# On-line Partial Discharge source identification in medium and high voltage cable networks

Demetres Evagorou

Department of Electrical and Computer Engineering

University of Cyprus

A thesis submitted for the degree of

*Doctor of Philosophy*

October 2011

**APPROVAL PAGE FOR THESIS DISSERTATION**

SUBMITTED IN FULFILMENT OF REQUIREMENTS FOR DEGREE OF DOCTOR OF PHILOSOPHY AT UNIVERSITY OF CYPRUS BY

**Demetres Evagorou**
Candidate

TITLE: **On-line Partial Discharge source identification in medium and high voltage cable networks**

APPROVED:

| | |
|---|---|
| **Prof. Charalambos Charalambous** | |
| Committee Chairperson | Signature |
| **Dr George E. Georghiou** | |
| Research Supervisor | Signature |
| **Dr Hans Edin** | |
| Committee Member | Signature |
| **Dr Andreas Kyprianou** | |
| Committee Member | Signature |
| **Dr George Mitsis** | |
| Committee Member | Signature |
| **Dr Andreas Starvrou** | |
| Committee Member | Signature |

DATE: **October 26, 2011**

# Abstract

On-line condition monitoring provides information about the existence of PDs in cables and their accessories under normal operating conditions and lies at the heart of a condition-monitoring program of power equipment. Identification of the source of a PD and the evolution of the measured quantity of PD activity over time are of paramount importance for the assessment of the insulation integrity of power equipment. In such a continuous monitoring system the feature vector has to have the minimum possible number of dimensions otherwise classification times can be very slow for real-time implementation.

The motivation behind this work, and hence its contribution is to propose and evaluate two complementary lower dimensional feature vectors based on time and phase-resolved measurements, which can be exploited in on-line scenarios. Towards this end the methodology that is followed is the creation of an experimental database of known PD sources, its analysis using statistical and signal processing methods, the proposition of the feature vectors, and their assessment using supervised, unsupervised, experimental and field data.

Firstly a database of experimental PDs of different sources is built, based on an experimental set-up in a high voltage laboratory under controlled conditions. Based on analysis of this database two feature vectors of lower dimensionality, one using time-resolved and another using phase-resolved measurements, for use in a continuous on-line system, are proposed. The tools employed in achieving a compact characterisation of different PD sources are the Wavelet Packets Transform and Higher Order Statistics. The WPT provides an adaptive time frequency decomposition of the PD signals, resulting in a compact representation, and specific nodes of the best basis tree are selected using a similar procedure for the two proposed features. The HOS are employed to derive statistical descriptors of the wavelet coefficients of the selected nodes, reducing even further the dimensionality.

While the proposition of these proposed features is one novelty of this work, their evaluation using different algorithms and data is another. Firstly they are evaluated using the experimental database and two supervised algorithms that utilise different training principles, the Support Vector Machines (SVM) and the Probabilistic Neural Network (PNN). The SVM is used to compare the proposed features to already existing works that employ the raw wavelet coefficients as features. Despite the slight reduction in classification accuracy the proposed features show comparable rates to existing solutions, taking into account the significant reduction in the dimensionality. The proposed feature vectors are then employed to directly compare the SVM and the PNN, which provide similar results, irrespective of the training method employed. In all the cases statistical analysis by the use of tolerance intervals is used to support the results.

Next the proposed time-resolved feature is evaluated using an unsupervised algorithm simulating on-line conditions where the use of supervised algorithms poses a problem as the training class labels and the number of PD sources are not a priori known. The use of unsupervised or clustering algorithms can provide this information, or even alleviate the need for training data. The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm using the time-resolved proposed feature vector is used on both experimental, and field data acquired from the network of the Electricity Authority of Cyprus. The results demonstrate the use of this feature in on-line field measurements, as a pre-processing step to Phase Representation Partial Discharge (PRPD) analysis.

# Abstract

Η παρακολούθηση για ύπαρξη Μερικών Ηλεκτρικών Εκκενώσεων (ΜΗΕ) σε πραγματικές συνθήκες προσφέρει σημαντικές πληροφορίες για την κατάσταση του διηλεκτρικού των υπογείων καλωδίων μεταφοράς και διανομής και συνεπώς βρίσκεται στην πρώτη γραμμή ενός προγράμματος συντήρησης του εξοπλισμού ισχύος. Η αναγνώριση της πηγής της ΜΗΕ και της εξέλιξης της τιμής της με το χρόνο είναι υψίστης σημασίας για την εκτίμηση της ακεραιότητας του διηλεκτρικού σε εξοπλισμό ισχύος. Προς επίτευξη τούτου χρειάζεται ένα συνεχές σύστημα παρακολούθησης, μια από τις ιδιότητες του οποίου πρέπει να είναι η ελαχιστοποίηση των διαστάσεων του χαρακτηριστικού διανύσματος, ούτως ώστε ο αλγόριθμος κατηγοριοποίησης να μην είναι πολύ αργός και άρα αδύνατο να χρησιμοποιηθεί σε ένα σύστημα πραγματικού χρόνου.

Αυτή η ερευνητική εργασία δημιούργησε μια πειραματική βάση δεδομένων βασισμένη σε μια σε εργαστηριακή διάταξη κάτω από ελεγχόμενες συνθήκες. Ακολούθως, με βάση την ανάλυση της βάσης δεδομένων, προτείνονται δύο χαρακτηριστικά διανύσματα μικρότερων διαστάσεων, ένα για δεδομένα με βάση το χρόνο και άλλο για δεδομένα με βάση την φάση, τα οποία έχουν την δυνατότητα να χρησιμοποιηθούν σε ένα συνεχές σύστημα παρακολούθησης πραγματικού χρόνου. Για να επιτευχθεί ο συμπαγής χαρακτηρισμός των διαφορετικών πηγών ΜΗΕ οι μέθοδοι ανάλυσης των πακέτων κυματιδίων και της ανώτερης τάξης στατιστικής χρησιμοποιούνται. Τα πακέτα κυματιδίων αποσυνθέτουν το σήμα σε μια χρονική και φασματική απεικόνιση, προσαρμοσμένη στα χαρακτηριστικά του σήματος επιτυγχάνοντας συμπαγέστερη απεικόνιση. Συγκεκριμένοι κόμβοι του δέντρου της καλύτερης βάσης επιλέγονται από μια παρόμοια διαδικασία για τα δυο προτεινόμενα χαρακτηριστικά διανύσματα. Κατόπιν μέσω της χρήσης της ανώτερης τάξης στατιστικής πραγματοποιείται στατιστικός χαρακτηρισμός των συντελεστών κυματιδίων των επιλεγμένων κόμβων, επιτυγχάνοντας περαιτέρω μείωση της διάστασης.

Τα προτεινόμενα χαρακτηριστικά διανύσματα αξιολογούνται από δύο εποπτικούς αλγόριθμους (τα πιθανολογικά νευρωνικά δίκτυα και τις μηχανές υποστήριξης

διανυσμάτων), οι οποίοι εκπαιδεύονται βασιζόμενοι πάνω σε δύο διαφορετικές αρχές, χρησιμοποιώντας την πειραματική βάση δεδομένων. Οι Μηχανές Υποστήριξης Διανυσμάτων (ΜΥΔ) χρησιμοποιούνται για να συγκριθούν τα προτεινόμενα χαρακτηριστικά διανύσματα με παρόμοιες εργασίες οι οποίες χρησιμοποιούν τους συντελεστές των κυματιδίων ως χαρακτηριστικά διανύσματα. Παρ όλη την μικρή μείωση στην ακρίβεια της κατηγοριοποίησης, τα προτεινόμενα διανύσματα παράγουν αποδεκτά αποτελέσματα, δεδομένης και της αισθητής μείωσης στις διαστάσεις. Ακολούθως τα προτεινόμενα διανύσματα χρησιμοποιούνται στην απευθείας σύγκριση των ΠΝΝ και των ΜΥΔ, οι οποία δίδει παρόμοια αποτελέσματα, ανεξαρτήτως της μεθόδου εκπαίδευσης. Σε όλες της περιπτώσεις τα αποτελέσματα επιβεβαιώθηκαν μέσω της στατιστικής ανάλυσης χρησιμοποιώντας τα διαστήματα ανοχής.

Σε συνθήκες πραγματικών δεδομένων οι εκπαιδευόμενοι αλγόριθμοι μπορεί να αντιμετωπίσουν προβλήματα καθώς οι ετικέτες τάξης των δεδομένων που χρησιμοποιούνται για εκπαίδευση και ο αριθμός των διαφορετικών πηγών δεδομένων δεν είναι εκ των προτέρων γνωστά. Ευτυχώς η χρήση αλγόριθμων χωρίς εποπτεία ή αλγόριθμων ομαδοποίησης δίνει αυτές τις πληροφορίες ή μπορεί ακόμη και να καταργήσει την ανάγκη για δεδομένα εκπαίδευσης. Συνεπώς χρησιμοποιείται ο αλγόριθμος DBSCAN με το προτεινόμενο χαρακτηριστικό διάνυσμα που χρησιμοποιεί χρονικά δεδομένα πάνω σε πειραματικά και πραγματικά δεδομένα από το δίκτυο της Αρχής Ηλεκτρισμού Κύπρου (ΑΗΚ). Τα αποτελέσματα καταδεικνύουν τη δυνητική χρησιμότητα του προτεινόμενου χαρακτηριστικού διανύσματος ως εργαλείο προ επεξεργασίας της ανάλυσης σε σχέση με την φάση (PRPD).

# Acknowledgements

*Κι αν πτωχική την βρεις, η Ιθάκη δεν σε γέλασε. Έτσι σοφός που έγινες, με τόση πείρα, ήδη θα το κατάλαβες η Ιθάκες τι σημαίνουν.*

*-Κ. Καβάφης*

I would like to acknowledge the support, encouragement and feedback of a number of people that made this work possible. Firstly many thanks to my supervisor Dr George E. Georghiou for his guidance, encouragement and motivation. He has done everything possible to see that this work had a successful outcome and his behaviour has set an example for me. Working with his team has been a beneficial learning experience both for my personal and professional development. Without his contribution none of this would have been possible.

A special acknowledgement to Dr Andreas Kyprianou for his contribution in the signal processing part of this work and for our discussions that led to fruitful development of ideas. His guidance was more than valuable and without it this work would had been so much less interesting.

Special thanks to Prof. Paul L. Lewin for his feedback throughout this work and for granting us access to The Tony Davies High Voltage laboratory at the University of Southampton, where part of the experiments was performed. Also the help of Dr Liewi Hao in setting up the experiments and providing all the necessary help during our visit at the lab is acknowledged.

The involvement of the Electricity Authority of Cyprus was an integral part of this work, both because of its financial contribution, as well as for allocating personell that escorted us during the field measurements, as well as granting us access to its T&D network. Special attribute is made to Dr Venizelos Efthymiou and Dr Andreas Stavrou from the EAC for their trust in this work.

The financial support of the EAC and the Cyprus Research Promotion Foundation for funding this work under contract TEXNO/1104/10 and ΤΕΧΝΟΛΟΓΙΑ/ΜΗΧΑΝ/0308(BIE)/05 is greatly appreciated.

Last, but not least, I thank my wife and my family for their understanding, tolerance and support during the years required to complete this work.

To my parents, my wife, and to my daughter.

Στους γονείς, την γυναίκα και κόρη μου.

A mis padres, mi mujer y mi hija.

# Contents

# List of Figures

# List of Tables

# Nomenclature

**Acronyms**

ACF Autocorrelation Function

ANN Artificial Neural Network

CWT Continuous Wavelet Transform

DBSCAN Density-Based Spatial Clustering of Applications with Noise

DSO Digital Signal Oscilloscope

EAC Electricity Authority of Cyprus

EPR Ethylene Propylene Rubber

EPR Ethylene Propylene Rubber

ERM Empirical Risk Minimisation

GIS Gas Insulated Switchgear

HF High Frequency range, 3-30 MHz

HFCT High Frequency Current Transformer

HOS Higher Order Statistics

HV High Voltage

MRA Multiresolution Analysis

MV Medium Voltage

OPTICS Ordering Points To Identify the Clustering Structure

PD Partial Discharge

pdf probability density function

PE Polyethylene

PILC Paper Insulated Lead Covered

PRPD Phase Representation Partial Discharge

QMF Quadrature Mirror Filters

RV Random Variable

$SF_6$ Sulphur Hexafluoride

SRM Structural Risk Minimisation

STFT Short time Fourier Transform

SVM Support Vector Machine

T&D Transmission and Distribution

UHF Ultra High Frequency range, 300 MHz-3 GHz

VC Vapnik-Chervonenkis

VHF Very High Frequency range, 30-300 MHz

WA Wavelet Analysis

WPT Wavelet Packet Transform

WT Wavelet Transform

XLPE Cross-Linked Polyethylene

**Mathematical**

$\epsilon$ Minimum distance used to define a core point in the DBSCAN

$\gamma$ Smoothing parameter of the exponential kernel in SVM

$\mathbf{x}$ Vector representation (column vector)

$\omega_m$ A priori probability of class $m$

$\rho_j(\mathbf{x})$ Decision risk of of classifying the input $\mathbf{x}$ into class $j$

$C$ Regularisation coefficient indirectly controlling the number of misclassifications in SVM ($C = \infty$ in the linearly separable case)

$c_{jm}$ Loss associated with classifying an input vector $\mathbf{x}$ into class $j$ given that it belongs to class $m$

$D_\omega^r$ Differential operator $\frac{d^r}{d\omega^r}$

$f(x, y)$ Probability density function of two RVs $X$ and $Y$

$f(x, \beta)$ Function parameterised by $\beta$ where $\beta \in \Lambda$ and $\Lambda$ is a set of parameters

$f(x, \beta_0)$ Function that minimises the risk functional $R(\beta)$ over the class of functions $f(x, \beta)$, $\beta \in \Lambda$

$f(x, \beta_l)$ Function that minimises the empirical risk based on a training set of $l$ samples

$F(x, y)$ Probability distribution function of two RVs $X$ and $Y$

$f_j(\mathbf{x})$ pdf of class $j$

$G_j$ No of components in the approximated pdf of class $j$

$g_j(\mathbf{x})$ General mixture approximating the pdf of class $j$

$h_k$ VC dimension

$k$ Capacity of the learning machine

$K(\mathbf{x}_i, \mathbf{x}_j)$ Kernel function

$k_{\min}$ Minimum number of points used to define a core point in the DBSCAN

$l$ Number of training samples

$L(y, f(x, \beta))dF(x, y)$ Loss function measuring the discrepancy between the actual response $y$ to a given input $x$ and the response $f(x, \beta)$ provided by a learning machine

$L_d(\alpha)$ Dual Lagrange function

$L_d(\alpha)$ Primal Lagrange function

$R(\beta)$ Risk functional, the expected value of the loss $L(y, f(x, \beta))dF(x, y)$

$R(\beta_l^k)$ Risk functional implementing SRM where $k$ is related to the VC dimension $h_k$

$R_{\text{emp}}(\beta)$ Empirical risk functional

$S_k$ Subsets of functions $S_k = \{f(x, \beta),\ \beta \in \Lambda_k\}$

$x$ Scalar representation

$\mathcal{W}x(a, b)$ Continuous Wavelet Transform with dilation $a$ and scaling $b$

$f \star g(t)$ Continuous convolution

$f \star g[n]$ Discrete convolution

$\int_{-\infty}^{+\infty} x(t)g(t)dt$ Correlation

$\langle x, g \rangle$ Inner product

$\mathbb{C}$ Complex numbers

$\mathbb{Z}$ Integers

$\mathbb{N}$ Positive integers including 0

$\mathbb{R}$ Real numbers

$\mathbb{R}+$ Positive real numbers

$P_{\mathbf{v}_j} x$ Projection of signal $x(t)$ onto subspace $\mathbf{V}_j$

$P_{\mathbf{W}_j} x$ Projection of signal $x(t)$ onto subspace $\mathbf{W}_j$

$\Phi(\omega)$ Characteristic function of a RV

$\Psi(\omega)$ Second characteristic function of a RV

$M(\omega)$ Moment generating function of a RV

$\kappa_r$ Cumulant of order r

$\mu'_r$ Moment of order r

$\mu_r$ Central moment of order r

$\phi_{a,b}$ Scaling function

$\mathbf{V}_j$ Vector space spanned by an orthonormal basis $\{\phi_{j,n}\}_{n\in\mathbb{Z}}$ for any $j \in \mathbb{Z}$

$\mathbf{W}_j$ Vector space spanned by an orthonormal basis $\{\psi_{j,n}\}_{n\in\mathbb{Z}}$ for any $j \in \mathbb{Z}$

$\mathbf{U} \oplus \mathbf{V}$ Direct sum of two vector spaces

$\psi_{a,b}$ Wavelet function

$a_j[n]$ Wavelet "approximation" coefficients at scale $j$

$d_j[n]$ Wavelet "detail" coefficients at scale $j$

$d_j^p[k]$ Wavelet Packet coefficients at scale $j$ and position $p$

**Physical Constants**

$\eta$ Electron attachment coefficient

$\alpha$ Electron ionisation coefficient (number of electrons produced by an electron per unit length in the direction of the field)

$E_0$ Externally applied electric field across the electrodes of a dielectric specimen

$E_{inc}$ Critical or inception electric field above which a PD can occur

$E_i$ Internal electric field across a PD

$E_{res}$ Residual electric field

$\epsilon_0$ permittivity of free space $\approx 8.854 \times 10^{-12}\mathrm{Fm}^{-1}$

$\epsilon_r$ relative permittivity of material

# Chapter 1

# Introduction

## 1.1 Motivation

On-line condition monitoring provides information about the existence of PDs in cables and their accessories under normal operating conditions and lies at the heart of a condition-monitoring program of power equipment. Sole identification of a PD, without any information about its source of origin, can be utilised for placing the equipment under monitoring and assessing it as having an increased risk of failure. However some PD types have a detrimental effect on the integrity of the insulation while others will pose no risk, taking into account the equipment under test. For example a corona discharge in oil filled switchgear can be classified as relatively harmless while the same type of discharge can be deemed as having a detrimental effect on Sulphur Hexafluoride ($SF_6$) switchgear. Therefore identification of the PD source is of paramount importance for the assessment of the insulation integrity of power equipment.

The evolution of the amplitude of PD activity has to also be recorded over time through several measurements before an evaluation can be made. This is due to the fact that the evolution (trending) provides more information about the integrity of the insulation rather than an absolute measurement. Once the presence of a PD and its source has been identified then further measurements will have to be taken at regular time intervals. Since calibration is not possible in on-line measurements, trending will also allow correlation between measurements of different equipment with the loading cycle. Generally the time to failure is very difficult to predict and can be years or seconds from the time of the first measurement. Therefore valuable information may be lost in between measurements (such as PD activity prior to failure) or even cable failure may occur between the measurement intervals. The implementation of an early warning

system to detect trend to failure and thus enable predictive maintenance can thus only be achieved through continuous on-line monitoring, the ultimate target of a complete PD monitoring system. Such a system poses difficulties that need to be addressed before its implementation becomes practical, especially when used for example in real-time mode. If the raw single pulse data delineating the characteristics of a PD in the time domain is used then the amount of data that needs to be processed can soon become unmanageable and lead to data storage and processing problems. Also input data to the classification algorithm has to have the minimum possible number of dimensions otherwise classification times can be very slow for real-time implementation.

The application of on-line PD measurements often gives a more descriptive status as the cable is operating under normal conditions, providing information about progressing degradation under operational stress and taking into account service factors that may affect the PD activity. Such factors can be the operating temperature and thermal expansion characteristics of the different materials, the mechanical loading, the current loading effects as well as long-term energisation effects (i.e. insulation responding differently because it has been energised for a long time, compared to initial energisation) which are not captured by off-line PD testing.

## 1.2 Aim of the thesis

The aim of this thesis is to propose and evaluate feature extraction methods of low dimensions that will be implemented in a continuous on-line system where minimisation of data storage and computational times are important. The proposed features are to employ complementary time and phase-resolved data and will be evaluated using a supervised algorithm with experimental data, and subsequently a supervised and unsupervised algorithm using on-line, field data.

## 1.3 Literature Review

### 1.3.1 Physical models

Partial Discharge measurement has emerged as the dominant insulation assessment method in medium and high voltage power cables. Although other methods such as the low frequency testing method [6] and the oscillating wave test method [7] exist, PD monitoring is by far the most widely adopted. One of the main reasons is that alternative methods

require the application of a voltage at a different frequency than the mains, which has been shown to provide PD patterns that are easily misinterpreted when compared to patterns using the 50 Hz cycle [8–10].

Partial Discharge is a complex physical process that can be physically explained using the theory of electrical breakdown in gases [11, 12]. For a PD to occur two conditions have to be fulfilled. Firstly an initiating electron has to be available and the electric field has to be above the inception field. Several attempts have been performed to model the physical behaviour of PDs [13], the majority of which have been performed in order to provide a better understanding of the underlying phenomena and to be able to interpret the different PD patterns. Some of the physical factors affecting the PD behaviour [13] are the interfaces involved in the PD, the electron generation, the physical properties of the streamer and the electrical field. Although there have been attempts to relate the parameters of physical models to actual measurements, such as [1] where the PD is treated as a stochastic process described by a differential equation, these models are valid for very specific cases and rely on unknown physical parameters. One such parameter relating to the model is the surface conductivity that has to be selected a priori based on knowledge of the discharge source, for example that it is a needle in air, or a void discharge and can be fine tuned using the observed pattern by trial and error. This information is not available in field scenarios where no knowledge about the discharge type is available and moreover such parameters do change during the course of PD activity.

Another model where PD activity has been treated as a stochastic process that can show memory effects due to the internal field alteration from space charge detrapment and decay related to surface conductivity is found in [14, 15]. In these works it was attempted to unravel those memory effects through the use of conditional amplitude and phase-of-occurrence distributions. In Pulse Sequence Analysis [16, 17] these space charge memory effects are taken into account by the considering the correlation between consecutive discharges. It relies on a plot of the phase difference between two successive discharges at time n ($\Delta U_n$) against ($\Delta U_{n-1}$) as a parameter set where the point to dielectric, electrode bounded cavity and electrical treeing discharges produced from experimental data have been employed to produce distinct PD fingerprints.

In contrast to attempts performed to directly relate physical parameters to actual measurements that can be a daunting task, physical models have been used in order to gain an understanding into the physical mechanisms underlying PD activity that can lead to descriptive parameters extracted from PD measurements. For example, rise times when measured at the discharge site [18] give an inside into the discharge mechanisms

involved and are affected by factors such as the overvoltage and space charges present at the discharge site. In voids where a short gap is involved lower amplitude longer rise-time pulses are referred as Townsend-like discharges while higher amplitude shorter rise-time pulses are referred to a streamer-like from streamers in long gap theory [19]. However both types of discharges are actually Townsend discharges since both involve the cathode feedback mechanism but with different mechanisms involved in the emission of electrons from the cathode [20]. In the case of low overvoltage conditions the classical ion impact is the primary cathode emission mechanism where in the case of high overvoltage and space charge build up the mechanism involved is photon impact followed by field and thermal emission resulting in lower rise times.

Such descriptive methods have been widely used in PD feature extraction and identification, due to the fact that they can be easily extracted from PD measurements. In this work this descriptive approach is followed after reviewing some of the most important descriptors of the discharge pulses such as their amplitude, rise time, recurrence rate, phase relationship of occurrence with respect to the applied alternating voltage and time interval to the preceding pulses [20].

### 1.3.2 PD Identification

One of the early and most successful analysis methods employed are the so called $\phi-q$ and $\phi-n$ representations where the peak ($q_p$) charge, the average ($q_m$) charge or the number of discharges, $n$, per phase window are plotted as a univariate distribution against the phase. Dividing the extracted quantity (for example the maximum charge) into windows as well, allows the calculation of the number of discharges for each phase and charge window. This can be plotted as a bivariate distribution against the charge and the phase, giving the 3-dimensional $\phi - q - n$ plot, one of the most complete PD data representations [21–27]. These representations have traditionally been referred to as Phase Representation PD (PRPD), a terminology that is adopted here.

Statistical analysis tools were one of the first methods applied successfully on phase-resolved data to get descriptive parameters for the 2-dimensional PRPD distributions and many such attempts can be found in the literature [21, 22, 24, 27, 28]. For example in [28] the complete feature vector was derived from the estimation of the skewness and kurtosis of the positive and negative half cycle of the $\phi - q_m$ and $\phi - n$ distributions as well as other statistical operators such as the correlation coefficient. This 9-dimensional feature was tested on sixteen full scale samples and showed a good potential for classification of

discharging defects in insulation structures using contour score. In [21] minimum distance classification using statistical data on pulse quantities such as apparent charge, energy and phase has been performed using the $L_1$ and $L_2$ norm metrics. In the minimum distance classifier each class is represented by a cluster and a reference point for that cluster, where a data point is classified into the cluster with the nearest reference point. A 15 dimensional feature derived from the application of statistical tools to analyse the properties of the mean pulse height and pulse count phase distributions was employed in [22]. These descriptors were used as input patterns to a back-propagation network, Kohonen self-organizing map and learning vector quantization network. All three neural networks, as used in this work, recognized fairly well the PD patterns of those insulation defects for which they were trained.

In addition to statistical analysis tools, signal processing on 2-dimensional PRPD or image processing on 3-dimensional PRPD were employed. The $\phi - q - n$ was transformed into a 2-dimensional image where one of the quantities (usually the $n$) was represented by the pixel values of the image. Tools such as the 2D wavelet transform [26] or fractal analysis [29, 30] were used to extract descriptive features.

In the above approaches the data was acquired using the IEC 60270 [31] standard which defines the upper bandwidth of the measurements to be below 1 MHz. The use of PRPD distributions to produce interpretable representations is based on the assumption that the recorded data is free from interference and there is only one source of PD active. In on-line field conditions this is often not the case and therefore wideband measurements employing higher frequency ranges (3 MHz-3 GHz) have been employed in order to be able to pre-process a PD signal to remove interferences and separate multiple PD sources concurrently active. The employment of such methods allows the analysis of the waveforms of PD defects which forms a recent development in PD identification. The two methods can complementary be combined as in [32], an approach that will be followed in this thesis.

Most non-conventional electric PD systems are based on detecting the HF properties of PD processes. As a result, to sensitively detect the PD signals, higher frequency contents of PD signals are processed in time- or frequency domain. Time and frequency content of time-resolved PD signals is related to the type of PD as different sources of PD produce different time and frequency characteristics. Moreover, the low-pass characteristics of underground power cables result in distortion as a function of the distance travelled, affecting the frequency content of the pulse. In [33] different classification algorithms have been applied on PD time parameters such as the apparent charge, the rise time,

fall time, area under the PD pulse and product of the pulse width and apparent charge. The pulse shape recognition capabilities using these time parameters were assessed using the nearest neighbour classifier, learning vector quantization and multilayer perceptron paradigms, on artificial cylindrical cavities of different sizes and provided similar results. An extension of the above is found in [34] where similar parameters are classified using the fuzzy decision tree based on different cavity sizes. The fuzzification is based on three void sizes, small (1 mm), medium (1.5 mm) and large (2 mm).

To analyse the frequency content of a PD signal the Fourier Transform was employed in [35] to extract features to identify fixed (on conductors or spacer surfaces) or free moving particles in $SF_6$ gas-insulated substations (GIS). However one of the weak points of the Fourier Transform (FT) is that its basis functions extend over the whole time interval without any localisation properties, and therefore time information is completely lost. To overcome this problem the Short Time Fourier Transform (STFT) is employed that uses a fixed window to capture the signal allowing a simultaneous time-frequency analysis of the signal.

An implementation of the above transform can be found in [36–38], where the signal was captured in windows of $K$ samples. The acquired pulses were mapped into a time (T) and frequency (F) equivalent bandwidth calculated as the second central moment of the square of the time and frequency coefficients of the data, creating a time-frequency representation of the energy of the signal. The detailed procedure for the T-F map was explained in [38] where the signals have been identified using a fuzzy classification. A distance function is used to to derive the membership of data as a function of the distance from each cluster and formed clusters are classified as different PD sources, where clustering is done in an unsupervised mode.

The Wavelet Transform provides a variable time-frequency decomposition and has been employed in PD classification using experimental data [39]. The detail wavelet coefficients at level 3 are utilised as features and an SVM is employed in the classification. The training data was provided by experimental set-ups with different dimensions and shapes than the test data which proved the robustness of the use of the WT with an SVM. In [40] the WT has been used in obtaining a time-frequency map representation of different PD types utilising time data. A neuro-fuzzy approach achieves automatic PD identification based on slow and fast rise times of the PD pulses. The Wavelet Packet Transform (WPT) due to its inherent adaptive structure and ability to capture the signal's characteristics in a few coefficients was employed in [41] where the kurtosis and skewness of selected coefficients were used to identify experimental PDs in GIS.

The statistical analysis of PD pulses is able to capture the variability in PD properties and have thus been employed in feature extraction. Such a method was employed in [42] on time-resolved measurements, where different PD sources were identified on the basis of the shape parameter of the Weibull distribution of the pulse height. The Autocorrelation Function (ACF) was also used in [43] for clustering time-resolved data (identified by the K-means clustering algorithm). The utilisation of statistical analysis using moments was confined up to the second order and the use of Higher Order Statistics (HOS) received limited attention. The application of HOS on time-resolved data was examined on acoustic and electric data taken from experimental measurements of contaminating particles in a tank of oil [44]. Features such as the collision time interval between successive particle-tank impact, and the impact magnitude were extracted and it was demonstrated that higher order moments and (histogram entropy) of the features do improve classification results.

This research attempts to employ the concept of statistical analysis through the application of HOS (up to fourth order) on selected wavelet coefficients, in order to explore the potential of the combination of WPT and HOS methods on electrical measurements in power cables. The proposed features are evaluated using experimental and field data and supervised (such as the PNN and the SVM) and unsupervised (DBSCAN) algorithms. The next section elaborates on the motivation behind this thesis and the following section on the work done in this thesis.

## 1.4 Work done towards the solution of the problem

The correct classification of different PD types is of vital importance in a condition-monitoring program and the evolution of PD activity over time gives valuable information about the integrity of the system. This improves the decision-making and thus the work done in this research implemented algorithms that can be part of a continuous on-line condition monitoring system. The target of PD source identification using a lower dimensional feature was broken into different tasks which when combined together gave an overall solution to the problem as shown in figure 1.1.

The PD source identification problem is approached using both phase and time resolved data which provide a complementary and thus a more complete representation of the PD data. Traditional phase resolved PD data representation of the measured quantities is regarded as one of the most complete forms of graphical representation of

a PD pulse distribution [21–24, 26, 27]. However its use in on-line, on-site implementations suffers from the contamination of the measurements by noise and the need to set a threshold, below which all measurements are discarded. Also the simultaneous activity from multiple PD sources that can potentially be present at the equipment under test makes the implementation of such phase resolved methods much more prone to erroneous conclusions. Specific time characteristics of the waveform such as the rise time, fall time and shape constitute important parameters that can be correlated to the PD activity. The recording of a single pulse waveform requires sufficient sampling rate which when whole data cycles are sampled the amount of data becomes prohibitively expensive in terms of storage and processing. The capture of single pulse data together with its phase of occurrence allows the data to be processed in the time domain taking advantage of the time resolved data analysis benefits. Both types of data were acquired in a laboratory under controlled conditions from different PD sources (corona in air, floating in oil, internal in oil and surface in air). This data formed part of a database that is analysed using statistical and signal processing tools to extract the most representative features of the different PD sources.

Feature extraction algorithms for both phase and time resolved data are proposed utilising the Wavelet Packet Transform (WPT) as a pre-processing step and following a procedure for the selection of particular nodes in the tree decomposition that are found

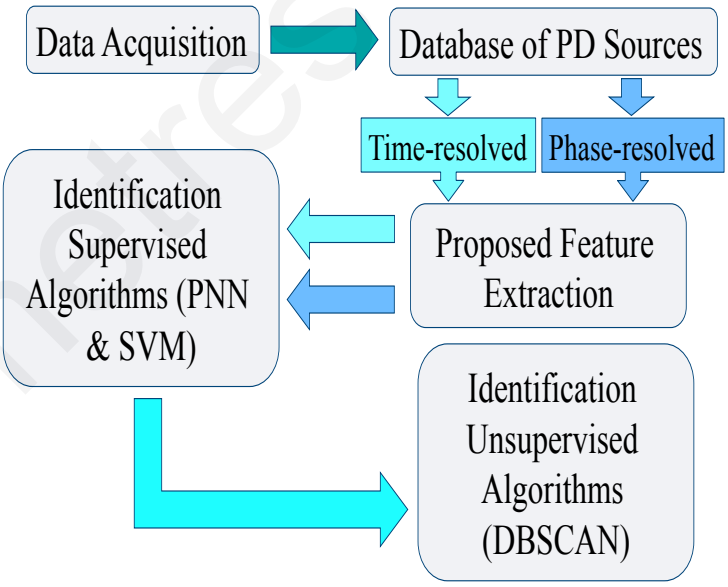



Figure 1.1: Flow chart of the work done in this thesis towards the proposal and evaluation of lower dimensional feature vectors for use in continuous on-line monitoring systems.

to best represent the characteristics of the signal.

Once the specific nodes of the WPT tree were selected Higher Order Statistics (HOS) were used to approximate the coefficients of these nodes. The first four cumulants were used for each node resulting in 4 dimensions per node included in the feature. For phase resolved data all the nodes in the best wavelet packet tree decomposition were selected and HOS of the nodes at each level were added together to reduce the number of dimensions. This feature can be compared to the raw use of the un-processed wavelet coefficients at specific nodes where a feature vector of 1468 dimensions was created. The new feature yields gives similar results but with a dimensionality of the order of 36, a significant improvement [45]. This low dimensional feature vector alleviated the numerical problems associated with the use of a Probabilistic Neural Network (PNN) which was compared to the use of the Support Vector Machine (SVM) and features of much higher dimensions, providing similar results.

For the time resolved data the use of HOS on the wavelet coefficients of nodes at level 3 position 1 ($d_3^1[k]$), level 4 position 1 ($d_4^1[k]$), level 5 position 0 ($d_5^0[k]$) and level 5 position 1 ($d_5^1[k]$) produced a feature vector of 16 dimensions while if the un-processed, wavelet coefficients at the same nodes were used 1304 dimensions would have been created. The proposed feature was compared to the un-processed feature using the SVM and the results were identical (98%). Next the PNN was compared to the SVM, which was made possible through the reduction in dimensions and the alleviation of problems in the PNN associated with the dimensionality. The two algorithms produced similar classification results although implementing different training methods.

Once evaluated using these supervised algorithms the potential use of the time-resolved proposed feature in on-line scenarios was evaluated. In such cases the presence of an unknown number of PD sources simultaneously active and the fact that the training data class labels are not a priori known called for the use of an unsupervised or clustering algorithm. Such an algorithm can provide the class labels and the number of PD sources, under the assumption that each source maps into a single cluster, enabling the use of supervised algorithms or even totally alleviating the need for training data. The DBSCAN was utilised firstly on experimental data acquired from PILC cable loops with known defects, simulating on-line operating conditions as close as possible, and then extended to field data acquired from the EAC distribution network. The results proved promising, demonstrating the use of the proposed time-resolved feature vector as a pre-processing step to PD source clustering. Subsequently PRPD plots can be extracted for each cluster which can help in the identification of PD sources.

## 1.5 Novelty of the research

The research in this thesis is an attempt to approach the PD classification in cable networks from a continuous on-line monitoring perspective. The main novelty of this research is the proposal and evaluation of feature extraction methods for both phase and time-resolved data that incorporate an important reduction in the number of dimensions without any significant drop in the performance. The proposed feature extraction methods are evaluated and compared with existing methods by utilisation of supervised and unsupervised algorithms, and experimental and field on-line data.

Firstly the comparison of the proposed feature extractor for continuous on-line PD identification is performed to a similar feature extractor of higher dimensionality using the SVM. Results of both the phase and time resolved data show slightly inferior but still acceptable performance between the already used, higher dimensional and the proposed feature vectors.

As a consequence of this lower dimensional feature the use of the Probabilistic Neural Network (PNN) is made possible as this method suffered from numerical problems associated with the dimensionality of the data as well as speed problems as its implementation involves the computation of the inverse of a matrix whose dimension is directly proportional to the dimensionality of the feature vector. As the complexity of inverting an $N$-dimensional matrix is $O(N^2)$, a significant reduction in the number of dimensions results in an even higher reduction in the computational complexity for the PNN. The SVM and the PNN are directly compared using the same feature vector and demonstrated the use of the proposed features with supervised algorithms, that rely on two distinct training principles.

Finally the DBSCAN is applied on the proposed time-resolved feature vector utilising both experimental and field, on-line data evaluating the use of the proposed feature vector in pre-processing real data to extract different clusters and then PRPD can be complementary used to identify the different sources.

## 1.6 Outline of the proposal

The aim of this thesis is to propose and evaluate feature extraction methods of low dimensions that will be implemented in a continuous on-line system where minimisation of data storage and computational times are important.

In chapter 2 the experimental setup in a high voltage laboratory to generate different PD sources under controlled conditions is described. The experiments are performed in order to generate the most frequently encountered PD types in cable networks, namely the corona, the floating, the internal and the surface discharges. The database of different PD sources includes both time and phase resolved measurements, which are considered complementary to each other. The apparatus and the sensor used are analysed, together with a treatment of the two main cable types used in the Cypriot T&D network and their characteristics to PD propagation.

