



TU Clausthal

ADVANCES TO ENERGY INFORMATICS

ON THE COLLECTION, PROCESSING, AND PRIVACY
PROTECTION OF ELECTRICITY CONSUMPTION DATA

KUMULATIVE
HABILITATIONSSCHRIFT

ANDREAS REINHARDT

Andreas Reinhardt

*Advances to Energy Informatics: On the Collection, Processing, and Privacy
Protection of Electricity Consumption Data*

Von der Fakultät für Mathematik/Informatik und Maschinenbau der
Technischen Universität Clausthal genehmigte Habilitationsschrift

Clausthal-Zellerfeld, 2019

CONTENTS

1	INTRODUCTION	1
1.1	Energy informatics	3
1.2	State of the art	5
1.3	Contributions and outline	10
2	COLLECTION OF ELECTRICITY CONSUMPTION DATA	13
2.1	Load signatures	13
2.2	Collecting load signatures using embedded systems	15
2.3	Load signature compression	17
2.4	Remote re-configuration of data collection devices	19
2.5	In-network data processing	22
3	PROCESSING ELECTRICITY CONSUMPTION DATA	25
3.1	Appliance recognition	25
3.2	Consumption prediction	29
3.3	Anomaly detection	32
3.4	Load modeling for automated evaluations	34
4	PROTECTING USER PRIVACY IN SMART GRIDS	37
4.1	Intentional data perturbation	37
4.2	Load hiding through renewable power generation	40
5	CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS	43
6	REFERENCES	45
7	LIST OF ACRONYMS	61
A	PUBLICATIONS CONTRIBUTING TO THIS HABILITATION THESIS	63

INTRODUCTION

The almost ubiquitous availability of electrical power is one of the key accomplishments of the 20th century. This *electrification* has not only led to a significantly increased comfort for users of electrical energy, but also enabled formerly inconceivable technological progress. The global population has since become widely accustomed to the omnipresence of electricity; almost 80% of the planet's population have access to electrical power today [IEA10]. In order to provide a reliable and resilient energy supply, electrical energy generation, transmission, and distribution systems are being steadily improved. As a result of these efforts, most power supply grids worldwide already attain availability levels in excess of 99% [CEE16]. The German power grid, being particularly well known for its stability, ranks among the most reliable power grids, with an average unavailability time of less than 12 minutes for the year 2015 [VDE16].

Besides the evolution of electrical power grids, strong research activities towards the *digitalization* have been conducted throughout the past decades. Thanks to continuous innovations in computer systems and communication technologies, innumerable digital devices have found their way into private homes, offices, and industrial plants worldwide. While initially operated in a standalone fashion, their interconnection in a world-spanning network has enabled formerly unimaginable novel systems and services, including the World Wide Web [BLCG92], Cloud computing [MG11], and the Internet of Things [Ash09]. In analogy to electrical power grids, a world without ubiquitous access to networked resources has become truly unthinkable.

The symbiotic combination of *electrification* and *digitalization* has led to the emergence of *smart grids* [AKM12]. Smart grids augment conventional electrical power grids by their instrumentation with networked data collection points. This decentralized deployment of data collection devices allows for a detailed monitoring of the grid status in real-time. Diverse application scenarios exist for the use of data collected this way, including the early detection of transitions into unstable regions of operation and the enablement of detailed consumption analytics for better capacity planning. A second characteristic property of smart grids is the presence of decentralized data processing and actuator devices, which allow stakeholders to exert fine-grained control over the grid's operation, e.g., to quickly respond to changing operating conditions by applying corresponding countermeasures. The principal features that set smart grids apart from conventional power grids are the availability of a large volume of information provided by the deployed sensing points, and the range of services enabled by appropriately combining available data from such sources.

Electricity metering infrastructure has seen a constantly progressing trend towards its digitalization. Electromechanical induction meters are gradually being replaced by digital metering devices (*smart meters*) with communication interfaces. More and more electrical in-home appliances are equipped with network adapters to enable their interconnection and create *smart homes*. Even commercial and industrial sites are gradually being retrofitted with the sensing infrastructure necessary to gain a deeper understanding of energy consumption and allow for its optimization, as part of the transition towards *Industrie 4.0* [KWH13]. The notion of *smart cities* is a continuation of these concepts at greater scale, and combines information from electrical distribution and transmission grids with data from other modalities in urban regions.

This range of deployment scenarios for measurement devices already hints at the heterogeneity of the landscape of data collection points in smart grids. The high information requirement in some locations may necessitate the use of Phasor Measurement Units (PMUs), i.e., devices that capture a range of electrical parameters at high temporal and amplitude resolutions. In contrast, other sensing points may have less stringent requirements to the rate and resolution of collected data, so they can be realized using smart meters or even uncalibrated low-cost plug-level power sensors. Regardless of the technological realization of measurement devices, however, the principal components in an Advanced Metering Infrastructure (AMI) consistently integrate three functionalities synergistically: Sensors to measure power-related physical quantities, a data processing component to extract relevant features from collected data, and a bidirectional communication interface to relay such information and receive configuration and actuation commands.

The benefits of a fine-grained data collection in smart grids become apparent when considering the services they enable for the different stakeholders. Transmission System Operators (TSOs) and Distribution System Operators (DSOs) can gain deeper insights into the evolution of consumer characteristics over time and use this information for the appropriate dimensioning of the infrastructure they operate. Through the detection and localization of fluctuations and disturbances in the power grid, decentralized generators, such as photovoltaic (PV) systems or wind turbines, as well as controllable storage systems can be operated accordingly to minimize the risk of propagating faults into the grid and to avoid peak loads. Electricity consumers can also benefit from the correlation of grid-wide information with the data they have provided. On the one hand, detailed insights into generation and demand allow utility companies to offer flexible tariff structures to incentivize “grid-friendly” user behavior. On the other hand, data collected within industrial sites or private homes can be used to create awareness for a building’s energy consumption and to achieve savings, e.g., through energy-aware building automation. In conclusion, the holistic collection of generation and demand data allows for an unprecedented view of the power grid and the enablement of many novel services.

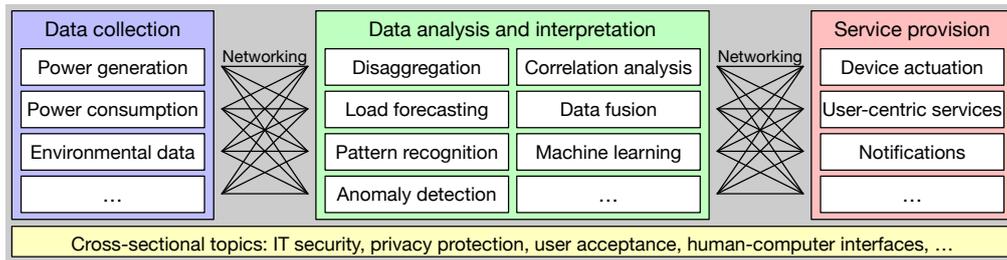


Figure 1.1: Delineation of the scope of energy informatics.

1.1 ENERGY INFORMATICS

The technological challenges and opportunities arising from the intersection of electrical power grids with networked systems that collect, relay, and process information from/within them, represent the foundation of *energy informatics*, the principal subject of this habilitation. Energy informatics is an interdisciplinary research field that encompasses all aspects of smart grids that can be addressed through the application of Information and Communication Technology (ICT). The focus of energy informatics thus lies on the integration of computer networking, data processing, and systems engineering within the power grid. A visualization of the core components of energy informatics is shown in Figure 1.1 and to be understood as follows. Data sources (left) provide readings of grid properties, analysis methods (center) cater to their interpretation and fusion, and services are provided to stakeholders (right) based on the insights gained. Computer networks are employed to interface the components with each other.

Its dependency on requirements and boundary conditions imposed by several involved disciplines sets energy informatics apart from many other application domains of ICT. Given that transmission and distribution grids build up the core of the smart grid, knowledge of power electronics and electrical power engineering is vital to understand design decisions, valid operating condition ranges, and characteristics of the components deployed in power grids. Moreover, smart grids are also tightly coupled to measurement engineering, given that AMI devices need to be carefully designed and correctly deployed to collect meaningful data. Lastly, adequate solutions for ensuring data security and privacy protection are mandatory to cater for user acceptance and fulfill legal requirements for data collection, transmission, and storage in smart grids. For the realization of energy informatics functionalities, these aspects need to be adequately considered.

Energy informatics will gradually transform the traditional power grid, in which electrical power and energy consumption used to be the principal quantities of interest, into a smart power grid. Readings which used to be solely metered at the customer handover point and merely sufficed to enable services like billing and capacity planning, will be enriched by a broad range of other sensed modalities. Even today, the increasing deployment of further sens-

ing devices in the power grid has already led to an unprecedented degree of additional information. In addition to monitoring electrical power flows, currently sensed modalities also include power quality [Bol99], Electromagnetic Interference (EMI) emissions in substations [Smo12], or environmental conditions [HBA+12]. Thus, a comprehensive set of data describing the status of a power grid in a detailed fashion can already be retrieved on demand, and this holistic situative picture is constantly refined due to the perpetually growing number of deployed data collection systems.

Meaningful ways to analyze and correlate collected data, with the objective of extracting higher-level information and providing services based thereupon, are at the core of energy informatics. Consequently, this research area has seen an almost exponential increase of scientific attention in the last decade. One of the fundamental ideas behind energy informatics has been highlighted in seminal work by George Hart in [Har85]: Using observed changes to a household's aggregate load (collected at the electricity meter) to identify appliance usage and operation modes. Since then, the task of *disaggregating* the household total load into information about the individual contributing appliances, often referred to as Non-Intrusive Load Monitoring (NILM), has evolved into a research challenge addressed by numerous researchers worldwide, including Hart himself [Har89; Har92]. Many more works that extend and/or refine this concept have been presented subsequently, differing in the sensor deployment locations, their temporal and amplitude resolutions, or the extent of supplementary data collected. Moreover, numerous prototypes have been implemented to underpin the practical relevance of energy informatics, such as through energy audits [BGM+01], the identification of user types [CDW+11; AXR+15], the detection of household characteristics [WHM+12; BSS+14; BSS+15] and occupancy [KBS15], the detection of anomalous consumption [APR15; AKG+15], and the prediction of future demand [BMN09; RCK14]. While the majority of these approaches is centered around the inhabitants of a dwelling, the underlying data processing operations are often generally applicable across application scenarios, e.g., to determine an appliance's mechanical wear [ORL+10] for predictive maintenance purposes. This brief excerpt of the state of the art already shows that significant research activities are being conducted on converting conventional electrical power grids into smart grids through the application of ICT, with contributions ranging from generic architectures (e.g., [CEN12]) to complex technical solutions for specific challenges.

A contributing factor for the wide scientific and industrial uptake of energy informatics is its practicality due to the large-scale rollout of *smart metering* infrastructure [EU12]. Millions of homes and industrial sites worldwide have already been equipped with such networked electricity meters (cf. [EC14; EF14]). Besides smart electricity meters, many dwellings are also being fitted out with smart home and/or building automation systems, which often comprise energy monitoring devices for electrical circuits or individual wall outlets. An unprecedented and fast-growing amount of fine-grained

energy consumption data is thus continually becoming available for analysis. Moreover, supplementary data to increase the data processing accuracy and performance in energy informatics are gradually being provided thanks to the emergence of smart cities and industrial automation networks. The rise of the Internet of Things (IoT), with its envisioned 50 billion terminal devices by the year 2020 [Eva11], is fueling this process even more. A holistic architecture for collecting, processing, and interpreting data related to the generation and/or consumption in a power grid in meaningful ways is thus strongly required. Once it is in place, a wide range of novel services will provide benefits to all stakeholders, including electricity consumers, operators (including DSOs and TSOs), and the power grid itself.

1.2 STATE OF THE ART

It is the objective of energy informatics to consider the collection, preprocessing, and storage of electricity consumption data in a holistic and all-encompassing way. Contributions to several, partially orthogonal, research topics must hence be considered in union. The current state of the art for each of the relevant fields is summarized as follows.

Technically, the collection of *energy consumption data* can be realized in one of the following four ways [FLG+11]:

1. *Distributed direct sensing*: This approach relies on (retro-)fitting individual electrical appliances with sensing devices. The currently prevalent approach to facilitate distributed direct sensing is the insertion of a sensor into the appliance's power cord. To this end, both commercially available solutions (*smart plugs*, often part of home automation solutions) as well as research prototypes like Plug [LFO+07], ACme [JDHD+09], SEM [WBD+11], or SmartMeter.KOM [RBM+11] exist. While allowing for the data collection at appliance-level granularity, the need for an individual sensor per appliance makes this approach expensive. To cut this cost, sensors with low resolution are often being used, resulting in lower achievable signal information.
2. *Single-point sensing*: Given the presence of electricity meters in virtually all homes, leveraging them to collect measurements of voltage and current (and thus the extraction of phase information) is the most direct and trivial way of gathering information. As smart meters are generally supplied by a customer's utility company and need to be installed by certified technicians, however, their technical specifications are usually pre-determined (cf. [DoE11]) and non-changeable. The absence of a unified interface to collect (high-resolution) sampling data from smart meters additionally complicates the actual retrieval of data.
3. *Intermediate sensing methods*: While the two aforementioned methods reflect the extremal points of potential sensor locations in dwellings, several intermediate options exist as well. Marchiori et al. have proposed

the installation of metering equipment on circuit level in [MHH+10]. In [JS10], Jung and Savvides present a methodology for the progressive deployment of sensors in order to minimize the risk of ambiguous power attribution to appliances. A versatile hardware platform for intermediate sensing is Monjolo [DCD13], which can be clipped to any cable (e.g., on the distribution board or in conduit boxes). Monjolo can determine the presence of flowing currents, yet by design it is neither able to quantify their exact amplitude nor whether they result in a real or reactive power consumption.

4. *Indirect sensing*: In contrast to the direct monitoring of electrical power flows, approaches have also emerged to indirectly infer power flow information based on different modalities. Concepts leveraging sensors for vibration, magnetic fields, or thermal imaging have been published in ViridiScope [KSC+09], ANNOT [SGD+10], TinyEARS [TGM10], Heat-Probe [HKC+11], Supero [PTM+13], Deltaflow [CCD14]. Such surrogate indicators are often suitable to infer the operational state of an appliance, although they are commonly not able to exactly quantify the resultant energy demand.

Besides the sensor deployment locations, their *sampling rate* is equally important to consider. Across the practical deployments referenced above, it can be noted that the majority of existing approaches technically only allow for the collection of samples at a sampling rate 1 Hz, or even less. An analysis of spectral components in appliances' current waveforms is hence implicitly excluded because this sampling frequency does not satisfy the Nyquist-Shannon sampling theorem. In order to overcome this limitation, several research platforms have been presented which allow to capture voltage and current signatures at higher sampling rates (e.g., 15 kHz [BGS+11; CLY08; CLL10], or 8 kHz [LFO+07; LLC+03; RPL+10]). However, such systems are mostly not available commercially, particularly not in larger quantities. Even though general-purpose power analyzers could be used to facilitate this data collection, economic reasons hinder their wide deployment (e.g., for distributed direct sensing). Research has nonetheless unambiguously demonstrated that higher sampling frequencies – up to the megahertz range – are required for capturing transient events or even characteristic noise signatures backscattered by electrical consumers onto the power lines [FLG+11]. Assessments of single point sensing deployments with sampling rates of 36 kHz [GRP10] and even 100 kHz [BGM+09] have been conducted to this end. When recording current flows at such sampling rates, the authors of the related studies found that transition noises do not only occur for device de-/activation transitions, but also for internal appliance state changes [CLS+06; PRK+07]. This observation is also well aligned with earlier work by Marubayashi [Mar97], which indicated that each appliance (de)activation leads to minimal changes of the mains voltage within the household. Another general observation was that switching power supplies, such as the ones typically shipped with laptop com-

puters, produce considerable characteristic noise in the range of hundreds of kilohertz [FLG+11]. A dedicated consideration of high-frequency sampling, at least at selected points within a building, thus appears as a promising approach to infer additional information on electrical energy consumption, as confirmed by preliminary studies in [GRS14; KUK+16].

Data collection campaigns in the real-world incur significant monetary expenses for sensor purchases, installation, and maintenance. This is particularly aggravated by the fact that no universally applicable collection methodology has been proposed to this day, thus the longevity of installations is not ensured. However, the *use of real-world data* is indispensable for the development of practically applicable analytics algorithms. Consequently, researchers around the globe have started to collect power and/or energy consumption data by means of measurement campaigns. In the last decade, a growing number of energy consumption data sets has been collected and made available for download. A selection of such data sets is listed in Table 1.1 on page 8, annotated by their year of collection and instrumentation methodology. Moreover, the number of buildings monitored, sensor sampling rates, and geographic locations of collection are reported.

When reviewing the characteristics of the listed data sets, it becomes apparent that no coherent data collection strategy can be observed. Measurement devices are often deployed opportunistically, and the re-use of existing devices (e.g., smart meters) is paramount in most scenarios. As a result, several limitations of the published data sets can be summarized:

- ▷ Only a small subset of the available data sets feature information on a household's aggregate demand in conjunction with the power consumptions of all present appliances (i.e., the combination of SPS with DDS or PDS). Only two data sets encompass full-building instrumentations, yet their data has not been collected at a sampling rate sufficient for spectral analysis.
- ▷ User actions that correlate to appliance actuation decisions are rarely recorded and published with the data (with Smart* [BMI+12] and ECO [BKC+14] being exceptions).
- ▷ Additional annotations like the ground truth of an appliance's actual mode of operation are absent from the aforementioned data sets. Only the PLAID data set [GGK+14] contains different entries for start-up and steady-state current traces.

