J. Shanghai Jiao Tong Univ. (Sci.), 2019, 24(4): 459-470
https://doi.org/10.1007/s12204-019-2095-1

Logistic Regression Based Arc Fault Detection in Photovoltaic Systems Under Different Conditions

JIA Fan[1] (贾　帆), *LUO Liwen*[1*] (罗利文), *GAO Shiyue*[2] (高诗悦), *YE Jian*[3] (叶　健)
(1. School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; 2. University of Michigan - Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China; 3. Analog Devices, Inc., Shanghai 200021, China)

Abstract: This paper investigates direct current (DC) arc fault detection in photovoltaic system. In order to avoid the risk of fire ignition caused by the arc fault in the photovoltaic power supply, it is urgent to detect the DC arc fault in the photovoltaic system. Once an arc fault is detected, the power supply should be cut off immediately. A lot of field experiments are carried out to obtain the data of arc fault current of the photovoltaic system under different current conditions. Cable length, arc gap, and the effects of different sensors are tested. These three conditions are the most significant features of this paper. Four characteristic variables from both the time domain and the frequency domain are extracted to identify the arc fault. Then the logistic regression method in the field of artificial intelligence and machine learning is originally used to analyze the experimental results of arc fault in the photovoltaic system. The function between the probability of the arc fault and the change of the characteristic variables is obtained. After validating 80 groups of experimental data under different conditions, the accuracy rate of the arc fault detection by this algorithm is proved to reach 100%.
Key words: photovoltaic, arc fault, fast Fourier transform, logistic regression
CLC number: TM 914　　**Document code:** A

0 Introduction

Due to human's increasing demand on energy and increasing realization of environmental pollution, solar energy as sustainable energy is getting wider attention. Sustainable energy provides energy with manageable environmental effects as well as consuming nonrenewable resources like fossil fuels. A photovoltaic system converts solar energy into electricity without polluting the environment or emitting greenhouse gases. Direct current (DC) arc faults, however, is one of the major potential dangers of the photovoltaic systems. As solar cells, cables and other components are exposed in the open air, arc fault is likely to occur because of the corrosion of these components. Unlike the alternating current (AC) arc faults, there are no zero-crossing currents in DC arc faults. DC arc faults usually last for longer time and generate large amount of heat, and finally conflagrations break out. Thus, it is important to detect DC arc faults. In 2011, article 690.11 of the National Electrical Code required that DC photovoltaic arc-fault circuit protection devices must be used in solar photovoltaic electrical energy systems with rated voltage over 80 V[1]. In the same year, the Underwriters Laboratories developed UL1699B for testing and certification of DC arc-fault protection devices in photovoltaic systems[2].

Currently, the major methods used to detect DC arc faults are listed below. Fourier transform is widely used to analyze waves in the frequency domain. In Refs. [3-6], fast Fourier transform (FFT) is used to figure out the frequency domain spectrum. In Ref. [5], the sampling frequency is 250 kHz, as the frequency of the arc fault usually occurs between 40 and 100 kHz, and FFT is done for every 1 024 data points. After obtaining the frequency spectrum of the current using FFT, it can be judged whether an arc exists by comparing the amplitude with that of a normal state current or power. In Ref. [6], a DC arc fault detecting algorithm is provided. A threshold value is defined by calculating the mean and standard deviation of each frequency component of the normal current. When the magnitude of the frequency component exceeds the threshold value for several continuous sample periods, it will be detected as an arc fault.

However, FFT cannot provide enough information in the time domain. Short time Fourier transform (STFT) is similar to the FFT method, which provides a frequency domain signal representation, with an addition

Received date: 2018-10-10
∗**E-mail:** lwluo@sjtu.edu.cn

of time domain signal representation. In Ref. [7], STFT is introduced to do the time-frequency domain analysis. However, due to the limited window size, we have to balance between the time resolution and the frequency resolution[8]. If the window size is too small, it is difficult to figure out the frequency precisely. If the window size is too large, it cannot present the time domain information in the signal precisely. In Ref. [7], it is claimed that the calculation of STFT is relatively less than that of discrete wavelet transform (DWT), which will be introduced later.

Wavelet transform (WT) can resolve signal into smaller frequency bands and present the signal in both time and frequency domains. Compared to STFT, the signal is decomposed to diminish wavelets. Thus, the time domain information and the frequency domain information are obtained simultaneously. DWT provides signal analysis in multiple time resolutions[9]. It decomposes the waveform like a series of high-pass filters. In Ref. [9], DWT is used in real-time detection of photovoltaic arc signal. And wavelet packet decomposition (WPD) collects more information than DWT, as it works like a series of both high-pass and low-pass filters and collects all coefficients of the decomposition [10]. WPD is suitable to discover short abrupt changes. In Ref. [11], six levels of decomposition are done and arc faults are successfully detected.

Statistic-based method is based on recording statistic characters like the mean, the standard deviation and the root mean square (RMS) value of the input signal[12]. In Ref. [13], the range of current value in a certain time resolution is used to detect the arc fault. On the one hand, this method is quite efficient when detecting the arc fault at the instant it generates. On the other hand, the peaks of random disturbance can easily affect this detection method, resulting in an alarm when no arc fault happens. In Ref. [10], even if some algorithms get rid of the influence of random impulses, the maximum power point tracking (MPPT) operation from inverter may still influence the detection result.

Artificial intelligence based methods are also used in arc fault detection. In Ref. [14], artificial neural network (ANN) is used. Using a set of high frequency spectrum components of DC systems with different load configurations and different arcing conditions, the neural network is trained. The only problem is that the amount of data needed is quite large. Support vector machine (SVM) is used in Ref. [8]. By using SVM, the algorithm can set a threshold value automatically based on the current data with arc faults and normal current data.

1 Experimental Setup and Conditions

The purpose of this research is to collect DC arc fault data in different working conditions and extract the features of the data. Therefore, this experimental platform contains a real photovoltaic power generation system and the arc generator which is recommended in UL1699B. Because of this, the data of this experiment is more reliable and applicable. The overall schematic diagram of the system is shown in Fig. 1.

The solar panel used in this experiment is CEC6-60-255MA produced by CECEP. The main specifications of the solar panel are listed in Table 1.

