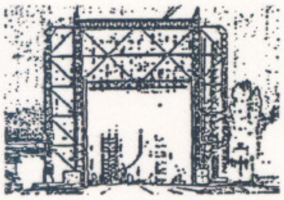

NINTH INTERNATIONAL SYMPOSIUM
ON HIGH VOLTAGE ENGINEERING

August 28 - September 1, 1995

GRAZ CONVENTION CENTER
AUSTRIA, EUROPE

# A Standardised Computer Data File Format for Storage, Transport, and Off-line Processing of Partial Discharge Data

T. Hücker*, P. von Glahn**, H.-G. Kranz*, T. Okamoto***

*Laboratory of High Voltage Engineering, Bergische University of Wuppertal, Germany
**National Institute of Standards & Technology, Gaithersburg, MD, USA
***Central Research Institute of Electric Power Industry, Japan

**Abstract**

We present an overview of a proposed data file format for digitised partial discharge data. The proposed format will permit investigators at different institutions to exchange PD data and collaborate on the analysis and understanding of the PD phenomenon. We include an example in which investigators at all three institutions evaluated the same data record and report their analysis results.

## 1. Introduction

In recent years, digital partial discharge (PD) recording systems have been developed that store the magnitude, polarity, and phase position of each PD event in computer files for subsequent analysis. As different investigators develop the capability to digitally record PD data, multiple PD data file formats are emerging. To facilitate co-operation between investigators, we propose a standardised data file format. Recorded PD data will be converted from the unique local file format to the standardised format using custom software. The standard format data files will be used to create databases of digitised PD measurements representing many different PD defects and measurement conditions. By making these databases accessible to PD investigators throughout the world by means of computer networks, individual investigators will have access to documented PD data from a much broader range of PD sources than they can locally reproduce. Alternatively, an investigator with PD recording capability and a unique experimental apparatus can establish a publicly accessible database of recorded PD data for others to use in their research. Thus the individual researcher will no longer have to develop an extensive private database of PD measurements but will be able to concentrate on developing improved PD diagnosis tools.

In the following sections, we discuss the characteristics of the proposed file standard. We then present a summary of the standard and conclude with an example of the type of insights into PD that are possible when different investigators examine the same PD data file.

## 2. Characteristics of the Proposed Standard

The proposed file standard was designed to be a long-term solution to the need to share PD data among members of the PD research community. As such, it possesses a number of key characteristics.

**Space efficient:** Because data exchange may take place via either physical media or electronic networks, the format has to be compact to minimise the size of the media or the transfer time.

**Extensible:** Digitisation of PD is now in its infancy, with current digitisers storing only the amplitude and the phase-of-occurrence of each PD pulse. Future systems may well add other parameters such as pulse width. Therefore, the data format must be easily extensible too accommodate new parameters.

**Contains test voltage data:** Different experiments use different test voltages and waveforms. Further, for some experiments, the voltage level changes with time. Therefore, the format must include provisions for representing both the time-varying amplitude and waveform of the test signal.

**Multiple time segment digitisation:** While some PD measurements take place over a short time, others such as ageing experiments span periods measured in hours or days. To keep the size of the data file to a reasonable level, PD is measured on an intermittent basis (e.g., 2 minutes of recording every 30 minutes). Each recording period is referred to as a *segment*. The data format must provide for multiple segments within one file.

**Self-documenting:** Many variables effect the characteristic of the experimental PD data. These must be documented within the data format so that other experimenters have access to this information.

## 3. Data File Format:

The digitised PD data are stored in a binary file according to the format described below. An additional textual information file accompanies the binary data file. This textual file documents the experimental conditions and significance of the data.

