Expert Systems with Applications 31 (2006) 345–350

Expert Systems with Applications

www.elsevier.com/locate/eswa

ART 2—an unsupervised neural network for PD pattern recognition and classification

B. Karthikeyan [a,*], S. Gopal [b], S. Venkatesh [c]

[a] *Electrical and Electronics Department, SASTRA Deemed University, Thanjavur 613 402, Tamilnadu, India*
[b] *M/s W.S Test Systems Ltd, 27th kM, Bellary Road, Doddajala Post, Bangalore 562 157, Karnataka, India*
[c] *SRM Deemed University, Kattankulathoor, Chennai, Tamilnadu, India*

Abstract

This paper introduces a method of classifying partial discharges of unknown origin. The innovative trend of using Artificial Neural Network (ANN) towards classification of Partial Discharge (PD) patterns is cogent and discernible. The Adaptive Resonance Theory (ART), a type of neural network which is suitable for PD pattern recognition is explained here. To ensure the suitability and reliability of chosen network for PD pattern recognition, the network is tested with the well known Iris plant database and alphabet character for recognition & classification. Further more the network is trained with various combinations of ϕ–q–n distributions of PD patterns and tested. It is shown that the ART 2 network is able to classify the PD patterns. The paper ends with analyzing the efficacy of multifarious features selected in the measurement space. Also the validation of input features is done using 'Hold-One-Out' method and partial set training technique

Keywords: Partial discharge (PD); Adaptive Resonance Theory (ART 2); Pattern recognition

1. Introduction

One of the prime causes of electrical insulating system failure in high-voltage power equipment arises from partial discharges that occur in gas filled cavities that undergo ionization and subsequent partial discharge when subjected to elevated electrical stresses. The current process of assessing the dielectric integrity of insulating systems is by conducting a Non-Destructive Test i.e. Partial discharge test which gives information regarding the healthiness of insulation which is quantified by the magnitude of apparent charge (Pedersen, Crichton, & McAllister, 1991; Van Brunt, 1991). The existing PD meters and the software developed so far, displays the ϕ–q–n patterns and the magnitude of PD. The IEC's has specified limits of PD for every power equipment, failing, which has to be replaced without knowing the type of partial discharge as each PD activity has its own degradation effects. However, the existing PD meters do not give any information regarding the source of PD. It is only relatively recently that a considerable effort has been devoted to apply neural networks to the recognition of partial discharge pulse patterns in order to identify various discharge sites on electrical power apparatus, that may eventually lead to electrical breakdown, as a result of exposure to the chemical and physical degradation effects induced by the partial discharges. In order to assist the neural networks with the pattern recognition task, some of the partial discharge data were statistically preconditioned.

2. Partial discharge

Partial discharge is an electrical breakdown confined to the localized regions of the insulating medium used in power apparatus. They are characterized by pulsating currents which has a very short time period varying between a few nanoseconds (Van Brunt, 1991) up to a few micro seconds. PD may be classified based on the site of occurrence as voids, inclusions, surface discharge etc. PD diagnosis techniques are used as Non-Destructive Test method to detect insulation flaws/faults. The diagnosis involves PD patterns captured during PD measurement unique to each fault. The display of PD pulses in elliptical time base as indicated in Fig. 1 is usually preferred to the sinusoidal time base system for representation of PD pulses.

A variety of PD analysis techniques have been applied on the data to extract characteristic information of PD as indicated in Fig. 2. The distribution analysis approach follows probabilistic models and is popularly used for ageing characterization.

* Corresponding author. Tel.: +91 4362 264101; fax: +91 4362 264120.
E-mail addresses: karthik@eee.sastra.edu (B. Karthikeyan), mwbil@bgl.vsnl.net.in (S. Gopal), venkatsri_73in@yahoo.com (S. Venkatesh).