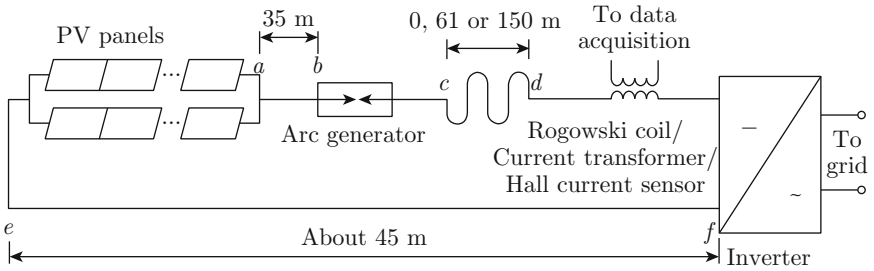


Fig. 1 Overall system schematic diagram

Table 1 Specifications of the solar panel

Parameter	Value	Parameter	Value
Rated maximum power ($P_{\max}$)/W	255	Temperature of normal operating cell (T_{NOC})/°C	47±2
Tolerance/W	0—5	Maximum system voltage/kV	1 (DC)
Voltage at $P_{\max}$(V_{mp})/V	30.9	Maximum series fuse rating/A	15
Current at $P_{\max}$(I_{mp})/A	8.26	Operating temperature/°C	−40—85
Open-circuit voltage (V_{oc})/V	38.1	Weight/kg	19.5
Short-circuit current (I_{sc})/A	8.95	Dimension/(m × m × m)	1.650 × 1.992 × 0.040

Note: application class is Class A; cell technology is Mono-Si.

a–b and e–f are the positive and negative bus cables respectively. The arc generator is shown in Fig. 2. The electrodes, one moveable and one stationary, are made of solid copper with a diameter of 6 mm. The electrodes can be separated by using a lateral adjustment means to position the moveable electrode to a desired gap.

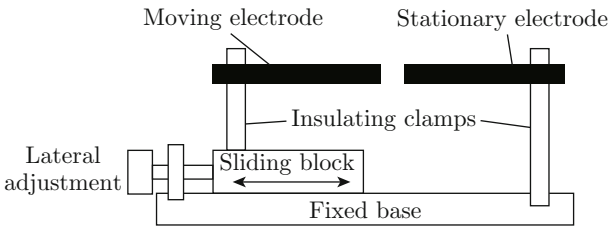


Fig. 2 Diagram of the arc generator

c–d is the additional cable, which is used to increase the distance between the detection device and the place the arc fault generates. The length of this cable is 0, 61 or 150 m, and arranged in S-shape in accordance with UL1699B. In order to measure the effect of different current sensors, three different sensors are used: Rogowski coil, current transformer (CT) and Hall current sensor. The Rogowski coil and CT are used to measure the AC component of the current. The Hall current sensor is used to make up the former two sensors' inability of measuring the DC component of the current, as it is able to measure the instantaneous current value, including both the AC and the DC component of the current. The main specifications of the three current sensors are listed in Tables 2, 3 and 4. The main specification of SUNGROW's photovoltaic inverter SG5K-D are listed in Table 5.

Table 2 Specifications of Rogowski coil

Parameter	Value
Accuracy class	0.2
Rated transformation constant/($\mu H \cdot Hz^{-1}$)	7.66
Primary-second isolation/kV	6
Secondary winding inductance/mH	1.14
Secondary winding resistance/Ω	37.6
Self-resonance frequency/kHz	200

Note: part number is PA3206NL.

The data acquisition card is the model USB-4600 produced by Smacq company with a resolution of 16 bits, 8 channels and a maximum sampling rate of 1 million samples per second of each channel.

To cover all kinds of actual working conditions in the experiment, which conditions will affect the current in an arc fault should be determined first. An arc fault current can be regarded as a superposition of two currents. One is the current that always exists in the circuit, that is, the normal working current, which will be called the background current in this paper. The other current will only occur when an arc is generated, which will be called the arc current in this paper.

Table 3 Specifications of CT

Parameter	Value
Turn ratio	1 : 200 : 200
DC resistance/Ω	$\leqslant 3.3$
Dielectric withstand	6 kV, 50 Hz
Insulation resistance/GΩ	$\geqslant 1$
Environment temperature/°C	−55—85

Note: insulation class is Class B.

Table 4 Specifications of Hall current sensor

Parameter	Value
Primary nominal RMS current/A	20
Reference voltage/V	2.5
Theoretical sensitivity/($mV \cdot A^{-1}$)	40
Accuracy/%	±1
Frequency bandwidth/kHz	400

Note: part number is HLSR 20-P.

Table 5 Specifications of photovoltaic inverter

Parameter	Value
Maximum input voltage/V	600
Rated input voltage/V	360
Amount of MPPT	2
Rated input current/A	20 (10/10)
Maximum input current/A	24 (12/12)
Rated output power/kW	5.0
Maximum output power/kW	5.5
Maximum output current/A	25
Rated output voltage/V	230 (AC)
Rated output frequency/Hz	50/60

First, the factors that affect the background current should be considered. In one electrical system, only three factors affect the working current: power, load and cable. The DC working current in the photovoltaic system is directly related to the output power of the photovoltaic panels, which is also the most important factor affecting the DC component of the background current. The output current of photovoltaic panels varies with the light intensities and temperatures, so it needs to be measured under different light conditions and temperatures. But photovoltaic panels are a large-scale facility on the roof, and cannot be placed in a container to set a certain light intensity and temperature artificially. Therefore, in order to experiment under different light intensities and

temperatures, we can only choose different dates and time for experiment. The light intensity and temperature are different on different days and in different time, such as morning, noon and evening. The output voltage-ampere characteristics of photovoltaic panels are different on different dates and in different time, so the different working currents are obtained. No matter how the weather conditions and other environmental conditions change, the final result is the different values of the working currents, which is the direct factor affecting the current in an arc fault. Weather condition, light intensity and temperature are indirect factors, which affect the arc characteristics indirectly by affecting the working current. Therefore, only the working current is distinguished in this paper, rather than the specific weather, time, light intensity and temperature. In this paper, the working current is divided into three classes. Class I is 2—4 A, class II is 6—8 A, and class III is 13—15A. Summarizing the measured current from photovoltaic panels on sunny days, the working current can reach class III at noon, class II at about 9∶00 or 15∶00, and class I at about 8∶00 or 16∶00.

Different models of inverters from different companies have different switching frequencies, internal circuits and control algorithms which may affect the noise in the current. However, due to the limited experimental condition, we only use one inverter in this experiment, which is the one produced by SUNGROW mentioned before. Different inverters can be used in future study. Because of the MPPT in the photovoltaic inverter, the DC current in the photovoltaic system will always be around the current at the maximum point. Therefore, the load at the AC side of the circuit will not influence the DC current.

The main influence of the cable to the current is the length of the cable. The longer the wire is, the larger the cable inductance and its capacitance to ground will be. This will lead to a more serious decay in the high frequency current component, making the cable a natural low-pass filter. Considering the UL1699B standard and the user's feedback in application, the lengths of the additional cables are 0, 61 or 150 m.

