



**HAL**  
open science

# Non-Intrusive Information Sources for Activity Analysis in Ambient Assisted Living Scenarios

Philipp Klein

► **To cite this version:**

Philipp Klein. Non-Intrusive Information Sources for Activity Analysis in Ambient Assisted Living Scenarios. Other. Université de Haute Alsace - Mulhouse, 2015. English. NNT : 2015MULH8932 . tel-01526695

**HAL Id: tel-01526695**

**<https://theses.hal.science/tel-01526695>**

Submitted on 23 May 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**UNIVERSITÉ DE HAUTE-ALSACE**

## **Thèse**

*Présentée pour obtenir le grade de*

DOCTEUR DE L'UNIVERSITÉ DE HAUTE-ALSACE

**Discipline: Électronique, Électrotechnique et Automatique**

*par*

**Philipp KLEIN**

---

# **Non-Intrusive Information Sources for Activity Analysis in Ambient Assisted Living Scenarios**

---

(Arrêté Ministériel du 30 mars 1992)

Soutenue publiquement le 19. novembre 2015 devant le jury composé de:

<i>Rapporteurs</i>	Thierry CECCHIN	Professeur, Université de Lorraine
	Sébastien CAUET	MCF HDR, Université de Poitiers
<i>Examineurs</i>	Patrice WIRA	Professeur, Université de Haute-Alsace
	Djaffar OULD	MCF HDR, Université de Haute-Alsace

Thèse préparée au sein du laboratoire MIPS  
à l'Université de Haute-Alsace sous la direction de  
Professeur Jean MERCKLE

et au sein de Signal Processing Research Group (ReSP)  
à Furtwangen University sous la co-direction de  
Professeur Dirk BENYOUCEF.



# Acknowledgment

The work presented in this thesis was funded by *ZAFH-AAL*, a project of the Zentrum für Angewandte Forschung an Hochschulen (*ZAFH*)<sup>1)</sup> research funding program of the Ministry of Science, Research and the Arts of the State of Baden-Württemberg, Germany. Its goal is to research and develop assistive technologies to help people live autonomously and maintain their social relations and participation in everyday life.

The thesis on hand was made possible by a collaboration of the Modélisation, Intelligence, Processus et Systemes (MIPS) Laboratory at Université de Haute-Alsace, Mulhouse, France, and the Signal Processing Research Group (ReSP) at Furtwangen University, Furtwangen, Germany. The international, French-German exchange was an interesting and informative experience for me.

First and foremost I wish to thank my advisor Professor Jean Mercklé, MIPS laboratory, and my co-advisor Professor Dirk Benyoucef, ReSP Group, for their constant support. Without their commitment and international partnership, this thesis would not have been possible. They were always there to help me when technical or administrative difficulties had to be mastered.

I would also like to thank Thierry Cecchin, Professeur des Universités, and Sébastien Cauet, Maître de Conférences HDR, for their support by reviewing my thesis and being members of my jury. Thanks also go to Patrice Wira, Professeur des Universités, for being president of my jury, and to Djaffar Ould, Maître de Conférences HDR, for completing the circle of members of the jury.

Not mentioned yet are the people who supported me with their expertise and patience during long discussions. Thanks go to Thomas Bier, Frederik Laasch, Matthias Wöhrle, and Pirmin Held. Especially Thomas Bier's work should be mentioned. By pioneering his Ph.D. thesis in the international partnership of our two universities, he provided me with lots of experience in administrative and technical concerns, as well as an important set of data to base my investigations on.

Finally, I am very grateful to my parents, family, and friends for all their love and encouragement during all highs and lows of this Ph.D. thesis.

---

<sup>1)</sup>The word-by-word translation is “center for applied research at universities of applied sciences”.



# Contents

<b>1</b>	<b>Topic &amp; Motivation</b>	<b>7</b>
1.1	Interpretation of Information . . . . .	8
1.2	Application Scenario . . . . .	9
1.3	Organization of this Thesis . . . . .	9
1.4	Contributions . . . . .	9
<b>2</b>	<b>Relevant Work &amp; Overview of State-of-the-Art Methods</b>	<b>13</b>
2.1	Activity Monitoring . . . . .	13
2.2	Device-Free Passive Presence Detection . . . . .	14
2.3	Non-Intrusive Load Monitoring . . . . .	15
<b>3</b>	<b>Device-Free Passive Presence Detection</b>	<b>21</b>
3.1	Introduction . . . . .	21
3.2	Proposed Detection System . . . . .	26
3.3	Conclusion . . . . .	36
<b>4</b>	<b>Non-Intrusive Load Monitoring for AAL Applications</b>	<b>37</b>
4.1	Introduction . . . . .	37
4.2	Appliance Model . . . . .	38
4.3	Choice of Features . . . . .	39
4.4	Machine Learning . . . . .	39
4.5	Signals & Signal Choice . . . . .	44
4.6	Evaluation Methods . . . . .	54
4.7	Public Data Sets . . . . .	58
4.8	ReSP Data Sets . . . . .	59
<b>5</b>	<b>Event Detection</b>	<b>63</b>
5.1	Introduction . . . . .	63
5.2	Event Detection using Weighted Difference Edge Detection . . . . .	64
5.3	Event Detection using Clustering . . . . .	69
5.4	Event Detection using Steady State Approximation . . . . .	71
5.5	Results . . . . .	72
5.6	Conclusion . . . . .	76
<b>6</b>	<b>Feature Extraction &amp; Appliance Classification</b>	<b>79</b>
6.1	Introduction . . . . .	79
6.2	Steady State Power Features . . . . .	79
6.3	Feature Extraction with Steady State Approximation . . . . .	84
6.4	Results . . . . .	85
6.5	Time-Series Fingerprint Classification . . . . .	92
6.6	TSFP Data Volume Reduction . . . . .	101
6.7	Appliance Turn-On Fingerprint Catalog . . . . .	104
6.8	Conclusion . . . . .	108

<b>7 Case Studies</b>	<b>111</b>
7.1 Pattern Discovery on Appliance Switching Profiles . . . . .	111
7.2 Detecting Overheating Heaters in Load Monitoring Data . . . . .	123
7.3 Conclusion . . . . .	124
<b>8 Conclusion &amp; Perspectives</b>	<b>127</b>
<b>A Notation, Definitions, and Elementary Functions</b>	<b>133</b>
<b>B Discrete Transforms</b>	<b>135</b>
B.1 Orthogonal Series Expansion . . . . .	135
B.2 Continuous Wavelet Transform . . . . .	135
B.3 Discrete Dyadic Wavelet Series: The Discrete Wavelet Transform . . . . .	136
<b>C Additional Daily Profiles</b>	<b>139</b>

## Chapter 1 | Topic & Motivation

Most European, and other Western societies, are going through a phase of demographic change [1, 2, 3, 4]. Due to decreasing birth rates and steady medical progress, the average age is increasing. This leads to new challenges for health care and especially elderly care. There will be less qualified caregivers to take care of an ever increasing number of old people. Rising labor costs add to this problem. This has led to more and more technology being applied in the elderly care sector. In fact, a totally new field of technology called *ambient assisted living (AAL)* was created. “AAL aims at extending the time older people can live in their preferred home environment by increasing their autonomy and assisting them in carrying out activities of daily living, but also by the use of ICT<sup>2)</sup> products and the provision of remote services including care services that will assist them to achieve the autonomy, independence and dignity appropriate to their needs and conditions.” [5]

Typical application scenarios are detection of dangerous situations (fall, fire), reminders (intake of medication), monitoring of health and activity, or playful training of physical and cognitive fitness. Other benefits of technology in elderly care are the possibility to objectively discuss changes or anomalies, and to monitor long-term treatments conveniently. The demographic change presents a huge commercial potential for these technologies because of the number of older people growing rapidly, and their endeavor to live in their own homes for as long as possible [6]. AAL is also an emerging field for research and technology. The special challenge is that there are often suitable technologies to solve AAL problems available from other areas, but they need to be adapted to the special requirements inherent in the field.

An important goal of AAL is the automatic analysis and prediction of residents’ behavior and activity inside their homes [7, 8]. This shall provide caregivers with a cheap and always available monitoring instance. A possible application concept is described in section 1.2 in more detail. In order to analyze human activity, a variety of data regarding the person(s) to be monitored must be available. Data acquisition often requires powerful and expensive sensor technology that needs to be installed inside the monitored home. This is not only a complex and expensive approach, but also means intruding on the residents’ private lives.

The basic motivation behind the research presented in this thesis is therefore to extract activity information from indirect measurements, or from sensors that are already available for other purposes. This approach is non-intrusive and inexpensive, two crucial arguments for acceptability. The pioneering and running costs of the overall system are very important in the private health care sector. Two approaches to non-invasive and non-intrusive information acquisition for activity monitoring in AAL scenarios are examined in this thesis.

The first part deals with device-free passive (DFP) human presence detection. This is an active location monitoring technology which incorporates networks of radio sensors that are installed inside the home. The advantage over other approaches is that the person being monitored is not required to wear or carry a piece of equipment with them. However, this still requires the installation of dedicated equipment.

Hence, non-intrusive load monitoring (NILM) data, i.e. the disaggregation of individual appliances from the total consumption of electric energy, is used in the second part to infer human activity. Whenever an appliance is operated, it leaves a trace in the home’s electricity

---

<sup>2)</sup>[ICT stands for *information and communication technology*.]

profile. Detecting and classifying these traces can be done without installing sensors to each appliance by implementing them in smart meters. Hence, NILM is a convenient, unobtrusive, and cheap solution because smart meters are likely to become mandatory.

Another benefit of using NILM as a monitoring source is its inherent detection of an appliance that could pose a threat to human health, or life. About 30% of all accidents with devices in private homes involve electric appliances [9]. An automatic supervision from a central point in the smart meter helps to reduce the risk and prevent residents from harm.

## 1.1 Interpretation of Information

Indirect sensor information cannot be directly used for activity monitoring. It needs to be interpreted in order to derive instructions regarding actions to be taken by caregivers. Some examples are given in Fig. 1.1. The two raw data sources in this thesis, DFP presence detection and NILM, provide two basic types of information. Firstly, since many appliances are operated manually, NILM is an ideal source of information on the *daily activities* of residents. Secondly, most appliances are stationary in the home. Hence, if an appliance is operated manually, the *approximate location of a resident* in the corresponding time instant is known. This is supported by DFP presence detection which gives a direct position estimate within its detection zones. The two types can be interpreted individually or in combination with each other. Immediate decisions are based on instantaneous data. These are especially useful for direct intervention, should a dangerous situation occur in the home. On a second level, the information can be aggregated leading to activity or position profiles. These allow for a more long-term analysis of residents' lives, and reveal if habits change over time.

The interpretation of raw sensor data is a complex process that can be done automatically or with human aid. The latter opens the door to a networked elderly care concept, which will be discussed in the next section.

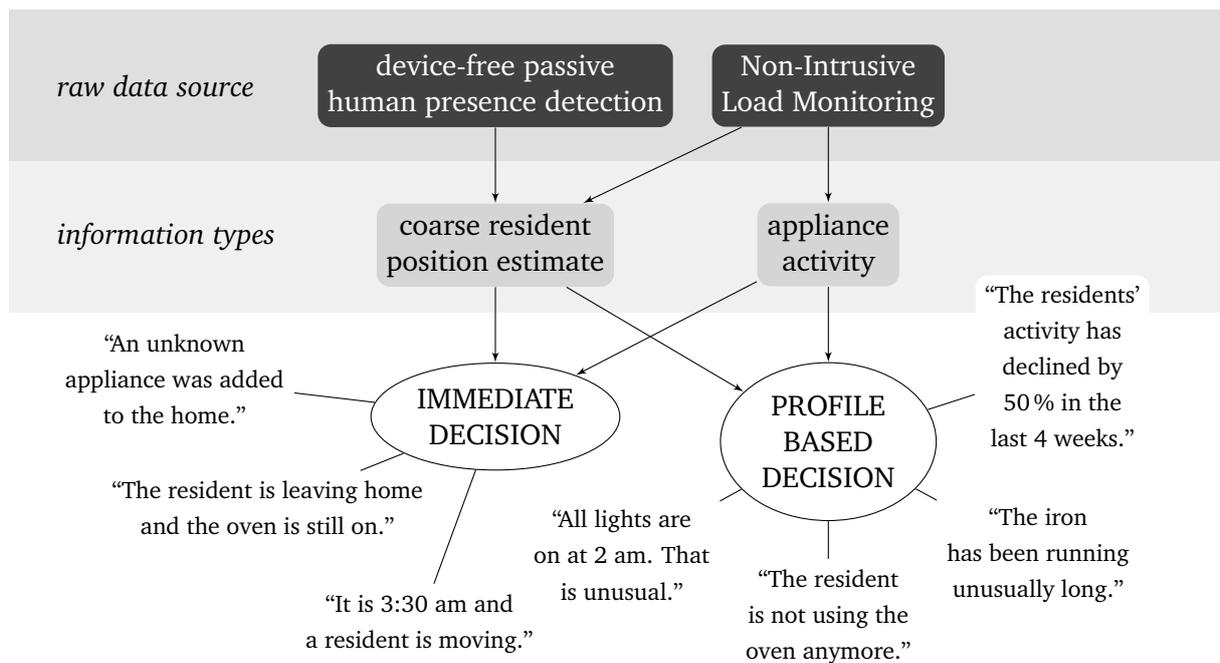


Fig. 1.1: examples of conclusions that can be drawn from indirect activity data

## 1.2 Application Scenario

The widespread availability of personalized and accurate information of a person's life can be an important factor for a better and more cost effective health care system. Having a tool for reliable, cheap and always available monitoring enables a larger circle of people than before to give help and care. It appeals to the networking habits of humans. The overall concept is drawn in Fig. 1.2. Raw sensor information is acquired inside the home, then filtered and processed in a data gateway. From here, information is transferred to different circles of people who interpret the data. According to their background and profession, these people give input to the residents in the form of advice, instructions, other immediate action, or a reflection of the residents' behavior. The circles of people are:

**Professional care services** that have employees visit, nurse, and look after elderly people are the traditional form. In severe cases, their expertise and professionalism are indispensable. However, due to rising labor costs and the shift in population structure (a decreasing number of young and increasing number of older people) in many Western countries [2], professional elder care is becoming more expensive.

**Relatives** have always been an important factor for elder care. In modern societies however, it is not as common for multiple generations to live under one roof, or at least in the same village or town. This makes it harder and more stressful for children and other relatives to stay informed. Technology can, to some degree, bridge the distance by allowing information to be transferred automatically.

**Informal caregivers** represent a new circle of care providers. Giving away personal information that relates to one's physical, psychological, or mental state to relatives or close friends is not something people are fond of. Opening themselves up to a group of anonymous but nevertheless interested people might be a more acceptable solution. This could be a circle of people who do not necessarily know each other personally. They are connected in a *social care network*. The task of data interpretation is shared, and input to one member is provided by the group.

**Residents** themselves profit from the data, as they are provided with a tool to analyze and reflect on their own behavior. Being monitored can create a sense of being safe in the own home; residents do not need to be afraid of forgetting actions that protect them from harm anymore.

The data gateway is a very important component to deal with privacy concerns. It manages which kind of information is passed to which circle. If it contains interpretation logic, the information can be passed to the residents directly.

## 1.3 Organization of this Thesis

This document is organized in eight chapters. An overview of relevant state-of-the-art work is given in chapter 2. There are two topics, in which research is presented. Chapter 3 describes the radio signal presence detection. Chapters 4, 5, and 6 deal with different aspects of non-intrusive load monitoring for AAL applications. Two case studies regarding the information content of Load Monitoring data are presented in chapters 7. An overall conclusion and outlook are given in chapter 8.

## 1.4 Contributions

The major contributions made in this thesis are as follows:

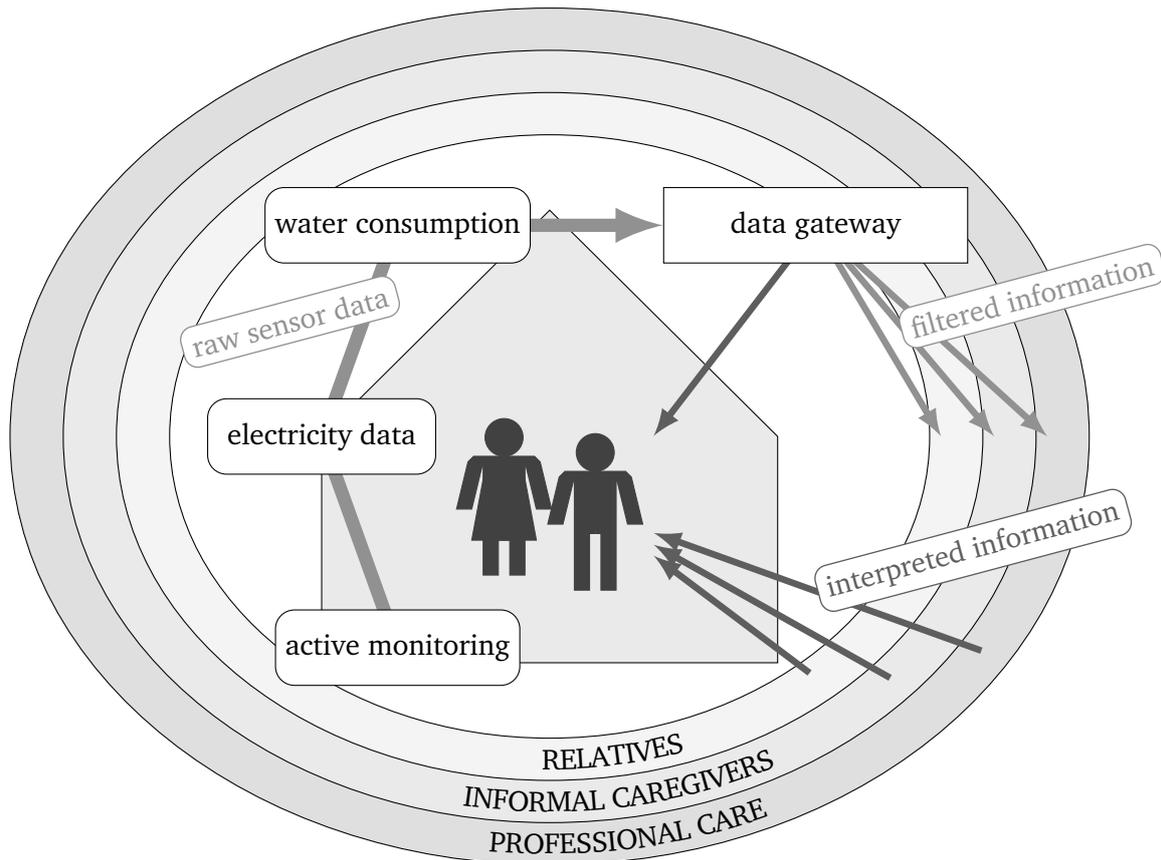


Fig. 1.2: visual representation of an elderly care concept that relies on networks including a circle of “informal caregivers”

1. A novel system of Bluetooth Low Energy radio nodes and an algorithm to detect human presence without transponders (3.2).
2. A new method for detecting switching events which significantly reduces the number of false positives (chapter 5).
3. An enhanced feature extraction method which leads to more reliable features for power difference classification (chapter 6).
4. An alternative approach to features for appliance classification which requires a minimal amount of training data (chapter 6).
5. A study of what sampling rate is necessary for successful application of these new features (chapter 6).
6. A catalog of characteristic behavior that can be exploited by these new features (section 6.7).
7. A study of residents’ behavior using raw activity profiles and appliance specific curves. The study is based on the public data set UK-DALE and own data that was specifically acquired for this thesis (chapter 7.1).
8. Finally, an evaluation of how threats from overheating electric appliances can be detected and avoided (chapter 7.2).

Contributions were taken from the following theses, which were co-advised by the author:

1. In his Bachelor’s thesis, Simon Gleissner provided data on the properties of Bluetooth Low Energy links for sections 3.1 and 3.2 [10].

2. A contribution to the fingerprint synchronization in section 6.5.6 and the appliance fingerprint catalog in section 6.7 were made by Alexander Bartler in his Bachelor's thesis [11].

During the work on this thesis the following publications were made:

- Philipp Klein, Dirk Benyoucef, Jean Merckle, and Djaffar Ould Abdeslam. Analysis of fingerprints of electric appliances as starting point for an appliance characteristics catalog. In IECON 2014 - 40th Annual Conference of the IEEE Industrial Electronics Society, pages 3517-3521, October 2014 [12].
- P. Klein, J. Merckle, D. Benyoucef, and T. Bier. Test bench and quality measures for non-intrusive load monitoring algorithms. In IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society, pages 5006-5011, November 2013 [13].
- T. Bier, D. Benyoucef, D. Ould Abdeslam, J. Merckle, and P. Klein. Smart meter systems measurements for the verification of the detection & classification algorithms. In IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society, pages 5000-5005, November 2013 [14].

The following papers are prepared for publication:

- P. Klein, J. Merckle, and D. Benyoucef. Device-free Passive Human Presence Detection in Ambient Assisted Living Environments, 2015. Prepared for publication in IEEE Transactions on Mobile Computing.
- P. Klein, J. Merckle, and D. Benyoucef. Device-free Human Presence Detection for Human Activity Recognition in Ambient Assisted Living Applications, 2015. Prepared for publication in IEEE Sensors Letters.



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

This chapter gives an overview of state-of-the-art research that is relevant to the topics of this thesis. A detailed introduction into individual approaches or methods is given alongside the other chapters wherever useful.

### 2.1 Activity Monitoring

Activity monitoring is the process of acquiring sensor data, inferring information on the residents' activity, and transmission of this information to the residents themselves, caregivers, or other concerned parties. Activity monitoring can have many goals:

- *Inactivity detection* detects when a person has been inactive for an unusually long period. Since this could be a sign of fainting, falling, or unexpected leaving the home, an alarm is triggered when a too long inactivity period is detected [15]. The problem is that the threshold has to be very high, since there are many normal situations where no activity can be detected. This can be avoided, to some degree, by learning the (in)activity profile for each person individually. For inactivity detection, the extracted features have to be represented in a way such that anomalies can be detected. The use of multiple data sources requires that each source's data be aligned and normalized in order to result in a meaningful profile [16, 17].
- The second goal of activity monitoring is *Human Activity Recognition (HAR)*. In contrast to inactivity detection, a classification of *what* activities residents are performing is of interest. HAR aims at the recognition of predefined activities, which can be hard to define a priori. Obviously, this requires much higher detailed information than inactivity detection.
- The third goal of activity monitoring is therefore *activity pattern discovery (APD)*. This does not require predefined models. Instead, sensor information is scanned for unknown regular patterns or, to the contrary, irregularities or random events in the residents' life.

The fundamental step for activity monitoring is data acquisition. Barger et al. used a vector of sensor location, timestamps, duration of events and activity measures acquired from passive infrared (PIR) sensors and appliance level power meters in order to detect several Activities of Daily Living (ADLs)<sup>3)</sup> [18]. Another approach, using PIR sensors with the addition of a temperature sensor for the stove and a sensor pad in the bed, was presented by Virone et al. [19]. Brdiczka et al. evaluated 3D video trackers to classify actions such as “walking” or “interaction with table” [20]. Amor and James used a set of sensors for PIR signatures, ambient light, ambient noise levels, as well as a camera and an infrared (IR) remote sensor capturing the TV remote signals for activity monitoring [16]. All these methods required intrusive and sometimes very expensive sensors that are difficult to accept in a private home. Acquiring information non-intrusively is therefore an important improvement for activity monitoring.

---

<sup>3)</sup>ADLs are daily routines like cleaning, preparing meals, eating, sleeping, etc.

## 2.2 Device-Free Passive Presence Detection

A first step towards non-intrusive information acquisition is DFP human presence detection. As the name suggests, it detects human beings passively (i.e. they do not actively participate in their monitoring). Instead, wireless sensor networks are employed to do this [21]. The basic principle is the molecular resonance of the human body which causes variations of electromagnetic fields in wireless networks. Several efforts were done to model the impact of the human body on electromagnetic waves, e.g. by Ghaddar et al. [22] for 10 GHz and Nechayev et al. [23] for the common 2.45 GHz band. Other studies analyzed the absorption behavior for mobile communication links [24] and for 2.4 GHz [25]. The effect of the human body on electromagnetic radiation was exploited for a variety of applications in the past. Most DFP applications can be classified into one or more of the following categories:

**Presence detection:** this refers to the sole detection of human presence within a monitored area. The result is a binary output (“presence” or “no presence”), or, depending on the algorithm, a probability of human presence.

**Localization:** in this category, the goal is to determine the location of a human inside a monitored area. The output is a coordinate, and possibly a probability. Note that localization can either be achieved by actually estimating the position, or by performing a presence detection for multiple cells. The precision of the latter is limited by the spacial cell size.

**Tracking:** for some applications it is interesting not only to localize a person but to track its path through a monitored area. The output is a sequence of coordinates, as well as velocity or direction information.

The following survey of relevant papers is roughly sorted into the applications envisioned by their authors and the base technology employed. Intrusion detection and intruder tracking can be regarded as the first application of the DFP methodology. In the beginning, communication links of existing wireless local area network (WLAN) links (IEEE 802.11) were analyzed [26, 27, 28, 21, 29]. Since WLAN is designed to require only a few access points for large areas the precision of the localization is not optimal. Therefore, additional wireless networks based on the wireless personal area network (WPAN)/ZigBee standard (IEEE 802.15.4) were analyzed [30, 31, 32, 33, 34]. It describes meshed grids of radio nodes for sensor networks. The mesh of nodes allows for an easy transfer of radio link information, and a multitude of detection cells can be created. The size of each cell is much smaller than a WLAN cell which increases the sensitivity and precision of intrusion detection and localization.

Gradually, the focus shifted from specialized intruder detection towards detection of humans in general scenarios. Relevant work can be found for the three DFP categories mentioned in the introduction section. Human presence detection using WLAN can be found in [35, 36, 37, 38, 39, 40, 41, 42]. WPAN networks were also used for presence detection, e.g. in [43, 44, 45, 46]. DFP localization was introduced using WLAN [47, 48, 49, 50, 51, 52], WPAN [53, 54, 55, 56, 57], as well as a combination of these [58]. For high localization accuracy, a high density of nodes is required, which is why additional radio signal standards were analyzed; examples are active radio frequency identification (RFID) nodes [59, 60]. Zhang et al. proposed a combination of 802.15.4 and 303.825 MHz active RFID [61]. Other localization work was done using proprietary protocols on 433.1 MHz and 909.1 MHz [62, 63], as well as frequency modulation (FM) radio (87 MHz to 111 MHz) [64]. Tracking people is to some degree related to localization. Nevertheless, some papers were presented specifically targeting this field. Again, WLAN [65, 52] or WPAN [66, 55, 67] were used as common base technologies. Other authors made use of the 868 MHz to 928 MHz frequency band [68, 69].

The DFP approach was also used in more specialized applications, like people counting [37, 70], gesture recognition [71], velocity estimation [72], and interaction with mobile electronic devices [73]. In the recent years a few studies were published that are especially interesting

for AAL applications: Patwari et al. showed how to detect human breathing and estimate the breathing rate in a DFP way [74]. The recognition of human activities using WLAN [75] or 802.15.4 [76] has also been a research topic.

The simple principle of operation makes DFP an interesting technology for HAR. Through its localization and presence detection abilities, it can provide vital information on where inhabitants are, and what activities they perform. There are, however, some requirements that are especially important in AAL applications. The components need to be discrete in a sense that, when mounted, they are invisible to the inhabitants to make sure they do not behave abnormally with fear of being monitored. Additionally, it is important that the system topology is easy to install and maintain. For the private health care sector, it is also important that all required hardware is cheap to buy, and integrates well into standard communication systems. The technology used so far does not meet these requirements. WLAN covers large areas, thus reducing accuracy, and requires high energy levels for operation. WPAN components have never conquered the consumer market. They are still comparably expensive and inflexible to integrate into more common communication networks. For AAL applications more research is required to address these issues.

## 2.3 Non-Intrusive Load Monitoring

The next step towards even less intrusion is NILM. The operation of an electric appliance is an activity that can be inferred from the total electricity consumption. If implemented in the smart meter, this does not require any additional sensors inside the home at all. Clement et al. proposed an approach where smart meter data is evaluated in order to deduce the ADLs performed by the residents [77, 78]. Patel et al. analyzed high frequency components in the voltage signal for a decentralized classification of switching events [79]. Their results showed that it is possible to find out which switch was flipped if say, a room light is connected to multiple switches. Although this is very useful information for the activity domain, the authors mention that these features are very sensitive. Due to the high selectivity, there are a large number of classes to be trained, and the features are likely to change when an appliance is moved to a different location (than the training situation) or used at a different time of the day. In the context of AAL a lot of research is still required regarding a suitable overall structure with respect to implementation and usability, an event detector for activity monitoring purposes, and an appropriate choice of features.

### 2.3.1 Disaggregation Procedure & Implementation

NILM is the disaggregation of the total electricity consumption of a home into individual appliances. It can be divided into several sub-problems, each of which is discussed in one of the following subsections. The origins of the research area date back to the 1980s. Ever since, disaggregation has almost always been regarded as a linear sequence of measurement, event detection, feature extraction and classification, and energy tracking [80], as depicted in Fig. 2.1. This is a straightforward structure that can be easily implemented on any kind of computing machine. The downside is that errors propagate throughout the whole process. In 2010 Liang et al. developed a generalization and formalization of features and the disaggregation process. Their goal was to combine a multitude of features and algorithms using a “committee decision mechanism” in order to improve performance. They showed that, in selected use cases (i.e. for single appliances), the combination yields better results than one of the individual algorithms alone is able to achieve [81, 82].

Due to its simple yet established structure, and the lack of alternative research, the linear sequence is adopted to this thesis as well. However, the two important parts, event detection and feature extraction, will be analyzed independently in chapters 5 and 6, respectively. The last part, energy tracking, is responsible for an estimation of how much energy an appliance

has consumed in a given period. This is not relevant for this thesis and will therefore not be mentioned any further.

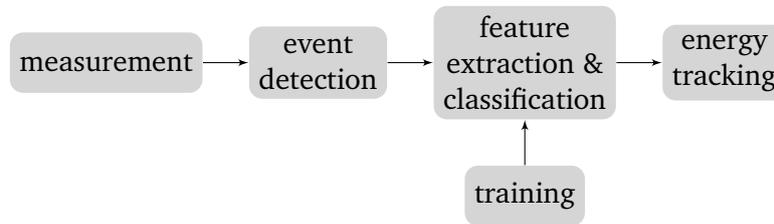


Fig. 2.1: traditional linear structure for the disaggregation of electric appliances

### 2.3.2 Event Detection

For disaggregation, a set of features is computed from the measured data under the assumption that they are characteristic for the connected appliances. In order to minimize the search space and the number of features to classify, an event detector is used prior to feature extraction and classification in the classic linear structure of operation. At this stage, features are only computed for switching events instead of the whole signal. Although important, event detection was never a prominent topic of NILM research. Only in recent years have some specialized methods been developed. Before that, event detection was a simple differentiation of the power signal combined with a fixed threshold [83, 84]. A newer approach was presented by Anderson et al. They defined a set of quality measures based on event counting mechanisms, and a measure that uses the step height (power levels) to weigh the events. This is more suitable in the NILM context, because it reduces the influences of minor events of low step heights [85]. An interesting approach to event detection was published by Barsim et al. [86, 87, 88]. This will be described and used as a reference method in chapter 5. Further research is required regarding a performance comparison of the different approaches and their suitability in the context of Activity Monitoring for AAL applications.

### 2.3.3 Features & Feature Extraction

The second disaggregation step is the actual extraction of features. A group at the Massachusetts Institute of Technology (MIT) proposed a first approach in 1986 [84, 80]. The appliance power consumption during operation was modeled by linear time-invariant (LTI) complex impedances, corresponding to their steady state power consumption (rectangular approximation), at a rate of 1 Hz. Complex appliances were divided into LTI loads by using finite state machines (FSMs). The features used for classification were the power differences for each event, called *steady state power features (SSPFs)*. Another topic was the implementation on computers available at that time. This required a linear structure of operation steps. This steady state power approach was later used by Pihala to assign appliance labels to smart meter data [89]. Marceau and Zmeureanu were able to detect the three highest load appliances in a household (water heater, electric room heater, and refrigerator) based on their steady state power consumption. The error was around 10% [90]. Baranski also used steady state features, which were acquired from an optical sampling of the electromechanical Ferraris counter that were very common in Germany before 2006 [91]. The problem is that the low number of appliances mentioned in the papers casts doubts on the credibility and significance of the results.

NILM, using SSPFs, is especially successful in scenarios with a small number of individual appliances, which is the case in the residential homes. In more complex scenarios, some of the simplifications for appliance behavior have to be abandoned. Cole and Albicki acquired current harmonics for the separation of 10 appliances (mostly heating, ventilation and air condition-

ing (HVAC)<sup>4)</sup> and refrigerators/freezers) [92]. Laughman et al. showed that it is possible to separate appliances by their current harmonics, even if the corresponding SSPFs overlap in the  $PQ$  plane<sup>5)</sup> [93].

Another feature for NILM is transients. Transients are very short sections of the instantaneous current or power, that reoccur in every operation cycle. The assumption of time-invariant loads is disregarded. A number of papers dealing with specially tailored problems were published. Norford and Leeb worked on a combination of SSPFs and transient analysis to centrally detect HVAC equipment in business buildings [94]. Whichakool et al. presented a method to distinguish variable speed drives (VSDs) from constant loads [95]. Transients were also considered for the disaggregation of appliances in general. Their temporal sequence was used by Leeb et al. as a feature to classify switching events [96]. However, the vast number of appliances makes it difficult to generalize transients across classes or groups of appliances. This lead Chang et al. to analyze not the time-series signal, but the energy of transient current pulses. They were successful in a limited number of test cases [97, 98]. The work of Wang took the next step; he approximated the power consumption of appliance by series of two geometric objects, namely “triangles” (transient ramps) and “rectangles” (steady state power phases) [99]. This dramatically reduces the number of possible signal shapes. However, since there was only a small volume of data (three appliances), the results presented are not significant. Yang et al. proposed a separation of steady state and transient periods by applying a cluster analysis to the power signal. Steady state periods are characterized by conventional SSPFs. Transient signals are approximated by sixth order polynomials. The feature vector combines the SSPFs and the polynomials’ coefficients [100, 86, 101].

Another approach requiring high-grained data uses steady state high-frequency features. Gupta et al. found that signatures can be found in the frequency band between 36 kHz and 500 kHz, making them suitable to reliably disaggregate appliances. However, they found that the crosstalk (inductive or capacitive coupling) between phase wires has a negative influence on the features [102].

In all approaches described thus far, the idea was to regard the disaggregation problem as classifying a sequence of switching events. The power consumption was implicitly assumed to be constant between switching events. A different method was chosen by Dong et al. They modeled appliances by using dynamic systems whose transfer function brings the user activity and the power signal into relation. Each model is trained off-line during a “system identification phase”. The disaggregation problem is then to identify the most probable combination of models for each new value that is recorded by the smart meter [103]. In their paper, the authors assumed LTI systems, an extension to different models is claimed to be possible. They also showed how to solve the problem using adaptive filters [104], but only three appliances were used for verification. In another analysis, they deduced some fundamental limitations of NILM algorithms [105]. As a result, they state that a sampling rate of at least  $1/5$  Hz is required for successful disaggregation. This will be of special interest in chapter 6.5 of this thesis.

A systematic comparison of the most important approaches described before is given in Table 2.1. From this table, it can be seen that a variety of algorithms exists, but they were applied to a very limited number of data sets. For Activity Monitoring, it is important to find features that can be used for a separation of manually and automatically operated loads. A variety of methods was described in this subsection. A selection and test of suitable features, leading to sufficient reliability, is still an open question for research.

---

<sup>4)</sup>The term HVAC has two meanings: in this thesis, HVAC always stands for “heating, ventilation and air conditioning”. If “high-voltage alternating current” is meant instead, this will always be written out.

<sup>5)</sup>The  $PQ$  plane refers to a representation of real and reactive power, a concept of alternating current (AC) theory. It will be described in section 4.5.4.

Table 2.1: comparison of NILM feature extraction approaches

author	year	references	sample rate	appliances	features & details
Hart	1986	[80]	1 Hz		steady state ( $P$ , $Q$ )
Pihala	1998	[89]	1 Hz		steady state ( $P$ , $Q$ )
Marceau	2000	[90]	16 s	3	steady state ( $P$ )
Ito	2004	[106]	$\geq 50$ Hz	33	steady state (current waveform)
Baranski	2006	[91]			steady state ( $P$ )
Reeg	2010	[107]	1 Hz	ca. 7	steady state ( $P$ , $Q$ )
Jiang	2012	[108]	$\geq 50$ Hz	11	steady state (current waveform)
Basu	2013	[109]	1 h	4	steady state ( $P$ )
Cole	2000	[92]	$\geq 1.8$ kHz	10	steady state (current) harmonics
Laughman	2003	[93]	$\geq 8$ kHz	single	steady state ( $P$ , $Q$ , current harmonics)
Norford	1996	[94]		single	steady state ( $P$ , $Q$ , harmonics); only HVAC
Wichakool	2007	[95]	$\geq 1.5$ kHz	3	steady state (harmonics); only VSDs
Gupta	2010	[102]	$\geq 1$ MHz	94	high frequency steady state features
Leeb	1995	[96]			transient time series
Chang	2008	[97, 98]	$\approx 15$ kHz	3	transient energy
Wang	2011	[99]	10 Hz	3	transients triangles & steady state rectangles
Dong	2013	[103, 104]	12 Hz	3	dynamical (LTI) models

### 2.3.4 Data Sets

Towards the end of the 2000s, the construction of data sets containing so called “ground truth data” became a field of research in the NILM context. Ground truth data is recorded manually or automatically during the measurement of voltage and current. They indicate the actual state of the connected appliances. Without this data, it is impossible to assess and validate existing or new NILM approaches. The start was made by Berges et al., who, in 2009, developed a measuring system especially tailored to the problem of ground truth data collection [110]. This resulted in the publicly available data set BLUED [111]. Other examples of data sets are REDD [112], tracebase [113], Smart\* [114], and UK-DALE [115]. They will be described in section 4.7 in more detail. During the work on this thesis, some own work was published in this area of study as well [14, 13].

Since 2010, a multitude of data sets have been published. They were created with different backgrounds, resulting in different data formats and contents. In order to standardize procedures and allow for comparability of the results, an international research group of Batra, Kelly, Parson, et al. created the “Open Source Toolkit for Non-Intrusive Load Monitoring” (NILMTK) [116]. An important feature is the Python script which is an introduction to a standardized data format, NILMTK-DF. Other features are functions for analysis and preprocessing of the data, benchmark algorithms, and methods for the performance assessment of NILM algorithms. NILMTK not only supports electrical signals, it is also able to take meta data, like information about the measurement setup, into account [117].

The different backgrounds might be a problem for a valid evaluation of Activity Monitoring using NILM. A suitable selection therefore must be made, before using any data sets.

### 2.3.5 Training & Setup

Independent of the chosen features, training of the classifiers is a special problem. A data base of features to compare with must be created prior to actually using a classification algorithm.

Goncalves et al. analyzed unsupervised training methods, which resulted in the possibility of performing a successful clustering of the  $PQ$  data for appliances with high power consumption [118]. Reeg and Overbye showed that automatic k-means clustering of SSPFs needs manual tuning in order to yield correct training data [107]. In his PhD thesis, Parson developed models based on Hidden Markov Chains, which are able to automatically parametrize themselves in an installation [119, 120].

For a robust disaggregation, it is necessary to organize, and systematically analyze the signatures of electric appliances, as stated by Lee in 2004 [121]. He built a catalog of 11 appliances based on their current signal time series. Another attempt was made by Lam et al., who grouped typical home appliances by a cluster analysis, according to voltage-current trajectories<sup>6</sup> and traditional SSPFs [122]. The efforts are by far not exhaustive, so the algorithms have to be trained locally each time, leaving room for optimization.

---

<sup>6</sup>Those trajectories occur when instantaneous voltage and current are plotted against each other period-wise.



## Chapter 3 | Activity Monitoring Based on device-free passive Presence Detection Using Radio Signal Distortion

### 3.1 Introduction

The major benefit of activity detection using NILM is that the only sensor required is mounted in a place that is usually hidden from the residents' eyes. It is the most non-intrusive solution possible. However, there are a few limitations. The information provided by NILM is of little use if people do not actively use electric appliances. In other words, NILM data is a superposition of some human activity patterns and data from continuously running appliances (refrigerators, heatings, air conditioning, etc.). This is still vital information for the detection of hazardous situations, but it may limit the use for activity pattern discovery. In this case, secondary technology is required: a more intrusive system, with dedicated components, needs to be installed in the home. DFP human presence detection is discussed in this chapter as a solution that does neither incorporate visible equipment, nor require inhabitants to carry components with them. A system example will be presented in chapter 3.2.

Applying a DFP presence detection network in Activity Monitoring scenarios requires some special constraints that are different from other applications. The goal is to detect human movement or presence as an indication of activity, with a single binary output for each monitored area. For the use in residential environments, low cost and easy installation of the required system components are essential. The system design should be flexible enough to allow one or multiple monitored areas. To ensure easy installation, the sensor nodes should be low energy types, able to be powered by batteries. If possible, low complexity detection algorithms should be used so that they can be implemented on one of the low energy sensor nodes. The training effort should be as low as possible, again to simplify installation. For this thesis, the received signal strength indicator (RSSI) is chosen as the only activity indicator, so that cheap and simple radio nodes can be used.

These constraints eliminate conventional presence detection systems, like light barriers, PIR sensors, or cameras because they need a visual connection to the monitored area. This makes it hard to mount them inconspicuously. Furthermore, they are usually quite expensive. Radio waves, as a detector medium, are an interesting alternative because they travel through materials, and are easy to generate and evaluate.

The remainder of this chapter is dedicated to an introduction to radio waves and state-of-the-art detection algorithms.

#### 3.1.1 Radio Waves and Other Relevant Terms

*Radio waves* are electromagnetic waves whose frequency is in a selected range, the so called *radio spectrum* from 3 Hz to 3 THz [123]. The frequency of the radiation used for DFP detection is an important factor for the system design. Its value determines the physical size, as well as the operating distance. For applications in the private sector, the selection of the frequency is limited to industry, science, and medicine (ISM) bands. Those are frequency ranges that are allocated

for free use in the mentioned fields of application. Frequencies outside these bands are usually strictly allocated to other technical areas, like military communication, civil communication (GSM<sup>7</sup>), UMTS<sup>8</sup>), etc. The ISM bands are listed in Table 3.1.

Table 3.1: frequencies of the ISM bands (from [124]).

lower limit	center	upper limit
13.530 MHz	13.560 MHz	13.567 MHz
26.975 MHz	27.120 MHz	27.283 MHz
40.660 MHz	40.680 MHz	40.700 MHz
2.400 GHz	2.450 GHz	2.500 GHz
5.725 GHz	5.800 GHz	5.875 GHz
24.000 GHz	24.125 GHz	24.250 GHz

The most popular radio signal standards in the non-military consumer area are WLAN<sup>9</sup> and Bluetooth<sup>10</sup>. WLAN is used for wireless local area networks. Bluetooth is intended for short ranges and medium data transfer rates. Both standards use frequencies in the ISM band from 2.4 GHz to 2.5 GHz. Another known standard using this band is WPAN/ZigBee<sup>11</sup>. Due to its small range of applications in the consumer area, this is not considered in this thesis. An especially interesting standard is Bluetooth Low Energy (BLE), an extension to the original Bluetooth standard that was included in version 4.0. BLE is an independent standard marketed under the same Bluetooth trade label. It was developed for small, low-energy devices that are powered by coin cells, or other low capacity energy storage units. This makes this technology an ideal choice for solving the problems, given the requirements that were discussed at the beginning of this chapter.

The propagation and interaction of radio waves with objects is a very complicated matter that exceeds the scope of this thesis. An overview of the topic can be found in the book by Werner Bächtold [125]. Fortunately, a simplified model for the radio transmission between a transmitter and a receiver exists. The so called *Fresnel zone* defines an ellipsoid space, which has to be clear of any absorbing matter in order to achieve undisturbed communication [126]. The model is sketched in Fig. 3.1a). The Fresnel radius is given by

$$r_f(d_1) = \sqrt{\frac{n \cdot \lambda \cdot d_1 \cdot d_2}{D}} = \sqrt{\frac{n \cdot \lambda \cdot d_1 \cdot (D - d_1)}{D}} \quad (3.1)$$

with the Fresnel zone number  $n$ ,  $\lambda$  is the wavelength of the radiation<sup>12</sup>), and  $D$  is the distance between transmitter and receiver.  $d_1$  and  $d_2$  mark the distance from either component to the point where the radius is sought. The maximum Fresnel zone radius is given at

$$\arg \max_{d_1} \{r_f(d_1)\} = \frac{D}{2}, \quad (3.2)$$

i.e.  $d_1 = d_2$ , by

$$b_f = \max_{d_1} \{r_f(d_1)\} = \frac{1}{2} \cdot \sqrt{n \cdot \lambda \cdot D}. \quad (3.3)$$

The maximum radius as a function of  $D$  is shown in Fig. 3.1b) for different values of  $b_f$ .

<sup>7</sup>)The global system for mobile communication (GSM) is a standard for cell phone mobile communication.

<sup>8</sup>)The universal mobile telecommunications system (UMTS) is the successor of GSM.

<sup>9</sup>)Defined in the standard IEEE 802.11.

<sup>10</sup>)Defined in the standard IEEE 802.15.1.

<sup>11</sup>)Defined in the standard IEEE 802.15.4.

<sup>12</sup>)Wavelength  $\lambda$  and frequency  $f$  are related via  $\lambda = c/f$ , where  $c$  is the propagation speed.

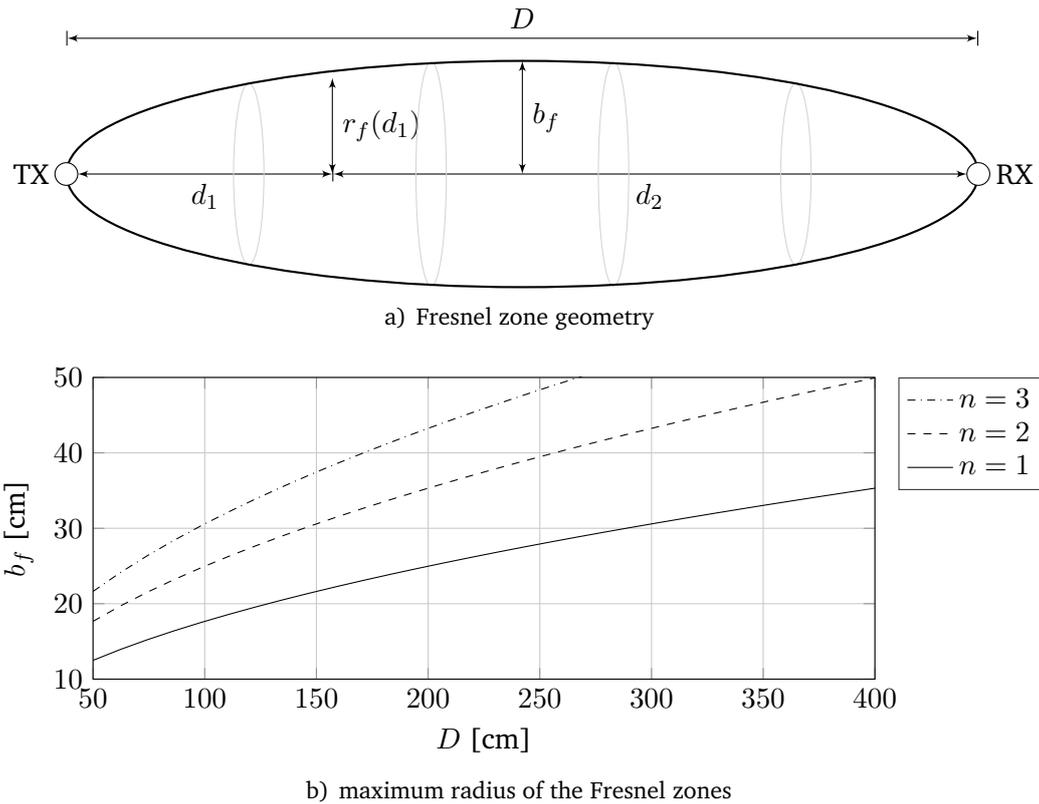


Fig. 3.1: the Fresnel zone as a model of the electromagnetic field between sender and receiver

If an electromagnetic wave hits a metallic surface, it induces eddy currents. These eddy currents retransmit an electromagnetic wave at the edges of the material, which can be interpreted as a *reflection* [125]. This means that an electromagnetic wave does not necessarily arrive at a receiver exclusively via the shortest direct path, or the *line of sight*. It may hit some reflection points, which means that the receiver captures multiple rays from the same source. This is called *multipath propagation*.

If the eddy currents are conducted away from an object's corners (e.g. by a low-resistance connection to electric ground), the magnitude of the reflected wave can be strongly reduced. This means that the incoming wave is *absorbed*, rather than being reflected. This means that, under some conditions, objects are able to shadow the radiation.

Another absorption mechanism is the transformation of energy into heat. The electromagnetic radiation is able to excite molecules that are able to move (typically in liquids or gases). The absorption reaches a maximum when the molecules' resonance frequency is equal or very close to the radiation frequency. Due to its high water content, the human body has a measurable influence on electromagnetic fields [24]. If an electromagnetic wave hits the human body, it is partially or fully absorbed by heat dissipation inside the tissue. This is the basic concept that is exploited for DFP detection using radio signals.

### 3.1.1.1 Attenuation Characteristics of Human Skin

The amount of energy absorbed by an object depends on two parameters, the *penetration depth*, and the *effective surface* that is exposed to the radiation. The penetration depth measures the distance an electromagnetic wave travels into a material until its intensity reaches  $1/e \approx 37\%$  of the intensity just below the surface. It depends on the frequency, as well as material parameters. If the object is thicker than the penetration depth, most of the energy it is exposed to is absorbed. In the other case, if the object is thinner than the penetration depth, a significant part of the radiation passes through it, decreasing the amount of energy absorbed. The penetration depth

$\delta = 1/\beta$  is the reciprocal of the attenuation constant  $\beta$ , which is given by [127]

$$\beta = \sqrt{\pi \cdot f \cdot \kappa \cdot \mu}, \quad (3.4)$$

where  $\kappa$  is the conductivity of the material, and  $\mu$  is its permeability. The permeability  $\mu = \mu_r \cdot \mu_0$  is often given by the product of material dependent relative permeability  $\mu_r$  and vacuum permeability  $\mu_0 = 4 \cdot \pi \cdot 10^{-7} \text{ N/A}^2$ . The parameters of human skin are given in Table 3.2. From these parameters the attenuation constant can be computed to be

$$\beta_{\text{skin}} = 721.44 \text{ m}^{-1}. \quad (3.5)$$

This yields a penetration depth of

$$\delta_{\text{skin}} = 1.37 \text{ mm}. \quad (3.6)$$

This means that human skin can be treated as a two-dimensional object, because it will completely absorb electromagnetic radiation at 2.4 GHz.

Table 3.2: electric and magnetic parameters of the human skin for  $f = 2.4 \text{ GHz}$  (from [25])

symbol	parameter	value
$\mu_r$	relative permeability	38.0629
$\mu$	permeability	$4.783 \cdot 10^{-5} \text{ N/m}^2$
$\kappa$	conductivity	$1.4408 \text{ S/m}$

### 3.1.1.2 Signal Quality Estimation by Received Signal Strength Indication

The RSSI gives an approximate indication of the radio link quality by measuring the power level received by a receiver [128, 129]. It is usually expressed as a ratio of the received power  $P_{\text{RX}}$  to 1 mW

$$\text{RSSI}_{\text{dBm}} = 10 \cdot \log \left( \frac{P_{\text{RX}}}{1 \text{ mW}} \right), \quad (3.7)$$

and expressed in dBm units. If known, it can also be expressed as a ratio of received and transmitted power  $P_{\text{TX}}$

$$\text{RSSI}_{\text{dB}} = 10 \cdot \log \left( \frac{P_{\text{RX}}}{P_{\text{TX}}} \right), \quad (3.8)$$

where it is usually measured in dB. As discussed in section 3.1.1.1, at 2.4 GHz, the human body absorbs most of the radiation it is exposed to. Therefore the RSSI, as defined in eq. (3.7) or (3.8), can be used to indicate when a person is crossing a transmission line.

### 3.1.2 Reference Presence Detection Methods

In this section, relevant detection methods are described in more detail. They will be used in the next chapter as a benchmark for an own algorithm. To begin the section, two algorithms that were presented by Youssef et al. [21, 35] are discussed.

Let  $x[k]$  be the signal of measured RSSI values, and

$$d[k] = x[k+1] - x[k] \approx \frac{d}{dk} x[k] \quad (3.9)$$

be the sequence of corresponding fluctuations, i.e. an approximation of the derivative of  $x[k]$ .

The first method uses a long and a short sliding window to compute the averages

$$\bar{x}_l[k] = \frac{1}{w_l} \cdot \sum_{i=k}^{k+w_l-1} x[i] \quad \bar{x}_s[k] = \frac{1}{w_s} \cdot \sum_{i=k+w_l}^{k+w_l+w_s-1} x[i], \quad (3.10)$$

where  $w_l, w_s \in \mathbb{N}$  are the corresponding window sizes and  $w_l > w_s$ . The two averages are compared at each time step  $k$ . If

$$\left| \frac{\bar{x}_l[k] - \bar{x}_s[k]}{\bar{x}_l[k]} \right| > x_{\text{thr}} \Big|_{k=k_0} \quad (3.11)$$

an event is generated at  $k_0 + w_s$ .  $x_{\text{thr}}$  is a predefined threshold, which has to be determined from training. The optimal parameter values given by the authors are  $w_l = 20$ ,  $w_s = 5$ , and  $x_{\text{thr}} = 0.03$ . This method is called ‘‘Youssef I’’ in this thesis.

In the second method, the detection is based on the RSSI signal’s statistics. Hence, the sliding variance

$$v[k] = \frac{1}{w-1} \cdot \sum_{i=k}^{k+w-1} \left( x[i] - \frac{1}{w} \cdot \sum_{\ell=k}^{k+w-1} x[\ell] \right)^2, \quad (3.12)$$

where  $w$  is the window size, is used for detection. During a training period from  $k = k_1$  to  $k_2$ , in which no event may occur, the average training variance,  $\bar{v}_t$ , and its standard deviation,  $\sigma_t$ , are computed by

$$\bar{v}_t = \frac{1}{k_2 - k_1 + 1} \cdot \sum_{k=k_1}^{k_2} v[k] \quad \sigma_t = \sqrt{\frac{1}{k_2 - k_1 + 1} \cdot \sum_{k=k_1}^{k_2} (v[k] - \bar{v}_t)^2}. \quad (3.13)$$

After training ( $k > k_2$ ), an event is generated at  $k_0 + w$  if

$$v[k] > \bar{v}_t + r \cdot \sigma_t \Big|_{k=k_0} \quad (3.14)$$

with  $r \in \mathbb{R}$ . The optimal parameters were given by  $w = 40$  and  $r = 4$ . This method is identified by ‘‘Youssef II’’ here.

Another interesting approach is to compare the distributions of samples under human presence and undisturbed signals. Lee et al. use histograms as an approximation of the distributions [31, 33]. Their method uses the derivative of the RSSI signal  $d[k]$ . Let  $h_{\Delta x}(x, k, w)$  be the histogram with bin width  $\Delta x$  of a section of  $x[k]$  that is limited by the sliding window from  $k$  to  $k + w - 1$ <sup>13)</sup>. Human movement is indicated at time  $k_0$  if

$$p[k] = \frac{\sum_{i=-1}^1 h_{\Delta d}(d, k, w)[i]}{\sum h_{\Delta d}(d, k, w)} < p_{\text{thr}} \Big|_{k=k_0} \quad (3.15)$$

where  $p_{\text{thr}}$  is a threshold that needs to be set up in advance from observations of undisturbed and disturbed histograms. The motivation behind this is that the undisturbed signal is represented by a single group, which is sharply concentrated around a mean<sup>14)</sup> in the histogram. Hence, the ratio of the sum of a few bins around this mean, and the sum of the whole histogram, is almost 1. The presence of a human will introduce fluctuations to the RSSI signal. This leads to an

<sup>13)</sup>A histogram is a method to visually estimate the distribution of values in  $x$ . Suppose  $x \in [a, b]$ . This interval is discretized into sub-intervals, so called *bins*, of width  $\Delta x$ , i.e. there are  $\frac{b-a}{\Delta x}$  bins. The histogram is the sequence of numbers of samples that fall into each bin.

<sup>14)</sup>Since the base signal is the derivative of the RSSI values the mean will be zero.

increased width of the group in the histogram. Consequently, the ratio  $p$  will decrease. Typical values were given by  $\Delta d = 1$ ,  $w = 100$ , and  $p_{\text{thr}} = 0.65$ .

Lin et al. enhanced this method by assuming that the RSSI fluctuations are normally distributed [37]. Using sliding windows, they computed the average,  $\bar{d}[k]$ , and standard deviation,  $\sigma[k]$ , of the derivative by

$$\bar{d}[k] = \frac{1}{w} \cdot \sum_{i=k}^{k+w-1} d[i] \quad \sigma[k] = \sqrt{\frac{1}{w-1} \cdot \sum_{i=k}^{k+w-1} (d[i] - \bar{d}[k])^2}, \quad (3.16)$$

where  $w$  is again the window size. For each time step, the normal distribution  $\mathcal{N}(\mu, \sigma^2)$ , parametrized by  $\mu = \bar{d}[k]$  and  $\sigma^2 = (\sigma[k])^2$ , is integrated from  $-1$  to  $1$ , i.e.

$$p[k] = \int_{-1}^1 \mathcal{N}(\bar{d}[k], (\sigma[k])^2) dd. \quad (3.17)$$

It is the probability of an RSSI fluctuation value falling into the interval  $[-1, 1]$ . From sample measurements, the authors derived an expected value for this probability with human movement, and a value without, leading to a threshold  $p_{\text{thr}}$ . If  $p[k] < p_{\text{thr}}|_{k=k_0}$ , then an event is generated at  $k_0$ .  $p_{\text{thr}} = 0.3$  and  $w = 10$  were given as optimal values by the authors. They claim that their method produces less false positives than Lee's algorithm.

## 3.2 Proposed Detection System

In this chapter, a new system for the detection of human presence using the BLE radio communication standard is described. The general system structure is discussed in section 3.2.1. A study of the system's parameters is presented in section 3.2.2. Section 3.2.3 introduces the setup that is used for an evaluation of the system. A new presence detection algorithm is presented in section 3.2.4. The results of a performance evaluation are given in section 3.2.5.

### 3.2.1 Radio Node Network

The DFP detection methods introduced in section 3.1.2 either used equipment already installed in the home (WLAN), or additional radio nodes, usually based on WPAN/ZigBee. Both were developed as communication standards for high data rate (WLAN), or high flexibility (ZigBee). Designing low energy components of small size was not a high priority for their design. This means, that employing them in an AAL context is an inconvenient compromise. Here, low cost and low installation effort are essential for a good user acceptance.

In this thesis, BLE is proposed as an alternative solution. BLE was introduced as an addition to the classic Bluetooth standard in version 4.0, with special emphasis on low energy consumption devices [130]. It has found its way into the consumer market, and it is included in more and more smartphones. In contrast to ZigBee, BLE equipment is likely to be available at extremely low prices and a large variety of components. The proposed solution uses miniaturized standard BLE components, that are powered by CR 2032 type coin cells<sup>15)</sup>. BLE uses a frequency of 2.45 GHz. It passes through many construction and furniture materials. This makes it easy to integrate the components into homes. The main information carrier is the RSSI, that is easily available through commercial BLE integrated circuits. In its simplest form, the system consists of two BLE components, one transmitter and one receiver. The operation characteristic is comparable to conventional light barriers. A typical application for this is the monitoring of linear movements, e.g. through corridors or along staircases. In more complex scenarios, setups with one transmitter and multiple receivers can be installed to simultaneously monitor multiple areas with minimum energy consumption.

<sup>15)</sup>These coin cells are 20 mm in diameter and 3.2 mm thick, which defines the components' minimal size.

### 3.2.1.1 Bluetooth Low Energy

A short introduction to BLE is given in this subsection. A full description can be found in the specification documents for version 4.0 [130] and version 4.1 [131]. BLE is a wireless, packet-based communication protocol. It defines several modes of operation for different purposes. One of these modes is called *advertising*. All transmitting nodes (TX) emit data packets in a predefined interval, the so called *advertising interval*. Its inverse is the *advertising rate*. The advertising packets contain no address, i.e. they can be received by any BLE receiver (RX) that is in range. To avoid collisions with other 2.4 GHz communication, the advertising packets are transmitted sequentially on three channels, each with a different center frequency. An example timing sequence is drawn in Fig. 3.2. The BLE center frequencies can be found on its vertical axis. The slight differences in center frequency lead to different propagation characteristics. This means that there are three information carriers, that can be exploited without implementing proprietary protocols.

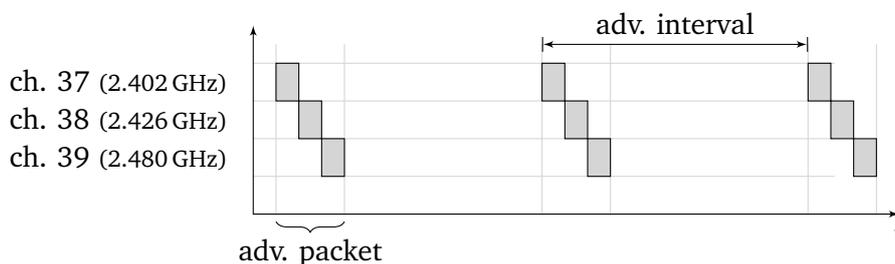


Fig. 3.2: BLE advertising timing pattern. Frequencies from [131, p. 2502].

Since BLE advertising has no addressing, there is a variety of possible topologies. Simple configurations of one transmitter and one receiver are possible, as well as multiple component setups. The only requirement is, that they stay within range of each other. As was shown in section 3.1.1.1, the human body absorbs electromagnetic radiation at 2.4 GHz. The mounting configuration should therefore be such that the receiver is able to catch all advertising packets, even in the most disturbed case (i.e. the disturbed RSSI values should not fall below the receiver's threshold). If the distance is too large, a detection algorithm must be able to use packet loss as an activity indicator.