0957-4174/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2005.09.029

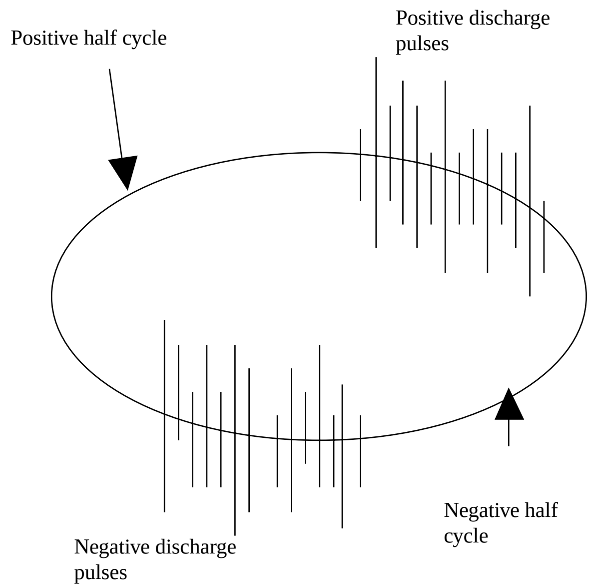

Fig. 1. PD pulse representation on elliptical time base.

The pulse sequence analysis method characterizes the temporal behavior of the data. This approach is useful for understanding the physical process involved in the system. Further, the characterization based on time information has measurement advantages. (Pulse sequence analysis can lead to instrument independent characterization.) The transform/state based techniques such as Fast Fourier Transform (FFT), Artificial Neural Network (ANN), Fuzzy logic etc. are mainly used for classification & recognition of PD defects (Gulski & Kreuger, 1992; Gulski & Krivida, 1993; Kranz, 1993; Kranz & Krump, 1992; Mazroua & Salama, 1993; Satish & Zaengl, 1994; Mazroua, Bartnikas, & Salama, 1994; Krivda, Gulski, Satish, & Zaengl, 1995; Candela, Mirelli, & Schifani, 2000).

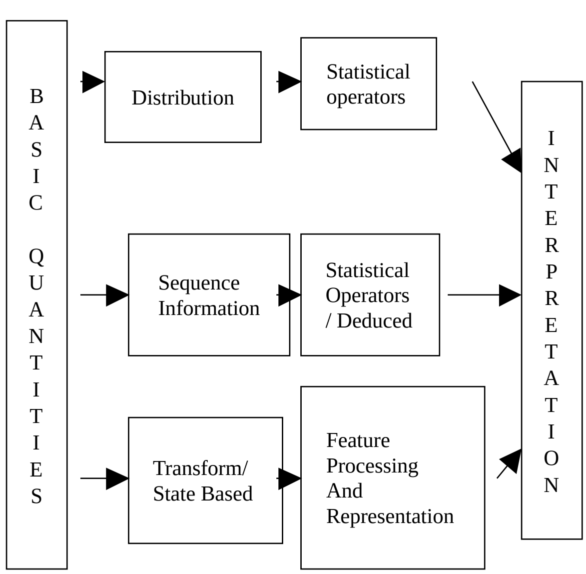

Fig. 2. Techniques used in PD analysis.

3. Why ANN for PD recognition?

Partial discharges in insulation system comprise a large variety of physical phenomena, ranging from low intensity phenomena like charge carrier emission from surfaces over leakage currents along insulator surfaces, glow discharges and charge carrier injection into liquids and solids over a medium intensity such as electrical treeing and streamers to intense discharge types such as leaders, sparks and partial arcs (Van Brunt, 1991). Few of these discharge types may contribute to insulation degradation and some of them may trigger breakdown. Each discharge has its own pattern and there are many cases where the patterns remain same for different faults. The complexity of analyzing PD patterns obtained from digital computer acquisition system is evident as this is a complex non-linear problem (Gulski & Kreuger, 1992; Gulski & Krivida, 1993; James & Phung, 1995; Kranz, 2000). The process being stochastic, the associated effects of memory propagation with the influence of residues from previous PD pulses etc have made the classification of such PD patterns in terms of ϕ–q–n even more complex.

