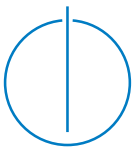
Anwar Ul Haq

Appliance Event Detection for Non-Intrusive Load Monitoring in Complex Environments

Technische
Universität
München



Technische Universität München

Fakultät für Informatik
Lehrstuhl für Wirtschaftsinformatik

Appliance Event Detection for Non-Intrusive Load Monitoring in Complex Environments

Anwar Ul Haq

Vollständiger Abdruck der von der Fakultät für Informatik der Technische Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr. Ing)

genehmigten Dissertation.

Vorsitzende(r): Prof. Dr. Hans Michael Gerndt

Prüfer der Dissertation:

1. Prof. Dr. Hans-Arno Jacobsen
2. Prof. Dr. -Ing. Georg Carle

Die Dissertation wurde am 30.10.2018 bei der Technische Universität München eingereicht und durch die Fakultät für Informatik am 06.12.2018 angenommen.

Abstract

Recently, several load monitoring techniques have been introduced, but non-intrusive load monitoring (NILM) stands out due to single-point (aggregate) energy measurement to detect individual appliance and estimate its power consumption. The accurate appliance detection primarily depends on the density and diversity of appliances under observation. For commercial and industrial environments, there are multiple appliances such as laptops (office), refrigerators (supermarket), and motors (factory) operating in parallel, which usually require high-frequency data acquisition (DAQ) for anomaly detection and simultaneous event reduction. The non-invasive appliance recognition through load disaggregation also supports effective scheduling of appliances for the demand-side management programs.

This work proposes, develops, and evaluates a fine-grained high-frequency DAQ solution to accurately detect individual appliances from the aggregate load, even when several appliances are operating simultaneously. As most disaggregation inaccuracies stem from DAQ, we initiated our investigation by exploring and experimenting with different low-cost, off-the-shelf energy monitors to gather the required aggregate energy data. Similarly, we examined the use of sound card, due to its better quality analog to digital converters and built-in filters to screen-out unwanted low-frequency noise, to measure energy at high sampling frequency. Unfortunately, inadequate low-level signal handling capabilities and lack of multiple inputs for the three-phase system were key deficiencies for effective NILM utilization. To handle NILM DAQ limitations and fulfill other appliance specific disaggregation requirements, we have developed a circuit-level electric appliance radar (CLEAR). CLEAR is a customizable energy monitoring solution capable of providing a cost-effective mechanism to simultaneously gather accurate energy data at up to 250 kHz from three circuits (or six channels).

Similarly, this work also evaluates the appliance-switching (on/off) event detection capability to isolate high-resolution appliance event signatures (start-up current waveform pattern) from the aggregate data acquired by CLEAR. We assessed CLEAR in an office environment with an abundance of switched-mode power supplies (SMPS) to observe the effect of high-resolution on simultaneous event detection. A novel event detection algorithm using the Hilbert-Huang transform is proposed to demonstrate the effectiveness of high-resolution data to accurately detect SMPS-equipped appliance events from the aggregate load. The results indicate that higher sampling frequency readily reduces simultaneous events and helps detect most of the low-powered office appliance switching events from the aggregate

load.

Furthermore, eight different audio, sliding-window, and dictionary-based lossless compression techniques were compared based on metrics such as compression ratio and time to compress/decompress energy data. For high-frequency energy data, the audio-based lossless compression algorithms showed better results, essentially due to the periodic nature of energy data.

Zusammenfassung

In den letzten Jahren wurden verschiedene Techniken zur Lastgangüberwachung vorgestellt. Die nicht-invasive Lastgangmessung (NILM) zeichnet sich durch ihre Einzelpunktmessung (aggregiertes Signal) der elektrischen Energie aus, um einzelne Geräte zu erkennen und deren individuellen Leistungsverbrauch zu ermitteln. Eine fehlerfreie Geräteerkennung hängt stark von der Häufigkeit und Diversität der einzelnen eingesetzten Geräte ab. Kommerzielle und industrielle Umgebungen beinhalten eine Vielzahl an ähnlichen Verbrauchern, wie Laptops (Büros), Kühlschränke (Supermärkte) und Motoren (Fabriken), welche alle gleichzeitig aktiv sein können. Solch komplex überlagerte Umgebungen erfordern hochfrequent erfassende Messsysteme (DAQ) zur Signalerfassung und weiteren Anomalieerkennung um zeitgleiche Ereignisse zu unterscheiden. Die nicht-invasive Geräteerkennung mithilfe der Lastgangdisaggregation ermöglicht Informationen aus dem aggregierten Signal zu gewinnen und damit eine effektivere Planung von Geräten für Demand-Side Management (DSM) Programme.

Diese Arbeit definiert, implementiert, und evaluiert ein feingranulares hochfrequentes DAQ zur präzisen Erkennung von Einzelverbrauchern aus dem Gesamtlastgangs, mit besondern Fokus wenn mehrere Verbraucher gleichzeitig betrieben werden. Ein Großteil von Ungenauigkeiten im Bereich der Disaggregation lässt sich auf die Datenerfassung zurückführen. Daher starten wir mit der experimentellen Untersuchung von verschiedenen preiswerten und handelsüblichen Energiemessgeräten um die notwendigen Gesamtlastgänge zu ermitteln.

Im nächsten Schritt, untersuchen wir die Verwendung von handelsüblichen PC Soundkarten zur hochfrequenten Energiemessung um ungewünschte niederfrequente Störsignale aufgrund der besseren Qualität der Analog-zu-Digital Wandler und eingebauten Filter herauszufiltern. Allerdings sind die Fähigkeiten zur Signalverarbeitung auf unterster Ebene ungenügend und die Anzahl der Signaleingänge für ein 3-Phasen-System wiederum nicht ausreichend für effektives NILM. Um die NILM- und dazu verwandte Anforderungen an ein DAQ System zu erfüllen, haben wir CLEAR entwickelt: ein circuit-level appliance radar (etwa: Schaltkreis-bezogenes Geräteradar). CLEAR ist eine konfigurierbare Energiemesslösung welche eine kosten-effektive gleichzeitige Messung von Energiedaten mit bis zu 250 kHz von 3 Schaltkreisen (6 Kanälen) ermöglicht.

Anhand der mit CLEAR aufgezeichneten Messdaten evaluieren wir eine Geräte An-/Ausschalt-Eventerkennung mithilfe von hoch aufgelösten Eventsignaturen (beinhalten

Wellenform des Stroms und Einschaltspitzen). Wir testen CLEAR in einer Büroumgebung mit einem großen Anteil an Schaltnetzteilen um den Effekt von hoch aufgelösten Messdaten mit einer simultanen Eventerkennung zu evaluieren. Wir stellen einen neuartigen Eventerkennungsalgorithmus vor, basierend auf Hilbert-Huang Transformationen, um die Vorteile von hochfrequenten Messdaten zur Erkennung von Schaltnetzteilen zu demonstrieren. Die Ergebnisse zeigen, dass höhere Abtastraten die Anzahl an gleichzeitig erkannten Events reduzieren und dabei helfen kann, einen Großteil der Kleinverbraucher in Büros anhand des Gesamtlastgangs zu erkennen.

Die verfügbaren NILM Messdaten wurden mit acht verschiedenen Komprimierungsverfahren, basierend auf Audio, Sliding-Window und Wörterbuch, größenreduziert und in weiterer Folge die Kompressionsrate und Rechenzeit ermittelt. Hochfrequente Energiedaten zeigten die besten Ergebnisse mit Audio-basierten verlustfreien Kompressionsverfahren, aufgrund der Periodizität in Energiedaten.

In the loving memory of my beloved father...

Acknowledgement

This dissertation presents the research conducted during my four year stay at the Chair of Application and Middleware Systems (I-13), Department of Informatics, Technical University Munich. This work would not have been possible without the amazing people who supported and advised me throughout my stay in Germany and to whom I feel deeply indebted.

First and foremost, I express my sincere gratitude to my doctoral supervisor Prof. Hans-Arno Jacobsen for allowing me the freedom in various aspects of scientific research. I am thankful for his enduring support, persistent guidance, valuable suggestions, and encouragement throughout my stay at I-13. It has been a privilege to work under his supervision.

I am deeply grateful to the rest of committee members for their support and guidance. I would like to sincerely thank Prof. Georg Carle for his support as the secondary supervisor and Prof. Michael Gerndt for accepting to chair my PhD defence. A special thanks to Ms. Manuela Fischer for her unconditional support during the submission of dissertation.

During my stay at I-13, I had the privilege to work with some really smart people. I wish to acknowledge the support from my colleagues (alphabetically-listed) Amir, Christoph Goebel, Danial, Elias, Jan, Jeeta, Jose, Martin, Mo-Reza, Pezhman, Pooya, Ruben, and Svenja. I would like to especially thank Matthias and Thomas, my collaborators on the NILM project. Special thanks to Victor del Razo and Christoph Doblander for their support throughout my stay. Lastly, I would also like to thank Kaiwen and Tobbias for introducing me to the world of board games.

I also had the opportunity to work and supervise some excellent students. I would especially like to thank Benjamin, Masha, Linus, and Christian for their helpful input to compression and event detection work. Special thanks to Dr. Boehmer for his valuable input to the design and prototyping of CLEAR and MEDAL.

I am deeply indebted to my wonderful family for their endless support and prayers. Thank you for always being there when I needed your support. I owe my deepest gratitude to my late father- my sole inspiration and role model throughout my life. Thank for always believing in me. Its only because of you and mom that I made it this far.

On the non-scientific front, I would like to thank all my good friends Rameez, Safi, Suli, Sardar, Abdul, Irfan, Esra, and Raziye for their unconditional help and support. Special

thanks to Christina for inspiring me to learn German and also encouraging me during difficult times.

Finally, I express my profound gratitude to the Higher Education Commission of Pakistan for the financial support through the Faculty Development Program (UESTP/UETS) in cooperation with German Academic Exchange Service (DAAD), and Alexander von Humboldt foundation for the research and travel grants during my PhD studies.

List of Publications

Publications resulting from work performed during the doctoral research at TUM with first authorship:

- A. U. Haq, T. Kriechbaumer, M. Kahl, and H.-A. Jacobsen. "CLEAR - A circuit level electric appliance radar for the electric cabinet." In: *2017 IEEE International Conference on Industrial Technology (ICIT)*. Toronto, Canada, Mar. 2017, pp. 1130–1135. ISBN: 978-1-5090-5319-3/17. DOI: 10.1109/ICIT.2017.7915521
- A. U. Haq and H.-A. Jacobsen. "Prospects of appliance-level load monitoring in off-the-shelf energy monitors: A technical review." In: *Energies* 11.1 (2018), p. 189. DOI: 10.3390/en11010189
- A. U. Haq, M. Kahl, and H.-A. Jacobsen, "A high-frequency appliance event detector for non-intrusive load monitoring," submitted to *IEEE Transactions on Industrial Informatics*.
- A. U. Haq, B. A. Degenhart, M. B. Heravi, Nikola Dinev, and H.-A. Jacobsen, "Analysis of lossless compression algorithms and selective compressed sensing approach for non-intrusive load monitoring," submitted to *IEEE Access*.

Publications resulting from collaboration work performed during the doctoral research at TUM with co-authorship:

- M. Kahl, C. Goebel, A. U. Haq, T. Kriechbaumer, and H.-A. Jacobsen. "NoFaRe: A non-intrusive facility resource monitoring system." In: *Energy Informatics*. Vol. 9424. Lecture Notes in Computer Science. Springer International Publishing, 2015, pp. 59–68. DOI: 10.1007/978-3-319-25876-8_6
- T. Kriechbaumer, A. U. Haq, M. Kahl, and H.-A. Jacobsen. "MEDAL: A cost-effective high-frequency energy data acquisition system for electrical appliances." In: *Proceedings of the 2017 ACM Eighth International Conference on Future Energy Systems*. e-Energy '17. Hong Kong, Hong Kong: ACM, May 2017. ISBN: 978-1-4503-5036-5/17/05. DOI: 10.1145/3077839.3077844
- M. Kahl, A. U. Haq, T. Kriechbaumer, and H.-A. Jacobsen. "A Comprehensive feature study for appliance recognition on high frequency energy data." In: *Proceedings*

of the Eighth International Conference on Future Energy Systems. ACM. 2017, pp. 121–131. DOI: 10.1145/3077839.3077845

- M. Kahl, T. Kriechbaumer, A. U. Haq, and H.-A. Jacobsen. "Appliance classification across multiple high frequencyeEnergy datasets." In: *2017 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. 2017. DOI: 10.1109/smartgridcomm.2017.8340664

- M. Kahl, V. Krause, R. Hackenberg, et al. "Measurement system and dataset for in-depth appliance energy consumption analysis in industrial environments." In: *tm - technisches messen* (2018). DOI: 10.1515/teme-2018-0038

- M. Kahl, A. U. Haq, T. Kriechbaumer, and H.-A. Jacobsen. "WHITED- A worldwide household and industry transient energy dataset." In: *3rd International Workshop on Non-Intrusive Load Monitoring*. 2016. URL: https://www.i13.in.tum.de/index.php?id=114&L=0

Contents

CHAPTER 1

Introduction

Energy systems around the globe are evolving with a transition from fossil fuels towards clean energy technologies. Through smart grid initiative, we have witnessed a historic transformation in the energy sector over the last decade. Some tremendous technological innovations have kick-started a positive trend to achieve climate change ambitions in accordance with the *Paris Agreement* [9]. Still, more significant efforts are required as the sustainable energy future relies heavily on the interaction between different energy technologies. An integrated approach is required with the collaboration of all the key stakeholders to reduce the overall energy demand.

The CO_2 emission reduction is the most pressing requirement for the climate change initiative. This requirement has already encouraged the reduction in greenhouse gas emissions. According to 2DS paradigm [10], scaling up renewable energy generation by 74% alone can help achieve the net-zero CO_2 emissions from the power sector by 2060. Hence, the renewable energy resources provide a realistic alternative to mitigate the climate change challenges. Unfortunately, due to high dependency on weather, these renewable energy resources cannot be entirely controlled. In order to increase grid reliability, these weather driven resources require flexibility for a consistent generation, transmission, and distribution of power across the grid [11].

To some extent, the grid flexibility required by renewable energy resources can be provided through improved weather prediction, increased energy conversion efficiency through technological advancement, and with the integration of grid storage mechanisms [12]. Similarly, with efficient use of resources, the effective prosumer (producer and consumer)

participation in the demand side management programs can also provide the inherent flexibility. As the renewable energy resources are expected to increase on the demand side, suitable measures are required to encourage active consumer partition in the demand side management programs. According to Lord Kelvin: *"To measure is to know. If you cannot measure, you can not improve it"*. Hence, to ensure effective participation, consumers require in-depth knowledge about appliance-level energy consumption statistics and identification of power hungry appliances.

In this work, we present a comprehensive roadmap to implement non-intrusive load monitoring (NILM) technique, especially for the complex and challenging environments. Although NILM techniques for appliance event detection and load estimation from aggregate load have been around for nearly three decades, they mostly focus on the residential environment. Even to date, there is little insight to address the challenges associated with the commercial and industrial environment. In an industrial environment, for instance, multiple parallel processes and sub-events are concurrently occurring from different motors and other sub-machine components. The detection of each sub-event such as heating, lighting, and controller event is critical for proper identification and maintenance of appliances. Hence, it is of utmost importance to avoid or minimize these simultaneous events. Other factors such as appliance type and the number of appliances also impact the event detection capability during multiple simultaneous events.

1.1 Motivation

The smart grid initiative has led to tremendous innovations to shape the future of energy usage. Smart meters, equipped with bi-directional communication and power transfer capabilities, are deployed to measure the overall energy consumed in a household [13]. Traditionally, utility companies use energy meters to measure the aggregate energy for a household or building. The indicators such as weekly, monthly, and yearly energy consumption are primarily used for billing purposes. Unfortunately, such consumption indicators are not appropriate to infer any useful information regarding individual appliance consumption patterns.

For the future electric grid, consumers play an important role in conserving electricity usage to help decrease the overall energy consumption. For effective participation in demand-side management (DSM) programs, consumers need to be aware of the overall and

appliance-specific consumption details, ideally in real time. Similarly, the incorporation of weather-driven renewable resources on the distribution grid also requires flexibility to optimize the energy utilization through DSM. Considerable savings can be achieved by sharing fine-grained real-time energy consumption at the appliance level with consumers. According to a recent study [14], around 4.5% savings can be achieved through proper consumption feedback.

To further encourage consumer participation, the smart homes and buildings are expected to embed the necessary intelligence to automatically manage the necessary operations within a home or building. These operations range from the incorporation of intermittent renewable generation and storage systems to the intelligent building management systems. To effectively perform these basic energy management operations, it is essential to measure the energy consumed by the individual appliance. With appliance-level consumption information, the consumers can distinguish between power-hungry appliances and perform demand response (DR) for local load management.

For accurate measurement of overall energy consumption, the energy monitors can be broadly categorized into two categories; multi-point and single-point energy monitors. The multi-point energy monitors require the installation of some dedicated sensors for individual appliances. Multi-point metering can therefore be termed as intrusive load monitoring (ILM), as each appliance to be measured must be attached to a dedicated sensing unit (a smart plug or a smart power strip). This technique is useful and accurate for acquiring appliance-specific energy information. On the contrary, the maintenance of these sparsely distributed energy sensors can be labor intensive, even if the sensors are cost-effective (mass produced) and long-lasting.

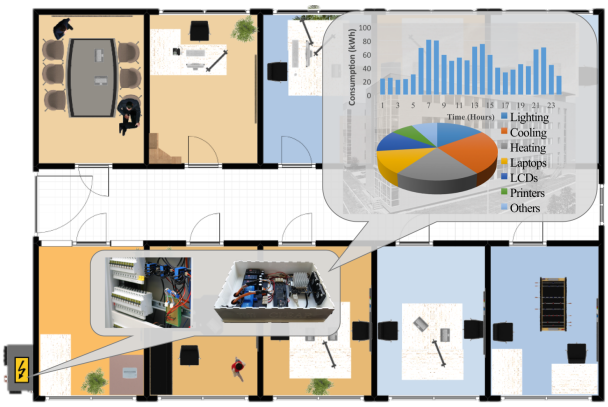


Figure 1.1.1: Non-intrusive load monitoring in office environment

Similarly, the single-point energy monitors take measurement at a single location, usually at electricity mains, to acquire aggregate energy consumption. The traditional utility energy meters fall into this category as they obtain energy data at a fixed location (main circuit board) and provide the aggregate weekly, monthly, or yearly energy consumption. The recent advances in informatics, especially in machine learning, have fostered the growth of the non-intrusive load monitoring technique [15, 16, 17]. NILM, as shown in Fig. 1.1.1, breaks down the aggregate energy consumption to essentially differentiate between individual loads.

1.2 Problem Statement

With this research, we focus on the design and evaluation of NILM hardware to suggest a better way of acquiring energy dataset, especially for the complex and challenging environments. The goal of this study is to improve the data acquisition (DAQ) and event detection capability of NILM-based energy monitoring systems. The proper DAQ and event detection for NILM requires simultaneous acquisition of high-frequency voltage and current measurements from multiple circuits to satisfy disaggregation requirements. This work concentrates on following research objectives/challenges.

Challenge 1: How to design a suitable and cost-effective NILM hardware capable of performing accurate energy disaggregation under multiple operating environments such as residential, commercial, and industrial?

Challenge 2: Is it possible to detect and capture appliance switching (on/off) events for complex environments such as offices and factories with low appliance diversity and high appliance density? If adequately detected, can we accurately estimate the instantaneous energy of the appliance event?

Challenge 3: How does sampling frequency effect appliance event detection accuracy? Does the high-frequency DAQ help minimize simultaneous events?

Challenge 4: How to take advantage of high-frequency DAQ while minimizing potential disadvantages related to data quality and storage?

1.3 Approach

The overall objectives of this work are fourfold. First, some off-the-shelf energy monitors are compared using a technical survey. The idea is to evaluate different energy monitors available on the market and check their suitability with NILM. Second, a state-of-the-art DAQ hardware capable of acquiring high-resolution energy data from challenging environments is presented. Third, a novel event detection algorithm based on Hilbert-Huang transform is presented to evaluate the acquired data for switching event detection. Fourth, we compare eight different audio, sliding-window, and dictionary-based lossless compression algorithms based on compression ratio and time to compress/decompress energy data. Let us discuss these approaches in detail.

1.3.1 Technical Evaluation of Off-The-Shelf Energy Monitors

The smart grid initiative has encouraged utility companies worldwide to roll-out new and smarter versions of energy meters. Before an extensive roll-out, which is both labor-intensive and incurs high capital costs, consumers need to be incentivized to reap the long-term benefits of such smart meters. Off-the-shelf energy monitors (e-monitors) can provide consumers with insight into such potential benefits. As the consumer owns e-monitors, the consumer has greater control over the data, which significantly reduces the privacy and data confidentiality concerns. Because only limited online technical information is available about e-monitors, we evaluate several existing e-monitors using an online technical survey directly from the vendors.

Besides automated e-monitoring, the use of different off-the-shelf e-monitors can also help us demonstrate the state-of-the-art techniques such as non-intrusive load monitoring, data analytics, and predictive maintenance of appliances. Our survey indicates a trend towards the incorporation of such state-of-the-art capabilities, particularly the appliance-level e-monitoring and load disaggregation. We have also discussed some essential requirements to implement load disaggregation in the next generation e-monitors. In future, these intelligent e-monitoring techniques will encourage effective consumer participation in the

DSM programs.

1.3.2 Design and Evaluation of NILM DAQ Hardware

Depending on the restrictions and application requirements, multiple NILM approaches have been proposed in the literature [15]. Unfortunately, none of these approaches have considered the requirements and challenges of detecting multiple appliance-switching events, a typical scenario in an industrial environment where many motors are operating in parallel. Similarly, in supermarkets and offices, a large number of similar appliances such as refrigerators and low-powered appliances are working together. Due to their unpredictable switching (on/off), it is almost impossible to avoid some simultaneous events. High-frequency DAQ can extract more information regarding these appliances and help reduce (or even prevent) the simultaneous events. Besides improving appliance detection accuracy, high-frequency DAQ can also assist in predictive maintenance (suggest appliance maintenance before it fails) and anomaly detection (detect unknown or malfunctioning appliances) inside a building.

1.3.3 NILM Event Detection in Complex Environments

For accurate NILM disaggregation, precise event detection is of paramount importance. The accurate appliance switching (on/off) event detection mainly depends on factors such as data resolution, appliance density, and diversity. In general, the appliance detection challenge using load disaggregation techniques increases as the percentage of power consumption decreases.

In a study by Wong et al. [18], the authors have introduced a categorization of NILM approaches into event-based methods and non-event-based methods. Event-based methods first determine the time instants when the aggregate power consumption changes drastically. Later, the responsible appliances are identified, and the state of each appliance is monitored to calculate the event duration. On the other hand, non-event-based methods completely refrain from detecting events and instead try to detect the state of appliances from instantaneous properties of the current and voltage signals. The accurate event detection of low-powered SMPS-equipped office appliances can be performed using event-based approach. We present a Hilbert-Huang transform based event detection approach to

detect micro-events from an office environment with an abundance of SMPS-equipped appliances. The challenge lies in accurately determining the switching events caused by these low-powered supplies. For this purpose, we perform our proposed event detection algorithm on high-frequency aggregate energy data from BLOND [19].

1.3.4 Analysis of Lossless Compression Algorithms on Energy Data

The high-frequency data acquisition significantly improves the event detection accuracy and recognition of appliances [20]. It also helps to minimize simultaneous appliance switching events [21]. The collection of high-frequency energy data also facilitates fault-detection and predictive maintenance of appliances through smart monitoring [22]. In other words, evaluating electro-mechanical systems through analysis of electrical power data could lead to the detection of malfunctioning appliances in a network. Unfortunately, high-frequency sampling also elevates the storage requirements making error-free data handling both process intensive and challenging. For example, high-frequency DAQ for a three-phase system requires 56 GB/day storage when sampled at 50 kHz. The storage requirement is elevated to 281 GB/day when sampling at 250 kHz [1], which makes storage as one of the main drawbacks when considering high-frequency DAQ.

One way to reduce the system load and network traffic is to apply compression algorithms to reduce data size and hence, the storage requirement. The primary motivation of this work is to explore different lossless compression algorithms developed over the years. The energy data typically consists of periodic voltage and current waveforms centered at a fundamental frequency of 50 Hz (60 Hz for the US). The audio-based lossless compression algorithms can be readily applied due to very nature of these periodic AC waveforms. Similarly, as energy monitoring is a continuous process, large volumes of data are expected to be accumulated over time, especially at higher sampling rates. The sliding-window and dictionary-based lossless compression techniques are also explored in this study to find a suitable match.

1.3.5 Known Use Cases

Anomaly Detection

Often appliance health monitoring is critical to detect faults and suggest maintenance before complete breakout [23], especially when the appliance is not easily accessible. Most faults in electric machines develop gradually and take some time to grow and completely break the machine. Motor current signature analysis (MCSA) method is widely utilized for fault detection in induction motors [24]. A similar analysis can be performed by the proposed hardware to closely monitor the AC waveform of the appliance in high-frequency. Later, machine learning techniques can be used to detect any unusual changes in the AC waveform and provide online appliance diagnosis.

Wide Area NILM

Since NILM can be easily utilized in a house or building, it makes sense to utilize such monitoring on a broader level such as neighborhood and district-level. By using NILM on district-level, we can overcome the privacy concerns of the consumers as the only information available to the Utilities will include the number and type of appliances on a higher district-level without giving away the consumer identity. This will help add the flexibility in the distribution grid through better renewable energy prediction and will also encourage effective consumer participation.

Energy Leakage Detection

The proposed hardware and event detection scheme can also be utilized to better manage the operating appliances in a building. It can identify the consumption patterns of appliances, which can be used to check the efficiency of the operating appliances. It can also help identify the building managers detect unauthorized appliance use and even unintentional energy leakage (caused by appliances accidentally left operating). Similarly, the proposed hardware is capable enough to identify the speed of motor [25].

1.4 Contributions

Some contributions for the technical survey include:

- Through our online survey, we directly contacted 54 different companies to obtain technical information regarding 79 different e-monitors.
- We received useful information regarding the architecture and operation of 27 e-monitors, a response ratio of 34.1%. Some technical information was not publicly available.
- We also explored the available online literature from 9 companies for 14 different products. These companies did not participate in the survey but provided enough technical information online as technical-notes.
- We provide an in-depth analysis of NILM and highlight key requirements for NILM-enabled DAQ systems. This knowledge helped us to understand how the state-of-the-art e-monitors can be upgraded to perform load disaggregation.
- Our survey indicates a trend towards the incorporation of such state-of-the-art capabilities, particularly the appliance-level e-monitoring and load disaggregation.

Some challenges associated with high-frequency DAQ include increased complexity to handle simultaneous data streams from all three phases. To overcome NILM DAQ challenges, we propose a state-of-the-art DAQ system with high temporal and amplitude accuracy. The main contributions in design and evaluation of proposed DAQ hardware include:

- A reproducible and purpose-built hardware design for the NILM application is presented, which satisfies the general safety requirements
- Due to its customizable features, the proposed DAQ system is operable under various operating environments (residential, commercial, and industrial)
- The proposed hardware is capable of acquiring and handling six simultaneous high-frequency sample streams (up to 250 kHz) from electricity mains, which improves the probability of anomaly detection while reducing simultaneous events

- As a test case, the proposed hardware is installed in an office with an abundance of low-powered switched mode power supplies. In such an environment, the challenge is to minimize event detection errors caused by simultaneous switching appliances. The successful event detection in an office environment using high-frequency DAQ will enable us to detect events in the industrial environment, where many motors and other machines are operating simultaneously

Event detection from aggregate load is a daunting task, especially in complex industrial and commercial environments, where many similar appliances are operating simultaneously. The main contributions of our NILM-based event detection approach include:

- We present a novel technique to detect low power appliance switching events from an office environment with high appliance density (upto 50 appliances) and low appliance diversity (mostly SMPS-equipped appliances)
- The approach is based on Hilbert-Huang transform (HHT) which provides time-frequency-energy analysis on the acquired/recorded AC waveform. The HHT method is also useful to estimate the instantaneous amplitude (power) of the AC waveform
- We perform empirical evaluation on BLOND dataset and also provide runtime considerations
- Similarly, we present the effect of sampling frequency on the simultaneous events detection

The last challenge is related to the massive amount of data that resulted from 250 kHz sampling rate from six channels. The main contributions for the comparison of lossless compression algorithms for energy data include:

- We present a comparison of state-of-the-art lossless compression algorithms on high-frequency energy data
- We inspect eight different audio, sliding-window, and dictionary-based compression algorithms
- Using different lossless audio compression techniques, we try to find the correlation between compression ratio and sampling frequency.

- Similarly, we present the compression efficiency of periodic energy data, especially with the music based algorithms. In general, a smooth and stable periodic waveform (voltage) compresses well as compared to non-smooth and unstable periodic waveform (current)

1.5 Organization

The rest of dissertation is organized as follows. Chapter 2 presents the background knowledge necessary to understand the basics of energy monitoring, especially the non-intrusive load monitoring technique. Chapter 3 explores the related work regarding the types of NILM enabled DAQ hardware followed by different event detection techniques used in NILM literature with a focus on high-frequency energy DAQ. We also discuss different compression techniques, especially for energy data.

Chapter 4 partly deals with challenge 1. In this chapter, we discuss the results of the online technical survey conducted to compare different off-the-shelf energy monitors available on the market. The purpose of the online survey was to obtain detailed technical information regarding e-monitors, as limited information is publicly available. A total of 54 different companies were shortlisted and invited to participate in the survey.

Chapter 5 also addresses challenge 1. Here, we propose hardware design for a circuit level electric appliance radar (CLEAR), and briefly discuss the hardware and software components. We also discuss the approaches we tried before the development of customized CLEAR hardware architecture. Similarly, we explain the process of how the data is acquired, processed, and forwarded for persistence. Chapter 6 evaluates the CLEAR hardware based on the key requirements we defined for state-of-the-art NILM DAQ system.

Chapter 7 briefly addresses challenge 2 and 3. This chapter elaborates different event detection techniques utilized in NILM. Later, we propose a novel event detection technique based on Hilbert-Huang transform to detect switching events of SMPS-equipped appliances. Apart from processing complexity, the initial tests are encouraging as we are able to segregate most of the SMPS events from aggregate load. We believe that such a technique can also be applied to a higher district-level to support demand side management programs for the smart grid.

Chapter 8 addresses challenge 4. Here, we compare eight different lossless compression

techniques for aggregate CLEAR data. The comparison is based on compression ratio and time to compress/decompress. One interesting result is the relatively better performance of music-based compression techniques. Chapter 9 presents the conclusions and future work covering different aspects of design and evaluation of DAQ hardware for NILM and appliance event detection in challenging and complex environments.

CHAPTER 2

Background

The concept of electricity monitoring emerged immediately after the inception of electricity generation and distribution systems during the late 19th century. The first commercial use of electricity was direct current (DC), and electrochemical meters were introduced initially to measure electricity consumption [26]. These meters were labor-intensive as they required the periodic removal and weighing of plates from an electrolytic cell. Electrochemical meters were then replaced by electromechanical meters, also known as induction meters or Ferraris meters [27]. The early electromechanical meters measured charge in ampere-hours and calculated energy consumed during the billing period.

In the beginning, electricity was primarily utilized by lighting systems and to a lesser extent for operating electric loads such as electric motors. As more industries shifted from oil and gas to electricity, there was an enormous increase in energy demand and hence the need to measure electricity use accurately. Modern buildings, both residential and commercial, constitute a major portion of the electricity demand. It is estimated that around 73% of electricity in the United States is consumed by buildings [28]. From 1999 to 2004, the consumption of electricity in the residential sector of European Union (EU) alone has increased by 10.8% [29]. In Europe, the energy consumption in buildings accounts for 41% of the primary energy consumption, for which a major chunk of this primary energy (85%) is utilized to achieve a comfortable room temperature (mostly through oil and gas heating), and the remaining 15% is consumed as electrical energy [30].

In this chapter, we explore the different energy monitors (e-monitors) currently available to consumers. The main goal of our work is to help researchers, building managers, and

consumers choose the e-monitor best suited for their specific applications. Although some of these e-monitors are appropriate to manage renewables and can provide added features, such as load disaggregation for appliance-level monitoring, they are often overlooked as a result of a lack of available technical data about their capabilities. Similarly, as compared to a smart meter, which is owned by the utility company, the e-monitor is bought and managed by the consumer. The e-monitor allows consumers to have more control over data. The consumer can even share non-private data collected by the e-monitor with the utility to facilitate in load-forecasting.

2.1 Energy Monitoring Overview

Before going into detail, it is necessary to differentiate between a smart meter and an e-monitor. A smart meter is the next-generation meter capable of linking a building with the utility company to enable two-way communication and power exchange between them [31]. A smart meter also assists in remote billing and instant load feedback to the utility for load forecasting. As it is owned by the utility, the smart meter comes with inherent drawbacks related to data confidentiality and privacy [32, 33, 34]. On the contrary, an e-monitor is owned by the consumer and works independently alongside existing energy meters, without any direct effect on the billing. E-monitors are preferred because of their ability to observe energy consumption patterns in real-time through a user-friendly visual interface, and they are helpful in making informed energy-conservation decisions. They can be easily installed by clipping their current sensors around a current-carrying wire or directly inserting them into a power plug. As a result of local and private cloud storage, the e-monitor can minimize privacy concerns and added features, such as disaggregation and the efficient integration of renewables, and can encourage consumers and building managers to participate in DSM effectively.

For a fair comparison, it is important to view how different vendors and platforms measure and calculate energy consumption. For load monitoring, there are two main categories of e-monitors available on the market: single- and multi-point e-monitors. The single-point e-monitors capture the aggregate energy consumption of the whole house, building or industrial facility. The multi-point e-monitors constantly capture measurement data at several locations and are preferred for detailed load monitoring, such as monitoring the power usage of individual appliances. Monitoring at the appliance-level can result in more

engaged consumer participation as consumers can better identify power-hungry appliances and accordingly manage their peak load. For a rational comparison of e-monitors, we have outlined six dimensions, the types of parameters, the sampling frequency, the accuracy, the resolution, the application area, and the cost of monitoring equipment on which we base our comparisons.

