind processing;
- TF map;
- PD pattern representation;
- other signal processing techniques.

5.4.1 Signal Decomposition

Since a PD signal is a type of transient signal, conventional time or frequency domain processing techniques may not represent the PD signals accurately. Consequently, these techniques cannot always effectively extract PD signals from extensive noise. Therefore, signal decomposition in joint time and frequency domains has been developed for PD signal analysis [31, 32]. In the following sections, wavelet transform (WT), empirical mode decomposition (EMD), ensemble EMD (EEMD), and MM are introduced.

5.4.1.1 Wavelet Transform (WT)

WT decomposes a signal into time and frequency domains with multi-resolutions. In WT, a wavelet is a waveform having zero-mean amplitude and short time duration [33]. This wave-like signal is named a wavelet function (also called a

“mother wavelet”). It generally satisfies $\int_{-\infty}^{\infty} \Psi(t) dt = 0$ and $\int_{-\infty}^{\infty} |\Psi(t)|^2 dt < \infty$,

where Ψ is the wavelet function [34]. Among various wavelet functions,

Daubechies (db), biorthogonal (bior), and Coiflets (coif) are adopted widely in PD signal processing [11–14, 35–41].

After selecting a proper wavelet function, WT applies shifting and scaling operations on the wavelet function, which in turn decompose the original signal into a series of wavelet coefficients [34]:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \Psi\left(\frac{t-b}{a}\right) \quad \text{s.t. } a, b \in \mathbb{R} \quad (5.3)$$

where a is the scale factor, adopted to compress and stretch the wavelet function; b is the translation factor, adopted to shift the wavelet function along the time axis. Through the scale and translation factors, WT can achieve both time and frequency representations of the signal.

In PD analysis, Discrete WT (DWT) is normally used in the form [34]

$$W_{j,k} = \sum_{n \in \mathbb{Z}} X(n) 2^{(-j/2)} \Psi(2^{-j} n - k) \quad (5.4)$$

where $X(n)$ is the discrete function of a signal, j and k are integers. For signal reconstruction, inverse DWT is performed as [34]

$$X(n) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} W_{j,k} \Psi_{j,k}(n) \quad (5.5)$$

5.4.1.2 Empirical Mode Decomposition (EMD) and Ensemble EMD (EEMD)

EMD is an adaptive decomposition technique. Without specifying any wavelet function as WT does, EMD transforms the original signal into a number of mono-component signals (termed intrinsic mode functions, IMFs) with different frequency scales, using the original signal's characteristic time scale [42]. The decomposed signals (i.e., IMFs) can later be used to rebuild the original signal precisely. To remove riding waves and smooth uneven amplitudes that might be embedded in IMFs, a shifting process is used in EMD. Therefore, an IMF can be treated as a mono-component signal [43].

It is assumed that $x(t)$ is a signal being processed in the sifting process. First, local extrema of $x(t)$ are defined. Then, cubic spline interpolations are performed on both maxima and minima to calculate the upper and lower envelopes $[x_U(t)$ and $x_L(t)$]. The mean of the above two envelopes is denoted $m(t) = (x_U(t) + x_L(t))/2$. The iteration for deciding IMF is $I_{ni}(t) = x(t) - m_{ni}(t)$, where $I_{ni}(t)$ is the n th IMF at the i th iteration. $I_{ni}(t)$ is the first IMF if it complies with the mono-component condition. Otherwise, the iteration will be continued on $I_{ni}(t)$. After extracting the first IMF, which embraces the original signal's highest-frequency component, the iteration will be continued on the residual $r_n(t) = x(t) - I_{ni}(t)$, replacing $x(t)$ by $r_n(t)$ and repeating the above process until the last IMF is obtained. A precise reconstruction $x_R(t)$ can be obtained by a linear combination:

$$x_R(t) = \sum_{(n=1)}^{(N-1)} I_{ni}(t) + r_N(t) \quad (5.6)$$

where N is the total number of IMFs and r is the last IMF or residual.

Ideally, an IMF is a signal with mono-component. However, due to the problem of mode mixing in the above sifting process, an IMF often embraces signal components with dramatically disparate frequency scales. The signal components with similar frequency scales may also reside in different IMFs [44]. Therefore, EEMD was proposed.

Through adding white noise to the original signal, EEMD forces the sifting process to make different frequency scales collate in the proper IMFs dictated by a dyadic filter bank [44]. After adding the white noise, EEMD extracts IMFs as EMD does. Once the extraction of the first ensemble is complete, a set of IMFs with noise is generated. Then, a new white noise with the same standard deviation of the amplitude as the previously added noise is injected into the original signal again and the EMD process is continued on this new noisy signal. The resulting IMFs are the average value of the ensembles on each IMF. Since white noise is a zero-mean random signal, the average value can cancel the errors produced by the noise.

5.4.1.3 Mathematical Morphology (MM)

MM applies mathematical operators between the original signal and a structure element, which is a predefined geometric shape with finite length as shown in Figure 5.5. Morphological features, which are the result of MM operation on the original signals, are then extracted.

Two basic operators in MM are erosion and dilation, which can be expressed by Minkowski subtraction and addition, respectively. By combining these two operators, another two operators (opening and closing) are created. Assuming x_n is a signal with a discrete function in the domain $X = (0, 1, 2, \dots, N - 1)$ and s_m is a structure element with a discrete function over the domain $S = (0, 1, 2, \dots, M - 1)$ [45], the four operators are [46]

$$(X \ominus S)_n = \min(X_{(n+m)} - S_m), \quad m \in 0, 1, 2, \dots, M - 1 \quad (5.7)$$

$$(X \oplus S)_n = \max(X_{(n-m)} + S_m), \quad m \in 0, 1, 2, \dots, M - 1 \quad (5.8)$$

$$(X \circ S)_n = (X \ominus S \oplus S)_n \quad (5.9)$$

$$(X \cdot S)_n = (X \oplus S \ominus S)_n \quad (5.10)$$

where \ominus , \oplus , \circ , and \cdot denote erosion, dilation, opening, and closing operators, respectively. Among the above four operators, opening and closing operators can preserve negative and positive peaks of signals, respectively. This means that the amplitudes of negative (positive) peaks of the original signals will be unchanged before and after execution of opening (closing) operation. Therefore, they are chosen for signal decomposition to maintain the integrity of decomposed signals. Also, a flat shape structure element is used due to easy implementation. Previous research proved that the shapes of structure elements do not affect signal

Figure 5.5 Types of structure element: (a) flat; (b) sinusoidal; and (c) triangular structure elements.



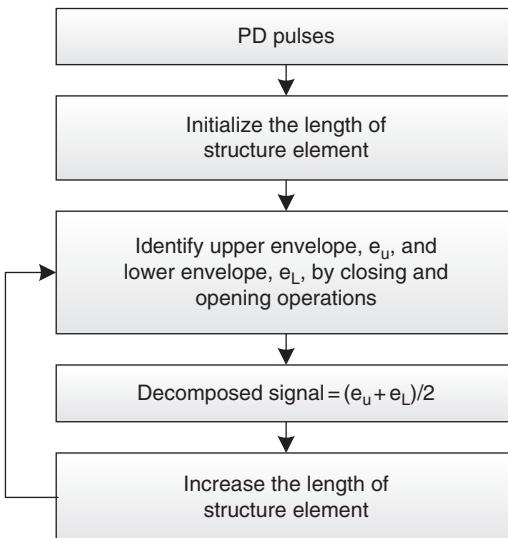


Figure 5.6 The MM-based signal decomposition method.

analysis much [47]. The flowchart of MM-based signal decomposition for individual PD pulses is shown in Figure 5.6.

As shown in this figure, the MM-based signal decomposition starts with an initialization of the structure element length, which will be used to construct upper and lower envelopes with closing and opening operators, respectively. The decomposed signal at the current length of the structure element is obtained by averaging the upper and lower envelopes. Then, the length of the structure element is increased by one and the above process is repeated. Since more pulses (including both shorter and longer intervals) are covered by the envelopes if the lengths of structure elements become longer, changing the lengths could alter the frequencies of envelopes. In this decomposition method, the lengths were set from 1 to 20 samples to provide 20 different frequency bands of PD pulses.

5.4.2 Blind Processing

Blind processing has been applied to diverse applications, but there are a few applications in PD analysis. In this section, blind equalization (BE), which is one of the blind processing techniques, is introduced.

5.4.2.1 Blind Equalization (BE) System

BE can recover a source signal without requiring a training process to identify the signal. Figure 5.7 depicts the schematic model of a single-input single-output (SISO) BE system (dotted line) [48]. BE only requires the received signal and some statistical parameters of the source signal. From the SISO BE system, the equalized signal is the output and it can reveal the source signal.

In this figure, the source signal is an independent and identically distributed (i.i.d.) non-Gaussian random signal having zero mean, second-order cumulant $\sigma^2 = E[s^2(k)]$, third-order cumulant $c_3 = E[s^3(k)]$, and fourth-order cumulant

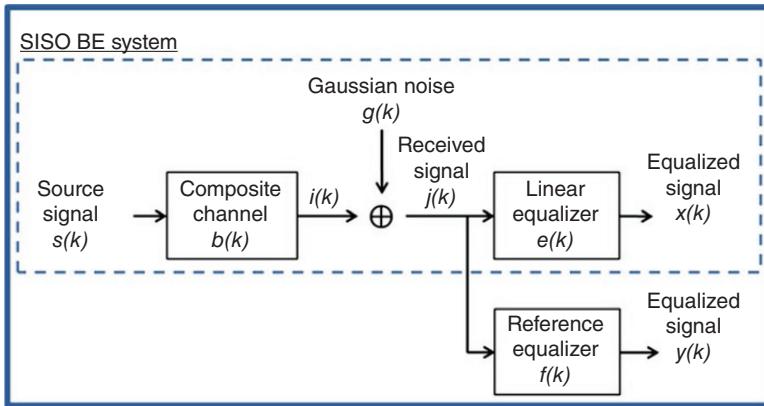
EVA system

Figure 5.7 Schematic models of SISO BE and EVA systems.

$c_4 = E[s^4(k)] - 3\{E[s^2(k)]\}^2$, where $E[\cdot]$ denotes expectation. It assumes the fourth-order cumulant is nonzero ($c_4 \neq 0$) [49].

The source signal in Figure 5.7 is influenced by an unknown composite channel, which is considered as a time-invariant system with finite impulse response (FIR) $b(k) = [b(0), \dots, b(n)]$, where n is the filter order. After the signal is linearly distorted, the distorted signal is mixed with zero-mean Gaussian noise, which is statistically independent of source signal, to produce a received signal

$$j(k) = b(k)^* s(k) + g(k) \quad (5.11)$$

where * denotes the convolution operator. For signal recovery, a linear equalizer with FIR $e(k) = [e(0), \dots, e(n)]$ is used to reconstruct the equalized signal

$$x(k) = e(k)^* j(k) \quad (5.12)$$

To obtain accurate signal recovery, a combined impulse response of equalization should only contain one nonzero unit magnitude component [50]

$$u(k) = b(k)^* e(k) = \delta(k - k_0) \quad (5.13)$$

where $u(k)$ is the combined system, k_0 is the time delay, and $\delta(k)$ is the Dirac delta function. After equalization, the obtained equalized signal is closed to the delayed source signal in terms of mean square error (MSE), defined as [49]

$$\text{MSE}(e, k_0) = E\{[x(k) - s(k - k_0)]^2\} = \min \quad (5.14)$$

5.4.2.2 Eigenvector Algorithm (EVA) for Blind Equalization (BE)

The structure of EVA [49] is similar to that of BE (Figure 5.7), but a reference equalizer is added. This reference equalizer can generate implicit signals as a reference for an iteration process to find an optimal equalized signal. EVA seeks the maximum fourth-order cross-cumulant in the equalization:

$$\max |c_4^{xy}(0,0,0)| \quad \text{s.t. } r_{xx}(0) = \sigma_s^2 \quad (5.15)$$

Replacing $x(k)$ by $j(k)$ [i.e., $x(k) = j(k)^* e(k)$], equation (5.15) becomes

$$\max \left| e^* C_4^{yj} e \right| \quad \text{s.t. } e^* R_{jj} e = \sigma_s^2 \quad (5.16)$$

where $*$ denotes the conjugate transpose function, R_{jj} is the autocorrelation matrix of the received signal, and C_4^{yj} is an $(n+1) \times (n+1)$ Hermitian cross-cumulant matrix:

$$C_4^{yj} = E\{[y(k)]^2 j j^*\} - E\{[y(k)]^2\} E\{j j^*\} - E[y(k)j] E[y^*(k)j^*] \\ - E[y^*(k)j] E[y(k)j^*] \quad (5.17)$$

The above approach makes use of a statistical property (i.e., the fourth-order cumulant) of both the output of the reference equalizer and the received signal. By optimizing equation (5.16), this leads to a generalized eigenvector problem

$$C_4^{yj} e_{\text{EVA}} = \lambda R_{jj} e_{\text{EVA}} \quad (5.18)$$

The coefficient vector $e_{\text{EVA}} = [e_{\text{EVA}}(0), \dots, e_{\text{EVA}}(n)]^T$ can be obtained by selecting the eigenvectors of $R_{jj}^{-1} C_4^{yj}$ with the largest eigenvalue λ .

EVA is implemented by an iteration process as follows [49]:

- Step 1. *Initialization.* Set reference equalizer $f^{(0)}(k) = \delta(k - \lfloor n/2 \rfloor)$ and iteration number $i = 0$, where $\lfloor \cdot \rfloor$ is the floor function and n is the equalizer length.
- Step 2. Estimate autocorrelation matrix R_{jj} .
- Step 3. Find $y(k) = j(k)^* f^{(i)}(k)$ and estimate cross-cumulant matrix C_4^{yj} .
- Step 4. Find coefficient $e^{(i)}_{\text{EVA}}$ by choosing eigenvectors of $R_{jj}^{-1} C_4^{yj}$ with the largest eigenvalue λ in $C_4^{yj} e^{(i)}_{\text{EVA}}(k) = \lambda R_{jj} e^{(i)}_{\text{EVA}}(k)$.
- Step 5. Let the reference equalizer equal e_{EVA} [i.e., $f^{(i)}(k) = e^{(i)}_{\text{EVA}}(k)$], increase $i = i + 1$, and go to Step 3 until i reaches its maximum value (e.g., $i = l$).

The above signal decomposition and blind processing techniques are used mainly for PD signal de-noising. In the next section, a TF map is detailed for PD source separation.

5.4.3 Time/Frequency (TF) Map

Conventionally, a TF map is constructed based on equivalent time length and bandwidth of individual PD pulses for multiple PD source separation [51–54]. However, the TF map may not always capture accurately the characteristics embedded in the shape of a PD pulse. Therefore, in this chapter, a TF sparsity map is proposed [32].

5.4.3.1 Equivalent Time and Bandwidth

A TF map by means of equivalent time length and bandwidth of a PD pulse is a widely adopted approach for separating multiple PD sources [51–54]. For a PD pulse p , with N samples, a time reference is defined as

$$t_0 = \frac{\sum_{i=0}^N t_i p_i(t_i)^2}{\sum_{i=0}^N p_i(t_i)^2} \quad (5.19)$$

where t_i denotes the time instant of acquiring the i th sample. The equivalent time length of a PD pulse is defined as

$$T^2 = \frac{\sum_{i=0}^N (t_i - t_0)^2 p_i(t_i)^2}{\sum_{i=0}^N p_i(t_i)^2} \quad (5.20)$$

If P is the frequency value obtained from Fourier transform, the equivalent bandwidth of a PD pulse is defined as

$$F^2 = \frac{\sum_{i=0}^{N/2} f_i^2 |P_i(f_i)|^2}{\sum_{i=0}^{N/2} |P_i(f_i)|^2} \quad (5.21)$$

where f_i denotes the frequency at the i th sample.

A TF map is constructed by projecting T^2 and F^2 of each PD pulse onto a two-dimensional (2D) plane. On this plane, the projected data points, which belong to PD pulses having similar shape, will merge together to form a distinct group. Thus, a number of groups can be formed and each group corresponds to a unique PD source. This method provides a mechanism for describing an individual PD pulse in terms of time and frequency representations. However, as shown in equations (5.20) and (5.21), pulse shapes cannot always be represented precisely, since only the overall amplitude/magnitude in the time/frequency domain of a PD pulse is considered. As a consequence, sometimes the PD pulses generated by the same PD source, which are supposed to have similar values of T^2 and F^2 and locate closely on the TF plane, can be scattered in different groups. In the next section, an improved version of the TF map, the TF sparsity map, is described.

5.4.3.2 Time/Frequency Sparsity Map

Figure 5.8 presents a conceptual diagram [Figure 5.8(a)] and a flowchart of the TF sparsity map [Figure 5.8(b)].

The construction of a TF sparsity map has three stages. The first stage is extracting PD pulses from measured signals at one acquisition. The second stage is the MM-based signal decomposition for decomposing the extracted PD pulses in time and frequency domains. The third stage is the calculation of sparsity values, sparsity trends, and roughness values for constructing a TF sparsity map.

Sparsity is defined as

$$S(p) = \frac{\sqrt{1/N \sum_{i=1}^N p^2(i)}}{\sqrt{1/N \sum_{i=1}^N |p(i)|}} = \sqrt{N} \frac{\sqrt{\sum_{i=1}^N p^2(i)}}{\sqrt{\sum_{i=1}^N |p(i)|}} \quad (5.22)$$

where p is the PD pulse with N samples. $\sqrt{\sum_{i=1}^N p^2(i)}$ and $\sqrt{\sum_{i=1}^N |p(i)|}$ denote norms.