The major variable parameters which characterize the PD pulses and which define the basis of the physical phenomena of PD are the time of occurrence (ϕ), magnitude of discharge (q) and the number of discharges (n) which are represented as three dimensional pattern is taken up for recognition and classification task (Satish & Zaengl, 1994). Neural networks have been investigated as a method for recognizing partial discharge characteristics of several kinds of electrode systems by many researchers (Gulski & Krivida, 1993; Krivda, 1995; Mazroua & Salama, 1993; Salama, 2002; Satish & Zaengl, 1994). As the manual classification requires skilled personnel, use of computer aided measurement and processing technique given by IEC 60270 serve as an aid for recognition and classification of PD patterns.

The pattern recognition basically involves the identification of similar data within a collection, which resembles the new input. Since, Artificial Neural Network (ANN) has the ability to learn from examples (Cachin & Wiesman, 1995; Hong, Fang, & Hilder, 1996; Meijer, Gulski, & Smit, 1998), generalize well from training, handle noisy data conveniently, create their own relationship amongst information and hence no equations, it has become an innovative technique suitable for PD pattern recognition and classification.

3.1. Why ART 2?

Several types of ANNs have been used till date for the classification of PD patterns (Hoof, Freisleben, & Patsch, 1997; Kai, Kexiong, Fuqi, & Chengqi, 2002; Mazroua & Salama, 1993). All the networks used until now for PD recognition gain efficiency with the plasticity of the inner connections, which means that they can change their inner weights and there by processing information. The desirable plasticity is still contrary to the stability of the networks. The modification of the connections must be prevented when the network comes across an accidental value of the input signal. As a result of it, the networks used so far, for the PD recognition task is entirely divided into two phases—a training phase and a recognition phase. Once the network is trained, the network does not change any connections. The network classifies the input patterns with the help of the trained learning patterns. However, if some new signals appear that were unknown during training, the network cannot react to them. The consequence is that the network has to be trained completely anew, if the risk of failure by the classification has to be made as small as possible.

3.2. Adaptive Resonance Theory (ART)

For the solution of Plasticity–Stability dilemma, a theory, which describes some special self-organizing neural networks with a system of differential equations called the Adaptive Resonance Theory, was developed. The name is derived from the fact that the connections of the neurons would only be modified if the input pattern is similar enough to an already known pattern. In such a case there is so-called resonance of a pattern in the network. If the input pattern is not similar, it generates a new category by itself. Therefore, the concept of the ART model allows it to learn new patterns without forgetting already known ones. Thus, the ART network needs no separate training period for the networks used for classification because it learns continuously.

Adaptive Resonance Theory's indicated in Fig. 3 makes much use of a competitive learning paradigm. A criterion is developed to facilitate the occurrence of winner take all phenomenon. A single node with the largest value for the set criterion is declared the winner within its layer and it is said to classify a pattern class. If there is a tie for the winning neuron in a layer, then an arbitrary rule, such as the first of them in a serial order, can be taken as the winner.

The neural network developed for this theory establishes a system that is made up of two subsystems, one being the attentional subsystem, which contains the unit for gain control. The other is an orienting subsystem, which contains the unit for reset. During the operation of the network modeled for this theory, patterns emerge in the attentional subsystem and are called traces of STM (short-term memory) and those of oriental subsystem are called traces of LTM (Long-term memory) are in the connection weights between the input layer and the output layer.

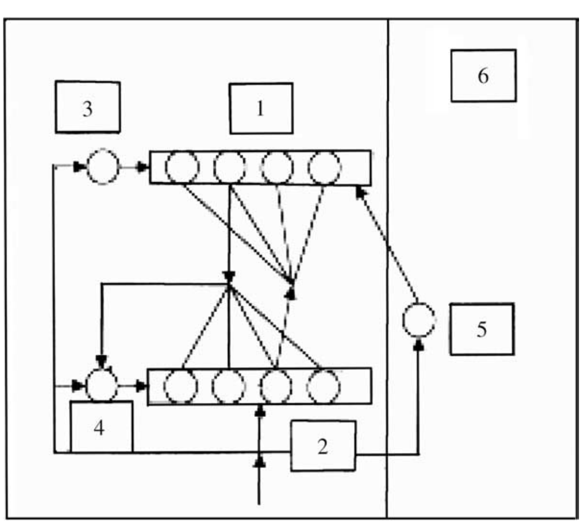

Fig. 3. The Adaptive Resonance Theory neural network architecture.

