GCU
Glasgow Caledonian University

University for the Common Good

Rough set theory applied to pattern recognition of partial discharge in noise affected cable data

Peng, Xiaosheng; Wen, Jinyu; Li, Zhaohui; Yang, Guangyao; Zhou, Chengke; Reid, Alistair; Hepburn, Donald M.; Judd, Martin D.; Siew, W. H.

Rough Set Theory Applied to Pattern Recognition of Partial Discharge in Noise Affected Cable Data

Xiaosheng Peng, Jinyu Wen, Zhaohui Li, Guangyao Yang
State Key Laboratory of Advanced Electromagnetic Engineering and Technology
School of Electrical and Electronic Engineering
Huazhong University of Science and Technology, Wuhan, China, 430074

Chengke Zhou, Alistair Reid, Donald M. Hepburn
School of Engineering and Built Environment
Glasgow Caledonian University
Cowcaddens Road, Glasgow, UK, G4 0BA

Martin D. Judd and **W. H. Siew**
Department of Electronic & Electrical Engineering
University of Strathclyde
Royal College, 204 George Street, Glasgow, UK, G1 1XW

ABSTRACT

This paper presents an effective, Rough Set (RS) based, pattern recognition method for rejecting interference signals and recognising Partial Discharge (PD) signals from different sources. Firstly, RS theory is presented in terms of Information System, Lower and Upper Approximation, Signal Discretisation, Attribute Reduction and a flowchart of the RS based pattern recognition method. Secondly, PD testing of five types of artificial defect in ethylene-propylene rubber (EPR) cable is carried out and data pre-processing and feature extraction are employed to separate PD and interference signals. Thirdly, the RS based PD signal recognition method is applied to 4000 samples and is proven to have 99% accuracy. Fourthly, the RS based PD recognition method is applied to signals from five different sources and an accuracy of more than 93% is attained when a combination of signal discretisation and attribute reduction methods are applied. Finally, Back-propagation Neural Network (BPNN) and Support Vector Machine (SVM) methods are studied and compared with the developed method. The proposed RS method is proven to have higher accuracy than SVM and BPNN and can be applied for on-line PD monitoring of cable systems after training with valid sample data.

Index Terms — **Partial Discharges, Pattern Recognition, Signal Processing, Rough Set, Knowledge Acquisition, Cables.**

1 INTRODUCTION

AUTOMATIC interference signal recognition and PD recognition from different sources play important roles in PD based condition monitoring of cable systems [1-3]. A recommended flow chart for automatic PD and interference signal recognition for an on-line condition monitoring system, containing four working steps, is shown in Figure 1.

Step 1 is named as Training Step. After data denoising, pulse extraction and signal identification, training data sets

are divided into PD signals and interference signals. Then the data containing known signal types are fed into a pattern recognition algorithm, for example, RS [3,4], Neural Network [5] or SVM [6,7], to obtain the criteria, which are fed into step 2.

Step 2 is named as Application Step. The criteria are transformed into programmable codes, which are applicable for autonomous classification of large volumes of application data.

Step 3 is named as Validation Step. Validation is carried out according to signal visualisation criteria, e.g. parameter clustering analysis, to evaluate the decision rules and identify the errors of signal classification. These are fed into

step 4.

Step 4 is named as Adjustment Step. Effective PD evaluation methods are applied for further identification and adjustment, e.g. Phase Resolved PD (PRPD) Pattern analysis [2], to create new criteria, which are added to the criteria in step 1 to improve system effectiveness.

In Figure 1, methodology selection for pattern recognition in step 1 is significant for the success application of on-line pattern recognition of PD and interference signals. Generally speaking, there are two types of pattern recognition methodologies. In "Black Box" systems, e.g. SVM, BPNN, the assessment criteria are not available. In "White Box" systems, e.g. RS, Decision Tree, the criteria are available [8-10]. It is important to select a "White Box" method for on-line recognition of PD and interference signals for two major reasons. Firstly, it is easy to transform the criteria of "White Box" into IF THEN rules, which could be easily integrated into the on-line monitoring programs. Secondly, it is easy to modify the "White Box" criteria using the feedback from step 4, which will improve the accuracy of the pattern recognition through the application of on-line condition monitoring systems.

In comparison with other pattern recognition methods, RS has two significant advantages. Firstly, RS is a "White Box" method, which could transfer information into decision rules being applied for further pattern recognition and signal classification. Secondly, RS is able to identify and remove redundant information and is able to identify the most effective parameters from the input data [8, 9].

In this paper, RS based pattern recognition for interference rejection and PD recognition from different sources is studied, based on experimental data obtained in high voltage laboratory studies of PD from different cable fault types.

2 ROUGH SET THEORY

Rough Set (RS) is a mathematical tool that extracts information automatically from data by using matrix algebra and Set Theory to investigate correlations amongst data [4]. It has been used in many areas, including approximate classification, pattern recognition, process control, expert systems and data mining, e.g. [11-14]. Some background knowledge of RS is provided, including Information System, Lower and Upper Approximation, Signal Discretisation, Attribute Reduction, flowchart of RS based Pattern recognition, but fuller information can be gained from a paper presented at CIRED [4] on the application of RS Theory to PD assessment.

2.1 INFORMATION SYSTEM

An Information System (IS) can be defined as containing a finite set of objects with defined attributes, i.e. IS = (U, A), where U is a non-empty, finite sets of objects which require to be searched though to develop relationships and A is non-empty, finite sets of attributes which can be assigned to the objects [15]. For every object X_n in the data set, there is a set of defined attributes, $a \in A$, where V_a expresses the set of values a, called the domain of attribute a. One of the attributes which is assigned in the data pertaining to a particular object, X_n, is a decision as to the "condition" of the object, based on the other attributes.

