UK-DALE: A dataset recording UK Domestic Appliance-Level Electricity demand and whole-house demand

Jack Kelly and William Knottenbelt
Department of Computing, Imperial College London, UK
jack.kelly@imperial.ac.uk

***Abstract*—Many countries are rolling out smart electricity meters. These measure a home's total power demand. However, research into consumer behaviour suggests that consumers are best able to improve their energy efficiency when provided with itemised, appliance-by-appliance consumption information. Energy disaggregation is a computational technique for estimating appliance-by-appliance energy consumption from a whole-house meter signal.**

To conduct research on disaggregation algorithms, researchers require data describing not just the aggregate demand per building but also the 'ground truth' demand of individual appliances. We present 'UK-DALE': an open-access dataset from the UK recording Domestic Appliance-Level Electricity at a sample rate of 16 kHz for the whole-house and at $^{1}/_{6}$ Hz for individual appliances. This is the first open access UK dataset at this temporal resolution. We recorded from four homes, one of which was recorded for 499 days, the longest duration we are aware of for similar datasets. We also describe the low-cost, open-source, wireless system we built for collecting our dataset.

I. BACKGROUND & SUMMARY

Energy disaggregation researchers require access to large datasets recorded in the field in order to develop and test disaggregation algorithms but it is not practical for every researcher to record their own dataset. Hence the creation of open access datasets is key to promote a vibrant research community.

Researchers at MIT led the way by releasing an open-access smart meter dataset in 2011 [4] and more datasets have subsequently been released by researchers around the world. At the time of writing, the only open-access disaggregated dataset recorded in the UK is the DECC/DEFRA Household Electricity Study [2] which has a sample period of 2 minutes.

We present the first open access UK dataset with a high temporal resolution. We recorded from four homes. Every six seconds we recorded the active power consumed by individual appliances and the whole-house apparent power consumption. Additionally, in two homes, we sampled the whole-house voltage and current at 44.1 kHz (down-sampled to 16 kHz for storage) and also calculated the real power, reactive power and RMS voltage at 1 Hz. In home one, we recorded for 499 days and individually recorded from almost every single appliance in the home resulting in a recording of 54 separate channels (although less channels were recorded towards the start of the dataset). We will continue to record from this home for the foreseeable future. We recorded from three other homes for several months; each of these homes recorded between 5 to 20 channels. Figure 1 provides an overview of the system design and Table I provides summary statistics about the dataset.

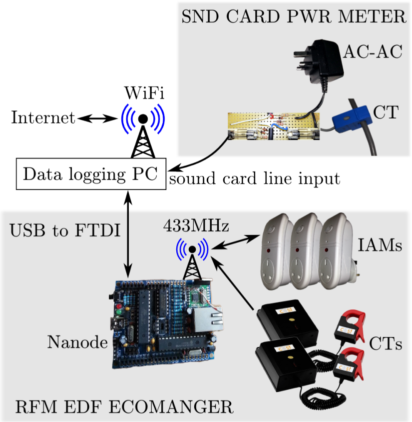

Fig. 1. The system diagram and the three major components of the system: (top left) the data logging PC; (top right) the sound card power meter (which uses the sound card on the data logging PC to record the output from an AC-AC adaptor and a current transformer (CT)) and (bottom) the 'RFM EDF ecomanager' which uses a Nanode to communicate over the air with a set of individual appliance monitors (IAMs) and current transformer (CT) sensors.

This dataset may also be of use to researchers working on modelling the electricity grid; exploring the potential for automated demand response or researching appliance usage behaviour.

A longer version of this paper is available [6].

II. METHODS

We first describe our approach to monitoring individual appliances once every 6 seconds and then describe how we record whole-house mains power at 44.1 kHz.

A. *Individual appliance monitoring*

In UK houses such as those in our dataset, mains 'rings' extend from the fuse box. Many sockets share the same ring. Hence, in order to measure individual appliances in the UK,

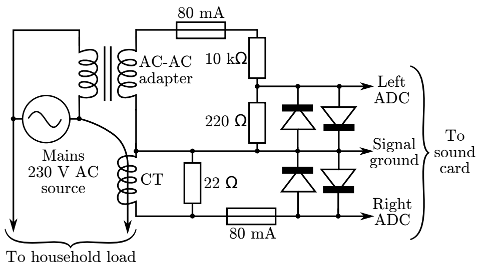

Fig. 2. Circuit diagram for interfacing a sound card to a CT clamp and AC-AC adaptor to measure mains current and voltage, respectively. Adapted from [7]. Each diode is a 1N5282 (1.3 V forward voltage bias).

we must install plug-in individual appliance monitors (IAMs) between each appliance and its wall socket.

