

UNIVERSITÉ DE HAUTE-ALSACE

Thèse

Presentée pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE HAUTE-ALSACE

Discipline: Électronique, Électrotechnique et Automatique

par

Pirmin Held

Frequency Invariant Transformation of Periodic Signals for Non-Intrusive Load Monitoring

(Arrêté Ministériel du 30 mars 1992)

Soutenue publiquement le 28. mars 2019 devant le jury composé de:

<i>Rapporteurs</i>	Yassine RUICHEK	Professeur, Université de Technologie de Belfort-Montbéliard (UTBM)
	Nicolas RETIÈRE	Professeur, Université Grenoble Alpes (G2Elab)
<i>Examineurs</i>	Nadine LE-FORT PIAT	Professeur, Ecole Nationale Supérieure de Mécanique et des Microtechniques de Besançon (ENSMM)
	Jean MERCKLÉ	Professeur, Université de Haute-Alsace (UHA)

Thèse préparée au sein de l'Institut de Recherche en Informatique,
Mathématiques, Automatique et Signal (IRIMAS) (ex MIPS)
à l'Université de Haute-Alsace sous la direction de

MCF HDR Djaffar OULD ABDESLAM

et au sein de Institute for Smart Systems (ISS)
à Furtwangen University sous la co-direction de

Professeur Dirk BENYOUCEF.

Acknowledgment

I would like to thank all those who contributed to the successful completion of this work.

The work of this thesis was funded by the BMBF¹⁾ as part of the *I-Mon*²⁾ project at Furtwangen University. The aim of this project is to determine the power consumption of individual devices from the total electricity consumption of a household or a company.

The thesis was made possible by a collaboration of the Institut de Recherche en Informatique, Mathématiques, Automatique et Signal (IRIMAS) Laboratory at Université de Haute-Alsace, Mulhouse, France, and the Institute for Smart Systems (ISS) at Furtwangen University, Furtwangen, Germany. The international cooperation was an interesting and instructive experience.

My special thanks go to my advisor Professor Djaffar Ould Abdeslam from IRIMAS at Université de Haute-Alsace (UHA) and to my co-advisor Professor Dirk Benyoucef from ISS at Hochschule Furtwangen University (HFU). I would like to thank them for their excellent advice and constant support.

I would like to thank Professor Yassine Ruichek for his valuable comments during the different stages of this work and for being a member of the jury. Furthermore, I would like to thank Professor Nicolas Retière, Professor Nadine Le-Fort Piat, and Professor Jean Mercklé for reviewing this thesis and being a member of the jury.

In addition, I would like to acknowledge my colleagues Frederik Laasch, Daniel Weißhaar, Alaa Saleh, and Judith Jakob for the good cooperation and the numerous discussions. Especially, I would like to thank Steffen Mauch for proofreading this work. Furthermore, I would like to thank Phillip Klein and Thomas Bier for their previous work which was the basis of my work. I would like to thank the Language Centre of the HFU and in particular Andrew McDouall, Ozan Varli for their linguistic support in this work and the previous publications.

I would like to thank my strictest proofreader Corinna Brichta for her patience, support and love. Finally, I am very grateful to my family for their support, especially to my brother Thorbjörn Held for his tireless corrections. My parents have supported me throughout my career. I am very grateful to them and dedicate this work to them.

¹⁾BMBF: Federal ministry of education and research (in German: BundesMinisterium für Bildung und Forschung)

²⁾I-Mon: Intelligent energy monitoring of electrical consumers

Contents

1	Topic & Motivation	7
1.1	Application Scenario	8
1.2	Contributions	9
2	Relevant Works & Overview of State-of-the-Art Methods	11
2.1	Historical Overview	11
2.2	State of the Art Methods in Non-Intrusive Load Monitoring (NILM)	16
2.2.1	Measuring Systems and Datasets	19
2.2.2	Waveform	24
2.2.3	Event Detection	33
2.2.4	Features & Feature Extraction	35
2.2.5	Classification	37
3	Objectives and Hypotheses	41
4	Home Equipment Laboratory Dataset	43
4.1	Introduction	43
4.2	Measurement System	44
4.3	Measurement Scenarios	46
4.4	Power Distribution of the Appliances	46
4.5	Conclusion	48
5	Waveform	49
5.1	Introduction	49
5.2	Influence of the Amplitude of the Voltage Signal in NILM	50
5.2.1	Conventional Power Supplies	51
5.2.2	Switching Power Supplies	51
5.2.3	Standardization of Passive Appliances	52
5.3	Frequency Invariant Transformation of Periodic Signals	53
5.3.1	FIT-PS with Higher Sampling Frequency	60
5.3.2	Reduction of the Sampling Frequency	61
5.3.3	Dimension Reduction in k-Dimension	63
5.4	Results	63
5.5	Discussion	64
5.6	Conclusion	68
6	Different Frequency Invariant Transformation of Periodic Signals (FIT-PS) Methods	69
6.1	FIT-PS with Interpolation	69
6.2	FIT-PS without Interpolation	70
6.3	Additional Curve Fitting for FIT-PS	71
6.4	Results	73
6.5	Discussion	77
6.6	Conclusion	78

7	Event Detection	79
7.1	Introduction	79
7.2	Event Detection Using FIT-PS	80
7.2.1	Event Detection Based on a Sampling Frequency of 1.2 kHz	80
7.2.2	First Stage	80
7.2.3	Second Stage	83
7.2.4	Sampling Frequency of 12 kHz	87
7.2.5	Principal Component Analysis for Dimension Reduction	88
7.3	Parameter Optimization	91
7.3.1	Simulated Annealing	93
7.3.2	Generalized Pattern Search	94
7.4	Results	95
7.4.1	Investigation of Different Numbers of Components Resulting from the Principal Component Analysis (PCA)	95
7.4.2	Comparing Different Sampling Rates and Waveforms	95
7.4.3	Results of Parameter Optimization	97
7.5	Discussion	100
7.6	Conclusion	101
8	Feature Extraction for Appliance Classification	103
8.1	Introduction	103
8.2	Steady State Features	104
8.3	Transient State Feature	109
8.4	Discussion	112
8.5	Conclusion	112
9	Classification	113
9.1	Introduction	113
9.2	k-Nearest Neighbor Classifier	114
9.3	Naive Bayes Classifier	115
9.4	Support Vector Machine	116
9.5	FeedForward Neural Network	119
9.6	Long Short-Term Memory Network	121
9.7	Results of Validation and Test Data	122
9.7.1	Validation Results of the k-Nearest Neighbor	123
9.7.2	Test Results of the k-Nearest Neighbor	125
9.7.3	Validation Results of the Bayes Classifier	126
9.7.4	Test Results of the Bayes Classifier	126
9.7.5	Validation Results of the Support Vector Machine	128
9.7.6	Test Results of the Support Vector Machine	129
9.7.7	Validation Results of the Neuronal Network	131
9.7.8	Test Results of the Neuronal Network	133
9.7.9	Comparison of the Different Classification Methods and Features	135
9.8	Interpretation of the Harmonic Results	139
9.9	Discussion	141
9.10	Conclusion	143
10	Summary & Perspectives	145
10.1	Summary	145
10.2	Perspectives	147
A	Exemplary Representation of FIT-PS Waveform	151

Chapter 1 | Topic & Motivation

Reducing CO₂ emissions is a global challenge. In addition to the introduction of renewable energy, CO₂ emissions can also be reduced through lower energy consumption. Improved management of energy production and energy consumption can also lead to a more effective use of existing resources and thus to a reduction in CO₂ emission. Saving costs through lower energy consumption is an additional motivation for private or industrial consumers. However, since the customer is usually confronted with his power consumption only once a year and no further information such as, for example, which device has consumed how much power, it is challenging for the consumer to change his behavior or renew devices with unnecessary high power consumption. Determining the power consumption of individual appliances requires an energy meter in front of each appliance. Since this is complicated and expensive, this is not realized in practice. With Non-Intrusive Load Monitoring (NILM) the total power consumption of a house, a flat, or an industrial building is determined only at one central location. Out of this total energy consumption, the individual consumption of each device is specified. From this original idea, applications in various areas of life can be derived which are explained in this section.

In the classical approach, multiple sensors are needed to measure the individual power consumption of devices. Because of the high installation effort and the intrusion into the grid, this approach is also described as intrusive monitoring, as shown in Fig. 1.1. This method has the disadvantage that there is a high effort at installation because the number of sensors increases with the number of devices which have to be measured. Often it is not possible to install the measurement system without disconnecting the appliances. Especially in the industrial context, a disconnection of machines is problematic. In case several appliances are measured at the same time many meters are required. Apart from the hardware costs, additional problems like communication between the meters for time synchronization have to be solved. Synchronization is required, for example, when power peaks of running appliances should be minimized, or for comparing the sum of the individual power consumption with the total measured power consumption.

In 1982 G. W. Hart from the Massachusetts Institute of Technology (MIT) [61] presented the idea of Non-Intrusive Load Monitoring (NILM). In NILM only one measurement system at a central position (in Fig. 1.2 in front of the electricity meter) of the grid is used. Here, the total power consumption of all devices connected to the grid is measured. With this information, a disaggregation of the measured total power consumption, into the individual power consumption is performed as shown in Fig. 1.2. The comparison of Fig. 1.1 and 1.2 shows the advantage

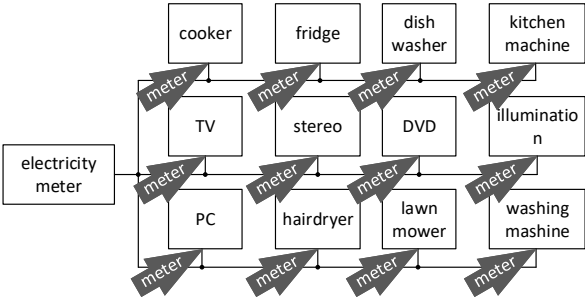


Figure 1.1: Intrusive monitoring

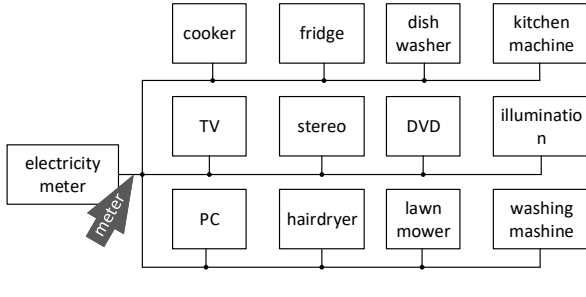


Figure 1.2: Non-intrusive monitoring

of NILM. Here, the effort for installation and measurement hardware is much less than in an intrusive measurement setting. This idea also avoids problems concerning the communication between the different meters for example time synchronization.

At the start of the millennium, the interest in a detailed energy consumption analysis increased. More and more intelligent electricity meters were installed in the industry, as well as, in private households. However, the higher resolution of the total power consumption leads only to a small benefit for the customer so far. The more detailed time curve of the total power consumption is usually not very meaningful and challenging to interpret for the customer. The customer would have a much greater benefit if the actual consumption could be assigned to the individual devices. This additional information can be supplied by NILM.

The reactor disaster in Fukushima, Japan in 2011, led to a rethinking of the energy policy. In Germany and other countries, the energy transition (the transition from conventional to renewable energy production) was politically initiated. Furthermore, at the United Nations Climate Change Conference in Paris, France an agreement to reduce emissions and reduce the effects of global warming below 1.5°C was signed by 195 countries. This crucial event is essential for the increased expansion of renewables across the globe. The regulation of the grid needs to be changed from controlling a small number of conventional power plans to a significantly bigger number of small power plans. For this challenge, it is necessary to know where, at which time, and how much energy is required. Through the use of smart meters, it is already possible to inform the grid operator about the actual power consumption of individual customers. The use of NILM could provide more detailed information about the power consumption of individual types of appliances. With this information a more precise forecast of the power consumption is possible.

One way of cutting costs and resources for electricity production is to use the generated power as efficiently as possible and to reduce over-capacities. An effective supplement is the introduction of time variable electricity prices. This provides incentives to the customer to use smart devices, for example, washing machines that automatically start when the electricity price is low because more current is generated than consumed (for example at night times). NILM offers the possibility to determine the individual consumption of the devices used in the household. With this, an individual energy concept can be created for the individual household, which includes the time-optimized use of the devices. Smart devices are particularly suitable for this because they can be started automatically. However, smart devices are usually unable to determine their energy consumption. Thus, NILM remains indispensable for creating the individual energy concept of a household or a company.

NILM offers various more applications, which are described in the next section.

1.1 Application Scenario

In addition to the possibilities mentioned above for cost reduction, there are other scenarios in which NILM can be used. The permanent monitoring of the power consumption allows detecting deviations in the user behavior. In the doctoral thesis of Phillip Klein [87], NILM was applied to the context of Ambient Assisted Living (AAL). In this work, the purpose was to monitor user behavior. On the one hand, atypical user behaviors could be identified by the use of NILM. For example, NILM can detect when people are in a situation of need due to an accident or a health issue. On the other hand, electrical consumers with potential risks are monitored by NILM. This could be the stove or oven but also large industrial machines. One application could be to detect if a stove is switched on, but not off, for a long time. Depending on the situation, either a message could be sent to the residents themselves, or a caregiver could be informed.

Another application scenario for NILM is the detection of failures or misoperation. As an example, the icing of freezing equipment in private households or the industry could be recognized with a NILM system. The use of NILM here is very similar to the previously described task in the AAL context since the aim is to detect unusual situations whereby the trigger is induced

by appliances and not by a human.

Human error can also be detected more efficiently by using NILM. An industrial dishwasher, for example, is already started by the kitchen staff in the morning. Consequently, high energy consumption is generated over the entire day. The dishwasher could be started a few hours later. Typically, such errors are complicated to comprehend, since the person, responsible for the reduction of energy consumption, does not know exactly which devices are relevant for power consumption and how they are used in practice. On the other hand, the persons responsible for the early switching on of the machine are not aware of the costs associated with the operation of such a machine. NILM can provide this information more transparently for non-experts.

Early detection of faults in machines is a different use-case for NILM. A typical current consumption can be detected and corresponding actions can be taken before a complete failure of the device occurs. For example, NILM can be used for fault detection and diagnosis of cooling units [7]. The results showed that some errors could better detected by NILM. These are primarily electrical or electromechanical faults which are more difficult to detect via a thermal measuring. However, the use of additional sensors is advantageous for the detection of other errors. This article concludes that a NILM system connected with existing sensors leads to improved detection and analysis of errors. Another example where a NILM system can be advantageous is in the car. The error “exhaust gas purification” can have multiple causes, from a problem on the engine over the exhaust gas purification to the sensor. Using NILM can readily be determined the case when the heating element of the sensor is defect. This limits troubleshooting, saves time, and reduces costs.

NILM could also be used to optimize the efficiency of electrical storage. The idea is to use stored energy if larger loads are switched on over a more extended period. In the case of big loads, the efficiency of inverters is significantly better compared to small loads. If only small loads are connected, the power supply of these appliances are realized exclusively via the power grid and not with the inverter of the energy storage system because the energy from the battery can be used more effectively in the case of large consumers. With NILM, an improved prediction of the load states is possible, whereby the efficiency of inverters can be increased. The funding of solar energy production in Germany was reduced from 50 ct/kWh (2005) [57], to currently 12.70 ct/kWh (2017) [39]. Therefore, in the future, a large number of solar energy systems will be used together with energy storage systems. This is a very interesting application for NILM since the average electricity price of about 20 ct/kWh (in Europe, as at January 2019) makes it worthwhile for many private solar energy operators to store cost-efficient energy excess and consume the electricity themselves when required.

1.2 Contributions

Following the most important contributions during this work are listed below. The publications are arranged with decreasing date so that the most recent publication is on the top. Parts of these works are presented in this thesis.

- Pirmin Held, Steffen Mauch, Alaa Saleh, Djaffar Ould Abdeslam, Dirk Benyoucef. Frequency Invariant Transformation of Periodic Signals (FIT-PS) for Classification in NILM. IEEE Transactions on Smart Grid, December 14, 2018 [68].
- Daniel Weißhaar, Pirmin Held, Steffen Mauch, Dirk Benyoucef. Device Classification for NILM using FIT-PS compared with Standard Signal Forms. IEEE CANDO EPE 2018, International IEEE Conference and workshop in Óbuda on Electrical and Power Engineering, Budapest, Hungary, November 20-21, 2018 [144].
- Pirmin Held, Steffen Mauch, Daniel Weißhaar, Djaffar Ould Abdeslam, Dirk Benyoucef. Parameter Optimized Event Detection for NILM using Frequency Invariant Transformation of Periodic Signals (FIT-PS). IEEE ETFA 2018, 2018 IEEE 23rd International Conference

on Emerging Technologies and Factory Automation, Torino, Italy, September 4th-7th, 2018 [70].

- Alaa Saleh, Pirmin Held, Dirk Benyoucef, Djaffar Ould Abdeslam. A Novel Procedure for Virtual Measurements Generation suitable for Training and Testing in the context of Non Intrusive Load Monitoring. SIGNAL 2018 - The Third International Conference on Advances in Signal, Image and Video Processing, Nice, France, May 2018 [130].
- Pirmin Held, Steffen Mauch, Alaa Saleh, Djaffar Ould Abdeslam, Dirk Benyoucef. HELD1: Home Equipment Laboratory Dataset for Non-Intrusive Load Monitoring. SIGNAL 2018 - The Third International Conference on Advances in Signal, Image and Video Processing, Nice, France, May 2018 [67].
- Alaa Saleh, Pirmin Held, Dirk Benyoucef, and Djaffar Ould Abdeslam. EMD inspired Filtering algorithm for signal analysis in the context of Non Intrusive Load Monitoring. IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society. Beijing, China, November 2017 [131].
- Patrick Walter Baumann, Pirmin Held, Andreas Heinzlmann, Dirk Benyoucef, Franz Baumgartner. Effizienzsteigerung von Photovoltaik Batteriesteuerung durch geschicktes Load Monitoring. Smart Home and Living Bodensee, Konstanz, November 24, 2017 [19].
- Patrick Baumann, Pirmin Held, Andreas Heinzlmann, Dirk Benyoucef. Increasing the Efficiency of Photovoltaic (PV) Batteries through Non-Intrusive Load Monitoring. EU PVSEC 2017, Amsterdam, September 28, 2017 [18].
- Alvaro Ortiz Pérez, Vera Kallfaß-de Frenes, Alexander Filbert, Janosch Kneer, Benedikt Bierer, Pirmin Held, Philipp Klein, Jürgen Wöllenstein, Dirk Benyoucef, Sigrid Kallfaß, Ulrich Mescheder, and Stefan Palzer. Odor-Sensing System to Support Social Participation of People Suffering from Incontinence. Sensors, December 2016 [119].
- Pirmin Held, Alaa Saleh, Djaffar Abdeslam Ould, and Dirk Benyoucef. Frequency Invariant Transformation of Periodic Signals (FIT-PS) for high frequency Signal Representation in NILM. In 3rd Baden-Württemberg Center of Applied Research Symposium on Information and Communication Systems, pages 1-6, Karlsruhe, December 2016 [69].
- Pirmin Held, Frederik Laasch, Djaffar Ould Abdeslam, and Dirk Benyoucef. Frequency invariant transformation of periodic signals (FIT-PS) for signal representation in NILM. In Industrial Electronics Society, IECON 2016-42nd Annual Conference of the IEEE, pages 5149-5154. Florence, Italy, October 23-26, 2016 [66].

Chapter 2 | Relevant Works & Overview of State-of-the-Art Methods

In order to get an overview of the state of the art, this chapter initially presents the historical development of Non-Intrusive Load Monitoring (NILM). In the first section, an overview of the different areas of NILM is given. Following the overview, the most cited publications are presented in a summary.

Sec. 2.2 describes a detailed state of the art of NILM, where the structure of NILM is presented. Subsequently, the individual tasks of NILM are described in detail.

2.1 Historical Overview

The idea of measuring the total current and voltage of a house and deducting the individual energy consumption of the appliances was published by George W. Hart, working for the Massachusetts Institute of Technology (MIT) and funded by the Electric Power Research Institute (EPRI). In 1985, Hart et al. wrote the report: “Prototype Nonintrusive Appliance Load Monitor” [60]. Some further publications [61] and a patent registration (US4858141 A) [64] followed. Hart et al. already described the essential steps of NILM on which today’s procedures are based, see Fig. 2.1.

For the implementation of the entire NILM process, the tasks shown in Fig. 2.1 have to be solved. For data generation, a measurement system is required. Since the success of all following tasks depends on the quality of the measurements, this is a crucial part. Voltage and current are the measured quantities. The waveform or signal form is calculated out of the measured signals. Usually, the periodical voltage and current signals are transformed into non-periodically waveform like active or reactive power. Additionally, the HARMONICS (HARs) are commonly used. With this non-periodic waveform, the event detection is realized. The task of event detection is to detect if an appliance is switched on or off. With the information about the position of an event, further features can be calculated out of the existing signal forms. These features are used to differentiate the individual consumers through the classification. The task of classification itself is similar to other classification problems from other areas like image or speech recognition. The tracking aims to assign the total power consumption to the individual devices. A more detailed overview of the individual tasks of NILM is given in Sec. 2.2.

The different tasks of NILM, belong to different engineering disciplines, this increases the challenge of developing an overall functional system. First of all, a measuring system has to be developed. This falls into the field of electrical engineering or the special discipline of measurement engineering and embedded programming. Alternatively, public records that are suitable for the development of NILM procedures can be identified. For the selection or construction of a suitable waveform, knowledge about the electrical behaviors of the grid and the individual appliances is required. Additionally, experience how to handle the enormous amount of data is necessary. The event detection, feature extraction, classification, and tracking are common subtasks of machine learning. Due to the many different sub-problems, there are a large number of publications in the area of NILM.

In Fig. 2.2, the number of publications found by the terms “non-intrusive load monitoring”, “NILM” and “Non-Intrusive Appliance Load Monitoring (NIALM)” are plotted over the years. In

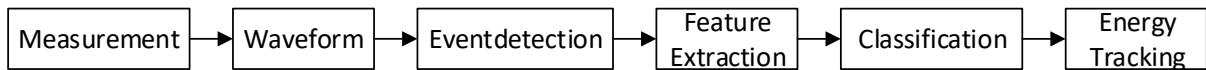


Figure 2.1: Classic approach of the processing chain of NILM

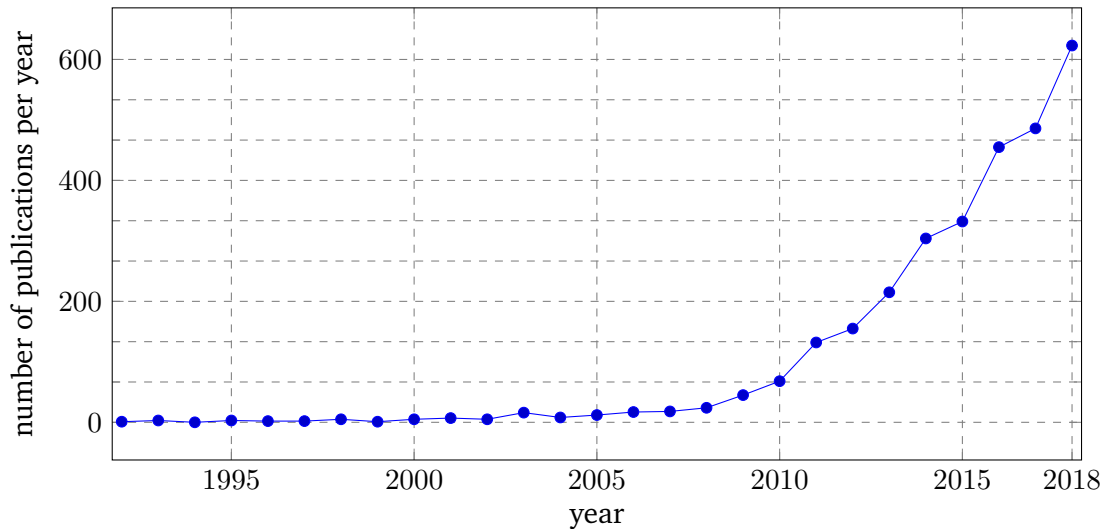


Figure 2.2: Number of publications concerning NILM

The data have been generated by using the software Publish or Perish, product version: 5.25.2.6208.

the years after the first publications of Hart et al., until 2008, the small number of publications, lead to the conclusion, that there was only a small research interest on this topic. After 2008 the number of publications has risen significantly for a variety of different reasons.

During the past two decades, energy awareness has changed. With an increased awareness of energy consumption, a growing interest in how much power is consumed from which device exists. Because of the high effort and costs, this problem cannot be solved with the traditional methods, which require a measuring unit for each individual appliance. The advantage of the idea of NILM is that only one measuring system is required for measuring the total power consumption of all appliances. Out of this total power consumption the individual power consumption of each appliance is calculated by NILM methods.

There are different reasons why up to now no complete NILM system could be developed. For example, the computational power which is required for a successful disaggregation could only be provided by comparatively expensive hardware. The cost for hardware exceeds the cost benefits by optimizing the energy consumption. In the past, it was difficult to use high sampled measuring data for NILM since the computational effort to handle high sampled measurements was too large for low-cost hardware. Also, no public datasets were available. During the last decade, this has changed. With the technical development of powerful single board computers, solving the required task of NILM on a small and cheap system.

With the installation of the smart meter (electronic power meter) in the industry, as well as in private households, the interest of the NILM technology increases further. Besides, there is currently a very strong interest in artificial intelligence in general, as well as in smart home applications. NILM includes with the event detection, classification and partly the tracking very popular topics of current research areas.

To name all relevant papers is difficult because of the large number of publications. For an overview, the most cited publications are listed in Tab. 2.1. The table is sorted by appearance and contains the information on how often the individual publications were cited. The aim is to provide an overview of the most important publications published between 1992 and 2013. More recent publications will be found in Sec. 2.2.

Table 2.1: List of the most frequently cited publications from NILM (01.02.2017)

author	year	references	cites	publication
Hart	1992	[61]	1204	Proceedings of the IEEE
Norford and Leeb	1996	[117]	239	Energy and Buildings
Farinaccio and Zmeureanu	1999	[50]	190	Energy and Buildings
Marceau and Zmeureanu	2000	[108]	146	Energy Conversion and Management
Baranski and Voss	2004	[12]	124	IEEE
Paatero and Lund	2006	[121]	356	International Journal of Energy Research
Lam et al.	2007	[94]	186	IEEE Transactions on Consumer Electronics
Shaw et al.	2008	[134]	167	IEEE Transactions on Instrumentation and Measurement
Kim	2009	[85]	244	Proceedings of the 11th international conference on Ubiquitous computing
Efthymiou and Kalogridis	2010	[46]	408	IEEE International Conference on Smart Grid Communications (SmartGridComm)
Gupta et al.	2010	[58]	297	Proceedings of the 12th ACM international conference on Ubiquitous computing
Kalogridis et al.	2010	[81]	260	First IEEE International Conference on Smart Grid Communications
Molina-Markham et al.	2010	[114]	252	Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building
Liang et al.	2010	[101]	253	IEEE Transactions on Power Delivery
Froehlich et al.	2011	[53]	201	IEEE Pervasive Computing
Kim et al.	2011	[84]	256	Proceedings of the 2011 SIAM International Conference on Data Mining
Kolter and Johnson	2011	[89]	460	Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA
Zeifman and Roth	2011	[151]	428	IEEE transactions on Consumer Electronics
Barker et al.	2012	[13]	166	SustKDD
Kolter and Jaakkola	2012	[88]	217	AISTATS
Zoha et al.	2012	[155]	239	Sensors
Fan et al.	2013	[49]	309	IEEE Communications Surveys and Tutorials
Carrie Armel et al.	2013	[6]	194	Energy Policy

Older publications are favored in this review because of the higher likelihood that they have been cited. Since NILM was not always evenly present in the research community, it is also problematic to use citations per year as a criterion.

In the following lines, a brief overview of the publications from Tab. 2.1 is given.

Norford and Leeb presented in their paper, the difficulties that arise for a NILM system in a commercial building [117]. As an application, they investigated how NILM can monitor the correct power-up function of multiple devices. Also, power oscillations of a poorly turned chiller could be detected.

In 1999, Farinaccio and Zmeureanu described a process which determines the electricity consumption of a domestic water heater and a refrigerator from the total consumption of a domestic building. For both devices, a separate process was developed [50]. The method requires sub-metering of the target appliances for the training to find the individual characteristics. For classification, the power consumption is analyzed over several seconds. Through device analysis during training, both devices can be detected. The error for calculating the energy consumption is between -10.5% and 15.9% .

In their work, Marceau and Zmeureanu [108] described an algorithm which extracts two features of a device, in a learning phase. In a second evaluation phase, these devices are extracted from the characteristics of the overall consumption. As an example a realization for a water heater and a baseboard heater, as well as for a refrigerator is presented. The error of the estimated energy for the three appliances is less than 10% .

Baranski and Voss described in their paper an optical measuring system which can be placed on an existing Ferraris meter ³⁾. Hence, recording of the active power every second can be realized with fewer effort [12]. Unknown applications can be detected by using clustering in combination with genetic algorithms. The genetic algorithm is used to vary the parameters for the event detection.

Paatero and Lund [121] describe a procedure in which a household consumption can be simulated, hourly based on statistical probabilities and external influences, for example, temperature or day of the week. The different consumption patterns between business days and weekends can be distinguished. At weekends, most people sleep longer than during the week, and most of them are away from home during the day. Also, on weekends people spend more time with cooking than on weekdays. Of course, this depends very much on the residents.

Nevertheless, most people show a different behavior on the weekends than during the week which can be analyzed with the power consumption. A case study with a demand side management system was applied with the goal of reducing peak consumption. A reduction of the peak consumption of 7.2% was realized by using demand side management on cooling appliances like freezers or refrigerators.

In the publication [94] of 2007, Lam et al. examined the load signatures of individual appliances by recording the current and voltage signal with a high sampling frequency. The signals are displayed in voltage-current (V-I) trajectories. Due to the comparatively high sampling frequency, device-specific characteristics occurring within a period can also be taken into account with this approach. Therefore, new features for the classification have to be developed. Lam et al. show that an improvement in device detection is possible with V-I trajectories.

In their work, Shaw et al. focused on the transient state changes of individual devices [134]. In addition to cost reduction, this also includes the examination of diagnostics and the monitoring of electromechanical systems.

Kim et al. described in their work a measuring system which determines the power consumption of individual devices utilizing external sensors (such as magnetic, light and acoustic sensors) [85]. The additional sensors provided an alternative solution approach for finding the ground truth data⁴⁾.

³⁾The Ferraris meter, named after Galileo Ferraris are still widespread in private households even if they are gradually replaced by smart meters. The Ferraris meter is based on the induction principle.

⁴⁾Ground truth data or reference data are necessary for the NILM context to evaluate the algorithms. Depending on the concrete task the requirer information of the ground truth varies.

With higher sampling frequencies of the smart meters, more information about the devices and the user behavior can be obtained. A critical issue of NILM technology is the question of data protection. Currently, only low frequencies (usually one per year) information is used to calculate the costs of energy. Working with higher sampled data allows reconstructing the user behavior. The problem of data protection in NILM is investigated in [46]. The proposed method contributes an additional layer for security. So the privacy protection is not realized directly on the smart meter.

The active and reactive power is most frequently used in signal representation for NILM. To make use of high-frequency components other waveforms like Electro-Magnetic Interference (EMI), which arises through Switch Mode Power Supply (SMPS)⁵⁾, can be used [58]. Therefore, a frequency spectrum of up to 500 kHz is investigated. Particularly small loads (like mobile phone chargers or energy saving lamps) can be distinguished from each other by applying this waveform.

With the use of NILM the user behavior can be analyzed, because of this, the data protection has to be discussed. Theoretically, sensible information about the behavior of people living in a household can be generated from a measurement system outside the house. One way of solving the problem of data protection in NILM is to use a battery [81]. The battery is used to obscure the use of individual devices. Thus, the battery can be charged to replicate the consumption of specific devices; for example, it could simulate a TeleVision (TV) in the evening. The effort of this method primarily depends on the size of the consumption of the appliances. The power consumption of large consumers, for example, a washing machine is difficult and expensive to hide from the smart meter, since batteries with high capacity are required.

The work of Molina-Markham et al. investigates the extent to which smart meters can obtain private data and user content [114]. A sample rate of one second and measurements over two months are used. The information generated by smart meters can be used to obtain information about the behavior of people living in a house. Information about how many people live in a house, their sleeping or eating routines, is extracted.

Different waveforms and features are compared in [101], where an allocation of the load signatures in micro and macro level is investigated. In this work, waveforms with a sampling rate higher than the mains frequency are classified at the micro level. If the sampling rate is lower than the network frequency (for example, 1 sample/s), then, the features are assigned to the macro level.

The advantages and disadvantages of different waveforms are discussed in [53]. The paper uses the continuous voltage noise signature for the classification. A sampling rate of up to 500 kHz is used. An investigation of SMPS devices shows that there is a minimum of overlapping in the frequency spectrum because of variety during manufacturing and the power requirements of the SMPS devices.

In the classical approach, training data are recorded, and the NILM system is trained with these recorded data. In practice, the training effort is very high since it is difficult to generate the reference data. Unsupervised learning or classification do not require previous information or training data. Since the training process is a difficult task unsupervised methods, which do not require training, are advantageous. Using unsupervised disaggregation methods can provide appliance power usage information [84]. Here, different measurements from seven homes are used. The proposed algorithm achieved an accuracy of up to 99.8% for four devices and up to 87.4% for up to ten devices.

For a better comparability of the different NILM approaches, the Reference Energy Disaggregation Data Set (REDD) [89] was created. The dataset is taken from five houses. It is measured with a sampling frequency of 15 kHz during the turn on and off positions of appliances. During the Steady State (SSt), where in general no significant change in the signal exists, the available

⁵⁾Switch Mode Power Supply (SMPS) belongs to the group of the converter. SMPSs are used to convert an Alternating Current (AC) or Direct Current (DC) into a DC. The high efficiency and a compact method of construction are the main advantages compared to power supply transformers.

sample frequency is 1 Hz.

In the overview given in [151] from the year 2011, different waveforms, features, and algorithms are considered. The authors conclude that no NILM system is available which recognizes all devices in a household. There is also no feature that can be used reliably for all devices. By combining multiple features, the recognition rate of the event detection can be increased.

Another public dataset is “smart” [13], which provides a sampling frequency of 1 Hz. Additionally, data from over 400 anonymous homes with a sample rate of one minute is available.

In [88], an extension to the Hidden Markov Model (HMM) is introduced. The used classifier works unsupervised. This results in a good classification of most devices.

In [155] a review of NILM methods has been presented. One essential conclusion of this work is, that no feature or disaggregation algorithm which applies to all types of appliances.

The problem of data protection in intelligent energy networks are discussed in the publication [49]. The paper presents an overview of the smart grid communication its challenges and opportunities.

A detailed description of the state of the art of NILM is described in [6]. An overview of the increased performance due to higher sampling rates is given. Depending on the sampling rate of the smart meters, different hardware is required. A higher sampling rate leads to increasing costs of the overall system. Using sampling frequencies of up to 2 kHz can be realized cost efficiently. For higher sampling rate a hardware update will be required.

Summary

NILM becomes more and more popular over the last years. The different papers shows that the research field of NILM is widely spread. The most cited papers handle topics from data generation, measurement systems, feature extraction, classification, and data protection. On the one hand, there are publications from people with electrical background about the measurement system and the data generation. On the other hand scientists with a computer science background, publishing in the area of event detection, feature extraction and classification.

To solve all different tasks of NILM is difficult since the individual tasks are complex and can only be solved with the necessary basic knowledge from the relevant special fields. Therefore, a specialization was made to the individual areas. However, the problem with isolated processing of the individual thematic tasks is that under certain circumstances a priori knowledge from another area of specialization is not sufficiently used. For example, an electronics engineer who builds the measuring system knows little about the following tasks in the field of machine learning. The focus of a computer scientist who is an expert in machine learning lies mainly on the development of event detection or classification methods. Therefore, complex machine learning methods often presented in combination with low-sampled standard waveforms.

The problem of developing a suitable waveform for NILM is not finally answered. A detailed description of the current waveforms and their problems follow in Sec. 2.2.2. This section is not exhaustive but should give a first overview of the various work in the field of NILM. Now that some of the most important publications on NILM have been presented, a state of the art of the individual NILM editing steps follows.

2.2 State of the Art Methods in NILM

This section provides the state of the art of NILM, which introduces the individual tasks in detail. The mathematical modeling of the fundamental problem of NILM is described in Equ. (2.1) [42].

$$y(t) = \sum_{d=1}^D y_d(t) \quad \text{for} \quad t = 1, \dots, T \quad (2.1)$$

The sum of power consumption of several devices d at a time t , is given by $y(t) \in \mathbb{R}$. D describes the number of devices in a building and the power consumption of the device d at time t by $y_d(t)$.

In reality, neither the number of devices nor their properties are known. From a mathematical point of view, there is a poorly set inverse problem which is described in Equ. (2.1).

The inverse problem cannot be solved directly without a-priori information. Different information or assumptions can be used in order to solve the problem. The individual devices usually are not switched on and off at the same time. The runtime and the frequency of the appliances are most of the time different. Besides, the devices have commonly different power consumptions. Depending on the device, a positive or negative phase shift occurs with capacitive or inductive consumers while the current leads in the case of capacitive and the voltage for inductive loads. Other appliances are operated with a phase-cut control. Nonlinear loads in example transformers or fluorescent lamp generate HARs. Switching power supplies can also feedback high-frequency interferences to the mains. With these different characteristics, the operation of individual appliances from the total consumption can be determined.

The solutions in the literature very often handle only one task of NILM. For example, only the event detection or classification. Commonly, the results reach a very high accuracy. A more sophisticated analysis of the individual paper shows the current limitations. The number of devices is eminently reduced, or individual devices are excluded from the statistics because of problems of their detection or disgusting. Moreover, the number of simultaneously active appliances is often low or only one active device at a time is assumed. The particular limitations of the existing literature create difficulties of comparing the different methods. One reason for this limitation is the available datasets which will be discussed in more detail in Sec. 2.2.1. Another reason is the fact that there are multiple application scenarios (as described in Sec. 1.1) by weighting the different priorities. While for one application the recognition of a low number of wrong detected events might be more important, another application has the focus on detecting not too many events.

However, to construct a NILM system in general, several tasks from different engineering disciplines have to be solved. The whole linear process chain is visualized in Fig. 2.1. The disaggregation process can be described as a linear process chain which consists of the following six phases:

- ① measurement
- ② waveform
- ③ event detection
- ④ feature extraction
- ⑤ classification
- ⑥ tracking

In the literature of NILM the linear structure of the individual tasks is used, almost exclusively. Usually, the information flow is described as a unidirectional process of the different tasks. The information goes from the event detection to the feature extraction but not vice versa. The linear structure is well suited to present the different tasks and a categorization of the current topic.

Nevertheless, on closer examination, it turns out that this structure is not quite right in the sense of the information flow. A more detailed illustration of the information flow is presented in Fig. 2.3

Current and voltage are the measured quantities provided by the measurement system ①. Waveforms or signal forms ② are calculated out of the voltage and current signal. Usually, these waveforms are non-periodically like active and reactive power, see Equ. (2.1) or other waveforms like V-I trajectories [11] where current and voltage is used directly. Transforming the signal into the frequency domain is also a common method. The S-Transform can be used in order, to contain the time information [120]. In the literature, the terms waveform, signal form,

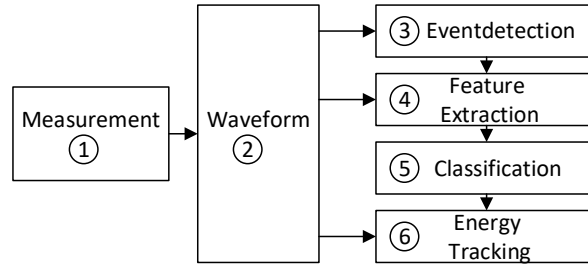


Figure 2.3: Common information flow in NILM. In comparison to the usual linear presentation of the processing chain of the NILM tasks, see Fig. 2.1, the information flow is not purely linear in a row.

and feature are used ambiguously. Therefore, in this work, the signal which is calculated from current and/or voltage, without further information, such as the position of an event, is defined as a waveform.

The event detection ③ is used to identify on and off times of the individual appliances [148]. Therefore, one or more waveforms are used.

The information of the event is used for the feature extraction ④. With the position of the event, new features can be calculated out of the waveform, for the subsequent classification.

The term feature extraction is used in literature for both the signal representation directly after the measurement and for the extraction of information after the event detection. Nevertheless, from now on the acquisition of information after the event detection is described as feature extraction which is generating one or more “features”. The directly used or calculated signal after the measurement and before the event detection is described as “signal form” or “waveform”.

The classification is one of the most important tasks in NILM since here the identification of the individual devices are realized ⑤. The aim of classification is to allocate the occurring events to the specific appliances. Therefore, unsupervised and supervised approaches have to be distinguished. For supervised learning training data are available in order to train the classifier with the training features provided by the feature extraction. Unsupervised learning does not require training data. The features are assigned according to their similarity. Ideally, this results in a cluster⁶⁾ for each device. Depending on the application the classification is the last task in NILM. This applies for all applications with the focus on extracting information of running appliances and user behavior like in the Ambient Assisted Living (AAL) context. Finally, energy tracking ⑥ is carried out and the actual energy demand of the individual device is determined.

Depending on the concrete application of a NILM system, different goals have to be achieved. Fig. 2.4 describes the different ways how in NILM the error can be calculated. In the classical use of NILM, the aim is to allocate the power consumption of each device out of the total power consumption $Y(t)$. The individual errors of the power consumption of the devices are calculated by $e_D(t) = Y_D(t) - \hat{Y}_D(t)$, where $D = 1, 2, \dots, d$. The sum of the absolute values of $e_D(t)$ results in the sum error $e_{sum}(t)$ of all appliances. In addition, the sum of all detected consumptions \hat{Y} is often compared to the original total consumption $Y(t)$. This results in the error $e_{tot}(t)$. Often the customer of a NILM is interested in the active power consumption. In this case, Y is set equal to the active power consumption. In the industry additionally, costs can arise for reactive power. Therefore, active and reactive power can be calculated individually, or the apparent power can be used.

There are different ways to indicate the accuracy of the classification. In applications where the aim is to identify the individual devices, the error is calculated by comparing the classified devices with the reference data. Another method to calculate the accuracy, for this scenario, is

⁶⁾A cluster consists of objects that have similar properties and that differ from objects that are not in the same cluster.

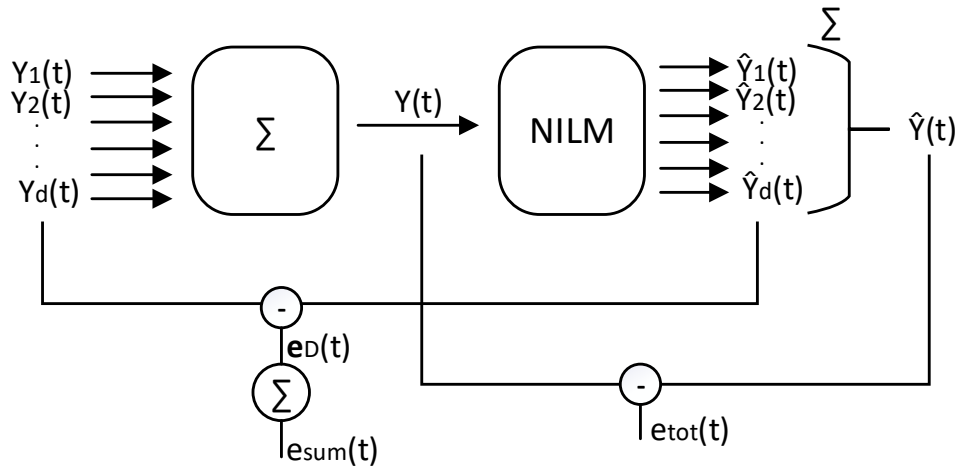


Figure 2.4: Presentation of the different error calculations in NILM

sometimes realized with the correct assignment of the power consumption. The problem with this rating criterion is that devices with high power consumption are weighted more heavily than devices with low power consumption. With the goal of getting the operation information about the appliances the power consumption, usually plays a subordinate role and should not be used for the evaluation.

In the following sections, the state of the art of the individual tasks of NILM is presented.

2.2.1 Measuring Systems and Datasets

To develop and test algorithms for NILM different datasets have been created. Therefore, suitable measuring equipment is required. There are different approaches. Some follow the approach to embedding the NILM algorithms in the measurement system. Others merely measure with the embedded system and realize the appropriate calculations relevant to NILM in the cloud, as there is significantly more computing power available. Here, the advantage is that the algorithms can continually be adapted or improved. The data available in the cloud can also be used to improve other NILM systems. Since this is difficult at higher sampling rates, because of the amount of data, this procedure is usually used only at lower sampling rates (from 1 second up to a few minutes). Also, the protection of private data is more challenging for cloud-based systems.

To realize an investigation of different waveforms the initially measured voltage and current is required. With a higher sampling frequency, more details about the specific power consumption of the individual measured appliances can be generated. More detailed information is advantageous for the subsequent event detection and classification. The necessary sampling rate depends on the application and the appliances which have to be disaggregated by NILM. During this work, predominantly sampling rates between 1 kHz and 12 kHz are used. A realization on comparatively inexpensive embedded systems is currently possible with this sampling rates. On the other hand, the sampling frequency is high enough to calculate different waveforms and features like HARs.

An investigation of existing measuring systems for NILM is presented in the next section, followed form an examination of different publicly available datasets for NILM.

Measuring Systems

Commercially available measuring systems often work internally with a very high sampling rate. However, only the processed waveforms are stored and provided to the user. The “Fluke 1738 Advanced Power Energy Logger”, for example, uses a sampling frequency of 10.24 kHz. The high sampling frequency is used for the calculation of HARs. The minimal interval of the logger,

where the different waveforms like active, reactive, apparent power, or HARs are stored, is one second [52]. Unfortunately, the data logger does not provide the possibility to log the current and voltage directly. Also, there is usually no way to export this high-frequency data. Typically, these power meters or analyzer allow only to extract specific signal forms like active or reactive power or the different HARs.

To measure the original current and voltage signal with a higher sampling rate, of several kHz, new measuring systems are developed [113, 122, 22]. These measurement systems use an analog part where commonly a level adjustment, filtering, and galvanic insulation is realized. While the filters ensure that no aliasing occurs, the galvanic isolation is used for safety reasons. The conversation from analog into digital values is realized with Analog-to-Digital Converters (ADCs).

For example, a commercially available ADC card can be used. Also, audio input devices are commonly used. The high-pass filter which is used in this kind of measuring equipment causes that there are no DC components in the signal. Therefore, for example, the waveform of a half-wave rectifier is not displayed correctly.

The documentation of non-commercial measuring systems often does not provide all information about the measuring system. Sometimes, there is no information available whether the measuring system can measure current and voltage synchronously. The design of the analog part and especially the filtering is often not clearly documented.

Nevertheless, datasets generated out of such measurement systems are used for the development of NILM algorithms. These uncertainties must be taken into account in order not to draw false conclusions. For example, a low-pass filter which was designed in such a way that already high frequencies are strongly suppressed leads to the fact that high-frequency signal components are filtered out of the signal. If this is not known because of a lack of documentation of the measuring system, the impression could arise that the consideration of high-frequency signal components for NILM does not cause any improvement. There is also a risk that aliasing can occur due to an incorrectly designed low-pass filter. In this case, the results of the different NILM can be worse than without aliasing. For this reason, the generation of data needs to be considered.

Another problem of measurement systems existing of several subcomponents is the difficulty to use these measuring systems in real-world environments, outside a laboratory. As a rule, the measuring system must be accommodated in the control cabinet. If additionally, a computer or a laptop is needed for the operation of the measuring equipment (ADC card or audio card) there is often not enough space in the control cabinet. For a clean installation, a compact measurement system is required.

During this work, a compact measuring system including a data logger for high frequencies of up to 10 kHz on three phases was developed from the Institute for Smart Systems (ISS) working group. The measuring system can measure on three phases voltage and current simultaneously. Algorithms developed during this work are implemented on the measuring hardware with the goal to develop a real-time NILM system.

For this work, public datasets will be investigated on their suitability for the different tasks. These public data sets will be added with own measurements which will be measured with the measurement created by Thomas Bier during his doctoral thesis [21]. Depending on the available measurement system, different sampling frequencies are available in public datasets.

Datasets

For the development of NILM algorithms, different public datasets are available. As can be seen in the following list, there are datasets from real-world scenarios as well as laboratory measurements available. The two types of datasets are suitable for different NILM applications.

- Real-world scenario
 - REDD (1 Hz/15 kHz) [89]
 - BLUED (12 kHz) [4]
 - UK-DALE (16 kHz) [82]
 - ECO (1 Hz) [20]
- Laboratory measurements
 - PLAID (30 kHz) [54]
 - WHITED (44 kHz) [80]
 - COOLL (100 kHz) [124]

Datasets measured in real-world scenarios have the advantage that they contain statistical information of the user behavior. Usually, the measurement is done continuously over several days, months, or even years. This creates enormous amounts of data. For example the Building-Level Fully-labeled dataset for Electricity Disaggregation (BLUED) dataset, in which two phases are measured with 12 kHz over one week, has a size of 230 GB (converted in Network Common Data Format (NetCDF)⁷⁾). This amount of data needs to be handled during the development of NILM algorithms.

Existing laboratory datasets store only the measurement during a switching event. Compared to real-world measurements, the amount of data is much lower. The position of the event in the ground truth data is in laboratory datasets much more exact. Usually, there are no errors in the ground truth data of laboratory datasets. This is because under laboratory conditions all appliances are entirely under control. Therefore, the exact time when an appliance is switched on or off is also known.

Real-world datasets have problems with unlabeled devices. These devices lead to False Positive (FP) in the evaluation of the event detection algorithm which is problematic since the algorithm detect the event of the unlabeled appliances correctly. With datasets which contain unlabeled appliances, it is theoretically possible that a poor event detection method leads to better results compared to a correctly working event detection.

The development of an event detection requires a continuous measurement, to ensure that the algorithm does not produce too many FP, during the steady states. Therefore, real-world measurements should be preferred for testing event detection algorithms. Additionally, they have the advantage that the appliances are used in the correct probability. The event detection requires, only the position of an event in the dataset from the ground truth data. This is necessary for both supervised (methods which require training and reference data) and unsupervised (methods which learn without reference data) methods since even an unsupervised event detection need the references during the test to decide if an event was calculated correctly or not, or if an event was not recognized.

For classification, the ground truth data require additional information like the ID of the appliances. Also, the information if the current event is a switching on or off event is useful. In case of training a classifier (supervised classification) several training features of each appliance are necessary. Real-world measurements have the problem that several appliances are used very rarely. Hence, training and testing a classifier is difficult with real-world measurements since there is not enough statistical information about all individual appliances for training. Also for the testing, a low number of switching events of some appliances is problematic. If, for example, the refrigerator has by far most events, as it constantly turns on and off all day long and the classifier detects only the refrigerator correctly the total achieved performance could be very good. The refrigerator has in some datasets much more on and off events than all other appliances together [4]. This needs to be considered when using real-world datasets for event detection and/or classification.

Another problem of real-world datasets are problems in the reference data. For the calculation of the features for classification, the exact position of the events is required. Apart from the BLUED dataset in which the events are precisely labeled, there is usually an inaccuracy of several seconds in the reference data. The calculation, of required features for the classification,

⁷⁾NETwork Common Data Format (NetCDF) is a file format for the exchange of scientific data. There are interfaces for the common programming languages. More information can be found at the at the University Corporation for Atmospheric Research (UCAR) homepage where also the project homepage is hosted [2]

is problematic since the events are not or only partly available as a reference. The reference data needs to be extracted from the usually available information of the continuous power measurement of each appliance. For the REDD and United Kingdom Domestic Appliance-Level Electricity dataset (UK-DALE) datasets, the sampling frequency of these reference data are two and six seconds [89, 82]. Additionally, the reference data of each device and the total power consumption available with a high resolution needs to be synchronized. Since some appliances have more than one SSt or a dynamic power consumption errors can occur during the calculation of their events for the reference data.

For classification, laboratory datasets are commonly used. For these kinds of datasets, it is easier to ensure that the number of events for all devices is more or less the same. Furthermore, the handling of the data is more comfortable since the ground truth data are more precise and the amount of data is smaller.

Up to now, there are mainly laboratory datasets which include continuous measurements. Also, the existing laboratory measurements do not contain simultaneously active devices. Since the simultaneous operation of appliances is very often the case in practice, a NILM system, for practical use, has to handle this.

The generation of data records for NILM is complex since an intrusive measurement environment with all its disadvantages is required for generating reference data. A data logger recording the current and voltage is needed. Depending on the used sampling frequency this must be developed since most commercially available measurement systems are not able to lock current and voltage with a high frequency. Storage is often not possible with standard systems or only for a few hours.

In addition to the total consumption, the individual devices have to be monitored or measured, too. This is necessary to generate reference data. These reference data are required for the evaluation of the different stages of NILM (event detection, classification, and tracking). The event detection requires only the information of the position of an event. For classification additionally, the information about the device and ideally whether it was switched on or off is required. The tracking requires at least the sum of the power consumption during each active state of a device. However, to have the power consumption over the total time of each appliance would be optimal for the evaluation of the tracking.

Tab. 2.2 provides an overview of the currently available datasets. The basis of this was the compilation of [26] which was supplemented with current datasets. Most records work with a low sampling frequency. There are a number of professional measuring devices that support data logging with this resolution. For high-frequency resolutions, custom data loggers have been used. Tab. 2.2 present six data sets including high resolution data of several kHz (REDD [89], BLUED [4], UK-DALE [82], PLAID [54], WHITED [80], and COLL [54]). For this work, exclusive data sets with a sampling rate of at least 1 kHz are of importance, and only these are described in more detail. Datasets with a lower sampling rate do not provide the initially measured voltage and current signal. Therefore, only the waveform which is provided by these datasets can be used for all further tasks of NILM.

The REDD dataset is sampled with 15 kHz but only at the transients (when a device has been turned on or off). The other area of the signal is sampled with one second. For event detectors which make use of high-resolution signals, the use of this dataset is unfavorable.

The reference data were recorded as a power signal with a sampling rate of two seconds. This is especially useful for later tracking since the actual power requirement of each device is available here. This is particularly advantageous for devices which have a dynamic consumption. However, for event detection, the on/off points must first be calculated for each device. For events which occur with a short time distance, this is partially fault-prone.

The structure of the UK-DALE dataset is similar to the REDD dataset with the difference that a sampling rate of 16 kHz over the whole time and not only during the events is used. Similar as in the REDD dataset the reference data are generated by measuring the individual appliances. These measurements have a sample rate of six seconds. The low sampling rate of reference data

produces the same problems as described already for the REDD dataset.

The BLUED dataset uses a sample rate of 12 kHz. The turn on or off events of the different appliances is marked with indices. This data set is very suitable for the development or testing of event detection algorithms. A weakness of this dataset is the limited measuring time of only eight days. Classification with this data set is problematic since individual devices are switched only rarely. Some devices like the “Kitchen Hallway light” are switched on and off only a few times. Other devices like a sub-woofer or an Apple TV in the basement have not switched on or off at all. Hence, the dataset is not ideal for classification. Other problems occur since only appliances with a power consumption higher than 30 W are labeled in the reference data. In the high-frequency measurement, the devices with low power consumption are still available. These cause events without corresponding ground truth data and occurs mainly in the second phase of the data set.

PLAID, WHITED, and COOLL are laboratory datasets with a sampling frequency up to 100 kHz. They contain individual measurements of each device but no measurements where multiple numbers of appliances are active at the same time.

In addition to these datasets, our working group has created own datasets in recent years. They were recorded with our measuring system which allows a maximum resolution of 4 kHz. However, they were down-sampled to 50 Hz. The measurement system and also some of the datasets are created during the doctoral theses of Thomas Bier [21] and Philipp Klein [87]. The measurements consist of individual records of different household appliances which have been switched on and off automatically, as well as a multi-device measurement in which several household appliances were switched on and off randomly. In contrast to the publicly available data sets, the advantage here is that the ground truth data have a higher accuracy. This is not always given for the publicly available measurements and leads to considerable problems during the training and evaluation.

Depending on the individual NILM task, different kind of dataset have to be selected. After a suitable dataset is chosen, the measured voltage and current signal need to be represented in a suitable waveform. Therefore the state of the art of waveforms used in NILM is described in Sec. 2.2.2.

Table 2.2: Comparison of NILM datasets

dataset	location	reference	year	sample rate aggregate measurement	sample rate ground truth	duration	number of houses
REDD	USA	[89]	2011	1 sec / 15 kHz	2 sec	3 to 19 days	5
BLUED	USA	[4]	2012	12 kHz	events are la- beled	8 days	1
UMass Smart	USA	[13]	2012	1 sec	1 sec	3 months	3
Tracebase	DE	[129]	2012	N/A	1-10 sec	N/A	15
Pecan	USA	[72]	2012	1 min	1 min	7 days	10
Street Sample HES	UK	[154]	2012	2 min or 10min	2 min or 10 min	1 to 12 months	251
AMPds	CDN	[107]	2013	1 min	1 min	1 year	1
iAWE	IND	[16]	2013	1sec	1 or 6 sec	73 days	1
BERDS	USA	[106]	2013	20 sec	20 sec	1 year	N/A
UK-DALE	UK	[82]	2014	16 kHz	6 sec	3-17 months	4
GreenD	AT/IT	[115]	2014	1 sec	1 sec	1 year	9
COMBED	IND	[17]	2014	30 sec	30 sec	18 months	8
ECO	CH	[20]	2014	1 sec	1 sec	8 months	6
SustData	PT	[123]	2014	50 Hz	50 Hz	5 years	50
PLAID	USA	[54]	2014	30 kHz	transients ⁵⁾	5 s (per device)	56
WHITED	GER	[80]	2016	44 kHz	transients ⁵⁾	5 s (per device)	N/A
COOLL	USA	[54]	2016	100 kHz	single devices	N/A	N/A

2.2.2 Waveform

In this work, the terms waveform and signal form are used with the same meaning. They describe the result of the transformation from the measured current and/or voltage signal into a new waveform which has, compared to the originally measured waveforms, advantages. For example, out of two measured signals (voltage and current), one waveform is generated which contains the information of both signals at the same time. This reduces the complexity of the subsequent event detection since only one signal needs to be considered.

The measured voltage and current signal of the grid are periodic signals. For simplification of the further stages of NILM, the signals are converted into a non-periodic waveform such as active, reactive, or apparent power. The advantage of these type of waveform is that during a SSt where no device is switched on or off the amplitude of the waveform is relatively constant. To see multiple events in a figure, an observation time of usually several seconds or minutes is necessary. Because of the grid frequency of 50 Hz, the single periods are not visualized in detail as illustrated in Fig. 2.5. The further tasks of NILM, for example, the event detection, or the feature extraction can handle a non-periodic signal with less effort.

Another advantage of non-periodic waveforms is the lower sampling rate which can be used for standard power signal forms, like for example (active power P , reactive power Q , and apparent power S). On the other hand, not all information from the initially measured current and voltage signal is included in the calculated power signal forms which will be described in detail in Sec. 2.2.2.

⁵⁾PLAID and WHITED are laboratory datasets. They include measurements of the transients of single devices.

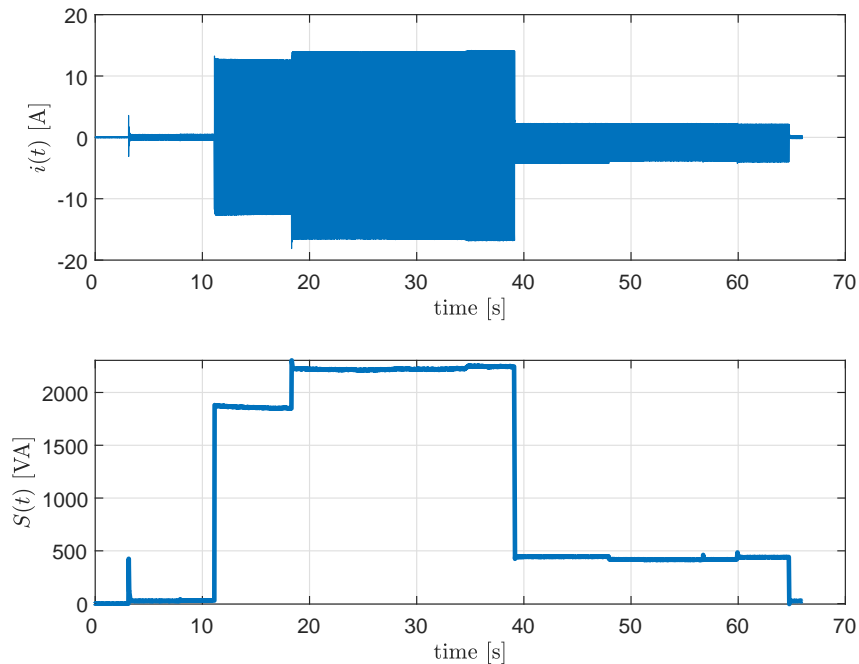


Figure 2.5: Comparison of the periodical, original current signal ($i(t)$) to the non-periodical complex power (S) signal form

For selecting a suitable waveform, the available sampling rate must be taken into account. In the beginning of NILM in 1992, a low sampling frequency was used [61]. In newer literature much higher frequencies, up to several kHz, but also methods which require a low sampling rate, are used.

The advantage of higher sample rates is that additional waveforms can be used for the NILM process. Common waveforms which can represent the information of high sampling rates are, for example, HARs in the frequency spectrum. In [97] for example, this higher HARs are used to detect the power consumption of variable-speed drives used in the industrial field.

The necessary sampling rate depends on the number of appliances, the kind of appliances, and their respective differences. Generating a general statement about the number of devices that can be distinguished at a particular sampling rate is difficult, as it depends on many factors such as the distinctiveness of each device or its operation. Nevertheless, there are rough guidelines from the literature [6].

- sample rate up to 1 Hz → No. of maximal distinguished devices: ≈ 10
- sample rate up to 60 Hz → No. of maximal distinguished devices: ≈ 20
- sample rate up to 40 kHz → No. of maximal distinguished devices: ≈ 40
- sample rate up to 1 MHz → No. of maximal distinguished devices: ≈ 100

The goal of the enumeration is not to give an exact statement about how many devices can be distinguished with a sampling frequency. Further more, it shows that with more details that can be detected with a higher sampling rate, a more accurate distinction is possible.

The exact number of appliances which can be classified with a specific sampling rate depends very much on the combination and the appliances types. For example, appliances with similar power consumption are harder to distinguish for a classifier, which uses only the active and reactive power, than appliances with a clear difference in consumption. Therefore other or additional waveforms containing high-frequency information can be useful. The higher the sampling rate, the more details in the signal can be recognized, which is helpful for the classification of appliances since there is a higher chance of differentiation between appliances with a similar consumption.

However, this requires a corresponding waveform and feature extraction, such as HARs in the frequency space [97] which make use of the additional information.

Nevertheless, the high-frequency signal components are only suitable in combination with other features. They include comparatively little information which can be used for device recognition if they are used exclusively as shown in [140].

The commonly used NILM waveforms need to be combined to minimize the loss of information. This combination of different waveforms was already shown in different publications [151, 75]. A problem of combining different waveforms is that for each device setting other kinds of waveforms are required, and also the weight of different waveforms depends on the device setting.

Another positive aspect of a high sampling rate is that a more significant number of features can be created. For example, the use of transient states as a feature requires a sample rate much higher than 1 Hz because the maximum length of a transient state is only up to three seconds for a few devices. Most appliances used in a private household do have a transient state shorter than one second.

For the NILM process, known waveforms such as active power, reactive power, and HARs are commonly used because of different reasons. Commercially available energy meters can store these type of waveforms [1]. Usually, a reduction of the sampling rate to 1 Hz is realized. This allows easy entry to the topic of NILM since no development for specific measuring systems is necessary and first results can be achieved easily. Due to the reduced sampling rate, the computational effort is comparatively low, and the data handling is simple. In general, these realizations have different limitations. Usually, they are only able to detect a limited number of devices. Similar devices are summarized into groups. A distinction of the individual appliances in a group is not realized. The performance of the classification is often judged according to the correctly assigned power consumption. This already suggests good results if only the appliances with the largest consumption were recognized correctly.

The available literature of NILM deals much more with new event detections, classifications, or tracking methods and only a limited number of working groups deals with the development of waveforms for NILM.

The use of V-I trajectory as waveform for NILM process, achieve good results [65, 75, 43]. [75] uses trajectories combined with traditional features, [65] and [43] only use trajectory as input waveforms. The use of trajectories as in [43] provides the advantage that no transformation from the time domain to the frequency domain is necessary. Furthermore, no additional waveforms are needed. All these advantages reduce the calculation effort. However, compared to traditional features, the disadvantage of trajectories is the complex analysis of the signal. The graphical signature has to be extracted as described in [43]. In Sec. 2.2.2 the use of trajectories for NILM is discussed more in detail.

Since the measured voltage and current signal are of particular importance for the calculation of waveforms, they are presented in detail in the following section.