Figure 5.9 shows sparsity values calculated from time and frequency domains of the decomposed signals. In this figure, the signals were generated by three

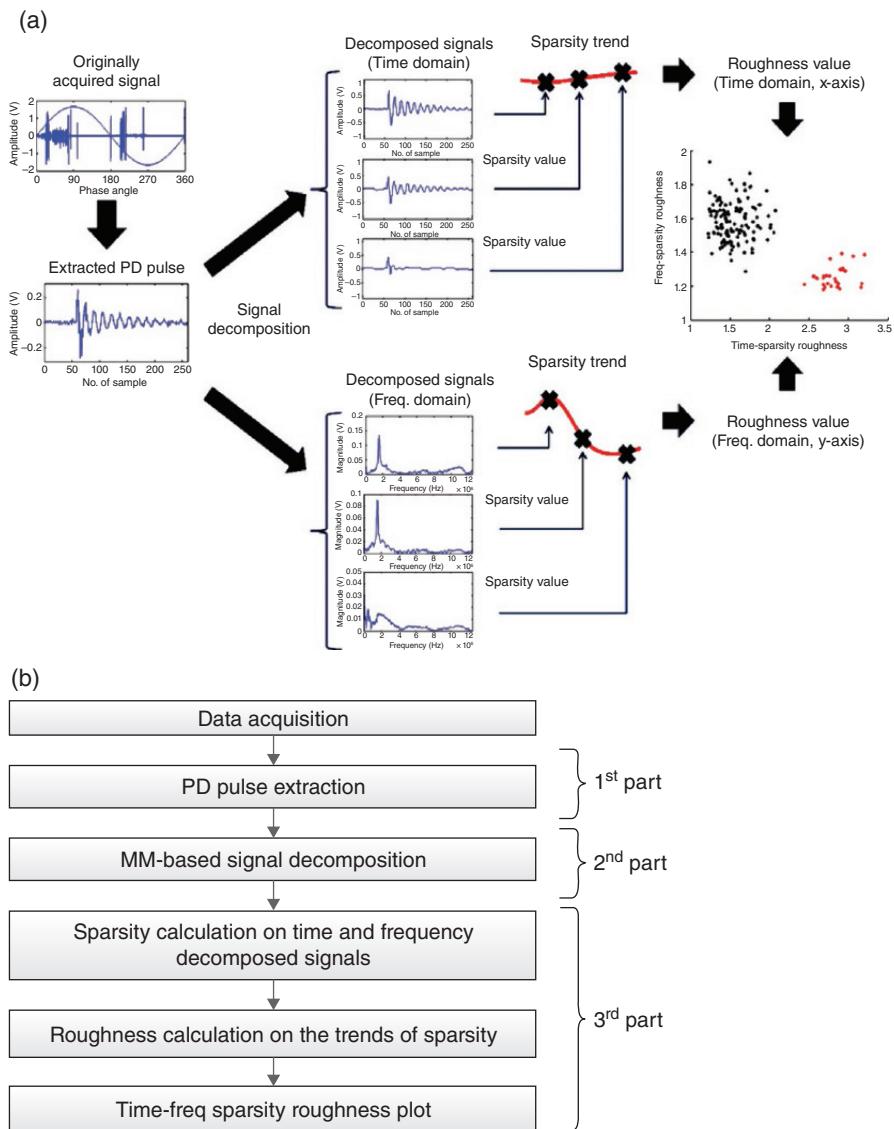


Figure 5.8 TF sparsity map for multiple PD source separation: (a) conceptual diagram and (b) flowchart. (See insert for color representation of the figure.)

different PD sources including discharge in transformer oil, corona, and surface discharge. After considering sparsity values for all decomposed signals, each PD source exhibits a particular trend of sparsity values (solid gray lines). This trend is named the sparsity trend. In the figure, solid gray lines represent the average of three different tendencies of the trends, with each trend related to one type of PD source.

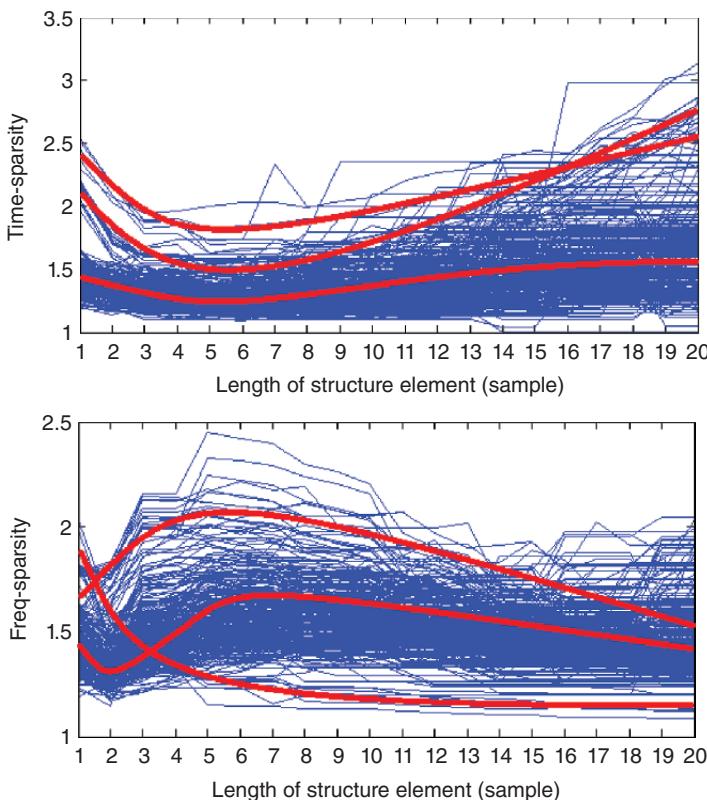


Figure 5.9 Sparsity trends for time (top) and frequency (bottom) domains of decomposed signals. Each gray line represents a sparsity trend of each PD pulse. Solid gray lines represent the average of three different tendencies of the trends. (See insert for color representation of the figure.)

To quantify the sparsity trends, roughness, which is defined as an average of absolute values, can be adopted as

$$R = \frac{1}{L} \sum_{i=1}^L |t(i)| \quad (5.23)$$

where $t(i)$ is the sparsity trend for the i th length of the structure element and L is the number of lengths of structure elements.

The values of roughness are used to describe variation of sparsity trends in time and frequency domains to represent PD pulses. Therefore, each PD pulse is represented by two roughness values (one in the time domain and another in the frequency domain). After projecting roughness values of each PD pulse onto a 2D map, the TF sparsity map is obtained. After PD pulses are represented by the roughness values and the values are projected onto a TF sparsity map, each group of PD pulses on the TF sparsity map represents a single PD source. Then, a density-based spatial clustering of application with noise (DBSCAN) is used for clustering PD pulses on the TF sparsity map [55, 56]. DBSCAN performs clustering

based on the density of data points in a volume, which is decided by the minimum number of data points N_{\min} and neighborhood radius ϵ . $N_{\min} = 2$ is set to ensure even a small amount of PD pulses belonging to the same PD source can be clustered. For a data set $X(m \times 2)$, its neighborhood radius is defined as [57]

$$\epsilon = \sqrt{\frac{\text{prod} \left(\max(X) - \min(X) \right) N_{\min} \Gamma(2)}{m\pi}} \quad (5.24)$$

where prod is the product of the array and Γ is the gamma function,

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

After PD source separation, each individual PD source can have its particular pattern (mixture PD sources also have their own patterns). This pattern can be represented by various methods, which are introduced in the next section.

5.4.4 PD Pattern Representation

5.4.4.1 Phase-Resolved Pulse Sequence

In PD measurements, PD activity is represented as a sequence of pulses with their intensity and time of occurrence. For PD source classification, PD patterns should exhibit a clear consistency and be independent of the types of PD sensor used in the measurements. Conventional PRPD diagrams can be suitable for PD source classification, given that test arrangements are standardized and the relation between test voltages and discharge activities is known *a priori*. However, PRPD diagrams can be influenced by a number of factors.

PD signals are wide-band signals and different types of PD sensor can have different measuring bandwidths. This can influence PD pulse resolutions. Noise resistivity and sensitivity also depend strongly on the measuring bandwidths. Therefore, discrepancy may exist among PRPD diagrams obtained from different types of PD sensor for the same PD source.

Therefore, a phase-resolved pulse sequence (PRPS) diagram has been investigated as a complement to the PRPD diagram for PD source classification [48]. Figure 5.10(a) shows parameters that are used for pulse sequence (PS) analysis in a single power cycle. There are a number of methods to analyze PS. In Ref. [58], PS was analyzed based on voltage differences Δh between sequential PD pulses. This indicates the amount of voltage required to ignite a PD pulse. In Ref. [59], a PRPS diagram was developed based on the slope $m = \Delta h / \Delta w$ to examine voltage gradients for initiating a PD pulse. Although these methods can improve the consistency of PD pattern analysis, the amplitudes of applied voltages are required.

Figure 5.10(b) illustrates a recently developed PRPS diagram [48], which uses the phase angle between two consecutive PD pulses as a quantitative measure (“amplitude”). It does not require the amplitude values of applied voltages. Each power cycle is divided into 360 windows and the PD pulse amplitudes are averaged in each window. Then, in each cycle, the differences in pulse phase angles are

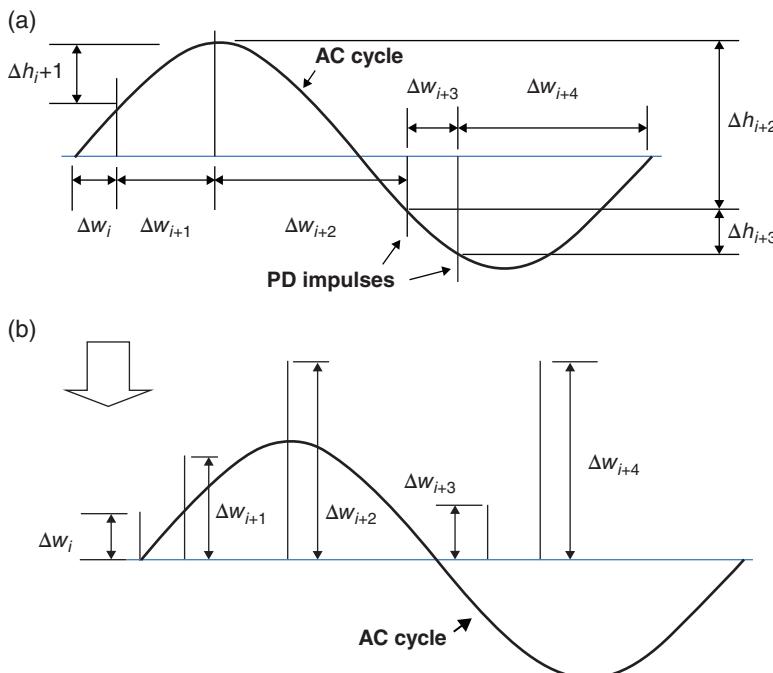


Figure 5.10 (a) Parameters of PS analysis and (b) the recently developed PRPS diagram.

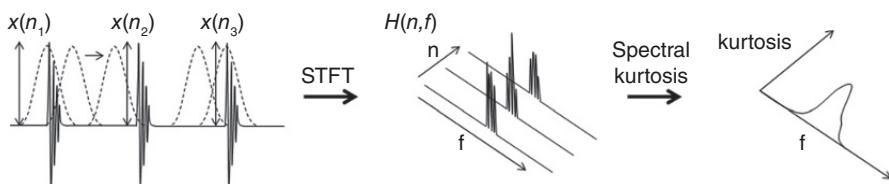


Figure 5.11 Interpretation of spectral kurtosis [62].

calculated and converted into amplitudes in a PRPS diagram. Finally, these amplitudes are normalized to the $[-1, 1]$ interval.

5.4.4.2 Kurtogram Construction Using Spectral Kurtosis

A kurtogram, which is constructed by spectral kurtosis, is another PD pattern representation method. Spectral kurtosis describes the impulsiveness of a signal in the frequency domain [60]. To compute spectral kurtosis, a short-time Fourier transform (STFT) with a window shifting along the time domain of a PD signal is used to obtain a complex envelope function $H(n, f)$ (Figure 5.11) [61]. The complex envelope function can be defined as the decomposition of a non-stationary random signal $x(n)$, where $n \in \mathbb{Z}$, which matches the characteristic of a PD signal by the Wold–Cramer representation [61].

$$x(n) = \int_{-1/2}^{+1/2} H(n,f) e^{j2\pi fn} dZ_x(f) \quad (5.25)$$

where $dZ_x(f)$ is a spectral increment. Taking the square of the complex envelope function represents the power spectrum at each time instant (i.e., n) of a PD signal. The power spectrum of the whole time period of the signal is obtained by taking the average of all the power spectra [62].

A kurtosis value for each frequency scale (right graph of Figure 5.11) is obtained by calculating the fourth moment of $H(n,f)$ [as shown in the numerator of equation (5.26)] at each time position and then taking the average of these values. The final kurtosis value is a normalization of the average value by the square of the mean square value [62]. In equation (5.26), the minus 2 makes the value of kurtosis zero for a Gaussian signal. The above process is performed based on a window shifting along a PD signal to find the spectral kurtosis for one level. By repeating the process for different sizes of window, different levels of spectral kurtosis are formed and a kurtogram is constructed.

$$K(f) = \frac{\langle |H(n,f)|^4 \rangle}{\langle |H(n,f)|^2 \rangle^2} - 2 \quad (5.26)$$

In addition to the aforementioned signal processing techniques, there are several more recently developed techniques for PD signal analysis. These techniques are presented in the next section and their applications are shown in later sections of this chapter.

5.4.5 Other Signal Processing Techniques

5.4.5.1 Fractal Dimension and Entropy

Fractal dimension and entropy are used to remove stochastic noise embedded in the measured PD signal. The fractal dimension is referred to as the changes of a pattern's details with respect to the scales used for measuring this pattern. A number of methods including box-counting, variance, and spectral methods have been used to calculate fractal dimension [63]. In this chapter, box counting is introduced for its simplicity and efficiency [64, 65]. The fractal dimension (FD) of a signal in Euclidean space is defined as [66]

$$FD = \lim_{\epsilon \rightarrow 0} \frac{\log N(\epsilon)}{\log(1/\epsilon)} \quad (5.27)$$

where $N(\epsilon)$ is the least number of boxes with side length ϵ to cover the signal.

A PD source caused by a particular insulation defect can exhibit a unique PD pattern. Owing to the fractal dimension's ability to represent various patterns and describe complex shapes using different sizes of box (i.e., side lengths of boxes), it has been applied to feature extraction and subsequent PD source recognition [66, 67]. In Ref. [65], the fractal dimension was used to quantify the severity of stochastic noise when the noise and PD signals were located in different power cycles. After the quantification, the noise and PD signals can have various values of fractal

dimension. This enables the separation of the noise and PD signals and subsequent removal of the noise.

When the noise and PD signals appear in the same power cycles, fractal dimension alone cannot extract PD signals. To deal with this situation, entropy is introduced [65]. Entropy measures the disorder in a random variable. A larger value of entropy relates to more chaotic data [68]. For a signal $X = [x_1, x_2, \dots, x_n]$, its entropy $H(X)$ is defined as

$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (5.28)$$

where $p(x_i)$ is the probability mass function of x_i .

5.4.5.2 Pre-whitening

The pre-whitening process is used to enhance the blind processing technique in PD signal de-noising. It is performed based on an autoregressive (AR) model, which provides a linear prediction of the noise in the acquired signals in PD measurements. The predicted value (or amplitude) of the noise $\hat{y}(n)$ at the current time step is calculated based on the weighted sum of p values (or amplitudes) of an acquired signal $x(n)$ at previous time steps [62]:

$$\hat{y}(n) = - \sum_{k=1}^p a(k)x(n-k) \quad (5.29)$$

where k denotes previous time steps from 1 to p , and the $a(k)$ are coefficients which can be calculated from the solution of Yule–Walker equations. The actual value of the acquired signal $x(n)$ at the current time step is the summation of the predicted value of the noise $\hat{y}(n)$ and a residual part $e(n)$ [62]:

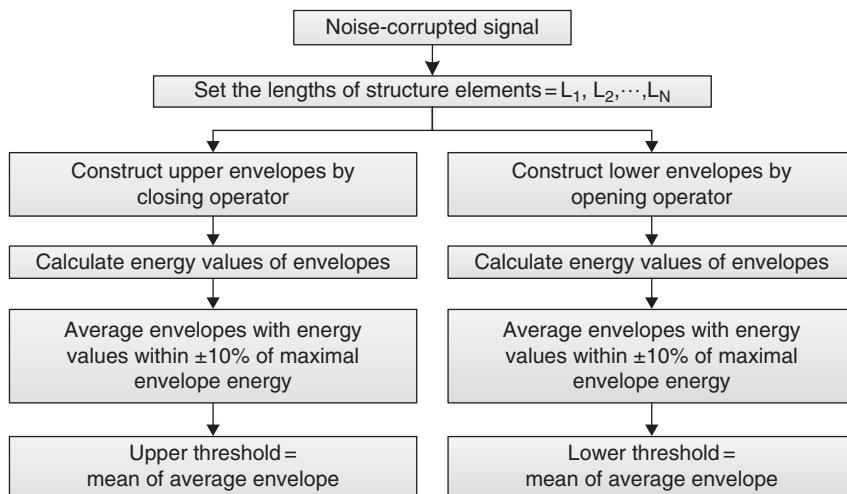
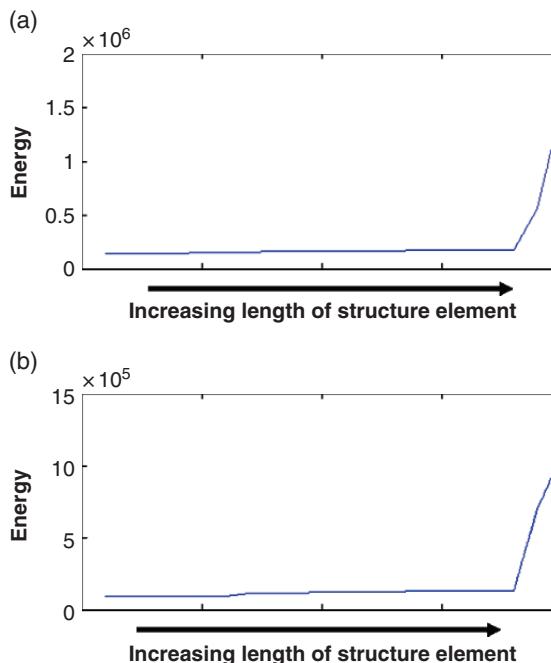
$$x(n) = \hat{y}(n) + e(n) \quad (5.30)$$

The aforementioned process is based on a linear prediction of the noise. After subtracting the predicted value of noise $\hat{y}(n)$ from the acquired signal $x(n)$, the term $e(n)$ in equation (5.30) contains the remaining stationary white noise and transient pulses, which are related to PD pulses. Thus, it can be considered that the impulsiveness of a signal (i.e., transient PD pulses) is enhanced [69], since a large portion of noise can be removed by the above pre-whitening process. In equation (5.30), since $e(n)$ is considered as a pre-whitened signal, this process is named a pre-whitening process.

5.4.5.3 Mathematical Morphology for Thresholding in PD Signal Extraction

MM is used not only for signal decomposition, but also for defining thresholds in PD signal extraction. Based on the theory of MM, AMT is developed to automatically define the lengths of flat structure elements for creating thresholds, as shown in Figure 5.12 [31].

A series of lengths L_1, L_2, \dots, L_N , is selected for flat structure elements to form upper and lower envelopes by using closing and opening operators, respectively. The lengths are multiples of the sample number in one AC power cycle divided by

**Figure 5.12** AMT method for thresholding.**Figure 5.13** Energy values of (a) upper and (b) lower envelopes.

32 to maintain a reasonable computational time. After generating the envelopes, energy values – defined as the sum of amplitudes squared for each sample of envelopes – are calculated. An example of the energy values for both upper and lower envelopes using flat structure elements with different lengths is shown in Figure 5.13.

