

Partial discharge pattern classification using multilayer neural networks

L. Satish
B.I. Gururaj

Indexing terms: Neural networks, Multilayer. Partial discharge, Recognition process

Abstract: Partial discharge measurement is an important means of assessing the condition and integrity of insulation systems in high voltage power apparatus. Commercially available partial discharge detectors display them as patterns by an elliptic time base. Over the years, experts have been interpreting and recognising the nature and cause of partial discharges by studying these patterns. A way to automate this process is reported by using the partial discharge patterns as input to a multilayer neural network with two hidden layers. The patterns are complex and can be further complicated by interference. Therefore the recognition process appropriately qualifies as a challenging neural network task. The simulation results, and those obtained when tested with actual patterns, indicate the suitability of neural nets for real world applications in this emerging domain. Some limitations of this method are also mentioned.

1 Introduction

Electrical insulation plays an important role in any high voltage power apparatus. Partial discharges (PDs) occur when the local electric field exceeds the threshold value and results in a partial breakdown of the surrounding medium. Its cumulative effect leads to degradation of the insulation. PDs are initiated by the presence of defects in manufacture or choice of higher stresses dictated by design considerations. Measurements are made to detect these PDs and monitor the soundness of insulation during service life of the apparatus. PDs manifest as sharp current pulses at the terminals, and its nature depends on the type of insulation, defect present and measuring circuit and detector used [1].

Commercially available PD detectors generally display PD pulses on an elliptic time base. These detectors are widely used in manufacturing, research and testing establishments. The phase position and spread of the pulses depend on the nature of PD source and thus provide important features for interpretation. These patterns are, at times, complex depending on the type of defect producing PD, and can further be complicated owing to interference. A few typical patterns are shown in Fig. 1.

A systematic presentation of different types of PD display patterns, along with explanations, were first reported by CIGRE Working Group 21.03 in 1969 [2] and recently, in 1988, by Natrass [3]; these form the knowledge base in this domain. Over the years, experts have used this pictorial knowledge base and associated patterns obtained during PD tests to the ones which yielded the best match to arrive at conclusions about the behaviour of the insulation system under test.

Availability of personal computers and expert system building tools brings up possibilities of automating this process. A study of the literature shows, however, that no attempt has been made to use this pictorial knowledge base for PD interpretation. In recent years, pulse height analysis data (pulse height versus pulse phase and pulse count versus pulse phase) along with related statistics, obtained during PD tests in laboratories, have been used in conjunction with expert systems [4]. This approach builds up a statistical database for PD interpretation. Full success was reported for only simple cases (point and surface discharges). Gulski and Krueger [5], instead of using expert systems, stored templates comprising statistical parameters relating physical defect models (eight in number), and compared them with those obtained in further tests. However, results given by them for two cases clearly show the low discrimination of the method.

Thus, there have been no attempts to codify all the available pictorial knowledge so as to consider all the PD patterns for classification. One such effort was reported by the authors to classify most of the PD patterns using an expert system based pattern recognition approach [6]. Lack of a learning phase, and a need to manually extract and codify rules, were its drawbacks. Hence newer approaches were studied.

Multilayer neural networks (NNs) with back propagation algorithms have been successfully used in various applications as pattern classifiers (Speech, Sonar, EEG, ECG, etc.) [7, 8]. The advantage of NN over other classifiers is its ability to learn from examples, thereby avoiding explicit coding of knowledge as rules, as in rule based systems. Knowledge in the training set is extracted and stored in the connection weights and neuron biases during the learning phase. Such an implicit rule generation mechanism enables use of real world signals or patterns for on-line training, which is difficult to implement with rule based systems when considered alone. However, Gallant [9] presents arguments regarding possibilities for

circumventing this problem by the combined use of neural nets and expert systems, which he calls 'connectionist expert systems'. Secondly, the PD pulses and the resulting patterns often change, being functions of voltage, duration of its application, nature of defect, amount of ageing etc. Hence patterns only resemble each

other broadly, even though they are due to the same cause. Lastly, the ability of NN to generalise and be invariant to some transformations after training are some of the highly desirable features which justify its use in the current problem.

During the latter half of 1991, a few researchers have also used NN for PD recognition. Murase *et al.* [10] used real world signals from artificially created defects in insulation models as inputs. They employed NN with one hidden layer, and reported success for three types of patterns (protrusions on high voltage conductors, grounding enclosures and voids in insulators). Hiroshi and Takeshi [11] and Hozumi [12] describe NN based systems, used primarily as PD alarm processors for specific applications such as PD in XLPE cables, and tree initiation in needle shaped voids in epoxy resin, respectively. Therefore it emerges that, apart from Murase *et al.*, no researcher seems to have reported on the subject of this paper.

