FAULT DETECTION, CLASSIFICATION AND PROTECTION IN SOLAR PHOTOVOLTAIC ARRAYS

A Dissertation Presented

by

Ye Zhao

to

The Department of Electrical and Computer Engineering

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in the field of

Electrical Engineering

Northeastern University
Boston, Massachusetts

August 2015
Acknowledgments

First of all, I would like to express my deepest appreciation to my research advisor, Prof. Brad Lehman, who gave me this great research opportunity. I am heartily thankful to Prof. Lehman for his tireless guidance, extraordinary patience and constant encouragement during my doctoral journey. His knowledge, creativity and passion not only inspire me during my graduate studies, but also enlighten my future career. Without his persistent help, this dissertation would not have been possible.

I would like to thank my committee members, Professor Jennifer Dy and Dr. Peng Li who examine my dissertation and provide me so many valuable suggestions.

I am grateful to Mersen USA Newburyport-MA, LLC for the research grant that funded my work for all these years. I would like to thank the people at Mersen, Dr. Jean-François de Palma, Roy Ball, Jerry Mosesian, Ed Knapp, Craig McKenzie and Robert Lyons, who gave me valuable suggestions and great help in my research. In addition, I sincerely thank the people at Mersen France SB SAS, including, but not limited to, Thierry Arnaud, Matthieu Laroche, and Andras Pozsgay for their kind support during my internship in Annecy Le Vieux, France. Special thanks go to Florent Balboni who offered me great support on the solar photovoltaic experiments in this dissertation.

I am very fortunate to meet great friends during my graduate studies in Boston. I would like to thank all my colleagues from the power electronics research group:
Song Chen, Renato Nakagomi, Chung-Ti Hsu, Stephanie Quinn, Su Sheng, Qian Sun, Jen-Hung Huang, Dawood Talebi, Yue Zheng and Prof. Dorin O. Neacsu, who provided me with their friendly help. I would like to thank my friends Chenfu Guo, Gen Wang, Haixiao Yan, Xia Li, Boyang Hou, Kaiyu Zhao, Zijun Yao, Chao Chen, Yue Huang, Mert Korkalı, Liuxi (Calvin) Zhang and many more, for having numerous memorable moments in my life.

I am also grateful to Lindsay Sorensen for her advice about writing and proofread of my publications.

Finally, I am grateful to all my family members, especially my wife, Ling Yang, my son, Luke Zhao, my father, Heping Zhao, my mother, Ling Li, and my father-in-law, Zhongqing Yang for always believing in me and encouraging me to do my best. All of their love, support, and sacrifices over these past years gave me the strength and determination to make my dreams come true.
# Contents

List of Figures vii

List of Tables x

List of Acronyms xi

Abstract xiv

1 Introduction 1

1.1 Motivation and Background ........................................ 1
1.2 Problem Statement .................................................. 4
1.3 Dissertation Organization ......................................... 6
1.4 Contributions ....................................................... 8

2 Literature Review 11

2.1 Overview of Faults in Solar PV Arrays .......................... 11
2.2 Protection Gap of Conventional Solutions ....................... 13
  2.2.1 Ground Fault Detection Interrupters ...................... 13
  2.2.2 Overcurrent Protection Devices ........................... 17
  2.2.3 Summary ...................................................... 18
2.3 Existing Fault Detection, Classification and Location Solutions 19
  2.3.1 Quantitative-Model Based Solutions ....................... 21
    2.3.1.1 Current-Voltage (I-V) Analysis .................. 21
    2.3.1.2 Performance Comparison .......................... 21
    2.3.1.3 Performance Ratio (PR) .......................... 24
    2.3.1.4 Capture Loss Analysis ............................ 26
  2.3.2 Process-History Based Solutions ........................... 27
    2.3.2.1 Statistical Methods ................................. 27
    2.3.2.2 Machine Learning .................................... 28
  2.3.3 Signal-Processing Based Solutions ........................ 29
    2.3.3.1 Dc Arc-Fault Circuit Interrupter (AFCI) .......... 29
    2.3.3.2 Insulation Resistance Monitor ..................... 31
    2.3.3.3 Residual Current Detector (RCD) .................. 32
2.3.3.4 Time Domain Reflectometry (TDR) . . . . . . . . 34
2.3.3.5 Spread Spectrum Time Domain Reflectometry (SSTDR) 36
2.3.3.6 Infrared (IR) Thermography . . . . . . . . . . . . . 37
2.3.3.7 Internal Series Resistance . . . . . . . . . . . . . . 38

2.4 Existing Fault Protection Solutions . . . . . . . . . . . . . . . . . 39
2.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41

3 Fault Studies in a Commercial-Scale PV Laboratory 42
3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42
3.2 Simulation of Line-Line Faults in PV systems . . . . . . . . . . . 44
  3.2.1 Typical Grid-Connected PV Systems . . . . . . . . . . . . . 44
  3.2.2 Simulation Parameters . . . . . . . . . . . . . . . . . . . . . . 46
  3.2.3 Fault Analysis using $I-V$ Curves and MPPT Effects . . . . 47
  3.2.4 PV Dynamic Characteristics . . . . . . . . . . . . . . . . . . 50
  3.2.5 Fault Evolution in Time Domain . . . . . . . . . . . . . . . 51
  3.2.6 More Simulation Results and Discussion . . . . . . . . . . . 53
3.3 Experimental Setup in the PV Laboratory . . . . . . . . . . . . . . 55
3.4 Experimental Results . . . . . . . . . . . . . . . . . . . . . . . . . . 58
  3.4.1 Line-Line Fault with 2-Module Difference ($LL-2$) . . . . . 60
  3.4.2 Line-Line Fault with 5-Module Difference ($LL-5$) . . . . . 62
  3.4.3 Line-Line Fault with 7-Module Difference ($LL-7$) . . . . . 64
  3.4.4 More Experimental Results and Discussion . . . . . . . . . . 66
3.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 68

4 Fault Detection Using Outlier Detection Rules 70
4.1 Features for Fault Detection . . . . . . . . . . . . . . . . . . . . . . 71
  4.1.1 String Current . . . . . . . . . . . . . . . . . . . . . . . . . . 72
  4.1.2 String Current Change Rate . . . . . . . . . . . . . . . . . . 72
4.2 Proposed ODR for PV systems . . . . . . . . . . . . . . . . . . . . 73
  4.2.1 Statistical ODR . . . . . . . . . . . . . . . . . . . . . . . . . 73
  4.2.2 Local Outlier Factor (LOF) . . . . . . . . . . . . . . . . . . 77
4.3 Fault Detection in the PV Laboratory . . . . . . . . . . . . . . . . 79
  4.3.1 Statistical ODR Results . . . . . . . . . . . . . . . . . . . . . 79
    4.3.1.1 ODR on String Current $i$ . . . . . . . . . . . . . . . . 80
    4.3.1.2 ODR on String Current Change Rate $di/dt$ . . . . . . 80
    4.3.1.3 False alarms . . . . . . . . . . . . . . . . . . . . . . . 80
  4.3.2 LOF Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . 84
  4.3.3 More Results under Various Solar Irradiance . . . . . . . . . 86
4.4 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 86
4.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 91

5 Fault Classification Using Machine-Learning Algorithms 92
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>92</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Existing Solutions and their limitations</td>
<td>93</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Research Contributions Obtained in this Chapter</td>
<td>95</td>
</tr>
<tr>
<td>5.2</td>
<td>Fundamentals of Machine-Learning Algorithms</td>
<td>97</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Fundamentals of Machine Learning Techniques</td>
<td>97</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Machine Learning Techniques in PV Systems</td>
<td>98</td>
</tr>
<tr>
<td>5.3</td>
<td>Improved Data Visualization using Normalized Parameters</td>
<td>101</td>
</tr>
<tr>
<td>5.4</td>
<td>Graph-based Semi-supervised Learning (GBSSL) Algorithm in Solar PV Arrays</td>
<td>104</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Introduction of Graph Data in PV Arrays</td>
<td>104</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Detailed GBSSL Algorithm for FDC</td>
<td>106</td>
</tr>
<tr>
<td>5.5</td>
<td>Simulation Results</td>
<td>110</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Simulated PV Systems</td>
<td>110</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Faults in PV Systems</td>
<td>111</td>
</tr>
<tr>
<td>5.5.3</td>
<td>Simulation of GBSSL for FDC</td>
<td>113</td>
</tr>
<tr>
<td>5.5.4</td>
<td>Simulation Results</td>
<td>119</td>
</tr>
<tr>
<td>5.5.5</td>
<td>Discussion</td>
<td>122</td>
</tr>
<tr>
<td>5.6</td>
<td>FDC Experiments in PV Systems</td>
<td>123</td>
</tr>
<tr>
<td>5.6.1</td>
<td>Experimental Setup</td>
<td>123</td>
</tr>
<tr>
<td>5.6.2</td>
<td>Experimental Results</td>
<td>126</td>
</tr>
<tr>
<td>5.6.3</td>
<td>Discussion</td>
<td>129</td>
</tr>
<tr>
<td>5.7</td>
<td>Conclusion</td>
<td>130</td>
</tr>
<tr>
<td>6</td>
<td>Conclusions and Future Work</td>
<td>131</td>
</tr>
<tr>
<td>6.1</td>
<td>Conclusions</td>
<td>131</td>
</tr>
<tr>
<td>6.2</td>
<td>Future Work</td>
<td>139</td>
</tr>
</tbody>
</table>

Bibliography                                                                                      142

List of Publications                                                                             156
List of Figures

1.1 Fire hazards in PV installations caused by faults. .......................... 3
1.2 Excessive data are available in PV installations. ......................... 4
1.3 Schematic diagram of a grid-connected PV system, including various types of faults in the PV array. .................................................. 5

2.1 Common faults in PV arrays (dc side). ........................................ 12
2.2 Illustrative example of GFDI in grounded PV systems [12]. ........... 13
2.3 The first undetected ground fault between a grounded current-carrying conductor (i.e., the negative conductor) and the equipment-grounding conductor [3]. .................................................. 15
2.4 The second ground fault between a ungrounded current-carrying conductor (i.e., the positive conductor) and the equipment-grounding conductor [3]. .................................................. 16
2.5 The photo and melting time vs. current curve of fuses. ................. 18
2.6 Overview of general solutions to increase PV-system reliability and safety. ................................................................. 19
2.7 The I-V curve reveals salient PV characteristics. ....................... 21
2.8 Performance comparison of a PV system. ................................. 22
2.9 System overview of PV fault detection using performance comparison [27] ................................................................. 23
2.10 Performance ratio in a grid-connected PV system. ....................... 24
2.11 Daily, weekly, and monthly PR values for a PV system for 2001 [31]. 25
2.12 Supervised learning. ........................................................... 28
2.13 The dc arc experimental setup at Northeastern University. .......... 30
2.14 Normal vs. series dc-arc current in frequency domain. ............. 31
2.15 Insulation resistance monitoring for ground-fault detection in ungrounded PV systems [31] ............................................................... 32
2.16 Concept of differential relays for internal fault detection [51] ....... 33
2.17 RCD application in solar PV arrays. ......................................... 34
2.18 Schematic diagram for TDR setup in the PV field. [52] ............... 35
2.19 Step input and the reflected waveforms with cable. [53] ............. 35
2.20 TDR on open-circuit and short-circuit PV strings (Use a negative step input). ................................................................. 36
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.21</td>
<td>Fault detection in PV arrays using Infrared (IR) thermography.</td>
<td>37</td>
</tr>
<tr>
<td>2.22</td>
<td>The equivalent circuit of the one-diode model for a PV module.</td>
<td>38</td>
</tr>
<tr>
<td>3.1</td>
<td>Line-line faults in a PV array with a centralized PV inverter.</td>
<td>45</td>
</tr>
<tr>
<td>3.2</td>
<td>Schematic diagram of the simulated PV system.</td>
<td>46</td>
</tr>
<tr>
<td>3.3</td>
<td>Simulated I-V curves of the PV array under line-line faults (LL-2 and LL-5).</td>
<td>48</td>
</tr>
<tr>
<td>3.4</td>
<td>Simulated $v_{pe}$ vs. $g_{pe}$ curves of PV strings under normal, and line-line faults, such as LL-1 to LL-5.</td>
<td>51</td>
</tr>
<tr>
<td>3.5</td>
<td>Simulated PV string current $i_j$ ($j = 1$ to $12$) evolution during LL-2.</td>
<td>52</td>
</tr>
<tr>
<td>3.6</td>
<td>The photo of the PV laboratory located in France.</td>
<td>55</td>
</tr>
<tr>
<td>3.7</td>
<td>Schematic diagram of the PV laboratory.</td>
<td>56</td>
</tr>
<tr>
<td>3.8</td>
<td>The photo of major experimental equipments.</td>
<td>57</td>
</tr>
<tr>
<td>3.9</td>
<td>The photo of PV monitoring board [71].</td>
<td>58</td>
</tr>
<tr>
<td>3.10</td>
<td>Experimental results of LL-2 (2-module LL fault).</td>
<td>61</td>
</tr>
<tr>
<td>3.11</td>
<td>Experimental results of LL-5 (5-module LL fault).</td>
<td>63</td>
</tr>
<tr>
<td>3.12</td>
<td>Experimental results of LL-7 (7-module LL fault).</td>
<td>65</td>
</tr>
<tr>
<td>4.1</td>
<td>Univariate normal distribution of PV string currents ${x_i}$.</td>
<td>75</td>
</tr>
<tr>
<td>4.2</td>
<td>Apply statistical ODR on string current $i_1$ to $i_6$ of LL-2 experimental results (which cannot be cleared by fuses).</td>
<td>81</td>
</tr>
<tr>
<td>4.3</td>
<td>Apply statistical ODR on $di_1/dt$ to $di_6/dt$ of LL-2 experimental results (which cannot be cleared by fuses).</td>
<td>82</td>
</tr>
<tr>
<td>4.4</td>
<td>False alarms caused by ODR on string current.</td>
<td>83</td>
</tr>
<tr>
<td>4.5</td>
<td>False alarms caused by ODR on current change rate.</td>
<td>83</td>
</tr>
<tr>
<td>4.6</td>
<td>Fault detection using LOF on experimental results of LL-2 and normal conditions.</td>
<td>85</td>
</tr>
<tr>
<td>4.7</td>
<td>LL-2 under medium solar irradiance ($\sim 410W/m^2$).</td>
<td>87</td>
</tr>
<tr>
<td>4.8</td>
<td>LL-2 under low solar irradiance ($\sim 170W/m^2$).</td>
<td>88</td>
</tr>
<tr>
<td>4.9</td>
<td>False alarms may be caused by partial shading on PV strings.</td>
<td>90</td>
</tr>
<tr>
<td>5.1</td>
<td>Limitation of supervised learning: SVM model trained in summer will not work properly in winter.</td>
<td>94</td>
</tr>
<tr>
<td>5.2</td>
<td>Simplified flowchart of machine learning techniques: (a) supervised, (b) unsupervised learning, (c) semi-supervised learning.</td>
<td>98</td>
</tr>
<tr>
<td>5.3</td>
<td>Overview of the proposed fault detection and classification (FDC) model for PV systems.</td>
<td>99</td>
</tr>
<tr>
<td>5.4</td>
<td>The overlapping MPPs of a PV array over a wide range of irradiance and temperature. This is difficult for FDC.</td>
<td>102</td>
</tr>
<tr>
<td>5.5</td>
<td>$V_{NORM}$ vs. $I_{NORM}$ of a PV array over a wide range of irradiance and temperature. This is better for FDC.</td>
<td>103</td>
</tr>
<tr>
<td>5.6</td>
<td>A graph data consisting of edges and nodes.</td>
<td>104</td>
</tr>
</tbody>
</table>
List of Figures

5.7 An illustrative example of GBSSL for PV: (a) Initial labels $x_1$ and $x_2$ with unlabeled PV data. (b) Unlabeled data are classified by GBSSL. .................................................. 105
5.8 Overview of the proposed GBSSL model for FDC in PV systems. . 107
5.9 Line-line faults and open-circuit faults in a simulated PV system. . 111
5.10 An illustrative example: new data points $x_1$ and $x_3$ will be classified as classes NORMAL and OPEN respectively, by the proposed GBSSL model. Similarly, data points $x_2$ and $x_4$ will be classified as LL. ................................................................. 118
5.11 New fault data: “OPEN3” and “OPEN4” can be identified successfully, but “LL 10% $R_f=0$”, “LL 10% $R_f=10$”, “LL 20% $R_f=0$” and “LL 20% $R_f=10$” may be difficult to detect since they are overlapping with “NORMAL”. ................................................................. 120
5.12 Schematic diagram of the experimental PV system. .................. 124
5.13 Photo of the experimental PV system. ................................. 125
5.14 Initial labels for GBSSL in experiments. ............................... 127
5.15 GBSSL results of NORMAL, LL and OPEN1. ....................... 128

6.1 Schematic diagram of a grid-connected PV system, including conventional GFDI and OCPD. ..................................................... 139
List of Tables

2.1 Maximum allowable ground current detection settings [15]. . . . . . 14
2.2 Summary of existing PV fault detection, classification and location 20
2.3 Summary of existing PV fault protection . . . . . . . . . . . . . . . 40

3.1 Simulation results of various line-line faults at String #1 under STC 54
3.2 Main Parameters of Experimental PV Laboratory . . . . . . . . . . 57
3.3 Summary of Experimental Results . . . . . . . . . . . . . . . . . . . 67

4.1 A brief comparison between the statistical ODR and LOF in PV fault detection . . . . . . . . . . . . . . . . . . . . . 89

5.1 Summary results of fault detection and classification . . . . . . . . 121
5.2 Experimental PV Components . . . . . . . . . . . . . . . . . . . . . 125
5.3 Experimental Results of Fault Detection and Classification . . . . 129
List of Acronyms

AFCI  Arc-Fault Circuit Interrupter  
ANN  Artificial Neural Network  
BNN  Bayesian Neural Network  
DWT  Discrete Wavelet Transform  
FDC  Fault Detection and Classification  
FFT  Fast Fourier Transform  
GBSSL  Graph-Based Semi-Supervised Learning  
GFDI  Ground Fault Detection Interrupter  
GUI  Graphical User Interface  
IR  Infrared  
I-V  Current vs. Voltage  
LL  Line-Line fault  
LOF  Local Outlier Factor  
MPP  Maximum Power Point  
MPPT  Maximum Power Point Tracking
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEC</td>
<td>National Electrical Code</td>
</tr>
<tr>
<td>OCPD</td>
<td>Overcurrent Protection Device</td>
</tr>
<tr>
<td>ODR</td>
<td>Outlier Detection Rule</td>
</tr>
<tr>
<td>PR</td>
<td>Performance Ratio</td>
</tr>
<tr>
<td>PV</td>
<td>Photovoltaic</td>
</tr>
<tr>
<td>RCD</td>
<td>Residual Current Detector</td>
</tr>
<tr>
<td>SSL</td>
<td>Semi-Supervised Learning</td>
</tr>
<tr>
<td>SSTDR</td>
<td>Spread Spectrum Time Domain Reflectometry</td>
</tr>
<tr>
<td>STC</td>
<td>Standard Test Condition</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>TDR</td>
<td>Time Domain Reflectometry</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
</tbody>
</table>
Dedicated to Ling and Luke
Abstract

Fault analysis and fault detection are important to the efficiency, safety and reliability of solar photovoltaic (PV) systems. Despite the fact that PV systems have no moving parts and usually require low maintenance, they are still subject to various fault conditions. Especially for PV arrays (dc side), it is difficult to shut down PV modules completely during faults, since they are always energized by sunlight in daytime. Furthermore, conventional series-parallel PV configurations increase voltage and current ratings, leading to higher risk of large fault currents or dc arcs.

This dissertation reviews the challenges and limitations of existing fault detection and protection solutions in solar PV arrays. For the first time, a 35kW commercial-scale PV laboratory is designed to study faults under real-working conditions and to discover the “blind spots” in conventional fault protection schemes. It is shown that the line-line fault may not be detectable by traditional overcurrent protection devices (OCPD) under certain conditions. Therefore, the fault may remain in the PV system as a safety concern.

