Impulsive noise reduction for transient Earth voltage-based partial discharge using Wavelet-entropy

Luo, Guomin; Zhang, Daming; Tseng, King Jet; He, Jinghan

2016

Luo, G., Zhang, D., Tseng, K. J., & He, J. (2016). Impulsive noise reduction for transient Earth voltage-based partial discharge using Wavelet-entropy. IET Science, Measurement & Technology, 10(1), 69-76.

https://hdl.handle.net/10356/82896

https://doi.org/10.1049/iet-smt.2014.0203

IET Science, Measurement & Technology

Research Article

IET Journals
The Institution of Engineering and Technology

Impulsive noise reduction for transient Earth voltage-based partial discharge using Wavelet-entropy

ISSN 1751-8822
Received on 23rd January 2014
Revised on 14th January 2015
Accepted on 23rd January 2015
doi: 10.1049/iet-smt.2014.0203
www.ietdl.org

Guomin Luo[1] ✉, *Daming Zhang*[2], *King Jet Tseng*[3,4], *Jinghan He*[1]

[1]*School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, People's Republic of China*
[2]*School of Electrical Engineering and Telecommunications, New South Wales University, Sydney 2052, Australia*
[3]*SinBerBEST Program, Berkeley-Singapore Alliance for Research in Singapore, Singapore*
[4]*School of EEE, Nanyang Technological University, Singapore*
✉ *E-mail: gmluo@bjtu.edu.cn, guominluo@hotmail.com*

Abstract: Partial discharge (PD) is caused by the localised electrical field intensification in insulating materials. Early detection and accurate measurement of PD are very important for preventing premature failure of the insulating material. Detection of PDs in metal-clad apparatus through the transient Earth voltage method is a promising approach in non-intrusive on-line tests. However, the electrical interference from background environment remains the major barrier to improving its measurement accuracy. In this study, a wavelet-entropy-based PD de-noising method has been proposed. The unique features of PD are characterised by combining wavelet analysis that reveals the local features and entropy that measures the disorder. With such features, a feed-forward back-propagation artificial neural network is adopted to recognise the actual PDs from noisy background. Comparing with other methods such as the energy-based method and the similarity-comparing method, the proposed wavelet-entropy-based method is more effective in PD signal de-noising.

1 Introduction

Partial discharge (PD) measurement is an effective means of insulation monitoring. A lot of PD measuring methods have been proposed and reported to have good performance in detecting and identifying PDs in metal-clad apparatus such as the gas insulated switchgear and so on [1]. However, most existing methods cannot provide both good sensitivity and convenient installation at the same time. Traditional PD sensors such as coupling capacitors are usually used in off-line applications. Internal ultra-high frequency (UHF) couplers mounted on the inner side of the metal-cladding can provide real-time monitoring, but arranging a shutdown specifically to fit such couplers can rarely be justified for operational equipment [2]. In such cases, the non-intrusive measurement is more preferred. The acoustic detection method is a well-known non-intrusive technique. It has the advantages of being no interruption of operation, no response to electromagnetic interferences and capable of locating PD sources, but its sensitivity decreases greatly when apparatuses have more complex structure and metallic shield [3]. Unlike acoustic wave which can be easily attenuated, the radiating electromagnetic wave from PDs can induce a small pulse-like voltage on the inner surface of cladding and propagate a longer way. Such signal which is known as transient Earth voltage (TEV) leaks into the external space through dielectric openings or joints on the cladding and transmits to the earth through external cladding surface, on to which the non-intrusive TEV sensors are mounted [4, 5]. The implementation of the TEV-based technique is convenient to use and no electrical power shutdown is needed. It has been widely applied in high voltage (HV) and medium voltage apparatuses which are not fully shielded or with dielectric openings on their metalwork.

As a non-intrusive on-line measurement, the major problem needs to be addressed is the interferences from surroundings. Although the improvement of sensor design can shield many external noises, the TEV signals may still be polluted during propagation. The ordinary hardware-based de-noising methods such as differential circuit and noise gating cannot perform effectively since the TEV sensors should be relocated from one panel to another during detections [6]. The software-based methods provide more flexible solutions, among which recognising pulse features by AI classifier is the most popular way in extracting real PDs [7, 8]. It should be noticed that the way how the signals are characterised has direct influence on the de-noising results. The statistical evaluation such as pulse fingerprint and probability analysis which studies the φ-q-n distribution patterns extract the common features of a group of pulses [6, 9]. These methods cannot distinguish the noisy pulses from PDs when they are mixed in a same signal. To reject the impulsive interferences one by one, some methods such as the waveform analysis and the frequency filtering techniques were proposed, but the time-domain methods cannot distinguish pulses occurring at the same time, while the frequency-domain methods cannot reject interferences that fall into the same frequency band with PDs [10]. Time-frequency analysis which reveals the energy variations with both time and frequency is more effective in representing the unique features of PDs and noises. Wavelet analysis with a varying resolution was proved to be an ideal tool in extracting time-frequency PD features [11–14]. Since the commonly used wavelet features, that is, the energy-based features, would be inconsistent and unstable to represent pulses from a same source if the pulses' energy varies a lot, a more fuzzy and intelligent classifier is necessary to obtain better results. To simplify the design of the classifier and make the PD recognition less dependent on the AI technique, a reliable and robust wavelet-based characterisation which can stand the variations of PDs and provide unique features should be explored.

In this paper, we propose a novel approach, which adopts the wavelet-entropy and artificial neural network ANN to reject impulsive noises in TEV-based PD measurement. The basic idea is to characterise the pulses by wavelet-entropy and recognise real PD pulses from noisy signals. Comparing with the traditional magnitude-based features, the wavelet-entropy only measures the disorder of wavelet distributions and is thus stable to represent

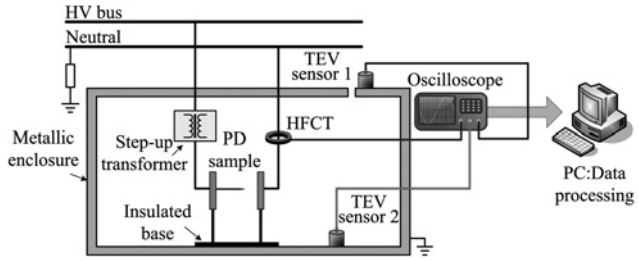

Fig. 1 *Measurement setup of laboratory test*

PDs. A series of experiments were performed to prove the feasibility and effectiveness of our approach.

