

A fault diagnosis method for photovoltaic module current mismatch based on numerical analysis and statistics

Zhixiang Zhang^a, Mingyao Ma^{a,*}, Hai Wang^b, Haisong Wang^a, Wenting Ma^a, Xing Zhang^a

^a National and Local Joint Engineering Laboratory for Renewable Energy Access to Grid Technology, Hefei University of Technology, Hefei 230009, China

^b Discipline of Engineering and Energy, Murdoch University, Murdoch, WA 6150, Australia



ARTICLE INFO

Keywords:

Photovoltaic (PV) module
Current mismatch
I-V characteristic
Data statistics
Curve division
Fault diagnosis

ABSTRACT

Photovoltaic (PV) module faults will not only reduce the power generation efficiency of PV modules, but also cause a series of safety problems. As the most common fault type, current mismatched fault leads to the decrease of the output current of the PV module resulting in a step in the I-V characteristic curves and multiple peaks in the P-V curves, such that the output power of the PV modules will be greatly affected. This paper focuses on current mismatched faults caused by partial shading, hot spot and crack through the investigation of faulty PV modules in actual PV power plants. The I-V characteristics of PV modules with current mismatch type faults are tested, and their fault characteristics are extracted. Not only can the current mismatch make the I-V characteristic curve of the module a step, but also the cause of the I-V curve of each current mismatched fault is analyzed in combination with the reverse bias model of the PV cell. In order to further decouple the features of different faults in the I-V curve step, a numerical analysis and statistical method is proposed for diagnosing PV module mismatch faults, which divides the I-V curve into a high voltage area and a low voltage area. The detection line composed of key points in the low voltage area is used to detect mismatch, while the current drop and linear regression fitting of the step data are used to identify the specific fault types in the high voltage area. Combined with the actual I-V data of PV modules under different working conditions, four case studies involving various types of faults are demonstrated to show that the proposed fault diagnostic method exhibits strong discriminating power and adaptability, and high practical application value.

1. Introduction

Among renewable energy sources, photovoltaic (PV) power generation with the fastest development rate is experiencing the fastest industrialization and the largest scale in the industry after wind power generation (REN21, 2019). As PV modules are important components of PV systems, their reliability is a key factor to ensure the performance of the entire system. Although the service life of PV modules is 20 to 25 years in theory, their life will be seriously affected by various fault problems due to the outdoor installation of PV modules and continuous exposure to harsh environmental conditions (Dhanup et al., 2018; Asma et al., 2018). As seen from the results of a study having a two-year monitoring of multiple PV systems, the average annual energy loss due to various faults is approximately 18.9 % (Firth et al., 2010). It is known that common causes of module failure are shading, hot spot, potential-induced degradation (PID), diode failure, crack, etc (Shifeng Deng et al., 2017). Among these faults, the partial shading, hot spot and

crack faults lead to the PV module current mismatch, which results in the rising temperature of the PV module. If these faults are not eliminated in time, they will directly affect the efficiency of the PV systems, and may even lead to devastating disasters such as fires. Thus, as long as the on-line fault diagnosis of PV modules can be realized, not only can the cost of manual maintenance be reduced, but also the power generation efficiency of PV systems will be improved and serious consequences caused by faults will be avoided resulting in considerable social benefits.

Most scholars are committed to large-scale PV systems using centralized inverters or series inverters. This paper proposes an effective online fault detection method for medium and small PV systems connected by power optimizers which are especially popular for buildings and households. The small and medium power distributed systems equip each PV module with an optimizer with functions such as DC-DC and maximum power point tracking (MPPT) and I-V curve scanning. When one PV module in the system is not working properly, only this PV

* Corresponding author.

E-mail address: miyama@hfut.edu.cn (M. Ma).

module will be affected and other PV modules will work normally, which further improves the efficiency of the system. Note that the accurate detection and replacement of faulty PV modules during the operation of the PV system can further enhance the efficiency of the system rapidly and effectively. For the typical PV module current mismatch type faults, namely partial shading, hot spot and crack, the PV module fault diagnosis technology can be roughly divided into the following methods: electrical measurement method, mathematical model method, infrared image analysis method, intelligent detection method, and monitoring system method (Mellit et al., 2018). In detail, the electrical measurement method is to measure the operating parameters of the modules and to detect faulty modules by comparing and analyzing the operating states or electrical parameters of the modules, which may require a large number of sensors and increase the cost of the system (Tingting et al., 2020; Siva et al., 2017). The model-based methods compare the parameters of the PV module with the actual operating module by establishing an accurate PV module model (Haizheng et al., 2019; Aref et al., 2020). Although these methods are computationally efficient, their reliability highly depends on the accurate PV models. Moreover, the modeling process is complex and the identification of model parameters is difficult. The infrared image analysis method that uses the temperature difference between the normal module and the faulty module for fault diagnosis can accurately locate the fault (Akram et al., 2019; Sheikh et al., 2020), but the installation is impractical for household and small distributed PV systems due to the high cost of the infrared equipment. Intelligent detection methods are used for PV module fault diagnosis to separate faults from operating data through intelligent classification algorithms, i.e., random forest, artificial neural network, decision tree, etc. (Zhicong et al., 2018; Li et al., 2021; Rabah et al., 2018). These methods have relatively low cost and good real-time performance, but they rely on expertise knowledge and require a lot of training data. The monitoring system methods that diagnose system faults by monitoring data changes in the system are generally used in large PV systems (Stegner et al., 2018). However, the types of faults that can be determined are limited, which is difficult to locate module-level faults. Among them, the I-V based electrical measurement method is the most simple and effective one, because the I-V curve characteristics of the PV module can not only reveal the operating information of the module, but also correspond well to its fault types (Sarikh et al., 2018).

As the standard IEC62446 points out, the shape of the I-V curve provides information about faults, including damaged cells, short-circuit bypass diodes, partial shading, module mismatch, reduced shunt resistance and increased series resistance (IEC62446, 2016). The standard IEC62446-1 divides the failure of the I-V curve into six shapes, and the step is one of them indicating the mismatch of the PV modules. Partial shading will lead to a step change of the PV module I-V curve, and the partial shading fault can be diagnosed by detecting the step of the I-V curve by deriving the I-V curve (Spataru et al., 2015). However, the I-V curve of faults such as crack and hot spot also have steps, which will affect the diagnosis of partial shading. By using the current and voltage of the module at the maximum power point, the shading effect can be effectively detected (Manit et al., 2020). However, this method requires real-time temperature parameters to correct the maximum power point voltage and current data, and does not consider the impact of other mismatches on the maximum power point data of the module. The principal component analysis (PCA) method of extracting features from the I-V curve was used to detect some partial shading patterns due to the advantage of the simplicity for its model (Fadhel et al., 2019). However, it is necessary to perform some shading condition measurements to establish its PCA model, which makes the training process a non-simple task. According to the standard IEC61215 10.9.2, when the operating current of the PV module exceeds the photo-generated current of a single cell or cell group in the module, the cell or cell group is in a reverse bias state at this time and consumes power, which causes the hot spot effect (IEC61215, 2005). Hot spot causes the I-V curve of the PV module to

have an inclined step, which can be diagnosed using the current change rate at two points on the step of the I-V curve (Bakhsh et al., 2016). Unfortunately, the current changing rate of the cracked module in the I-V curve step will also affect the hot spot detection. Hot-spot free module with integrated bypass diodes for every single PV cell can effectively avoid the hot spot effect and improve the performance of the module, but the cost of the module will be increased, which is not applicable in practice (Hamed et al., 2019). By adding a bypass circuit to improve the structure of the sub-module of the PV module, hot spots can be avoided, but this method will increase the cost and bring challenges to the manufacture of the PV module (Pierluigi et al., 2019). At present, there is no effective diagnostic method for the crack problem of PV modules. Although one or more stress conditions can be predicted for PV modules at the laboratory scale, such as thermal cycling, bending tests, vibrations, and simulated front snow loads (Assmus et al., 2011; Kajari-Schröder et al., 2011), these studies are limited by the rare consideration of the actual factors and difficult collection of reliable statistics for lab-sized samples. A statistical algorithm based on T-test and F-test to identify the significant impact of cracks on the output power performance of the PV module was proposed (Mahmoud et al., 2017). This method compares the theoretical I-V curve with the I-V curve of the cracked module, and judges the crack by power loss, but the influence of the power loss caused by the coupling of other faults is not considered. Electroluminescence (EL) technology can well detect the internal defects of PV cells, such as cell cracks in PV modules (Sara et al., 2020). However, this method requires expensive equipment and needs to be performed in dark conditions, which cannot meet the real-time requirements of fault diagnosis. Note that the existing fault diagnosis strategies are mostly directed to one of the above-mentioned faults and do not consider the coupling between current mismatch faults. In fact, the above three faults will cause the I-V curve to step, which may be difficult to distinguish the specific faults for the above diagnostic methods. Aiming at the widely used PV modules with three bypass diodes, this work proposes a data-driven PV module current mismatch fault diagnosis method based on I-V data. For the distributed PV system with small and medium power class with power optimizer, this paper proposes a diagnostic method for PV module current mismatch faults based on numerical analysis and statistics of I-V data. The method combines the data of PV module I-V curve and divides the I-V curve into two areas. Then, the steps are judged in the low voltage area, and the above three faults are decoupled using polynomial regression fitting in the high voltage area. Finally, combined with the power optimizer with I-V scanning function for verifications, the results from the case studies show that the proposed method has high accuracy and good fault diagnosis performance.

The paper is organized as follows. Section 2 establishes the reverse bias model of PV cells and analyzes the composition of I-V characteristics of PV modules. In Section 3, the faulty PV modules collected at the actual PV power plants are tested, the I-V curves of three common current mismatch faults, (namely partial shading, hot spot and crack) are analyzed in combination with the reverse bias model of PV cells, and the fault features are extracted. Section 4 proposes a novel PV module fault diagnosis method in detail. Section 5 gives the results and discussion, where the detection results are given and instances under different cases are studied to evaluate the performance of the method. Finally, Section 6 draws the conclusion.

2. PV module output characteristics

2.1. PV cell model

A PV cell is essentially a p-n junction. When the illumination is constant, the photo-generated current does not change. It is regarded as a constant current source in the equivalent circuit. Considering the reverse breakdown effect of the diode, the Bishop model with reverse bias characteristics is a single diode model (Bishop, 1988). As shown in

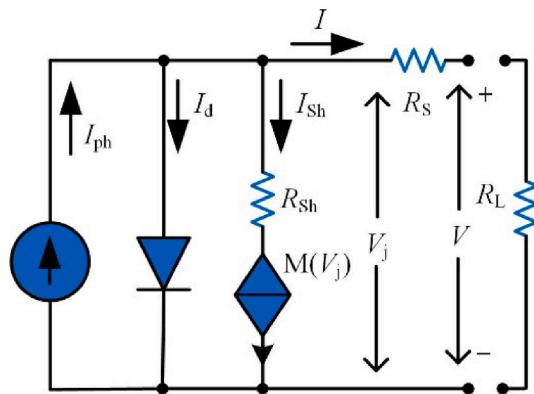


Fig. 1. PV cell equivalent circuit.

Fig. 1, the model consists of five parameters.

