ADVANCED CLASSIFICATION AND IDENTIFICATION OF PLUGGED-IN ELECTRIC LOADS

A Dissertation
Presented to
The Academic Faculty

by

LIANG DU

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
December 2013

ADVANCED CLASSIFICATION AND IDENTIFICATION OF PLUGGED-IN ELECTRIC LOADS

Approved by:

Dr. Ronald G. Harley, Advisor
School of ECE
Georgia Institute of Technology

Dr. Thomas G. Habeter, Co-Advisor
School of ECE
Georgia Institute of Technology

Dr. Miroslav M. Begovic
School of ECE
Georgia Institute of Technology

Dr. Jennifer E. Michaels
School of ECE
Georgia Institute of Technology

Dr. David G. Taylor
School of ECE
Georgia Institute of Technology

Dr. Lei Zhu
School of ME
Georgia Institute of Technology

Dr. Yi Yang
Global Research and Technology
Eaton Corporation

Date Approved: [October 31, 2013]

To

my mother, Guo-Lian Liu,

my father, Rong-Guang Du,

my lovely wife, Wei Liu, and

my beloved daughter, Lindsay Du,

for their love and support.

献给我的父母, 妻子, 和孩子

ACKNOWLEDGEMENTS

A doctoral dissertation is usually considered to be a personal accomplishment. However, it would not have been possible for me to finish this dissertation without the inspiration, encouragement, and support from many people.

First of all, I would like to express my most sincere thanks to my dissertation advisor Dr. Ronald G. Harley. He has been a wise and trusted advisor throughout the entire process. It is due to his constant inspiration and encouragement that I have gained a deeper understanding of engineering and made progress toward solving problems and improving my communication skills as a researcher. Had it not been for his vision, encouragement, and his confidence in my ability, much of this dissertation would not have been completed. I am deeply grateful for his guidance.

I would also like to express my gratitude to my co-advisor Dr. Thomas G. Habetler. His invaluable guidance and constant encouragement provide me with tremendous motivation. Special appreciation shall also be devoted to my committee members, Dr. Jennifer E. Michaels, Dr. Miroslav M. Begovic, Dr. Lei Zhu, and Dr. David G. Taylor.

I am indebted to Dr. Bin Lu and Dr. Yi Yang their support. Bin gave me this great opportunity to work on this research project, and this dissertation would not exist without his trust in the beginning. Yi is always there to discuss every technical detail with me and provide valuable comments. I had an opportunity to spend the summer of 2011 and work with Yi as an intern at Eaton Innovation Center (now Eaton Global Research and Technology). A large part of this dissertation was formulated during that period under her guidance.

I am also indebted to Dr. Jose Restpetro for his help and assistance to my experimental work. Jose is a great mentor to work with and learn from. I am fortunate to work with many exceptionally brilliant colleagues, include but not limit to, Hao Chen, Zhaoyu Wang, Yi Du, Dr. Pinjia Zhang, Dr. Jing Dai, Dr. Siwei Cheng, Dr. Jiaqi Liang, Andrew Paquette, Dr. Diogenes Molina, Dustin Howard, Jorge Hernandez, Dr. Wei Qiao, Dr. Zhi Gao, Nan Liu, Dawei He, Lijun He, Taizi Liu, Jie Dang, Zhenyu Tan, Bai Cui, Yi Deng, Dongbo Zhao, Qin Sun, Dr. Yao Duan, Zhenkai Wu, and Rui Fan for their friendship and support.

There are numerous names of faculty, family and friends that I should mention here, who have helped me during my six years and ten months at Georgia Tech. I want to express my gratitude to all of the people I know. I should thank Dr. Wassim Haddad to bring me to Georgia Tech. Life goes up and down, and I had my difficult time in my 2nd and 3rd years. Dr. Eric Feron supported me with patience. Thank you, Eric.

Most of all, I owe the greatest debt of gratitude to my family. My parents and parents-in law have always been the source of encouragement and support throughout my life. My parents always understand me and support me with no conditions. My dear wife, Wei Liu, has shared every single step in this long journey with me. Without their great love, encouragement, and understanding, everything would not have been possible.

The financial supports from the following institutions/organizations are gratefully acknowledged:

(1) U.S. Department of Energy

(2) Eaton Corporation

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS ..... **iv**

LIST OF FIGURES ..... **xiii**

LIST OF TABLES ..... **xvii**

LIST OF ABBREVIATIONS ..... **xix**

SUMMARY ..... **xxii**

Chapter 1 Introduction and Objectives **1**

1.1 Electricity Consumption of Plugged-In Electric Loads 1

1.2 Needs and Opportunities for Plugged-In Load Management 5

1.2.1 Energy saving by regulations and direct PEL control 5

1.2.2 Management of PELs in smart buildings 6

1.2.3 PEL management for demand response 7

1.2.4 Needs for smart power outlets 8

1.3 Non-Intrusive PELs Identification 10

1.4 Challenges of Plugged-In Electric Load Identification 12

1.4.1 Diversity within each PELs type and similarity between different PELs types 13

1.4.2 Utilizing long-term waveforms for PEL identification 16

1.5 Problem Statement 17
1.6 Hierarchical Identification Framework 18
1.7 Dissertation Outline 20
1.7.1 Feature extraction for PELs 20
1.7.2 Classification of PELs into categories 20
1.7.3 Identification of PELs in each category 21

Chapter 2 Literature Review 22

2.1 Introduction 22
2.2 General Framework for Electric Load Identification Systems 23
2.3 Review of Event Detection Methods 25
2.3.1 Event detection using steady-state values 25
2.3.2 Event detection using transient characteristics 25
2.4 Review of Features for Electric Loads 26
2.4.1 Active and reactive power 27
2.4.2 Peak, average, and RMS current values 27
2.4.3 Instantaneous values 27
2.4.4 Harmonics spectrum of the current waveform 28
2.4.5 Total harmonic distortion (THD) 28
2.4.6 Power factor 29
2.4.7 Crest factor or peak-to-average ratio 29
2.4.8 Transient features 30
2.4.9 Graphical features 31
2.4.10 Summary of features 35

2.5 Review of Identification Methods 35
2.5.1 P-Q plane 35
2.5.2 Decision tree 36
2.5.3 Optimization methods 37
2.5.4 Expert system 37
2.5.5 Artificial Neural Networks (ANNs) 38
2.5.6 Summary of identification methods 38
2.6 Review of Operating Mode Identification Methods 39
2.7 Summary of Chapter 41

Chapter 3 Plugged-In Electric Loads Classification by Features from V-I Trajectories 42

3.1 Introduction 42
3.2 Classification of Plugged-In Electric Loads by Front-End Power Supply Circuit Topology 43
3.3 Typical V-I Trajectories of Each PEL Category 45
3.4 Feature Extraction by Mapping V-I Trajectories to Cell Grids with Binary Values 48
3.4.1 Limitation in performance of existing graphical load signatures 48
3.4.2 Binary mapping from V-I trajectories to cell grids 50
3.4.3 Application of proposed mapping algorithm to collected data 56
3.4.4 Features extracted from the binary cell grid 57
3.5 Determining the Number of Self-Crossing Intersections 60
3.6 Computational Complexity Analysis for Proposed Signatures 62

3.7 Expected Values of Proposed Features for All PELs Categories 65

3.8 Chapter Summary 66

Chapter 4 Self-Organizing Classification and Identification of Plugged-in Electric Loads 67

4.1 Introduction 67

4.2 Self-Organizing Maps 67

4.2.1 Neurons and their assigned values 68

4.2.2 Sequential training algorithm 69

4.2.3 Batch training algorithm 70

4.2.4 Representations of the SOM 71

4.3 Supervised SOM (SSOM) for PELs Classification and Identification 73

4.3.1 Supervised SOM 73

4.3.2 SSOM for PELs classification and identification 75

4.3.3 Discussion on performance of the SSOM identifier 76

4.4 Tests on Performance of the SSOM on PELs Classification 77

4.4.1 PELs features used in tests 78

4.4.2 Test on the necessary amount of data to train SSOM 81

4.4.3 Performance of the proposed graphical features 82

4.5 Chapter Summary 85

Chapter 5 Probabilistic Identification of Plugged-in electric Loads 87

5.1 Introduction 87

5.2 Overview of Bayesian Decision Theory and Probability Estimation 87

5.3 Probabilistic Identification of an Unknown PEL Represented by a Single Feature Vector 89
5.4 Hybrid SSOM / Bayesian Classifier 92
5.5 Estimation of Conditional Probabilities with an Unknown PEL Represented by a Set of Feature Vectors 93
5.6 Direct Estimation 94
5.7 Tests on the Hybrid SSOM/Bayesian Identifier 95
5.8 Chapter Summary 102

Chapter 6 Multi-Class Identification of Plugged-In Electric Loads by Support Vector Machine 104

6.1 Introduction 104
6.2 In-Category PEL Identification by SVM 106
6.2.1 Introduction to Support Vector Machine (SVM) 106
6.2.2 Multi-class identification by SVM 109
6.3 In-Category PEL Identification by One-Against-All SVM 110
6.4 Comparison of Time-Domain and Frequency-Domain Features for SVM 112
6.5 Testing the Performance of the Hybrid SSOM/SVM Classifier on a Large Number of PELs 116
6.5.1 Comparison of performance of the SSOM identifier and the hybrid SSOM/SVM identifier on in-category PEL identification 116
6.5.2 Testing the performance of the hybrid SSOM/ SVM classifier on a large number of PELs 119
6.6 Discussion on the Performance of the Hybrid SSOM/SVM identifier 120

6.7 Chapter Summary 120

Chapter 7 Identification of Plugged-In Electric Loads by Long-Term Waveforms 122

7.1 Introduction 122

7.2 Representing Long-Term Waveform by Finite-State Machine 123

7.2.1 Finite-state machines 123

7.2.2 Representation of long-term current waveforms 125

7.2.3 Definition of states 126

7.3 Reduction of States and Definition of Elemental States and Events 129

7.4 Classification and Identification of Repeating Actions and Modes 134

7.4.1 Repeating patterns in FSM representations 135

7.4.2 Almost-identical repeating patterns 137

7.4.3 Step up/down repeating patterns 138

7.4.4 Spike-lead repeating patterns 140

7.4.5 Summary of repeating patterns 142

7.5 Features Extracted from the FSM Representation 143

7.6 Chapter Summary 145

Chapter 8 Conclusions and Contributions 147

8.1 Introduction 147

8.2 Conclusions of This Dissertation 147

8.3 Contribution of This Dissertation 149

8.4 Outcomes of This Dissertation 150

8.4.1 U.S. Patents 150

8.4.2 Journal Paper 151
8.4.3 Conference Paper 152
8.4.4 Contribution of chapters 153
8.4.5 Other Contribution 153
8.5 Future Work 154
8.5.1 Implementation and field tests 154
8.5.2 PEL disaggregation 154

Appendix A Data Acquisition System and Lab Environment 160
A.1 PCB Sensors Module 161
A.2 BNC Connector and Cable 164
A.3 NI DAQ Device 165
A.4 LabView Program 167
A.5 Calibration 168
A.6 Harmonics in the Supply Voltage 168

Appendix B Mapping of Representative V-I Trajectories to Binary Cell Grids 171

Appendix C SOM Toolbox for MATLAB 180

Appendix D SVM Toolbox for MATLAB 184

BIBLIOGRAPHY 186

VITA 198

LIST OF FIGURES

Page

Figure 1.1 U.S. electricity retail sales by sectors from 1949 to 2009 [1]. 1

Figure 1.2 Classification of electric loads in buildings 2

Figure 1.3 Deploying smart outlets and smart power strips in buildings [39] 9

Figure 1.4 Commercially available smart outlets and smart power strips 9

Figure 1.5 General framework of non-intrusive PELs identification 12

Figure 1.6 Current waveforms to illustrate the diversity in types and similarity between types of PELs 14

Figure 1.7 Illustration of the *Navetas*™ energy management system 15

Figure 1.8 The *enPowerMe*™ load monitoring system 15

Figure 1.9 Long-term and short-term current waveforms of an LCD TV and a laptop computer 16

Figure 1.10 Hierarchical PEL identification framework 19

Figure 2.1 General framework for electric load identification systems [42-44]. 23

Figure 2.2 Transient features in active power of (a) a lamp bank and (b) an induction motor [65]. 30

Figure 2.3 V-I trajectories of (a) a desktop computer and (b) a refrigerator [44]. 32

Figure 2.4 Illustration of V-I trajectory segments and measurements [68]. 33

Figure 2.5 Using graphical features for load disaggregation [69] 34

Figure 2.6 Relative positions of a group of appliances in the complex P-Q plane [49]. 36

Figure 2.7 Current waveforms of office appliances in different operating modes 40

Figure 3.1 Typical normalized V-I trajectories of the seven load categories.....47
Figure 3.2 Normalized V-I trajectories of four plugged-in electric loads to illustrate the limitation in performance of existing graphical load signatures 50
Figure 3.3 Mapping V-I trajectory to a binary cell grid 51
Figure 3.4 Eight neighbors of a cell in the grid 54
Figure 3.5 Illustration of mapping a V-I trajectory to a binary cell grid.....55
Figure 3.6 Two key cells (C1 and C2) and three key lines (L1-L3) in a cell grid 57
Figure 3.7 Example of a continuum of occupied cells in the grid.....59
Figure 3.8 Illustration of graphical features for PELs from Category PAC 60
Figure 3.9 Determining the number of self-crossing intersections 61
Figure 3.10 Checking adjacent unoccupied cells in one direction 64
Figure 4.1 The U-matrix visualizes distances between neighboring neurons, and helps to show the cluster structure. 72
Figure 4.2 Labeled neuron grid of the SOM representing ten PELs. 74
Figure 4.3 The cross-validation framework of the SSOM identifier 76
Figure 4.4 The comparison of $MSE_{I_{RMS}}$ of a DVD player and a STB. 80
Figure 5.1 Current waveforms of test PELs 97
Figure 6.1 Current waveform of a copier in operation for 60 seconds 105
Figure 6.2 The SVM framework utilizing kernel functions. 108
Figure 6.3 The hybrid SSOM/SVM identifier framework 111
Figure 6.4 Current and voltage profiles of three PELs with quite different characteristics: Fan (top left), TV (top right), and DVD (bottom) 113

Figure 6.5 Current and voltage profiles of LCD TV (top left), LED TV (top right), and plasma TV (bottom) 117
Figure 7.1 An example of a FSM illustrating its elements 124
Figure 7.2 Actual and quantized operating current waveforms of a plasma TV for 60 seconds 128
Figure 7.3 Extracted states with associated current and time values from the operating waveform in Figure 7.2 129
Figure 7.4 Peak-to-peak operating current waveform of a plasma TV for 60 seconds 131
Figure 7.5 Corresponding states with associated values extracted from Figure 7.4 132
Figure 7.6 Corresponding spikes, semi-steady states, steady states, and oscillation states extracted from Figure 7.4 132
Figure 7.7 Corresponding spike events, semi-steady states, steady states, and oscillation states extracted from Figure 7.6 133
Figure 7.8 Current waveforms of four office appliances in different operating modes 135
Figure 7.9 Four PELs in active mode as examples of non-existence of repeating patterns 143
Figure 8.1 Measured aggregated current waveform 157
Figure A.1 Two views of the portable PCB sensors module 161
Figure A.2 PCB circuit schematic 162
Figure A.3 PCB design schematic 163

Figure A.4 BNC connectors and cables 165
Figure A.5 NI SCXI-1000 DAQ system 165
Figure A.6 NI USB-6008 DAQ system 166
Figure A.7 Block diagram of the LabView data acquisition program. 167
Figure A.8 Utility voltage waveform in the Lab 168
Figure A.9 Total harmonic distortion in voltage 169
Figure A.10 Harmonic spectrum of the distorted ac utility-supplied voltage waveform 170
Figure C.1 Table format data for MATLAB SOM toolbox.....180

LIST OF TABLES

Page

Table 1.1 List of PELs considered in this dissertation....4

Table 3.1 Comparison of the number of real multiplications and real additions needed by different algorithms 63

Table 3.2 Expected signatures of all 7 PELs categories 65

Table 4.1 Test Success Rates of the SSOM Identifier 81

Table 4.2 Tests on the performance of the proposed graphical features....82

Table 4.3 Comparison of proposed graphical features and conventional features....84

Table 5.1 Number of PELs from each group 95

Table 5.2 Test results for 8 PELs in 3 Scenarios 98

Table 5.3 How probabilities are estimated by the hybrid identifier....101

Table 6.1 Comparison of testing success rate of different feature sets using multi-class one-against-all SVMs 114

Table 6.2 Testing success rate of the multi-class one-against-all SVM and SSOM identifiers with different amount of data....115

Table 6.3 Testing success rate of the hybrid SSOM/SVM identifier and the SSOM identifier for in-category PELs 118

Table 6.4 Testing success rate of typical PELs using the multi-class one-against-all SVM identifier....119

Table 7.1 Associated I_k and T_k values of states in the FSM of operating waveform in Figure 7.8(a) 136

Table 7.2 Associated I_k and T_k values of states in the FSM of operating waveform in Figure 7.8(d) 138

Table 7.3 Associated I_k and T_k values of states in the FSM of operating waveform in Figure 7.8(b) 139

Table 7.4 Associated I_k and T_k values of states in the FSM of operating waveform in Figure 7.8(c) 141

Table 7.5 Summary of Features of a selected set of important office electric loads 144

Table 8.1 Summary of testing results in this dissertation 149

Table 8.2 Contribution of each chapter of this dissertation 153

Table 8.3 Timed sequence of semi-steady and steady states represent Figure 8.1.....158

Table 8.4 Harmonics in utility single phase voltage in the lab 170

Table B.1 42 representative V-I trajectories mapped to binary cell grids 171

LIST OF ABBREVIATIONS

ANN	Artificial neural network
BMS	Building management system
BMU	Best matching units
BNC	Bayonet Neill–Concelman
CF	Crest factor
CFL	Compact fluorescent light
CVD	Continuously variable device
DAQ	Data acquisition
DFT	Discrete Fourier transformation
DIF	(Per-cycle) difference
DOE	Department of Energy
EMI	Electromagnetic interference
FFT	Fast Fourier transformation
GMM	Gaussian mixture model
HVAC	Heating, ventilation, and air conditioning
IEC	International Electrotechnical Commission
kNN	K-nearest neighbor
LCD	Liquid-crystal display
LED	Light emitting diode
MAP	Maximum *a priori* probability

MEL	Miscellaneous electric load
MFD	Multi-function device
MLE	Maximum likelihood estimation
MLP	Multi-layer perceptron
MRI	Magnetic resonence imaging
MSE	Mean square error
NI	National Instrument Inc.
NILM	Non-intrusive load monitoring
PAC	Phase angle controller
PCA	Principle component analysis
PCB	Printed circuit board
PEL	Plugged-in electric load
PF	Power factor
PF_{disp}	Displacement power factor
PF_{dist}	Distortion power factor
PFC	Power factor correction
P-Q plane	Active power (P) versus reactive power (Q) plane
RBF	Radial-basis function
RMS	Root mean square
SMPS	Switching mode power supply
SOM	Self-organizing map
SSOM	Supervised self-organizing map
STB	Set-top box

SVM	Support vector machine
THD	Total harmonic distortion
U-matrix	Unified distance matrix
V-I trajectory	Voltage versus current trajectory

SUMMARY

The total electricity consumption of plugged-in electric loads (PELs) currently accounts for more usage than any other single end-use service in residential and commercial buildings. Compared with other categories of electric loads, PELs possess significant potential to be efficiently controlled and managed in buildings. Therefore, accurate and reliable PEL identification methods that are used to collect identity and performance information are desired for many purposes. However, few existing electric load identification methods are designed for PELs to handle unique challenges such as the diversity within each type of PEL and similarity between different types of PELs equipped by similar front-end power supply units.

The objective of this dissertation is to develop non-intrusive, accurate, robust, and applicable PEL identification algorithms utilizing voltage and current measurements. Based on the literature review of almost all existing features that describe electric loads and five types of existing methods for electric load identification, a two-level framework for PELs classification and identification is proposed.

First, the supervised self-organizing map (SSOM) is adopted to classify a large number of PELs of different models and brands into several groups by their inherent similarities. Therefore, PELs with similar front-end power supply units or characteristics fall into the same group. The partitioned groups are verified by their power supply unit topology. That is, different groups should have different topologies. This dissertation proposes a novel combination of the SSOM framework and the Bayesian framework. Such a hybrid identifier can provide the probability of an unknown PEL belonging to a specific type of load.

Within each classified group by the SSOM, both *static* and *dynamic* methods are proposed to distinguish PELs with similar characteristics. Static methods extract steady-state features from the voltage and current waveforms to train different computational intelligence algorithms such as the SSOM itself and the support vector machine (SVM). An unknown PEL is then presented to the trained algorithm for identification. In contrast to static methods, dynamic methods take into consideration the dynamics of long-term (minutes instead of milliseconds) waveforms of PELs and extract elements such as spikes, oscillations, steady-state operations, as well as similarly repeated patterns.

CHAPTER 1 INTRODUCTION AND OBJECTIVES

1.1 ELECTRICITY CONSUMPTION OF PLUGGED-IN ELECTRIC LOADS

In the United States, electric loads in residential and commercial buildings accounted for around 75% of the total electricity consumption in 2012 [1]. Moreover, the total consumption by residential and commercial buildings has been increasing for the past six decades [1], as shown in the following figure. The economic, operational, and environmental impacts of increasing electric power consumption have drawn world-wide attention to the need for better energy consumption management and direct control of electric loads in not only residential houses but more importantly also commercial buildings such as hospitals, schools, and data centers.

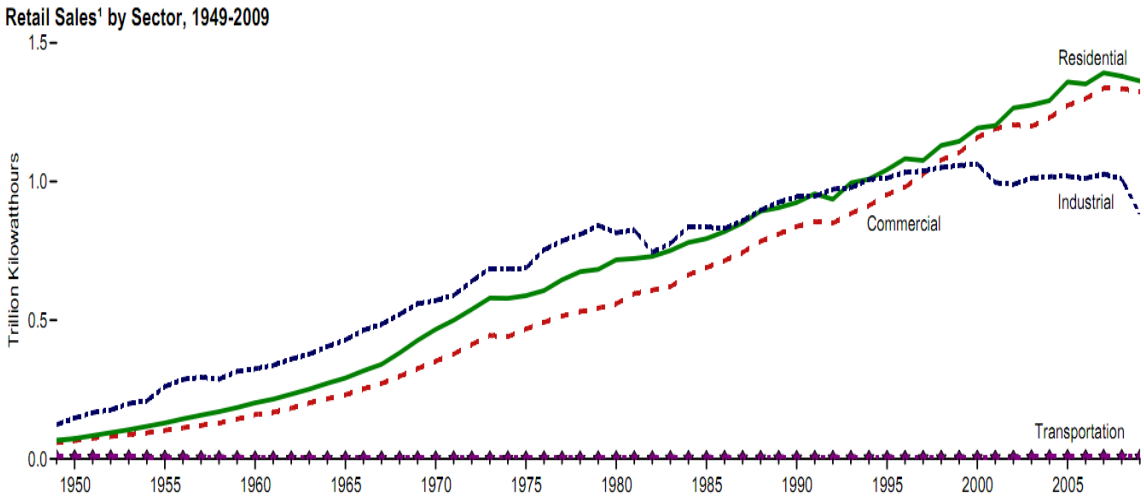


Figure 1.1 U.S. electricity retail sales by sectors from 1949 to 2009 [1].

Electric loads in residential and commercial buildings are commonly divided into groups such as space conditioning, water heating, ventilation, lighting, major appliances, and miscellaneous [2]. Miscellaneous electrical loads (MELs) are the diverse collection of electricity-consuming devices including portable loads which are electronic appliances plugged into sockets, along with all hard-wired loads that do not fit into other major end-use categories [3]. The suggested partition of all electric loads by [2], as well as some examples, is shown in the following figure.

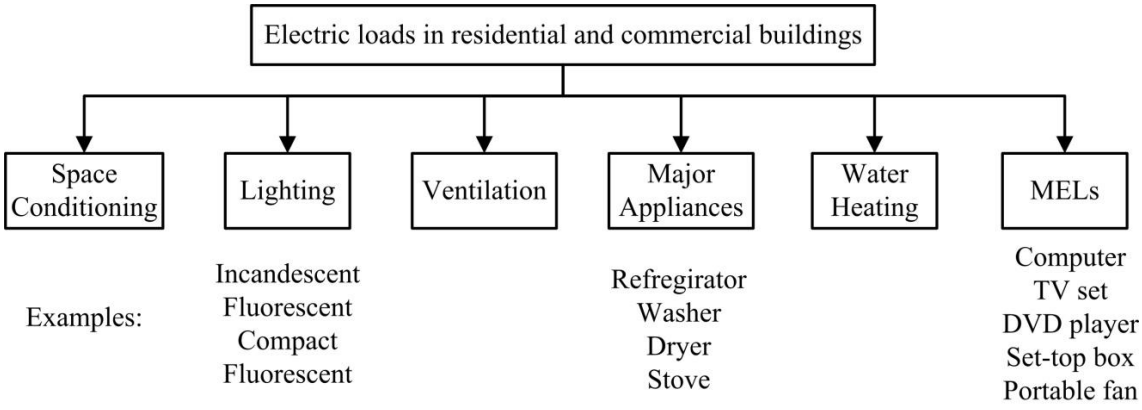


Figure 1.2 Classification of electric loads in buildings

It is reported that MELs currently consume more electricity than any other single end-use service in residential and commercial buildings [3]. Furthermore, a recent report from the United States Department of Energy (DOE) [4] indicates that "miscellaneous uses dominate growth in electricity demand" in residential buildings. For example, the electricity consumption of TV sets and set-top boxes surpassed that of refrigerators in 2010. It is also predicted in [4] that MELs' consumption (e.g., video displays and medical

devices) will increase by an average of 2.3 percent per year and, in 2035, will account for about 40 percent of total electricity consumption in the commercial sector. The rapid growth in both residential and commercial buildings is commonly considered to be driven by consumer electronics. It is predicted that the growth will continue and even accelerate due to network connections of MELs in the future [5].

Portable MELs, which account for the majority of all MELs, are of special interest in this dissertation for the following several reasons.

(1) Non-portable MELs, such as distribution transformers, non-road electric vechiles including electric trams, electric locomotives, and wheeled vehicles that are not intended for use on public roads (such as airport ground support equipment), magnetic resonence imaging (MRI), and elevators [5] are less frequently installed, not as easily accessible and controllable compared with portable MELs. Note that here "controllable" means real-time direct load control accordingly to different needs and scenarios.

(2) A large number of electric loads in other categories are also portable, such as refrigerators, washers and driers, air conditioners, and lighting appliances. These portable loads can be controlled in the same manner as portable MELs.

(3) A large number of portable electric loads are *vampire* loads [6, 7]. In other words, they are defined by DOE as "electronic devices which still consume electricity while in standby mode or being switched off" [7]. Such vampire energy should be efficiently managed to reduce the amount of wasted energy.

This dissertation focuses on portable MELs and other portable major appliances, which will be referred to as *plugged-in electric loads* (PELs) within this dissertation.

Specifically, the PELs considered in this dissertation include, but are not limited to, the appliances listed in Table 1.1.

Table 1.1 List of PELs considered in this dissertation

	Residential	Commercial
Home entertainment	TV: LED, LCD, plasma, and CRT	
	TV accessories: set-top box (STB), DVD player, video cassette recorder (VCR), and audio devices	
	Video game consoles: PlayStation, Xbox, Wii, etc.	
Home appliances	Washer and dryer, Portable Spa	
Public appliances	Lighting: dimmer, incandescent, fluorescent, and compact fluorescent lamps Space conditioning: portable fan, space heater, humidifier, dehumidifier, and portable air conditioner	
		Vending machine, Water dispenser
Network	Modem, Router	Server
Kitchen appliances	Cooker, Stove Dish washer	
	Microwave oven Coffee brewer Portable refrigerator Toaster Hot water kettles	
Computer	Desktop, laptop, and (external) monitor	
Office appliances	Projector Fax machine Copy machine Multi-function device (MFD) Shredder Cordless phone and answering machine	
Other	Charger: any with battery External hard drive Home security system Clock radio/small stereo Portable electric space decoration device	

1.2 NEEDS AND OPPORTUNITIES FOR PLUGGED-IN LOAD MANAGEMENT

The large portion of the total electricity consumption by PELs offers opportunities to manage PELs usage and consumption, reduce energy wasted by vampire loads, and regulate PELs operation for a sustainable future. Compared with other major high power electric loads such as water heating and space conditioning appliances, PELs possess great and unique potentials to be efficiently managed in buildings as they can be directly controlled (e.g., turned ON/OFF) by the switches in power strips, main sockets, and power outlets in which PELs are plugged into. Furthermore, the controllability of PELs results in a large number of ongoing work for many purposes including energy saving, building management, and demand response.

1.2.1 Energy saving by regulations and direct PEL control

Energy Star indicates that in United States on average it costs each household $100 per year for PELs while they are off or in standby mode. On a national basis, standby PELs consumes more than 100 billion kilowatt hours annually and contributes to more than $10 billion in annual energy costs. Proper PELs consumption management can result in as much as 75% standby power savings [8] and 40 million tons of carbon emission reduction expected per year in United States [9].

Current work on reducing the amount of energy consumed by vampire loads can be summarized as follows:

(1) Introduction of regulations to reduce the energy consumption by PELs in standby or OFF mode. For example, *Energy Star* standard 5.1 requires that qualified TV sets must consume no more than one watt while in sleep mode

[10], which has been introduced as a regulation by the California Energy Commission in 2011 [11].

(2) Direct control (e.g., turned ON/OFF) of PELs when they are in standby or OFF mode. For example, a recent effort by DOE, *Building America*, has started to identify and reduce PELs consumption [12] and aims at 50% energy savings in new homes by 2015.

1.2.2 Management of PELs in smart buildings

For the purpose of a sustainable future, DOE has announced its goal of achieving market ready net-zero energy residential and commercial buildings by 2020 and 2025 [13]. This requires a centralized management of electric loads, renewable energy sources, and possibly energy storages. The zero net energy consumption of these buildings are achieved by harvesting energy from renewable energy sources such as solar panels and wind generators, utilizing high-efficiency electric loads, and reducing the amount of wasted energy through proper load and building management.

Recently, a new building management scheme called “appliance commitment” has been proposed in [14], which aims at scheduling thermostatically controlled household electric loads based on price and consumption forecasts to meet specified optimization objectives such as maximum users' comfort level. Similar electric load management schemes are investigated in [15] via binary on-off policies of the smart flexible devices with user’s comfort considered. Furthermore, with the expanding deployment of plugged-in medical equipment and electric vehicles [3, 16], certain types of PELs are expected to be managed with specific requirements.

Besides energy saving to achieve net zero energy consumption, PEL management can also enhance the capability of building management systems (BMS), introduce more intelligence into PEL operations, and improve building occupant experience. For instance, the protection device on an uninterruptible power supply (UPS) or a power strip cuts off all connection when over-current happens but such an unexpected power cutoff will cause a plugged-in desktop computer to loss all its current work. In this case of over-current event, proper load management should disconnect/turn off noncritical loads and keep the desktop computer on. This example can be extended to a more general application to keep a selected set of critical loads (such as network servers and computers) on under all circumstances. These two examples illustrate that incorporation of more intelligence into load management can help to improve building occupant experience and enhance BMS capabilities.

1.2.3 PEL management for demand response

For the purpose of demand response, many efforts have been devoted by others to the demand-side management (DSM) of electric loads in residential and commercial buildings [17-25]. DSM of electric loads typically aims at improving system reliability, dynamic pricing [26], reducing energy consumption [27, 28], and introducing advanced real-time control [29-31], and load balancing [32, 33]. With the fast deployment of plugged-in electric vehicles (PEVs), new demand response schemes [34] with large numbers of PEVs at homes as shiftable electric loads as well as energy storages [35] are still under investigation.

Typical demand response in buildings to reduce energy consumption during peak energy-consumption hours is achieved by a centralized building automation system with

time scheduling. A number of such building automation systems have been designed and are available, such as Siemens "*Demand Response Solutions for Commercial Buildings*" [36], Lawrence Berkeley National Laboratory's automated demand response system [37], Pacific Northwest National Laboratory's facility energy decision system (FEDS) [38].

A major problem within these automated building demand response system is that they highly rely on time signal from utilities to start and end demand responses. In other words, these systems do not perform load management and energy saving during normal hours. Therefore, besides centralized building-level building management system for demand response during certain peak hours, distributed outlet-level load management systems are also desired for building occupants to meet different occupants' different individual needs.

1.2.4 Needs for smart power outlets

To summarize previous discussions, due to their special characteristics in universality, flexibility, and controllability, PELs possess unique potentials not only in energy saving but also in many other purposes such as intelligent building energy management, granular consumption information collection for building efficiency certification, and demand response for reliable and economical operation. Furthermore, a centralized building management system cannot meet the needs of PEL control and management in many cases. Therefore, smart power outlets (or smart power strips) are desired by many applications to collect usage information and perform control actions on individual PEL. The general framework of deploying smart outlets and smart power strips in a distributed manner is shown in the following figure.

Figure 1.3 Deploying smart outlets and smart power strips in buildings [39]

Figure 1.3 shows that smart outlets and smart power strip can collect information of PELs plugged into them, communicate with either local-level or building-level management system, and perform control actions. Several examples of commercially available smart outlets and smart power strips are shown in the following figure.

Figure 1.4 Commercially available smart outlets and smart power strips

To summarize, most smart strips typically have controllable sockets and uncontrollable sockets. Users can plug the loads that they would like to manually control into controllable sockets and turn ON/OFF PELs through wireless communication between the remote and the power strip. Moreover, loads plugged-into the uncontrollable sockets stay connected to the utility network all the time as there are no switches in these sockets to control PELs.

More intelligence is desired to be incorporated into current smart power outlets and smart power strips because all control actions need to be performed manually by users. Instead of manually control PELs every day, users may need to have programmable smart outlets such that they are define certain rules for the smart outlet to carry out in an automatic manner. In order to achieve automatic PEL management through smart power strips, it is necessary for the smart power strips to have the capability of knowing what is the identity (model, type, and operating status) of the plugged-in load, follow the pre-defined management rules, and perform necessary actions to corresponding PELs. In other words, without knowing the PEL identity without ambiguity, smart power strips may perform undesired actions to PELs.

1.3 NON-INTRUSIVE PELS IDENTIFICATION

As discussed above, in order to achieve the various PELs management prospects discussed above, the information of PELs identity, consumption, and performance is required. Specifically, PELs identity information (i.e., the type or model of each PEL) is the most important part because the consumption and performance information should be credited to specific PELs and the building management system should know without

ambiguity which PELs are under control. Therefore, reliable and accurate PELs identification methods are the foundation of all PELs management prospects.

However, the majority of electric loads in residential and commercial buildings still remain unidentified due to the lack of embedded identity labels inside electric loads as well as communication between electric loads and a building management system. This is particularly true for PELs due to their low costs, gigantic total-number, and dynamic portability. The relatively low costs of PELs make it not economic to embed internal identity signal generator with communication capability.

There are two kinds of load identification approaches, intrusive and non-intrusive. A physically intrusive approach is proposed in [40] where sensors are installed on every electric load to monitor status of the loads, and signals are sent to data processor through a power line. However, the intrusive approach needs the cooperation of manufacturers and users. Furthermore, the communication of signals and information is also demanding.

As a result, it is more realistic to design a PELs identification algorithm in a non-intrusive manner. In other words, installation of extra, interior, or intrusive wiring or sensors into any PELs or existing plugged-in sockets in buildings is not required.

The only available information for non-intrusive PELs identification includes voltage and current waveforms collected from sockets or outlets. PELs often present unique characteristics in these electric signals, which are discussed in more details in later chapters. Such load characteristics provide a viable means to identify the type of a PEL (e.g., computer, TV, or lamp, etc.) and even possibly its operation status (e.g., startup, normal, standby, etc.) by analyzing these electric signals.

The general framework of the non-intrusive PELs identification problem is illustrated in the following figure.

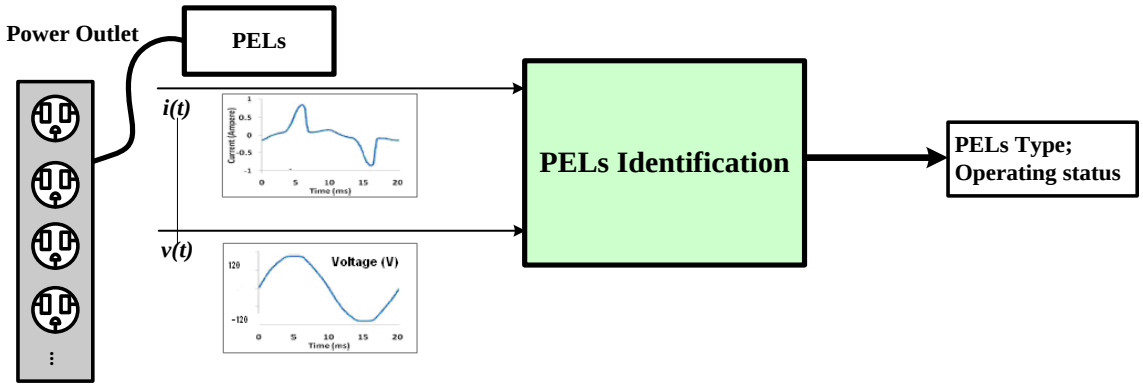


Figure 1.5 General framework of non-intrusive PELs identification

Note that the PELs identification problem is non-intrusive because the voltage and current waveforms are measures externally without intrusive wiring or sensors into the PEL. Also, the only source of information for the non-intrusive PELs identification problem is contained in the voltage and current waveforms.

1.4 CHALLENGES OF PLUGGED-IN ELECTRIC LOAD IDENTIFICATION

Starting with the original idea of non-intrusive load monitoring (NILM) by Hart in the late 1980s [41], many methods have been proposed to monitor and identify electric loads over the past twenty years. A comprehensive review of existing work is provided in Chapter Two of this dissertation. However, few methods are designed specifically for PELs and have addressed the following challenges.

1.4.1 Diversity within each PELs type and similarity between different PELs types

The fast development of front-end power supply units and wide deployment of personal electronic devices such as tablet computers and smart phones bring challenges to PELs identification. Some of the most challenging problems are listed as follows.

(1) Different types of PEL are equipped with similar front-end power supply units and thus have similar characteristics;

(2) Each type of PEL could be equipped with different front-end power supply units as more efforts have been devote to regulate PEL power consumption. Therefore, PELs of the same type may have quite different characteristics;

(3) A PEL may show quite different characteristics in different operating modes. For example, current waveforms of a PEL with a power factor correct (PFC) unit are quite different when the PFC unit is turned on or off;

(4) Intelligent PEL identification methods should have the capability of receiving inputs or feedback from users or building managers to improve their robustness. No identification method can guarantee 100% success rate or no error under all scenarios. However, inputs or feedback can help when identification algorithms cannot tell apart certain PELs without ambiguity.

(5) Several PELs are typically connected into one power outlet. In this case a single current waveform would consist of mixed signals of multiple PELs.

Several plots of real-world current waveforms are shown in the following figure to illustrate the above challenges. Three cases are considered including

(1) Characteristics of a PEL can be different: an LED TV in active mode (a) and in energy saving mode (b);

(2) Characteristics within a type of PELs can be different: LED TVs of two different manufactures: (a, b) and (c);

(3) Characteristics of different types of PELs can be similar: an LED TV (c) and a set-top box (d).

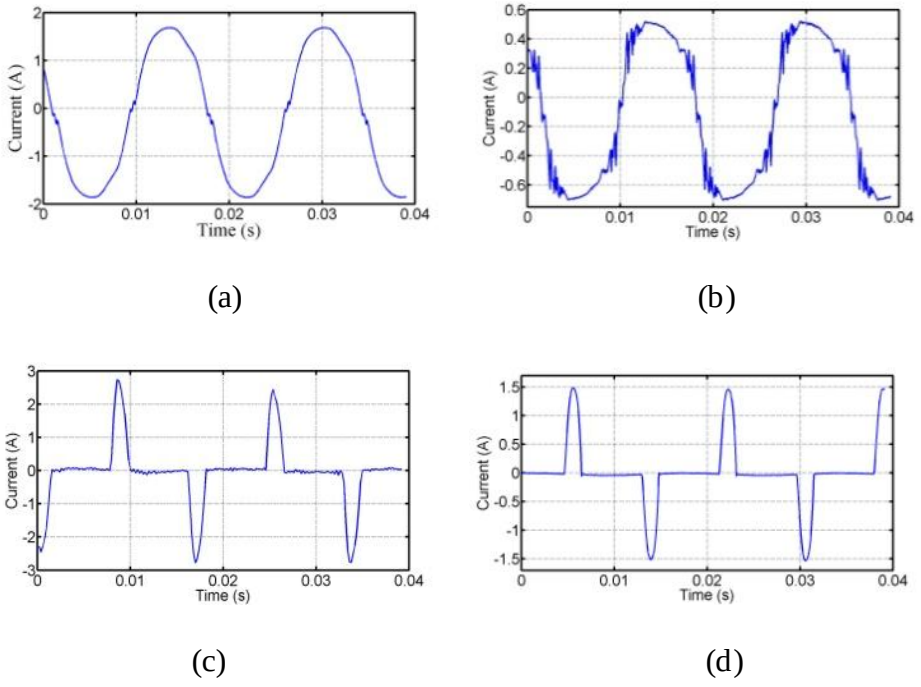


Figure 1.6 Current waveforms to illustrate the diversity in types and similarity between types of PELs

The above example shows that the diversity within *each* type of PELs and the similarity between *different* types of PELs significantly complicate the identification. Few existing methods have addressed these challenges. As a result, available commercial load identification and monitoring products have limited capabilities to consider only several PELs with quite different power ratings.

For example, the home energy management system developed by *Navetas*™ only considers coffeemakers, TVs, refrigerators, lamps, and vacuum sweepers, as shown in the following figure [42].

Figure 1.7 Illustration of the *Navetas*™ energy management system

Also, the home energy management system developed by *enPowerMe*™ cannot identify low-power (less than 100 W) loads, as illustrated in the following figure.

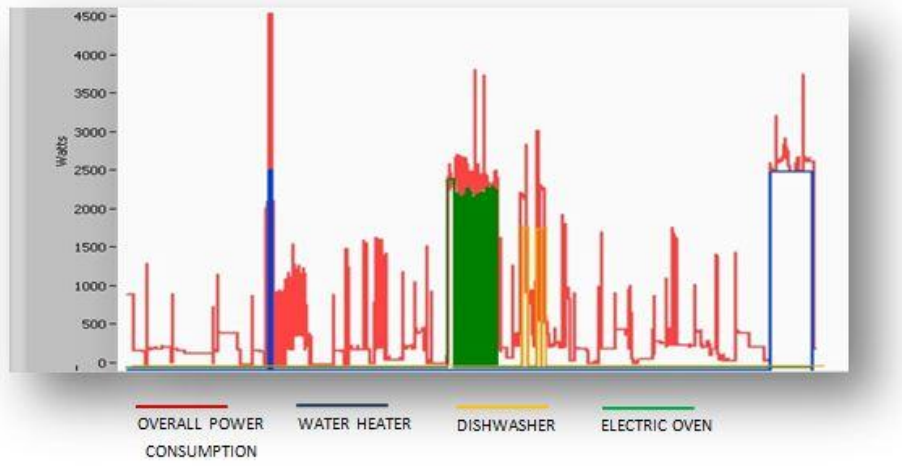
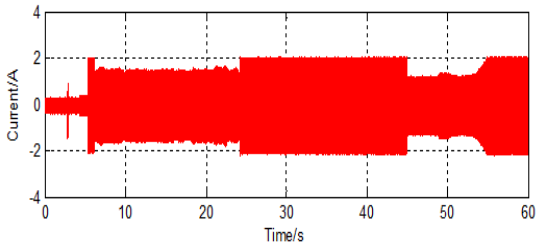


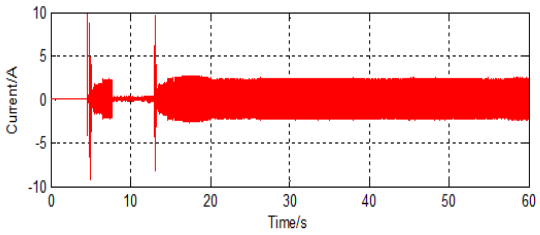
Figure 1.8 The *enPowerMe*™ load monitoring system

1.4.2 Utilizing long-term waveforms for PEL identification

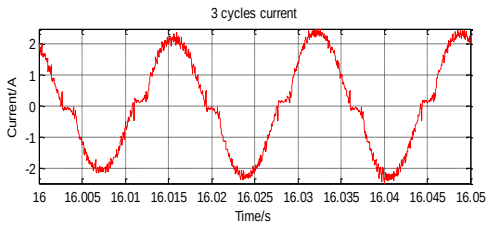
As reviewed and summarized in Chapter Two, most existing load identification methods in literature utilize only short-term voltage and current waveforms (for instance, typically several electrical cycles), which is not so reliable in some cases when applied to PEL identification. For example, the following figures show the short-term (several electrical cycles) and long-term current waveforms (several seconds) of an LCD TV and a laptop computer.