It is imperative to *evaluate the information content* of collected data in order to determine the efficacy of a sensor instrumentation methodology. Despite the importance of this research challenge, however, it has received only little scientific attention to date. In contrast to this barely covered research topic, *energy data processing* (also named *energy analytics*) has been intensively studied and numerous algorithms have been proposed to this end. Even though they do not primarily target to evaluate the information content in consumption data, they allow for its indirect inference; input data of higher quality generally

NAME	YEAR	GRANULARITY	# BUILDINGS	SAMPLING RATE	LOCATION
PowerNet [KGH+10]	2010	DDS	1	1 Hz	United States
REDD [KJ11]	2011	SPS + PDS + ISM	6	15 kHz (SPS) $1\text{-}1/3$ Hz (PDS/ISM)	United States
BLUED [AOB+12]	2012	SPS	1	12 kHz	United States
Tracebase [RBB+12]	2012	PDS	20+	1 Hz	Germany
Smart* [BMI+12]	2013	SPS + DDS + ISM	3	1 Hz (SPS/ISM) $\sim 1/5$ Hz (DDS)	United States
AMPds [MPB+13]	2013	SPS + ISM	1	$1/60$ Hz	Canada
ACS-F2 [RGH13]	2013	PDS	<i>n/a</i>	$1/10$ Hz	Switzerland
OCTES [EST13]	2013	SPS	33	$1/7$ Hz	Scotland
iAWE [BGS+13]	2013	SPS + PDS + ISM	1	1 Hz	India
GREEND [MEE+14]	2014	PDS	9	1 Hz	Austria/Italy
ECO [BKC+14]	2014	SPS + PDS	6	1 Hz	Switzerland
PLAID [GGK+14]	2014	PDS	55	30 kHz	United States
UK-DALE [KK14]	2015	SPS + PDS	5	16 kHz (SPS) $1/8$ Hz (PDS)	United Kingdom
DRED [URP15]	2015	SPS + DDS	1	1 Hz	Netherlands
REFIT [MLS+15]	2015	SPS + PDS	20	$1/8$ Hz	United Kingdom
WHITED [KUK+16]	2016	PDS	<i>n/a</i>	44 kHz	worldwide
COOLL [PNMR+16]	2016	PDS	1	100 kHz	France

Table 1.1: Selection of existing energy informatics data sets. Sensor instrumentations: *DDS*: distributed direct sensing of all individual loads; *PDS*: partial distributed direct sensing (selected loads only); *SPS*: single-point sensing at the meter; *ISM*: intermediate sensing methods, e.g., circuit-level monitoring.

also lead to higher accuracy values of the energy analytics algorithms. The primary focus of early energy analytics research has been the disaggregation of a household’s energy demand based on the observed changes of its real – and often also reactive – power consumption [Har92; DK99; FZ99; Pru02; LLC+03; BV04; Cox04]. While initial solutions were predominantly focused on binary-state appliances (i.e., devices with a deterministic and repeatable change of real and reactive power when turned on or off), many subsequent projects have also looked at appliances with more complex consumption patterns. In the majority of solutions presented to date (cf. [ZR11; ZGI+12; AGS+13] for comprehensive surveys of NILM approaches), low-frequency data have been used, i.e., samples collected at a rate below the frequency of the mains voltage (often 1 Hz or less). However, other approaches have also looked into the use of

input data containing spectral components in order to disaggregate household totals (e.g., [LSJ95; CA98; CA00; IUI+04; BGM+09; PGW+11; DDD+17]). Besides household load disaggregation, the detection of operational state and anomalous consumption behavior in industrial settings (e.g., electrical motors) has emerged as a secondary application scenario of energy analytics [LLN+05; SLN+08; ORL+10]. Once initial solutions had demonstrated the possibility and viability of energy analytics, subsequent approaches were designed to cater to the disaggregation of appliances with variable power consumptions during their operation (such as variable-frequency drives) as well as the usage of more sophisticated techniques for their use disaggregation algorithms, e.g., Bayesian networks [PGW+14] or neural networks [Pru02; RNS+10; KK15]. Domain-specific preprocessing steps have also been investigated in [QG14; RK14], targeting the elimination of spurious data or the conversion of raw sensor samples into data structures that allow for faster processing.

Despite this positive prospect for energy analytics, two more aspects are vital for its success. Firstly, a trade-off between the amplitudinal and temporal resolution of collected power data and the desired accuracy of the data analytics needs to be determined to optimize their *demand for storage space*. Correlation analyses were conducted in [EK10; LES+14; TSB+14], hinting at a non-linear relation between the granularity at which data are collected and the resultant service quality. Requirements to data storage systems also show a major dependency on the sampling resolution; in [MCB+15], annotated power flows from seven channels captured at 12.8 kHz sampling rate have been shown to produce a daily volume of almost 20 gigabytes. Even when applying lossless compression, the storage requirement for data collected at 24 bits resolution and 16 kHz sampling rate sums up to 4.8 gigabytes per appliance per day [KK14]. More recently, a storage demand of 286 gigabytes per day has been determined when sampling a three-phase system at 250 kHz [HKK+17]. Besides the data preprocessing solutions presented above, orthogonal approaches to maintain manageability have hence also been presented in literature. For example, proposals have been made to downsample data recorded at 100 kHz to a mere 20 Hz in order to mitigate the storage requirement [BGM+09], or to compress smart meter data [RRR+12; EEK+15]. Storage demands are also strongly influenced by the way in which consumption data are being recorded. While IBM recommends database solutions [IBM12], the NILMTK project [BKP+14] natively supports input data in the HDF5 file format. The data sets listed in Table 1.1 mostly use plain comma-separated value files (e.g., in [KGH+10; RBB+12; GGK+14]) or XML-based representations (such as [RGH13]), neither of which is well-suited to achieve high performance.

Secondly, besides purely technological constraints, the *user acceptance of smart metering* and other forms of energy data collection is crucial for the success and adoption of energy analytics. Adequate privacy protection of collected data is thus mandatory, especially as numerous campaigns against

the installation of smart meters have already formed worldwide¹. Existing approaches mostly rely on cryptography to protect the confidentiality of collected data [JAL12; WL13], but are partially also complemented by means for local data obfuscation [REC15] or the use of local storage or renewable generation to cover up power consumption [ZJW+14; EPE14; RKE+15] and maintain user privacy this way.

1.3 CONTRIBUTIONS AND OUTLINE

To realize truly smart grids, research and development activities in all of the underlying domains need to be conducted, and their results must be synergistically combined. Energy informatics is one of the pivotal areas for the success of smart grids. Predominantly concerned with the application of ICT in the domain of electrical power systems, energy informatics encompasses a range of topic fields, as shown in the overview diagram in Figure 1.1 on page 3. My recent research activities have been focused on this domain, and I have made several scientific contributions to the areas depicted in the figure. These contributions are presented in more technical depth in the following chapters of this manuscript, yet briefly categorized in this section to cater for a common understanding of their relation to each other and their interconnections.

The first set of contributions presented in this habilitation thesis address the field of collecting electricity consumption information and ensuring their lossless storage and forwarding. The concerted collection of sensor data related to electrical quantities as well as supplementary sensor data about the prevailing environmental conditions serves as the foundation for many services, and is thus an important foundation for energy informatics. Unless sensor data of appropriate resolution and granularity are available, the processing and resulting services provision is severely hampered. Thus, a first set of my contributions concerns the creation of infrastructures to collect data from decentralized sources in a scalable fashion and ensure their timely and accurate transmission. My contributions to this domain include the design and implementation of embedded systems (in hard- and software) to collect electricity consumption information, as well as their efficient coding to alleviate bandwidth requirements and cater for scalability. The contributions I have made to the field of energy information collection, targeting to capture a holistic and comprehensive picture of a part of the power grid, are presented in Chapter 2.

Besides the collection of data, their processing takes an integral role in energy informatics and smart grids. Raw electricity consumption data are generally large in volume, repetitive in their nature, and often contain only few events of importance. In order to fully exploit the information content in such data, it is thus necessary to eliminate redundancies in the data and extract

¹Examples include <http://www.stopsmartmeters.org> and <http://stopsmartmeters.org.uk>

patterns of significance. Due to their nature of being periodically sampled, electricity consumption information can be modeled as time-series data. Thus, techniques from the domain of time series analysis be applied to extract knowledge from the input data. I have investigated several alternative ways to extract high-level information from raw electricity consumption data in my research. By considering electricity consumption data in both the time and the frequency domain, the approaches allow to determine characteristic power consumption patterns when appliances are activated, deactivated, or change their mode of operation. Based on such information, it becomes possible to detect anomalies by identifying deviations from previously observed patterns, or to predict an appliance's future power demand when such recurring patterns are detected. My research results on the extraction of information content from raw electricity consumption data are summarized in [Chapter 3](#).

Besides purely technological considerations, user acceptance is vital for the success of smart grids and energy informatics. In order to cater to the users' needs, an adequate protection of the sensitive data collection from their metering points is indispensable. This is also of great relevance to ensure compliance with current legislation (e.g., the European Union's General Data Protection Regulation [[EU16](#)]). In essence, the protection of user privacy is a key contributor to the acceptance of load monitoring in private homes. Ensuring a strict compliant to privacy regulations and the sufficient protection of electricity consumption data thus represents a cross-sectional topic of high practical relevance. Consequently, I have investigated mechanisms to protect user privacy in smart grids, which are presented in [Chapter 4](#) of this work. Given that it is not possible to measure the degree of achieved privacy protection in absolute numbers, in-depth analyses of the parameter spaces have been conducted in order to derive recommendations for configuring the designed protection algorithms.

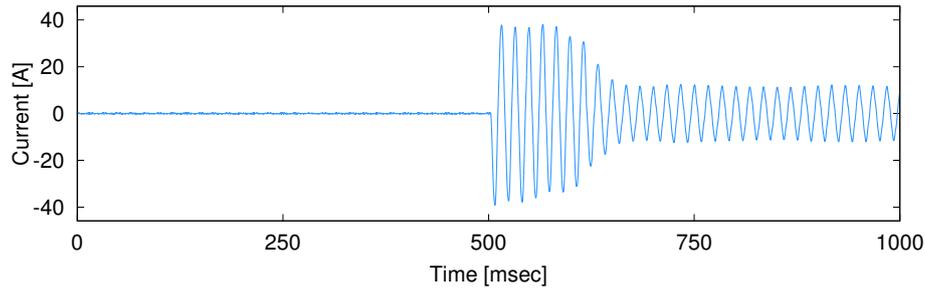
This habilitation thesis concludes with a summary of the presented contributions and possible future research directions in [Chapter 5](#).

COLLECTION OF ELECTRICITY CONSUMPTION DATA

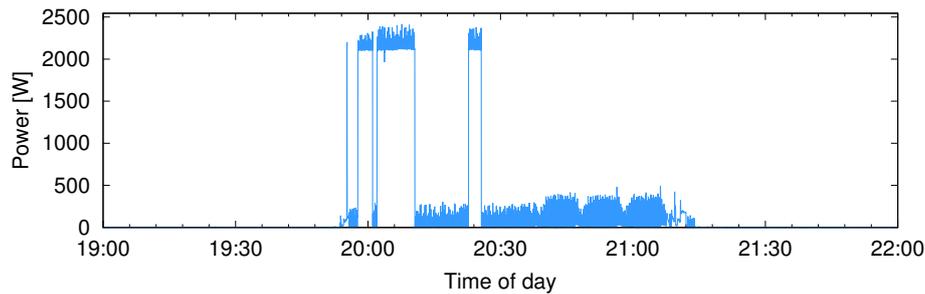
The availability of energy consumption data at the required resolution for adequate processing is one of the principal pillars of energy informatics. While a plethora of devices to collect energy information in a scalable manner have been presented in prior research, they most often feature limited data reporting rates to ensure their operability over bandwidth-limited communication channels. In fact, it is not uncommon to see reporting intervals as long as 15 minutes between samples (as defined in §60 MsbG), even though most approaches presented in scientific publications rely on readings reported once per second. Besides the different data reporting intervals, the nature of the collected readings also often differs between sensor deployments. In a large number of systems, only a single sample of the value of interest (usually the averaged consumption since the last report) is being communicated. Substituting a possibly large number of collected voltage and current waveform samples by a single scalar value leads to a loss of information, though. Thus, it severely hampers the operation of many energy analytics algorithms, or at least affects their accuracy negatively. Nonetheless, reasons for the widespread collection of such downsampled data exist, most importantly the significantly reduced buffer and bandwidth requirements for the data transmissions. The transmission of low-rate low-detail readings makes it possible to use tightly resource-constrained embedded systems for data collection, while still allowing for the observation of general trends. In order to fully exploit the potential of energy informatics, innovative contributions are needed to collect electricity consumption data at sufficiently high resolutions to enable the extraction of their full information content, yet keep the volume of data manageable at the same time. After a brief introduction of the notion of load signatures, four of my contributions that address this challenge are presented in the remainder of this section.

2.1 LOAD SIGNATURES

Load signatures represent the fundamental quantities of interest in energy analytics. They model an electrical appliance's characteristic power consumption during the course of its operation, determined by the current drawn as well as the voltage simultaneously applied to its terminals. The analysis of load signatures has been extensively studied in research, e.g., in [LNK+10; CCT+12; DHH+16; DRD+18], with the objective of extracting high-level information from collected data. My own contributions to the field of load signature analysis are presented in [Chapter 3](#) of this work. Two principal types of load sig-



(a) Microscopic load signature of a washing machine's activation current [Rei17a].



(b) Macroscopic load signature of a washing machine's power consumption [Rei17a].

Figure 2.1: Macroscopic and microscopic load signatures of a washing machine.

natures have been defined by Zeifman and Roth in [ZR11], which are briefly summarized as follows to cater for a common understanding.

Load signatures that contain voltage and current signals captured at frequencies much greater than the frequency of the mains voltage are generally referred to as *microscopic* load signatures. PLAID [GGK+14] is a data set of microscopic appliance inrush signatures collected at 30 kHz sampling rate, i.e., 500 times the mains frequency of its collection location (Pittsburgh, USA). A sample load signature from the PLAID data set is shown in Figure 2.1a; note that only the current consumption waveform is shown for visual clarity. It shows a washing machine's current consumption in the moment the appliance is switched on. The sinusoidal nature of the current is clearly visible, as well as the fact that the power consumption in the initial mains periods after activation is largely different than the comparably steady waveform afterwards, due to the initially low resistance of internal heating elements. While this level of detail appears promising for energy data analytics, a key downside of microscopic load signatures is their storage demand.

Alternatively, load signatures can be composed of averaged readings of a load's power consumption, reported at low frequencies, such as once per second. This lowered resolution makes it impossible to reconstruct the actual shape of the appliance's current intake. However, such *macroscopic* load signatures cater to the limitations imposed by wireless communications and storage systems. Even though the level of detail achievable from microscopic load signatures is no longer available, macroscopic load signatures still allow

for the analysis of long-term trends and patterns on time scales of minutes, hours, or even days. An example for this type of load signature is shown in [Figure 2.1b](#), which visualizes the power consumption of a washing machine from the Tracebase data set [RBB+12], in which appliance-level data has been collected in Germany at 1 Hz sampling rate. The washing machine's heating cycles with power consumptions in excess of 2,000 W can be easily distinguished from rinsing phases with considerably lower power demand.

The conversion of a microscopic load signature into its macroscopic representation is easily possible, yet a lossy process. For example, the multiplication of the current waveform shown in [Figure 2.1a](#) with the corresponding voltage samples (also part of the data set) yields the washing machine's power consumption which can then be integrated over the course of one second. The three largely different phases visible in [Figure 2.1a](#) (off, initial inrush, steady state), however, all take place within one second. Consequently, they can no longer be discerned in the macroscopic load signature once the data have been converted, despite their potential relevance for load signature analysis algorithms.

2.2 COLLECTING LOAD SIGNATURES USING EMBEDDED SYSTEMS

Electricity consumption data can be collected in many distributed locations within the power grid. Many more meaningful insights can, however, be gained by correlating data simultaneously captured in different locations. A network of data collection devices must hence be established to create such an infrastructure for energy data analytics. As many potential metering locations are located in places where no wired data connection is available, wireless solutions are frequently used to this end. A key advantage is their quick and simple installation, and due to their installation at a live mains connection there is virtually no installation effort to supply them with power. From a technological point of view, networked metering devices can be seen as an instantiation of Wireless Sensor Networks (WSNs) [ASS+02], in which each distributed sensor is responsible for the collection, on-board processing, and reporting of electricity consumption data. Their miniaturized form factor and the low monetary cost of WSN devices makes them highly viable for deployments at scale. The strong research activities that have been ongoing in the domain of WSNs for almost two decades moreover provide an excellent foundation for the practical use of wireless sensors for energy monitoring.

The majority of today's WSN platforms do not feature sensors for mains voltage amplitudes or current flows by default. Versatile expansion headers, however, make it possible to interface external peripherals. To foster research on energy analytics I have thus designed TUCap, an embedded microcontroller-based system that allows for the collection and processing of consumption data. Technical details about TUCap's features and its data processing capabilities are presented in the following publication.

A. Reinhardt. "TUCap: A Sensing System to Capture and Process Appliance Power Consumption in Smart Spaces." In: *Proceedings of the 16th GI/ITG KuVS Fachgespräch "Sensornetze" (FGSN)*. 2017, pp. 31–34.

At the heart of TUCap is an extension board that relies on an Analog-to-Digital Converter (ADC) which allows for the synchronous sampling of voltage and current signals at 36 kHz sampling rate and 24 bits of resolution. Thus, it enables the collection of microscopic load signatures, and harmonics up to the 360th order can be analyzed. Through the choice of transducers (a clip-on current transformer and a voltage transformer with CEE 7/16 "Europlug" connector), its non-intrusive installation is facilitated. The TUCap sensor interface is connected to an attached microcontroller board via a Serial Peripheral Interface (SPI) bus connection. A photography of the sensor interface is shown in Figure 2.2.

The buffering, processing, and transmission of collected voltage and current readings is the task of an attached microcontroller platform. Without loss of generality, a Teensy 3.2² device has been used in the prototypical evaluation of TUCap. Based on a 72 MHz ARM Cortex-M4 microcontroller with 256 kB of Flash memory and 64 kB of RAM, the Teensy platform natively supports SPI communications at up to 25 MHz frequency. Thus, it can retrieve and buffer sensor readings in real-time; even real-time data processing is possible to a limited extent. In fact, a supplementary study on the similarity of consecutive voltage and current waveforms, also part of the publication, has demonstrated their high resemblance, making them well-suited for compression. In summary, TUCap is a key enabling component for load signature collection on embedded systems and the basis for the contributions presented in Section 2.3.