According to the equivalent circuit of the PV cell, the I-V output characteristic equation of the PV cell can be obtained:

$$I = I_{ph} - I_d - \frac{(V + R_S I)}{R_{Sh}} \quad (1)$$

$$a \frac{(V + R_S I)}{R_{Sh}} \left(1 - \frac{V + R_S I}{V_{br}} \right)^{-m}$$

$$I_d = I_0 \left[\exp \left(\frac{q(V + R_S I)}{n k T} \right) - 1 \right] \quad (2)$$

$$I_{Sh} = \frac{V_j}{R_{Sh}} \left[1 + a \left(1 - \frac{V_j}{V_{br}} \right)^{-m} \right] \quad (3)$$

$$V = V_j - I R_S \quad (4)$$

where I is the PV cell output current, I_{ph} is the PV cell photo-generated current, V is the output voltage, V_j is the voltage across the junction, $M(V_j)$ is the non-linear current source controlled by the junction voltage, I_0 is the reverse saturation current, I_d is the dark current (current flowing through the diode), q is the charge of the electron with the value of 1.6×10^{-19} , k is the Boltzmann's constant with the value of 1.38×10^{-23} , T is the cell temperature, n is the diode ideality factor, R_S is the cell series resistance, R_{Sh} is the cell parallel resistance, a is the current coefficient associated with avalanche breakdown, m is the avalanche breakdown factor of the diode, V_{br} is the p-n junction breakdown voltage.

According to the above-mentioned PV cell output I-V characteristic equation (1), the cell output characteristics under different parameters are obtained by using Matlab/Simulink, which is shown in Fig. 2. As the

parallel resistance decreases, the degree of reverse bias increases and the reverse leakage current increases. As the irradiance increases, the short circuit current of the PV cell increases.

2.2. PV module I-V output characteristics

The I-V curve represents all the possible operating points of a PV cell, module or string under a certain irradiation and temperature in the existing environment, that is, voltage and current points. From the I-V curve, parameter information of the PV module that describes the operating status of the PV module can be obtained, such as open circuit voltage, short-circuit current, maximum power, and fill factor. Since the voltage of a single PV cell is small, a number of PV cells is typically connected in series to form a substring. In order to reduce the effects of partial shading and hot spots, and to increase power generation efficiency, a bypass diode is usually connected in parallel to each substring (Shimizu et al., 2011). The target PV modules tested in this paper are all formed by three PV substrings in series, while each substring is composed of 20 PV cells connected in series.

As the cells in the PV module are connected in series, their output current should be equal, and the output characteristics of each cell are identical under normal operating conditions. The PV module output I-V curve is a composite of each substring I-V curve which is further a composite of the I-V curves of each series of PV cells, as shown in Fig. 3. Note that V_{OCi} is the open circuit voltage of normal PV cells and V_{oc} is the open circuit voltage of the PV module.

The following relationship between V_{OC} and V_{OCi} is satisfied:

$$V_{OC} = 60V_{OCi} \quad (5)$$

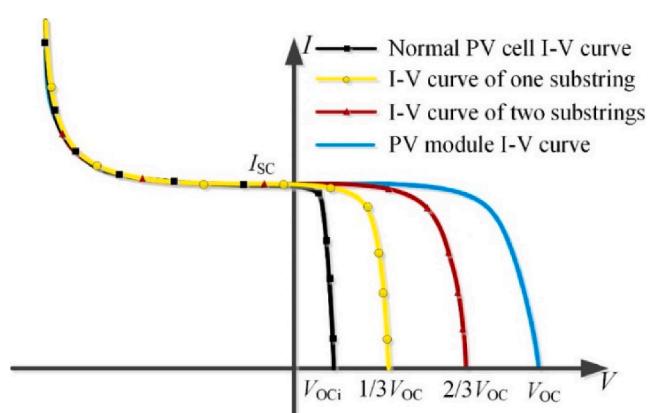


Fig. 3. Schematic diagram of the composition of the I-V curve.

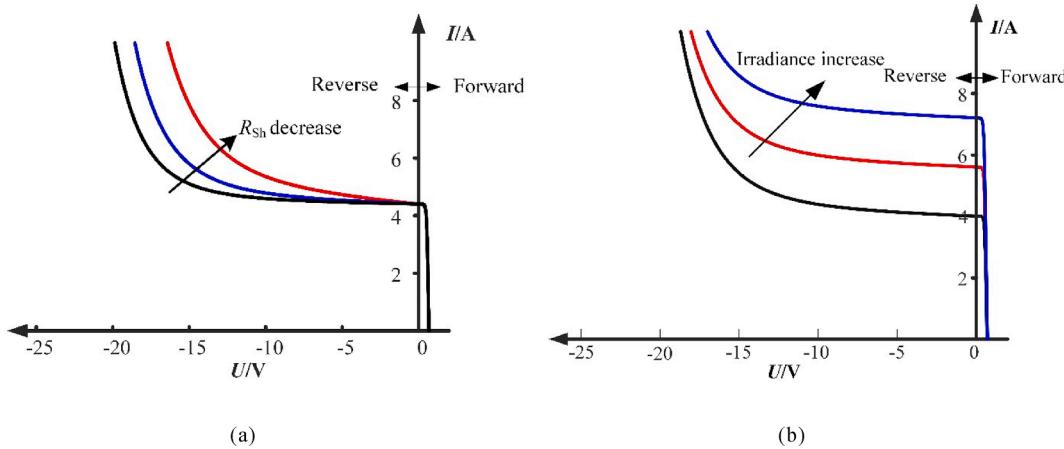


Fig. 2. I-V curves of PV cells under different parameters. (a): I-V curves with different parallel resistance; (b): I-V curves with different irradiance.

3. Fault analysis and feature extraction

3.1. Faulty PV module data collection

It is known that collecting data from faulty PV modules using installed systems is a challenging task. In this work, the collected PV modules are constituted of three 20-cell substrings connected in series. Each substring is connected in parallel to a bypass diode, which can bypass the current of a severely damaged string. A total of 2382 PV modules with current mismatch faults were collected in a 120 MW PV power plant in Lingbi County, Anhui Province, China ($33^{\circ}32'N$ $117^{\circ}33'E$). The power station is a distributed PV power plant on a flat ground. The local environment is ideal, with the lowest temperature within a year of minus $10^{\circ}C$ and the highest temperature of $38^{\circ}C$. A PV string in the distributed PV system includes 22 PV modules connected in series which are connected to the 1000 V DC system. The nameplate rated parameters of the PV modules are shown in Table 1, while the distribution and fault features of PV modules with different fault types are shown in Table 2.

3.2. Fault analysis

The I-V characteristic curve of the PV module represents all possible operating current and voltage points. When the PV module is faulty, its internal electrical parameters will change, which can be directly reflected on the I-V curve (Hirata et al., 2012). Therefore, I-V curve features can be used to diagnose PV modules of different fault types. In order to study the effects of the above different current mismatch fault types on the I-V characteristics, the normal, partial shading, hot spot, and cracked PV modules were tested experimentally in this work. Note that the experimental conditions are chosen to be sunny with the irradiance tested above $800 W/m^2$, and the I-V data of the PV modules are scanned using the HT415 solar cell tester.

3.2.1. Partial shading

Partial shading is the most common form to make the PV module current mismatched. For instance, PV modules are often shaded by buildings, bottom dirt, etc. When shaded, the PV cell is in a reverse bias state. The shaded cell is seriously heated, and the bypass diode is turned on. The experimental scene under the 10% shading of a single cell is shown in Fig. 4, where a cardboard with a light transmittance of 0 is used to shade 1/10 of the area of a single cell in the PV module. Different forms of shading are tested with the test results shown in Fig. 5.

When the PV module is shaded, the shaded cell is reverse biased. Thus, the bypass diode of the substring is turned on, and the I-V curve appears a flat step. As the degree of shadow increases, the step current decreases. The short circuit current of the shaded PV cell is reduced, which is lower than that of the normal PV cell. Due to the series relationship between the cells, the output current of one substring will be determined by the PV cell with the smallest output current.

Assuming that the shaded substring operates at current I , the following expression of the diode turn-on voltage threshold V_c is satisfied:

Table 1
PV module nameplate rated parameters.

| Parameter | Value |
|--|------------|
| Open circuit voltage V_{oc}/V | 37.4 |
| Short circuit current I_{sc}/A | 8.42 |
| Voltage at the maximum power point V_m/V | 30.1 |
| Current at the maximum power point I_m/A | 7.81 |
| Peak power P_m/W | 235 |
| Normal operating temperature $T_c/{}^{\circ}C$ | 48 ± 2 |

Table 2

PV module fault type and distribution.

| Fault type | Number of fault PV modules | Fault feature |
|-----------------|----------------------------|----------------------------|
| Partial shading | 960 | Flat step |
| Hot spot | 790 | Slanted straight-line step |
| Crack | 632 | Convex feature step |



Fig. 4. Photograph of the experimental scene under one cell 10% shading.

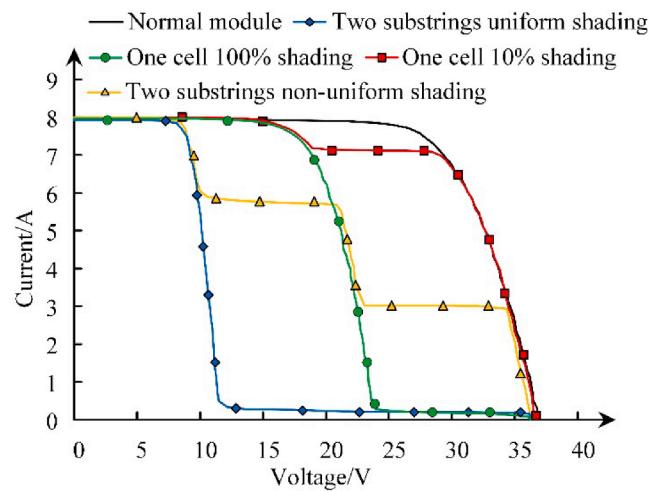


Fig. 5. I-V curves of PV modules under different degrees of shading.

$$V_r - \sum_{m=1}^{19} V_i \geq V_c \quad (6)$$

where the normal power PV cell voltage is V_i , and V_r is the reverse bias voltage which the shaded PV cell is subjected to. As shown in Fig. 6, only one shaded PV cell exists in the PV module, and the remaining 59 PV cells are normal. Note that I_C is the corresponding PV module output current when the diode is turned off, and I_{SC}' is the short circuit current of the shaded PV cell. V_{rc} is the reverse bias voltage that the shaded PV cell is subjected to when the expression (6) is satisfied.

When the PV module operates in a large current range, that is, $I > I_C$, since the reverse bias voltage of the shaded PV cell always satisfies the equation (6), the diode is turned on, and the shaded substring is bypassed. When the PV module operates in the current range of $I_{SC}' < I \leq I_C$, the diode is turned off and the shaded PV cell operates in the

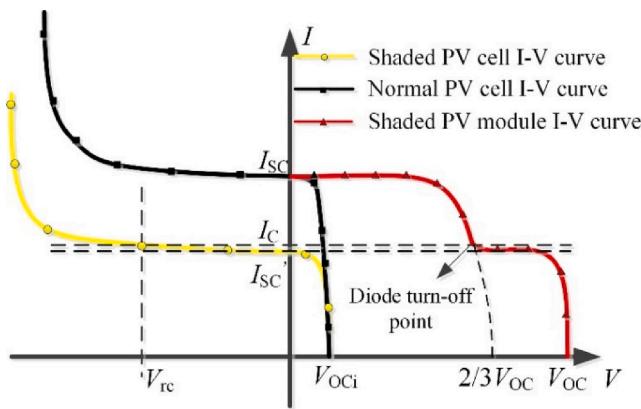


Fig. 6. I-V characteristics of shaded PV module.

reverse bias state. The PV module I-V curve is a composed of 1 shaded PV cell operating in a reverse bias state and the remaining 59 normal PV cell I-V curves, exhibiting a flat step feature. When the PV module is operating in the $I < I_{SC}$ ' current range, the shaded cell is in a forward power state and outputs power in conjunction with the remaining 59 normal cells.

3.2.2. Hot spot

Hot spot is a common phenomenon in PV systems. During operation, when a cell in a PV module is shaded leading to the reverse bias state of the shaded single cell, the heat is generated as a load power consumption. The local temperature of the cell is raised, which may cause permanent damage to the PV cell to form hot spot and even cause a fire. Hot spot in this work refers to the permanent damage of the cell caused by the long-term heating, which is irreversible. In addition to shading, internal defects such as cracking, desoldering, etc. can also cause hot spot (Simon et al., 2010). Several sets of hot spot PV modules are selected for testing with the corresponding results shown in Fig. 7.