Table 1 presents a typical IS for data used to classify the flower Iris into 1 of the 3 types of iris genus, i.e. Setosa, Versicolor and Virginica, based on defined approximations of attributes [16]. The full data set contains 150 objects in the set U, i.e. types of iris to be classified. Each of the objects has 5 attributes in set A, namely sepal length and width, petal length and width and the decision on classification. Thus, for this
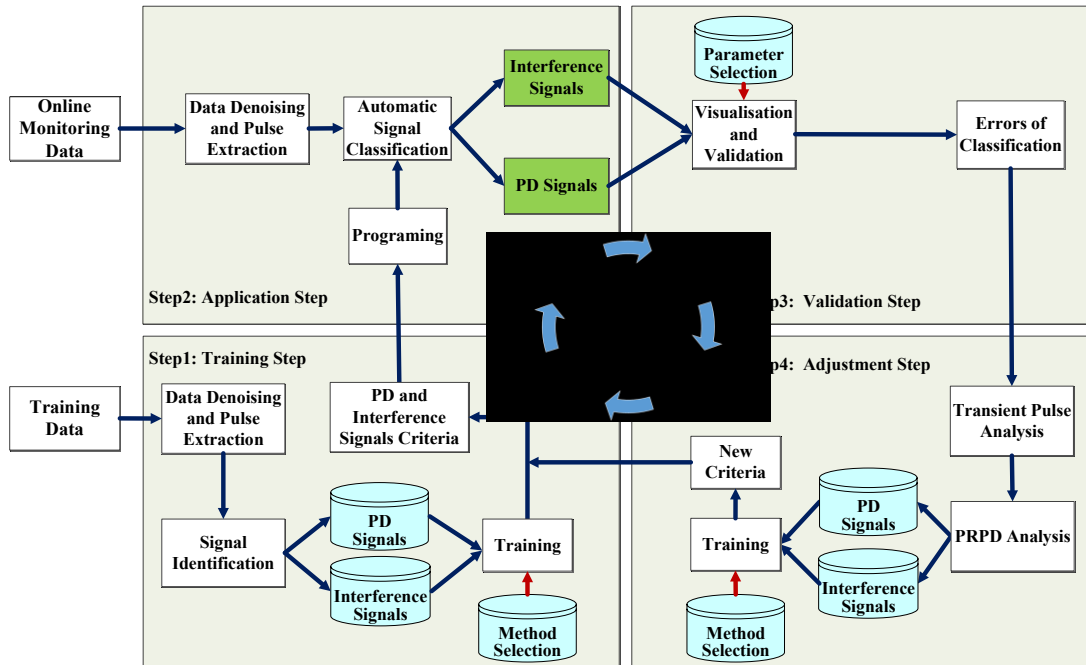

Figure 1. Flowchart of automatic PD recognition of online condition monitoring systems

example, the data sets are defined as: U = {X$_1$, X$_2$, … X$_{150}$}, A = {Sepal length, Sepal Width, Petal Length, Petal Width, Decision}. The decision state will assign each object to 1 of the 3 types of iris genus based on defined approximations of the 4 types of attributes. The approximation process is covered in section 2.2.

Table 1. Information system of iris data.

U	Sepal Length	Sepal Width	Petal Length	Petal Width	Decision
x1	5.1	3.5	1.4	0.2	Setosa
…	…	…	…	…	…
x51	7.0	3.2	4.7	1.4	Versicolor
…	…	…	…	…	…
x101	6.3	3.3	6.0	2.5	Virginica
…	…	…	…	…	…

2.2 LOWER AND UPPER APPROXIMATION

For objects in an Information System, for every subset of attributes B ⊆ A, an associated equivalence relation, called the B-indiscernibility relation, and denoted by IND (B), is defined as follows:

$$IND(B) = \{(x, y) \in U \times U : \forall a \in B, a(x) = a(y)\}$$

The equivalence classes of the indiscernibility relationship form upper and lower approximations to the values which can be used to define whether an object fits within a decision class, these are used in inductive learning to differentiate certain and uncertain knowledge [14]. As in section 2.1, X denotes a subset of elements in the finite sets of objects $U(X \subset U)$, which is the union of the training objects.

The lower approximation of X in B space, denoted as $\underline{BX}$, is defined as the elements that certainly belong to the elements of X. The upper approximation of X in B space, denoted as $\overline{BX}$, is defined as the union of these elements which possibly belong to the set X.

Similarly, all of the training data that can be potentially distinguished from others, conditioned on one or more extracted features (B) of all of the attributes (A), is denoted as the upper approximation of X in the space of these PD features (B). The classified objects can be further confirmed as a member that is potentially related to a classification.

The principle of set X, upper approximation, lower approximation, boundary and the demonstration of the application of RS theory to PD recognition are shown in Figure2.
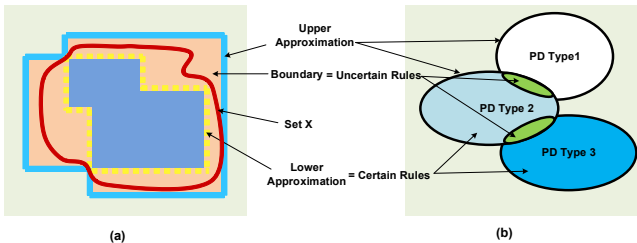

Figure 2. (a) Schematic demonstration of the upper (blue solid line) and lower (yellow broken line) approximation of set X, (b) the demonstration of the application of RS theory to PD recognition.

2.3 SIGNAL DISCRETISATION

Signal discretisation is used to transform continuous attributes into discrete ones. Taking PD rise time as an example, if the PD rise time for a given detection set-up ranges from 10 ns to 1000 ns, the continuous data can be converted into discrete values for use in RS analysis. If the objective is to identify the relationship between PD rise time and the PD sources, e.g. from cable insulation and cable termination, two types of attribute discretisation is required.

First, decision attributes are labeled as:

- PD from cable insulation is assigned a value 0;
- PD from cable termination is assigned a value 1;
- PD from cable joints is assigned a value of 2;
- PD from switchgear box is assigned a value of 3;

and so on.

The second attribute to be assigned could relate to, for example, PD signal rise time (RT), e.g. assignments could be:

- RT between 10 to 100 ns are assigned a value of 0;
- RT between 100 to 200 ns are assigned a value of 1;
- RT between 200 to 500 ns are assigned a value of 2; and
- RT between 500 to 1000 ns are assigned a value of 3.

Signal discretisation is one of the biggest challenges in application of RS theory, as RS cannot deal with continuous attributes data [13]. Unsupervised signal discretization and supervised signal discretization methodologies are applied in practical cases. Unsupervised signal discretisation includes Equal Interval Width, Equal Frequency Intervals, etc. [14,15]. Supervised signal discretisation includes One-Rule Discretizer (1RD), Recursive Minimal Entropy Partition (RMEP), Naive Scaler, Semi-Naive Scaler, etc. [14,15]. In this paper Equal Frequency Intervals, Naive Scaler, Semi-Naive Scaler and Recursive Minimal Entropy Partition are competitively employed for signal discretisation. More information of the methods can be found in references [14] and [15].

2.4 ATTRIBUTES REDUCTION

If the set of attributes which is being examined contains inter-related information then it should be possible to reduce the number of factors that are considered in the analysis. A knowledge representation system containing the condition attributes A and the decision attributes d is called a decision table. In the decision table, if a set of attributes a$_i$ is superfluous in A, i.e. if it exerts no influence on the lower approximation of one or more focused objects. This can be expressed as IND(B) = IND(B − a$_i$) indicating that the attribute a$_i$ can be removed from the decision table. In this case, the attribute set is reduced. If removal of the attribute set decreases the ability of the system to correctly relate data to a decision then the attribute a$_i$ is indispensable. There are several kinds of method for attribute reduction, e.g. Genetic Algorithm, Johnson's Algorithm, Holte's One Rule, etc. [11,14]. In this paper Genetic Algorithm and Johnson's Algorithm are competitively applied for attribute reduction.