To eliminate the detection “blind spot,” outlier rules, such as statistical outlier detection rules (ODRs) and local outlier factors (LOFs) are proposed in PV-string monitoring systems. To further identify the fault types (or so-called fault classification), machine learning algorithms are studied in solar PV arrays. To overcome the drawbacks of supervised learning algorithms, a semi-supervised learning algorithm is proposed. The dissertation demonstrates the effectiveness in fault detection and classification in both simulation and experimental results.
Chapter 1

Introduction

1.1 Motivation and Background

Fault analysis, detection and protection are essential to prevent unexpected events in solar photovoltaic (PV) systems. Despite the fact that solar PV systems have no moving parts and usually require low maintenance, they are still subject to various failures or faults along the PV arrays, power conditioning units, batteries, wiring, and utility interconnections [1, 2]. Especially for PV arrays (dc side), it is difficult to shut down PV modules completely during faults, since they are energized by sunlight in daytime. Furthermore, PV is scalable and modular technology that can build a PV power plant by connecting a large number of PV modules in series and parallel configuration. Once PV modules are electrically connected, any fault among them can affect the entire system performance. This means the PV system
is only as robust as its weakest link (e.g., the faulted PV components). In a large PV array, it may become difficult to properly detect or identify a fault, which can remain hidden in the PV system until the whole system breaks down. In addition, conventional series-parallel PV configurations increase voltage and current ratings, leading to higher risk of large fault currents or dc arcs.

Due to faults occurring within PV arrays, several fire hazards have been reported in PV installations [3–6]. Fig. 1.1(a) shows the results of a fire hazard in a 383 kW PV array in Bakersfield, California in 2009 [3, 4]. Another fire hazard is illustrated in Fig. 1.1(b), which occurred in a 1 MW PV power plant in Mount Holly, North Carolina, in 2011 [6]. In these cases, the fault remained unnoticed and hidden in the system until the hazard caused catastrophic fire. These fire hazards not only show the weakness in conventional fault detection and protection schemes in PV arrays, but also reveal the urgent need of a better way to prevent such issues.

Nowadays, due to the growing capacity of PV systems, there has been a proliferation of power conversion units, monitoring systems, protection devices and communication equipment being added to PV installations. As a result, excessive PV data becomes available (both instantaneous and historical). For example, as shown in Fig. 1.2 as a typical grid-connected PV system, various PV data are available from the weather station, PV arrays, PV inverters and utility grid. These PV data are mainly used to evaluate the PV system performance and calculate the energy losses over long periods of times. Although fault detection has been developed using historical PV data, it requires a long process time (at least a
Chapter 1. **Introduction**

(a) Fire hazards in a 383kW PV array, in Bakersfield, California, in 2009 [3].

(b) Fire hazards in a 1,208kW PV array, in Mount Holly, North Carolina, in 2011 [6].

![Image of PV array damage]

**Figure 1.1:** Fire hazards in PV installations caused by faults.

Few hours or days) that may hinder the fault detection response and effectiveness. Hence, it is necessary to develop more responsive fault-detection algorithms that can make better use of these readily available PV data.
Fig. 1.2: Excessive data are available in PV installations.

1.2 Problem Statement

Fig. 1.3 demonstrates a typical grid-connected PV system including a PV array with a number of PV modules in series and parallel connection, a central PV inverter with maximum power point tracker (MPPT), overcurrent protection devices (OCPD) and ground fault detection interrupters (GFDI). Several types of fault could happen inside PV arrays, such as line-line faults, ground faults, open-circuit faults, and mismatch faults. Among these faults, line-line faults and ground faults are the most common faults in solar PV arrays, which potentially involve large fault current or dc arcs. As shown in Fig. 1.3, conventional fault detection and
Chapter 1. **Introduction**

Protection methods usually add OCPD (e.g., fuses) and GFDI within PV components [7] to prevent PV components from large fault current. However, it has been shown that certain faults in PV arrays may not be cleared by OCPD or GFDI due to non-linear output characteristics of PV arrays, PV current-limiting nature, high fault impedances, low irradiance conditions, PV grounding schemes, or MPPT of PV inverters [1, 8]. This difficulty brings “blind spots” in the protection schemes, leading to reduced system efficiency, accelerated system aging, dc arcs and similar fire hazards reported previously in [3, 8].

**Figure 1.3:** Schematic diagram of a grid-connected PV system, including various types of faults in the PV array.
1.3 Dissertation Organization

Motivated by the previously discussed research problems, this dissertation aims to analyze and explain the limitations of conventional OCPD and to propose new fault detection and classification methods. Therefore, we can eliminate the fault detection gap in solar PV arrays. Specifically, the dissertation is organized as follows.

The results for the literature review are presented in Chapter 2. Chapter 3 discusses the fault experimental results in the PV laboratory. Chapter 3 focuses on the proposed fault detection methods using statistical outlier detection rules \((ODRs)\) and local outlier factors \((LOFs)\). To classify the fault types in PV arrays, Chapter 5 develops the machine learning algorithms that detect the fault occurrence, and also classify the fault types. In Chapter 6, a brief summary of the research results is presented and future research works are addressed.

Chapter 2: Literature review

- The existing fault detection and classification solutions for solar PV arrays are studied.
- The benefits and drawbacks of existing approaches are explored and compared.

Chapter 3: Fault Studies in a Commercial-Scale PV Laboratory
• The fault behavior in a large grid-connected PV system consisting of more than 100 PV modules is simulated and analyzed.

• The fault experiments are implemented in a commercial-scale PV system, which is able to create and record faults.

• The transient response and post-fault behavior of a PV system are studied during the fault evolution, including PV array, overcurrent protection devices (OCPD, i.e., fuses) and grid-connected PV inverter.

• The success and failure of fault clearance by OCPD (i.e., fuses) are analyzed.

• The challenges to fault detection and protection and their blind spots are discussed.

**Chapter 4: Fault Detection Using Outlier Detection Rules**

• Statistical outlier detection rules are proposed for real-time fault detection on PV-string level, such as $3\sigma$ rule, Hampel identifier and Boxplot rule. The results of statistical outlier detection rules and their limitations (such as false alarms) are discussed.

• The proposed fault detection algorithms have been integrated with commercial products using Matlab graphical user interface (GUI).
Chapter 1. Introduction

- Another method using local outlier factor (LOF) is used and implemented for solar PV arrays. The performance of the statistical ODR and LOF are discussed and compared.

Chapter 5: Fault Classification Using Machine-Learning Algorithms

- The limitations of supervised learning algorithms in for PV fault detection and classification are explained.

- Normalized parameters are created as the classification features, which can result in better data clustering.

- A graph-based semi-supervised learning algorithm is proposed for fault detection and classification.

- The proposed method is evaluated in PV simulation.

- The proposed method is implemented and verified on experimental PV data.

Chapter 6: Conclusion and Future Research

- The conclusions and future research are presented.

1.4 Contributions

This dissertation demonstrates several major research contributions and scientific advancements over the existing solutions.
Contribution 1: Discovery of OCPD “blind spots” in PV systems. The fault transient response and post-fault behavior in PV arrays (dc side) are implemented and then demonstrated in a 35kW, 500Vdc commercial-scale PV laboratory, including PV array, overcurrent protection devices (OCPD, i.e., fuses) and grid-connected PV inverter. Meanwhile, “blind spots” in conventional fault detection and protection schemes are discovered and explained in both simulation and experimental results.

Contribution 2: Outlier detection rules are proposed for fault detection in PV arrays. New fault detection using outlier detection rules are proposed to fill the fault detection gap of PV systems, which are only based on the PV-string current measurement. Compared to traditional methods, the proposed methods have several advantages, such as no requirement of weather information, quick response and easy implementation.

Contribution 3: Machine learning algorithms are proposed and developed for fault detection and classification in solar PV arrays. The proposed machine learning algorithm focuses on semi-supervised learning (SSL) algorithms, showing several advantages over existing supervised learning algorithms. The proposed method is able to be integrated within most PV inverter topologies in order to take
advantage of readily available measurements in existing PV systems.
Chapter 2

Literature Review

This chapter presents the literature review of various types of faults in solar PV arrays, protection gaps of conventional solutions, as well as existing approaches of fault detection, classification, location and protection solutions. The benefits and limitations of these existing approaches are also explored and compared.

2.1 Overview of Faults in Solar PV Arrays

A PV system may be subjected to a variety of faults, including those in PV array, power conditioning unit, and utility grid. This dissertation focuses on PV arrays, which exhibit non-linear and limited current vs. voltage (I-V) characteristics that are differ from other traditional dc or ac power sources. Therefore, faults in PV arrays require careful evaluation and special consideration. As shown in Fig. 2.1,
a variety of faults are commonly found inside the PV array. In this dissertation, it is considered that the PV array is the only source of fault current, since most PV inverters provide galvanic isolation between PV arrays and utility grids. Among these faults, ground faults and line-line faults have the most potential to cause large fault current through the fault path [9]. Without proper fault detection or protection, they could cause severe problems in the PV array, such as dc arcs and even fire hazards [3]. In addition, series or parallel dc arcs may occur along these four fault categories [10]. Especially for the series arcs, since they behave similar to inserted variable resistance, they may be difficult to identify or extinguish [11].
1. Detect ground faults in PV arrays mounted on the roofs of dwellings
2. Interrupt the fault current
3. Indicate that a ground fault had occurred
4. Disconnect the faulted part of the PV array
5. "Crowbar" (short-circuit) the PV array

The original GFPD prototype was developed in two versions that were similar except for voltage rating. The basic concept was to insert a 0.5- or 1-amp circuit breaker in the DC system-bonding conductor. This device connected the grounded circuit conductor (usually the negative) to the grounding system (the point where equipment-grounding conductors and the grounding-electrode conductor are connected together). Any ground-fault currents must flow through this bond on their way from the ground-fault point back to the driving source—the PV module or array (see illustration). When the current in this bond exceeds the rated 0.5 or 1 amp, the circuit breaker trips to the open position. This action interrupts the fault current, even when the fault is in 1984, engineers at the National Fire Protection Association (NFPA) directed the PV industry to propose Section 690.5 for the 1987 National Electrical Code (NEC), which would require a ground-fault protection device (GFPD) on PV systems installed on dwellings. This requirement resulted from a presentation by engineers from a national laboratory that showed a PV module that had been subject to a ground fault and had subsequently caught fire and melted down. The engineers failed to mention that this was a prototype, unlisted PV module, the module was on a concrete pad, and that ground faults in PV systems were somewhat rare. So it’s not surprising that these firefighters concluded that PV ground faults lead to fires, and directed the engineers to submit the proposal. However, when the requirement was established, no GFPDs existed for PV systems.

In 1989, I joined the PV industry as a full-time employee at the Southwest Technology Development Institute. One of my first projects was to develop prototype hardware that could be used to meet the new Section 690.5 requirement. In the 1987 Code, the requirements for this fire-reduction device were to:

2.2 Protection Gap of Conventional Solutions

Conventional fault detection and protection uses ground-fault detection interrupters (GFDI) and overcurrent protection devices (OCPD), as required by National Electrical Code (NEC) [7]. Their basic functions and limitations are introduced as follows.

2.2.1 Ground Fault Detection Interrupters

Ground fault detection interrupters (GFDI) (sometimes also called ground fault protection devices, GFPD) are a conventional solution to detect and interrupt ground faults in PV arrays in grounded PV systems. As shown in Fig. 2.2, the PV system is system-grounded as the negative conductors are intentionally grounded...
Table 2.1: Maximum allowable ground current detection settings [15].

<table>
<thead>
<tr>
<th>Device dc rating (kW)</th>
<th>Maximum ground-fault current detecting settings (Amperes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 – 25</td>
<td>1</td>
</tr>
<tr>
<td>25 – 50</td>
<td>2</td>
</tr>
<tr>
<td>50 – 100</td>
<td>3</td>
</tr>
<tr>
<td>100 – 250</td>
<td>4</td>
</tr>
<tr>
<td>&gt;250</td>
<td>5</td>
</tr>
</tbody>
</table>

[13]. Another grounding in Fig. 2.2 is called equipment grounding: non-current carrying conductive parts, such as metallic module frames, equipment, and conductor enclosures, should be grounded [7, 14].

A small fuse (e.g., 1A) is usually integrated inside PV inverters to detect ground faults by sensing the ground-fault leakage current. The UL 1741 standard requires that the settings of GFDI as given in Table 2.1 [15].

In Fig. 2.2, no matter whether there is a positive ground fault (in red) or a negative ground fault (in blue), the ground-fault current in the closed fault path will always return through the GFDI. When the fault current is large enough, the GFDI (e.g., fuse) will be blown. Then the PV inverter will shut down and the PV array will be de-energized into open-circuit conditions.

However, “blind spots” of GFDI have been discovered in [8, 16] in a double-ground fault case, which is believed to be the cause of the Bakersfield PV fire hazards that have been reported in [3]. In the double-ground fault in the Bakersfield fire, two ground faults occurred one after another. The first ground happened between the negative conductor and the ground, which is illustrated in Fig. 2.3. The fault
STAGE 1: Undetected first fault

Figure 2.3: The first undetected ground fault between a grounded current-carrying conductor (i.e., the negative conductor) and the equipment-grounding conductor [3].

current remained smaller than the GFDI setting so that the fault was hidden in the PV system.

In Fig. 2.4, the second ground fault occurred between the ungrounded current-carrying conductor (i.e., the positive conductor) and the equipment-grounding conductor. This resulted in a high-magnitude ground-fault current so that the GFDI fuse was quickly blown in the inverter. Unfortunately, the first and the second ground fault points contributed to a closed fault path so that the high-magnitude fault current flowed continuously without interruption. Since the fault current was much higher than the current rating of the conductors, the fault current generate excessive heat in the conductor, leading to conductor insulation
as shown in Figure 2, then the GFP fuse instantly opens, removing the main ground connection on the array. However, this is exactly what should not happen if there is already a fault in the array. Now, instead of having a large equipment-grounding conductor to carry the fault current, a 10 or 12 AWG source-circuit conductor has to carry the entire return current. This means that the short-circuit current now has to return through the point where the first undetected fault occurred, as shown in Figure 2. Within several seconds that conductor can get so hot that the insulation on the conductor melts or catches fire. The fire is then driven into whatever flammable material is in proximity to the wire, such as the flammable PV module backsheet or any flammable roof materials.

As the Bakersfield Fire persisted, it appears that multiple faults occurred in the source-circuit wiring, which caused several string fuses to see excessive reverse currents. These fuses would have been clearing as the sun was beginning to go down, which could have reduced the fault-current flow below the threshold at which it was capable of sustaining a fire. However, without corrective action, new fires could have started the following day as sunlight levels exceeded the condition when the initial fire occurred.

One might ask, how did such a dangerous situation go unnoticed? Until the Bakersfield Fire, very few people believed that this set of events could happen. In fact, many still believe that this is a "two-fault" event that engineers generally are not required to design for. I disagree. The idea that this scenario could repeat itself many times is not only plausible, but it is very likely. A dangerous blind spot exists—making up one quarter to one half of all large system wiring—insofar as it is possible for a first fault to be present in a grounded source-circuit conductor yet invisible to the GFP device. The first fault could exist at installation if proper commissioning tests are not done. Since ground faults are the most common faults in an array—and 50% of these faults may go undetected in the source-circuit conductors on larger PV systems—we have a real problem. The Bakersfield Fire is very easy to repeat experimentally. Unfortunately, the

The dangerous fault scenario of double-ground faults has also been found as the reasons of other fire cases, such as the PV solar fire in Mount Holly, North Carolina in 2011 [6]. In this case, the first ground fault remained undetected since the current level was much lower than the rating of the GFDI. The second ground fault occurred afterwards, and tripped the GFDI. However, the ground-fault current could flow continuously in the closed fault path, and therefore, the ground faults could not be cleared. Eventually, the fire occurred.
2.2.2 Overcurrent Protection Devices

The U.S. National Electrical Code (NEC) requires that overcurrent protection devices (OCPD) (e.g., fuse) shall be in series with each PV string to protect modules and wirings from overcurrent caused by the fault [7]. The rating current of OCPD \(I_n\) shall be no less than 156\% of PV modules rated short-circuit current \(I_{SC}\) at standard test condition (STC, 1000 W/m\(^2\), 25\(^\circ\)C, 1.5 air mass) [7]. Besides, \(I_n\) should comply with both the UL standard 2579 [17] and European IEC standard 60269-6 [18]. The non-fusing current \(I_{nf}\) of PV fuse is required as \(1.13 \times I_n\) (at least 1.76\(I_{SC}\)), according to IEC standard 60269-6 [18]. The detailed comparison between NEC, IEC and UL standards about PV fuse selection has been discussed in [19].

In addition, the fuse has a specific non-linear melting time (the required time to melt the fuse elements) vs. current characteristics (see Fig. 2.5). Generally, a larger current above \(I_{nf}\) usually requires a shorter melting time. For instance, it may need a few hours to melt a fuse if the current is only slightly above \(I_{nf}\). If the PV fault current stays below \(I_{nf}\), the fuse will not be blown and therefore, the fuse protection gap exists.

For solar PV arrays, since they are the only source of fault current, the magnitude of the backfed current into the faulted string \(I_{back}\) may be greatly reduced by low solar irradiance, maximum power point tracker (MPPT), high fault resistance, or
small-mismatched fault location [1]. Therefore the fuse protection gap may exist if the fault current stays below $I_{nf}$.

### 2.2.3 Summary

As previously discussed, the protection gap exists in both GFDI and OCPD, when certain fault scenarios occur inside the PV array. As a result, the fault may not be cleared successfully, and it can remain undetected in the PV system until the whole system fails (e.g., fire hazards).

To fill the protection gap of conventional fault detection and protection devices, various solutions have been proposed using different techniques. These solutions may depend on PV system grounding and the transformer isolation in PV inverters. As shown in Fig. 2.6, various existing fault detection, classification, protection and location have been proposed for solar PV arrays to increase the PV system
efficiency, safety and reliability. The following part of this chapter summarizes the existing solutions in each category, and briefly discusses their advantages and disadvantages.

2.3 Existing Fault Detection, Classification and Location Solutions

*Fault detection* is defined as “the indication that something is going wrong in the monitored system” [21]. In addition to fault detection, *fault classification* can automatically identify the type of fault. To further help maintenance people to look for the fault, *fault location* can estimate the cable fault location so as to expedite the system’s recovery from fault. Therefore, fault detection, classification and location are essential to monitor and identify unexpected issues in solar PV systems. Existing solutions are briefly discussed below and summarized in Table 2.2.
Table 2.2: Summary of existing PV fault detection, classification and location

<table>
<thead>
<tr>
<th>Fault Type</th>
<th>Method</th>
<th>Fault Detection?</th>
<th>Fault Classification?</th>
<th>Fault Location?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground Faults (or line-line fault with ground point)</td>
<td>GFDI</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Insulation resistance monitor</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>RCD</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>TDR</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>SSTDR</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>I-V curve analysis</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Statistical method</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Line-line faults</td>
<td>OCPD</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>RCD</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>TDR</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>I-V curve analysis</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Statistical method</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Performance comparison</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Capture loss analysis</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Performance ratio</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Machine learning</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>Open-circuit faults</td>
<td>I-V curve analysis</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Performance comparison</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Capture loss analysis</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Performance ratio</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Machine learning</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>Mismatch faults</td>
<td>I-V curve analysis</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Performance comparison</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>IR thermography</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Internal series resistance</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Machine learning</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>De-arc faults</td>
<td>AFCI</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td></td>
<td>Machine learning</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>
2.3.1 Quantitative-Model Based Solutions

2.3.1.1 Current-Voltage (I-V) Analysis

Current vs. voltage (I-V) curve analysis can provide most of the operating points of a PV module, string or array. As illustrated in Fig. 2.7, the I-V curve reveals salient PV characteristics. Therefore, it is possible to detect and classify PV faults, such as series losses, shunt losses, mismatch losses, reduced current and reduced voltage based on I-V characteristics [22–25].

2.3.1.2 Performance Comparison

In addition to fire hazards and safety issues, faults in PV systems may cause a large amount of energy loss. The energy lost due to PV faults has been analyzed and categorized in UK domestic PV systems [26] and it was estimated that the
annual energy loss was up to 18.9% in PV systems due to faults. Therefore, it is necessary to monitor PV system performance, study the fault pattern and develop the fault detection methods.

*Performance comparison* compares the actual PV performance with the simulated performance under real-time operation [23, 27]. Recently, performance comparison has been proposed for fault detection. Generally, it compares the actual performance with the expected performance. The fault detection rule is straightforward: significant difference in produced and measured output performance may indicate a fault. As shown in Fig. 2.8, the performance evaluation usually has several components, including weather information such as solar irradiance and temperature, expected PV performance as a benchmark, actual measured PV performance, performance comparison and fault detection.
Figure 2.9: System overview of PV fault detection using performance comparison [27].