It can be seen that the I-V curves of the hot spot assembly have a step in the high voltage range, and the step exhibits a slanted straight-line feature. It is assumed that one PV cell is a hot spot cell, and the remaining 59 cells are normal, as shown in Fig. 8.

Note that I_C is the corresponding PV module output current when the diode is turned off, and I_{SC}' is the short circuit current of the hot spot PV cell. V_{rc} is the reverse bias voltage that the hot spot PV cell is subjected to when it satisfies the equation (6). The hot spot PV cell has large leakage current, and the parallel resistance is smaller than that of the normal PV cell (Bishop, 1989), so its I-V characteristic has a slanted straight-line

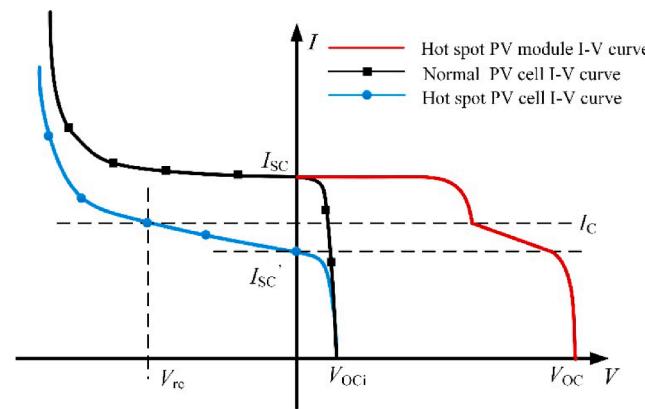


Fig. 8. I-V characteristics of hot spot PV module.

characteristic.

When the PV module operates in a large current range, that is, $I > I_C$, since the reverse bias voltage of the hot spot PV cell always satisfies the equation (6), the diode is turned on, and the hot spot substring is bypassed. The analysis of other current intervals in the I-V curve of the hot spot module is similar to the above partial shading fault. Note that when the PV cell is shaded for a long time, it may cause hot spot. Unlike the partial shading fault, the reverse leakage current of the hot spot cell is large, so that the reverse bias characteristic of the PV cell exhibits a slanted straight-line feature.

3.2.3. Crack

Since cracks in PV cells are difficult to avoid, they are regarded as a serious problem for PV modules and basically impossible to quantify in their impact on the efficiency of the module during its lifetime (Pingel et al., 2009; Morlier et al., 2015; Köntges et al., 2014). In this work, crack fault refers to the damage or cracking of PV cells in PV modules. Cell cracks normally appear in crystalline silicon PV modules during their transportation from the factory to their installation place and subsequent exposure to repeated climatic events such as snow loads, hailstorms or strong wind blows (Köntges et al., 2013; Käsewieter et al., 2014). Cracks in PV modules may cause disconnection between cell parts, resulting in a decrease in output power, insulation failure, non-compliance with safety regulations, and potential safety hazards such as electric leakage of PV modules. The I-V characteristics of the tested cracked PV modules with different degrees of crack are shown in Fig. 9.

It can be concluded that the I-V curves of the cracked PV modules have a step exhibiting a convex feature. Different from the I-V curve step characteristics of partial shading and hot spot PV modules, the I-V curve

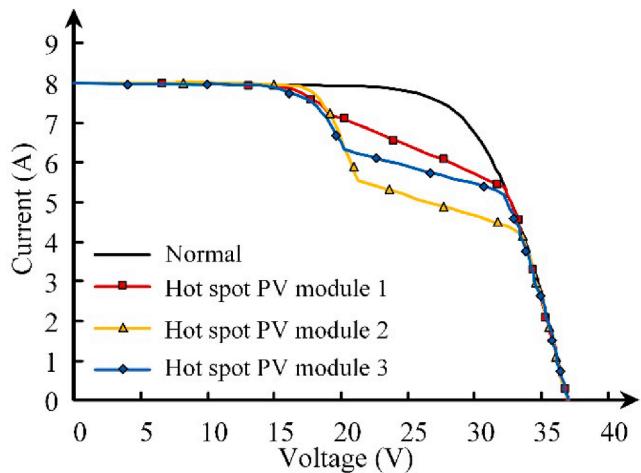


Fig. 7. I-V curves of different hot spot PV modules.

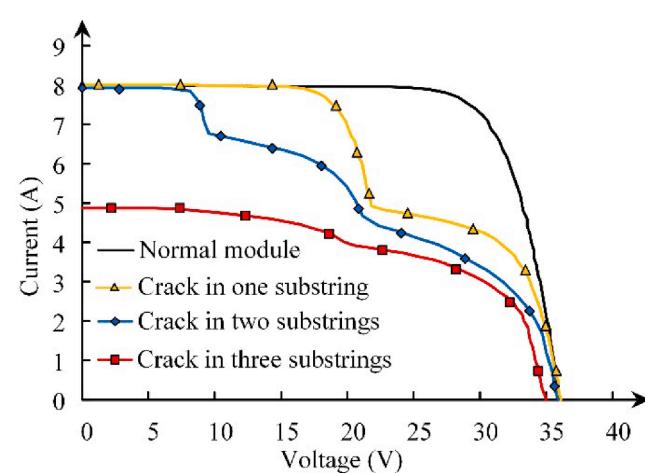


Fig. 9. I-V curves of PV module with different degrees of crack.

step characteristics of cracked PV modules are non-linear, and the step exhibits the characteristics of an upward convex function. The short circuit current I_{SC} and the open circuit voltage V_{OC} of the not all substring cracked PV modules are basically unchanged. As the crack degree is deepened leading to the severe cell damage, the current generating capacity is drastically reduced, and the short circuit current is greatly reduced. The step section of the cracked PV module I-V curve exhibits a convex feature, and the I-V characteristic of the cracked module is demonstrated in Fig. 10. Assuming that there are 3 cracked cells in the same substring, the remaining 57 cells are normal. The cracked cells are different in degree of cracking, so their reverse bias characteristics are also different. Note that I_{SC} is the short circuit current of the normal PV cell, I_1 is the current corresponding to the diode turn-off point, and I_{SC1} , I_{SC2} , and I_{SC3} are the short circuit currents of the 3 cracked cells, respectively.

When the PV module current is I_1 , the diode is critically turned off, and the normal PV cell voltage is V_1 . The reverse bias voltages of the three cracked PV cells are V_{b1} , V_{b2} , and V_{b3} , respectively, satisfying the following relationship:

$$V_{b1} + V_{b2} + V_{b3} - \sum_{m=1}^{17} V_1 = V_C \quad (7)$$

where V_C is the diode conduction threshold voltage.

When the current of the PV module is in the range of $I_1 < I \leq I_{SC}$, the bypass diode where the cracked substring is located is turned on, and the I-V curve of the PV module is a composite of 40 normal PV cells of the remaining substrings. In the interval of $I_{SC1} < I \leq I_1$, the diode is turned off, and at this time, the 3 cracked PV cells are all operating in the reverse bias state. Corresponding to the L_1 segment on the I-V curve is the result of a combination of 3 cracked PV cells operating in the reverse bias state and the remaining normal PV cells. The analysis of other current intervals in the I-V curve of the cracked module is similar. In the actual cracked PV module, since there are many cracked cells with non-uniform crack degree, the I-V curve of the cracked module has more inclined sections with different slopes, and composite curve exhibits convex feature.

3.3. Fault features

It can be concluded from the above three current mismatch type faults that the partial shading, hot spot and crack all will cause the I-V curve of the PV module to step and the current in the step to drop. The difference among them lies in that the shadow fault is a flat step, the hot spot has a slanted straight-line feature in the step, and the step of the crack fault exhibits a convex feature. Note that the step features of the

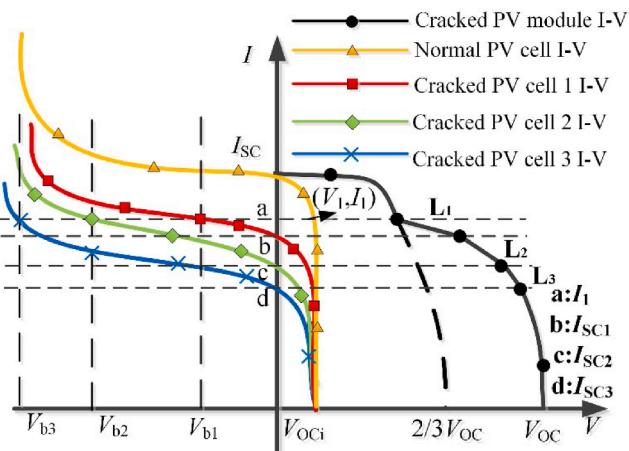


Fig. 10. I-V characteristics of cracked PV module.

three mismatch faults are summarized in Table 2.

4. Fault diagnosis process

4.1. Data acquisition and preprocessing

In order to reduce the mismatch effect and increase the power of the system, household, industrial and commercial rooftop distributed power stations are equipped with a power optimizer for each PV module of the PV string. When the module in the PV string is shaded, only the shaded module will be affected, while the output of other modules will not be affected. The cask effect of the PV string is avoided, and the modules can work in an optimal state. The system structure is shown in Fig. 11. The I-V scanning function of the optimizer can obtain the I-V data of the series PV modules within 1 s, including 32 sets of voltage and current data points with different voltages. The I-V data is recorded as (U_i, I_i) in the order of voltage from high to low, where $i = 0, 1, 2, \dots, 31$. Because the scanning time is very short, the irradiance change and power loss during the scanning process can be ignored.

For a certain temperature and irradiation, to derive the voltage of equation (1), we can obtain that the derivative value of I is less than 0, indicating that I decreases monotonically, i.e., I decreases as V increases. In practice, due to the instability of data sampling, the obtained I-V data may have a large jitter, which may affect the subsequent use of fault diagnosis. In theory, the current of I-V data decreases as the voltage increases. Therefore, in order to eliminate the outliers, the raw I-V data is preprocessed. The voltage data corresponding to the I-V data of the PV module remains unchanged, when the current data meets the following conditions:

$$\begin{cases} I_i > I_{i-1} \\ I_i > I_{i+1} \end{cases} \quad (8)$$

or meet the following conditions:

$$\begin{cases} I_i < I_{i-1} \\ I_i < I_{i+1} \end{cases} \quad (9)$$

Then the current value I_i is interpolated processed, satisfying the following equation:

$$I_i = \frac{I_{i-1}(U_i - U_{i+1}) + I_{i+1}(U_{i-1} - U_i)}{U_{i-1} - U_{i+1}} \quad (10)$$

Since the preprocessed data still contains random noise, they should be filtered prior to data analysis. To reduce the random noises, the median filtering technique is adopted in this work (Arce, 2005). For preprocessed current data I_i , the median filter f_m is applied on I_i to produce smoothed current data I_n :

$$I_n = f_m[I_{i-N}, \dots, I_i, \dots, I_{i+N}] \quad (11)$$

where N is a nonnegative integer. In most cases, since the smooth window is symmetric, we filter the current data using the same interval. After many tests, we set $N = 1$ in this paper to ensure best filtering effect. The above process is performed to obtain the smoothed data (U_n, I_n) , where $n = 0-31$. A set of raw I-V data and the data after interpolation and smoothing are compared as shown in the Fig. 12. It can be seen that after processing, the abnormal data on the original I-V curve can be eliminated and the curve is smoother.