Originally Measured Waveforms

The voltage and current signal are the measured quantities, thus, all available information is contained in both signals. Further signals are derived from these variables. In Equ. (2.2) and (2.3), the current $i(t)$ and the voltage $v(t)$ are described. $I(t)$ and $V(t)$ represent general nonlinear signals. The nonlinear signals lead to higher HARs of the mains frequency. The time-dependent frequency is denoted by $f(t)$ and the phase angle between $v(t)$ and $i(t)$ is given by $\varphi(t)$ [66]. Both signals ($v(t)$ and $i(t)$) include noise which can be decried by $n_v(t)$ for the voltage and $n_i(t)$ for the current. Since the phase shift can only be calculated between two signals, the phase shift of the voltage is assumed to be zero.

$$v(t) = V(t) \cdot \cos(2\pi f(t) \cdot t + \phi(t)) + n_v(t) \quad (2.2)$$

$$i(t) = I(t) \cdot \cos(2\pi f(t) \cdot t + \varphi(t) + \phi(t)) + n_i(t) \quad (2.3)$$

The frequency f , as described in Equ. (2.2) and Equ. (2.3), is not constant. The frequency is subject to two main factors of influence. The variation from the 50 Hz are relatively low, (<0.2 Hz) in the European interconnection network [47]. Additionally, the measurement system has a variation of the sampling rate. Drifts of the oscillator usually cause these variations. In addition to a constant error, temperature changes can lead to variations, too. Therefore a error $\phi(t)$ occurs.

As mentioned non-periodic waveforms have advantages compared to periodic waveforms. Therefore, different waveforms which are commonly used in the NILM context are presented in the following sections.

Active and Reactive Power

The most common waveforms in NILM are active $P(t)$ and reactive power $Q(t)$. The power signals can be calculated out of the Root Mean Square (RMS) values of the current and voltage signal. For time-continuous signals, the RMS for current and voltage are:

$$i_{rms}(t) = \sqrt{\frac{1}{t_1 - t_0} \cdot \int_{t_0}^{t_1} i(t)^2 dt} \quad (2.4)$$

$$v_{rms}(t) = \sqrt{\frac{1}{t_1 - t_0} \cdot \int_{t_0}^{t_1} v(t)^2 dt} \quad (2.5)$$

where $t_1 - t_0$ is the time over one or multiple periods.

Equ. (2.6) describes the phase shift between voltage and current.

$$\Delta\varphi(t) = \varphi_v(t) - \varphi_i(t) \quad (2.6)$$

Active power (P) and reactive power (Q) can be calculated using the RMS values $i_{rms}(t)$ and $v_{rms}(t)$.

$$P(t) = v_{rms}(t) \cdot i_{rms}(t) \cdot \cos(\Delta\varphi(t)) \quad (2.7)$$

$$Q(t) = v_{rms}(t) \cdot i_{rms}(t) \cdot \sin(\Delta\varphi(t)) \quad (2.8)$$

Through the averaging by the integral in Equ. (2.4) and (2.5) high-frequency signal components are lost. Even if averaged over only one period, the information is low-pass filtered and highly compressed. The information, of the shape of the waveform, within one period of the current signal, is not contained in these power signals. This is unfavorable since this information can help to distinguish the different appliances during the classification process.

The calculation of the phase is especially for non-sinusoidal signals not a trivial task. The direct calculation of the phase can be avoided by calculating the apparent power S and the active power P . The reactive power Q can be calculated by Equ. (2.11).

$$S = \frac{1}{T} \sqrt{\int_{t_0}^{t_1} u^2(t) dt \cdot \int_{t_0}^{t_1} i^2(t) dt} \quad (2.9)$$

$$P = \frac{1}{T} \int_{t_0}^{t_1} u(t) \cdot i(t) \quad (2.10)$$

$$Q = \sqrt{S^2 - P^2} \quad (2.11)$$

High-frequency signal components are not considered. Nevertheless, the power waveforms are the most common waveforms in the NILM context. Active and reactive power have already

been used by Hart [62, 61] where a normalization method of the power metrics (2.12) is presented.

$$P_{norm} = P \cdot \left(\frac{120}{v} \right)^2 \quad (2.12)$$

v is the measured voltage of one phase with respect to neutral and 120 is the nominal voltage, typical in America. The advantage of this normalization is that voltage alterations, present on the power network, have less effect on the subsequent event detection.

An additional benefit of this waveform is the availability of numerous professional data loggers like the Fluke 1730 (a three-phase energy logger) on the market. This measuring system allows recording directly into active and reactive power. These data loggers are usually only capable of recording data in the range of seconds, even if the initial measurement has a significantly higher sampling frequency. Since positive results can be obtained with these waveforms, they are used very frequently [15, 40, 138, 41, 87, 21].

One of the most common, additional to P and Q used waveforms, are the HARs. Commercial measurement systems, which use internal a high sampling rate, are usually not able to log the measured voltage and current directly but often can log further calculated signal forms like the HARs. The Fluke 1730 energy logger, for example, can log active power, reactive power, and the HARs with a maximum logging frequency of 1 Hz.

Harmonics

In addition to active and reactive power the HARmonics (HARs) are popular waveforms in the NILM context [133, 79, 132]. Modern appliances have a nonlinear power consumption. These appliances use a switching power supplies which creates HAR frequencies [135]. These frequencies are multiples of the fundamental frequency (which is equal to the grid frequency) of 50 Hz or 60 Hz [9]. The signal of an ideal oscillator is given by

$$v(t) = \hat{v} \sin(2\pi ft + \Delta\varphi(t)) \quad (2.13)$$

with the amplitude \hat{v} , the frequency f and the phase-shift $\Delta\varphi(t)$.

In practice there are variations of the amplitude and the phase. The oscillation of a real oscillator is presented in [149] by

$$v(t) = (V + \epsilon(t)) \sin(2\pi ft + \Delta\varphi(t) + \phi(t)). \quad (2.14)$$

$\epsilon(t)$ represents the variation of the amplitude and $\phi(t)$ is the variation of the phase also called phase noise.

The power grid itself causes the phase noise. The frequency of the power grid cannot be regulated exactly constant because the frequency depends on appliances, as well as energy suppliers. Besides, an error arises due to the measurement hardware. As a rule, a constant sampling rate is assumed. However, due to different influences, such as the temperature, small variations of the sampling frequency can occur.

The HARs can be calculated using the Fourier Transform (FT)[112]. The complex FT is presented by

$$X(f) = \int_{-\infty}^{\infty} x(t) \cdot e^{-j2\pi ft} dt. \quad (2.15)$$

The Discrete Fourier Transform (DFT) is described as

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j \frac{2\pi nk}{N}} \quad k = 0, 1, 2, \dots, N-1 \quad (2.16)$$

$$(2.17)$$

Table 2.3: Values of individual HARs according to EN 50160: 2005 [29]

odd HARs				even HARs	
no multiples of 3		multiples of 3		order HAR	HARV in %
order HAR	HARV in %	order HAR	HARV in %		
5	6.0	3	5.0	2	2.0
7	5.0	9	1.5	4	1.0
11	3.5	15	0.5	6...24	0.5
13	3.0	21	0.5		
17	2.0				
19	1.5				
23	1.5				
25	1.5				

Since $e^{j\varphi} = \cos \varphi + j \sin \varphi$ can be written as $e^{-j\varphi} = \cos \varphi - j \sin \varphi$ applies

$$X[k] = \sum_{n=0}^{N-1} x[n] \left(\cos \left(\frac{2\pi nk}{N} \right) - j \sin \left(\frac{2\pi nk}{N} \right) \right) \quad (2.18)$$

For the transformation from the time domain into the frequency domain, different effective algorithms like the Fast Fourier Transform (FFT) [112, 71, 136] or the Goertzel-Algorithm [127, p. 480-481] can be used to improve the performance.

For the calculation of the HARs, the Goertzel-Algorithm may advantage because of the reduced calculation effort. In comparison to the FFT algorithm the Goertzel algorithm is faster when the number of spectral components which have to be calculated is below $5/6 \log_2(N)$, where N is the number of sampling points [127, p.481].

In NILM, HARs occur if an appliances has a nonlinear power consumption. Periodic waveforms can be described as the sum of several sinusoidal sequences, where the individual sinusoidal oscillations are multiples of the fundamental frequency. Symmetrical waveforms, in which positive and negative characteristics are identical, produce odd HARs. For regulatory reasons most appliances burdening the power line symmetrically. However, there are also some appliances which utilize a half-wave rectifier and, therefore, cause additionally odd HARs as these appliances only draw current in half of the period [28]. Unbalanced waveforms cause even and odd HARs.

In order to obtain a rough overview of the permissible device behavior in the power grid, some legal limits are shown. For ensuring the network quality, the government has set up some rules which the device manufacturers must comply. Tab. 2.3 shows the allowed HARmonic Voltages (HARV) according to EN 50160: 2005 [29].

“Under normal operating conditions, during each period of one week, 95 % of the 10 minute mean RMS values of each individual HAR voltage shall be less than or equal to the value given in Tab. 2.3. Resonances may cause higher voltages for an individual HAR.” [29]

The maximum allowed Total Harmonic Distortion (THD) of the supply voltage, calculated from all HARs up to 40th order, is 8 % [29].

Tab. 2.3 indicates that deviations from the sinusoidal shape are permissible under certain conditions. For this reason, the use of these signal forms for event detection or classification in the context of NILM seems to be advantageous. Under the law provisions, appliances which produce HAR can be sold and applied. For this reason, different works have investigated the use of HAR in NILM [28, 30, 31, 32, 35, 100, 140].

In the context of NILM, different assumptions are made when using HARs. Equ. (2.19) describes, how an arbitrary periodic signal, can be generated from sinusoidal oscillations, of a multiple of the fundamental frequency. The restriction of a steady state waveform, where both

negative and positive half-wave are point symmetric, is made [9].

$$i[k] = \sum_{n=1}^{\infty} \hat{x}[n] \cdot \cos\left(\frac{2\pi nk}{N}\right) \quad (2.19)$$

Where $i(k)$ is the discrete time domain function, n is the HAR number, \hat{x}_n the amplitude of the n^{th} HAR component and the window length N . Due to the limitation that no even HARs are considered, the simplification to Equ. (2.19) should only be used in case of appliances which have the same power consumption on the negative and positive half-wave.

In [51] a simple test set-up was realized that shows the detection of different devices on the HARs. In addition to classification, the HARs can be used also for event detection [77].

In [118] an algorithm is presented which detects the speed-related slot HARs in NILM using a notch filter, a trigger detector and a speed estimator. Therefore, this approach allows to examined particular complex machines with motors use different speeds drives.

The transformation from time into the frequency space requires a certain window-length including multiple samples of the current signal. On the one hand, the order of HARs which can be used depends on the sampling frequency. On the other hand, it depends on the window length which is used for calculation. The sampling frequency is limited due to the measurement hardware. The window length depends on the switching distance of the different appliances because the signal within one window length should not contain switching processes. A detailed explanation of the feature extraction will be presented in Sec. 2.2.4. Usually, the minimum distance between the switching events of the appliances is not known. Hence, the window length should be taken as short as possible to avoid the problem of detecting multiple states.

Therefore, the number of samples, for the calculation of the frequency spectrum, is limited. This limitation causes the problem that the frequency spectrum is blurred, because of the small window-length. From Tab. 2.3, it can be assumed that the energy of higher HARs is smaller, because of regulatory reasons. Thus, the quality of the higher HARs is worse. The order of HARs which can be used for classification is limited with the consequence of losing information.

HARs can only be calculated for analyzing steady states since, during a transient state, the signal is not necessarily of a periodical form. Therefore, the use of HAR is limited to the use of SSts.

Current Waveform (CW)

The CW extracts the current signal over one period [95, 74]. Since only the current signal is required for this waveform, the hardware complexity is reduced. This reduces the costs of the measurement system since the measurement and signal processing costs for the mains voltage is eliminated. When using a galvanic insulated current sensor, like sensors relying on the hall effect or Rogowski coils, no direct contact with the grid voltage is needed. Therefore, the installation costs are significantly lower because no change or additions like additional fuses are required. The use of CW requires a measurement system with a sample rate of several kHz to ensure that sufficient sampling points per period are available. However, a significant disadvantage is the loss of the phase information between current and voltage. In the literature, the CWs are often used in combination with other features [74, 126].

Trajectories

An additional waveform, getting more and more popular in the NILM context, is the use of trajectories [83, 38, 11]. The trajectories are generated directly from the current and voltage signal. After voltage and current have been measured over a period, a two dimensional fingerprint of each appliances can be created [43, 38, 11]. New features can be calculated out of the two-dimensional signal form [65]. The performance depends mainly on the feature extraction methods. Compared with other waveforms, the feature extraction is more complicated.

For illustration, four trajectories of a radio, a vacuum cleaner, a refrigerator, and a kettle are presented in Fig. 2.6 to 2.9.

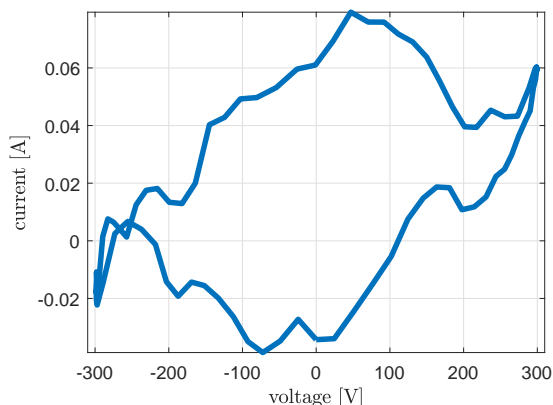


Figure 2.6: Radio (ID:003)

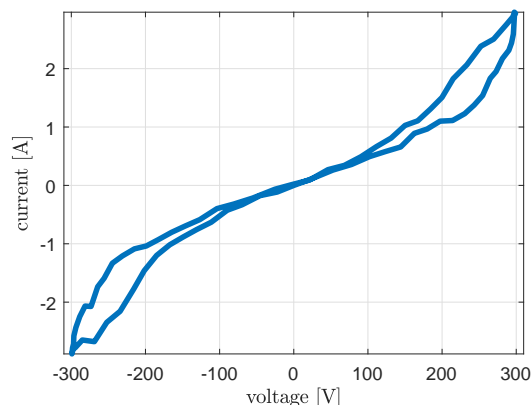


Figure 2.7: Vacuum Cleaner (ID:005)

Trajectories of two appliances of the HELD1 [67] dataset

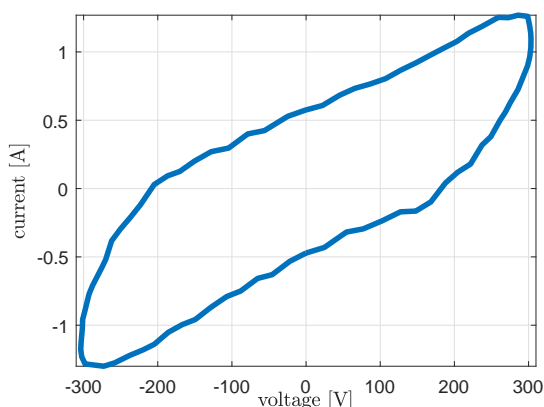


Figure 2.8: Refrigerator (ID:013)

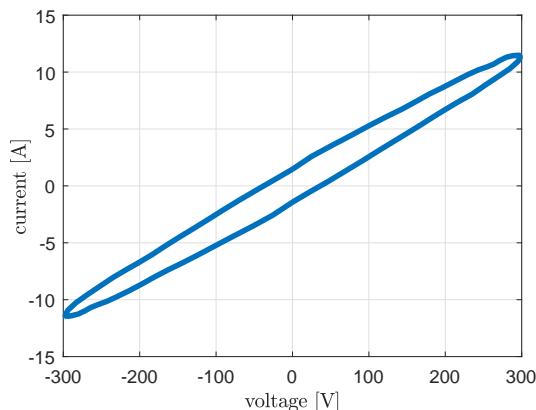


Figure 2.9: Kettle (ID:017)

Trajectories of two appliances of the HELD1 [67] dataset

Different features can interpret V-I trajectories. For example, different parts of the trajectories are analyzed [94]. The device properties are extracted from a graphical representation. For the classification of features generated out of trajectories, commonly classification methods from image processing are used. If V-I trajectories can be successfully used, for the classification, depends on the further feature extraction. The feature extraction of trajectories is comparatively complex, as well as the methods for the analysis of the features. The complexity of trajectories, compared to CW, is adverse. However, the additional use of voltage compared to CW allows using the phase information between current and voltage.

The additional use of the voltage, not only convert more information into the waveform, but it also adds noise. The amplitude of the effective value of the voltage is not constant. Variations of up to $\pm 10\%$ are possible. A more detailed discussion of the use of voltage information for NILM is provided in Sec. 5.2.

Further Time Domain Waveforms

Additionally, to the already described waveforms, more waveforms in the time domain exist. These, in NILM used, waveforms are instantaneous power waveforms (IPWs) and Instantaneous

Admittance Waveforms (IAWs) [140]. The time domain waveforms, commonly used in NILM, can be described as

$$CW(t) = i(t), \quad (2.20)$$

$$IPW(t) = i(t) \cdot v(t), \text{ and} \quad (2.21)$$

$$IAW(t) = \frac{i(t)}{v(t)}. \quad (2.22)$$

Where $i(t)$ is the instantaneous current and $v(t)$ the instantaneous voltage. While $CW(t)$ and $IPW(t)$ are commonly used in the NILM context, the use of $IAW(t)$ is problematic because of the division by zero, which causes numerical instability. The current waveform is superior to the HARs; thus, the distinction of individual devices can be improved, which leads to better results [140].

There exist some other waveforms which are not very frequently used in NILM context. A more common method is the combination of different waveforms.

Combination of Signal Forms

In context of NILM, no waveform containing all information and allow a simple feature extraction. For different appliances, different waveforms are superior. Therefore, several combinations of signal forms have been used in NILM to improve the results of the classification [44, 101, 102]. In [75], the use of multiple features, created from the trajectories, P , and Q , leads to better results in classification than individual features. Choosing the right combination of waveforms and the resulting features is difficult as the selection depends on numerous factors, such as the individual device types and the sampling rate. Furthermore, with changing appliances, the optimal combination of the selected waveform is subject to changes. Additionally, the weighting factors of the different used waveforms can vary. Due to the circumstance that several combinations exist, the optimization problem is quite complex. Furthermore, the computational effort needs to be considered when multiple waveforms have to be calculated in parallel.

The combination of different signal forms or the resulting features, can hardly be realized in a low-cost embedded system as the computation power for the calculation of different waveforms is too low. For the event detection, the total current and/or voltage signal needs to be analyzed. Using multiple waveforms may improve the results of specific event detection task, but the computational effort is very high. Important to note is that there is no feature combination which is suitable for all measurement scenarios. The achieved improvements via feature selection are commonly for one specific measurement scenario.

Summery

The investigating of the literature about waveforms in NILM show that many publications [10, 15, 21, 87, 91, 109, 128, 137, 153] use a similar approach, as the original idea from Hart et al. [63]. Due to their advantages primarily the signal forms of active and reactive power were used. These can be calculated internally in the measuring system and stored with a low resolution, for example 1 Hz. Furthermore, the use of power waveforms simplifies the computational effort for the subsequent processing steps of event detection, feature extraction, and classification. Due to the technical progress in the last 30 years, significantly higher-frequency signals can now be processed on more efficient hardware. Many studies use approaches to map the high-frequency signal components, that are not considered when relying on power, by adding for example HARs [28, 51, 77, 78, 79, 118, 132, 133, 135].

Since the use of active and reactive power, even in combination with the HARs, did not always lead to more satisfactory results, further signal forms were investigated. Different signal forms, like HAR, CW, IAW, and IPW, have been investigated in the NILM context [101, 102].

The selection of the signal shape is of particular importance for the following NILM process but especially for the feature extraction and classification. The aim is to keep as much as possible of the relevant information contained in the original current and voltage signal. At the same time, the information should be compressed so that correlated information and noise are minimized. Having a non-periodic signal is worthwhile for further processing.

The use multiple waveforms has become more widespread in recent years. However, this leads to disadvantages since different combinations of appliances require different signal forms or variable weights. The feature selection depends on the specific device setting. The calculation effort due to combined waveforms is high; thus, more expensive hardware will be required. Therefore, the focus of this work is to develop a suitable waveform for NILM.

2.2.3 Event Detection

The purpose of event detection is to determine the switching on and off positions of devices in the signal. In general, this corresponds to a state detection where the different SSts are distinguished.

A common definition of events in the context of NILM is:

“An event is a transition from one steady state to another steady state which definitely differs from the previous one.”[148]

Most common are the power signal forms $P(t)$, $Q(t)$, and $S(t)$. In most publications, only one of the waveforms is used.

There are different motivations for using event detection in NILM. The event detection reduces the computation effort for the feature extraction and classification. It is sufficient to determine these only at a few specific positions of the signal instead of calculating features over the entire signal. Thus, a classification has to be performed only for these features. Without event detection features have to be calculated over the entire signal. The classification has, in addition to the classes which are assigned to the individual devices, a class, representing features without the activity change of devices.

The waveform calculated from the current and voltage signal can be divided into different states. A Steady State (SSt) is a signal part without significant change of the amplitude. In these signal sections, devices are neither turned on nor off. The Transient State (TSt) denotes the section of a signal in which a device has switched on or off. In general, the amplitude of the waveform varies during a TSt. The different states are illustrated in Fig. 2.10a), using exemplary the active power as waveform. For real-world measurements a distinction is sometimes more difficult. Fig. 2.10b) is showing a real measured signals of a hair straightener. The power consumption of this device represents, especially well, a widespread behavior of various devices. After switching on, the power consumption increases to a maximum. Subsequently, the power consumption decreases exponentially. The position of the start of the first TSt can be identified relative exactly. The end position of the TSt, which is also the start of the second SSt, is not clear. The TSt of the switching of process, is for most appliances very short. This is the typical behavior of many appliances and particularly pronounced in the case of the hair straightener.

The goal of the event detection is to determine all existing or occurring events. In the case of NILM, disturbances are caused by the mains, the measuring instruments or the measuring system itself. The noise causes non-existent events to be detected. Therefore, the thresholds are increased, which leads to the fact, that under certain circumstances, appliances with a small consumption are no longer detected. On the other hand, not too much events have to be detected. Since limiting the number of unrecognized events and the number of too many recognized events are two opposing optimization goals, a trade-off must be found.

Events which are detected correctly are called True Positives (TPs). Not recognized events are called False Negatives (FNs) and falsely detected events are FPs. Fig.2.11 visualize the possible errors during an event detection.

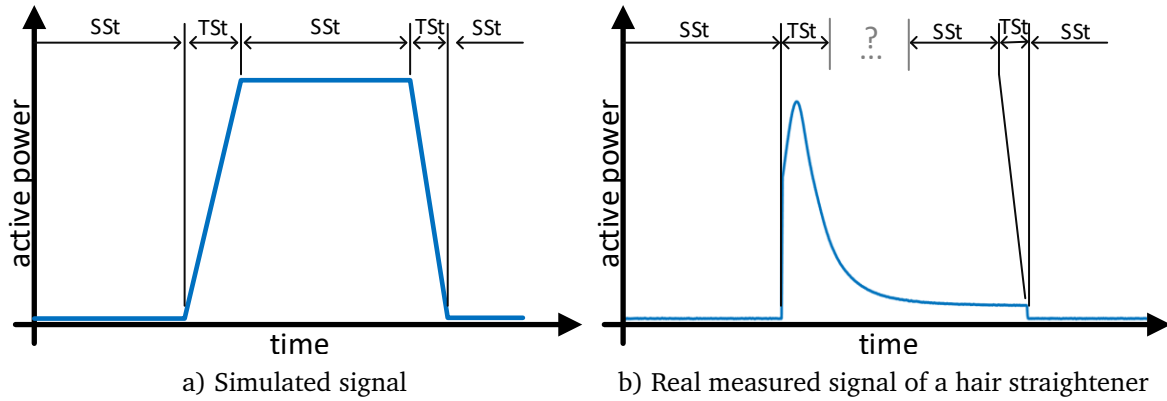


Figure 2.10: The different signal states (SSt and TSt) of NILM are shown on an active power signal

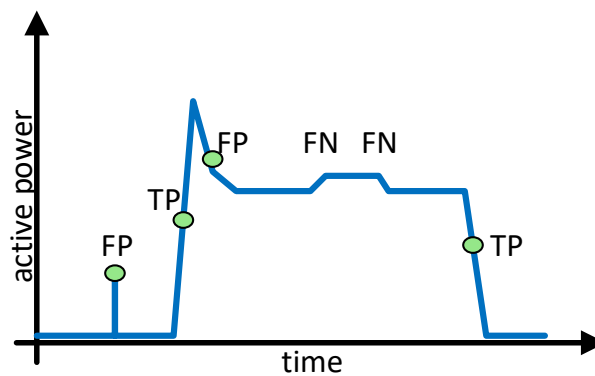


Figure 2.11: Example of possible results for the event detection; detected events are marked green.

In this case, four existing events are presented. The first marked event is a FP. The algorithm detected this event, for example, because of the occurring noise in the signal. The second detected event is a TP. Here, an existing event is correctly recognized. The next event is a special case of a FP; the existing event is recognized multiple times. The following two events are not recognized by the event detector which leads to two FNs. The last event is detected correctly and is a TP.

In order to provide a better comparability, the following performance metrics

$$\text{recall} = \frac{TP}{TP + FN} \quad (2.23)$$

$$\text{precision} = \frac{TP}{TP + FP}, \quad (2.24)$$

which depend on TP, FN and FP, are used.

Many event detection methods are based on low sampled signal forms [87, 90, 105, 141]. For example, an adaptive event detection method called Steady State Approximation (SSA), analyses the waveform in order to find the different SSts [87]. Consequently, between the different SSts, TSts or events must exist. This approach has the aim of reducing the number of FPs. Since this algorithm is used in AAL context where a high number of FPs is not acceptable. In the SSA approach, the active power is used as a waveform. An automatic adaption of the applied window length of the waveform is implemented. The advantage of this approach is that no training data is required and a real-time implementation is possible. Compared to other algorithms, the SSA produces much less FPs while the decrease of TPs is hardly significant. This approach uses active and reactive power at a sampling frequency of 50 Hz for the event

detection. The algorithm achieved a high accuracy for appliances with high power consumption. The number of FP increases significantly when small appliances should be detected too.

A further event detection method is based on a General Likelihood Ratio (GLR) algorithm [105]. In this statistical approach, higher accuracy is attempted by calculating different probabilities. High-frequency sampling rates (of several kHz) reduces the error probability since events, which are close to each other, can be distinguished using a higher sampling rate. Nevertheless, the sampling rate of maximum 8 Hz is comparatively low and no other waveforms were examined.

The main advantage of event detection methods, which work with low sampling rates, is the low computational effort. For example, in embedded hardware, this is an important advantage. However, the performance of such systems has been increased, in recent years. The three-phase 24 bit ADC from Analog devices, for example, “ADE7878AACPZ” has a sampling rate of 8 kHz and is available for 6.18 \$⁸⁾. Depending on the resolution and the sampling rate, there are several ADC on the market.

There is a higher chance to miss events, by using a waveform with a lower sampling rate because of the reduced information in the signal. For example, two events which are close together, can be mistakenly recognized as two events. Appliances with small power consumption, use often switching power supplies. This creates a non-sinusoidal power consumption. These devices use only a small part of the period for their power supply. A 50 Hz or less sampled waveform is averaged over the entire period. This adversely affects the detection. For this reason, event detection methods rely on the high sampled current [148]. To compare the performance with respect to other methods, the BLUED dataset [4] is used. The detection method is based on kernel Fisher discriminant analysis of the current HARs features. For the BLUED dataset, an accuracy of 98.16 % recall and 97.94 % precision for phase A, and 70.41 % recall and 87.29 % precision for phase B could be achieved [4].

The use of different datasets, for event detection in the context of NILM, causes difficulties when comparing the individual results. Results based on a low sampled datasets, cannot be compared with methods using higher sampling rates. Using limited signal sections of a dataset or defining different restrictions as for example a certain power level of the devices, also causes difficulties when comparing algorithms. Furthermore, the implementation of existing event detection methods is complicated since different parameters will affect the performance.

The event detection needs to handle the complete waveform. Therefore, the computational effort of the event detection should be as low as possible, especially when targeting the implementation in a low-cost embedded system.

Due to the linear structure of the different NILM tasks, the errors made in the individual processing steps will accumulate. Therefore, a high recognition rate of the event detection is favorable. In order to investigate, if a new waveform containing more information compared to existing waveforms is advantageous for the event detection, the task of event detection was investigated in detail during this work, Chap. 7.

2.2.4 Features & Feature Extraction

The information provided for classification is denoted as a feature. In literature, common expressions are fingerprints or signatures instead of feature [26]. The feature extraction can be implemented by Steady State (SSt) and Transient State (TSt) signal components. Thus, features can be distinguished in Steady State Feature (SSF) and Transient State Feature (TSF) [61]. SSFs are more commonly used in NILM since good results, especially with a limited number of devices and even at a low sampling rate, can be achieved. Compared to TSF, SSF leads to less variation. TSF requires a higher sampling rate of several kHz because the time of the TSt of most applications is only a few seconds.

⁸⁾14.01.2019 Analog Devices: ADE7878AACPZ <https://www.analog.com/en/products/ade7878a.html#product-discussions>

In practice, several devices are running at the same time. Therefore, the difference of the waveform, before and after the event, is calculated. By using the difference of the waveform, one feature for each device is calculated. Without using the difference, a feature for every combination of devices running at the same time exist. The number of possible combination can be calculated as

$$N = \sum_{k=1}^n \binom{n}{k}, \quad (2.25)$$

where n is the number of applications and N is the number of possible combinations of devices which are used simultaneously. For the example, $n = 4$ devices, there are $N = 15$ possible combinations of devices running at the same time.

$$N = \sum_{k=1}^4 \binom{4}{k} \quad (2.26)$$

$$N = \frac{4!}{1! \cdot 3!} + \frac{4!}{2! \cdot 2!} + \frac{4!}{3! \cdot 1!} + \frac{4!}{4! \cdot 0!} \quad (2.27)$$

$$N = 4 + 6 + 4 + 1 = 15 \quad (2.28)$$

In a household usually more than 20 devices have to be distinguished. Hence, 1,048,575 possible combinations of devices running at the same time need to be considered. The probability of overlapping for such a high number of clusters in the feature space is very likely. Therefore, it is better to subtract the offset, which exists due to already active devices, by calculating the difference of the signal from before and after an occurring event. Thus, only one clusters for each appliance exist in the feature space; therefore the probability of overlapping is less.

In Tab. 2.4, an overview of different features for the use in NILM is given. A first approach of SSF in NILM was published by Hart [61, 62]. A rectangular approximation of the SSF power consumption was presented to calculate the power consumption of the individual devices. For classification, the difference of active and reactive power of two SSts was used at a sample rate of 1 Hz.

In 1995, Leeb et al. [98] used the TSF for event detection and to classify these switching events in NILM. One year later, Norford and Leeb published an algorithm which uses a combination of SSF and TSF for the detection of Heating, Ventilation and Air Conditioning (HVAC) equipment [117].

First feasible procedures were described by Pahiala [125] and Marceau [108]. A steady state approach for device recognition was presented, allowing the classification of a water heater, an electric room heater, and a refrigerator.

In [36], a method is presented for the recognition of complex loads like heat pumps, dishwashers or refrigerators. The turn-on process is more difficult to analyze than the turn off event. Turn off events have usually a fast decrease in power; thus, the distinction between SSt and TSt is less error-prone compared to on events. In [33], a combination of SSF and TSF is proposed.

With an increasing number of devices, the likelihood that active and reactive power will be unsatisfactory, due to a similar power consumption of some devices, increases. In order to be able to distinguish a larger number of devices, the HARs can be included for the classification [35] and [96]. For this purpose, a sampling rate of 1.8 kHz and 8 kHz is used.

A further significant increase in the sampling rate was investigated by Gupta [58], where a sample rate of 1 MHz is used. Many modern loads have a switching power supply to achieve high efficiency. As these devices generate EMI during operation, these inferences are used for device recognition. For evaluation, the frequency band between 36 kHz and 500 kHz was used. During the investigations, a correlation between the different phases could be recognized. Guzel and Ustunel also used EMI for the classification of 14 different devices [59]. A principal null space analysis was used for classification which achieved an accuracy of 79.9%. However, the very high effort due to the sampling rate in the MHz range is the reason for the use of much lower frequencies (of several kHz) in succeeding publications.

In [78], Jonetzko et al. presented a measurement system for NILM which uses a sample frequency of 4 kHz. The work examined whether the voltage measurement is necessary or if information of the current is sufficient for NILM. However, due to the missing phase information, the results of the measurements without voltage are worse.

Another method is presented by Du et al. [43]. Here, features are extracted from trajectories which are mapped to a binary cell grid. Whether the individual devices of the data set are used exclusively individually or also operated simultaneously is not described.

Summary

In the context of NILM, the generated features are the essential part for a successful classification. During the years several features have been investigated; most commonly used features are the SSF. The principle of the generation of these SSF has not changed over the years. The difference before and after an event of a non-periodic waveform is determined. By calculating the mean value over several samples before and after the event, a noise reduction is realized.

The commonly used features determined out of active power, reactive power, and HARs do not represent all information in an appropriate manner. Therefore, different waveforms like trajectories have been used to include the information of the signal form into the feature. By using features calculated from trajectories, different difficulties may occur during the classification. The use of trajectories is problematic when several devices are running at the same time since larger devices can influence the grid voltage.

Currently, there is no feature or even a combination of features which are superior with respect to the others. For each dataset or new combination of appliances, another feature or combination of features needs to be calculated to achieve the best results. Most features are calculated directly out of the used waveforms. The calculation effort of calculating several different waveforms is high. In case the waveform is only required for the feature extraction and not for the event detection, the effort could be reduced since the waveform needs only be calculated before and after an event. Nevertheless, the calculation effort is increasing with each feature, requiring an additional waveform.

A classification procedure needs to handle the different features. Here, the same problems as in the case of the different waveforms occur. A feature selection is only possible for a known collection of devices. Additional, influences like variations of the mains frequency can influence the best feature selection. For these reasons a new signal representation is needed, which contains as much as possible information for classification in NILM. New features have to be developed from the new signal form. An investigation of these features can be realized, by applying them to different classification methods and by comparison with existing features.

2.2.5 Classification

Supervised classification uses training data with reference data to train the classifier. For using supervised classification, several training data are required for each device. On the one hand, the class structures made available in the training should be learned as precisely as possible. On the other hand “learning by heart” needs to be avoided since the classification method needs to learn the general structure of the classification problem and not merely the individual training examples, which is also called overfitting. In this case, the classifier can assign the training examples correctly, but no further data. The classification method needs to generalize from the training examples to unseen test data. There are several ways to ensure a good generalization ability. Depending on the classification method used, different methods exist to improve the generalization. Another option that has been used in this work is the splitting of training data into training and validation data. After learning the classifier with the training data, the validation data are used to verify whether the classifier has actually learned the pattern or over-fits the individual training examples.

Table 2.4: Comparison of NILM feature extraction approaches, modified from [87]

author	year	references	sample rate	appliances	feature & remarks
Hart	1989	[62]	1 Hz		steady state (P, Q)
Leeb	1995	[98]			transient time series
Norford	1996	[117]		single	steady state (P, Q, HARs); only HVAC
Pihala	1998	[125]	1 Hz		steady state (P, Q)
Marceau	2000	[108]	16 s	3	steady state (P)
Cole	2000	[35]	≥ 1.8 kHz	10	steady state (current) HARs
Laughman	2003	[96]	≥ 8 kHz	single	steady state ($P, Q, \text{current HARs}$)
Wichakool	2007	[147]	≥ 1.5 kHz	3	steady state (HARs); only VSDs
Chang	2008	[33, 34]	≈ 15 kHz	3	transient energy
Gupta	2010	[58]	≥ 1 MHz	94	steady state features (EMI)
Guzel and Ustunel	2015	[59]	2 MHz	14	steady state features (EMI)
Jonetzko et al.	2015	[78]	4 kHz	9	steady state features (voltage, current)
Du et al.	2016	[43]	30.72 kHz	75	steady state features (V-I trajectories)

One advantage of supervised classification is the possibility to distinguishing between similar devices since the different classification methods have the chance to learn the differences of similar data. A disadvantage of supervised classification is the difficulty of generating training data. In the NILM context the training data needs to be labeled with the event position, the information about switching on or off event, and the device name or device ID. Especially in real-world applications, with several appliances, this is a difficult task. Industrial machines often cannot be turned on and off, as desired for the training of the classifier.

To reduce the problem of generating training data, Florian Liebgott and Bin Young presented a method to reduce the number of necessary training data without significant loss of performance. The results depend on the selected data for training. By clustering the unlabeled data and selecting the sample close to the centroid, good results could be achieved [103].

Another approach is to use sub-meter to generate training data [109]. The advantage of this supervised method is that only low-frequency reference data are required (1/3 Hz of the REDD dataset [89]). Additionally, variable loads are supported. However, the synchronization of the different training measurements is difficult. Because of the lower sampling rate of the reference measurement, the position of an occurring event is not defined precisely. Thus, an allocation to the individual appliances is difficult, if multiple appliances are switching close to each other.

For applications where only a limited number of devices needs to be detected, supervised learning can be used. In [56], a method to detect a refrigerator out of the total power consumption is presented. Furthermore, Thomas Bier presented in his doctoral thesis a method to classify a refrigerator [21].

In [10], up to nine devices are classified by a supervised method. The presented results show that automatically running appliances like refrigerators, freezer, and dishwasher can be classified with higher accuracy compared to manually switched devices, for example, lamps, TVs, or stereos.

Since supervised learning algorithms usually need a large set of labeled training data, Florian Liebgott and Bin Yang use active learning for a supervised classification method to compensate this disadvantages [103]. A reduction of up to 90% of necessary training data could be achieved.

Unsupervised Classification

Unsupervised classification does not use training information about individual devices. By clustering, the features are analyzed to find similarities in order to cluster the features.

Fully unsupervised methods have the aim to find a cluster for each appliance or type of appliances [24]. An assignment to the devices or their corresponding names is not required necessarily. In some approaches, appliances of the same type are clustered together.

In [111], a cloud-based disaggregation procedure is presented. The method does not attempt to generalize information of device classes of other databases. For device types with high variability, this is a difficult task. However, depending on the specific application, devices of the same type should be distinguished or grouped.

In some application, not all devices need to be clustered. In [116], for example, an unsupervised method to extract the charging loads of electric vehicles is presented.

Semi-Supervised Classification

Semi-supervised classification can be seen as a method between supervised and unsupervised learning. To generate training sequences of each device is commonly difficult. Finding the correct device, for each cluster, like in the case of unsupervised learning, is also difficult. Several training sequences for several appliances exist. The problem of unlabeled appliances is also considered. Thereby a higher accuracy, compared to only supervised methods, can be achieved [56]. Another advantage of semi-supervised classification is the reduction of labeled data that are required for the training [14]. Semi-supervised classification can be used, to learn with a generalized set of appliances types in order to adapt, for example, device types of an unknown household [76]. The advantage of this is that the complex training is not necessary for each household. However, results obtained by semi-supervised classification are worse than the results of a classification process in which all devices are trained explicitly. In [99], an Expectation Maximization (EM)-based semi-supervised classifier is presented and compared to a multi-label algorithm [152].

Summary

The classification problem in NILM does not differ from other common classification problems. Depending on the application of NILM, a supervised, unsupervised, or a semi-supervised classification is required. Regardless of the classification methods, additional reference data about the device is required in the data records of the classification. This is necessary for supervised, unsupervised, or a semi-supervised classification since even a completely unsupervised classifier needs to be tested and evaluated. Because of the different applications, a comparison of different classification methods is often challenging. The individual classification methods, used during this work, are described in Chap. 9.

Chapter 3 | Objectives and Hypotheses

The questions investigated in this thesis are based on the state of the art and are listed in this chapter. For each question, a hypothesis is set which is examined during this work.

In Non-Intrusive Load Monitoring (NILM) commonly standard waveforms (P, Q, and HARMONICS (HARs)) are used. A variety of methods in the field of event detection, classification, and tracking were applied to these signal forms or their combination. The signal representation is an essential part of NILM since all other processing steps are based on it. So far, no signal representation exclusively suitable for NILM has been developed. Rather popular is the combination of different signal forms or their features. This work aims to generate a signal representation from the current and voltage signal which is especially suitable for NILM. Since the signal representation has influences on all tasks of NILM it has to be applied to the whole NILM process to examine the advantages and disadvantages of such a signal representation.

1. Is it possible to develop a specific waveform for NILM that combines the relevant information of the current and voltage signal?

Hypotheses 1: Transforming the current signal according to the phase shift between current and voltage produces a NILM specific waveform that contains the relevant information of the current and voltage signal.

2. Is the new signal representation advantageous compared to existing signal forms?

Hypotheses 2: The new signal representation contains more information than the previously used signal forms and achieves better results in event detection.

3. What influence does the sampling rate have on event recognition?

Hypotheses 3: Waveforms with higher resolution contain more information, because of this, they can produce better results using a suitable event detection technique.

4. Is it possible to generate specific features for the new signal representation?

Hypotheses 4: The new signal representation which is developed primarily for NILM allows generating features with simple procedures.

5. Can the new signal representation provide better classification results than standard waveforms or combinations of different waveforms/features?

Hypotheses 5: The new signal representation achieves better classification rates compared to standard waveforms.

Chapter 4 | Home Equipment Laboratory Dataset

For the development and research of Non-Intrusive Load Monitoring (NILM) algorithms, datasets are required for training and testing. In Sec. 2.2.1 different existing public datasets, presenting the state of the art, have been presented. During this work, it turned out that the existing data records were not sufficient, therefore a new dataset was recorded. Parts of this chapter have been published in [67].

This chapter is structured as follows: first, an introduction into the topic of datasets is given followed by Sec. 4.2 which describes the measurement system developed by Thomas Bier [22]. The different measurement scenarios are described in Sec. 4.3 and the power distribution of the different appliances are described in Sec. 4.4.

4.1 Introduction

Existing datasets can be separated into two categories: real-world datasets and laboratory datasets. Real-world datasets have the disadvantage of different, frequently encountering problems concerning the reference or ground truth data. Laboratory datasets only provide the measurement during the switching on or off process, but the reference data is usually very reliable. On the other hand, there are no public laboratory measurements which include simultaneously active appliances.

The individual tasks of NILM lay down different demands on the data sets. Compared to the event detection, where only the information is required when or at which position an event occurs, the classification additionally requires the information which appliances were switched on or off. Furthermore, each device requires a minimum number of switching on and off cycles. This requirement becomes especially important when using supervised procedures. The proposed dataset is divided into training, validation, and test data. Compared to other classification problems like image processing or recognition of handwriting, where several thousands of training and test data is available, the available number of training or test data is much lower for NILM.

In real-world measurement scenarios with a power meter installed for a certain time, the problem occurs that some devices, like a refrigerator, have many switching cycles, whereas a vacuum cleaner is used very rarely. For datasets only recorded over a few days, this represents a major problem. In the Building-Level fully-labeled dataset for Electricity Disaggregation (BLUED) dataset [4], for example, some appliances were not used at all during the entire measurement campaign. The UK-Dale dataset solves this problem since its measurements last for more than four years. The problem of this dataset is the considerable amount of generated data. The dataset is generated using a sampling frequency of 16 kHz and a resolution of 24 bit. Thus, about 8.3 GB per day will be created and accordingly 3.03 TB in one year. Handling this amount of data is difficult and time-consuming. Besides, the reference data has a low sampling frequency, so the exact position of an event remains unknown. Because of this, a feature extraction and subsequent classification with this dataset is difficult.

Laboratory measurements of individual devices have the advantage that significantly more events can be generated in a shorter time. Additionally, the reference data are usually more reliable, since error sources can be excluded more easily. Due to the fact, that only one device

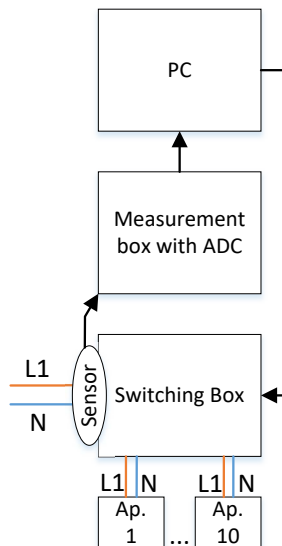


Figure 4.1: Block diagram of the measurement system; Analog-to-Digital Converter (ADC), appliance (Ap.), phase 1 (L1), neutral line (N), Personal Computer (PC) [67]

can be active at the same time, laboratory measurements depict reality only to a minimal extent.

For this reason, the Home Equipment Laboratory Dataset 1 (HELD1) [67] was created during this work. The measurement box [22] created by Thomas Bier during his PhD [21] was used. Compared to the old measurement campaign where only the active and reactive power have been stored, in the new measurement campaign, the measured current and voltage have been stored directly with the complete sample rate of 4 kHz.

The goal was to create a dataset primarily for the classification task of NILM. Particular attention is paid to the reference data. In order to ensure that every event is properly marked, only simple devices with one state were used. To generate measurements, reflecting problems of real scenarios, up to six devices were operated in parallel during a single measurement. The HELD1 dataset [67] is the first public available laboratory dataset containing individual measurements of devices for training, as well as measurements including several active devices at the same time.

4.2 Measurement System

The measuring system of Thomas Bier [21] was used to create the dataset. The used measuring system shown in Fig. 4.1 is described in detail in [22]. The sampling rate for current and voltage was increased from 1 kHz to 4 kHz. To avoid aliasing, a low-pass filter with a cutoff frequency of 1.3 kHz was applied before sampling. The analog-digital converter operates with 16 bits which corresponds to a theoretical resolution of $\sqrt{2} \cdot 63 \text{ A}/2^{16} = 0.961 \text{ mA}$ being equal to 0.61 W. Actually, current and voltage measurements have a noise of $\approx 16 \text{ mA}$ and $\approx \pm 2 \text{ V}$ for short-circuited inputs. This corresponds to $\pm 8 \text{ mA} \cdot 230 \text{ V} = \pm 1.84 \text{ W}$. In the HELD1 dataset [67], 18 different appliances are used. Their power consumptions are shown in Tab. 4.1. The selected appliances are simple appliances without any major internal load switching. Appliances which did not work reliably during the measurement campaign have been excluded from the dataset. For example, a toaster broke during the measurements. As a result, all measurements in which the toaster was involved, were not used for the published version. Thus, it was ensured that the same devices could be used for the different scenarios and the numbering of the device ID in Tab. 4.1 is not continuous.

Fig. 4.2 shows a selection of the different appliances used for the dataset.

Table 4.1: Device list of the HELD1 dataset [67]

ID	Name	P(W)
001	Toaster	998
002	Hairdryer (setting 2)	1155
003	Radio	6.2
005	Vacuum Cleaner (red)	424
007	Hair Straightener	56
009	Heat Gun (setting 1)	820
010	Router	9.2
011	Desk Lamp	20
013	Refrigerator (white)	170
014	Refrigerator (blue)	190
015	Fluorescent Lamp	40
016	Light Bulb Box	20
017	Kettle	2100
019	Hairdryer (setting 1)	500
020	Heat Gun (setting 2)	1603
021	Fan	22
022	Multifunction Tool (Dremel [®])	30
023	LED Lamp	1



Figure 4.2: Illustration of random selected individual appliances with corresponding device ID [67]

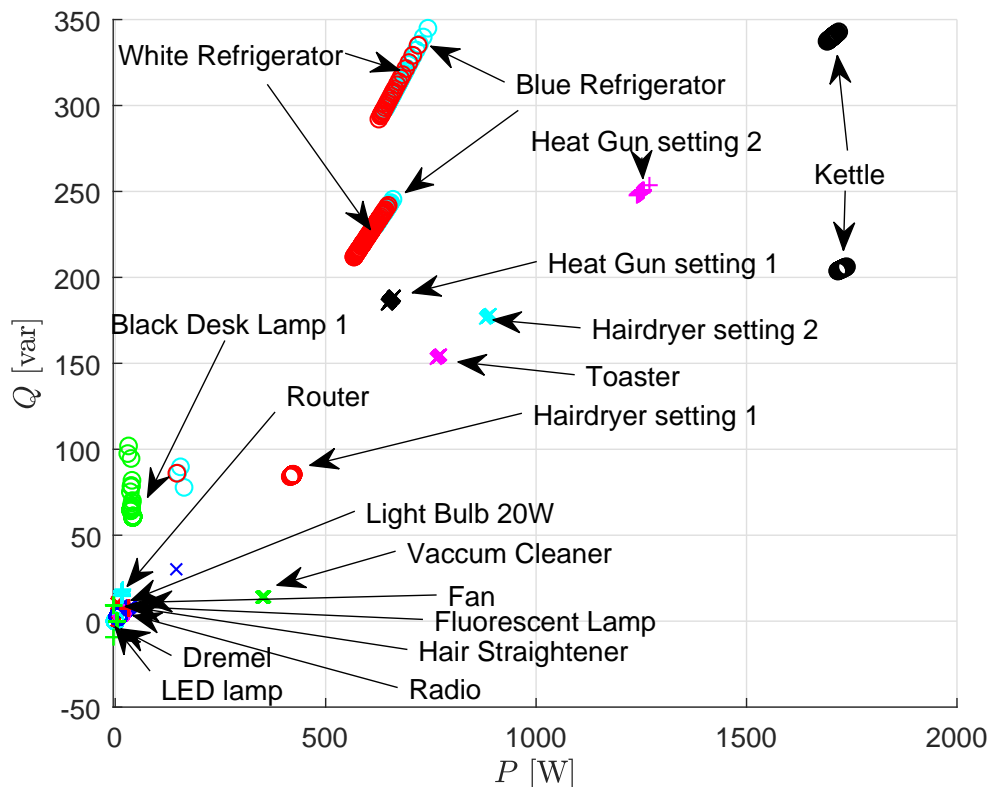


Figure 4.3: Active and reactive power distribution of the appliances [67]

4.3 Measurement Scenarios

Different complex measurements have been performed by varying the number of simultaneously active devices. The dataset is divided into four consecutive measurements [67]:

- Training data (consisting of individual measurements, 100 on/off events per device)
- Test measurement one (one device being active at the same time)
- Test measurement two (up to four active devices being active at the same time)
- Test measurement three (up to six devices being active at the same time)

In order to facilitate the training process of unsupervised learning, 100 on and 100 off operations from each individual device have been recorded. Overall, the dataset provides 3,530 training events and 13,200 test events.

During test measurements, the devices were randomly switched on and off under consideration of the corresponding scenario. For all scenarios, the minimum time distance between two events are three seconds.

4.4 Power Distribution of the Appliances

Fig. 4.3 shows the different power distributions of the used appliances. A few individual consumers (for example a refrigerator or an electric kettle) have two different clusters of power distribution. There are different reasons for this behavior. Considering the electric kettle, one reason is the difference in power consumption between the switching on and the switching off steady state. This is caused by the influence of the water temperature which has a slight impact on the power consumption of the electric kettle. Therefore, the power consumption immediately

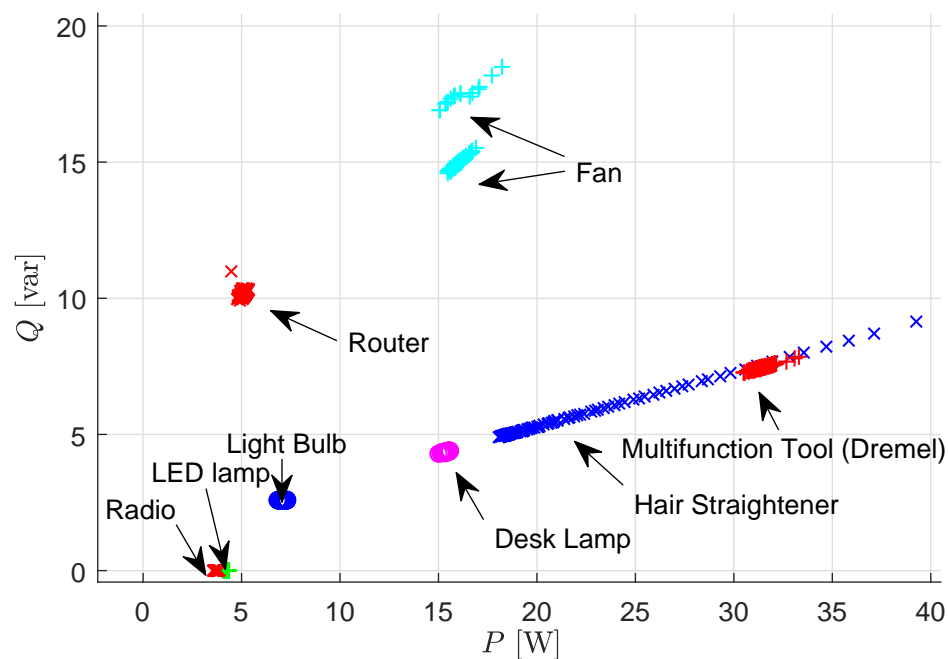


Figure 4.4: Detailed illustration of the lower power range from Fig. 4.3 [67]

after starting the electric kettle and when the water is cold, is higher compared to the power consumption when the water is hot.

In the case of the refrigerator sometimes only the pump turned on but not the whole cooling unit. This also results in the formation of two clusters at different positions. Fig. 4.4 shows the clusters of appliances with lower power consumption in more detail. While it is difficult to distinguishing appliances with a small consumption in Fig. 4.3, they can be separated very well from each other in the detailed analysis shown in Fig. 4.4. Only the clusters of the radio and the LED lamp, as well as, the hair straightener and the multifunctional tool, are very close together. Compared to the other appliances, the hair straightener exhibits the most significant drift of the individual training features. The Transient State (TSt) of the switching on-process of the hair straightener was already presented in Fig. 2.10b), since it has an extremely long transient during the switching on-process. Because of the exponential signal, the position in the signal where the TSt ends and the Steady State (SSt) starts, is not clearly defined. As a result, the calculated active and reactive power of the hair straightener will vary considerably more than those of other appliances.

Fig. 4.3 and Fig. 4.4 show the results of the training data record. In this scenario, only one device is active at the same time. If several devices are active simultaneously, the overall noise level is increased. This, however, results in larger clusters of the individual appliances, leading to a higher probability of overlapping. The test datasets two and three used up to four and up to six appliances running simultaneously.

Most appliances can already be separated by P and Q (see Fig. 4.3 and Fig. 4.4). However, some clusters show overlapping data, for example, the clusters of the hair straightener, the multifunctional tool (Dremel[®]), radio, LED lamp, or the two refrigerators. The idea is that the distinctiveness of these devices can be improved by adding the HARMONICS (HARs) to the feature space of P and Q .

Table 4.2: Comparison of the HELD1 dataset [67] with commonly used NILM datasets

dataset	samplerate aggregate measurement	samplerate ground truth	duration	single / multiple devices / laboratory data set
REDD [89]	1sec / 15kHz	3sec	3 to 19 days	-/yes/no
BLUED [4]	12kHz	event based	8 days	-/yes/no
PLAID [54]	30kHz	-	5sec	yes/no/yes
WHITED [80]	44kHz	-	5sec	yes/no/yes
COOLL [124]	100kHz	-	6sec	yes/no/yes
UK-DALE [82]	16kHz	6sec	3-17 months	-/yes/no
ECO [20]	1sec	1sec	8 months	-/yes/no
SustData [123]	50Hz	50Hz	5 years	-/yes/no
HELD1 [67]	4kHz	event based	scenario	yes/yes/yes

4.5 Conclusion

The HELD1⁹⁾ is the first public available laboratory dataset for NILM which contains continuous measurements with simultaneously active devices. For more comfortable handling, separate training sequences for each device have been created. For each appliance, there are 100 on and off events available for training. The clearly defined training scenario enables an improved comparison of the individual methods applied to the dataset.

First classification results have been published in order to provide reference data [67]. A sampling rate of 4 kHz is used for the voltage and current measurement. The reference data is provided similar to the BLUED dataset. The index of the individual event is provided in the reference data including the information of the device ID and the type of the event (switching on or off).

⁹⁾Home Equipment Laboratory Dataset 1 (HELD1) dataset online available at <http://141.28.27.61/>.

Chapter 5 | **Waveform**

A waveform is calculated from the originally measured current and voltage signal. These waveforms are easier to handle for the following Non-Intrusive Load Monitoring (NILM) tasks. In general, these waveforms are non-periodic. Additionally, the converted waveform contains information from voltage and current in only one signal; therefore, the further processing is easier. The aim is to develop a waveform for the requirements of NILM. The new waveform should contain as much information as possible that is relevant for the event detection and the classification. Since the calculation effort of the waveform is high, both tasks (event detection and classification respectively feature extraction) should be manageable with one waveform.

This chapter describes the NILM part of generating a waveform from the original sampled and recorded current and voltage signals. After an introduction, an investigation of the influence of voltage fluctuations of the grid follows. The focus of this chapter is a new signal representation which is tailored primarily for NILM. Different sampling frequencies have been investigated on the new waveform. The chapter concludes with the results and a discussion. Parts of this chapter have already been published during this work [66, 68, 69].

5.1 Introduction

One of the most critical tasks, of the NILM process chain, is to convert the measured current and voltage signal into a non-periodic waveform. Non-periodic waveforms are commonly used since they are easier to handle for event detection. A first segmentation can already be achieved by applying a threshold. The underlying waveform is crucial for feature extraction and the following classification. The most common features in NILM is the Steady State Feature (SSF). The feature is calculated by subtracting the difference of the waveform, before and after an event. In the case of multiple active appliances at the same time, the noise level in the signal is increased as the power consumption of the individual appliances varies. When using a non-periodic waveform, white noise can be reduced by calculating the mean value over a specified window length of the waveform. Furthermore, the computational effort should be kept low such that the conversion can be realized with low-cost and low-performance hardware.

Most waveforms used up to now, carried out a substantial compression of the information. For example, for the calculation of active and reactive power, which is described in Sec. 2.2.2, the Root Mean Square (RMS) of the values (see Equ. (2.4) and Equ. (2.5)), an averaging over at least one period is carried out. Therefore, white noise is suppressed, which is advantageous. After calculating the mean value a downsampling is usually carried out which reduces the computational effort for all subsequent tasks because of the reduction of sampling points. The disadvantage of this averaging and downsampling is that some information is lost. To circumvent the possible information loss, additional waveforms are used. The requirement for the waveform is to contain as much information as relevant for successfully applying NILM.

Existing waveforms presented in Sec. 2.2.2, have not been designed with the explicit goal of event detection or classification for NILM. Nevertheless, in the NILM context, these waveforms are predominantly used. In order to improve the results of classification, different waveforms are combined. The use of multiple features, created out of numerous waveforms, leads to better results compared with the use of one single waveform. This fact shows the limits of the currently

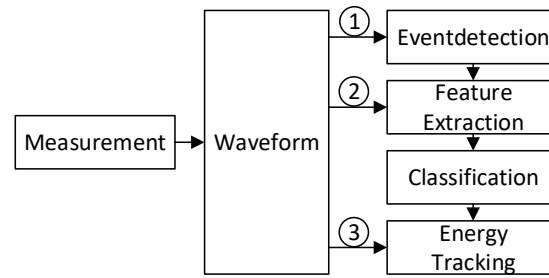


Figure 5.1: In comparison to the usual linear presentation of the processing chain of the NILM tasks, see Fig. 2.1, the information flow is not ordered linearly.

used waveforms.

The logical processing chain of NILM differs from the flow of information. Fig. 5.1 presents the flow of information in NILM. Compared to Fig. 2.1, where the single NILM tasks are ordered as a series of operations, the extracted waveforms are used for each further NILM task, parallel. The generated waveform is used for the event detection, the feature extraction, and the tracking. Since the waveform is used to calculate the features for the classification, the waveform is of particular importance for the classification. A combination of waveforms, as described in Chap. 2, requires high computational effort since usually for each feature a different waveform needs to be calculated.

In principle, a individual waveform ① to ③ can be generated for each subsequent task. Additionally, the disadvantage is the resulting calculation effort. The measured voltage and current are sampled usually with a high sampling rate to measured also the frequency content of an appliance. The task of converting waveforms is usually computationally expensive. Therefore, a waveform that can be used for succeeding tasks is highly preferable.

The initial current and voltage signal includes all available information. The trajectory waveform (described in Sec. 2.2.2) utilize already all the given information. Because of the two dimensional structure, the waveform is compared to other waveforms more complicated when trying to extract the features [140]. Additionally, the variation of the voltage amplitude causes problems with the reproducibility. On the one hand, the calculated waveform must contain as much information as possible, on the other hand, the relevant information must be suited for further processing. Unneeded information and noise should be suppressed as much as possible to minimize the required succeeding processing power. The primary information is contained in the current signal as the voltage signal allows mainly the calculation of the phase shift between current and voltage. The influence of the amplitude of the voltage signal is discussed in the next section.

5.2 Influence of the Amplitude of the Voltage Signal in NILM

The amplitude of the grid voltage, as well as the tolerances, vary from country to country. In Europe, the voltage of the power grid varies between 207 V and 253 V ($230 \text{ V} \pm 23 \text{ V}$) according to EN 50160:2005 [29]. Depending on the load, these voltage fluctuations in the grid affect the power or current consumption of the individual appliances. Therefore, a closer look at the influence of the voltage, respectively the voltage variations seems to be required.

Hart [61] has already mentioned the voltage fluctuations in the power grid as a problem for NILM. The problem was solved by normalizing the power (see Equ. (2.12)). In the case of passive loads including resistive, inductive or capacitive components, the variance of the normalized power in a device is significantly lower. The situation is different for devices that contain a switching power supply. Without taking the slightly modified efficiency of these power supplies into account, the voltage variation does not lead to a change in the power consumption as a higher current consumption is the result; thus, the power remains the same.

Today, the use of switching power supplies is much more popular than a few decades ago due to the need of higher efficiency and better semiconductor components such as Metal-Oxide-Semiconductor Field-Effect Transistors (MOSFETs). The main advantages of switching power supplies compared to conventional Alternating Current (AC)/Direct Current (DC) converters are, a higher efficiency especially for appliances with smaller power consumption, lower size and weight because of smaller transformers, a lower standby consumption, and lower costs. Devices with a switching power supply replaced devices using a conventional AC/DC converters. For example, light bulbs are replaced by Light-Emitting Diodes (LEDs) with a switching power supply. Other appliances like Personal Computers (PCs), laptops, or mobile devices of all types are added due to the technical progress. Today, switching power supplies are installed in many devices and these should also be considered for NILM. The differences, between a conventional power supply and a switching power supply, are analyzed in detail in the following subsections.

5.2.1 Conventional Power Supplies

The relationships between load, voltage, and current for passive appliances or appliances using a conventional power supply can be described as in Equ. (5.1) and (5.2). Therefore, a sinusoidal waveform of voltage and current is assumed. $P(t)$ is the active power, $v_{rms}(t)$ and $i_{rms}(t)$ are the RMS values of $v(t)$ Equ. (2.2) and $i(t)$ Equ. (2.3) and Z is the impedance of the appliance.

$$P(t) = i_{rms}(t) \cdot v_{rms}(t) \cdot \cos(\varphi) \quad (5.1)$$

$$i_{rms}(t) = \frac{v_{rms}(t)}{Z} \quad (5.2)$$

The power consumption depends on the actual voltage $v_{rms}(t)$. Z is constant, thus, $P(t)$ and $i_{rms}(t)$ are proportional to the voltage $v_{rms}(t)$.

$$P(t) = \frac{v_{rms}^2(t)}{Z} \quad (5.3)$$

5.2.2 Switching Power Supplies

Equ. (5.4) describes the relationship between voltage, current, and time-variable load for switching appliances. The equation for the power consumption is equivalent to Equ. (5.1). The difference compared to passive appliances is that $P(t)$ is independent from voltage variations. This is realized by adjusting the load. $Z(t)$ is the complex time-dependent impedance of the switching device.

$$i_{rms}(t) = \frac{v_{rms}(t)}{Z(t)} \quad (5.4)$$

Excluding the power losses, the power consumption of the switching load is independent of the voltage tolerances since the switching power supply compensates voltage changes, thus, the current consumption increase or decrease. As a result, the power consumption of such devices is constant over the time (independent from the voltage). The current consumption of switching loads is inversely proportional to the voltage changes. This is realized by the complex time-dependent resistance $Z(t)$ as given in Equ. (5.4).

In Tab. 5.1, the theoretical influence of the voltage change of active and passive devices is given. Both types of devices have the same nominal power consumption $P_{nen} = 230 \text{ W}$. In the first case, the grid voltage is precisely 230V. thus, devices types have the same power and current consumption. In the second case, a voltage increase of 10% of the nominal voltage. Thereby, the current and power consumption of the device with the conventional power supply increase, while the power consumption of the device with the switching power supply is constant

Table 5.1: Comparison of the influence of voltage fluctuations of resistive and active load.

case	grid voltage	passive load ($P_{nen} = 230 \text{ W}$)		active load ($P_{nen} = 230 \text{ W}$)	
		current consumption	power consumption	current consumption	power consumption
1	230 V	1 A	230 W	1 A	230 W
2	253 V	1.1 A	278.3 W	0.9091 A	230 W
3	207 V	0.9 A	186.3 W	1.1111 A	230 W

and thus, have a lower current consumption. For lower voltages as described in the third case, the behavior is reversed.

Because of this different behavior, a simple normalization of the power consumption is not constructive, since this is usually done for either simple resistive load or active load, but not for both at the same time. In the following formulas, Equ. (5.5) to (5.8), a standardization is shown for both appliances types individually.