The binary data file is organised into a sequence of records of different types illustrated in Figure 1. It will begin with a FILE_HEADER record defining the structure of the file. The file is divided into one or more segments, each corresponding to a continuous recording period. Each segment begins with a CALIBRATIONS record containing information needed to convert the binary PD data into engineering units (seconds, degrees, charge, etc.). Following the CALIBRATIONS record are a sequence of PD_CYCLE records containing the actual PD data. Each PD_CYCLE record also contains the current rms value of the test voltage which is adequate to reconstruct either DC or sinusoidal AC test voltages. If another waveform is used for the test voltage, the file will end with a record containing samples of one cycle of the test voltage.

| |
|---|
| FILE_HEADER |
| CALIBRATIONS segment 1 |
| PD_CYCLE (first of segment 1)<br>...<br>PD_CYCLE (last of segment 1) |
| CALIBRATIONS segment 2 |
| PD_CYCLE (first of segment 2)<br>...<br>PD_CYCLE (last of segment 2) |
| ... |
| CALIBRATIONS last segment |
| PD_CYCLE (first of last segment )<br>...<br>PD_CYCLE (last of last segment ) |
| TEST_VOLTAGE_SAMPLE (optional) |

Figure 1: Data file format

### 3.1. Data Types

The data records are composed of a sequence of bytes that are interpreted according to the formats given in this section. The types are those used in the C programming language. Table 1 gives the storage requirements, and data ranges for the types used in this proposed standard. Data is stored in the file in the byte order used by the Intel x86 microprocessor. Negative numbers are represented with twos complement notation.

Table I. Data Type Definitions

| | |
|---|---|
| **ASCII Data Type:** | |
| Size: | Variable (field zero padded) |
| Value Range: | Standard ASCII Characters |
| **Byte Data Type:** | |
| Size: | 1 byte |
| Value Range: | 0 to 255 |
| **Unsigned Integer Data Type:** | |
| Size: | 2 byte |
| Value Range: | 0 to 65,535 |
| **Signed Integer Data Type:** | |
| Size: | 2 byte |
| Value Range: | -32,768 to +32,767 |
| **Long Unsigned Integer Data Type:** | |
| Size: | 4 byte |
| Value Range: | 0 to 4,294,967,295 |
| **Double Data Type:** | |
| Size: | 8 byte |
| Value Range: | ±1.7E308 with ≥ 15 digits precision |

### 3.2. Data Record Formats

The following subsections provide a summary description of the data file format. Each record format is described in terms of the data types defined in Table I. More details on the format are available from the main authors.

#### 3.2.1. FILE_HEADER Record Format

The FILE_HEADER record defines the format of the data file as shown in Table II. It begins with a minimal description of the experiment to aid in file identification and includes information on the location of the test voltage sample (if one is present).

In the case of intermittently-recorded PD data, many segments can be stored sequentially in the file. The first byte location and starting time of each segment is included in the FILE_HEADER record.

Many different quantities can be measured for a PD pulse depending on the capabilities of a given PD recording system (e.g., amplitude, pulse width, rise time, fall time, asymmetry, phase position, etc.). The name of the quantities represented in the PD_CYCLE record are included in the FILE_HEADER record together with information on how each is represented (e.g., as a signed or unsigned 16-bit integer). All data in the PD_CYCLE records are stored as integer quantities. A conversion equation must be applied to each measurement quantity to convert it to a useful quantity (i.e., charge in picocoloumbs). The FILE_HEADER defines the form of this conversion equation in the PD Record Field Description. Both polynomial and logarithmic calibration functions are supported with the needed calibration coefficients stored in the CALIBRATIONS record at the start of each segment.

Table II. FILE_HEADER Record Format Summary

| RECORD FIELD | | DATA TYPE | COMMENT |
|---|---|---|---|
| Record Identification | | ASCII (15) | Identification text: "PD-DATA Ver. 1" |
| PD-Description | | ASCII (50) | Short textual measurement description |
| Date | | ASCII (10) | Local date measurement started |
| Time | | ASCII (8) | Local time measurement started |
| Institution | | ASCII (15) | Name of recording institution |
| Operator | | ASCII (15) | Name of recording person |
| Frequency | | unsigned int | Test voltage frequency in Hertz |
| Pulse Resolution | | unsigned int | Pulse resolution time of apparatus |
| Test Voltage Sample | | long unsigned | Location of test voltage record in file. |
| Segment Number | | unsigned int | Number of segments in file |
| Segment Start Locations | | structure | Segment starting locations within file (one per segment) |
| | Starting Point | long unsigned | position of first byte of segment in file |
| | Cycle No. | long unsigned | AC cycles since start of the measurement |
| PD Record Field Number | | byte | Number of fields in PD Record structure |
| PD Record Field Description | | structure | Description of each field in PD Record structure (one per field) |
| | Quantity | ASCII (25) | Name of the quantity |
| | How stored | byte | Signed or unsigned integer |
| | Conversion type | byte | Form of conversion function (Table Ib) |