2.1.1 Parameter Type

Except for voltage and current, most of the parameters (if utilized) are calculated using standard mathematical formulations. These parameters are derived internally, and for the most part, a subset of these parameters is utilized and displayed to consumers. Some basic parameters are described below.

Voltage and Current Waveform

Voltage waveform measurement assists in making corrective measures against harmful low and high voltage levels. Usually, the voltage transformers (AC–AC adaptors) are used to measure the peak and root-mean-square (RMS) voltages of the line. Unlike the voltage waveform, the current waveforms are not stable sine waves; they vary considerably depending upon the type of operating load, as illustrated in Fig. 2.1.1. Each load type (resistive, inductive or capacitive load) has a different influence on the current curve, and often the inrush current features are used for appliance segregation using NILM.

Power and Power Factor

The main feature used by almost every energy-metering device is the real power. This is the true rate at which energy is used and is calculated through the voltage and current measurements [35]. Similarly, the power factor is used to distinguish between resistive, inductive and capacitive appliances. It determines the phase difference caused by the inductive and capacitive components. A positive phase angle indicates a net inductive reactance of the circuit, where the current lags voltage. On the contrary, a negative phase angle indicates a net capacitive reactance of the circuit as the current leads the voltage.

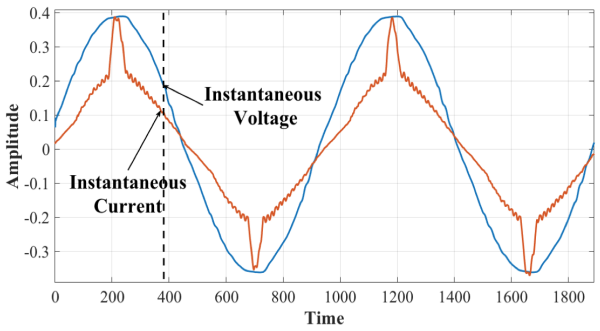


Figure 2.1.1: Instantaneous voltage and current waveforms.

Harmonics

Harmonics or higher-frequency components occur as a result of pulsating devices (such as frequency drive, electric welders, etc.), resulting in system heating and overvoltage [36]. The harmonics are created by different electronic components present in the appliance circuitry. They produce a new distinct waveform as a result of the superposition of different harmonics. The fast Fourier transform (FFT) resolves the superimposed waves into their constituent waves. In e-monitors, another term commonly associated with harmonics is the total harmonic distortion (THD). This refers to the presence of harmonic distortion caused by the non-linear loads. THD determines the power quality of the system, where a lower THD indicates a reduction in heating, peak currents, and losses [37]. As a result of these distinctive features, the higher-order harmonics are useful for power disaggregation applications.

2.1.2 Resolution

The resolution is determined by the number of bits of the analog to digital converters (ADCs) and defines the number of codes that can be formed digitally using these bits. Because the voltage and current signals are continuous in nature (analog), to calculate the other set of features (e.g., real power, RMS voltage and current, power factor, etc.), analog signals need to be converted to digital signals. The uncertainty in the digital signal is determined by the measurement accuracy in the analog input and is known as the resolution of the signal. The resolution is determined by the number of bits used to represent each variable (bits of ADCs), which defines the quantization levels and hence the uncertainty.

2.1.3 Sampling Frequency

For e-monitoring, the choice of any specific sampling frequency or sampling rate depends upon the amount of information we are interested in obtaining from these signals. The sampling frequency may range from the hourly reading to the high-frequency (MHz) range. In general, to observe the harmonics and transient switching response of the appliances, it is better to utilize a higher sampling frequency. It is also important to mention that the sampling done for analog to digital conversion might not be the same as samples reported for display. Although all e-monitors have a sampling rate sufficient to satisfy the Nyquist criteria and accurately calculate the consumed power, most of the modern e-monitors downsample to lower sampling rates to reduce storage requirements.

2.1.4 Accuracy

The accuracy is determined by the difference between the measured and the true consumption. A study on commercial smart meters indicated an accuracy of around 99.96% within a $+/-2\%$ accuracy range [38]. Generally, the accuracy is considered the most specified feature for any meter, and often a 0.5% minimum accuracy is considered adequate for revenue billing [36]. The inaccuracies mainly stem from the ADCs and transformers (both voltage and current). The ADCs introduce a quantization error, which can be reduced using a higher-bit ADC corresponding to the smaller step-size.

2.1.5 Application Environment

The e-monitors are utilized almost everywhere electricity monitoring is required, but how they are utilized differs on the basis of the area of application. The residential sector consists of housing units; the commercial sector consists of non-manufacturing business establishments (e.g., warehouses, hotels, restaurants, etc.), and the industrial sector consists of manufacturing units with fixed machinery (e.g., motors, drives, generators, etc.) [39]. In residential and commercial buildings, the aggregate load is mostly monitored using electromechanical meters. This single-point sensing can be single- or three-phase monitoring at the whole house or building-level. If one is interested in more detailed energy consumption information, circuit-level monitoring can be applied, which can be termed as multi-point energy sensing.

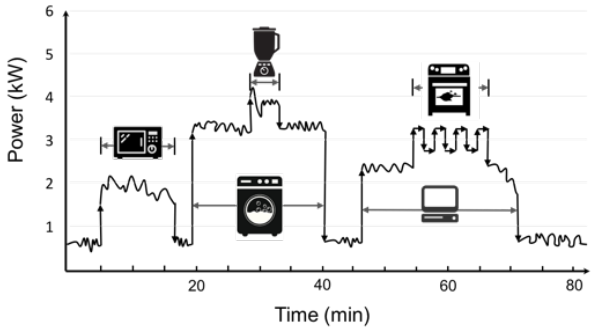


Figure 2.2.1: Disaggregation using non-intrusive load monitoring (NILM).

2.1.6 Cost of Energy Monitors

Cost is one of the most important factors when purchasing an e-monitor. A large-scale longitudinal survey was carried out by the Department of Energy and Climate in the United Kingdom to estimate the cost of different monitoring solutions for electricity and gas. The survey results recommend three different e-monitoring packages ranging from £210 to £950 per dwelling [40].

2.2 Non-Intrusive Load Monitoring

Recently, the use of NILM has increased, as it takes advantage of single-point sensing (i.e., at electricity mains) to identify the operating electrical appliances [41, 42]. NILM, as shown in Fig. 2.2.1, breaks down aggregate power consumption to essentially differentiate between specific individual loads. As compared to the ILM or other traditional approaches, NILM helps to achieve an enhanced appliance load profile at a reduced cost. Since its inception in the 1980's, NILM research has evolved considerably and has now produced new tools for feature extraction and load-disaggregation algorithms [43].

NILM is a combination of four main modules, as shown in Fig. 2.2.2. The DAQ module is responsible for measuring the aggregate load of a house or building. Depending on the disaggregation algorithm and the area of application, data can be acquired at a low frequency (smart meter) or a high frequency (in kHz to MHz range) [42]. In the feature extraction module, raw data is processed to detect and extract individual appliance events (on/off). There exist two main classes of feature extraction:

The steady-state load signatures concentrate on the signal amplitude and its smooth variations from high to low and vice versa. These amplitude changes are not abrupt and hence do not require fast sampling; they are preferred for appliances with a high power rating. The load identification module analyzes these features through the application of different disaggregation algorithms [44]. On the other hand, the transient features capture the abrupt changes in the current waveform to identify appliances. Transient load signatures essentially capture the unique pattern an appliance follows, particularly when it is switched on/off [45].

Earlier NILM approaches focused on feature selection and extraction with little emphasis on learning and inference techniques [46, 47]. The advances in computer science and machine learning techniques have led to innovations in data prediction and disaggregation techniques. Existing machine learning algorithms such as the support vector machine (SVM) [48], *k*-nearest neighbor (k-NN) [49], and artificial neural network (ANN) [50, 51, 52] algorithms have had a significant impact on the development of NILM. However, much effort is still required to bring the error caused by different prediction and disaggregation algorithms to within an acceptable range. Once data is acquired, proper handling and screening is required to appropriately present it for disaggregation. Different compression and storage methods have been suggested in the literature [53, 54]. Apart from acquiring transient features, using high-frequency DAQ also enables load disaggregation in near real-time [42]. The consumers can take advantage of near-real-time disaggregation feedback to adequately utilize DSM programs and reduce their load during peak hours. Such data can also be used for occupancy detection and can capture the occupant specific energy consumption [55].

The load or energy consumption profiles help to determine the pattern of energy usage with respect to time [56]. For consumers, these patterns are helpful in finding energy leakage. Utility companies use these patterns as a statistical tool for load forecasting. A precise and appliance-level consumption profile can be produced either through multi-point

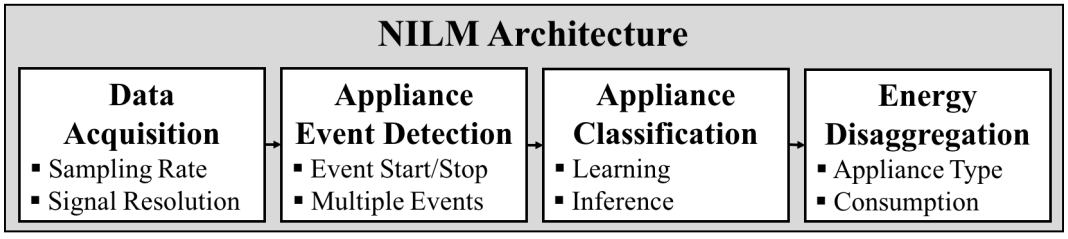


Figure 2.2.2: Non-intrusive load monitoring (NILM) architecture.

e-monitoring by using smart plugs or through load disaggregation techniques using NILM [57]. In addition to calculating power consumed by the appliance at a particular point in time, the shape of the power consumption profile also provides useful insight into the energy usage behavior. Consumption profiles are also useful in distinguishing the multi-state appliances on the basis of identical patterns of peaks during their operation.

2.3 NILM DAQ Requirements

To incorporate NILM and load disaggregation techniques in the next-generation e-monitors, we have listed some key requirements for NILM. The specific requirements may vary according to the specific application area of these e-monitors.

2.3.1 Sampling Frequency

One key requirement for any NILM system is to detect the appliances from the aggregate load accurately. Accurate appliance detection requires an adequate sampling rate to detect appliance switching events from the aggregate load. For precise appliance identification, many factors, such as appliance diversity, the number of appliances, the operating states of the appliances (e.g., washing machine, dishwasher, etc.), and the least amount of power consumed by an appliance under observation, need to be considered. As a general rule for disaggregation, a higher sampling frequency allows us to distinguish more appliances in near real-time. For example, with a sampling frequency of 10 to 40 kHz, one can differentiate 20 to 40 appliances. Increasing the sampling frequency beyond 1 MHz can help to distinguish 40 to 100 unique appliances [42]. To be considered for NILM-related operation, the sampling frequency can either be high or adjustable to fit the specific application.

2.3.2 High-Resolution

As discussed earlier, the resolution of any DAQ system is determined by the number of bits of the ADC. In terms of NILM, this resolution determines the uncertainty introduced by the DAQ system and hence decides the accuracy of event detection. In general, the

sampling frequency is responsible for the uncertainty along the x-axis, whereas the number of ADC bits defines the uncertainty along the y-axis during analog to digital conversion. Thus, a high resolution is important to accurately detect the events from the aggregate load and reduce the chances of simultaneous events.

Similarly, the resolution of ADCs defines the minimum change in the signal level (e.g., power, current, etc.) detectable from the aggregate load. Assuming no external noise, a house with 50 A demand can detect a minimum load of 11.23 W using a 10 bit ADC, whereas a minimum detectable load using a 16 bit ADC for the same house is 0.17 W. The survey results indicate most of the available e-monitors are capable of measuring a 1–5 W load, which is quite suitable for load disaggregation.

2.3.3 Accuracy

Throughout the NILM literature, various accuracy definitions have been used and are broadly covered in some recent studies [58, 42, 59, 60]. At times, accuracy is defined in terms of the fraction of correctly recognized events, while sometimes the fraction of correctly explained total energy determines the accuracy [46]. Norford et al. [44] determined the accuracy by utilizing the difference in the estimated and apparent power drawn by an appliance.

Similarly, some other NILM accuracy definitions include classification accuracy [61, 62], the appliance-wise fraction of load duration and the fraction of correctly identified or missed switching events [63], and the receiving operating characteristic (ROC) curve [58]. Makonin et al. [64] utilized classification and estimation performance (at both the overall and appliance-level) for reporting the NILM accuracy. To compare the accuracies of different NILM studies, Batra et al. [43] developed a toolkit to check the quality and accuracy of different datasets compared against predefined NILM algorithms.

2.3.4 Multi-Environment Operability

Another requirement for NILM is the ability to acquire the consumption data from multiple channels. The commercial and industrial environments require at least three channel measurements for measuring three-phase systems. Similarly, sometimes the residential households are also equipped with a three-phase system; thus a three-channel DAQ system

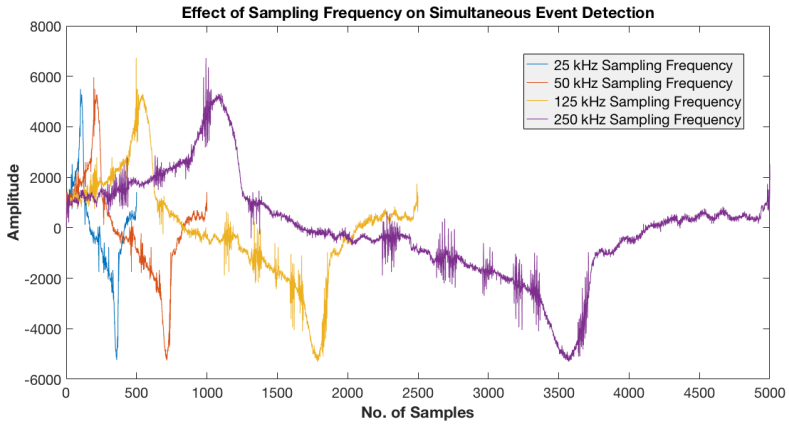


Figure 2.3.1: Current waveform acquired at 250 kHz using CLEAR and downsampled at different sampling frequencies.

is required to adequately handle the simultaneous measurements. The number of inputs is doubled if both the voltage and current are measured simultaneously. Apart from a couple of e-monitors and the smart plugs, the surveyed e-monitors fulfil the criteria of a three-phase measurement. Some NILM-enabled e-monitors such as Smappee [65], Verdgris, and CURB acquire the aggregate load at the circuit-level, which facilitates in appliance classification and detection using the NILM algorithms as a result of a reduced appliance number in a single circuit.

2.3.5 Simultaneous Event Detection

Most NILM studies follow the switch continuity principle (SCP) [21], which states that at a given instant, only one appliance is switched (on/off). Such an assumption may lead to error when a number of appliances are operating, such as in the office environment, or when more than one multi-state appliance, such as a washing machine and dish-washer, is operating together. One way to deal with this issue is to increase the resolution and sampling frequency of the aggregate DAQ. The effect of data granularity in an office environment can be observed in Fig. 2.3.1. The spikes indicate the switching events caused by the switched-mode power supply (SMPS)-equipped office appliances. For a single current waveform at 250 kHz (5000 sample points), the switching events are easy to detect, whereas the downsampled 25 kHz (500 sample points) current waveform introduces two or more simultaneous events, which are difficult to detect using available disaggregation algorithms.

2.3.6 Appliance Identification Parameters

Apart from basic parameters such as the current, power, energy, and harmonics [66], the new range of e-monitors are expected to be equipped with advanced sensors and a high processing power to acquire transient appliance features. These features or parameters are utilized to detect appliance switching from the aggregate load using machine learning algorithms. Khal et al. [5] have identified 36 such features, including wavelet analysis, voltage-current (*V-I*) trajectory, inrush current ratio, waveform approximation, and log attack time, along with other spectral and temporal features. When considering load disaggregation, it is always better to incorporate more parameters, as certain parameters work better for particular load types [67, 68].

Similarly, the instantaneous admittance waveform (*IAW*) [59] is a robust feature, as it simplifies the calculations because small differences in impedance are harder to observe as compared to admittance (inverse of impedance) [69]; this can introduce some numerical instability as a result of sharp spikes as the voltage approaches zero. Similarly, the current waveform for dynamic loads such as air-conditioners varies from cycle to cycle. To capture these variations, we usually perform eigenvalue (*EIG*) analysis by rearranging the time-series current waveform into matrix form. The study on appliance load signatures [59] indicates that power-hungry appliances usually have higher first *EIG* features. Even the second and third *EIG* features of these appliances show a good correlation and can be utilized as a feature for appliance identification.

In addition to the main features or parameters discussed above, external parameters are also helpful in the e-monitoring of individual appliances and are known as side-channel features. Features such as the time of day, weather information, the appliance location in the circuit (single- or three-phase), and the appliance usage pattern can help to boost the appliance detection process [70]. The side-channel-assisted NILM can significantly enhance the ground truth verification capabilities of the NILM-based systems. In addition to external parameters, light, sound, and the electromagnetic field (*EMF*) are also utilized as side-channel features. To obtain the appliance switching information, the electromagnetic sensors are placed in close proximity to the appliance under observation (usually a 5–10 cm range) [71]. Similarly, the channel electrical noise can also assist in the appliance identification, but this has a strong dependency on the electrical wiring system—the main drawback of using this approach [72].

2.3.7 NILM Scalability

Scalability is one basic requirement for any DAQ system, and in terms of NILM, scalability can be considered as the ability of the e-monitor to detect newly added appliances. This can be achieved either through supervised or unsupervised learning techniques [73]. Generally, appliances are identified through their unique load signature in the current waveform during their start-up. Disaggregation algorithms scan for these abrupt changes to begin the process of classification, inference, and learning. Once a new appliance is added to the system, the disaggregation algorithms try to match the features of these newly added appliances with the already developed appliance feature database. Another use-case regarding scalability is to apply the disaggregation at a district-level [74]. This can help to detect the power-hungry appliances in the district and help the utility companies to manage the distribution by incorporating more renewable energy.

2.3.8 NILM Reliability

One of the main requirements regarding NILM-enabled e-monitors is to reliably scan for appliance switching events from the aggregate load. As a result of the unpredictable nature of appliance switching, the e-monitors are required to acquire the measurement data around the clock for a long time (ideally forever). To ensure reliability, an e-monitor is expected to withstand small network and power outages. The use of the on-board buffers, mass storage devices, and battery banks for backup power is encouraged. The increase in the sampling frequency and the number of measurement channels adds to the challenge of maintaining reliability.

2.3.9 Privacy and Data Confidentiality

Besides many benefits regarding NILM, one drawback often associated with the NILM technique is the lack of consumer privacy. The fine-grained energy utilization information cannot only reveal one's presence in the house, but such data can also help to deduce the activities and habits of the consumer [75, 76]. Lisovich et al. [77] experimented to determine what kind of information can be extracted from the energy consumption data. They concluded that even with just a 15 s data resolution, they were able to accurately identify the major operating appliances to infer the eating habits and sleeping cycles of

the residents. One way to solve this problem is to use a local storage and utilize on-site disaggregation algorithms to build load profiles.

2.3.10 Efficient Data Storage and Analytics

As most of the available disaggregation algorithms work on the precollected measurement data, the measurement data needs to be collected and stored by the e-monitor. With the simultaneous measurement of multiple channels, collecting error-free data is a major challenge. The data collection and storage challenge increases further with the high-frequency measurements requiring large volumes of data to be stored at a steady rate. It is also important to use well-established file formats to store the data. HDF5 is a commonly used data format in the NILM research community [43] as a result of its superior data handling. HDF5 is compatible with input from multiple simultaneous streams and supports large, complex, and heterogeneous data. Once the data is collected and stored, different machine learning and data analytic techniques are applied to accurately detect appliances from the aggregate load.

2.3.11 Cost-Effective and User-Friendly

The adaptability of new technology in the public domain is mainly attributed to factors such as the equipment cost, ease of installation, and user-friendly operation. Our survey (see Chapter 4) indicates an average price of € 375 for NILM-enabled e-monitors. The cost of equipment depends upon the single- or three-phase system and utilized accessories. Similarly, the NILM-enabled e-monitors in the survey mostly come with split-core CTs, which can be easily installed by clipping around the mains cable of an electric meter. The appliance-level energy consumption information captured by these e-monitors is available to the users through dedicated apps. Disaggregation is an automated process and depends on whether supervised or unsupervised learning approaches are used, as discussed in Section 2.3.7.

CHAPTER 3

Related Work

The origin of energy monitoring dates back to 19th century, shortly after the inception of electricity. Traditionally, energy is monitored at a single point to calculate aggregate consumption on a whole house or building level. Sub-metering is applied to measure energy consumption on circuit or appliance level. Several similar data acquisition hardware solutions have been suggested in literature. Many variations of low and high granularity data acquisition hardware [78], with as high as 24-bit analog to digital converters, have been utilized for collection of different data sets [79]. Some of the high-resolution and high-frequency approaches are discussed here.

Since its inception in the mid 1980's, NILM research has evolved considerably and has produced new tools for feature extraction and load disaggregation [43]. NILM takes advantage of different machine learning techniques to predict the operating appliances from aggregate load and estimate their energy consumption. Earlier NILM approaches focused on feature selection and extraction with little emphasis on learning and inference techniques [80, 44]. The advances in computer science and machine learning techniques have led to innovations in data prediction and disaggregation techniques. Existing machine learning algorithms such as support vector machine (SVM) [48], k-nearest neighbor (k-NN) [81], and artificial neural networks (ANN) [51] have had a great impact on the development of NILM. However, further research is required to bring the error caused by different prediction and disaggregation algorithms within an acceptable range.

As compared to the ILM or other traditional approaches, NILM helps to achieve superior appliance load profiles at reduced cost. The limitations with NILM approach are dominant

in case of appliances with multiple operating states (e.g., motors, washing machines, LEDs, etc.) and lower power office appliances (e.g., laptops, LCDs, etc.,) equipped with switched mode power supplies (SMPS) which can be easily confused with noise. The appliance state changes can be analyzed using statistical characteristics such as shape, size, duration and other abrupt variations in the appliance signatures [5]. In order to observe these sudden variations in the signal, fairly high sampling frequency is required to reveal detailed and precise load signature characteristics. The disadvantage of using high-frequency sampling is extra processing power, data transmission, and storage requirements which adds to the DAQ hardware complexity and cost.

3.1 NILM DAQ Hardware

As the proposed work is focused on high-frequency DAQ, Table 3.1.1 lists some high-frequency DAQ hardware used for NILM research to acquire aggregate voltage and current. Some commercially available NILM compatible hardware solutions are also listed in Table 3.3.1. There are other high-frequency DAQ solutions such as the electrical data recorder (EDR) proposed by Maass et al. [82] which is capable of recording three-phase voltage time series data at 25 kHz. Although the acquired data has adequate resolution, the absence of current measurement limits their use as the disaggregation algorithms, which rely on load signature from current waveform. Similarly, there exist some other high-frequency DAQ solutions for per appliance energy consumption measurement (e.g., smart plugs, smart power strips, etc.) but are not considered in this study as we are only interested in NILM.

Some non-conventional hardware approaches have also been proposed in the literature. In [83], the authors have presented a new sensing approach that decouples dynamic range and resolution during current waveform measurement. This division of measuring range facilitates an optimized sensor design for specific applications. The windowing sensor approach divides the measurement in two subsystems, the compensation current and residual current measurement, which demonstrates accuracy over a wide range. Similarly, some math-based approaches with special hardware have also been proposed in the literature for accurate estimation of the spectral content of the measured voltage and current waveforms [84, 85]. We are also not considering smart meters due to lower sampling frequency, which makes it difficult to avoid simultaneous evens. Similarly, the data from commercial meters

is owned by the utility company.

3.2 NILM Event Detection

An event-based NILM approach consists of following main steps: Power measurement, event detection, classification, and energy disaggregation [91]. Besides reducing errors resulting from data acquisition (power measurement), the event detection step is yet another important step to ensure accurate appliance disaggregation. Since disaggregation accuracy is directly related to detection accuracy, accurate event detection is of paramount importance for proper classification of appliances. Similarly, the detection accuracy is also directly related to sampling frequency, where higher sampling results in higher accuracy detection [92]. Additionally, the higher sampling frequency also helps minimize simultaneous events.

Over the years many event detection approaches have been proposed in the literature. Wild et al. [93], propose an unsupervised approach based on kernel Fisher discriminant analysis (KDFA) for non-linear and variable loads. They define event as an active session which is marked as a deviation between two consecutive steady states. These active sessions vary in duration so an important task for NILM detector is to accurately identify the start and end duration. On the other hand, Meziane et al. [94, 95] present a high accuracy NILM detector (HAND), an unsupervised event-based algorithm which even performs better than KFDA approach. The HAND algorithm is simple, iterative, and fast as it uses standard deviation (SD) of the current signals envelope as a feature for disaggregation.

Similarly, Baets et al. [96] present a statistical event detection based on chi-squared goodness of fit (χ^2 GOF) approach and claim a 7-12 percent performance increase in F-measure as compared to state of the art. In another work, Baets et al. [97] present a Cepstrum smoothing method to eliminate noise for better event detection and claim the results to be compatible with χ^2 GOF. Trung et al. [98] present an improved cumulative sum (CUSUM) approach for event detection, especially for multi-state appliances. The other highlight of their approach as the use of FPGA (Spartan 6) to accelerate the CUSM detection from about 186 us in the CPU to about 1 us in the FPGA.

Similarly, some other statistical approaches based on different transforms have also been suggested in literature for NILM event detection . The Fourier transform has been applied by Li et al. [99] to detect appliances and the obtained harmonics are used as input for

Table 3.1.1: List of high-frequency DAQ (aggregate and circuit-level) solutions for NILM

NILM Research Studies					
Author/Study	**Hardware**	**Sampling**	**Resolution**	**Current Sensor**	**Voltage Sensor**
REDD [79]	NI-9239	15 kHz	24-bits	TED (200 A)	PICO (TA041)
BLUED [86]	NI-USB-9152A	12 kHz	16-bits	TED (QX 201-CT)	PICO (TA041)
UK-DALE [87]	Soundcard	16 kHz	24-bits	YHDC SCT-013-000	Ideal Power (77DB-06-09)
Clifford et al. [88]	NerdJack	8 kHz	16-bits	Inductive CT	-
Proposed Study	Custom Design	250 kHz	16-bits	LEM (HAL-50 S)	BLOCK (VB 1.5/2/6)
NILM Compatible Commercial Products					
Company	**Hardware**	**Sampling**	**Resolution**	**Current Sensor**	**Voltage Sensor**
LabJack [89]	UE-9	250 Hz	12 to 16-bits	LEM LA-55/205	Proprietary
Smappee Pro [2]	Proprietary	2-16 kHz	-	magLAB SCT-T24	Proprietary
Verdigris [2]	Proprietary	7.68 kHz	16-bits	Third Party	Proprietary
CURB [2]	Proprietary	8 kHz	1 W	Proprietary	-
Sense [90]	Proprietary	1 MHz	-	Proprietary	-

the classification using support vector machine. In another approach, Su et al. [100] compared the wavelet transform to the short time Fourier transform (STFT). The authors concluded that the wavelet transform is better suited to detect start-up transients, because the STFT necessitates a fixed window-size, which could either be too short or too long for some events. However, the study covered the start-up events of only three appliances (non-aggregate data) in a laboratory setting.

In a similar study, the event-based detection approach by Alcala et al. uses Hilbert transform (HT) to detect the envelope of transient states [101]. The main disadvantage of using only HT lies in the fact that it is not always possible to obtain a physically meaningful instantaneous frequency [102]. Hence, reducing a function in multiple intrinsic mode funtions (IMFs) using HHT increases the probability to obtain meaningful instantaneous frequency.

Although some of above mentioned techniques provided high event detection accuracy, all of these techniques were developed for residential environment where appliance density is generally low. The main event detection challenges for commercial building, as pointed out by Norford et al., are caused by periodic overlapping events [47]. These challenges make the normal event-based NILM approaches for residential environment unsuitable for commercial settings [103]. Besides NILM, the Hilbert Huang transform (HHT) has been applied to various fields such as analyzing seismic data and earthquakes, ocean waves, image processing, and ECG changes detection in medical studies [104, 105]. Similarly, HHT has also already been used in the context of the power system for short-circuit detection, bearing fault detection, power quality classification, broken rotor bar detection and load forecasting [106]. Also the HHT was used in a number of papers to detect and analyze low frequency oscillations in the power system[107, 108, 109] and to analyze transient disturbance signals in the power system [110].

3.3 Compression of Energy Data

Over the years, many approaches have been presented in literature, as shown in Table 3.3.1, but in reality, they are all variations of Shannon-Fano coding and Lempel-Ziv77 (LZ77). Earlier compression approaches focused on hardware coding until Huffman implemented dynamically generated codes based on input data. Many variations of the Lempel-Ziv77 algorithm have been derived, but only a few such as Lempel-Ziv-Markov chain algorithm

(LZMA), Lempel-Ziv-Welch (LZW), and DEFLATE with its variants such as GZIP and Bzip2 have survived. Similarly, with the launch of internet and incorporation of more images on web pages, several new formats such as ZIP, GIF, and PNG were introduced.

For decades, different audio and other data compression techniques have been developed and utilized. While most of these techniques are purpose-built for specific use-cases, some are more general and find their application in multiple fields. Recently, some of these techniques have been utilized to compress energy data with the focus on the aggregate load (data collected by smart meter representing the whole house load). The aggregate load compression is less efficient as compared to appliance level compression, since some individual appliances such as refrigerator tend to follow a periodic pattern, and a repeated pattern can be compressed efficiently [111].

Ringwelski et al. [112] compared lossless compression algorithms for smart plugs and smart meter based energy data in terms of compression ratio and processing times. For low sampling frequencies (1 s), they were able to achieve average compression rates between 75% and 95% with relatively modest execution time. For lower execution time, they achieved 40-60% compression rate. In another study, Zeinali et al. [113] proposed applying adaptive Huffman (AH) and Lempel-Ziv Welsh (LZW) algorithms for wireless transmission of low-frequency smart grid data (10 minutes to one-hour). Using LZW, they achieved 74-88% compression rates on average. Although they claim to achieve 98-99% compression ratio using double compression approach (both LZW and AH), the sampling rate is too low for considering load disaggregation. Similarly, Unterweger et al. [114] presented a compression approach tailored for low complexity encoding, decoding, and transmission of smart meters load profile data in their study. As compared to other approaches, the proposed method reduces the average bit string length per value. They assume that if smart meter data is similar to their test data, the data volume can be reduced by almost 90%, but again only low-frequency (1 s) data were analyzed with their approach.

Most of the energy data compression studies concentrate on the low-frequency data, but since we are interested in detecting the appliance switching events from the aggregate load, our focus in this work is to discover compression techniques suitable for aggregate load, especially considering the high-frequency data. Some high-frequency datasets employ compression techniques on the aggregate load data. It should be mentioned that data compression is generally more effective on a single big file, than on individual files.

J. Z Kolter et al. [115] published a high-frequency dataset at 15 kHz aggregate data using

Table 3.3.1: History of prominent compression techniques

Compression History				
Technique	**Author**	**Year**	**Type**	**Prominent Features**
Morse Code	S. B. Morse	1836	Code	Frequently used characters such as 'e' and 't' had shorter codes
Shannon-Fano Coding	C. Shannon, R. Fano	1949	Code	Start of main frame computing by assigning codes to symbols
Huffman Coding	D. A. Huffman	1951	Code	Modified the bottom-up probability tree to top-down approach to increase compression efficiency
LZ77	A. Lempel & J. Ziv	1977	Sliding-Window	First time utilized dynamic dictionary
LZSS	J. Storer & T. Szymanski	1982	Dictionary	Modified LZ77 (references omitted if length less than break-even point
LZW	T. Welch	1984	Dictionary	Other variations include LZMW, LZAP, and LZWL
LZMA	Numerous	1996	Dictionary	Other implementations include XZ, LZIP, ZIPX,LZHAM, and DotNetCompression
PKZIP2	P. Katz	1996	Dictionary	Combination of LZ77 and Huffman coding

24-bit analog to digital converter (ADC). The data was recorded locally using a mass storage device which reduced network traffic but resulted in 11 GB of data per day. The data was compressed using Bzip2, which resulted in 1.5-3 times reduction in the data volume. Similarly, J. Kelly et al. [87] presented an approach to acquire high-frequency aggregate data using the sound card. The uncompressed 16 kHz data using 24-bit files accumulated 28.3 GB[1] of data per day. Since the data was in audio format, they utilized free lossless audio codec (FLAC) to reduce the data volume to 4.8 GB per day.

[1]The author suggests 8.3 GB of data per day reduced to 4.8 GB per day (57% of its original size) on his blog (http://jack-kelly.com/data/)

CHAPTER 4

Energy Data Acquisition

The residential and commercial buildings account for a fair share of the grid load [116]. Traditionally, buildings are considered as passive grid component, only consuming the power from the grid, but smart homes and buildings have given a new life to this concept. The smart homes and buildings have enormous potential in terms of energy generation through renewables and energy conservation through the demand side management programs (DSM) [117, 118]. Hence smart homes and buildings are regarded as an essential component of future grid.

The smart grid initiatives allow energy providers and consumers to intelligently manage their energy needs through real-time monitoring. Around the globe, the smart meters are increasingly penetrating the energy market[119]. This penetration results in enormous data volumes to be transferred, stored and analyzed. Perhaps the most critical use of collecting energy data is to empower energy providers, policymakers, and governments to implement energy management programs and consistently achieve energy performance improvements. The smart grid support for demand response provides strategies for the electricity service provider to shed loads during peak load periods with minimal consumer inconvenience [120]. Considering a large number of grid consumers, the raw data may reach in terabytes (TB) per day under different scenarios [121].