There are two possible factors, the shape of the electrode and the gap between the electrodes, affecting the arc current. In real cases, the shape of the electrodes where an arc generated is too complicated to imitate. So the electrodes are all replaced by copper rods in this experiment. The gap distances between the arc electrodes are 1, 2 or 3 mm.

The above discussions on the factors affecting the arc fault current can be summarized in Fig. 3.

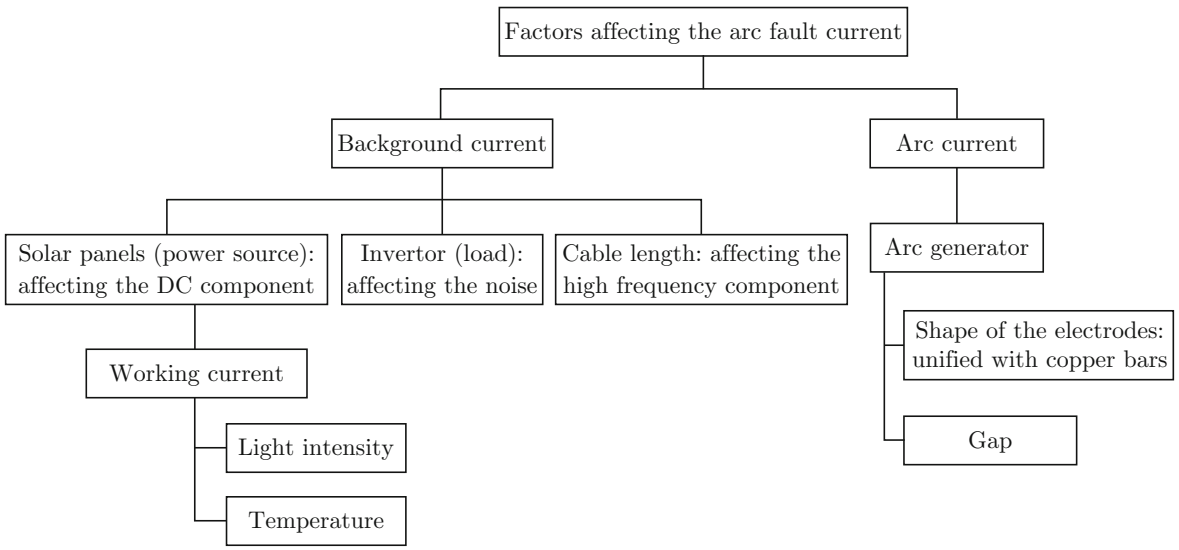


Fig. 3 Factors affecting the arc fault current

2 Experimental Data and the Extraction of Characteristic Variables

Because of the nonlinearity and limitation of frequency band of sensors and the insufficiency of sampling rate, the signals obtained by data acquisition (DAQ) cannot fully reflect the actual current waveforms. Therefore, a set of experiments with the shunt has been done to measure the actual current, and an oscilloscope has been used to collect the voltage across the shunt. In this way, the nonlinear and frequency band limitation of sensors is eliminated, and the sampling rate can reach 4×10^9 samples per second, thus the actual current can be seen when the arc fault happens. Figure 4 shows the waveforms of this set of experiment (t is the time).

The resistance of the shunt is 1 Ω, so the value of voltage across it is equal to the value of the current passing it, which is also the actual current in the system. It can be seen directly from Fig. 4:

(1) When there is an arc fault, there will be a drop in the DC current.

(2) At the instant the arc is generated, there are series of high frequency impulses. The amplitude of the impulses can reach to several amperes, and the pulse width is about ten nanoseconds to tens of nanoseconds.

(3) Figure 4(a) shows that after the arc generates, the current when the arc steadily combusts is quite similar to that of the normal current, except that the burr is more obvious, which is shown intuitively that the waveform is "thicker".

(4) From Figs. 4(a) and 4(b), frequency components at 100 Hz and 40 kHz can be obviously observed. The explanation of these two frequency components will be discussed in detail in the following sections.

In the different conditions introduced before, plenty of experiments are carried out and dozens of experimental data are obtained. All conclusions in this paper are based on the whole data. But due to the limit of the length of the paper, it's hard to present more figures. So, experimental diagrams of only one typical condition are shown below to explain the regularities sought out from the whole data.

Figure 5 shows the typical waveforms of voltage and current before and after an arc fault. The whole process can be divided into five stages:

(1) Stage 1 is the stage of the normal situation.

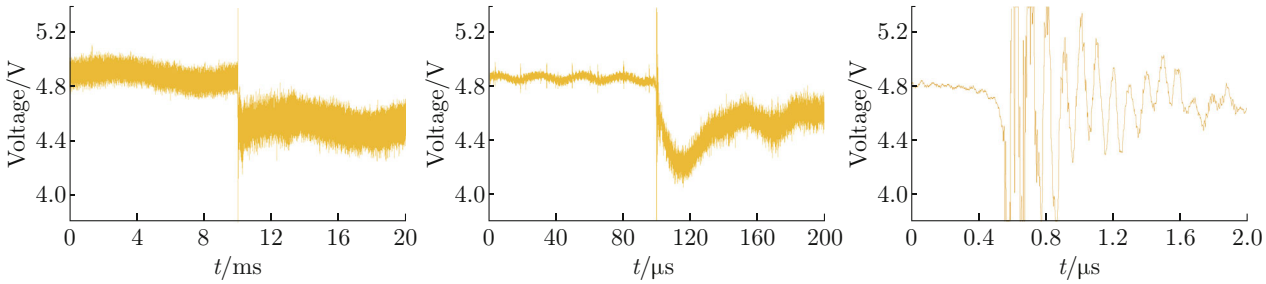
(2) Stage 2 is the stage at the moment of the arc occurrence.

(3) Stage 3 is the stable combustion stage of the arc.

(4) Stage 4 is the stage at the moment of the arc extinction.