For example, to prevent energy and subsequent financial losses in PV systems, an automatic PV performance comparison has been proposed in [27], which monitors the difference between the simulated and actual energy yield in real time (see Fig. 2.9). The fault detection system gathers satellite-derived solar irradiance and ambient temperature, which are fed into the simulation model to predict the PV’s AC output power ($P_{\text{sim}}$). Meanwhile, it monitors the actual AC output power ($P_{\text{actual}}$) and compares it with $P_{\text{sim}}$. Four general fault categories are: Constant energy loss, changing energy loss, snow cover and total blackout. The limitation exists: since it monitors the PV power over a period of time (equivalent to energy yield), it has a slow response. For example, this fault detection method may take at least one day.
2.3.1.3 Performance Ratio (PR)

Performance ratio (PR) is proposed in [28–30] as a normalized parameter of the PV system energy yield to evaluate the system performance. Independent of the orientation and inclination of the panel, PR considers the overall effects of system losses so that it can be used for fault detection. It is usually defined as (2.1) using the final yield ($Y_f$) over the reference yield ($Y_r$).

$$PR = \frac{Y_f}{Y_r} \quad \text{(dimensionless)}$$
Fig. 2.10 illustrates how to obtain $Y_f$ and $Y_r$ in a grid-connected PV systems. Specifically, $Y_f$ represents the normalized AC energy output to the utility grid. It is defined in (2.2), where $E$ is the net AC power output and $P_0$ is the nominal PV power.

$$Y_f = \frac{E}{P_0} \text{ (kWh/kW) or (hours)} \quad (2.2)$$

On the other side, $Y_r$ represents the normalized solar irradiation conditions. It is defined in (2.3), where $H$ is the total in-plane solar irradiation ($kWh/m^2$) and $G_{STC}$ is 1 kW/m$^2$.

$$Y_r = \frac{H}{G_{STC}} \text{ (hours)} \quad (2.3)$$

Figure 2.11: Daily, weekly, and monthly PR values for a PV system for 2001 [31].
PR values for PV systems are commonly reported on a monthly or yearly (long-term) basis. For short-term basis, such as daily or weekly, PR gives better resolution and it can be used for fault detection in PV systems. For instance, as shown in Fig. 2.11, most of the PR values lie between 0.65 and 0.8 when the PV system is normally operating. However, the significantly reduced PR represents faults/anomalies in the PV system, such as partial shadings, inappropriate system sizing, MPPT errors, inverter failures, and faults within PV arrays. In addition, Fig. 2.11 shows that PR values for smaller intervals (such as daily) give better fault-detection resolution and quicker response than monthly or weekly data [30].

2.3.1.4 Capture Loss Analysis

To overcome the limitation of previously discussed PR (performance ratio), the system capture loss ($L_c$) has been proposed in [28, 32] to further understand the abnormality on the system level. $L_c$ represents the losses due to PV array operation. It is defined in (2.4), where $Y_r$ is the reference yield and $Y_A$ is the daily array energy output per kW of the installed PV array. In addition, $L_c$ can be divided into two types of losses, namely the thermal capture loss ($L_{ct}$) and miscellaneous capture loss ($L_{cm}$) in (2.5).

$$L_c = Y_r - Y_A$$ (2.4)
\[ L_c = L_{ct} - L_{cm} \]  

(2.5)

The measured capture loss \((L_{c,mes})\) should remain in theoretical boundary if the PV array is normal. Therefore, (2.6) can be used for fault detection, where \(\delta\) is the standard deviation of the simulated capture losses \(L_{c,sim}\).

\[ L_{c,sim} - 2\delta < L_{c,mes} < L_{c,sim} + 2\delta \]  

(2.6)

As an extended research to [32], fault classification algorithm based on power loss analysis was proposed [33], which is based on the deviation of the measured parameters from the simulated ones.

2.3.2 Process-History Based Solutions

2.3.2.1 Statistical Methods

Statistical methods are proposed to detect abnormality in PV systems based on energy generation [34]. Specifically, descriptive and inferential statistics are applied on the measured energy generation of each subarray of a PV plant. The experimental results show that the proposed method can successfully detect a wiring mistake at a single PV panel out of 22 normal panels.
Multivariate outlier rules using \textit{minimum covariance determinant} (MCD) has been proposed for PV fault detection \cite{35}. Specifically, based on a number of voltage and current measurements of PV modules at each specific time, the MCD is used to calculated the robust distance (RD). The fault detection rule becomes straightforward: if the calculated RD is larger than a threshold, the fault occurs in the PV module.

\subsection*{2.3.2.2 Machine Learning}

\textit{Machine learning} is a subarea of artificial intelligence, which automatically extracts knowledge from the given PV data set. One category of the machine learning uses supervised learning approaches. As shown in Fig. 2.12, depending on a large amount of labeled data, supervised learning algorithms can learn the system and make the prediction after it is trained. A variety of supervised learning models have been proposed in PV installations. Artificial Neural Networks (ANN) is
developed for PV performance evaluation under partial shadings [36], PV health status monitoring [37] and short-circuit fault detection in PV arrays [38]. Bayesian Neural Network (BNN) and regression polynomial models have been proposed to predict the soiling effects on large-scale PV arrays [39]. PV fault detection and classification use decision-tree model in [40], K-nearest neighbor and support vector machine (SVM) in [41].

2.3.3 Signal-Processing Based Solutions

2.3.3.1 Dc Arc-Fault Circuit Interrupter (AFCI)

_Dc arc-fault circuit interrupter (AFCI)_ is required by _NEC 2014 Edition_ [7] to fill the protection shortcomings of conventional GFDI and OCPD. If the fault is not properly cleared by GFDI and OCPD, dc arc may occur in PV arrays [42]. An electric arc is defined as “an electrical breakdown of a gas which produces an ongoing plasma discharge, resulting from a current flowing through normally nonconductive media such as air” [43]. In Fig. 2.13(a), a dc arc is created using dc arc generator and a PV simulator in the Power Electronics lab at Northeastern University. The schematic of the dc arc experimental setup is given in Fig. 2.13(b). In this low-power case, the arc voltage $V_{arc} = 30Vdc$ and arc current $I_{arc} = 3A$. At this moment of the arc fault, the PV simulator is still working around the maximum power point (MPP) so that the dc arc is continuous without interruption. In real
PV fields, as the current and voltage rating increases, dc arcs may have a much higher power level that can easily ignite fire hazards.

(a) Dc arc is created by a dc arc generator and a dc power supply.

(b) The schematic diagram of a dc arc experimental setup.

Figure 2.13: The dc arc experimental setup at Northeastern University.

As shown in Fig. 2.14, the dc arc will generate a significant amount of ac noise that can be used as a fault detection signature. AFCI should detect and interrupt series and parallel arcing faults in dc PV sources and output circuits. In addition, several dc-arc detection methods have been developed using Fast Fourier Transform (FFT) [44], Artificial Neural Network (ANN) [44] and Discrete Wavelet Transform (DWT) [45].
2.3.3.2 Insulation Resistance Monitor

*Insulation resistance monitor* has been used as a ground-fault protection in ungrounded PV systems, when the system is de-energized [46–48]. Insulation resistance of the PV array ranges from $k\Omega$ to $M\Omega$, and it decreases dramatically when insulation faults or direct contact faults (e.g. hazardous ground fault) occur. Therefore, a significantly reduced insulation resistance may indicate a ground fault.

Solar PV insulation model has been proposed in [47] that includes PV modules, series and parallel insulation resistance, leakage capacitance of PV modules, and PV working conditions (e.g., short-circuit or open-circuit). In addition, the experimental results show that insulation resistance of a PV array decreases as PV module temperature rises or relative humidity (RH) increases. Fig. 2.15 shows a commercial product of insulation resistance monitor. It monitors both the positive-ground
2.3.3.3 Residual Current Detector (RCD)

*Residual current detector (RCD)* acts like a differential relay in power system, by monitoring the current difference between input and output terminals of the protected zone (see Fig. 2.16). If the current difference is higher than a threshold, the fault alarm will be sent out. However, parasitic capacitance of the PV array may cause considerable leakage current when transformerless PV inverters are used. This may lead to nuisance tripping on RCD [50].

The differential protection is known as the best protection technique now and for more than 50 years for transformers, motors, generators, and reactors in power systems [51]. The idea of differential relays is to monitor the electrical quantities...
Figure 2.16: Concept of differential relays for internal fault detection [51]

(usually current) entering and leaving the protected zone. As shown in Fig. 2.16, the primary current coming in and leaving from the protected zone are $I_{1a}$ and $I_{1b}$, respectively. The corresponding secondary current are $i_{1a}$ and $i_{1b}$. Their difference is the relay’s operating current $i_{op} = i_{1a} - i_{1b}$. If the magnitude $|i_{op}|$ is higher than zero or a small threshold, a fault alert will be sent out.

Similar to the differential relay in power systems, residual current detector (RCD, or residual current monitoring unit, RCMU) has been proposed for fault detection in PV arrays (see Fig. 2.17). RCD monitors the sum of currents coming in and going out from the protected zone, which can be an entire PV array as RCD#1 or an individual PV string as RCD#2. Once $|i_{op}|$ is larger than their separate threshold, a fault will be detected. For example, a ground fault occurs in String 1 at point $G1$, who has a ground fault current $I_g$. Therefore, $i_{1a} \neq i_{1b}$ and $i_{pos} \neq i_{neg}$. And the fault will be detected by both RCD#1 and RCD#2 properly.

However, the protection “blind spot” of RCD exists when no external fault point involves with the protected zone. For example, short-circuit faults, open-circuit
faults, or degradations inside the protected zones. In these cases, $i_{pos}$ should be identical to $i_{neg}$, and $i_{op}$ should be around zero.

### 2.3.3.4 Time Domain Reflectometry (TDR)

*Time domain reflectometry (TDR)* is a measurement approach to determine the electrical wire or cable characteristics by injecting certain waveforms (usually step or impulse signals) and observing reflected waveforms. TDR compares the reflections from the unknown line environment to those generated by standard or given
impedance. Therefore, TDR can be used to troubleshoot the faulty wires or cables. In addition to fault detection and classification, TDR is able to locate the fault, which is especially useful in a large electrical system.

Fault location using TDR has been proposed in PV fields [52–54]. The schematic diagram is shown in Fig. 2.18. For example, Fig. 2.19 shows a step input into a PV string and its reflected waveforms. The waveforms differs with the type of conditions, such as open-circuit fault, short-circuit fault, and resistance load.
Fig. 2.20 shows the TDR results on open-circuit and short-circuit faults in PV strings. Note that a negative input step is used this case. The different TDR response may be helpful to further identify the fault location.

However, TDR has several drawbacks. First, the PV system under test must be off-line, which will affect the energy yield. Second, it requires human input (labor cost) to observe and learn the reflected waveforms, which hinders its effectiveness of automatic fault detection.

2.3.3.5 Spread Spectrum Time Domain Reflectometry (SSTDR)

As an improved version of TDR, spread spectrum time domain reflectometry (SSTDR) has been commercially proposed to detect aircraft wiring faults [55]. SSTDR has low test signal levels and high noise immunity, which make it a viable solution to detect live wires fault (in energized systems). For PV ground-fault detection, the autocorrelation peaks generated by SSTDR under ground-fault condition is
significantly higher than the normal one. According to this, a fault detection algorithm has been proposed in [56]. However, SSTDR has not been proposed for fault location or classification.

2.3.3.6 Infrared (IR) Thermography

Infrared (IR) thermography has been used to identify certain mismatch faults, such as hot spots on PV modules, which may cause irreversible damage and power loss on PV modules [57, 58]. Since IR inspection is periodic and repetitive, it may be costly if it requires PV maintenance personnel increased labor costs over the lifespan PV modules.
2.3.3.7 Internal Series Resistance

The schematic diagram of the widely used one-diode model is given in Fig. 2.22. The current equation of the equivalent circuit can be written in (2.7) [14].

$$I_{pv} = I_L - I_S \left[ \exp \left( \frac{V_{pv} + I_{pv}R_s}{A \cdot k \cdot T \cdot N_S \cdot q} \right) - 1 \right] - \frac{V_{pv} + I_{pv}R_s}{R_{sh}}$$  \hspace{1cm} (2.7)

where $I_{pv}$ is the output current of PV module, $I_L$ is the light-generate current, $I_S$ is the saturation current of the diode, $V_{pv}$ is the module voltage, $R_s$ is the equivalent internal series resistance of module, $R_{sh}$ is the equivalent shunt resistance of module, $A$ is the diode ideal factor, $q$ is the electron charge $(1.6 \times 10^{-19} C)$, $k$ is the Boltzmann constant $(1.38 \times 10^{-23} J/K)$, $T$ is the PV module temperature $(K)$, and $N_S$ is the number of series solar cells in the module.

The internal series resistance ($R_s$) of PV modules can be used for degradation detection inside PV modules. The actual $R_s$ can be estimated by using either I-V
curve analysis [60] or the measurement around open-circuit voltage [61]. Compared with the normal $R_s$, degraded PV modules usually exhibit increased $R_s$, which provide a useful feature for fault detection.

### 2.4 Existing Fault Protection Solutions

Once the fault is detected, a fault signal can activate fault protection devices to interrupt the fault, in order to prevent PV components from damage. “The fundamental objective of system protection is to provide isolation of a problem area in the power system quickly, so that the shock to the rest of the system is minimized and as much as possible is left intact” [51]. Fault protection can be also called as fault clearance, or fault interruption.

In PV applications, the fault area in the solar PV arrays should be isolated so that the impact to the rest of the PV system is minimized. The existing fault protection methods for PV arrays are summarized in Table. 2.3. Passive methods use ground-fault detection interrupters (GFDI), overcurrent protection devices (OCPD) or blocking diodes [1, 7]. On the other hand, active methods use more complex sensing circuitry to detect the fault, and rely on circuit breakers, contactors, or semiconductors switches to de-energize and isolate the affected PV components [62–64].
Passive protection devices have obvious limitations. The “blind spots” of GFDI and OCPD have been discussed previously. Furthermore, blocking diodes are not a substitute of OCPD and they can prevent OCPD from normal operation [1]. To improve fault protection, active fault protection devices have been developed and shown the advantages over passive ones. But they greatly dependent on the fault detection methods (i.e., decision-making algorithms). Therefore, there is still a great need of fault detection methods that can provide responsive and reliable tripping signal to active protection devices.

<table>
<thead>
<tr>
<th>Protection Type</th>
<th>Method</th>
<th>Cost</th>
<th>Complexity</th>
<th>Limitation/Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive</td>
<td>Passive protection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GFDI (Required by NEC)</td>
<td>Low</td>
<td>Low</td>
<td>Blind spots</td>
</tr>
<tr>
<td></td>
<td>OCPD (Required by NEC)</td>
<td>Low</td>
<td>Low</td>
<td>Blind spots</td>
</tr>
<tr>
<td></td>
<td>Blocking diode (Not required by NEC)</td>
<td>Low</td>
<td>Low</td>
<td>Low reliability</td>
</tr>
<tr>
<td>Active</td>
<td>Dc arc-fault circuit interrupter (AFCI)</td>
<td>Medium</td>
<td>Medium</td>
<td>High-frequency interference from power condition units</td>
</tr>
<tr>
<td></td>
<td>Contactor + protection relay</td>
<td>Medium – High</td>
<td>Medium</td>
<td>Bulky and costly for high-voltage application</td>
</tr>
<tr>
<td></td>
<td>Semiconductor switch + protection relay</td>
<td>Low – Medium</td>
<td>Medium</td>
<td>PV-module level only</td>
</tr>
</tbody>
</table>
2.5 Conclusion

In this chapter, we present the literature review of existing fault detection, classification, location and protection solutions for solar photovoltaic (PV) arrays (dc side). Also, the benefits and drawbacks of existing approaches are explored and compared. Generally, fault detection, classification and location methods can be divided into three categories, which are quantitative model-based methods, process-history based methods, and signal-processing based methods. Required by the National Electrical Code (NEC), ground-fault detection interrupters (GFDI) and overcurrent protection devices (OCPD) are widely used for fault protection in PV installations. However, their weakness and limitations have been discovered, which may lead to the previously reported fire hazards.

To address this issue, various methods have been proposed using different techniques. However, they have limitation that may hinder their effectiveness in real-world PV application. Therefore, there is an urgent need of better fault detection methods to prevent PV systems from fault hazards.
Chapter 3

Fault Studies in a Commercial-Scale PV Laboratory

3.1 Introduction

Fault analysis is a fundamental task to prevent the aforementioned faults and associated fire hazards from PV installations. Since PV arrays are unique power sources that have non-linear output characteristics and current-limiting features, special attention is needed to study and understand the fault behavior of PV systems, including PV arrays, PV inverters, and protection devices (e.g., OCPD and GFDI).

Fault analysis in PV systems have been studied in the literature [1, 3, 8, 23, 26, 65–69]. However, these fault analyses have obvious limitations: 1) Some of the studies
are only focusing on mismatch faults, which do not include the dangers associated with large fault current or possible fire hazards [23, 69]; 2) Some PV fire hazards are reported on a case-by-case basis. They often lack the verification of real PV data or any detailed transient fault behavior [3, 26]; and 3) Experimental results under real-working conditions are limited by the size of the PV systems (installed power capacity < 200 W) [68]. No ground faults or line-line faults have ever been implemented on PV systems at commercial-scale level, which are typically above 20 kW installed power with 3-phase ac output connection. Therefore, the fault current in reported experiments is usually only a few amps, which is much less than real-world fault current (ranging from a few amps to hundreds of amps) [1].

In order to solve these limitations of the prior art, this chapter focuses on the fault studies of solar PV arrays and discovers the shortcomings of OCPD in a typical 35kW solar PV systems in both simulations and experiments. Specifically, this chapter of the dissertation includes the following two research contributions:

**Contribution 1**: The line-line faults are simulated and analyzed in a typical commercial-scale PV system of 35 kW.

- The success and failure of fault clearance by OCPD (i.e., fuses) are studied. The $I-V$ curve analysis shows that “blind spots” in OCPD exist if the current magnitude is too low. It is shown that fault current magnitude varies greatly with the changing fault locations.
Contribution 2: Fault experiments are implemented in a typical commercial-scale PV system of 35 kW that is able to create and record faults.

- Together with international partner Mersen Corp., we have designed and built an experimental PV testing facility, located in Saint-Bonnet-de-Mure, France. This is the first time line-line fault measurement in a commercial-scale PV system have been presented. PV strings are specially designed to create line-line faults. Oscilloscope and PV monitoring system are used to capture the transient and post-fault scenarios in the PV arrays, which provide well sampled data for fault analysis. The fault analysis covers both the transient response and post-fault behavior of the PV system (dc side), including PV array, overcurrent protection devices (OCPD, i.e., fuses) and grid-connected PV inverter.

3.2 Simulation of Line-Line Faults in PV systems

3.2.1 Typical Grid-Connected PV Systems

A typical grid-connected PV system includes a PV array of $M \times N$ modules ($M$ number of series modules in every string and $N$ number of parallel PV strings), a central PV inverter with MPPT, overcurrent protection devices (OCPD) and a
ground fault detection interrupter (GFDI). As shown in Fig. 3.1, line-line faults could happen inside PV arrays and potentially may involve large fault current or dc arcs. This research focuses on line-line faults, which are defined as an accidental short-circuiting between two points in the array with different potentials. Illustrated in Fig. 3.1, a line-line fault may be caused by a short-circuit fault between two different points or a double-ground fault in PV arrays.
3.2.2 Simulation Parameters

As the schematic diagram shown in Fig. 3.2, a grid-connected PV system with a centralized inverter is developed in MATLAB/Simulink. Using the widely used one-diode PV model, a solar PV array with $12 \times 12$ PV modules (rated at $34.56\, kW$) is simulated, which consists of 12 modules in series per string and 12 strings in parallel ($M = 12$ and $N = 12$). In the simulation, a 2-stage PV inverter has a front-end dc-dc converter cascaded with a 3-phase inverter. The dc-dc converter uses a boost topology to step up the PV voltage to the dc bus voltage. Meanwhile, the dc-dc converter is integrated with the MPPT algorithm, which extracts the maximum possible power from the PV array. The dc-ac 3-phase inverter is playing the role of converting the dc power in a proper form into the ac utility grid. At the same time, the inverter maintains the dc bus voltage at a predefined value.