5.2.3 Standardization of Passive Appliances

Voltage changes lead to a varying power consumption of passive devices. This is unfavorable as it can cause problems when detecting devices. The aim is to normalize $P(t)$ and $i(t)$ so that they are independent of voltage changes. The normalized power $P_{norm}(t)$ and current $i_{norm}(t)$ can be written as

$$P_{norm}(t) = P(t) \cdot \left(\frac{V_{nen}}{v_{rms}(t)} \right)^2 \quad (5.5)$$

$$i_{norm}(t) = i_{rms}(t) \cdot \frac{V_{nen}}{v_{rms}(t)}, \quad (5.6)$$

where V_{nen} denote the nominal voltage which is in Europe 230 V. Thus, $P_{norm}(t)$ and $i_{norm}(t)$ are independent from voltage variations.

Standardization of Switching Appliances

The normalized power and current, in case of active appliances, are described by

$$P_{norm}(t) = P_{nen} = v_{rms}(t) \cdot i_{rms}(t) \quad (5.7)$$

$$i_{norm}(t) = i_{rms}(t) \cdot \frac{v_{rms}(t)}{V_{nen}}. \quad (5.8)$$

Since P is independent of the voltage P_{norm} can be calculated by multiplication of current and voltage. The factor for calculating i_{norm} in is inversely to Equ. (5.6).

The comparison of the power consumption of both types of appliances shows that a normalization is only required for passive appliances. Nevertheless, to receive a voltage-independent normalized current, a normalization is required for both types of appliances, as shown in Equ. (5.8) and (5.6).

In order to take advantage of the normalization, the information is required whether the device has a conventional or a switching power supply since the normalization of the current signal is inverse for both types of appliances.

In scenarios where only passive or active appliances are running, the normalization with respect to voltage level can be advantageous. In general, it is difficult to decide in advance to the classification whether a device is an active or a passive appliances. A wrong decision leads to a higher deviation of the standardized values. Therefore, without the information whether a

device is a active or a passive appliances, the normalization cannot be applied. Moreover, some devices may work with a combination of both power supplies, for example, complex devices. Due to this, the normalization to a constant voltage amplitude was not investigated further.

The result of this investigation is that the amplitude of the voltage contains relevant information for NILM. Up to now, no method exists which allow the efficient use of this information since there is no method to analyses the characteristics of the power supply before the classification is applied. Without this information, the voltage mainly provides the information of the phase shift between current and voltage. This knowledge has been considered in the development of a new waveform for NILM, which is presented in the following section.

5.3 Frequency Invariant Transformation of Periodic Signals

The amplitude of the voltage cannot easily be used to improve NILM. Voltage changes, caused by events of appliances are commonly very low. The primary information, including the voltage signal, is the phase shift between current and voltage. The phase shift between voltage and current is an important information and should remained in the waveform.

In this thesis, a new waveform was developed that was specifically designed for event detection [66, 69] and classification [68, 144] for NILM. The new waveform is generated directly by utilizing $i(t)$ and $v(t)$ with the goal of containing as much relevant information as possible. Therefore, the initially measured signal forms $i(t)$ and $v(t)$ are examined more closely. The measured signals are periodically. Only the current signal can lose its periodic characteristic when all devices are disconnected or an event occurs. Furthermore, a constant sampling frequency of $i(t)$ and $v(t)$ is assumed. Since succeeding processing of a periodic signal is difficult, the periodicity is removed from the signal. For this purpose, both current and voltage are decomposed into their periods. Therefore, only the voltage signal is used as a reference for determining the starting or ending point of a period. Since, the voltage signal has less variations, while the current signal varies depending on the number and type of loads. The start and end points of a voltage period are used as reference points for the current signal. Since the mains frequency is not constant, the current signal between the start and the end position of a period is interpolated with the number of sampling points resulting from the sampling rate and the mains frequency. As a result, a re-sampled current signal is available for each period of the voltage. Moreover, these sampling points are always placed at the same position with respect to the period of the voltage.

The interpolation is necessary since the mains frequency is subject to negligible fluctuations. Although the fluctuation is relatively low, commonly <0.2 Hz, in the European interconnection network. For a very short time (10 s), deviations of up to 0.8 Hz are possible without load shedding. Lower frequencies 49.0 Hz, 48.8 Hz, 48.6 Hz, and 48.4 Hz lead to load shedding of approximately 12.5 % for each frequency threshold [47]. But even small frequency variations lead to considerable deviations due to the comparatively long measuring time if these deviations are not taken into account.

The interrelation between mains frequency, grid load and grid feed should be considered. The mains frequency decreases if insufficient energy is produced in the mains. If to much energy is in the grid, the frequency will rise. The fact that the grid frequency is subject to fluctuations was also apparent to normal electricity customers at the beginning of 2018. The mains frequency is averaged over a long time span of hours and day assumed to be 50 Hz. This allows to operate low-cost clocks with the mains frequency. Permanent deviation of the mains frequency affects the time of various devices connected to the grid which using the mains frequency for tracking the time.

The European Network of Transmission System Operators for Electricity (ENTSO-E) published a press release [52] where the cause of the frequency deviation is attributed to the countries of Kosovo and Serbia. Because of political problems between Kosovo and Serbia, to less energy was supplied by these countries into the European grid. Finally, this had only a small

effect on the actual mains frequency. However, this continued for a longer period of months and results in a network time delay of more than five minutes, in the whole European grid. As mentioned, the mains frequency can only be regulated to 50 Hz over a long period of time. Accordingly, deviations of 50 Hz occur over shorter time periods of seconds or minutes. For this reason, the frequency tolerances of the grid should be taken into account for a waveform for NILM, to judge how much this will affect the repeatability.

The new signal has the same number of sampling points over each periods. Moreover, the position of the individual sampling points is not changing (compared to the voltage signal) over the periods. Hence, each of these sampling points can be regarded as a separate dimension. High-frequency and low-frequency signal components are taken into account in the proposed waveform. Since the the voltage is used as a reference to identify the start and end point of a period, all phase-related information is retained in the interpolated current signal. The voltage waveform is not used in the further steps of NILM, thus, the number of data is reduced by half.

The individual steps of this signal transformation have been represented graphically in Fig. 5.5, for better illustration.

- a) The voltage signal is used to calculate the start and end of a period by identifying, the zero crossing from the negative to the positive half-wave. The position of the zero crossing is determined by identifying the sampling point before and after the zero crossing. Using the information of these two points, the zero crossing is calculated by linear approximation

$$f(x) = m \cdot x + b \quad (5.9)$$

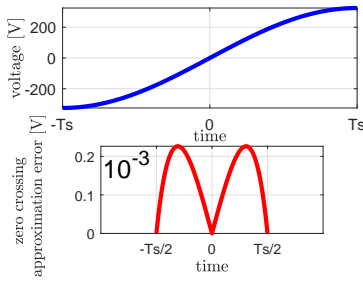
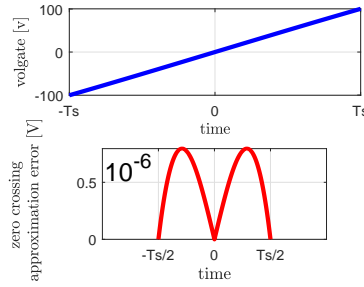
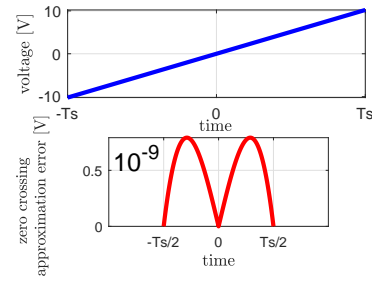
$$f(x) = \frac{f_1 - f_0}{x_1 - x_0} \cdot x + f_0 \quad \left| \quad m = \frac{f_1 - f_0}{x_1 - x_0} \quad , \quad x_0 := 0 \quad (5.10)$$

$$x_z = (-f_0) \cdot \frac{x_1}{f_1 - f_0} \quad \left| \quad f(x) = 0 \quad , \quad f_0 \neq f_1 \quad (5.11)$$

$f(x) = 0$ representing the vertical position of the zero crossing. x_z is the horizontal position of the zero crossing. x_0, f_0 respectively x_1 and f_1 are the sample points before and after the zero crossing. The sine function, in the considered window, is monotonically increasing, so $f_0 \neq f_1$ can be assumed. This assumption can be made since the grid voltage is very similar to a sinusoidal function and the mains frequency f_g is much lower compared to the sampling frequency f_s . The result x_z contains the calculated distance from x_1 to the zero crossing.

Since the linear approximation of the sine function is error-prone, this is considered more closely. Considering the sine from a minimum to the following maximum, the signal is point symmetrical to the zero crossing. This causes a low error by the linear approximation. In case the sampling pint before and after the zero crossing are point symmetrical to the zero crossing, the error made by the linear approximation is zero. Moreover in case that x_0 or x_1 are equal to the zero crossing of the signal form, no error occurs by using the linear approximation.

Fig. 5.2, 5.3, and 5.4 show the error of the linear approximation with sampling frequencies of 100 Hz, 1 kHz and 10 kHz. The segment of the sinusoidal signal is shown from $-T_s = -\frac{1}{f_s}$ to $T_s = \frac{1}{f_s}$. The corresponding error caused by the linear approximation is visualized red. For $f_s = 200$ Hz, the nonlinear behavior of the sine can be seen between $-T_s$ and T_s . The error is plotted in the center of the two sampling positions x_0 or x_1 between $-\frac{T_s}{2}$ and $\frac{T_s}{2}$. For example, in case $x_0 = -T_s$ and $x_1 = 0$, the error is zero. Compared to the higher sampling rates, the error at lower sampling rates is higher since the distance between x_0 and x_1 is bigger. The deviation resulting of the linear approximation of the sine is higher and thus the error too. Of particular note is that the error axes are scaled different since the error differs in increments of power of ten. The method has the advantage that the turning point of the sinusoid is identical to the zero crossing. Since the derivation at this position is very small, compared to the rest of the signal, the zero crossing is more appropriate for

Figure 5.2: $f_s = 200$ HzFigure 5.3: $f_s = 1$ kHzFigure 5.4: $f_s = 10$ kHz

Error of determining zero crossings using a linear approximation for different sampling intervals.

defining the beginning and the end of a period. The maximum error resulting from the linear interpolation of $f_s = 10$ kHz is less than 10^{-9} V.

- b) The new horizontal position for the re-sampled current values, is calculated according to

$$I_{m,n} = \frac{I_{m,0} - I_{m+1,0}}{n_k} \cdot n \quad \left| \quad n = 0 \dots n_k - 1, \quad (5.12)$$

where $I_{m,n}$ is the index of the m -th period at the n -th index and n_k is the number of sample points for each period.

The number of sample points is calculated by $n_k = \frac{f_s}{f_g}$.

- c) The current signal is interpolated on the calculated indices by using a linear interpolation. In addition to the low computational complexity, the non-sinusoidal form of the current is the reason to use a linear interpolation. A detailed discussion of different interpolation methods is presented in Chap. 6.

The interpolation, results with a current signal, including a identical number of sampling points, for each period. Furthermore, the position of a sampling point within a period is identical for all periods, related to the zero crossing of the voltage signal.

- d) The one-dimensional signal is converted into a multi-dimensional signal. The result is that there are as many dimensions as sampling points in a period. In this example, each period exists of 21 sampling points. The two presented periods are shown in l direction. The result is a $n_l \times n_k$ dimensional signal where L is the number of periods and K is the number of sampling points of each period.

The transformation from FIT-PS into $i(t)$ can be written as

$$\mathbb{N}, \mathbb{N} \rightarrow \mathbb{R} \\ l, k \mapsto i \left(\underbrace{(l-1) \cdot T_g}_{\text{period}} + \underbrace{T_s \cdot k}_{\text{inside one period}} \right), \quad (5.13)$$

where l describes the numbers of the periods and k is the sampling number within one period of the FIT-PS signal.

In Fig. 5.6, a section of the signal from the Building-Level fully-labeled dataset for Electricity Disaggregation (BLUED) dataset [4] is shown in FIT-PS signal representation. There are 600 periods with a sample rate of 1.2 kHz and a grid frequency of 60 Hz, resulting with 20 samples per period. The amplitude of the current is shown on the vertical axis, as well as with colors. Fig. 5.6 shows a Steady State (SSt) of a refrigerator.

Since a 3D representation is sometimes difficult to interpret, the amplitude of the current can be visualized with colors. Fig. 5.7 shows the top view of Fig. 5.6, wherein a 3D representation was reduced to a 2D representation.

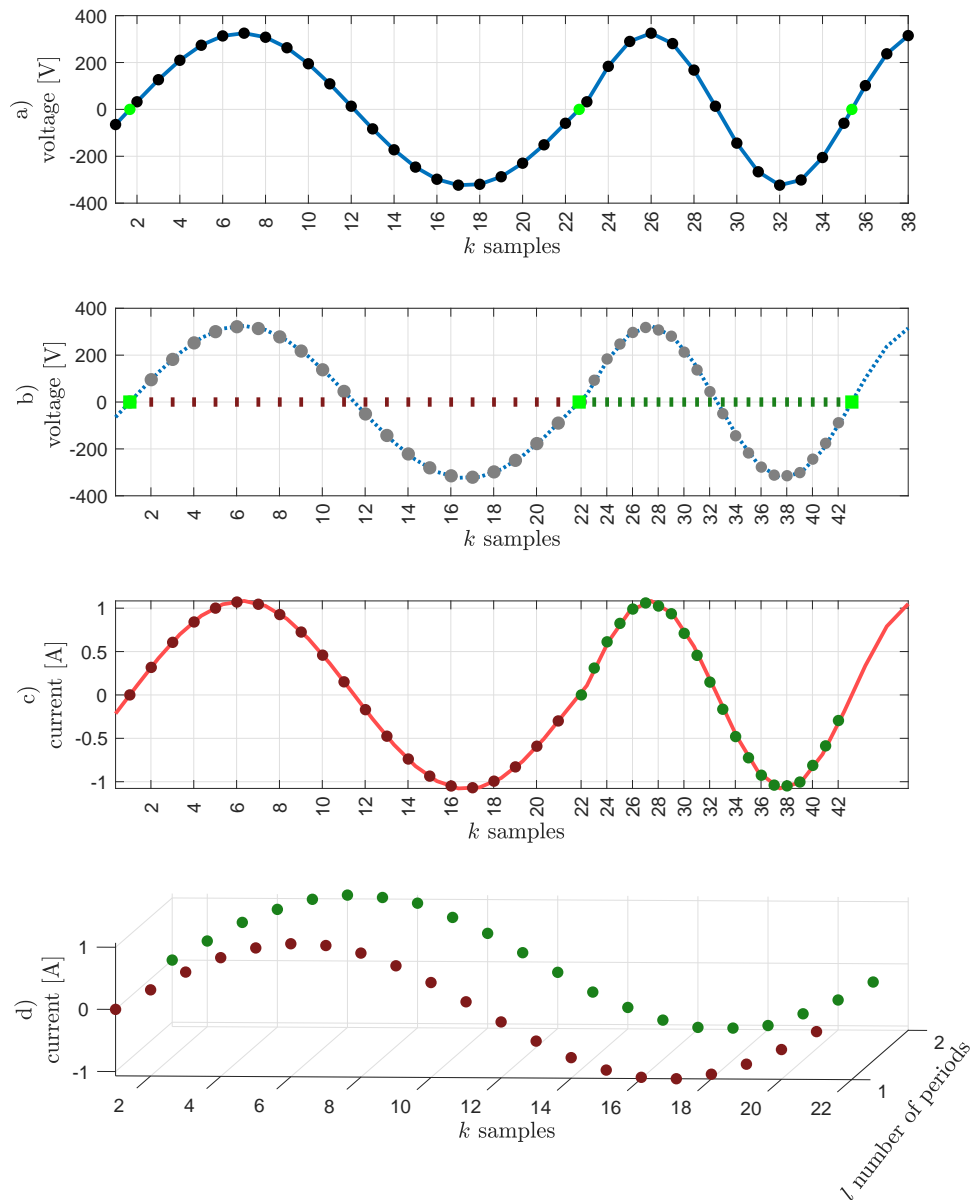


Figure 5.5: Graphical representation of FIT-PS [68]. a) finding zero crossings, b) calculating new sampling positions, c) re-sampling and interpolation of the current signal based on the new sampling positions, d) periodical signal representation

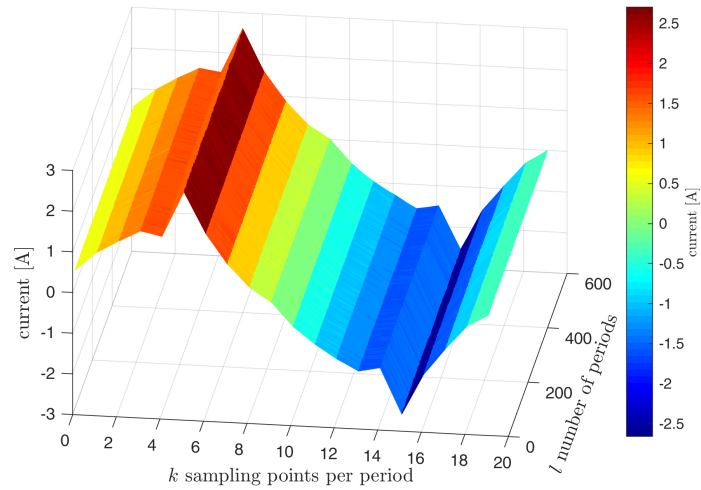


Figure 5.6: FIT-PS of a signal with 20 samples per period showing a SSt of a refrigerator

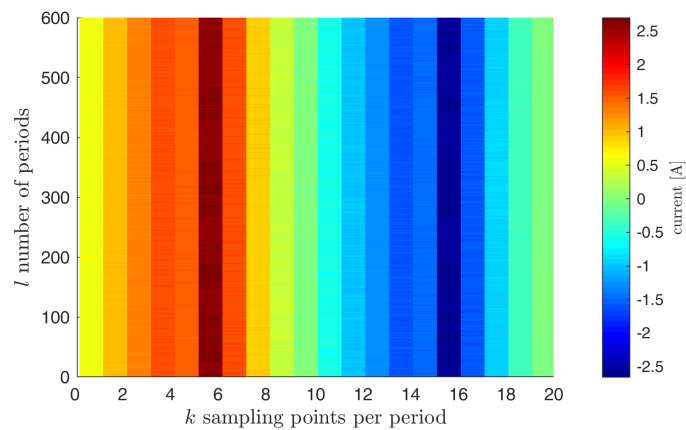


Figure 5.7: 2D signal visualization of Fig. 5.6. The current amplitude is colored.

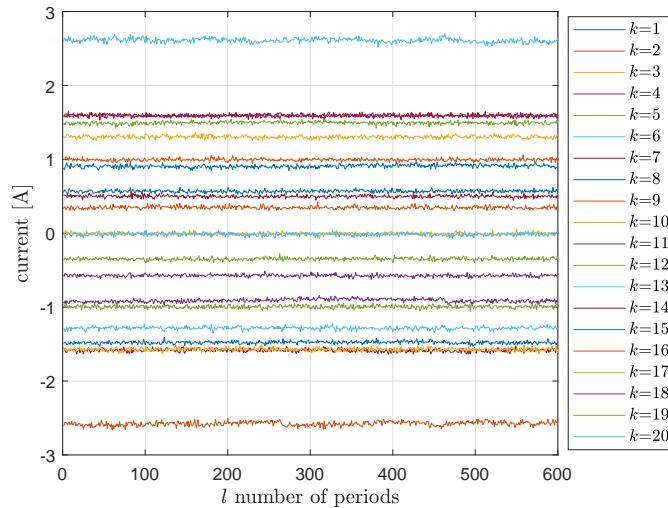


Figure 5.8: FIT-PS signal representation of the individual dimensions.

A significant advantage of the new signal representation is that the periodicity of the signal is omitted without the need for averaging. Therefore, no essential information is lost, since only a adjustment of the sampling times is carried out. Fig. 5.8 shows a SSt signal converted with FIT-PS. The k dimensions are plotted individually over the periods l . Since FIT-PS ensures that the position of each sampling point during a period of the voltage is at the same position, the current value for each period is nearly constant during a SSt. The FIT-PS waveform converts two periodic signals into a non-periodic waveform which simplifies the subsequent tasks without losing relevant information.

Fig. 5.8 visualize that most of the k dimensions have a low absolute amplitude value. At maximum or minimum of the current amplitudes (between 2 A to 3 A respectively -2 A to -3 A) there is only one sampling point per period. The reason for this is not obvious, as for an equidistant sampled sinusoidal signal, a concentration of sample points to the maximum, as well as to the minimum is expected, because of lower signal change at these range in the signal. Fig. 5.9 illustrates a simulated sinusoidal signal. On the left, a period of sine is plotted with 200 sampling points and an amplitude of one ampere. On the right, the histogram of this signal is shown. The samples are sorted into 50 equally spaced columns. The most sampling points are in the area of the maximum and the minimum.

The reason why this is not the case in Fig. 5.8 is that the current is not sinusoidal. This is more obvious in Fig. 5.7. The temporally short peak in the period leads to only one measuring point in the area of the maximum and the minimum. To get a more detailed waveform in the area of occurring current peaks, a higher sampling rate is required. Especially in case of lower sampling rates, it must be ensured that the Nyquist sampling theorem is fulfilled. Since the signal form of the current depends on the active devices, a sinusoidal waveform cannot be assumed. High-frequency signal components needs to be sufficiently low-pass filtered. Decisive for this are the sampling frequency f_s and the resolution of the Analog-to-Digital Converter (ADC). The system, developed during this work, uses an ADC with $f_s = 10$ kHz and a resolution of 16 bit. To ensure that no aliasing occurs, the low-pass filter before the ADC must be designed accordingly. The signal components with a frequency greater than half of the sampling frequency have to be filtered in such a way that the Least Significant Bit (LSB) is not effected by higher frequency components. Therefore, a low pass filter is required, reaching at least $20 \cdot \log(2^{-16}) = -96.33$ dB at the half sampling frequency.

The information content of the signal representation is shown in Fig. 5.10, where a Transient State (TSt) of a refrigerator, as well as the SSt are illustrated. At the $l = 300$ the refrigerator is turned on. The switch-on process can be seen in the TSt. During the transient, an apparent

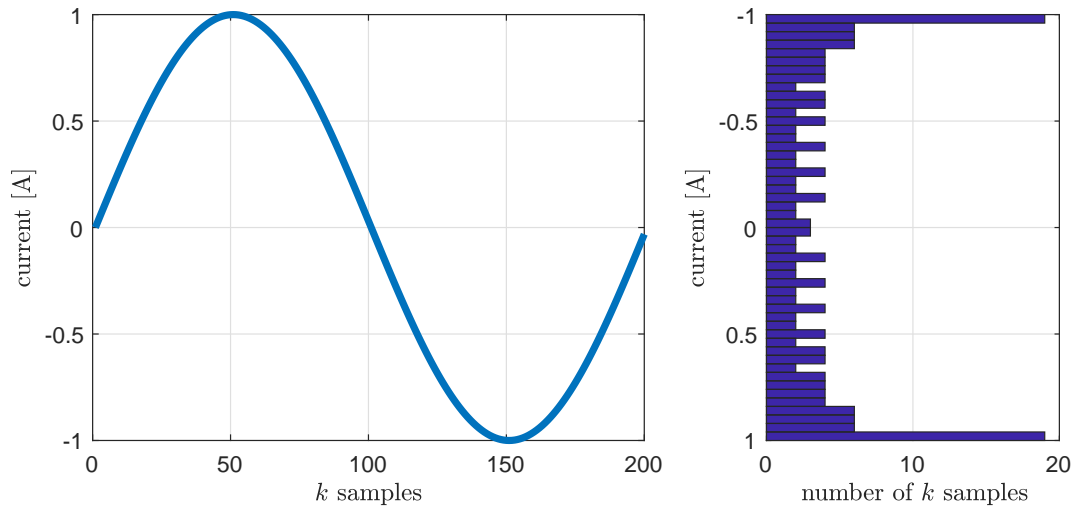


Figure 5.9: On the left a sine with 200 equidistantly samples and on the right a histogram of the samples from the sinusoidal signal is visualized.

change of the amplitude (presented with colors) of all dimensions can be seen. Additionally, a phase shift during the TSt occurs. The representation of the phase shift between current and voltage signal is possible since the reference between current and voltage signal is maintained, even if the voltage signal is not used further. After the TSt a second (SSt z_{i+1}) follows.

The signal shape and the phase of SSt z_i and z_{i+1} differ. At the SSt z_{i+1} , the maximum and minimum are visible at $k = 6$ respectively $k = 16$. While at SSt z_i , the maximum and minimum are not obvious. The zero crossing from positive to negative is for the first SSt at $k = 7$ and for the second SSt at $k = 9$. Additionally, the zero crossing from the negative to the positive occurs for SSt z_i at $k = 17$, compared to SSt z_{i+1} at $k = 18$.

The original sampling frequency of the BLUED dataset is 12 kHz. As mentioned in the state of the art, higher sampling rates can significantly improve the results of the event detection and classification. Therefore, in the subsequent chapter, the FIT-PS waveform is applied on a 12 kHz signal.

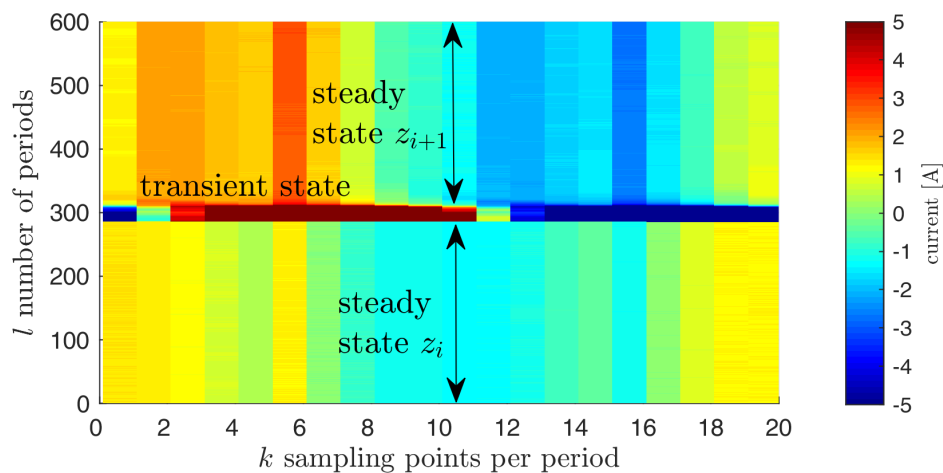


Figure 5.10: FIT-PS waveform with 20 samples per period SSt z_i : device one; TSt: switching on procedure of device two; SSt z_{i+1} : device one and device two (refrigerator).

5.3.1 FIT-PS with Higher Sampling Frequency

The FIT-PS signal representation was applied successfully on signal forms in a range below 4 kHz. In this section, FIT-PS is applied on measurement signals with up to 12 kHz. Parts of this section have already been published [69].

Fig. 5.11 presents the FIT-PS converted waveform, calculated from a originally current and voltage measurement which contains a sampling rate of 12 kHz. Thus, the resolution in the k direction is ten times higher compared to Fig. 5.10 where a sample frequency of 1.2 kHz, at a mains frequency of 60 Hz, is used. Due to the high sample frequency, additional characteristics in the signal are visible. For example, the waveform differs obvious from a sinusoid function. In the first and second SSt at $k = 158$, a small peak can be observed. The peak only occurs in the negative half-wave. For most devices, this is uncommon as these devices have a point-symmetric signal form of positive and negative half-wave. Additionally a different noise level on the single dimensions k along l are visible. This information can be used for the following NILM tasks.

Fig. 5.11 illustrates the magnified TSt of Fig. 5.12. This specific part shows the startup process of the refrigerator in detail. The length, phase shift, and shape of the transient can be used for event detection or classification. The exact position within a period, where the device turns on, can be identified. In case a device is switched on always at the same position within a period, this information can be used for the classification.

For purposes of illustration the amplitude of Fig. 5.11 and 5.12 are truncated above respective below ± 5 A.

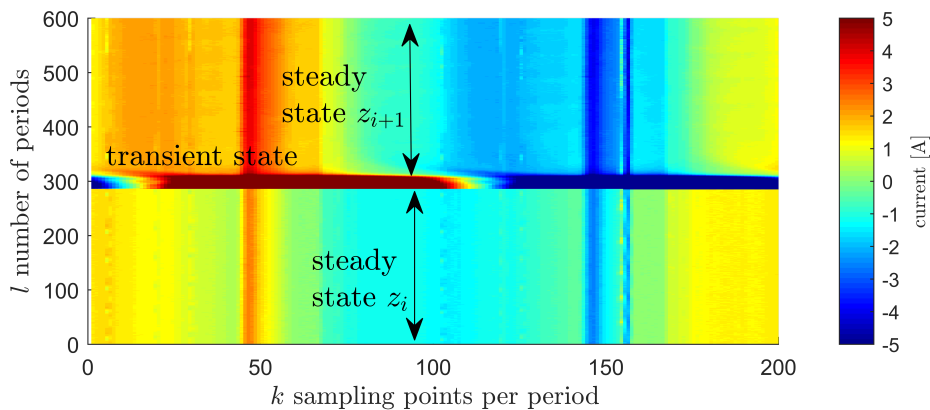


Figure 5.11: FIT-PS waveform with 200 sample points per period; SSt z_i : device one; TSt z_{i+1} : switching on procedure of a refrigerator; SSt z_{i+2} : device one and two. Published in [66].

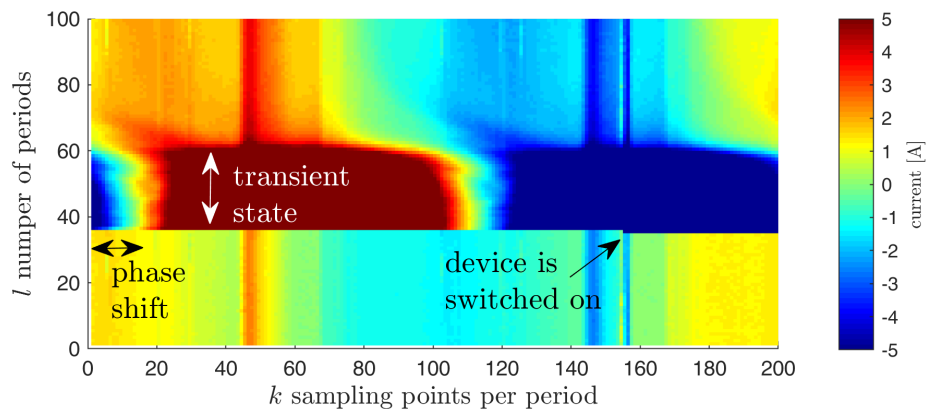


Figure 5.12: FIT-PS waveform with 200 sample points per period illustrating a turn on TSt of a refrigerator.

5.3.2 Reduction of the Sampling Frequency

FIT-PS applied on SSt signal forms enables the reducing of the sampling frequency with almost no loss of information. This is possible because the frequency f_d of switching devices on or off, of each appliance d is much lower than the grid frequency f_g

$$f_d \ll f_g. \quad (5.14)$$

A SSt is defined as a signal region without significant change in the signal. At the same time, a SSt can prevail for a very long time. For example, during the night only a few devices are switched on and off automatically. Therefore, the amount of useful information, compared to the amount of data is low. This is shown during the SSts of Fig. 5.10 and Fig. 5.11. The different periods are very similar during the SSt; thus, they contain the same information.

The mains frequency f_g is usually much faster than the switching frequency of the appliances f_d . Assuming that Equ. (5.14) is valid, FIT-PS allows reducing the amount of data because the sampling frequency can be reduced without losing information during a SSt. During the SSt, the individual periods from the FIT-PS converted signal contain the same information. Therefore, one measuring point $i[k, l]$ per period l can be used. With each additional period, the measuring point is shifted by $k + 1$ as illustrated in Fig. 5.13, whereby $l \in \{1, \dots, L\}$ and $k \in \{1, \dots, K \left(= \frac{f_s}{f_g} \right)\}$. For better visualization, a sampling frequency of 1.2 kHz, at a mains frequency of 60 Hz, is used. Only one sampling position per period is used, independent of the resolution, respectively the sampling frequency of the conventional method. Therefore, a data reduction of a factor equal to the number of samples per period is achieved. The amount of data, which can be reduced with this method, increases proportionally with the sampling frequency since the grid frequency is usually fixed. However, the time required to represent a complete period changes with the utilized resolution since for each sampling point another period is used.

With this method less powerful hardware is needed for processing; thus the measurement system needs less memory space. For a practical realization of this method, a zero detection of the voltage signal, an accurate timer for triggering the ADC and a compensation of mains frequency fluctuations are required.

In Fig. 5.14, the variance of the downsampled signal with respect to the FIT-PS representation is presented. The downsampled signal is shown in red color. The variance of the FIT-PS waveform over 20 periods is shown in gray. There is only a small variation between the downsampled signal and the original FIT-PS representation.

Fig. 5.15 present the variation of a signal, measured with 12 kHz. Additionally the resulting downsampled signal measured over 200 periods is presented. Since the sampling frequency of the downsampled signal depends mainly on the grid frequency f_g , a data reduction of more than factor 200 is possible in this case. The number of periods over which the information is collected, is extended with the number of sampling points. For the illustration of one period, 200 periods are required.

The higher sampling rate leads to more variations of the different periods during the SSt. Fig. 5.15 illustrates this at $k = 153$ and $k = 155$, where the original measured signal in gray differs from the constructed period in red. The difference can be explained by the longer time required to display a complete period. Additionally, the higher resolution of the FIT-PS waveform contains more device-specific variations on individual dimensions. Nevertheless, the reconstructed signal represent the original signal very well.

The distance between the events of the appliances needs to be larger than the periods required for the reconstruction of the signal. To ensure that devices are switching on or off with a minimal distance is typically complicated. An event detection, based on this waveform, will produce a slightly higher error in case of short event intervals. Nevertheless, this approach offers different possibilities for investigating the SSt. The signal shape can be represented by a sampling rate less than the mains frequency. In Chap. 7, this method was applied for the event detection. A summary of the results can be found in Tab. 7.1 and 7.2.

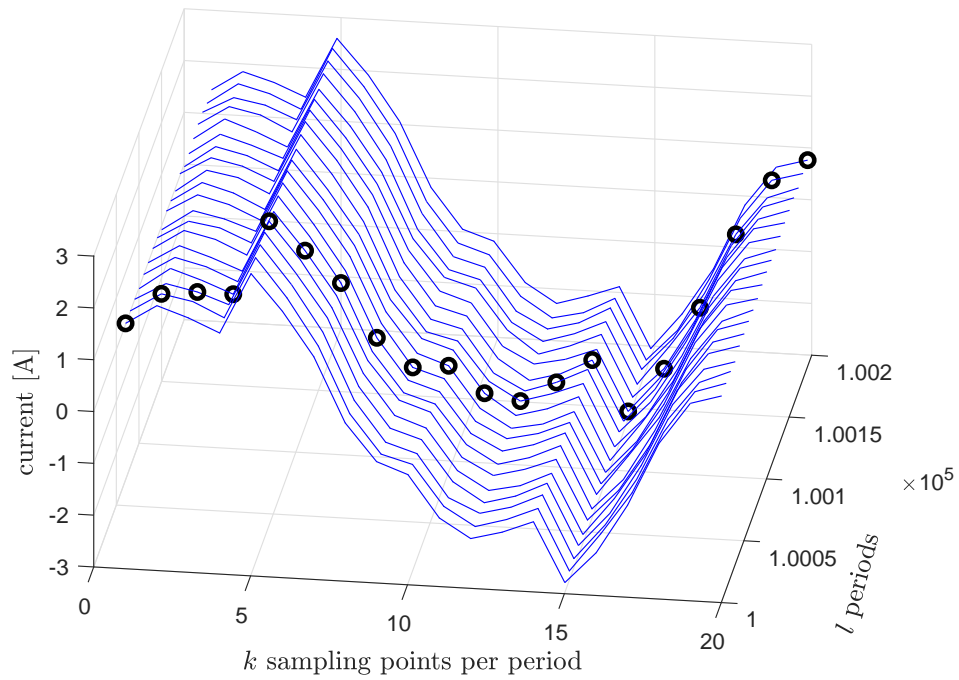


Figure 5.13: Representation of a downsampling without losing the signal shape, during a SSt. 60 Hz periodic signal (continuous lines) under sampled at 57 Hz sample frequency (black circle).

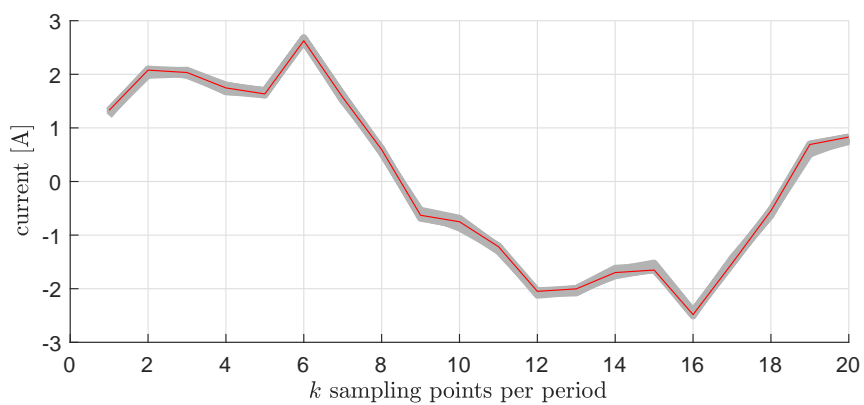


Figure 5.14: Comparison between the variance of the signal (gray), sampled with 1.2 kHz and the downsampled signal (red).

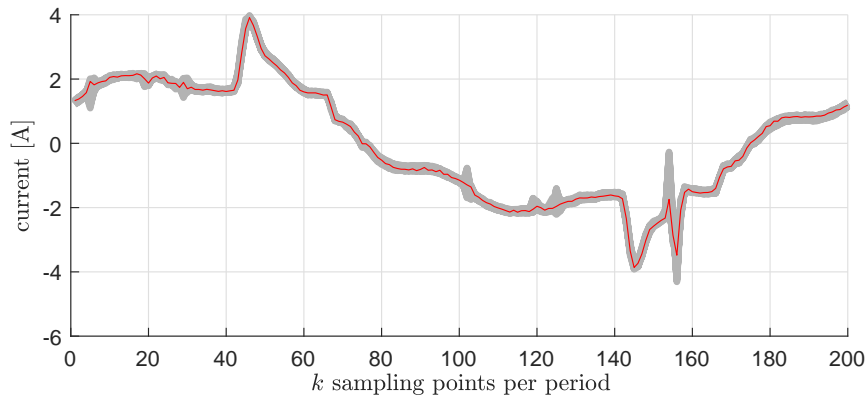


Figure 5.15: Comparison between the variance of the signal (gray), sampled with 12 kHz and the downsampled signal (red).

5.3.3 Dimension Reduction in k -Dimension

The number of dimensions corresponds directly with the sampling frequency of the measured voltage and current. In general, the question about the required sampling frequency for the event detection and a subsequent classification in NILM cannot be answered. As mentioned in the Chap. 2, this depends strongly on the measuring scenario and individual appliances. The aspect of the sampling frequency needed for the FIT-PS waveform is analyzed in more detail for the BLUED dataset. At the beginning 50% of the dimensions are used, followed by 25% and 12.5%. The percentage reduction of dimensions is equal to the reduction of the sampling frequency. A visual illustration of the use of different sampling frequencies and correspondingly changing number of samples per period is presented in Fig. 5.16. Further examples of FIT-PS waveform, applied on different appliances of the BLUED dataset with different frequencies, can be found in the appendix, Fig. A.4 to A.9.

Multiple combinations are possible for the selection of the individual k dimensions. The optimal selection of dimensions could lead to improved results of specific applications, since noise affected dimensions and correlated information can be removed in advance. Furthermore, an optimization by weighting or reducing dimensions is possible. However, this results in an optimization for a specific question, which needs to be recalculated by slight changes of the framework conditions, for example when a new device is added. At this stage of the work only scenarios are considered which are resulting from a reduced sampling rate.

5.4 Results

For a final performance evaluation of the new waveform, an application on the event detection and the classification is necessary. The results of the event detection are presented in Sec. 7.4. The results of the classification are given in Sec. 9.7. Since event detection and classification have not been introduced so far, the new signal representation can be only evaluated manually based on visualization. In Fig. 5.16, different sample rates of the measurement signal are shown. The FIT-PS waveform, calculated from a higher sampled current and voltage signal, contains more details. However, the noise increases too since a stronger averaging, thus, low-pass filtering is applied with a lower sampling. The device illustrated in Fig. 5.16 is an energy saving lamp. The switching power supply operates only in a limited area of the period which is represented precisely with the FIT-PS signal representation. FIT-PS allows evaluating the change of each k dimension. This is an advantage compared to commonly used power waveforms calculating a RMS value over the whole period. In particular, a subsequent classification can benefit from the information within a period. The prerequisite for this, is a suitable feature extraction and the use of suitable classification methods.

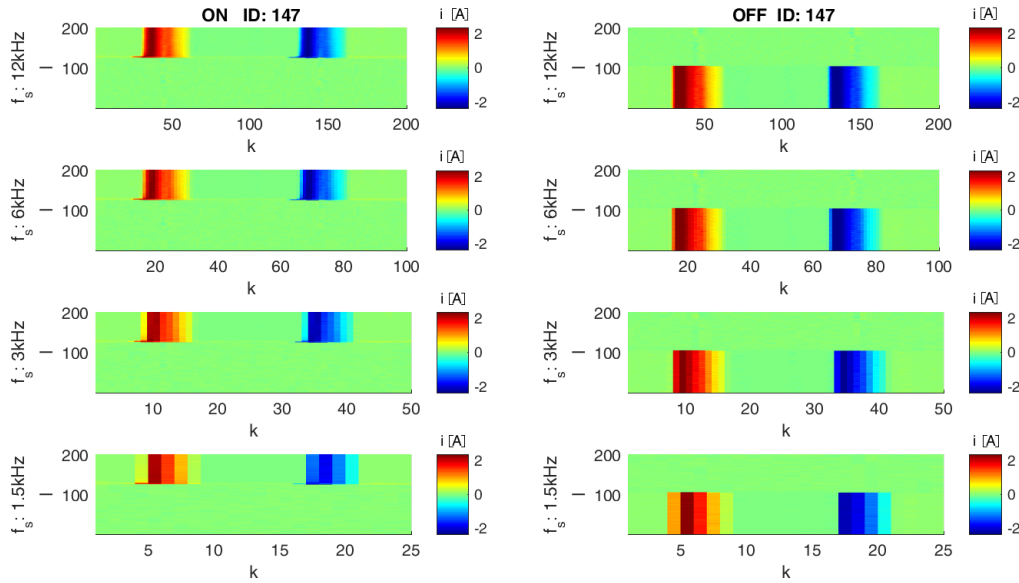


Figure 5.16: Different sampling rates of a turning on (left side) and off (right side) process of an energy saving lamp, from BLUED dataset [4], presented with FIT-PS. Sampling rates from 12 kHz to 1.2 kHz are illustrated. Thereby the numbers of dimensions k are varying from 200 to 25 while the number of 200 periods in l direction is constant. The amplitude of the current is represented colored.

5.5 Discussion

The waveforms used so far, in the NILM context, have not been developed explicitly to fulfill the needs of NILM. The fact that a combination of the existing waveforms leads to the best result summarize the key problem that no individual waveform includes enough information or does not represent the information suitable for the used classification method. The combination of waveforms leads to higher computational effort. Furthermore, the weighting of the waveforms must be determined to the specific application scenario.

The waveform should contain, as much as possible information, which is relevant for NILM and prepare the signals in such a way that they can be handled successfully in the subsequent tasks. FIT-PS represents the current signal with regards to the zero crossing of the voltage signal; thus, the phase information remains. The FIT-PS waveform allows to analyze each sampling point of a period individually. This k dimensions are non-periodic which simplifies the further processing steps.

The main information from the voltage signal, which can be used for NILM, is the phase shift of the current compared to the voltage signal, described in Sec. 5.2. More relevant information of the voltage amplitude can be used by distinguishing between appliance with a switching or a conventional power supply. With the use of an additional classification task, this problem could be solved. However, there are also devices which are not clearly assigned to the one or other category. Not considering the voltage also reduces the data amount by 50%. In case the voltage signal should be used as well, it needs to be interpolated too. This could be also necessary for applications, where the exact power consumption should be calculated.

The reduced noise is the main effect by relying on the real or reactive power. The non-periodic waveform of real and reactive power is an additional reason why these waveforms are commonly used in NILM. They allow a good graphical overview of the measurement. Switching events can be visualized more accurate with non-periodic signal compared to periodical signals like the current $i(t)$.

With the FIT-PS approach, the individual dimensions are also non-periodic. An averaging in k direction is possible to reduce noise. However, this has negative effects since the high frequency information within the waveform is lost. The individual k dimensions of FIT-PS are non-periodically, thus, the mean value over the l periods can be calculated without losing the information of the waveform. As for active or reactive power, the result is a reduced noise level of the FIT-PS waveform. The advantage compared to other waveforms is that the current signal form is still available in the individual k dimensions of FIT-PS.

With the help of a hardware zero-crossing detector and a Phase-Locked Loop (PLL), the measured signal can be measured directly in the FIT-PS waveform. A measurement instrument capable of adapting the sampling rate and equipped with a zero-crossing detector directly in the hardware could nearly reduce the total calculation effort to generate FIT-PS.

The use of FIT-PS increases the requirements for the event detection and classification, as they have to deal with higher dimensional input signals or features. A more detailed signal also contains more noise, thus, an adaption of the event detection and classification algorithm to the corresponding waveform is necessary. Compared to an one or two-dimensional 50 Hz sampled signal, the complexity of a FIT-PS converted waveform is higher. Common datasets such as the the BLUED dataset [4] uses a sample rate of 12 kHz or more at a mains frequency of 60 Hz. Using FIT-PS on this dataset results in a 200-dimensional signal, illustrated in Fig. 5.12. A reduction of the number of dimensions can be realized by reducing the number of re-sampled positions (during the interpolation), see Fig. 5.5 b). A reduction of the sample rate during the interpolation of FIT-PS, has a similar effect on the resulting waveform, like a lower sampling frequency of the measured current and voltage signal. To avoid aliasing and the lose of high-frequency information, the sampling frequency after the interpolation should not be below, the sampling frequency of the measuring system or an appropriate filtering needs to be provided.

Fig. 5.17 presents a measurement of the Home Equipment Laboratory Dataset 1 (HELD1) dataset. The top diagram of Fig. 5.17 shows the active power. On the bottom, the FIT-PS signal interpretation is presented whereas events are marked with an red 'x'. For the FIT-PS representation, each dimension is plotted independently, representing the current over the periods. The dataset has a sampling frequency of 4 kHz at a mains frequency of 50 Hz; thus, 80 sampling positions per period exist. This results in 80 lines which are colored in Fig. 5.17.

Events from appliances with a large power consumption are visualized with both waveforms obviously. In order to compare the waveform of switching events of small appliances, a more detailed visualization is presented in Fig. 5.18. The event next to the period 6,200, is shown in more detail and marked with a red 'x'.

The active power drops at the position where the event occurs. The FIT-PS signals have only in some dimensions a significant change. Furthermore, the noise of the different dimensions differs. In case of the event detection, this can be advantageous since only dimensions with a low noise level can be used. Additionally, an adaption of the threshold to the noise level of the individual dimensions could be implemented. Since several dimensions are available, a 'majority voting' could also be used.

A derivation is commonly used, for a first analysis of the waveform, for event detection. This is advantageous for the presentation of FIT-PS since the offset of the individual dimensions is eliminated. Fig. 5.19 shows the derivation of active power and FIT-PS. The event is well visualized for both waveforms. Since FIT-PS presents dimensions from the positive and negative half-wave at the same time, the derivation of the dimensions can be positive or a negative. Furthermore, dimensions with a low noise level could be preferred.

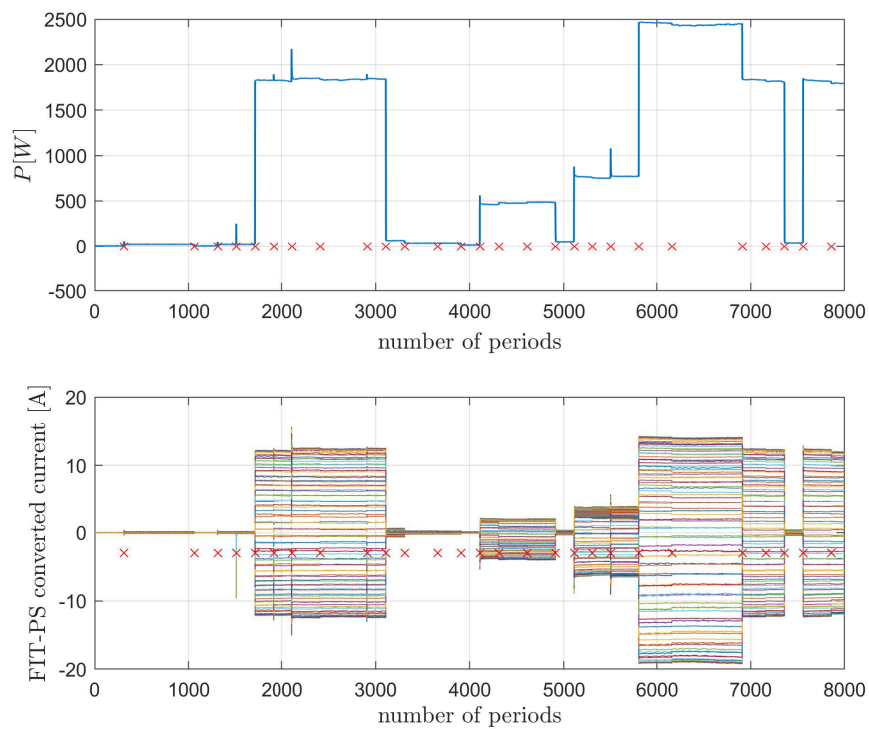


Figure 5.17: Comparison of the active power waveform P (top) with the FIT-PS representation (bottom). The waveform is selected from the HELD1 dataset (measurement 0242); events are marked with red 'x'.

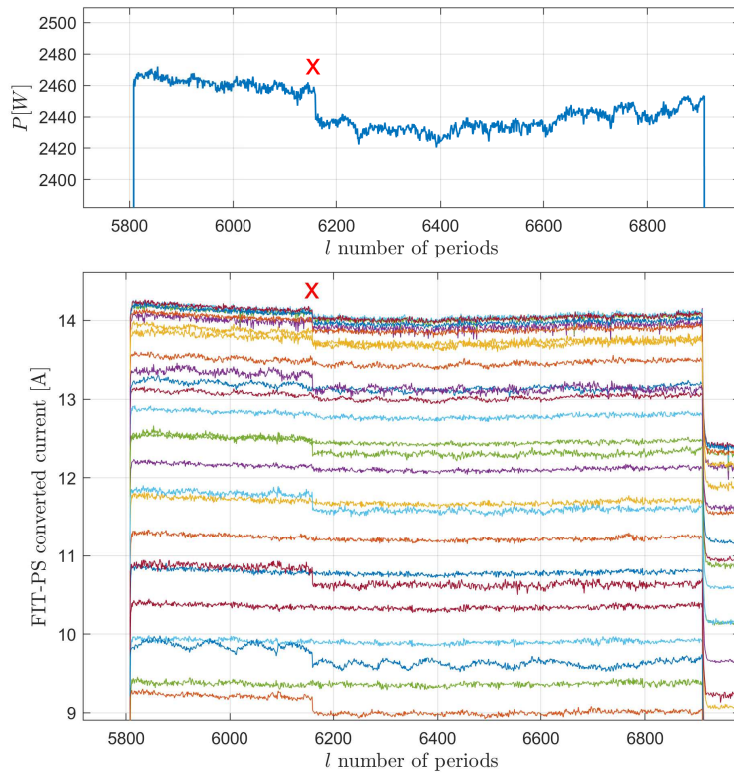


Figure 5.18: Detailed comparison of the active power waveform P (top) with the FIT-PS presentation (bottom). Signal section of the HELD1 dataset (measurement 0242).

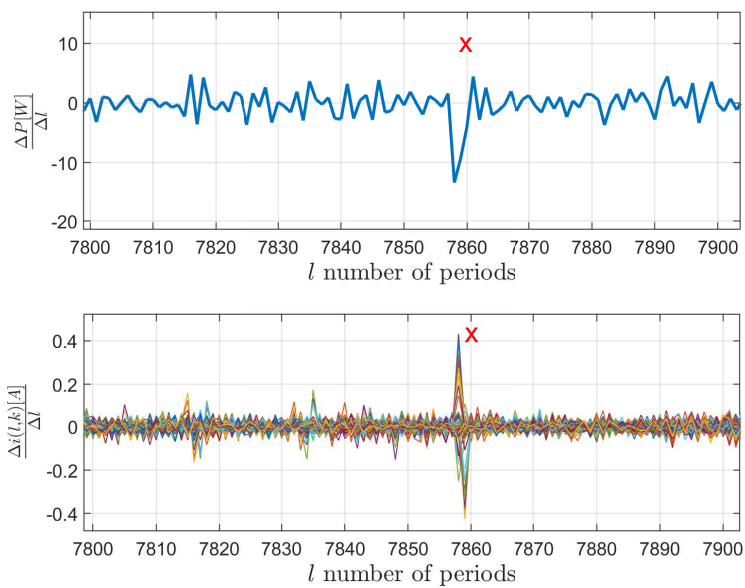


Figure 5.19: Detailed comparison of the derivative of the apparent power waveform P with the derivative of the FIT-PS presentation. Signal section of the HELD1 dataset (measurement 0242).

5.6 Conclusion

The FIT-PS waveform represents the information, relevant for NILM, from the initially measured voltage and current signal in one waveform. As the information is relevant for NILM, the amplitude characteristic of the current, as well as the phase shift between current and voltage, are presented. Whether FIT-PS contains for NILM relevant information and represents this information adequate, cannot be answered presently. Therefore, hypothesis 1 from Chap. 3: “Transforming the current signal according to the phase shift between current and voltage produces a NILM specific waveform that contains the relevant information of the current and voltage signal.”, cannot be answered presently. Further investigations of the subsequent event detection and classification with the new waveform are necessary, in order to compared the results with other waveforms.

Chapter 6 | Different FIT-PS Methods

In this chapter, the different methods for Frequency Invariant Transformation of Periodic Signals (FIT-PS) are described in detail. In chapter 5.1, the question arose about the optimal method concerning accuracy and calculation effort of the FIT-PS procedure.

Three different approaches are discussed in the following. Sec. 6.1 presents the pseudo-code of the FIT-PS method presented in Sec. 5.1. In this method the zero crossing is determined by linear approximation between the value before and after the zero crossing. With the knowledge about the exact zero crossing, the current signal can be interpolated and re-sampled.

In case of general low-cost embedded systems, interpolating and re-sampling requires processing power as floating point is typically required. Sec. 6.1 presents the FIT-PS method without resembling and interpolating the current signal. The current signal is split into the individual periods, where the zero crossing, from negative to positive half-wave of the voltage signal, defines start and end of a period. As a result, the computational effort can be minimized since no interpolation is required for this method.

The third Sec. 6.3 investigates if a more complex interpolation can improve the resulting waveform. Therefore the zero crossing of the voltage are calculated by using a fitting.

The effects to the succeeding interpolation, are discussed in Sec. 6.4 and evaluated by a classification method. Especially the classification task relies on the best possible input signals since noise on the inputs will increase the size of the different clusters. Thus, the probability of overlapping clusters increases significantly. Due to this circumstance, the classification is used to evaluate the accuracy of the different FIT-PS methods. The used classification method is described in Chap. 9 in detail. However, the classification method do not have to be understood in detail for the interpretation of the results of this chapter.

6.1 FIT-PS with Interpolation

In order to be able to compare the different possibilities of implementing FIT-PS, the method is described as pseudo code. The following pseudo code (Alg. 1 to Alg. 4) describe the FIT-PS method which has been presented in Sec. 5.1. This algorithm was already published in [68, 66].

“The FIT-PS method divides the sampled signal with respect to its fundamental frequency, thus, the line frequency. Therefore, the first step is the detection of the zero crossings:

Algorithm 1: Finding zero crossings

```
nn = 1
for n = 1 : N - 1 do
    if (v(n) < 0) AND (v(n + 1) > 0) then
        IDXZero(nn++) = n
```

The variable ‘IDXZero’ has as many entries as the signal has zero crossings. The pseudo code neglects additional checks to eliminate multiple detections of zero-crossings for one period due to simplicity.

Based on the absolute position of the zero crossing for the given data, Alg. 2 describes the linear approximation of the zero crossing.

Algorithm 2: Calculation of the exact position of zero crossings

```

for  $l = 1 : \text{length}(\text{IDXZero}) - 1$  do
   $n = \text{IDXZero}(l)$ 
   $\text{IDXZeroShift}(l) = \left( -v(n) \cdot \frac{1}{v(n+1)-v(n)} \right)$ 

```

‘IDXZeroShift’ is, for a given zero crossing, the estimated shift of the crossing itself. ‘IDXZeroShift’ ranges from zero to one; thus, the maximum deviation is less than one sample. Due to a possible variation of the amplitude, the maximum steepness and almost constant derivative of the sinusoidal signal, the zero crossing is chosen to define the length of a period.

After the zero crossings have been found, the indices can be assigned to the individual periods. So far, this has been done by linear interpolation as described in Alg. 3.

Algorithm 3: Periodical allocation by interpolation

```

 $X_{l,k} = 0$ 
for  $l = 1 : \text{length}(\text{IDXZeroShift}) - 1$  do
   $\text{Len} =$ 
   $(\text{IDXZero}(l+1) + \text{IDXZeroShift}(l+1)) - (\text{IDXZero}(l) + \text{IDXZeroShift}(l))$ 
   $\text{Dis} = \frac{\text{Len}}{\#\text{SampPerPeriod}}$ 
  for  $k = 2 : \#\text{SampPerPeriod}$  do
     $k1 = \text{IDXZero}(l) + \text{IDXZeroShift}(l) + \text{Dis} \cdot (k - 1)$ 
     $k2 = \text{floor}(k1)$ 
     $k3 = \text{ceil}(k1)$ 
     $X_{l,k} = \text{intp}(\text{IDXZeroShift}(l), \text{current}(k2), \text{current}(k3))$ 

```

$X_{l,k}$ depicts the resulting matrix as having the dimension of $n_l \times n_k$ where n_l is the number of periods and n_k is the number of sampling points of each period. The number of sampling points per period is given in k direction whereas the number of periods is given in l direction. The following parameters have been selected for the application of the HELD1 data set: $n_k = \frac{fs}{fg} = \frac{4 \text{ kHz}}{50 \text{ Hz}} = 80$ and $n_l = 75$ periods. The minimum distance between the events is 3 seconds. While calculating the steady state, 1.5 seconds before and after the event are ignored to exclude the transient state.

The interpolation function (‘intp’) performs a linear interpolation.

Algorithm 4: Function of the linear interpolation

```

float intp( interval, f1, f2)
return  $f1 + (f2 - f1) \cdot \text{interval}$ 

```

The argument ‘interval’ has a range of zero to one, whereas $f1$ and $f2$ are two values. ‘intp’ returns the calculated value at position ‘interval’; thus, interpolates the value between $f1$ and $f2$ at position ‘interval’. Due to the linear interpolation, a slight deviation from the original signal may occur. To keep this as low as possible, the aliasing filter of the measurement system must be designed correctly.” Published in [68].

6.2 FIT-PS without Interpolation

One possibility to minimize the calculation effort is to skip the interpolation. Instead of calculating the zero crossing, the position of the closest sampling position of the voltage waveform is selected as zero crossing. The position of the zero crossing define the start of the period whereas the previous sampling position of the following period defines the end of the period.

A disadvantage of the excluded interpolation is the occurring error, caused by the distance

of the selected sampling position to the correct zero crossing. As the sampling frequency is in general not synchronized to the mains frequency, the number of sampled points per period can vary. Additionally, the variance over the periods of all dimensions increases because instead of the linear interpolation, the index which is the closest to the zero crossing is used for dimension one. Furthermore, the error can increase with every index of a period, if the sampling frequency is not exactly a multiple of the grid frequency.

Nevertheless, this implementation method was investigated to examine if these errors have a significant influence on the results regarding Non-Intrusive Load Monitoring (NILM). The following lines describe the pseudo-algorithm of this implementation.

First, the position of the zero crossings is determined. This is equivalent to the description in Alg. 1. Since the interpolation is ignored, in contrast to Alg. 2, the sampling point with the smallest distance to the actual zero crossing is used.

Algorithm 5: Calculate the closest position to the zero crossing

```

for  $l = 1 : \text{length}(\text{IDXZero}) - 1$  do
   $\text{IDXZeroShift}(l) = \begin{cases} \text{IDXZero}(l) & \text{for } v(\text{IDXZero}(l) + 1) - v(\text{IDXZero}(l)) \geq 0 \\ \text{IDXZero}(l) + 1 & \text{for } v(\text{IDXZero}(l) + 1) - v(\text{IDXZero}(l)) \leq 0 \end{cases}$ 

```

In Alg. 6, the sampling points of each period are allocated. Instead of the linear interpolation, the indices of a period are used without further calculations. However, this leads to the possibility of different amount of sampling points for each period.

Algorithm 6: Rearrange periodically without interpolation

```

 $X1_{l,k} = 0$ 
for  $l = 1 : \text{length}(\text{IDXZeroShift}) - 1$  do
   $\text{Len} = \text{IDXZeroShift}(l + 1) - \text{IDXZeroShift}(l)$ 
  for  $k = 2 : \text{Len}$  do
     $X1_{l,k} = \text{current}(\text{IDXZeroShift}(l) + (k - 1))$ 

```

The dimension of $X1_{l,k}$ does not have to be identical with $X_{l,k}$ as the number of samples per period are not fixed as long as the sampling rate is not synchronized to the mains frequency. Additionally, the mains frequency is allowed to vary to some extend [29].

6.3 Additional Curve Fitting for FIT-PS

In order to improve the robustness with respect to the noise, the question arises if the determination of the zero-crossing can be improved. The approach presented in Sec. 6.1 used only two values; before and after the detected zero crossing. In case the two samples before and after the zero-crossing are superposed with noise, the estimation of the zero-crossing is error-prone. A more reliable result can be achieved by considering additional sampling points. One possibility is for example the use of curve fitting with multiple samples. The assumption for this approach is that the voltage signal is a sine wave with a constant frequency for the length of one period.

The determination of the shift is implemented as a function 'lsqSinusShift', thus, Alg. 2 is replaced by Alg. 7.

Algorithm 7: Find zero crossing by applying curve fitting

```

 $\text{IDXZeroShift}(l) = \text{lsqSinusShift}(\text{current}(\text{IDXZero}(l) : \text{IDXZero}(l + 1)))$ 

```

The function takes as argument the number of lines of $X_{l,k}$. Each line can be interpreted as

vector and for each line, the shift is estimated and returned. The problem of estimating the shift is in general a nonlinear problem as

$$y(t) = A \cdot \sin(ft + \varphi). \quad (6.1)$$

The frequency can be assumed as constant whereas the phase φ and amplitude A are unknown. Nevertheless, the function

$$y(t) = A \cdot \sin(ft + \varphi) \quad (6.2)$$

can be written as

$$y(t) = A \cdot \sin(ft) \cdot \cos(\varphi) + A \cdot \cos(ft) \cdot \sin(\varphi) \quad (6.3)$$

by applying the angle addition formulas express [145]. Thus, the nonlinear problem of estimating the phase can be converted into a linear problem

$$y(t) = \beta_1 \cdot \sin(ft) + \beta_2 \cdot \cos(ft) \quad (6.4)$$

by introducing $\beta_1 = A \cdot \cos(\varphi)$ and $\beta_2 = A \cdot \sin(\varphi)$. The determination of β_1 and β_2 can be performed easily when the metric minimizing the quadratic error.

A nonlinear problem would require a nonlinear solver whereas a standard microcontroller is not suited for such a task. Besides, a constant runtime cannot be guaranteed for nonlinear problems in general. Furthermore, the convergence is guaranteed for a nonlinear problem.

Alg. 8 describes the function for sine approximation. As shown in Alg. 8, the over-determined linear equation system can be solved by calculating the pseudoinverse ‘ $\text{pinv}(X)$ ’. The multiplication with the pseudoinverse minimizes the quadratic error as the problem can be written in the form $A \cdot \beta = b$. As A is not a quadratic matrix, the underdetermined system can be solved by calculation $\text{pinv}(A) \cdot b = \beta$. For detailed information and calculations, please refer to [93]. The signal **inpS**, for which the zero crossing needs to be calculated, as well as the number of sampling points per period ‘NoSpP’ are passed to the function **lsqSinusShift**. Nevertheless, φ cannot be calculated directly, instead β is determined. Therefore, the phase shift is determined using four-quadrant inverse tangent, abbreviated with ‘ atan2 ’.

Algorithm 8: function of the curve fitting

```

float lsqSinusShift(inpS, NoSpP)
timeInt = 0 : 1/fs : (NoSpP - 1) · 1/fs
X = ones(NoSpP, 2)
X(:, 1) = sin((2 · π) · fg · timeInt)
X(:, 2) = cos((2 · π) · fg · timeInt)
β = pinv(X) · inpS
return  $\frac{\text{atan2}(\beta(2), \beta(1))}{2 \cdot \pi} \cdot \text{NoSpP}$ 

```

Computationally, the pseudoinverse can be extracted out of the ‘for loop’ in Alg. 1 if f_g is assumed as constant. Thus, the pseudoinverse can be precalculated. The benefit is that no effort is required for this task during run-time. The return value of ‘lsqSinusShift’ is in the range of zero to one, similar to ‘intp’, Alg. 4.