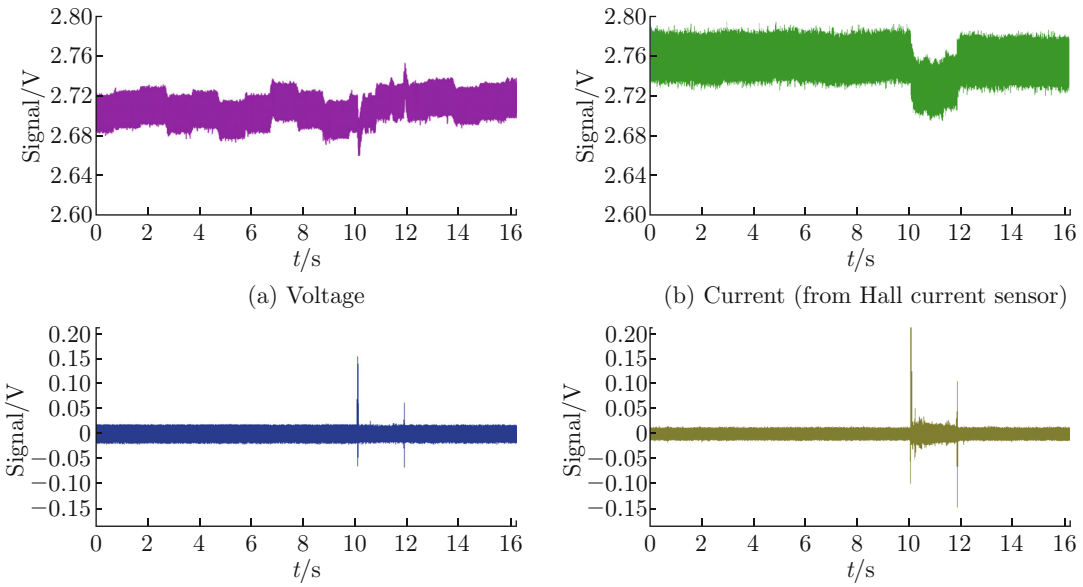
(5) Stage 5 is the stage returned to normal.

Obviously, only the first three stages can be used to distinguish the arc fault. The sampling rate is 5.12×10^5 samples per second (the following experimental data of this paper are all obtained under this sampling rate). In Fig. 5, the arc fault generated at around the 10th second and ends at around the 12th second. Figure 5(a) is the signal of the voltage on the DC side of the



(a) 2 ms/div (b) 20 μs/div (c) 200 ns/div

Fig. 4 Voltage waveform across the sampling resistor under different time scales



(a) Voltage (b) Current (from Hall current sensor)

(c) Current (from Rogowski coil) (d) Current (from CT)

Fig. 5 Typical waveforms of voltage and current before and after an arc fault

inverter through a voltage divider and a linear amplification. Figure 5(b) is the output signal of the Hall current sensor. Figure 5(c) is the output signal of the Rogowski coil. Figure 5(d) is the output signal of CT. Figure 5(a) represents the actual voltage of the inverter DC side. Figure 5(b) represents the actual current in the photovoltaic system. Figures 5(c) and 5(d) represent the AC component of the current.

Comparing Fig. 5 to Fig. 4, we can find that the key features in Fig. 4, like the drop of DC current and the high frequency impulse, are also shown in Fig. 5. This proves the feasibility of using current sensors to detect arc faults. From Fig. 5(a), it can be seen that the change of voltage is disorder, and step changes exist in the normal working state. The difference between the fault and the normal stage is not explicit, so it is not suitable to use it as the characteristic variable to detect whether an arc occurs. By comparing Figs. 5(c) and 5(d), it can be seen that when CT is used as the sensor, the difference of burr between fault stage and normal stage is more explicit. Therefore, in the time domain, it is more applicable to use CT as a sensor, rather than using Rogowski coil.

Figure 6 shows the spectrums obtained after FFT for the experimental data above (the first point representing the DC component is skipped to make the result clearer), and the number of points used in FFT is 4 096.

As can be seen from Fig. 6, there is a significant peak at the frequency of 40 kHz, which is the main switching frequency of the inverter, and peaks at the frequency of its harmonics. This is in accordance with the conclusion obtained from Fig. 4. The frequency component of 100 Hz mentioned in Fig. 4 can be seen in Fig. 6, at the very first point, which represents the frequency of 125 Hz. However, due to the problem of spectrum resolution, it is not clear with the DC component. The frequency of 100 Hz is the result of the MPPT algorithm of the inverter, which produces periodic fluctuation of current for tracking the maximum power point. In addition, there are some smaller peaks at frequencies of 8, 24 kHz, etc., which are caused by the switching frequencies of the Boost, Buck and other DC-DC switching power supplies inside the inverter to power the controlling circuits.