In the following, a single point-to-point connection is assumed<sup>16)</sup>. The step towards multi-component networks is mostly a matter of communication firmware, which is not a topic of this thesis.

## 3.2.2 System Parameters

The system parameters were evaluated with the help of Simon Gleissner, who wrote his Bachelor's thesis at Furtwangen University [10]. The results are concluded in this section.

### 3.2.2.1 Advertising Channel Differences

With three advertising channels and a fixed receiver, there are six setup modes in total, three for the setup where the transmitting antenna is facing the receiver, and three for the situation where the transmitting antenna is facing the mounting post or wall. In order to find the optimal setup, disturbed measurements of all six variants are evaluated in this subsection. The effect of the different conditions is shown by the histograms in Fig. 3.3. Each subplot corresponds to one measurement of one channel and one TX-RX configuration. The differences between the advertising channels are clearly visible. The most dominant RSSI value, the largest bin of the

<sup>16)</sup>Technically, BLE advertising is not a real connection because communication is only possible in TX → RX direction. However, for the sake of simplicity, a single TX-single RX configuration is called point-to-point connection in this thesis.

histogram, is different for each channel. There are two reasons, that have to be considered. Firstly, the antenna frequency response is not flat for all frequencies. This is a component specific limitation that has to be accepted as it is. Secondly, the environment responds differently to different frequencies. This can be exploited for inferring more information from the communication, or for suppression of disturbances other than those caused by humans. An evaluation of this effect is not in the scope of this thesis.

In Fig. 3.3c) and 3.3d), it can be seen that channel 38 tends to form multiple local maxima. This means, that its undisturbed RSSI level fades more over time than the other two channels. Since this could lead to erroneous detections it might be advisable to avoid using this channel.

Mounting the TX antenna towards the receiver seems to lower the link attenuation by 3 to 5 dB, compared to the TX antenna facing the mounting post. Therefore, the TX antenna-to-RX configuration is preferable in order to maximize the usable distance between the two.

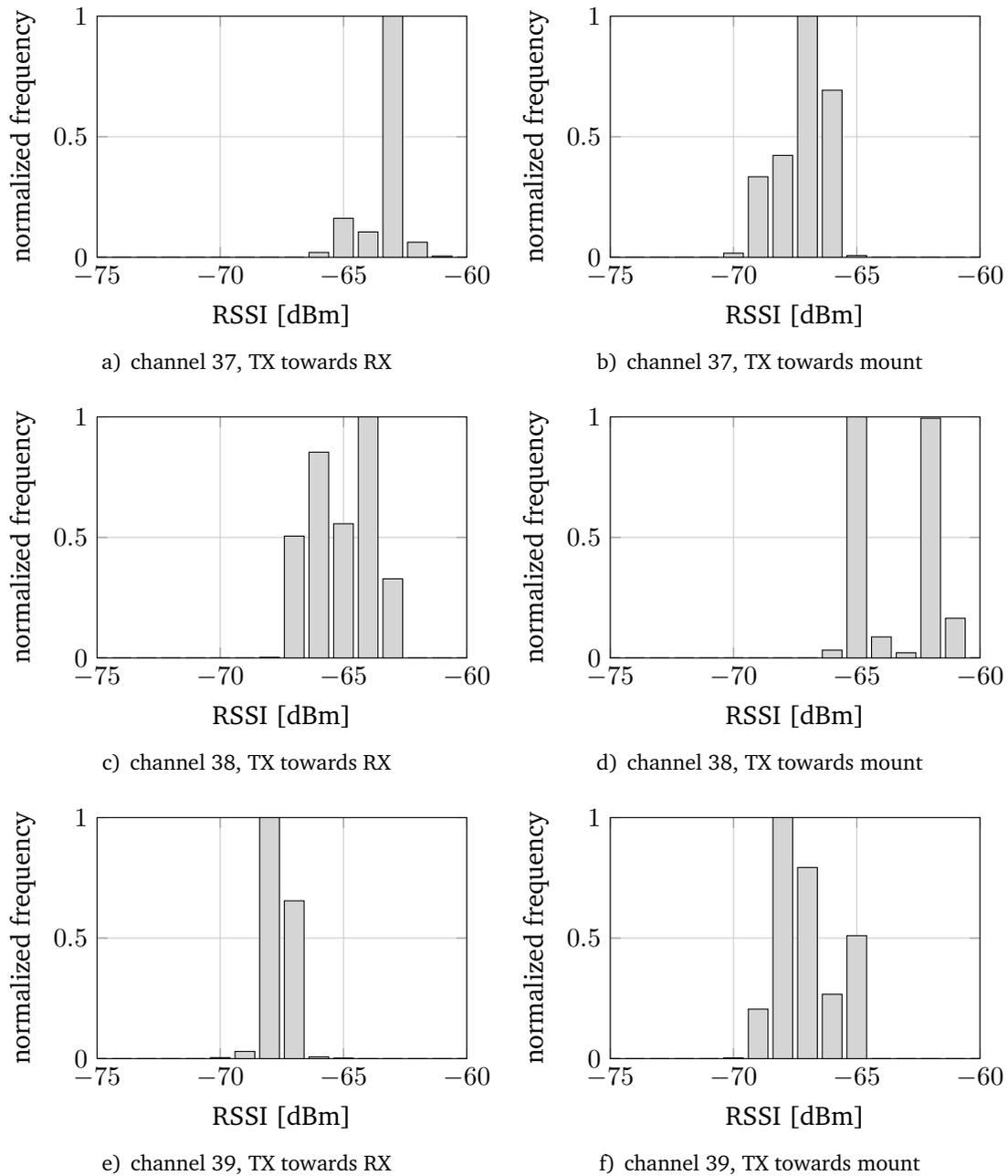


Fig. 3.3: histograms of disturbed RSSI measurements for different channels and mounting configurations

### 3.2.2.2 Maximum Distance Between Sender and Receiver

With the components used, the maximum distance between transmitter and receiver under free field conditions is  $D_{\max} = 20$  m. This is sufficient to cover normal corridors, small rooms, and specific detection areas like a bed or a door frame. If a distance greater than  $D_{\max}$  is chosen, advertising packets are frequently lost and the overall performance of the system will suffer because of the lower and unpredictable resulting packet rate.

### 3.2.2.3 Detection Area Geometry

In section 3.1.1, a simplified spatial model of the propagation of electromagnetic waves in the area between a transmitter and a receiver was given. The relevance of this model is evaluated in this section using a point-to-point setup, as sketched in Fig. 3.4. In order to visualize the shape of the detection area, a person approached the transmission area from a far distance in  $y$  direction. Repeating this procedure for several values of  $d_1$  and  $D$ , the horizontal spatial geometry of the detection zone was estimated. Assuming isotropic antennas, the results should be valid for the vertical direction as well. For three experiments with  $D = 1.5$  m,  $D = 2$  m, and  $D = 3$  m, the measured values are shown in Fig. 3.5. A point of the solid “measured” lines was made once the RSSI dropped by 3 dB. The plot of the second Fresnel zone, marked by the dashed line, reveals a good match of the measured data, at least in the center region of the Fresnel zone. This means that the zone monitored by a single BLE link is almost an ellipsoid, and its maximum radius can be derived from the transmitter-receiver distance  $D$  via eq. (3.3) for  $n = 2$ .

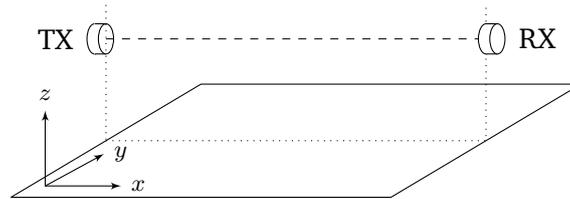


Fig. 3.4: setup of the detection system for a single monitored area. The dashed line marks the line of sight.

### 3.2.2.4 Packet Transmission Rate

The packet transmission rate, or, in the case of BLE the advertising rate, is a crucial parameter for successful presence detection. A higher packet transmission rate means higher probability of a human disturbing a packet which, in turn, leads to RSSI changes. On the other hand, a high packet rate requires high energy consumption, which is a limiting factor for battery powered devices. In order to estimate the required packet rate, a person moving with constant normal walking speed ( $v = 1.39$  m/s = 5 km/h) is considered. The person passes the detection zone on a path perpendicular to the line of sight. Let the radius of the detection zone along the person’s path be  $r$ . The maximum packet interval results from the fact that at least one packet must be emitted while the person is located inside the detection zone. Hence, the maximum possible interval is given by

$$\Delta t_{\max} = \frac{2 \cdot r}{v}. \quad (3.18)$$

The packet rate  $f_p$  is limited by  $f_p \geq \frac{1}{\Delta t_{\max}}$ . A packet rate of 10 Hz ( $\Delta t = 100$  ms advertising interval) was chosen for the system proposed here. According to eq. (3.18), this results in a detection zone radius of approximately 7 cm at walking speed. Evaluating Fig. 3.1b), this corresponds to a transmitter-receiver distance of  $> 0.5$  m, so the system is well suited for corridors, door frames and open rooms.

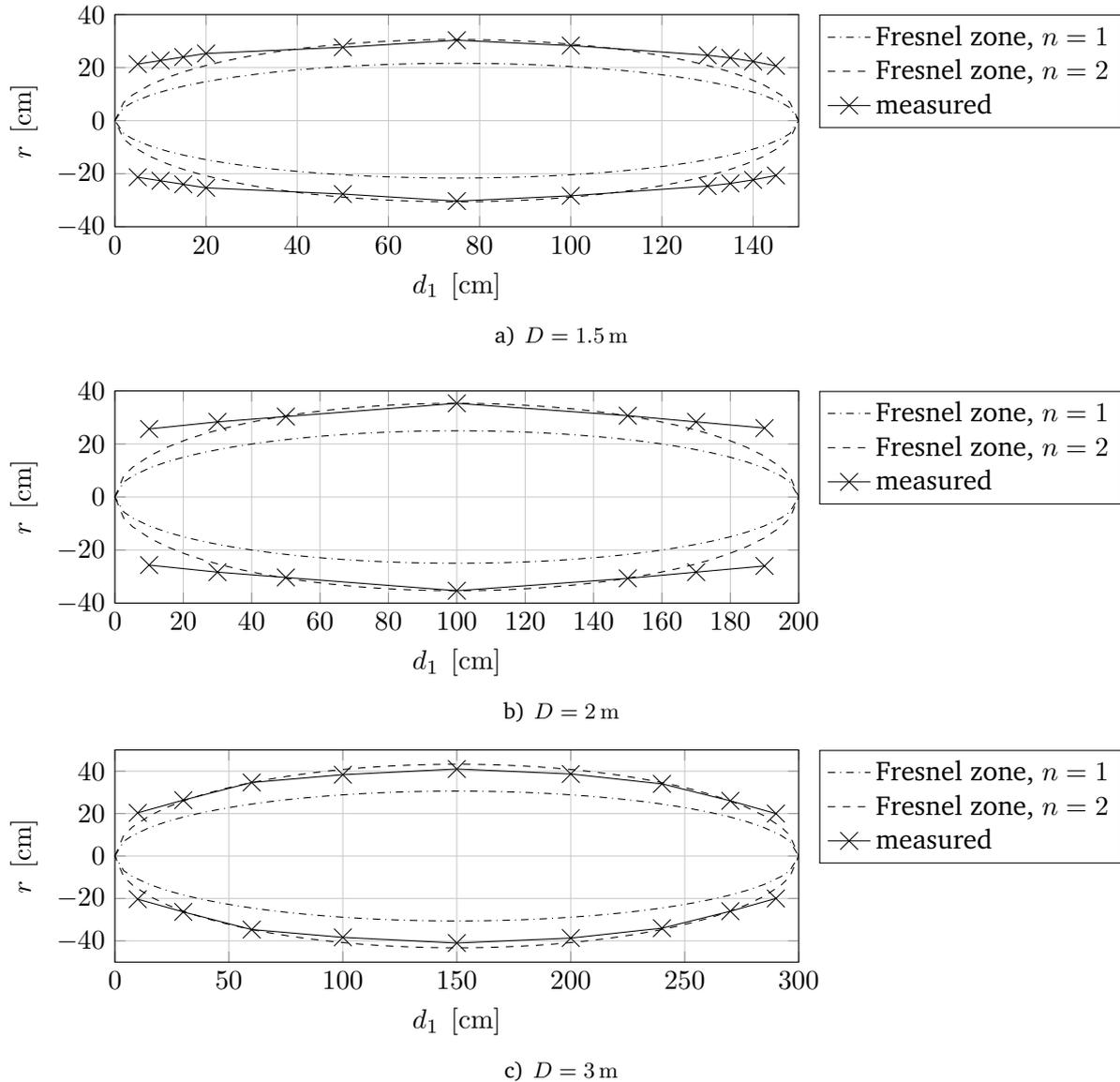


Fig. 3.5: comparison of measured RSSI values and the Fresnel zones in order to verify the geometry of the detection area

The packet intervals used in related studies can be roughly grouped into three groups, as shown in Table 3.3. The list shows that the medium range  $100 \text{ ms} \leq \Delta t < 1 \text{ s}$  is the most common choice.

### 3.2.3 Performance Assessment Setup

The results presented in state-of-the-art papers often show singular snapshots of a detection phase [43, 44, 134]. They do not convey an evaluation of how well the methods perform over time. This information is essential, because in an AAL scenario, the detection system is constantly monitoring people. For AAL it is especially challenging to perform a field test. In order to record referenced data, that gives some ground truth information on the residents' activities, a lot of sensing equipment has to be installed. To simplify this procedure, the use of a conventional light barrier<sup>17)</sup>, which directly evaluates what should be detected by the DFP system, is evaluated in this thesis. It detects movement by an IR light beam that is interrupted by non-transparent objects. The system used here includes an integrated process and display unit,

<sup>17)</sup><http://www.elv.de/ir-lichtschanke-mit-alarmgeber.html>

Table 3.3: packet transmission intervals used in related DFP systems

interval range	references
$\Delta t < 100$ ms	[26, 28, 32, 132, 34, 76, 41]
$100$ ms $\leq \Delta t < 1$ s	[21, 31, 35, 69, 47, 53, 133, 48, 37, 70, 65, 44, 45, 50, 63, 60, 74, 57]
$\Delta t \geq 1$ s	[30, 29, 33, 59, 61, 39, 64, 52]

that counts the interruptions. The overall number of events is shown on a display. Measurements were recorded in two scenarios:

- 1. Laboratory (lab):** both activity detection systems were mounted in the frame of a laboratory door, as sketched in Fig. 3.6a). During 48 hours a total of 277 events were recorded by the reference light barrier.
- 2. Home:** to confirm the operation in a more realistic AAL scenario, the setup was mounted in the main entry corridor of a three-person home, as depicted in Fig. 3.6b). In order to get intermediate information from the light barrier, the 43 h measurement period is split into six separate measurements.

In both scenarios, the radio components and the IR light barrier were mounted such that their detection zones overlap as much as possible.

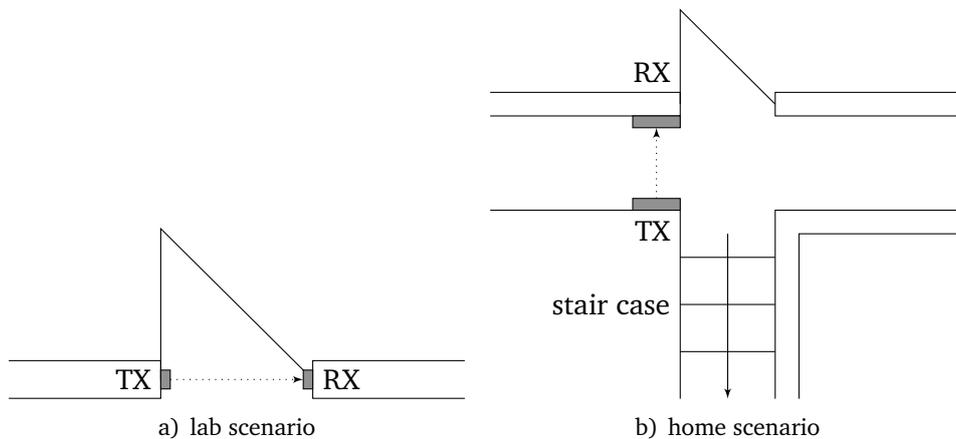


Fig. 3.6: experiment setups for the performance evaluation. TX = transmitter, RX = receiver. The dotted line marks the direct line of sight.

### 3.2.3.1 Performance Measure

The use of the conventional light barrier which only shows the total number of events requires a special quality indicator. Traditional values like detection, miss, or false positive rate cannot be computed from the data. Instead, the error of absolute numbers of events is used. Let  $N_d$  be the number of events detected by a DFP algorithm, and  $N_{gt}$  be the number of events indicated by the light barrier. The deviation or error is

$$e = \frac{N_d - N_{gt}}{N_{gt}}. \quad (3.19)$$

It may be both positive and negative. A value  $e > 0$  means that the number of detected events is greater than the reference, and vice versa for errors  $e < 0$ .

### 3.2.4 Proposed Presence Detection Algorithm

The methods of Youssef et al. are low complexity event detection algorithms, which makes them the predestined choice for implementation on low power sensor nodes. However, they require training in each application scenario in order to derive values for the threshold parameters. In this thesis, a method that simplifies the training procedure by deriving the threshold continuously from the RSSI signal is proposed. The base signal is the absolute derivative  $|d[k]|$ , that was introduced in eq. (3.9) on page 24. Again, average

$$\mu_d[k] = \frac{1}{w} \cdot \sum_{i=k}^{k+w-1} |d[i]| \quad (3.20)$$

and standard deviation

$$\sigma_d[k] = \sqrt{\frac{1}{w-1} \cdot \sum_{i=k}^{k+w-1} (|d[i]| - \mu_d[i])^2} \quad (3.21)$$

are computed from sliding windows of length  $w$ . The condition for an event at time  $k_0$  is

$$|d[k]| > \mu_d[k] + \sigma_d[k] + d_{\text{offs}}|_{k=k_0}. \quad (3.22)$$

The difference to the ‘‘Youssef II’’ method is that equations (3.20) and (3.21) are computed at each time step during the detection phase, prior to evaluating (3.22). By doing so, the requirement for a dedicated training phase is avoided. The parameter  $d_{\text{offs}}$  is intended for noise cancellation and depends on the DFP system components, not on the operating environment.

To show the dependency on the parameters, a full parameter search was done using the lab and home measurements. The resulting average error is presented in Fig. 3.7. It can be seen, that at  $d_{\text{offs}} \approx 2.2$  the error becomes zero for almost all window sizes. The BLE receivers used here provide integer values of the RSSI in dBm. Hence,  $d_{\text{offs}} \approx 2.2$  corresponds to a typical noise amplitude of  $\pm 1$  dBm. The window size does not have much influence, so the a value of  $w = 10$  can be chosen for this method.

During algorithm development, it became clear that the raw detected events need to be filtered because a human movement frequently triggered a series of consecutive events. Therefore, consecutive events are eliminated prior to any performance evaluation. The implementation is a straightforward loop structure, that eliminates events that immediately follow each other (i.e. no advertising packet was received between two events that did not fulfill condition (3.22)).

### 3.2.5 Results & Discussion

Results regarding the evaluation of RSSI signals are presented in this section. First, an excerpt from a home scenario measurement is analyzed in subsection 3.2.5.1. Next, in subsection 3.2.5.2, the detection algorithms are compared using the error measure that was introduced in eq. (3.19). Finally, an evaluation and comparison of the computation effort is given in subsection 3.2.5.3.

#### 3.2.5.1 Observations on Signal Level

From the experiments, two fundamental observations regarding the RSSI values of a BLE communication link were made. Humans passing by, or moving inside the detection zone leave negative spikes of height  $\geq 4$  dB. This is a good indication of an activity that involves movement of the human body. On the other hand, the steady presence of humans imposes more long-term fluctuations on the signal when they stay in the detection field for calm activities. An example can be seen in Fig. 3.8. The thick black line marks the sliding average of the RSSI values. At

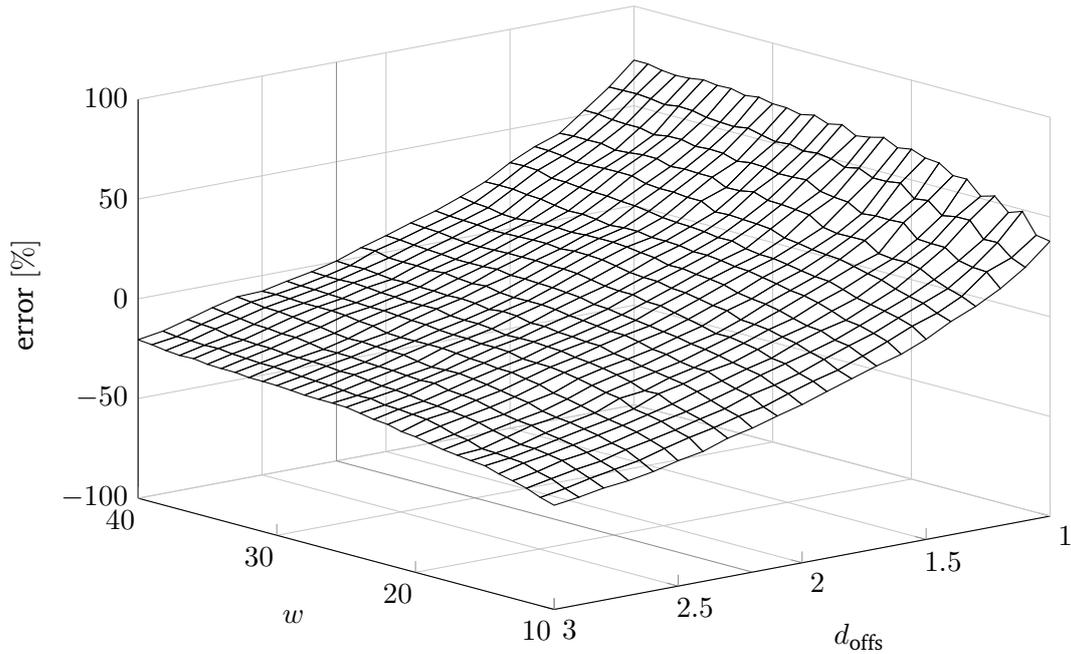


Fig. 3.7: results of the parameter optimization for the proposed presence detection algorithm

$t \approx 1$  min, a person moved into the vicinity of the transmitter, which can be seen by a downward slope. The spikes marked by black crosses are transient events resulting from other people crossing the line of sight.

This insight leads to a redefinition of the goal for the DFP algorithms. For Activity Monitoring, the activity in terms of human movement is most relevant. A person standing steadily in the detection area does not provide much information for activity analysis. Therefore, an algorithm should be able to detect spikes caused by the movement of human tissue inside a radiation zone. It should not be disturbed by drifts that result from steady presence or altered environmental conditions. This confirms the necessity for an algorithm that derives its threshold from the current signal, rather than from training. It also means that the detection of activity is possible even if another person is located inside the detection zone.

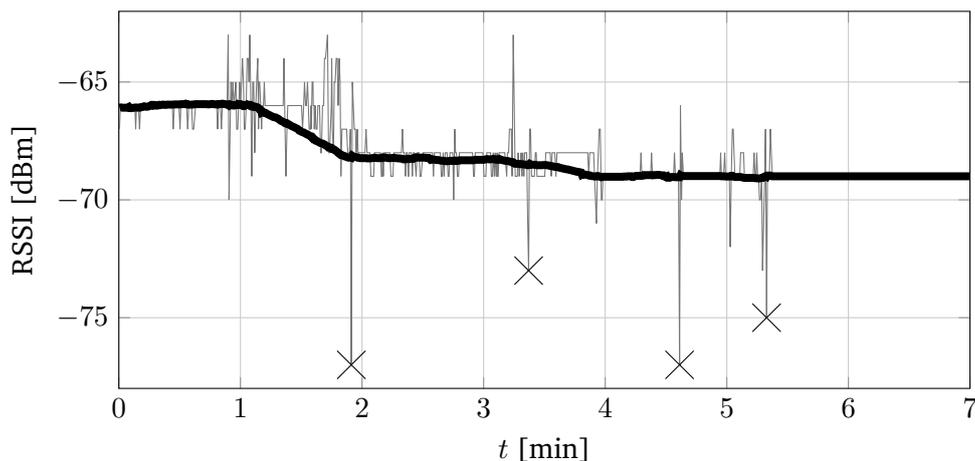


Fig. 3.8: extract from an RSSI measurement in the home scenario

### 3.2.5.2 Quantitative Evaluation

For a quantitative result, the error in eq. (3.19) is evaluated for all measurements described in section 3.2.3. The average error for both scenarios is shown in Fig. 3.10a). The method

“Lin” misses a considerable amount of events so their parametrization makes them insensitive to the RSSI spikes in the measurements. The “Youssef II” algorithm is too sensitive because its absolute error is very high and positive. “Youssef I” and “Lee” perform better but they also miss some events. The proposed method seems to find the best compromise between sensitivity and robustness.

The reason for the high positive error of the “Youssef II” method can be seen in Fig. 3.9a). The method does not have any means to cancel noise in the signal. This means the threshold sometimes drops down to zero. Any noise then immediately triggers an event. An additional source of error is that the method does not compute the absolute value of the RSSI derivative. Hence, it misses the event at  $\approx 21$  s. The proposed method detects this. It is also more immune to noise because of its adaptive threshold and its noise suppression factor  $d_{\text{offs}}$ . The threshold in Fig. 3.9b) is well above the detection signal.

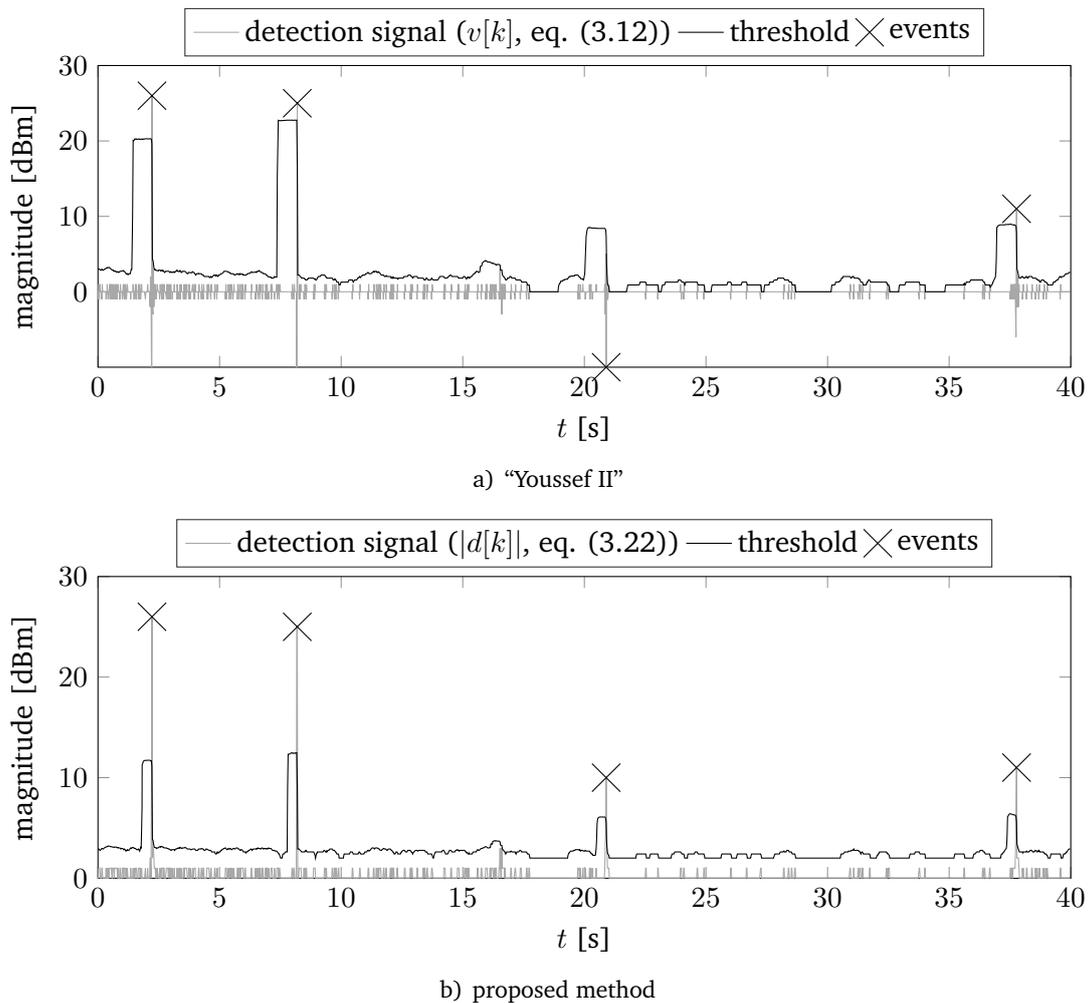


Fig. 3.9: example of detection signal (gray) and threshold (black) from a home measurement. Ground truth events are marked by black crosses.

Dividing the errors into lab and home, as shown in Fig. 3.10b), reveals that the methods perform differently in the two scenarios. A visual inspection of signals from both scenarios showed no obvious difference of the detection results. To the contrary, the IR light barrier showed a significantly different detection characteristic than the radio network during the measurement phase. It is much more sensitive due to its very narrow detection zone. A person, whose silhouette has gaps e.g. by carrying objects, or spreading the arms away from the body, is likely to trigger multiple events when crossing the IR beam. The DFP detection zone has a much wider cross section, which usually covers the entire silhouette of a human. From the measuring expe-

rience with the proposed detection system, it can be stated that the amount of errors due to the different detection characteristics of the light barrier is in the range of the error of the proposed method.

Another possible source of error is the high selectivity the reference light barrier. It does not detect the presence of humans close to it, but not in the line of sight. In the radio signal localization case, a person has an impact on the electromagnetic field even if it is not directly blocking the line of sight, which may include standing behind one of the radio components. In these cases, the radio signal localization method will find more events than the light barrier. Since these events provide more useful information than the light barrier, this is considered another major benefit of using radio signals.

These two reasons are fundamental limitations of the verification using a system with a different detection characteristic. If a more reliable evaluation was required, a reference system that is able to detect events under the same conditions as the tested system should be used. In the AAL context, a single event is usually not used as a singular piece of information. When the event data is used for profile analysis or fusion with other data sources, the radio signal detection system should be parametrized such that it detects slightly too many events, rather than too few.

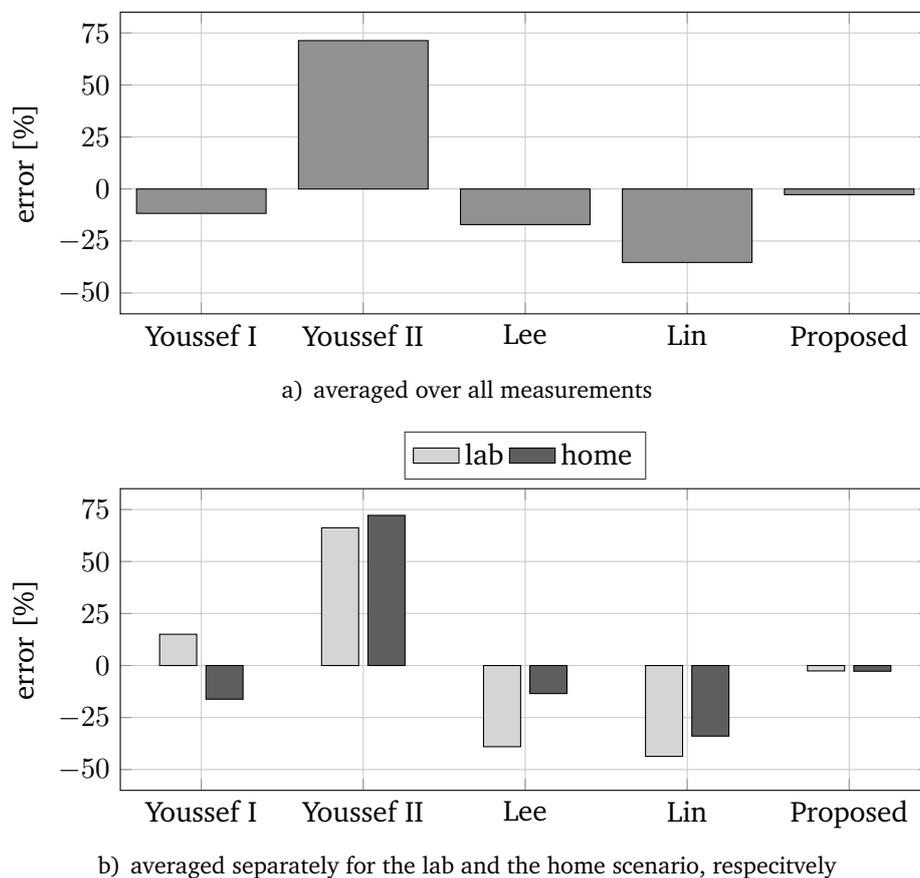


Fig. 3.10: detection errors of various DFP presence detection algorithms

### 3.2.5.3 Comparison of the Computation Effort

In order to estimate the computation effort, the execution time of the implemented algorithms is measured in Matlab while processing both the lab and the home data set. The time needed by the described methods to process all measurements is shown in Fig. 3.11. The graph is normalized to the time the “Youssef I” algorithm needed. This algorithm turned out to be the slowest one. “Youssef II” and the proposed method are almost evenly fast, and faster than the

others because they do not need complex computations such as histograms or probability density estimations.

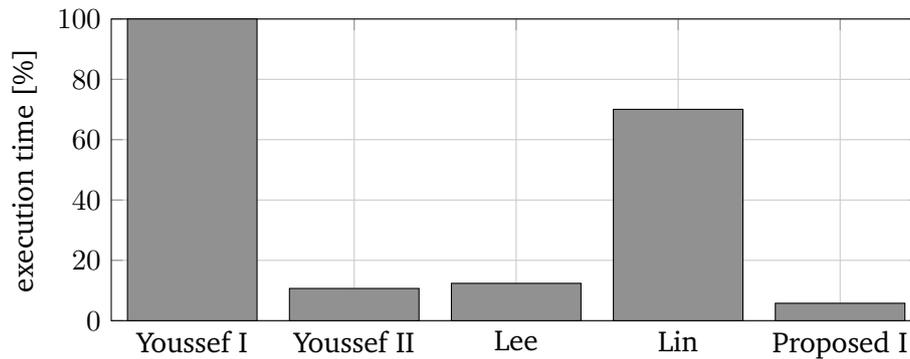


Fig. 3.11: comparison of the execution time of various presence detection methods

### 3.3 Conclusion

In this part of the thesis, DFP human presence detection was introduced as a solution of the problem that activity monitoring requires intrusive and expensive sensor equipment. DFP detection utilizes electromagnetic waves to passively localize human beings within detection zones. Common suitable radio technology is expensive or hard to integrate into the home. As a solution, a BLE based system approach was presented. It consists of standard coin-cell sized and powered radio nodes, that can be easily mounted and hidden from view. BLE is much more suitable for consumer market applications, because its components are cheaper, smaller, and more energy efficient than existing equipment based on WLAN or WPAN/ZigBee. To the best of the author's knowledge, this is the first time that BLE was used for DFP detection.

State-of-the-art detection algorithms were described. An enhanced algorithm tailored to the specific needs of AAL was proposed. The performance of the reference algorithms and the proposed method was compared based on signal analysis and a quantitative evaluation under laboratory and private home conditions. The proposed method outperforms the reference methods in both scenarios. A conventional IR light barrier was employed as reference data source in the test scenarios. It is easier to mount and evaluate than e.g. cameras. However, it has a serious limitation, in that it is more sensitive than DFP solutions. This makes it harder to evaluate the test signals correctly. Regarding the collection of reference data, there is room for optimization for future investigations.

# Chapter 4 | Non-Intrusive Load Monitoring for AAL Applications

## 4.1 Introduction

The goal of NILM is the identification of electric appliances in the total power consumption of a home. Traditionally, the question to answer is, how much energy each of the appliances has consumed over an accounting period, without needing to modify the in-house electric installation by mounting measuring equipment at each of the appliances. The fundamental idea of NILM is that each appliance leaves a characteristic fingerprint in its power consumption, which identifies it. In simplified terms, the load identification is the extraction, analysis, and classification of a set of features from the signals measured at the smart meter. As described in chapter 1, this information is further exploited for activity analysis and detection of hazardous situations in this thesis. Since the consumption of individual appliances is not relevant for these fields, all topics regarding energy computation will not be covered. The following abilities have to be analyzed:

1. Detection of switching events with a power difference greater than 15 W for Activity Monitoring,
2. Classification of appliances consuming 15 W or more for the separation of automatic and manually operated loads, and
3. Classification of appliances consuming more than 200 W for the detection of dangerous situations caused by heating components.

The lower threshold of 15 W is chosen because relevant appliances are considered to have a power consumption greater than this value, and because this allows for a margin to the resolution of the ReSP measurement equipment of approximately 5 W.

Applying NILM to AAL brings special requirements, that are not as relevant in its traditional field of application, the analysis of electric energy. The most critical factor is reliability of the information, that is passed to residents or caregivers. False alarms are to be avoided as much as possible. On the other hand, missed action on actual incidents worth reporting quickly diminishes the credibility of the overall system.

As with DFP systems, hardware cost and installation effort are to be kept as low as possible. Especially in the sector of private or informal health care, price is a decisive argument (compare the introduction of chapter 3.1). The ultimate goal behind the NILM idea is to integrate the algorithms into smart meters. Once this is achieved, the NILM technology can be rolled out to a large group of people because of the compulsory installation of smart meters in many countries [135, 136].

Another special framework condition applies in the AAL application area. Manual setup of appliance features during the installation is acceptable, whereas generic NILM systems usually aim at automatic setup, which does not require interaction. However, the algorithm design should require as little training effort as possible to allow for an economic installation.

The two major components of a NILM system, event detection and feature extraction and classification, are analyzed separately in this thesis. Each analysis shall highlight the prospects

of successful application in AAL. The separation of the analyses is done to avoid any influence of one component on the other. Thus, should the performance of one stage prove to be insufficient, the other stage can be evaluated independently.

Three event detectors are described and compared in chapter 5. Two feature extraction and classification algorithms, one using steady state power features, the other using time-series fingerprints (TSFPs), are compared based on classification rates in chapter 6. The remainder of this chapter is dedicated to an appliance model in section 4.2, a detailed description of signals for the event detection and feature extraction, and their estimation from smart meter data (section 4.5). The chapter is concluded by a presentation of public (section 4.7) and ReSP data sets (section 4.8), that will be used for the evaluations.

## 4.2 Appliance Model

The multitude of different appliances requires a universally valid model of the operation pattern. The appliances considered in this thesis are characterized as follows. The power consumption is distinctly limited by two events. The first event is called *turn-on event*, the second one *turn-off event*. The time between these events is called *operation cycle*. The operation cycle is divided into four sections:

1. The cycle commences when the appliance is put into operation by a switch (manual appliance), or by its internal electronics or electricians (automatic appliance). This is characterized by a steep increase of power, and sometimes a short transient spike in the power signal.
2. In the second section, the internal components of the appliance settle to the steady state operating conditions. The power signal is expected to fluctuate due to the settling processes in this section.
3. The appliance then reaches its steady state operation phase, which is the third section. This is the phase in which the appliance performs its designated operation. The power is assumed to be constant in this section.
4. At some point in time, the appliance is turned off again, either by the user, or by internal control logic. This is characterized by a rapid power decrease. Decay or swing off processes are not to be expected in this fourth section because the appliance is usually disconnected from the mains. This means that the smart meter is unable to record these processes.

A graphic example of the power consumption of an appliance fulfilling these requirements is given in Fig. 4.1.

The current  $i(t)$ , that is measured by the smart meter, is modeled as

$$i(t) = s(t) \cdot c(t) \cdot v(t), \quad (4.1)$$

where  $v(t)$  is the sinusoid that represents the grid voltage, and  $c(t)$  contains components for the transformation from voltage to current (i.e. the appliance's conductance), as well as a model for the current dynamics due to time-dependent operating conditions. Its contents are non-periodic. The signal  $s(t) \in \{0, 1\}$  is a piece-wise constant function, where 0 denotes that the appliance is in the "off" state, and 1 indicates the "on" state.

### 4.2.1 Compound Appliances

Obviously, not all possible appliances can be modeled by the previous definition. This applies especially to appliances where multiple, high-power components are integrated into one device (e.g. the heaters and motors in dish washers, washing machines, clothes driers, etc.). If each component is considered an independent appliance, those types of compound appliances can

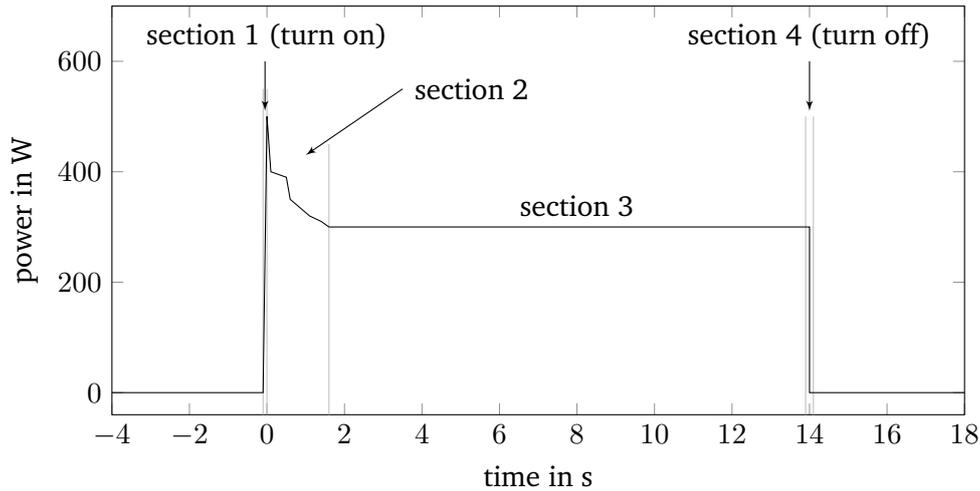


Fig. 4.1: power consumption of an appliance in accordance with the standard model with four sections

be modeled under the assumptions of section 4.2. A subsequent logic then has to match the components with higher level appliance labels after their individual disaggregation.

More complex appliances, where individual components are not distinguishable (e.g. personal computers (PCs), VSDs, electronically controlled machines), require even more sophisticated models, and are not regarded in this thesis. In this thesis, compound appliances are assumed to be separable into dominant components. Accordingly, these are treated as individual appliances. Obviously, this needs to be accounted for by reference data accompanying a measurement.

### 4.3 Choice of Features

As described in section 2.3, there are many different approaches and features for solving the disaggregation problem. The choice for this thesis is made according to the general nature regarding the robustness of the features. Transient features, that are extracted from very short periods of a signal, are very selective and sensitive to the structure of the power grid, and to the time and place of operation. Therefore, they are not considered in this thesis. SSPFs, on the other hand, can be computed from much longer time periods. It is therefore self-evident to assume that they are more robust in the face of fluctuations or superpositions, even if this might require some preprocessing of the signals. A new type of features are so called “time-series fingerprints”, where characteristic signal sections are extracted and used for classification. This approach is presented in section 6.5.

### 4.4 Machine Learning

In this section, a short introduction to important terms and algorithms of machine learning are given. The interested reader is referred to specialized machine learning literature, e.g. the book by Bishop [137], for more details.

#### 4.4.1 Data Representation and Notation

Let a feature be represented by an  $N \in \mathbb{N}$ -dimensional vector  $\vec{x} \in \mathbb{R}^N$ . This type is called *real-valued*. In general, complex-valued features  $\vec{x} \in \mathbb{C}^N$  are also possible, but this case is not taken into account in this section. The set

$$X = \{\vec{x}_1, \dots, \vec{x}_{N_t}\} \quad (4.2)$$

is called the *test set*. It contains unknown features, that need to be classified. Nevertheless, a vector

$$\vec{c}^{(gt)} = \left[ c_1^{(gt)} \quad \dots \quad c_{N_t}^{(gt)} \right]^T \quad (4.3)$$

of appliance IDs for each test feature is required for verification of the classification result. Furthermore, let there be a *training set*

$$Y = \{ \vec{y}_1, \dots, \vec{y}_{N_{tr}} \} \quad (4.4)$$

of  $N_{tr} \in \mathbb{N}$  training features  $\vec{y}_j \in \mathbb{R}^N$ ,  $j = 1, \dots, N_{tr}$ . Depending on the machine learning method, a label vector

$$\vec{c} = [c_1 \quad \dots \quad c_{N_{tr}}]^T \quad (4.5)$$

accompanying the training data may be required. This vector assigns a class label  $c_i$  to each training vector  $\vec{y}_i$ ,  $i = 1, \dots, N_{tr}$ . The  $c_i$  are typically integers, but any set of unambiguous symbols can be used, depending on the actual implementation. The set

$$U = [u_1 \quad \dots \quad u_{N_u}]^T \quad (4.6)$$

contains  $N_u$  unique class labels. Obviously,  $c_i^{(gt)} \in U$ ,  $i = 1, \dots, N_t$ , and  $c_j \in U$ ,  $j = 1, \dots, N_{tr}$ .

#### 4.4.2 Dissimilarity Measures

If a function  $d : (\mathbb{R}^N, \mathbb{R}^N) \rightarrow \mathbb{R}$  fulfills the following conditions,

$$d(\vec{x}, \vec{y}) = d(\vec{y}, \vec{x}), \quad (4.7)$$

$$d(\vec{x}, \vec{y}) = 0 \Leftrightarrow \vec{x} = \vec{y}, \quad (4.8)$$

$$d(\vec{x}, \vec{z}) \leq d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z}), \quad (4.9)$$

it is called *dissimilarity* or *distance measure*. These conditions are fulfilled by the class of norms  $\|\cdot\|$ , which again fulfill the following requirements

$$\|\vec{x}\| = 0 \Leftrightarrow \vec{x} = [0 \quad \dots \quad 0]^T, \quad (4.10)$$

$$\|a \cdot \vec{x}\| = |a| \cdot \|\vec{x}\| \quad \forall a \in \mathbb{R}, \vec{x} \in \mathbb{R}^N, \quad (4.11)$$

$$\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\| \quad \forall \vec{x}, \vec{y} \in \mathbb{R}^N. \quad (4.12)$$

A frequently used class of norms are the *inner product norms*, which are defined by

$$\|\vec{x}\|_{\mathbf{A}} = \sqrt{\vec{x} \mathbf{A} \vec{x}^T}, \quad (4.13)$$

where  $\mathbf{A} \in \mathbb{R}^{N \times N}$  is the *norm inducing matrix*. The most important special case of this matrix is

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}. \quad (4.14)$$

With eq. (4.14), eq. (4.13) defines the *Euclidean norm*<sup>18)</sup>.

<sup>18)</sup>Other well known norms are Lebesgue, Minkowski, and Mahalanobis norms. They are not relevant for the topics discussed in this thesis.

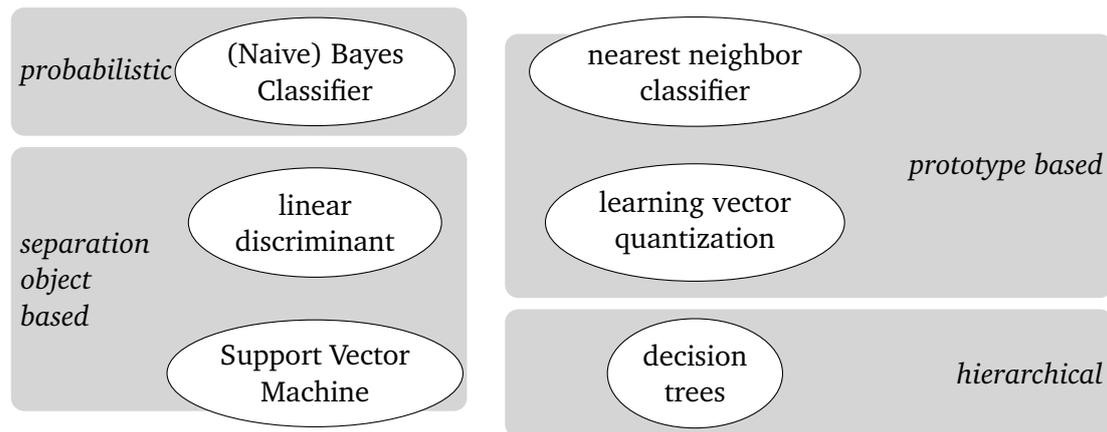


Fig. 4.2: classification methods grouped according to the way they separate the data. Graph adapted from [138, p. 89].

### 4.4.3 Concepts of Machine Learning

The area of machine learning can be divided into the four groups correlation, regression, classification, and clustering, according to the desired application. They are shortly described in the following.

The goal of the *correlation analysis* is to find connections or relations in data sets, and quantify them using statistics. It is related to *regression*. This tries to find functional dependencies in the data. This is used to understand systematic links, and to modify systems in order to achieve a target measure. *Classification* tries to find correlations in data sets, with the goal to assign a class label to an unknown feature. Training a classifier using a labeled training data set is called *supervised learning*. *Clustering* is a mechanism to divide a data set into groups (clusters) without a priori knowledge of the class labels. It may be used prior to classification in order to acquire training data. This is called *unsupervised learning*.

Machine learning is applied in this thesis for two purposes. In chapter 6, it is as a mechanism to analyze the quality of a feature set that is extracted from smart meter data. Sufficiently labeled training data is assumed to be available, so the algorithm analysis is a supervised learning classification task. The employed classification methods are introduced in subsection 4.4.3.1. An activity profile study will be presented in chapter 7.1. Here, unsupervised learning will be required for the separation of unusual profiles. An introduction to the k-means clustering algorithm, that is used in this chapter, is given in subsection 4.4.3.2.

#### 4.4.3.1 Supervised Learning (Classification)

The task of a classifier is to label an unknown feature vector. To do so, it computes statistics from the training data set with the goal to find decision borders in the feature space. The field of classifiers can be divided according to their classification principles. A few examples are given in Fig. 4.2.

**Probabilistic Classification** Probabilistic approaches try to estimate a probability density function for each cluster. A test feature is assigned to the cluster which best explains its location in the feature space. An example is the (Naive) Bayes Classifier, which uses the a posteriori probability of a test feature as an indicator for cluster membership.

**Support Vector Machines** The second important class are linear classifiers. They try to fit linear separation objects (lines in 2D, (hyper-) planes in higher dimensional spaces) to the data. For each separation object, a binary decision can be made for a test feature. It lies either on

one side of the object or the other. A common example are support vector machines (SVMs)<sup>19)</sup>. They use a linear object (line, plane, hyperplane)

$$f(\vec{y}) = \langle \vec{w}, \vec{y} \rangle + b, \quad (4.15)$$

where  $\vec{w}$  is the normal vector of the hyperplane, and  $b$  is a bias parameter, to separate data of two classes, i.e.  $c_i \in \{-1, 1\}$  in eq. (4.5). The decision of the classification is binary, i.e. SVM is a non-probabilistic approach. A solution is found if  $f(\vec{y}_i) > 0$  for all training samples with  $c_i = 1$ , and  $f(\vec{y}_i) < 0$  for  $c_i = -1$ . Important to the SVM is the concept of the *margin*, which is the shortest distance between the training samples and the hyperplane. It is a means to find the best solution if there are several possible options. Samples of  $Y$  that lie on the margin are called *support vectors*. There are several extensions of the simple SVM, like kernel functions for non-linear separating objects, or soft margins for dealing with overlapping clusters. These go beyond the scope of this thesis and are not considered here.

**Nearest Neighbor Classifier** A disadvantage of the two previous approaches is that they need the data to be fittable to their models (distributions, separation objects, etc.). Therefore, another classification principle is *prototype based learning*. As the name suggests, samples are taken from the training data to directly classify a test feature. This is also referred to as *lazy learning*. It requires a lot more memory for training, because all training samples need to be stored, and makes the classification more computationally complex, but it allows for the data to be distributed arbitrarily.

A prominent member of this class of classifiers is the *Nearest Neighbor method*. It uses the dissimilarity of the test feature  $\vec{x}_i$  and the training data as basis for the decision function. The classification rule is

$$j = \arg \min_{n=1, \dots, N_{tr}} \{ \|\vec{x}_i - \vec{y}_n\| \}, \quad (4.16)$$

i.e.  $\vec{x}_i$  is assigned the class label  $\vec{c}[j]$  of its closest neighbor. This makes it clear that the Nearest Neighbor classifier simply “memorizes” the training data. It does not actually generalize the data by extracting statistics. This makes it very sensitive in the face of outliers. An example is illustrated in Fig. 4.3. By evaluating the Euclidean distances, the test sample, marked by the large cross, is assigned to the “black dot” cluster, although the visual impression tells the reader to assign it to the cluster marked by the small crosses.

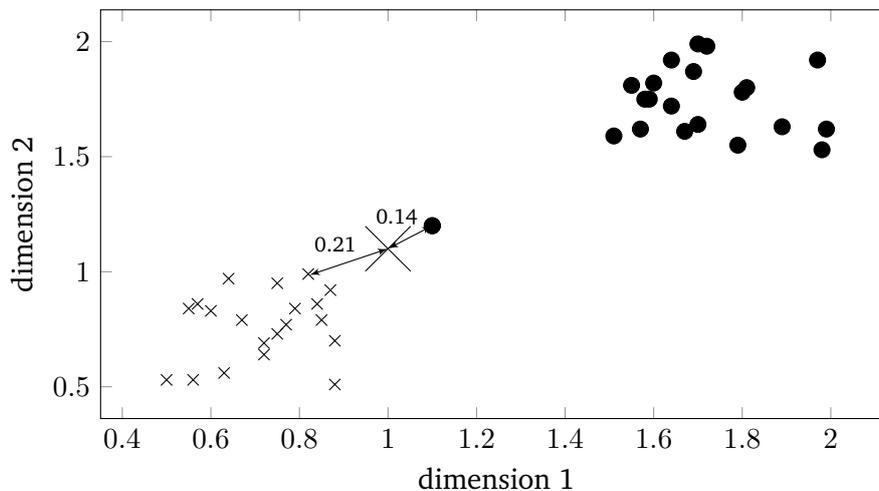


Fig. 4.3: sensitivity of the Nearest Neighbor classifier to outliers. The numbers above the arrows indicate the Euclidean distances.

<sup>19)</sup>This introduction is based on [137, p. 326 ff.].

For this reason, the k-Nearest Neighbor (kNN), an extension of the Nearest Neighbor classifier, takes the distribution<sup>20)</sup> of the training samples around the test feature into account. To classify an unknown test sample, the set

$$Y_k = \{\tilde{y}_1, \dots, \tilde{y}_\kappa\} \subset Y \quad (4.17)$$

of the  $\kappa \in \mathbb{N}$  nearest neighbors around this sample is sought. Let the corresponding class labels be given in  $\vec{c}_k$ . The test sample is assigned to the class which has the maximum number of samples in  $Y_k$ . The classification rule is: assign an unknown vector  $\vec{x}$  to the class  $\vec{c}[j_k]$  with

$$j_k = \arg \max_{j=1, \dots, N_u} \left\{ \sum_{i=1}^{\kappa} \begin{cases} w_i & \forall \vec{c}_k[i] = U[j] \\ 0 & \text{else} \end{cases} \right\} \quad (4.18)$$

and individual weights  $w_i$ .

When applying kNN classifiers to real data, there may be the problem that there are more training samples in  $Y_k$  far away from  $\vec{x}$  than close to it. It is reasonable to assume that training samples far away from  $\vec{x}$  belong to a class that  $\vec{x}$  unlikely belongs to. Therefore, a function of the distance between training and test data

$$w_i = f(d(\vec{x}, \vec{y}_i)). \quad (4.19)$$

can be introduced as for the weights  $w_i$ . The functional dependency is frequently chosen  $w_i = \frac{1}{d(\vec{x}, \vec{y}_i)}$  or  $w_i = \frac{1}{d(\vec{x}, \vec{y}_i)^2}$ .

#### 4.4.3.2 Unsupervised Learning (Clustering)

Unsupervised learning tries to find relations in the (training) data without a priori knowledge of the class labels. Two relevant algorithms are introduced in this subsection.

**k-Means Clustering** In this subsection, an introduction to the k-means clustering algorithm is given based on Bishop's book [137, p. 424 ff.]. The task for a clustering algorithm is to divide the (training) data set  $Y$  into  $\kappa \in \mathbb{N}$  clusters or groups. The data are assumed to be a mixture of Gaussians

$$p(\vec{y}) = \sum_{\ell=1}^{\kappa} \pi_\ell \cdot \mathcal{N}(\vec{y} | \vec{\mu}_\ell, \Sigma_\ell), \quad (4.20)$$

with mixture coefficients  $\pi_\ell$ ,  $\pi_\ell \geq 0 \forall \ell$ , and  $\sum \pi_\ell = 1$ . For the k-means algorithm, each cluster is assumed to be represented by its *centroid*  $\vec{\mu}_\ell$ ,  $\ell = \{1, \dots, \kappa\}$ . The cluster assignment is determined by minimizing the function

$$J = \sum_{n=1}^{N_{tr}} \sum_{\ell=1}^{\kappa} r_{n\ell} \cdot \|\vec{y}_n - \vec{\mu}_\ell\|^2 \quad (4.21)$$

with two unknowns, the assignment variables  $r_{n\ell}$ , and the cluster centroids  $\vec{\mu}_\ell$ . The search for these two parameters is done iteratively using a method equivalent to the Expectation Maximization (EM) algorithm.

<sup>20)</sup>Distribution in the sense of spatial location, not a probability distribution.

**Gaussian Mixture Modeling** This introduction is taken and shortened from [137, p. 430 ff.] and [139]. The Gaussian Mixture Model (GMM) approach is a generalization of the k-means clustering technique. To do this, a binary latent variable  $\vec{z} = [z_1 \dots z_\kappa]^T$  with  $z_\ell \in \{0, 1\}$ ,  $\ell = 1, \dots, \kappa$ , and  $\sum z_\ell = 1$  is introduced. By defining  $p(z_\ell = 1) = \pi_\ell$ , i.e.

$$p(\vec{z}) = \prod_{\ell=1}^{\kappa} \pi_\ell^{z_\ell}, \quad (4.22)$$

the conditional distribution

$$p(\vec{y} | \vec{z}) = \prod_{\ell=1}^{\kappa} \mathcal{N}(\vec{y} | \vec{\mu}_\ell, \Sigma_\ell)^{z_\ell} \quad (4.23)$$

can be derived from eq. (4.20). Marginalizing out the  $z_\ell$  yields

$$p(\vec{y}) = \sum_{\vec{z}} p(\vec{z}) \cdot p(\vec{y} | \vec{z}) = \sum_{\ell=1}^{\kappa} \pi_\ell \cdot \mathcal{N}(\vec{y} | \vec{\mu}_\ell, \Sigma_\ell). \quad (4.24)$$

Now, the Gaussian mixture of eq. (4.20) can be expressed in terms of the joint distribution  $p(\vec{y}, \vec{z})$ , a property that is required for the parameter estimation via the EM algorithm. As a last step before the EM parameter search, the variable

$$\begin{aligned} \gamma(z_\ell) = p(z_\ell = 1 | \vec{y}) &= \frac{p(\vec{y} | z_\ell = 1) \cdot p(z_\ell = 1)}{\sum_{j=1}^{\kappa} p(\vec{y} | z_j = 1) \cdot p(z_j = 1)} \\ &= \frac{\pi_\ell \cdot \mathcal{N}(\vec{y} | \vec{\mu}_\ell, \Sigma_\ell)}{\sum_{j=1}^{\kappa} \pi_j \cdot \mathcal{N}(\vec{y} | \vec{\mu}_j, \Sigma_j)} \end{aligned} \quad (4.25)$$

is introduced. With this definition, the  $\pi_\ell$  can be considered the prior probability of  $z_\ell = 1$ , and  $\gamma(z_\ell)$  the posterior probability for a given  $\vec{y}$ . Finally, the EM algorithm is used to iteratively search for the optimal values for  $\vec{\mu}_\ell$  and  $\Sigma_\ell$ .

There are two major differences of the GMM algorithm, compared to k-means clustering. Firstly, because the  $\gamma(z_\ell)$  are not binary, the data assignment is soft (i.e. a data point in  $Y$  can contribute to several clusters). Secondly, the GMM estimates covariance matrix  $\Sigma_\ell$  and mean  $\vec{\mu}_\ell$  for each cluster, instead of just the centroids.

## 4.5 Signals & Signal Choice

Generally speaking, smart meter signals are required for two areas of interest, event detection and appliance component classification. In this section, a definition of these signals and deduced information is given. To begin with, the smart meter measuring principle is modeled in subsection 4.5.1, followed by a direct current (DC) model and the traditional network analysis of electrical circuits in subsection 4.5.2. The section is concluded by the introduction of alternating voltages and currents as instantaneous quantities in subsection 4.5.3, and in their complex representation in subsection 4.5.4.

### 4.5.1 The Smart Meter View

A smart electricity meter is usually connected to the mains at the point of common coupling (PCC) of a building. Here, it measures voltage and current on all present wires (phases). Therefore, the basis for all further investigations is a set of voltage and current signals. Voltage and

current are related via Ohm's law  $v(t) = r(t) \cdot i(t)$ , where  $v$  is the electric voltage,  $i$  is the electric current, and  $r$  is the resistance. This means that voltage and current, the signals measured by the smart meter, depend on each other. The actual interest for NILM is the resistance  $r$ , which is characteristic for a connected load.

Meters are usually intended to provide data for the electricity bill. To do this they compute the electric power  $p(t) = v(t) \cdot i(t)$  and integrate over this to measure the consumed energy. Power and resistance are related via

$$p(t) = v(t) \cdot i(t) = \frac{v(t)^2}{r(t)} = i(t)^2 \cdot r(t). \quad (4.26)$$

This means that, if a resistance  $r$  is to be identified, and only the power  $p$  is provided by the smart meter, at least either the voltage or the current need to be measured or estimated. The power alone is not sufficient for an unambiguous identification of a load.