This paper reports simulation results of a system capable of automatically classifying the different PD patterns including those imposed by interference using a feedforward NN with two hidden layers. The displayed patterns are captured with a CCD camera, processed to extract the PD signal to form the input. Alternatively, PD signals can be digitised and used as input. The network output indicates the class to which the input belongs. A brief summary of results for actual PD patterns is also given.

2 Network topology and training set

2.1 Choice of NN structure

Essentially, NN evolve a nonlinear mapping between the input and output in the training phase during which the weights and biases are iteratively updated using the back propagation algorithm [8]. In this supervised learning, input-output pattern pairs are repeatedly fed until the squared error over the entire training set is a minimum. The factors influencing convergence are

- (i) complexity of the decision regions to be formed
- (ii) network topology and size
- (iii) initial choice of weights and thresholds
- (iv) values of learning and momentum rates, δ and a
- (v) input pattern presentation order.

A single layer NN can classify only linearly separable signals, and is inadequate to achieve complex mappings as required in the present problem. This limitation is overcome by introducing additional hidden layers between input and output. The number of hidden layers, and the nodes they must contain to achieve a desired input-output mapping, depends on the complexity of the input pattern space to be partitioned. Lippmann [7] has demonstrated that, with three layers (two hidden and one output), any complex decision region can be generated. For simple problems, like XOR, the number of hidden layer nodes required for obtaining convergence with near certainty is related to the number of input presentations [8]. For NN with one hidden layer, bounds and conditions on the number of hidden layer nodes have been derived [13-15]. It is shown that any arbitrary training set with ' p ' patterns can be 'exactly' implemented with $(p - 1)$ hidden layer nodes. Sartori and Antsaklis [14] caution that this bound is sure to implement the training set 'exactly' but may not be the best way to ensure correct classification (which is most essential in any classifier). In the absence of such a theory for NN with more than one hidden layer, it seems logical to use results

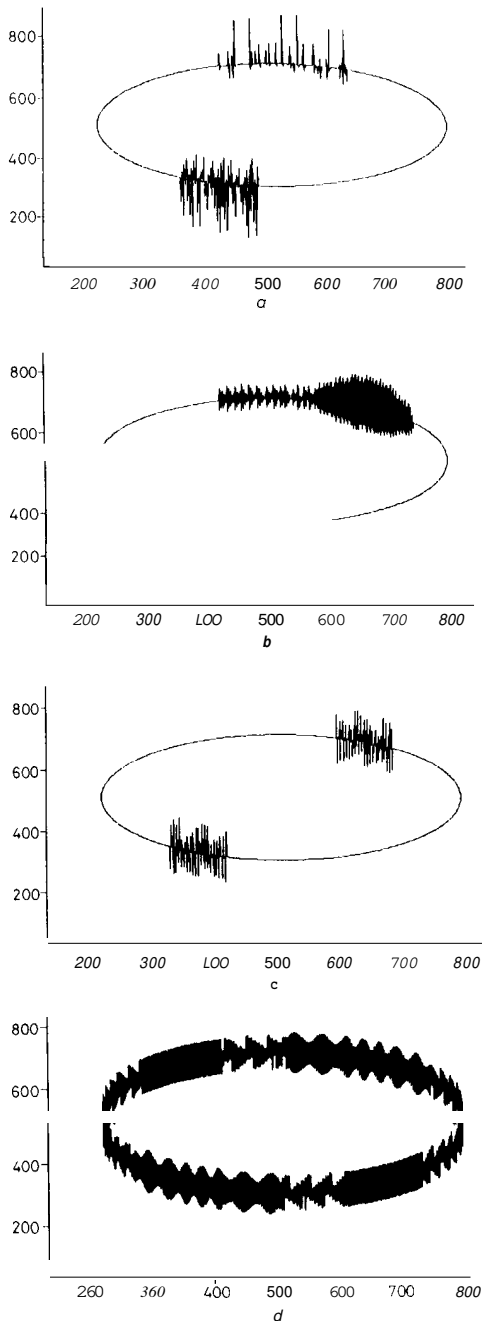


Fig. 1 Sample PD and interference patterns

a PD in dielectric bounded cavity
 b PD due to gas bubbles in insulating liquid in contact with moist cellulose
 c and d interferences due to asynchronous machine and radio station emission

presently available along with experience in the problem domain to estimate the number of hidden layer nodes required.

At the time of writing this paper, the authors came across the conditions on bounds for the number of samples needed for neural learning [16]. However, for this method to be implemented, it is necessary to ensure sufficient number of 'boundary samples' (i.e. those close to the cluster boundary). This involves preprocessing of the training set to compute the number of clusters and selection of those close to its boundary. These computations, being large and involved, have not been attempted.