The main parameters of the PV modules under Standard Test Conditions [(STC), $1000\, W/m^2$, $25\, ^\circ C$, $1.5$ air mass] are as follows: the maximum power $P_{MP} = 242\, W$, the open-circuit voltage $V_{OC} = 37.2\, V$, the maximum power voltage

![Schematic diagram of the simulated PV system.](image)
$V_{MP} = 30\text{ V}$, the short-circuit current $I_{SC} = 8.6\text{ A}$, the maximum power current $I_{MP} = 8.1\text{ A}$, the number of series solar cells per module $N_s = 60$, and three bypass diodes per module. The MPPT [e.g., perturb and observe (P&O)] voltage of the PV inverter ranges from $V_{MPPT_{min}} = 230\text{ V}$ to $V_{MPPT_{max}} = 500\text{ V}$. At each voltage perturbation, the MPPT voltage step is $\hat{v}_{MPPT} = 2\text{ V}$.

The simulated PV system is capable of studying faults among them. In this dissertation, a line-line fault with $x$ number module difference is annotated as “LL-$x$”, where $x$ can be any integer between $1$ and $M$, where $M$ is the number of PV modules per string. Assume that the whole PV array is operating under STC. Note that we only consider solid faults, which means zero fault impedance.

In the following simulations, two representative cases, such as LL-2 and LL-5 are studied. Namely, a line-line fault with 2 modules and a line-line fault with 5 modules. In the following part of this chapter, fault scenarios in solar PV array are analyzed and explained through fault evolution in time domain and $I$-$V$ curve analysis.

### 3.2.3 Fault Analysis using I-V Curves and MPPT Effects

First, let us assume that the string current $i_j$ ($j = 1 \ldots 12$) consists of a dc component ($I_j$) and a ac component ($\hat{i}_j$), which are defined as

$$i_j = I_j + \hat{i}_j \quad \text{for} \quad j = 1 \ldots 12 \quad (3.1)$$
Specifically, $I_j$ is the steady-state dc current, which is determined by the specific $I$-$V$ curve and the array’s operating voltage $v_{pv}$. On the other hand, $\hat{i}_j$ is mainly caused by the array voltage perturbation $\hat{v}_{MPP}$ when the MPPT is working.

The $I$-$V$ curves of the PV array and String #1 are plotted in Fig. 3.3 for fault analysis. Here we only consider the steady state of PV string current $I_j$, which is mainly determined by weather conditions, its $I$-$V$ curve of the specific string, and the operating voltage $v_{pv}$. For this reason, the $I$-$V$ curve analysis is fundamental to understand the fault scenarios among PV strings. When a $LL$-$x$ fault occurs, the $I$-$V$ curve of the faulted PV string will change accordingly. Since the faulted string has $x$ number of modules less, it will have an open-circuit voltage reduced...
by $x \cdot V_{oc}$. But the short-circuit current remains the same as other normals strings at $I_{SC}$.

The $I-V$ curves of the PV array and strings under the previously discussed $LL-2$ is plotted in Fig. 3.3. Before $LL-2$ occurs, the PV array is operating around the normal maximum power point (MPP), $MPP1$. At the moment of $LL-2$, the $I-V$ curve of the array is changed suddenly with a reduced open-circuit voltage and a new MPP $MPP2$. Since $v_{pv}$ follows the P&O MPPT’s command ($V_{CMD}$) and oscillates around $V_{MPP1}$, it is found that $I_1$ can lie between $0.2I_{SC}$ and $0.3I_{SC}$ at the moment of the fault. After that, once $V_{CMD}$ is updated by MPPT, $v_{PV}$ will decreases gradually until it reaches $V_{MPP2}$. As a result, $I_1$ increases to $0.51I_{SC}$. For normal String #2 to #12, as $v_{pv}$ decreases, their current $I_j$ ($j = 2$ to 12) is slightly increased toward $I_{SC}$. It is worth mentioning that this fault has no reverse current and will not be cleared by OCPD.

Similarly, $I_1$ under $LL-5$ can be predicted by Fig. 3.3. It is found that $I_1$ has a large reverse current $-4.4I_{SC}$ at $V_{MPP1}$ at the moment of the fault. This reverse current is coming from other normal strings into String #1. At this point, String #1 is dissipating $4.4 \times 8.6 A \times 361 V = \sim 13.7 kW$ power as a load. Since the reverse current ($-4.4I_{SC}$) is much larger than the PV fuse’s $I_{nf}$ (at least $1.76I_{SC}$), the fault has a better chance to be cleared by the fuse before $v_{pv}$ reaches a new global MPP at $V_{MPP3}$. 
3.2.4 PV Dynamic Characteristics

The dynamic conductance $g_{pv}$ (unit: $S = \Omega^{-1}$) of PV strings is defined as

$$g_{pv} = \left. \frac{\hat{i}}{\hat{v}_{pv}} \right|_{\text{at certain } v_{pv}}$$

representing the string current change $\hat{i}$ under a small voltage perturbation $\hat{v}_{pv}$. In other words, a PV string with a larger $g_{pv}$ will have a larger current change under the same voltage perturbation. The $v_{pv}$ vs. $g_{pv}$ curves under normal conditions and various line-line faults are plotted in Fig. 3.4 for further discussion. Note the x-axis is the array voltage $v_{pv}$ that ranges from 0 to $V_{OC}$.

For example at the knee of the I-V curves (around $V_{MPP1}$), according to (3.2), different conditions have the corresponding current perturbation under the same voltage perturbation $\hat{v}_{MPPT}$. By comparing $g_{pv}$ of various conditions (see Fig. 3.4), it is found that line-line faults involved with more modules tend to have a more negative $g_{pv}$ at $V_{MPP1}$ and therefore a higher magnitude of $\hat{i}$. This research finding predicts that as more PV modules are involved at line-line faults, the faulted string will have larger magnitude of current variation $|\hat{i}|$ under the MPPT perturbation at the post-fault steady state.

The second observation in Fig. 3.4 is that $g_{pv}$ is a function of the dc operating point $v_{pv}$. In other words, $g_{pv}$ increases as $v_{pv}$ is decreased. For example, under the same voltage perturbation $\hat{v}_{MPPT}$, $|\hat{i}|$ decreases as the dc operating point $v_{pv}$ is
Figure 3.4: Simulated $v_{pv}$ vs. $g_{pv}$ curves of PV strings under normal, and line-line faults, such as LL-1 to LL-5.

reduced. It is necessary to mention that $\hat{i}$ caused by $\hat{v}_{MPPT}$ can only be observed under P&O or other similar MPPT techniques.

### 3.2.5 Fault Evolution in Time Domain

In this subsection, the same fault scenarios are simulated and studied in the time domain, which can be predicted and explained by the previously studied $I$-$V$ curve analysis and MPPT effects. Assume LL-2 fault occurs on String #1 at time $t_1$, which will cause mismatch between String #1 and other normal strings (String #i for $i = 2 \ldots 12$). Since the faulted PV array has larger open-circuit voltage than $V_{MPPTmin}$, the inverter’s operation is sustained and the MPPT is still working.
The current evolution of $LL-2$ on String #1 is simulated in Fig. 3.5. The current axis (vertical) is normalized by $I_{SC}$ for clarity of explanation, where $I_{SC} = 8.6 \, \text{A}$ under $STC$. Current of String #1 ($i_1$) and current of other normal strings ($i_j$, for $j = 2$ to 12) are plotted in Fig. 3.5 for comparison. Before the fault (time $t < t_1$), the string current $i_1$ to $i_{12}$ have identical values and are overlapping in Fig. 3.5. When $LL-2$ occurs at time $t_1$, PV array voltage $v_{pv}$ is found as 363.7 V and $i_1$ is greatly reduced to $0.2I_{SC}$, because String #1 has a changed $I-V$ curve (according to Fig. 3.3). After time $t_2$, the P&O MPPT begins to respond by gradually decreasing $v_{pv}$ from the pre-fault MPPT command $V_{CMD}$ at $V_{MPP1}$ towards a new MPP at $V_{MPP2}$. During this process, $i_1$ is continually increasing.
as \( v_{pv} \) keeps decreasing. Finally at time \( t_3 \), \( i_1 \) reaches a post-fault steady state and begins to oscillate around \( 0.51I_{SC} \). Notice that during the fault evolution, \( i_1 \) is always positive and never has any overcurrent. This brings the protection challenges to the OCPD, since OCPD never has a chance to clear the fault. Note that as previously discussed, the non-fusing current \( (I_{nf}) \) of PV fuses is required as \( 1.13 \times I_n \) (at least \( 1.76I_{SC} \)). Therefore, to successfully clear the fault, the fault current magnitude must be larger than \( I_{nf} \).

The current ripple caused by the MPPT in the time domain can be successfully predicted by the PV dynamic conductance \( g_{pv} \). There are two aspects regarding the current ripple: 1) The faulted string #1 tends to have a larger current ripple under the same MPPT voltage perturbation. For instance in Fig. 3.5, during post-fault steady state (after time \( t_3 \)), the peak-peak current ripple \( \hat{i}_{pp} \) of String #1 is greatly increased to \( 0.059I_{SC} \), compared to the pre-fault \( \hat{i}_{pp} = 0.009I_{SC} \). 2) Since \( v_{pv} \) is reduced to \( V_{MPP2} \), normal strings are expected to have smaller current ripple. For example in Fig. 3.5, after time \( t_3 \), \( \hat{i}_{pp} \) of normal strings is further reduced to \( 0.0047I_{SC} \), compared to the pre-fault value \( 0.009I_{SC} \).

### 3.2.6 More Simulation Results and Discussion

More simulation results under \( STC \) are summarized in Table. 3.1 for comparison, which has shown that the line-line fault with more modules tend to have a more reduced \( dc \) current component \( I_j \). The conclusion is that the OCPD (e.g., fuses)
Table 3.1: Simulation results of various line-line faults at String #1 under STC

<table>
<thead>
<tr>
<th>Fault Type</th>
<th>$I_1$ during Transient</th>
<th>$I_1$ at Post-Fault Steady State</th>
<th>Fault Clearance by OCPD?</th>
</tr>
</thead>
<tbody>
<tr>
<td>LL-1</td>
<td>$0.77I_{SC}$</td>
<td>$0.81I_{SC}$</td>
<td>No</td>
</tr>
<tr>
<td>LL-2</td>
<td>$0.25I_{SC}$</td>
<td>$0.51I_{SC}$</td>
<td>No</td>
</tr>
<tr>
<td>LL-3</td>
<td>$-0.76I_{SC}$</td>
<td>$-0.076I_{SC}$</td>
<td>No</td>
</tr>
<tr>
<td>LL-4</td>
<td>$-2.28I_{SC}$</td>
<td>$-0.286I_{SC}$</td>
<td>No</td>
</tr>
<tr>
<td>LL-5</td>
<td>$-4.4I_{SC}$</td>
<td>$-0.438I_{SC}$</td>
<td>Most likely</td>
</tr>
<tr>
<td>LL-6</td>
<td>$-7.321I_{SC}$</td>
<td>$-0.4179I_{SC}$</td>
<td>Yes</td>
</tr>
<tr>
<td>LL-7</td>
<td>$-10.61I_{SC}$</td>
<td>$I_1 = 0$ A after fuse is blown.</td>
<td>Yes</td>
</tr>
<tr>
<td>LL-8</td>
<td>$-10.96I_{SC}$</td>
<td>$I_1 = 0$ A after fuse is blown.</td>
<td>Yes</td>
</tr>
<tr>
<td>LL-9</td>
<td>$-10.97I_{SC}$</td>
<td>$I_1 = 0$ A after fuse is blown.</td>
<td>Yes</td>
</tr>
<tr>
<td>LL-10</td>
<td>$-11I_{SC}$</td>
<td>$I_1 = 0$ A after fuse is blown.</td>
<td>Yes</td>
</tr>
<tr>
<td>LL-11</td>
<td>$-11I_{SC}$</td>
<td>$I_1 = 0$ A after fuse is blown.</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*a*In all cases, the pre-fault current $I_1 = 0.94I_{SC}$ under STC.

can successfully clear the line-line faults if enough PV modules (5 modules or more in this specific case) are involved. However, the protection gap exists when OCPD fails to work under line-line faults with less than 4 modules, since the faulted string does not have overcurrent to melt fuses.

Moreover, PV string current under LL-1 and LL-2 remains positive, which may show similar values to normal string current. It becomes difficult to differentiate between faulted and normal string current. Therefore, there is an urgent need of
simple and responsive fault detection methods to detect such faults in PV systems.

3.3 Experimental Setup in the PV Laboratory

To verify the research findings by the previous simulation results, for the first time, a 35kW commercial-scale PV laboratory (shown in Fig. 3.6) is designed to study a series of line-line faults under real-working conditions, including transient behavior of grid-connected PV inverters, MPPT response, and failure of OCPD protection. The experimental results verify our previous simulation results, showing that the line-line fault with involving 4 modules or less in a 12-module string may remain in the PV array without OCPD interruption and become a safety issue.

A specialized 35 kW, 500 Vdc grid-connected PV system (also called the PV laboratory) has been designed and built that is capable of creating different types of PV faults, as shown in the photograph (see Fig. 3.6) [70]. Fig. 3.7 illustrates the
schematic diagram of the PV lab. A photograph of major experimental equipments is given in Fig. 3.8. Specifically, the current probe amplifier (TCPA300), current probe (TCP303, measurement range dc–15 MHz, 150 Adc) and voltage differential probe (5200A) by Tektronix are used with the oscilloscope for transient measurement. Additionally, the PV string monitoring system by Mersen (see Fig. 3.9) is installed on every six PV strings to accurately record the string current and the PV array voltage [71] at 0.5 Hz sampling frequency. The main parameters of PV components are given in Table 3.2 in detail.

In Fig. 3.7, the first two strings are specially designed to create line-line faults.
### Table 3.2: Main Parameters of Experimental PV Laboratory

<table>
<thead>
<tr>
<th>Equipment</th>
<th>Model</th>
<th>Detailed parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV module</td>
<td>TE240-60M+</td>
<td>At standard test conditions (STC): $V_{OC}=37.2$ V, $I_{SC}=8.6$ A, $V_{MPP}=29.55$ V, $I_{MPP}=8.2$ A, $P_{MPP}=240$ W</td>
</tr>
<tr>
<td>PV array</td>
<td>144 modules, 12 strings, 12 modules per string</td>
<td>At MPP under STC: $P_{PV}=34.56$ kW, $V_{PV}=446.4$ V, $I_{PV}=103.2$ A</td>
</tr>
<tr>
<td>Grid-connected inverter</td>
<td>Fronius CL 36.0 Center inverter (with transformer isolation)</td>
<td>Max. output power 36 kW, DC startup voltage: 245 V, MPP voltage range: 230 V to 500 V</td>
</tr>
<tr>
<td>PV string monitoring system</td>
<td>Mersen HMMC6A</td>
<td>String voltage measurement range ±1000 V DC, accuracy ±0.5% (±5 V); nominal string current measurement range ±20 A, accuracy ±0.5% (±100mA). Sampling frequency in experiments: 0.5 Hz</td>
</tr>
<tr>
<td>PV fuses</td>
<td>Mersen HP10M15</td>
<td>Rated voltage 1000 Vdc, rated current 15 Adc</td>
</tr>
</tbody>
</table>

![Figure 3.8: The photo of major experimental equipments.](image-url)
The switches among these two strings are controllable so that it is easy to make a short circuit between specific points. For example, the line-line fault with one module difference is defined as “LL-1” in Fig. 3.7. A line-line fault with two modules difference is denoted as “LL-2”. Similarly, the faults “LL-3”, “LL-4”, “LL-5” and “LL-7” can also be created by closing the corresponding switches in the experimental test bed. Note that we only consider solid faults, which means zero fault impedance. Besides, this dissertation focuses on line-line faults, which may be caused either by a single short-circuit fault in grounded or ungrounded systems, or a double-ground fault in ungrounded PV systems.

### 3.4 Experimental Results

In the experiments, line-line faults involved with various number of PV modules (i.e., LL-x, x = 1, 2, 3, 4, 5 or 7) are implemented in the PV lab under real-working conditions. The fault analysis of transient and post-fault PV system behavior will be given as follows.
As shown in Fig. 3.7, the line-line fault with 2 module difference is defined as “LL-2” in this dissertation. It has a 16.67% location mismatch since it involves two-module mismatch between the fault points in the faulted string (normally, 12 modules per string). Similarly, a line-line fault with 5-module and 7-module difference are annotated as “LL-5” with 41.67% mismatch and “LL-7” with 58.3% mismatch, respectively. This section focuses on three representative faults “LL-2”, “LL-5” and “LL-7”. As predicted in the previous simulation results, the MPPT of the inverter is still able to work, as long as the faulted PV array can sustain the working condition of the PV inverter. As a result, the MPPT tends to gradually increase the faulted string current by decreasing the PV array voltage [1].

In summary, the experimental results have verified our previous research findings and simulation results.

- **LL-2** will remain undetectable by fuse since the faulted string still has a positive current, which could be difficult to differentiate from other normal-working strings. This brings the fault detection and protection challenge to the PV system.

- **LL-5** is successfully cleared by fuse in this specific example, which is triggered by a larger reverse current when the PV inverter checks the open-circuit voltage of the PV array.
• *LL-7* causes a large reverse current into the faulted string, which comes from other normal PV strings. This large current will be cleared by the fuse immediately.

In addition, more experimental results including “LL-1”, “LL-3”, and “LL-4” are given and summarized in Section 3.4.4. It is shown that the more modules involved in a line-line fault, makes it more likely that it can be cleared by fuses. Specifically in our example, *LL* faults with 4 modules or less may be unable to be cleared by fuses.

### 3.4.1 Line-Line Fault with 2-Module Difference (*LL-2*)

The transient behavior of *LL-2* fault is captured by the oscilloscope in Fig. 3.10(a). Meanwhile, the long-term fault behavior is recorded by the PV monitoring system in Fig. 3.10(b), which is especially useful for post-fault steady-state analysis. When *LL-2* occurs at time $t_1$, String #1 current ($i_1$) drops from 7.2 A to 2.5 A, and reaches 2.75 A after 80 ms. Note that the currents of Strings #7 through #12 ($i_7$ to $i_{12}$) are not plotted, since they have similar currents to normal string currents $i_2$ to $i_6$ in the experiments.

Fig. 3.10(b) shows that MPPT has an effect on the faulted PV array: the array voltage ($v_{PV}$) is reduced gradually in order to find an optimal operating point. Due to this effect, $i_1$ increases accordingly until a new MPP is reached. Compared to the pre-fault MPP for the PV array (12 parallel strings; 320 V, 81 A, and 25.9
Fault Studies in a Commercial-Scale PV Laboratory

Chapter 3.

### Transients

**PV array voltage**

- **$v_{PV}$**
- **$i_1$**

- Fault occurs at time $t_1$

### (a) Transient behavior captured by the oscilloscope (Upper trace $v_{PV}$, 50 V/div; lower trace $i_1$, 2.5 A/div; time/division is 10 ms/div).

### (b) Post-fault steady state captured by the PV monitoring board.

**Figure 3.10:** Experimental results of \(LL-2\) (2-module LL fault).
kW), the post-fault MPP has a 7.34% power loss at the operating point 312 V, 77 A, and 24 kW.

The protection challenge for PV fuses is obvious: the faulted string current (varying between 2.75 A and 4 A in Fig. 3.10(b)) is too low and will never blow a fuse (rating current 15 A). As a result, the fault remains in the PV array. When the fault is manually cleared at time $t_2$, the MPPT increases $v_{PV}$ until another optimal operating point is found. The detailed $I$-$V$ curve analysis of similar simulated fault scenarios has been given previously in Section 3.2 in the dissertation.

### 3.4.2 Line-Line Fault with 5-Module Difference ($LL-5$)

Fig. 3.11 illustrates the transient of $LL-5$ fault and its subsequent fault clearance by fuses. The PV array has a pre-fault MPP at 327 V, 62.4 A, and 20.4 kW (under lower solar irradiance than $LL-2$). At time $t_1$, $i_1$ has an immediate reverse current (-23 A) at String #1. Notice $V_{PV}$ drops because the PV array has a changed $I$-$V$ curve, which has been previously explained. But the MPP voltage command ($V_{CMD}$) given by the MPPT remains the same as the one before the fault. Therefore, $v_{PV}$ keeps increasing to $V_{CMD}$ after the fault, until $V_{CMD}$ is updated at $t_2$. This leads to a maximum reverse current (-25 A) at String #1 in Fig. 3.11(b). At this moment, the PV power loss caused by the fault is about 40% of the pre-fault MPP, since the faulted array is operating at 327 V, 37.4 A, and 12.23 kW. Meanwhile, the power dissipation on String #1 is approximately 8.17
Chapter 3. *Fault Studies in a Commercial-Scale PV Laboratory*  

(a) Transient behavior captured by the oscilloscope (Upper trace $v_{PV}$, 50 V/div; lower trace $i_1$, 5 A/div; time/division is 10 ms/div).