The network uses the processing with the feedback between its two layers, until the resonance occurs. Resonance occurs when the output in the first layer after the feed back connection from the second layer matches the original pattern used as input for the first layer in that processing cycle. The degree of match, when measured suitably, exceeds a predetermined level, termed the vigilance parameter. Just as a photograph matches the likeliness of the subject to a greater degree when the granularity is higher, the pattern match gets finer when the vigilance parameter is closer to 1.

4. Initial task—verification of the neural network: ART 2

The capability of the ART 2 network for recognition process is verified by alphabet character recognition which has only binary inputs and by Fishers Iris plant data base which has analog inputs. It is inferred that ART 2 network adapts to both binary and analog inputs for recognizing the patterns effectively. The verification methods carried for the ART network proved that the network is highly suitable for the actual PD Pattern classification task and can be implemented for the problem considered here.

4.1. Preliminary considerations

Preliminary considerations involve two important stages. They are the features extraction and classification. The necessity of feature extraction is to capture the distinctive attributes that correlate to each discharge. Hence this step may be called a preprocessing step wherein an m-dimensional vector is mapped into a reduced n-dimensional feature vector where n is equal to the number of extracted features. The feature vector is applied to the ANN to perform the classification task. Thus the concept is that of defining the boundary surface that divides the feature space into a number of disjoint regions that represent the different classes. Thus the problem is now one of making a decision as to which side of the boundary the new input falls.

4.2. Application of inputs to neural networks

The input data obtained from the computer aided PD measurement and acquisition system is provided to the ANN as input in the form necessary for feature extraction so as to capture the distinctive attributes that correlate to the discharge. Such preprocessed data input called the fingerprints now becomes the basis used for feature extraction.

Databank of such several PD data specific to a single type flaw (void, corona, surface discharge and oil corona) that have been extracted from the computer based PD measurement and acquisition system is now used as input. The training phase thus involves training from such a database, which now is preprocessed in the form of fingerprints for each type of pattern. In the test phase an unknown PD pattern is given as input, which is converted in the form of such fingerprints, and the classification is now to one of a particular class of PD.

4.3. Phase window concept

Because of the phase dependent behavior of the PD generated under AC Voltage, several quantities as function of phase angle can be used to describe the PD phenomena. For this purpose the voltage cycle should be divided into phase windows representing the phase angle axis (0–360). If the observation takes place over several voltage cycles, in each phase window the statistical distribution of the individual PD events can be determined. For this purpose in each phase window three quantities are determined.

- Sum of discharge magnitudes observed in one phase window $Q_i = \sum Q_i$
- The number of discharges observed in one phase window, $N = \sum i$
- The average value of discharges observed in one phase window $Q_\mathrm{n} = Q_\mathrm{s}/N$

4.4. Various type of input data

In this case, database of 10 PD patterns for each kind of void, surface discharge, corona and oil corona is used. The database include inputs based on the 'measures of statistical distribution' and 'measures of maximum and minimum values' A partial set from the available pattern group are taken as training patterns and the remaining patterns of the group are used as test patterns. From this set, different combinations of set are taken as input for training and the remaining for testing.

The inputs are based on the following measures such as measures of maximum and minimum value, measures of central tendency, and measures of dispersion. In case of measures of maximum and minimum value, the features are taken from a phase window width of 30 and 10 degrees window width in the distribution. In other cases of measure, measurements from the pattern are taken for every 30 degree window width in the distribution.