The primary aim of any household or building manager is to intelligently utilize appliances with regard to user comfort and preferences while emphasizing energy efficiency. To manage the amount of energy spent, it is necessary to measure how and where this energy is consumed. Under the smart grid paradigm, the next-generation smart buildings will require

bidirectional power and data communication to reduce demand during high wholesale market prices or grid malfunctions [122]. This situation calls for highly interactive metering technologies that act as middleware to seamlessly gather data regardless of the vendor or communication protocol.

4.1 Technical Survey Overview

For this research, we conducted a comprehensive online survey [123] of various e-monitoring solutions available on the market. The purpose of the online survey was to obtain detailed technical information regarding e-monitors, as limited information is publicly available. A total of 54 different companies were shortlisted and invited to participate in the survey. The survey included 79 different e-monitors, all of which were off-the-shelf monitors and hence owned directly by the customers. For three respondents, we were not allowed to publish the data, but their information is included in the results. We received responses from 18 companies for 27 e-monitors through the online forms, a response ratio of 34.1%. We further collected information from 9 companies on their 14 e-monitors through online literature. In the survey, we grouped similar monitors from the same vendor together. For complete data, please refer to our technical note in Appendix A.

4.1.1 Application Environment

We identified the applications of the available e-monitors and divided these into three main categories: residential, commercial (including buildings and offices) and industry. Some of these e-monitors could be deployed in multiple environments. According to the survey, more than 90% of the e-monitors could be utilized in the residential sector, while over 60% could be utilized for commercial use. Similarly, more than 30% of the e-monitors could be utilized for industrial use.

4.1.2 Monitor Categories

For our survey, we categorized different monitors on the basis of their installation and measurement position in the electrical network of buildings. These included smart plugs,

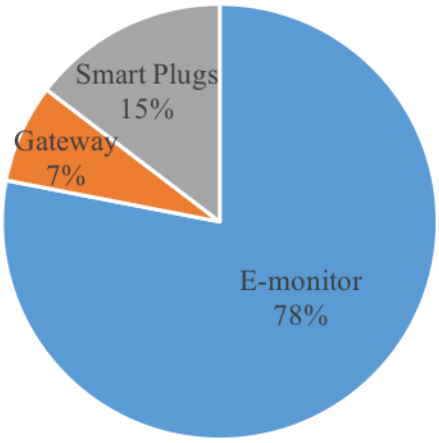


Figure 4.1.1: Different categories of e-monitors.

which are mounted on the wall outlet to measure individual end appliances. These smart plugs are utilized for the collection and validation of turn-on/off events to establish ground truth, to verify the load disaggregation algorithms. The smart e-monitors, such as smart-me, are installed between the electricity mains and distribution box inside the building. Because these monitors were owned and controlled directly by the customer, they were included in the survey. The majority of the e-monitors included in this study were installed in the fuse box or attached directly to the electric main and meters. These e-monitors usually incorporated electricity monitoring and analytics aimed at reducing the monthly electric bill through effective customer participation. This study also includes the gateways installed between the e-monitoring unit and the Internet to upload information directly to a cloud. Over 75% of the surveyed e-monitors included e-monitors followed by the smart plugs (Fig. 4.1.1).

4.1.3 System Compatability

We surveyed the different monitoring solutions and their compatibility with either single- or three-phase systems. According to our survey, over 70% of the e-monitors were compatible with both single- and three-phase systems (Fig. 4.1.2). The single-phase e-monitors could be scaled to three phases by using multiple units, and they could be calibrated using a pure resistive load so that the voltage and current curves did not mismatch.

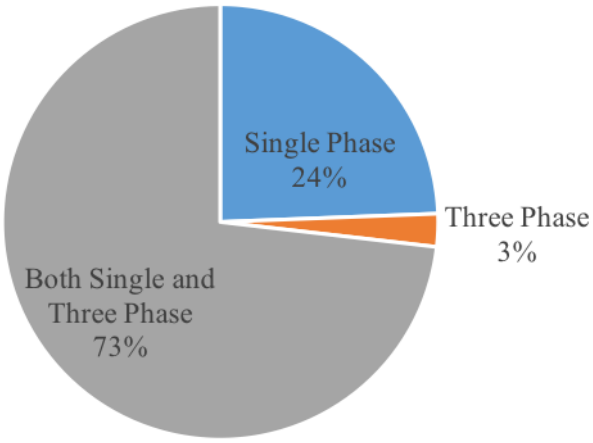


Figure 4.1.2: E-monitor utilization system.

4.1.4 Sensor Type

Some e-monitors only measure the current of the system while assuming a constant voltage; for three-phase systems, it is essential to include the voltage for at least one phase, if not for all phases. The three phases are ideally considered balanced, but if the load is not evenly distributed in each phase, which is very often the case, then different phases tend to have different voltages. The survey results indicate that nearly 60% of the monitoring systems used current transformers (CTs) and some utilized Rogowski coils for the measurement of the current (Fig. 4.1.3).

Although Rogowski coils are safer to use than regular CTs and offer a broader measurement range, they are still underutilized. A recent study [124] compared Rogowski coil-equipped digital meters with Ferraris principle-based electromechanical meters. The experiments indicated an increased reading of 376% as compared to conventional meters, which was mainly caused by electromagnetic interference in digital meters. A higher cost, as compared to the CT, is also a factor for the underdeployment of Rogowski coils. It is also noteworthy that only 21% of the e-monitoring solutions independently measured voltage. Some monitoring solutions directly used a shunt, pulse count, and optical measurement from the meter (see Appendix A).

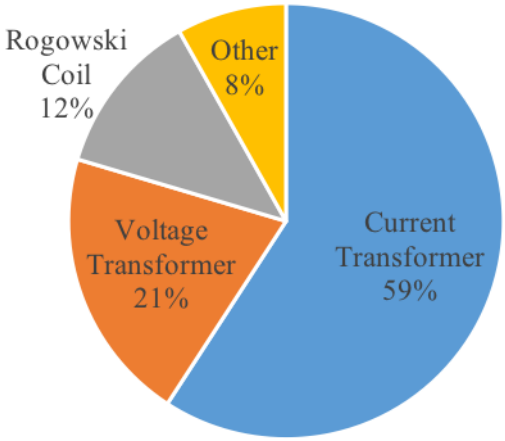


Figure 4.1.3: Types of sensor used by e-monitors.

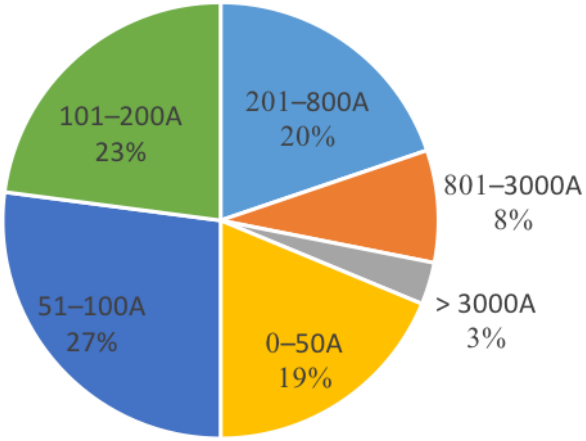


Figure 4.1.4: Rating of current transformers.

4.1.5 Sensor Rating

The type of sensor utilized depends on the application and is defined by the maximum load to be measured. A CT consists of an iron core with primary and secondary coils wrapped around it. The survey results indicate a wide variety of CTs utilized by different e-monitors (Fig. 4.1.4). The results also indicate that around 70% of the CTs were rated up to 200 A. This was due to the extensive use of e-monitors in the residential sector, for which the maximum load at any given time does not typically exceed 200 A.

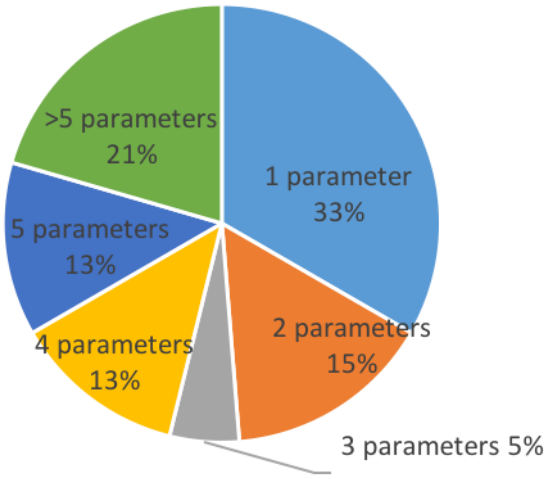


Figure 4.1.5: Number of e-monitor parameters used.

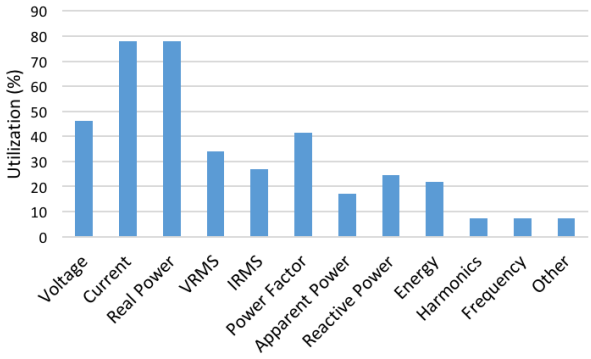


Figure 4.1.6: Parameters used by the e-monitors.

4.1.6 Parameter Type

E-monitors primarily measure the system voltage and current passing through a point at any given point in time. On the basis of these measurements, many different parameters can be calculated. According to the survey results, most of the e-monitors utilized a single parameter, followed by the use of five or more parameters (Fig. 4.1.5). Although almost all of the e-monitors measured the voltage and current (except when the voltage was assumed constant), these measurements were not necessarily displayed to the user. Approximately 80% of e-monitors utilize current and real power to indicate load, followed by voltage (Fig. 4.1.6). The inclusion of load-specific parameters enhances the distinction among the appliances. In our survey, some e-monitors utilized up to nine distinct parameters.

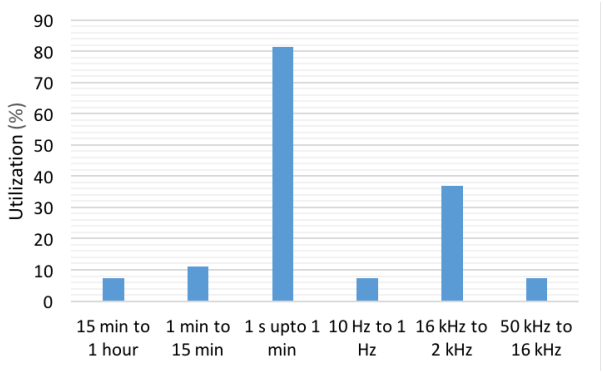


Figure 4.1.7: Sampling frequency used by e-monitors.

4.1.7 Sampling Frequency

The sampling frequency is an important dimension for comparison and is required for the proper conversion of an analog signal to digital. The voltage waveform is usually quite stable and can be reconstructed easily, but the current waveform is not even close to a proper sine-wave and hence requires increased sampling for proper digital reconstruction. Our survey indicated that a 1 s to 1 min sampling rate was commonly used by these e-monitors (Fig. 4.1.7).

Not all the participants disclosed the number of bits, but most ADCs lay between 10 and 16 bits. From the received data, most of the monitors were using 16-bit ADCs. Another important parameter associated with resolution is the power resolution, that is, the minimum level of power measured by these appliances. Most e-monitors have a power resolution of between 1 and 5 W, making them capable of exact and accurate metering.

4.1.8 Measurement Channels

Some load appliances rely on three-phase measurement, while others operate on a single phase, consequently affecting the number of channels that can be monitored. More than 75% of appliances support the measurement of one to three distinct channels (Fig. 4.1.8). With three input channels, one can either measure three separate single phases or one three-phase. Some e-monitors can simultaneously measure multiple channels (at the circuit and breaker level) in an electric cabinet. Circuit-level energy measurement can help in the disaggregation process, as an individual circuit has fewer appliances as compared to an entire house. As a result of the reduced set of appliances, there is also a smaller

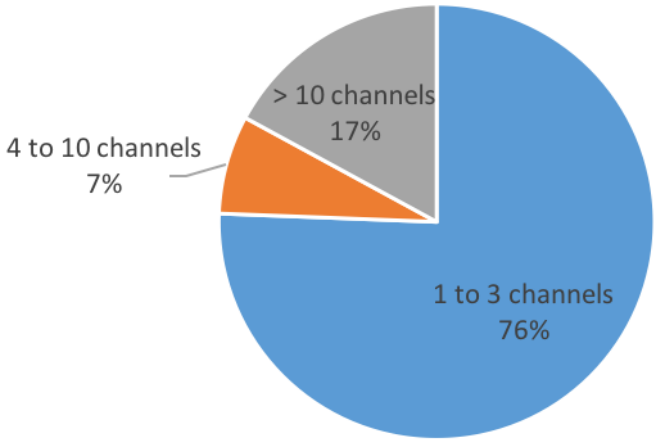


Figure 4.1.8: Number of channels measured by e-monitor.

probability of appliances switching on or off at the same time. With Verdigris, one can accurately measure about 42 different channels/circuits [125]. GridSpy [126] is another example of a system capable of measuring six circuits per node (wireless data collector) and 30 circuits per hub (collects and uploads data) and can scale up to 600 circuits per site. CURB Pro is also capable of monitoring 18 breakers per hub and this breaker level measurement (hardware disaggregation) facilitates disaggregation algorithms, as the type of load appliance on a particular breaker is already known [127].

4.1.9 Storage Type

E-monitors are capable of storing data either locally or by uploading to a cloud to perform further analytics. Most e-monitors prefer to upload data to a cloud, while others have the dual capability to store data locally and, at the same time, upload it to a cloud (Fig. 4.1.9). Issues such as data privacy and confidentiality can be decreased by using local or private cloud storage. With such data, arrangements can be made for the consumer to take advantage of load disaggregation and compute the appliance-level power consumption.

4.1.10 Equipment Cost

To compare costs, we converted all prices into euro to help consumers find the best-suited solution according to their application requirements and budget constraints. According to our survey, the cost of e-monitors varies between €38 and €3220 for a single product

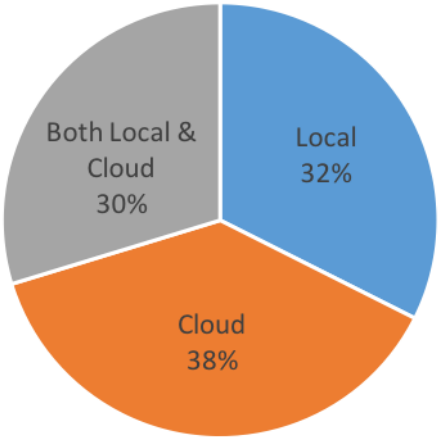


Figure 4.1.9: Different storage options for e-monitors.

according to its application. The typical price range is €452 to around €655, depending upon single- or three-phase systems and accessories utilized with the e-monitors. The prices of smart plugs range between €15 and €79, with an average price of around €48.

4.2 Findings, Observations, and Recommendations

The primary purpose of this study was to gather technical information to facilitate researchers, facility managers, and general consumers in selecting an e-monitoring system that best fits their requirements. Commonly, e-monitors are used to track and display the amount of utilized and conserved energy. The critical differences in e-monitors originate from the application area, the sampling frequency, the resolution, the system configuration, and the sensor type. Because the power consumed by e-monitors is quite small, we have not considered it in our study.

We believe that consumers can participate efficiently in DSM programs once they are provided with real-time energy consumption information, particularly at the appliance level. Information regarding appliance-level energy consumption can help to identify energy-hungry appliances and facilitate demand response. Some of the surveyed e-monitors, such as Smappee, Smappee Pro, Neurio, Verdigris, CURB Pro, and CURB Duo, which made up around 18% of the surveyed e-monitors, already claim to utilize NILM techniques. Similarly, most of the other e-monitors possess enough resolution, parameter diversity, processing power, and sampling frequency to employ disaggregation.

In some cases, monitoring appliance health is critical to the overall system operation, particularly for industrial applications. The load disaggregation techniques can facilitate the prediction of faults and recommend appliance maintenance before complete breakdown. Similarly, in addition to being a labor-intensive task, some of the most significant hurdles in the speedy roll-out of smart meters are data confidentiality and privacy concerns. As a result of the private storage and ownership of both the e-monitor and data, consumers can virtually experience smart grid benefits without compromising on privacy.

Most e-monitors can be used in multiple settings and configurations, as they come with numerous options regarding the sensor rating, the number of inputs, and the application area. Furthermore, the utility companies can also take advantage of the data from e-monitors (if allowed) to obtain detailed information regarding high-power appliances operating in an area and enhance the renewable integration through DSM programs.

Although NILM has been around for three decades, the technology never made its way into the public domain until recently. This was mainly due to a high equipment cost and a lack of disaggregation accuracy. Our survey indicates the presence of a new and affordable range of e-monitors, most of which can be easily upgraded to support NILM and disaggregation.

CHAPTER 5

Circuit Level Electric Appliance Radar

Before the extensive roll-out of smart meters, it is important to realize the potential benefits of smart metering to consumers. Circuit level electric appliance radar (CLEAR) is developed to tap these benefits and even expand the operational capabilities of available smart meters. In a bid to find a potential DAQ solution satisfying NILM disaggregation requirements, an extensive online technical survey was conducted to check capabilities of different off-the-shelf energy monitors [2]. Although some energy monitors fulfilled the requisite capabilities such as sensor rating, measured electrical parameters (e.g., voltage, current, power, power factor, etc.), accuracy, and number of inputs, but often lacked high frequency sampling required for analyzing transients. The lack of customizable DAQ features limits the application of these energy monitors in the context of NILM.

Due to its configurable sampling frequency, local/cloud data storage, multiple inputs, and ease of installation, CLEAR is applicable to various operating environments. We have initially installed CLEAR in an office environment to capture building-level aggregate energy data as, to the best of our knowledge, no such high resolution dataset is publicly available for the office environment. The first prototype is operating at our institute. One challenge associated with office environments is the presence of multiple similar appliances such as laptops, LEDs, desktop PCs, servers, and printers. The use of SMPS also limits most disaggregation algorithms due to the use of switching regulators to increase energy efficiency. A key requirement is to use high frequency DAQ to uniquely capture the start-up transients. In addition to observing transients, the high frequency data also facilitates predictive maintenance of appliances, by observing degradation of appliance load signature over time.

Apart from error introduced by disaggregation algorithms, typical uncertainty issues stem from DAQ system. The DAQ module is expected to acquire the aggregate load at an adequate rate to distinctly segregate different operating appliances. Usually, voltage and current sensors are installed in the main distribution panel to accurately measure the voltage and current. As primary current is relatively high and dangerous to measure directly, the common current sensing techniques lower these high currents by using shunt resistor to measure voltage drop or by using magnetic sensing devices such as current transformers (CT), Rogowski coils, or Hall-effect current transducers. We will now describe our design approach which led to the develop our customized DAQ system.

5.1 Earlier Approaches

5.1.1 Open Energy Monitor

We started with an off-the-shelf and cost-effective metering solution provided by open energy monitor (OEM) [2]. OEM, a platform that provides services for energy monitoring, control, and analysis of the energy data. These devices are compatible with Arduino, a well- known open source sensor-actuator platform. In the first phase, all the devices were assembled, soldered (GLCD) and tested with offline storage (microSD card) and online storage (emoncms.org). The main components of OEM include:

- Monitoring unit (emonTx V .3)
- Current transformer (CT)
- Voltage transformer (9V AC-AC)
- Base unit (Raspberry pi)
- Display (emonGLCD)

The main data processing unit is capable of monitoring four separate circuits simultaneously. We used emonTx V.3, capable of measuring three CT inputs with 33kW maximum power and a single 4.5 kW CT for precision measurements. This version is equipped with Atmega 328P, a 10 bit analog to digital converter and a surface mount RF antenna to increase the range of transmission as most energy meters are deployed in the basement.

The CT is a type of instrument transformer used for the measurement of current passing through a wire at a given time. A CT produces smaller output current for a given input, as secondary current is directly proportional to the primary current. For our experiment, we tested different split-core type CTs with turn ratio of 1:2000 and conversion ratings of 100A: 50A, 100A:1V, and 30A:1V. The main difference is the absence (for output in amperes) and presence (for output in voltage) of burden resistance. In order to capture line voltage samples, we used a 240V-9V AC-to-AC adapter to derive real power, power factor and line frequency.

The raspberry pi acts as a web-connected base station, which directly receives data from energy monitor and posts it to an online database. We have used both online web logging on emoncms.com and offline logging on its internal microSD card or an external hard drive. Raspberry pi is equipped with an external RFM12Pi radio frequency transceiver for communicating with energy monitor and display unit. In order to display the real-time energy consumption from emonTx, we used Arduino compatible wireless graphical display unit, also known as emonGLCD. It also includes a built-in temperature and LDR light sensor. It is also equipped with RFM12Pi transceiver to exchange data with emonTx. Since the monitoring unit lacks a hardware clock, time synchronization was performed by raspberry pi.

A prototype OEM V3 was installed in a kitchen with multiple appliances, as shown in Fig. 5.1.1. Although OEM is capable of recording 50 sample pairs (voltage and current) per second at 2.5 kHz, the reporting rate is up to 1 s (10 s by default). We configured OEM to increase the reporting rate but after 2 samples/s, packets started to dropout. OEM is accurate for low-frequency metering to acquire aggregate power data but, due to lack of samples at high-frequency, it was inappropriate for accurate disaggregation. Since emonTx-V3 is equipped with a 10-bit ADC (Atmega 328P), it is difficult to detect small loads. Similarly, CT and AC-AC adapter are other sources of error resulting in the phase-angle mismatch.

In order to reduce uncertainty caused by OEM, some researchers have explored the possibility of adding a separate ADC to enhance the capability of the current emonTx. OEM is an Arduino compatible platform which provides services for energy monitoring and analysis of energy data. The main data processing unit (emonTx V.3) is capable of simultaneously monitoring current for four separate circuits using CTs. Out of these, three CT inputs can measure a maximum power of 33 kW whereas the remaining 4.5 kW input is utilized for precision measurements. This version is equipped with Atmega 328P, a 10

Figure 5.1.1: Appliances connected to open energy monitor in kitchen

bit analog to digital converter, and a surface mount RF antenna to increase the range of transmission as most energy meters are deployed in the basement of the residence.

5.1.2 Sound Card Energy Monitor

Although OEM was accurate for low frequency metering to acquire aggregate power data but due to lack of samples at high frequency, it was not appropriate for accurate disaggregation in office environment. As emonTx-V3 comes with a 10-bit ADC, so theoretically, we are unable to detect small loads by using these standard OEM devices. For example, consider a small home with a peak requirement between 0 A and 30 A at any given time. As emonTx uses only 9 bits to capture both positive and negative curves over the range of 0 to 30 A, so theoretically, minimum detable load is 46.8 W.

Typical office appliances such as laptops, LCD's, LED's and florescence tubes with lower power rating cannot be detected with the 10-bit ADC. In order to tackle this problem, we explored means to increase the accuracy of signal read by using a sound card to capture a high-resolution signal using the card's 16-bit ADC operating at 48 kHz. Low cost sound cards have been around for years with built-in filters to remove noise below 20 Hz and many lossless compression techniques are also available to reduce data size.

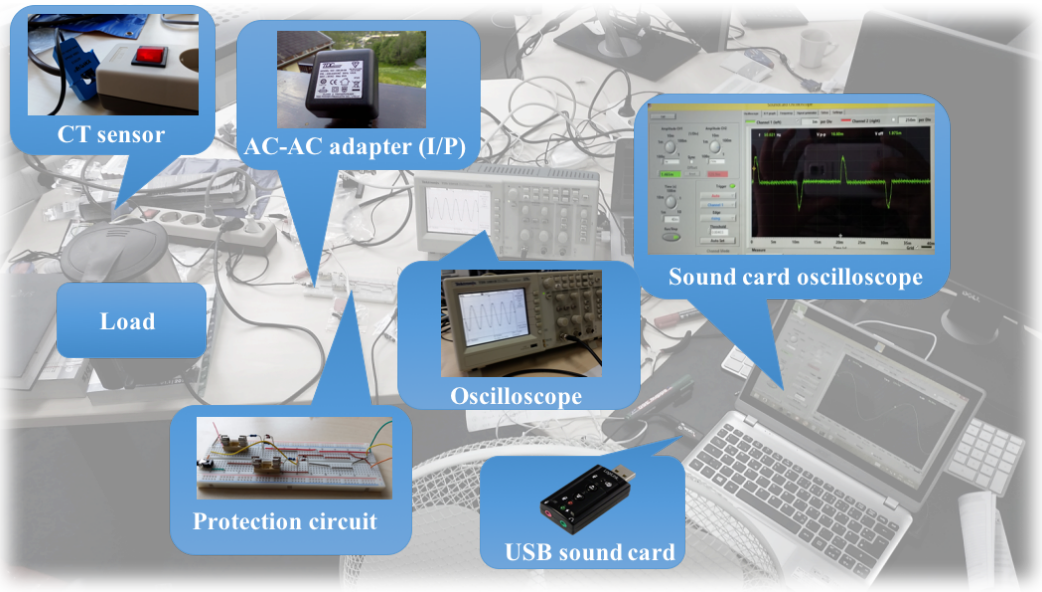


Figure 5.1.2: Sound card energy meter (mic input)

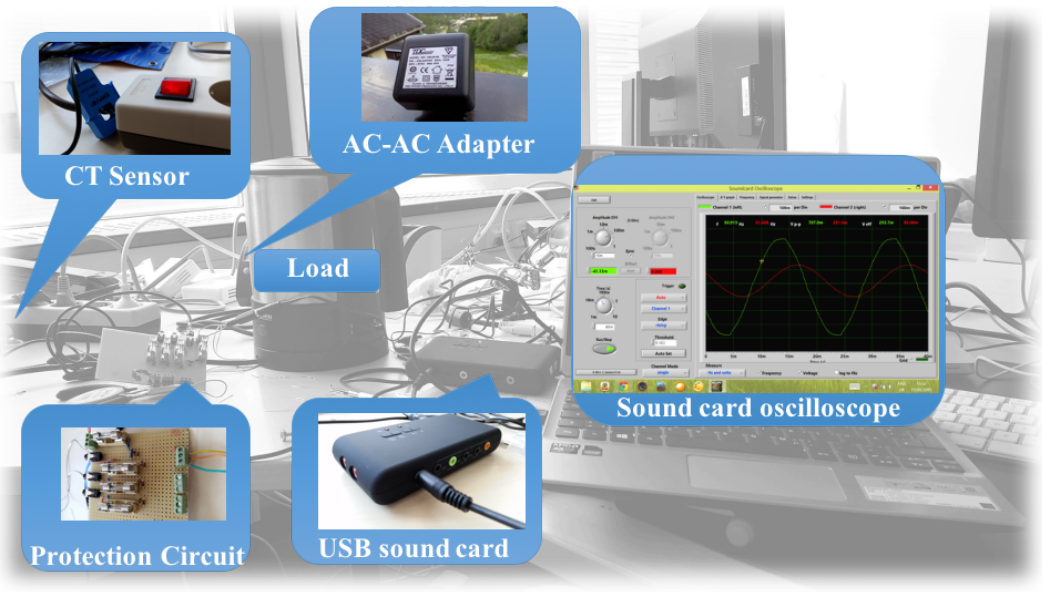


Figure 5.1.3: Sound card energy meter (line input)

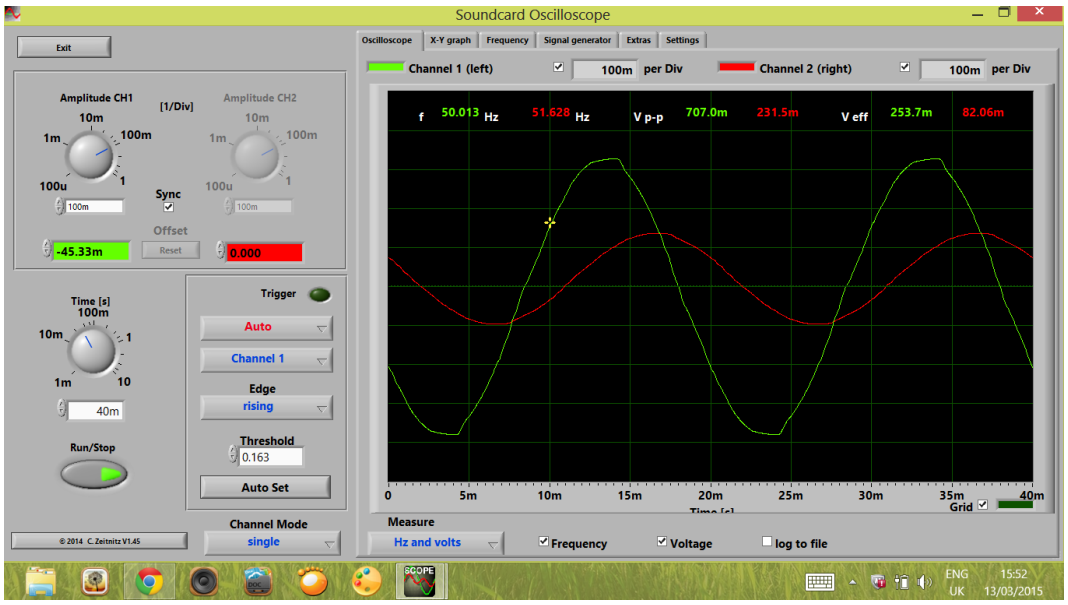


Figure 5.1.4: Voltage (green) and current (red) waveforms for electric kettle

The experimental setup for the USB sound card with mic input is demonstrated in Fig. 5.1.2 whereas Fig. 5.1.3 shows an extension with line-input. In our experimental setup, we opted for different USB sound cards instead of laptops built-in sound card. The important aspect was the use of line-in port, having high impedance stereo input, to capture both voltage and current simultaneously. The microphone input has lower impedance mono-input which tends to mix both voltage and current signals. A closer look on the voltage and current signals for electric kettle, as shown in Fig. 5.1.4.

These unique power signatures help in appliance recognition from aggregate measurements. To extend the experimental setup for 3-phase measurement, we used three different USB sound cards with line-in and captured both voltage and current for each phase simultaneously as shown in the Fig. 5.1.5. The AC-AC adaptors were used to extract the information (e.g., phase difference, line voltage etc.) from voltage signals. The adaptor model we used (ideal power, 900 VAC and 600mA) had nearly zero quiescent power and phase variation of about 4 to 7.5 degrees. The system was calibrated to minimize the error. Some peak shaving of the voltage signal was observed but it was mainly due to the protection diodes.

To protect the sound card from dangerous voltage and current surges, we developed a simple protection circuit to keep the current and voltage measurements within the required

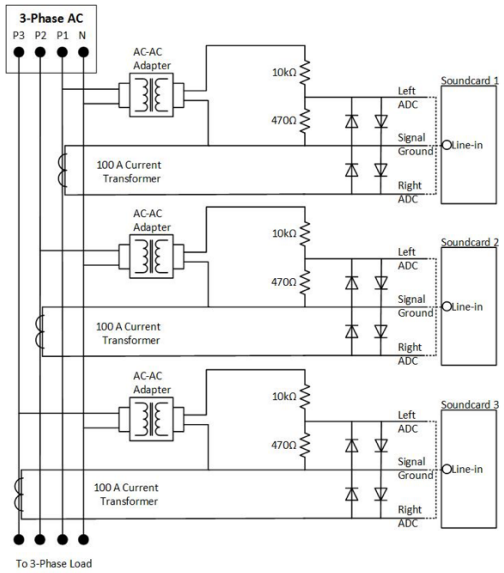


Figure 5.1.5: Thee phase sound card measurement system

operating range. We also utilized a software-based sound card oscilloscope[1] to observe the signal waveform in real-time. A similar study was carried out by Kelly et al. [87] to measure the single-phase voltage and current measurement for domestic households using the built-in sound card of a laptop. Our approach enhanced that research further by utilizing three USB sound cards to measure the voltage and current for each phase using line-in port (a high impedance stereo input to simultaneously capture both voltage and current). So, instead of just measuring a single-phase, our approach enabled us to simultaneously measure either three single-phase circuits or a three-phase circuit.

An important consideration was the use of line input (line-in) port, a high impedance stereo input, to simultaneously capture both voltage and current. In contrast to line-in, the microphone input (mono) has lower impedance and eventually mixes both voltage and current signals which makes it impossible to obtain the original values.

Due to better data resolution, the use of sound card demonstrated fairly accurate disaggregation by detecting appliance transient features from switching appliances. The preliminary testing with some disaggregation algorithms provided encouraging results. Although the sound card measurement system provided a cost-effective solution to simultaneously measure the voltage and current, it is associated with few drawbacks such as lack of multiple inputs, limited low-level signal processing, and lack of long-term

[1]https://www.zeitnitz.eu/scope_en

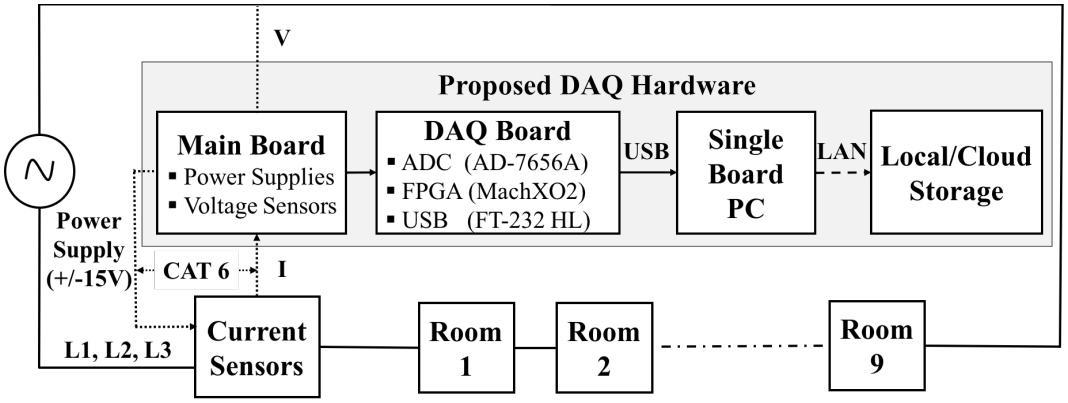


Figure 5.1.6: Design architecture of proposed hardware

error-free measurement for multiple phases.