The NN has the capacity to 'memorise' and 'generalise' to some extent. It is said to generalise when the input-output relation is nearly correct for stimuli never encountered during training. Highly layered networks tend to generalise better [17, 18], owing to robust internal representation and are more fault-tolerant due to wider spread of mapping information. PD phenomena are statistical in nature and hence the phase position and individual pulse amplitudes vary within some limits. NN must be made invariant to these variations. Among the methods cited [19], invariance is achieved by including more than one example per class in the training set. This choice is justified as each class has many exemplar patterns, and inclusion of as many as possible (limited by training time, memory requirements etc.), will result in better exposure of NN to different inputs, and hence develop far greater ability for generalisation.

2.2 Formation of training set

A study of the pictorial knowledge base, and accompanying literature, resulted in identification of 21 key patterns. Of them, seven are due to interferences only. These patterns are often present with no test voltage applied and sometimes rotate continuously owing to their asynchronism with the time base. These distinguishing features are used by experts to identify and minimise them with suitable line filters and shielded test enclosures. If all the patterns are considered together, there is a high likelihood of the pattern features getting mixed up leading to convergence problems. Also, information about stationarity of the pattern cannot be conveyed in a single capture of the image, and comparison of multiple raw video images is cumbersome. Therefore they have to be considered separately. Hereinafter, these will be referred to as Categories A (14 in number) and B (7 in number), respectively. Table 1 contains their descriptions. Fig. 2 shows the logic used for discrimination.

The pattern generating program also takes into account the possible random variations of pulse amplitudes and pulse phase positions. These ranges were derived after studying the literature on the pictorial knowledge base. The details of PD generation, display

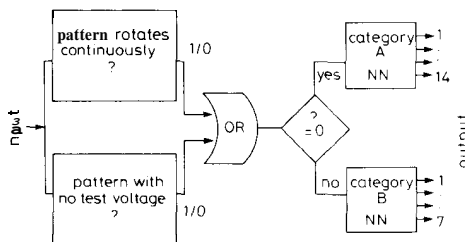


Fig. 2 Logic to classify input into Categories A or B

and signal extraction are given in Appendix 7. The positive part of the extracted signal is averaged every 10° (electrical) and then scaled to yield 36 continuous valued samples. The amount of information lost by not including the negative part of the PD signal in the training set is minimal, as in most cases, the positive and negative parts are symmetrical. If desired, another NN can be used for it, and the resulting output vectors are to be logically combined to obtain the final output. The examples belonging to a particular class, if more than one, are all put together in the training set.

Table 1: Description of PDs of categories A and B

| Category A | | |
|------------|--|-------------------------------|
| Class | Description of PD | Fig. no. in Natrass paper [3] |
| 1 | PD in solid dielectric bounded cavity | 5 |
| 2 | Gas bubbles in insulating liquid in contact with moist cellulose (e.g. oil impregnated paper) | 15 |
| 3 | Cavity between metal on one side and dielectric on the other | 17 |
| 4 | PD in number of cavities of various shapes, or on external dielectric surface, or high tangential stress | 9 |
| 5 | PD at fissures in elastomeric insulation in the direction of the field | 7 |
| 6 | Active growth of a carbon track in organic material | 24 |
| 7 | Sharp point in an insulating liquid | 42 |
| 8 | Sharp point on a high voltage conductor | 40 |
| 9 | Sharp point on grounded enclosures | — |
| 10 | Conducting particles formed in voids in cast resin insulation | 11 |
| 11 | Laminar cavities in machine insulation | 13 |
| 12 | Surface discharges between external metal and dielectric surfaces | 19 |
| 13 | Tracking of contaminated organic insulation originating from a cavity | 26 |
| 14 | Contact noise | 38 |
| Category B | | |
| Class | Description of interference | Fig. no. in Natrass paper [3] |
| 1 | Six pulse thyristor or mercury arc rectifier | 44 |
| 2 | Asynchronous machine | 46 |
| 3 | Fluorescent lamps | 45 |
| 4 | High frequency power amplifier or oscillator | 41 |
| 5 | Radio broadcast | 48 |
| 6 | Amplifier noise | — |
| 7 | Induction heater and ultrasonic generators | 49 |

2.3 Program validation

The program for back propagation was tested with the XOR problem using two input nodes, two nodes in first and second hidden layers and one output node (2-2-2-1). Though this structure is superfluous, it helped study of the network behaviour with respect to various parameters. With the tolerance fixed at 0.1%, the network converged in 21 out of 30 trials within 2000–4000 sweeps for δ values of 0.9 and 0.8. When tested with $\delta = 0.3$, it took 11000–15000 sweeps to converge (26 out of 30 trials), but it failed on all 25 trials except one with $\delta = 0.1$, even after 32500 sweeps, taking about 260 s of CPU time on a 386 running at 16 MHz (error being computed every 250 sweeps).

Initial connection weights and neuron biases were chosen randomly in the range: 0.0–0.5, and a value of 0.9 was used throughout. Weights and biases are updated

after presentation of each pattern. Neurons in all the layers are fully connected. One sweep implies one iteration over the entire training set. The error is computed as the total squared difference between the desired and actual outputs over the entire training set divided by the total number of patterns.

