Meas. Sci. Technol. **19** (2008) 055707 (10pp) doi:10.1088/0957-0233/19/5/055707

Improving detection sensitivity for partial discharge monitoring of high voltage equipment

L Hao, P L Lewin and S G Swingler

The Tony Davies High Voltage Laboratory, Electrical Power Engineering Group, School of Electronics and Computer Science, University of Southampton, Highfield, Southampton SO17 1BJ, UK

E-mail: haoliwei@soton.ac.uk

Received 20 September 2007, in final form 13 March 2008
Published 21 April 2008
Online at stacks.iop.org/MST/19/055707

Abstract

Partial discharge (PD) measurements are an important technique for assessing the health of power apparatus. Previous published research by the authors has shown that an electro-optic system can be used for PD measurement of oil-filled power transformers. A PD signal generated within an oil-filled power transformer may reach a winding and then travel along the winding to the bushing core bar. The bushing, acting like a capacitor, can transfer the high frequency components of the partial discharge signal to its earthed tap point. Therefore, an effective PD current measurement can be implemented at the bushing tap by using a radio frequency current transducer around the bushing-tap earth connection. In addition, the use of an optical transmission technique not only improves the electrical noise immunity and provides the possibility of remote measurement but also realizes electrical isolation and enhances safety for operators. However, the bushing core bar can act as an aerial and in addition noise induced by the electro-optic modulation system may influence overall measurement sensitivity. This paper reports on a machine learning technique, namely the use of a support vector machine (SVM), to improve the detection sensitivity of the system. Comparison between the signal extraction performances of a passive hardware filter and the SVM technique has been assessed. The results obtained from the laboratory-based experiment have been analysed and indicate that the SVM approach provides better performance than the passive hardware filter and it can reliably detect discharge signals with apparent charge greater than 30 pC.

Keywords: partial discharge, power transformer, radio frequency current transducer, electro-optic modulation, wavelet analysis, data mining, support vector machine

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Field measurements of partial discharge (PD) are becoming an important tool for quality control and risk assessment of operational electrical insulation systems. With the development of sensors and measurement techniques, condition-based maintenance is attracting more and more interest. Research into continuous on-line condition monitoring has considered problems such as remote measurement and transmission, noise immunity, maintenance convenience and operator safety. The use of optical transmission techniques therefore has clear advantages. Work at Southampton has concentrated on implementing an electro-optic modulator (EOM) system to PD continuous on-line monitoring of cables [1]. Preliminary research [2] has indicated that the measurement sensitivity may be influenced by optical noise. Research to date has considered using additional hardware filters, active amplifiers and radio frequency transformers to improve the sensitivity of the EOM system [2, 3].

The characteristic output signals from the radio frequency current transducer (RFCT) for PD detection within a power

transformer are different from the signals from the capacitive coupler (CC) for cable PD detection. The RFCT current measurement has a lower bandwidth (up to 200 MHz) than a CC voltage measurement (up to 400 MHz) [4] and is comparably weaker in terms of signal energy, which can in turn reduce the measurement sensitivity. Therefore, the useful PD signals may be buried by background noise due to a low signal-to-noise ratio (SNR) and approaches proposed in [2] may not be applicable. In particular, the use of an active amplifier requires an extra power supply which causes inconvenience in application and maintenance and may mean that the measurement system is no longer 'passive' at the bushing tap point.

This paper, taking the bushing-tap RFCT system for PD measurement and the EOM system for signal transmission as an example, compares the signal extraction performances of a passive band-stop filter and a machine learning technique based on different feature parameters. Analysis of the 'filtered' results indicate that it is possible using the support vector machine (SVM) approach to reliably detect discharge signals with an apparent charge greater than 30 pC.

2. Detection methods

In this research, the aim has been to develop a system capable of monitoring PD activity of electrical transmission assets over significant periods of time. Therefore, the developed system should not require additional power sources at the measurement points. The proposed PD measurement system comprises a PD detection system and an EOM-based transmission system.

2.1. The PD measurement system

A transformer bushing-tap system for PD measurement has been developed and is used as the experimental model, as shown in figure 1. The 60 kV bushing used in this experiment has a capacitance of 235 pC. A connection from the tap point to ground passed through the RFCT three times. The RFCT has a useful measurement bandwidth of up to 200 MHz. Three turns on the primary side is used to improve the overall measurement gain of the sensor within its bandwidth. This measurement system has been used on-site for on-line PD monitoring of power transformers [5]. The RFCT output signal is fed into the EOM.

A Robinson Type 5 Model 700 conventional PD electrical detector with 40–300 kHz bandpass frequency was used to detect and quantify the apparent charge of simulated PD event. A digital signal oscilloscope (DSO), LeCroy LC684DXL with a bandwidth of 1.5 GHz, was also used to display, store and analyse the signals from the Robinson detector, RFCT and photoreceiver. The data were also displayed and saved in a personal computer via a GPIB card for further processing.

Simulated PD pulses were generated by a LDIC LDC-5/RUF UHF calibrator which working like a pulse generator can generate two UHF calibration pulses synchronized with the zero-crossing points of the 50 Hz AC power supply. A typical pulse is shown in figure 2, the pulse has a rise time of

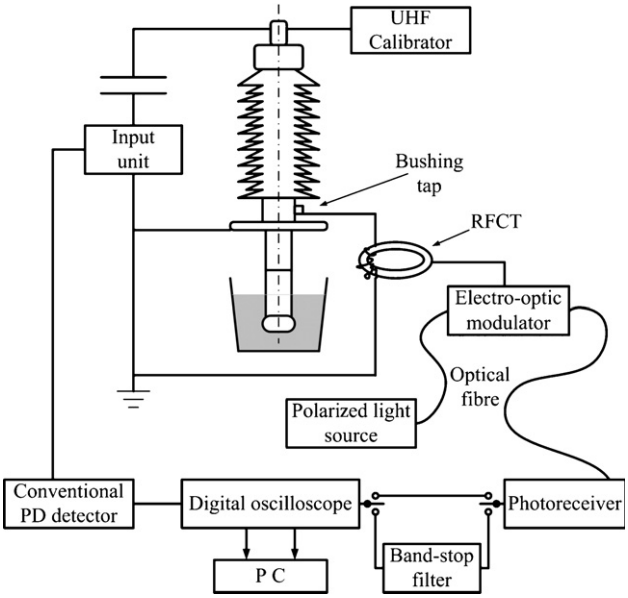

Figure 1. Schematic diagram of the experimental circuit.