4.5. Inputs based on maximum and minimum values

The extracted data in each phase window is divided into eight types based on the measures of maximum and minimum values and the representation is listed below:

1. Maximum Q − Maximum No. of pulses − angle— for 30 degree phase window

(Intrepretation: The maximum value of apparent charge (Q) in the window, the maximum number of pulses and its corresponding instant of phase angle in the phase window width of 30 degree)

2. Maximum Q − angle − No. of pulses— for 10 degree phase window
3. Maximum I − angle − No. of pulses— for 30 degree phase window
4. Minimum Q − angle − No. of pulses— for 30 degree phase window
5. Minimum Q − angle − No. of pulses— for 10 degree phase window
6. Maximum No. of pulses − angle − Q— for 30 degree phase window
7. Maximum No. of pulses − angle − Q— for 10 degree phase window
8. Maximum Q − Minimum Q − angle − No. of pulses— for 30 degree phase window

4.6. Framing the network

ART 2 network includes additional preprocessing systems, which carry out contrast amplification and a noise suppression that enhance the stability of pattern recognition. The classification and efficiency of the network for the inputs is listed below:

The ART network is framed in such a manner that the number of neurons in F1 layer and F2 layer are selected according to the following considerations. The number of data that represents one exemplar pattern (ex: -void) is the number of neurons in the F1 layer and the total number of patterns used for training constitutes the number of neurons in F2 layer. The weights are initialized and the processing ensures where the network tries to learn the distinctive features of the input for each pattern and during testing an unknown pattern is given as input where the network classifies it to the kind of defect. Such a process is carried out for all the different forms of input and the efficiency of the network is justified by the misclassification done by the network. The observations are listed in tabular column and shown below.

4.7. Inferences

It is observed that the classification task was accurate and the speed was fast since it is a non-iterative procedure. The value of the vigilance parameter in the network measures the degree to which the system discriminates between different classes of input patterns. Also the value of the vigilance parameter (ρ) determines the granularity with which input patterns are classified by the network. For a given set of patterns to be classified, a large value of ρ will result in finer discrimination between classes than with a smaller value of ρ.

4.8. Statistical operators

PD phenomena fit into the category of a stochastic process because their physical properties are desirable in terms of a set of time-dependent random variables (Van Brunt, 1991). The PD phenomena that occur in dielectric media are inherently complex stochastic processes that exhibit significant statistical variability in such properties as pulse amplitude; shape and time of occurrence. Statistical measures such as measures of central tendency that include mean, mode, median, standard deviation and variance and measures of dispersion which comprises range, mean deviation, quartile deviation are extracted from the PD data. The ART-2 network is trained with these forms of inputs and recognition capabilities are observed.

4.9. Inputs to the network

Measures based on maximum and minimum values are fed as inputs to the network and the observations are tabulated in Table 2. The available statistical measurement vector of the pattern is divided into two sets for training. The first four patterns in all cases viz. void 1–4, corona 1–4, surface 1–4, oilcorona 1–4 and noise 1–4, forms the first set and the remaining patterns in all the cases as the second set. The network is trained with the sets and tested for the remaining paterns and observed for results. The observations are tabulated in Table 2 and shown below.

It is inferred that the ART network is able to classify the patterns more efficiently with the 'measures of maximum and minimum values' as compared to the other modes of input. This seems to provide good classification of PD patterns. Whatever may be type of input mode to the network, the number of misclassifications is less when the number of training exemplar is more or at least 75% of the total exemplars

5. Training and performance evaluation

The performance of a trained network after training has been completed is to be evaluated. There are several methods adopted for the validation of the network and the input viz. Partial Set Training, Hold One-Out Training and Pathology Analysis. The Hold One-Out and the partial set training method is being adopted to evaluate the same.

5.1. Hold-One-Out Training

A technique used to measure the effectiveness of network training is called appropriately enough, the hold-one-out technique. In this approach, one exemplar is extracted from the training set, the network is trained using the partial training set, and the withheld pattern is used to determine the effectiveness of the training. However unlike the partial-set training approach, this process is repeated for every exemplar in the training set. Thus for a training set containing n exemplars, the network must be trained n times.

The benefit of this approach is that, once the network can successfully recognize the withheld exemplar on all trials, the training set is complete, and then the network will be able to recognize the similarities in any new pattern. The problem with the approach unfortunately, is time. Repeating the entire training process n times, where n is the number of exemplars in the training set; will often makes this approach untenable.

This method is carried to validate the effectiveness of the network with 'measures of minimum and maximum' as inputs to the network and the observations are shown in Table 3. It is observed that there are some misclassifications while carrying out the Hold-One-Out. H is inferred that the method so it requires the input to be calibrated much more, so that the network can to classify the patterns much better than before.

