Towards a Cost-Effective High-Frequency Energy Data Acquisition System for Electric Appliances

Thomas Kriechbaumer, Anwar Ul Haq, Matthias Kahl, and Hans-Arno Jacobsen
Technische Universität München
Email: thomas.kriechbaumer@in.tum.de

***Abstract*—Traditional energy measurement fails to support consumers in making intelligent decisions to save energy. Non-intrusive load monitoring is one solution to provide disaggregated power consumption profiles. Machine learning approaches rely on public data sets to train parameters for their algorithms – most of which only provide low-frequency appliance-level measurements, thus limiting the available feature space for recognition.**

In this paper, we propose a low-cost measurement system for high-frequency energy data. It extends an off-the-shelf power strip with voltage & current sensors and a single-board PC as data aggregator. We show the architecture and describe a centralized data collector for multiple systems recording concurrently. The self-contained unit for six monitored outlets can achieve up to 25 kHz for all signals simultaneously. A full unit is expected to cost less than 100 €.

I. Introduction

Due to an increased number of household and office appliances, it is infeasible to attach energy meters to each device. Lately, Non-Intrusive Load Monitoring (NILM) has gained wide popularity due to its single point sensing, user-friendly deployment, and power consumption estimations. Appliance-level load profiles allow the user to make intelligent decisions in order to reduce the overall energy usage.

In order to properly validate power disaggregation and appliance identification algorithms, a ground truth is required to provide information regarding actual switching events and power consumption to train models and tune parameters.

Usually smart plugs are utilized to collect appliance-level measurements as ground truth for larger data sets. Unfortunately, most of these low-cost smart plugs come with the inherent problem of having only low-frequency measurement capabilities. This makes it difficult to reconstruct the actual current and voltage waveforms – which is required to track on/off events of appliances with precision. Assuming that only a single appliance is switched on within a small time window can be misleading, and may introduce errors in the case of larger office buildings (large number of appliances) and especially with switch-mode power supplies (fast periodic current spikes).

This paper proposes a high-frequency Data Acquisition system (DAQ) for energy monitoring and ground truth collection. Our proposed system can capture startup & switch-off transients, which can be used for feature extraction, appliance identification, and fault detection. The design is based on an off-the-shelf power strip, a single-board PC for data collection, and sensors to measure electricity signals. All components are contained in a custom enclosure to provide a self-contained high-frequency energy DAQ system for voltage and six current signals.

The rest of this paper is structured as follows. We discuss related work and existing systems in Section II, and subsequently introduce a new system architecture in Section III. The data collection and storage details are explained in Section IV. We show preliminary results and design evaluations in Section V. This paper concludes in Section VI.

II. Related Work

Research in the field of NILM is mostly based on public data sets to analyze and evaluate new algorithms for the task at hand. Multiple research groups have created and published data sets, which can be categorized into low- and high-frequency measurements [1]. In recent years high-frequency measurements have become more popular, due to technological advances in the field of electronics. Still, specialized hardware is required, which usually is quite expensive (up to thousands of Euros).

Correctly identifying appliance transients and signatures is a key step for power disaggregation and appliance identification [2]. This typically requires high sampling frequencies in the range of Kilo-Hertz [3]. Different appliance types can have distinguishing features hidden in the electrical signal waveform. High-frequency voltage and current data carry a vast amount of information depending on the appliance type. Such information was collected and visualized to compare different appliance types on various feature metrics [4].

A high-quality data set for household energy consumption was provided by [5]. The mains signal was sampled with 16 kHz. Appliance-level data (ground truth) were sampled at 1 s – a significant lower frequency, not capable of reconstructing the mains waveforms. This poses problems in the correct matching of appliance transients close to violating the switch continuity principle [6]. This is a common concern with many existing data sets [7] [8] [9].