(b) Post-fault steady state captured by the PV monitoring board.

**Figure 3.11:** Experimental results of *LL-5* (5-module LL fault).
64 kW ($= -25 \times 327 V$), which is likely to cause module degradations or even fire hazards if the fault is not cleared promptly.

After $t_2$, the MPPT begins to continuously look for an optimal working point by reducing $v_{PV}$. Similar to the $LL-2$, $i_1$ is increasing (the magnitude is reducing) gradually by the MPPT. In our specific example, $|i_1|$ decreases to 18.4 A at time $t_3$. At time $t_4$, this particular PV inverter checks the open-circuit (called “$V_{OC}$ check”) under poor-performance conditions of the PV array, resulting in a negative peak of the reverse current, as well as a current drop on other normal strings. The fuse is triggered by this current peak and it is blown at $t_5$, showing that the fault is cleared successfully by the fuse.

3.4.3 Line-Line Fault with 7-Module Difference ($LL-7$)

LL faults with 7 modules ($LL-7$) has been implemented as shown in Fig. 3.12. Since String #1 has 7 modules less than other normal strings, it has a much reduced open-circuit voltage ($V_{OC}$), leading to a significantly changed $I-V$ curve of the whole PV array. The transient behavior is captured by the scope in Fig. 3.12(a). After the fault occurs at $t_1$, the reverse current into String #1 has a steady current (-42 A), which comes from other normal strings, such as String #2 to String #12. This means that most of PV array current is flowing into faulted String #1 instead of into the PV inverter. A negative current spike is found as -57 A at $t_1$. The additional reverse current other than the steady current (-42 A)
Fault occurs at $t_1$

Fault occurs at $t_1$

Fault occurs at $t_1$

(a) Transient behavior captured by the oscilloscope (Upper trace $v_{PV}$, 100 V/div; lower trace $i_1$, 10 A/div; time/division is 500 ms/div).

(b) Post-fault steady state captured by the PV monitoring board.

Figure 3.12: Experimental results of LL-7 (7-module LL fault).
may be caused by the discharging at input capacitor of the PV inverter. Notice that the array voltage \( V_{PV} \) drops to 280 V after \( t_1 \). This is the maximum voltage the faulted PV array can sustain (probably close to \( V_{OC} \) of the PV array). During \( t_1 \) to \( t_2 \), the power dissipated on String #1 is about 11.76 kW \( (= -42A \times 280V) \), which could damage the PV wiring and PV modules.

Under this circumstances, the top fuse at String #1 is blown by this large reverse current (-42A) at \( t_2 \). Thus, the fault is cleared successfully. After the fault clearance, the \( I-V \) curve of the whole array goes back to normal, except for losing one string. Meanwhile, the PV array voltage immediately increases to normal and causes a voltage “overshoot” at \( t_3 \). The control loop of the PV inverter regulates \( V_{PV} \) to its MPPT command \( (V_{CMD}) \) at \( t_4 \), which is similar to the one before \( t_1 \). After the fault clearance, \( v_{PV} \) goes back to normal.

### 3.4.4 More Experimental Results and Discussion

In addition to the previously discussed faults, three more faults are implemented in the PV laboratory, including \( LL-1 \), \( LL-3 \), and \( LL-4 \). All experimental results are summarized in Table 3.3 for comparison and discussion. Because PV faults occur under different solar irradiance levels, the pre-fault current of the faulted string varies in Table 3.3. But it does not affect the conclusion that the faulted string current decreases as more modules are involved in the fault. In addition, normal strings in the PV array contribute to the reverse current into the faulted string.
### Table 3.3: Summary of Experimental Results

<table>
<thead>
<tr>
<th>Fault Type</th>
<th>Pre-fault steady state</th>
<th>Transient</th>
<th>Post-fault steady state (new MPP)</th>
<th>Fault clearance by OCPD?</th>
</tr>
</thead>
<tbody>
<tr>
<td>LL-1 (at String #2)</td>
<td>$I_2 = 4.2$ A</td>
<td>No reverse current.</td>
<td>New MPP at $I_2 = \sim 3.3$ A</td>
<td>×</td>
</tr>
<tr>
<td>LL-2</td>
<td>$I_1 = 7.2$ A</td>
<td>No reverse current. $I_1 = 2.5$ A</td>
<td>New MPP at $I_1 = \sim 3.5$ A</td>
<td>×</td>
</tr>
<tr>
<td>LL-3</td>
<td>$I_1 = 6.4$ A</td>
<td>Small reverse current. $I_1 = -3$ A</td>
<td>New MPP at $I_1 = \sim 0.8$ A</td>
<td>×</td>
</tr>
<tr>
<td>LL-4</td>
<td>$I_1 = 5.7$ A</td>
<td>Medium reverse current. $I_1 = -12.7$ A</td>
<td>New MPP at $I_1 = \sim -5.5$ A</td>
<td>×</td>
</tr>
<tr>
<td>LL-5</td>
<td>$I_1 = 5$ A</td>
<td>Large reverse current. $I_1 = -23$ A</td>
<td>$I_1$ increases up to $-18.4$ A before fuse is blown.</td>
<td>✓</td>
</tr>
<tr>
<td>LL-7</td>
<td>$I_1 = 5$ A</td>
<td>Large reverse current. $I_1 = -42$ A</td>
<td>$I_1 = 0$ A after fuse is blown.</td>
<td>✓</td>
</tr>
</tbody>
</table>

Therefore, for a line-line fault involving certain modules, more parallel strings in the PV array tend to cause more reduced string current. Some key remarks from the experiments include:

- LL faults involving a few modules (e.g. LL-1 to LL-4) are difficult to clear by OCPD (i.e., fuses), since either the faulted string still has a positive current (LL-1 and LL-2) or a small negative current (LL-3 and LL-4). Therefore, theses faults can remain in the PV system and become a safety hazard.

- LL-5 fault is cleared by fuse successfully, which is triggered by the “$V_{OC}$ check” of the PV inverter. It is necessary to mention that the “$V_{OC}$ check”
depends on the specific PV inverters being used. Even without this “VOCC check,” the fuse will still be blown as long as the magnitude of the faulted string current is larger than $I_{nf}$, but it may take a longer time (up to hours) according to its time-current characteristics [72]. More importantly, the speed of MPPT plays a key role in the changing rate of faulted string current ($i_1/dt$). If the MPPT is updated too fast, $|i_1|$ may be reduced below $I_{nf}$ so quickly that it never becomes large enough to blow a fuse [1].

- LL-7 has a large reverse current that is easy to clear by fuses. Since the fault is cleared quicker (within 2 secs) than the MPPT’s response, $i_1$ remains the same during fault.

### 3.5 Conclusion

A 35 kW, 500 Vdc commercial-scale PV laboratory has been simulated and utilized for fault experiments under real-working conditions. The OCPD protection gap, MPPT response, and PV inverters transient behavior under faults are analyzed and discussed. Both the simulation results and experimental results show that the “blind spot” of OCPD exists when the line-line fault only has a few number of PV modules in the fault path. For example, the fault can be cleared by OCPD when the fault mismatch is higher than $\sim 41\%$ (e.g. at least 5 modules of a 12-module string) in this typical example. Thus the uncleared fault may remain in the system and become a safety hazard.
Detecting the changed direction of PV-string current may be an convenient way identify the reverse-current fault (e.g., \( LL-3 \) and \( LL-4 \) in our specific case). However, this approach may fail to identify \( LL-1 \) and \( LL-2 \) faults, since the faulted string has similar positive current as other normal strings. This brings difficulty in PV-string monitoring system for fault detection. Therefore, there is an urgent need of new fault detection method to detect line-line faults which only involve a few modules.
Chapter 4

Fault Detection Using Outlier Detection Rules

In order to eliminate the previously discussed “blind spots” in overcurrent protection devices (OCPD), outlier detection rules (ODR) are proposed for fault detection in this chapter. By measuring each PV string current, the proposed algorithms can differentiate the faulted string from the normal strings, even under the condition when the faulted string still has the positive current. (Note that the OCPD fails in this case.) Specifically, this chapter of the dissertation demonstrates the following achievements.

**Contribution 1:** Requiring only PV string current, two types of fault detection using ODR are developed that provide quantitative approaches to discover the faulted string (called outliers).
uses the statistical ODR. An alternative approach is called local outlier factor (LOF). Both proposed methods show several advantages over traditional fault detection methods: 1) simple to implement, 2) does not rely on weather information, and 3) suitable for real-time operation.

**Contribution 2**: The proposed methods are used on the experimental results in the PV lab under various conditions, such as line-line fault with 1 module (LL-1) and 2 modules (LL-2), showing that these faults (undetectable by OCPD) can be successfully detected. In addition, the performance of the proposed outlier rules are compared and discussed in detail.

### 4.1 Features for Fault Detection

To develop a responsive and reliable fault detection algorithm, the first step is to choose appropriate PV features that are convenient to measure and analyze. Although the previously discussed PV dynamic conductance $g_{pv}$ may be used for fault detection (see (3.2) and Fig. 3.4), it can only be found if the global I-V characteristic of the PV array is scanned when the PV system is off-line. Thus, it is not easy to observe or use $g_{pv}$ directly in real-time operation. Instead of $g_{pv}$, the following two features are more observable and viable for fault detection among PV strings.
4.1.1 String Current

PV string current is the most convenient feature to detect a fault in a series-parallel connected PV array. Besides, this feature is usually available when a PV string monitoring system is installed. Compared with its neighboring strings, a reduced string current $i_j$ on the $j$th String ($j = 1$ to $N$) might indicate a fault.

4.1.2 String Current Change Rate

A sudden current change at a particular string may indicate an unexpected operating condition. It represents the PV dynamics that are useful for fault detection and fault analysis. String current change rate $\frac{di_j}{dt}(k)$ of String #$j$ at the $k$th sample is defined as

$$\frac{di_j}{dt}(k) = \frac{i_j(k) - i_j(k-1)}{t(k) - t(k-1)} \quad (4.1)$$

where $i_j(k)$ is the $k$th sample of String #$j$ current and $t(k) - t(k-1)$ is the sampling interval. At the moment of the fault, the faulted string current $i_j$ usually reduces and leads to a significant drop of $\frac{di_j}{dt}$, which may be a fault indication.

Based on these two features, ODR for fault detection will be introduced as follows, which does not require weather information.
4.2 Proposed ODR for PV systems

4.2.1 Statistical ODR

According to [73], an outlier is usually defined as anomalous data with respect to the majority of the data set. A data set \( \{x_i\} \) with \( i = 1...n \) can represent PV string current, where \( n \) is the number of strings in parallel. The most common PV array configurations is the series-parallel connection (see Fig. 3.7). Assuming each parallel PV string has identical electrical ratings and similar environmental conditions, then each string should have similar current under normal conditions. In other words, the majority of PV string currents should have similar values, except for the outliers. This is the essential idea to identify outliers among PV strings.

Therefore, given certain environmental conditions, each string current becomes a univariate variable \( \{x_i\} \). The entire set of string currents can be viewed as random samples coming from a normal distribution \( N(\mu, \sigma^2) \), with two parameters such as mean \( \mu \) and variance \( \sigma^2 \) (\( \sigma \) is the standard deviation). Assume \( x_1...x_n \) are independent and identically distributed (i.i.d.). The normal probability density function \( p(x) \) is defined in (4.2).

\[
p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\} \tag{4.2}
\]
Given a PV array made of a number of PV strings, $p(x)$ is strongly related to electrical parameters, solar irradiance level and operating temperature on every PV string. For example, the $I$-$V$ curves shown in the Chapter 3 are mainly dependent on solar irradiance level, such that higher irradiance can lead to larger short-circuit current ($I_{SC}$) and therefore higher output power. As solar irradiance is changing continuously during a day, the parameters of $p(x)$ vary as well. Furthermore, the true values of $\mu$ and $\sigma^2$ are unknown in real-world conditions because of limited samples and measurement noise. Instead, $p(x)$ is estimated using the sample mean $\hat{\mu}$ in (4.3) and the unbiased sample variance estimate $\hat{\sigma}^2$ in (4.4), based on maximum likelihood estimation [74].

\[
\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{4.3}
\]

\[
\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\mu})^2 \tag{4.4}
\]

For demonstration purposes only, a normal distribution of PV string currents under certain environmental conditions is drawn in Fig. 4.1. Every dot represents a measurement of string current at specific time. Notice that most data values are located near the true mean $\mu$, except for two values $x_1$ and $x_2$. By common sense, $x_1$ may be selected as an outlier, since it is far away (more than $3\sigma$) from $\mu$. But
it is difficult to identify $x_2$. This leads to the need of a quantitative approach to set the outlier detection rules.

For the data set $\{x_i\}$, assume a reference value $x_0$, a measure of variation $\zeta$ and a threshold parameter $\alpha$. The statistical outlier rule (ODR) determines $x_i$ as an outlier if it satisfies the condition below:

$$|x_i - x_0| > \alpha \zeta \tag{4.5}$$

Then the ODR defined in (4.5) has the upper bound and lower bound as $x_0 + \alpha \zeta$ and $x_0 - \alpha \zeta$, respectively. Since the PV string usually has a reduced $i$ or negative $di/dt$ at the moment of the fault, for PV systems it is possible to ignore the upper bound and simplify the ODR to the lower bound only, as shown in (4.6).

$$x_j < x_0 - \alpha \zeta \tag{4.6}$$
The success of outlier detection greatly depends on the contamination level of the data set \( \{ x_i \} \), which is defined as the fraction of outliers in the data set (4.7) [73]. Intuitively, a higher contamination level in the data set tends to cause the deviation in parameter estimation and is likely to bring more difficulty to outlier detection.

\[
\text{contamination level} = \frac{\text{the number of outliers}}{\text{total number of samples}} \quad (4.7)
\]

Assuming the distribution of nominal data is symmetric and assigning different variables into (4.5), three most commonly used statistical ODR are obtained as follows [73].

- **The 3\( \sigma \) rule:** \( x_0 = \hat{\mu}, \; \zeta = \hat{\sigma} \), where \( \hat{\mu} \) is the sample mean, \( \hat{\sigma} \) is the sample standard deviation and \( \alpha = 3 \). Since \( \hat{\mu} \) is sensitive to outliers, the 3\( \sigma \) rule may break down at the contamination level greater than 10%.

- **The Hampel identifier:** \( x_0 = \tilde{x}, \; \zeta = S \), where \( \tilde{x} \) is the sample median, \( \alpha \) is usually chosen as 3, and \( S \) is defined as below. The Hampel identifier breaks down at the contamination level greater than 50%.

\[
S = \frac{1}{0.6745} \text{median}\{ |x_i - \tilde{x}| \} \quad (4.8)
\]

- **The standard Boxplot outlier rule:** \( x_0 = \tilde{x}, \; \alpha \) is usually chosen as 2, and \( \zeta = Q \), where \( Q = x_U - x_L \), \( x_L \) is the lower quartile (the 25th percentile) and
$x_U$ is the upper quartile (the 75th percentile). Besides, the Boxplot outlier rule can be extended as (4.9) for an asymmetric distribution of data set as below. It breaks down at the contamination level greater than 25%.

$$x_i > x_U + 1.5Q \text{ or } x_i < x_L - 1.5Q$$ (4.9)

4.2.2 Local Outlier Factor (LOF)

The second approach uses a spatial proximity-based local outlier factor (LOF, first proposed in [75]) for fault detection. The main difference between the statistical ODR and LOF is that LOF is a scoring technique that assigns an outlier score to each instance or sample. Therefore, it gives users the freedom to choose the outlying threshold. Intuitively, LOF is a density-based outlier detection rule showing that the density around an outlier is significantly lower than the density of its local neighbors. Quantitatively, a degree of being an outlier is defined as LOF, depending on how isolated the outlier is away from its neighbors [75].

At first, several basic notions/definitions are introduced.

- $\text{dist}(p,q)$: The distance between object $p$ and $q$. For a data set of one-dimensional PV string current $x_i$, Euclidean distance (2-norm distance) simply becomes the absolute value of $x_p - x_q$.  
• $k$-dist($p$): The distance of the object $p$ to the $k$-th nearest neighbor, where $k$ is a positive integer.

• $N_k(p)$: A set of $k$ nearest neighbors of an object $p$ contains every object in a given data set $D$, whose distance from $p$ is no greater than its $k$-distance.

$N_k(p) = \{ q \in D \setminus \{p\} \mid dist(p,q) \leq k$-dist($p$)\}$

• $reach$-$dist(p,q)$: The reachability distance of an object $p$ with respect to object $q$ (i.e., from $p$ to $q$) is defined as: $reach-dist(p,q) =$ $\max\{k$-dist($q$), $dist(p,q)\}$

• Local reachability density of $p$ is defined as:

$$lrd_k(p) = \frac{\|N_k(p)\|}{\sum_{q \in N_k(p)} reach-dist(p,q)} \quad (4.10)$$

where $\|N_k(p)\|$ represents the number of objects in $N_k(p)$. Intuitively, it shows that a larger reachability distance will lead to lower local reachability density. Note that $\|N_k(p)\|$ could be larger than the number $k$ since multiple objects may have same distance to the object $p$.

• $LOF$ of an object $p$ is defined as:

$$LOF_k(p) = \frac{\sum_{q \in N_k(p)} \frac{lrd_k(q)}{lrd_k(p)}}{\|N_k(p)\|} = \frac{\sum_{q \in N_k(p)} lrd_k(q)}{\|N_k(p)\|} \cdot \frac{1}{lrd_k(p)} \quad (4.11)$$
Fault Detection Using Outlier Detection Rules

Note that $LOF_k(p)$ is a degree of being an outlier. If $LOF_k(p)$ is significantly higher than its neighbors', the object $p$ is more likely to be an outlier. $LOF_k(p)$ only has one parameter $k$ and may vary as $k$ changes. The proper range of $k$ has been discussed in [75] and a large $k$ is usually recommended to highlight the outlier.

4.3 Fault Detection in the PV Laboratory

4.3.1 Statistical ODR Results

For PV applications, an underperforming or faulted PV string could be viewed as an outlier if its current $i$ or current change rate $di/dt$ deviates markedly from other normal ones, assuming PV strings have identical output current under similar environmental conditions. To verify the effectiveness of ODR in a commercial-scale PV system, we apply them in the previous $LL$-2 experimental results shown in Fig. 3.10.

The estimated solar irradiance is about 830 $W/m^2$. The outlier rules are applied to PV string monitoring board #1, leading to the 16.67% contamination level ($= 1/6$, one faulted string out of 6 strings per monitoring board). The $3\sigma$ rule is not considered in the following part of this dissertation, since it breaks down at contamination level higher than 10%, and it will surely fail in this specific example.
4.3.1.1 **ODR on String Current** $i$

The lower bound of $ODR$ is calculated at each sampling point in Fig. 4.2. The zoom-in window shows that the LL-2 fault (undetectable by fuses) can be successfully detected by both the Hampel identifier (4.8) and the Boxplot rule (4.9) at time $t_1$, since the faulted string current drops below the lower bound of these two rules.

4.3.1.2 **ODR on String Current Change Rate** $di/dt$

$ODR$ is applied on the previously defined $di/dt$ in Fig. 4.3. Similar to $ODR$ on $i$, only the lower bound violation is considered. Fig. 4.3 shows a significant negative spike on $di_1/dt$ below the lower bound of Hampel identifier and the Boxplot rule, which indicates a fault on String #1. After the fault occurs, $|di_1/dt|$ is often larger than normal strings. This is explained by its degraded $g_{pv}$ and the changed $I-V$ curve discussed in Chapter 3.

4.3.1.3 False alarms

Using $ODR$ on $i$ or $di/dt$ alone may cause false alarms since the statistical $ODR$ becomes aggressive when the majority of samples $\{x_i\}$ are nearly identical, according to (4.5) and (4.6). For example, $ODR$ causes false alarms on string current at time 1532 s in Fig. 4.4. Similarly, $ODR$ causes false alarms on $di/dt$ at time 1540 s and 1542 s in Fig. 4.5. To reduce false alarms, it is possible to send out
Chapter 4. Fault Detection Using Outlier Detection Rules

Figure 4.2: Apply statistical ODR on string current $i_1$ to $i_6$ of LL-2 experimental results (which cannot be cleared by fuses).
Chapter 4. *Fault Detection Using Outlier Detection Rules*

Figure 4.3: Apply statistical ODR on $\frac{di_1}{dt}$ to $\frac{di_6}{dt}$ of LL-2 experimental results (which cannot be cleared by fuses).
alarms only when $ODR$ on $i$ and $di/dt$ are both violated at the same time. Doing so can potentially eliminate the false alarms in Fig. 4.4 and 4.5.