Genetic Algorithm for attribute reduction was first introduced by Jakub Wroblewski in reference [17], which has support for both cost information and approximate solutions [14]. Johnson's Algorithm is a simple greedy algorithm to compute a single reduct only [14]. The algorithm has a natural bias towards finding a single prime implicate of minimal length [14]. More information of the algorithms can be found in references [14] and [15].

2.5 FLOWCHART OF APPLICATION OF ROUGH SET THEORY FOR PATTERN RECOGNITION

The flowchart of application of RS theory for development of PD pattern recognition is shown in Figure 3.

Raw data should first be pre-processed, e.g. to remove noise and interference signals. Thereafter, feature extraction will be carried out on the PD pulses identified. The data from the features then has a discretisation procedure applied, changing the continuous information into discrete condition attributes. To allow the RS algorithm to be trained, data partition divides the data into training and testing data sets.

Using the training data sets, procedures for attribute reduction and decision rules generation are applied in order to generate initial decision rules.

Thereafter, the decision rules are reviewed to remove isolated rules with low support. Using test data sets, rules validation is applied to assess whether the decision rules generated by training data sets are sufficiently robust. If the accuracy is higher than threshold, the final set of decision rules is generated: if the outcome is not satisfactory then the procedure in the flowchart is repeated using other data pre-processing and feature extraction.
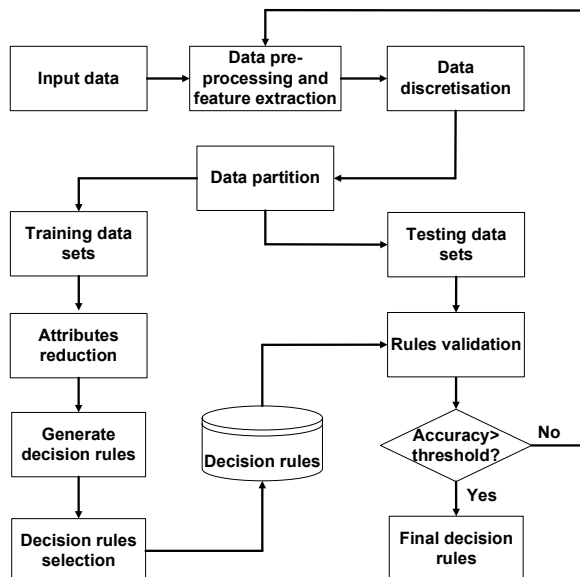

Figure 3. Flowchart of RS based pattern recognition

3 PD DATA GENERATION AND FEATURE EXTRACTION

3.1 EPR CABLE AND ARTIFICIAL DEFECTS

One type of 11kV EPR cable, which is from an in-service breakdown incident in a power generation, was applied for PD testing experiments. The cable has only one point of breakdown. Cable sections which were free of defects were taken to produce artificial defects and then for PD testing.

The dimensions and layers of the EPR cable sample are shown in Figure 4.

Figure 4. The dimensions and layers of the EPR cable sample. Units: mm.

Five types of artificial defect are created in the cable for PD testing: these are shown schematically in Figure 5 and Figure 6. The length of each EPR cable sample is 1.5 m to 2 m. All five defect types are found in in-service cable systems [18].

To generate defect type 1, void crossing PVC over sheath, aluminium armour, sheath, outer semiconductor, EPR insulation layer, a precision 0.4 mm diameter printed-circuit-board drill was applied to the EPR cable sample to generate a hole. Then a copper tape is fixed over the hole to seal the defect, as shown in Figure 5a.
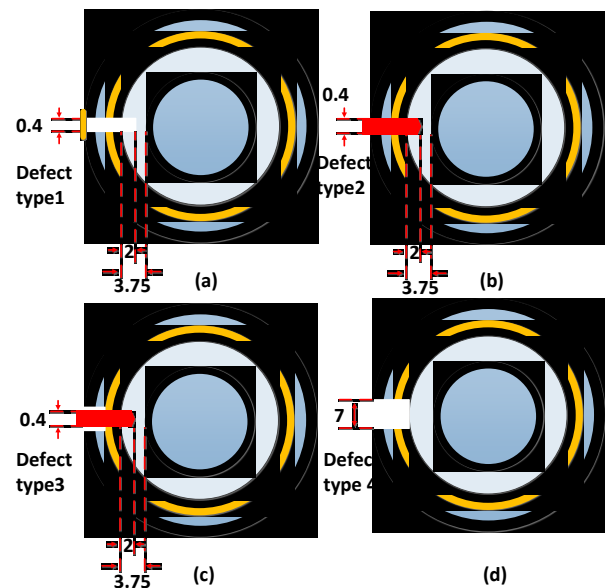
Figure 5. Defect type and dimension: (a) 1: void in insulation; 0.4mm diameter, 2mm depth. (b) 2: protrusion on outer conductor; 0.4mm diameter, 2mm depth. (c) 3: floating protrusion; 0.4mm diameter, 2mm depth. (d) 4: breech in outer conductor; 7mm*7mm.

To generate defect type 2, protrusion on outer conductor [18], a printed-circuit-board drill is inserted into the cable, Figure 5b.

To generate defect type 3, floating protrusion on outer

conductor, a printed-circuit-board drill was inserted into the cable, as shown in Figure 5c. The difference between defect type 2 and 3 is: for type 2 the protrusion is in contact with the outer conductor; for type 3 the protrusion is not in contact with the outer conductor.

To generate defect type 4, breech in outer conductor, a 7 mm*7mm area was cut off from PVC over sheath, aluminium armour, sheath and outer semiconductor, as shown in Figure 5d.

To generate defect type 5, surface discharge around end termination, part of the outer copper sheath of the cable sample is exposed and connected to the earth, as shown in Figure 6.

Figure 6. Defect type 5: Surface discharge around end termination.