5.2 CLEAR Hardware Design

Initial success with the sound card energy monitor encouraged us to develop a reproducible, high-frequency and high-resolution DAQ system for the cyber-physical systems. The design and customizable features make our proposed solution applicable to many different working environments. Our first prototype is installed in an office environment where one major challenge lies in detecting and isolating multiple switching events caused by SMPS-equipped appliances. The overall design architecture of the proposed energy monitor is shown in Fig. 5.1.6. The custom-made PCBs are housed in a laser-cut enclosure and are designed to fit together with the single-board PC as a single unit (Fig. 5.1.7).

5.2.1 Main Board

Being the point of contact between the electric cabinet and digital circuitry, the main board houses sensing units, power supplies, and auxiliary connections between the analog and digital part of energy monitor (Fig. 5.2.1). The primary function of the main board is to acquire an error-free, high-quality, and continuous stream of analog signals (line voltage and current) and deliver them to the DAQ board for further processing. For our prototype, we have chosen Hall-effect-based current transducers from LEM (HAL 50-S) capable of handling a primary nominal RMS current of 50 A with a measuring range of ±150 A. As

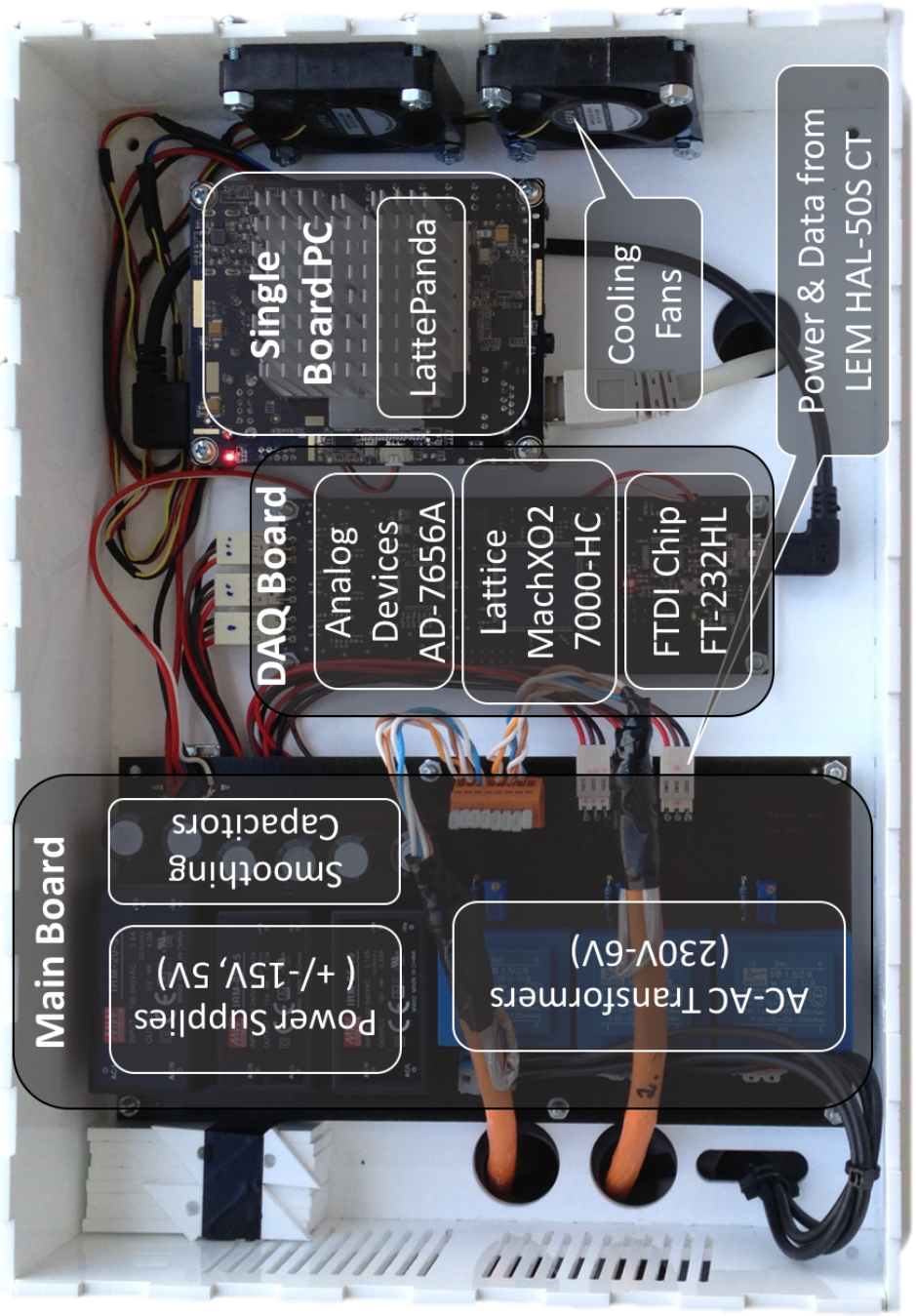


Figure 5.1.7: Prototype for data acquisition system

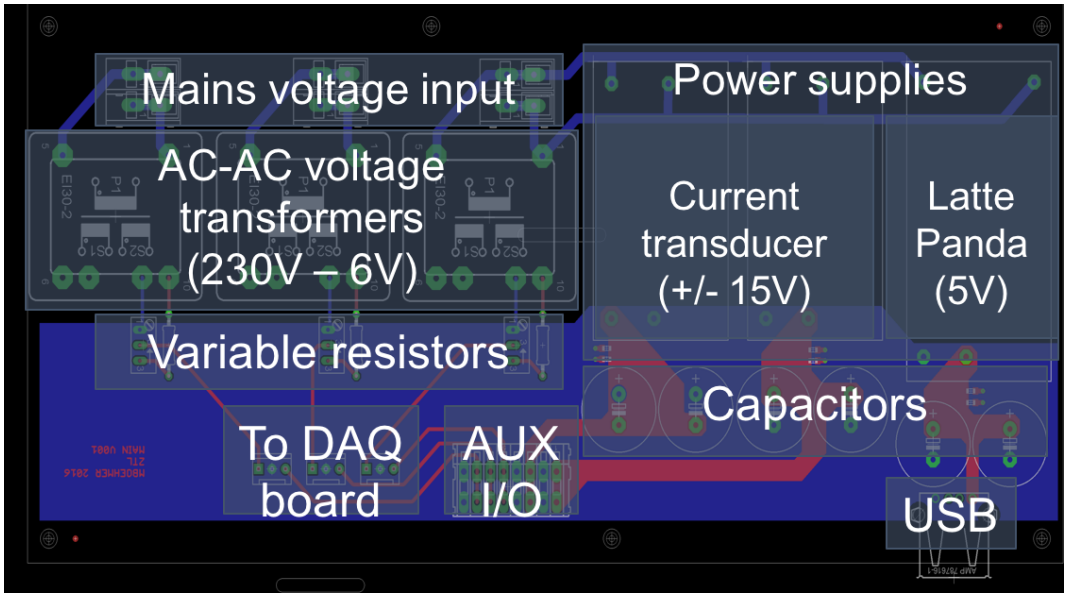


Figure 5.2.1: Main board consisting of power supplies and voltage transformers

the CTs tend to show better result near the full load, multiple turns on the primary winding increases the overall measurement efficiency. Since each line has a 16 A circuit breaker, we introduced three turns on the primary winding to boost the effective signal bandwidth while staying in the 50 A nominal rms current range.

The Hall-effect-based CTs also require an external voltage of ±15 V, which is provided by two power supplies on the main board. Additionally, the external power supplies keep the secondary winding of CTs energized to gather accurate calculations as no power is consumed by the components. Main features of HAL 50-S CT are shown in Table. 5.2.1. Similarly, the external power supplies also help prevent any potential accident caused by open secondary winding. Besides the 15 V power supplies for CTs, a 5 V power supply is also added to operate the single-board PC and cooling fans. Smoothing capacitors are added to ensure a clean power signal to CTs and help prevent any fluctuations from power supplies to affect the raw voltage and current measurements.

The line voltage for each phase is fed directly to one of three AC-AC voltage transformers which steps-down the voltage. Variable resistors are used to calibrate the voltage for each phase according to the required input level for the DAQ board. Two separate CAT-6 cables are used to connect the main board with CTs in electric cabinet. The first cable provides power (±15 V) to the CTs whereas the second cable transfers the raw measurements to the main board. The use of separate CAT-6 cables helps prevent cross-talk between power and

Table 5.2.1: Summary of LEM (HAL-50 S) CT parameters

ELECTRICAL DATA			
Notation	Parameter	Value	Unit
I_{PN}	Primary nominal rms current	50	A
I_{PM}	Primary current, measuring range	±150	A
V_{out}	Output voltage (Analog) @ $\pm I_{PN}$	±4	V
V_C	Supply voltage	±15	V
ACCURACY			
Notation	Parameter	Value	Unit
X	Accuracy @ I_{PN},T_A=25°	<±1	%
BW	Frequency bandwidth (-3db)	DC..50	kHz

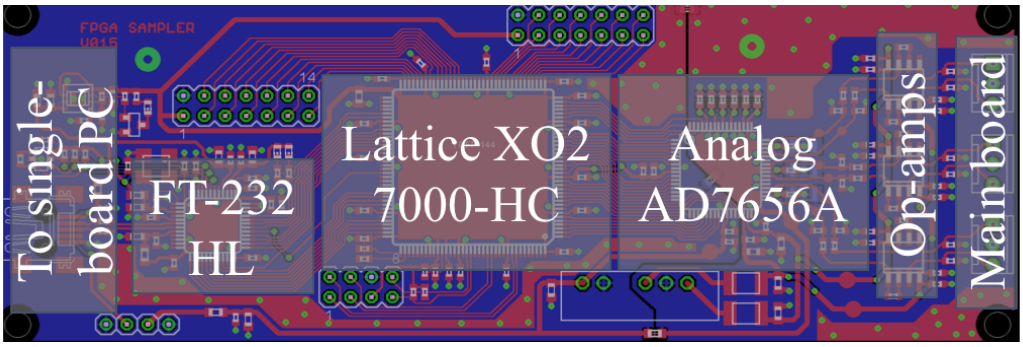


Figure 5.2.2: DAQ board consisting of ADC, FPGA and USB conversion chip

signal channel. The AC-AC voltage transformer and Hall-effect current transducers also provide galvanic isolation to improve the overall safety and equipment protection.

5.2.2 Data Acquisition Board

The DAQ board, as shown in Fig. 5.2.2, is tasked to convert the raw analog measurement data into digital format and forward it to the single-board PC for post-processing and persistence. The DAQ board contains no high-voltage component to isolate raw measurement data from any electromagnetic distortion caused by high power components. We have utilized a 16-bit, bipolar, successive approximation ADC from Analog Devices (AD-7656A), which is capable of handing six simultaneous channel streams at 250 kHz. A true bipolar signal in ±4 V range is accommodated with a 2.5 V on-chip reference. The six ADCs are grouped

as pairs of three to initiate simultaneous sampling of voltage and current signal pair for each phase.

After we acquire the high frequency data, proper handling is required to collect, secure, and store these simultaneous data streams. We have utilized a high-performance FPGA (Lattice MACH XOR 7000-HC) which triggers single shot to read data into the memory for buffering. Some special FPGA characteristics include high performance, instant power-on (microseconds), low power consumption, flexible on-chip clocks (8 primary clocks and 2 phase-locked loops per device), and infield logic update during system operation. To avoid momentary processing delays, first-in first-out (FIFO) buffers are added to stabilize data streams. A USB conversion chip from FTDI (FT-232 HL) operating in an asynchronous FIFO mode is added for USB data transfers to the single board PC.

5.2.3 Single Board PC

The prime job of single board PC is to collect the energy measurement data from the DAQ board and securely transfer it to persistent storage. With our proposed measurement system, both local and cloud storage options are available. The local storage provision is encouraged for privacy-conscious consumers where the disaggregation algorithms are applied locally and no private data is sent over the internet. Currently, we have attached a USB storage device which also serves as a backup to store high-frequency data for a couple of days. At 250 kHz, the measurement data from the DAQ board is around 281 GB/day. For the proposed hardware, we have utilized LattePanda due to its superior processing (quad-core 1.8 GHz) and memory (4 GB with upgradeable eMMC up to 64 GB) characteristics. Due to the high cost of LattePanda, we are also experimenting with Raspberry Pi 3 as a replacement.

5.2.4 Housing

The main board, the DAQ board and the single-board PC are housed together in a custom-built, laser-cut and non-conductive acrylic glass casing, capable of working over a broad temperature range. As the measuring unit is expected to be operational 24/7 for a long duration (ideally forever), two cooling fans are added to avoid over-heating.

5.2.5 Software Architecture

The software architecture is designed to operate the metering unit as a standalone DAQ unit. Initial experiments aim to implement the disaggregation algorithms on an external server and experiments will be carried out for on-site storage and disaggregation using different state-of-the-art NILM algorithms.

5.2.6 Energy DAQ Software

The energy DAQ (e-DAQ) is performed on a single-board PC running generic Linux distribution. The software architecture provides a common command and control infrastructure to enable bulk transfers via USB. The custom e-DAQ is initialized as a systemd service to start rest of the service and keep track of the Linux processes. Using command line switches and environment variables, the systemd services initialize the FPGA and other low-level parameters such as file names, chunk size, and sampling frequency configuration.

The initialization phase is followed by the USB bulk transfer requests to fill up the queue and initiate the transfer requests. A full buffer on the sampler board triggers a callback in the e-DAQ software as the data is forwarded to the Linux kernel. The triggered callback also includes a buffer containing the acquired data. Since the hardware supports customizable sampling frequency, the buffer size on the USB interface chip may not always align with our standard 14-byte packet size. Due to this size mismatch, the callback buffer might contain fragmented samples, which should be defragmented to align the samples before validating the missing samples using checksum identifier. This validation ensures error-free and consistent DAQ, hence the data gets written into a file.

Each file is labeled with a sequence number with a corresponding time-stamp to facilitate identification and reassembly process. The appropriate file size depends upon sampling rate and available bandwidth. A reasonable method is to align the file size according to the sampling time such as 2, 5, 10, 15, 30, or 60 minutes, depending on the samples acquired per second. At 250 kHz, we have chosen a 2 min duration for each file and 50 min data (25 files) chunks are batch transferred for storage, where each file is properly time-stamped to pinpoint the switching event during the disaggregation. Once the maximum file size is achieved, new data are written into a new file. The completed file is handled through

an asynchronous procedure to transform sequential ADC values into 16-bit little-endian integer values. Currently, we are collecting data for offline disaggregation and since the task is not time critical, it is performed asynchronously with actual data acquisition.

5.2.7 Collector Service

Concurrently, a collector service is operating to transfer finalized data files to a persistent storage (cloud) and perform a list of tests to validate certain system checks. With high sampling rate, huge volumes of data are generated at a firm rate, the storage backend must be capable of managing this consistently large flow of streaming data. The two-way bulk transfers via USB-ethernet interface (system commands and sampled data) may cause bottlenecks and hence require a trade-off between transfer speed and network bandwidth utilization to ensure a sustainable data flow.

The DAQ board consists of three local buffers to cope with short outages (minutes up to hours depending upon sampling frequency). A USB storage is also provided on the single-board PC in case of network outage. This allows uninterrupted measurements and files on the single-board PC until the collector service kicks in again after failure. After the network connection is re-established, the buffered data are transferred to a persistent storage.

During normal operation, the buffered files can be transmitted as soon as the collector service is activated, but since writing files to persistent storage is costly, files are batched up before persistence. This requires keeping all the files in RAM until the collector receives them. However, RAM is a scarce resource and most single-board PCs have limited memory. This problem is countered through a watchdog, which regularly moves the files from RAM to a local mass storage device. The mass storage device also helps prevent potential data loss in case of system failure or power outage.

CHAPTER 6

Evaluation

To test the effectiveness of available NILM algorithms and develop new algorithms for the office environment, we have installed the first prototype at our institute. The test environment consists of 9 offices with a maximum 51 different appliance models at a given time (since the number of occupants changed during the six month monitoring period). Due to special protection requirements, we were not allowed to install the monitoring hardware in the electric cabinet room (Fig. 6.0.1). So, the CTs are the only external components present in the electric cabinet (Fig. 6.0.2) and connected with the monitoring unit in the adjacent room through the CAT-6 cables.

Figure 6.0.1: CAT-6 cables to transfer power and current data

6.1 Evaluation Criteria

Although the particular requirements of any DAQ system vary with the type of application, we have listed a few key requirements for DAQ hardware with special emphasis on load disaggregation. The precise event detection from aggregate load requires adequate sampling frequency. A high sampling rate (R1) is required when the appliance diversity is high or in presence of low-power appliances such as laptops and other SMPS-equipped appliances. Similarly, higher resolution (R2) ensures proper matching and extraction of appliance load signature from aggregate load for accurate appliance detection. The higher resolution also helps to reduce simultaneous events due to the higher sample count. The DAQ system should be stable enough to acquire non-stop and simultaneous data recording (R3) for multiple channels (R4) using well-established file formats to support interoperability (R5).

The DAQ system should be scalable (R6) and reliable (R7) to handle additional appliances, circuits, and even withstand short network outages. The design should also ensure data privacy and confidentiality (R8) through local data handling. The DAQ system should be capable of efficiently storing (R9) the energy data to perform post processing and data analytics (R10). It should also support multiple operating environments (R11) and be capable enough to detect simultaneous appliance-switching events (R12) to deal with switch continuity principle. Finally, the system should be independent and user friendly (R13) to ensure large-scale deployment.

Figure 6.0.2: Increase in data sensitivity using loops

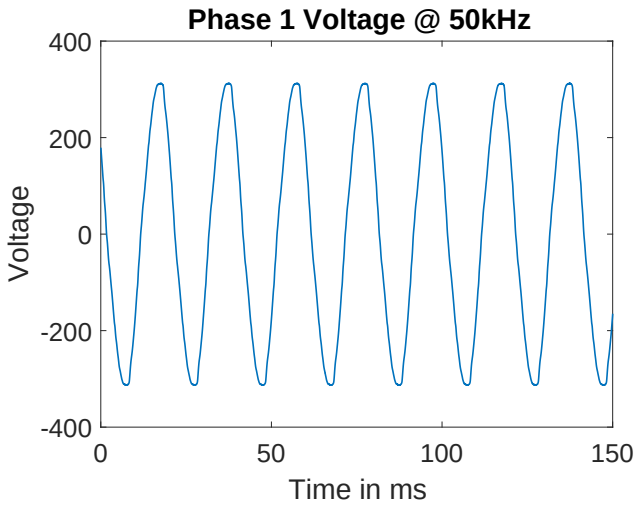


Figure 6.1.1: Voltage signal for phase 1 (peak-to-peak voltage)

6.2 Evaluation at 50 kHz

The voltage and current waveforms acquired by CLEAR for the first phase are shown in Fig. 6.1.1 and Fig. 6.2.1, respectively. Apart from the 120° phase shift, the other two phases show similar waveforms. The graphs show 0.15 sec of mains voltage and current sampled at 50 kHz for a single phase. As expected, the voltage signal is smooth and shows peak-to-peak voltage for the first phase. On the contrary, the current signal shows some underterministic spikes. These spikes are mainly due to switched-mode power supplies which dominate office environments. Due to the presence of these transients, the challenge lies in how to use these underterministic spikes to train disaggregation algorithms.

Event detection is an important step to accurately single out appliances from the aggregate load. These events are detected by observing start-up appliance transients which contain appliance specific unique identifiers. The start-up transients of electric kettle and multitool are shown in Fig. 6.2.2 and Fig. 6.2.3, respectively. Due to higher power requirement of kettle (1800 W), much more current is drawn during start-up and the event is easily visible from the aggregate current curve. On the other hand, multitool consumes much less power (135 W) and hence has a much smaller impact on the aggregate curve, as shown in Fig. 6.2.3. The high probability of simultaneous events due to increased number of electrical appliances in office environments also contribute to disaggregation challenges. Besides 50 kHz data-acquisition, we have collected data at 100 kHz, 150 kHz, 200 kHz, and 250 kHz to check the hardware performance and data handling capabilities of CLEAR. We will

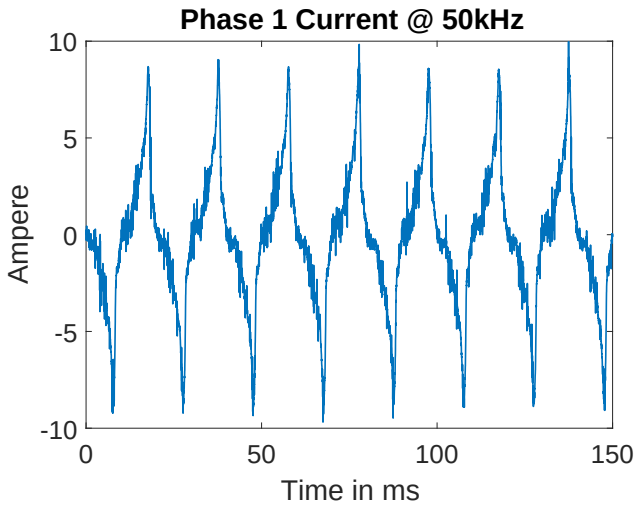


Figure 6.2.1: Current signal for phase 1 (spikes due to presence of SMPS)

utilize these sampling frequencies to analyze any improvement in appliance detection.

At 50 kHz, we obtain 1000 sample points per cycle which is much better than REDD (250 sample points at 15 kHz) and BLUED (200 sample points at 12 kHz) datasets collected in US (60 Hz grid frequency). Similarly, the UK-DALE dataset, which was collected at 50 Hz grid frequency, provides 320 sample points at 16 kHz. For 250 kHz, these sample points increase to 5000 per cycle, which helps to clearly observe even smallest spikes caused by SMPS. As a first step, we have installed CLEAR in our institute to acquire high-resolution energy data. This will lead us to develop accurate disaggregation algorithms for office environments. We are also interested to explore different compression techniques to reduce the storage requirement. Another research direction would be to apply some on-board low level signal processing to facilitate event detection for NILM disaggregation algorithms.

6.3 Evaluation at 250 kHz

Since, to the best of our knowledge, no high-frequency data set representing an office environment exists, we have collected the data using our proposed hardware and will be openly available for the research community. To make a more reasonable contribution for the NILM community, we have also gathered high-frequency ground truth data for developing and testing new disaggregation algorithms. The aggregate data is gathered over a period of 6 months at different sampling frequencies (50 kHz and 250 kHz) to check

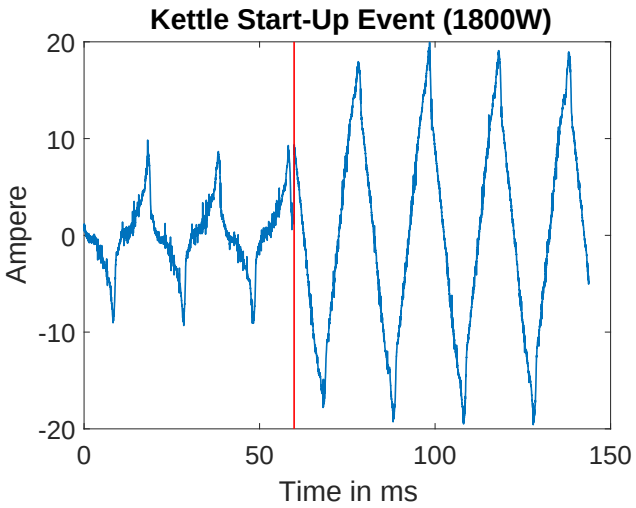


Figure 6.2.2: Start-up transient response of electric kettle

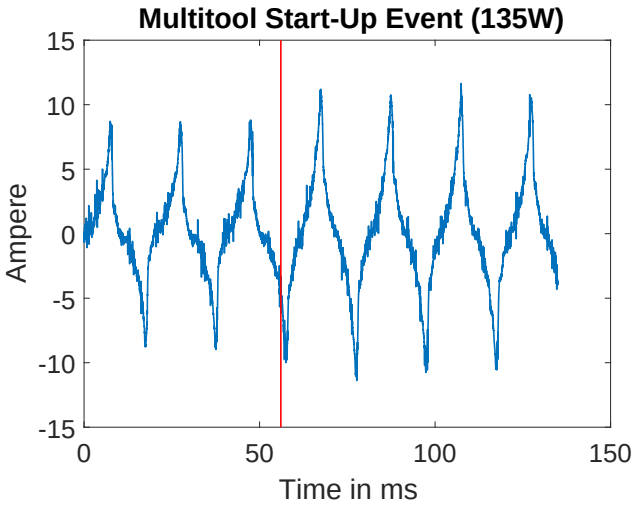


Figure 6.2.3: Start-up transient response of multitool

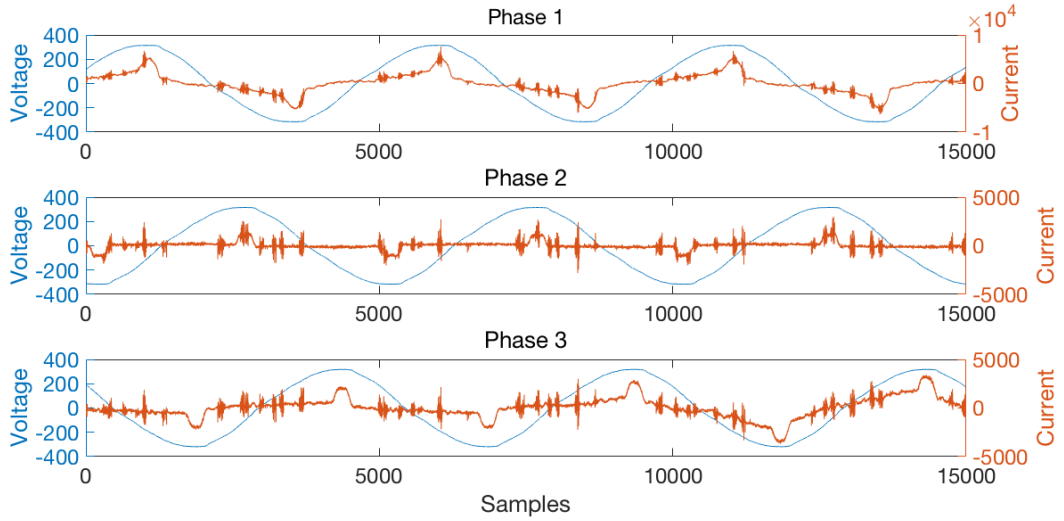


Figure 6.3.1: Voltage and current waveforms (raw values) of the three-phase system at 250 kHz

the performance of our proposed hardware. Apart from increasing the chances to detect multiple events, the higher frequency results in better bandwidth utilization and anomaly detection. The three-phase voltage and current for each phase at 250 kHz are shown in Fig. 6.3.1. Phase 1 has a higher current range (around 9 A) due to more appliances on this phase.

The periodogram in Fig. 6.3.2 shows that most of the signal is centered at 50 Hz, the fundamental frequency. It is a common scenario for periodic signals where the rest of the signal's power is distributed throughout the frequency domain with small peaks at the harmonic frequencies. Although at 250 kHz sampling, the harmonics are theoratically possible upto 125 kHz (Nyquist criterion), but only first few harmonics are shown here as a reference. Similarly, crest factor (ratio of peak to rms current) indicates how extreme the peaks/spikes are in the waveform. A linear sinusoidal waveform has the crest factor of 1.414 and increases with the number of random peaks in the waveform [128]. For our measurements, the higher crest factor indicates the presence substantial number of spikes due to the abundance of SMPS-equipped appliances.

6.3.1 Sampling, Resolution, and Accuracy

The sampling rate (R1) determines the amount of information that can be extracted from the aggregate load. Armel et al. [17] suggest that higher sampling can significantly

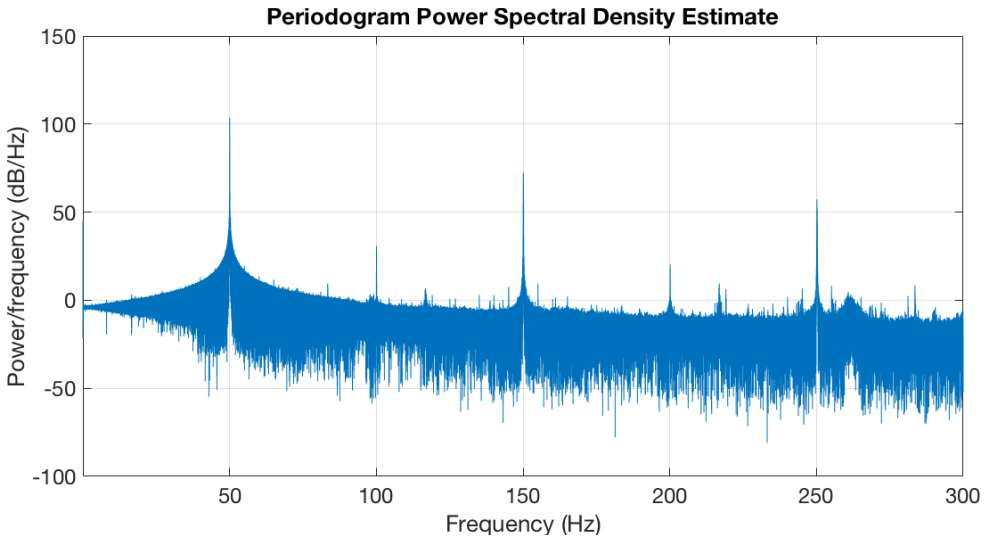


Figure 6.3.2: Fundamental frequecy and harmonics of voltage signal

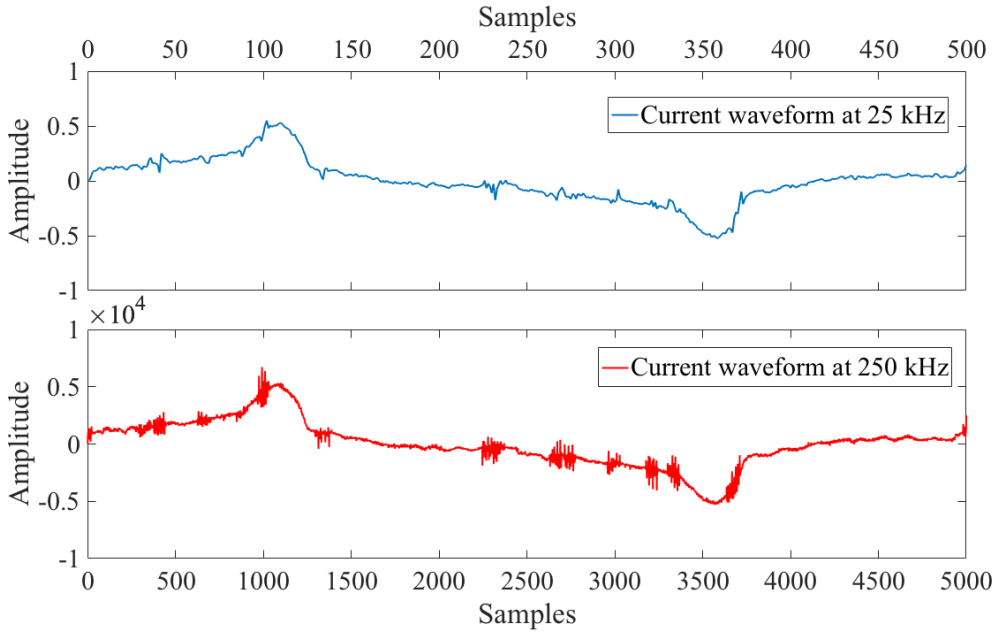


Figure 6.3.3: High sampling rate and resolution can increase the accuracy to detect switching events (indicated by spikes)

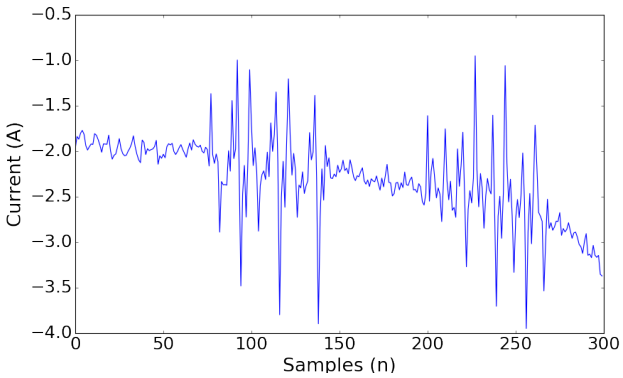


Figure 6.3.4: Two independent events casued by the SMPS equipped office appliances observed at 250 kHz from aggregate load

improve the recognition of both number and type of appliances during the disaggregation process. Less granular data can still disaggregate higher load appliances but the accuracy of identifying lower power appliances will be limited. The typical interesting range is between 10 kHz to 40 kHz; a range providing medium order harmonics and differentiation between 20 to 40 different appliances. Usually, higher energy savings are achieved with accurate measurement system, as precise enough energy breakdown (at appliance level) can be achieved. This also enables us to segregate the power-hungry appliances from the rest of appliances for effective participation in the DSM programs.

The disaggregation accuracy depends on sampling frequency and resolution during the analog to digital conversion. Higher resolution along the y-axis depends upon the ADC bits to determine the quantization levels, where higher quantization levels achieve higher accuracy by reducing uncertainity. Similarly, the x-axis resolution depends on the number of samples captured from the raw voltage and current signals. With 250 kHz sampling, the proposed energy monitor has a wide operation range to capture overall signal behavior while being sensitive enough to cover the minute details (spikes), as shown in Fig. 6.3.4. Theoratically, it is capable of measuring a minimum isolated load of 0.17 W but because of base noise, we consider a minimum of 30 W as potential switching event. This fulfills the requirement of basic office appliances such as laptops and LCDs which have a power consumption of 45 W or more.