1 ns and its bandwidth is greatly in excess of the 200 MHz bandwidth of the RFCT sensor. The pulses generated by the UHF calibrator were injected into the bushing core, which were then simultaneously measured by both the Robinson detector and the RFCT/photoreceiver.

An Agilent 4395A Network/Spectrum/Impedance Analyser with a 10 Hz to 500 MHz measurement bandwidth is used to measure the frequency response of the electro-optic system and the band-stop filter.

2.2. The electro-optic modulation technique

The PD on-line monitoring technique is based on the use of a lithium niobate ($LiNbO_3$) electro-optic modulator. The measurement mechanism uses the measured PD signal and applies it across an optical-fibre-coupled $LiNbO_3$ waveguide modulator, as shown in figure 3, which modulates the intensity of the transmitted laser light as an approximately linear function of the voltage applied across it, as shown in figure 4. The optical network supplies the polarized laser light via polarization maintaining optical fibre to the $LiNbO_3$ modulator input and monitors the optical output from the modulator using an optical receiver. The EOM is compact and passive requiring no additional power to operate. The laser source, which is controlled by a temperature and current laser diode controller, has a wavelength of 1550 nm and maximal power of 10 mW. A polarization tuner was used to ensure that the input light for the modulator was linearly polarized. The optical receiver has a bandwidth of 1 GHz. The schematic diagram of the EOM transmission system is shown in figure 5.

As shown in figure 6(*a*), the relationship between a single pulse from an HP 8082A pulse generator applied across the EOM and the resultant output from the optical receiver is represented. These two signals have similar rise and fall times. The slight delay between two signals is caused by

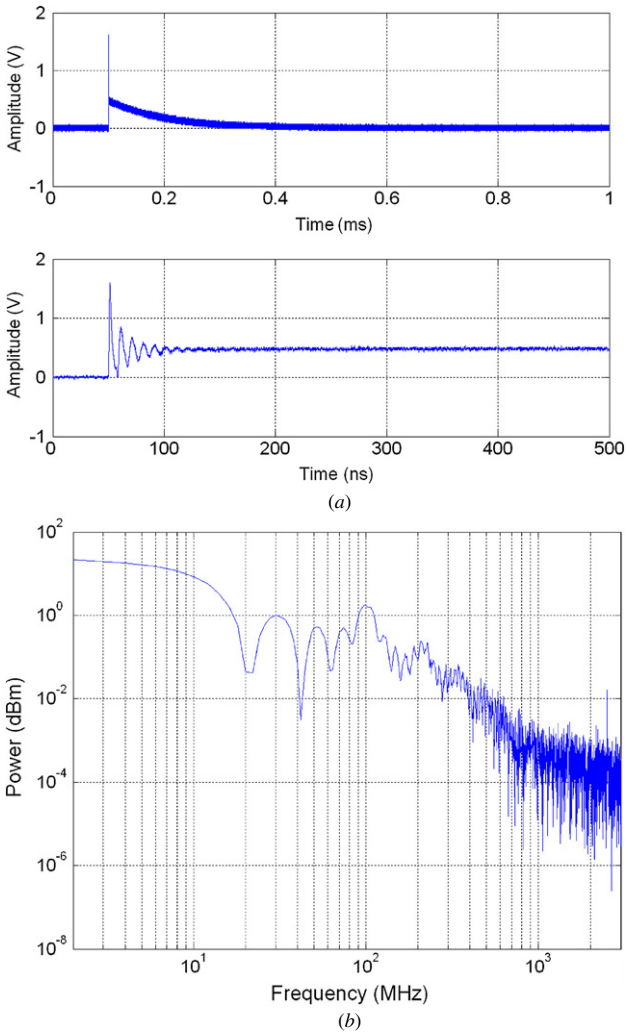

Figure 2. Characteristic of the signal from the UHF calibrator: (*a*) output signal of the UHF calibrator and (*b*) calibration pulse frequency spectrum.

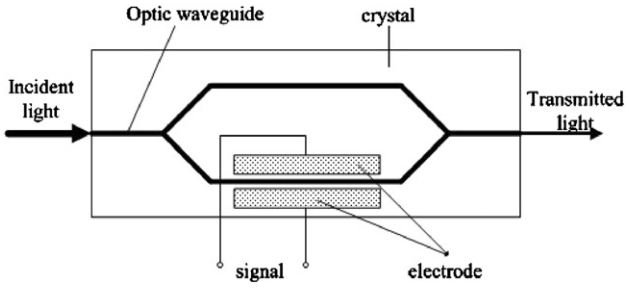

Figure 3. Configuration of the $LiNbO_3$ waveguide modulator.

the different transmission path of the measured pulses and the delay from the EOM system. The overall frequency response of the system is measured by the Agilent 4395A Network Analyser. The results are shown in figure 6. It can be initially assumed that the method of transmission will not significantly alter the detected PD signal information.

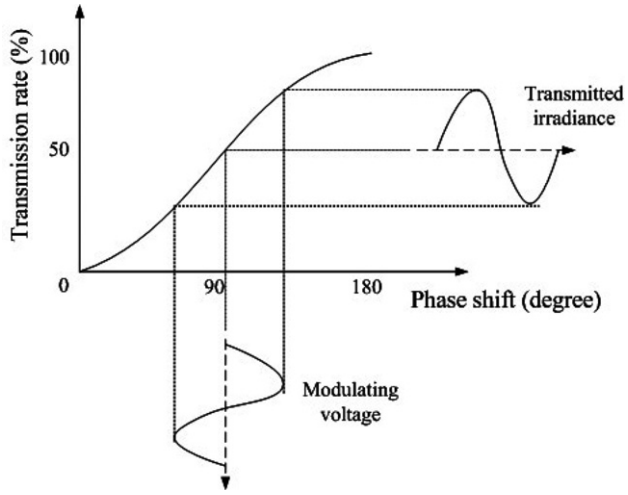

Figure 4. Principle of modulation.