3.2 EXPERIMENTAL SETUP

A standard IEC 60270 system was set up for PD testing, based on the artificial defects presented in section 3.1. A commercial High Frequency Current Transformer (HFCT), which is widely applied for on-line PD monitoring of cable systems [1, 2], was integrated into the IEC 60270 system. The bandwidth of the HFCT, as shown in Figure 7, is from 20 kHz to 20 MHz. A LeCroy 104Xi digital oscilloscope is used for data acquisition. A voltage probe connected to the EPR cable conductor gave the 50 Hz sinusoidal phase signal, the trigger signal for the data acquisition. Wide-band amplifiers and wide-band attenuators are applied during testing to adjust the signals into the range of the oscilloscope ($\pm$10V). The PD testing and detection system schematic is given in Figure 8.
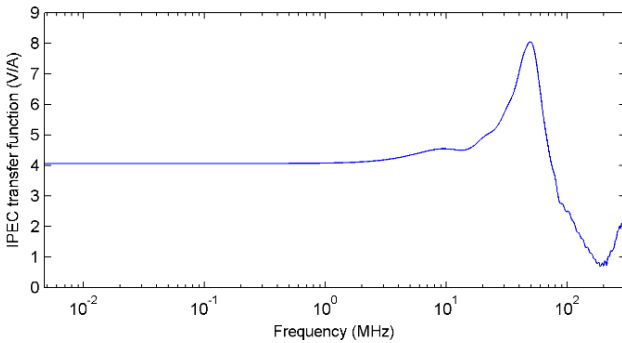

Figure 7. HFCT frequency response.

The sample rate of the data acquisition is 100 MS/s. The sample time length for each data acquisition is 20 ms, one power cycle. Signal from three channels, namely 50 Hz phase

signal, signal from the HFCT and signal from the detection impendence Z_m of the IEC 60270 system, are synchronously logged for further analysis.

During PD testing of each artificial defect, the voltage was increased from 0 to PD Inception Voltage (PDIV), in steps of 1 kV, to 11kV, 12 kV or 13 kV and data from three channels was logged for each voltage level. PD test voltage and the number of sets of 20 ms data for each defect type are shown in Table 2. Although the numbers of data sets for the five defect types are different, the number of PD pulses per defect is similar. More than 400 sets of PD pulses were extracted from each defect and data analysis and pattern recognition applied.
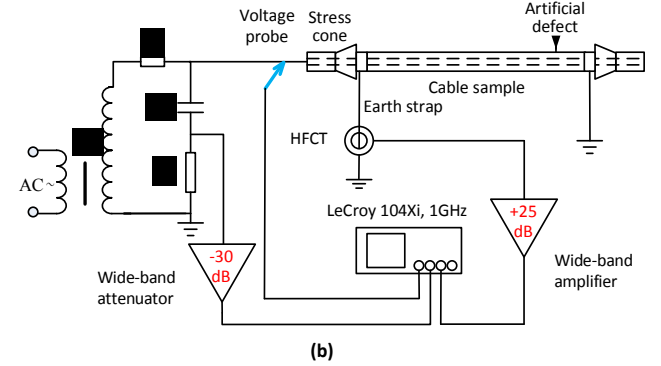

(b)

Figure 8. PD detection system combining a commercial HFCT and an IEC60270 system. Data was measured relative to the 50 Hz AC supply. C_k represents the IEC coupling capacitor and Z_m the measurement impedance.

Table 2. PD test voltage and number of sets of 20 ms data for each defect.

Defect Type	5 kV	6 kV	7 kV	8 kV	9 kV	10 kV	11 kV	12 kV	13 kV	SUM
Type 1	50	50	50	50	50	50	50	0	0	350
Type 2	0	50	50	50	50	50	54	53	52	409
Type 3	0	51	50	50	26	50	52	61	0	340
Type 4	0	0	50	50	51	52	51	0	0	254
Type 5	0	0	10	10	10	10	10	0	0	50

3.3 DATA PRE-PROCESSING AND FEATURE EXTRACTION

Data pre-processing and feature extraction are applied to the raw data to extract pure PD and interference signals for pattern recognition.

As the Signal to Noise Ratio (SNR) of PD detected by IEC 60270 system is high, the threshold is set to two times that of the white noise (recommended by IEC 60270) to rejected noise and interference signals [19]. The SNR of PD detected by HFCT is low, it is challenging for effective PD pulse extraction detected by HFCT. As in Figure 8, data acquisition of the IEC 60270 system and HFCT are fully synchronised with the AC signal, collected in Channel 1. Location information of PD pulses extracted in IEC 60270 system are applied to identify PD pulses detected by HFCT as the timing of PD pulses from the two channels are close (less than 500 ns). Details of PD identification by HFCT, previously named as Synchronous Detection and Multi-information Fusion (SDMF) based signal identification method, is presented in reference [20], and is not fully discussed here. After data pre-processing,

PD and interference signals detected by HFCT are extracted from the raw data.

PD feature extraction was then applied to describe PD activity from different faults in terms of different PD pulse shape and distribution parameters. The three sets of features that could be used for this investigation span different time-bases, i.e. PD features of individual PD pulses (in the nano-second scale), PD features of cumulative data over one power cycle (20 ms) and PD features of cumulative data (over an extended period of time). The relationships of the three kinds of PD features with respect to defect classification are shown in Figure 9.
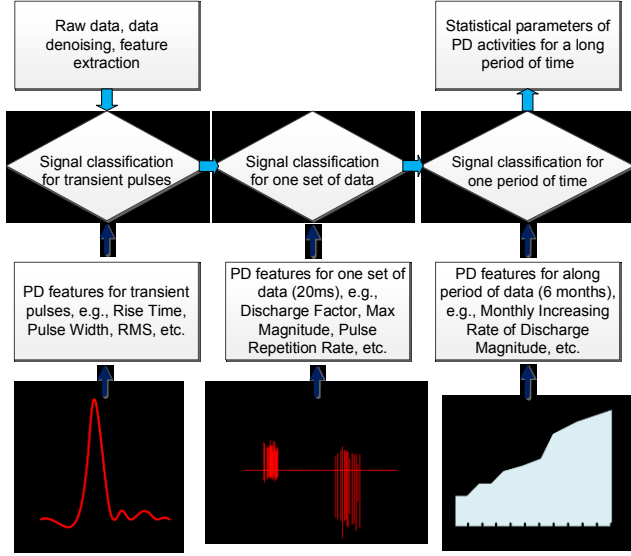

Figure 9. Relationships of three kinds of PD features.

Features of single pulses describe physical characteristics of individual PD pulses. Features for 20 ms data sets describe distribution of pulses in a power cycle and can be used based on phi-q-n identification: low numbers of pulses from certain fault types reduces the effectiveness of these measures. Features over a longer time describe the trend of PD activity and could provide asset managers with information on development of faults.

In this paper, however, due to the limitation on the test duration in laboratory studies, only features for single pulses are applied for pattern recognition. The sample size and defined characteristics of the faults are sufficient for pattern recognition and evaluation. The other sets of features are also applicable for RS based pattern recognition and knowledge discovery but a longer study would need to be undertaken with known sample faults to allow the knowledge rules to be identified with any confidence.