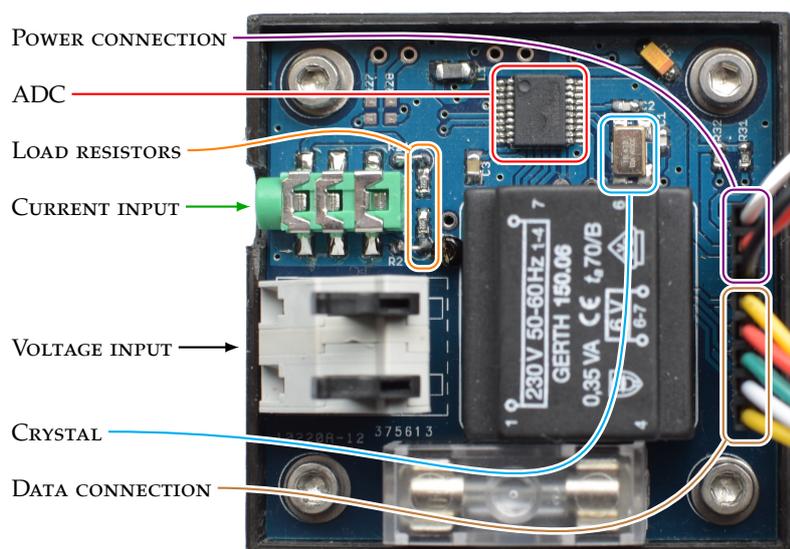


Figure 2.2: Key components on the TUCap sensor interface board [Rei17b].

²A technical description of the device is available at <https://www.pjrc.com/teensy/>

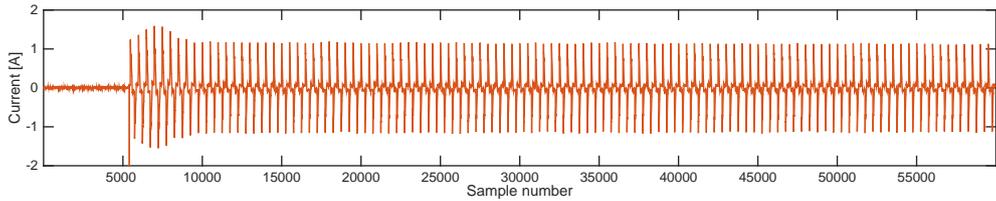
2.3 LOAD SIGNATURE COMPRESSION

The information content in microscopic load signatures is significantly greater than in macroscopic load signatures [LES+14; TSB+14]. Thus, the use of microscopic features is strongly favorable to tap the full potential of energy informatics. Their demand for storage, however, is non-negligible. While readings collected once per second amount to a total of 86,400 samples per day, more than 3 billion data points are collected when collecting samples at a rate of 36 kHz. At an amplitude resolution of 24 bits per sample, this amounts to a storage demand of almost 10 gigabytes per day, as compared to slightly more than 250 kilobytes for the secondly interval. At the same time, the repetitive nature of current and voltage waveforms (cf. Figure 2.1) shows only little variations between successive mains periods. This has motivated the investigation to what extent a compression of subsequent waveforms is possible, which has been documented in the following publication:

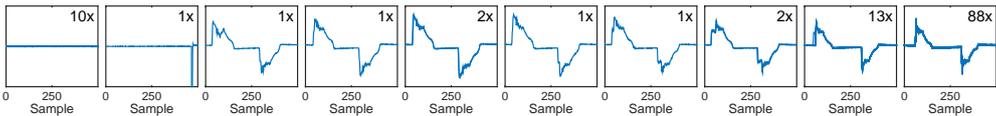
A. Reinhardt. “Adaptive Load Signature Coding for Electrical Appliance Monitoring over Low-Bandwidth Communication Channels.” In: *Proceedings of the 5th IFIP Conference on Sustainable Internet and ICT for Sustainability (SustainIT)*. 2017, pp. 1–8.

The key contribution of aforementioned paper is ALSCEAM, a scheme for the adaptive compression of microscopic load signatures. In the paper, I initially conducted an assessment of the compressibility of load signatures by converting them into different time series data representations: IEEE 754 floating-point numbers, as Comma-Separated Values (CSV), and RIFF WAVE (PCM data). A subsequent assessment of their compressibility with established data compression tools (ZIP, bzip2, FLAC) has demonstrated savings in excess of 99.5%, yet only when sufficiently many waveform periods were available for compression. When only a single waveform or the concatenation of 120 mains periods were used for analysis, the savings dropped to 15–80%. In fact, in some configurations compression losses even occurred, i.e., the compressed output was larger in size than the input data. Moreover, the bzip2 compression algorithms is documented to require at least 400 kilobytes of RAM; it thus vastly exceeds the capacity available on the TUCap platform (cf. Section 2.2).

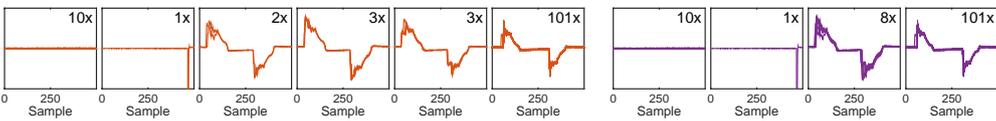
ALSCEAM instead relies on the concept of omitting repetitive waveforms without significant variations from their transmission altogether. To this end, the waveform of one period of the quantity of interest (voltage or current) is buffered initially. A dissimilarity metric based on the Root-Mean-Square Error (RMSE) to the stored data is computed for all subsequently recorded waveforms. Newly collected waveforms are only transmitted when their dissimilarity to the previously recorded waveform exceeds a threshold ρ_{th} . The effect of applying ALSCEAM is visualized in Figure 2.3. Instead of transmitting data for all 120 mains periods, only values for 10, 6, or 4 mains periods are reported when applying ALSCEAM, leading to a reduction of the data by 91.7–96.7%.



(a) Current consumption of a compact fluorescent lamp for a duration of two seconds, taken from the PLAID data set [GGK+14] (file 3.csv). Note that a negative peak overshoot at sampling point 5,399 has been omitted for improved visual clarity.



(b) Application of ALSCEAM to decompose the trace shown above into repetitive current consumption waveform periods using sensitivity value $\rho_{th}=0.25$.



(c) Decomposition of the trace with a less strict similarity requirement ($\rho_{th}=0.45$). (d) Decomposition of the trace with $\rho_{th}=0.6$.

Figure 2.3: Application of ALSCEAM to separate a microscopic load signature into similar repetitive waveforms. The number of occurrences of the waveforms is annotated in the top-right corner of each diagram [Rei17a].

A complementary approach to reduce the traffic resulting from the transmission of uncompressed macroscopic load signature data is presented in:

A. Reinhardt and S. Koessler. "PowerSAX: Fast Motif Matching in Distributed Power Meter Data Using Symbolic Representations." In: *Proceedings of the 9th IEEE International Workshop on Practical Issues in Building Sensor Network Applications (SenseApp)*. 2014, pp. 531–538.

The proposed approach, named PowerSAX, is conceptually similar to ALSCEAM in the sense that it is based on the detection of previously observed consumption patterns. Its operation on macroscopic data, however, imposes different boundary conditions to be considered. While the periodic nature of current and voltage waveforms can be analyzed in-depth when microscopic detail is available, macroscopic load signatures lack such short-term patterns. Instead, a different type of repetitive patterns can be observed for numerous device types: Many appliances, particularly white goods, operate according to a pre-defined sequence of (internal) states. The macroscopic load signature of the washing machine visualized in Figure 2.1b can, e.g., be expected to repeat in the same fashion for every operation of the appliance. Such devices also often have operation durations on the order of minutes or even hours (the washing cycle in Figure 2.1b takes over two hours to complete).

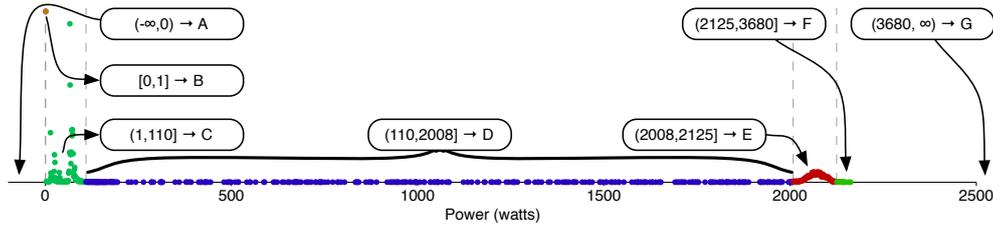


Figure 2.4: Histogram of the power consumption values of a dishwasher appliance, with cluster boundaries and corresponding symbol annotations [RK14].

Like ALSCEAM, PowerSAX has been designed with its applicability on embedded sensing platforms in mind. In particular, it mitigates memory constraints by operating on symbolic approximations [LKL+03] of the macroscopic load signature data instead of processing raw consumption readings. The computation of the mapping between consumption data and corresponding symbolic value is determined through a density-based clustering of the observed readings. To this end, the histogram of an appliance’s power consumption is being computed, and clusters of similar consumption values are determined using the MeanShift algorithm [CM02]. An example for this internal conversion process is shown in Figure 2.4, which shows the clustered histogram of a dishwasher’s power consumption and the mapping of clusters to a symbolic alphabet with characters A–G. This alphabet generation process is executed during an initialization phase and adapts the symbol mapping to the consumption characteristics of the attached appliance(s). As a positive side effect of histogram clustering, PowerSAX is also mostly immune against measurement noise.

During its regular operation, newly collected macroscopic consumption values are converted to their symbolic representations according to the previously determined mapping. Thereafter, the most recent symbols in a sliding time window are compared to a set of previously stored symbolic sequences. Once a known symbolic sequence pattern is found in the stream of symbols, a corresponding message is transmitted to notify stakeholders about the occurrence of the underlying event (e.g., “dishwasher cycle completed”). The potential for efficient binary encodings of the symbol sequences, combined with the early termination of the symbol sequence matching as soon as a mismatching symbol is detected further minimize PowerSAX’s resource demand and accelerate its operation.

2.4 REMOTE RE-CONFIGURATION OF DATA COLLECTION DEVICES

When data sources for energy analytics applications are connected through low-power wireless links, bandwidth utilization considerations are of major concern. The data rates of protocols currently used for networked embedded systems (such as IEEE 802.15.4 [IEEE06]) often range below 1 Mbps, and packet

sizes are often confined to small values (127 bytes in IEEE 802.15.4). Consequently, a single sensing device that transmits microscopic load signatures in real-time can already completely exhaust the available bandwidth, leaving no transmission opportunities for other wireless terminals. Sharing the wireless channel between multiple devices thus requires all participating stations to lower their number of transmitted packets. While local preprocessing mechanisms like ALSCEAM (cf. Section 2.3) provide a viable way to reduce traffic during periods of constant appliance power consumption, bursty traffic can still occur when appliance operation modes change. It is thus necessary to have a means to limit the maximum data generation rate of sensors, e.g., by relaying commands to device groups in order to avoid congestion on the wireless channel. The RoCoCo protocol facilitates the dissemination of control messages into networks of embedded sensing systems while minimizing the resulting energetic overhead, and is presented in the following publication:

A. Reinhardt and C. Renner. "RoCoCo: Receiver-initiated Opportunistic Data Collection and Command Multicasting for WSNs." In: *Proceedings of the 12th European Conference on Wireless Sensor Networks (EWSN)*. 2015, pp. 218–233.

RoCoCo is a symbiosis between data collection and data dissemination. Built on the collection protocol ORiNoCo [URT12], it enables the collection of sensor readings at a destination device, e.g., a border router, at very low energetic overhead. In contrast to collection protocols, however, RoCoCo also offers a way to disseminate data (e.g., control messages) to individual nodes of groups thereof. This is accomplished by synergistically integrating additional fields into the messages exchanged by ORiNoCo. Most importantly, a Bloom filter [Blo70] is used to store the addresses of all destination devices, such that an arbitrary number of recipients can be specified at constant overhead.

The low-power operation of ORiNoCo is only minimally affected by these additions. In fact, testbed evaluations have demonstrated that the network-wide average energy consumption has only increased by 0.11% when RoCoCo is being executed on top of ORiNoCo. Further evaluation results from practical testbed experimentation with 36 embedded systems are shown in Figure 2.5. The first subdiagram shows the average distance between nodes and the edge router. Due to the nature of the underlying collection protocol, routes can change between packet transmissions, such that corresponding topology changes take place. In all cases, nodes were not located further than five hops away from the edge router. In the second subdiagram, the average interval between data transmissions is shown. Given that nodes located in proximity to the edge router need to relay more data on behalf of nodes further into the network, their intervals are mostly smaller than for far-away nodes. The third subdiagram shows the delay experienced when configuration messages are being disseminated into the network. While not as quick as the data collection (shown in the fourth subdiagram), the figure still demonstrates that

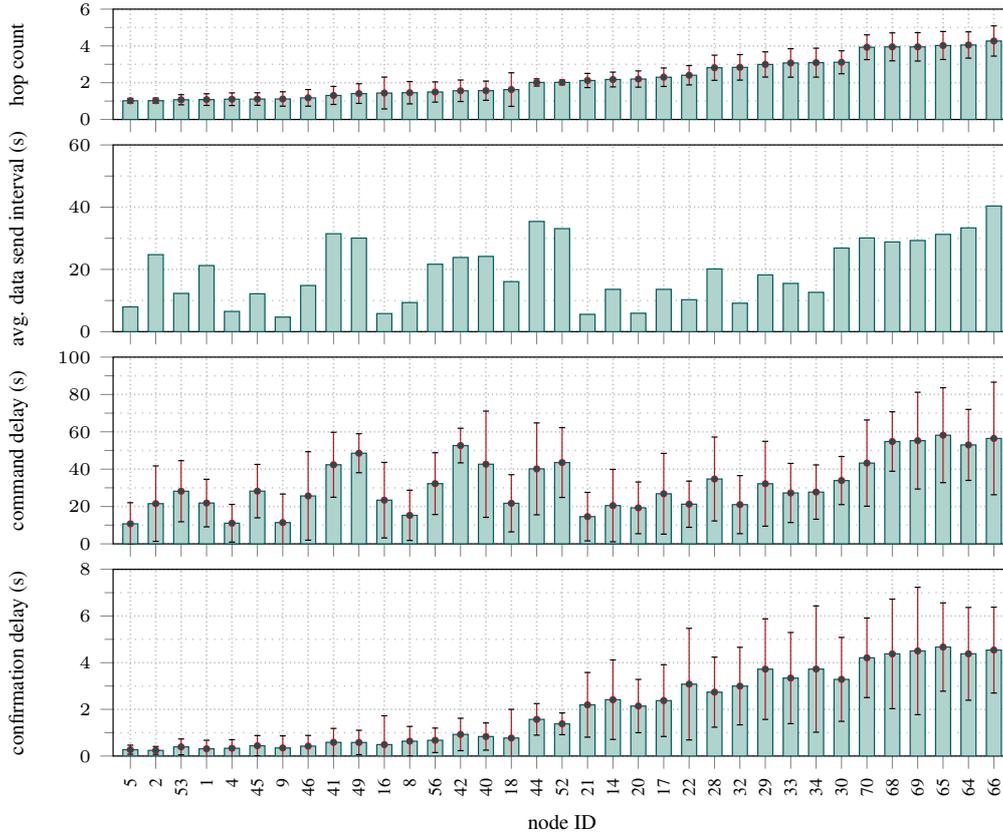


Figure 2.5: Synoptic visualization of four key evaluation parameters when running RoCoCo in a testbed setting with 36 devices. Nodes are ordered by their average hop count to the network edge (first subfigure). The second subfigure shows the average interval at which packets (i.e., collection messages and command confirmations) are being sent. The remaining two subfigures visualize the time required to disseminate the command to each device, as well as the time taken to confirm receipt of the command to the edge router. Error bars indicate the standard deviation [RR15].

dissemination of control messages for up to four hops is mostly accomplished in less than one minute. The fourth and last subfigure finally shows the time taken to confirm the receipt of the configuration message back to the edge router. Given that this traffic follows the same direction as the data that is being collected by the sensor devices, the forwarding process terminates quickly, in less than eight seconds for the given topology.

In conclusion, RoCoCo serves as a substrate for wireless networks of embedded systems which facilitates quick data collection while allowing for the configuration of nodes at extremely low energy overhead. It is sufficiently lightweight to be executed on sensor devices with only 48 kilobytes of program memory and 10 kilobytes of RAM, thus well-suited for its application in energy informatics data collection scenarios.

2.5 IN-NETWORK DATA PROCESSING

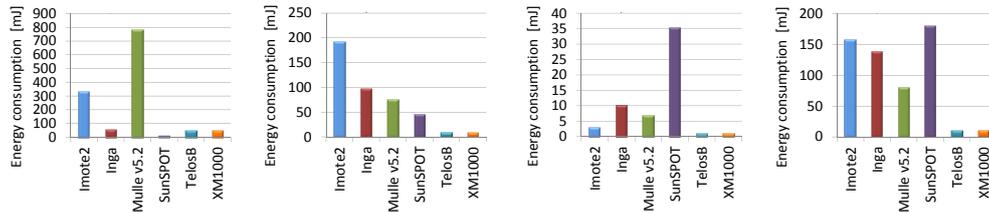
Up to this point, only devices with identical hardware configurations have been considered in the presented contributions. The upcoming vision of the Internet of Things (IoT) will entail a drastic change to this assumption. Smart objects with embedded networking capabilities will permeate our everyday lives and enable many currently unthinkable services and novel interaction paradigms. Heterogeneous in their nature, IoT devices will not only serve different purposes, but be equipped with different sensing, processing, storage, and communication capabilities.

