Research on improved time difference estimation algorithm based on fourth-order cumulant in UHF PD location

Wenbing Zhu¹, Wei Xu, Jiabin Zhou, Ran Xu and Feng Yi
State Grid Shandong Electric Power Research Institute, Jinan, Shandong 250002, China
¹ E-mail: 2020202070053@whu.edu.cn

Abstract. Ultra-high frequency (UHF) detection technology is a fast-developing method for partial discharge (PD) diagnosis of electrical equipment in recent years. Extracting accurate time difference is the priority for PD location in electrical equipment using UHF technology. An improved time difference estimation algorithm based on fourth-order cumulants was first proposed in this article. It mainly included the following two steps, that is using the "correlation-shift-superposition" transformation method to process multiple samples collected from a PD source at first, and then utilizing the statistical characteristics of the fourth-order cumulant to calculate the time difference. In addition, the accuracy and stability of the improved algorithm and the other three common methods were compared. The calculation results show that the improved time difference estimation algorithm based on the fourth-order cumulant can obtain a more accurate and stable time difference than the other three methods. Improved time difference estimation algorithm based on fourth-order cumulant can approximately meet the requirements for reliable location of PD sources.

1. Introduction
Partial Discharge (PD) is an early discharge phenomenon caused by internal insulation defects in electrical equipment [1, 2]. PD is a significant reason for the insulation breakdown of high-voltage electrical equipment, also a sign of insulation deterioration. How to achieve fast, efficient, accurate, and reliable PD location has always been a major difficulty in the field of online monitoring and fault diagnosis of electrical equipment.

UHF detection technology has developed rapidly in PD fault diagnosis of electrical equipment in recent years [3], has the advantages of strong anti-interference ability, broad frequency band, and high detection accuracy. However, in practical applications, the UHF-based PD location method has a series of problems to be solved, such as the accuracy of the time difference estimation algorithm of the UHF signal. In the current PD location research and application, the common methods of the time difference estimation mainly include the initial peak method, the generalized correlation method, and the energy accumulation method. The initial peak method and the energy accumulation method require a high signal-to-noise ratio (SNR) of the signal. The generalized correlation method requires high similarity of the signal, stable and uncorrelated noise of the two channels. These factors limit the accuracy of time difference estimation.

In this paper, an improved UHF PD location time difference estimation algorithm based on fourth-order cumulant was proposed, and the accuracy and stability of the improved algorithm and the other
three common time difference estimation algorithms were compared and analyzed using the PD measured signal.

2. Improved time difference estimation algorithm

When the signal is a non-Gaussian process and the noise obeys the Gaussian distribution, the high-order statistical characteristics of the signal can be used to estimate the time difference [4].

The improved UHF PD location time difference estimation algorithm based on fourth-order cumulant mainly included the following two steps, that is using the "correlation-shift-superposition" transformation method to process multiple samples collected from a PD source at first, and then utilizing the statistical characteristics of the fourth-order cumulant to calculate the time difference.

2.1. Correlation-shift-superposition transformation

Correlation refers to the degree of internal correlation of one or two signals at different moments [5].

The definition of the cross-correlation function is

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t)x(t+\tau)dt$$

The waveform of the PD sampling signal is discrete. The discrete cross-correlation function is defined as

$$R_{ij}(m) = \sum_{k=1}^{s}x_i(k)x_j(k+m)$$

Where $m$ is the number of shifting point, $s$ is signal length, $k$ is the numerical order of sampling points; $x_i$ and $x_j$ are the waveform data collected by two different UHF sensor arrays.

For the $n$ waveforms collected by each UHF sensor array at different times, if the $k$th waveform among the $n$ waveforms is the most similar one to the other $n-1$ waveforms, the $k$th waveform will be considered to be the best one in reflecting the discharge characteristics, and the $k$th sample should be used as the shift standard. Shift refers to changing the relative distance between the standard and other samples until the cross-correlation function of the two samples takes the maximum value. Superposition refers to taking the average after superposition. The specific steps of correlation-shifting-superposition transformation are as follows.

1. Calculate the cross-correlation function value between each sample and the other $n-1$ samples and sum them up (the number of shifting is zero).
   $$\begin{align*}
   M_1 &= R_{12}(0) + R_{13}(0) + \cdots + R_{1n}(0) \\
   M_2 &= R_{21}(0) + R_{23}(0) + \cdots + R_{2n}(0) \\
   &\vdots \\
   M_n &= R_{n1}(0) + R_{n2}(0) + \cdots + R_{n(n-1)}(0)
   \end{align*}$$

2. Compare the $n$ sum values above, and use the sample corresponding to the maximum value as the shift standard.

3. The $m$ value corresponding to the maximum cross-correlation function between the shift standard and the sample to be shifted is taken as the number of shifting point.