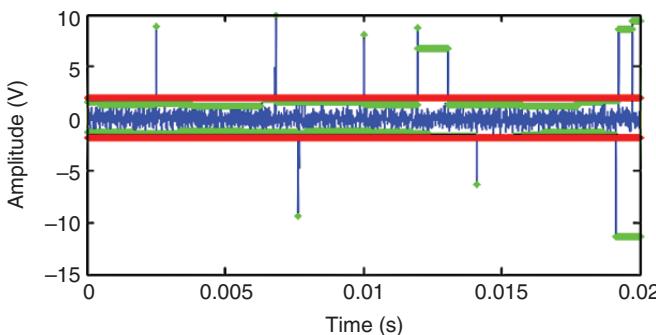


Figure 5.14 Results of AMT for random pulses [average upper and lower envelopes (gray dots) and optimal thresholds (dark gray lines)].

It can be seen from this figure that the energy values increase slowly with increasing lengths of structure elements until the lengths attain relatively large values. The relatively small increase of energy values at the initial and middle stages in Figure 5.13 are due to the envelopes covering pulses with either small intervals or large intervals but small amplitudes. However, when the lengths become longer, the envelopes start to cover pulses with both large intervals and amplitudes. This leads to a significant increase in energy values at the later stage. After calculating the energy values of envelopes, the envelopes whose energy values are equivalent to $\pm 10\%$ of maximal envelope energy are averaged to form an overall envelope. Thresholds are obtained by taking the mean value of each average envelope and used to remove white noise, as shown in Figure 5.14.

After describing the advanced signal processing techniques, in the next section their applications in PD signal analysis are provided with several case studies. This is to verify the effectiveness of these techniques in practice.

5.5 Application of Advanced Signal Processing Techniques for PD Analysis

As shown in Figure 5.1, PD signal analysis includes (1) PD signal de-noising, (2) PD source separation, and (3) feature extraction for PD source classification. In this section, the application of the signal processing techniques on these three steps is demonstrated.

5.5.1 PD Signal De-noising

In this section, the de-noising techniques are focused on the three major types of noise: white noise, DSI, and stochastic noise.

5.5.1.1 De-noising of Low-Frequency DSI and White Noise

DSI can have a wide range of frequency bands generated from communication systems and radio transmissions. In the following sections, a PD signal de-noising method is presented to remove low-frequency DSI and white noise.

5.5.1.1.1 De-noising Method

To remove low-frequency DSI and white noise, a signal decomposition-based PD signal de-noising method can be adopted, as shown in Figure 5.15 [31].

Firstly, EEMD is executed to decompose a noise-corrupted signal. The ensemble number was set to 300 without sacrificing computational time, while the decomposition performance can be maintained; the amplitude of the injected white noise was set to 0.2 standard deviations of the signal, as suggested in Ref. [44] to maintain the decomposition performance. According to Ref. [70], an IMF generally satisfies the Cauchy condition after five iterations in the sifting process. More iterations may not change the IMFs significantly. Thus, 10 iterations were used to guarantee the stability and convergence of IMFs. After decomposition by EEMD, kurtosis was applied to select the IMFs with PD pulses due to the sensitivity of transient signals. In other words, a larger kurtosis value refers to a signal consisting of abruptly changed pulses, while a smaller kurtosis value refers to a slowly fluctuating signal or a signal consisting of evenly distributed amplitudes (e.g., white noise and periodic signal). The kurtosis k is defined as

$$k = \frac{\sum_{i=1}^L (I_i - \bar{I})^4}{(L-1)\sigma^4} \quad (5.31)$$

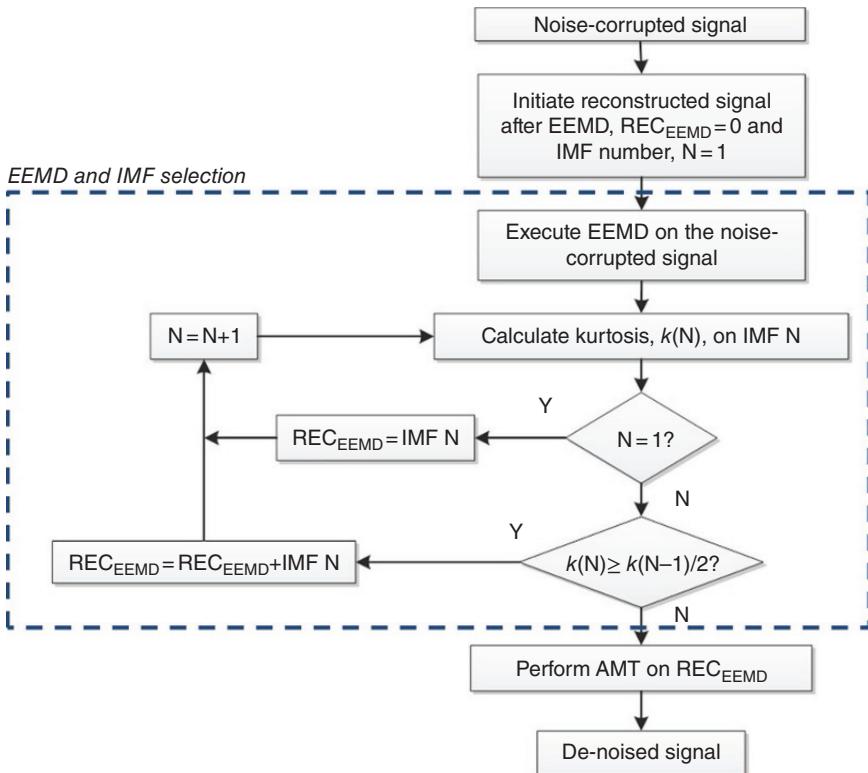


Figure 5.15 The signal decomposition-based (EEMD-based) PD signal de-noising method.

where \bar{I} , L , and σ are the mean, length, and standard deviation of an IMF, respectively.

An IMF selection process starts from the first IMF, which comprises the highest-frequency component of a noise-corrupted signal. If the kurtosis of an IMF is suddenly dropped to half that of the previous IMF, this implies that the IMF with the lower kurtosis value consists of fluctuations in low-frequency noise and/or white noise. Therefore, it is considered as pure noise without containing any PD pulses and discarded. Otherwise, it is added to the previous IMF for reconstruction and the selection will be continued on the remaining IMFs. Such kurtosis-based IMF selection can eliminate the low-frequency IMFs generated from the end effect of EMD.

After the above selection process, signal reconstruction can be performed. Since some noise with the same frequency scales of the selected IMFs may be included in the reconstructed signal, the AMT can be employed to create upper and lower thresholds of the signal. A de-noised signal can then be produced by applying hard thresholding.

5.5.1.1.2 De-noising Results

To evaluate the performance of the signal decomposition-based PD signal extraction, three measures including MSE M , correlation coefficient γ , and pulse number error PE were adopted and defined [31]:

$$M = \frac{1}{L} \sum_{i=1}^L (O_i - D_i)^2 \quad (5.32)$$

$$\gamma = \frac{\sum_{i=1}^L (O_i - \bar{O})(D_i - \bar{D})}{\sqrt{\sum_{i=1}^L (O_i - \bar{O})^2 \sum_{i=1}^L (D_i - \bar{D})^2}} \quad (5.33)$$

$$PE = |P_o - P_d| \quad (5.34)$$

where L is the length of signal, O denotes the original signal, D denotes the de-noised signal, \bar{O} and \bar{D} are the mean values of O and D , respectively, P_o is the number of pulses in the original signal, and P_d is the number of pulses in the de-noised signal. The measures M and γ are used to indicate the similarity of PD patterns, whereas PE is for examining the integrity of PD pulses.

The de-noising methods introduced in the above sections are used to remove both DSI and white noise in the following case studies. In the case studies, both DSI and white noise are injected into the measured PD signal. The DSI consists of three sinusoidal signals with randomly selected amplitudes and frequencies. The frequencies of DSI are up to 1 kHz to simulate signal fluctuations that could appear in online PD measurements. White noise is simulated with different SNRs. The de-noising results are compared using three different types of signal decomposition technique: EMD, EEMD, and DWT. The results and evaluations of the three methods are shown in Figure 5.16 and Table 5.3.

The DWT-based method applied a commonly used threshold to wavelet coefficients with five decomposition levels [11]. The EMD-based method used the same approach as the EEMD-based method (Figure 5.15), but the decomposition

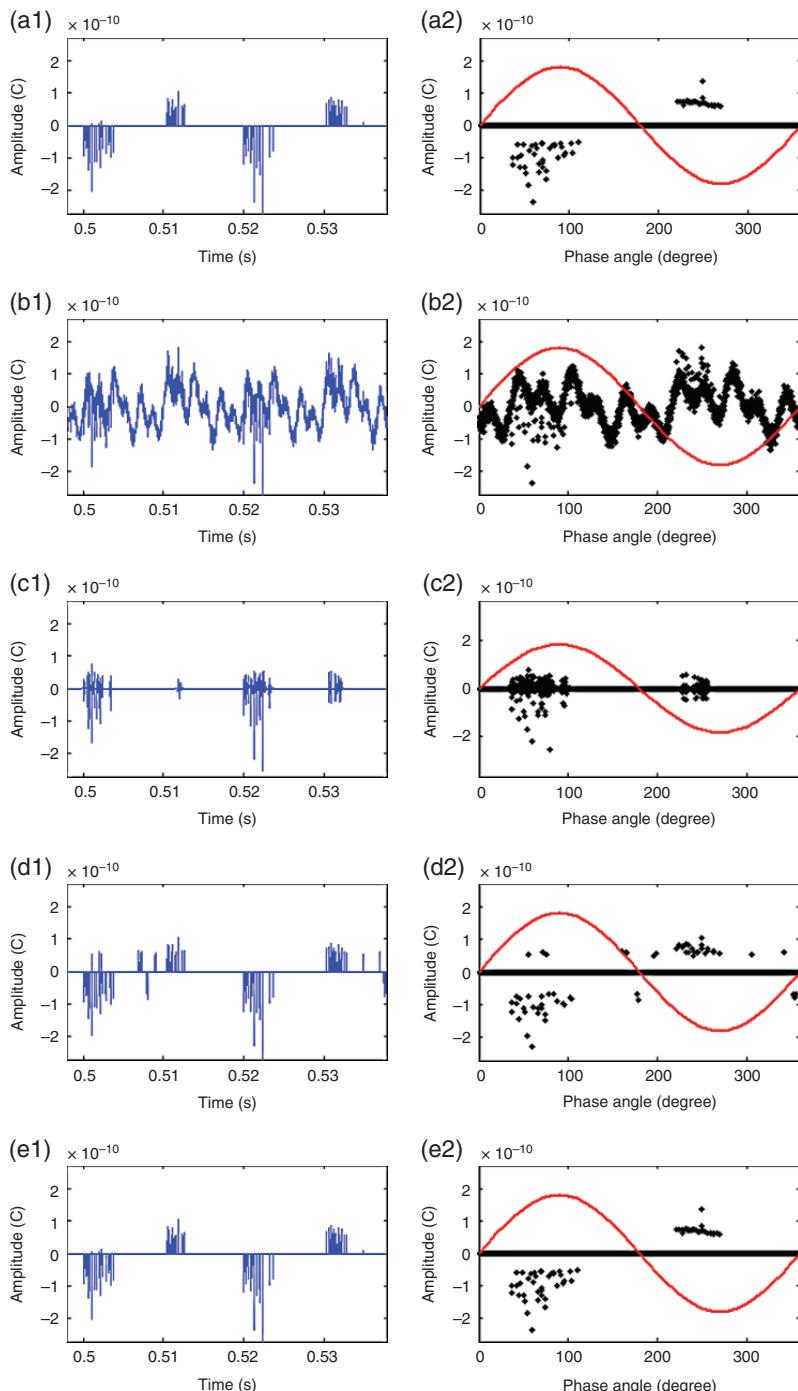


Figure 5.16 De-noising results of internal discharge. (a1)–(e1) Original signal, noise-corrupted signal ($\text{SNR} = -5 \text{ dB}$), and de-noised signals of DWT (db5), EMD + AMT, and EEMD-based method. (a2)–(e2) Corresponding PRPD diagrams. (See insert for color representation of the figure.)

Table 5.3 Comparisons of measures on PD signals acquired from PD source models

Model no.	Comparison with DWT			Comparison with EMD + AMT		
	$\downarrow M$ (%)	$\uparrow \gamma$ (%)	$\downarrow PE$ (%)	$\downarrow M$ (%)	$\uparrow \gamma$ (%)	$\downarrow PE$ (%)
1	74	32	99	44	6	0
2	81	14	99	87	24	92
3	81	29	96	81	27	40
4	91	23	96	87	12	14

Note: Model 1 – internal discharge, Model 2 – discharge due to floating particles, Model 3 – discharge in transformer oil, Model 4 – corona.

of EEMD was replaced by EMD. This is termed EMD + AMT. Figure 5.16 presents both time and PRPD diagrams of original signal, noise-corrupted signal, and de-noised signals in two power cycles for internal discharge. Table 5.3 compares the de-noising performances of DWT, EMD + AMT, and the EEMD-based method on signals ($SNR = -5$ dB) obtained from different PD source models for 100 power cycles. The signals were acquired from the capacitive measurement system.

From Figure 5.16(c1), it can be seen that the polarities of most pulses after de-noising by DWT are blurred. A whole cluster of PD pulses at the right-hand side shift from positive to the middle position; and a large number of pulses disappear. The results can also be observed in the corresponding PRPD diagram [Figure 5.16 (c2)]. It can also be seen that some pulses are missing and noise still remains in the de-noised signal obtained by EMD + AMT [Figure 5.16(d1)]. The results reveal the superiority of the EEMD-based method. It can effectively extract PD signals from noise and preserve the amplitudes and locations of PD pulses.

Table 5.3 also shows that the EEMD-based method still outperforms the others in all four PD source models. A higher reduction of pulse number error can be noticed in the DWT compared to EMD + AMT. The results imply that the DWT-based technique may not be suitable for PD signal de-noising, since additional pulses can be induced in the de-noised signals.

From the above results, the signal decomposition-based de-noising method is proven to be able to remove low-frequency DSI and white noise. In the next section, two methods based on the blind processing technique are verified to remove high-frequency DSI and white noise. PRPS and kurtogram are also examined for accurate PD pattern representation.

5.5.1.2 De-noising of High-Frequency DSI and White Noise with PD Pattern Representation

5.5.1.2.1 De-noising and PD Pattern Representation Methods – First Method

To remove high-frequency DSI and white noise as well as to provide an accurate PD pattern representation after de-noising, two BE-based PD signal de-noising and pattern representation methods can be used [48]. The flowchart of the first method is shown in Figure 5.17 [48]. It uses an advanced EVA for PD signal de-noising on DSI and AMT for removing white noise. A PRPS diagram is then constructed from the de-noised signals and used for PD pattern representation.

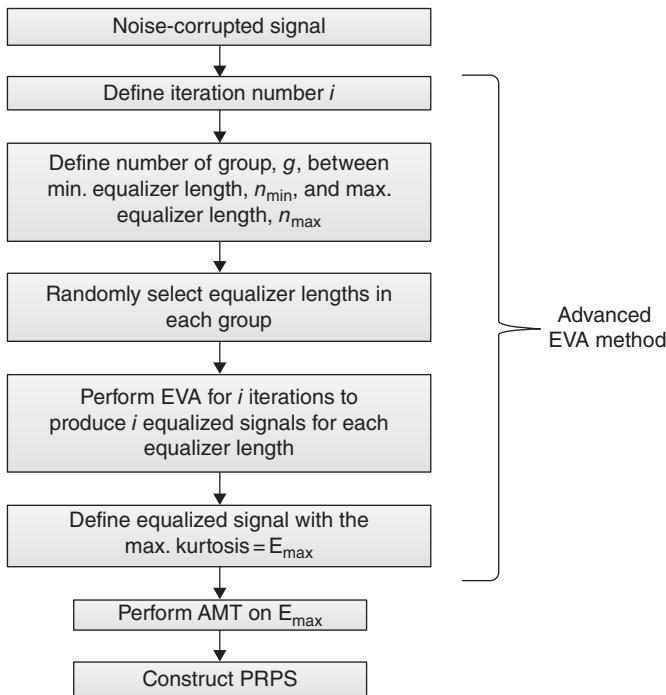


Figure 5.17 BE-based PD signal de-noising and pattern representation (first) method.

Equalized signals (recovered signals) in EVA are obtained by an iteration process. In this process, the selection of equalizer length is crucial. Improper selection of the length can incur the generation of unwanted equalized signals [71, 72]. In Ref. [71], all possible equalizer lengths were employed to seek optimal results on signal recovery. Although the results can reveal proper equalizer lengths, it is time-consuming and may not be suitable for online PD measurements of field transformers.

The BE-based method includes a novel approach for providing an adaptive selection of equalizer lengths. In this approach, a noise-corrupted signal is processed by EVA. The autocorrelation matrix R_{jj} and cross-cumulant matrix C_4^{yy} in EVA are defined as

$$R_{jj} = \frac{\sum_{k=n}^{N-1} jj^T}{N-n} \quad (5.35)$$

$$C_4^{yy} = \frac{\sum_{k=n}^{N-1} y^2(k) jj^T - \left(\sum_{k=n}^{N-1} y^2(k) \right) R_{jj}}{N-n} - \left(\frac{\left(\sum_{k=n}^{N-1} y(k) j \right) \left(\sum_{k=n}^{N-1} y^T(k) j^T \right)}{(N-n)^2} + \frac{\left(\sum_{k=n}^{N-1} y^T(k) j \right) \left(\sum_{k=n}^{N-1} y(k) j^T \right)}{(N-n)^2} \right) \quad (5.36)$$

where $j^T = [j(k), j(k-1), \dots, j(k-n)]$, T is the transpose function, N is the sample number, and n is the equalizer length. Equation (5.36) is derived based on Ref. [49]. The iteration number i for EVA is set to 10 to produce a reasonable number of equalized signals without sacrificing computational time. Then, a number of groups g are defined and evenly distributed between the minimum equalizer length n_{\min} and the maximum equalizer length n_{\max} to allocate the equalizer lengths in different ranges. Here, n_{\min} and n_{\max} were set to 10 and 100, respectively, and g was set to four. Three equalizer lengths were randomly selected in each group and the selection was subject to uniform distribution [72]. Such an arrangement is to ensure that the selected equalizer lengths cover the optimal equalizer length, which can be located in different groups.