The assumption of a constant frequency can be loosened by estimating the net grid frequency. The simplification might be harder to be implemented but the pseudoinverse can be pre-calculated for different net grid frequencies in advance. Otherwise, the pseudoinverse might be even calculated in real-time but might require, at least some, additional computational power.

6.4 Results

To compare the results of the three FIT-PS methods first, a visual comparison is presented in Fig. 6.1. The plot shows the three different FIT-PS approaches in the time domain. While there are no apparent discrepancies between FIT-PS with interpolation and FIT-PS with fitting, there are recurring distortions in FIT-PS without interpolation which are marked with arrows in the figure. Additionally, the frequency spectrum of the different FIT-PS methods and the original

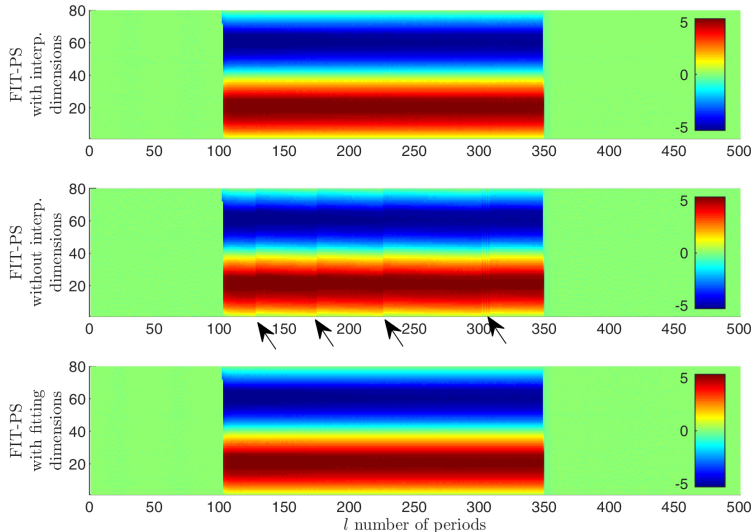


Figure 6.1: Comparison of the three FIT-PS methods in the time domain. The waveforms represent a switching on and off cycle of the heat gun (setting2) from the HELD1 dataset.

signal are analyzed in Fig. 6.2. The energy of the spectrum is indicated with $P(f)$. Please note that $P(f) \neq P(t)$. Therefore, the multidimensional FIT-PS waveform is reshaped into a one-dimensional waveform like the original current signal $i(t)$. In contrast to the originally measured current signal, the sampling rate of the signal converted with FIT-PS always corresponds to a constant integer multiple of the mains frequency.

In Fig. 6.2, the odd-numbered of HARMONICS (HARs) can be identified obviously. The level of HARs of FIT-PS without interpolation is similar than those from the original signal, whereas the level of HARs with FIT-PS and interpolation respectively fitting is significantly higher. Therefore, the waveform which was converted by FIT-PS is for further processing more suitable.

The differences of the signal forms, respectively the implementation methods, can be seen in more detail, at a smaller section of the frequency spectrum. As an example, a section of the fifth HAR at 250 Hz was chosen. The result is given in Fig. 6.3.

Both FIT-PS methods with interpolation and fitting show a significantly improved spectrum. The energy in the frequency domain is concentrated on the frequency of the spectral line of the HAR. This is the result due to adjusting the varying mains frequency with FIT-PS including an interpolation. Therefore, an easier extraction of relevant information, like the HAR is possible. Particularly unfavorable is the FIT-PS method without interpolation. Here, no improvements can be seen compared to the original signal. Furthermore, peaks which occur at a minimal distance to the HARs can be seen in Fig. 6.3 (only for the FIT-PS signal with interpolation and fitting). The reason for this can be traced back to the measurement. The measurement, which was transfer into the frequency domain, is the training scenario of 'Dev001' of the HELD1 [67] dataset, see Fig. 6.4. In this training scenario, only one device was switched on and off. Because the on and off time was always the same, this can be interpreted as a convolution with a rectangular function.

The active time of the appliance was five seconds and the non-active time ten seconds. The

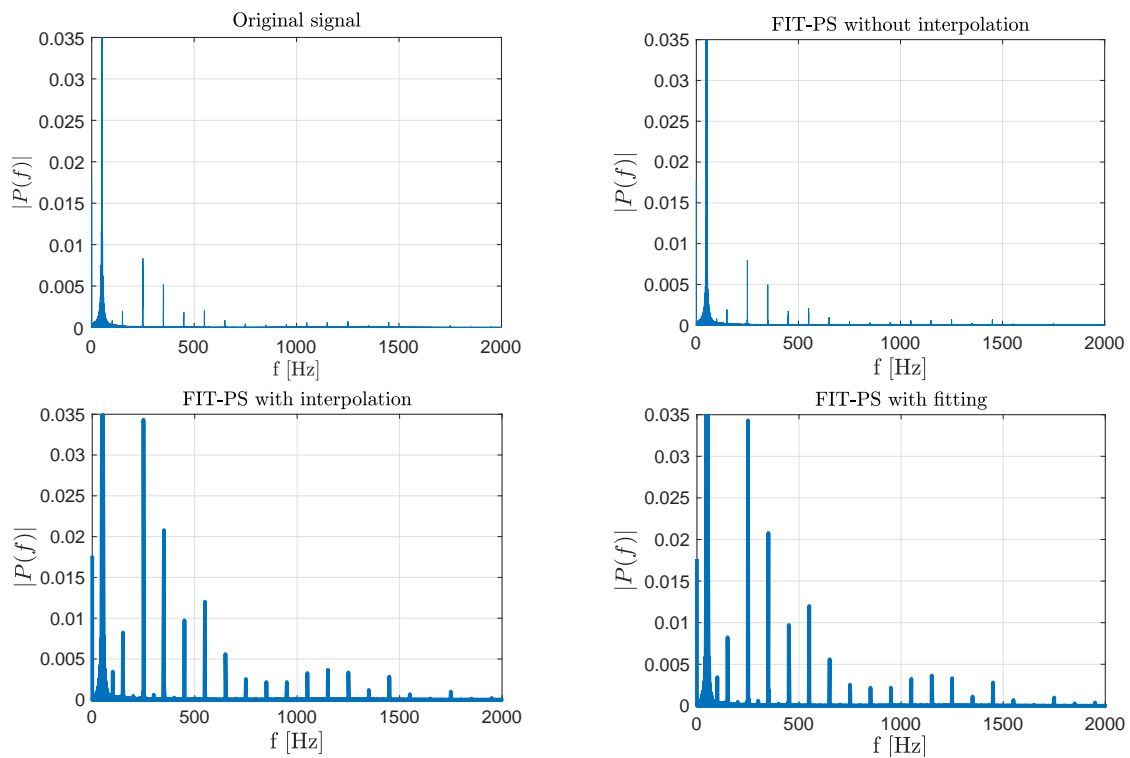


Figure 6.2: Frequency spectrum of the original current signal compared with the different FIT-PS methods (vertical axis clipped at 0.035)

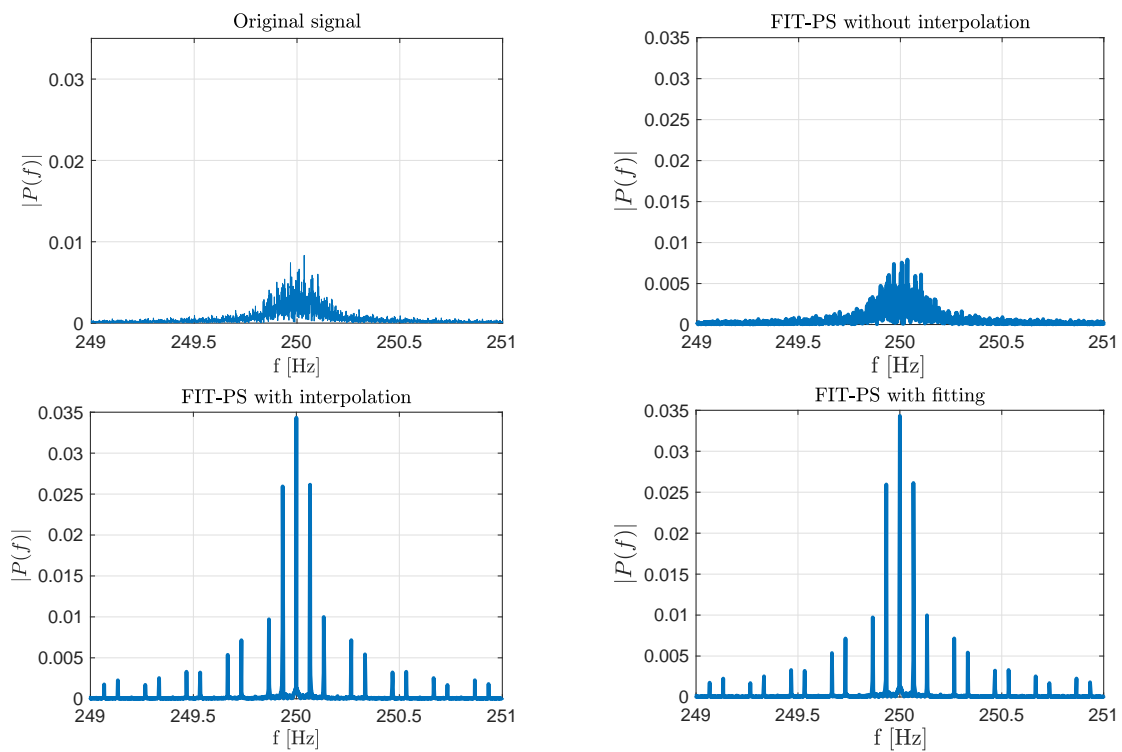


Figure 6.3: Section of the spectrum of the original current signal compared with the different FIT-PS methods

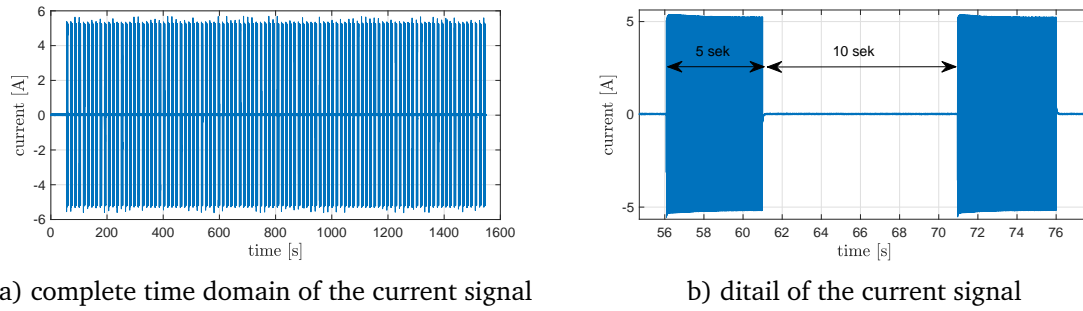


Figure 6.4: Current signal of the measurement which was used for the analysis of the different FIT-PS approaches

distance between the peaks in the frequency spectrum is exact 0.0665 Hz. This corresponds to one period in the time domain $\frac{1}{0.0665 \text{ Hz}} = 15.0340 \text{ s}$. By calculating the Fast Fourier Transform (FFT) over a correspondingly large window, the switching on and off processes can have a significant influence on the frequency spectrum. For appliances which have a periodic on and off behavior like a refrigerator, this can be used as additional information for off-line classification. In case of using a smaller window length for the FFT, which is necessary for classifying shortly after an occurring event, this effect does not appear.

Furthermore, the use of FIT-PS can also have a positive effect on classical signal forms such as the HARs, since the energy is more precisely distributed to the individual frequencies. Therefore, Sec. 6.5 examines whether FIT-PS also has a positive effect on the classic power signals.

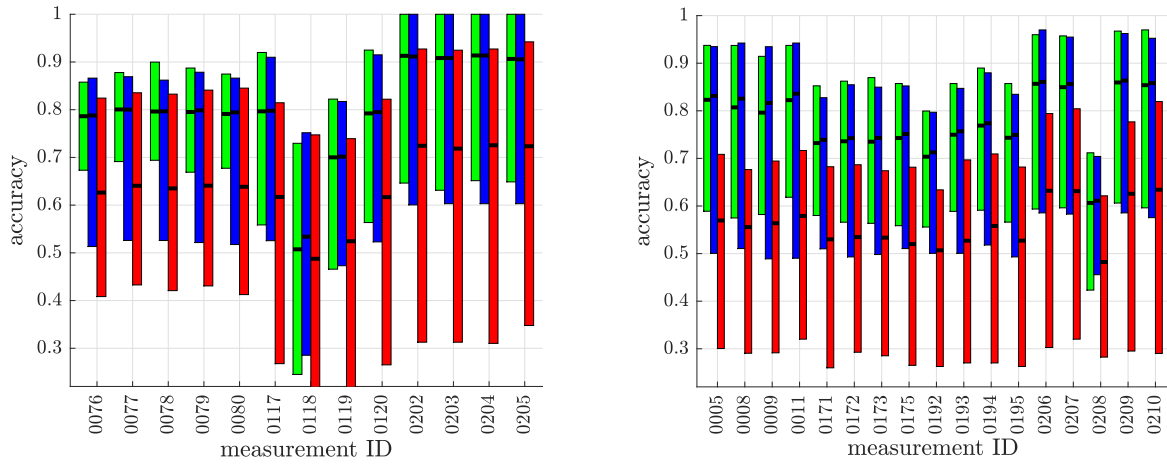
Classification Results of the Different FIT-PS Methods

In order to confirm the previously made visual evaluation in the time and frequency domain, the results of classification with the different FIT-PS methods are presented in this section. The classification is described in detail in Chap. 9. The results of the classification can be understood widely regardless of the Chap. 9 because the results depend on the implementation of the signal form. Since the different waveforms are still very similar, the method of classification is affecting the accuracy of all waveforms in the same way.

The individual results of each implementation method can be seen in Fig. 6.5 a) to Fig. 6.6 b). The different FIT-PS methods are evaluated with the HELD1 dataset [67]. The unsupervised classifier is trained with the training data of the dataset. For classification, a FeedForward Neural Network (FFNN) has been applied. The results of the neural net depend on the initial values because of local minimum in the objective function. Therefore, each neural net was trained 100 times. More details about the classifier and the neural net are described later in this work and can be found in Sec. 9.5.

Fig. 6.5 to 6.6 show the accuracy of the three different FIT-PS methods. FIT-PS with interpolation is shown in green; FIT-PS with interpolation and fitting is shown in blue; FIT-PS without interpolation is shown in red. Each figure show a set of measurements for a specific measurement scenario. In Fig. 6.5 a), only one appliances is switched on at the same time. In Fig. 6.5 b) and Fig. 6.6 a), up to four respectively up to six appliances are active at the same time. Fig. 6.6 b) visualize the over all results of the different scenarios. Different results are achieved by different randomly chosen initial parameters of the FFNN. Therefore, the accuracy varies which is presented with the bar along the achieved accuracy. The accuracy is defined as the number of correctly classified events divided by the total number of existing events Equ. (9.1). The mean accuracy is marked in all rectangles with a black horizontal bar. The 100 different training initializations lead to different results. For each measurement scenario the range of the achieved result is presented by a bar.

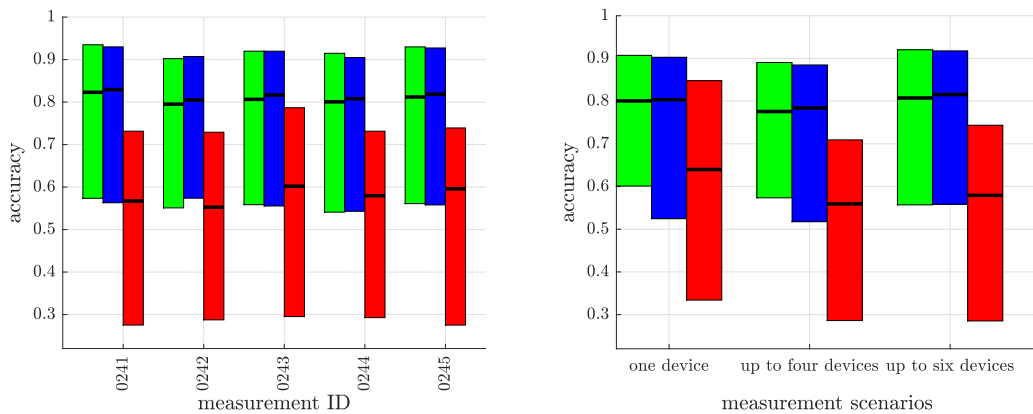
The figures show that the accuracy deteriorate in case of FIT-PS without interpolation (red bar). The achieved minimum, maximum and the mean accuracy of FIT-PS without (red) in-



a) Test scenario one: one device active at the same time.

b) Test scenario two: up to four device active at the same time.

Figure 6.5: Results of the different FIT-PS approaches for one and up to four active appliances. green: FIT-PS with simple zero detection (one value before and one after the zero crossing); blue: FIT-PS with curve fitting for zero detection; red: FIT-PS without interpolation



a) Test scenario three: up to six device active at the same time.

b) Overview of the different scenarios.

Figure 6.6: Results of the different FIT-PS approaches for up to six active appliances and an overview of the three different scenarios. green: FIT-PS with simple zero detection (one value before and one after the zero crossing); blue: FIT-PS with curve fitting for zero detection; red: FIT-PS without interpolation

terpolation is worse than interpolation (green and blue). Comparing the FIT-PS method with interpolation and with fitting and interpolation the, results are much more similar. The minimal accuracy of the fitted signal form is in most cases worse than the signal form using a linear interpolation for the zero detection. One reason for this might be that the shape of voltage signal it not a perfect sinusoidal. This causes problems for the calculation of the zero crossing using the fitting under the assumption that the voltage is sinusoidal.

Fig. 6.6 b) presents the different scenarios of the dataset with a different number of simultaneously active appliances. In case the number of active appliances rises at the same time, the noise in the signal usually increases, too. The comparison of the FIT-PS method with interpolation (green) with the fitting and interpolation (blue) shows that with increasing number of active devices, the implementation with fitting performs slightly better.

6.5 Discussion

Each method implementing FIT-PS offers various advantages and disadvantages. The first implementation in Sec. 6.2 is a combination of the second and the third implementation. The effort to calculate the zero crossing is low since the calculation depends on only two sampling positions. The subsequential re-sampling of all current positions needs more computational power since each sampling point is affected. However, the use of a linear interpolation limits the effort.

The second method of Sec. 6.3, is developed to minimize the calculation effort. In this approach no re-sampling is required, which reduces the computational effort. On the other hand, little variations in the signal can occur over time. The reason is that the sampling frequency is not forced to be integer multiple of the mains frequency. This method does not take care of this circumstance since no interpolation is realized compared to the first method. With a higher sampling frequency, the signal variations are getting smaller. Nevertheless, this variations in the signal is problematic for event detection, where events can be falsely detected but also for the classification since a lower accuracy is achieved with this method compared to the others.

The aim of the second implementation of Sec. 6.1 is to archive a higher accuracy and higher robustness of the transformed signal. Therefore, the zero crossing was calculated by using fitting methods for the voltage signal. Here the advantage is that the resulting zero-crossing position does not only depend on the two sampling positions before and after the zero crossing. The effort of the fitting could be minimized by calculating the necessary parameters offline such to save computational power.

However, the main problem remains as the voltage signal is not a perfect sinusoidal curve. Because of this, an additional error during the calculation of the zero crossing occurs. Therefore, no obvious improvement of the waveform was achieved with this method.

In general, this depends on the sampling frequency. Measurement methods with a lower sampling rate could benefit from the method because the error of the linear approximation of the zero crossing is higher as shown in Fig. 5.2 to 5.4.

Furthermore, in case of more variations in the measurement, the third implementation using the fitting for zero detection can be advantageous. The noise level rises with the number of simultaneously active appliances. In this context, the zero detection method with fitting benefits slight compared to the first implementation without fitting, see Fig. 6.6.

Improving Conventional Power Signal Forms using FIT-PS

FIT-PS can improve the conventional power signals like active power and reactive power, which are very commonly used in NILM context.

Usually, a constant window length is used to calculate active and reactive power. Since the grid frequency is not constantly at 50Hz or 60Hz, additional noise is generated. This effect could be reduced by adapting the sampling rate to a multiple of the mains frequency, realized by the help of FIT-PS. Additionally to the interpolated current signal, the voltage signal needs to be

interpolated too. Apparent power S , real power P , and reactive power Q can be calculated from a FIT-PS converted waveform.

$$S[l] = \frac{1}{K} \sqrt{\sum_{k=1}^K i^2[l, k] \cdot \sum_{k=1}^K v^2[l, k]} \quad (6.5)$$

$$P[l] = \frac{1}{K} \sum_{k=1}^K (i[l, k] \cdot v[l, k]) \quad (6.6)$$

$$Q[l] = \sqrt{S[l]^2 - P[l]^2} \quad (6.7)$$

$$(6.8)$$

The k samples over a period l of length K are used to calculate the power signals S and P .

In case the voltage, after the FIT-PS transformation, is not available, the voltage signal can be estimated. The voltage signal has only small variations from a sinusoidal curve, in contrast to the current signal. The first dimension $v[l, k] \mid k = 1$ is the zero crossing of the voltage. The main disadvantage is that the exact amplitude of the voltage is no longer given since the voltage of the grid is varying, for example in Europe up to $\pm 10\%$. Therefore, at least the Root Mean Square (RMS) voltage $V_{rms}[l]$ must be measured in order to be able to calculate an accurate power calculation. This allows recalculating the voltage for each dimension k of every period l .

$$v[l, k] = V[l] \cdot \sin(2\pi \cdot k/K) \quad \forall k \in \{1, 2, 3, \dots, K\}, \quad l \in \{1, 2, 3, \dots, L\} \quad (6.9)$$

Without the information regarding the voltage, the RMS voltage can be estimated; in general the RMS voltage is 230V. However, the calculated power would then be error-prone because voltage variations cannot be taken into account.

Whether the power is needed at depends on the specific application. In the Ambient Assisted Living (AAL) context, for example, the goal of NILM is to detect the status of the appliances to determine whether the concerned person is in a needy situation. For this application, the exact power consumption of the individual appliances is not the main focus.

In the classical approach of NILM which has the aim to disaggregate the total power consumption into the individual power consumption in order to save energy, the power consumption of the individual appliances is of significant interest.

The focus of most other applications is on the information if an appliance is in a specific state. Therefore, the information on the power consumption is not necessary.

6.6 Conclusion

The implementation of FIT-PS significantly influences the achieved result. With a sampling rate of 4 kHz and a mains frequency of 50 Hz, only a minimal error occurs when the linear interpolation of the zero-crossing is applied. Fitting with the assumption of a sinusoidal function cannot clearly improve the results of the in Chap. 9 discussed classification method. Nevertheless, the result of the investigation is that interpolation is necessary since the achieved classification results without the interpolation are significantly worse. The advantage of the interpolation has been confirmed by the analysis of the frequency spectrum, where the energy could be concentrated to the corresponding HAR.

Chapter 7 | Event Detection

In this chapter, an event detection method for Non-Intrusive Load Monitoring (NILM) is presented. After an introduction, a description of an event detection method, which is optimized for the use of Frequency Invariant Transformation of Periodic Signals (FIT-PS), is presented. Simulated annealing and global pattern search are applied for automatic parameter selection of the event detection method. The results regarding the event detection of the proposed FIT-PS waveform are compared with results from the literature. Some parts of this chapter have been presented at various conferences [66, 69]. For evaluation of the event detection, the Building-Level fully-labeled dataset for Electricity Disaggregation (BLUED) and the Home Equipment Laboratory Dataset 1 (HELD1) dataset are used.

7.1 Introduction

The event detection for NILM is used after the preprocessing and before the feature extraction. The aim of event detection is to determine the device's change of state, for example when an appliance is switched on or off, see Sec. 2.2.3. The detected changes of state, the events, provide the timing for the feature extraction. The position of an event is required for the calculation of Steady State Feature (SSF) as well as for Transient State Feature (TSF).

When knowing the events, the computational cost is less since only features are forwarded to the classification algorithm whereby not the whole data stream needs to be analyzed. The overall computational cost for the NILM process is reduced by the event detection as the decision of whether or not an event has occurred is usually of lower complexity than the classification.

In the amount of data that is recorded over several days, there are usually only a few hundreds switching activities of different devices. For example in the BLUED dataset, which was measured over one week in a private household, 2,418 events occur. This results in about 345 events per day and only 14.4 events per hour. The distribution of the events is not uniform over the time, for example, at night a decreasing number of events is recognized while during the day, when the residents are at home, more events occur. The total BLUED dataset being stored in the NETWORK Common Data Format (NetCDF), has a size of about 230 GB. In one hour, more than one gigabyte of data has to be processed. Applying an event detector, only the positions in the signal are analyzed which are relevant for the subsequent classification. This is advantageous because the calculation effort is reduced since the data-amount which has to be analyzed by the classification is decreased drastically.

Appropriate event detection is of particular importance due to the the linear structure of NILM which leads to error propagation. A higher recognition rate during the event detection improves directly the results of the whole NILM process. Most of the event detection methods used in the NILM context, described in Sec. 2.2.3, use one or a limited number of waveforms. As a result, the complexity of the event detection method and the computational effort is kept low. However, the used waveforms usually do not contain all the information of the original signal, see Chap. 5. Subsequently, an event detection method is presented which is based on the FIT-PS signal representation being described in Chap. 5.3. The use of FIT-PS has the advantage that all information which is relevant for event detection remains in the signal.

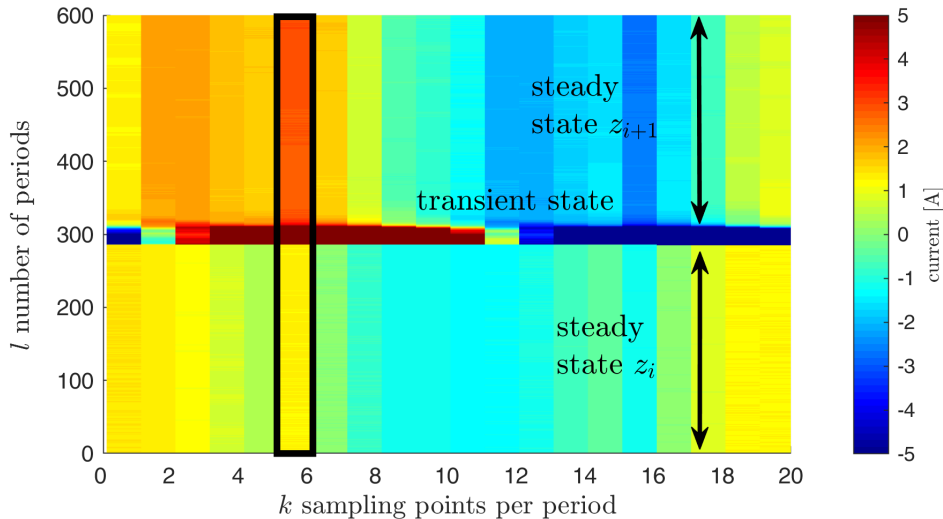


Figure 7.1: Event detection applied on FIT-PS

7.2 Event Detection Using FIT-PS

The event detector is based on the FIT-PS signal representation, see Chap. 5.3. Fig. 7.1 visualize the current signal which has been converted into FIT-PS signal representation. The individual dimensions k over the periods l , of the FIT-PS converted signal, are non-periodical as illustrated in the sixth dimension.

Depending on the available sampling frequency, the visualization of the FIT-PS signal presents multiple details. This affects the development of the event detection algorithm. Depending on whether these details of the signal are present or not, the event detection algorithm must be adjusted. An event detection method for a sampling frequency of 1.2 kHz is presented in the following section.

7.2.1 Event Detection Based on a Sampling Frequency of 1.2 kHz

The utilized BLUED dataset has a sampling frequency of 12 kHz whereas the mains frequency is 60 Hz. In previous publications, often a lower sampling frequency was used due to the limitation of a low-cost embedded system. To achieve better comparability with existing methods, the development of a FIT-PS based event detection method relies on a low frequency as well. With a sampling frequency of 1.2 kHz at a mains frequency of 60 Hz, 20 sampling points per period exist. An example of a resulting FIT-PS signal representation is given in Fig. 7.1.

The event detection method presented consists of two stages. In the first stage, the signal is examined for irregularities. Conspicuous signal sections are further investigated; thus, in a second stage the decision is done. For this purpose, the supposed start and end point of the Transient State (TSt) has to be determined. Thus, the signal can be distinguished in Steady State (SSt) and TSt. Here, two methods are presented to determine the individual state.

7.2.2 First Stage

The original measured current and voltage signals are converted into FIT-PS signal representation. To reduce the influence of disturbances, a low-pass filter, Equ. (7.1), is applied in each dimension $k \in \{1, \dots, K\}$ over each period $l \in \{1, \dots, L\}$ of the signal $i[l, k]$. K and L denote the maximal number of dimensions of the signal in FIT-PS representation. The low-pass filter is used for calculating $i_{L,P}$, see Equ. (7.2), where M_{LP} is the order of the filter and b_g are the filter coefficients.

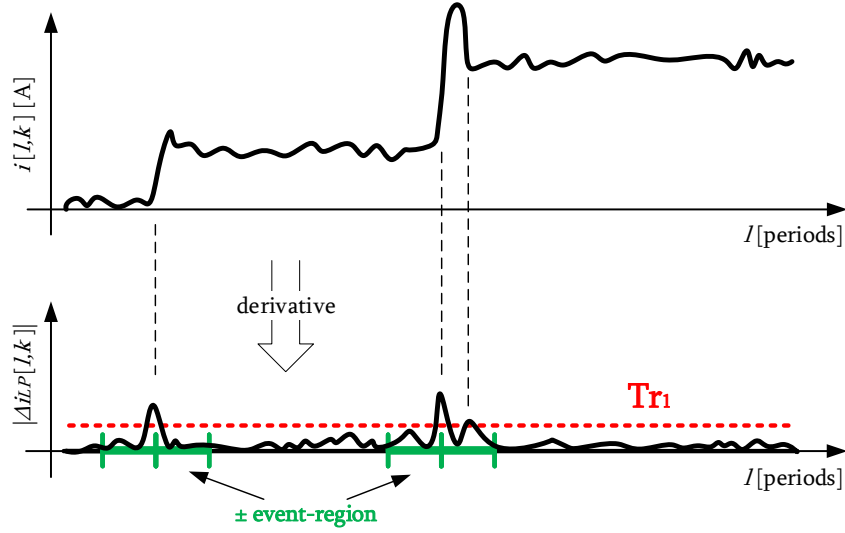


Figure 7.2: First stage of the event detection. The derivative is calculated from each dimension k of the FIT-PS concerted current signal over the periods l .

$$\mathbf{h}_{LP}[l] = \sum_{g=1}^{M_{LP}} b_g \delta[l-g] = \begin{cases} b_g & \text{if } 1 \leq g \leq M_{LP} \\ 0 & \text{else} \end{cases} \quad (7.1)$$

$$\mathbf{i}_{LP}[l, k] = \mathbf{h}_{LP} * \mathbf{i}[\cdot, k] \quad \forall k \in \{1, \dots, K\}, \quad l \in \{1, \dots, L\} \quad (7.2)$$

The aim is to reduce the noise of the current signal converted to FIT-PS, where \mathbf{h}_{LP} is defined as a rectangular window of length M_{LP} .

The event detector is based on a two-step process, to reduce the computational effort and to decrease the complexity. In the first stage, each possible event is detected with a simple threshold which is applied to the derivate of the signal. The derivative is used to eliminate the offset caused by other appliances which are already switched on before an event occurs. Fig. 7.2, visualizes one dimension of the current signal over the time in periods, as well as its derivative. The red dotted line represents the threshold Tr_1 of the first stage. This threshold is adjusted in order to detect all True Positive (TP) events. This leads to higher numbers of False Positive (FP) events which will be handled in the second stage of the event detection. Therefore, a more complex method can be applied for the second stage since only a selection of the total signal, of the possible signal ranges, needs to be analyzed.

The discrete derivative of the low pass filtered signal \mathbf{i}_{LP} is denoted with $\Delta \mathbf{i}_{LP}$ ¹⁰⁾. The forward difference quotient is

$$\begin{aligned} \Delta \mathbf{i}_{LP}[l, k] &= \frac{\mathbf{i}_{LP}[l + \Delta l, k] - \mathbf{i}_{LP}[l, k]}{\Delta l} \\ &= \mathbf{i}_{LP}[l + 1, k] - \mathbf{i}_{LP}[l, k] \quad \forall k \in \{1, \dots, K\}, \quad l \in \{1, \dots, L\}. \end{aligned} \quad (7.3)$$

The partial derivative is calculated of the signal \mathbf{i}_{LP} with respect to l for all k dimensions. Since $\Delta l = 1$, \mathbf{i}_{LP} is calculated by the difference of $\mathbf{i}_{LP}(l + 1, k)$ and $\mathbf{i}_{LP}[l, k]$. For simplification, the derived signal is represented by $\Delta \mathbf{i}_{LP}$.

The maximum of $\Delta \mathbf{i}_{LP}$ is determined for each period l

$$\Delta \mathbf{i}_{max}[l] = \max_{k \in \{1, \dots, K\}} \{ \Delta \mathbf{i}_{LP}[l, k] \}. \quad (7.4)$$

¹⁰⁾not to be confused with the Laplace operator

Equ. (7.4) yields a vector in l direction and contains the the maximum value for each individual column k . The advantage of using the maximum compared to the average is that devices using a phase-fired controller can be detected more efficiently. Those devices often use only a part of the period. By using the maximum of the current $i[l, k]$ a dimension reduction is performed which do not consider dimensions without relevant information. The signal quality is improved since noise from signal contents without relevant information is not considered.

Another property of building the derivative is the elimination of the offset. Therefore, a more straightforward definition of the threshold, for the recognition of the event, can be defined. The aim is to find all indices of $\Delta i_{max}[l]$ over a fixed threshold Tr_1 where the total number of indices over the threshold is N . $\#x$ symbolizes the number of the set x .

$$\Delta i_{max}[l] > Tr_1 \quad \forall l \in \{1, \dots, L\} \quad (7.5)$$

$$N = \#\left\{l \mid \Delta i_{max}[l] > Tr_1\right\} \quad (7.6)$$

The index vector Φ_{temp} is determined by the constraints from Equ. (7.5).

$$\Phi_{temp}[n] = l \mid \Delta i_{max}[l] > Tr_1. \quad (7.7)$$

Φ_{temp} describes the location of a possible event after the first stage of the event detection. The position l is stored in $\Phi_{temp}[n]$, where n starts with $n = 1$ and is incremented by one, with each new entry.

$$1 < n \leq N \quad (7.8)$$

The boundary condition in Equ. (7.5) describes that a potential event is detected when Δi_{max} is above the threshold Tr_1 . All indices of possible events are listed in Φ_{temp} . The information with respect to the maximum value is described in $D[n]$.

$$D[n] = \operatorname{argmax}_{k \in \{1, \dots, K\}} \{\Delta i_{LP}[\Phi_{temp}[n], k]\} \quad \forall n \in \{1, \dots, N\} \quad (7.9)$$

The problem in event detection is the suppression of multiple detections of the same event while maintaining the ability to differentiate between events which are located very close to each other. To avoid multiple detections of the same event, an ‘event-region’ is defined. Only the first event in the event region is used. An example for such an event-region is presented in Fig. 7.2.

The derivative of the indices is calculated. As presented in Equ. (7.3), Δ represents here the partial derivative.

$$\begin{aligned} \Delta \Phi_{temp}[n] &= \frac{\Phi_{temp}[n + \Delta n] - \Phi_{temp}[n]}{\Delta n} \\ &= \Phi_{temp}[n + 1] - \Phi_{temp}[n] \quad \forall n \in \{1, \dots, N\} \end{aligned} \quad (7.10)$$

Since $\Delta n = [n + 1] - n = 1$, Equ. (7.10) follows. $\Delta \Phi_{temp}$ is a vector which contains the distances between all two consecutive indices of Φ_{temp} . In case the distance of the $\Delta \Phi_{temp}$ indices is lower than the event-region, the index is not used further. Therefore, two different indices n and m are used.

$$\Phi_1[m] = \begin{cases} [] & \text{if } \Delta \Phi_{temp}[n] < \text{event-region} \\ \Phi_{temp}[n] & \text{else} \end{cases} \quad \forall n \in \{1, \dots, N\} \quad (7.11)$$

Φ_1 contains all indices of probable events Φ_{temp} with a greater distance than the event-region. Due to the removal of samples, the number of indices for Φ_1 is smaller or equal to the length of Φ_{temp} .

This event detection method excludes the detection of multiple events, close to each other. Depending on the parameter setting of event-region, events which are close to each other can be detected. The window length of event-region is usually only a few microseconds. Events which occur within a shorter distance than event-region are missed. However, turning on different appliances within such a short time window is very unlikely.

7.2.3 Second Stage

The second stage of the event detection investigates the signal sections with a potential event calculated in the first stage. Since not the whole signal needs to be analyzed, a more complex method, compared to the first stage, can be applied. In the second stage, the difference between the stationary level before and after the event is calculated and compared with a second threshold. Therefore, the start and end position of the TSt is required to avoid using a window with a fixed length. An adaptive window-length is advantageous because the length of the TSt depends on the individual appliances. To recognize events which occur close behind each other, the window-length should be as short as possible. Nevertheless, the window-length must be as long as the current TSt of an appliance; otherwise, multiple events are likely to be detected.

For the final decision, the signal is analyzed over a wider window range compared to the first stage. A trade-off between multiple detections of the same event, on the one hand, and the detection of adjacent events, on the other hand, has to be found. A large window length has the disadvantage that events close to each other may be detected falsely as one event. Therefore, a variable window length where no additional event is detected is introduced. In order to minimize the risk of losing events, the start and end point of a window is the same as the start and end point of a transient state.

For this purpose, the signal is decomposed into stationary and transient sections. State Selection V1 and State Selection V2 describe two different methods to find the transition between SSt and TSt and vice-versa.

State Selection V1

The start $\Phi_{1,s}[n]$ and end index $\Phi_{1,e}[n]$ of the transient states are calculated, thus, the resulting window is shorter than the maximum window length G_2 . This case is illustrated in Fig. 7.3. For simplicity, n is now on the index and N the total number of recognized event from the first stage. The distances $\Phi_1[n] - \Phi_{1,s}[n]$ and $\Phi_{1,e}[l] - \Phi_1[l]$, respectively, depend on the specific devices. The vectors w_s and w_e contain the range of the samples in which the start and end sample of the transient state is expected to be found.

$$w_s[n] = [\Phi_1[n] - G_2 + 1, \dots, \Phi_1[n]] \quad \forall n \in \{1, \dots, N\} \quad (7.12)$$

$$w_e[n] = [\Phi_1[n], \dots, \Phi_1[n] + G_2 - 1] \quad \forall n \in \{1, \dots, N\} \quad (7.13)$$

The start and end point of the n -th potential event is determined based on the window $w_s[n]$ or $w_e[n]$ of the FIT-PS converted and low pass filtered current signal i_{LP} . Thereby only the k dimension which is maximal at the derivative Δi_{max} (7.9) is used. The moving mean value $\mu_{1,s}$ is used to calculate the moving variance $\sigma_{1,s}^2$ with the window length N_σ . Equ. (7.14) and (7.15) describe the calculation providing the first sample of the transient, while Equ. (7.16) and (7.17) show the calculation of the end position of the transient. The variance is calculated for each sample $l \in w_s[n]$ respectively $l \in w_e[n]$. The procedure is repeated for each possible event $n \in \{1, \dots, N\}$.

$$\mu_{1,s}[l, n] = \frac{1}{N_\sigma} \sum_{\nu=0}^{N_\sigma-1} i_{LP}[l - \nu, D[n]] \quad \forall l \in w_s[n], \quad n \in \{1, \dots, N\} \quad (7.14)$$

$$\sigma_{1,s}^2[l, n] = \frac{1}{N_\sigma} \sum_{\nu=0}^{N_\sigma-1} \left(i_{LP}[l - \nu, D[n]] - \mu_{1,s}[l, n] \right)^2 \quad \forall l \in w_s[n], \quad n \in \{1, \dots, N\} \quad (7.15)$$

$$\mu_{1,e}[l, n] = \frac{1}{N_\sigma} \sum_{\nu=0}^{N_\sigma-1} i_{LP}[l - \nu, D[n]] \quad \forall l \in w_e[n], \quad n \in \{1, \dots, N\} \quad (7.16)$$

$$\sigma_{1,e}^2[l, n] = \frac{1}{N_\sigma} \sum_{\nu=0}^{N_\sigma-1} \left(i_{LP}[l - \nu, D[n]] - \mu_{1,e}[l, n] \right)^2 \quad \forall l \in w_e[n], \quad n \in \{1, \dots, N\} \quad (7.17)$$

As start and end positions, the samples with the smallest variances are selected. Again, for each possible event $n \in \{1, \dots, N\}$, the smallest variance is determined.

$$\Phi_{1,s}[n] = \operatorname{argmin}_{l \in \mathbf{w}_s[n]} \{\sigma_{1,s}^2[l, n]\} \quad \forall n \in \{1, \dots, N\} \quad (7.18)$$

$$\Phi_{1,e}[n] = \operatorname{argmin}_{l \in \mathbf{w}_e[n]} \{\sigma_{1,e}^2[l, n]\} \quad \forall n \in \{1, \dots, N\} \quad (7.19)$$

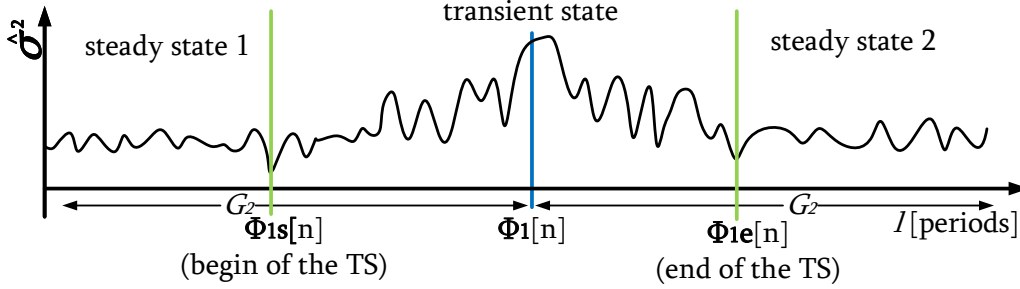


Figure 7.3: State selection V1: Visualizing the calculation of the start and end sample of a transient. The proposed event from stage one is marked blue whereas the calculated start and end positions are marked in green.

This method is described in detail in [66] and [69]. However, the disadvantage of this procedure is the possibility that the transients are too long, due to the noise sensitivity of the scheme. Therefore, the transient range is slightly larger than necessary. To circumvent the limitation, a second method has been developed.

State Selection V2

To minimize the length of the transient area, a thesis [143] was advertised and supervised for carrying out various simulations in this context. An improved detection method for detecting the transition between SSt and TSt is presented. This approach replaces Equ. (7.14) to (7.19). The result of the first stage, as shown in Equ. (7.11), are the indices Φ_1 of the assumed events. For the final decision, the area around these events will be examined further. For this purpose, $\mathbf{w}_g[n]$ for each possible event $n \in \{1, \dots, N\}$, contains the samples around the assumed event.

$$\mathbf{w}_g[n] = (\Phi_1[n] - G_2, \dots, \Phi_1[n] + G_2) \quad \forall n \in \{1, \dots, N\} \quad (7.20)$$

The variance using a sliding window of the length N in l direction is calculated for all k dimensions, where K is the total number of dimensions. Different variances are calculated for each dimension $k \in K$. For each dimension k , the window of length N is slid over each sample contained in $\mathbf{w}_g[n]$.

$$\mu[l, k, n] = \frac{1}{N_v} \sum_{\nu=0}^{N_\sigma-1} \mathbf{i}_{LP}[l - \nu, k] \quad \forall k \in \{1, \dots, K\}, \quad l \in \mathbf{w}_g[n], \quad (7.21)$$

$$n \in \{1, \dots, N\}$$

$$\sigma^2[l, k, n] = \sum_{\nu=0}^{N_\sigma-1} (\mathbf{i}_{LP}[l - \nu, k] - \mu[l, k, n])^2 \quad \forall k \in \{1, \dots, K\}, \quad l \in \mathbf{w}_g[n], \quad (7.22)$$

$$n \in \{1, \dots, N\}$$

For each possible event $n \in \{1, \dots, N\}$, the k th dimension is selected that holds the maximum variance.

$$\hat{\sigma}^2[l, n] = \max_{k \in \{1, \dots, K\}} \{\sigma^2[l, k, n]\} \quad \forall l \in \mathbf{w}_g[n], \quad n \in \{1, \dots, N\} \quad (7.23)$$

Different active appliances cause a different noise level, thus, a fixed threshold does not lead to good results. Hence, a dynamic threshold $\mathbf{Tr}_{TSt}[n]$ is proposed. Due to determining the dynamic threshold, the maximum and the mean of $\hat{\sigma}^2$ over the periods l of the window $w_g[n]$ is calculated. The threshold value $\mathbf{Tr}_{TSt}[n]$ is calculated as

$$\hat{\sigma}^2[n] = \max_{l \in \{1, \dots, L\}} \{\hat{\sigma}^2[l, n]\} \quad \forall n \in \{1, \dots, N\} \quad (7.24)$$

$$\bar{\mu}[n] = \frac{1}{L} \sum_{l=1}^L (\hat{\sigma}^2[l, n]) \quad \forall n \in \{1, \dots, N\} \quad (7.25)$$

$$\mathbf{Tr}_{TSt}[n] = \bar{\mu}[n] + \zeta \cdot (\bar{\mu}[n] - \hat{\sigma}^2[n]) \quad \forall n \in \{1, \dots, N\}. \quad (7.26)$$

ζ has been determined as 0.63 by performing various simulations. The derived adaptive threshold takes into account the samples before an event and updates the threshold based on the mean and variance of the signal. With the threshold ζ , the start and end index of the transients is calculated. As illustrated in Fig. 7.5, the start index of the TSt $\Phi_{2,s}[l]$ is the first index, before the index $\Phi_1[l] - n_1$, where all values from $\Phi_1[l] - n_1 - G_1$ to $\Phi_1[l] - n_1$ of $\hat{\sigma}^2$ are smaller than the threshold $\mathbf{Tr}_{TSt}[n]$. For simplicity, the indices of the window-length G_1 are combined in the vector w_s .

$$w_s[n] = (\Phi_1[n] - G_1 + 1, \dots, \Phi_1[n]) \quad \forall n \in \{1, \dots, N\} \quad (7.27)$$

In Equ. (7.28), the window w_s is shifted by $-n_1$ until the constraint is satisfied and all entries in the window are below the threshold $\mathbf{Tr}_{TSt}[n]$. When the condition is fulfilled, the corresponding index is calculated for $\Phi_{2,s}[n]$, for all n possible events.

$$\begin{aligned} \Phi_{2,s}[n] = \Phi_1[n] - n_1 \quad & \text{if} \quad \max_{l \in w_s[n]} \{\hat{\sigma}^2[l - n_1]\} < \mathbf{Tr}_{TSt}[n] \quad (7.28) \\ & \forall n_1 \in \{0, \dots, G_2 - G_1 - 1\}, \\ & n \in \{1, \dots, N\} \end{aligned}$$

The end of the transient state $\Phi_{2,e}[l]$ is determined equivalent to the calculation of the start of the transient $\Phi_{2,s}[l]$.

$$w_e[n] = [\Phi_1[n], \dots, \Phi_1[n] + G_1 - 1] \quad (7.29)$$

$$\begin{aligned} \Phi_{2,e}[n] = \Phi_1[n] + n_2 \quad & \text{if} \quad \max_{l \in w_e[n]} \{\hat{\sigma}^2[l + n_2]\} < \mathbf{Tr}_{TSt}[n] \quad (7.30) \\ & \forall n_2 \in \{0, \dots, G_2 - G_1 - 1\}, \\ & n \in \{1, \dots, N\} \end{aligned}$$

Fig. 7.4 and 7.5 describe the method graphically. The threshold $\mathbf{Tr}_{TSt}[n]$ is represented by a red line. The area of the windows $w_s[n]$ and $w_e[n]$ are boxed in yellow. Fig. 7.4 present the start of the method. Here, the constraint is not fulfilled since several values of $\hat{\sigma}^2$ in both windows are above the threshold. Fig. 7.4 illustrates the situation that the constraint of Equ. (7.28) and Equ. (7.30) is fulfilled because all indices of the windows have a value below the threshold. The start and the end position of the TSt are represented in green lines. Therefore, the signal can be separated into SSt and TSt.

The new approach is used for the further calculation since the results of $\Phi_{2,s}[n]$ and $\Phi_{2,e}[n]$ outperform the results generated with $\Phi_{1,s}[n]$ and $\Phi_{1,e}[n]$.

$$\Phi_s[n] = \Phi_{2,s}[n] \quad \forall n \in \{1, \dots, N\} \quad (7.31)$$

$$\Phi_e[n] = \Phi_{2,e}[n] \quad \forall n \in \{1, \dots, N\} \quad (7.32)$$

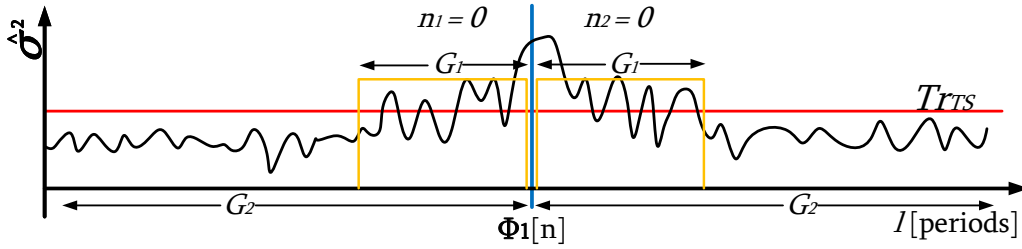


Figure 7.4: State selection V2: Exemplary representation of $\hat{\sigma}^2$, visualizing a transition between a SSts and a TSts. The yellow windows w_s and w_e with the length G_1 start at Φ_1 .

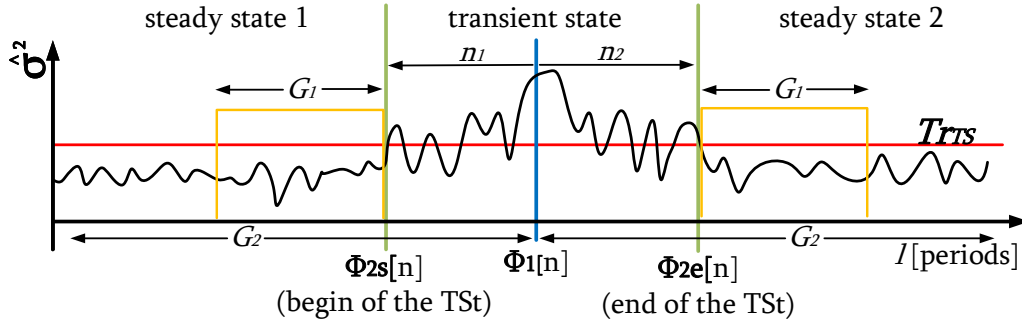


Figure 7.5: State selection V1: Exemplary representation of $\hat{\sigma}^2$, visualizing a transition between a SSt and a TSt. The windows have been moved by n_1 or n_2 values until all values are below the threshold Tr_{TSt} .

Final Decision

Since the exact position of the transient sections are known, the difference τ from the SSt before and after the event can be calculated for all k dimensions, see Equ. (7.33). In order to calculate the difference, the mean value of the signal $i_{LP}[\nu, k]$ in the range of $\Phi_s - M$ to Φ_s is calculated, first. The use of the mean value prevents that peaks caused by noise, influence the result significantly. The number of, FP which was comparatively high after the first stage, is now reduced by using more values of the signal, in order to make the final decision.

$$\tau[n, k] = \frac{1}{M} \left(\sum_{\nu=\Phi_s[n]-(M-1)}^{\Phi_s[n]} i_{LP}[\nu, k] - \sum_{\nu=\Phi_e[n]}^{\Phi_e[n]+(M-1)} i_{LP}[\nu, k] \right) \quad (7.33)$$

$$\forall n \in \{1, \dots, N\}, \quad k \in \{1, \dots, K\}$$

The threshold of stage two is given by a red dotted line in Fig. 7.6. Taking into account the difference calculated in Equ. (7.33), the event detection is finalized by comparing τ to a second threshold T_2 . Therefore, the maximum value resulting from the absolute value of the k dimension from τ is used. The absolute value needs to be calculated to cope with dimensions containing a negative amplitude. Equ. (7.34) denote the final result of the indices of the events E .

$$E[m] = \Phi_1[n] \quad \text{if} \quad \max_{k \in \{1, \dots, K\}} \left\{ |\tau[n, k]| \right\} \geq Tr_2 \quad \forall n \in \{1, \dots, N\} \quad (7.34)$$

The event detection method described in this section was applied to the BLUED dataset by using a sampling frequency of 1.2 kHz. At higher frequencies, an adjustment of the event detection is necessary, which is described in the following subsection.

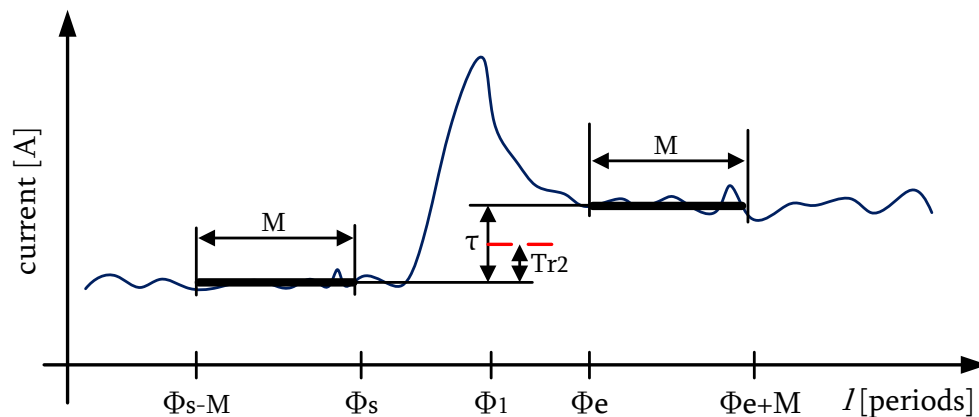


Figure 7.6: Second stage of the event detection. The difference before and after the suspected event is calculated.

7.2.4 Sampling Frequency of 12 kHz

The event detection method developed for 1.2 kHz causes significant deterioration when applied to data having 12 kHz sampling rate. Thus, the number of FPs increases when using a higher sampling frequency.

The deteriorations are caused by small current peaks specific to some appliances. These peaks always occur in the same position of a period but not continuously. When viewing the FIT-PS converted current signal, these current peaks occur only in individual dimensions of the high-sampled signal marked with arrows in Fig. 7.7. Fig. 7.8 is showing the three dimensional plot of Fig. 7.7 where the peaks are presented.

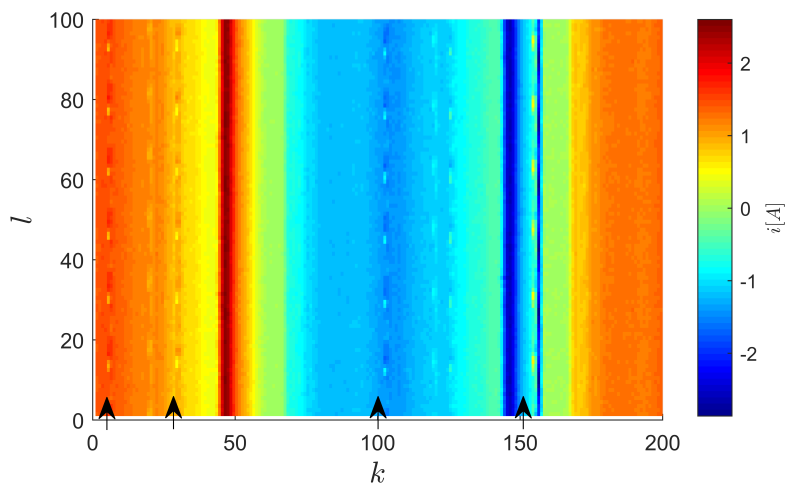


Figure 7.7: Current peaks at a SSt measured with high sampling rate and converted into FIT-PS signal representation. [66]

For signals which are recorded at a lower frequency, for example 1.2 kHz, these peaks cannot be seen because of the low-pass characteristics. Compared to the FIT-PS representation with higher sampling rate, an averaging over several dimensions was realized. In comparison to the low-pass filter from Equ. (7.1), which is applied in l direction of the FIT-PS signal, the low-pass characteristic, caused by a lower sampling rate, works in k direction. Hence, also the number of k dimensions is smaller with a lower sampling frequency.

However, the event detection method based on FIT-PS needs to be modified, if applied to signals with higher sampling rates. Other signal forms like for example active power P do not have

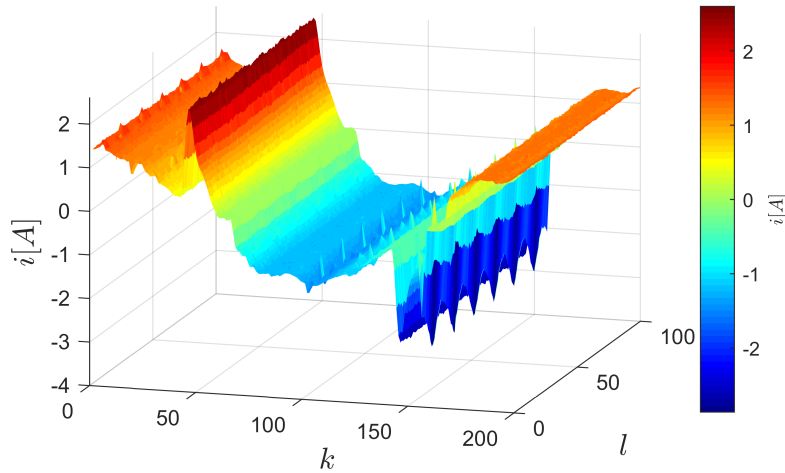


Figure 7.8: 3D view of Fig. 7.7. [66]

this problem because they calculate a mean value over at least one period. The occurring peaks, which are not corresponding to switching on or off events, can be reduced by a low-pass filter. The problem can also be solved by finding dimensions which are affected to a higher noise level; therefore, these are then not taken into account during the event detection process. Since the number of dimensions of the FIT-PS converted signal increases with the sampling rate, a suitable method is required for dimension reduction. Hence, the Principal Component Analysis (PCA) was considered more closely.

7.2.5 Principal Component Analysis for Dimension Reduction

The PCA, also known as Karhunen-Loeve transform, is a technique for dimension reduction, lossy data compression or feature extraction. There are two common definitions of the PCA, the maximum variance formulation and the minimum-error formulation [23, p.561].

“PCA can be defined as the orthogonal projection of the data onto a lower dimensional linear space, known as the principal subspace, such that the variance of the projected data is maximized (Hotelling, 1933). Equivalently, it can be defined as the linear projection that minimizes the average projection cost, defined as the mean square distance between the data points and their projections (Pearson 1901).” [23, p.561]

The PCA itself is well described in the literature, for example in [23, p.561-589].

In the proposed method, the PCA is used to reduce the number of dimensions of a signal in FIT-PS representation which has a high sampling frequency. The PCA gives the k -dimensional projection matrix $i_{PCA}[l, k]$ where $k \ll l$. Most NILM applications do not require real-time capabilities. Therefore, a high number of periods l can be measured before calculating the PCA, such that $k \ll l$ is satisfied. The transformation from the low-pass filtered signal i_{LP} to the new uncorrelated signal i_{PCA} is realized by the multiplication with the weights W .

$$i_{PCA} = i_{LP} \cdot W \quad (7.35)$$

W is a k by k matrix whose columns are the eigenvectors of $i_{LP}^T \cdot i_{LP}$.

For most appliances, there are correlations between the individual dimensions of i_{LP} . Nevertheless, the results achieved by the PCA lead to a higher recognition rate. Because not all k dimensions of i_{LP} contain independent information, the required number of dimensions of i_{PCA} is below the number of dimensions from i_{LP} . Reduction of dimensions is achieved as the PCA weights the dimensions by their variance and only the first κ principal components are

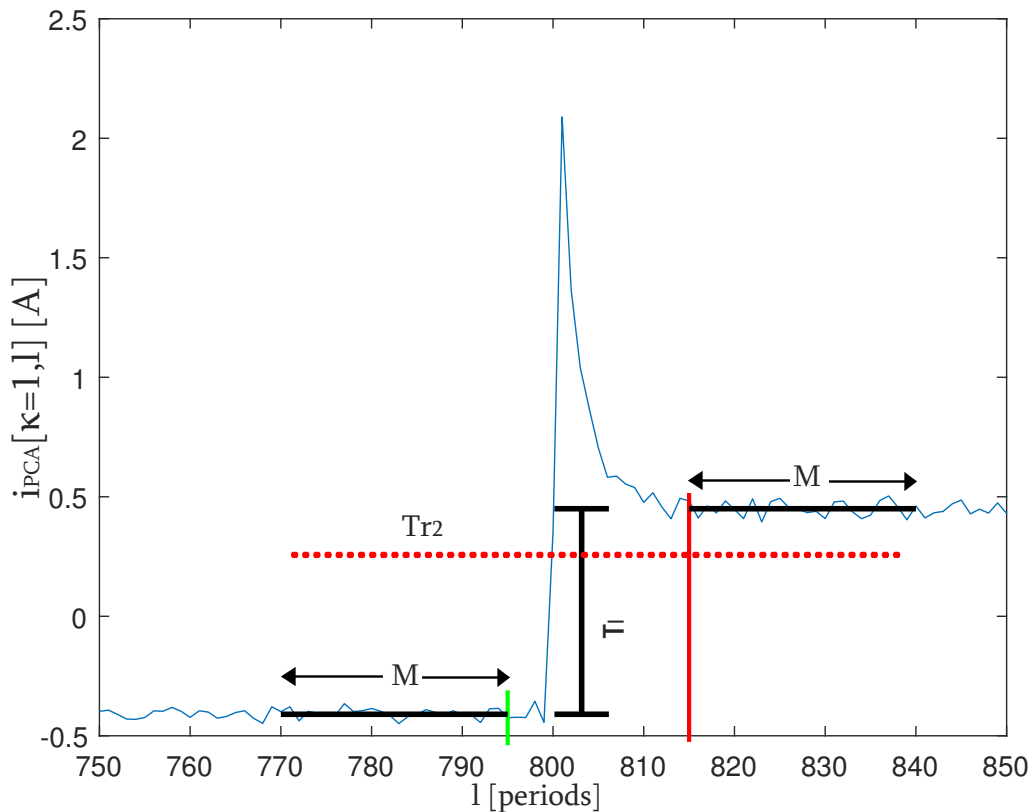


Figure 7.9: Second stage of the event detection. The difference before and after the suspected event is calculated. Modified from [143].

used. By relying only on the κ principal components $i_{PCA\kappa}$, a dimension reduction is achieved. $i_{PCA\kappa}$ contain the uncorrelated information.

$$i_{PCA\kappa} = i_{LP} \cdot \mathbf{W}_{\kappa} \quad (7.36)$$

The signal range, which is used for the calculation of the PCA, must contain events. Otherwise not the dimensions, which contain the information of the event, will be weight highly but the dimensions containing the noise. Hence, a high number of measuring positions is required to ensure that at least one event is included in the measured data block.

Fig. 7.9 presents an event of a real measurement. Here, the second stage of the event detection method is visualized. The different components calculated by the PCA are used for the decision if an event should be recognized. The first component of the PCA are shown in Fig. 7.9.

The event detection method, which is applied on the 12 kHz data, uses the PCA for dimension reduction. Hence, Equ. (7.3), (7.4), and (7.9) of the first stage of the event detection method have to be modified as given in Equ. (7.37) to (7.42).

$$\begin{aligned} \Delta i_{PCA\kappa}[l, k] &= \frac{i_{PCA\kappa}[l + \Delta l, k] - i_{PCA\kappa}[l, k]}{\Delta l} \\ &= i_{PCA\kappa}[l + 1, k] - i_{PCA\kappa}[l, k] \quad \forall k \in \{1, \dots, \kappa\}, \quad l \in \{1, \dots, L\} \end{aligned} \quad (7.37)$$

$$\Delta i_{max}[l] = \max_{k \in \{1, \dots, \kappa\}} \{ \Delta i_{PCA\kappa}[l, k] \} \quad \forall l \in \{1, \dots, L\} \quad (7.38)$$

$$D[n] = \operatorname{argmax}_{k \in \{1, \dots, \kappa\}} \{ \Delta i_{PCA\kappa}[\Phi_{temp}[n], k] \} \quad \forall n \in \{1, \dots, N\} \quad (7.39)$$

In Equ. (7.37) to (7.39), the number of dimensions K is reduced to κ components. Equ. (7.14) to (7.17), respectively Equ. (7.21) and (7.22) are modified by Equ. (7.40) and (7.41), when using the PCA.

$$\boldsymbol{\mu}[l, k, n] = \frac{1}{N_\nu} \sum_{\nu=0}^{N_\sigma-1} \mathbf{i}_{PCA\kappa}[l - \nu, k] \quad \forall k \in \{1, \dots, \kappa\}, \quad l \in \mathbf{w}_g[n], \quad (7.40)$$

$$n \in \{1, \dots, N\}$$

$$\boldsymbol{\sigma}^2[l, k, n] = \sum_{\nu=0}^{N_\sigma-1} (\mathbf{i}_{PCA\kappa}[l - \nu, k] - \boldsymbol{\mu}[l, k])^2 \quad \forall k \in \{1, \dots, \kappa\}, \quad l \in \mathbf{w}_g[n], \quad (7.41)$$

$$n \in \{1, \dots, N\}$$

The final decision of the event detection in Equ. (7.33) is replaced by

$$\tau[n] = \frac{1}{M} \left(\sum_{\nu=\Phi_s[n]-(M-1)}^{\Phi_s[n]} \mathbf{i}_{PCA\kappa}[\nu, \mathbf{D}[n]] - \sum_{\nu=\Phi_e[n]}^{\Phi_e[n]+(M-1)} \mathbf{i}_{PCA\kappa}[\nu, \mathbf{D}[n]] \right) \quad (7.42)$$

$$\forall n \in \{1, \dots, N\}.$$

Since the signal for calculating the PCA must contain events, a large part of the signal needs to be investigated.