The other approach is to use the $LOF$ to improve the fault detection performance, which will be given in the following subsection.
4.3.2 LOF Results

Each PV string current (e.g., \(i_1\ldots i_6\)) are viewed as an object in LOF. As shown in Fig. 4.6, the LOF results are plotted in 3 dimensions for better visualization. The \(x\), \(y\) and \(z\) axes are representing time, the string number, and LOF, respectively. Note that the parameter \(k\), the number of nearest neighbors, could be application-dependent. The practical guidelines for choosing \(k\) have been discussed in [75]. Basically, a relatively large \(k\) compared to the sample size can highlight the outlier and give a stable LOF outcome.

In this particular example, \(k\) is chosen as 4 (relatively large compared to a data set of 6 PV strings) for best performance. By setting an alarm threshold \(\theta\), the fault detection rule can be simply as: PV String \(#j\) is an outlier if its current \(i_j\) satisfies \(LOF_k(i_j) > \theta\). However, choosing \(\theta\) is not straightforward and often requires special consideration [76]. In our specific example, since \(LOF_k(i_j)\) of normal strings are usually under 5, we simply choose \(\theta = 5\) as the threshold.

Now, the LOF is applied on the previously discussed LL-2. Fig. 4.8(a) shows that during fault interval from \(t_1\) to \(t_2\), LL-2 has a significant increase in LOF, which is higher than \(\theta\) so that it can be easily identified as a fault. Under normal conditions, Fig. 4.8(b) shows that LOF can avoid false alarms caused by the statistical ODR (previously shown in Fig. 4.4 and 4.5). The reason is that LOF uses reach-
\[
\text{dist}(p,q) = \max\{k\cdot \text{dist}(q), \text{dist}(p,q)\}
\]
that can reduce the statistical fluctuations of \(\text{dist}(p,q)\) [75]. This may avoid the false alarm found in statistical rules.
(a) LOF on LL-2 fault.

(b) LOF on normal conditions.

Figure 4.6: Fault detection using LOF on experimental results of LL-2 and normal conditions.
4.3.3 More Results under Various Solar Irradiance

In addition to experimental results under high irradiance, $LL-2$ under medium and low solar irradiance ($\sim 410\text{W/m}^2$ and $\sim 170\text{W/m}^2$) are also studied and shown in Fig. 4.7 and 4.8. The figures illustrate that both of the statistical $ODR$ and $LOF$ are able to detect the fault at time $t_1$.

4.4 Discussion

The experimental results verify the effectiveness of the proposed statistical $ODR$ (e.g., Hampel identifier and Boxplot rule) and local outlier factor ($LOF$) for fault detection using PV string currents and current change rate. Both of them can successfully detect the line-line fault with small mismatch that may not be cleared by OCPD. The advantages such as simplicity, responsiveness and requiring no weather information make them prime candidates for real-time fault detection in PV string monitoring systems. Furthermore, a comparison between the statistical $ODR$ and $LOF$ is briefly summarized in Table 4.1.

For both the statistical $ODR$ and $LOF$, the accuracy of outlier detection relies on the number of samples $\{x_1...x_N\}$. The intuition is that if more samples/measurements are available, the true distribution $p(x)$ can be better estimated. Therefore, increasing the number of PV strings for outlier detection may decrease the contamination level in the data set and improve the overall performance.
Chapter 4. *Fault Detection Using Outlier Detection Rules*

3.2

3.3

3.4

3.5

3.6

String current (A)

-0.05

0

0.05

Time (second)

di/dt (A/sec)

String #1

String #2

String #3

String #4

String #5

String #6

Hampel identifier

Boxplot rule

Faulted String #1

Normal strings

Lower bound of ODR

Fault occurs at time $t_1$

(a) Statistical ODR under LL-2 fault at time $t_1$.

(b) LOF under LL-2 fault at time $t_1$.

Figure 4.7: LL-2 under medium solar irradiance ($\sim 410 W/m^2$).
Chapter 4. Fault Detection Using Outlier Detection Rules

Figure 4.8: LL-2 under low solar irradiance (∼170W/m²).
Table 4.1: A brief comparison between the statistical ODR and LOF in PV fault detection

<table>
<thead>
<tr>
<th>Outlier Detection Rules</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical ODR</td>
<td>Outlier rules in meaningful units. Lower bounds provide easy data interpretation.</td>
<td>Binary property. May become aggressive and cause false alarms.</td>
</tr>
</tbody>
</table>

One limitation of the proposed method is that it may cause false alarms if unknown partial shading occurs on PV strings. For example in Fig. 4.9, both the statistical ODR and LOF may send out alarms under partial shading on String #9. Therefore, by measuring string current alone, the proposed methods are difficult to differentiate line-line faults from partial shadings. In this case, other artificial intelligence techniques with additional measurements should be adopted [40]. On the other hand, if the partial shading’s pattern is periodic and can be estimated, such as using artificial neural network (ANN) [36], the shaded string current may be corrected to a unshaded value. Then the proposed method should still work on corrected PV data.
Chapter 4. Fault Detection Using Outlier Detection Rules

Figure 4.9: False alarms may be caused by partial shading on PV strings.
4.5 Conclusion

To eliminate same “blind spots” in fault detection, this chapter proposes two types of outlier detection rules (\(ODR\)) that are suitable for real-time operation in a PV-string current monitoring system. The first one uses the statistical \(ODR\), such as the Hampel identifier and the Boxplot rule. But sometimes they may become aggressive in fault detection and cause false alarms. To avoid false alarms, one proposed solution is to use the statistical \(ODR\) on multiple features, such as string current or current change rate. The other proposed solution is to use a machine-learning outlier rule—the local outlier factor (\(LOF\)), which is the second type of \(ODR\) in this dissertation. \(LOF\) has settable outlying threshold and shows better reliability in fault detection. The advantages of the statistical \(ODR\) and \(LOF\) over traditional fault detection are: 1) simple to implement, 2) independent of weather information, and 3) suitable for real-time operation.

The effectiveness of the proposed methods are verified by the experimental results in the PV laboratory under real-working conditions, including the MPPT response from the PV inverter. The proposed statistical \(ODR\) and \(LOF\) can detect the fault successfully under various solar irradiance levels (e.g., high, medium and low solar irradiance). A brief comparison between the statistical \(ODR\) and \(LOF\) and their limitations under partial shading conditions are also discussed.
Chapter 5

Fault Classification Using Machine-Learning Algorithms

5.1 Introduction

In addition to fault detection (mainly discussed in Chapter 4), this dissertation proposes fault classification as another useful feature in PV systems. Fault classification can indicate the type of fault and further help maintenance people to expedite the system’s recovery. Based on the prior knowledge of the systems, machine learning algorithms are usually used to achieve automatic fault classification. This chapter studies fault classification methods and focuses on graph-based semi-supervised learning (GBSSL) algorithms for PV fault detection and classifications (FDC).
5.1.1 Existing Solutions and their limitations

Machine learning techniques have been proposed for both fault detection and classification (FDC) in PV systems [27, 40, 41, 77, 78]. A decision-tree model has been proposed to detect and classify fault types in PV arrays [40]. A clustering-based method is used for quantifying PV system’s effects on utility grids [78]. K-nearest neighbor and support vector machine are used for FDC in solar panels [41].

For example, Fig. 5.1 shows a classifier using Support Vector Machine (SVM) model that has been trained to identify the open-circuit fault in one string ("OPEN1") from the normal condition ("NORMAL"). The classification features use $V_{NORM}$ and $I_{NORM}$, which will be explained in the following section. The SVM model is trained using the PV data collected at temperature $30^\circ C$ in summer. However, Fig. 5.1 shows that the trained model will mistakenly classify "OPEN1 at temperature $0^\circ C" as normal conditions, since the operating points are shifted away in winter. Because the operating points of solar arrays are dependent on weather, they may change over a year. To solve this problem, the fault detection algorithm needs to be self-learning and should be updated continuously.

Therefore, these supervised-learning classification models have several drawbacks as:
Figure 5.1: Limitation of supervised learning: SVM model trained in summer will not work properly in winter.

Drawback 1: As pointed out in Fig. 5.1, the challenge lies in the non-linear operating point of the PV array (dc side), which varies widely as the environment changes or solar cells degrade. This means that a trained model in summer (high irradiance, high temperature) might mistakenly classify normal PV operation as a fault during winter (low irradiance, low temperature). Therefore, the trained model needs updating continuously over time.

Drawback 2: The supervised learning models may require a large amount of training data, which could hinder their effectiveness. Since the training data are collected in typical faults under real-working conditions, it
is difficult or expensive to obtain.

Drawback 3: The supervised learning models are designed only based on specific operating data on certain PV installations. This means the model is built on a case-by-case basis, and it may not be possible to have a general fault detection model. In addition, the models are difficult to visualize by PV maintenance people.

5.1.2 Research Contributions Obtained in this Chapter

To address the issues above, this dissertation proposes graph-based semi-supervised learning (GBSSL) for fault detection and classification (FDC) in PV arrays. The proposed method can detect faults that are unnoticeable to conventional OCPD or GFDI, as well as identify the fault type. Instead of requiring additional sensing circuit in [27, 40, 41, 77, 78], the proposed method only uses readily available measurements, such as PV array voltage and array current (commonly used for MPPT in PV inverters), PV operating temperature and solar irradiance (often available at weather stations). Therefore, the proposed method avoids additional hardware installation or extra labor cost to conventional PV systems.

This chapter focuses on two categories of common faults that may be difficult to detect or clear by conventional OCPD. 1) Line-line fault, defined as an accidental short-circuiting between two points in the array with different potentials. It may be caused by a short-circuit fault between two different points or a double ground
fault in PV fields, and 2) Open-circuit fault, defined as an accidental disconnection at a normal current-carrying conductor, which can be caused by cracking PV cells or modules, loose string connections or blown PV fuses. Note that ground fault is not directly discussed in this dissertation, since essentially it is a special case of line-line faults involving a ground point. Also, it is relatively easy to detect and clear by GFDI or residual current device (RCD).

In summary, this chapter presents the following research contributions:

**Contribution 1:** A new method/new algorithm is presented that normalizes measured data in a specific manner that allows machine learning to detect and classify faults that previously were not detectable.

**Contribution 2:** For the first time, a graph-based semi-supervised learning (GB-SSL) model is developed for fault detection and classification in solar PV arrays. The model has several advantages over the previously reported solutions, such as low model training cost, self-learning ability over time in real-time operation, and better data visualization.

**Contribution 3:** The fault detection methods are designed to be included within the controller of existing PV inverter technologies. The approach is able to work with most any PV inverter topology and takes advantage of readily available measurements in existing PV systems.
Contribution 4: The fault detection methods are able to detect dangerous line-line faults, which cannot be detected by most known methods. The effectiveness of the proposed GBSSL model is validated in both a simulation platform and a small-scale PV experimental system under real working conditions.

5.2 Fundamentals of Machine-Learning Algorithms

5.2.1 Fundamentals of Machine Learning Techniques

As a subarea of artificial intelligence, machine learning is a learning algorithm that automatically extracts knowledge from a given data set. The data set usually is composed by a number of instances (similar to measurements), and each instance has two following parts. The first part is called attribute, which is a feature of an instance that determines its classification. The second part is called class label, identifying which category this instance belongs to. For example, a data set may include n number of instances $x_i$ ($i = 1$ to $n$). Each instance $x_i$ is annotated in a set of $\{x_{1,i}, x_{2,i}, \ldots, x_{d,i}, y_i\}$, including $d$ number of attributes (d-dimensional data) and a class label $y_i$. After fault classification implementation, class labels $y_i$ will be identified as normal condition or any specific fault type. The data set is called labeled data if its class labels $y_i$ are known. Otherwise, it is unlabeled data.
Chapter 5. Fault Classification Using Machine-Learning Algorithms

![Diagram of machine learning techniques: (a) supervised, (b) unsupervised learning, (c) semi-supervised learning.](image)

Figure 5.2: Simplified flowchart of machine learning techniques: (a) supervised, (b) unsupervised learning, (c) semi-supervised learning.

Generally, machine learning techniques can be divided into three categories: supervised learning, unsupervised learning and semi-supervised learning (see Fig. 5.2) [79]. As seen in Fig. 5.2(a), supervised learning uses completely labeled training data with known class labels. However, labeling data is expensive, requiring human effort and expertise to classify. In contrast, unsupervised learning in Fig. 5.2(b) uses only unlabeled training data. In Fig. 5.2(c), semi-supervised learning (SSL) falls between supervised learning and unsupervised learning, since it uses both labeled and unlabeled data for training. This chapter focuses on SSL since it only requires a few of labeled data and can improve learning accuracy considerably using inexpensive unlabeled data.

5.2.2 Machine Learning Techniques in PV Systems

The overview of the proposed PV system is shown in Fig. 5.3 schematically, including a typical grid-connected PV system and the proposed machine-learning
GBSSL model for FDC.

- The typical PV system includes: a PV array, a PV inverter, utility grid and conventional fault protection devices (i.e. OCPD and GFDI). The PV inverter harvests the maximum output power from the PV array using the MPPT algorithm, and feeds the power into the utility grid. The PV array exhibits non-linear current vs. voltage ($I-V$) curves, whether the PV array is under normal or fault conditions. The previous chapters in this dissertation have discussed that when the PV array is faulted, it has a changed configuration, resulting in changed $I-V$ curves and reduced maximum power points (MPP). After that, if the fault is not cleared properly, it is likely that the PV inverter will still work, as long as the PV array can achieve the minimum operating voltage of the PV inverter [1]. Consequently, the faulted PV array is expected to work at a new, but sub-optimal, possibly hazardous MPP along with its faulted $I-V$ curves. It is worth mentioning that the resulting
deviated PV-array MPPs can provide helpful information for the proposed GBSSL model.

- The proposed GBSSL model uses readily available PV data. The GBSSL model for FDC measures the instantaneous short-circuit current reference ($I_{SC-REF}$) and the open-circuit voltage reference ($V_{OC-REF}$) of the reference modules (small, lower power), and receives the PV-array current ($I_{MPP}$) and voltage ($V_{MPP}$) at PVs MPP from the inverter. Alternatively, instead of using reference modules, it is also possible to obtain $I_{SC-REF}$ and $V_{OC-REF}$ from simulation models using instantaneous solar irradiance and solar cell temperature monitored by weather stations that are commonly installed in PV fields. Therefore, the proposed method can take advantage of readily available PV data without adding significant hardware costs.

- The proposed GBSSL model can be integrated into PV inverters. The GBSSL model only monitors the PV array (dc side) at MPP under steady state, so it does not rely on any particular power conversion unit. Thus, another advantage of the proposed method is the ease of integration within a PV inverter of any circuit topology.
5.3 Improved Data Visualization using Normalized Parameters

The operating MPPs of PV arrays–PV voltage ($V_{MPP}$) and PV current ($I_{MPP}$)–range widely over a year. These fluctuations are caused by their dependence on environmental conditions, such as solar irradiance and temperature. For example, in Fig. 5.4 the MPPs of a specific PV array under normal (NORMAL) and fault conditions are simulated and plotted in two dimensions–$V_{MPP}$ vs. $I_{MPP}$. The fault conditions (as class labels) are known and include line-line faults (LL) and open-circuit faults (OPEN). In Fig. 5.4, even though LL faults tend to have lower $V_{MPP}$ than NORMAL and OPEN, it demonstrates the common occurrence of overlapping on MPPs. In other words, the faulted PV array may have similar $V_{MPP}$ and $I_{MPP}$ to a normal array, which creates challenges in identifying faults from normal conditions. To better visualize and identify the PV faults, the normalized parameters are introduced in Fig. 5.5 as the data attributes, which show better data clustering and easier data visualization than the original attributes ($V_{MPP}$ vs. $I_{MPP}$ in Fig. 5.4). Detailed discussion about Fig. 5.5 will be given in Section IV. Two new normalized parameters are:

- The normalized PV voltage $V_{NORM}$. Defined as $V_{MPP} / (N_{MOD} \ast V_{OC-REF})$, where $N_{MOD}$ is the number of modules in series per PV string, and $V_{OC-REF}$ is the open-circuit voltage reference of the PV module.
Chapter 5. *Fault Classification Using Machine-Learning Algorithms*

![Figure 5.4: The overlapping MPPs of a PV array over a wide range of irradiance and temperature. This is difficult for FDC.](image)

- The normalized PV current $I_{NORM}$. Defined as $I_{MPP} / (N_{STR} \times I_{SC-REF})$, where $N_{STR}$ is the number of strings in parallel in the array, and $I_{SC-REF}$ is the short-circuit current reference of the PV module.

The helpfulness of the visualization in Fig. 5.5 depends on the accuracy and lifecycle degradation of the reference measurements. It is assumed that the reference modules and the PV arrays degrade similarly or at insignificant rates. However, since the reference modules and the PV array have identical working environments, such as ultraviolet (UV) exposure, thermal stress, or humidity, the PV array and reference modules are expected to have similar front surface soiling or optical degradation over time [66, 80]. In addition, the effect on $V_{OC-REF}$ and $I_{SC-REF}$ caused
by the cell temperature difference associated with short-circuit and open-circuit conditions is insignificant [81]. As a result, it is advantageous to have normalized parameters $V_{\text{NORM}}$ and $I_{\text{NORM}}$, which would remain constant, even as the PV modules degrade uniformly. In this dissertation, the reference modules are chosen to have the same PV model parameters as other operating modules in the PV array. However, any partial shadings or mismatch on the reference modules or over the PV array may cause bad data in $V_{\text{NORM}}$ and $I_{\text{NORM}}$, leading to incorrect fault detection and classification. For this reason, the partial shading on PV arrays is not considered in the dissertation.
5.4 Graph-based Semi-supervised Learning (GB-SSL) Algorithm in Solar PV Arrays

5.4.1 Introduction of Graph Data in PV Arrays

![Graph Diagram]

**Figure 5.6:** A graph data consisting of edges and nodes.

The proposed GBSSL presents PV data in an undirected graph $G = (X, E)$ as shown in Fig. 5.6, which consists of two types of elements, namely vertices $X$ and edges $E$. If there is always a path for every vertex to any other vertex, the graph is said to be connected. It will be assumed that there are no internal loops or multiple edges in the graph.

Graphs can be used to represent real-world problems, in that graph nodes are data points, and graph edges are weighted so that they can encode similarities between points [82]. Therefore, graphs can also be used to represent operating points (voltage and current) in solar PV arrays.
Labels of NORMAL                                  Labels of FAULT
Unlabeled MPPs of NORMAL                  Unlabeled MPPs of FAULT
I-V curves of NORMAL                     I-V curves of FAULT

Figure 5.7: An illustrative example of GBSSL for PV: (a) Initial labels $x_1$ and $x_2$ with unlabeled PV data. (b) Unlabeled data are classified by GBSSL.

The GBSSL algorithm proposed in [82, 83] will be extended to be applied to solar PV arrays. For illustration purpose, in Fig. 5.7, PV normal condition (NORMAL) and fault condition (FAULT) are considered in a normalized $I-V$ curve (in 2-D), which has possible normalized MPPs of a PV array. Note that normalized MPPs vary slightly during the irradiance and temperature changes, since they tend to cluster in their own category (NORMAL or FAULT). Also, the GBSSL assumes that nearby points are likely to have the same class label. The goal of the GBSSL algorithm is to let the labeled data, for example $x_1$ and $x_2$ in Fig. 5.7(a), spread their class label information to unlabeled neighboring points in order to achieve a global stable state. As a result of GBSSL in Fig. 5.7(b), unlabeled data will be classified by the initial labels accordingly. This demonstrates the growing label size of the proposed algorithm and its self-learning ability in real-time operation. As more data are collected and labeled, the GBSSL for FDC on this specific PV installation is improved. Thus, a few initial labeled data allow a large amount of
unlabeled data to be classified.

