

# REDD: A Public Data Set for Energy Disaggregation Research

J. Zico Kolter  
Computer Science and Artificial Intelligence  
Laboratory  
Massachusetts Institute of Technology  
Cambridge, MA  
kolter@csail.mit.edu

Matthew J. Johnson  
Laboratory for Information and Decision Systems  
Massachusetts Institute of Technology  
Cambridge, MA  
mattjj@csail.mit.edu

## ABSTRACT

Energy and sustainability issues raise a large number of problems that can be tackled using approaches from data mining and machine learning, but traction of such problems has been slow due to the lack of publicly available data. In this paper we present the Reference Energy Disaggregation Data Set (REDD), a freely available data set containing detailed power usage information from several homes, which is aimed at furthering research on energy disaggregation (the task of determining the component appliance contributions from an aggregated electricity signal). We discuss past approaches to disaggregation and how they have influenced our design choices in collecting data, we describe the hardware and software setups for the data collection, and we present initial benchmark disaggregation results using a well-known Factorial Hidden Markov Model (FHMM) technique.

## 1. INTRODUCTION

Energy and sustainability problems represent one of the greatest challenges facing society. More than 83% of the world's energy comes from (unsustainable) fossil fuels, with renewable energy from wind, solar, geothermal and biomass making up only approximately 2% of the total [11]. Meanwhile, the demand for energy is constantly growing: worldwide energy production grew by 46% in the the 20 years from 1987 to 2007 [11]. The simple physical limits of our current energy resources, as well as the environmental and climate impact of burning massive amounts of fossil fuels, make a research focus on issues of sustainability imperative. Furthermore, there are numerous problems in sustainability that are fundamentally data analysis and prediction tasks, areas where techniques from data mining and machine learning can prove invaluable.

Despite the importance of sustainability research and the relevance of data mining and machine learning techniques, there has been relatively little work in these areas, at least compared to other applications areas such as computational

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SustKDD 2011 August 2011, San Diego, CA, USA  
Copyright 2011 ACM 978-1-4503-0840-3 ...\$10.00.

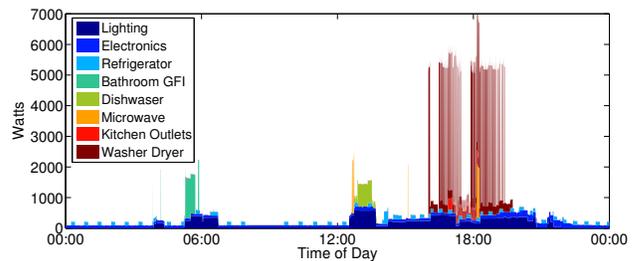


Figure 1: An example of energy consumption over the course of a day for one of the houses in REDD.

biology or machine vision. We argue that this situation is at least partly due to the scarcity of *publicly available data* for such domains. For example, although there are vast amounts of data relevant to energy domains (the energy consumption of each individual building and household in the country, the loading of each electrical transmission and distribution line) the majority of this data is unavailable to researchers. Furthermore, there is significant evidence that publicly available data sets have spurred previous applications areas in machine learning and data mining: biological applications have been aided greatly by the data sharing mandates of biological journals and government organizations [16, 12]; many early successes in natural language processing were spawned by the now-classic Wall Street Journal corpus [10]; and machine vision research has been aided greatly by common benchmark datasets such as MNIST digit recognition [9], CalTech 101 [3], and the PASCAL challenge [2]. Despite some initial progress towards this same goal for energy and sustainability domains [17], there are currently few such data sets geared to the ML and data mining communities.

In this paper, we present our work on developing a public data set of this type, termed the Reference Energy Disaggregation Data Set (REDD). The data is specifically geared towards the task of *energy disaggregation*: determining the component devices from an aggregated electricity signal. REDD consists of whole-home and circuit/device specific electricity consumption for a number of real houses over several months' time. For each monitored house, we record (1) the whole home electricity signal (current monitors on both phases of power and a voltage monitor on one phase) recorded at a high frequency (15kHz); (2) up to 24 individual circuits in the home, each labeled with its category of appliance or appliances, recorded at 0.5 Hz; (3) up to 20

plug-level monitors in the home, recorded at 1 Hz, with a focus on logging electronics devices where multiple devices are grouped to a single circuit. An example of this type of data is shown in Figure 1. As of the time of writing (June 15th, 2011), we have 10 homes monitored, with a total of 119 days of data (combined over all homes), 268 unique monitors, and more than 1 terabyte of raw data. To the best of our knowledge, REDD represents the largest publicly available data set for disaggregation with the true loads of each house identified. The entirety of the data as well as code for parsing the data and running basic algorithms is publicly available on the web: <http://redd.csail.mit.edu>.