As a result of the exponentially growing number of devices, an increased saturation of wireless communication links is highly likely to be observed. Besides the option to transition to new communication technologies like the fifth generation of mobile communication systems (5G) [PSS16], technologies to integrate existing devices and infrastructures are vital for the success of the IoT. One concept from the domain of WSNs stands out in this regard, given its implicit applicability on heterogeneous systems: *In-network processing* relies on the use of computational power available within the network [CLX+06]. Leveraging processing capabilities inside the network of embedded systems dispenses with the need to fully forward a sensor's collected data to the network edge, possibly relayed by multiple intermediate stations. Instead, devices with sufficient computational energy reserves can apply data processing algorithms within the network to reduce the traffic volume. As a side effect, the negative effects of energy holes [LM07] can be reduced when in-network processing is applied.

Existing works on in-network processing partially consider device heterogeneity, yet primarily with regard to a device's residual energy. Through shifting resource-intensive processing tasks away from nodes that are approaching the depletion of their energy budget, a prolonged network runtime can often be accomplished. However, methods to compensate for other dimensions of device heterogeneity have rarely been considered in related work, despite the expectation that networks composed of devices with entirely different specifications are expected to emerge. This motivated the contributions of the following paper:

A. Reinhardt and D. Burgstahler. "Exploiting Platform Heterogeneity in Wireless Sensor Networks by Shifting Resource-Intensive Tasks to Dedicated Processing Nodes." In: *Proceedings of the 14th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. 2013, pp. 1–9.

The paper presents an investigation into the facets of device heterogeneity, by conducting an in-depth comparison of six widely used hardware platforms for embedded sensing systems with regard to their computational capabilities. Four data processing algorithms that can be employed to process and/or reduce wireless traffic have been implemented on each of the target systems,



(a) Fast Fourier Trans- (b) LZSS compression. (c) Twofish encryption. (d) Recursive Fibonacci form (FFT). series calculation.

Figure 2.6: Energy demand for the execution of four benchmarking functions on each of the six embedded sensing platforms [RB13].

and the corresponding energy consumptions for their execution been used as a benchmark. Results are visualized in Figure 2.6 and demonstrate the hugely varying energy demands to perform the same operation on different devices. In the paper, this observation is followed by a simulative analysis to determine situations in which task migration between deployed devices is meaningful, and to what extent energy can be saved through collaborative data processing. A key insight was that there are many situations where lightweight systems with inefficient processors can benefit from the data processing in their vicinity, leading to network-wide energy savings from 46–77%.

This result also allows for conclusions to be drawn on the successful application of energy informatics. As soon as buildings will experience a stronger penetration with IoT systems, setting up dedicated networks to collect electricity consumption information will no longer be necessary. Instead, sensors for voltage, current, and relevant environmental parameters can be expected to seamlessly integrate into this IoT infrastructure. The study conducted in the presented paper has effectively proven that there will be potential in in-network data processing, particularly when heterogeneous devices are available. The resultant possible traffic reductions will strongly cater to the requirements of energy informatics.

Besides investigating methodologies and instrumentation approaches to collect electricity consumption information in an efficient and scalable way, their processing also takes a key role in energy informatics. In existing literature, both macroscopic and microscopic load signatures have been found to contain a large information content. By extracting such information, a variety of services for all involved stakeholders in the smart grid are conceivable. Firstly, the user's comfort and convenience can be improved by learning typical usage patterns of appliances and actuating them autonomously according to the user's preferences. At the same time, the automated deactivation of appliances to reduce a household's energy consumption while maintaining the user's satisfaction becomes possible. Secondly, services to the advantage of utility companies can be realized, such as the prediction of future electricity demand. Such information can be used in the capacity planning of transmission lines and to estimate the demand for generation. Thirdly and lastly, functionalities to ensure and improve safety and security in both residential and industrial settings can be based on consumption data. By finding outliers in consumption data, anomalies in household activities and manufacturing processes alike can be recognized early and relevant stakeholders be notified. I have made contributions to three current research challenges in energy data processing: Appliance recognition, load forecasting, and anomalous consumption detection. The contents of the corresponding scientific publications are summarized in the following sections. Lastly, [Section 3.4](#) introduces AMBAL, a tool to synthetically generate load signatures, and thus enlarge the body of available data sets (cf. [Table 1.1](#) on [page 8](#)) significantly. While not directly a technique for load signature processing, AMBAL facilitates the generation of large and well-annotated data sets for the evaluation of such algorithms.

3.1 APPLIANCE RECOGNITION

One of the principal services in energy informatics is the capability to recognize electrical appliances based on their microscopic and/or macroscopic consumption characteristics. Expressive features are required to fully distinguish consumption patterns between appliances. Most often, these are based on the characteristic power consumption patterns, the temporal patterns of consumption changes, and possibly the prevailing environmental conditions during appliance use. The most prominent application scenario for appliance recognition is its use in [NILM](#), i.e., the recognition of currently operated appliances from the aggregate household load.

The extraction of features from collected load signatures to is also the principal contribution of the following publication:

A. Reinhardt, P. Baumann, D. Burgstahler, M. Hollick, H. Chonov, M. Werner, and R. Steinmetz. "On the Accuracy of Appliance Identification Based on Distributed Load Metering Data." In: *Proceedings of the 2nd IFIP Conference on Sustainable Internet and ICT for Sustainability (SustainIT)*. 2012, pp. 1–9.

Contrary to the majority of research works that try to tackle NILM from a holistic viewpoint (i.e., through the analysis of aggregate consumption traces), the paper addresses an important sub-challenge, namely the recognition of single appliances from their macroscopic load signatures. Even though this problem is less complex to solve than full NILM, several research challenges have been determined during the research undertaken for this paper. First and foremost, characteristic and expressive features are required that allow machine learning algorithms to discriminate between appliances, with ideally no confusion between appliance types. Thus, a key contribution of the paper is the investigation of potential features. Thorough feature engineering from the domains of electrical engineering, information technology, and computer science and engineering, a selection of 517 candidate features has been determined.

Features have been considered in four principal areas; a selection of the features in each category is briefly summarized as follows:

1. *Temporal appliance behavior*: Features in this category describe appliance activities with regard to their timings. To this end, both the durations of active periods as well as the intervals in-between active periods are considered. Moreover, features like the number of appliance activations per day fall into this category.
2. *Power and energy consumption*: For both power and energy (i.e., the integration of power consumption over time) features are extracted that model average and extremal values. Also, a division of the 24 hours of a day into 144 segments of ten minutes length each is performed, and power and energy consumption values for these segments are considered as features.
3. *Power consumption shape*: The way how an appliance's power consumption evolves over time during its activity is also considered as a feature to describe appliances. In particular, the smoothness of a device's load signature and the amplitude changes of power consumption are considered to discriminate between appliances.
4. *Noise level and other statistical features*: Some appliances exhibit periodicities in their power consumptions. Thus, through a transformation to the frequency domain, such recurrences are identified and used to recognize the type of the underlying appliance.

Table 3.1: Ten most relevant macroscopic features according to their information gain.

RANK	FEATURE
1	maximum power consumption in last daily activity phase
2	maximum power consumption throughout the day
3	average power consumed during all daily activity phases
4	average power demand for the complete day
5	average energy demand for appliance activity phases
6	highest power consumption observed during activity phases
7	lowest power consumption observed during activity phases
8	average power consumption during activity phases
9	median duration of activity phases
10	highest encountered drop in power consumption

In order to assess to what extent these features are expedient to recognize appliance types, a supplementary contribution was needed. At the time of the paper’s publication, only a very small number of data sets of macroscopic load signatures was available (cf. Table 1.1). Also, macroscopic traces were only available for a small subset of the appliance types, predominantly IT systems (desktop computers, monitors, printers, and network switches) [KGH+10]. To overcome this limitation, a data set called *tracebase*³ has been collected within the scope of the paper’s research. Composed of more than 1,200 macroscopic load signature traces from more than 100 appliances, the data set has established itself in the NILM community and been used as the foundation for numerous subsequent research works.

Through the combination of the collected load signatures with the candidate features, a system design for appliance recognition has been developed that succeeded in discriminating between 33 appliance types at an accuracy of 95.5%. A supplementary analysis of the information gain of the used candidate features has moreover been conducted, with its results shown in Table 3.1. Thus, a classification of appliances has been demonstrated to be possible, yet insights for future improvements were also determined. For example, devices with similar operational times and consumption ranges (such as various lamps or computer monitors of different types) were often confused with each other. Also, the time window used in the presented approach has always been chosen as one day, i.e., 24 hours. This might be sufficient for many practical application scenarios, given that it is rather uncommon to connect electrical appliances to different outlets every time they are being used. However, other applications might require faster reactions to changed consumption characteristics.

³Available for download at <http://www.tracebase.org>

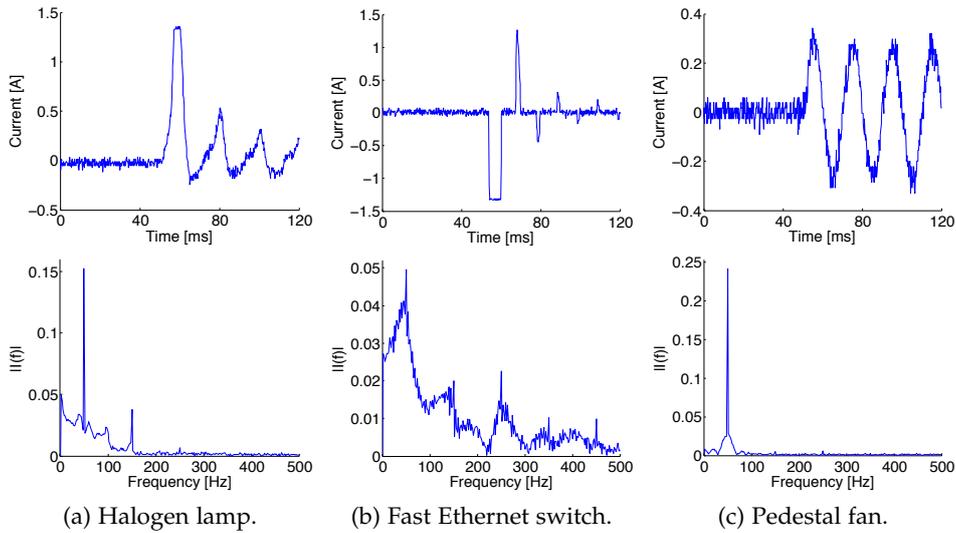


Figure 3.1: Inrush current waveforms and corresponding frequency spectra (as computed by the data collection platform) of three appliances [RBZ+12].

To address the aforementioned limitation, I have conducted further research into accelerating the appliance recognition step. For this purpose, the use of microscopic load signatures has emerged as a viable option. Collected at a much higher temporal resolution, more information content can be extracted from short segments already. An corresponding approach towards the appliance detection based on high-resolution data is documented in:

A. Reinhardt, D. Burkhardt, M. Zaheer, and R. Steinmetz. “Electric Appliance Classification Based on Distributed High Resolution Current Sensing.” In: *Proceedings of the 7th IEEE International Workshop on Practical Issues in Building Sensor Network Applications (SenseApp)*. 2012, pp. 1003–1009.

The paper describes how microscopic current consumption waveforms are being collected by means of a SmartMeter.KOM [RBM+11] sensor at a sampling rate of 1.6 kHz. The resource limitations of the used platform (32 kilobytes of program memory and 2.56 kilobytes of RAM), however, limit the data processing to the computation of a Fast Fourier Transform (FFT) over a duration of 16 mains periods. Subsequently, the amplitudes of the fundamental frequency as well as the first four odd harmonics (i.e., 150 Hz through 450 Hz for a mains frequency of 50 Hz) are forwarded to an external system for storage and analysis. The analysis of spectral information during an appliance’s initial activation and its steady-state operation showed that both greatly vary between devices, and thus qualify as distinctive features for the application of machine learning. Based on further results of the analyses conducted in the paper, it could be shown that the spectra of inrush currents are slightly more expressive than those during steady-state operation. Their visualization in the time and frequency domain is thus shown in Figure 3.1.

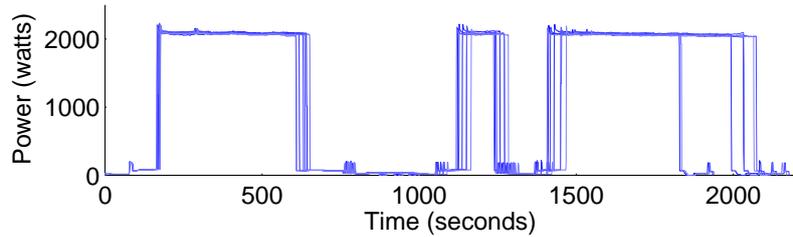
Table 3.2: Ten most relevant microscopic features according to their information gain.

RANK	FEATURE
1	Average current
2	Magnitude of the fundamental frequency (50 Hz) in the spectrum
3	RMS value of the current
4	Peak current
5	Magnitude of the 3rd harmonic (150 Hz) in the spectrum
6	Magnitude of the 5th harmonic (250 Hz) in the spectrum
7	Magnitude of the 7th harmonic (350 Hz) in the spectrum
8	Magnitude of the 9th harmonic (450 Hz) in the spectrum
9	Phase shift
10	Magnitude of DC offset

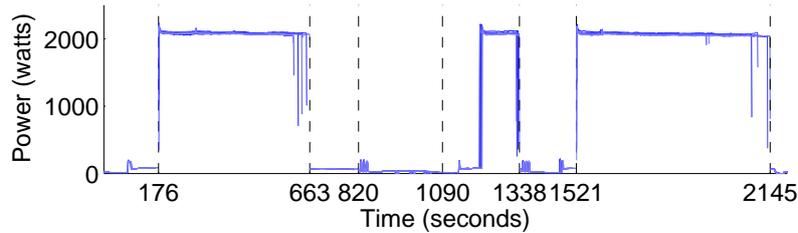
It can be observed that amplitudes in the frequency domain are not only of varying amplitudes (note the different axis scaling), but the odd harmonics also have different amplitudes. Again, an evaluation of the information content of extracted features has been conducted, with results shown in [Table 3.2](#). Features similar to those relevant for macroscopic load signature analysis (as presented above and published in [\[RBB+12\]](#)) again rank among the top four values. However, the harmonic content has been determined to contain a greater information value than, e.g., the phase shift. Executing appliance recognition based on microscopic signatures has a major advantage in terms of recognition delays: Less than one second was required to achieve accuracy levels of up to 98.7%. A more expensive sensor instrumentation is, however, needed in comparison to the operation on macroscopic signatures, which can generally be supplied with data from any off-the-shelf plug-level meter.

3.2 CONSUMPTION PREDICTION

A second application of load signature analysis is the case of predicting future power consumption of appliances. Given the stabilizing effects flexible energy tariffs can have on the power grid [\[SWH+15\]](#), it is important for customers to understand their consumption behavior. In particular, appliances with long operational durations (particularly white goods, see also [Section 2.3](#)) may be started when energy prices are low, but only complete their working cycle at a much later time. Information about a household's expected future consumption can thus allow users to make informed decisions about when appliances should be activated. Knowledge of future demand is also of key interest for large consumers and utility companies alike. While the former class of consumers are often billed by their peak loads, techniques like load shedding or



(a) Traces aligned by starting time.



(b) Traces aligned at each point where major changes in consumption occur.

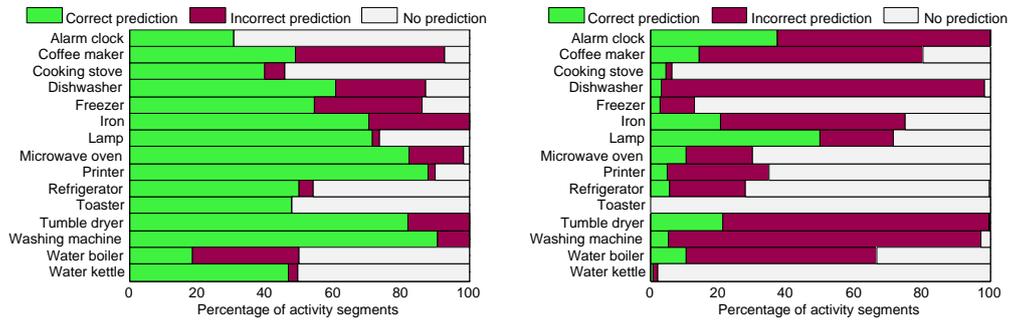
Figure 3.2: Superimposed macroscopic load signatures of seven operation cycles of the same dishwasher appliance [RR16].

on-demand activation (“peak shaving” and “valley filling” [BNJLo8]) are essential to sustain their economical operation. Utility companies, in turn, can leverage information about future electricity demands to optimize their generation infrastructure and reduce the capacity requirements of standby power plants (spinning reserve). To address these requirements, a load forecasting algorithm based on time series analysis was presented in:

A. Reinhardt, D. Christin, and S. S. Kanhere. “Can Smart Plugs Predict Electric Power Consumption? A Case Study.” In: *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous)*. 2014, pp. 257–266.

The rationale behind using characteristics in macroscopic load signatures for load forecasting becomes clear when considering the seven superimposed power consumption traces of a dishwasher shown in Figure 3.2a. The high coherence between operation cycles is obvious, and can be attributed to the predefined sequence of steps in the dishwasher’s program. This becomes particularly obvious when re-aligning the trace at points in time at which significant changes to the power consumption occur, as visualized in Figure 3.2b. Such change points consequently represent potential indicators of future power consumption, at least for a limited duration.

The approach presented in the paper exploits this property of macroscopic load signatures by identifying major changes in consumption and annotating them by the appliance’s subsequent power demand. The detected load changes, termed *prediction signatures* in the paper, are subsequently analyzed with regard to potential ambiguities. Prediction signatures that lead to



(a) Distribution of predictions when using prediction signature databases trained to individual appliance types. (b) Distribution of predictions when using a common database of prediction signatures for all appliance types.