Table Ib: Form of Conversion Functions

| VALUE | CONVERSION TYPE | CONVERSION FUNCTION | No. OF COEFF. |
|---|---|---|---|
| 1 | Linear Without Offset | $q[pC] = f_0 \cdot Q_{int}$ | 1 |
| 2 | Linear | $q[pC] = f_1 \cdot Q_{int} + f_0$ | 2 |
| 3 | Quadratic | second order polynomial | 3 |
| 4 | Cubic | third order polynomial | 4 |
| ... | ... | ... | ... |
| 100 | Logarithmic Without Polarity | $q[pC] = f_1 \cdot e^{(f_2 \cdot Q_{int})} + f_0$ | 3 |
| 101 | Logarithmic With Polarity | $q[pC] = \frac{Q_{int}}{\lvert Q_{int} \rvert} \cdot f_1 \cdot e^{(f_2 \cdot \lvert Q_{int} \rvert)} + f_0$ | 3 |

#### 3.2.2. PD_CYCLE Record Format

Each PD pulse can be represented by its phase-of-occurrence and one or more additional parameters (e.g., amplitude). The minimal data record for describing the PD on a single AC cycle is stored in a PD_CYCLE record as defined in Table III. The record begins with a constant record identifier byte used by the file reading program for error control. Each record contains a byte indicating the number of elapsed cycles since the last PD_CYCLE record so that cycles with no PD pulses can be properly accounted for. Next in the record is a integer giving the current root mean square (RMS) value of the excitation voltage. The CALIBRATION record at the beginning of the segment contains the coefficient of the conversion equation relating this integer value to the actual RMS value.

Table III: PD_CYCLE Record Format

| RECORD MEMBER | | DATA TYPE | COMMENT |
|---|---|---|---|
| Record Identification | | byte | Identification value (always $55_{16}$) |
| Incremental AC Cycle Counter | | byte | Number of AC cycles since the last PD |
| Test Voltage | | unsigned int | RMS test voltage value of this cycle |
| PD Number | | unsigned int | Number of discharges in this cycle |
| PD Record | | structure | PD Data (one per discharge) |
| | PD Value | signed | A/D converter values for pulse charge |
| | Phase Position | unsigned int | Pulse phase [Note 1] |

Note 1: The phase position is recorded as a 16-bit unsigned number with $00000000_{16}$ = 0 deg and $ffffffff_{16}$ = $(360 - 1/2^{16})$deg.

Because PD_CYCLE records are only stored for AC cycles with partial discharges, the test voltage is unknown for time intervals without discharges (e.g. below PD inception voltage). Therefore after 50 cycles without PD a PD_CYCLE record with the PD number = 0 should be stored to both record the current excitation voltage and maintain the incremental cycle count. Finally, the record contains the actual PD data as defined in the FILE_HEADER record.

### 3.2.3. CALIBRATIONS Record Format

All PD measurements in the PD_CYCLE records are given as 16-bit integer quantities. The FILE_HEADER record defines the conversion equations needed to convert these integer values to useful, floating point values. The CALIBRATIONS record shown in Table IV is located at the beginning of each data segment. It contains the coefficients for these conversion equations. Since the conversion equation coefficients may change during the course of the segment, multiple conversion coefficient sets can be stored in the record, each with an accompanying effective time definition.