The IAMS we use are ‘EcoManager Transmitter Plugs’ developed by Current Cost and distributed by EDF Energy. The standard base station for these IAMs is the EcoManager. The EcoManager can only handle a maximum of 14 transmitter plugs and only provides data once a minute via its serial port. We need up to 54 appliances monitored per home and data every 10 seconds or faster. So, we set about building our own base station. With the help of others in the community, we reverse-engineered the EcoManager protocol.

With the reverse engineered protocol in hand, we built our own base station by programming an open-source, rapid-development platform called the ‘Nanode’[1]. The Nanode includes an ATmega328P microcontroller running at 16MHz (the same microcontroller as used on several Arduinos) and a HopeRF RFM12b radio frequency module[2]. The EcoManager products appear to use the same (or similar) RF module tuned to the 433MHz ISM band[3]. We found the appropriate starting point for our RF configuration settings by using a Bus Pirate[4] to sniff configuration packets from the Serial Peripheral Interface (SPI) connecting the EcoManager’s microcontroller to its RF module. Our base station polls each IAM in sequence.

To measure power from hard-wired appliances such as boilers and kitchen ceiling lights, we used Current Cost transmitters with current transformer (CT) clamps[5]. These transmitters use the same radio frequency as the EDF IAMs but a different protocol. In particular, the Current Cost transmitters cannot *receive* RF data. Instead they transmit a data packet once every 6 ± 0.3 seconds, whether or not the RF channel is clear. Hence RF collisions are inevitable and there is no mechanism to request re-transmission of lost data. As such, our base station minimises the chance of packet collisions by learning the transmit period of each Current Cost TX and ensuring that the base station does not transmit for a short window of time prior to the expected arrival of a Current Cost TX packet. Current Cost transmitters do not use a checksum but instead use Manchester encoding hence corrupted packets are unlikely but by no means impossible.

B. *Measuring whole-house power consumption*

The latest iteration of the UK smart meter specification [1] is detailed enough to allow us to build our own metering system which closely mimics what a UK smart meter is likely to provide. The latest specifications [1] state that smart meters are required to connect to the Home Area Network (HAN) using ZigBee Smart Energy Protocol v1. A disaggregation system would probably access smart meter data by way of a ‘Consumer Access Device’ (CAD) connected to the HAN. CADs can request instantaneous active power data from the meter once every 10 seconds.

We set out to build a metering system that would collect active power once a second, as well as to sample the voltage and current waveforms at 44.1 kHz.

One solution would be to use an off-the-shelf Current Cost whole-house transmitter with a current transformer (CT) clamp. These work with our wireless base station. We used this solution in several houses where our bespoke solution was impractical. There are several disadvantages to using a CT clamp connected to a wireless transmitter:

- CT clamps measure current (I). The transmitter usually has no way to measure voltage and so must use a hard-coded value for voltage (V) to calculate a power reading (P) using $P = I \times V$. However, mains voltage in the UK is allowed to vary by +10 % to -6 % (sometimes quite abruptly) so power readings for a linear resistive load could vary over the range +20 % to -12 % (as noted by Hart [3]). To see this, note that $P = I \times V$ and current changes with voltage for resistive loads according to Ohm’s law ($I = {}^{V}\!/_{R}$). These abrupt changes in reported power due to external noise are problematic for disaggregation algorithms because disaggregation algorithms tend to rely on changes in power consumption to detect appliances turning on or off.
- Battery powered transmitters tend to sparsely sample from their CT clamp in order to minimise battery usage. Hence rapid changes may be missed.
- Without instantaneous measurements of both voltage and current, it is not possible to measure active power or reactive power. Hence CT clamps without voltage measurements only report an estimate for apparent power.
- The OpenEnergyMonitor emonTx[6] is capable of getting round all three disadvantages mentioned above. However, the emonTx uses an analogue to digital converter with only 10 bits of resolution. If we want to measure a primary current which varies from, say, 0 to 30 amps then the emonTx can only resolve changes larger than 14 watts[7]. ‘Real’ smart meters will almost certainly have considerably higher resolution so, unfortunately, the current generation of the emonTx is not a suitable proxy for a ‘real’ smart meter[8].