The acquired data is analysed in chapter 3 in an effort to arrive at a compact feature representation. Two features are proposed, one for time-resolved and another for phase-resolved data. Both employ the Wavelet Packets Transform (WPT) to achieve an adaptive time-frequency decomposition of a signal. Specific nodes are selected that represent the characteristics of the different PD sources and the wavelet coefficients at these nodes describe their characteristics. In an effort to reduce the number of dimensions the first four cumulants from Higher Order Statistics (HOS) are used as statistical descriptors of the wavelet coefficients at the selected nodes.

The proposed vectors are evaluated in chapter 4 using the experimental dataset and two supervised algorithms that employ a set of labelled training data in order to classify previously unseen data. Firstly the time-resolved proposed feature is compared to a higher dimensional feature vector [39] using a Support Vector Machine (SVM) and the classification rate as a performance criterion. Once shown that the reduction in the number of dimensions does not significantly affect the performance, the SVM is compared to the Probabilistic Neural Network using the proposed feature. The two algorithms utilise different principles during the training phase, and it is investigated whether this has an effect on the classification rate. The same procedure is followed for the phase-resolved proposed feature vector and the reproducibility of the results for both features is verified by the use of tolerance intervals.

Finally in chapter 5 the proposed time-resolved feature vector is evaluated as a potential tool in pre-processing of on-line data to extract clusters and their respective PRPD plots which can be utilised in PD identification. In on-line scenarios the presence of an unknown number of PD sources and the lack of knowledge about the class label of the training data poses a problem to the use of supervised algorithms. Fortunately unsupervised or clustering algorithms can provide such class labels or even alleviate the need for training data. The DBSCAN is employed on the proposed feature vector calculated from

experimental data on PILC cable loops, simulating operating conditions as close as possible. Once found to identify different PD sources on-line field data from the distribution network of the EAC are analysed, showing promising results and the potential that this feature can be used as a pre-processing tool in on-line PD identification.

At the end the conclusions drawn from this work are summed up in chapter 6 and some thoughts on future work are reflected in chapter 7.

## 1.7 List of publications out of the thesis

### Conferences

I D. Evagorou, A. Kyprianou, P.L. Lewin, A. Stavrou, V. Efthymiou, and G.E. Georghiou. An investigation into the parameters affecting the success of the wavelet packets method for denoising partial discharge signals. In *5th Mediterranean Conference on Power Generation, Transmission and Distribution*, November 2006.

II Demetres Evagorou, Andreas Kyprianou, Paul L. Lewin, Andreas Stavrou, Venizelos Efthymiou, and George E. Georghiou. Classification of partial discharge signals using probabilistic neural network. In *IEEE International Conference on Solid Dielectrics, ICSD*, pages 609–615, July 2007.

III D. Evagorou, A. Kyprianou, P.L. Lewin, A. Stavrou, V. Efthymiou, and G.E. Georghiou. Effect of the choice of the decomposition level on the success of the wavelet packets method for denoising partial discharge signals. In *6th Mediterranean Conference on Power Generation, Transmission and Distribution*, November 2008.

IV D. Evagorou, A. Kyprianou, P.L. Lewin, A. Stavrou, V. Efthymiou, and G.E. Georghiou. Evaluation of Partial Discharge Denoising using the Wavelet Packets Transform as a Preprocessing Step for Classification. In *Electrical Insulation and Dielectric Phenomena (CEIDP), 2008*, October 2008.

V J.A. Hunter, P.L. Lewin, D. Evagorou, A. Kyprianou, and G.E. Georghiou. Comparison of two partial discharge classification methods. In *IEEE International Symposium on Electrical Insulation*, June 2010.

VI D. Evagorou, A. Kyprianou, A. Stavrou, J.A. Hunter, L. Hao, P.L. Lewin, and G.E. Georghiou. Performance of the support vector machine partial discharge classifi-

cation method to noise contamination using phase synchronous measurements. In *Electrical Insulation and Dielectric Phenomena (CEIDP), 2010*, October 2010.

VII Demetres Evagorou, Andreas Kyprianou, Andreas Stavrou, Venizelos Efthymiou, Paul L. Lewin, and George E. Georghiou. Setup and preliminary results of a pilot on-line cable pd monitoring system on an 11 kv switchboard using capacitive couplers. In *7th Mediterranean Conference and Exhibition on Power Generation, Transmission, Distribution and Energy Conversion (MedPower 2010)*, November. 2010.

VIII D. Evagorou, A. Kyprianou, L. Hao, P.L. Lewin, and G.E. Georghiou. Multisource PD identification based on phase synchronous and asynchronous data. In *Electrical Insulation and Dielectric Phenomena (CEIDP), 2011*, October 2011.

## Journals

I D. Evagorou, A. Kyprianou, P.L. Lewin, A. Stavrou, V. Efthymiou, A.C. Metaxas, and G.E. Georghiou. Feature extraction of partial discharge signals using the wavelet packet transform and classification with a probabilistic neural network. *IET Science, Measurement & Technology*, 4(3):177–192, May 2010.

II D. Evagorou, A. Kyprianou, P.L. Lewin, A. Stavrou, and G.E. Georghiou. A lower dimensional feature vector for identification of partial discharges of different origin using time measurements. *IOP Measurement, Science & Technology*, Under Review.

# Chapter 2

# Building a database of experimental PD sources

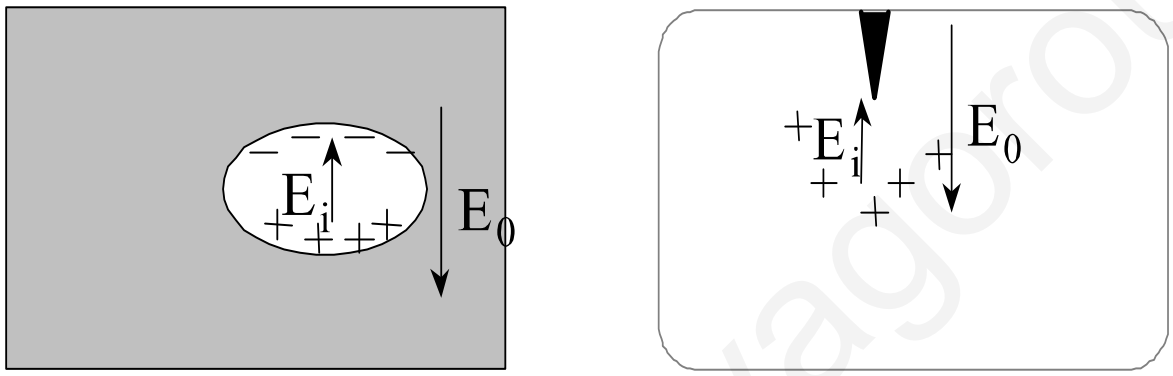
In this chapter an attempt to build a database of PD sources using an experimental arrangement in a high voltage laboratory is presented. Firstly a brief overview of how PDs are generated and the equivalent electrical model that can be used to simulate their recurrence using a set of capacitors is given. Next two common types of cables, installed at the Cypriot T&D network, are briefly reviewed as their high-frequency response characteristics can affect the bandwidth of the PD and the choice of sensor. The frequency response of a commercially available clamp-on High Frequency Current Transformer (HFCT) sensor is measured using a Network Analyser in order to determine its bandwidth and in effect its suitability in PD measurements for cables and their accessories. Once the sensor is characterised an experiment is designed to create a database of experimental PDs composed of phase synchronised as well as single pulse data. The equipment used in the experiment and the different set-ups for building the database of PD sources is explained next. This chapter presents the first step of the methodology that will lead to the development of a proposed feature extraction method for PD identification.

## 2.1 PD generation

A partial discharge (PD) is a localised electrical discharge that only partially bridges the insulation [31]. Although a PD does not indicate a complete failure of the system or collapse of the voltage between the energised electrodes, it indicates a weak point in the installed system. The physical principles governing the generation and evolution of partial discharges lie on the study of electrical breakdown in gases, and especially in air

which has been around for a long time. A rather simple overview of how these discharges are generated and repeated is presented next while more detailed theoretical background can be found in textbooks such as [11],[12].

For a PD to take place the total electric field ($E_{tot}$) must be above the critical electric field, known as the inception electrical field ($E_{inc}$), and an initiating electron must be available to start the first avalanche of the ionisation process. The total electric field is defined as the sum of the internal field ($E_i$) and the externally applied electrical field




(a) (b)

Figure 2.1: External ($E_0$) vs internal ($E_i$) electric field for (a) gas-filled void in an insulator and (b) corona in air configuration (taken from [1])

($E_0$) (figure 2.1). In general the fraction of the externally applied electric field that will appear across the void gap depends on the geometry of the void [13]. The critical electric field is a function of the material properties and depends on the choice of the gaseous and solid insulation materials during the insulation design as well as the pressure of the gas. The availability of initiating electrons controls the statistical characteristics of the discharges while in the limiting case where there is ample availability of electrons, PDs can show deterministic behaviour [46].

### 2.1.1 Recurrence of discharges

A PD can be detected as a sequence of current impulses across the leads of the sample. The PD process leading to these impulses is qualitatively described with the aid of figure 2.2, in relation with the behaviour of the total electric field for this particular void. Although the electric field across the dielectric is a fraction of the externally applied voltage and the constant of proportionality is related to the shape of the void, here, without loss of the qualitative description it is assumed to be unity. Therefore $E_o$ appears

across the gap. Assuming that this is the first discharge in the sample no space charges

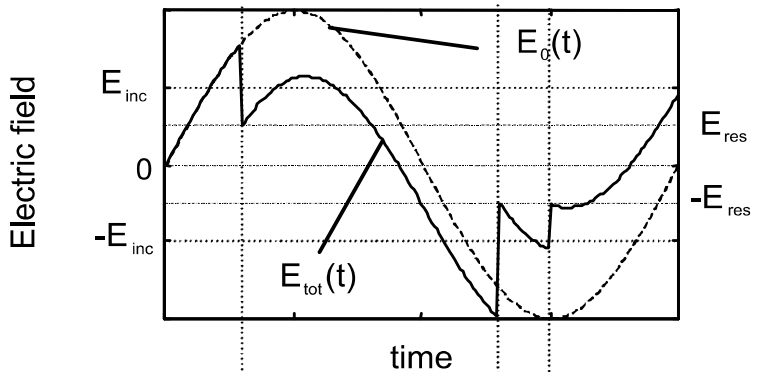



Figure 2.2: Behaviour of $E_{tot}(t)$ under the application of a 50 Hz voltage (taken from [1]). The recurrence of PDs is explained with reference to the electric fields. The total electric field $E_{tot}(t)$ is the sum of the externally applied $E_0(t)$ and the internal $E_i(t)$ fields. When it reaches a value above $E_{inc}$ and provided that a free electron is available then a PD occurs. The total electric field immediately after a PD occurrence drops to $E_{res}$, and subsequently rises approximately as $E_0(t)$. When it is raised again above $E_{inc}$ another discharge occurs and this pattern repeats itself.

exist and $E_{tot}$ has the same value as $E_o$ up to the first PD occurrence. When $E_{tot}$ surpasses the inception field ($E_{inc}$) and an initiating electron becomes available then a PD occurs, dropping $E_{tot}$ to the residual field. This electric field drop (and thus the voltage drop) takes place in a very short period of time (of the order of $10^{-7}$s), which can be picked up at the leads of the samples as a breakdown current. After the discharge the total electric field (being the total of the applied electric field and the internal electric field of the surface discharges deposited at the cavity walls after the first discharge, figure 2.1) begins to increase again. When the electric field reaches the inception field again then a new PD occurs. Space charges left over by previous discharges can become trapped at shallow traps and provide additional electron supply through their detrapping [**?** ]. When the total electric field starts to fall a PD occurs when it reaches the $-E_{inc}$ value. Therefore a pattern of voltage drops takes place during the discharge. The magnitude of the pulses depends on $\pm E_{inc}$ and $\pm E_{res}$. The positive and negative values being equal pulses of the same amplitude are created while asymmetrical discharges occur at cavities with an inception field $E_{inc}$ being unequal to $-E_{inc}$. Moreover in a practical case the generation and decay rate of space charges from previous PDs will affect the regularity of the PD sequence. These PD sequences can theoretically be picked up by different measurement sensors based on the various physical quantities generated by PD activity.

## 2.2 PD detection

PD activity generates physical phenomena which can cause the transmission of acoustic, electrical and optical energy, as well as chemical changes within the dielectric material. Detection methods are based on sensors that can measure these emissions which are very small, typical signal magnitudes are often close to background noise levels. Optical and chemical sensors find use mainly in laboratory measurements with the optical method being characterised by a high sensitivity (less than 1 pC) [47]. Practical detection relies on acoustic and electrical methods the latter being the most commonly used in power cables.

The acoustic emission technique [48] has the advantage of being immune to electrical interference but attenuation of the acoustic waves significantly reduces measurement sensitivity (around 100 pC). It is more suitable for PD monitoring in power transformers [49], switchgear or GIS, due to the better transmission coefficient in oil compared to power cables. Since the ultrasonic signals travel from the origin of a discharge to the sensor along different paths and at different speeds in a power transformer, distinguishing the signal coming from the origin and those corresponding to alternative paths of the reflected signals is one of the main limitations of its external use in PD location.

Electrical methods constitute the most common of the detection methods and rely on two different physical principles which can be used to classify them into two broad categories. Sensors forming a capacitive link [50],[51] with the equipment under test have been used and provide the most direct way to measure PD activity, giving the highest sensitivity. However they require an interruption of the supply to the equipment in

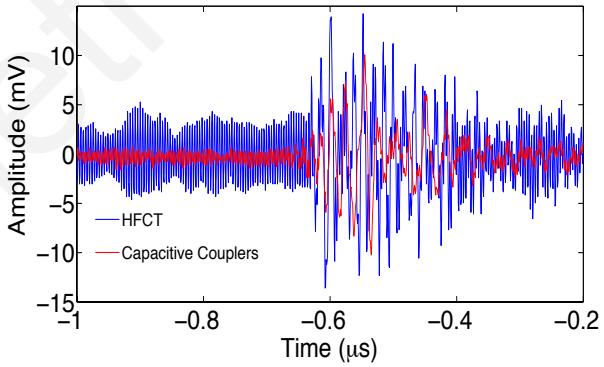



Figure 2.3: Comparison of a commercial HFCT sensor and an epoxy mica capacitor measurement at an on-line PD monitoring system

order to be installed which makes them less attractive for on-line measurements. On the

other hand inductive sensors constitute a large part of the electrical detection method and their success lies on the fact that they can be installed non-invasively. They are clamped around the earth wire of the equipment under test measuring the current to earth. Data acquired from an on-line cable using a commercially available HFCT sensor (IPEC 140/100 HFCT) and a capacitor consisting of 80 mm layer of epoxy impregnated mica splittings as the main dielectric (referred to as epoxy mica capacitive coupler) is shown in figure 2.3 where the lower noise floor demonstrates a higher sensitivity for the capacitive coupler. However a drawback of both methods is that they are susceptible to electrical interferences and noise, which usually have to be removed or reduced by pre-processing the measurements using signal processing techniques. The choice of an HFCT in this work is based on the ease of installation of the HFCT clamp-on sensor on measuring medium voltage cables, which becomes important when field measurements as a screening method for PD activity are performed on a systematic basis.

Electrical coupling techniques work on various frequencies, ranging from a few MHz to several hundred MHz. Sensors are being classified according to their bandwidth ranging from High Frequency (HF, 3-30 MHz), Very High frequency (VHF, 30 -300 MHz) all the way up to the Ultra High frequency (UHF, 300 MHz-3 GHz) range. The choice of a sensor with a sufficiently high bandwidth depends on the equipment under test and affects the detected PD. For cable circuits HF sensors are used due to the high frequency attenuation characteristics of the cable [52], [53] while UHF sensors are used in Sulphur Hexafluoride ($SF_6$) equipment which has a lower high frequency attenuation factor. In the next section an overview of the two types of underground cables encountered in the T&D network of the Electricity Authority of Cyprus (EAC) is given and how their high frequency attenuation affects PD detection is presented.

## 2.3 Underground cables

When considering the propagation characteristics of underground cables the type of cable must be taken into account as different types have different loss mechanisms. The principal design elements of Medium (MV) and High Voltage (HV) are depicted in figure 2.4. With reference to this figure they are classified in two general categories according to the dielectric used: impregnated paper and extruded synthetic materials.

Impregnated paper cables consist of narrow strips of paper lapped spirally over each other [2]. In order to ensure that different layers do not have any air gaps between them (and thus prevent pre-discharges) a high-viscosity insulating oil is used, which

at operating temperature becomes less viscous thus filling any air gaps. This change in viscosity with operating temperature produces a"self healing" effect for this type of cable, which means that an air void forming a source of PD at low operating temperature (and therefore high viscosity) can be filled with oil when the cable reaches a higher operating temperature (at which the oil has a lower viscosity) which temporarily prevents this source from discharging.

Solid dielectric cables require no impregnating medium and the dielectric can be extruded directly on the conductor, resulting in a less time-intensive operation to construct this type of cables. The two main dielectric materials used are the cross-linked polyethylene (XLPE) and synthetic rubber (Ethylene Propylene Rubber, EPR). XLPE has prevailed as the most commonly used solid dielectric material for voltages above 150 kV as the EPR has comparatively higher dielectric losses. One of the things that has to be kept in mind is that in polymeric cables the level of PD activity permitted is much lower than in PILC as they lack the stability provided by the oil viscosity in the presence of PDs. Although XLPE cables have replaced PILC ones since many years, large lengths of PILC cables are still in operation in many countries. The brief analysis here aimed at understanding the structural and electrical (table 2.1) differences between the two cable types, as the cable type can be taken into account when interpreting PD measurements due to their slightly different HF behaviour, as explained next.

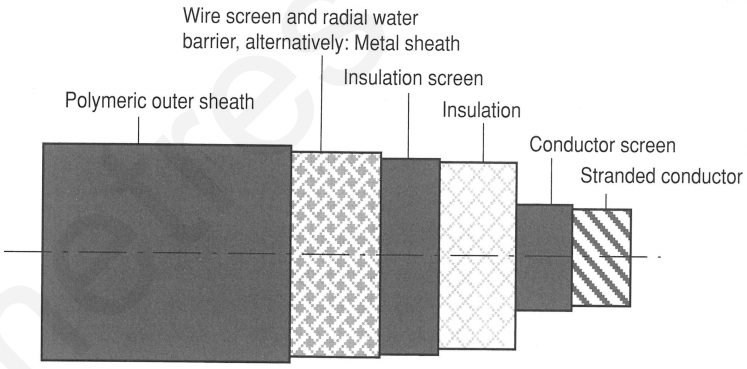



Figure 2.4: Explanation of the main principal design elements in High Voltage cables (taken from [2]). The figure illustrates the stranded conductor that can be made of aluminium or copper, together with the conductor and insulation screens (semiconductor shields) which differs for paper-insulated and XLPE cables. In paper-insulated they consist of copper tape interlapped with kraft or metalised kraft paper tape while in polymeric dielectrics they consist of extruded semiconducting layers of polyethylene (PE) which ensures good bonding between this layer and the insulation. The metal and outer sheaths are not considered here.

Table 2.1: Comparison of Electrical and Dielectric Properties of paper-insulated and XLPE-insulated High Voltage (HV) cables(taken from [2])

| Property | Test Conditions | Paper Value | XLPE Value |
|---|---|---|---|
| Dielectric dissipation Factor tan $\delta$ | 50 Hz/20 °C | 1.5-3×$10^{-3}$ | 0.3-0.5×$10^{-3}$ |
| Permittivity $\epsilon_r$ | 50 Hz/20 °C | 3.5-3.8 | 2.3-2.4 |
| Dielectric loss coefficient $\epsilon_r \cdot \tan\delta$ | 50 Hz/20 °C | 5-12×$10^{-3}$ | 0.7-1.2×$10^{-3}$ |

### 2.3.1 High frequency behaviour of cables

Power cables act as low-pass filters therefore high frequency signals are being attenuated when propagating through them. This attenuation is frequency dependent and can alter the waveform of a signal travelling down the cable. The mechanisms dominating the high frequency losses for paper-insulated and XLPE cables are slightly different at higher frequencies. At low frequencies (<1MHz) the loss mechanism is common to both types of cables and is dominated by the skin effect of the conductor. This loss is frequency dependent and varies roughly as the square root of the frequency. Dielectric losses increase roughly in proportion with the frequency and thus dominate at higher frequencies. For PILC cables with relatively high dielectric losses (see table 2.1) this is the dominating mechanism at high frequencies (MHz range). For cables with low dielectric losses, such as XLPE, the radial capacitive current passing through the conductor and ground shield layers is the dominating high frequency loss mechanism. This current is dependent on the frequency related conductivity and dielectric constant of the shield [52]. The loss mechanisms decrease the bandwidth (-6 dB) of a 270 MHz pulse to 10 MHz after around 100 m of propagation [54] which explains the fact that a HF sensor is sufficient for PD measurements in cables. In the next section the bandwidth of the sensor used in acquiring the PD signals in this work is measured by a network analyser.

## 2.4 Bandwidth of HFCT

A commercial HFCT with internal and external diameters 100 mm and 140 mm respectively (figure 2.5) is used in acquiring the PD data and its frequency response is analysed to evaluate its suitability for PD measurements. The magnitude and phase frequency responses captured by the network analyser, are shown in figure 2.6. One port of the

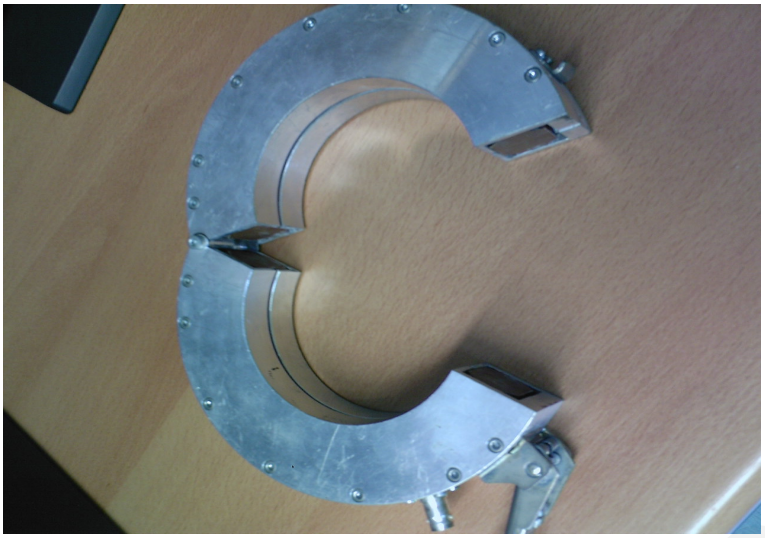
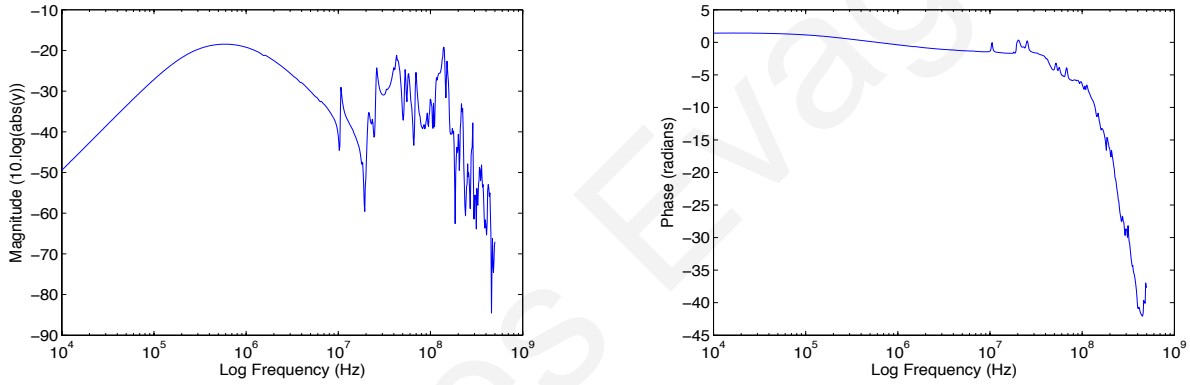


Figure 2.5: Commercial HFCT (IPEC 140/100) analysed




(a) HFCT Magnitude (b) HFCT Phase

Figure 2.6: HFCT Power Spectrum

network analyser is connected to a wire terminated by a 50 Ω BNC connector passing through the centre of the sensor (simulating the earth carrying conductor) while the second port is connected to the BNC of the sensor. From the frequency response plot above, it can be deduced that the bandwidth of the transformer at 25 dB below the peak value, defined as the useful bandwidth, is around 200 MHz. The actual frequency response of the sensor defined as the point of occurrence of the first resonance, is around 10 MHz. At higher frequencies a series of resonances and anti-resonances can be observed as for example at 20 MHz, which is due to the presence of parasitic inductances and capacitances in the sensor. The same response has been observed in [55] where the higher frequency response is dominated by the capacitances in the system. Also according to [56] the

difference in the value of the coupling coefficients between different sections of the CT (depending on the position of the primary coil in the transformer window) can be one of the reasons for the occurrence of resonances and antiresonances at higher frequencies. Similar frequency behaviour has been predicted (although at higher frequencies) for a Rogowski coil modelled as a distributed-element transmission line [57].

With a useful bandwidth of around 200 MHz and a first resonance at 10 MHz the HFCT has a sufficient bandwidth for experimental PD data collection as well as on-line PD data from cable loops. Therefore one can proceed to the build-up of a database of various PD sources under laboratory controlled conditions that will be used in the characterisation of PDs. In order to generate partial discharge data from different sources a simple experiment is set up in a high voltage laboratory. A diagram of the experimental arrangement is shown in figure 2.7. The 60 kV bushing tap of a large auto-transformer used (model 60HC755) has a 235 pF nominal capacitance, and is PD free under its nominated operating condition. The clamp-type split core HFCT IPEC model 140-100 is

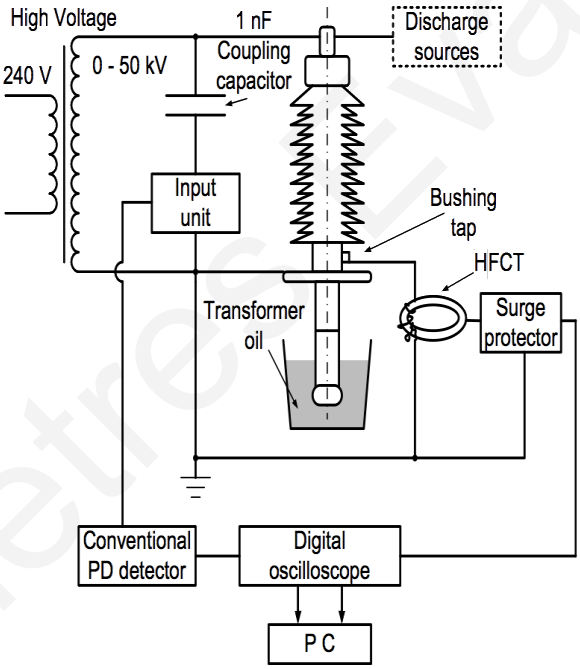



Figure 2.7: Diagram of the experimental setup

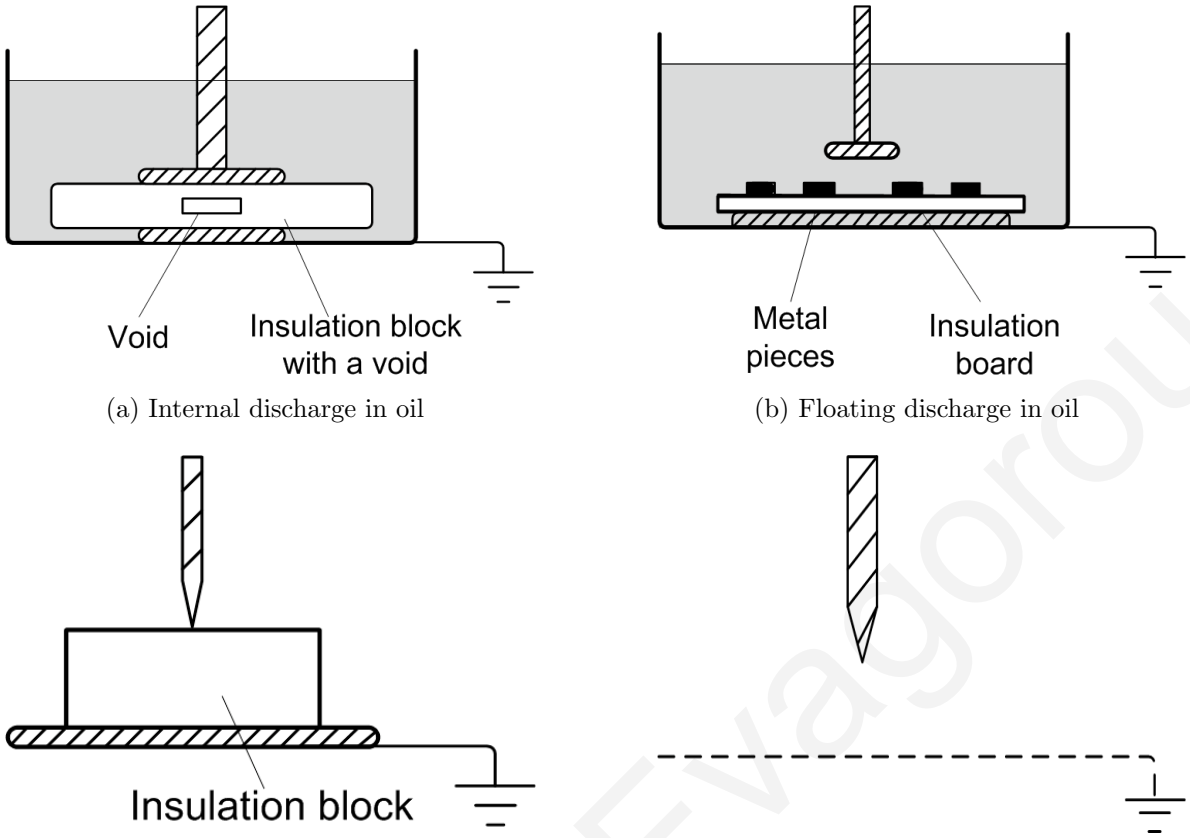
used as a sensor with a measurable frequency range from 10 kHz to 200 MHz. A digital oscilloscope, Tektronix DPO7254 with a bandwidth of 2.5 GHz and 400 MSample memory is used to display, analyze and store the obtained signals. A Robinson type 5 model 700 conventional electrical PD detection (in accordance to IEC 60270 [31]) with 40-300 kHz

bandpass frequency is also used to quantify the apparent charge of the simulated PD activity, through a blocking capacitor ($C_K$) of 1 nF and a measuring impedance (input unit). The different discharge sources are generated by artificial defects introduced to high voltage equipment under controlled conditions in order to form a database of different PDs. As explained in the next section this database consists of PD sources that exists in field conditions and the purpose of this database is to be analysed in the following chapter where a reduced set of features will be deduced based on this experimental database. Once this is evaluated on experimental data then field data are employed to evaluate the identification performance of the proposed feature. The different PD sources that form part of this database are described next.

### 2.4.1 Artificial PD Sources

Taking into account the majority of discharges found in cable networks and their accessories, four types of discharges are considered, namely corona discharge, floating discharge, internal discharge and surface discharge. The occurrence of internal discharges is the most common encountered in power cables and can be present in XLPE and PILC cables, as well as their accessories. They can occur in the metal semiconductor interface due to the result of the different thermal expansion coefficients of the two materials and thermal loading. The surface discharge is more common in cable joints close to the field grading cone, when for example there is a tracking along an interface. The floating discharge requires the inclusion of a metal particle in the dielectric and is more unusual to exist in power cables but can exist in nearby transformers or GIS equipment and propagate into cables. Similarly a corona discharge can exist where sharp object are subject to an electric as for example a protrusion on a high voltage electrode surrounded by air or a loose screw, and do propagate into cables from nearby equipment. The arrangement shown in figure 2.8 is used to generate these PDs and each set-up is explained briefly below.

Before analysing each discharge type it is noteworthy to mention that the set-up for the internal and the floating discharge types are immersed in mineral oil. This can be explained by the fact that the inception voltage of both corona and surface discharges in air is much lower than the respective inception voltage of the floating and the internal discharges in air. Immersion in oil increases the inception voltage of the corona and surface discharges and therefore the inception voltage of the floating and the internal discharges can be reached without the unwanted occurrence of any corona or surface discharges.




(a) Internal discharge in oil

(b) Floating discharge in oil

(c) Surface discharge in air

(d) Corona discharge in air

Figure 2.8: Experimental setup for generating PDs of different types

Internal discharges are by far the most common family of discharges in the cable dielectric which can be further subdivided into categories depending whether an insulating or conducting interface is involved. Figure 2.9 shows common types of internal discharges in a cable which includes protrusions at the semiconductive layers, microvoids in the dielectric and conducting or non-conducting inclusions. To simulate an internal discharge in a gas-filled inclusion of low dielectric strength (microvoid in the dielectric) a void of 5 mm (diameter) by 1 mm (depth) is embedded between two pieces of perspex, which is placed between two symmetric planar electrodes. The HV source is connected to the upper electrode and the lower electrode is earthed while the whole arrangement is immersed in transformer oil, as shown in figure 2.8a. The recorded environmental conditions were recorded prior to commencing the data acquisition and found to be 18°C at a pressure of 761 mmHg.

Floating discharges are examined next as they can be considered a special case of internal discharges. As shown in figure 2.9 internal discharges can occur at the presence

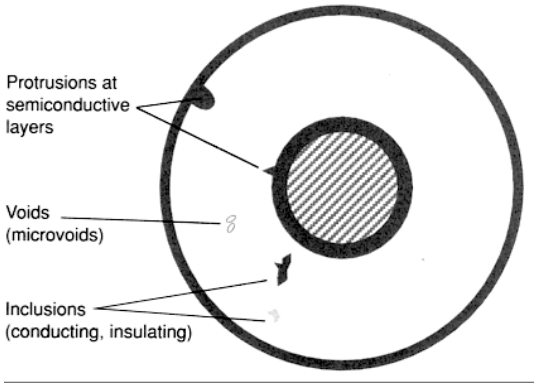



Figure 2.9: Common defects in polymeric cables (taken from [2])

of contaminating particles. When these contaminating particles are conductive, the resulting discharge is included here as a separate discharge, the floating discharge. Floating discharge in oil is simulated by inserting a 5 mm thick Perspex sheet with coils of tinned copper wire on its surface between the two planar electrodes and the whole arrangement immersed in oil (figure 2.8b). The upper electrode is at a HV potential while the lower electrode is earthed, and the temperature is 20.3°C and the pressure 743 mmHg. When a voltage is applied to the arrangement the copper wires are at a floating potential and thus the name for this discharge.

Surface discharges can exist in cable accessories for example in cable joints or terminations and generally where dielectrics of different permittivity are joined together or at the end of a conductor-insulator interface. Surface discharges occur when there exists a stress component parallel to the dielectric surface. These discharges affect the electric field extending beyond the region where the discharge originated from. To simulate surface discharge behaviour, a perspex block is inserted between a pair of electrodes, the upper electrode is connected to the high voltage power supply, and the lower electrode is grounded, as shown in figure 2.8c. The temperature prior to performing the experiment is 21°C and the pressure 761 mmHg. The upper electrode is sharp, as this reduces the inception voltage compared to a planar electrode. Application of a HV creates a discharge along the insulator block that propagates along the insulator surface, and is thus considered as a conductor-insulator discharge. Electron avalanches now start both in the gas (with an electron ionisation coefficient $\alpha$) as well as along the insulator surface (with an electron ionisation coefficient $\alpha_s$).

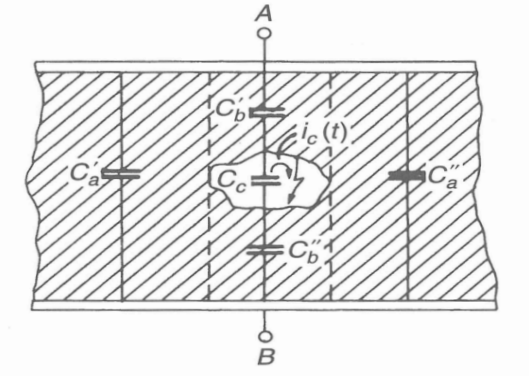
Corona discharges are not usually found in cables but they can propagate from neighbouring equipment or from overhead lines and be detected in PD measurements. Corona

occurs at sharp points in gasses, usually at the high voltage electrode. The distance between the electrodes is large but due to the field enhancement at a sharp point, a partial breakdown does occur. In order to simulate the corona discharge, a piece of thin aluminium wire is suspended from the high voltage conductor, as shown in figure 2.8d, with ambient conditions being 15.5°C at a pressure of 748mmHg. Even though the distance between the electrodes is large, the field concentration at the sharp edge of the HV electrode causes a partial breakdown of the surrounding air. The length of the wire, as well as the radius of the sharp point, have a significant effect on the inception voltage, and by adjusting these two parameters different PD inception voltages can be realized. Due to the distribution of space charges negative coronas (when the HV electrode is at a negative potential) appear at a much lower voltage than positive coronas, and show a repetitive pattern.