### 5.4.2 Detailed GBSSL Algorithm for FDC

A data set $X = \{x_1, \ldots, x_2, x_{l+1}, \ldots, x_n\}$ is represented by a $d \times n$ matrix as (5.1) shown below, in which each column vector $x_i = [x_{1,i} \ x_{2,i} \ \ldots \ x_{d,i}]^T$ is an instance in the data set $X$. A label set $L = \{1, \ldots, c\}$ has the number of $c$ class labels. A label column vector is $y_i = [y_1 \ y_2 \ \ldots \ y_n]^T_{1 \times n}$, in which each entry $y_i \in L$ representing the class label of the instance $x_i$. The first $l$ instances $x_i$ ($i \leq l$) in $X$ are labeled data with labels $y_i$. The remaining instances $x_u$ ($l + 1 \leq u \leq n$) in $X$ are unlabeled data which have $y_u = 0$ ($l + 1 \leq u \leq n$).

$$X = \begin{bmatrix}
  x_{1,1} & \ldots & x_{1,l} & x_{1,l+1} & \ldots & x_{1,n} \\
  x_{2,1} & \ldots & x_{2,l} & x_{2,l+1} & \ldots & x_{1,n} \\
  \vdots & \ddots & \vdots & \ddots & \ddots \\
  x_{d,1} & \ldots & x_{d,l} & x_{d,l+1} & \ldots & x_{1,n}
\end{bmatrix}_{d \times n}$$  \hspace{1cm} (5.1)

Specifically, by using the proposed normalized attributes shown in Fig. 5.5, the matrix $X$ has two rows ($d = 2$, two dimensions for $V_{NORM}$ and $I_{NORM}$) and $n$ number of columns. The two entries for each column vector $x_i$ are defined in (5.2) as an instance of PV measurement at time $t_i$. 
$x_i = [V_{\text{NORM}_i}, I_{\text{NORM}_i}]^T$  

As shown in Fig. 5.8, there are five steps in the flowchart of the proposed GBSSL algorithm for FDC. Detailed explanations are given as follows.

Step 1): First, the prior knowledge about the system is given, including the initial labeled data $x_1, \ldots, x_i$ and their class labels $y_1, \ldots, y_i$. For example, the number of class labels is $c = 3$ in our PV application. Thus $y_i = 1, 2$ or 3 represent NORMAL, LL or OPEN conditions.

Step 2): New instances are recorded with unknown labels ($y_u = 0$, $l+1 \leq u \leq n$). Then the data set matrix $X$ and the label column $Y$ are constructed.
accordingly. Using the label column $Y$, a label matrix $Z$ is established of size $n \times c$ ($c = 3$ in our PV application), whose elements $z_{ij}$ at the $i$th row and the $j$th column are defined in (5.3).

$$z_{ij} = \begin{cases} 
1, & \text{if } y_i = j \in \{1, \ldots, c\} \\
0, & \text{otherwise}
\end{cases} \quad (5.3)$$

After that, a weight matrix $W$ (size $n \times n$) is introduced to represent the closeness of its entries $w_{ij}$ as defined in (5.4). Thus, the matrix $W$ is a pair-wise relationship matrix on the data set $X$ with zero diagonal elements.

$$w_{ij} = \begin{cases} 
\exp(-\|x_i - x_j\|^2/2\sigma^2), & \text{if } i \neq j \\
0, & \text{if } i = j
\end{cases} \quad (5.4)$$

where $\sigma = 1$ as the bandwidth parameter. For example, in Fig. 5.6 the edges between the points $x_i$ and $x_j$ in the graph are weighted by the element $w_{ij}$. Equation (5.4) implies that if $i \neq j$, $w_{ij}$ is large when $x_i$ and $x_j$ are close to each other, and small when they are far way. Next, a matrix $S = D^{-1/2}WD^{-1/2}$ (size $n \times n$) is established, where $D$ is a diagonal matrix with the diagonal element $(d_{ii})$ equal to the sum of the $i$th row of the matrix $W$: $d_{ii} = \sum_j w_{ij}$. Note that the weight matrix $W$ is normalized symmetrically.
Step 3): Once the matrices $Z$ and $S$ are built, the proposed GBSSL can be implemented to find the classification solution matrix $F$ according to (5.5), where $0 < \alpha < 1$ and $(I - \alpha S)^{-1}$ can be viewed as a graph or diffusion kernel [82].

$$F = \lim_{t \to \infty} F(t) = (I - \alpha S)^{-1}Z \quad (5.5)$$

The solution matrix $F$ is a $n \times c$ matrix (the same size as the matrix $Z$) with nonnegative entries $f_{ij}$, where the row vector is $F_i = [f_{i,1} \, f_{i,2} \, \ldots \, f_{i,c}]$.

$$F = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,c} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,c} \\ \vdots & \vdots \\ f_{n,1} & f_{n,2} & \cdots & f_{n,c} \end{bmatrix}_{n \times c} \quad (5.6)$$

Step 4): The label matrix $Z$ and label column vector $Y$ can be updated based on the classification result—the matrix $F$ found previously. The rule of classification is straightforward: The matrix $F$ corresponds to a classification criteria on the point set $X$ by labeling each instance $x_i$ as a label $y_i = \arg \max_{j \leq c} f_{ij}$. In other words, the class label for each point $x_i$ will be chosen as the class $j \ (1 \leq j \leq c)$ if $f_{i,j}$ is the maximum entry in the corresponding row vector $F_i$. For example, there are 3 classes in the PV application ($c = 3$), in relation to class labels Class1–NORMAL,
Class2–LL, and Class3–OPEN. For the instance $x_i$, if the corresponding row vector $F_i$ equals to $[0.1, 0.2, 1.3]$, then $x_i$ will be classified into Class3–OPEN, so that $y_i = 3$, since the 3rd entry in $F_i$ has the maximum value.

Step 5): Fault detection and classification results will be indicated continuously. If any faults are identified, fault alarms will be sent out.

### 5.5 Simulation Results

#### 5.5.1 Simulated PV Systems

Using the widely used one-diode model [84] for each individual solar module/panel, this dissertation builds another simulation PV system (17.6kW) in MATLAB/Simulink consisting of $10 \times 10$ PV modules (monocrystalline silicon) that is capable of studying faults among modules. Note that this simulation system is smaller than the simulation system presented in Chapter 3 (35kW, $12 \times 12$ PV modules). The schematic diagram is shown in Fig. 5.9. The number of series modules per string is $N_{MOD} = 10$, and the number of parallel strings is $N_{STR} = 10$. The main parameters of each PV module at standard test conditions (STC) are as follows: the maximum power $P_{MPP} = 176W$, the open-circuit voltage $V_{OC} = 44.4V$, the maximum power voltage $V_{MPP} = 35.7V$, the short-circuit current $I_{SC} = 5.4A$, the maximum power current $I_{MPP} = 4.95A$. 
5.5.2 Faults in PV Systems

As shown in Fig. 5.9, there are two categories of faults in the PV systems: line-line faults (LL) and open-circuit faults (OPEN). The normalized MPPs under a variety of normal and fault conditions have been plotted in Fig. 5.5.

1) Normal condition (NORMAL)

Under a wide range of environmental conditions of changing solar irradiance and temperature, the normalized MPPs have the following operating range: \( V_{\text{NORM}} \in (0.77, 0.90) \) and \( I_{\text{NORM}} \in (0.86, 0.91) \).
2) **Line-line fault (LL)**

A variety of line-line faults without or with fault resistance \( (R_f = 0 \Omega \text{ or } 10 \Omega) \) are studied. The fault resistance for typical line-line PV faults is considered as \( 0\Omega \) at solid faults, or as \( 10 \Omega \) which may be caused by poor connections or dc arcs [28]. The fault between the fault point Fault1 and negative conductor (Fault1-Neg) in Fig. 5.9 is defined as “30% location mismatch \( (LL \ 30\%) \),” since it involves 3-module mismatch between the fault points in the faulted string (normally 10 modules per string). Similarly, the Fault2-Neg fault is defined as “40% location mismatch \( (LL \ 40\%) \)”. Compared with NORMAL, \( I_{NORM} \) under \( LL \) is slightly reduced to a range of \( (0.81, 0.91) \), but \( V_{NORM} \) is significantly decreased which lies in a wide range of \( (0.57, 0.78) \). The reason is that the MPPT tends to reduce \( V_{MPP} \) to reach the sub-optimal MPP under \( LL \) faults, leading to a reduced \( V_{NORM} \). Meanwhile, \( I_{NORM} \) may be reduced as well, since the fault tends to reduce \( I_{MPP} \).

3) **Open-circuit fault (OPEN)**

Open-circuit faults on one string (OPEN1) and two strings (OPEN2) are included in Fig. 5.5. Notice that \( V_{NORM} \) of OPEN faults remains the same as the one in NORMAL conditions. But \( I_{NORM} \) is reduced proportionally by the number of lost strings, due to the parallel connection of PV strings.
5.5.3 Simulation of GBSSL for FDC

1) Initial Labels

As shown in Fig. 5.5, the normal condition is annotated as NORMAL; the line-line faults include “LL 30% $R_f=0$,” “LL 30% $R_f=10$,” “LL 40% $R_f=0$,” and “LL 40% $R_f=10$;” the open-circuit faults include “OPEN1” and “OPEN2.” The number of initial labels is only 10 (approximately 1.4% of 697 instances in each specific fault) for each of the previously mentioned PV conditions. The initial label data are PV MPPs under limited environmental conditions: solar irradiance ranging from 450 to 900 W/m$^2$ and ambient temperature fixed at 20°C. One advantage of the GBSSL is that it uses only use a few initial labels to classify a large amount of new PV operating data, which are called test data as below.

2) Measurement Uncertainty

The accuracy and effectiveness of any GBSSL methods depend on the reliability of the measured signals. Measurement can be expressed as the following equation.

$$\text{measurement} = \text{best estimate} \pm \text{uncertainty} \quad (5.7)$$

To successfully identify the fault type, the measured instances in three main fault categories (NORMAL, LL and OPEN) should have no overlapping in the normalized parameters (i.e., $V_{NORM}$ and $I_{NORM}$ in Fig. 5.5). In our specific
example shown in Fig. 5.5, the minimum necessary distance between any two
categories in the horizontal axis \(V_{\text{NORM}}\) and the vertical axis \(I_{\text{NORM}}\) are
annotated as \(dV\) and \(dI\), respectively. To avoid overlapping, the measurement
uncertainty in \(V_{\text{NORM}}\) and \(I_{\text{NORM}}\) should be no more than \(dV/2\) and \(dI/2\),
respectively. From the simulations in Fig. 5.5, \(dV = 0.02\) and \(dI = 0.035\). This
means that the measurement uncertainty in \(V_{\text{NORM}}\), denoted by \(\Delta V_{\text{NORM}}\), must
satisfy \(\Delta V_{\text{NORM}} < 0.01\). Similarly, \(\Delta I_{\text{NORM}} < 0.0175\) must be satisfied for the
GBSSL method to work. If it is assumed that the reference modules have higher
accuracy measurement equipment, then the reference signal measurement error
is relatively small compared to the array voltage and current measurements.
Then in our example for the \(N_{\text{MOD}} = 10\) modules in series for each string
with each panel having \(V_{\text{OC-REF}}=44.00\text{V}\) (at STC), this would lead to the
requirement that \(\Delta V_{\text{MPP}}\) (the PV array voltage measurement error) should be
less than \(4.40\text{V} = 0.01 \times N_{\text{MOD}} \times V_{\text{OC-REF}}\). Similarly, since each PV panel
is assumed in the simulation to have \(I_{\text{SC-REF}} = 5.40\text{A}\) (at STC), and there
are \(N_{\text{STR}} = 10\) paralleled strings, the acceptable PV array current (into the
inverter) measurement error \(\Delta I_{\text{MPP}}\) should be less than \(0.945\text{A} = 0.0175 \times
N_{\text{STR}} \times I_{\text{SC-REF}}\).

However, it is possible that the measurement error for the reference modules
may not be negligible. In that case, the approaches of [85] may be used to
provide bounds for the individual measurement errors: Consider a normalization function as \( f(a, b) = \frac{a}{N \cdot b} \), where \( N \) is a constant, \( a \) and \( b \) are measurements with uncertainties \( \Delta a \) and \( \Delta b \). In our PV application, the normalization function is either \( V_{NORM} \) or \( I_{NORM} \); the parameter \( a \) can represent \( V_{MPP} \) or \( I_{MPP} \); and the parameter \( b \) can represent \( V_{OC-REF} \) or \( I_{SC-REF} \). The constant \( N = N_{MOD} = N_{STR} \) is 10. Then the propagation of the measurement uncertainty in \( f(a, b) \) is defined as \( \Delta f \) in the following equation \([85]\).

\[
\Delta f^2 = \left( \frac{\partial f}{\partial a} \right)^2 \Delta a^2 + \left( \frac{\partial f}{\partial b} \right)^2 \Delta b^2
\]

\[
= \frac{\Delta a^2}{(N \cdot b)^2} + \left( \frac{a}{-N \cdot b^2} \right)^2 \Delta b^2
\]

(5.8)

If we assume \( \Delta a \geq N \cdot \Delta b \) and \( a \leq N \cdot b \), we can simply find the maximum \( \Delta f \) in (5.9).

\[
\Delta f_{\text{max}} = \frac{\sqrt{2} \cdot \Delta a}{a}
\]

(5.9)

For this specific example with uncertainty in both the reference modules and PV array measurements, \( V_{MPP} \) should be less than 3.10V and similarly \( I_{MPP} \) should be less than 0.668A. It is worth mentioning that the calculated measurement uncertainties are only good to identify the specific faults in Fig. 5.5.
For a different PV system, the normal and fault conditions may exhibit various $dV$, $dI$ or other parameters. This is likely to result in changed measurement uncertainties.

3) **Test data**

In order to test the robustness of the proposed GBSSL model for PV FDC, two types of PV data are utilized that are not included in initial labels, namely new environmental conditions and new fault conditions.

- First, we use the PV data under wide-ranging environmental conditions that the GBSSL has hardly seen before. To cover all possible working conditions, the PV simulation covers the combinations of module-plane solar irradiance ($G_T$) extensively varying from 200$W/m^2$ to 1000$W/m^2$ with step change of 50$W/m^2$, and ambient temperature ($T_{amb}$) changing from -10°C to 30°C by 1°C. It is useful to test the robustness of the proposed GBSSL. The solar cell temperature ($T_{cell}$) can be estimated by a simplified empirical function in (5.10), where the nominal operating cell temperature ($NOCT$) is chosen as 50°C [86].

\[
T_{cell} = T_{amb} + \frac{NOCT - 20°C}{800W/m^2} \cdot G_T
\]  

(5.10)

- Second, new fault conditions (could be more challenging) are added into the test data, including “LL 10% $R_f=0$,” “LL 10% $R_f=10$,” “LL 20%
Chapter 5. Fault Classification Using Machine-Learning Algorithms

$R_f=0$, “LL 20% $R_f=10$,” open-circuit faults on 3 strings (OPEN3) and 4 strings (OPEN4). These types of faults are not included in the initial data so they are “unseen” for the GBSSL model. They are meant to test the performance of the GBSSL on new data set.

4) How GBSSL works at PV Test Data

Test data are fed into the GBSSL model one after another as real-time operation. This means only one new instance at every GBSSL iteration, which implies that $n = l + 1$ in (5.1). The initial labels for the GBSSL are illustrated in the normalized parameters shown in Fig. 5.10. The FDC procedure is explained using the following example: At time $t = t_1$, there is a new instance ($x_1$) entering the proposed model, which indicates a sudden change of PV operating point. Based on the given initial labels as well as the similarity between $x_1$ and various classes, it is clear to see that $x_1$ is closer to “NORMAL” than any other cluster. Therefore, GBSSL will classify $x_1$ as class “NORMAL” and store it as a new label. Thus, the size of labeled data can increase as more data is collected and identified, demonstrating the self-learning ability of the GBSSL. Similarly, new data points $x_2$ at $t = t_2$ and $x_3$ at $t = t_3$ will be correctly classified as “LL” class and “OPEN” class, respectively. As previously discussed, a gradual and uniform PV aging over years is likely to be classified as “NORMAL,” since it is a slow deviation from “NORMAL” working MPPs. Therefore, the proposed GBSSL may avoid nuisance tripping on such events.
In the previous discussion the proposed GBSSL always utilized the PV arrays MPP under both normal and fault conditions, which could be achieved by the MPPT of the inverter. It is worth mentioning that the GBSSL can also successfully identify the fault even when the PV array deviates from its true MPP to a “sub-MPP.” For example, in some severe fault circumstances, the PV array’s MPP voltage ($V_{MPP}$) may become below the minimum MPP voltage of the PV inverter ($V_{INV-MIN}$). As a result, instead of $V_{MPP}$, the PV array is controlled to work at $V_{INV-MIN}$ as a sub-MPP by the inverter. In particular, the fault “LL 40% $R_f=0$” under $G_T=800W/m^2$ and $T_{amb} = 20^\circ C$ has the following parameters: $V_{MPP}=240$ V, $I_{MPP}=38.9$ A, $V_{OC-REF} = 40.28$ V and $I_{SC-REF} = 4.374$ A. If the PV array voltage is clamped by the inverters $V_{INV-MIN}$ at 250V, then PV array current
(I_{PV}) is found as 37.08 A (smaller than I_{MPP}). Therefore, V_{NORM} is redefined as V_{INV-MIN}/(N_{MOD} \cdot V_{OC-REF}) = 0.621. Similarly, I_{NORM} becomes I_{PV}/(N_{STR} \cdot I_{SC-REF}) = 0.848, which is smaller than the I_{NORM} at NORMAL. This operating point is annotated as point \( x_4 \) in Fig. 5.10. In the similar classification process as \( x_2 \), it is obvious to see \( x_4 \) will be classified as line-to-line fault, in “LL” class, successfully.

### 5.5.4 Simulation Results

To fairly test the effectiveness of the proposed GBSSL and avoid the effects from data occurring sequences, the simulation focuses on the worst-case scenario of fault detection and classification (FDC)—the GBSSL only uses initial labels for FDC and does not update the calculated labeled data (the Step 4 in Fig. 5.8 is neglected). Thus, any particular test data can be the first input data that the GBSSL learns in the PV system. The detection accuracy and classification accuracy are defined in (5.11) and (5.12). The simulation results of FDC for a variety of PV conditions are summarized in Table 5.1.

\[
\text{Detection Accuracy} = \frac{\text{correct \# of detection}}{\text{total \# of instances}} \tag{5.11}
\]
Chapter 5. Fault Classification Using Machine-Learning Algorithms

Figure 5.11: New fault data: “OPEN3” and “OPEN4” can be identified successfully, but “LL 10% $R_f=0$”, “LL 10% $R_f=10$”, “LL 20% $R_f=0$” and “LL 20% $R_f=10$” may be difficult to detect since they are overlapping with “NORMAL”.

Classification Accuracy = $\frac{\text{correct \# of classification}}{\text{total \# of instances}}$ \hspace{1cm} (5.12)

By using only 1.43\% ($=70/4879$) of the total data as initial labels, the proposed GBSSL model can detect and classify PV data under various classes, including
Table 5.1: Summary results of fault detection and classification

<table>
<thead>
<tr>
<th>Fault Category</th>
<th>Detailed PV Conditions</th>
<th>Graph-Based SSL Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Initial of Total Data</td>
</tr>
<tr>
<td>NORMAL</td>
<td>NORMAL</td>
<td>100%</td>
</tr>
<tr>
<td>LL</td>
<td>LL30% $R_f = 0$</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>LL30% $R_f = 10$</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>LL40% $R_f = 0$</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>LL40% $R_f = 10$</td>
<td>100%</td>
</tr>
<tr>
<td>OPEN</td>
<td>OPEN1</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>OPEN2</td>
<td>100%</td>
</tr>
<tr>
<td>LL (new)</td>
<td>LL10% $R_f = 0$</td>
<td>1.43%</td>
</tr>
<tr>
<td></td>
<td>LL10% $R_f = 10$</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>LL20% $R_f = 0$</td>
<td>78.19%</td>
</tr>
<tr>
<td></td>
<td>LL20% $R_f = 10$</td>
<td>37.3%</td>
</tr>
<tr>
<td>OPEN (new)</td>
<td>OPEN3</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>OPEN4</td>
<td>100%</td>
</tr>
</tbody>
</table>

NORMAL, LL and OPEN. Specifically, the GBSSL is able to identify LL faults successfully that are undetectable by conventional OCPD, such as “LL 30% $R_f=0$,” “LL 30% $R_f=10$,” “LL 40% $R_f=0$,” and “LL 40% $R_f=10$.”

