Enhancing Electricity Audits in Residential Buildings with Nonintrusive Load Monitoring

Mario E. Berges, Ethan Goldman, H. Scott Matthews, and Lucio Soibelman

Summary

Nonintrusive load monitoring (NILM) is a technique for deducing the power consumption and operational schedule of individual loads in a building from measurements of the overall voltage and current feeding it, using information and communication technologies. In this article, we review the potential of this technology to enhance residential electricity audits. First, we review the currently commercially available whole-house and plug-level technology for residential electricity monitoring in the context of supporting audits. We then contrast this with NILM and show the advantages and disadvantages of the approach by discussing results from a prototype system installed in an apartment unit. Recommendations for improving the technology to allow detailed, continuous appliance-level auditing of residential buildings are provided, along with ideas for possible future work in the field.

Address correspondence to:
Mario Berges
113 Porter Hall
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213
USA
marioberges@cmu.edu

Introduction

There are many opportunities for reducing electricity consumption in buildings, but identifying and quantifying them is often perceived to be too time-consuming or too expensive to justify, particularly in single-family homes. The average consumer currently receives a monthly bill as an indicator of his or her consumption. However, not only has energy-metering hardware become cheaper and more easily available in recent years, but the "smart meter" installations proposed by many utilities may provide much higher-resolution electricity consumption data than monthly bills currently do. This creates an opportunity to provide accurate and building-specific energy conservation recommendation and verification without costly submetering hardware or expert assistance. Given that residential buildings account for as much as 37% of the total electricity use in the United States (Energy Information Administration [EIA] 2008), it is well worth pursuing.

Which conservation opportunities have the most impact may seem obvious, but most building owners don't have a good sense of how much energy different appliances and activities consume. In fact, people consistently overestimate the impact of less energy-consuming appliances and less effective conservation activities, and underestimate the impact of more energy-consuming appliances and more effective conservation activities (Kempton et al. 1985). Research on energy metering has shown that targeted feedback can be an effective way to remedy this problem, by providing specific and timely information (Darby 2006; Parker et al. 2006; Fischer 2008).

Energy audits are one way to obtain accurate and objective assessments of how to achieve savings. An energy audit is a process by which a building is inspected and analyzed by an experienced technician to determine how energy is used in it, with the goal of identifying opportunities for reducing the amount needed to operate the building while maintaining comfort levels (Thumann and Younger 2003). These audits, particularly when focused on electricity, can identify two different types of conservation opportunities: equipment upgrades and altering usage patterns. Assessing the value of either type requires a base-line measurement of individual end loads. For residential audits, and specifically the case of electricity audits, these measurements may consist of simply multiplying an appliance's wattage by its estimated hours of use per year. The wattage may be based on a nameplate rating, which can differ from actual power levels, or by connecting a portable power meter to the equipment, which is time consuming and not possible for all appliances. Estimating the number of hours that a device is in use can be difficult, particularly for thermostatically controlled loads like refrigerators and air conditioners. Thus, more granular feedback on appliance-level electricity consumption is needed to validate the effectiveness of a proposed opportunity.

There is a clear need, and a good value proposition, for providing building owners, professional energy auditors, and other interested parties, including homeowners, with simpler tools that would produce more accurate estimates of the energy consumption of the individual electrical loads in a building at a reasonable cost.

In this article, we present an overview of the available hardware tools for supporting residential energy (electricity) audits, analyzing the advantages and disadvantages of each type of technology. We then discuss how nonintrusive load monitoring (NILM), a technique for identifying individual loads from the total power consumption of the building, can be used to support and enhance the audit process. Our goal is not to provide highly accurate consumption information for individual appliances in the home, but rather to help auditors and building owners prioritize by providing them relevant information. Results from a prototype NILM system currently deployed in an occupied residential building in Pittsburgh, PA, are used to support the claims. We conclude with a discussion of the advantages of NILM and the necessary improvements, along with a description of possible future work.

Electricity Audit Alternatives

While electricity metering systems for commercial and industrial buildings have been available for many years–partly due to the higher returns on investment possible by better understanding the energy use of large pieces of

electrical equipment–a number of smaller-scale residential products have emerged. Residential electricity sensors are typically promoted as a means to save energy, for financial and often environmental reasons. Most are either plug-load or whole-house meters, with a few exceptions. Neither option can both accurately measure individual appliances as well as track the changes in consumption over time.

We will explore some of the general features of these meters by grouping them into three types: whole-house, plug-level, and packaged solutions. Examples of commercially available meters in these three categories are shown in table 1. Our goal is not to provide a comprehensive or extensive review of the available technologies, but rather to provide some context for our discussion about how NILM algorithms can benefit the audit process. For a recent review of electricity use feedback devices, the reader may refer to work by the Electric Power Research Institute (EPRI 2009) and by Berges and colleagues (2010).

Whole-House Meters

Whole-house meters are typically installed on or near the utility meter or in the home's breaker panel, before the power is distributed to separate circuits. A few meters are capable of simply relaying the utility meter's reading after capturing it either using optical sensors (Blue Line Innovations 2009) or decoding the radio signal by which the meter is communicating with the utility. More commonly, one or more current transducers are placed around the main electric feed lines to measure the electric current. In some meters, the voltage is measured as well, in order to

calculate power more accurately. Metrics such as the home's instantaneous power in watts or the estimated monthly energy use in watt-hours are typically displayed on a dedicated display unit, and in some cases a data port allows a connected computer to log data, transmit readings over the Internet, or provide a richer graphical interface.