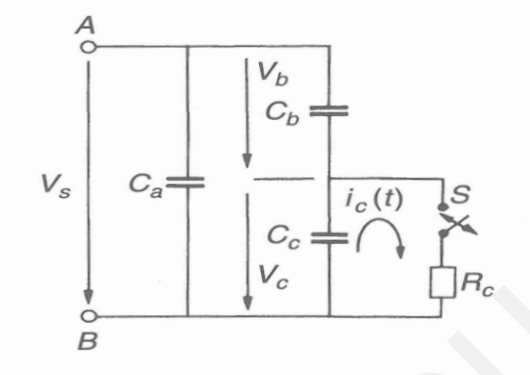
The experimental set-up and measuring equipment is used to acquire phase and time-resolved data from the PD sources described above in order to build a database of different PD sources that will be used in the development of a feature extractor for PD characterisation, and its assessment using classification algorithms. Although the dielectric material, the distance between the electrodes, the ambient humidity and temperature all have an effect on the amplitude of the discharges, their effect is not investigated in this work. Before proceeding to examine the acquisition of the various PD sources the whole set-up needs to be calibrated in order to be able to record the apparent charge for each discharge. The notion of apparent charge plays a crucial role in quantifying the magnitude of the discharges and is explained below.

### 2.4.2 Apparent charge

The actual charge transfer locally involved at the PD site is the variable of interest but since this is not an observable quantity the apparent charge is used instead. It is a quantity that can be related to the actual charge at the site and can be understood through the so called a-b-c model of PDs using capacitors to approximate PD currents at the terminals of laboratory test objects. The electric field distribution across the terminals (A-B) is simulated by partial capacitances as shown in figure 2.10a, provided that no space charges are present to disturb this distribution. The field lines inside the cavity are simulated by $C_c$ while the lines above and below the cavity wall to the upper and lower electrode by $C_b = C_b' C_b'' / (C_b' + C_b'')$. The field distribution on the rest of the specimen outside the cavity is represented by $C_a'$ on the left and $C_a''$ on the right. These




(a) Capacitances used in the model




(b) Electrical model of PD across the test specimen using capacitors

Figure 2.10: PD explanation through an electrical model using capacitors (reproduced from [3])

two capacitances are in parallel and can therefore be represented by $C_a$ $(= C_a' + C_a''')$. All of the above capacitances, provided that the field distribution inside them is uniform, can be approximated by the well known parallel plate capacitor equation

$$C = \epsilon_r \epsilon_0 \frac{A}{d} \tag{2.1}$$

where $A$ is the area of overlap and $d$ the separation between the two plates. With reference to this equation and taking into account the realistic geometries involved we have $A_c = A_b$ and $d_b \gg d_c$ which gives the relation $C_c \gg C_b$. Considering the dimensions involved in $C_a$ it can be deduced that $A_a \gg A_c$ which translates into $C_a \gg C_c$. Combining all the above relations the following inequality relates the magnitudes of the capacitances involved

$$C_a \gg C_c \gg C_b. \tag{2.2}$$

PD phenomena can be simulated by the equivalent circuit shown in figure 2.10b. The voltage controlled (by the voltage across the capacitor $C_c$ given by $V_c$) switch is closed for a short period of time, during which the capacitor discharges through resistor $R_c$. The shape of the current through this resistor, $i_c(t)$ is related to the discharge process but unfortunately it is not a measurable quantity. Instead by employing the principle of conservation of charge before and after the discharge occurrence it can be related to the voltage drop across the terminals $\delta V_a$. Assuming that the sample was charged to $V_a$ and when the switch closes the terminals are removed from the voltage source $(V_s)$ then

the current $i_c(t)$ releases a charge $\delta q_c = C_c \delta V_c$, provided $C_c$ completely discharges. As a result a voltage drop $\delta V_a$ appears across the terminal given by

$$\delta V_a = \frac{C_b}{(C_a + C_b)} \delta V_c. \tag{2.3}$$

This quantity is measurable and proportional to $(C_b \delta V_c)$, which can be vaguely related to the charge released by current $i_c(t)$, $\delta q_c$. This is achieved through the increase of $C_c$

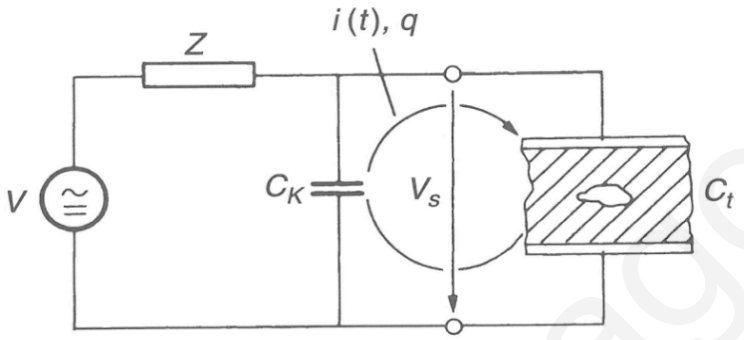



Figure 2.11: PD test circuit where $C_t$ is the total capacitance across the specimen (taken from [3])

with the geometric dimensions of $C_b$. However the magnitude of $\delta V_a$ will be rather small (the ratio $C_b/C_a$ will be very small) and practical problems in measuring such small magnitudes exist, let alone the fact that this ratio is not known. Therefore practical detection circuits are based on the current $i(t)$ and charge $q$ from the circuit shown in figure 2.11. The specimen in figure 2.10a is now connected to an ac voltage source through an impedance Z and a 'coupling capacitor' $C_K$ has been placed in parallel with the test specimen $C_t$. During the short duration of the PD current pulse, $C_K$ can be regarded as quite a stable voltage source, releasing a charging current, $i(t)$, between $C_K$ and $C_t$ to cancel the voltage drop $\delta V_a$ across $C_t \approx (C_a + C_b)$. If $C_K \gg C_t$, $\delta V_a$ is completely compensated and the charge transfer provided by $i(t)$ is given by

$$q = (C_a + C_b)\delta V_a = C_b \delta V_c \tag{2.4}$$

through application of equation 2.3. This charge is the so called apparent charge, a fundamental quantity in PD measurements. It is not the actual charge involved locally at $C_c$ but a more realistic quantity, independent of $C_a$, that measures the interaction of dipole moments of local charge with the electrodes.

However for finite values of $C_K$ the voltage across it will also drop during the PD, by a value $\delta V^*$. Assuming that the same charge has been transferred in the PD test circuit and the PD test object ($i_c(t) = i(t)$) we have the charge transferred from $C_K$ to $C_t$ by the reduced $i(t)$, $q_m$ given by

$$q_m = C_K \delta V^* = \frac{C_K}{C_a + C_b + C_K} q \approx \frac{C_K}{C_a + C_K} q. \tag{2.5}$$

This charge is related to $q$ by a constant of proportionality which needs to be calculated for a given experimental arrangement by a calibration procedure in order to get indicative magnitudes of the various PD sources and establish the noise floor of the arrangement in coulombs.

### 2.4.3 Calibration of experimental arrangement

To quantify the apparent charge during a PD event the Robinson detector is calibrated using two injected pulses of known charge (50 or 500 pC) from a LDIC LDC-5RUF UHF calibrator. Once the output of the Robinson detector is recorded on a channel of the oscilloscope, the HFCT is connected to another channel (as shown in the actual experimental setup in figure 2.12). The HV power supply is switched on and raised to the desired voltage level, generating PD activity. Both channels (HFCT and Robinson detector) are displayed and recorded giving quantitative information about the apparent charge in the PD when compared to the reading of the calibration pulse.

The measurements taken during the acquisition of an internal discharge at an applied voltage of 26 kV are shown in 2.13. A 50 pC charge is first injected into the set-up of figure 2.7 at the UHF calibrator point and the recorded pulse at the output of the Robinson detector (channel 1, yellow trace of the oscilloscope) is shown in figure 2.13a. From this figure it can be seen that a charge of 50 pC produces a reading of 1.9 Volts on the channel measuring the output of the Robinson detector (for the pulse at the positive half of the voltage). Without changing the experimental set-up or the Robinson detector settings the HFCT is connected to channel 2 (blue trace), the HV power supply is switched on and the voltage raised to 26 kV. At this level the two channels are recorded simultaneously (figure 2.13b) and the reading of the output of the Robinson detector is mapped to picocoulombs according to the previous calibration (the measured 1.8 V map to $1.8/1.9 \times 50 \approx 47$ pC). Subsequently the reading of the HFCT channel is read to be 8.1 mV which translates to approximately 17.23 mV/100 pC. Following the same procedure for all the PD sources and voltage levels indicative values in pC are obtained

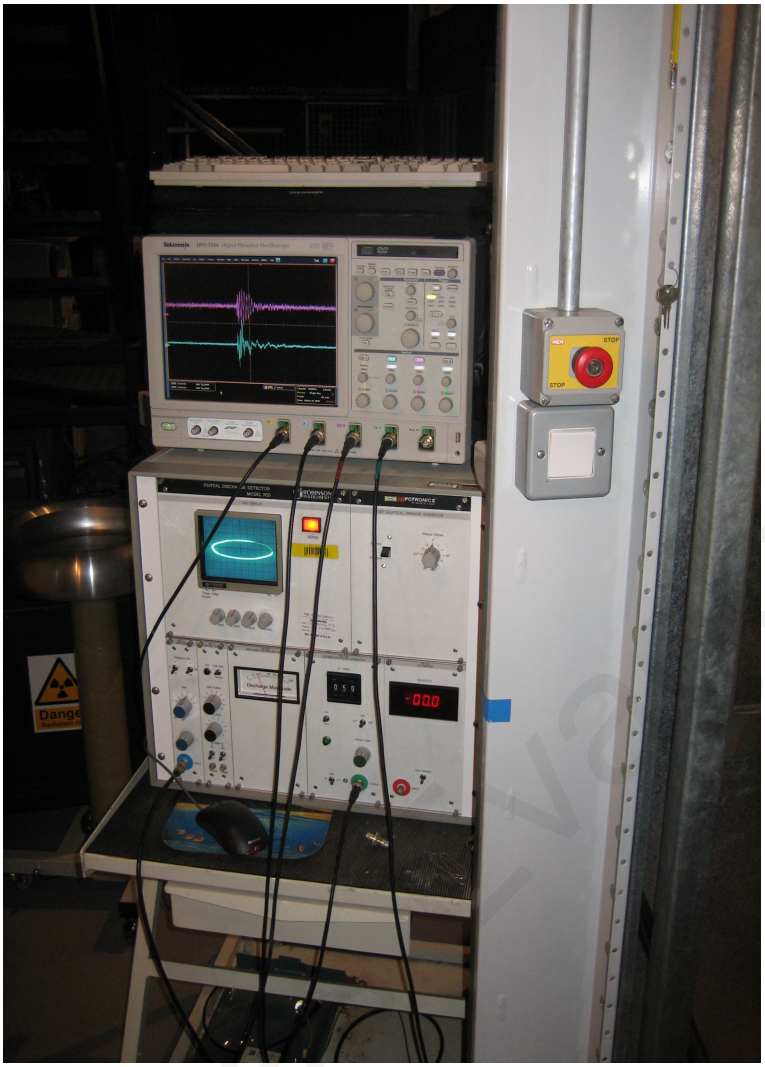
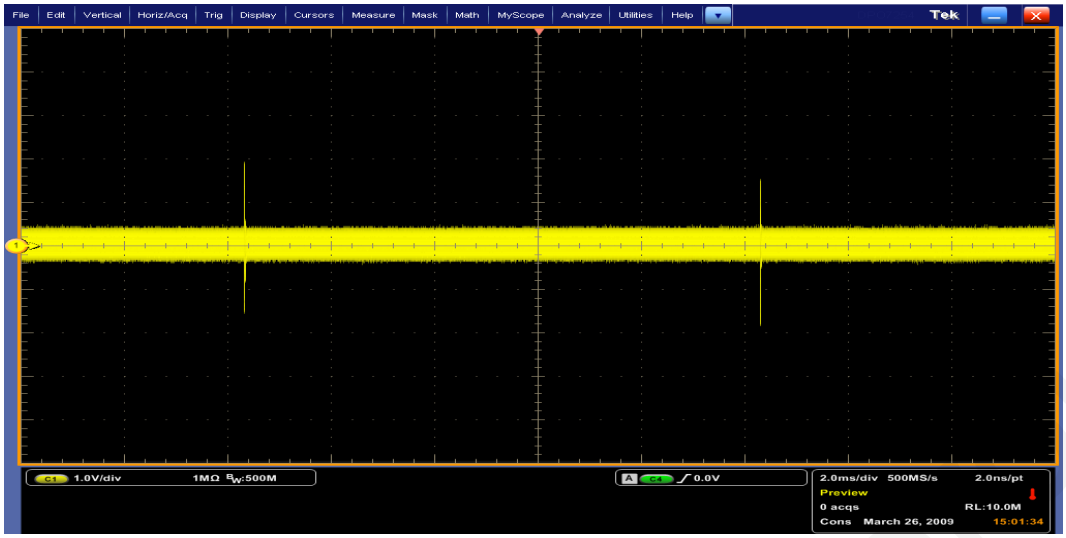
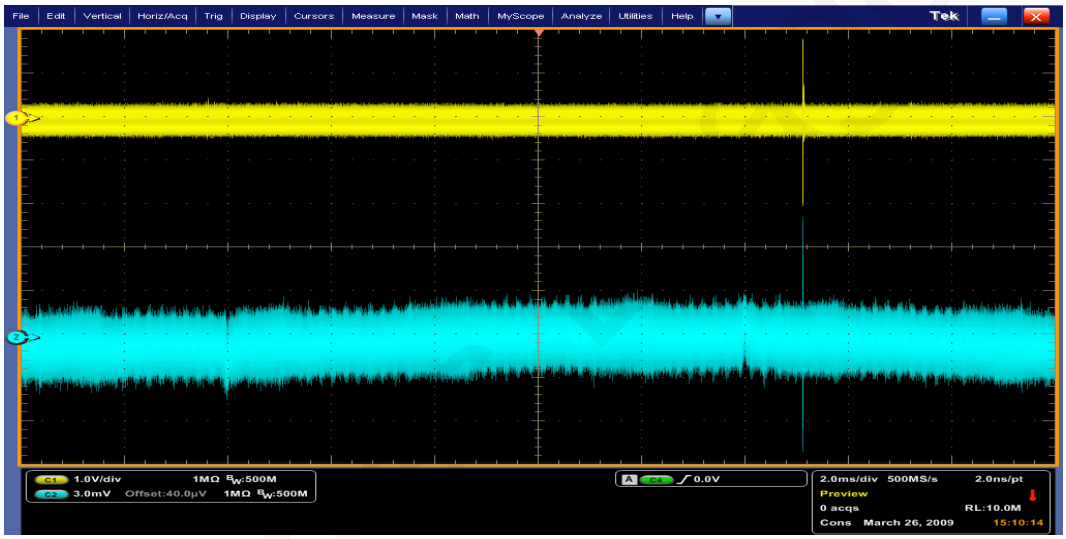


Figure 2.12: Robinson detector (bottom) and oscilloscope (top) setup to measure the output of the detector and the reading from the HFCT on different channels

for the HFCT sensor readings. Typical measurements of the various discharge types, show that the captured data suffers from background noise which affects the sensitivity of the measurements. Discharge activity having magnitude greater than 45 pC can be reliably detected by the measurement system, while discharge activity with a charge lower than this is buried in background noise, as shown from the acquired data in the following section.




(a) Robinson detector output for 50 pC pulse calibration




(b) Robinson detector and HFCT sensor output for an internal PD at an applied voltage of 26 kV

Figure 2.13: (a) Plot of the oscilloscope reading for the Robinson detector calibration pulse in and (b) simultaneous reading of the Robinson detector and the HFCT sensor output for an internal discharge

## 2.5 Data acquisition

The experimental arrangement analysed in section 2.5 is used to acquire phase as well as single pulse data for the four PD sources at different applied voltages. This data forms a database of different PDs that will be analysed in order to extract a representative feature vector, able to compactly describe the characteristics of each PD. Subsequently the dataset will also be employed in evaluating the proposed feature using the classification

performance as a criterion.

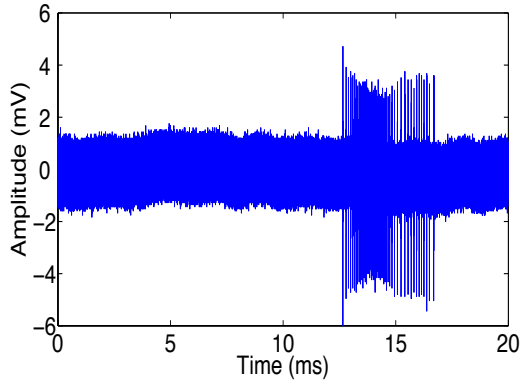
Data acquired in synchronisation with the ac cycle can be used to assess the recurrence of discharges as well as to extract the so called Phase Resolved Partial Discharge (PRPD) data representations. These representations constitute one of the most complete representations of PD data since they can be directly related to physical aspects of the discharge such as for example the symmetry of a void discharge. The sampling rate is chosen to be 500 MHz, as the -25 dB high frequency point for the HFCT used is around 200 MHz. To reduce the amount of data the peak detection function of the oscilloscope is used in the phase synchronised acquisition.

### 2.5.1 Phase data using Peak Detection

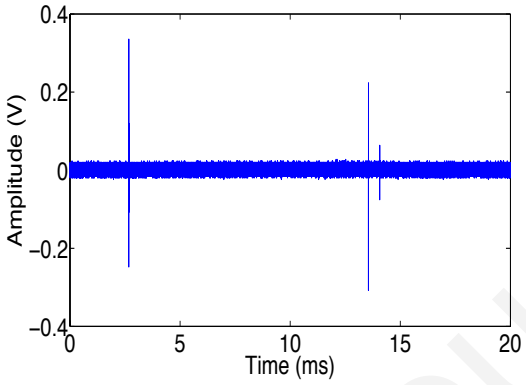
A broadband HFCT sensor of 200 MHz useful bandwidth implies a sampling frequency of at least 500 MHz. If used for PD detection, such an approach will very quickly create a significant volume of data that would introduce processing delays which are not generally acceptable. Therefore the envelope of the PD discharge signal measured by the HFCT is captured using the digital storage oscilloscope's peak detect function. The peak detection function of the oscilloscope is useful for catching high frequency glitches but at a lower sampling rate. This is achieved by internally sampling at a high enough frequency so as to avoid aliasing but only recording the highest and lowest of all the samples in two consecutive intervals. This approach reduces the amount of data stored per cycle of applied voltage by a factor of $10^3$. A key issue with such an approach, is whether post-processing of the enveloped data can discriminate between PD sources since information about the individual pulses is lost using this acquisition mode. The use of peak detection to reduce the volume of data allows the post processing of whole cycles of data which are used in the calculation of traditional PD descriptors such as the pulse count distribution,($\phi - n$ plot, $H_n(\phi)$) and the mean pulse height distribution ($\phi - q - n$ plot, $H_{qn}(\phi)$) [58]. Several cycles, at a sampling rate of 500 kS/s using the peak detection, for different applied voltages and in synchronisation with the 50 Hz AC mains frequency, are captured as seen in table 2.2 while plots for the different PD sources are shown in figure 2.14.

### 2.5.2 Single Pulse data

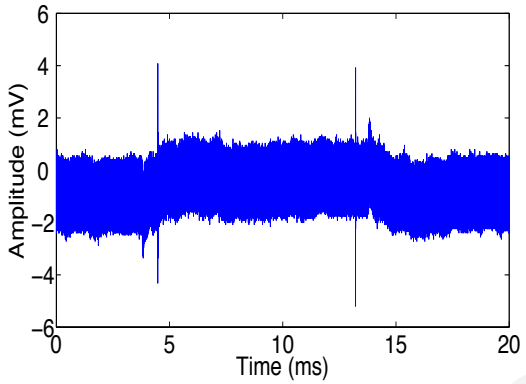
Although phase resolved measurements can be utilised to extract patterns related to the physics of the discharge, when applied to on-line data they suffer from noise and



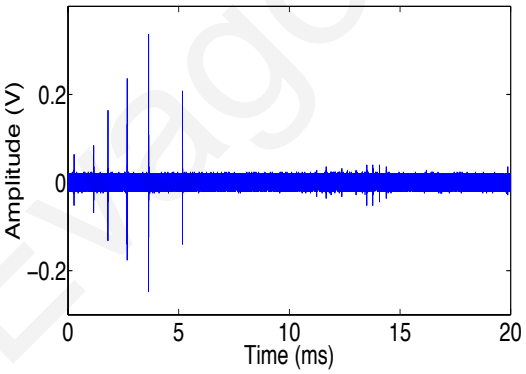

(a) Corona discharge in air at an applied voltage of 9 kV. From calibration of the Robinson detector, the sensitivity of this measurement is 11.25 pC/mV.




(b) Floating discharge in oil at an applied voltage of 23.4 kV. From calibration of the Robinson detector, the sensitivity of this measurement is 2.47 pC/mV.




(c) Internal discharge in oil at an applied voltage of 24.5 kV. From calibration of the Robinson detector, the sensitivity of this measurement is 10.75 pC/mV.




(d) Surface discharge in air at an applied voltage of 9 kV. From calibration of the Robinson detector, the sensitivity of this measurement is 3.68 pC/mV.

Figure 2.14: One cycle phase-resolved measurements for various PD types, at different voltages sampled at 500 kS/s (using the peak detection function)

interferences. Measured quantities make no distinction between a true PD and a pulse shape interference. Moreover the occurrence of multiple discharge sources simultaneously active in on-line measurements further complicates the extracted patterns. Since the success of PRPD methods is based on laboratory measurements where the assumption that a single PD source is active and no interferences present such methods find limited applicability in field measurements. However using signal processing techniques to pre-process the data enables the extraction of homogeneous (in the sense that they consist of a PD from a single source and is interference free) phase patterns and thus the use of PRPD. The signal processing part is done on single pulse data sampled at a sufficiently

Table 2.2: Details of simulated PD data acquired in synchronisation with phase and using the peak detection function of the oscilloscope (500 kS/s).

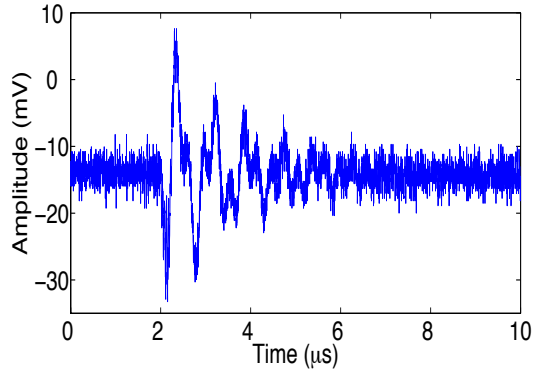
| **Type** | **Voltage** | **No of cycles** | **Sample length (per cycle)** |
|---|---|---|---|
| Corona in air | 7 kV | 500 | 9992 pts |
| Corona in air | 8 kV | 500 | 10 002 pts |
| Corona in air | 9 kV | 500 | 10 011 pts |
| Corona in air | 11 kV | 500 | 9998 pts |
| Floating in oil | 23.4 kV | 500 | 10 000 pts |
| Floating in oil | 25.5 kV | 500 | 10 000 pts |
| Internal in oil | 21 kV | 500 | 10 000 pts |
| Internal in oil | 24.5 kV | 500 | 10 000 pts |
| Surface in air | 6 kV | 500 | 10 000 pts |
| Surface in air | 9 kV | 500 | 10 000 pts |

high sampling rate, where features such as the rise and fall times relate to the physics, as well as the location of the PD. Single pulse signals are acquired in batches of 500 pulses per file and 10 files are recorded for each applied voltage. This generates 5000 pulses per applied voltage as shown in table 2.3. Plots of randomly chosen time-resolved data are presented in figure 2.15 where the different waveforms produced by each PD source can be seen.

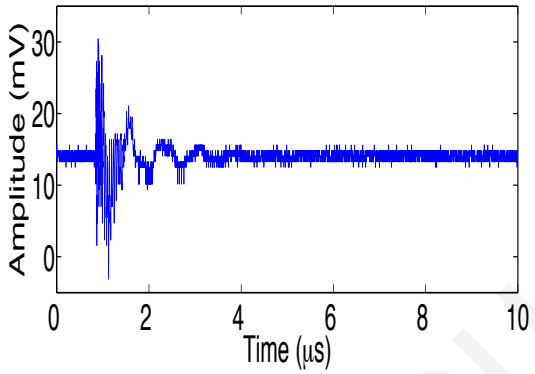
The acquired phase and time-resolved data form a database of PD sources that will be analysed in the subsequent chapters in order to propose and evaluate feature vectors of low dimensionality that can find application in on-line measurements.

## 2.6 Conclusions

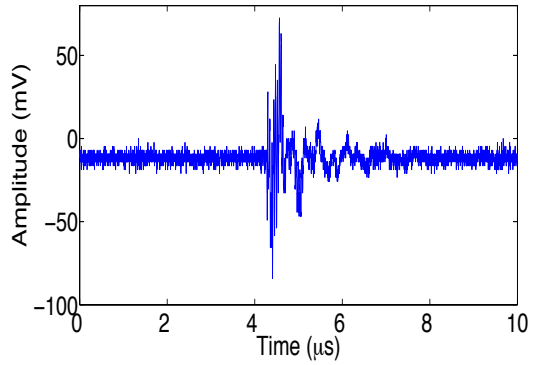
In this chapter the occurrence and recurrence of PD activity was briefly reviewed followed by an explanation of the different types of cables and their effect on PD measurements. Next the frequency response of the HFCT sensor used in the experiments was measured and found to be adequate for use in the experiments. The equipment used and the experimental set-up to build a database of different sources of PD was explained. The database consists of phase and single pulse data, as they complement each other. The



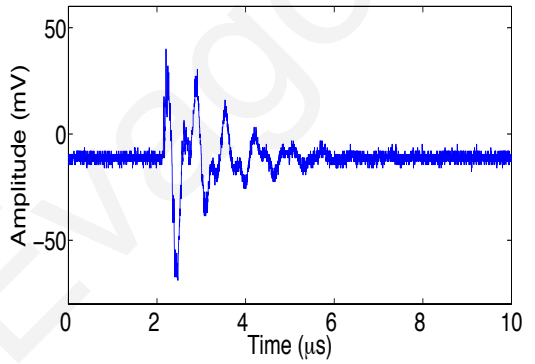

(a) Corona discharge in air at an applied voltage of 6 kV. From calibration of the Robinson detector, the sensitivity of this measurement is 3.3 pC/mV.




(b) Floating discharge in oil at an applied voltage of 24 kV. From calibration of the Robinson detector, the sensitivity of this measurement is 0.12 pC/mV.




(c) Internal discharge in oil at an applied voltage of 26 kV. From calibration of the Robinson detector, the sensitivity of this measurement is 1.27 pC/mV.




(d) Surface discharge in air at an applied voltage of 6 kV. From calibration of the Robinson detector, the sensitivity of this measurement is 3.39 pC/mV.

Figure 2.15: Single pulse data measurements for various PD types, at different voltages sampled at 500 MS/s

data collected here forms the basis for the following work and in the next chapter two feature extraction methods are proposed.

Table 2.3: Details of simulated single pulse PD data acquired at a sampling rate of 500 MHz

| **Type** | **Voltage** | **Sample Duration** | **Sample length** | **Sample Quantity** |
|---|---|---|---|---|
| Corona in air | 6 kV | 10 $\mu$s | 5000 pts | 5000 |
| Corona in air | 9 kV | 10 $\mu$s | 5000 pts | 5000 |
| Floating in oil | 24 kV | 10 $\mu$s | 5000 pts | 5000 |
| Floating in oil | 28 kV | 10 $\mu$s | 5000 pts | 5000 |
| Internal in oil | 26 kV | 10 $\mu$s | 5000 pts | 5000 |
| Internal in oil | 36 kV | 10 $\mu$s | 5000 pts | 5000 |
| Surface in air | 6 kV | 10 $\mu$s | 5000 pts | 5000 |
| Surface in air | 9 kV | 10 $\mu$s | 5000 pts | 5000 |

# Chapter 3

# Feature Extraction

In this chapter a feature vector of low dimensionality is proposed, which reduces the storage requirements and computational burden of the classification algorithm. Two variants of the proposed feature are presented, one using time and another using phase resolved data, as the use of both type of data complements each other facilitating the PD identification method. In order to deduce a feature that best characterises each discharge with a minimum number of dimensions, the data collected under laboratory controlled conditions from various sources of PD in chapter 2 are analysed. To tackle this lower dimensionality constraint the Wavelet Packet Transform (WPT) and Higher Order Statistics (HOS) are employed. The WPT expands the signal using an adaptive time-frequency decomposition, delineating characteristics that are not immediately obvious from a representation in the time domain. The selection of specific time-frequency regions that contain most of the information in the signal and the use of HOS as a statistical descriptor of the wavelet coefficients in these tiles reduces the dimensions of the characterisation. The reduction in the number of dimensions achieved through the combination of the above tools can find potential application in continuous on-line PD monitoring systems. Such systems enable trending analysis of the measured PD data that can reveal important information about the severity of the deterioration caused by the discharge.

## 3.1 Literature Review

The traditional goal of a feature extractor is to characterize an entity to be recognised by very similar values for objects in the same category but very distinct for different categories [59]. Some of the most important descriptors of the discharge pulses are their amplitude, rise time, recurrence rate, phase relationship of occurrence with respect to

the applied alternating voltage and time interval to the preceding pulses [20]. As some of these attributes are extracted from time-resolved while others from phase-resolved data, different feature extraction methods [60] are broadly categorised according to the representation utilised.

Both phase and time-resolved (also called single pulse) data can be related to aspects of the physics of the discharges and this can explain their extended use in attempts to describe PD behaviour. Phase-resolved data representations give information about the phase of occurrence of PD activity and can provide the intervals between successive discharges. Such information can be related to electron availability or the decay time constants of space charges. On the other hand time-resolved data takes into account quantities such as the rise-time giving an insight into the discharge mechanisms involved in the emission of electrons at the cathode [18, 20]. The two approaches are reviewed in the following two sections and as phase-resolved representations were historically the first to be employed in PD analysis they are reviewed first.

### 3.1.1 Phase-Resolved Representations

Phase-resolved data involves acquisition of several cycles in synchronisation with the mains frequency. One of the early and most successful analysis methods employed are the so called $\phi - q$ and $\phi - n$ representations where the phase axis is divided into small bins and extracted quantities are plotted against it. In the former case the peak ($q_p$) or the average ($q_m$) charge while in the latter the number of discharges, $n$, per phase window are plotted as a univariate distribution against the phase. Dividing the extracted quantity (for example the maximum charge) into windows as well, allows the calculation of the number of discharges for each phase and charge window. This can be plotted as a bivariate distribution against the charge and the phase, giving the 3-dimensional $\phi - q - n$ plot, one of the most complete PD data representations [21–27]. These representations have traditionally been referred to as Phase Representation PD (PRPD), a terminology that is adopted here.

Although these 2 or 3-dimensional plots can be utilised in themselves as feature vectors, their dimensionality depends on the product of amplitude and phase bins. In an attempt to reduce the number of dimensions the PRPD were used to extract features. Towards this end two main paths were followed, the descriptive and the physical modelling. In the descriptive approach mathematical and data analysis tools were employed to derive a set of values that describe the PRPD (or the data series), while in the physical

modelling approach a physical model or a set of equations was employed to analyse the observed phase-resolved data series.

Statistical analysis tools were one of the first methods applied successfully on phase-resolved data to get descriptive parameters for the 2-dimensional PRPD distributions and many such attempts can be found in the literature [21, 22, 24, 27, 28]. For example in [28] the complete feature vector was derived from the estimation of the skewness and kurtosis of the positive and negative half cycle of the $\phi - q_m$ and $\phi - n$ distributions as well as other statistical operators such as the correlation coefficient. This 9-dimensional feature was tested on sixteen full scale samples and showed a good potential for classification of discharging defects in insulation structures.

In addition to statistical analysis tools, signal processing on 2-dimensional PRPD or image processing on 3-dimensional PRPD were employed. The $\phi - q - n$ was transformed into a 2-dimensional image where one of the quantities (usually $n$) was represented by the pixel values of the image. Tools such as the 2D wavelet transform [26] or fractal analysis [29, 30] were used to extract descriptive features.

The utilisation of statistical, image, or signal processing tools results in a set of parameters that are not directly physically interpretable. For example a change in the applied voltage can result in a significant change in the value of the parameter in spite of the fact that the physics of the discharge remains the same. To extract more meaningful descriptors, whose interpretation is directly related to the physics of the actual discharge, attempts to employ physical models were undertaken. However very few approaches exist in the literature, possibly due to the complexity of the PD process and its dependence on the initial conditions.

A stochastic model was considered in [1] where three physical parameters (the initiation probability, the decaying function of the internal field, and the residual field) were used to describe an observed phase-resolved series. Another physical model approach is the Pulse Sequence Analysis (PSA) [61]. It relies on a plot of the phase difference between two successive discharges at time n ($\Delta U_n$) against ($\Delta U_{n-1}$) as a parameter set. One of the benefits of this approach is that it takes into account the space charges developed after the PD has been initiated and the phase shift as a result of the presence of these space charges. However a general drawback of these methods is that due to the complexity of the discharge process it is difficult to estimate the physical parameters directly from the measurements and a trial and error procedure must be followed.

In this research the descriptive approach is used, despite its aforementioned drawback, due to its simplicity in estimating a set of descriptive parameters from the measurements.

The WPT and HOS are combined to derive a set of features and detailed explanation of the phase-resolved proposed feature is given section 3.5.

PRPD distributions produce interpretable representations under the assumption that the recorded data is free from interference and there is only one source of PD active. However in on-line measurements these two assumptions do not hold which results in descriptors extracted from PRPD distributions and physical parameters from phase-resolved series, that will not be able to adequately characterise the PD process. Fortunately phase-resolved methods can still be employed provided that a pre-processing step to separate different sources of PD is applied [37]. Single pulse data implements this based on pulse shape analysis and is reviewed in the next section.

### 3.1.2 Time Resolved Representations

In time-resolved measurements data representing the waveforms of single pulses are captured for a sufficiently long time at a sufficiently high sampling rate. The "sufficiently long" requirement is needed in order to capture quantities such as the rise-time, fall-time and waveform of the PD without truncating them, and a "sufficient sampling rate" ensures that no distortion will be introduced due to the digitisation (sampling). A variety of discriminatory features derived from the actual measurement or by the application of signal processing techniques exist in the literature, and publications relevant to the work in this thesis are reviewed.

Time and frequency content of time-resolved PD signals is related to the type of PD as different sources of PD produce different time and frequency characteristics. Moreover, the low-pass characteristics of underground power cables result in distortion as a function of the distance travelled, affecting the frequency content of the pulse. Adding the two, time or frequency analysis of PD signals can result in identifying PDs, under the assumption that PDs originating from the same source will have similar time and frequency characteristics.

To analyse the frequency content of a PD signal the Fourier Transform was employed in [35] to extract features to identify fixed (on conductors or spacer surfaces) or free moving particles in $SF_6$ gas-insulated substations (GIS). However one of the weak points of the Fourier Transform (FT) is that its basis functions extend over the whole time interval without any localisation properties, and therefore time information is completely lost. To overcome this problem the Short Time Fourier Transform (STFT) is employed that uses a fixed window to capture the signal allowing a simultaneous time-frequency analysis of the signal.

An implementation of the above transform can be found in [36–38], where the signal was captured in windows of $K$ samples. The acquired pulses were mapped into a time (T) and frequency (F) equivalent bandwidth calculated as the second central moment of the square of the time and frequency coefficients of the data, creating a time-frequency representation of the energy of the signal. However the STFT employs a fixed time-frequency plane, whose area is limited by the Heisenberg uncertainty principle. Wavelet Analysis (WA) overcomes this limitation by providing a time-frequency representation that is dependent on its scaling parameters.

The Wavelet Transform, a discretisation of WA, provides a variable time-frequency decomposition that delineates signal characteristics not immediately obvious in the original signal. The WT was firstly applied in denoising PD signals [62–64] but was extended to the utilisation of the wavelet coefficients as features [39].

The Wavelet Packets Transform, an extension of the WT, expands the signal into a set of adaptive bases (time-frequency tiles adapted according to the characteristics of the signal) that potentially results in a more compact representation. This transform is employed here and a procedure is proposed where the nodes representing the characteristics of the different PD sources are selected. The wavelet coefficients of these nodes constitute descriptors and can be used as features. However in order to reduce the dimensionality of the representation statistical analysis is employed to approximate these coefficients.

The statistical analysis of PD pulses is able to capture the variability in PD properties. Such a method was employed in [42] on time-resolved measurements, where different PD sources were identified on the basis of the shape parameter of the Weibull distribution of the pulse height. The Autocorrelation Function (ACF) was also used in [43] for clustering time-resolved data (identified by the K-means clustering algorithm). The utilisation of statistical analysis using moments was confined up to the second order and the use of Higher Order Statistics (HOS) received limited attention. The application of HOS on time-resolved data was examined on acoustic and electric data taken from experimental measurements of contaminating particles in a tank of oil [44].

This research attempts to employ the concept of statistical analysis through the application of HOS (up to fourth order) on selected wavelet coefficients, in order to explore the potential of the combination of WPT and HOS methods on electrical measurements in power cables. To the author's best knowledge the only study in the literature where similar work was performed on time-resolved data is [41] but substantial differences exist between that work and this thesis, such as the selection of the WPT nodes, the HOS and the number of PD samples used. As both time and phase-resolved proposed features

employ these analysis tools, they are theoretically analysed in the next section, starting from Wavelet Analysis.

Using the wavelet analysis signals with form changes, characteristic of those generated by PDs, can be better analysed with a decaying oscillatory wavelet signal (due to the admissibility condition) than with a uniform Fourier sinusoid of infinite extend. The theoretical analysis starts with a redundant transform, the Continuous Wavelet Transform (CWT) to explain how the wavelet coefficients are calculated and why they can enhance specific characteristics of a signal. Redundant approximations employ dictionaries of vectors that are larger than bases to build sparse representations of complex signals.