While we present some basic results on disaggregation here, the focus of this paper is the data set itself: the design decisions that went into the data collection, as well as the hardware and software system. We begin by presenting a brief overview of existing work on disaggregation and discuss how this influenced our choices of which data to collect from each home and at what frequency. We then describe the software and hardware systems we have built for this task, and discuss their strengths and limitations. Finally, we present brief results on the data, and highlight several directions for future algorithmic work.

## 2. ENERGY DISAGGREGATION

Energy disaggregation, also referred to as a non-intrusive load monitoring (NILM),<sup>1</sup> is the task of using an aggregate energy signal, such as that coming from a whole-home power monitor, to make inferences about the different individual loads of the system. The value of this technology is that information about individual appliances is much more useful to consumers than simply total electricity usage; studies have shown that user feedback of this type can induce behavior changes that improve user efficiency by 15% [1, 13]. Disaggregation technology is also seen as an intermediate between existing electricity meters (which merely record whole-home power usage at some frequency) and a fully energy-aware home appliance network, where each device reports its consumption to a central location; an oft-stated goal of disaggregation research is to push energy awareness to a ubiquitous level, paving the way for more detailed energy monitoring in the future.

Academic work on energy disaggregation began with the work of Hart et. al [6] in the 1980s and 1990s. The initial approaches look for sharp edges (corresponding to device on/off events) in both the real and reactive power signals, and cluster devices according to these changes in consumption. Later work has explored a number of different directions: Using more complex device models with multiple states, integrating frequency analysis and other features of the AC waveforms, and making use of external features such as time of day or weather conditions. A recent review of numerous existing techniques for energy disaggregation can be found in [18]. In this paper, we highlight some of the key *distinctions* which have characterized past work in energy disaggregation and how they have informed our choices for REDD.

**Frequency of Measurements.** Past work has spanned a broad range in terms of the frequency of energy measure-

ments used for disaggregation: some work has used average power measurements over periods as long as an hour [8], while others have analyzed the harmonics of AC waveforms using MHz resolutions [14, 5].<sup>2</sup> Most approaches fall somewhere in between these two extremes, with many studies either using power readings on the order of a 1 Hz rate or AC current measurements on the order of several kilohertz. Since higher-frequency measurements can be sub-sampled to produce lower-frequency data, for our purposes of data collection it makes sense to collect data at the highest frequency possible up to the feasibility of storing the data. We chose 15kHz monitoring (for the whole-home data) as a trade-off between these factors.

**Real / Reactive Power.** Past work has also differed in whether the methods consider only the real power signal or both the real and reactive powers.<sup>3</sup> This decision is connected to the point above, since real and reactive powers can be computed using measurements of the AC waveform, but reactive power is a common enough quantity to merit its own distinction. For REDD, since we are collecting the AC waveform itself, we can easily compute both real and reactive powers.

**Use of External Features.** Some past approach use external features such as time of day, day of year, or weather information, whereas some merely use the power signal itself. All data in REDD is recorded with UTC time stamps, along with general geographical information (only up to a city level, for privacy reasons), so that it can be associated with such external features.

**Supervised / Unsupervised Training.** Most approaches to energy disaggregation have been *supervised*, in that the system is trained on individual device power signals (or is given manually identified device change-points in a whole-home energy signal). Alternatively, some recent work has advocated *unsupervised* approaches that consider the whole home signal without labeling, and automatically separate different signals [7]. To facilitate supervised approaches and to aid in evaluating all approaches, REDD includes as much “supervised” information as possible: we monitor each individual circuit in the home (especially important for large loads that cannot be easily monitored by a plug load) as well as many large plugs loads as is feasible.