#### 4.5.2 Direct Current Mains Model

In an electricity distribution grid, a great multitude of loads is connected to a source, or a grid of sources in parallel. For NILM it is important that external loads in other homes or business facilities are eliminated from the measured signals. In order to find out how this can be done, the diagram drawn in Fig. 4.4 is analyzed. The load to be identified is  $G_L$ . It is the reciprocal of the load's resistance  $R_L = \frac{1}{G_L}$ . It is fed by the constant voltage  $v_s$  via a non-ideal power line with resistance  $R_m$ <sup>21)</sup>. Here, the voltage source  $v_s$  represents the grid transformer closest to the measuring point (the last one in the transformer chain seen from the power plant).  $v_m$  and  $i_m$  are measured with the smart meter. For the moment, all quantities are assumed stationary, except an additional external load  $g_e(t)$  which behaves as follows

$$g_e(t) = \begin{cases} 0 & \forall t < t_0 \\ G_1 & \forall t \geq t_0 \end{cases}, \quad (4.27)$$

where  $t_0 \in \mathbb{R}$  is an arbitrary switch-on time. It will cause a voltage drop  $\Delta v$

$$\Delta v := v_m(t \geq t_0) - v_m(t < t_0) \quad (4.28)$$

compared to  $t < t_0$ . The voltage drop will result in a lower current  $i_m$

$$i_m(t \geq t_0) = G_L \cdot (v_m(t < t_0) - \Delta v). \quad (4.29)$$

With

$$v_m(t < t_0) = v_s \cdot \left( 1 - \frac{R_m}{R_m + \frac{1}{G_L}} \right) \text{ and} \quad (4.30)$$

$$v_m(t \geq t_0) = v_s \cdot \left( 1 - \frac{v_s \cdot R_m}{R_m + \frac{1}{G_L + G_1}} \right), \quad (4.31)$$

the voltage drop becomes

$$\Delta v = v_s \left( \frac{1}{R_m (G_L + G_1) + 1} - \frac{1}{R_m \cdot G_L + 1} \right). \quad (4.32)$$

Computing the instantaneous power before

$$p_m(t < t_0) = v_m(t < t_0) \cdot i_m(t < t_0) = G_L \cdot (v_m(t < t_0))^2 \quad (4.33)$$

<sup>21)</sup>In this model,  $R_m$  represents the actual resistance of the wires as well as the internal resistance of the voltage source. The influence of the wiring between the PCC and individual homes is neglected, because the line length between the power station and the PCC is considered much longer.

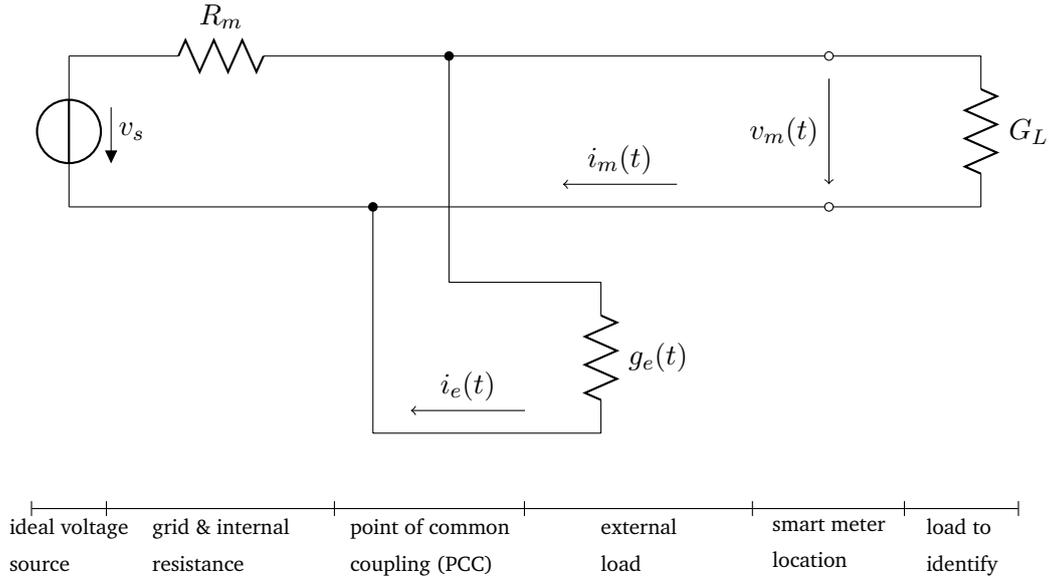


Fig. 4.4: DC model for the analysis of external load influences

and after the switching event

$$p_m(t > t_0) = v_m(t \geq t_0) \cdot i_m(t \geq t_0) = G_L \cdot (v_m(t < t_0) - \Delta v)^2, \quad (4.34)$$

it becomes clear that the voltage difference is squared in eq. (4.34) and  $p_m(t < t_0) \neq p_m(t > t_0)$ . The external load will have a considerable impact on the estimated power that the load  $G_L$  consumes. If the power signal was used for NILM it would have to be normalized with respect to a constant mains voltage in order to eliminate the external influence. This is done in state-of-the-art methods, as will be shown in section 6.2.1 later.

The external influence can be directly eliminated by computing the conductance

$$g_m(t < t_0) = \frac{i_m(t < t_0)}{v_m(t < t_0)} = G_L, \quad (4.35)$$

$$g_m(t \geq t_0) = \frac{i_m(t \geq t_0)}{v_m(t \geq t_0)} = \frac{G_L \cdot (v_m(t < t_0) - \Delta v)}{v_m(t < t_0) - \Delta v} = G_L, \quad (4.36)$$

$$g_m(t < t_0) = g_m(t > t_0), \quad (4.37)$$

where the voltage difference vanishes due to the division. The resistance  $r_m = \frac{v_m}{i_m}$  can also be used. However, in many situations the current may become 0 ( $r_m \rightarrow \infty$ ). Then the division is not allowed any more<sup>22)</sup>, which makes  $r_m$  a worse choice than  $g_m$ .

Although constant quantities are assumed, equations (4.33) through (4.37) are valid for arbitrary  $v$ ,  $i$ , and loads. If the signals or components become time dependent, i.e.  $v(t)$ ,  $i(t)$ ,  $g_L(t)$ , the *instantaneous power*  $p_m(t)$  and the *instantaneous conductance*  $g_m(t)$  are computed by

$$p_m(t) = v_m(t) \cdot i_m(t) \quad (4.38)$$

$$g_m(t) = \frac{i_m(t)}{v_m(t)}. \quad (4.39)$$

### 4.5.3 Instantaneous Quantities with Alternating Current

The voltage fed into AC power networks is a harmonic wave. Its magnitude and frequency are regulated such that they are constant within standardized tolerances<sup>23)</sup>. In this section, the

<sup>22)</sup>The voltage may also become 0 if the grid fails or in alternating current situations. This has to be kept in mind, especially if AC is applied.

<sup>23)</sup>The nominal RMS voltage in Central Europe is 230 V  $\pm$  10 % for single phase systems according to IEC 60038. The frequency is typically controlled within 50 Hz  $\pm$  0.02 Hz. The maximum allowed difference from the nominal

instantaneous power and conductance in the presence of a harmonic oscillating voltage will be analyzed. For the moment, it is assumed that magnitude  $V_1$ , frequency  $\omega_1$ , and voltage zero phase angle  $\varphi_v$  are constant<sup>24)</sup>. Then, the voltage delivered by the power plant is an ideal steady state harmonic function

$$v(t) = V_1 \cdot \cos(\omega_1 \cdot t + \varphi_v). \quad (4.40)$$

If only one linear load  $G_L$  is connected, the current will be ideal, too

$$i(t) = I_1 \cdot \cos(\omega_1 \cdot t + \varphi_i) = V_1 \cdot G_L \cdot \cos(\omega_1 \cdot t + \varphi_i). \quad (4.41)$$

The *instantaneous power* consumed by the load is

$$\begin{aligned} p(t) &= v(t) \cdot i(t) = V_1 \cdot I_1 \cdot \cos(\omega_1 \cdot t + \varphi_v) \cdot \cos(\omega_1 \cdot t + \varphi_i) \\ &= \frac{1}{2} \cdot V_1 \cdot I_1 \cdot [\cos(\omega_1 \cdot t + \varphi_v - \omega_1 \cdot t - \varphi_i) + \cos(\omega_1 \cdot t + \varphi_v + \omega_1 \cdot t + \varphi_i)] \\ &= \frac{1}{2} \cdot V_1 \cdot I_1 \cdot [\cos(\varphi_v - \varphi_i) + \cos(2\omega_1 \cdot t + \varphi_v + \varphi_i)] \\ &= \frac{1}{2} \cdot V_1 \cdot I_1 \cdot \left( \underbrace{\cos(\Delta\varphi)}_{\text{constant offset}} + \underbrace{\cos(2 \cdot \omega_1 \cdot t + \varphi_v + \varphi_i)}_{\text{double frequency}} \right), \end{aligned} \quad (4.42)$$

where  $\Delta\varphi = \varphi_v - \varphi_i$  is the phase difference between voltage and current. The instantaneous power oscillates with twice the mains frequency, and has a non-zero mean if  $\Delta\varphi \neq n \cdot 90^\circ$ ,  $n \in \mathbb{N}$ . The *instantaneous conductance* is

$$g(t) = \frac{i(t)}{v(t)} = \frac{I_1 \cdot \cos(\omega_1 \cdot t + \varphi_i)}{V_1 \cdot \cos(\omega_1 \cdot t + \varphi_v)}. \quad (4.43)$$

Here, the problem with this quantity becomes obvious. It is undefined in intervals of  $\pi$  because  $\cos(x) = 0$  for  $x = \{\frac{\pi}{2}, \frac{3\pi}{2}, \dots\}$ . With the instantaneous power this problem does not arise because of the multiplication. Therefore, it is better to use quantities that average one or multiple periods (e.g. the root mean square (RMS)), and compute them before the division. By doing so, power equivalent quantities, that do not oscillate around zero anymore, are used. This is discussed in the next subsection.

#### 4.5.4 Complex Quantities with Alternating Current

As indicated in section 4.5.2, the instantaneous conductance is free of external load influences. The main problem of the undefined division, that arose in the previous subsection, can be eliminated if (time-dependent) RMS values are computed. If used for classification, though, this RMS conductance is a one-dimensional feature. This may be too little information if e.g. similar loads<sup>25)</sup> are connected to the grid. If a linear but not completely ohmic load (e.g. an electric motor) is connected to an alternating current grid, it will introduce a phase difference between voltage and current. This phase shift may be characteristic for the load. The traditional way of describing AC circuits is to express the harmonic voltage (4.40) as the real part of the *complex voltage* [141]

$$\underline{v}(t) := V_1 \cdot e^{j(\omega_1 \cdot t + \varphi_v)} = V_1 \cdot e^{j \cdot \omega_1 \cdot t} \cdot e^{j \cdot \varphi_v} \quad (4.44)$$

$$v(t) = \Re\{\underline{v}(t)\}. \quad (4.45)$$

value is  $\pm 0.2$  Hz [140].

<sup>24)</sup>The index 1 is used here because this oscillation will later be the fundamental wave, and it is usually indexed by 1.

<sup>25)</sup>Similar with respect to their conductance.

The same can be done for the current, which becomes the *complex current*

$$\underline{i}(t) := I_1 \cdot e^{j\omega_1 t} \cdot e^{j\varphi_i}. \quad (4.46)$$

In the complex plane, the complex voltage and current can be thought of as pointers rotating with the same angular frequency  $\omega_1$  and an angle difference of  $\varphi_v - \varphi_i$ .

In order to overcome the problem of undefined quotients, the periodicity of the instantaneous quantities can be eliminated by introduction complex RMS voltages and RMS currents<sup>26)</sup>, which take into account amplitude and phase only [142]. These two parameters allow for a separation of ohmic from non-ohmic loads. Assuming steady state periodic signals  $v(t)$  and  $i(t)$ , their *complex RMS representations* are defined as

$$\underline{V}_{\text{rms}} := V_{\text{rms}} \cdot e^{j\varphi_v} = \frac{V_1}{\sqrt{2}} \cdot e^{j\varphi_v} \quad (4.47)$$

$$\underline{I}_{\text{rms}} := I_{\text{rms}} \cdot e^{j\varphi_i} = \frac{I_1}{\sqrt{2}} \cdot e^{j\varphi_i}. \quad (4.48)$$

Note, that the complex RMS voltage and current are only a symbol for their complex, time-dependent representation [142]. It has to be kept in mind that they both refer to the same frequency  $\omega_1$ . The complex RMS value with time-independent magnitude and phase is also called a *phasor* [143]. Consequently, the *complex apparent power* is defined as [144]<sup>27)</sup>

$$\underline{S} := \underline{V}_{\text{rms}} \cdot \underline{I}_{\text{rms}}^* = \frac{1}{2} \cdot V_1 \cdot I_1 \cdot e^{j\Delta\varphi}. \quad (4.49)$$

It is often divided into the *real power*  $P$  and *reactive power*  $Q$  with

$$P = \Re\{\underline{S}\} = \frac{1}{2} \cdot V_1 \cdot I_1 \cdot \cos(\Delta\varphi), \quad (4.50)$$

$$Q = \Im\{\underline{S}\} = \frac{1}{2} \cdot V_1 \cdot I_1 \cdot \sin(\Delta\varphi), \text{ and} \quad (4.51)$$

$$\underline{S} = P + j \cdot Q. \quad (4.52)$$

The physical reasons for these powers are important for the layout and dimensioning of power grids, but for this thesis it is only important to know that both components exist and that they are characteristic for a load. Similarly as before, the *complex admittance* is defined as

$$\underline{Y} := \frac{\underline{I}_{\text{rms}}}{\underline{V}_{\text{rms}}} = \frac{I_1}{V_1} \cdot e^{-j\Delta\varphi}. \quad (4.53)$$

Its components are *conductance*  $G$  and *susceptance*  $B$  with

$$G = \Re\{\underline{Y}\} = \frac{I_1}{V_1} \cdot \cos(\Delta\varphi) \quad (4.54)$$

$$B = \Im\{\underline{Y}\} = \frac{I_1}{V_1} \cdot \sin(-\Delta\varphi) = -\frac{I_1}{V_1} \cdot \sin(\Delta\varphi), \text{ and} \quad (4.55)$$

$$\underline{Y} = G + j \cdot B. \quad (4.56)$$

As discussed before, the complex admittance is more robust in practical scenarios than the impedance ( $\underline{Z} := \frac{\underline{V}_{\text{rms}}}{\underline{I}_{\text{rms}}}$ ) because the current is in the nominator and it is more likely to become zero than the voltage. Note, that, in eq. (4.53), it might be easier to compute  $\frac{I_{\text{rms}}}{V_{\text{rms}}}$  of a signal section, instead of  $\frac{I_1}{V_1}$  by taking the maximum of this section. Due to the averaging character of the RMS, distortions like spikes or drops are handled more robustly.

<sup>26)</sup>RMS and magnitude are handled inconsistently in the literature. Therefore, the subscript rms is always used when referring to RMS values.

<sup>27)</sup>The apparent power is *only* defined as a product of RMS values. Therefore, the rms index is omitted for powers.

### 4.5.5 RMS Magnitude Estimation

One important parameter required for the computation of complex AC quantities is the magnitude of the corresponding harmonic function. In this section, the properties of the RMS operation for the estimation of the magnitude are analyzed. The RMS of a signal  $x(t)$  is defined as

$$x_{\text{rms}} = \lim_{T \rightarrow \infty} \sqrt{\frac{1}{T} \cdot \int_0^T (x(t))^2 dt}. \quad (4.57)$$

If  $x(t)$  is a steady state periodic signal,  $T$  is usually chosen to be the cycle duration. In this case, the periodic signal  $x(t)$  can be represented by one single characteristic number. For sinusoid signals

$$x(t) = \hat{x}_1 \cdot \cos(\omega_1 \cdot t + \varphi_1) \quad (4.58)$$

with  $\omega_1 = \frac{2\pi}{T_1}$ , the RMS is

$$\begin{aligned} x_{\text{rms}} &= \sqrt{\frac{1}{T_1} \cdot \int_0^{T_1} \hat{x}_1^2 \cdot \cos^2(\omega_1 \cdot t + \varphi_1) dt} \\ &= \hat{x}_1 \cdot \sqrt{\frac{1}{T_1} \cdot \left[ \frac{t}{2} + \frac{\sin(\omega_1 \cdot t + \varphi_1) \cdot \cos(\omega_1 \cdot t + \varphi_1)}{2 \cdot \omega_1} \right]_{t=0}^{T_1}} \\ &= \frac{\hat{x}_1}{\sqrt{2}}, \end{aligned} \quad (4.59)$$

so RMS and magnitude are related via the factor of  $\sqrt{2}$ <sup>28)</sup>. For discrete-time signals, the RMS becomes

$$x_{\text{rms}} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (x[i])^2} = \sqrt{\frac{1}{N} \cdot \langle \vec{x}, \vec{x} \rangle}, \quad (4.60)$$

where  $\vec{x}$  is a series of  $N \in \mathbb{N}$  samples of the signal  $x(t)$ .

#### 4.5.5.1 RMS Signals

The RMS, as it was defined in eq. (4.57), was computed over an infinite time period. This is a result of the traditional AC theory, in which all signals are steady state sinusoids (of infinite length). In this section, the original definition is adapted so that it represents a signal that is changing over time. The *RMS signal* is defined as

$$x_{\text{rms}}(t) = \sqrt{\frac{1}{T} \cdot \int_t^{t+T} (x(t))^2 dt}, \quad (4.61)$$

where  $T \in \mathbb{R}_+$  is the integration period. When dealing with non-steady state AC signals, this period is usually chosen equal to the period of the fundamental oscillation. The definition (4.61) is a postulated proposition for a different representation of a given signal. For a detailed analysis of the spectral influence of the RMS computation, the procedure is split into three steps, as illustrated in Fig. 4.5.

<sup>28)</sup>A common example is the magnitude of a 230 V RMS voltage. Its peak voltage is  $230 \text{ V} \cdot \sqrt{2} \approx 325 \text{ V}$ .

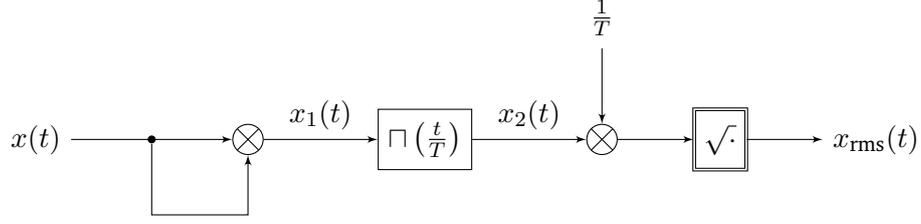


Fig. 4.5: three step model of the computation of an RMS signal

In equations, the three steps are

$$x_1(t) = (x(t))^2, \quad (4.62)$$

$$x_2(t) = \int_t^{t+T} x_1(t) dt = x_1(t) * \Pi\left(\frac{t}{T}\right), \text{ and} \quad (4.63)$$

$$x_{\text{rms}}(t) = \sqrt{\frac{1}{T} \cdot x_2(t)}, \quad (4.64)$$

where  $\Pi(\cdot)$  is the rectangle function, that is defined in eq. (A.3) in the appendix. Each step is analyzed individually in the following.

**Step 1: squaring.** In the first step, the signal is squared. Assuming that  $\omega_1$  is the lowest frequency present in  $x(t)$ , the input signal is represented by its Fourier series

$$x_{\mathcal{F}}(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cdot \cos(n \cdot \omega_0 \cdot t) + b_n \cdot \cos(n \cdot \omega_1 \cdot t)) \quad (4.65)$$

with the Fourier coefficients

$$a_n = \frac{2}{T} \cdot \int_{t_0}^{t_0+T} x(t) \cdot \cos(n \cdot \omega_1 \cdot t) dt, \quad (4.66)$$

$$a_0 = \frac{2}{T} \cdot \int_{t_0}^{t_0+T} x(t) dt, \quad (4.67)$$

and

$$b_n = \frac{2}{T} \cdot \int_{t_0}^{t_0+T} x(t) \cdot \sin(n \cdot \omega_1 \cdot t) dt. \quad (4.68)$$

For periodic signals,  $t_0 \in \mathbb{R}$  can have arbitrary values. For a real-valued sinusoid signal with frequency  $\omega_1 \in \mathbb{R}_+$  and magnitude  $x_1 \in \mathbb{R}_+$ , i.e.

$$x(t) = \hat{x}_1 \cdot \cos(\omega_1 \cdot t), \quad (4.69)$$

all  $b_n$  are zero because  $x(t)$  is symmetric. Furthermore, let  $T_1 = \frac{1}{f_1}$  with  $\omega_1 = 2\pi \cdot f_1$ . Then

$$a_0 = \frac{2}{T_1} \cdot \underbrace{\int_{t_0}^{t_0+T_1} (\hat{x}_1 \cdot \cos(\omega_1 \cdot t))^2 dt}_{= \frac{T_1}{2} \cdot \hat{x}_1^2} = \hat{x}_1, \quad (4.70)$$

$$a_n = \frac{2}{T} \cdot \underbrace{\int_{t_0}^{t_0+T_1} (\hat{x}_1 \cdot \cos(\omega_1 \cdot t))^2 \cdot \cos(n \cdot \omega_1 \cdot t) dt}_{= \begin{cases} 0 & , n \neq 2 \\ \frac{T_1}{4} \cdot \hat{x}_1^2 & , n = 2 \end{cases}} = \begin{cases} 0 & , n \neq 2 \\ \frac{\hat{x}_1^2}{2} & , n = 2 \end{cases}, \quad (4.71)$$

and

$$x_1(t) = \frac{\hat{x}_1^2}{2} \cdot (1 + \cos(2 \cdot \omega_1 \cdot t)), \quad (4.72)$$

which means that the square operation equally divides and shifts the spectral component of  $x(t)$  to the direct component and a component of doubled frequency  $2 \cdot \omega_1$ .

**Step 2: averaging.** Averaging over a window of length  $T$  is equivalent to the convolution of  $x_1(t)$  with a rectangular pulse of width  $T$ , so

$$\begin{aligned} x_2(t) &= x_1(t) * \square\left(\frac{t}{T}\right) \\ \circ & \quad \quad \quad \circ & \quad \quad \quad \circ \\ X_2(\omega) &= X_1(\omega) \cdot \text{si}\left(\frac{\omega T}{2}\right) \end{aligned} \quad (4.73)$$

The resulting frequency response function,  $\text{si}(\omega)$ , has zeros at  $x = \{\pm\pi, \pm 2\pi, \pm 3\pi, \dots\}$ . In the current example, the function parameter is  $\text{si}(\omega T/2)$  which corresponds to zeros at  $f = \frac{\omega}{2\pi} = \{f_1, 2 \cdot f_1, 3 \cdot f_1, \dots\}$ , as shown in Fig. 4.6. Consequently, the averaging leads to an elimination of the spectral components at  $\omega = j \cdot \omega_1$  with  $j \in \mathbb{N}$ , which means that only the steady component

$$x_2(t) = \frac{\hat{x}_1^2}{2}. \quad (4.74)$$

of the sinusoid signal  $x(t)$  defined in eq. (4.69) remains. Another conclusion is that the RMS signal oscillates if the input contains frequency components at  $\omega = r \cdot \omega_1$ ,  $r \in \mathbb{R}$ ,  $r \notin \mathbb{N}$ .

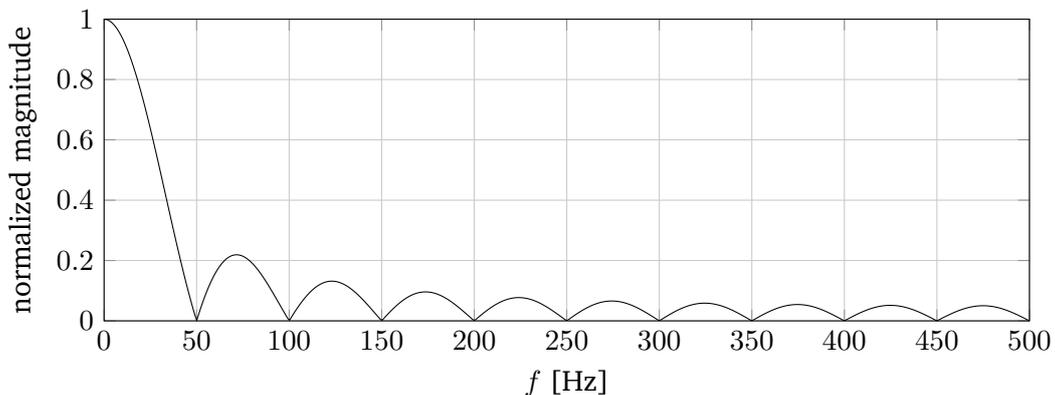


Fig. 4.6: frequency response of  $\square\left(\frac{t}{T}\right)$  with  $T = 20$  ms

**Step 3: square root.** In general, the spectral influence of a square root operation is hard to determine. However, by plugging eq. (4.64) into the Fourier coefficient (4.66)

$$a_n = \frac{2}{T} \cdot \int_{t_0}^{t_0+T} \sqrt{\frac{1}{T} \cdot x_2(t)} \cdot \cos(n \cdot \omega_1 t) dt, \quad (4.75)$$

it can be deduced that a spectral component  $X(\omega)$  with  $\omega > 0$  will introduce harmonics. This means that  $x_{rms}(t)$  will have a higher bandwidth than  $x_2(t)$ . An example for this is given in Fig. 4.7. Here, the signal  $x(t) = 1 + \cos(\cdot)$  was added a direct component in order to make sure that  $\sqrt{x(t)} \in \mathbb{R}_+$  for all  $t$ .

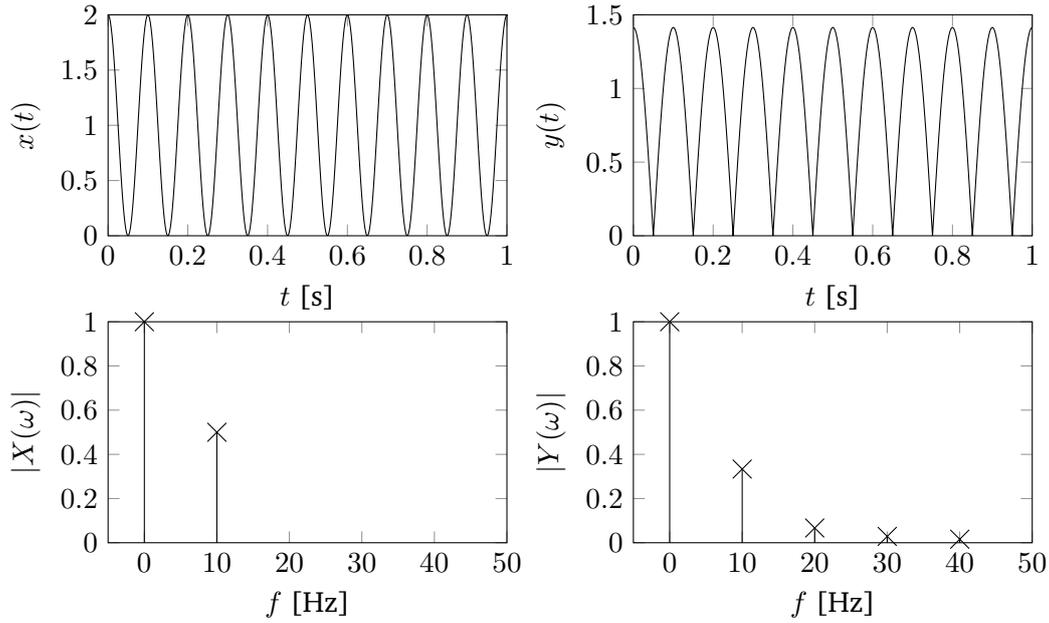


Fig. 4.7: example for the generation of harmonics by taking the square root of  $x(t) = 1 + \cos(2\pi \cdot 10 \text{ Hz} \cdot t)$ ,  $y(t) = \sqrt{x(t)}$ .

For the example signal of eq. (4.69), the square root is trivial because it only contains the direct component after the averaging stage

$$x_{rms}(t) = \sqrt{x_2(t)} = \frac{\hat{x}_1}{\sqrt{2}}. \quad (4.76)$$

The scaling factor  $\frac{1}{T}$  vanishes due to the integration over  $T$ .

#### 4.5.5.2 Signal Dynamics under the Influence of the RMS

Let  $c(t)$  represent the physical behavior of an appliance, i.e. an equivalent to its time-dependent electric resistance. The measured current signal can be interpreted as the modulation of a carrier signal  $v(t) = \hat{v} \cdot \cos(\omega_1 \cdot t)$  with the load behavior as the wanted signal  $c(t)$ . It is

$$x(t) = v(t) \cdot s(t) = \hat{v} \cdot \cos(\omega_1 \cdot t) \cdot c(t). \quad (4.77)$$

In this subsection, the bandwidth limits of  $c(t)$  are analyzed. In the frequency domain, the modulation (4.77) is a shift of the spectral components of  $c(t)$  towards the carrier frequency  $\omega_1$ :

$$\begin{aligned} X(\omega) &= \hat{v} \cdot \mathcal{F}\{\cos(\omega_1 \cdot t)\} * C(\omega) \\ &= \frac{1}{2} \cdot \hat{v} \cdot (\delta(\omega - \omega_1) + \delta(\omega + \omega_1)) * C(\omega) \\ &= \frac{1}{2} \cdot \hat{v} \cdot (C(\omega - \omega_1) + C(\omega + \omega_1)) \end{aligned} \quad (4.78)$$

In order to avoid aliasing, the condition

$$C(\omega) = 0 \Big|_{|\omega| > \frac{\omega_1}{2}} \quad (4.79)$$

must be fulfilled for  $c(t)$ . Hence,  $c(t)$  can only be reconstructed from  $x(t)$  if it does not contain spectral components with  $|\omega| \geq \frac{\omega_1}{2}$ . The corresponding bandwidth  $B_C$  of  $c(t)$  has to fulfill

$$B_C < \omega_1. \quad (4.80)$$

The squaring operation (step 1) of the RMS yields

$$x_1(t) = x(t) \cdot x(t) = \hat{v}^2 \cdot \cos(\omega_1 t) \cdot \cos(\omega_1 t) \cdot c(t) \cdot c(t) \quad (4.81)$$

$$X_1(\omega) = \frac{1}{4} \cdot \hat{v}^2 \cdot (\delta(\omega - 2\omega_1) + 2\delta(\omega) + \delta(\omega + 2\omega_1)) * C(\omega) * C(\omega) \quad (4.82)$$

Following the rules of the convolution, the bandwidth  $B_{C*C}$  of  $C(\omega) * C(\omega)$  is limited by<sup>29)</sup>

$$B_C \leq B_{C*C} \leq 2 \cdot B_C. \quad (4.83)$$

Using this and eq. (4.82), the bandwidth of  $c(t)$  after the squaring is limited by

$$2 \cdot B_C < 2 \cdot \omega_1. \quad (4.84)$$

Consequently, the squaring has no influence on the requirements regarding the bandwidth of  $c(t)$ . The averaging with length  $T_1$  (step 2) eliminates all spectral components at  $\omega = \{\omega_1, 2 \cdot \omega_1, 3 \cdot \omega_1, \dots\}$  (cf. Fig. 4.6). This means that the requirement  $B_C < \omega_1$  remains unchanged after the averaging and, approximately, also after the square root (step 3). The RMS operation shifts spectral components of  $X(\omega)$ , the spectrum of the measured signal, towards the frequency region around  $\omega = 0$ , yielding an approximation of  $C(\omega)$ , the spectrum of the wanted signal. The RMS can therefore be interpreted as a *demodulation method*.

Common AC electricity systems use a fundamental frequency of  $f_1 = 50$  Hz. Hence, evaluating condition (4.79), the wanted signal  $c(t)$  in eq. (4.77) cannot have patterns shorter than

$$t_{\min} \approx \frac{2\pi}{\frac{\omega_1}{2}} = \frac{2}{50 \text{ Hz}} = 40 \text{ ms} \quad (4.85)$$

or, in frequency terms, components at frequencies  $f > 25$  Hz. For discrete-time (power) signals, this means that a sampling rate of 50 Hz after the RMS computation is sufficient. This rate will be assumed for all RMS signals in the following.

A real data example is given in Fig. 4.8. The upper graph shows the turn-on settling period of a refrigerator, which is sampled at a rate of 1 kHz. The gray line marks the measured current. The black line is its RMS signal computed according to equations (4.62) through (4.64) and an averaging period of  $T = 20 \text{ ms} = \frac{1}{50 \text{ Hz}}$ . The lower graph shows the corresponding amplitudes of the spectra, that were computed from this section of the signal at the various stages of the RMS computation. It is obvious that the appliance turn-on settling process is limited to a frequency range of  $0 < f < 20$  Hz.

#### 4.5.6 Phase Difference Estimation

The phase difference of voltage and current is the second piece of information that required for the complex RMS representation. Using the Fast Fourier Transform (FFT), it can be conveniently computed in the frequency domain. It is well known that the Fourier transform of a sinusoid  $x(t) = \hat{x} \cdot \cos(\omega_1 \cdot t + \varphi)$  is

$$\underline{X}(\omega) = \mathcal{F}\{x(t)\} = \sqrt{\frac{\pi}{2}} \cdot \hat{x} \cdot (\delta(\omega - \omega_1) \cdot e^{-j\varphi} + \delta(\omega + \omega_1) \cdot e^{j\varphi}). \quad (4.86)$$

<sup>29)</sup>The lower limit can be explained using a rectangular pulse  $\Pi(t)$  because  $(\Pi(t))^2 = \Pi(t)$ . Hence,  $\mathcal{F}\{\Pi(t)\} * \mathcal{F}\{\Pi(t)\} = \mathcal{F}\{\Pi(t)\}$ , i.e. the bandwidth required for the squared rectangular pulse is equal to its original.

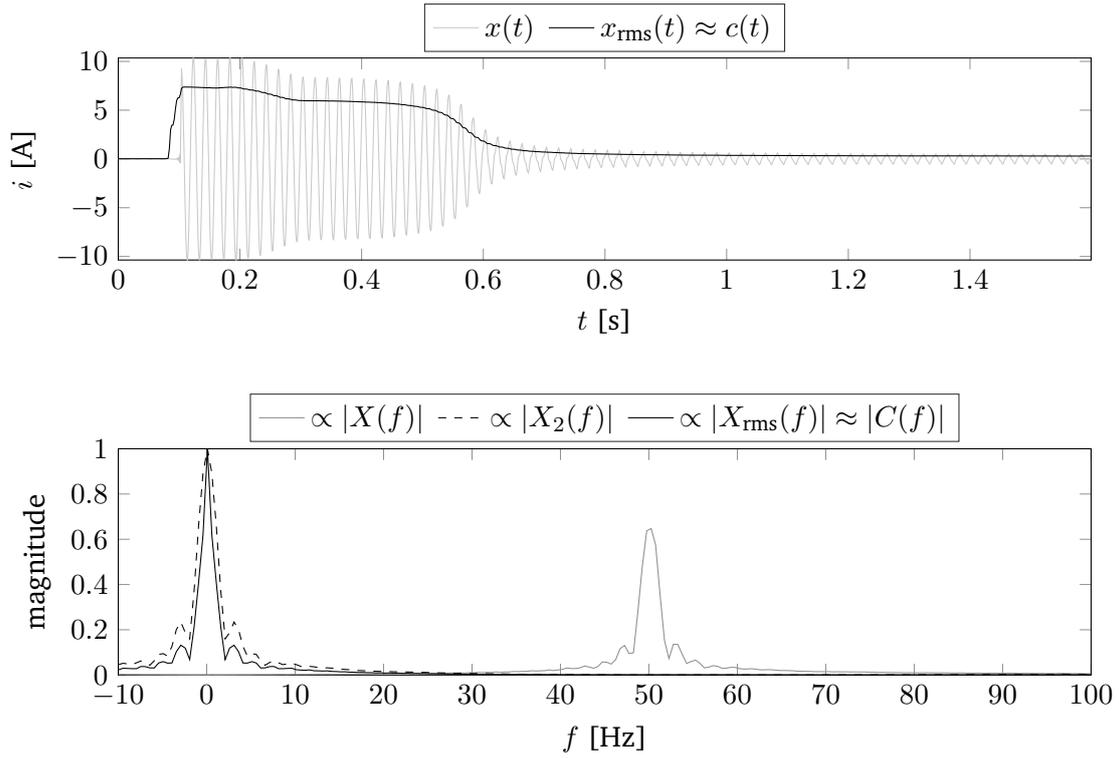


Fig. 4.8: real data example of the RMS operation in time and frequency domain

With

$$\underline{X}_1 = \underline{X}(\omega) |_{\omega=\omega_1} = \sqrt{\frac{\pi}{2}} \cdot \hat{x} \cdot \delta(\omega - \omega_1) \cdot e^{-j\varphi}, \quad (4.87)$$

the zero phase angle of  $x$  is given by

$$\varphi = -\arg\{\underline{X}_1\}. \quad (4.88)$$

This can be used directly in the complex RMS representations in equations (4.47) and (4.48).

#### 4.5.7 Conclusion

The complex admittance (or its components) is considered the most robust signal for NILM applications, because it eliminates voltage differences, that result from the operation of external appliances. In the complex apparent power, this voltage difference is squared. Therefore, the complex apparent power could be used as well, but a prior normalization of the voltage is necessary to eliminate influences of other homes. The normalization will be described in section 6.2.1 on page 80 in more detail. The power is especially helpful for training of unknown loads. The identification may then be supported by the nominal power given on their type plate.

Admittance and apparent power should be computed from RMS signals and an independent phase estimation, rather than directly from  $v_m(t)$  and  $i_m(t)$ , because the periodicity makes some computations invalid. The complex admittance can be used as a symbol only. Details on the computation of the RMS and the phase difference were given in sections 4.5.5 and 4.5.6, respectively.

## 4.6 Evaluation Methods

In order to compare disaggregation algorithms, a set of quality numbers is required to measure their performance. In sections 4.6.1 and 4.6.2, the numbers that will be used for performance

measurements of load monitoring algorithms are defined. The contents of this section are taken from a paper that was written for the IECON conference in Vienna in November 2013 by the author of this thesis [13]. A survey of public data sets of NILM data is given in section 4.7. Finally, ReSP data sets, that were created at Furtwangen University, are presented in section 4.8.

#### 4.6.1 Quality Measures for Event Detection

In the event detection phase, the base information is the number of events. It is assumed that, by appropriate ground truth measurement, a list of  $N_e$  true switching events  $e_i^{(gt)}$

$$\vec{e}^{(gt)} = \left[ e_1^{(gt)} \quad e_2^{(gt)} \quad \dots \quad e_i^{(gt)} \quad \dots \quad e_{N_e}^{(gt)} \right]^T \quad (4.89)$$

is available, where  $e_i^{(gt)}$  is the time stamp of the  $i$ -th event. Additionally, the ground truth data contain a vector of appliance IDs  $c_i^{(gt)}$

$$\vec{c}^{(gt)} = \left[ c_1^{(gt)} \quad c_2^{(gt)} \quad \dots \quad c_i^{(gt)} \quad \dots \quad c_{N_e}^{(gt)} \right]^T, \quad (4.90)$$

which is required in the next section. An event detection algorithm generates a list of detected event time stamps  $e_j^{(d)}$

$$\vec{e}^{(d)} = \left[ e_1^{(d)} \quad e_2^{(d)} \quad \dots \quad e_j^{(d)} \quad \dots \quad e_{N_d}^{(d)} \right]^T. \quad (4.91)$$

$N_d$  is the number of detected events. Note, that it may happen that  $N_d > N_e$  or even  $N_d \gg N_e$ , depending on signal noise and the parametrization of the event detection algorithm.

##### 4.6.1.1 True Positives & True Positive Rate

A detected event  $e_j^{(d)}$  is considered a *true positive (TP)*, when there exists an  $e_i^{(gt)}$  such that

$$\left| e_i^{(gt)} - e_j^{(d)} \right| < \Delta t_{\text{tol}}, \quad i = 1, \dots, N_e, \quad (4.92)$$

where  $\Delta t_{\text{tol}} \in \mathbb{R}$  is a time tolerance. It allows for the correction of small errors between the ground truth data and the actually detected events. Let  $N_{TP}$  be the number of true positives. Obviously, this is limited by the number of true switching events, i.e.  $N_{TP} \leq N_e$ . The *true positive rate (TPR)* is the relation of number of true positives and true number of events (ground truth), i.e.

$$\text{TPR} = \frac{N_{TP}}{N_e}. \quad (4.93)$$

Its value depends on the quality of the algorithm under test (AUT), which is what the evaluation is about, and on the quality of the ground truth data (and therefore the tolerance  $\Delta t_{\text{tol}}$ ). Bad ground truth data will result in bad algorithm performance according to this measure.

##### 4.6.1.2 False Negatives & False Negative Rate (Misses)

It may happen that the event detector misses an event, i.e. that there exists at least one event  $e_i^{(gt)}$  for which no detected event  $e_j^{(d)}$  was found according to eq. (4.92). These events are called *false negatives (FNs)*. The number of false negatives is  $N_{FN} = N_e - N_{TP} \geq 0$ , and the *false negative rate (FNR)* is

$$\text{FNR} = \frac{N_{FN}}{N_e}. \quad (4.94)$$

Note that, by this definition,

$$\text{TPR} + \text{FNR} = 1, \quad (4.95)$$

meaning that each ground truth event is either a true positive, if it is found by the algorithm under test, or a false negative, if it is not found. An optimal event detector achieves  $\text{TPR} = 1$  and  $\text{FNR} = 0$ .

#### 4.6.1.3 False Positives & False Positive Rate

Measured signals may contain edges, ramps, or peaks that are not described by the ground truth data. Nevertheless, an event detector is very likely to find an event here. These events are called *false positives (FPs)*. So, if no match was found for a detected event  $e^{(d)}$  according to eq. (4.92), it is counted as a false positive. The number of false positives is  $N_{FP} = N_d - N_{TP}$ , with the corresponding *false positive rate (FPR)*

$$\text{FPR} = \frac{N_{FP}}{N_e}. \quad (4.96)$$

The false positive refers to the number of ground truth events. Therefore it is possible to have  $\text{FPR} > 100\%$ , depending on the signal's quality and the sensitivity of the algorithm under test. An optimal event detector would not find any additional events, therefore the optimal FPR is 0.

A graphic overview of the event detection parameters is given in Fig. 4.9.

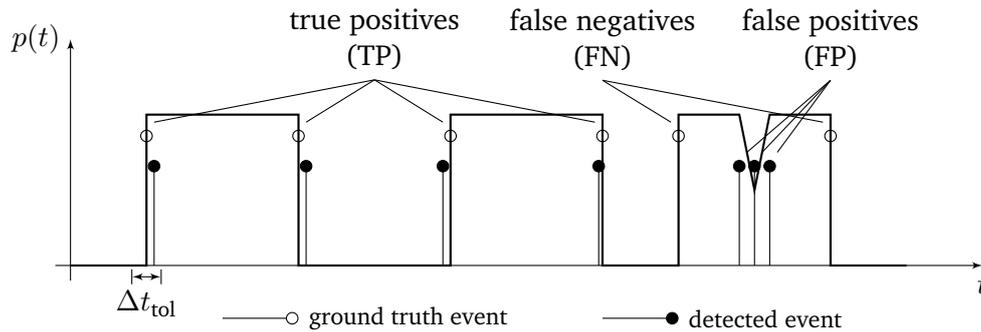


Fig. 4.9: definitions for the evaluation of event detection algorithms

## 4.6.2 Quality Measures for Load Identification

In the load identification case, there are several ways to measure an algorithm's performance. The measures may be based on the number of appliance *events* correctly classified, the number of appliance *cycles* correctly classified, or the energy correctly assigned (especially interesting for NILM with the goal to disaggregate the consumed energy).

Appliance events are single events caused by an appliance. Turning on an appliance and turning it off again are two events. Appliance cycles are pairs of turn-on and turn-off events. See Fig. 4.10 for an example. Consequently, multiple absolute and relative quality measures are defined here.

### 4.6.2.1 Appliance Events

Let there be  $N_u$  unique appliances, and let

$$\vec{a}^{(gt)} = \left[ a_1^{(gt)} \quad a_2^{(gt)} \quad \dots \quad a_\ell^{(gt)} \quad \dots \quad a_{N_u}^{(gt)} \right]^T \quad (4.97)$$

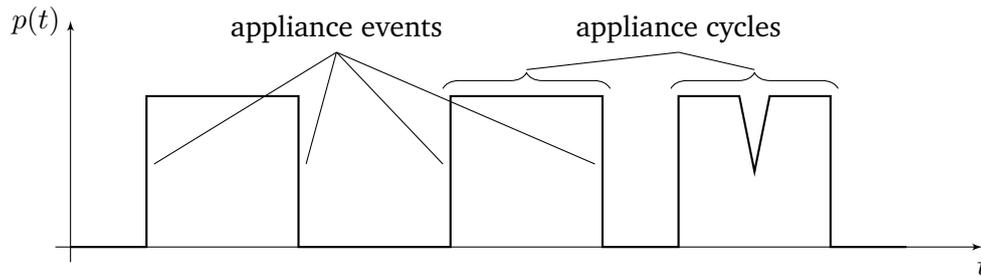


Fig. 4.10: definitions for evaluation of the appliance classification stage

be the vector of numbers of events for each appliance. The numbers of appliance events that were classified correctly by the AUT are given in the vector

$$\vec{a}^{(d)} = [a_1^{(d)} \quad a_2^{(d)} \quad \dots \quad a_\ell^{(d)} \quad \dots \quad a_{N_u}^{(d)}]^T. \quad (4.98)$$

It is computed by counting matching appliance IDs for all true positives (as defined in section 4.6.1.1). The quality of this vector therefore depends on the time tolerance  $\Delta t_{\text{tol}}$ , as well. The *appliance event classification rate (AECR)* is defined as

$$\text{AECR} = \frac{\sum_{\ell=1}^{N_u} a_\ell^{(d)}}{\sum_{\ell=1}^{N_u} a_\ell^{(gt)}} = \frac{\sum_{\ell=1}^{N_u} a_\ell^{(d)}}{N_e}. \quad (4.99)$$

The total number of correct appliance events cannot be greater than the number of true positives. This means that the TPR is the limit for the AECR. Of course, the AECR could be referred to the number of TPs instead of the total number of events. This would increase the resulting numbers but it would lead to confusion: an AECR of 100% would mean that all TPs were classified correctly but not all ground truth events if  $\text{TPR} < 100\%$ .

This value may be computed for each appliance individually by

$$\text{AECR}_\ell = \frac{a_\ell^{(d)}}{a_\ell^{(gt)}}, \quad \ell = 1, \dots, N_u. \quad (4.100)$$

Here it is wise to take the individual appliance's ground truth events as a basis, therefore the index  $\ell$  in the denominator of eq. (4.100). The appliance event classification rate measures how good a classifier for load signatures is on single event basis. It may also be used to measure the quality of the signatures themselves, if the AUT's performance is known for another set of signatures.

#### 4.6.2.2 Appliance Cycles

Let  $N_c$  be the total number of appliance cycles (combination of on- and off events) that were caused by any of the connected appliances. Further, let

$$\vec{b}^{(gt)} = [b_1^{(gt)} \quad b_2^{(gt)} \quad \dots \quad b_\ell^{(gt)} \quad \dots \quad b_{N_u}^{(gt)}]^T \quad (4.101)$$

be the vector that lists the numbers of cycles for each appliance. The numbers of appliance cycles that were classified correctly by the AUT are represented by the vector

$$\vec{b}^{(d)} = [b_1^{(d)} \quad b_2^{(d)} \quad \dots \quad b_\ell^{(d)} \quad \dots \quad b_{N_u}^{(d)}]^T. \quad (4.102)$$

The *appliance cycle classification rate (ACCR)* is

$$\text{ACCR} = \frac{\sum_{\ell=1}^{N_u} b_{\ell}^{(d)}}{\sum_{\ell} b_{\ell}^{(gt)}} = \frac{\sum_{\ell=1}^{N_u} b_{\ell}^{(d)}}{N_c}. \quad (4.103)$$

As before, this value may be computed for each appliance individually by

$$\text{ACCR}_{\ell} = \frac{b_{\ell}^{(d)}}{b_{\ell}^{(gt)}}, \quad \ell = 1, \dots, N_u. \quad (4.104)$$

For complex (autonomously controlled) appliances, such as televisions, washing machines, or personal computers, it may be very difficult to find appropriate turn-on and turn-off pairs. Although even complex appliances are turned on or off eventually, there may be many events between the turn-on and the turn-off event due to their autonomous behavior. Therefore, it may be difficult to compute the appliance classification rate.

For a fully working NILM system, the appliance cycle classification rate is very important. It is not enough to have an algorithm which is able to classify events correctly. Only if it is able to detect whole appliance cycles, it can provide useful information on the appliance's duration of usage and energy consumption. For Activity Monitoring on the other hand, the AECR is sufficient, because each single appliance switching event provides activity information.

#### 4.6.2.3 Feature Specific Quality Measures

The quality of features can be measured by applying a event detection or machine learning algorithm, and compute the numbers mentioned before. This means that the algorithm may falsify the results if it is not able to handle the data correctly. Therefore, special quality measures that bypass this problem are introduced in the chapters they are required in.

## 4.7 Public Data Sets

Having an extensive pool of data is important for a meaningful evaluation of NILM algorithms. Several researchers were concerned with this fact and published specific data sets. In this section, some of these data sets are discussed. Where applicable, their purpose for this thesis is described.

**REDD:** The REDD data set by J. Z. Kolter and M. J. Johnson [112] contains electricity data of 6 real homes. This could be an interesting choice for the activity profile analysis in chapter 7.1. However, it turned out that the data contains too many gaps so that the overall volume is too low for a meaningful result.

**BLUED:** under the title “BLUED: A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research”, this data set was created at Carnegie Mellon University, USA. It targets event based approaches and provides high-grained voltage and current data, as well as a ground truth event list [111]. The publicly available part contains measurement data from a single-family home in Pittsburgh, PA, USA of a period of one week. Voltage and current were sampled at a rate of 12 kHz<sup>30)</sup>. The ground truth data was acquired in a manual process using plug-level meters, environmental sensors, and circuit panel meters. They were manually tuned to match the time scale of the voltage and current data. This data set is used in this thesis to verify the results of the event detection and appliance identification stages using real measurement data.

<sup>30)</sup>In their paper, the authors mention that, due to a faulty sensor dimensioning, the current signals contain only components up to 300 Hz [111, p. 3].

**tracebase:** tracebase was created by a work group at Darmstadt University, Germany. The authors' goal was to analyze the behavior of electric equipment in general [113]. The data was recorded using plug level meters (intrusive load monitoring). The data set contains more than 1,000 traces of the real power, sampled at an interval of about 1 Hz. Since this data set only provides real power, the data it will not be used in this thesis.

**Smart\* Data Set for Sustainability:** over 400 households were analyzed for the Smart\* data set. Not only does it contain electricity information in different levels of detail (there is sub-metered data available for one household), it also provides environment and activity information [114]. Since it contains only values on a minute time scale, this data set cannot be used for the purposes of this thesis.

**UK-DALE:** the UK-DALE data set contains measurements from five houses, three of which were sub-metered with up to 54 individual meters. The aggregate signals were sampled with a 44.1 kHz audio analog-to-digital converter (ADC), and stored at a rate of 16 kHz (voltage & current) and 1 Hz (voltage, real and reactive power). The sub-metered signals are available in 6 s intervals. In total, 910 days of data are available [115]. This is an interesting data set for activity profile analysis and will be used in chapter 7.1 in comparison with own data.

## 4.8 ReSP Data Sets

After the public data, some ReSP data sets, that were recorded at Furtwangen University for the purposes of NILM research, are presented in this section. Parts of them were acquired by Thomas Bier, who published his PhD thesis in this area as well [145]. Other parts are the work of the author of this thesis. The data sets are labeled with roman numbers. Each one is described in its following subsection.

### 4.8.1 ReSP Data Set I

Data set I contains superposed appliance cycles, that were recorded under laboratory conditions by Thomas Bier. It is split into 10 files, where each file contains 1000 switching events of 10 individual appliances, so each appliance has 50 turn on and 50 turn off events in each file. The appliances are: (ID = 1) water kettle, (2) hand-held electric mixer, (3) freezer, (4) hair-dryer 1, (5) mixer, (6) lamp, (7) room heater, (8) energy saving lamp\*, (9) hair-dryer 2, and (10) radio\*. Appliances marked with an asterisk (\*) cannot be taken into account because their power consumption falls below the resolution of the measuring equipment. The time between consecutive events was chosen randomly. The distribution is shown in Fig. 4.11.

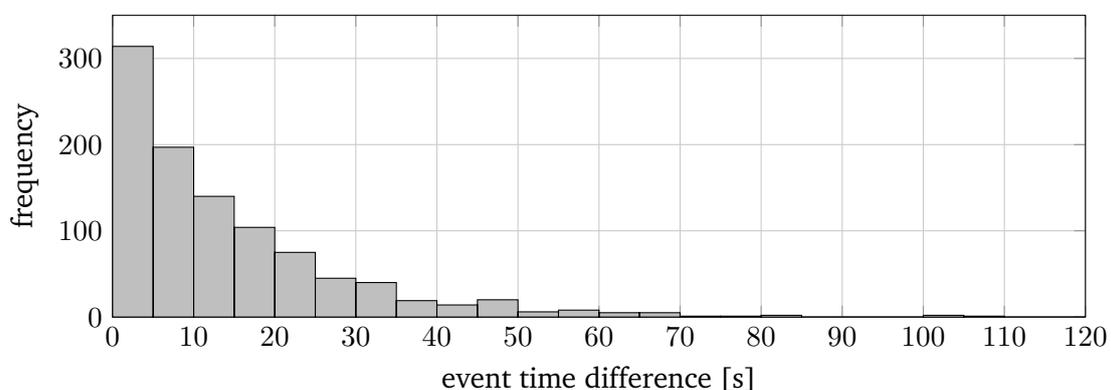


Fig. 4.11: time difference between consecutive events in data set I

### 4.8.2 ReSP Data Set II

This data set was also acquired by Thomas Bier. It was specifically designed for event detection evaluations. It contains several individual files, that were recorded under laboratory conditions without superpositions. It is divided into three subsets, according to the active appliance. The three subsets are summarized in Table 4.1.

Table 4.1: appliances and switching events in data set II

file no.	no. of events	scenario
<b>II-1: electric room heater (<math>P = 1600</math> W)</b>		
1	400	random events
2	1,200	random events
3	480	random events with $\Delta t > 1$ s
4	9,000	$\Delta t$ decrease from 1 s to 0.2 s in 0.1 s steps
5	1,500	$\Delta t$ decrease from 15 s to 1 s in 1 s steps
<i>sum</i>	<i>12,580</i>	
<b>II-2: incandescent lamp (<math>P = 40</math> W)</b>		
1	1,200	random events
2	400	random events
3	400	random events
4	9,000	$\Delta t$ decrease from 1 s to 0.2 s in 0.1 s steps
5	1,500	$\Delta t$ decrease from 15 s to 1 s in 1 s steps
<i>sum</i>	<i>12,500</i>	
<b>II-3: freezer</b>		
1	92	random events
2	200	random events
<i>sum</i>	<i>292</i>	
<b>total sum</b>	<b>25,372</b>	

### 4.8.3 ReSP Data Set III

This data set represents appliances that could pose a danger to humans. It was recorded by Philipp Klein. Special emphasis lies on heating elements. The measurements are augmented by hand mixer events to increase the variety of the data. The intention behind this data set is to specifically test the functionality of event detection algorithms for the detection of hazards that are caused by electric appliances.

Table 4.2: appliance switching events in data set III

appliance class	no. of appliances	no. of events
cooling devices	2	178
immersion heaters	1	122
irons	1	90
water kettle	1	66
hand mixers	1	300
stoves/ovens	3	1,962
superpositions of the aforementioned (IIIs)		2,703
<b>sum</b>	<b>9</b>	<b>4,420</b>

#### 4.8.4 ReSP Data Set IV

This data set is a collection of single appliance measurements. It was created by Thomas Bier. Each file contains the trace of a single appliance, so there are no superposed operation cycles. The event times of the ground truth list were extracted by manually fine-tuning an event detector after the measurement phase with a precision of 10 ms (sub-period time scale). In the data set, there are five different classes. The set contains 112 events of 31 different appliances. It is summarized in Table 4.3.

#### 4.8.5 ReSP Data Set V

Data set V is similar to IV, but the number of appliance classes was increased to 10. The set contains 577 events of 137 different appliances. It is also summarized in Table 4.3. The ground truth events were detected automatically by an event detection algorithm, and verified manually. The time stamp precision is in the range of some periods of the 50 Hz fundamental component of the grid.

Table 4.3: appliance switching events in data sets IV and V

id	appliance class	data set IV		data set V	
		# appliances	# events	# appliances	# events
1	cooling devices	3	10	9	36
2	water kettle	12	23	39	41
3	microwave	3	43	28	189
4	toaster	12	24	25	26
5	heating pump	1	12	3	13
6	TV	—	—	10	19
7	PC	—	—	9	13
8	room heater	—	—	1	215
9	TFT monitor	—	—	12	24
10	washing machine	—	—	1	1
<b>sum</b>		<b>31</b>	<b>112</b>	<b>137</b>	<b>577</b>

#### 4.8.6 ReSP Data Set VI

This data set is described in section 7.1.2 on page 111.



## Chapter 5 | Event Detection

### 5.1 Introduction

Event detection is an extensive term. In NILM, it is the process of finding time instants in an electricity measurement, at which a load or a load component was turned on or off. Considering the electric power during this period, the event detection can also be considered a step or edge detection, as it is known in the image processing field. Therefore, methods of the two-dimensional image domain can be used for the one-dimensional power signals. This is discussed in the following section 5.2. There are methods that were developed specifically for the event detection in NILM applications. One example is discussed in section 5.3. In section 5.4 a new algorithm that adapts itself to the power signal is introduced. The performance of these three algorithms is compared in various scenarios in section 5.5. Section 5.6 gives a conclusion of the chapter.

For the description of the event detection problem, the model from eq. (4.1) is used here. The task for the event detection is to find the time instances where  $s(t)$  changes from 0 to 1 and vice versa, i.e. the event detector output  $ed(t)$  should be<sup>31)</sup>

$$ed(t) = \frac{ds(t)}{dt}. \quad (5.1)$$

Through smart meter measurement, only the signal  $p(t) = s(t) \cdot c(t) \cdot v(t)$  can be acquired. This means that  $s(t)$  has to be estimated from it. To this end, the assumption, that the periodic signal  $v(t)$  is eliminated, e.g. by computing the RMS, is made. In an ideal scenario, the power directly follows the input which means that the appliance is linear and time-invariant (LTI), i.e.  $c(t)$  is the constant admittance of the appliance. In this case,  $ds(t)/dt$  can be directly estimated from the power signal by

$$ed(t) \approx \frac{dp(t)}{dt}. \quad (5.2)$$

This case is sketched in Fig. 5.1b). In reality, not all appliances are LTI systems, which means that  $p(t)$  does not have to be directly proportional to  $s(t)$ . In those cases, the event detector should work like in the ideal scenario, i.e. find only the edges in the input signal  $s(t)$ . Two examples are given in Fig. 5.1c) and 5.1d).

The straightforward way of recovering  $s(t)$  from  $p(t)$  would be to find the inverse appliance transfer function  $c^{-1}(t)$ , apply this to the measured signal, and estimate  $s(t)$ , as depicted in the vertical branch of Fig. 5.1a). This requires the knowledge of all possible appliance transfer functions. Since the variety of appliances is vast, a more general approach is sought. The task in this chapter is therefore to find a general approach to event detection without appliance specific models. Nevertheless, the event detection should be able to detect all switching events, regardless of an appliance's power consumption, while keeping the number of false positives as low as possible.

---

<sup>31)</sup>A word on the notation:  $ed(t) \in \{-1, 0, 1\}$  (or  $ed[k]$  in for discrete-time signals) is the *event indication signal*, i.e.  $|ed(t)| = 1|_{t=t_0}$  indicates an event at  $t_0$ . The value -1 indicates a turn-off event, turn-on events are marked by the value +1. In contrast,  $ei[\ell]$ ,  $\ell \in \mathbb{N}$  is the *event index signal* containing the time instances or indexes at which  $|ed[k]| = 1$ .

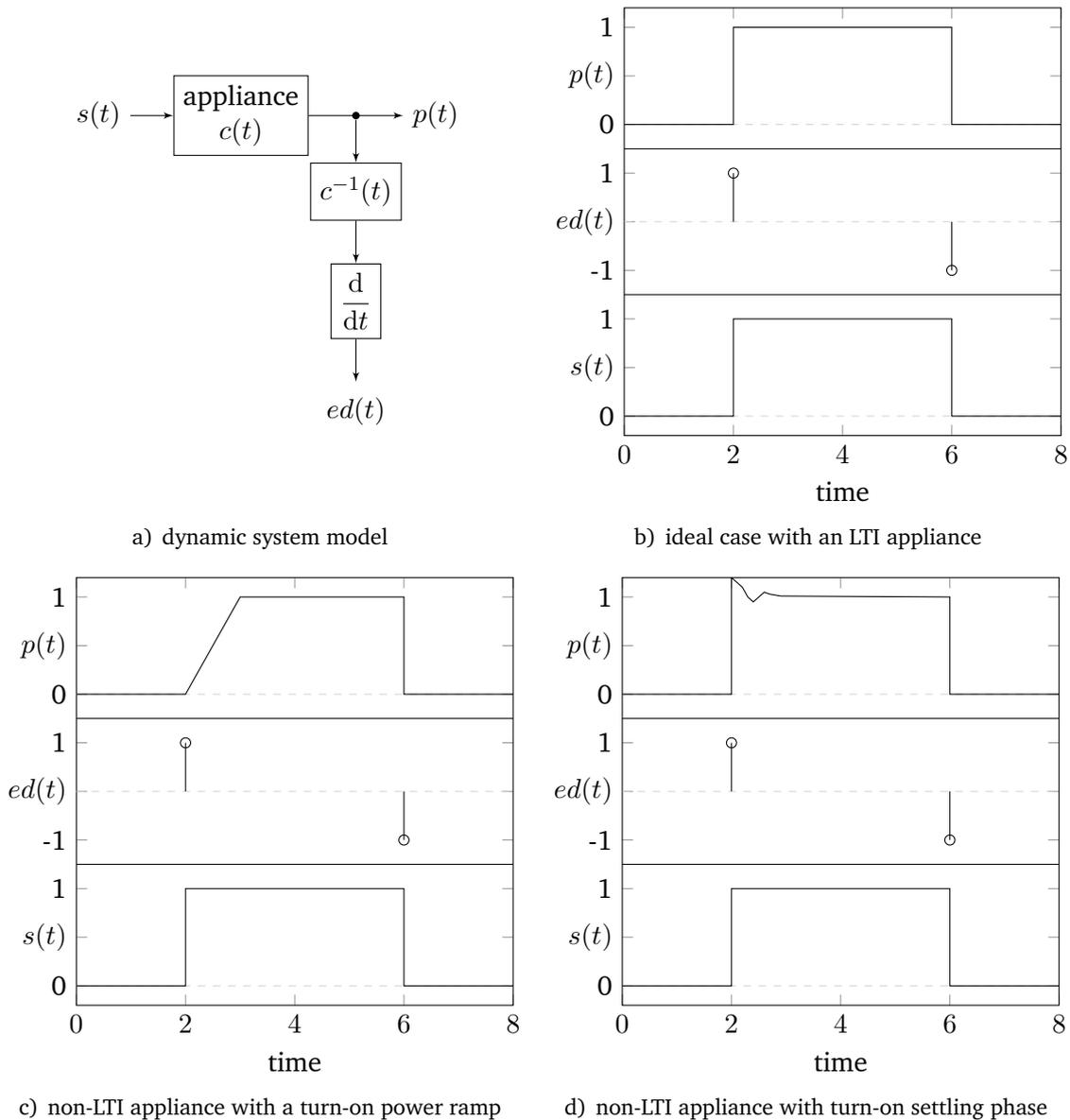


Fig. 5.1: event model, and examples for several special cases

## 5.2 Event Detection using Weighted Difference Edge Detection

The most common way to edge detection is to approximate the derivative  $ds(t)/dt$  by a weighted difference of the power signal. This is depicted in Fig. 5.2. First, the signal is smoothed by a low-pass filter to reduce noise. Then, a derivative, or an equivalent of a derivative, is applied for transforming edges into peaks. Finally, the derived signal  $d(t) \in \mathbb{R}$  is mapped to  $\{-1, 1\}$  by a threshold decision, indicating a turn off or a turn on event, respectively.