Once the equalizer lengths are determined, EVA is performed to produce possible equalized signals aimed at removing the noise. After that, the equalized signal with the maximum kurtosis, E_{\max} , is chosen as an optimal equalized signal to extract PD pulses from the noise-corrupted signal. Since EVA seeks the maximum fourth-order cross-cumulant which refers to peakedness of signal distributions, PD pulses can be extracted from the equalized signals. After the EVA process, the noise can be removed even if its amplitude is larger than that of PD pulses. EVA can also remove white noise by reselecting the equalized signals. However, the de-noising performance of EVA on white noise is not consistent. Therefore, in the BE-based method, AMT is applied to eliminate the remaining white noise. Finally, a PRPS diagram is constructed from the de-noised signal.

5.5.1.2.2 De-noising and PD Pattern Representation Results – First method

According to Ref. [12], a variety of high-frequency DSI up to about 2 GHz could exist during field PD measurements at substations. Therefore, in case studies (using data collected from laboratory measurements), the noise was generated from a linear combination of 10 random frequencies up to 2 GHz with different amplitudes. The white noise was generated from the measurement environment. Here, it is worth mentioning that no artificial noise was added to the case studies of field measurements.

Figure 5.18 shows results of each processing step of the first proposed method on the PD signals acquired from the PD model of discharges in transformer oil using two HFCTs. CT1 has a frequency range from 350 kHz to 35 MHz and CT2 has a frequency range from 500 kHz to 50 MHz. After the noise-corrupted signals are processed by EVA with selected equalizer lengths, a series of equalized signals are generated for extracting PD pulses from the signals. Among all the equalized signals, some of them cannot reveal PD pulses. However, when the selection criterion is based on the kurtosis, an optimal equalized signal with the highest kurtosis value is chosen and can effectively extract PD pulses as shown in Figure 5.18(c). It can be seen that the phase angles (or time intervals) of the extracted pulses are the same as the original PD signals [Figure 5.18(a)].

Since white noise is still present in the optimal equalized signals, AMT is then applied for adaptively selecting upper and lower thresholds [Figure 5.18(d)]. The results of de-noised signals after AMT are shown in Figure 5.18(e), in which the phase angles of PD pulses are clearly identified. Figures 5.18(f) and (g) are PRPD

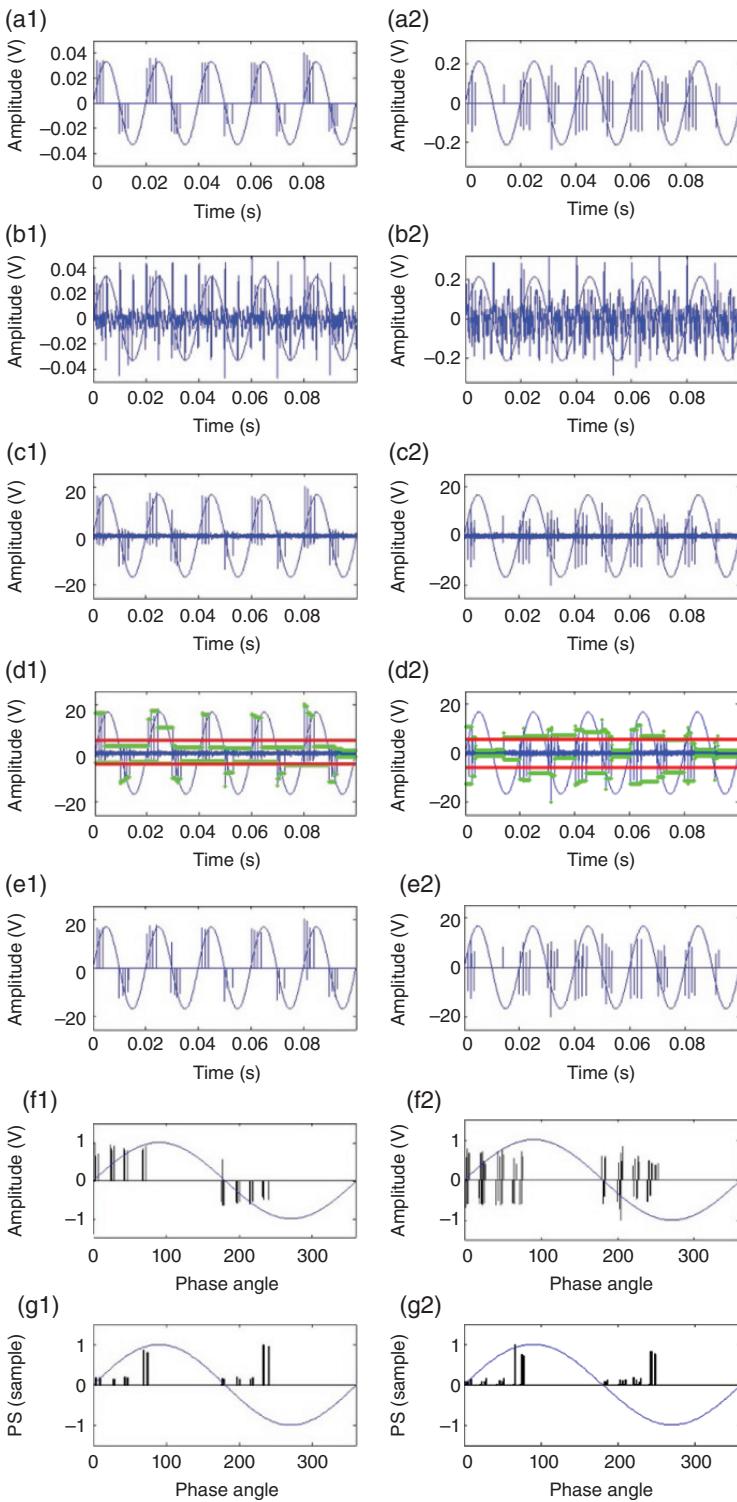


Figure 5.18 Results of discharge in transformer oil using (a1)–(g1) CT1 and (a2)–(g2) CT2.
 (a) Original signals, (b) noise-corrupted signals, (c) optimal equalized signals, (d) results of AMT on equalized signals, (e) de-noised signals, (f) PRPD diagram of original signals, and (g) PRPS diagram of de-noised signals. (See insert for color representation of the figure.)

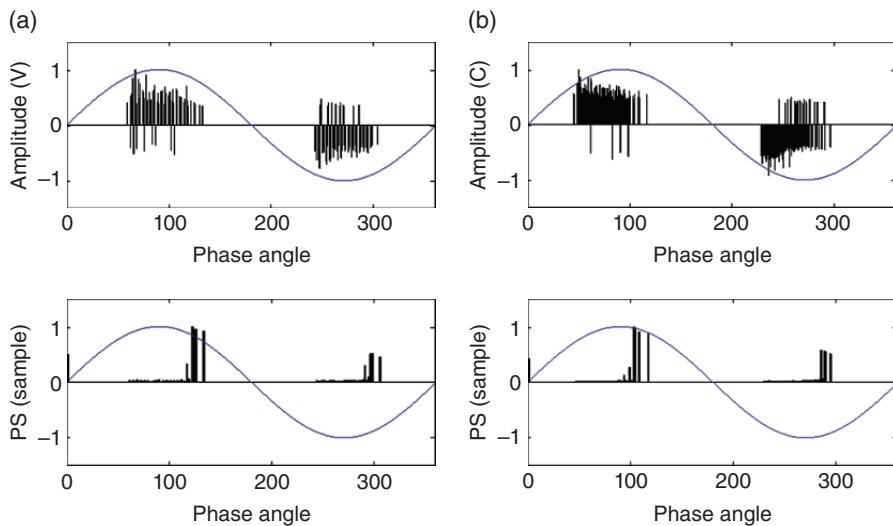


Figure 5.19 Results of PRPD and PRPS diagrams for a distribution transformer using (a) CT1 and (b) capacitive measurement.

diagrams of the original signals and PRPS diagrams of the de-noised signals, respectively. It can be seen that the PRPD diagrams fluctuate between the signals of the two CTs, while the PRPS diagrams are similar.

Field measurement and validation. To further validate the BE-based method, PD signals were acquired from a 10 kVA single-phase distribution transformer using the two HFCTs and capacitive measurement system. The results are shown in Figure 5.19 [48]. It can be seen that the PRPD patterns are different, while the PRPS patterns are similar between the PD signals of CT1 and capacitive measurement.

5.5.1.2.3 De-noising and PD Pattern Representation Methods – Second Method

In the second method for PD signal de-noising, a pre-whitening technique is applied before executing the advanced EVA method. This is to enhance the de-noising ability of EVA.

The flowchart of the second BE-based PD signal de-noising and pattern representation method is shown in Figure 5.20 [73].

As shown in Figure 5.20, the BE-based method starts with pre-whitening on a noise-corrupted PD signal. This is due to pre-whitening being able to enhance the impulsiveness of PD signals and thus facilitate the subsequent EVA process. Though the pre-whitening process can remove most of DSI, it may encounter difficulties if the noise amplitude is too large. This is due to the AR model used in the pre-whitening process not being able to provide an accurate prediction of the noise. In practical situations, the noise level is not always predictable. If small-amplitude PD signals mix with heavy noise, the pre-whitening process cannot always fulfil the task of removing the noise. Therefore, EVA is used to extract PD pulses from the remaining noise. After pre-whitening and EVA processes, a

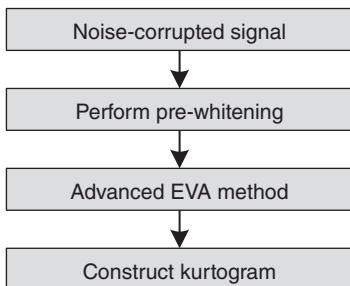


Figure 5.20 BE-based PD signal de-noising and pattern representation (second) method.

de-noised signal is plotted using a spectral kurtosis-based kurtogram for accurate PD pattern representation.

PD signals generated by different PD sources exhibit different patterns and are normally represented by PRPD diagrams [51, 74]. However, the patterns in PRPD diagrams can be overlapped if more than one PD source co-exist and occur simultaneously in a transformer [75, 76]. The patterns can also be influenced by different PD sensors and measurement systems. As such, in a PRPD diagram, PD patterns can be distorted and consequently PD sources causing discharges in HV apparatus cannot be identified and separated. Therefore, some extent of ambiguity can be induced in assessing HV apparatus conditions [75].

The kurtogram can be adopted to represent patterns of PD signals which have been de-noised by pre-whitening and EVA. Such a representation can be used to identify different PD sources. Though DSI can be removed by pre-whitening and EVA, white noise, which is normally found in field PD measurements, can still be present in the de-noised signals. However, a kurtogram is capable of representing PD patterns accurately even in the presence of white noise.

5.5.1.2.4 De-noising and PD Pattern Representation Results – Second Method

The noise investigated in the second method was the same as in the first method except that 10 more noise components of DSI with random frequencies and amplitudes were generated. The white noise was generated from the measurement environment. Figure 5.21 shows results of the proposed method for a signal acquired from the PD model of corona. The original and noise-corrupted signals are shown in Figures 5.21(a) and (b), respectively. As can be seen from the two figures, the amplitude of noise imposed on the original signal is very high. This demonstrates that the proposed method can extract PD signals and represent PD patterns when heavy noise is present. Figure 5.21(c) presents the signal obtained after a pre-whitening process on the noise-corrupted signal. It can be seen that after pre-whitening, the noise amplitude is much reduced. However, noise still suppresses the PD signal across the whole time span. Though the pre-whitening process cannot remove the entire noise, it does reveal PD pulses to some extent, which makes the subsequent EVA process easier. Figure 5.21(d) shows the de-noised signal (i.e., equalized signal) after EVA. It can be seen that the PD pattern of the corona is extracted, although the remaining noise amplitude is higher than that of the original signal shown in Figure 5.21(a).

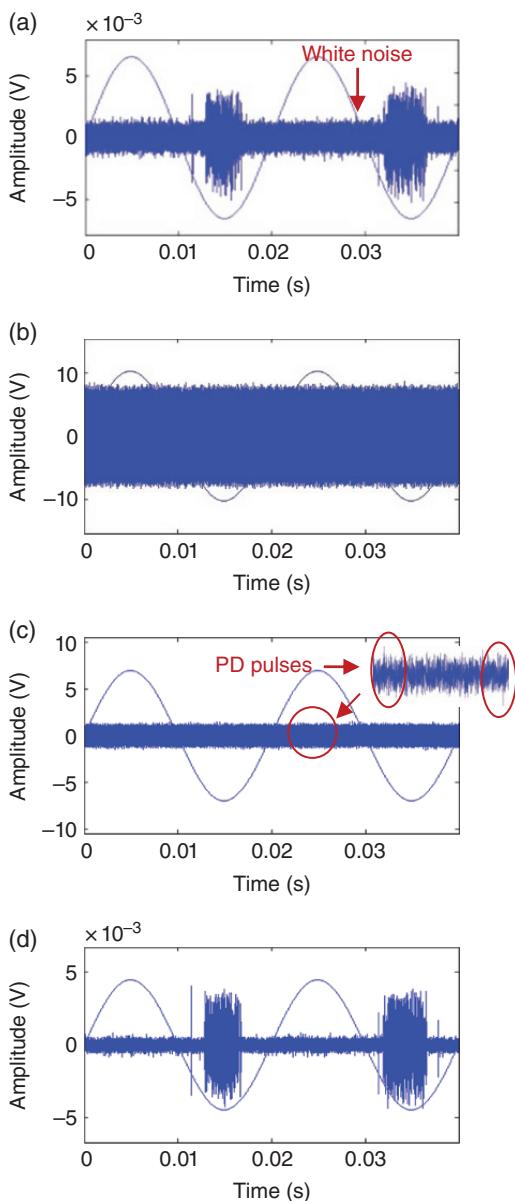


Figure 5.21 Results of corona. (a) Original signal, (b) noise-corrupted signal, (c) pre-whitened signal with magnified portion, (d) de-noised (equalized) signal, (e) kurtogram of original signal, and (f) kurtogram of de-noised signal.

Figure 5.21(e) shows a kurtogram constructed from the original signal [Figure 5.21(a)], while Figure 5.21(f) shows a kurtogram of a de-noised signal [Figure 5.21(d)]. Comparing these two kurtograms, there is not much difference, although the noise amplitude in the de-noised signal is higher than that in the original signal. This demonstrates the consistency of kurtograms in representing PD patterns.

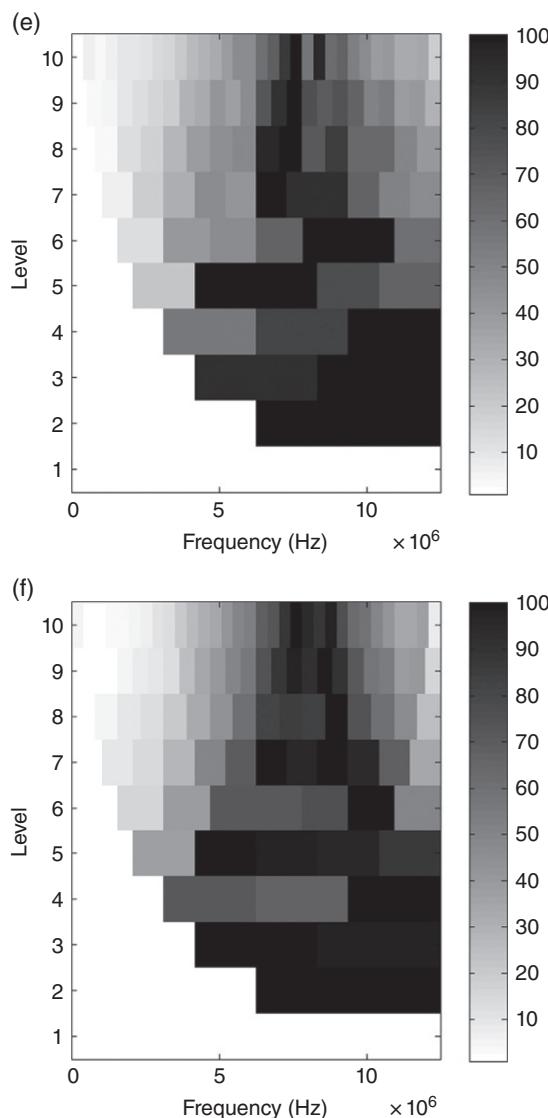
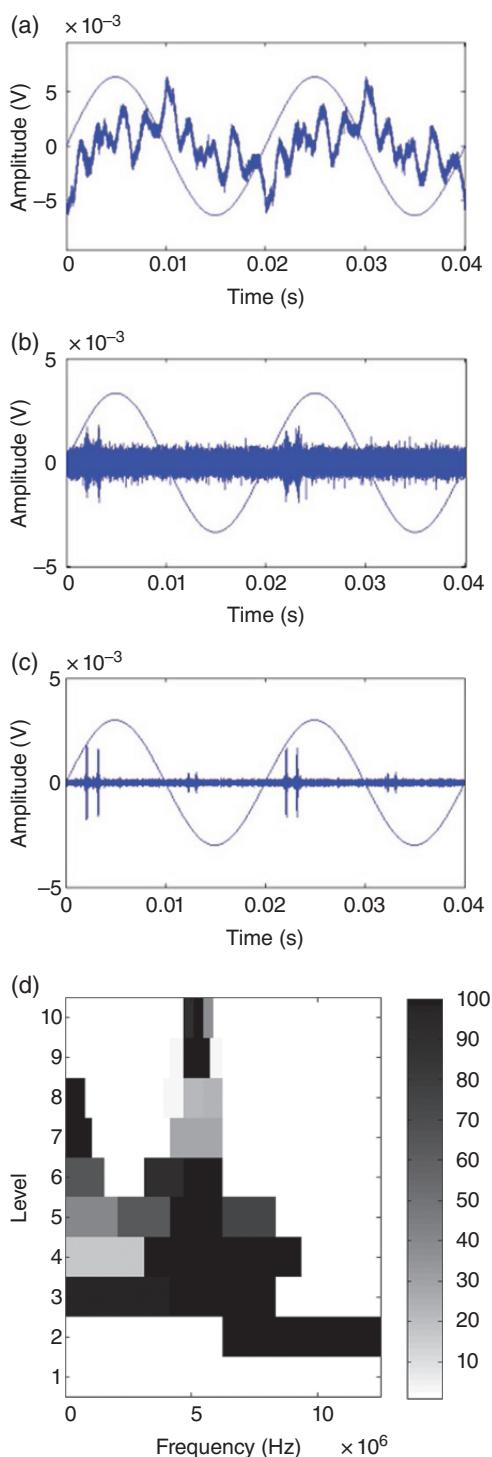


Figure 5.21 (Continued)

Field measurement and validation. The BE-based method has also been applied to signals acquired from a field PD measurement on a 10 MVA transformer (33 kV/11 kV) at the substation of a local distribution authority. As shown in Figure 5.22, the originally acquired signal is severely corrupted by a relatively low-frequency noise [Figure 5.22(a)]. This low-frequency noise was caused by AC harmonics. After applying pre-whitening, two pulse-shaped signals were uncovered in the positive power cycles [Figure 5.22(b)]. By applying EVA to the pre-whitened signal, two more pulses were revealed in each negative power cycle [Figure 5.22(c)]. The periodic signal may indicate that a single PD source exists

Figure 5.22 Results of a 10 MVA transformer. (a) Original signal, (b) pre-whitened signal, (c) de-noised (equalized) signal, and (d) kurtogram of de-noised signal.