Table IV. CALIBRATIONS Record Format

| RECORD MEMBER | DATA TYPE | COMMENT |
|---|---|---|
| Record Identification | byte | Identification value (always $66_{16}$) |
| Time Interval Number | unsigned int | Number of measuring range changes |
| Calibration Record | structure | Calibration factors as defined in header structure (one per time interval) |
| Switching Time Point | long unsigned | AC cycle number in the segment were the measuring range begins |
| Conversion Coefficient | double | Lowest order coefficient of 1st quantity |
| ... | ... | ... |
| Conversion Coefficient | double | Highest order coefficient of last quantity |
| Voltage Coefficient | double | Test voltage conversion coefficient |

### 3.2.4. TEST_VOLTAGE_SAMPLE Record Format

For DC or sinusoidal AC excitation, the RMS value stored in each PD_CYCLE record are adequate to fully define the excitation voltage for each PD pulse. If some other excitation waveform is used, samples of this waveform can be stored in a TEST_VOLTAGE_SAMPLE record defined in Table V. This record will be the last record in the data file. The record begins with a constant record identifier used for playback error detection. It includes the conversion coefficients needed to interpret each voltage sample.

Table V: TEST_VOLTAGE_SAMPLE Record Format

| RECORD MEMBER | DATA TYPE | COMMENT |
|---|---|---|
| Record Identification | byte | Identification value (always $99_{16}$) |
| Calibration Factor | double | For converting samples to engineering units |
| Calibration Offset | double | |
| Test Voltage Samples | unsigned int | Number of test voltage samples |
| TV Record | structure | Samples of test voltage |
| Voltage Value | signed int | A/D converter values for test voltage |
| Phase | unsigned int | Phase of A/D value [Table III, Note 1] |

### 3.3. Documentation File Format

An additional text file will be created to document and accompany the data file. This file is an ASCII text file describing the experiment and the contents of binary file. It is normally created using a text editor. It is intended to be human readable so the formatting is not critical. As a minimum, it will contain the following information:

- Experiment condition documentation
- Organisation generating the data file
- Geographic location, date, and time of the experiment
- AC Excitation frequency and wave shape used
- A description of the significance of each segment in the file

## 4. Example of Co-operative PD Data Analysis

To illustrate the insights possible when multiple investigators analyse the same data file, we generated a data file at NIST using the NIST PD Digitizer /2/. The file was generated from a point-plane geometry in room air with a gap of 1 cm and sinusoidal excitation ($V_{rms}$ = 3.1 kV, $f$ = 200 Hz). The PD was sensed with a 300 Ω resistor between the plane and ground. Digitisation lasted for 297 seconds. The resulting 6.1 megabyte data file was analysed at NIST, the Laboratory of High Voltage Engineering in Wuppertal, and CRIEPI in Japan. The analysis results are presented below.

### 4.1. NIST Analysis Results

Stochastic analysis was performed on the PD data at NIST as discussed in /2/. The PD source was very stable over time, showing little change in the number of pulses or integrated charge over a time of 300 s (Figure 2). This indicates that the PD source was not ageing during this period.

Figure 2 shows an absence of positive pulses on many cycles whereas negative pulses were always present. This is characteristic of a sharp object on supply side of the PD site opposed to a smoother surface on the ground side (the more intense field around the sharp object more easily removes electrons to initiate a PD pulse).

The relationship between the positive and negative integrated charge $Q^{\pm} = \sum_i q_i^{\pm}$ (where $q_i^{\pm}$ is the magnitude of the $i_{th}$ positive (negative) pulse on a cycle and the summation is over one cycle) gives an indication of the source geometry. Figure 2 shows a substantial difference between the magnitudes of $Q^{+}$ and $Q^{-}$ characteristic of PD in an air gap with no dielectric material present.