4.2. Current mismatch detection

Based on the fault features of the above PV module current mismatch types, a partition-based mismatch detection method is proposed in this work. Taking the experimental PV module as an example, the PV module has three substrings. Because the PV module I-V curve is composed of the I-V curves of each substring, the I-V curve of the PV module is divided into two sections, namely the low voltage area $[0, 2V_{OC}/3]$ and

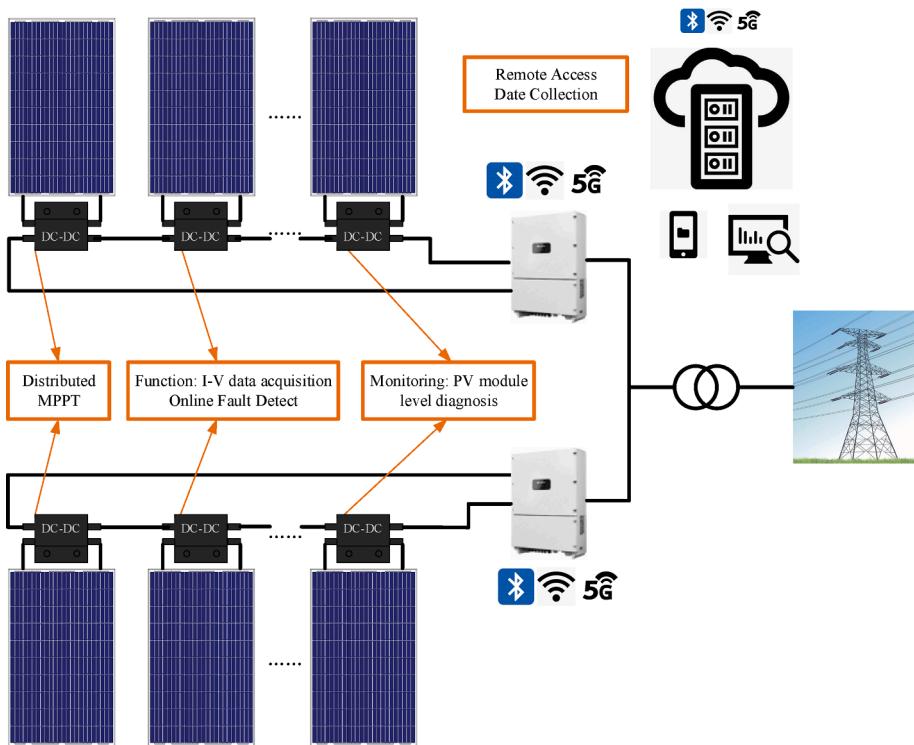


Fig. 11. Schematic diagram of data acquisition and fault diagnosis system.

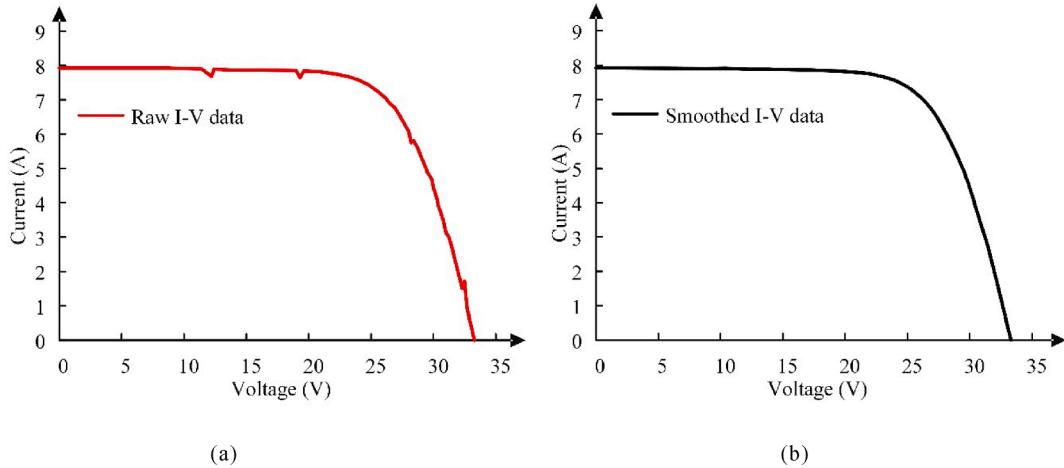


Fig. 12. Comparison of raw and smoothed I-V data. (a): Raw I-V curve; (b): Smoothed I-V curve.

high voltage area [$2V_{OC}/3, V_{OC}$]. The step inflection point that appears on the I-V curve corresponds to the moment when the bypass diode connected in parallel with the substring is from turned on to off. When there is a mismatch in the substrings of the PV module, the current at the two-thirds open circuit voltage will drop due to the conduction of the bypass diode. Therefore, by detecting the distortion of the low voltage section I-V curve, we can determine whether the current mismatch occurs in the PV module. Take the partial shading fault as an example, the specific method is shown in Fig. 13.

Taking two-thirds open circuit voltage point as the starting point marked as $P_1(2V_{OC}/3, I_k)$, here I_k is obtained by linear interpolation between two adjacent points. The first point on the I-V curve that meets the following conditions is found in the direction of voltage reduction, while the other point is marked as $P_2(U_t, I_t)$. The condition is given as follows:

$$I_{SC} - I_t \leq \varepsilon \quad (12)$$

where ε is the critical value of normal PV module current drop. There is a section of the I-V curve in the low voltage region that is approximately constant current. After statistical analysis of a large number of normal PV module data, the optimal value of ε is set to 0.1. Points P_1 and P_2 form a detection straight line, which is expressed as:

$$I = f(U) = \frac{I_k - I_t}{2V_{OC}/3 - U_t}(U - U_t) + I_t \quad (13)$$

Then the relative position of the I-V curve and the detection line can be judged. As shown in Fig. 14, it is a set of actual measured I-V curve of the module without mismatch, and mismatch detection is performed on it according to the method of this work. The points on the I-V curve in the range of P_1 to P_2 are all above the detection line. At this time, the

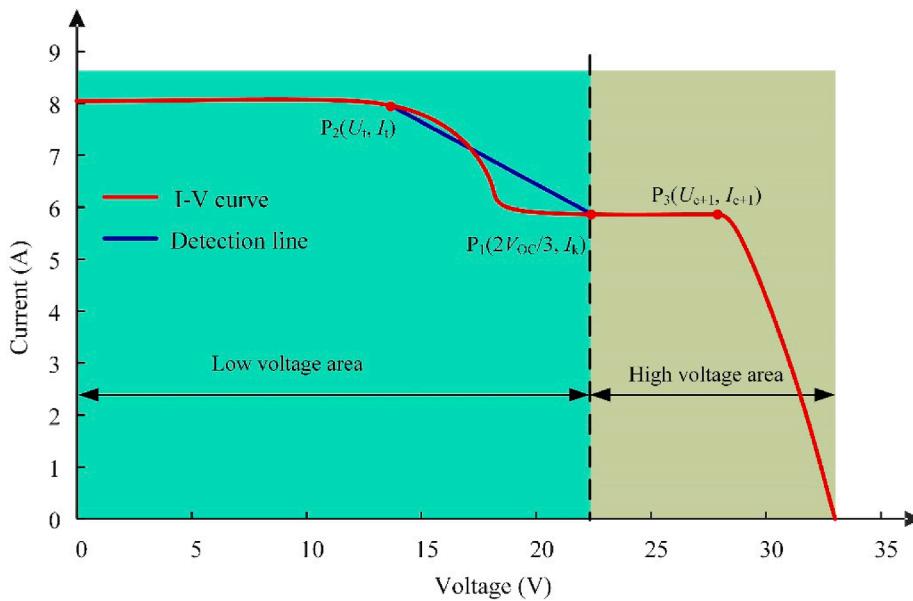


Fig. 13. Schematic diagram of I-V curve division and feature points selection.

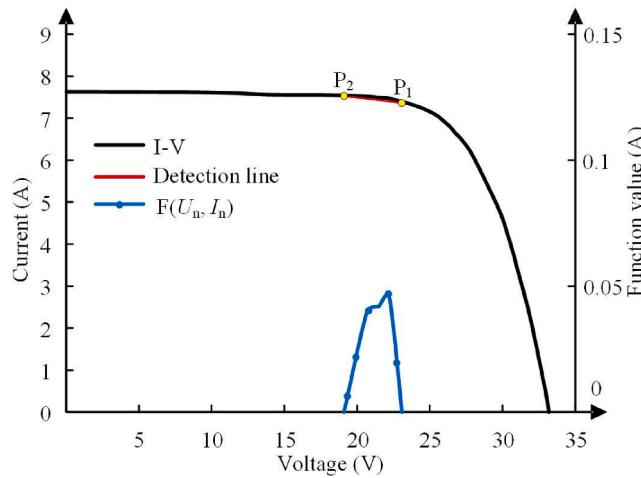


Fig. 14. Characteristics of the non-mismatch PV module.

detection line does not indicate mismatch. The setting function $F(U_n, I_n)$ is the current difference between the I-V curve and the detection line at the same voltage. Because the normal I-V curve is approximately a constant current state in the low voltage section, when the point on the I-V curve is below the straight line, the curve is distorted and the PV module is mismatched. The expression of $F(U_n, I_n)$ is as follows:

$$F(U_n, I_n) = I_n - f(U_n) \quad (14)$$

where $n = k, k + 1, \dots, t$. According to the above method, whether the PV module is mismatched can be determined.

When the PV module is mismatched, the fault type is further decoupled according to the I-V data of the high voltage area. In order to determine the end of the step, a linear interpolation method is used for the current data of every three points in the high voltage area, and the interpolated data is compared with the original I-V data. The interpolated current data is defined as I_{int_n} , and the function $G(I_{int_n}, I_n)$ is the difference between the original I-V current data in the high voltage area and the interpolated current data. The voltage and current point corresponding to the maximum point of $G(I_{int_n}, I_n)$ is recorded as (U_e, I_e) , which is the end of the step. The expressions for I_{int_n} and $G(I_{int_n}, I_n)$ are:

$$I_{int_n} = \frac{I_{n-1}(U_n - U_{n+1}) + I_{n+1}(U_{n-1} - U_n)}{U_{n-1} - U_{n+1}} \quad (15)$$

$$G(I_{int_n}, I_n) = I_n - I_{int_n} \quad (16)$$

4.3. Fault decoupling of step feature

In order to further determine the fault type of the PV module, it is necessary to decouple the step feature of the mismatched module. The I-V data of the high voltage area from point $P_1(2V_{OC}/3, I_k)$ to a point after the end of the step $P_3(U_{e+1}, I_{e+1})$ is selected as the data for step identification. Through the above experimental test, it is found that the step of partial shading fault is relatively flat, while the step current drop for hot spot and crack is larger. By defining the variable ΔI as the current drop from point $P_1(2V_{OC}/3, I_k)$ to $P_3(U_{e+1}, I_{e+1})$, the expression of ΔI is given by:

$$\Delta I = I_k - I_{e+1} \quad (17)$$

The values of ΔI for different fault types of multiple random samples are statistically analyzed, and the corresponding results are shown in Fig. 15. The abscissa indicates the number of samples, and the current drop of the PV module with partial shading fault falls within the range of [0, 0.18 A]. The hot spot and crack fault exceed this range, so the partial

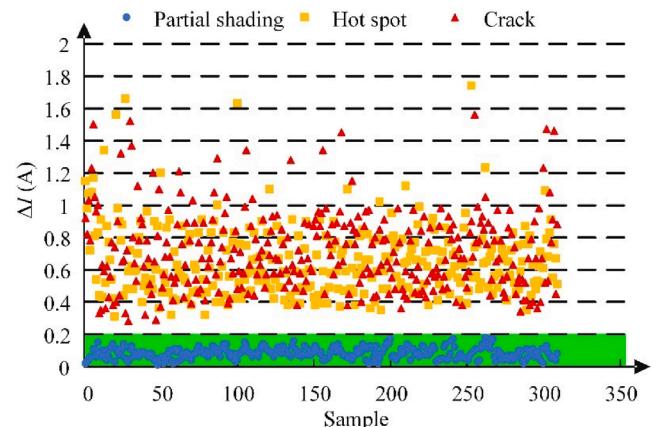


Fig. 15. Statistics of the ΔI values of different fault types.

shading fault can be identified by ΔI .