3 Results and discussions

3.1 Graphical internal representation

Efforts have been made by various workers to explain how the information is internally coded. Pao and Sobajic [20] propose 'that, in NN computing, the very processing itself encodes knowledge'. Other explanations [18, 21] are based on strong and weak correlations formed by computing the angle between n-dimensional vectors comprising the weight matrices. In this paper, another depiction is presented.

The connections between active neurons in the three layers of a converged network for two different classes of input are given in Figs. 3a-d. The active neuron is defined as one whose output is greater than 0.9 if in the first and second hidden layer, and greater than 0.8 if in the output layer. Initially, all the active neurons are found in the three layers for the given input. Only those connections between active neurons are considered for which

$$w(i, j) * \text{neuron}(i) + \text{bias}(j) > 3.0$$

where $w(i, j)$ is the weight between active neurons i and j , $\text{neuron}(i)$ is the output of active neuron i , and $\text{bias}(j)$ is the threshold of active neuron j . A value of 3.0 for the above limit is chosen because, then, the sigmoid function

has a value of 0.95. The active neurons and pathways shown in Figs. 3a-d are the major contributors for yielding a 'high' at the appropriate output node. The rest of the connections (not shown) are also equally important because they have to interfere cooperatively to produce a 'low' at the rest of the output nodes. The figures also illustrate the way in which the input-output mapping information is realised and the generalisation achieved.

3.2 Results for Category A patterns

Case 1: Initially, four distinct classes among Category A were chosen with the aim of probing the program further. Their input signals had the following characteristics

- (i) a pulse group at each voltage zero
- (ii) one pulse group at the positive voltage peak
- (iii) one pulse group at the negative voltage peak
- (iv) a pulse group at each voltage peak.

As the inputs were distinct, the NN converged on all trials.

Case 2: Here, with all classes included, various δ values (0.95, 0.9, 0.8, 0.75, 0.6, 0.45, 0.25) were tried which led to failures on all trials. On examination of the pertinent learning curves, it was found that they were very slow in learning or oscillatory. Values of δ of 0.6 and 0.45 resulted in reduced oscillations, and showed only signs of convergence. In these trials, 4000 sweeps took about 180 minutes (error being computed every 25 sweeps). A tolerance of 0.5% is used in all these cases.

Case 3: Attributing these to overfitting effects, a reduced network with 6 chosen as 0.6 and 0.5 were tried, and it

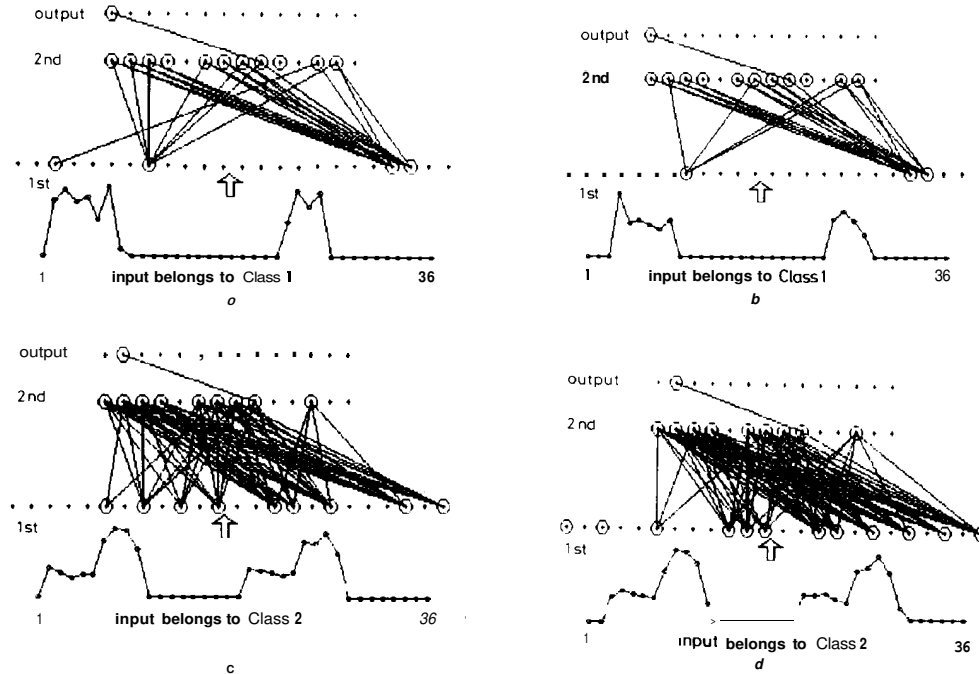


Fig. 3 Depiction of generalisation in NN after learning in Case 4

(five examples per class, Table 2)