Figure 3.3: Fraction of correct, incorrect, and missing predictions for 15 appliance types; in both cases, the prediction horizon was set to 120 seconds [RCK14].

inconsistent subsequent demands are purged, given that they do not allow for forecasts. In turn, prediction signatures that are always followed by the same (or a sufficiently similar) power consumption pattern, are retained. Due to the limited scale of the data set used in the evaluations, a third class of prediction signatures exist, namely those that occur exactly once in macroscopic load signatures. Such signatures can be seen as an artifact of the data set's limitations. Thus, they are not considered in the evaluation to avoid a falsification of the results.

It also needs to be noted that variations in appliance activity durations limit the time for which predictions can be made reliably. Thus, different prediction horizons, i.e., times for which future consumption annotations are being made, have been evaluated. Selected results for two scenarios are shown in Figure 3.3. In Figure 3.3a, the results are plotted when prediction signatures are extracted and evaluated for each considered appliance type separately. The analysis is hence based on 15 sets of signatures that occur more than once in the training data and lead to consistent forecasting results. The attained results show that a large fraction of correct forecasts are possible for many appliance types, under the precondition that only prediction signatures for the given type are being used. In contrast to the class-wise analysis, Figure 3.3b shows the outcomes when a single prediction signature database is created for all appliance types. Once all ambiguous prediction signatures have been purged, many more incorrect predictions result than in the class-wise analysis. The reason lies within the similarity of power consumption changes across appliance types, yet with different subsequent consumption behavior. An accurate fitting of the prediction signature database to the appliance(s) under consideration is thus crucial for successful predictions.

3.3 ANOMALY DETECTION

A third use case for energy analytics is the detection of consumption anomalies. Analogous to the forecasting future electricity demands, anomaly detection is of relevance for both consumers and suppliers of electrical energy. To TSOs, DSOs, and utility companies, anomalous consumption can be an indicator of faults in the power grid. In industrial sites, anomalies and in particular gradual shifts of consumption patterns are often indicative of the mechanical wear of machinery [ORL+10]; thus, their detection is an enabling element for predictive maintenance. Even in private homes, anomalies often either indicate imminent danger or at least an unusual behavior of the residents. The latter is particularly of interest in Ambient Assisted Living (AAL) scenarios.

From a technical perspective, anomaly detection and load prediction share many commonalities. An expectation of future consumption is mandatory to determine to what extent the actual consumption deviates from it. I have consequently presented insights into the recognition of consumption anomalies based on the observation of regularities in load signatures in the following work:

A. Reinhardt and D. Reinhardt. "Detecting Anomalous Electrical Appliance Behavior based on Motif Transition Likelihood Matrices." In: *Proceedings of the IEEE International Conference on Smart Grid Communications (SmartGrid-Comm) Symposium on Data Management, Grid Analytics and Dynamic Pricing*. 2016, pp. 704–709.

In the work, again prediction signatures are being discovered in macroscopic load signature data. In contrast to the sole matching of such prediction signatures, as performed in the work discussed in Section 3.2, the forecasts are now being compared to an appliance's actual consumption. In case projected and actual power demand match up, the appliance is considered to exhibit a regular operation pattern. In contrast, once a mismatch between prediction and actual demand is detected, an event is signaled to inform interested stakeholders about the occurrence. The paper builds on the concept of a *motif transition likelihood matrix*, a multidimensional structure that maps prediction signatures to the probabilities of future demand. Again, a conversion into symbolic representations (cf. Section 2.3) is applied to macroscopic load signatures to reduce computational efforts. Symbol sequences of a fixed length, referred to as *motifs*, are subsequently extracted, and an analysis is conducted on the temporal distances of their occurrences. During a training phase, the system computes the probabilities of transitions between motifs and inserts them into the transition likelihood matrix. A graphical visualization of the probabilities for motif transitions from the motif AAAB is shown in Figure 3.4. The temporal offset between the motif occurrences is shown on the x-axis, whereas the probabilities of the given transition are cumulatively plotted on the ordinate axis.

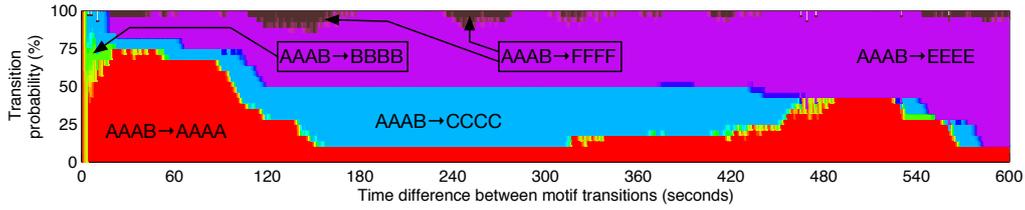
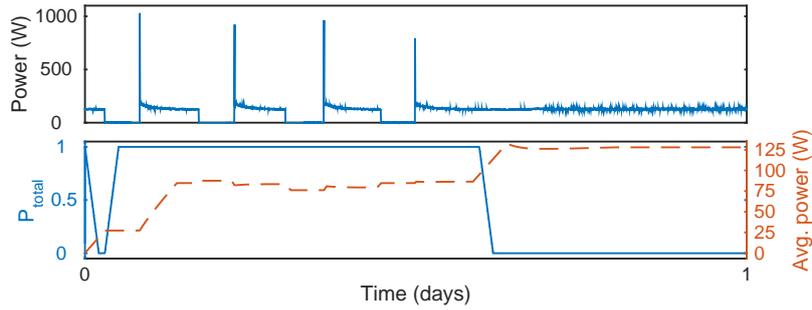
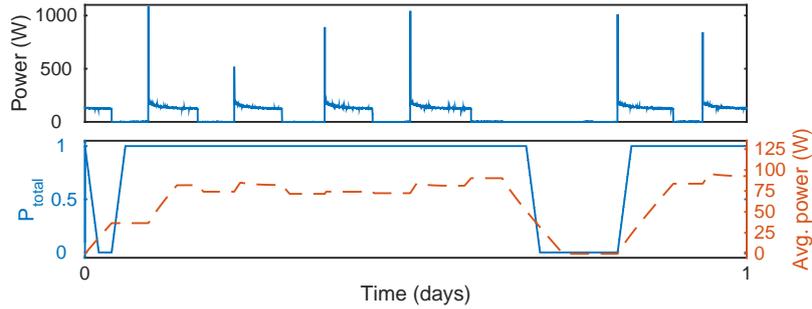


Figure 3.4: Sample symbol transition probability diagram, showing the likelihood of transitions from the motif AAAB to other motifs plotted over the time distance between these motif occurrences [RR16].



(a) Refrigerator with continuous motor operation.



(b) Refrigerator with unexpected stopping and resuming of motor.

Figure 3.5: Assessment of the motif-based anomaly detection algorithm’s output P_{total} in comparison to the appliance’s averaged power consumption for two sample anomaly cases [RR16].

Once trained and adapted to the regular patterns of the appliance under consideration, the system is ready to operate. For each newly incoming power consumption sample, its motif representation is computed, and the probability of regular appliance operation, P_{total} , is being computed. An example for the anomaly detection performance is given in Figure 3.5, where two faults have been simulated for a refrigerator appliance. P_{total} , plotted as a continuous line in the lower subdiagrams, outputs an almost binary representation of the likeliness of normal operation ($P_{total} = 1$) during the regular cooling cycles, yet quickly drops to zero when irregularities are encountered. For comparison, the mean values of the power consumptions (dashed lines) show strong fluctuations, thus their potential for anomaly detection is limited.

3.4 LOAD MODELING FOR AUTOMATED EVALUATIONS

Many energy informatics research challenges are encountered by using tools from the domain of artificial intelligence, in particular machine learning. Most often, one or more of the publicly available data sets (cf. [Table 1.1](#)) are being used to evaluate the performance of newly developed algorithms and systems. Undiscovered annotation errors or incomplete data thus manifest themselves not only in the data sets, but also in all papers based upon them. In many cases, the limited scale of the data sets also expresses itself in an effect known as *overfitting*. Overfitting takes place when machine learning algorithms adapt too well to the characteristics of the training data set and can as a result no longer be generalized to other, yet unseen, data. Mitigating the shortage of data sets is not trivial, given the efforts required to set up, operate, and maintain energy data collection campaigns. Thus, other means to source input data for research and development purposes are needed. An approach that relies on the modeling of macroscopic load signatures and their subsequent synthesis into representative and annotated traces is presented in:

N. Buneeva and A. Reinhardt. “AMBAL: Realistic Load Signature Generation for Load Disaggregation Performance Evaluation.” In: *Proceedings of the IEEE International Conference on Smart Grid Communications (SmartGridComm) Symposium on Big Data Management and Analytics*. 2017, pp. 443–448.

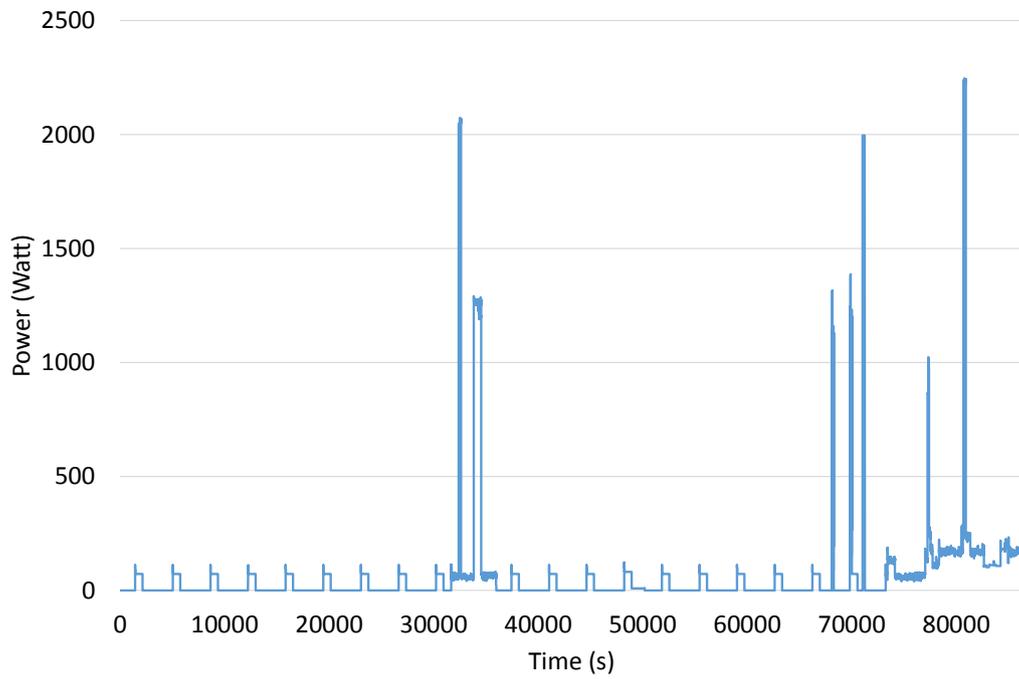
The developed system is a tool to analyze macroscopic load signatures and derive accurate appliance-level power consumption models. Generated models are composed of parametrizable basic building blocks, a concept initially presented by Barker et al. in [\[BKI+13\]](#). Through their combination, appliance power consumption can be approximated to the desired accuracy level. A generalization step used in the model-building process ensures that representative, yet compact, descriptions of consumption patterns are being extracted for each modeled appliance.

AMBAL has been tested by modeling the consumption behavior of 24 appliances (14 from Tracebase [\[RBB+12\]](#) and 10 from ECO [\[BKC+14\]](#)). It could be observed that for many appliance types (e.g., water kettle, refrigerator, freezer, microwave oven) models can be composed of less than five components, while still not exceeding a modeling error of 6%.

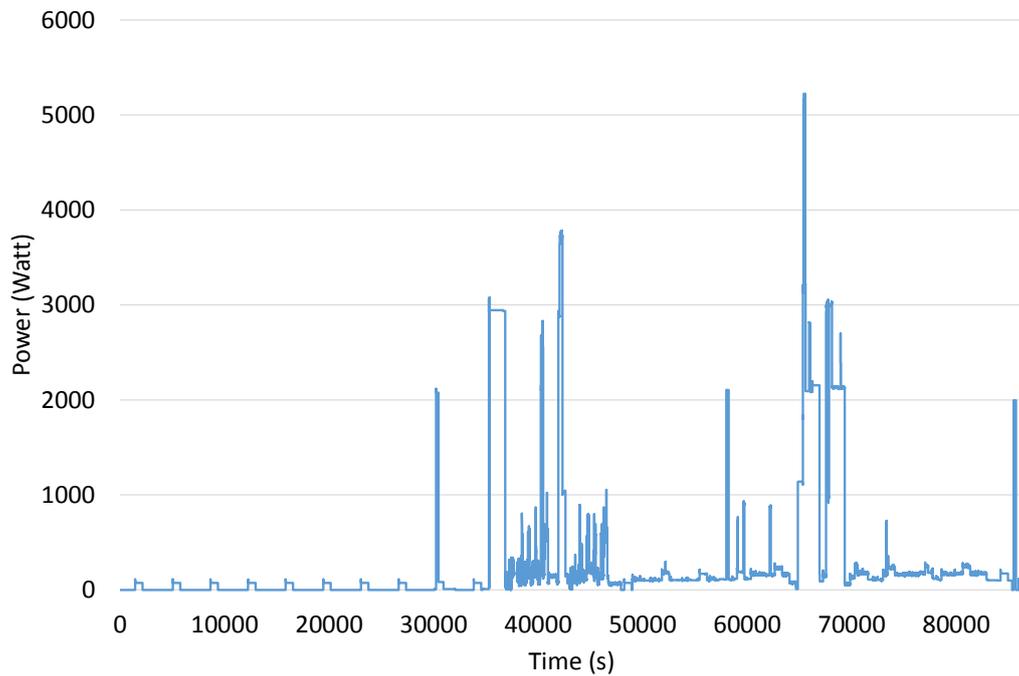
Furthermore, AMBAL has been augmented by a tool to synthesize formerly extracted appliance consumption models into artificial, yet realistic-looking, traces for their use in testing and improving energy analytics algorithms. Moreover, by coupling AMBAL with a user activity models (following the work of Richardson et al. [\[RTI08\]](#)), the foundation for the generation of fully synthetic, yet realistic-looking, aggregate consumption traces has been created. A visualization of the potential of this synergistic combination is shown in [Figure 3.6](#). Models for 14 electrical household appliances have been derived using AMBAL, and recombined according to the activity models for a single-

person household as well as for a household that accommodates three persons. Besides the generation of time series data, AMBAL also outputs annotations about appliance de-/activations. The workflow for evaluating NILM algorithms using AMBAL was chosen as follows: A large number of appliance-level traces are generated through AMBAL and forwarded to a software NILM tool called nilmtk⁴ for training its internal models. After that step, AMBAL is executed once again to create a new set of synthetic macroscopic load signatures as well as an aggregated trace that contains the summation of the loads of all contributing devices. The latter trace is then forwarded to nilmtk for disaggregation. In a final step, the disaggregated components are compared against the appliance-level traces created by AMBAL. Through changing the number of traces that are generated and aggregated, input data of different complexity levels [EPE15] can be used in the evaluation of NILM algorithm performance.

⁴nilmtk is available for download at <http://nilmtk.github.io>



(a) Synthetic aggregate consumption trace for a household with one occupant.



(b) Synthetic aggregate consumption trace for a household with three occupants.

Figure 3.6: Examples of aggregated daily power consumption traces generated using a set of 14 appliances with appliance usage times defined by user activity models [Bun16].

As highlighted in the previous sections, the analysis and interpretation of electricity consumption data is a groundbreaking technology to enable many novel services. One major obstacle on the path towards a widespread adoption of energy analytics is, however, its acceptance by customers. More precisely, the threat of inferring the presence of appliances, detecting daily routines and habits, or flagging times during which the customer is out of the house potentially represent serious intrusions into user privacy. While privacy considerations are usually not imminent when instrumenting transmission and distribution grids with sensing equipment, their threat grows when metering devices are deployed closer to the customer premises. Residential characteristics can already be inferred when aggregated consumption data for housing blocks are available, and become increasingly detailed once individual homes are monitored in a fine-grained fashion [WHM+12; BSS+15].

Within the field of IT security, several solution approaches to this problem are being investigated. However, schemes such as differential privacy [Dwo06] or the use of homomorphic encryption [RAD78], can only partly mitigate privacy considerations. The fundamental assumption for their use is the presence of a trusted third party, usually the metering point operator. In addition to billing customers for their consumed electricity, however, business models for this stakeholder also include other ways of exploring and exploiting the information content in consumption data. In fact, a strong growth in the number of companies who build their business models on the analysis of consumption data can be observed. Thus, it ensues that orthogonal approaches to protect user privacy are required. Two such approaches, based on the intentional reporting of falsified data and the use of domestic PV generation, are presented in the remainder of this section.

4.1 INTENTIONAL DATA PERTURBATION

The non-linear relation between the granularity at which data are collected and the resultant energy analytics accuracy has been demonstrated in [EK10; LES+14; TSB+14]. An intuitive way to minimize the impact on user privacy when electricity consumption data are being collected is thus the reduction of the data's resolution. Such resolution reductions can be performed in two possible ways: Either the data's temporal resolution is lowered, e.g., by down-sampling data, or its amplitude resolution is reduced, commonly accomplished through quantization. To assess the impact of such data processing on user privacy, I have conducted a comprehensive study, published in:

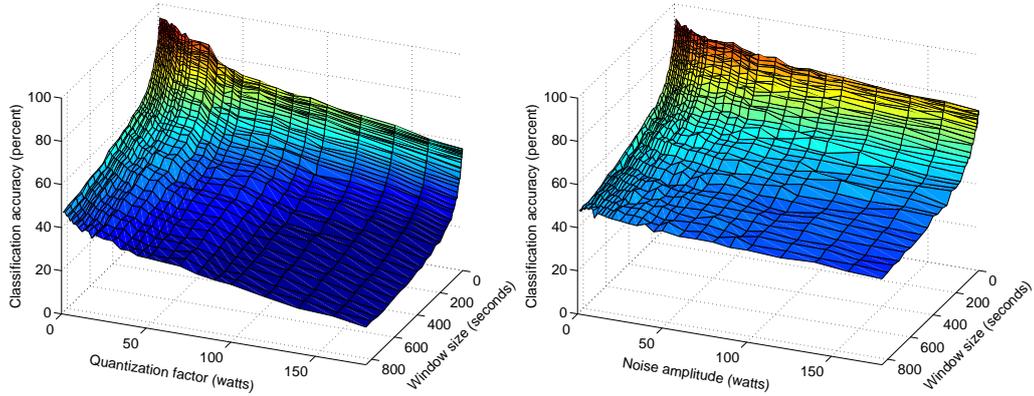
A. Reinhardt, F. Englert, and D. Christin. “Averting the Privacy Risks of Smart Metering by Local Data Preprocessing.” In: *Pervasive and Mobile Computing (PMC)* 16 (2015), pp. 171–183.