6.3.2 Appliance Switching

The higher resolution (R2) also increases the probability to identify simultaneous switching events. The event detection complexity increases with the number of appliances and can pose a major challenge when numerous similar appliances are operating in parallel, a common scenario in an office. A comparison at 25 kHz and 250 kHz sampling frequency for current of phase 1 is shown in Fig. 6.3.3. Single current waveform at 25 kHz (500 sample points), down-sampled from 250 kHz, is compared with the corresponding waveform at 250 kHz (5000 sample points) sampling rate. Similarly, using the same 250 kHz waveform, Fig 6.3.4 shows two switching events caused by SMPS-equipped office appliances. These two switching events happen within a space of 300 samples and are around 50 samples apart from eachother. It can be observed that higher sampling frequency offers more likelihood to detect the start-up transients (indicated here by spikes) of low powered office appliances and minimizes the probability of simultaneous events. Never the less, detecting the appliance associated with each spike is a challenge as each SMPS includes a switching regulator to reduce energy wastage by storing excess energy and utilzing it in next cycle [129].

Most NILM studies follow switch continuity principle (SCP) [21], which assumes that at a given time, only one appliance changes its state (on/off). For office environments, multiple appliance-switching events at an instant are common due to high appliance diversity. The office appliances often work in pair (laptop and LCD) and this assumption can lead to error. A similar situation exists for industrial environment where multiple motors are working together. Fig. 6.3.3 and Fig. 6.3.4 indicate that higher sampling frequencies help better distinguish these switching events (R11). Once an event is detected, these appliance start-up features (load signature) are extracted by the disaggregation algorithms to perform inference and learning.

6.3.3 Reliable and Simultaneous DAQ

Due to the unpredictable nature of appliance usage, one major requirement is to acquire non-stop voltage and current measurement data from each measured phase (R3). For three-phase systems, multiple inputs are required to simultaneously monitor the individual voltage and current streams for each phase. The cost-effective approach proposed in [87] only supports single-phase measurement. This approach lacks the multiple input feature

and hence not practical for three-phase systems. The proposed design is capable of handling six measurement channels (R4) using Analog Devices (AD7656A) chip.

To ensure reliable data handling (R7), a Lattice XO2 7000HC FPGA chip is used to collect and forward the data to the single-board PC (LattePanda) via USB (FTDI-232HL) chip. Here data is converted into the appropriate format (HDF5) and forwarded to the storage server. To avoid any data loss due to network connectivity issues, a mass storage device is attached as backup with the single-board PC. During its six month operation, only one major error was observed which resulted in data loss of about 2 hours.

6.3.4 Scalability and Interoperability

From prototype to actual deployment, one challenge regarding energy monitoring is its interoperability with existing infrastructure and how well it can expand to meet future requirements. In terms of NILM, a system can be termed as scalable if it is able to accurately detect a newly added appliance category. To detect newly added appliances, we rely on the turn on/off transients (Fig. 6.3.4) caused by appliance switching. Once these events are detected, appliance classification algorithms are used to point out the appliance type.

Scalability and interoperability usually go hand in hand and require well-known standard measurement and data formats. Our hardware, is capable of operating under different voltage and current rating. The customizable features and replaceable external components (CTs, single-board PC) make the system deployment user friendly (R13) and extendable. Due to its independent DAQ and persistence, our hardware can scale well with multiple units operating in parallel (R5, R6) for different floors of a building or for different buildings altogether, only requiring basic internet connectivity.

6.3.5 Data Processing, Storage, and Privacy

Disaggregation algorithms require high-frequency and high-resolution data, especially with large number of low-powered appliances present in an office environment. Capturing this huge amount of data for three phases requires special emphasis for error-free simultaneous data collection. Two different approaches were used to collect and process the data. At 50 kHz, the data was converted to HDF5 format, compressed and transferred to storage

cluster in chunks. Due to limited resources of single-board PC, the raw data at 250 kHz was directly transferred in chunks (50 min) to the storage cluster and then converted to HDF5 format and compressed. In order to perform load disaggregation, data needs to be securely stored in an appropriate format (R10). We utilized HDF5 as it is a well-established and widely accepted file storage format in the NILM community (R9).

Besides major advantages, data confidentiality is one main concern regarding NILM as energy usage patterns provide a detailed and in-depth analysis about consumer's life style. To apprehend these concerns, we have provided multiple storage options to the consumers (R8). Our monitoring system is capable of storing both on- and off-site data storage (cloud). So far, the on-site data storage is only used as a backup but the future designs will incorporate on-board NILM algorithms to perform disaggregation locally.

CHAPTER 7

NILM Event Detection

In today's digital age, the ever growing demand for the collection and analysis of big data has steered many techniques for discovering certain patterns within the data. The prime objective is to extract useful information for proper interpretation and analysis of the data under study. In information processing, this knowledge helps us develop different data trends for condition monitoring and later identification of minor deviations for anomaly detection [130]. Although these minor deviations or micro-events vary for different applications, they form the basis of any data driven analysis approach.

For NILM, the event-based detection approaches can be further sub-divided into two categories. The first category consists of macro-events which occupy from one up to multiple AC cycles. To detect such inter-cycle events, low-sampling frequency (in seconds/Hertz) is adequate. The low-sampling frequency usually works well for a household or a small building, where limited appliances are operating (low density) and each appliance has very distinctive features (high diversity) for effective classification of these appliances from aggregate load. For the second category, several micro-events occur within an AC cycle. To detect such small intra-cycle events, high-sampling frequency (in milliseconds/kHz) is essential. Hence, for the commercial and industrial environments such as an office, a supermarket, or a factory with high appliance density and low appliance diversity, high-frequency sampling is an absolute necessity. Without high-frequency sampling, event detection is not accurate as simultaneous switching events are almost impossible to avoid.

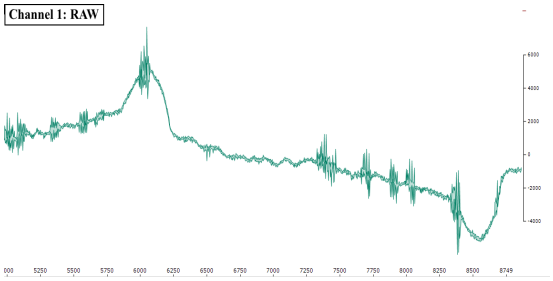
During most NILM studies, the switch continuity principle (SCP) [21] is assumed by the researchers. According to SCP, only a single appliance changes its state (on/off) at

a given time. This assumption is not valid when multiple low-powered appliances are operating in parallel. For instance, consider an environment where several multi-state appliances such as freezers or motors with multiple sub-machine components (e.g., heating, lighting, controllers, etc.), are operating simultaneously. Using low-sampling frequency will certainly merge some micro-events and with multiple micro-events stacked together, the appliance classification error is likely to increase.

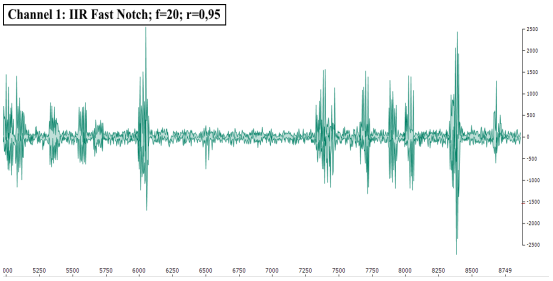
This chapter presents a Hilbert-Huang transform based event detection approach to detect micro-events from an office environment with an abundance of SMPS-equipped appliances. The challenge lies in accurately determining the switching events caused by these low-powered supplies. For this purpose, we perform our proposed event detection algorithm on high-frequency aggregate energy data from BLOND [19]. Unlike conventional analysis methods, most of the real world energy data represents non-linear and non-stationary systems. This limits the application of Fourier transform which is mostly suited for analyzing linear systems with stationary signals [102]. The Hilbert-Huang transform approach provides time-frequency-energy analysis of the data and helps estimate the instantaneous energy, an important requirement for NILM. We also discuss why the Hilbert-Huang transform based approach is preferred instead of a discrete Fourier transform and a discrete wavelet transform.

7.1 SignalPlant-based NILM Event Detection

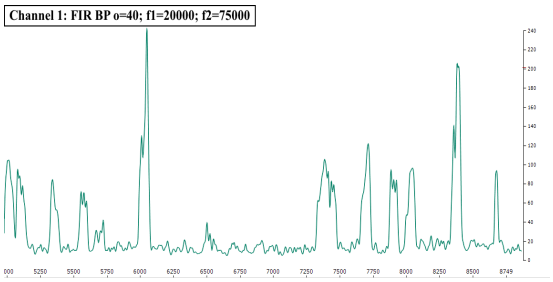
To detect events from aggregate load, we first employed some already existing tools to check how well we can detect these events, if at all possible. Initially, we utilized a freely available statistical event detector tool called SignalPlant [131]. SignalPlant was primarily developed for bio-medical signals (e.g., ECG, EEG, etc.), where a lot of low-powered waveforms are expected and it is very challenging to remove the base noise from the signal. We used a combination of built-in filters such as infinite impulse response (IIR) and finite impulse response (FIR) to remove the unwanted signal noise and obtain smooth curves for the current waveform. Later, we utilized the built-in threshold detector to identify these events from the aggregate load. The event detection process using the SignalPlant is shown in the Fig. 7.1.1. We tested for a small window to determine the number of events from the aggregate load. The initial results indicate that the aggregate data acquired by CLEAR can be accurately utilized to precisely identify the low-powered operating appliances in



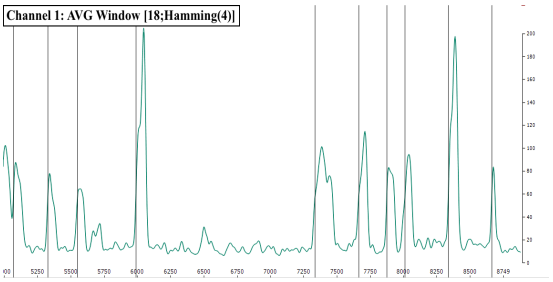
(a) Raw samples from office at 250 kHz



(b) Samples after application of IIR filter



(c) Samples after application of FIR filter



(d) Signal smoothing and event detection (vertical black lines)

Figure 7.1.1: Event detection process using SignalPlant

a building. This success led us to develop our own event detection algorithm using the Hilbert-Huang transform.

7.2 Event Detection Techniques

In this section, we will discuss the two common methods for detecting changes in information processing systems, namely the discrete Fourier transform and the discrete wavelet transform. In the end, we will introduce the Hilbert-Huang transform, its advantages, and disadvantages followed by the event detection results.

7.2.1 Discrete Fourier Transform (DFT)

In general, the DFT is a transformation of a time domain signal into the frequency domain [132]. This is done by interpreting the space of all possible complex signals $S = \mathbb{C}^{N+1}$ of length $N + 1$ as a vector space, and defining the dot product between two elements of this vector space $a, b \in S$ as $a \cdot b = \sum_{k=0}^{N} a[k]\overline{b[k]}$. It turns out that using this definition of the

dot product, the following set of signals forms an orthonormal basis of S.

$$B := \{b_n[k] = \exp(\frac{2\pi i}{N+1}kn) | n \in \{0; ..; N\}\}$$

Using this basis, the DFT is nothing more than a transformation from the natural basis to the basis B. The coefficient x_n of the base-vector b_n can then be interpreted as the prevalence of the frequency $\frac{n}{N+1}$ in the signal $s \in S$. As the coefficients x_n can be complex-valued (even if $s \in \mathbb{R}$), they are usually not directly displayed, but instead as $20 \log_{10} |x_n|$dB.

Two parameters, the sampling frequency f_s and the number of used samples $N + 1$, are required when applying the DFT to actual signals. Both parameters influence the set of frequencies analyzed when applying the DFT. The lowest analyzed frequency, at $n = 0$, is always 0, i.e., a constant signal. The highest frequency, according to the Nyquist-Shannon sampling theorem, will be $\frac{f_s}{2}$, which when equating with $\frac{n}{N+1}$ yields $n = \frac{f_s \cdot (N+1)}{2}$. As all remaining analyzed frequencies linearly interpolate between the lowest and highest frequency, the final set of analyzed frequencies is:

$$F = \{f_n = \frac{f_s n}{N+1} \mid n \in \{0, ..., \frac{N+1}{2}\}\}$$

The main disadvantage of the DFT lies within the high difficulty of detecting when a certain instantaneous frequency is observed, as this would require interpretation of the phase spectrum of a signal. In other words, the DFT is not suitable for analyzing non-stationary signals. Unfortunately, most signals acquired from real word physical processes will have at least some non-stationary content. The main advantages of the DFT include the computational efficiency and strong mathematical proofs.

7.2.2 Discrete Wavelet Transform (DWT)

The DWT also uses a transformation to an orthonormal basis to draw conclusions about an input signal. In the case of the DWT, this new basis consists of wavelets. Usually, one mother-wavelet ψ is used to construct all base-wavelets by dilating and translating the mother-wavelet. This way a wavelet can be dilated by a factor w and translated by τ in the following way, to generate the wavelet [133].

$$\psi_{w,\tau}(t) = \psi(\frac{t-\tau}{w})/\sqrt{w}$$

Another key feature of wavelets is that their features are localized [133], i.e., the wavelets are almost always zero except for one short period of time in which they exhibit a characteristic signal-shape. This time-restricted nature has the effect that a basis-coefficient $x_{w,\tau}$ for a wavelet $\psi_{w,\tau}$ after a transformation of a signal $s \in S$ not only yields information concerning the frequency $f \backsim 1/w$ of s, but also about the location τ within s, where this frequency is prevalent. It should be noted that when using the DWT, the period and not its inverse (frequency), is implicitly calculated. This is because one cannot directly choose the frequency f of any given wavelet. Instead, the width of the wavelet $w \backsim 1/f$ is chosen. Hence, the ordinate of wavelet-spectrograms is usually a width and not a frequency.

When applying the DWT to signal measurements, three parameters are required. First, the range of the wavelet widths is chosen, which implicitly changes the range of analyzed frequencies. Second, the number of samples used for the DWT are chosen. But unlike the DFT, this does not directly influence the highest analyzable frequency. It only gives an upper bound to the range of the wavelet widths, as matching a wavelet longer than the base signal would make little sense. The third parameter is the sampling frequency f_s, which directly correlates to the frequencies corresponding to the wavelets.

The main disadvantage of the DWT is its inability to determine the instantaneous frequency of the signal at every point in time. This discreteness is the main weakness of the DWT. Furthermore, DWT is suitable for *non-stationary* and *linear* signals, whereas most of the naturally occurring signals are *non-stationary* and *non-linear* signals. On the bright side, the DWT is capable of detecting a change in frequency over time and does so at a reasonable computational complexity.

7.2.3 Hilbert-Huang Transform (HHT)

The HHT, first described by Huang et al. [134], takes a different approach than the previous two methods. HHT provides an empirical method to examine the time-series data for time-frequency-energy analysis. Instead of dictating what the orthogonal basis vectors should look like, it empirically generates its own set of orthogonal basis vectors from a given base-signal through empirical mode decomposition (EMD). The resulting basis vectors are then called intrinsic mode functions (IMFs). Thereafter, the Hilbert transform is applied to the IMFs to analyze the instantaneous energy and frequency of the IMFs at any point in time. Contrary to the other methods explained so far, the HHT was explicitly developed to

deal with *non-stationary* and *non-linear* signals. The HHT method is explained in detail below.

Empirical Mode Decomposition

The first step in the two-step Hilbert-Huang transform is the EMD. EMD essentially splits the given signal into multiple IMFs.

Intrinsic Mode Function – IMFs decompose the times-series data into a set of functions based on the different frequencies. So, IMFs provide a pre-processing step to remove the already known unwanted noise from any signal. Any signal is an intrinsic mode function if it satisfies two conditions:

- The number of *maxima* and *minima* together must be equal to the number of zero-crossings or atmost differ by one.
- The mean of the local envelopes defined by its local *maxima* and *minima* should be zero at all times to obtain meaningful instantaneous frequency.

A graphical representation for these conditions is shown in Fig. 7.2.1. In the first example, the region around x=1900 has a number of local $extrema$ with value below zero. For those *maxima* and *minima*, the zero crossings are not balanced and hence, the total number of zero crossings will not match the number of $extrema$. Similarly, the second example seems to have enough zero crossings, but the upper envelope has a greater magnitude than the lower envelope. A proper IMF satisfying both conditions is shown in the last example. If the two conditions are satisfied, it is ensured that at every point in time an instantaneous frequency of the signal can be defined and computed by applying the Hilbert transform.

The EMD algorithm for NILM – As energy data acquired from real world measurements are unlikely to fulfill the strict IMF-conditions, it needs to be transformed first. This can be achieved by applying the EMD algorithm, as shown in Fig. 7.2.2. The EMD algorithm does not transform the energy data into a single IMF, but instead splits it into several IMFs. This corresponds to separating physical effects on different time-scales.

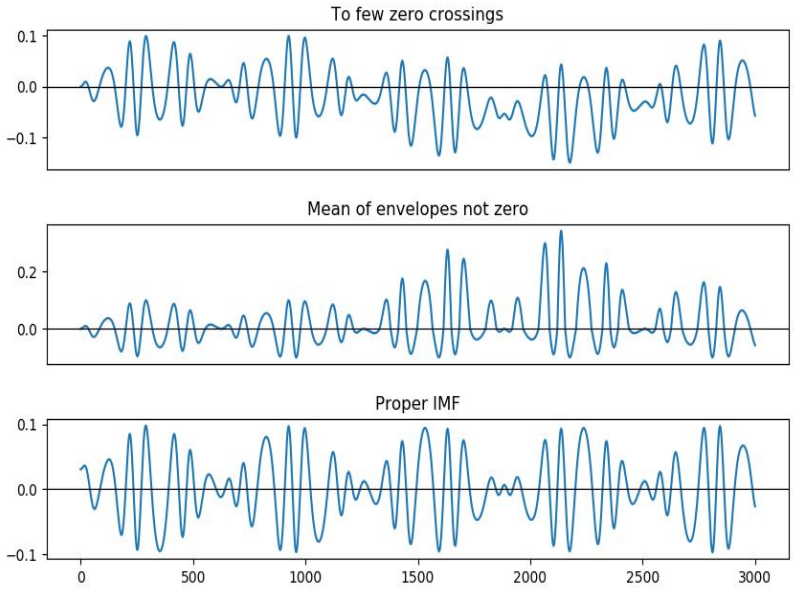


Figure 7.2.1: Counterexamples for IMF conditions

The algorithm begins with the input data. After the first IMF is obtained, it is subtracted from the input and the algorithm continues to produce the next IMF from the result of this subtraction. The first IMF contains information regarding the highest frequency components present in the measured data (often the first IMF contains high-frequency noise). The following IMFs contain information about effects on larger timescales. The process continues until the result of the aforementioned subtraction contains no more local extrema, but instead a monotonic function. This is called the *residue*. Mathematically [104], we denote the input signal as $X(t)$, the kth IMF as c_k and the final residue as r. The decomposition can then be written as

$$X(t) = \sum_{k=1}^{n} c_k + r$$

The construction of IMFs is carried out by an iterative process called *sifting*. An intermediate IMF that still needs more steps of sifting is denoted as h_{ki}, where k is the IMF about to be constructed and i is the number of sifting iterations already performed. In the same fashion, the mean of the upper and the lower envelope of the signal is denoted as m_{ki}. One step of sifting can be written as:

$$h_{k(i+1)} = h_{ki} - m_{ki}$$

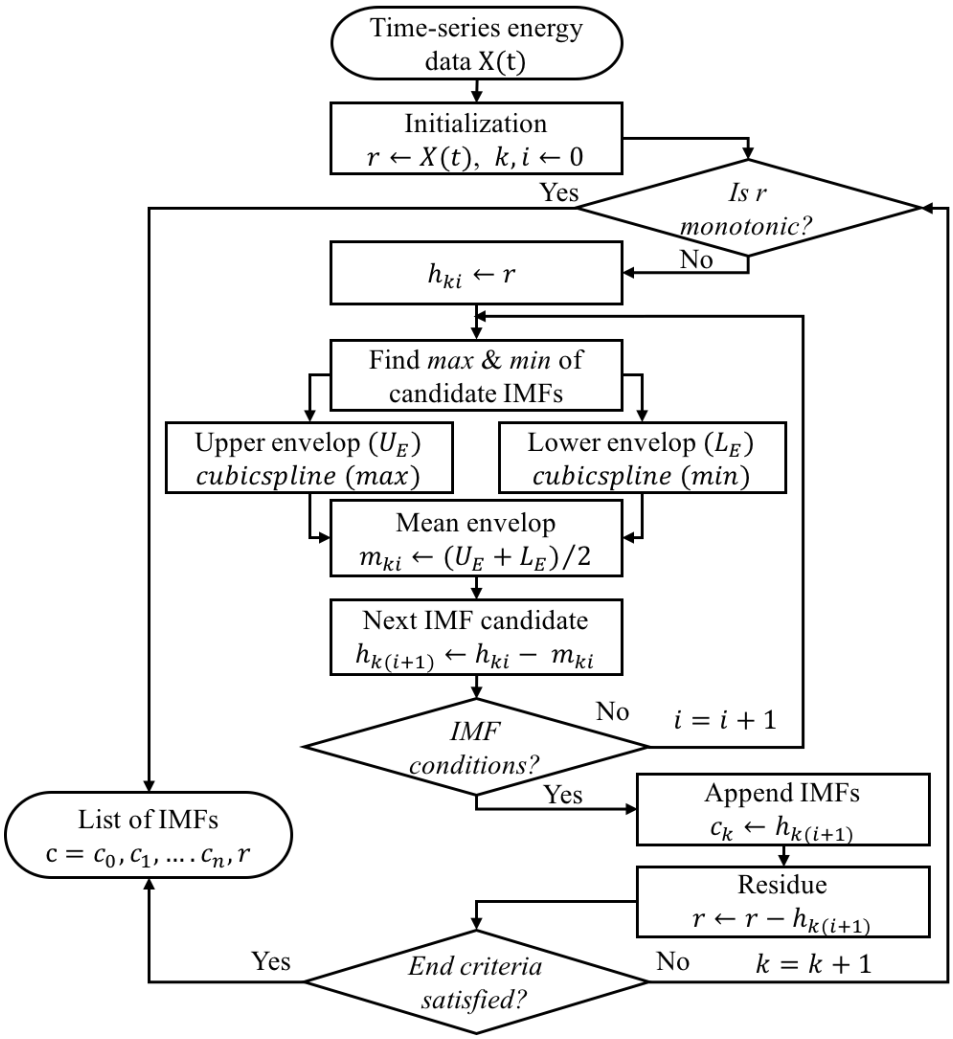


Figure 7.2.2: EMD algorithm for NILM event detection

First, all the local *maxima* (*max*) and *minima* (*min*), as indicated in Fig. 7.2.2, have to be identified. Then an upper and a lower envelope is created by fitting a cubic spline through all the *max* and *min*, respectively. In the first iteration, the mean of the upper and lower envelope is subtracted from the input data. The sifting process is repeated until the aforementioned IMF conditions are met. So in total, the EMD algorithm consists of an outer loop, which produces one IMF per iteration and an inner loop, in which multiple iterations of sifting are performed to obtain a candidate IMF.

Stoppage criteria – An important decision that has to be made is when to stop the sifting process. Ideally, the process should stop when data series satisfying IMF conditions are produced. In practice, it is not easy to end the sifting process because the second condition (mean value of upper and lower envelope must be zero) cannot be easily fulfilled due to numerical instabilities of calculations with floating-point numbers. Similarly, too many sifting steps could obliterate amplitude fluctuations that carry important information about the underlying process [134].

A more practical stopping criteria would require the mean of the envelopes to be close to zero, but using a fixed threshold for this condition limits different amplitudes the input signal might have over time. A better approach is to leverage the fact, that sifting is a converging process, i.e., in each iteration of sifting, the changes made to the data should be smaller than in the previous iteration. Therefore, Huang [135] proposed the standard deviation between the results of two consecutive sifting steps as stoppage criterion:

$$SD_k = \sum_{t=0}^{T} \frac{|h_{k(i-1)}(t) - h_{ki}(t)|^2}{h^2_{k(i-1)}(t)}$$

Sifting stops if the value calculated for SD is below a predefined threshold value. This value is dependent on the length of the input data T.

Hilbert Transform

The second step in the HHT-based event detection method is to take the computed IMFs and extract their instantaneous amplitude and frequency. It is achieved by first expanding the IMFs to the complex number plane in such a way that they become analytic, in other words, their negative frequencies become zero. For this purpose, the Hilbert transform is

applied to take in a signal $x[k]$ and output a signal $y[k]$, such that the signal $x[k] + iy[k]$ is analytic. It turns out that the transformation to achieve this is a convolution with the function $\frac{1}{\pi t}$. So, the Hilbert transform can be written as:

$$H(u)(t) = (\frac{1}{\pi t} * u)(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{u(\tau)}{t - \tau} d\tau$$

According to Huang et al. [134], the result $X_j(t)$ of the sum of the jth IMF $x_j(t)$ and its Hilbert transformation $iH(x_j)(t)$ can be written as:

$$\begin{aligned} X_j(t) &= x_j(t) + iH(x_j)(t) \\ &= a_j(t) \exp(i\phi_j) = a_j(t) \exp(i \int \omega_j(t) dt) \end{aligned}$$

7.3 HHT-based NILM Event Detection

In this section, we will explain the implementation and analysis work in detail. As we are interested in detecting and isolating micro-events, we will be using the BLOND energy dataset [19], mainly due to a sufficiently high sampling-rate. The event-detection techniques discussed here were implemented in Python, and are built to read `.hdf5` files from BLOND.

7.3.1 Building Level Office eNvironment Dataset (BLOND)

In contrast to other existing NILM datasets, which mostly cover residential environments, BLOND contains measurements for an office building in Germany using CLEAR and MEDAL. The appliances present in an office environment are different from typical households. Most appliances have low power-consumption (e.g., when compared to an oven) and there exist multiple appliances of the same type such as laptops, PCs, and LEDs. To detect these appliances based on the patterns they cause in the aggregated measurements, it is necessary to increase the sampling frequency. BLOND also provides an excellent environment to observe the effect of simultaneous events, as a number of low-powered SMPS-equipped appliances are operating on each phase. Due to abundance of these low-powered appliances, the probability of simultaneous events is quite high.

BLOND-250 uses a high sampling frequency of 250 kHz for 50 days of aggregated

recordings using CLEAR [19]. Similarly, recordings at individual power plugs are also available with a sampling frequency of 50 kHz using MEDAL. For this study, we are not considering BLOND-50 which contains aggregate data with a sampling frequency of 50 kHz at the mains and 6.4 kHz for the individual power plugs and covers 213 days of measurements. Even at 50 kHz, some of the low-powered events from SMPS-equipped appliances get merged together (see 7.3.5). All files are saved in the $.hdf5$ format, which is a common format for storing large amounts of scientific data. At 250 kHz, each aggregate load file (CLEAR) has a file size of around 220 MB and contains 6 channels, corresponding to the voltage and current of each of the three phases and represents 2 minutes worth of data. All the experiments were conducted on 250 kHz data.

7.3.2 Preliminary Data Analysis

HHT-based Event Detector – The proposed event detector first applies the Hilbert-Huang transform to the input signal, and then detects events in that signal by analyzing the output of the HHT. To detect the events, the algorithm uses the sum of instantaneous amplitudes of some chosen IMFs to generate a signal-energy-curve. On this signal-energy-curve, a standard peak-detection-algorithm is applied.

As soon as the optimal parameters for detecting events in a specific dataset have been manually optimized, they can be applied to an entire file at once. In the end, a CSV-file is created containing the outcome of the HHT event detection. Each row contains individually detected event with following information:

1. Event number (detected for input file)
2. Start sample (to mark start of event)
3. End sample (to mark end of event)
4. Event duration (in number of samples)
5. Overlap: a boolean value describing if this event was considered to consist of two overlapping events by the peak-detection.
6. Event-Type: ’1’ if there is a single non-overlaping event (’2’ otherwise).

Table 7.3.1: Parameters of Peak-Detection

Parameter	Description
Height (h)	Peaks smaller than this value are ignored.
Prominence (p)	Prominence is a relative parameter, that determines that a peak has to be larger that its neighboring peaks. It can be used to suppress small peaks that lie on the shoulder of larger peaks.
Width (w)	Peaks narrower than this value are ignored.
Distance (d)	The minimal distance between two peaks. If two peaks are too close, the smaller one will be rejected.
Rel_height (rh)	Used to define at which height of the peak the width is measured. 1.0 means measure the width at the absolute lowest point, which will lead to very broad event-borders.

Parameters of the Peak Detection – The choice of IMFs is an important part of the event detection algorithm because the next steps only use the sum of instantaneous amplitudes of specific IMFs to detect event peaks. The parameters utilized for the peak-detection are shown in Table 7.3.1. The Python package `scipy` is used for peak-detection [136].

Instantaneous frequencies – Analyzing the instantaneous frequencies of IMFs can reveal which of them might correspond to actual events. Generally, the instantaneous frequency of the IMF becomes less variable (more smooth) with higher IMF number (lower frequency), as seen in Fig. 7.3.2. Hence, to detect micro-events, one has to primarily use the low-numbered IMFs, because micro-events tend to have a high-frequency. At the same time, due to a high noise probability, we should also avoid the lowest IMFs. Using the above criteria and visual inspection of IMFs with 250 kHz data, we selected $IMF1$ and $IMF2$ as candidates for event detection, as they feature both high, and fairly narrow-banded frequencies. $IMF0$ was not utilized due to high presence of noise.

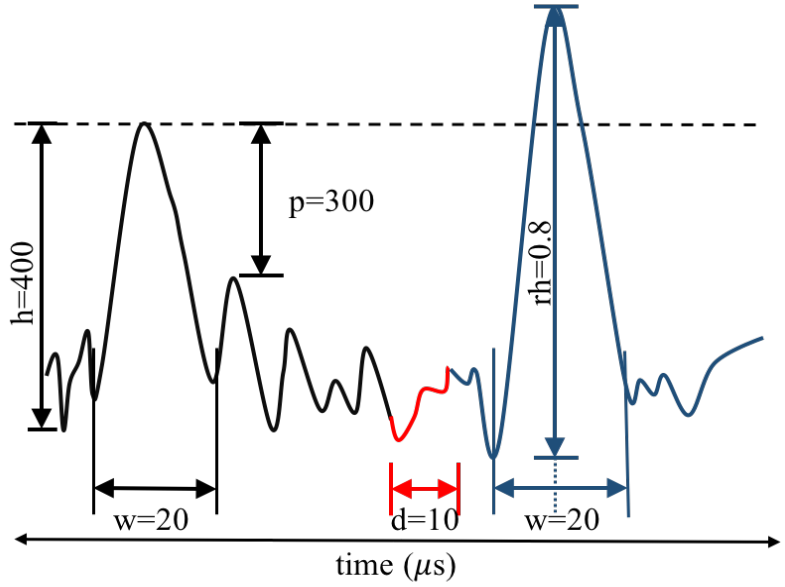


Figure 7.3.1: Visual representation of parameters listed in Table 7.3.1

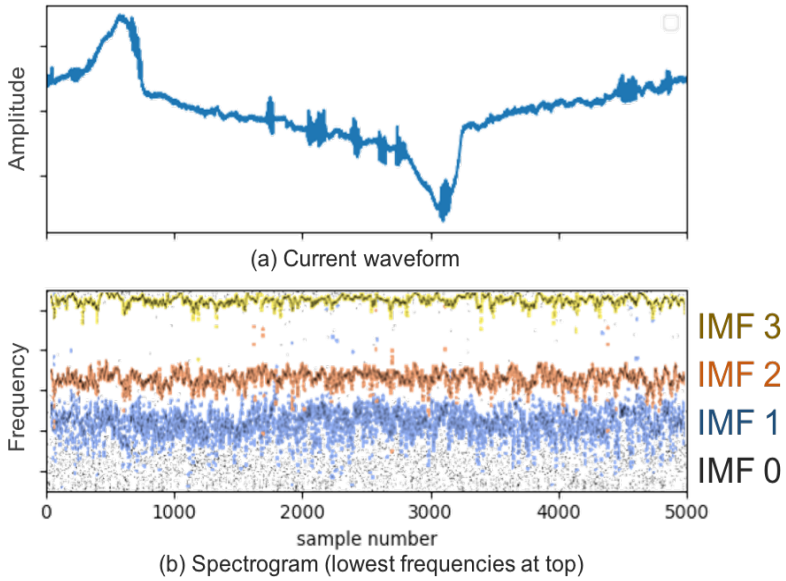


Figure 7.3.2: IMFs generated using EMD algorithm on aggregate data from CLEAR (a) input, (b) spectrogram showing $IMF0$ (black), $IMF1$ (blue), $IMF2$ (red), $IMF3$ (yellow)

7.3.3 Analysis of Micro-Bursts

The output of our HHT event-detector algorithm is shown in Fig. 7.3.3. The sum of amplitudes of the chosen IMFs are displayed. Similarly, the boundaries of the detected peaks are used to mark the detected events in the original data. The amplitude of the sum of IMFs can be up to 10 times higher than their surrounding background noise. Comparing this result to the preliminary result of the DWT, the HHT not only extracts the events from the original signal, but also achieves a good signal-to-noise ratio (about a factor of 3x higher).

In general, the proposed event detector can reliably detect the position and length of events. However, the user has to define what actually classifies as an event. For example, the last detected event in Fig. 7.3.3 might not be considered a real event because it is too short. Some fine-tuning can be done by adjusting the parameters defined in Table 7.3.1. When adjusting these parameters, the user should keep in mind that this modification can also cause some peaks to be suppressed, which actually do correspond to real events. The rationale for only including $IMF1$ and $IMF2$ in the sum as input to the peak detection algorithm is as follows:

- When including high numbered IMFs in the sum, the peaks tend to get very broad, up to the point where entire periods of the mains-power-frequency will be considered as an event.
- Similarly, the lowest IMF ($IMF0$) contains a lot of noise at 250 kHz, which leads to a noisy input to the peak-detection, resulting in more false positives.