Note consistency of output even with realistic changes in amplitude and phase in b and d wrt a and c, respectively

+ neuron ○ active neuron

Table 2: Details of training for Category A patterns

| Case | Number of classes | Structure of NN used | Examples per class | δ | Number of trials | Number of successes | Number of sweeps* |
|------|-------------------|----------------------|--------------------|----------|------------------|---------------------|-------------------|
| 1 | 4 | 36-12-8-4 | 1 | 0.9 | 16 | 16 | 1000 |
| 2 | 14 | 36-30-22-14 | 1 | 0.95 | 7 | 0 | # |
| | | | | to 0.25 | | | |
| 3 | 14 | 36-24-14-14 | 1 | 0.6 | 6 | 4 | 3500 |
| | 14 | 36-24-14-14 | 1 | 0.5 | 9 | 6 | 3750 |
| 4 | 12 | 36-24-12-12 | 2 | 0.6 | 9 | 7 | 2500 |
| | 14 | 36-24-14-14 | 5 | 0.6 | 9 | 6 | 780 |

* results give average of 10 trials
 # failed even after 4000 sweeps in each trial
 - 1 trial only, as first

yielded repeatable results. When tested for generalisation, it was found that all the converged NNs had implemented the training set, but invariance to realistic changes in phase and amplitude were just satisfactory which called for further improvement.

Case 4: This improvement was achieved by including initially two examples per class, and then increasing them to five examples per class. In addition to this, it resulted in a reduced number of sweeps to obtain convergence. Figs. 4a and b show sample learning curves depicting them. Table 2 summarises these results.

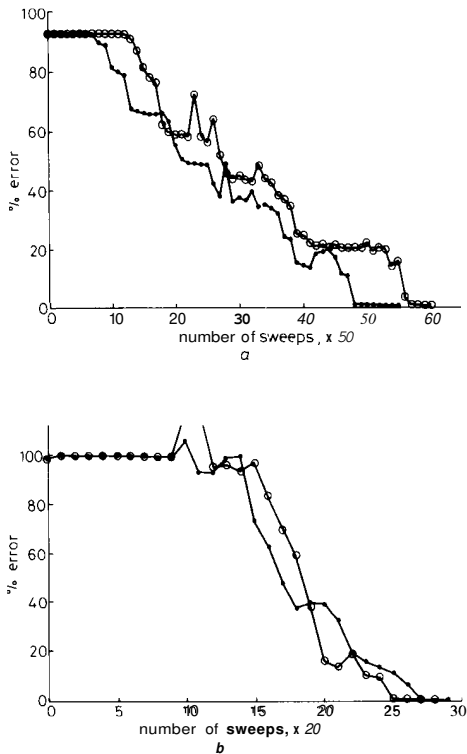


Fig. 4 Sample learning curves
 Note reduction in number of sweeps for convergence
 a Case 4, 2 examples/class 2750 sweeps ●-●-●
 3000 sweeps ○-○-○
 b Case 4, 5 examples/class 580 sweeps ●-●-●
 560 sweeps ○-○-○

Extensive recognition tests were carried out on four converged NNs (Category A) to check their generalisation abilities. It involved feeding inputs different from those used during training. Table 3 summarises these results. The overall recognition rate lies in the range of 83 to 88%.

On 9 out of 30 trials with $\delta = 0.9$ and 4 out of 30 trials with $\delta = 0.3$, in the XOR problem, plateaux were encountered. Most of the connection weights and biases of the network, after failure, had attained high values. These high values cause saturation of the activation function, resulting in back propagating a very small error, and hence causing learning to come to a standstill. The algorithm failed, even with an increased iteration count (25 000). A similar condition was observed on one occasion (in nine trials) in Case 4. Both these learning curves are shown in Fig. 5. Watrous [22] has drawn attention to such plateau regions. A very similar situation is reported by Wasserman [23] which he calls 'paralysis'.

Table 3: Recognition results for Categories A and B

| Category A | | | | | |
|--------------|-------------------------|-----------------------------------|------------|------------|------------|
| Class | Number of test patterns | Number of correct classifications | | | |
| | | NN-1 | NN-2 | NN-3 | NN-4 |
| 1 | 86 | 69 | 77 | 77 | 67 |
| 2 | 71 | 60 | 51 | 54 | 55 |
| 3 | 83 | 76 | 68 | 69 | 72 |
| 4 | 91 | 73 | 86 | 87 | 79 |
| 5 | 77 | 71 | 57 | 65 | 69 |
| 6 | 82 | 61 | 64 | 70 | 60 |
| 7 | 75 | 64 | 71 | 61 | 73 |
| 8 | 45 | 43 | 45 | 38 | 45 |
| 9 | 50 | 50 | 41 | 50 | 49 |
| 10 | 58 | 52 | 38 | 13 | 31 |
| 11 | 62 | 58 | 62 | 49 | 61 |
| 12 | 59 | 58 | 58 | 59 | 56 |
| 13 | 56 | 50 | 50 | 39 | 40 |
| 14 | 58 | 56 | 55 | 56 | 57 |
| Total | 953 | 88% | 86% | 83% | 85% |