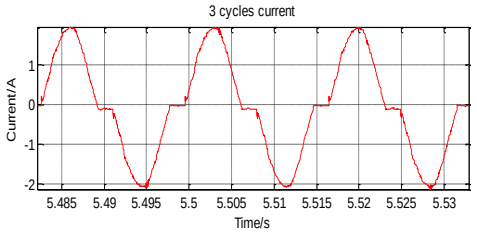
(a) Long-term current waveform of an LCD TV



(b) Long-term current waveform of a laptop



(c) Short-term current waveform of an LCD TV



(d) Short-term current waveform of a laptop

Figure 1.9 Long-term and short-term current waveforms of an LCD TV and a laptop computer

The short-term current waveforms of the LCD TV and laptop computer are quite similar, which makes it difficult to tell them apart. However, the long-term operating

current waveforms of these two PELs are quite different and should be used to get more accurate PEL identification.

Thus, a reliable method is needed to model or represent the shapes of long-term voltage and current waveforms with the capability to extract information about the operating status of PELs for the purpose of PELs identification. Some recent work has started to identify operating modes from long-term (hours or days) waveforms utilizing the active power with a low resolution (e.g., one data point every hour) and it focuses on the total energy consumed over a given time period. However, the following issues still remain unsolved:

(1) Identify load operating modes in real-time from high resolution data (e.g., 10^2-10^3 data points per second) for real-time direct load control and energy management;

(2) Identify the steady-state operation as well as the transient operating modes during startup in real-time;

(3) Report not only the total amount of power consumed at each operating mode but also the total amount of time that the PEL is operating at this mode over a certain time period;

(4) Detect certain operating modes in real-time from long-term voltage and current waveforms for the purpose of intelligent electric load identification.

1.5 PROBLEM STATEMENT

This dissertation aims at developing accurate, reliable, efficient, and robust PELs identification using load features extracted from electric signals such as voltage and

current measurements. The proposed research focuses on the following four aspects to provide solutions for advanced PEL identification.

(1) Robustness: achieve meaningful classification and identification of PELs listed in Table 1.1 based on front-end power supply unit circuit topology and electrical operation principles to handle the diversity within each type of PELs and the similarity between different types of PELs;

(2) Accuracy: achieve certain identification success rates under all scenarios and provide solutions when the identification cannot be made without ambiguity;

(3) Adaptiveness: learn from user inputs or feedback, update classification and identification rules if necessary, and include *a priori* information and required identification granularity;

(4) Intelligence: extract signatures/patterns when multiple PELs are connected into a single outlet or power strip such as startup transients and steady-state features, investigate the applicability of the extracted signatures/patterns for effective PELs activity recognition, and identify the unknown PELs to a certain level of granularity.

1.6 HIERARCHICAL IDENTIFICATION FRAMEWORK

Considering the diverse nature of PEL, the enormous number of PELs, and the challenging aspects of advanced PEL identification, this dissertation follows existing work [2, 43-45] which have developed meaningful taxonomy of typical PELs in commercial buildings, uses the suggested taxonomy in [43], and proposes a hierarchical PEL identification framework as shown in the following figure.

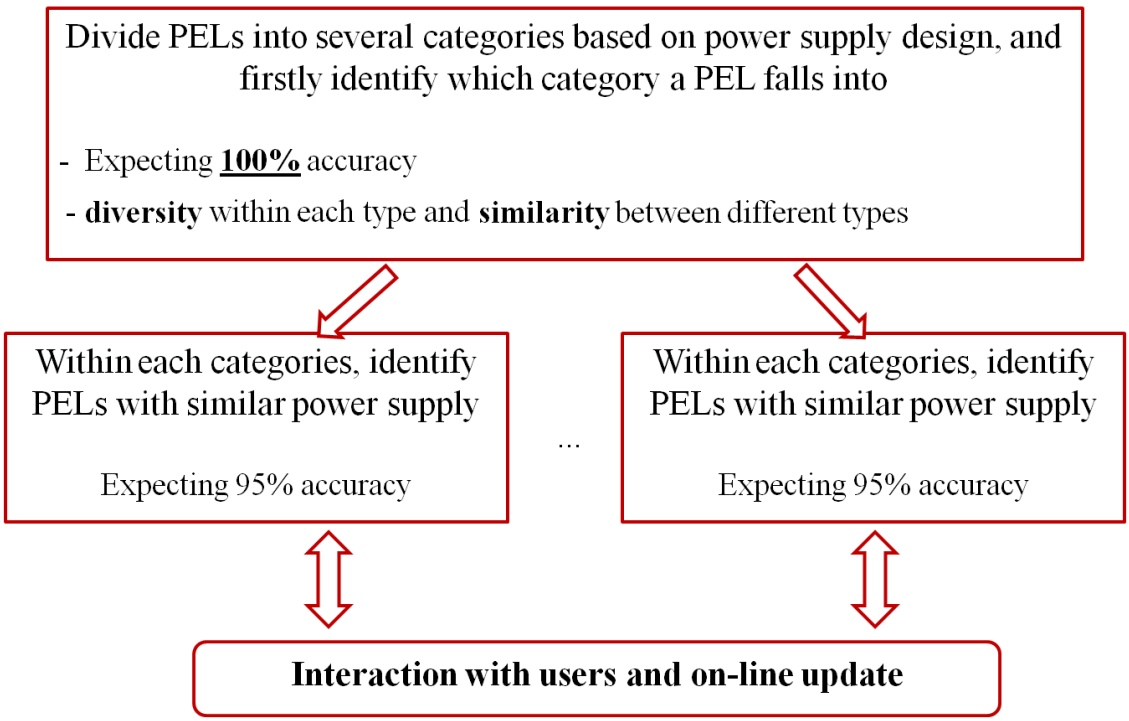


Figure 1.10 Hierarchical PEL identification framework

As shown in Figure 1.8, the proposed hierarchical (multi-level) PEL identification framework consists of three steps:

(1) (Top level) Classification of PELs into a number of categories based on their front-end power supply units topology. In this step, an unknown PEL is first specified into one of the PEL categories.

(2) (Middle level) Within the specified PEL category classified by step (1), the next step is to indicate the actual identity (and operating status if possible) of the unknown PEL.

(3) (Bottom level) If necessary, the proposed PEL identification framework could interact with users and receive inputs to update its identification rules.

1.7 DISSERTATION OUTLINE

A comprehensive literature review of the existing methods and techniques pertinent to this dissertation is summarized in Chapter Two. Chapters Three to Seven constitute the main body of this dissertation, which can be divided into three parts:

1.7.1 Feature extraction for PELs

Chapter Three proposes a low computational-cost but yet efficient method to extract load signatures for PELs classification and identification. Instead of carrying out frequency domain analysis such as DFT and FFT, Chapter Three proposes to extract the similarity of voltage-current (V-I) trajectories between loads by mapping V-I trajectories to a grid of cells with binary cell values. A novel set of graphical signatures extracted from the grid cells with V-I trajectories mapped on is presented, which can be utilized for many applications.

1.7.2 Classification of PELs into categories

Chapter Four introduces the fundamental framework of the self-organizing map (SOM) and the extension of SOM to a supervised manner for classification and identification of PELs. The supervised SOM (SSOM) can classify a large amount of PELs into several groups. Different sets of PEL features, including both time-domain and frequency-domain feature, are considered to be used in SSOM. Chapter Five presents a novel combination of the SSOM framework and the Bayesian identifier framework to function as a hybrid identifier and provide the probability of an unknown PEL belonging to a known category.

1.7.3 Identification of PELs in each category

Chapters Six and Seven discuss in-category identification of PELs, i.e., identifying similar PELs within each PEL category.

For static method, Chapter Six presents a novel hybrid SSOM/SVM identifier for the multi-class in-category PEL identification problem. The proposed hybrid identifier utilizes the power of previously supervised Self-Organizing Map (SSOM) classifier for PELs proposed in Chapters Four and Five to first classifie an unknown PEL into one of the seven PEL categories discussed in Chapter Three. Within each cluster, a more accurate identification decision is made by the well establish multi-class one-against-all SVM classifier. The results are satisfactory for the testing purpose.

For dynamic methods, Chapter Seven proposes a novel finite-state-machine (FSM) representation of long-term operating waveforms for the purpose of indicating load identity and operating modes. The operating current or voltage waveform is converted into a quantized sequence of states. A set of elemental states and events are defined to reduce the number of states and extract numerical features to represent and identify PELs under different operating modes. Three major categories of repeating patterns in waveforms that correspond to repeating operating actions are summarized, and identification methods are proposed for each such category.

Finally, Chapter Eight summarizes the main contributions and lists outcomes of this dissertation.

CHAPTER 2 LITERATURE REVIEW

2.1 INTRODUCTION

Started with its first introduction by Hart [41] in the late 1980's, the non-intrusive load monitoring (NILM) problem has attracted wide range of attentions and interests globally. A large amount of work has been reported on electric load identification by worldwide researchers. Most existing work in the literature follows a common process which is summarized as a general framework and presented in section 2.2.

This general framework for electric load identification contains three main modules/steps: event detection, feature extraction, and load identification using extracted features. The load identification process starts if a turn ON/OFF event is detected. How to detect ON/OFF events is reviewed in section 2.3.

A set of features of an electric load is defined as its unique signature which can represent its characteristics. With an electric load with unknown identity represented by a pre-defined set of features, the identification decision is then made by comparing the features with a reference database and finding out a known load with most similar features. Major existing electric load features and feature comparison methods in the literature are reviewed in sections 2.4 and 2.5, respectively.

Some recent work aims at determining not only the identity of electric loads but also the operating status of electric loads, which is not a part of the original NILM problem. Related work in the literature is discussed in section 2.6.

Finally, section 2.7 summarizes this chapter.

2.2 General Framework for Electric Load Identification Systems

Starting with the original idea of non-intrusive load monitoring (NILM) by Hart in the 1990's [41], many methods have been proposed to monitor and identify electric loads over the past 20 years. Many electric load identification systems have been proposed, built, and tested based on these methods. Most existing electric load identification systems follow the general framework shown in the following figure [46-48].

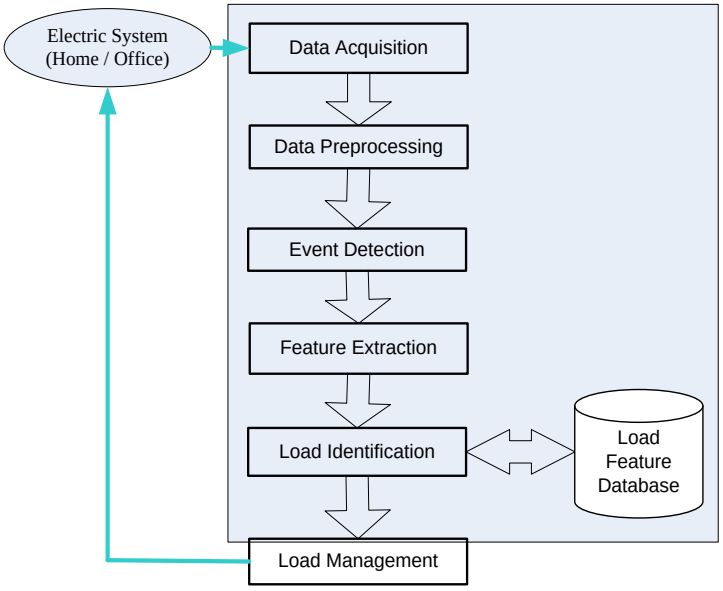


Figure 2.1 General framework for electric load identification systems [42-44].

In Figure 2.1, the data acquisition (DAQ) module captures steady-state raw data as well as transient signals if necessary, and then the data preprocessing module carries out predefined data conditioning and processing actions such as filtering, normalization, and frequency spectrum calculation by the discrete Fourier transformation (DFT) or the fast

Fourier transformation (FFT). Some load identification methods depend on the event detection module to detect whether there is an actual electric load being switched on/off. The event detection module can be implemented by triggering simple thresholds such as instantaneous active power or sophisticated thresholds such as root-mean-square (RMS) values, harmonics, and/or transient values.

The feature extraction module is the key part, which varies notably in different systems as it determines the accuracy and performance of the overall load identification system. Features can be either time-domain (from voltage and current waveforms) or frequency-domain (from harmonic spectrum of steady-state signals). The extracted features represent the characteristics of electric loads. The load identification module utilizes the extracted features, compares them with a database of features of known electric loads, and identifies the unknown load based on pre-defined rules such as maximum similarity.

The load management module, usually decoupled from the load identification section composed of the previous five modules, utilizes the information generated from the load identification module and provides granular load energy consumption and performance details to drive various building energy management tasks such as energy intensity reduction, demand reduction, peak shaving, energy optimization, and proactive equipment maintenance.

The major differences between the various load identification systems mainly fall into the so-called adopted *features* and *feature comparison* method reviewed in the rest of this chapter.

2.3 REVIEW OF EVENT DETECTION METHODS

Event detection based methods have typically been adopted in the earlier load identification systems and have later been replaced by other more advanced load identification methods.

2.3.1 Event detection using steady-state values

Early work [49-51] proposes to continuously monitor the operation of electric loads and search for changes in steady-state active and reactive power. A significant change of exceeding a predefined threshold of adopted features is considered as an indication of an electric load being switched on/off and the differences in steady-state active and reactive power values are considered to be the distinguishing characteristics of that load. Subsequent identification is then made by comparing the distinguishing characteristics of that load with a library of known characteristics of typical loads.

However, this method works only for a limited number of scenarios with only a few quite different electric loads. It is shown in [49] that this method can identify the switching of a refrigerator, an oven element, and a stove burner. Furthermore, the steady-state real and reactive power are even less informative in commercial buildings where substantial efforts, such as power factor correction and load balancing, are made to homogenize the steady-state behavior of different loads.

2.3.2 Event detection using transient characteristics

In order to overcome the limitations of steady-state values, some later work suggest considering transient characteristics. In [52], a multi-scale transient event detection algorithm is introduced to identify individual loads in buildings by examining measured

transient profiles observed in the aggregated current waveforms available at the service entry. This algorithm can be used to identify observed transient waveforms even when multiple transients overlap. In [53], a transient event detection method using voltage distortion is proposed. The implementation of transient event detector using a multiprocessor is explained in [54]. However, these methods are designed for major appliances with distinguishing characteristics but cannot be directly applied to PELs with similar characteristics.

The event detection module is typically not included in later load identification systems, the majority of which directly extract time-domain or frequency-domain electric features as the characteristics of electric loads. Compared with events, the electric features of electric loads are of higher dimension and thus possess a better descriptive capability.

2.4 REVIEW OF FEATURES FOR ELECTRIC LOADS

The performance of almost all existing load identification methods in the literature highly depends on the electrical features (also called signatures in some context) of the loads, which are defined to be "an electrical expression that a load device or appliance distinctly possesses" [44].

Assume that the voltage and current waveforms can be represented by the following equations:

$$V(t) = \sum_{k=1}^{\infty} V_k \sin(k\omega_0 t + \delta_k), \tag{1}$$

$$I(t) = \sum_{k=1}^{\infty} I_k \sin(k\omega_0 t + \theta_k), \tag{2}$$

where ω_0 is the fundamental frequency, V_k and δ_k denote the magnitude and phase angle of the *k*-th harmonic in voltage, and I_k and θ_k denote the magnitude and phase angle of the *k*-th harmonic in current, respectively.

The following steady-state features for electric loads [48-51, 55] can be deduced from the voltage and current waveforms and are widely used in various electric load identification systems.

2.4.1 Active and reactive power

The amount of active power an electric load consumes in real-time or the average amount of active power it consumes over a certain period of time is probably the most straightforward and intuitive feature of this load. Furthermore, the amount of reactive power can roughly indicate whether this load is resistive, inductive, or capacitive.

2.4.2 Peak, average, and RMS current values

Peak current and average current are proposed for load identification in [56]. Furthermore, the root mean square (RMS) value I_{RMS} of the current measurement is also considered [57, 58]. However, I_{RMS} gives equivalent information on the active power but needs no additional multiplication (with voltage).

2.4.3 Instantaneous values

Instantaneous values such as instantaneous active power p_{inst} [59], current i_{inst}, and admittance y_{inst} can serve as features. Some electronic converter connected loads may have huge spikes of instantaneous admittance, which separate them from other loads.

The main disadvantage of instantaneous values is that a proper time scale should be defined because there cannot be either too many or too few number of instantaneous features.

2.4.4 Harmonics spectrum of the current waveform

The harmonic spectrum of the current is proposed to identify loads in [47, 48, 50, 60] as the current waveform in the time domain provides one of the most complete sets of information to describe the behavior of electric loads. The main advantage of using current harmonics lies in the high resolution of the signal which can reflect detailed characteristics of the appliance.

More specifically, the 3^{rd} and 5^{th} harmonics are more informative than others. Any single phase device (such as desktop, laptop, TV, and LCD monitor) which contains a switching mode power supply (SMPS) contains high percentages of the 3^{rd} and 5^{th} harmonics in the current waveform. Therefore, the magnitude and phase of the 3^{rd} and 5^{th} harmonics in the current waveform can also be considered as features.

2.4.5 Total harmonic distortion (THD)

The total harmonic distortion of the current waveform is widely [59, 61, 62] adopted to describe the linearity of an electric load as well as power quality. Linear loads draw current that is sinusoidal while nonlinear loads draw a current that is not perfectly sinusoidal, i.e., distorted. With the harmonic components $I_k, k = 1,...,\infty,$ of the current waveform calculated by the Fourier transformation, the THD in the current waveform shown by equation (2) is defined as

$$THD_I = \frac{\sqrt{\sum_{k=2}^{\infty} I_k^2}}{I_1} \cdot 100\% \tag{3}$$

2.4.6 Power factor

There are several different power factor definitions available such as displacement power factor PF_{disp}, distortion power factor PF_{dist}, and power factor PF [63, 64]. They provide equivalent information and are defined as follows.

$$PF_{disp} = \cos(\delta_1 - \theta_1) \tag{4}$$

$$PF_{dist} = \frac{1}{\sqrt{1 + THD_I^{\,2}}} \tag{5}$$

$$\begin{aligned} PF &= \frac{\cos(\delta_1 - \theta_1)}{\sqrt{1 + THD_I^{\,2}}} \\ &= PF_{disp} PF_{dist} \end{aligned} \tag{6}$$

where δ_1 and θ_1 are the fundamental voltage and current angles, respectively.

Typically, only one or two of the above three definitions are used in one system to avoid redundancy. In this dissertation, two different definitions of the power factor by (4) and (5) are adopted.

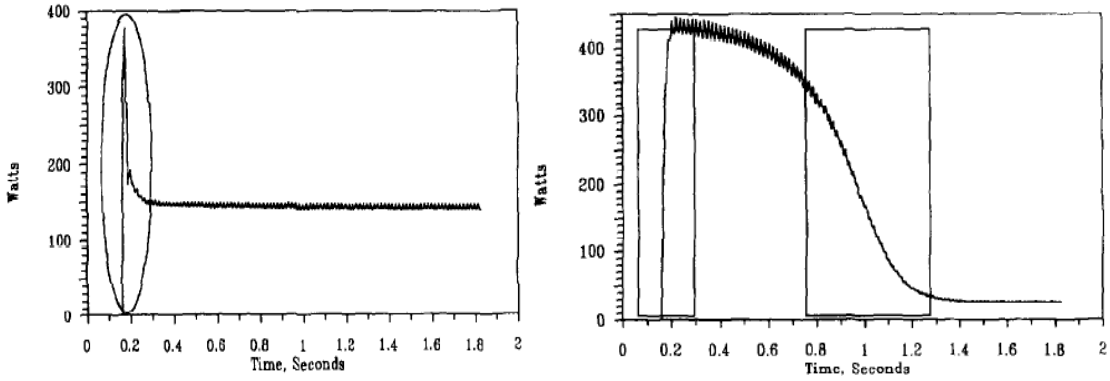
2.4.7 Crest factor or peak-to-average ratio

The crest factor (CF), also called peak-to-average ratio, is defined to be the ratio of peak value to the RMS value of a certain waveform. For example, the crest factor for current CF_I is defined as

$$CF_I = \frac{|I_{peak}|}{I_{RMS}} \tag{7}$$

2.4.8 Transient features

The transient power is also used as a feature for variable electric loads [60, 65-67]. The following figure is taken from [65] as an example.



(a) (b)

Figure 2.2 Transient features in active power of (a) a lamp bank and (b) an induction motor [65].

The transient power is suggested in [53, 65] to be calculated for every half electric cycle (each electric cycle is 1/60 seconds in U.S.) and the resulting switching transient waveforms are shown in Figure 2.2. It is straightforward to observe that the active power transient of the lamp bank has a sharp rise to its peak value and then drops to its steady-state value in less than 0.1 second. However, the active power transient of the induction motor drops much more slowly (in around one second).

Other work by Leeb and his team [52-54, 65] proposes to use shapes of the transient waveforms to distinguish different loads. This approach can identify simultaneously switched loads when the transients do not overlap. However, when the loads are switched on too frequently so that their transients overlap significantly, the loads may not be identified.

The major problem of using transient features in real-world applications is that they may not be able to be observed or detected repetitively for different models or brands of a certain electric load because a type of electric loads may have similar not identical transient profiles. In other words, the transient feature of each load is typically concluded within a certain range instead of a certain value. Thus, the identification decision would be inaccurate if different electric loads have overlapping ranges of transient features. This issue has not been well addressed in the literature for the purpose of electric load identification.

2.4.9 Graphical features

Instead of using numerical values as features for electric loads, it is proposed in [44] to use graphical signatures in the two-dimensional voltage-current (V-I) trajectory as electric load features. A V-I trajectory is plotted in a two-dimensional figure with voltage values on the horizontal axis and current values on the vertical axis.

It is claimed in [44, 68] that V-I trajectories of different types of electric loads have distinct graphical shapes, which are related to the operating characteristics of the loads. Two examples of V-I trajectories of different electric loads are shown in the following figure.

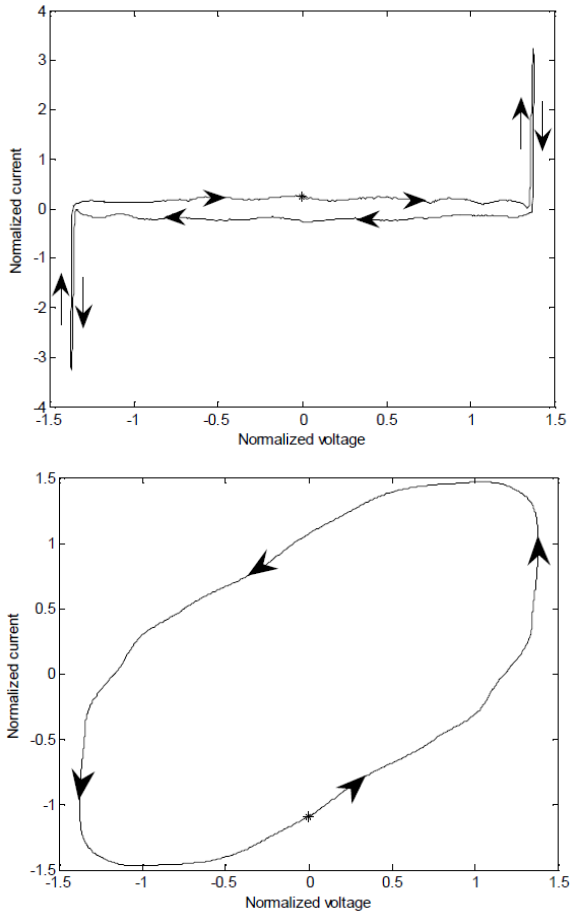


Figure 2.3 V-I trajectories of (a) a desktop computer and (b) a refrigerator [44].

It is summarized in [68] that there are eight shape features that can describe the V-I trajectory: asymmetry, looping direction, area, curvature of the mean line, self-intersection, slope of middle segment, area of left and right segments, and peak of middle segment. For instance, the asymmetry property and the looping direction can be observed in Figure 2.3. Moreover, the following figure is taken from [68] to illustrate how to divide the V-I trajectories into several segments (left, middle, and right) and extract graphical features from each segment.

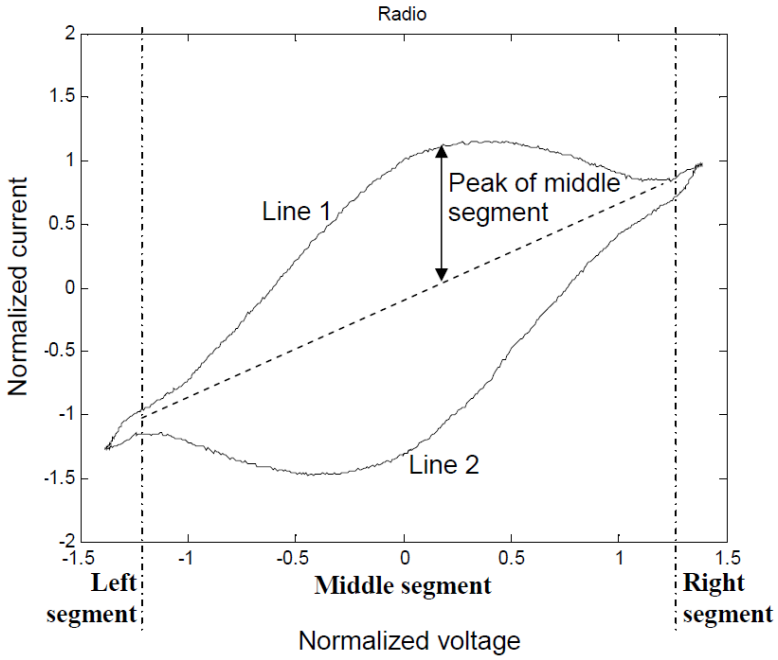


Figure 2.4 Illustration of V-I trajectory segments and measurements [68].

Furthermore, all loads are classified according to the shape features, and the taxonomy of all loads is constructed and then compared to the taxonomies based on traditional features such discussed above.

Note that these graphical features proposed by [68] are extended and used in [69] to study the load disaggregation problem as shown in the following figure. The load disaggregation problem aims at identifying multiple electric loads (which are connected to the same power supply source and thus) from mixed voltage/current waveforms The load disaggregation problem does not fall into the scope of this dissertation and thus is not discussed in details. Some recent survey papers [70, 71] can be referred to as summaries of load disaggregation methods.

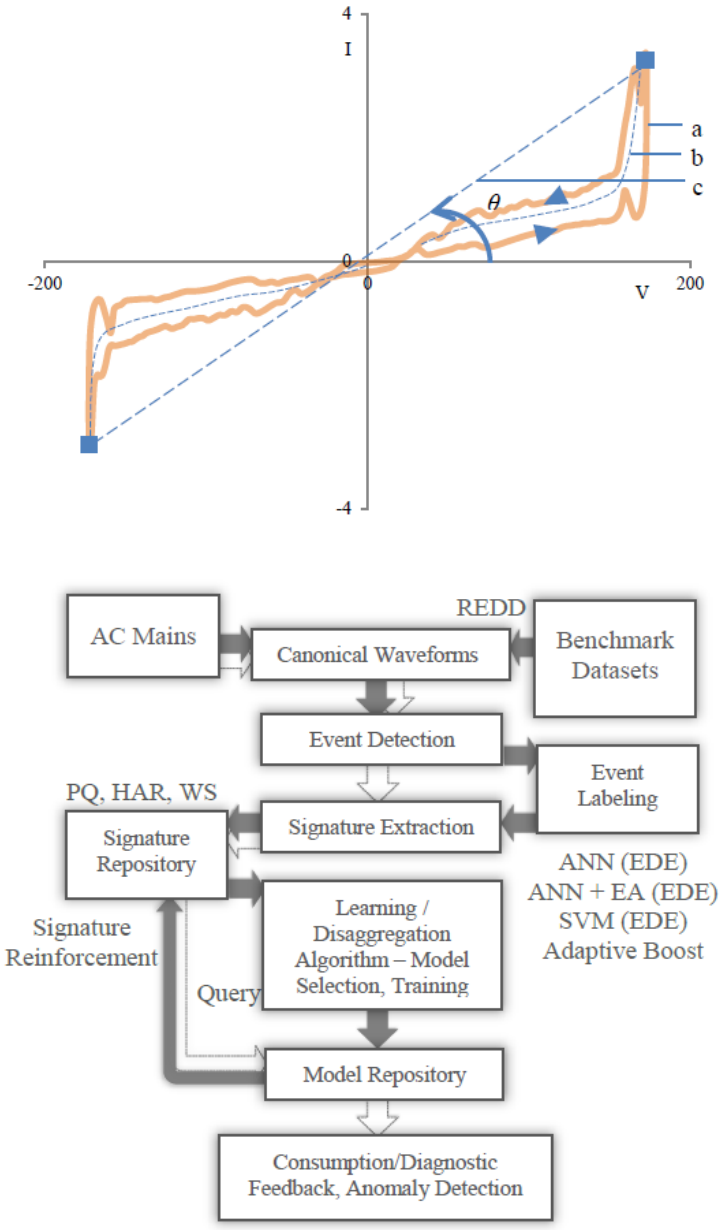


Figure 2.5 Using graphical features for load disaggregation [69]

Like other features, different loads with similar front-end power supply units would possess similar V-I trajectories, especially after normalization on the voltage and current data.

2.4.10 Summary of features

To summarize, many time-domain, frequency-domain, and graphical-based features have been proposed in the literature to characterize electric loads. The major source of information is the measured voltage and current waveforms for non-intrusive load identification. Therefore, some features provide similar information and characteristics as others and thus it is redundant to use all available existing features.

Furthermore, the major criteria to evaluate different features include the complexity of computing features, the similarity between features of loads of the same type, and the diversity between features of loads of different types. Unfortunately, there is no existing set of features that can distinguish all electric loads without ambiguity. A comparison of existing features can be found in [12, 24, 26, 31, 42].

2.5 REVIEW OF IDENTIFICATION METHODS

The load identification module takes extracted features as its inputs and compares the features of an unknown load to a database containing features of known loads. The general principle for identification is that the unknown load is identified as the one of the known loads when the unknown load has features that are most similar to those of one of the known library loads. Many methods have been proposed in the literature to describe how to measure the similarity between two sets of features, which are summarized as follows.

2.5.1 P-Q plane

Hart [49] proposes to use a two-dimensional complex power plane (P-Q plane) to locate relative positions of different appliances, as shown in the follow figure. The real and imaginary axes in the complex P-Q plane denote the value of active and reactive

power values, respectively. Loads that lie far away from each other in the plot can be identified using only real and reactive power, as shown in the following figure [49].

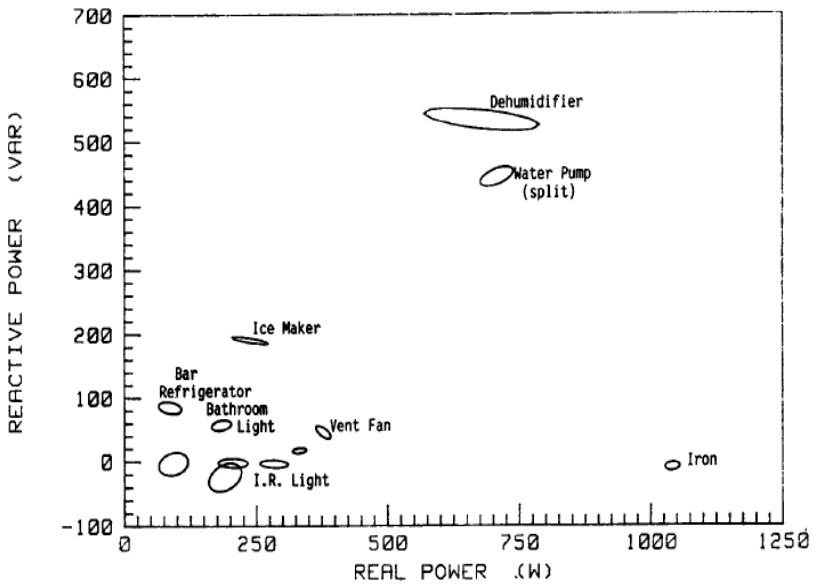


Figure 2.6 Relative positions of a group of appliances in the complex P-Q plane [49].

This method has certain drawbacks as indicated in [60]. For example, this method only works for electric loads that are located far away from each other in the P-Q plane, which may not be true of all electric loads especially in commercial buildings. Also, the P-Q plane becomes crowded with indistinguishable loads as the number of loads increases.

2.5.2 Decision tree

Assume that a set of features has been chosen to set up a database of known loads, and that the value range of each feature of each load can be concluded from the database [43, 72]. When the features of an unknown load come in, an identification decision can

be done by comparing the incoming features with the value ranges of database features step by step, with one feature at each step.

2.5.3 Optimization methods

The similarity problem between unknown and known features is also formulated to be solved as an optimization problem in [47]. The objective function is defined as the minimum difference while comparing an electric load with unknown identity with a set of loads with known identity from a database, i.e.,

$$\underset{j}{\operatorname{argmin}} \sum_{k=1}^{N} w_k (\hat{y}_{(k,j)} - y_k)^2 \tag{8}$$

where $\hat{y}_{(k,j)}$ is the k-th feature of the feature vector j in the known database of loads, y_k is the k-th feature extracted from measurement of the unknown load, w_k is the weight of feature k, N is the total number of feature. The weight w_k can help to adjust the significance of each feature.

2.5.4 Expert system

It is proposed in [50] to utilize the expert system to identify different household appliances. The features adopted include current values, voltage values, active power, duration and shape of the current transient, and harmonics in the current waveform. The household appliances used for experiment are divided into categories such as resistive, pump-operated, motor-driven, electronically-fed, and fluorescent lighting. Test results for selected cases are acceptable.

The disadvantage of using the expert system for load identification is that the expert system depends on the engineer's domain knowledge and requires accurate knowledge of the electric loads being considered.

2.5.5 Artificial Neural Networks (ANNs)

Artificial neural networks (ANNs) can be used to identify electric loads by training ANNs to learn features of known electric loads. Through the training process, the structure and parameters of ANNs are built to capture different features of loads [56, 73].

Different types of ANNs, such as multi-layer-perceptron (MLP), radial-basis-function (RBF), and support vector machines (SVM) are applied in [74, 75]. The ANN is first trained by a database of features of known loads. Once trained, the ANN can perform identification tasks when it is presented with the same set of features of the unknown load. A comparison of performance shows that MLP and SVM-based models are both able to determine the presence of particular devices based on their harmonic signatures [74].

To summarize, the major advantage of ANNs lies in their capability to evolve and learn without extra knowledge. The training process of ANNs is statistical in nature. Therefore, the ANNs are able to extract the statistical information of features from the database and utilize this information to do identification.

2.5.6 Summary of identification methods

Existing methods can be divided into two major categories: methods comparing similarities between extracted steady-state or transient features and their variations with a predefined database as well as computational intelligence algorithms.

Methods in the former category cannot distinguish between different electric loads without ambiguity when extracted features of the unknown load are very similar to

several known loads in the database. On the other hand, ANNs are powerful tools but they also suffer from problems including lack of knowledge during the training, computational cost, convergence criteria, and initial parameter selection.

To summarize, most existing methods cannot efficiently handle the diversity within each type of loads and similarity between similar types of loads and thus cannot be directly applied to the identification of PELs.

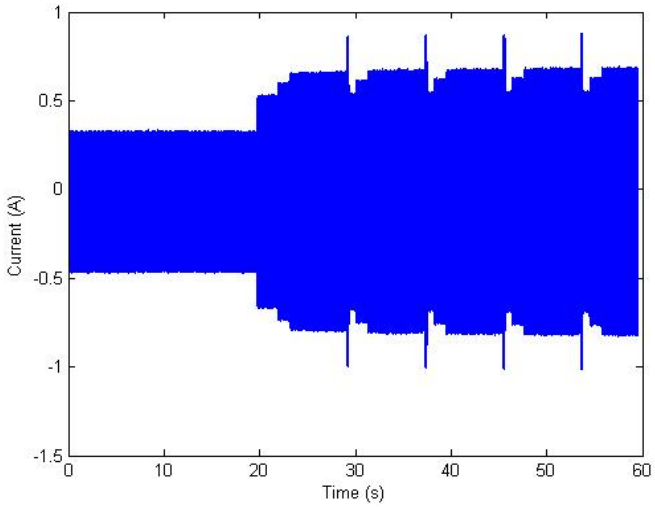
2.6 REVIEW OF OPERATING MODE IDENTIFICATION METHODS

Instead of identifying electric loads based on features extracted from short-term waveforms, recent efforts have started to identify operating modes from long-term (hours or days) waveforms. A recent report by the German Federal Ministry [76] analyzed 4 operating modes of communication devices: normal, standby, off-mode, and off.

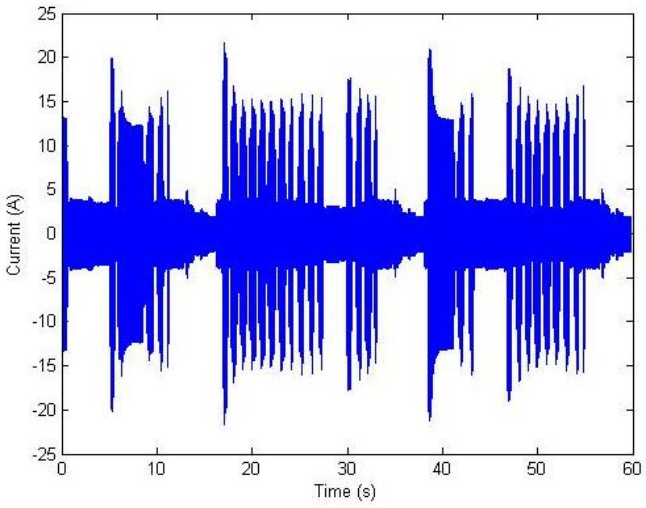
In the U.S., a study by the Lawrence Berkeley National Laboratory [77] employs a non-intrusive inventory-based method to study the power status of office appliances during night-time. It only considers snapshots at single points in time and thus does not provide the time spent in each power status. The National Renewable Energy Laboratory presented a histogram heuristic clustering technique to divide a data set of electric loads operation for several days into clusters based on similarity criteria and extracted operating modes [78].

To summarize, these efforts mainly utilize the active power with a low resolution (e.g., one data point every hour) and focus on the total energy consumed in a given time period. However, some technical problems still remain unsolved, such as identifying real-time operating modes using high resolution data and reporting the total amount of time operating at a certain mode.

The following figure shows the current waveform (of 60 seconds) of two electric loads in offices. Figure 2.6(a) shows the transition from standby mode to faxing (active) mode of a fax machine, and Figure 2.6(b) represents a multi-functional device (MFD) in double sided photocopying mode.



(a) Transition between operating modes of a fax machine



(b) A multi-function-device in recurrent operating mode

Figure 2.7 Current waveforms of office appliances in different operating modes

The detection of the transition from a standby to an active mode in Figure 2.6(a) is a crucial step for energy management, which should not only rely on detecting the change in power. Also, in Figure 2.6(b) the instantaneous peak current is time-varying and typical identifying features in the literature vary from cycle to cycle. Thus, existing methods may fail to correctly identify this multi-functional device (MFD). Therefore, a method is needed which can extract features from long-term and time-varying operations.

2.7 SUMMARY OF CHAPTER

This chapter first presents a general framework for the electric load identification problem, which has been widely used by most existing work in the literature. Furthermore, major existing methods for different modules in this framework, such as feature extraction, event detection, and load identification are reviewed and compared. Advantages and disadvantages of most reviewed methods in this chapter have been presented.

CHAPTER 3 PLUGGED-IN ELECTRIC LOADS CLASSIFICATION BY FEATURES FROM V-I TRAJECTORIES

3.1 INTRODUCTION

As discussed in Chapters One and Two, a primary factor that determines the performance of any electric load identification system is the set of features selected to represent electric loads. Therefore, a large number of work on different electric load features in the literature has been reported by researchers as reviewed in section 2.4.

This chapter proposes a set of computationally efficient but yet accurate features to represent PELs for the purpose of PEL classification. Section 3.2 presents a classification of PELs into seven categories by their front-end power supply circuit topology. Based on the power supply circuit topology, V-I trajectories of PELs within the same category are very similar in shape. Typical V-I trajectories of each PEL category are shown in section 3.3.

Based on the analysis in section 3.3, a set of graphical features are then proposed in section 3.4 by first mapping a V-I trajectory onto a grid of cells with binary values and then extract certain graphical features from the mapped cell grid.

The computational cost of the proposed features is analyzed in section 3.5, which shows that they require less computational resources than features in the literature. Expected values of the proposed features for each PEL category are summarized in section 3.6 for the purpose of PEL classification.

Finally, section 3.7 summarizes this chapter.

3.2 CLASSIFICATION OF PLUGGED-IN ELECTRIC LOADS BY FRONT-END POWER SUPPLY CIRCUIT TOPOLOGY

The number of types and models of commercially available PELs currently used in residential and commercial buildings is enormous. Furthermore, considering the fact that voltage and current waveforms are the only source of information available for PELs feature extraction and identification, front-end power supply units of PELs play a key role as they directly determine the characteristics of the current waveform. For instance, as discussed in section 1.4.1, PELs within the same type (i.e., flat-panel TVs) could be equipped with different power supply units and thus present quite different current waveforms. On the other hand, different types of PELs may be equipped with similar power supply units.

Therefore, it is neither feasible nor necessary to characterize and identify each PEL individually in many applications. Instead, it is sometimes more practical and robust to first classify all PELs into several categories by their front-end power supply topology and then extract common signatures for PELs in each category as shown in the hierarchical identification framework shown in Figure 1.8.

Based on a study on over 95% of all commercially available front-end power supply topologies, it is proposed in [43] to divide PELs into the following seven categories based on their front-end power supply circuit topology:

(1) *Resistive loads* (Category R): a typical PEL in this category contains a resistance directly connected to the front-end terminal and thus there is no phase angle difference between its current and voltage waveforms;

(2) *Reactive loads* (Category X): a typical PEL in this category contains an inductance directly connected to the front-end terminal through a rectifier and thus there is a large phase angle difference between its current and voltage waveforms;

(3) *Electronic loads without power factor correction* (Category NP): a typical PEL in this category consists of a front-end electromagnetic interference (EMI) filter, a rectifier, a voltage or current filter, and a DC-DC converter. There is typically a very small phase angle difference (close to zero) between its current and voltage waveforms but the current waveform contains a notable amount of harmonics;

(4) *Electronic loads with power factor correction* (Category P): a typical PEL in this category consists of a front-end EMI filter, a rectifier, a voltage regulator, a power factor correction (PFC) module, and a DC-DC converter. Its current waveform is similar to resistive loads, but notable current discontinuity and switching noise can be observed;

(5) *Complex structure loads* (Category M): a typical PEL in this category consists of multiple circuits supplied by independent front-end power supply units and thus its overall current waveform is composed of current waveforms from one or more of the above four categories;

(6) *Linear loads* (Category T): a typical PEL in this category consists of a transformer, a rectifier, and electronic components. Its current waveform is highly distorted due to transformer saturation. Notable phase angle difference can also be observed;

(7) *Phase angle controllable loads* (Category PAC): a typical PEL in this category continuously adjusts its current waveform by controlling the firing angle of a thyristor.

Furthermore, category M loads also include PELs that operate at several different power levels and switch between these power levels repeatedly during usage. These PELs are programmed in a pre-defined manner to operate in this repeated switching-mode manner because their functional performance may require repeated processes in a certain sequence.

For example, most high volume printers have two (or more) printing engines/motors in a single device and are able to print both sides of the paper in a single pass, i.e., double-sided printing. A double-sided printing job is a repeated process of feeding a sheet of paper, printing and rolling the paper forward, holding the paper for the ink to dry, reversing the paper to print on the other side, and then feeding the next sheet of paper. The two engines are programmed to operate in different combinations with different power consumption levels during this repeated process, and these combinations could fall into one or several other categories.

Note that the front-end power supply circuit topology of categories T and PAC are no longer adopted in modern power supply industry, but these two categories are still included in this dissertation for completeness.

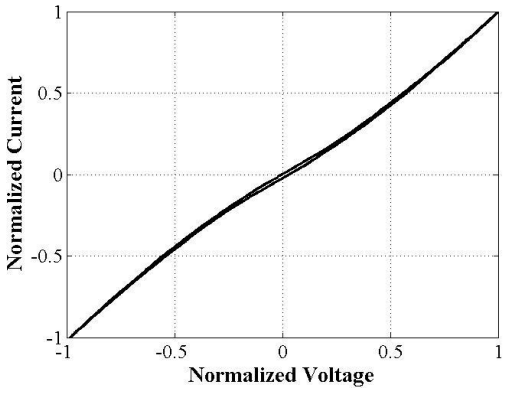
3.3 TYPICAL V-I TRAJECTORIES OF EACH PEL CATEGORY

In the literature, a large number of existing works on the characterization and identification of PELs use features extracted from the harmonic spectrum of current waveforms derived by discrete Fourier transformation (DFT) or fast Fourier

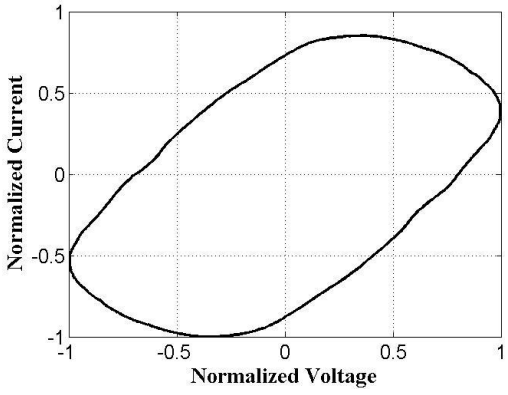
transformation (FFT). However, for the purpose of practical applications with only limited computational capability or hardware capability, the computational cost of DFT or FFT is probably too high in some cases. For example, if a PEL identifier is desired in a power strip or a power outlet, these applications may only have a micro-processor with very limited amount of memory. Fourier transforms may not be desired in these applications.

