Study on time-frequency extractors for partial discharge pulse current used in sequence grouping and identification

Abstract. This paper introduces a general way to make feature extraction for partial discharge (PD) pulse waveshape used in sequence grouping resorting with equivalent time-frequency method. A digital ac PD pulse detection system with bandwidth of 10 kHz - 40 MHz and sampling rate of 100 MS/s is introduced, which was developed to study those feature extractors. Simulation and test results with gas insulated switch (GIS) show that they are practical and effective to develop PD detection system with grouping and identification technique.

Streszczenie. Dokonano ekstrakcji impulsu wyładowania niezupelnego przy wykorzystaniu odpowiedniej metody czasowo-częstotliwościowej. Wykorzystano system cyfrowy z pasmem 10 kHz – 40 MHz i częstotliwością próbowania 100 MS/s. Badania wyłącznika gazowego potwierdziły skuteczność metody. (Czasowo-częstotliwościowa ekstrakcja i identyfikacja impulsu wyładowania niezupelnego)

Keywords: partial discharge; pulse current; feature extraction; grouping; identification.

Słowa kluczowe: wyładowanie niezupelne, impuls prądowy, identyfikacja.

Introduction

PD detection has almost become the most popular method to evaluate dielectric insulation condition at ac voltage applications [1-5]. For this, there are different kinds of PD measurement systems developed based on pulse current detection (the standard method) [6], UHF detection [7], acoustic detection [8] and even optical detection [9], with advantages and disadvantages respectively. Most of those conventional PD detection systems make the analysis of PD data, with assuming implicitly the presence of only one PD phenomenon (Fig. 1). But in operation conditions, more than one pulse source may be active due to the presence of either a plurality of defects or electrical noise, sometimes even both [10-12]. Different PD phenomena simultaneously active can combine in a mixed PD pulse sequence. To this problem, conventional PD detection systems suffer a major drawback. They only collect incoming pulse signals and compress the information in two parameters: the arrival time (phase) and the magnitude. In order to avoid noise and external interference pulses to be detected, these detection systems allow the operator to select different input filters. However, even in the ideal case of complete noise suppression, they may provide PD patterns that can be made of the superposition of different PD and noise sources. Thus, the different PD phenomena simultaneously active can be mixed in the different histograms deduced from the analysis of pulse height and phase distributions, which is currently used for diagnosis purpose. So those devices could fail to deal with the multi-PD sources.

Presently, a novel measurement system for ac PD that uses a pulse sequence grouping technique (Fig. 2) before the PD identification was presented in references [10-12]. It fully exploits the information carried by the PD pulse waveshape. On-board electronics and artificial intelligence tools enable to sort out sub-groups of pulse sequence, each of them relevant to one PD signal or noise source. Then some post-processing is applied to those sub-groups of different pulse signals, including noise rejection, identification and diagnosis. A simple dc PD pulse wideband detection system was developed with pulse source separation technique [13], which is also capable of handing noise suppression and multi-PD sources classification at dc voltage. Besides the grouping technique developed resorting to current pulse with wideband detection, reference [14] introduced remote radiometric measurements for detecting PD behavior in power transformer. All of those works [10-14] give a result that the pulse sequence grouping is useful for the PD measurement system, which is capable of handling noise rejection and multi-PD sources.

In design of a PD measurement system with pulse sequence grouping technique, finding a set of features corresponding to an optimal grouping performance
(accuracy and reliability) is critical. A PD diagnostic system designer typically does not have much difficulty to obtain a large number of features by applying different feature extraction methods on PD pulse waveshapes. However, the designer often faces challenges in finding a set of features which may give the optimal grouping performance. The primary reasons for that may be [15]: a) features cannot be evaluated individually since feature interaction may affect the grouping performance more significantly than features themselves; b) optimal features cannot be obtained by simply combining all features from different feature extraction methods since there exist redundant and irrelevant features; and c) a separability criterion should be applied to the feature extraction methods for analyzing their performance on the grouping objects before integrated into the measurement system.