4. After all the samples have been shifted, $n$ samples are superimposed and then averaged to obtain statistically significant waveform sample $\bar{x}_a$

$$\bar{x}_a = \frac{1}{n} \sum_{k=1}^{n} x_k$$

2.2. Basic properties and mathematical models of higher-order cumulants

Noise is generally a Gaussian process and the signal is a non-Gaussian random process [6]. The zero-mean Gaussian noise usually has the following three basic properties: (1) The second-order moment is
the same as the second-order cumulant of the zero-mean Gaussian noise, and equal to its variance. (2) the odd-order moment of the zero-mean Gaussian noise is identically equal to zero, and the even-order moment is not equal to zero; (3) The third-order and above cumulant of the zero-mean Gaussian noise is identically equal to zero [7].

\[ X = \begin{bmatrix} X_1, X_2, \ldots, X_n \end{bmatrix}^T \]

is a random vector, \( \omega = [\omega_1, \omega_2, \ldots, \omega_n]^T \), the moment generating function of the random vector \( X \) is

\[
\Phi(\omega_1, \omega_2, \ldots, \omega_n) = E\{\exp[j(\omega_1 X_1 + \omega_2 X_2 + \ldots + \omega_n X_n)]\}
\]

(5)

The cumulant generating function is

\[
\Psi(\omega_1, \omega_2, \ldots, \omega_n) = \ln[\Phi(\omega_1, \omega_2, \ldots, \omega_n)]
\]

(6)

Expanding Equations (5) and (6) according to Taylor series, then the \( r \)-th order \((r = r_1 + r_2 + \ldots + r_k)\) joint moment and joint cumulant of random vector \( X \) are defined as:

\[
m_{r_1, r_2, \ldots, r_k} = E\{X_1^{r_1} \ldots X_k^{r_k}\} = (-j)^r \left| \frac{\partial^r \Phi(\omega_1, \ldots, \omega_k)}{\partial \omega_1^{r_1} \ldots \partial \omega_k^{r_k}} \right|_{\omega_1 = \ldots = \omega_k = 0}
\]

(7)

\[
c_{r_1, r_2, \ldots, r_k} = (-j)^r \left| \frac{\partial^r \Psi(\omega_1, \ldots, \omega_k)}{\partial \omega_1^{r_1} \ldots \partial \omega_k^{r_k}} \right|_{\omega_1 = \ldots = \omega_k = 0}
\]

(8)

Where \( E\{\cdot\} \) is the mathematical expectation. \( r_1, r_2, \ldots, r_k \) are the order of the sequence. \( \omega(i=1, 2, \ldots, L) \) represents the angular frequency component of continuous time.

Let \( \{y(n)\} \) be a \( k \)-order stationary random process with zero mean, its \( k \)-order moment \( M_k(y, \tau_1, \tau_2, \ldots, \tau_{k-1}) \) and \( k \)-order cumulant \( C_k(y, \tau_1, \tau_2, \ldots, \tau_{k-1}) \) are respectively defined as

\[
M_k(y, \tau_1, \tau_2, \ldots, \tau_{k-1}) = m_k \{y(n), y(n + \tau_1), \ldots, y(n + \tau_{k-1})\}
\]

(9)

\[
C_k(y, \tau_1, \tau_2, \ldots, \tau_{k-1}) = c_k \{y(n), y(n + \tau_1), \ldots, y(n + \tau_{k-1})\}
\]

(10)

Where \( \tau_m(m=1, 2, \ldots, r-1) \) is the time difference.

2.3. The principle of the fourth-order cumulant used in time difference estimation

If the probability density distribution of a random process is symmetric, the third-order cumulant of the random process is always equal to zero [7, 8]. In practical applications, the time difference estimation algorithm based on the fourth-order cumulant signal is generally selected for calculation [9].

The relative time difference of two signals \( x(t) \) and \( y(t) \) excited by a same excitation source after passing through different propagation paths is

\[
\begin{cases}
x(t) = s(t) + w_1(t) \\
y(t) = As(t-D) + w_2(t)
\end{cases}
\]

(11)

Where \( s(t) \) is an unknown signal, \( s(t-D) \) is an unknown signal with a delay of \( D \), \( A \) is the gain of the signal, \( w_1(t) \) and \( w_2(t) \) are unknown additive noise sources.