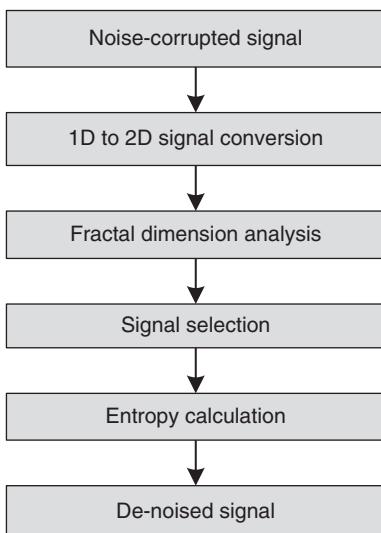


Figure 5.23 Graphics-based PD signal de-noising method.

inside the transformer. To further verify whether the pulses are related to PD events, oil samples were collected from this transformer and DGA was conducted. The DGA analysis using IEC 60599 indicates that there was a possibility of PD occurrence. Moreover, the result from Doernenburg ratios provided an indication of arcing. Thus, the pulses in Figure 5.22(c) have a high possibility of being generated by a PD event. After de-noising, a kurtogram was plotted in Figure 5.22(d). It can be seen that the pattern is different from those in Figures 5.21(e) and (f). This means that the PD event may not be related to the corona. To verify this pattern, a comparison with other kurtogram patterns from different PD sources should be performed.

5.5.1.3 De-noising of Stochastic Noise

To remove stochastic noise, a graphics-based PD signal de-noising method can be adopted. This method is a hybrid of fractal dimension and entropy, as shown in Figure 5.23 [65]. Firstly, an acquired one-dimensional (1D) noise-corrupted signal is transformed into a 2D signal. In this transformation, the original signal in a power cycle is converted into 512×512 pixels. Then, the fractal dimension is calculated on the transformed 2D signal in each power cycle with different side lengths of boxes, 2^x , where $x = [1, 2, \dots, n]$. If severe noise is present in a particular power cycle, then the value of fractal dimension in the power cycle is different from that of others. Thus, it is able to indicate the presence of noise. Subsequently, the noise can be removed.

Since noise may be located in every or most of the power cycles with similar fractal dimension values, using fractal dimension alone may not be able to completely distinguish and separate PD signals from noise. Moreover, even in situations where fractal dimension can distinguish noise appearing at most of the power cycles, it is still not appropriate to discard all signals located in those power cycles. Therefore, entropy is calculated in each box within each power cycle.

For a signal generated by stochastic noise, if it locates in a particular box in a power cycle, it is unlikely that it will be present in the same box in other power cycles when compared with PD signals, due to the random nature of stochastic noise. Based on this observation, the entropy value of the noise in each box is not as high as that of PD signals. After adding entropy values of the same box for every power cycle together and using colors to represent the sum (the highest to lowest values are represented by a color map ranging from red, orange, yellow, cyan, to blue), PD signals (i.e., de-noised signals) can be identified and subsequently extracted.

5.5.1.3.1 De-noising Results

Figure 5.24 presents de-noising results for a PD measurement performed on the PD source model of internal discharge. From Figure 5.24(a) (original PD signal), it can be observed that the PD signal locates periodically in most of the power cycles. Severe stochastic noise occurs at the first and second power cycles, while relatively insignificant noise can be found at the third power cycle. The stochastic noise was generated by occasional discharges from the HV transformer supplying voltage to the PD measurement and the coupling capacitor in the measurement circuit. This is because both the HV transformer and the coupling capacitor have been used for more than 20 years and cannot be regarded as discharge-free.

Figure 5.24(b) presents normalized results of fractal dimension in each power cycle with different side lengths of boxes. The numbers in the figure represent the power cycle numbers. It can be seen that fractal dimensions from the first to third cycles are separated from those of the other cycles in most of the side lengths. The fractal dimension of the third cycle containing minor noise is close to those of the fourth to tenth cycles. By contrast, the fractal dimensions of the first and second cycles containing severe noise are relatively distant from those of the fourth to tenth cycles. These results can be used to distinguish and separate PD signals from noise.

By rejecting the signals in the first three power cycles and combining the signals in the remaining cycles into one cycle, a PRPD pattern is obtained [Figure 5.24(d)]. Figure 5.24(c) is the original PRPD diagram without applying the proposed de-noising method. It can be seen that the PD pattern in the original PRPD diagram is contaminated by stochastic noise. This demonstrates that the original PRPD diagram cannot fully represent the PD pattern caused by the corresponding insulation defect (i.e., internal discharge), and thus may not be suitable for transformer insulation diagnosis.

The results in Figure 5.24 prove that the fractal dimension used by the proposed method is capable of identifying and removing stochastic noise. It can also indicate the severity of noise by comparing the signal value in each power cycle. For those signals where noise is located in every power cycle and the signals themselves have similar values of fractal dimension, entropy can further eliminate the noise.

Figure 5.25 presents de-noising results for a PD measurement performed on the PD source model of the corona. From the original signal and its magnification [Figures 5.25(a) and (b)], noise can be observed in most of the power cycles. By contrast, the discharge signals [indicated by red arrows, Figure 5.25(b)] are hardly

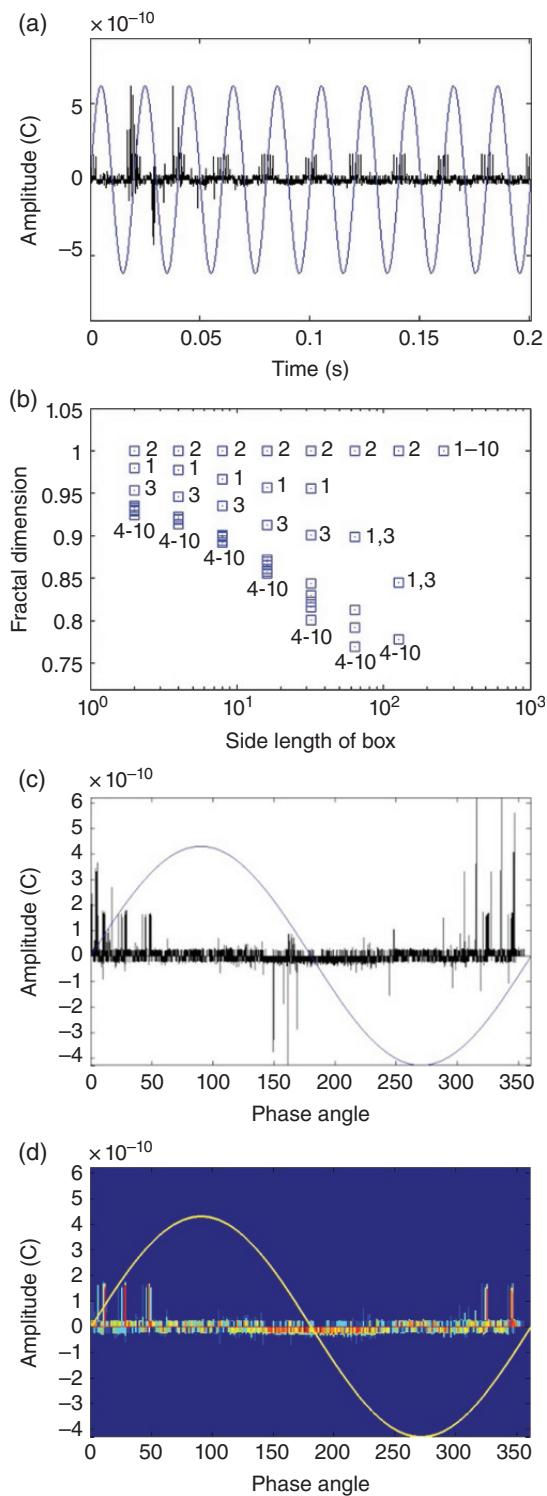
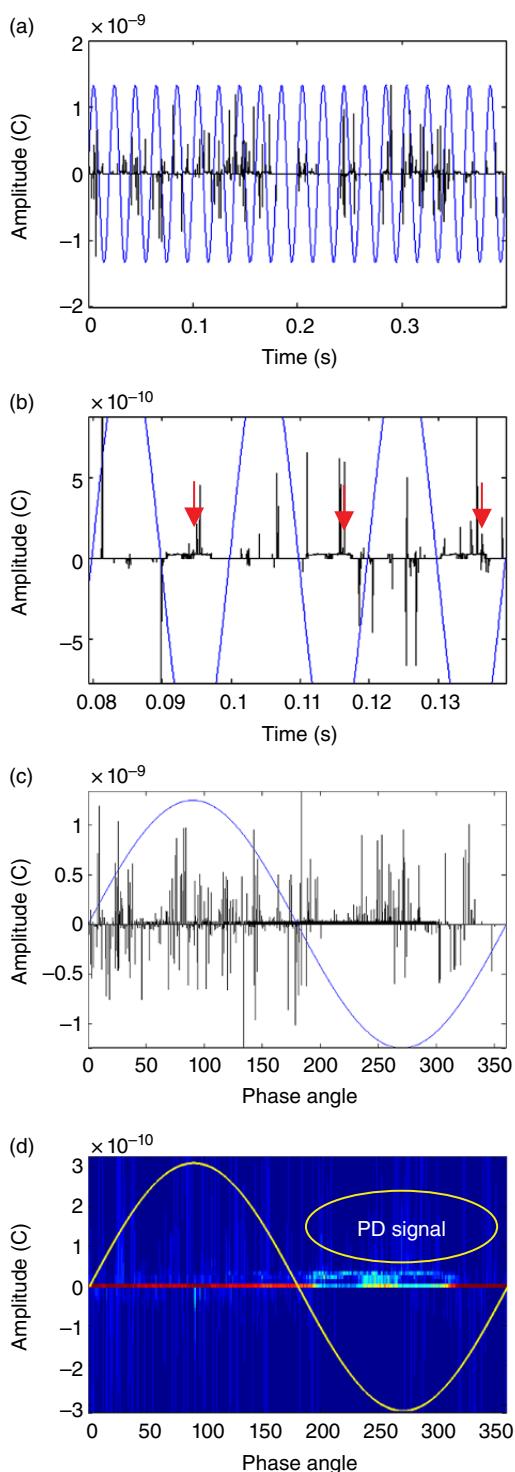


Figure 5.24 De-noising results of internal discharge. (a) Original signal, (b) fractal dimension, (c) original PRPD diagram, and (d) de-noised PRPD diagram.

Figure 5.25 De-noising results of corona. (a) Original signal, (b) original signal (magnified), (c) original PRPD diagram, and (d) de-noised PRPD diagram (magnified).



identified due to their relatively small amplitudes. Though some noise can be identified in some power cycles and removed by fractal dimension calculation, other noise is still present in the remaining power cycles (since the signals in these power cycles have similar fractal dimension values). Through calculating entropy in each box of fractal dimension in each cycle and then combining all entropy values into a single cycle, a de-noised PRPD diagram is obtained [Figure 5.25 (d)]. In the figure, the PD signal is clearly identified (the pulses represented in light gray are stochastic noise). However, in the original PRPD diagram without applying the proposed method, the PD signal is submerged in noise [Figure 5.25(c)].

Field measurement and validation. The proposed method was then applied to extract PD signals acquired from a PD measurement on a 5 MVA power transformer at a substation with 22 kV test voltage, as shown in Figure 5.26. It can be seen that the original PD signal is immersed in noise and can hardly be identified [Figures 5.26(a) to (c)]. However, the PD signals can be revealed by using the proposed method [Figure 5.26(d)]. The PD signal located near the negative cycle may be related to the corona, while the PD signal located in both the negative and positive cycles may be generated by internal discharge.

From the above verification of the signal processing techniques on PD signal de-noising, it is showed that a variety of noise can be removed. In the next section, signal processing techniques are applied to separate multiple PD sources that can possibly be embedded in a transformer.

5.5.2 Multiple PD Source Separation

After PD signal extraction (de-noising), multiple PD source separation is performed. In this section, both the TF sparsity map and the conventional TF map (based on equivalent time length and bandwidth) are applied to signals acquired from PD source models and a 100 kVA transformer (10.5 kV/415 V).

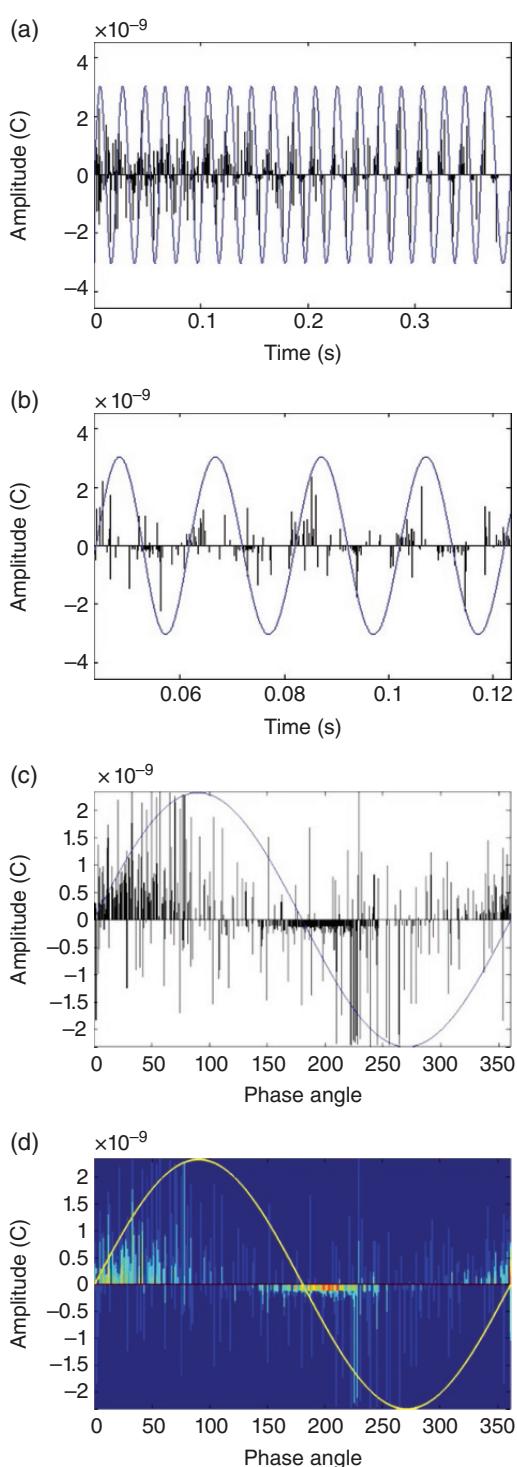
Figure 5.27 presents the results of TF sparsity maps and a conventional TF map for separating three PD sources (corona, surface discharge, and discharge in transformer oil). As can be observed from Figure 5.27(a), the conventional TF map is unable to distinguish these three types of PD source. Most of the data points belonging to different types of PD source are merged together. This result indicates that PD pulses cannot be represented accurately by equations (5.20) and (5.21) adopted by the conventional TF map, since some information regarding the shapes of PD pulses is lost. By contrast, the proposed TF sparsity map achieves a very clear separation of the three types of PD source [Figure 5.27(b1)].

In Figure 5.27(b2), the corresponding PRPD diagram indicates that PD pulses in gray are generated by discharge in transformer oil, PD pulses in black are generated by corona, and PD pulses in light gray are generated by surface discharge.

Figure 5.28 shows the results of separating PD sources consisting of corona, surface discharge, and internal discharge. It can be seen from Figure 5.28(a) that the PD sources cannot be clearly distinguished and separated by a conventional TF map. Although three clusters have been formed by DBSCAN, some data points (in light gray) cannot be classified.

By contrast, the three PD sources can clearly be distinguished by the TF sparsity map [Figure 5.28(b1)], in which the data points belonging to different types of PD

Figure 5.26 De-noising results of PD signals acquired from a transformer. (a) Original signal, (b) original signal (magnified), (c) original PRPD diagram, and (d) de-noised PRPD diagram.



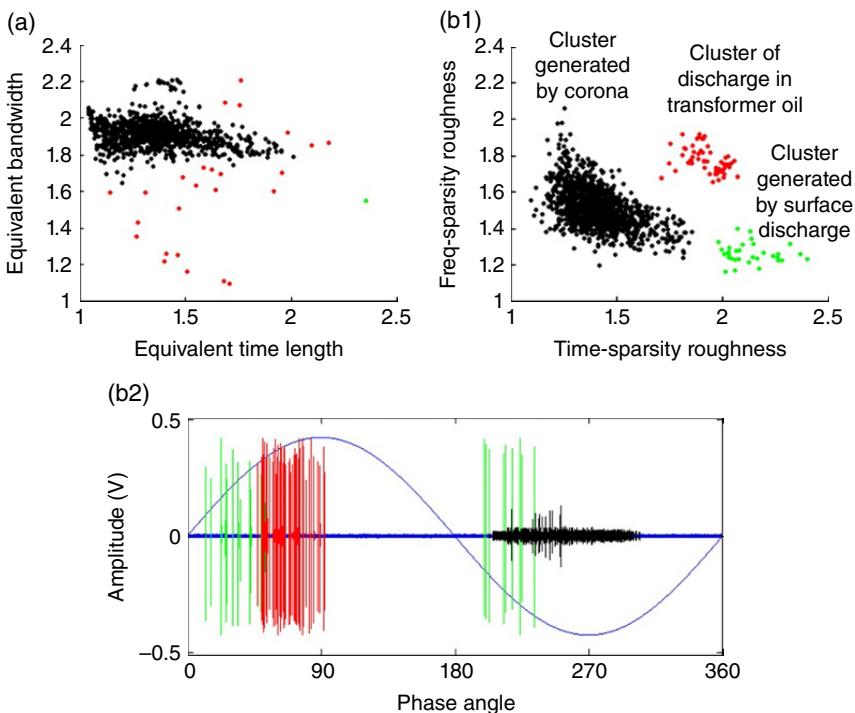


Figure 5.27 Results of multiple PD sources combining corona, surface discharge, and discharge in transformer oil. (a) Result of conventional TF map, (b1) result of TF sparsity map, and (b2) PRPD diagram based on the result of TF sparsity map.