The organisation of the remaining parts is as follows: Section 2 presents the experimental setup and introduces our non-intrusive sensor. Time-frequency characteristics of the noises and the wavelet features of the pulses are investigated in Section 3. The entropy theory and its advantages in representing PDs are illustrated in Section 4. Section 5 portrays the details of the proposed de-noising procedure. The parameters of ANN selection is introduced in Section 6 according to the wavelet-entropy features. Finally, the performance of the proposed PD de-noising method and comparisons with the prior ones are discussed in Section 7.

2 Experimental setup

In TEV measurement, the non-intrusive sensors are mounted to the external surface of the grounded metallic enclosure. They detect TEV signals that are produced by radiating electromagnetic waves of PDs and leak out the metal enclosure [4]. To demonstrate the feasibility of the TEV measurement and collect TEV-based PD signals, the following experimental test setup (as shown in Fig. 1) is adopted.

This system includes four parts: a metallic enclosure which simulates the cladding of electrical apparatus, a PD generation device which is placed inside the enclosure, the PD sensors that include two non-intrusive sensors and one commercial high frequency current transformer (HFCT) and the signal display (Tektronix TDS7104) and processing equipment.

The enclosure is an aluminium box with size of 1 m × 1 m × 2 m. The PD generator is with a needle-to-plane structure and simulates the PD sources in HV equipment. The detailed structure of the non-intrusive sensors was described in [15] and is also shown in Fig. 2*a*. The sensor is composed of four parts: part 1 is the female BNC interface, part 2 is a metal cover that shields environment interferences, part 3 is an outer conductor and part 4 is an inner conductor. The frequency response of the non-intrusive sensor is from 80 Hz to 9 MHz. Its amplitude response is portrayed in Fig. 2*b*. Here, $|U_o/U_i|$ equals to the absolute value of output voltage divided by input voltage at different frequencies. In such frequency range, the impulsive PD energy is less likely to attenuate [16].

Table 1 Pulse responses of industrial HFCT and proposed non-intrusive TEV sensor

Sensors	Parameters			
	Rise time	Fall time	Low freq. –3 dB, Hz	High freq. –3 dB, MHz
TEV sensor	≃38 ns	≃2.9 ms	80	9
HFCT	27.2 ns	2.09 us	around 110 k	around 12

The performance of the proposed non-intrusive sensor is demonstrated by comparing with an industrial HFCT (HFCT1400–100, produced by High Voltage Partial Discharge Ltd., U.K.). Their parameters are listed in Table 1.

Owing to the small lower cut-off frequency (80 Hz), the PDs detected by the proposed TEV sensor would reveal long damping waveforms which can cause pulse-overlapping and difficult recognition. Accordingly, a Butterworth high-pass filter is employed before data analysis. The filter order is 3, and the cut-off frequency is selected to be 100 kHz [1, 16]. By adding a filter, the proposed PD measuring system can be as effective as industrial HFCTs. A number of experimental tests done in [15] have proved its performance. Fig. 3 shows three signals before and after filtering. Two of them are from the TEV sensors, one being attached to the inner surface of the enclosure and the other being attached to the outer surface of the enclosure. The other signal is from HFCT. Before filtering, the TEV-based PDs contain low frequency variations because of the fluctuation of sinusoidal energy. After filtering, the three signals are almost the same. These tests verify the theory of TEV and provide the basis for field test using non-intrusive PD sensing technique.

3 Features of PDs and noises

3.1 Noises in on-line PD measurement

In on-line PD measurement, noises can be produced by many kinds of sources and coupled with systems in different ways, showing different features. Therefore noise rejection has no omnipotent solution and is best approached by devising several techniques, each of which is tailored for a specific kind of noise. Many previous works and field tests suggest that the noises that may probably be rejected during on-site PD measurements are: the white noise, the sinusoidal and harmonic interferences, the repetitive pulses and the random pulses [17]. Those noises have different patterns and can be classified into two groups: the non-impulsive noise and the impulsive noise.

The non-impulsive noise includes the white noise and the sinusoidal interference. The white noise can be commonly found in on-site PD measurement. It generally comes from the measuring

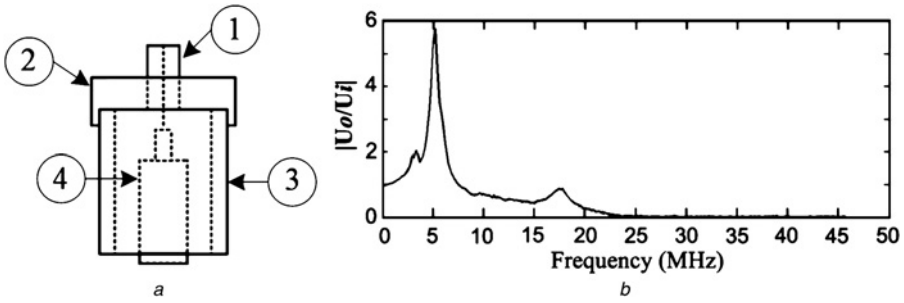

Fig. 2 *Coaxial PD sensor developed for non-intrusive PD measurement*

a Construction of the sensor
b Frequency response

IET Sci. Meas. Technol., 2016, Vol. 10, Iss. 1, pp. 69–76

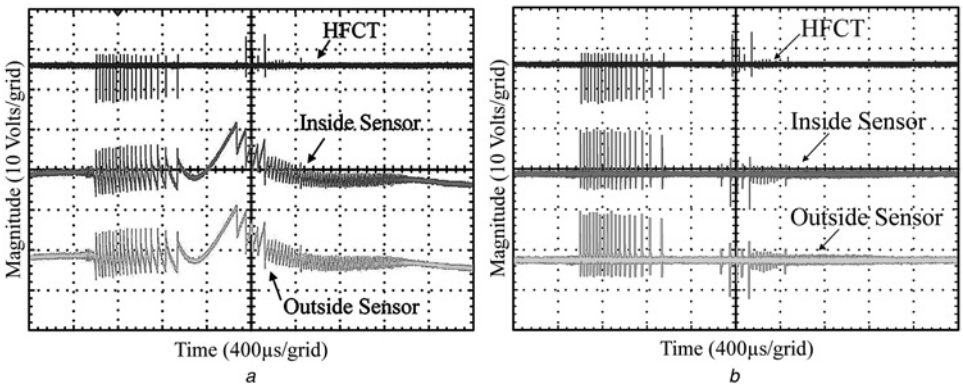

Fig. 3 *Measured PD signals from different sensors before and after filtering*

a Measured signals before filtering
b Measured signals after filtering

equipment or noisy environment. Its time-frequency spectrum in Fig. 4*a* shows that the white noise has equal power density throughout the whole frequency range. The sinusoidal interferences are usually generated by the communication equipment, and often reveal oscillating-frequency-dependent lines in their time-frequency spectrums such as the one shown in Fig. 4*b*. The non-impulsive noises can be effectively rejected by the wavelet thresholding techniques because of their time-invariant energy distributions in each frequency band [18].