The dimensions that are relevant for the event detection are weighted greater. In case the signal does not include events, the PCA weighted the dimensions including the most noise. Because of this limitations, an alternative method using the envelope of the FIT-PS waveform is investigated.

Envelope for Dimension Reduction

Another way of reducing the dimensions is to use the envelope of the FIT-PS waveform. This is illustrated in Fig. 7.10, where the individual dimensions k are plotted for each period l . The envelope of the radio, which has one of the smallest power consumption of the appliances in the HELD1 dataset, is colored red in Fig. 7.10 b). For the calculation of the envelope, the maximum

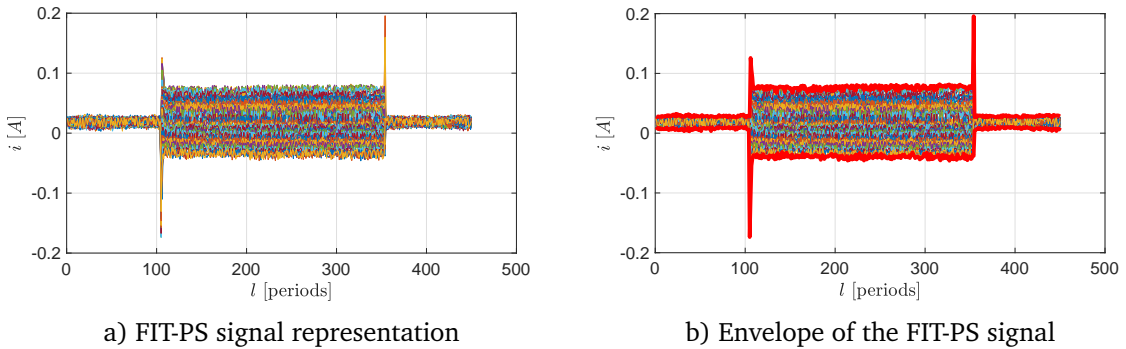


Figure 7.10: FIT-PS signal representation of a radio (ID003) of the HELD1 dataset

and minimum of i_{LP} overall k dimensions is calculated.

$$\mathbf{i}_{env}[l, 1] = \max_{k \in \{1, \dots, K\}} \{i_{LP}[l, k]\} \quad \forall l \in \{1, \dots, L\} \quad (7.43)$$

$$\mathbf{i}_{env}[l, 2] = \min_{k \in \{1, \dots, K\}} \{i_{LP}[l, k]\} \quad \forall l \in \{1, \dots, L\} \quad (7.44)$$

For the event detection, the absolute difference of $\mathbf{i}_{env}[l, 1]$ and $\mathbf{i}_{env}[l, 2]$ is used

$$\mathbf{i}_{envD}[l, 1] = |\mathbf{i}_{env}[l, 2] - \mathbf{i}_{env}[l, 1]| \quad \forall l \in \{1, \dots, L\}. \quad (7.45)$$

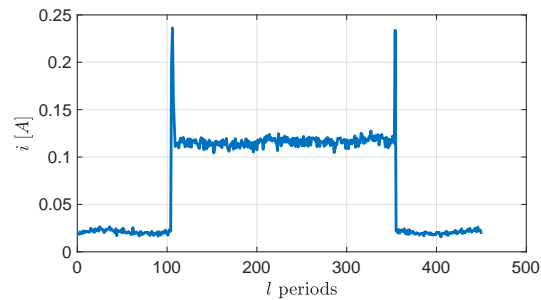
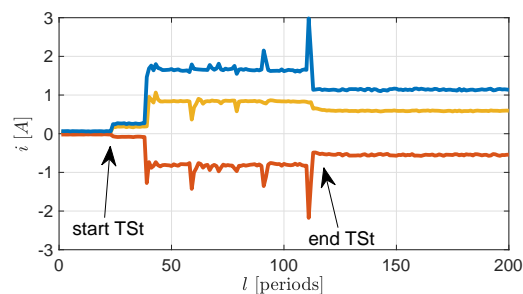


Figure 7.11: Difference of the envelope of Fig. 7.10

The result of the envelope i_{envD} can be used equivalent to the result of the PCA. The result of the enveloped is a one-dimensional vector, as illustrated in Fig. 7.11.

A significant advantage of this method is the reduced noise. Fig. 7.12 present the minimum, maximum, and the resulting difference from the TSt of a fluorescent lamp. During the TSt,

Figure 7.12: Envelope of a fluorescent lamp; maximum of the envelope $i_{env}[l, 1]$ (orange), minimum of the envelope $i_{env}[l, 2]$ (red), absolute difference i_{envD} (blue).

several noise peaks can be seen on the waveforms, especially on the maximum and minimum of the envelope. Calculating the difference of the envelope can reduce the noise. The noise which affects both waveforms in the same direction is compensated.

Different parameters need to be adjusted for a satisfying result of the presented event detection method. Threshold values and window lengths have to be adapted individually for each scenario. Since these parameters depend on each other, a multidimensional optimization problem needs to be solved.

7.3 Parameter Optimization

The presented event detection method has different parameters which need to be adjusted at the same time. Since those parameters depend on each other, the number of possible combinations is too high to test each combination to find the optimum¹¹⁾. Also, a manual parameter selection is difficult and time-consuming. A thesis [143] was advertised and supervised for carrying out various simulations and investigate possibilities for an automatically parameter adaption in this context. Parts of this section have been published in [70]. Parameter selection in the NILM context and especially for the event detection was hardly studied in the past. In most publications, the determination of parameters is not discussed at all. However, the performance is heavy affected by a good choice of parameters.

Algorithms with a low number of parameters have the advantage of being able to adapt to other applications quickly and without complex adjustments. In general, most methods that perform a dynamic or adaptive adjustment of parameters cannot operate without parameters.

¹¹⁾also referred as “brute-force search” in the literature

For example, dynamic detection of the start and end position of the transient was determined in Equ. (7.26). The threshold value is adjusted according to the current noise in the signal. However, an adaptation with a parameter was achieved indirectly by using ζ . Therefore, an automatic parameter selection for event detection in NILM is presented. In computer science, approximation algorithms are used to solve complex optimization problems. Two standard approximation algorithms have been applied to solve the optimization problem for event detection.

Fig. 7.13 and 7.14 show the resulting precision and recall of the two parameters; window-length M and threshold value of the first stage Tr_1 . The second parameter Tr_1 has a higher influence on the precision, as well as the recall.

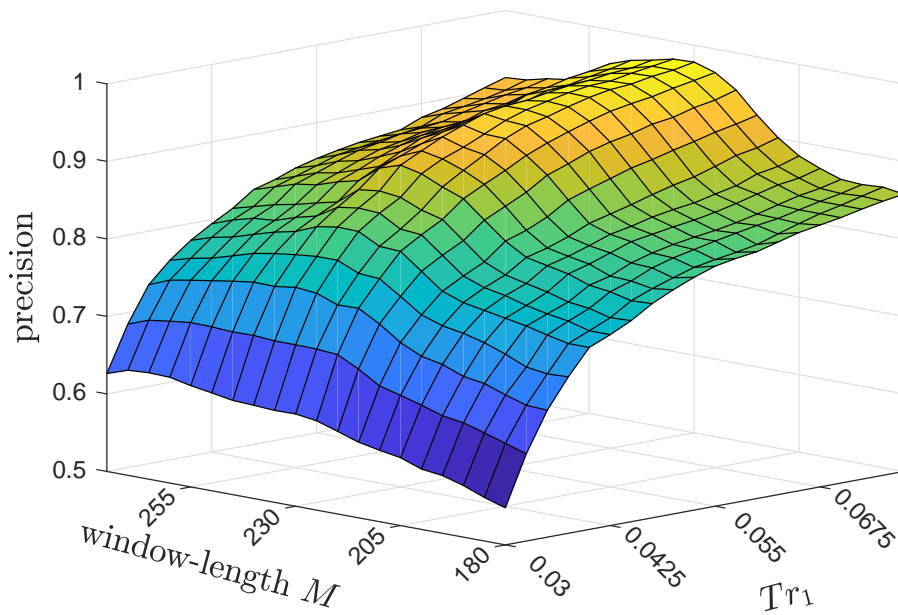


Figure 7.13: Representation of the precision dependency of two parameters of the event detection using the PCA. The results are based on phase A of the BLUED dataset. Modified from [70].

For event detection in the NILM context, there is not only one target value to optimize. The number of False Negative (FN), as well as the number of FP, should tend to zero. Unfortunately, optimizing one target value usually causes the other to increase. For example, in the case of increasing the threshold Tr_1 , the number of FP decreases, but in the same time the number of FN increases. This is illustrated in Fig. 7.13 and 7.14. As a reminder, the relationship between precision, recall, TP, FP, and FN is as follows

$$\text{recall} = \frac{TP}{TP + FN} \quad (7.46)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (7.47)$$

For simplification, the number of FP and FN are equally weighted. Therefore, the result of the detection can be specified with the total error E_{tot} . E_{tot} can be used as a cost function which should be minimized.

$$E_{tot} = FP + FN \quad (7.48)$$

Thus, the aim is to find a parameter setting which minimized the sum of FP and FN . After defining the target value which is to be optimized, the parameters are considered in more detail. The following parameters will be optimized:

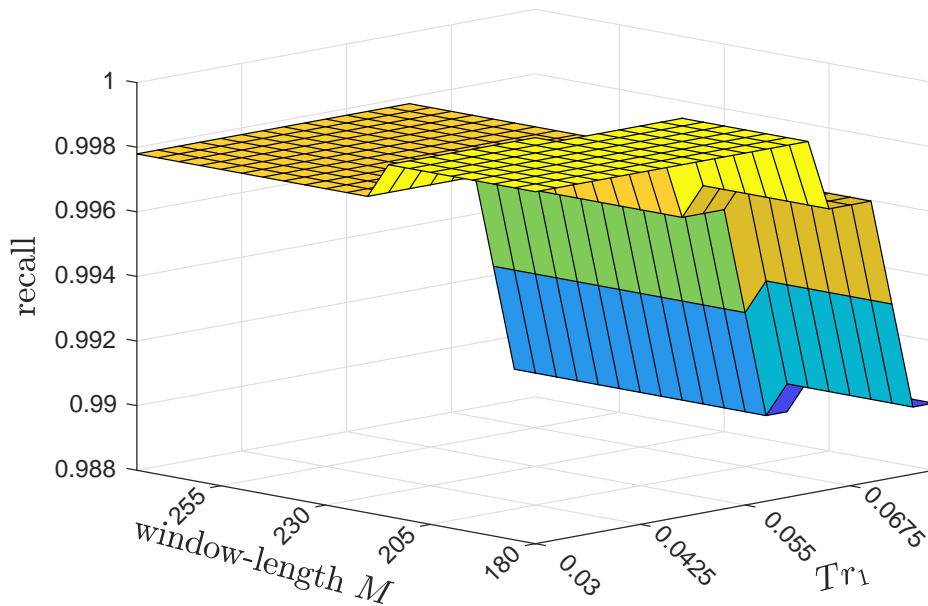


Figure 7.14: Representation of the recall dependency of two parameters of the event detection using the PCA. The results are based on phase A of the BLUED dataset. Modified from [70].

- event-region: only one event is accepted in this region (see Fig. 7.2)
- Tr_1 : threshold of the first stage (see Fig. 7.2)
- transient-region G_2 : surrounding region at the potential event for searching the start and end of the event (see Fig. 7.4)
- Tr_{TSt} : threshold to find the start and end position of the transient (see Fig. 7.5)
- Tr_2 : threshold of the second stage (see Fig. 7.9)
- M : window-length for the comparison of the two SSt (see Fig. 7.9)

Simulated Annealing (SA) and Generalized Pattern Search (GPS) have been chosen for solving the optimization problem. The cost functions are not analytically known precisely. Therefore, fast gradient-based methods may have problems in case of multiple local maximum and minimum. SA can deal with arbitrary cost functions and statistically guarantees to find an optimal solution. Also, GPS does not require a gradient and can be used for optimization problems with multiple extrema.

A short description of both methods is presented in the following two subsections.

7.3.1 Simulated Annealing

SA is an approximation technique to solve complex problems where brute-force-methods are not manageable and therefore cannot be applied. Simulated annealing has been introduced in 1983 by S. Kirkpatrick et al. [86]. The imitation of a cooling process, such as annealing in metallurgy is the fundamental idea of SA. The slow cooling of the heated liquid metal ensures that the atoms have enough time to align and form stable crystals. As a result, a low energy state near the optimum is achieved. Applied to the optimization process, this means that, with a certain probability, even worse results are accepted. The temperature curve of this method corresponds to the probability of accepting worse results. Accepting worse results by some probability enables the algorithm to escape local minimums [86].

The start position of the algorithm is selected randomly. Then, a new position is chosen randomly. In case the new position leads to a better result, the algorithm jumps to this position. If the new parameter setting is worse than the old one, the former setting is chosen, according to the probability of the acceptance \mathcal{P}_A

$$\mathcal{P}_A = \frac{1}{1 + \exp\left(\frac{\Delta D}{T_{temp}}\right)}. \quad (7.49)$$

ΔD defines the difference from the new to the old objective and T_{temp} the current temperature. As the temperature value T_{temp} is a monotonously decreasing function, the probability of accepting a decline decreases over time. The advantage of SA is that not every parameter combination, thus, the total parameter space, has to be tested and still the optimum can be found. This is possible since SA can escape local minimums and thus finds the global optimum.

7.3.2 Generalized Pattern Search

GPS is a method for solving optimization problems without any information about the gradient of the target function. GPS is also known as direct, derivative-free, or black-box search. The idea of GPS is to search around the currently given parameter for a new parameter setting that gives a better result for the defined target function. Therefore, GPS can be applied for example to functions which are non-continuous and/or non-differentiable. Furthermore, when the gradient can hardly be calculated due to numeric problems, GPS can be used. The advantages of GPS are flexibility, robustness, and simplicity.

The main idea of pattern search is to compute the objective function at the mesh points with a certain distance to the current position. The first position, which leads to a better result than the current position, is selected. In case the algorithm can find a better solution, the distance of the mesh is multiplied by two.

If the new parameters do not lead to a better solution, the current position is retained. The distance of the meshed will be multiplied by 0.5.

There are different ways to place the mesh, for example, the so-called poll method which is the standard method and has been applied for the proposed method. For each parameter dimension, the mesh is aligned in the positive and negative direction [73, 142].

Results of optimization problems including many local minimums, can be improved by calculating all-new mesh points around the current position, even if a better result has already been found. Therefore, not the first better position rather the best position around the current location is selected.

The algorithm used in this work has the following termination conditions:

- The mesh size becomes too small.
- The maximum number of iterations has been reached.
- A timeout has been exceeded.
- After a successful iteration, if the distance of the two previously found points and the mesh size are smaller than a predetermined tolerance.
- After a successful iteration, if the change of the result of the two previous iterations, is below a predetermined tolerance.

More details about the different GPS algorithms and implementations can be found in the literature [8, 25, 73, 110, 142, 146, 150].

7.4 Results

In this section, the results of the different investigations are presented. Therefore, the BLUED dataset [4] was especially used since this dataset is suitable for event detection. The data set fulfills various requirements, such as precisely labeled reference data and a consistently high sampling rate (not only in the region of events). The measurement is performed in a real environment and not under laboratory conditions. Thus, the statistical distribution of the events in the dataset corresponds to a real scenario. Additionally, the dataset has been used many times in the literature. This allows to compare the proposed method with existing methods. At last, results of the event detection method using the HELD1 dataset [67] are presented.

7.4.1 Investigation of Different Numbers of Components Resulting from the PCA

High sampling rates of the measurement system lead to a huge number of dimensions in FIT-PS signal representation. The BLUED dataset provides a sampling frequencies of 12 kHz which leads, at a mains frequency of 60 Hz, to 200 dimensions. Thus, the PCA is used for dimension reduction in the context of event detection. The influence of the number of components which is used for the event detection is investigated. Simulations are made using different numbers of components.

In Fig. 7.15, recall and precision are shown by using different numbers of principal components of the PCA. The number of errors in phase A, of the BLUED dataset, is quit small. The results do not differ very much but the best result can be achieved with three components.

The recall of phase B increases from one to two used dimensions. More dimensions do not improve the results for the recall significantly. The precision of phase B is slightly decreasing from using one to two dimensions. Additional dimensions causes deterioration until the fourth component. This indicates that the additional principal components contribute more noise than useful information for the event detection algorithm. Using more than four components, the results for the precision stabilize at a comparatively low level. The best results in case of phase B is achieved when using two components.

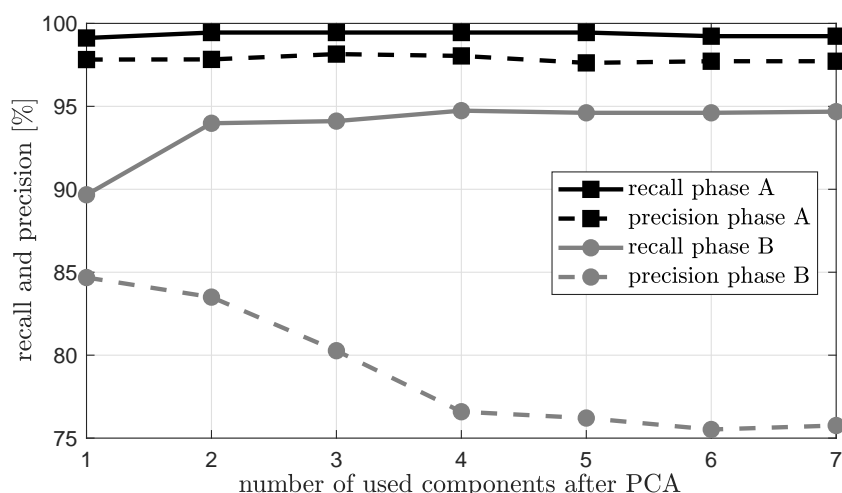


Figure 7.15: Comparison of the results of the event detection with different numbers of dimensions after PCA.

7.4.2 Comparing Different Sampling Rates and Waveforms

In this section, the results achieved with an event detection based on FIT-PS are compared to results achieved with a state of the art method based on waveforms. In order to generalize the result of the comparison, the event detector using the FIT-PS approach is applied to signals

with different sampling frequencies. The compared method uses active power P and a modified generalized likelihood ratio for event detection [5].

At first, the parameters of the proposed event detection method based on FIT-PS signal representation, are adjusted manually. The results of this approach were published in [66] and [69]. In [66], a sample frequency of 1.2 kHz was used compared to 12 kHz in [69].

In Tab. 7.1 and 7.2, the event detection results, based on FIT-PS (sample frequency of 12 kHz, 1.2 kHz and 57 Hz), are presented and compared to the method proposed in literature [5] while the results, achieved with the BLUED dataset, are presented in [4]. The performance of all event detectors depends highly on the considered phase. Phase B event detection is more complicated due to the device types, since some individual appliances have a very low consumption. Furthermore, non-labeled events caused from complex appliances lead to a higher number of FP.

Table 7.1: Detection performance BLUED phase A

BLUED (A)	FIT-PS 12 kHz	FIT-PS 1.2 kHz	FIT-PS 57 Hz	[4]
recall	99.45 %	99.31 %	94.80 %	98.16 %
precision	98.15 %	97.51 %	91.63 %	97.94 %

Table 7.2: Detection performance BLUED phase B

BLUED (B)	FIT-PS 12 kHz	FIT-PS 1.2 kHz	FIT-PS 57 Hz	[4]
recall	93.98 %	87.37 %	78.41 %	70.41 %
precision	83.50 %	82.08 %	78.62 %	87.29 %

The results of FIT-PS with 12 kHz at phase A are superior compared to the method presented in [4] in regards to recall. Precision reaches a similar result as in [4]. The results of FIT-PS with 1.2 kHz are only slightly worse than the results of [4] and FIT-PS with 12 kHz. FIT-PS with 57 Hz, as described in Sec. 5.3.2, produced a higher error rate. The sampling rate is much lower compared to the other methods. The reason for the higher error rate is due to the occurring peaks in l direction, which can be seen in Fig. 7.7 and 7.8. In the high sampled signal, the mean value over a certain window length is taken to reduce the influence of these peaks. With the low sample rate, either the number of measurement points in the window is less, or the length of the window must be increased. Both is disadvantageous for the event detection and leads to a higher number of FP and FN.

In phase B, the recall of FIT-PS outperforms the compared event detector [4] significantly even with the lower sampling rate. The precision of FIT-PS is slightly inferior compared to [4]. By comparison of both FIT-PS event detectors, a reduction in recall and precision is shown for FIT-PS at 1.2 kHz. The reason for this is the highly reduced amount of data and the resulting difficulty of the noise filtering. In general, the results of phase A are significantly better than the results of phase B because phase B has smaller electronic appliances which run at the same time and have switching events close to each other [4].

A comparison of the total results of the BLUED dataset including phase A and B are presented in Tab. 7.3. Additionally, the table includes results from the previous doctoral thesis [87] for comparison. A sampling rate of only 50 Hz was used in this work. Additionally, the method was primarily developed for high power household appliances. The smallest device which was selected for this event detection method has a power consumption of 530 W. For this types of appliances 100 %, of both recall and precision could be achieved. The significantly worse results

Table 7.3: Detection performance BLUED phase A and B

BLUED (A&B)	FIT-PS 12 kHz [69]	FIT-PS 1.2 kHz	FIT-PS 57 Hz	[4]	[87, p. 77]
recall	95.97 %	91.65 %	84.28 %	80.39 %	81.40 %
precision	88.49 %	87.45 %	83.39 %	91.67 %	20.85 %

for the precision can partly be explained by the fact that the parameters were not optimized separately for both phases. This work has focused on identifying potentially dangerous devices, so that the parameters have been optimized for the detection of these devices. The parameters of the proposed event detection method have been selected manually.

7.4.3 Results of Parameter Optimization

The different parameters of the event detection method relying on FIT-PS have to be adjusted. For a different appliance setup, the parameters need to be adjusted again. This task is demanding since the parameter depend on each other and have several local minima. As described in Sec. 7.3, there are too many possible parameter settings to try every combination. Instead, two standard optimization techniques are used. The optimal result, which can be achieved with this event detection method, is unknown. Therefore, a comparison with the results of the two optimization algorithms and the manually selected parameters is realized. Tab. 7.4 and 7.5 present the results of the event detection for the different parameter selection techniques. Since, recall and precision are close to 100 %, the absolute number of FPs and FNs is presented additionally, for a better comparison. The sum of FPs and FNs is denoted as E_{tot} . The BLUED dataset has 906

Table 7.4: Detection performance BLUED Phase A, with optimized parameters

BLUED (A)	SA	GPS	manually
recall	99.78 %	99.12 %	99.45 %
precision	99.45 %	99.78 %	98.15 %
FN	2	8	5
FP	5	2	17
E_{tot}	7	10	22

Table 7.5: Detection performance BLUED Phase B, with optimized parameters

BLUED (B)	SA	GPS	manually
recall	95.44 %	94.99 %	93.98 %
precision	94.42 %	94.87 %	83.50 %
FN	72	79	95
FP	89	81	293
E_{tot}	161	160	388

events on phase A, 1,578 on phase B and in total 2,484 events. The results are achieved with the following parameter setting in Tab. 7.6 and 7.7. As presented, the results of the two optimization methods differ not very much. Furthermore, the proposed parameters of both methods are close to each other as shown in Tab. 7.6 and 7.7.

Table 7.6: Parameter setting of the different methods for phase A

BLUED (A)	SA	GPS
event-region	367	369
Tr_1	0.070	0.072
G_2	84	79
Tr_2	2.75	3.24
M	65	56

Table 7.7: Parameter setting of the different methods for phase B

BLUED (B)	SA	GPS
event-region	443	443
Tr_1	0.053	0.057
G_2	78	80
Tr_2	3.27	3.36
M	119	119

Results of the HELD1 Dataset

The HELD1 is a public dataset which was created especially for the classification task of NILM. Nevertheless, the dataset can be used for event detection. In contrast to the BLUED dataset,

the HELD1 dataset is a laboratory dataset, where many events are near to each other to avoid high measuring time and a high amount of data. Therefore, a reduced possibility of FP exists in the dataset because there is a high chance that a randomly selected index is close enough to an occurring event. The results of the event detection of the HELD1 dataset are presented in Tab. 7.8. As a show case, Tab. 7.8 illustrates the results of three measurement scenarios

Table 7.8: Detection performance HELD1 dataset using FIT-PS waveform

HELD1 scenarios:	0003	0116	0201
recall	100 %	100 %	100 %
precision	99.5 %	97.3 %	99.8 %
FN	0	0	0
FP	2	11	1
E_{tot}	2	11	1

of the HELD1 dataset. Due to a lack of reference results from the literature for this data set, own reference results were calculated, given in Tab. 7.9. Comparing the results from Tab. 7.8

Table 7.9: Detection performance HELD1 dataset using P, Q , and S

HELD1 scenarios:	0003	0116	0201
recall	100 %	100 %	98 %
precision	96.6 %	93.2 %	94 %
FN	0	0	8
FP	14	29	25
E_{tot}	14	29	33

and 7.9, the event detection using the FIT-PS signal representation achieves better results with respect to precision. On all three measurement scenarios, the number of FPs is clearly lower by using the FIT-PS converted signal.

The training scenarios of the HELD1 dataset are used for the parameter optimization with SA and GPS. The HELD1 dataset has a total number of 3,530 training events. Tab. 7.10 presents the results achieved by optimizing the parameters of the event detection applied to the signal in FIT-PS representation. The parameters used to achieve those results are presented in Tab. 7.11.

Table 7.10: Results of the optimization for the HELD1 dataset using FIT-PS (envelope)

HELD1 training:	SA	GPS
recall	99.6 %	99.8 %
precision	99.2 %	99.2 %
FN	14	6
FP	29	28
E_{tot}	43	34

Table 7.11: Parameter corresponding to the results in Tab. 7.10

HELD1 parameter:	SA	GPS
event-region	178	180
Tr_1	0.0404	0.0284
G_2	127	38
Tr_2	0.0472	0.0486
M	6	6

The parameter selection was realized by using SA and GPS in combination with the training measurements of the HELD1 dataset. There are three different test scenarios which differ in the maximal number of simultaneously active appliances (one, up to four, and up to six). In total, there are 13,200 events in the test scenarios available. Fig. 7.16 a) presents the results of the test measurement achieved with the training parameters, optimized with SA. In order to test the robustness of the optimization methods, the optimization was applied ten times with different randomly selected start parameters. Fig. 7.16 b) illustrates the results of the test scenarios with the calculated training parameters optimized for 100 iterations using the GPS. The range of the recall is presented in green, while the precision is colored in blue. The dotted lines present the result of the best training parameter (with the lowest number of E_{tot}). The continuous lines present the median of all test results.

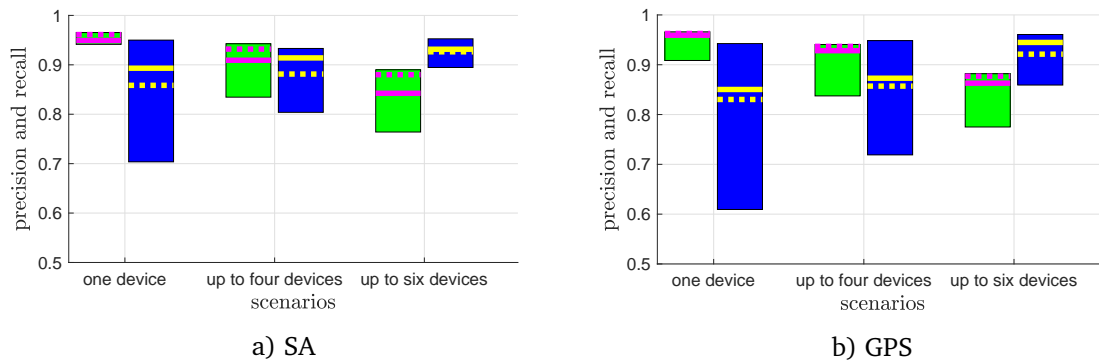


Figure 7.16: Results of the test measurements of the HELD1 dataset achieved with the parameters from 100 training iterations of the GPS or ten training parameters of SA. Recall: green, precision: blue; corresponding result to the best training iteration are marked pink respectively yellow with a dotted line; the median is marked with a solid line.

Both optimization methods achieve a similar result. While the recall slightly get worse with an increasing number of simultaneously active devices, the precision behaves converse. The parameter settings often cannot accommodate recall and precision at the same time. While for recall, the best training result (shown with the dotted line) is always above the median and is close to the best parameter setting, this cannot be achieved for precision. Here, the results of the selected training parameters are always just below the median. The comparison between SA and GPS shows that SA produces more stable results. By calculating several iterations, this can be compensated by GPS.

The results of the test measurement are significant below the results, achieved with a direct optimization of the measurement. The results of parameters generated, with SA and GPS, is very similar, but the variation of the results achieved with SA is lower than those from GPS.

Runtime of SA and GPS

Comparing the runtime of SA and GPS, a significant difference is visible. SA needed 2 h and 48 min compared to GPS which required about 17 min, for solving the optimization problem for the BLUED dataset on phase A. Tab. 7.12 presents the runtime of the two optimization methods. The mean time for the two different approaches to accomplish 100 iterations of training is presented. On phase B of the BLUED dataset, both optimization methods require more time to

Table 7.12: Parameter setting of the different methods for phase B

Dataset	runtime SA [min]	runtime GPS[min]
BLUED phase A	168.11	16.93
BLUED phase B	314.44	31.05
HELD1	177.15	13.37

converge. With 1,578 events, phase B has significantly more events than phase A which has 906 events. The training scenarios of the HELD1 dataset includes 3,530 events. Nevertheless, the time required for both optimization methods is similar to the time of the BLUED dataset on phase A. The events of the HELD1 dataset are closer to each other; thus, the amount of data which needs to be processed is lower compared to the BLUED dataset. On all datasets, GPS is about ten times faster compared with SA.

The optimization has been run on the following system:

i7-6800K CPU @ 3.40 GHz, 64 GB RAM, 64-Bit Windows 10 and Matlab 2018a.

7.5 Discussion

The presented event detection method uses the FIT-PS waveform introduced in Chap. 5. The non-periodic form of the signal simplifies the event detection process since signal changes can be detected more easily. In comparison to the dependence of active and reactive power, which is typically calculated for each period of the mains frequency, the sampling rate in the signal converted with FIT-PS is significantly higher. A higher sampling rate leads to more sampling points per period and thus, to a higher number of dimensions in the FIT-PS signal representation. The higher number of dimensions allows a more detailed signal representation. However, this does not automatically lead to an improvement of the event detection. On the contrary, if the event detection algorithm is not adapted accordingly, the higher sampling rate can lead to even worse results since the higher resolution can contain internal switching processes of the appliances. Internal switching operations of the individual devices can be detected as an event which might be interpreted as a FP. This specific information are not available when using a lower sampling frequencies or waveforms which form an averaging.

However, especially events generated by smaller appliances can be better recognized by a higher sampling rate. Many low power devices have a switching power supply which only affects a small part of the period. Using active P or reactive power Q , the signal is averaged over at least one period. As a result, it is likely that the change in the signal is neglected. In contrary, the FIT-PS waveform allows the analysis of the change of each sampling point of a period. Each dimension is analyzed independently which is advantageous for appliances using a switching power supply. Dimensions without signal change do not influence the evaluation of the event detection.

The event detector can analyze the high dimensional signal directly. Therefore, the event detection proposed in Sec. 7.2 uses the dimension with the highest amplitude deviation over the periods. Additionally, in case of a higher sampling frequency, the PCA is used to reduce the dimensions. The results which are achieved, with the PCA and FIT-PS, are superior compared to methods presented in [4, 87]. Nevertheless, the use of the PCA for dimension reduction has disadvantages. The PCA is based on the variance of the individual dimensions. In case there is no event in the area over which the coefficients of the PCA are calculated, the noise in the signal is increased. Hence, a large window-length for calculating the PCA is used. This ensures that there are events in the part of the signal which are used for the calculation.

The parameter optimization methods improved the results compared to the manually selected parameters. The main advantage of this parameter optimization is that the parameter can be optimized for individual appliances settings. Especially the GPS converge fast to a satisfactory parameter setting. For the BLUED dataset, which contains measurements for one week, only 16 minutes are required for the parameter selection. The use of an optimization method for parameter selection allows adapting the parameters to different measurement scenarios without great effort. However, the requirement is that suitable reference data is available.

In many measurement scenarios, the reference data is not or not fully available or is not reliable and precise enough. Therefore, an event detection method is advantageous which can be trained without the reference data.

For the HELD1 dataset, high accuracy of recall and precision can be achieved by using the FIT-PS waveform. The envelope based on FIT-PS outperforms results based on active, reactive, and apparent power. Parameters are optimized relying on the training data and applied to the test data. Precision and recall of the test data cannot reach the values of the training data. This indicates that the parameter setting depends strongly on the individual measurement scenarios. Further investigations are required in order to develop methods with greater robustness or adaptively adjusting the parameter. Nevertheless, the results achieved with separate training and test data of the HELD1 dataset can be used for further improvements.

In case of poorly chosen parameters, the first stage of the event detection algorithm may find an unusually high number of possible events. Since every possible detected event is investigated

with a more complex second algorithm, the runtime of the detection is likely to increase significantly. Therefore, an abort criterion was introduced during the parameter optimization. The termination criterion is met, if the number of FPs exceeds the number of TPs. Thereby, the time for parameter search is reduced significantly.

The results of the HELD1 dataset illustrate that the results of precision and recall strongly depends on the number of simultaneously active appliances. Therefore, investigations of parameter adjustment, depending on the number of simultaneously active appliances, may improve the event detection. A feedback structure which utilizes the results from classification and tracking can be used for adaptive parameter selection. However, the implementation of a feedback structure is difficult as after the classifier, only FPs but no FNs can be detected. The FNs need to be identified by the tracking. Therefore, the use of a feedback structure requires an implementation of the whole NILM process, including the tracking.

Distinguishing between switching on and off events may increase precision and recall. For most appliances, the TSt of switching on are larger compared to the TSt of switching off events. For a better assignment of the state, distinguishing between on and off events is beneficial. However, this would result in an additional number of parameters, which are also subject to optimization.

7.6 Conclusion

During this work, an event detection method, which uses a two-step decision procedure, was successfully applied with different sampling frequencies from 1.2 kHz up to 12 kHz. An event detection was performed with the proposed FIT-PS waveform and with active power waveforms. The investigation with respect to different sampling frequency leads to the outcome that higher sampling frequencies achieves better results. Different variants of the event detection method have been evaluated. The PCA was investigated to reduce the number of dimensions of the FIT-PS waveform. Additionally, an envelope of the FIT-PS waveform was used at the second stage of the event detection. Using FIT-PS to under-sample the signal down to 57 Hz cannot reach the performance of the higher sampling rate. Compared to the literature, where 50 Hz waveforms have been used for event detection, a significant improvement is achieved by using FIT-PS.

Hypotheses 2: “The new signal representation contains more information than the previously used signal forms and achieves better results in event detection.”; Hypotheses 3: “Waveforms with higher resolution contain more information, therefore, they can produce better results using a suitable event detection technique.” from Chap. 3, can be confirmed since better results can be generated with the new waveform.

By introducing optimization methods for the event detection, on the one hand significantly better results can be achieved, on the other hand, the parameters of the event detection methods can be optimized more efficiently for new measurements. Furthermore, the comparability of different event detection methods or waveforms for event detection can be increased, since the optimized parameters for each method can be calculated more efficiently and more independent.

Chapter 8 | Feature Extraction for Appliance Classification

This chapter describes the feature extraction in the context of Non-Intrusive Load Monitoring (NILM) with focus on the proposed Frequency Invariant Transformation of Periodic Signals (FIT-PS) waveform. There are many ways to generate features from the different waveforms. For example, a complete signal section between switching a device on and off can be used as a feature.

Features using a sampling interval that is slower than one sample per period of the grid voltage can be described as macro level feature. Features based on a faster sampling frequency than the grid frequency can be described as micro level features [101]. The focus of this thesis is on the micro level features since micro level features are more universally applicable compared to macro level features. For better comparability the commonly used Steady State Feature (SSF) and Transient State Feature (TSF) are applied to FIT-PS, and to active power, reactive power in combination with and without HARMONICS (HARs).

The power and Root Mean Square (RMS) values of the current are used to describe the SSF and TSF. A graphical representation of the features for different devices from the Building-Level fully-labeled dataset for Electricity Disaggregation (BLUED) and Home Equipment Laboratory Dataset 1 (HELD1) datasets is shown. The evaluation of the different features is presented in Chap. 9, by the results from the classification.

8.1 Introduction

There are different reasons to use a feature extraction before the classification task in NILM. Depending on the sampling frequency and the used waveform a large amount of data is acquired. In order to reduce the complexity of the classification algorithm, a feature extraction or fingerprint is carried out beforehand. The advantage of another preprocessing technique is the increasing robustness of the classification, since most classification algorithms are lacking the ability to handle the complexity of the whole signal.

In the case of multiple active appliances, the feature extraction is used to eliminate the offset of other active devices. This is necessary because otherwise, a very high number of combinatoric possibilities needs to be handled during the classification. More information about this combinatorial problem can be found in Sec. 2.2.4.

For example Fig. 8.1 illustrates, using the active power waveform, a switching on and off process which can be seen as representative event for many devices. Between the events, there are no significant changes in the signal. Therefore this area in the waveform is defined as a Steady State (SSt). Switching on the device results in a peak. Height and length of this peak depend on the device. For most devices the Transient State (TSt) of the switching on process can be detected between a few milliseconds and a few seconds. Compared to the switching on-process the switching off-process is significantly shorter for most devices. A more detailed consideration of the Transient State Feature (TSF) and SSF follows. In particular the SSF can be calculated by many signal shapes. The following explanation is based first on the example of one dimensional active power.

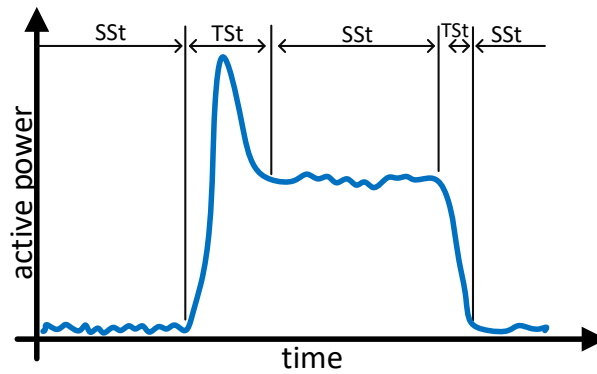


Figure 8.1: SSt and TSt illustrated on a active power waveform.

8.2 Steady State Features

The SSFs are the most common used features for NILM in literature. One reason for this is their applicability even with a low sampling rate. Another reason is that many household appliances have a constant consumption which changes hardly over time. Therefore, features that are formed from the difference between a SSt before and after an event can easily be reproduced.

An exemplary representation, of how SSFs are calculated, is shown by means of the active power in Fig. 8.2. Two devices are switched on and off during the depicted time. The SSF is determined by calculating the difference from before and after the transient area. Calculating the mean value of the SSt before and after the TSt reduces noise in the resulting SSF.

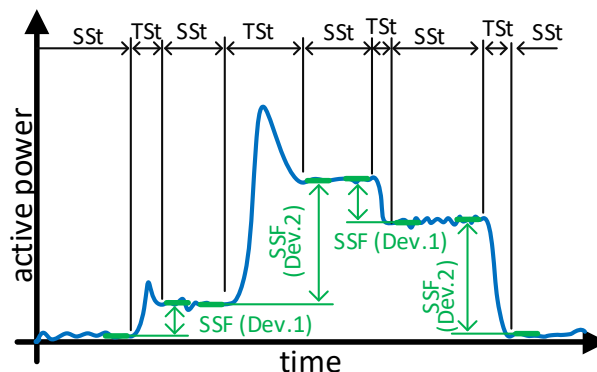


Figure 8.2: Illustration of the calculation of SSF from a power waveform.

A further advantage of SSFs is the possibility to use them with all common signal representations. A prerequisite for the simple differentiation is that the used signal representation does not have any periodicity. For each waveform, one SSFs per event is calculated. Therefore, the resulting information is highly compressed. Hence the classifier can be trained with a small number of data, calculated of different waveforms, only at the relevant positions in the signal when events occur.

Steady State Feature Extraction with FIT-PS

In this section, a steady state feature extraction method for FIT-PS is described. For the FIT-PS waveform the SSFs are calculated for each dimension k . Since the waveform of the individual dimensions is not periodic, the standard feature extraction method from literature can be used. The resulting SSF vector is used for the classification process.

Fig. 8.3 presents the FIT-PS waveform containing $k = 80$ dimensions from a section of the HELD1 dataset. The different dimensions are colored differently. Off events are marked with a

red and on events are marked with a green line. Between the events, no significant changes occur in the signals. The hairdryer, which is switched on first, uses a half-wave rectifier. Therefore, the dimensions of the negative half-wave have a higher amplitude compared to dimensions of the positive half-wave. The heat gun on the first level is the second appliance switched on and reinforces the described effect. The third device switched on is the router, followed by turning off the Light-Emitting Diodes (LED) lamp and turning on the multifunction tool. A significant change in the signal can be seen when the hairdryer and the heat gun are turned off, respectively (at period 2750 and 3050).

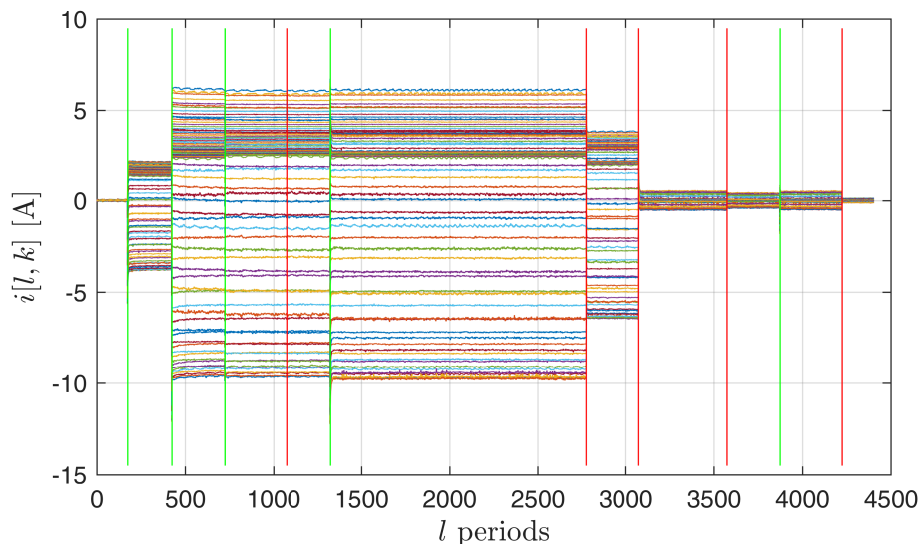


Figure 8.3: Different SSFs of the FIT-PS waveform. The off events are marked red while on events are marked with green line.

The SSF are calculated by the difference of each dimension k from before and after an event. The result is a vector with the length equal to the number of k dimensions. The resulting SSF for the HELD1 dataset are presented in Fig. 8.4, where 99 training features of each appliance are plotted. Each SSF is plotted with a transparency of 20%. In case of multiple overlaid SSF, the lines are darker. The feature can be interpreted as the current over one period concerning the zero crossing of the voltage signal. Therefore, the feature includes the phase-shift between current and voltage. This can be seen, for example in case of the radio, the router or the fluorescent lamp.

For most devices the SSF only has a small variance. Only the two refrigerators and the hair straightener have a small number of significantly different SSF. In case of the hair straightener, the first SSF has a higher current consumption because the device was cold. The refrigerators have a compressor and a pump, which can be active independently. The LED lamp's variations are caused by the noise induced by its very low power consumption.

In case of a half-wave rectifier such as the heat gun (setting 1) or the hair dryer (setting 1) only one half-wave is used. Since the measurement system has a high-pass filter, the measured signal is not correct. However, this does not affect the classification as long as the measurement system is not changed.

Depending on the country, half-wave rectifier must be considered special, since the user may change the phase with the neutral when plugging in. In France, a CEE 7/5 socket is commonly used. This socket does not allow a change of the neutral or phase when connecting the appliances to the same plug. In Germany, the CEE 7/3 socket is used which allows changing the neutral with the phase by rotating the plug 180° .

Hence, the SSF must consider this, especially for hand equipment which is plugged and unplugged frequently. Therefore, a rotation of the plug was simulated for all appliances. This ensures that appliances with an unsymmetrical power consumption can be classified indepen-

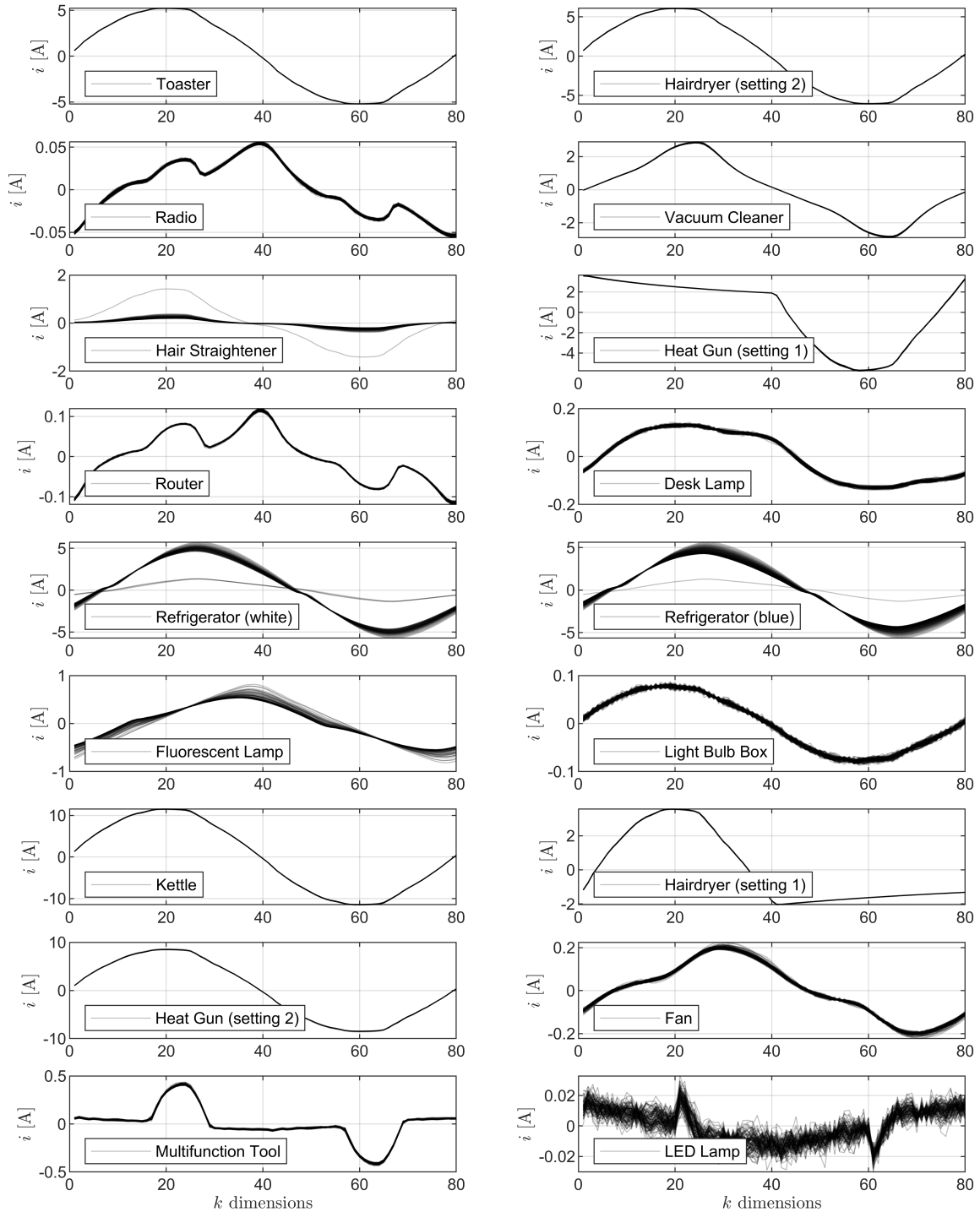


Figure 8.4: FIT-PS SSFs of 18 appliances of the HELD1 dataset. 100 features of each device are plotted behind each other.

dently of the position of the plug. Exemplary this simulated rotation is presented in Fig. 8.5 for the fan. The original measured signal is presented on the left, while the simulated change of neutral and phase is shown on the right.

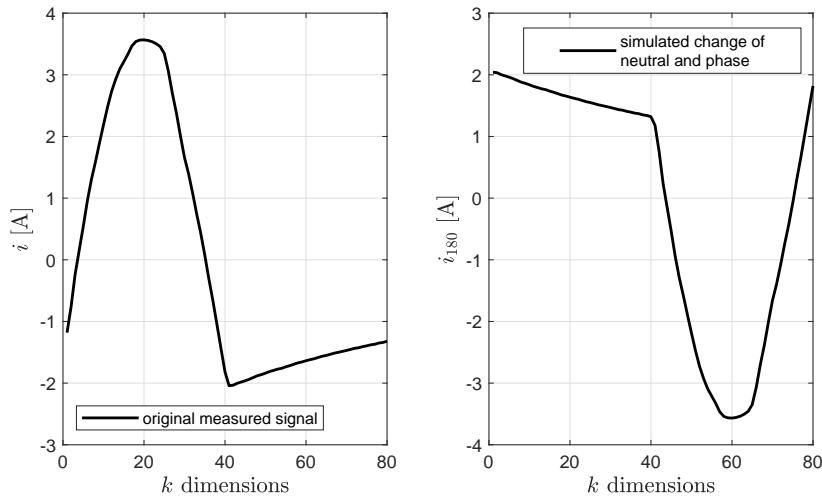


Figure 8.5: Simulation of a 180° change of the plug, of the fan.

FIT-PS allows the simulation of a 180° change of the plug very efficiently. For example, a signal with $K = 80$ dimensions is shifted by half a period, resulting in the first half-wave becomes the second and vice versa. Finally, the new signal is multiplied by minus one.

$$i_{180}(l, \cdot) = \left(i \left[l, \frac{K}{2} + 1 \dots K \right], i \left[l, 1 \dots \frac{K}{2} \right] \right) \cdot (-1) \quad \forall l \in 1 \dots L \quad (8.1)$$

$$i_{180}(l, \cdot) = (i[l, 41 \dots 80], i[l, 1 \dots 40]) \cdot (-1) \quad \Big|_{K=80} \quad \forall l \in 1 \dots L \quad (8.2)$$

Fig. 8.6 presents the variance of the appliances of the HELD1 dataset. For each device the variance is calculated over ten periods of the SSt. Compared to the current signal over one period, the variance of the SSt includes much more noise. Nevertheless, for some applications the variance contains useful information. In this scenario, only single active appliances are used. In case of multiple active appliances, the noise of each active appliances is added to the measured variance of all devices. Therefore, the variance is not used as a feature for the classification.

Additionally, the resulting SSF of selected appliances of the BLUED dataset [4] are presented in the appendix in Fig. A.1. The variance of the individual SSt are illustrated in Fig. A.2 and the median in Fig. A.3.

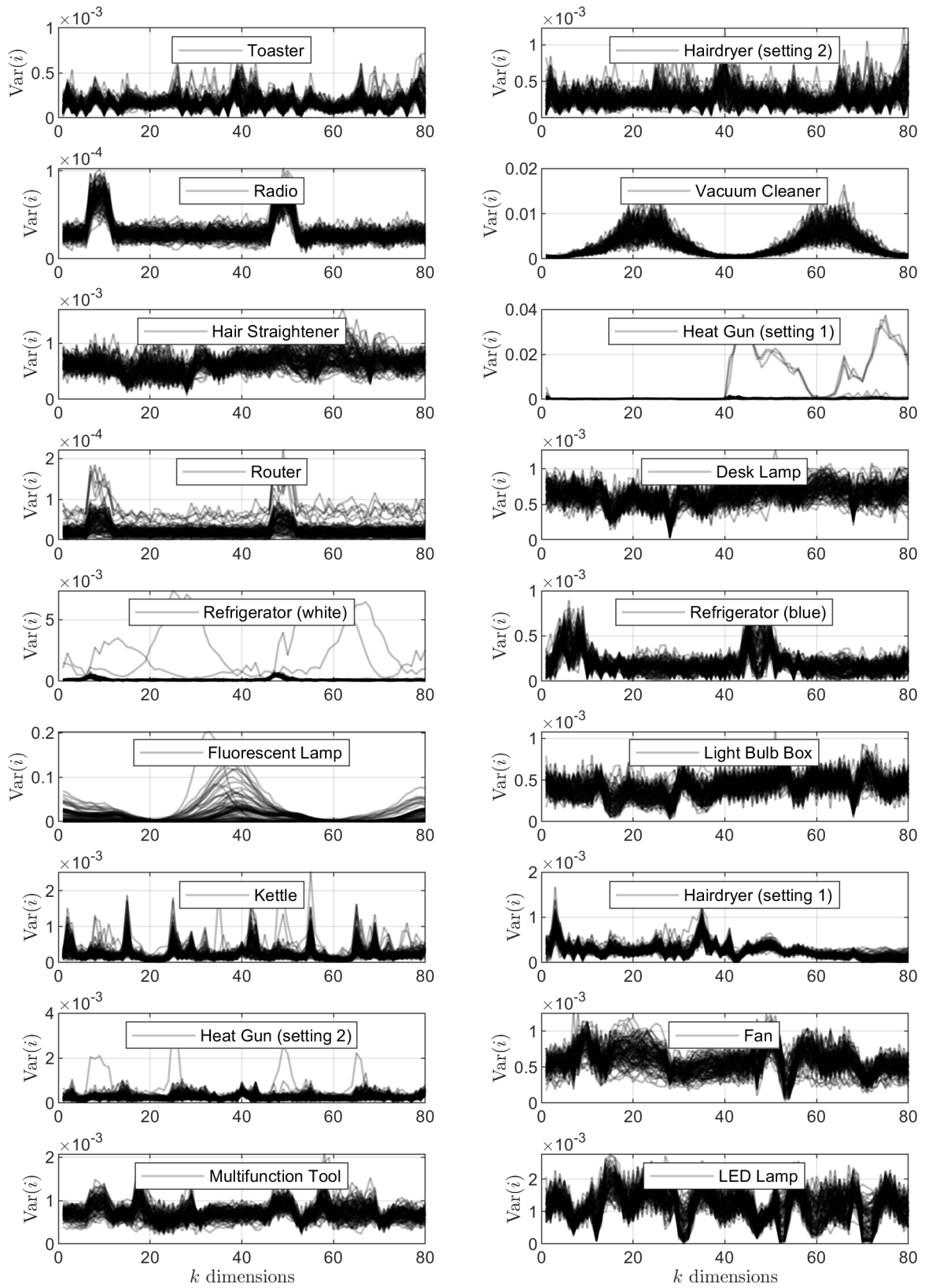


Figure 8.6: Variance of the SSts of the FIT-PS waveform from 18 appliances of the HELD1 dataset.

Mathematical Description of SSF

The features used in the further course of the work are calculated with the waveforms P , Q , the HARs of the current HAR_i , and the FIT-PS waveform $i[l, k]$. Additionally, the event $E[m]$, the window length w_1 , and w_2 are used for calculation. w_1 the distance from the from the recognized event $E[m]$ to the start or the end of the transient while w_2 is the length of the SSt which is used to calculate the mean value.

$$P_{SSt}[m] = \frac{1}{w_2} \sum_{n=1}^{w_2} P[E[m] + w_1 + n] - \frac{1}{w_2} \sum_{n=1}^{w_2} P[E[m] - w_1 - n] \quad (8.3)$$

$$Q_{SSt}[m] = \frac{1}{w_2} \sum_{n=1}^{w_2} Q[E[m] + w_1 + n] - \frac{1}{w_2} \sum_{n=1}^{w_2} Q[E[m] - w_1 - n] \quad (8.4)$$

$$\text{HAR}_{SSt}[m, h] = \frac{1}{w_2} \sum_{n=1}^{w_2} \text{HAR}_i[E[m] + w_1 + n, h] - \frac{1}{w_2} \sum_{n=1}^{w_2} \text{HAR}_i[E[m] - w_1 + n, h] \quad \forall h \in 1, \dots, H \quad (8.5)$$

$$\text{FIT-PS}_{SSt}[m, k] = \frac{1}{w_2} \sum_{n=1}^{w_2} i[E[m] + w_1 + n, k] - \frac{1}{w_2} \sum_{n=1}^{w_2} i[E[m] - w_1 - n, k] \quad \forall k \in 1, \dots, K \quad (8.6)$$

P_{SSt} and Q_{SSt} result in a vector with the length of the number of events. The HARs and the FIT-PS features have a feature vector for each event, thus, they form a two-dimensional matrix, where H describes the number of HARs and K the number of sampling positions per period.

8.3 Transient State Feature

TSFs use the waveform of the transient area during a switching on- or off-event. The prerequisite for this is that the signal form is suitable for the transient range. HAR of the current, for example, cannot be used because the transient region is not periodic. Fig. 8.7 presents the first 70 periods of i_{rms} , exemplary for four appliances. The vacuum cleaner has a high peak after turning on, while the refrigerator has only a very short TSt. The end of the transient in case of the vacuum cleaner and the fan is not clear. The current consumption of the vacuum cleaner decreases exponentially. Also, the current consumption of the fan decreases slightly. An unusual behavior is shown by the the fluorescent lamp. The current consumption increases after a delay of about 20 periods, then the current consumption remains at a relatively constant level before it drops slightly to the next constant level.

In the case of multiple active appliances the signal level of the SSt has to be subtracted to eliminate the offset. Therefore, the SSF have to be calculated, too. This is usually unproblematic and also possible with FIT-PS. The following sub-chapter describes the use of FIT-PS for the TSt.

Transient State Feature Extraction with FIT-PS

FIT-PS has the advantage that the transient region of the FIT-PS signal, received from the converted current signal $i[l, k]$, can be used directly. In case of multiple active appliances, the SSt

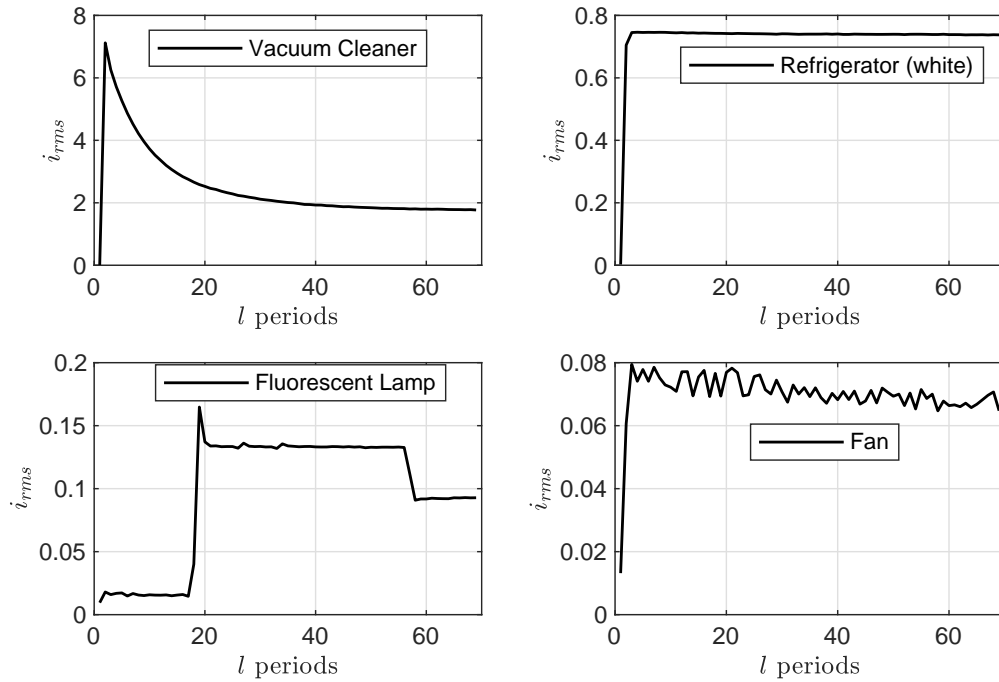


Figure 8.7: TSts of the four appliances of the HELD1 dataset illustrated with i_{RMS} .

of the FIT-PS converted signal needs to be subtracted. To get a better idea of the TSt of FIT-PS Fig. 8.8 shows the first ten periods of each appliance of the HELD1 dataset. The limitation to ten periods is only for visualization reasons. For the classification, 70 periods after the switching event are used. Several appliances show a peak during the switching on-process. Hence, the following lower amplitudes of the signal cannot be displayed in an appropriate way. This is the case for the radio, router, desk lamp, light bulb box, multifunction tool, and the LED lamp. For illustration reasons, each appliance has an individual adapted color scale for the current. Otherwise small consumptions cannot be illustrated. Nevertheless, the amplitude of the current is an important criterion for the distinction of the different appliances.

FIT-PS allows to determine the position where an appliance is switched on or off. Therefore, appliances which have an electronic to control the start process can be classified better. The phase-shift during the TSt as well as during the SSt can be used. The different length of the TSt is also a distinctive feature.

As described for the SSt, appliances with a non-symmetrical power consumption, like the heat gun and the fan (setting 1), have to be considered. This is done for the TSt the same as for the SSt, in Equ. (8.2). This is a significant benefit of FIT-PS since it avoids further training with turned plug.

The TSF of the FIT-PS waveform can be described as

$$\begin{aligned} \mathbf{FIT-PS}_{TSt}[m, k, \cdot] = & \mathbf{i}[\mathbf{E}[m] - w_1, \dots, \mathbf{E}[m] + w_1 + w_2, k] \\ & - \frac{1}{w_2} \sum_{n=1}^{w_2} \mathbf{i}[\mathbf{E}[m] - w_1 - n, k] \quad \forall k \in 1, \dots, K \end{aligned} \quad (8.7)$$

The use of several periods leads to the fact that the FIT-PS feature is shown here as a three-dimensional matrix.

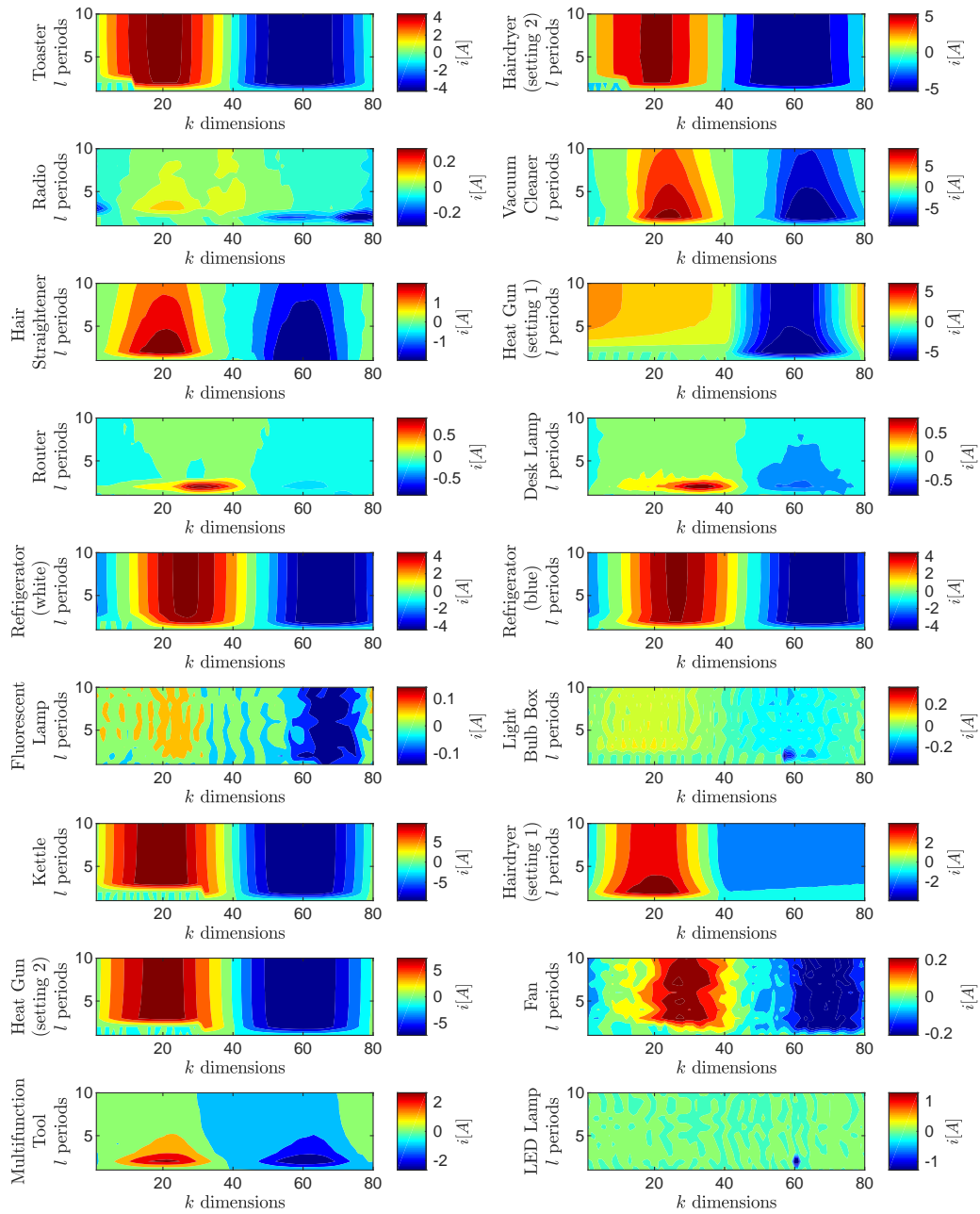


Figure 8.8: Illustration of the TSts with FIT-PS signal representation, for all appliances of the HELD1 dataset.