Using a custom circuit board for voltage and current sensing allows for a compact recording device and provides configurable parameters for rated current and mains frequency [10]. However, this approach still requires an external analog-to-digital converter and is only capable of monitoring a single appliance. Combining multiple monitored power outlets into a single unit would improve cost efficiency, reduce complexity

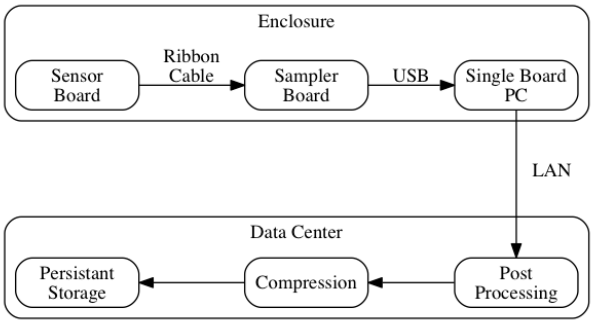

Fig. 1. The architecture of the measurement unit shows the data flow through multiple stages.

of analog-to-digital conversion, and allow for easier data collection and storage.

Lights, refrigerators, printers, and other consumer electronics are not strict resistive loads and generate different patterns. Using an oscilloscope, it is possible to visualize these patterns and analyze the data digitally [11]. However, most oscilloscopes are only capable of displaying values with 8- or 10-bit resolutions for a low number of parallel signals. Oscilloscopes have limited processing power and storage capacities, which is a key requirement for long-time continuous recordings.

Household energy consumption data are an important source for smart grid solutions. Using new forecasting algorithms and disaggregated energy data, these predictions can be of great value to grid operators and expansion plans. Generating forecasts for a large number of households (thousands) is feasible on commodity hardware [12] and can be accurate to a $15\,\mathrm{min}$ time window [13].

III. Architecture

The key requirements for the proposed system are: high sampling frequency (to reconstruct waveforms and harmonics), multiple monitored outlets (in a combined unit), integrated processing unit (with network connectivity), and a centralized data collection and storage back end. The proposed high-level architecture can be seen in Figure 1. This paper focuses on the physical data acquisition, collection, and storage – a real-time data analytics pipeline is deferred to future work.

The design of physical components is based on an off-the-shelf six-port power strip. The current design and all prototypes make use of the *Schuko* socket/plug system (used in most EU countries). The power strip is used as foundation and center piece – all components are designed around it.

The architecture for the DAQ system distinguishes between three main components, which are designed to fit together into a single assembly. Using a custom-built enclosure, the necessary safety requirements are met: touch protection, electric isolation, and rigidity. Once the analog signals are converted to digital values, the samples are packed together and sent to the data center for post-processing and persisting.

The main control unit is a single-board PC, which is connected to the local network via WiFi or ethernet. The data acquisition, including the analog-to-digital conversion (ADC), is handled by custom circuitry attached via USB. Voltage and current signals are generated with an independent circuit board to isolate high- and low-voltage domains.

A. Sensor Board

The sensor board is the only high-voltage component in the system. It contains current sensors and circuitry for voltage sensing. The power strip supplies a live and neutral wire connection. This external power is used to provide a common $5\,\mathrm{VDC}$ power supply for all sensors, ICs, and the single-board PC.

The voltage signal is generated by an AC-AC transformer which acts as galvanic isolator and step-down converter from $230\,\mathrm{V_{RMS}}$ to $6\,\mathrm{V_{RMS}}$. The low-voltage signal is then fed through a series of voltage dividers and a DC-offset injector. Adding a DC-offset is necessary to produce a strictly positive signal for the unipolar ADCs placed on the sampler board. The final signal has a mean of $2.5\,\mathrm{V}$, with peaks at $1.27\,\mathrm{V}$ and $3.39\,\mathrm{V}$.