### 3.1.3 The Continuous Wavelet Transform (CWT)

The continuous wavelet transform of a signal $x(t)$ can be considered as its correlation (or inner product) with a wavelet function. The transform is given by

$$\mathcal{W}x(a,b) = \int_{-\infty}^{+\infty} x(t)\psi^*_{a,b}(t)dt = \langle x, \psi_{a,b}\rangle \tag{3.1}$$

where

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{3.2}$$

is the wavelet function which is a function $\psi \in \mathbf{L^2}(\mathbb{R})$ that has to fulfil certain mathematical properties [65]. For real wavelets it has to have a finite energy and obey the admissibility condition given by

$$C_\psi = \int_0^{+\infty} \frac{\|\hat{x}(\omega)\|^2}{\omega} d\omega < \infty \tag{3.3}$$

where $\hat{x}(f)$ is the Fourier Transform of $x(t)$. The admissibility condition is satisfied if $\hat{x}(\omega) = 0$ and $\hat{x}(\omega)$ is continuously differentiable. It can be verified that $\hat{x}(\omega)$ is continuously differentiable if $\psi$ has a sufficient time decay:

$$\int_{-\infty}^{+\infty} (1+|t|)|\psi(t)|dt < \infty. \tag{3.4}$$

The above conditions translate into a real wavelet being a decaying oscillatory function with zero mean. Introducing a scaling function $\phi(t)$ allows the aggregation of wavelet at scales larger than $a_0$ ($a > a_0$) and reduces the number of scales needed to represent the

signal. This scaling function is defined by

$$|\hat{\phi}(\omega)|^2 = \int_{\omega}^{+\infty} \frac{|\hat{\psi}(\xi)|}{\xi} d\xi. \tag{3.5}$$

Also $\|\phi\| = 1$ and from the admissibility condition

$$\lim_{x \to 0} |\hat{\phi}(\omega)|^2 = C_{\psi} \tag{3.6}$$

which points to the fact that the scaling function has to have a low-pass response. The importance of the behaviour of the wavelet function in enhancing the local characteristics of a signal can best be explained with the aid of figure 3.1. The CWT is clearly a function of two variables, one controlling the dilation ($a$) and the other the translation ($b$) of the wavelet. At a particular scale the transform measures the correlation of the wavelet function with the signal $x(t)$ and therefore large coefficients are produced at translations where the local characteristics of the two signals match. Increasing the scale results in

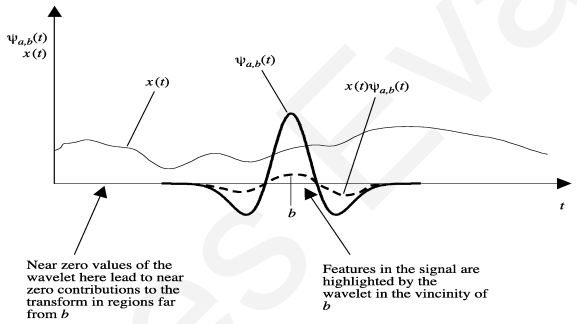
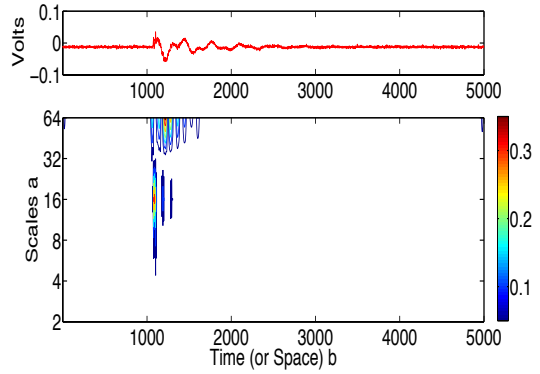
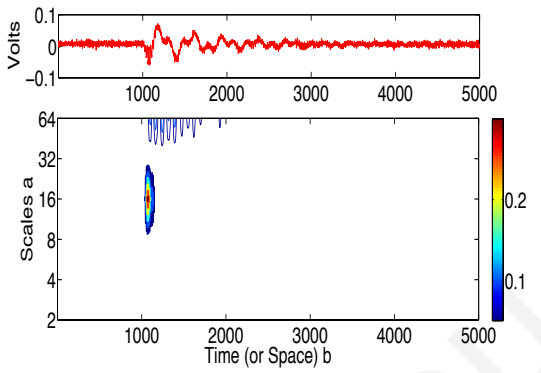



Figure 3.1: Explanation of the capability of the CWT to highlight the local characteristics of a signal based on dilations and translations. The transform is an inner product of the signal with the wavelet function and at dilations and translations where there is similarity between the two large coefficients are being generated (figure taken from [4]).

a dilated wavelet which correlates to different time characteristics of the signal. Since the wavelet at different dilations and locations has a given time and frequency support it results in a continuous time-frequency distribution. The scalogram, which is the energy of the CWT coefficients as a fraction of the energy of the signal, shown in figure 3.2 for two different experimental signals, is such a distribution. Looking at this plot it can be seen that the peaks occur at different scales which indicates different time-frequency content for the experimental surface and corona discharges both at an applied voltage of 6 kV, not immediately obvious from their time plots.




(a) Scalogram for a single pulse experimental surface discharge at an applied voltage of 6 kV




(b) Scalogram for a single pulse experimental corona discharge at an applied voltage of 6 kV

Figure 3.2: Scalogram plots of the CWT coefficients for two different experimental PD sources. The scalogram is given as the energy of the coefficients $P_W x(a,b) = \|\mathcal{W}x(a,b)\|^2$ and the plots here show the $P_W x(a,b)$ values normalised to the signal energy ($\|x(t)\|^2$). The actual time plots of the signals on top of the scalograms show that the CWT delineates differences in the signals to produce different coefficients in the time-frequency decomposition.

The CWT is a redundant representation as the dilation and translation parameters are continuous variables producing a higher number of wavelet coefficients than critically sampled parameters. The Wavelet Transform through a dyadic logarithmic discretisation produces such a representation resulting in a more concise characterisation. It is equivalent to the mathematical theory of Multiresolution Analysis (MRA), which links the transform to signal processing theory of digital filters, where iterative computation through a series of Quadrature Mirror Filters (QMF) is achieved. Next we look at the theory behind this analysis and its implications in processing PD signals.

### 3.1.4 The Wavelet Transform

The Wavelet Transform emerges through a discretisation of the dilation and translation parameters to give a less redundant representation. One such discretisation scheme is the logarithmic which gives the following wavelet function

$$\psi_{j,n}(t) = 1\frac{1}{\sqrt{a_0^j}}\psi\left(\frac{t - nb_0a_0^j}{a_0^j}\right) \tag{3.7}$$

where the integers $j$ and $n$ control the dilation and translation respectively. Setting this logarithmic scaling to a dyadic grid by letting $a_0 = 2$ and $b_0 = 1$ a critical sampling is achieved in the sense that a unique inverse transformation exists. The dyadic wavelet function is given by

$$\psi_{j,n}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t - n2^j}{2^j}\right). \tag{3.8}$$

The formalisation of the link between the discretisation of the CWT and the sampling used can be found in frame theory [65, 66].

By employing Multiresolution Analysis (MRA) the discrete Wavelet Transform (WT) can be calculated in a completely iterative manner, which is ideal for computations. In MRA the approximation of a function $f$ at a resolution $2^{-j}$ is defined as the orthogonal projection $P_{\mathbf{v}_j} f$ on a vector space $\mathbf{V}_j \subset \mathbf{L^2}(\mathbb{R})$. According to the axioms of multiresolution [65, page 264], a sequence $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $\mathbf{L^2}(\mathbb{R})$ is a multiresolution approximation if the following six properties are satisfied

$$\forall (j,n) \in \mathbb{Z}^2, \quad f(t) \in \mathbf{V}_j \Leftrightarrow f(t - 2^j n) \in \mathbf{V}_j \tag{3.9}$$

$$\forall j \in \mathbb{Z}, \quad \mathbf{V}_{j+1} \subset \mathbf{V}_j \tag{3.10}$$

$$\forall j \in \mathbb{Z}, \quad f(t) \in \mathbf{V}_j \Leftrightarrow f\left(\frac{t}{2}\right) \in \mathbf{V}_{j+1} \tag{3.11}$$

$$\lim_{j\to+\infty} \mathbf{V}_j = \bigcap_{j=-\infty}^{+\infty} \mathbf{V}_j = 0 \tag{3.12}$$

$$\lim_{j\to-\infty} \mathbf{V}_j = \text{Closure}\left(\bigcup_{j=-\infty}^{+\infty} \mathbf{V}_j\right) = \mathbf{L^2}(\mathbb{R}) \tag{3.13}$$

and there exists $\theta$ such as $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ is a Riesz basis of $\mathbf{V}_0$. The orthogonal projection on $\mathbf{V}_j$ is included in $\mathbf{V}_{j-1}$ and letting the orthogonal complement of $\mathbf{V}_j$ in $\mathbf{V}_{j-1}$ be $\mathbf{W}_j$ then we have

$$\mathbf{V}_{j-1} = \mathbf{V}_j \oplus \mathbf{W}_j. \tag{3.14}$$

where the $\oplus$ denotes a direct sum of vector spaces. This can be expressed as the sum of the orthogonal projections of $f$ on $\mathbf{V}_j$ and $\mathbf{W}_j$:

$$P_{\mathbf{V}_{j-1}} f = P_{\mathbf{V}_j} f + P_{\mathbf{W}_j} f. \tag{3.15}$$

The complement $P_{\mathbf{W}_j} f$ is the detail of $f$ at scale $2^{j-1}$, which disappears at the coarser

scale $2^j$. To compute these projections an orthonormal basis of $\mathbf{V}_j$ is needed. A scaling function constructed through orthogonalisation of the Riesz basis given by [65]

$$\phi_{j,n} = \frac{1}{\sqrt{2^j}}\phi\left(\frac{t-n}{2^j}\right) \tag{3.16}$$

provides such a basis. The family $\{\phi_{j,n}\}_{n\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{V}_j$ for any $j \in \mathbb{Z}$. Similarly, the family of wavelet functions $\{\psi_{j,n}\}_{n\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{W}_j$ for any scale $2^j$ and $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ an orthonormal basis of $\mathbf{L^2}(\mathbb{R})$ for all scales. The coefficients of the projection in each of the $\mathbf{V}_j$ and $\mathbf{W}_j$ vector spaces are defined as

$$a_j[n] = \langle f, \phi_{j,n}\rangle \quad \text{and} \quad d_j[n] = \langle f, \psi_{j,n}\rangle \tag{3.17}$$

where $a_j[n]$ are the "approximation" and $d_j[n]$ the "detail" wavelet coefficients at scale $j$. Successive application of 3.15 results in a fast wavelet transform that can successively decompose each approximation into a coarser approximation plus the detail coefficients which are calculated with a cascade of discrete convolutions and subsamplings [67, 68]

$$a_{j+1}[p] = \sum_{n=-\infty}^{+\infty} h[n-2p]a_j[n] = a_j \star \overline{h}[2p] \tag{3.18}$$

$$d_{j+1}[p] = \sum_{n=-\infty}^{+\infty} g[n-2p]a_j[n] = a_j \star \overline{g}[2p] \tag{3.19}$$

where $\overline{x}[n] = x[-n]$.

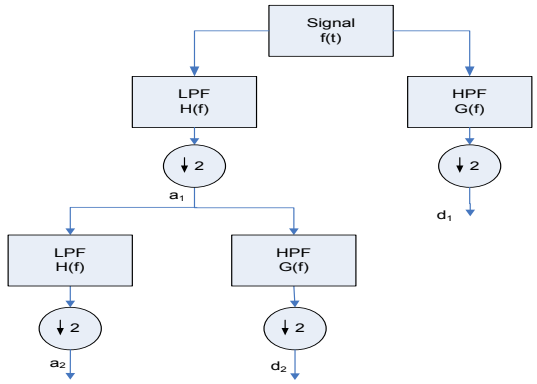
The coefficients $h[n]$ and $g[n]$ are given by the following relations

$$h[n] = \frac{1}{\sqrt{2}}\langle\phi(\frac{t}{2}), \phi(t-n)\rangle \tag{3.20}$$

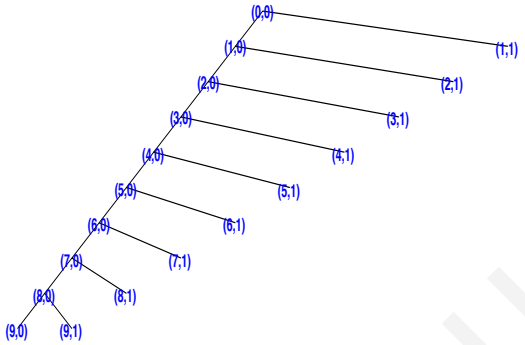
$$g[n] = \frac{1}{\sqrt{2}}\langle\psi(\frac{t}{2}), \phi(t-n)\rangle \tag{3.21}$$

$$g[n] = (-1)^{1-n}h[1-n] \tag{3.22}$$

and can be regarded as the coefficients of discrete conjugate mirror filters [65]. The original signal passes through two filters and two signals emerge, the approximation and detail coefficients corresponding to low-frequency and high-frequency components respectively. An orthogonal representation of $a_L = \langle f, \phi_{L,n}\rangle$ is composed of the detail coefficients of $f$ at scales $2^L < 2^j < 2^J$ plus the remaining approximation coefficients at the largest scale $2^J$. The whole procedure of iteratively filtering the signal from the




(a) The Wavelet Transform implemented as a series of filters




(b) The resulting Wavelet Transform tree

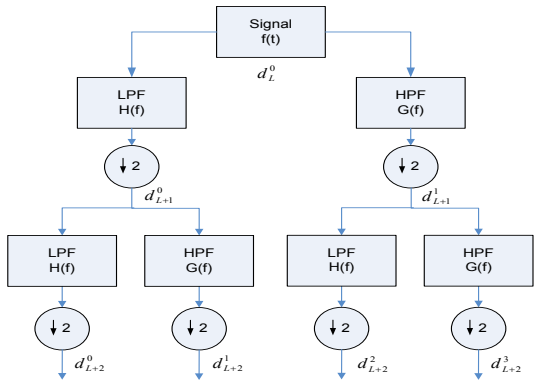
Figure 3.3: The implementation of a fast transform for the WT ideal for computations is shown as a series of filter operations with subsamplings (↓2). The resulting tree from the implementation is also shown which has been expanded up to level 9.

previous level to produce the detail and approximation coefficients at the next level is shown in figure 3.3a. The resulting wavelet tree in figure 3.3b is a projection of a signal in bases that are a priori known and does not depend on the signal's characteristics. Through the use of Wavelet Packets however the projection of a signal is done on bases that are adaptively selected according to an entropy criterion. Therefore Wavelet Packets are better suited than the WT for providing a more compact representation of the signal, a desired property for feature extraction algorithms.

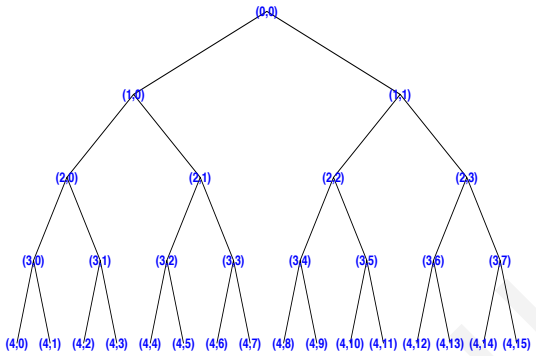
### 3.1.5 Wavelet Packets Transform (WPT)

The WPT, extends the WT, by further projecting the signal in subspace $\mathbf{W}_j$ onto two new subspaces at scale $j+1$, thus deriving new bases [69]. In comparison with the WT where only the "approximation" coefficients $a_j$ are further decomposed at scale $j+1$, now the "detail" coefficients $d_j$ can be further decomposed at scale $j+1$. This decomposition is given by the following projection

$$\mathbf{W}_j^p = \mathbf{W}_{j+1}^{2p} \oplus \mathbf{W}_{j+1}^{2p+1}. \tag{3.23}$$




(a) The Wavelet Transform implemented as a series of filters



(b) The resulting Wavelet Transform tree

Figure 3.4: Implementation of the WPT as a series of filter operations with subsampling. The full resulting tree from the implementation is also shown which has been expanded up to level 5.

where the following two wavelet functions

$$\psi_{j+1}^{2p} = \sum_{n=-\infty}^{+\infty} h[n]\psi_j^p(t - 2^j n) \tag{3.24}$$

$$\psi_{j+1}^{2p+1} = \sum_{n=-\infty}^{+\infty} g[n]\psi_j^p(t - 2^j n) \tag{3.25}$$

constitute orthogonal bases for the two subspaces. Similarly to the wavelet transform $h[n]$ and $g[n]$ are given by

$$h[n] = \langle \psi_{j+1}^{2p}(u), \psi_j^p(u - 2^j n) \rangle \tag{3.26}$$

$$g[n] = \langle \psi_{j+1}^{2p+1}(u), \psi_j^p(u - 2^j n) \rangle \tag{3.27}$$

where the relation of 3.22 still holds and they can be viewed as the coefficients of discrete filters recursively calculated by subsampling and convolving the discrete filters coefficients $h[n]$ and $g[n]$ with the coefficients at a lower scale

$$d_{j+1}^{2p}[k] = d_j^p \star \overline{h}[2k] \quad \text{and} \quad d_{j+1}^{2p+1}[k] = d_j^p \star \overline{g}[2k] \tag{3.28}$$

as shown in figure 3.4a. In figure 3.4b the full WPT tree is shown where it can be seen how the wavelet coefficients at node $(j, p)$, $d_j^p[n]$, representing the projection in subspace $\mathbf{W_j^p}$ are projected into their two children giving wavelet coefficients $d_{j+1}^{2p}[n]$ and $d_{j+1}^{2p+1}[n]$.

This full tree decomposition provides a larger choice of bases that the signal can be expanded into and can provide an adaptive decomposition for a particular signal. Therefore a measure is needed as a criterion for the selection of the bases to be included in the decomposition. The requirement is that this cost function is concave and that it is minimised by the selected basis. Several measures exist in the literature the most commonly used being Shannon's entropy of the energy of the WP coefficients [65] defined as

$$C(f, B) = -\sum_{m=1}^{N} \frac{|\langle f, g_m \rangle|^2}{\|f\|^2} \ln \left( \frac{|\langle f, g_m \rangle|^2}{\|f\|^2} \right) \tag{3.29}$$

where $g_m$ is the dictionary of wavelet bases. For a length $N$ signal, the dictionaries of these wavelet packets include more than $2^{\frac{N}{2}}$ bases and finding the minimum by brute force comparison of the cost of all Wavelet Packets is computationally prohibitive. Fortunately the fast dynamic algorithm of Coifman and Wickerhauser [70] finds the best basis with $O(N \log_2 N)$ complexity, by taking advantage of the tree structure of these dictionaries and the additivity of the cost function. In figure 3.5 the best basis tree is shown for an

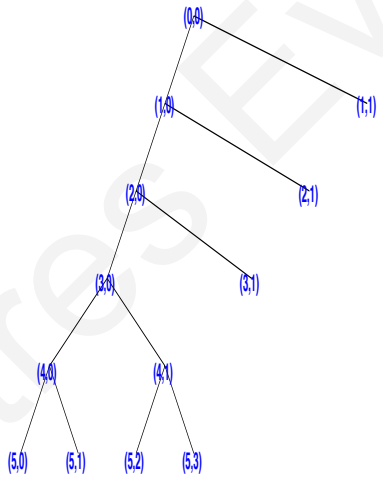
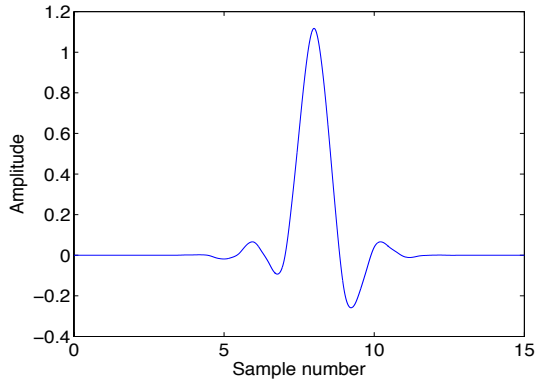


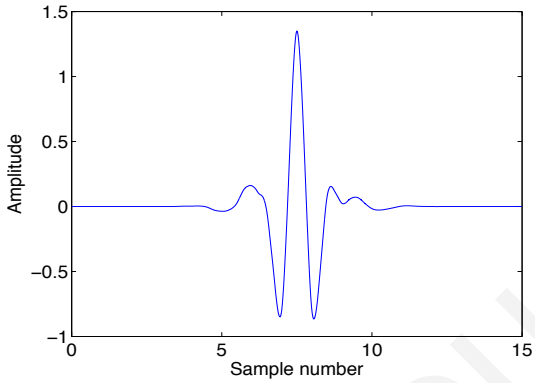

Figure 3.5: Best Basis tree representation for an experimental corona single pulse at an applied voltage of 6 kV using Shannon's entropy as a cost function.

experimental corona at an applied voltage of 6 kV computed by employing the fast basis search algorithm and Shannon's entropy cost function.

The cost function is not the only parameter that has to be chosen a priori by the user but the number of vanishing moments and the family of the wavelet used can have an impact on the decomposition, as explained next.




(a)




(b)

Figure 3.6: The Symmlets 8 (a) scaling and (b) wavelet functions chosen as the functions to be used in the wavelet analysis of PDs.

### 3.1.6 WPT parameters selection

Wavelets are defined by the number of vanishing moments and the family they belong to, which have to be selected according to the application at hand. A wavelet having n vanishing moments, defined as

$$\int_{t=-\infty}^{\infty} t^k \psi(t) dt = 0 \quad \text{for} \quad 0 \leq k < n \tag{3.30}$$

suppresses polynomials of degree n-1. Although the support size of a function and the number of vanishing moments are a priori independent, due to the constraints imposed on orthogonal wavelets if $\psi$ has $p$ vanishing moments then its support is at least $2p-1$ [65]. This imposes a trade-off between the number of vanishing moments and the support. The number of vanishing moments relates to the frequency behaviour of the wavelet filter, which in turn can be linked to the steepness of the filter. It has been shown [71] that provided the order of the filter is above a certain value (where this value depends on the wavelet packet) and a certain resolution level then the wavelet coefficients within a particular node decorrelate. In effect choosing higher wavelet moments can have a "whitening" effect on the coefficients. The two most widely used family of wavelets in PD are the Daubechies and the Symmlets wavelets. Daubechies are very asymmetric because they are constructed by selecting the minimum-phase response while the Symmlets provide a linear phase response. The choice of the Symmlets wavelet with 8 vanishing moments (shown in figure 3.6) is found to be a good compromise between a high enough number

of moments and an acceptable level of support. This choice is based on reported work in the literature [72] as well as work by the author in denoising simulated PD signals [62, 73] where the performance of the wavelet family and order was evaluated based on the correlation coefficient, the SNR and the Mean Squared Error (MSE) between the original and the denoised signal. Unless otherwise explicitly stated throughout this work the Symmlets 8 wavelet is used and the wavelet packet tree obtained by minimisation of Shannon's entropy function. Also the effect of the levels of decomposition on the denoising performance was examined in [74] where the inclusion of 9 levels produced better results.

In this section the WPT is selected to adaptively decompose the PD signal and is applied on experimental PD data resulting in a best basis tree, using the parameters above. Subsequently specific nodes of this tree that are believed to best represent the various PD sources are selected, using a procedure that is explained in section 3.4.1. Although the actual wavelet coefficients of the selected nodes can be used as features [39], this creates a feature vector of high dimensionality which is undesirable due to its higher computational and storage burden. Instead the application of statistical analysis tools to describe the probability density function of the respective wavelet coefficients, is examined. Application of Higher Order Statistics as statistical descriptors is one possible solution, and the theoretical ground for their use is set in the next section.

## 3.2 Higher Order Statistics (HOS)

As stated in sections 3.1.1 and 3.1.2 statistical analysis methods were utilised in order to capture the statistical nature of PD signals and their amplitude variations. This statistical nature stems from the fact that the initiation and development of a PD depends on a number of physical mechanisms (for example the presence of a starting electron in order for a PD to take place is governed by a probability law). The probability density function (pdf) of PD amplitudes was used for example in [42, 75] where a Weibull Distribution was fitted to the pulse amplitudes which in turn could be described by the scale and shape parameters of the distribution. Such parametric approaches require the a priori selection of an underlying distribution and the calculation of its parameters from the measurements through an appropriate estimator. As long as this assumption holds true the parameters can accurately represent the process but if this is not the case then the performance will deteriorate. The use of HOS approximates a distribution in a least

squares sense without making an assumption about its shape, through the calculation of the statistical cumulants.

The basic principles for the description of a random variable revolve around the interplay between four statistical descriptors: the probability density (pdf), the characteristic, the moment and the cumulant generating functions. If two distributions have a certain number of cumulants in common they will bear some resemblance to each other and a mathematical justification of this can be found in [76, page 111]. The cumulants, have properties that are more useful from a theoretical point of view. The cumulant of two statistically independent random variable equals the sum of the cumulants of the individual random variables, whereas this is not true for higher order moments. This allows us to treat cumulants as operators simplifying matters [77]. Furthermore from the additivity property third and higher-order cumulants are blind to Gaussian noise. All of the above properties make the cumulants the chosen set of descriptors used in this work and are thus described in the next section.

In the univariate case $X$ is a real random variable with probability density $f(x)$ and the characteristic function $\Phi(\omega)$ given by

$$\Phi(\omega) = \int_{-\infty}^{+\infty} e^{j\omega x} f(x) dx \tag{3.31}$$

### 3.2.1 Use of cumulants to describe the pdf

The cumulants, $\kappa_1$, $\kappa_2$, ..., $\kappa_r$ are defined as the derivatives of the second characteristic function $\Psi(\omega)$ defined as

$$\Psi(\omega) = \ln(\Phi(\omega)). \tag{3.32}$$

where $\Phi(\omega)$ is the characteristic function given by

$$\Phi(\omega) = \int_{-\infty}^{+\infty} e^{j\omega x} f(x) dx. \tag{3.33}$$

The cumulant function, $\kappa_r$ is the coefficient of $\omega^r/r!$ in $\ln \Phi(\omega)$, if an expression in power series exists in the same way that $\mu'_r$ is the coefficient of $\omega^r/r!$ in $\Phi(\omega)$. Relations between cumulants and central moments exist and are given by

$$\begin{aligned} \kappa_2 &= \mu_2 \\ \kappa_3 &= \mu_3 \\ \kappa_4 &= \mu_4 - 3\mu_2^2. \end{aligned} \tag{3.34}$$

Cumulants except the first are invariant under change of origin, a property that is shared by central moments ($\mu_r$) as well.

The notion of cumulants as a set of descriptive constants can be extended to the multivariate case where a collection of random variables, $\mathbf{X} = [X_1\ X_2\ .....\ X_n]^T$ is defined by the joint characteristic function of the multivariate distribution $F(x_1,\ x_2,\ \cdots\ , x_n)$ given by

$$\Phi(\omega_1, \omega_2, ..... , \omega_n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{j\omega_1 x_1 + j\omega_1 x_1 + \cdots + j\omega_n x_n}. \tag{3.35}$$

The multivariate equivalent of the cumulant expansion as a power series is given by

$$\begin{aligned}\exp\left(\sum_{r_1=0}^{\infty}\sum_{r_2=0}^{\infty}\cdots\sum_{r_n=0}^{\infty}\frac{\kappa_{r_1 r_2\cdots r_n}\omega_1^{r_1}\omega_2^{r_2}\cdots\omega_n^{r_n}}{r_1!r_2!\cdots r_n!}\right)\\ = \sum_{r_1=0}^{\infty}\sum_{r_2=0}^{\infty}\cdots\sum_{r_n=0}^{\infty}\frac{\mu'_{r_1 r_2\cdots r_n}\omega_1^{r_1}\omega_2^{r_2}\cdots\omega_n^{r_n}}{r_1!r_2!\cdots r_n!}\end{aligned} \tag{3.36}$$

where $k_{00}$ is defined as equal to 0. The calculation of product cumulants, $\kappa_{r_1 r_2\cdots r_n}$ and higher order terms when n is larger than 2 in 3.36 becomes a tedious task and the series expansion can be approximated by partial terms such as for example $r_i = 1\ \forall i$.

Having reviewed how the cumulants can be a set of descriptive constants, the normalised cumulants (normalised to the energy of the signal, $\kappa_2$) are chosen, as they are invariant to changes of scale and translations, to approximate the pdf of the wavelet coefficients at selected nodes of the WPT tree. These cumulants have to be estimated from the available samples $d_j^p[k]$ and the following equations [76] are used in their calculation:

$$\eta = \mu'_1 = \frac{1}{N_j}\sum_{k=1}^{N_j}(d_j^p[k]) \tag{3.37}$$

$$\sigma = \sqrt{\kappa_2} = \left[\frac{1}{N_j - 1}\sum_{k=1}^{N_j}(d_j^p[k] - \eta)^2\right]^{\frac{1}{2}} \tag{3.38}$$

$$\gamma_1 = \frac{\kappa_3}{\kappa_2^{3/2}} = \frac{\sqrt{N_j(N_j-1)}}{N_j - 2}\frac{1}{\sigma^3}\sum_{k=1}^{N_j}(d_j^p[k] - \eta)^3 \tag{3.39}$$

$$\begin{aligned}\gamma_2 = \frac{\kappa_4}{\kappa_2^2} = \frac{N_j(N_j+1)}{(N_j-1)(N_j-2)(N_j-3)}\frac{1}{\sigma^4}\\ \sum_{k=1}^{N_j}(d_j^p[k] - \eta)^4 - 3\frac{(N_j-1)^2}{(N_j-2)(N_j-3)}.\end{aligned} \tag{3.40}$$

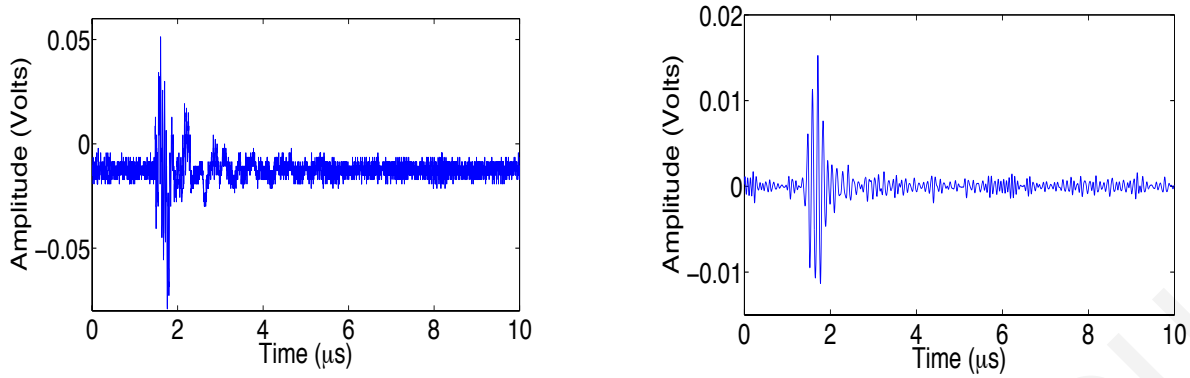
In all the above equations $N_j$ is the number of wavelet packet coefficients at scale $j$, which halves each time the scale increases by one. They are used in the calculation of four statistical descriptors for each wavelet packet tree node included in the representation.

Now that the theoretical background behind the use of HOS as statistical descriptors has been layed, two feature extractors based on a combination of the tools analysed in the previous two sections are proposed. The first proposed feature is based on time-resolved measurements while the second one on phase-resolved data.

## 3.3 Time-resolved PD proposed feature

The applicability of WPT and HOS in getting compact, descriptive representations of PD signals of different sources was explained theoretically in the previous two sections. Here a proposed feature vector based on time measurements and the two data analysis tools is proposed. The measurements shown in table 2.3 are used in the analysis and are processed using the WPT to reduce the size of the representation and recover the useful information, using an adaptive time-frequency decomposition. Previous research [73, 78] showed the advantages of wavelet decomposition of PD signals in both the time and frequency domain. Through the use of the Wavelet Packet Transform (WPT) some of the signal characteristics can be enhanced and represented in a more compact form resulting in a feature vector of lower dimensions. For example figure 3.7 shows the original signal of an internal discharge at an applied voltage of 26 kV (3.7a) against the reconstructed signal from wavelet packet coefficients at node (5,0) (3.7b) which demonstrates the fact that each node highlights different time frequency characteristics that are not immediately obvious in the original signal.

Specific nodes of the WPT have been reported to provide good results in identifying PDs, for example in [39] the wavelet coefficients of node at level 3 position 1 ($d_3^1[k]$) expanded using a Symmlet 8 were scaled and used as a feature vector in a SVM classification machine. After processing, the data length was reduced to approximately 1/8 of its original size [39] which supports the ability of the transform to compact useful information in a signal. In this research a combination of different nodes of the wavelet packet tree decomposition are used and their selection procedure is outlined in the next section.




(a) Original

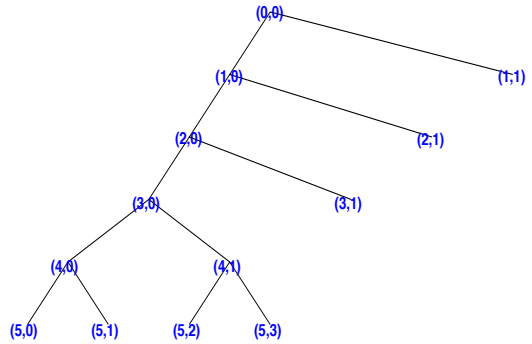
(b) Reconstructed

Figure 3.7: Time plot of experimentally acquired internal discharge at an applied voltage of 26 kV against its reconstruction using the wavelet coefficients $d_1^5[k]$ which shows that coefficients in specific nodes can be used to delineate different time-frequency characteristics of a signal.

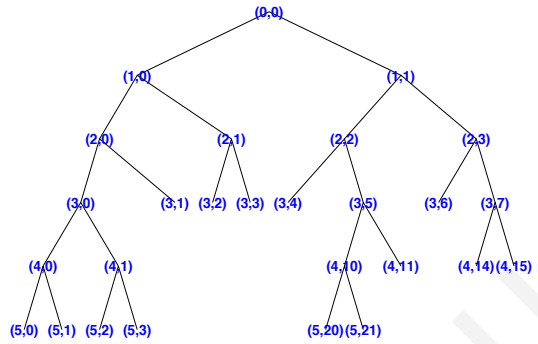
### 3.3.1 Selection of the Wavelet Packet Transform Nodes

All the single pulse experimentally acquired PD signals of table 2.3 are expanded into the best basis tree that minimises Shannon's entropy using a Symmlet wavelet with 8 vanishing moments. Typical wavelet packet trees for the different discharge types can be seen in figure 3.8 where the signals are expanded up to level 5. The number of wavelet coefficients at a particular scale ($N_j$) are used in the formula of the calculation of the cumulants (3.37-3.40). Since a low value can result in unreliable estimates (high bias of the estimator) and the number of coefficients halves with an increase of scale, level 5 is considered a good compromise ($N_5 = 170$). The coefficients are not exactly half each other because 5000 is not an integer of a power of 2 and the algorithm takes this into account. The selection procedure followed is to extract a common best basis tree whose coefficients adequately describe and differentiate the different PD sources.

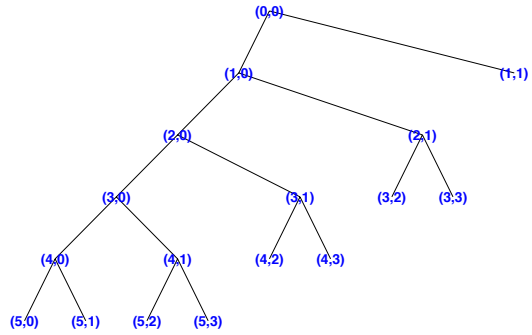
For each discharge type the terminal nodes in the best basis tree expansion for each signal are recorded. The ten terminal nodes that are most frequently included in the best basis tree are selected for each discharge type. Terminal nodes that appear on all four "common" trees are selected to represent the "common" wavelet packet tree for all discharges, shown in figure 3.9. From this common tree nodes (3,1), (4,1), (5,0) and (5,1) are selected as the nodes whose coefficients pdf will be approximated by HOS. The frequency content of these nodes is shown in table 3.1 and is seen that they capture different frequency content of the PD signals. Since the field data is expected to be



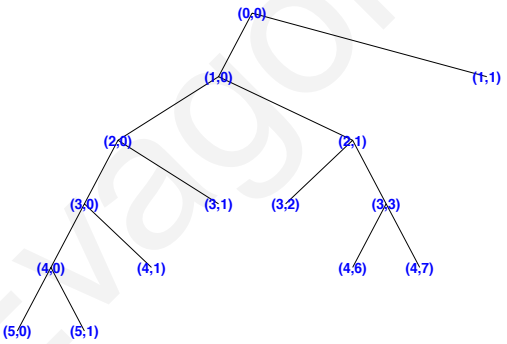

(a) Corona discharge in air at an applied voltage of 6 kV.




(b) Floating discharge in oil at an applied voltage of 24 kV




(c) Internal discharge in oil at an applied voltage of 26 kV




(d) Surface discharge in air at an applied voltage of 6 kV

Figure 3.8: Best Basis tree expansion for single pulse experimental PD data of different sources (sampled at 500 MS/s) calculated through mimimisation of Shannon's entropy cost function and using a Symmlet wavelet with 8 vanishing moments.