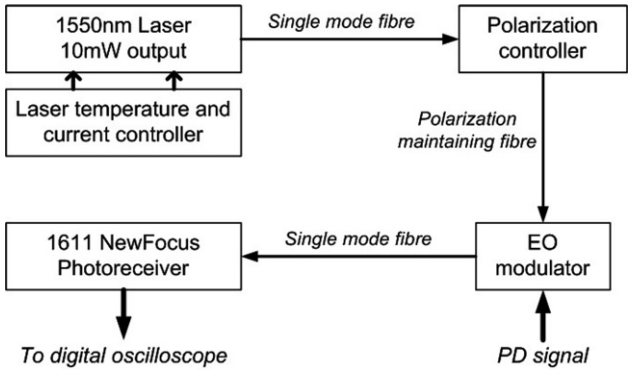

Figure 5. EOM-based transmission of PD signal data.

3. Wavelet decomposition filtering

The combination of time domain and frequency domain information, using the wavelet transform, provides a suitable tool for PD signal analysis. Wavelet decomposition works as a pair of high-pass and low-pass filters which decompose the original signal (S) into a series of detail coefficients (cD) and approximate coefficients (cA), as shown in figure 7. The detail and approximate coefficients represent the components after the high-pass filter and the low-pass filter, respectively. The decomposition process is an iterative process and can be described in a tree structure, as shown in figure 8. The coefficients of each level have half of the bandwidth and length than the level from which they were decomposed.

Wavelet filtering techniques have been successfully applied to PD data from high voltage cables [6, 7] as well as high voltage transformers [8] and GIS [9].

4. Support vector machine identification

As a pattern recognition tool, the support vector machine (SVM) is based on statistical learning theory which has been researched since 1960s. The SVM was first proposed by

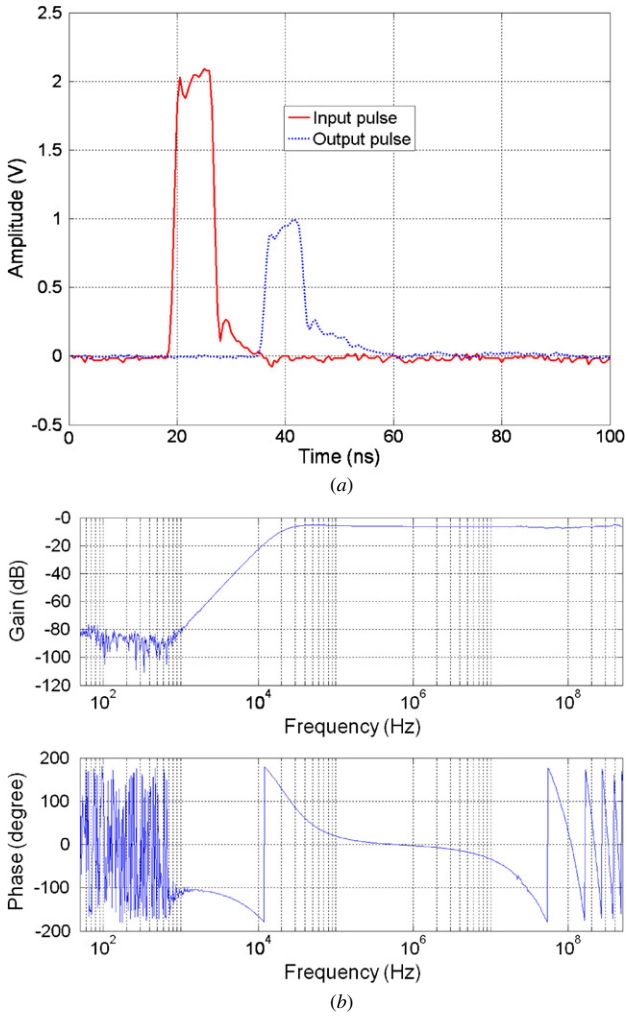

Figure 6. Performance of the EOM system: (*a*) input signal of the EOM and output signal from the photoreceiver and (*b*) overall frequency response of the EOM and photoreceiver.

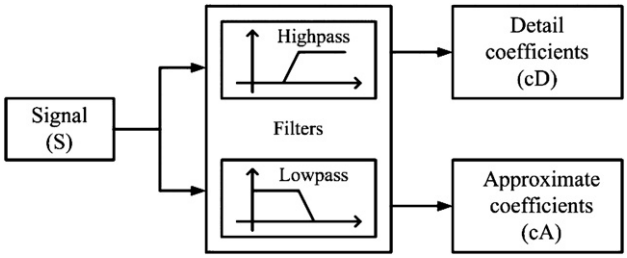

Figure 7. Concept of the wavelet decomposition.

V N Vapnik in 1995 [10]. This learning machine uses a central concept of SVMs–kernels for a number of learning tasks [11]. Based on kernel methods, SVMs can be adapted to different tasks and domains by the choice of the kernel function and base algorithm [11]. They represent great advantages in small sample quantity, nonlinear and high-dimensionality pattern recognition problems. Successful applications have demonstrated that SVMs can perform as well or better

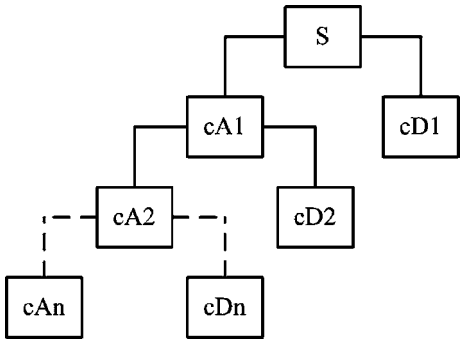

Figure 8. Structure of the decomposition process.

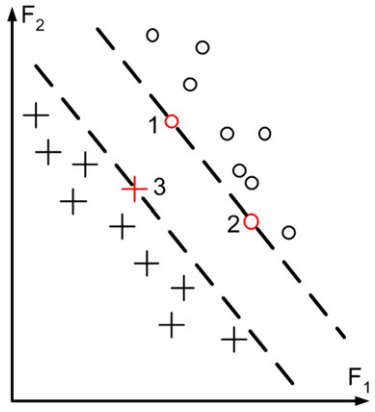

Figure 9. Support vectors.

than neural networks in a wide variety of fields, including engineering, information retrieval and bioinformatics [11, 12]. Relevant research on the application of SVMs to high voltage engineering can be found in [13, 14].