8.4 Discussion

The feature extraction with FIT-PS can be solved by using the FIT-PS waveform directly. This is the case for SSFs, as well as for TSFs. Hence, no information is lost, while on the other hand, unnecessary information is still in the signal. Therefore, the following classifier must be able to differentiate between necessary and unnecessary information. This more complex classification and training problem procedure is a challenge. Especially in case of the use of TSFs, a larger number of information is transferred to the classifier. For example, the HELD1 dataset the TSFs with FIT-PS represents a matrix with 70 periods \times 80 dimensions. The SSF with FIT-PS is a vector of length 80.

FIT-PS allows to analyze the SSt in more detail since the signal can be analyzed at each position or dimension of the period over the time. This is shown exemplary in Fig. 8.6, where the variance of the different dimensions for all appliances is calculated. The generation of additional information for the classification process can be enabled by this approach. When calculating the variance of each dimension continuously, and not only when an event occurs, the event detection could use the information about the noise in each dimension. By weighting or omitting dimensions, an improvement of classification results could be achieved.

The switching off TSt of most appliances differs significantly from the switching on TSt. Because of this, on and off TSts have to be handled separately. For the SSF it might be possible to use them independently to decide if a device was turned on or off. Indeed the SSt often reveals a small drift. Especially heating equipment like the electric kettle has a decreasing power consumption during the time, because the resistance of the heating coils increases over time. Therefore, also the SSFs are distinguished in on and off features. Unfortunately, this results in a double number of clusters and subsequently in a higher chance of misclassifications. Since on and off events can be distinguished easily, this distinction is executed before the classification. Thus, two different classifications are required, but the number of clusters for each classifier remains the same as the number of devices.

The problem of a feature extraction with a limited number of features is that this provides optimal results for only one scenario. With one additional appliance, different features can lead to better results. Thus, for each scenario a new feature extraction is necessary. With the use of FIT-PS, all relevant information from the waveform remains in the feature. However, the problem addressed is not eliminated, since the following classification algorithm must be re-learned in the case of changed scenarios. This is also necessary if traditional features are used.

More examples of TSFs with FIT-PS are visualized in the appendix from Fig. A.4 to Fig. A.9. To be able to compare the suitability of different features, they have to be tested on a classification procedure. For this reason, the results of the different features, applied on various classifiers, are presented in the following Chap. 9.

8.5 Conclusion

SSF and TSF can be calculated from the FIT-PS waveform without a significantly higher effort compared to other waveforms. The dimensions of FIT-PS are higher compared to the power signal with and without HARs, but no additional calculation effort, as for example for the calculation of the HARs, is required. Because FIT-PS can distinguish between symmetric and unsymmetrical power consumption, a point reflection of the features is realized. In addition to the commonly used SSF, FIT-PS allows the extraction of additional features like the variance of a SSt. Additionally, FIT-PS allows using the transient directly. This is an advantage compared to the HARs for example, which cannot be used since the current signal during the transient is not periodically. Hypotheses 4 from Chap. 3 “The new signal representation which is developed primarily for NILM allows generating features with simple procedures.” can therefore be confirmed.

Chapter 9 | Classification

The aim of classification in Non-Intrusive Load Monitoring (NILM) is the identification of the individual appliances. The information for the classification is generated by different previous steps of the NILM chain. The event detection recognizes signal changes caused by switching on or off operations of individual appliances. This information is used by the feature extraction which collects relevant information from the signal range where an event occurs. Subsequently, the calculated features are used for the classification. The introduction presents an overview of the classification in the context of NILM.

To verify the suitability of the proposed Frequency Invariant Transformation of Periodic Signals (FIT-PS) waveform for NILM, the features extracted from FIT-PS are applied to various classification methods and compared to commonly used power features in combination with HARmonics (HARs).

Different features are applied on standard classifiers, for example k-Nearest Neighbor (kNN), naive Bayes classifiers, Support Vector Machine (SVM), FeedForward Neural Network (FFNN) and Long Short-Term Memory (LSTM) neural network.

The classifiers are trained with the single measurements based on the Home Equipment Laboratory Dataset 1 (HELD1) dataset [67], where each device is switched on and off 100 times. Measurement 0003, 0116 and 0201 are used for validation since they contain most of the appliances used in the HELD1 dataset. Validation data are needed to test the parameter settings of the different classification methods. After a short explanation of the individual classification procedures, the presentation of the results obtained by using different parameters of the validation and test data follows in Sec. 9.7.

9.1 Introduction

The classification in context of NILM is realized by using different kinds of features. Depending on the specific application, different requirements are made on the classification. On the one hand, there are applications, where the result of the classification is time critical. One example for this kind of classification problems is the calculation of the individual power consumption of devices. Here the results are required weekly or monthly. On the other hand, there are applications, where a fast evaluation is required. Applications which need the result of classification immediately after the switching on or off are for example the use of NILM in the Ambient Assisted Living (AAL) context. If the stove is not switched off, this must be detected without using the corresponding switch-off event for classification. For predicting energy consumption, the devices must also be identified immediately after they have been switched on or off.

Therefore, classification methods can be distinguished between classification methods which have a certain real-time capability and those which use features, calculated over a longer time. Classification procedures, that use features generated over a short time, can perform a classification close to the event.

In the case that features can be used over a long time, the classification can theoretically achieve a higher accuracy due to more information of the features. The use of interleaved logical relationships between on and off operations is possible. When a device is turned off, there must exist a turn on event too. Thus, the classification method can make a better decision since both

on and off characteristics are taken into account for the classification. Furthermore, the time between on and off events, as well as the complete waveform between on and off can be considered. In this thesis, classification procedures that are real-time capable are investigated, so a classification result is available soon after the detected event.

The accuracy in context of classification is calculated as follows

$$\text{accuracy} = \frac{\text{number of correctly classified events}}{\text{total number of events}}. \quad (9.1)$$

The focus of this chapter is to investigate the suitability of the FIT-PS feature in comparison with standard power features combined with HARs, with respect to the classification in the context of NILM. Therefore, supervised classification methods have been investigated. Since on or off events can be distinguished easily, two different classifiers, one for on and one for off events, are used. To distinguish between on and off events, the Root Mean Square (RMS) value of the current is calculated from the FIT-PS converted signal before and after the event

$$\text{diff}_{\text{event_temp}} = \sqrt{\frac{1}{K} \sum_{k=1}^K i[l_a, k]^2} - \sqrt{\frac{1}{K} \sum_{k=1}^K i[l_b, k]^2} \quad (9.2)$$

where l_b is the period before and l_a the period after the Transient State (TSt) of the event. For computational reasons, the equation can be simplified to

$$\text{diff}_{\text{event}} = \sum_{k=1}^K i[l_a, k]^2 - \sum_{k=1}^K i[l_b, k]^2 \quad (9.3)$$

For off events, $D_{\text{on/off}} = 0$ and for on events, $D_{\text{on/off}} = 1$.

$$D_{\text{on/off}} = \begin{cases} 0 & \text{if } \text{diff}_{\text{event}} < 0 \\ 1 & \text{if } \text{diff}_{\text{event}} \geq 0 \end{cases} \quad (9.4)$$

The classifiers have to distinguish between 18 different appliances of the HELD1 dataset [67], presented in Tab. 4.1.

9.2 k-Nearest Neighbor Classifier

In the kNN classification, the assignment to a class is performed by considering the k nearest neighbors. k defines how many neighbors are included in the decision. The decision of which class an unknown point in the feature space belongs to is made by the majority membership of the k neighbors. Different distance metrics such as the Euclidean distance or the Manhattan metric can be used. In the simplest case $k = 1$, the classification is solved based on the nearest neighbor. The result of the classification mainly depends on the parameter k . However, the right choice depends on the particular dataset and classification problem. If k is chosen too small, this can lead to misclassification due to the prevalent noise. In the case of a k that is too large, neighbors with a long distance can have too much influence on the decision.

The training of the classification method is achieved by simple storing the training samples. This is also referred as lazy learning while the modeling takes place during the classification process.

Fig. 9.1 illustrates the kNN classifier on a two-dimensional feature-space including two classes. The new measuring point needs to be classified. With $k = 1$, the sampling point is classified to the blue cluster. For $k = 4$, the unknown position is classified to the green cluster since three neighbors belong to the green and only one to the blue cluster. The blue neighbor might be an outlier, so at this point, the assignment to the green cluster seems more appropriate.

The advantage of this classification method is that there is almost no training effort. On the other hand, all training samples need to be stored. For classification problems with a high number of training samples, possibly in high dimensional spaces, this may lead to memory problems. Also, the computational effort during the classification process can become problematic.

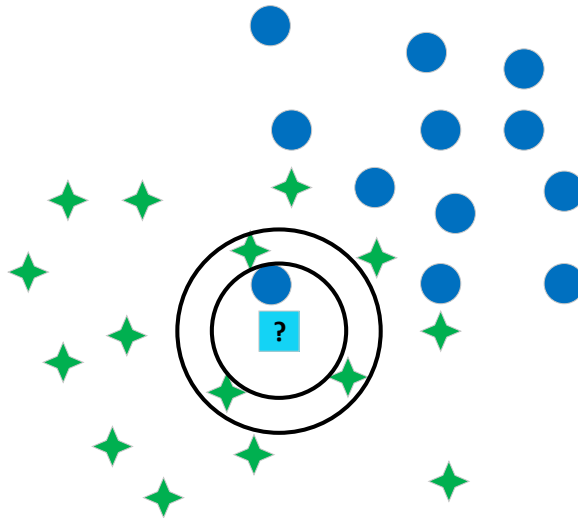


Figure 9.1: Exemplary representation of a kNN with $k = 1$ and $k = 4$ neighbors applied on a two-dimensional classification problem with two classes.

9.3 Naive Bayes Classifier

The Bayes classifier uses statistical characteristics for the classification. Each point in the feature space can be assigned to a class. The assignment of an object to a class is based on the greatest probability. In general, the probability that a point in feature space belongs to a class is highest in the center of a class. In case the point is at the edge of the class, the probability is comparable low. For the calculation of the probability, the probability density function of the existing classes is required. In order to avoid confusion with the active power P , the probability is denoted by the character \mathcal{P} . Usually, an explicit specification of the density function is difficult. For the following explanation, a normally distributed density function is assumed.

$\mathcal{P}(\mathbf{x}|c)$ does not take into account that some classes may occur more often than other classes. Therefore, the problem description is augmented with the a-priori class likelihood $\mathcal{P}(c)$. In practice, the a priori probability from the learning sample may be difficult to estimate if, for example, only a fixed number of training features per device is available. This subsection summarizes the Bayesian classifier written in [71, 306-311].

The decision is based on the knowledge of the feature vector \mathbf{x} . Without this knowledge, the decision could only be made on the basis of $\mathcal{P}(c)$, which is therefore called a-priori probability. With the knowledge of the feature vector \mathbf{x} (a-posteriori), the class which maximizes $\mathcal{P}(c|\mathbf{x})$ will be chosen. According to the decision rule D_{Bay} , described in Equ. (9.5), a classification should be made with the number of classes C .

$$D_{\text{Bay}} = \underset{c=1, \dots, C}{\operatorname{argmax}} \{ \mathcal{P}(c|\mathbf{x}) \} \quad (9.5)$$

The Bayes classifier is based on Bayes's theorem which describes conditional probabilities. For two events A and B, the probability of A provided that B has already occurred can be calculated by the probability of B on the condition that A has occurred, where $\mathcal{P}(B) > 0$ must apply.

$$\mathcal{P}(A|B) = \frac{\mathcal{P}(B|A) \cdot \mathcal{P}(A)}{\mathcal{P}(B)} \quad (9.6)$$

The conditional probability is $\mathcal{P}(A|B)$ of the event A under the condition that B has occurred. The conditional probability is $\mathcal{P}(B|A)$ of the event B under the condition that A has occurred, the a-priori probability $\mathcal{P}(A)$ of the event A and the a-priori probability $\mathcal{P}(B)$ of event B.

Applying this to the classification problem results in

$$\mathcal{P}(c|\mathbf{x}) = \frac{\mathcal{P}(\mathbf{x}|c) \cdot \mathcal{P}(c)}{\mathcal{P}(\mathbf{x})}. \quad (9.7)$$

The denominator can be replaced by

$$\mathcal{P}(\mathbf{x}) = \sum_{c=1}^C \mathcal{P}(\mathbf{x}|c) \cdot \mathcal{P}(c), \quad (9.8)$$

which enables a calculation from the training patterns. The denominator is irrelevant for maximizing the discriminator function since the denominator is class independent. Therefore, the decision rule can be reduced to quantities that can be determined on the basis of training data.

$$D_{\text{Bay}} = \operatorname{argmax}_{c=1, \dots, C} \{d_i(\mathbf{x})\} \quad (9.9)$$

$$d_i(\mathbf{x}) = \mathcal{P}(\mathbf{x}|c) \cdot \mathcal{P}(c) \quad (9.10)$$

$\mathcal{P}(c)$ is the a-priori class probability and corresponds to the probability of how often an individual devices is switched on or off. $\mathcal{P}(\mathbf{x}|c)$ is the conditional probability of the feature vector \mathbf{x} . In case $\mathcal{P}(c)$ is unknown, the class probability must be assumed to be the same.

$$D_{\text{Bay}} = \operatorname{argmax}_{c=1, \dots, C} \{\mathcal{P}(\mathbf{x}|c)\} \quad (9.11)$$

There are several options for selecting a suitable density function. The most widespread density function is the Gaussian or normal distribution. Therefore, an example of a probability densities of the feature vectors \mathbf{x} , represented by a simple normal distribution, is presented. Equ. (9.12) specifies the multidimensional probability density of a normal distribution,

$$\mathcal{P}(\mathbf{x}|c) = \frac{1}{(2\pi)^{N/2} \|\mathbf{K}_c\|^{1/2}} \cdot \exp\left(-\frac{1}{2} \cdot (\mathbf{x} - \boldsymbol{\mu}_c)^T \cdot \mathbf{K}_c^{-1} \cdot (\mathbf{x} - \boldsymbol{\mu}_c)\right), \quad (9.12)$$

whereas $\boldsymbol{\mu}_c$ is the mean value of class c and \mathbf{K}_c is the covariance matrix of class c . For simplification the transformation

$$d_c(\mathbf{x}) = \ln(d_c(\mathbf{x})(2\pi)^{N/2}) \quad (9.13)$$

is used. Applying Equ. (9.12) to Equ. (9.9) and using the simplification $d_c(\mathbf{x})$, results in the distinction function

$$d_c^*(\mathbf{x}) = \ln(\mathcal{P}(m)) - \frac{1}{2} \ln(\|\mathbf{K}_c\|) - \frac{1}{2} \cdot (\mathbf{x} - \boldsymbol{\mu}_c)^T \cdot \mathbf{K}_c^{-1} \cdot (\mathbf{x} - \boldsymbol{\mu}_c). \quad (9.14)$$

The simplification of Equ. (9.12) to (9.14) is not only for simplifying the mathematical notation and reducing the computational effort. The fact that the exponential function does not has to be calculated explicitly increases the numerical stability. The density function can be assigned with higher accuracy to the most critical clusters.

Compared to distance classifiers, the Bayes classifier considers the respective density function of the classes. An optimal classification by the Bayes classifier can be carried out if the density function is known. Therefore, the Bayes classifier is used in combination with simulated classification problems with known distribution. Here, the Bayes classifier achieves an optimal result, thus, other classification methods can be compared. More details about the Bayes classifier can be found in the literature [45] and [71, p. 306-311].

9.4 Support Vector Machine

The idea of the SVM was developed in 1992 by Vapnik et al. [27]. In order to deal with nonlinear classification problems, the data is transferred to a higher dimensional space that theoretically reaches infinite order. The aim is to place a separation plane between two classes. An example is illustrated in Fig. 9.2 where two clusters in the two-dimensional space are separated from

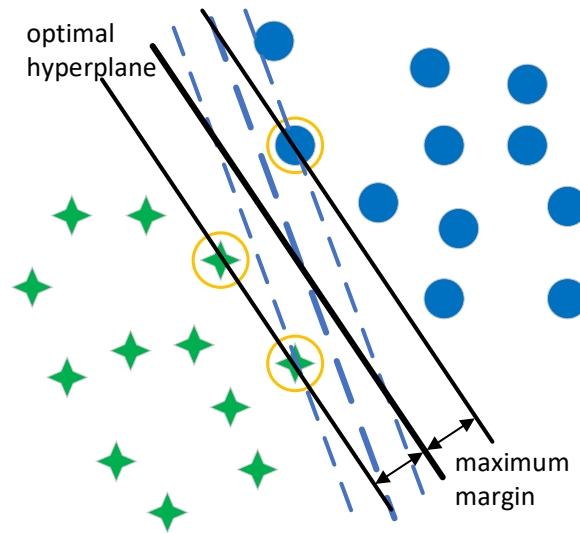


Figure 9.2: Linear separation of two clusters in two dimensions

each other. The thin lines represent the hyperplane closest to the clusters. The boundary margin is the distance from the two outer hyperplanes to the separation plane, where two separation planes are illustrated. The dotted blue line is not optimal since its margin is not maximal. The black continuous lines present the optimal hyperplane, maximizing the distance between the two clusters.

Mathematically the SVM uses a linear separation plane, also called hyperplane, described by the decision function

$$D(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b, \quad (9.15)$$

where \mathbf{x} is the training vector, \mathbf{w} is a vector of the length n and $b \in \mathbb{R}$ is a bias parameter. In the case of a two-class problem, the hyperplane can be defined as

$$\mathbf{w}^T \mathbf{x}_i + b \begin{cases} \geq 1, & \text{for } y_i = 1 \\ \leq -1, & \text{for } y_i = -1. \end{cases} \quad (9.16)$$

Equ. (9.16) can be expressed as follows

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \quad \text{for } i = 1, \dots, n \quad (9.17)$$

$$\frac{y_i D(\mathbf{x}_i)}{\|\mathbf{w}\|} \geq M, \quad (9.18)$$

where $M \in \mathbb{R}$ is the margin. From Equ. (9.18) follows that in order to maximize the margin M , the minimum for \mathbf{w} must be found. The optimal hyperplane can be obtained by minimizing

$$Q(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2, \quad (9.19)$$

with respect to the constraint given in Equ. (9.17). The calculation of the hyperplane does only require those training data for which $y_i D(\mathbf{x}_i) / \|\mathbf{w}\| = M$. They are called support vectors and are marked orange in Fig. 9.2.

Previously, it was assumed that the data could be separated linearly. The area between the two clusters without data points is called hard-margin. In case that an error-free separation of the data is not possible, a slip variable ξ can be introduced. The idea is to allow a certain error of the training data.

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i \quad \text{for } i = 1, \dots, n \quad (9.20)$$

Unlike a hard-margin, a soft-margin, allows a small number of training values to be within the margin or on the other cluster. Fig. 9.3 exemplary presents the use of a soft margin. In practice, a soft margin can be advantageous in case of noise in the training. The influence of outliers in the dataset is reduced by the soft-margin.

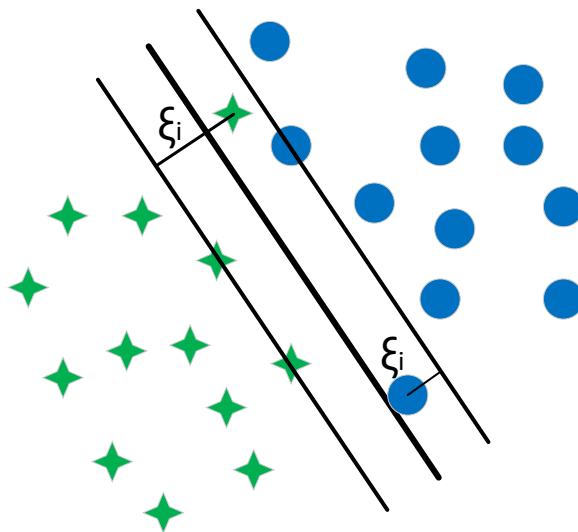


Figure 9.3: Linear separation of two clusters with a slip variable

Linear SVMs have only limited possibilities to separate clusters. To solve nonlinear problems, a kernel is used for the SVM. The data are transformed to a higher-dimensional feature space. The computational effort increases with the number of dimensions. When choosing the number of dimensions, over-fitting should be avoided, since this unnecessarily increases the computational effort without any positive impact on the results. The goal is to transform a non-linear problem into a linearly problem. This is presented in Fig. 9.4 where in a) a nonlinear problem of two classes in a one-dimensional space is illustrated. Fig. 9.4 b) presents the feature-space transformed in a higher dimension by a quadratic kernel. Thus, a linear separation of both classes is possible. The following kernel functions are used commonly:

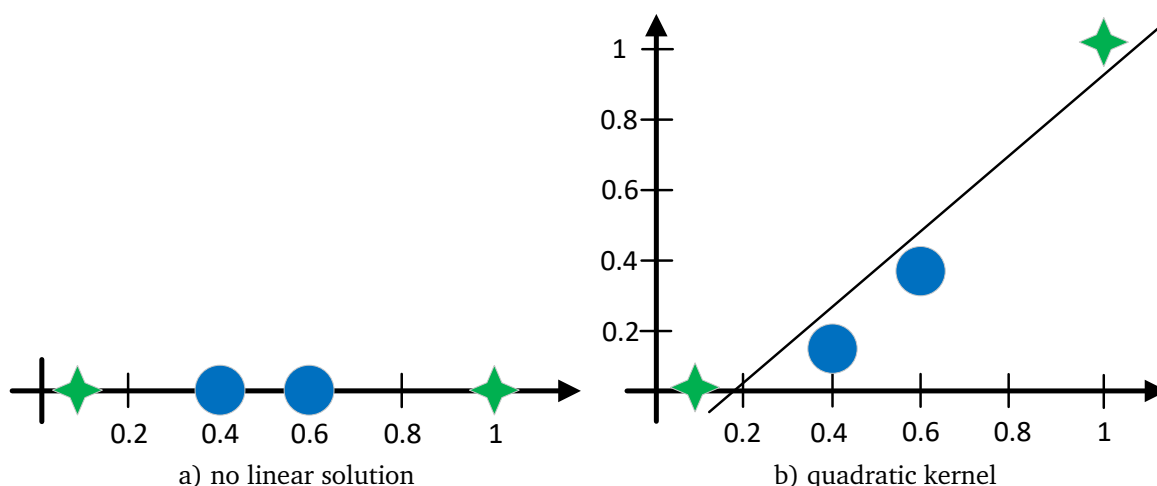


Figure 9.4: Illustration of the kernel trick for SVM classification

- o polynomial kernel:

$$K_p^\Phi(x, x') = (x^T x' + 1)^p \tag{9.21}$$

where p describes the order of the polynomial

- Gaussian kernel:

$$K_{\sigma}^{\Phi}(\mathbf{x}, \mathbf{x}') = e^{-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma^2}} \quad (9.22)$$

with σ as the standard deviation

- sigmoid kernel:

$$K_{k,\Theta}^{\Phi}(\mathbf{x}, \mathbf{x}') = \tanh(k(\mathbf{x}^T \mathbf{x}') + \Theta) \quad (9.23)$$

k determines the gain and θ the offset.

The advantage of the kernel function K^{Φ} lies in the low computation effort. The decision function is as follows

$$D(\mathbf{x}) = \sum_{i \in S} \alpha_i y_i K^{\Phi}(\mathbf{x}_i, \mathbf{x}) + b, \quad (9.24)$$

where b is given as

$$b = y_i - \sum_{i \in S} \alpha_i y_i K^{\Phi}(\mathbf{x}_i, \mathbf{x}_j). \quad (9.25)$$

SVMs have been developed to solve two-class problems. However, there are ways to apply the SVM to multi-class problems. In the one-against-all method, shown in Fig. 9.5 a), the problem is split into $c - 1$ two-class problems, where c indicates the number of classes. The orange area marks the area in the feature space which is not assignable.

A slightly more elegant method is the one-against-one method, shown in Fig. 9.5 b), to reduce the non-classifiable area. Here the problem is divided into $c(c - 1)$ two-class problems.

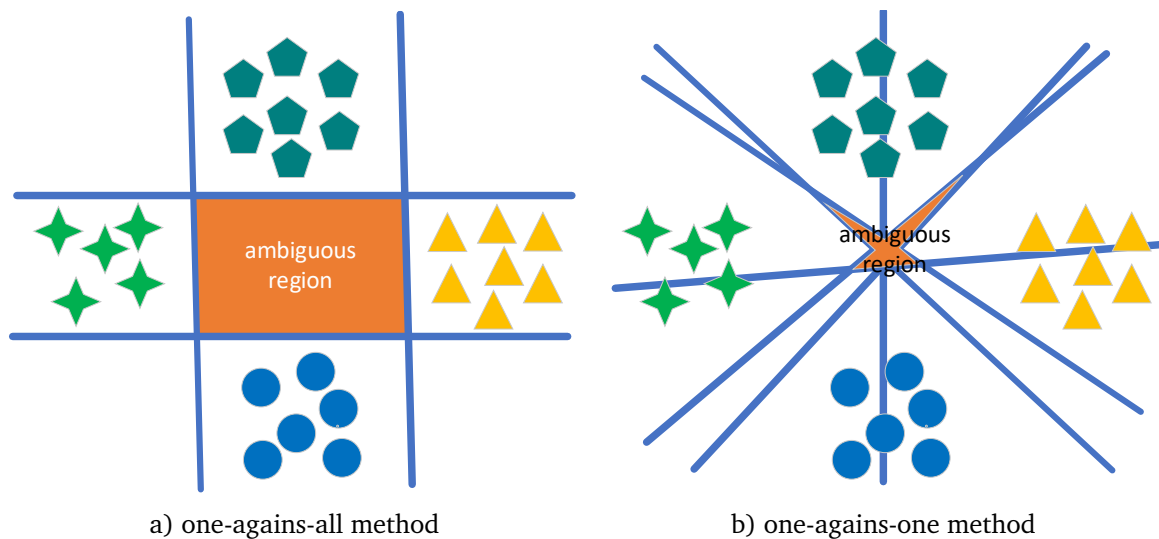


Figure 9.5: Linear separation of four classes in the two dimensional space.

The SVM is advantageous when dealing with high-dimensional spaces. Compared to the kNN, not all training data have to be stored. For the following classification, the stored parameters of the different hyperplanes can be used. The calculation effort during the classification is low compared to other investigated classification methods. However, more time is needed for the training than for the kNN, which does not lead to a problem in case of supervised learning. Nevertheless, finding a suitable kernel can be difficult and computationally intensive. For more detailed information about SVM, the literature [92, 3] is recommended.

9.5 FeedForward Neural Network

In recent years, the use of neural networks has become more and more important for various scientific questions. Therefore, this thesis examines the suitability of different neural network

structures for NILM classification. First of all, an investigation of a FeedForward Neural Network (FFNN) is realized which is a comparatively simple structure.

The FFNN exists of multiple artificial neurons illustrated in Fig. 9.6.

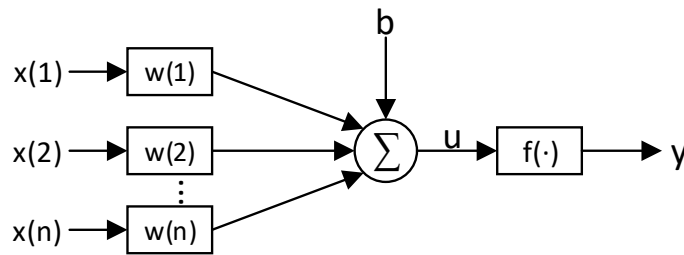


Figure 9.6: Model of an artificial neuron

The inputs $x(1), x(2), \dots, x(n)$ can be supplied by other neurons or by a signal input. The input signals x are multiplied with the individual weights $w(1), w(2), \dots, w(n)$ and summed up. If this result is higher than the activation threshold, also denoted as bias b , a signal is given in the direction of the output. The intermediate result u is mapped via a nonlinear element with the function f and returns the value for y at the output which may be used as input for other neurons.

Mathematically, this can be described as follows.

$$u = \sum_{n=1}^N w(n) \cdot x(n) - b \quad (9.26)$$

$$y = f(u) \quad (9.27)$$

Different activation functions are for example the linear functions, or different sigmoid functions like the hyperbolic tangent, or the logistic functions. A detailed explanation of these activation functions can be found in the literature [37, p.13-18].

An Artificial Neural Network (ANN) consists of at least two layers (input and output layers). These layers can be supplemented by so called hidden layers, as shown in Fig. 9.7. A multiple-layer neuronal net consists of one or more hidden layers. In the feed forward structure, the information moves without loops or feedback from the input to the output.

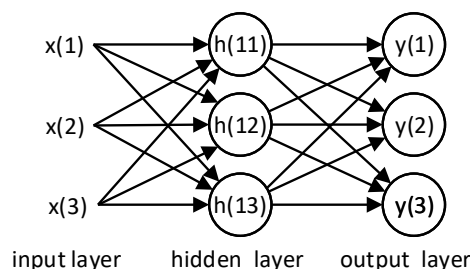


Figure 9.7: Structure of a FFNN with one hidden layer

The advantage of FFNNs is that no explicit knowledge about the problem to be solved is required. As a result, complex problems can be solved better than with conventional approaches. Due to the simple structure, the required computational effort of a FFNN is lower compared to other network structures. Nevertheless, the training of neurons compared to the previously presented classification methods is very computationally-intensive. Furthermore, an ANN requires many training examples. More details about FFNNs can be found for example in [23, p.227-232] or in [37].

9.6 Long Short-Term Memory Network

The LSTM network is able to store information of previous signals. Three types of gates are used for this purpose: input-gate, memory-gate, forget-gate, and output-gate. The memory of the LSTM network allows to classify based on time series data. Since the network structure has a memory, these changes can take the waveform over time into account. The structure of a single LSTM cell is shown in Fig. 9.8.

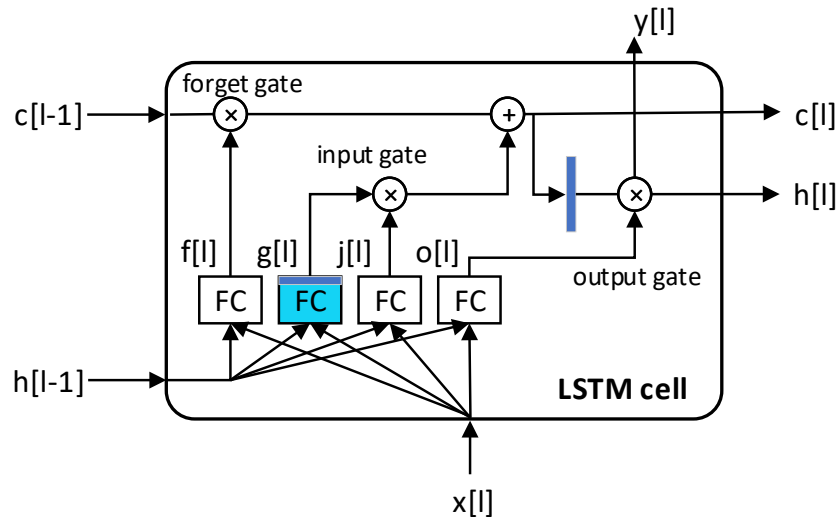


Figure 9.8: Structure of a LSTM cell

Like any other conventional cell, the LSTM cell, has an input x and an output y . The input is periodically loaded whereby $x[l] \in \{1 \dots K\}$, with l periods, and K is the number of sampling positions per period. In addition, the cell has more inputs and outputs on its sides to control adjacent cells. $h[l]$ can be interpreted as a short-term memory and $c[l]$ as a long-term memory.

Inside the LSTM net, four Fully Connected (FC) layers are shown. These receive the input signal x and the short-term memory $h[l-1]$. The blue marked cell which outputs $g[l]$ differs from the other cells, since the input value x and the short-term memory $h[l-1]$ are analyzed and processed further within the LSTM network. The three other layers are gate controllers, processing input and memory values. In contrast to the layer marked in blue, they only pass on a logical signal (represented by the gray arrows). The gates controlled by these logical outputs are closed at 0 and opened at 1. \otimes describes the element-by-element multiplication and \oplus the addition of the signals.

The forget-gate is controlled by $f[l]$ and determines which parts of the long-term memory should be deleted. The input-gate is driven by $j[l]$ and decides whether $g[l]$ should be passed to the output and memory. The output-gate $o[l]$ decides whether the sum of $c[l-1]$ and $g[l]$ should finally be passed to the output $y[l]$ and short-term memory $h[l]$.

LSTM networks can distinguish between important and unimportant inputs. Due to their memory, LSTM networks are suitable for analyzing time series, for example, in pattern recognition, audio signals and texts. In NILM, the memory of the LSTM network allows to use the TSt for classification. LSTM networks are very powerful methods to solve time series problems. Compared with the other classification methods, they allow to analyze the TSt. On the other side, the training effort is much higher, compared to the methods presented here. Further details on the function and structure of LSTM networks are described in [55, 406-416].

Long Short-Term Memory Network Applied on NILM

This section was published by the author [68].

“In order to use additional information from the Transient State Feature (TSF) for classification, FIT-PS was extended to a LSTM network, Fig. 9.9. The used LSTM network consists of $n_l = 100$ units. Thus, five periods before and 95 periods during and after the event can be used for classification. The complete structure of the applied

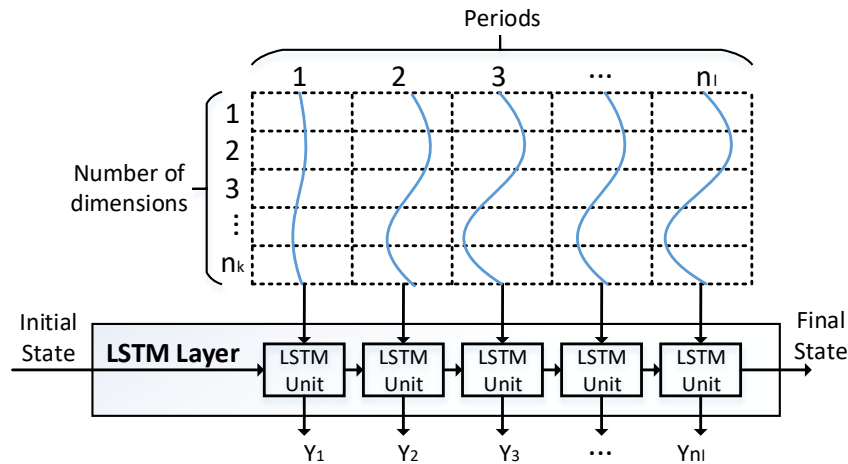


Figure 9.9: LSTM layer [68]

neural net can be seen in Fig. 9.10. At first, an 80-dimensional sequential input layer is used. The LSTM layer with 80 dimensions and a length of 100 LSTM units follows. The output of the LSTM net is connected to the fully connected layer which has an output size of 18 (equal to the number of classes). Here, the previously learned features from the LSTM net are combined and thus, used for the classification. A softmax activation function calculates the probability between zero and one for each class. The classification, or output layer, finalizes the network.” [68]

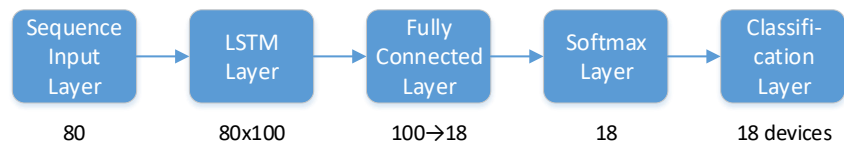


Figure 9.10: Complete structure of the LSTM net. [68]

9.7 Results of Validation and Test Data

This section presents the validation and test results of the different classification methods using different features. The simulations are based on test measurements of the HELD1 dataset. To define a suitable setting or structure of the different classification methods, the different basic approaches have been applied on validation data. The approach achieving the best result during validation is selected for the test. Nevertheless, all approaches are applied on the test data in order to confirm the most suitable classification setting, determined by applying validation data. For testing the complete test measurements of the HELD1 dataset, 13,200 on and off events are used.

The classifier is trained using the examples provided by the training data. The validation data are used for an initial evaluation to verify the suitability of the parameter settings. Thereby, the validation data have to be used, in order to avoid overfitting. The parameter setting of the classifier, which achieves the best results based on the validation data, will subsequently be used for the classification of the test data. In order to verify whether suitable parameters have been selected in combination with the validation data, the test data are verified with the different

parameter settings.

Fig. 9.11 illustrates the distribution of the classification data. The dataset is split into training, validation, and test data. For training, 100 switching on and off cycles for each of the 18 appliances are available. For validation, 1,200 switching on and off events in total are used. The different test scenarios contain 13,200 events in total.

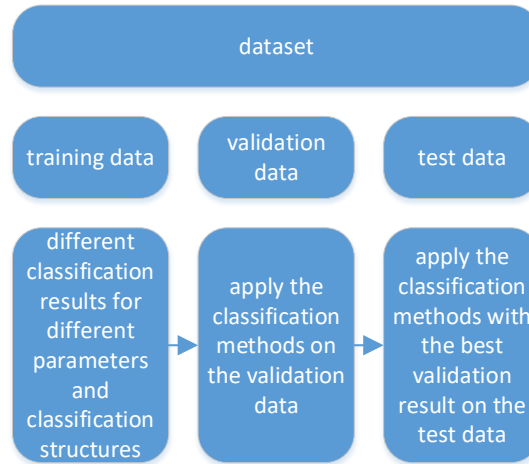


Figure 9.11: Segmentation of the data for classification

9.7.1 Validation Results of the k-Nearest Neighbor

The kNN classifier has the number of k neighbors as degree of freedom. To find a suitable k , the training data are split into training and validation data. The choice of k also depends on the used feature. Therefore, different features (from the waveform P and Q ; P , Q , and HARs; and FIT-PS) have been applied to the kNN with a changing value for k . When using the power signal in combination with the HARs, an appropriate number of HARs must be selected additionally. Therefore, with the validation data, as shown in Fig. 9.12, all possible combinations of $k = 1$ to $k = 40$ combined with HARs from 1 to 39 are calculated. Using five to ten HARs result in the

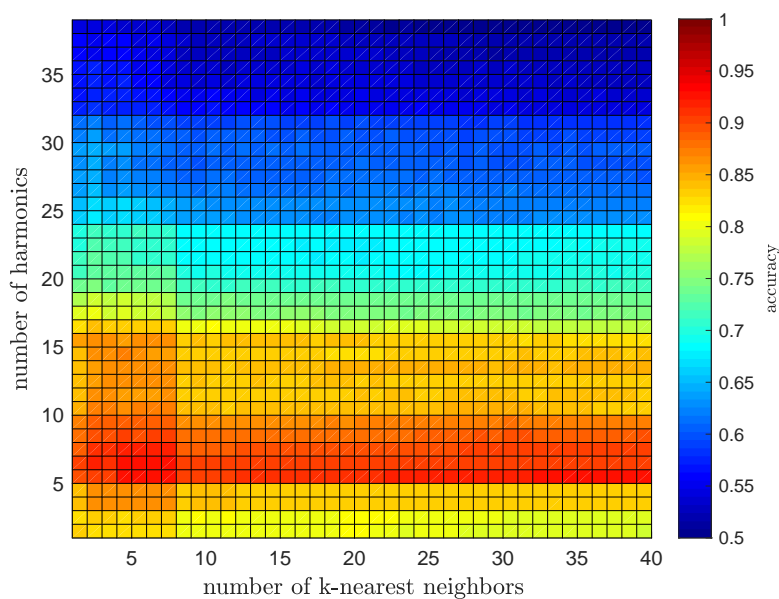


Figure 9.12: Validation results of the kNN for different number of k and HARs

best performance. A further increase of HARs leads to deterioration of the results. Additionally,

the used HARs do not contain additional new information that are useful for the classifier. Due to the additional noise, the results deteriorate.

Compared to the number of applied HARs, the number of utilized k neighbors is less relevant for the result. For $3 < k < 8$, good results can be achieved. Using more k neighbors ($7 < k < 20$) leads to a deterioration of the results. However, a further increasing of k , also leads to an increased accuracy. The best number of k neighbors depend on the structure of the individual clusters, as well as, on the position and distance to each other. Therefore, the selection of parameters depends on the available appliances.

The best values for the validation data, can be achieved with $k = 6$ and five HARs.

The different features, investigated during this work, have been applied on the kNN with different values for k , illustrated in Fig. 9.13. The number of HARs is chosen from the previously calculated best results (HAR= 5). The best result can be achieved with the FIT-PS wave-

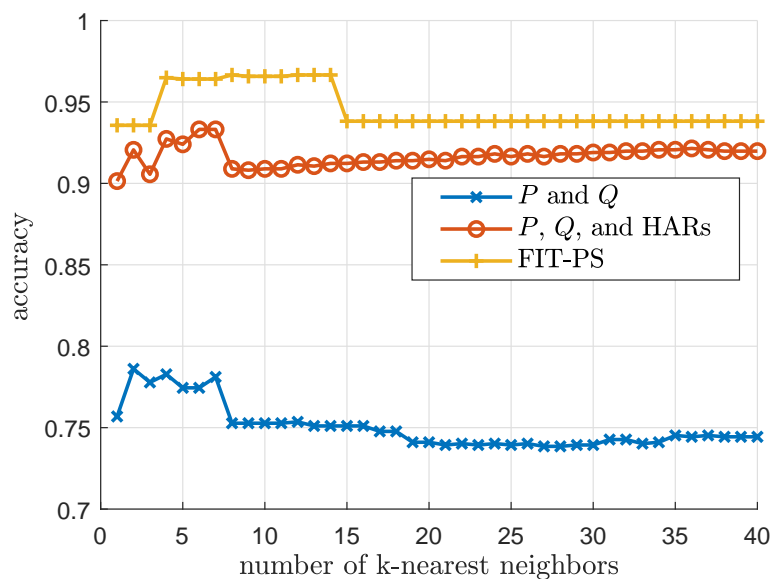


Figure 9.13: Validation results of the kNN for different features and a varying number of k -neighbors

form relying on the feature calculated from this waveform. The FIT-PS feature outperforms the power without additional HARs, as well as with different numbers of additionally HARs. FIT-PS achieves the best result with $k = 8, 12, 13, 14$ with 96.66%. The power waveform and the HARs perform best with $k = 6, 7$ where a accuracy of 93.32% is achieved. By using only the active and reactive power waveform, a maximal accuracy of 78.61%, if combined with two HARs, can be achieved.

The steep increase of the accuracy from $k = 4$ and the similar steep decrease at $k = 15$ for the FIT-PS feature can be explained by the comparatively low amount of validation data. The same applies to the fluctuations of the accuracy of power features in combination with HARs. For validation, 1,200 events of the HELD1 dataset are used. However, these data have to be split to 18 different appliances, resulting in about 66 events per device. As a result, even slight changes of the parameters may have a significant impact on the accuracy.

However, when selecting a small number for k , the results will be worse, as the classifier is overfitted. The classifier does not learn to identify the appliances, but only the specific training examples. In case that k is chosen too large, the results will also deteriorate since the generalization is now very pronounced. Thus, detailed appliance-specific characteristics are no longer considered.

9.7.2 Test Results of the k-Nearest Neighbor

The test results of the kNN with different number of HARs and k -neighbors are illustrated in Fig. 9.14. The best result using the kNN on the power waveform combined with HARs achieves

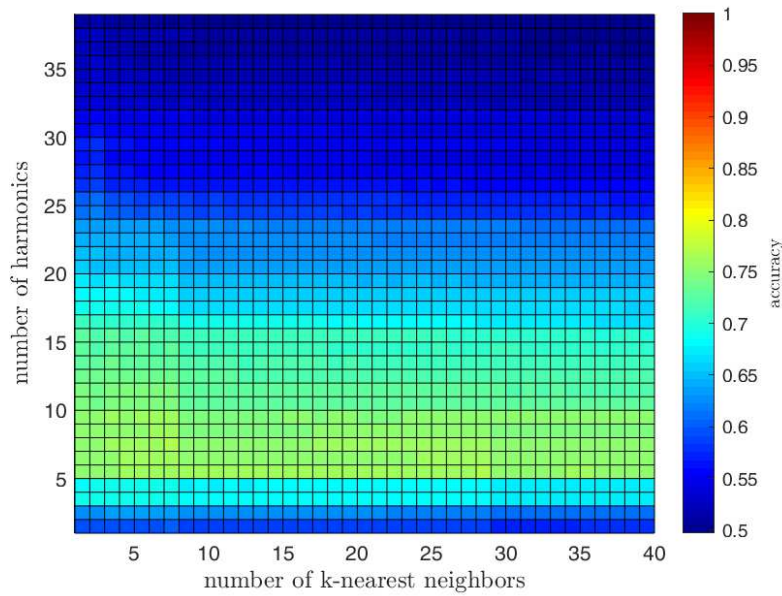


Figure 9.14: Test results of the kNN for varying k and number of HARs

an accuracy of 76.49% with seven HARs and $k = 7$. The test result for the parameters selected with the validation data (five HAR and $k = 6$) achieves an accuracy of 76.36% and is close to the best result which is achieved with kNN for the test data.

Fig. 9.15 presents the test results of all compared features for different number of k -nearest neighbors. The best result with 91.22% is achieved with FIT-PS and $k = 14$. Since all results for

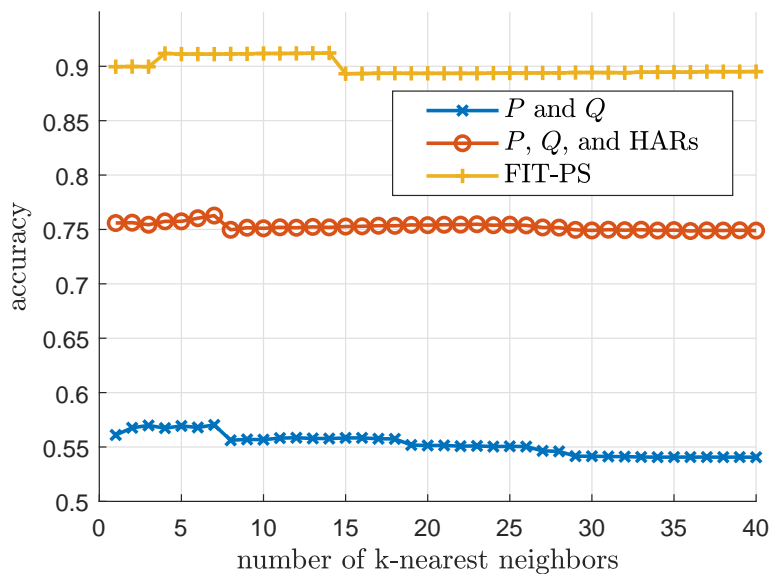


Figure 9.15: Test results of the kNN for varying number of k -neighbors

k , between four and fourteen, are above 91 %, these results can be achieved with the selected parameters from the validation data. Compared to the validation data, the difference of the results achieved with FIT-PS and the power waveform with HARs is higher. A reason for this is

that HARs have difficulties with multiple active appliances at the same time. The exclusive use of the active P and reactive power Q leads to significantly worse results.

9.7.3 Validation Results of the Bayes Classifier

During the analysis of the validation data, the optimal number of HARs for the use of the Bayes classifier was examined and is illustrated in Fig. 9.16. At first, a normal distribution is assumed. The assumption of a normal distribution leads to detection rates of 55.07% with 16 HARs. The use of a Gaussian kernel as kernel density estimation did not improve either. Here, the best result is achieved with five HARs with 49.3%. The use of Bayesian optimization for the parameters of the kernel density estimation could improve the results for active and reactive power in combination with HARs. The best result with 85.38% for the validation data could be achieved with ten HARs. The addition of individual HARs leads to significant variations in accuracy. Often, a deterioration can be seen when an even HARs is added. Most appliances have the same power consumption for positive and negative half wave. Therefore, the investigation of even HARs does not contain any further information. An further improvement of the results can be achieved if suitable HARs are selected specifically. However, since all possible combinations must be simulated, this increases the computing effort enormously.

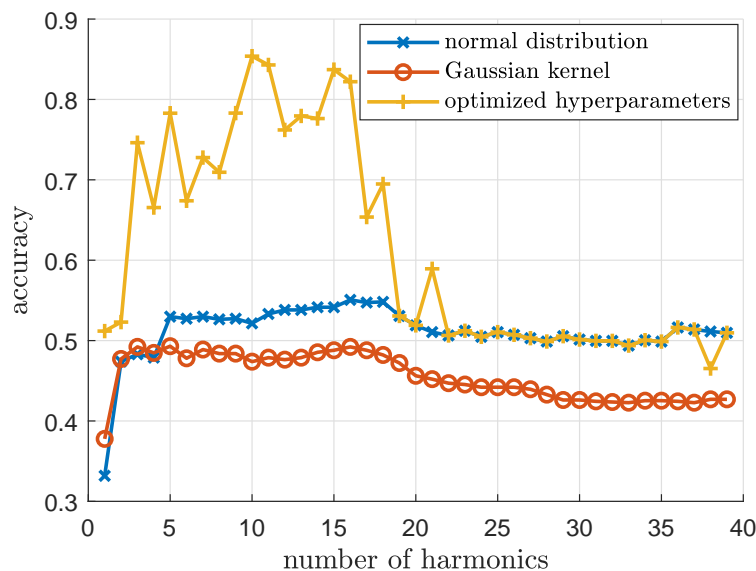


Figure 9.16: Validation results of the Bayes classifier for the feature combination P , Q , and HARs for varying number of HARs

Fig. 9.17 presents the results of different features with different assumptions for the Bayes classifier. The use of the active and reactive power only, is clearly surpassed by the other two features. For Gaussian kernel and a normal distribution, FIT-PS reaches with 73.76% and 82.03% the best result. The results of FIT-PS decrease when using optimized hyper-parameters. Here, the active and reactive power in combination with ten HARs lead to the best result with an accuracy of 85.38%. The high-dimensional FIT-PS feature cannot be mapped optimal in the hyperspace. Due to the phase shift, a normalization of the individual dimensions of the FIT-PS signal shape is problematic. This has a negative effect on solving an optimization problem, since the size of the amplitude in the different dimensions is very different.

9.7.4 Test Results of the Bayes Classifier

The results of the test data, achieved with the Bayes classifier for different number of HARs, is presented in Fig. 9.18. The best result is achieved with the Bayes classifier using an optimized

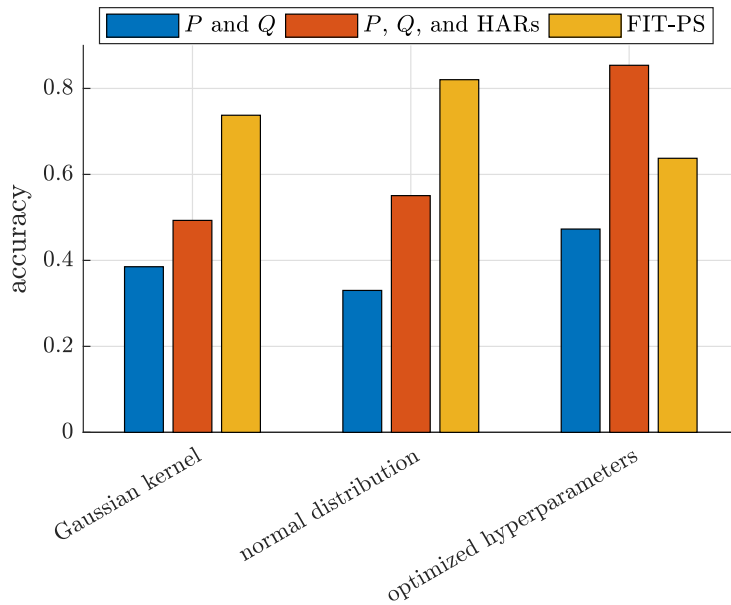


Figure 9.17: Validation results of the Bayes classifier using different features

calculation of the hyper-parameters. Therefore, the training data are divided into training and validation data in order to avoid over-fitting. The best result with 61.93 % is achieved with 15 HARs. The best result during the validation was achieved with ten HARs resulting in an accuracy of 60.38 % for the test data.

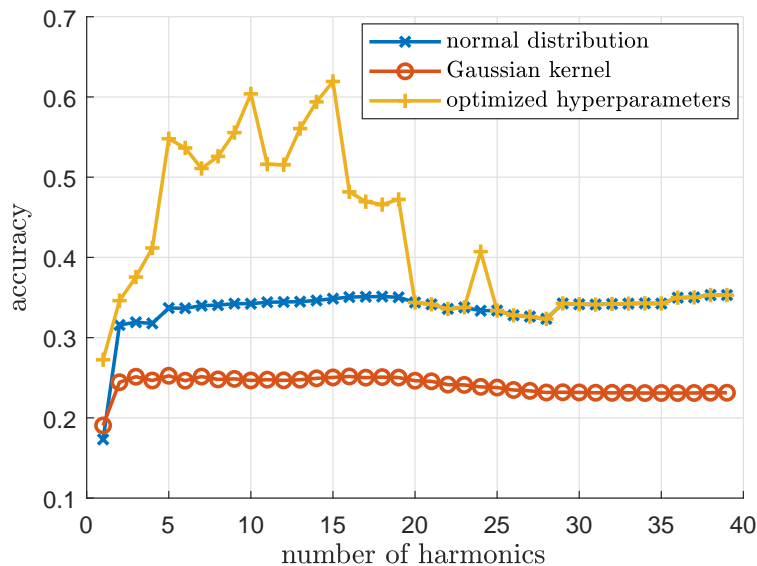


Figure 9.18: Test results of the Bayes classifier for the feature combination P , Q , and HARs for varying number of HARs

A comparison of the different parameter settings of the Bayes classifier is illustrated in Fig. 9.19. The use of a Gaussian kernel results in a poor accuracy, especially for the feature combination with HARs which is below 30 %. Assuming a normal distribution, FIT-PS achieves the best result with an accuracy of 68 %. By using the optimization, significantly better results of 47.29 % or 61.93 % are achieved for the power feature alone and in combination with the HARs, respectively. However, the results are slightly worse when using the optimization on the FIT-PS feature.

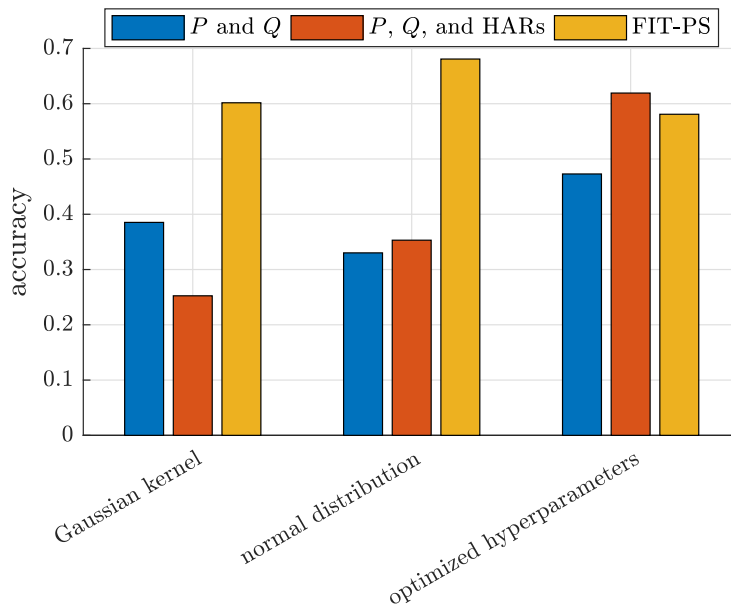


Figure 9.19: Test results of the Bayes classifier using different features

The choice of a suitable density distribution turns out to be difficult because the features consist of a very large number of dimensions. Therefore, a suitable density distribution for each individual dimension must be estimated. The effort could be reduced by an additional dimension reduction. Since the classifier is tested on a laboratory data set, the probability of switching on or off is for all appliances more or less equal. Therefore, the Bayesian classifier cannot take advantage of different appearance probabilities for different devices.

9.7.5 Validation Results of the Support Vector Machine

As with the previous classification methods, the validation data is used to examine the optimal number of HARs for the different SVMs, illustrated in Fig. 9.20.

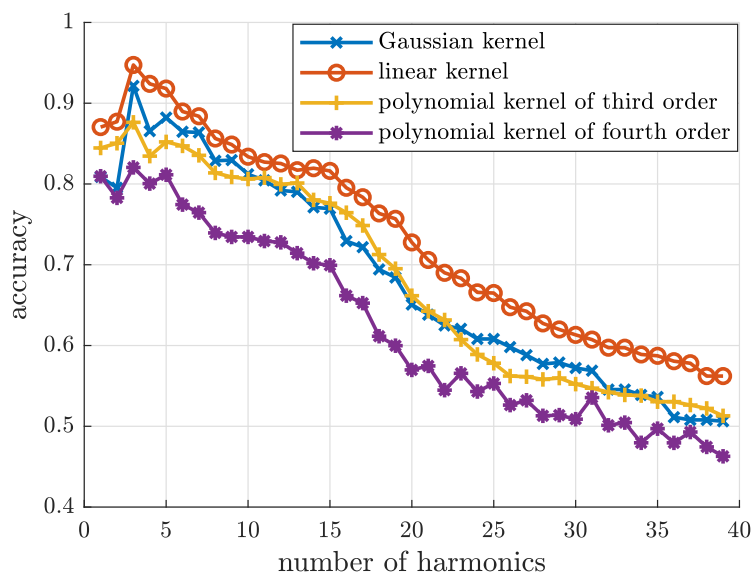


Figure 9.20: Validation results of the SVM for the feature combination P , Q , and HARs and a varying number of HARs

All validated kernels achieve the best result with three HARs. The addition of HARs of

higher order leads to a deterioration of the results. Only for individual positions, a relative improvement can be seen when increasing the number of HARs to the next higher number. This effect is visible exemplary in case of the change from four to five HARs when a Gaussian or the polynomial kernels is applied. In most cases, only the additional use of an odd number of the HARs lead to an improvement. This indicates that they contain information relevant for the classifier. On the other hand, the deterioration observed in the addition of even numbers of HARs indicates that they contribute more noise than information to the classification.

For the validation of the HARs with the SVM, the best result is achieved by the linear kernel with 94.74%, followed by the Gaussian kernel with 92.14%, the polynomial kernel of third order with 87.63% and fourth order with 82.05%.

The parameters of the different kernels are not optimized for the specific number of HARs. An parameter optimization needs to be applied for the different kernels to improve the results further. However, the focus of this chapter is on the comparison of FIT-PS and P , Q with or without HARs. Therefore, the standard parameters are applied to all features.

The four different kernels have been applied on all investigated features, see Fig. 9.21. For

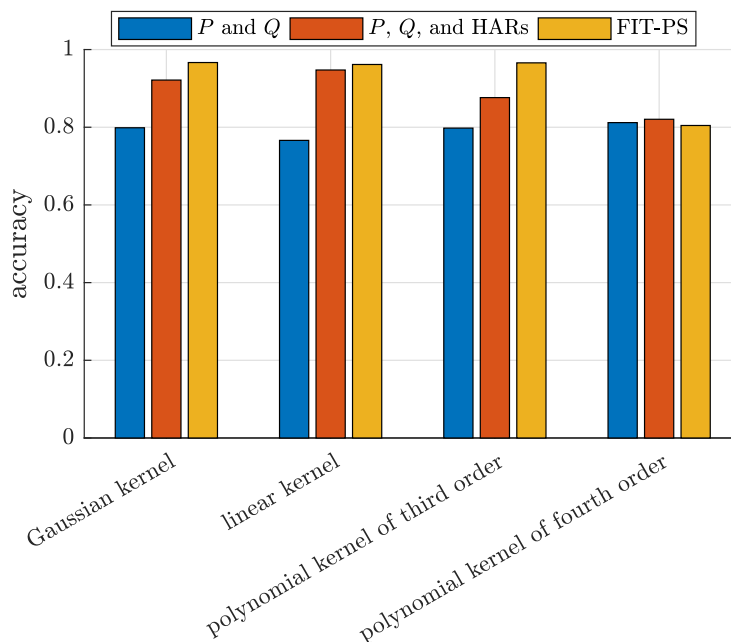


Figure 9.21: Validation results of the SVM different features

the exclusively use of active P and reactive power Q , the polynomial kernel of fourth order leads to the best accuracy with 81.19%. The variation of accuracy, when other kernels are used, is small. If the HARs are used additionally, the highest accuracy is reached with the linear kernel as illustrated in Fig. 9.20. Using FIT-PS as feature, the Gaussian kernel leads with 96.66% to the best result closely followed by the polynomial kernel of third order with 96.57%.

9.7.6 Test Results of the Support Vector Machine

The test results achieved with SVM for different number of HARs are visualized in Fig. 9.22. The best result with 72.57% can be achieved with a linear kernel using five HARs. During the validation, three HARs lead to the best result. This selection would result in an accuracy of 69.09%. With the use of more complex kernel functions, the results of the power feature in combination with the HARs is getting worse.

Fig. 9.23 presents the different results achieved using different kernel and feature combinations. The best result with 89.42% is achieved with FIT-PS and the Gaussian kernel. The power feature, in combination with HARs, reaches an accuracy of 72.57% with a linear kernel. Without the HARs a accuracy of 59.64% is achieved by a polynomial kernel of third order.

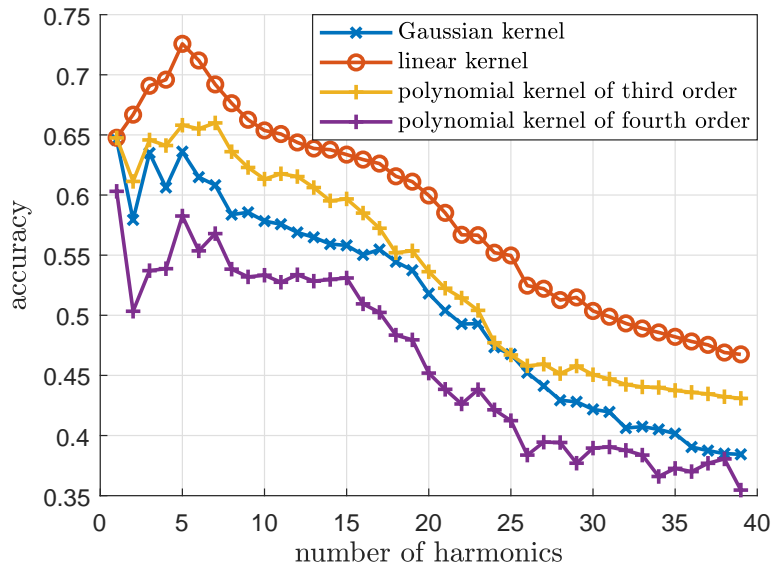


Figure 9.22: Test results of the SVM classifier for the feature combination P , Q , and HARs for varying number of HARs

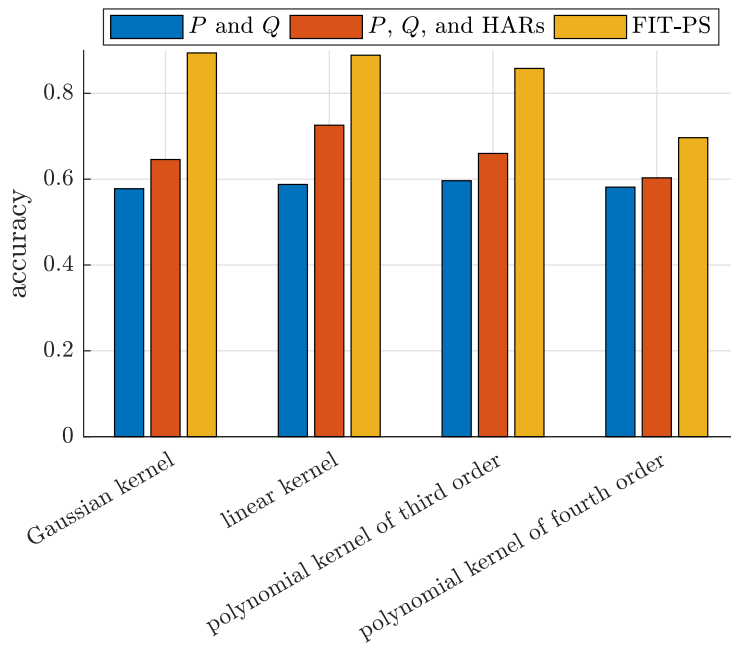


Figure 9.23: Test results of the SVM using different features

9.7.7 Validation Results of the Neuronal Network

The analysis of different numbers of HAR on the FFNN promises that 15 HARs lead to the best result for this classification method on the HELD1 dataset. This section is part of a paper published by the author [68].