It is observed that the normalized V-I trajectories of PELs within each category share very similar shapes, which can be used to describe and represent PELs within each category. Furthermore, PELs of different categories possess quite different shapes of normalized V-I trajectories. In other words, normalized V-I trajectories described in a properly defined metric space can be used as features to distinguish different categories as they are close within-category but quite far away between-categories in the manner of distances.

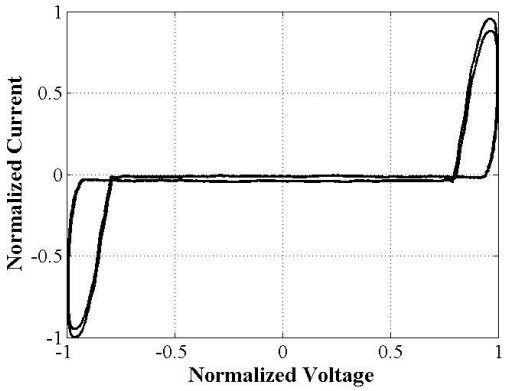
Typical normalized V-I trajectories of the seven load categories discussed above are shown in the following figure.



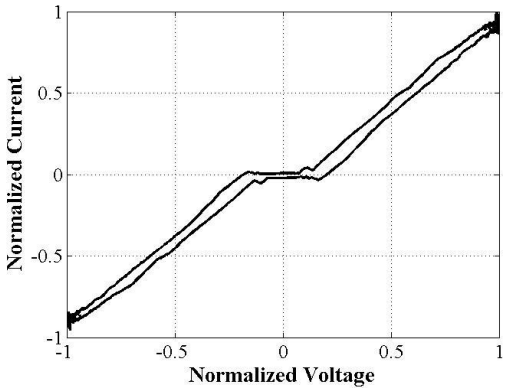
(1) Category R



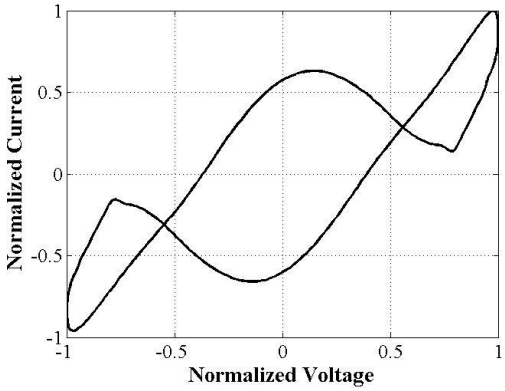
(2) Category X



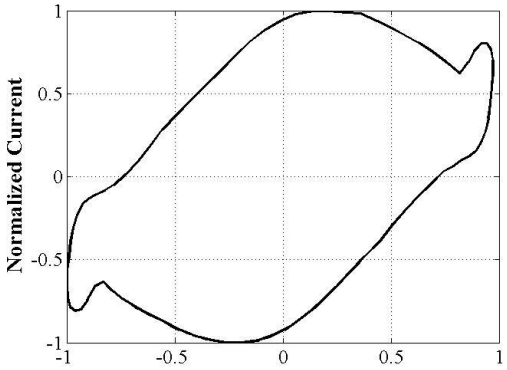
(3) Category NP



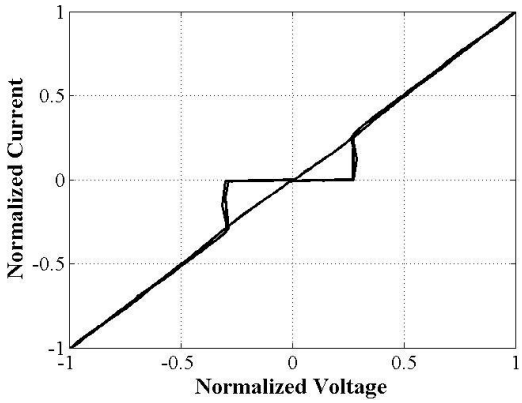
(4) Category P



(5) Category M



(6) Category T



(7) Category PAC

Figure 3.1 Typical normalized V-I trajectories of the seven load categories

Note that here "normalized" means that the discrete-time sampling measurements of the voltage and current waveform are normalized by their maximum values. Figure 3.1 shows that normalized V-I trajectories from different load categories appear different. Some recent studies focus on developing graphical (load) features from V-I trajectories to represent different loads whose voltage and current measurements form the corresponding V-I trajectories. A notable study based on 126 sets of operating data of different PEL types and modes summarizes that there are eight shape signatures that can be considered to describe the V-I trajectory: asymmetry, looping direction, area, curvature of the mean line, self-intersection, slope of middle segment, area of left and right segments, and peak of middle segment [44].

However, calculating these graphical features still requires a large amount of computational resources as the entire V-I trajectory needs to be traversed in a certain order or direction. Also, these features are designed for a taxonomy of loads that is similar to the load groups defined by IEC Standard 61000-3-2. Therefore, they are not suitable for the proposed seven PELs categories. For example, PELs from Category R and Category P have very similar such features and thus cannot be effectively distinguished without ambiguity.

3.4 Feature Extraction by Mapping V-I Trajectories to Cell Grids with Binary Values

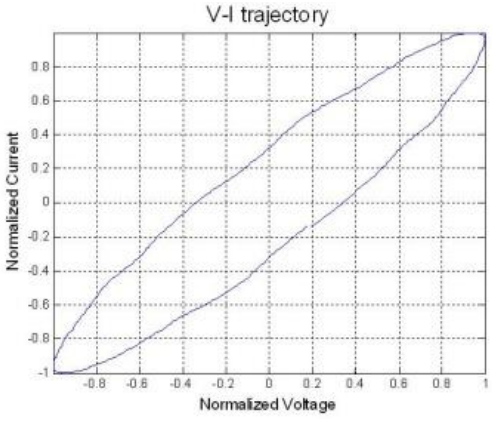
3.4.1 Limitation in performance of existing graphical load signatures

The existing graphical load features discussed above are purely based on the shape of a V-I trajectory. However, as discussed in Chapter One, different models of PELs within the same category can be equipped with similar (but not identical) front-end power

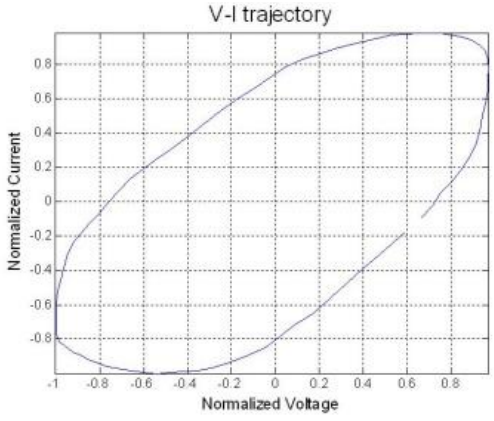
supply topology. Therefore, such PELs present similar (but not identical) current waveforms as well as V-I trajectories. In this case, there can be significant differences in some of the existing graphical load features, which are supposed to be identical as these PELs belong to the same type or category. Moreover, some existing graphical features may no longer be true or useful.

Several examples are presented in Figures 3.2(a) and 3.2(b), which show the V-I trajectories of two portable fans. These two V-I trajectories have similar shapes but quite different area values of both the entire V-I trajectory and of the left and right segments, as well as the peak values of the middle segments.

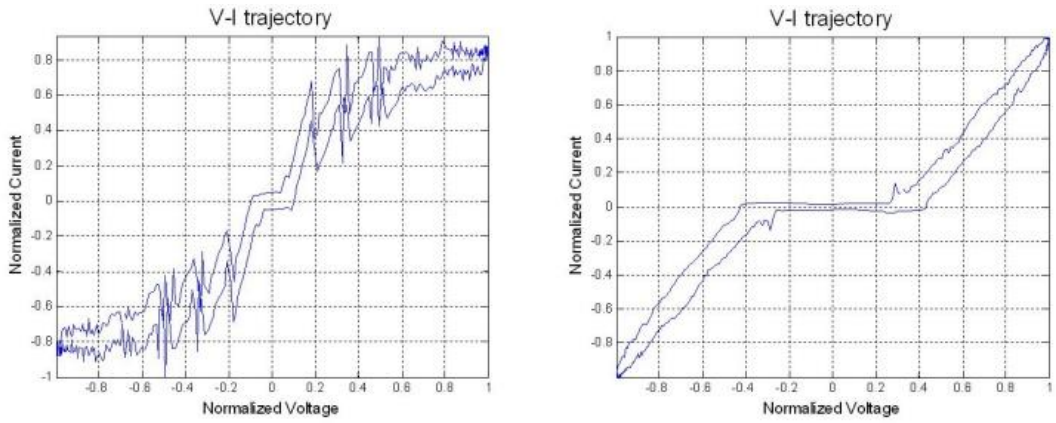
As another example, Figures 3.2(c) and 3.2(d) show the V-I trajectories of two flat-panel TV sets, which have similar shapes but quite different zero crossing times (and thus left, middle, and right segment values). Also, determining the asymmetry, looping direction, and area of Figure 3.2(c) is complicated and consumes much computation effort due to the oscillations in the V-I trajectory.



(a) A 32-inch portable fan



(b) A 9-inch portable fan



(c) An LED TV (d) An LCD TV

Figure 3.2 Normalized V-I trajectories of four plugged-in electric loads to illustrate the limitation in performance of existing graphical load signatures

3.4.2 Binary mapping from V-I trajectories to cell grids

In order to effectively handle the variance in the current waveform as well as the difference between V-I trajectories of PELs within the same PELs category and reduce the error as well as computational cost as discussed earlier, this dissertation proposes to first map a V-I trajectory to a (square) grid of cells. Each cell is assigned a binary weight value.

Furthermore, if the V-I trajectory crosses though a cell in the grid, this cell is considered to be occupied by this V-I trajectory and then assigns its binary weight value to be 1. An occupied cell is shown as a solid black cell as shown in Figure 3.3. If not, the cell is then assigned a binary weight value of 0 and shown as an empty cell as shown in Figure 3.3.

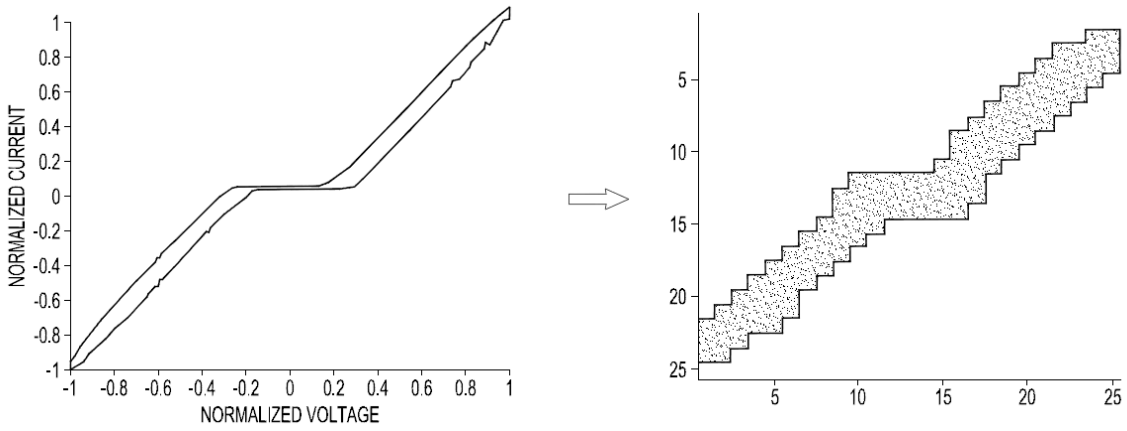


Figure 3.3 Mapping V-I trajectory to a binary cell grid

The binary cell grid is a generalization of V-I trajectories. V-I trajectories with similar but not identical shapes can have identically mapped binary cell grids. This is because two V-I trajectories can pass through a cell along slightly different paths but the cell is yet considered as occupied and assigned a value of 1. Also, the distortion in a V-I trajectory that is caused by the harmonic present in the current waveforms can be smoothed out as the part within an occupied cell is represented by this cell, no matter how distorted this part of the V-I trajectory might be. For example, the V-I trajectory of a PEL from category P shown in Figure 3.2(c) is quite distorted. On the other hand, the V-I trajectories of other PELs from category P shown in Figure 3.2(d) and Figure 3.3 are much more smoother with no notable distortion. However, after mapping to a grid of cells, all of these three V-I trajectories look the same, i.e., similar (if not identical) to the mapped cell grid in Figure 3.3.

The binary cell grid mapping algorithm is defined as follows.

(1) Load the voltage and current data: assume that there are a total of K data points of the form (v_k, i_k), where $k = 1,\ldots,K$. Also, v_k and i_k are the voltage and current values of sampled data point k, respectively;

(2) Calculate

$$\begin{aligned} v_{\max} &= \max v_k, \\ v_{\min} &= \min v_k, \\ i_{\max} &= \max i_k, \\ i_{\min} &= \min i_k, \\ v_0 &= \frac{1}{2}(v_{\max} + v_{\min}), \text{ and} \\ i_0 &= \frac{1}{2}(i_{\max} + i_{\min}). \end{aligned} \tag{9}$$

Note that in general $|v_{\max}| \neq |v_{\min}|$ and $|i_{\max}| \neq |i_{\min}|$. Thus (v_0, i_0) are used as the origin of the cell grid instead of (0, 0).

(3) Read input N which defines the size of the grid in the horizontal direction and calculate

$$\begin{aligned} \Delta v &= \frac{v_{\max} - v_0}{N} \\ \Delta i &= \frac{i_{\max} - i_0}{N} \end{aligned} \tag{10}$$

and generate two sequences

$$\{v_0 - N\cdot\Delta v,\ v_0 - (N-1)\cdot\Delta v,\ \ldots\ , v_0 - \Delta v, v_0 + \Delta v, \ldots\ , v_0 + (N-1)\cdot\Delta v,\ v_0 + N\cdot\Delta v\}$$

and

$$\{i_0 - N\cdot\Delta i,\ i_0 - (N-1)\cdot\Delta i,\ \ldots\ , i_0 - \Delta i, i_0 + \Delta i, \ldots\ , i_0 + (N-1)\cdot\Delta i,\ i_0 + N\cdot\Delta i\},$$

which both have $2N$ elements. Note that the former sequence is from voltage sampling measurements and the latter one is from current. Also, both sequences use the same N because the V-I trajectory considered is normalized and the range of both horizontal and vertical direction are between [-1, 1] and symmetrical.

(4) Define an $2N \times 2N$ square cell grid, and the $\left(x^{\text{th}},\ y^{\text{th}}\right)$ cell is assigned with a position value $\left(v_0 + \Delta v \cdot (x - N),\ i_0 + \Delta i \cdot (y - N)\right)$ and a binary model value $B_{x,y}$ which is initialized to be 0;

(5) Load data points of one-half cycle, starting from the zero crossing point from negative to positive to another zero crossing point from positive to negative;

(6) Start with the first data point of the data points loaded in step (5), which is denoted by $\left(v_1^h, i_1^h\right)$, and execute the following loop (described in pseudo code):

for every cell $\left(N+1,\ y\right),\ y = N+1, N+2,\ \ldots,\ 2N$

if $\left(v_1^h - v_0\right) < \frac{\Delta v}{2}$ and $\left(i_1^h - (i_0 + (y - N) \cdot \Delta i)\right) < \frac{\Delta i}{2}$

cell $\left(N+1,\ y\right)$ is occupied and $B_{\Delta+1,\ y} = 1$;

cell $\left(N+1,\ y\right)$ is stored as the <u>*winner*</u> of $\left(v_1^h, i_1^h\right)$;

BREAK;

end

end

(7) For the remaining half-cycle data points from step (5), repeat step (6) but only search the 8 adjacent cells (as shown in the following Figure 3.4) of the previous winner indicated by the above pseudo code.

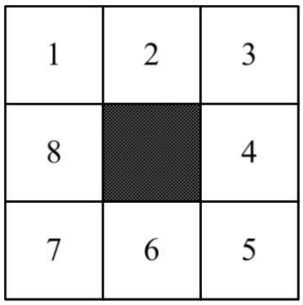


Figure 3.4 Eight neighbors of a cell in the grid

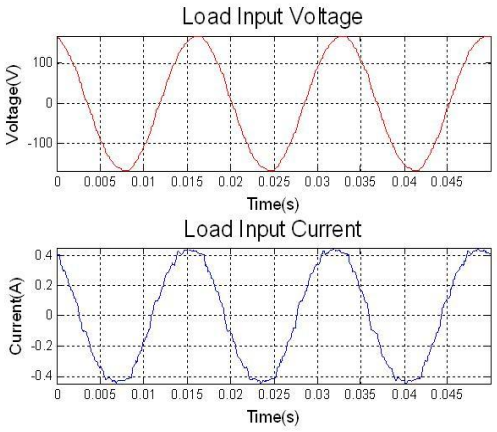
(8) Repeat from Step (6) for a pre-defined number of times.

(9) End.

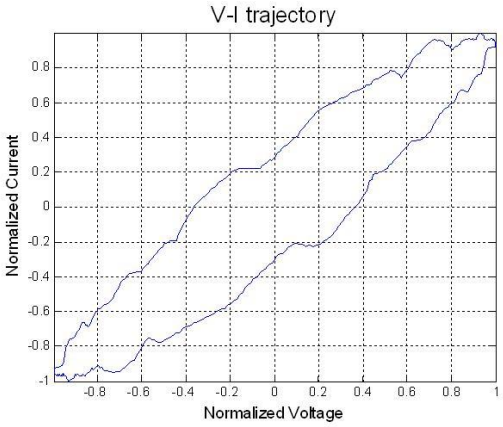
Note that besides defining the size of the cell grid ($N \times N$, $N=2\Delta$), the parameter N can also be interpreted as the width of each cell because the cell grid is square, normalized, and defined over [-1, 1] by [-1, 1] (in other words fixed overall width). Therefore, N should be carefully chosen based on different applications. If there are too many cells, the mapping of V-I trajectories to the binary cell grid may not effectively handle the variance of similar V-I trajectories. However, the mapped binary cell grid may not correctly represent the V-I trajectories if the number of cells is insufficient.

An example is given in the following figures to illustrate the mapping of a V-I trajectory to a binary cell by the above algorithm. The voltage and current waveforms of

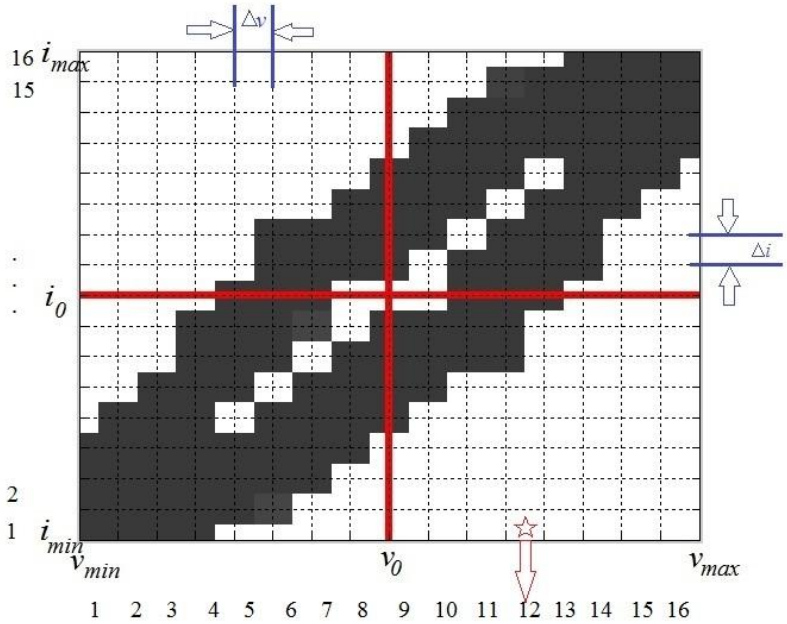
a portable fan in medium-speed steady operating state is shown first in Figure 3.5(a). The corresponding V-I trajectory from this set of voltage and current waveforms as well as how it is mapped to a cell grid are shown in Figure 3.5(b) and 3.5(c), respectively.



(a) Voltage and current profile of a fan



(b) Corresponding V-I trajectory



(c) The cell grid after mapping a V-I trajectory

Figure 3.5 Illustration of mapping a V-I trajectory to a binary cell grid

In this example, sampled voltage and current measurements of one cycle (1/60 seconds in U.S.) are mapped to a cell grid with binary values by the algorithm proposed above. For the purpose of illustration, in Figure 3.5 the parameter N is set to be 8 and thus the generated cell grid is of the size 16×16.

From the sampled data for this example,

$$
\begin{aligned}
v_{\max} &= \ 169.2 \text{ V}, \\
v_{\min} &= -168.3 \text{ V}, \\
i_{\max} &= \ 0.45 \text{ A}, \\
i_{\min} &= -0.44 \text{ A}, \\
v_0 &= 0.45 \text{ V}, \\
i_0 &= 0.05 \text{ A}, \\
\Delta v &= 21.1 \text{ V}, \\
\Delta i &= 0.05 \text{ A}.
\end{aligned}
$$

which can also observed in Figure 3.5(a) and are labeled in Figure 3.5(c).

To illustrate the position value $\left(v_0 + \Delta v \cdot (x - N),\ i_0 + \Delta i \cdot (y - N)\right)$ and the binary model value $B_{x,y}$ of a cell $\left(x^{\text{th}},\ y^{\text{th}}\right)$, the ($12^{\text{th}}$, 1^{st}) cell marked by the star in Figure 3.5 is selected as an example. It has the positional value (0.45+4 × 21.1, 0.05+4× 0.05)=(84.85, 0.25) and the binary model value 0 (i.e., not occupied).

3.4.3 Application of proposed mapping algorithm to collected data

In order to validate the techniques proposed in this dissertation, a lab environment has been setup with a set of data acquisition device, with details shown in Appendix A. Over 200 real-world data files of over 50 PELs have been collected to form a database. A set of

46 representative V-I trajectories from this database and their mapped cell grids using the proposed algorithm above are shown in Appendix B.

3.4.4 Features extracted from the binary cell grid

Besides reducing the variance by the difference between V-I trajectories of PELs within the same load category, the mapping of V-I trajectories onto binary cell grids can also reduce the effect of distortion but preserve the graphical characteristics. For each category of PELs, a novel set of signatures that can be directly identified from the binary cell grid according to two key cells and three key lines, as shown in Figure 3.6. Note that compared with Figure 3.5, Figure 3.6 uses the same set of voltage and current measurements but the parameter N is now set to equal to 12 instead of 8 to illustrate the values of certain features.

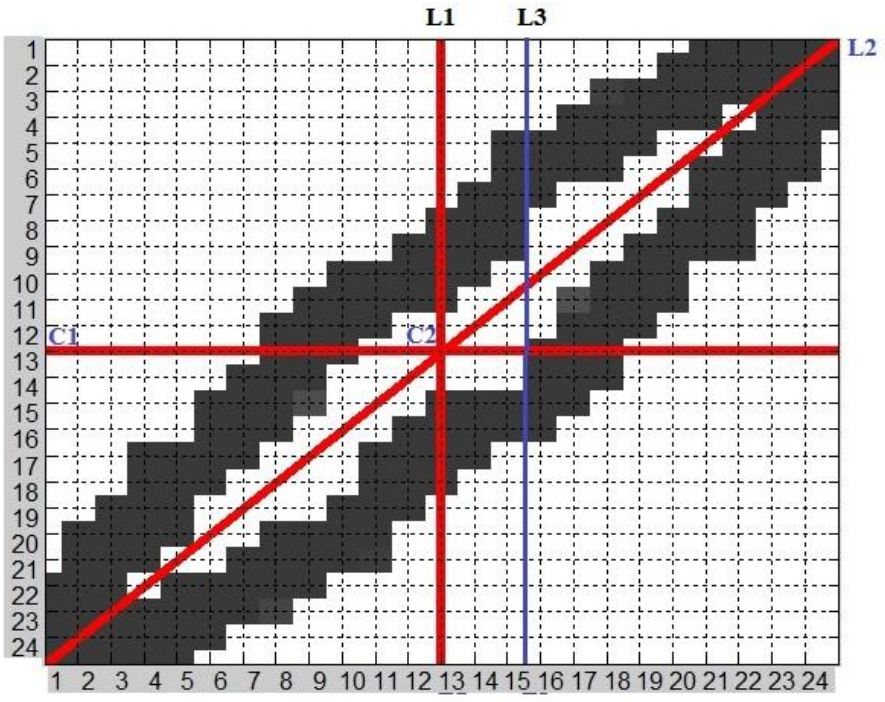


Figure 3.6 Two key cells (C1 and C2) and three key lines (L1-L3) in a cell grid

The following set of signatures is proposed to represent each load category:

(1) The binary value of the left horizontal cell $(1, N)$, marked as cell "C1". For instance, in Figure 3.6 the binary model value $B_{1,12}$ of cell C1, i.e., cell (1, 12), is 0.

(2) The binary value of the central cell (N, N), marked as cell "C2". For instance, in Figure 3.6 the binary model value $B_{12,12}$ of cell C2, i.e., cell (12, 12), is 1.

(3) The multiplication of the binary values of all anti-diagonal grid cells, i.e., all cells traversed by line "L2". This value is also a binary value and indicates whether the V-I trajectory is linear. For instance, in Figure 3.6 multiplication of the binary values of all anti-diagonal grid cells equals to 0 as some anti-diagonal cells are not occupied (and thus are of 0 value).

(4) The number of *continuum*s of occupied cells (i.e., with binary value 1) within all cells $(N, [1:2N])$, which indicates the number of intersections of the V-I trajectory and the base voltage v_0 line (marked as line "L1").

Note that the notation $[1:2N]$ denotes all integers from 1 to $2N$ and a continuum of occupied cells means a set of cells that are occupied and adjacent to one another in the grid, as shown in the following figure. Also, the number of continuums of occupied cells, instead of the number of occupied cells, is counted to make the proposed PELs feature more robust to variations and measurement noise.

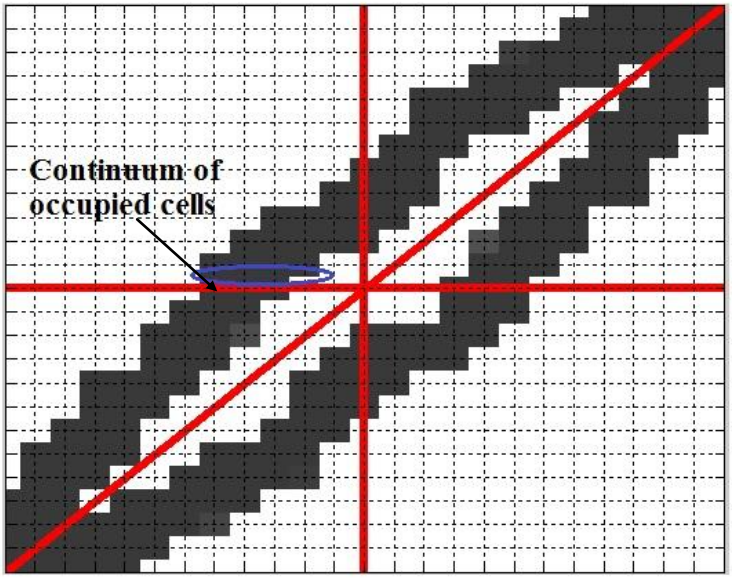


Figure 3.7 Example of a continuum of occupied cells in the grid

(5) Whether there exist any self-crossing intersections of the V-I trajectory itself. As shown in Figure 3.1(5), there are self-crossing intersections in the V-I trajectory for PELs from Category M. On the other hand, there are no there are self-crossing intersections in the V-I trajectory for PELs from Categories R and X as shown in Figures 3.1(1) and 3.1(2).

(6) The number of intersections of the V-I trajectory with the $v_0 + 0.2N \times \Delta v$ line (marked as line "L3").

This feature is mainly designed for PELs in Category PAC as their V-I trajectories are basically an anti-diagonal straight line with triangles as shown in the following figure. Note that the representation of the triangles in the V-I trajectories of PELs in Category PAC by this feature is not unique. There can be other similar features.

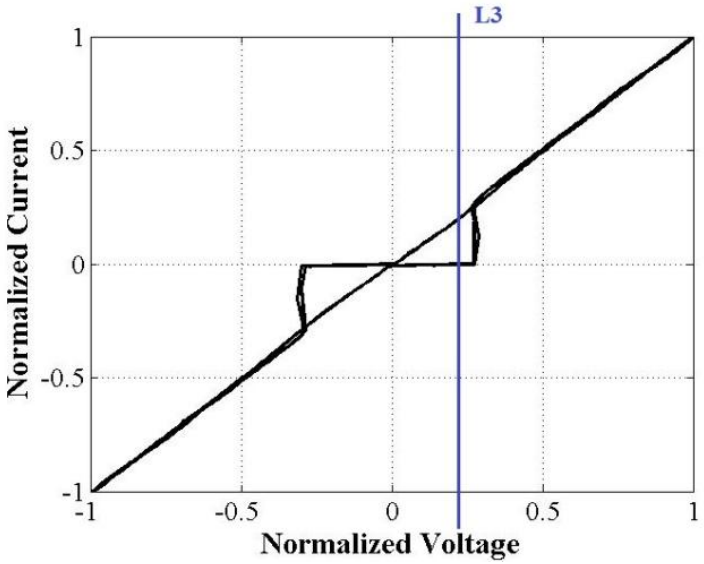


Figure 3.8 Illustration of graphical features for PELs from Category PAC

3.5 DETERMINING THE NUMBER OF SELF-CROSSING INTERSECTIONS

The V-I trajectories of some electric loads cross-intersect themselves, as shown in Figure 3.1(7). It is suggested in [44] that the number of self-crossing intersections contained by a V-I trajectory could be related to the order of harmonics. A simulated load with a significant 3rd (or 5^{th}) harmonic component in the current has two (or four) self-crossing intersections. However, it can also be caused by loads in Category M, i.e., loads with multiple independent front-end power supply units.

Therefore, this dissertation proposes a general but yet low-cost algorithm to determine the number of self-crossing intersections contained in a V-I trajectory, as shown in the following figure.

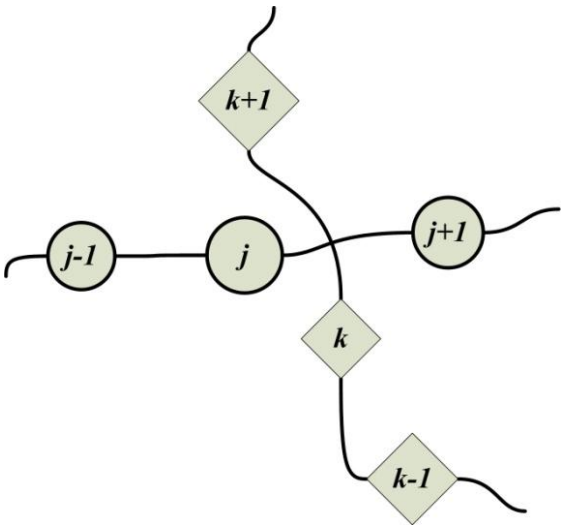


Figure 3.9 Determining the number of self-crossing intersections

(1) Read one-half cycle ($\frac{1}{120}$ seconds) of sampled data points [0-, 0+], starting with the zero-crossing data point from negative voltage values to positive voltage values (denoted by 0-) and ending with the zero-crossing data point from positive voltage values to negative voltage values (denoted by 0+);

(2) For every data point j within the region [0-, peak+] where peak+ denotes data point in [0-, 0+] with the maximal positive voltage value, find the data point *k* whose voltage value is closest to point *j*;

(3) Denoting a data point *j* with voltage value v_j and current value i_j by a vector $\underline{\underline{j}} = (v_j, i_j)$, check whether values of the sampled current sequence $\{\underline{\underline{j}}-1, \underline{\underline{j}}, \underline{\underline{j}}+1\}$ and $\{\underline{\underline{k}}-1, \underline{\underline{k}}, \underline{\underline{k}}+1\}$ are monotonically increasing. If yes, go to step (4); if not, repeat step (3) and start with $\underline{\underline{j}}+1$;

(4) Check whether data points $\underline{\boldsymbol{k}\text{-}1} = (v_{k-1}, i_{k-1})$ and $\underline{\boldsymbol{k}+1} = (v_{\boldsymbol{k}+1}, i_{\boldsymbol{k}+1})$ are on different sides of the line determined by $\underline{\boldsymbol{j}\text{-}1} = (v_{j-1}, i_{j-1})$ and $\underline{\boldsymbol{j}+1} = (v_{j+1}, i_{j+1})$ using the following criterion [79]:

$$\left\{\left(\underline{\boldsymbol{j}+1} - \underline{\boldsymbol{j}\text{-}1}\right) \times \left(\underline{\boldsymbol{j}+1} - \underline{\boldsymbol{k}\text{-}1}\right)\right\} \cdot \left\{\left(\underline{\boldsymbol{j}+1} - \underline{\boldsymbol{j}\text{-}1}\right) \times \left(\underline{\boldsymbol{j}+1} - \underline{\boldsymbol{k}+1}\right)\right\} < 0 \qquad (11)$$

where $\times$ denotes the cross product and $\cdot$ denotes the dot product of two vectors. In other words, for any j and k, an instance when (11) is satisfied is considered as a self-crossing intersection.

(5) End.

3.6 COMPUTATIONAL COMPLEXITY ANALYSIS FOR PROPOSED SIGNATURES

Most existing load signatures are extracted from the harmonic spectrums of the voltage and current waveforms, which require computing the discrete Fourier transformation (DFT) or the fast Fourier transformation as well as their inverse. Though many fast Fourier transform (FFT) algorithms [80, 81] have been proposed and implemented in practice to reduce the computational cost of computing the DFT [82, 83], computing the FFT may be still be too expensive in computing cost in many low-cost applications.

The graphical features for PELs proposed in section 3.3, i.e., from the binary mapping of V-I trajectories, can greatly reduce the computational cost but nevertheless achieve more than an average of 99% accuracy in field tests.

The following table compares the number of real multiplications and real additions needed by the proposed binary mapping method, with several widely used DFT and FFT algorithms.

Table 3.1 Comparison of the number of real multiplications and real additions needed by different algorithms

Algorithms	Goertzel [84]	Radix-2 FFT [85]	Proposed binary mapping
	N-point data M DFT terms	N-point data	N-point data Any size cell grid
Complexity	$O(NM)$	$O(N\log_2 N)$	$O(N)$
Real multiplications	$(2N+4)M$	$2N\log_2 N$	~4N
Real additions	$(4N+4)M$	$2N\log_2 N$	~8N

Note that determining the minimum, maximum, and mean values in equation (9) does not require much computational effort as data points are proposed in this dissertation to be taken into computation half-cycle by half-cycle (corresponding to each electric cycle of 1/60 seconds in the U.S.) and within each half-cycle the data points are ordered by the magnitude of their voltage values.

Furthermore, for the mapping to the binary grid, only one half-cycle data points are necessary, and each is compared with adjacent unoccupied cells in one direction, as shown in the following figure.

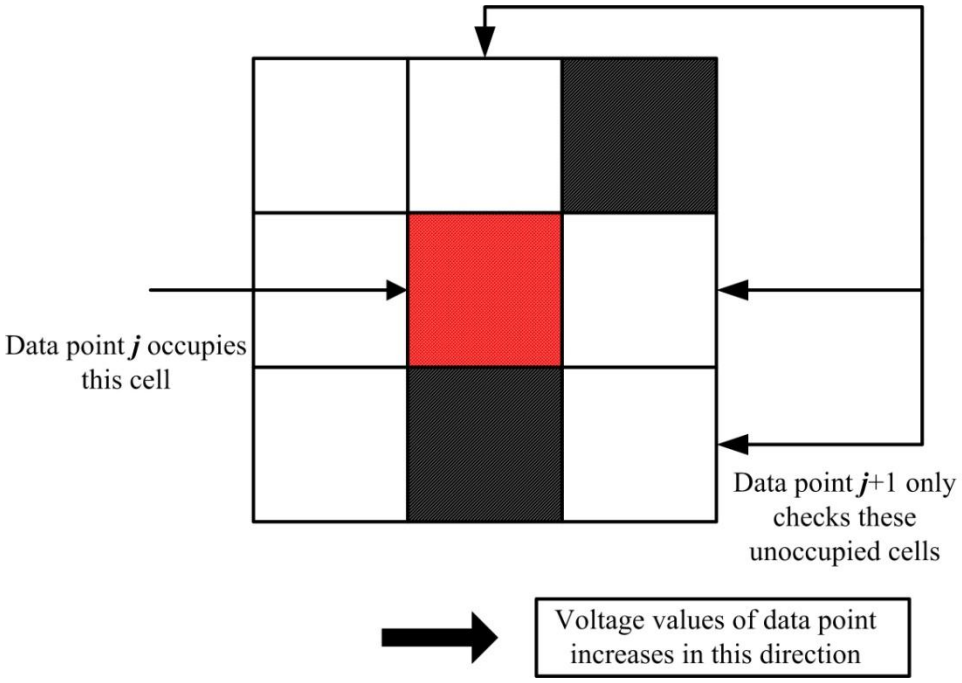


Figure 3.10 Checking adjacent unoccupied cells in one direction

Note that in Figure 3.10 filled cells denote occupied cells with binary model value $B=1$, while unfilled cells denote unoccupied cells with $B=0$. In this case, each data point may compare with 0 to 5 adjacent cells as the algorithm searches in one direction. A rough estimate of the average numbers of cells that each data point checks would be 4 cells. Furthermore, each such comparison needs 2 real multiplications and 4 real additions, which are independent of the grid size. In other words, the size of the cell grid has no impact on the complexity of the proposed mapping algorithm.

In the extreme case of only computing the fundamental I_1 and V_1 as well as the 3rd and 5th harmonics in current (I_3 and I_5), i.e., M=4, the proposed algorithm still has computational advantages over Goertzel's algorithm [84]. In typical cases that a fuller

harmonic spectrum is present, the proposed algorithm needs the same number of multiplications and additions when N=4. However, in practical cases N is much greater than 4 and thus the proposed set of features from V-I trajectories are expected to save a large amount of computation efforts.

3.7 EXPECTED VALUES OF PROPOSED FEATURES FOR ALL PELS CATEGORIES

The proposed graphical features from binary mapping of V-I trajectories for the 7 categories of PELs are expected to have values in the following Table 3.2. It can be observed that the proposed set of features is distinct for each PEL category. Test results of the performance of the proposed set of features using real-world data are presented in later chapters of this dissertation, after the introduction of the supervised Self-Organizing Map (SSOM) in Chapter Four.

Table 3.2 Expected signatures of all 7 PELs categories

Category	Feature 1 value of left horizontal cell	Feature 2 value of central cell	Feature 3 multiplication of anti-diagonal cells	Feature 4 number of continuums of occupied cells	Feature 5 self-crossing intersections	Feature 6 number of intersections with line L3
R	0	1	1	1	0	1
X	0	0	0	2	0	2
NP	1	1	0	1	0	1
P	0	1	X	1	0	1
M	0	X	0	2	1 or more	2
T	0	0	0	2	0	2
PAC	X	1	1	1	0	2

where “X” means either 0 or 1.

3.8 CHAPTER SUMMARY

This chapter proposes a low computational-cost but yet accurate algorithm for signature extraction from the voltage and current waveforms of PELs for classification and identification. Instead of utilizing digital signal processing and frequency domain analysis, this dissertation abstracts the similarity of voltage-current (V-I) trajectories between loads and proposes to map the V-I trajectories to a cell grid with binary cell values. Graphical signatures can then be extracted for many purposes. The proposed method significantly reduces the computational cost compared to existing frequency-domain signature extraction and analysis methods. It is shown in Chapter Four that an average of over 99% of success rate can be achieved using the proposed features.

The proposed graphical features from mapped cell grid are designed for PEL classification. In other words, the proposed features can achieve an average of over 99% of accuracy to assign an unknown PEL into one of the seven PEL categories. However, these features cannot be used to distinguish PELs that are in the same category with similar characteristics. Other features and methods are discussed in Chapters Six and Seven to identify PELs in the same category.

CHAPTER 4 SELF-ORGANIZING CLASSIFICATION AND IDENTIFICATION OF PLUGGED-IN ELECTRIC LOADS

4.1 INTRODUCTION

The diversity within each type of PEL and the similarity between different types of PELs significantly complicate the identification of PELs as shown by examples in Chapter One. However, very few existing methods have addressed these challenges as shown in Chapter Two. This chapter proposes to apply a supervised version of the Self-Organizing Map (SOM) to classify and identify PELs. The SOM, invented by Kohonen [86], is an unsupervised artificial neural network (ANN) mapping from a high-dimensional data space to a low-dimensional neuron grid while preserving statistical and topological information.

A major advantage of the SOM is that it classifies all training data into several groups by their inherent relationships, known as "clustering by nature". Therefore, for a large number of different PELs of different brands and models, an SOM can divide them into several groups such that PELs with a similar power supply and features are clustered into the same group. For the purpose of load identification, the basic SOM is extended to a supervised SOM (SSOM) in this dissertation to function as a classifier.

4.2 SELF-ORGANIZING MAPS

The SOM is an unsupervised ANN trained by competitive learning. Its output is a low-dimensional (typically two-dimensional), discretized grid of neurons with similar characteristics as the input data. Necessary notations and concepts are introduced as follows.

4.2.1 Neurons and their assigned values

Consider an SOM with K neurons and a set of training data consisting of l-dimensional vectors $\underline{\boldsymbol{x}}_q = [x_{q1}, x_{q2}, x_{q3}, \ldots, x_{ql}]$. Each neuron n_i is assigned with [86] the following values:

(1) a time-invariant topological position (i.e., an two-dimensional coordinate in the output grid);

(2) a time-varying parametric weight (also called *reference*, *model* or *codebook*) vector $\underline{\boldsymbol{m}}_i = [m_{i1}, m_{i2}, m_{i3}, \ldots, m_{il}]$ of the same dimension as the input data;

(3) a predefined function which defines a neighborhood (e.g., a circle or a square in two-dimensions) centered at the neuron.

All neurons compete to respond to the input but only one neuron wins at each time in the sense of minimum distance. Denoted by the subscript c, the winning neuron is called the *Best Matching Unit* (BMU):

$$c = \arg\min_i \left\{ \left\| \underline{\boldsymbol{x}}_q - \underline{\boldsymbol{m}}_i \right\| \right\} \tag{12}$$

where $\|\cdot\|$ is a distance function. In many applications, the Euclidean distance function d is used to measure closeness.

$$\mathrm{d}_E(\underline{\boldsymbol{m}}_i, \underline{\boldsymbol{x}}_q) = \sqrt{\sum_{k=1}^{l} (m_{ik} - x_{qk})^2}. \tag{13}$$

Furthermore, there are two types of training algorithms for the SOM: *sequential* and *batch*. Sequential training takes one single input vector at each training step and updates the weights, whereas the batch training presents all input data vectors to the neuron grid before any updates are made at each training step.

The processes of both sequential training and batch training are provided as follows.

4.2.2 Sequential training algorithm

The basic sequential SOM training algorithm is defined as follows [86]:

(1) All neurons' weight vectors m_i, where $i=1,2,\ldots, K$, are initialized;

(2) A training data vector is chosen randomly from the training data and presented to all the neurons;

(3) Find the Best Matching Unit (BMU) c to this input training vector;

(4) Calculate the radius of the neighborhood $N_c(t)$ of the BMU c according to the neighborhood function $h_{ci}(t)$, where i denotes any other neuron than the BMU. Also, $h_{c,i}(t)$ is usually chosen as a function of the distance between r_c and r_i such that $h_{c,i}(t)\rightarrow 0$ when $t\rightarrow\infty$.. For example, $h_{c,i}(t)$ can be in Gaussian form:

$$h_{c,i}(t) = \alpha(t)\cdot \exp\left(-\frac{\left\|r_c - r_i\right\|^2}{2\sigma(t)}\right) \tag{14}$$

where $\alpha(t)$ is a scalar *learning rate factor* that is monotonically decreasing with time, which is usually set to a high value early in the training to produce a rough training phase. $\sigma(t)$ defines the width of the kernel corresponding to the radius of the neighborhood $N_c(t)$.

(5) All neuron that has a Euclidean distance to the BMU less than the neighborhood radius update their weights. The closer a neuron is to the BMU, the more its weights get altered during the training process. That is,

$$m_i(t+1) = m_i(t) + h_{c,i}(t)[x(t) - m_i(t)] \tag{15}$$

where $t = 0, 1, 2, \ldots$ is an integer denoting the discrete time index.

(6) Return to step 2 and repeat for N iterations where N is the total number of input training vectors presented to the SOM. N may exceed the number of data vectors in the data base, in which case the training vector is each time selected randomly from the data vector base. Past experience shows that for good performance N should be at least 500 times the number of neurons.