The smoothing filter and the derivative are often combined in a single convolution operation, such that

$$d(t) = x(t) * h_d, \quad (5.3)$$

where  $h_d$  is the transfer function of the combined filter. Usually, in NILM, discrete-time signals are used, so the derivative is approximated by a finite difference. Hence the term *weighted difference edge detection*. Examples are

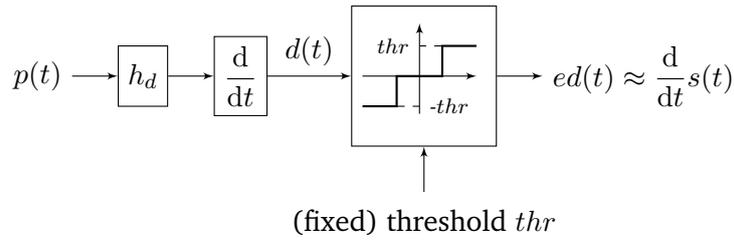


Fig. 5.2: fixed threshold WDE event detection

1. the simple difference

$$\vec{h}_{\text{diff}} = [-1 \ 1]^T \Rightarrow d[k] = x[k] - x[k-1], \quad (5.4)$$

2. the Prewitt operator

$$\vec{h}_{\text{Prewitt}} = [-1 \ 0 \ 1]^T, \quad (5.5)$$

3. and the Sobel operator

$$\vec{h}_{\text{Sobel}} = [-2 \ 1 \ 2]^T. \quad (5.6)$$

### 5.2.1 Full System Model

The system model in Fig. 5.3a) contains all components required for a discussion of the search for optimal values of the parameters mentioned in the previous section. The components are from left to right: the input is the piece-wise constant signal  $s(t)$  which indicates the operating state of an appliance. The appliance transfer function  $c(t)$  transforms the input signal into a power signal. This enters the smart meter, where measurement noise  $n_m(t)$  originating from the smart meter electronics is added. This noise is assumed to be normally distributed with

$$n_m(t) \sim \mathcal{N}(\mu_m, \sigma_m^2) \quad (5.7)$$

and  $\mu_m = 0$ . The analog signal  $p(t)$  is converted to a stream of discretized values using  $N_b$  bits. The discretization noise is modeled by  $n_d(t)$ . The influence of current transformers or voltage transducers is neglected. The discrete-time discretized power signal  $p[k]$  now leaves the smart meter and enters the event detector.

The number of bits,  $N_b$ , is assumed to be large enough for the measurement noise to be dominant over the discretization noise, and therefore  $n_d(t)$  is assumed zero. This leads to the simplified model sketched in Fig. 5.3b). Here, all signals are assumed to be discrete-time representations from the beginning.

### 5.2.2 Proposed WDE Algorithm

The event detector proposed here is a more general form of the simple difference in eq. (5.4), the Prewitt operator in eq. (5.5), and a difference that was formulated by Rosenfeld [83]. It approximates the derivative by averaging and subtracting two sections of the real power signal

$$d[k] = \underbrace{\frac{w_3}{k_3} \cdot \sum_{i=0}^{k_3-1} p[k-i]}_{\text{right window}} - \underbrace{\frac{w_1}{k_1} \cdot \sum_{i=k_2+k_3}^{k_1+k_2+k_3-1} p[k-i]}_{\text{left window}}. \quad (5.8)$$

An event is indicated if

$$|d[k]| > P_{\text{thr}}, \quad (5.9)$$

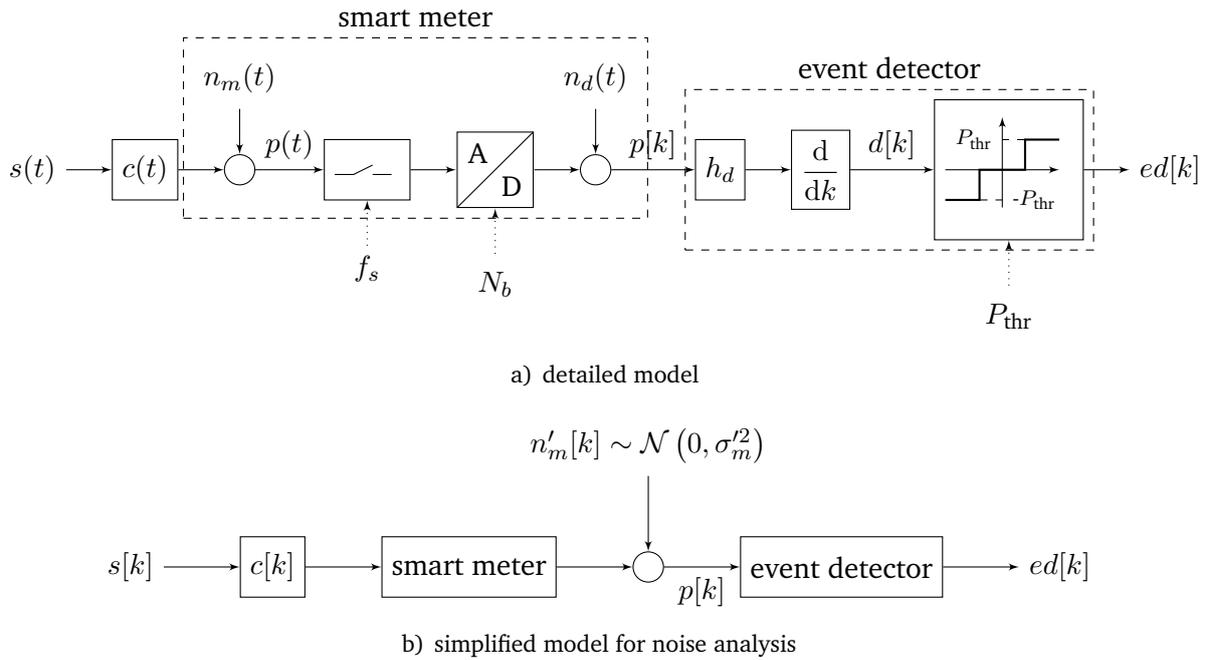


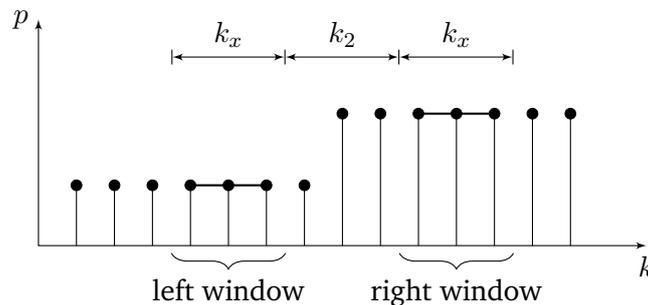
Fig. 5.3: system model of the event detection

where  $P_{\text{thr}} \in \mathbb{R}_+$  is the power threshold. If not stated otherwise,  $P_{\text{thr}} = 15 \text{ W}$  is chosen as a reasonable value.

The window from which the derivative is computed is parametrized by three values  $k_1, k_2, k_3 \in \mathbb{N}$  and two weights  $w_1, w_3 \in \mathbb{R}$ . When expressed as a discrete-time filter impulse response, this results in

$$\vec{h}_d = \left[ \underbrace{-\frac{w_1}{k_1} \quad \dots \quad -\frac{w_1}{k_1}}_{k_1} \quad \underbrace{0 \quad \dots \quad 0}_{k_2} \quad \underbrace{\frac{w_3}{k_3} \quad \dots \quad \frac{w_3}{k_3}}_{k_3} \right]^T. \quad (5.10)$$

For the sake of simplicity,  $k_1 = k_3 = k_x$  is assumed for all further analyses. A graphic example is given in Fig. 5.4.

Fig. 5.4: proposed event detection scheme for  $k_x = k_2 = 3$ 

A differentiating filter has a high-pass characteristic, which means that it lets high frequency components pass while attenuating lower frequency components. For  $(k_x, k_2) = (1, 0)$ , the filter is a simple differentiator and has a high-pass frequency response, as shown by the solid curve in the left subplot of Fig. 5.5. As  $k_x$  grows, the averaging character becomes more dominant, transforming the filter to stronger resemble a low-frequency band-pass (dash-dotted and dotted curves), which suppresses high frequency noise. The Prewitt operator ( $k_x = k_2 = 1$ ) results in a symmetric band-pass with center frequency  $\Omega/2$  (dashed curve).

From the step responses in the right subplot of Fig. 5.5, it can be concluded that two corrections have to be applied to the results when a filter with the transfer function defined in eq. (5.10) is used as an event detector.

**Correction 1:**  $k_2$  and  $k_x$  produce an offset to the event index, so the results need to be corrected by subtracting  $k_x + \frac{k_2}{2}$ .

**Correction 2:** the averaging over  $k_1 + k_x$  values results in  $d[k]$  to be of a piecewise constant type, rather than just showing peaks. This means that after each step from 0 to 1 or 0 to -1, the following  $k_1 + k_x - 1$  indexes have to be deleted, or otherwise there will be a lot of false positives in the event detection results. The piecewise constant nature of the step response for  $k_x > 1$  means that  $k_x$  is the limit to the minimum time that has to pass between two consecutive events. If two events fall into a time slot of  $k_1 + k_x$  samples, they can only be detected as one event.

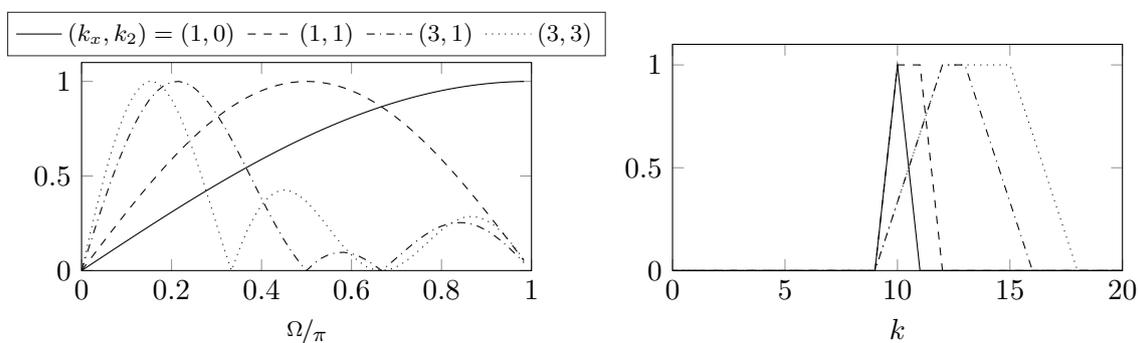


Fig. 5.5: filter frequency responses for different parametrizations ( $w_1 = w_3 = 1$ )

It is important to note, that all filters have no zero-frequency gain, i.e. no integrating character. This is very important and will be shown next. Taking the Prewitt operator as a basis, and modifying either *only*  $w_1$  or *only*  $w_3$ , is another way of designing the filter ( $\vec{h}_d$  becomes asymmetric in this case). The frequency responses for some values of  $w_3$  are shown in the left subplot in Fig. 5.6.  $w_1$  is fixed to a value of 1. It can be seen that, with growing  $w_3$ , the filter more and more loses its attenuation of the low and high frequency components and will eventually become an all-pass filter. Although this might be interesting for some applications, here it has to be kept in mind that with growing  $w_3$ , the integrating character is becoming dominant over the differentiation because of the non-zero zero-frequency gain. This can be clearly seen by the step responses shown in the right subplot in Fig. 5.6. It can be concluded that this kind of parametrization is not suitable for the event detection process.

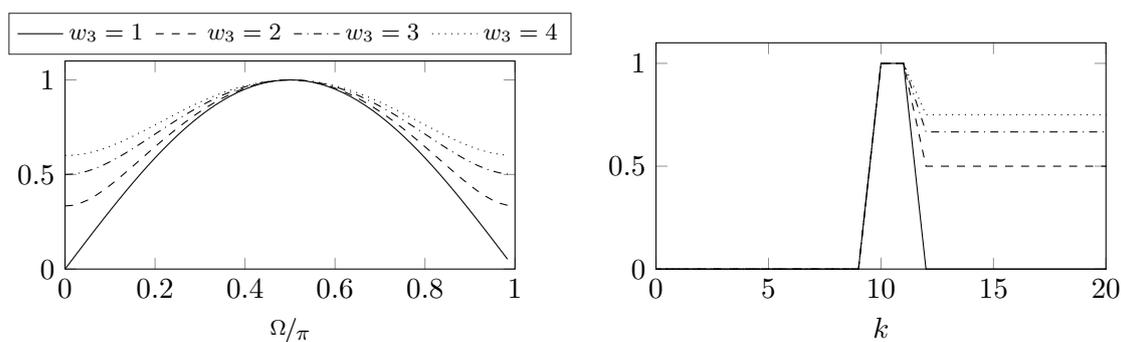


Fig. 5.6: filter frequency responses for different parametrizations ( $k_x = k_2 = w_1 = 1$ )

### 5.2.3 Parameter Optimization using Test Signals

In this section, the optimal parameters for the event detector proposed in section 5.2.2 are sought using test signals. The test signals are sequences of 100 zeros followed by 100 ones (rectangular sequence), corresponding to 2 s on and 2 s off at a data rate of 50 Hz.

#### 5.2.3.1 Noise Sensitivity

The test signal is added normally distributed noise ( $n'_m(t)$  in Fig. 5.3b)) with given signal-to-noise ratio (SNR) values. SNR and noise variance  $\sigma_m'^2$  are related via

$$\text{SNR} = \frac{\frac{1}{N} \sum_{\ell=0}^{N-1} p[\ell]^2}{\sigma_m'^2}, \quad (5.11)$$

where  $N$  is the number of samples in  $p[k]$ . As the event detection tends to be oversensitive, i.e. due to the high-pass characteristic it will output too many events rather than too few in the presence of noise, this test should only affect the false positive rate. For low SNR values it is very probable that it will find many of the ground truth events by chance. Indeed, for all simulation runs, the true positive rate is  $\geq 95\%$  for each parameter step. The false positive rate, on the other hand, highly depends on the SNR, as is shown in Fig. 5.7 for different values of the averaging window length  $k_x$ . At  $k_x = 1$ , the false positive rate never falls below 100% in the analyzed range of SNR, i.e. the algorithm is by far too sensitive without averaging. For values  $k_x > 1$ , the FPR behaves as to be expected: as the SNR grows it eventually goes down to 0. This shows the low-pass influence, that is achieved by a  $k_x > 1$ , which suppresses high frequency noise.

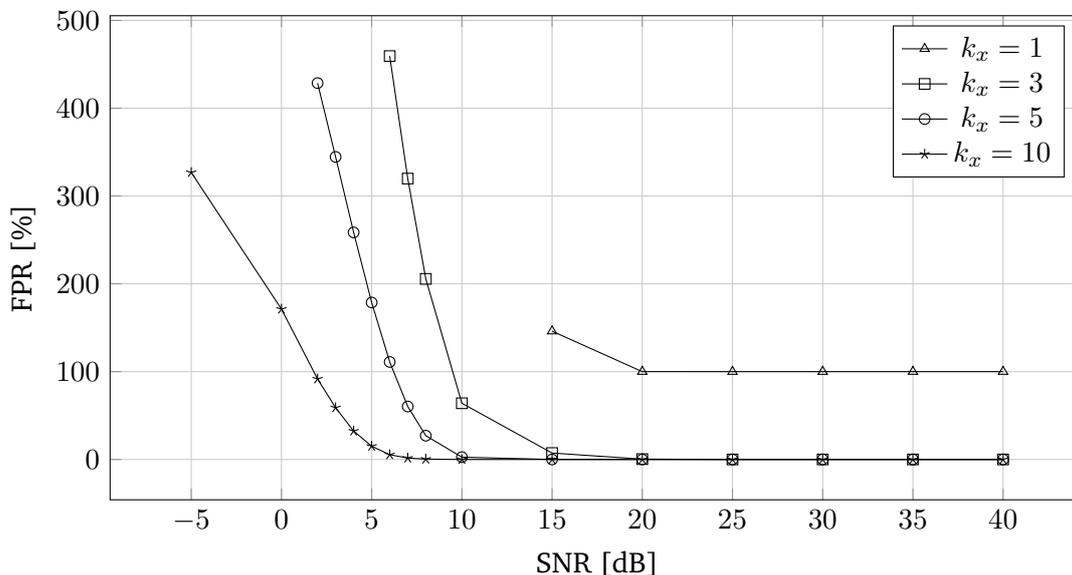


Fig. 5.7: sensitivity of the proposed WDE detector to noise (expectation of 100 runs with 200 events, each),  $k_2 = 3$ ,  $w_1 = 1$

#### 5.2.3.2 Averaging Window Length Sensitivity

In the previous section the data showed that a large value of  $k_y$  suppresses noise and reduces the FPR. There is an upper limit though, when  $2 \cdot k_x$  reaches the time between consecutive events. For the test signal, this time is 100 samples. If the SNR is fixed to a value such that optimal detection is possible (15 dB seems a good guess from the solid line marked by circles in

Fig. 5.7), this can be verified: as soon as  $k_x$  reaches 50, the solid TPR line in Fig. 5.8 starts to decrease. It can also be seen, that the solid FPR line is not linear with respect to  $k_x$ , a fact that was not immediately evident in Fig. 5.7. Another interesting finding is, that the parametrization  $k_x = k_2$  is an unfortunate choice. The dashed FPR line does not drop as quickly and the dashed TPR line degrades even worse for  $k_x > 50$ . When  $k_x = k_2$  is chosen, the overall length of the detection window increases faster than a fixed value for  $k_2$  with increasing  $k_x$ . This also increases the chances that unwanted signal parts are included in the detection window.

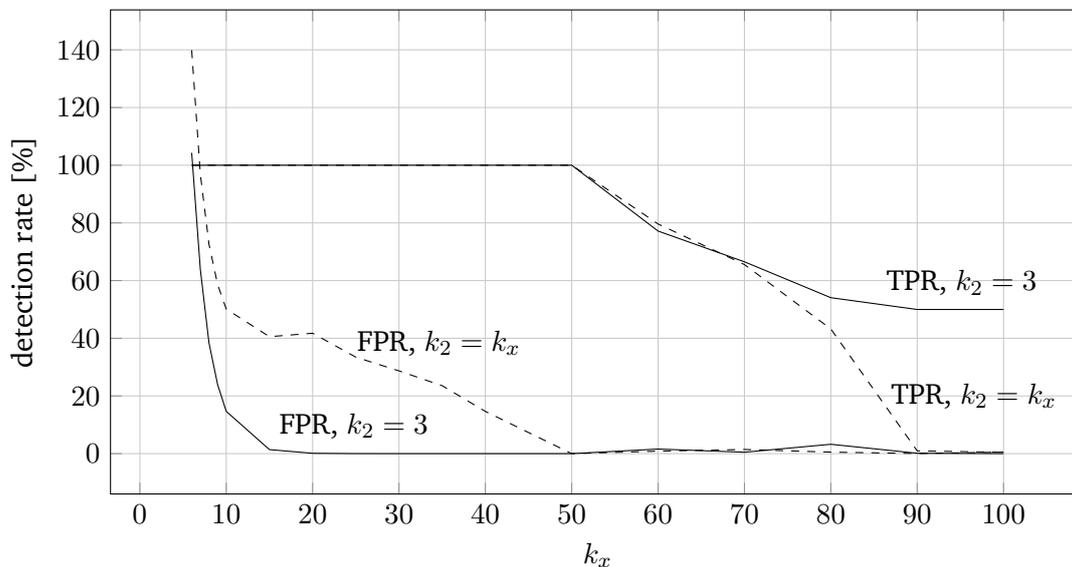


Fig. 5.8: sensitivity to the value of  $k_x$ , SNR = 5 dB

#### 5.2.4 Parameter Analysis using Real Signals

When using real measured signals, the lower bound of the SNR is fixed. Here, the threshold value and  $k_x$  are parameters to analyze, because they depend on the different power steps in the measurements. A full parameter search is required here, because the measured signal is not as well defined as the test signals before. Also, the test signal results made it clear that there is no single optimal point for the parameter search when either the TPR or the FPR are chosen as optimization parameters. This behavior has to be expected for measured signals as well, and can be shown by evaluating the detection rates for a real data file. Fig. 5.9 shows the TPR for the ReSP data set I in the upper subplot, and the FPR in the lower subplot. The TPR reaches its maximum for the smallest value of the threshold, which is intuitively correct because this yields the most sensitive event detector. The  $k_x$  parameter does not have a big impact on the detection rate. Unfortunately, though, the FPR also reaches its highest values around small values of the threshold and  $k_x$ . This shows that a compromise between true positives and false positives must be found. They cannot be tuned independently.

Based on these analyses, the choice  $P_{\text{thr}} = 15$  W for the threshold parameter is confirmed. Therefore, it is also chosen for all following performance comparisons. Furthermore,  $k_x = 15$  and  $k_2 = 3$  are chosen for the time parameters.

### 5.3 Event Detection using Clustering

The method in section 5.2 assumes, that an event is represented by a steep change in power. This means that potentially any edge or power change may be classified as an event, leading to false positives. Streubel and Yang proposed a radically different approach. Their event detector tries to divide the power signal into transient and steady state sections [86]. An event is represented

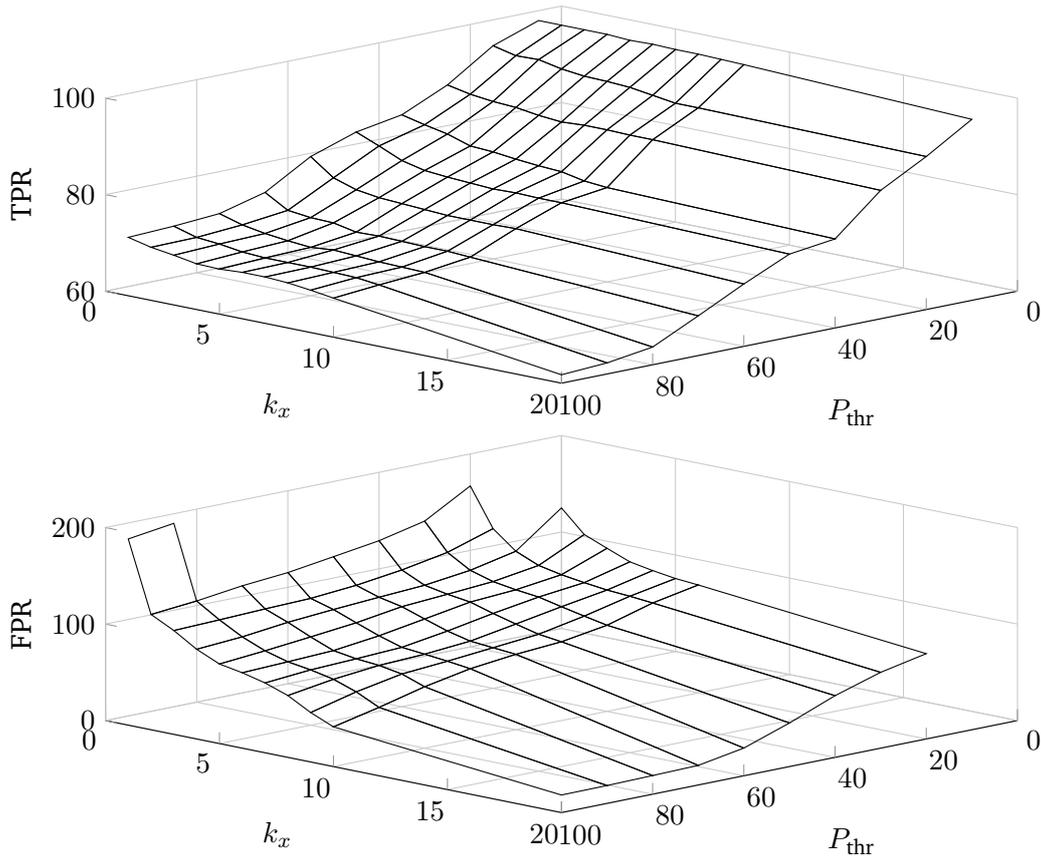


Fig. 5.9: full parameter search for data set I, file 10, with  $k_2 = 3$

by the transition from one steady-state period to the other. This is supposed to result in a more robust event detector, if the detection of steady states is reliable. The idea for this is to perform a cluster analysis of a measurement in the  $PQ$  plane. Clusters are assumed to be normally distributed, so GMMs are estimated from a measurement of  $\vec{P} \in \mathbb{R}^N$  and  $\vec{Q} \in \mathbb{R}^N$  containing  $N$  values, using the EM algorithm. Afterwards, the vector  $\vec{v} \in \mathbb{N}^N$ , which indicates the cluster index for each sample, is computed. In order to detect steady states, the authors compute the “second order difference vector” [86]  $\vec{w} = |\vec{v}''|$ . The binary vector

$$\vec{m}[k] = \begin{cases} 1, & \vec{w}[k] = 0 \\ 0, & \vec{w}[k] > 0 \end{cases} \quad 1 \leq k \leq N \quad (5.12)$$

indicates steady state sections ( $\vec{m}[k] = 1$ ) and transient sections ( $\vec{m}[k] = 0$ ). For the final result, it is filtered such that it does not contain single ones, i.e.  $[\dots 0 1 0 \dots]^T \mapsto [\dots 0 0 0 \dots]^T$ . Note, that the model order (i.e. the number of states to expect) needs to be estimated prior to clustering.

A major disadvantage of this method is that it requires a considerable amount of training data in order to get a reliable clustering result. This might be a limiting factor, especially on low-power hardware. Furthermore, if, during the normal operation phase, a new appliance is added to the household, a new training phase might be necessary to update the clusters. The estimation of the model order is essential. If the feature space is divided into too few clusters, some transitions might be missed. If, on the other hand, there are too many clusters, the algorithm might identify too many transitions (false positives due to overfitting).

## 5.4 Event Detection using Steady State Approximation

In this chapter, the idea of chapter 5.3 to divide the power signal into transient and steady-state periods is picked up. The goal is to design an algorithm that shrinks the memory footprint, is applicable to real-time applications, and allows for a convenient implementation on low power computing equipment. In section 5.2, the event detection was done by analyzing signal segments of constant length  $2 \cdot k_x + k_2$ . On the one hand, the segment length had to be rather short in order to react to events following each other within a short period of time. On the other hand, this yielded a too sensitive algorithm when there are spikes or ramps in the signal instead of perfect rectangular structures.

The proposed method is called Steady State Approximation (SSA). It is inspired by the way humans would partition a power signal: when viewing the signal in a plot we divide it into blocks of low slope (constant power) and “everything else” (i.e. the transient periods). The goal is therefore to divide the power signal  $p[k]$  into arbitrarily long segments, in which the trend of the power remains under a given threshold. To do this, the computation of the derivative is adapted to the currently present signal<sup>32)</sup> by defining

$$d[k] = p[k] - p[k - k_{st}[i]], \quad (5.13)$$

where  $k_{st}[i]$  is the starting index of the  $i$ -th segment.  $d[k]$  is computed according to eq. (5.13) every time a new value of  $p[k]$  is recorded. If the value of  $d[k]$  exceeds a (predefined) threshold  $d_{thr,1} \in \mathbb{R}$  at time  $k_0[i]$ , the period between  $k_{st}[i]$  and  $k_0[i]$  is marked a steady state. The algorithm then turns into transient state observation mode, which means that  $d[k]$  is observed until it falls below a second threshold  $d_{thr,2} \in \mathbb{R}$ . This time instance is  $k'_0[i]$ . It marks the beginning of a new steady state, so

$$k_{st}[i + 1] = k'_0[i]. \quad (5.14)$$

To find out whether an event was rising or falling, the average power

$$P[i] = \frac{1}{k_0[i] - k_{st}[i]} \cdot \sum_{\ell=k_{st}[i]}^{k_0[i]} p[\ell] \quad (5.15)$$

is computed for each steady state  $i$ . If the difference  $P[i + 1] - P[i]$  is positive or negative, the  $i$ -th event is marked rising or falling, respectively.

The simplicity of the algorithm is revealed in an example for an on-line implementation. See Listing 5.2 for the pseudocode. The implementation can straightforwardly be done in a microprocessor requiring only four scalar variables, and one matrix for the storage of the steady state blocks. The initialization in Listing 5.1 needs to be done once at the start of a running period.

Listing 5.1: initialization of the SSA algorithm

```

1 mode = steady state observation % detection mode
2 k_st = 0 % steady state block start index
3 k_0 = 0 % steady state block end index
4 p_st = p[0] % block start power
5 blkTbl = [] % list (matrix) of steady-state blocks

```

<sup>32)</sup>Here the on-line processing is discussed. The method can also be implemented to run off-line, which means that the time index  $k$  has to be replaced by a window size  $w[k]$ .

Listing 5.2: SSA processing steps for every new value of  $p[k]$ 

```

1   $d = p[k] - p_{st}$  % update the slope
2
3  if mode == steady state observation
4      if  $d > d_{thr,1}$ 
5          mode = transient state observation
6           $k_0 = k$ 
7           $p_{st} = p[k]$ 
8          Add  $[k_{st} \ k_0]$  to blkTbl
9      end
10 else if mode == transient state observation
11     if  $d < d_{thr,2}$ 
12         mode = steady state observation
13          $k_{st} = k$ 
14          $p_{st} = p[k]$ 
15     end
16 end

```

### 5.4.1 Correction for Settling Processes

Slowly settling processes and sequences of peaks may result in several steady states after each other, although a human would consider the whole ramp to be only one state. Therefore, the events need to be post-processed. If, for two adjacent steady state blocks  $i$  and  $i + 1$ , the power difference  $|p[k_{st}[i + 1]] - p[k'_0[i]]|$ <sup>33)</sup> is smaller than a threshold  $P_{thr}$ , i.e. the power signal in the transient area between the two blocks is continuous, the two blocks are merged. A graphic example is given in Fig. 5.10. Without the filtering, the two turn-on settling processes at  $0.5\text{ s} < t < 4\text{ s}$  and  $4.5\text{ s} < t < 8\text{ s}$  are split into multiple blocks. With the post-processing, these short blocks are merged into two large ones.

### 5.4.2 Discussion

In contrast to the non-adaptive algorithm in section 5.2.2, the SSA adapts itself to the measured signal. The combination of adaptive derivative and post-processing is a highly non-linear noise suppressing filter, but may also miss events of low step height. It can be expected that the number of false positives is drastically reduced for the kind of signals plotted in Fig. 5.1c) and 5.1d). The SSA approach does not require a real parameter optimization procedure. Its window lengths are automatically adapted to the signal. The threshold  $P_{thr}$  can be set a priori for all systems. The implementation in real-time applications is straightforward because neither does SSA require any pattern recognition libraries, nor appliance specific models which have to be trained in advance. It is solely based on an adaptive derivative of the power signal.

## 5.5 Results

Three event detection approaches were described in the previous three sections. They are now tested and compared in three different scenarios. The general validity of the event detection is tested in subsection 5.5.1 on three different appliance types. The detection of high power loads for hazard monitoring is tested in subsection 5.5.2. A general comparison based on a large set of data is given in subsection 5.5.3.

<sup>33)</sup>Note that this is not the difference of the steady state powers. It is the difference of power signal  $p[k]$  evaluated at the first value of the block  $i + 1$  and the last value of the  $i$ -th block.

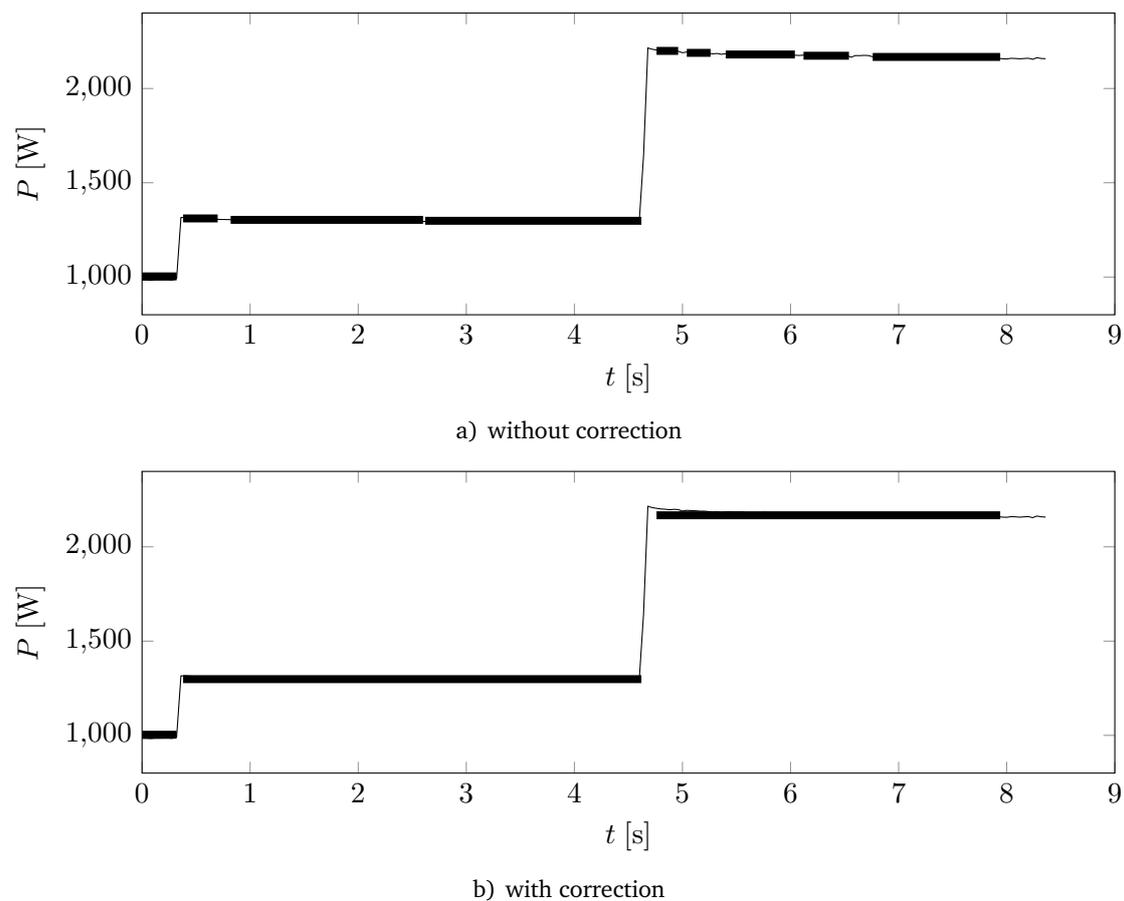


Fig. 5.10: effect of the correction of settling processes on the SSA event detection (indicated by horizontal bars)

### 5.5.1 Test for Generality

The intention of the event detection chapter was to find a method that is able to detect as many true switching events as possible without requiring specialized models for certain appliances. Additionally, a low false positive rate is required for an event detection algorithm to be acceptable in Activity Monitoring applications. In this subsection, this ability is tested using the measurements contained in data set II. The division of the data set into three subsets allows for a diagnostic test under different conditions:

- II-1:** this contains switching events of an electric room heater. Here, a high-power appliance with fluctuating power consumption due to temperature changes is represented.
- II-2:** here an incandescent lamp rated 40 W was measured. This represents a low-power load with a power consumption close to the desired detection threshold (15 W).
- II-3:** this subset contains 292 events of a freezer. The turn-on events of freezers are more complex than simple two-state loads, which is a challenge for event detectors, especially when analyzing false positives. The steady-state period is shortened to get more events into the measurement.

An example of one operation cycle is given in Fig. 5.11 for each subset.

The true positive rates achieved by the three event detectors are shown in Fig. 5.12a). They perform almost equally, with a small disadvantage for the SSA method. The differences of the FPR performance indicated in Fig. 5.12b) are much more distinct. Here, both the clustering and the SSA algorithm outperform the WDE approach, with a slightly better overall performance of

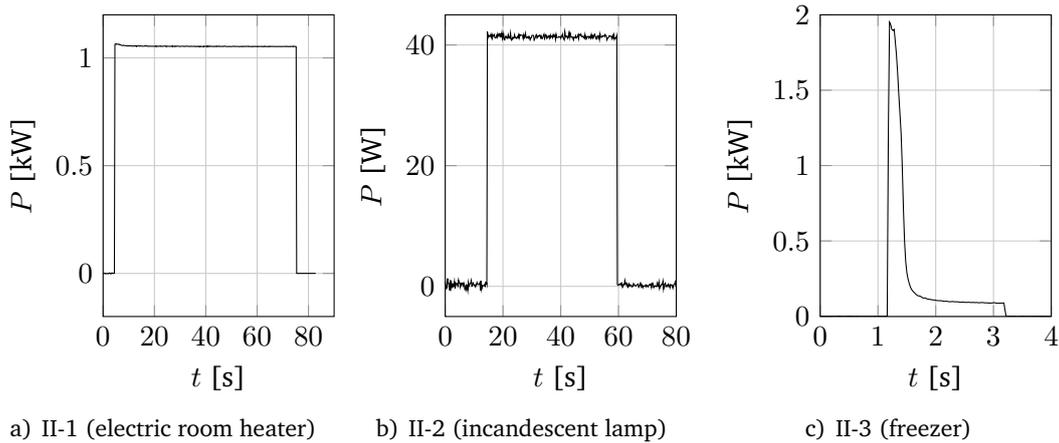


Fig. 5.11: examples of the appliance power consumption pattern in the three subsets of data set II

the SSA event detection. The high false positive rate produced by the WDE method shows that the short window size of the derivative and the transient slope as a feature characterizing an event are prone to misdetections. Although they miss some events, the methods incorporating more knowledge of the signal (cluster information, variable window size for SSA) are much more robust.

Best results are obtained for the II-2 (incandescent lamp) subset in all three cases. The lamp consumes a little amount of power (approximately 40 W), so the turn-on and turn-off transients are very short. This reduces the risk of multiple events being detected at each transient. Its settling period, which could also lead to false positives, is very short, because its filament heats up almost immediately. This can be shown by the freezer measurement (II-3). Here, all methods produce the highest FPR. This is not surprising, because a freezer has a very distinct settling behavior with usually a short steady-state period before reaching its long-term power consumption. False positives are the result. The freezer (or in general cooling devices) is an example where a type specific model of the settling behavior would decrease the false positive rate.

The electric room heater (II-1) shows a distinct exponential settling phase, which is sometimes detected as false positive. Here, the methods which consider some knowledge of the signal's past or statistics are better, because they assign the settling phase to the following steady-state phase.

Overall, the analyzed event detection methods were able to detect more than 95% of the ground truth events without requiring appliance specific models. The number of false positives is a critical factor in AAL applications. Hence, the proposed SSA method is favorable in this use case.

### 5.5.2 Test of the Hazard Detection

For a detection of hazards, it is important to reliably detect appliances consuming large amounts of power. In this section, the superposed measurements of ReSP data set III are used. As described in section 4.8.3, this contains switching events of high power appliances. The achieved true positive rate is shown in Fig. 5.13. It can be seen that all three event detection approaches perform comparably well with detection rates above 95%, so the detection of high power appliances is not a problem under laboratory conditions.

As real-world example, the BLUED data set is used to test whether events of appliances with a power consumption of 200 W or more are detected correctly. To do this, the ground truth events are filtered for the appliance IDs listed in Table 5.1. Those events are expected to lead to a power change of more than 200 W. Fig. 5.13 shows the true positive rate, that is achieved by comparing the detected events with these filtered ground truth data. The false positive rate is

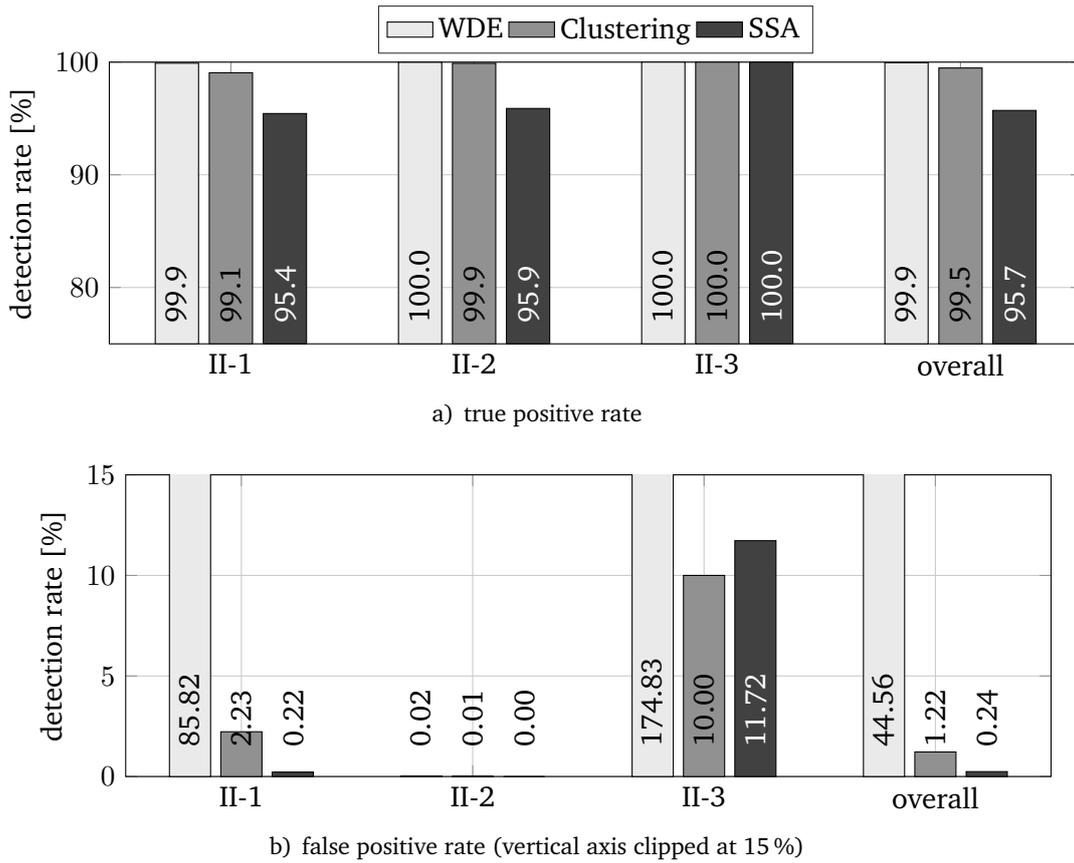


Fig. 5.12: event detection results for data set II

not significant here, because not all ground truth events are considered for the evaluation. On phase A, both the WDE and the SSA method detect all events correctly, the clustering method misses some of them. Of the events on phase B, approximately 8% are not detected. Those are caused by the garage door (ID 103). It is rated 530 W in the appliance table. Evaluating the power signal, its actual consumption was found to be at around 150 W. Once the events of the garage door are removed from the ground truth events, the true positive rate increases to 100% for all three methods. Since the garage door does not actually contain a heating element, this is an acceptable restriction. All in all, the three algorithms proved to be successful in detecting all events that are considered an indication of possible endangerment of residents.

Table 5.1: selected appliances from the BLUED data set for high power events

id	appliance	rated power	# events	phase
103	Garage Door	530 W	24	B
108	Kitchen Aid Chopper	1,500 W	16	A
127	Air Compressor	1,130 W	20	A
132	Hair Dryer	1,600 W	8	A
134	Iron	1,400 W	40	B
Sum			108	

### 5.5.3 Performance Comparison

To get a more profound impression of the algorithms' performance, they are evaluated on a combination of data sets I, II, and III, totaling 37,492 events, and on the BLUED data set with

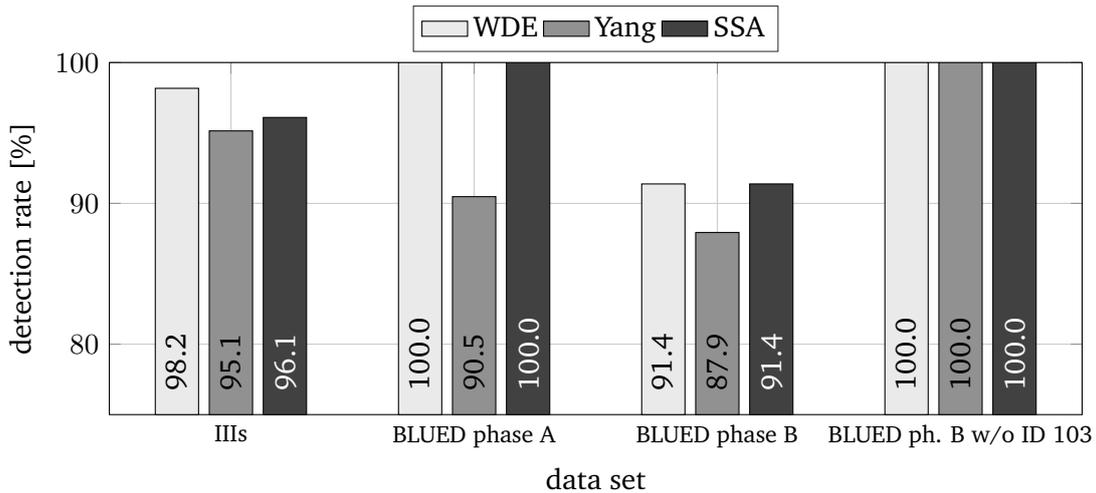


Fig. 5.13: true positive rate for the detection of high power appliances in the data sets IIIs and BLUED

2,418 events. The detection rates are visualized in Fig. 5.14. Best TPR results are achieved by the WDE algorithm for both data sets. Again, this is due to a very sensitive detector using a small window size. The price for the high true positive rate is a likewise high FPR which makes this kind of event detection too sensitive for AAL requirements.

With both data sets, the TPR performance of the other event detection methods is almost equal. The false positive rate results draw a different picture. Here, the SSA event detection performs significantly better in reducing the FPR by orders of magnitude. The clustering method highly depends on the estimation of the number of clusters in the data set (cf. section 5.3). The results presented here suggest that further optimization is possible, though it is unlikely that the false positive rate can be decreased down to the levels of the SSA.

It should be noted here that the false positive rate reaches absurd values for the BLUED data set. This is due to distortions that can be found on phase B. They introduce a lot of intermediate (unknown) steady-states and switching events, which forces up the number of false positives. Some of these events are attributed to the fact that the BLUED reference data only contains switching events for whole appliances, not for their dominant components. To some degree, the data set violates the assumption made in section 4.2.1.

## 5.6 Conclusion

Two state-of-the-art event detection methods incorporating a generally valid appliance model for NILM were described in this chapter. The first, called WDE, detects switching events by searching for rapid changes in the power consumption. On the one hand, this is an approach that leads to very good detection of true events, even for small power changes. On the other hand, the high sensitivity automatically leads to an amount of false positives that is unacceptable for AAL. The second method was proposed by Streubel and Yang [86]. It divides the signal into transient and steady-state sections using a clustering technique. This approach is much less likely to produce false detections, while maintaining a high rate of detection of actual events. However, the rate of false positives leaves room for improvement. A drawback of this approach is that it requires a considerable amount of training data which raises the hurdle for implementation.

Based on the idea to divide the signal into transient and steady-state sections, a new method called SSA was introduced. It is based on the observation of the signal's slope. Hence, it does not require training data. Its parameters are automatically adapted to the signal. Due to its iterative design, it is convenient to implement in real-time applications while requiring minimal amounts of memory and computation time.

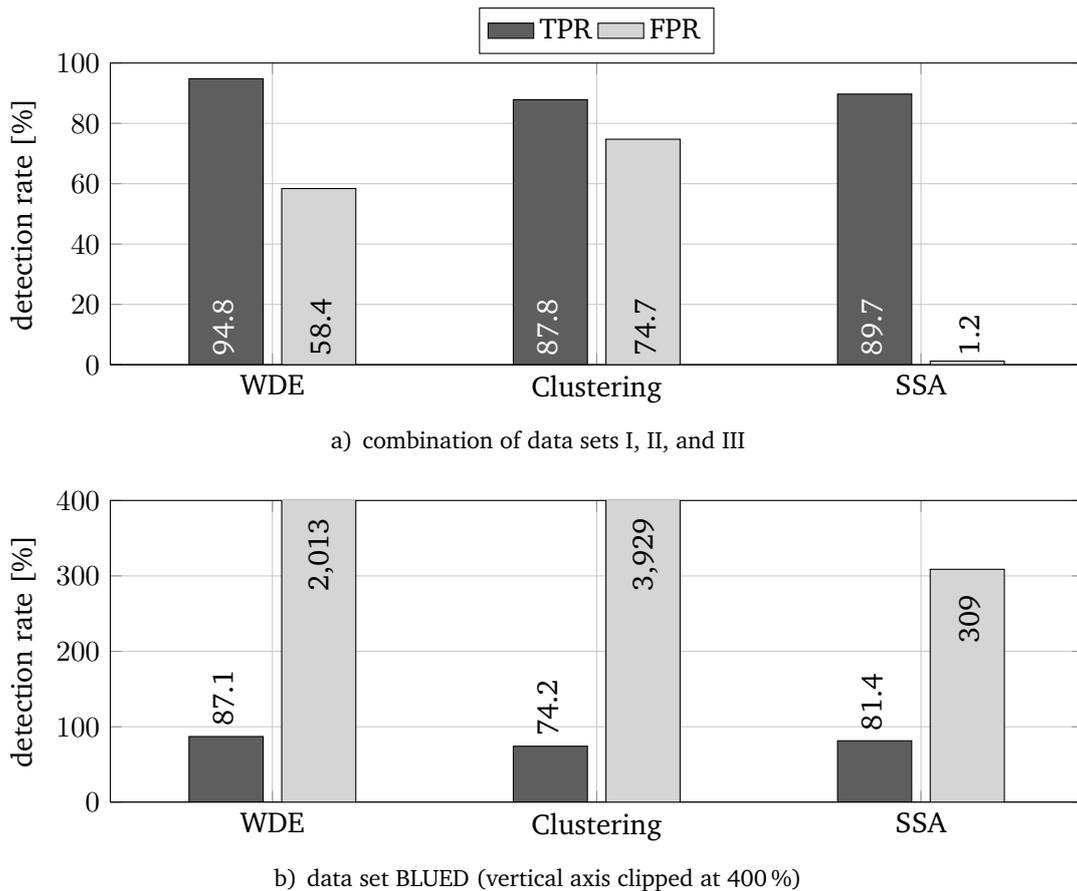


Fig. 5.14: event detection performance comparison on large data sets

The validity of the general appliance model was analyzed based on specialized ReSP measurements. Although applicable in many situations, this model might require a downstream identification of some special appliances to correct the event detection results.

When used for the detection of switching events of appliances that could pose a danger to residents due to heat, all approaches proved to be successful. This was shown with the public data set BLUED.

Finally, the algorithms were compared on the basis of a large number of events, both under laboratory (ReSP measurements) and real-world (BLUED) conditions. While performing similarly with respect to the true positives, the three methods differ greatly when it comes to false positives. Here, the clustering method performs better than WDE. Even more improvement is achieved by SSA, which decreases the false positive rate to levels such that, perhaps with an identification of problematic loads, it could be used for the detection of potentially dangerous appliances and as information source for activity monitoring.



# Chapter 6 | Feature Extraction & Appliance Classification

## 6.1 Introduction

NILM tries to estimate the amount of energy each appliance in a living or business compartment consumes by identifying their active periods and individual power consumption from aggregate data. Hence, the extraction and classification of characteristic features is an essential part. The goal in this chapter is to assess the prospects of successful classification using two different kinds of features. Sections 6.2 through 6.4 address steady state power features (SSPF). Sections 6.5 and 6.6 introduce a new kind of features, so called time-series fingerprints (TSFP). The applicability of these features for many different appliances is evaluated in section 6.7 in form of an appliance fingerprint catalog. A conclusion of the chapter is given in section 6.8.

In order to avoid any errors due to event detection problems, only measurements containing precise ground truth data are used. The test bench is sketched in Fig. 6.1. The algorithms to test are provided with preprocessed smart meter data, as well as a list of precise ground truth event time stamps. The classification results (lists of IDs) are compared with the ground truth data in order to compare the algorithms by computing the AECR, which was defined in section 4.6.2.1.

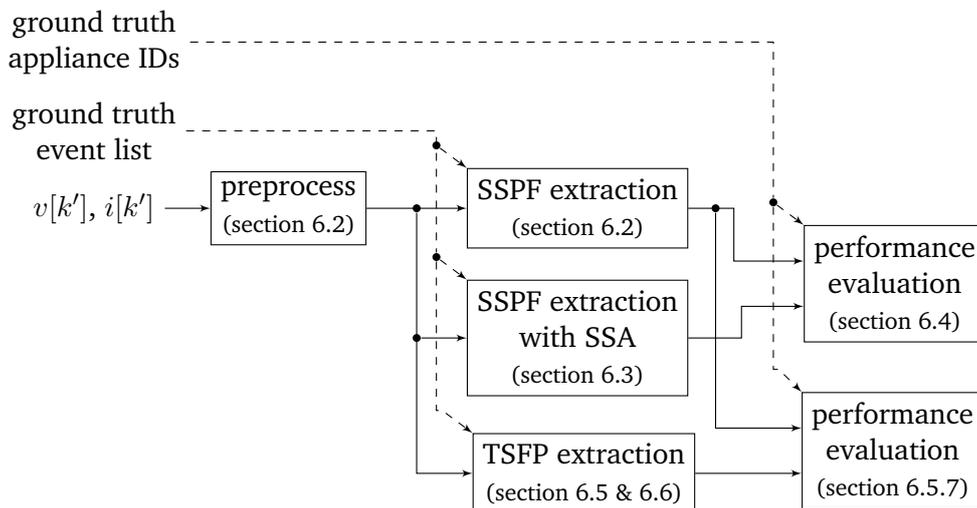


Fig. 6.1: test bench setup for the evaluation of feature extraction algorithms

## 6.2 Steady State Power Features

As mentioned in section 2.3.3, the beginnings of the NILM research field date back to the 1980s. A basic yet powerful approach remains from this time. The assumption is that appliances are turned on at a given moment in time, are being operated for a while, and turned-off in the end. This model is similar to the appliance model that was defined in section 4.2. The power

consumption between the switching events is assumed to be constant (rectangular signal approximation). The difference of power levels at each event is the feature, that is used for load identification. It may be impossible to distinguish loads of the same type using these features. Since this is considered a rare case, it is acceptable for two identical loads to be classified as the same appliance.

Let  $P[k]$  be the real power signal, and let the  $i$ -th event be marked by the index  $k = e_i$ . A SSPF feature is the difference of the power signal after the event  $P[e_i + 1]$ , and before the event  $P[e_i - 1]$ . SSPF features are denoted by a  $\Delta$  prefix:

$$\Delta P_i = P[e_i + 1] - P[e_i - 1]. \quad (6.1)$$

The feature vector  $\vec{f}_i$  of the  $i$ -th event contains the differences of real power  $P$  and reactive power  $Q$

$$\vec{f}_i = [\Delta P_i \quad \Delta Q_i]^T. \quad (6.2)$$

Each appliance is represented by two points in the two-dimensional feature space. One point marks the turn-on events, the other one the turn-off events. If the rectangular signal approximation holds, the points are point-symmetric with respect to the origin. See Fig. 6.2 for an example.

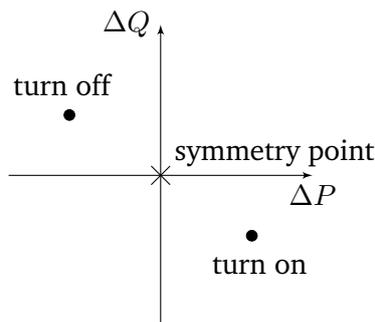


Fig. 6.2: symmetry of the steady state power feature for real power  $P$  and reactive power  $Q$  difference features, when appliances are modeled by steady state constant power consumption

If there are fluctuations in the signals, the points are enlarged to “clouds” or clusters. This means that a data set is good if the following two standard pattern recognition conditions are fulfilled.

1. The clusters of an event type (on or off) belonging to an appliance are concentrated in a small region of the feature space.
2. There are no overlapping clusters, i.e. the distances of the clusters’ centroids is large compared to their spatial extent<sup>34</sup>).

### 6.2.1 Reference Algorithm for the Feature Extraction

The first attempt to load identification using SSPFs used real power  $P$  and reactive power  $Q$ . This two-dimensional feature vector was used because of the problem of appliances having similar real power consumption. The separation of  $P$  and  $Q$  takes the phase difference between voltage and current into account. This is the essentially new information, that was added to the feature space. In contrast to e.g.  $[\Delta I \quad \Delta \varphi]^T$ , the values in the pair  $[\Delta P \quad \Delta Q]^T$  have the same dimension (power), which makes it easier to get an intuitive understanding of the values. Originally, this method was applied to signals sampled at 1 Hz [80]. In this subsection,

<sup>34</sup>“spatial extent” here refers to the volume a cluster takes up inside the feature space.

the original algorithm is described as a reference implementation for further analyses. The description is followed by an optimization of the required parameters.

In order to eliminate errors due to voltage fluctuations, the power signals are normalized with respect to a constant mains voltage

$$\underline{S}[k] = \tilde{S}[k] \cdot \left( \frac{V_n}{V_{\text{rms}}[k]} \right)^2, \quad (6.3)$$

where  $\tilde{S}[k]$  is the apparent power measured by the smart meter.  $V_n$  is the normalization voltage<sup>35</sup>). The reason for the normalization with  $\left( \frac{V_n}{V_{\text{rms}}[k]} \right)^2$  is that the admittance is a quantity that intrinsically eliminates the influence of the mains voltage level (cf. section 4.5.2). With the relation

$$\underline{S} = \frac{V^2}{\underline{Z}} \Rightarrow \underline{Z} = \frac{V^2}{\underline{S}} \Rightarrow \frac{1}{\underline{Z}} = \underline{Y} = \frac{\underline{S}}{V^2}, \quad (6.4)$$

the normalized apparent power

$$\underline{S}[k] = \frac{\tilde{S}[k]}{(V[k])^2} \cdot V_n^2 = \underline{Y}[k] \cdot V_n^2 \quad (6.5)$$

is a quantity free of voltage fluctuation influences. The multiplication with  $V_n^2$  is a step that was introduced to make sure that the normalized apparent power is still measured in power units ( $V \cdot A$ ), which helps to intuitively interpret the values. This last step is not actually necessary for the classification.

Eq. (6.1) can be interpreted as an approximation of the derivative  $\frac{d}{dk}$ , i.e. the derived signal

$$x'[k] = \frac{dx[k]}{dk} \quad (6.6)$$

is approximated by

$$\hat{x}'[k] = x[k+1] - x[k-1] \quad (6.7)$$

and evaluated at time indexes  $k = e_i$ . In order to reduce the influence of fluctuations and noise, the approximation of the derivative (6.1) is modified by introducing three parameters to the algorithm. A feature is the difference of averages of two windows of length  $w_{\text{cl}} \in \mathbb{N}$ . One window is positioned  $k_l \in \mathbb{N}$  sampling values before the event  $e_i$ , the other window  $k_r \in mn$  values after the event. The feature is the difference of right and left window, i.e.

$$\Delta \underline{S}_i = \underbrace{\frac{1}{w_{\text{cl}}} \cdot \sum_{\ell=1}^{w_{\text{cl}}} \underline{S}(e_i + k_r + \ell)}_{\text{right window}} - \underbrace{\frac{1}{w_{\text{cl}}} \cdot \sum_{\ell=1}^{w_{\text{cl}}} \underline{S}(e_i - k_l - \ell)}_{\text{left window}}. \quad (6.8)$$

A graphical explanation is given in Fig. 6.3.

The parameters  $k_l$  and  $k_r$  are allowed to be set up independently because of the turn-on dynamics in the appliance model (section 4.2). A suitable, independent choice of parameters should help to eliminate variations in the features by shifting the averaging windows to stable steady state sections. The search for optimal parameters is the topic of the following subsection.

<sup>35</sup>Note that  $\underline{S}[k]$  is a time-dependent RMS signal. Consequently,  $V_n$  is the RMS of the idealized mains voltage.

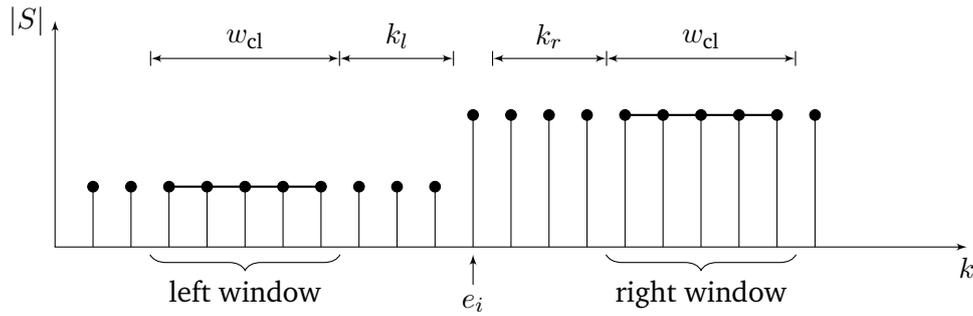


Fig. 6.3: illustration of the parameters for the extraction of SSPFs

## 6.2.2 Parameter Analysis

When computing the features according to eq. (6.8), there are the three parameters  $k_l$ ,  $k_r$ , and  $w_{cl}$ , that need to be set up. To get an impression of reasonable ranges, the kNN classifier is applied to the extracted feature data, because it does not need to extract statistics from the training data. The search for optimal parameters is based on data set I<sup>36</sup>). Due to the fact that there are event time differences  $< 1$  s, the 50 Hz RMS signals are chosen for the analyses instead of the 1 Hz sampling rate, which was originally used by Hart [80]. Like in [80], features are the differences of real power  $\Delta P$  and reactive power  $\Delta Q$ .