The study is based on macroscopic load signatures and the appliance recognition approach presented in [RBB+12] (cf. Section 3.1). Instead of using the input data in its unaltered form, i.e., one sample per second, the following data preprocessing steps to reduce the data resolution are considered:

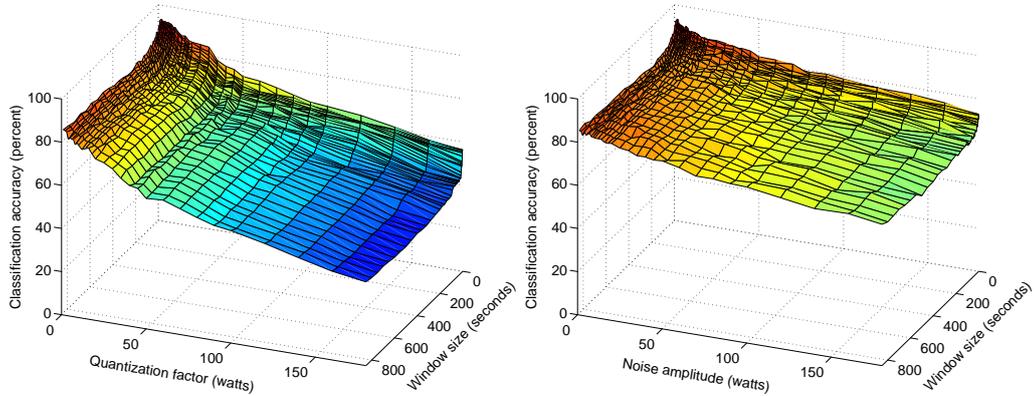
- ▷ *Temporal averaging* lowers the temporal resolution of load signatures by reporting the average power consumption of a window containing a fixed number of most recent values whenever a new sample is collected.
- ▷ In *temporal down-sampling*, a sample of the data is taken at a rate lower than the native data resolution and repeatedly reported for a fixed time duration; all intermediate load changes remain unreported.
- ▷ *Noise addition* is a way of modifying the amplitude of power consumption readings by overlaying random fluctuations on the reported data. It is particularly beneficial when characteristic patterns in appliance power consumption have small amplitudes, thus they can be covered in the added noise, which leads to a potentially higher privacy protection.
- ▷ *Amplitude quantization* is a way to alter the load signature’s amplitude by rounding each consumption sample to a multiple of a pre-defined quantization factor.

All combinations of the two amplitude alteration schemes with the two approaches to alter the temporal resolution of data have been analyzed with regard to their impact on the appliance recognition accuracy. Moreover, *averaged down-sampling*, i.e., the combination of an averaging step with a subsequent down-sampling, has been considered in the analysis. The impact of these data filters on the device recognition accuracy is shown in Figure 4.1. Across all diagrams, an appliance recognition rate of 93.5% is achieved when feeding unaltered data to the appliance recognition system. Note that this is slightly lower than the 95.5% determined in Section 3.1 due to the fact that an extended version of the data set with more appliance diversity has been used here. The figure shows strongly differing accuracy degradations for different choices of the perturbation methods’ parameterizations. The combination of temporal down-sampling with value quantization (shown in Figure 4.1a; lowest observed recognition rate at 36.9%) provides much better privacy protection than the choice of noise addition and temporal averaging (see Figure 4.1d and Figure 4.1f; lowest observed recognition rate 56.1%).

Reporting of intentionally perturbed data may lead to inaccuracies in billing, yet such errors can be bounded by the choice of the quantization and downsampling values. At the same time, however, making less information available to the operator of the data collection infrastructure implicitly lowers the risk of intrusions into user privacy.

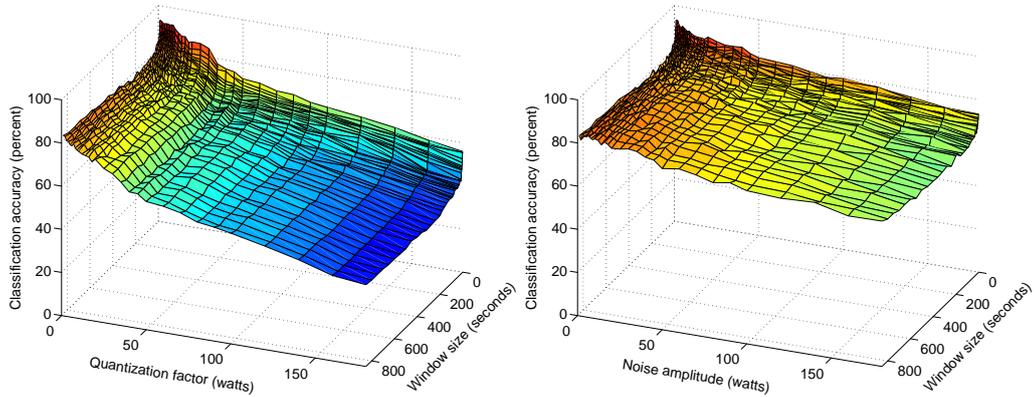


(a) Temporal down-sampling vs. quantization. (b) Temporal down-sampling vs. noise addition.



(c) Averaging vs. quantization.

(d) Averaging vs. noise addition.



(e) Averaging/down-sampling vs. quantization. (f) Averaging/down-sampling vs. noise addition.

Figure 4.1: Resulting device recognition accuracies when the preprocessing filters have been applied to the input data [REC15].

4.2 LOAD HIDING THROUGH RENEWABLE POWER GENERATION

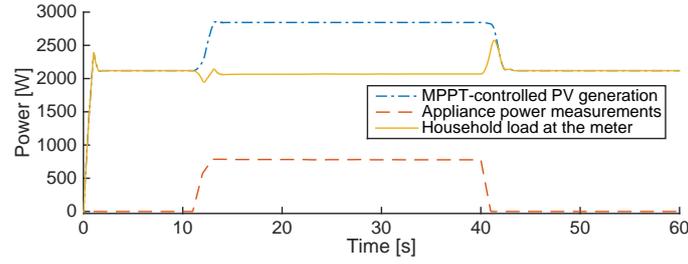
An intentional degradation of data quality by means of the removal or perturbation of characteristic signals can be used to limit the potential of energy analytics. While this generally leads to improvements in user privacy, it may not always be possible to report modified data. Particularly when data shall be used for billing purposes, legislative requirements usually impose strict accuracy requirements. Alternatives to the intentional reporting of falsified data have thus emerged through the local installation of controllable consumers [EPE14] or energy storages [KED+10]. By using energy stored in local batteries to supply appliances with electricity, their existence can effectively be obfuscated or even completely covered. As a downside, potentially large banks of batteries might be required to satisfactorily hide the presence of large consumers. This is partly mitigated when following the approach of load-based load hiding, i.e., the load signature perturbation through the addition of controllable consumption. However, besides the monetary cost incurred when operating appliances for the purpose of hiding consumption patterns, high maintenance efforts also result from the requirement to keep appliances ready at all times.

Photovoltaic (PV) generation, a renewable energy source, is widely deployed in Germany. In 2018, almost 24% of the country's total energy generation were supplied by PV installations⁵. The strong dependency on the prevailing environmental conditions renders the output PV installations volatile by nature. Their generation is strongly dependent on the current weather conditions, and small variations in insolation and cloud coverage can already lead to drastic changes in the output power. Moreover, the fact that the solar radiation can even vary even between neighboring homes when shattered clouds are present, makes solar generation hard to predict; a strong asset when leveraging it for privacy protection. It is thus conceivable to use PV systems to improve user privacy protection, an analysis of which is presented in:

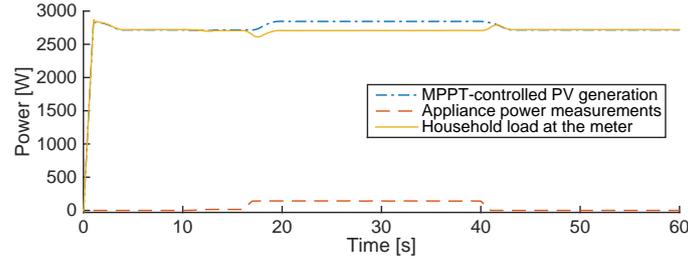
A. Reinhardt, G. Konstantinou, D. Egarter, and D. Christin. "Worried About Privacy? Let Your PV Converter Cover Your Electricity Consumption Fingerprints." In: *Proceedings of the IEEE International Conference on Smart Grid Communications (SmartGridComm) Symposium on Cyber Security and Privacy*. 2015, pp. 25–30.

In the work, a coupling is created between macroscopic load signatures collected in real-time and the operating point of a PV converter. A converter is a device that transforms the Direct Current (DC) output of solar cells into an Alternating Current (AC) voltage of utility frequency that can be supplied to the power grid. During regular operation of the converter, a Maximum Power Point Tracking (MPPT) algorithm is being executed to maximize its delivered power. Thus, no further increases to its output power are possible, yet controlled reductions are viable and easy to realize. By modifying the MPPT

⁵Up-to-date charts are available at https://www.energy-charts.de/energy_pie.htm



(a) Hidden toaster load signature.



(b) Hidden LCD TV load signature.

Figure 4.2: Household aggregates when appliance operations are hidden by using real-time measurements to control a PV converter’s MPPT algorithm [RKE+15].

algorithm to intentionally deviate from the optimum operating point, it is possible to feed less power into the grid. Assuming a household’s consumption remains unchanged, this reduction in generated power will appear identical to an appliance’s operation when considering the aggregate balance. It thus become possible to “play back” the macroscopic load signatures of appliances and simulate the presence of appliances, even though they may not even be physically present in the dwelling. Furthermore, when operating the converter in a region below its maximum power point, controlled generation changes become possible in both directions. Thus, it becomes possible to completely hide the operation of appliances. Through a direct coupling of load signature samples sensed at appliance-level with the converter’s MPPT algorithm, the power intake of the appliance can be fully compensated for through an increased generation.

An example for this is shown in Figure 4.2, in which the operations of two appliances (a toaster and an LCD TV) shall be hidden from the aggregate load reported from a smart meter. Through operating the converter at an output power slightly below its maximum of 3,000 watts, sufficient power swings in both the positive and then negative direction are possible. Consequently, the aggregate load (shown as a continuous line) only experiences slight peaks during the moment the appliance is (de)activated. This effect could be traced back to communication delays between the actual moment of measurement and the reaction of the PV converter. Using microscopic load signatures would allow for a faster reaction and likely mitigate these artifacts altogether. A very high degree of privacy protection can thus be accomplished through the application of the designed system.

CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

Energy informatics is an emerging research field that has sparked interest in many different research communities. Across all involved disciplines, the unequivocal objective is to create a *smart* power grid, i.e., an infrastructure for the transport of electrical power that is aware of its current status and can adapt and react to changes. To turn this vision into reality, many contributions along the flow of electrical consumption data are required. Some of them have already been extensively considered in research for several decades, e.g., investigations into ways to accomplish Non-Intrusive Load Monitoring. In contrast, many other fundamental challenges, e.g., how to minimize the number of sensing devices in a building while maximizing the information content in their collected data, have barely received any scientific coverage yet. Energy informatics is thus still dependent on more novel contributions to make smart grids a reality.

Twelve such contributions were presented in this manuscript, all of which have been summarized in the previous chapters. Their order of presentation has been aligned with the flow of electricity consumption data in electrical power grids: From their collection (e.g., by means of a smart meter), via their networked transmission to a data collection device (e.g., a concentrator node), to the processing at their destination (e.g., a notification system that alarms an operator in case of consumption anomalies). Boundary conditions from the multitude of involved domains, including electrical power engineering, measurement engineering, information and communication technology, computer science and engineering, and economics, were considered in the designs and implementations of the contributions presented in this habilitation thesis. Only by considering energy informatics from a holistic and interdisciplinary viewpoint, practical and usable solutions with a high user acceptance will emerge.

The way to truly smart power grids is, however, still long and winding, and the following research challenges remain as major obstacles on this path.

- ▷ Collecting data in a methodological way: How and where must sensors be deployed to capture all necessary data in the required resolution?
- ▷ Networking between data-collecting devices: How can the data collected from distributed sensors be exchanged in a scalable and timely fashion?
- ▷ Universally applicable data processing algorithms: Which methods can extract the greatest information content from all collected data, while being resilient to varying data quality levels and providing adequate means for user privacy protection?

Expressive data, combined with powerful analysis methods, will enable novel opportunities for energy informatics. Let us see what the future has in store.

REFERENCES

-
- [ASS+02] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. "A Survey on Sensor Networks." In: *IEEE Communications Magazine* 40.8 (2002).
- [APR15] J. Alcalá, O. Parson, and A. Rogers. "Health Monitoring of Elderly Residents via Disaggregated Smart Meter Data and Log Gaussian Cox Processes." In: *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys)*. 2015.
- [AXR+15] A. Alhamoud, P. Xu, A. Reinhardt, F. Englert, P. Scholl, D. Böhnstedt, and R. Steinmetz. "Extracting Human Behavior Patterns from Appliance-level Power Consumption Data." In: *Proceedings of the 12th European Conference on Wireless Sensor Networks (EWSN)*. 2015.
- [AOB+12] K. Anderson, A. Ocneanu, D. Benitez, D. Carlson, A. Rowe, and M. Bergés. "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)*. 2012.
- [AKM12] H.-J. Appelrath, H. Kagermann, and C. Mayer, eds. *Future Energy Grid. Migrationspfade ins Internet der Energie (acatech STUDIE)*. Heidelberg u.a.: Springer Verlag, 2012.
- [AKG+15] P. Arjunan, H. D. Khadilkar, T. Ganu, Z. M. Charbiwala, and A. S. P. Singh. "Multi-User Energy Consumption Monitoring and Anomaly Detection with Partial Context Information." In: *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys)*. 2015.
- [AGS+13] K. C. Armel, A. Gupta, G. Shrimali, and A. Albert. "Is Disaggregation the Holy Grail of Energy Efficiency? The Case of Electricity." In: *Energy Policy* 52.1 (2013).
- [Ash09] K. Ashton. "That 'Internet of Things' Thing." In: *RFID Journal* (2009). Available online at: <http://www.rfidjournal.com/articles/pdf?4986>.
- [BNJLo8] A. Bar-Noy, M. Johnson, and O. Liu. "Peak Shaving through Resource Buffering." In: *Proceedings of the 6th Workshop on Approximation and Online Algorithms (WAOA)*. 2008.

- [BV04] M. Baranski and J. Voss. "Genetic Algorithm for Pattern Detection in NIALM Systems." In: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (ICSMC)*. 2004.
- [BKI+13] S. Barker, S. Kalra, D. Irwin, and P. Shenoy. "Empirical Characterization and Modeling of Electrical Loads in Smart Homes." In: *Proceedings of the 4th International Green Computing Conference (IGCC)*. 2013.
- [BMI+12] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht. "Smart*: An Open Data Set and Tools for Enabling Research in Sustainable Homes." In: *Proceedings of the Workshop on Data Mining Applications in Sustainability (SustKDD)*. 2012.
- [BGS+13] N. Batra, M. Gulati, A. Singh, and M. B. Srivastava. "It's Different: Insights into Home Energy Consumption in India." In: *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings (BuildSys)*. 2013.
- [BKP+14] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava. "NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring." In: *Proceedings of the 5th ACM International Conference on Future Energy Systems (e-Energy)*. 2014.
- [BKC+14] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, and S. Santini. "The ECO Data Set and the Performance of Non-Intrusive Load Monitoring Algorithms." In: *Proceedings of the 1st ACM International Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys)*. 2014.
- [BSS+15] C. Beckel, L. Sadamori, S. Santini, and T. Staake. "Automated Customer Segmentation Based on Smart Meter Data with Temperature and Daylight Sensitivity." In: *Proceedings of the 6th IEEE International Conference on Smart Grid Communications (SmartGridComm)*. 2015.
- [BSS+14] C. Beckel, L. Sadamori, T. Staake, and S. Santini. "Revealing Household Characteristics from Smart Meter Data." In: *Elsevier Energy* 78 (2014).
- [BGM+01] M. E. Bergés, E. Goldman, H. S. Matthews, and L. Soibelman. "Enhancing Electricity Audits in Residential Buildings with Nonintrusive Load Monitoring." In: *Journal of Industrial Ecology* 14.5 (2001).