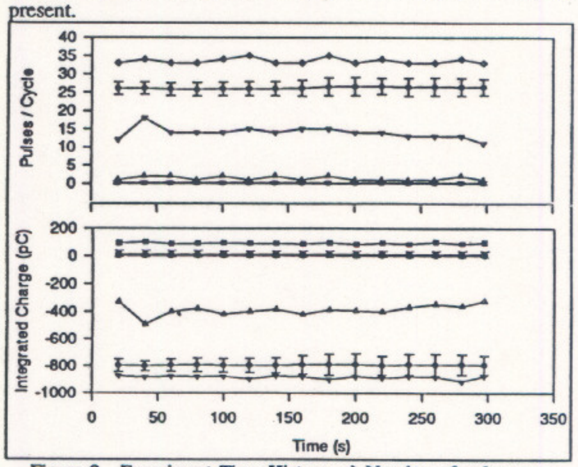


Figure 2. Experiment Time History: a) Number of pulses per cycle *vs* time, b) Integrated charge *vs* time. Upper and lower lines are maxima, centre line is average, error bars are one standard deviation (20 s processing period).

To obtain more insight into the PD source, one can consider memory effects. Figure 3 shows no correlation between the phase of the first negative pulse and the integrated charge on the previous half cycle such as seen when dielectric materials are present in the gap /2/. Since there is no noticeable change of phase with changes in integrated charge, no phase-to-phase charge storage is present. This indicates that the PD occurs in a metal-to-metal geometry as opposed to one incorporating one or more insulators.

### 4.2. Wuppertal Analysis Results

The partial discharge (PD) signals measured and recorded at NIST were converted to a local format at Wuppertal and classified by an automated PD diagnosis system. This PD diagnosis system employs multiple pattern recognition concepts with different feature extraction and classification algorithms /1,3/.

The fingerprints are extracted through Fourier analysis. The distance classifier used here is based on the Euclidean distance with respect to the physical traceability of the decision /4/. With these algorithms the degree of conformity between the NIST-measurement and the different PD faults from the entire reference library of the Wuppertal Institute is determined (Figure 4).

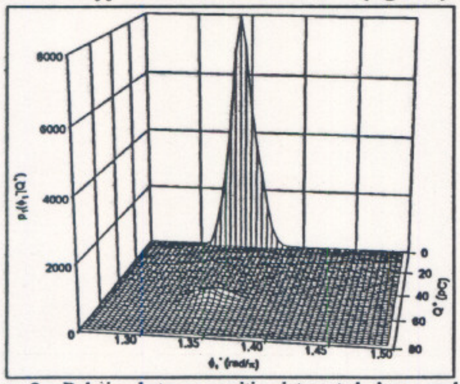


Figure 3. Relation between positive integrated charge and the phase-of-occurrence of the first pulse on the succeeding negative half-cycle. (Z axis not normalised)

The PD defect with the highest recognition rate is identified as the most probable cause of PD.

The NIST-measurement was correctly identified from 15 different PD reference faults as TIP ON HIGH POTENTIAL. In Figure 4, the 12 most probable PD faults are shown. In this figure the conformity is described as the conditional probability of occurrence of the different PD reference faults.

Considering the fact that the reference database was measured at a completely different test set-up as described in /3/ and at a different frequency (50 Hz instead of 200 Hz) the conformity of the NIST-measurement with the corresponding PD defect in the Wuppertal database is very high. The correct PD defect can be identified definitely by such a procedure.

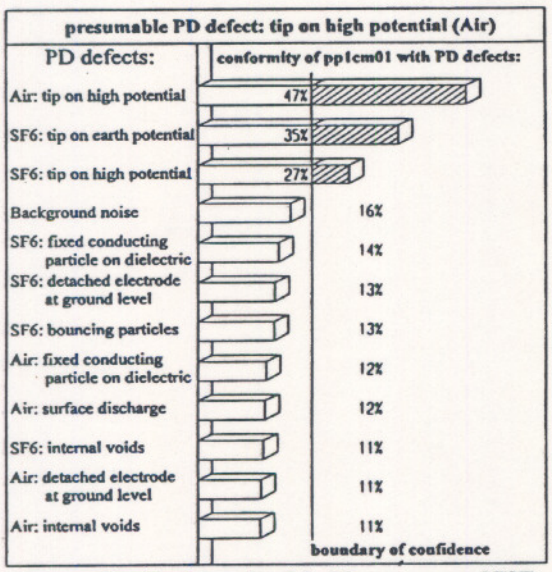


Figure 4. Classification results of the PD measurement (NIST) by an Euclidean distance classifier (reference database recorded in Wuppertal at 50 Hz, different test set-up)