This paper attempts to address the challenge by the following content. Firstly, according to the PD wideband detection system with sequence grouping technique, based on those pioneered works [10-15], and some works related to PD pulse waveshape, a general way to make feature extraction for PD pulse waveshape is provided. As an example, three feature extractors are obtained with post-processing methods applied to PD pulse waveshapes in time and frequency domains, which are equivalent evaluation, autocorrelation and information measure. And a separability criterion for pulse feature extractors is defined and analyzed with PD pulse sequences generated by the combination of exponential models and Monte Carlo simulation. Then, a digital ac PD pulse detection system with bandwidth of 10 kHz – 40 MHz and sampling rate of 100 MS/s is introduced, which was developed to study of those feature extractors used in PD pulse grouping. At last, Experimental verification is made resorting to the PD pulse detection system for GIS artificial models and Monte Carlo simulation. Then, a digital ac PD pulse detection system with bandwidth of 10 kHz – 40 MHz and sampling rate of 100 MS/s is introduced, which was developed to study of those feature extractors used in PD pulse grouping. At last, Experimental verification is made.

PD pulse feature extraction

Actually, PD pulse feature extraction is a mapping tool, which provides a compact and meaningful representation of the measured pulses embedded in the waveshapes in time domain and allows fast post-processing. In other words, the samples relevant to a single pulse waveshape should be compressed into a few numbers (two or three parameters), each of them bringing essential information. In Fig. 3, a general way to make feature extraction for PD pulse is shown. PD pulses are characterized by points in 2D Cartesian parameters plane \((\sigma_1, \sigma_2)\) or 3D Cartesian parameters space \((\sigma_1, \sigma_2, \sigma_3)\) [13], then visual inspection of results can be achieved with image representation. The parameters \(\sigma_1, \sigma_2, \sigma_3\) and \(\sigma_4\) may or not have relationship of each other. The values of them are decided by the transform \(T\) and functions \(F\) in Fig. 3 chosen by the PD detection system designer. A lot of feature extraction methods are available, if different transforms and linear or nonlinear functions are used. But an efficient mapping tool methods are available, if different transforms and linear or nonlinear functions are used. But an efficient mapping tool.

\[ R_p(t) = \lim_{\tau \to \infty} \left[ \sum_{k=0}^{n-1} p(t_k) p(t_k + \tau) / n \right] \]

For the PD pulse signal \(p(t_k), k = 0, 1, 2, \ldots, n-1\) sampled by the detection system with \(k\) points, its function \(R_p(m)\) can be defined as: \(R_p(m) = \sum_{i=1}^{n-1} p(t_i) p(t_{i+m}) / n\). Apparently, \(R_p(m)\) can be chosen as the parameters to make representation of the measured pulses. Based on this, a

![PD pulse waveform](Image)

Fig. 3. A general way of feature extraction for PD pulse waveshape

A simple way to characterize a signal \(p(t)\) simultaneously in time and frequency is to consider its mean localizations and dispersions. This can be obtained by considering \(|p(t)|^2\) and \(|P(f)|^2\) as probability distributions, and looking at their mean values and standard deviations [16]:

\[
\begin{align*}
\langle t \rangle &= \left( 1 / E_p \right) \int_{-\infty}^{\infty} t |p(t)|^2 dt \\
\langle f \rangle &= \left( 1 / E_p \right) \int_{-\infty}^{\infty} f |P(f)|^2 df \\
\langle T \rangle &= \left( 4 \pi / E_p \right) \int_{-\infty}^{\infty} (t - t_0)^2 |p(t)|^2 dt \\
\langle F \rangle &= \left( 4 \pi / E_p \right) \int_{-\infty}^{\infty} (f - f_0)^2 |P(f)|^2 df
\end{align*}
\]

where: \(E_p = \int_{-\infty}^{\infty} |p(t)|^2 dt\).