In order to further identify hot spot and crack fault, polynomial regression fitting is performed on the step data of hot spot and cracked PV modules, as the polynomial regression fitting method is one of the methods of curve fitting. The basic idea is to approximate the functional relationship between the discrete point coordinates of the discrete data with a continuous curve (Helzer et al., 2004). By fitting the I-V data of the step and analyzing the correlation between them, a suitable expression can be found. The accuracy of polynomial fitting is usually evaluated by the following indicators. (i) Root mean squared error (RMSE), also called the standard deviation of the regression fitting system, is used to measure the deviation between the observed value and the true value. The smaller the value, the closer the fitted data is to the true value. (ii) Coefficient of determination (R^2) is also known as goodness of fit. In fact, the coefficient of determination is used to characterize the degree of fitting through the changes of data. The normal range of R^2 is [0, 1]. When it is much closer to 1, it indicates that the independent variable of the equation has a stronger ability to explain the dependent variable, and the change caused by the independent variable accounts for a higher percentage of the total change. The denser the observation points are near the regression line, the better the model fits the data. The larger R^2 is, the more explanatory the predicted data can be through the model. In this work, statistical parameters RMSE and R^2 are selected to evaluate the accuracy of polynomial regression fitting. Through the fitting of multiple sets of data, it is found that the step data of the hot spot module has a strong linear characteristic and the step data of the cracked module has a good parabolic characteristic. Fig. 16 gives an example of the fitting results of the step data from P_1 to P_3 of a set of hot spot and cracked modules. According to the results of the step fitting, the I-V curve of the hot spot module exhibits a linear characteristic from P_1 to P_3 , while the cracked module exhibits a parabolic characteristic from P_1 to P_3 . Note that Table 3 shows an example of the step data fitting results for a set of hot spot and cracked PV modules among many fitting results.

The step data of the hot spot module has a stronger linear characteristic than that of the cracked module. Therefore, hot spot and cracked can be distinguished by linear regression fitting. Through experiments with large amounts of data, the R^2 statistical results of the linear fitting of hot spot and cracked modules are shown in Fig. 17, and the R^2 values of the linear fitting of the hot spot modules in the sample are all above 95%. Therefore, when the goodness of fit of the linear polynomial is above 95%, the fault is considered to be hot spot, otherwise it is crack fault.

4.4. Algorithm

In order to ensure the accuracy of the I-V data of the PV modules, the I-V curves are obtained when the irradiance is high in sunny weather. In this work, the I-V curves are scanned when the irradiance is greater than 500 W/m². When having irradiance meter, the irradiance information is read directly, and the MPPT current I_m of the string is obtained from the inverter operation information without the irradiance meter to estimate the irradiance. Specifically, three parts are included: I-V data acquisition and smoothing, current mismatch detection, and step decoupling. The algorithm flow chart is shown in Fig. 18.

5. Results and discussion

5.1. Test results

The proposed fault diagnosis method is verified in a 10.2 MW distributed PV power system on the roof of Sungrow Power Co., Ltd. located in Hefei, Anhui, China (31°52'N 117°17'E) using the fault modules collected from the actual PV power plants. A PV string in the system is formed by 22 PV modules connected in series and connected to the 1000 V DC system. The PV modules in the string of the distributed system are equipped with power optimizers to obtain I-V data, and the corresponding diagnosis system is shown in Fig. 11. The power optimizer is SP-375 W provided by Sungrow Power Supply Co., Ltd. with the specifications shown in Table 4.

The test sample data is randomly selected from a large number of faulty modules and has no overlap with the above fault cases. Among them, the non-mismatch, partial shading, hot spot and crack modules are 500, 500, 480, and 325 respectively. For the setting of partial shading in the experiment, we use cardboard with light transmittance of 0 and mesh cloth with light transmittance of 1/3 to cover in different shapes. The partial shading in the experiment are shading patterns of different sizes and shapes. Fewer samples of cracked modules are collected mainly because the crack in the actual power plants mostly occurs in the initial installation of the modules and the double glazing modules. In the later operation, the probability of crack fault is small. In this work, combined with the characteristics of the optimizer, the online fault diagnosis is performed according to the above diagnostic algorithm. In order to specifically evaluate the performance of the fault diagnosis method, two kinds of indicators are defined: fault detection rate and false detection rate. The fault detection rate is used to evaluate the performance of the method to correctly detect faults and identify specific fault types, while the false detection rate is used to evaluate the performance of faults being falsely detected or identified. The definitions of fault detection rate and false detection rate are given as follows:

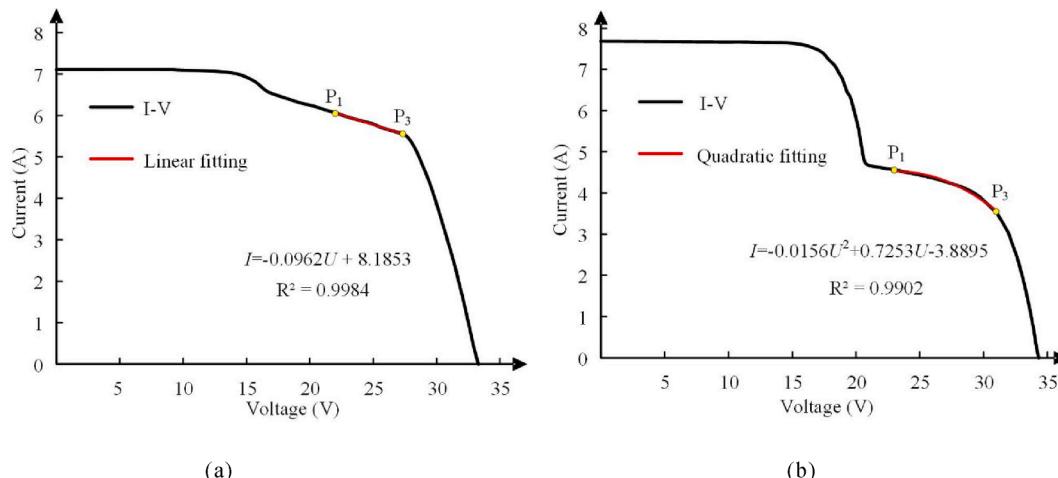
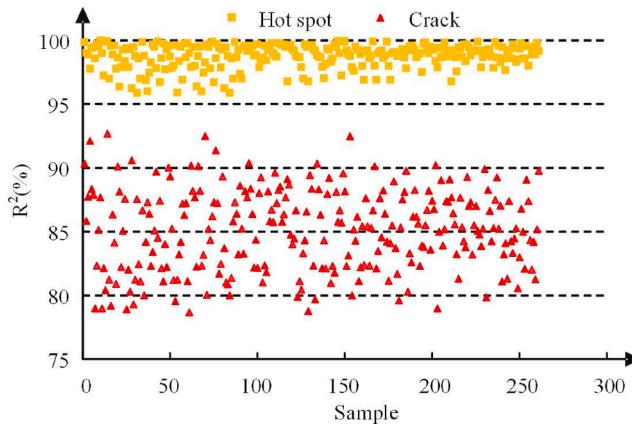
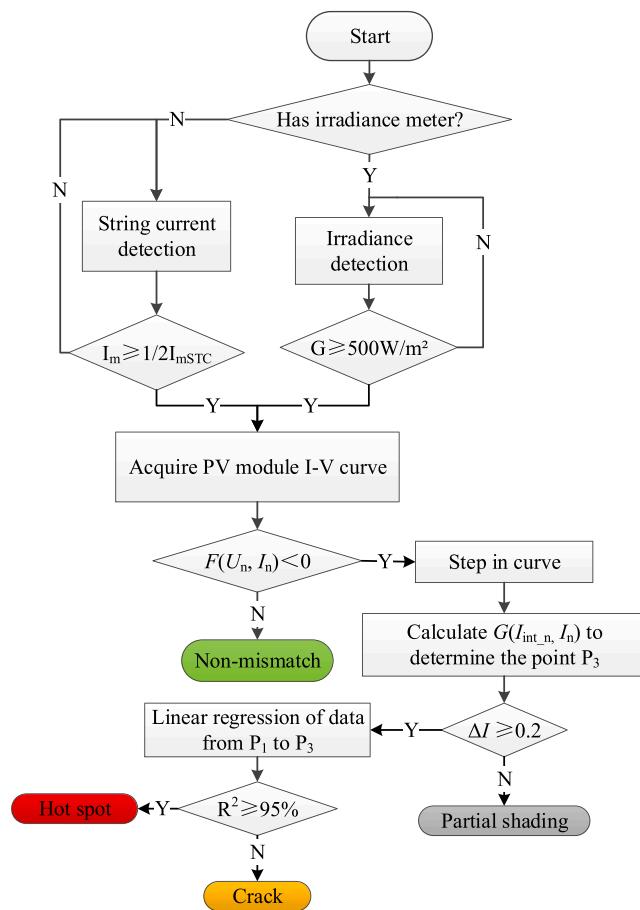


Fig. 16. Step fitting results of a set of hot spot and cracked modules. (a): Hot spot; (b): Crack.

Table 3

Fitting results of the step data of the hot spot and cracked PV modules.

| Module type | Linear fitting | | | Quadratic fitting | | |
|-------------|-----------------------|---------|----------------|----------------------------------|----------|----------------|
| | Fitting equation | RMSE | R ² | Fitting equation | RMSE | R ² |
| Hot spot | $I = -0.0852U + 8.16$ | 0.01975 | 0.9982 | $I = -0.024U^2 + 0.0273U + 6.84$ | 0.003041 | 0.9998 |
| Crack | $I = -0.0865U + 4.26$ | 0.05874 | 0.9034 | $I = -0.0117U^2 + 0.541U - 4.06$ | 0.02082 | 0.9910 |

**Fig. 17.** Statistics of R^2 values of linear fitting of hot spot and cracked modules.**Fig. 18.** Fault diagnosis flow chart.**Table 4**
Specifications of the used power optimizer.

| Specifications | Parameters |
|------------------------|------------------|
| Operating temperature | -40 °C to +85 °C |
| Relative humidity | 0–100 %RH |
| Altitude requirements | 0–4000 m |
| Input voltage | 8–80 V DC |
| Maximum output current | 12A |
| Maximum input power | 375 W |
| Output voltage range | 4–80 V DC |
| Maximum output current | 15A |
| Maximum efficiency | 99.5% |

$$\text{Fault Detection Rate} = \frac{\text{Correct instances of detection and identification of L}}{\text{Total instances of L}}$$
 (18)

$$\text{False Detection Rate} = \frac{\text{False instances of detection or identification of L}}{\text{Total instances of fault}}$$
 (19)

where L is the label of a specific fault type. The fault type label consists of non-mismatch, partial shading, hot spot, and crack faults.

Confusion matrix, as a visualization tool, can intuitively compare the detected type with the actual type. Each column of the confusion matrix represents the detected type, and the total number of each column represents the number of data (samples) diagnosed as the category; each row represents the true attribution category of the data, and the total number of data in each row represents the number of data instances of that category. The confusion matrix of the test results of the above samples is shown in Fig. 19. The fault detection rate and false detection rate are calculated according to the confusion matrix to evaluate the accuracy of the fault diagnosis method. The fault diagnosis results are shown in Table 5. It can be seen that the adopted method is effective for fault detection and identification. The lower the false detection rate, the better the ability of the method to identify specific faults.

5.2. Discussion

In this section, four case studies involving different types of faults are conducted, including non-mismatch, partial shading, hot spot, and crack faults. Based on the sample data of the above diagnosis results, the adaptability of the method in each case is discussed in detail.

| | Detected type | | | |
|-------------|-----------------|-----------------|----------|-------|
| | Non-mismatch | Partial shading | Hot spot | Crack |
| Actual type | Non-mismatch | 500 | 0 | 0 |
| | Partial shading | 0 | 500 | 0 |
| | Hot spot | 19 | 0 | 461 |
| | Crack | 13 | 10 | 6 |
| | | | | 296 |

Fig. 19. Confusion matrix of sample test results.