In the present investigation 17 features of transient pulses were adopted including pulse width [21], rise time [21], fall time [21], peak voltage of transient pulse, polarity of transient pulse, mean voltage of transient pulse, root mean square of transient pulse, standard deviation of transient pulse, skewness of transient pulse, kurtosis of transient pulse, crest factor of transient pulse (maximum magnitude over root mean square), form factor of transient pulse (root mean square over average

value), main frequency of transient pulse, phase angle, equivalent time length (T) [22, 23], equivalent bandwidth (W) [22, 23], discharge magnitude in picoCoulomb. The details, including units and example of the features data from a PD pulse, are shown in Table 3.

2000 interference signals and 2000 PD signals from five types of artificial defects in 11kV EPR cables, detected by HFCT, are processed and selected using the system outlined above. The 17 parameters of the 4000 samples, shown in Table 2, are calculated and used as input for investigation of RS based pattern recognition.

Table 3. PD features of individual pulses.

Index	Name	Unit	Example
Feature 1	peak voltage	mV	3.65
Feature 2	polarity	-1 or +1	-1
Feature 3	mean voltage	mV	2.28
Feature 4	root mean square	mV	2.61
Feature 5	standard deviation	mV	1.31
Feature 6	pulse width	10 ns	15
Feature 7	rise time	10 ns	7
Feature 8	fall time	10 ns	8
Feature 9	skewness	-	0.35
Feature 10	kurtosis	-	0.53
Feature 11	crest factor	-	1.39
Feature 12	form factor	-	1.14
Feature 13	main frequency	Hz	0
Feature 14	phase angle	degree	175.3
Feature 15	T	-	23.5
Feature 16	W	-	8.37
Feature 17	discharge magnitude	pC	72.9

4 ROUGH SET BASED INTERFERENCE SIGNAL REJECTION

It is challenging to differentiate PD and interference pulses as the SNR of signals detected by HFCT is low. To consider parameters which can be used to separate the two signal types, an example of a PRPD from defect type 2, using maximum voltage, rise time, equivalent bandwidth [22] and equivalent time length [22] as discriminators, is shown in Figure 10. The visualisation of PD and interference signals shows that it is difficult to distinguish PD and interference signals through the features.

RS based pattern recognition of PD and interference signals is carried out using the process in Figure 3. 1500 PD and 1500 interference signals are applied to train the system: the other 500 PD signals and 500 interference signals are applied to test the system. 4 signal discretisation methods, namely Naive, Semi-Naive, Entropy and Equal Frequency, are applied to the PD and interference samples. Genetic and Johnson attribute reduction methods are applied to the training data sets. To consider the impact of the different discretisation and reduction methods, a matrix of methods was undertaken, as outlined in Table 4. Once the knowledge rules are generated these are then evaluated by applying the test samples to the system. The results of knowledge rules evaluation for the combinations, which are also the results of signal classification, are shown in Table 4.

In Table 4, of all combinations considered, combinations C1 (Naive and Genetic) and C3 (Semi-Naïve and Genetic) have

the highest signal evaluation accuracy, with a value of 99.9%. Processing time includes both training and testing times. Note that combination C1 has the longest processing time, 24.3 s, and C3 has a shorter time, 17.5 s.
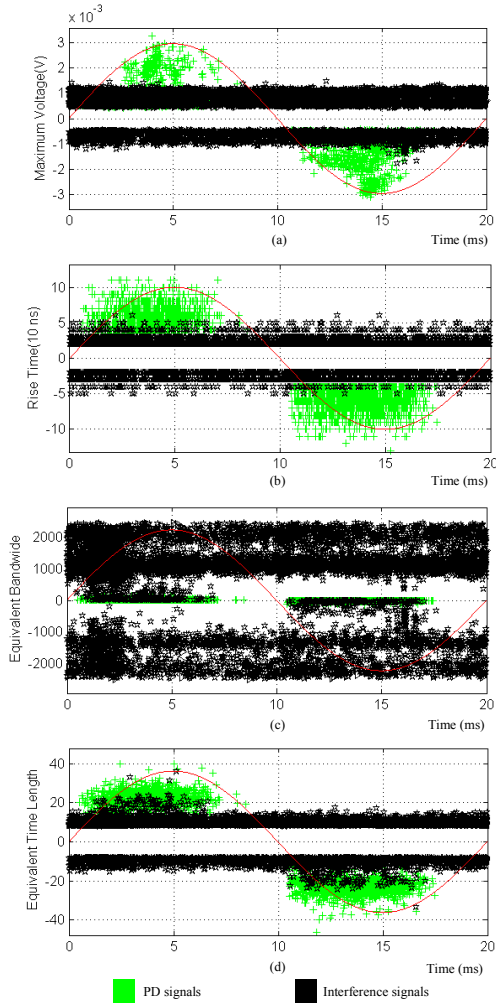

Figure 10. PRPD of four pulse signal parameters from defect type 2 (a) maximum magnitude, (b) rise time, (c) equivalent bandwidth, (d) equivalent time length.

Table 4. Results of decision rules evaluation of PD and interference signals recognition.

Signal Discretisation Algorithm	Attribute Reduction Method	Combination Type	Results	Processing Time(s)
Naive	Genetic	C 1	99.9%	24.3
	Johnson	C 2	95.6%	6.6
Semi-Naive	Genetic	C 3	99.9%	17.5
	Johnson	C 4	95.6%	7.7
Entropy	Genetic	C 5	99%	14.6
	Johnson	C 6	35.6%	5.7
Equal Frequency	Genetic	C 7	67.9%	2.1
	Johnson	C 8	67.9%	1.6

Assessment of the signal evaluation using combination C1 is shown in Table 5. These indicate that only 1 PD pulse is not recognised, i.e. an accuracy of 99.8%, but the accuracy of interference signals identification is 100%. It can be seen that the signal classification accuracy is very robust. As presented in Figure 10, some interference signals have similar magnitude to PD pulses, but the RS based interference signal rejection has

high accuracy despite SNR being low.

Table 5. Results of decision rules evaluation of combination type C1.