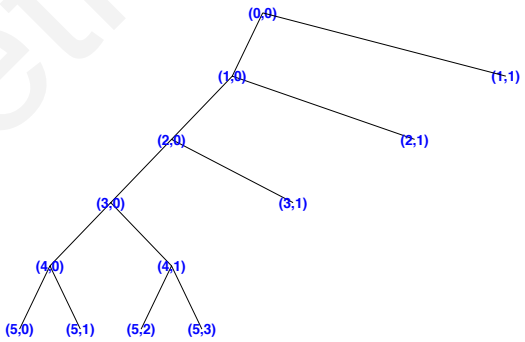



Figure 3.9: The "common" best basis tree selected using the proposed procedure for node selection.

Table 3.1: Frequency content of specific nodes of the WPT.

| **Wavelet Node** | **Relative Frequencies** | **Absolute Frequencies (MHz)** |
|---|---|---|
| 3,1 | $f_s/16$-$2f_s/16$ | 31.25-62.50 |
| 4,1 | $f_s/32$-$2f_s/32$ | 15.625-31.25 |
| 5,0 | 0-$f_s/64$ | 0-7.8125 |
| 5,1 | $f_s/64$-$2f_s/64$ | 7.8125-15.625 |
| 5,2 | $3f_s/64$-$4f_s/64$ | 23.4375-31.25 |
| 5,3 | $2f_s/64$-$3f_s/64$ | 15.625-23.4375 |

affected by the propagation characteristics of the signal which attenuates high frequency content (the frequency content of a PD in a cable is expected to be around 10 MHz) and the same sensor is used it can be expected that field data can be adequately described by the selected nodes, which will be verified in chapter 5 when field data are examined. To verify that this selection of nodes will adequately represent the signal and how many of the nodes have to be selected an error analysis is performed in the following section.

#### 3.3.1.1 Error analysis on the selected nodes

An error analysis is performed on the wavelet coefficients of the selected nodes to verify that they can represent the PD signal satisfying a performance criterion and that selection of more nodes does not significantly improve the performance. Several criteria have been used in this thesis which are the entropy of the energy of the wavelet coefficients at the selected nodes plus the mean square error (mse) and the relative energy (to the signal's energy) of the reconstructed signal using only the wavelet coefficients at each node.
Results of this error analysis are shown for each PD type in tables 3.2, 3.3, 3.4, and 3.2 for the corona, floating, internal and surface discharges respectively. From the analysis it seems to exist a dominant node in the examined nodes, where the coefficients of node (5,0) ($d_5^0[k]$) significantly improve all three error criteria for all PD types. For example for the corona discharge the reconstruction of the signal using coefficients $d_5^0[k]$ retains around 86% of the energy of the signal. The rest of the nodes have a much lower contribution and they can be ordered with a decreasing contribution as (4,1), (5,2), (5,3) and (3,1). A compromise between the number of nodes and the quality of the representation on one side with the number of dimensions of the representation has to be reached. As the nodes (5,2) and (5,3) are children nodes of (4,1) which means that it contains all the information of its children nodes, the inclusion of the four selected wavelet nodes (3,1),

Table 3.2: Error analysis for various nodes of the Wavelet Packet Transform tree for the corona discharge. The values are given as the mean and the standard deviation of the respective error measure, with 10000 pulses being analysed.

| Variable | Property | WPT Node index | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | [0,0] | [3,1] | [3,2] | [3,3] | [4,1] | [5,0] | [5,1] | [5,2] | [5,3] |
| entropy | mean | 7.59 | 0.15 | 0.11 | 0.12 | 0.40 | 3.87 | 0.04 | 0.24 | 0.17 |
| | std | 0.39 | 0.06 | 0.04 | 0.04 | 0.17 | 0.73 | 0.02 | 0.11 | 0.07 |
| energy (%) | mean | 100.00 | 1.48 | 1.08 | 1.13 | 5.43 | 86.23 | 0.45 | 3.09 | 2.33 |
| | std | 0.00 | 0.62 | 0.44 | 0.47 | 2.67 | 5.76 | 0.21 | 1.73 | 1.20 |
| mse | mean | 0.00 | 0.99 | 0.99 | 0.99 | 0.95 | 0.14 | 1.00 | 0.97 | 0.98 |
| | std | 0.00 | 0.01 | 0.00 | 0.00 | 0.03 | 0.06 | 0.00 | 0.02 | 0.01 |

Table 3.3: Error analysis for various nodes of the Wavelet Packet Transform tree for the floating discharge. The values are given as the mean and the standard deviation of the respective error measure, with 10000 pulses being analysed.

| Variable | Property | WPT Node index | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | [0,0] | [3,1] | [3,2] | [3,3] | [4,1] | [5,0] | [5,1] | [5,2] | [5,3] |
| entropy | mean | 8.10 | 0.13 | 0.06 | 0.12 | 0.24 | 4.74 | 0.03 | 0.14 | 0.11 |
| | std | 0.60 | 0.13 | 0.07 | 0.12 | 0.25 | 0.95 | 0.03 | 0.15 | 0.13 |
| energy (%) | mean | 100.00 | 1.67 | 0.63 | 1.65 | 4.72 | 87.32 | 0.34 | 2.58 | 2.14 |
| | std | 0.00 | 2.08 | 0.83 | 2.04 | 6.06 | 15.15 | 0.41 | 3.52 | 3.16 |
| mse | mean | 0.00 | 0.98 | 0.99 | 0.98 | 0.95 | 0.13 | 1.00 | 0.97 | 0.98 |
| | std | 0.00 | 0.02 | 0.01 | 0.02 | 0.06 | 0.15 | 0.00 | 0.04 | 0.03 |

(4,1), (5,0) and (5,1) seems a good compromise, retaining more than 90% of the energy while the addition of more nodes does not significantly improve this.

The pdf of the wavelet coefficients at each selected node is approximated by their mean $\eta$, standard deviation $\sigma$, normalised skewness $\gamma_1$ and normalised kurtosis $\gamma_2$, and will be used as a fingerprint for each type of PD and as an input to the classifier. These quantities are estimated from the available samples using equations 3.37-3.40. To investigate the ability of this feature extraction method to produce significantly different feature values for different sources of PD a confidence interval plot for the estimated HOS is plotted

Table 3.4: Error analysis for various nodes of the Wavelet Packet Transform tree for the internal discharge. The values are given as the mean and the standard deviation of the respective error measure, with 10000 pulses being analysed.

| Variable | Property | WPT Node index | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | [0,0] | [3,1] | [3,2] | [3,3] | [4,1] | [5,0] | [5,1] | [5,2] | [5,3] |
| entropy | mean | 7.92 | 0.19 | 0.07 | 0.09 | 0.63 | 4.29 | 0.07 | 0.25 | 0.37 |
| | std | 0.46 | 0.12 | 0.05 | 0.06 | 0.42 | 0.92 | 0.05 | 0.22 | 0.26 |
| energy (%) | mean | 100.00 | 2.47 | 0.67 | 0.92 | 11.38 | 81.01 | 1.02 | 4.75 | 6.62 |
| | std | 0.00 | 1.86 | 0.58 | 0.68 | 9.10 | 13.67 | 0.93 | 5.42 | 5.33 |
| mse | mean | 0.00 | 0.98 | 0.99 | 0.99 | 0.89 | 0.19 | 0.99 | 0.95 | 0.93 |
| | std | 0.00 | 0.02 | 0.01 | 0.01 | 0.09 | 0.14 | 0.01 | 0.05 | 0.05 |

Table 3.5: Error analysis for various nodes of the Wavelet Packet Transform tree for the surface discharge. The values are given as the mean and the standard deviation of the respective error measure, with 10000 pulses being analysed.

| Variable | Property | WPT Node index | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | [0,0] | [3,1] | [3,2] | [3,3] | [4,1] | [5,0] | [5,1] | [5,2] | [5,3] |
| entropy | mean | 7.61 | 0.09 | 0.05 | 0.07 | 0.12 | 4.22 | 0.03 | 0.06 | 0.07 |
| | std | 0.50 | 0.06 | 0.04 | 0.05 | 0.07 | 0.67 | 0.02 | 0.04 | 0.04 |
| energy (%) | mean | 100.00 | 0.97 | 0.43 | 0.74 | 1.81 | 94.08 | 0.30 | 0.75 | 1.06 |
| | std | 0.00 | 0.73 | 0.38 | 0.66 | 1.18 | 4.41 | 0.22 | 0.55 | 0.64 |
| mse | mean | 0.00 | 0.99 | 1.00 | 0.99 | 0.98 | 0.06 | 1.00 | 0.99 | 0.99 |
| | std | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 | 0.04 | 0.00 | 0.01 | 0.01 |

in figure 3.10. In the calculation of the confidence intervals for the estimation of each parameter in the feature vector, 500 pulses per PD source are used. From the plot it can be seen that the normalised kurtosis at the selected wavelet nodes produces distinct values for different PD sources, at the 0.95 confidence interval. Therefore this feature seems to have the potential to be used as a fingerprint in PD source identification.

If the procedure of selecting the WPT coefficients of these nodes is followed, a feature vector of 1304 dimensions will be created. Through the use of HOS on each of these four nodes, a feature vector of 16 dimensions (4 dimensions per node) is constructed. This

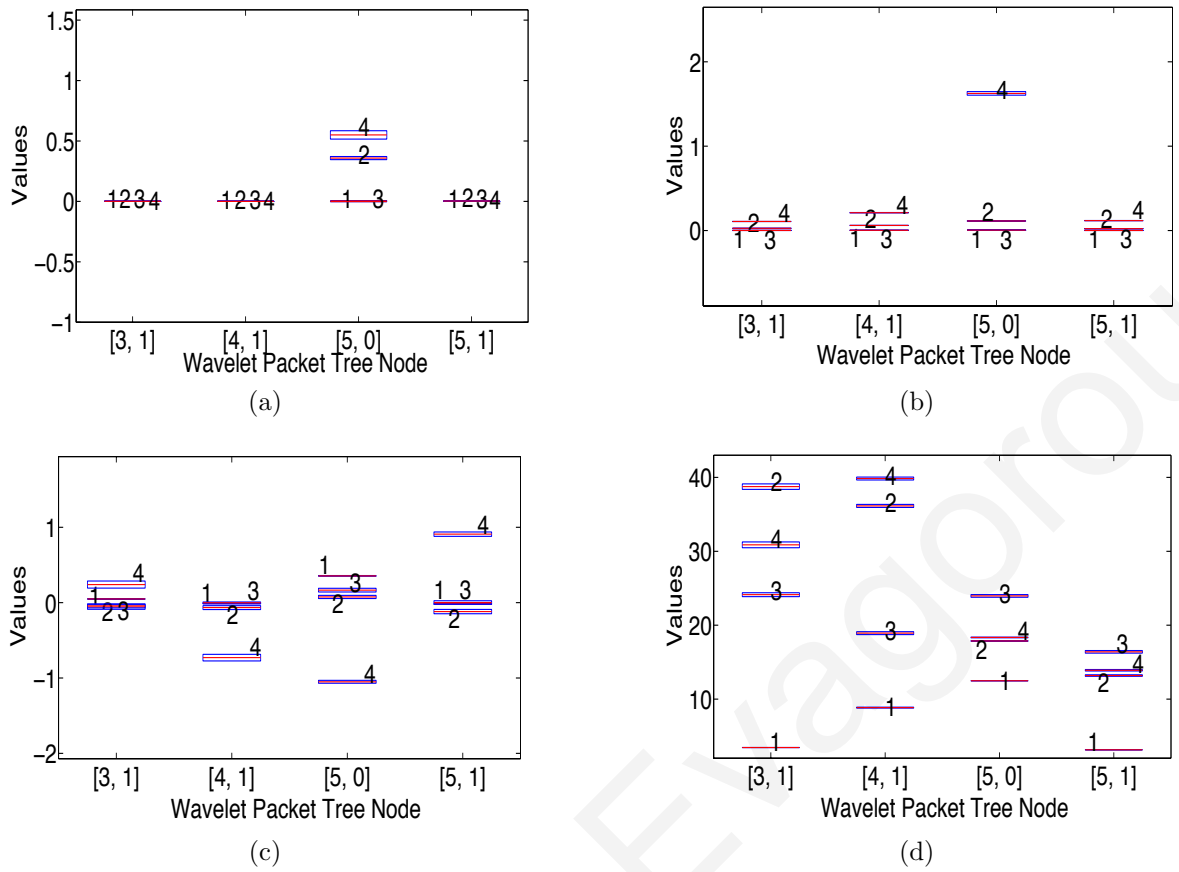



Figure 3.10: Confidence interval for the feature vector. The feature vector has been broken down in order to make more visible the differences between the different statistical descriptors used. In 3.10a $\eta$, estimated from the wavelet coefficients of the selected nodes is plotted. In 3.10b $\sigma$ and in 3.10c $\gamma_1$ are shown. Finally 3.10d plots $\gamma_2$. The middle line is the value of the mean of the estimated quantity while the box enclosing it is the 0.95 confidence interval, with 500 samples used. The plot resulted in separate values for the majority of the dimensions, especially when $\gamma_2$ values are considered, showing the potential of this feature to be used as a fingerprint in PD source identification. The numbering in the figure stands for 1: Corona, 2: Floating, 3: Internal, 4: Surface.

reduction in the number of dimensions translates into reduced computational effort both in training and classification times, as well as lower storage requirements. The proposed feature vector for single pulse data will be considered in subsequent chapters and its use as a pre-processing step to create homogeneous phase resolved data patterns will be assessed. In the next section we look at the proposed phase resolved feature extractor that is also based on the WPT and HOS.

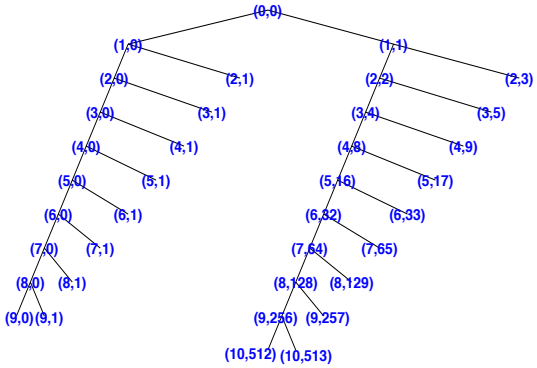
## 3.4 Phase Resolved PD Proposed Feature

Traditional phase resolved representations have been used in PD identification, in the early days of PD, mainly in laboratory measurements. However with the advent of computational power and the use of non-conventional sensors they can be used as a complementary PD identification tool. Since they rely on the assumption of a single PD source concurrently active and no interference present, pre-processing using time-resolved features and clustering techniques is required.

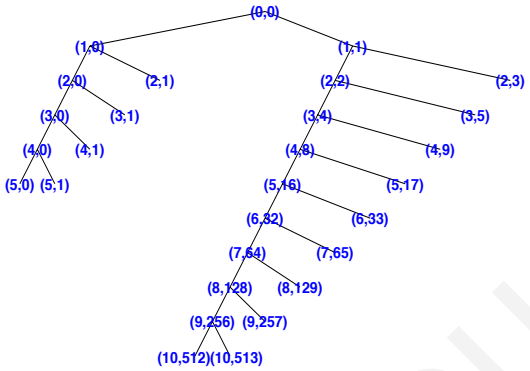
Statistical analysis and signal processing tools were utilised in an attempt to describe the characteristics of PRPD with a set of descriptive parameters, as reviewed in section 3.1.1. In this section the experimental PD signals collected under controlled conditions are processed using the WPT to obtain a feature vector based on HOS. The approach of using the WPT to extract a feature vector, utilises the properties of the transform itself in representing the signal, combined with the use of HOS as statistical descriptors.

The phase resolved data acquired using the peak detection function of the oscilloscope (table 2.2 and figure 2.14) is expanded into the WPT using a Symmlet 8 wavelet. The best basis trees obtained by the minimisation of Shannon's entropy function for different PD sources are shown in figure 3.11 with 10 levels of decomposition used. The wavelet packet coefficients of each terminal node are approximated by the mean $\eta$, standard deviation $\sigma$, normalised skewness $\gamma_1$ and normalised kurtosis $\gamma_2$. The statistical operators for terminal nodes at the same decomposition level are added together and thus only four statistical descriptors are used at each decomposition level. As 10 levels are used in the decomposition and level 1 does not include any terminal nodes, this procedure results in 9 levels to be represented by four statistical descriptors. To estimate the statistical descriptors the available samples from table 2.2 are used in equations 3.37-3.40.

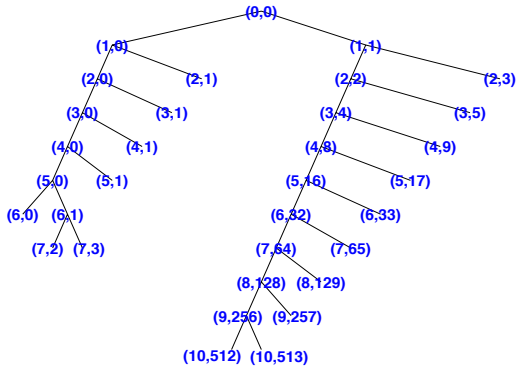
The proposed feature uses four descriptors for each scale, therefore a feature vector of 36 dimensions is needed to represent the data. If the histogram of the wavelet coefficients at each level is used as a feature vector, then for 100 bins used in the calculation of the histogram, a feature vector of dimension 900 will emerge for each PD signal. Use of the proposed method reduces the dimensions of the feature vector from 900 to 36, a reduction by a factor of 25. This makes the input vector more tractable and the classification less resource demanding. The use of HOS in dimensionality reduction is demonstrated [45] when compared to a similar feature extraction method that utilises specific nodes of the WPT tree decomposition (the $d_3^1$, $d_6^1$ and $d_9^1$ coefficients) [79]. As this has a dimension of 1468 then the proposed feature with only 36 dimensions offers a reduction of the order



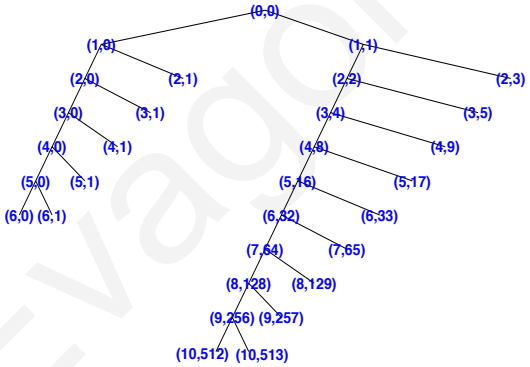

(a) Corona discharge in air at an applied voltage of 6 kV.




(b) Floating discharge in oil at an applied voltage of 24 kV




(c) Internal discharge in oil at an applied voltage of 26 kV




(d) Surface discharge in air at an applied voltage of 6 kV

Figure 3.11: Best Basis tree expansion for phase resolved experimental PD data of different sources (sampled at 500 kS/s through the peak detection function of the oscilloscope) calculated through mimimisation of Shannon's entropy cost function and using a Symmlet wavelet with 8 vanishing moments.

of 40. In the next chapter the classification rate of this lower dimensional feature will be evaluated and its suitability in identifying different PD sources assessed.

## 3.5 Conclusions

In this chapter two feature vectors, one for time and another for phase-resolved data, of lower dimensionality were proposed. Due to this property they have the potential to be applied on data collected from continuous on-line systems where the storage and computation complexity need to be minimised. In both methods the WPT was firstly applied on the data resulting in an adaptive time-frequency decomposition of the PD signal. This

exploited the different frequency content between different PD sources resulting in different coefficients at specific nodes of the Wavelet Packet tree. A procedure was followed to select particular nodes in the tree that contained discriminatory characteristics. Following the application of statistical analysis tools, HOS were employed to get a set of statistical descriptors of the wavelet coefficients at each of the selected nodes. From the analysis in this section, it seems that the proposed feature vectors have the potential to be used as fingerprints in PD characterisation. In order to verify this, the performance of the proposed features in classifying PDs of different sources using supervised classification algorithms is considered in the following chapter.

# Chapter 4

# PD Classification

In this chapter two supervised classification algorithms are used to assess the classification performance of the proposed feature vectors in chapter 3 using the experimental database of different PD sources. Although the two algorithms are based on a training set to provide a category label in order to estimate the weights of the discriminant function, they rely on inherently different principles in order to calculate these weights. The Probabilistic Neural Network (PNN) is based on the Empirical Risk Minimisation (ERM) and the Support Vector Machines (SVM) on the Structural Risk Minimisation (SRM) principle, each minimising a different term of the risk functional generalisation learning formula. The reproducibility of their performance on both single pulse and phase resolved data is statistically evaluated using the concept of tolerance intervals. Both the single pulse and phase resolved data proposed features show marginally inferior classification rates to similar works [39, 79] but with a significant reduction in the number of dimensions, which can be exploited in scenarios where the computational complexity and the storage requirements need to be minimised.

## 4.1 Overview

The various PD classification algorithms in the literature are based on statistical analysis, signal processing analysis, fuzzy logic, time-series analysis, artificial neural networks (ANN), and the more recent Support Vector Machine (SVM) [60, 80]. Although these algorithms make use of different fundamental concepts in achieving the classification of PD data, they all rely on a set of labelled data to estimate a set of weights in the algorithm. Such supervised methods usually achieve this through the minimisation of a

the error between the estimated and the provided labels. In the following section a brief review of the different methods employed in PD identification is provided.

Artificial neural networks (ANN) have been successfully applied in various pattern classification problems including PD identification [33, 81–83]. The basic advantage of ANN over other distance classifiers is its ability to learn from examples, where knowledge in the training set is extracted and stored in the connection weights and neuron biases during the learning phase. The Multilayer Perceptron (MLP) [84, 85] and Radial Basis Function (RBF) neural networks are known for their ability to universally approximate any non-linear function, through the presence of a hidden layer with a differentiable non-linear activation function. Such networks usually follow a winner takes it all strategy where different PD sources are discriminated into a single class and the information provided is only that class id. However there are cases when two classes are marginally close that more information, such as for example the probability of belonging to a specific class, will be useful if conveyed to the user. PNNs constitute a flavour of ANN that have a meaningful interpretation of their outputs in terms of probabilities.

A much more recent topic used in the identification of PD patterns is the so called machine learning methods and in particular the Support Vector Machine (SVM). It was firstly described by V.N. Vapnik in 1995 [86] as an application of statistical learning theory. SVM uses the concept of Kernels for a number of learning tasks and has shown better performance than neural networks in a variety of fields [78, 85, 87]. The SVM has been used in a number of PD classification tasks including PDs where results of using the SVM with several feature extractors have been published in [78]. Its advantages include the fact that the network topology need not be selected a priori but is done during the training phase, and that optimisation is performed on a quadratic convex function that has a single minimum alleviating issues with local minima.

The algorithms chosen to be used in this research, namely the Probabilistic Neural Network (PNN) and the Support Vector Machine (SVM) are based on the theory of statistical learning but use different underlying principles to produce discriminant functions. This theory is briefly outlined in the following section in order to give a background insight into the differences and similarities between the two algorithms that will facilitate the comparison of the results.

## 4.2 Statistical Learning Theory

The notions of Empirical Risk Minimisation (ERM) and Structural Risk Minimisation (SRM) lie at the heart of the implementation of the PNN and the SVM respectively. The theory of statistical learning theory developed by Vapnik [86] formally defines these two principles.

The problem of learning is that of choosing from a set of functions $f(x, \beta)$, $\beta \in \Lambda$ the one that minimises the following expected risk:

$$R(\beta) = \int L(y, f(x, \beta))dF(x, y) \tag{4.1}$$

where $L(y, f(x, \beta))$ is a loss function between the actual response of the system $y$ and the expected value of the response $f(x, \beta)$. Usually $F(x, y)$ is unknown and the only available information is a training set of $l$ observations drawn from $F(x, y) = F(x)F(y|x)$ which are assumed to be independent and identically distributed (i.i.d.), and given below

$$(x_1, y_1), \ldots, (x_l, y_l). \tag{4.2}$$

In order to minimise the risk functional based on the empirical data of (4.2) the expected risk is replaced by the empirical risk

$$R_{\text{emp}}(\beta) = \frac{1}{l}\sum_{i=1}^{l} L(y, f(x_i, \beta))dF(x, y) \tag{4.3}$$

and the function $L(y, f(x, \beta_0))$, that minimises the expected risk, by $L(y, f(x, \beta_l))$. This principle is called the Empirical Risk Minimisation (ERM) inductive principle and plays a crucial role in learning theory.

In spite of the consistency of the ERM (tends to the expected risk as $l$ tends to infinity), it is intended to deal with large samples, where the empirical risk asymptotically attains the actual risk. To avoid the risk of overfitting when sets of smaller sizes are considered the complexity of the learning machine must also be controlled, which is based on the bound formula [86, page 123]

$$R(\beta_l^k) \leq R_{\text{emp}}(\beta_l^k) + \Phi(\frac{l}{h_k}) \tag{4.4}$$

where the first term is the empirical risk and the second one the confidence interval. The term on the left of the inequality is the expected risk while the index $l$ is the number of

training samples, $k$ the capacity and $h_k$ the Vapnik-Chervonenkis (VC) dimension of the learning machine. The Structural Risk Minimisation (SRM) principle minimises this risk by controlling both factors of the inequality in contrast to the ERM that minimises only the first term. For a given number of training samples $l$, the minima of the empirical risks decrease while $\Phi(\frac{l}{h_k})$ increases, as the index $k$ increases, which explains the fact that the SRM principle defines a trade-off between the quality of the approximation of the given data and the complexity of the approximating function.

The two classification methods used in this thesis differ in the constructive approach to minimise (4.4). The Probabilistic Neural Network approach minimises the first term (the number of errors in the training set) through Empirical Risk Minimisation (ERM) while the second term is a priori chosen and kept fixed. The approach taken by the SVM is exactly the opposite, minimisation of the second term through Structural Risk Minimisation (SRM) while the first term is a priori set to a value [86]. For example the structure of the PNN (number of hidden layers neurons) has to be selected a priori, while a regularisation parameter can be incorporated to penalise the complexity of the function. In fact Radial Basis Functions (RBF) were developed as part of this regularisation theory. In the case of the SVM the user selects the Kernel of the nonlinear mapping from the input to the feature space and the penalty on errors.

### 4.2.1 Probabilistic Neural Network (PNN)

PNNs are part of the larger family of feedforward RBF Neural Networks that respond to an input pattern by processing the input data from one layer to the next with no feedback paths. The PNN is based on the estimation of the conditional probability density functions (pdf) from a training set using a sum of Gaussian kernels, and choosing the class that minimises the expected risk. Considering the general case, one is faced with the problem of classifying a given set of p-dimensional input vectors (the feature vectors) $\mathbf{x}^T = [x_1....x_j....x_p]$ in one of M classes $\theta_1$, $\theta_2$, ... $\theta_M$.The decision risk of classifying the input $\mathbf{x}$ into class $j$ is given by (4.5). Here $\omega_m$ denotes the a priori probability of class m and $c_{jm}$ the loss associated with classifying an input vector $\mathbf{x}$ into class $j$ given that $\mathbf{x}$

$$\rho_j(\mathbf{x}) = \sum_{m=1}^{M} c_{jm}\omega_m f_m(\mathbf{x}) \tag{4.5}$$

belongs to class $m$. The minimum decision risk is to classify $\mathbf{x}$ into that class $j$ having the minimum risk, that is $j = arg\ min\{\rho_j(\mathbf{x})\}$. The decision $j$ is the optimum Bayesian

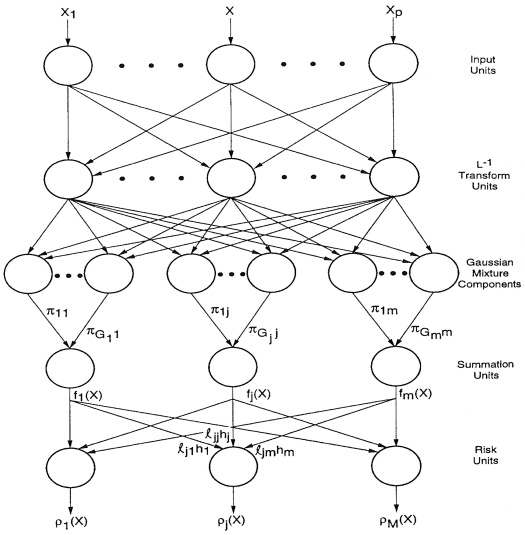



Figure 4.1: Probabilistic Neural Network using a mixture of Gaussian densities (after Streit and Luginbulh [5]).

classification decision [88]. The discriminant function can be arbitrarily complex and it can approximate any polynomial discriminant function [89]. The key to using (4.5) is the ability to estimate pdfs based on training patterns. According to [90] and [91] the multivariate estimate of the pdf of class $j$ can be expressed by $g_j$ as

$$g_j(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}\sigma^p}\frac{1}{T_j} \sum_{i=1}^{T_j} \exp\left\{-\frac{(\mathbf{x}-\mathbf{x}_{ji})^T(\mathbf{x}-\mathbf{x}_{ji})}{2\sigma^2}\right\} \tag{4.6}$$

where $j$ is the class, $i$ the pattern number in the $j^{\text{th}}$ class, $T_j$ the total number of training patterns in class $j$, $\mathbf{x}_{ji}$ the $i^{\text{th}}$ training sample from class $j$, $\sigma$ is a smoothing parameter and $p$ the dimensionality of the space. The pdf estimate for class $j$ is simply the sum of all multivariate Gaussian distributions centred at each training sample. This means that for every training sample a node is needed in the hidden layer of the NN. As all the training samples have to be stored and used in the calculation of the expected risk, the PNN suffers from large computation times. However if the maximum likelihood training of the PNN is employed, which is based on the Generalised Fisher (GF) training [5] then the number of nodes can be reduced significantly, which reduces the computation time. Figure 4.1 shows the structure of the PNN trained with this method. The pdf of class $j$,

$f_j(\mathbf{x})$, is approximated by a general mixture pdf, denoted by $g_j(\mathbf{x})$, that is,

$$f_j(\mathbf{x}) \approx g_j(\mathbf{x}) = \sum_{i=1}^{G_j} \pi_{ij} p_{ij}(\mathbf{x}), \qquad j = 1, ...., M \tag{4.7}$$

where $\pi_{ij}$ denotes the proportion of component $i$ in class $j$ and $G_j$ the number of different components in the $j$th class mixture pdf. The pdf $p_{ij}$ is assumed to have the form

$$p_{ij}(\mathbf{x}) = (2\pi)^{-N/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_{ij})^T \Sigma^{-1} (\mathbf{x} - \mu_{ij})\right\} \tag{4.8}$$

where $\mu_{ij}$ is the mean vector and $\Sigma$ is the positive definite covariance matrix of $p_{ij}$. Now the problem of the training algorithm is to estimate the parameters, $\lambda$, defining the homoscedastic mixture of mixtures, which comprises of the following variables:

- $\omega_j$ = the a priori probability of class $j$,
- $\pi_{ij}$ = the mixing proportion of component $i$ in class $j$,
- $\mu_{ij}$ = the mean vector of component $i$ in class $j$, and
- $\Sigma$ = the common covariance matrix of all Gaussians.

Estimation of $\lambda$ can be seen as estimation of incomplete data as the component $i$ in which a training vector of class $j$ belongs to is not known but has to be estimated from the observed data. Therefore the Expectation Maximisation (EM) algorithm [92] computes the maximum likelihood parameter estimates. The number of components $G_j$ in class $j$ is a parameter that has to be selected by the user since NNs are based on ERM to minimise the error between the estimated and labelled data. This is equivalent to the selection of $h_k$ in equation 4.4 that affects the generalisation ability of the PNN. A high value will result in zero training error (first term in 4.4) but high variance while a low value gives a low variance but a high bias. Methods for selecting such parameters from a data set do exist in the literature and are employed in the relevant section of parameter selection. Support Vector Machines do not have such a model order selection problem but have other user selected parameters as seen in the following section.

### 4.2.2 Support Vector Machine (SVM)

The SVM approaches the solution to classification by using a linear function to separate the two classes. The construction of maximum margin separating hyperplanes between

two classes plays a very important role related to the fact that due to mathematical theories of maximisation a function called the dual function can be maximised instead. This maximisation relates not to the dimensionality of the feature space but to the number of training vectors that reside on the maximum margin separating hyperplanes, the so called support vector machines. For example, figure 4.2 shows two linearly separable clusters of data that are described by two features ($x_1$ and $x_2$). The construction of an optimal canonical separating hyperplane (OCSH) with the largest margin is the ultimate learning goal in statistical learning theory underlying SVM [85]. A canonical hyperplane is the one satisfying

$$\min_{\mathbf{x}_i \in \mathbf{x}} |\mathbf{w}^T \mathbf{x}_i + b| = 1. \tag{4.9}$$

The geometrical margin $M$ is given by

$$M = \frac{2}{\|\mathbf{w}\|}. \tag{4.10}$$

The points closest to the OCSH, that satisfy

$$y_j |\mathbf{w}^T \mathbf{x}_i + b| \equiv 1, \quad i = 1, N_{sv} \tag{4.11}$$

are support vectors, where $N_{sv}$ is the number of support vectors. This hyperplane obtained from limited training samples must have a maximal margin because it will probably better classify new data and must be in canonical form because this will ease the quest for significant patterns, here called support vectors. During training, a minimization of the canonical hyperplane weight norm $\|w\|$ maximizes the margin given by (4.10) and minimizes the VC dimension at the same time [86], implementing a SRM.

Thus, in order to find the optimal separating hyperplane having a maximal margin, a classic non-linear optimization problem with inequality constraints must be solved: a learning machine should minimize $\|w\|$ subject to the following inequality constraints

$$y_j (\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i = 1, l. \tag{4.12}$$

The optimisation of such a problem with equality and inequality constraints can be solved using the Lagrange formalism, which leads to the primal form of the objective function

$$L_p(\mathbf{w}, b, \alpha) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \sum_{i=1}^{l} \alpha_i \{ y_j [\mathbf{w}^T \mathbf{x}_i + b] - 1 \} \tag{4.13}$$

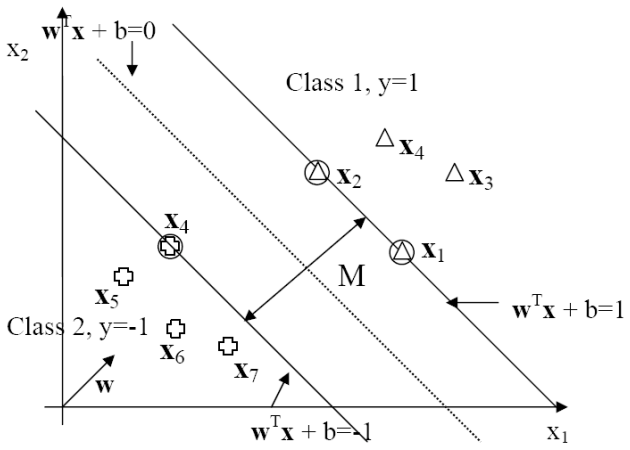



Figure 4.2: Optimal Canonical Separating Hyperplane (OCSH).

where $\alpha_i$ are the Lagrange multipliers. Lagrangian treatment of convex optimisation problems leads to an alternative dual description where the solution of the problem of minimising $\mathbf{w}^T\mathbf{w}$ subject to constraints 4.12 is equivalent to determining the saddle point of the function $L_p(\mathbf{w}, b, \alpha)$, at which it is minimised with respect to $\mathbf{w}$ and maximised with respect to $\alpha_i$ which gives the dual form of the Lagrangian

$$L_d(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j. \tag{4.14}$$

In order to find the optimal hyperplane, the dual Lagrangian must be maximised with respect to non-negative $\alpha_i$

$$\alpha_i \geq 0 \qquad i = 1, l \tag{4.15}$$

subject to constraints

$$\sum_{i=1}^{l} \alpha_i y_i x_i. \tag{4.16}$$

The importance of the solution of this optimisation problem in the so called dual space is that it is expressed in terms of the training data and depends only on the scalar products of input patterns $(\mathbf{x_i}, \mathbf{x_j})$. Furthermore utilising the Karush-Kuhn-Tucker (KKT) complementary conditions which state that only Lagangian multipliers corresponding to points on the maximal margin hyperplane $(\alpha_{oi})$ are non-zero, then the dual representation can

be expressed as

$$\sum_{i=1}^{l} y_i \alpha_{oi} \langle \mathbf{x_i} \cdot \mathbf{x_j} \rangle \tag{4.17}$$

and

$$\sum_{i \in \mathrm{SV}}^{l} y_i \alpha_{oi} \langle \mathbf{x_i} \cdot \mathbf{x_j} \rangle. \tag{4.18}$$

The optimal hyperplane is now defined by

$$\mathbf{w}_o = \sum_{i=1}^{N_{\mathrm{sv}}} \alpha_{oi} y_i \mathbf{x}_i \tag{4.19}$$

$$b_o = \frac{1}{N_{sv}} \left( \sum_{s=1}^{N_{\mathrm{sv}}} \frac{1}{y_s} - \mathbf{x}_s^T \mathbf{w}_o \right). \tag{4.20}$$

The summation goes over only support vectors since the Lagrange multiplier of non support vectors vanishes and can be calculated using the very few support vectors of the training data set and not the whole training data. The decision hyperplane is now

$$d(\mathbf{x}) = \sum_{i=1}^{l} y_i \alpha_{oi} \mathbf{x}^T \mathbf{x}_i + b_o. \tag{4.21}$$

With reference to figure 4.2, the support vectors are enclosed in circles ($\mathbf{x}_1, \mathbf{x}_2$ for class 1 and $\mathbf{x}_4$ for class 2).