Remark: one important property of the SOM is that it provides a topology preserving mapping from the input space to the two-dimensional neuron grid. This means that data points that are close or share many common features in the input space are mapped to neurons that are positioned close to one another to form a so-called *cluster*.

The SOM therefore converts complex, nonlinear statistical relationships between high-dimensional data items into simple geometric relationships on a two-dimensional grid. As it thereby compresses information while preserving the most important topological and metric relationships, the SOM can also be considered to produce some degree of abstractions.

4.2.3 Batch training algorithm

Batch training is an improved version of the above sequential training algorithm. Instead of process input data vector one-by-one in a sequential manner, batch training is significantly faster, especially with MATLAB [86]. In this dissertation, the batch training is adopted using the MATLAB SOM toolbox [87]. Details on how to use this toolbox are provided in Appendix C.

The batch SOM training algorithm proceeds as follows:

(1) Initialize weight vectors $\underline{\boldsymbol{m}}_i$;

(2) Partition the input data set into the *Voronoi* regions [86] of the weight vectors, i.e., each input vector $\underline{\boldsymbol{x}}_q$ belongs to the region of its closest neuron n_i;

(3) Update $\underline{\boldsymbol{m}}_i$ according to

$$\underline{\boldsymbol{m}}_i(t+1) = \frac{\sum_{q=1}^{N} h_{q,c}(t)\underline{\boldsymbol{x}}_q}{\sum_{q=1}^{K} h_{q,c}(t)} \tag{16}$$

(4) where c is the BMU of the input vector $\underline{\boldsymbol{x}}_q$, $h_{q,c}(t)$ is the neighborhood function, and K is the number of neurons [86].

(5) Return to step (2) and repeat a number of times. ■

Note that data vectors with similar values in the input space are mapped to neurons that are positioned close to one another and thus form a cluster. Thus, different brands and models of a type of PEL are mapped into the same cluster. Different types of PELs are expected to be mapped into different clusters.

4.2.4 Representations of the SOM

The unified distance matrix (U-matrix) [88] is widely used to represent an SOM. The U-matrix illustrates weight vectors by showing the distances between adjacent neurons. For each neuron, the distance between itself and its adjacent neurons (the number of which depends on its neighborhood topology) is calculated and presented with different colorings or with a gray scale image. Light colors depict the closely spaced neurons and darker colors indicate the more distant neurons. Groups of light colors can be roughly considered as clusters, with dark parts considered as the boundary regions.

As an example, the feature vectors of ten PELs from three PEL categories, including two set-top boxes (STBs), one DVD player, one space heater, one plasma TV, one LCD TV, one LED TV, one LCD monitor, one laptop computer, and one desktop computer, are presented to train a 50-by-50 SOM for classification. The corresponding U-matrix is shown in the following figure.

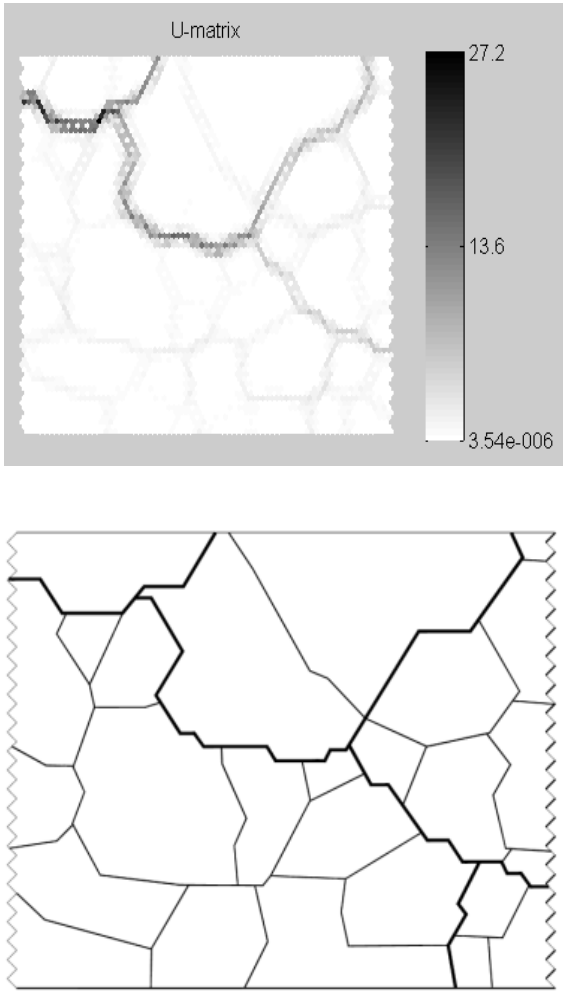


Figure 4.1 The U-matrix visualizes distances between neighboring neurons, and helps to show the cluster structure.

These ten PELs are classified into four clusters. However, the U-matrix does not tell which PEL falls into which cluster, which is mainly because the training data is not labeled and the training is unsupervised. Other tools, such as the labeled neuron grid, are usually used together with the U-matrix to provide extra information.

4.3 SUPERVISED SOM (SSOM) FOR PELS CLASSIFICATION AND IDENTIFICATION

In this dissertation, the basic SOM is extended to a supervised SOM for PELs identification. The SSOM utilizes the SOM's capability of clustering by similarities, preserves statistical structure, and uses the trained map to do identification.

4.3.1 Supervised SOM

The SOM introduced so far is unsupervised as the training vectors are not labeled and no class identity information is attached or used during the training. Such unsupervised SOMs are not intended for identification.

Assume that there are M known classes, $\omega_1, \omega_2, \ldots, \omega_M$, and each input feature vector $\underline{\boldsymbol{x}}_q$ is pre-assigned to one of the classes. Each $\underline{\boldsymbol{x}}_q$ remains unchanged in its numerical values but labeled by a string containing its pre-assigned class identity. Only the numerical values of $\underline{\boldsymbol{x}}_q$ are used in the training, and the BMU for each $\underline{\boldsymbol{x}}_q$ is labeled the same as $\underline{\boldsymbol{x}}_q$. When the training is finished, neurons that have become the BMU to one or more input vectors are classified into one of the classes by a *voting mechanism*. That is, if a neuron has been the BMU to multiple input classes (each for probably multiple times), then it is classified into the class for which it has been the BMU for the greatest number of times. Neurons that have never been a BMU to any input vector are marked as "*unclassified*" after the training has been completed.

After labeling the training data and having completed the training, the SOM can be turned into a classifier or predictor by partitioning the grid neurons into clusters. Revisit the above example in section 4.2.3 with the PELs labeled as

(1) two set-top boxes and 1 DVD player: labeled by "1";

(2) one Space heater: labeled by "2";

(3) one plasma TV: labeled by "3";

(4) one LCD TV, one LED TV, one LCD monitor, one laptop computer, and one desktop computer: labeled by "4",

and the labeled neuron grid of the trained SOM is shown in the following figure.

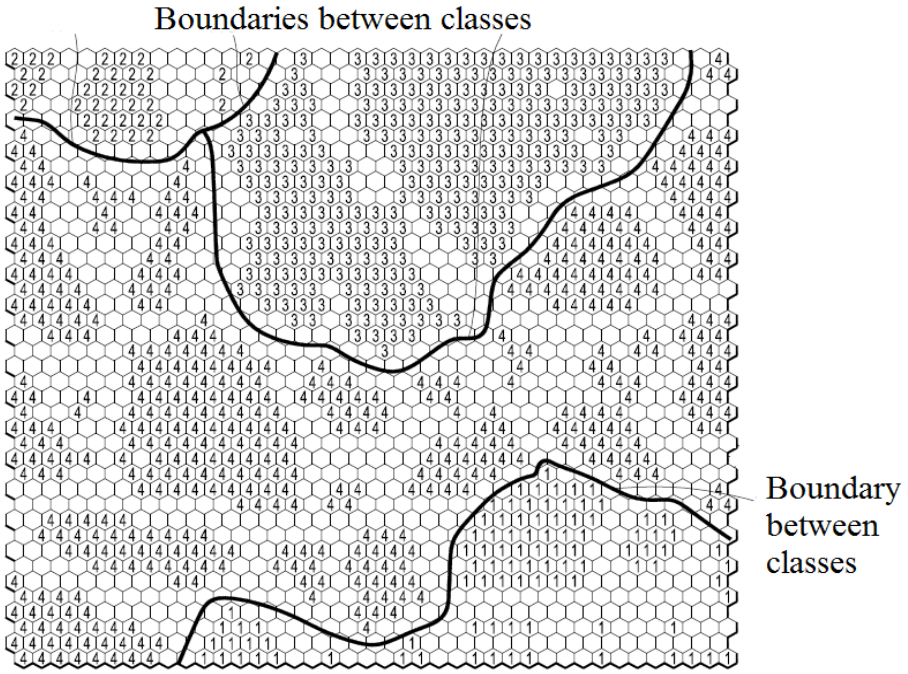


Figure 4.2 Labeled neuron grid of the SOM representing ten PELs.

In Figrue 4.2, solid lines represent boundaries shown in Figure 4.1. Note that the observable clusters in the labeled neuron grid match the observable boundaries in the U-matrix in Figure 4.1. Moreover, the labeled neuron grid or the U-matrix gets very crowded and the boundaries are difficult (if not impossible) to observe with large numbers of PELs, classes, or types. However, it does not affect the PELs identification as these representations are just for illustration and the identification is carried out using the trained weight vectors of the neurons.

4.3.2 SSOM for PELs classification and identification

After training, there are usually some neurons which have never been BMUs to any inputs and thus only a subset of $K_s < K$ neurons are classified by the voting mechanism and classified into one of the pre-defined known classes. When an unknown feature vector $\underline{\boldsymbol{x}}$ is presented to the SOM, $\underline{\boldsymbol{x}}$ is compared to all K_s classified neurons and is classified to be within that same class as the neuron with the minimum distance to x in the vector space. The Euclidean distance is used in most applications of the SOM.

The performance of the SSOM identifier can be verified by *cross validation*. The input data set is divided into two subsets, performing the classification on one subset (called the training set) and validating the identification on the other subset (called the testing set). Both subsets are labeled by class identity labels. A correct identification is said to be made if the actual class identity label of a testing feature vector matches the class identity label assigned by the SSOM identifier. Otherwise, it is an incorrect identification. The success rate is then defined as

$$success\ rate = \frac{\text{number of correct identifications}}{\text{number of testing vectors}} \tag{17}$$

The cross-validation framework is shown in the following figure.

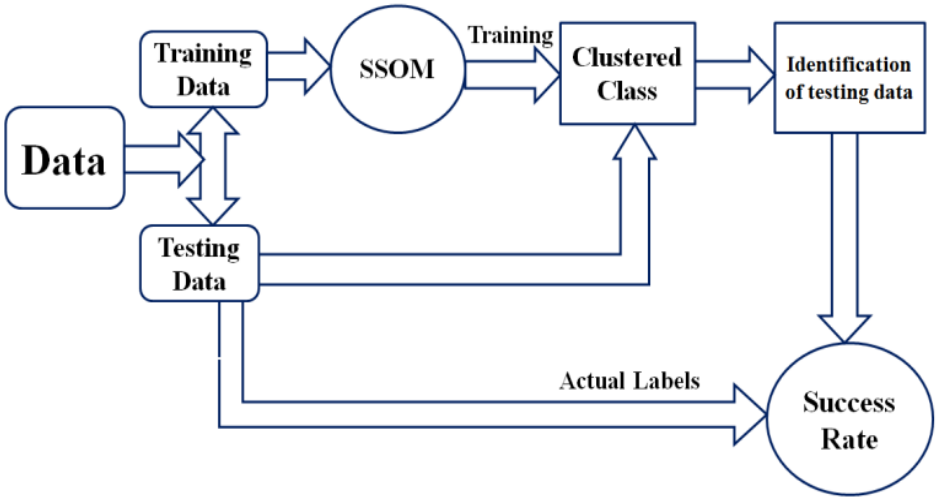


Figure 4.3 The cross-validation framework of the SSOM identifier

4.3.3 Discussion on performance of the SSOM identifier

To summarize, the SSOM identifier is first trained by collected PELs data. Different brands and models of each type of PELs are classified together to form clusters. The SSOM identifier does not only achieve high accuracy when handling known PEL models but also performs well to identify unknown PEL models. An unknown model to the SSOM identifier falls into the cluster in which other models of this PEL type do. Therefore, the SSOM identifier extracts information contained in the large amount of training data and store simplified information in the trained neuron grid. For practical purposes, the trained neuron grid, instead of the training data set, is sufficient to be stored, updated, and used for identification. The only on-line computation required is to compute the feature vectors representing incoming unknown PELs.

The SSOM identifier presented so far can only identify PELs to a certain granularity because different types of PELs with similar features could also be clustered together. For example, in Figure 4.2, LCD TVs, LED TVs, and LCD monitors are labeled the same since most flat-panel display devices are equipped with front-end power supply units of similar design topology. An incoming unknown PEL which falls into this cluster is identified as an LCD TV, LED TV, or LCD monitor. It is sometimes difficult or unnecessary to distinguish LCD TVs from LED TVs. More discussions on how to achieve a finer granularity level are provided in later chapters.

Furthermore, as shown in Figure 4.2, there are a number of neurons that have never been BMUs to any input and thus not labeled. During the test, similar to the training phase, there is a BMU to each testing feature vector. If some BMUs to the testing data are not labeled, identification decisions can be made based on those labeled BMUs. In the extreme case that all BMUs are not labeled, the unknown PEL would be identified as "*New*".

4.4 TESTS ON PERFORMANCE OF THE SSOM ON PELS CLASSIFICATION

In this section, results of several tests on performance of the SSOM on PELs classification and identification are presented. A lab environment has been established to collect data from available commercial PELs in a non-intrusive manner. The data is collected using a printed circuit board (PCB) sensor module with current and voltage sensors and a set of data acquisition devices. The circuit schematic of the PCB sensor module is given in Appendix A. It is designed with the capability of being placed as accessories as well as the purpose of introducing no influence on the operation of PELs. For accuracy and convenience, the sampling frequency is set to 30.72 kS/s, i.e., 1024

sampled data points per electricity cycle (1/60 seconds in U.S.). Lower sampling frequencies such as 7.68 kS/sec and 3.84 kS/sec are also tested and the cross validation results of the SSOM identifier remain relatively the same.

4.4.1 PELs features used in tests

Besides the proposed graphical features from mapping of normalized V-I trajectories to cell grid as introduced in Chapter Three, another set of PEL features from harmonic analysis is also considered for the purpose of completeness and comparison. The proposed graphical features achieve an average of over 99% accuracy on PELs classification, but may need to be combined with other features for PELs identification. With the voltage and current waveforms given by equations (1) and (2) in Chapter Two, the following features can be utilized from harmonic analysis:

(1) The RMS current I_{RMS};

(2) The total harmonic distortion (THD) in current THD_I as defined by Equation (3);

(3) The power factor *pf* as defined by Equation (6);

(4) The crest factor as defined by Equation (7) ;

(5) The 3rd and 5th harmonics (amplitude and phase) in current;

(6) Per-cycle difference in I_{RMS}

$$DIF_{I_{RMS}}(T) = I_{RMS}(T) - I_{RMS}(T-1); \tag{18}$$

(7) Mean square error (MSE) in I_{RMS} defined to be

$$MSE_{I_{RMS}}(T) = \left(Trace\left[DIF_{I_{RMS}}(T)^{T} * DIF_{I_{RMS}}(T) \right] \right)^{\frac{1}{2}}. \quad (19)$$

where T denotes the index of cycles of voltage and current waveforms (i.e., each cycle is 1/60 seconds in the U.S.).

The per-cycle difference $DIF_{I_{RMS}}(T)$ in I_{RMS} (as well as in other features) is equivalent to the first-order rate of change, and the MSE in I_{RMS} (as well as in other features) is equivalent to the second-order rate of change. The $MSE_{I_{RMS}}$ can be used to distinguish PELs have similar characteristics, as shown in the following example.

Based on a state-of-the-art study, DVD players and set-top boxes (STBs) have similar features listed as follows:

(1) A small power factor around 0.5.

(2) The THD in the current varies between 1.5-1.8.

(3) Similar power consumption between 25 to 50 W.

(4) A high crest factor: around 3.5.

(5) A very high ratio of 3rd harmonic amplitude compared to the fundamental: close to 1.

(6) A very high ratio of 5rd harmonic amplitude compared to the fundamental: around 0.8-0.9.

In other words, conventional features in literature cannot distinguish DVD players from STBs without ambiguity. However, the $MSE_{I_{RMS}}$ values of DVD players and STBs are different as shown in Figure. The $MSE_{I_{RMS}}$ of STB remains close to 0.1 without any

sudden spikes. On the other hand, for DVD players, spikes can be observed and detected in the $MSE_{I_{RMS}}$.

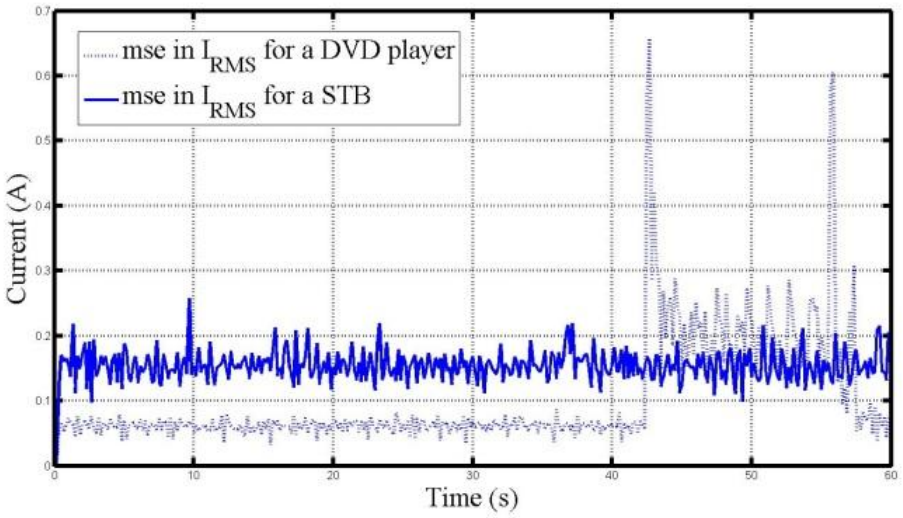


Figure 4.4 The comparison of $MSE_{I_{RMS}}$ of a DVD player and a STB.

Furthermore, features (6) and (7) in the above definition can be extended to any other features (1)-(5), such as, DIF_{THD_I} , $DIF_{3rd\ in\ current}$, and $DIF_{5th\ in\ current}$.

Note that insufficient number of features reduces the performance of the SSOM, but too many features also imply redundancy in information and extra computational cost. The above set of features was selected for best performance after many tests.

Also, as the selected features have values of different ranges and the distance function used in forming the SOM clusters is Euclidean, features with large values may have a larger influence in the distance/difference function than features with small values, although this does not necessarily reflect their respective significance in the design of the

classifier. However, tests show that normalization of data significantly reduces the PELs identification accuracy. It is possibly because of the strong representation of certain PELs by several features. Therefore, in this dissertation all features are not normalized

4.4.2 Test on the necessary amount of data to train SSOM

One of the many advantages that the SSOM possesses is that it requires only a small amount of data to train compared with other ANNs [86], such as the Support Vector Machine (SVM) and Radial Basis Function (RBF) networks.

Several scenarios are considered to investigate the amount of data necessary to train the SSOM. Test results of three typical PELs and different choices are shown in Test Success Rates of the SSOM Identifier. For all the tests presented in this Chapter, 512-point FFTs are employed to calculate the features as there are 1024 sampled data points per cycle. Also, the total amount of feature vectors (for training and testing) for each tested PEL is 3600, as each data file is of time duration 60 seconds (i.e., 3600 cycles). Each row represents a different set of tests with different percentage of data used for training and testing. For example, 5%/95% means 5% of the available data is used for training and the remaining 95% is used for testing.

Table 4.1 Test Success Rates of the SSOM Identifier

	DVD	TV	Fan	Overall
5% / 95%	49.36%	81.99%	97.34%	76.23%
10% / 90%	91.30%	87.81%	97.69%	92.26%
20% / 80%	96.74%	94.03%	93.92%	94.90%
67% / 33%	99.75%	99.83%	98.00%	99.19%

Table 4.1 shows that with sufficient training data, the SSOM identifier can reach a 99% success rate for PELs known to the SSOM. i.e., PELs whose data files have been used for training the SSOM.

4.4.3 Performance of the proposed graphical features

In this test, a 50-by-50 SSOM is first trained with a database of a total of 627 sets of real-world PELs data. For each set of PELs data, its steady-state current and voltage waveforms for 60 seconds are stored and converted to a set of graphical features as proposed in Chapter Three. For accuracy tests, a total of 75 sets of PELs (of 23 types) steady-state current and voltage waveforms are used, converted to the proposed graphical features, and tested. For each testing data set, a total number of between 900 to 3000 V-I trajectories are selected and mapped to a 64-by-64 cell grid.

Two tests of different scenarios are carried out and the results are shown in the following tables.

Table 4.2 Tests on the performance of the proposed graphical features

Target Load Category	Load Type	Total Models	Total Number of Tests	Success Rate (%)
NP	Battery Charger	1	3000	83.4
	DVD Player	4	3000	100
	Desktop Computer	2	3000	99.8
	LCD Monitor	7	3000	99.5
	Printer	1	3000	99.9
	Electronic Circuit Board	1	3000	98.7

Table 4.2 continued

P	LCD TV	8	3000	98.5
	LED TV	3	3000	99.2
	Plasma TV	2	3000	99
	Multi Function Device	3	3000	93
	Projector	4	3000	99.9
M	Microwave Oven	4	1800	99
R	Space Heater	4	1800	93
	Coffee Maker	2	1800	98
	Incandescent Lamp	4	1800	99.2
	Electric Skillet	2	1800	98.6
T	Stapler	1	1800	98.9
	Adapters	5	1800	100
X	Fan	5	3600	98.5
	Refrigerator	4	3600	100
	Water Dispenser	1	3600	100
	Shredder	2	3600	65
PAC	Incandescent Lamp with Dimmer	1	1800	50

Table 4.2 validates that the proposed graphical features by binary mapping of V-I trajectories in Chapter Three used with the SSOM classifier can achieve an average of over a 90% success rate with a relatively large load set and with seven PELs categories.

Furthermore, the major cases with low success rates are from some PELs in Category PAC; such PELs are mistakenly categorized as being in the Category R due to the similarity between their normalized V-I trajectories. Increasing the sampling rate may

help to improve the performance. Meanwhile, from the application point of view if an incandescent lamp with dimmer with a relatively small phase angle is identified as a resistive load, then the resulting classification is still acceptable.

The following table shows the results of another test to validate the performance of of the proposed graphical features in Chapter Three compared with conventional features from harmonic analysis. The same database from previous example is used to train a 50-by-50 SSOM and a total of 30 sets of PELs (of 13 types) are tested. For each testing data set, 100 V-I trajectories are selected and mapped to a 256-by-256 cell grid.

This test aims at comparing the performance of the SSOM used with proposed graphical features as well as the set of selected features defined in section 4.4.1. The comparison is shown in the following table.

Table 4.3 Comparison of proposed graphical features and conventional features

Test Loads	Total Number of Loads	Total Number of Tests	Success Rate by Graphical Feature	Success Rate by Conventional Feature
LCD TV	3	300	100%	100%
Set-top Box	4	400	100%	100%
Space Heater	2	200	99%	96%
Laptop	2	200	99%	98.7%
LED TV	2	200	99%	100%
Microwave Oven	4	400	92%	91%
Desktop	2	200	97%	47%
Bluetooth Charger	1	100	99%	66.4%
Multi-Function Device	3	300	100%	76.8%
LED Light	1	100	99%	96%

Table 4.3 shows that the proposed graphical PELs features achieved greater accuracy on PELs classification (i.e., into one of the PEL categories) compared to conventional features from harmonic analysis that are designed for general electric loads.

4.5 CHAPTER SUMMARY

This chapter has explained the extension of the standard SOM to a supervised SOM to function as a classifier for PELs. With large amounts of real-world data, the supervised SOM (SSOM) identifier first classifies a large number of PELs into several clusters. An classification decision is made to locate the neuron which has the closest weights to an input feature vector within the trained neuron grid. The SSOM identifier has the advantage that it achieves high success rate in different cases:

(1) If the testing PELs have been used for training the SSOM, the SSOM can achieve almost 100% success rate;

(2) If the testing PELs have not been used for training but other similar PELs have, the SSOM can achieve an average of over 95% success rate.

Moreover, the SSOM is robust to variance in features and requires limited online computational effort.

Furthermore, the next chapter extends the SSOM framework presented in this chapter and combines it with the Bayesian identification framework. The proposed framework can indicate the probability of an unknown PEL belonging to each category as well as each type.

Note that the SSOM is originally designed as a classification tool. Therefore, with selected sets of features, the SSOM classifier can achieve an average of over 95% success rate to assign an unknown PEL into one of the seven categories proposed in Chapter

Three. However, if directly applied to identification of PELs, in some cases the SSOM classifier can only achieve a success rate lower than 60% as PELs with similar characteristic fall into the same cluster and thus stay closely in the trained SSOM. Other features and methods are discussed in Chapters Six and Seven to identify PELs with similar characteristic.

CHAPTER 5 PROBABILISTIC IDENTIFICATION OF PLUGGED-IN ELECTRIC LOADS

5.1 INTRODUCTION

In the proposed PELs classification and identification framework, the identification decision made by the SSOM identifier discussed so far is *hard* or *absolute*. In other words, an unknown PEL is assigned into one (and only one) of the known classes. As discussed in Chapter One, the electrical signatures that represent PELs from the same class may have statistical variations. Therefore, an absolute decision is sometimes not a desired result as errors often exist and no classifier can give 100% success rate.

Instead, a *soft* or *probabilistic* decision may be desired which indicates the probability of an unknown PEL belonging to a particular type. A hard decision can be made based on the probabilities if needed. That is, the task now is to design PEL identifiers that classifies an unknown PEL "in the most probable of classes" [89].

This dissertation proposes to combine the SSOM with the Bayesian framework to estimate the *a posterior* probabilities by utilizing some information which is generally ignored in literature during the training of the SSOM. In the identification step, feature vectors of an unknown PEL are provided to the proposed identifier to get the probability of the unknown load belonging to a specific class of loads.

5.2 OVERVIEW OF BAYESIAN DECISION THEORY AND PROBABILITY ESTIMATION

The Bayesian decision theory is a statistical framework which takes variation of the patterns as well as costs and risks into consideration and then classifies an unknown pattern in the most probable sense. Specifically, given a classification task of M classes,

$\omega_1, \omega_2, \ldots, \omega_M$, and an unknown PEL which is represented by a feature vector $\boldsymbol{x}$, the Bayes classifier assigns $\boldsymbol{x}$ to be class ω_j by maximizing the conditional *a posteriori* probabilities $\Pr(\omega_i \mid \boldsymbol{x})$ [89, 90] for all *i*. That is,

$$\omega_j = \arg\max_i \Pr(\omega_i \mid \boldsymbol{x}), i = 1, 2, ..., M, \tag{20}$$

in which each $\Pr(\omega_i \mid \boldsymbol{x})$ represents the probability of the unknown PEL falls into each respective class ω_i, given that the observed feature vector of the corresponding PEL takes the value $\boldsymbol{x}$. Therefore, this type of Bayesian classifier is usually called a *maximum a posteriori (MAP)* classifier.

Therefore, the problem of PEL classification and identification is converted to calculate the conditional probabilities. Unfortunately, the true values of $\Pr(\omega_j \mid X)$ are generally inaccessible [91] and thus are usually estimated from available feature vectors corresponding to the database of known PELs by methods such as statistical learning [92, 93] and ANNs [94, 95].

One of the most commonly encountered probability density functions in practice is the Gaussian or normal density function. The major reasons for its popularity are its computational tractability and the fact that it models adequately a large number of cases.

In general, different methods for probability estimation can be divided into three types [89]: *parametric*, *nonparametric*, and *semi-parametric*. In the parametric method the probability density function (pdf) is assumed to be of a standard form (e.g., Gaussian or uniform) and its parameters can then be estimated either using maximum likelihood estimation (MLE) [96] or other methods. The nonparametric methods, such as histogram and Parzen's windows [97], generally require large amounts of data and appropriate

parameters which are usually set by trial-and-error [89]. The semi parametric-methods, such as the mixture of known distributions [98], offer a tradeoff between the simple and limited parametric methods and the computational intensive nonparametric methods. For instance, the Gaussian mixture model (GMM) has been quite popular and widely used for speaker identification and face recognition for its simple but powerful setup [99]. The parameters of GMM can be estimated either using the gradient descent method or the expectation maximization (EM) algorithm [100].

Two scenarios, with an unknown PEL $\boldsymbol{X}$ represented by a single feature vector and by a set of feature vectors, are discussed as follows.

5.3 Probabilistic Identification of an Unknown PEL Represented by a Single Feature Vector

Given an unknown feature vector $\underline{\boldsymbol{x}}_q = [x_{q1}, x_{q2}, x_{q3}, \ldots, x_{ql}]$, the Bayes' theorem states that

$$\Pr\left(\omega_j \mid \underline{\boldsymbol{x}}_q\right) = \frac{p\left(\underline{\boldsymbol{x}}_q \mid \omega_j\right)\Pr\left(\omega_j\right)}{p\left(\underline{\boldsymbol{x}}_q\right)} \tag{21}$$

where $\Pr(\omega_j)$ denotes the *a priori* probability of class ω_j and $p(\underline{\boldsymbol{x}}_q)$ is the probability density function (pdf) of $\underline{\boldsymbol{x}}_q$. As data is digitally sampled and stored, it is not necessary to estimate a continuous pdf but rather the values of the pdf at given conditions, which is much faster and easier. That is,

$$\Pr\left(\omega_j \mid \underline{\boldsymbol{x}}_q\right) = \frac{\Pr\left(\underline{\boldsymbol{x}}_q \mid \omega_j\right)\Pr\left(\omega_j\right)}{\Pr\left(\underline{\boldsymbol{x}}_q\right)}. \tag{22}$$

The *a priori* probabilities $\Pr(\omega_j)$, $j = 1, 2, \ldots, M$, can be estimated from the training data. If the total number of available feature vectors for training is N and N_j of them belong to the class ω_j, then $\Pr(\omega_j) \approx N_j / N$, which is a widely used approximation in many applications [89]. For instance, if the training data set contains 18432 samples from class ω_1 (for example, DVD players) and 55296 samples from class ω_2 (for example, TVs), then $\Pr(\omega_1) \approx 0.25$ and $\Pr(\omega_2) \approx 0.75$.

If the actual numerical value of each individual $\Pr(\omega_i \mid \boldsymbol{x})$, is not required, $p(\boldsymbol{x})$ is not needed because it appears in every $\Pr(\omega_i \mid \boldsymbol{x})$, represented by (5.2), as the denominator. In other words, all $\Pr(\omega_i \mid \boldsymbol{x})$ have the same denominator, which can be normalized and thus not taken into account during the classification. In this case, the *maximum a posteriori* (MAP) classifier returns the same identification result, which is the class that gives the maximum *a posteriori* probability $\Pr(\omega_i \mid \boldsymbol{x})$. The relative relationships or ranking based on numerical value of all $\Pr(\omega_i \mid \boldsymbol{x})$ remain the same, with and without the normalization by $p(\boldsymbol{x})$.

On the other hand, even if the actual numerical value of an individual $\Pr(\omega_i \mid \boldsymbol{x})$ is required, the common denominator $p(\boldsymbol{x})$ can be calculated by (5.3), as shown in the following equation:

$$\Pr(\omega_i \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid \omega_i)\Pr(\omega_i)}{\sum_{i=1}^{M} p(\boldsymbol{x} \mid \omega_i)p(\omega_i)}. \tag{23}$$

If an individual $\Pr(\omega_j \mid \underline{\boldsymbol{x}}_q)$ is not required, $\Pr(\underline{\boldsymbol{x}}_q)$ can be ignored. Otherwise, $\Pr(\underline{\boldsymbol{x}}_q)$ is either still not needed as $\Pr(\omega_j \mid \underline{\boldsymbol{x}}_q)$ can be normalized or $\Pr(\underline{\boldsymbol{x}}_q)$ can be calculated by

$\Pr\left(\underline{\boldsymbol{x}}_q\right) = \sum_{j=1}^{M} \Pr\left(\underline{\boldsymbol{x}}_q \mid \omega_j\right)$. Therefore, $\Pr(\underline{\boldsymbol{x}}_q \mid \omega_j)$ is the only difficulty to estimate $\Pr\left(\omega_j \mid \underline{\boldsymbol{x}}_q\right)$. Note that this method should work with a sufficient amount of samples for adequate accuracy.

Therefore, $p(\boldsymbol{x}|\omega_i)$, the likelihood function of $\boldsymbol{x}$ with respect to ω_i, is the only difficulty in the MAP classifier. Note that when the feature vectors take only discrete values, the density function $p(\boldsymbol{x}|\omega_i)$ becomes probabilities and is usually denoted by $\Pr(\boldsymbol{x}|\omega_i)$. In other words, an accurate estimation of the underlying environmental pdf needs to be derived from the available data.

The probability $\Pr(\underline{\boldsymbol{x}}_q \mid \omega_j)$ can be estimated from the training data. For each class ω_j, suppose there are T_j feature vectors available for training and $Y_j(\underline{\boldsymbol{x}}_q)$ denotes the set of training feature vectors that are close to $\underline{\boldsymbol{x}}_q$

$$Y_j\left(\underline{\boldsymbol{x}}_q\right) = \{\underline{\boldsymbol{y}} \in \omega_j, \left\|\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}_q\right\| < \xi\} \tag{24}$$

where ξ is a pre-defined parameter describing the acceptable error in difference. Then

$$\Pr(\underline{\boldsymbol{x}}_q \mid \omega_j) \approx \frac{card\{Y_j\left(\underline{\boldsymbol{x}}_q\right)\}}{T_j} \tag{25}$$

where *card* denotes cardinality of a set.

The accuracy of (25) depends on factors such as T_j, the resolution of features, and the parameter ξ. With a higher resolution (e.g., the power factor is stored to be 0.9088 instead of 0.91), the number of possible $\boldsymbol{x}$ increases exponentially fast. For instance, suppose that a set of 13 features is adopted and every feature is normalized to be within the range of [0,1] with the resolution equal to 10^{-2}. In this case, there are 1026 possible values of $\boldsymbol{x}$.

Even this number can be significant reduced by a proper ξ, but may still too large to handle.

5.4 HYBRID SSOM / BAYESIAN CLASSIFIER

Compared to utilizing the statistical information contained in the labeled training data, $\Pr(\underline{\boldsymbol{x}}_q \mid \omega_j)$ can also be estimated using the statistical information contained in the SSOM neuron grid and the BMU history. As the SSOM has already extracted statistical information from the training data set during the training, the latter one greatly simplifies the estimation process while preserving the accuracy. In fact, there exists information in the training process of the SSOM that is generally ignored in literature.

Recall that one neuron is selected to be the BMU at each training step and labeled to be the same class as the incoming feature vector. When the training is complete, each neuron could have been the BMU to feature vectors from several different pre-known classes and thus labeled differently or have never been a winner. This observation contains rich and important information which can be utilized to estimate the conditional probabilities. The reason lies in the fact that the voting mechanism is applied when the training is completed to determine each neuron's final class label and thus the minorities are ignored. This dissertation proposes to include all information contained in the training process.

Assume that there are T feature vectors used in the training, the *activation* $A_j(n_i)$ [101] for a neuron n_i from the class ω_j is defined as

$$A_j(n_i) = card\left\{\underline{\boldsymbol{x}} \in \omega_j \mid n_i = \arg\min_i \left\{\left\|\underline{\boldsymbol{x}} - \underline{\boldsymbol{m}}_i\right\|\right\}\right\} \tag{26}$$

which denotes how many times this neuron n_i has been the BMU to training feature vectors from class ω_j. Then

$$\Pr(\underline{\boldsymbol{x}}_q \mid \omega_j) \approx \frac{A_j(c)}{\sum_{j2=1}^{M} A_{j2}(c)}, \tag{27}$$

where c is the BMU to $\underline{\boldsymbol{x}}_q$.

In the case that n_i has never been the BMU to class ω_j, $\Pr(\underline{\boldsymbol{x}}_q \mid \omega_j)$ is defined to be 0.

5.5 Estimation of Conditional Probabilities with an Unknown PEL Represented by a Set of Feature Vectors

A more practical problem is to estimate $\Pr(\omega_j \mid X)$ of an unknown PEL $\boldsymbol{X}$ represented by a set of feature vectors. As the number of testing samples increases, the estimation in this case becomes more precise. Under an analysis similar to Chapters 5.1 and 5.2, the identification is made by (21) and (22) and the problem can also be transformed into estimating $\Pr(X \mid \omega_j)$ where $\boldsymbol{X}$ consists of a set of feature vectors $\underline{\boldsymbol{x}}_q$, where q is an index. For each $\underline{\boldsymbol{x}}_q$, the individual probability $\Pr(\underline{\boldsymbol{x}}_q \mid \omega_j)$ can be estimated using statistical information contained in the labeled training data as shown in (24) and (25) or utilizing the SSOM neuron grid and BMU history as shown in (26) and (27).

For the set of BMUs corresponding to $\boldsymbol{X}$, let K^L denote the number of neurons that are labeled during the training. For each labeled neuron n_i that has been the BMU to $\boldsymbol{X}$, it may have been the BMU to feature vectors from different classes during the training and thus labeled to be different classes for multiple times. The labels of each n_i are stored and ordered according to their *frequency* (total number of times), which are reflected by its activations (26) from different classes. Furthermore, $\Pr(X \mid \omega_j)$ can be the average,

maximum, or other function of all activations. A weighted average of all activations, defined as

$$\Pr\left(X \mid \omega_j\right) \approx \begin{cases} \dfrac{1}{\sum_{i=1}^{K^L} T_i} \sum_{i=1}^{K^L} \left(\dfrac{A_j\left(n_i\right)}{\sum_{j=1}^{M} A_j\left(n_i\right)} \times T_i \right), & \text{if } K^L \neq 0 \\ \textit{New PEL}, & \text{if } K^L = 0 \end{cases}, \tag{28}$$

is proposed and utilized in this dissertation, where

$$T_i = card\left\{ \underline{\boldsymbol{x}} \in X \mid n_i = \arg\min_i \left\{ \left\| \underline{\boldsymbol{x}} - \underline{\boldsymbol{m}}_i \right\| \right\} \right\} \tag{29}$$

denotes the frequency of n_i being BMU to $\boldsymbol{X}$.

5.6 DIRECT ESTIMATION

It is observed in a number of tests of training an SSOM with different data sets that with proper choice of features, most neurons have a surjective correspondence with classes. That is, each neuron has only been the BMU to one class. In this case, the probabilities (26) from activations would be 1 and a fast but accurate estimation of $\Pr(\omega_j \mid X)$ can be made by directly utilizing the distribution of all BMUs to the training feature vectors.

Assume that the T BMUs appear T_1 times as the BMU to class ω_1, T_2 times as the BMU to class ω_2, …, and T_M times as the BMU to class ω_M, then (27) can be simplified as

$$\Pr(\omega_j \mid X) \approx \frac{T_j}{T}. \tag{30}$$

These estimated probabilities are in fact percentages. If all BMUs are labeled, they are equivalent to the success rates. This estimation seems simple but quite accurate in tests where not all BMUs are labeled.

5.7 TESTS ON THE HYBRID SSOM/BAYESIAN IDENTIFIER

For practical purposes, the SSOM should be trained by sufficient amounts of data to minimize the chance of encountering unknown PELs types. However, for the purpose of illustration, in this test an SSOM is trained by the data of 48 representative PELs from 12 types, which are labeled into 7 groups, listed as follows.

(1) Cellphones;

(2) Computers: desktop and laptop;

(3) DVD players and Set-top boxes;

(4) Resistive PELS: space heater and portable fan;

(5) LCD TVs, LED TVs, and LCD Monitors;

(6) Microwave ovens

(7) Plasma TVs.

The number of PELs models from each group is listed as follows, which is used for the estimation of $\Pr(\omega_j)$.

Table 5.1 Number of PELs from each group

Group	1	2	3	4	5	6	7
Number of PELs	5	7	10	10	8	5	3

In this test, 8 PELs are used to provide three testing scenarios. For each PEL, 360 independent tests are carried out.

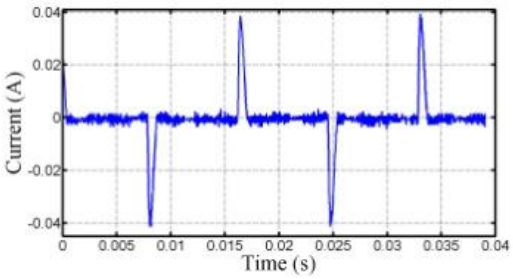
(1) Test PELs numbered by 1-4: data of these 4 PELs used in training. For this case, a 100% accuracy is expected because the statistical information of the data should have already been stored in the SSOM.

(2) Test PELs numbered by 5 and 6: PELs included in the training but a different set of data of these 2 PELs not used in training are used for testing. For this case, a high accuracy (for example, greater than 90%) is expected because different samples of the same load may contain noise and fluctuations.

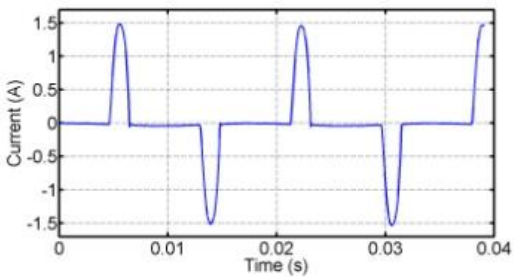
(3) Test PELs numbered by 7 and 8: PELs whose models not considered in the training.

Note that, due to space limitations, Scenario 3 considers only cases of unknown models. In other words, the test PELs types are assumed known to the proposed identifier.

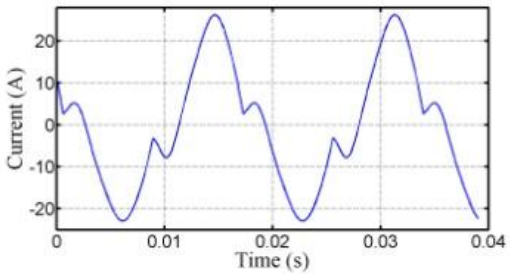
The current waveforms of test PELs are as follows.



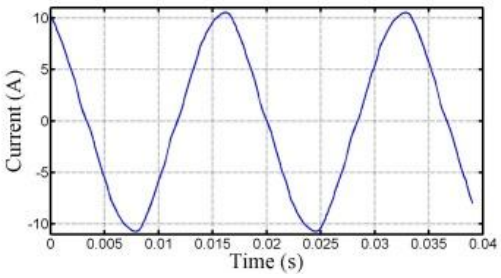
1. Cellphone



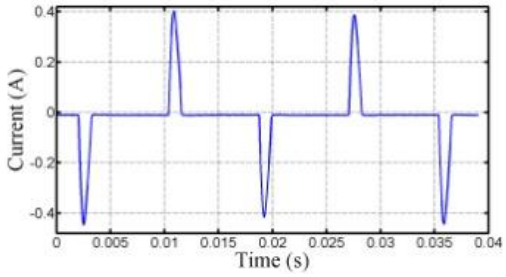
2. Set-top box



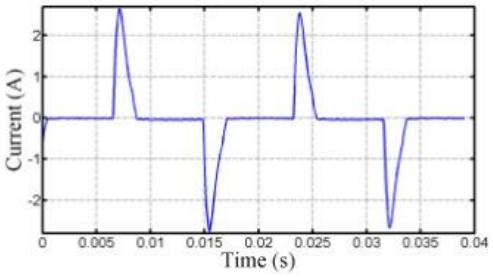
3. Type-A microwave



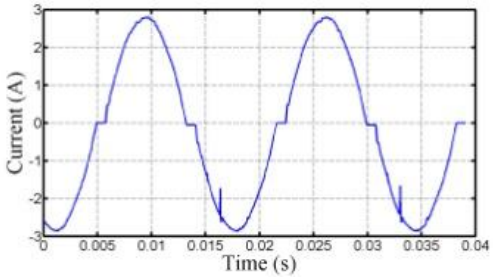
4. Space Heater



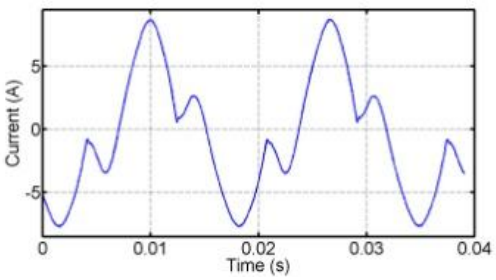
5. DVD player



6. Laptop computer



7. Type-B LCD TV



8. Type-B microwave

Figure 5.1 Current waveforms of test PELs

The testing results are summarized in the Table 5.2. The actual label of each test PEL is shown in the 2nd column from the left. For each test PEL, the BMU history from the training is listed in the 3rd column from left. Based on the BMU history, the estimated probabilities by the hybrid SSOM/Bayesian method and the direct method are listed in the 4th and 5th columns from left, respectively. Based on the estimated probability, an absolute decision can be made based on the greatest probability, which is shown in the 2nd column from the right. Finally, the most right column shows the success rate based on absolute decisions.