Then a signal can be characterized in the time-frequency plane by its mean position \((t_0, f_0)\) and a domain of main energy localization whose area is proportional to \(T \times F\).

According to Fig. 3, the \((T,F)\) and \((T,F,T,F)\) named equivalent time-frequency method (ETFM) are chosen to be the feature extractors for PD pulse signals, where \(TF = T \times F\). The \((T,F)\) plane is called "equivalent time-length and bandwidth" plane in reference [10-13], has already been verified that it can be successfully employed for achieving noise rejection and sources separation of ac PD and dc PD.

Autocorrelation is a special digital signal processing method [17] that has the ability to extract a measurement signal when it is completely swamped by noise, i.e. when the noise amplitude is larger than the signal amplitude. Unfortunately, phase information in the measurement signal is lost during the autocorrelation process, but the amplitude and frequency can be extracted accurately. For a measurement signal \(p(t)\), the autocorrelation coefficient \(R_p(t)\) is the average value of the product of \(p(t)\) and \(p(t - \tau)\), where \(p(t - \tau)\) is the value of the measurement signal delayed by a time \(\tau\). The autocorrelation function \(R_p(t)\) describes the relationship between \(R_p\) and \(\tau\) as \(\tau\) varies:

\[ R_p(\tau) = \lim_{N \to \infty} \left[ \sum_{t=1}^{N} p(t) p(t - \tau) dt / (2T) \right] \]
representation of the measured pulses named time-frequency autocorrelation method (TFACM) is described as:

\[
\begin{align*}
R_i(t) &= \frac{1}{n} \sum_{k=0}^{n-1} p(t_k) p(t_{k+i}) \quad (3)
R_i(0) &= \frac{1}{n} \sum_{k=0}^{n-1} p(t_k) / (n-1)
R_j(0) &= \frac{1}{n} \sum_{k=0}^{n-1} p(t_k) p(t_{k+j}) / (n-1)
R_j(0) &= \frac{1}{n} \sum_{k=0}^{n-1} p(t_k) p(t_{k+j}) / (n-1)

\end{align*}
\]

where: \(R_i(0)\) and \(R_j(0)\) – the maximum value of autocorrelation coefficients of \(p(t)\) in time domain and \(P(f)\) in frequency domain, respectively.

Then a PD pulse signal can be characterized in the time-frequency autocorrelation plane by \((R_i(m), R_j(m))\). The \((R_i(m), R_j(m), R_k(m))\) is also chosen to be the feature extractors for PD pulse signals according to Fig. 3, where \(R_k(m) = R_i(m) \times R_j(m)\).

Is there a simple definition which can describe the pulse signal for its special characteristic? A solution is given by applying an information measure to the signal \(p(t)\). This information, known as Renyi information [17], is given by

\[
R_i^c(t) = \log_2 \left[ \int_{-\infty}^{\infty} p^c(t) dt \right] / (n-1)
\]

where: the result by this measure is expressed in bits.

So here comes another compact and meaningful representation of the measured PD pulses named time-frequency entropy method (TFEM) described as:

\[
\begin{align*}
R_i^c &= -\log_2 \left[ \sum_{k=0}^{n-1} (t_k - \bar{t})^2 \cdot \rho^2(t_k) / \sum_{k=0}^{n-1} \rho^2(t_k) \right] \\
R_j^c &= -\log_2 \left[ \sum_{k=0}^{n-1} (j_k - \bar{f})^2 \cdot \rho^2(f_k) / \sum_{k=0}^{n-1} \rho^2(f_k) \right]
\end{align*}
\]

where: \(R_i^c\) and \(R_j^c\) – the second order Renyi information of \(p(t)\) in time domain and \(P(f)\) in frequency domain.

Then a PD pulse signal can be characterized in the time-frequency entropy plane by \((R_i^c, R_j^c)\) and space by \((R_i^j, R_j^i, R_k^i)\), where \(R_k^i = R_i^c \times R_j^i\).