The six independent current signals (one for each socket) are generated by a Hall effect-based integrated circuit, which generates an analog voltage proportional to the current flowing across the chips primary pins. Depending on the use case, the system can be built with ICs calibrated to different sensitivity settings, allowing sensing of currents from 5, 20, and $30\,\mathrm{A_{peak}}$. The chip we selected, Allegro ACS712, yields $2.5\,\mathrm{V}$ if no current is flowing, and it can range from $0.5\,\mathrm{V}$ to $4.5\,\mathrm{V}$ for the negative and positive AC half-waves.

A 10-pin connector (ribbon cable) is used to wire all sensor signals to the sampler board on the opposite side of the power strip. This connector is also used to provide a common supply voltage for all components. This separates the high-voltage connections (on the sensor board) from the low-voltage components (on the sampler board and single-board PC).

B. Sampler Board

The main tasks of the sampler board are to convert the analog sensor output into digital data and transfer them via USB to a single-board PC. The involved components are eight individual ADCs, a microcontroller, and a dedicated USB connectivity chip. This board is designed to fit on top of the single-board PC and make use of the general purpose I/O pins. This creates a compact unit without any loose wires or plugs.

The ADCs treat each signal independently, while the digital output of all converters is grouped into an 8-bit bus directly attached to the microcontroller. Using a 12-bit ADC resolution, the microcontroller needs to read 12 bytes. Prepending a 16-bit checksum, each sample is transferred to the USB connectivity chip without any further processing. The microcontroller's sole responsibility is to get data from the ADCs and send them to the USB chip.

The USB connection from the sampler board to the single-board PC is deliberately kept simple, to reduce any additional computationally heavy-lifting on the single-board PC. Recorded sample packets are received via USB bulk transfer

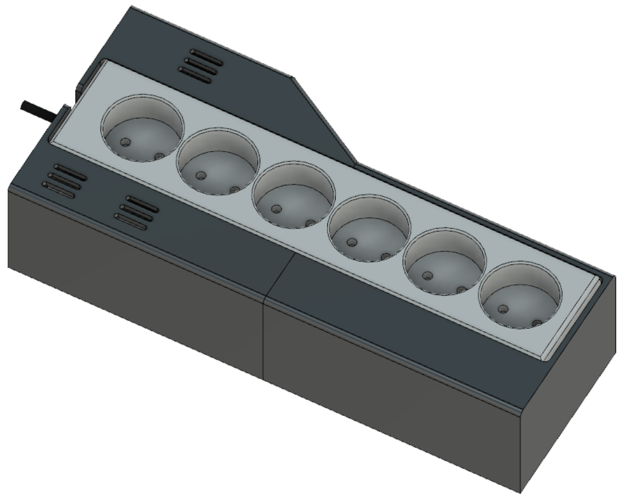

Fig. 2. The fully assembled measurement unit with the power strip in the center and the 3D-printed enclosure.

in the Linux kernel. Control commands are sent over a serial communication link, based on a simplified Hayes command set[1].

The main components are eight MCP3201 for analog-to-digital conversion (12-bit, unipolar, $0\,\mathrm{V}$ to $4.096\,\mathrm{V}$), an ATmega324PA microcontroller for the control logic, an FTDI232H for USB data transfers, and a Raspberry Pi B+ for buffering and network connectivity. The ICs and other electronics have an estimated price of less than 65 €.

C. Enclosure

The six-port power strip in the center is implanted within a custom two-part enclosure to protect and shield the electronics inside. Fig. 2 shows a rendering of the fully assembled measurement unit. On the top left corner is the Raspberry Pi, with the network port accessible. On the bottom (over the full length of the power strip) is the sensor board enclosed with only two small cooling openings in the case.

The case design makes use of existing design features of the single-board PC and power strip. Existing mounting holes and screws can be reused to assemble all components. Special seals and plastic lips around the seam act as touch guards and protect from high-voltage contacts. The internals of the power strip are incorporated in the sensor board layout to minimize wire lengths and copper paths.