- [BGS+11] M. E. Bergés, E. Goldman, L. Soibelman, H. S. Matthews, and K. Anderson. "User-centered Non-Intrusive Electricity Load Monitoring for Residential Buildings." In: *Journal of Computing in Civil Engineering* 25.1 (2011).
- [BGM+09] M. Bergés, E. Goldman, H. Matthews, and L. Soibelman. "Learning Systems for Electric Consumption of Buildings." In: *Proceedings of the ASCE International Workshop on Computing in Civil Engineering (IWCCE)*. 2009.
- [BLCG92] T. J. Berners-Lee, R. Cailliau, and J.-F. Groff. "The World-Wide Web." In: *Computer Networks and ISDN Systems* 25 (1992).
- [BMNo9] V. Bianco, O. Manca, and S. Nardini. "Electricity Consumption Forecasting in Italy using Linear Regression Models." In: *Energy* 34.9 (2009).
- [Blo70] B. H. Bloom. "Space/Time Trade-offs in Hash Coding with Allowable Errors." In: *Communications of the ACM* 13.7 (1970).
- [Bol99] M. H. Bollen. *Understanding Power Quality Problems: Voltage Sags and Interruptions*. Wiley-IEEE Press, 1999.
- [Bun16] N. Buneeva. "Automated Modeling of Appliance Signatures for Realistic Consumption Trace Synthesis." M.Sc. Thesis. Technische Universität Clausthal, 2016.
- [CEN12] CEN-CENELEC-ETSI Smart Grid Coordination Group. *Smart Grid Reference Architecture*. Online: http://gridscientific.com/images/Smart_Grid_Reference_Artitecture.pdf. 2012.
- [CCT+12] H. H. Chang, K. L. Chen, Y. P. Tsai, and W. J. Lee. "A New Measurement Method for Power Signatures of Nonintrusive Demand Monitoring and Load Identification." In: *IEEE Transactions on Industry Applications* 48.2 (2012).
- [CLL10] H. Chang, C. Lin, and J. Lee. "Load Identification in Non-intrusive Load Monitoring Using Steady-State and Turn-On Transient Energy Algorithms." In: *Proceedings of the 14th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. 2010.
- [CLYo8] H.-H. Chang, C.-L. Lin, and H.-T. Yang. "Load Recognition for Different Loads with the Same Real Power and Reactive Power in a Non-intrusive Load-monitoring System." In: *Proceedings of the 12th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. 2008.

- [CDW+11] F. Chen, J. Dai, B. Wang, S. Sahu, M. Naphade, and C.-T. Lu. "Activity Analysis Based on Low Sample Rate Smart Meters." In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 2011.
- [CLX+06] Y. Chen, H. V. Leong, M. Xu, J. Cao, K. C. C. Chan, and A. T. S. Chan. "In-Network Data Processing for Wireless Sensor Networks." In: *Proceedings of the 7th International Conference on Mobile Data Management (MDM)*. 2006.
- [CCD14] M. Clark, B. Campbell, and P. Dutta. "Deltaflow: Submetering by Synthesizing Uncalibrated Pulse Meter Streams." In: *Proceedings of the 5th ACM International Conference on Future Energy Systems (e-Energy)*. 2014.
- [CA00] 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 (IMTC)*. 2000.
- [CA98] A. I. Cole and A. Albicki. "Data Extraction for Effective Non-Intrusive Identification of Residential Power Loads." In: *Proceedings of the 15th Instrumentation and Measurement Technology Conference (IMTC)*. 1998.
- [CM02] D. Comaniciu and P. Meer. "Mean Shift: A Robust Approach Toward Feature Space Analysis." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.5 (2002).
- [CEE16] Council of European Energy Regulators (CEER). *6th CEER Benchmarking Report on the Quality of Electricity and Gas Supply*. Online: https://www.nve.no/Media/4862/1-c16-eqs-72-03_ceer-6thbr_intro-ch1-2-2.pdf. 2016.
- [CLS+06] R. Cox, S. Leeb, S. Shaw, and L. Norford. "Transient Event Detection for Nonintrusive Load Monitoring and Demand Side Management using Voltage Distortion." In: *Proceedings of the Applied Power Electronics Conference and Exposition (APEC)*. 2006.
- [Cox04] R. W. Cox. *A Field-based Approach to Non-Intrusive Load Monitoring*. Proceedings of the "Inference from Data and Models" Workshop at the Massachusetts Institute of Technology, USA. 2004.
- [DDD+17] L. De Baets, C. Develder, D. Deschrijver, and T. Dhaene. "Automated Classification of Appliances using Elliptical Fourier Descriptors." In: *Proceedings of the IEEE International Conference on Smart Grid Communications (SGC)*. 2017.

- [DRD+18] L. De Baets, J. Ruyssinck, C. Develder, T. Dhaene, and D. Deschrijver. "Appliance Classification using VI Trajectories and Convolutional Neural Networks." In: *Energy and Buildings* 158 (2018).
- [DCD13] S. DeBruin, B. Campbell, and P. Dutta. "Monjolo: An Energy-harvesting Energy Meter Architecture." In: *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (SenSys)*. 2013.
- [DoE11] Department of Energy and Climate Change and the Office of Gas and Electricity Markets. *Functional Requirements Catalogue*. Online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/42737/1480-design-requirement-annex.pdf. 2011.
- [DK99] S. Drenker and A. Kader. "Nonintrusive Monitoring of Electric Loads." In: *IEEE Computer Applications in Power* 12.4 (1999).
- [DHH+16] L. Du, D. He, R. G. Harley, and T. G. Habetler. "Electric Load Classification by Binary Voltage-Current Trajectory Mapping." In: *IEEE Transactions on Smart Grid* 7.1 (2016).
- [Dwo06] C. Dwork. "Differential Privacy." In: *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (ICALP)*. 2006.
- [EF14] The Edison Foundation – Institute for Electric Innovation. *Utility-scale Smart Meter Deployments: Building Block of the Evolving Power Grid*. IEI Report. 2014.
- [EK10] C. Efthymiou and G. Kalogridis. "Smart Grid Privacy via Anonymization of Smart Metering Data." In: *Proceedings of the 1st IEEE International Conference on Smart Grid Communications (SmartGridComm)*. 2010.
- [EPE14] D. Egarter, C. Prokop, and W. Elmenreich. "Load Hiding of Household's Power Demand." In: *Proceedings of the 5th IEEE International Conference on Smart Grid Communications (SmartGridComm)*. 2014.
- [EPE15] D. Egarter, M. Pöchacker, and W. Elmenreich. "Complexity of Power Draws for Load Disaggregation." In: *arXiv Computing Research Repository (CoRR)* abs/1501.02954 (2015).
- [EEK+15] F. Eichinger, P. Efron, S. Karnouskos, and K. Böhm. "A Time-Series Compression Technique and Its Application to the Smart Grid." In: *The VLDB Journal* 24.2 (2015).

- [EST13] Energy Saving Trust. *Opportunities for Communities Through Energy Storage (OCTES) – Final Project Report*. Online: <http://octesnpp.eu/documents/SMAPReportMarch2013.pdf>. 2013.
- [EC14] The European Commission. “Benchmarking Smart Metering Deployment in the EU-27 with a Focus on Electricity.” In: *Report from the Commission COM/2014/356/final* (2014).
- [EU12] The European Parliament and the Council of the European Union. “Directive 2012/27/EU on Energy Efficiency.” In: *Official Journal of the European Union L 315* (2012).
- [EU16] The European Parliament and the Council of the European Union. “Regulation (EU) 2016/679: Protection of Natural Persons with regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation).” In: *Official Journal of the European Union L 119/1* (2016).
- [Eva11] D. Evans. *The Internet of Things – How the Next Evolution of the Internet is Changing Everything*. Cisco Internet Business Solutions Group (IBSG) White Paper. 2011.
- [FZ99] L. Farinaccio and R. Zmeureanu. “Using a Pattern Recognition Approach to Disaggregate the Total Electricity Consumption in a House into the Major End-uses.” In: *Elsevier Energy and Buildings* 30.3 (1999).
- [FLG+11] J. Froehlich, E. Larson, S. Gupta, G. Cohn, M. S. Reynolds, and S. N. Patel. “Disaggregated End-Use Energy Sensing for the Smart Grid.” In: *Pervasive Computing* 10.1 (2011).
- [GGK+14] J. Gao, S. Giri, E. C. Kara, and M. Bergés. “PLAID: A Public Dataset of High-resolution Electrical Appliance Measurements for Load Identification Research.” In: *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys)*. 2014.
- [GRS14] M. Gulati, S. S. Ram, and A. Singh. “An in Depth Study into Using EMI Signatures for Appliance Identification.” In: *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys)*. 2014.
- [GRP10] S. Gupta, M. S. Reynolds, and S. N. Patel. “ElectriSense: Single-point Sensing Using EMI for Electrical Event Detection and Classification in the Home.” In: *Proceedings of the 12th ACM International Conference on Ubiquitous Computing (UbiComp)*. 2010.

- [HKK+17] A. U. Haq, T. Kriechbaumer, M. Kahl, and H. A. Jacobsen. "CLEAR: A Circuit Level Electric Appliance Radar for the Electric Cabinet." In: *IEEE International Conference on Industrial Technology (ICIT)*. 2017.
- [Har85] G. W. Hart. *Prototype Nonintrusive Appliance Load Monitor*. Tech. rep. MIT Energy Laboratory and Electric Power Research Institute, 1985.
- [Har89] G. W. Hart. "Residential Energy Monitoring and Computerized Surveillance via Utility Power Flows." In: *IEEE Technology and Society Magazine* 8.2 (1989).
- [Har92] G. W. Hart. "Nonintrusive Appliance Load Monitoring." In: *Proceedings of the IEEE* 80.12 (1992).
- [HBA+12] L. Hernández, C. Baladrón, J. M. Aguiar, L. Calavia, B. Carro, A. Sánchez-Esguevillas, D. J. Cook, D. Chinarro, and J. Gómez. "A Study of the Relationship between Weather Variables and Electric Power Demand inside a Smart Grid/Smart World Framework." In: *MDPI Sensors* 12.9 (2012).
- [HKC+11] B.-J. Ho, H.-L. C. Kao, N.-C. Chen, C.-W. You, H.-H. Chu, and M.-S. Chen. "HeatProbe: A Thermal-based Power Meter for Accounting Disaggregated Electricity Usage." In: *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp)*. 2011.
- [IBM12] IBM Software. *Managing Big Data for Smart Grids and Smart Meters*. Information management white paper. Available for download at http://www-935.ibm.com/services/multimedia/Managing_big_data_for_smart_grids_and_smart_meters.pdf. 2012.
- [IEEE06] IEEE Std. 802.15.4 Part 15.4: *Wireless medium access control (MAC) and Physical layer (PHY) specifications for low-rate wireless personal area networks (LR-WPANs)*. 2006.
- [IEA10] International Energy Agency. *World Energy Outlook*. Available at <http://www.worldenergyoutlook.org/media/weo2010.pdf>. 2010.
- [IUI+04] 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: *Proceedings of the International Symposium on Applications and the Internet (SAINT)*. 2004.
- [JDHD+09] X. Jiang, S. Dawson-Haggerty, P. Dutta, and D. E. Culler. "Design and Implementation of a High-Fidelity AC Metering Network." In: *Proceedings of the 8th International Conference on Information Processing in Sensor Networks (IPSN)*. 2009.

- [JAL12] P. Jokar, N. Arianpoo, and V. C. M. Leung. “A Survey on Security Issues in Smart Grids.” In: *Security and Communication Networks* 9.3 (2012).
- [JS10] D. Jung and A. Savvides. “Estimating Building Consumption Breakdowns Using ON/OFF State Sensing and Incremental Sub-meter Deployment.” In: *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems (SenSys)*. 2010.
- [KWH13] H. Kagermann, W. Wahlster, and J. Helbig, eds. *Deutschlands Zukunft als Produktionsstandort sichern: Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0, Abschlussbericht des Arbeitskreises Industrie 4.0*. acatech – Deutsche Akademie der Technikwissenschaften e.V., 2013.
- [KUK+16] M. Kahl, A. Ul Haq, T. Kriechbaumer, and H.-A. Jacobsen. “WHITED – A Worldwide Household and Industry Transient Energy Data Set.” In: *Proceedings of the 3rd International Workshop on Non-Intrusive Load Monitoring (NILM)*. 2016.
- [KED+10] G. Kalogridis, C. Efthymiou, S. Z. Denic, T. A. Lewis, and R. Cepeda. “Privacy for Smart Meters: Towards Undetectable Appliance Load Signatures.” In: *First IEEE International Conference on Smart Grid Communications (SmartGridComm)*. 2010.
- [KGH+10] M. Kazandjieva, O. Gnawali, B. Heller, P. Levis, and C. Kozyrakis. *Identifying Energy Waste through Dense Power Sensing and Utilization Monitoring*. Tech. rep. CSTR 2010-03. Stanford University, 2010.
- [KK14] J. Kelly and W. J. Knottenbelt. “UK-DALE: A Dataset Recording UK Domestic Appliance-Level Electricity Demand and Whole-House Demand.” In: *arXiv Computing Research Repository (CoRR) abs/1404.0284* (2014).
- [KK15] J. Kelly and W. Knottenbelt. “Neural NILM: Deep Neural Networks Applied to Energy Disaggregation.” In: *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys)*. 2015.
- [KSC+09] Y. Kim, T. Schmid, Z. M. Charbiwala, and M. B. Srivastava. “ViridiScope: Design and Implementation of a Fine Grained Power Monitoring System for Homes.” In: *Proceedings of the 11th International Conference on Ubiquitous Computing (UbiComp)*. 2009.

- [KBS15] W. Kleiminger, C. Beckel, and S. Santini. "Household Occupancy Monitoring Using Electricity Meters." In: *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*. 2015.
- [KJ11] J. Z. Kolter and M. J. Johnson. "REDD: A Public Data Set for Energy Disaggregation Research." In: *Proceedings of the Workshop on Data Mining Applications in Sustainability (SustKDD)*. 2011.
- [LLC+03] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong. "Power Signature Analysis." In: *IEEE Power & Energy Magazine* 1.2 (2003).
- [LLN+05] K. D. Lee, S. B. Leeb, L. K. Norford, P. R. Armstrong, J. Holloway, and S. R. Shaw. "Estimation of Variable-Speed-Drive Power Consumption From Harmonic Content." In: *IEEE Transactions on Energy Conversion* 20.3 (2005).
- [LSJ95] S. B. Leeb, S. R. Shaw, and J. L. K. Jr. "Transient Event Detection in Spectral Envelope Estimates for Nonintrusive Load Monitoring." In: *IEEE Transactions on Power Delivery* 10.3 (1995).
- [LM07] J. Li and P. Mohapatra. "Analytical Modeling and Mitigation Techniques for the Energy Hole Problem in Sensor Networks." In: *Pervasive and Mobile Computing* 3.3 (2007).
- [LNK+10] J. Liang, S. K. K. Ng, G. Kendall, and J. W. M. Cheng. "Load Signature Study — Part I: Basic Concept, Structure, and Methodology." In: *IEEE Transactions on Power Delivery* 25.2 (2010).
- [LES+14] J. Liao, G. Elafoudi, L. Stankovic, and V. Stankovic. "Non-intrusive Appliance Load Monitoring using Low-resolution Smart Meter Data." In: *Proceedings of the 5th IEEE International Conference on Smart Grid Communications (SmartGridComm)*. 2014.
- [LFO+07] J. Lifton, M. Feldmeier, Y. Ono, C. Lewis, and J. Paradiso. "A Platform for Ubiquitous Sensor Deployment in Occupational and Domestic Environments." In: *Proceedings of the 6th International Symposium on Information Processing in Sensor Networks (IPSN)*. 2007.
- [LKL+03] J. Lin, E. Keogh, S. Lonardi, and B. Chiu. "A Symbolic Representation of Time Series, with Implications for Streaming Algorithms." In: *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD)*. 2003.

- [MCB+15] H. Maaß, H. K. Cakmak, F. Bach, R. Mikut, A. Harrabi, W. Süß, W. Jakob, K.-U. Stucky, U. G. Kühnapfel, and V. Hagenmeyer. "Data Processing of High-Rate Low-Voltage Distribution Grid Recordings for Smart Grid Monitoring and Analysis." In: *EURASIP Journal on Advances in Signal Processing* 1 (2015).
- [MPB+13] S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. V. Bajic. "AMPds: A Public Dataset for Load Disaggregation and Eco-Feedback Research." In: *Proceedings of the Electrical Power and Energy Conference (EPEC)*. 2013.
- [MHH+10] A. Marchiori, D. Hakkarinen, Q. Han, and L. Earle. "Circuit-Level Load Monitoring for Household Energy Management." In: *IEEE Pervasive Computing* 10.1 (2010).
- [Mar97] G. Marubayashi. "Noise Measurements of the Residential Power Line." In: *Proceedings of the 1st International Symposium on Power Line Communications and Its Applications (ISPLC)*. 1997.
- [MG11] P. Mell and T. Grance. *The NIST Definition of Cloud Computing*. Tech. rep. 800-145. National Institute of Standards and Technology (NIST), 2011.
- [MEE+14] A. Monacchi, D. Egarter, W. Elmenreich, S. D'Alessandro, and A. M. Tonello. "GREEND: An Energy Consumption Dataset of Households in Italy and Austria." In: *Proceedings of the 5th IEEE International Conference on Smart Grid Communications (SmartGridComm)*. 2014.
- [MLS+15] D. Murray, J. Liao, L. Stankovic, V. Stankovic, R. Hauxwell-Baldwin, C. Wilson, M. Coleman, T. Kane, and S. Firth. "A Data Management Platform for Personalised Real-time Energy Feedback." In: *Proceedings of the 8th International Conference on Energy Efficiency in Domestic Appliances and Lighting (EEDAL)*. 2015.
- [ORL+10] U. A. Orji, Z. Remscrim, C. Laughman, S. B. Leeb, W. Wichakool, C. Schantz, R. Cox, J. Paris, J. L. Kirtley Jr., and L. K. Norford. "Fault Detection and Diagnostics for Non-intrusive Monitoring Using Motor Harmonics." In: *Proceedings of the 25th Annual IEEE Applied Power Electronics Conference and Exposition (APEC)*. 2010.
- [PSS16] N. Panwar, S. Sharma, and A. K. Singh. "A Survey on 5G." In: *Physical Communication* 18 (2016).