Starting point of the optimization is the parameter set given in Table 6.1. It is chosen according to the event time differences of data set I. The features are classified using these parameters to get a benchmark result. The classifier is trained with the data of file 1 (100 events per appliance). Fig. 6.4 shows the AECR for each of the 10 files, using 20 neighbors and a weight function  $w_i = \frac{1}{d(\vec{x}, \vec{y}_i)}$ . The classification rates point out two interesting facts.

1. **Test with file 1:** this means that training and test set are equal. In theory, if all features were perfect, the classifier should be able to achieve AECR = 100 %. In the present case, the test result is very close to the optimum, so the kNN classifier is well suited to label the data.
2. **Test with files 2 to 10:** when testing with the other files (training  $\neq$  test), the results are similar for all of them. This means that the quality of the features is consistent throughout the files. The average rate of 76.6 %, which is considerably lower than the optimum, means that there remain uncertainties in the ground truth data that cannot be corrected. On the other hand this could also mean that not all clusters are congruent with the test data, although they were acquired under the same conditions. This will be revisited in section 6.3.

Table 6.1: starting point parameters

parameter	samples	time units
$k_l$	10	200 ms
$k_r$	10	200 ms
$w_{cl}$	5	100 ms

### 6.2.2.1 Optimization of the Feature Extraction Parameters

For optimization of the feature extraction parameters, the training set is taken from file 1. The kNN classifier is tested with file 10. The three parameters to optimize span a three-dimensional

<sup>36</sup>)As described in section 4.8.1, this data set consists of 10 individual files of the same setting, so it can easily be divided into chunks of 100 events per appliance.

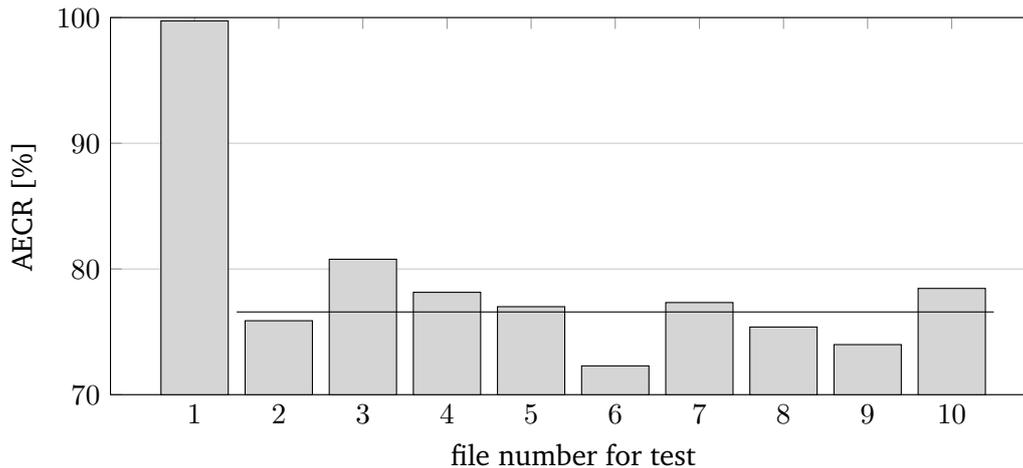


Fig. 6.4: test results with the starting point parameters. The black line marks the average of files 2 to 10

search space. In order to visualize the results, the maximum along each dimension is sought. The results presented in Fig. 6.5 allow for the following conclusions to be made.

1. The room for optimization is very narrow. Maximum and minimum of the AECR curves in Fig. 6.5 are separated by just 4.5%. The features depend only slightly on the parameters in this data set.
2. The features are optimal with short averaging windows. A greater window size takes more of the section 2 dynamics of the appliance model of section 4.2 into account which makes the features instable.
3. The  $k_l$  curve has its maximum at low parameter values. With larger  $k_l$ , the chances of overlapping with adjacent events rise and the classification performance decreases.
4. The  $k_r$  parameter is optimal at higher values than  $k_l$ . This can also be explained by the power dynamics in section 2 of the appliance model. If  $k_r$  is increased, less of the dynamic part is included in the following averaging window and the features become more stable. This is irrelevant for turn-off events because there are no power dynamics to expect. However, the fact that the optimum is found at  $k_r = 800$  ms indicates that the errors due to power dynamics after turn-on events are more dominant. This confirms the necessity for independent time shift parameters  $k_l$  and  $k_r$ .

The optimal parameters from this analysis are shown in Table 6.2.

Table 6.2: optimal parameters for the feature extraction algorithm

parameter	samples	time units
$k_l$	15	300 ms
$k_r$	40	800 ms
$w_{cl}$	3	60 ms

### 6.2.2.2 Optimization of the kNN Parameters

When using the kNN classifier, it may be beneficial to introduce a weight function for computing the dissimilarity measures (cf. section 4.4.3.1), especially if the training clusters are spread over

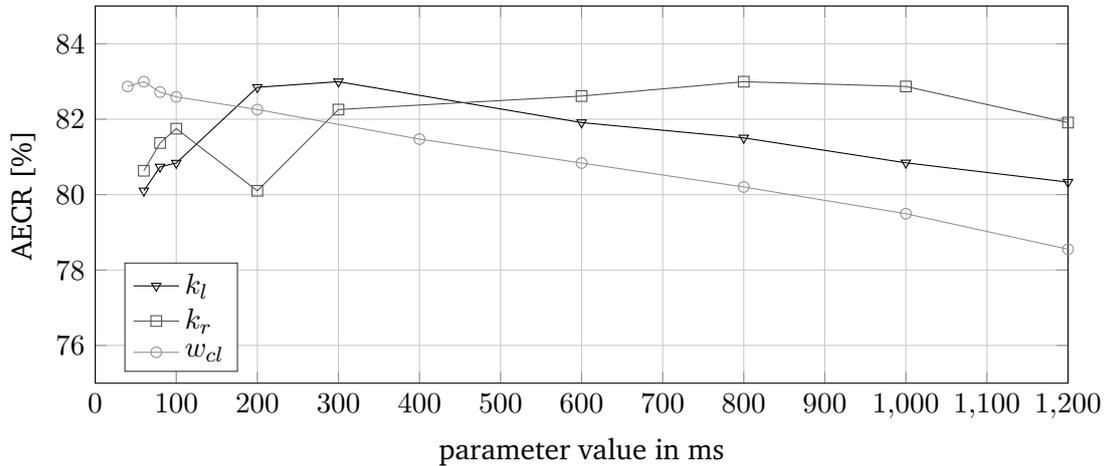


Fig. 6.5: maximum of the AECR along each dimension for different parameter values (data set I; training: file 1; test: file 10)

wide ranges or divided into sub-clusters. In this section, a parametric search is performed in order to find out which weight function to use. The basic function is chosen to be

$$w_i = \frac{1}{d(\vec{x}, \vec{y}_i)^p} \quad p \in \mathbb{N}. \quad (6.9)$$

Using file 1 of data set I for training and file 10 for the test, the AECR is shown in Fig. 6.6 for different values of  $p$  and  $\kappa$ , the number of neighbors. In the range of  $1 \leq \kappa \leq 17$ , there is no evident influence of the weight exponent  $p$ .  $\kappa$  itself does not affect the AECR either in this range. This indicates that there are many outliers or widespread clusters that cannot be classified correctly, even if some neighbors are taken into account. For  $\kappa > 17$ , the weighting with  $p = 1$  or  $p = 2$  yields better results than no weighting at all ( $p = 0$ , so all  $w_i = 1$ ). This is another argument for the assumption that some of the features are instable and spread over a large area of the feature space. This will be verified in the next section.

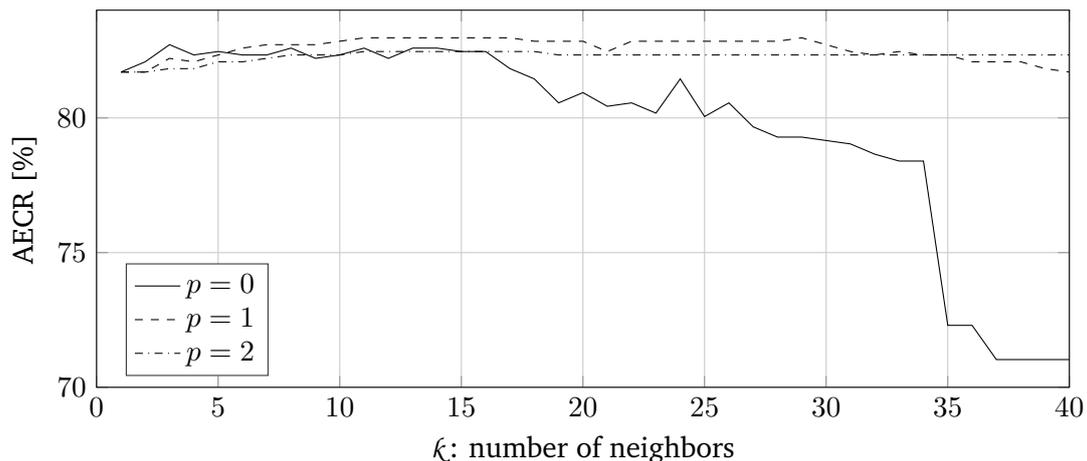


Fig. 6.6: influence of the weighting function and the number of neighbors on the kNN classification results

### 6.3 Feature Extraction with Steady State Approximation

The classification rates shown in Fig. 6.4 raise suspicions about the stability of the features that are extracted by the reference method. If tested with the training data, the classification rate

is close to 100%. If, on the other hand, other data from the same measurement setting is used for testing, the results fall below an average of 85%. The reason for this is that the reference method uses a fixed time difference between detected event and averaging window, given by the parameters  $k_l$  and  $k_r$  in eq. (6.8). An appliance that complies with the 4-section model, introduced in section 4.2, is allowed to have a high degree of power dynamics in section 2, right after the switching-on transient. If, under this assumption, the turn-on event is inaccurate by a small amount of time, the averages, from which the features are computed, may become unpredictable. This increases the cluster variance, and leads to a division into subclusters, increasing the chances of superpositions in the feature space.

A graphic example for the instability is given in Fig. 6.7. Here, only the window after the event is considered. The right average is computed from a window of size is  $w = 4$  at  $k_r = 2$  samples after (right of) the event. For the first event  $e_1$ , the window average is 2.5. The second event  $e_2$  is shifted by two samples to the right. The average drops to 1.5. This illustrates, how a slight inaccuracy of the event index distorts the features computed with the reference algorithm.

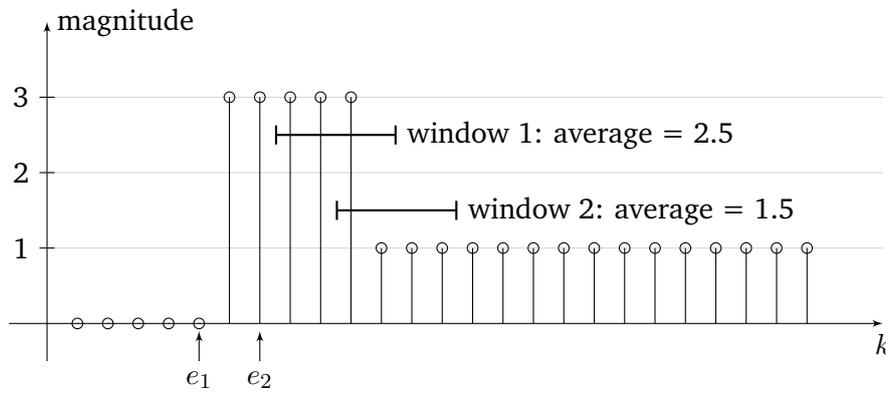


Fig. 6.7: example for inaccurate features due to inexact event positions

In order to stabilize the feature extraction periods, the Steady State Approximation method, which was introduced in section 5.4, is proposed here once again. Here, it is used to find the largest steady-state section between two events. The average of this section is used for the computation of the steady state power features. Theoretically, this should result in more stable features, i.e. a reduction of cluster variance and a better concentration of all features of one appliance.

## 6.4 Results

In this section the standard SSPF feature extraction and the SSA refinement are compared in order to find the best alternative. In subsection 6.4.1, this is done using feature specific quality measures and scatter plots of the feature space. The AECR performance is evaluated in subsection 6.4.2.

### 6.4.1 Feature Analysis

The proximity of clusters in the feature space can be quantified by dissimilarity measures [146]. For a test set  $X$  with  $N_t$  features, the *dissimilarity matrix* is given by

$$\mathbf{D} = \begin{bmatrix} 0 & d(\vec{x}_1, \vec{x}_2) & d(\vec{x}_1, \vec{x}_3) & \dots & d(\vec{x}_1, \vec{x}_{N_t-1}) & d(\vec{x}_1, \vec{x}_{N_t}) \\ 0 & 0 & d(\vec{x}_2, \vec{x}_3) & \dots & d(\vec{x}_2, \vec{x}_{N_t-1}) & d(\vec{x}_2, \vec{x}_{N_t}) \\ 0 & 0 & 0 & \dots & d(\vec{x}_3, \vec{x}_{N_t-1}) & d(\vec{x}_3, \vec{x}_{N_t}) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & d(\vec{x}_{N_t-1}, \vec{x}_{N_t}) \end{bmatrix}. \quad (6.10)$$

Because of the symmetry condition for a dissimilarity measure in eq. (4.7), it is sufficient to compute the elements of  $\mathbf{D}$  for  $b > a$  only, and  $\mathbf{D}$  will contain

$$N_D = \sum_{i=1}^{N_t-1} i = \frac{1}{2} \cdot (N_t - 1) \cdot N_t \quad (6.11)$$

elements which are unequal to zero. Depending on the choice of features, the elements of the dissimilarity matrices do not necessarily have an upper bound<sup>37)</sup>. The lower bound for the elements is 0, indicating that there are identical elements of two subsets. The physical units of the values in  $\mathbf{D}$  may be absurd, if the feature vectors contain elements that are not comparable. For example, the Euclidean distance of feature vectors containing real power, measured in Watts, and the phase difference, measured in radians, will have the unit  $\sqrt{(\text{W} - \text{rad})^2}$ , which has no physical meaning. The cluster proximity analysis, as defined here, can therefore be used to qualitatively analyze the differences between experiment outcomes. The absolute values of the distances are useless.

In order to get a one-dimensional indication of the quality of a given data set, the histogram is computed for all elements of  $\mathbf{D}$  combined. Well confined clusters will be marked by local maxima, because there will be a lot of features with similar distance from the other clusters. Outliers will produce a “noise floor”, i.e. a base line  $> 0$  in the histogram.

Using data set I, the Euclidean distances of all features from file 1 are compared in Fig. 6.8. With SSA, the spikes are indeed higher and there are several regions in which the base level of the histogram is lower. This indicates that some instability of the features is attributed to inaccurate event positions. However, because both curves are similar, the majority of the SSA stabilized features will be the same as those extracted by the reference method. Outliers seem to be in the minority, so they are hard to spot in the dissimilarity plots.

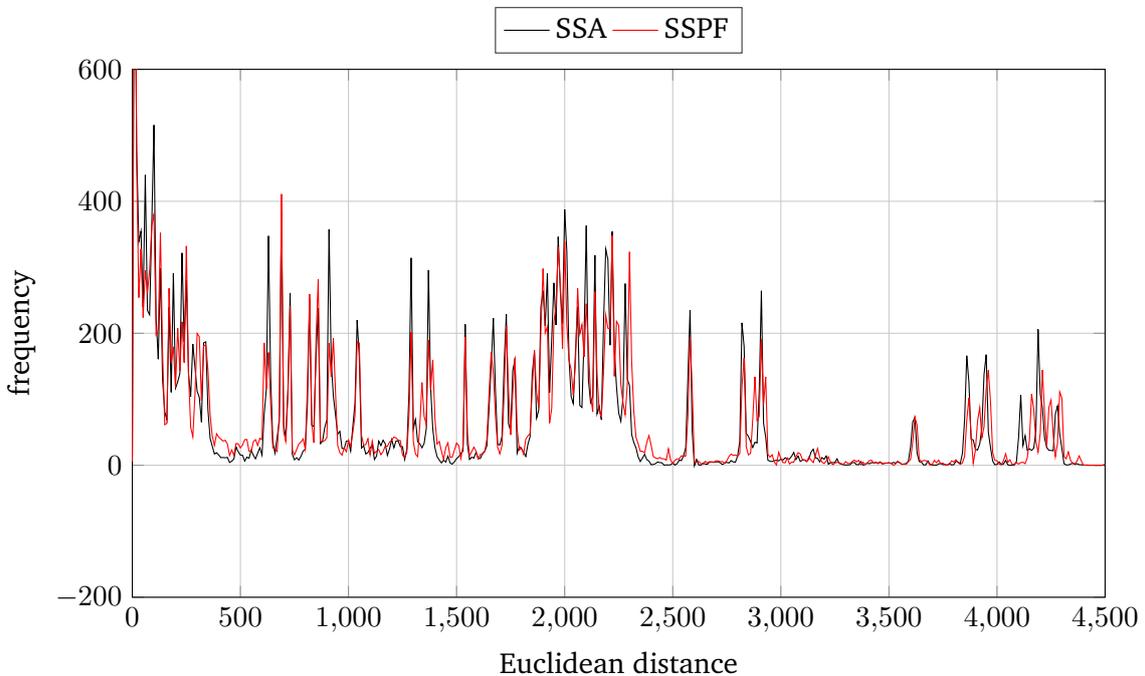
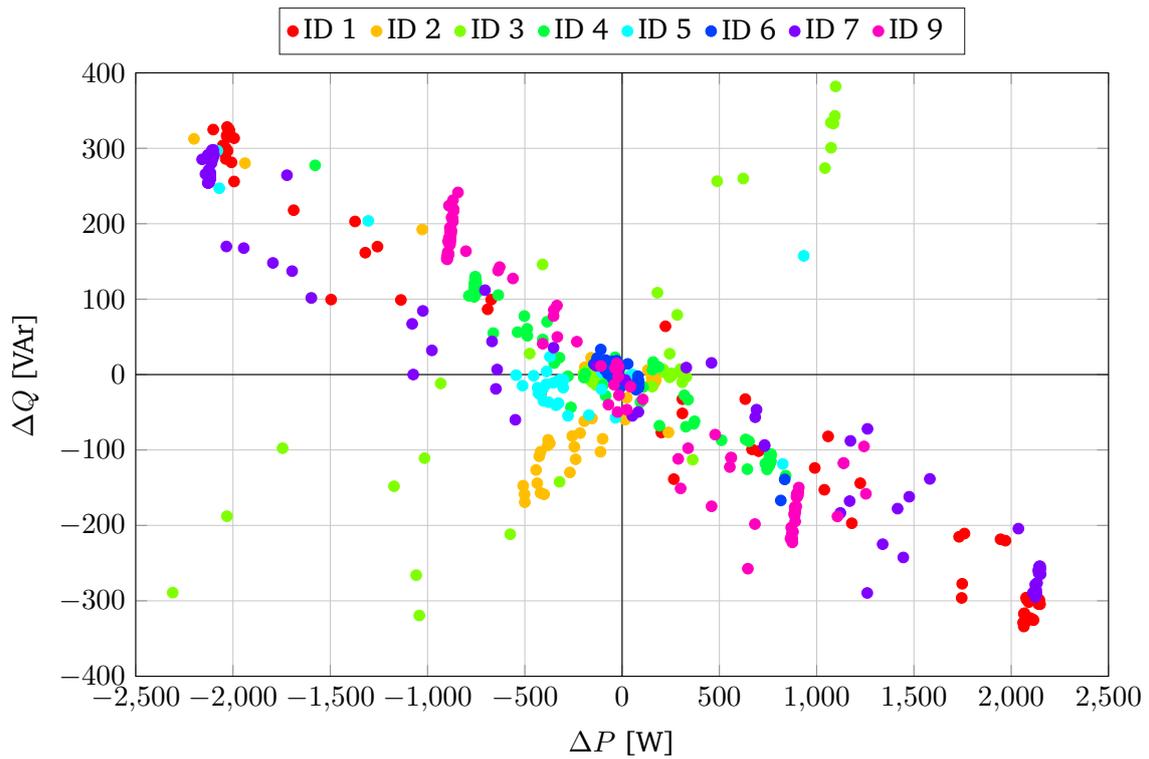


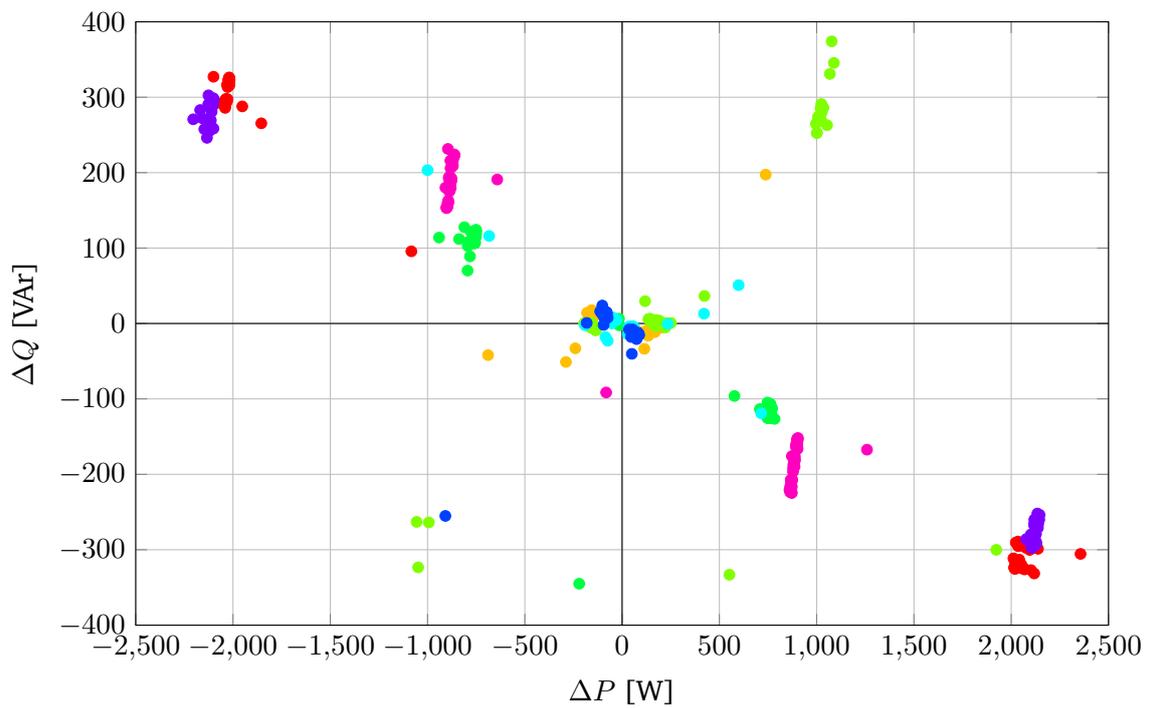
Fig. 6.8: histogram of the Euclidean distances for the features in data set I, file 1. Note that the usual order of plots was changed to make the differences visible.

<sup>37)</sup>If, for example, the feature vector contains real and reactive power differences, there is no real upper limit, because, in theory, any amount of power can be drawn. Under realistic conditions, there is an upper bound, though. It is given by the rating of the fuses in an electric installation. The phase difference between voltage and current, on the other hand, has natural limits at e.g.  $-\pi$  and  $\pi$ .

A scatter plot of the features is better suited to do this. With the reference feature extraction method, there are at least two more or less confined clusters in Fig. 6.9a). It can also be seen that there are several outliers that make the plot look cluttered. With SSA in Fig. 6.9b), the feature space is much cleaner. There are distinct clusters for each appliance, and there are drastically fewer outliers. It can be expected that this leads to a better classification result, especially for non-prototype based classifiers.



a) without SSA



b) with SSA

Fig. 6.9: SSPF extracted from file 1 of data set I

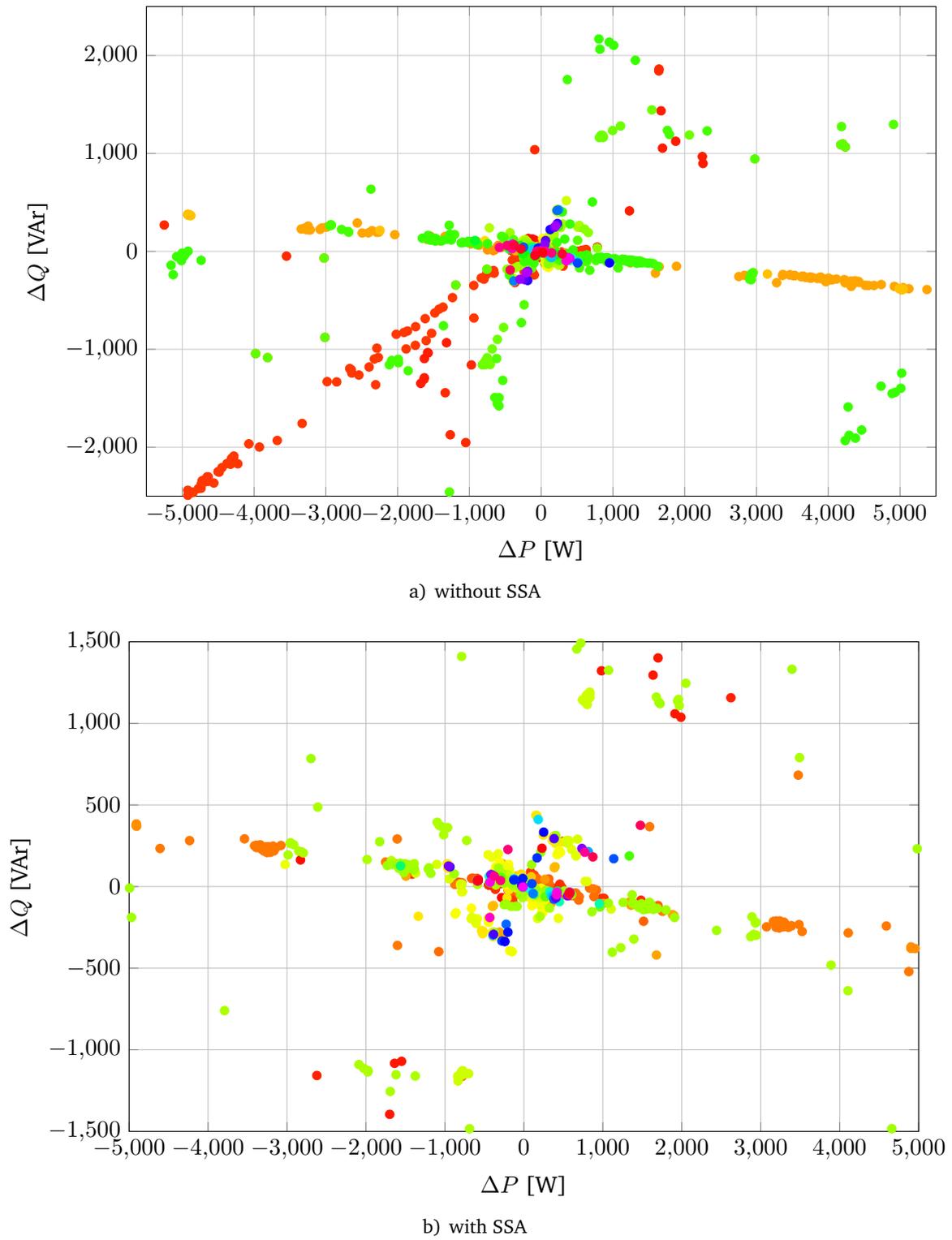


Fig. 6.10: SSPF extracted from the BLUED data set

### 6.4.2 Performance Comparison

Data set I is used in this subsection to compare the classification performance of three classification methods, when using the features of the reference algorithm and SSA features. The training data is taken from file 1, so there are 100 events per appliance. Similar to the Fig. 6.4, the test is performed with all 10 files contained in data set I. The individual classification rates and overall averages are shown in Fig. 6.11.

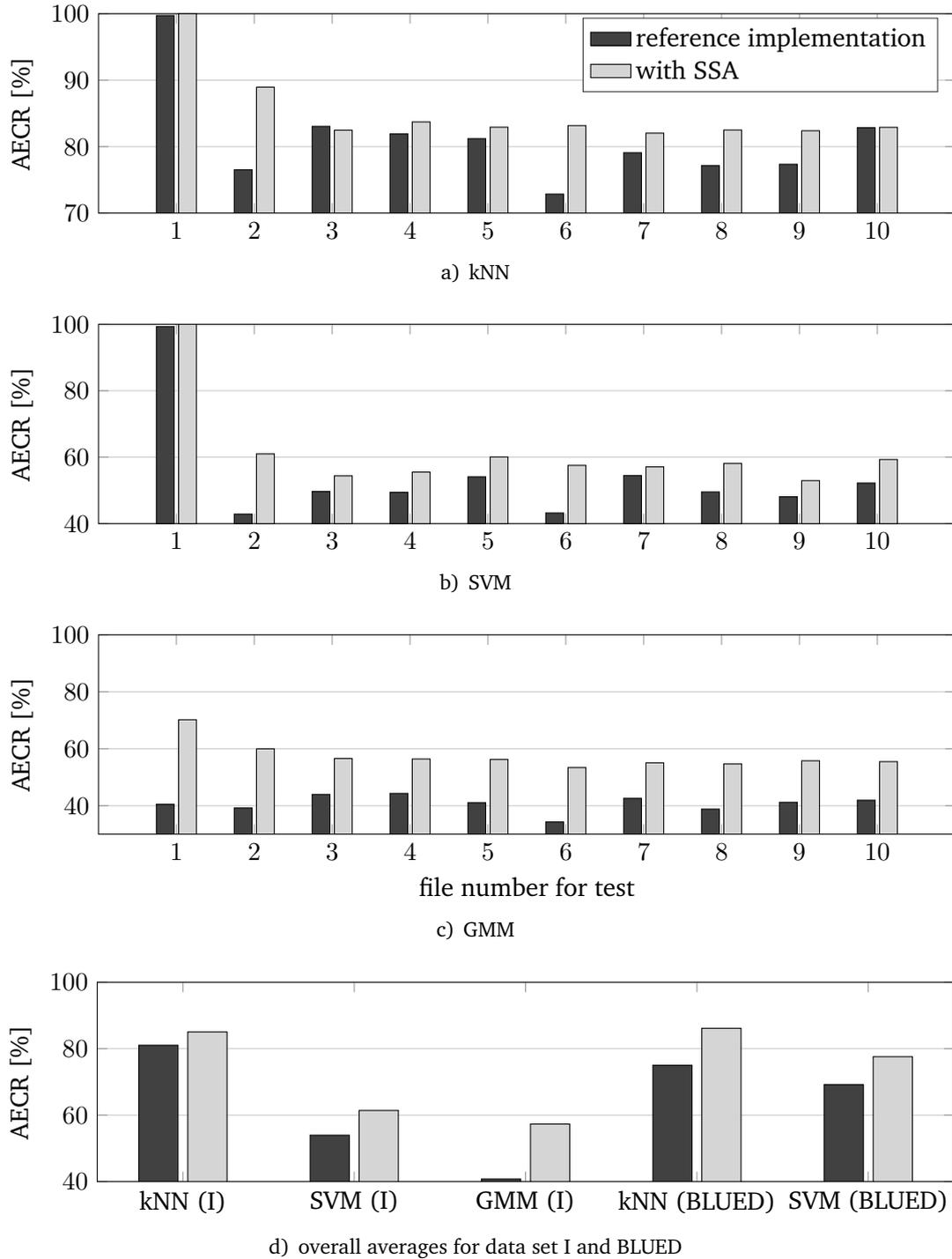


Fig. 6.11: test results for the 10 measurements in data set I, and overall averages

The benchmark is set by the kNN classifier with the optimal parameter set obtained in subsection 6.2.2, and a weight function  $w_i = \frac{1}{d(\vec{x}, \vec{y}_i)}$ , where  $d(\cdot, \cdot)$  is the Euclidean distance. The

kNN should be able to deal with the outliers, that were visible in Fig. 6.9, because it uses lazy learning. On the other hand it could be hard to reliably classify overlapping clusters, if the number of training samples is too small. The classification rates are visualized in Fig. 6.11a). It can be seen that with the SSA features the AECR is higher than the reference algorithm results for all test files. Overall, the SSA has an advantage of  $\approx 2.9\%$  over the reference implementation for the extraction of SSPFs.

Next, in order to put these results into perspective, a linear SVM (i.e. without kernel functions) is trained with the same data. An evaluation of suitable kernel functions goes beyond the scope of this thesis. From Fig. 6.11b), it can be seen that the classification rates are considerably lower. This means that the data are hard to divide with linear separation objects. The most important factor for the bad classification performance is that some appliances form multiple sub-clusters in the  $PQ$  plane. If this is not considered during training, which is the case here, any learning approach except prototype based methods will have trouble. This means that, in the field, the training must be long enough for all sub-clusters to be represented in the data. A linear learning mechanism has more trouble with outliers than a prototype based one because the outliers smudge the clusters' distributions. Hence, the linear SVM profits more from the stable features that are obtained by applying the SSA than the kNN. This can be seen by the higher performance difference between SSA and reference features in Fig. 6.11b) than in Fig. 6.11a).

Finally, unsupervised learning in the form of a GMM with single-mode multivariate normal distributions is analyzed. Its initial guesses for the appliance IDs, i.e. the latent variables  $z$  for all  $\vec{y} \in Y$ , are set to the ground truth data of the training set. After training with the training file, the classification result is obtained by finding the maximum of the a posteriori probability for each test sample. If the initial guesses are a good explanation of the training data, the test with the same set will yield almost 100 % AECR. As shown in Fig. 6.11c), this is not the case. To the contrary, about 50 % of the data is explained best by a different mapping of appliance IDs. This means that unsupervised learning as a simplification of the training procedure has little chances of success. This confirms the results by Reeg and Overbye [107]. Again, the SSA refinement brings considerable improvement because of the lower numbers of outliers.

The classifiers are also applied to a real world example in form of the BLUED features. See Fig. 6.11d) for the detection rates. They are comparable to the laboratory data set I results. Again, the SSA features lead to a  $\approx 6\%$  better classification performance.

### 6.4.3 Influence of the Training Set Size

An important factor for the usability of a NILM system is the number of training samples it requires. To analyze this, 20 % of data set I are used for testing in this subsection. The remaining events are chosen randomly for training. This allows for a range of 2 (1 for turn-on, 1 for turn-off) to 800 training samples per appliance. The experiment is repeated 500 times for each feature extraction method. In Fig. 6.12a), the average classification rates of the kNN classifier are plotted over the number of training samples. With both feature extraction methods, the classification rate increases with increasing training set size. The conclusion is that the appliance behavior is more complex than assumed by the standard model and it requires a lot of knowledge in order to classify them correctly. Again, the margin between reference method and SSA is around 3 % with the better end for SSA.

With the BLUED data set the AECR drop-off towards small training set sizes in Fig. 6.12b) is much more prominent. Reasonable classification rates of 70 % or higher are achieved not before 60 training samples per appliance are used. Here, the fact that the numbers of events per appliance are not equally distributed has to be kept in mind; the number of training samples can only be increased such that approximately 10 % of the samples remain for testing. The SSA feature extraction performs better than the reference implementation by approximately 6 %.

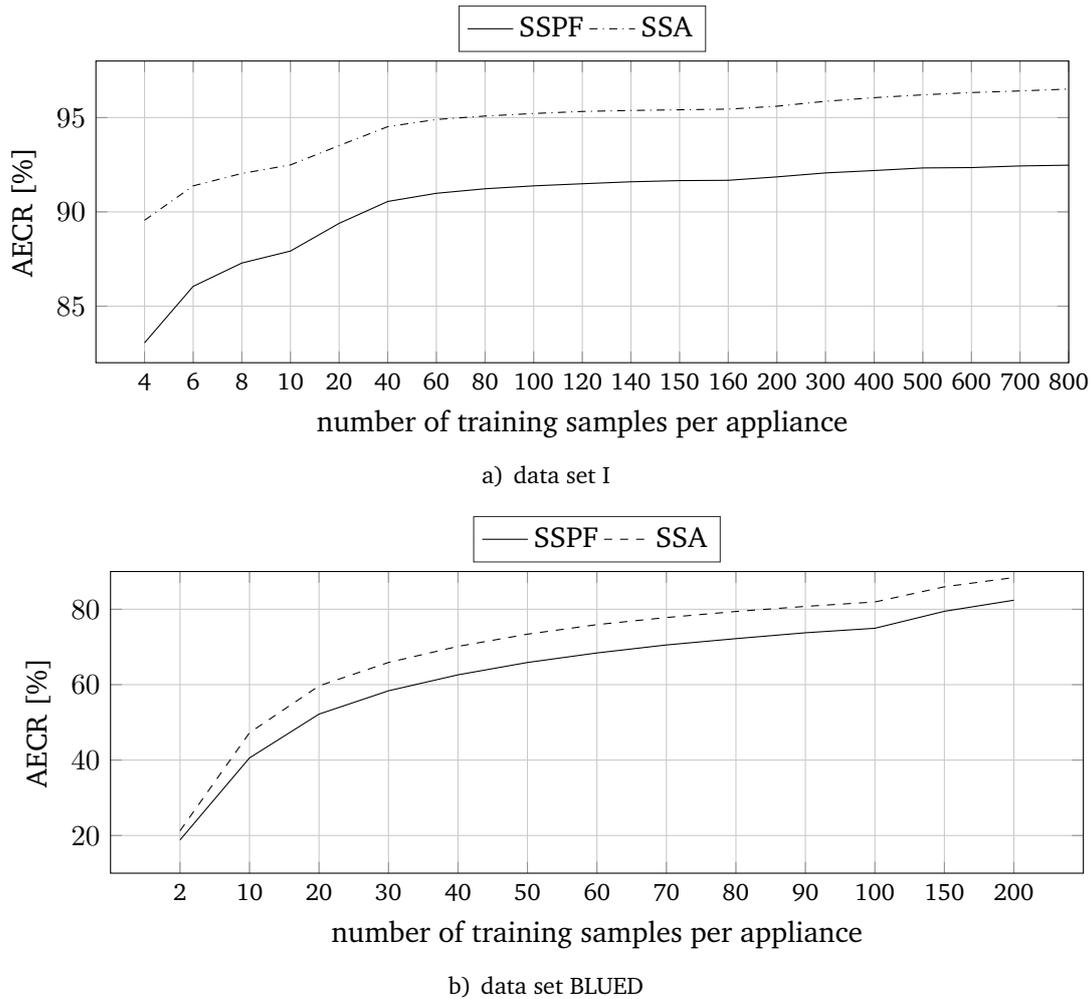


Fig. 6.12: influence of the number of training samples on the classification results

## 6.5 Time-Series Fingerprint Classification

In chapter 6.2, appliances were classified according to a feature vector, that consisted of a number of power or other signal differences computed at switching events. It was shown that, even with a limited collection of 10 appliances, there were overlapping clusters in the feature space, making it impossible to classify them correctly. An important insight was the fact that features computed during the settling phase, right after turning on an appliance, lead to unpredictable features. This thought is developed further in this section. The idea is to extract not only signal level differences, but characteristic time-series sequences from the RMS current signal. This is the continuation of an idea by Vogts, who wrote his Master’s thesis at Furtwangen University in 2011 [147]. It should help to identify appliances according to their typical settling phase. Another goal is to reduce the number of required training samples, ultimately reaching a situation where only one universal training fingerprint of an appliance class needs to be extracted.

The feature extraction is a straightforward procedure. Its fundamental steps are illustrated in Fig. 6.13. The RMS current  $I_{\text{rms}}[k]$  at a data rate of 50 Hz, as defined in eq. (4.61), is used as input signal<sup>38)</sup>. For each event  $e_i$ , a section of this signal is extracted as “fingerprint” or feature vector  $\vec{x}_i \in \mathbb{R}^{L_{\text{fp}}}$ . This is referred to as *time-series fingerprint* (TSFP). The fingerprint length,  $L_{\text{fp}}$ , will be defined later. An error function  $J : (\mathbb{R}^{L_{\text{fp}}}, \mathbb{R}^{L_{\text{fp}}}) \rightarrow \mathbb{R}$  is then applied to all combinations of known training fingerprints  $\vec{y}_j$  to compare them. This function maps the time domain to a quantity that measures the similarity of the fingerprints. The more similar two fingerprints are,

<sup>38)</sup>This is a consequence of the RMS demodulation of harmonic currents and will be addressed in section 6.5.1.

the higher  $J$  should be. In this thesis, the assumption is made that the fingerprints are actually characteristics of the appliances. A statistical distribution is not considered.

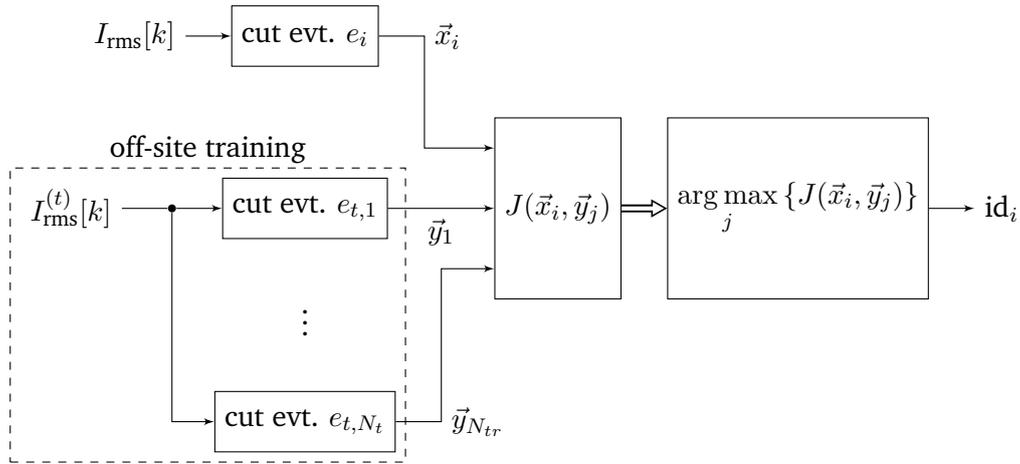


Fig. 6.13: classification procedure for the test feature  $\vec{x}_i$  using TSFPs

### 6.5.1 Fingerprint Model

Here, the model defined in eq. (4.1) is used as a basis. Let the measured signals be of the form

$$v[k'] = v_1 \cdot \cos(\omega_1 \cdot k') \quad (6.12)$$

$$i[k'] = s[k'] \cdot c[k'] \cdot v[k'], \quad (6.13)$$

where  $v[k']$  is the carrier that represents the fundamental grid component,  $s[k']$  is a piece-wise constant signal modeling the appliance state (on or off), and  $c[k']$  is a model for the appliance characteristic.  $c[k']$  is the signal that is used as TSFP.  $p[k']$  is demodulated using the RMS, as shown in section 4.5.5.2, in order to eliminate the periodic component in  $v[k']$ . After applying the RMS operation, and downsampling by the factor  $2\pi \cdot f_s / \omega_1$  (transition from  $k'$  to  $k$ ), the characteristic sequence becomes

$$I_{\text{rms}}[k] \propto \hat{c}[k] \cdot \hat{s}[k]. \quad (6.14)$$

It is an estimate of the appliance state sequence  $\hat{s}[k]$  and the appliance characteristic sequence  $\hat{c}[k]$  with a sample interval of  $2\pi/\omega_1$ . Assuming that  $s[k] \in \{0, 1\}$ , eq. (6.14) simplifies to

$$I_{\text{rms}}[k] \propto \hat{c}[k]. \quad (6.15)$$

This means that a section of  $I_{\text{rms}}[k]$  can directly be used as TSFP, when transitions in  $s[k]$  are known from the event detection. Let  $k = e_i$  mark the time equivalent signal index at which the  $i$ -th event has happened. Let there be  $N_e$  events, i.e.  $i = 1, \dots, N_e$ . The  $i$ -th fingerprint is defined as

$$\vec{x}_i = [I_{\text{rms}}[e_i - k_1] \quad \dots \quad I_{\text{rms}}[e_i] \quad \dots \quad I_{\text{rms}}[e_i + k_2]]^T \quad (6.16)$$

with two scalars  $k_1 \in \mathbb{N}$  and  $k_2 \in \mathbb{N}$ , which still need to be defined. They determine the time that is extracted before and after the event index  $e_i$ , so  $k_1$  and  $k_2$  define the length of the TSFPs. The fingerprint length is

$$L_{\text{fp}} = k_1 + k_2 + 1. \quad (6.17)$$

Examples for these time-series fingerprints are given in Fig. 6.14 for data sets IV and V. They will be used in section 6.5.7 for an evaluation of this type of features.

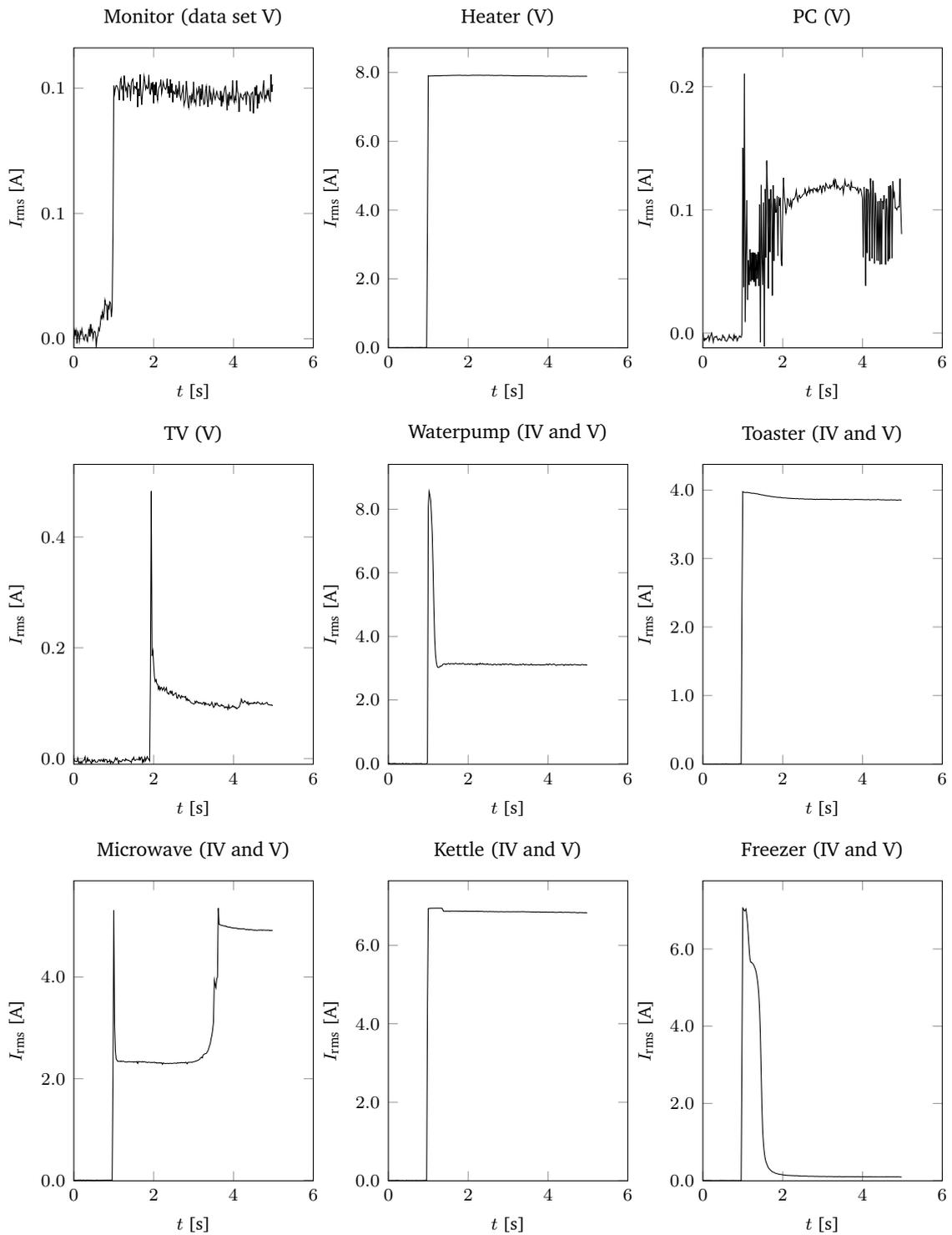


Fig. 6.14: TSFP fingerprint examples. The data sets they are included in are given in the each subplot's title.

### 6.5.2 Discussion

A fingerprint, as it is defined in this section, is a part of the current signal, that corresponds to the settling phases of electric appliances. The time this process usually takes needs to be found in order to get a good value for all classes. In general, the larger  $k_1$  and especially  $k_2$ , the larger the chance that the fingerprint actually covers the turn-on dynamics. However, there are two limiting factors.

1. If there are fluctuations in the duration of the period between turn-on and turn-off, i.e. the cycle duration, or the steady state period is short, the fingerprint should not expand over the whole operating period. As shown in Fig. 6.15, this would involve the danger of including the turn-off event which would result in wrong fingerprints.
2. In real-world scenarios, superposed appliance power consumptions are to be expected. To increase the chance of undisturbed clean fingerprints,  $L_{fp}$  should be chosen as short as possible.

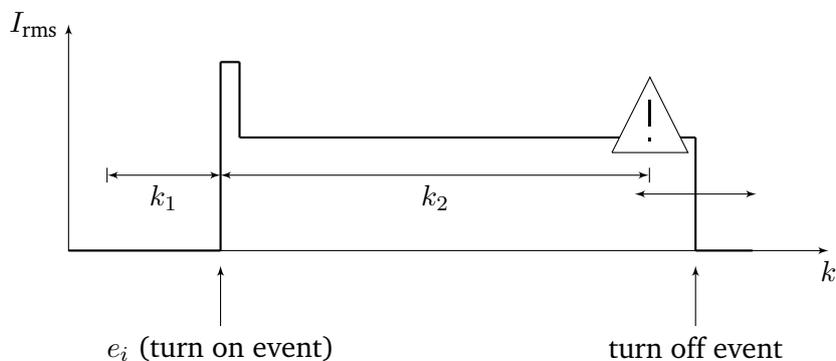


Fig. 6.15: the choice of  $k_2$  is limited by fluctuations of the cycle durations

$k_1$  defines the time period that is extracted before the event time stamp  $e_i$ . Therefore, the choice of  $k_1$  depends strongly on the precision that is achieved by the event detector. The observations, that will be presented in the appliance fingerprint catalog in section 6.7, indicate that a range of  $t_1 \geq 400 \text{ ms}$ <sup>39)</sup> is an appropriate choice to make sure that the fingerprints cover the turn-on transients. For the second parameter, a value of  $t_2 \approx 3 \text{ s}$ <sup>40)</sup> seems to be a good choice. Since, under laboratory conditions, superpositions can be ruled out,  $t_1 = 1 \text{ s}$  and  $t_2 = 4 \text{ s}$  are chosen for the following sections. This yields a fingerprint length of  $L_{fp} = 1 \text{ s} \cdot 50 \text{ Hz} + 4 \text{ s} \cdot 50 \text{ Hz} + 1 = 251$ .

### 6.5.3 Fundamental Assumptions

In this thesis, the following assumptions are made regarding the use of TSFPs.

1. There are no superpositions, i.e. no two appliances are operated simultaneously.
2. All fingerprints are of equal length, so they can be directly compared. This means that, in some cases, the fingerprint extraction period may protrude beyond the dynamic section into the steady state period.
3. Only turn-on events can be covered. The turning-off procedure is usually not visible to the smart meter. This was explained in section 4.2.

<sup>39)</sup>  $t_1$  is the equivalent of  $k_1$ , but  $k_1$  refers to a number of sampled values. This is determined by the signal data rate  $f_s$  via  $k_1 = f_s \cdot t_1$ . The same holds for  $k_2 = f_s \cdot t_2$ .

<sup>40)</sup> See footnote 39).

### 6.5.4 Classifiers for Time-Series Fingerprints

A special challenge for this type of features is its high dimensionality ( $L_{fp}$  dimensions) and the low number of training samples. Ideally, only one training sample per appliance class is sufficient. Two classifiers, that take these requirements into account, are described in this subsection and compared in subsection 6.5.5.

#### 6.5.4.1 Dissimilarity Classifier

This first option uses dissimilarity measures, computed by the Euclidean norm, to classify TSFPs. This means that a fingerprint is interpreted as a point in an  $L_{fp}$ -dimensional feature space. Each element of the fingerprint vector is considered a dimension. For the classification, a test fingerprint  $\vec{x}_i \in \mathbb{R}^{L_{fp}}$  is compared to  $N_{tr}$  training fingerprints  $\vec{y}_j \in \mathbb{R}^{L_{fp}}, j = 1, \dots, N_{tr}$  by the function

$$J(\vec{x}_i, \vec{y}_j) = d(\vec{x}_i, \vec{y}_j) = \sqrt{\langle \vec{x}_i - \vec{y}_j, \vec{x}_i - \vec{y}_j \rangle} = \sqrt{\sum_{m=1}^{L_{fp}} (\vec{x}_i[m] - \vec{y}_j[m])^2}. \quad (6.18)$$

The classification rule for the  $i$ -th event is

$$\text{id}_i = \arg \min_{j=1, \dots, N_{tr}} \{J(\vec{x}_i, \vec{y}_j)\}, \quad (6.19)$$

i.e. the  $i$ -th event is assigned to the one class for which the Euclidean distance is minimal. This is the same principle that is used in the nearest neighbor classifier.

#### 6.5.4.2 Cross-Correlation Classifier

As an alternative to the similarity classifier, an approach using the cross-correlation principle is described in this section. The measure for the similarity of two fingerprints is the maximum of the cross-correlation

$$J(\vec{x}_i, \vec{y}_j) = \max_{k=e_i-k_1, \dots, e_i+k_2} \left\{ \left( \frac{\vec{x}_i}{\|\vec{x}_i\|} \star \frac{\vec{y}_j}{\|\vec{y}_j\|} \right) [k] \right\}, \quad \begin{array}{l} i = 1, \dots, N_t \\ j = 1, \dots, N_{tr} \end{array}. \quad (6.20)$$

Because of the integrating character of the cross-correlation

$$(x \star y)[k] = \sum_{\ell=-\infty}^{\infty} x^*[\ell] \cdot y[k + \ell], \quad (6.21)$$

it is necessary to normalize the signals by

$$\frac{\vec{x}}{\|\vec{x}\|} = \frac{\vec{x}}{\sqrt{\sum_{i=1}^{L_{fp}} \vec{x}[i]^2}} \quad (6.22)$$

to unit energy. The normalization assures the comparability of the possibly highly different magnitudes of the fingerprints. The classification rule is

$$\begin{aligned} \text{id}_i &= \arg \max_{j=1, \dots, N_{tr}} \{J(\vec{x}_i, \vec{y}_j)\} \\ &= \arg \max_{j=1, \dots, N_{tr}} \left\{ \max_{k=e_i-k_1, \dots, e_i+k_2} \left\{ \left( \frac{\vec{x}_i}{\|\vec{x}_i\|} \star \frac{\vec{y}_j}{\|\vec{y}_j\|} \right) [k] \right\} \right\}. \end{aligned} \quad (6.23)$$

### 6.5.5 Comparison & Discussion

The computation of the Euclidean distance  $d(\vec{x}_i, \vec{y}_j) = \sqrt{\langle \vec{x}_i - \vec{y}_j, \vec{x}_i - \vec{y}_j \rangle}$  results in a distance larger than zero if a time-shifted signal  $y[k] = x[k - K]$ ,  $K \in \mathbb{N}$  and its original  $x[k]$  are considered, although their shapes are equal. A classifier using these distances as a (dis)similarity measure is therefore translation variant and prone to false classifications, if the event detector computes inaccurate time stamps. The cross-correlation is not affected by this because the exemplary function  $y$  is shifted over the whole length of the test function  $x$

$$(x \star y)[k] = \sum_m x^*[m] \cdot y[k + m], \quad (6.24)$$

making the cross-correlation classifier invariant to translation. It can be expected that the cross-correlation method is more robust than the dissimilarity classifier. The drawback is that the cross-correlation classifier needs to normalize the fingerprint to make sure they are comparable. The dissimilarity classifier, which does not require normalization, might be superior when there are appliances with similar dynamic sequence but different power consumption. The problem is illustrated in Fig. 6.16.

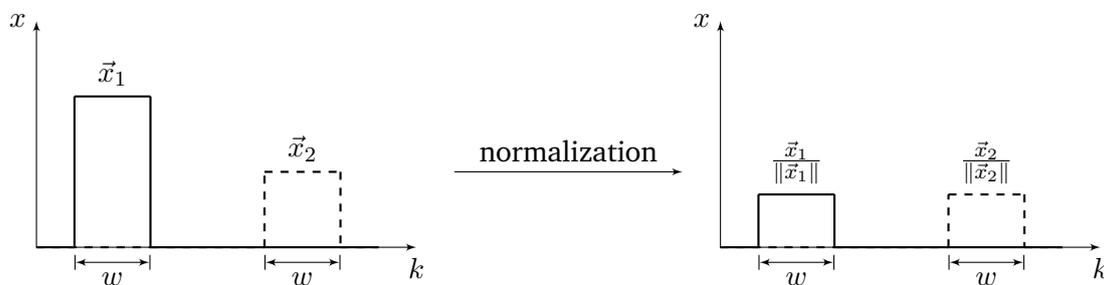


Fig. 6.16: influence of the energy normalization on two rectangular pulses

The Euclidean distance, as defined in eq. (6.18), requires both the test fingerprint and the training fingerprint to be of equal length. Equal length is not required for the cross-correlation. Hence, the cross-correlation classification method is preferred in this thesis.

### 6.5.6 Fingerprint Synchronization

This section is based on a Bachelor's thesis that was written during the course of this PhD thesis by Alexander Bartler [11]. When using the temporal sequence of power values as a feature, it is important to make sure that they are compared with the same relative time shift, i.e. a point in time at which all fingerprints of an appliance are most similar. This is especially true for the dissimilarity classifier, because it has no inherent time shift mechanism. The cross-correlation classifier, on the other hand, is invariant to translation. However, a synchronization to reference points within each fingerprint may help to make the cutting out of relevant sections of the signal more precise.

In order to find reference points in fingerprints, the features are compared to a generic function  $x_s : \mathbb{R} \rightarrow \mathbb{R}$ . To find the amount of time  $t_{\text{corr}}$ , by which a test feature  $\vec{x} \in \mathbb{R}^{L_{\text{fp}}}$  has to be shifted, the maximum of the cross-correlation

$$t_{\text{corr}} = \arg \max_k \left\{ \left( \frac{\vec{x}}{\|\vec{x}\|} \star \frac{\vec{x}_s}{\|\vec{x}_s\|} \right) [k] \right\} \quad (6.25)$$

is sought<sup>41)</sup>.  $\vec{x}_s \in \mathbb{R}^{N_s}$  is a sequence of  $N_s \in \mathbb{N}$  samples of  $x_s$ . The best universal function was found to be a linear function of the type

$$x_s(t) = m_s \cdot t + c_s. \quad (6.26)$$

<sup>41)</sup>In a real implementation,  $t_{\text{corr}}$  has to be corrected for the length of either  $\vec{x}$  or  $\vec{x}_s$ , depending on the order of arguments to the cross-correlation function.

For turn-on events, a negative slope  $m_s$  and  $c_s > 0$  are used. Turn-off events are best synchronized by a function with  $m_s > 0$  and  $c_s = 0$ . The two versions are sketched in Fig. 6.17.

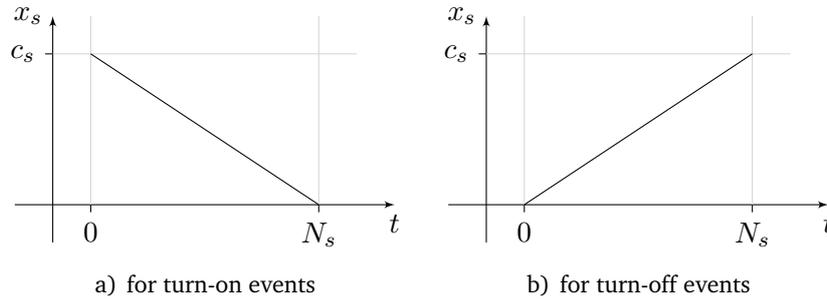


Fig. 6.17: linear synchronization functions

A necessary condition for the synchronization to be successful is that a global maximum of the cross-correlation can be found. This is achieved if

$$N_s \geq L_{fp}, \quad (6.27)$$

i.e. if the generic function is longer than the fingerprint to synchronize, or if they are both of equal length. The optimum value for the ratio  $\frac{N_s}{L_{fp}}$  was found to be  $\geq 2.5$ . Referring back to Fig. 6.13, the synchronization of fingerprints is included in the “cut evt.  $e_i$ ” blocks.

### 6.5.7 Performance Comparison on Real Measurements

The performance of the TSFP classifiers is assessed using data sets IV (section 4.8.4) and V (section 4.8.5). For training, one fingerprint per appliance is used. The remaining data is used for testing. The SSPF algorithm defined in 6.2.1 is used as a benchmark. The numbers presented in Table 6.3 indicate that at a low number of appliances (data set IV), both TSFP classifiers are equally good and superior to the SSPF method, even though the margin is small. If the number of appliances is increased (set V), the TSFP methods fall behind the SSPF algorithm. Here, the cross-correlation method has an advantage over the dissimilarity measure approach. The reason for this is investigated in the next section 6.5.8.

Table 6.3: classification results

algorithm	data set IV	data set V	data set I*	data set I**
TSFP with dissimilarity classification	97.3 %	80.3 %		
TSFP with cross-correlation classification	97.3 %	82.7 %	35.2 %	74.0 %
SSPF with kNN	96.4 %	85.7 %	81.0 %	

A comparison with the TSFP results, obtained in section 6.4, is given in the last two columns of Table 6.3. If all appliance events including turn-off events are classified (marked by data set I\*) the TSFP performance is poor. This is owed to the relatively high number of simple heating appliances which, with normalized fingerprints, will turn out very similar. Reducing the classified appliances to those with characteristic fingerprints, the classification rate is improved (data set I\*\*). However, it does not reach the performance of the SSPF classifier. The cause for this is the missing handling of the superpositions that are present in data set I. Even though the superpositions are a violation of the requirements from section 6.5.2 the TSFP classification yields reasonable performance.