**Training / Testing Generalization.** Another key distinction (which has not been greatly considered in past energy disaggregation work) is generalizing from training data to test data. The vast majority of previous disaggregation approaches (at least those with rigorous quantitative evaluation) have typically evaluated the algorithms on the same devices (but in different conditions) as they were trained on; that is, they attempt to build a model that can disaggregate a given appliance even in new conditions, but do not attempt to build models that explicitly *generalize* across multiple different devices of the same category. In our own

<sup>2</sup>The work of Patel et al. is substantially different from most other approaches to disaggregation, as they use high frequency measurements to look for transients of the *voltage* signal of the home, and not necessarily the current.

<sup>3</sup>Since the data mining community may not be familiar with this terminology, briefly, real power corresponds to the power that is actual consumed by an appliance, whereas reactive power corresponds to current that flows through a circuit, but is put back into the system typically via an inductive load in the appliance. Any text on AC power will include a rigorous description of these quantities.

<sup>1</sup>Some authors make subtle distinctions between energy disaggregation and NILM, but for our purposes we treat these terms as synonymous.

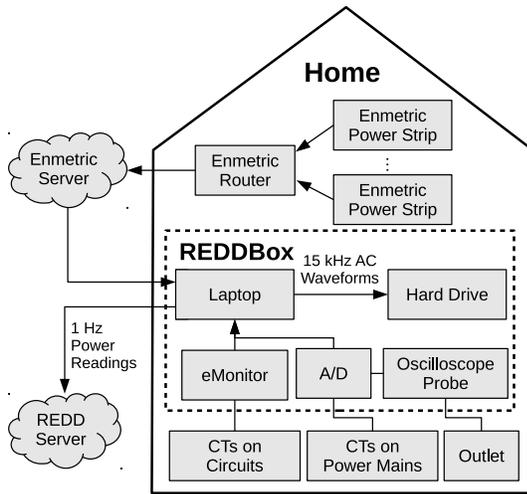


Figure 2: Schematic of the different components of the REDD hardware and software system.

past work [8] we have considered this challenge of generalizing across multiple homes, but the data used in that work was only available at an extremely low resolution (one hour), and was not permitted to be made publicly available, greatly limiting the ability of researchers to directly compare to the approach. In contrast, a goal of REDD is to consider several different homes, such that work that attempts to generalize across device types can be rigorously evaluated using this data set. As expected, and as we illustrate concretely in Section 4, generalization across homes and device categories makes disaggregation a much more challenging problem.

**Evaluation Metrics.** Finally, previous work in power disaggregation has used different metrics for evaluating performance: initial work typically focused only on on/off changes for devices, and the natural metric here is whether the algorithm can correctly classify which device is turning on or off given a change point in the whole home signal. An alternative approach is to look at the percentage of energy correctly classified (the original work by Hart et al., [6] considered both these metrics). The latter has the advantage that it is more generally applicable to disaggregation tasks, since it does not rely on extracting edges in the aggregate power signal, and applies to devices with multiple states or “smooth” power ons. This metric naturally weights high-power devices more heavily than low-power devices. While we argue that this feature is often desirable, since the absolute power consumed is the ultimate quantity we hope to influence, the metric may indicate good performance even when low-power devices are classified poorly, and in some cases these low-power devices are those over which the user has greatest control. Thus, while we will consider the “total energy properly classified” metric in our experiments, REDD can accommodate many performance metrics.

### 3. THE REDD HARDWARE AND SOFTWARE SYSTEMS

We developed the REDD hardware and software systems with the considerations of the previous section in mind. The hardware system in each house logs data from the whole-home current and voltage (at high-frequency) from each in-



Figure 3: Enmetric router and Power Port, designed and built by Enmetric Systems, Inc.



Figure 4: The eMonitor, designed and built by Powerhouse Dynamics, Inc.

dividual circuit and from selected plugs. The data is logged both locally and to central database, which stores information from all the houses and can be accessed via a web interface. A schematic of the system is shown in Figure 2.

#### 3.1 Hardware Setup

For plug-level data, we use a wireless plug monitoring system developed by Enmetric (<http://www.enmetric.com>), shown in Figure 3. The system consists of several power strips, each containing four independently monitored outlets, and a router that connects to the home’s internet connection via DHCP and processes the reading from each of the wireless devices. Energy information is then sent to a central server at a rate of 1Hz. Because the system reports at a sufficient rate and is fairly easy to install in most homes, we use the system as-is for the plug level data collection.