**Separability criterion for PD pulse feature extractions**

In this section, a separability criterion for PD pulse feature extractors is defined. Before its application to the feature extraction results of the parameter extractors introduced in section 2 for an online field PD pulse signals, the ones of simulation PD pulse sequence are analyzed to show the grouping performance, theoretically. So in the following, simulation PD pulse sequence used in this section will be given before the definition of separability criterion.

Past investigations carried out to analyze PD signals were shown that different PD sources generated current pulses with a considerable variety of shapes. And the PD signals detected at the terminals of tested object are the convolution of the signal generated by PD source with the dynamic impedance of transmission path, which makes their shapes distorted in an unpredictable way. But most of the detected PD pulse signal can be described approximately with different exponential models (EMs) in the following [18, 19]:

\[
\begin{align*}
\hat{P}(x) &= Ae^{-\lambda x} \\
\hat{P}(y) &= Ae^{-\lambda y} \sin 2\pi f t \\
\hat{P}(z) &= Ae^{-(\lambda x - \lambda y)} \\
\hat{P}(w) &= Ae^{-(\lambda x - \lambda y)} \sin 2\pi f t
\end{align*}
\]

where: \(A\) – the amplitude, \(\tau\) – the attenuation rate and \(f_s\) – the oscillation frequency of the PD pulse signal, respectively.

In Fig. 4, noised simulation PD pulse signals generated using the expression (6) is shown.

![Fig.4. Simulation PD pulse signals generated with EMs](image)

Four PD pulse sequences generated by combination of EMs and Monte Carlo simulation (MCS) are shown in Fig. 5, where each one is corresponding to the pulse in Fig. 4. MCS is a method for iteratively evaluating a deterministic mode using sets of random numbers as inputs. It was once used in reference [20] to simulate the PD distribution of GIGRE Method II electrode system. Here, the way of MCS with combination of the EMs to generate PD pulse sequence is defined as five steps:

1): Choose a parametric model, such as \(P_i(x)\) in (6).

2): Generate a set of random inputs \(z_1, z_2, z_3\) with a certain distribution, such as hypodispersion. Redefine the parametric model:

\[
P_i(x) = (A + z_1) e^{-(\lambda x - z_2)} \sin 2\pi f(x + z_3) t
\]

3): Evaluate the model and store the results as \(P_i^c(x)\).

4): Repeat steps 2 and 3 for \(i = 1, \ldots, n\).

5): Make feature extraction for \(P_i^c(x), i = 1, \ldots, n\),

where in 5), feature extraction is made to the simulation results (PD pulse sequence) instead of conventional analysis such as histogram, summary statistics and confidence intervals, etc. The feature extraction results with application of ETFM, TFACM and TFEM are shown in the Fig. 6, which will be evaluated by the separation criterion defined in following content.

For the set \(\{\sigma\}_{i=1,2,\ldots,k}\) \(k = 1,2,\ldots, k\) can be treated as one of the most important measures to analyze the performance. The mean value of the distance between all elements in set \(\{\sigma\}_{i=1,2,\ldots,k}\) can be defined as

\[
D_i^2((\sigma)_{i=1,2,\ldots,k}) = \sum_i \sum_{j<i} D_j^2((\sigma)_{i=1,2,\ldots,k}) / (k(k-1))
\]

where: \(D_j^2((\sigma)_{i=1,2,\ldots,k})\) – the distance between element \(\sigma\) and \(\sigma_j\).
PRZEGLĄD ELEKTROTECHNICZNY (Electrical Review), ISSN 0033-2097, R. 87 NR 10/2011

343

The distance $D^3(\{\sigma_i\}_i, \{\sigma_j\}_j)$ can be defined as the 3D feature extractor applied to PD pulse waveshape:

$$D^3(\{\sigma_i\}_i, \{\sigma_j\}_j) = (\sigma^1_i - \sigma^1_j)^2 + (\sigma^2_i - \sigma^2_j)^2 + (\sigma^3_i - \sigma^3_j)^2$$