### 6.5.8 Analysis of the Classification Errors

The TSFP classifiers have shown to misclassify some events when the number of appliance classes is large. The reasons for this are investigated in this section.

The few errors, that occurred with data set IV, result from small differences between the signal shape of the training sample, and some test samples of the same class. Another reason is that there are two similar classes in this data set. A typical error is shown in Fig. 6.18a). The solid line shows the test sample, which belongs to a water kettle (class 2). It obviously deviates from the trained sample (dashed), making the fingerprint of the toaster (dash dotted, class 4) more similar. This leads to a misclassification and is observable with both the cross-correlation and the dissimilarity measures (Fig. 6.18b)). The latter reveals another problem: Not only is the kettle's signal shape different from the training sample, but the power consumption is also very different. In fact, it is much more similar to the toaster than to the trained kettle fingerprint. This only comes to light because the dissimilarity classifier does not use normalization. This insight means that the SSPF algorithm will misclassify this event, too, and the attempt to use a universal training set fails in this case. The training data would have to be recorded for both kettle variations.

The errors illustrated in Fig. 6.19a) and 6.19b) are similar to the previous one. Here, the signal shape deviates from trained fingerprints in a way that successful classification is a rather random incident. The proposed methods are not able to deal well with these situations. At best, an on-site training of specific appliances could help to reduce false classifications.

An error due to too long fingerprints is shown in Fig. 6.20a). If  $k_2$  is chosen too large, the fingerprint extraction periods overlaps the turn-off transient, or, in this case, the operation cycle is unexpectedly short. Note, that the magnitude in Fig. 6.20a) is distorted due to the normalization. Without normalization, the shapes are more similar (Fig. 6.20b)). The SSPF algorithm does not run into this problem here, because it computes its feature much closer to the event. To remedy this situation, the TSFP length should be made variable. Another approach to solving these situations is to introduce classification probabilities and combine classifier results.

In the last error scenario, the time scale of the fingerprints is different. This is clearly visible in Fig. 6.21, and it may lead to false classifications if the differences are significant. Here, an on-site training could help, which again is a situation in which the universal training approach seems to fail. Another possibility is to divide the fingerprint into sections so that fixed-points can be extracted.

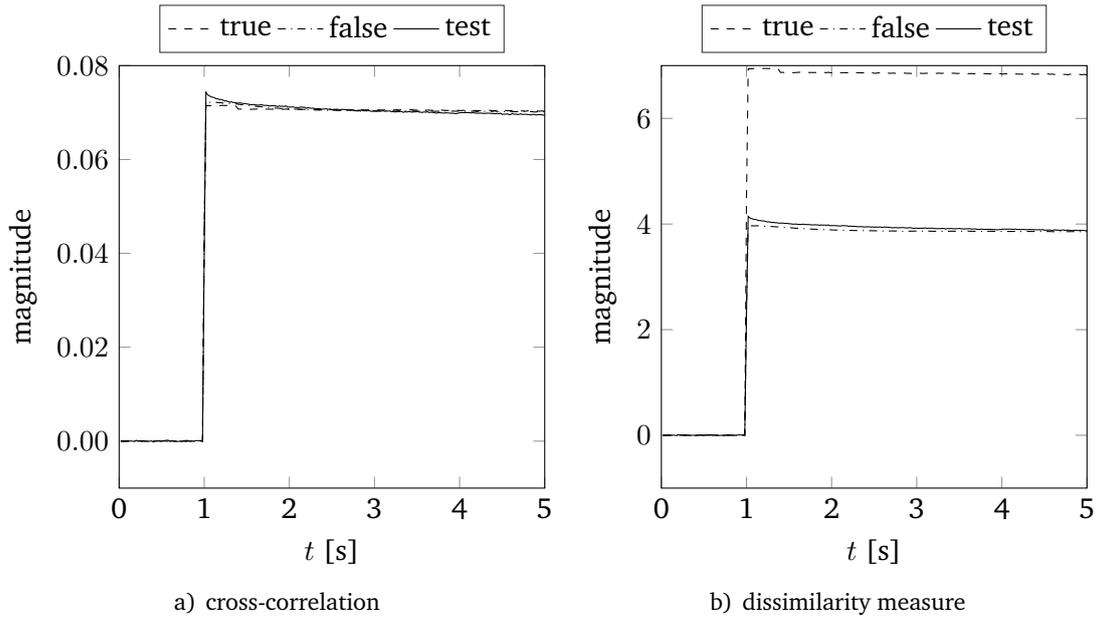


Fig. 6.18: misclassification due to deviating signal shapes (data set IV)

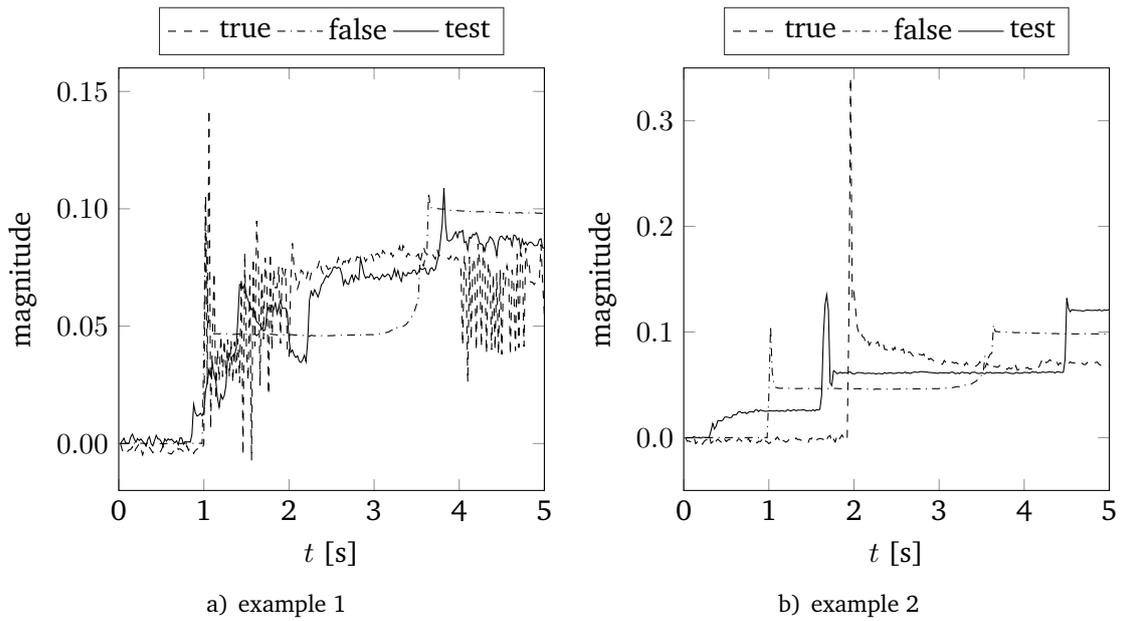


Fig. 6.19: misclassification due to unexpected signal shape (data set V)

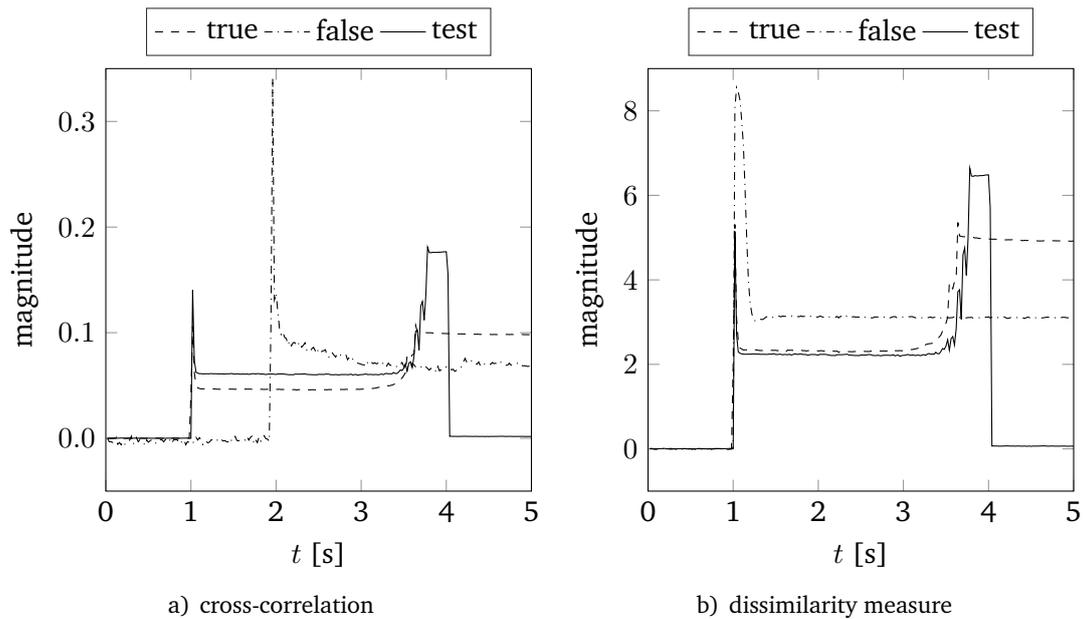


Fig. 6.20: misclassification due to fingerprints overlapping the turn-off transient (data set V)

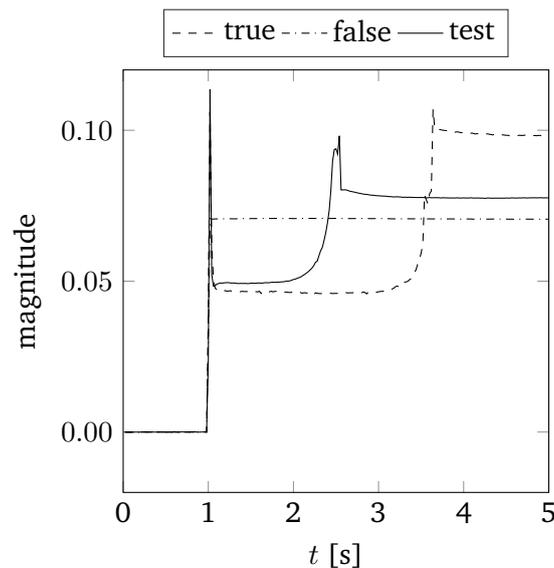


Fig. 6.21: misclassification due to different time scales (data set V, cross-correlation)

### 6.5.9 Recommended Method

From the theoretical reflection and the classification results it is advisable to use the cross-correlation method for TSFP classification. If there are only few classes of appliances, and all requirements from section 6.5.3 are met, both methods may be equally good, but under more realistic conditions the cross-correlation outperforms the dissimilarity measures. It is therefore used exclusively throughout the remainder of this chapter.

## 6.6 TSFP Data Volume Reduction

The fingerprints introduced in the previous section are vectors of length  $L_{\text{fp}} = k_1 + k_2 + 1$ . As described before in section 6.5.2, there is a lower limit for both  $k_1$  and  $k_2$ . For the numbers

chosen in this chapter, the fingerprint length, or number of dimensions, is

$$L_{\text{fp}} = 1 \text{ s} \cdot 50 \text{ Hz} + 4 \text{ s} \cdot 50 \text{ Hz} + 1 = \underline{\underline{251}}, \quad (6.28)$$

which requires much more memory than a SSPF feature vector  $[\Delta P \ \Delta Q]^T$ . Therefore, this section is dedicated to a search for the minimum signal frequency at which TSFPs can be used as a feature, thus reducing the memory required for a feature. As illustrated in Fig. 6.22, the AECR is used again as an indication of the chances of success. In comparison to Fig. 6.13, downsampling units  $d \downarrow$  and low-pass filters  $h_d$  to avoid aliasing are added. The downsampled fingerprints are computed by first applying the low-pass filter

$$\vec{x}'_i = \vec{x}_i * \vec{h}_d \quad (6.29)$$

with a not yet defined low-pass impulse response  $h_d[k]$ . Afterwards,  $\vec{x}'_i$  is downsampled by the factor  $d \in \mathbb{N}$

$$\vec{x}'_i[j] = \vec{x}'_i[j \cdot d] \quad j \in \mathbb{N}, \quad e_i - k_1 \leq j \cdot d \leq e_i + k_2, \quad i = 1, \dots, N_e. \quad (6.30)$$

Hence, the filter has to be designed such that the corner frequency is

$$\Omega_{-3\text{dB}} \leq \frac{\Omega_{\text{max}}}{d}, \quad (6.31)$$

where  $\Omega_{\text{max}}$  is the upper limit of the bandwidth of  $\vec{x}_i$ . The resulting signal rate is

$$f_{\text{lo}} = \frac{f_s}{d}. \quad (6.32)$$

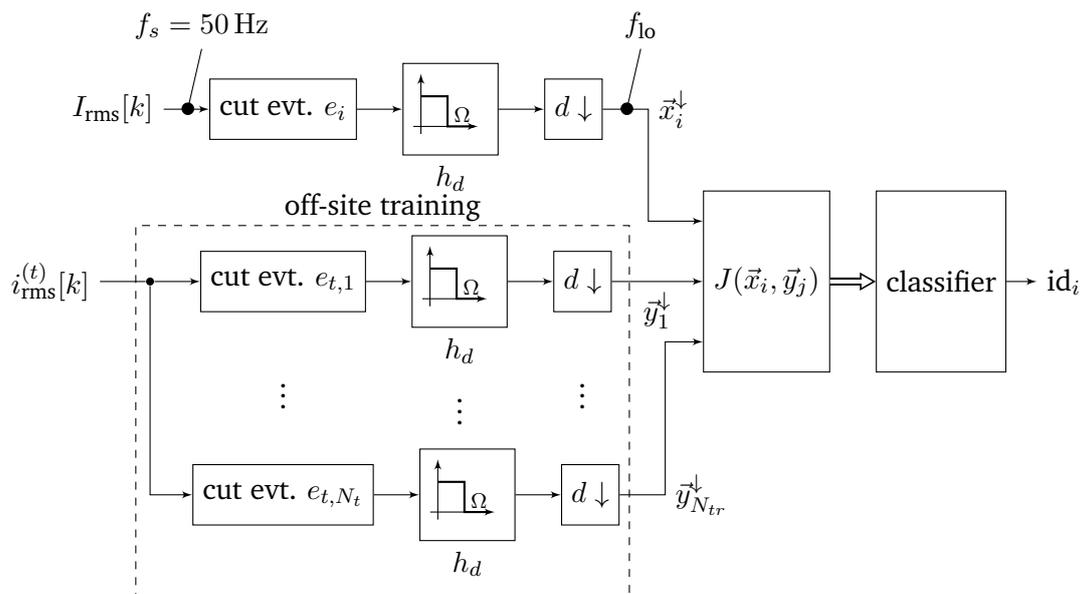


Fig. 6.22: TSFP classification using downsampled fingerprints (cf. Fig. 6.13)

### 6.6.1 Filter Design

The Mallat expansion of the Wavelet Transform is an elegant and efficient way to obtain a filtered downsampled versions of a signal. See section B.3.1 in the appendix on page 137 for a short introduction. Each sampling interval level can be achieved by evaluating the approximation path. The effect of each downsampling step is shown in Fig. 6.23. In the left column, a fingerprint of a refrigerator is downsampled down to a resulting sampling rate of 1.67 Hz. In the right column, the low-pass filter is applied *without* downsampling. It becomes clear that a significant part of the characteristic behavior vanishes with decreased sampling rate. From a visual point of view, this is a severe limitation of the fingerprints' quality. This is why a classification comparison is done in the next subsection.

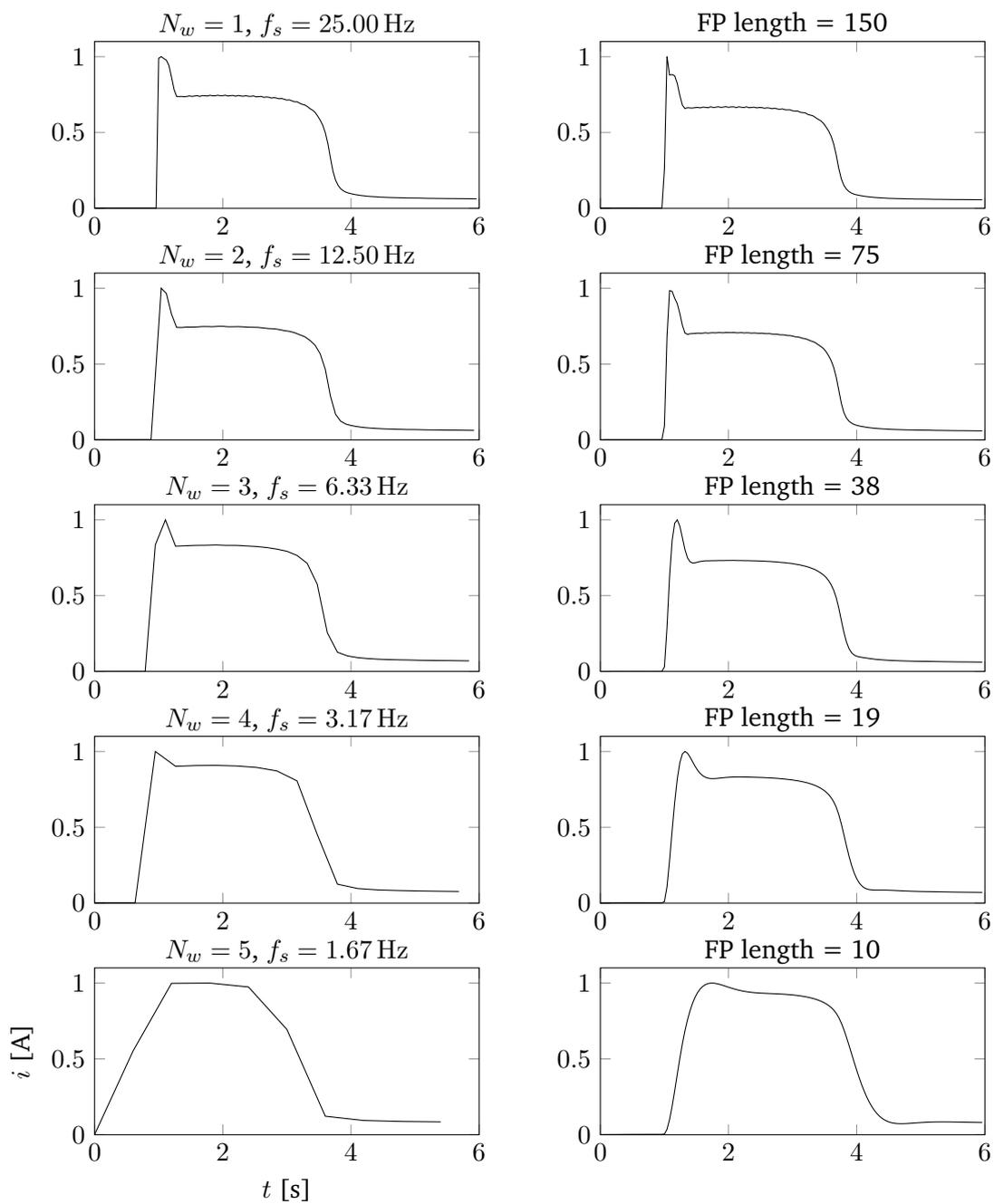


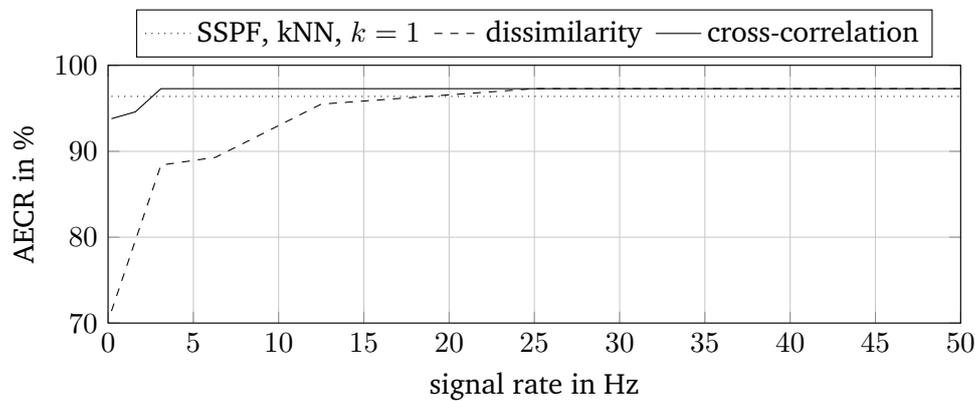
Fig. 6.23: fingerprint of a freezer at different levels of signal rate

### 6.6.2 Minimum Acceptable Signal Rate

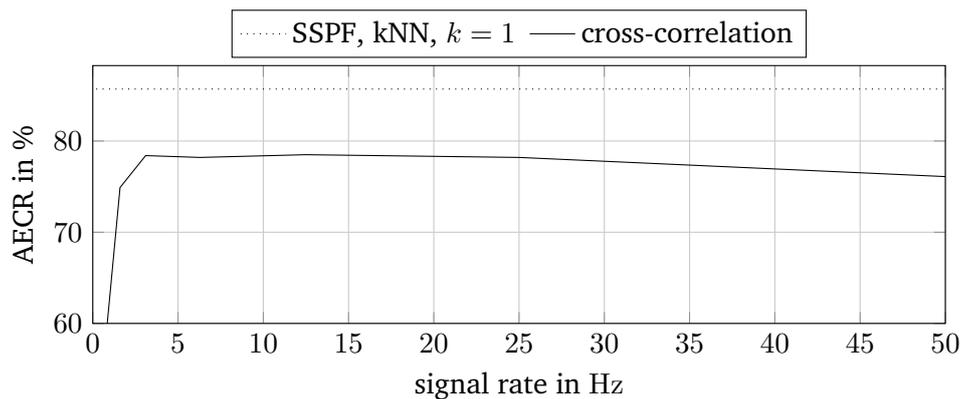
The experiment from section 6.5.7 is repeated here with different levels of signal sampling rate. The classification rates for data set IV are shown in Fig. 6.24a). It can be confirmed that the cross-correlation bank outperforms the dissimilarity measure method, because it is able to achieve its maximum classification rate at much lower signal rates. The differences result from the fact that the downsampling may introduce a time shift to the fingerprints, which poses a problem to the translation variant dissimilarity classifier. With the cross-correlation bank, a downsampling by the factor of 16 is possible without losing classification performance. A signal rate of 3 Hz should therefore not be undercut.

In the case of data set V (Fig. 6.24b)), the AEGR even increases a little bit when reducing the sampling rate starting from 50 Hz. This means that there are parts of the signal that are better eliminated, because they cause the classifier to get some fingerprints confused. Again, the classification rate drops drastically below a sampling rate of 3 Hz.

If the signal rate can be reduced to around 3 Hz, this means that the relevant information that is extracted with the fingerprint as defined in this thesis is spread over a spectral range of  $0 < f \leq 1.5$  Hz. All higher frequency components could not be represented by a signal that is sampled with 3 Hz.



a) data set IV



b) data set V

Fig. 6.24: influence of the signal rate on the classification performance

## 6.7 Appliance Turn-On Fingerprint Catalog

In section 6.5, current signal sections were introduced as a new feature type for appliance classification. In this chapter, a multitude of different appliances is analyzed to form a catalog. The results were published in the paper [12]. Because the catalog is based on a variety of

different appliances, a suitable display method is to plot the average turn-on characteristics for all appliances of one class into one plot. Let

$$X = \{\vec{x}_1, \dots, \vec{x}_N\} \quad (6.33)$$

be the set of  $N$  signal traces for an appliance class, where  $\vec{x}_i$  is a column vector containing the fingerprint time series values of the  $i$ -th event. Then the average fingerprint  $\vec{x}_{\text{avg}}$  is the expectation of  $X$

$$\vec{x}_{\text{avg}} = \mathcal{E}\{X\} = \frac{1}{N} \sum_{i=1}^N \vec{x}_i. \quad (6.34)$$

The uncertainty of the fingerprints of one class is measured by the bias-corrected variance

$$\vec{x}_{\text{var}}[k] = \frac{1}{N-1} \sum_{i=1}^N (\vec{x}_i[k] - \vec{x}_{\text{avg}}[k])^2. \quad (6.35)$$

A graphical representation is obtained by plotting the *standard deviation band*, which is limited by the vectors

$$\vec{x}_{\text{SD-}} = \vec{x}_{\text{avg}} - \vec{x}_{\text{var}}^{\frac{1}{2}} \quad (6.36)$$

and

$$\vec{x}_{\text{SD+}} = \vec{x}_{\text{avg}} + \vec{x}_{\text{var}}^{\frac{1}{2}}. \quad (6.37)$$

The catalog is based on data set V, which contains 137 different appliances with 577 switching events in total. Two classes of appliances stand out by showing a characteristic turn-on settling procedure: microwave ovens and cooling devices (refrigerators, freezers). Two examples are given in Fig. 6.25 for a microwave oven, and Fig. 6.26 for a refrigerator, respectively. It can be seen that both follow a recurring pattern. For the two classes in question, this is shown in Table 6.4. The exact parametrization of each section may be different for each appliance, but the fundamental structure is similar within each class. An example for this is given in Fig. 6.27 for a different microwave. Note, that the uncertainty in the transition states and the steady state is much higher than in Fig. 6.25.

Table 6.4: sequence of distinct sections in the fingerprints of microwaves and cooling devices

section	description
1	High power spike (may sometimes be small or missing at all)
2	Decrease to medium power
3	Medium power state
4	Transition to steady state
5	Steady state

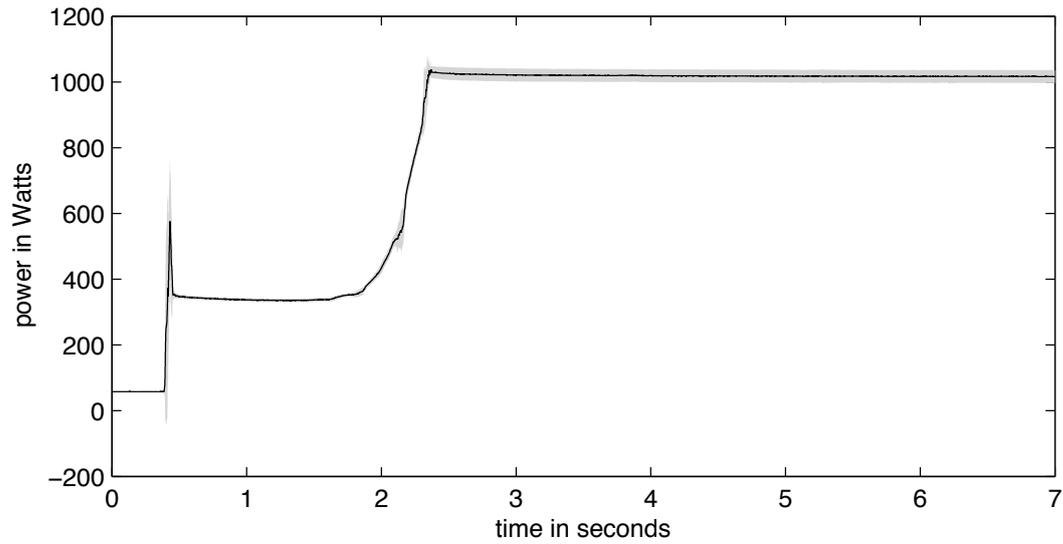


Fig. 6.25: mean (black line) and uncertainty band (gray area) for a microwave oven averaged over 10 events (manufacturer: Sharp)

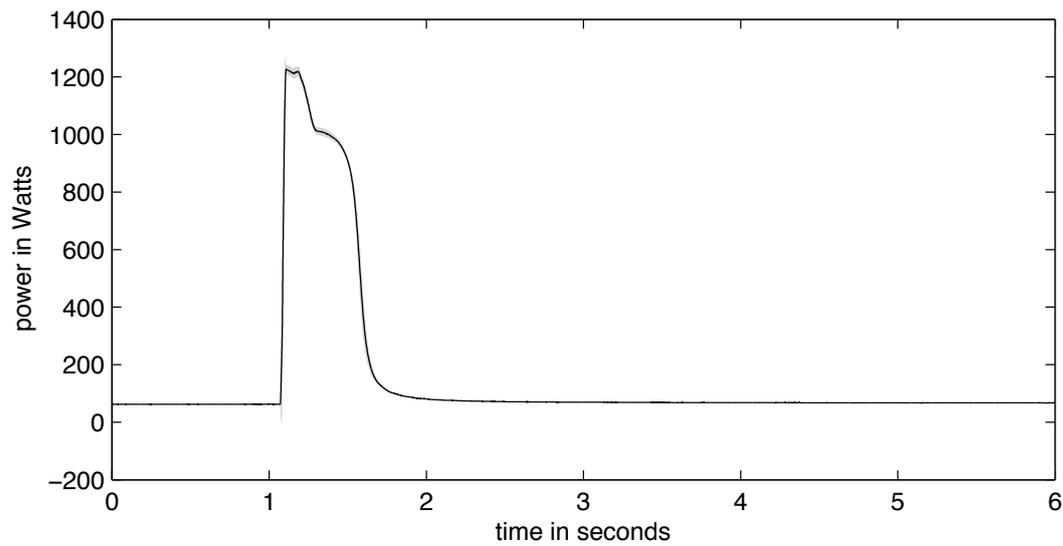


Fig. 6.26: mean (black line) and uncertainty band (gray area) for a refrigerator averaged over 30 events (manufacturer: Bomann)

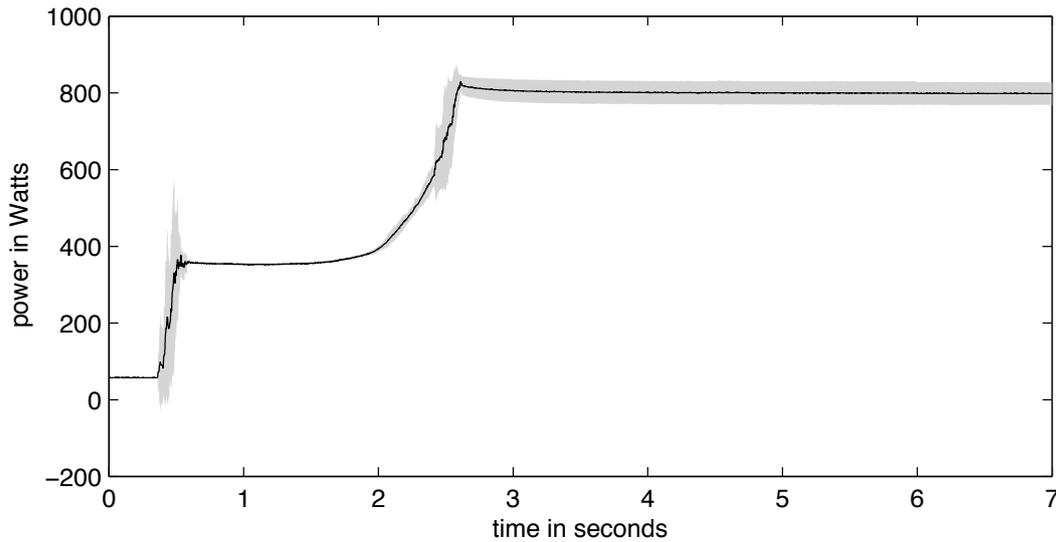


Fig. 6.27: mean (black line) and uncertainty band (gray area) for a microwave oven averaged over 10 events (manufacturer: micromaxx)

The situation is different for heaters. There is no sequence of distinct states, as seen before. However, there are more subtle characteristics. Heating components usually have two parameters special to their power consumption: the steepness of the power increase after the power switch is turned on, and the power draw pattern due to the conductor's temperature-dependent resistance. This can be seen in Figures 6.28, 6.29, and 6.30. The figures show the switching cycles of a room heater, and the heater elements of two washing machines, respectively. The time constant of the settling process after reaching maximum power consumption is much longer for the Privileg washing machine (Fig. 6.29) than the one of the Miele machine (Fig. 6.30). The space heater shows no observable settling process, which means that its filament almost immediately reaches its final constant temperature. The washing machine heaters, on the other hand, need to heat up water and parts of the machine, which takes a certain amount of time.

The change of the settling parameters over time will be used for a study of how to detect overheating appliances in chapter 7.2.

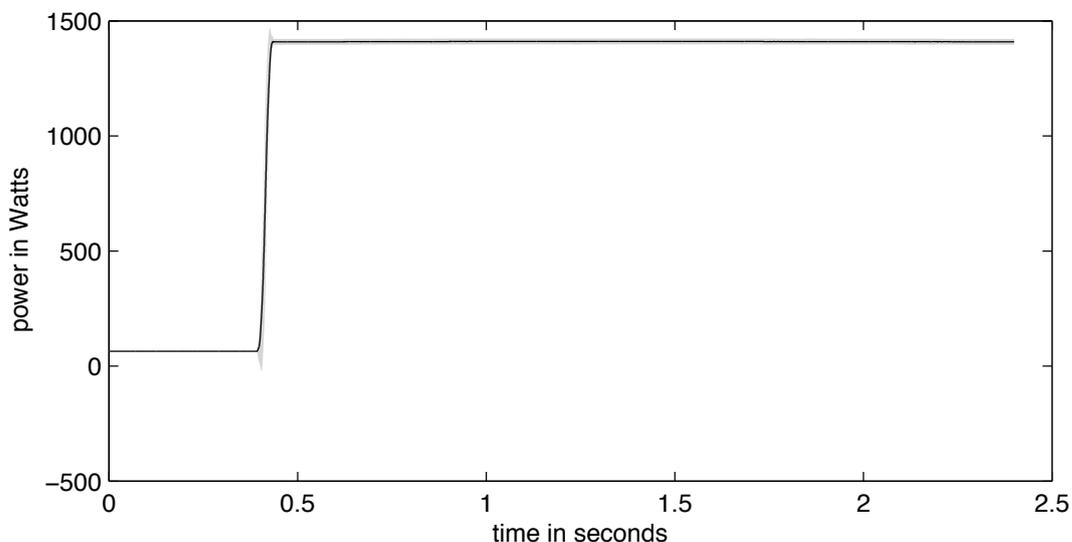


Fig. 6.28: mean (black) and uncertainty band (gray area) for a space heater averaged over 215 events (manufacturer: FiF)

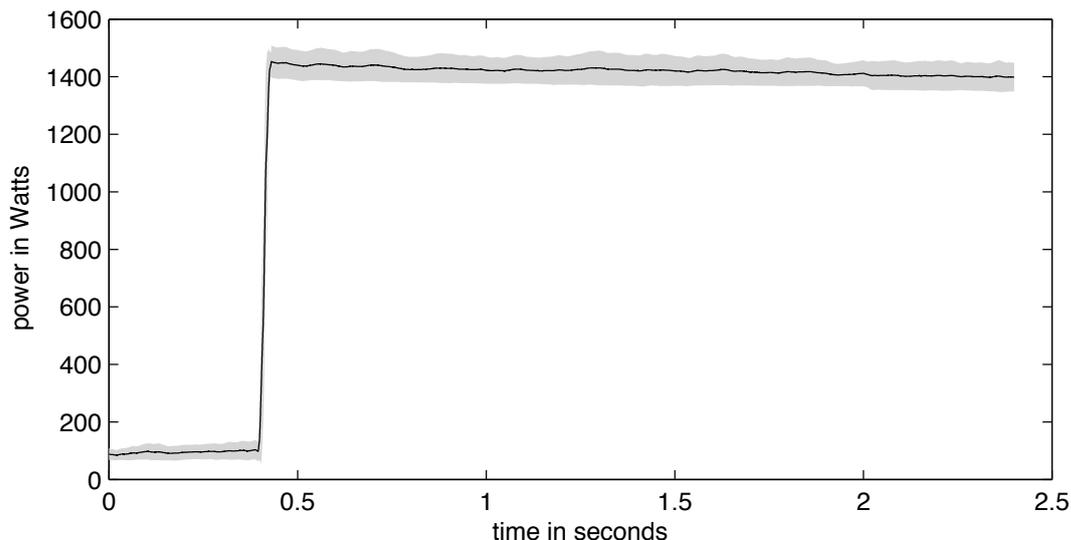


Fig. 6.29: mean (black) and uncertainty band (gray area) for the heater component of a washing machine averaged over 7 events (manufacturer: Privileg)

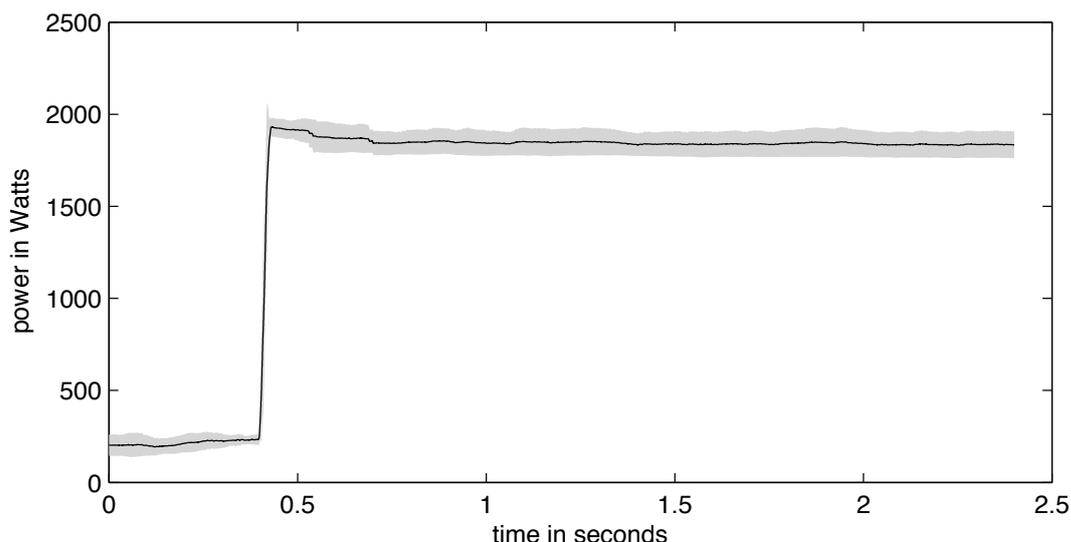


Fig. 6.30: mean (black) and uncertainty band (gray area) for the heater component of a washing machine averaged over 6 events (manufacturer: Miele)

## 6.8 Conclusion

In this chapter, two approaches to feature extraction were described. The first makes use of differences of short averaging periods to compute a feature for a switching event. This method is very easy to implement. In this thesis, differences of real and reactive power were used to construct the feature vector. An implementation for 50 Hz RMS signals was described, together with an optimization of the involved parameters. The features can be classified using standard classification methods. Especially the kNN algorithm has proven to be successful. Nevertheless, due to fluctuations in the appliance behavior, the features tend to be unstable, which has a negative effect on the classification performance. Due to the instability of the features, a considerable amount of training data is required, which is a downside for convenient implementation in the field.

The Steady State Approximation method, that was introduced in section 5.4, is able to eliminate power fluctuations due to the settling of electric appliances. It was shown to reduce the variance of the appliances' clusters by computing actual steady state features. This led to an improvement of 3 % to 6 % with respect to the appliance event classification rate.

The second class of features is a new addition to the research field. Using time-series fingerprints comes with the benefit that one training sample is sufficient for certain appliances to be correctly classified. Another benefit is that, for some types of appliances, a class specific training feature can be used a priori, i.e. no on-site training is required at all. This makes the TSFP method an ideal complement for the SSPF method by providing a convenient method for partial training. In special cases, the TSFP method outperforms the SSPFs, because it allows for a differentiation of appliances with equal or very similar power consumption. It was also shown that the sampling rate can be reduced down to 3 Hz without losing crucial information.

A short introduction to an appliance fingerprint catalog using these TSFP features was presented. This showed that for some classes characteristic fingerprints can be used as features and that they can be extracted a priori. This reduces the training efforts for these features.



## Chapter 7 | Case Studies

Two experiments regarding the information acquisition from load monitoring data are presented in this chapter. First, activity profiles from electricity data are analyzed in section 7.1. Secondly, in section 7.2, it is shown how the overheating of electric stoves can be detected non-intrusively by analyzing their characteristic switching behavior.

### 7.1 Pattern Discovery on Appliance Switching Profiles

In this section, a pattern discovery study is presented to show how the daily life of residents is reflected in the switching behavior of electric appliances. It is expected that the switching of appliances, in contrast to their power demand, reveals more information on how active the residents are over time. The goal is to find a representation of appliance switching events that is suitable for the detection of abnormal or changing life patterns. An identification of active appliances and ADLs is not intended.

#### 7.1.1 Definition of Characteristic Profiles

Assuming that a day is the most common recurring pattern in a person's life, a time span of 24 h is chosen as base interval in this chapter. Let  $\vec{t}_i \in \mathbb{R}^{N_e}$  be the list of  $N_e \in \mathbb{N}$  appliance switching events (time stamps), detected on day  $i$ . Let  $\vec{p}_i \in \mathbb{R}^{N_b}$  be the activity profile of this day, i.e.  $\vec{p}_i$  contains the number of events counted in time intervals  $\Delta t_p$  over a period of 24 hours. Therefore, a profile can be considered an integrated version of  $\vec{t}_i$  with equidistant samples on the time axis. Each profile contains

$$N_b = \frac{24 \text{ h}}{\Delta t_p} \quad (7.1)$$

samples. Furthermore, let  $P = \{\vec{p}_i \mid i = 1, \dots, N_p\}$  be the set of  $N_p \in \mathbb{N}$  daily profiles. The average profile is the expectation

$$\vec{p}_{\text{avg}} = \mathcal{E}\{P\} = \frac{1}{N_p} \cdot \sum_{i=1}^{N_p} \vec{p}_i. \quad (7.2)$$

The time interval  $\Delta t_p$  is an important factor for the information content that can be inferred from a profile. If  $\Delta t_p$  is chosen too large the profile becomes an average of one or more days. This does not convey activity information on a daily basis. If, on the other hand,  $\Delta t_p$  is too small, the profile is constituted of single (random) events and no statistical patterns can be deduced.

An example is given in Fig. 7.1. An interval of  $\Delta t_p = 1 \text{ h}$  results in a rough profile (dash-dotted line). It may be hard to determine the exact locus of peaks here. Counting the events per minutes yields a profile that shows only little resemblance of the other two. An interval of  $\Delta t_p = 10 \text{ min}$  seems to be a good compromise between resolution and integrating character.

#### 7.1.2 Data Set

Additionally to the UK-DALE data (cf. section 4.7 on page 58), a data set labeled “(ReSP) data set VI” was created by the author specifically for this thesis. The setting is an apartment of a

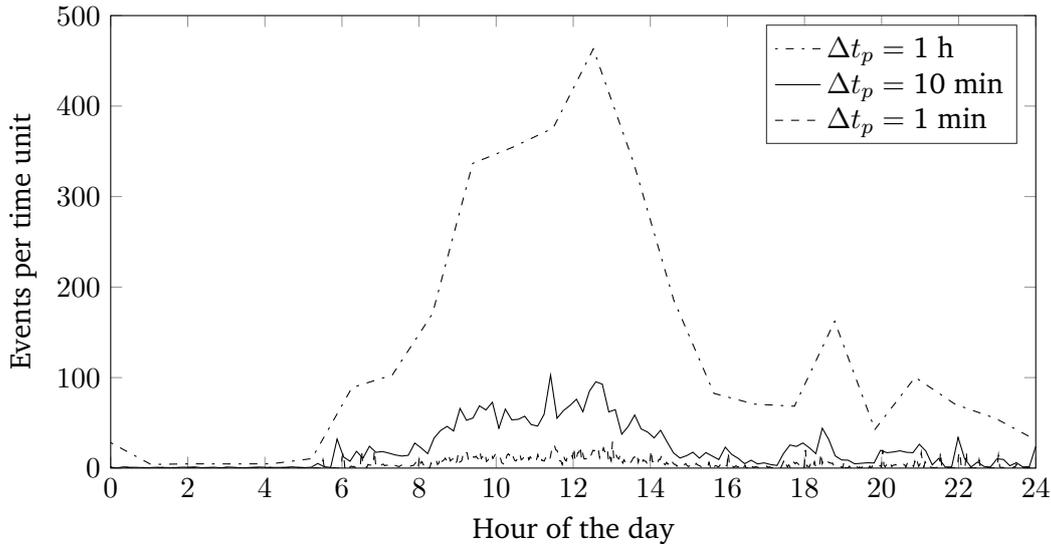


Fig. 7.1: day profiles for different intervals ( $N_p = 17$ )

multi-family house in Germany. The two married residents have both reached retirement age. Nevertheless, they are active members of the village’s community and they live an independent life. The data set contains voltage and current traces of three electric phases for a period of approximately 6 months, from August 2013 to March 2014. Activity information can be extracted for 161 days. A summary of all data is given in Table 7.1.

The switching events are detected using the Steady State Approximation event detection algorithm, that was presented in section 5.4 on page 71. It has proven to be the best compromise between sensitivity (true positives) and oversensitivity (false positives).

Table 7.1: data available for activity profile analyses

data set	residents	$N_p$
UK-DALE, house 1	student(s)	658 d
UK-DALE, house 2	student(s)	142 d
UK-DALE, house 5	student(s)	131 d
ReSP data set VI	2 retirees (f+m)	161 d
	total	1,092 d

### 7.1.3 Raw Activity Profiles

For a first investigation, all available days of own data set VI are averaged. The profile is indicated by the solid line in Fig. 7.2. At night-time (approximately from 22:00 to 05:30), autonomously operating appliances, such as refrigerators, telephone systems, or the central heating<sup>42)</sup>, are active alone. It can be seen that their influence is minimal in this household.

From this profile it can be inferred that the residents usually get up between 06:00 and 07:00. The first rise of activity in Fig. 7.2 is caused by the morning procedures, i.e. bathroom and breakfast activities. The following global maximum at around 09:00 could be caused by cleaning the apartment. The next local activity maximum at approximately 11:00 results from preparing lunch. After that, the average activity drops steadily down to almost none after 20:00. There seems to be no distinct evening activity, that is reflected in the operation of electric appliances.

<sup>42)</sup>The ReSP data set covers late summer, fall, and winter so all heating situations are included.

### 7.1.3.1 Comparison with the Standard Power Demand Curve

The Standard Power Demand Curve (SPDC) is a daily profile, that is used by electricity generating utilities to predict the grid's power demand in order to operate their power stations in the most economic and efficient way. The dashed line in Fig. 7.2 marks the SPDC for German private homes on a work day in the transition period between March 21 and May 14, as well as September 15 and October 31 [148]. Comparing the average activity profile with the SPDC reveals considerable differences. The first reason for this is that, obviously, the residents' daily life pattern regarding their electric appliances is atypical. This is dominant in the evening hours, where most working people return home and cause a lot of power demand. This is an important insight, which makes it clear that appliance switching profiles reveal critical personal information. They have to be interpreted for each person individually, and special care needs to be taken to protect the data from unauthorized access. Secondly, it has to be considered that power demand and switching activity do not necessarily result in the same curves. An appliance may have a high power demand but no significant activity when it is operated continuously. An example is a water pump, which is required for the circulation of warm water through a centrally heated building. On the other hand, a lot of activity results from comparably low power appliances, like freezers or refrigerators, which perform several cooling cycles per day.

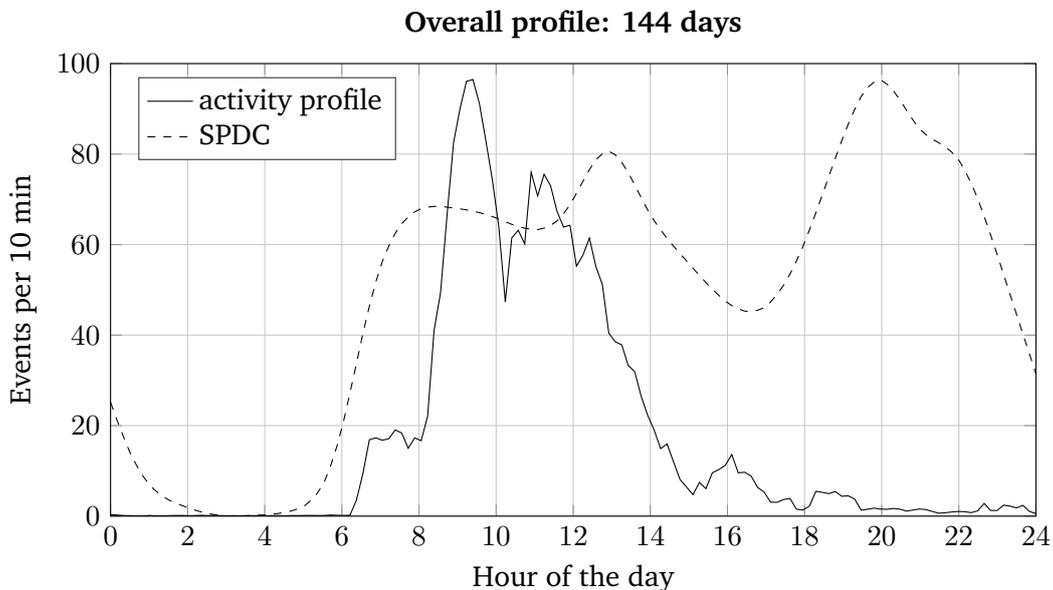


Fig. 7.2: average activity profile for ReSP data set VI

### 7.1.4 Profile Classification

Human life in general has many different rhythms and characteristics. When analyzing the activity at home, the life of an individual is expected to be different under various conditions, e.g. weekends and work days. In order to find out what to expect as “normal” profiles, i.e. daily patterns that represent ordinary days in a person's life, an unsupervised learning technique is applied. The goal is to discriminate and group similar daily profiles. In order to do this, each sample of a profile is treated as one dimension in an  $N_b$ -dimensional feature space. Note, that here, in contrast to section 6.5.4, this is a valid assumption, because each sample of  $\vec{p}_i$  represents a given time of the day, and there are no time shifts. Thus, standard clustering methods can be applied. Assuming that the profile clusters conform to multivariate normal distributions, the k-means clustering technique is applied here. An issue of the k-means algorithm is that the number of clusters,  $\kappa$ , has to be determined before the actual clustering. It affects both the algorithm performance and the clustering result [149]. Here, a visual inspection method is adopted. A

clustering run will result in  $\kappa$  clusters being filled with  $N_j$ ,  $j = 1, \dots, \kappa$  profiles, respectively. If, at a certain value of  $\kappa$ , the minimum

$$N_{\min} = \min_{j=1, \dots, \kappa} \{N_j\} \quad (7.3)$$

is 1, only one profile is assigned to a cluster. This means that overfitting of the features is beginning. At this point, the optimal value of  $\kappa$  is assumed to be found.

Using all days in data set VI, it can be observed in Fig. 7.3 that the minimum number of profiles per cluster drops with increasing  $\kappa$ . The minimum of 1 is reached for  $\kappa \geq 9$  clusters.

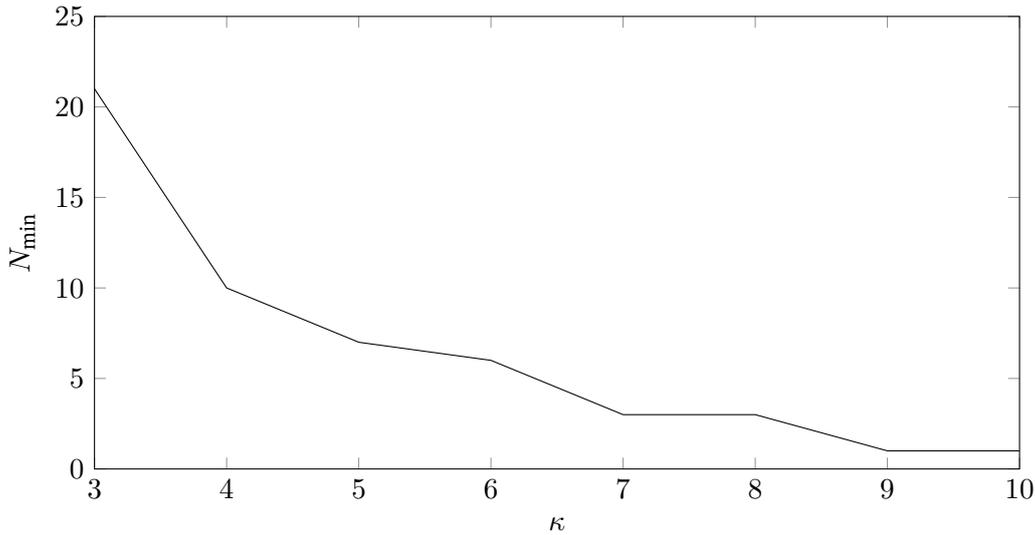


Fig. 7.3: effect of  $\kappa$  on the minimum number of profiles per clusters

### 7.1.5 Clustered Activity Profiles

In this section, clustered profiles of raw switching events are analyzed. The goal is to estimate what information can be inferred from this type of representation. As described in the previous section, the daily profiles are grouped into 9 clusters after event detection.

#### 7.1.5.1 Scenario 1: Senior Residents

For the ReSP data set, the average profiles are shown in Fig. 7.4. Each subplot represents an activity profile type. Each subplot's header shows the number of individual traces that were assigned to the corresponding type. The vertical axes are all limited to the same range in order to allow for a clear comparison.

The first thing that needs to be addressed is the exceptionally high number of events, and the fact that approximately 78 % of all daily curves are assigned to profile a). All profiles except number a) have a maximum of more than 500 events/10 min. It turned out that one single appliance is responsible for these exceptionally high maxima. It causes many switching events within a short period of time, although it is switched on only once every few days. Since this appliance is active on phase 3 only, it can be eliminated by removing this whole phases from the studies. The resulting clustered profiles are shown in Fig. 7.5. It is obvious that the resulting profiles are very different from the previous ones, which were obtained for phases 1 to 3. This makes it clear that raw activity profiles have to be interpreted or verified automatically or manually, before inferring any information on the residents' lives. From the reduced data set, using only phases 1 and 2, the following observations can be made.

1. The magnitudes of 10 to 30 events/10 min are reasonable numbers. As an example, it will be shown in chapter 7.2 that electric stoves periodically turn on and off their heating

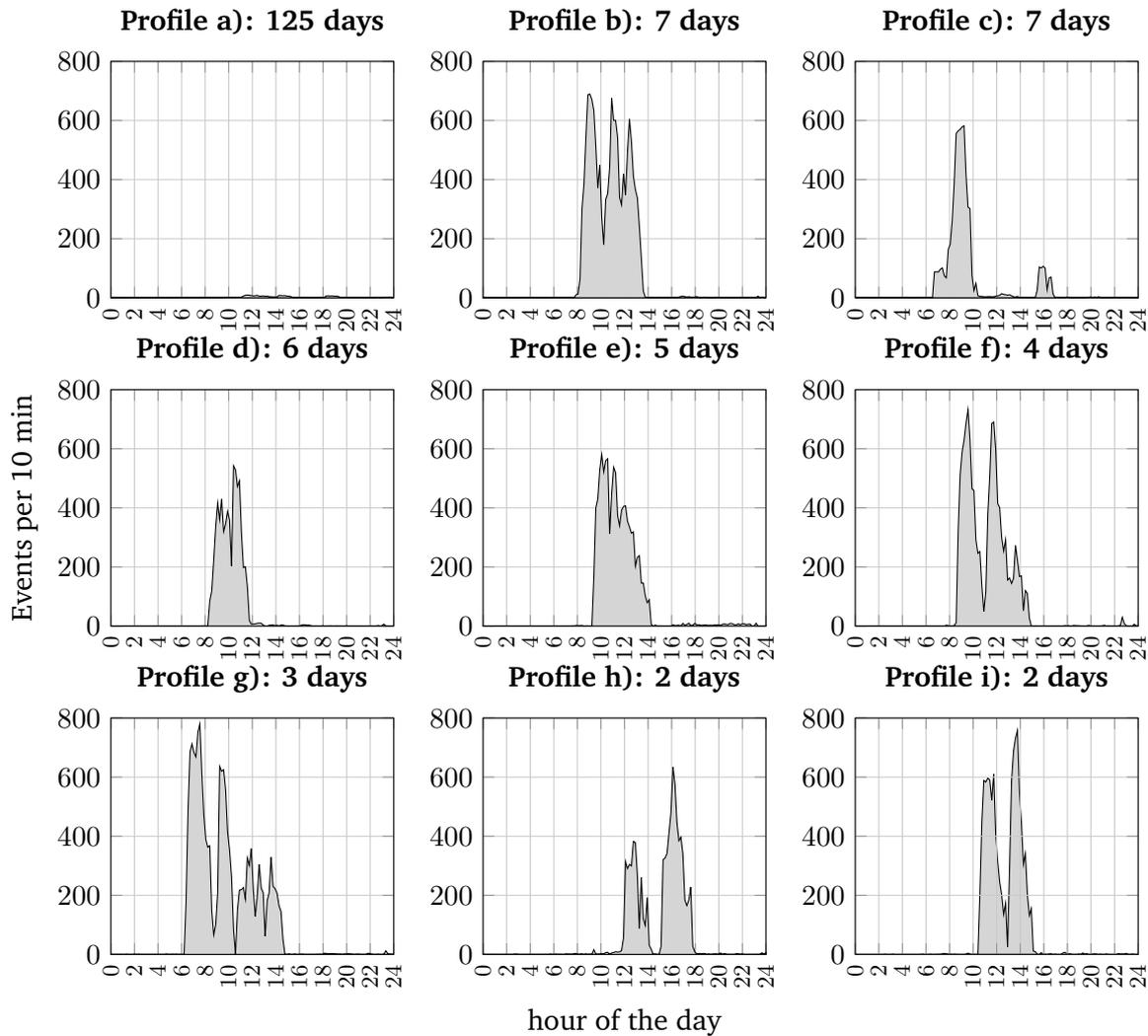


Fig. 7.4: clustered daily profiles for the own data set VI

elements with period times in the range of 30 s. They alone can generate up to 20 events per 10 min.

2. The most common profile (number a), 58% of all days) in Fig. 7.5 is comparably flat, even around lunchtime. Of the remaining, profiles b), c), d), and e) are the ones to which more than one day was assigned. Since their maxima are all located around noon, it can be concluded that the residents are not in their apartment in a considerable amount of time, and that they do not prepare hot meals for lunch very often. This suspicion can be confirmed by filtering the profiles for involved appliances. This will be shown in section 7.1.5.2.
3. After removing the events on phase 3, the overall activity curve, i.e. the expectation of all daily profiles, shows a closer resemblance of the Standard Power Demand Curve in Fig. 7.6. As before, the pattern in the evening is atypical with respect to the SPDC, which expresses the residents' individual life patterns.
4. The sequence of cluster indexes is given in Fig. 7.7. This shows that there is no evidence of a recurring pattern<sup>43)</sup>. The residents' life seems to be irregular with respect to e.g. week-weekend patterns.

<sup>43)</sup>Note that in October, November, and December, there are some days missing. Here, the measurements were invalid due to failing measurement equipment.

5. Profiles f), g), h), and i) are real exceptions, with only one day per profile, each. Together, they make up 2.5% of all data. These can be initial points for a discussion of irregular residents' activities.

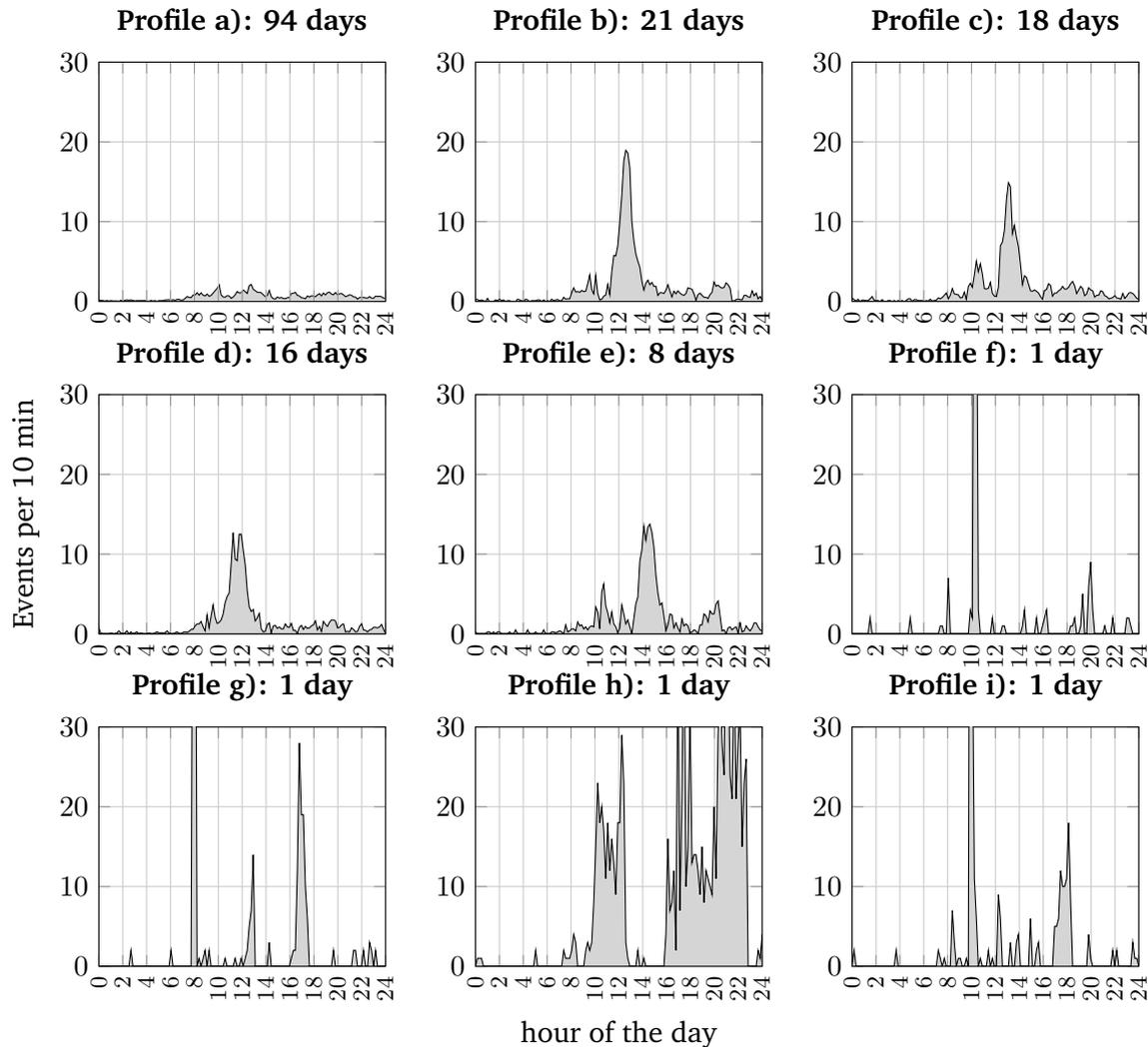


Fig. 7.5: clustered daily profiles for the ReSP data set VI (phases 1 and 2 only)

### 7.1.5.2 Appliance Specific Activity Profile

In the previous subsection, it was suspected that the electric stove is used in less than half of the time in the senior residents' home. In order to verify this, the switching events are classified using TSFP features. Although no training data is available, it is safe to assume that the stove will be standing out from all other appliances due to its power consumption. After eliminating all events except those supposedly caused by the stove, the daily profiles shown in Fig. 7.8 are obtained. Of the 161 total days, 89 (55%) are left out because the stove was not active at all. They are not shown in the graphs. On another 17 days (11%), it was active for short periods of time, which is why profile a) has a magnitude less than 1 and there is no distinct maximum. Profiles b), c), d), and e) are similar to Fig. 7.5, both with respect to their shape and clustering frequency. This confirms the suspicion, that was expressed in the beginning of this subsection.