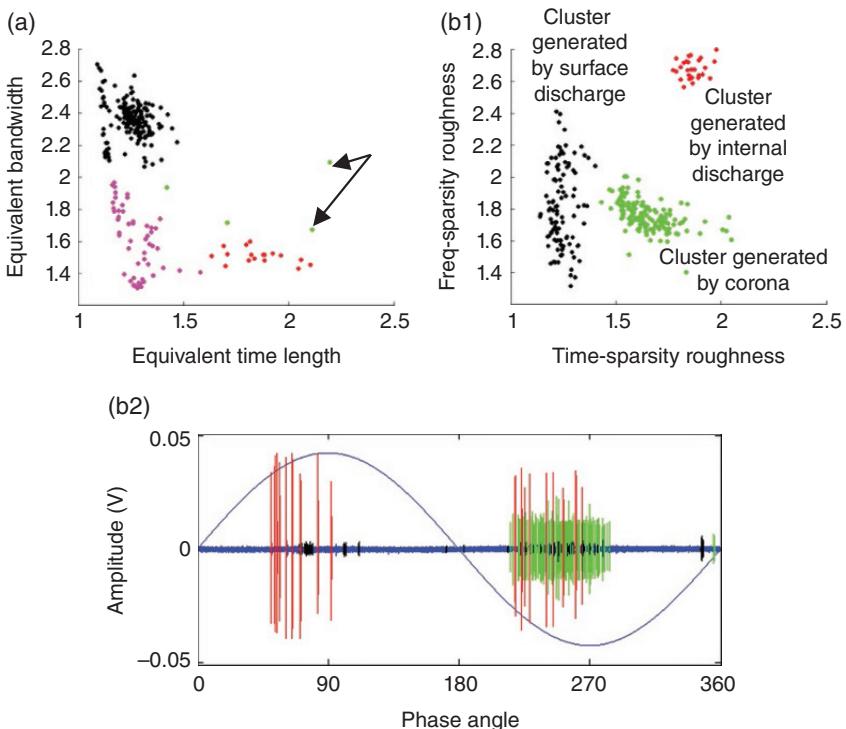


Figure 5.28 Results of multiple PD sources combining corona, surface discharge, and internal discharge. (a) Result of conventional TF map, (b1) result of TF sparsity map, and (b2) PRPD diagram based on the result of TF sparsity map. (See insert for color representation of the figure.)

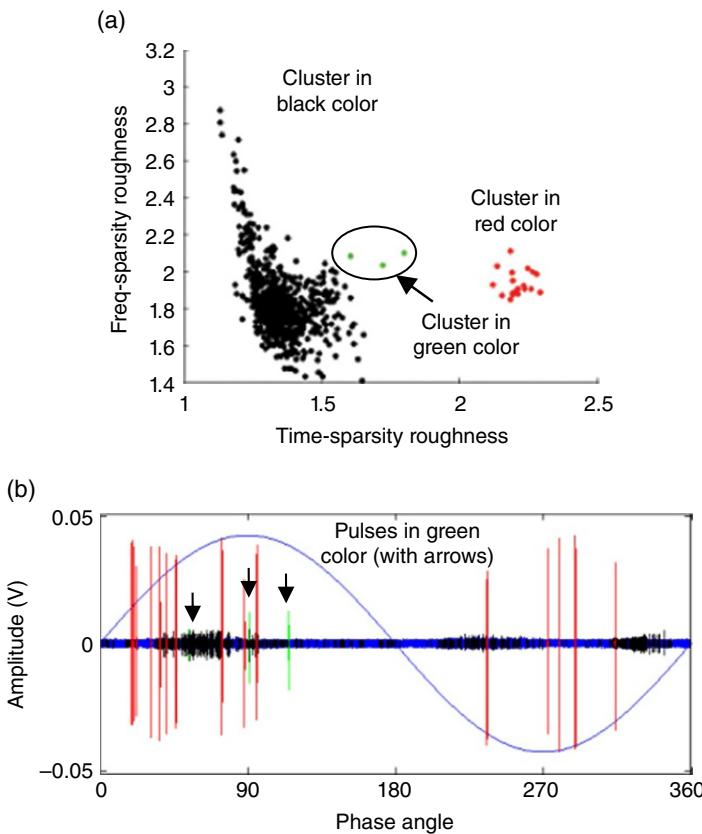


Figure 5.29 Results for a 100 kVA transformer. (a) TF sparsity map and (b) PRPD diagram.

source are well separated. From the PRPD diagram, the PD pulses generated by internal discharge, surface discharge, and corona can be clearly identified. In particular, all three types of PD pulse can be distinguished in the negative power cycle, even if they overlap. Moreover, the TF sparsity map is able to identify the PD pulses generated by surface discharge (in black), even when the amplitudes of these pulses are relatively small and embedded in the PD pulses generated by the corona.

Field measurement and validation. The TF sparsity map has also been applied to analyze signals acquired from an online PD measurement on a 100 kVA transformer (10.5 kV/415 V). Signals were acquired using CT1 clamping on the grounding wire of the transformer. The signals were acquired over 20 power cycles. The results of the 100 kVA transformer are shown in Figure 5.29. From Figure 5.29(a), it is clearly observed that two main clusters are formed on the TF sparsity map. The corresponding distribution of PD pulses shown in the PRPD diagram [Figure 5.29 (b)] may be due to internal and surface discharges. As shown in Figure 5.29(a), a very small cluster on the TF sparsity map between the two clusters (in light gray (circled)) also exists. Since DBSCAN is a density-based clustering method, except for the two clusters in black and gray that are dense enough to form

clusters, the data points in light gray are recognized as an individual cluster by DBSCAN. The pulses corresponding to this small cluster are located in the positive cycle, and these pulses may be generated by discharge in transformer oil. In Figure 5.29(b), the signals in dark black are white noise and are not extracted.

The results presented above prove that the TF sparsity map is capable of providing a clear separation of multiple PD sources. After source separation, the next step in PD analysis involves extraction and PD source classification (refer to Figure 5.1). This will be discussed in the next section.

5.6 PD Source Classification

PD source classification is aimed at identifying the types of defect (PD sources) causing discharges in a transformer. Different types of PD source can generate different patterns [77, 78]. The discharge pattern is represented by a set of features, which can be computed using PRPD or directly from measured PD signals. On the basis of these features, it is possible to develop computer algorithms to classify the types of PD source to assist transformer condition assessment. Over the past three decades, considerable efforts have been made to apply pattern recognition algorithms such as genetic algorithms, ANNs, knowledge-based systems, SVMs, and WTs for PD source classification [29, 68, 77–81].

5.6.1 Feature Selection

PD signals after de-noising can be represented as an $N \times D$ dataset $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T$, where N is the number of PD signals, and each signal is obtained from one PD acquisition and is a D -dimensional vector – i.e., $\mathbf{x}_i = [x_i^1, \dots, x_i^D]$. The feature selection stage transforms the dataset \mathbf{X} to an $N \times d$ dataset $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]$, $\mathbf{y}_i = [y_i^1, \dots, y_i^d]$, where $d \ll D$. Though the number of features in dataset \mathbf{Y} is much less than that in dataset \mathbf{X} , the features in dataset \mathbf{Y} can still provide essential features to discriminate different PD sources. Feature selection can direct the subsequent pattern recognition algorithm to focus on the most relevant but non-redundant information.

Most feature extraction techniques for PD signal analysis are based on the PRPD diagram [68]. However, the dataset extracted from the PRPD diagram still has considerably high dimensionality. For a PRPD diagram with 200 phase windows, the number of features will be $200 \times 3 = 600$. To deal with the high dimensionality, a number of feature selection approaches including statistical operators, principle component analysis (PCA), kernel principle component analysis (KPCA), and stochastic neighbor embedding (SNE) can be employed to extract representative features from PRPD.

5.6.1.1 Statistical Operators [79]

In this approach, a number of statistical operators are calculated from the three PRPD distributions. The feature set consists of 24 features: (1) skewness and kurtosis for both positive and negative cycles; (2) number of peaks for both positive

and negative cycles; (3) asymmetry between positive and negative cycles; and (4) cross-correlation factor of the three PRPD distributions.

5.6.1.2 Principle Component Analysis [79, 82]

PCA maps the original data into a lower-dimensional linear subspace while keeping as much of the variance in the data as possible. PCA solves an eigenvalue problem to find such a linear subspace:

$$\text{cov}(\mathbf{X})\mathbf{M} = \lambda\mathbf{M} \quad (5.37)$$

where $\text{cov}(\mathbf{X})$ is the covariance matrix of the dataset \mathbf{X} , λ are the d principal eigenvalues, and \mathbf{M} is a linear mapping formed by the d principle eigenvectors of $\text{cov}(\mathbf{X})$. The low-dimensional data representations \mathbf{y}_i of the data points \mathbf{x}_i are computed through a linear mapping $\mathbf{Y} = \mathbf{XM}$. The elements of \mathbf{Y} (e.g., \mathbf{y}_i) form the features set.

5.6.1.3 Kernel PCA [79, 82]

KPCA uses a kernel function to transform the original data into a feature space. The elements of this kernel are in the form

$$K_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j) \quad (5.38)$$

where κ is a kernel function (one commonly used kernel is the Gaussian kernel, refer to Section 5.6.2) and $\mathbf{x}_i, \mathbf{x}_j$ are different data points in dataset \mathbf{X} . KPCA computes the eigenvectors of the covariance matrix \mathbf{a}_i in the above feature space. To obtain the low-dimensional data representation \mathbf{Y} , the data points in the original dataset \mathbf{X} are projected onto the eigenvectors of the covariance matrix \mathbf{a}_i :

$$\mathbf{y}_i = \left\{ \sum_{j=1}^N a_i^{(j)} \kappa(\mathbf{x}_i, \mathbf{x}_j), \dots, \sum_{j=1}^N a_d^{(j)} \kappa(\mathbf{x}_i, \mathbf{x}_j) \right\} \quad (5.39)$$

As in PCA, the elements \mathbf{y}_i in dataset \mathbf{Y} form the features set.

5.6.1.4 Stochastic Neighbor Embedding [79, 82]

SNE is aimed at retaining the pairwise distances between the data points in high-dimensional space. In SNE, the probability p_{ij} that data points \mathbf{x}_i and \mathbf{x}_j are generated by the same Gaussian distribution is computed for all possible pairs of data points in the original dataset \mathbf{X} . The probabilities q_{ij} of the corresponding data points in low-dimensional space (i.e., $\mathbf{y}_i, \mathbf{y}_j$ generated by the same Gaussian distribution) are also computed. These two probability matrices are denoted \mathbf{P} and \mathbf{Q} , respectively. SNE minimizes the difference between the probability distributions \mathbf{P} and \mathbf{Q} by minimizing the sum of Kullback–Leibler divergences

$$\varphi = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (5.40)$$

As in PCA and KPCA, the elements \mathbf{y}_i form the features set.

Table 5.4 Information on feature datasets

Feature extraction	Information on dataset X	Feature size of dataset X	Feature size of dataset Y
Statistical operator	Phase-resolved PD data in 200 phase windows	600	24
PCA	Phase-resolved PD data in 200 phase windows	600	7
KPCA	Phase-resolved PD data in 200 phase windows	600	7
SNE	Phase-resolved PD data in 200 phase windows	600	7
DWT	PD pulses of 100 power cycles	around 40,000	36

5.6.1.5 Discrete Wavelet Transform [79, 82]

Instead of using the PRPD diagram, DWT utilizes PD signals directly for feature extraction. After being decomposed by DWT, a PD signal is transformed to a set of wavelet coefficients at nine decomposition levels. Accordingly, nine wavelet coefficient distributions are formed. Then, the first four moments (i.e., mean, variance, skewness, and kurtosis) on the distributions of these wavelet coefficients are adopted to form the feature set, which consists of $4 \times 9 = 36$ features [79].

Information on the above datasets X and Y is presented in Table 5.4.

5.6.2 Pattern Recognition [79]

The pattern recognition algorithm consists of training (learning) and testing (classification) stages. In the training stage, the relationship between data points (e.g., selected features) and their corresponding types of PD source is explored and then approximated using a training database. In the testing stage, a new input data point (not included in the training database) is classified into one type of PD source [79]. Typical pattern recognition algorithms are fuzzy support vector machine (FSVM), radial basis network (RBF), multi-layer perceptron (MLP), Bayesian classifier, k nearest neighbor (k NN), and two-layer (input–output) network. The formulation of FSVM is introduced briefly in the following section. The formulations of RBF, MLP, Bayesian classifier, k NN, and two-layer network can be found in Chapter 4. For a detailed derivation of these algorithms, readers may refer to Refs [82–84].

FSVM is a variant of SVM [84]. In FSVM, each data point is assigned a weight, which describes the degree to which this data point belongs to different types of PD source (classes). These weights are augmented to the original training dataset Y in the form $\{\mathbf{y}_k, C_k, \rho_k\}$, where $0 < \rho_k < 1 < 1$ are the weights, determined using an exponentially decaying function [79, 84]

$$\rho_k = \frac{2}{1 + \exp(\beta d_k^{\text{cen}})}, \beta \in [0, 1] \quad (5.41)$$

where $d_k^{\text{cen}} = \mathbf{y}_k - \mathbf{v}_i^{1/2}$ is the Euclidean distance between a data point \mathbf{y}_k and its cluster center \mathbf{v}_i , which is the geometrical mean of the data points belonging to the same type of PD source. β determines the steepness of the decay.

The FSVM algorithm converts the input data from its original data space R^m to a higher-dimensional space through a nonlinear function $h = \Phi(x)$. Then, in that space the FSVM searches for an optimal separation hyperplane by solving a quadratic programming (QP) problem [84]

$$\text{Minimize} \quad \frac{1}{2} \mathbf{w}^T \mathbf{w} + G \sum_{k=1}^N \rho_k \xi_k \quad (5.42)$$

$$\text{Subject to} \quad z_k (\mathbf{w}^T \cdot \Phi(\mathbf{y}_k) + b) \geq 1 - \xi_k, \quad k = 1, \dots, N \quad (5.43)$$

$$\text{and} \quad \xi_k \geq 0, \quad k = 1, \dots, N \quad (5.44)$$

where the pair (\mathbf{w}, b) defines the separation hyperplane, in which \mathbf{w} is the hyperplane's normal vector while b is the bias. G , the regularization parameter, is used to balance the margin maximization and misclassification; $\xi_k \geq 0$ is the error term due to misclassification.

The dual form of the above QP problem is [84]

$$\text{Maximize} \quad \sum_{k=1}^N \alpha_k - \frac{1}{2} \sum_{k=1}^N \sum_{j=1}^N \alpha_k \alpha_j z_k z_j K(\mathbf{y}_k, \mathbf{y}_j) \quad (5.45)$$

$$\text{Subject to} \quad \sum_{k=1}^N \alpha_k z_k = 0 \quad (5.46)$$

$$\text{and} \quad 0 \leq \alpha_k \leq \rho_k G \quad (5.47)$$

where α_k is the Lagrange multiplier and $K(\mathbf{y}_k, \mathbf{y}_i)$ is the kernel function in the form $K(\mathbf{y}_k, \mathbf{y}_i) = \Phi(\mathbf{y}_k)^T \Phi(\mathbf{y}_i)$. The Gaussian kernel $K(\mathbf{y}_k, \mathbf{y}_i) = e^{-\gamma \|\mathbf{y}_k - \mathbf{y}_i\|^2}$ is adopted in PD source classification, where γ is the variance parameter.

After obtaining the maximized hyperplane, for a new data point \mathbf{y}_* the FSVM can predict its class label z_* :

$$z_* = \text{sgn} \left[\sum_{k=1}^N \alpha_k z_k K(\mathbf{y}_k, \mathbf{y}_*) + b \right] \quad (5.48)$$

5.6.3 Case Studies

5.6.3.1 Feature Selection and Data Preparation [79]

In the case studies, the original dataset is obtained from PD measurements (capacitive measurement system) on a number of different experimental PD source models, which include corona, discharge in oil, surface discharge, internal discharge, and discharge due to floating particles. The database has 1000 data points, with each data point corresponding to one PD acquisition, which is made up of PD pulses of 100 AC cycles. Firstly, the original dataset is converted to the PRPD dataset (i.e., dataset X), which consists of average discharge pulse magnitude, maximum discharge pulse magnitude, and discharge number count in each of the

200 phase windows. Then, the PRPD dataset is processed by feature extraction algorithms to form a feature dataset (i.e., dataset Y). Such a feature dataset is then used by pattern recognition algorithms for PD source classification. For feature extraction using DWT, the original dataset is processed directly without converting to PRPD, and the first four moment statistics are calculated to form the feature dataset. Information on the above PRPD dataset and feature dataset can be found in Table 5.4.

To evaluate the pattern recognition algorithms, dataset Y is randomly split into two parts: a training dataset (having 70% data points in dataset Y) and a testing dataset (having 30% data points in dataset Y). The training dataset is used to find the model parameters in FSVM and ANNs. These parameters are: number of neighbors in k NN, number of hidden nodes in MLP and RBF, and number of mixture components in Bayesian classifiers, G and γ in FSVM. Tenfold cross-validation on the training dataset is conducted to obtain these parameters [79]. Finally, the trained algorithms make classification on the data points in the testing dataset into one of the PD sources (classes).

5.6.3.1.1 Results

Tables 5.5 to 5.9 summarize the classification results of six pattern recognition algorithms integrated with different feature extraction approaches. Each table presents the overall classification accuracy (%) and the classification accuracy with respect to each type of PD source.

From these tables, it can be seen that FSVM outperforms the other five ANN algorithms. Amongst the ANN algorithms, MLP also attains desirable and consistent classification accuracy. It is also not strange that the two-layer network attains lower classification accuracy, since its simple architecture may not be able to capture the nonlinearity of the dataset. It is interesting to see that k NN performs well with the feature extraction approaches PCA, KPCA, and DWT. A similar phenomenon of k NN has also been reported in Ref. [85].

The tables also reveal that k NN, two-layer network, MLP, RBF, and FSVM integrated with DWT attain higher classification accuracy compared to these algorithms integrated with the other four feature extraction approaches. k NN,

Table 5.5 Classification results (%) of algorithms (with statistical operators) [79]

Algorithm	Overall	Class 1	Class 2	Class 3	Class 4	Class 5
k NN	90.8	99.4	83.8	86.7	99.5	85.2
Two-layer	90.6	94.7	87.2	86.1	97.6	88.1
RBF	92.5	98.5	87.4	90.1	97.8	89.2
MLP	92.0	98.1	84.8	91.6	99.1	86.6
Bayesian	93.2	97.5	85.6	91.0	99.3	93.3
FSVM	95.1	99.5	92.0	92.2	99.4	92.7

Note: Class 1 – corona; Class 2 – discharge in oil; Class 3 – surface discharge; Class 4 – internal discharge; Class 5 – discharge due to floating particle.