Circuit-level data and whole-home data require a more involved setup. For circuit level data, we again make use of an off-the-shelf solution: the eMonitor, developed by Powerhouse Dynamics (<http://www.powerhousedynamics.com>), shown in Figure 4. The eMonitor comes with current transformers (CTs) that attach to each individual circuit of the home in a house’s circuit breaker panel; the version we use monitors up to 24 circuits independently. However, the eMonitor reports power consumption to a central server at a maximum rate of once per minute. Since we are looking for more frequent power readings, we directly request measurements from the monitor using its API at the highest rate possible (for the current hardware, about one reading for all the circuits every 3 seconds).

To measure whole-home AC waveforms at high frequency, we use CTs from a TED (<http://www.theenergydetective.com>) to measure current in the power mains, a Pico TA041 oscilloscope probe (<http://www.picotechnologies.com>) to measure voltage for one of the two phases in the home, and a National Instruments NI-9239 analog to digital converter to transform both these analog signals to digital readings. This A/D converter has 24 bit resolution with noise of approximately 70  $\mu V$ , which determines the noise level of our current and volt-



Figure 5: The REDDBox, installed in a home (left) and showing internals (right).

age readings: the TED CTs are rated for 200 amp circuits and a maximum of 3 volts, so we are able to differentiate between currents of approximately  $((200)(70 \times 10^{-6})/(3)) = 4.66\text{mA}$ , corresponding to power changes of about 0.5 watts. Similarly, since we use a 1:100 voltage stepdown in the oscilloscope probe, we can detect voltage differences of about 7mV. All the data is sent to a laptop, which logs the data and sends a subset of the raw data to our central server.

Finally, since the system contains a number of electronics in close proximity to the circuit breaker box (the eMonitor, A/D converter, oscilloscope proper, computer, external hard drive, and various power supplies/cables), we have built small boxes, dubbed “REDDBoxes,” to contain all these parts in a single unit. A picture of the complete system as it would be installed near a circuit breaker box is shown in Figure 5.

### 3.2 Software System

The software system on the laptop in each REDDBox contains all the logic to query data from each of the monitors, store the readings locally, and send processed information (power from each of the monitors at a maximum of 1Hz) to a central database. Recall that we log two phases of current and one phase of voltage at 15kHz; readings from the A/D converter are 24-bit, resulting in  $\approx 11\text{GB}$  of data logged from each home per day (which we can compress 1.5-3X using bzip2). It is infeasible to send this much information over the network, so we log the data locally to an external hard disk, which we manually collect periodically and copy to the REDD server. Since developing new features from AC waveforms is a principal research direction for disaggregation methods, we include this full data for researchers to analyze if desired, though we also compute simple power information from the signal.

In addition to the software running locally at each home, we have a central database that stores power readings from all the homes, as well as a web interface that displays the real-time status of the system and allows users to see recent data from the houses. A view of the web interface is shown in Figure 6.

### 3.3 Privacy Considerations

Finally, given the nature of this public data set, we want to briefly discuss the privacy concerns involved. Sharing someone’s real-time power data in addition to their identity is potentially quite harmful: in addition to simply being able to estimate private information such as the amount of time someone spends watching television, it would be quite easy

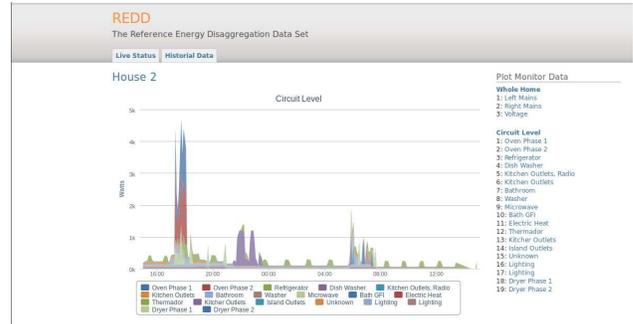


Figure 6: Live web interface displaying power consumption at the circuit level in a home.