		Predicted			
		PD	Interference	Ud*	
Actual	PD	499	0	1	99.8%
	Interference	0	500	0	100%
		100%	100%	0	99.9%

Ud* is for Undefined

5 ROUGH SET BASED PD SIGNAL RECOGNITION FROM DIFFERENT SOURCES

Now that confidence in the ability of the system to separate PD and noise has been established, it is necessary to explore the ability to differentiate different fault types. Figure 11 overlays the PRPD from four PD parameters for five types of defects: defect type is discriminated by color as shown in the key.
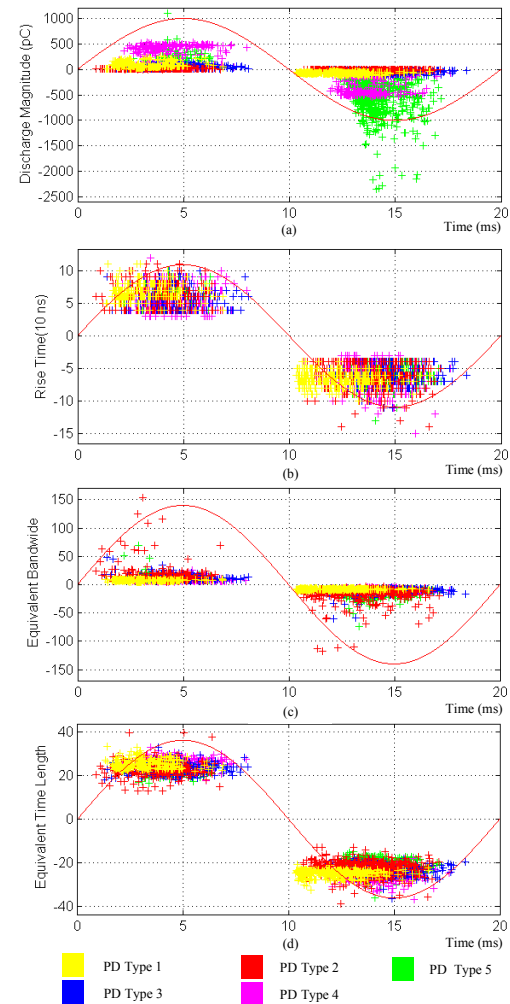
Figure 11. PRPD of four parameters of five types of PD signals (a) PRPD of discharge magnitude of pC, (b) PRPD of rise time, (c) PRPD of equivalent band width, (d) PRPD of equivalent time length.

RS based pattern recognition for PD from different sources is carried out according to the flowchart in Figure 3, based on the 2000 PD samples. 1500 samples PD signals from 5 defect types are randomly selected to train the system, 300 samples

for each fault type; the other 500 PD samples are applied for testing, 100 samples for each fault type. Signal discretisation methods and attribute reduction methods are the same that used for RS based interference signal rejection. The results of decision rules evaluation are shown in Table 6.

Table 6. Results of decision rules evaluation of PD from different sources

Signal Discretisation Algorithm	Attribute Reduction Method	Combination Type	Results	Processing Time (s)
Naive	Genetic	C 1	93%	26.4
	Johnson	C 2	41%	22.7
Semi-Naive	Genetic	C 3	89.2%	26.2
	Johnson	C 4	41%	19.8
Entropy	Genetic	C 5	32.2%	13.0
	Johnson	C 6	1.2%	10.2
Equal Frequency	Genetic	C 7	34.4%	1.5
	Johnson	C 8	34%	1.1

In Table 6, the combination C1 of Naive and Genetic has the highest signal evaluation accuracy, 93%, but combination C1 has the longest processing time, 26.4 s, this includes both training and testing times. The assessment of the signal evaluation for combination C1 is shown in Table 7. It is shown that the average signal classification accuracy is 93%: 1 sample of PD Type 4 is misjudged as Type 1; 1 sample of PD Type 4 is misjudged as PD Type 3.

Table 7. Results of decision rules evaluation of combination type C1.

		Predicted PD Type						
		T1	T2	T3	T4	T5	Ud*	
Actual PD Type	T1	95	0	0	0	0	5	95%
	T2	0	93	0	0	0	7	93%
	T3	0	0	91	0	0	9	91%
	T4	1	0	1	89	0	9	89%
	T5	0	0	0	0	97	3	97%
	Ud*	0	0	0	0	0	0	Ud
		98.9%	100%	98.9%	100%	100%	0	93%

Ud* is for Undefined

6 SVM AND BPNN BASED PD RECOGNITION FROM DIFFERENT SOURCES

To further demonstrate the advantages of RS based PD pattern recognition, two popular PD pattern recognition methods, SVM and BPNN, are applied to PD from different sources and their determination compared with RS method.

SVM was first proposed by Vapnik in 1995 [7]. The input vectors of SVM are mapped to a high-dimensional space by a non-linear mapping in order to identify an optimal hyperplane which makes the gaps between samples of different classes largest [6, 7]. There are four factors, signal normalisation, kernel function, penalty factor C and kernel function paramount γ, which significantly affect the accuracy of SVM based pattern recognition [6, 7]. In this work, normalisation is applied to fit the signals into 0 to +1 range; Radial Basis Function (RBF) is chosen as the kernel function, based on references [5], [6] and [7]; a recommended Grid-Search and Cross-Validation method, reference [22], is employed for optimisation of C and γ. C= 2^{-10}, 2^{-9},…,2^{9},2^{10}, γ=2^{-10}, 2^{-9},…,2^{9},2^{10} are applied for grid-search. The PD recognition accuracies of 1500 training samples under different C and γ are
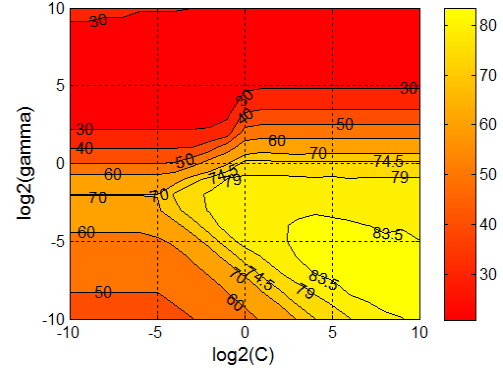
shown in Figure 12.

Figure 12. The training accuracy of different C and γ combination.

The values of 30 to 80 represent the Cross-Validation accuracy from 30% to 80%. The best training accuracy is 86.5%, when C=256 and γ =0.0039063. The Grid-Search time is 331.58 s, which is the training time of SVM based PD recognition. Then the 500 testing samples are input to the trained SVM with C=256 and γ=0.0039063. The results of PD recognition from different sources are shown in Table 8.

Table 8. Results of SVM based PD recognition.

		Predicted PD Type					
		T1	T2	T3	T4	T5	Accuracy
Actual PD Type	T1	100	0	0	0	0	100%
	T2	0	70	30	0	0	70%
	T3	4	17	79	0	0	79%
	T4	0	0	0	95	5	95%
	T5	0	0	0	5	95	95%
							87.8%

The average accuracy is 439/500=87.8%. The time for testing is 0.73 s. So the total time, including training and testing, for the SVM based PD recognition from different sources is 332.31 s.