Table 5.6 Classification results (%) of algorithms (with DWT) [79]

Algorithm	Overall	Class 1	Class 2	Class 3	Class 4	Class 5
kNN	94.2	99.2	91.4	96.0	98.3	85.4
Two-layer	92.6	93.3	80.8	94.2	95.3	91.8
RBF	97.3	98.8	97.5	97.2	98.3	94.3
MLP	98.5	97.9	99.6	99.4	98.2	97.2
Bayesian	92.0	92.8	94.2	87.8	99.7	85.8
FSVM	98.8	97.4	99.4	98.9	100.0	98.3

Table 5.7 Classification results (%) of algorithms (with PCA) [79]

Algorithm	Overall	Class 1	Class 2	Class 3	Class 4	Class 5
kNN	93.7	100.0	86.4	98.1	100.0	83.3
Two-layer	81.4	100.0	69.7	75.9	100.0	61.0
RBF	91.2	93.9	86.5	91.7	99.8	84.5
MLP	93.6	100.0	83.3	97.8	99.8	81.2
Bayesian	90.1	99.9	88.2	90.6	100.0	74.3
FSVM	96.5	99.9	93.4	99.7	99.8	89.1

Table 5.8 Classification results (%) of algorithms (with KPCA) [79]

Algorithm	Overall	Class 1	Class 2	Class 3	Class 4	Class 5
kNN	93.9	100.0	90.3	96.0	99.8	82.8
Two-layer	83.2	98.9	74.2	82.8	99.3	59.6
RBF	91.7	96.9	87.8	91.7	100.0	82.0
MLP	93.8	100.0	87.9	99.0	99.5	81.3
Bayesian	92.4	99.9	87.9	95.3	99.1	78.4
FSVM	96.1	99.8	92.5	99.7	99.9	88.1

Table 5.9 Classification results (%) of algorithms (with SNE) [79]

Algorithm	Overall	Class 1	Class 2	Class 3	Class 4	Class 5
kNN	87.2	97.6	79.4	85.8	93.2	79.9
Two-layer	28.7	43.2	55.1	6.9	34.7	0
RBF	85.2	92.4	76.7	82.5	91.7	83.7
MLP	89.4	99.4	85.7	83.3	99.3	79.3
Bayesian	85.9	94.1	83.8	80.0	94.5	77.0
FSVM	92.3	100.0	86.3	91.3	99.2	84.5

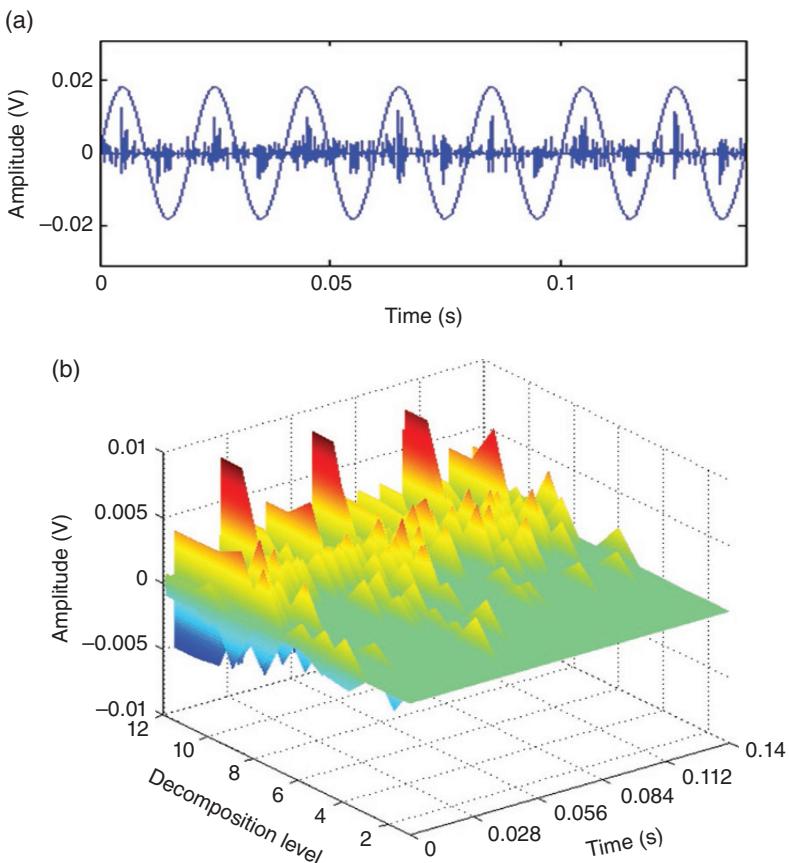


Figure 5.30 PD measurement results of a 10 kVA distribution transformer: (a) measured PD signals; (b) DWT decomposition.

MLP, RBF, Bayesian classifier, and FSVM integrated with KPCA and PCA attain comparable classification accuracy as they are integrated with statistical operators. It is worth mentioning that KPCA and PCA only use seven features, while the statistical operators approach uses 24 features. The pattern recognition algorithms integrated with SNE could not attain desirable classification accuracy. The two-layer network integrated with SNE even fails to make classification. This might be due to the optimization in SNE being trapped into local minima in the cost function [86].

In Figure 5.30, PD patterns of a 10 kVA (10.5 kV/415 V) distribution transformer are presented. An inductive PD measurement system using HFCT was employed for PD measurement of this transformer. The above trained FSVM algorithm (using features extracted by DWT) was employed to recognize the types of PD source inside this transformer. Results indicate that this transformer has 80% possibility of internal discharge, 15% possibility of surface discharge, and 5% possibility of other type discharges.

To achieve the desirable capability of the pattern recognition algorithms in PD source classification, the training database needs to include as many types of PD source as possible. It is also worth mentioning that some other factors also need to be considered when applying the algorithms to PD source classification. For example, the resolution of the PD measurement system has an influence on the overall performance of the algorithms. If the algorithms are trained with PD signals of full resolution and tested on the squashed (distorted) PD signals, then the overall performance of the algorithms will be decreased significantly.

5.7 Acoustic and UHF Methods for PD Signal Detection and Localization

5.7.1 Acoustic PD Measurement for PD Source Localization [87, 88]

5.7.1.1 Sound Wave Propagation

In acoustic PD measurement, a number of acoustic sensors are employed to estimate the location of the PD source inside a transformer. The estimation is made based on the distance between the PD source and the acoustic sensors mounted on the transformer's tank through measuring the time of arrival (TOA) – that is, the elapsed time of sound (generated by discharge) traveling from the PD source to the sensors.

To measure TOA, it is important to identify the starting time instant of the arriving sound waveform acquired by the acoustic sensors. This requires a trigger signal to evoke acoustic sensors to start recording PD events. As shown in Figure 5.31, an HFCT-based PD measurement system can be used to provide such a trigger signal. This is because HFCT measures the electrical signals generated by a PD source, which are much faster than the sound signals generated by the same PD source. Then, the absolute TOA of an acoustic sensor can be obtained by subtracting the arrival time measured by this sensor from the reference time at which the HFCT detects the PD event.

The sound wave generated by a PD source can propagate along different paths inside a transformer and thus there may exist different TOAs [89]. In Figure 5.32, two extreme sound wave travel paths are depicted: (1) the shortest path, in which a sound wave travels only via transformer oil and reaches the acoustic sensors; and (2) the longest path, in which the sound wave first travels in transformer oil (perpendicular to the transformer's tank) and then enters the transformer tank wall and propagates until it reaches the acoustic sensors. Other possible sound wave propagation paths can exist between the above longest and shortest wave paths.

5.7.1.2 Spatial Intersection Technique for Locating PD Source

In the spatial intersection technique, two spheres for each acoustic sensor can be formed. The radii of these two spheres are decided by: (1) the longest distance that the sound wave can travel; and (2) the shortest distance that the sound wave can travel. The sound wave propagation velocity in oil is set to 1400 m/s, and it is set to

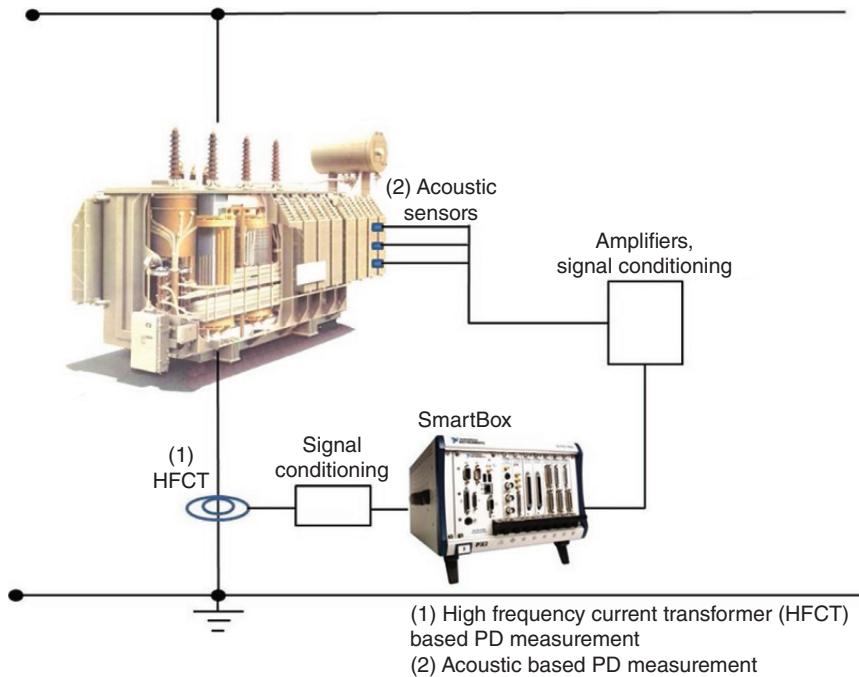


Figure 5.31 Setup of acoustic PD measurement of a transformer.

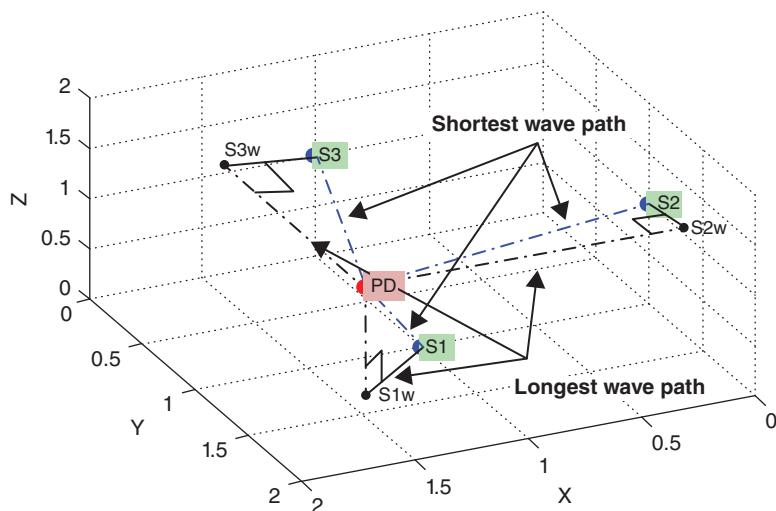


Figure 5.32 Possible sound wave travel paths from PD source to three acoustic sensors.

5900 m/s for longitudinal waves and 3200 m/s for shear waves in the transformer tank wall [89].

Accordingly, each acoustic sensor will have two spheres, one sphere has a small diameter and the other sphere has a large diameter. The intersection area of three acoustic sensors will be the possible location of a PD source.

5.7.1.3 Case Study

The above technique was adopted for localizing a PD source inside a transformer, of which the dimensions are 78 cm in the X -axis, 36.2 cm in the Y -axis, and 119.4 cm in the Z -axis. The transformer and the locations of three acoustic sensors are shown in Figure 5.33. Each AE sensor had an operating frequency range of 50–400 kHz and the output signal of the sensor was amplified by an external pre-amplifier by 60 dB. The measurement setup was the same as that depicted in Figure 5.31. An internal discharge experimental model was placed inside the transformer tank to generate PD signals.

The graphical visualization is shown in Figure 5.34. Figure 5.34(a) depicts the X - Y - Z 3D view of the intersection area of three acoustic sensors. The projections of the intersection areas on three planes (X - Y , Y - Z , and X - Z) are depicted in Figures 5.34(b)–(d), respectively. In the figures, the gray line denotes the center of the intersection volume by each set of two sensors. Line ① is a central line of an overlapped volume by sensors 1 and 2, line ② by sensors 1 and 3, and line ③ by sensors 2 and 3. The dash line arrow indicates the possible PD location. Among these, line ② by sensors 1 and 3 is the most significant since it appears on all three projection planes.

From the above results, the PD source is estimated at the point $X = 25$ cm, $Y = 18$ cm, $Z = 30$ cm. In Figure 5.35, this point is drawn on the transformer tank as a gray star mark. The error lies within 10% of full length in each axis. The error is thought to be caused mainly by the inaccuracy of sound wave velocity and triggering time. However, the result indicates a reasonably close estimation of the location to the actual PD source.

5.7.2 UHF PD Measurement

Except for the acoustic methods, UHF can also be used for PD source localization. There are different types of UHF sensor that can be used in power transformers.

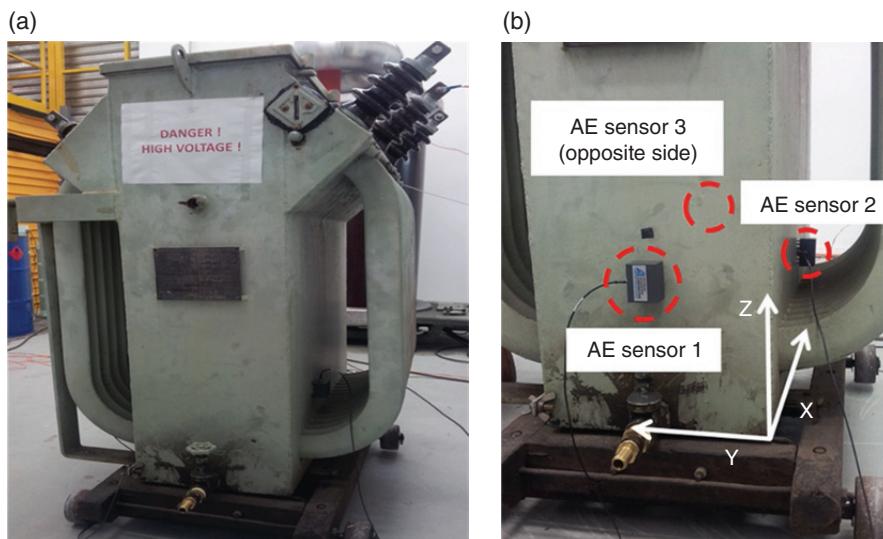


Figure 5.33 Tested transformer tank and AE sensor installation.

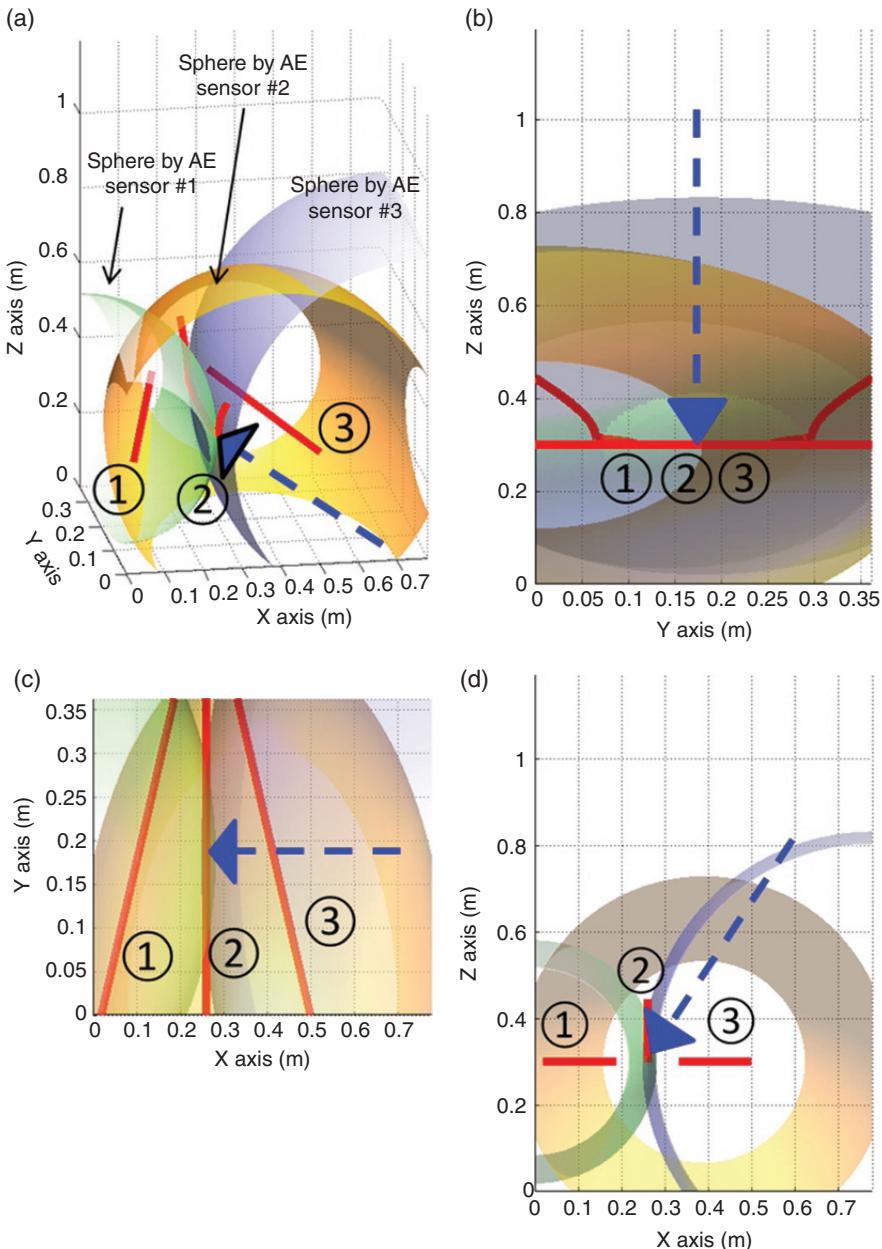


Figure 5.34 The results of spatial intersectional PD localization: (a) X-Y-Z axis view; (b) Y-Z axis view; (c) X-Y axis view; and (d) X-Z axis view.