The impulsive noise includes the repetitive pulse and the random pulse. The repetitive pulses are usually generated by electronics equipment. The random pulses generally come from switch operations, fault transients, lightning disturbances and so on. These noises also have impulsive waveforms and time-variant spectrums in time-frequency domain, as shown in Figs. 4*c* and *d*. Obviously, they cannot be removed by the level-dependent wavelet thresholding techniques and more effective methods should be explored.

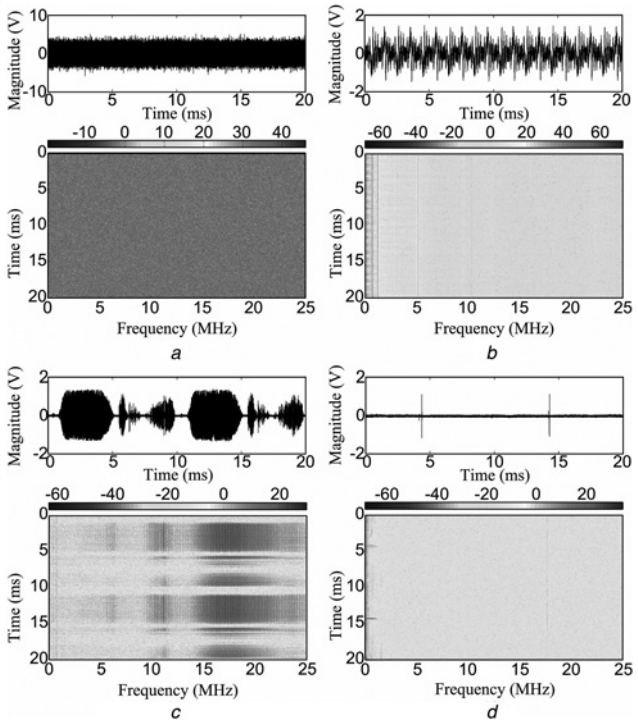

Fig. 4 *Noise types and their time-frequency spectrums*

a White noise and its time-frequency spectrum
b Sinusoidal interference and its time-frequency spectrum
c High-frequency repetitive pulses and its time-frequency spectrum
d Random pulses and its time-frequency spectrum of random pulses

3.2 Wavelet features of PD and the impulsive noises

Wavelet transform is effective in identifying sharp edge transitions of impulsive signals: the amplitudes of coefficients at fine scales are very small (almost zero) when the signal is piecewise regular, and the large magnitude coefficients only occur exclusively near the areas of major spatial activities [17, 19]. Using TEV sensor described in Section 2, the PD pulses generally have a short rise time (about 40 ns) and a wide spectrum (about 9 MHz), while the impulsive noises cannot have such features because of their mechanisms and propagation losses. To show the differences between PDs and noisy pulses, their wavelet coefficients of all scales are shown in Fig. 5. According to the frequency response range of the measuring system, the second coiflet is adopted and the decomposition level is selected to be 9. All scales are plotted to a same time axis for easier comparison.

Fig. 5*a* portrays the wavelet coefficients of a single PD pulse collected in the test shown in Fig. 1. Large-amplitude coefficients can be found in all decomposition levels. Such phenomenon suggests a wide frequency range of the PD pulse. By comparing the maximum magnitude of the coefficients on each level, it is easy to conclude that most energy of the PD pulse is concentrated at lower frequency bands such as the seventh (d7: 195.3–390.6 kHz) and the eighth (d8: 97.65–195.3 kHz) levels. Accordingly, the PD pulse has a less oscillating waveform.

Figs. 5*b* and *c* show the wavelet coefficients of two typical examples of the impulsive noises: one being a repetitive pulse from the electronics equipment and the other being a random noise from the switching operation. TEV measurement is a local PD detecting method, and the interference sources are often far away from TEV sensors in practical applications. Owing to the energy attenuation and loss during propagation, the TEV-measured impulsive noises usually have narrower frequency spectrums than the local PD pulses. To simulate such energy loss, the TEV sensor is located a few metres away from the pulse sources. Thus, the wavelet coefficients of the repetitive noise in two higher frequency bands (d1 to d2: 6.25–25 MHz) are almost zero, and three high-frequency wavelet decomposition levels (d1–d3: 3.125–25 MHz) of the random pulse contain a large number of zeroes. Moreover, because of the different mechanisms and the resonance or distortion during propagation, the impulsive noises may have large-magnitude energy in high frequency bands which can cause oscillating waveforms. For example, the largest coefficients of the repetitive noise are in the fifth (d5: 781.25 kHz–1.56 MHz) and the sixth (d6: 390.625–781.25 kHz) levels, and the largest energy of the random pulse is in the sixth (d6: 390.625–781.25 kHz) level. The large-energy-containing frequency bands of both noisy pulses are higher than those of PDs.