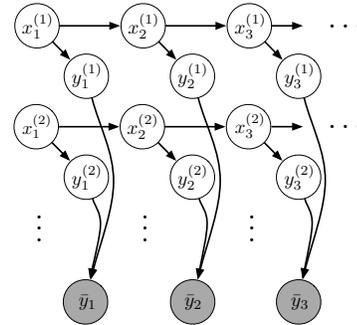


Figure 7: Graphical model representation of the Factorial Hidden Markov Model (FHMM).

to determine if someone was home or not based upon their power usage. For these reasons, we (1) store no identifying information about the houses in the database, and disclose only that they are in the greater Boston area, and (2) release only historical data, and keep the live portion of the website for private use alone. (All participants in the study are made aware of these stipulations.) Although privacy concerns are still an issue that requires constant monitoring, our hope is that these safeguards greatly decrease the risk of disclosing or identifying personal information from the data.

## 4. EXPERIMENTAL RESULTS

Here we present examples of simple algorithms applied to REDD. The goal of this section is not to present state-of-the-art performance results, but rather to demonstrate the performance of a well-studied algorithm for this task and highlight some of the challenges for future work.

We focus on the Factorial Hidden Markov Model (FHMM) [4], which has been considered recently as a method for disaggregation [7]. In the FHMM, each of the  $n$  devices (or circuits) in the home is described via a Hidden Markov Model (HMM). Each device has a discrete hidden state, denoted  $x_t^{(i)} \in \{1, \dots, N_i\}$  for the state at time  $t$  for device  $i$ , which corresponds roughly to the internal state of the device (“off”, or in one of several possible “on” states). At each time  $t$ , given the internal state, the  $i$ th device emits a Gaussian-distributed power, denoted  $y_t^{(i)}$ , with state-specific mean and variance parameters. However, we only observe the sum of all the power outputs at each time,  $\bar{y}_t = \sum_{i=1}^n y_t^{(i)}$ . The disaggregation task can then be framed as an inference problem: given an observed sequence of aggregate en-

House	Monitors	Device Categories
1	20	Electronics, Lighting, Refrigerator, Disposal, Dishwasher, Furnace, Washer Dryer, Smoke Alarms, Bathroom GFI, Kitchen Outlets, Microwave
2	19	Lighting, Refrigerator, Dishwasher, Washer Dryer, Bathroom GFI, Kitchen Outlets, Oven, Microwave, Electric Heat, Stove
3	24	Electronics, Lighting, Refrigerator, Disposal, Dishwasher, Furnace, Washer Dryer, Bathroom GFI, Kitchen Outlets, Microwave, Electric Heat, Outdoor Outlets
4	19	Lighting, Dishwasher, Furnace, Washer Dryer, Smoke Alarms, Bathroom GFI, Kitchen Outlets, Stove, Disposal, Air Conditioning
5	10	Lighting, Refrigerator, Disposal, Dishwasher, Washer Dryer, Kitchen Outlets, Microwave, Stove

**Table 1: Description of the houses and devices used in the evaluation.**

ergy  $\bar{y}_1, \dots, \bar{y}_T$ , we aim to compute the posterior probability of the individual device consumptions  $y_t^{(i)}$ ,  $i = 1, \dots, n$ ,  $t = 1, \dots, T$ . A graphical model depicting representing this FHMM is shown in Figure 7.

Although training and inference in an FHMM are nontrivial, the algorithms are described in detail in other work, and so we only discuss them briefly here and include code for the algorithm in the REDD release. To build the model from data we use the individual appliance energy sequences, as collected by the individual device monitors, and train HMMs using the standard Baum-Welch (EM) algorithm (thus, the algorithm we are describing falls under the “supervised” designation of Section 2). Exact posterior inference in the FHMM model is not tractable (we use 4 states per device, and typically around 20 devices per home for a total of  $4^{20} \approx 1 \times 10^{12}$  different combinations of hidden states), so we use a blocked Gibbs sampling scheme: we fix the hidden states of all but one of the chains, resulting in a Gaussian posterior over the emissions over the remaining chain; at this point, we can efficiently sample over hidden states for the held-out chain, and repeat the process until the distribution over all hidden states mixes. (We also anneal the sampling procedure by artificially inflating the variance of the observed aggregate outputs during the early iterations of Gibbs sampling.)