The Hold-One-Out technique was applied to the network with statistical operators. However it was observed that, there are few misclassifications. In this manner the features or combinations of features in the input pattern that excited each unit can be identified. By performing such an inspection after training the network, the characteristics of the input that the network deemed important to successfully recognizing the exemplars is identified. This knowledge can be used further to refine the training set (Table 1).

Table 1
Observations on the measures of maximum and minimum values

Type of input	Phase window width	Number of pattern misclassified	$\eta\%$
Q max-N max-angle-26[a]	30	1	97
Q max-angle-$N-21$[a]	10	3	91
Q max-angle-N-26[a]	30	0	100
Q min-angle-$N-24$[a]	10	2	94
Q min-angle-N-26[a]	30	2	94
N max-angle-$Q-21$[a]	10	1	97
N max-angle-Q-26[a]	30	0	100
Q max-Q min angle-$N-21$[a]	30	3	91

η is the ratio of patterns classified to the total patterns presented expressed in percentage.
[a] Denotes the number of patterns used for training.

5.2. Partial set training

Apart from the Hold-One-Out technique, the effectiveness of the network is evaluated by presenting the network that were not a part of the training set. This is to evaluate how well the network can successfully learn from training exemplars. The evaluation of the network is done by presenting the network with various exemplar patterns which were not used during training, but which were similar to the training patterns. This technique is adopted for the 'measures based on central tendency and dispersion and their inferences are also shown in Table 2'.

5.3. Inference

It is found that the classification of the network is good and the level of misclassification is less. It makes clear that the data extracted on the above measures is more appropriate, so that

Table 2
Observations for statistical operators as input (partial set training)

Sl. No	Type of input	Phase window width	Number of training patterns	Pattern misclassified
1	Measures of dispersion (set-I)	30	20	1
2	Measures of dispersion (set-II)	30	23	Nil
3	Central tendency (set-I)	30	20	1
4	Central tendency (set-II)	30	23	Nil

Table 3
Observations for hold-one-out method

Sl. no	Type of input	Phase window width	Pattern misclassified
1	*Q* max-*N* max-angle	30	1
2	*Q* max-angle-*N*	10	3
3	*Q* max-angle-*N*	30	0
4	*Q* min-angle-*N*	10	2
5	*Q* min-angle-*N*	30	2
6	*N* max-angle-*Q*	10	1
7	*N* max-angle-*Q*	30	0
8	*Q* max-*Q* min angle-*N*	30	3

the network can learn the distinct features of the respective pattern more effectively and hence the classification efficiency is more (Tables 1 and 2).

The validation techniques carried out proves that the input is somewhat linear so that the network is able to recognize the underlying similarity of identical patterns superimposed on constant backgrounds having different levels.

The ART 2 network is able to classify the patterns of training very accurately but the limitation is that minimal variations in the fingerprints lead to categorization into a new pattern as per the ART 2 network configuration. This may find application in identifying and classification of continuous noises in on-line application for use in PD diagnostics

6. Conclusion

Since the case of PD pattern classification is one of classifying the pattern that is always not same or completely identical, it is important that the ANN is able to provide a reasonable generalization of such patterns characterizing the defect or flaw. Hence, first a sufficient number of training exemplars are required to describe the function. Obviously, there is no way a network can learn the variations in the function that are not reflected in the training exemplars. That is why, the future exemplars should be from the same probability distribution function. Secondly, if the number of training exemplars used is not sufficient to pin down the free parameters in the network to capture the regularity in the data, the best the network can do is assign some random component to some parameters. Thus, it is meaningless to evaluate the capability for generalization using exemplars that reflect variations not captured in the training exemplars.

It is observed that the ART 2 network has its own specialties and novelties and helps in identifying the needs of PD pattern classification in its own distinct means. It is however noted that further research and validation is essential before concrete conclusions are made. Further, this paper only deals with single defect types while in practical cases the PD patterns may involve multiple defects and including the effect of background noises, which thus necessitates further procedures.