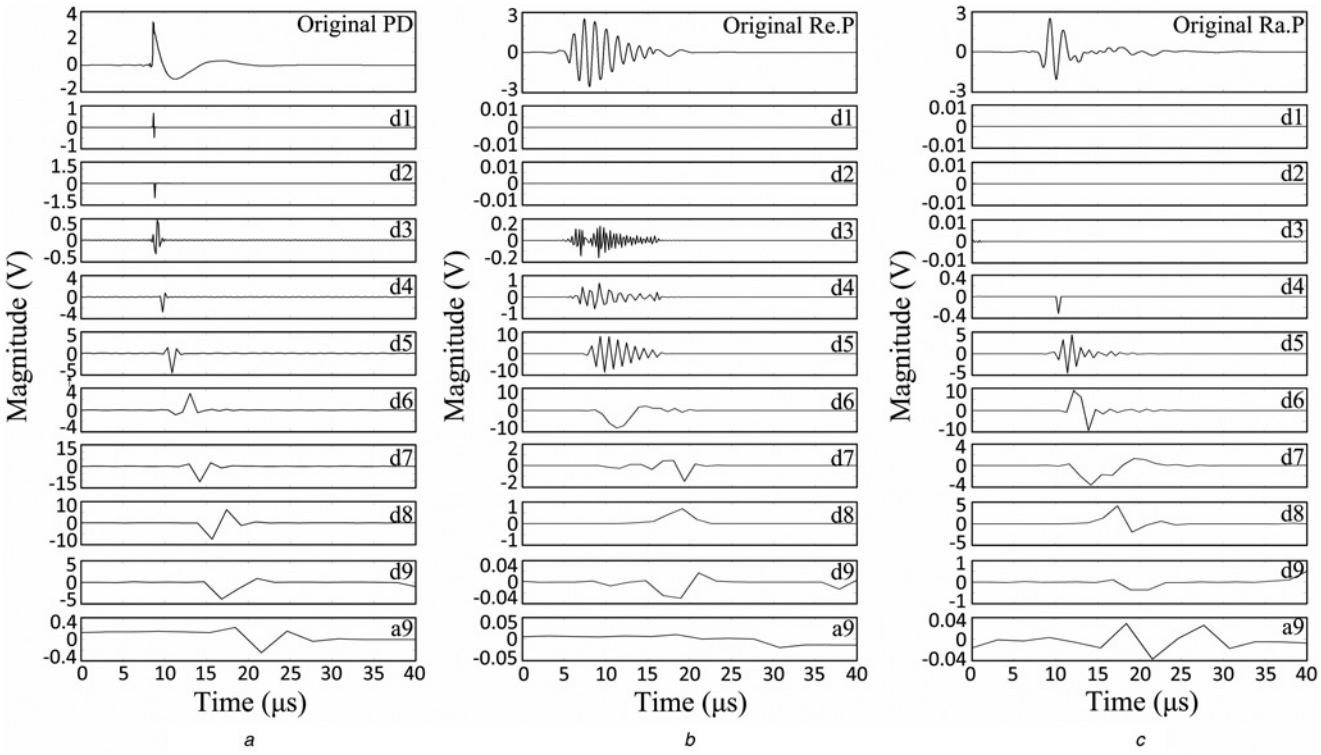

Fig. 5 *Wavelet coefficients of three typical pulses*

a PD pulse and its wavelet coefficients
b Repetitive pulse and its wavelet coefficients
c Random pulse and its wavelet coefficients
'Re.' is short for 'repetitive pulse', and 'Ra.' is short for 'random pulse'
Sampling rate is 50 MSamples/s

Comparing with noisy pulses, the PD pulse has more non-zero coefficients on high-frequency scales (wider frequency spectrum) and less oscillating waveforms, since most large-magnitude coefficients are contained in lower frequency bands. It is possible to adopt these features to classify PDs and impulsive noises. However, the number of wavelet coefficient is too large to represent the features for further classification. An operator is thus needed to reduce the feature dimension and characterise pulses properly. Describing the disorder of wavelet coefficient distributions seems to be a possible solution.

4 Entropy representation

4.1 Fundamentals of entropy

Entropy is a popular measure of uncertainty and disorder. It was first applied in power system for the control of generators [20]. Thereafter, many different applications of entropy were reported in various aspects of power system such as transient signal analysis and power quality evaluation [21, 22]. These applications of entropy suggest its potential in characterising PD signals. The uncertainty or disorder of the energy distribution of a single wavelet decomposition level contains a lot of pulse information. Entropy is thus adopted to reveal those underlying features. All the wavelet coefficients of ith detail level is considered to be a set X_i. The entropy H_i of set X_i with possible values $\{x_{i1}, x_{i2}, \ldots, x_{in}\}$ is defined as follows [23]

$$H_i = -\sum_{j=1}^{n} p(x_{ij}) \log_b p(x_{ij}) \tag{1}$$

where $p(x_{ij})$ is the probability of x_{ij}. The probabilities are estimated by a statistic histogram whose x-axis contains 10 equal intervals of wavelet coefficients' magnitude. Here, b is the base of logarithm. It is usually set to 2, and the unit of entropy is 'bit' accordingly. This wavelet coefficient set is defined to be ordered when all the coefficients are zero. Almost all the wavelet coefficients concentrate in one or two bins of the statistic histogram for entropy calculation, and a small entropy is thus produced. When pulse energy appears, such coefficient set varies from order to disorder. Meanwhile, the distribution of statistic histogram becomes more uniform and produces a greater entropy value. Therefore the value of entropy can denote the order of the energy distribution of wavelet coefficients. Furthermore, its value only depends on the distribution or probability rather than the amplitude of the coefficients.

When characterising a single pulse, the entropy values of all decomposition scales form an entropy vector H. To ensure all the entropy values fall into a common range [0 1] such that the comparison and illustration will be more straightforward, the entropy ratio ρ_{Ei}, as shown in (2), is adopted in the following analysis

$$\rho_{Ei} = \frac{H_i}{||H||} \tag{2}$$

4.2 Effectiveness demonstration of entropy

High-frequency attenuation of the impulsive signals often happens in TEV measurement. The degree of this attenuation depends on many factors, such as the distance between pulse source and sensor, the surface impedance of metal cladding and so on. For a robust representation, the characteristics of a certain pulse detected by a same sensor should be similar no matter where it is collected. In other words, the features of a pulse should be insensitive to the attenuations during propagation.