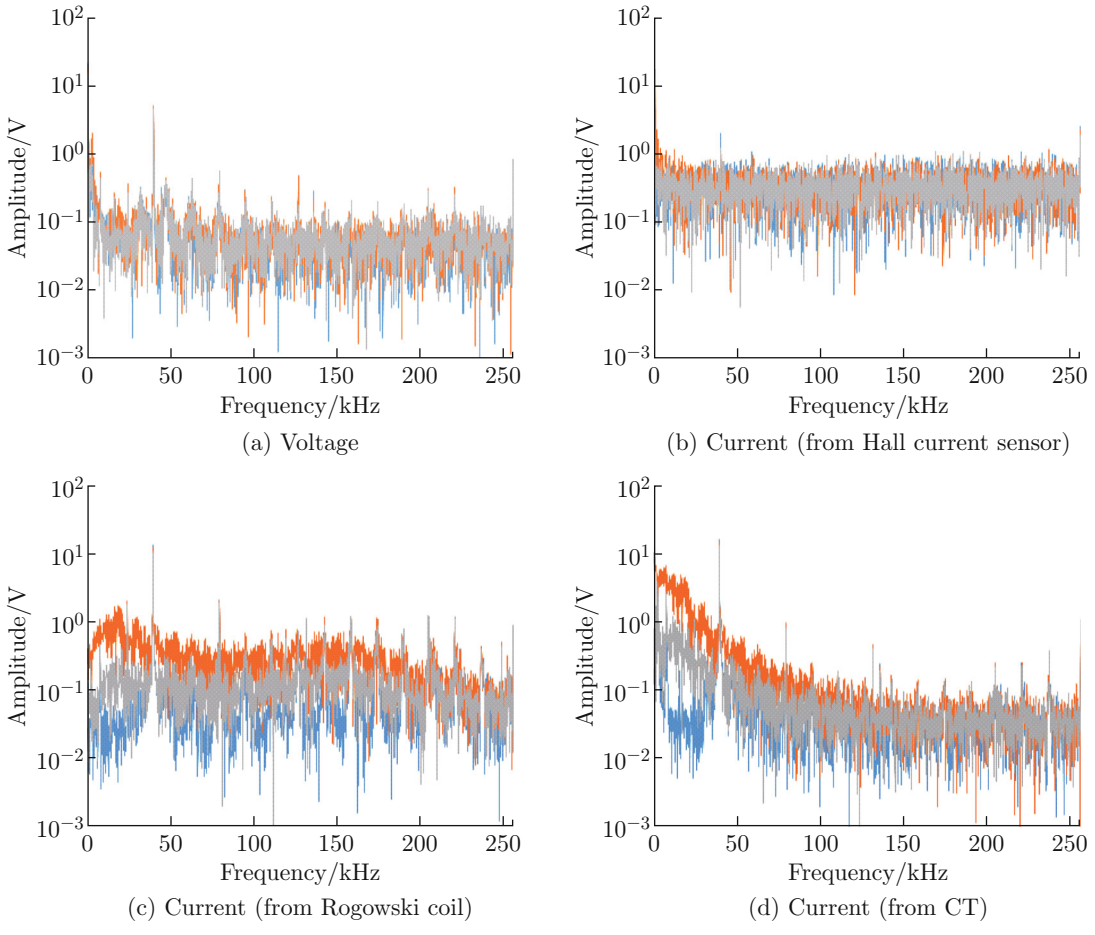


Fig. 6 Typical frequency spectrums of voltage and current

In addition, the output of Rogowski coil is the voltage u which is proportional to the derivative of primary current rather than the primary current itself. Mathematically, secondary side opening voltage is given by

$$u = M\frac{\mathrm{d}i}{\mathrm{d}t}, \tag{1}$$

where, u is the open-circuit voltage on the secondary side of the Rogowski coil; i is the primary side current; M is the mutual inductance coefficient. For a signal of a certain frequency f, the amplitude after derivation is $2\pi f$ times the original amplitude. For example, a current signal at 100 kHz with an amplitude of 1 in the primary side of a Rogowski coil, will cause an open-circuit voltage signal on the secondary side with an amplitude of $10^5 \times 2\pi M$. So the amplitude of the spectrum is magnified, and the higher the frequency, the larger the magnification. However, this problem does not exist for CT. The current ratio between primary and secondary is maintained near the ratio of the CT, as long as it is in its frequency band range. It can be considered that the secondary side current is directly proportional to the primary side current. It can be seen by comparing Figs. 6(c) and 6(d) that in low frequency (below 20 kHz), the spectrum of CT is significantly decreasing with the increase of frequency, but the spectrum of Rogowski coil is basically flat. In the high frequency part (above 100 kHz), the spectrum of the Rogowski coil has a slight upward trend and overruns the spectrum of CT. This is an evidence of the previous analysis. Therefore, CT is more suitable for arc fault detection than Rogowski coil from the frequency domain perspective.

It can be seen from Figs. 6(a) and 6(b) that there is no obvious difference between the frequency spectrums of the voltage signal and the output signal of the Hall current sensor before and during the arc fault. Therefore, both signals can be excluded from the signals that need to do spectrum analysis.

It can be seen from Figs. 6(c) and 6(d) that the frequency component below 100 kHz at the moment of the arc generation and the stable combustion stage of the arc, especially at the moment of the arc generation, will be obviously higher than the normal condition. This feature is very helpful to identify the arc immediately, and the circuit is cut off when the arc fault occurs, so as to avoid the fire threat caused by the further combustion of the arc.

The time constant of DC current drop caused by arc fault is about tens of microseconds. As mentioned earlier, due to the role of MPPT in the inverter, the sudden change of AC load does not affect the DC current. The noise frequency brought by the inverter is between thousands of hertz and tens of kilohertz. Other factors that cause the DC current change include the influence of MPPT and the change of environment. The period of MPPT is 10 ms, and the time constants of environmental changes are longer than 1 s. Therefore, by choosing the right length of time window, the arcing can be identified by the sudden drop of the DC current. The length of time window selected in this paper is 10ms. The average current caused by the switching noise of the inverter and the influence of MPPT is about zero in this time window, and the change of the environment can be ignored in this time scale.

To sum up, combining the effects of different sensors in time and frequency domain, a Hall current sensor and CT are recommended to detect the DC current drop and arcing current respectively. The following analysis in the time domain takes the data from Hall current sensor and CT as an example. The frequency domain analysis takes only data from CT as an example.

Based on the above analysis and experimental results, the following four characteristic variables are selected.

(1) Sum of logarithms in frequency domain a. The modulus of the data recorded by CT after a 4 096 points FFT, which is the amplitude in the spectrum, is marked as $\{F_k\}, k = 1, 2, \cdots, 4\,096$. As the sampling rate of 5.12×10^5 samples per second, points $\{F_k | k = 17, 18, \cdots, 721\}$ are in correspondence to frequency 2 to 90kHz. Sorting these 705 points from the largest to the smallest, we can obtain $\{F'_j\}, j = 1, 2, \cdots, 705$. The Sum of logarithms is calculated by

$$a = \sum_{j=51}^{685} \lg F'_j. \tag{2}$$

This characteristic variable can reflect the amplitude of the frequency component of arc fault current between 2 and 90 kHz. By eliminating 50 maximum and 20 minimum values, the switching noise from the inverter is eliminated, and the interference caused by the frequencies with extremely low or high amplitude is eliminated.

(2) The time domain range b. The data acquired by CT in 10 ms is recorded as $\{u_n\}, n = 1, 2, \cdots, 5\,120$. The range is the difference between the maximum and minimum values of the time domain data:

$$b = \max\{u_n\} - \min\{u_n\}. \tag{3}$$

This characteristic variable can reflect the feature of the high frequency impulse at the instant when an arc occurs.

(3) Sum of the absolute values in the time domain c. The sum of the absolute values in the time domain is given by

$$c = \sum_{n=1}^{5\,120} |u_n|. \tag{4}$$

This characteristic variable reflects the increase in the burr of the current. In order to reduce the amount of

calculation, the variance and the standard deviation of the data are not used here.

(4) The DC current d. The data acquired by Hall current sensor in 10 ms is recorded as $\{i_n\}$. According to the transfer function of the Hall current sensor, the primary side current is $(i_n - 2.515)/0.04$. The DC current is the mean value of the data in 10 ms, which is calculated by

$$d = \frac{1}{5\,120} \sum_{n=1}^{5\,120} \frac{i_n - 2.515}{0.04}. \tag{5}$$

This characteristic variable can reflect the drop in DC current when an arc occurs.