---

[1] http://www.nanode.eu/

[2] http://www.hoperf.com/rf_fsk/fsk/21.htm

[3] It is not legal to use the 433MHz band without a license in some countries.

[4] http://dangerousprototypes.com/docs/Bus_Pirate

[5] http://www.currentcost.com/product-transmitter.html

[6] http://openenergymonitor.org/emon/emontx

[7] The emonTx uses 10 bits of resolution to capture both the +ve and -ve sides of the AC signal so in effect it uses only 9 bits to cover a range of 30 amps. $30\,\mathrm{A} \div 2^9$ ADC steps $= 0.06\,\mathrm{A}$ per ADC step so it can resolve changes in current $\geq 0.06\,\mathrm{A}$. And $0.06\,\mathrm{A} \times 230\,\mathrm{V} = 13.8\,\mathrm{W}$.

[8] The OpenEnergyMonitor developers are working on a higher resolution monitor.

No existing home energy monitor that we are aware of will provide an accurate proxy for UK smart meters. Expensive power quality monitors costing at least several hundred UK pounds can measure with the accuracy we require but these are prohibitively expensive and some require CT sensors *without* a split core, hence requiring the installer to disconnect the meter tails from the utility company's meter, which can only be done with permission from the utility company.

We propose a low-cost, high resolution, easy to install technique for recording whole-house mains power using a computer sound card, a CT clamp and an AC-AC adapter.

Typical sound cards have remarkably good analogue to digital converters (ADCs). Typical specifications of a modern sound card include 96 kHz sample rate; simultaneous recording of at least 2 channels; 90 dB signal to noise; 20 bits per sample; built-in high-pass filter and built-in anti-alias filter. Given that each bit provides 6 dB of dynamic range, we effectively have $^{90}\!/_{6} = 15$ bits of 'signal' and $20 - 15 = 5$ bits of 'noise'.

To record mains voltage and current waveforms we built a simple circuit to connect the sound card to an AC-AC adapter and a CT clamp (Figure 2). This circuit does not require the user to handle any hazardous voltages. We used the line-input of the sound card rather than the microphone input because the line-input should provide a lower noise signal path than the sound card's microphone pre-amplifier. The standard maximum peak-to-peak voltage for consumer audio equipment line-input is 0.89 volts. Hence the aim of our circuit must be to reduce the output voltage of each sensor so that we never deliver more than 0.89 volts to the sound card.

To measure mains voltage as safely as possible, we used a standard AC-AC adapter[9] (a 'wall wart'). This provides a peak open-circuit output voltage of approximately 11 volts. Research done by the Open Energy Monitor project suggests that the output of the AC-AC adapter should track the mains input voltage linearly over the range 185.5 V to 253 V[10]. We reduce the AC-AC adapter's output voltage with a voltage divider circuit (with two resistors: $10\,\mathrm{k\Omega}$ and $220\,\Omega$) to produce about 0.7 V peak-to-peak which is fed into one channel of the sound card's line input.

To measure mains current, we used a current transformer (CT) clamp[11]. The CT is connected in parallel to a $22\,\Omega$ burden resistor. This configuration produces about 0.89 V peak-to-peak across the burden resistor when the CT is presented with a primary current of 30 amps RMS which, we believe, is the most current that any of our homes under study will pull.

To protect the sound card against overload, both channels include an 80 mA quick-blow fuse and a pair of 1N5282 diodes (with a 1.3 V forward voltage bias) to ensure that the circuit is unlikely to ever deliver more than 1.3 V to the sound card.

Let us calculate a rough estimate for our measurement resolution. If we want to measure a primary current with a range of 0 to 30 $\mathrm{A_{rms}}$ then we should be able to resolve changes in primary current of approximately $3\,\mathrm{mA}$ per sample[12]. For the voltage measurement, if we want a range of 0 to $253\,\mathrm{V_{rms}}$ ($230\,\mathrm{V_{rms}} + 10\,\%$) then we should be able to resolve changes of approximately $22\,\mathrm{mV}$ per sample[13]. Given that the sensors are likely to be noisy and given that we are only providing $0.7\,\mathrm{V_{peak\text{-}to\text{-}peak}}$ to the ADC for the voltage measurements, we should downgrade our resolution per sample to about $30\,\mathrm{mV}$ and $5\,\mathrm{mA}$ for voltage and current respectively. This gives us a resolution for power of approximately $30\,\mathrm{mV} \times 5\,\mathrm{mA} = 150\,\mathrm{mW}$.