It turns out that using $IMF1$ and $IMF2$ for further processing is a good compromise. Incidentally, the choice of not including/calculating higher IMFs also has a positive impact on processing speed and overall processing time.

7.3.4 Empirical Evaluation of BLOND Energy Dataset

As the micro-events considered in this work have not been analyzed before and also the BLOND dataset is not labeled with respect to those, it is not possible to evaluate the actual performance of the event-detection implemented. But it is possible to provide some

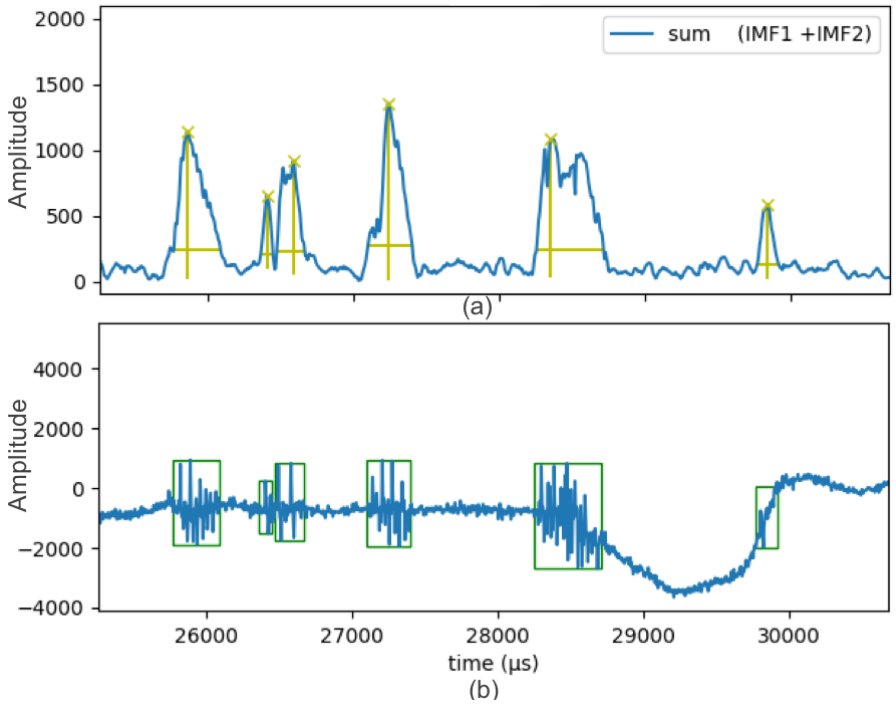


Figure 7.3.3: Input ($IMF1 + IMF2$) to find (a) peaks and (b) detected events

empirical results gained when applying the algorithm to sample files from the dataset. For this section, we consider the following five HDF5 sample files.
Sample 1: (`clear-2017-06-21T06-07-20.122303T+0200-0028508`)
Sample 2: (`clear-2017-06-21T07-51-51.046155T+0200-0028560`)
Sample 3: (`clear-2017-06-21T13-01-22.636113T+0200-0028714`)
Sample 4: (`clear-2017-05-16T11-39-20.114921T+0200-0002881`)
Sample 5: (`clear-2017-06-03T16-34-36.724819T+0200-0015924`)
Sample 1 to 3 were chosen because they cover different times of the same day. Sample 1 covers 2 minutes of measurement in the early morning at around 6 o'clock, where power-consumption is not altered much by activity of the researchers in the building. Sample 2 covers two minutes at around 8 o'clock in the morning. Here the power-consumption is changed a little due to someone switching on a device. This was judged by looking at the daily summary provided by BLOND. Sample 3 was recorded on the same day at 13 o'clock, when power-consumption is high. Sample 4 and 5 were additionally chosen randomly from two different days to compare the results with sample 1 to 3.

The average event-rate in events per 100 ms for above five files is shown in Fig. 7.3.4.

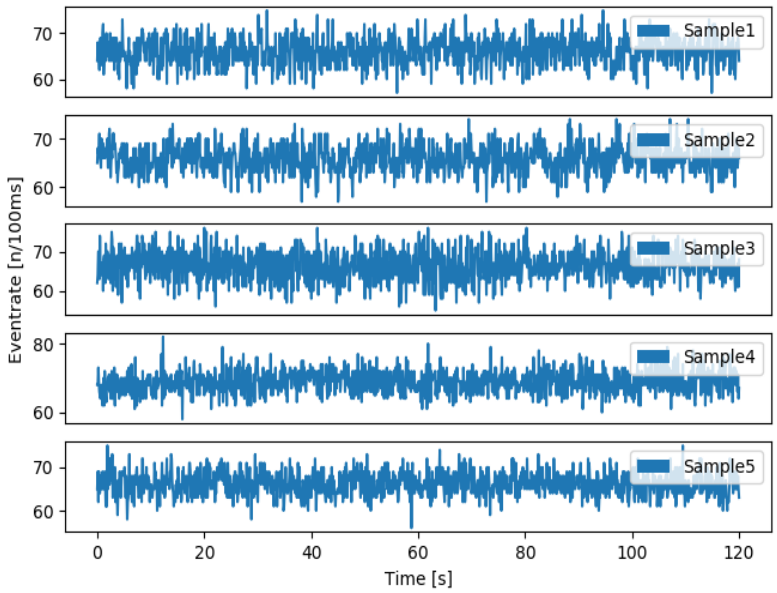


Figure 7.3.4: Event rate for selected sample data

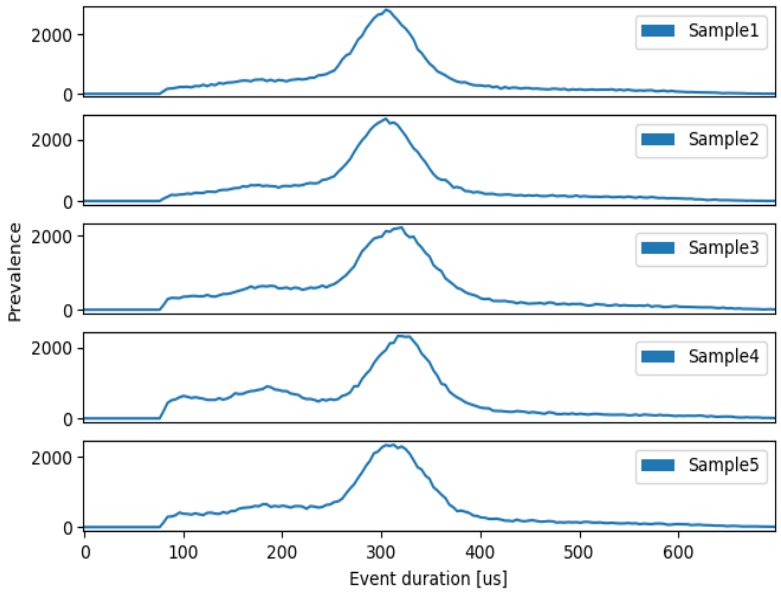


Figure 7.3.5: Event-histogram based on event-duration

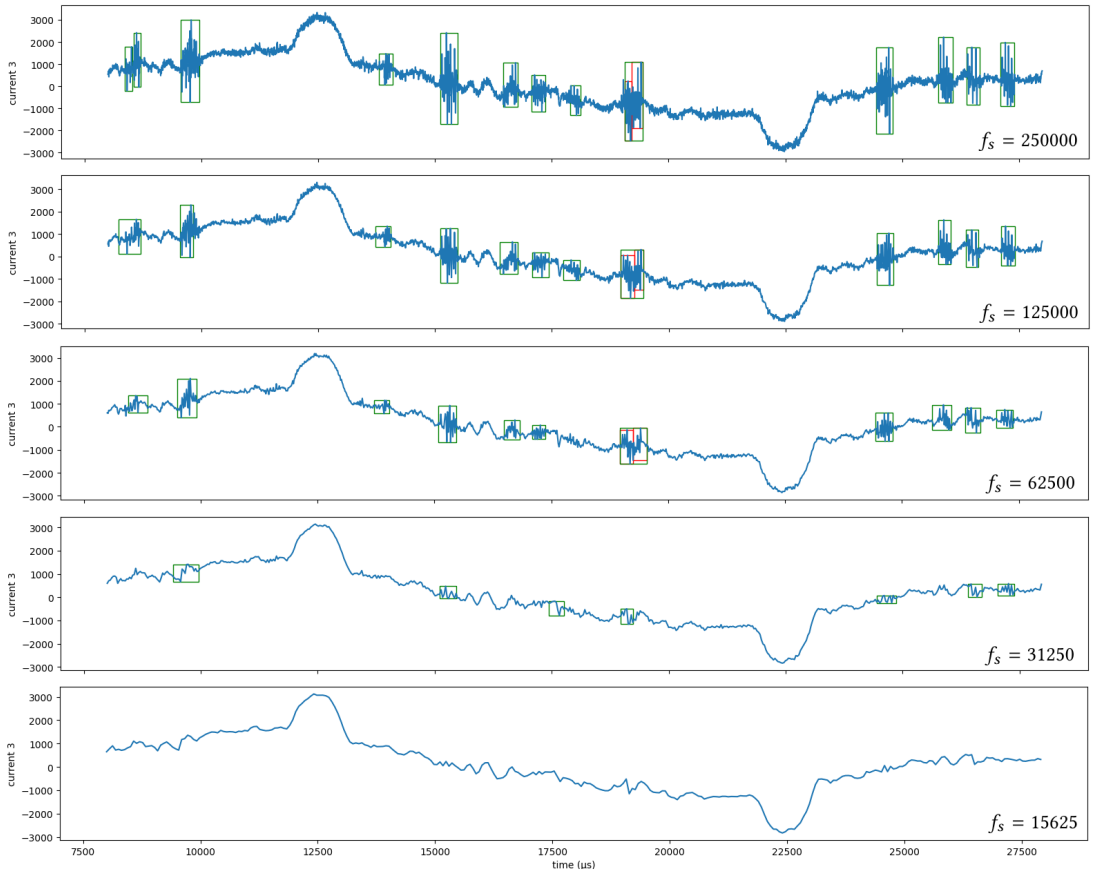


Figure 7.3.6: Event detection at different sampling rates using Hilbert-Huang transform

Although it seems reasonable to expect more events during busy office hours the comparison of the first three samples, the samples differ only slightly.

Similarly, Fig. 7.3.5 shows the histogram of the event-duration. Here, a slight difference is noticeable during the office times. There seem to be slightly more short events with a duration of around 180 µs as compared to night time. This especially applies to sample 4, which contains data from 11:39 AM. This is essentially due to the fact that SMPS-equipped appliances show more rapid charging-discharging outside the stand-by mode. In all samples, most events have a duration of around 300 µs which represent a typical SMPS-equipped appliance. Similarly, there are no events that are shorter than 80 µs. This is due to the peak-detection algorithm, which is set to a minimal peak-width of 20 samples and one sample covers 4 µs.

7.3.5 Effects of Reduced Sampling Frequency

To check the effect of the sampling frequency on event detection, we downsampled the aggregate data at four different sampling frequencies of $f_s = [125,000; 62,500; 31,250; 15,625]$, as shown in Fig. 7.3.6. We observed that as the sampling frequency decreases, events can still be detected, however, some events got coalesced/merged together and became indistinguishable. For different sampling rates, the parameters for the event-detection (see 7.3.2) had to be adjusted manually. Interestingly, with the decrease of the sampling frequency ($f_s = [125,000; 62,500]$), the event-detector had to be adjusted to include $IMF0$ and $IMF1$ to detect events as now they contain the useful information regarding SMPS-equipped appliances. So, by using minor adjustments with the HHT event detector algorithm, we can accurately detect events from multiple frequency ranges.

It can be observed that a high sampling frequency is beneficial for detecting micro-events in current measurements. However, the sampling frequency below 250 kHz might still be sufficient in some cases. Both performance of the algorithm and visual inspection of events in general, begin to quickly degrade at sampling frequencies below 62.5kHz. The main reason for this lies not in the HHT-based method itself, but in the fact that micro-events from an office environment have a certain frequency, thereby putting a lower bound on the sampling frequency.

7.3.6 Runtime Considerations

Although the proposed EMD algorithm can accurately detect events from complex environments, it is computationally expensive. To improve the runtime, some theoretical considerations are discussed here to identfy the bottlenecks.

- The runtime of the algorithm should increase linearly with the number of input data-points
- The runtime of the algorithm should increase linearly with the number of IMFs to be calculated
- The number of sifting steps should be constant with respect to the number of input data-points

Table 7.3.2: Runtime measurements (in s) for EMD algorithm

No. of samples	250,000			500,000			1,000,000		
	t_{IMF}	n_{sift}	t/n_{sift}	t_{IMF}	n_{sift}	t/n_{sift}	t_{IMF}	n_{sift}	t/n_{sift}
$IMF0$	12.1	23	0.53	20.7	18	1.15	34.4	14	2.46
$IMF1$	11.9	26	0.46	49.9	51	0.98	65.1	31	2.10
$IMF2$	15.1	36	0.42	27.8	30	0.93	45.1	23	1.96
$IMF3$	15.0	37	0.41	22.3	25	0.89	51.7	28	1.85
$IMF4$	14.1	35	0.40	31.0	36	0.86	71.9	35	2.05
$IMF5$	16.8	43	0.39	29.5	35	0.84	52.4	30	1.75
$IMF6$	14.4	36	0.40	31.0	37	0.84	79.9	47	1.70
$IMF7$	22.0	57	0.39	123.0	146	0.84	96.5	57	1.69
$IMF8$	18.9	49	0.39	39.1	46	0.85	97.7	58	1.68
Total	140.3			374.2			594.6		
Mean		38	0.42		47.1	0.91		35.9	1.92

When implementing the EMD-algorithm naively and choosing a fixed threshold for the stoppage criteria (see section 7.2.3), the sifting process depends on the length of the input, due to the definition of the stoppage criteria.

Runtime Evaluation – To improve intuition for the behavior of the EMD-algorithm, more IMFs than necessary for peak-detection were calculated in the following experiment. The algorithm was set to produce 9 IMFs and input sizes of 250,000, 500,000, and 1,000,000 samples representing 1s, 2s, and 4s duration aggregate energy data. Initially, only one core of the CPU was used.

Table 7.3.2 shows the results of this experiment. The number of sifting steps (n_{sift}) performed until the stoppage criteria is reached can not be predicted exactly, as this process is completely data-dependent. For example, for IMF 7 in the 500,000 sample test, 146 steps of sifting were performed. This also has a big influence on the overall runtime. Maybe small perturbations in the input data in this example led to bad behavior of the cubic spline estimation.

Similarly, the processing time per sifting step (t/n_{sift}) decreases gradually as higher IMFs are calculated. The main reason is that higher IMFs contain fewer *minima* and *maxima* (smoother waveform) which results in lower time per sifting step. With respect to the size of the input, the processing time increases almost linearly. This can be seen by computing

Table 7.3.3: Runtime measurements (in s) for NILM

No. of samples	2,000,000 EMD	2,000,000 Total	4,000,000 EMD	4,000,000 Total
Sequential	402.7	406.9	783.7	791.7
Parallel (4 Cores)		177.3		347.5

the factor of 2.17 between the mean of (t/n_{sift}) for input-size 250,000 and 500,000 for example. The computing factor from input-size 500,000 to 1,000,000 is 2.11. Hence, as expected, the computation time for sifting step of increases almost linearly with input data.

Runtime Evaluation for NILM Event Detection – Overall, the total processing time for the entire algorithm is dependent on the number of IMFs to be calculated, the processing time for one step of sifting (dependent on the size of the input), and the steps of sifting until the stoppage criterion is reached. For our NILM event-detection test, the algorithm was configured to only use $IMF1$ and $IMF2$. So, in total three IMFs had to be generated. As it is not possible to process the complete data in one pass, because most machines do not have sufficient RAM, the algorithm is implemented to divide the data into smaller chunks. Here a chunk-size of 500,000 (2s) was used. Measurements were taken for input-sizes of 2,000,000 (8s) and 4,000,000 (16s). Also the benefits of using multiple cores were explored.

Table 7.3.3 shows the results, from which it can be seen that the EMD-algorithm takes the most time of the overall process. The Hilbert transform and finding the peaks takes less than 1% of the total time. By extrapolating to a full 120 s sample of the BLOND-250, the following statements about the runtime can be made:

- In sequential mode, a full 2 min (120 s) CLEAR file should take around 100 minutes to process.
- In parallel mode with 4 cores, a full 120 s CLEAR file should take around 40 minutes to process.

In general, inspite of being complex and computationally expensive, HHT proves to be an effective method to detect the SMPS appliance events. The proposed method can effectively detect low-powered signals with a good signal to noise ratio (SNR), as shown in Fig. 7.3.3.

This boost in SNR is probably due to the EMD sifting out the useful parts of the signal before analyzing the energy of the signal. Due to this boost in SNR, the event-detector based on the HHT achieved better results.

CHAPTER 8

Energy Data Compression

In this chapter, we compare state-of-the-art audio, sliding-window, and dictionary-based lossless compression algorithms to increase reliability, reduce transmission time, and hence decrease the storage requirement. Since these compression algorithms are openly available for decades, this work explores the use of already available algorithms to check their suitability with the real energy data acquired at a high sampling frequency.

Our work mainly contributes to finding any correlation between compression ratio and sampling frequency. Similarly, how well a smooth and stable periodic waveform (voltage) compresses as compared to a non-smooth and unstable periodic waveform (current). We also try to find if periodic energy data has better compression efficiency with audio compression techniques due to its similarity with music data. Since we are interested in NILM using high-frequency DAQ, our work represents a scenario where data is regularly acquired, compressed, and transferred to persistent storage where disaggregation algorithms can be readily applied.

8.1 Compression Algorithm Classification

Compression represents the art of condencing the information in a compact form. This results in a significant size reduction of particular file to support processing, transferring, or storing huge chunks of data [112]. Coding, linear prediction, and pattern recognition are among the most common techniques used in compression algorithms. The process to

compress a file is called encoding, while its decompression is termed as decoding [137].

8.1.1 Sampling Frequency

The sampling frequency is an important factor to accurately disaggregate the load. Since we are interested to detect the switching events of appliances from aggregate load and compare them with the standard appliance load signature, we require high-resolution data to increase the overall appliance detection accuracy. The data size is also linearly related with the sampling frequency. Consider, for example, that we acquire one-hour energy data for two channels (voltage and current) with 1 Hz, 1 kHz, and 100 kHz, and 32 bit per value (4 Bytes), the corresponding file sizes would be as follows:

$$1\text{Hz: } \frac{1}{\text{s}} * 3600\text{s} * 2 * 4\text{Byte} = 28.125 \text{ kByte}$$

$$1\text{kHz: } \frac{1000}{\text{s}} * 3600\text{s} * 2 * 4\text{Byte} = 27.466 \text{ MByte}$$

$$100\text{kHz: } \frac{100,000}{\text{s}} * 3600\text{s} * 2 * 4\text{Byte} = 2.682 \text{ GByte}$$

Although higher frequencies lead to larger file sizes, the appliance detection algorithms require kHz to MHz sampling frequency to accurately disaggregate [138]. The accuracy of transient-state based NILM algorithms also increases at higher sampling frequency and creates a lot of possibilities for further computation. Using high-frequency enables us to precisely detect appliance switching and predict its power consumption by utilizing a single power meter per household.

8.1.2 Compression Techniques

There are two main types of compression techniques used in literature. The lossy compression, as the name suggests, makes it impossible to recover the original signal waveform and hence cannot accurately identify the appliance during load disaggregation. Although lossy data formats such as the commonly known and widely spread MPEG-2 Layer III (better known as MP3), have a good compression rate but comes with some inherent issues which make it unsuitable for energy data. The main drawback is the inability to retrieve original data after lossy compression. Lossy compression algorithms focus on

Table 8.1.1: Summary of utilized lossless compression techniques

Audio-Based Compression			
Technique	**Year**	**Levels**	**Main Feature**
FLAC	2001	8	Rice code
OptimFrog	2001	10	Optical decorrelation
Monkeys Audio	2000	5	Linear prediction
Sliding-Window & Dictionary-Based Compression			
Technique	**Year**	**Levels**	**Main Feature**
LZMA	1998	9	Delta encoder
LZ4	2011	12	LZ77 with fixed byte encoding
Zstandard	2016	22	Asymemetric numeral system (ANS)
Gzip	1992	9	LZ77 & Huffman coding
Bzip2	1996	9	Burrows-Wheeler algorithm

discarding and de-emphasizing pieces of digital data. Data is compressed until it reaches a target file size. Lossy codecs usually take longer to compress since they have the added responsibility to decide which information can be permanently removed. A lossy algorithm can typically achieve 5-20% of the original size [137].

On the contrary, lossless compression algorithms do not irreversibly transform the data. After lossless compression, it is still possible to reproduce an exact duplicate of the original data by decoding. Thus the signal information does not change during the compression and decompression process. The lossless compression algorithms are preferred for energy data compression as they retain the signal spikes caused by the appliance switching. These unique transients are considered an essential feature for the load disaggregation algorithms. A typical lossless algorithm generally reduces the file size to about 50-60% of the original size [137]. The summary of all the lossless compression techniques used in our study is shown in Table 8.1.1. It also lists the main features and number of compression levels for each algorithm.

8.2 Audio Compression

Audio compression techniques consist of widely accepted lossless compression algorithms, as they work well with oscillating high-frequency data because the oscillating signal can be

well approximated due to repeating values. As a result, most lossless audio compression techniques have excellent compression ratios. Like audio data, high-frequency energy data is also oscillating, so audio compression should work well on energy data. In this section, we look at the audio formats utilized in this study and shortly discuss their internal configuration.

8.2.1 WAVE

WAVE [139] is an audio file format developed by IBM and Microsoft for saving raw audio data. Although it is uncompressed, WAVE is the most common raw audio format. For our audio compression technique, we have used a wav-file as an input. The WAVE file format, commonly stored within a $RIFF$ chunk, contains a 12-byte $RIFF$-header containing the string $RIFF$, file size, and the string $wave$. The small header consists of a 24-byte format section, which contains mainly information such as the number of channels and sampling rate.

8.2.2 FLAC

The Xiph.org foundation developed free lossless audio codec (FLAC) [140]. It is a lossless audio compression technique also adapted for the UK-DALE energy dataset. For FLAC audio compression, the first step is to divide the audio frame into several blocks. The number of blocks depends on the compression stage but usually lie in the range of 2000-6000 blocks. Afterward, data-coding techniques are applied to these data blocks. In FLAC, multi-channel coding is not possible as it only supports stereo channel coupling. For the stereo channel, the use of Mid-Side-/Left-Right-Coding or leave it unchanged is automatically decided.

$$\text{Left-/Right-Coding: } Channel_L \text{ - } Channel_R$$

$$\text{Mid-/Side-Coding: } M = \frac{Channel_L + Channel_R}{2}$$

$$S = \frac{Channel_L - Channel_R}{2}$$

In the next step, the linear predictive coding scheme is applied for signal approximation.

Afterward, the *residuum* is calculated, which is essentially the difference between the signal and its approximation.

$$Residuum = \text{Signal - Approximation}$$

The residuum is then compressed using Rice codes. In the last step, a header and footer are added with a 16-bit CRC-checksum for synchronization. There are four modes, again depending on the compression stage:

- Verbatim: zero order predictor → uncompressed
- Constant: used for constant values, which appear for a certain time
- Fixed linear prediction: rather restrictive linear predictor, which is limited to the fourth- order
- FIR linear prediction: up to 32nd order, Levinson-Durbin algorithm for calculating the LPC coefficients, precision can be varied from sub-frame to sub-frame

8.2.3 OptimFROG

OptimFROG (ofr) is proprietary lossless audio codec with high compression ratio and was developed by Florin Ghido in 2001. It uses generalized stereo decorrelation concept, a new audio compression technology, together with an optimal predictor to achieve superior compression [141]. The global minimum is obtained by merging the stereo decorrelation and prediction into one step. As compared to other audio codecs, one main drawback of using OptimFROG is relatively more time required for encoding and decoding the audio files.

8.2.4 Monkeys Audio

Monkeys Audio (mac) is a freely available fast, efficient, and lossless audio compression algorithm developed by Mathew T. Ashland in 2000 [142]. Mac employes a symmetric compression algorithm, where compression takes comparable time and resources as decompression. Although it has one of the highest compression ratios as compared to other codecs, it comes at the price of extra processing time [143].

The compression process starts with a transform of left and right stereo channels to mid-channel (X) and side channel (Y). The inter-correlation indicates that X is similar to both channels and Y consists of smaller numbers. A constant appears for a specified time before a somewhat restrictive linear predictor (limited to the fourth-order) is applied.

8.3 Non-Audio Compression Techniques

Most of sliding-window and dictionary-based compression techniques are based on the algorithm developed by Jacob Ziv and Abraham Lempel in 1977 called LZ77 or LZ1 [144]. Lets have a look at the very basic pseudocode of the encoder:

8.3.1 LZMA

Lempel-Ziv-Markov chain algorithm (LZMA) was developed in 1998 and is similar to LZ77 algorithm [145]. It has following main steps:

- Delta Encoding: In the first step, the data is saved in a more efficient way for the next step. The first byte is stored as it is and the following bytes are saved as the difference between the previous and the current byte.
- Sliding Dictionary Algorithm: This algorithm is a lot more complex than using a static dictionary as it always tries to find the longest match.
- Range Coder: As the last step, the range coder encodes all symbols into one using probability estimation.

8.3.2 Lz4

Lz4 is a lossless compression algorithm developed by Yann Collet [146]. It is extremely fast as it is not aiming for the best compression rate. So, the algorithm doesn't have to compute the longest match as in LZMA, which significantly increases the compression speed at a decent compression ratio. Lz4 utilizes little small-integer coding on data blocks, where a block mainly comprises a $Token$ and $Literals$.

8.3.3 Zstandard

Zstandard (Zstd) is a lossless compression algorithm also developed by Yann Collet [147]. It aims to combine the compression speed of Lz4 with a high compression ratio of LZMA to achieve high efficiency, especially for the real-time applications. This high efficiency is achieved by a combination of the LZ77 algorithm and finite state entropy, which is based on asymmetric numeral systems (ANS)[148]. ANS again is a compromise of the speed between Huffman coding and arithmetic coding.

8.3.4 Gzip

Gzip is another dictionary-based lossless compression algorithm developed by Jean-loup Gailly and Mark Adler in 1992. It is based on DEFLATE and was developed to be used for free by GNU. It is a combination of LZ77 and Huffman coding. With gzip, text files such as source code are reduced by 60-70% [149]. In general, compression achieved by gzip is much better than LZW, Huffman coding, and adaptive Huffman coding.

8.3.5 Bzip2

Bzip2 is free and open-source lossless compression technique developed and maintained by Hulian Seward. Bzip2 utilizes Burrows-Wheeler algorithm and was initially released in 1996. Compared to its predecessors such as LZW and DEFLATE, bzip2 has more efficient compression but is also relatively slower. The Burrows-Wheeler method utilizes a sorting block, where the input is read block by block and is separately encoded as one string [142].

8.4 Data Structure

8.4.1 Data Source: BLOND

The BLOND data is captured from a typical three-phase office environment and stored in HDF5-files of approximately 220 MB file size (gzip compression applied). Each file represents two minutes of data for six simultaneous channels at 250 kHz. Each file contains

30 million 16-bit integer entries for each of six sub-datasets representing three phases for current and three phases for voltage. To balance error caused by hardware components such as voltage and current transformers, a calibration factor is included to restore the original measurement-value. For our experiment, we utilized one-hour data (30 such files) recorded on the 19th of June 2017 from 11 AM to 12 PM.

8.4.2 Data Preprocesing

Data preparation and comparison of compression algorithms are implemented in Python. The high-frequency energy data were down-sampled to get eight sample rates: 250000, 125000, 50000, 25000, 10000, 5000, 2500, and 1000, which helped to test the performance of disaggregation algorithms at different sampling rates. Similarly, to ensure the same entries (same binary file size) for test-files at all sample rates, the number of entries were calculated using the smallest sample rate, which is 1kHz: (30E6 * 1000/250000 = 120000). The down-sampled data were later saved as 16-bit integer values to a *.bin* file. Additionally, a *.wav* file was created for each rate by collecting the minimum and maximum value and scaling all integer values into float values ranging from -1 to 1. These minimum and maximum values were stored to revert to the original values later for validation.

So, for each HDF5 file and therein each sub-dataset (current1, current2, current3, voltage1, voltage2, voltage3), the *.bin* and *.wav* files were prepared at eight sample rates. As measurements are from a three-phase system, 1,2, and 3 here represent current and voltage for phase 1, phase 2, and phase 3. Additionally, while reading in the files, the power for each phase (power1, power2, power3) was calculated on the fly by multiplying the respective current and voltage entries. This data generated (30 files) * (9 datasets) * (6 sample rates) = 1620 *.bin* and *.wav* file pairs.

8.5 Evaluation

In this section, we will compare eight state-of-the-art audio, sliding-window, and dictionary-based lossless compression algorithms. The comparison is based on compression ratio and processing time to figure out which technique performs better. Although several compression levels are available for each compression algorithm, as indicated in Table 2, we have only included the best compression ratio of each algorithm for our results. Before

going into detail regarding different compression algorithms, lets discuss how we compare different compression algorithms in our study.

8.5.1 Compression Ratio

Compression ratio is an important factor to classify different compression algorithms. Compression ratio is obtained by comparing the initial uncompressed file with the compressed file and can be calculated as follows [142].

$$\text{Compression ratio} = \frac{size\,of\,compressed\,data}{size\,of\,initial\,data}$$

All our compression ratios are compared with a binary file to have a good base value for the comparison.

Various factors can affect the compression ratio. If the dataset includes a few symbols that occur with high-frequency, this pattern is beneficial for a compression algorithm, and the compression performance would be relatively good. Same applies to data that has low entropy. Entropy is a measure of uncertainty. High entropy means the data has high variance and thus contains much information or noise. In the case of data with high entropy, the compression algorithm has a hard time finding redundancies, and therefore the performance of the algorithm in terms of compression ratio is not good. Additionally, the variety of reading types and dataset sizes impact compression performance. Some algorithms have shown better performance for a larger dataset, whereas others exhibit higher compression ratio for small datasets.

8.5.2 Processing Time

The processing time indicates the time required to execute the compression algorithms. Typically, it depends upon the complexity of the compression algorithms and hardware capabilities of the energy monitoring equipment. The only reason for introducing processing time here is to compare the complexity of the compression algorithms. Processing time includes both compression and decompression time and is affected by many factors including data structure, amplitude (or volume), and the compression ratio. Some experimental studies [112] have shown that the compression speed of some algorithms is not affected by

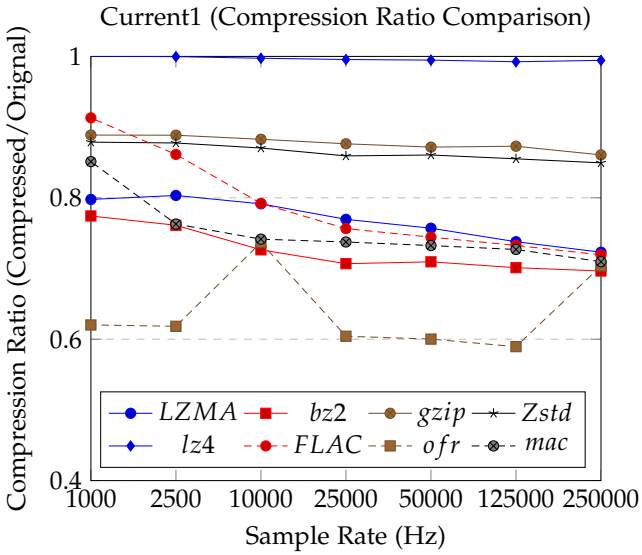


Figure 8.5.1: Compression ratio for current at different sampling frequencies

the size of the dataset. Similarly, some compression algorithms perform faster on larger data sets. Compression time is a very important criterion to choose compression algorithms in different applications, especially when considering real-time applications.

8.5.3 Comparison Results

Firstly, lets compare the compression ratio of all eight lossless compression techniques for current (I), voltage (V), and power (P) of channel 1, as shown in Fig. 8.5.1, 8.5.2, and 8.5.3, respectively. The compression ratio on y-axis determines the how much file size is reduced as compared to original. Here, a value of 1 indicates no compression whereas a value of 0.6 indicates data occupies 60% of its original size after compression (reduced by 40%). In general, the voltage data compresses more efficiently for most of the lossless algorithms, mainly due to the periodic waveform and repeating values. Similarly, the compression ratio for V and I tends to improve gradually, especially above 10 kHz. This improvement in compression ratio suggests that the increase in sampling frequency results in better compression. In terms of compression ratio, OptimFROG (ofr) shows better results for both current and voltage signals at almost all sampling frequencies, as shown in Fig. 8.5.1 and 8.5.2. It is interesting to see that bzip2 (bz2) also performs well for current waveform which contains a lot of spikes due to presence of switched mode power supplies (SMPS) in an office environment. For power signal, opr performs better at most sampling frequencies.