Take test PEL 5 for example, for which the test PEL is labeled as group 3. For 360 inputs, 329 BMUs are labeled and all of them are labeled as group 3. Therefore, the estimated probabilities by both methods are both 100% to be group 3, which is correct. Since not all neurons are labeled, the success rate based on individual decisions is 329/360=91.39%.

Table 5.2 Test results for 8 PELs in 3 Scenarios

Test PEL index	Test PEL group	BMU history	Estimated probability by hybrid classifier	Estimated probability by direct method	Decision based on estimated prob.	Success rate based on absolute decision
For 360 inputs:						
1	1	• There are 360 BMUs • all BMUs are labeled to be Group 1	100% in Group 1	100% in Group 1	Correct	100%

Table 5.2 continued

2	3	• all 360 BMUs are labeled. • all of them labeled as Group 3.	100% in Group 3	100% in Group 3	Correct	100%
3	6	• all 360 BMUs are labeled. • all of them labeled as Group 6.	100% in Group 6	100% in Group 6	Correct	100%
4	4	• all 360 BMUs are labeled. • all of them labeled as Group 4.	100% in Group 4	100% in Group 4	Correct	100%
5	3	• 329 BMUs are labeled. • all of them labeled as Group 3.	100% in Group 3	100% in Group 3	Correct	91.39%
6	2	• 353 BMUs labeled. • 304 BMUs labeled as Group 2; • 48 BMUs labeled as Group 3; • 1 BMU labeled as Group 5;	81.4% in Group 2 18.35% in Group 3 0.25% in Group 5	86.12% in Group 2 13.6% in Group 3 0.28% in Group 5	Correct	94.44%

Table 5.2 continued

7	5	• 88 BMUs are labeled. • 1 BMU labeled as Group 4; • 87 BMUs labeled as Group 5;	95.24% in Group 5	95.24% in Group 5	Correct	24.17%
8	6	• all 334 BMUs are labeled. • 11BMUs labeled as Group 9; • 322 BMUs labeled as Group 6;	96.58% in Group 6	96.58% in Group 6	Correct	86.39%

Table 5.2 shows that

(1) Tests 1-4 all have 100% success rate. Other tests also get satisfactory success rate except 7. The reason for low success rate (24.17%) based on absolute decisions for PEL 7 is that only a small number of BMUs are labeled and thus a large number of tests are identified as “new” as their BMUS are not labeled during training.

(2) Estimation by the hybrid and the direct method are the same in cases where all BMUs are labeled;

(3) Test 7 has a low success rate as most BMUs are not labeled. Thus, a probabilistic decision becomes meaningful in this case to make identification decision.

Furthermore, two representative examples are shown in the following table to explain details on estimation by the hybrid identifier.

Table 5.3 How probabilities are estimated by the hybrid identifier

Test	BMU neuron Num*	K^L	T_i	Activation $A_j(n_i)$ '*group label* (*frequency*)'		$\Pr(X \mid \omega_j)$
				highest freq	other freq	
3	4 198 2400 2500	4	339 18 1 2	'6(339)' ** '6(18)' '6(1)' '6(2)'		As all BMUs labeled as '6', $\Pr(X \mid \omega_6) = 1$
6	2016 2065 2066 2120 2259 2260 2261 2262 2267 2309 2310 2311 2312 2313 2361 2363 2364 2422 2473 2474	19	2 1 7 1 4 153 21 19 1 36 2 8 6 23 4 36 10 13 11 2	2(1) 5(18) Null '2(3)' '2(3)' '2(9)' '3(11)' '3(7)' '2(3)' '2(54)' '2(21)' '3(9)' '2(6)' '2(8)' '2(14)' '2(24)' '2(4)' '2(7)' '2(15)' '2(4)'	'2(4)' Null	$\Pr(X \mid \omega_2) = (2 + 1 \times \frac{4}{18+4} + 1 + 4 + 153 + 1 + 36 + 2 + 6 + 23 + 4 + 36 + 10 + 13 + 11 + 2) / 353 = 0.8617$ $\Pr(X \mid \omega_3) = (21 + 19 + 8) / 353 = 0.136$ $\Pr(X \mid \omega_5) = \frac{18}{18+4} / 360 = 0.0023$ $\Pr(\omega_3 \mid X) = \frac{0.8617 \times 7}{0.8617 \times 7 + 0.136 \times 10 + 0.0023 \times 8} = 0.814$

* The indexing number of neuron in the 50-by-50 neuron grid map, not related to other numerical values.

** Note that for data that has been used in the training, the testing BMU frequencies for winning neurons are exactly the same as the those in the training. For example, 339 matches 6(339).

Note that in Table 5.3 the activation $A_j(n_i)$ is denoted in the format of 'group label (frequency)'. For example, 6(339) corresponding to the neuron number 4 in the 2nd row of Table 5.3 means that the neuron number 4 has been BMU to group label 6 for a frequency of 339 times.

5.8 CHAPTER SUMMARY

This chapter proposes a simple yet efficient and practical method for the probabilistic identification of PELs. The SSOM is combined with the Bayesian decision making method to perform as a hybrid identifier. The history of the best matching units (BMUs) during the training is used to estimate the probability of an unknown PEL belonging to each known class and achieves high accuracy.

The proposed hybrid classifier achieved correct classification decision in the tests. The proposed framework is also quite robust to how many BMUs are labeled. Remaining question related to the hybrid classifier is that the confidence level of the estimated probability, which requires another large set of data to calculate. Also, in all the tests the estimated *a posterior* probabilities are quite different in magnitudes. For instance, in Table 5.2 the estimated probabilities are 81.4% in Group 2, 18.35% in Group 3, and 0.25% in Group 5. The errors in estimation have no impact on the hard decision based on

maximum *a posterior* probabilities as 81.4% is much greater than 18.35%. However, if in some cases that the estimated *a posterior* probabilities of several different categories are quite close, the errors in estimation may cause incorrect hard decision.

CHAPTER 6 MULTI-CLASS IDENTIFICATION OF PLUGGED-IN ELECTRIC LOADS BY SUPPORT VECTOR MACHINE

6.1 INTRODUCTION

As discussed earlier in Chapters Four and Five, the supervised SOM (SSOM) identifier can classify a large number of PELs into a number of clusters in its neuron grid. However, the SSOM framework has been originally designed as a data classification and visualization algorithm rather than an identification method. Therefore, the SSOM identifier could encounter problems to distinguish some PELs equipped by similar front-end power supply units (and thus similar characteristics) within each of the seven PEL categories proposed in Chapter Three. For example, in Figure 4.2 an LCD TV and a desktop computer both belong to Category NP, are all labeled as "4", and thus fall into the same cluster. In this case, the SSOM identifier has a low success rate in identifying each PEL within this cluster. Therefore, other methods are desired to solve the problem of identifying similar PELs within each category, referred to as the *in-category* PEL identification in this dissertation.

In this dissertation, it is proposed to use two different methods for in-category PELs identification: namely static and dynamic. The static method is data-based. That is, the set of training data for the SSOM is again used for (steady-state or static) feature extraction and algorithm training to achieve more precise identification results. However, many PELs, such as office appliances and personal computers, exhibit time-varying operations with respect to time and usage. Therefore, these PELs may possess time-varying steady-state features. Another case in which steady operating states are not well defined is

during the startup and transient operating modes. For example, the current waveform of a copier during the copying process is shown in Figure 6.1.

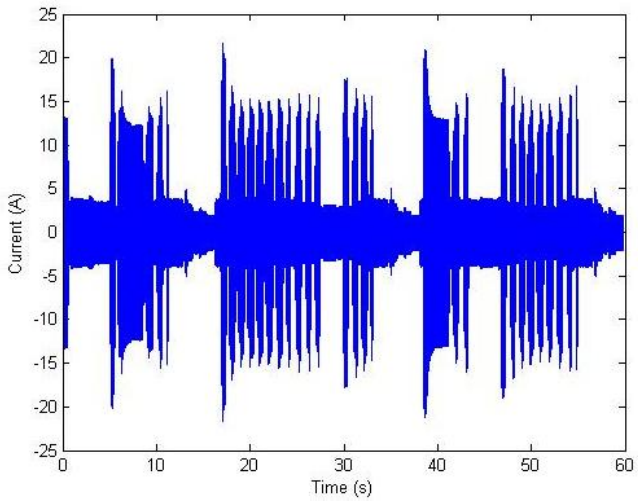


Figure 6.1 Current waveform of a copier in operation for 60 seconds

On the other hand, the dynamic method is more model-based rather than data-based. In other words, the in-category PELs identification is carried out based on the understanding of the operating principles and mechanisms of target PEL models. In this case, this dissertation proposes to model long-term (in minutes or hours, compared to short-term which is mainly in cycles or seconds) waveforms to extract information and patterns and perform identification.

For the static method, the support vector machine (SVM) is considered for in-category PELs identification and presented in this chapter. On the other hand, the proposed dynamic method is discussed in Chapter Seven.

6.2 IN-CATEGORY PEL IDENTIFICATION BY SVM

6.2.1 Introduction to Support Vector Machine (SVM)

The basic SVM [102] is inherently designed for classification of two classes, ω_1 and ω_2, of patterns which are represented by a dataset of labeled feature vectors. An SVM first train itself with the dataset, constructs hyperplanes in a high-dimensional (possibly infinite dimensional) vector space, and then assigns each input feature vector x with unknown identity into one of two known classes.

Given a set of data in which the two classes ω_1 and ω_2 are assumed to be linearly separable, the first step an SVM carries out is to find a (linear) hyperplane that separates the training data, denoted by

$$g(x) = \boldsymbol{\omega}^T x + \boldsymbol{\omega}_0 = 0 \tag{31}$$

in which $x \in \mathbb{R}^l$ is an l-dimensional feature vector in the training data set, $\boldsymbol{\omega} \in \mathbb{R}^l$ is known as the *weight* of the hyperplane, and $\boldsymbol{\omega}_0$ is the *threshold* of the hyperplane [89].

Formally, a hyperplane g(x) (typically denoted by its parameters ($\boldsymbol{\omega}$, $\boldsymbol{\omega}_0$)) separates the training dataset if feature vectors belong to ω_1 fall into one side of ($\boldsymbol{\omega}$, $\boldsymbol{\omega}_0$) and those feature vectors belong to ω_2 fall into the other side. Such a separating hyperplane ($\boldsymbol{\omega}$, $\boldsymbol{\omega}_0$) is not unique for a training dataset. Instead, there typically exist many such separating hyperplanes. Therefore, an SVM searches for pairs of separating hyperplanes such that they separate the training dataset and there are no training feature vectors fall between them. The region bounded by a pair of separating hyperplanes is called the *margin* of this

pair. The output of a basic SVM is a pair of separating hyperplanes with greatest distance, typically known as the *maximum-margin* hyperplanes.

To summarize, the basic SVM is a linear classifier. When the dataset of two classes is not linear separable, it is proposed in 1992 [103] to extend the basic SVM to a nonlinear classifier by applying the *kernel*. First, an SVM utilizes a (typically nonlinear) mapping Φ to map the training dataset from the original feature vector space to a high-dimensional (possibly infinite dimensional) Euclidean space H:

$$\Phi : x \in \mathbb{R}^l \to \Phi(x) \in H, \tag{32}$$

in which the two classes ω_1 and ω_2 of patterns can be separated by a hyperplane denoted by

$$g(x) = \boldsymbol{\omega}^T x + \boldsymbol{\omega_0}. \tag{33}$$

Generally, the SVM only depends on the training dataset through inner products in H, i.e. on functions of the form

$$K(x_i, x_j) = \left\langle \Phi(x_i), \Phi(x_j) \right\rangle. \tag{34}$$

where K is usually called a *kernel* function. It is common that only K needs to be specified instead of knowing the explicit form of Φ when training a SVM.

Let $x_i, i = 1, 2, ...,$ be feature vectors in the training data. For each x_i denote the corresponding class indicator by y_i (e.g., +1 for ω_1 and -1 for ω_2). Once an appropriate kernel has been selected, the optimal hyperplane ($\boldsymbol{\omega}$, $\boldsymbol{\omega_0}$) can be determined by solving the constrained optimization problem [89]

$$\max_{\lambda}\left(\sum_{i}\lambda_i - \frac{1}{2}\sum_{i,j}\lambda_i\lambda_j y_i y_j K(x_i, x_j)\right) \tag{35}$$

subject to

$$\begin{aligned} &0 \le \lambda_i \le C, \qquad i = 1, 2, \ldots N \\ &\sum_{i}\lambda_i y_i = 0 \end{aligned} \tag{36}$$

where $\boldsymbol{\lambda}$ is the vector of nonnegative Lagrange multipliers λ_i and C is a parameter to be chosen by the user with larger C corresponding to assigning a higher penalty to errors.

The resulting classifier assigns x to ω_1 (ω_2) if

$$g(x) = \sum_{i=1}^{N}\lambda_i y_i K(x_i, x) + \omega_0 > (<)\ 0 \tag{37}$$

The SVM utilizing a kernel function K is shown in the following figure [89].

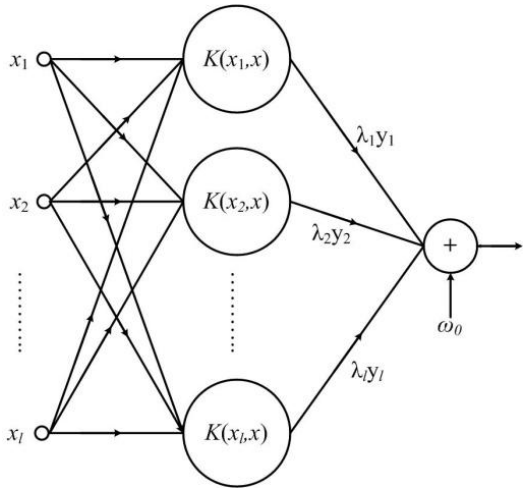


Figure 6.2 The SVM framework utilizing kernel functions.

6.2.2 Multi-class identification by SVM

The basic SVM introduced in section 6.2.1 is designed for two class identification problem, i.e., one-against-one. When the training dataset consists of feature vectors from a number of *M* classes (*M* is greater than two) and an input feature vector needs to be assigned into one of the classes, the basic SVM cannot be directly applied to this multi-class identification problem. A recent empirical study [104] indicates that most efficient method for extending SVM to multi-class identification is to decompose it into multiple one-against-one problems which can be handled by basic SVMs. In an *M*-class identification problem, there are two common extensions of the SVM [105]:

(1) The first extension considers the *M*-class identification problem as a set of *M* two-class problems, known as *one-against-all*.

(2) The other extension aims at training *M*(*M*-1)/2 basic SVM classifiers, called *one-against-one*.

Furthermore, researchers are also investigating other extensions of the basic SVM to efficiently handle the multi-class identification problem. For instance, Crammer and Singer [106] proposed a multiclass SVM method which casts the multiclass classification problem into a single optimization problem, rather than decomposing it into multiple binary classification problems. Other notable extensions of the binary SVM to multiclass identification problems include fuzzy logics [107, 108], decision tree architecture [109], and adaptive directed acyclic graph [110].

However, there is no conclusion yet in literature which extension works better. Different extension methods are used in different applications such as protein fold

identification [111], image classification [112], document categorization [113], and text classification [114, 115].

In this dissertation, the one-against-all multi-class SVM is adopted. In other words, for each ω_i of the M classes, the one-against-all SVM aims at determining optimal hyperplanes, $g_i(x)$, $i = 1,2,\ldots,M$, so that $g_i(x) > g_j(x)$ for all $j \neq i$ and $x \in \omega_i$. The classification rule is given by [89]

$$x \in \omega_i \quad \text{if} \quad i = \arg\max_k \{g_k(x)\} \tag{38}$$

Many kernels are available for use in an SVM, such as polynomials, radial basis function (RBF), and hyperbolic tangent kernels. In this dissertation the Gaussian RBF kernel [116] is selected, which is also the most commonly adopted kernel in pattern recognition problems.

6.3 IN-CATEGORY PEL IDENTIFICATION BY ONE-AGAINST-ALL SVM

As discussed above, the SVM identifier is known to perform well when handling similar but not identical sets of feature vectors [115] and thus can be applied to the in-category PEL identification problem. Furthermore, the SVM can be either directly applied for multi-class in-category PEL identification or combined with the SSOM framework presented in Chapters Four and Five. The latter combination is expected to have better performance because it is based on the *a priori* category information and deals with fewer numbers of PELs to be identified. This combination of the SVM and the SSOM forms a hybrid SSOM/SVM framework which is studied and adopted in this dissertation for multi-class in-category PEL identification [117].

The architecture of this hybrid SSOM/SVM PEL identifier is shown in the following

figure.

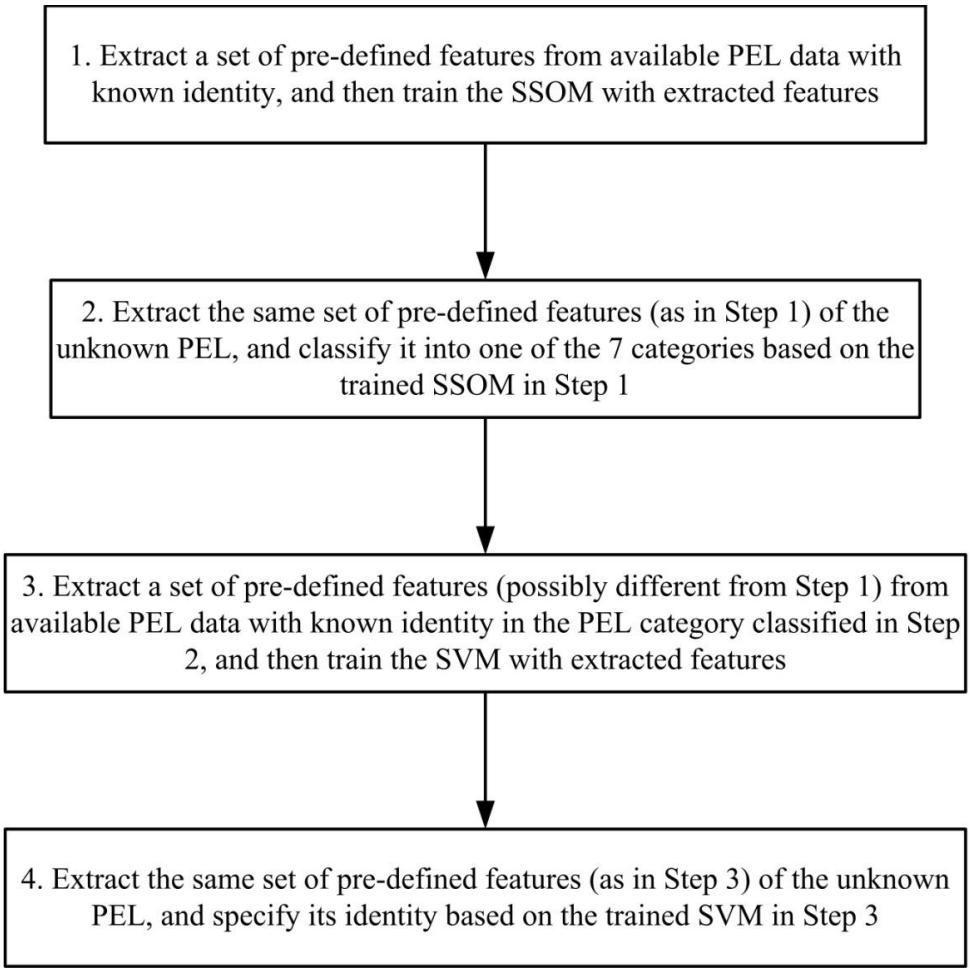


Figure 6.3 The hybrid SSOM/SVM identifier framework

Note that in Figure 6.3 steps 1 and 3 may use different sets of features because the SSOM and the SVM utilize different principles in the training process and thus the training outputs are also different. For instance, the training output of an SVM is a set of parameters describing a high-dimensional hyper-plane, which cannot be visualized in a two-dimensional as it can be in the SSOM framework. Therefore, a set of features that

works well for the SSOM may not achieve the same accuracy for the SVM.

Based on the above discussion, it is necessary to compare the performance of different steady-state features in the SVM that are widely used in the literature. Considering the large number of available features as discussed in section 2.3 and the even larger number of combinations, it is unlikely and unnecessary to test all combinations. Furthermore, as discussed in Chapter Four, a set of time-domain features are selected for the SSOM after a number of tests.

In the literature, several researchers [74, 75] have used the SVM for signature recognition and harmonic source identification in power distribution networks, in which the harmonic spectrum (i.e., frequency-domain features) is considered for SVM training and identification. Following these existing work using frequency-domain features and other work in the literature using time-domain features as reviewed in Chapter Two, this dissertation compares the performance of time-domain and frequency-domain features used in SVM for PEL identification.

6.4 COMPARISON OF TIME-DOMAIN AND FREQUENCY-DOMAIN FEATURES FOR SVM

This section compares two sets of features that have been widely used to represent PELs:

(1) Time-domain features: as listed in section 4.4.1;

(2) Frequency-domain features found in the harmonic spectrum of the current waveform. For example, the features used in this chapter are all the odd harmonics from the fundamental to the 25^{th}.

Consider a dataset consists of feature vectors for 3 selected PELs with quite different current waveform characteristics (such as shape, total harmonic distortion, power factor,

and real power). The three selected PELs include a DVD player (labeled as D), an LCD TV (labeled as T), and an oscillating fan (labeled as F). The length of measurement is 60 seconds for all three PELs, i.e., 3600 electrical cycles of data for each PEL. Therefore, the total number of available feature vectors for each PEL for training and testing is 3600 if harmonic components are calculated by an FFT for each cycle.

Their current and voltage profiles are shown in the following figure.

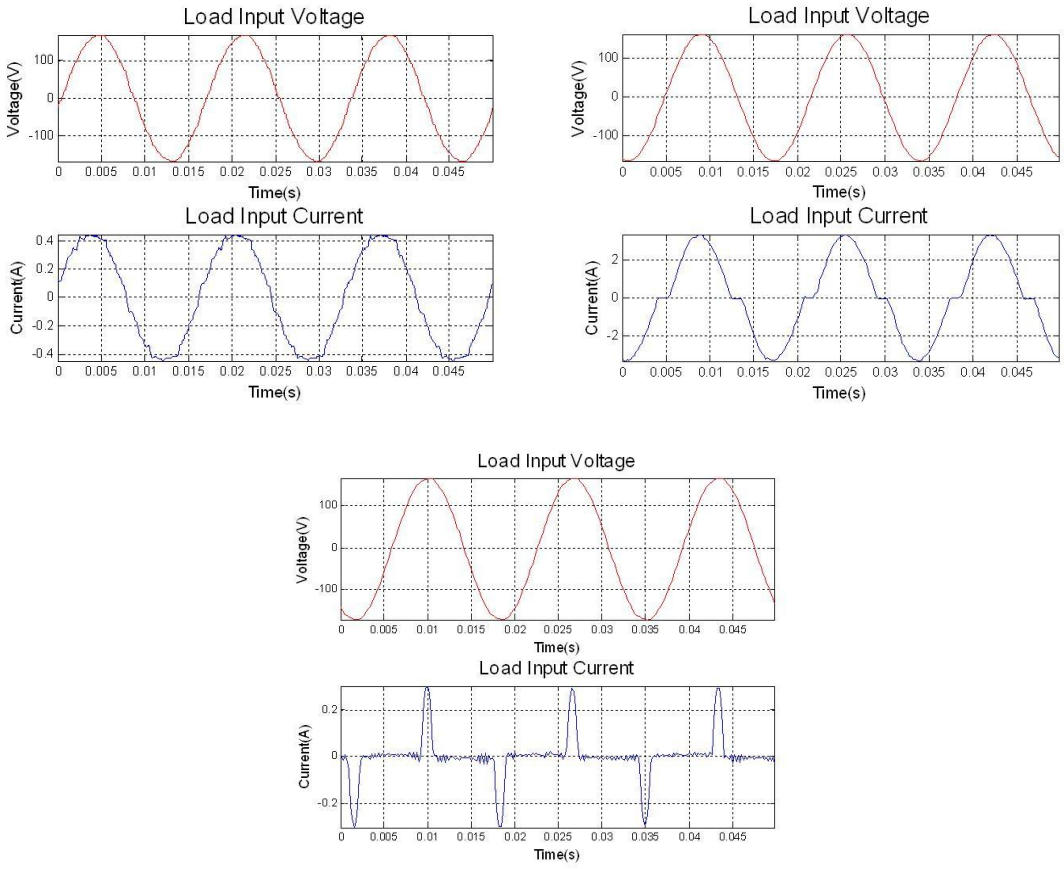


Figure 6.4 Current and voltage profiles of three PELs with quite different characteristics: Fan (top left), TV (top right), and DVD (bottom)

Figure 6.4 shows that the selected three PELs have quite different current characteristics: the current waveform shape of the fan (top left) is significantly different from the TV(top right) and the DVD player (bottom). Also, the peak current value of the TV (top right) is around 2 A, which is significantly greater than the fan (top left) and the DVD player (bottom). The purpose of selecting distinguishable PELs is to show that in this simple case the time-domain features with the SVM can guarantee 100% accuracy but the frequency-domain features cannot.

Furthermore, three different cases with different ratios of training and testing data are tested and compared, and the testing results are shown in the following table. Note that the results are generated by solving multi-class one-against-all SVMs using the SVM-KM toolbox [118] for MATLAB. More details on the inputs and outputs of the MATLAB program are provided in Appendix D.

Note that the cross validation mechanism shown in Figure 4.3 is also adopted to test the performance of the one-against-all SVM identifier. The identification success rate is defined and calculated in a similar manner as in Figure 4.3.

Table 6.1 Comparison of testing success rate of different feature sets using multi-class one-against-all SVMs

Success rate with different training/testing ratio and features	270 points for training , 3330 for testing	540 points for training , 3060 for testing	1080 points for training , 2520 for testing
Time-domain features	100%	100%	100%
Frequency-domain features	99.56%	99.53%	99.43%

Table 6.1 shows that simply using harmonics cannot guarantee a success rate of 100% even with only the three relatively distinct PELs with the quite different current waveform characteristics of Figure 6.4. In other words, errors may occur even for a simple case of multi-class PEL identification if PELs are represented by frequency-domain features. Moreover, these errors can be avoided if the PELs are represented by time-domain features. Therefore, in this dissertation only the set of (time-domain) features listed in section 4.4.1 is used to represent PELs for in-category PEL identification by SVM, similar to what has been used for the SSOM.

Furthermore, one well known advantage of the multi-class one-against-all SVM classifier is that it requires a relatively small amount of training data compared with other notable classifiers. Some test results for the three PELs shown in Figure 6.4 but with different choices of cross-validation are shown in the following table, in which the ratios between training data and testing data are indicated in the first column.

In these tests, 512-point FFTs are carried out to calculate the harmonics if necessary.

Table 6.2 Testing success rate of the multi-class one-against-all SVM and SSOM identifiers with different amount of data

Success Rate	DVD	TV	Fan	Total
SVM (5% for training, 95% for test)	100%	93.77%	100%	97.92%
SVM (10% for training, 90% for test)	100%	93.43%	100%	97.81%
SVM (20% for training, 80% for test)	100%	100%	100%	100%
SVM (30% for training, 70% for test)	100%	100%	100%	100%

Table 6.2 continued

SSOM (5% for training, 95% for test)	49.36%	81.99%	97.34%	76.23%
SSOM (10% for training, 90% for test)	91.30%	87.81%	97.69%	92.26%
SSOM (20% for training, 80% for test)	96.74%	94.03%	93.92%	94.90%
SSOM (67% for training, 33% for test)	99.75%	99.83%	98.00%	99.19%

As another advantage of combining SSOM and SVM, it is clear in Table 6.2 that the multi-class one-against-all SVM classifier can get 100% success rate with only 20% of the total data. On the other hand, the SSOM identifier trained by the same amount of data only achieves an average success rate of 94.9%, which is much lower than the success rate of the SVM.

6.5 TESTING THE PERFORMANCE OF THE HYBRID SSOM/SVM CLASSIFIER ON A LARGE NUMBER OF PELS

Two tests are carried out in this section to show the performance of proposed methods for multi-class in-category PELs identification. These two tests are carried out with the SSOM trained by feature vectors (as listed in section 4.4.1) from 32 different PELs of 12 types mixed together. Moreover, each PEL has 3600 feature vectors.

6.5.1 Comparison of performance of the SSOM identifier and the hybrid SSOM/SVM identifier on in-category PEL identification

In this case, three different types of TVs (one LCD TV, one LED TV, and one plasma TV) are tested. The voltage and current profiles of these three TVs are shown in

the following figure. These three PELs share similar front-end power supply units as well as feature vectors, and thus all belong to Category P. Therefore, this case represents a multi-class in-category PEL identification problem. Both the SSOM identifier and the hybrid SSOM/SVM identifier are used to compare their performance on this in-category PEL identification problem.

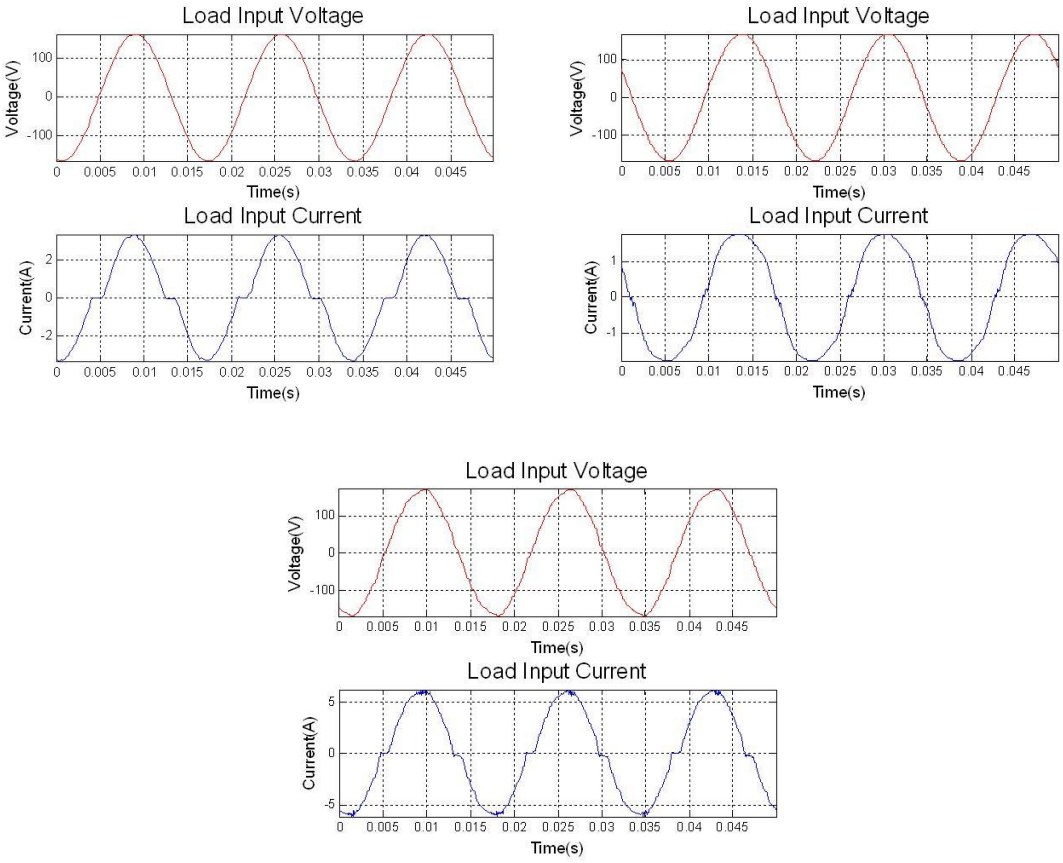


Figure 6.5 Current and voltage profiles of LCD TV (top left), LED TV (top right), and plasma TV (bottom)

Figure 6.5 shows that these three TVs have very similar current waveform

characteristics. In the trained SSOM, these three TVs fall into the same cluster and are thus very difficult to be identified without ambiguity using the SSOM identifier.

The success rate of in-category PEL identification by the SSOM identifier and the hybrid SSOM/SVM identifier are listed in the following table. Table 6.3 shows that the SSOM identifier achieves an average success rate around only 85% to identify each type of TVs. On the other hand, the hybrid SSOM/SVM classifier can achieve an average testing success rate greater than 95%.

Table 6.3 Testing success rate of the hybrid SSOM/SVM identifier and the SSOM identifier for in-category PELs

Success Rate	LCD TV	LED TV	Plasma TV	Average
SSOM identifier	80.17%	97.85%	85.25%	85.28%
Hybrid SSOM/SVM identifier (20% data for training)	98.30%	78.89%	98.96%	92.05%
Hybrid SSOM/SVM identifier (30% data for training)	95.99%	90.95%	98.85%	95.26%

Furthermore, Table 6.3 shows that the more training data for the SVM in the hybrid SSOM/SVM identifier, the better performance it has. However, the SVM training in the hybrid SSOM/SVM identifier still requires far less data than the pure SSOM classifier as discussed in section 6.3.

6.5.2 Testing the performance of the hybrid SSOM/ SVM classifier on a large number of PELs

The following Table 6.4 summarizes test results to show the performance of the hybrid SSOM/SVM identifier for in-category PEL identification. A SVM is first trained by the same data that is used to train the SSOM, and 13 PELs from 5 major categories are test to check the performance of the multi-class one-against-all SVM identifier.

Table 6.4 Testing success rate of typical PELs using the multi-class one-against-all SVM identifier

PEL	Category	Success Rate (%)
Compact Fluorescent Lights	P	98.67
Fluorescent Lights	R	100
Incandescent Lights	R	100
Fan	X	100
Printer	M	99.66
Cellphone Charger	NP	100
DVD player	NP	98.66
Heater	R	100
LCD TV	P	99.72
LED TV	P	93.33
Microwave	M	100
Plasma TV	P	89.66
Set Top Box	NP	100

6.6 DISCUSSION ON THE PERFORMANCE OF THE HYBRID SSOM/SVM IDENTIFIER

Compared with existing methods such as only SSOM or SVM identifiers, the proposed hybrid SSOM/SVM identifier performs better in the sense of accuracy, robustness, and applicability. The SSOM identifier first extracts information from the large amount of training data and store that simplified information in the trained neuron grid. When an input feature vector is presented, the hybrid SSOM/SVM identifier first determines which load category the input feature vector falls into and then utilizes the SVM discriminant function for each category to get a robust and correct identification decision. Tests results based on real-world data show that an average accuracy of over 99% can be achieved by the hybrid SSOM/SVM identifier.

However, the in-category PEL identification by SVM (combined with the SSOM or not) is still based on the representation of PELs by steady-state (or static) features. The SVM identifier may have difficulty in handling PELs with highly dynamic activities over time. For instance, as shown in Table 6.3 the hybrid SSOM/SVM identifier achieved a lower success rate in identifying LED TVs (93.33%) and plasma TVs (89.66%). As shown in the Chapter Seven, the dynamic characteristics of some PELs should be addressed for reliable PEL identification as well as operating mode identification.

6.7 CHAPTER SUMMARY

This chapter presents a novel hybrid SSOM/SVM identifier for the multi-class in-category PEL identification problem. The proposed hybrid identifier utilizes the power of previously supervised Self-Organizing Map (SSOM) classifier for PELs proposed in Chapters Four and Five to first classifier an unknown PEL into one of the seven PEL categories discussed in Chapter Three. Within each cluster, a more accurate identification

decision is made by the well establish multi-class one-against-all SVM classifier. The average success rate based on different tests is over 95%, which is satisfactory.

The proposed hybrid SSOM/SVM identifier still suffers from several disadvantages, such as only handling steady-state conditions and still having a high computational cost especially with large number of classes in each category. Moreover, the features used for the hybrid SSOM/SVM are still extracted from cycle-by-cycle waveform, i.e., short-term observations. Chapter Seven proposes a method to extract features and information from long-term observations.

CHAPTER 7 IDENTIFICATION OF PLUGGED-IN ELECTRIC LOADS BY LONG-TERM WAVEFORMS

7.1 INTRODUCTION

As discussed in section 1.4.2, it is desired by various applications to have a reliable and efficient method to identify not only the type but also the operating mode of PELs. Since the introduction of non-intrusive load monitoring (NILM) [49] in the 1980s, much effort has been devoted to identify the type and model of different electric loads. A comprehensive survey of existing load identification methods can be found in [46]. However, very few existing work has considered identifying the operating status or modes (i.e., standby, active, or sleep) of loads in real-time; ignoring such information may lead to incorrect load identification as shown in section 2.5.

Furthermore, without considering the operating modes and status of PELs, existing load identification methods may assign an incorrect identity to an electric load that exhibits different current waveforms (and thus different current characteristics) under different operating modes. The rapidly developing designs of front-end power supply units and worldwide implementation of regulations on energy efficiency further complicate the load identification. Moreover, many electronic appliances are currently equipped with power factor correction (PFC) units which may be turned on or off automatically while in different operating modes. Examples can be found in Figures 1.4 and 1.7 in Chapter One.

This chapter proposes a novel finite-state-machine (FSM) representation of long-term operating waveforms of electric loads for the purpose of indicating load identity and operating modes. The operating waveform is first converted into a quantized sequence of

states. Each state is assigned with two-dimensional numerical values: RMS current values and staying-time values. A set of elemental states and events are defined to reduce the number of states and extract numerical features to represent and identify electric loads under different operating modes. Three major categories of repeating patterns in waveforms that correspond to repeating operating actions are summarized, and identification methods are proposed for each such category.

7.2 REPRESENTING LONG-TERM WAVEFORM BY FINITE-STATE MACHINE

7.2.1 Finite-state machines

A finite-state machine (FSM) consists of a finite number of states, a set of actions, and a set of state transitions between states. A state transition is an action that starts from one state and ends in another state. If the start state and the end state are the same, it is called a *self*-state transition. A state transition is triggered by a pre-defined event or a condition. In some context, the FSM framework is also known as the *automata* theory [119].

Formally, a finite-state machine F is a 4-tuple where F is defined as

$$F = (Q, \Sigma, \delta, q_0) \tag{39}$$

where Q is the nonempty set of *states*, Σ is the set of *events*, $\delta : Q \times \Sigma \to Q$ is the *state transition function*, and $q_0 \in Q$ is the *initial state*. Note that $\times$ denotes the Cartesian product of two sets and $\to$ denotes an *onto* mapping.

The following figure illustrates the elements of a FSM:

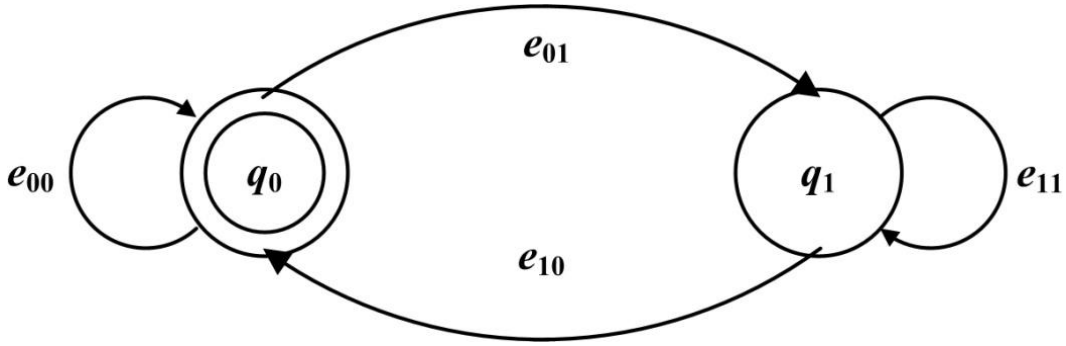


Figure 7.1 An example of a FSM illustrating its elements

In Figure 7.1, a FSM consists of the following elements:

(1) A set of states $Q = \{q_0, q_1\}$;

(2) A initial state q_0;

(3) A set of events $\Sigma = \{e_{00}, e_{01}, e_{10}, e_{11}\}$;

(4) A state transition function δ such that:

$\delta(q_0, e_{00}) = q_0$;
$\delta(q_0, e_{01}) = q_1$;
$\delta(q_1, e_{10}) = q_0$;
$\delta(q_1, e_{11}) = q_1$;
$\delta(q_0, e_{10}), \delta(q_0, e_{11}), \delta(q_1, e_{01}), \delta(q_1, e_{00})$ not defined

In other words, starting from the initial state q_0, the FSM either transits to state q_1 as defined by $\delta(q_0, e_{01})$ or stays at q_0 as defined by the self-state-transition $\delta(q_0, e_{00})$. When the FSM is at state q_1, the FSM either transits to state q_0 as defined by $\delta(q_1, e_{10})$

or stays at q_1 as defined by the self-state-transition $\delta(q_1, e_{11})$. Other state transitions are not defined.

7.2.2 Representation of long-term current waveforms

Existing representation methods of operational waveforms do not perform well for the identification of electric load operating modes in real-time. For PELs identification, waveforms are typically represented by a set of time-domain or frequency-domain features, which requires high resolution data (e.g., at least 10^3 samples per second). However, recent efforts [76-78] represent long-term operation of electric loads only by very low resolution data (e.g., one active power measurement every 5 minutes or even an hour).

This dissertation represents the long-term operation of any electric load by a sequence of RMS current values $I_{RMS}(\cdot)$, where $I_{RMS}(n)$ denotes the RMS current value over a pre-defined time window n. The length of the pre-defined time window can be from one cycle (e.g., 1/60 seconds in U.S.) to minutes or hours, depending on the purpose of different applications.

The representation by I_{RMS} has many advantages compared with other options. For instance, the sequence of peak current values can also represent long-term operations but a high sampling rate is typically required to maintain accurate measurement of peak values. Also, $I_{RMS}(n)$ provides equivalent information about the average active power at time window n but needs no additional multiplication (with RMS voltage) and thus reduces computational costs.

7.2.3 Definition of states

For the purpose of real-time identification of operating modes, sequences of $I_{RMS}(\cdot)$ representing long-term operation of electric loads are then sent into the proposed algorithm for real-time process and analysis. An $I_{RMS}(\cdot)$ sequence is transformed into an event sequence which is represented by a FSM, from which a set of features can be extracted to identify different operating modes. Besides the capability of extracting features from numerical variances in $I_{RMS}(\cdot)$ values, another advantage of FSM modeling is its capability to efficiently handle the concept of time. Specifically, if an electric load is in steady operating mode and consumes an almost constant amount of power, the corresponding FSM then stays at a certain state enabled by self-state transition as defined in section 7.1.1. In other words, the FSM representation also records how long it stays at each state.

To summarize, the representation by $I_{RMS}(\cdot)$ is actually two-dimensional, i.e., includes time durations and values of current. Therefore, each state $S_k(I_k, T_k)$ within an associated FSM is assigned two values, one corresponds to the current value (denoted by I_k) and the other one corresponds to the total amount of time (denoted by T_k), where k is an index for states.

Let $I_{RMS}(i)$ be a sequence of RMS current values, where $i = 1, 2, \ldots, N$ denotes an index for RMS current values and N is the total number of available RMS current values. A set of states $S_k(I_k, T_k)$ with assigned values I_k and T_k can be extracted from $I_{RMS}(i)$ by the following algorithm.

(1) Start with $k = 1$, assign $I_k = I_{RMS}(1)$, and let $T_k = 1$;

(2) Continue with $i = 1$ and compute the stepwise difference $\Delta_{i+1,i} = \left| \frac{I_{RMS}(i+1) - I_{RMS}(i)}{I_{RMS}(i)} \right|$ between $I_{RMS}(i+1)$ and $I_{RMS}(i)$. If $\Delta_{i+1,i} > \Delta$, where Δ is a pre-defined threshold value, then $k = k + 1$ and go to step (3). Otherwise $i = i + 1$, $T_k = T_k + 1$, and repeat step (2);

(3) Assign $I_k = I_{RMS}(i+1)$, let $T_k = 1$, and compute the stepwise difference $\Delta_{i+1,i}$. If $\Delta_{i+1,i} > \Delta$, then $k = k + 1$ and repeat step (3). Otherwise, $i = i + 1$ and $T_k = T_k + 1$, and repeat step (3);

(4) Stop when $i = N$, where N is the total number of RMS data points.

Note that the above algorithm analyzes operating current waveforms in real-time. In other words, RMS values of current of each time window are analyzed and compared with those of a previous time window to represent the operating patterns with respect to time.

The following figure shows an example**.** Consider the operating current waveform of a desktop computer for 60 seconds shown in the following figure, which includes the transient period from OFF to ON (from around 2 seconds to 3 seconds) as well as start-up period during active ON mode lasting for more than 55 seconds. The time window is set to be one cycle in this example (1/60 seconds in U.S.) and thus $N = 3600$.