We now describe the software for our 'sound card power meter'. We simultaneously record vectors of voltage and current readings (we record in 1 second chunks; this time period was chosen simply because REDD uses this sample period for mains data). To calculate $|S|$ (apparent power) we use $|S| = I_{\mathrm{rms}} \times V_{\mathrm{rms}}$ where I_{rms} and V_{rms} are the root mean squared values for the current and voltage vectors respectively. To calculate P ('real' or 'active' power) we use $P = \frac{1}{N}\sum_{i=1}^{N} I_i V_i$ where N is the number of samples; I_i and V_i are the i^{th} samples of the current and voltage vectors respectively.

The conclusion is that we achieve a resolution greater than that required to provide a good proxy for 'real' smart meters[14]. We save P, $|S|$ and V_{rms} to disk once a second with a precision of 2 decimal places in a CSV file.

We also save the raw ADC data to disk. To reduce the space required, the ADC data are down-sampled using the open-source audio tool 'sox'[15] to 16 kHz. The uncompressed 16 kHz 24-bit files would require 28.8 GBytes per day so we compress the files using the Free Lossless Audio Codec (FLAC) to reduce the storage requirements to ≈ 4.8 GBytes per day.

To convert the raw ADC values into voltage and current readings, we first need to find appropriate conversion constants. We need to calibrate each data collection system separately to compensate for manufacturing variability in the components. We calibrate each system once when the system is first setup. We connect a 'Watts up? PRO meter'[16] to the data logging PC via USB during setup to automatically calibrate voltage and current conversion factors. We typically use a resistive load like a kettle to calibrate the system.

We have implemented the power monitoring system described as five software projects. All software packages are available from https://github.com/JackKelly/<*name*> where *name* is one of {`rfm_edf_ecomanager`, `rfm_ecomanager_logger`, `powerstats`, `babysitter`, `snd_card_power_meter`}.

---

[9] the 'Ideal Power 77DB-06-09'

[10] http://openenergymonitor.org/emon/buildingblocks/report-mascot-9v-acac-adaptor

[11] The 'YHDC SCT-013-000'. Both the CT clamp and AC-AC adapter were sourced from the Open Energy Monitor shop: http://shop.openenergymonitor.com/components/

[12] $30\,\mathrm{A_{rms}} \times \sqrt{2} \times 2 \approx 85\,\mathrm{A_{peak\text{-}to\text{-}peak}}$ and $85\,\mathrm{A_{peak\text{-}to\text{-}peak}} \div 2^{15}$ADC steps $\approx 3\,\mathrm{mA}$.

[13] $253\,\mathrm{volt_{rms}} \times \sqrt{2} \times 2 \approx 716\,\mathrm{V_{peak\text{-}to\text{-}peak}}$ and $716\,\mathrm{V_{peak\text{-}to\text{-}peak}} \div 2^{15}$ADC steps $\approx 22\,\mathrm{mV}$.

[14] Although we do not know the precise resolution of 'real' smart meters. This decision is likely to be left to the manufacturers [personal communication with DECC, March 2013].

[15] http://sox.sourceforge.net

[16] http://www.wattsupmeters.com/secure/products.php?pn=0

C. Complete metering setup

To collect our own dataset, we installed the following equipment in each home:

- Multiple EDF Individual Appliance Monitors.
- A CurrentCost CT clamp & transmitter to measure whole-house apparent power. Home 1 used additional CC CT clamps to measure the lighting circuit, kitchen ceiling lights, boiler and solar hot water pump.
- A Nanode running our rfm_edf_ecomanager code.
- A small-footprint Atom PC[17] using the Intel DN2800MT motherboard (with a Realtek ALC888S audio codec capable of sampling at 96 kHz at 20-bit resolution with a signal to noise ratio of 90 dB; and a line-input socket on the rear) and a 320 GB HDD; runs Ubuntu Linux Server; consumes 14 watts active power.
- Homes 1 and 2 had the 'sound card power meter' system installed to measure whole-home active and reactive power and voltage.

A system diagram is shown in Figure 1.

III. DATA FORMAT

UK-DALE uses a data similar format to REDD [4].

One way in which UK-DALE differs from REDD is that UK-DALE includes a detailed metadata file which follows the NILM Metadata schema [5]. This metadata describes properties such as the specifications of most appliances; the mains wiring between the meters and between meters and appliances; exactly which measurements are provided by each meter; which room each appliance belongs in etc. The `labels.dat` file in each directory is redundant and is included to provide compatibility with REDD.