The above analysis can be extended to cases where the data are not separable. If the procedure above is applied to such data then the algorithm chooses almost all training points as support vectors. In order to avoid this a soft margin classifier is implemented instead where all the data in this margin, whether on the correct or wrong side of the separating line are ignored. This is achieved by the introduction of non negative slack variables in the optimisation problem [93].

A design penalty parameter C, is introduced to control the penalty assigned to errors. Higher C means higher penalty to misclassifications, simultaneously resulting in larger weights. This parameter controls the maximum allowable value for the Lagrange multipliers $a_i$ in the dual optimisation process. The previous case of linearly separable data can be considered a special case of the non separable case with the value $C = \infty$. This parameter has to be selected by the user during the construction of an SVM and together with the kernel are the only parameters of the algorithm.

For non-linearly separable data sets, the classification ability of the above optimal

hyperplane is limited. A solution is to map the input vectors to a high-dimensional feature space, $F$, by using a non-linear mapping $\mathbf{z} = \mathbf{\Phi}(\mathbf{x})$ (where $\mathbf{\Phi}$ represents a mapping $X \mapsto F$) and use linear classification analysis in this feature space. Unfortunately this mapping creates a feature space of very high dimensionality and hence poses a computational problem when working in this space. Recalling that in (4.21) training samples appear only in the form of scalar products $\mathbf{x}_i^T\mathbf{x}_j$ then replacing them by $\mathbf{z}_i^T\mathbf{z}_j$ in the feature space $F$, can be expressed through a kernel function

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{z}_i^T\mathbf{z}_j = \mathbf{\Phi}^T(\mathbf{x}_i)\mathbf{\Phi}(\mathbf{x}_j). \tag{4.22}$$

The advantage of this kernel is that it is a function in the input space and thus avoids a mapping, $\mathbf{\Phi}(\mathbf{x})$, in the higher dimensional feature space. Instead, the required scalar products in a feature space $\mathbf{\Phi}^T(\mathbf{x}_i)\mathbf{\Phi}(\mathbf{x}_j)$ are calculated directly by computing kernels $K(\mathbf{x}_i, \mathbf{x}_j)$ for given training data vectors in an input space. In this way, one bypasses the possibility of an extremely high dimensional feature space $F$. Thus, using the chosen kernel $K(\mathbf{x}_i, \mathbf{x}_j)$, an SVM can be constructed that operates in an infinite dimensional space [85]. The decision surface now is given by

$$d(\mathbf{x}) = \sum_{i=1}^{l} y_i\alpha_i K(\mathbf{x}, \mathbf{x}_i) + b \tag{4.23}$$

where it can be seen that the computations are done in the lower dimensional input space. A Kernel has to follow certain mathematical conditions (Mercer's conditions) and the choice of the kernel used in this work is explained in the choice of the parameters section. Next the methodology followed to select the relevant parameters for each algorithm is explained.

## 4.3 Methodology for Parameter Selection

As explained in the previous section both classification algorithms utilised in this work have free parameters which have to be selected by the designer of the system a priori. Generally there is no approved method for selecting them and various techniques can be utilised. In this work a grid-search algorithm is selected to find their optimal values. In order to compare the results produced by different parameters a criterion has to be selected such as for example the training error. However, in classification algorithms a low training error on a given dataset does not guarantee good generalisation ability.

Therefore the grid-search algorithm needs to take into account the estimation error on previously unseen data as a criterion for the parameter selection.

A solution to the above problem can be the use of the holdout method where the available data is separated into two sets, the training and test set. As implied by the names the training data set is used for training the algorithm and the test set is used for estimating the classification accuracy of the trained classifier (its generalisation ability). However this method is a single train-and-test experiment and the estimate of the classification error will be misleading if it happened to get an unfavourable split. The $K$-fold cross validation overcomes this problem by resampling at the expense of more computations. A $K$-fold partition of the dataset is created and $K$ experiments are carried out. For each of the $K$ experiments $K-1$ folds are used for training and the remaining one for testing. Now the classification error is estimated as the average classification error of the $K$ experiments. The choice of the number of folds, $K$, is a trade off between the variance and the bias of the estimator as well as of the structure and sparsity of the dataset. A large number of folds reduces the bias of the estimator (very accurate estimator) at the expense of computational time but increases the variance of the true classification error. On the other hand a small number of folds reduces the computational time, decreases the variance of the estimator but increases the bias resulting in a conservative or higher classification error than the true error. A compromise between the computation time and the number of folds is reached by setting $K$=5. This methodology is applied to the selection of the relevant parameters both in the time and phase resolved proposed features. Therefore next the time-resolved proposed feature and the relevant parameters selection is discussed.

## 4.4 Time Resolved Data Results

The classification performance of the proposed lower dimensional feature vector in section 3.4 is used to evaluate the suitability of this feature in separating PDs of different sources. The evaluation is done by comparing the proposed feature to a feature used in similar works that uses the un-processed, scaled wavelet coefficients [39, 78, 79, 94, 95]. Since in these works the SVM was employed, it is chosen to be used in the comparison, ensuring that the effect of the choice of a classification algorithm on the results is minimised. However before proceeding to evaluate the proposed feature vector the parameters of the algorithms are selected following the procedure in section 4.3.

The available time-resolved dataset (table 2.3) is composed of 40000 vectors. The dataset is divided into 5 sets with 8000 pulses each where four folds are used to train the algorithm and one to test it. After the parameters selection the data is divided into a training and a test data set of 4000 and 36000 pulses respectively. The same division is used in both the SVM and the PNN classification methods to be able to compare the results and eliminate the effect of any unfavourable division affecting one of the algorithms. Therefore in the next section the classification results for the test data set, trained using the training set and the optimum parameters calculated using the above procedure, are presented.

### 4.4.1 Proposed vs higher dimensional feature vector using the SVM

The RBF kernel is the most widely chosen kernel in the nonlinear mapping from the input to the feature space in SVM, mainly due to the fact that it provides a similarity measure with tractable properties, generating a zero value when the inputs are apart and one when close together, as deduced from 4.24. Moreover it has been used in previous works showing good performance in PD classification [39] and is chosen to be used with the SVM.

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{||\mathbf{x} - \mathbf{x}_i||^2}{2\gamma^2}\right) \tag{4.24}$$

The two parameters that have to be optimised are $\gamma$, the smoothing parameter of the exponential kernel and $C$, the bound on the Laplace coefficient indirectly controlling the number of misclassifications. Both are grid-searched using cross validation where the large range of values of both $C$ and $\gamma$ are best exploited by an exponential search. The base two exponential is chosen and the value of C is set to $2^{19}$ and $\gamma$ to $2^{-14}$ giving a cross validation accuracy of around 96% as shown in figure 4.3.

The proposed lower dimensional feature for time resolved data is validated against a similar feature extraction method used in [39, 79] in order to demonstrate whether this lower dimensional feature can be used in identifying PDs of different origin. The results of the proposed feature vector using the SVM with an RBF kernel, trained using 4000 data (1000 from each source) and tested using the remaining 36000, are shown in table 4.1.

The results of the proposed feature vector are compared to a feature extraction method that uses the actual wavelet packet coefficients, after being normalised so at each node

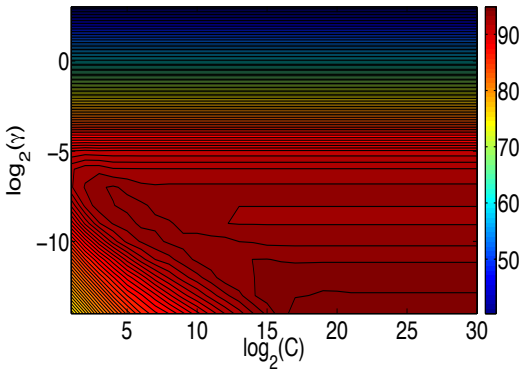



Figure 4.3: Plot of the optimum parameters search using the cross validation technique for the SVM using the proposed feature.

Table 4.1: Misclassification matrix using the SVM with $C = 2^{19}$ and $\gamma = 2^{-14}$ for the proposed feature that utilises the HOS for dimensionality reduction.

| | Input Class | | | | Sample Size | |
|---|---|---|---|---|---|---|
| Decision | Corona | Floating | Internal | Surface | Test | Training |
| Corona | 99.90% (8991) | 0.09% (8) | 0.16% (14) | 0.06% (5) | 9000 | 1000 |
| Floating | 0.01% (1) | 97.33% (8760) | 1.44% (130) | 2.47% (222) | 9000 | 1000 |
| Internal | 0.09% (8) | 1.83% (165) | 98.37% (8853) | 0.82% (74) | 9000 | 1000 |
| Surface | 0.00% (0) | 0.74% (67) | 0.03% (3) | 96.67% (8699) | 9000 | 1000 |

they have a range between +1 and -1, as features [79]. The results of employing the raw coefficients of the same nodes as the proposed vector, are shown in table 4.2, where the optimum parameters ($C = 2^{30}$ and $\gamma = 0.0156$) are used.

The use of the un-processed, scaled wavelet coefficients achieves slightly better results than the HOS of the same coefficients (98.47% and 98.06% respectively). This deterioration in performance of the proposed lower dimensional vector is small and cannot be considered detrimental. The significance of the results lies in the fact that the classification performance is kept almost at the same level while incorporating the advantages of the proposed feature extractor. The most important improvement is the dimension-

Table 4.2: Misclassification matrix using the SVM with $C = 2^{30}$ and $\gamma = 2^{-6}$ on the feature vector that utilises the scaled, raw, wavelet coefficients at specific levels.

| | Input Class | | | | Sample Size | |
|---|---|---|---|---|---|---|
| Decision | Corona | Floating | Internal | Surface | Test | Training |
| Corona | 100.00% (9000) | 0.10% (9) | 0.28% (25) | 0.06% (5) | 9000 | 1000 |
| Floating | 0.00% (0) | 95.43% (8589) | 0.99% (89) | 0.00% (0) | 9000 | 1000 |
| Internal | 0.00% (0) | 4.47% (402) | 98.61% (8875) | 0.11% (10) | 9000 | 1000 |
| Surface | 0.00% (0) | 0.00% (0) | 0.12% (11) | 99.83% (8985) | 9000 | 1000 |

ality reduction from 1304 down to 16 (a reduction of about 82 times) which can play a significant role in the quest to move to on-line, real-time PD monitoring where the reduction of classification time and storage requirements are very important. In order to demonstrate the improvement in classification complexity the time to classify the test dataset is compared for the two features. The number of support vectors (SV) for the two cases is 438 for the novel feature and 1842 for the un-processed wavelet coefficients. The number of support vectors relates to the generalisation ability of the algorithm through an inverse relationship [93, page 101], [85, page 162] and to the classification time. The proposed feature resulted in a faster classification time (3.00 seconds compared to 636.90 seconds on the same PC) for the whole test dataset. In an on-line, real-time implementation scenario this improvement in classification time is very important and the successful classification of PDs using feature vectors of such low dimensions is a push towards its actual implementation.

In the next section two classification algorithms are compared using the proposed feature vector. The choice of these two algorithms is based on the fact that although under some circumstances they can be structurally equivalent, their training methods employ different criteria as explained in section 4.2. The comparison is possible due to the fact that the proposed lower dimensional feature vector alleviated singularity problems associated with the PNN.

### 4.4.2 PNN and SVM comparison using the proposed feature

The PNN is chosen as an alternative classification algorithm for the evaluation of the proposed vector. The two algorithms are based on different training principles and the comparison of the two algorithms demonstrates the effect of the training methods. As with the SVM, the PNN has free parameters that have to be a priori selected by the user. The selection of parameter $G_j$ is an important one and unfortunately this parameter cannot be estimated from the Expectation Maximization (EM) algorithm used to train the PNN but it has to be selected a priori. It represents the number of nodes used to approximate the pdf of class $j$ in 4.7. Selection of this parameter will have an effect on the computational burden of the algorithm as well as on the classification results, and its choice is based on grid-search using cross validation. Values of $G_j$ from 1 to 5 for each class are considered and optimum results are obtained for $G = [1\ 2\ 3\ 4]$ with a cross validation classification accuracy of around 98%. The classification results are shown in table 4.3 and can be directly compared with those in table 4.1. Hence a direct comparison between the SVM method used in [39, 78, 94] and the PNN used in [45, 96] is performed, where the results are truly comparable since the same feature vector and dataset are used.

The two methods provide very similar results indicating that they can be considered two equivalent methods as far as the classification rate of PD of different sources is concerned. In order to investigate whether the partition of the data is more favourable for one of the methods and whether these results are reproducible, the method of tolerance intervals is employed.

### 4.4.3 Classification Accuracy Assessment by Tolerance Intervals

Tolerance intervals assess statistically the range of values that some specified proportion of the population falls in, with some probability. Reproducibility is examined by removing a random uniform set of training samples, $S(= 4000)$, from the available samples $T(= 40000)$, training the algorithms with $S$ and then using the remaining samples to evaluate the performance. Iterative application of this procedure results in a number of training trials which are considered independent realisations of a multivariate discrete random variable. The realisations are independent because of the independence of the subsets $S$ drawn from $T$, and discrete since there is only a finite number of different possible subsets $S$ that can be drawn from $T$ [5]. The univariate case was extended to the multivariate case

Table 4.3: Misclassification matrix using the PNN with $G_1 = 1$, $G_2 = 2$, $G_3 = 3$ and $G_4 = 4$.

| | Input Class | | | | Sample Size | |
|---|---|---|---|---|---|---|
| Decision | Corona | Floating | Internal | Surface | Test | Training |
| Corona | 99.97% (8997) | 0.00% (0) | 0.24% (22) | 0.23% (21) | 9000 | 1000 |
| Floating | 0.00% (0) | 96.96% (8726) | 1.19% (107) | 1.43% (129) | 9000 | 1000 |
| Internal | 0.03% (3) | 3.03% (273) | 98.56% (8870) | 0.02% (2) | 9000 | 1000 |
| Surface | 0.00% (0) | 0.01% (1) | 0.01% (1) | 98.31% (8848) | 9000 | 1000 |

through the concept of statistically equivalent blocks [97, 98] and to the discontinuous case [99]. From the tables in [100] 413 training trials are performed so that a population coverage of 0.9 with a 0.95 confidence between the maximum and minimum value in the training trials is observed after removing the appropriate statistically equivalent blocks [101].

The tolerance intervals for the SVM and the PNN are shown in Tables 4.4 and 4.5 respectively. The way the tolerance intervals are presented is the inclusion of a mean value with a deviation value. Addition of the deviation value to the mean value gives the maximum of the tolerance interval while subtraction of the deviation value gives the minimum one. The analysis confirms that the results in section 4.4.2 are within the

Table 4.4: Tolerance Intervals for SVM with $C = 2^{19}$ and $\gamma = 2^{-14}$.

| | Input Class | | | |
|---|---|---|---|---|
| Decision (%) | Corona | Floating | Internal | Surface |
| Corona | 99.61±0.36 | 0.30±0.28 | 0.46±0.36 | 0.07±0.07 |
| Floating | 0.13±0.13 | 96.97±0.99 | 1.83±0.93 | 2.37±0.88 |
| Internal | 0.32±0.31 | 2.27±0.99 | 97.63±1.02 | 0.88±0.61 |
| Surface | 0.08±0.08 | 0.76±0.56 | 0.31±0.29 | 96.92±1.17 |

Table 4.5: Tolerance Intervals for PNN with $G_1 = 1$, $G_2 = 2$, $G_3 = 3$ and $G_4 = 4$.

| | Input Class | | | |
|---|---|---|---|---|
| Decision (%) | Corona | Floating | Internal | Surface |
| Corona | 99.88±0.09 | 0.01±0.01 | 0.43±0.40 | 0.43±0.42 |
| Floating | 0.00±0.00 | 97.06±1.96 | 2.21±1.82 | 2.44±2.06 |
| Internal | 0.12±0.09 | 2.93±1.97 | 97.41±2.07 | 0.04±0.04 |
| Surface | 0.00±0.00 | 0.24±0.24 | 0.07±0.07 | 97.36±2.13 |

confidence interval for all of the discharge types. This translates into good reproducibility of the results.

Next the actual tolerance intervals for the PNN and the SVM are compared. Ideally a superior classification algorithm should have higher middle values and lower deviation values at the same time indicating a higher average classification value and tighter limits respectively. The analysis has concluded that both algorithms provided reproducible, consistent and similar results with no algorithm showing overall superiority.

In this section the proposed feature vector for use in continuous, on-line PD monitoring was compared to the use of the un-processed wavelet coefficients and showed a slight drop in classification performance from around 98.47% to 98.06%. However, this is considered acceptable taking into account the massive reduction in the number of dimensions and the considerably reduced classification time. Moreover through the use of the proposed lower dimensional feature vector a direct comparison of the performance of the PNN against the SVM in classifying the PD data was made possible and the method of tolerance intervals was used to assess the consistency and reproducibility of the classification results. The results demonstrated that both methods provided consistent and reproducible results with very similar identification rates.

Once the time-resolved proposed feature vector is evaluated and found to produce comparable results in PD identification with existing methods when trained using a supervised algorithm noise is added in an effort to relate the performance on the experimental measurements to field conditions.

### 4.4.4 Performance with Additive White Gaussian Noise (AWGN)

The experimental data utilised in evaluating the performance of the proposed time-resolved feature vector is acquired under controlled conditions and is not affected by

noise. In field conditions however noise can pose a problem and in this section AWGN is added to the experimental data and the identification performance of the proposed method quantified with respect to the Signal to Noise Ratio (SNR). The SNR is calculated in dB, using

$$(\mathrm{SNR})_{\mathrm{dB}} = 10 \log \left( \frac{x_s^2}{x_n^2} \right) \tag{4.25}$$

where $x_s^2$ and $x_n^2$ are the variances of the signal and noise respectively.

Table 4.6: Classification rate for the proposed time-resolved feature vector with the addition of White Gaussian Noise using the SVM.

| | SNR (dB) | | | |
|---|---|---|---|---|
| Input Class | 10 | 7 | 3 | 0 |
| Corona | 99.96 | 99.98 | 99.97 | 99.98 |
| Floating | 95.60 | 92.91 | 82.40 | 56.20 |
| Internal | 96.84 | 93.16 | 80.23 | 61.85 |
| Surface | 89.30 | 75.32 | 40.46 | 20.76 |
| **Overall** | **95.43** | **90.34** | **75.77** | **59.70** |

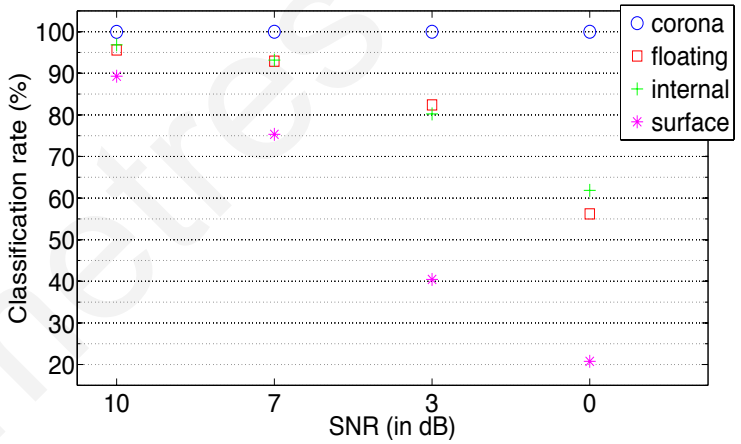



Figure 4.4: Plot of the classification rates for the different PD sources with varying SNR values, using a SVM.

Firstly the SVM is employed, trained using the optimum parameters obtained using the 5 fold cross validation on the experimental data. Then AWG noise of different variances is added to the experimental data, giving different SNR values, and classified using

Table 4.7: Classification rate for the proposed time-resolved feature vector with the addition of White Gaussian Noise using the PNN.

| | SNR (dB) | | | |
|---|---|---|---|---|
| Input Class | 10 | 7 | 3 | 0 |
| Corona | 99.96 | 97.54 | 49.68 | 6.78 |
| Floating | 97.09 | 97.13 | 96.87 | 96.57 |
| Internal | 97.77 | 96.41 | 85.20 | 63.25 |
| Surface | 96.24 | 95.42 | 90.82 | 47.40 |
| **Overall** | **97.77** | **96.63** | **80.64** | **53.50** |

the already trained model. The results are shown in table 4.4 and figure 4.6. A similar procedure is followed using the PNN where it is trained using the noise free experimental data and tested using data corrupted by AWG noise as shown in table 4.5 and figure 4.7. As expected the overall classification rate deteriorates with decreasing SNR indicating that in field conditions for an SNR value below 3 dB denoising algorithms will need to be considered in order to improve the results. Denoising can be incorporated as a pre-processing step on the time-resolved measurements due to the wider bandwidth of this data, allowing noise to be differentiated from useful signals. Since the WPT has been extensively utilised in successfully denoising PD signals through thresholding of the wavelet coefficients [62, 63, 102–104], this pre-processing step can be easily incorporated

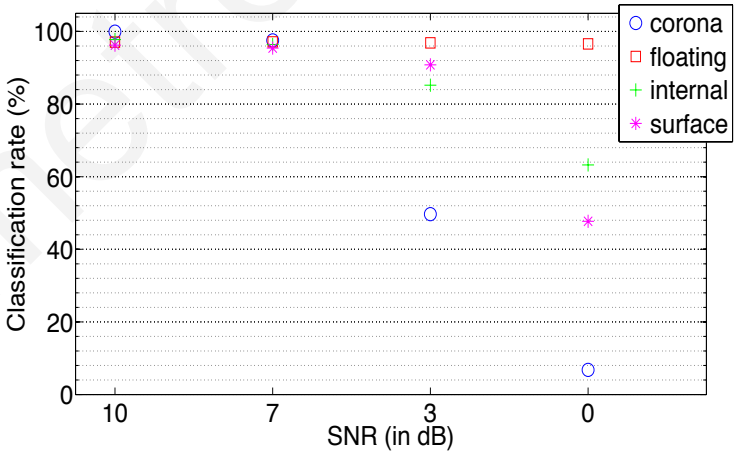



Figure 4.5: Plot of the classification rates for the different PD sources with varying SNR values, using a PNN.

into the proposed features.

In the following section the phase-resolved feature is examined to assess whether it is a valid complementary PD identification method, through the comparison of the results given by the two algorithms already utilised in this section.

## 4.5 Phase Resolved Data Results

The feature extraction algorithm for PD classification proposed in section 3.5 is used as an input to the PNN and SVM. This feature extraction algorithm uses the mean, standard deviation, skewness and kurtosis of the wavelet coefficients at each decomposition level. For the PD signals used, each proposed feature has a dimension of 36. In training the two algorithms, feature vectors from the database of laboratory simulated PD data of section 2.6.1 are used. These training vectors are calculated using 100 randomly chosen measurements of corona, floating in oil, internal in oil and surface in air discharge. There is no reason to assume that one class is more probable than the other, and this translates into setting $\omega_l$ to 0.25. There is also no reason to assign a higher risk in misclassifying a sample coming from one class into any other class. Therefore the cost of misclassification was set to one and the cost of correctly classifying a sample was taken to be zero.

### 4.5.1 Proposed vs higher dimensional feature vector using the SVM

The proposed lower dimensional feature is validated against a similar feature extraction method used in [39, 79], where the un-processed wavelet coefficients at selected nodes, scaled to have values between -1 and +1 are used as inputs to the classification algorithm. The SVM is the chosen method to validate the novel feature vector as the SVM is the method used in [39]. The question that is answered here is whether this lower dimensional feature can be used as a feature in identifying PDs of different origin using phase resolved data.

When using RBF kernels there are two parameters to be optimised, the bound on the Laplace optimisation coefficient indirectly controlling the number of misclassifications that can be tolerated, $C$, and the smoothing parameter of the exponential in the RBF used as the kernel function, $\gamma$. In this work, $C$ and $\gamma$ are grid-searched using cross validation. The large range of values of both $C$ and $\gamma$ is best exploited by an exponential search. The base two exponential is chosen and the value of C was set to 2 and $\gamma$ to $2^{-13}$

giving a cross validation accuracy of around 94.86% for the proposed feature. For the feature utilising the un-processed wavelet coefficients the value of C is set to $10^7$ and $\gamma$ to $10^{-2}$ giving a cross validation accuracy of around 96.86%. Results using the RBF kernel

Table 4.8: Misclassification matrix for the proposed feature vector using the SVM with optimum parameters $C = 2$ and $\gamma = 2^{-13}$.

| | Input Class | | | | Sample Size | |
|---|---|---|---|---|---|---|
| Decision | Corona | Floating | Internal | Surface | Test | Training |
| Corona | 98.10% (880) | 0.00% (0) | 2.56% (23) | 0.00% (0) | 996 | 99 |
| Floating | 0.00% (0) | 84.54% (760) | 0.44% (4) | 1.00% (9) | 1000 | 101 |
| Internal | 1.78% (16) | 0.00% (0) | 95.11% (856) | 0.00% (0) | 1000 | 100 |
| Surface | 0.11% (1) | 15.46% (139) | 1.89% (17) | 99.00% (891) | 1000 | 100 |

Table 4.9: Misclassification matrix for the un-processed wavelet coefficients $d_3^1$, $d_6^1$ and $d_9^1$ feature vector using the SVM with optimum parameters $C = 10^7$ and $\gamma = 10^{-2}$.

| | Input Class | | | | Sample Size | |
|---|---|---|---|---|---|---|
| Decision | Corona | Floating | Internal | Surface | Test | Training |
| Corona | 100.00% (897) | 0.00% (0) | 0.00% (0) | 0.00% (0) | 996 | 99 |
| Floating | 0.00% (0) | 88.09% (792) | 0.11% (1) | 1.89% (17) | 1000 | 101 |
| Internal | 0.00% (0) | 0.56% (5) | 99.89% (899) | 0.00% (0) | 1000 | 100 |
| Surface | 0.00% (0) | 11.35% (102) | 0.00% (0) | 98.11% (883) | 1000 | 100 |

SVM trained using 400 data (100 from each source) with the optimum parameters found

from a 5 fold cross validation and tested using the remaining 3596 are shown in table 4.8 and table 4.9. The overall identification rate for the proposed feature vector is 94.19% which is very close to the rate 96.52% provided by the un-processed wavelet coefficients at specific nodes. However the dimensionality of the proposed feature is of the order of 36 which compared to the dimensionality of the un-processed feature of the order of 1468, signifies a significant reduction (almost 40 times). This is exactly what is sought, a feature vector of lower dimensionality but with no significant drop in performance, to be used in a continuous on-line monitoring where the reduction in computational burden can be significant. Once the proposed feature has been validated as having an acceptable classification performance two classification algorithms are compared, the PNN and SVM using this feature.

### 4.5.2 PNN and SVM using the proposed feature

The two classification algorithms are based on different training principles and are compared using the proposed feature, in order to investigate whether one of the training methods gives superior results. The PNN has a set of a priori selected parameters, the number of nodes needed to represent the pdf of class $j$ in 4.7. The method of cross validation is employed considering values of $G_j$ from 1 to 5 for each class and optimum results are obtained for $G$=[1 2 1 1] with a cross validation classification accuracy of around 84.40%.

The proposed feature vector lives in a 36-dimensional space that is very difficult to be visualised by humans that are accustomed living in a 3-dimensional world. Therefore a transformation into a 2 or 3-dimensional plot can facilitate the understanding of how the different PD sources are represented by the proposed feature. Principal Component Analysis (PCA) is such an effective linear transformation that projects high dimensional data onto a lower dimensional space.

#### 4.5.2.1 Principal Component Analysis (PCA)

The Principal Component Analysis is a transformation that projects the data $\mathbf{x}$ into the directions defined by its eigenvectors where the coefficients of this projection are the eigenvalues. The $p$-dimensional mean value $\mu$ and the covariance matrix $\mathbf{\Sigma}$ $(= (\mathbf{x}-\mu)^T(\mathbf{x}-\mu))$ are calculated from the full data set $\mathbf{x}$. The computed eigenvalues of the covariance matrix are sorted in an ascending order and the $m$ largest values are selected with their

corresponding eigenvectors. The PCA is then the projection onto the $m$-dimensional subspace given by

$$\mathbf{x}' = \mathbf{A}^T(\mathbf{x} - \mu) \tag{4.26}$$

where $\mathbf{A}$ is the $p \times m$ matrix of the $m$ eigenvectors [59]. In our case the 36-dimensional phase-resolved feature vector is projected onto its two dimensional subspace ($m = 2$) spanned by the two eigenvectors with the largest eigenvalues of the covariance matrix of the input data. The plot of the first 2 principal components, used as a visualisation

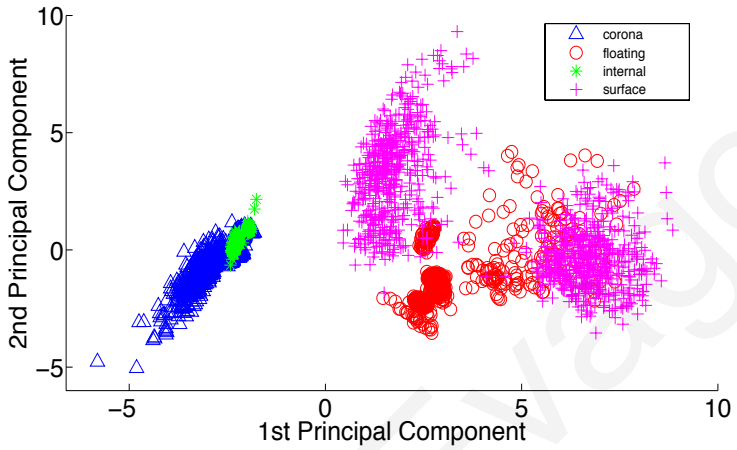



Figure 4.6: Plot of the first two Principal Components of the phase-resolved feature extraction for PD data of different sources (variables normalised) for data visualisation.

tool for the data, can be seen in figure 4.6. Although it merely represents a projection of the proposed feature vector onto a plane, the results obtained in table 4.8 and table 4.10 for the SVM and PNN respectively, can be explained with the use of this figure. It is expected that the highest misclassification rates will be between the classification of floating and surface discharges, due to the floating discharge data scatter into the surface discharge data. Indeed the floating discharge data are misclassified as surface discharge data at a rate of 15.46% using the SVM, and surface discharge data are misclassified at a rate of 6.89% using the PNN, both being the worst classifications for each algorithm. The internal discharge, that exhibits the highest rate of correct classification, is projected onto a very highly concentrated space on the plane.

The PNN provides marginally better overall classification results (96.92%) compared to the SVM (94.19%) but the difference is small and in order to investigate whether the partition of the data is more favourable for one of the methods and whether these results are reproducible, the method of intervals is employed.

Table 4.10: Misclassification matrix for the proposed vector using the PNN with optimum parameter $G$=[1 2 1 1].

| | Input Class | | | | Sample Size | |
|---|---|---|---|---|---|---|
| Decision | Corona | Floating | Internal | Surface | Test | Training |
| Corona | 99.22% (890) | 0.11% (1) | 0.00% (0) | 0.78% (70) | 996 | 99 |
| Floating | 0.56% (5) | 96.22% (865) | 0.00% (0) | 6.89% (62) | 1000 | 101 |
| Internal | 0.22% (2) | 0.11% (1) | 100.00% (900) | 0.11% (1) | 1000 | 100 |
| Surface | 0.00% (0) | 3.56% (32) | 0.00% (0) | 92.22% (830) | 1000 | 100 |

### 4.5.3 Classification Accuracy Assessment by Tolerance Intervals

Using the method of tolerance intervals the consistency and reproducibility of the classification algorithms can be evaluated following the same procedure as in section 4.4.3.

Reproducibility is examined by removing a random uniform set of training samples,

Table 4.11: Tolerance Intervals for the proposed feature vector using the SVM with optimum parameters $C = 2$ and $\gamma = 2^{-13}$.

| | Input Class | | | |
|---|---|---|---|---|
| Decision (%) | Corona | Floating | Internal | Surface |
| Corona | 97.49±2.40 | 0.06±0.06 | 4.61±4.06 | 0.00±0.00 |
| Floating | 0.06±0.06 | 84.67±1.56 | 0.83±0.83 | 2.28±2.28 |
| Internal | 2.45±2.34 | 0.39±0.39 | 94.22±3.89 | 0.00±0.00 |
| Surface | 0.61±0.61 | 15.23±1.45 | 2.11±1.67 | 97.72±2.28 |

$S(= 400)$, from the available samples $T(= 3996)$, training the algorithms with $S$ and then using the remaining samples to evaluate the performance. From the tables in [100] 918 training trials are performed so that a population coverage of 0.9 with a 0.95 confi-

Table 4.12: Tolerance Intervals for the proposed feature vector using the PNN with optimum parameters $G$=[1 2 1 1].

| | Input Class | | | |
|---|---|---|---|---|
| Decision (%) | Corona | Floating | Internal | Surface |
| Corona | 98.38±1.62 | 1.33±1.33 | 0.28±0.28 | 1.28±1.28 |
| Floating | 1.00±1.00 | 92.22±5.22 | 0.11±0.11 | 6.23±6.01 |
| Internal | 1.45±1.45 | 0.33±0.33 | 99.72±0.28 | 0.50±0.50 |
| Surface | 0.00±0.00 | 6.89±5.33 | 0.00±0.00 | 93.33±6.12 |

dence between the maximum and minimum value in the training trials is observed after removing the appropriate statistically equivalent blocks [101].

The results shown in Table 4.11 and Table 4.12 support the results obtained in Table 4.8 and 4.10 for the SVM and PNN respectively. The analysis using tolerance intervals indicates reproducible results where the larger performance variations are observed when floating discharges were identified as surface discharges with the SVM and internal discharges correctly identified with the PNN.

Using phase resolved data the SVM was used to assess the performance of the feature vector extracted using the WPT and HOS. The proposed feature vector resulted in identical results with higher dimensional feature vectors but has the advantage of a significantly lower dimensionality which can be applied in continuous on-line applications where computational times must be minimised. The success rate of the SVM with this feature vector was compared to the PNN and in both cases this was above 90%, which is very promising. The PNN gave marginally better identification rates compared to the SVM. The tolerance intervals were used to show that the classification rates for both methods were reproducible. The importance of the results lies in the fact that the use of HOS resulted in a feature vector that achieves marginally inferior results compared to similar works but at a significantly lower number of dimensions which further supports the use of WPT and HOS in characterising PD signals of different sources.

## 4.6 Conclusions

In this chapter two supervised algorithms, the SVM and the PNN were used to evaluate the classification performance of the time and phase-resolved proposed features. The

experimental database of PD data was partitioned into training and testing datasets, used to train and test the algorithms respectively. The SVM was used to compare the proposed, time-resolved feature vector to a higher dimensional one giving slightly lower classification rates but with a significantly lower dimensionality. Once the proposed feature showed acceptable results, it was employed to compare the PNN to the SVM. The two algorithms provided similar results and the method of tolerance intervals was used to assess their reproducibility. As these algorithms are based on different training principles, the results showed acceptable performance for the proposed feature, irrespective of the training method used. A similar procedure to the above was followed using the phase-resolved data where the SVM was used to compare it against higher dimensional features, and also to the PNN. The phase-resolved proposed feature showed comparable results to other approaches but with a lower dimensionality.

The use of WPT and HOS, employed in the proposed features, have the potential to be used in PD source separation, using supervised algorithms. Indeed the results proved to be encouraging, achieving a significant dimensionality reduction at the expense of a slight drop in performance. This can find potential use in on-line measurements were lower storage and computational times are required. However in such scenarios it can sometimes be difficult to label the training data, as the different PD sources are a priori unknown. Unsupervised or clustering algorithms divide the data into different clusters based on some metric which can be used to provide labels to the training data or even alleviate the use of training data. In the next chapter the potential use of the proposed feature with a clustering algorithm to separate PDs of different sources from actual field measurements will be demonstrated, using time-resolved data.

# Chapter 5

# Clustering using the Proposed Time-resolved feature

In this chapter the use of the time-resolved feature vector proposed in chapter 3 and evaluated in chapter 4 is extended to on-line scenarios. In such cases the collected data used as training samples in a supervised algorithm have no a priori class labels. To overcome this problem clustering techniques can be applied to group data together based on a similarity measure, which can provide the class labels or even alleviate the need for training data. Such methods are applied to time-resolved data, as a preprocessing step and can be complemented with PRPD. Therefore the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm [105] is employed here on the proposed time-resolved feature vector to evaluate its performance as a potential pre-processing tool in on-line measurements. The evaluation is performed on a set of laboratory data and once satisfactory results obtained, field data from the distribution network are examined.

## 5.1 Introduction

In on-line scenarios supervised algorithms suffer from the fact that the training labels or the different PD sources are not a priori known, as in the case of experimental data. However the use of clustering algorithms forms "natural grouping" of the input patterns which can provide both the labels and the number of clusters for use with a supervised algorithm. Moreover the data in each cluster can be identified as originating from a different PD source, alleviating the need for training data. The formation of a cluster depends on a distance metric or the minimisation of a cost function and is classified into the partitioning and hierarchical approaches. In the former all the points in the data

set form disjoint clusters while in the latter each point can belong to several clusters depending on the level of clustering. In this work a partitioning algorithm is used where different choices of the distance metric and cost function exist.

One of the most widely used algorithms relying on a distance metric is the $K$-means clustering, used in [43] with the autocorrelation as a feature. This algorithm relies on estimating the euclidean distance of a sample from the $K$ means representing the clusters and then recomputes the means. The number of clusters is usually a priori chosen according to some knowledge about the data or through a pre-processing procedure of the data.