Table 5
Fault diagnosis results.

| Module type | Number of PV module samples | Fault detection rate/% | False detection rate/% |
|-----------------|-----------------------------|------------------------|------------------------|
| Non-mismatch | 500 | 100 | 2.45 |
| Partial shading | 500 | 100 | 0.77 |
| Hot spot | 480 | 96.04 | 0.46 |
| Crack | 325 | 91.08 | 0 |

5.2.1. Case study of non-mismatch

Under normal operating conditions, the I-V curve has no distortion, which is the characteristic of a convex function without inflection points and steps. Therefore, when performing mismatch detection on normal modules, the points on the I-V curve are above the detection straight line. The value of the mismatch detection function $F(U_n, I_n)$ of the normal module is non-negative, and the module does not have a mismatch. When PV modules have non-current mismatch faults such as degradation or potential induced degradation (PID), although the I-V curve of the module is distorted, there is no bypass diode in the module to conduct, which will not affect the mismatch fault diagnosis result of the proposed method. Fig. 20 shows the mismatch detection results of normal module and PID fault, where the detection function $F(U_n, I_n)$ value is all non-negative, and the diagnosis result is non-mismatch. Therefore, the non-mismatch type will not affect the diagnosis of current mismatch faults.

5.2.2. Case study of partial shading

According to the characteristics of the partial shading PV modules, as the degree of shading increases, the step current of the module becomes smaller. In this case study, the accuracy of the method for partial shading fault is verified according to different degrees of shading.

(1) A single substring with small shading: In the test samples, the smallest shading is that the bottom of the PV module is soiled by mud. The appearance and I-V curve of the PV module are shown in Fig. 21. A substring is slightly soiled in the PV module, and the step feature is also obvious. According to the above method, the mismatch detection and step decoupling of the shading module are performed, with the distribution of the value of $F(U_n, I_n)$ and $G(I_{int,n}, I_n)$ shown in Fig. 21 (b). It can be seen that $F(U_n, I_n)$ has negative value indicating a mismatch in the shaded PV module. The coordinates of point P_1 and point P_3 of the step data are (22.3, 6.64) and (25.8, 6.49) respectively. After calculating the value of ΔI as 0.15 A, within the threshold range of the partial shading, the PV module is diagnosed as partial shading fault. Therefore, this method is effective for slight shading.

(2) Multiple substrings with different degrees of shading: When multiple substrings in PV module are non-uniformly shaded, multiple bypass diodes in the module are turned on, causing multiple steps in the I-V curve. As shown in Fig. 22 (a), two substrings in the module have different degrees of shading, as shown in Fig. 22 (b), the two substrings of the PV module are mismatched at this time, and two steps appear on the I-V curve. The mismatch detection and step decoupling of the module with two substrings shaded are performed, and the distribution of the value of $F(U_n, I_n)$ and $G(I_{int,n}, I_n)$ is shown in Fig. 22 (b). The $F(U_n, I_n)$ has negative value, so there is a mismatch in the PV module. The coordinates of point P_1 and point P_3 of the step data are (23, 0.21) and (33.8, 0.17) respectively. After calculating the value of ΔI as 0.04 A, within the threshold range of the partial shading, the PV module is diagnosed as partial shading fault. This method is also effective when multiple substrings are non-uniformly shaded.

5.2.3. Case study of hot spot

The leakage current of the hot spot PV cell is larger than that of the normal cell. Since the I-V curve of the PV module is composed by the I-V curves of the cells, the I-V curve of the hot spot module has a sloping step. For general hot spot modules, due to the fact that the I-V curve is distorted and the concavity and convexity of the curve changes, there is an obvious concave inflection point. The infrared image of the hot spot module is shown in Fig. 23 (a), while the I-V curve is shown in Fig. 23 (b). Note that (i) The operating voltage and current of the hot spot module are 27.5 V and 6.15 A, respectively. (ii) The ambient temperature is 36.5 °C and the wind speed is 2.4 m/s. The bypass diode of the substring is turned on, and the I-V curve has obvious inflection point and step. The mismatch diagnosis of the module with hot spot is performed, and the distribution of the value of $F(U_n, I_n)$ and $G(I_{int,n}, I_n)$ is shown in Fig. 23 (b). The $F(U_n, I_n)$ has negative value, so there is a mismatch in the PV module. This mismatch detection method is also applicable to hot spot fault. The coordinates of point P_1 and point P_3 of the step data are (22.1, 6.73) and (27.5, 6.15) respectively. After calculating the value of ΔI as 0.58 A, which exceeds the threshold range of partial shading, linear regression fitting is performed on the step data with the corresponding result shown in Fig. 23 (c). The R^2 of the linear polynomial fitting reaches 99%, indicating that the step data has good linear characteristic, so the PV module is diagnosed as hot spot fault.

There are modules with insignificant hot spot effect in the test samples. One set of I-V curve and infrared image are shown Fig. 24. In the infrared image shown in Fig. 24(a), the operating voltage and current of the hot spot module are 26.7 V and 5.02 A, respectively. The ambient temperature is 29.3 °C, and the wind speed is 1.5 m/s. It is found that the temperature of the hot spot cell is no more than 10 °C which is higher than the normal cell, and the hot spot heating is not

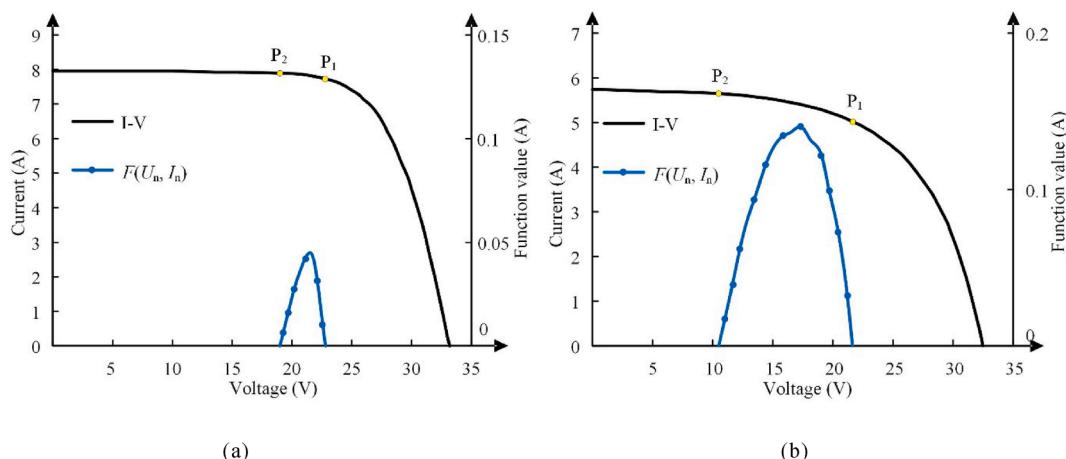


Fig. 20. Non-mismatch type detection results. (a): Normal module detection result; (b): PID module detection result.

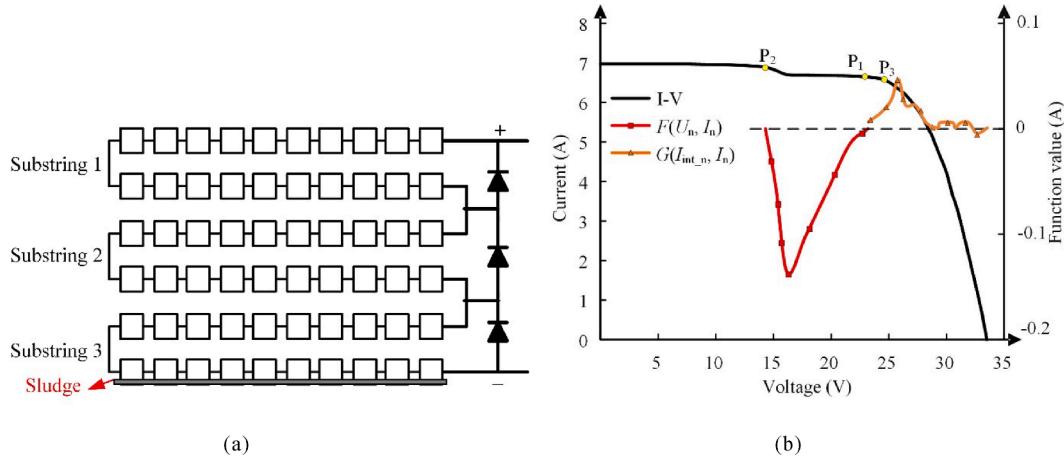


Fig. 21. Case study of the module under one substring soiled. (a): Schematic diagram of shading; (b): Characteristics of the module.

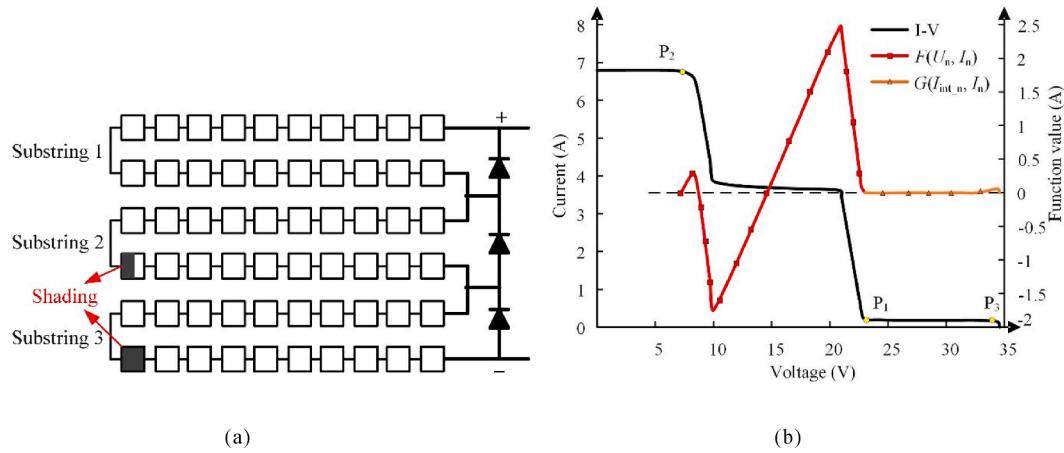


Fig. 22. Case study of the module under multiple substrings with different degrees of shading. (a): Schematic diagram of shading; (b): Characteristics of the module.

serious. The I-V curve does not show a concave inflection point. The mismatch diagnosis of the module is performed, and the distribution of the value of $F(U_n, I_n)$ and $G(I_{int_n}, I_n)$ is shown in Fig. 24 (b). We can see that the $F(U_n, I_n)$ values are all non-negative, because the points on the I-V curve are all above the detection line. Therefore, the mismatch detection method fails for this slight degree of hot spot. The module is detected as a non-mismatch type.

5.2.4. Case study of crack

The crack degree of the cell in the cracked PV module is different, which makes the step of the I-V curve present a convex feature. In this case study, the adaptability of the proposed mismatch diagnosis method to PV modules with different crack degrees is investigated.

- (1) Single substring crack: When a substring in the PV module is cracked, its I-V curve has an obvious convex step feature. Fig. 25 (a) shows the I-V curve of the one substring cracked module, and mismatch diagnosis is performed. The distribution of the value of $F(U_n, I_n)$ and $G(I_{int_n}, I_n)$ is also shown in Fig. 25 (a), where the coordinates of point P₁ and point P₃ of the step data are (23.8, 6.46) and (31.2, 5.60) respectively. After calculating the value of ΔI as 0.86 A, which exceeds the threshold range of partial shading, linear regression fitting is performed on the step data of the module with the corresponding result shown in Fig. 25 (b). The R^2 result of linear regression fitting is 90.93%, which does not reach the preset strong linear correlation threshold $R^2 \geq 95\%$, so the PV module is diagnosed as crack fault.