The custom 2-part enclosure is 3D-printed (see Fig. 3) and costs less than 15 €. The manufacturing for a fully self-contained measurement system are expected to be below 100 €.

IV. Data Collection and Storage

Since most single-board PCs are equipped with a single USB interface, all data has to pass through this bottleneck. This means receiving data from the sampler board must not interfere with sending the recorded data over the network. With careful selection of buffer sizes and queue lengths the transfer in both directions can be guaranteed to be without data loss. This could also be achieved by using a real-time operating system, but this requires additional efforts for network connectivity and USB bulk transfer handling.

Assuming a sampling frequency of $5\,\mathrm{kHz}$, the input from the sampler board is approx. $70\,\mathrm{KiB/s}$. The single-board PC stores the data in chunks (e.g., $128\,\mathrm{MiB}$, $256\,\mathrm{MiB}$, $512\,\mathrm{MiB}$) in the internal memory before periodically sending them to the data center. In order to assure uninterrupted transfer, the outgoing bandwidth must be limited to not impair the incoming data from the sampler board. Preliminary tests show that a bandwidth corridor between $500\,\mathrm{KiB/s}$ and $1000\,\mathrm{KiB/s}$ yields the expected result.

Since computational power is limited on a single-board PC, all post-processing of data has to be done after collecting the data off-board. This creates the opportunity to convert the data into a multiple common file formats, apply compression, and add redundancy for error correction.

The microcontroller stores each sample as 14-byte packets: 12-bit for eight ADCs generate 12 bytes, plus 2 bytes for the current sample ID. This structure needs to be converted with multiple bit operations to extract the actual digital values from the analog signal. The sample ID overflows at a modest rate (depending on the sampling frequency), so it can be used to detect data loss.

Finding a suitable data format to store high-frequency data is an interesting research topic in its own right, however, based on the NILMTK project [14], we currently support HDF5 as the default storage format. Using HDF5 allows for a structured and easy to read file structure. Most common programming

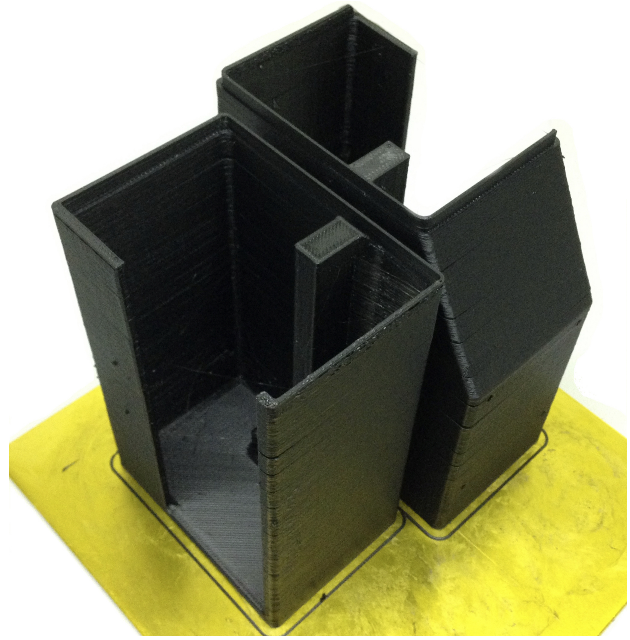

Fig. 3. The enclosure printed in two parts with an FDM-based printer – a Prusa i3 derivative – a full print takes approx. 10 hours.