An application of clustering algorithms is the vector quantisation [106] where a $p$ dimensional vector $\mathbf{x}$ can be represented by a codebook of $g$ vectors using a minimum distortion criterion to select the code vectors.

The nearest-neighbour algorithm assigns each sample to the cluster its closest point belongs to, where again the distance function chosen affects the resulting partitioning. A large variety of similarity (or distance) measures and criterion functions exist for clustering which can be found in [59, 107]. The Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise (DBSCAN) [105] is used for the evaluation of the time-resolved proposed feature, in on-line scenarios using both experimental and real data.

## 5.2 Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

The DBSCAN belongs to the family of algorithms that rely on a distance metric to form clusters of arbitrary shape (in contrast to many algorithms that assume an elliptic cluster) and can identify noise. One of its advantages is that it does not require the number of clusters to be a priori selected but instead it is defined by two parameters, $\epsilon$, the minimum distance, and $k_{\min}$, the minimum number of points. The $\epsilon$-neighbourhood of a point p is defined as all the points, q, that have a distance less that $\epsilon$ between p and q. The distance metric employed is the Euclidean distance although other metrics can be used (e.g. Mahalanobis). Two types of points can be identified within a cluster, core points and border points. The DBSCAN is based on the definition of density reachability and density connectivity therefore these notions have to be defined:

- A point p is a core point if more than $k_{\min}$ points are included in a hypersphere (of volume V), with radius $\epsilon$ centred at p.
- A point q is directly density-reachable from p with respect to $\epsilon$ and $k_{\min}$ if it is included in V and p is a core point.
- A point r is density-reachable from p if there is a chain of objects $p_1$, $p_2$,...,$p_n$, where $p_1$=p and $p_n$=r, such that, for $1 \leq i \leq n$, $p_{i+1}$ is directly density-reachable from $p_i$.
- A point s is density-connected to p, if there is an object t such that both s and p are density-reachable from t.

The algorithm starts processing points randomly checking if they are core points. Once a core point is discovered it then chooses this point as the seed. Next all the points that are density-reachable from the seed are retrieved forming the cluster containing the seed. The algorithm then chooses a new unprocessed point and the whole procedure is repeated. The process terminates when no new points can be added to any of the clusters. The essence of the algorithm is that a density-based cluster is being formed based on the notions of density-connected and density-reachable points. Data not contained in a sufficiently populated cluster will be identified as noise. This clustering algorithm could serve as a pre-processing step to interference rejection as randomly occurring pulse shaped noises are expected to either be located sparsely so that no cluster will be formed or be identified as a different cluster and be rejected. Moreover identified clusters can be used as training data to a supervised algorithm or identified as different PD sources, under the assumption that each cluster represents a PD source. This clustering algorithm has been used in [108] on on-line data and has shown promising results. A similar approach will be followed here to evaluate the feature vector presented in 3.4, on experimental and on-line data for source identification and interference rejection. Due to the fact that the clustering algorithm is based on a distance metric (in this case the Euclidean distance) when the input data has a high number of data dimensions it can suffer from the curse of dimensionality. In order to minimise the number of dimensions and also make it easier to visualize the data the Principal Component Analysis (PCA) is chosen to reduce the number of dimensions.

The PCA was analysed in section 4.5.2.1 and it basically provides an effective linear transformation to project high dimensional data onto a lower dimensional space. The eigenvalues of the covariance matrix of the input data are calculated and sorted in a descending order. The data is then projected into the $m$-dimensional space spanned by

the eigenvectors corresponding to the $m$ largest eigenvalues. The projection of the 16-dimensional proposed time-resolved feature vector onto the 3-dimensional space spanned by the 3 largest eigenvalues, constitutes the input to the DBSCAN.

## 5.3 Evaluation on multiple sources using laboratory data

The use of time resolved data using the WPT and HOS on specific nodes of the wavelet packet tree which correspond to specific frequencies was proposed in section 3.4. The method provided high enough identification rates of different PD sources from laboratory measurements under controlled conditions, when trained using a supervised algorithm (chapter 4). Time domain data with a clustering algorithm can be used as a pre-processing step for separating different PD sources. The procedure followed is that PD data are acquired in synchronization with the phase at a high enough sampling rate and processed with a peak extraction algorithm to obtain the location and waveform of the pulses. Then the pulses can be analysed in the time domain using the proposed feature vector and an appropriate clustering algorithm. Different identified clusters can be considered as different subpatterns that can each correspond to a discharge source. From these clusters the PRPD data can be utilised in the identification, facilitating PD recognition. In order

Table 5.1: Details of simulated single cycle PD data acquired at a sampling rate of 500 MHz.

| **Type** | **Voltage** | **Sample Duration** | **Sample length** | **Sample Quantity** |
|---|---|---|---|---|
| Corona in air | 6 kV | 20 ms | 5000 pts | 40 |
| Corona in air | 9 kV | 20 ms | 5000 pts | 40 |
| Floating in oil | 24 kV | 20 ms | 5000 pts | 40 |
| Floating in oil | 28 kV | 20 ms | 5000 pts | 40 |
| Internal in oil | 26 kV | 20 ms | 5000 pts | 40 |
| Internal in oil | 36 kV | 20 ms | 5000 pts | 40 |
| Surface in air | 6 kV | 20 ms | 5000 pts | 40 |
| Surface in air | 9 kV | 20 ms | 5000 pts | 40 |

to test the proposed time-resolved feature vector as a pre-processing tool eighty two cycles of data for each of four PD sources at two different voltage levels (41 measurements voltage

per level) are taken in the laboratory as shown in table 5.1. The data is processed through a peak extraction algorithm and time resolved data is extracted. The mean of each cycle is subtracted from the data in order to facilitate the peak detection algorithm. Therefore the feature extraction is modified to exclude the mean from the HOS.

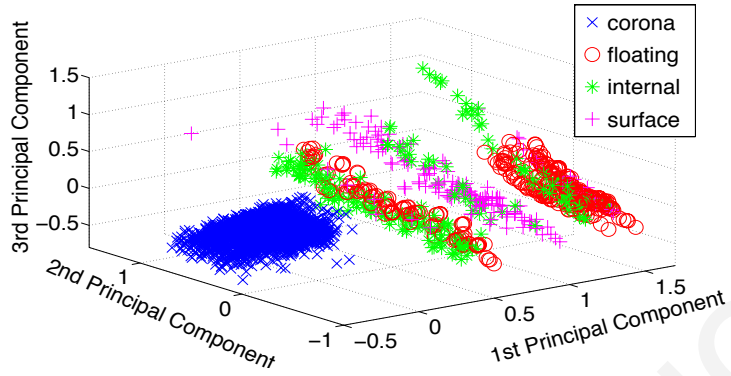



Figure 5.1: Plot of the first three principal components of the HOS feature vector on selected nodes of the wavelet tree of experimental data.

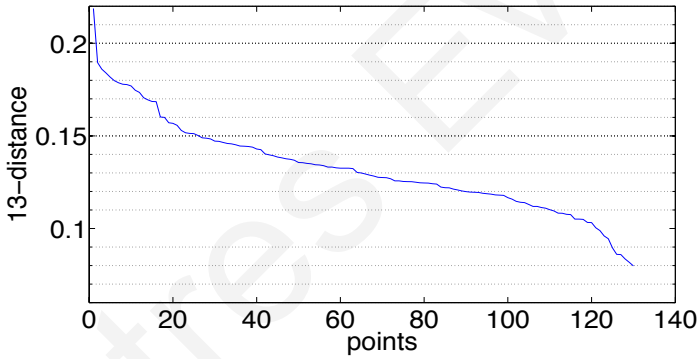
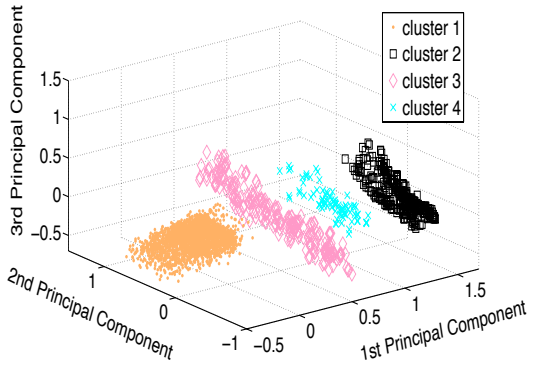
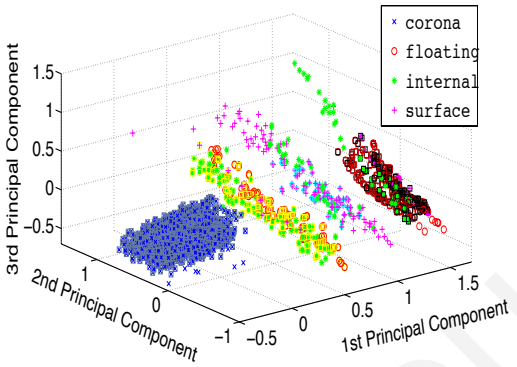



Figure 5.2: Plot of the k-distance graph, for $k$=13, for DBSCAN $\epsilon$ and $k_{\min}$ parameter selection.

In figure 5.1 the 3-dimensional plot of the PC for the different sources of PD is plotted where the first three PC retain around 89% of the signal's energy. Before running DBSCAN the parameters of the algorithm have to be selected as they can have a significant effect in the creation of clusters. The procedure in [105] is followed where a $k$-distance graph is plotted in descending order, with $k_{\min} = 13$. From this plot (figure 5.2) the parameter $\epsilon = 0.14$ is selected and the results of the clustering shown in figure 5.3a. The superimposed results of the DBSCAN and the experimental data is included in figure 5.3b. The results of the DBSCAN algorithm presented in a visual form have




(a) Plot of the identified clusters by the DBSCAN run on the first three principal components of the proposed time-resolved feature vector using experimental data




(b) Plot of the first three principal components of the HOS feature vector on selected nodes of the wavelet tree of experimental data superimposed on the result of the DBSCAN algorithm

Figure 5.3: Plots of the output of the DBSCAN algorithm on the left and superimposed with the experimental data on the right. The squares around the experimental data indicate the points that have been included in a cluster by the algorithm while their colour identifies the cluster number.

also been included in table 5.2. It can be seen that the clustering identifies well some of the sources such as the corona which forms a distinct dense cluster and shows acceptable performance for the floating (68.25%) and the internal (59.79%) discharges. The surface discharge shows the lowest performance of all discharge types (41.05%), due to the fact that the global parameters can not handle clusters of significantly different densities. The surface discharge data as seen in figure 5.1 scatters within the internal and floating discharges and forms three subclusters that can not be captured using global parameters.

To improve the classification accuracy the clustering algorithm can be combined with the SVM, a supervised algorithm considered in chapter 4. Data from the identified clusters using the DBSCAN are used as the training data where each cluster provides a class label. The optimum parameters using the 5 fold validation are selected ($C = 2^{20}$ and $\gamma = 2^{-10}$) and the data identified as outlier is processed by the SVM giving the classification rates shown in table 5.3. From this results it can be seen that the combination of the clustering and unsupervised algorithms increases the identification rate for the surface discharge to 71.05% and the overall rate from 67% to around 75%. In order to examine the efficiency of the DBSCAN algorithm on data acquired under conditions that resemble on-line conditions as close as possible an experiment is designed using a PILC cable loop.

Table 5.2: Identification matrix for the DBSCAN algorithm with parameters $k_{\min} = 13$, $\epsilon = 0.14$ evaluating the whole dataset.

| | Input Class | | | |
|---|---|---|---|---|
| Decision | Corona | Floating | Internal | Surface |
| Corona | 99.62% (2656) | 0.00% (0) | 0.00% (0) | 0.00% (0) |
| Floating | 0.00% (0) | 68.25% (288) | 14.43% (28) | 17.89% (34) |
| Internal | 0.00% (0) | 30.09% (127) | 59.79% (116) | 1.58% (3) |
| Surface | 0.00% (0) | 0.00% (0) | 5.67% (11) | 41.05% (78) |
| Outliers | 0.38% (10) | 1.66% (7) | 20.10% (39) | 39.47% (75) |

Table 5.3: Identification matrix for the outliers using an SVM with parameters $C = 2^{20}$ and $\gamma = 2^{-10}$ trained using identified clusters by the DBSCAN as training data (89 samples from each cluster randomly chosen).

| | Input Class | | | |
|---|---|---|---|---|
| Decision | Corona | Floating | Internal | Surface |
| Corona | 100.00% (10) | 0.00% (0) | 0.00% (0) | 1.33% (1) |
| Floating | 0.00% (0) | 100.00% (7) | 46.15% (18) | 5.33% (4) |
| Internal | 0.00% (0) | 0.00% (0) | 2.56% (1) | 17.33% (13) |
| Surface | 0.00% (0) | 0.00% (0) | 51.28% (20) | 76.00% (57) |

## 5.4 Evaluation using laboratory data collected from PILC loops

An experiment is designed at a high voltage laboratory where three phase belted Paper Insulated Lead Covered (PILC) cables are utilised to study the effects of introducing known defects. The actual setup of the experiment is described in detail in [109] and is briefly presented here. The aim of the experiment is to simulate operating conditions as close as possible therefore allowing control of the voltage, current and cable temperature. Since the PILC cables are operating at 11 kV then the voltage during the experiment is set to this value and the temperature is controlled by injecting a circulating current through two Current Transformers (CT). To investigate the PD signals produced by common fault mechanisms, a 10 m cable section was terminated and a straight lead joint that contains a known defect is introduced to the sample. When constructing a straight lead joint in a PILC cable section, the sweating process involves applying a molten,"solder" consisting of aluminum and tin to ensure an effective connection of the conductors within the joint. After the application of the solder material, it is imperative to wipe away any excess material and sand down the area to reduce the number of sharp points on the conductor that would lead to a high stress point and reduce the insulating performance of the joint. To mimic the effects of poor ferrule preparation, a metal spike is attached to phase 2 and orientated in the direction of phase 1. The spike is manufactured by wrapping a metal wire around the conductor and covering it with a layer of solder as shown in figure 5.4.

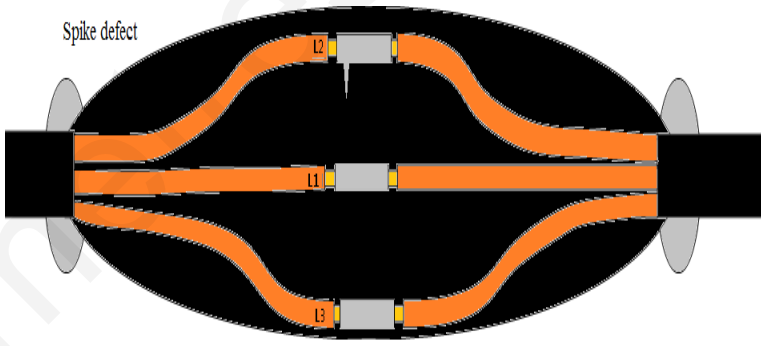



Figure 5.4: Spike induced defect in a joint/splice of a 3 phase belted PILC cable (185 $mm^2$) loop in phase 2 (L2) pointing towards phase 1 (L1).

The cable under test is thermally cycled for an extended period of time and activity due to the spike gradually reduced until it is indiscernible with respect to background

noise. To simulate another source of PD mechanical crushing is applied to the cable, as shown in figure 5.5, until PD activity is detected. However the cable deteriorates rapidly

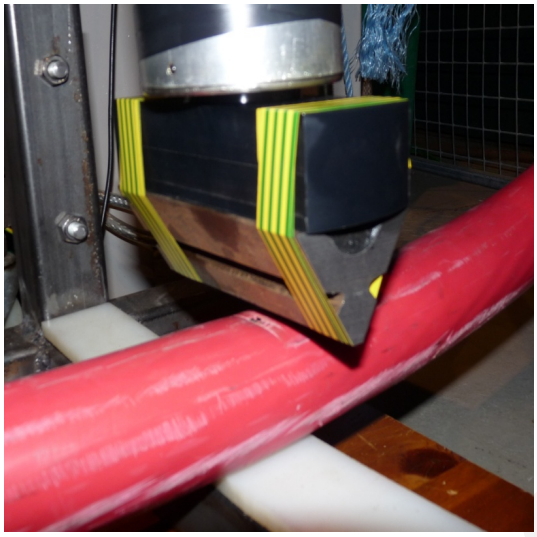


Figure 5.5: Mechanical crushing of a PILC using a sharp edge.

and fails before a large amount of PD data could be collected.

After the cable fails it is maintained at a rated temperature using two CT until the insulation could support the rated voltage. At this point a number of cycles are captured at a rate of 100 MHz using a commercially available HFCT around the earthing conductor. In figure 5.6 the PRPD data of 15 cycles after the cable could again support the rated voltage are shown and are analysed using the time-resolved proposed feature and the DBSCAN to extract the different PD sources. The system used to acquire the data triggers on the positive edge of the low voltage side therefore depending on the transformer configuration there is a phase shift between the voltage of the PRPD plot and each of the three phases on the high voltage side. The parameters $\epsilon$ and $k_{\text{min}}$ are selected according to the procedure described above and set to 5 and 4 respectively. Figure 5.7 shows the two classes identified by the DBSCAN algorithm. The respective PRPD and typical waveform plots of classes 1 and 2 are shown in figure 5.8. From this plot it can be deduced that class 1 represents the current transformer pulses and class 2 the pulses after the application of the mechanical crushing, as the cable was monitored prior to the application of the crushing and no pulses having the time-frequency distribution of the pulses in 5.8d were observed.

The proposed time-resolved feature vector and the DBSCAN appear to give meaningful results and can identify different PD sources and interference, using experimental

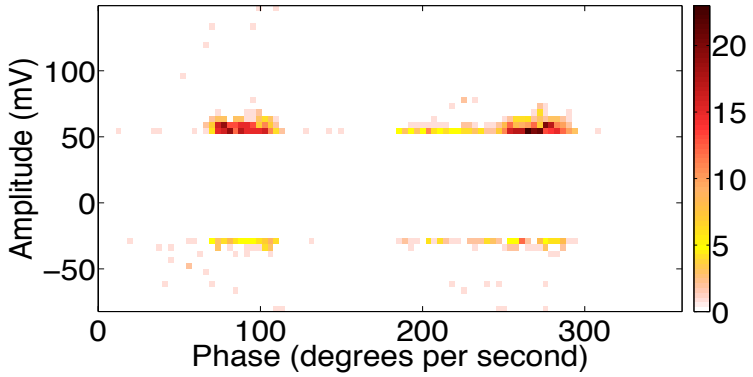



Figure 5.6: PRPD data for 15 cycles when the cable is able to withstand the rated voltage after its failure.

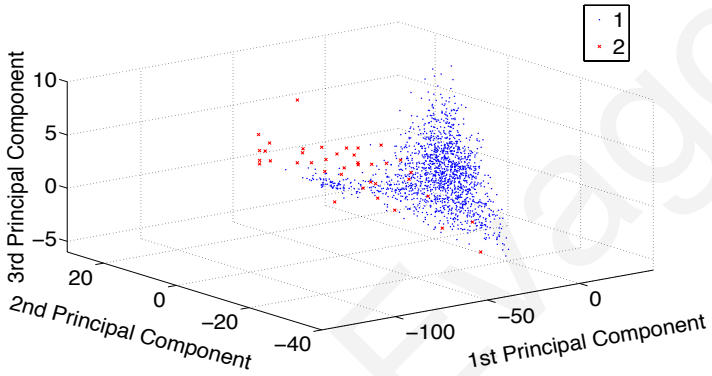
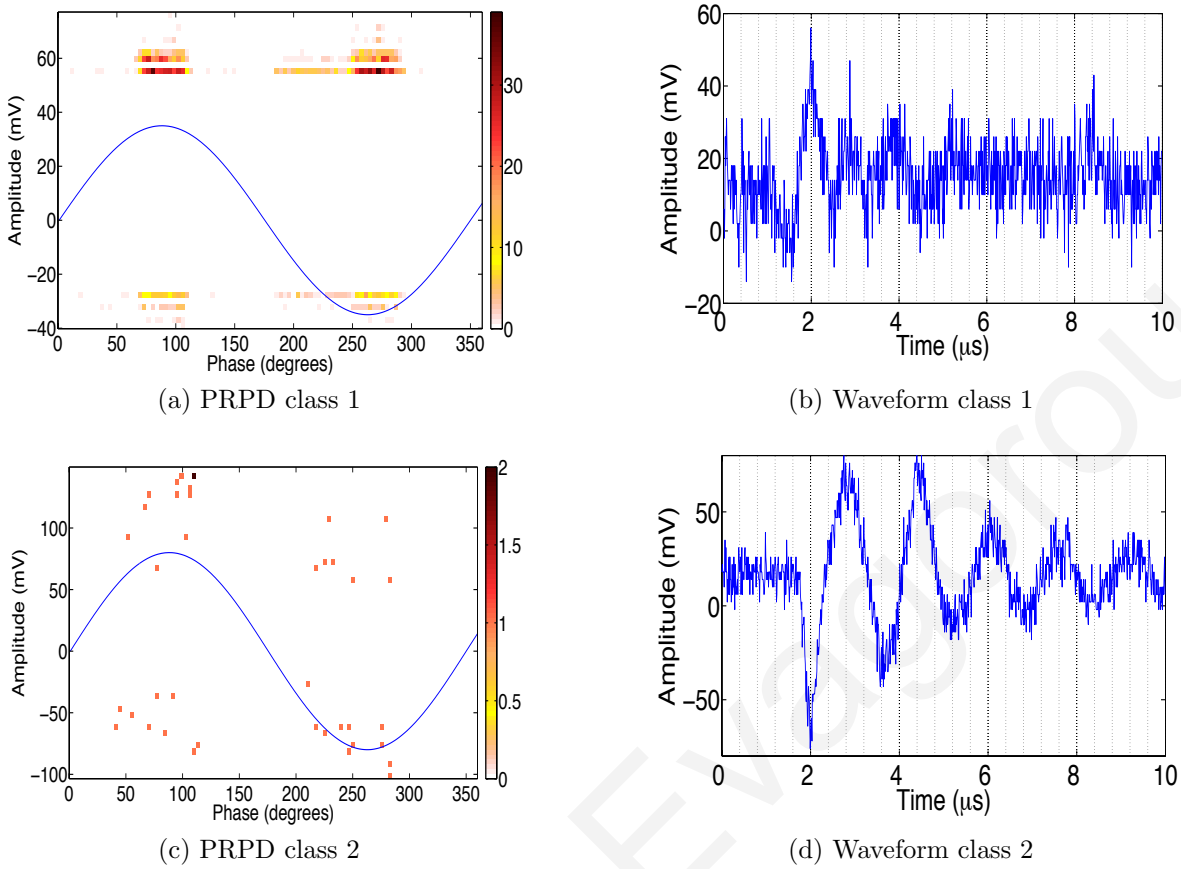



Figure 5.7: Clustering for PILC spike and mechanical crushing data.

data of PILC cable loops, designed to simulate operating conditions as close as possible. Therefore next step is to evaluate the proposed feature vector on field data acquired from the Electricity Authority of Cyprus (EAC) distribution network.

## 5.5 Evaluation using field data

The proposed feature vector is evaluated using on-line, field data that are acquired from the Cyprus distribution network belonging to the EAC. The cable loops in this network consist of combinations of lengths of PILC and XLPE cables. The PILC cables were laid many years ago, and have started to show signs of ageing. A systematic procedure of screening cable loops categorised as increased risk, based on the details and previous behaviour of the cables, was undertaken based on PD measurements. Analysis of these measurements designated specific locations that are believed to exhibit increased PD activity. The locations are under monitoring and analysed using the proposed time-




(a) PRPD class 1

(b) Waveform class 1

(c) PRPD class 2

(d) Waveform class 2

Figure 5.8: PRPD and typical single pulse data for classes 1 and 2 identified by the DBSCAN algorithm with $k_{\min} = 4$ and $\epsilon = 5$.

resolved feature to provide more insight into PD activity and as a means of evaluating the feature in situations where PD activity is believed to be present. Three examples from such loops are presented.

### 5.5.1 Field data from Pavlou Nirvana substation

The particular cable loop under investigation consists of runs of PILC and XLPE cables connected together by transitional and straight through joints, with a total length of around 400 m. The HFCT sensor analysed in section 2.4 is placed around the earthing wire at the PILC cable termination end and 120 cycles are recorded at 50 MHz in synchronization with the mains frequency. The mains is fed from an 11kV/415V delta/star Dy11 configuration transformer, which means that the secondary voltage leads the primary by 30 °C. This is taken into account when interpreting the PRPD plots, where the

sine waves of the three phases on the HV side are shown. The data is processed by a peak extraction algorithm and the feature vector proposed is calculated. After applying the PCA to reduce the number of dimensions the DBSCAN is run using the parameters $\epsilon$=0.7 and $k_{\min}$= 4.

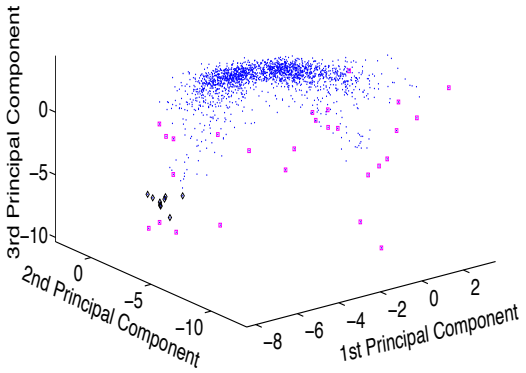



Figure 5.9: Plot of the first three principal components of the feature vector of the processed by the DBSCAN algorithm with parameters $\epsilon$ set to 0.7 and $k_{\min}$ to 4. The points marked with x are the points identified as class 1 while those marked with an x enclosed by a diamond as class 2. Outliers have been marked with an x and enclosed by a square.

The outcome is shown in figure 5.9 where it can be seen that the data is identified as two classes with some outlier points. The waveforms of selected data points are plotted in figure **??**. From this plot it can be seen that classes 1 (figure 5.10a) and 2 (figure 5.10b) contain very similar pulses and can be merged into a single class. The points identified as outliers (figure 5.10c) consist of signals that can be regarded as interference, and therefore being correctly rejected. By examination of the waveforms it can be concluded that merged classes 1 and 2 resemble the waveforms of PD activity. Figure 5.11 displays the PRPD plot from data that have been identified to belong to classes 1 and 2 and can be used as a complementary identification tool. This specific loop has been placed under monitoring and the evolution of PD activity is being recorded.

### 5.5.2 Field data from Kato Lakatamia substation

The second loop analysed also consists of mixed PILC and XLPE cables with several joints. The same procedure is followed where a peak detection algorithm is run and the extracted waveforms together with their phase of occurrence are recorded. The time-resolved proposed feature vector is extracted and processed through PCA to extract the first three principal components. The outcome of the DBSCAN algorithm using the

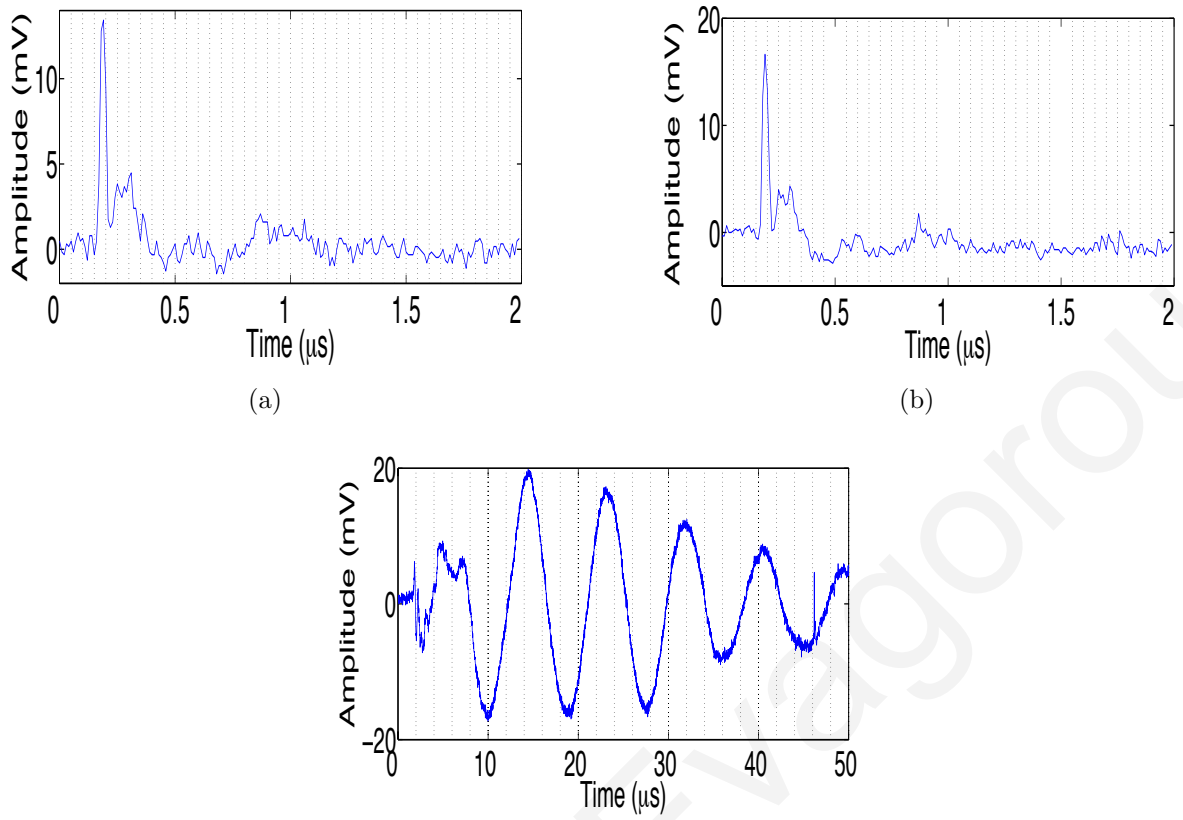



Figure 5.10: Plots of the waveforms of data clustered into class 1 (a), class 2 (b) and outliers (c).

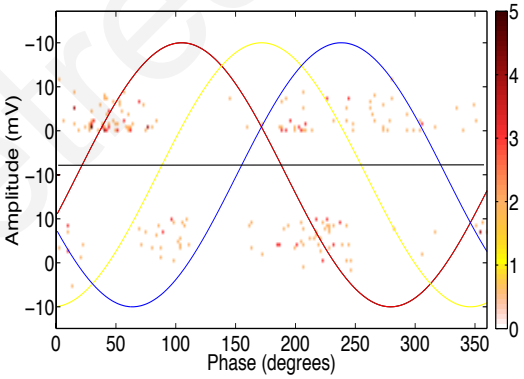



Figure 5.11: PRPD data plot of the on-line field, data at cable loop of the EAC network for classes 1 and 2 identified by the DBSCAN algorithm with $\epsilon$ set to 0.7 and $k_{\min}$ to 4.

parameters $\epsilon = 1.5$ and $k_{\min} = 4$, is shown in figure 5.12, where it can be seen that the data is clustered into a single class plus the outliers.

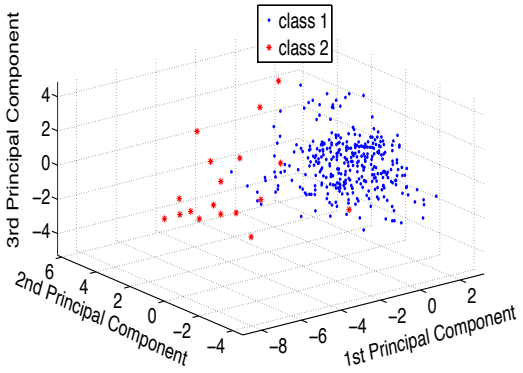
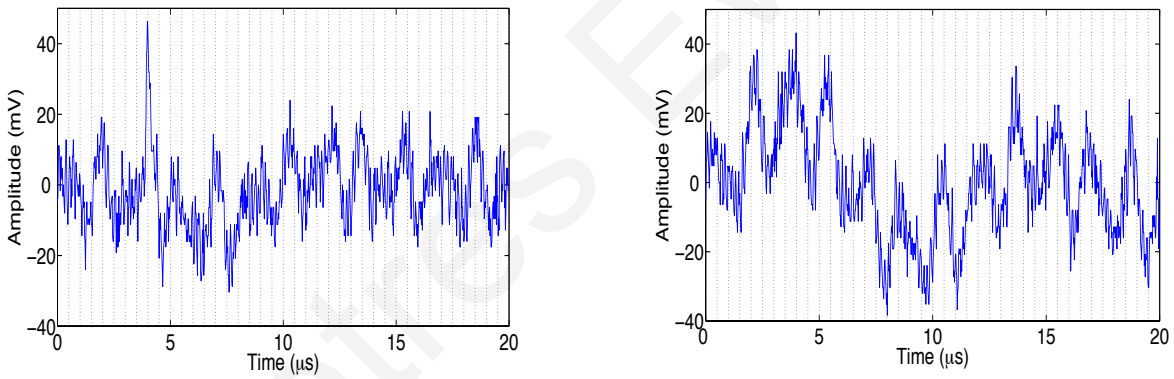



Figure 5.12: Plot of the first three principal components of the feature vector of the processed by the DBSCAN algorithm with parameters $\epsilon$ set to 1.5 and $k_{\min}$ to 4. The points marked with $\cdot$ are the ones identified in class 1 while outliers have been marked with an $*$ .

Looking into the waveforms of the identified classes it can be concluded that a signal that resembles a PD pulse is identified in class 1 (figure 5.13a) while indeed the outliers class identifies interference pulses (figure 5.13b). The PRPD plot from data identified as




(a) (b)

Figure 5.13: Plots of the waveforms of data clustered into class 1 (a) and outliers (b).

class 1 is plotted in figure 5.14 where it can be seen that the majority of the pulses occur on the rising edge of the positive ac cycle and the falling edge of the negative ac cycle (first and third quadrants) with respect to the yellow phase, which can support the fact that the data in class 1 represents actual PD events. The loop has been placed under monitoring and the evolution of PD activity is being recorded.

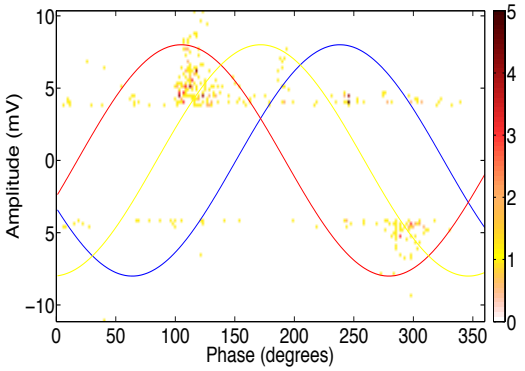



Figure 5.14: PRPD data plot of the on-line field data at cable loop of the EAC network for class 1 identified by the DBSCAN algorithm with $\epsilon$ set to 1.5 and $k_{\min}$ to 4.

### 5.5.3 Field data from Santa Rosa substation

The third loop analysed also consists of mainly PILC cable with just a small section of XLPE. The same procedure is followed as with the previous cases and the outcome of the DBSCAN algorithm using the parameters $\epsilon = 1.5$ and $k_{\min} = 4$, is shown in figure 5.15, where it can be seen that the data is clustered into a two classes plus the outliers. The waveforms of the identified classes are shown in figure 5.16 where class 1 is shown in

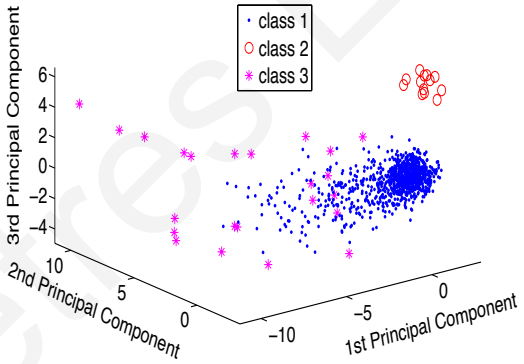



Figure 5.15: Plot of the first three principal components of the feature vector of the processed by the DBSCAN algorithm with parameters $\epsilon$ set to 1.5 and $k_{\min}$ to 4. The points marked with $\cdot$ are the ones identified in class 1 while class 2 is identified by $\circ$. Outliers have been marked with an $*$ .

figure 5.16a and 5.16b. The PRPD plots for the two classes shown in figures 5.17a and 5.17b respectively verify the fact that the activity in this specific loop is not related to PD activity.