The SVM is a method for finding functions from a set of labelled training data. The function can be either a classification function or a regression function. SVM earns its name by constructing the solution to the learning problem in terms of a subset of the training data; this subset is referred to as the support vectors (SVs) [15]. For example, figure 9 shows two linearly separable clusters of data that are described by using two features (F_1 and F_2). Three SVs (numbered and illustrated in red) that lie on the critical decision boundary of the two groups of data can be used to determine the hyperplane and margin of the two clusters. Several distinct concepts contribute to the theory behind SVM. Linear learning machines, the technique of kernel functions and feature spaces, the use of a sophisticated learning bias derived from statistical learning theory and optimization theory are all combined to create the SVM learning system [11].

Based on structural risk minimization (SRM) principles, the SVM is equipped with a greater capability of generalization than traditional neural network approaches. The kernel mapping provides a unifying framework for most of the commonly employed model architectures, enabling comparisons to be performed. In classification problems generalization control is obtained by maximizing the margin,

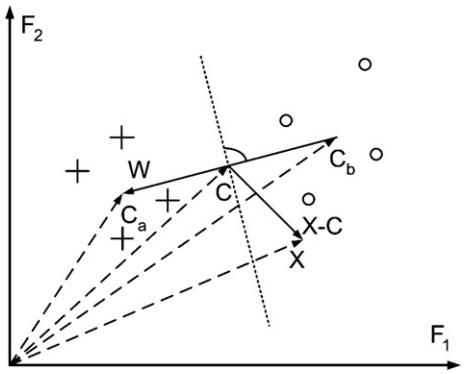

Figure 10. Similarity assessment by dot product.

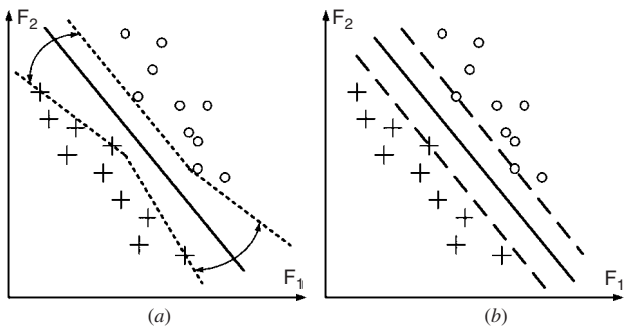

Figure 11. The hyperplanes: (*a*) a range of hyperplanes and (*b*) the unique optimal hyperplane and its margin.

which corresponds to minimization of the weight vector in a canonical framework. The solution is obtained as a set of support vectors that can be sparse. These lie on the boundary and as such summarize the information required to separate the data [13, 15].

4.1. Similarity assessment

Figure 10 represents a simple two-dimensional pattern recognition example where each dimension (F_1, F_2) is a measure of a specific feature of the input data. The crosses and circles represent two different classes (class a and class b), respectively. In order to evaluate the attribution of a new point X to these two clusters, a similarity assessment algorithm needs to be addressed. An obvious criterion is by 'measuring' the distances between the point X and both clusters.

As shown in figure 10, the centres of the two clusters, C_a and C_b, can be computed. The attribution of the assigned point X can be obtained by calculating its distances to C_a and C_b. Assume that the pattern X belongs to the circle class (class b), it means

$$\|X - C_b\| < \|X - C_a\|. \tag{1}$$

This is corresponding to the sign of the dot product between $\mathbf{X} - \mathbf{C}$ (where $\mathbf{C} = (\mathbf{C}_a + \mathbf{C}_b)/2$) and $\mathbf{W}$ ($\mathbf{W} = \mathbf{C}_a - \mathbf{C}_b$), which will change sign as the enclosed angle passes through $\pi/2$. Therefore, the decision function can be written as

$$y = \mathrm{sgn}((\mathbf{X} - \mathbf{C}) \cdot \mathbf{W}). \tag{2}$$

This means that the corresponding decision boundary is a hyperplane (dotted line in figure 10) orthogonal to $\mathbf{W}$.

4.2. Hyperplane classifiers

Figure 11 shows a linearly separable two-dimensional condition. The circles and crosses denote the training data of two classes, respectively. The plane that can separate the two clusters without error is defined as the hyperplane. It is obvious that no hyperplane is unique. However, among all these separating hyperplanes there exists a unique optimal hyperplane which is distinguished by the maximum margin of separation between any training point and the hyperplane itself. It is an optimal solution of [11]:

$$\max \min\{\|X - X_i\| | X \in \mathcal{H}, \mathbf{W} \cdot \mathbf{X} + b = 0, i = 1, \ldots, n, \mathbf{W} \in \mathcal{H}, b \in \mathbf{R}\} \tag{3}$$

where $\min\{\cdot\}$ represents the distance between training points and hyperplanes and the optimal solution is where this distance is a maximum in (3). X exists in the feature space, namely Hilbert space ($\mathcal{H}$). $\mathbf{W}$ and b are parameters of the optimization problem, n is an integer representing the point number and $\mathbf{R}$ represents the real domain. To construct the optimal hyperplane, following (3), equation (4) needs to be solved:

$$\min \tau(\mathbf{w}) = \tfrac{1}{2}\|\mathbf{w}\|^2 \tag{4}$$

$$\text{subject to } y_i[(\mathbf{X}_i \cdot \mathbf{w}) + b] - 1 \geqslant 0, \qquad i = 1, \ldots, n. \tag{5}$$

An example of the optimal hyperplane and its margin is shown as the solid line and dashed line in Figure 11(*b*), respectively. The function τ and inequation in (4) are called the objective function and inequality constraints, respectively. The constrained optimization problem has now been formed and expressed by (4). By introducing Lagrange multipliers $\alpha_i \geqslant 0$ and a Lagrangian

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i [y_i[(\mathbf{X_i} \cdot \mathbf{w}) + b] - 1], \tag{6}$$

this kind of 'saddle' problem can be dealt with. The Lagrangian has to be minimized with respect to the primal variables $\mathbf{w}$ and b and maximized with respect to the dual variables α_i. Those support vectors lying on the margin satisfy the equality constraints as a consequence of Karush–Kuhn–Tucker (KKT) conditions [11].