| Category B | | | | | |
|--------------|-------------------------|-----------------------------------|------------|------------|------------|
| Class | Number of test patterns | Number of correct classifications | | | |
| | | NN-5 | NN-6 | NN-7 | NN-8 |
| 1 | 55 | 34 | 36 | 48 | 44 |
| 2 | 58 | 53 | 52 | 52 | 52 |
| 3 | 55 | 46 | 40 | 50 | 50 |
| 4 | 40 | 37 | 13 | 26 | 33 |
| 5 | 60 | 50 | 51 | 59 | 60 |
| 6 | 55 | 21 | 50 | 53 | 32 |
| 7 | 30 | 30 | 30 | 30 | 30 |
| Total | 353 | 77% | 77% | 90% | 85% |

This is a drawback of NN; suggestions or methods for overcoming such situations are, unfortunately, not given in the literature.

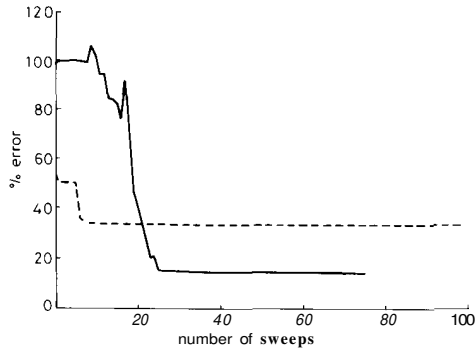


Fig. 5 Plateaux obtained during learning in XOR problem and in Case 4

(Five examples per class, Table 2)
 - - - 25000 sweeps ($\times 250$) XOR
 — 1500 sweeps ($\times 20$) Case 4

3.3 Results for Category B patterns

Out of the seven patterns, three are quite similar (interferences due to radio stations, amplifier noise and high frequency sources). To correctly differentiate among them, the samples have to be large enough to reflect their characteristics distinctly, since averaging smoothens out finer details. Hence 90 sample points were taken for these patterns (averaging the extracted signal every 4" (electrical)).

The network structure of 90-16-8-7 with five examples per class was considered (in all, 35 patterns). δ values of 0.6, 0.4, and 0.3 were tried, resulting again in highly oscillatory error curves. These reduced with a decrease in δ value. Repeatable results were obtained with $\delta = 0.25$, converging on all five trials, with an average sweep of about 1800. Tests for invariance showed good results. To improve these further, eight examples per class was considered with the same topology (in all, 56 patterns), but repeatable results were obtained only with a further reduced δ value of 0.15, converging in five out of six trials, with an average sweep of about 1400.

Recognition tests were done on four converged NNs to ascertain their generalisation capability. Table 3 contains these results. As can be seen, NNs 7 and 8 yield better results indicating superior generalisation. This clearly indicates that not all converged networks generalise to the same extent. It is believed that the networks might have reached different optimal points on the error surface, owing to different starting points with all implementing the training set to the same extent (tolerance value). Lack of an index or norm to define and quantify the amount of generalisation, makes such generalisation tests necessary to ascertain this aspect. Such tests were done and Table 3 gives the results.

4 Recognition of actual PD patterns

Even though the NNs had performed satisfactorily in the generalisation tests, since they were trained with synthetic patterns, their performance when fed with actual PD patterns had to be ascertained. So five types of Category A, and 3 types of Category B patterns, were obtained and displayed on the PD detector screen in the conventional

way. They comprised both complex and simple patterns. Recognition tests with these actual patterns yielded an overall recognition rate of 84% (NN-3) for Category A patterns and 79% (NN-7) for Category B patterns, for the best performing NNs. Further details can be found [24]. Fig. 6 shows the processed camera output for PD due to surface discharge on a cable specimen, along with the extracted pattern. The increased levels of noise and pulsive interferences (not trained for) in the actual patterns, as is evident in Fig. 6, may be the reason for the lower levels of performance of the NNs.

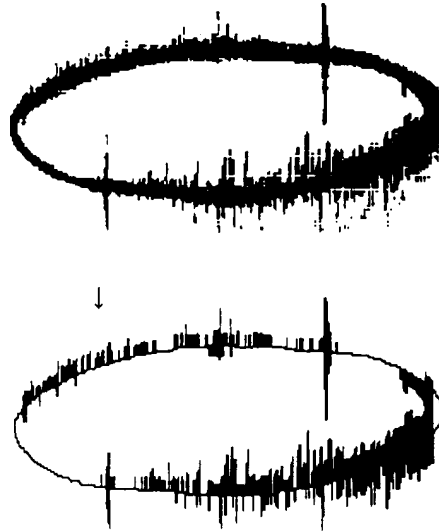


Fig. 6 Sample of processed camera output and extracted pattern for actual PD due to surface discharge on cable

Two large and symmetrical pulsive interference present during measurements have also been captured

5 Conclusions and future prospects

The success obtained in classifying both synthetic and some actual PD patterns using NN demonstrates the utility of this approach and, in particular, its structure with two hidden layers. It is believed that the network parameters arrived at will be of help to researchers proceeding to use NN for this application. An element of empiricism, however, exists in the selection of required constants owing to lack of a theoretical foundation for NN with more than one hidden layer. Perhaps these issues will be addressed by NN experts in years to come.