A three-layered BPNN, which is widely adopted for pattern recognition, is also applied for PD recognition from different sources. As the number of input parameter is 17, the number of nodes in the BPNN input layer is 17. Similarly, as the expected number of PD defects to be recognised is 5, the number of nodes in the BPNN output layer is 5. Although the number of nodes in the hidden layer plays an important role in the success of BPNN based pattern recognition, there is no available theory on how the number of nodes in the hidden layer should be chosen [24]. To compensate for this, several numbers of nodes in the hidden layer, namely 10, 15, 20, 25 and 30, are employed to achieve ideal signal classification accuracy. Signal normalisation is also applied to limit the signals into 0 to +1 range before BPNN based signal classification. The learning rate of the BPNN is 0.01. The 2000 PD samples have the same division of 1500 training samples and 500 testing samples as used for both RS and SVM methods.

The relationships of the numbers of hidden layers, the processing time and the signals classification accuracy are shown in Table 9. The best signal classification is 84.8% when the number of hidden layer is 10 and the training and testing

processing time is 3.4 second.

Table 9. Results of BPNN based PD recognition.

Numbers of hidden layers	Learning rate	Signal classification accuracy	Processing time (s)
30	0.01	81.0%	18.8
25	0.01	81.4%	11.1
20	0.01	78.8%	9.1
15	0.01	84.2%	5.4
10	0.01	84.8%	3.4

The comparison of the three PD pattern recognition methods is shown in Table 10. As the training and testing time evaluation of three methods, RS, SVM and BPNN, are under the same computer, the comparison of computational time is valid.

Table 10. Comparison of three kinds of pattern recognition methods.

	Accuracy	Time	Key steps	Advantages	Disadvantages
RS	93 %	26.4s	Signal Discretisation, Attribute Reduction	"White Box", High accuracy	processing time will increase with number of samples
SVM	87.8%	332.3s	Signal normalisation, kernel function, C, G	High accuracy, adaptive generation ability	"Black Box"
BP NN	84.8%	1.3s	Signal normalisation, numbers of hidden Layers	High accuracy, adaptive generation ability	"Black Box"

In Table 10, RS shows the best signal classification accuracy among three methods, which proves the effectiveness of RS based PD recognition.

A disadvantage of RS is that the method takes more processing time for training and testing than BPNN. The reason for this increasing time for RS is that as the number of sample increases the number of decision rules will also increase. As the testing samples are evaluated with all the decision rules, increasing the number of decision rules increases the time taken.

Although both SVM and BP are capable of adaptive generation for large samples and have high signal classification accuracy, the biggest disadvantage of the methods is that they are "Black Box" and do not allow the user to understand the decision making process.

7 CONCLUSIONS AND DISCUSSIONS

RS based pattern recognition has been shown capable of interference signal rejection and, when compared with SVM and BPNN methods, higher accuracy for PD signal recognition from different sources. Conclusions can be drawn as following:

- RS has high signal classification accuracy: 99.9% for interference signal rejection and 93% for PD recognition of different sources, when the Naive algorithm based signal discretisation and Genetic algorithm based attribute reduction is applied.
- Signal discretisation and attribute reduction methods play important roles in RS based PD and interference signals pattern recognition.
- Analysis has shown that, based on the experimental PD data, the Naive algorithm based signal discretisation is one of the most effective signal discretisation methods among 4 different methods.
- Analysis has also shown that, based on experimental data, Genetic Algorithm based attribute reduction is an effective method for RS based pattern recognition.
- RS is proven to provide higher signal classification accuracy than SVM and BPNN, based on 2000 PD samples from laboratory studies of 5 fault types.
- RS is shown to require more processing time than BPNN, and less processing time than SVM, based on the 2000 PD samples.

REFERENCES

[1] C. Zhou, M. Michel, D.M. Hepburn and X. Song, "On-line partial discharge monitoring in medium voltage underground cables", IET Sci. Measurement Technology, Vol. 3, No. 5, pp. 354–363, 2009.

[2] X. Peng, C. Zhou, D. M. Hepburn, M. D. Judd and W. H. Siew, "Application of K-Means Method to Pattern Recognition in On-line Cable Partial Discharge Monitoring", IEEE Trans. Dielectr. Electr. Insul., Vol. 20, No. 3, pp. 754-761, 2013.

[3] D. Dey, B. Chatterjee, S. Chakravorti and S. Munshi, "Cross-wavelet transform as a new paradigm for feature extraction from noisy partial discharge pulses", IEEE Trans. Dielectr. Electr. Insul., Vol. 17, No. 1, pp. 157-166, 2010.

[4] C. Zhou, M. Michel, D. M. Hepburn and G. Zhang, "Rough Set Theory for Data Mining in an On-line Cable Condition Monitoring System", Congrès International des Réseaux Electriques de Distribution (CIRED), Vienna, Austria, Paper No. 0068, 2007.

[5] K. X. Lai, B. T. Phung and T. R. Blackburn, "Application of data mining on partial discharge part I: predictive modelling classification", IEEE Trans. Dielectr. Electr. Insul., Vol. 17, No. 3, pp. 846-854, 2010.

[6] L. Hao and P. L. Lewin, "Partial Discharge Source Discrimination using a Support Vector Machine", IEEE Trans. Dielectr. Electr. Insul., Vol. 17, No.1, pp. 189 -197. 2010.

[7] J. Tang, F. Liu, X. Zhang, X. Liang and Q. Fan, "Partial discharge recognition based on SF6 decomposition products and support vector machine", IET Sci. Measurement Technology, Vol. 6, No. 4, pp. 198-204, 2012.

[8] Z. Wang, C. Guo, Q. Jiang and Y. Cao, "A Fault Diagnosis Method for Transformer Integrating Rough Set with Fuzzy Rules", Trans. of Institute of Measurement and Control, Vol. 28, No. 30, pp. 243-251, 2006.

[9] G. Lambert-Torres and A. P. Alves, "Classification of Power System Operation Point Using Rough Set Techniques", IEEE International Conference on Systems Man and Cybernetics, Beijing, China, Vol. 3, pp. 1898-1903, 1996.

[10] H. Hirose, M. Hikita, S. Ohtsuka, S. Tsuru and J. Ichimaru, "Diagnosis of Electric Power Apparatus using the Decision Tree Method", IEEE Trans. Dielectr. Electr. Insul., Vol. 15, No. 5, pp. 1252-1260, 2008.

[11] C. L. Hor, P. Crossley, S. Watson and D. Millar, "Rough Set Theory: A True Landmark in Data Analysis Studies in Computational Intelligence", Springer Berlin Heidelberg Press, pp. 279-320, 2009.

[12] S. Mitra, M. Mitra and B. B. Chaudhuri, "A Rough-Set-Based Inference Engine for ECG Classification", IEEE Trans. on Instrumentation and Measurement, Vol. 55, No. 6, pp. 2198-2206, 2006.

[13] S. Dalai, B. Chatterjee, D. Dey, S. Chakravorti and K. Bhattacharya, "Rough-Set-Based Feature Selection and Classification for Power Quality Sensing Device Employing Correlation Techniques", IEEE Sensors J., Vol. 13, No. 2, pp. 563-573, 2013.