In order to facilitate the representation, a characteristic vector $\boldsymbol{x}$ is used to express a set of characteristic variables:

$$\boldsymbol{x} = [a \quad b \quad c \quad d]^{\mathrm{T}}. \tag{6}$$

The four characteristic variables under 27 typical conditions are summarized in Tables 6 and 7. It can be seen from Tables 6 and 7 that each characteristic value shows a significant upward trend with the increase of

Table 6 Typical experimental data for the characteristic variables a and b

Experimental condition			a			b		
Current level	Cable length/m	Gap/mm	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
Class I	0	1	−900.4	181.3	−365.8	0.020	0.821	0.079
		2	N/A	N/A	N/A	N/A	N/A	N/A
		3	N/A	N/A	N/A	N/A	N/A	N/A
	61	1	−906.9	−86.9	−374.2	0.016	0.601	0.058
		2	N/A	N/A	N/A	N/A	N/A	N/A
		3	N/A	N/A	N/A	N/A	N/A	N/A
	150	1	−884.4	−55.9	−625.9	0.015	0.444	0.059
		2	−914.3	16.7	−673.7	0.014	1.029	0.025
		3	−893.9	−95.6	−673.9	0.013	0.610	0.039
Class II	0	1	−814.9	−101.6	−531.2	0.025	0.372	0.04
		2	−763.4	121.7	−522.7	0.030	1.942	0.062
		3	−769.7	268.3	−523.9	0.028	2.470	0.113
	61	1	−788.8	83.6	−689.8	0.021	1.135	0.027
		2	−788.1	−28.9	−674.6	0.021	0.718	0.047
		3	−784.3	−21.2	−481.2	0.023	0.657	0.085
	150	1	−789.2	−199.2	−698.5	0.020	0.728	0.041
		2	−840.9	73.2	−608.7	0.018	1.330	0.049
		3	−810.5	91.9	−774.5	0.019	1.133	0.026
Class III	0	1	−669.3	27.0	−473.1	0.040	0.577	0.062
		2	−648.1	184.0	−497.3	0.045	1.553	0.083
		3	−622.9	131.2	−345.4	0.045	2.202	0.105
	61	1	−698.6	−37.8	−536.6	0.038	1.049	0.069
		2	−705.2	222.4	−601.8	0.032	5.371	0.043
		3	−693.9	51.3	−627.3	0.036	0.810	0.058
	150	1	−675.6	−93.5	−585.6	0.035	1.055	0.053
		2	−692.7	−144.4	−589.9	0.032	1.224	0.054
		3	−699.4	−118.2	−621.6	0.033	1.301	0.071

Note: the current of class I is too low that the arc of 2 and 3 mm cannot be generated in some conditions; therefore, the sets of data are not applicable (N/A).

Table 7 Typical experimental data for the characteristic variables c and d

Experimental condition			c			d		
Current level	Cable length/m	Gap/mm	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
Class I	0	1	19.02	150.48	33.29	3.090	2.677	1.972
		2	N/A	N/A	N/A	N/A	N/A	N/A
		3	N/A	N/A	N/A	N/A	N/A	N/A
	61	1	12.81	62.95	25.92	2.578	2.395	1.734
		2	N/A	N/A	N/A	N/A	N/A	N/A
		3	N/A	N/A	N/A	N/A	N/A	N/A
	150	1	8.54	148.45	23.35	3.402	2.691	2.821
		2	8.41	136.91	10.88	3.521	2.962	2.748
		3	8.65	196.51	18.03	3.713	2.590	2.691
Class II	0	1	19.32	58.37	24.71	7.232	6.792	6.219
		2	19.98	239.63	30.00	7.340	5.352	5.907
		3	20.23	512.42	35.70	7.451	4.873	5.384
	61	1	13.61	235.14	14.39	7.788	6.861	6.443
		2	13.27	236.95	22.98	7.828	5.976	6.827
		3	13.48	242.29	45.81	7.768	5.642	5.853
	150	1	10.01	138.36	17.35	6.908	6.014	6.105
		2	9.30	244.79	25.12	6.958	5.282	5.850
		3	9.40	449.24	11.56	6.952	4.578	5.491
Class III	0	1	30.07	138.47	32.69	13.992	13.118	12.159
		2	29.92	330.46	39.75	13.966	11.450	11.478
		3	29.27	487.23	45.85	14.012	9.656	11.087
	61	1	19.48	223.72	33.61	13.100	11.303	11.208
		2	18.50	915.94	20.38	12.910	8.838	10.448
		3	18.61	404.90	22.91	13.162	10.259	10.714
	150	1	15.99	223.27	20.44	14.810	12.759	12.524
		2	14.29	203.07	25.19	14.105	11.140	11.397
		3	15.50	320.42	31.64	13.686	9.503	10.458

Note: the current of class I is too low that the arc of 2 and 3 mm cannot be generated in some conditions; therefore, the sets of data are not applicable (N/A).

of the current. This is because the increase of the current results in an overall enlargement of the time-domain waveform, and each characteristic quantity will naturally increase. Except for the DC drop, the other characteristic variables are decreasing with the increase of cable length. This is consistent with the analysis of the previous analysis that the increasing cable length will make the attenuation of high-frequency signals more severe. The change of the arc gap has no obvious effect on the characteristic variables.

3 Detection Algorithm

From Tables 6 and 7, it can be seen that under the same conditions, when an arc occurs, each characteristic variable will increase (or decrease, for DC current) in different degrees. It is easy to misjudge whether there is an arc fault, relying on only a single set of characteristic values, due to the fluctuation of the characteristic variables caused by the change of the environment. For example, when the working current is high (such as at

12∶00), the sum of logarithms in frequency domain in the normal condition may have exceeded that in the fault condition when the working current is low (such as at 17∶00). For another example, at 12∶00, even if the DC current drops due to an arc fault, the current is still greater than the normal current at 17∶00. Therefore, relying on only one set of characteristic variables to identify the arc fault, the accuracy of the arc fault recognition is bound to be low. So, the change between two sets of characteristic variables is used to identify arc faults in this paper. The arcing detector needs to update the normal characteristic vectors as a reference vector $\boldsymbol{x}_0$ periodically, such as every 50 ms. During this period of time, the characteristic vector $\boldsymbol{x}$ obtained from each sampling are made different from this reference vector, and the variation of the characteristic vector is calculated as $\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_0$. Whether the arc fault occurred is determined according to the value of $\Delta\boldsymbol{x}$.

In order to judge whether an arc occurs according to the change of characteristic vectors, this paper uses the logistic regression method in the artificial intelligence machine learning method. The logistic regression method is a machine learning method used to determine which class, whether 0 or 1, the object belongs to according to some characteristic values of the object. For example, doctors judge whether a patient has a disease based on some of his physical indicators. Arc fault identification is also judged by a number of characteristic variables and the result belongs to only two classes: yes (1) or no (0). Therefore, the logistic regression method is especially suitable for arc fault detection in this paper. And the logistic regression can also give the probability P of the object as 1, which is also very beneficial to the detection of the arc fault, because further actions can be determined based on how large the probability is. Only conclusion is presented in this paper, for more information on the principles and inferences of logistic regression, please refer to Ref. [15]. The calculation of P is given by

$$P = \frac{1}{1 + \mathrm{e}^{-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{X}}}, \tag{7}$$

where $\boldsymbol{w}$ is the parameter that needs to be obtained by machine learning using the logistic regression, and $\boldsymbol{X} = \begin{bmatrix} 1 \\ \Delta\boldsymbol{x} \end{bmatrix}$.