IV. DATASET

Table I shows statistics describing the UK-DALE dataset. This table shows that house 1 has the most coverage both in terms of numbers of meters and in terms of time span.

The data are freely available for download[18].

Figure 3 shows the measurement errors for our 'sound card power meter' and the Current Cost Current Transformer (CT) sensor across a range of resistive loads. The ground truth was measured using a Watts Up meter. The range of loads were created by using three incandescent lamps and by changing the number of primary turns on the CT from one to seven. For each load, we recorded one minute of data and took the largest error from that minute of data. The sound card power meter was calibrated 6 months prior to the test. This illustrates that our 'sound card power meter' consistently produces a relative error of less than 2% and that the Current Cost CT meter produces errors of less than 6% as long as the power is above 100 watts (the whole-home power consumption very rarely drops below 100 watts).

TABLE I. SUMMARY OF THE DATA AVAILABLE FOR EACH HOUSE.

ID	Number of meters[a]	Mains[b] sample rate	Time span[c] (days)	Uptime[d] (days)
1	54	16 kHz & 6 sec	499	470
2	20	16 kHz & 6 sec	234	199
3	5	6 sec	39	36
4	6	6 sec	205	205

[a] The maximum number of meters used (including mains meters). Some houses started with a small number of meters to test the system and then added more.

[b] Indicates whether we recorded the mains current and voltage waveform at 16 kHz using our sound card power meter (as well as at 6-second intervals using the Current Cost CT clamp).

[c] Difference between the first and last timestamps.

[d] Total duration that the system was recording.

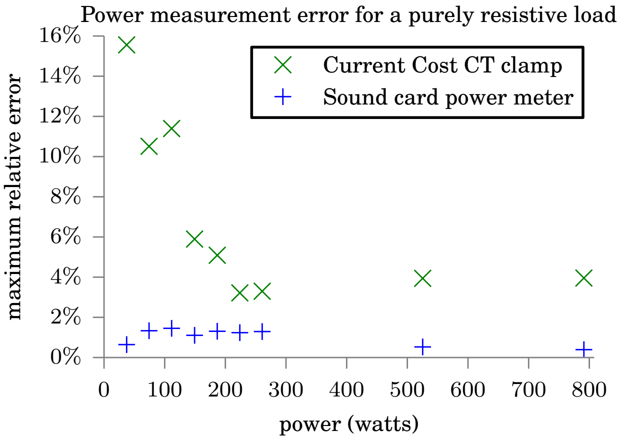

Fig. 3. A plot of the maximum relative measurement error for power measurements across a range of loads.

REFERENCES

[1] DECC. Smart Metering Equipment Technical Specifications: version 2. Technical report, UK Government, London, 2012. URL: http://bit.ly/1401o61.

[2] DEFRA. Household Electricity Study - EV0702. Technical report, UK Government, London, 2012. URL: http://bit.ly/MnIwGE.

[3] G. W Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, December 1992. `doi:10.1109/5.192069`.

[4] J. Zico Kolter and Matthew J. Johnson. REDD: A public data set for energy disaggregation research. San Diego, California, 2011. URL: http://redd.csail.mit.edu/kolter-kddsust11.pdf.

[5] Jack Kelly and William Knottenbelt. Metadata for Energy Disaggregation. In *The 2nd IEEE International Workshop on Consumer Devices and Systems (CDS 2014)*, Västerås, Sweden, July 2014. `arXiv:1403.5946`.

[6] Jack Kelly and William Knottenbelt. UK-DALE: A dataset recording UK Domestic Appliance-Level Electricity demand and whole-house demand. *ArXiv e-prints*, 2014. `arXiv:1404.0284`.

[7] Robert Wall. Report: Ideal Power 9V AC/AC Adaptor | OpenEnergyMonitor. URL: http://openenergymonitor.org/emon/buildingblocks/report-idealpower-9v-acac-adaptor.

---

[17] Full component listing of the Atom PCs we built can be found at http://jack-kelly.com/intel_atom_notes and a guide to setting up a complete data logging system can be found at https://github.com/JackKelly/rfm_ecomanager_logger/wiki/Build-a-complete-logging-system

[18] http://data.ukedc.rl.ac.uk/cgi-bin/dataset_catalogue/view.cgi.py?id=18 and http://data.ukedc.rl.ac.uk/cgi-bin/dataset_catalogue/view.cgi.py?id=19