(2) Multiple substrings with different degrees of crack: When the cells of different substrings in the PV module are cracked to different degrees, the I-V curve appears multiple steps. As shown in Fig. 26 (a), the cells of the three substrings in the PV module are cracked showing different degree of crack, the I-V curve of the cracked module has multiple convex feature steps. The mismatch diagnosis of the multi-substring cracked module is then performed. The distribution of the value of $F(U_n, I_n)$ and $G(I_{int_n}, I_n)$ is also shown in Fig. 26 (a), where the coordinates of point P₁ and point P₃ of the step data are (21.9, 1.64) and (28.6, 1.32) respectively. The linear regression fitting is performed on the step data of the module with two substrings cracked with the corresponding result shown in Fig. 26 (b). Finally, according to the calculated characteristic parameters for diagnosis, the module is diagnosed as crack fault.

(3) Uniform crack of cells in PV module: When there is crack fault in the PV module, if the fault is not eliminated in time, the cells of the entire module may eventually be damaged, causing the degree of crack tend to be uniform, as the cracked module continues to operate. The number of cracked cells in the PV module at the initial stage of crack may be relatively small, and the convex feature of the steps on the I-V curve are not obvious. When only one cell in the module is cracked, its I-V curve is shown in Fig. 27 (a). Because there is only one cell cracked in the module, the step on the I-V curve of the module after compounding is approximately an inclined straight line, similar to the hot spot module. The mismatch diagnosis of the cracked module is performed, and

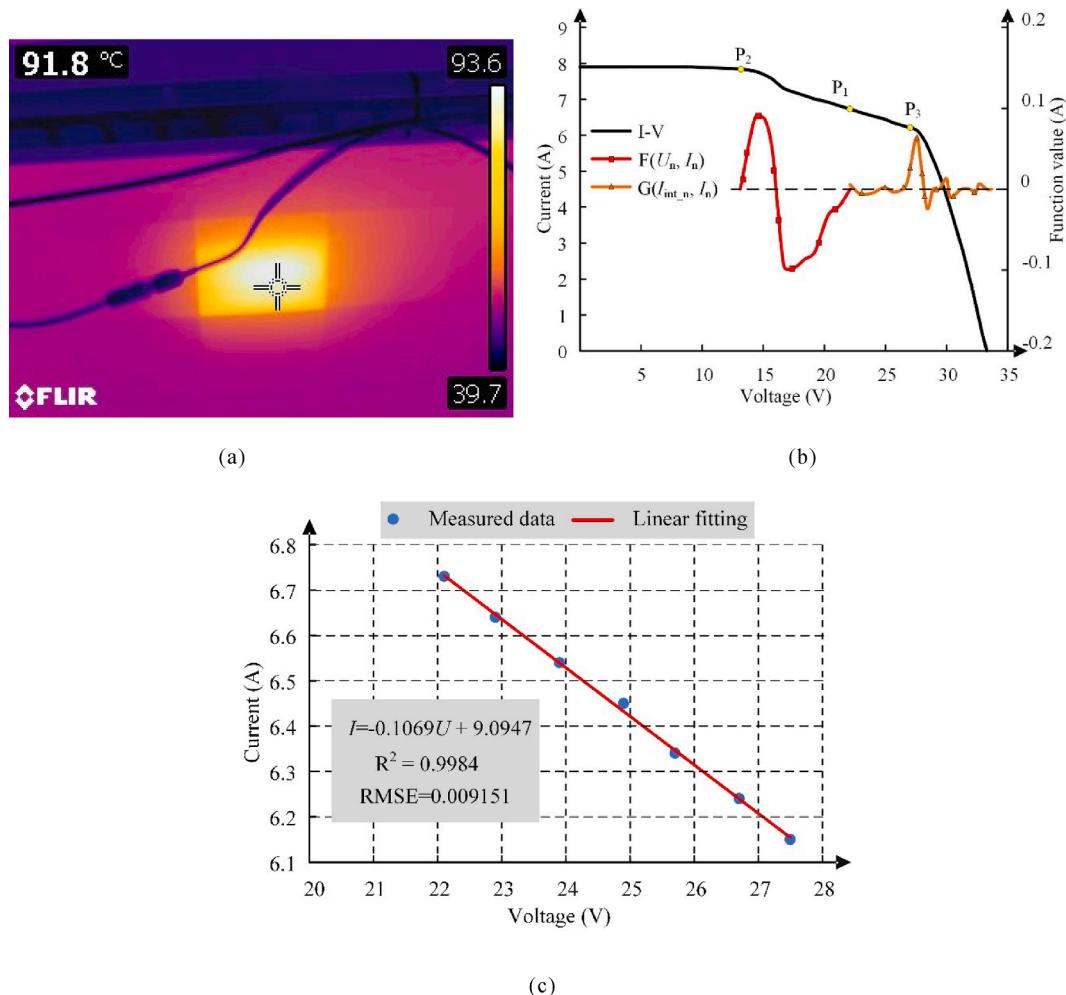


Fig. 23. Case study of a general hot spot module. (a): IR image of the hot spot module; (b): Characteristics of the module; (c): Linear regression fitting of step data.

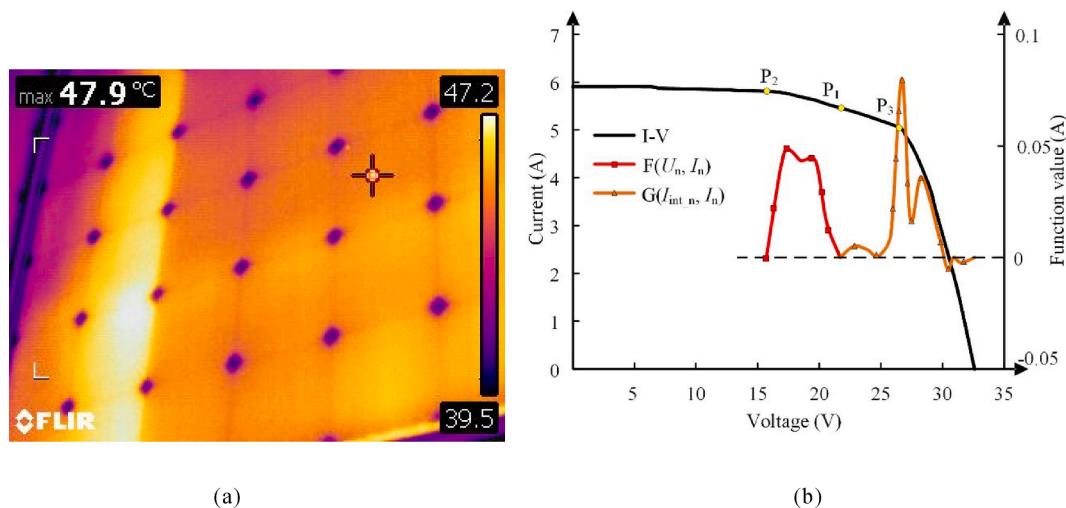


Fig. 24. Case study of an insignificant hot spot module. (a): IR image of the hot spot module; (b): Characteristics of the module.

the distribution of the value of $F(U_n, I_n)$ and $G(I_{int_n}, I_n)$ is also shown in Fig. 27 (a). The $F(U_n, I_n)$ has negative value, indicating a mismatch in the PV module. After obtaining the value of ΔI as 0.61 A, which exceeds the threshold of the partial shading fault, the linear regression fitting is also used with the corresponding result of the step data shown in the Fig. 27 (b). Since the fitting

result R^2 is 98.74%, it is misjudged as hot spot fault. In addition, when the cells in the PV module are slightly and uniformly cracked, the steps on the I-V curve are relatively flat, which may be misjudged as partial shading fault.

The I-V curve of the other PV module with all cells cracked is shown

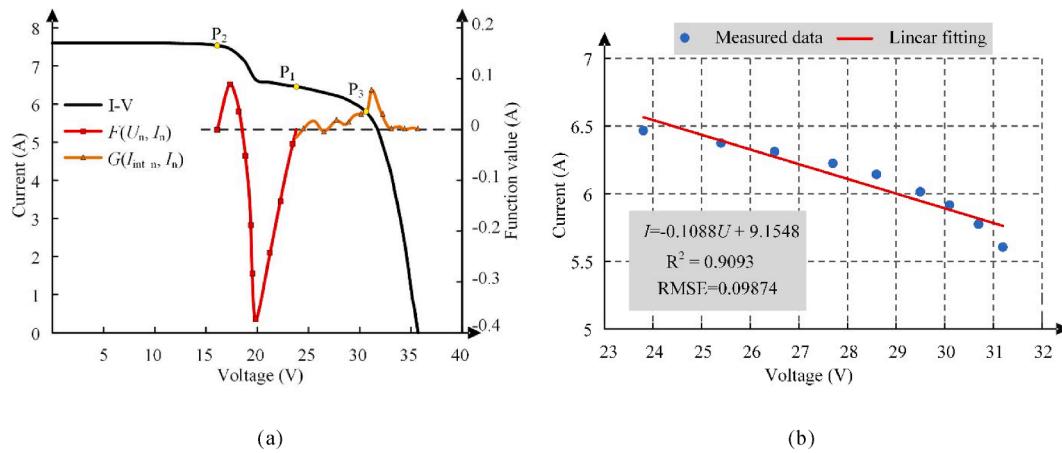


Fig. 25. Case study of a cracked module with a single substring crack. (a): Characteristics of the module; (b): Linear regression fitting of step data.

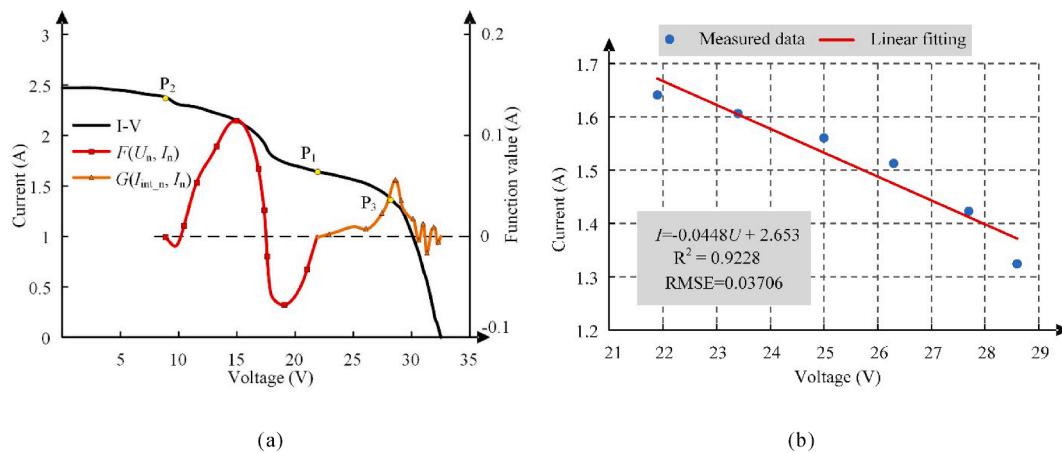


Fig. 26. Case study of a cracked module with multiple substrings cracked in different degrees. (a): Characteristics of the module; (b): Linear regression fitting of step data.