When $P > 0.5$, it can be concluded that an arc fault occurs. When $P < 0.5$, it can be concluded that there are no arc fault in the system. The closer P is to 1, the possibility of having an arc is higher. The closer P is to 0, the possibility of not having an arc is higher.

Using the logistic regression algorithm in MATLAB to study 60 sets of experimental data under different conditions, parameter $\boldsymbol{w}$ is obtained:

$$\boldsymbol{w} = \begin{bmatrix} -149.276\,8 \\ 0.266\,7 \\ 56.301\,5 \\ -0.904\,3 \\ -1\,492.7 \end{bmatrix}. \tag{8}$$

80 sets of data are used to validate the model given by Eq. (7) where $\boldsymbol{w}$ is given by Eq. (8), the validating result is shown in Table 8. It can be seen that for these 80 sets of data, the accuracy rate of arc fault recognition is 100%. The divergence is very large as P is extremely close to 1 when an arc occurs, and P is extremely close to 0 when there is no arc in the system.

The flowchart of arc fault detection algorithm is shown in Fig. 7. Firstly, a characteristic vector is collected under normal condition as the reference vector $\boldsymbol{x}_0$. Then the data are continuously collected and the characteristic vector $\boldsymbol{x}$ is computed. The difference of $\boldsymbol{x}$ and $\boldsymbol{x}_0$ is calculated, and then $\Delta\boldsymbol{x}$ and $\boldsymbol{X}$ is computed. Substituting $\boldsymbol{X}$ into Eq. (7), the probability of arc fault P will be obtained. If $P > 0.5$, it indicates that the arc fault may have occurred, then the breaker should cut off the circuit. If $P < 0.5$, it indicates that there is no arc fault, and if it's time to update the reference vector, then the $\boldsymbol{x}_0$ will be replaced by $\boldsymbol{x}$, otherwise, the data are acquired and the characteristic vector $\boldsymbol{x}$ is calculated again.

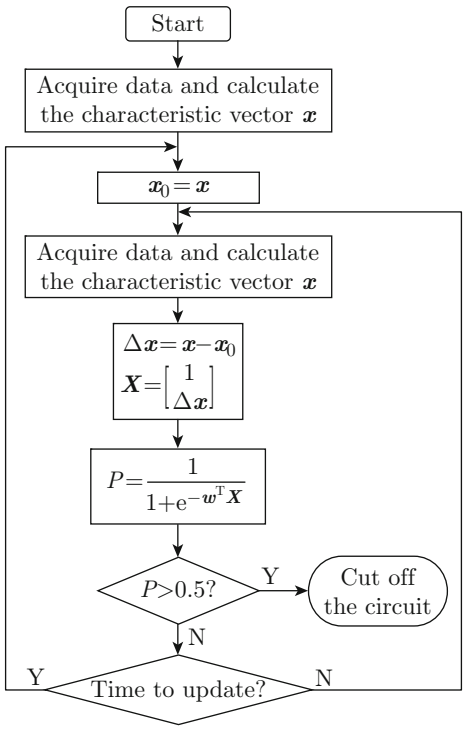


Fig. 7 Flowchart of the arc fault detection algorithm

Table 8 Validating results of logistic regression

Δa	Δb	Δc	Δd	Actual value	P	Δa	Δb	Δc	Δd	Actual value	P
1 081.7	0.801	131.46	−0.134	1	1	800.5	2.162	457.16	−0.310	1	1
534.5	0.059	14.26	−0.362	1	1	196.2	0.022	2.62	−0.131	1	1
820.0	0.585	50.13	−0.071	1	1	172.1	0.043	9.69	−0.180	1	1
532.7	0.042	13.10	−0.328	1	1	323.9	0.065	15.78	−0.208	1	1
−29.8	−0.001	−0.13	0.035	0	1.04×10^{-91}	−6.6	−0.006	−0.98	−0.015	0	1.21×10^{-56}
−9.5	−0.002	0.11	0.092	0	4.10×10^{-126}	4.7	−0.002	−0.87	0.005	0	8.82×10^{-68}
828.6	0.429	139.91	−0.209	1	1	660.8	1.011	204.23	−0.137	1	1
901.1	1.014	128.38	−0.129	1	1	921.0	5.333	896.46	−0.325	1	1
788.9	0.595	187.97	−0.239	1	1	750.0	0.772	385.42	−0.217	1	1
258.5	0.044	14.81	−0.171	1	1	162.1	0.031	14.13	−0.144	1	1
210.8	0.010	2.35	−0.192	1	1	96.8	0.005	0.90	−0.202	1	1
210.5	0.024	9.49	−0.209	1	1	71.3	0.020	3.43	−0.182	1	1
51.5	0.005	0.66	0.015	0	1.84×10^{-69}	−17.1	−0.003	−1.69	−0.048	0	4.49×10^{-36}
45.2	0.003	0.90	0.030	0	3.30×10^{-80}	−23.8	−0.002	−0.49	−0.076	0	5.97×10^{-19}
713.3	0.347	39.05	−0.061	1	1	582.1	1.020	207.28	−0.138	1	1
936.6	1.917	220.30	−0.260	1	1	531.2	1.189	187.08	−0.248	1	1
1083.2	2.445	493.09	−0.326	1	1	557.5	1.266	304.43	−0.358	1	1
283.7	0.024	5.39	−0.140	1	1	90.0	0.018	4.46	−0.154	1	1
292.2	0.037	10.68	−0.183	1	1	85.7	0.019	9.21	−0.230	1	1
291.0	0.088	16.38	−0.256	1	1	54.0	0.036	15.65	−0.294	1	1
0.7	0.000	−0.34	0.005	0	1.12×10^{-68}	555.9	0.374	159.88	−0.228	1	1
4.5	0.002	−0.13	−0.003	0	2.94×10^{-63}	86.4	0.021	11.11	−0.260	1	1
872.4	1.114	221.53	−0.119	1	1	−5.4	0.001	0.30	0.001	0	4.43×10^{-67}
759.9	0.697	223.35	−0.233	1	1	578.0	0.399	28.40	−0.029	1	1
767.6	0.636	228.68	−0.276	1	1	162.3	0.010	−0.15	−0.102	1	1
99.0	0.006	0.78	−0.173	1	1	−9.2	0.309	13.14	0.006	0	4.51×10^{-68}
114.3	0.026	9.37	−0.123	1	1	563.1	0.013	−0.16	−0.033	1	1
307.6	0.064	32.20	−0.249	1	1	237.5	0.020	6.57	−0.136	1	1
−51.7	−0.002	−0.72	0.007	0	4.63×10^{-76}	−21.7	−0.001	−0.01	−0.006	0	2.28×10^{-64}
−21.3	−0.001	−0.62	0.006	0	5.90×10^{-72}	635.5	0.144	10.72	−0.066	1	1
590.0	0.708	128.35	−0.129	1	1	494.3	0.073	7.47	−0.159	1	1
862.4	1.310	234.77	−0.235	1	1	13.9	0.003	-0.70	−0.006	0	7.37×10^{-60}
881.1	1.113	439.22	−0.337	1	1	581.7	0.306	34.45	−0.088	1	1
90.7	0.021	7.33	−0.116	1	1	272.1	0.025	5.00	−0.182	1	1
180.5	0.029	15.11	−0.153	1	1	−6.1	0.000	0.00	−0.001	0	8.40×10^{-66}
14.7	0.006	1.54	−0.205	1	1	588.4	0.345	34.34	−0.075	1	1
21.2	0.005	−0.15	−0.002	0	1.07×10^{-61}	314.4	0.037	10.00	−0.139	1	1
46.4	0.005	−0.80	0.001	0	1.08×10^{-60}	1.4	−0.001	0.22	0.004	0	5.02×10^{-68}
696.4	0.537	108.40	−0.062	1	1	485.5	0.258	18.95	−0.078	1	1
853.3	1.513	300.39	−0.182	1	1	344.3	0.038	13.48	−0.117	1	1