[1]Hayes Command Set Reference: http://nemesis.lonestar.org/reference/telecom/modems/at/summary-at.html

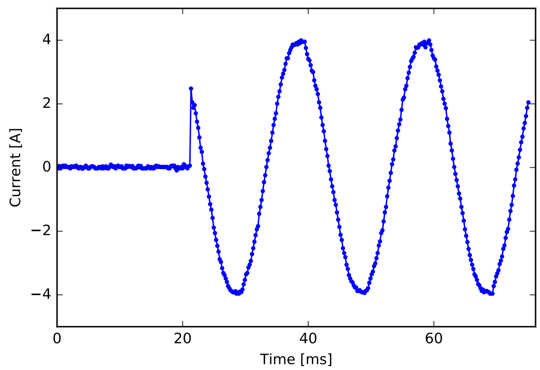

Fig. 4. This graph shows the recorded current trace of a room heater with approx. $650\,\mathrm{W}$. After $20\,\mathrm{ms}$ the appliance was switched on, resulting in an almost instantaneous current draw.

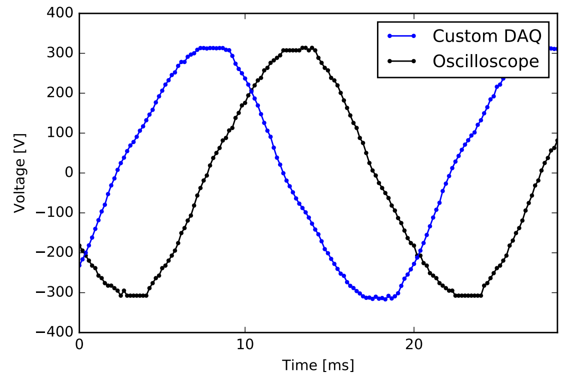

Fig. 5. The recorded high-frequency voltage signal by our unit compared to an off-the-shelf oscilloscope recording. For easier comparison, the oscilloscope signal is shifted in time by +7 ms.

languages and scientific computing software packages support HDF5 files natively or provide libraries to import them efficiently. Depending on post-processing pipeline, it might be reasonable to compress all files before persisting them to disk, or use the intrinsic HDF5 data set compression.

Each file carries a list of attributes: timestamp (UNIX time with microsecond precision), sampling frequency (Hertz), a measurement unit identifier (UUID), and a sequence number of the total recording this file is part of. Every signal stream (voltage and current signals) is stored as individual data set with corresponding attributes: calibration factor and removed DC-offset.

Assuming a raw chunk of $256\,\mathrm{MiB}$ (approx. 1 hour @ $5\,\mathrm{kHz}$) the resulting HDF5 file is the same size. During file parsing and converting to HDF5, the samples are checked for gaps and data loss, before being compressed and persisted to permanent storage.

V. Evaluation

Each ADC uses a unipolar fixed reference of $4.095\,\mathrm{V}$. This allows a value range from 0 to 4095 using a 12-bit resolution. The current sensors are rated for $1.575\,\mathrm{V}$ to $3.425\,\mathrm{V}$ (ACS712-05B) and $0.52\,\mathrm{V}$ to $4.48\,\mathrm{V}$ (ACS712-30A). This results in clipping of the positive half-wave (only on the ACS712-30A), reducing the maximum measurable current to $27.4\,\mathrm{A}$. Since the mains fuse triggers at $16\,\mathrm{A}$, this is not an issue. The average noise of $0.034\ \mathrm{A_{RMS}}$ (ACS712-05B) and $0.046\ \mathrm{A_{RMS}}$ (ACS712-30A) is sufficient to detect appliances as low as $10\,\mathrm{W}$.

As a benchmark test, a room heater with approx. $650\,\mathrm{W}$ was used to test the current signal accuracy. The reconstruction of the mains frequency ($50\,\mathrm{Hz}$ sinusoidal) is clearly visible in Fig. 4. The Hall effect-based sensor has a range of $\pm 5\ \mathrm{A_{peak}}$ according to the data sheet. However, our results show that current spikes up to $\pm 8.63\ \mathrm{A_{peak}}$ can be measured accurately. The ADC starts to clip above that threshold.