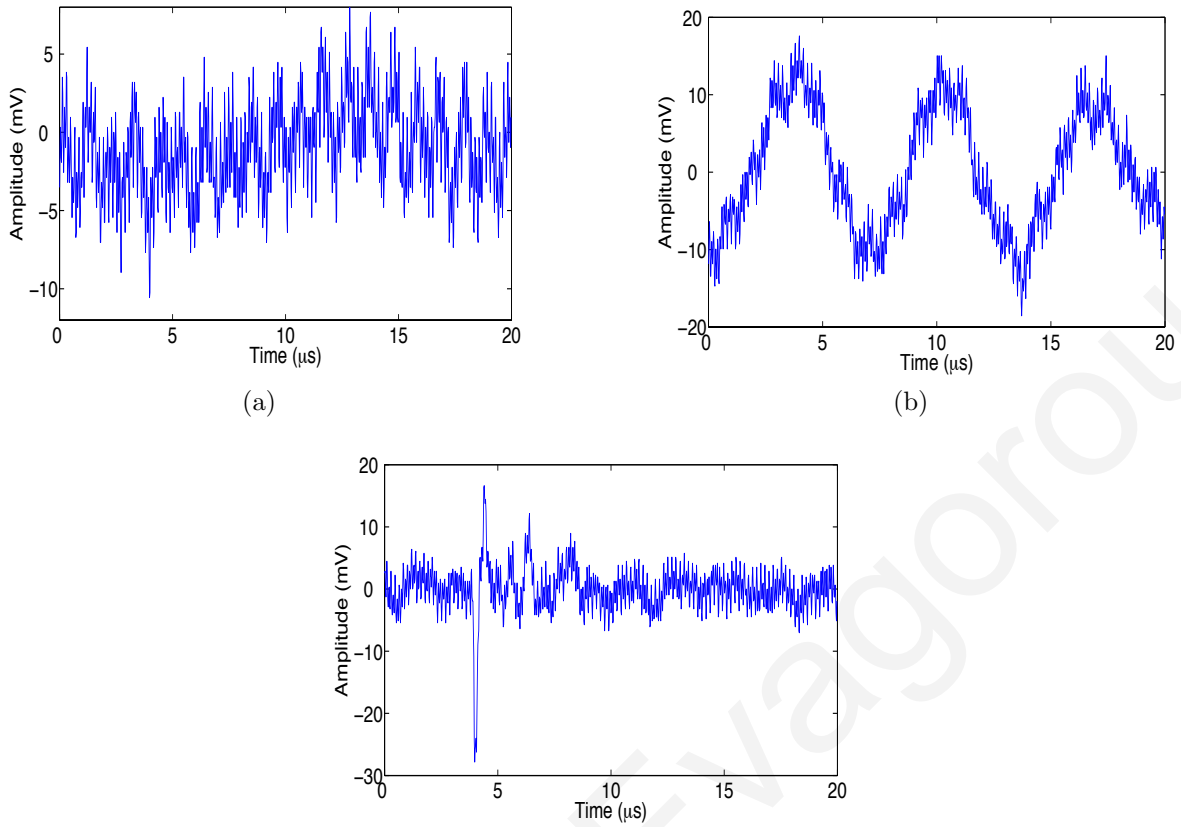



Figure 5.16: Plots of the waveforms of data clustered into class 1 (a), class 2 (b) and outliers (c).

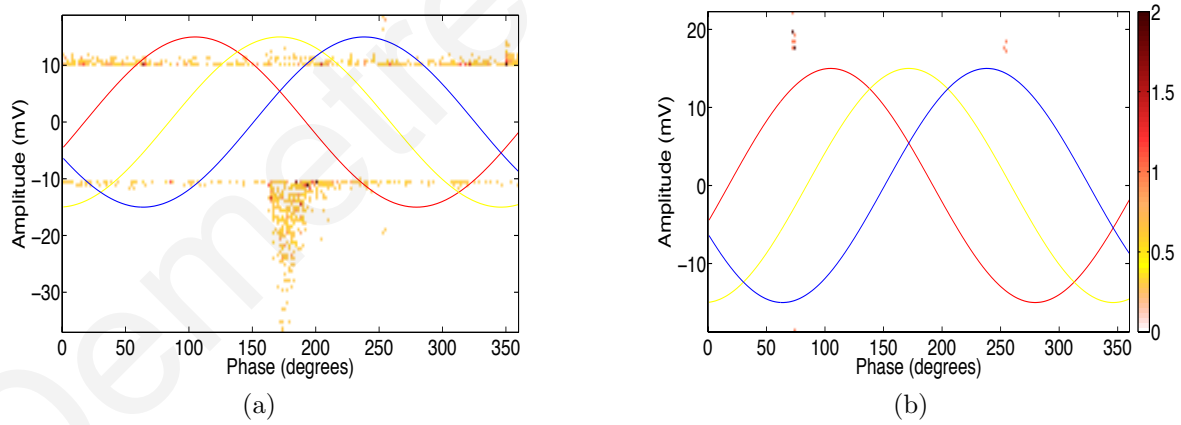



Figure 5.17: PRPD data plot of the on-line field data at cable loop of the EAC network for (a) class 1 and (b) class 2 identified by the DBSCAN algorithm with $\epsilon$ set to 1.5 and $k_{\min}$ to 4.

## 5.6 Conclusions

In this chapter the application of the proposed time-resolved feature vector on on-line data was evaluated. In such a scenario the lack of class labels for the training data, as well as the lack of any a priori knowledge about the number and type of PD sources, pose a difficulty to the use of supervised algorithms. Fortunately the use of clustering or unsupervised algorithms provides the class labels or even alleviates the use of training data. The DBSCAN was evaluated on experimental data from PILC cables with artificial defects where the effect on PD activity of the application of mechanical crushing on a spike defect was identified and the relevant PRPD extracted. Next on-line measurements from the Cyprus distribution network at locations suspected of increased PD activity with mixed PILC and XLPE cables were utilised. The analysis identified a potential PD sources, clustering pulses belonging to this source close together and rejected interference pulses by not including them in a specific cluster. It can be concluded that the proposed feature vector has the potential to be used in on-line applications, either through a combination of supervised or clustering algorithms or through the sole employment of a clustering algorithm.

# Chapter 6

# Conclusions

Classification into sources of different origin is essential in evaluating the severity of the damage caused by PD activity on the insulation of power cables and their accessories. The correct classification of different PD types is of vital importance in a condition monitoring program. The work in this thesis was an attempt to approach the PD classification in cable networks from a continuous on-line monitoring perspective. Such an approach imposes additional constraints on the implementation of the algorithms such as the need of an algorithm that will create manageable storage requirements and computational effort. Both phase and time resolved features were employed in order to give a more complete representation of the data.

In order to study the characteristics of different PD sources an experiment was designed under laboratory controlled conditions. The experiment provided phase and time-resolved data for four types of discharges, believed to more frequently occur in power cable networks. The corona, floating, internal and surface discharge data was acquired using an HFCT sensor, simulating on-line acquisition where this sensor is clamped around the earth wire. The data formed the database on which the development of the proposed features and their evaluation was based.

Both the phase and time-resolved features employed the WPT to expand the signal into wavelet coefficients that represent different time-frequency characteristics of the signal. In phase resolved data the probability distribution (pdf) of the wavelet coefficients at each scale provided a description of the main features of each PD type. In order to reduce the dimension of the pdf statistical descriptors such as the mean, variance, skewness and kurtosis were used producing a feature vector of 36 dimensions. Using the experimental database the proposed feature vector was validated against a feature extraction method of much higher dimensionality (1468) using a supervised algorithm, the SVM. The success rate was similar for both features but with a reduction factor of the order of 40, which

has a significant impact on the computational burden. The SVM was compared to the PNN using the proposed feature. Since their training is based on different principles, the results proved to be acceptable irrespective of the training method used.

Using time-resolved data four nodes of the WPT tree decomposition were selected and the coefficients at each of the selected nodes found to describe different PD sources. Statistical descriptors, provided by HOS, on the selected coefficients of each node resulted in a significant dimensionality reduction (1304 to 16). Comparison of this proposed feature with the use of the raw wavelet coefficients using the SVM showed a slight drop in performance from around 98.47% to 98.06%, which was considered acceptable given the gain in the computational load resulting from the dimensionality reduction.

The direct comparison of the performance of the PNN against the SVM in classifying PD data using the time-resolved proposed feature was performed, which became possible through the reduction in dimensionality as it alleviated numerical problems in the PNN associated with the number of dimensions. The two algorithms employ different training principles and the results showed that both methods provided consistent and reproducible results with very similar identification rates.

One of the problems facing on-line PD source separation is the presence of PD activity from multiple sources. This poses a problem to the use of supervised algorithms since no knowledge about the class labels or the number of PD sources in the training data is available. The use of unsupervised or clustering algorithms can provide this information or even alleviate the use of training data. The proposed time-resolved feature was evaluated on experimental data acquired from PILC cable loops, simulating operating conditions as close as possible. The DBSCAN was used and identification of an artificially created PD source was achieved. Field data from the EAC network from locations suspected of showing PD activity was processed by the proposed feature and the DBSCAN algorithm. The analysis identified a potential PD source and also rejected data that resembled interference pulses. This evaluation performed showed that the time-resolved proposed feature vector has the potential to be used in an on-line application.

The end goal of this work was to achieve an acceptable identification rate for PD of different sources for a continuous on-line application. Such an application needs feature vectors of lower dimensional space resulting in lower computational effort and storage burden. This requirement was incorporated into the design of the feature extractor through the use of HOS. The lower dimensionality allows the capture of data at more regular intervals which is a potential step closer to continuous on-line PD monitoring

that allows trending of data, giving an insight into the behaviour of PD activity from inception to failure.

# Chapter 7

# Future Work

This work proposed two feature extraction methods, one using time and another using phase-resolved data, that achieved a characterisation of different PD sources using a compact representation. This lower dimensionality can be exploited in an on-line environment where the lower storage requirements of the proposed feature vectors allow the capture of data at more regular intervals. Therefore suspected locations of the EAC distribution network can be monitored on an almost real-time basis thus drawing conclusions about the evolution of the PD with time.

Supervised and unsupervised algorithms were used in evaluating the proposed feature vectors. The supervised algorithms were first employed on experimental data where laboratory controlled set-ups were used to build a database of different PD sources. This database can be expanded to include more PD sources or further subdivide the existing PD sources. For example the internal discharge simulated was a void discharge located midway between the electrodes. This can further subdivided into three different discharges one when the void is close to the high voltage electrode, one the void in the middle and a third one with the void close to the earthed electrode. Similarly the rest of the PD sources can be subdivided into more specific ones and the proposed algorithms can be evaluated if they can achieve good separation of these sources.

In this thesis a database was acquired and separated into a training and a test dataset to train and evaluate the algorithms. Another improvement of the proposed feature vectors is their evaluation on a different training and testing database. The training and test databases could perhaps be acquired using slightly different experimental set ups. As an example the training database for the surface discharge can be acquired using a needle shaped high voltage electrode and the test using a disc shaped one. By this procedure the robustness of the proposed algorithms will be evaluated as well as the results will

be more indicative of real scenarios where the training and test data sets are completely independent and not part of the same dataset.

In identifying PD sources using the field data in an unsupervised mode the DBSCAN algorithm has been employed. One of the drawbacks of this algorithm is that the parameters of the algorithm are determined from the thinnest cluster in the dataset. Although this results in clusters of arbitrary shape to be identified when the underlying data has clusters with different density then the algorithm can not identify the true structure. As an alternative the OPTICS algorithm can be employed which creates an augmented ordering of the database representing its density based clustering structure. The use of this algorithm can be considered in field data and compared to the results presented here to evaluate any increase in identification.

The proposed feature vectors are to be employed in a continuous on-line condition monitoring system where trending will be employed to evaluate the insulation integrity of the equipment under test. However the interpretation of the trending has not received much attention and is a field that can give vital contributions to the PD community once exploited.

In general the PD community is very close to successful implementation of PD diagnostics. Several research groups have proposed different algorithms for identifying PDs in field conditions, including the work done in this thesis. From the results it seems that we are in a position to identify PD activity with a high probability of being correct. However the ultimate target of the PD community should be to move towards prognostics, which not only identifies the presence of PD activity but also can provide a time frame towards failure. This can be achieved through a series of new research activities such as more modelling of the PD process in order to better understand the underlying process per specific case. Moreover monitoring only PD activity can not provide us enough information about the insulation integrity as this activity depends on a number of additional factors . In order to arrive at more informative conclusions about the PD activity at a particular location it has to be coupled to the ambient and soil temperature, the soil thermal conductivity, the temperature of the core conductor, the load as well as meteorological conditions such as the speed of the wind. Existing models should be modified to incorporate all this information in order to enforce PD prognostics.

# Bibliography


[1] Heitz C. A generalized model for partial discharge processes based on a stochastic process approach. *Journal of Physics D: Applied Physics*, 32(9):1012, 1999. xii, 3, 15, 16, 39

[2] Egon Peschke and Rainer von Olshausen. *Cable Systems for High and Extra-High Voltage: Development, Manufacture, Testing, Installation and Operation of Cables and their Accessories*. Publicis MCD Verlag, 1999. xii, xiii, xvii, 18, 19, 20, 25

[3] W.S. Zaengl E. Kuffel and J. Kuffel. *High Voltage Engineering: Fundamentals*. Newnes, 2 edition, 2006. ISBN 0750636343. xiii, 27, 28

[4] P.S. Addison. *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance*. Institute of Physics Publishing, 2002. xiii, 43

[5] R. L. Streit and T. E. Luginbuhl. Maximum likelihood training of probabilistic neural networks. *IEEE Transactions on Neural Networks*, 5:764–783, September 1994. xiv, 68, 78

[6] E. Gockenbach and W. Hauschild. The selection of the frequency range for high-voltage on-site testing of extruded insulation cable systems. *IEEE Electrical Insulation Magazine*, 16(6):11 –16, nov.-dec. 2000. ISSN 0883-7554. doi: 10.1109/57.887600. 2

[7] E. Gulski, J.J. Smit, P.N. Seitz, J.C. Smit, and M. Turner. On-site pd diagnostics of power cables using oscillating wave test system. In *Eleventh International Symposium on High Voltage Engineering, 1999*, volume 5, pages 112 –115, 1999. doi: 10.1049/cp:19990898. 2

[8] A. Cavallini and G.C. Montanari. Effect of supply voltage frequency on testing of insulation system. *IEEE Transactions on Dielectrics and Electrical Insulation*, 13 (1):111 – 121, February 2006. ISSN 1070-9878. doi: 10.1109/TDEI.2006.1593409. 3

[9] C. Forssen and H. Edin. Partial discharges in a cavity at variable applied frequency part 1: measurements. *IEEE Transactions on Dielectrics and Electrical Insulation*, 15(6):1601 –1609, december 2008. ISSN 1070-9878. doi: 10.1109/TDEI.2008.4712663.

[10] C. Forssen and H. Edin. Partial discharges in a cavity at variable applied frequency part 2: measurements and modeling. *IEEE Transactions on Dielectrics and Electrical Insulation,*, 15(6):1610 –1616, december 2008. ISSN 1070-9878. doi: 10.1109/TDEI.2008.4712664. 3

[11] J.M. Meek and J.D. Craggs. *Electrical breakdown of gases*. Wiley series in plasma physics. Wiley, 1978. ISBN 9780471995531. 3, 15

[12] E. Nasser. *Fundamentals of gaseous ionization and plasma electronics*. Wiley series in plasma physics. Wiley-Interscience, 1971. ISBN 9780471630562. 3, 15

[13] L. Niemeyer. A generalized approach to partial discharge modeling. *IEEE Transactions on Dielectrics and Electrical Insulation*, 2(4):510 –528, August 1995. ISSN 1070-9878. doi: 10.1109/94.407017. 3, 15

[14] R.J. Van Brunt. Stochastic properties of partial-discharge phenomena. *IEEE Transactions on Electrical Insulation*, 26(5):902–948, 1991. ISSN 00189367. doi: 10.1109/14.99099. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=99099. 3

[15] R.J. Van Brunt, E.W. Cernyar, and P. von Glahn. Importance of unraveling memory propagation effects in interpreting data on partial discharge statistics. *IEEE Transactions on Electrical Insulation*, 28(6):905–916, 1993. ISSN 00189367. doi: 10.1109/14.249364. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=249364. 3

[16] M. Hoof and R. Patsch. Pulse-sequence analysis: a new method for investigating the physics of pd-induced ageing. *IEE Proceedings - Science, Measurement and Technology*, 142(1):95–101, 1995. doi: 10.1049/ip-smt:19951563. 3

[17] Rainer Patsch and Farhad Berton. Pulse sequence analysis - a diagnostic tool based on the physics behind partial discharges. *Journal of Physics D: Applied Physics*, 35(1):25, 2002. URL http://stacks.iop.org/0022-3727/35/i=1/a=306. 3

[18] R. Bartnikas. A comment concerning the rise times of partial discharge pulses. *IEEE Transactions on Dielectrics and Electrical Insulation*, 12(2):196–202, April 2005. ISSN 1070-9878. doi: 10.1109/TDEI.2005.1430390. 3, 38

[19] R. Bartnikas and J.P. Novak. On the character of different forms of partial discharge and their related terminologies. *IEEE Transactions on Electrical Insulation*, 28(6): 956–968, 1993. ISSN 00189367. doi: 10.1109/14.249369. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=249369. 4

[20] R. Bartnikas. Partial discharges. their mechanism, detection and measurement. *IEEE Transactions on Dielectrics and Electrical Insulation*, 9(5):763– 808, October 2002. Digital Object Identifier 10.1109/TDEI.2002.1038663. 4, 38

[21] H.-G. Kranz. Diagnosis of partial discharge signals using neural networks and minimum distance classification. *IEEE Transactions on Dielectrics and Electrical Insulation*, 28(6):1016 – 1024, December 1993. Digital Object Identifier 10.1109/14.249375. 4, 5, 8, 38, 39

[22] E. Gulski and A. Krivda. Neural networks as a tool for recognition of partial discharges. *IEEE Transactions on Electrical Insulation*, 28(6):984 – 1001, December 1993. Digital Object Identifier 10.1109/14.249372. 4, 5, 39

[23] T. Hucker and H.-G. Kranz. Requirements of automated pd diagnosis systems for fault identification in noisy conditions. *IEEE Transactions on Dielectrics and Electrical Insulation*, 2(4):544–556, August 1995. Digital Object Identifier 10.1109/94.407020.

[24] R.E. James and B.T. Phung. Development of computer-based measurements and their application to pd pattern analysis. *IEEE Transactions on Dielectrics and Electrical Insulation*, 2(5):838 – 856, October 1995. Digital Object Identifier 10.1109/94.469978. 4, 8, 39

[25] M.K. Abdul Rahman, R. Arora, and S.C. Srivastava. Partial discharge classification using principal component transformation. *IEE Proceedings - Science, Measurement*

*and Technology*, 147(1):7, 2000. ISSN 13502344. doi: 10.1049/ip-smt:20000074. URL http://link.aip.org/link/ISMTEV/v147/i1/p7/s1&Agg=doi.

[26] E.M. Lalitha and L. Satish. Wavelet analysis for classification of multi-source pd patterns. *IEEE Transactions on Dielectrics and Electrical Insulation*, 7(1):40–47, 2000. 5, 8, 39

[27] E. Gulski. Discharge pattern recognition in high voltage equipment. *IEE Proceedings Science, Measurement and Technology-*, 142(1):51 – 61, January 1995. 4, 8, 38, 39

[28] F.H. Kreuger, E. Gulski, and A. Krivda. Classification of partial discharges. *IEEE Transactions on Electrical Insulation*, 28(6):917 – 931, December 1993. Digital Object Identifier 10.1109/14.249365. 4, 39

[29] R. Candela, G. Mirelli, and R. Schifani. Pd recognition by means of statistical and fractal parameters and a neural network. *IEEE Transactions on Dielectrics and Electrical Insulation*, 7(1):87 – 94, February 2000. Digital Object Identifier 10.1109/94.839345. 5, 39

[30] B.X. Du, Wei Guozhang Wei, Wu Yuan, and M. Ouyang. Wavelet-based and fractal theory on partial discharge classification. In *Proceedings of 2005 International Symposium on Electrical Insulating Materials*, volume 2, pages 463 – 466, June 2005. 5, 39

[31] *IEC 60270 Partial Discharge Measurements*. International Electrotechnical Commission, 2000. 5, 14, 23

[32] Alexander Kraetge, Stefan Hoek, and Kay Rethmeier. Advanced possibilities of synchronous conventional and uhf pd measurements for effective noise suppression. In *Conference Record of the 2010 IEEE International Symposium on Electrical Insulation (ISEI)*, pages 1–4, 2010. 5

[33] A.A. Mazroua, M.M.A. Salama, and R. Bartnikas. Discrimination between pd pulse shapes using different neural network paradigms. *IEEE Transactions on Dielectrics and Electrical Insulation*, 1(6):1119 – 1131, December 1994. Digital Object Identifier 10.1109/94.368651. 5, 65

[34] M.M.A. Salama and Bartnikas R. Fuzzy logic applied to pd pattern classification. *IEEE Transactions on Dielectrics and Electrical Insulation*, 7(1):118–123, February 2000. doi: 10.1109/94.839349. 6

[35] S. Meijer and J.J. Smit. Uhf defect evaluation in gas insulated equipment. *IEEE Transactions on Dielectrics and Electrical Insulation*, 12(2):285 – 296, April 2005. ISSN 1070-9878. doi: 10.1109/TDEI.2005.1430398. 6, 40

[36] A. Cavallini, G.C. Montanari, A. Contin, and F. Pulletti. A new approach to the diagnosis of solid insulation systems based on pd signal inference. *IEEE Electrical Insulation Magazine*, 19(2):23–30, March-April 2003. ISSN 0883-7554. doi: 10.1109/MEI.2003.1192033. 6, 41

[37] A. Cavallini, G.C. Montanari, F. Puletti, and A. Contin. A new methodology for the identification of pd in electrical apparatus: properties and applications. *IEEE Transactions on Dielectrics and Electrical Insulation*, 12(2):203–215, April 2005. doi: 10.1109/TDEI.2005.1430391. 40

[38] A. Contin, A. Cavallini, G.C. Montanari, G. Pasini, and F. Puletti. Digital detection and fuzzy classification of partial discharge signals. *IEEE Transactions on Dielectrics and Electrical Insulation*, 9(3):335–348, June 2002. ISSN 1070-9878. doi: 10.1109/TDEI.2002.1007695. 6, 41

[39] L. Hao and P.L. Lewin. Partial discharge source discrimination using a support vector machine. *IEEE Transactions on Dielectrics and Electrical Insulation*, 17(1): 189 –197, February 2010. ISSN 1070-9878. doi: 10.1109/TDEI.2010.5412017. 6, 11, 41, 51, 54, 64, 74, 75, 78, 83

[40] E Carminati, L Cristaldi, M Lazzaroni, and A Monti. A neuro-fuzzy approach for the detection of partial discharge. *IEEE Transactions on Instrumentation and Measurement*, 50(5):1413 – 1417, October 2001. 6

[41] J. Jin, C.S. Chang, C. Chang, T. Hoshino, M. Hanai, and N. Kobayashi. Classification of partial discharge events in gas insulated substations using wavelet packet transform and neural network approaches. *IEE Proceedings Science, Measurement and Technology*, pages 55 – 63, March 2006. 6, 41

[42] A. Contin, G.C. Montanari, and C. Ferraro. Pd source recognition by weibull processing of pulse height distributions. *IEEE Transactions on Dielectrics and Electrical Insulation*, 7(1):48 –58, February 2000. ISSN 1070-9878. doi: 10.1109/94.839341. 7, 41, 51

[43] A. Contin and S. Pastore. Classification and separation of partial discharge signals by means of their auto-correlation function evaluation. *IEEE Transactions on Dielectrics and Electrical Insulation*, 16(6):1609 –1622, December 2009. ISSN 1070-9878. doi: 10.1109/TDEI.2009.5361581. 7, 41, 91

[44] RM. Sharkawy, RS. Mangoubi, TK. Abdel-Galil, MMA. Salama, and R. Bartnikas. Svm classification of contaminating particles in liquid dielectrics using higher order statistics of electrical and acoustic pd measurements. *IEEE Transactions on Dielectrics and Electrical Insulation*, 14(3):669–678, June 2007. 7, 41

[45] D. Evagorou, A. Kyprianou, P.L. Lewin, A. Stavrou, V. Efthymiou, A.C. Metaxas, and G.E. Georghiou. Feature extraction of partial discharge signals using the wavelet packet transform and classification with a probabilistic neural network. *IET Science, Measurement & Technology*, 4(3):177–192, May 2010. 9, 61, 78

[46] R. Patsch and M. Hoof. Physical modeling of partial discharge patterns. In *Proceedings of the 1998 IEEE 6th International Conference on Conduction and Breakdown in Solid Dielectrics*, pages 114 –118, June 1998. 15

[47] F.H. Kreuger. *Partial discharge detection in high-voltage equipment*. Butterworths, 1989. ISBN 9780408020633. URL http://books.google.com/books?id=GOdSAAAAMAAJ. 17

[48] Y. Tian, P.L. Lewin, A.E. Davies, and G. Hathaway. Acoustic emission techniques for partial discharge detection within cable insulation. In *Eighth International Conference on Dielectric Materials, Measurements and Applications*, pages 503–508, September 2000. Digital Object Identifier 00888169. 17

[49] E. Grossmann and K. Feser. Sensitive online pd-measurements of onsite oil/paper-insulated devices by means of optimized acoustic emission techniques (aet). *IEEE Transactions on Power Delivery*, 20(1):158 – 162, January 2005. ISSN 0885-8977. doi: 10.1109/TPWRD.2004.835288. 17

[50] H. Ota, M. Ichihara, N. Miyamoto, S. Kitai, Y. Maruyama, M. Fukasawa, and H. Takehana. Application of advanced after-laying test to long-distance 275 kv xlpe cable lines. *IEEE Transactions on Power Delivery*, 10(2):567–579, April 1995. Digital Object Identifier 10.1109/61.400876. 17

[51] Y. Tian, P.L. Lewin, A.E. Davies, S.J. Sutton, and S.G. Swingler. Partial discharge detection in cables using vhf capacitive couplers. *IEEE Transactions on Dielectrics and Electrical Insulation*, 10(2):343–353, April 2003. Digital Object Identifier 10.1109/TDEI.2003.1194121. 17

[52] J.Y. Zhou and S. Boggs. High frequency properties of shielded power cable - part 1: overview of mechanisms. *IEEE Electrical Insulation Magazine*, 21(6):24–28, November 2005. ISSN 0883-7554. doi: 10.1109/MEI.2005.1541486. 18, 20

[53] S. Boggs, A. Pathak, and P. Walker. Partial discharge. xxii. high frequency attenuation in shielded solid dielectric power cable and implications thereof for pd location. *IEEE Electrical Insulation Magazine*, 12(1):9–16, January-February 1996. Digital Object Identifier 10.1109/57.484104. 18

[54] S. Boggs. High frequency attenuation in transmission class solid dielectric cable. *IEEE Transactions on Power Delivery*, 23(4):1713–1719, October 2008. ISSN 0885-8977. 20

[55] E. Laboure, F. Costa, and F. Forest. Current measurement in static converters and realisation of a high frequency passive current probe (50 a-300 mhz. In *Fifth European Conference on Power Electronics and Applications*, volume 4, pages 478–483, September 1993. 21

[56] P. Poulichet, F. Costa, and E. Labour. High-frequency modeling of a current transformer by finite-element simulation. *IEEE Transactions on Magnetics*, 39(2):998–1007, March 2003. 21

[57] Valentinas Dubickas and Hans Edin. High frequency model of rogowski coil with small number of turns. *IEEE Transactions on Instrumentation and Measurements*, 56(6):2284–2288, December 2007. 22

[58] E. Gulski. Computer-aided measurement of partial discharges in hv equipment. *IEEE Transactions on Electrical Insulation*, 28(6):969–983, December 1993. 32

[59] Richard O. Duda, Peter E. Hart, and David G. Stork. *Partern Classification*. John Wiley & Sons, Inc, second edition, 2000. 37, 86, 91

[60] N.C. Sahoo, M.M.A. Salama, and R. Bartnikas. Trends in partial discharge pattern classification: a survey. *IEEE Transactions on Dielectrics and Electrical Insulation*, 12(2):248 – 264, April 2005. doi: 10.1109/TDEI.2005.1430395. 38, 64

[61] M. Hoof, B. Freisleben, and R. Patsch. Pd source identification with novel discharge parameters using counterpropagation neural networks. *IEEE Transactions on Dielectrics and Electrical Insulation*, 4(1):17 –32, February 1997. ISSN 1070-9878. doi: 10.1109/94.590861. 39

[62] X. Zhou, C. Zhou, and B.G. Stewart. Comparisons of discrete wavelet transform, wavelet packet transform and stationary wavelet transform in denoising pd measurement data. In *Conference Record of the IEEE International Symposium on Electrical Insulation*, pages 237 –240, 11-14 2006. 41, 51, 82

[63] A. Kyprianou, P.L. Lewin, V. Efthimiou, A. Stavrou, and G.E. Georghiou. Wavelet packet denoising for online partial discharge detection in cables and its application to experimental field results. *Measurement Science and Technology*, 17(9):2367–2379, July 2006. doi:10.1088/0957-0233/17/9/001. 82

[64] Hao Zhang, T Blackburn, B Phung, and D Sen. A novel wavelet transform technique for on-line partial discharge measurements. 1. wt de-noising algorithm. *IEEE Transactions on Dielectrics and Electrical Insulation*, 14(1):3 – 14, February 2007. 41

[65] Stéphane Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 525 B Street, Suite 1900, San Diego, California 92101/4495, USA, third edition, 2009. ISBN 9780123743701. 42, 45, 46, 49, 50

[66] Gerald Kaiser. *A friendly guide to wavelets*. Birkhäuser, 1994. 45

[67] S. G. Mallat. An efficient image representation for multiscale analysis. In *Optical Society of America, Topical Meeting on Machine Vision p 172-175 (SEE N89-19145 11-74)*, pages 172–175, 1980. 46

[68] S.G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11 (7):674–693, July 1989. Digital Object Identifier 10.1109/34.192463. 46

[69] R. R. Coifman. Wavelet analysis and signal processing. In Louis Auslander, Tom Kailath, and Sanjoy K. Mitter, editors, *Signal Processing, Part I: Signal Processing Theory*, pages 59–68. Springer-Verlag, New York, NY, 1990. 47

[70] R.R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713 – 718, March 1992. Digital Object Identifier 10.1109/18.119732. 49

[71] Abdourrahmane M. Atto, Dominique Pastor, and Alexandru Isar. On the statistical decorrelation of the wavelet packet coefficients of a band-limited wide-sense stationary random process. *Elsevier Signal Processing*, 87(10):2320–2335, 2007. 50

[72] D. Evagorou, A. Kyprianou, P.L. Lewin, A. Stavrou, V. Efthymiou, and G.E. Georghiou. An investigation into the parameters affecting the success of the wavelet packets method for denoising partial discharge signals. In *5th Mediterranean Conference on Power Generation, Transmission and Distribution*, November 2006. 51

[73] L Hao, P L Lewin, and S G Swingler. Improving detection sensitivity for partial discharge monitoring of high voltage equipment. *Measurement Science and Technology*, 19(5):055707, 2008. URL http://stacks.iop.org/0957-0233/19/i=5/a=055707. 51, 54

[74] D. Evagorou, A. Kyprianou, P.L. Lewin, A. Stavrou, V. Efthymiou, and G.E. Georghiou. Effect of the choice of the decomposition level on the success of the wavelet packets method for denoising partial discharge signals. In *6th Mediterranean Conference on Power Generation, Transmission and Distribution*, November 2008. 51

[75] M. Cacciari, A. Contin, and G.C. G.C Montanari. Use of a mixed-weibull distribution for the identification of pd phenomena [corrected version]. *IEEE Transactions on Dielectrics and Electrical Insulation*, 2(6):1166 – 1179, December 1995. Digital Object Identifier 10.1109/94.484322. 51

[76] A. Stuart and K. Ord. *Kendall's Advanced Theory of Statistics: Distribution Theory (Volume 1.* Wiley & Sons, 6 edition, 1994. 52, 53

[77] J.M. Mendel. Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. *Proceedings of the IEEE*, 79(3):278 –305, March 1991. ISSN 0018-9219. doi: 10.1109/5.75086. 52

[78] L. Hao, P. L. Lewin, and S. J. Dodd. Comparison of support vector machine based partial discharge identification parameters. In *IEEE International Symposium on Electrical Insulation*, pages 110–113, Toronto, Canada, June 2006. 54, 65, 74, 78

[79] J.A. Hunter, P.L. Lewin, D. Evagorou, A. Kyprianou, and G.E. Georghiou. Comparison of two partial discharge classification methods. In *IEEE International Symposium on Electrical Insulation*, June 2010. 61, 64, 74, 75, 76, 83

[80] J. T. Tou and R. C. Gonzalez. *Pattern Recognition Principles*. Addison-Wesley, 1974. 64

[81] A.A. Mazroua, M.M.A. Salama, and R. Bartnikas. Pd pattern recognition with neural networks using the multilayer perceptron technique. *IEEE Transactions on Dielectrics and Electrical Insulation*, 28(6):1082 – 1089, December 2003. Digital Object Identifier 10.1109/14.249382. 65

[82] M.M.A. Salama and R. Bartnikas. Determination of neural-network topology for partial dischargepulse pattern recognition. *IEEE Transactions on Neural Networks*, 13(2):446–456, March 2002. doi: 10.1109/72.991430.

[83] A.A. Mazroua, M.M.A. Salama, and R. Bartnikas. Neural network system using the multi-layer perceptron technique for the recognition of pd pulse shapes due to cavities and electrical trees. *IEEE Transactions on Power Delivery*, 10(1):92 – 96, January 1995. Digital Object Identifier 10.1109/61.368411. 65

[84] S. Haykin. *Neural Networks, a Comprehensive Foundation*. Prentice-Hall, 2 edition, 1999. 65

[85] Vojislav Kecman. *Learning and Soft Computing: Support Vector Machines, Neural Networks and Fuzzy Logic Models*. MIT Press, 2001. 65, 70, 73, 77

[86] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995. 65, 66, 67, 70

[87] B. Scholkopf and A.J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, 2002. 65

[88] S. J. Press. *Applied Multivariate Analysis*, chapter 13. Kreiger Publishing, 1982. 68

[89] D.F. Specht. Generation of polynomial discriminant functions for pattern recognition. *IEEE Transactions on Electronic Computers*, EC-16(3):308–319, July 1967. 68

[90] E. Parzen. On the estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962. URL `citeseer.ist.psu.edu/parzen62estimation.html`. 68

[91] T. Cacoullos. Estimation of a multivariate density. *Annals of Mathematical Statistics*, 18(2):179–189, 1966. 68

[92] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Annals of Mathematical Statistics*, 39:1–38, 1977. 69

[93] John Shawe-Taylor and Nello Cristianini. *Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000. 72, 77

[94] L. Hao, P. L. Lewin, Y. Tian, and S. J. Dodd. Partial discharge identification using a support vector machine. In *IEEE Conference on Electrical Insulation and Dielectric Phenomena*, pages 414 – 417, March 2005. 74, 78

[95] L. Hao, P. L. Lewin, and S. G. Swingler. Identification of multiple partial discharge sources. In *International Conference on Condition Monitoring and Diagnosis*, pages 118 – 121, April 2008. 74

[96] Demetres Evagorou, Andreas Kyprianou, Paul L. Lewin, Andreas Stavrou, Venizelos Efthymiou, and George E. Georghiou. Classification of partial discharge signals using probabilistic neural network. In *IEEE International Conference on Solid Dielectrics, ICSD*, pages 609–615, July 2007. 78

[97] Abraham Wald. An extension of wilks' method for setting tolerance limits. *The Annals of Mathematical Statistics*, 14(1):45–55, March 1943. 79

[98] J.W. Tukey. Nonparametric estimation, ii. statistically equivalent blocks and multivariate tolerance regions–the continuous case. *The Annals of Mathematical Statistics*, 18(4):529–539, December 1947. 79

[99] J.W.Tukey. Nonparametric estimation, iii. statistically equivalent blocks and multivariate tolerance regions–the discontinuous case. *The Annals of Mathematical Statistics*, 19(1):30–39, March 1948. 79

[100] H. Ackerwn and K. Abt. Designing the sample size for non-parametric, multivariate tolerance regions. *Biometrical Journal*, 26(7):723–734, 1984. 79, 87

[101] R. B. Murphy. Non-parametric tolerance limits. *Annals of Mathematical Statistics*, 19(4):581–589, December 1948. 79, 88

[102] Y. Tian, P.L. Lewin, P. Wang, S. J. Sutton, and S. G. Swingler. Application of wavelet-based denoising to online measurement of partial discharges. In *Proceedings of International Conference on Solid Dielectrics*, July 2004. 82

[103] P.D. Agoris, S. Meijer, E. Gulski, and J.J. Smit. Threshold selection for wavelet denoising of partial discharge data. In *Conference Record of the 2004 IEEE International Symposium on Electrical Insulation*, pages 62 – 65, September 2004. Digital Object Identifier 10.1109/ELINSL.2004.1380450.

[104] C.S. Chang, J. Jin, S. Kumar, Q. Su, T. Hoshino, M. Hanai, and N. Kobayashi. Denoising of partial discharge signals in wavelet packets domain. *IEE Proceedings Science, Measurement and Technology-*, 152(3):129 – 140, May 2005. Digital Object Identifier 10.1049/ip-smt:20041315. 82

[105] Martin Ester, Hans peter Kriegel, Jörg S, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *International Conference on Knowledge Discovery and Data Mining*, pages 226–231. AAAI Press, 1996. 90, 91, 94

[106] T.K. Abdel-Galil, Y.G. Hegazy, M.M.a. Salama, and R. Bartnikas. Fast Match-Based Vector Quantization Partial Discharge Pulse Pattern Recognition. *IEEE Transactions on Instrumentation and Measurement*, 54(1):3–9, February 2005. ISSN 0018-9456. doi: 10.1109/TIM.2004.839762. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1381791. 91

[107] Andrew Webb. *Statistical Pattern Recognition*. John Wiley & Sons, Inc, second edition, 2005. 91

[108] L. Hao, P.L. Lewin, J.A. Hunter, D.J. Swaffield, A. Contin, C. Walton, and M. Michel. Discrimination of multiple pd sources using wavelet decomposition and principal component analysis. *IEEE Transactions on Dielectrics and Electrical Insulation*, 18(5):1702 –1711, october 2011. ISSN 1070-9878. doi: 10.1109/TDEI.2011.6032842. 92

[109] J A Hunter, L Hao, D J Swaffield, P L Lewin, and M Michel. Partial discharge in medium voltage three-phase cables. In *IEEE International Symposium on Electrical Insulation*, pages 1–5, 2010. ISBN 9781424463015. 97