Internal disc couplers are used only for pre-existing locations of transformers specified by manufacturers, while external disc couplers can be installed in existing dielectric windows [90]. An oil-valve probe provides a convenient and online means of installation by putting the sensor into the transformer tank through

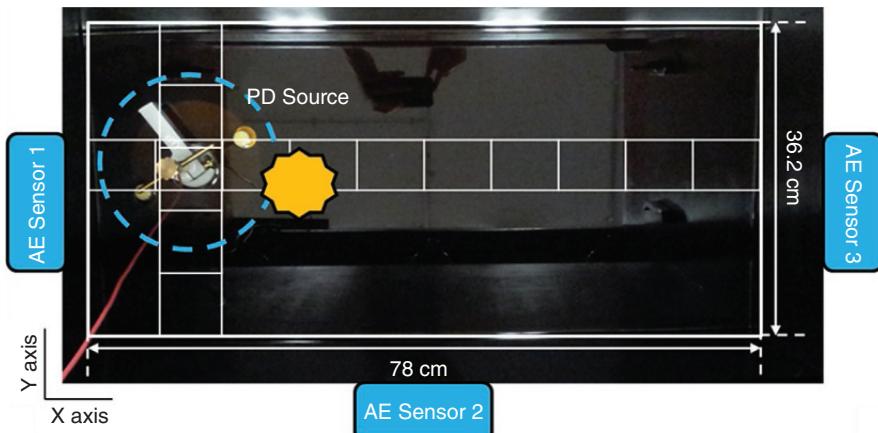


Figure 5.35 The final PD source localization result.

the oil valve. One of the advantages of adopting UHF detection is associated with greater immunity to noise, since the noise in substations is primarily below a few megahertz [26, 91, 92], except for the noise sources from communication systems. A disadvantage of UHF detection is that it is less flexible, since UHF sensors and associated accessories are usually specific to a particular type of equipment [28].

5.8 Summary

In this chapter, a comprehensive PD signal analysis from various PD measurement methods to PD signal de-noising and finally PD source classification and localization has been presented. Several newly developed signal processing techniques have also been detailed. The signal processing methods have been verified by case studies using PD signals acquired from PD source models and substation transformers to indicate their effectiveness on PD signal analysis in practice.

References

- 1 Jarman P, Hooton R, Walker L, Zhong Q, Ishak T, Wang Z. Transformer life prediction using data from units removed from service and thermal modelling. CIGRE Session 2010, Paris.
- 2 Picher P, Riendeau S, Gauvin M, Léonard F, Dupont L, Goulet J *et al.* New technologies for monitoring transformer tap-changers and bushings and their integration into a modern IT infrastructure CIGRE Session 2012, Paris.
- 3 CountryEnergy's Electricity Network Regulatory Proposal 2009–2014, 2008.
- 4 Grechko O, Kalacheva I. Current trends in the development of in-service monitoring and diagnostic systems for 110–750 kV power transformers. *Applied*

- Energy Research: Russian Journal of Fuel, Power, and Heat Systems* 1996; 34: 84–97.
- 5 Lee K, Schneider J. Rockport transient voltage monitoring system: Analysis and simulation of record waveforms. *IEEE Transactions on Power Delivery* 1989; 4: 1794–1805.
 - 6 Bartley W. Analysis of transformer failures. Paper presented at the 36th Annual Conference of the International Association of Engineering Insurers, Stockholm, Sweden, 2003.
 - 7 Minhas M, Reynders J, Klerk PD. Failures in power system transformers and appropriate monitoring techniques. Paper presented at the 11th International Symposium on High Voltage Engineering, London, 1999.
 - 8 Khawaja R. Monitoring of high voltage oil-impregnated transformer insulation. Ph.D. dissertation, University of New South Wales, 2006.
 - 9 Lai K, Phung B, Blackburn T. Application of data mining on partial discharge. Part I: Predictive modelling classification. *IEEE Transactions on Dielectrics and Electrical Insulation* 2010; 17: 846–854.
 - 10 IEEE Std. 60270. High-Voltage Test Techniques – Partial Discharge Measurements, 2000.
 - 11 Ma X, Zhou C, Kemp I. Automated wavelet selection and thresholding for PD detection. *IEEE Electrical Insulation Magazine* 2002; 18: 37–45.
 - 12 Zhou X, Zhou C, Kemp I. An improved methodology for application of wavelet transform to partial discharge measurement denoising. *IEEE Transactions on Dielectrics and Electrical Insulation* 2005; 12: 586–594.
 - 13 Ma X, Zhou C, Kemp I. Interpretation of wavelet analysis and its application in partial discharge detection. *IEEE Transactions on Dielectrics and Electrical Insulation* 2002; 9: 446–457.
 - 14 Zhang H, Blackburn T, Phung B, Sen D. A novel wavelet transform technique for on-line partial discharge measurements. Part 1: WT de-noising algorithm. *IEEE Transactions on Dielectrics and Electrical Insulation* 2007; 14: 3–14.
 - 15 Zhang H. Advanced techniques application of on-line partial discharge detection in power cables. Ph.D. dissertation, University of New South Wales, 2006.
 - 16 Sriram S, Nitin S, Prabhu K, Bastiaans M. Signal denoising techniques for partial discharge measurements. *IEEE Transactions on Dielectrics and Electrical Insulation* 2005; 12: 1182–1191.
 - 17 Cho N, Choi C, Lee S. Adaptive line enhancement by using an IIR lattice notch filter. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 1989; 37: 585–589.
 - 18 Happe S, Kranz H, Krause W. Advanced suppression of stochastic pulse shaped partial discharge disturbances. *IEEE Transactions on Dielectrics and Electrical Insulation* 2005; 12: pp. 265–275.
 - 19 Muhr M, Strehl T, Gulski E, Feser K, Gockenbach E, Hauschild W *et al.* Sensors and sensing used for non-conventional PD detection. CIGRE Session 2006, Paris.
 - 20 Lai K. *Application of data mining techniques for characterisation and recognition of partial discharge*. Ph.D. dissertation, University of New South Wales, 2010.
 - 21 Macedo E, Araújo D, Costa E, Freire R, Lopes W, Torres I *et al.* Wavelet transform processing applied to partial discharge evaluation. Paper presented at the 25th International Congress on Condition Monitoring and Diagnostic Engineering Management, Huddersfield, 2012.

- 22 WG21.16. Partial Discharge Detection in Installed HV Extrude Cable Systems, 2001.
- 23 Kheirmand A, Leijon M, Gubanski S. New practices for partial discharge detection and localization in large rotating machines. *IEEE Transactions on Dielectrics and Electrical Insulation* 2003; 10: 1042–1052.
- 24 Liu Y, Lin F, Zhang Q, Zhong H. Design and construction of a Rogowski coil for measuring wide pulsed current. *IEEE Sensors Journal* 2011; 11: 123–130.
- 25 Rubio-Serrano J, Rojas-Moreno M, Posada J, Martinez-Tarifa J, Robles G, Garcia-Souto J. Electro-acoustic detection, identification and location of partial discharge sources in oil–paper insulation systems. *IEEE Transactions on Dielectrics and Electrical Insulation* 2012; 19: 1569–1578.
- 26 Sinaga H. Detection, identification and localization of partial discharges in power transformers using UHF techniques. Ph.D. dissertation, University of New South Wales, 2012.
- 27 Stone G. Partial discharge diagnostics and electrical equipment insulation condition assessment. *IEEE Transactions on Dielectrics and Electrical Insulation* 2005; 12: 891–903.
- 28 Chakravorti S, Dey D, Chatterjee B. *Recent Trends in the Condition Monitoring of Transformers: Theory, Implementation and Analysis*. Berlin: Springer, 2013.
- 29 Jin X. Classification of partial discharges of power transformers with neural network applications. Ph.D. dissertation, University of Southern California, 2000.
- 30 Su C. *Electromagnetic Transients in Transformer and Rotating Machine Windings*. Pennsylvania, PA: IGI Global, 2013.
- 31 Chan J, Ma H, Saha T, Ekanayake C. Self-adaptive partial discharge signal denoising based on ensemble empirical mode decomposition and automatic morphological thresholding. *IEEE Transactions on Dielectrics and Electrical Insulation* 2014; 21: 294–303.
- 32 Chan J, Ma H, Saha T. Time-frequency sparsity map on automatic partial discharge sources separation for power transformer condition assessment. *IEEE Transactions on Dielectrics and Electrical Insulation* 2015; 22(4): 2271–2283.
- 33 Mallat S. *A Wavelet Tour of Signal Processing: The sparse way*. New York: Academic Press, 2009.
- 34 Salomon D, Motta G, Bryant D. *Data Compression: The complete reference*, 4th edn. London: Springer, 2007.
- 35 Gaouda A, Ei-Hag A, Abdel-Galil T, Salama M, Bartnikas R. On-line detection and measurement of partial discharge signals in a noisy environment. *IEEE Transactions on Dielectrics and Electrical Insulation* 2008; 15: 1162–1173.
- 36 Li J, Jiang T, Grzybowski S, Cheng C. Scale dependent wavelet selection for denoising of partial discharge detection. *IEEE Transactions on Dielectrics and Electrical Insulation* 2010; 17: 1705–1714.
- 37 Zhang H, Blackburn T, Phung B, Sen D. A novel wavelet transform technique for on-line partial discharge measurements. Part 2: On-site noise rejection application. *IEEE Transactions on Dielectrics and Electrical Insulation* 2007; 14: 15–22.
- 38 Song X, Zhou C, Hepburn D, Zhang G. Second generation wavelet transform for data denoising in PD measurement. *IEEE Transactions on Dielectrics and Electrical Insulation* 2007; 14: 1531–1537.

- 39 Zhongrong X, Ju T, Caixin S. Application of complex wavelet transform to suppress white noise in GIS UHF PD signals. *IEEE Transactions on Power Delivery* 2007; 22: 1498–1504.
- 40 Dey D, Chatterjee B, Chakravorti S, Munshi S. Cross-wavelet transform as a new paradigm for feature extraction from noisy partial discharge pulses. *IEEE Transactions on Dielectrics and Electrical Insulation* 2010; 17: 157–166.
- 41 Chang C, Jin J, Kumar S, Su Q, Hoshino T, Hanai M et al. Denoising of partial discharge signals in wavelet packets domain. *IET Science, Measurement and Technology* 2005; 152: 129–140.
- 42 Huang N, Shen Z, Long S, Wu M, Shih H, Zheng Q et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society London, Series A* 1998; 454: 903–995.
- 43 Chan J, Tse P. A novel, fast, reliable data transmission algorithm for wireless machine health monitoring. *IEEE Transactions on Reliability* 2009; 58: 295–304.
- 44 Wu Z, Huang N. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis* 2009; 1: 1–41.
- 45 Zhang L, Xu J, Yang J, Yang D, Wang D. Multiscale morphology analysis and its application to fault diagnosis. *Mechanical Systems and Signal Processing* 2008; 22: 597–610.
- 46 Nikolaou N, Antoniadis I. Application of morphological operators as envelope extractors for impulsive-type periodic signals. *Mechanical Systems and Signal Processing* 2003; 17: 1147–1162.
- 47 Dong Y, Liao M, Zhang X, Wang F. Faults diagnosis of rolling element bearings based on modified morphological method. *Mechanical Systems and Signal Processing* 2011; 25: 1276–1286.
- 48 Chan J, Ma H, Saha T. Automatic blind equalization and thresholding for partial discharge measurement in power transformer. *IEEE Transactions on Power Delivery* 2014; 29: 1927–1938.
- 49 Jelonnek B, Boss D, Kammeyer K. Generalized eigenvector algorithm for blind equalization. *Signal Processing* 1997; 61: 237–264.
- 50 Shalvi O, Weinstein E. New criteria for blind deconvolution of nonminimum phase systems (channels). *IEEE Transactions on Information Theory* 1990; 36: 312–321.
- 51 Cavallini A, Chen X, Montanari G, Ciani F. Diagnosis of EHV and HV transformers through an innovative partial-discharge-based technique. *IEEE Transactions on Power Delivery* 2010; 25: 814–824.
- 52 Cavallini A, Montanari G. A new methodology for the identification of PD in electrical apparatus: Properties and applications. *IEEE Transactions on Dielectrics and Electrical Insulation* 2005; 12: 203–215.
- 53 Contin A, Cavallini A, Montanari G, Pasini G, Puletti F. Digital detection and fuzzy classification of partial discharge signals. *IEEE Transactions on Dielectrics and Electrical Insulation* 2002; 9: 335–348.
- 54 Cavallini A, Montanari G, Contin A, Puletti F. A new approach to the diagnosis of solid insulation systems based on PD signal inference. *IEEE Electrical Insulation Magazine* 2003; 19: 23–30.
- 55 Ester M, Kriegel H, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, 1996.

- 56** Hao L, Lewin P, Hunter J, Swaffield D, Contin A, Walton C *et al.* Discrimination of multiple PD sources using wavelet decomposition and principal component analysis. *IEEE Transactions on Dielectrics and Electrical Insulation* 2011; 18: 1702–1711.
- 57** Daszykowski M, Walczak B, Massart D. Looking for natural patterns in data. Part 1. Density-based approach. *Chemometrics and Intelligent Laboratory Systems* 2001; 56: 83–92.
- 58** Hoof M, Patsch R. Pulse-sequence analysis: A new method for investigating the physics of PD-induced ageing. *IET Science, Measurement and Technology* 1995; 142: 95–101.
- 59** Kranz H. PD pulse sequence analysis and its relevance for on-site PD defect identification and evaluation. *IEEE Transactions on Dielectrics and Electrical Insulation* 2005; 12: 276–284.
- 60** Antoni J. The spectral kurtosis: A useful tool for characterising non-stationary signals. *Mechanical Systems and Signal Processing* 2006; 20: 282–307.
- 61** Antoni J. Fast computation of the kurtogram for the detection of transient faults. *Mechanical Systems and Signal Processing* 2007; 21: 108–124.
- 62** Randall R. *Vibration-Based Condition Monitoring: Industrial, Aerospace and Automotive Applications*: A. New York: Wiley, 2011.
- 63** IEEE St. 61198. Mineral Insulating Oils – Methods for the Determination of 2-Furfural and Related Compounds, 1993.
- 64** Li J, Du Q, Sun C. An improved box-counting method for image fractal dimension estimation. *Pattern Recognition* 2009; 42: 2460–2469.
- 65** Chan J, Ma H, Saha T, Ekanayake C. Stochastic noise removal on partial discharge measurement for transformer insulation diagnosis. Paper presented at the IEEE Power & Energy Society General Meeting, Washington, D.C., 2014.
- 66** Li J, Jiang T, Harrison R, Grzybowski S. Recognition of ultra high frequency partial discharge signals using multi-scale features. *IEEE Transactions on Dielectrics and Electrical Insulation* 2012; 19: 1412–1420.
- 67** Gu F, Chang H, Kuo C. Gas-insulated switchgear PD signal analysis based on Hilbert–Huang transform with fractal parameters enhancement. *IEEE Transactions on Dielectrics and Electrical Insulation* 2013; 20: 1049–1055.
- 68** Gulski E. Digital analysis of partial discharges. *IEEE Transactions on Dielectrics and Electrical Insulation* 1995; 2: 822–837.
- 69** Randall R. Applications of spectral kurtosis in machine diagnostics and prognostics. *Key Engineering Materials* 2005; 293/294: 21–32.
- 70** Wu Z, Huang N. A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings of the Royal Society London, Series A* 2004; 460: 1597–1611.
- 71** Li Y, Tse P, Wang X. Recovery of vibration signal based on a super-exponential algorithm. *Journal of Sound and Vibration* 2008; 311: 537–553.
- 72** Wang D, Tse P. A new blind fault component separation algorithm for a single-channel mechanical signal mixture. *Journal of Sound and Vibration* 2012; 331: 4956–4970.
- 73** Chan J, Ma H, Saha T. Hybrid method on signal de-noising and representation for online partial discharge monitoring of power transformers at substations. *IET Science, Measurement and Technology* 2015; 9(7): 890–899.
- 74** Mazzetti C, Mascioli FF, Baldini F, Panella M, Risica R, Bartnikas R. Partial discharge pattern recognition by neuro-fuzzy networks in heat-shrinkable joints

- and terminations of XLPE insulated distribution cables. *IEEE Transactions on Power Delivery* 2006; 21: 1035–1044.
- 75 Venkatesh S, Gopal S. Orthogonal least square center selection technique – A robust scheme for multiple source partial discharge pattern recognition using radial basis probabilistic neural network. *Expert System Applications* 2011; 38: 8978–8989.
- 76 Wang K, Liao R, Yang L, Li J, Grzybowski S, Hao J. Optimal features selected by NSGA-II for partial discharge pulses separation based on time-frequency representation and matrix decomposition. *IEEE Transactions on Dielectrics and Electrical Insulation* 2013; 20: 825–838.
- 77 Bartnikas R. Partial discharges: Their mechanism, detection and measurement. *IEEE Transactions on Dielectrics and Electrical Insulation* 2002; 9(5): 763–808.
- 78 Sahoo NC, Salama MMA, Bartnikas R. Trends in partial discharge pattern classification: A survey. *IEEE Transactions on Dielectrics and Electrical Insulation* 2005; 12(2): 248–264.
- 79 Ma H, Chan J, Saha T, Ekanayake C. Pattern recognition techniques and their applications for automatic classification of artificial partial discharge sources. *IEEE Transactions on Dielectrics and Electrical Insulation* 2013; 20: 468–478.
- 80 Chan J, Ma H, Saha T. Partial discharge pattern recognition using multi-scale feature extraction and support vector machine. Paper presented at the IEEE Power & Energy Society General Meeting, Vancouver, Canada, 2013.
- 81 Gulski E, Kreuger F. Computer-aided recognition of discharge sources. *IEEE Transactions on Electrical Insulation* 1992; 27: 82–92.
- 82 Bishop CM. *Pattern Recognition and Machine Learning*. Berlin: Springer, 2006.
- 83 Lin CF, Wang SD. Fuzzy support vector machines. *IEEE Transactions on Neural Networks* 2002; 13: 464–471.
- 84 Bishop CM. *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press, 1995.
- 85 Jain AK, Duin RPW, Mao J. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2000; 22: 4–37.
- 86 Hinton G, Roweis S. Stochastic neighbor embedding. *Advances in Neural Information Processing Systems* 2002; 15: 833–840.
- 87 Seo J. Advanced signal processing techniques for online partial discharge monitoring and assessment of power transformers. M.Phil. dissertation, University of Queensland, 2014.
- 88 IEC 62478. High Voltage Test Techniques – Measurement of Partial Discharges by Electromagnetic and Acoustic Methods, draft version.
- 89 Phung B, Blackburn TR, Liu Z. Acoustic measurements of partial discharge signals. *Journal of Electrical & Electronics Engineering* 2001; 21(1): 41–47.
- 90 Chen H. Partial discharge identification system for high-voltage power transformers using fractal feature-based extension method. *IET Science, Measurement and Technology* 2012; 7: 77–84.
- 91 IEEE Std. 60505. Evaluation and Qualification of Electrical Insulation Systems, 2011.
- 92 Koo J, Jung S, Ryu C, Lee S, Lee B. Identification of insulation defects in gas-insulated switchgear by chaotic analysis of partial discharge. *IET Science, Measurement and Technology* 2010; 4: 115–124.