For the set $\{\sigma_i\}_i, i = 1, 2, \ldots, k_a$ and set $\{\sigma_j\}_j, i = 1, 2, \ldots, k_b$ formed by the same feature extractor applied to two different pulse sequences A and B, the mean value of the distance between all elements in sets $\{\sigma_i\}_i, i = 1, 2, \ldots, k_a$ and $\{\sigma_j\}_j, i = 1, 2, \ldots, k_b$ can be defined as

$$D^3(\{\sigma_i\}_i, \{\sigma_j\}_j) = \frac{1}{k_a} \sum_{i=1}^{k_a} \sum_{j=1}^{k_b} D^3(\sigma^1_i, \sigma^1_j)$$

(10)

where $D^3(\sigma^1_i, \sigma^1_j) = (\sigma^1_i - \sigma^1_j)^2$ – the distance between element $\{\sigma_i\}_i$ and $\{\sigma_j\}_j$.

It has the same definition as the expression (9). Apparently, if the $D^3(\{\sigma_i\}_i, \{\sigma_j\}_j)$ or $D^3(\{\sigma_i\}_i, \{\sigma_j\}_j)$ are with small values and $D^3(\{\sigma_i\}_i, \{\sigma_j\}_j)$ is with big value, the feature extractor applied to PD pulse sequence A and B is effective. So with the expressions (9) and (10), the separability criterion to evaluate the performance of feature extraction methods for PD pulse sequence grouping can be defined as

$$J^2_{A,B} = \frac{D^3(\{\sigma_i\}_i, \{\sigma_j\}_j) + D^3(\{\sigma_i\}_i, \{\sigma_j\}_j)}{D^3(\{\sigma_i\}_i, \{\sigma_j\}_j) + D^3(\{\sigma_i\}_i, \{\sigma_j\}_j)}$$

(11)

where: $J^2_{A,B} > 1$ means that the feature extractor applied to pulse sequences A and B for grouping has good property, while $J^2_{A,B} \leq 1$ means poor property.

Table 1 to 3 shows the corresponding values of separability measure with (2D, 3D) form, which are obtained by application of expressions (8), (9) and (11) to the feature extraction results shown in Fig. 6. The mean values of separability measure using ETFM, TFACM and TFEM are (5.59, 8.94), (5.58, 6.52) and (2.15, 2.40), which gives a result that all feature extractor have good performance used in PD pulse sequence grouping.

Table 1. Values of separability measure of ETFM (2D, 3D)

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.18, 29.1</td>
<td>2.81, 29.6</td>
<td>1.12, 2.65</td>
</tr>
<tr>
<td>B</td>
<td>14.5, 493</td>
<td>5.42, 6.42</td>
<td>5.27, 10.5</td>
</tr>
<tr>
<td>C</td>
<td>64.6, 226</td>
<td>48.0, 1573</td>
<td>6.92, 10.0</td>
</tr>
<tr>
<td>D</td>
<td>272, 1194</td>
<td>46.0, 1407</td>
<td>6.74, 7.61</td>
</tr>
</tbody>
</table>

Table 2. Values of separability measure of ETFM (2D, 3D)

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>7.5, 7.9</td>
<td>14.7, 18.7</td>
<td>2.26, 1.55</td>
</tr>
<tr>
<td>B</td>
<td>91.8, 122</td>
<td>12.8, 137</td>
<td>3.43, 3.25</td>
</tr>
<tr>
<td>C</td>
<td>55.5, 80.2</td>
<td>14.7, 18.7</td>
<td>1.29, 6.07</td>
</tr>
<tr>
<td>D</td>
<td>171, 211</td>
<td>142, 201</td>
<td>1.29, 6.07</td>
</tr>
</tbody>
</table>

Table 3. Values of separability measure of ETFM (2D, 3D)