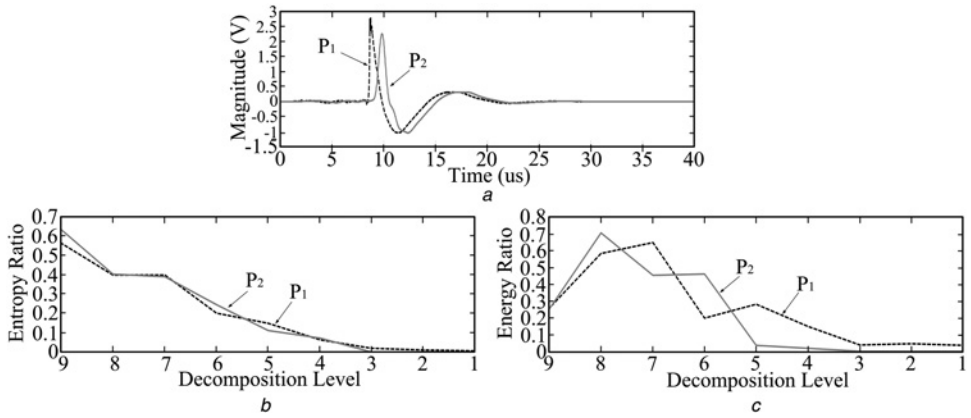

Fig. 6 *Comparison of entropy distributions and energy distributions*

a Original PD pulse P_1 and the filtered pulse P_2
b Entropy ratios of two pulses
c Energy ratios of two pulses

According to the definition of entropy in Section 2, the entropy value can be insensitive to the magnitude attenuation during propagation. This advantage is illustrated by comparing with another popular representation: the energy ratio [24]. The definition of the energy ratio can be found in Appendix.

A low-pass Butterworth filter (order 8, cut-off frequency 0.5 MHz) is adopted to simulate the high-frequency attenuation during propagation. The original PD pulse P_1, as shown in Fig. 6*a*, is collected by the TEV sensor and sampled at the rate of 25 MSamples/s. The pulse P_2 is generated by filtering P_1. The second coiflet is used and the decomposition level is 9. Figs. 6*b* and *c* illustrate the entropy ratios and the energy ratios of P_1 and P_2, respectively. As demonstrated in Fig. 6*c*, the energy of filtered pulse P_2 of levels 1–5 (d1–d5: 390.625 kHz–12.5 MHz) decreases greatly when compared with the energy ratios of P_1. The largest difference between two energy ratio vectors is around 0.26 at level 5. On the other hand, the entropy ratios of P_1 and P_2 are more similar. The largest difference is around 0.07 at level 9. Therefore when a same classifier is employed, the two pulses P_1 and P_2 are more likely to be recognised as one if the entropy-based representation is adopted.

5 PD recognition procedure

Although entropy is a robust representation, it is still not easy to judge pulse type by a simple rule. The entropy ratios of pulses of a same type, or even from a same source, may vary with many factors such as environment, propagation and so on. A classifier is thus needed to improve the de-noising efficiency. Just like the brain of human being, ANN provide a brain-like capability for solving problems. Owing to their excellent classification and recognition ability, ANNs are very popular and have been applied in many different areas, especially, the PD diagnosis [14, 25].

The proposed PD recognition procedure includes three steps: measurement, characterisation and classification.

(1) *Measurement:* Both the PDs and the impulsive noises are measured. The PD pulses are collected by the TEV measuring system in experiments shown in Fig. 1. The influence of the applied voltage is considered by carrying out tests under different applied voltages. The applied voltages are randomly selected and vary from inception voltage to the voltage that keeps stable PD occurrence. The noisy pulses were collected from an operating switchgear located in the Paya Ubi Industrial Park, Singapore. This switchgear is absolutely exclusive of PD occurrence. Two types of noisy pulses are measured: the repetitive pulses generated by the switching operation of electronics equipment, and the random pulses produced by breakers that operate occasionally. All these signals were sampled at a rate of 50 MSamples/s.

(2) *Characterisation:* As illustrated in Fig. 7, the characterisation procedure contains two steps: the wavelet thresholding and the entropy calculation.

The wavelet thresholding method is adopted here to remove the non-impulsive noises and generate a higher signal-to-noise ratio. The details of the wavelet thresholding technique have been discussed by the authors in [26]. The universal threshold with hard thresholding function is used to achieve the largest estimation risk. The level-wise threshold is $T = \sigma\sqrt{2\log_e N}$, where σ is the estimation of noise and N is the size of signal.

After rejecting non-impulsive noises, the pulse energy is left. The de-noised signal can be divided into many pulse-containing segments. One segment is assumed to include only one pulse. The segments with overlapped pulses will not be discussed in this paper. The wavelet-entropy ratio vector of each pulse-containing segment is calculated according to (1) and (2). Each segment generates an entropy ratio vector with a size J, which represents the number of decomposition level.

(3) *Classification:* All the pulse-containing segments are analysed one-by-one. The wavelet coefficients of segments classified as real PDs are kept and reconstructed while those regarded as noises are rejected and deleted. Neural networks which can 'learn' from the given patterns is employed in this research. Since the wavelet-entropy produces a more effective representation of pulses and only a limited number of samples are available, a simple network is preferred. Therefore a three-layer feed-forward back-propagation (BP) neural network which can map any non-linear inputs to their output with minimal computational overheads is used to attain satisfying recognition. Here, 30 PDs, 80 repetitive pulses and 20 random pulses are used in the training process. The training is terminated when the number of iterations reaches one thousand or when the training error is smaller than 1×10^{-7}. The whole PD recognition procedure has been realised on a MATLAB-based platform.

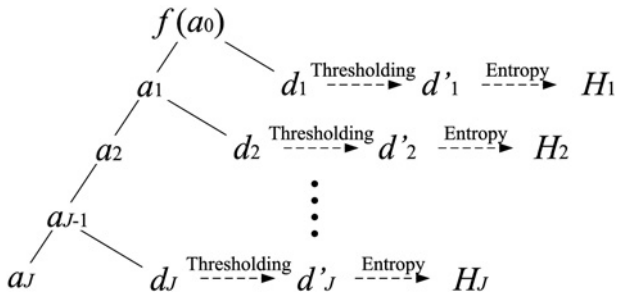

Fig. 7 *Fundamental of entropy based feature extraction*

6 Selection of architecture and functions of ANN used in classification

To reach their best performance, suitable parameters and functions of ANN are selected for a particular application. In the following contents, the architecture and functions of the network are selected according to the requirements of wavelet-entropy-based PD recognition.

6.1 Selection of network architecture

The selection of network architecture is an important issue of the ANN application. A feed-forward BP network with only one hidden layer is effective enough for PD extraction [27]. The size of the network is discussed here according to the wavelet-entropy features.

(1) *Size of input and output layers:* Generally, smaller size of network leads to less calculation, shorter training duration and faster response. As long as the input layer and output layer can represent the patterns of external environment, the number of nodes should be as few as possible [28].