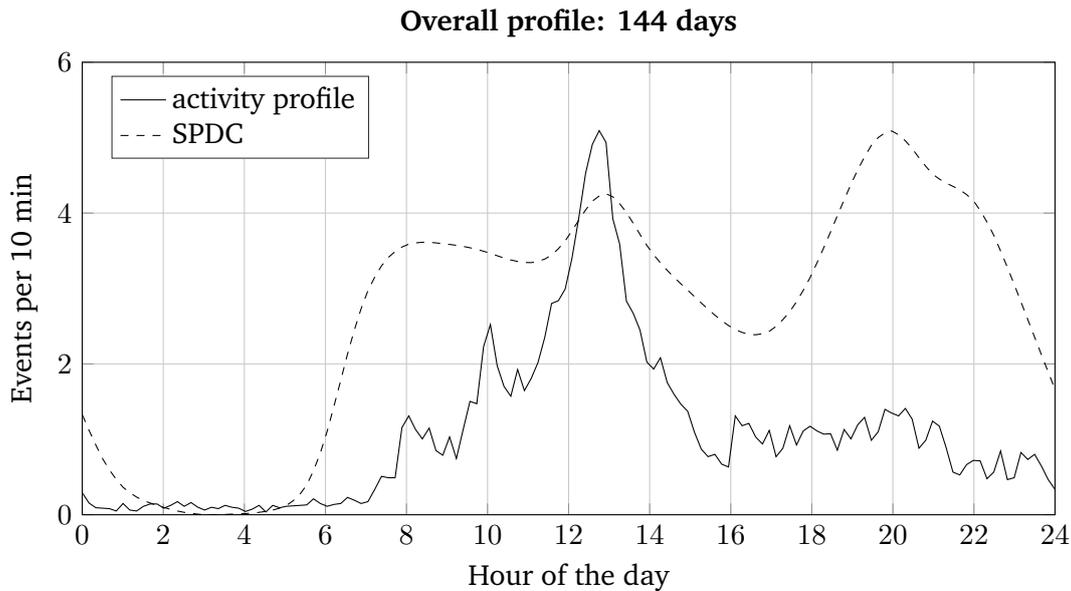


Fig. 7.6: average activity profile for the ReSP data set VI (phases 1 and 2 only)

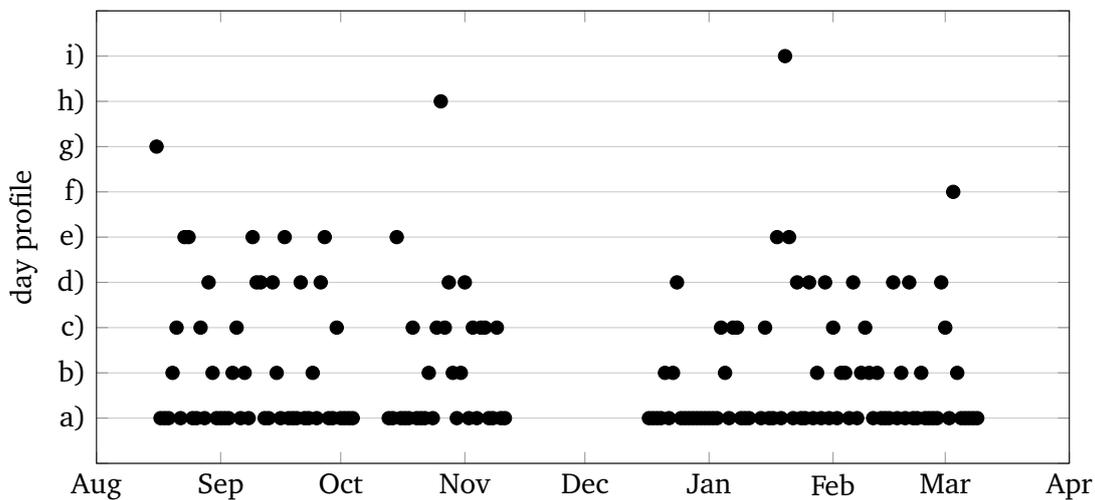


Fig. 7.7: cluster index sequence for the ReSP data set VI (phases 1 and 2 only)

**7.1.5.3 Scenario 2: Student Residents**

The clustering results for the available houses 1, 2, and 5 of the UK-DALE data set are shown in Fig. 7.9, 7.11, and 7.13, respectively. The corresponding cluster index sequence plots are plotted in Fig. 7.10, 7.12, and 7.14 for house 1, 2, and 5, respectively.

In all three UK-DALE homes, the most common profile is one that has no distinct accumulation of events, and the overall number of events is lower than in the other profiles. In this concern, the two scenarios are similar. Differences occur on days with higher rates of activity. For house 1 in Fig. 7.9, there are 8 profile types with similar number of occurrences, but different times of their maxima. Together, they make up 52% of all days. The fact that each of these profiles has only one distinct global maximum indicates that the student uses the apartment only for short periods of time. Furthermore, the cluster index sequence in Fig. 7.10 indicates that the student’s life is very irregular.

The situation is similar in house 2 in Fig. 7.11. However, the low activity profile a) is much more dominant. The days showing activity (Profiles b) to i)) make up only 24% of the total data volume. Consequently, the sequence plot in Fig. 7.12 shows less exceptions from profile a).

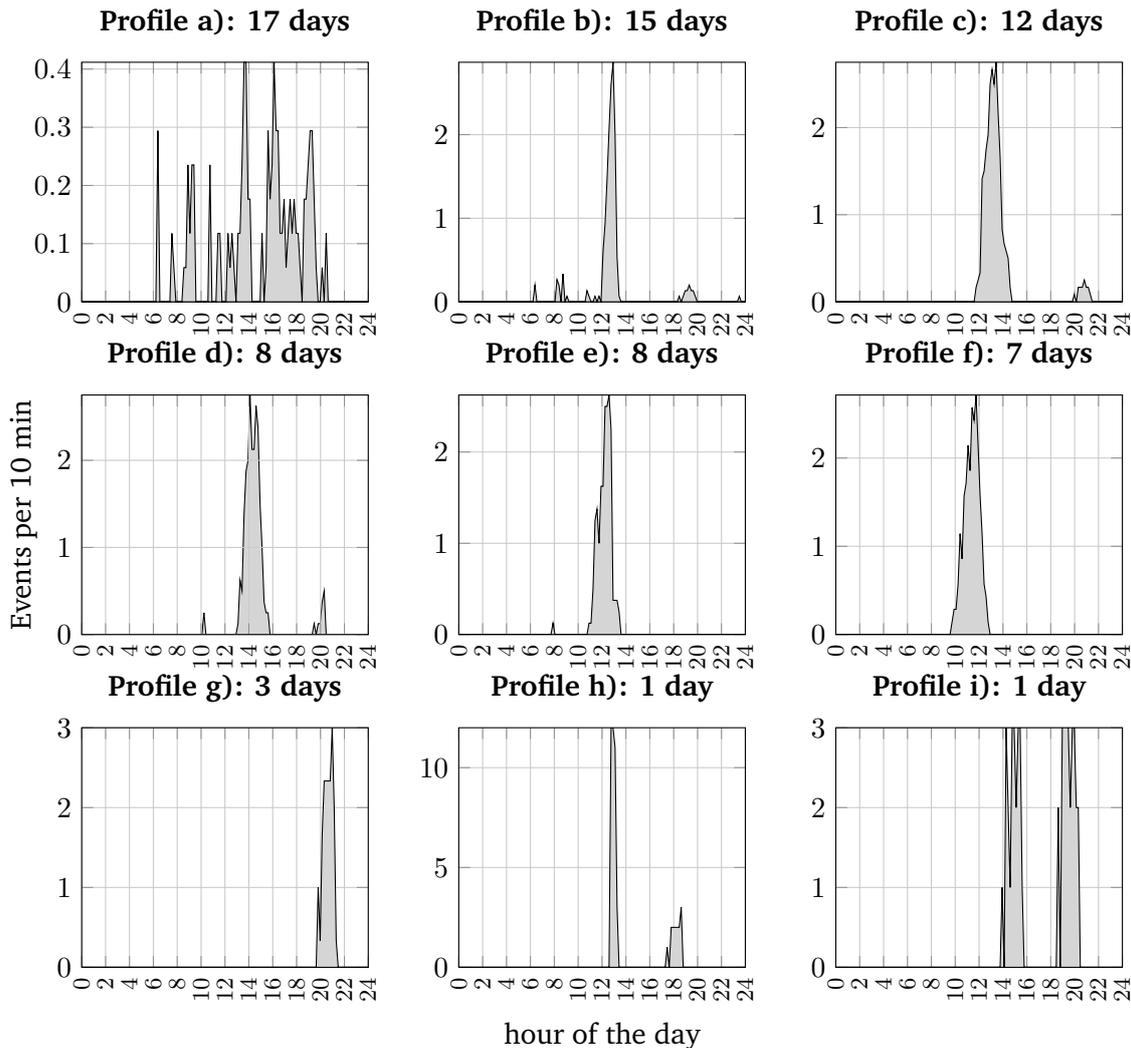


Fig. 7.8: daily profiles of the stove in ReSP data set VI

In house 5, on 44% of all days a clear-cut profile can be found. In the sequence of profile indexes in Fig. 7.14, a hint of recurring structures can be discovered. However, this is not sufficient for a deduction of a regular pattern in the student's life.

### 7.1.6 Influence of False Positives

In chapter 5, it was shown that false positives are an event detection phenomenon that cannot be completely avoided. The goal in this section is to check what influence false positives have on the daily activity profiles. To do this, the switching events are detected with the WDE method, that was introduced in section 5.2. This produces considerably more false positives than the SSA method. For the ReSP data set, the 9 average profiles are shown in Fig. 7.15. It is immediately evident that the absolute number of events is much higher than in Fig. 7.5. Although there is some characteristic shape in some of the profiles, lots of the original information is lost due the high number of false positives.

The situation is different for the UK-DALE measurements. Here, the false positives do not lead to a drastically increased number of events, which means that the measurement is less disturbed than the ReSP data set. The clustering results are slightly different from the corresponding SSA results, but the fundamental profile types are comparable. The corresponding graphs can be found in chapter C in the appendix.

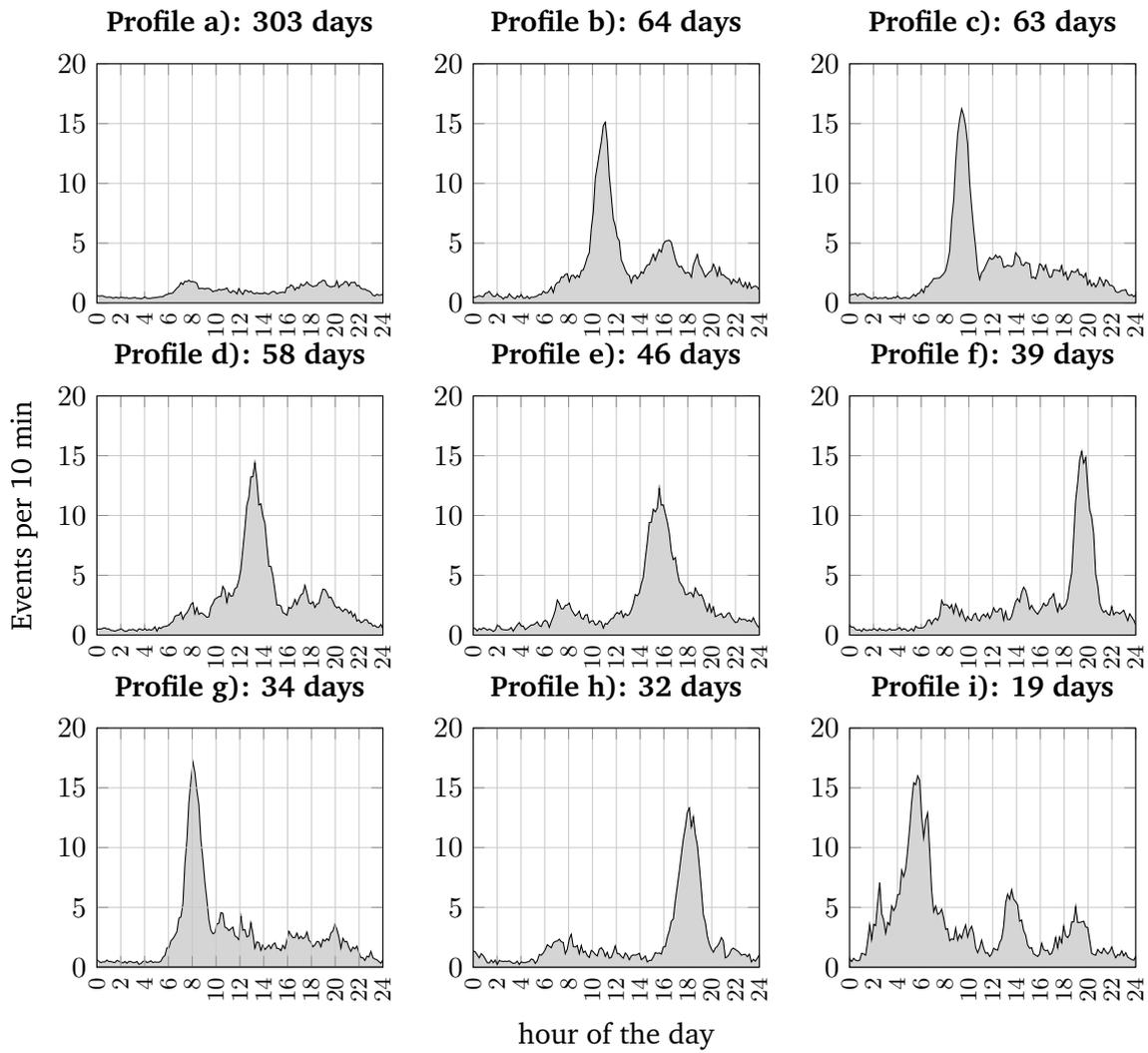


Fig. 7.9: clustered daily profiles for the UK-DALE data set, house 1

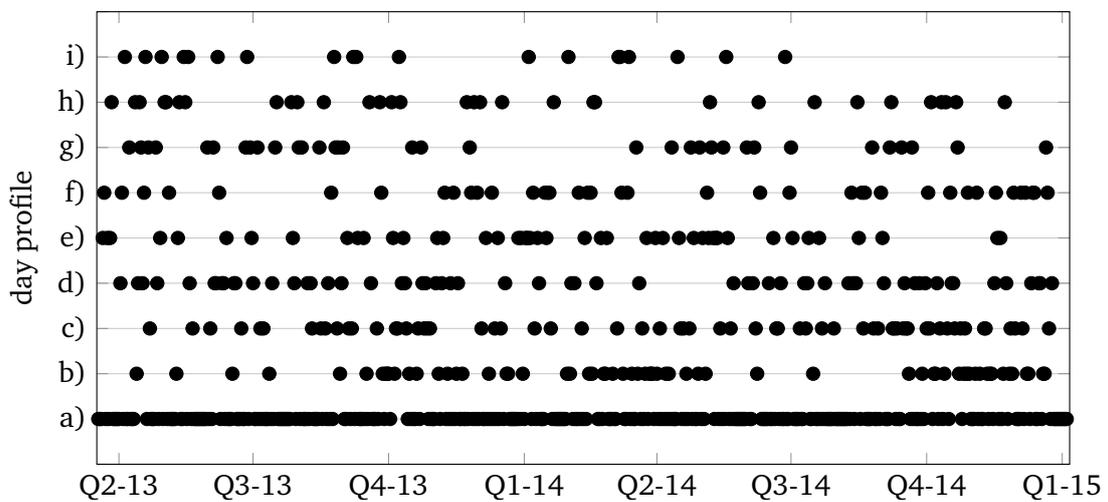


Fig. 7.10: cluster index sequence for the UK-DALE data set, house 1

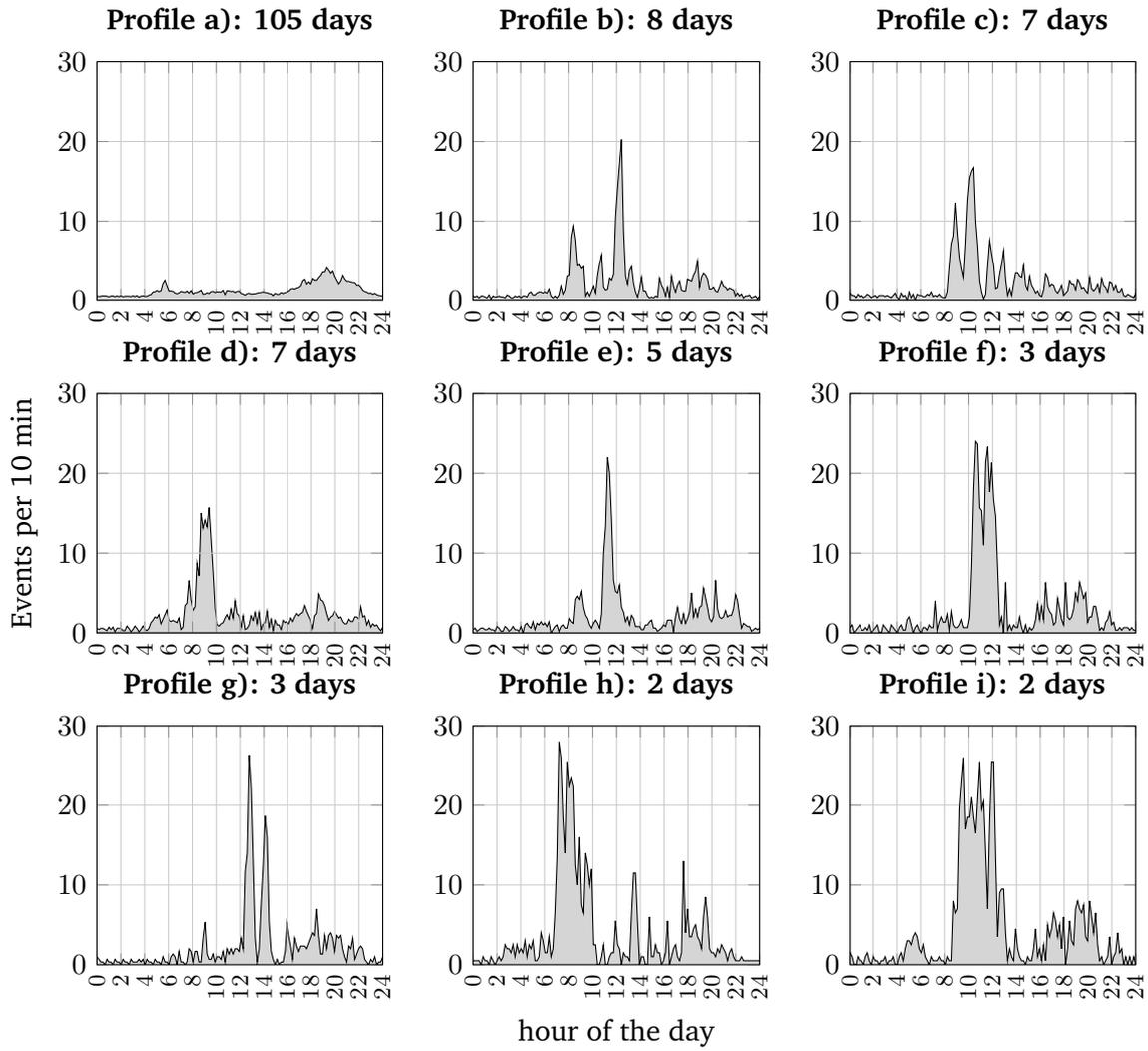


Fig. 7.11: clustered daily profiles for the UK-DALE data set, house 2

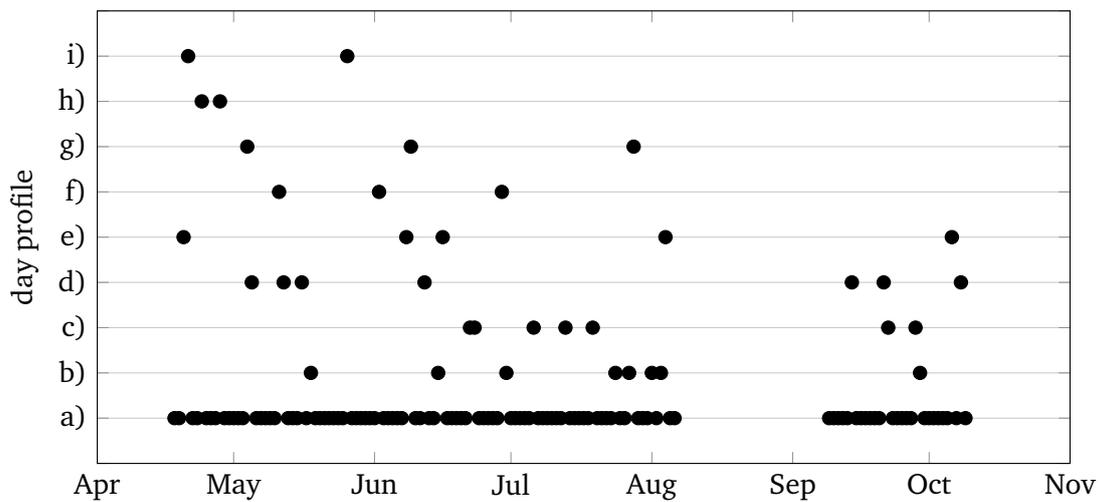


Fig. 7.12: cluster index sequence for the UK-DALE data set, house 2

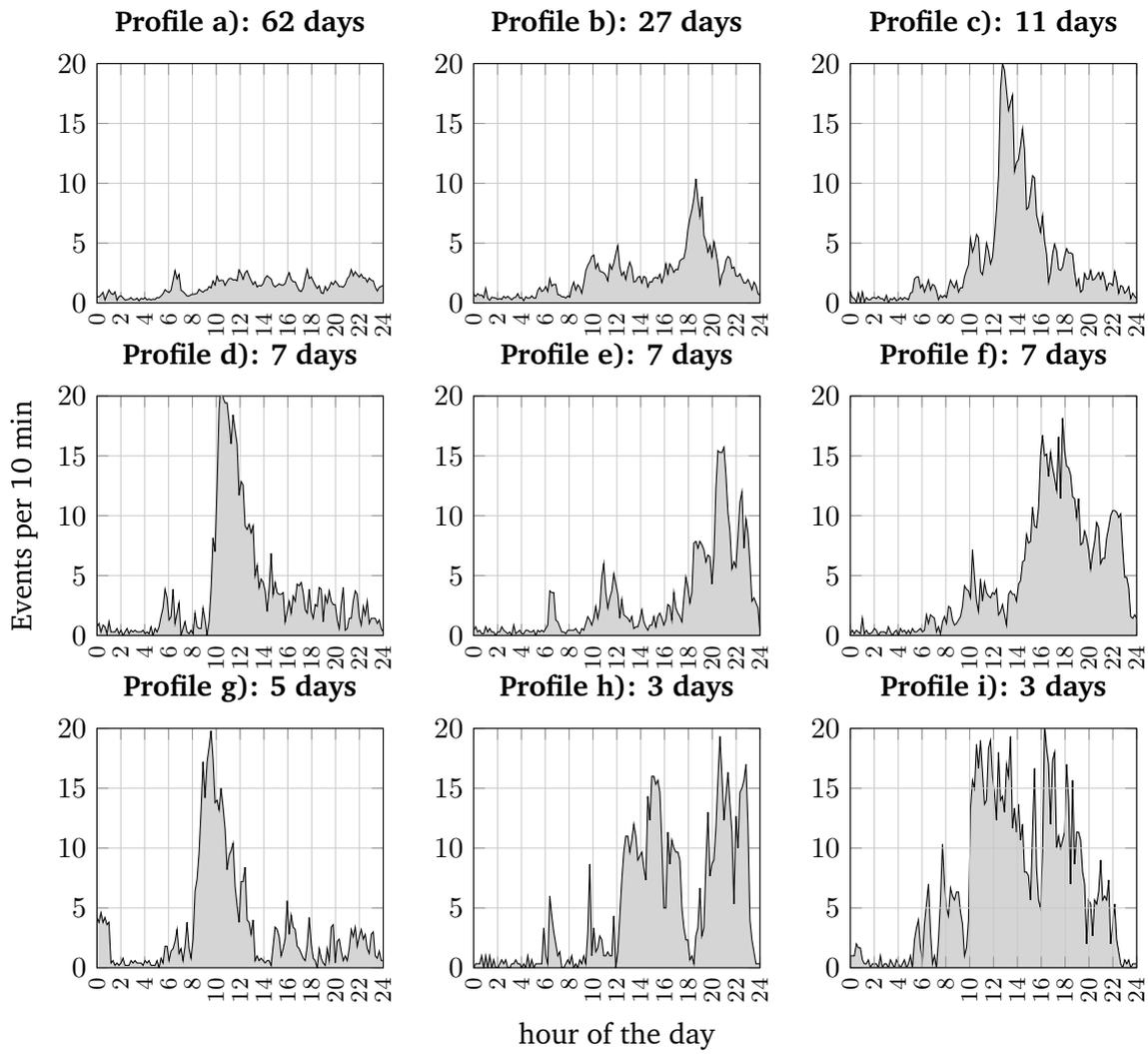


Fig. 7.13: clustered daily profiles for the UK-DALE data set, house 5

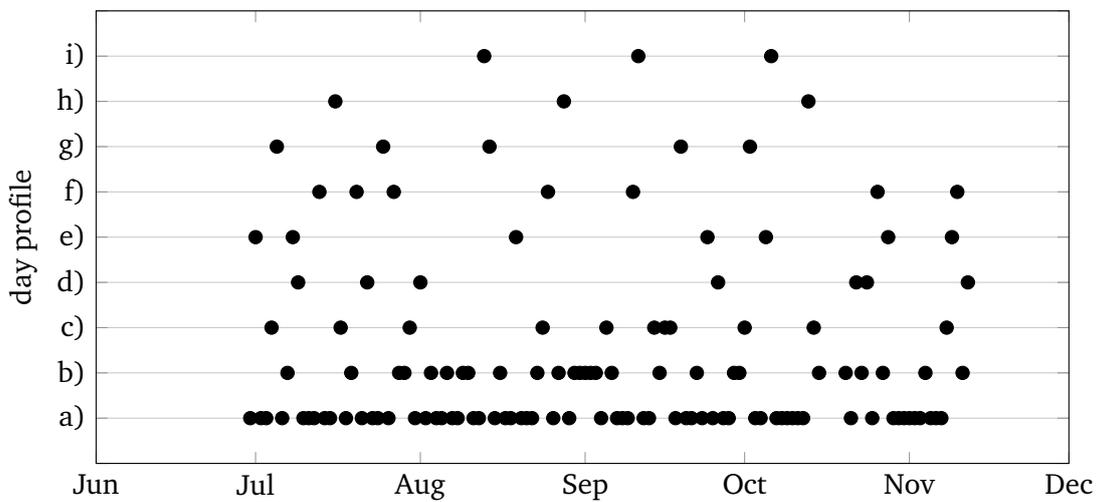


Fig. 7.14: cluster index sequence for the UK-DALE data set, house 5

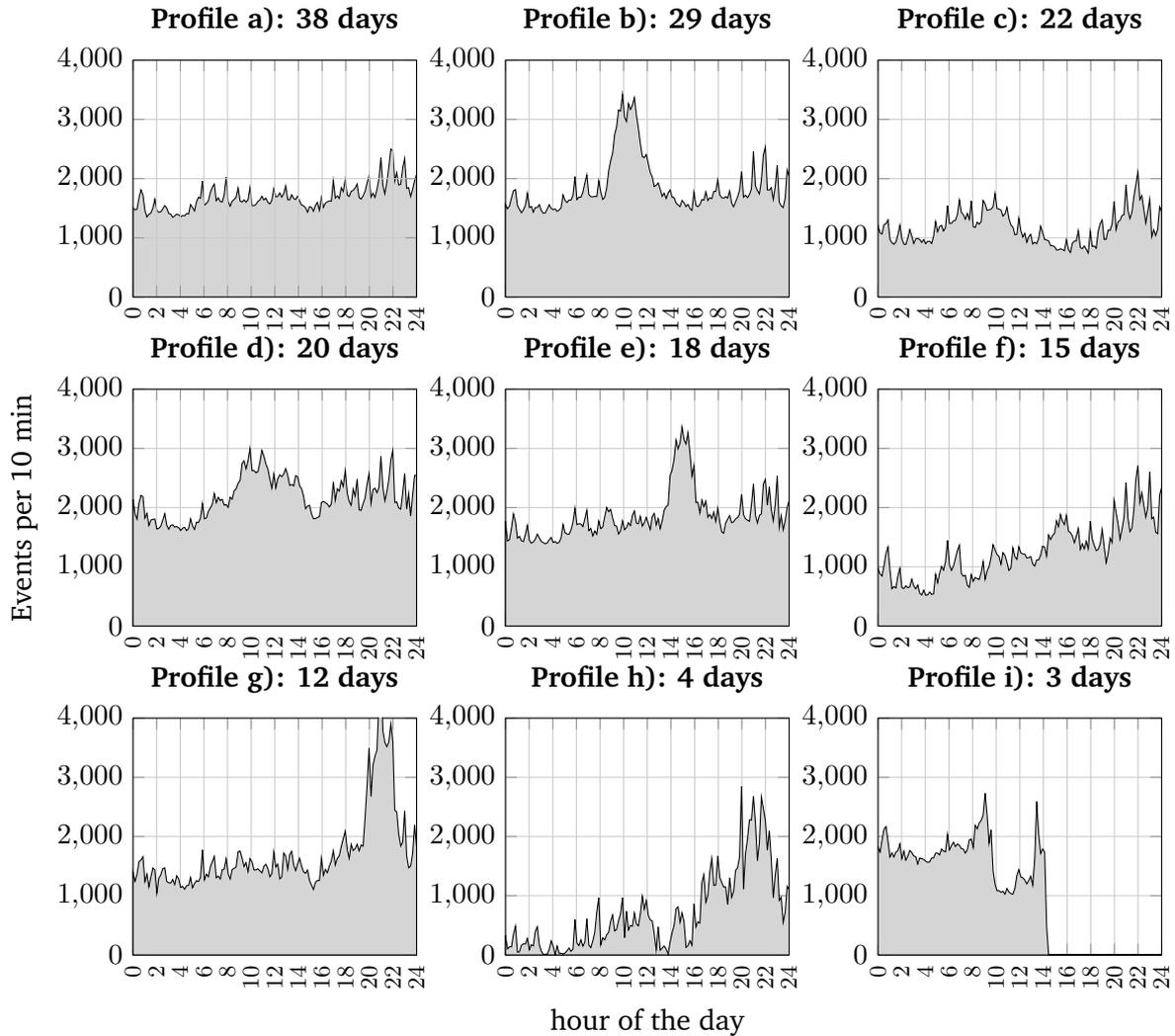


Fig. 7.15: clustered daily profiles for the ReSP data set VI using the WDE event detector (phases 1 and 2 only)

## 7.2 Detecting Overheating Heaters in Load Monitoring Data

In this section, a case study is presented to show how NILM can be used to detect possibly hazardous situations caused by the overheating of electric stoves. For the experiment, a common cooking scenario is considered. A pot of cold water is heated and boiled until all water is evaporated. Once the pot is empty, the temperature of the hotplate increases due to the reduced heat dissipation. This poses a threat to people and the surrounding kitchen furniture by an increased amount of heat spreading into the surroundings.

### 7.2.1 Model

Most stoves control the temperature of their hotplates and the amount of dissipated heat by periodically turning on and off the heating element, according to its temperature. See Fig. 7.16 for a real data example. The heating element is switched on for about half a minute before it is idle for a few minutes.

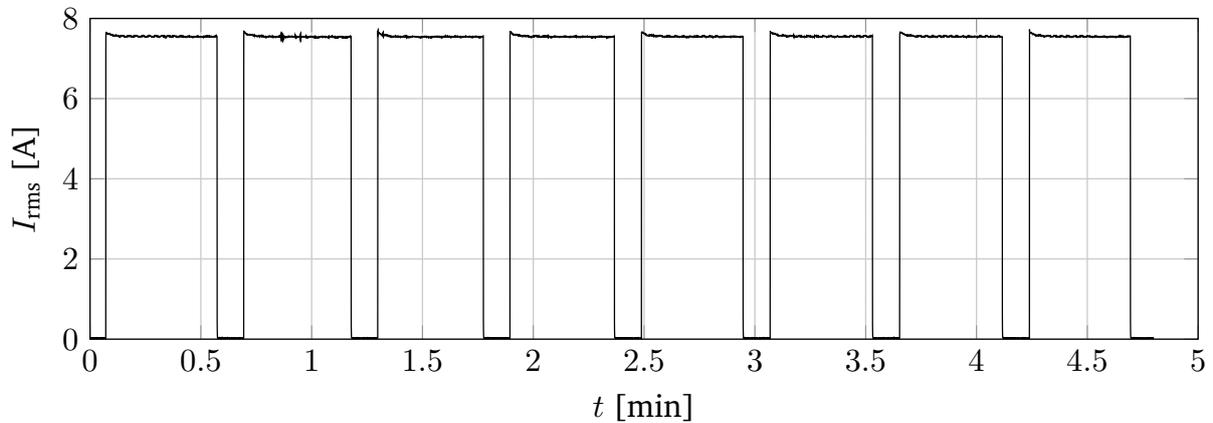


Fig. 7.16: excerpt of a real current signal of a stove heating a pot of water

The resistance of a conducting material depends on its temperature. Assuming a constant temperature coefficient  $\alpha$ , the electric resistance is

$$R(\vartheta) = R_0 \cdot (1 + \alpha \cdot (\vartheta - \vartheta_0)), \quad (7.4)$$

where  $\vartheta$  is the temperature in  $^{\circ}\text{C}$ ,  $\vartheta_0 = 20^{\circ}\text{C}$ , and  $R_0$  is the resistance at  $\vartheta_0$  [150]. During operation, a heating element is supplied with electric energy. A part of this energy is dissipated into the pot, which passes the energy on to the water and heat radiation. At a certain temperature the system reaches equilibrium, i.e. the amount of energy added is equal to the amount dissipated. The electric current drawn during this process can be modeled by an exponential function [151]

$$i(t) = i_1 \cdot e^{-\frac{t}{\tau}} + i_0, \quad (7.5)$$

where  $\tau$  is the time constant. Here, it is assumed that the grid voltage is constant (or normalized, like shown in section 6.2.1 on page 80).  $i_0$  is the current that flows under equilibrium conditions. The sum  $i_0 + i_1$  is the current that flows right after the heating element is turned on. For the experiment in this chapter, this model is fitted to each operation cycle of the hotplate. Two additional time parameters,  $\Delta t_{\text{op}}$  and  $\Delta t$ , are used to measure the active time of the heating element, and the time between consecutive turn-on events, respectively. A sketch of the parameters is given in Fig. 7.17.

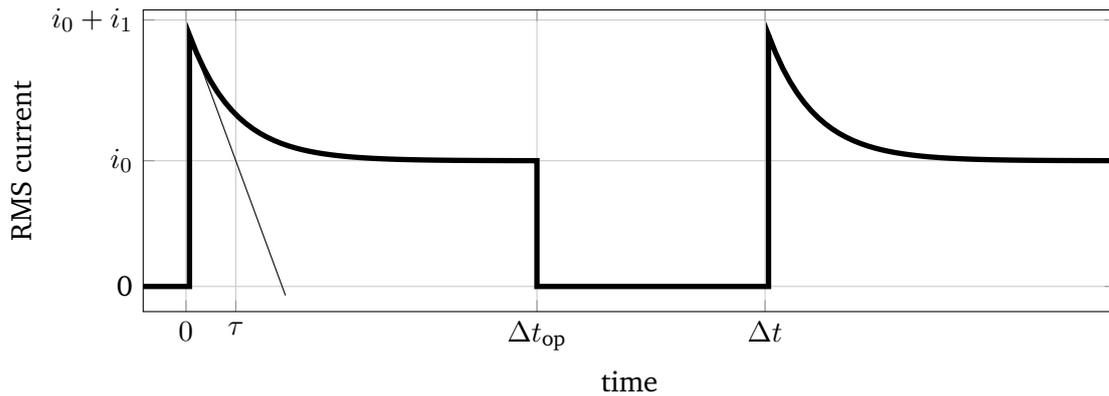


Fig. 7.17: idealized model of the operation of an electric stove with model parameters

## 7.2.2 Results

The parameters found for stove manufactured by Bosch are shown in Fig. 7.18. Both, the time between turn-on events and the operation time of the heating element, change at  $t \approx 22$  min (Fig. 7.18d)), which indicates the time at which the water in the pot was completely evaporated. By reducing  $\Delta t_{op}$  and  $\Delta t$  the device reacts to the reduced heat dissipation with the goal to keep the temperature constant. With the Bosch stove, these time parameters are therefore a good indicator for overheating, if the “normal” operating range is known from training data. The current  $i_0$  (Fig. 7.18b)) drops slightly throughout the whole measurement but there is no obvious reaction to the pot running dry. This means that this particular stove controls its temperature very well to minimize the risk of fire. The parameters  $i_1$  (Fig. 7.18c)) and  $\tau$  (Fig. 7.18a)) do not show any reaction to the changed operating conditions at  $t \approx 22$  min.

The situation is different with a different stove model, shown in Fig. 7.19. The  $\Delta t$  parameter (solid line in Fig. 7.19d)) of the second model drops only slightly once the pot is empty at  $t \approx 30$  min. The active time of the heating element,  $\Delta t_{op}$ , experiences a slight decrease, not as high as the Bosch stove.  $i_1$  (Fig. 7.19c)), on the other hand, shows a distinct increase after this event. This means that this specific model has is less successful in controlling its temperature and the hot plate is getting hotter under extreme conditions. Both remaining parameters,  $\tau$  (Fig. 7.19a)) and  $i_0$  (Fig. 7.19b)), show only a slight change at the overheating event. Overall, with this model, the overheating condition can be better detected by evaluating the peak current, rather than the time parameters  $\Delta t_{op}$  or  $\Delta t$ .

## 7.3 Conclusion

In this chapter, the feasibility of pattern discovery for Activity Monitoring using appliance switching events was shown. Using the k-means clustering algorithm, it is possible to discriminate several different types of daily profiles. This is an objective base for a retrospective discussion of resident activity and health. The discussion could lead to some profiles being declared normal or abnormal, which, using a classifier, can be automatically detected. Based on electricity data from senior and student residents, it was shown that activity profiles are personal information and that the curves assume different shapes for different people. A matching of activity profiles and ADLs can be used in the future to achieve deeper insights into personal habits. It can be expected that this requires fusion with other (possibly more intrusive) monitoring data sources.

Furthermore, a hazardous situation, the overheating of electric stoves when no load is present, was investigated. The analysis showed, that it is possible to detect the danger through data that is available on a smart meter. Although limited to specific parameters of some models, this is an interesting possibility that can be implemented and marketed as an extra feature in a smart meter with NILM.

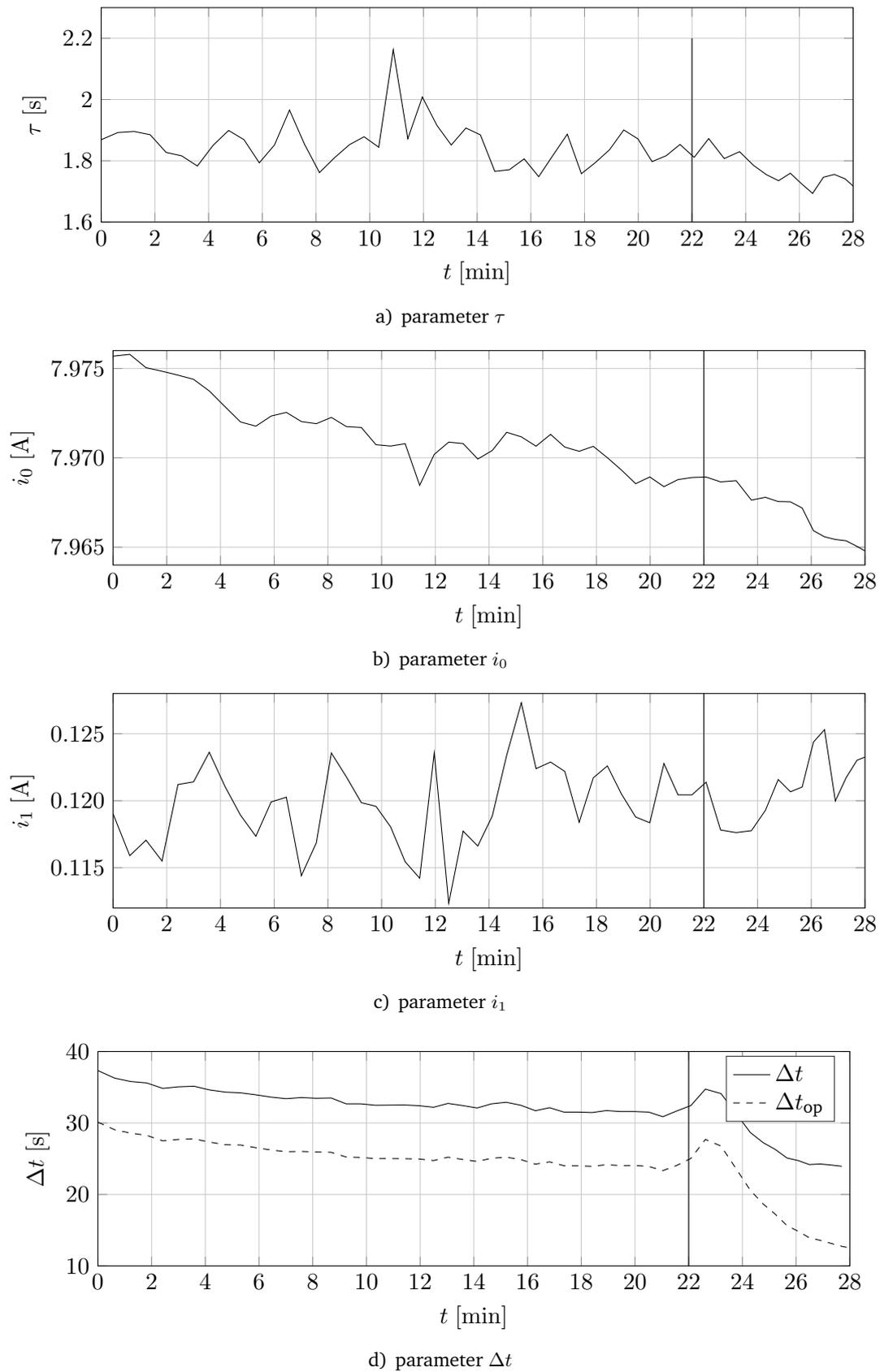


Fig. 7.18: operation cycle parameters of a stove while evaporating all water in a pot (model: Bosch HSS 322020)

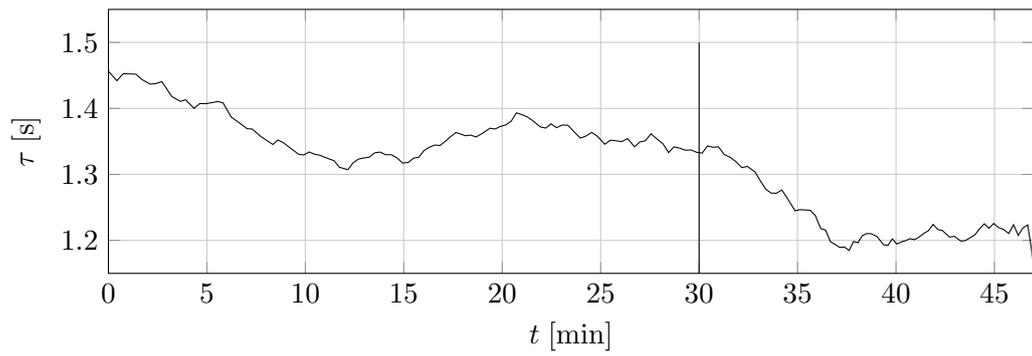
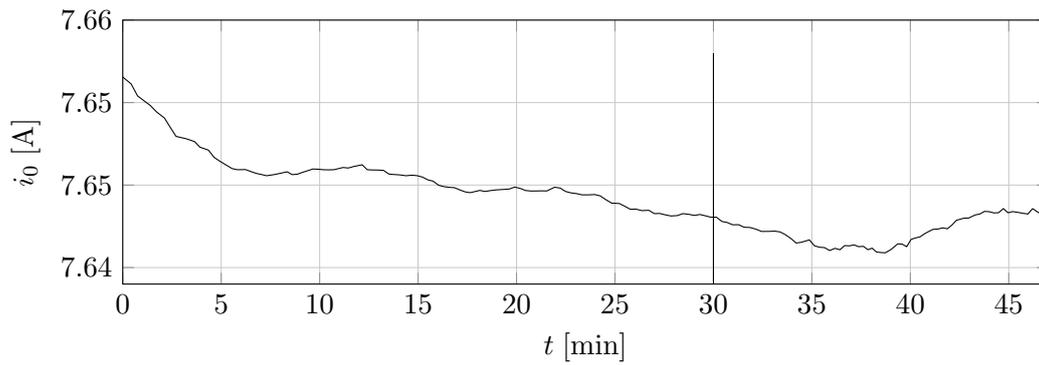
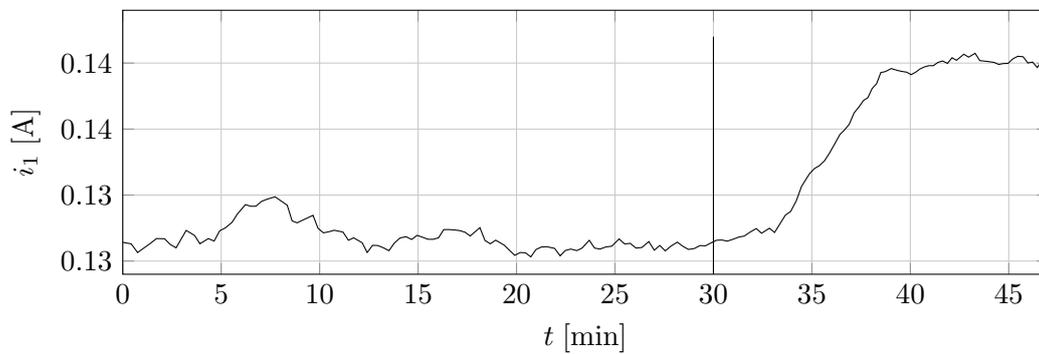
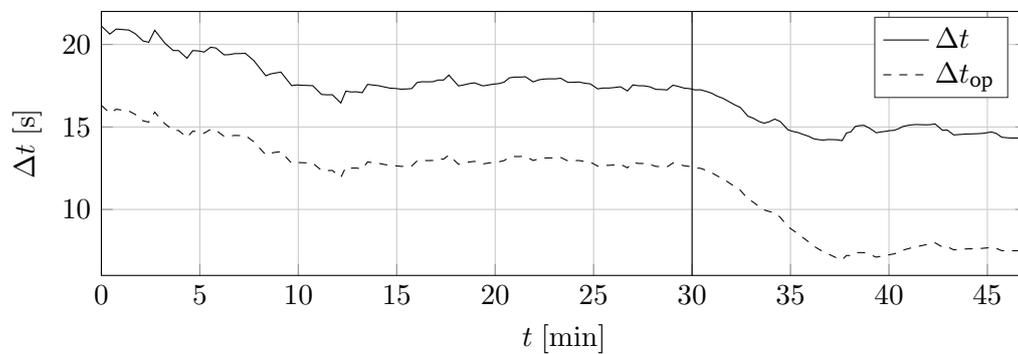
a) parameter  $\tau$ b) parameter  $i_0$ c) parameter  $i_1$ d) parameter  $\Delta t$ 

Fig. 7.19: sliding averages of the operation cycle parameters of a stove while evaporating all water in a pot (model: Gorenje E 54V1-E4)

## Chapter 8 | Conclusion & Perspectives

The automatic analysis of human activity will be a key component for future elderly care services. It helps to reduce costs and provide higher care quality. Activity analysis requires the acquisition of suitable information from the elderly residents' environment, i.e. the home. All technology in the context of elderly care should be designed such that it is of actual use to all people involved and that it helps to reduce pioneering and running costs. This requires the acquisition of information, i.e. the sensors, to be unobtrusive and cheap. Conventional approaches to activity analysis use self-contained sensors like cameras, pressure-sensitive floors, PIR sensors, or specialized sensors for singular pieces of information. Most of these sensors are clearly visible. Therefore, they are an obvious intrusion into the residents' home and privacy. Additionally, their installation requires a lot of work and adjustment, which makes them an inappropriate choice.

In this thesis, two approaches to less intrusive acquisition of information regarding residents' activity were investigated. The first is called *device-free passive human presence detection*. It exploits human interference with electromagnetic waves to detect when a person is present inside a detection zone. The particularly interesting feature of this approach is that detected persons do not have to carry a transponder with them. The field is known from other applications like intruder detection or automatic people counting. A system specifically designed for residential applications in the AAL was presented. It incorporates BLE radio components, which are very cheap and easy to integrate into standard consumer electronic systems. It was shown that, using these components, a simple adaptive derivative thresholding algorithm is suitable to detect human activity in terms of transversing a detection area. Comparable reference algorithms perform worse, because they lack the ability to adapt their detection thresholds.

The disadvantage of the proposed BLE detection system is that it requires dedicated sensors which have to be installed inside the home. Hence, non-intrusive load monitoring was proposed as an even less intrusive approach. In its traditional field, NILM is used to compute the amount of energy each electric appliance in a home consumed. Since this requires the identification of each individual appliance and their active operating periods, and many residential appliance are operated manually, human activity can be deduced from this data. The benefit is that NILM can be implemented in smart meters thus reducing installation effort and intrusion into the actual living environment to a minimum. Traditionally, NILM is a linear sequence of (1) event detection, (2) feature extraction, (3) feature classification, and (4) energy tracking. Parts 1 through 3 were discussed in this thesis; energy tracking is not required for activity analysis.

In the NILM context, event detection tries to detect moments in time in which an electric appliance was turned on or off. It is thus the main source of activity information. A good event detector is able to find all events of any appliance (generally valid appliance model) without producing false positives. Three approaches to event detection were compared in this thesis:

1. *Weighted difference edge detection* approximates the derivative of the electric power signal and applies a predefined fixed threshold in order to find switching events. Its straightforward implementation makes it a low-complexity algorithm that is interesting for low-power computing systems. However, it needs a low threshold in order to detect switching events of low-power appliances. This automatically leads to a very high level of false positives.

2. A more complex algorithm uses an *unsupervised machine learning* approach. It divides a stream of power samples into clusters. The model order of this clustering problem is the indicator of a switching event. This makes this method more sensitive and more robust to false positives. However, experiments showed that there is room for optimization.
3. A new method called *Steady State Approximation* (SSA) was proposed. It iteratively analyzes the slope of the power signal, and it can be implemented in a straightforward way, similarly to the WDE algorithm. In contrast to WDE, the SSA approach yields less false positives by orders of magnitude. It is therefore a significant enhancement of the event detection for NILM.

Feature extraction and classification are the parts of NILM that identify each switching event. For activity analysis, this is required to discriminate manually and automatically operated appliances. The features should be data that corresponds to characteristics of electric appliances. Two types of features were analyzed in this thesis:

1. The first type are *steady state power features* (SSPF). As the name suggests, the steady state power consumption of an appliance is used as its characteristic feature. In order to distinguish linear ohmic from other loads, real and reactive power are used. They are computed by subtracting the power levels after and before a switching event. Using this implementation, it was shown that the classification of these features is prone to errors because some electric appliances show a distinct dynamic settling behavior after switching them on. More reliable features were obtained by applying the SSA method, that was introduced in the event detection part. It allows for an identification of the actual steady state periods and thus an elimination of the dynamic settling phases. This leads to an up to 15 % better classification performance. Three standard classification algorithms were compared with these features. Supervised learning using the kNN algorithm was the best option because it is able to deal with outliers, and it can handle different distributions simultaneously.
2. The disadvantage of SSPF is that a lot of training data is required for successful classification. Thus, the time-series fingerprint approach was proposed. In contrast to the combination of SSA and SSPF, it takes the dynamic settling phase after a turning on event as actual feature; the feature vector is a section of the RMS current signal. It was shown that this type of feature can be used to generate a priori training features, so the on-site training procedure can be greatly simplified. An excerpt of a fingerprint catalog was presented to show positive examples of similarities within different classes of appliances. Finally, it was shown that a sampling rate of approximately 3 Hz is sufficient for these features.

The thesis is completed by two experiments using real-world data. They give an impression of the information that can be inferred from the NILM data. Firstly, the activity of the residents reflects in activity profiles, i.e. daily sequences of activity. Different characteristic profiles can be distinguished automatically by an unsupervised learning method. These daily activity profiles proved to be highly personal information that is characteristic for each resident. Hence, NILM data is a vital source of information for activity analysis. The second experiment showed how a direct threat by electric appliances can be detected in NILM data. When analyzing the operating characteristics of electric stoves, the overheating as a possible hazard can be detected by observing specific parameter values.

For both information acquisition methods, future research should proceed in the following directions:

- A full system development is required in order to equip private homes with non-intrusive activity information acquisition on a larger scale. The development needs to include the development of required smart meters, computing platforms, and data gateways. This is also an important for successful field tests that yield large amounts of useful data for further evaluations.

- The information channels DFP presence detection and NILM data can be investigated further to gain insights into specific personal customs. A guided interpretation of the data and a deeper evaluation can show how future elderly can profit from the data.
- Human activity and customs are of a manifold nature. Activity analysis might profit from a multitude of information sources, so the proposed approaches can be considered a key component of a superordinate data fusion system, that takes various information channels into accounts. This can also help to adjust the overall monitoring to the personal situations of various residents.
- Future research should also address data privacy as a key topic. A distributed elderly care system is only possible if the highly personal data is treated with respect, and confidential transmission, storage, and processing can be guaranteed. It will be an interesting challenge to integrate cloud services without the data leaks that can be observed with other internet-based services.



# Appendix



## Appendix A | Notation, Definitions, and Elementary Functions

Thousands separator is the comma, the decimal mark is the dot. Notation conventions regarding mathematic equations are listed in Tables A.1 and A.2.

The delta function (Dirac pulse) is defined as

$$\delta(t - t_0) = \begin{cases} \infty & x = t_0 \\ 0 & x \neq 0 \end{cases} . \quad (\text{A.1})$$

The discrete-time version is

$$\delta[k - k_0] = \begin{cases} 1 & k = k_0 \\ 0 & k \neq 0 \end{cases} . \quad (\text{A.2})$$

The rectangle function (rect function) is defined as

$$\square(t) = \begin{cases} 0 & |t| > 1/2 \\ 1 & |t| \leq 1/2 \end{cases} . \quad (\text{A.3})$$

Table A.1: notation overview

symbol	details
$x(t)$	continuous-time signal in time domain
$t$	continuous time variable
$x[k]$	discrete-time signal in time domain
$k$	discrete time index
$X(f)$	continuous-frequency signal in frequency domain
$f$	continuous frequency variable
$X[n]$	discrete-frequency signal in frequency domain
$n$	discrete frequency index
$\vec{x}$	vector as a section of a discrete-time signal in time domain
$\vec{X}$	vector as a section of a time-discrete signal in frequency domain
$\mathbf{X}$	matrix
$N$	number of elements in a vector or matrix
$\vec{x}^T$	transposed vector of $\vec{x}$
$\vec{x}(k)$	vector at time instant $k$
$\vec{x}[i]$	$i$ -th element of a vector
$\vec{x}(k)[i]$	$i$ -th element of a vector at time instant $k$
$\langle \vec{x}, \vec{y} \rangle$	$= \sum_{i=1}^N \vec{x}[i] \cdot \vec{y}[i]$ : scalar product of the vectors $\vec{x}$ and $\vec{y}$
$X$	$= \{x_1, x_2, \dots, x_N\}$ : set of $N$ elements
$X(i)$	$i$ -th element of the set $X$
$\underline{x}$	complex quantity
$j$	imaginary unit with $j^2 = -1$
$\Re\{\cdot\}, \Im\{\cdot\}$	real and imaginary part of a complex quantity
$\underline{x}^*$	$= \Re\{\underline{x}\} - j \Im\{\underline{x}\}$ : complex conjugate of $\underline{x}$
$x \propto y$	$x$ is proportional $y$
————	solid line style
-----	dashed line style
-.-.-.-.	dash dotted line style
.....	dotted line style

Table A.2: set &amp; interval notations

symbol	range	details
$(a, b) = ]a, b[$	$= \{x \in \mathbb{R}   a < x < b\}$	
$[a, b) = [a, b[$	$= \{x \in \mathbb{R}   a \leq x < b\}$	
$(a, b] = ]a, b]$	$= \{x \in \mathbb{R}   a < x \leq b\}$	
$[a, b] = [a, b]$	$= \{x \in \mathbb{R}   a \leq x \leq b\}$	
$\mathbb{N}$	$\{1, 2, 3, \dots\}$	set of natural numbers without 0
$\mathbb{N}_0$	$\{0, 1, 2, 3, \dots\}$	set of natural numbers with 0
$\mathbb{N}_{>n}$	$\{n+1, n+2, n+3, \dots\}$	set of natural numbers with a lower limit
$\mathbb{R}$	$[-\infty, \infty]$	set of real numbers
$\mathbb{R}_+$	$]0, \infty]$	set of real positive numbers without 0
$\mathbb{Z}$	$\{\dots, -2, -1, 0, 1, 2, \dots\}$	set of integers

## Appendix B | Discrete Transforms

A discrete transform is a function that maps a real or complex valued signal  $x$  by an  $n$ -tuple, i.e.  $n$  discrete samples. The signal itself can be either in continuous or discrete-time representation. If the signal is represented by a finite number of values  $N \in \mathbb{N}$  the transform is called a *block transform*.

For discrete transforms the signal  $\vec{x}$  is assumed to be from the signal space  $X = \text{span} \{\vec{\varphi}_1, \dots, \vec{\varphi}_n\}$ , i.e.  $\vec{x}$  can be represented by a linear combination

$$\vec{x} = \sum_{i=1}^n \alpha_i \cdot \vec{\varphi}_i \quad (\text{B.1})$$

of basis vectors  $\{\vec{\varphi}_1, \dots, \vec{\varphi}_n\}$  of the signal space. They can be either linearly dependent or linearly independent. The representation of the signal (its *transform*) is the vector

$$\vec{\alpha} = [\alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_N]^T. \quad (\text{B.2})$$

It completely describes  $\vec{x}$  with respect to the basis  $\{\vec{\varphi}_1, \dots, \vec{\varphi}_n\}$ .

### B.1 Orthogonal Series Expansion

In order to compute the representation of an unknown signal  $\vec{x}$  with respect to a basis  $\{\vec{u}_1, \dots, \vec{u}_n\}$ , it is assumed that the signal can be represented by

$$\vec{x} = \sum_{i=1}^n \alpha_i \cdot \vec{u}_i. \quad (\text{B.3})$$

The basis vectors are required to be orthonormal, i.e.

$$\langle \vec{u}_i, \vec{u}_j \rangle = \delta_{ij} \quad \text{mit } \delta_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{else} \end{cases}. \quad (\text{B.4})$$

Using this requirement the representation of  $\vec{x}$  can be computed from eq. (B.3) using the scalar product

$$\langle \vec{x}, \vec{u}_j \rangle = \alpha_i \cdot \langle \vec{u}_i, \vec{u}_j \rangle \quad j = 1, \dots, n \quad (\text{B.5})$$

$$\Rightarrow \alpha_j = \langle \vec{x}, \vec{u}_j \rangle. \quad (\text{B.6})$$

### B.2 Continuous Wavelet Transform

The continuous Wavelet transform is an integral transform that uses a function  $\psi\left(\frac{t-b}{a}\right)$  with to parameters  $a, b \in \mathbb{R}, a > 0$  as a kernel. The Wavelet transform of a signal  $x(t)$  is

$$\mathcal{W}_\psi \{x(t)\} (b, a) = \frac{1}{|a|^{\frac{1}{2}}} \cdot \int_{-\infty}^{\infty} x(t) \cdot \psi^* \left( \frac{t-b}{a} \right) dt \quad (\text{B.7})$$

where  $\psi^*(\cdot)$  is the complex conjugate of  $\psi(\cdot)$  [152]. The kernel function  $\Psi$  is called the *wavelet*.

The Wavelet transform is *affine invariant* because the Wavelet transform of  $x(t)$  remains unchanged when scaling  $x(t) \rightarrow x(\frac{t}{c})$  except for the scaling itself. In addition it is *translation invariant* because a time shift of the signal  $x(t) \rightarrow x(t-t_0)$  causes a time shift of the transformed version of  $t_0$  [153, p. 144].

### B.3 Discrete Dyadic Wavelet Series: The Discrete Wavelet Transform

This section is based on the books by Mallat [154] and Mertins [153]. Details regarding the steps described in this section can be found in these books.