In practice, a separating hyperplane may not exist, for example high noise levels could cause an overlap of the classes. By adding a slack variable $\xi_i \geqslant 0$, where $i = 1, \ldots, n$, the so-called soft margin hyperplane or generalized hyperplane can be achieved, which can minimize the empirical risk [11]:

$$y_i[(\mathbf{X_i} \cdot \mathbf{w}) + b] - 1 + \xi_i \geqslant 0, \qquad i = 1, \ldots, n. \tag{7}$$

In the simplest case, referred to as the C-SV classifier, this is done by solving (8), for some $C > 0$,

$$\min \tau(\mathbf{w}, \xi_i) = \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^{n} (\xi_i). \tag{8}$$

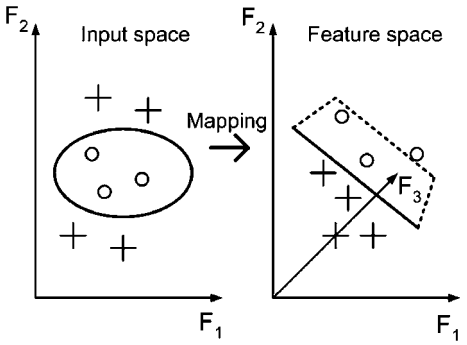

Figure 12. Mapping the training samples from input space to feature space.

Another possible realization of a soft margin variant of the optimal hyperplane uses the more natural ν-parameterization. In it, the parameter C is replaced by a parameter $\nu \in (0, 1]$ which can be shown to provide lower and upper bounds for the fraction of examples that will be SVs and those that will have non-zero slack variables. It uses a primal objective function with the error term $\left(\frac{1}{\nu n}\sum_i \xi_i\right) - \rho$ instead of $C\sum_i \xi_i$ and separation constraints that involve a margin parameter ρ:

$$y_i[(\mathbf{X}_i \cdot \mathbf{w}) + b] - \rho + \xi_i \geqslant 0, \qquad i = 1, \ldots, n. \tag{9}$$

4.3. Kernel methods

For nonlinearly separable data sets, the classification ability of the above optimal hyperplane is limited. To allow for a more general decision function, one alternative is to map the input vectors to a high-dimensional feature space by using a nonlinear mapping $\Phi(x) : \mathcal{X} \to \mathcal{H}$. A complex pattern classification problem in a higher dimensional space is more likely to be linearly separable than in lower dimensional space [11, 15]. Figure 12 shows the basic concept of mapping the samples into some other dot product space called the feature space. This is expressed as

$$f = \{\phi(x)|x \in \mathcal{X}\}, \qquad \mathcal{X} = \{x_1, x_2, \ldots, x_n\}$$
$$\to \Phi(x) = \{\phi_1(x), \phi_2(x), \ldots, \phi_n(x)\}. \tag{10}$$

The idea of the SVM is to map the training samples into a higher dimensional feature space via ϕ and construct a separating hyperplane with maximal margin there. In order to construct the mapping, a kernel function is used. The kernel function is defined as a way of computing the inner product $\langle\phi(x), \phi(x_i)\rangle$ in feature space directly as a function of the original inputs [11], that is

$$k(x, x_i) = \langle\phi(x), \phi(x_i)\rangle. \tag{11}$$

The application of various kernels to PD data have been assessed previously [13]. It has been found that the Gaussian radial basis function (Gaussian-RBF) kernel is the most effective for PD data, where the Gaussian-RBF kernel is defined as

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{12}$$

where $\sigma > 0$ is the kernel parameter controlling the flexibility of classifiers.

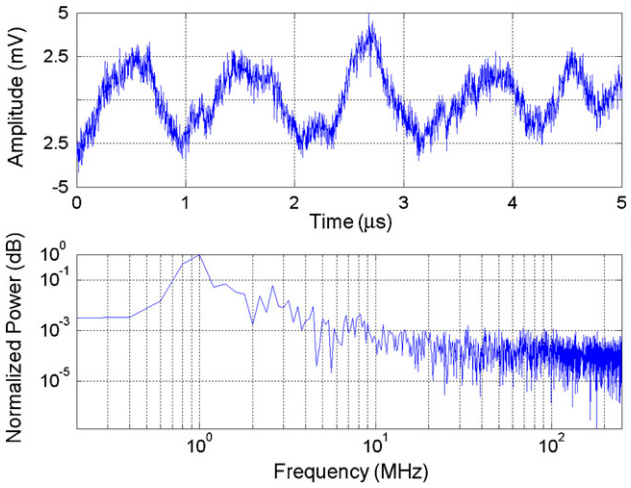

Figure 13. EOM system noise and its frequency content.

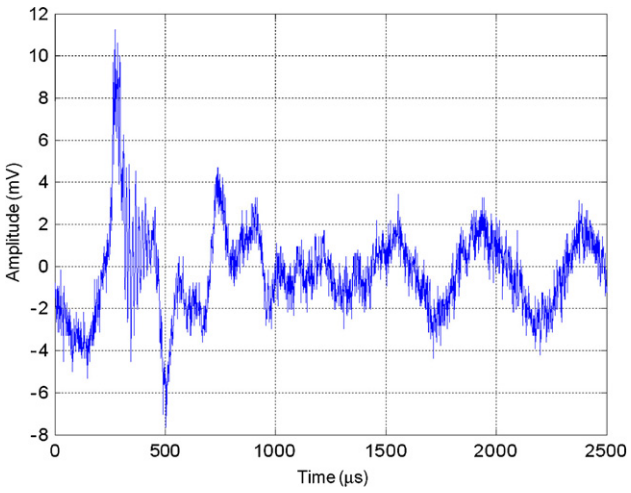

Figure 14. The minimum detectable pulse (160 pC) from the photoreceiver.

5. Experiment

5.1. Band-stop filter for denoising

For the laboratory-based experiment, the peak background noise level of the EOM-based measurement (figure 1) system is around 5 mV, with most significant frequency components concentrated under 1 MHz, as shown in figure 13.

Under these conditions, the minimum detectable PD level of the measurement from the photoreceiver without any filter or additional data processing is around 160 pC, as shown in figure 14.

The background noise from the EOM system is mainly due to the optical circuit itself and makes it possible to apply a static band-stop filter to eliminate the known frequency components of the noise. A passive filter with a band-stop frequency of 540 kHz to 1.6 MHz manufactured by Physical and Electronic Laboratories Ltd was used to depress the background noise. The frequency response of the filter has been measured by using the Agilent 4395A Network Analyser, as shown in

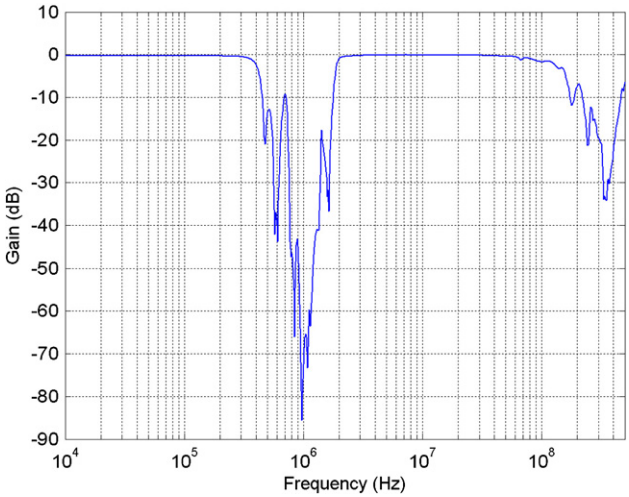

Figure 15. Frequency response of the band-stop filter.