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.44, 3.55</td>
<td>1.38, 8.39</td>
<td>0.44, 3.55</td>
</tr>
<tr>
<td>B</td>
<td>1.29, 6.07</td>
<td>1.38, 8.39</td>
<td>1.29, 6.07</td>
</tr>
<tr>
<td>C</td>
<td>2.78, 8.12</td>
<td>1.38, 8.39</td>
<td>1.29, 6.07</td>
</tr>
<tr>
<td>D</td>
<td>2.20, 14.49</td>
<td>1.38, 8.39</td>
<td>1.29, 6.07</td>
</tr>
</tbody>
</table>

A lot of values of separability measure were obtained resorting to feature extraction results from different simulation PD pulse sequences, which shows almost the same analysis result shown in the above. This gives a theoretical approve that those feature extractors (ETFM, TFACM and TFEM) are capable of process the PD pulse.
AC PD measuring system description

For many applications, it has been observed that a bandwidth up to 40 MHz, associated with a digitizer as fast as 100 MS/s is sufficient to provide enough information for PD signals detection [10, 13]. In Fig. 7, a simple measuring system for the digital acquisition of AC PD current pulse signals was being developed resorting to a fast digitizer of NI5112 type (its analog bandwidth is 100 MHz, the maximum sample rate is 100 MS/s and the memory length is 16 MS in each channel). A 10 kHz to 40 MHz hardware bandwidth limit is imposed to remove unwanted low-frequency and high-frequency noise for the signal detection Channel A. Channel A is connected to the measuring impedance in the test circuit, while Channel B is connected to a voltage divider allows the synchronization of PD signals acquisition with the applied voltage.

In Fig. 7, the trigger source, trigger type, trigger level, trigger slope, min sample rate and min sample length of the digitizer are set to be Channel B, trigger Edge, 0, Positive, 100 MS/s and 2 MS, respectively. These parameters mean that the detection system only works when Channel B gets a coming AC voltage signal at its zero phase, then after Channel A samples signals from detection impedance with 2MS (100 MS/s×20 ms = 2 MS), the system stops and will work again until the coming AC voltage signal at its another zero phase. For the frequency of AC voltage is 50 Hz in China, so 20 ms is one period of AC voltage.

The next work is to extract PD pulse waveshapes and their corresponding phase information within the 2 MS sampled by Channel A, using a moving extraction window (MEW) shown in Fig. 7. The phase value is easy to get for the ith point within 2 MS: phase(i)=360*i/2M (Φ in Fig. 7). The PD pulse signals can be extracted using MEW with a threshold value and a simple algorithm. The length of MEW can be 100 (1 us), 200 (2 us), 300 (3 us), ……, 1000 (10 us) points should be done is defined as:

\[ \text{بك} = \frac{1}{2} \sum_{i=1}^{N} P_i \]

for \( j = 1, 2, \ldots, N \) (\( N \) is the total number of pulse waveshapes) and \( i = 1, 2, \ldots, k \), where: \( p_i \) and \( \Delta t(i-1) \) – the magnitude and time of the ith date point of the pulse waveshape \( p_i \), \( \Delta f \) and \( \Delta f(i-1) \) – the magnitude and frequency component of the ith date point of \( P_j \) (\( \Delta t = 1/f_s \), \( \Delta f = f_s/(k+1) \) is the sampling rate).

Front-ends are specifically designed for AC PD measurement to present the detection and analysis results. One of them is shown in Fig. 8. It provides a set of push-buttons to do the pulse feature extraction and unsupervised clustering of the acquired pulse sequence.

Experimental verification

To investigate the grouping performance of the detection system with those feature extractors for online field PD signals, the PD signals are sampled form a GIS device and two PD calibrators, respectively. The GIS device is filled with the gas SF\textsubscript{6} and its inside pressure strength is 0.35MPa [21]. It is shown in Fig. 9 with the four typical PD phenomena including pulse height and phase distributions, with the optimum data partition which consists of four sub-groups, are reported in Fig. 10a.