The four-order cumulants of the signals \( x(t) \) and \( y(t) \) are

\[
C_{xxyy}(\tau_1, \tau_2, \tau_3) = c_4 \{x(t), x^*(t+\tau_1), y(t+\tau_2), y^*(t+\tau_3)\}
\]

(12)

Where "*" means conjugate. Substituting Equation (11) into Equation (12), we can get

\[
C_{xxyy}(\tau_1, \tau_2, \tau_3) = A^4 c_4 \{s(t), s^*(t+\tau_1), s(t+\tau_2-D), s^*(t+\tau_3-D)\}
\]

(13)

Where \( C_{x4}(\tau_1, \tau_2-D, \tau_3-D) \) is the fourth-order self-cumulant of signal \( s(t) \), and the fourth-order self-cumulant of signal \( x(t) \) can be defined as

\[
C_{x4}(\tau_1, \tau_2, \tau_3) = c_4 \{x(t), x^*(t+\tau_1), x(t+\tau_2), x^*(t+\tau_3)\}
\]

(14)

Similarly
From Equation (13) and Equation (15), we can get
\[ C_{xy}(\tau_1, \tau_2, \tau_3) = A^2 C_{xyy}(\tau_1, \tau_2 - D, \tau_3) \] (16)

From Equations (15) and (16), we can see that the time difference between \( C_{xyy}(\tau_1, \tau_2, \tau_3) \) and \( C_{xy}(\tau_1, \tau_2, \tau_3) \) is \( D \). The cross-correlation function of \( C_{xyy}(\tau_1, \tau_2, \tau_3) \) and \( C_{xy}(\tau_1, \tau_2, \tau_3) \) is
\[ R_{xy}(\tau) = \int \int \int C_{xyy}(\tau_1, \tau_2, \tau_3) C_{xyy}^*(\tau_1, \tau_2 + \tau, \tau_3 + \tau + D) d\tau_1 d\tau_2 d\tau_3 \] (17)

Substituting Equations (16) into Equations (17), we can get
\[ R_{xy}(\tau) = A^4 \int \int \int C_{xyy}(\tau_1, \tau_2, \tau_3) C_{xyy}^*(\tau_1, \tau_2 + D, \tau_3 + D) d\tau_1 d\tau_2 d\tau_3 \] (18)

So there is
\[ |R_{xy}(\tau)| \leq A^4 \int \int \int |C_{xyy}(\tau_1, \tau_2, \tau_3)|^2 d\tau_1 d\tau_2 d\tau_3 \] (19)

When \( \tau = D \), \( R_{xy}(\tau) \) takes the maximum value. At this time, \( \tau \) is used as the estimated value of time difference \( D \), so that the time difference value between signals \( x(t) \) and \( y(t) \) can be determined.

### 2.4. Time difference estimation algorithm based on fourth-order cumulant

The discretization of the model described in Equation (11) can be defined as
\[
\begin{align*}
x(n) &= s(n) + w_1(n) \\
y(n) &= s(n - D) + w_2(n)
\end{align*}
\] (20)

The following assumptions are made on the discretized signal propagation model described in Equation (20).

1. Suppose the excitation source \( \{e(n)\} \) is a zero-mean non-Gaussian independent and identically distributed sequence. The signal \( s(n) \) is generated by using \( \{e(n)\} \) to excite an exponentially stable linear system, that is
\[ s(n) = \sum_{h=-\infty}^{\infty} f(h)e(n - h) \] (21)

Where \( f(h) \) represents the impulse response system.

2. \( E(\{e^2(n)\}) - 3E(\{e^4(n)\})^2 = 0 \) and \( E(\{e^4(n)\}) < 0 \).

3. The noise \( w_1(n) \) and \( w_2(n) \) are stationary zero-mean Gaussian processes, independent of the excitation source \( \{e(n)\} \) and the signal \( \{s(n)\} \). And the following Equation (22) holds
\[ \cos \{w_1(n), w_2(n)\} \leq M \beta^{n-\lambda} \] (22)

Where \( 0 < M < \infty \) and \( 0 < \beta < 1, b = 1, 2, l = 1, 2, n_1 \) represents the \( n_1 \)-th sampling point.