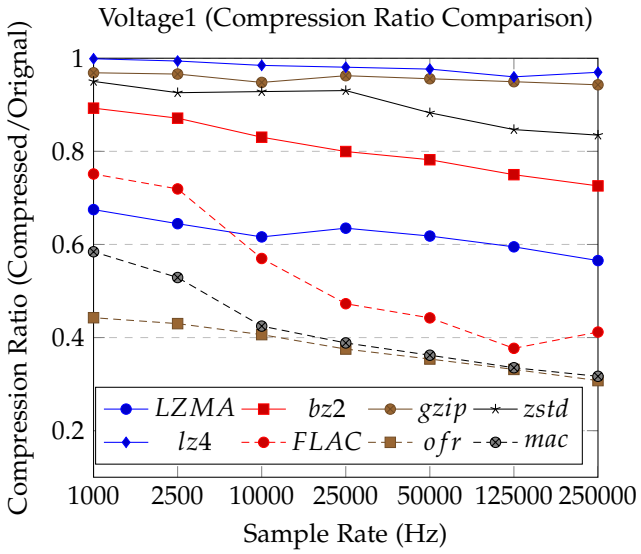


Figure 8.5.2: Compression ratio for voltage at different sampling frequencies

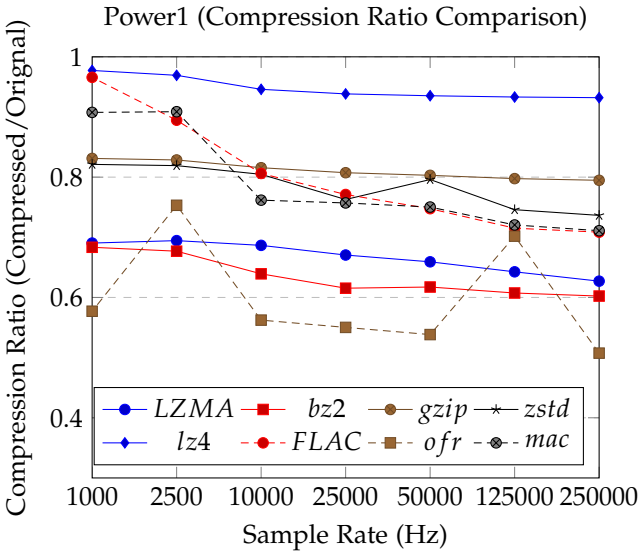


Figure 8.5.3: Compression ratio for power at different sampling frequencies

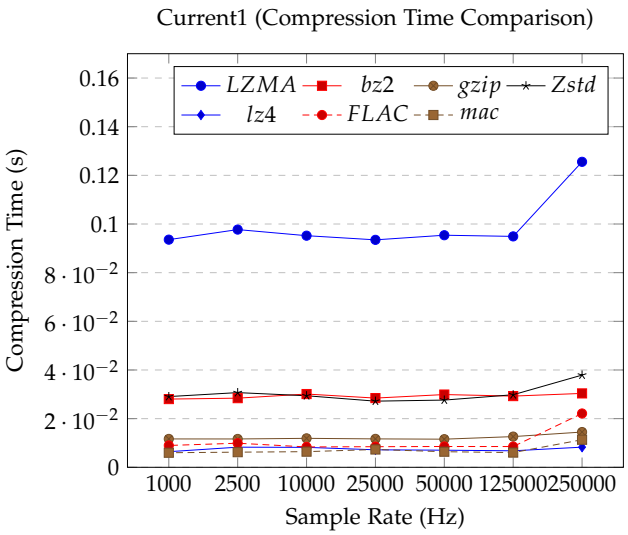


Figure 8.5.4: Compression time for current at different sampling frequencies

Overall, the audio-based techniques show much better compression ratios as compared to the sliding-window and dictionary-based techniques.

For compression time, the results are shown below. We have excluded ofr from the results because, as expected, it takes a long time to compress (see Fig. 8.5.10). In terms of time to compress, mac and lz4 show better performance for current (Fig. 8.5.4), voltage (Fig. 8.5.5), and power (Fig. 8.5.6). It is interesting to see that apart from LZMA, most algorithms show reasonable performance.

Similarly for compression time, the results are shown below. In General, decompression takes much less time as compared to compression. In terms of time to decompress, lz4 and Zstd show identical performance for all current (Fig. 8.5.7), voltage (Fig. 8.5.8), and power (Fig. 8.5.9) waveforms.

8.5.4 Key Findings for Energy Data Compression

We know that the choice of compression algorithm varies with application. For energy data using NILM, if near real-time disaggregation is required, processing time is the defining factor whereas, for offline disaggregation, the compression ratio is of paramount importance. One can, therefore, argue that for near real-time disaggregation, compression can be directly performed at the remote server (with better computation power) by transferring the uncompressed data. Unfortunately, that would be a constraint for restricted bandwidth

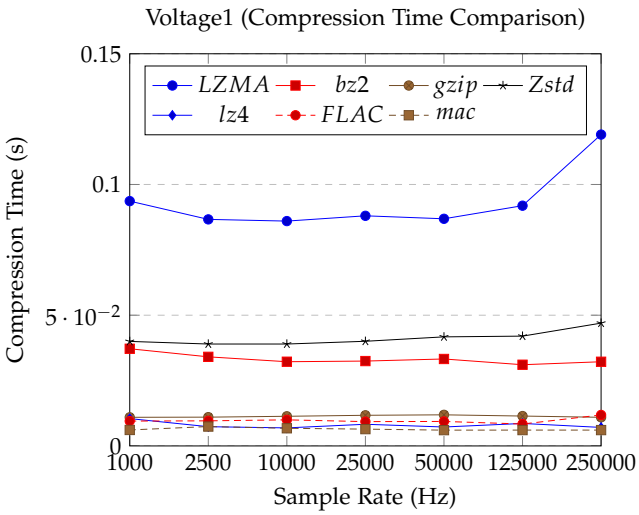


Figure 8.5.5: Compression time for voltage at different sampling frequencies

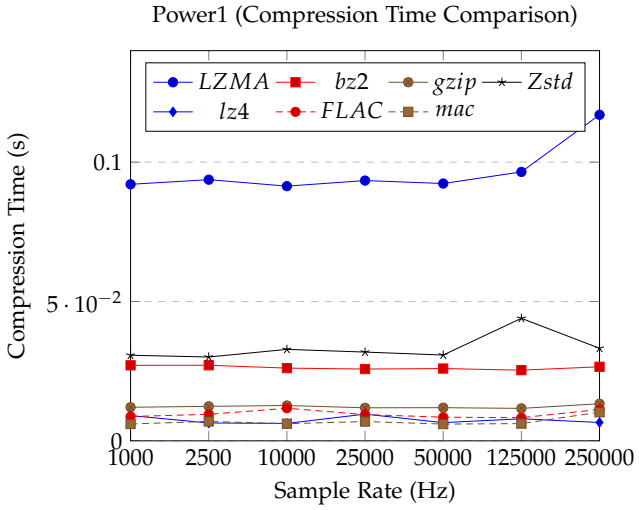


Figure 8.5.6: Compression time for power at different sampling frequencies

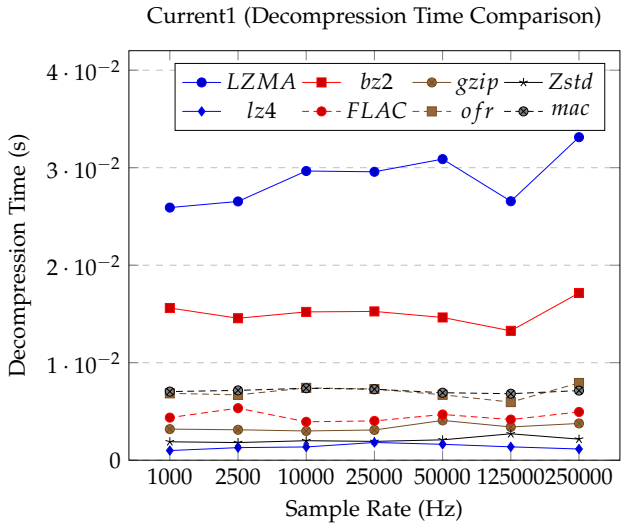


Figure 8.5.7: Decompression time for current at different sampling frequencies

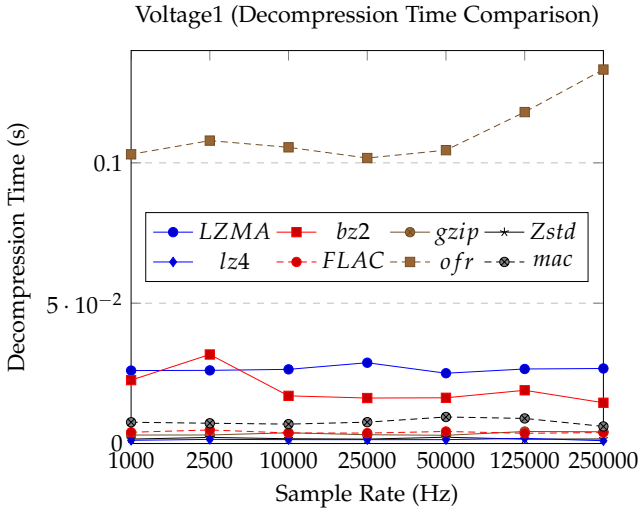


Figure 8.5.8: Decompression time for voltage at different sampling frequencies

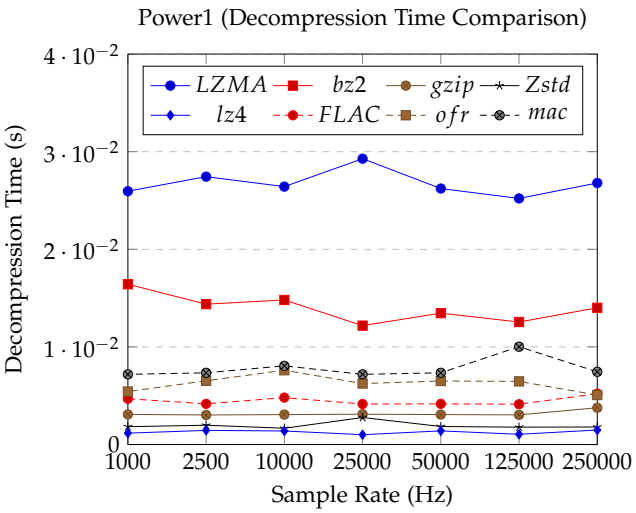


Figure 8.5.9: Decompression time for power at different sampling frequencies

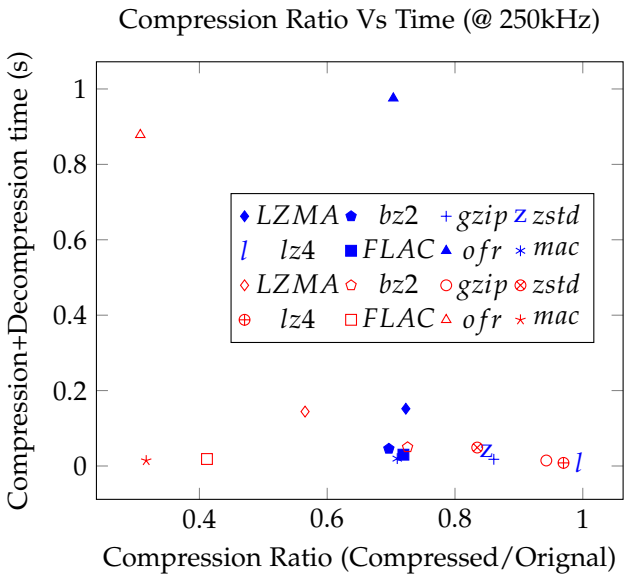


Figure 8.5.10: Comparison of compression ratio vs time for voltage (red) and current (blue) at 250 kHz

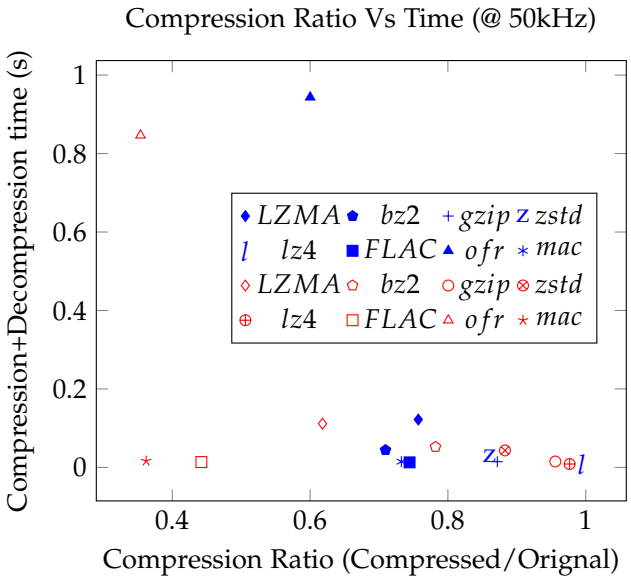


Figure 8.5.11: Comparison of compression ratio vs time for voltage (red) and current (blue) at 50 kHz

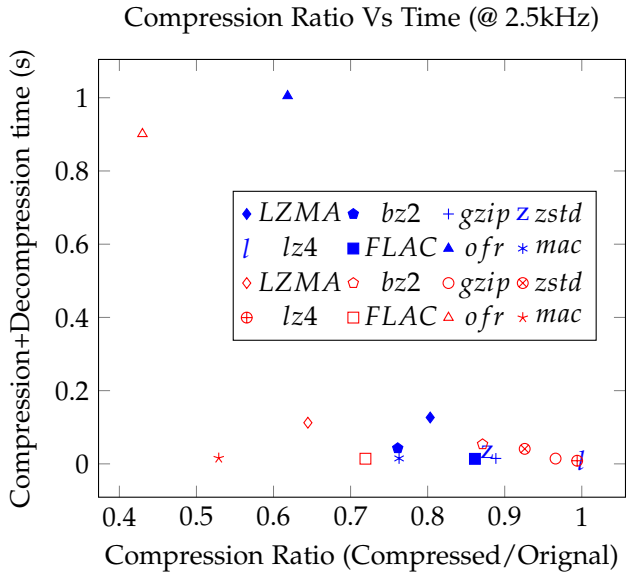


Figure 8.5.12: Comparison of compression ratio vs time for voltage (red) and current (blue) at 2500 Hz

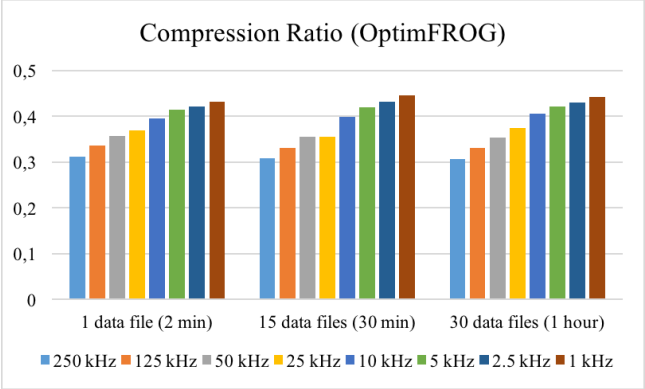


Figure 8.5.13: Comparison of compression ratio with sampling frequency and data volume for voltage1 (OptimFROG)

networks.

We compare both voltage and current of all eight lossless compression techniques at 250 kHz, 50 kHz, and 2500 Hz, as shown in Fig. 8.5.10, 8.5.11, and 8.5.12, respectively, to determine the best all-around performance. Clearly, mac is the overall winner at almost all the frequencies closely followed by FLAC. The results indicate that audio compression techniques such as mac and FLAC perform better at good compression speeds (less compression time). Their best performance at around 50 kHz (and above) is justified as these formats are originally designed to compress audio files, which also lie at a similar frequency range. Similarly, since audio compression algorithms perform better with smooth changes, they show superior performance for the periodic energy data. On the contrary, if compression time is not an issue, OptimFROG outperforms the other compression techniques for both current and voltage at high- and low-frequency data.

To observe the effect of compression ratio on sampling frequency and data volume, we have compared the compression ratio of OptimFROG voltage waveform, as shown in Fig. 8.5.13. We can observe that compression ratio increases with the sampling frequency. Hence, the sampling and compression ratio are positively correlated in almost all compression algorithms we tested, most significantly notable in the audio-based compression techniques. Similarly, we observed no significant change in compression ratio based on data volume, at least not for one-hour data.

To compare overall performance, we added the compression and decompression time to compare them directly to the compression ratio. Our experiment indicates that decompression is generally much faster as compared to compression. Similarly, there is a

significant difference, again most notable in the audio-based algorithms, in the compression ratios of current as compared to the voltage. The smoother voltage waveform is usually better compressed than the uneven and spikier current waveform.

CHAPTER 9

Conclusions

Modern buildings contribute a significant portion of electric load to the grid. In order to increase energy efficiency and participate effectively in demand side management programs, consumers need access to real-time energy consumption information, preferably at the appliance-level. Recently, there has been a growing interest in appliance-level energy monitoring to help consumers view fine-grained energy consumption information. The NILM approach utilizes single-point sensing and machine learning techniques to help disaggregate energy data and estimate the appliance-specific energy consumption. NILM provides one such appliance-level energy monitoring solution at a considerably reduced cost.

In this work, we first compare different state-of-the-art e-monitors available on the market and determine their ability to utilize load disaggregation. Through the online technical survey and detailed product review, we compared 41 e-monitors on the basis of several dimensions, including measured and derived parameters, the sampling frequency, the accuracy, the resolution, the area of application and the cost. The comparison suggests that most e-monitors possess enough capabilities and processing power to incorporate advanced monitoring techniques for the residential environment. In the future, these intelligent meters can act as the point of contact between smart buildings for local demand response and renewable resource sharing. Before the complete roll-out of smart meters, consumers can realize the offered advantages by selecting and using intelligent off-the-shelf e-monitors.

Since most challenges in NILM stem from DAQ hardware, this work proposes a fine-grained and high-frequency data acquisition solution to accurately detect individual appliances

from the aggregate load, even when several appliances are operating simultaneously. Accurate appliance detection mainly depends on the number and type of appliances under observation. For commercial and industrial environments, there are multiple appliances such as laptops (office), refrigerators (supermarket), and motors (factory) operating in parallel, which usually require high-frequency data acquisition for anomaly detection and simultaneous event reduction. To handle appliance specific disaggregation requirements for challenging commercial and industrial environments, we have developed CLEAR, a customizable and NILM-enabled energy monitoring solution. CLEAR is capable of providing a cost-effective mechanism to simultaneously gather accurate energy data at up to 250 kHz from multiple circuits.

Most of the available NILM studies focus on the residential environment with little information regarding complex non-residential environments such as an office or a factory. Unlike a residential environment, most commercial and industrial settings consist of a large number of appliances containing multiple sub-machine components such as controllers, heating, and lighting components operating simultaneously. Due to low power consumption and variable switching characteristics, there is no cost-effective solution to isolate individual sub-components. With its modular and customizable DAQ features, CLEAR provides an ideal solution to record high-quality data from multiple environments. CLEAR was tested in an office environment to observe the effect of high-resolution on simultaneous event detection.

Similarly, we also developed a novel event detection approach for non-intrusive load monitoring based on the Hilbert-Huang transform. Even after decades of NILM research, NILM event detection remains a daunting task, especially for complex non-residential environments. For accurate detection of micro-events, efficient detection algorithms along with high-resolution data acquisition are necessary. To evaluate our event detection technique, we apply our proposed algorithm to BLOND, a building level office environment dataset collected using CLEAR. Due to an abundance of SMPS-equipped appliances, simultaneous micro-events are almost inevitable in such an environment. The HHT-based event detection approach on high-frequency energy data suggests that the micro-events can be detected.

Lastly, data compression is employed to reduce the bandwidth requirement, improve file transfer time, and eventually minimize the storage requirement when CLEAR or similar DAQ hardware is utilized. We have evaluated eight lossless compression algorithms

including FLAC, OptimFROG, Monkeys Audio, LZMA, Zstd, bz2, gzip, and Lz4 to check their suitability with the high-frequency energy data for NILM. We compared them considering different aspects such as the compression/decompression speed as well as their compression ratios to pinpoint some suitable compression techniques for high-frequency energy data.

According to our comparison, OptimFROG shows an excellent compression ratio, even better than the best versions of LZMA and FLAC. Overall, Monkeys Audio and FLAC show dominant all-around performance with reasonable compression ratio and superior processing speed. Our study indicates that audio-based lossless compression techniques outperform other methods, primarily due to steady changes in the periodic waveform. Similarly, the smoother voltage waveform compresses much more efficiently as compared to the AC waveform. Moreover, the compression ratio and the sampling frequency are positively correlated, where higher frequency leads to a slightly better compression ratio. In general, it can be safely stated that audio-based lossless compression algorithms give better performance and can be employed for energy data compression.

There are a lot of possibilities for further work, especially the HHT event detection method could be valuable for NILM. Besides improvements of the developed algorithm, a deeper understanding of the BLOND dataset and the events it contains can also help fine-tune the expected outcomes. A primary component in the EMD algorithm is the creation of the upper and lower envelopes by using cubic spline interpolation. Other interpolation methods such as B-splines and min/max filters can be used to estimate the envelopes [150] and improve the performance and accuracy of event detection. Similarly, besides optimizing the threshold value for the currently used stoppage criteria (standard deviation), one should also look more deeply into other criteria to stop the HHT algorithm [151]. The threshold method [152] and S-number criterion are other possible methods to stop the sifting process.

Another possibility to improve the event detection accuracy is by combining the HHT with more classical signal processing approaches. One might apply a high-pass filter before using the HHT, thereby removing the highest frequency (noise) which contains no useful appliance information. This would help optimize the overall process as it is easier to create envelopes for low-frequency data (due to fewer spikes). Similarly, to perform the HHT in near real-time conditions, the most promising direction is to look into hardware assisted methods such as digital signal processing techniques [153, 154]. The HHT event detection method should be applied to other NILM datasets such as BLUED [86] (residential, 60

Hz), UK-DALE [87] (residential, 50 Hz), and LILAC [7] (industrial, 50 Hz) to compare the detection performance of energy data from different settings.

The wide area NILM approach, as discussed in Chapter 1 (see 1.3.5), is also a viable extension to this work. In a district-level NILM approach, all the major residential appliances in the district such as electric kettle, microwave, refrigrator, and electric stove might be detected using the HHT-based approach. The district-level NILM approach will help both utility and consumer to estimate the local demand and make informed decision regarding renewable integration in the micro-grid. Similarly, we also explored the use of NILM toolkit (NILMTK) [43] with CLEAR data. NILMTK is an open-source project to benchmark different NILM datasets using a standard data format. It comes with pre-processing algorithms, statistical metrics to describe a dataset, and four state-of-the-art disaggregation algorithms: factorial hidden markov model (FHMM), combinatorial optimization, Hart85, and maximum likelihood estimation. Initially, we experimented with downsampled 1 s data as NILMTK currently only supports low-frequency. An extension to higher sampling rate would be good in the future to compare the benefits of high-frequency datasets.

List of Figures

List of Tables

Bibliography

[1] A. U. Haq, T. Kriechbaumer, M. Kahl, and H.-A. Jacobsen. "CLEAR - A circuit level electric appliance radar for the electric cabinet." In: *2017 IEEE International Conference on Industrial Technology (ICIT)*. Toronto, Canada, Mar. 2017, pp. 1130–1135. ISBN: 978-1-5090-5319-3/17. DOI: 10.1109/ICIT.2017.7915521.

[2] A. U. Haq and H.-A. Jacobsen. "Prospects of appliance-level load monitoring in off-the-shelf energy monitors: A technical review." In: *Energies* 11.1 (2018), p. 189. DOI: 10.3390/en11010189.

[3] M. Kahl, C. Goebel, A. U. Haq, T. Kriechbaumer, and H.-A. Jacobsen. "NoFaRe: A non-intrusive facility resource monitoring system." In: *Energy Informatics*. Vol. 9424. Lecture Notes in Computer Science. Springer International Publishing, 2015, pp. 59–68. DOI: 10.1007/978-3-319-25876-8_6.

[4] T. Kriechbaumer, A. U. Haq, M. Kahl, and H.-A. Jacobsen. "MEDAL: A cost-effective high-frequency energy data acquisition system for electrical appliances." In: *Proceedings of the 2017 ACM Eighth International Conference on Future Energy Systems*. e-Energy '17. Hong Kong, Hong Kong: ACM, May 2017. ISBN: 978-1-4503-5036-5/17/05. DOI: 10.1145/3077839.3077844.

[5] M. Kahl, A. U. Haq, T. Kriechbaumer, and H.-A. Jacobsen. "A Comprehensive feature study for appliance recognition on high frequency energy data." In: *Proceedings of the Eighth International Conference on Future Energy Systems*. ACM. 2017, pp. 121–131. DOI: 10.1145/3077839.3077845.

[6] M. Kahl, T. Kriechbaumer, A. U. Haq, and H.-A. Jacobsen. "Appliance classification across multiple high frequencyEnergy datasets." In: *2017 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. 2017. DOI: 10.1109/smartgridcomm.2017.8340664.

[7] M. Kahl, V. Krause, R. Hackenberg, et al. "Measurement system and dataset for in-depth appliance energy consumption analysis in industrial environments." In: *tm - technisches messen* (2018). DOI: 10.1515/teme-2018-0038.

[8] M. Kahl, A. U. Haq, T. Kriechbaumer, and H.-A. Jacobsen. "WHITED- A worldwide household and industry transient energy dataset." In: *3rd International Workshop on Non-Intrusive Load Monitoring*. 2016. URL: https://www.i13.in.tum.de/index.php?id=114&L=0.

[9] J. Rogelj, M. Den Elzen, N. Höhne, et al. "Paris agreement climate proposals need a boost to keep warming well below 2 C." In: *Nature* 534.7609 (2016), p. 631.

[10] E. R. Masanet. *Energy technology perspectives 2017: Catalysing energy technology transformations*. OECD, 2017.

[11] M. Milligan, B. Frew, E. Zhou, and D. J. Arent. *Advancing System Flexibility for High Penetration Renewable Integration*. Tech. rep. NREL (National Renewable Energy Laboratory), 2015.

[12] J. M. Carrasco, L. G. Franquelo, J. T. Bialasiewicz, et al. "Power-electronic systems for the grid integration of renewable energy sources: A survey." In: *IEEE Transactions on Industrial Electronics* 53.4 (2006), pp. 1002–1016.

[13] X. Fang, S. Misra, G. Xue, and D. Yang. "Smart grid- The new and improved power grid: A survey." In: *IEEE Communications Surveys & Tutorials* 14.4 (2012), pp. 944–980.

[14] J. Kelly and W. Knottenbelt. "Does disaggregated electricity feedback reduce domestic electricity consumption? A systematic review of the literature." In: *arXiv preprint arXiv:1605.00962* (2016).

[15] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar. "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey." In: *Sensors* 12.12 (2012), pp. 16838–16866.

[16] J. M. Gillis, J. A. Chung, and W. G. Morsi. "Designing new orthogonal high order wavelets for non-intrusive load monitoring." In: *IEEE Transactions on Industrial Electronics* (2017).

[17] 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 (2013), pp. 213–234.

[18] Y. F. Wong, Y. A. Şekercioğlu, T. Drummond, and V. S. Wong. "Recent approaches to non-intrusive load monitoring techniques in residential settings." In: *2013 IEEE Computational Intelligence Applications in Smart Grid (CIASG)*. Apr. 2013, pp. 73–79. DOI: 10.1109/CIASG.2013.6611501.

[19] T. Kriechbaumer and H.-A. Jacobsen. "BLOND, a building-level office environment dataset of typical electrical appliances." In: *Scientific Data* 5.180048 (Mar. 2018). DOI: 10.1038/sdata.2018.48.

[20] 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 (2013), pp. 213–234. ISSN: 0301-4215.

[21] S. Makonin. "Investigating the switch continuity principle assumed in non-intrusive load monitoring (NILM)." In: *Electrical and Computer Engineering (CCECE), 2016 IEEE Canadian Conference on*. IEEE. 2016, pp. 1–4.

[22] P. Lindahl, G. Bredariol, J. Donnal, and S. Leeb. "Noncontact electrical system monitoring on a US Coast Guard cutter." In: *IEEE Instrumentation & Measurement Magazine* 20.4 (2017), pp. 11–20.

[23] T. S. Abdelgayed, W. G. Morsi, and T. S. Sidhu. "Fault detection and classification based on co-training of semi-supervised machine learning." In: *IEEE Transactions on Industrial Electronics* (2017).

[24] A. Aboulian, D. H. Green, J. F. Switzer, et al. "NILM dashboard: A power system monitor for electromechanical equipment dåiagnostics." In: *IEEE Transactions on Industrial Informatics* (2018).

[25] U. A. Orji, Z. Remscrim, C. Schantz, et al. "Non-intrusive induction motor speed detection." In: *IET Electric Power Applications* 9.5 (2015), pp. 388–396.

[26] T. A. Edison. *Electric Meter*. Google Patents. 1881. URL: http://www.google.com/patents/US251545.

[27] O. Blitiiy. *Electric Meter for Alternating Current*. Google Patents. 1890. URL: https://www.google.com/patents/US423210.

[28] J. J. Conti and P. D. Holtberg. *Annual energy outlook 2015 with projections to 2040*. Tech. rep. U.S. Energy Information Administration (EIA), April, 2015.

[29] P. Bertoldi and B. Atanasiu. *Electricity consumption and efficiency trends in the enlarged European Union*. Tech. rep. Institute for Environment and Sustainability, 2007.

[30] Siemens. *Building automation impact on energy efficiency*. Tech. rep. Siemens Building Technologies, 2007. URL: http://w3.siemens.com/market-specific/global/en/data-centers/Documents/BAU-impact-on-energy-efficiency.pdf.

[31] "Smart Meter Systems: A metering industry perspective." In: *A Joint Project of the EEI and AEIC Meter Committees* (2011).

[32] R. Li, G. Dane, C. Finck, and W. Zeiler. "Are building users prepared for energy flexible buildings?—A large-scale survey in the Netherlands." In: *Applied Energy* 203 (2017), pp. 623–634.

[33] D. Wei, Y. Lu, M. Jafari, P. M. Skare, and K. Rohde. "Protecting smart grid automation systems against cyberattacks." In: *Smart Grid, IEEE Transactions on* 2.4 (2011), pp. 782–795. ISSN: 1949-3053.

[34] H. Zou, H. Jiang, J. Yang, L. Xie, and C. Spanos. "Non-intrusive occupancy sensing in commercial buildings." In: *Energy and Buildings* 154 (2017), pp. 633–643.

[35] TI-Designs. *Single-phase electric meter with isolated energy measurement*. Tech. rep. Taxes Instruments, 2016. URL: http://www.ti.com/lit/ug/tidu455a/tidu455a.pdf.

[36] R. Schlobohm. *Electronic power meters guide for their selection and specification*. Tech. rep. GE Specification Engineer, 2005.

[37] G. J. Wakileh. *Power systems harmonics: fundamentals, analysis and filter design*. Springer Science & Business Media, 2001. ISBN: 3540422382.

[38] *Evaluation of advanced meter system deployment in Texas – Meter accuracy assessment*. Tech. rep. Navigant Consulting (PI) LLC, , July 30, 2010.

[39] H. Joshi. *Residential, commercial and industrial electrical systems: Network and installation*. Vol. 2. Tata McGraw-Hill Education, 2008. ISBN: 0070620970.

[40] J. Palmer and N. Terry. *Costing monitoring equipment for a longitudinal energy survey*. Tech. rep. Department of Energy and Climate Change, UK, 2015.

[41] S. Giri and M. Bergés. "An energy estimation framework for event-based methods in non-intrusive load monitoring." In: *Energy Conversion and Management* 90 (2015), pp. 488–498.

[42] 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 (2013), pp. 213–234. ISSN: 0301-4215.

[43] N. Batra, J. Kelly, O. Parson, et al. "NILMTK: An open source toolkit for non-intrusive load monitoring." In: *Proceedings of the 5th International Conference on Future Energy Systems*. ACM. 2014, pp. 265–276.

[44] L. K. Norford and S. B. Leeb. "Non-Intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms." In: *Energy and Buildings* 24.1 (1996), pp. 51–64.

[45] H.-H. Chang. "Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses." In: *Energies* 5.11 (2012), pp. 4569–4589.

[46] G. W. Hart. "Non-intrusive appliance load monitoring." In: *Proceedings of the IEEE* 80.12 (1992), pp. 1870–1891. ISSN: 0018-9219.

[47] L. K. Norford and N. Mabey. "Non-Intrusive electric load monitoring in commercial buildings." In: (1992).

[48] T. Onoda, G. Rätsch, and K.-R. Müller. "Applying support vector machines and boosting to a non-intrusive monitoring system for household electric appliances with inverters." In: (2000).

[49] T. B. Fomby and T. Barber. "K-nearest neighbors algorithm: Prediction and classification." In: *Lecture notes in Southern Methodist University, Dallas, TX* (2008), pp. 1–5.

[50] L. Gomes, F. Fernandes, T. Sousa, et al. "Contextual intelligent load management with ANN adaptive learning module." In: *Intelligent System Application to Power Systems (ISAP), 2011 16th International Conference on*. IEEE. 2011, pp. 1–6.

[51] 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 Built Environments*. ACM. 2015, pp. 55–64.

[52] A. G. Ruzzelli, C. Nicolas, A. Schoofs, and G. M. O'Hare. "Real-time recognition and profiling of appliances through a single electricity sensor." In: *2010 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*. IEEE. 2010, pp. 1–9.

[53] M. Ringwelski. "The effect of data granularity on load data compression." In: *Energy Informatics: 4th DA-CH Conference, EI 2015, Karlsruhe, Germany, November 12-13, 2015, Proceedings*. Vol. 9424. Springer. 2016, p. 69.

[54] A. Unterweger and D. Engel. "Resumable load data compression in smart grids." In: *IEEE Transactions on Smart Grid* 6.2 (2015), pp. 919–929.

[55] H. N. Rafsanjani, C. R. Ahn, and M. Alahmad. "A review of approaches for sensing, understanding, and improving occupancy-related energy-use behaviors in commercial buildings." In: *Energies* 8.10 (2015), pp. 10996–11029.

[56] M. Heidarinejad, J. G. Cedeño-Laurent, J. R. Wentz, et al. "Actual building energy use patterns and their implications for predictive modeling." In: *Energy Conversion and Management* 144 (2017), pp. 164–180.

[57] K. S. Barsim, R. Streubel, and B. Yang. "An approach for unsupervised non-intrusive load monitoring of residential appliances." In: *Proceedings of the 2nd International Workshop on Non-Intrusive Load Monitoring*. 2014.

[58] M. Zeifman and K. Roth. "Non-intrusive appliance load monitoring: Review and outlook." In: *IEEE Transactions on Consumer Electronics* 57.1 (2011).

[59] J. Liang, S. K. Ng, G. Kendall, and J. W. Cheng. "Load signature study—Part I: Basic concept, structure, and methodology." In: *IEEE Transactions on Power Delivery* 25.2 (2010), pp. 551–560.