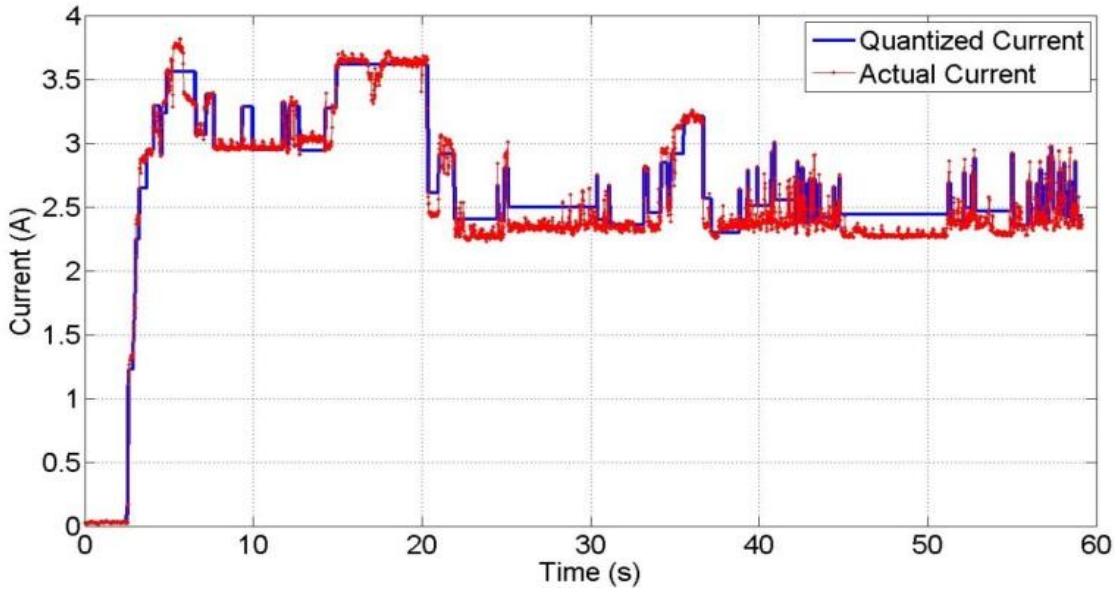


Figure 7.2 Actual and quantized operating current waveforms of a plasma TV for 60 seconds

For comparison, the actual $I_{RMS}(i)$, $i = 1, 2, \ldots, 3600$, are plotted as the dashed line (in red color) and the associated state values T_k are plotted as the solid line (in blue color) in Figure 7.2. Furthermore, extracted states $S_k(I_k, T_k)$ are shown in the following figure. Note that k is typically far less than N (in this example $k = 112$) because each state S_k corresponds to a total number of T_k RMS values of current that are within the region $(I_k - \Delta, I_k + \Delta)$.

Note that in Figure 7.3 the upper bars (above the zero-valued horizontal axis) represent the values of associated RMS current value I_k of each state and the lower bars (below the zero-valued horizontal axis) represent the values of associated staying-time T_k of each state.

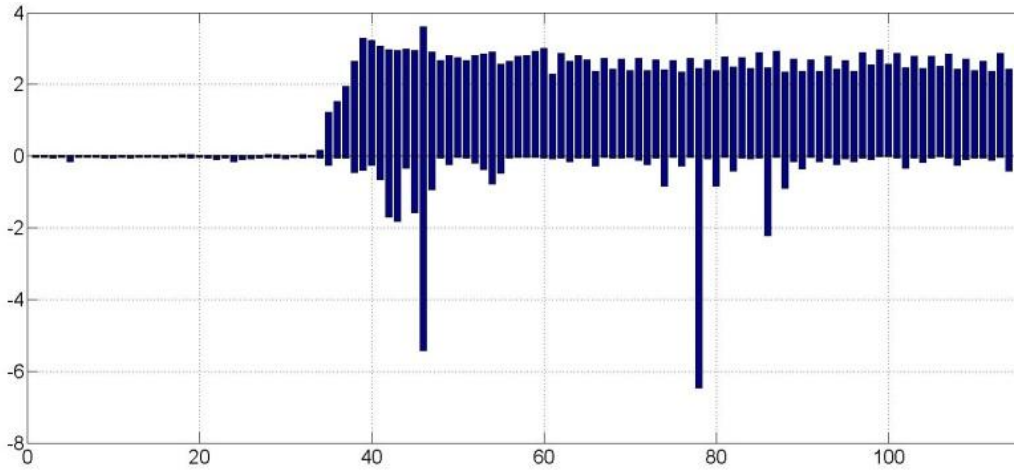


Figure 7.3 Extracted states with associated current and time values from the operating waveform in Figure 7.2

The start-up process of the desktop computer only takes a few seconds. The power consumption of the computer in active mode (when running) varies for a large part of time except the two time periods between around 15 to 20 seconds and between 45 and 52 seconds. These shapes can be observed within the corresponding extracted states.

However, the relevant FSM contains too many states and does not explicitly include the start-up transients such as spikes. The next section defines several elemental states and events to reduce the number of states, highlight states of interests, and extract describing features.

7.3 Reduction of States and Definition of Elemental States and Events

For the purpose of reducing the number of states and extracting useful information, the following four elemental states as well as the spike event are defined.

(1) Standby state: if the associated I_k value for a state S_k is less than a predefined threshold value (e.g., 0.1A), then S_k is defined as a *standby* state. Furthermore, if a standby state is the initial state (or the first state in sequence), it is called the *start* state.

(2) Steady state: if a representing FSM stays at a certain state for at least 5 seconds, then the state is defined to be a *steady* state;

(3) Semi-steady state: if a representing FSM stays at a certain state for at least 1 second but less than 5 seconds, then this state is called a *semi-steady* state (sometimes all called *semi* states for abbreviation in this dissertation);

(4) Oscillation state: if a representing FSM stays at a certain state for less than 1 second, then this state is called an *oscillation* state;

(5) Spike event: if $\frac{I_{RMS}(i)}{I_{RMS}(i-1)} > \varepsilon_1$ and $\frac{I_{RMS}(i)}{I_{RMS}(i+1)} > \varepsilon_2$, where ε_1 and ε_2 are pre-defined threshold values, then a *spike* event is defined to occur at time step *i*. Note that a spike is actually an event instead of a state.

Remark: Although the spike event is defined over RMS current values, it can be equivalently defined over I_k values because each state S_k corresponds to a total number of T_k sequential RMS current values with a small variance in magnitude. That is, a spike event does not occur between associated RMS current values of any state but only possibly occur between the T_k-th I_k value of state S_k and the first I_{k+1} value of state S_{k+1}. Therefore, a spike event indeed triggers transition between states.

Remark: the threshold values ε_1 and ε_2 determine the performance of the proposed FSM presentation and thus should be properly selected according to applications and comparisons of different choices. The number of spike events decreases and some obvious spike events could be missed if ε_1 and ε_2 are too large. On the other hand, there might be too many spike events if ε_1 and ε_2 are too small. Therefore, selecting proper threshold values ε_1 and ε_2 is a trade-off between different factors. There is no general rule on how to select threshold values ε_1 and ε_2 but it is suggested to choose ε_1 and ε_2 between 1 and 3 based on a large number of trials and tests.

Continuing with the above example in section 7.1.3, Figure 7.4 plots the peak-to-peak envelop of the operating current waveform of a plasma TV for 60 seconds. Furthermore, Figure 7.5 shows the corresponding extracted states with their associated current and staying-time values from the current waveform shown in Figure 7.4.

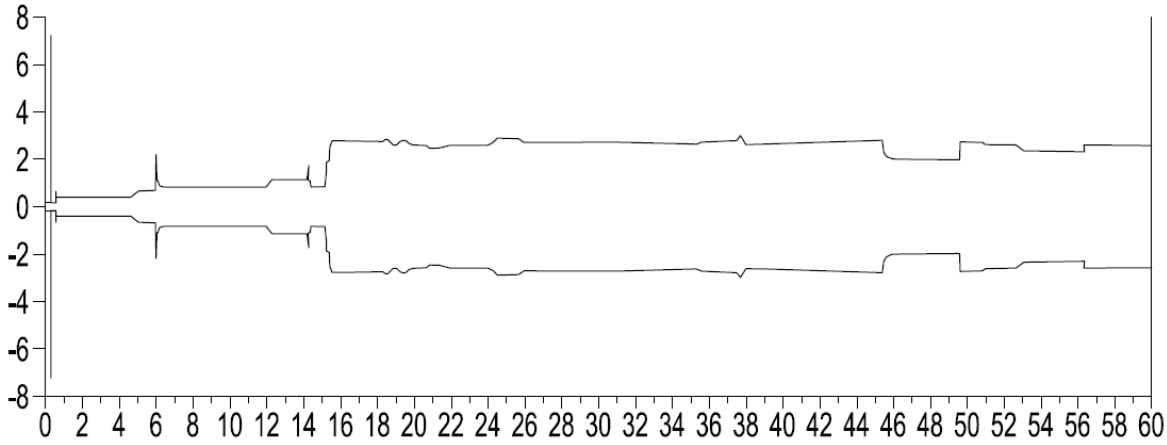


Figure 7.4 Peak-to-peak operating current waveform of a plasma TV for 60 seconds

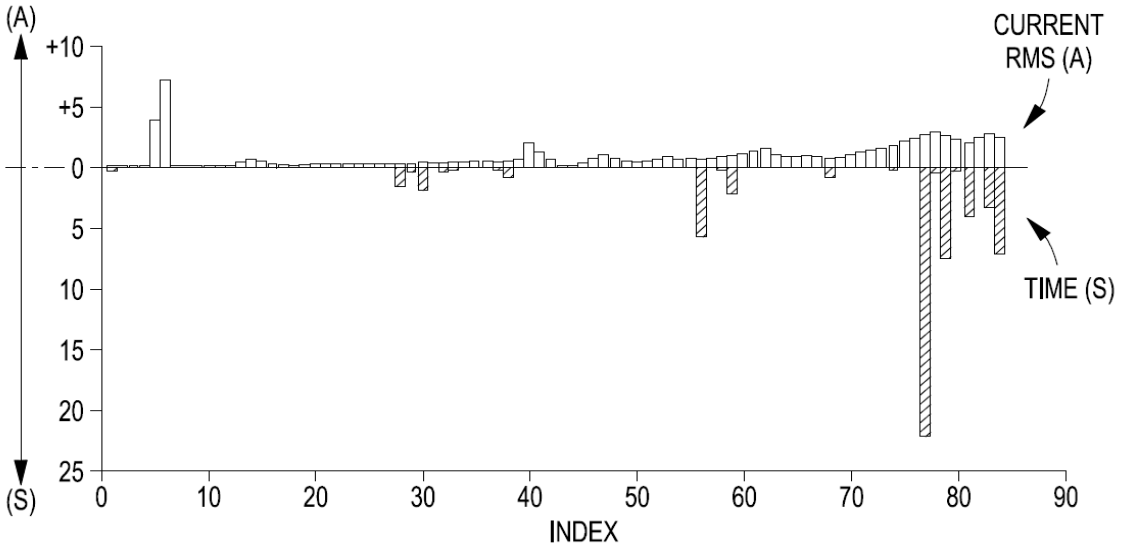


Figure 7.5 Corresponding states with associated values extracted from Figure 7.4

The following figure illustrates corresponding spike events, semi-steady states, steady states, and oscillation states extracted from the operating current waveform in Figure 7.4.

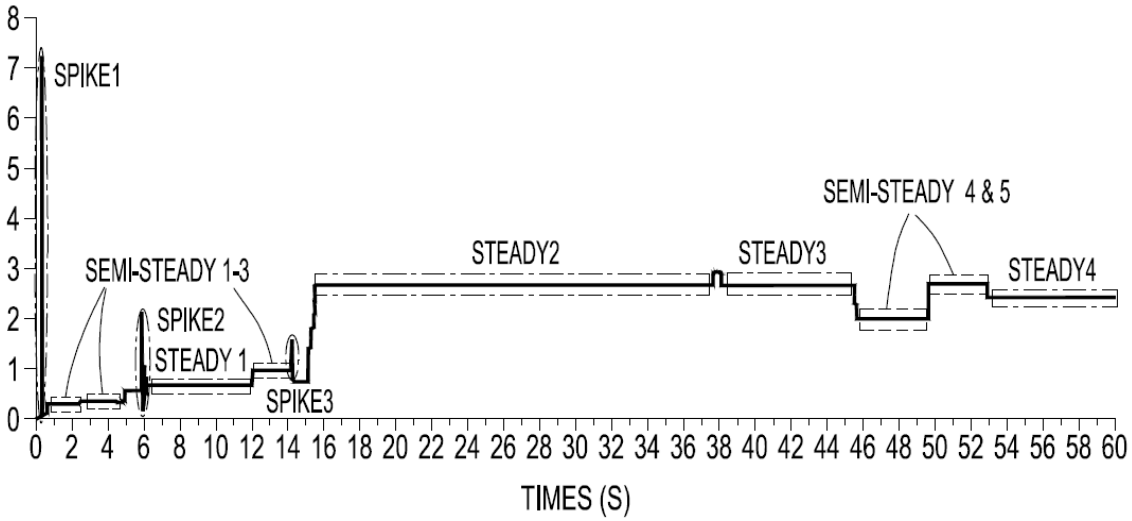


Figure 7.6 Corresponding spikes, semi-steady states, steady states, and oscillation states extracted from Figure 7.4

Note that the elemental states can be directly identified by the T_k values shown in Figure 7.5.

It can be observed in Figure 7.6 that there are several spikes when the plasma TV is turn on (between 0-15 seconds), which should be detected in order to correctly identify the startup mode. After a notable step raise of the RMS current from around 1 A to 3 A, the plasma TV gradually operates in active mode but the RMS current values vary with time. Accordingly, in Figure 7.5 the representing FSM stays in one steady state for over 20 seconds, in 3 steady states for over 5 seconds, in 5 semi-steady states for over 1 second, and in over 70 states for a very short period of time.

The corresponding FSM representation of the operating waveform of the plasma TV is shown in the following figure.

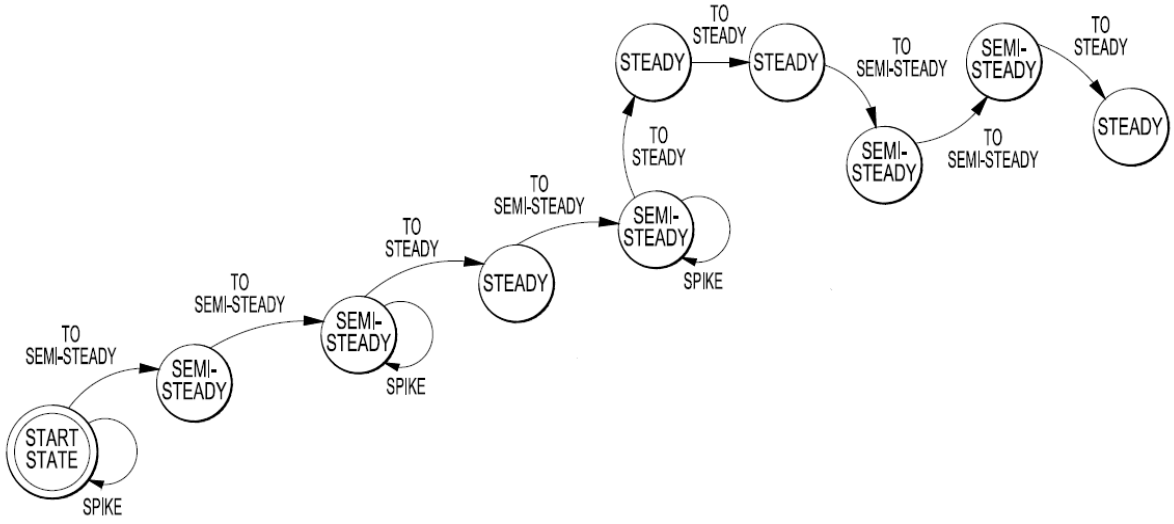
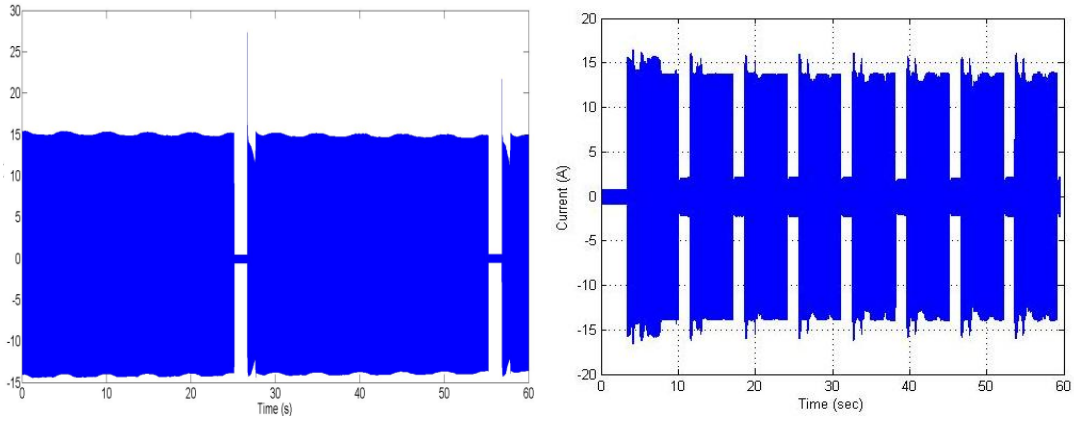


Figure 7.7 Corresponding spike events, semi-steady states, steady states, and oscillation states extracted from Figure 7.6

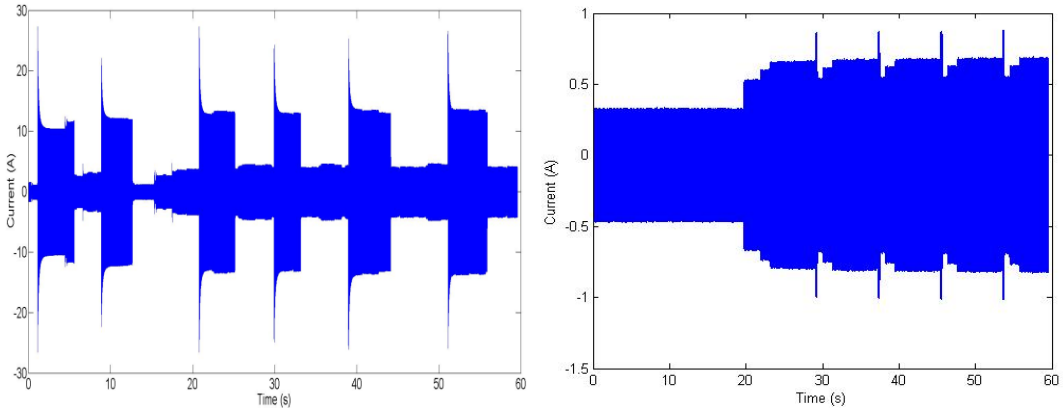
7.4 CLASSIFICATION AND IDENTIFICATION OF REPEATING ACTIONS AND MODES

As illustrated in the above examples, many electric loads consume a time-varying amount of power in active mode, which poses challenges to the identification of electric loads from long-term operating waveforms. Furthermore, many electric loads repeat certain actions which are separated by time intervals with non-identical length. In other words, there exist repeating patterns that represent operating actions that are similar but not identical. For various applications discussed in Chapters One and Two, it is desired to have a comprehensive understanding of repeating actions and operating modes during PEL operations as well as an efficient method to classify and identify these repeating patterns.

Several examples of repeating similar patterns in long-term operating waveforms are shown in the following figure.



(a) A microwave oven in reheat mode; (b) A printer in double-sided printing mode;



(c) A multi-function-device in printing mode; (d) A fax machine in faxing mode

Figure 7.8 Current waveforms of four office appliances in different operating modes

These results show that some electric loads possess almost identical waveforms when repeating certain operating actions or operating modes, such as the loads shown in Figures 7.8(a), 7.8(b), and 7.8(d). On the other hand, some electric loads repeat non-identical but similar waveforms as well as patterns, such as shown in Figure 7.8(c). For the purpose of accurate identification of electric loads and their operating modes, it is necessary to develop a set of features as well as an identification mechanism to detect the existence or non-existence of repeating similar patterns (and possibly how many times they repeat).

7.4.1 Repeating patterns in FSM representations

In order to reliably identify repeating patterns, it is necessary to first understand how they are repeated in the form of states (as well as their associated values) and events in the FSM representations. To extract information from a large number of states with

different values, the following characteristics are useful based on a study of a large set of real-world operating waveforms.

(1) The type of the states, i.e., semi-steady or steady.

(2) Associated I_k values for all steady and semi-steady states. Repeating similar patterns are observed to have similar I_k values.

(3) Associated T_k values for all steady and semi-steady states. Repeating similar patterns should spend similar amounts of time staying in each steady or semi-steady state.

(4) The occurrence of spikes before and after steady and semi-steady states.

As an example, the operating waveform of a microwave oven in reheat mode as shown in Figure 7.8(a) can be represented by a FSM with 3 semi-steady states and 2 steady states. The associated I_k (in A) and T_k (in seconds) are listed as follows:

Table 7.1 Associated I_k and T_k values of states in the FSM of operating waveform in Figure 7.8(a)

State-Type	I_k (A)	T_k (s)
Steady	15.36	25.1
Semi-Steady	0.47	1.5
Steady	15.37	27.6
Semi-Steady	0.47	1.6
Semi-Steady	14.6	2.1

The similar patterns "15.36-0.47" and "15.37-0.48" in I_k can be considered to be candidates that represent repeated operating actions, which match the waveform shown in Figure 7.8(a). Moreover, these two candidate patterns also have similar staying-time "25.1-1.5" and "27.6-1.6".

To summarize, a pattern which repeats in a similar manner during a fixed time length of observation consists of a sequence of steady and semi-steady states, each of which has similar values of current and staying-time as well as the same state types. For example, a repeating pattern in a sequence of two states "steady (15.36 A, 25.1 s) – semi-steady (0.47 A, 1.5 s)" can be observed twice in Figure 7.8(a).

Moreover, the spike events also play an important role in identifying repeating patterns as discussed in the next section.

Based on evaluating a large number of real-world PEL current waveforms as well as a comprehensive study on the front-end power supply topology design of most commercially available PELs, it is observed that there are three types of repeating patterns, described as follows.

7.4.2 Almost-identical repeating patterns

This type of repeating patterns contain repeating semi-steady or steady states whose associated I_k and T_k values have similar numerical values. In other words, a certain number of semi-steady or steady state repeats several times. Oscillation states and spike events may occur between these repeating semi-steady or steady states but they are not needed to identify this type of repeating patterns and thus generally not used for this pattern type.

A typical example is given in Figure 7.8(d), which can be modeled by a FSM with 1 semi-steady and 6 steady states. The associated I_k (in A) and T_k (in seconds) are listed in the following table.

Table 7.2 Associated I_k and T_k values of states in the FSM of operating waveform in Figure 7.8(d)

State Type	I_k (A)	T_k (s)
Steady	0.33	19.7
Semi	0.53	2.1
Steady	0.66	5.8
Steady	0.67	5.8
Steady	0.67	5.8
Steady	0.68	5.8
Semi	0.67	3.7

The results in Table 7.2 show that the steady state with I_k around 0.67 A and staying-time T_k around 5.78 seconds repeats four times. Note that the final steady state with peak current 0.67 A is not considered as repeated because its staying-time is only 3.73 s, which is caused by the cutoff of the observation waveform.

7.4.3 Step up/down repeating patterns

This type of recurring similar patterns usually repeats in a sequence of step up/down waveforms which can be represented by a sequence of semi-steady and steady states with step up/down I_k and T_k values. Similar to the case of almost identical repeating patterns

discussed above, oscillation states and spike events may occur but they are not utilized for identification in this case.

The operating waveform of a laser printer in double-sided printing mode shown in Figure 7.8(b) falls into this category, and can be modeled by a FSM with associated I_k and T_k values listed in the following table.

Table 7.3 Associated I_k and T_k values of states in the FSM of operating waveform in Figure 7.8(b)

State Type			After combining similar states		Sub-sequence
	I_k (A)	T_k (s)	I_k (A)	T_k (s)	
semi*	14.70	4.93	14.70	4.93	Sub-sequence-1
semi	1.96	1.23	1.96	1.23	
semi	13.96	1.1	13.96	5.13	Sub-sequence-2
semi	13.71	4.03			
semi	2.13	1.22	2.13	1.22	
semi	14.09	1.42	14.09	5.17	Sub-sequence-3
semi	12.58	3.75			
semi	2.00	1.47	2.00	1.47	
semi	13.72	4.22	13.72	4.22	Sub-sequence-4
semi	2.17	1.20	2.17	1.20	
semi	13.72	3.25	13.72	3.25	Sub-sequence-5
semi	1.88	1.17	1.88	1.17	
semi	13.54	3.2	13.54	3.2	Sub-sequence-6
semi	2.00	1.2	2.00	1.2	
semi	13.57	3.27	13.57	3.27	Sub-sequence-7
semi	1.98	1.23	1.98	1.23	
semi	13.42	3.2	13.42	3.2	

* *semi* denotes semi-steady states

After combining adjacent semi-steady states with almost identical state values (i.e., semi-steady states with state values 13.96 A and 13.71 A), the repetitive sub-sequences of states indicated in Table 7.3 represent the recognizable repetitive patterns.

In this example, seven sub-sequences with a step-down pattern in both state values and state durations can be observed and detected. The first several step-down sub-sequences have relatively higher pre-step state values and state durations, i.e., 14.70 A, 13.96 A, and 14.09 A, as the printer transits from standby mode to active mode. The latter three sub-sequences have relatively smaller but almost identical pre-step state values, 13.72 A/3.25 s, 13.54 A/3.2 s, and 13.57 A/3.27 s, as the printer is in a stable active mode. The post-step state values remain close to 2 A and the post-step state durations remain close to 1.2 s.

7.4.4 Spike-lead repeating patterns

Some current waveforms include repeating similar pattern for which the "in-between" time durations of their occurrences are not uniform. These types of repeating patterns usually starts with a notable spike and thus can be detected by comparing the variance in the ratio of the step up/down current values of the semi-steady and steady states after spike events.

An example of spike-lead repeating patterns is given in Figure 7.8(c), in which six printing jobs are performed during this observation. Accordingly, six repeating patterns can be detected in the representation of a sequence of a spike event followed by a semi-steady state (with similar but not identical associated I_k and T_k values).

Table 7.4 Associated I_k and T_k values of states in the FSM of operating waveform in Figure 7.8(c)

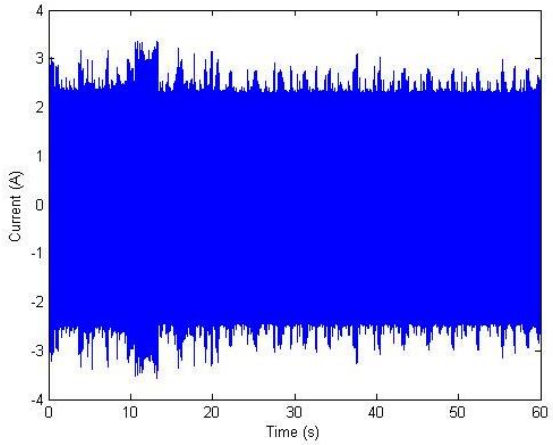
State-Type	I_k	T_k	Repetitive Sub-sequences
spike	23.21	0	Subsequence-1
semi*	11.08	2.87	
semi	11.02	1.07	
semi	2.98	2.18	
spike	20.32	0	Subsequence-2
semi	12.75	3.32	
semi	1.18	2.38	
semi	2.69	1.88	
semi	3.36	1.75	
semi	3.76	1.48	
spike	23.22	0	Subsequence-3
semi	12.96	3.55	
semi	4.24	4.25	
spike	20.80	0	Subsequence-4
semi	13.04	2.38	
semi	4.08	2.45	
semi	4.50	2.53	
spike	21.97	0	Subsequence-5
semi	13.85	4.5	
semi	4.10	4.5	
semi	4.60	2.43	
spike	22.97	0	Subsequence-6
semi	14.17	4.45	
semi	4.15	3.63	

* *semi* denotes semi-steady states

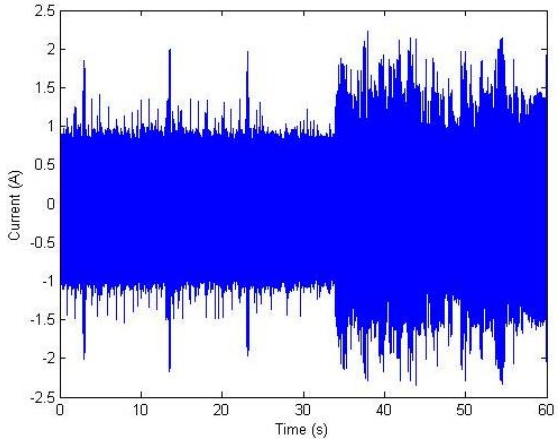
7.4.5 Summary of repeating patterns

To summarize, typical repeating similar patterns shown in Figure 7.9(a)-(d) illustrate cases from three major classes of patterns. The existence and non-existence of these three classes of repeating patterns can be accurately determined using I_k and T_k values of sequences of spike events, semi-steady and steady states. In this dissertation, it is proposed to start with each spike and check the existence of a repeated sequence of states. This method works well for most electric loads. For example, within the sequence "spike-steady-spike-spike-spike-steady-spike-steady" the subsequence "spike-steady" occurs three times. In addition, to make sure that this repeated sequence of states is indeed generated from repeated operation of electric loads, the associated I_k and T_k values of each semi-steady or steady state are also compared.

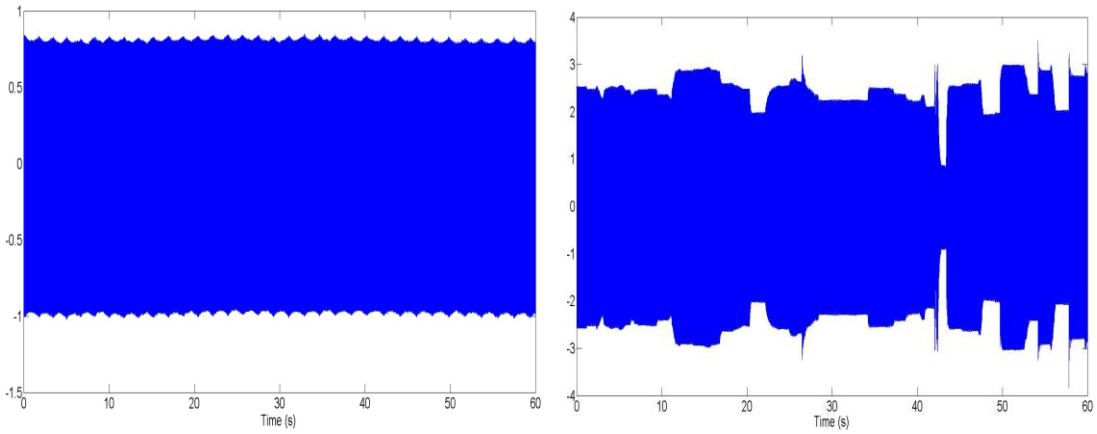
Several examples of operating waveforms that do not contain notable repeating pattern are given in the following figure.



(a) A desktop computer in active mode;



(b) A laptop in active mode;



(c) An LCD monitor in active mode; (d) An LCD TV in active mode.

Figure 7.9 Four PELs in active mode as examples of non-existence of repeating patterns

7.5 FEATURES EXTRACTED FROM THE FSM REPRESENTATION

The next problem is how to distinguish between different FSMs in a numerical way. Different FSM representations should preferably be converted into a set of numerical features that can be used to indicate the identity and operating modes of electric loads in real-time. A novel set of numerical features is proposed and listed as follows.

(1) Number of spikes

(2) Number of semi-steady states

(3) Number of steady states

(4) $$\frac{\text{Total time in semi-steady states}}{\text{Length of the operating waveform}}$$

(5) $$\frac{\text{Total time in steady states}}{\text{Length of the operating waveform}}$$

(6) Number of states per time window

(7) Existence or non-existence of repeating patterns

In contrast with *absolute* values that are directly measured from relevant FSM representations such as features (1)-(3), features (4)-(6) are *relative* values that correspond to the intrinsic characteristics of the operating waveforms. Features (4) and (5) reflect how "steady" an operating waveform is. For an operating waveform that holds relatively constant values of RMS current, features (4) and (5) should be quite a small number. Otherwise, features (4) and (5) would be quite large. Moreover, feature (6) indicates how long oscillations last in an operating waveform. For a waveform that oscillates all the time, such as Figure 7.9(b), feature (6) would be a large number.

The above seven features of a selected set of important office electric loads are summarized in the following table. Over 99% identification accuracy rate is achieved in tests using these proposed features.

Table 7.5 Summary of Features of a selected set of important office electric loads

	Computer	Office appliance		Microwave	Monitors and TV	
		Pattern 1	Pattern 2		Start-up transient	Steady-State
Number of Spikes	0-11	Typically >10	<10	<5	<5	Only one state
Number of Semi-steady states	<25, nonzero	Typically around 10	1-16	1-5	5-10	
Number of Steady states	<5, typically 0	0	1-10	1-5	0-2	
Time in semi-steady/Total	< 0.6	Typically <0.7	0.4	<0.2	<0.2	
Time in steady/Total	<0.4 Typically 0	0	0.15-0.8	>0.8	>0.8	
Number of states per time window	>5	Typically 6-10	Typically <1.5	<1	~1	
Repeated Patterns	None	Yes		2-5	None	

7.6 CHAPTER SUMMARY

This chapter has proposed an efficient method to represent long-term operating waveforms of electric loads by finite-state machines (FSMs). The major advantage of a FSM representation lies in its capability to detect repeated patterns, reduce duplicate states and transitions by allowing self state transition, and extract time values associated with each state. Each repeated pattern may not be exactly the same and the time durations in-between the repeated patterns are also not exactly identical in practice, but the FSM representation can extract the common pattern by state transitions and eliminate the effect of time by self state transitions.

A set of numerical features are proposed for indicating the identity and operating modes of electric loads in real-time. A high identification accuracy rate is achieved in tests on a large set of real-world date using the proposed features.

There are two potential problems of the proposed state-machine representation of long-term waveforms. First, the performance of the proposed method may be affected by the accuracy of sampled data. For example, if the sampled data points are only with accuracy of one decimal place, the difference between adjacent RMS current values may be affected and the proposed algorithm may give a different set of states with different staying time values. Second, detecting spike events depends highly on parameters. In this dissertation, the ratio between two adjacent RMS current values is set to be 1.85 to detect spikes, which is based on trials. If this ratio is set to be much larger, there would be only a very small number of spikes detected and important spike events may be missed. On the other hand, if this ratio is set to be a smaller number, there could be too many spike

events. For a different data set, this ratio needs to be carefully selected based on trials and applications.

CHAPTER 8 CONCLUSIONS AND CONTRIBUTIONS

8.1 INTRODUCTION

This chapter concludes this dissertation by first reviewing the objective of this research and proposed solution framework, summarizing the technical development and major contributions of Chapters Three to Seven, and finally presenting the list of outcomes such as U.S. patent applications, journal papers, and conference papers directly contributed by Chapters Three to Seven of this dissertation.

Finally, section 8.5 presents several future research directions that can extend the PEL identification problem as well as proposed techniques to other applications.

8.2 CONCLUSIONS OF THIS DISSERTATION

The objective of this dissertation is to develop non-intrusive, accurate, robust, and applicable PELs identification algorithms based on voltage and current measurements. A two-level framework for PELs identification is proposed: first classify an unknown PEL into one of the predefined classes and then identify the unknown PEL within this class.

Chapter Three proposes a low computational-cost but yet accurate algorithm to extract features from the voltage and current waveforms of PELs for classification and identification. Instead of utilizing digital signal processing and frequency domain analysis, the proposed feature extraction algorithm first abstracts the similarity of voltage-current (V-I) trajectories between different PELs and then maps the V-I trajectories to cell grids with binary cell values. Graphical features can then be extracted for many purposes. The proposed method significantly reduces the computational cost compared to existing frequency-domain signature extraction and analysis methods. It is

shown in Chapter Four that an average of over 99% of success rate can be achieved using the proposed features.

Chapter Four proposes the extension of the basic SOM to a supervised manner for classification of PELs. The supervised SOM (SSOM) can classify a large amount of PELs into several groups. Chapter Five extends the framework in Chapter Four and presents a novel combination of the SSOM and the Bayesian framework to function as a hybrid identifier and provide the probability of an unknown PEL belonging to a known group. Extensive research has been carried out to test the performance of PELs classification and identification using the SSOM, and the testing results are satisfactory.

In Chapters Six and Seven, static and dynamic methods are considered for in-category PEL identification, respectively. For the static method, a hybrid SSOM/SVM identifier is proposed. The proposed hybrid identifier utilizes the power of the supervised Self-Organizing Map (SSOM) classifier for PELs to first classify the large amount of PELs models into several clusters. Within each cluster, a more accurate identification decision is made by the well establish multi-class one-against-all SVM classifier.

For dynamic modeling, a continuous current waveform is first converted into a quantized sequence of current values. This sequence is then transformed into a finite-state machine (FSM) consists of different types of states and transitions between states. A set of features can be extracted from the formulated finite-state machine, which is utilized to represent and identify electric appliances with different operating principles and modes.

The overall testing results of different features and proposed methods are summarized in the following table.

Table 8.1 Summary of testing results in this dissertation

Chapter	Method	Purpose	Features	Average Success Rate
Chapter Four	SSOM	Classification into categories	- Proposed graphical features in Chapter Three - A set of 13 selected features	> 95%
Chapter Five	SSOM	Probabilistic classification into categories	- A set of 13 selected features	>95% based on MAP
Chapter Six	SVM	In-Category identification	- A set of 13 selected features	>95%
Chapter Seven	State Machine	In-Category identification	- A set of 7 features proposed in Chapter Seven	>95%

8.3 CONTRIBUTION OF THIS DISSERTATION

The main contributions of this dissertation lie in several different aspects of a novel two-level PELs classification and identification algorithm, summarized as follows:

(1) Robustness: the proposed SSOM identifier first classifies the large amount of real-world PELs into clusters and then identifies their types within each class. In this manner, the diversity within each type of PELs and the similarity between different types of PELs can be handled in a robust manner.

Another contribution is a novel hybrid method by combining the SSOM framework and the Bayesian framework to utilize the training information to provide probabilistic identification results;

(2) Accuracy: with each classified class, both static and dynamic methods are proposed to achieve high successful identification rates under all scenarios but modeling the steady-state characteristics as well as dynamic performance and operating principles of PELs;

(3) Adaptiveness: the proposed framework can learn from user inputs or feedback, update classification and identification rules if necessary, and include *a priori* information and required identification granularity;

(4) Intelligence: the proposed method also has the capability to investigate the applicability of the extracted signatures/patterns for effective PELs disaggregation, and identifying an unknown PEL to a certain level of granularity.

8.4 OUTCOMES OF THIS DISSERTATION

The literature review and research work presented in Chapters Two to Seven of this dissertation have resulted in a number of publications and U.S. patent applications, listed as follows:

8.4.1 U.S. Patents

[1] Bin Lu, Ronald G. Harley, **Liang Du**, Yi Yang, and Sharma K. Santosh, Prachi Zambare, and Mayura Madane, U.S. Patent Application 13/304,758, Publication US20130138651, "*System and method employing a self-organizing map load feature database to identify electric load types of different electric loads*", filed on November 28, 2011. Patent publication date: May 30, 2013.

[2] Bin Lu, Ronald G. Harley, **Liang Du**, Yi Yang, and Sharma K. Santosh, Prachi Zambare, and Mayura Madane, U.S. Patent Application 13/597,324, "*System and method for electric load identification and classification employing support vector machine*", filed on August 29, 2012.

[3] Yi Yang, **Liang Du**, and Dawei He, U.S. Patent Application 13/920,602, "*System and method for instantaneous power decomposition and estimation*", filed on June 18, 2013

[4] Yi Yang, **Liang Du**, and Dawei He, U.S. Patent Application 13/912,819, "*System and method employing graphical electric load categorization to identify one of a plurality of different electric load type*", filed on June 7, 2013

[5] **Liang Du**, Yi Yang, Ronald G. Harley, Thomas G. Habetler, and Dawei He, U.S. Patent Application 13/908,263, *"Method and system employing finite state machine modeling to identify one of a plurality of different electric load types"*, filed on June 3, 2013

8.4.2 Journal Paper

[1] **Liang Du**, Jose A. Restrepo, Yi Yang, Ronald G. Harley, and Thomas G. Habetler, "*Nonintrusive, Self-Organizing, and Probabilistic Classification and Identification of Plugged-In Electric Loads*", IEEE Transactions on Smart Grid, vol. 4, issue 3, pp. 1371-1380, 2013

[2] Dawei He, **Liang Du**, Yi Yang, Ronald G. Harley and Thomas G. Habetler, "*Front-End Electronic Circuit Topology Analysis for Model-Driven Classification of Appliance Loads*," IEEE Transactions on Smart Grid, vol. 3, no. 4, pp.2286-2293, Dec 2012

[3] **Liang Du**, Yi Yang, Ronald G. Harley and Thomas G. Habetler, "*Real-Time Identification of Electric Loads Using Long-Term Operating Waveforms*," under review, IEEE Transactions on Smart Grid

[4] **Liang Du**, Yi Yang, Ronald G. Harley and Thomas G. Habetler, "*Characterization and Identification of Electric Loads by Binary Voltage-Current Trajectory Mapping*," to be submitted to IEEE Transactions on Smart Grid

8.4.3 Conference Paper

[1] **Liang Du**, Yi Yang, Dawei He, Ronald G. Harley, Thomas G. Habetler, and Bin Lu, "*Support Vector Machine Based Methods For Non-Intrusive Identification of Miscellaneous Electric Loads*", In Proceedings of the 38th Annual Conference of the IEEE Industrial Electronics Society (IECON 2012), Oct 25-28, Montreal, Quebec, Canada

[2] Dawei He, **Liang Du**, Yi Yang, Ronald G. Harley, and Thomas G. Habetler,, "*A Model-Driven Taxonomy of Appliance Loads: Front-End Electronic Circuit Topology Analysis*", In Proceedings of the IEEE Energy Conversion Congress and Exposition (ECCE), p.1228-1232, Raleigh, NC, USA, September 15-20, 2012

[3] **Liang Du**, Dawei He, Yi Yang, Jose A. Restrepo, Ronald G. Harley, and Thomas G. Habetler, "*Self-Organizing Classification and Identification of Miscellaneous Electric Loads*", In Proceedings of the IEEE Power & Energy Society General Meeting (PES-GM 2012), July 22-26, San Diego, CA,USA

[4] Yi Du, **Liang Du**, Bin Lu, Ronald G. Harley, and Thomas G. Habetler, "*A Review of Identification and Monitoring Methods for Electric Loads in Commercial and Residential Buildings*", In Proceedings of the IEEE Energy Conversion Congress and Exposition (ECCE 2010), p.4527-4533, Atlanta, GA, USA, September 12-16, 2010

8.4.4 Contribution of chapters

The contribution of each chapter of this dissertation is listed in the following table.

Table 8.2 Contribution of each chapter of this dissertation

Chapters	Patents Listed in Section 8.4.1	Journal Papers Listed in Section 8.4.2	Conference Papers Listed in Section 8.4.3
Two			Paper [2]
Three	Patents [3] and [4]	Papers [2] and [4]	Paper [4]
Four	Patent [1]	Paper [1]	Paper [3]
Five			
Six	Patent [2]		Paper [1]
Seven	Patent [5]	Paper [3]	

8.4.5 Other Contribution

The following journal and conference papers, as outcomes of class projects during my graduate study, have been published or under review.

[1] **Liang Du**, Santiago Grijalva, and Ronald G. Harley, "*Distributed Potential-Game Formulation of Constrained Economic Dispatch with Guaranteed Convergence*", under review, IEEE Transactions on Control of Network Systems

[2] **Liang Du**, Lijun He, and Ronald G. Harley, "*A Survey of Methods for Shunt Capacitor Banks Placement in a Distorted Power Network*", In Proceedings of

The 38th Annual Conference of the IEEE Industrial Electronics Society (IECON 2012), Oct 25-28, Montreal, Quebec, Canada [120]

[3] **Liang Du**, Santiago Grijalva, and Ronald G. Harley, "*Potential-Game Formulation of Optimal Power Flow problems*", In Proceedings of the IEEE Power & Energy Society General Meeting (PES-GM 2012), July 22-26, San Diego, CA [121]

8.5 FUTURE WORK

8.5.1 Implementation and field tests

Techniques proposed in Chapters Three, Four, and Seven have been implemented in different product prototypes developed at Eaton Global Research and Technology center. Field tests of prototypes with techniques from this dissertation have been scheduled to take place in the National Renewable Energy Laboratory (NREL). Test results and feedback can be analyzed to further improve the proposed framework.

8.5.2 PEL disaggregation

The PEL identification problem studied in this dissertation is based on the assumption that a set of voltage and current waveforms is available for each PEL. However, in some cases, such waveforms for each PEL are not available. For example, in the case of multiple PELs are plugged into the same outlet or strip, measured current waveform is an aggregated signal of multiple PELs. The problem of identifying multiple plugged-in electric loads from mixed voltage/current waveforms is typically known as "load disaggregation" in the literature [70, 71, 122]. Given only this aggregated current waveform, the PEL disaggregation problem differs from the (single) PEL identification

problem in the sense that it is not practical or possible to follow the PEL identification procedure as shown in Figure 2.1 to solve the PEL disaggregation problem due to the following reasons:

(1) The number of commercial PELs being used in real-world is enormous. Moreover, the number of different combinations of PELs increases exponentially with the number of available PELs. Unlike in the PEL identification problem, it is impossible to build and maintain a database of aggregated waveforms of PELs with known identities (model and type) for the PEL disaggregation problem.