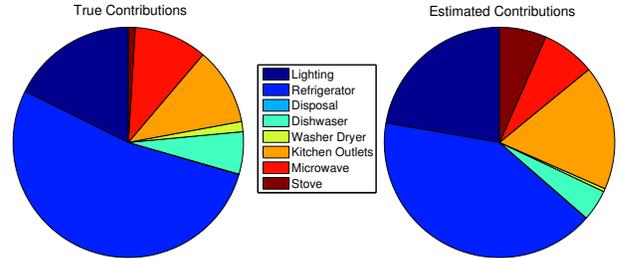
To evaluate the method, we used 2 weeks of data from 5 of the houses in REDD; since the plug-level monitors had not yet collected sufficient data at the time of writing, we use the whole-home and circuit level data. A description of the devices in each of these homes is given in Table 1. In the presented experiments we sub-sampled the data to 10 second intervals using a median filter. To evaluate the performance of the method, we used the “total energy correctly assigned” metric described in Section 2, defined formally as

$$\text{Acc} = 1 - \frac{\sum_{t=1}^T \sum_{i=1}^n \left| \hat{y}_t^{(i)} - y_t^{(i)} \right|}{2 \sum_{t=1}^T \bar{y}_t} \quad (1)$$

where  $\hat{y}_t^{(i)}$  denotes the algorithm’s prediction for the  $i$ th de-

House	FHMM		Simple Mean	
	Train	Test	Train	Test
1	71.5%	46.6%	41.4%	21.5%
2	59.6%	50.8%	39.0%	36.7%
3	59.6%	33.3%	46.7%	18.8%
4	69.0%	52.0%	52.7%	32.5%
6	62.9%	55.7%	33.7%	19.8%
Total	64.5%	47.7%	42.7%	25.9%

**Table 2: Percentage of total energy classified correctly for different houses, using FHMM disaggregation and a simple model that predicts the device’s average consumption percentage at each time.**



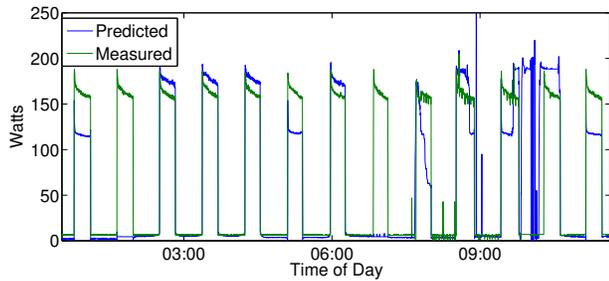
**Figure 8: Pie chart showing predicted and actual energy consumed for House 5 (when the FHMM is trained only on houses 1-4), averaged over the course of two weeks.**

vice at the  $t$ th time step, and where the 2 factor in the denominator comes from the fact that the absolute value will “double count” errors, since  $\sum_{i=1}^n y_t^{(i)} = \sum_{i=1}^n \hat{y}_t^{(i)}$ .

Table 2 shows the disaggregation performance of the FHMM model on the five houses we consider. We focus on two testing procedures: in the first case we build HMM models from devices in a given house, and then attempt to disaggregate energy in that house; in the second case, we train on four of the houses and test on the remaining held-out house. This procedure is analogous to “training” versus “testing” error, and we label the results accordingly in Table 2. For comparison we also show the performance of a simple mean prediction algorithm, which estimates the total percentage that each device type consumes and predicts that the total energy breaks down according to this percentage at all times.

As seen in Table 2, the FHMM is able to disaggregate the power data reasonably well; as expected, there is a significant drop in accuracy when moving from training prediction to test prediction, but the FHMM method still works substantially better than simple mean prediction. Although average accuracies of around 50% may seem low, we emphasize that this is for the case of predicting a previously unseen set of devices, and this metric measures the percentage of total energy correctly classified at each 10 second interval. If we aggregate the predictions over a longer time horizon, then errors tend to “cancel out,” and we often obtain much higher accuracy. Figure 8, for example, shows the total true and predicted energy for house 5 (training only on houses 1-5), summed over two weeks. At this level of aggregation, the method classifies 82% of the energy correctly, and such “aggregated” charts have significant value for user feedback.