The input of ANN has a same size with the entropy ratio vector, which equals to the wavelet decomposition level. Its size should be large enough to cover the whole pulse spectrum. As aforementioned, the signals are sampled at 50 MSamples/s and the lower cut-off frequency of TEV measuring system is 100 kHz. Therefore the smallest decomposition scale is 8 and the widest frequency range of the final approximation is from DC to 97.65 kHz ($50\,\text{MHz}/2^{8+1}$) accordingly. However, to cover the PD energy that is not fully removed by the 3-order Butterworth high-pass filter in the proposed measuring system, the minimum level is chosen as 9 and the frequency band covered by the entropy vector is from 48.83 kHz to 25 MHz.

The output layer presents a pattern of the pulse types. As the unknown pulses only needs to be distinguished as PD or not, the smallest output layer size can be one: the output '1' represents PDs and '0' denotes interferences including both the repetitive pulses and the random pulses.

(2) *Size of hidden layer:* The selection of hidden layer size is very important because too few neurons will result in under-fitting while too many nodes may cause over-fitting and huge calculation complexity [27]. A simple way to determine a suitable size is to try the possible numbers one by one [27]. For a selected ANN with size of 9-x-1, possible candidates of x are chosen from 2 to 8. Their performances are evaluated by mean-squared-errors (MSE) between the trained results and the desired responses. The average MSE is defined as

$$E_{\text{MSE}} = \frac{1}{j}\sum_{j\in C} e_j^2 = \frac{1}{j}\sum_{j\in C}(y_j - d_j)^2 \tag{3}$$

where C includes all the nodes at the output layer. The E_{MSE}s of ANN with different sizes of x are shown in Table 2. As the initial condition of ANN training is randomly selected, the MSE values are the averages of 20, 50 and 100 networks, respectively. All those networks are trained separately.

Table 2 Performance of NNs with different sizes of hidden layers

Scaled average E_{MSE}s	Size of hidden layer						
	2	3	4	5	6	7	8
20 NNs	1.038	1.029	0.94	1.057	1.142	1.101	1.117
50 NNs	1.099	1.045	1.038	1.14	1.137	1.196	1.121
100 NNs	3.399	3.301	3.198	3.346	3.41	3.792	3.842

The actual E_{MSE}s are the numbers in Table 2 multiplied by 1×10^{-7}.
The training will not stop until its training performance is smaller than 1×10^{-6} or the number of training epochs is larger than 1000.

Table 3 Mean MSEs of different activation functions

Functions	Sigmoid function	Log-sigmoid functions	Linear function
Scaled average E_{MSE}s	0.179	8.48	1.546

The actual E_{MSE}s are the numbers in Table 3 multiplied by 1×10^{-8}.
The training will not stop until its training performance is smaller than 1×10^{-7} or the number of training epochs is larger than 1000.

Table 4 Mean MSEs of different functions with different neural networks

Functions	QN	LM	LM with BR
Scaled average E_{MSE}s	0.338	1.951	0.179

The actual E_{MSE}s are the numbers in Table 4 multiplied by 1×10^{-8}.
The training will not stop until the training performance is smaller than 1×10^{-7} or the number of training epochs is larger than 1000.

It can be seen from the table that the average MSE distributions always reach their minimums when the size of hidden layer is 4. The hidden layer size is thus set to 4 and the structure of network is 9-4-1.

6.2 Selection of functions

When selecting the network functions, both the activation function for neurons and the training function for network should be considered.

(1) *Activation function:* For neurons, three kinds of transfer functions: sigmoid function, log-sigmoid function and linear function are often used. The training performance of 50 ANNs with these different activation functions are shown in Table 3. These networks have a same size of 9-4-1. According to the average E_{MSE}s, the sigmoid function which increases monotonically and exhibits smoothness is adopted as the activation function [27].

(2) *Training function:* For a whole network, the training function indicates the ways that ANN updates the weights and bias. Three commonly used algorithms are discussed here: the quasi-Newton (QN) optimisation, the Levenberg–Marquardt (LM) algorithm and the Bayesian regulation (BR). The mean values of the MSEs of 50 ANNs with same architecture are calculated and shown in Table 4.

Obviously, the networks trained with LM algorithm and BR can generate smaller errors than the others. Therefore the LM algorithm with BR is selected to produce a better recognition and classification result.

7 PD recognition results and comparisons

In this section, PD recognition is performed using the proposed system. To illustrate its effectiveness, the proposed method is compared with the other two methods: the energy distribution with ANN and the wavelet-entropy with similarity comparison.

7.1 Recognition results of the proposed method

Six groups of PD and noise pulses are recognised and their results are listed in Table 5. Each group contains all the PD and noisy pulses occurring in one measuring cycle (20 ms). The PD signals were collected in laboratory tests and noisy pulses were collected in

Table 5 Recognition results of proposed method

Group no.	Noise type	Real PDs	Recognition		Misjudgments	
1	Re.	20	16	(80%)	0	(0%)
2	Ra. + Re.	23	19	(82.61%)	0	(0%)
3	Ra. + Re.	19	15	(78.95%)	0	(0%)
4	Ra. + Re.	22	20	(90.91%)	0	(0%)
5	Ra. + Re.	31	26	(83.87%)	3	(9.68%)
6	Ra.	36	35	(97.22%)	1	(2.78%)
total		151	131	(86.75%)	4	(2.65%)

field tests. These signals were added directly to simulate the original noisy signals.

Here, 'Ra' and 'Re' stand for random pulse and repetitive pulse, respectively. The results in Table 5 suggest two conclusions. First, the recognition rates are very good although there are some fluctuations under different conditions. The lowest recognition rate is 78.95% of group 3, and the highest is 97.22% of group 6. The average recognition rate of six groups is 86.75% which should be enough for practical use in most cases. Second, misjudgments are rare. Four out of six groups have no misjudgments. Only four noise pulses in total are misjudged as PDs, and the total misjudgment rate is 2.65%.