(2) Continuing with the above discussion with database of PELs with known identities, this database provides *a priori* information which relates voltage/current waveforms to PEL identities. The solution to the PEL identification problem proposed in Chapters Three to Seven of this dissertation does not require other *a priori* information.

On the other hand, the PEL disaggregation problem does require other *a priori* information including, but not limited to, the number of PELs that are connected at a common supply source and which PELs are currently in active mode. Without this *a priori* information, it is not reasonable to adopt the techniques developed for the PEL identification, i.e., comparing with a database of PELs with known identities even if such a database exists.

For instance, an aggregated current waveform may be contributed by four PELs in active mode and two PELs in standby/sleep mode (and thus not contributing much). This aggregated waveform is then compared to a known database

(assume it exists) and found to be very similar to an aggregated current waveform consisting of five PELs in active modes. However, it is not a reasonable or convincing solution to match five PELs with known identities to six unknown PELs.

(3) A unique difficulty in the PEL disaggregation problem is that it is sometimes not possible to identify several PELs of the same category which are plugged into a common supply source.

For example, considering the case that one space heater (500 W), one portable fan (60 W), and one incandescent light (100 W) are plugged into a common power strip. These three PELs all belong to category R, i.e., resistive or linear load. The aggregated current waveform is almost sinusoidal. In this case, it is not possible to tell whether it is a portable fan (60 W) or another incandescent light (60W) because either of them aggregated with the other two PELs would present the same sinusoidal aggregated current waveform.

(4) Another difficulty in the PEL disaggregation problem is that some PELs that consume a small amount of power may be dominated by other PELs consuming much higher amount of power.

For instance, if a space heater (800 W) is aggregated with a cell phone (10 W) or a laptop computer (50 W), the aggregated waveform is dominated by the space heater (and thus almost sinusoidal). In this case it is nearly not possible to detect the existence of the cell phone or the laptop computer.

To summarize, the PEL disaggregation problem is much more complicated and challenging than the PEL identification problem. As an extension of this dissertation, the

FSM modeling method proposed in Chapter Seven can be extended to a timed FSM for the purpose of event detection for PEL disaggregation, in which the starting time and ending time associated with each state are also recorded besides the overall staying-time in each state. An example of aggregated current waveform is given in the following figure.

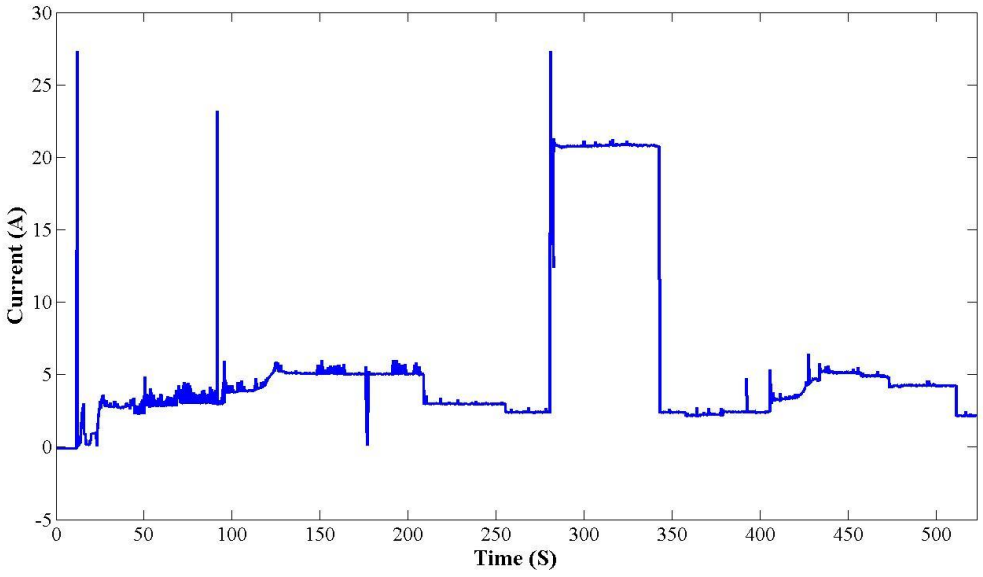


Figure 8.1 Measured aggregated current waveform

Using the previously proposed FSM presentation of long-term observations in Chapter Seven, the above current waveform can be represented by a sequence of semi-steady and steady states, whose associated current values and staying-time values are shown in the follow table.

Note that the spike events are not shown in the following figure for notational convenience.

Table 8.3 Timed sequence of semi-steady and steady states represent Figure 8.1

States	Current (A)	Staying-time (s)	Staying from (s)	Staying until (s)	Events
standby	n/a	11.88	0.017	11.9	<= At 11.9 s, LCD Monitor ON
semi	0.931	2.9	20.12	23.0	<= At 23.6 s, desktop computer ON
semi	3.364	1.467	26.22	27.67	
semi	2.825	4.317	28.77	33.07	
steady	**2.82**	**8.25**	33.95	42.18	
semi	2.334	1.067	45.18	46.23	
semi	2.434	1.383	46.55	47.92	<= At 50.3 s, lamp ON
semi	2.928	2.133	56.15	58.27	
semi	3.115	1.817	61.75	63.55	
semi	3.143	2.067	63.62	65.67	
semi	3.122	1.183	66.15	67.32	
semi	3.107	1.75	84.78	86.52	
semi	3.096	1.083	89.73	90.8	
semi	3.007	2.95	91.55	94.48	<= At 95.6 s, portable fan ON
semi	3.85	2.25	96.67	98.9	
semi	3.801	1.2	98.95	100.1	
semi	3.805	2.883	100.2	103.1	
steady	**4.223**	**8.617**	104.8	113.4	
semi	4.091	4.567	113.5	118.1	<= At 113.3 s, projector ON
semi	4.512	3.883	118.1	122.0	
semi	4.978	2.033	122.0	124.0	
steady	**5.521**	**45.28**	124.0	169.3	
steady	**4.963**	**6.867**	169.3	176.1	
steady	**5.183**	**14.3**	177.4	191.7	
semi	5.183	1.317	191.7	193.0	
semi	5.207	1.267	193.5	194.8	
semi	5.169	3.567	194.8	198.4	
steady	**5.13**	**5.35**	198.5	203.8	
semi	5.649	1.183	203.8	205.0	
semi	5.051	1.383	205.0	206.4	

Table 8.3 continued

semi	4.998	2.2	206.5	208.7	<= At 208.7 s, projector OFF
semi	3.209	1.517	208.7	210.2	
steady	**3.058**	**8.083**	210.3	218.4	<= LCD monitor OFF
steady	**3.012**	**36.68**	218.4	255.1	<= At 255.1 s, lamp OFF
steady	**2.459**	**13.27**	255.1	268.3	
steady	**2.424**	**9.983**	268.4	278.4	
semi	2.389	2.467	278.4	280.9	<= At 280.88 s, a microwave ON
semi	14.31	1.533	281.0	282.5	
steady	**19.46**	**60.08**	282.7	342.7	<= At 340.72 s, the microwave OFF
steady	**2.482**	**14.65**	342.8	357.4	<= At 357.4 s, the fan OFF
steady	**2.176**	**6.433**	357.4	363.8	
steady	**2.265**	**6.817**	363.9	370.7	
steady	**2.255**	**6.733**	371.0	377.7	
steady	**2.382**	**14.52**	377.8	392.2	<= At 398.9 s, LCD monitor ON
steady	**2.473**	**13.32**	392.3	405.6	<= At 405.6 s, fan ON
semi	3.321	3.867	407.5	411.4	
steady	**3.284**	**8.4**	411.4	419.8	<= At 419.8 s, projector ON
steady	**3.651**	**5.4**	419.8	425.2	
semi	4.044	1.017	425.2	426.2	
semi	4.509	1.4	426.2	427.6	
steady	**4.447**	**6.233**	427.7	433.9	
steady	**5.208**	**39.27**	434.0	473.2	<= At 473.2S, fan OFF
steady	**4.216**	**38.07**	473.3	511.3	<= At 511.3S, projector OFF
steady	**2.33**	**11.33**	511.3	522.7	

Another possible extension of the features proposed in Chapter Three is to develop similar graphical features based on V-I trajectories to help to detect ON/OFF events for PELs disaggregation.

More details on these extensions to study the PEL identification problem is currently under development and is expected to be reported in the near future.

APPENDIX A DATA ACQUISITION SYSTEM AND LAB ENVIRONMENT

All the tests in this dissertation are based on real-world PEL data collected for this research. A data acquisition (DAQ) system designed with help from Dr. Jose A. Restrepo and built for data collection contains the following components:

(1) A printed circuit board (PCB) sensors module with current and voltage sensors, which measures the actual voltage and current waveforms of connected PELs. The measurements of the PCB sensors module are sent out as outputs through two BNC (Bayonet Neill–Concelman) type connectors and two cables, one for the voltage waveform and the other one for the current waveform.

(2) A set of *National Instrument* (NI) analog/digital data acquisition (DAQ) devices to sample the voltage and currents waveforms measured by the PCB sensors module in step (1). The input channels of the NI DAQ device are connected to the PCB sensors module through the two BNC cables. The output channels of the NI DAQ device are connected to computers through high-speed USB cables.

(3) A *National Instrument* (NI) LabView program running on computers to store the sampled data in a pre-defined format. This LabView program also has the capability of plotting real-time waveforms in programmed user interfaces (UIs).

A.1 PCB SENSORS MODULE

Two figures of the PCB sensors module viewed from different angles are shown in Figure A.1. The design details of this PCB sensors module presented in this section are described by courtesy of Dr. Jose A. Restrepo.

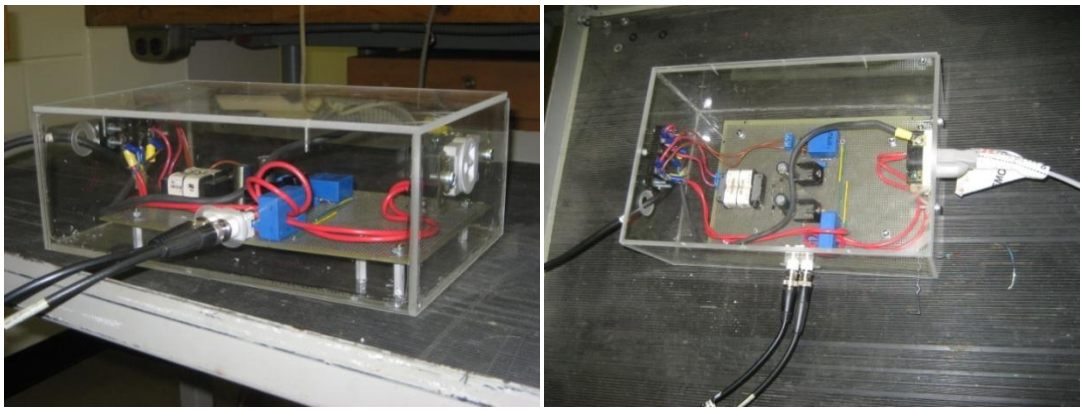


Figure A.1 Two views of the portable PCB sensors module

The PCB sensors module is contained in a plexi-glass box for insulation, protection, and portability. It has one input cable which can be plugged into any 120 V outlet and one output electrical socket into which a load can be plugged-in. This permits data to be captured at any location. Also, the module has two BNC output connectors to the NI DAQ device. The left BNC connector provides the voltage waveform and the right one provides the current waveform.

The PCB circuit schematic is given in the following figure.

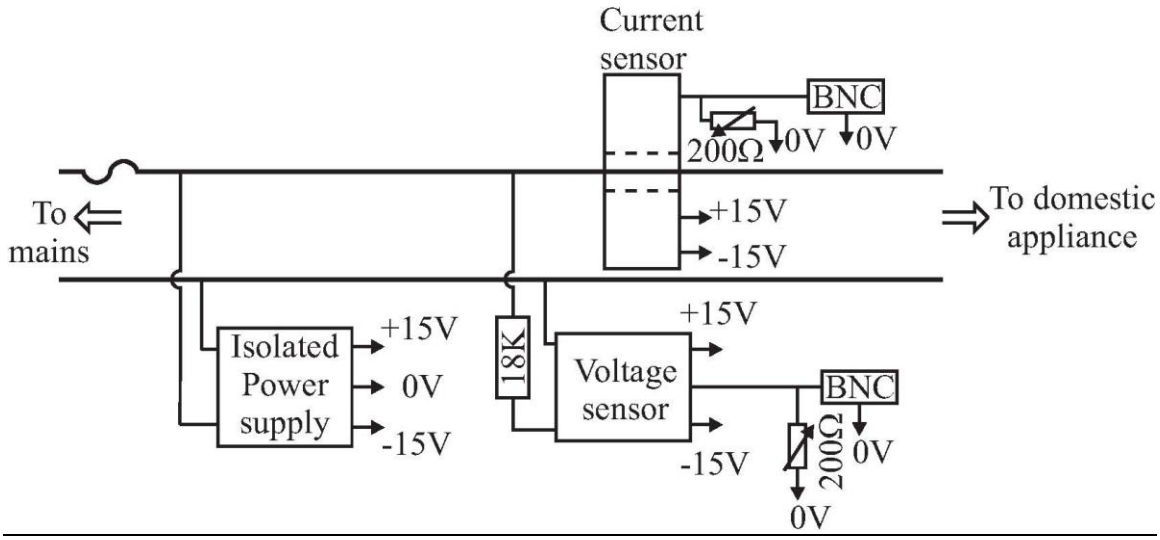
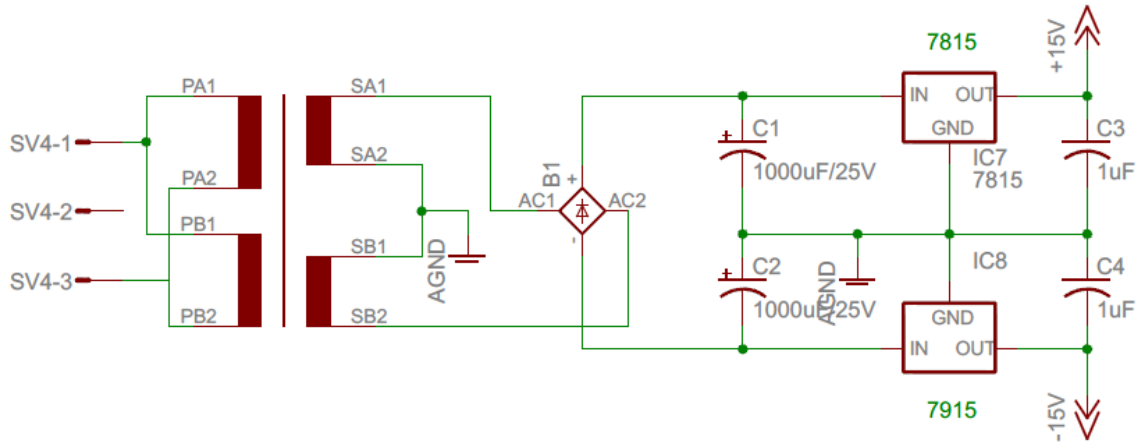


Figure A.2 PCB circuit schematic

A more detailed PCB design schematic is show in the following figure.



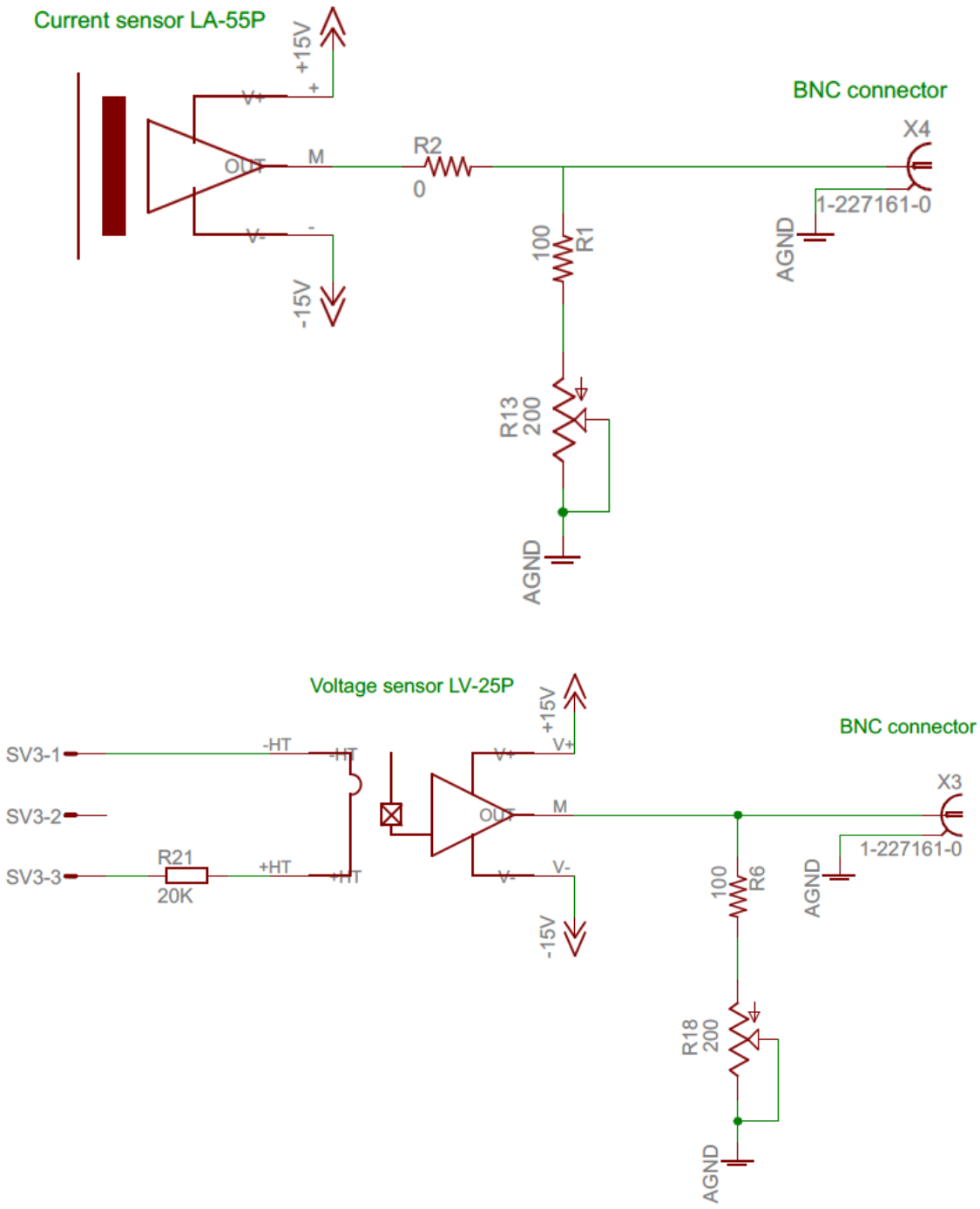


Figure A.3 PCB design schematic

A list of components appears below:

- (1) Tamura 3FS-328 transformer
- (2)73100-0069 BNC connector
- The rectifier bridge is made with four diodes (1n4007)
- (1) LEM LA-55P current sensor
- (1) LEM LV-25P voltage sensor
- (2) 100 Ohm ¼ W resistors
- (2) 200 Ohm ¾ W multiturn trimming potentiometer.
- (1) 18K, 3 W resistor
- (2) 1000uF/25V radial capacitor
- (2) 1uF/25V ceramic capacitor
- (1) LM7805 (1A positive regulator)
- (1) LM7905 (1A negative regulator)
- (2) TERMINAL BLOCK 3.5MM 3POS PCB.

Note that the wire through the current sensor has two loops or turns to increase the conversion ratio.

A.2 BNC CONNECTOR AND CABLE

The PCB sensors module and the NI DAQ device are connected through BNC connectors and cables, as shown in the following figure.

Figure A.4 BNC connectors and cables

A.3 NI DAQ DEVICE

Two different sets of NI DAQ devices have been used for data collection. In the early stage of this research, a set of NI SCXI-1000 DAQ system was used, which consists of the following three components:

(1) NI SCXI-1000: compact 4-slot AC-powered chassis

(2) NI SCXI-1305: 8 channel AC/DC coupling BNC terminal block

(3) NI SCXI-1141: 8 channel elliptic low-pass filter

The NI SCXI-1000 DAQ system is shown in the following figure.

Figure A.5 NI SCXI-1000 DAQ system

Some of the target PELs are large in size and weight, such as household refrigerators and TVs over 40 inches. It turned out that the NI SCXI-1000 DAQ system was physically too large and too heavy to be carried to different test sites. Therefore, a mobile DAQ system (see Figure A.6), the NI USB-6008 DAQ system was acquired for convenience and mobility.

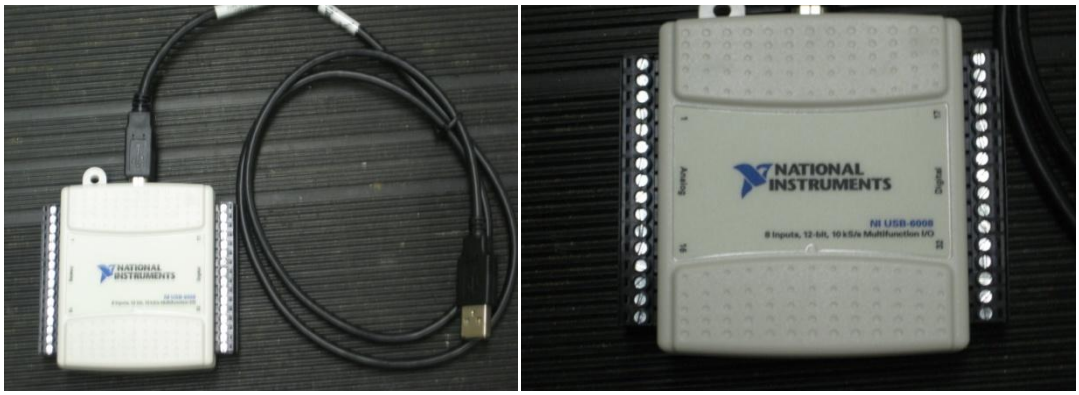


Figure A.6 NI USB-6008 DAQ system

The NI USB-6008 DAQ system has 12-bit accuracy with 8 analog inputs and 2 analog outputs, an interface of 12 digital I/O slots, is powered by a USB cable from computers and is compatible with LabView. This USB DAQ device supports single channel (at 10000 samples per second) or multiple (up to 8) channels of sequential sampling, where each channel is sampled at 1250 samples per second. For example, for this dissertation, two channels (current and voltage waveforms) are used and thus each channel has a max sampling rate 5000 samples per second. The actual time difference

between sampling of the two channels is not given by NI but is suggested to be negligible.

A.4 LABVIEW PROGRAM

A program was written in the NI LabView software (version Professional 8.5) to configure and acquire current and voltage measurements from the PCB sensors module. The following figure shows the block diagram of the LabView program written for such purposes. The sampling frequency is chosen as 30.72 k-Hz and the data is saved to multiple files on the computer.

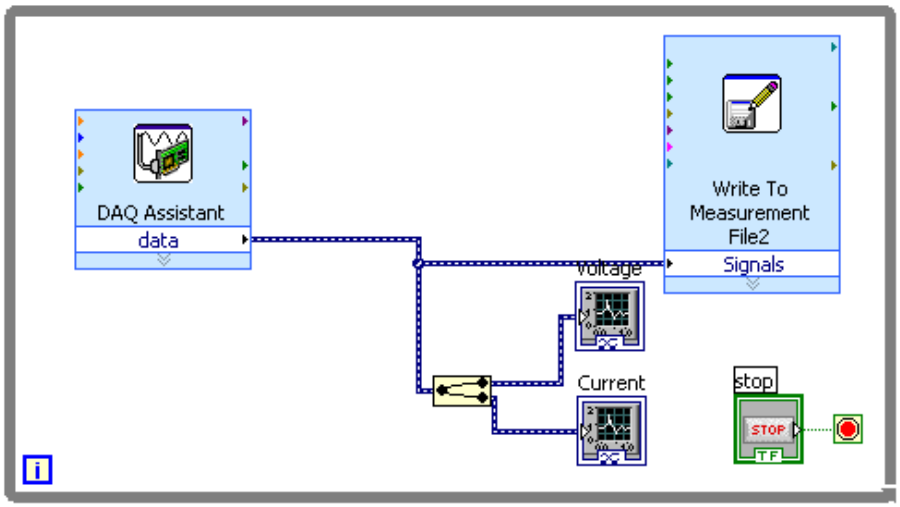


Figure A.7 Block diagram of the LabView data acquisition program.

There are usually two stages to sample the waveform. First the current and voltage are sampled for one minute to record the transition when the load is turned on. Then, after fifteen minutes, the current and voltage are sampled again for one minute to record the dynamic changes affected by any thermal or other issues.

A.5 CALIBRATION

The calibration of this entire DAQ system was validated by connecting the PCB sensor module to the utility. The voltage output channel coming from the PCB sensor module through the LabView system on a computer was compared with a direct measurement of this same output voltage using a *Tektronix* differential voltage probe. Results agree well.

The supplying voltage in calibration was provided by a calibrated *FLUKE* 177 True RMS multimeter which has an error of less that 1%, the offset of the current sensor was measured as -11mv with a maximum output capacity 7 V; the offset of the voltage sensor was measured as -6mv with a maximum output capacity 7 V. The final measuring scale of sensitivity is 1 V in the output for 5 A in the input of the current sensor, and 1 V in the output for 50 V in the input of the voltage sensor.

A.6 HARMONICS IN THE SUPPLY VOLTAGE

The 120 V ac utility power source in the lab testing environment contains noise and distortion, as shown in the following figure.

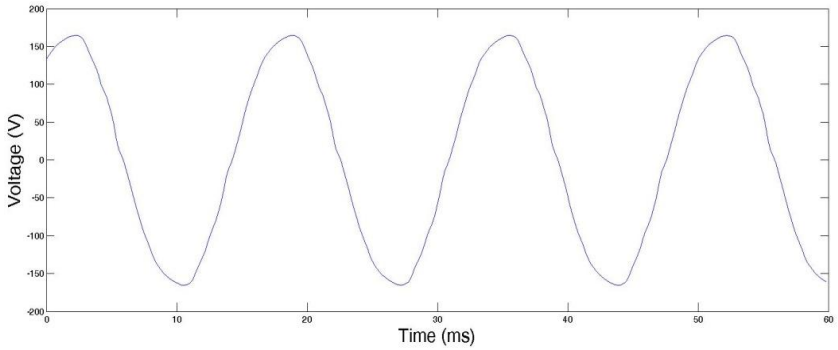


Figure A.8 Utility single phase voltage waveform in the Lab

Figure A.8 shows that the single-phase voltage supplied by utilities is not pure sinusoidal and notable distortion can be observed in the voltage waveform. Therefore, it is necessary to verify that the harmonic distortion in the utility supply voltage is within a certain limit defined by utility standards (e.g., IEEE standards [123, 124]).

A measurement of this voltage waveform without any load connected and using a *Yokogawa* WT1600 digital power meter revealed that the total background harmonic distortion (THD) is around 3.13%, as shown in the following figure.

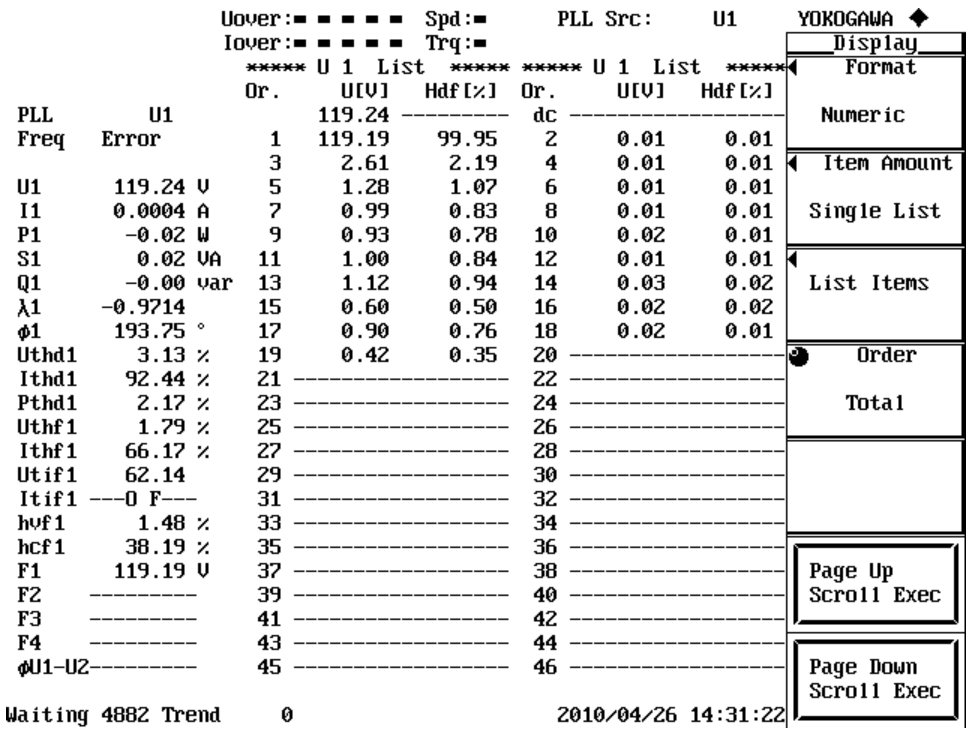


Figure A.9 Total harmonic distortion in voltage

The harmonic spectrum of the distorted utility-supplied voltage waveform is shown in Figure A.10, in which the vertical axis is of logarithm scale to better show harmonics with small magnitudes.

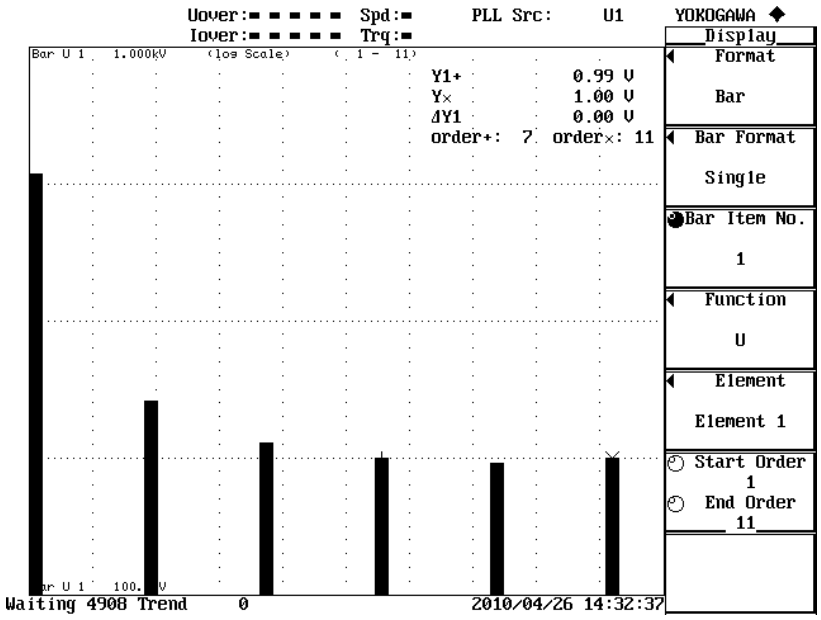


Figure A.10 Harmonic spectrum of the distorted ac utility-supplied voltage waveform

The following table summarizes Figures A.9 and A.10 and verifies that the harmonic distortion in utility supply voltage in the lab meets IEEE standards.

Table A.1 Harmonics in utility single phase voltage in the lab

Harmonic order	Harmonic magnitude (V)	Harmonic compared to RMS (%)	Harmonic compared to fundamental (%)
1st	119.19	99.95	100
3rd	2.61	2.19	2.19
5th	1.28	1.07	1.07
7th	0.99	0.83	0.83
9th	0.93	0.78	0.78
11th	1.00	0.84	0.84
13th	1.12	0.94	0.94
15th	0.60	0.50	0.50
17th	0.90	0.76	0.75
19th	0.42	0.35	0.35

APPENDIX B MAPPING OF REPRESENTATIVE V-I TRAJECTORIES TO BINARY CELL GRIDS

A set of V-I trajectories of 42 representative PELs from the database collected by the DAQ system presented in Appendix A and their mapped cell grids using the proposed algorithm in Chapter Three are shown in this section.

Table B.1 42 representative V-I trajectories mapped to binary cell grids

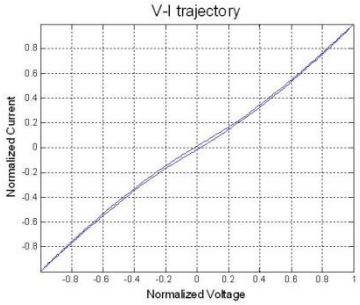
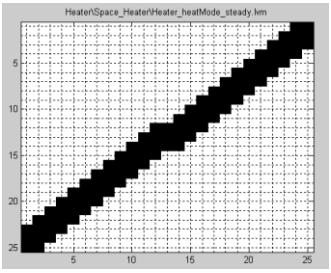
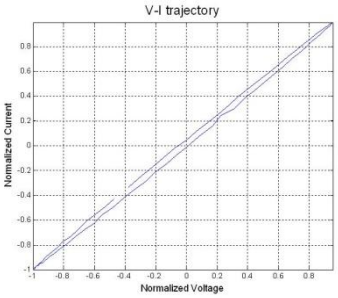
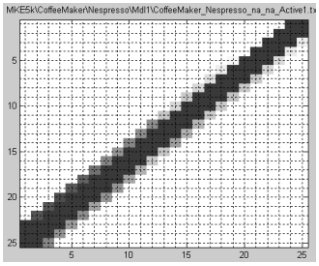
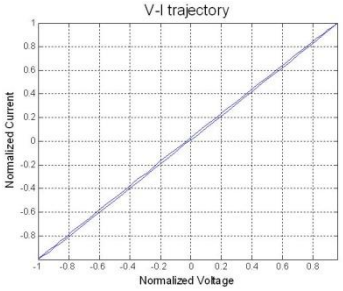
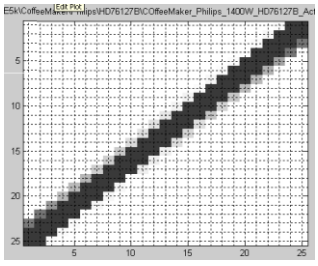
Number	PEL	V-I Trajectory	Mapped Binary Cell Grid
1	Space Heater		
2	Bread toaster		
3	Coffee Maker		

Table B.1 continued

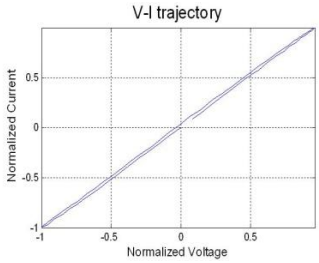
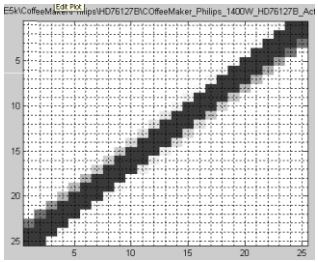
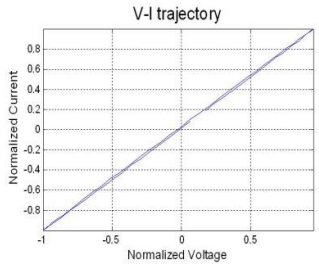
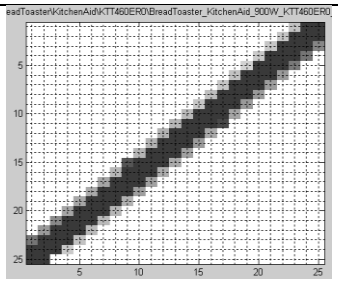
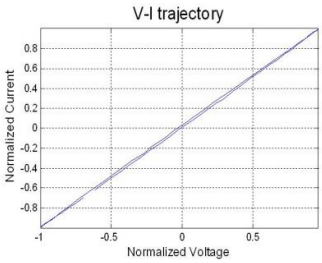
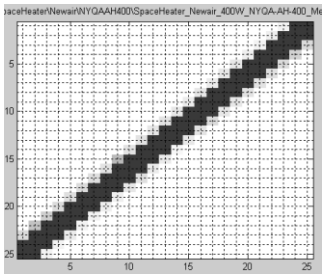
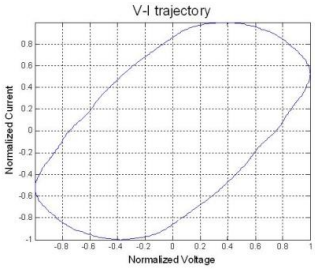
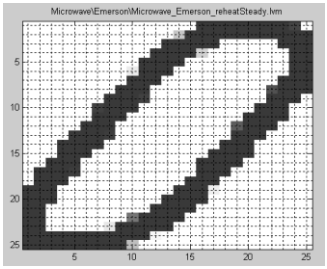
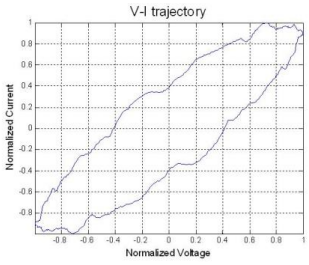
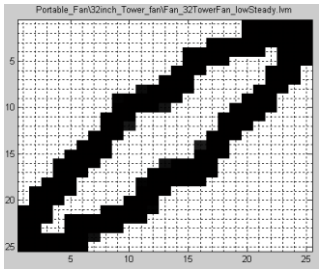
4	Space Heater 2		
5	Space Heater 3		
6	Space Heater 4		
7	9 Inch Portable Fan		
8	32 Inch Tower Fan		

Table B.1 continued

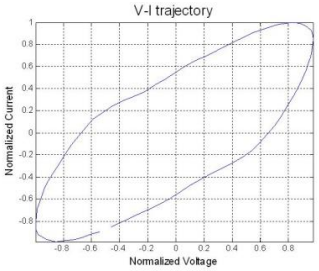
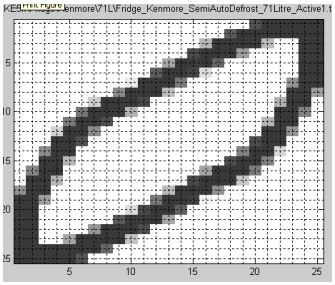
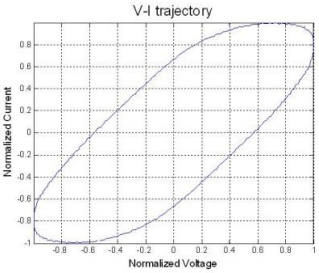
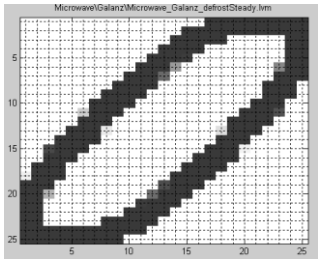
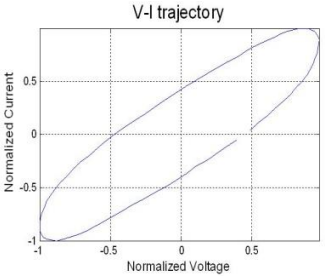
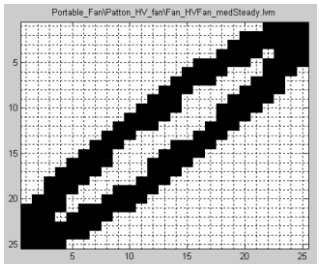
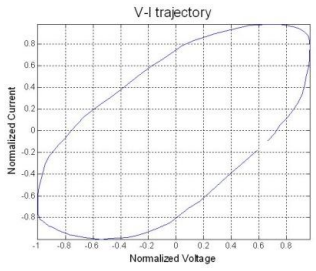
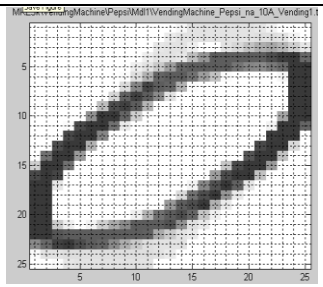
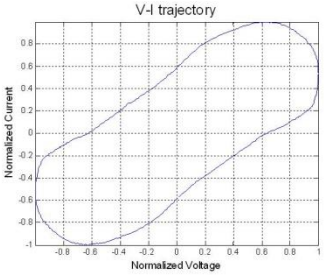
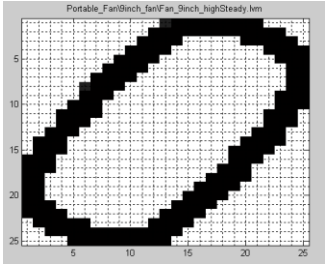
9	Portable Refrigerator		
10	Microwave		
11	Portable Fan		
12	Vending Machine		
13	Space Heater 5		

Table B.1 continued

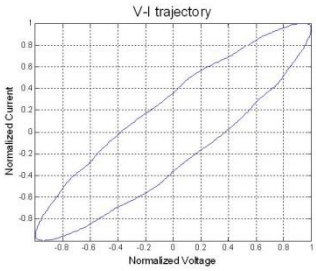
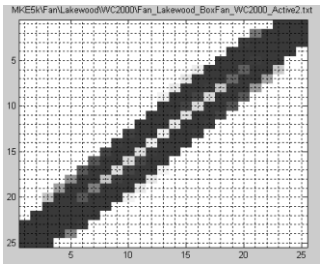
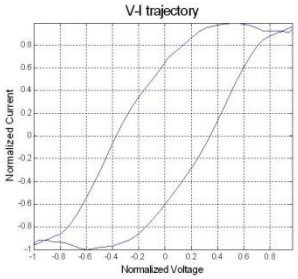
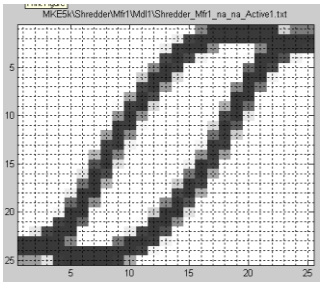
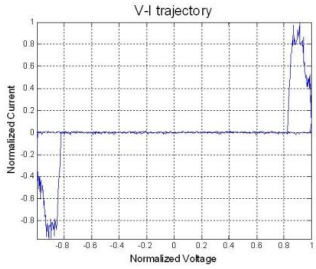
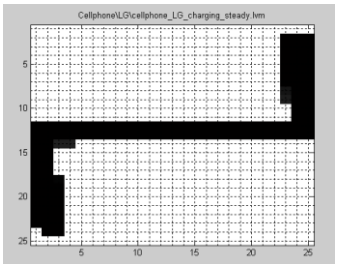
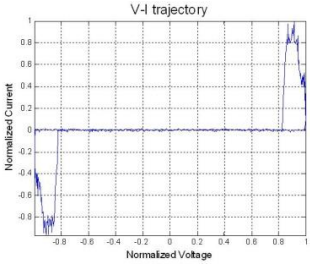
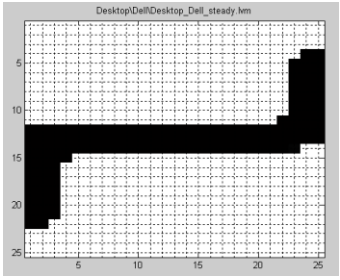
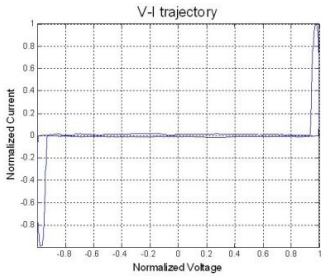
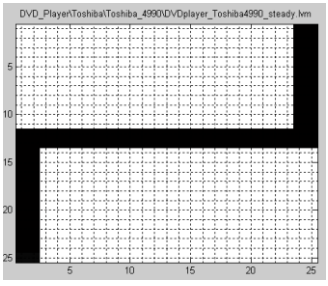
14	Portable Fan 2		
15	Shredder		
16	DVD Player		
17	Cellphone		
18	Laptop Computer		