- [PGW+11] O. Parson, S. Ghosh, M. Weal, and A. Rogers. "Using Hidden Markov Models for Iterative Non-intrusive Appliance Monitoring." In: *Proceedings of the Workshop on Machine Learning for Sustainability Workshop (MLSUST)*. 2011.
- [PGW+14] O. Parson, S. Ghosh, M. Weal, and A. Rogers. "An Unsupervised Training Method for Non-intrusive Appliance Load Monitoring." In: *Artificial Intelligence* 217 (2014).
- [PRK+07] S. Patel, T. Robertson, J. Kientz, M. Reynolds, and G. Abowd. "At the Flick of a Switch: Detecting and Classifying Unique Electrical Events on the Residential Power Line." In: *Proceedings of the 9th International Conference on Ubiquitous Computing (UbiComp)*. 2007.
- [PTM+13] D. Phillips, R. Tan, M. Moazzami, G. Xing, J. Chen, and D. Yau. "Supero: A Sensor System for Unsupervised Residential Power Usage Monitoring." In: *Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 2013.
- [PNMR+16] T. Picon, M. Nait Meziane, P. Ravier, G. Lamarque, C. Novello, J.-C. Le Bunetel, and Y. Raingeaud. "COOLL: Controlled On/Off Loads Library, a Public Dataset of High-Sampled Electrical Signals for Appliance Identification." In: *arXiv Computing Research Repository (CoRR)* abs/1611.05803 (2016).
- [Pru02] A. Prudenzi. "A Neuron Nets Based Procedure for Identifying Domestic Appliances Pattern-of-Use from Energy Recordings at Meter Panel." In: *Proceedings of the IEEE Power Engineering Society Winter Meeting*. Vol. 2. 2002.
- [QG14] F. L. Quilumba-Gudino. "Using Advanced Metering Infrastructure Data for Smart Grid Development." PhD thesis. The University of Texas at Arlington, 2014.
- [Rei17a] A. Reinhardt. "Adaptive Load Signature Coding for Electrical Appliance Monitoring over Low-Bandwidth Communication Channels." In: *Proceedings of the 5th IFIP Conference on Sustainable Internet and ICT for Sustainability (SustainIT)*. 2017.
- [Rei17b] A. Reinhardt. "TUCap: A Sensing System to Capture and Process Appliance Power Consumption in Smart Spaces." In: *Proceedings of the 16th GI/ITG KuVS Fachgespräch "Sensornetze" (FGSN)*. 2017.

- [RBB+12] A. Reinhardt, P. Baumann, D. Burgstahler, M. Hollick, H. Chonov, M. Werner, and R. Steinmetz. "On the Accuracy of Appliance Identification Based on Distributed Load Metering Data." In: *Proceedings of the 2nd IFIP Conference on Sustainable Internet and ICT for Sustainability (SustainIT)*. 2012.
- [RB13] A. Reinhardt and D. Burgstahler. "Exploiting Platform Heterogeneity in Wireless Sensor Networks by Shifting Resource-Intensive Tasks to Dedicated Processing Nodes." In: *Proceedings of the 14th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. 2013.
- [RBM+11] A. Reinhardt, D. Burkhardt, P. S. Mogre, M. Zaheer, and R. Steinmetz. "SmartMeter.KOM: A Low-cost Wireless Sensor for Distributed Power Metering." In: *Proceedings of the 6th IEEE International Workshop on Practical Issues in Building Sensor Network Applications (SenseApp)*. 2011.
- [RBZ+12] A. Reinhardt, D. Burkhardt, M. Zaheer, and R. Steinmetz. "Electric Appliance Classification Based on Distributed High Resolution Current Sensing." In: *Proceedings of the 7th IEEE International Workshop on Practical Issues in Building Sensor Network Applications (SenseApp)*. 2012.
- [RCK14] A. Reinhardt, D. Christin, and S. S. Kanhere. "Can Smart Plugs Predict Electric Power Consumption? A Case Study." In: *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous)*. 2014.
- [REC15] A. Reinhardt, F. Englert, and D. Christin. "Averting the Privacy Risks of Smart Metering by Local Data Preprocessing." In: *Pervasive and Mobile Computing* 16 (2015).
- [RK14] A. Reinhardt and S. Koessler. "PowerSAX: Fast Motif Matching in Distributed Power Meter Data Using Symbolic Representations." In: *Proceedings of the 9th IEEE International Workshop on Practical Issues in Building Sensor Network Applications (SenseApp)*. 2014.
- [RKE+15] A. Reinhardt, G. Konstantinou, D. Egarter, and D. Christin. "Worried About Privacy? Let Your PV Converter Cover Your Electricity Consumption Fingerprints." In: *Proceedings of the 6th IEEE SmartGridComm (SGC) Symposium on Cyber Security and Privacy*. 2015.
- [RR16] A. Reinhardt and D. Reinhardt. "Detecting Anomalous Electrical Appliance Behavior based on Motif Transition Likelihood Matrices." In: *Proceedings of the IEEE International Conference on Smart Grid Communications (SmartGridComm), Sym-*

- posium on Data Management and Grid Analytics and Dynamic Pricing*. 2016.
- [RR15] A. Reinhardt and C. Renner. “RoCoCo: Receiver-initiated Opportunistic Data Collection and Command Multicasting for WSNs.” In: *Proceedings of the 12th European Conference on Wireless Sensor Networks (EWSN)*. 2015.
- [RPL+10] Z. Remscrim, J. Paris, S. Leeb, S. Shaw, S. Neuman, C. Schantz, S. Muller, and S. Page. “FPGA-based Spectral Envelope Preprocessor for Power Monitoring and Control.” In: *Proceedings of the 25th Annual IEEE Applied Power Electronics Conference and Exposition (APEC)*. 2010.
- [RTIo8] I. Richardson, M. Thomson, and D. Infield. “A High-Resolution Domestic Building Occupancy Model for Energy Demand Simulations.” In: *Energy and Buildings* 40.8 (2008).
- [RGH13] A. Ridi, C. Gisler, and J. Hennebert. “ACS-F2 – A New Database of Appliance Consumption Signatures.” In: *Proceedings of the 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*. 2013.
- [RRR+12] M. Ringwelski, C. Renner, A. Reinhardt, A. Weigel, and V. Turau. “The Hitchhiker’s Guide to Choosing the Compression Algorithm for Your Smart Meter Data.” In: *Proceedings of the 2nd IEEE ENERGYCON Conference and Exhibition / ICT for Energy Symposium (ENERGYCON)*. 2012.
- [RAD78] R. L. Rivest, L. Adleman, and M. L. Dertouzos. “On Data Banks and Privacy Homomorphisms.” In: *Foundations of Secure Computation* 4.11 (1978).
- [RNS+10] A. Ruzzelli, C. Nicolas, A. Schoofs, and G. O’Hare. “Real-Time Recognition and Profiling of Appliances through a Single Electricity Sensor.” In: *Proceedings of the 7th Annual IEEE Communications Society Conference on Sensor Mesh and Ad Hoc Communications and Networks (SECON)*. 2010.
- [SGD+10] A. Schoofs, A. Guerrieri, D. T. Delaney, G. M. P. O’Hare, and A. G. Ruzzelli. “ANNOT: Automated Electricity Data Annotation Using Wireless Sensor Networks.” In: *Proceedings of the 7th Annual IEEE Communications Society Conference on Sensor Mesh and Ad Hoc Communications and Networks (SECON)*. 2010.
- [SWH+15] M. Schreiber, M. E. Wainstein, P. Hochloff, and R. Dargaville. “Flexible Electricity Tariffs: Power and Energy Price Signals Designed for a Smarter Grid.” In: *Energy* 93 (2015).

- [SLN+08] S. Shaw, S. Leeb, L. Norford, and R. Cox. “Nonintrusive Load Monitoring and Diagnostics in Power Systems.” In: *IEEE Transactions on Instrumentation and Measurement* 57.7 (2008).
- [Sm012] R. Smolenski. “EMI Measuring Procedures in Smart Grids.” In: *Conducted Electromagnetic Interference (EMI) in Smart Grids*. Springer London, 2012.
- [TSB+14] M. G. Tauber, F. Skopik, T. Bleier, and D. Hutchison. “A Self-organising Approach for Smart Meter Communication Systems.” In: *Proceedings of the 7th IFIP TC 6 International Workshop on Self-Organizing Systems (IWSOS)*. 2014.
- [TGM10] Z. C. Taysi, M. A. Guvensan, and T. Melodia. “TinyEARS: Spying on House Appliances with Audio Sensor Nodes.” In: *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building (BuildSys)*. 2010.
- [URT12] S. Unterschütz, C. Renner, and V. Turau. “Opportunistic, Receiver-Initiated Data-Collection Protocol.” In: *Proceedings of the 9th European Conference on Wireless Sensor Networks (EWSN)*. 2012.
- [URP15] A. S. N. Uttama Nambi, A. Reyes Lua, and V. R. Prasad. “LocED: Location-aware Energy Disaggregation Framework.” In: *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments (BuildSys)*. 2015.
- [VDE16] VDE Verband der Elektrotechnik Elektronik Informationstechnik e.V.; Forum Netztechnik/Netzbetrieb im VDE (FNN). *Störungs- und Verfügbarkeitsstatistik, Berichtsjahr 2015*. 2016.
- [WL13] W. Wang and Z. Lu. “Cyber Security in the Smart Grid: Survey and Challenges.” In: *Elsevier Computer Networks* 57.5 (2013).
- [WHM+12] M. Weiss, A. Helfenstein, F. Mattern, and T. Staake. “Leveraging Smart Meter Data to Recognize Home Appliances.” In: *Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 2012.
- [WBD+11] T. Weng, B. Balaji, S. Dutta, R. Gupta, and Y. Agarwal. “Managing Plug-loads for Demand Response Within Buildings.” In: *Proceedings of the 3rd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (BuildSys)*. 2011.
- [ZR11] M. Zeifman and K. Roth. “Nonintrusive Appliance Load Monitoring: Review and Outlook.” In: *IEEE Transactions on Consumer Electronics* 57.1 (2011).

- [ZJW+14] J. Zhao, T. Jung, Y. Wang, and X. Li. "Achieving Differential Privacy of Data Disclosure in the Smart Grid." In: *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*. 2014.
- [ZGI+12] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar. "Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey." In: *MDPI Sensors* 12 (2012).

LIST OF ACRONYMS

5G	fifth generation of mobile communication systems
AAL	Ambient Assisted Living
AC	Alternating Current
ADC	Analog-to-Digital Converter
AMI	Advanced Metering Infrastructure
CSV	Comma-Separated Values
DC	Direct Current
DSO	Distribution System Operator
EMI	Electromagnetic Interference
FFT	Fast Fourier Transform
ICT	Information and Communication Technology
IoT	Internet of Things
MPPT	Maximum Power Point Tracking
NILM	Non-Intrusive Load Monitoring
PMU	Phasor Measurement Unit
PV	photovoltaic
RMSE	Root-Mean-Square Error
SPI	Serial Peripheral Interface
TSO	Transmission System Operator
WSN	Wireless Sensor Network

PUBLICATIONS CONTRIBUTING TO THIS HABILITATION THESIS

The following scientific contributions have been presented in this manuscript.

A. Reinhardt, P. Baumann, D. Burgstahler, M. Hollick, H. Chonov, M. Werner, and R. Steinmetz. “On the Accuracy of Appliance Identification Based on Distributed Load Metering Data.” In: *Proceedings of the 2nd IFIP Conference on Sustainable Internet and ICT for Sustainability (SustainIT)*. 2012, pp. 1–9, cited on page 26 of this thesis.

Contribution: A load signature data set is presented in this paper, which has been collected as part of Hristo Chonov’s B.Sc. thesis, which I supervised. The data set has undergone post-processing by Burgstahler and Werner. As its second contribution, the paper presents a machine learning-based data processing solution to identify appliances based on their load signatures. This algorithm has been implemented by myself, based on student contributions. I was supported by Baumann in the evaluations. Hollick and Steinmetz contributed to the writing. Overall, I contributed more than half of this paper.

A. Reinhardt, D. Burkhardt, M. Zaheer, and R. Steinmetz. “Electric Appliance Classification Based on Distributed High Resolution Current Sensing.” In: *Proceedings of the 7th IEEE International Workshop on Practical Issues in Building Sensor Network Applications (SenseApp)*. 2012, pp. 1003–1009, cited on page 28 of this thesis.

Contribution: This paper summarizes the outcomes of Dominic Burkhardt’s B.Sc. thesis, which he wrote under my supervision. All conceptual considerations were jointly developed in weekly meetings and implemented by Burkhardt. Co-author Zaheer has been responsible for the practical evaluations. Steinmetz made textual contributions to the manuscript. Burkhardt and I equally shared the workload and put the same efforts into this paper.

A. Reinhardt and D. Burgstahler. “Exploiting Platform Heterogeneity in Wireless Sensor Networks by Shifting Resource-Intensive Tasks to Dedicated Processing Nodes.” In: *Proceedings of the 14th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. 2013, pp. 1–9, cited on page 22 of this thesis.

Contribution: This paper presents the results of Daniel Burgstahler’s M.Sc. thesis, which he wrote under my supervision. I proposed the paper’s fundamental concept of understanding and exploiting system heterogeneity in

networks of embedded systems, and held weekly meetings to discuss concepts to utilize it in practice. Burgstahler was responsible for the implementations and evaluations on different hardware platforms and contributed paragraphs to the paper. The workload was equally shared between the two authors.

A. Reinhardt and S. Koessler. “PowerSAX: Fast Motif Matching in Distributed Power Meter Data Using Symbolic Representations.” In: *Proceedings of the 9th IEEE International Workshop on Practical Issues in Building Sensor Network Applications (SenseApp)*. 2014, pp. 531–538, cited on page 18 of this thesis.

Contribution: This publication is an outcome of Sebastian Kößler’s student research project, conducted at UNSW Australia under my supervision. Conceptual considerations for solving the motif matching problem in load signature data were drafted in weekly discussions and subsequently implemented and evaluated by the student. Each co-author contributed about the same amount of work to this publication.

A. Reinhardt, D. Christin, and S. S. Kanhere. “Can Smart Plugs Predict Electric Power Consumption? A Case Study.” In: *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous)*. 2014, pp. 257–266, cited on page 30 of this thesis.

Contribution: This work on assessing the potential of predicting future appliance electricity consumption has been completely conducted by myself. Minor language revisions were proposed by Christin and Kanhere.

A. Reinhardt, F. Englert, and D. Christin. “Averting the Privacy Risks of Smart Metering by Local Data Preprocessing.” In: *Pervasive and Mobile Computing (PMC)* 16 (2015), pp. 171–183, cited on page 38 of this thesis.

Contribution: The impact of load obfuscation on appliance classification accuracy is studied in this work. I am responsible for the scientific approach taken and its implementation. Co-author Englert has supported this publication by conducting all necessary evaluations, co-author Christin has proposed selected initial ideas and contributed paragraphs to the paper. Overall, more than two thirds of this study were based on my contributions.

A. Reinhardt and C. Renner. “RoCoCo: Receiver-initiated Opportunistic Data Collection and Command Multicasting for WSNs.” In: *Proceedings of the 12th European Conference on Wireless Sensor Networks (EWSN)*. 2015, pp. 218–233, cited on page 20 of this thesis.

Contribution: This paper is joint work with Bernd-Christian Renner with whom all stages of the research have been conducted conjointly. We both contributed to an equal extent to the implementation, evaluation, and writing.

A. Reinhardt, G. Konstantinou, D. Egarter, and D. Christin. “Worried About Privacy? Let Your PV Converter Cover Your Electricity Consumption Fingerprints.” In: *Proceedings of the IEEE International Conference on Smart Grid Communications (SmartGridComm) Symposium on Cyber Security and Privacy*. 2015, pp. 25–30, cited on page 40 of this thesis.

Contribution: In this work, we use active components to enhance user privacy in smart grids. While I have taken primary responsibility for the work, the co-authors have contributed at different stages. Egarter has been active during the evolution of the concept, given his own prior work on related topics. Konstantinou has contributed simulation models of photovoltaic converters and evaluated how their intentional deviation from the maximum power point impacts their operation in practice. Co-author Christin has assisted in determining to what extent privacy protection can be improved by means of the proposed approach. In total, I have contributed about half of this paper.

A. Reinhardt and D. Reinhardt. “Detecting Anomalous Electrical Appliance Behavior based on Motif Transition Likelihood Matrices.” In: *Proceedings of the IEEE International Conference on Smart Grid Communications (SmartGridComm) Symposium on Data Management, Grid Analytics and Dynamic Pricing*. 2016, pp. 704–709, cited on page 32 of this thesis.

Contribution: Anomalous appliance consumption behavior is often an indicator for mechanical wear and thus hints at the risk of device failure. This paper thus presents an anomaly detection algorithm based on transition matrices which I have devised, implemented, and evaluated. I am the paper’s principal contributor, and co-author Delphine Reinhardt has primarily supported me in writing the paper.

A. Reinhardt. “Adaptive Load Signature Coding for Electrical Appliance Monitoring over Low-Bandwidth Communication Channels.” In: *Proceedings of the 5th IFIP Conference on Sustainable Internet and ICT for Sustainability (SustainIT)*. 2017, pp. 1–8, cited on page 17 of this thesis.

Contribution: I am the only author of this paper and fully responsible for all contributions of this work.

N. Buneeva and **A. Reinhardt**. “AMBAL: Realistic Load Signature Generation for Load Disaggregation Performance Evaluation.” In: *Proceedings of the IEEE International Conference on Smart Grid Communications (SmartGridComm) Symposium on Big Data Management and Analytics*. 2017, pp. 443–448, cited on page 34 of this thesis.

Contribution: This thesis summarizes the key contributions of Nadezda Buneeva’s M.Sc. thesis which she wrote under my supervision. I have contributed the initial idea for this work and to writing the paper. In turn, the

concept of this work has been conjointly developed during weekly meetings and was fully implemented and evaluated by the student. Each co-author contributed about the same amount of effort to this work.

A. Reinhardt. "TUCap: A Sensing System to Capture and Process Appliance Power Consumption in Smart Spaces." In: *Proceedings of the 16th GI/ITG KuVS Fachgespräch "Sensornetze" (FGSN)*. 2017, pp. 31–34, cited on page 16 of this thesis.

Contribution: I am the only author of this paper and fully responsible for all contributions of this work.