7.2 Comparison with wavelet-energy-based method

Wavelet energy was proved to be effective in representing and recognising PDs by many research works. Among them, the method proposed by Zhou *et al.* [29] was very popular. That method describes the wavelet energy distributions of single PD pulses. Such distribution consists of two parts: energy in approximations *Ca* and energy in details *Cd*. The *Ca* and *Cd* are both vectors with a same size which is actually the number of decomposition level. By analysing the *Ca* and *Cd* of real PD pulses, the noisy pulses with different energy distributions can be rejected and the noises contained in PD pulses can also be removed. This method reveals good de-noising results. To provide a further illustration of wavelet-entropy's effectiveness, the wavelet energy feature is adopted to de-noise the signals in Table 5 as a comparison. To compare with the previous method, the detail coefficients are considered here. The energy ratio vectors defined in Appendix and the BP network introduced in Section 6 are employed. The PD extraction results are listed in Table 6.

Comparing with the proposed method, the energy-based one shows a less effective performance. Its average recognition rate is only 60.27% which is about one third lower than that of the entropy-based method. However, the energy-based method produces less misjudgments: only one pulse is misjudged and the total misjudgment rate is 0.66%. Although a higher misjudgment rate (2.65%) of the proposed method is found in Table 5, such rate still remains at a very low level and is acceptable in practical applications.

7.3 Comparison with the similarity-comparing method

Besides ANN, it is also possible to reject impulsive noises by using some simple rules. As the wavelet-entropy distributions of the pulses from a same type are similar, comparing the similarities between the unknown pulse and the known types is a possible way in pulse classification. Yang *et al.* calculated similarities between the wavelet distributions of different PD pulses to recognise their sources [30]. In their research work, two UHF sensors were used to capture the PD signals within a transformer tank. There were three PD sources inside the tank, and the two sensors that were located on different panels measured those signals at the same time. The wavelet distributions of each pulse from different sensors was compared. The ones with larger similarities (>0.4 in [30]) are considered to be from a same source. This method was demonstrated to have good performance in PD source recognition.

Table 6 Recognition results of the wavelet energy based method

Group no.	Real PDs	Recognition		Misjudgments	
1	20	15	(75%)	0	(0%)
2	23	14	(60.87%)	0	(0%)
3	19	7	(36.84%)	0	(0%)
4	22	13	(59.09%)	0	(0%)
5	31	13	(41.94%)	0	(0%)
6	36	29	(80.56%)	1	(2.78%)
total	151	91	(60.27%)	1	(0.66%)

Table 7 Recognition results of similarity-comparing method

Group no.	Real PDs	Recognition		Misjudgments	
1	20	18	(90%)	2	(10%)
2	23	20	(86.96%)	0	(0%)
3	19	17	(89.47%)	1	(5.26%)
4	22	19	(86.36%)	1	(9.09%)
5	31	19	(61.29%)	3	(9.68%)
6	36	27	(75%)	2	(5.56%)
total	151	120	(79.47%)	9	(5.96%)

The performance of the similarity-based method is compared with the proposed method to illustrate the proper use of ANN. The signals in Table 5 are also processed by the similarity-comparing method. The procedure is as follows. First of all, a standard data base is found. The average entropy ratio vectors ρ_{Ek} of 20 pulses from each type are taken as the standard distributions. After that, the similarities between the entropy ratio vector ρ_{Eu} of any unknown pulse and all three standard entropy ratio vectors ρ_{Ek} are calculated. Here, k stands for the type of pulses such as 'PD', 'Re' and 'Ra'. The largest similarity Sim_k corresponds to the k type of pulses that the unknown pulse can be recognised as with highest confidence. The similarity Sim_k is calculated as (4)

$$\mathrm{Sim}_k = \left\langle \frac{\rho_{Eu}}{||\rho_{Eu}||} \cdot \frac{\rho_{Ek}}{||\rho_{Ek}||} \right\rangle \tag{4}$$

Since $||\rho_E|| = 1$, the formula in (4) can be rewritten to be

$$\mathrm{Sim}_k = \langle \rho_{Eu} \cdot \rho_{Ek} \rangle \tag{5}$$

The PD extracting results with similarity-comparing method are illustrated in Table 7. By simply comparing the similarities between the unknown pulse and the known features, this method requires less calculations and simpler judgments. However, the recognition rates vary a lot. The lowest recognition rate (group 5, 61.29%) is about one third less than the highest one (group 1, 90%). Furthermore, the similarity-comparing method produces a lower total recognition rate of 79.47%, and a higher total misjudgment rate of 5.96% than the proposed ANN-entropy-based method.

8 Conclusions

In this paper, the combination of the wavelet-entropy and neural network is used to solve the problem of impulsive noise reduction of TEV-based PD measurement. Based on the studies of pulse properties and comparisons with energy distribution, the wavelet-entropy is proved to be effective in characterising PD and noise pulses and reducing the feature dimension. Furthermore, with careful selection of the parameters, a neural network which is suitable for entropy-based PD recognition was constructed and trained. The test results demonstrated that the proposed wavelet-entropy-based PD recognition with neural network can effectively recognise most PD pulses in different tests and show a better performance than previous methods including the energy-based method and similarity-comparing method. So far, the

proposed system is applied to recognise the needle-to-plane PD pulses that are generated in experimental environment. Its applications on different types of PDs from practical environment will be studied in future works.

9 Acknowledgment

This research is funded by the Republic of Singapore's National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. BEARS has been established by the University of California, Berkeley as a centre for intellectual excellence in research and education in Singapore.