The recorded voltage signal was compared against data from an oscilloscope (Hantek 6022BE). The waveform is again clearly visible in Fig. 5, and matches the oscilloscope's data. The sampling frequency of our system can be adapted in the range of $250\,\mathrm{Hz}$ up to $25\,\mathrm{kHz}$ of simultaneously sampling for all signals with 12-bit resolution.

Preliminary compression tests (with lossless data compression algorithms) showed that the `lzma2` algorithm with a `xz` container yields the best compression factor. A $256\,\mathrm{MiB}$ HDF5 file in its compressed state only requires approx. $70\,\mathrm{MiB}$ – only 30% of the original size. Since high-frequency data signals have a repeating structure (12-bit groups, limited range of input values), tweaking compression parameters has potential and is deferred to further research.

VI. Conclusions

We have shown a new high-frequency data acquisition system design, which can be used as a cost-effective measurement system for electric appliances. This is especially useful in the context of NILM and appliance identification.

The presented modular design allows the research community to adapt and customize to their specific needs (power outlet shape, mains voltage and frequency). The sampling frequency and signal resolution show a vast improvement compared to existing data sets, which only provide ground truth data with a second-interval.

We would like to encourage the use of a high-frequency ground truth for future research and new data sets. This could improve data quality for various NILM-related subtasks and similar research fields.

References

[1] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey," *Sensors*, vol. 12, no. 12, p. 16838, 2012.

[2] K. C. Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? The case of electricity," *Energy Policy*, vol. 52, pp. 213–234.

[3] A. S. Bouhouras, A. N. Milioudis, and D. P. Labridis, "Development of distinct load signatures for higher efficiency of NILM algorithms," *Electric Power Systems Research*, vol. 117, pp. 163 – 171.

[4] J. Gao, S. Giri, E. C. Kara, and M. Bergés, "PLAID: A Public Dataset of High-resolution Electrical Appliance Measurements for Load Identification Research: Demo Abstract," in *ACM BuildSys '14*, 2014, pp. 198–199.

[5] J. Kelly and W. Knottenbelt, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," *Scientific Data*, vol. 2, no. 150007.

[6] S. Makonin, "Investigating the Switch Continuity Principle Assumed in Non-Intrusive Load Monitoring," in *IEEE CCECE '16*, 2016.

[7] J. Z. Kolter and M. J. Johnson, "REDD: A Public Data Set for Energy Disaggregation Research," in *SustKDD '11*, vol. 25, pp. 59–62.

[8] K. Anderson, A. Ocneanu, D. Benítez, D. Carlson, A. Rowe, and M. Bergés, "BLUED: A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research," in *SustKDD '12*, pp. 1–5.

[9] M. Bergés, E. Goldman, H. S. Matthews, and L. Soibelman, *Learning Systems for Electric Consumption of Buildings*.

[10] F. Englert, T. Schmitt, S. Kößler, A. Reinhardt, and R. Steinmetz, "How to Auto-configure Your Smart Home? High-resolution Power Measurements to the Rescue," in *ACM e-Energy '13*, pp. 215–224.

[11] W. Lee, G. Fung, H. Lam, F. Chan, and M. Lucente, "Exploration on Load Signatures," in *ICEECS '04*, vol. 2.

[12] A. Veit, C. Goebel, R. Tidke, C. Doblander, and H.-A. Jacobsen, "Household electricity demand forecasting: Benchmarking state-of-the-art methods," in *Proceedings of the 5th International Conference on Future Energy Systems*, ser. e-Energy '14. New York, NY, USA: ACM, 2014, pp. 233–234.

[13] H. Ziekow, C. Doblander, C. Goebel, and H.-A. Jacobsen, "Forecasting household electricity demand with complex event processing: Insights from a prototypical solution," in *Proceedings of the Industrial Track of the 13th ACM/IFIP/USENIX International Middleware Conference*, ser. Middleware Industry '13. New York, NY, USA: ACM, 2013, pp. 2:1–2:6.

[14] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava, "NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring," in *ACM e-Energy '14*.