In addition, Fig. 5.11 shows that the proposed model can work with new types of faults, even if they are not listed in the initial label set. For “OPEN3” and “OPEN4” PV faults, the proposed model can identify them successfully, since they are more similar to “OPEN” in the initial labels—similar $V_{NORM}$ but reduced $I_{NORM}$. However, for “LL 10% $R_f=0$,” “LL 10% $R_f=10$,” “LL 20% $R_f=0$” and
“LL 20% \( R_f=10 \)” of LL faults, the proposed model may misclassify them as NORMAL, because they are overlapping with NORMAL in the initial labels. The more overlapping between the LL fault and NORMAL, the more likely the LL is misclassified as NORMAL. Also, since “LL 10% \( R_f=0 \) or \( R_f=10 \)” and “LL 20% \( R_f=0 \) or \( R_f=10 \)” can only be misclassified as NORMAL, their detection accuracy and classification accuracy are the same (see Table 5.1).

### 5.5.5 Discussion

The success of FDC relies on the distinguishable data points in the 2-D normalized parameters (i.e., \( V_{NORM} \) vs. \( I_{NORM} \)). The simulation results show that sometimes certain aspects of a PV array’s size, fault type and fault resistance (\( R_f \)) can make the operating point difficult or even impossible to identify. Detection and classification errors may be caused by overlapping between NORMAL and LL. This brings difficulty to the GBSSL model to identify the new PV data near their border. For instance, under a wide range of solar irradiance and ambient temperature, the overlapping may occur between the LL with 20% mismatch or less (LL≤20%) and NORMAL in the plotting of \( V_{NORM} \) vs. \( I_{NORM} \) (see Fig. 5.11). For this reason, the LL≤20% may be misclassified as NORMAL by the GBSSL, leading to a protection challenge in the PV array, especially when the fault resistance (\( R_f \)) becomes significant, e.g. 10 Ω or more. One possible way to eliminate this “blind spot” in such a large PV array to increase the number of voltage or current
measurement in the PV array, such as using additional current sensors in sub-arrays/strings [70].

Another limitation in GBSSL classification is that it may become vulnerable to bad data, which could be caused by measurement noise, communication error or poor electrical connections. If the GBSSL uses a bad data, it may have incorrect FDC result (i.e., wrong class label) and cause false alarms. During the GBSSL’s self-learning process, the incorrect class labels may be propagated to the next data point. To solve this problem, digital filter, data pre-treatment and outlier detection rules have been proposed to eliminate the bad data [70, 73, 77, 87]. These types of technical challenges are typical of many machine learning methods, particularly for GBSSL.

5.6 FDC Experiments in PV Systems

5.6.1 Experimental Setup

A small-scale grid-connected PV system is built to create and record both normal and fault experiments under real-working conditions, such as varying irradiance and temperature (see Fig. 5.12 and Fig. 5.13). The PV array consists of 8 PV strings in parallel and each string has 2 modules in series.

The detailed parameters of main PV components are given in Table 5.2. Note that the reference modules have identical electrical parameters and environmental
conditions as the PV array. In the experiments, a moving average filter with a window size of 9 samples is used to reduce the measurement noise. Alternatively, the Kalman filter can be developed for measurement noise reduction [87].

Similar to previous simulation results, two types of faults have been created in the PV array: 1) A solid line-line fault between the middle of one string and negative conductor (LL 50% $R_f=0$), and the same LL with fault resistance $R_f = 20\Omega$ (LL 50% $R_f=20$); 2) An open-circuit fault on a string (OPEN1). During the experiment, the solar irradiance is ranging between $320\, W/m^2$ and $500\, W/m^2$, and the solar cell temperature is changing between $26^\circ C$ and $33^\circ C$.

It is necessary to mention that the fault current of all faults is sufficiently small that it cannot be detected by OCPD (e.g., fuses). As previously explained, though, it still remains a safety hazard [1]. If the inverter can detect the fault with the proposed algorithm, it might be able to send an alarm signal to users or shut

**Figure 5.12:** Schematic diagram of the experimental PV system.
down automatically. Therefore, by detecting such faults, the proposed method can significantly improve the overall reliability and safety of the PV system.

The flowchart of the GBSSL model for FDC has been given in Fig. 5.8. The first step in the flowchart is to obtain the PVs initial labels. We let the PV array
operate for a short period of time (e.g. less than one minute) in each condition. Meanwhile, the GBSSL model records the operating data as initial labels in the normalized parameters, such as $V_{\text{NORM}}$ and $I_{\text{NORM}}$ (see Fig. 5.14). Since the GBSSL requires only a few data sets (e.g., 10 points for each condition) as the initial labeled data, it saves time and avoids tedious training processes reported in the literature [40, 78]. Once the initial labels are found, the GBSSL is ready for real-time FDC. The PV experimental events follow the sequence of “NORMAL,” “OPEN1,” “NORMAL,” “LL 50% $R_f=20$,” “NORMAL,” “LL 50% $R_f=0$” and “NORMAL.” As new data are recorded, the GBSSL will categorize each new data point by using the proposed algorithm, update their classification labels, and send out fault alarms if necessary. In the experiments, there are approximately 1000 experimental data points associated with each condition, and the initial labels are only 1% of their dataset. The sampling frequency is arbitrarily chosen as 5 Hz in the experiment, but it can be increased to perform more responsively in FDC.

### 5.6.2 Experimental Results

As shown in Fig. 5.15, the experimental PV data under normal conditions or post-fault steady state are classified one after another by the GBSSL. As discussed in previous simulation results, the experimental instances under the same condition have distinct clusters in groups that do not overlap with another category. The variation of the operating points in the same cluster is caused by changing environmental conditions and steady-state oscillations due to the MPPT (e.g., perturb
Figure 5.14: Initial labels for GBSSL in experiments.

\& observe). Table 5.3 summarizes the experimental results and verifies the FDC effectiveness of the proposed GBSSL under real-working conditions. As have been noted in [1, 77], these LL faults often do not have large reverse current so that they cannot be identified or cleared with known commercial OCPD.

The “NORMAL,” “LL 50% \(R_f=0\),” “LL 50% \(R_f=20\)” and “OPEN1” in our particular case can be identified 100% correctly. This high accuracy is achieved thanks to the distinguishable clusters in the Fig. 5.15. Specifically, the PV array at “NORMAL,” “OPEN1” and “LL 50% \(R_f=20\)” is able to work at its own MPP. However at “LL 50% \(R_f=0\),” the \(V_{MPP}\) of the faulted PV array is significantly reduced below the inverters \(V_{MIN-IINV}\). Therefore, the \(V_{PV}\) will be clipped at \(V_{MIN-IINV}\).
as a sub-MPP (similar to the point $x_4$ in Fig. 5.10). In summary, the experimental results are consistent with the previous simulation findings:

- Compared with “NORMAL,” “OPEN1” has similar $V_{NORM}$ but $I_{NORM}$ is decreased proportionally to the number of lost strings.

- “LL 50% $R_f=20$” has reduced $V_{NORM}$ and slightly reduced $I_{NORM}$.

- Since the PV array at “LL 50% $R_f=0$” is working at sub-MPP, both $V_{NORM}$ and $I_{NORM}$ are reduced compared to “NORMAL” conditions.
### Table 5.3: Experimental Results of Fault Detection and Classification

<table>
<thead>
<tr>
<th>Fault Category</th>
<th>Detailed PV Conditions</th>
<th>Solar Irradiance</th>
<th>Graph-Based SSL Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>NORMAL</td>
<td>No fault ((V_{PV} = V_{MPP}))</td>
<td>320–500 W/m²²</td>
<td>100% (100%) labels of 1000 for each condition.</td>
</tr>
<tr>
<td>LL</td>
<td>LL50% (R_f = 0) ((V_{PV} \text{ clipped by } V_{INV-MIN}))</td>
<td>~335 W/m²²</td>
<td>100% (100%)</td>
</tr>
<tr>
<td>OPEN</td>
<td>OPEN1 ((V_{PV} = V_{MPP}))</td>
<td>~420 W/m²²</td>
<td>100% (100%)</td>
</tr>
</tbody>
</table>

**5.6.3 Discussion**

To apply GBSSL, initial labeled data was generated by measuring data from the intentionally created experimental PV faults. This was simple and safe for the small-scale benchmark system. However, for larger PV systems, the generation of real PV faults could associate with large fault current, potential safety issue and additional labor cost, which may hinder its practical application in PV field. To avoid these issues, it might be possible to create initial label data based on PV simulation models. This may bring safety, flexibility, portability, and cost-reduction to the proposed GBSSL. Without loss of generality, the simulated initial labels should follow the fault characteristics discovered by this dissertation.
5.7 Conclusion

Some solar PV faults have small fault current so they may not be cleared by over-current protection devices (OCPD). Thus, the faults can remain hidden in the PV system, resulting in possible dangers associated with it (sub-optimal performance, dc arcing, fire hazard, etc.). To identify these hidden faults, for the first time a graph-based semi-supervised learning (GBSSL) has been propose for fault detection and classification (FDC) in solar PV arrays. It increases the PV’s safety and reliability by detecting these faults (unnoticeable by OCPD). In addition, the proposed method is able to identify the specific fault type so that PV users can expedite the system restoration procedure.

To better visualize the PV data under normal and fault conditions, this chapter first develops new attributes using the normalized voltage and normalized current. Different from previous works, the proposed GBSSL model only requires a few points of the costly labeled data (∼1% of the total data set), while making use of inexpensive unlabeled data. The GBSSL algorithm is first analyzed and explained in detail through simulation results. In addition, the self-learning ability of GBSSL is proved as the label set can be updated over time during weather changes or PV arrays degrade. Furthermore, the proposed method does not depend on any particular PV inverter topologies and only uses readily available measurements in existing PV systems—such as PV-array voltage, array current, operating temperature and irradiance—requiring no additional hardware installations.
Chapter 6

Conclusions and Future Work

The research presented in this dissertation has reviewed the existing fault detection and protection solutions and their limitations, discovered the shortcomings in overcurrent protection devices (OCPD) in PV systems, and developed new fault detection and classification solutions to eliminate the fault protection gap. The main contributions of the research are presented below as well as the recommendations for future work.

6.1 Conclusions

The achieved results and major research contributions of this dissertation include:
• In Chapter 2, the existing solutions of fault detection, protection, classification and location are reviewed and their limitations are identified.

  – Literature review about fault detection, classification, and location solutions is presented in three categories, such as quantitative model-based solutions, process-history based solutions, and signal-processing based solutions. Their advantages and limitations are compared and discussed in detail.

  – Among these solutions, ground fault detection interrupters (GFDI) and overcurrent protection devices (OCPD) are widely used in the U.S. PV market, since they are required by U.S. PV standards. However, the protection gaps in GFDI and OCPD may exist under certain conditions, that can lead to reported fire hazards in the literature. For GFDI, it is shown that it may be vulnerable to double-ground fault in PV arrays, leading to previously reported fire hazards in the U.S. Specifically, when the first ground fault occurs between the negative conductor and the ground, GFDI may not be able to detect or clear the fault since the ground-fault current is relatively low. After that, if the second ground faults happens between the ungrounded current-carrying conductor (i.e., the positive conductor) and the ground, a high-magnitude ground-fault current will be created and it is likely to blow the GFDI (e.g., fuses). Unfortunately, the first and the second ground fault points
contribute to a closed fault path so that the high-magnitude fault current can flow continuously without interruption. Since the fault current can be much higher than the current rating of the conductors, the fault current may generate excessive heat in the conductor, leading to conductor insulation breakdown and fire hazards.

- The shortcomings of OCPD have been discussed in this dissertation. Specifically, according to the standards, OCPD is rated based on the maximum current rating of the PV modules. However, since solar PV arrays are current-limited power sources and their output characteristics are highly dependent on the solar irradiance level and fault locations, fault current may become below the OCPD rating. Therefore, faults may remain in the PV system without being cleared and become a problem to system efficiency, reliability and safety. Therefore, there is an urgent need of fault detection solutions to solve these problems.

- In Chapter 3, line-line fault inside PV arrays has been studied in simulation and experimental platforms.

- Since a PV array usually consists of a number of PV modules that are electrically connected, the PV array can be severely affected by ground faults or line-line faults among PV modules. If one PV module is under a fault, the whole system may be significantly impacted and could have reduced efficiency, degraded lifetime and safety hazards. This means
that the PV system is only as robust as its weakest link (e.g., the faulted PV components).

– Chapter 3 focuses on line-line faults, which are defined as an accidental short-circuiting between two points in the array with different potentials. Also, a line-line fault can be caused by a double-ground fault in PV arrays. A simulation platform has been built for this chapter that is capable of creating different fault scenarios inside PV arrays. It is shown that the fault current is greatly dependent on the interconnection of PV modules, fault location, solar irradiance level, and maximum power point tracking (MPPT) of PV inverters. The fault scenarios have been analyzed and explained using current vs. voltage (I-V) curve analysis, showing that under certain circumstances, the fault current may be so small that OCPD fails in fault protection. Therefore, the fault may be hidden in the PV array without being noticed, until the whole PV system fails. This could lead to previously reported fire hazards, property damage, and associated safety issues.

– The PV dynamic conductance is introduced in this chapter, which can explain the PV dynamic behavior under the small voltage perturbation by the MPPT. In addition, the transient behavior and MPPT effects on the PV array are discussed. It is shown that under the fault, if the MPPT is still working, MPPT tends to limit the faulted string current so that it becomes difficult for OCPD to clear the fault.
To prove the simulation results, line-line fault experiments are implemented in a commercial grid-connected PV system of 35kW rated output power. The experimental results verify the simulation results and show that line-line faults with small mismatch may not be detected or cleared successfully by conventional solutions (i.e., OCPD). Therefore, the line-line fault may remain uninterrupted in the PV system until the whole system fails. On the other hand, the experimental results also show that if the line-line fault involves large location mismatch, the OCPD has a better chance to clear the fault.

- In Chapter 4, outlier detection rules (ODR) are proposed in solar PV string monitoring systems that are able to detect the unnoticeable faults by OCPD.

- Compared with normal strings, the faulted string usually has a reduced output current. Therefore, PV string current and its rate of change can be used as features for fault detection. They are relatively convenient to measure and process. Based on these two features, two approaches—statistical ODR and local outlier factor (LOF)—are proposed separately to eliminate the protection gap of OCPD.

- The first method presented is called statistical outlier detection rule (ODR) which includes the $3\sigma$ rule, the Hampel identifier and the Box-plot outlier rules. Basically, these rules are quantitative methods to
define the boundaries of normal instances. For example, if a measurement lies outside of the boundaries, it will be identified as an outlier (e.g., fault). The second approach uses a spatial proximity-based local outlier factor (LOF) for fault detection. The main difference between the statistical ODR and LOF is that LOF is a scoring technique that assigns an outlier score to each instance or sample. Therefore, it gives users the freedom to choose the outlying threshold. The methods are applied to fault detection in PV arrays.

- The effectiveness of fault detection is verified on the experimental results in the 35kW PV laboratory under various solar irradiance conditions. The proposed statistical ODR’s (e.g., the Hampel identifier and the Boxplot outlier rules) are applied on the PV data of line-line faults with 1 or 2 modules at fault, which cannot be cleared by conventional OCPD. The experimental results have shown that both the Hampel identifier and the Boxplot outlier rule can identify the fault successfully after fault occurs. However, if the majority of the measurements are close to each other, the statistical ODR may become aggressive and could cause false alarms. To solve this problem, LOF is used for fault detection, which shows improved performance. Also, the comparison of statistical ODR and LOF and their advantages and disadvantages are presented. The proposed methods are more responsive than traditional methods, and it is suitable for real-time operation. Unlike the existing
solutions, the proposed methods do not require weather information, but only rely on the current measurement on each PV string.

- **In Chapter 5**, graph-based semi-supervised learning (GBSSL) method is proposed for fault classification in solar PV arrays.
  
  - In addition to fault detection, this dissertation proposes fault classification as another useful feature to indicate the fault type and further help expedite the system recovery from fault. The limitations of traditional methods, such as supervised learning methods, have been discussed. For example, the trained supervised learning model may fail when weather conditions of the PV array keep changing. Furthermore, labeled fault data in solar PV arrays might be difficult and costly to obtain.

  - Chapter 5 has shown that the $I$-$V$ output characteristics of a PV array can widely change as temperature and solar irradiance vary. As a result, it is possible that the faulted PV array has similar and even overlapping operating points as the normal PV array, making fault detection difficult. To solve this issue, two new parameters are first introduced, namely the normalized PV voltage and the normalized PV current. Taking weather information and PV array configurations into consideration, the new parameters tend to remain constant for each condition. That is PV data with same condition tend to cluster together. Thus, the proposed new parameters can improve PV data visualization and show better data clustering.
For the first time, GBSSL is proposed for fault detection and classification (FDC) in solar PV arrays. The proposed GBSSL is a semi-supervised learning algorithm, which exhibits self-learning ability and lower model training cost. It only uses a few labeled PV data and makes good use of a large amount of unlabeled PV data that can be easily obtained during PV system’s operation. By spreading the label information from labeled data to unlabeled data (e.g., new measurements), fault classification can be achieved. Also, the proposed method is utilizing readily available PV measurements in PV systems (i.e., MPPT voltage and current, and weather information) so that it saves the hardware upgrades and related labor costs. Besides, the proposed method can work with most any PV inverter topology that exists in the solar market.

The fault detection methods are able to detect open faults and dangerous line-line faults, which may not be detected by most known methods. The effectiveness of the proposed GBSSL model is validated in both a simulation platform and a small-scale PV experimental system under real working conditions.
6.2 Future Work

As shown in Fig. 6.1, conventional fault detection and protection methods use overcurrent protection devices (OCPD) and ground fault detection interrupters (GFDI) in solar PV arrays. However, this dissertation has shown that OCPD may fail in fault protection when the fault current is not high enough. This is caused by the irradiance-dependent and current-limit feature of PV arrays. Since OCPD (e.g., fuses or circuit breakers) are passive components that can only be blown or tripped at certain current/energy level, their limitations can be found in the solar PV arrays, leaving faults unnoticed or uncleared.

In this dissertation, we have proposed several decision-making algorithms to detect
and identify faults using different techniques, such as outlier detection rules or graph-based semi-supervised learning algorithms. Their performance has been validated in both simulation and experimental platforms, making them a promising choice for fault detection and classification.

To continue this research, future work may explore new active fault protection solutions, such as how to clear the fault actively, responsively and safely. Based on the tripping signal generated from the proposed methods, the active fault protection solution should increase the system efficiency, reliability, safety and fault immunity. Therefore, an integration of active fault protection approaches with the proposed methods would be a nice future research topic.

As the PV penetration level becomes more widespread, PV inverters take more and more responsibilities, not only in power conversions and maximum power point tracking, but also in fault detection and protection. As the most intelligent component in the PV system, PV inverters have the potential to provide more safety features. For example, recent PV inverters are featured with several different fault detection solutions, such as ground fault detection, dc-arc fault detection, insulation detection in ungrounded PV arrays, and residual current detection.

However, the existing fault detection solutions of PV inverters are usually based on signal-processing, merely relying on the instantaneous measurement. For this reason, fault classification is still not available for PV inverters. To make good use
of readily available historical data, future research may be focused on how to integrate the proposed fault classification methods into the PV inverters. This would provide better fault detection features and potentially increase the PV system reliability and safety.
Bibliography


[22] A. Abete, E. Barbisio, P. Cane, and P. Demartini, “Analysis of photovoltaic modules with protection diodes in presence of mismatching,” in *the Twenty
vol.2.

[23] D. Stellbogen, “Use of PV circuit simulation for fault detection in PV array
1307.

sis of a power output lowering of PV array with a (-dI/dV)-V characteristic,”
in Photovoltaic Energy Conversion, Conference Record of the 2006 IEEE 4th


and its use in fault detection,” Solar Energy, vol. 84, no. 4, pp. 624 – 635,
2010.

[27] A. Drews, A. de Keizer, H. Beyer, E. Lorenz, J. Betcke, W. van Sark, W. Hey-
denreich, E. Wiemken, S. Stettler, P. Toggweiler, S. Bofinger, M. Schneider,
G. Heilscher, and D. Heinemann, “Monitoring and remote failure detection
of grid-connected PV systems based on satellite observations,” Solar Energy,


List of Publications

The following list includes all the papers published by the author during his graduate studies. Papers designated by “⋆” are directly related with research results presented in this dissertation.

Journal Papers


Conference Papers


**Filed Patent**