In an interesting approach published very recently by Low, Lui *et al.* [25], use of a connectionist expert system with adaptive learning abilities has been proposed. This system's incremental learning ability is very useful when knowledge about newer examples are to be included without resorting to a fresh training cycle. Utility of such systems in this domain is yet to be explored fully.

Of the other alternatives being examined by the authors, hidden Markov models (HMMs) emerge as the most appropriate. As HMM has been very successfully applied in speech recognition and, because the shape and characteristics of the PD signals closely resemble speech signals, it is expected to perform well in this domain too. Preliminary results for a 4 class PD problem (Case 1, Table 2) has yielded encouraging results. Full details will be reported separately.

6 References

- 1 KREUGER, F.H.: 'Partial discharge detection in high-voltage equipments' (Butterworths & Co. Ltd., London, 1989)
- 2 CIGRE WORKING GROUP 21.03: 'Recognition of discharges', *Electra*, 1969, 11, pp. 61-98
- 3 NATTRASS, D.A.: 'Partial discharge measurement and interpretation', *IEEE Elect. Insul. Mag.*, 1988, 4, (3), pp. 10-23
- 4 GASSAWAY, J.D., JACOB, P.B., VASSILIADIS, C.A., and REYNOLDS, P.H.: 'Computer-aided partial discharge measurement and recognition'. 5th International Symposium on *High Voltage Engineering*, Braunschweig, 1987, Paper 41.03
- 5 GULSKI, E., and KREUGER, F.H.: 'Computer-aided recognition of discharge patterns'. 7th International Symposium on *High Voltage Engineering*, Dresden, 1991, Paper 71.01
- 6 SATISH, L., and GURURAJ, B.I.: 'Application of expert system to partial discharge pattern recognition'. CIGRE Study Committee 33 Colloquium, Leningrad, 1991, Paper CIGRE SC 33.91 (Coll) 1.16 IWD
- 7 LIPPMANN, R.P.: 'An introduction to computing with neural nets', *IEEE ASSP Mag.*, 1987, pp. 4-22
- 8 RUMELHART, D.E., and MCCLELLAND, J.L.: (Eds.): 'Parallel distributed processing: explorations in the microstructure of cognition' (MIT Press, 1986), vol. 1
- 9 GALLANT, S.I.: 'Connectionist expert systems', *Comm. ACM*, 1988, 31, (2), pp. 152-169
- 10 MURASE, H., OYAMA, M., HANAI, E., OHSHIMA, I., and SAKAKIBARA, T.: 'Application of neural networks to partial discharge pattern recognition'. CIGRE Study Committee 33 Colloquium, Leningrad, 1991, Paper CIGRE SC 33.91 (Coll) 1.7 IWD
- 11 HIROSHI, S., and TAKESHI, E.: 'Pattern recognition of partial discharge in XLPE cables using a neural network'. Proceedings of the 3rd International Conference on *Properties and Applications of Dielectric Materials*, Tokyo, 1991, pp. 43-46
- 12 HOZUMI, N., OKAMOTO, T., and IMAJO, T.: 'Discrimination of partial discharge patterns using neural network'. Proceedings of the 3rd International Conference on *Properties and Applications of Dielectric Materials*, Tokyo, 1991, pp. 47-50
- 13 HUANG, S.C., and HUANG, Y.F.: 'Bounds on the number of hidden neurons in multilayer perceptrons', *IEEE Trans.*, 1991, NN-2, (1), pp. 47-55
- 14 SARTORI, M.A., and ANTSAKLIS, P.J.: 'A simple method to derive bounds on the size and to train multilayer neural networks', *IEEE Trans.*, 1991, NN-2, (4), pp. 467-471
- 15 MIRCHANDANI, G., and CAO, W.: 'On hidden nodes for neural nets', *IEEE Trans.*, 1989, CS-36, (5), pp. 661-664
- 16 MEHROTRA, K.G., MOHAN, C.K., and RANKA, S.: 'Bounds on the number of samples needed for neural learning', *IEEE Trans.*, 1991, NN-2, (6), pp. 548-558
- 17 WIELAND, A., and LEIGHTON, R.: 'Geometric analysis of neural network capabilities'. IEEE International Conference on *Neural Networks*, 1987, 3, pp. 385-392
- 18 TILLERY, S.I., and COMBS, N.: 'The problem of internal representations in neural nets: concepts, implications, and a proposed metric'. IEEE International Conference on *Neural Networks*, 1987, 2, pp. 585-590
- 19 WIDROW, B., and LEHR, M.A.: '30 years of adaptive neural networks: perceptron, madaline, and backpropagation', *Proc. IEEE*, 1990, 78, (9), pp. 1415-1442
- 20 PAO, Y.H., and SOBAJIC, D.J.: 'Neural networks and knowledge engineering', *IEEE Trans.*, 1991, KDE-3, (2), pp. 185-192
- 21 MICHAELS, D.F.: 'On information representation in back-propagation classifier networks'. IEEE Conference on *Systems, Man, and Cybernetics*, 1990, pp. 41-45
- 22 WATROUS, R.L.: 'Learning algorithms for connectionist networks: applied gradient methods of nonlinear optimization'. IEEE International Conference on *Neural Networks*, 1987, 2, pp. 619-627
- 23 WASSERMAN, P.D.: 'Neural computing: theory and practice' (Van Nostrand Reinhold: New York, 1989)
- 24 SATISH, L.: 'Study of some artificial intelligence paradigms for partial discharge pattern recognition'. PhD thesis, Indian Institute of Science, Bangalore, 1992
- 25 LOW, B.T., LUI, H.C., TAN, A.H., and TEH, H.H.: 'Connectionist expert system with adaptive learning capability', *IEEE Trans.*, 1991, KDE-3, (2), pp. 200-207