Unknown external noise also exists during the data processing. Fig. 10b and 10e can be verified by their repetition rates and corresponding pulse waveshapes, which are generated by the two PD Calibrators. But for no repetition reproducibility of the waveforms, some ambiguities were obtained. 2D feature distributions shown in Fig. 10b to 10e are overlapped. The corresponding 2D and 3D feature distributions, with the optimum data partition which consists of four sub-groups, are shown in Fig. 11, which give the same separation results. The typical pulse waveshapes relevant to the A, B, C and D sub-groups are reported in Fig. 12. Four different PD pattern plots can be obtained, eventually, associating each point in the A, B, C and D sub-groups are shown in Fig. 10b to 10e.

Fig. 10b to 10e show four PD patterns obtained for each sub-group, which is relevant to each PD phenomenon and external noise. Fig. 10b and 10e can be verified by their repetition rates and corresponding pulse waveshapes, which are generated by the two PD Calibrators. But for no characteristics for B and C sub-groups with their PD.
patterns (Fig. 10c and 10d) and corresponding pulse waveshapes, they can be judged as unknown external noise signals.

PD measurements were performed adopting the MEW with a high threshold value (10mV) in order to avoid the pulsating background noise. The sampled PD data is shown in Fig. 13a, which was collected at a certain test voltage for 200 power cycles. Fig. 13a provides the original PD pattern, where PD phenomena are mixed with random pulsed signals. The corresponding 2D and 3D feature distributions, with the optimum data partition which consists of three subgroups, are show in Fig. 14. The typical PD pulse waveshapes in time and frequency domains relevant to the A, B and C sub-groups are reported in Fig. 15. Three different PD pattern plots can be obtained, eventually, associating each point in the A, B and C sub-groups are shown in Fig. 13b, 13c and 13d.

Fig. 13c and its waveshapes shown in Fig. 15b can be verified by its pulse height and phase distribution, which belongs to a PD phenomenon of corona discharge. So it can be judged as a PD source of the protrusion on bus bar (see Fig. 9a).

The typical pulse waveshape shown in Fig. 15a and its corresponding PD pattern in Fig. 13b both indicate that it is generated by a pulsed interference signals source, and it needs the length \( k \) of MEW set to be a larger number to record its whole pulse waveshape in time domain. While the typical pulse waveshapes shown in Fig. 15c also have the same characteristic as a PD pulse, but its corresponding PD pattern (Fig. 13d) shows that the pulse sequence has a stochastic phase distribution which indicates it must be a random pulsed interference signal source.

PD tests were performed adopting the MEW with a high threshold value (10mV) in order to avoid the pulsating
background noise. The length of MEW \( k \) was set to be 200 points. An interesting sampled PD data is shown in Fig. 16a for an unknown corona discharge source also active, which was collected at a certain high test voltage for 200 power cycles. Fig. 16a provides the original PD pattern. The corresponding 2D and 3D feature distributions, with the optimum data partition which consists of two sub-groups, are shown in Fig. 17. The typical PD pulse waveshapes in time and frequency domains relevant to the A and B sub-groups are reported in Fig. 18. Two different PD pattern plots can be obtained, associating each point in the A and B sub-groups are shown in Fig. 16b and 16c.

Fig. 16b and its waveshapes shown in Fig. 18a can be verified by its pulse height and phase distribution, which belongs to a PD phenomenon generated by the protrusion on earth potential (see Fig. 9b). Fig. 16c and its waveshapes shown in Fig. 18b are from a PD phenomenon generated by the protrusion on bus bar, but it is an unexpected result.
judged as a PD phenomenon of floating discharge. But the corresponding 2D and 3D feature distributions, with the optimum data partition which consists of two sub-groups, are shown in Fig. 20. The typical pulse waveshapes in time and frequency domains relevant to the A and B sub-groups are reported in Fig. 21. Two similar PD pattern plots are obtained, eventually, associating each point in the A and B sub-groups are shown in Fig. 19b and 19c.