### 4.3. CRIEPI Analysis Results

The data format developed at NIST and Wuppertal University, as described in section 3, was further validated. The PD measurements recorded at NIST were converted to a local data format at CRIEPI. The data were plotted as a reduced resolution $\phi$-q-n pattern in Figure 5. This format is useful for getting the feeling of the pattern differences among different electrode systems to enable rapid PD diagnosis by a human expert.

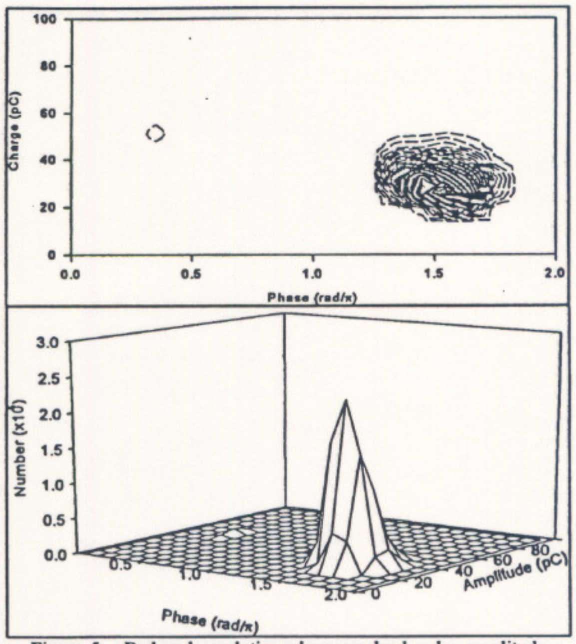


Figure 5. Reduced resolution phase-resolved pulse amplitude plot. Top: contour plot, bottom: three-dimension plot

### 5. Conclusion

An international standardised PD data format enables a world wide exchange of PD data. This opens the possibility of mutual assistance to identify the PD defects as shown in the example above. Additionally, the exchange of PD measurements can help to create a more general PD reference databases in future. Furthermore, the data format proposed here can be easily extended for future requirements.

**Data Format Support**

On the following FTP servers support for the standard PD data format is available (example programs in ANSI C and a full format description as postscript file):

wehs38.elektro.uni-wuppertal.de, login: anonymous /pub/pdf

### References

/1/ H.-G. Kranz, T. Hücker, A. Lapp, "A Partial Discharge Defect Identification System with Increased Diagnosis Reliability", 9th ISH, Graz, Austria, 1995

/2/ P. von Glahn, R.G. Van Brunt, "Continuous recording and stochastic analysis of partial discharge," IEEE Trans. Dielectrics & Elect. Insul., (in press).

/3/ T. Hücker, H.-G. Kranz, "The Requirement of automated PD Diagnosis Systems for Fault Type Identification under noisy Condition", IEEE Trans. Dielectrics & Elect. Insul., 3rd Volta Coll. issue 1995.

/4/ H.-G. Kranz, "Diagnosis of Partial Discharge Signals using Neural Network and Minimum Distance Classification", IEEE Trans.Elect. Insul., Vol. 28, No. 6, pp. 1016-1024, 1993.

**Addresses of main authors:**

Prof. Dr.-Ing. H.G. Kranz,
Laboratorium für Hochspannungstechnik
Fuhlrottstraße 10
42119 Wuppertal, Germany
TEL: ++49 / 202 439-3027, FAX: ++49 / 202 439-3026

Dr. Peter von Glahn
NIST, Bldg 220, Room B344
Gaithersburg, MD 20899, USA
TEL: (301) 975-2427, FAX: (301) 948-5796