“The structure of the feedforward net being applied is shown in Fig. 9.24. Depending on the waveform, a 17 (P , Q and 15 HARs) or 80 (FIT-PS with 80 samples per period) dimensional input vector is prevalent. Since the tests with a higher number of hidden layers could not improve the accuracy of the classifier, only one hidden layer is used. The dimension of the output layer depends on the number of appliances to be distinguished. Since it is easy to distinguish between on or off events, two different classifiers, one for on events and one for off events, are used.

To find the optimal number of neurons, the validation data was applied to the net structures with one neuron up to 500 neurons. Therefore, the data 0003, 0116 and 0201 of the HELD1 dataset have been used as validation data. These measurements of the dataset have been selected for validation because they include all used devices. The results are shown in Fig. 9.25 and 9.26 with blue dots. Depending on the initialization values of the neural net, random numbers in general, different results can be achieved. Due to this, 100 iterations with different initialization settings and different numbers of neurons have been simulated. The lower (red line), upper (green line) and average (black line) limits of accuracy were calculated in Fig.9.25 and 9.26.

The maximum accuracy of both features can be achieved with approximately 200 neurons. A higher number of neurons does not effect the accuracy significantly. The accuracy achieved with FIT-PS is higher than the accuracy obtained with the standard signal forms P , Q , and HARs, regardless of how many neurons are used. Furthermore, the variance of the results, caused by different initializations is significantly lower with FIT-PS as input signal.” [68]

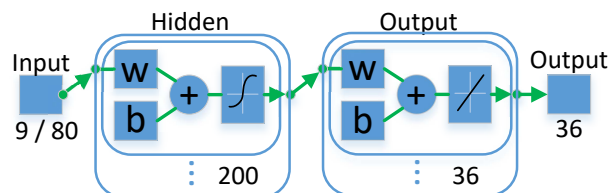


Figure 9.24: Structure of the feed forward neural net. The number of input dimensions depends on whether P , Q , and HARs are used as feature or FIT-PS [68]

Fig. 9.27 presents the validation results of the different features for different initialization settings, sorted from worst to best. For exclusively use of the active P and reactive power Q , the best result achieved an accuracy of 76.61 % for the validation data. With additionally use of HARs, an accuracy of 78.78 % is achieved. The use of FIT-PS as a feature for the FFNN shows significant improvement. The worst result of 100 iterations already achieves a significantly higher accuracy of 92.66 %. The best result achieves an accuracy of 96.32 % and the variation of the accuracy of FIT-PS is significantly lower compared to the other features.

Problems with local minimums lead to different results for different initialization seeds. To circumvent the local minimums, different initialization seeds are used on the set of validation data. The results are shown in Fig. 9.27. For the classification with the transient, only FIT-PS is used. P and Q are not used since they require a window length of at least one period for their calculation. Furthermore, since the current signal is not reliably periodic during the transient, the HARs are not used either. The best result of FIT-PS with a LSTM net using the TSt

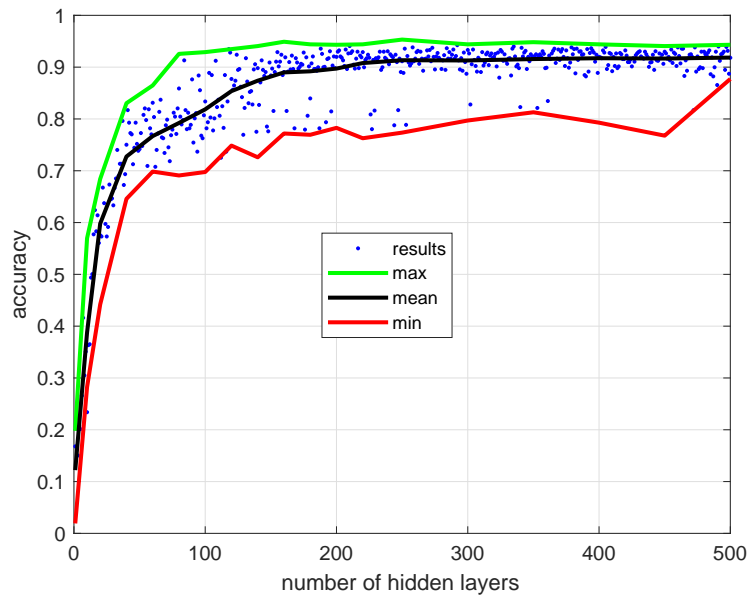


Figure 9.25: Validation results of a feed forward net using P, Q, and HARs for a different number of neurons. The limits of accuracy are calculated by 100 iterations with different initialization settings and different number of neurons. [68]

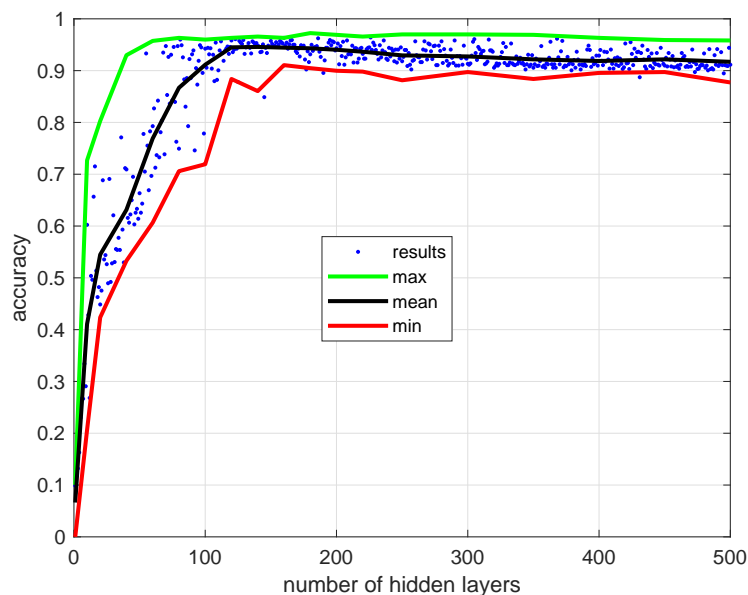


Figure 9.26: Validation result of a feed forward net using FIT-PS for a different number of neurons. The limits of accuracy are calculated by 100 iterations with different initialization settings and different number of neurons.[68]

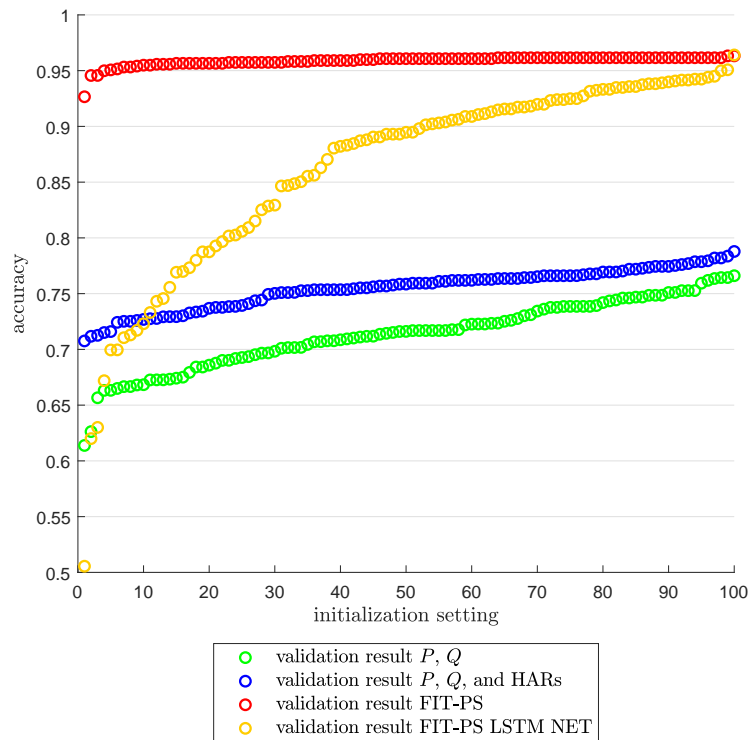


Figure 9.27: Results of a FFNN using different features for the validation data 0003, 0116, 0201. The accuracy is calculated by 100 iterations with different initialization settings and different number of neurons.[68]

reached an accuracy of 96.41 %. Compared to the results of FIT-PS on the FFNN, the variation of the accuracy is significantly higher for this setting. Therefore, it can be assumed that the training method is not adapted optimally. Nevertheless, as very good results can be achieved, the classifier is used for comparison with other classification methods and features.

9.7.8 Test Results of the Neuronal Network

In this subsection, the test results of the FFNN, as well as the results achieved with the LSTM net are presented. In contrast to the classification methods presented so far, multiple iterations are carried out during the training of the neural networks. Problems with local minimums lead to different results for different initialization seeds. To circumvent the local minimums, different initialization seeds are used on the set of validation data.

The results of the neural networks have been presented by the author in [68]

“Fig. 9.28 shows the accuracy of the validation data (‘o’) and the accuracy of the test results (‘x’) for various initialization settings. The figure shows the relationship between the accuracy of the validation data and the test data for each initialization. This relationship is most clearly demonstrated for FIT-PS with the LSTM network (orange): the better the initialization parameters were chosen, the higher the accuracy of the test results became, based on the results of the validation data. The results of the test data of P and Q (green) are widely spread. This means that the best initialization setting determined with the validation data does not necessarily represent a good initialization setting for the test data. Consequently, it is difficult to find the optimal initialization setting for P and Q . Although the additional use of the HARs (blue) leads to a reduction in the variance of the test results, the increase remains

low. This is also the reason why in Fig. 9.29 the selected initialization (dotted line) is only slightly better than the median (solid line) of all initialization settings. Thus, even with the additional use of HARs, it remains difficult to determine the optimal initialization setting through the validation data. The same applies to the results with FIT-PS, wherein FIT-PS achieves the lowest variance. However, for all combinations of scenarios and features the use of the best initialization setting based on validation leads to an accuracy similar or most often even superior to the median accuracy. Therefore, the use of the validation data to identify the optimal initialization, especially for the LSTM network, proves to be advantageous.” [68]

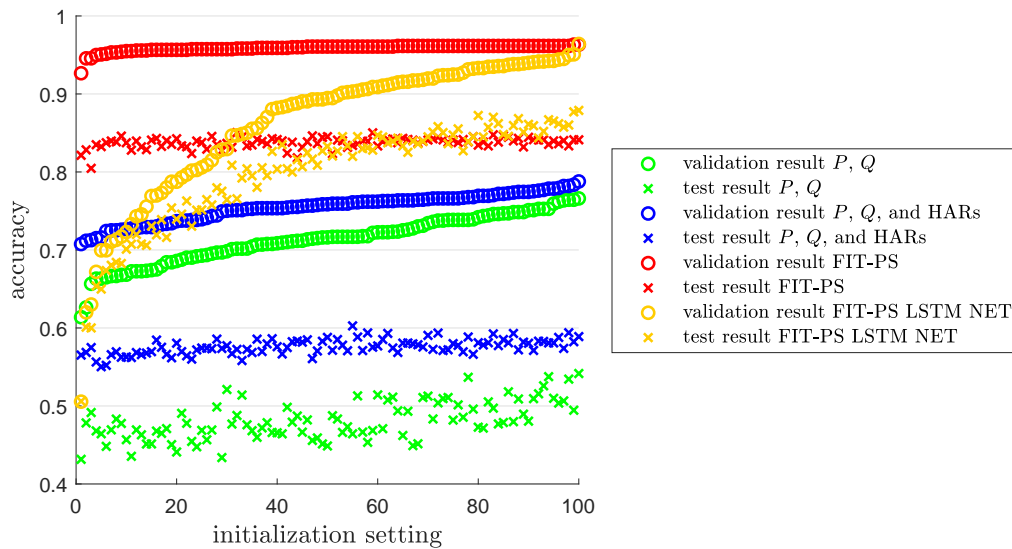


Figure 9.28: Comparison of validation results and test results on different initialization settings. The initialization settings are sorted according to increasing accuracy of the validation data. In general a higher accuracy of validation also leads to a higher accuracy of the test results. [68]

A detailed investigation of the results has been published by the author in [68].

“Table 9.1 and 9.2 show the results of measurement 0194 of the dataset using standard signal forms and the proposed signal form. The IDs of the devices in the measurement were listed horizontally and vertically. The devices with the ID 13, 15, 21, 23 did not appear in the measurement, but were falsely detected. The tables show the number of correctly classified devices in the diagonal (marked in green). For example, in Tab. 9.1 for the device ID 5, this means that it has been recognized correctly 36 times. However, it was incorrectly identified once as device ID 14 and three times as device ID 13.

The results of this measurement are not representative for the whole dataset because the average recognition rate for the dataset is higher. This measurement was selected to explain the following problem: Device ID 13 and ID 14 are refrigerators, which have a very similar power consumption behavior. Therefore, misclassification of both devices occurs frequently for standard and FIT-PS features. Using a FFNN net with P , Q , and HARs for this part of the dataset, an accuracy of 66% could be reached whereas FIT-PS reaches an accuracy of 75%.

Results of all different features are shown in Fig. 9.29. They are divided into measurements in which one, up to four and up to six appliances are active. The features and classifiers utilized are visualized in different colors. In order to test whether

Table 9.1: Conversion matrix of the classification results of measurement “0194” using P , Q , and HARs in a FFNN [68]

Dev. ID actual set														
	3	5	7	9	10	11	14	16	17	19	13	15	21	23
3	20	0	1	0	0	1	0	0	0	0	1	0	3	14
5	0	36	0	0	0	0	1	0	0	0	3	0	0	0
7	0	0	33	0	0	5	0	0	0	0	1	0	1	0
9	2	0	0	29	0	0	0	4	0	0	1	0	4	0
10	1	0	0	0	25	2	0	1	0	0	7	0	0	4
11	1	0	2	0	0	28	0	4	0	0	1	0	1	3
14	0	0	0	0	0	6	6	0	0	0	22	1	4	1
16	0	0	0	0	0	1	0	29	0	1	0	0	0	9
17	0	0	1	0	0	4	0	0	32	0	1	0	2	0
19	3	0	0	0	0	0	0	4	0	26	0	0	6	0

Table 9.2: Conversion matrix of the classification results of measurement “0194” using FIT-PS in a FFNN [68]

Dev. ID actual set														
	3	5	7	9	10	11	14	16	17	19	13	20	22	23
3	36	0	0	0	0	2	0	1	0	0	0	0	0	1
5	0	30	0	0	0	3	0	0	0	0	1	0	0	6
7	0	0	40	0	0	0	0	0	0	0	0	0	0	0
9	2	0	0	23	1	3	0	5	0	0	0	0	0	6
10	0	0	0	0	39	0	0	0	0	0	0	0	0	1
11	2	0	0	0	0	34	0	3	0	0	0	0	0	1
14	0	0	0	0	0	3	3	1	0	0	31	0	0	2
16	0	0	0	0	0	0	0	40	0	0	0	0	0	0
17	0	0	0	0	0	3	0	1	32	0	0	2	2	0
19	1	0	0	0	3	4	0	1	0	24	0	0	0	6

the initialization parameters selected during the validation also lead to good results for the test data, all 100 initialization parameters were used for the test data. The different initialization parameters lead to different results. These vary according to the feature utilized, classifiers and the number of simultaneously active devices. The LSTM net has the largest spread. During the validation, the classifier with the nearly optimal initialization seed can be selected. The dotted lines in the bars show the results reached with the best initial values from the validation. The solid lines in the bars show the median of all different test initial values of the classifier.

When one appliance is active at one time, the FIT-PS feature with a FFNN net reaches the highest accuracy with 90%. FIT-PS outperforms P , Q , and the HARs clearly in all scenarios. The differences become particularly evident if several devices are switched on at the same time. Here, the results remain most stable when using the LSTM network.” [68]

9.7.9 Comparison of the Different Classification Methods and Features

This section provides an overview of the results achieved with the different classification methods. Therefore, the best result of each feature and classification method is illustrated in Fig. 9.30. For each feature, different classification methods or parameter settings lead to the best result. Nevertheless, the kNN achieves the best result with the FIT-PS feature, as well as with P , Q , and HARs. Also, when using P and Q without HARs, its result is close to the best, which is obtained by applying the SVM. The accuracy of the SVM and the neural network (FFNN and LSTM net) is very similar for every feature. The use of the TSt increases the accuracy of the neural net but

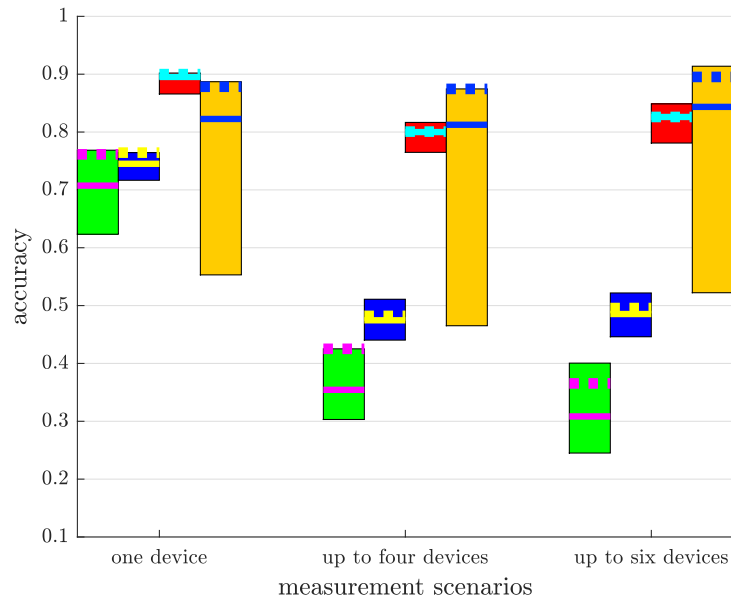


Figure 9.29: Accuracy of a FFNN using P and Q (green); P , Q , and HARs (blue); FIT-PS (red); LSTM net using FIT-PS (orange). The dotted lines represent the accuracy using the best initialization setting derived from the validation data. The solid lines represent the median of the accuracy of the test data.[68]

is still below the result of the kNN. The results achieved by the Bayes classifier cannot reach the accuracy of the compared classifiers.

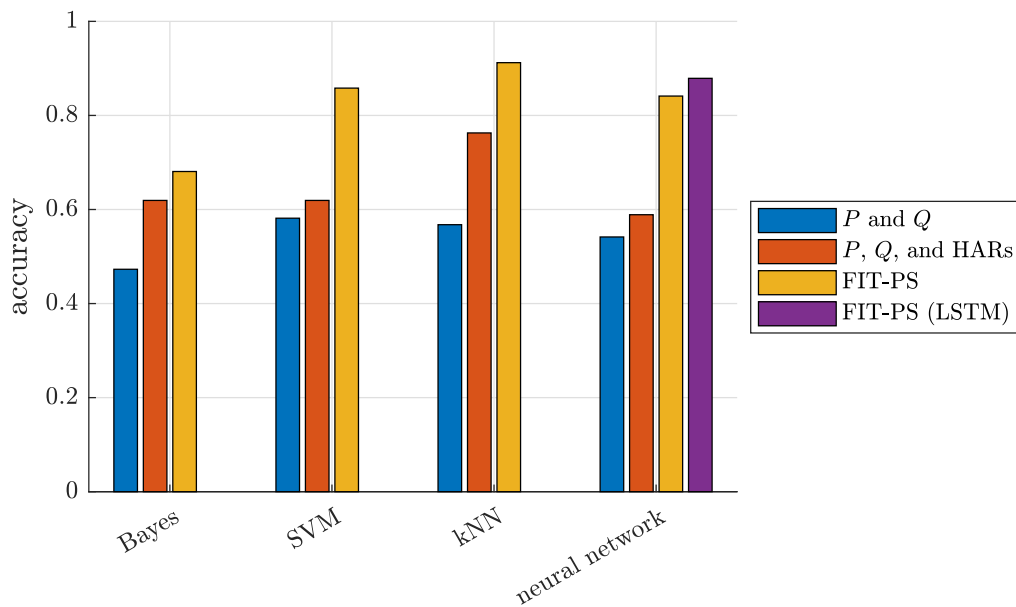


Figure 9.30: Best results of the different classifiers and features.

Comparison of the Calculation Time

Because of different dimensions of the features used for the classification, the training time varies. The achieved accuracy and the average training time are presented in Tab. 9.3 and 9.4. Due to different efficient implementations, the time required in MATLAB only provides limited

information about the computational effort. Nevertheless, both tables provide a rough estimation of the respective computational effort.

The simulations were performed on the following environment: Intel(R) Core(TM) i7-6800K CPU@3.40 GHz; RAM: 64 GB; Win 10 Pro 64-Bit, MATLAB 2018a. In order to get an impression about the effort of the different classification methods, the training and test time for the classification methods, with respect to the different features, are presented in Tab. 9.3 and 9.4.

First, the training and test time of the kNN is considered in relation to the number of HARs and the number of k . The training and test time for different k number of HARs and neighbors is presented in Fig. 9.31 and 9.32

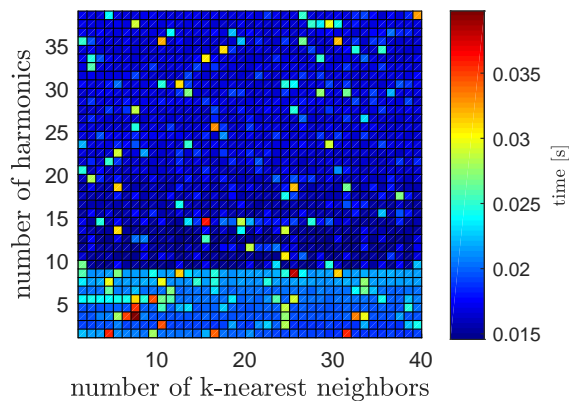


Figure 9.31: Training time of the kNN for different numbers of HAR and neighbors

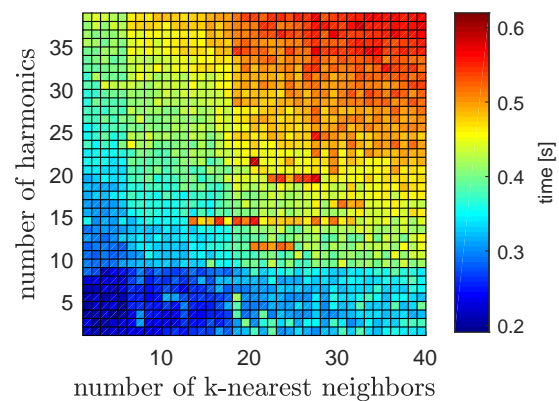


Figure 9.32: Time for classification using the kNN for different numbers of HARs and neighbors

Since the simulation runs on a windows 10 operating system which handles other additional tasks in the background during the simulation, a few time peaks occur. To a great extent, the training time is independent of the number of HARs or the number of neighbors used. This was expected because the training examples are just stored.

However, the time required for the classification reveals a strong dependence on the number of HARs as well as the number of neighbors. The required classification time correlates positively with the number of HARs or k -nearest neighbors.

Since the number of iterations for the training may differ, the presented training time takes only into account the time required for one iteration. In case of the neural network, where several iterations are calculated, the time needs to be multiplied by the corresponding number of iterations. To find the best number of HARs or k -nearest neighbors, additional training iterations are required.

For the different classification and feature combinations, the parameter setting with the highest accuracy during the validation is selected. Therefore, less complex power features may require a significantly longer training compared to the higher-dimensional FIT-PS feature, as the Bayes classifier illustrates. While FIT-PS with a normal distribution achieves the best result of the Bayes classifier, P and Q with and without HARs, obtain the best results when different parameters are set by an optimization procedure. The duration of the optimization process increases the required time significantly. Hence, Tab. 9.3 and 9.4 present the required time to achieve the best result with the respective classification/feature combination of validation data.

The fastest training time can be achieved if the kNN is used. The training time increases with increasing number of feature dimensions. This is not surprising since the training examples only

Table 9.3: Training time of the different classification methods and features in seconds

	Bayes	SVM	kNN	FFNN	LSTM
P, Q	0.301	2.360	0.0206	28.37	-
P, Q , and HARs	0.275	2.214	0.0241	58.81	-
FIT-PS	0.6337	119.8	0.0380	206.78	482.29

need to be saved. Additionally, the implementation of the kNN in MATLAB is very efficient. The naive Bayes classifier without optimizing parameters during the training allows time efficient training too. The training time of the SVM depends on the complexity of the kernel and the number of dimensions of the feature. The used neural networks require the longest training time since they need to run for several iterations (100 iterations for the results presented here). Especially the LSTM net with FIT-PS requires a comparatively long training time.

Tab. 9.4 presents the corresponding test time of the complete test measurements of the HELD1 dataset containing 13,200 features. The kNN has the fastest testing runtime, followed

Table 9.4: Test time of the test measurements of the HELD1 dataset in seconds

	Bayes	SVM	kNN	FFNN	LSTM
P, Q	0.7978	6.655	0.2243	0.4288	-
P, Q , and HARs	0.8418	5.665	0.2566	0.4538	-
FIT-PS	4.1615	25.83	1.0271	0.4886	5.897

by the FFNN. While the impact of the dimension of features in the neural networks show only negligible impact, it affects the runtime of the kNN significantly. Bayes classifier, SVM and LSTM network have a comparably long run time for the classification of the test data. The various classification methods are implemented different efficiently. This is explains for example the low runtime of the kNN during testing compared to other methods.

9.8 Interpretation of the Harmonic Results

The classification results show that FIT-PS compared to the combination of P , Q , and HARs achieve a higher accuracy especially for scenarios where multiple appliances are active at the same time. The reason for the poor result of the HARs when multiple devices are active at the same time, is not apparent at first glance. In order to investigate this behavior more closely, a mathematical investigation is carried out. A nonlinear load can be described as a nonlinear two-port [104, p.42-44]. The current $i_g(t)$ can be represented as a power series of the grid voltage $v_g(t)$ as described in

$$i_g(t) = \sum_{n=1}^{\infty} a_n \cdot v_g^n(t) = a_1 \cdot v_g(t) + a_2 \cdot v_g^2(t) + a_3 \cdot v_g^3(t) + \dots, \quad (9.28)$$

where a_n are the coefficients. During a Steady State (SSt), the voltage signal can be assumed as a periodical signal. With the limitation that voltage $v_g(t)$ is a periodic signal, $v_g(t)$ can be described as a sum of sinusoidal signals with different frequencies and amplitudes.

Due to nonlinear loads connected to the power grid, $v_g(t)$ deviates from an ideal sinusoidal shape. Thus $v_g(t)$ can be described by multiple sinusoidal curves. The following example based on [104, p.42-43], $v_g(t)$ consists of two sinusoidal signals with the amplitudes V_1 and V_2 and the frequencies $\omega_1 = 2\pi f_1$ and $\omega_2 = 2\pi f_2$

$$v_g(t) = V_1 \cdot \sin(\omega_1 \cdot t) + V_2 \cdot \sin(\omega_2 \cdot t) \quad (9.29)$$

Substituting Equ. (9.29) in. (9.28) and applying the addition theorems, one obtains:

$$\begin{aligned} i_g(t) &= \frac{1}{2} \cdot a_2 \cdot (V_1^2 + V_2^2) && \text{DC component} && (9.30) \\ &+ \left(a_1 \cdot V_1 + \frac{3}{4} \cdot a_3 \cdot V_1^3 + \frac{3}{2} \cdot a_3 \cdot V_1 V_2^2 \right) \cdot \sin(\omega_1 \cdot t) && \text{Fundamental} \\ &+ \left(a_1 \cdot V_2 + \frac{3}{4} \cdot a_3 \cdot V_2^3 + \frac{3}{2} \cdot a_3 \cdot V_1^2 V_2 \right) \cdot \sin(\omega_2 \cdot t) \\ &- \frac{1}{2} \cdot a_2 \cdot V_1^2 \cdot \cos(2 \cdot \omega_1 \cdot t) && \text{2nd HAR} \\ &- \frac{1}{2} \cdot a_2 \cdot V_2^2 \cdot \cos(2 \cdot \omega_2 \cdot t) \\ &+ a_2 \cdot V_1 \cdot V_2 \cdot \cos((\omega_2 - \omega_1) \cdot t) && \text{Intermodulation products} \\ &- a_2 \cdot V_1 \cdot V_2 \cdot \cos((\omega_2 + \omega_1) \cdot t) && \text{2nd order} \\ &- \frac{1}{4} \cdot a_3 \cdot V_1^3 \cdot \sin(3 \cdot \omega_1 \cdot t) && \text{3rd HAR} \\ &- \frac{1}{4} \cdot a_3 \cdot V_2^3 \cdot \sin(3 \cdot \omega_2 \cdot t) \\ &+ \frac{3}{4} a_3 \cdot V_1^2 \cdot V_2 \cdot \sin((2 \cdot \omega_1 - \omega_2) \cdot t) && \text{Intermodulation products} \\ &- \frac{3}{4} a_3 \cdot V_1^2 \cdot V_2 \cdot \sin((2 \cdot \omega_1 + \omega_2) \cdot t) && \text{3rd order} \\ &+ \frac{3}{4} a_3 \cdot V_1 \cdot V_2^2 \cdot \sin((2 \cdot \omega_2 - \omega_1) \cdot t) \\ &- \frac{3}{4} a_3 \cdot V_1 \cdot V_2^2 \cdot \sin((2 \cdot \omega_2 + \omega_1) \cdot t) \\ &\dots \end{aligned}$$

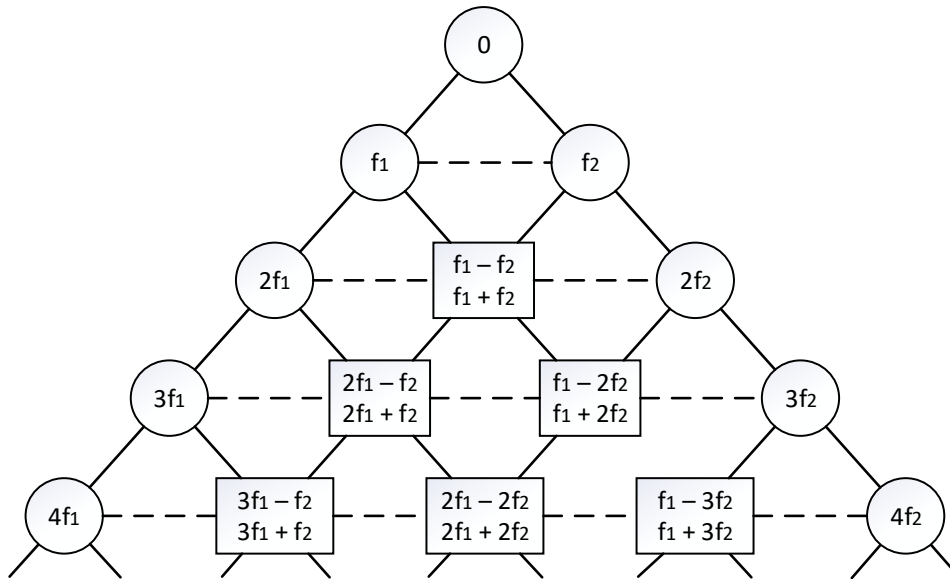


Figure 9.33: Pascal's triangle to form the combination frequencies. Modified from [104, p.44]

Here, the series is aborted after the 3rd order. Equ. (9.30) visualizes that the signal can be divided in Direct Current (DC) component, HARs with a multiple of the fundamental frequency and so-called intermodulation products in sums and differences of the fundamental frequencies and HARs.

Fig. 9.33 visualize the occurring frequencies which follow the educational law of Pascal's triangle. Provided that v_g consists of only one sinusoidal (or $V_1 = 0$ or $V_2 = 0$), there is no intermodulation product. Without DC component, the signal can be described by the HARs.

In practice, the shape of the mains voltage differs from the shape of a sine wave. The supply voltage consists of different sinusoidal oscillations with different frequencies. However, assuming that the waveform is constantly repeated after each period of the mains voltage (v_g), the frequency components can only be multiples of the mains frequency.

$$n \cdot f = f_n \quad (9.31)$$

Consequently, in the steady state only frequencies with integer numbers of multiples of the mains frequency can occur. Additionally, this is the case for appliances with a nonlinear power consumption too. In this case, the intermodulation products in Equ. (9.30) and Fig. 9.33 have the same frequency as the HARs.

This leads to two main results. On the one hand, all information on the voltage or current signal in the stationary state can be represented with HARs. On the other hand the occurring multiples of the mains frequency do not only depend on the properties of the appliance but also on the shape of the voltage signal of the grid. This means that for example, HARs can be measured even if only ohmic loads are operated in a household. The reason for this is the mains voltage being not a ideal sinus. This is visualized in Fig. 9.34 where the structure of the power grid is shown. The focus of NILM applications is on households and small businesses. These are connected to the same transformer. When HARs are transmitted from one household or company to the mains voltage, they affect all other households and companies. The main focus of applications for NILM are households and small businesses, since several households and small businesses are connected to the same transformer, HARs which affect the supply voltage of other appliances by nonlinear loads.

Depending on the distance of the different households or businesses to the transformer, this effect is more or less pronounced. The closer a device, which influences the form of the voltage signal, is placed to the measured household, the greater is the impact of the HARs. In the worst case, this device is located in the same household. Since the line resistance to the transformer

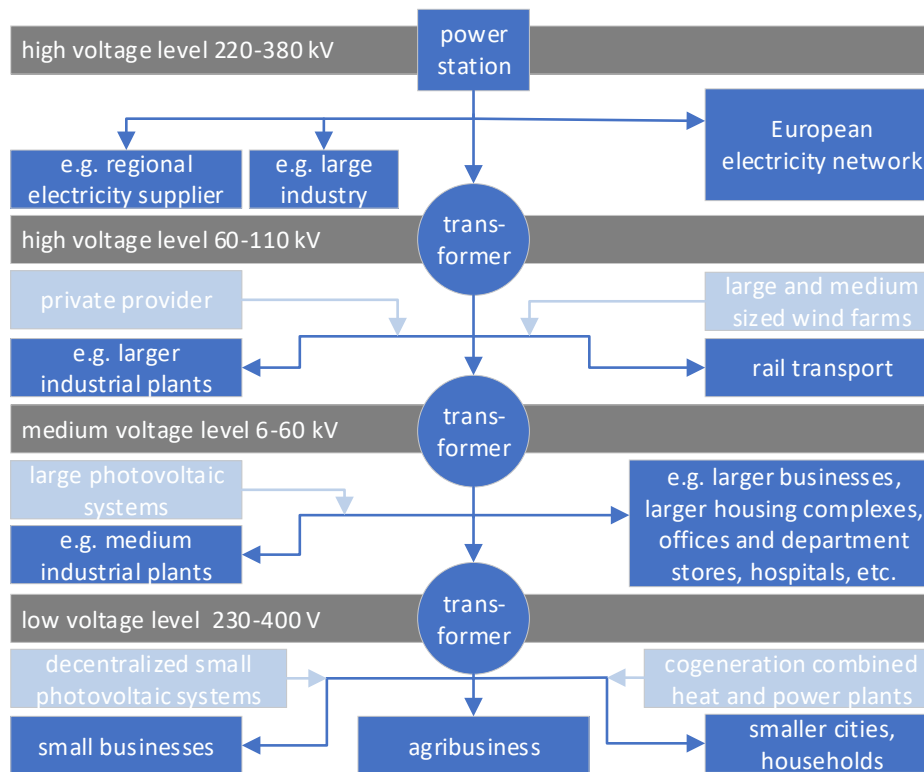


Figure 9.34: Structure of the power grid. Modified from [48]

cannot be neglected, these negative effects are greater if the measured household is connected to the transformer via a long power line.

For these reasons, the classification results are degraded when using HARs as features in case of multiple appliances that are operated simultaneously. An additional disadvantage of HARs as feature for the classification in NILM is the dependence of the results shown to the utilized number of HARs. To find the optimum, the training effort is multiplied by the number of HARs which is in the investigated case 39. In case of a changing appliances, setting the optimal number of HARs may change, leading to a higher computational training effort. For the compared FIT-PS waveform, all dimensions have been used. A specific selection of dimensions could improve the results but also needs to be adapted, like the number of HARs, to each measurement scenario.

9.9 Discussion

In this work, different classification methods have been applied using FIT-PS as feature and the power features P and Q , with and without HARs. The decision for a classification method is, in addition to the feature which is used, also crucially influenced by other boundary conditions. In NILM, usually a limited set of training data is available. Compared to other NILM data sets, the HELD1 data set includes with 100 switch-on and switch-off events a high number of training data. Nevertheless, compared to other classification problems like handwriting recognition where usually more than 10,000 training samples of each class are available, the number of training samples used for this application is still low.

The Bayes classifier assumes that the individual features respectively their dimensions are independent. Especially with the FIT-PS feature, for most classes a comparatively high correlation of the individual dimensions can be assumed. The assumption of a normal distribution seems to be incorrect for the power features since other distribution functions calculated during an optimization process lead to better results. To define a correct distribution function is difficult because of the low number of training examples. Under all features, the Bayes classifier leads to

the worst results.

The SVM is known to be able to deliver good results with high-dimensional data. By far the best result is achieved with FIT-PS. However, the use of additional HARs for P and Q leads to an improvement, but the number of HARs needs to be determined. A use of all available HARs degrades the classification result. Another advantage of SVM is the fast classification. However, this advantage is reduced by the number of devices that needs to be distinguished, whereby the SVM requires a comparatively long test time. To improve and therefore shorten the testing time, the one-against-one method instead of the one-against-all method could be used. The advantage is that instead of comparing one cluster with every other cluster individually, the comparison is performed with all other clusters at the same time. However, in general, this leads to worse results because the non classified area is extended. Especially for the high-dimensional FIT-PS waveform, the training time is comparatively high. Another difficulty is to find a suitable kernel, since this depends on the individual appliances. Adding additional appliances may require searching for a new kernel.

The best results for P , Q , and HARs, as well as for FIT-PS is achieved with the kNN. Since the kNN does not require any training, the training samples only need to be stored in the memory. The problem of most other classification methods for a low number of training samples is advantageous for the kNN. Because of the memory storage and the required classification effort, high numbers of training data are problematical for kNN. Therefore, in addition to the best results, the least computational effort for training and testing is required when using the kNN. While the number of dimensions of the features is negligible during the training, significantly more time is required when testing with the FIT-PS feature. Furthermore, it must be mentioned that several iterations of the training are required to find a good value for k .

The classification with the FFNN does also lead to good results. But compared to the other classification methods, a high training time is required. To achieve the presented results, 100 iterations of the training, with randomly chosen start parameters, had been run. Since the test time is comparatively short, the FFNN is advantageous for supervised classification where no or only limited number of new appliances are added to the measurement environment. For a changed measuring environment, the classifier must be relearned. This could be realized more easily in the cloud.

The advantage of the LSTM network is that transient state information can be used for classification. The TSt has a higher variation, compared to the SSt, since it may depend if an appliance is switched on at the zero crossing of the voltage or at the positive or negative peak of the power or in between. Therefore, more training samples are required for this classifier and feature. The results achieved with 100 training samples are slightly better compared to the FFNN which only uses the SSt. In addition to the problem of finding a suitable network structure, the problem of very high computational effort for the LSTM network during training is disadvantageous.

The search for suitable datasets for the classification in context of NILM is challenging. The features investigated during this work are based on the originally measured current and voltage signal, sampled with a frequency of at least 1 kHz. To generate training and test features, correct reference data are required. Since the reference data often have a much lower sampling rate of several seconds, the correct generation of features is challenging. Additionally, a certain number of switching events is required in order to generate training and test data. The Building-Level fully-labeled dataset for Electricity Disaggregation (BLUED) dataset [4], for example, has only three appliances with more than 100 switching events. Therefore, the presented classification method was only applied to the HELD1 dataset. Since this dataset has, in contrast to other laboratory datasets, also simultaneously active appliances, the difference to real world measurements is mainly due to the probability of occurring events.

9.10 Conclusion

An investigation of different supervised classification methods and different features is carried out. Therefore, the HELD1 dataset contains 18 different appliances. To find the best parameter setting of the individual classification method, several iterations of training are required. The best result is achieved by the FIT-PS feature independent of the classification method. Especially when multiple appliances are active at the same time, FIT-PS is advantageous compared to the use of HARs. The kNN leads to the best classification results of all compared classifiers. Furthermore, the training and test time of the kNN is shorter compared to the other classification methods. Hypothesis 5 in Chap. 3 “The new signal representation achieves better classification rates compared to standard waveforms.” can be confirmed. Therefore, hypothesis 1 “Transforming the current signal according to the phase shift between current and voltage produces a NILM specific waveform that contains the relevant information of the current and voltage signal.” can be confirmed too, since the results achieved with FIT-PS for event detection, as well as, for classification are better compared to standard waveforms or features.

Chapter 10 | Summary & Perspectives

10.1 Summary

The present work provides new and innovative approaches for the analysis of electrical energy consumption in order to detect electrical appliances robustly and reliably. The separation of the measured total current (including the voltage) at the energy meter into individual partial currents, which are assigned to individual devices, is called disaggregation and is known as Non-Intrusive Load Monitoring (NILM). The basic approach of disaggregation is to build a linear processing chain consisting of signal representation, event recognition, feature extraction, classification and tracking.

Numerous solutions for NILM are discussed in the literature. On closer inspection, however, it becomes apparent that the existing methods have different limitations. Typical limitations of the methods are a minimum energy consumption of the devices for event detection or the number of devices that can be distinguished in the classification. The problem of simultaneously active devices is rarely considered, although this is of crucial importance for the robustness of the methods.

As part of this work, an analysis of the existing approaches was carried out in order to gain a basic understanding. The main result was the strong orientation of most of the approaches towards the first published NILM solution by Hart et al. Essential differentiation features are mainly found in the methods of classification.

The signal types used have not usually been developed specifically for the NILM problem. This leads to reduced information content, lack of adaptation to the physical conditions of the electrical network and unfavorable representation of the signal characteristics. A combination of several signal types is used to reduce these disadvantages. However, this in turn leads to the problem of the appropriate weighting of the different signal forms. In addition, there are clear dependencies on the equipment used, which in turn leads to necessary optimizations of the weighting. In addition to this suboptimal signal representation, many of the proposed signal forms, such as HARMONIC (HAR) waves, have a noticeable complexity that leads to high computing power. Considering that disaggregation is a concept that should be integrated into a smart meter, high complexity is not an advantage.

For these reasons, a new signal representation was developed in this work, which is advantageous for the individual work steps of NILM. It is low in complexity, adapted to the problem of disaggregation, has no loss of information and nevertheless allows a clear representation of the signal.

The concept of Frequency Invariant Transformation of Periodic Signals (FIT-PS) is based on the idea to convert the current signal into a multidimensional signal in phase correct relation to the voltage. The new signal is invariant to any kind of fluctuations of the mains frequency or the jittering of the sampling frequency of the measuring electronics. The multi-dimensionality of the FIT-PS signal allows a more transparent, robust and cost-reduced evaluation of the subsequent event detection, feature extraction and classification.

This work focuses on the evaluation of FIT-PS for the stages event detection, feature extraction and classification. This is due to the fact that in many cases the information which device is switched on or off at a certain time is of central importance. In the first step the signal transformation from current and voltage signal to FIT-PS signal was investigated. Three approaches

for the realization of the invariance of the frequency fluctuations were considered. Different optimizations regarding the accuracy and the computational effort were carried out.

For the evaluation of FIT-PS in the stages of event detection and classification, it was necessary to generate a new data set Home Equipment Laboratory Dataset 1 (HELD1) that was published. The necessity resulted from the fact that in the publicly accessible data sets the reference data and the number of events per device did not meet the requirements. Using FIT-PS, accuracies of more than 98 % were achieved in event detection, so that the slightest inaccuracies in the reference data were clearly reflected in the accuracy. The HELD1 data set was recorded under laboratory conditions, so that the reference data of the 18 different devices are completely and correctly available. This is the first laboratory dataset for NILM that simultaneously contains several active appliances. There are 100 individual training events available for each device. The data set contains a total of 13,200 test events.

A new two-stage event detection method based on FIT-PS was developed and compared with other signal forms. This was necessary because previous event detection methods were usually based on a one-dimensional signal. For comparability with literature results, the Building-Level fully-labeled dataset for Electricity Disaggregation (BLUED) data set was used in addition to the HELD1 data set. The optimization of the parameters of the newly developed event detection method was carried out and compared with the two optimization methods Simulated Annealing (SA) and Generalized Pattern Search (GPS). These optimization methods provide an easy way to adapt the parameters to other data sets.

Following the event detection, the feature extraction was examined with the signal representation FIT-PS. FIT-PS provides a simple and computationally efficient feature extraction in Steady State (SSt) and Transient State (TSt). In addition to the FIT-PS features, the features of the active and reactive power, as well as the HARs were generated for evaluation purposes.

Several standard classifiers such as k-Nearest Neighbor (kNN), Support Vector Machine (SVM), Bayes and neuronal networks were considered for the final assessment and investigation of the suitability of FIT-PS. The classification methods are based on the comparison of FIT-PS with existing methods from the literature. The highest accuracy of all classification methods was achieved with FIT-PS. In addition to the Steady State Feature (SSF), the Transient State Feature (TSF) calculated by FIT-PS was also used for classification on an Long Short-Term Memory (LSTM) neural network. Especially for simultaneously active appliances, FIT-PS can achieve significantly better classification results than with the previously available features.

Within the scope of this work, a new signal representation FIT-PS and a new very robust event detection method were developed. It could be shown that FIT-PS is a suitable and advantageous signal representation in the context of disaggregation, since much better results could be achieved in both event detection and classification than with the standard methods.

10.2 Perspectives

NILM is a complex subject area with many different tasks. Therefore, it is not possible to work on all questions simultaneously. In this work, the focus was on the development of suitable signal representation and its implementation in the area of event detection and classification. This section outlines areas for further research.

The here presented FIT-PS waveform is implemented on a low-cost embedded system. A further task is to implement the event detection as well as the classification on the micro-controller. Since the measurements of the current and voltage depend significantly from the measuring instrument which is used. The properties of the measurement hardware are essential for the signal form. For example, if a low-pass filter is used, unbalanced loads are presented differently than without low-pass filters. Therefore, it is necessary to generate new measurements. It would also be desirable to have a conversion method that allows measurements from different measurement systems to be standardized.

The FIT-PS features improve the results of supervised classification in NILM. First approaches with unsupervised methods are promising. A further task is to apply FIT-PS on suitable unsupervised classification methods and compare them with other waveforms. Additionally, methods need to be investigated which allow identifying complex appliances.

The tracking is a necessary part of the classical NILM approach. The extent to which tracking is required, or whether it is sufficient to classify the appliances depends on the specific problem which needs to be solved.

An additional advantage of the FIT-PS signal representation is that individual current signals can be added, since the position of the samples within a period is independent of frequency variations of the grid. This allows generating simulated datasets out of individual measurements. Even the transient state of the different appliances can be considered directly. One advantage of simulated data is that different scenarios can be simulated with one set of individual measurements. The reference data are ideally located to the events, and the error rate is meager. Furthermore, scenarios which are difficult to build in practice can be simulated efficiently. The simultaneous switching on or off of devices poses problems for the classification. A real measuring environment which turns on and off more appliances at the same time or with small intervals is expensive to implement. Besides, some consumers do not turn on immediately after they are powered up. With simulated FIT-PS waveforms, the generation of simultaneously events or events which are close behind each other can be realized quickly. For further research, a simulation environment which allows creating different scenarios would be advantageous.

Appendix

Appendix A | Exemplary Representation of FIT-PS Waveform

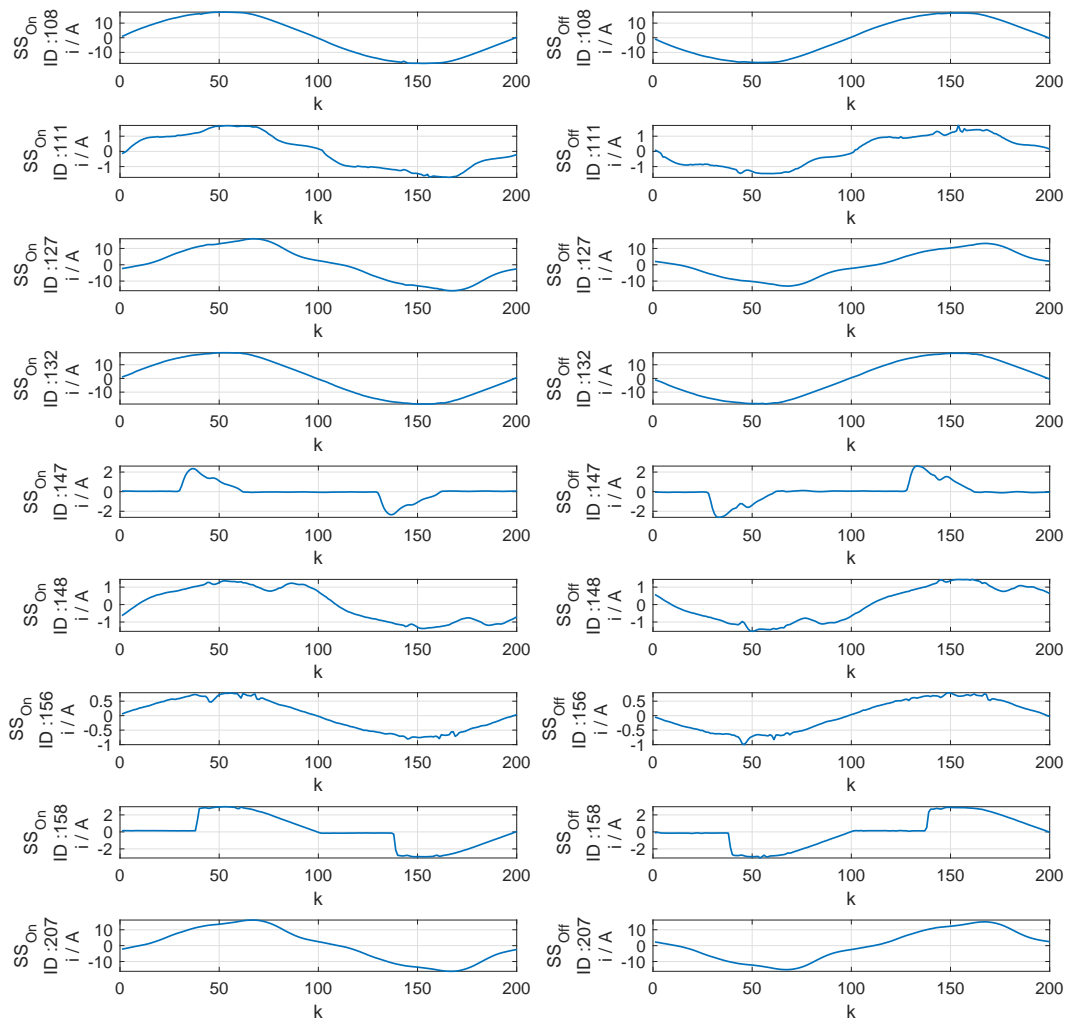


Figure A.1: Using the current amplitude transformed by FIT-PS as SSF of different devices of the BLUED dataset.

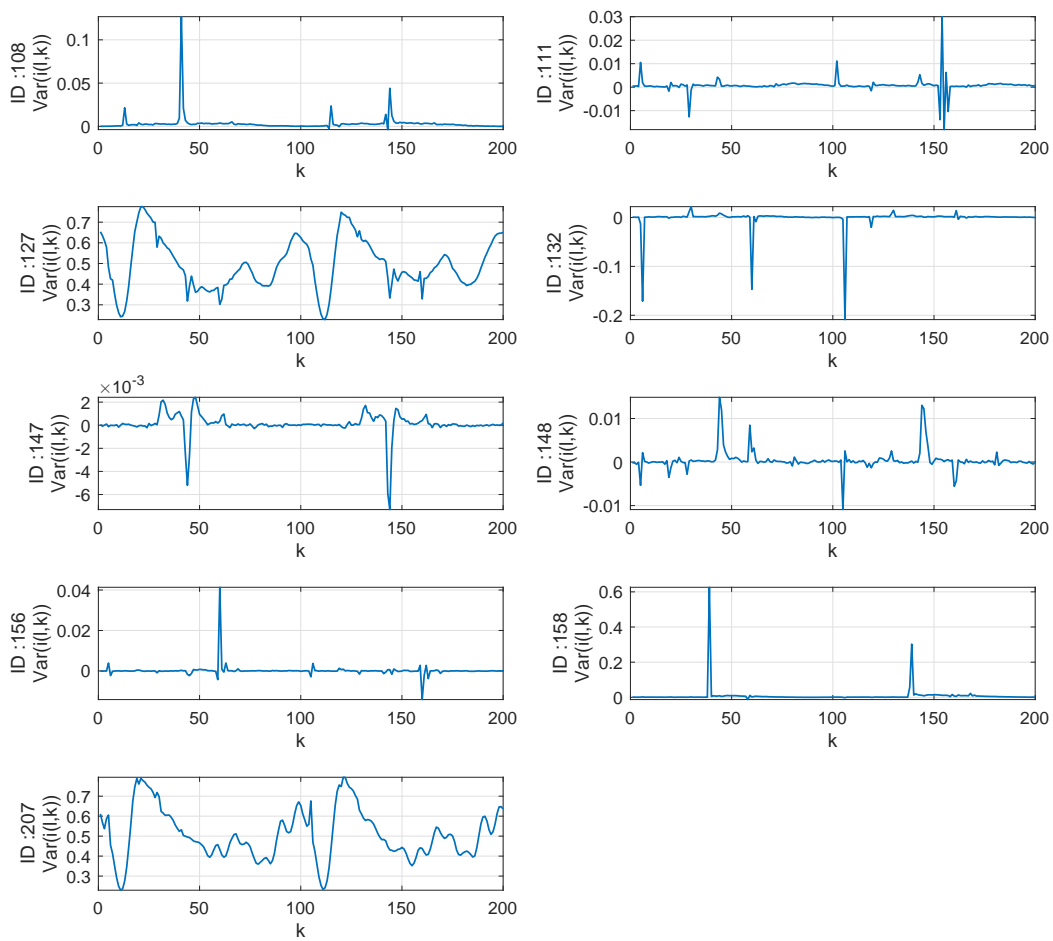


Figure A.2: Using the variance of the current amplitude transformed by FIT-PS as SSF of different devices of the BLUED dataset.

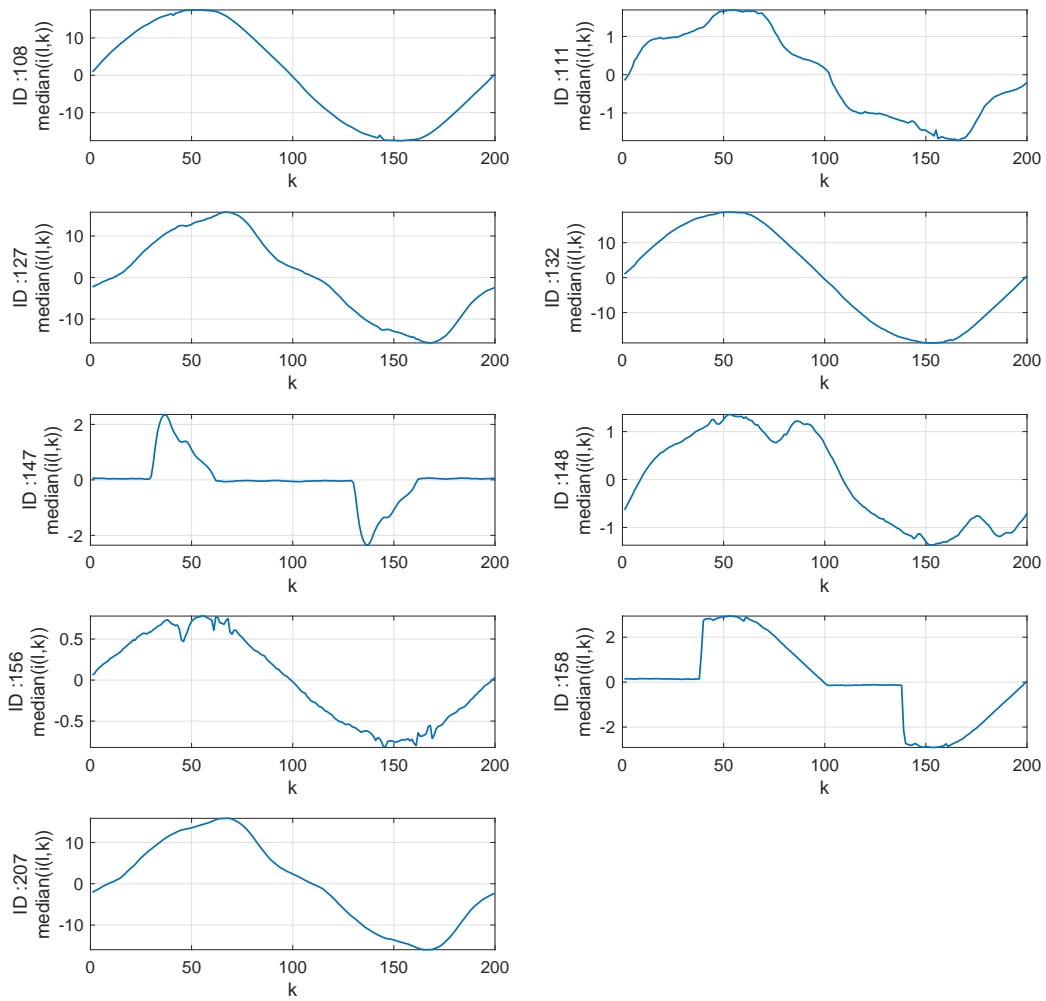


Figure A.3: Using the median of the current amplitude transformed by FIT-PS as SSF of different devices of the BLUED dataset

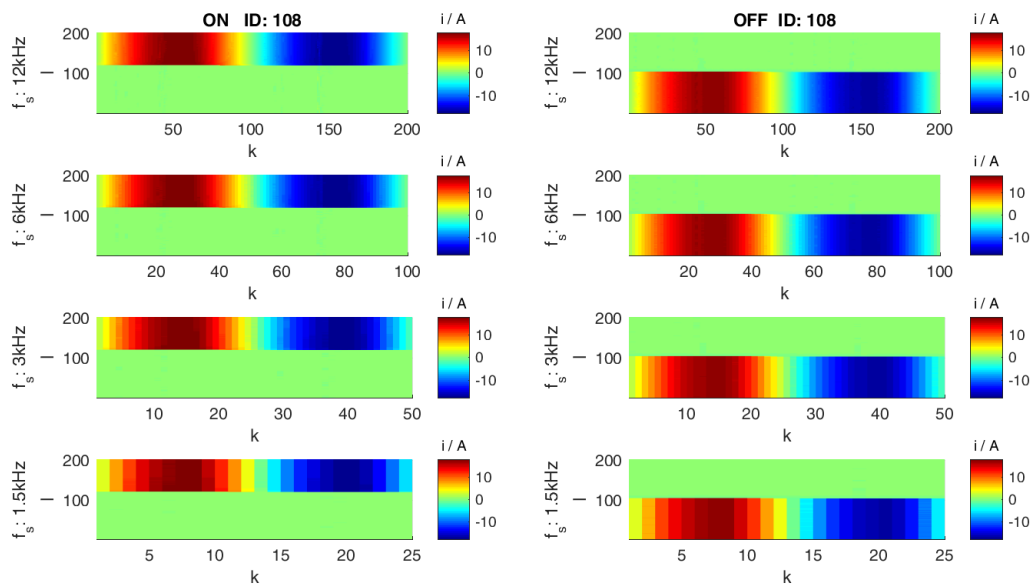


Figure A.4: Device ID108 from the BLUED dataset transformed by FIT-PS using different sample rates

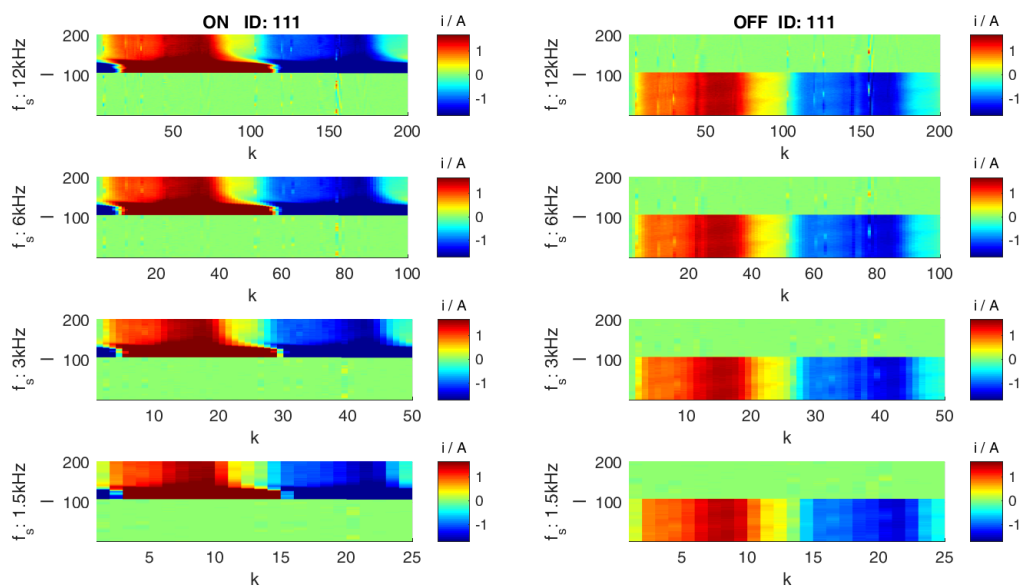


Figure A.5: Device ID111 from the BLUED dataset transformed by FIT-PS using different sample rates

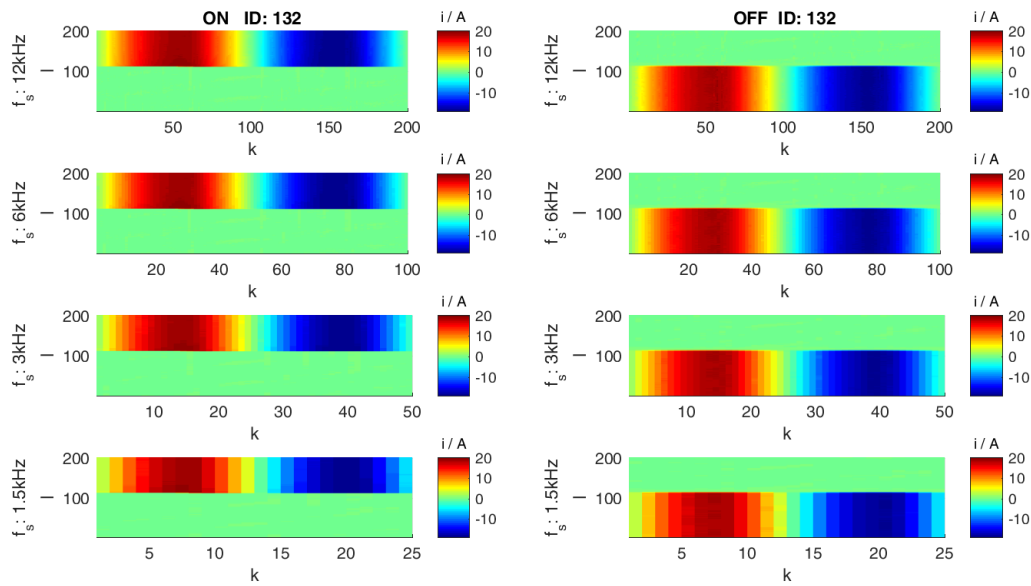


Figure A.6: Device ID132 from the BLUED dataset transformed by FIT-PS using different sample rates

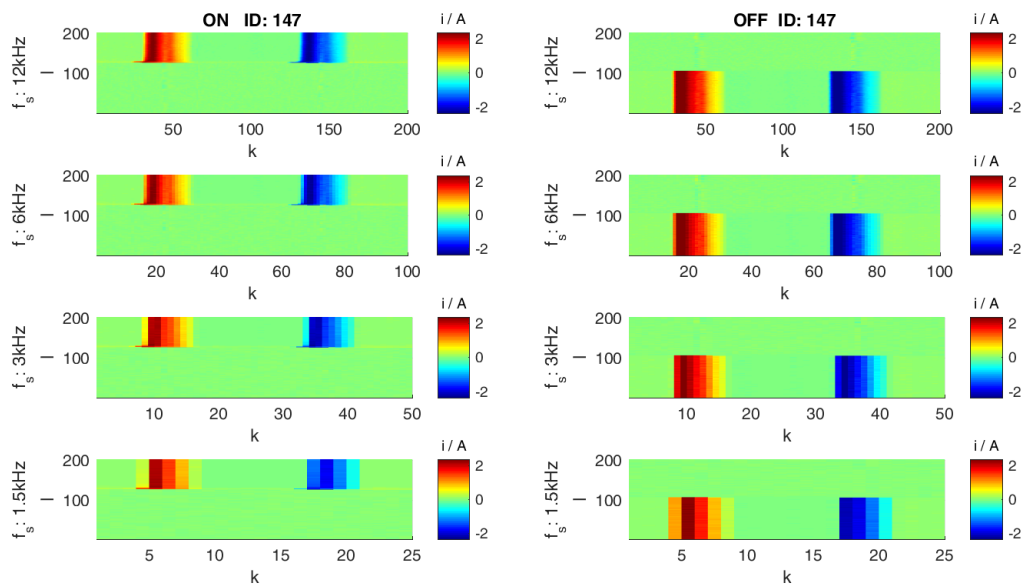


Figure A.7: Device ID147 from the BLUED dataset transformed by FIT-PS using different sample rates