[60] X. Hao, B. Tang, L. Hulu, and Y. Wang. "On the balance of meter deployment cost and NILM accuracy." In: *Proceedings of the 24th International Conference on Artificial Intelligence*. AAAI Press. 2015, pp. 2603–2609.

[61] D. Srinivasan, W. Ng, and A. Liew. "Neural-network-based signature recognition for harmonic source identification." In: *IEEE Transactions on Power Delivery* 21.1 (2006), pp. 398–405.

[62] M. Berges, E. Goldman, H. S. Matthews, and L. Soibelman. "Learning systems for electric consumption of buildings." In: *ASCI International Workshop on Computing in Civil Engineering*. Vol. 38. 2009.

[63] 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: *Energy and Buildings* 30.3 (1999), pp. 245–259.

[64] S. Makonin and F. Popowich. "Nonintrusive load monitoring (NILM) performance evaluation." In: *Energy Efficiency* 8.4 (2015), pp. 809–814.

[65] *Smappee*. (last accessed: last accessed: 05-01-18). URL: `http://www.smappee.com/us/`.

[66] A. S. Bouhouras, P. A. Gkaidatzis, K. C. Chatzisavvas, et al. "Load Signature Formulation for Non-Intrusive Load Monitoring Based on Current Measurements." In: *Energies* 10.4 (2017), p. 538.

[67] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar. "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey." In: *Sensors* 12.12 (2012), pp. 16838–16866.

[68] Y.-H. Lin and M.-S. Tsai. "Development of an improved time–frequency analysis-based non-intrusive load monitor for load demand identification." In: *Instrumentation and Measurement, IEEE Transactions on* 63.6 (2014), pp. 1470–1483. ISSN: 0018-9456.

[69] A. Dalen and C. Weinhardt. "Evaluating the impact of data sample-rate on appliance disaggregation." In: *Energy Conference (ENERGYCON), 2014 IEEE International*. IEEE, 2014, pp. 743–750.

[70] J. Z. Kolter and M. J. Johnson. "REDD: A public data set for energy disaggregation research." In: *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*. Vol. 25. Citeseer, 2011, pp. 59–62.

[71] A. Rowe, M. Berges, and R. Rajkumar. "Contactless sensing of appliance state transitions through variations in electromagnetic fields." In: *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*. ACM, 2010, pp. 19–24. ISBN: 1450304583.

[72] S. Gupta, M. S. Reynolds, and S. N. Patel. "ElectriSense: Single-point sensing using EMI for electrical event detection and classification in the home." In: *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*. ACM, 2010, pp. 139–148. ISBN: 1605588431.

[73] M. Aiad and P. H. Lee. "Unsupervised approach for load disaggregation with devices interactions." In: *Energy and Buildings* 116 (2016), pp. 96–103.

[74] *INTrEPID: INTelligent systems for Energy Prosumer buildIngs at District level, Tech. rep., FP7-ICT Project (317983).*

[75] C. E. Kement, H. Gultekin, B. Tavli, T. Girici, and S. Uludag. "Comparative Analysis of Load Shaping Based Privacy Preservation Strategies in Smart Grid." In: *IEEE Transactions on Industrial Informatics* (2017).

[76] D. Engel and G. Eibl. "Wavelet-based multiresolution smart meter privacy." In: *IEEE Transactions on Smart Grid* 8.4 (2017), pp. 1710–1721.

[77] M. Lisovich and S. Wicker. "Privacy concerns in upcoming residential and commercial demand-response systems." In: *IEEE Proceedings on Power Systems* 1.1 (2008), pp. 1–10.

[78] T. Babaei, H. Abdi, C. P. Lim, and S. Nahavandi. "A study and a directory of energy consumption data sets of buildings." In: *Energy and Buildings* 94 (2015), pp. 91–99.

[79] J. Z. Kolter and M. J. Johnson. "REDD: A public data set for energy disaggregation research." In: *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*. Vol. 25. 2011, pp. 59–62.

[80] G. W. Hart. "Non-Intrusive appliance load monitoring." In: *Proceedings of the IEEE* 80.12 (1992), pp. 1870–1891.

[81] K. Basu, V. Debusschere, S. Bacha, U. Maulik, and S. Bondyopadhyay. "Nonintrusive load monitoring: A temporal multilabel classification approach." In: *IEEE Transactions on Industrial Informatics* 11.1 (2015), pp. 262–270.

[82] H. Maass, H. Çakmak, W. Suess, et al. "Introducing the electrical data recorder as a new capturing device for power grid analysis." In: *Applied Measurements for Power Systems (AMPS), 2012 IEEE International Workshop on*. IEEE. 2012, pp. 1–6.

[83] W. Wichakool. "Advanced non intrusive load monitoring system." PhD thesis. Massachusetts Institute of Technology, 2011.

[84] Z. Clifford, J. J. Cooley, A.-T. Avestruz, et al. "A retrofit 60 Hz current sensor for non-intrusive power monitoring at the circuit breaker." In: *Applied Power Electronics Conference and Exposition (APEC), 2010 Twenty-Fifth Annual IEEE*. IEEE. 2010, pp. 444–451.

[85] K. D. Lee. "Electric load information system based on non-intrusive power monitoring." PhD thesis. Massachusetts Institute of Technology, 2003.

[86] K. Anderson, A. Ocneanu, D. Benitez, et al. "BLUED: A fully labeled public dataset for event-based non-intrusive load monitoring Research." In: *Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*. Beijing, China, Aug. 2012.

[87] J. Kelly and W. Knottenbelt. "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes." In: *Scientific Data* 2 (2015), p. 150007.

[88] Z. A. Clifford. "An analog and digital data acquisition system for non-intrusive load monitoring." PhD thesis. Massachusetts Institute of Technology, 2009.

[89] U. Series. *LabJack- Measurement and Automation*. URL: https://labjack.com/products/ue9.

[90] Sense. *How the sense home energy monitor works*. URL: https://blog.sense.com/how-the-sense-home-energy-monitor-works/.

[91] K. D. Anderson, M. E. Bergés, A. Ocneanu, D. Benitez, and J. M. Moura. "Event detection for non intrusive load monitoring." In: *IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society*. IEEE. 2012, pp. 3312–3317.

[92] R. Dong, L. Ratliff, H. Ohlsson, and S. S. Sastry. "Fundamental limits of non-intrusive load monitoring." In: *Proceedings of the 3rd International conference on High Confidence Networked Systems*. ACM. 2014, pp. 11–18.

[93] B. Wild, K. S. Barsim, and B. Yang. "A new unsupervised event detector for non-intrusive load monitoring." In: *Signal and Information Processing (GlobalSIP), 2015 IEEE Global Conference on*. IEEE. 2015, pp. 73–77.

[94] M. N. Meziane, T. Picon, P. Ravier, G. Lamarque, and J. Le. "A new measurement system for high frequency NILM with controlled aggregation scenarios." In: *Workshop on Non-Intrusive Load Monitoring (NILM), 2016 Proceedings of the 3rd International*. 2016.

[95] M. N. Meziane, P. Ravier, G. Lamarque, J.-C. Le Bunetel, and Y. Raingeaud. "High accuracy event detection for non-intrusive load monitoring." In: *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE. 2017, pp. 2452–2456.

[96] L. De Baets, J. Ruyssinck, C. Develder, T. Dhaene, and D. Deschrijver. "On the Bayesian optimization and robustness of event detection methods in NILM." In: *Energy and Buildings* 145 (2017), pp. 57–66.

[97] L. De Baets, J. Ruyssinck, D. Deschrijver, and T. Dhaene. "Event detection in NILM using cepstrum smoothing." In: *3rd International Workshop on Non-Intrusive Load Monitoring*. 2016, pp. 1–4.

[98] K. N. Trung, E. Dekneuvel, B. Nicolle, et al. "Event detection and disaggregation algorithms for nialm system." In: *Proceedings of 2nd International Non-Intrusive Load Monitoring (NILM) Workshop*. 2014.

[99] J. Li, S. West, and G. Platt. "Power decomposition based on SVM regression." In: *2012 Proceedings of International Conference on Modelling, Identification and Control*. June 2012, pp. 1195–1199.

[100] Y. C. Su, K. L. Lian, and H. H. Chang. "Feature Selection of Non-intrusive Load Monitoring System Using STFT and Wavelet Transform." In: *2011 IEEE 8th International Conference on e-Business Engineering*. Oct. 2011, pp. 293–298. DOI: 10.1109/ICEBE.2011.49.

[101] J. M. Alcalá, J. Ureña, and Á. Hernández. "Event-based detector for non-intrusive load monitoring based on the Hilbert Transform." In: *Emerging Technology and Factory Automation (ETFA), 2014 IEEE*. IEEE. 2014, pp. 1–4.

[102] N. E. Huang and Z. Wu. "A review on Hilbert-Huang transform: Method and its applications to geophysical studies." In: *Reviews of Geophysics* 46.2 (2008).

[103] N. Batra, O. Parson, M. Berges, A. Singh, and A. Rogers. "A comparison of non-intrusive load monitoring methods for commercial and residential buildings." In: *arXiv preprint arXiv:1408.6595* (2014).

[104] N. E. Huang. *Hilbert-Huang transform and its applications*. Vol. 16. World Scientific, 2014.

[105] H. Li, S. Kwong, L. Yang, D. Huang, and D. Xiao. "Hilbert-Huang transform for analysis of heart rate variability in cardiac health." In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 8.6 (2011), pp. 1557–1567.

[106] M. Zou and H. Zhang. "Load forecast based on least square SVR and HHT." In: *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. July 2017, pp. 106–110. DOI: 10.1109/FSKD.2017.8392910.

[107] P. h. Yang and L. h. Yue. "The study on monitor of power system low frequency oscillation base on HHT method." In: *2014 China International Conference on Electricity Distribution (CICED)*. Sept. 2014, pp. 305–307. DOI: 10.1109/CICED.2014.6991716.

[108] D. Y. Y. Li, C. Rehtanz, and R. Xiu. "Analysis of low-frequency oscillations in power system based on HHT technique." In: *2010 9th International Conference on Environment and Electrical Engineering*. May 2010, pp. 289–292. DOI: 10.1109/EEEIC.2010.5489995.

[109] Y. f. Ma, S. q. Zhao, and Y. q. Hu. "Identification of low-frequency oscillations in power systems based on improved HHT." In: *2011 Asia-Pacific Power and Energy Engineering Conference*. Mar. 2011, pp. 1–4. DOI: 10.1109/APPEEC.2011.5748812.

[110] C. Yang, S. Dong, Z. Tong, and Y. Wan. "Analysis and eesearch of the transient composite disturbance signal of power system based on HHT." In: *2008 International Symposium on Intelligent Information Technology Application Workshops*. Dec. 2008, pp. 1029–1032. DOI: 10.1109/IITA.Workshops.2008.268.

[111] F. Eichinger, P. Efros, S. Karnouskos, and K. Böhm. "A time-series compression technique and its application to the smart grid." In: *The VLDB Journal* 24.2 (2015), pp. 193–218.

[112] 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: *2012 IEEE International Energy Conference and Exhibition (ENERGYCON)*. Sept. 2012, pp. 935–940. DOI: 10.1109/EnergyCon.2012.6348285.

[113] M. Zeinali and J. S. Thompson. "Impact of compression and aggregation in wireless networks on smart meter data." In: *Signal Processing Advances in Wireless Communications (SPAWC), 2016 IEEE 17th International Workshop on*. IEEE. 2016, pp. 1–5.

[114] A. Unterweger and D. Engel. "Resumable load data compression in smart grids." In: *Smart Grid, IEEE Transactions on* 6.2 (2015), pp. 919–929. ISSN: 1949-3053.

[115] J. Z. Kolter and M. J. Johnson. "REDD: A public data set for energy disaggregation research." In: *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*. Vol. 25. Citeseer, pp. 59–62.

[116] D. R. Dunn. *Energy Consumtion by Sector in Monthly Energy Review, April 2018*. Tech. rep. US Energy Information Administration, 2018.

[117] G. T. Costanzo, G. Zhu, M. F. Anjos, and G. Savard. "A system architecture for autonomous demand side load management in smart buildings." In: *Smart Grid, IEEE Transactions on* 3.4 (2012), pp. 2157–2165. ISSN: 1949-3053.

[118] H. Shareef, M. S. Ahmed, A. Mohamed, and E. Al Hassan. "Review on home energy management system considering demand responses, smart technologies, and intelligent controllers." In: *IEEE Access* (2018).

[119] Q. Sun, H. Li, Z. Ma, et al. "A comprehensive review of smart energy meters in intelligent energy networks." In: *IEEE Internet of Things Journal* 3.4 (2016), pp. 464–479.

[120] K. Mahmud, M. Hossain, and G. E. Town. "Peak-load reduction by coordinated response of photovoltaics, battery storage, and electric vehicles." In: *IEEE Access* (2018).

[121] M. Aiello and G. A. Pagani. "The smart grid's data generating potentials." In: *Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on*. IEEE. 2014, pp. 9–16.

[122] X. Fang, S. Misra, G. Xue, and D. Yang. "Smart grid: The new and improved power grid: A survey." In: *Communications Surveys & Tutorials, IEEE* 14.4 (2012), pp. 944–980. ISSN: 1553-877X.

[123] *Energy monitors technical survey form*. (last accessed: 13-12-17). URL: `https://www.i13.in.tum.de/`.

[124] F. Leferink, C. Keyer, and A. Melentjev. "Static energy meter errors caused by conducted electromagnetic interference." In: *IEEE Electromagnetic Compatibility Magazine* 5.4 (), pp. 49–55. ISSN: 2162-2264.

[125] *Verdigris*. (last accessed: 13-12-17). URL: `http://verdigris.co`.

[126] *GridSpy*. (last accessed: 13-12-17). URL: `https://gridspy.com/devices.html`.

[127] *CURB Inc.* (last accessed: 13-12-17). URL: `http://energycurb.com/product/`.

[128] N. Iksan, J. Sembiring, N. Haryanto, and S. H. Supangkat. "Appliances identification method of non-intrusive load monitoring based on load signature of VI trajectory." In: *Information Technology Systems and Innovation (ICITSI), 2015 International Conference on*. IEEE. 2015, pp. 1–6.

[129] H. J. Zhang. "Basic concepts of linear regulator and switching mode power supplies." In: *Linear Technology* (2013).

[130] S. Agrawal and J. Agrawal. "Survey on anomaly detection using data mining techniques." In: *Procedia Computer Science* 60 (2015), pp. 708–713.

[131] F. Plesinger, J. Jurco, J. Halamek, and P. Jurak. "SignalPlant: an open signal processing software platform." In: *Physiological measurement* 37.7 (2016), N38.

[132] J. O. Smith. *Mathematics of the discrete Fourier transform (DFT)*. http://www.w3k.org/books/: W3K Publishing, 2007. ISBN: 978-0-9745607-4-8.

[133] L.-K. Shark and C. Yu. "Design of matched wavelets based on generalized Mexican-hat function." In: *Signal Processing* 86.7 (2006), pp. 1451–1469.

[134] N. E. Huang, Z. Shen, S. R. Long, et al. "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis." In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. Vol. 454. 1971. The Royal Society. 1998, pp. 903–995.

[135] N. E. Huang. "Introduction to the Hilbert–Huang transform and its related mathematical problems." In: *Hilbert–Huang transform and its applications*. World Scientific, 2014, pp. 1–26.

[136] T. S. community. *scipy.signal.findpeaks*. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html#scipy.signal.find_peaks.

[137] A. Benjamin. "Music Compression Algorithms and Why You Should Care." In: *Whitepaper: http://ese. wustl. edu/ContentFiles/Research/Unde rgraduateResearch/CompletedProjects/Web-Pages/s u10/AlexBenjamin_AudioCompression. pdf, created December* 9 (2010).

[138] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar. "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey." In: *Sensors* 12.12 (2012), pp. 16838–16866. URL: http://www.mdpi.com/1424-8220/12/12/16838/pdf.

[139] *Extensible Wave-Format Descriptors*. Accessed: 018-05-24. URL: %5Curl%7Bhttps://msdn.microsoft.com/en-us/windows/hardware/drivers/audio/extensible-wave-format-descriptors%7D.

[140] *FLAC - Free lossless audio codec*. Accessed: 2018-05-24. URL: %5Curl%7Bhttps://xiph.org/flac/%7D.

[141] F. Ghido. "An asymptotically optimal predictor for stereo lossless audio compression." In: *Data Compression Conference, 2003. Proceedings. DCC 2003*. IEEE. 2003, p. 429.

[142] D. Salomon and G. Motta. *Handbook of data compression*. Springer Science & Business Media, 2010.

[143] R. Parekh. *Principles of multimedia*. Tata McGraw-Hill Education, 2006.

[144] P. Nishad and R. Chezian. "A survey on lossless dictionary based data compression algorithms." In: *International Journal of Science, Engineering and Technology Research (IJSETR)* 2 (Feb. 2013), pp. 256–261.

[145] E. J. Leavlin and D. A. A. G. Singh. "Article: Hardware implementation of LZMA data compression algorithm." In: *International Journal of Applied Information Systems* 5.4 (Mar. 2013). Published by Foundation of Computer Science, New York, USA, pp. 51–56.

[146] *LZ4*. Accessed: 2018-05-24. URL: %5Curl%7Bhttp://lz4.github.io/lz4/%7D.

[147] *Smaller and faster data compression with Zstandard*. Accessed: 2018-05-24. URL: %5Curl%7Bhttps://code.facebook.com/posts/1658392934479273/smaller-and-faster-data-compression-with-zstandard/%7D.

[148] *Asymmetric numeral systems: entropy coding combining speed of Huffman coding with compression rate of arithmetic coding*. Accessed: 2018-05-24. URL: %5Curl%7Bhttps://dl.dropboxusercontent.com/u/12405967/ans.pdf%7D.

[149] *GNU Gzip: General file (de)compression*. Accessed: 2018-05-24. URL: %5Curl%7Bhttp://www.gnu.org/software/gzip/manual/gzip.html%7D.

[150] S. M. Bhuiyan, R. R. Adhami, and J. F. Khan. "Fast and adaptive bidimensional empirical mode decomposition using order-statistics filter based envelope estimation." In: *EURASIP Journal on Advances in Signal Processing* 2008.1 (2008), p. 728356.

[151] N. E. Huang, M.-L. C. Wu, S. R. Long, et al. "A confidence limit for the empirical mode decomposition and Hilbert spectral analysis." In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 459.2037 (2003), pp. 2317–2345. ISSN: 1364-5021. DOI: 10.1098/rspa.2003.1123. eprint: http://rspa.royalsocietypublishing.org/content/459/2037/2317.full.pdf. URL: http://rspa.royalsocietypublishing.org/content/459/2037/2317.

[152] G. Rilling, P. Flandrin, P. Goncalves, et al. "On empirical mode decomposition and its algorithms." In: *IEEE-EURASIP Workshop on Nonlinear Signal and Image PSrocessing*. Vol. 3. NSIP-03, Grado (I). 2003, pp. 8–11.

[153] P.-Y. Chen, Y.-C. Lai, and J.-Y. Zheng. "Hardware design and implementation for empirical mode decomposition." In: *IEEE Transactions on Industrial Electronics* 63.6 (2016), pp. 3686–3694.

[154] T.-C. Lu, P.-Y. Chen, S.-W. Yeh, and L.-D. Van. "Multiple stopping criteria and high-precision EMD architecture implementation for Hilbert-Huang transform." In: *Biomedical Circuits and Systems Conference (BioCAS), 2014 IEEE*. IEEE. 2014, pp. 200–203.

[155] *emonTx V3*. (last accessed: 13-12-17). URL: http://openenergymonitor.org/emon/modules/emonTxV3.

[156] *emonPi*. (last accessed: 13-12-17). URL: http://openenergymonitor.org/emon/modules/emonpi.

[157] *GreenEye Monitor*. (last accessed: 05-01-18). URL: http://www.brultech.com/greeneye/.

[158] *Energeno Wattson*. (last accessed: 13-12-17). URL: http://smarthomeenergy.co.uk/sites/smarthomeenergy.co.uk/files/Wattson_range_brochure_UK_1.2.pdf.

[159] *HIOKI Clamp on Power Logger*. (last accessed: 13-12-17). URL: https://www.hioki.com/en/products/detail/?product_key=5589.

[160] *TED- The Energy Detective*. (last accessed: 05-01-18). URL: http://www.theenergydetective.com/#.

[161] *Eco-Eye Elite-200*. (last accessed: 13-12-17). URL: http://www.eco-eye.com/product-commercial-monitor-elite-200.

[162] *Eco-Eye Plug-In*. (last accessed: 13-12-17). URL: http://www.eco-eye.com/product-mains-monitor-plug-in.

[163] *Eco-Eye Smart 600*. (last accessed: 13-12-17). URL: http://www.eco-eye.com/product-commercial-monitor-smart-600.

[164] *Blue Line Innovations*. (last accessed: 13-12-17). URL: http://www.bluelineinnovations.com/.

[165] *SEGmeter V2.5*. (last accessed: 05-01-18). URL: https://smartenergygroups.com.

[166] *EFERGY E2 Classic*. (last accessed: 05-01-18). URL: http://efergy.com/media/download/datasheets/e2classicv2_uk_datasheet_web2011.pdf.

[167] *EFERGY Energy Monitoring Socket 2.0*. (last accessed: 13-12-17). URL: http://efergy.com/media/download/manuals/ems_uk_instructions_web2011.pdf.

[168] *Tinytag Energy Logger Kit*. (last accessed: 13-12-17). URL: http://www.geminidataloggers.de/data-loggers/tinytag-energy-data-logger/tge-0001.

[169] *Episensor Wireless Three-Phase Electricity Monitor*. (last accessed: 13-12-17). URL: http://static.episensor.com/wp-content/uploads/ESD-003-00_Data_Sheet_ZEM-61.pdf.

[170] *EKM metering Omnimeter*. (last accessed: 05-01-18). URL: http://documents.ekmmetering.com/EKM_OmniMeter_UL_User_Manual_Spec_Sheet_Submeter.pdf.

[171] *EKM OmiMeter Pulse V.4*. (last accessed: 05-01-18). URL: http://documents.ekmmetering.com/EKM_Metering_LCD_Display_Value_Reading.pdf.

[172] *Smart-Me Metering*. (last accessed: 05-01-18). URL: http://smart-me.com/Description/Products.aspx.

[173] *Neurio Sensor W1*. (last accessed: 05-01-18). URL: http://support.neur.io/customer/en/portal/articles/1847880-neurio-user-manual.

[174] *Pikkerton ZBS 110V2*. (last accessed: 13-12-17). URL: http://www.pikkerton.com/_objects/1/16.htm.

[175] *Digi XBee Smart Plug*. (last accessed: 13-12-17). URL: http://www.digi.com/products/xbee-rf-solutions/range-extenders/xbee-smart-plug-zb#specifications.

[176] *Edimax Smart Plug Switch*. (last accessed: 05-01-18). URL: http://www.edimax.com/edimax/mw/cufiles/files/download/datasheet/SP-2101W_Datasheet_English_EU_type.pdf.

[177] *WattVision*. (last accessed: 13-12-17). URL: %5Curl%7Bhttp://www.wattvision.com/sensors%7D.

[178] *Energeno Wattson XL*. (last accessed: 13-12-17). URL: http://smarthomeenergy.co.uk/sites/smarthomeenergy.co.uk/files/Wattson_range_brochure_UK_1.2.pdf.

[179] *eGauge Main Units*. (last accessed: 13-12-17). URL: http://www.egauge.net/products/.

APPENDIX A

Technical Note

The abbreviations used in the technical note are explained in Table A.0.1. The technical note shown below lists all the information collected from different vendors [125, 126, 65, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179].

Company Name	Product Details	Application Area	Device Type	System Type	Sensor Type	Sensor Rating	Parameters	Sample Freq.	Resolution (bits/Power)	Accuracy	Channels	Storage	Cost[1]
OpenEnergy Monitor	emonTx V3	Residential	Energy Monitor	Single phase	CT, VT	100 A	V, I, P, V_{RMS},	1s-1 min	10 bits/ 10 W	< 100W (>10%) > 100W (<10%) > 150W (<6%) > 250W (<4%) > 500W (< 2%)	4	Requires Base Station	£60
OpenEnergy Monitor	emonPi	Residential	Energy Monitor	Single phase	CT, VT	100 A	V, I, P, V_{RMS}	1s-1 min	10 bits/ 10 W	Same as above	2	Local & Cloud	£155
Brultech Research Inc.	GreenEye Monitor	Residential Commercial	Energy Monitor	Single & three phase	CT, VT	50 A, 100 A, 200 A, 400 A, 600 A, 40	$V, I, P, V_{RMS}, I_{RMS}$	16kHz - 2 kHz	12 bits/ 1 W	1% Channel Accuracy + CT Accuracy (1-3%) 2% (at most)	32	Local, DashBox	$399-$597
GridSpy	GridNode + Hub	Residential Commercial	Energy Monitor	Single & three phase	CT, VT RC, Pulse inputs (from retail meter)	15A, 60A, 200 A, 400 A, 600 A, 800A, 1000A, 2000A, 5000A	$V, I, P, P_{app}, P_{reac}, V_{RMS}, I_{RMS}, \cos \Phi$, Analog inputs, temp.	1s-1 min	16 bits/ 1 W	+/- 1% current or voltage. +/- 2% for wattage	6 per node, 30 per hub, 600 per site	Local-1s data: 6 month Cloud-1 min: forever	NZD $1000 + $600+ = $1,600[2]
Smappee	Smappee	Residential	Energy Monitor	Single & three phase	CT, VT	50 A, 100 A, 200 A	$V, I, P, P_{app}, P_{reac}, V_{RMS}, I_{RMS}, \cos \Phi, E$, Harmonics	> 1KHz	N-A / 1 W	1% (Class 1)	3	Cloud	€229
Smappee	Smappee Pro	Commercial Industrial	Energy Monitor	Single & three phase	CT, VT	50A, 100A, 200A, 400A, 800A	Same as above	16 KHz- 2 kHz	N-A / 1 W	± 1%	9	Cloud	starting at €600
HIOKI E.E Corporation	Clamp on power Logger PW3360-21	Residential Commercial Industrial	Energy Monitor	Single & three phase	CT Voltage code	Many options	$V, I, P, P_{app}, E, P_{reac}, V_{RMS}, I_{RMS}, \cos \Phi$, Harmonics[3]	10.24 kHz	16 bits/ N-A	V: ±0.3%rdg. ±0.1%f.s. I: ±0.3%rdg. ±0.1%f.s.[4] P: ±0.3%rdg. ±0.1%f.s	1-3	Local	€2,629-€3,220[5]
Verdigris	Verdigris	Commercial	Energy Monitor	Single & three phase	CT	30A, 200A, 400A, 60A, 75A, 800A, other custom sizes	V, I, P	7.68 kHz	16 bits/ 10mW	+/-2 A	upto 42 breaker	Cloud storage, 4G LTE	$50-$250 per month
Energy, Inc.	TED - series	Residential Commercial Industrial	Energy Monitor	Single & three phase	VT, CT, RC	20 A, 200 A, 400 A, 60A, 2000A, 5000A	$V, I, P, P_{app}, P_{reac}, V_{RMS}, I_{RMS}, \cos \Phi, SC$	1s	N-A / 1 W	0.3-2%	32	Local & Cloud	$ 299-$ 1499.95
CURB Inc.	CURB Pro, CURB Duo	Residential Commercial Industrial	Energy Monitor[6]	Single & three phase	CT	30 A, 50 A, 100 A (up to 6000 A)	$V, I, P, P_{app}, P_{reac}, V_{RMS}, I_{RMS}, \cos \Phi, SC$	8 kHz (display 1s, 1 min, 1 hr)	N-A/ 1 W	2 %	upto 18 breaker /hub[7]	Local, Cloud, API	$399 (Pro) $749 (Duo)
Eco-Eye	Elite, Mini & Smart	Residential Commercial Industrial	Energy Monitor	Single & three phase	CT	100 A, 200 A	I	4s	12 bits/ 20 W	-	3	Local (up to 128 day)	£30-£100
Blue Line Innovations	PowerCost solution	Residential Commercial	Energy Monitor	Single phase	Optical reading	-	P	-	N-A	95-99%	1	Cloud, Real time display	$179
Smart Energy Groups	SEGmeter v2.5 (complete)	Residential Commercial	Energy Monitor	Single & three phase	CT, VT	60 A	E	1 min	12 bits/ N-A	-	8	Cloud	AUD 649.95

[1] Prices may vary over time
[2] Hub + half node + CTs (if below 60A)
[3] Also includes displacement cos Φ (with lead/lag display), active energy (consumption/regeneration), and reactive energy (lead/lag)
[4] For current (I) and active power (P), also include clamp sensor accuracy
[5] Based on 9661 and 9667-03 CT (discount included)
[6] Also act as gateway for IOT control
[7] Multiple hubs sync per location (e.g., 36, 54, 72, etc.)

Company Name	Product Details	Application Area	Device Type	System Type	Sensor Type	Sensor Rating	Parameters	Sample Freq.	Resolution (bits/Power)	Accuracy	Channels	Storage	Cost[8]
Smart Energy Groups	SEGmeter v2.5 (ready)	Residential Commercial	Energy Monitor	Single & three phase	CT, VT	100 A	E	1 min	12 bits/ N-A	-	8	Cloud	AUD 479.95
EFERGY Technologies Ltd.	E2 Classic, Elite Classic	Residential	Energy Monitor	Single & three phase	CT	100 A, 120	P, E	6s	N-A	> 90%	1-3	Cloud	€84 -€137
Energeno Ltd.	Wattson Classic	Residential	Energy Monitor	Single & three phase	CT	50 A	P	3-20s	N-A/ 1 W	-	1-3	Local up 28 days	£99.95
Energeno Ltd.	Watson XL	Commercial	Energy Monitor	Three phase	CT	200 A	P	3-20s	N-A/ 4 W	-	3	Local	€249. 95
Gemini	Tiny Tag Energy Logger Kit	Residential Commercial Industrial	Energy Monitor	Single & three phase	RC	Device compatible up to 2000 A	P, V_{RMS}, I_{RMS}, $\cos \Phi$	5kHz	N-A/ 0.1-0.01 kWh	I_{RMS}: 1% of the reading ±0.5A (>10A) V_{RMS}: 0.5% of reading, P: 2% of reading, $\cos \Phi$[9]: <0.02	1-3	Local(6 week @5 min)	€920
Episensor	ZEM-30XX, ZEM-61	Residential Commercial Industrial	Energy Monitor	Single & three phase	CT, RC	10A, 80A, 100A, 120A, 300A, 600A, 1000A, 3000A	V, I, P, P_{reac}, V_{RMS}, I_{RMS}, $\cos \Phi$, E	16 kHz - 2 kHz	14/ 1 W	0.50%	1-3	Local (Node 70,000 values), 4GB gateway	€269-€299 Single phase €575-€950 three phase
EKM Metering Inc.	OmniMeter Pulse V.4	Residential Commercial Industrial	Energy Monitor	Single & three phase	CT	100A, 200A, 400A, 600A, 800A, 1500A, 5000A	V, I, P, f, E, P_{reac}, $\cos \Phi$	2520.20 Hz	N-A/ under 50V, 0A, 0W	0.5 %	3	Cloud, local PC, Dash software	$220-$2,116 3-Phase 4-wire system
smart-me	smart-me Meter	Residential Commercial Industrial	Energy Monitor	Single phase	Shunt	32 A, 80 A	V, I, P, $\cos \Phi$	1-15 min, 1s-1 min	N-A/ 0.5 W	1 % (class 1)	1	Local (60 days) Cloud	€271
Neurio	W1, W13P	Residential	Energy Monitor	Single & three phase	CT	200 A	P, E	10 Hz – 1 Hz	N-A/ 1 W, 1 Wh	-	1-3	Cloud	$219.99-$289.99
Pikkerton	ZBS-110V2	Residential	Energy Monitor	Single phase	-	-	V, I, P, $\cos \Phi$, f	N-A	N-A	-	1	-	€180
Eco-Eye	Plug-In	Residential	Smart plug	Single phase	-	-	P	-	N-A/ 0.2 W	-	1	-	£11.88
Digi	XBee Smart Plug	Residential	Smart plug	Single phase	-	-	I, P	-	N-A	-	1	-	$84
EFERGY	Energy Monitoring Socket 2.0	Residential	Smart plug	Single phase	-	-	V, I, P, E, $\cos \Phi$, f	-	N-A	+/- 2%	1	-	€24.90
EDIMAX	SP-2101W	Residential	Smart plug	Single phase	-	-	I, P, E	5s	N-A	+/- 3%	1	Cloud	€43
smart-me	smart-me Plug	Residential Commercial	Smart plug	Single phase	Shunt	16 A	V, I, P, $\cos \Phi$	1s–1 min	N-A/ 0.1 W	1 %	1	Local (60 days) Cloud	€119
Wattvision	Wattvision	Residential Commercial Industrial	Gateway	Single & three phase	Pulse count	-	P	1s – 1 min	N-A / 2 W	2%	whole house	Cloud	$79
eGauge systems	Eg30xx series	Residential Commercial Industrial	Gateway	Single & three phase	CT, VT, RC	20 A, 30 A, 50 A, 100 A, 200 A, 400 A, 600A	V, I, P, P_{app}, P_{reac}, V_{RMS}, I_{RMS}, $\cos \Phi$	N-A	N-A	Overall systems (meter and CT) = - 0.5% accuracy compliant	12 inputs	Local	$500-$800 with (12 CTs)

[8] Prices may vary over time
[9] True above 1 kW

Table A.0.1: Abbreviations used in the technical note.

Parameter	Notation
Voltage	V
Current	I
Real power	P
Apparent power	P_{app}
Reactive power	P_{reac}
Power factor	$\cos\phi$
Energy	E
Frequency	f
RMS voltage	V_{RMS}
RMS current	I_{RMS}
Current transformer	CT
Voltage transformer	VT
Rogowski coil	RC
Side-channel information	SC