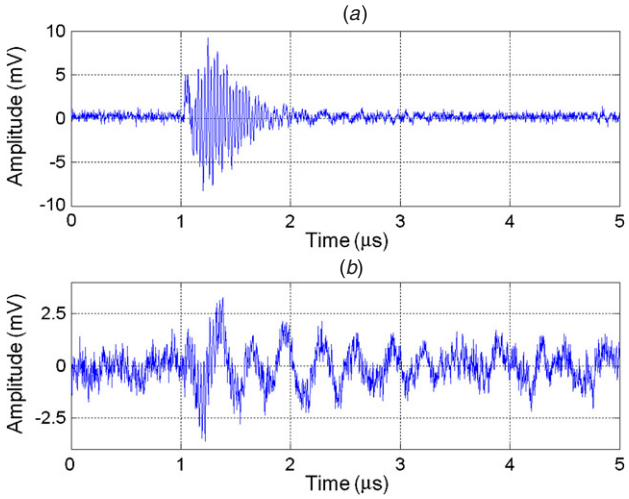

Figure 16. Filter result of a 50 pC level signal: (*a*) RFCT output and (*b*) photoreceiver output.

figure 15. The filter attenuates signals over the frequency range from 400 kHz to 2 MHz and above 80 MHz. The useful attenuation range (above 20 dB) is between 600 kHz and 1.8 MHz. By using the filter, the measurement sensitivity is improved to approximately a 50 pC level, as shown in figure 16.

5.2. Feature parameters for SVM discrimination

Due to the bandwidth of the measurement system, the sampling rate of the oscilloscope was set to 500 MSample s^{-1} for a period of 5 μs. Therefore, there are 2500 points in each acquisition. By segmenting the memory of the oscilloscope, 2800 individual pulses were recorded at each time. The data set which consists of 2800 pulses was stored in the memory of the DSO and also transferred to a personal computer via a GPIB card. Two sets of noise data and eight sets of PD data for 30 pC, 40 pC, 50 pC, 70 pC, 90 pC, 110 pC, 130 pC and

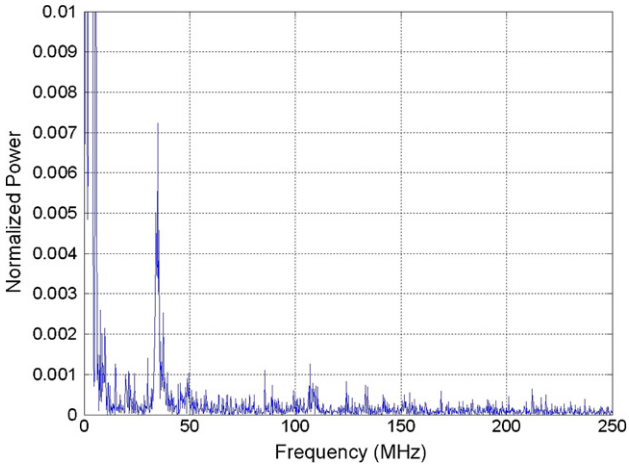

Figure 17. Frequency spectrum of a 160pC PD pulse.

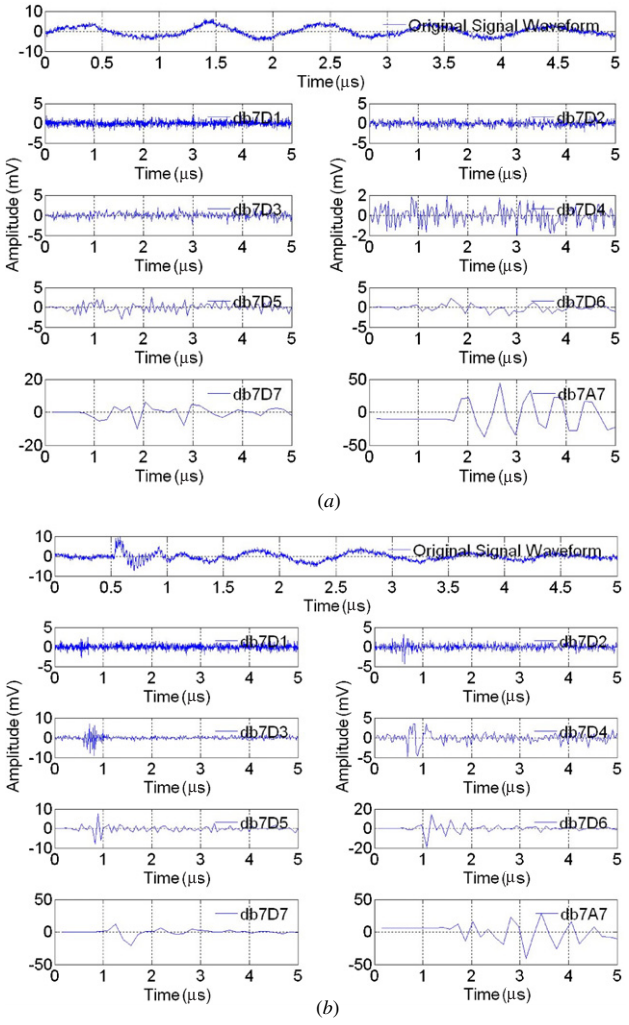

Figure 18. Wavelet decomposition coefficients: (*a*) noise signal and (*b*) 160 pC PD signal.