[14] A. Øhrn, "ROSETTA Technical Reference Manual", Department of Computer and Information Science, Norwegian University of Science and Technology , Trondheim, Norway, 2000.

[15] A. Øhrn, *Discernibility and Rough Sets in Medicine: Tools and Applications*, Ph.D. dissertation, Dept. Comput. Inf. Sci., Norwegian Univ. Sci. Technol., Trondheim, Norway, 1999.

[16] J. C. Bezdek, J. M. Keller, R. Krishnapuram, L. I. Kuncheva and N. R. Pal, "Will the real iris data please stand up? ", IEEE Trans. Fuzzy Systems, Vol. 7, No. 3, pp. 368-369, 1999.

[17] J. Wroblewski, "Finding minimal reducts using genetic algorithms", Proc. of the Int'l. Workshop on Rough Sets Soft Computing at the Second Annual Joint Conf. Information Sci., Wrightsville Beach, USA, pp.186-189, 1995.

[18] A. J. Reid, X. Peng, X. Hu, M. D. Judd, W. H. Siew, C. Zhou and D. M. Hepburn, "Comparison Of Partial Discharge Characteristics From Insulation Defects In 11kV EPR Cable", 17th Int'l. Sympos. High Voltage Eng., Hannover, Germany, 2011.

[19] IEC International Standard 60270, "High Voltage Test Techniques - Partial Discharge Measurements", Int'l Electrotechnical Commission (IEC), Geneva, Switzerland, 3rd edition, 2000.

[20] X. Peng, J. Wen, Z. Li, G. Yang, C. Zhou, A. Reid, D. M. Hepburn, M. D. Judd and W. H. Siew, "SDMF based Interference Rejection and PD Interpretation for Simulated Defects in HV Cable Diagnostics", IEEE Trans. Dielectr. Electr. Insul., Vol. 24, No. 1, 2017. TDEI Paper No. 6002.

[21] J. Densley, T. Kalicki and Z. Nodolny, "Characteristics of PD pulses in electrical trees and interfaces in extruded cables", IEEE Trans. Dielectr. Electr. Insul., Vol. 8, No. 1, pp. 48-57, 2001.

[22] G. C. Montanari and A. Cavallini, "A New Approach to Partial Discharge Testing of HV Cable Systems", IEEE Electr. Insul. Mag., Vol. 22, No. 1, pp. 14-23, 2006.

[23] W. Si, J. Li, P. Yuan and Y. Li, "Digital Detection, Grouping and Classification of Partial Discharge Signals at DC Voltage", IEEE Trans. Dielectr. Electr. Insul., Vol. 15, pp. 1663-1674, 2008.

[24] K. X. Lai, B. T. Phung and T. R. Blackburn, "Application of data mining on partial discharge part I: predictive modelling classification", IEEE Trans. Dielectr. Electr. Insul., Vol. 17, No. 3, pp. 846-854, 2010.

Xiaosheng Peng (M'11) received the B.Sc. and M.Sc. degrees from Huazhong University of Science and Technology, China in 2006 and 2009, respectively, and the Ph.D. in electrical engineering at Glasgow Caledonian University in 2012 funded by EPSRC. He has worked as a Post-Docroral Research in Glasgow Caledonian University funded by EDF Energy. He is currently a lecturer in School of Electrical and Electronic Engineering of Huazhong University of Science and Technology. His research interests are partial discharge signal processing and condition monitoring of power plant. He is a member of IET.

Chengke Zhou (M'06) received the B.Sc. and M.Sc. degrees in electrical engineering from Huazhong University of Science and Technology, China in 1983 and 1986, respectively, and the Ph.D. degree from the University of Manchester U.K., in 1994. Since then, he worked in Glasgow Caledonian University, U.K., as a Lecturer, Senior Lecturer and in Heriot-Watt University as a Reader until 2007 when he returned to Glasgow Caledonian University as a Professor. He has published over 100 papers in the area of PD based condition monitoring of MV/HV plant and power system analysis. He is member of IET.

Donald M. Hepburn received his B.A. (Hons) from the Open University in 1987 and the Ph.D. degree from Glasgow Caledonian University (GCU) in 1994. He has many years of industrial research experience and has been involved in research into high voltage insulation systems at GCU for over 20 years. His research interests cover monitoring of chemical changes to insulation materials, application of electrical, acoustic and RF monitoring equipment to HV components and application of advanced digital signal processing to information from the monitoring techniques. He is a Senior Lecturer at GCU and is involved in industrial and academic research projects.

Alistair J. Reid (M'11-SM'15) graduated from the University of Strathclyde, UK, in 2004 with a B.Eng. (Hons) degree in electrical and mechanical engineering and received the Ph.D. degree in 2007 for research on partial discharge monitoring. Subsequently, he has worked as a Post-Docroral Research at the University of Strathclyde, a Research Fellow at Glasgow Caledonian University and a Visiting Research Associate at California Institute of Technology. He is presently a Lecturer in the School of Engineering and Built Environment at Glasgow Caledonian University. His research interests include diagnostic monitoring techniques and systems, partial discharge, sensor design, embedded systems and energy harvesting. Alistair is a Senior Member of the IEEE and a committee member of the Universities High Voltage Network (UHVnet) in the UK.

Martin D. Judd (M'02-SM'04) is the Technical Director of High Frequency Diagnostics Ltd. He graduated from the University of Hull in 1985 with a first class (Hons) degree in Electronic Engineering, after which he gained 8 years of industrial experience, first with Marconi Electronic Devices and then with EEV Ltd. Martin received his PhD from the University of Strathclyde in 1996 for research into the excitation of UHF signals by partial discharges in gas insulated switchgear. He has worked extensively on UHF partial discharge location techniques for power transformers and was latterly Professor of High Voltage Technologies at the University of Strathclyde, where he managed the High Voltage Research Laboratory. In 2014 he founded High Frequency Diagnostics, a contracting and consultancy business that works in partnership with companies and universities to maximize the impact of R&D outputs by utilizing them in new technologies and applications.

W. H. Siew (M'95) is a Reader in the Department of Electronic & Electrical Engineering, University of Strathclyde, Glasgow, Scotland. He is a triple alumnus of the University of Strathclyde with degrees of B.Sc. (Hons) in electronic & electrical engineering; Ph.D. in electronic & electrical engineering; and Master of Business Administration. His areas of research interest include large systems electromagnetic compatibility; cable diagnostics; lightning protection; and wireless sensing systems. He is Convener of the CIGRE WG C4.208 and a member of the Technical Advisory Panel for the IET Professional Network on Electromagnetics. He is also a member of IEEE TC7. He is a Chartered Engineer and an MIEE.