## 5. CONCLUSION



**Figure 9: Predicted versus actual energy signal for the refrigerator in House 5.**

This paper has introduced REDD, a data set for research energy disaggregation. Energy disaggregation is an algorithmic challenge where advances can have a real impact on energy efficiency and sustainability. We have described the hardware and software setup and demonstrated a standard algorithm, the FHMM, for the disaggregation task.

Our ultimate goal in developing REDD, however, is to provide an easily-accessible data set for researchers working in data mining or machine learning. Thus, we highlight the fact that while FHMMs performed reasonably well in the experiments we presented, there is also much room for improvement. For example, Figure 9 shows an actual and predicted signal for the refrigerator in House 5; although the FHMM sometimes extracts the signal correctly, it also often fails to detect the refrigerator or estimates a noisy and unrelated signal. Many modifications, such including explicit durations via an HSMM [15], incorporating hard constraints on device signals, or looking at more complex features of the power signal can all help to improve this performance. Of particular interest to us is how such techniques could will extend to generalize across different devices in multiple homes. We are also excited by the prospect of semi-supervised techniques for disaggregation; while REDD aims to be a large resource, we can only outfit so many homes with such detailed sensing, and a great challenge that remains is to discover ways to merge this type of high-fidelity measurement with the massive amounts of (unlabeled) smart meter data that utility companies currently generate. Our hope is that the availability of a data set such as REDD can further motivate the machine learning and data mining communities to tackle this problem.

## Acknowledgments

This work was supported by ARPA-E (Advanced Research Projects Agency-Energy) under grant number DE-AR0000018. J. Zico Kolter is supported by a National Science Foundation Computing Innovation Fellowship. Matt J. Johnson is supported by an National Science Foundation Graduate Research Fellowship. We thank Carrie Armel and Mario Berges for helpful discussions. We thank Enmetric Systems and Powerhouse Dynamics for their assistance in using their hardware for this project.

## 6. REFERENCES

[1] S. Darby. The effectiveness of feedback on energy consumption. Technical report, Environmental Change Institute, University of Oxford, 2006.

[2] M. Everingham, L. V. Gool, C. Williams, J. Winn, and A. Zisserman. The PASCAL visual object classes (VOC) challenge. *International Journal for Computer Vision*, 88(2):303–338, 2010.

[3] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. *Computer Vision and Image Und.*, 106(1):59–70, 2007.

[4] Z. Ghahramani and M. I. Jordan. Factorial hidden markov models. *Machine Learning*, 29(2–3):245–273, 1997.

[5] S. Gupta, 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 Conference on Ubiquitous Computing*, 2010.

[6] G. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12), 1992.

[7] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han. Unsupervised disaggregation of low frequency power measurements. In *Proceedings of the SIAM Conference on Data Mining*, 2011.

[8] J. Z. Kolter, S. Batra, and A. Y. Ng. Energy disaggregation via discriminative sparse coding. In *Neural Information Processing Systems*, 2010.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[10] M. P. Marcus, M. A. Marcinkiewicz, and B. Santorini. Building a large annotated corpus of English: the Penn treebank. *Computational Linguistics: Special issue on using large corpora: II*, 19(2), 1993.

[11] Multiple. Annual energy review 2009. Technical report, U.S. Energy Information Administration, 2009.

[12] National Institute of Health. NIH data sharing policy and implementation guidance. [http://grants.nih.gov/grants/policy/data\\_sharing/data\\_sharing\\_guidance.htm](http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm), 2003.

[13] B. Neenan and J. Robinson. Residential electricity use feedback: A research synthesis and economic framework. Technical report, Electric Power Research Institute, 2009.

[14] S. N. Patel, T. Robertson, J. A. Kientz, M. S. Reynolds, and G. Abowd. At the flick of a switch: detecting and classifying unique electrical events on the residential power line. In *Proceedings of the Conference on Ubiquitous Computing*, 2006.

[15] L. Rabiner. A tutorial on hmm and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, February 1989.

[16] Science Magazine. Science magazine: General policies. <http://www.sciencemag.org/site/help/authors/policies.xhtml>, 2011.

[17] U. S. Department of Energy. Open energy info. <http://www.openei.org>, 2011.

[18] M. Ziefman and K. Roth. Nonintrusive appliance load monitoring: Review and outlook. *IEEE Transactions on Consumer Electronics*, 57(1):76–84, 2011.