Table B.1 continued

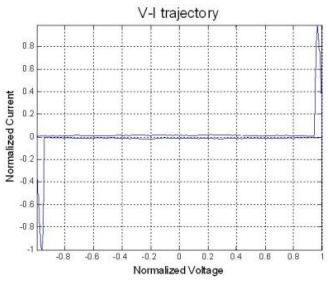
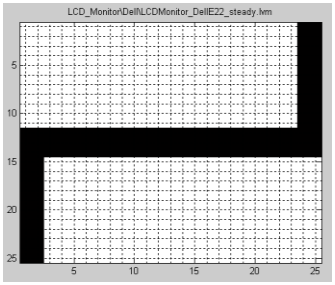
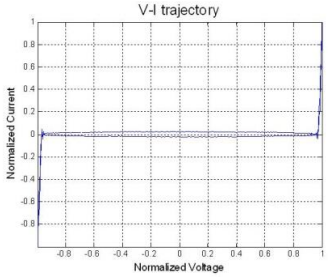
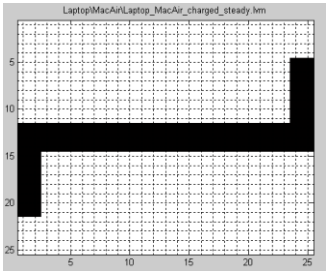
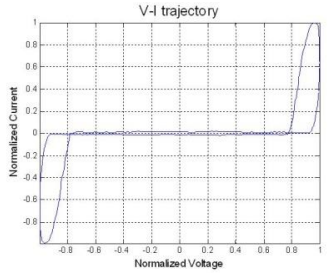
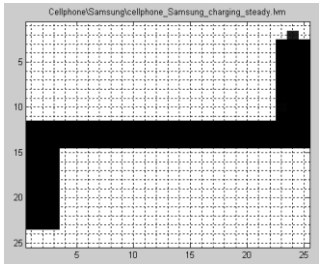
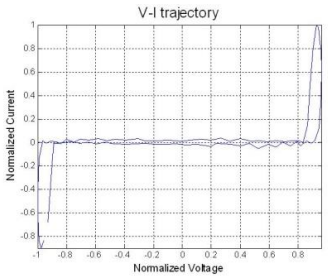
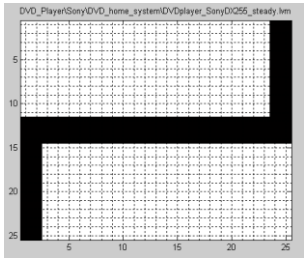
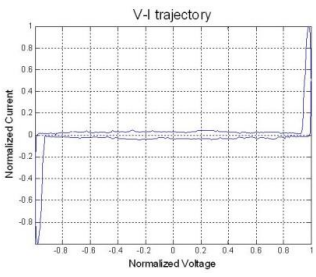
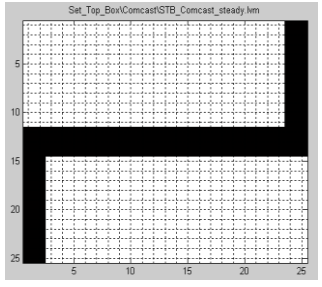
19	Laptop Computer 2		
20	DVD Player 2		
21	Desktop Computer		
22	LED TV		
23	LCD Monitor		

Table B.1 continued

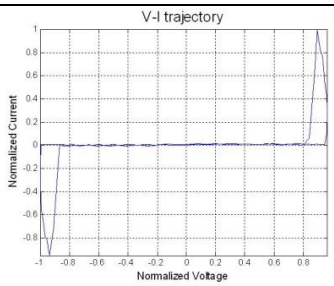
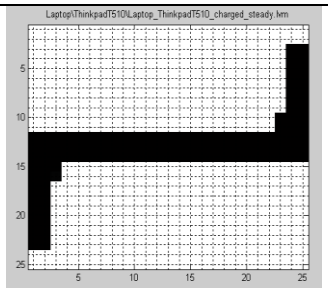
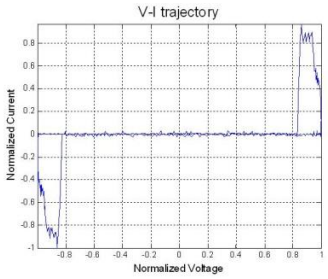
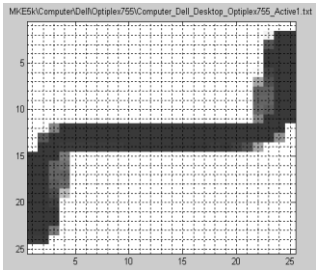
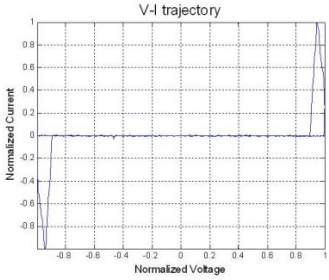
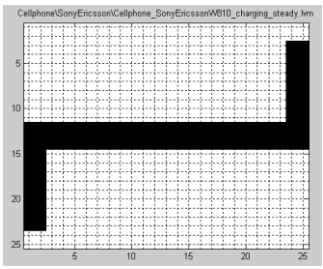
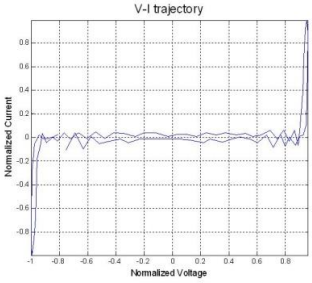
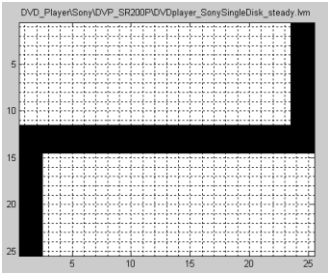
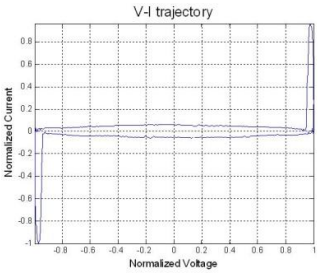
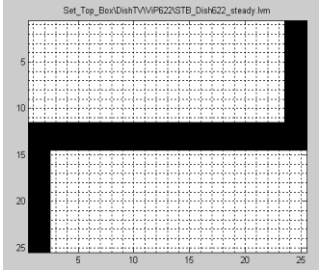
24	Cellphone 2		
25	Cellphone 3		
26	Set-Top Box		
27	DVD Player 3		
28	Set-Top Box 2		

Table B.1 continued

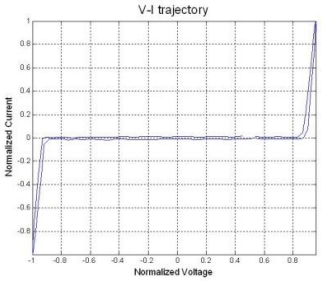
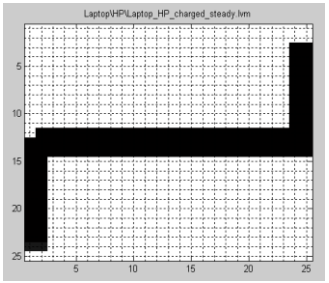
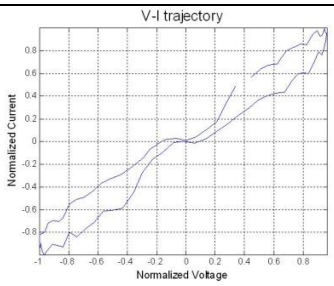
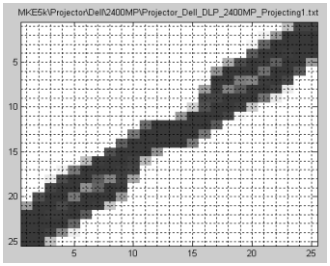
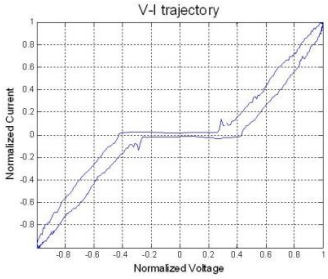
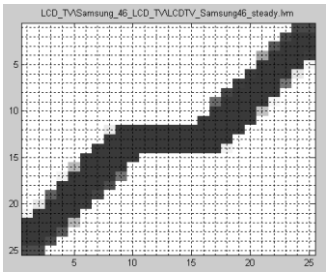
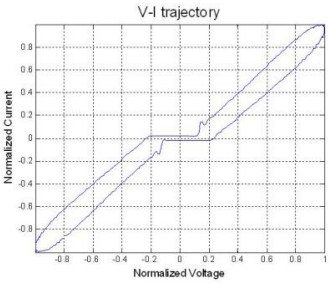
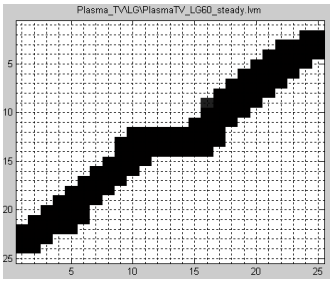
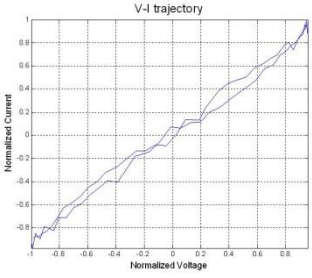
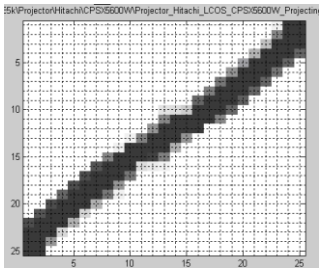
29	Electric Board		
30	Projector		
31	LCD TV		
32	LCD TV 2		
33	Projector 2		

Table B.1 continued

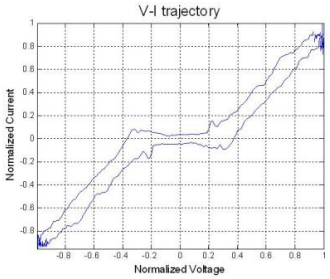
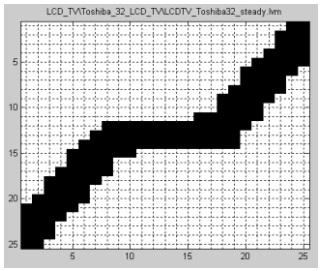
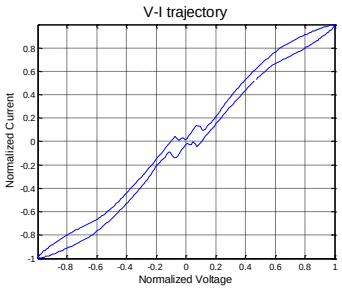
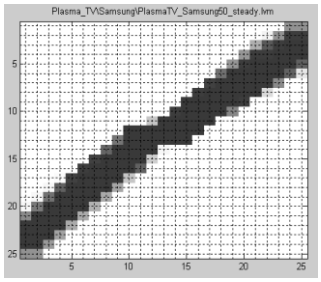
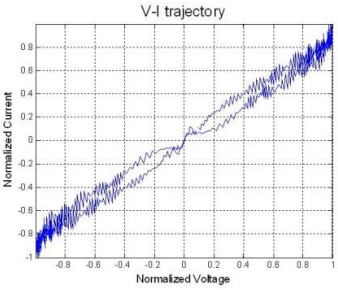
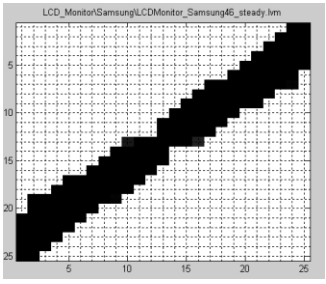
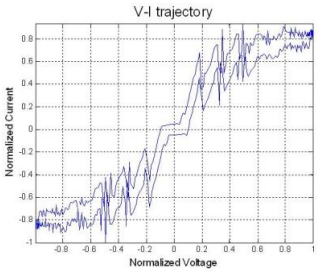
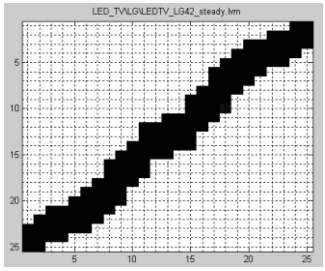
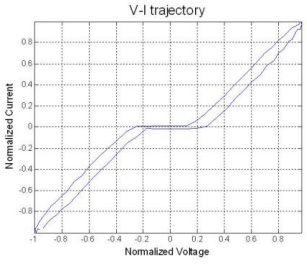
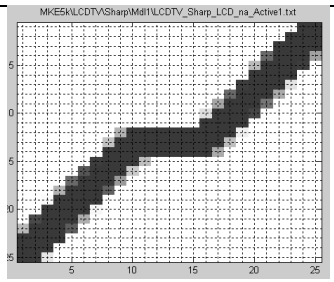
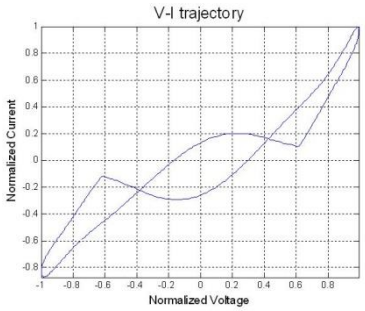
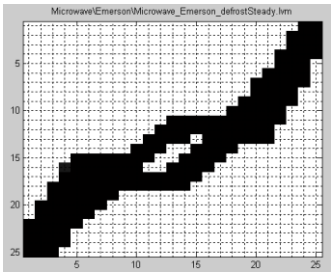
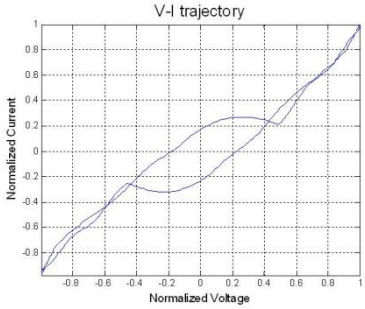
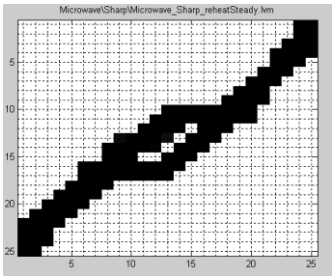
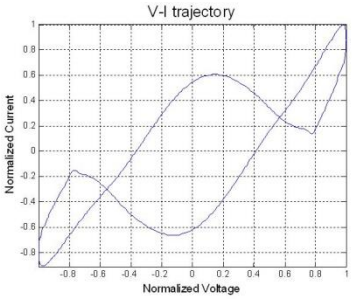
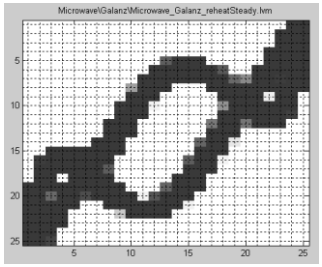
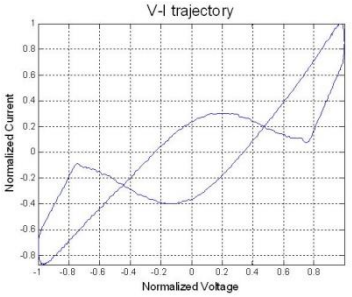
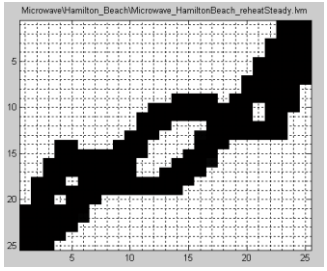
34	Plasma TV		
35	LED TV 2		
36	LCD Monitor 2		
37	Video Game		
38	LCD TV 3		

Table B.1 continued

39	Microwave 2 Defrost Mode		
40	Microwave 3 Defrost Mode		
41	Microwave 4 Reheat Mode		
42	Microwave 5 Reheat Mode		

APPENDIX C SOM TOOLBOX FOR MATLAB

This dissertation uses the MATLAB SOM toolbox developed by the the Laboratory of Information and Computer Science in the Helsinki University of Technology. This toolbox can be downloaded online and directly installed by placing the toolbox in the MATLAB toolboxes directory. This section only lists the key functions and necessary information to run tests. More details appear in the SOM toolbox tutorial and manual available at http://www.cis.hut.fi/somtoolbox/documentation.

The training data for the SOM toolbox is in tabular format, as shown in the following figure. In other words, there can be any number of training feature vectors but the length of each feature vector is fixed.

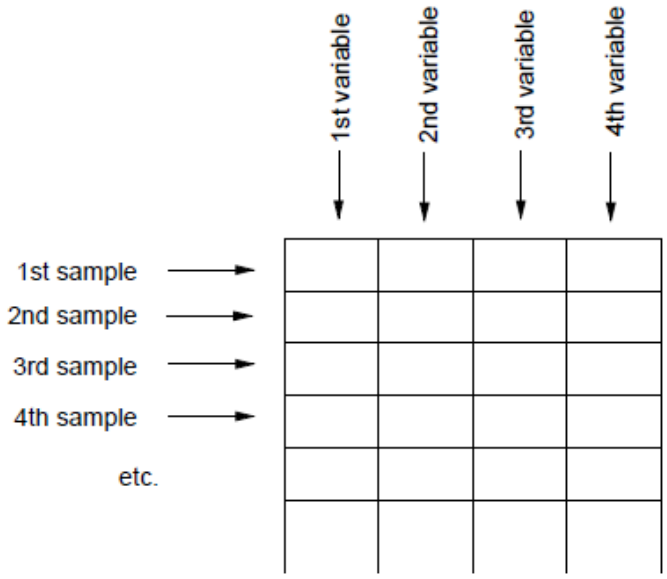


Figure C.1 Tabular format data for MATLAB SOM toolbox [87]

Figure C.1 shows that there can be many training feature vectors (called samples in Figure C.1) but each feature vector has only 4 features (called samples in Figure C.1), i.e., it has a length of 4. In this dissertation, the set of features listed in section 4.4.1 has 13 features, and there can be any number of such 1-by-13 features vectors for training dependsingon the number of PELs available for training.

Furthermore, the training data is converted to a so-called *data struct*, which is a MATLAB struct defined by the designers of this toolbox to group information together. Each such data struct has two major fields: numerical data (.data format) and label data (.labels format). Numerical data is the set of training feature vectors and the label data provides options to assign each training feature vector a label (or identity). The latter data format is very import for implementing the supervised SOM training and identification.

Similarly, the cell grid of an SOM is implemented by a MATLAB struct called *map struct*. Details on the map struct are not included here due to limited space.

There are many functions available in this toolbox, and the key ones that have been used are listed as follows.

(1) som_set: create, set, and check values to data structs

(2) som_data_struct: create and initialize a data struct

(3) som_map_struct: create and initialize a map struct

(4) som_normalize: normalize data set

(5) som_make: create, initialize, and train a SOM

(6) som_randinit: random initialization algorithm

(7) som_lininit: linear initialization algorithm

(8) som_seqtrain: sequential training algorithm

(9) som_batchtrain: batch training algorithm

(10) som_quality: quantization and topographic error of SOM

(11) som_bmus: calculates BMUs for given data vectors

(12) som_label: give labels to map units

(13) som_label2num: recodes string data labels to integer class labels

(14) som_autolabel: automatically labels the SOM based on given data

(15) som_neighborhood: calculates neighborhood matrix for the given map

(16) som_neighbors: calculates different kinds of neighborhoods

(17) som_show: basic visualization

(18) som_show_add: add labels, hits and trajectories

(19) som_show_clear: remove extra markers

Finally, the overall process of training an SOM and doing identification using the toolbox is summarized as follows.

(1) Read the input training data in a matrix *D*. Then call the som_data_struct function:

$$sD = \text{som\_data\_struct}(D);$$

where the output *sD* is a data struct containing the training data *D*.

(2) Label the data struct *sD* using som_label or som_autolabel.

(3) Update the data struct *sD* using som_set.

(4) Pre-process the data struct if necessary, e.g. using som_normalize to normalize the training data.

(5) Initialize and train the map using

sMap = som_make (sD);

After training, the output is a map struct named by sMap. More options on how to initialize and train the map can be implemented by calling functions som_lininit, som_raninit, som_seqtrain, and som_batchtrain.

(6) After training, read in the test data SOM_testData, find its BMUs according to the trained map

[Bmus] = som_bmus(sMap,SOM_testData),

and assign the labels of the BMUS sMap.labels(Bmus) to the test data, which completes the identification process.

(7) Compare the assigned labels and the true labels to calculate the identification success rate.

APPENDIX D SVM TOOLBOX FOR MATLAB

The SVM-KM toolbox for MATLAB was developed by Alain Rakotomamonjy from the University of Rouen, France. The toolbox can be downloaded online from the website http://asi.insa-rouen.fr/enseignants/~arakoto/toolbox/index.html and installed into the MATLAB toolboxes directory in a similar manner as the SOM toolbox.

The input training data format is the same as in the SOM toolbox, i.e., in a matrix or tabular format. That is, there can be any number of training feature vectors but the length of each feature vector is fixed.

The first step of training an SVM using the input data is to set the following parameters:

(1) c: bound of the Lagrangian multipliers;

(2) lambda: conditioning parameter for using quadratic programming (QP) method, typically very small;

(3) kernel: which kernel to use, e.g., 'poly' for polynomial kernel or 'gaussian' for Gaussian kernel;

(4) kerneloption: parameters of the selected kernel, e.g., bandwidth if the Gaussian kernel is selected;

(5) verbose: display outputs (default value is 0: no display);

(6) nbclass: the number of classes

The training of an SVM starts with calling the following function

```
svmmulticlassoneagainstall(SVM_trainData,SVM_trainDataLabel,nbclass,
                c,lambda,kernel,kerneloption,verbose);
```

where

- SVM_trainData: the training data
- SVM_trainDataLabel: lables of the training data

Other options include other functions such as svmmulticlassoneagainstone.

Finally, read the testing dataset SVM_testData and call the function svmmultival. This function assigns a (predicted) identity label to each training feature vector based on the support vectors of the trained SVM. Thus the output of this function is the set of predicted identity labels of the testing dataset. The identification success rate can be calculated by comparing the assigned labels with the true identity labels.

BIBLIOGRAPHY

[1] "Annual Energy Review 2011," Energy Information Administration, U.S. Department of Energy, Washington DC, DOC/EIA-0384(2011), 2012.

[2] R. Hendron and M. Eastment, "Development of an Energy-Savings Calculation Methodology for Residential Miscellaneous Electric Loads," in *ACEEE Summer Study on Energy Efficiency in Buildings*, Pacific Grove, California, 2006.

[3] U.S. Energy Information Administration. (2007). *Miscellaneous Electricity Services in the Buildings Sector*. Available: http://www.eia.gov/oiaf/aeo/otheranalysis/mesbs.html

[4] "Annual Energy Outlook 2010: With Projections to 2035," Energy Information Administration, U.S. Department of Energy, Washington DC, DOC/EIA-0383(2010), 2010.

[5] TIAX LLC, "Commercial and Residential Sector Miscellaneous Electricity Consumption: Y2005 and Projections to 2030," 2006.

[6] *ENERGY STAR: Standby Power and Energy Vampires*. Available: http://www.energystar.gov/index.cfm?c=about.vampires

[7] J. Rivas. (2009). *Managing Plug Loads*. Available: http://www.epa.gov/climateleadership/documents/events/11feb_plugloads.pdf

[8] *Lawrence Berkeley National Laboratory: Standby Power Home Page*. Available: http://standby.lbl.gov/

[9] C. D. Barley, C. Haley, R. Anderson, and L. Pratsch, *Building america system research plan for reduction of miscellaneous electrical loads in zero energy homes*: National Renewable Energy Laboratory, 2008.

[10] Energy Star. *ENERGY STAR Program Requirements for Televisions*.

[11] California Energy Commission. *California Approves New Energy Efficient TV Regulations*. Available: http://www.energy.ca.gov/releases/2009_releases/2009-11-18_tv_regulations.html

[12] *Building America: Resources for Energy Efficient Homes*. Available: http://www1.eere.energy.gov/buildings/building_america/

[13] National Renewable Energy Laboratory. (2009). *Getting to Net Zero*. Available: www.nrel.gov/docs/fy09osti/46382.pdf

[14] P. Du and N. Lu, "Appliance Commitment for Household Load Scheduling," *IEEE Transactions on Smart Grid,* vol. 2, pp. 411-419, 2011.

[15] G. Koutitas, "Control of Flexible Smart Devices in the Smart Grid," *IEEE Transactions on Smart Grid,* vol. 3, pp. 1333-1343, 2012.

[16] P. Zhang, C. Zhou, B. G. Stewart, D. M. Hepburn, W. Zhou, and J. Yu, "An Improved Non-Intrusive Load Monitoring Method for Recognition of Electric Vehicle Battery Charging Load," *Energy Procedia,* vol. 12, pp. 104-112, 2011.

[17] P. Palensky and D. Dietrich, "Demand Side Management: Demand Response, Intelligent Energy Systems, and Smart Loads," *IEEE Transactions on Industrial Informatics,* vol. 7, pp. 381-388, 2011.

[18] A. Mohsenian-Rad, V. W. S. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous Demand-Side Management Based on Game-Theoretic Energy Consumption Scheduling for the Future Smart Grid," *IEEE Transactions on Smart Grid,* vol. 1, pp. 320-331, 2010.

[19] J. L. Mathieu, P. N. Price, S. Kiliccote, and M. A. Piette, "Quantifying Changes in Building Electricity Use, With Application to Demand Response," *IEEE Transactions on Smart Grid,* vol. 2, pp. 507-518, 2011.

[20] Q. B. Dam, S. Mohagheghi, and J. Stoupis, "Intelligent Demand Response Scheme for Customer Side Load Management," in *Energy 2030 Conference, 2008. ENERGY 2008. IEEE*, 2008, pp. 1-7.

[21] T. Sousa, H. Morais, Z. Vale, P. Faria, and J. Soares, "Intelligent Energy Resource Management Considering Vehicle-to-Grid: A Simulated Annealing Approach," *IEEE Transactions on Smart Grid,* vol. 3, pp. 535-542, 2012.

[22] A. Sepulveda, L. Paull, W. G. Morsi, H. Li, C. P. Diduch, and C. Liuchen, "A novel demand side management program using water heaters and particle swarm

optimization," in *Electric Power and Energy Conference (EPEC), 2010 IEEE*, 2010, pp. 1-5.

[23] S. Lee, Y. Chon, Y. Kim, R. Ha, and H. Cha, "Occupancy Prediction Algorithms for Thermostat Control Systems Using Mobile Devices," *IEEE Transactions on Smart Grid*, vol. 4, pp. 1332-1340, 2013.

[24] Y. Guo, M. Pan, Y. Fang, and P. P. Khargonekar, "Decentralized Coordination of Energy Utilization for Residential Households in the Smart Grid," *IEEE Transactions on Smart Grid*, vol. 4, pp. 1341-1350, 2013.

[25] C. Chen, J. Wang, Y. Heo, and S. Kishore, "MPC-Based Appliance Scheduling for Residential Building Energy Management Controller," *IEEE Transactions on Smart Grid*, vol. 4, pp. 1401-1410, 2013.

[26] T. Hubert and S. Grijalva, "Modeling for Residential Electricity Optimization in Dynamic Pricing Environments," *IEEE Transactions on Smart Grid*, vol. 3, pp. 2224-2231, 2012.

[27] F. Corno and F. Razzak, "Intelligent Energy Optimization for User Intelligible Goals in Smart Home Environments," *IEEE Transactions on Smart Grid*, vol. 3, pp. 2128-2135, 2012.

[28] I. Georgievski, V. Degeler, G. A. Pagani, N. Tuan Anh, A. Lazovik, and M. Aiello, "Optimizing Energy Costs for Offices Connected to the Smart Grid," *IEEE Transactions on Smart Grid*, vol. 3, pp. 2273-2285, 2012.

[29] Y. Ozturk, D. Senthilkumar, S. Kumar, and G. Lee, "An Intelligent Home Energy Management System to Improve Demand Response," *IEEE Transactions on Smart Grid*, vol. 4, pp. 694-701, 2013.

[30] M. Pipattanasomporn, M. Kuzlu, and S. Rahman, "An Algorithm for Intelligent Home Energy Management and Demand Response Analysis," *IEEE Transactions on Smart Grid*, vol. 3, pp. 2166-2173, 2012.

[31] Y.-W. Chen, X. Chen, and N. Maxemchuk, "The Fair Allocation of Power to Air Conditioners on a Smart Grid," *IEEE Transactions on Smart Grid*, vol. 3, pp. 2188-2195, 2012.

[32] N. Lu, "An Evaluation of the HVAC Load Potential for Providing Load Balancing Service," *IEEE Transactions on Smart Grid*, vol. 3, pp. 1263-1270, 2012.

[33] N. Lu and Y. Zhang, "Design Considerations of a Centralized Load Controller Using Thermostatically Controlled Appliances for Continuous Regulation Reserves," *IEEE Transactions on Smart Grid*, vol. 4, pp. 914-921, 2013.

[34] G. T. Costanzo, G. Zhu, M. F. Anjos, and G. Savard, "A System Architecture for Autonomous Demand Side Load Management in Smart Buildings," *IEEE Transactions on Smart Grid*, vol. 3, pp. 2157-2165, 2012.

[35] Z. Xu, X. Guan, Q.-S. Jia, J. Wu, D. Wang, and S. Chen, "Performance Analysis and Comparison on Energy Storage Devices for Smart Building Energy Management," *IEEE Transactions on Smart Grid*, vol. 3, pp. 2136-2147, 2012.

[36] Siemens. *Demand Response Solutions for Commercial Buildings*. Available: http://w3.usa.siemens.com/buildingtechnologies/us/en/energy-efficiency/demand-response/Pages/demand-response-solution-commercial-buildings.aspx

[37] J. Page, S. Kiliccote, J. H. Dudley, M. A. Piette, A. K. Chiu, B. Kellow, *et al.*, "Automated Demand Response Technology Demonstration for Small and Medium Commercial Buildings," Lawrence Berkeley National Laboratory,2011.

[38] Pacific Northwest National Laboratory. *Energy Efficiency and Demand Response*. Available:http://eere.pnnl.gov/buildingtechnologies/energyefficiency_demandresponse.stm

[39] DOE National Renewable Energy Laboratory (NREL). *Residential Building Energy Efficiency Meeting 2010*. Available: http://apps1.eere.energy.gov/buildings/publications/pdfs/building_america/ns/plenary_2_emerging_tech.pdf

[40] G. Hancke and D. Vrey, "Electric load monitoring and control in the domestic environment," in *Instrumentation and Measurement Technology Conference, 1994 (IMTC/94)*, 1994, pp. 560-562 vol.2.

[41] G. W. Hart, E. C. Kern, and F. C. Schweppe, "Non-intrusive appliance monitor apparatus," U.S. Patent 4 858 141, Aug. 15, 1989.

[42] Navetas Energy Management. Available: http://www.navetas.com/

[43] D. He, L. Du, Y. Yang, R. G. Harley, and T. G. Habetler, "Front-End Electronic Circuit Topology Analysis for Model-Driven Classification and Monitoring of Appliance Loads in Smart Buildings," *IEEE Transactions on Smart Grid,* vol. 3, pp. 2286-2293, 2012.

[44] H. Y. Lam, G. S. K. Fung, and W. K. Lee, "A Novel Method to Construct Taxonomy Electrical Appliances Based on Load Signatures," *IEEE Transactions on Consumer Electronics,* vol. 53, pp. 653-660, 2007.

[45] D. He, L. Du, R. Harley, T. Habetler, and Y. Yang, "Electronic circuit survey for office load monitoring and identification," in *Energy Conversion Congress and Exposition (ECCE), 2012 IEEE*, 2012, pp. 1228-1232.

[46] Y. Du, L. Du, B. Lu, R. G. Harley, and T. G. Habetler, "A review of identification and monitoring methods for electric loads in commercial and residential buildings," in *Proc. 2010 IEEE Energy Conversion Conf. and Expo.*, Atlanta, GA, 2010, pp. 4527-4533.

[47] K. Suzuki, S. Inagaki, T. Suzuki, H. Nakamura, and K. Ito, "Nonintrusive appliance load monitoring based on integer programming," in *SICE Annual Conference, 2008*, 2008, pp. 2742-2747.

[48] J. Liang, S. Ng, G. Kendall, and J. Cheng, "Load Signature Study - Part I: Basic Concept, Structure, and Methodology," *IEEE Transactions on Power Delivery,* vol. 25, pp. 551-560, 2010.

[49] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE,* vol. 80, pp. 1870-1891, 1992.

[50] F. Sultanem, "Using appliance signatures for monitoring residential loads at meter panel level," *IEEE Transactions on Power Delivery,* vol. 6, pp. 1380-1385, 1991.

[51] S. Drenker and A. Kader, "Nonintrusive monitoring of electric loads," *Computer Applications in Power, IEEE*, vol. 12, pp. 47-51, 1999.

[52] S. B. Leeb, S. R. Shaw, and J. L. Kirtley, Jr., "Transient event detection in spectral envelope estimates for nonintrusive load monitoring," *IEEE Transactions on Power Delivery,* vol. 10, pp. 1200-1210, 1995.

[53] R. Cox, S. B. Leeb, S. R. Shaw, and L. K. Norford, "Transient event detection for nonintrusive load monitoring and demand side management using voltage distortion," in *Applied Power Electronics Conference and Exposition, 2006. APEC '06. Twenty-First Annual IEEE,* 2006, p. 7 pp.

[54] U. A. Khan, S. B. Leeb, and M. C. Lee, "A multiprocessor for transient event detection," *IEEE Transactions on Power Delivery,* vol. 12, pp. 51-60, 1997.

[55] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, *et al.*, "Power signature analysis," *IEEE Power and Energy Magazine,* vol. 2, pp. 56 - 63, 2003.

[56] T. Saitoh and Y. Aota, "Current Sensor-based Non-intrusive Appliance Recognition for Intelligent Outlet," presented at the The 23rd International Technical Conference on Circuits/Systems, Computers and Communications, 2008.

[57] L. Du, D. He, Y. Yang, J. R. Restrepo, R. G. Harley, T. G. Habetler, *et al.*, "Self-Organizing Classification and Identification of Miscellaneous Electric Loads," in *IEEE Power and Energy Society General Meeting,* San Diego, CA, 2012.

[58] L. Du, J. A. Restrepo, Y. Yang, R. G. Harley, and T. G. Habetler, "Nonintrusive, Self-Organizing, and Probabilistic Classification and Identification of Plugged-In Electric Loads," *IEEE Transactions on Smart Grid,* vol. 4, pp. 1371-1380, 2013.

[59] E. A. Cano Plata and H. E. Tacca, "Power quality assessment and load identification," in *Harmonics and Quality of Power, 2000. Proceedings. Ninth International Conference on,* 2000, pp. 840-845 vol.3.

[60] C. Laughman, L. Kwangduk, R. Cox, S. Shaw, S. Leeb, L. Norford, *et al.*, "Power signature analysis," *Power and Energy Magazine, IEEE,* vol. 1, pp. 56-63, 2003.

[61] R. K. Hartana and G. G. Richards, "Harmonic source monitoring and identification using neural networks," *IEEE Transactions on Power Systems,* vol. 5, pp. 1098-1104, 1990.

[62] A. Cole and A. Albicki, "Nonintrusive identification of electrical loads in a three-phase environment based on harmonic content," in *Proc. 17th IEEE Instrumentation and Measurement Technology Conf.* , 2000, pp. 24 - 29.

[63] M. H. Rashid, *Power Electronics: Circuits, Devices, and Applications*, 3rd ed.: Prentice Hall 2003.

[64] N. Mohan and T. M. Undeland, *Power Electronics: Converters, Applications, and Design*, 4th ed.: Wiley Press, 2007.

[65] L. K. Norford and S. B. Leeb, "Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms," *Energy and Buildings,* vol. 24, pp. 54 - 61, 1996.

[66] S. R. Shaw, S. B. Leeb, L. K. Norford, and R. W. Cox, "Nonintrusive Load Monitoring and Diagnostics in Power Systems," *IEEE Transactions on Instrumentation and Measurement,* vol. 57, pp. 1445-1454, 2008.

[67] K. D. Lee, S. B. Leeb, L. K. Norford, P. R. Armstrong, J. Holloway, and S. R. Shaw, "Estimation of Variable-Speed-Drive Power Consumption From Harmonic Content," *IEEE Transactions on Energy Conversion,* vol. 20, pp. 566-574, 2005.

[68] H. Y. Lam, "Voltage-Current Trajectory: A 2-Dimensional Approach to Understand Electrical Load Signatures," Master of Philosophy Thesis, The University of Hong Kong, 2007.

[69] T. Hassan, F. Javed, and N. Arshad, "An Empirical Investigation of V-I Trajectory Based Load Signatures for Non-Intrusive Load Monitoring," *IEEE Transactions on Smart Grid,* vol. PP, pp. 1-9, 2013.

[70] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey," *Sensors,* vol. 12, pp. 16838-16866, 2012.

[71] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Transactions on Consumer Electronics,* vol. 57, pp. 76-84, 2011.

[72] Z. Wang and G. Zheng, "Residential Appliances Identification and Monitoring by a Nonintrusive Method," *IEEE Transactions on Smart Grid*, vol. 3, pp. 80-92, 2012.

[73] L. Farinaccio and R. Zmeureanu, "Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses," *Energy and Buildings*, vol. 30, pp. 245-259, 1999.

[74] D. Srinivasan, W. S. Ng, and A. C. Liew, "Neural-network-based signature recognition for harmonic source identification," *IEEE Transactions on Power Delivery*, vol. 21, pp. 398-405, 2006.

[75] S. N. Patel, T. Robertson, J. A. Kientz, M. S. Reynolds, and G. D. Abowd, "At the flick of a switch: detecting and classifying unique electrical events on the residential power line," in *Proceedings of the 9th international conference on Ubiquitous computing*, Innsbruck, Austria, 2007, pp. 271-288.

[76] C. Cremer, W. Eichhammer, M. Friedewald, P. Georgieff, S. Rieth-Hoerst, B. Schlomann, *et al.*, "Energy consumption of information and communication technology (ICT) in Germany up to 2010: Summary of the final report to the german federal ministry of economics and labour," ed: Frauenhofer ISI and CEPE, Swiss Federal Institutes of Technology, Karlsruhe/Zurich, 2003.

[77] J. Roberson, C. Webber, M. McWhinney, R. Brown, M. Pinckard, and J. Busch, "After-hours Power Status of Office Equipment and Energy Use of Miscellaneous Plug-Load Equipment," *Lawrence Berkeley National Laboratory, Berkeley, CA, Tech. Rep. LBNL-53729*, 2004.

[78] S. Frank, L. G. Polese, E. Rader, M. Sheppy, and J. Smith, "Extracting Operating Modes from Building Electrical Load Data," in *Green Technologies Conference (IEEE-Green), 2011 IEEE*, 2011, pp. 1-6.

[79] R. Sedgewick, *Algorithms in C++, Part 5: Graph Algorithm*. Boston: Addison Wesley, 2002.

[80] C. Van Loan, *Computational frameworks for the fast Fourier transform*: SIAM Press, 1992.

[81] J. S. Walker, *Fast Fourier transforms*: CRC press, 1996.

[82] Z. Wang, "Fast algorithms for the discrete W transform and for the discrete Fourier transform," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, pp. 803-816, 1984.

[83] H. M. Ozaktas, O. Arikan, M. A. Kutay, and G. Bozdagt, "Digital computation of the fractional Fourier transform," *IEEE Transactions on Signal Processing*, vol. 44, pp. 2141-2150, 1996.

[84] A. V. Oppenheim and R. W. Schafer, *Discrete-time signal processing*, 3rd ed.: Prentice Hall, 2009.

[85] J. G. Proakis and D. K. Manolakis, *Digital signal processing: principles algorithms and applications*, 4th ed.: Prentice Hall, 2006.

[86] T. Kohonen, *Self-Organising Maps*, 3rd ed.: Springer, 2001.

[87] E. Alhoniemi, J. Himberg, J. Parhankangas, and J. Vesanto. (2005). *SOM Toolbox for Matlab*. Available: http://www.cis.hut.fi/somtoolbox/

[88] A. Ultsch and H. P. Siemon, "Exploratory Data Analysis: Using Kohonen Networks on Transputers," Technical Report 329, Univ. of Dortmund., Dortmund, Germany.1989.

[89] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Fourth ed.: Academic Press, 2008. ISBN: 978-1597492720.

[90] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*: John Wiley & Sons, 2012.

[91] K. Fukunaga, *Introduction to statistical pattern recognition*: Access Online via Elsevier, 1990.

[92] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning* vol. 1: springer New York, 2006.

[93] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 4-37, 2000.

[94] C. M. Bishop, *Neural networks for pattern recognition*: Oxford university press, 1995.

[95] B. D. Ripley, *Pattern recognition and neural networks*: Cambridge university press, 2007.

[96] J. Aitchison and S. Silvey, "Maximum-likelihood estimation of parameters subject to restraints," *The Annals of Mathematical Statistics,* vol. 29, pp. 813-828, 1958.

[97] B. Schölkopf and A. J. Smola, *Learning with kernels*: MIT Press, 2002.

[98] J. Diebolt and C. P. Robert, "Estimation of finite mixture distributions through Bayesian sampling," *Journal of the Royal Statistical Society. Series B (Methodological),* pp. 363-375, 1994.

[99] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital signal processing,* vol. 10, pp. 19-41, 2000.

[100] J. A. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," *International Computer Science Institute,* vol. 4, p. 126, 1998.

[101] G. Lefebvre and C. Garcia, "A probabilistic Self-Organizing Map for facial recognition," in *Pattern Recognition, 2008. 19th IEEE International Conference on*, 2008, pp. 1-4.

[102] V. N. Vapnik, *The Nature of Statistical Learning Theory*: Springer-Verlag, 1995.

[103] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144-152.

[104] K.-B. Duan and S. S. Keerthi, "Which is the best multiclass SVM method? An empirical study," in *Multiple Classifier Systems*, ed: Springer, 2005, pp. 278-285.

[105] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks,* vol. 13, pp. 415-425, 2002.

[106] K. Crammer and Y. Singer, "On the algorithmic implementation of multiclass kernel-based vector machines," *The Journal of Machine Learning Research,* vol. 2, pp. 265-292, 2002.

[107] T. Inoue and S. Abe, "Fuzzy support vector machines for pattern classification," in *Neural Networks, 2001. Proceedings. IJCNN'01. International Joint Conference on*, 2001, pp. 1449-1454.

[108] D. Tsujinishi and S. Abe, "Fuzzy least squares support vector machines for multiclass problems," *Neural Networks,* vol. 16, pp. 785-792, 2003.

[109] F. Takahashi and S. Abe, "Decision-tree-based multiclass support vector machines," in *Neural Information Processing, 2002. ICONIP'02. Proceedings of the 9th International Conference on*, 2002, pp. 1418-1422.

[110] B. Kijsirikul and N. Ussivakul, "Multiclass support vector machines using adaptive directed acyclic graph," in *Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on*, 2002, pp. 980-985.

[111] C. H. Ding and I. Dubchak, "Multi-class protein fold recognition using support vector machines and neural networks," *Bioinformatics,* vol. 17, pp. 349-358, 2001.

[112] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Transactions on Neural Networks,,* vol. 10, pp. 1055-1064, 1999.

[113] L. Cai and T. Hofmann, "Hierarchical document categorization with support vector machines," in *Proceedings of the thirteenth ACM international conference on Information and knowledge management*, 2004, pp. 78-87.

[114] K. I. Kim, K. Jung, S. H. Park, and H. J. Kim, "Support vector machines for texture classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 1542-1550, 2002.

[115] S. Abe, *Support vector machines for pattern classification*: Springer, 2010.

[116] I. Steinwart and A. Christmann, *Support vector machines*: Springer, 2008.

[117] L. Du, Y. Yang, Dawei, R. G. Harley, and T. G. Habetler, "Support Vector Machine based methods for non-intrusive identification of miscellaneous electric loads " in *The 38th Annual Conference of the IEEE Industrial Electronics Society*, Montreal, Canada, 2012.

[118] S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy. (2005). *SVM and Kernel Methods Matlab Toolbox*. Available: http://asi.insa-rouen.fr/enseignants/~arakotom/toolbox/index.html

[119] J. E. Hopcroft, R. Motwani, and J. D. Ullman, *Introduction to Automata Theory, Languages, and Computation*, 3rd ed.: Addison-Wesley, 2006.

[120] L. Du, L. He, and R. G. Harley, "A survey of methods for placing shunt capacitor banks in power network with harmonic distortion," in *IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society*, 2012, pp. 1198-1203.

[121] L. Du, S. Grijalva, and R. G. Harley, "Potential-game theoretical formulation of optimal power flow problems," in *Power and Energy Society General Meeting, 2012 IEEE*, 2012, pp. 1-6.

[122] M. Zeifman, "Disaggregation of home energy display data using probabilistic approach," *IEEE Transactions on Consumer Electronics,* vol. 58, pp. 23-31, 2012.

[123] "IEEE Recommended Practices and Requirements for Harmonic Control in Electric Power Systems," ed. IEEE Standard 519-1992, 1993.

[124] "IEEE Standard Definitions for the Measurement of Electric Power Quantities Under Sinusoidal, Nonsinusoidal, Balanced, or Unbalanced Conditions," *IEEE Std 1459-2010 (Revision of IEEE Std 1459-2000)*, pp. 1-40, 2010.

VITA

Liang Du was born in Ezhou, China in 1981. He received the dual Bachelor of Engineering degrees (with honor) major in automation and minor in computer science from Huazhong University of Science and Technology, Wuhan, China, in 2003. After working in the industry for a year, Liang moved to Canada in 2005 and received the Master of Applied Science degree in electrical and computer engineering from Concordia University, Montreal, Canada, in 2007. He joined Georgia Tech in 2007 and received the Master of Science degree in Aerospace Engineering with a focus on flight dynamics and control in 2010.

Since January 2010, Liang has been with ECE Power group, where he is currently a Ph.D. candidate and completing the Ph.D. degree in electrical engineering. His research interests include computational intelligence applications in power, smart grid, power system operations and control, and switching power converter technology.

He worked as a summer research intern at the Eaton Corporation Innovation Center in Milwaukee, WI, Mitsubishi Electric Research Labs in Cambridge, MA, and Philips Research North America in Briarcliff Manor, NY, in 2011, 2012, and 2013, respectively.

Liang Du is an IEEE student member since 2010. He is also a member of the IEEE Industry Applications, Power and Energy, Control Systems, Computational Intelligence, and Communications societies. He has authored one chapter, two IEEE Transactions papers, and nine IEEE conference papers. His research resulted in five U.S. patents, one published and four applied.