If the Hypothesis (1) and the Hypothesis (2) are both true, when \( n \geq 1 \), the signal \( \{s(n)\} \) is fourth-order generalized stationary, that is, the third-order and lower cumulants of the signal \( \{s(n)\} \) are time changeless. According to the time difference estimation criterion function based on the fourth-order cumulant given by Tugnait and other scholars \[4, 10\]
\[ J_1(d) = \frac{|C_4\{x(n-d), x(n-d), y(n), y(n)\}|}{\sqrt{|C_4\{x(n)\}||C_4\{y(n)\}|}} \] (23)

When the function \( J_1(d) \) takes the maximum value, the corresponding \( d \) at this time can be used as the estimated value of the time difference, that is, \( D = d \). When using this method to estimate the time difference, the Equation (23) needs to be converted into a discrete form (4, 8).
\[ J_{1N}(d) = \frac{|C_4\{x(n-d), x(n-d), y(n), y(n)\}|}{\sqrt{|C_4\{x(n)\}||C_4\{y(n)\}|}} \] (24)
Where $c_{4N}^\prime{}$ is the numerical estimate of the fourth-order cumulant, its calculation formula is as follow

$$c_{4N}^\prime{} = \frac{1}{N} \sum_{n=1}^{N} x^4(n) - 3 \left( \frac{1}{N} \sum_{n=1}^{N} x^2(n) \right)^2$$

(25)

Where $N$ is the sampling length. It can be proved that the following Formula (26) is true when the Hypotheses (1) ~ (3) are true.

$$J_{1N}(d) \xrightarrow{a.p.1} J_1(d) N \rightarrow \infty$$

(26)

Where $a.p.1$ represents the discretized time difference estimation criterion function $J_{1N}(d)$ converges to $J_1(d)$ with probability 1. It can be seen that the corresponding $d$ when the function $J_{1N}(d)$ takes the maximum value should be used as the time difference estimation value, that is, $D=d$.

3. Time difference calculation and analysis of PD actual measurement signal

3.1. The test platform of real measured PD signal

As shown in Figure 1, the experimental platform of the real measured PD signal is mainly composed of a large transformer box (240cm×310cm×200cm), an analog discharge source (needle-board defect model), a high-frequency coaxial transmission line, and a four-channel ultra-wideband LeCroy7100 high-speed oscilloscope (its analog bandwidth is 1GHz, maximum sampling rate is 20GS/s, and dual-channel storage capacity is 48 MB), UHF sensor array (its center frequency is 437.5MHz, and bandwidth is about 350MHz~525MHz), etc. In order to eliminate the influence of the system delay, the output of the sensor array is directly sent to the oscilloscope through the coaxial cable. The length of the coaxial transmission line of each channel is required to be equal. T1 is an electric voltage regulator (TEDGC-25), T2 is a corona-free test transformer (YDTW-25kVA/100kV), and R is a protection resistor (20kΩ). In the experiment, the position of PD source is fixed and the difference of interval between the two sensors and PD source is changed.

Figure 1. Schematic diagram of experimental system for PD location.

3.2. Time difference calculated by four different methods

Under the same range difference, the positioning time difference is estimated by the initial peak method, the generalized correlation method, the energy accumulation method and the improved time difference estimation algorithm based on fourth-order cumulant respectively. Before the calculation, the complex wavelet transform [11] is used to pre-process the signal to de-noise, so as to suppress the white noise and periodic interference of the PD signal. Figure 2 shows a group of original UHF signal waveforms collected by two sensors. The waveform contains many burrs because of the noise, and the
distortion of waveform is serious. Figure 3 shows the waveform after wavelet denoising and filtering. Obviously, after the elimination of low-frequency noise, the waveform burr is significantly reduced and the initial peak of the waveform is easier to judge.

**Figure 2.** The original signal waveform.  **Figure 3.** The waveform after wavelet denoising.

Set the range difference $\Delta d$ to 317 cm, its theoretical time difference is 10.566 ns. The results of time difference calculated by the initial peak method, the generalized correlation method, the energy accumulation method and the improved algorithm are shown in Table 1.

It can be seen from the data in Table 1 that

(1) When the initial peak value method is used to calculate the time difference, the average relative error is about 105.3%, the maximum relative error is about 309.0%, and the minimum relative error is about 3.6%. Its calculation results fluctuate greatly, and accuracy and stability are extremely low.