7 Appendix

7.1 Generation and display of PD patterns

A simple and effective, though artificial, method to display any complicated PD pattern on the oscilloscope

has been developed. This avoids use of high voltage transformers and components and the need for specimens with specific defects for generating required patterns. The idea is to use a real time digitiser in the reverse way (i.e. the required signals are computed and transferred to its memory; these are then used to create the pattern on an oscilloscope). To generate any pattern, the x- and y-axis signals corresponding to one 50 Hz cycle (20 ms) have to be computed. The x-signal (sine wave) is computed and stored while the y-signal (cosine wave plus PD pulses) is computed each time depending on the type of pattern to be displayed. The two signals are transferred via GPIB to the digitiser as 2 k points at 10 μ s sampling rate and 10-bit resolution. Sample patterns shown in Fig. 1 have been generated in this manner.

7.2 Extraction of input signal

A digitising camera is used to capture the displayed image. The compressed raw video file is then unpacked and displayed on the PC screen after normalisation and thresholding. A spot on the oscilloscope as seen by the eye does not correspond to a single pixel on the CCD array, but to many pixels (8 to 10), depending on its intensity. Therefore all displayed traces have to be corrected for this dispersion. First, a reference ellipse is captured and displayed on the PC screen from which a single pixel wide reference ellipse is extracted in the following manner. A boundary tracing algorithm was developed

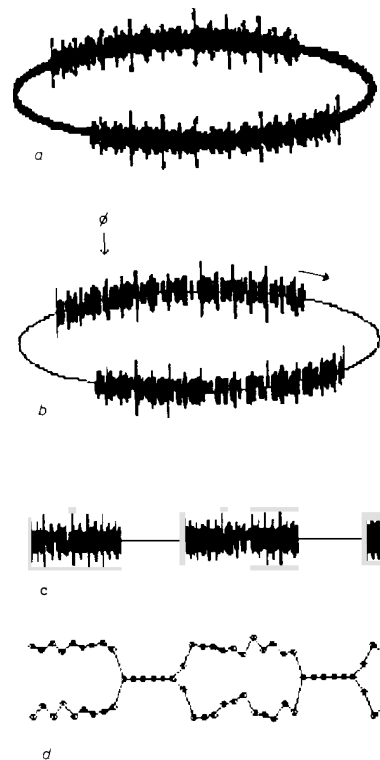


Fig. 7 Procedure for extracting signal patterns

- a processed camera output
- b extracted image pattern
- c signal pattern before averaging
- d scaled signal pattern 36 points long

and used to trace the inner and outer contours, from which the reference ellipse is computed and stored. Movable zero voltage markers on the elliptic trace are provided on the **PD** detectors for clarity in viewing patterns and hence made software selectable. The reference ellipse is then moved, if necessary, to coincide with the centre of the displayed pattern. This is to correct for any misalignment due to camera shake, shift of oscilloscope

trace etc. The pattern is then scanned vertically on either side along the reference ellipse, starting from the selected zero marker, in a clockwise direction to extract the **PD** pulses. The positive and negative portions are stored separately. As mentioned in Section 2.2, averaged versions of them form the input to **NN**. Fig. 7 illustrates a step-by-step procedure for extracting the input signal pattern.