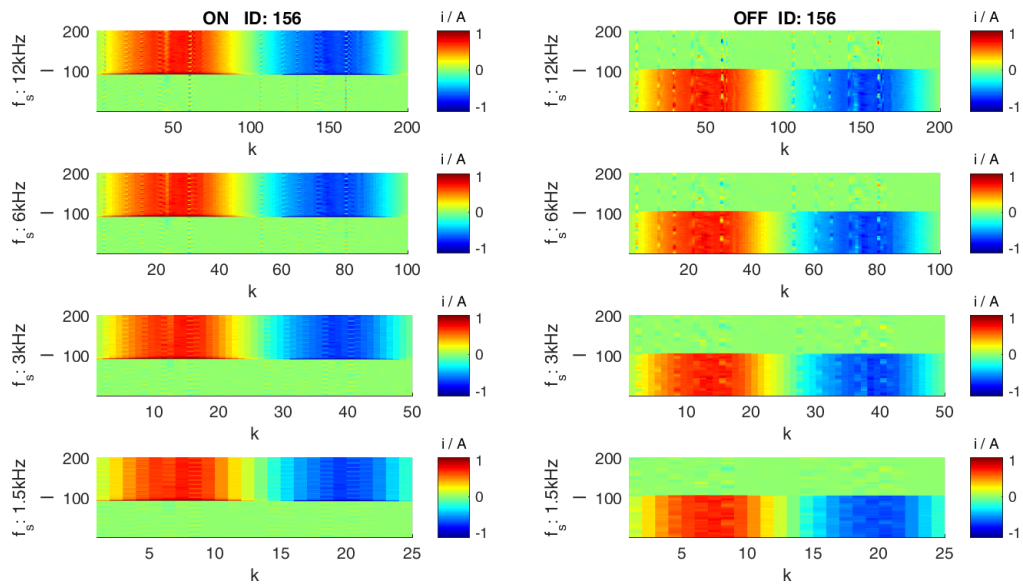


Figure A.8: Device ID156 from the BLUED dataset transformed by FIT-PS using different sample rates

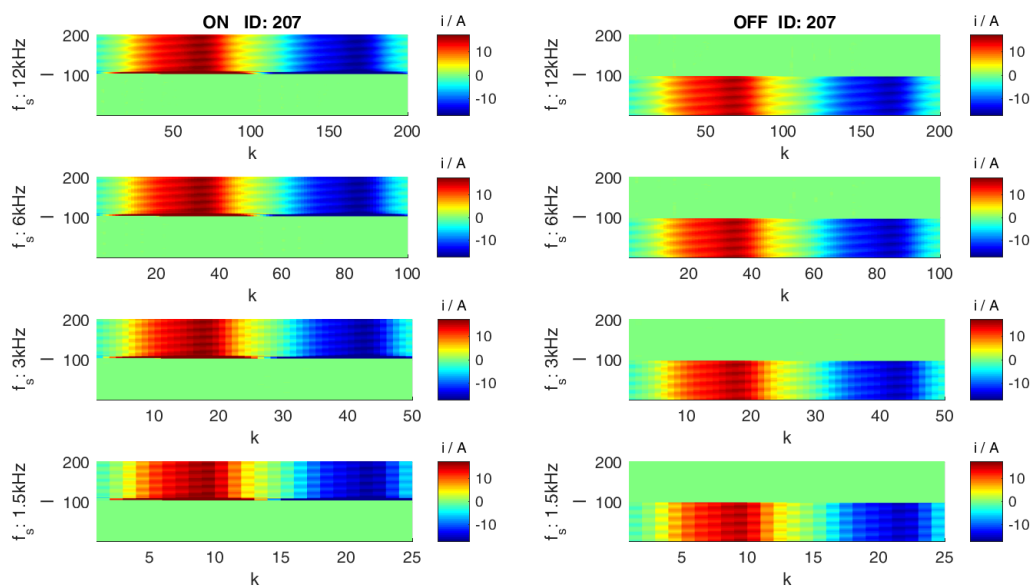


Figure A.9: Device ID207 from the BLUED dataset transformed by FIT-PS using different sample rates

Lists

List of Acronyms

AAL	Ambient Assisted Living	ISS	Institute for Smart Systems
AC	Alternating Current	kNN	k-Nearest Neighbor
ADC	Analog-to-Digital Converter	LED	Light-Emitting Diodes
ANN	Artificial Neural Network	LSB	Least Significant Bit
BLUED	Building-Level fully-labeled dataset for Electricity Disaggregation	LSTM	Long Short-Term Memory
CW	Current Waveform	MIPS	Modélisation, Intelligence, Processus et Systèmes
DC	Direct Current	MIT	Massachusetts Institute of Technology
DFT	Discrete Fourier Transform	MOSFET	Metal-Oxide-Semiconductor Field-Effect Transistor
EM	Expectation Maximization	NetCDF	NETwork Common Data Format
EMI	Electro-Magnetic Interference	NIALM	Non-Intrusive Appliance Load Monitoring
ENSM	Ecole Nationale Supérieure de Mécanique et des Microtechniques de Besançon	NILM	Non-Intrusive Load Monitoring
ENTSO-E	European Network of Transmission System Operators for Electricity	PC	Personal Computer
EPRI	Electric Power Research Institute	PCA	Principal Component Analysis
FC	Fully Connected	PLL	Phase-Locked Loop
FFNN	FeedForward Neural Network	REDD	Reference Energy Disaggregation Data Set
FFT	Fast Fourier Transform	RMS	Root Mean Square
FIT-PS	Frequency Invariant Transformation of Periodic Signals	SA	Simulated Annealing
FN	False Negative	SMPS	Switch Mode Power Supply
FP	False Positive	SSA	Steady State Approximation
FT	Fourier Transform	SSt	Steady State
GLR	General Likelihood Ratio	SSF	Steady State Feature
GPS	Generalized Pattern Search	SVM	Support Vector Machine
G2Elab	Laboratoire de Génie Electrique de Grenoble	THD	Total Harmonic Distortion
HAR	HARmonic	TP	True Positive
HARV	HARmonic Voltages	TSF	Transient State Feature
HELD1	Home Equipment Laboratory Dataset 1	TSt	Transient State
HFU	Hochschule Furtwangen University	TV	TeleVision
HMM	Hidden Markov Model	UCAR	University Corporation for Atmospheric Research
HVAC	Heating, Ventilation and Air Conditioning	UHA	Université de Haute-Alsace
IAW	Instantaneous Admittance Waveform	UK-DALE	United Kingdom Domestic Appliance-Level Electricity dataset
IPW	instantaneous power waveform	UTBM	Université de Technologie de Belfort-Montbéliard
IRIMAS	Institut de Recherche en Informatique, Mathématiques, Automatique et Signal	VSD	Variable Speed Drive

List of Figures

1.1	Intrusive monitoring	7
1.2	Non-intrusive monitoring	7
2.1	Classic approach of the processing chain of NILM	12
2.2	Number of publications concerning NILM The data have been generated by using the software Publish or Perish, product version: 5.25.2.6208.	12
2.3	Common information flow in NILM. In comparison to the usual linear presentation of the processing chain of the NILM tasks, see Fig. 2.1, the information flow is not purely linear in a row.	18
2.4	Presentation of the different error calculations in NILM	19
2.5	Comparison of the periodical, original current signal ($i(t)$) to the non-periodical complex power (S) signal form	25
2.6	Radio (ID:003)	31
2.7	Vacuum Cleaner (ID:005)	31
2.8	Refrigerator (ID:013)	31
2.9	Kettle (ID:017)	31
2.10	The different signal states (SSt and TSt) of NILM are shown on an active power signal	34
2.11	Example of possible results for the event detection; detected events are marked green.	34
4.1	Block diagram of the measurement system; Analog-to-Digital Converter (ADC), appliance (Ap.), phase 1 (L1), neutral line (N), Personal Computer (PC) [67]	44
4.2	Illustration of random selected individual appliances with corresponding device ID [67]	45
4.3	Active and reactive power distribution of the appliances [67]	46
4.4	Detailed illustration of the lower power range from Fig. 4.3 [67]	47
5.1	In comparison to the usual linear presentation of the processing chain of the NILM tasks, see Fig. 2.1, the information flow is not ordered linearly.	50
5.2	$f_s = 200$ Hz	55
5.3	$f_s = 1$ kHz	55
5.4	$f_s = 10$ kHz	55
5.5	Graphical representation of FIT-PS [68]. a) finding zero crossings, b) calculating new sampling positions, c) re-sampling and interpolation of the current signal based on the new sampling positions, d) periodical signal representation	56
5.6	FIT-PS of a signal with 20 samples per period showing a SSt of a refrigerator	57
5.7	2D signal visualization of Fig. 5.6. The current amplitude is colored.	57
5.8	FIT-PS signal representation of the individual dimensions.	58
5.9	On the left a sine with 200 equidistantly samples and on the right a histogram of the samples from the sinusoidal signal is visualized.	59
5.10	FIT-PS waveform with 20 samples per period SSt z_i : device one; TSt: switching on procedure of device two; SSt z_{i+1} : device one and device two (refrigerator).	59
5.11	FIT-PS waveform with 200 sample points per period; SSt z_i : device one; TSt z_{i+1} : switching on procedure of a refrigerator; SSt z_{i+2} : device one and two. Published in [66].	60
5.12	FIT-PS waveform with 200 sample points per period illustrating a turn on TSt of a refrigerator.	60
5.13	Representation of a downsampling without losing the signal shape, during a SSt. 60 Hz periodic signal (continuous lines) under sampled at 57 Hz sample frequency (black circle).	62
5.14	Comparison between the variance of the signal (gray), sampled with 1.2 kHz and the downsampled signal (red).	62
5.15	Comparison between the variance of the signal (gray), sampled with 12 kHz and the downsampled signal (red).	63
5.16	Different sampling rates of a turning on (left side) and off (right side) process of an energy saving lamp, from BLUED dataset [4], presented with FIT-PS. Sampling rates from 12 kHz to 1.2 kHz are illustrated. Thereby the numbers of dimensions k are varying from 200 to 25 while the number of 200 periods in l direction is constant. The amplitude of the current is represented colored.	64
5.17	Comparison of the active power waveform P (top) with the FIT-PS representation (bottom). The waveform is selected from the HELD1 dataset (measurement 0242); events are marked with red 'x'.	66
5.18	Detailed comparison of the active power waveform P (top) with the FIT-PS presentation (bottom). Signal section of the HELD1 dataset (measurement 0242).	67
5.19	Detailed comparison of the derivative of the apparent power waveform P with the derivative of the FIT-PS presentation. Signal section of the HELD1 dataset (measurement 0242).	67
6.1	Comparison of the three FIT-PS methods in the time domain. The waveforms represent a switching on and off cycle of the heat gun (setting2) from the HELD1 dataset.	73

6.2	Frequency spectrum of the original current signal compared with the different FIT-PS methods (vertical axis clipped at 0.035)	74
6.3	Section of the spectrum of the original current signal compared with the different FIT-PS methods	74
6.4	Current signal of the measurement which was used for the analysis of the different FIT-PS approaches	75
6.5	Results of the different FIT-PS approaches for one and up to four active appliances. green: FIT-PS with simple zero detection (one value before and one after the zero crossing); blue: FIT-PS with curve fitting for zero detection; red: FIT-PS without interpolation	76
6.6	Results of the different FIT-PS approaches for up to six active appliances and an overview of the three different scenarios. green: FIT-PS with simple zero detection (one value before and one after the zero crossing); blue: FIT-PS with curve fitting for zero detection; red: FIT-PS without interpolation	76
7.1	Event detection applied on FIT-PS	80
7.2	First stage of the event detection. The derivative is calculated from each dimension k of the FIT-PS concerted current signal over the periods l	81
7.3	State selection V1: Visualizing the calculation of the start and end sample of a transient. The proposed event from stage one is marked blue whereas the calculated start and end positions are marked in green.	84
7.4	State selection V2: Exemplary representation of $\hat{\sigma}^2$, visualizing a transition between a SSts and a TSts. The yellow windows w_s and w_e with the length G_1 start at Φ_1	86
7.5	State selection V1: Exemplary representation of $\hat{\sigma}^2$, visualizing a transition between a SSt and a TSt. The windows have been moved by n_1 or n_2 values until all values are below the threshold Tr_{TSt}	86
7.6	Second stage of the event detection. The difference before and after the suspected event is calculated.	87
7.7	Current peaks at a SSt measured with high sampling rate and converted into FIT-PS signal representation. [66]	87
7.8	3D view of Fig. 7.7. [66]	88
7.9	Second stage of the event detection. The difference before and after the suspected event is calculated. Modified from [143].	89
7.10	FIT-PS signal representation of a radio (ID003) of the HELD1 dataset	90
7.11	Difference of the envelope of Fig. 7.10	91
7.12	Envelope of a fluorescent lamp; maximum of the envelope $i_{env}[l, 1]$ (orange), minimum of the envelope $i_{env}[l, 2]$ (red), absolute difference i_{envD} (blue).	91
7.13	Representation of the precision dependency of two parameters of the event detection using the Principal Component Analysis (PCA). The results are based on phase A of the BLUEED dataset. Modified from [70].	92
7.14	Representation of the recall dependency of two parameters of the event detection using the PCA. The results are based on phase A of the BLUEED dataset. Modified from [70].	93
7.15	Comparison of the results of the event detection with different numbers of dimensions after PCA.	95
7.16	Results of the test measurements of the HELD1 dataset achieved with the parameters from 100 training iterations of the GPS or ten training parameters of SA. Recall: green, precision: blue; corresponding result to the best training iteration are marked pink respectively yellow with a dotted line; the median is marked with a solid line.	99
8.1	SSt and TSt illustrated on a active power waveform.	104
8.2	Illustration of the calculation of SSF from a power waveform.	104
8.3	Different SSts of the FIT-PS waveform. The off events are marked red while on events are marked with green line.	105
8.4	FIT-PS SSFs of 18 appliances of the HELD1 dataset. 100 features of each device are plotted behind each other.	106
8.5	Simulation of a 180° change of the plug, of the fan.	107
8.6	Variance of the SSts of the FIT-PS waveform from 18 appliances of the HELD1 dataset.	108
8.7	TSts of the four appliances of the HELD1 dataset illustrated with i_{RMS}	110
8.8	Illustration of the TSts with FIT-PS signal representation, for all appliances of the HELD1 dataset.	111
9.1	Exemplary representation of a kNN with $k = 1$ and $k = 4$ neighbors applied on a two-dimensional classification problem with two classes.	115
9.2	Linear separation of two clusters in two dimensions	117
9.3	Linear separation of two clusters with a slip variable	118
9.4	Illustration of the kernel trick for SVM classification	118
9.5	Linear separation of four classes in the two dimensional space.	119
9.6	Model of an artificial neuron	120
9.7	Structure of a FeedForward Neural Network (FFNN) with one hidden layer	120
9.8	Structure of a LSTM cell	121
9.9	LSTM layer [68]	122
9.10	Complete structure of the LSTM net. [68]	122
9.11	Segmentation of the data for classification	123
9.12	Validation results of the kNN for different number of k and HARs	123

9.13	Validation results of the kNN for different features and a varying number of k -neighbors	124
9.14	Test results of the kNN for varying k and number of HARs	125
9.15	Test results of the kNN for varying number of k -neighbors	125
9.16	Validation results of the Bayes classifier for the feature combination P , Q , and HARs for varying number of HARs	126
9.17	Validation results of the Bayes classifier using different features	127
9.18	Test results of the Bayes classifier for the feature combination P , Q , and HARs for varying number of HARs	127
9.19	Test results of the Bayes classifier using different features	128
9.20	Validation results of the SVM for the feature combination P , Q , and HARs and a varying number of HARs	128
9.21	Validation results of the SVM different features	129
9.22	Test results of the SVM classifier for the feature combination P , Q , and HARs for varying number of HARs	130
9.23	Test results of the SVM using different features	130
9.24	Structure of the feed forward neural net. The number of input dimensions depends on whether P , Q , and HARs are used as feature or FIT-PS [68]	131
9.25	Validation results of a feed forward net using P , Q , and HARs for a different number of neurons. The limits of accuracy are calculated by 100 iterations with different initialization settings and different number of neurons. [68]	132
9.26	Validation result of a feed forward net using FIT-PS for a different number of neurons. The limits of accuracy are calculated by 100 iterations with different initialization settings and different number of neurons.[68]	132
9.27	Results of a FFNN using different features for the validation data 0003, 0116, 0201. The accuracy is calculated by 100 iterations with different initialization settings and different number of neurons.[68]	133
9.28	Comparison of validation results and test results on different initialization settings. The initialization settings are sorted according to increasing accuracy of the validation data. In general a higher accuracy of validation also leads to a higher accuracy of the test results. [68]	134
9.29	Accuracy of a FFNN using P and Q (green); P , Q , and HARs (blue); FIT-PS (red); LSTM net using FIT-PS (orange). The dotted lines represent the accuracy using the best initialization setting derived from the validation data. The solid lines represent the median of the accuracy of the test data.[68]	136
9.30	Best results of the different classifiers and features.	136
9.31	Training time of the kNN for different numbers of HAR and neighbors	137
9.32	Time for classification using the kNN for different numbers of HARs and neighbors	137
9.33	Pascal's triangle to form the combination frequencies. Modified from [104, p.44]	140
9.34	Structure of the power grid. Modified from [48]	141
A.1	Using the current amplitude transformed by FIT-PS as SSF of different devices of the BLUED dataset.	151
A.2	Using the variance of the current amplitude transformed by FIT-PS as SSF of different devices of the BLUED dataset.	152
A.3	Using the median of the current amplitude transformed by FIT-PS as SSF of different devices of the BLUED dataset	153
A.4	Device ID108 from the BLUED dataset transformed by FIT-PS using different sample rates	154
A.5	Device ID111 from the BLUED dataset transformed by FIT-PS using different sample rates	154
A.6	Device ID132 from the BLUED dataset transformed by FIT-PS using different sample rates	155
A.7	Device ID147 from the BLUED dataset transformed by FIT-PS using different sample rates	155
A.8	Device ID156 from the BLUED dataset transformed by FIT-PS using different sample rates	156
A.9	Device ID207 from the BLUED dataset transformed by FIT-PS using different sample rates	156

List of Tables

2.1	List of the most frequently cited publications from NILM (01.02.2017)	13
2.2	Comparison of NILM datasets	24
2.3	Values of individual HARs according to EN 50160: 2005 [29]	29
2.4	Comparison of NILM feature extraction approaches, modified from [87]	38
4.1	Device list of the HELD1 dataset [67]	45
4.2	Comparison of the HELD1 dataset [67] with commonly used NILM datasets	48
5.1	Comparison of the influence of voltage fluctuations of resistive and active load.	52
7.1	Detection performance BLUED phase A	96
7.2	Detection performance BLUED phase B	96
7.3	Detection performance BLUED phase A and B	96
7.4	Detection performance BLUED Phase A, with optimized parameters	97
7.5	Detection performance BLUED Phase B, with optimized parameters	97
7.6	Parameter setting of the different methods for phase A	97
7.7	Parameter setting of the different methods for phase B	97
7.8	Detection performance HELD1 dataset using FIT-PS waveform	98
7.9	Detection performance HELD1 dataset using P, Q , and S	98
7.10	Results of the optimization for the HELD1 dataset using FIT-PS (envelope)	98
7.11	Parameter corresponding to the results in Tab. 7.10	98
7.12	Parameter setting of the different methods for phase B	99
9.1	Conversion matrix of the classification results of measurement "0194" using P , Q , and HARs in a FFNN [68]	135
9.2	Conversion matrix of the classification results of measurement "0194" using FIT-PS in a FFNN [68]	135
9.3	Training time of the different classification methods and features in seconds	138
9.4	Test time of the test measurements of the HELD1 dataset in seconds	138

Bibliography

- [1] “Fluke 1738 Advanced Power Energy Logger.” [Online]. Available: <https://www.fluke.com/en-us/product/electrical-testing/power-and-energy-loggers/three-phase-power-and-energy-loggers/fluke-1738>
- [2] “Unidata | NetCDF,” 2018. [Online]. Available: <https://www.unidata.ucar.edu/software/netcdf/>
- [3] S. Abe, *Support vector machines for pattern classification*, ser. Advances in pattern recognition. London: Springer, 2005, oCLC: ocm57475934. [Online]. Available: <https://www.springer.com/de/book/9781846282195#otherversion=9781852339296>
- [4] K. Anderson, A. Ocneanu, D. Benitez, A. Rowe, and M. Berges, “BLUED: A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research,” in *Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*, Beijing, China, 2012. [Online]. Available: <https://www.semanticscholar.org/paper/BLUED-%3A-A-Fully-Labeled-Public-Dataset-for-Load-Anderson-Ocneanu/ed1b8fc3074ec5d7bb7cf83e233d3b130637706f>
- [5] K. D. Anderson, M. E. Bergés, A. Ocneanu, D. Benitez, and J. M. Moura, “Event detection for non intrusive load monitoring,” in *IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society*. IEEE, 2012, pp. 3312–3317, 00032. DOI: <http://dx.doi.org/10.1109/IECON.2012.6389367>
- [6] K. C. ARMEL, A. GUPTA, G. SHRIMALI, and A. ALBERT, “IS DISAGGREGATION THE HOLY GRAIL OF ENERGY EFFICIENCY? THE CASE OF ELECTRICITY,” 2012, 00110. DOI: <http://dx.doi.org/10.1016/j.enpol.2012.08.062>
- [7] P. R. Armstrong, C. R. Laughman, S. B. Leeb, and L. K. Norford, “Fault detection based on motor start transients and shaft harmonics measured at the RTU electrical service,” 2004, 00011. [Online]. Available: <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1663&context=iracc>
- [8] C. Audet and J. Dennis, “Analysis of Generalized Pattern Searches,” *SIAM Journal on Optimization*, vol. 13, no. 3, pp. 889–903, Jan. 2002. DOI: <http://dx.doi.org/10.1137/S1052623400378742>
- [9] F. A. M. Awad, M. E. H. M. Ahmed, M. T. M. Salih, O. M. T. Yousif, and others, “DESIGN OF ACTIVE FILTER FOR POWER SYSTEM DISTRIBUTION,” Ph.D. dissertation, Sudan University of Science and Technology, 2015, 00000. [Online]. Available: <http://repository.sustech.edu/handle/123456789/12700>
- [10] M. Azaza and F. Wallin, “Supervised household’s loads pattern recognition.” IEEE, Oct. 2016, pp. 1–5. DOI: <http://dx.doi.org/10.1109/EPEC.2016.7771718>
- [11] L. D. Baets, T. Dhaene, D. Deschrijver, C. Develder, and M. Berges, “VI-Based Appliance Classification Using Aggregated Power Consumption Data,” p. 8, 2018. DOI: <http://dx.doi.org/10.1109/SMARTCOMP.2018.00089>
- [12] M. Baranski and J. Voss, “Genetic algorithm for pattern detection in NIALM systems,” in *Systems, man and cybernetics, 2004 IEEE international conference on*, vol. 4. IEEE, 2004, pp. 3462–3468, 00124. DOI: <http://dx.doi.org/10.1109/ICSMC.2004.1400878>
- [13] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht, “Smart*: An open data set and tools for enabling research in sustainable homes,” *SustKDD, August*, vol. 111, p. 112, 2012, 00000. [Online]. Available: http://wan.poly.edu/KDD2012/forms/workshop/SustKDD12/doc/SustKDD12_3.pdf
- [14] K. S. Barsim and B. Yang, “Toward a semi-supervised non-intrusive load monitoring system for event-based energy disaggregation,” in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2015, pp. 58–62, 00000. DOI: <http://dx.doi.org/10.1109/GlobalSIP.2015.7418156>
- [15] K. Basu, A. Hably, V. Debusschere, S. Bacha, G. J. Driven, and A. Ovalle, “A comparative study of low sampling non intrusive load disaggregation,” in *Industrial Electronics Society, IECON 2016-42nd Annual Conference of the IEEE*. IEEE, 2016, pp. 5137–5142, 00000. DOI: <http://dx.doi.org/10.1109/IECON.2016.7793294>
- [16] N. Batra, M. Gulati, A. Singh, and M. B. Srivastava, “It’s Different: Insights into home energy consumption in India.” ACM Press, 2013, pp. 1–8, 00048. DOI: <http://dx.doi.org/10.1145/2528282.2528293>
- [17] N. Batra, O. Parson, M. Berges, A. Singh, and A. Rogers, “A comparison of non-intrusive load monitoring methods for commercial and residential buildings,” *arXiv preprint arXiv:1408.6595*, 2014, 00009. [Online]. Available: <http://arxiv.org/abs/1408.6595>
- [18] P. Baumann, A. Heinzlmann, P. Held, and D. Benyoucef, “Increasing the Efficiency of Photovoltaic (PV) Batteries through Non-Intrusive Load Monitoring,” in *EU PVSEC*, Amsterdam, Sep. 2017. [Online]. Available: http://www.eupvsec-planner.com/presentations/c43755/increasing_the_efficiency_of_photovoltaic_pv_batteries_through_non-intrusive_load_monitoring.htm
- [19] P. W. Baumann, P. Held, A. Heinzlmann, D. Benyoucef, and F. Baumgartner, “Effizienzsteigerung von Photovoltaik Batteriesteuerung durch geschicktes Load Monitoring,” in *Smart Home and Living Bodensee*, Konstanz, Germany, Nov. 2017. [Online]. Available: <https://digitalcollection.zhaw.ch/handle/11475/1977>
- [20] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, and S. Santini, “The ECO data set and the performance of non-intrusive load monitoring algorithms.” ACM

- Press, 2014, pp. 80–89, 00049. DOI: <http://dx.doi.org/10.1145/2674061.2674064>
- [21] T. Bier, “Disaggregation of Electrical Appliances using Non-Intrusive Load Monitoring,” Ph.D. dissertation, Universite de Haute Alsace, Mulhouse, France, Dec. 2014. [Online]. Available: <https://tel.archives-ouvertes.fr/tel-01314432/>
- [22] T. Bier, D. Benyoucef, D. Abdeslam Ould, J. Merckle, and P. Klein, “Smart Meter Systems Measurements for the Verification of the Detection & Classification Algorithms,” 2013, 00003. DOI: <http://dx.doi.org/10.1109/IECON.2013.6699945>
- [23] C. M. Bishop, *Pattern recognition and machine learning*, 1st ed., ser. Information science and statistics. New York: Springer, 2006. [Online]. Available: <https://www.springer.com/de/book/9780387310732>
- [24] Bo Liu, Yinxin Yu, and Wenpeng Luan, “An unsupervised electrical appliance modeling framework for non-intrusive load monitoring,” in *IEEE Power & Energy Society General Meeting*. IEEE, 2017, pp. 1–5. DOI: <http://dx.doi.org/10.1109/PESGM.2017.8273794>
- [25] C. Bogani, M. Gasparo, and A. Papini, “Generalized Pattern Search methods for a class of non-smooth optimization problems with structure,” *Journal of Computational and Applied Mathematics*, vol. 229, no. 1, pp. 283–293, Jul. 2009. DOI: <http://dx.doi.org/10.1016/j.cam.2008.10.047>
- [26] R. Bonfigli, S. Squartini, M. Fagiani, and F. Piazza, “Unsupervised algorithms for non-intrusive load monitoring: An up-to-date overview,” in *Environment and Electrical Engineering (EEEIC), 2015 IEEE 15th International Conference on*. Rome, Italy: IEEE, 2015, pp. 1175–1180, 00000. DOI: <http://dx.doi.org/10.1109/EEEIC.2015.7165334>
- [27] 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*. New York, NY, USA: ACM, 1992, pp. 144–152. DOI: <http://dx.doi.org/10.1145/130385.130401>
- [28] A. S. Bouhouras, M. Vaggos, N. Poulakis, and G. C. Christoforidis, “Load signatures enhancement via odd-order harmonic currents.” IEEE, Jun. 2016, pp. 1–6. DOI: <http://dx.doi.org/10.1109/EEEIC.2016.7555497>
- [29] CENELEC, “EN 50160 Voltage characteristics of electricity supplied by public distribution systems,” 2005. [Online]. Available: https://www.schneider-electric.ae/library/SCHNEIDER_ELECTRIC/SE_LOCAL/APS/204836_1312/DraftStandard0026rev2-DraftEN501602005-05.pdf
- [30] W. Chan and A. T. So, “Power harmonics pattern recognition by solving fuzzy equations,” *Fuzzy Sets and Systems*, vol. 67, no. 3, pp. 257–266, Nov. 1994. DOI: [http://dx.doi.org/10.1016/0165-0114\(94\)90254-2](http://dx.doi.org/10.1016/0165-0114(94)90254-2)
- [31] W. Chan, A. So, and L. Lai, “Harmonics load signature recognition by wavelets transforms,” in *DRPT2000. International Conference on Electric Utility Deregulation and Restructuring and Power Technologies. Proceedings (Cat. No.00EX382)*. London, UK: IEEE, 2000, pp. 666–671. DOI: <http://dx.doi.org/10.1109/DRPT.2000.855745>
- [32] H.-H. Chang, K.-L. Lian, Y.-C. Su, and W.-J. Lee, “Power-Spectrum-Based Wavelet Transform for Nonintrusive Demand Monitoring and Load Identification,” *IEEE Transactions on Industry Applications*, vol. 50, no. 3, pp. 2081–2089, May 2014, 00016. DOI: <http://dx.doi.org/10.1109/TIA.2013.2283318>
- [33] H.-H. Chang, C.-L. Lin, and J.-K. Lee, “Load identification in nonintrusive load monitoring using steady-state and turn-on transient energy algorithms,” in *Computer supported cooperative work in design (cscwd), 2010 14th international conference on*. IEEE, 2010, pp. 27–32, 00050. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5472008
- [34] H.-H. Chang, C.-L. Lin, and H.-T. Yang, “Load recognition for different loads with the same real power and reactive power in a non-intrusive load-monitoring system,” in *12th International Conference on Computer Supported Cooperative Work in Design, 2008. CSCWD 2008*, Apr. 2008, pp. 1122–1127. DOI: <http://dx.doi.org/10.1109/CSCWD.2008.4537137>
- [35] A. Cole and A. Albicki, “Nonintrusive identification of electrical loads in a three-phase environment based on harmonic content,” in *Instrumentation and Measurement Technology Conference, 2000. IMTC 2000. Proceedings of the 17th IEEE*, vol. 1. IEEE, 2000, pp. 24–29, 00068. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=846806
- [36] A. I. Cole and A. Albicki, “Data extraction for effective non-intrusive identification of residential power loads,” in *Instrumentation and Measurement Technology Conference, 1998. IMTC/98. Conference Proceedings. IEEE*, vol. 2. IEEE, 1998, pp. 812–815, 00076. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=676838
- [37] I. N. da Silva, D. Hernane Spatti, R. Andrade Flauzino, L. H. B. Liboni, and S. F. dos Reis Alves, *Artificial Neural Networks*. Cham: Springer International Publishing, 2017, 00088. DOI: <http://dx.doi.org/10.1007/978-3-319-43162-8>
- [38] L. De Baets, C. Develder, T. Dhaene, and D. Deschrijver, “Automated classification of appliances using elliptical fourier descriptors.” IEEE, Oct. 2017, pp. 153–158. DOI: <http://dx.doi.org/10.1109/SmartGridComm.2017.8340669>
- [39] B. der Justiz und für Verbraucherschutz, “Gesetz für den Ausbau erneuerbarer Energien (Erneuerbare-Energien-Gesetz - EEG 2017) § 48 Solare Strahlungsenergie,” Jan. 2017, 00000. [Online]. Available: http://www.gesetze-im-internet.de/eeg_2014/_48.html

- [40] H. Dinesh, P. H. Perera, G. Godaliyadda, M. P. B. Ekanayake, and J. B. Ekanayake, "Individual power profile estimation of residential appliances using low frequency smart meter data," in *2015 IEEE 10th International Conference on Industrial and Information Systems (ICIIS)*. IEEE, 2015, pp. 140–145, 00001. DOI: <http://dx.doi.org/10.1109/ICIINFS.2015.7399000>
- [41] —, "Residential appliance monitoring based on low frequency smart meter measurements," in *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, 2015, pp. 878–884, 00001. DOI: <http://dx.doi.org/10.1109/SmartGridComm.2015.7436412>
- [42] R. Dong, L. Ratliff, H. Ohlsson, and S. S. Sastry, "Fundamental limits of nonintrusive load monitoring," in *Proceedings of the 3rd international conference on High confidence networked systems*. ACM, 2014, pp. 11–18, 00014. DOI: <http://dx.doi.org/10.1145/2566468.2566471>
- [43] L. Du, D. He, R. G. Harley, and T. G. Habetler, "Electric Load Classification by Binary Voltage-Current Trajectory Mapping," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 358–365, Jan. 2016, 00000. DOI: <http://dx.doi.org/10.1109/TSG.2015.2442225>
- [44] 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, no. 3, pp. 1371–1380, Sep. 2013. DOI: <http://dx.doi.org/10.1109/TSG.2013.2263231>
- [45] R. O. Duda, P. E. Harting, and D. G. Stork, *Pattern Classification*, 2nd ed. New Delhi: Wiley, 2000. [Online]. Available: <https://www.wiley.com/en-us/Pattern+Classification,+2nd+Edition-p-9780471056690>
- [46] C. Efthymiou and G. Kalogridis, "Smart grid privacy via anonymization of smart metering data," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*. IEEE, 2010, pp. 238–243, 00408. DOI: <http://dx.doi.org/10.1109/SMARTGRID.2010.5622050>
- [47] ENTSOE, "P5 - Policy 5: Emergency Operations," 2015. [Online]. Available: https://docstore.entsoe.eu/Documents/Publications/SOC/Continental_Europe/oh/20150916_Policy_5_Approved_by_ENTSO-E_RG_CE_Plenary.pdf
- [48] D. P. G. e.V., "Welt der Physik: Struktur des deutschen Stromnetzes," 2018. [Online]. Available: <https://www.weltderphysik.de/gebiet/technik/energie/strom/netzstruktur/>
- [49] Z. Fan, P. Kulkarni, S. Gormus, C. Efthymiou, G. Kalogridis, M. Sooriyabandara, Z. Zhu, S. Lambotaran, and W. H. Chin, "Smart Grid Communications: Overview of Research Challenges, Solutions, and Standardization Activities," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 21–38, 2013, 00309. DOI: <http://dx.doi.org/10.1109/SURV.2011.122211.00021>
- [50] 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, no. 3, pp. 245–259, 1999, 00190. DOI: [http://dx.doi.org/10.1016/S0378-7788\(99\)00007-9](http://dx.doi.org/10.1016/S0378-7788(99)00007-9)
- [51] C. Y. Feng, H. M. Hoe, M. P. Abdullah, M. Y. Hassan, and F. Hussin, "Tracing of energy consumption by using harmonic current." IEEE, Dec. 2013, pp. 444–449. DOI: <http://dx.doi.org/10.1109/SCORED.2013.7002628>
- [52] Fluke, "Continuing frequency deviation in the Continental European Power System originating in Serbia/Kosovo: Political solution urgently needed in addition to technical." Mar. 2018. [Online]. Available: <https://www.entsoe.eu/news/2018/03/06/press-release-continuing-frequency-deviation-in-the-continental-european-power-system-originating-in-serbia-kosovo-political-solution-urgently-needed-in-addition-to-technical/>
- [53] J. Froehlich, E. Larson, S. Gupta, G. Cohn, M. Reynolds, and S. Patel, "Disaggregated end-use energy sensing for the smart grid," *IEEE Pervasive Computing*, vol. 10, no. 1, pp. 28–39, 2011, 00201. DOI: <http://dx.doi.org/10.1109/MPRV.2010.74>
- [54] J. Gao, S. Giri, E. C. Kara, and M. Bergés, "PLAID: a public dataset of high-resolution electrical appliance measurements for load identification research: demo abstract." ACM Press, 2014, pp. 198–199, 00002. DOI: <http://dx.doi.org/10.1145/2674061.2675032>
- [55] A. Géron and K. Rother, *Praxiseinstieg Machine Learning mit Scikit-Learn und TensorFlow : Konzepte, Tools und Techniken für intelligente Systeme*. Heidelberg, GERMANY: O'Reilly, 2018. [Online]. Available: <http://ebookcentral.proquest.com/lib/hsfurtwangen/detail.action?docID=5215589>
- [56] J. M. Gillis and W. G. Morsi, "Non-Intrusive Load Monitoring Using Semi-Supervised Machine Learning and Wavelet Design," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2648–2655, Nov. 2017. DOI: <http://dx.doi.org/10.1109/TSG.2016.2532885>
- [57] H. T. GmbH, A. GmbH, T. GmbH, and T. T. GmbH, "Netztransparenz > EEG > Jahresabrechnungen," Jan. 2017. [Online]. Available: <https://www.netztransparenz.de/EEG/Jahresabrechnungen>
- [58] S. Gupta, M. S. Reynolds, and S. N. Patel, "ElectriSense: single-point sensing using EMI for electrical event detection and classification in the home," in *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, 2010, pp. 139–148, 00297. DOI: <http://dx.doi.org/10.1145/1864349.1864375>
- [59] T. Guzel and E. Ustunel, "Principal components null space analysis based non-intrusive load monitoring," in *Electrical Power and Energy Conference (EPEC), 2015 IEEE*. IEEE, 2015, pp. 420–423, 00000. DOI: <http://dx.doi.org/10.1109/EPEC.2015.7379987>

- [60] G. W. Hart, "Prototype Nonintrusive Appliance Load Monitoring," 1985. [Online]. Available: <http://www.georgehart.com/research/Hart1985.pdf>
- [61] —, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992, 00954. DOI: <http://dx.doi.org/10.1109/5.192069>
- [62] G. Hart, "Residential energy monitoring and computerized surveillance via utility power flows," *IEEE Technology and Society Magazine*, vol. 8, no. 2, pp. 12–16, Jun. 1989. DOI: <http://dx.doi.org/10.1109/44.31557>
- [63] —, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992. DOI: <http://dx.doi.org/10.1109/5.192069>
- [64] G. Hart, E. Kern, and F. Schweppe, "Non-intrusive appliance monitor apparatus," U.S. Patent US 4858 141; A, Aug., 1989. [Online]. Available: <https://www.osti.gov/biblio/7269283>
- [65] T. Hassan, F. Javed, and N. Arshad, "An empirical investigation of VI trajectory based load signatures for non-intrusive load monitoring," *Smart Grid, IEEE Transactions on*, vol. 5, no. 2, pp. 870–878, 2014, 00011. DOI: <http://dx.doi.org/10.1109/TSG.2013.2271282>
- [66] P. Held, F. Laasch, D. O. Abdeslam, and D. Benyoucef, "Frequency invariant transformation of periodic signals (FIT-PS) for signal representation in NILM," in *Industrial Electronics Society, IECON 2016-42nd Annual Conference of the IEEE*. Florence, Italy: IEEE, Oct. 2016, pp. 5149–5154, 00001. DOI: <http://dx.doi.org/10.1109/IECON.2016.7793617>
- [67] P. Held, S. Mauch, A. Saleh, D. Abdeslam Ould, and D. Benyoucef, "HELD1: Home Equipment Laboratory Dataset for Non-Intrusive Load Monitoring," in *The Third International Conference on Advances in Signal, Image and Video Processing SIGNAL 2018*, Nice, France, 2018. [Online]. Available: https://www.thinkmind.org/index.php?view=article&articleid=signal_2018_2_10_60009
- [68] P. Held, S. Mauch, A. Saleh, D. Ould Abdeslam, and D. Benyoucef, "Frequency Invariant Transformation of Periodic Signals (FIT-PS) for Classification in NILM," *IEEE Transactions on Smart Grid*, p. 8, Dec. 2018, accepted for publication. DOI: <http://dx.doi.org/10.1109/TSG.2018.2886849>
- [69] P. Held, A. Saleh, D. Abdeslam Ould, and D. Benyoucef, "Frequency Invariant Transformation of Periodic Signals (FIT-PS) for high frequency Signal Representation in NILM," in *3rd Baden-Württemberg Center of Applied Research Symposium on Information and Communication Systems*, Karlsruhe, Dec. 2016, pp. 1–6.
- [70] P. Held, D. Weißhaar, S. Mauch, D. O. Abdeslam, and D. Benyoucef, "Parameter Optimized Event Detection for NILM using Frequency Invariant Transformation of Periodic Signals (FIT-PS)," in *IEEE ETFA 2018 International Conference on Emerging Technologies and Factory Automation*. Turin, Italy: IEEE, Sep. 2018, p. 6. DOI: <http://dx.doi.org/10.1109/ETFA.2018.8502522>
- [71] R. Hoffmann, *Signalanalyse und -erkennung: eine Einführung für Informationstechniker*. Springer, 1998. [Online]. Available: <https://link.springer.com/book/10.1007/978-3-662-45323-0>
- [72] C. Holcomb, "A Test-bed for NILM," Pittsburgh, Pa, USA, 2012, 00000. [Online]. Available: http://nilmworkshop.org/2012/posters/holcomb_PecanStreet_NILM2012_poster.pdf
- [73] R. Hooke and T. A. Jeeves, "'Direct Search' Solution of Numerical and Statistical Problems," *Journal of the ACM*, vol. 8, no. 2, pp. 212–229, Apr. 1961. DOI: <http://dx.doi.org/10.1145/321062.321069>
- [74] T. D. Huang, W.-S. Wang, and K.-L. Lian, "A New Power Signature for Nonintrusive Appliance Load Monitoring," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1994–1995, Jul. 2015, 00000. DOI: <http://dx.doi.org/10.1109/TSG.2015.2415456>
- [75] N. Iksan, J. Sembiring, N. Haryanto, and S. H. Supangkat, "Appliances Identification Method of Non-Intrusive Load Monitoring based on Load Signature of V-I Trajectory," Bandung - Bali, Indonesia, 2015, 00000 OCLC: 951109734. DOI: <http://dx.doi.org/10.1109/ICITSI.2015.7437744>
- [76] A. Iwayemi and C. Zhou, "SARAA: Semi-Supervised Learning for Automated Residential Appliance Annotation," *IEEE Transactions on Smart Grid*, pp. 1–8, 2015. DOI: <http://dx.doi.org/10.1109/TSG.2015.2498642>
- [77] L. Jiang, S. Luo, and J. Li, "Automatic power load event detection and appliance classification based on power harmonic features in nonintrusive appliance load monitoring," in *Industrial Electronics and Applications (ICIEA), 2013 8th IEEE Conference on*. IEEE, 2013, pp. 1083–1088, 00004. DOI: <http://dx.doi.org/10.1109/ICIEA.2013.6566528>
- [78] R. Jonetzko, M. Detzler, K.-U. Gollmer, A. Guldner, M. Huber, R. Michels, and S. Naumann, "High frequency non-intrusive electric device detection and diagnosis," in *Smart Cities and Green ICT Systems (SMARTGREENS), 2015 International Conference on*. IEEE, 2015, pp. 1–8, 00001. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7297979
- [79] D. Jung, H. H. Nguyen, and D. K. Yau, "Tracking Appliance Usage Information Using Harmonic Signature Sensing," *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pp. 459–465, 2015, 00000. DOI: <http://dx.doi.org/10.1109/SmartGridComm.2015.7436343>
- [80] M. Kahl, A. U. Haq, T. Kriechbaumer, and H.-A. Jacobsen, "WHITED-A Worldwide Household and Industry Transient Energy Data Set," 2016, 00000. [Online]. Available: http://nilmworkshop.org/2016/proceedings/Poster_ID18.pdf

- [81] G. Kalogridis, C. Efthymiou, S. Z. Denic, T. A. Lewis, and R. Cepeda, "Privacy for smart meters: Towards undetectable appliance load signatures," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*. IEEE, 2010, pp. 232–237, 00260. DOI: <http://dx.doi.org/10.1109/SMARTGRID.2010.5622047>
- [82] J. Kelly and W. Knottenbelt, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," *Scientific Data*, vol. 2, p. 150007, Mar. 2015, 00034. DOI: <http://dx.doi.org/10.1038/sdata.2015.7>
- [83] A. A. Kholeif, H. A. A. El-Ghany, and A. M. Azmy, "Impact of supply voltage variation on V-I trajectory identification method." IEEE, Dec. 2017, pp. 839–844. DOI: <http://dx.doi.org/10.1109/MEPCON.2017.8301277>
- [84] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements," in *Proceedings of the 2011 SIAM International Conference on Data Mining*. SIAM, 2011, pp. 747–758, 00256. DOI: <http://dx.doi.org/10.1137/1.9781611972818.64>
- [85] Y. Kim, T. Schmid, Z. M. Charbiwala, and M. B. Srivastava, "ViridiScope: design and implementation of a fine grained power monitoring system for homes," in *Proceedings of the 11th international conference on Ubiquitous computing*. ACM, 2009, pp. 245–254, 00244. DOI: <http://dx.doi.org/10.1145/1620545.1620582>
- [86] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983. DOI: <http://dx.doi.org/10.1126/science.220.4598.671>
- [87] P. Klein, "Non-Intrusive Information Sources for Activity Analysis in Ambient Assisted Living Scenarios," Ph.D. dissertation, Universite de Haute Alsace, Mulhouse, France, Nov. 2015. [Online]. Available: <https://tel.archives-ouvertes.fr/tel-01526695/>
- [88] J. Z. Kolter and T. S. Jaakkola, "Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation." in *AISTATS*, vol. 22, 2012, pp. 1472–1482, 00217. [Online]. Available: <http://www.jmlr.org/proceedings/papers/v22/zico12/zico12.pdf>
- [89] J. Z. Kolter and M. J. Johnson, "REDD: A public data set for energy disaggregation research," in *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*, vol. 25. Citeseer, 2011, pp. 59–62, 00406. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.454.5796&rep=rep1&type=pdf>
- [90] L. Kong, D. Yang, and Y. Luo, "Non-Intrusive Load Monitoring and Identification Based on Maximum Likelihood Method." IEEE, Apr. 2017, pp. 268–272. DOI: <http://dx.doi.org/10.1109/ICEI.2017.54>
- [91] G. C. Koutitas and L. Tassioulas, "Low Cost Disaggregation of Smart Meter Sensor Data," *IEEE Sensors Journal*, vol. 16, no. 6, pp. 1665–1673, Mar. 2016, 00001. DOI: <http://dx.doi.org/10.1109/JSEN.2015.2501422>
- [92] K. Kroschel, G. Rigoll, and B. Schuller, *Statistische Informationstechnik: Signal- und Mustererkennung, Parameter- und Signalschätzung*, 5th ed. Berlin: Springer, 2011. [Online]. Available: <https://link.springer.com/book/10.1007/978-3-642-15954-1>
- [93] F. Kuhnert, *Pseudoinverse Matrizen und die Methode der Regularisierung*, 1st ed. Leipzig, Germany: Teubner, 1976. [Online]. Available: <https://doi.org/10.1002/zamm.19780581023>
- [94] H. Y. Lam, G. S. K. Fung, and W. K. Lee, "A novel method to construct taxonomy electrical appliances based on load signaturesof," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, 2007, 00186. DOI: <http://dx.doi.org/10.1109/TCE.2007.381742>
- [95] Z. Lan, B. Yin, T. Wang, and G. Zuo, "A Non-Intrusive Load Identification Method Based on Convolutional Neuronal Network," p. 5, Nov. 2017. DOI: <http://dx.doi.org/10.1109/EI2.2017.8245612>
- [96] C. Laughman, Kwangduk Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong, "Power signature analysis," *IEEE Power and Energy Magazine*, vol. 1, no. 2, pp. 56–63, Mar. 2003. DOI: <http://dx.doi.org/10.1109/MPAE.2003.1192027>
- [97] K. Lee, S. Leeb, L. Norford, P. Armstrong, J. Holloway, and S. Shaw, "Estimation of Variable-Speed-Drive Power Consumption From Harmonic Content," *IEEE Transactions on Energy Conversion*, vol. 20, no. 3, pp. 566–574, Sep. 2005, 00068. DOI: <http://dx.doi.org/10.1109/TEC.2005.852963>
- [98] S. B. Leeb, S. R. Shaw, and J. L. Kirtley, "Transient event detection in spectral envelope estimates for nonintrusive load monitoring," *IEEE Transactions on Power Delivery*, vol. 10, no. 3, pp. 1200–1210, 1995, 00254. [Online]. Available: <http://ieeexplore.ieee.org/abstract/document/400897/>
- [99] D. Li, K. Sawyer, and S. Dick, "Disaggregating household loads via semi-supervised multi-label classification." IEEE, Aug. 2015, pp. 1–5. DOI: <http://dx.doi.org/10.1109/NAFIPSWConSC.2015.7284144>
- [100] J. Li, S. West, and G. Platt, "Power decomposition based on SVM regression," in *Modelling, Identification & Control (ICMIC), 2012 Proceedings of International Conference on*. IEEE, 2012, pp. 1195–1199, 00013. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6260133
- [101] J. Liang, S. K. K. Ng, G. Kendall, and J. W. M. Cheng, "Load Signature Study-Part I: Basic Concept, Structure, and Methodology," *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 551–560, Apr. 2010, 00195. DOI: <http://dx.doi.org/10.1109/TPWRD.2009.2033799>
- [102] —, "Load Signature Study-Part II: Disaggregation Framework, Simulation, and Applications,"

- IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 561–569, Apr. 2010, 00083. DOI: <http://dx.doi.org/10.1109/TPWRD.2009.2033800>
- [103] F. Liebgott and B. Yang, “Active learning with cross-dataset validation in event-based non-intrusive load monitoring.” *IEEE*, Aug. 2017, pp. 296–300. DOI: <http://dx.doi.org/10.23919/EUSIPCO.2017.8081216>
- [104] D. Liebl, “Messen mit modernen Spektrumanalysatoren Educational Note,” Feb. 2013. [Online]. Available: https://cdn.rohd-schwarz.com/pws/dl_downloads/dl_application/application_notes/1ma201_1/1MA201_9d_spektrumanalysator_mess.pdf
- [105] D. Luo, L. K. Norford, S. R. Shaw, and S. B. Leeb, “Monitoring HVAC Equipment Electrical Loads from a Centralized Location Methods and Field Test Results,” 2003, 00000. [Online]. Available: <http://www.energy.ca.gov/2003publications/CEC-500-2003-097/CEC-500-2003-097F-A04.PDF>
- [106] M. Maasoumy, B. Sanandaji, K. Poolla, and A. S. Vincentelli, “Berds-berkeley energy disaggregation data set,” in *Proceedings of the Workshop on Big Learning at the Conference on Neural Information Processing Systems (NIPS)*, 2013, 00007. [Online]. Available: <https://www.semanticscholar.org/paper/BERDS-BERkeley-EneRgy-Disaggregation-Data-Set-Maasoumy/377e1010e90248a2280b2f4754b8f4a04fe414f9>
- [107] S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. V. Bajic, “AMPds: A public dataset for load disaggregation and eco-feedback research,” in *Electrical Power & Energy Conference (EPEC), 2013 IEEE*. IEEE, 2013, pp. 1–6, 00065. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6802949
- [108] M. L. Marceau and R. Zmeureanu, “Nonintrusive load disaggregation computer program to estimate the energy consumption of major end uses in residential buildings,” *Energy Conversion and Management*, vol. 41, no. 13, pp. 1389–1403, 2000, 00146. DOI: [http://dx.doi.org/10.1016/S0196-8904\(99\)00173-9](http://dx.doi.org/10.1016/S0196-8904(99)00173-9)
- [109] L. Mauch and B. Yang, “A new approach for supervised power disaggregation by using a deep recurrent LSTM network,” in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2015, pp. 63–67, 00000. DOI: <http://dx.doi.org/10.1109/GlobalSIP.2015.7418157>
- [110] Q. Meng and S. H. Ng, “Enhancing pattern search for global optimization with an additive global and local Gaussian Process model.” *IEEE*, Dec. 2017, pp. 2092–2103. DOI: <http://dx.doi.org/10.1109/WSC.2017.8247942>
- [111] M. A. Mengistu, A. A. Girmay, C. Camarda, A. Acquaviva, and E. Patti, “A Cloud-based On-line Disaggregation Algorithm for Home Appliance Loads,” *IEEE Transactions on Smart Grid*, pp. 1–1, 2018. DOI: <http://dx.doi.org/10.1109/TSG.2018.2826844>
- [112] M. Meyer, *Signalverarbeitung: analoge und digitale Signale, Systeme und Filter*, 8th ed., ser. Lehrbuch. Wiesbaden: Springer Vieweg, 2017, oCLC: 1001340202. [Online]. Available: <https://www.springer.com/us/book/9783658183202>
- [113] M. N. Meziane, P. Ravier, G. Lamarque, K. Abed-Meraim, L. Bunetel, Y. Raingeaud, and others, “Modeling and estimation of transient current signals,” in *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE, 2015, pp. 1960–1964, 00000. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7362726
- [114] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, “Private memoirs of a smart meter,” in *Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building*. ACM, 2010, pp. 61–66, 00252. DOI: <http://dx.doi.org/10.1145/1878431.1878446>
- [115] A. Monacchi, D. Egarter, W. Elmenreich, S. D’Alessandro, and A. M. Tonello, “GREEND: An Energy Consumption Dataset of Households in Italy and Austria.” Venice, Italy: IEEE, Nov. 2014, 00000. [Online]. Available: <http://ieeexplore.ieee.org/ielx7/6995184/7007609/07007698.pdf?tp=&arnumber=7007698&isnumber=7007609>
- [116] A. Munshi and Y. A.-R. Mohamed, “Unsupervised Non-Intrusive Extraction of Electrical Vehicle Charging Load Patterns,” *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2018. DOI: <http://dx.doi.org/10.1109/TII.2018.2806936>
- [117] 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*, pp. 51–64, 1996, 00000. DOI: [http://dx.doi.org/10.1016/0378-7788\(95\)00958-2](http://dx.doi.org/10.1016/0378-7788(95)00958-2)
- [118] U. A. Orji, Z. Remscrim, C. Laughman, S. B. Leeb, W. Wichakool, C. Schantz, R. Cox, J. Paris, J. L. Kirtley, and L. K. Norford, “Fault detection and diagnostics for non-intrusive monitoring using motor harmonics.” *IEEE*, Feb. 2010, pp. 1547–1554. DOI: <http://dx.doi.org/10.1109/APEC.2010.5433437>
- [119] A. Ortiz Pérez, V. Kallfaß-de Frenes, A. Filbert, J. Kneer, B. Bierer, P. Held, P. Klein, J. Wöllenstein, D. Benyoucef, S. Kallfaß, U. Mescheder, and S. Palzer, “Odor-Sensing System to Support Social Participation of People Suffering from Incontinence,” *Sensors*, vol. 17, no. 1, p. 58, Dec. 2016, 00000. DOI: <http://dx.doi.org/10.3390/s17010058>
- [120] D. Ould-Abdeslam, A. Amirou, and Z. Zidelmal, “Stockwell-transform for Electrical Defaults Localization.” Marrakech, Morocco: 2015 3rd International Renewable and Sustainable Energy Conference (IRSEC), Dec. 2015, p. 6. DOI: <http://dx.doi.org/10.1109/IRSEC.2015.7454937>
- [121] J. V. Paatero and P. D. Lund, “A model for generating household electricity load profiles,” *International Journal of Energy Research*, vol. 30,

- no. 5, pp. 273–290, Apr. 2006, 00356. DOI: <http://dx.doi.org/10.1002/er.1136>
- [122] 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,” *Lecture Notes in Computer Science*, vol. 4717, pp. 271–288, 2007, 00315. DOI: http://dx.doi.org/10.1007/978-3-540-74853-3_16
- [123] L. Pereira, F. Quintal, R. Gonçalves, and N. J. Nunes, “SustData: A Public Dataset for ICT4s Electric Energy Research.” in *ICT4S*, 2014, 00012. [Online]. Available: http://www.atlantispress.com/php/download_paper.php?id=13462
- [124] T. Picon, M. N. Meziane, P. Ravier, G. Lamarque, C. Novello, J.-C. L. Bunetel, and Y. Raingeaud, “COOLL: Controlled On/Off Loads Library, a Public Dataset of High-Sampled Electrical Signals for Appliance Identification,” *arXiv preprint arXiv:1611.05803*, 2016, 00001. [Online]. Available: <https://arxiv.org/abs/1611.05803>
- [125] H. Pihala, *Non-intrusive appliance load monitoring system based on a modern kWh-meter*, ser. VTT publications. Espoo: Technical Research Centre of Finland, 1998, no. 356, 00059 OCLC: 246146186. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.428.4015&rep=rep1&type=pdf>
- [126] R. S. Prasad and S. Semwal, “A simplified new procedure for identification of appliances in smart meter applications,” in *Systems Conference (SysCon), 2013 IEEE International*. IEEE, 2013, pp. 339–344, 00011. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6549903
- [127] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing - Principles, Algorithms, and Applications*, 3rd ed. United States of America: Prentice - Hall, 1995. [Online]. Available: <http://ultra.sdk.free.fr/docs/DxO/Digital%20Signal%20Processing%20-%20Principles,%20Algorithms%20&%20Applications%20Proakis%20&%20Dimitris%20G.%20Manolakis.pdf>
- [128] S. Rahimi, A. D. C. Chan, and R. A. Goubran, “Nonintrusive load monitoring of electrical devices in health smart homes.” *IEEE*, May 2012, pp. 2313–2316. DOI: <http://dx.doi.org/10.1109/I2MTC.2012.6229453>
- [129] A. Reinhardt, P. Baumann, D. Burgstahler, M. Hollick, H. Chonov, M. Werner, and R. Steinmetz, “On the Accuracy of Appliance Identification Based on Distributed Load Metering Data,” p. 9, 2012. [Online]. Available: <https://ieeexplore.ieee.org/document/6388037>
- [130] A. Saleh, P. Held, D. Benyoucef, and D. O. Abdeslam, “A Novel Procedure for Virtual Measurements Generation suitable for Training and Testing in the context of Non Intrusive Load Monitoring,” in *The Third International Conference on Advances in Signal, Image and Video Processing SIGNAL 2018*, Nice, France, 2018, p. 7. [Online]. Available: https://www.thinkmind.org/index.php?view=article&articleid=signal_2018_2_40_60080
- [131] A. Saleh, P. Held, D. Benyoucef, and D. Ould Abdeslam, “EMD inspired Filtering algorithm for signal analysis in the context of Non Intrusive Load Monitoring,” in *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*. Beijing, China: IEEE, Oct. 2017, pp. 3615–3620. DOI: <http://dx.doi.org/10.1109/IECON.2017.8216613>
- [132] J. I. Sanchez-Gomez, L. Morales-Velazquez, R. A. Osornio-Rios, and E. Guillen-Garcia, “Fuzzy c-means clustering for steady state events classification of electrical signals.” *IEEE*, 2016, pp. 1–5. DOI: <http://dx.doi.org/10.1109/CONIIN.2016.7498129>
- [133] S. Semwal, G. Shah, and R. Prasad, “Identification Residential Appliance Using NIALM,” 2014, 00000. DOI: <http://dx.doi.org/10.1109/PEDES.2014.7041965>
- [134] S. Shaw, S. Leeb, L. Norford, and R. Cox, “Nonintrusive Load Monitoring and Diagnostics in Power Systems,” *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 7, pp. 1445–1454, Jul. 2008, 00167. DOI: <http://dx.doi.org/10.1109/TIM.2008.917179>
- [135] R. Singh and A. Singh, “Energy loss due to harmonics in residential campus A case study,” in *Universities Power Engineering Conference (UPEC), 2010 45th International*. IEEE, 2010, pp. 1–6, 00017. [Online]. Available: <http://ieeexplore.ieee.org/abstract/document/5649824/>
- [136] S. W. Smith, *The scientist and engineer’s guide to digital signal processing*. San Diego (Calif.): California Technical Pub., 1999, oCLC: 493473234.
- [137] R. Streubel and B. Yang, “Identification of electrical appliances via analysis of power consumption,” in *Universities Power Engineering Conference (UPEC), 2012 47th International*. IEEE, 2012, pp. 1–6. DOI: <http://dx.doi.org/10.1109/UPEC.2012.6398559>
- [138] S. Su, Y. Yan, H. Lu, L. Kangping, S. Yujing, W. Fei, L. Liming, and R. Hui, “Non-intrusive load monitoring of air conditioning using low-resolution smart meter data,” in *Power System Technology (POWERCON), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–5, 00000. DOI: <http://dx.doi.org/10.1109/POWERCON.2016.7753952>
- [139] Y.-C. Su, K.-L. Lian, and H.-H. Chang, “Feature Selection of Non-intrusive Load Monitoring System Using STFT and Wavelet Transform.” *IEEE*, Oct. 2011, pp. 293–298, 00009. DOI: <http://dx.doi.org/10.1109/ICEBE.2011.49>
- [140] D. Teshome, T. D. Huang, and K.-L. Lian, “A Distinctive Load Feature Extraction Based on Fryze’s Time-domain Power Theory,” *IEEE Power and Energy Technology Systems Journal*, pp. 1–1, 2016, 00000. DOI: <http://dx.doi.org/10.1109/JPETS.2016.2559507>

- [141] N. K. Thokala, M. G. Chandra, and K. Nagasubramanian, "On load disaggregation using discrete events." IEEE, Nov. 2016, pp. 324–329. DOI: <http://dx.doi.org/10.1109/ISGT-Asia.2016.7796406>
- [142] V. Torczon, "On the Convergence of Pattern Search Algorithms," *SIAM Journal on Optimization*, vol. 7, no. 1, pp. 1–25, Feb. 1997. DOI: <http://dx.doi.org/10.1137/S1052623493250780>
- [143] D. Weißhaar, "Untersuchung und Parameteroptimierung eines neuen Eventdetektionsverfahrens für Non-Intrusive Load Monitoring," p. 134, 2018.
- [144] D. Weißhaar, P. Held, S. Mauch, and D. Benyoucef, "Device Classification for NILM using FIT-PS compared with Standard Signal Forms," in *International IEEE Conference AND workshop in Óbuda on Electrical and Power Engineering*. Budapest, Hungary: IEEE, 2018, p. 6. DOI: <http://dx.doi.org/10.1109/CANDO-EPE.2018.8601150>
- [145] E. Weisstein, "Trigonometric Addition Formulas – from Wolfram MathWorld," Jan. 2019. [Online]. Available: <http://mathworld.wolfram.com/TrigonometricAdditionFormulas.html>
- [146] L. Wen, L. Gao, X. Li, and L. Zhang, "Free Pattern Search for global optimization," *Applied Soft Computing*, vol. 13, no. 9, pp. 3853–3863, Sep. 2013. DOI: <http://dx.doi.org/10.1016/j.asoc.2013.05.004>
- [147] W. Wichakool, A.-T. Avestruz, R. Cox, and S. Leeb, "Resolving Power Consumption of Variable Power Electronic Loads Using Nonintrusive Monitoring," in *IEEE Power Electronics Specialists Conference, 2007. PESC 2007*, Jun. 2007, pp. 2765–2771. DOI: <http://dx.doi.org/10.1109/PESC.2007.4342456>
- [148] B. Wild, K. S. Barsim, and B. Yang, "A new unsupervised event detector for non-intrusive load monitoring," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2015, pp. 73–77, 00000. DOI: <http://dx.doi.org/10.1109/GlobalSIP.2015.7418159>
- [149] J. Wolf, "Application Note 1epan 16d," p. 7, 1995. [Online]. Available: https://cdn.rohde-schwarz.com/pws/dl_downloads/dl_application/application_notes/1epan16/1epan16d.pdf
- [150] Yao Nie and Kai-Kuang Ma, "Adaptive rood pattern search for fast block-matching motion estimation," *IEEE Transactions on Image Processing*, vol. 11, no. 12, pp. 1442–1449, Dec. 2002. DOI: <http://dx.doi.org/10.1109/TIP.2002.806251>
- [151] M. Zeifman and K. Roth, "Nonintrusive Appliance Load Monitoring: Review and Outlook," 2011. DOI: <http://dx.doi.org/10.1109/TCE.2011.5735484>
- [152] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognition*, vol. 40, no. 7, pp. 2038–2048, Jul. 2007. DOI: <http://dx.doi.org/10.1016/j.patcog.2006.12.019>
- [153] Z. Zhang, J. H. Son, Y. Li, M. Trayer, Z. Pi, D. Y. Hwang, and J. K. Moon, "Training-free non-intrusive load monitoring of electric vehicle charging with low sampling rate," in *Industrial Electronics Society, IECON 2014-40th Annual Conference of the IEEE*. IEEE, 2014, pp. 5419–5425, 00004. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7049328
- [154] J.-P. Zimmermann, M. Evans, J. Griggs, N. King, L. Harding, P. Roberts, and C. Evans, "Household Electricity Survey: A study of domestic electrical product usage," Tech. Rep. R66141, May 2012. [Online]. Available: <http://doi.org/10.5255/UKDA-SN-7874-1>
- [155] A. Zoha, A. Gluhak, M. Imran, and S. Rajasegarar, "Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey," *Sensors*, vol. 12, no. 12, pp. 16 838–16 866, Dec. 2012, 00243. DOI: <http://dx.doi.org/10.3390/s121216838>