160 pC apparent charge levels were recorded for processing. As stated in the previous section, the minimum detectable PD level without filtering is 160 pC. Therefore, the SVM was

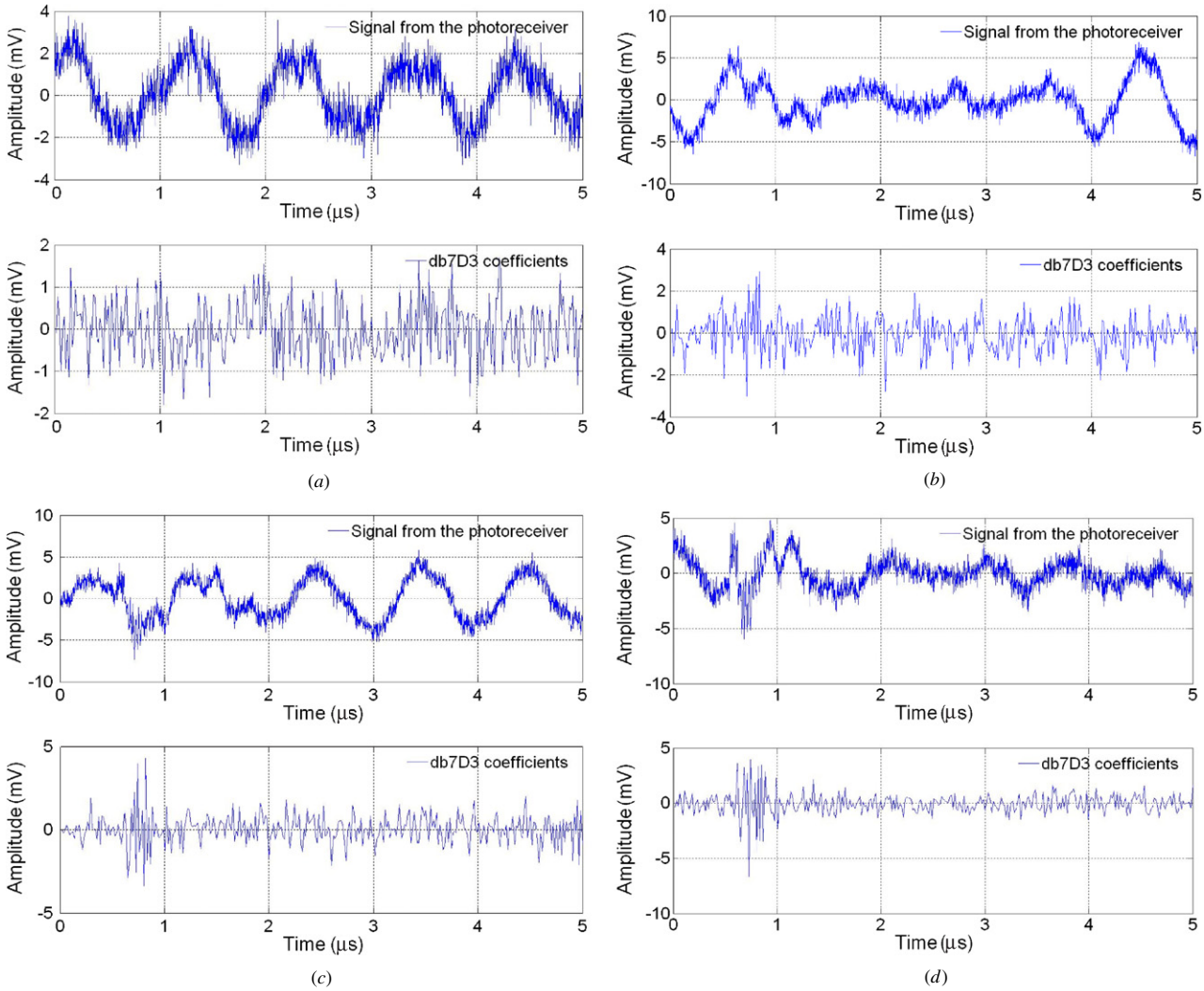

Figure 19. Original signal and Daubenchies 7 detail level 3 wavelet decomposition coefficients: (*a*) noise signal 2, (*b*) 30 pC PD signal, (*c*) 70 pC PD signal and (*d*) 130 pC PD signal.

trained with one set of noise data and 160 pC PD data and then tested using the other data sets.

The frequency spectrum and wavelet decomposition coefficients of the transmitted signal were chosen as possible feature parameters to characterize the PD signal and noise.

5.2.1. Frequency spectrum. As the time domain PD pulse and noise are not obviously separable, some analysis within the frequency domain is worth investigating. Figure 17 shows the frequency spectrum of 160 pC PD signal by using the fast Fourier transform (FFT).

The FFT results were taken as the input feature for the SVM. The training accuracy between 160 pC discharge data and noise is 98.1%. However, the identification results are not very satisfying, as shown in table 1. At the 130 pC level, the identification rate is no better than 66.5% (1861/2800). This is because the EOM system noise occupies a large number of low frequency components, which dominate the useful PD frequency components in this range and weaken the identification feature.

Table 1. Identification rates by using feature frequency spectrum.

Data sets	Frequency spectrum
Noise set 2	100% (2800/2800)
30 pC	14.4% (402/2800)
40 pC	27.4% (768/2800)
50 pC	34.1% (956/2800)
70 pC	37.0% (1037/2800)
90 pC	47.9% (1341/2800)
110 pC	54.4% (1523/2800)
130 pC	66.5% (1861/2800)

5.2.2. Wavelet decomposition coefficients. Wavelet analysis has been researched and applied to PD signal processing and noise reduction for many years. Some successful results have been achieved not only when applying to the simulated data but also the field data [7, 8, 16]. Research to date has

Table 2. Identification rates by using feature db7D3.

Data sets	db7D3
Noise set 2	100% (2800/2800)
30 pC	99.5% (2785/2800)
40 pC	99.8% (2795/2800)
50 pC	100% (2800/2800)
70 pC	100% (2800/2800)
90 pC	100% (2800/2800)
110 pC	100% (2800/2800)
130 pC	100% (2800/2800)

proposed that the wavelet Daubechies family order 7 (db7) may give good results in PD signal recovery and denoising [17]. Therefore in this application, the db7 was chosen as the mother wavelet. Figure 18 shows the examples of the decomposition coefficients at seven different levels for noise and 160 pC data, respectively.

As shown in figure 18, the detail coefficients at level 3 (referred to as db7D3) represent a better discrimination between the noise and useful PD signal from a visual point of view. The original signal from the photoreceiver and the db7D3 of noise set 2, 30 pC, 40 pC, 50 pC, 70 pC, 90 pC, 110 pC and 130 pC are collected for testing the performance of the SVM. db7D3 of noise set 2, 30 pC, 70 pC, 130 pC are also shown in figure 19 for visual comparison.