Fig. 19. PD patterns: (a) is the original one; (b) is the PD pattern of sub-group A; (c) is the PD pattern of sub-group B

The PD pulse sequence grouping gives an unexpected result for this artificial PD defect (see Fig. 9c), because the test designer only wants one PD phenomenon of floating discharge. But both sub-groups A and B in Fig. 19 can be verified by its pulse height and phase distribution, which belongs to a PD phenomenon of floating discharge.

PD measurements were also performed adopting the MEW with a high threshold value (10mV) and length of 200 points. One sampled PD data is shown in Fig. 23a, which was collected at a certain test voltage for 200 power cycles.
the pulse waveshapes in time and frequency domains are tiny.

The FFT is chosen as the transform to make feature extraction for PD pulse waveshape with consideration of a good compromise between complexity of computations and real-time requirements. This powerful tool lies at the base of the concepts of frequency spectrum and frequency response. As we all know, the FT is perfectly applicable to infinite and periodic signals since it is based on the assumption that any signal can be decomposed into a series of sine or cosine waveforms. Other popular and effective transforms, such as fast Mellin transform (FMT) \[22\], independent component analysis (ICA) \[19\], and Hilbert-Huang transform (HHT) \[23\], etc, can also be put into consideration for the PD pulse feature extraction, if the formed feature parameters are with good separability to stochastic pulse waveshapes of different PD sources or pulsed interference signals.

In this paper, how to make and evaluate the feature extraction methods of PD pulse waveshape used in sequence grouping are discussed, all of which meet the needs of online monitoring of PD data in real-time. For post processing of PD data not in real-time, some feature extractors such as principal component analysis (PCA), singular value decomposition (SVD), and Wigner-Ville distribution (WVD) \[18\], etc, can also be chosen as feature extractors of PD pulse waveshape, although they are time and memory consuming.

Conclusions

A general way to make feature extraction for PD pulse waveshape used in sequence grouping is presented. Three feature extractors are obtained with different post-processing to pulses in time and frequency domains, which are equivalent evaluation, autocorrelation and information measure. A separability criterion for PD pulse feature extractors is defined and how to generate PD pulse sequence with Monte Carlo simulation based on exponential models is presented, which is used to evaluate the grouping performance of those feature extractors in theoretically. The analysis results show that all feature extractors have good performance used in PD pulse sequence grouping. To investigate the grouping performance of the detection system with those feature extractors for an online field PD signals, a simple digital ac PD pulse measurement system with bandwidth of 10 kHz – 40 MHz and sampling rate of 100 MS/s is introduced. Experimental verification is made resorting to the ac PD detection system with GIS artificial defect models and two PD calibrators respectively, which shows that those feature extractors are practical and effective to develop the PD pulse measurement system with grouping technique to deal with noise rejection and multi-PD sources.

REFERENCES


---

**Discussion**

The above results show that this digital ac PD measurement system is capable of handling noise rejection for PD sources resorting to the recorded pulse waveshapes. Features related to pulse waveshapes introduced in section 2 (ETF, TFAC, and TFEM), can provide a significant contribution to solve the grouping problem when different kinds of pulse sources provide different signal waveshapes. Good results are obtained when the pulse height and phase distributions are overlapped, even the difference between


Authors: Dr. Si Wenrong, East China Electric Power Test & Research Institute Company Limited, Shanghai, China, 200437, E-mail: siwenrong@gmail.com; Dr. Fu Chenzhao, East China Electric Power Test & Research Institute Company Limited, Shanghai, China, 200437, E-mail: dsy_fucz@ec.sgcc.com.cn; Dr. Gao Kai, East China Electric Power Test & Research Institute Company Limited, Shanghai, China, 200437, E-mail: dsy_gaok@ec.sgcc.com.cn.