For the discrete Wavelet transform the transition from continuous parameters  $(a, b)$  to discrete values  $(j, k)$  is made;  $j$  and  $k$  are incremented in discrete steps. A special class of wavelets are functions that form an orthonormal basis of  $L_2(\mathbb{R})$ . They are of the form

$$\left\{ \psi_{jk}(t) = \frac{1}{\sqrt{2^j}} \cdot \psi \left( \frac{t - 2^j \cdot k \cdot T}{2^j} \right) \right\}_{(j,k) \in \mathbb{Z}^2}. \quad (\text{B.8})$$

$T$  is the sampling interval. In the following, it will be set to  $T = 1$  for simplicity. Using eq. (B.6) a signal  $x(t) \in L_2(\mathbb{R})$  can be represented by the series

$$\mathcal{W}_\psi \{x(t)\} (2^j \cdot k, 2^j) = \langle \vec{x}, \vec{\psi}_{jk} \rangle. \quad (\text{B.9})$$

Provided that a function system  $\left\{ \tilde{\psi}_{jk}(t) \right\}_{(j,k) \in \mathbb{Z}^2}$  dual to  $\left\{ \psi_{jk}(t) \right\}_{(j,k) \in \mathbb{Z}^2}$  exists,  $x(t)$  can be reconstructed by

$$x(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \langle \vec{x}, \vec{\psi}_{jk} \rangle \cdot \tilde{\psi}_{jk}(t). \quad (\text{B.10})$$

Since both function systems span the same signal space the reconstruction

$$x(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \langle \vec{x}, \vec{\tilde{\psi}}_{jk} \rangle \cdot \psi_{jk}(t) \quad (\text{B.11})$$

is admissible, as well. Let the Wavelet series expansion of  $x(t)$  with respect to the dual wavelet  $\tilde{\psi}_{jk}(t)$  be

$$d_j[k] = \mathcal{W}_{\tilde{\psi}} \{x(t)\} (k \cdot 2^j, 2^j) = \langle \vec{x}, \vec{\tilde{\psi}}_{jk} \rangle, \quad (j, k) \in \mathbb{Z}^2. \quad (\text{B.12})$$

Then, from eq. (B.11), the reconstruction becomes

$$x(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_j[k] \cdot \psi_{jk}(t). \quad (\text{B.13})$$

The functions  $\left\{ \psi_{jk}(t) \right\}_{(j,k) \in \mathbb{Z}^2}$  are linearly independent. Therefore, they span a subspace

$$W_j = \text{span} \left\{ \psi \left( \frac{t - k \cdot 2^j}{2^j} \right), k \in \mathbb{Z} \right\}, \quad j \in \mathbb{Z}. \quad (\text{B.14})$$

These subspaces are related to  $L_2(\mathbb{R})$  via

$$L_2(\mathbb{R}) = \cdots \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus \dots \quad (\text{B.15})$$

For each subspace an unambiguous signal

$$y_j(t) = \sum_{k=-\infty}^{\infty} d_j[k] \cdot \psi_{jk}(t), \quad y_j(t) \in W_j \tag{B.16}$$

can be determined from the representation of  $\vec{x}$ . It is called *detail signal* or short *detail*. The series  $d_j[k]$  are the *detail coefficients* for the resolution  $2^{-j}$ .

Complementary to each  $W_j$  a subspace  $V_j \subset L_2(\mathbb{R})$  is defined such that

$$V_j = V_{j+1} \oplus W_{j+1}. \tag{B.17}$$

Like  $W_j$  each subspace  $V_j$  is spanned by scaled and translated instances of one single function  $\phi(t)$ , the so called *scaling function*

$$V_j = \text{span} \left\{ \phi \left( \frac{t - k \cdot 2^j}{2^j} \right), k \in \mathbb{Z} \right\}, \quad j \in \mathbb{Z}. \tag{B.18}$$

It can be used to compute the *approximation* of  $x(t)$  with resolution  $2^{-j}$  by

$$x_j(t) = \sum_{k=-\infty}^{\infty} a_j[k] \cdot \phi_{jk}(t). \tag{B.19}$$

The  $a_j[k]$  are the *approximation coefficients*.

Each  $j \in \mathbb{Z}$  in eq. (B.9) corresponds to a frequency resolution level  $2^{-j}$ . The approximation of a signal  $x(t)$  with resolution  $2^{-j}$  is the orthogonal projection of  $x(t)$  onto a subspace  $V_j \subset L_2(\mathbb{R})$  [154, p. 221], denoted by  $P_{V_j}(x(t))$ . The set of all subspaces  $\{V_j\}_{j \in \mathbb{Z}}$  is called *multi-resolution approximation*.

To conclude, the most relevant signals and symbols are listed in Table B.1.

Table B.1: conclusion of the parameters for the discrete Wavelet transform

signal/symbol	signal space	description
$x(t)$	$L_2(\mathbb{R})$	Original signal to be decomposed
$x_j(t)$	$V_j$	Approximation signal at resolution $2^{-j}$
$y_j(t)$	$W_j$	Detail signal at resolution $2^{-j}$
$a_j[k]$		Approximation coefficients at resolution $2^{-j}$
$d_j[k]$		Detail coefficients at resolution $2^{-j}$

### B.3.1 Iterative Computation (Mallat Decomposition)

From eq. (B.17) it can be concluded that

$$x_j(t) = x_{j+1}(t) + y_{j+1}(t), \tag{B.20}$$

i.e. the approximation of a signal can be further divided into approximation and detail signals iteratively

$$\begin{array}{ccccccc}
 & & y_1(t) & & y_2(t) & & y_3(t) \\
 & \nearrow & & \nearrow & & \nearrow & \\
 x_0(t) & \rightarrow & x_1(t) & \rightarrow & x_2(t) & \rightarrow & x_3(t) \rightarrow \dots
 \end{array} \tag{B.21}$$

It  $a_0[k]$  is understood as a sampled version of  $x(t)$  which is filtered by  $\phi^*(-t)$  the decomposition in eq. (B.21) can also be applied to the coefficient series

$$\begin{array}{ccccccc}
 & & d_1[k'] & & d_2[k''] & & d_3[k'''] \\
 & \nearrow & & \nearrow & & \nearrow & \\
 a_0[k] & \rightarrow & a_1[k'] & \rightarrow & a_2[k''] & \rightarrow & a_3[k'''] \rightarrow \dots
 \end{array} \tag{B.22}$$

The length of the series  $a_{i+1}$  is decreased by a factor of 2 from  $a_i$ . This iterative computation scheme can be efficiently implemented using a multi-rate system of cascaded conjugated low-pass and band-pass filters. Using the wavelet function  $\psi$  and the scaling function  $\phi$  a signal  $x(t)$  can be analyzed (decomposition into approximation and detail) and reconstructed. This procedure is called the *Mallat Decomposition*.

### B.3.2 Discrete-Time Discrete Wavelet Transform

The theory from the previous subsection can be applied for discrete-time, i.e. sampled signals  $x[k]$ , by discretization of the integral in eq. (B.7). For dyadic scaled and translated wavelets this becomes

$$w_\psi \{x[k]\} (n \cdot 2^j, 2^j) = \frac{1}{\sqrt{2^j}} \cdot \sum_k x[k] \cdot \psi^* \left[ \frac{k - n \cdot 2^j}{2^j} \right]. \quad (\text{B.23})$$

The  $\psi^* \left[ \frac{k - n \cdot 2^j}{2^j} \right]$ ,  $j > 0$ ,  $(k, n) \in \mathbb{Z}^2$  are samples of the continuous-time complex conjugated wavelet function  $\psi^*(t)$ . This means that all properties of the discrete Wavelet transform obtained in the previous subsection can also be applied to the discrete-time domain. The only difference is that the discrete-time version is not translation invariant any more [153, p. 183]. The discrete-time discrete Wavelet transform can be efficiently computed by the *à trous algorithm* which is concordant with the Mallat algorithm [153, p. 187].

## Appendix C | Additional Daily Profiles

Figures C.1, C.2, and C.3 show clustered daily profiles of switching events that are extracted by the WDE method. It is supposed to produce more false positives than SSA. The corresponding section in the main part is 7.1.6 on page 118.

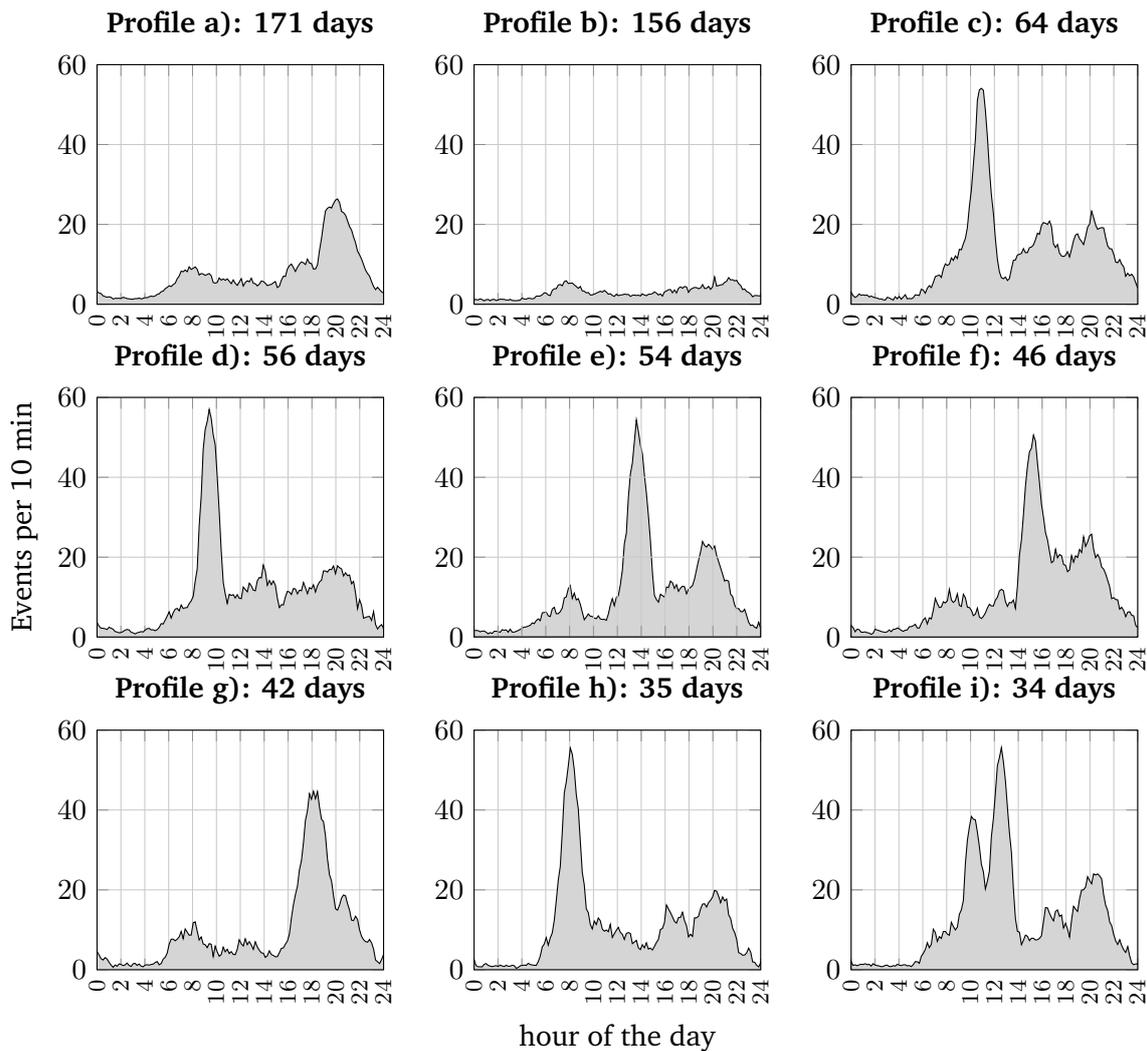


Fig. C.1: clustered daily profiles of WDE events for the UK-DALE data set, house 1

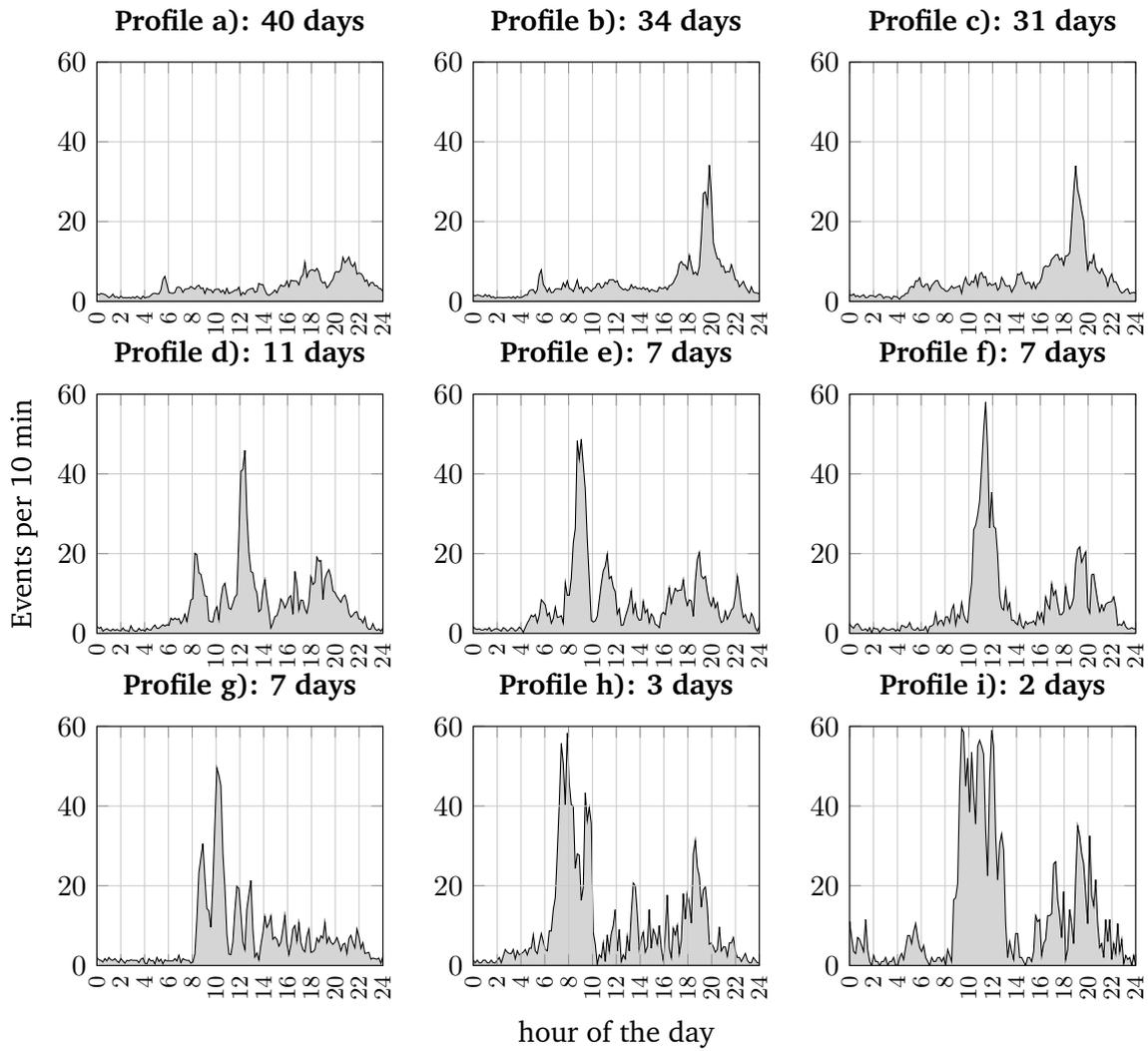


Fig. C.2: clustered daily profiles of WDE events for the UK-DALE data set, house 2

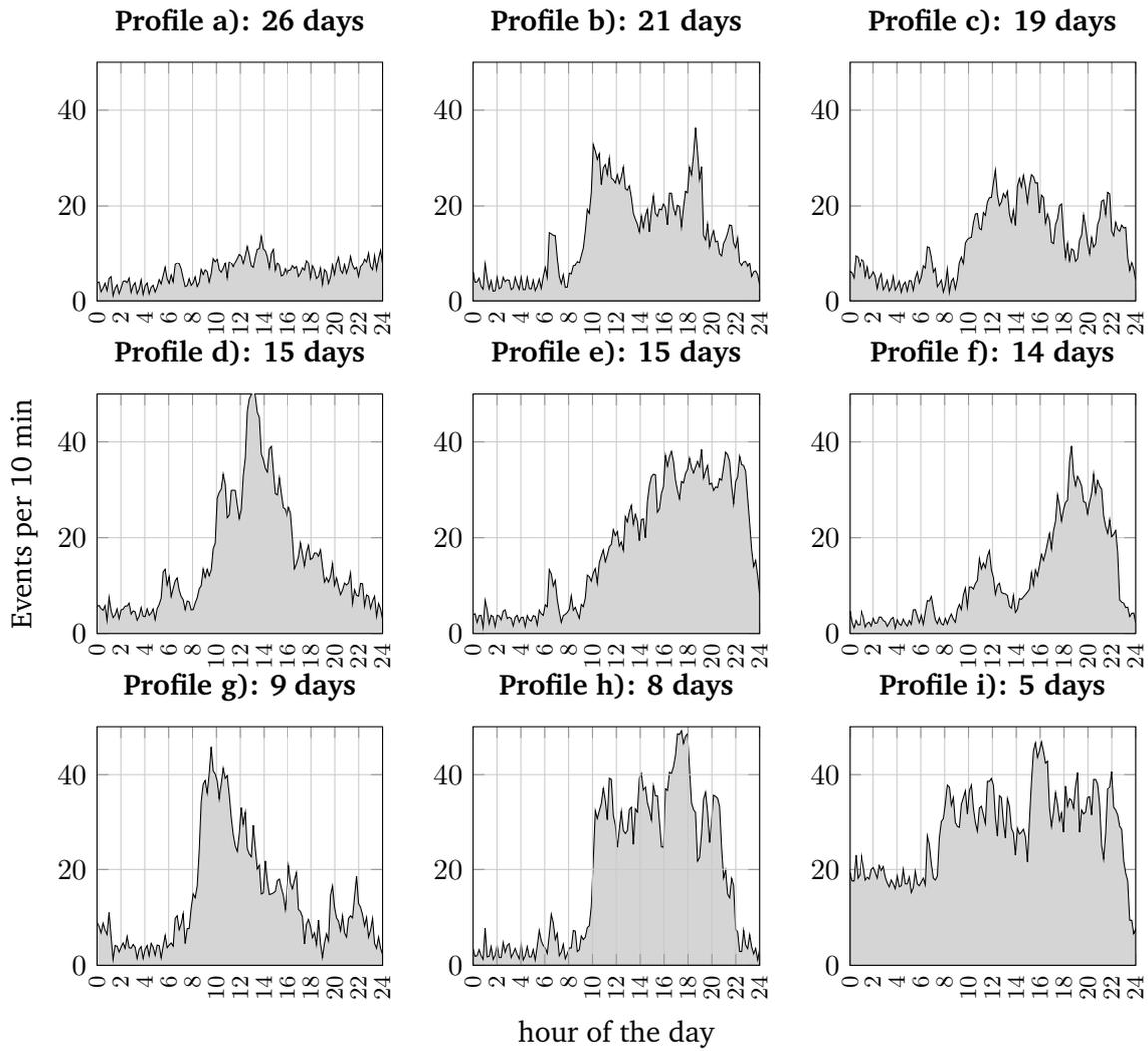


Fig. C.3: clustered daily profiles of WDE events for the UK-DALE data set, house 5



# Lists

## List of Acronyms

<b>AAL</b>	ambient assisted living	<b>LTI</b>	linear time-invariant
<b>AC</b>	alternating current	<b>MIPS</b>	Modélisation, Intelligence, Processus et Systemes
<b>ACCR</b>	appliance cycle classification rate	<b>MIT</b>	Massachusetts Institute of Technology
<b>ADC</b>	analog-to-digital converter	<b>NILM</b>	non-intrusive load monitoring
<b>ADL</b>	Activity of Daily Living	<b>PC</b>	personal computer
<b>AECR</b>	appliance event classification rate	<b>PCC</b>	point of common coupling
<b>APD</b>	activity pattern discovery	<b>PIR</b>	passive infrared
<b>AUT</b>	algorithm under test	<b>ReSP</b>	Signal Processing Research Group
<b>BLE</b>	Bluetooth Low Energy	<b>RFID</b>	radio frequency identification
<b>DC</b>	direct current	<b>RMS</b>	root mean square
<b>DFP</b>	device-free passive	<b>RSSI</b>	received signal strength indicator
<b>EM</b>	Expectation Maximization	<b>SNR</b>	signal-to-noise ratio
<b>FFT</b>	Fast Fourier Transform	<b>SPDC</b>	Standard Power Demand Curve
<b>FM</b>	frequency modulation	<b>SSA</b>	Steady State Approximation
<b>FN</b>	false negative	<b>SSPF</b>	steady state power feature
<b>FNR</b>	false negative rate	<b>SVM</b>	support vector machine
<b>FP</b>	false positive	<b>TP</b>	true positive
<b>FPR</b>	false positive rate	<b>TPR</b>	true positive rate
<b>FSM</b>	finite state machine	<b>TSFP</b>	time-series fingerprint
<b>GMM</b>	Gaussian Mixture Model	<b>UMTS</b>	universal mobile telecommunications system
<b>GSM</b>	global system for mobile communication	<b>VSD</b>	variable speed drive
<b>HAR</b>	Human Activity Recognition	<b>WDE</b>	weighted difference edge detection
<b>HVAC</b>	heating, ventilation and air conditioning	<b>WLAN</b>	wireless local area network
<b>ICT</b>	information and communication technology	<b>WPAN</b>	wireless personal area network
<b>IR</b>	infrared	<b>ZAFH</b>	Zentrum für Angewandte Forschung an Hochschulen
<b>ISM</b>	industry, science, and medicine		
<b>kNN</b>	k-Nearest Neighbor		

## List of Figures

1.1	examples of conclusions that can be drawn from indirect activity data . . . . .	8
1.2	visual representation of an elderly care concept that relies on networks including a circle of “informal caregivers” . . . . .	10
2.1	traditional linear structure for the disaggregation of electric appliances . . . . .	16
3.1	the Fresnel zone as a model of the electromagnetic field between sender and receiver . . . . .	23
3.2	BLE advertising timing pattern. Frequencies from [131, p. 2502]. . . . .	27
3.3	histograms of disturbed RSSI measurements for different channels and mounting configurations . . . . .	28
3.4	setup of the detection system for a single monitored area. The dashed line marks the line of sight. . . . .	29
3.5	comparison of measured RSSI values and the Fresnel zones in order to verify the geometry of the detection area . . . . .	30
3.6	experiment setups for the performance evaluation. TX = transmitter, RX = receiver. The dotted line marks the direct line of sight. . . . .	31
3.7	results of the parameter optimization for the proposed presence detection algorithm . . . . .	33
3.8	extract from an RSSI measurement in the home scenario . . . . .	33
3.9	example of detection signal (gray) and threshold (black) from a home measurement. Ground truth events are marked by black crosses. . . . .	34
3.10	detection errors of various DFP presence detection algorithms . . . . .	35
3.11	comparison of the execution time of various presence detection methods . . . . .	36
4.1	power consumption of an appliance in accordance with the standard model with four sections . . . . .	39
4.2	classification methods grouped according to the way they separate the data. Graph adapted from [138, p. 89]. . . . .	41
4.3	sensitivity of the Nearest Neighbor classifier to outliers. The numbers above the arrows indicate the Euclidean distances. . . . .	42
4.4	DC model for the analysis of external load influences . . . . .	46
4.5	three step model of the computation of an RMS signal . . . . .	50
4.6	frequency response of $\Pi\left(\frac{t}{T}\right)$ with $T = 20$ ms . . . . .	51
4.7	example for the generation of harmonics by taking the square root of $x(t) = 1 + \cos(2\pi \cdot 10 \text{ Hz} \cdot t)$ , $y(t) = \sqrt{x(t)}$ . . . . .	52
4.8	real data example of the RMS operation in time and frequency domain . . . . .	54
4.9	definitions for the evaluation of event detection algorithms . . . . .	56
4.10	definitions for evaluation of the appliance classification stage . . . . .	57
4.11	time difference between consecutive events in data set I . . . . .	59
5.1	event model, and examples for several special cases . . . . .	64
5.2	fixed threshold WDE event detection . . . . .	65
5.3	system model of the event detection . . . . .	66
5.4	proposed event detection scheme for $k_x = k_2 = 3$ . . . . .	66
5.5	filter frequency responses for different parametrizations ( $w_1 = w_3 = 1$ ) . . . . .	67
5.6	filter frequency responses for different parametrizations ( $k_x = k_2 = w_1 = 1$ ) . . . . .	67
5.7	sensitivity of the proposed WDE detector to noise (expectation of 100 runs with 200 events, each), $k_2 = 3, w_1 = 1$ . . . . .	68
5.8	sensitivity to the value of $k_x$ , SNR = 5 dB . . . . .	69
5.9	full parameter search for data set I, file 10, with $k_2 = 3$ . . . . .	70
5.10	effect of the correction of settling processes on the SSA event detection (indicated by horizontal bars) . . . . .	73
5.11	examples of the appliance power consumption pattern in the three subsets of data set II . . . . .	74
5.12	event detection results for data set II . . . . .	75
5.13	true positive rate for the detection of high power appliances in the data sets IIIs and BLUEd . . . . .	76
5.14	event detection performance comparison on large data sets . . . . .	77
6.1	test bench setup for the evaluation of feature extraction algorithms . . . . .	79
6.2	symmetry of the steady state power feature for real power $P$ and reactive power $Q$ difference features, when appliances are modeled by steady state constant power consumption . . . . .	80
6.3	illustration of the parameters for the extraction of SSPFs . . . . .	82
6.4	test results with the starting point parameters. The black line marks the average of files 2 to 10 . . . . .	83
6.5	maximum of the AEER along each dimension for different parameter values (data set I; training: file 1; test: file 10) . . . . .	84
6.6	influence of the weighting function and the number of neighbors on the kNN classification results . . . . .	84
6.7	example for inaccurate features due to inexact event positions . . . . .	85

6.8	histogram of the Euclidean distances for the features in data set I, file 1. Note that the usual order of plots was changed to make the differences visible. . . . .	86
6.9	SSPF extracted from file 1 of data set I . . . . .	88
6.10	SSPF extracted from the BLUED data set . . . . .	89
6.11	test results for the 10 measurements in data set I, and overall averages . . . . .	90
6.12	influence of the number of training samples on the classification results . . . . .	92
6.13	classification procedure for the test feature $\vec{x}_i$ using TSFPs . . . . .	93
6.14	TSFP fingerprint examples. The data sets they are included in are given in the each subplot's title. . . . .	94
6.15	the choice of $k_2$ is limited by fluctuations of the cycle durations . . . . .	95
6.16	influence of the energy normalization on two rectangular pulses . . . . .	97
6.17	linear synchronization functions . . . . .	98
6.18	misclassification due to deviating signal shapes (data set IV) . . . . .	100
6.19	misclassification due to unexpected signal shape (data set V) . . . . .	100
6.20	misclassification due to fingerprints overlapping the turn-off transient (data set V) . . . . .	101
6.21	misclassification due to different time scales (data set V, cross-correlation) . . . . .	101
6.22	TSFP classification using downsampled fingerprints (cf. Fig. 6.13) . . . . .	102
6.23	fingerprint of a freezer at different levels of signal rate . . . . .	103
6.24	influence of the signal rate on the classification performance . . . . .	104
6.25	mean (black line) and uncertainty band (gray area) for a microwave oven averaged over 10 events (manufacturer: Sharp) . . . . .	106
6.26	mean (black line) and uncertainty band (gray area) for a refrigerator averaged over 30 events (manufacturer: Bomann) . . . . .	106
6.27	mean (black line) and uncertainty band (gray area) for a microwave oven averaged over 10 events (manufacturer: micromaxx) . . . . .	107
6.28	mean (black) and uncertainty band (gray area) for a space heater averaged over 215 events (manufacturer: FiF) . . . . .	107
6.29	mean (black) and uncertainty band (gray area) for the heater component of a washing machine averaged over 7 events (manufacturer: Privileg) . . . . .	108
6.30	mean (black) and uncertainty band (gray area) for the heater component of a washing machine averaged over 6 events (manufacturer: Miele) . . . . .	108
7.1	day profiles for different intervals ( $N_p = 17$ ) . . . . .	112
7.2	average activity profile for ReSP data set VI . . . . .	113
7.3	effect of $\kappa$ on the minimum number of profiles per clusters . . . . .	114
7.4	clustered daily profiles for the own data set VI . . . . .	115
7.5	clustered daily profiles for the ReSP data set VI (phases 1 and 2 only) . . . . .	116
7.6	average activity profile for the ReSP data set VI (phases 1 and 2 only) . . . . .	117
7.7	cluster index sequence for the ReSP data set VI (phases 1 and 2 only) . . . . .	117
7.8	daily profiles of the stove in ReSP data set VI . . . . .	118
7.9	clustered daily profiles for the UK-DALE data set, house 1 . . . . .	119
7.10	cluster index sequence for the UK-DALE data set, house 1 . . . . .	119
7.11	clustered daily profiles for the UK-DALE data set, house 2 . . . . .	120
7.12	cluster index sequence for the UK-DALE data set, house 2 . . . . .	120
7.13	clustered daily profiles for the UK-DALE data set, house 5 . . . . .	121
7.14	cluster index sequence for the UK-DALE data set, house 5 . . . . .	121
7.15	clustered daily profiles for the ReSP data set VI using the WDE event detector (phases 1 and 2 only) . . . . .	122
7.16	excerpt of a real current signal of a stove heating a pot of water . . . . .	123
7.17	idealized model of the operation of an electric stove with model parameters . . . . .	124
7.18	operation cycle parameters of a stove while evaporating all water in a pot (model: Bosch HSS 322020) . . . . .	125
7.19	sliding averages of the operation cycle parameters of a stove while evaporating all water in a pot (model: Gorenje E 54V1-E4) . . . . .	126
C.1	clustered daily profiles of WDE events for the UK-DALE data set, house 1 . . . . .	139
C.2	clustered daily profiles of WDE events for the UK-DALE data set, house 2 . . . . .	140
C.3	clustered daily profiles of WDE events for the UK-DALE data set, house 5 . . . . .	141

## List of Tables

2.1	comparison of NILM feature extraction approaches . . . . .	18
3.1	frequencies of the ISM bands (from [124]). . . . .	22
3.2	electric and magnetic parameters of the human skin for $f = 2.4$ GHz (from [25]) . . . . .	24
3.3	packet transmission intervals used in related DFP systems . . . . .	31
4.1	appliances and switching events in data set II . . . . .	60
4.2	appliance switching events in data set III . . . . .	60
4.3	appliance switching events in data sets IV and V . . . . .	61
5.1	selected appliances from the BLUED data set for high power events . . . . .	75
6.1	starting point parameters . . . . .	82
6.2	optimal parameters for the feature extraction algorithm . . . . .	83
6.3	classification results . . . . .	98
6.4	sequence of distinct sections in the fingerprints of microwaves and cooling devices . . . . .	105
7.1	data available for activity profile analyses . . . . .	112
A.1	notation overview . . . . .	134
A.2	set & interval notations . . . . .	134
B.1	conclusion of the parameters for the discrete Wavelet transform . . . . .	137

## Bibliography

- [1] Elisabeth Bubolz-Lutz, Eva Gösken, Cornelia Kricheldorf, and Renate Schramek. *Geragogik - Bildung und Lernen im Prozess des Alterns. Das Lehrbuch*. Kohlhammer, September 2010.
- [2] Matthias Bartsch, Fidelius Schmid, and Antje Windmann. Der Pflegefall. *Der Spiegel*, 47, November 2012.
- [3] Charta der Vielfalt e.V. Jung - Alt - Bunt. Diversity & der demographische Wandel.
- [4] Das Demographie Netzwerk e.V. (ddn). Demographie Fakten.
- [5] Reiner Wichert and Birgid Eberhardt, editors. *Ambient Assisted Living*. 5. AAL-Kongress 2012. Springer, Berlin, 2012.
- [6] Jürgen Rüttgers. Jürgen Rüttgers: Sieben Schritte zu einer lebenswerteren Gesellschaft. *Spiegel Online*, December 2012.
- [7] Sang-Wan Lee, Yong Soo Kim, and Zeungnam Bien. A Nonsupervised Learning Framework of Human Behavior Patterns Based on Sequential Actions. *IEEE Transactions on Knowledge and Data Engineering*, 22(4):479–492, 2010.
- [8] C. Fernandez-Llatas, J.M. Garcia-Gomez, J. Vicente, J.C. Naranjo, M. Robles, J.-M. Benedi, and V. Traver. Behaviour patterns detection for persuasive design in Nursing Homes to help dementia patients. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC*, pages 6413–6417, 2011.
- [9] Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Hausunfälle 2000, August 2002.
- [10] Simon Gleissner. *Radio Signal Indoor Person Tracking*. Bachelor's thesis, Hochschule Furtwangen University, Furtwangen, Germany, December 2013.
- [11] Alexander Bartler. *Untersuchung der nichtlinearen gleitenden RMS-Berechnung sowie Parameterextraktion im Zuge von Non-Intrusive Load Monitoring*. Bachelor's thesis, Hochschule Furtwangen University, Furtwangen, Germany, December 2014.
- [12] Philipp Klein, Dirk Benyoucef, Jean Merckle, and Djaffar Ould Abdeslam. Analysis of fingerprints of electric appliances as starting point for an appliance characteristics catalog. In *IECON 2014 - 40th Annual Conference of the IEEE Industrial Electronics Society*, pages 3517–3521, October 2014.
- [13] P. Klein, J. Merckle, D. Benyoucef, and T. Bier. Test bench and quality measures for non-intrusive load monitoring algorithms. In *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, pages 5006–5011, November 2013.
- [14] T. Bier, D. Benyoucef, D. Ould Abdeslam, J. Merckle, and P. Klein. Smart meter systems measurements for the verification of the detection & classification algorithms. In *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, pages 5000–5005, November 2013.
- [15] M. Floeck and L. Litz. Inactivity patterns and alarm generation in senior citizens' houses. In *Control Conference (ECC), 2009 European*, pages 3725–3730, August 2009.
- [16] J.D. Amor and C.J. James. Behavioral pattern detection from Personalized Ambient Monitoring. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5193–5196, 2010.
- [17] Jennifer Listgarten. *Analysis of sibling time series data: alignment and difference detection*. PhD thesis, University of Toronto, 2006.
- [18] Tracy S. Barger, Donald E. Brown, and Majd Alwan. Health-Status Monitoring Through Analysis of Behavioral Patterns. *IEEE Transactions On Systems, Man, And Cybernetics - Part A: Systems and Humans*, Vol. 35(No. 1):22–27, January 2005.
- [19] Gilles Virone, Majd Alwan, Siddharth Dalal, Steven W. Kell, Beverely Turner, John A. Stankovic, and Robin Felder. Behavioral Patterns of Older Adults in Assisted Living. *IEEE Transactions on Information Technology in Biomedicine*, Vol. 12(No. 3):387–398, May 2008.
- [20] O. Brdiczka, M. Langet, J. Maisonnasse, and J.L. Crowley. Detecting Human Behavior Models From Multimodal Observation in a Smart Home. *IEEE Transactions on Automation Science and Engineering*, 6(4):588–597, 2009.
- [21] M. Youssef, Matthew Mah, and Ashok Agrawala. Challenges: Device-free passive localization for wireless environments. In *Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking, MobiCom '07*, pages 222–229, New York, NY, USA, 2007. ACM.
- [22] M. Ghaddar, L. Talbi, and T.A. Denidni. Human body modelling for prediction of effect of people on indoor propagation channel. *Electronics Letters*, 40(25):1592–1594, December 2004.
- [23] Y.I. Nechayev, P.S. Hall, and Z.H. Hu. Characterisation of narrowband communication channels on the human body at 2.45 GHz. *IET Microwaves, Antennas Propagation*, 4(6):722–732, June 2010.
- [24] P. Bernardi, M. Cavagnaro, S. Pisa, and E. Pizzuoli. Specific absorption rate and temperature increases in the head of a cellular-phone user. *IEEE Transactions on Microwave Theory and Techniques*, 48(7):1118–1126, July 2000.
- [25] P. Salonen, Y. Rahmat-Samii, and M. Kivikoski. Wearable antennas in the vicinity of human body. In *IEEE Antennas and Propagation Society International Symposium, 2004*, volume 1, pages 467–470 Vol.1, June 2004.
- [26] J.A. Soerensen, Z. Safar, Jianjun Chen, K.J. Kristoffersen, and M. Schioetz. Indoor surveillance with multimodal wireless networks. In *Proceedings of the Fourth IEEE International Symposium on Signal Processing and Information Technology, 2004*, pages 242–245, December 2004.

- [27] Jianjun Chen, Zoltan Safar, John Aa Soerensen, and Kare J. Kristoffersen. An rf-based surveillance system using commercial off-the-shelf wireless lan components. In *Proceedings of the 13th European Signal Processing Conference*, 2005.
- [28] Z. Safar, J.A. Sorensen, Jianjun Chen, and K.J. Kristoffersen. Multimodal wireless networks: distributed surveillance with multiple nodes. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05)*, volume 4, pages iv/853–iv/856 Vol. 4, March 2005.
- [29] A.E. Kosba, A. Abdelkader, and M. Youssef. Analysis of a device-free passive tracking system in typical wireless environments. In *2009 3rd International Conference on New Technologies, Mobility and Security (NTMS)*, pages 1–5, December 2009.
- [30] Sajid Hussain, Richard Peters, and Daniel L. Silver. Using received signal strength variation for surveillance in residential areas. volume 6973, page 69730L, March 2008.
- [31] P.W.Q. Lee, W.K.G. Seah, H. Tan, and Zexi Yao. Wireless sensing without sensors – an experimental approach. In *2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 62–66, September 2009.
- [32] Ossi Kaltiokallio, Maurizio Bocca, and Lasse M. Eriksson. Distributed RSSI processing for intrusion detection in indoor environments. In *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks, IPSN '10*, pages 404–405, New York, NY, USA, 2010. ACM.
- [33] Pius W. Q. Lee, Winston K. G. Seah, Hwee-Pink Tan, and Zexi Yao. Wireless sensing without sensors – an experimental study of motion/intrusion detection using RF irregularity. *Measurement Science and Technology*, 21(12):124007, December 2010.
- [34] F. Viani, M. Donelli, M. Salucci, P. Rocca, and A. Massa. Opportunistic exploitation of wireless infrastructures for homeland security. In *2011 IEEE International Symposium on Antennas and Propagation (APSURSI)*, pages 3062–3065, July 2011.
- [35] M. Moussa and M. Youssef. Smart devices for smart environments: Device-free passive detection in real environments. In *IEEE International Conference on Pervasive Computing and Communications, 2009. PerCom 2009*, pages 1–6, March 2009.
- [36] K. El-Kafrawy, M. Youssef, and A. El-Keyi. Impact of the human motion on the variance of the received signal strength of wireless links. In *2011 IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 1208–1212, September 2011.
- [37] Wei-Chuan Lin, W.K.G. Seah, and Wei Li. Exploiting radio irregularity in the internet of things for automated people counting. In *2011 IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 1015–1019, September 2011.
- [38] A.E. Kosba, A. Saeed, and M. Youssef. Robust WLAN device-free passive motion detection. In *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 3284–3289, April 2012.
- [39] A.E. Kosba, A. Saeed, and M. Youssef. RASID: A robust WLAN device-free passive motion detection system. In *2012 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 180–189, March 2012.
- [40] Jiang Xiao, Kaishun Wu, Youwen Yi, Lu Wang, and L.M. Ni. FIMD: Fine-grained device-free motion detection. In *2012 IEEE 18th International Conference on Parallel and Distributed Systems (ICPADS)*, pages 229–235, December 2012.
- [41] Zimu Zhou, Zheng Yang, Chenshu Wu, Longfei Shangguan, and Yunhao Liu. Towards omnidirectional passive human detection. In *2013 Proceedings IEEE INFOCOM*, pages 3057–3065, April 2013.
- [42] Zimu Zhou, Zheng Yang, Chenshu Wu, Longfei Shangguan, and Yunhao Liu. Omnidirectional coverage for device-free passive human detection. *IEEE Transactions on Parallel and Distributed Systems*, 25(7):1819–1829, July 2014.
- [43] B. Mrazovac, M.Z. Bjelica, B.M. Todorovic, A. Miljkovic, and D. Samardzija. Using radio irregularity for increasing residential energy awareness. In *Telecommunications Forum (TELFOR), 2011 19th*, pages 385–388, November 2011.
- [44] B. Mrazovac, M.Z. Bjelica, D. Kukulj, B.M. Todorovic, and D. Samardzija. A human detection method for residential smart energy systems based on zigbee RSSI changes. *IEEE Transactions on Consumer Electronics*, 58(3):819–824, August 2012.
- [45] Bojan Mrazovac and Milan Bjelica. Human presence detection using radio irregularity in wireless networks: Human detection in energy aware residential networks. 2012.
- [46] Bojan Mrazovac, Dragan Dr Kukulj, Dragan Dr Samardzija, Milan Bjelica, and Branislav Dr Todorovic. A method for human presence detection and zonal human subject localization in wireless sensor networks based on distributed processing of shannon entropy in the signal strength space, June 2014. CIB: H04B17/00; H04W24/08; H04W4/00; H04W64/00.
- [47] Moustafa Seifeldin and Moustafa Youssef. A deterministic large-scale device-free passive localization system for wireless environments. In *Proceedings of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments, PETRA '10*, pages 51:1–51:8, New York, NY, USA, 2010. ACM.
- [48] Gabriel Deak, Kevin Curran, and Joan Condell. History aware device-free passive (DfP) localisation. *Image Processing and Communications*, 16(3-4):21–30, 2011.
- [49] A. Eleryan, M. Elsabagh, and M. Youssef. Synthetic generation of radio maps for device-free

- passive localization. In *2011 IEEE Global Telecommunications Conference (GLOBECOM 2011)*, pages 1–5, December 2011.
- [50] M. Seifeldin, A. Saeed, A.E. Kosba, A. El-Keyi, and M. Youssef. Nuzzer: A large-scale device-free passive localization system for wireless environments. *IEEE Transactions on Mobile Computing*, 12(7):1321–1334, July 2013.
- [51] Jiang Xiao, Kaishun Wu, Youwen Yi, Lu Wang, and L.M. Ni. Pilot: Passive device-free indoor localization using channel state information. In *2013 IEEE 33rd International Conference on Distributed Computing Systems (ICDCS)*, pages 236–245, July 2013.
- [52] A. Saeed, A.E. Kosba, and M. Youssef. Ichnaea: A low-overhead robust WLAN device-free passive localization system. *IEEE Journal of Selected Topics in Signal Processing*, 8(1):5–15, February 2014.
- [53] J. Wilson and N. Patwari. Radio tomographic imaging with wireless networks. *IEEE Transactions on Mobile Computing*, 9(5):621–632, May 2010.
- [54] Federico Viani, Paolo Rocca, Manuel Benedetti, Giacomo Oliveri, and Andrea Massa. Electromagnetic passive localization and tracking of moving targets in a WSN-infrastructure environment. Departmental technical report, University of Trento, January 2011.
- [55] J. Wilson and N. Patwari. See-through walls: Motion tracking using variance-based radio tomography networks. *IEEE Transactions on Mobile Computing*, 10(5):612–621, May 2011.
- [56] J. Wilson and N. Patwari. A fade-level skew-laplace signal strength model for device-free localization with wireless networks. *IEEE Transactions on Mobile Computing*, 11(6):947–958, June 2012.
- [57] S. Savazzi, M. Nicoli, F. Carminati, and M. Riva. A bayesian approach to device-free localization: Modeling and experimental assessment. *IEEE Journal of Selected Topics in Signal Processing*, 8(1):16–29, February 2014.
- [58] Jie Yang and Yingying Chen. Indoor localization using improved RSS-based lateration methods. In *IEEE Global Telecommunications Conference, 2009. GLOBECOM 2009*, pages 1–6, November 2009.
- [59] Robert S. Moore, Richard Howard, Pavel Kuksa, and Richard P. Martin. A geometric approach to device-free motion localization using signal strength. Technical Report DCS-TR-674, Rutgers University, September 2010.
- [60] Xiuyuan Zheng, Jie Yang, Yingying Chen, and Yu Gan. Adaptive device-free passive localization coping with dynamic target speed. In *2013 Proceedings IEEE INFOCOM*, pages 485–489, April 2013.
- [61] Dian Zhang, Yanyan Yang, Dachao Cheng, Siyuan Liu, and L.M. Ni. COCKTAIL: An RF-based hybrid approach for indoor localization. In *2010 IEEE International Conference on Communications (ICC)*, pages 1–5, May 2010.
- [62] Chenren Xu, Bernhard Firner, Yanyong Zhang, Richard Howard, Jun Li, and Xiaodong Lin. Improving RF-based device-free passive localization in cluttered indoor environments through probabilistic classification methods. In *Proceedings of the 11th International Conference on Information Processing in Sensor Networks, IPSN '12*, pages 209–220, New York, NY, USA, 2012. ACM.
- [63] Chenren Xu, Bernhard Firner, Robert S. Moore, Yanyong Zhang, Wade Trappe, Richard Howard, Feixiong Zhang, and Ning An. SCPL: Indoor device-free multi-subject counting and localization using radio signal strength. In *Proceedings of the 12th International Conference on Information Processing in Sensor Networks, IPSN '13*, pages 79–90, New York, NY, USA, 2013. ACM.
- [64] Andrei Popleteev. Device-free indoor localization using ambient radio signals. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, UbiComp '13 Adjunct*, pages 549–552, New York, NY, USA, 2013. ACM.
- [65] Moustafa Anwar Seifeldin, Amr Fawzy El-keyi, and Moustafa Amin Youssef. Kalman filter-based tracking of a device-free passive entity in wireless environments. In *Proceedings of the 6th ACM International Workshop on Wireless Network Testbeds, Experimental Evaluation and Characterization, WiNTECH '11*, pages 43–50, New York, NY, USA, 2011. ACM.
- [66] F. Viani, L. Lizzi, P. Rocca, M. Benedetti, M. Donelli, and A. Massa. Object tracking through RSSI measurements in wireless sensor networks. *Electronics Letters*, 44(10):653–654, May 2008.
- [67] Dian Zhang, Yunhuai Liu, and L.M. Ni. RASS: A real-time, accurate and scalable system for tracking transceiver-free objects. In *2011 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 197–204, March 2011.
- [68] Dian Zhang, Jian Ma, Quanbin Chen, and L.M. Ni. An RF-based system for tracking transceiver-free objects. In *Fifth Annual IEEE International Conference on Pervasive Computing and Communications, 2007. PerCom '07*, pages 135–144, March 2007.
- [69] Dian Zhang and L.M. Ni. Dynamic clustering for tracking multiple transceiver-free objects. In *IEEE International Conference on Pervasive Computing and Communications, 2009. PerCom 2009*, pages 1–8, March 2009.
- [70] D. Puccinelli, A. Forster, A. Puiatti, and S. Giordano. Radio-based trail usage monitoring with low-end motes. In *2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pages 196–201, March 2011.
- [71] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. Whole-Home Gesture Recognition Using Wireless Signals, June 2013.
- [72] K. Woyach, D. Puccinelli, and M. Haenggi. Sensorless sensing in wireless networks: Implementation and measurements. In *2006 4th International Symposium on Modeling and Optimization*

- in *Mobile, Ad Hoc and Wireless Networks*, pages 1–8, April 2006.
- [73] S. Sigg, U. Blanke, and G. Troster. The telepathic phone: Frictionless activity recognition from WiFi-RSSI. In *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 148–155, March 2014.
- [74] N. Patwari, J. Wilson, S. Ananthanarayanan, S.K. Kasera, and D.R. Westenskow. Monitoring breathing via signal strength in wireless networks. *IEEE Transactions on Mobile Computing*, 13(8):1774–1786, August 2014.
- [75] Yu Gu, Lianghu Quan, and Fuji Ren. WiFi-assisted human activity recognition. In *2014 IEEE Asia Pacific Conference on Wireless and Mobile*, pages 60–65, August 2014.
- [76] Markus Scholz, Till Riedel, Mario Hock, and Michael Beigl. Device-free and device-bound activity recognition using radio signal strength. In *Proceedings of the 4th Augmented Human International Conference, AH '13*, pages 100–107, New York, NY, USA, 2013. ACM.
- [77] Jana Clement, Joern Ploennigs, and Klaus Kabitzsch. Smart Meter: Detect and Individualize ADLs. In Reiner Wichert and Birgid Eberhardt, editors, *Ambient Assisted Living, Advanced Technologies and Societal Change*, pages 107–122. Springer Berlin Heidelberg, January 2012.
- [78] Jana Clement, Joern Ploennigs, and Klaus Kabitzsch. Detecting Activities of Daily Living with Smart Meters. In Reiner Wichert and Helmut Klaus, editors, *Ambient Assisted Living, Advanced Technologies and Societal Change*, pages 143–160. Springer Berlin Heidelberg, January 2014.
- [79] Shwetak N. Patel, Thomas Robertson, Julie A. Kientz, Matthew S. Reynolds, and Gregory D. Abowd. At the Flick of a Switch: Detecting and Classifying Unique Electrical Events on the Residential Power Line (Nominated for the Best Paper Award). In John Krumm, Gregory D. Abowd, Aruna Seneviratne, and Thomas Strang, editors, *UbiComp 2007: Ubiquitous Computing*, number 4717 in Lecture Notes in Computer Science, pages 271–288. Springer Berlin Heidelberg, January 2007.
- [80] G.W. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, 1992.
- [81] Jian Liang, Simon K K Ng, G. Kendall, and John W M Cheng. Load signature study - part i: Basic concept, structure, and methodology. *IEEE Transactions on Power Delivery*, 25(2):551–560, April 2010.
- [82] Jian Liang, Simon K K Ng, G. Kendall, and John W M Cheng. Load signature study - part II: Disaggregation framework, simulation, and applications. *IEEE Transactions on Power Delivery*, 25(2):561–569, April 2010.
- [83] A. Rosenfeld. A nonlinear edge detection technique. *Proceedings of the IEEE*, 58(5):814–816, May 1970.
- [84] G.W. Hart. Residential energy monitoring and computerized surveillance via utility power flows. *IEEE Technology and Society Magazine*, 8(2):12–16, June 1989.
- [85] K.D. Anderson, M.E. Berges, A Ocneanu, D. Benitez, and J.M.F. Moura. Event detection for Non Intrusive load monitoring. In *IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society*, pages 3312–3317, 2012.
- [86] Roman Streubel and Bin Yang. Identification of electrical appliances via analysis of power consumption. In *Universities Power Engineering Conference (UPEC), 2012 47th International*, page 1–6. IEEE, 2012.
- [87] Karim Said Barsim, Roman Streubel, and Bin Yang. Unsupervised adaptive event detection for building-level energy disaggregation. 2014.
- [88] Karim Said Barsim, Roman Streubel, and Bin Yang. An approach for unsupervised non-intrusive load monitoring of residential appliances, 2014.
- [89] Hannu Pihala. *Non-intrusive Appliance Load Monitoring System Based on a Modern KWh-meter*. Technical Research Centre of Finland, 1998.
- [90] 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*, 41(13):1389–1403, September 2000.
- [91] Michael Baranski. *Energie-Monitoring im privaten Haushalt*. Cuvillier, E, Göttingen, auflage: 1., aufl. edition, 2006.
- [92] A Cole and A Albicki. Nonintrusive identification of electrical loads in a three-phase environment based on harmonic content. In *Proceedings of the 17th IEEE Instrumentation and Measurement Technology Conference, 2000. IMTC 2000*, volume 1, pages 24–29 vol.1, 2000.
- [93] C. Laughman, Kwangduk Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong. Power signature analysis. *IEEE Power and Energy Magazine*, 1(2):56–63, 2003.
- [94] Leslie K. Norford and Steven B. Leeb. Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms. *Energy and Buildings*, 24(1):51–64, 1996.
- [95] W. Wichakool, A-T. Avestruz, R.W. Cox, and S.B. Leeb. Resolving Power Consumption of Variable Power Electronic Loads Using Nonintrusive Monitoring. In *IEEE Power Electronics Specialists Conference, 2007. PESC 2007*, pages 2765–2771, June 2007.
- [96] S.B. Leeb, S.R. Shaw, and Jr. Kirtley, J.L. Transient event detection in spectral envelope estimates for nonintrusive load monitoring. *IEEE Transactions on Power Delivery*, 10(3):1200–1210, July 1995.
- [97] Hsueh-Hsien Chang, Ching-Lung Lin, and Hong-Tzer 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*, pages 1122–1127, April 2008.
- [98] Hsueh-Hsien Chang, Ching-Lung Lin, and Jin-Kwei Lee. Load identification in nonintrusive load monitoring using steady-state and turn-on transient energy algorithms. In *2010 14th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 27–32, April 2010.
- [99] Zhenyu Wang and Guilin Zheng. New method for non-intrusive data extraction and classification of residential appliances. In *Control and Decision Conference (CCDC), 2011 Chinese*, pages 2196–2201, 2011.
- [100] Bin Yang, Edward Eichstetter, and Zhou Zhou. Load measurement and analysis for inverse load reconstruction. In *Universities' Power Engineering Conference (UPEC), Proceedings of 2011 46th International*, pages 1–6, September 2011.
- [101] Bin Yang. An experimental study for inverse load reconstruction. *Journal of Energy and Power Engineering*, 6(11):1826–1834, 2012.
- [102] Sidhant Gupta, Matthew S. Reynolds, and Shwetak 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, Ubicomp '10*, pages 139–148, New York, NY, USA, 2010. ACM.
- [103] Roy Dong, Lillian Ratliff, Henrik Ohlsson, and S. Shankar Sastry. A dynamical systems approach to energy disaggregation. In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 6335–6340. IEEE, 2013.
- [104] Roy Dong, Lillian J. Ratliff, Henrik Ohlsson, and S. Shankar Sastry. Energy disaggregation via adaptive filtering. *arXiv preprint arXiv:1307.4132*, 2013.
- [105] Roy Dong, Lillian Ratliff, Henrik Ohlsson, and S. Shankar Sastry. Fundamental limits of nonintrusive load monitoring. In *Proceedings of the 3rd international conference on High confidence networked systems*, pages 11–18. ACM, 2014.
- [106] M. Ito, R. Uda, S. Ichimura, K. Tago, T. Hoshi, and Y. Matsushita. A method of appliance detection based on features of power waveform. In *2004 International Symposium on Applications and the Internet, 2004. Proceedings*, pages 291–294, 2004.
- [107] C.E. Reeg and T.J. Overbye. Algorithm development for non-intrusive load monitoring for verification and diagnostics. In *North American Power Symposium (NAPS), 2010*, pages 1–5, September 2010.
- [108] Lei Jiang, Suhuai Luo, and Jiaming Li. An Approach of Household Power Appliance Monitoring Based on Machine Learning. In *2012 Fifth International Conference on Intelligent Computation Technology and Automation (ICICTA)*, pages 577–580, January 2012.
- [109] K. Basu, V. Debusschere, and S. Bacha. Residential appliance identification and future usage prediction from smart meter. In *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, pages 4994–4999, November 2013.
- [110] Mario Berges, Ethan Goldman, H. Scott Matthews, and Lucio Soibelman. Learning Systems for Electric Consumption of Buildings. In *Computing in Civil Engineering*, pages 1–10. American Society of Civil Engineers, 2009.
- [111] 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.
- [112] J. Zico Kolter and Matthew J. Johnson. REDD: A Public Data Set for Energy Disaggregation Research. San Diego, CA, USA, 2011. Massachusetts Institute of Technology.
- [113] Andreas Reinhardt, Paul Baumann, Daniel Burgstahler, Matthias Hollick, Hristo Chonov, Marc Werner, and Ralf Steinmetz. On the Accuracy of Appliance Identification Based on Distributed Load Metering Data. In *Proceedings of the 2nd IFIP Conference on Sustainable Internet and ICT for Sustainability (SustainIT)*, pages 1–9, 2012.
- [114] Sean Barker, Aditya Mishra, David Irwin, Emmanuel Cecchet, Prashant Shenoy, and Jeannie Albrecht. Smart\*: An Open Data Set and Tools for Enabling Research in Sustainable Homes. In *Proceedings of the 2012 Workshop on Data Mining Applications in Sustainability (SustKDD 2012)*, Beijing, China, August 2012.
- [115] Jack Kelly and William Knottenbelt. 'UK-DALE': A dataset recording UK Domestic Appliance-Level Electricity demand and whole-house demand. *arXiv:1404.0284 [cs]*, April 2014. arXiv: 1404.0284.
- [116] Nipun Batra, Jack Kelly, Oliver Parson, Haimonti Dutta, William Knottenbelt, Alex Rogers, Amarjeet Singh, and Mani Srivastava. NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring. *arXiv:1404.3878 [stat]*, pages 265–276, 2014. arXiv: 1404.3878.
- [117] Jack Kelly and William Knottenbelt. Metadata for Energy Disaggregation. *arXiv:1403.5946 [cs]*, March 2014. arXiv: 1403.5946.
- [118] Hugo Goncalves, Adrian Ocneanu, and Mario Berges. Unsupervised disaggregation of appliances using aggregated consumption data. In *1st KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*, August 2011.
- [119] Oliver Parson. *Unsupervised Training Methods for Non-intrusive Appliance Load Monitoring from Smart Meter Data*. phd, University of Southampton, April 2014.
- [120] Oliver Parson, Siddhartha Ghosh, Mark Weal, and Alex Rogers. Non-intrusive load monitoring using prior models of general appliance types. pages 356–362, July 2012.
- [121] W. K. Lee, G. Fung, H. Lam, F. Chan, and M. Lucette. Exploration on load signatures. In *International Conference on Electrical Engineering (ICEE)*, 2004.

- [122] H. Y. Lam, G. S K Fung, and W. K. Lee. A Novel Method to Construct Taxonomy Electrical Appliances Based on Load Signatures. *IEEE Transactions on Consumer Electronics*, 53(2):653–660, 2007.
- [123] IEEE Standard Definitions of Terms for Radio Wave Propagation. *IEEE Std 211-1997*, 1998.
- [124] M. Willert-Porada. *Advances in Microwave and Radio Frequency Processing: Report from the 8th International Conference on Microwave and High-Frequency Heating held in Bayreuth, Germany, September 3-7, 2001*. Springer Science & Business Media, May 2007.
- [125] Werner Bächtold. *Mikrowellentechnik: Kompakte Grundlagen für das Studium*. Vieweg+Teubner Verlag, March 1999.
- [126] Roger L. Freeman. *Radio System Design for Telecommunications*. John Wiley and Sons, New York, 1st edition edition, 1987.
- [127] Karl Küpfmüller, Wolfgang Mathis, and Albrecht Reibiger. *Theoretische Elektrotechnik*. Springer-Lehrbuch. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [128] Giovanni Zanca, Francesco Zorzi, Andrea Zanella, and Michele Zorzi. Experimental Comparison of RSSI-based Localization Algorithms for Indoor Wireless Sensor Networks. In *Proceedings of the Workshop on Real-world Wireless Sensor Networks, REALWSN '08*, pages 1–5, New York, NY, USA, 2008. ACM.
- [129] Ambili Thottam Parameswaran, Mohammad Iftekhar Husain, and Shambhu Upadhyaya. Is RSSI a Reliable Parameter in Sensor Localization Algorithms – An Experimental Study. Niagara Falls, USA, September 2009.
- [130] Bluetooth Special Interest Group. Bluetooth Core Specification 4.0, June 2010.
- [131] Bluetooth Special Interest Group. Bluetooth Core Specification 4.1, December 2013.
- [132] Xi Chen, A. Edelstein, Yunpeng Li, M. Coates, M. Rabbat, and Aidong Men. Sequential monte carlo for simultaneous passive device-free tracking and sensor localization using received signal strength measurements. In *2011 10th International Conference on Information Processing in Sensor Networks (IPSN)*, pages 342–353, April 2011.
- [133] Jie Yang, Yong Ge, Hui Xiong, Yingying Chen, and Hongbo Liu. Performing joint learning for passive intrusion detection in pervasive wireless environments. In *2010 Proceedings IEEE INFOCOM*, pages 1–9, March 2010.
- [134] B. Mrazovac, B.M. Todorovic, M.Z. Bjelica, and D. Kukulj. Device-free indoor human presence detection method based on the information entropy of RSSI variations. *Electronics Letters*, 49(22):1386–1388, October 2013.
- [135] European Parliament, Council of the European Union. DIRECTIVE 2006/32/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL OF 5 April 2006 on energy end-use efficiency and energy services and repealing Council Directive 93/76/EEC, May 2006.
- [136] Bundesministerium der Justiz und für Verbraucherschutz. Gesetz über die Elektrizitäts- und Gasversorgung, July 2014.
- [137] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 1st edition edition, 2006.
- [138] Thomas A. Runkler. *Data Analytics - Models and Algorithms for Intelligent Data Analysis*. Springer Vieweg, Wiesbaden, 2012.
- [139] Thomas Brox. *Statistical Pattern Recognition*, 2013.
- [140] Thomas Gobmaier. *Netzfrequenzmessung*, December 2013.
- [141] Thomas Frey. *Signal- und Systemtheorie*. Springer DE, October 2008.
- [142] Marlene Marinescu and Jürgen Winter. *Basiswissen Gleich- und Wechselstromtechnik*. Vieweg, 2 edition, 2007.
- [143] A. G. Phadke and J. S. Thorp. *Synchronized Phasor Measurements and Their Applications*. Springer, August 2008.
- [144] J. Duncan Glover, Mulukutla S. Sarma, and Thomas J. Overbye. *Power System Analysis & Design*. Cengage Learning Emea, 5 edition, April 2011.
- [145] Thomas Bier. *Disaggregation of Electrical Appliances using Non-Intrusive Load Monitoring*. PhD thesis, Universite de Haute Alsace, Mulhouse, France, December 2014.
- [146] Brian S. Everitt, Sabine Landau, Morven Leese, and Daniel Stahl. *Cluster Analysis*. Wiley, Chichester, West Sussex, U.K, 5 edition edition, February 2011.
- [147] Stephan Vogts. *Smart Metering: Algorithmenentwicklung zur Merkmalsextraktion*. Master's thesis, Hochschule Furtwangen University, August 2011.
- [148] New Netz GmbH. *Standardlastprofile nach BDEW*.
- [149] D. T. Pham, S. S. Dimov, and C. D. Nguyen. Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1):103–119, January 2005.
- [150] Hering, Martin, and Stohrer. *Physik für Ingenieure*. Springer, Berlin, auflage: 7. aufl. edition, September 1999.
- [151] Erich Rummich. *Energiespeicher: Grundlagen, Komponenten, Systeme und Anwendungen*. expert, Renningen, auflage: 1 edition, January 2008.
- [152] G. Kutyniok. *Zeit-/Frequenzanalyse*, 2008.
- [153] Alfred Mertins. *Signaltheorie*. Teubner, Stuttgart, 1996.
- [154] Stephane Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, CA, USA, second edition edition, 1999.