(2) When the generalized correlation method is used to calculate the time difference, the average relative error is about 63.9%, the maximum relative error is about 166.0%, and the minimum relative error is less than 0.1%. Its calculation results still fluctuate greatly, and accuracy and stability are low, but better than the initial peak value method.

(3) When the energy accumulation method is used to calculate the time difference, the average relative error is about 17.3%, the maximum relative error is about 33.4%, and the minimum relative error is about 11.4%. Compared with the previous two methods, its calculation accuracy and stability are greatly improved, but are still not high enough to meet the actual needs of engineering.

(4) When the improved algorithm is used to calculate the time difference, the average relative error is about 2.5%, the maximum relative error is about 7.2%, and the minimum relative error is about 1.3%. Compared with the first three methods, its calculation accuracy and stability are significantly improved.

**Table 1.** Time difference calculated by four different methods.

| Sample number (n) | Initial peak method (ns) | generalized correlation method (ns) | energy accumulation method (ns) | improved algorithm (ns) |
|-------------------|--------------------------|-------------------------------------|---------------------------------|-------------------------|
| n=1               | 17.100                   | 16.450                              | 13.325                          | 10.225                  |
| n=10              | 16.050                   | 14.160                              | 12.140                          | 11.000                  |
| n=20              | 16.000                   | 16.950                              | 11.775                          | 11.075                  |
| n=30              | 23.300                   | 18.650                              | 14.095                          | 9.985                   |
| n=40              | 20.125                   | 15.300                              | 13.210                          | 10.900                  |
| n=50              | 17.145                   | 10.560                              | 12.415                          | 11.125                  |
| n=60              | 19.735                   | 15.250                              | 9.192                           | 10.835                  |
| n=70              | 43.210                   | 19.775                              | 12.375                          | 11.325                  |
| n=80              | 33.275                   | 28.102                              | 12.351                          | 11.100                  |
| n=90              | 10.950                   | 17.950                              | 13.010                          | 10.700                  |
4. Conclusions
This article mainly researches the time difference calculation issue in the time difference location method. An improved time difference estimation algorithm based on fourth-order cumulant is proposed. The accuracy and stability of the improved algorithm and the other three common algorithms are compared using the real measured signals of PD.

The results show that the average relative error of time difference calculated by improved algorithm is about 2.5%, is only one-seventh of the result calculated by the energy accumulation method which is the most accurate of the three common methods. And its maximum relative error is about 7.2%, the minimum relative error is about 1.3%. In conclusion, compared with the other three methods, the calculation accuracy and stability of the improved time difference estimation algorithm based on fourth-order cumulant are significantly improved, which can approximately meet the requirements for reliable location of PD sources.

Acknowledgements
All the work has been supported by the of Shandong Power Grid Co., Ltd. Science and Technology Project (520626200002).

References
[1] Wang Ke, Liao Ruijin, Zheng Shengxun, Gao Jun, Yang Lijun and Yan Jiaming 2013 Chaotic characteristic of partial discharge during electrical degradation of oil-impregnated paper High Voltage 39 3075-3081
[2] Wang Hui 2010 The evolution process of typical partial discharges in oil-paper insulation transformers Beijing: North China Electric Power University Master's Degree Thesis
[3] Gao Wensheng, Ding Dengwei, Liu Weidong and Feng Rui 2009 Location of PD by Searching in Space Using UHF Method High Voltage 35 2680-2684
[4] Tugnait J K 1993 Time delay estimation with unknown spatially correlated gaussian noise IEEE Transactions on Signal Processing 41 549-558
[5] Nikias C L and Mendel J M 1993 Signal Processing with Higher-Order Spectra IEEE Signal Processing Magazine 10 10-37
[6] Xu Zhongrong, Tang Ju, Zhang Xiaoxing and Sun Caixin 2008 Denoising of UHF partial discharge signal by complex wavelet transform for power transformer Electric Power Automation Equipment 28 27-32
[7] Goemans M X and Williamson D P 1995 Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming Journal of the Association for Computing Machinery 42 1115-1145
[8] Jones G, Willett P and Glen R C 1995 Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation Journal of Molecular Biology 245 43-53
[9] Willett P 1995 Genetic Algorithms in molecular recognition and design Trend. in Biote 13 516-521
[10] Kennedy J and Eberhart R 1995 Particle swarm optimization Proc. of IEEE Int. Conf. on Neural Networks Perth
[11] Chande S V and Sinha M 2011 Genetic Optimization for the Join Ordering Problem of Database Queries Annual IEEE India Conf. (INDICON) Hyderabad