10 References

1 IEC 60270: 'High-voltage test techniques – partial discharge measurements', 2001
2 Judd, M.D., Farish, O., Pearson, J.S., Hampton, B.F.: 'Dielectric windows for UHF partial discharge detection', *IEEE Trans. Dielectr. Electr. Insul.*, 2001, **8**, (6), pp. 953–958
3 Lazarevich, A.K.: 'Partial discharge detection and localization in high voltage transformers using an optical acoustic sensor'. Master of Science, Virginia Polytechnic Institute and State University, 2003
4 Brown, P.: 'Nonintrusive partial discharge measurements on high voltage switchgear'. IEE Colloquium on Monitors and Condition Assessment Equipment, Leatherhead, UK, 1996, pp. 10/1–10/5
5 Li, Y., Wang, Y., Lu, G., Wang, J., Xiong, J.: 'Simulation of transient earth voltages aroused by partial discharge in switchgears'. Int. Conf. on High Voltage Engineering and Application (ICHVE), New Orleans, USA, 2010, pp. 309–312
6 Bartnikas, R.: 'Partial discharges, their mechanism, detection and measurement', *IEEE Trans. Dielectr. Electr. Insul.*, 2002, **9**, (5), pp. 763–808
7 Okamoto, T., Tanaka, T.: 'Partial discharge pattern recognition for three kinds of model electrodes with a neural network', *IEE Proc. Sci., Meas. Technol.*, 1995, **142**, (1), pp. 75–84
8 Contin, A., Cavallini, A., Montanari, G.C., Pasini, G., Puletti, F.: 'Digital detection and fuzzy classification of partial discharge signals', *IEEE Trans. Dielectr. Electr. Insul.*, 2002, **9**, (3), pp. 335–348
9 Luo, G.: 'Transient earth voltage (TEV) based partial discharge detection and analysis'. Doctor of Philosophy, Nanyang Technological University, 2013
10 Florkowski, M., Florkowska, B.: 'Phase-resolved rise-time-based discrimination of partial discharges', *IET Gener. Transm. Distrib.*, 2009, **3**, (1), pp. 115–124
11 Dey, D., Chatterjee, B., Chakravorti, S., Munshi, S.: 'Cross-wavelet transform as a new paradigm for feature extraction from noisy partial discharge pulses', *IEEE Trans. Dielectr. Electr. Insul.*, 2010, **17**, (1), pp. 157–166
12 Jian, L., Tianyan, J., Harrison, R.F., Grzybowski, S.: 'Recognition of ultra high frequency partial discharge signals using multi-scale features', *IEEE Trans. Dielectr. Electr. Insul.*, 2012, **19**, (4), pp. 1412–1420
13 Chang, C.S., Jin, J., Chang, C., Hoshino, T., Hanai, M., Kobayashi, N.: 'Separation of corona using wavelet packet transform and neural network for detection of partial discharge in gas-insulated substations', *IEEE Trans. Power Deliv.*, 2005, **20**, (2), pp. 1363–1369
14 Evagorou, D., Kyprianou, A., Lewin, P.L., *et al.*: 'Feature extraction of partial discharge signals using the wavelet packet transform and classification with a probabilistic neural network', *IET Sci. Meas. Technol.*, 2010, **4**, (3), pp. 177–192
15 Luo, G., Zhang, D., Koh, Y., Ng, K., Leong, W.: 'Time-frequency entropy-based partial-discharge extraction for nonintrusive measurement', *IEEE Trans. Power Deliv.*, 2012, **27**, (4), pp. 1919–1927
16 Zhang, Z.S., Xiao, D.M., Li, Y.: 'Rogowski air coil sensor technique for on-line partial discharge measurement of power cables', *IET Sci. Meas. Technol.*, 2009, **3**, (3), pp. 187–196
17 Zhou, C., Song, X., Michel, M., Hepburn, D.M.: 'On-line partial discharge monitoring in medium voltage underground cables', *IET Sci. Meas. Technol.*, 2009, **3**, (5), pp. 354–363
18 Thanasopoulos, I., Avaritsiotis, J.: 'Wavelet analysis of short range seismic signals for accurate time of arrival estimation in dispersive environments', *IET Sci. Meas. Technol.*, 2011, **5**, (4), pp. 125–133
19 Donoho, D.L., Johnstone, I.M.: 'Ideal denoising in an orthonormal basis chosen from a library of bases', *Comptes Rendus De L Academie Des Sciences Serie I-Mathematique*, 1994, **319**, (12), pp. 1317–1322
20 Liu, R., Sun, X., Li, Z.: 'On the application of entropy in excitation control'. 2004 Int. Conf. on Power System Technology (PowerCon), Singapore, 2004, vol. 1, pp. 952–956
21 He, Z., Fu, L., Lin, S., Bo, Z.: 'Fault detection and classification in EHV transmission line based on wavelet singular entropy', *IEEE Trans. Power Deliv.*, 2010, **25**, (4), pp. 2156–2163
22 Dubey, R., Samantaray, S.R.: 'Wavelet singular entropy-based symmetrical fault-detection and out-of-step protection during power swing', *IET Gener. Transm. Distrib.*, 2013, **7**, (10), pp. 1123–1134
23 Verdú, S., McLaughlin, S.W., IEEE Information Theory Society: 'Fifty years of shannon theory', in Verdú, S., McLaughlin, S.W. (Eds.): 'Information theory: 50 years of discovery' (IEEE Press, 2000, 1st edn.), pp. 13–34
24 Dimopoulos, M.G., Spyronasios, A.D., Papakostas, D.K., Hatzopoulos, A.A.: 'Wavelet energy-based testing using supply current measurements', *IET Sci. Meas. Technol.*, 2010, **4**, (2), pp. 76–85
25 Mohanty, S., Ghosh, S.: 'Artificial neural networks modelling of breakdown voltage of solid insulating materials in the presence of void', *IET Sci. Meas. Technol.*, 2010, **4**, (5), pp. 278–288
26 Luo, G., Zhang, D.: 'Application of wavelet transform to study partial discharge in XLPE sample'. Australasian Universities Power Engineering Conf., AUPEC, Adelaide, Australia, 2009, pp. 1–6
27 Heaton, J.: 'Understanding backpropagation', in Smith, K. (Ed.): 'Introduction to neural networks with Java' (Heaton Research, 2005, 2nd edn.), pp. 125–154
28 Haykin, S.S.: 'Multilayer perceptrons', in Horton, M. (Ed.): 'Neural networks: a comprehensive foundation' (Prentice Hall, 1999, 2nd edn.), pp. 138–235
29 Zhou, X., Zhou, C., Kemp, I.J.: 'An improved methodology for application of wavelet transform to partial discharge measurement denoising', *IEEE Trans. Dielectr. Electr. Insul.*, 2005, **12**, (3), pp. 586–594
30 Yang, L., Judd, M.D.: 'Recognising multiple partial discharge sources in power transformers by wavelet analysis of UHF signals', *IEE Proc. Sci. Meas. Technol.*, 2003, **150**, (3), pp. 119–127

11 Appendix

Similar to the entropy ratio, the 'energy ratio' is adopted to normalise the wavelet energy distributions into the interval [0 1]. It is calculated as

$$\rho_{\text{engi}} = \frac{E_i}{||E||} \tag{6}$$

Here, the energy of each level equals to the Euclidean norm of wavelet coefficients, that is, $E_i = ||C_i|| = \sqrt{\sum_{j=1}^{J} c_{i,j}^2}$, where $c_{i,j}$ is the jth coefficient on the ith level.