Note: actual value 1 represents for arc fault, and 0 represents for none arc fault.

4 Conclusion

In this paper, the experimental platform contains a real photovoltaic power generation system and the arc generator which is recommended in UL1699B. A large number of experiments have been carried out under different conditions of current, temperature, cable length and arc gap. The current data with and without arc faults are obtained under different mentioned conditions. The effects of different sensors like Rogowski coil, Hall and CT are tested. The range, the sum of absolute value, the DC drop in the time domain and the sum in logarithm in the frequency domain are extracted as the characteristic variables in order to detect the DC arc fault. By using the algorithm of logistic regression in the field of artificial intelligence, the relationship between the change of characteristic variables and the occurrence of arc is studied. The probability

equation of the arc fault's existence by the change in the four characteristic variables is obtained. The whole process of the arc fault detection algorithm is given. 80 sets of experimental data under different conditions are used to validate the algorithm, and the accuracy of this detection algorithm is very high, reaching the detection rate of 100%. This proves that the arc fault detection method proposed in this paper is feasible and excellent.

There are still some shortcomings in the arc fault detection method provided in this paper. The parameters in the probability equation obtained in this paper are only applicable to the sensor model and DAQ used in this experiment. If another set of hardware detection device is used, the data obtained from the new device are needed to obtain the parameters in the probability equation that is suitable for the new set of hardware, using the logistic regression algorithm. In addition, due to the limitations of the experimental conditions, this paper does not include experiments under different types of inverters. Therefore, the performance of the arc fault detection algorithm in this article under different inverters' noises cannot be determined. This is the key content of the future work.

References

[1] National Fire Protection Association. National electrical code: NFPA 70 [S]. Quincy, MA, USA: NFPA, 2011.

[2] Underwriters Laboratories Inc. Outline of investigation for photovoltaic (PV) DC arc-fault circuit protection: UL 1699B [S]. Northbrook, IL, USA: UL, 2013.

[3] NOVAK B. Implementing arc detection in solar applications: achieving compliance with the new UL 1699B standard [EB/OL]. (2012-07-10) [2018-10-10]. http://www.ti.com/lit/wp/spry209/spry209.pdf.

[4] XIONG Q, JI S C, ZHU L Y, et al. A novel DC arc fault detection method based on electromagnetic radiation signal [J]. *IEEE Transactions on Plasma Science*, 2017, **45**(3): 472-478.

[5] XIA K, HE Z H, YUAN Y, et al. An arc fault detection system for the household photovoltaic inverter according to the DC bus currents [C]//*18th International Conference on Electrical Machines and Systems*. Pattaya, Thailand: IEEE, 2015: 1687-1690.

[6] CHAE S, PARK J, OH S. Series DC arc fault detection algorithm for DC microgrids using relative magnitude comparison [J]. *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 2016, **4**(4): 1270-1278.

[7] CHEN S L, LI X W, XIONG J Y. Series arc fault identification for photovoltaic system based on time-domain and time-frequency-domain analysis [J]. *IEEE Journal of Photovoltaics*, 2017, **7**(4): 1105-1114.

[8] WANG Z, BALOG R S. Arc fault and flash detection in photovoltaic systems using wavelet transform and support vector machines [C]//*43rd IEEE Photovoltaic Specialists Conference*. Portland, OR, USA: IEEE, 2016: 3275-3280.

[9] ZHU H Z, WANG Z, BALOG R S. Real time arc fault detection in PV systems using wavelet decomposition [C]//*43rd IEEE Photovoltaic Specialists Conference*. Portland, OR, USA: IEEE, 2016: 1761-1766.

[10] LU S B, PHUNG B T, ZHANG D M. A comprehensive review on DC arc faults and their diagnosis methods in photovoltaic systems [J]. *Renewable and Sustainable Energy Reviews*, 2018, **89**: 88-98.

[11] GUO Y M, WANG L, WU Z Q, et al. Wavelet packet analysis applied in detection of low-voltage DC arc fault [C]// *4th IEEE Conference on Industrial Electronics and Applications*. Xi'an, China: IEEE, 2009: 4013-4016.

[12] MENG Z, WANG L, SUN Q G. The characteristics of DC arc faults current [C]//*15th European Conference on Power Electronics and Applications*. Lille, France: IEEE, 2013: 1-9.

[13] YAO X, HERRERA L, JI S C, et al. Characteristic study and time-domain discrete-wavelet-transform based hybrid detection of series DC arc faults [J]. *IEEE Transactions on Power Electronics*, 2014, **29**(6): 3103-3115.

[14] MOMOH J A, BUTTON R. Design and analysis of aerospace DC arcing faults using fast Fourier transformation and artificial neural network [C]//*IEEE Power Engineering Society General Meeting*. Toronto, Canada: IEEE, 2003: 788-793.

[15] ROGERS S, GIROLAMI M. A first course in machine learning [M]. Boca Raton, FL, USA: Chapman and Hall/CRC, 2011.