References

Cachin, C., & Wiesman, H. J. (1995). PD recognition with knowledge based preprocessing and neural networks. *IEEE Transactions on Dielectrics and Electrical Insulation*, *2*(4), 578–589.

Candela, R., Mirelli, G., & Schifani, R. (2000). PD recognition by means of statistical and fractal parameters and a neural network. *IEEE Transactions on Dielectrics and Electrical Insulation*, *7*(1), 87–94.

Gulski, E., & Kreuger, F. H. (1992). Computer aided recognition of partial discharges. *IEEE Transactions on Electrical Insulation*, *27*(1), 82–92.

Gulski, E., & Krivida, A. (1993). Neural network as a tool for recognition of partial discharges. *IEEE Transactions on Electrical Insulation*, *28*(6), 984–1001.

Hong, T., Fang, M. T. C., & Hilder, D. (1996). PD classification by a modular network based on task decomposition. *IEEE Transactions on Dielectrics and Electrical Insulation*, *3*(2), 207–212.

Hoof, M., Freisleben, B., & Patsch, R. (1997). PD source identification with novel discharge parameters using counter propagation neural networks. *IEEE Transactions on Dielectrics and Electrical Insulation*, *4*(1), 17–32.

James, R. E., & Phung, B. T. (1995). Development of computer-based measurements and their application to PD pattern analysis. *IEEE Transactions on Dielectrics and Electrical Insulation*, *2*(5), 838–856.

Kai, Gao, Kexiong, Tan, Fuqi, Li, & Chengqi, Wu (2002). PD pattern recognition for stator bar models with six kinds of characteristics vector using BP network. *IEEE Transactions on Dielectrics and Electrical Insulation*, *9*(3), 381–389.

Kranz, Hans-Gerd (1993). Diagnosis of partial discharge signals using neural network and minimum distance classification. *IEEE Transactions on Electrical Insulation*, *28*(6), 1016–1024.

Kranz, H. (2000). Fundamentals in computer aided PD processing, PD pattern recognition and automated diagnosis in GIS. *IEEE Transactions on Dielectrics and Electrical Insulation*, *7*(1), 12–20.

Kranzand, Hans-Gerd, & Krump, Reiner (1992). Partial discharge diagnosis using statistical optimization on a PC based system. *IEEE Transactions on Electrical Insulation*, *27*(1), 93–105.

Krivda, A. (1995). Automated recognition of partial discharges. *IEEE Transactions on Dielectrics and Electrical Insulation*, *2*(5), 796–821.

Krivda, A., Gulski, E., Satish, L., & Zaengl, W. S. (1995). The use of fractal charts for recognition of 3-D discharge patterns. *IEEE Transactions on Dielectrics and Electrical Insulation*, *2*(5), 889–905.

Mazroua, Amira. A., & Salama, M. M. A. (1993). PD recognition with neural network using the multilayer perceptron technique. *IEEE Transactions on Electrical Insulation*, *28*(6), 1082–1087.

Mazroua, A. A., Bartnikas, R., & Salama, M. M. A. (1994). Discrimination between PD pulse shapes using different neural network paradigms. *IEEE Transactions on Dielectrics and Electrical Insulation*, *1*(6), 1119–1131.

Meijer, S., Gulski, E., & Smit, J. J. (1998). Pattern analysis of partial discharges in SF_6 GIS. *IEEE Transactions on Dielectrics and Electrical Insulation*, *5*(6), 830–842.

Pedersen, A., Crichton, G. C., & McAllister, I. W. (1991). The theory and measurements of partial discharge transients. *IEEE Transactions on Electrical Insulation*, *26*(3), 487–497.

Salama, M. M. A. (2002). Determination of neural-network topology for partial discharge pulse pattern recognition. *IEEE Transactions on Neural Networks*, *13*(2), 763–808.

Satish, L., & Zaengl, W. S. (1994). Artificial neural networks for recognition of 3-D partial discharge patterns. *IEEE Transactions on Dielectrics and Electrical Insulation*, *1*(2), 265–275.

Van Brunt, R. J. (1991). Stochastic properties of partial discharge phenomenon. *IEEE Transactions on Electrical Insulation*, *26*(5), 902–948.