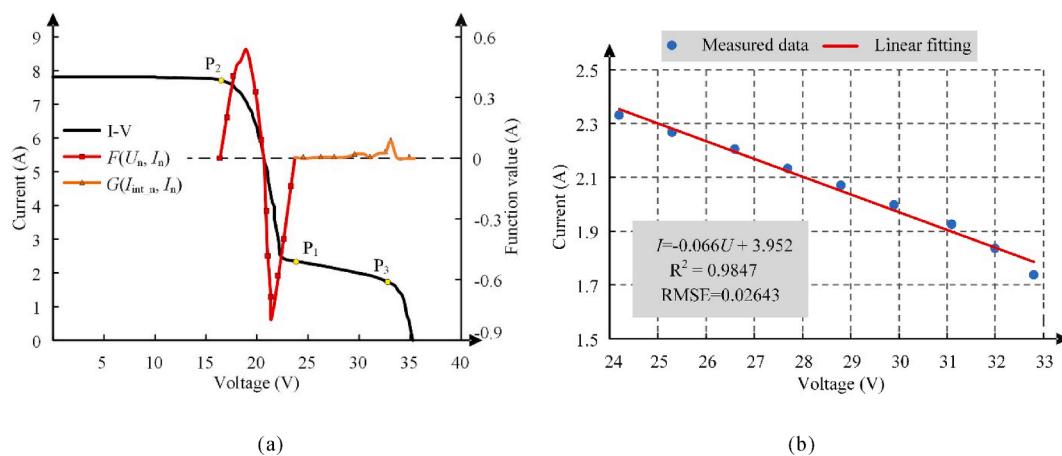


Fig. 27. Case study of a cracked module with only one cell cracked. (a): Characteristics of the module; (b): Linear regression fitting of step data.

in Fig. 28, where the degree of crack tends to be uniform, and the I-V curve has no step feature. The mismatch diagnosis is performed on the cracked module, and the values of $F(U_n, I_n)$ and $G(I_{int,n}, I_n)$ are also shown in Fig. 28. The result shows that $F(U_n, I_n)$ has no negative value and the module is diagnosed as non-mismatch. The crack fault of PV modules in long-term operation is more complicated and the mismatch fault can only be detected when the I-V curve has a concave feature.

Uniform crack will affect the accuracy of the diagnosis of cracked PV modules. At this time, it can be further considered to judge the crack fault based on the decrease of the short circuit current or the combination of multiple features.

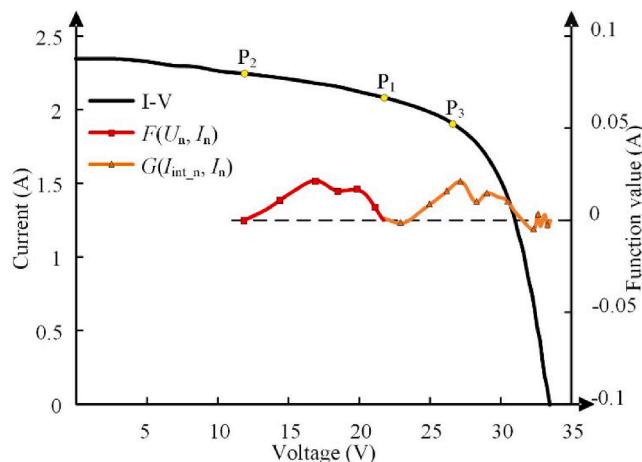


Fig. 28. Characteristics of the PV module with all cells cracked.

6. Conclusion

In this paper, several current mismatch faults in the actual PV power plant have been investigated, such as partial shading, hot spot and crack, and the I-V characteristic curves of each type of faulty PV module have been tested. When current mismatch fault exists in a certain substring, the bypass diode of the substring is turned on, and the I-V curve shows a step. The step features of different current mismatch types have been extracted. For widely used PV modules with three bypass diodes, a method for detecting mismatch by dividing the I-V curve and using feature points has been proposed. For the selected step data, statistical analysis and polynomial regression fitting methods are used to identify the partial shading, hot spot, and crack fault. Combined with the optimizer with I-V scanning function in the actual distributed system, the verification and case studies of different mismatch faults have been carried out. The experiment studies have well verified that the proposed fault diagnosis method has high accuracy as well as strong discrimination for PV module current mismatch fault diagnosis, and high sensitivity to small mismatch and significant practical application values.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Thanks to the funding support of the Project Supported by National Natural Science Foundation of China (Grant No.: 52061635101) and 111 Project (Grant No.: BP0719039). The authors would like to acknowledge the equipment provided by Sungrow Power Supply Co., Ltd. and the funding support of the No. 17 project of Central Research Institute.

References

- REN21, 2019. Renewables 2019 Global Status Report.
- Dhanup, S.P., Rajasekar, N., 2018. A comprehensive review on protection challenges and fault diagnosis in PV systems. *Renew. Sustain. Energy Rev.* 91, 18–40.
- Asma, T.L., Afef, B.B., Ilhem, S.B., 2018. Fault detection and monitoring systems for photovoltaic installations: a review. *Renew. Sustain. Energy Rev.* 82, 2680–2692.
- Firth, S.K., Lomas, K.J., Rees, S.J., 2010. A simple model of PV system performance and its use in fault detection. *Sol. Energy* 84 (4), 624–635.
- Deng, Shifeng, Zhang, Zhen, Chenhui, Ju, et al., 2017. Research on hot spot risk for high-efficiency solar module. *Energy Procedia* 130, 77–86.
- Mellit, A., Tina, G.M., Kalogirou, S.A., 2018. Fault detection and diagnosis methods for photovoltaic systems: a review. *Renew. Sustain. Energy Rev.* 91, 1–17.
- Tingting, P., Jiangfeng, Z., Li, L., et al., 2020. A fault locating method for PV arrays based on improved voltage sensor placement. *Sol. Energy* 201, 279–297.
- Siva, R.M., Singh, S.N., 2017. Online modular level fault detection algorithm for grid-tied and off-grid PV systems. *Sol. Energy* 157, 349–364.
- Haizheng, W., Jian, Z., Qian, S., et al., 2019. Probability modeling for PV array output interval and its application in fault diagnosis. *Energy* 189, 116248.
- Aref, E., Jafar, M., Mohammadreza, A., 2020. Line-line fault detection and classification for photovoltaic systems using ensemble learning model based on I-V characteristics. *Sol. Energy* 211, 354–365.
- Akram, M.W., Guiqiang, L., Yi, J., et al., 2019. Improved outdoor thermography and processing of infrared images for defect detection in PV modules. *Sol. Energy* 190, 549–560.
- Sheikh, A.R., Tania, U., David, A.P., 2020. PV system defects identification using Remotely Piloted Aircraft (RPA) based infrared (IR) imaging: a review. *Sol. Energy* 206, 579–595.
- Zhicong, C., Fuchuan, H., Lijun, W., et al., 2018. Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents. *Energy Convers. Manage.* 178, 250–264.
- Li, B., Delpha, C., Diallo, D., et al., 2021. Application of Artificial Neural Networks to photovoltaic fault detection and diagnosis: a review. *Renew. Sustain. Energy Rev.* 138, 110512.
- Rabah, B., Samir, M., 2018. Fault detection and diagnosis based on C4.5 decision tree algorithm for grid connected PV system. *Sol. Energy* 173, 610–634.
- Stegner, C., Dalsass, M., Luchascheider, P., et al., 2018. Monitoring and assessment of PV generation based on a combination of smart metering and thermographic measurement. *Sol. Energy* 163, 16–24.
- Sarikh, S., Raoufi, M., Bennouna, A., et al., 2018. Fault diagnosis in a photovoltaic system through I-V characteristics analysis. In: The 9th International Renewable Energy Congress (IREC 2018), Hammamet, Tunisia, pp. 1–6.
- IEC62446, 2016. Photovoltaic (PV) systems-Requirements for testing, documentation and maintenance.
- Spataru, S., Sera, D., Kerekes, T., et al., 2015. Diagnostic method for photovoltaic systems based on light I-V measurements. *Sol. Energy* 119, 29–44.
- Manit, S., Yoshihiro, H., Masahiro, Y., et al., 2020. Detection of shading effect by using the current and voltage at maximum power point of crystalline silicon PV modules. *Sol. Energy* 211, 1365–1372.
- Fadhel, S., Delpha, C., Diallo, D., et al., 2019. PV shading fault detection and classification based on I-V curve using principal component analysis: application to isolated PV system. *Sol. Energy* 179, 1–10.
- IEC61215, 2005. Crystalline silicon terrestrial photovoltaic modules-Design qualification and type approval.
- Bakhsh, H., Kazutaka, I., 2016. Real time hot spot detection using scan-method adopted with P&O MPPT for PV generation system. In: 2016 IEEE 2nd Annual Southern Power Electronics Conference (SPEC). Auckland, New Zealand, pp. 1–5.
- Hamed, H., Matthias, P., Bengt, J., et al., 2019. A novel electrical approach to protect PV modules under various partial shading situations. *Sol. Energy* 193, 814–819.
- Pierluigi, G., Pietro, T., Santolo, D., 2019. A bypass circuit for avoiding the hot spot in PV modules. *Sol. Energy* 181, 430–438.
- Assmus, M., Jack, S., Weiss, K.A., Koehl, M., 2011. Measurement and simulation of vibrations of PV-modules induced by dynamic mechanical loads. *Prog. Photovoltaics Res.* 19, 688–694.
- Kajari-Schröder, S., Kunze, I., Eitner, U., Köntges, M., 2011. Spatial and orientational distribution of cracks in crystalline photovoltaic modules generated by mechanical load tests. *Sol. Energy Mater. Sol. Cells*. 95, 3054–3059.
- Mahmoud, D., Violeta, H., Bruce, M., et al., 2017. The impact of cracks on photovoltaic power performance. *J. Sci.: Adv. Mater. Devices* 2, 199–209.
- Sara, G.S., Luis, H.C., Marfa, C.A., et al., 2020. Nondestructive characterization of solar PV cells defects by means of electroluminescence, infrared thermography, I-V curves and visual tests: experimental study and comparison. *Energy* 205, 117930.
- Bishop, J.W., 1988. Computer simulation of the effects of electrical mismatch in photovoltaic cell interconnection circuits. *Sol. Cells*. 25 (1), 73–79.
- Shimizu, T., Kamezawa, M., Kamezawa, T., et al., 2011. Generation control circuit for photovoltaic modules. *IEEE Trans. Power Electron.* 16, 293–300.
- Hirata, Y., Noro, S., Aoki, T., Miyazawa, S., 2012. Diagnosis photovoltaic failure by simple function method to acquire I-V curve of photovoltaic modules string. In: 2012 38th IEEE Photovoltaic Specialists Conference. Austin, Texas, USA, pp. 1340–1343.
- Simon, M., Meyer, E., 2010. Detection and analysis of hot spot formation in solar cell. *Sol. Energy Mater. Sol. Cells* 94 (2), 106–113.
- Bishop, J.W., 1989. Microplasma breakdown and hot-spots in silicon solar cells. *Solar Cells*. 26 (4), 335–349.
- Pingel, S., Zemen, Y., Frank, O., Geipel, T., Berghold, J., 2009. Mechanical stability of solar cells within solar panels. In: Proc. of Eur. Photovoltaic Sol. Energy Conf.. Dresden, Germany, pp. 3459–3464.
- Morlier, A., Haase, F., Köntges, M., 2015. Impact of cracks in multicrystalline silicon solar cells on PV module power-a simulation study based on field data. *IEEE J. Photovoltaics* 5 (6), 1735–1741.
- Köntges, M., Siebert, M., Illing, R., Wegert, F., 2014. Influence of photovoltaic module handling on solar cell cracking. In: Proc. of 28th Eur. Photovoltaic Sol. Energy Conf.. Munich, Germany, pp. 2276–2282.
- Köntges, M., Siebert, M., et al., 2013. Impact of transportation on silicon wafer-based PV Modules. In: Proc. of 28th Eur. Photovoltaic Sol. Energy Conf.. Paris, France, pp. 2960–2967.

Käsewieter, J., Haase, F., Haro, Larrode M., Köntges, M., 2014. Cracks in solar cell metallization leading to module power loss under mechanical load. *Energy Procedia* 55, 469–477.

Arce, G.R., 2005. Nonlinear Signal Processing: A Statistical Approach. John Wiley & Sons.

Helzer, A., Barzohar, M., Malah, D., 2004. Stable fitting of 2D curves and 3D surfaces by implicit polynomials. *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (10), 1283–1294.