Taking the db7D3 of the noise set and the 160 pC apparent charge level set as the feature for SVM discrimination, with a 100% training accuracy, the obtained identification rates by using the well-trained SVM are shown in table 2.

Compared with using the signal frequency spectrum as the input feature the db7D3-based SVM results are particularly impressive and give a measurement sensitivity better than that achieved using the static filter.

6. Discussion

The feasibility of using hardware and software to eliminate or discriminate electro-optic modulator system noise has been assessed. A passive band-stop filter and a data mining technique have been applied as the proposed hardware and software methods. Both methods can achieve promising results and are easy to implement in a practical application. By properly selecting the mother wavelet, decomposition level, kernel and parameters of the SVM, the wavelet decomposition coefficient based SVM represents an improved performance compared to the passive filter.

Some basic principles on selecting parameters used in the signal processing can be concluded at this stage of the research. Daubechies family and Symlet family mother wavelet at high level (such as higher than 6) can produce satisfactory processing results when applied to common PD pulses that have a non-symmetric Gaussian pulse shape. This is in agreement with the results published by other researchers [9, 17]. The decomposition level represents the bandwidth used to filter the signal. Therefore, the selection should be based on best frequency response performance of the PD sensor. In this paper, the Gaussian-RBF kernel function used with the SVM is chosen due to its good performance for general applications. However, there may exist other customized kernels that are more suitable for specific PD signal processing tasks, this is an area of future research.

Having improved the SNR of the optical system, the measurement sensitivity of the whole system is clearly dependent on the sensitivity and output power of the PD detection sensor itself.

7. Conclusions

A method to increase detection sensitivity for an optical phase modulator based PD monitoring system has been described. This new approach is entirely based on a machine learning technique, namely the support vector machine. As a practical application of the machine learning technique to data mining problems, a SVM has been chosen and used as the automatic identification tool. Signals from the EOM photoreceiver are filtered by a db7 wavelet filter. The obtained detail decomposition coefficient results at level 3 are selected as the identification feature for the SVM. Firstly, the SVM is trained by the PD data at a detectable apparent charge level, and then noisy data with buried PD signals are used to test the performance of the well-trained SVM. By choosing a proper kernel and adjusting the parameters used for the SVM, an optimized automatic identification rate has been achieved. This has a detection sensitivity of 30 pC and, if applied, this approach to denoising would greatly increase the capability of an on-line condition monitoring system for high voltage transmission assets.

Acknowledgments

This work is funded through the EPSRC Supergen V, UK Energy Infrastructure (AMPerES) Grant in collaboration with UK electricity network operators working under Ofgem's Innovation Funding Incentive Scheme—full detail is available at http://www.supergen-amperes.org.

References

[1] Tian Y, Lewin P L, Wilkinson J S, Shroeder G, Sutton S J and Swingler S G 2005 An improved optically based PD detection system for continuous on-line monitoring of HV cables *IEEE Trans. Dielectr. Electr. Insul.* **12** 1222–34

[2] Hao L, Lewin P L, Tian Y, Wilkinson J S, Swingler S G and Sutton S J 2006 Application of electro-optic modulation technique for PD monitoring of power transformers *Proc. 2006 IEEE Conf. on Int. Symp. on Electrical Insulation (Toronto)* pp 412–5

[3] Tian Y, Lewin P L, Wilkinson J S, Sutton S J and Swingler S G 2005 Improvement on measurement sensitivity of the optical remote sensing based PD continuous on-line monitoring system for HV cable joints *Proc. 2005 IEEE Int. Symp. on Electrical Insulating Materials (Kitakyushu)* pp 841–4

[4] Tian Y, Lewin P L, Davies A E, Sutton S J and Swingler S G 2003 Partial discharge detection in cables using VHF capacitive couplers *IEEE Trans. Dielectr. Electr. Insul.* **10** 343–53

[5] Han B 2004 Partial discharge monitoring of power transformers *PhD Thesis* University of Southampton

[6] Wang P, Lewin P L, Tian Y, Sutton S J and Swingler S G 2004 Application of wavelet-based de-noising to online measurement of partial discharges *Proc. 2004 IEEE Int. Conf. on Symp. on Solid Dielectrics (Toulouse)* pp 619–22

[7] Kyprianou A, Lewin P L, Efthimiou V, Stavrou A and Georghiou G E 2006 Wavelet packet denoising for online partial discharge detection in cables and its application to experimental field results *Meas. Sci. Technol.* **17** 2367–79

[8] Yang L and Judd M D 2003 Recognising multiple partial discharge sources in power transformers by wavelet analysis of UHF signals *IEE Proc. Sci. Meas. Technol.* **150** 119–26

[9] Chang C S, Jin J, Kumar S, Su Q, Hoshino T, Hanai M and Kobayashi N 2005 Denoising of partial discharge signals in wavelet packets domain *IEE Proc. Sci. Meas. Technol.* **152** 129–40

[10] Vapnik V N 1995 *The Nature of Statistical Learning Theory* (New York: Springer)

[11] Scholkopf B and Smola A J 2002 *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond* (Cambridge, MA: MIT Press)

[12] Bian Z and Zhang X 2000 *Pattern Recognition* 2nd edn (Beijing: Tsinghua University Press)

[13] Hao L, Lewin P L and Tian Y 2005 Partial discharge discrimination using a support vector machine *Proc. 14th Int. Symp. on High Voltage Engineering (Beijing)* (CD-ROM)

[14] Hao L, Lewin P L and Dodd S J 2006 Comparison of support vector machine based partial discharge identification parameters *Proc. 2006 IEEE Conf. on Int. Symp. on Electrical Insulation (Toronto)* pp 110–3

[15] Shawe-Taylor J and Cristianini N 2004 *Kernel Methods for Pattern Analysis* (Cambridge: Cambridge University Press)

[16] Chang C S, Jin J, Chang C, Hoshino T, Hanai M and Kobayashi N 2005 Separation of corona using wavelet packet transform and neural network for detection of partial discharge in gas-insulated substations *IEEE Trans. Power Deliv.* **20** 1363–9

[17] Ma X, Zhou C and Kemp I J 2002 Interpretation of wavelet analysis and its application in partial discharge detection *IEEE Trans. Dielectr. Electr. Insul.* **9** 446–57