Many studies have shown that whole-house electricity use feedback interfaces, even if displaying only instantaneous power, can motivate savings of 5% to 15%. (See Motegi et al. 2003; Parker et al. 2006; EPRI 2009; Granderson et al. 2009; Darby 2006 for a survey of this field.) However, the value of such feedback is limited by its lack of specificity. While the users might notice that the home is using a significantly higher amount of power at some point, they must test different appliances to see which one is responsible for the increase, switching each one on or off and observing the change in the home's power level. For all of the appliances in a house, this protocol can take 2–4 hours (Parker et al. 2006).

Because some large loads, such as hot water heaters, refrigerators, and ovens, are not able to be directly switched on and off by the user, it may be difficult to determine the power draw for those appliances and thus to give appropriate feedback if only a real-time whole-house wattage value is available. Further, because the user is manually noting the correlation between individual appliances' use and the resulting change in power level, they might be more concerned about the energy use of appliances like toasters and hairdryers, which draw high levels of power for a short time, than the energy use of refrigerators, which draw more modest amounts of power but are

Table 1 Examples of Commercially Available Residential Electricity Meters by Type

Meter type	Meter brand/model	Meter price
Whole-house	1. The Energy Detective (TED) (Energy Inc. 2009)	$200
	2. Power Cost Monitor[TM] (Blue Line Innovations 2009)	$109
	3. Powerkuff Monitor (Powerkuff LLC 2009)	$100
Plug-load meters	4. WattsUp? PRO (Electronic Educational Devices 2009)	$236
	5. Kill-a-watt (P3 International 2009)	$20
Packaged solutions	6. AlertMe (AlertMe.com 2009)	$600 + $160/year[a]

Note: Prices are in U.S. dollars.
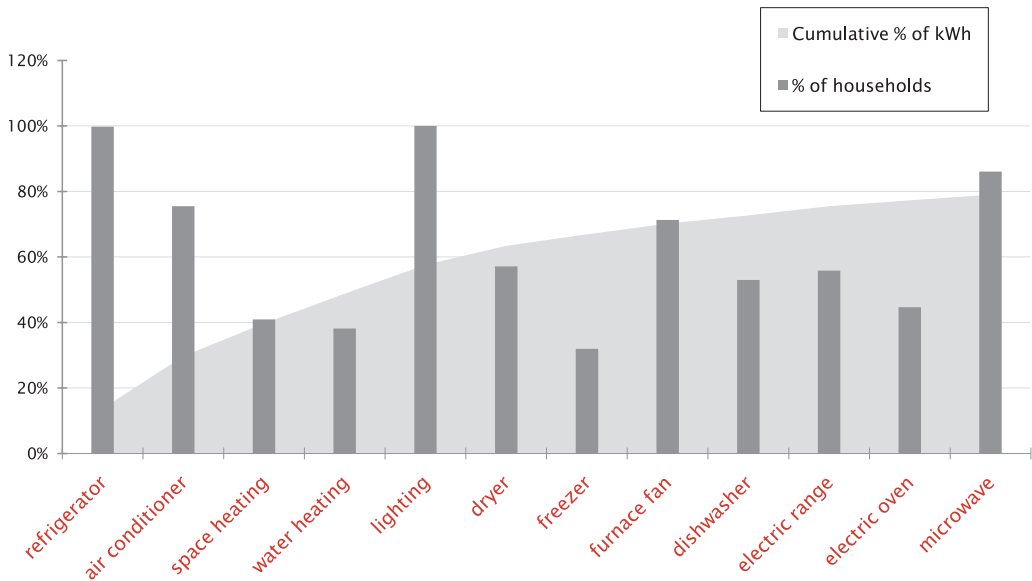[a] Assuming 12 appliances are being monitored.

Figure 1 Top 12 electricity-consuming appliances in the United States, ordered by their average annual kilowatt hour (kWh) use (EIA 2001). The bars indicate the percentage of households that have the appliance, while the light grey shading shows a cumulative graph of the percentage of energy they use in a household.

running for many hours each day and ultimately use far more energy. As a reference, figure 1 shows the top energy-consuming devices in a typical U.S. household (EIA 2001). Finally, even if the home's trend is tracked over time, it may be difficult to determine whether changes are due to behavior, new equipment, or seasonal variation, as multiple changes may overlap in time and either reinforce one another or cancel one another out.

Many metering technologies available require the help of an electrician to install. Using metering hardware that must be installed in the panel by a licensed professional adds time, and thus expense, to the process. It also has an effect on the return on investment (ROI) of the system. A subtype of these meters attaches an optical sensor to the existing utility meter and wirelessly relays kilowatt-hour (kWh) pulses to an in-home display. Though the cost is relatively low, it sends accumulated watt-hours as frequently as every 30 seconds (this value refers to meter number 2 in table 1). Other user-installable and inexpensive devices, like meter 3 in table 1, rely on an inductive sensor attached to the outside of the main

cables between the meter and the circuit breaker panel. These offer an even lower resolution, allowing for the identification of only large appliances when inspecting the overall power of the building. But at the same time they offer faster reporting rates, which could be used to support NILM algorithms.

If one intends to utilize whole-house data to understand the behavior of individual loads in a building, then another limitation of metering hardware such as these, which only report coarse real-power readings, is that appliances with a similar real-power consumption cannot be distinguished from one another. For example, a 50-watt incandescent light bulb may be confused with a 50-watt motor, or any other 50-watt load, even though their load dynamics are completely different.[1] It is therefore necessary to obtain other metrics, such as reactive power, the shape of start-up transients, or harmonic power components. Some appliances may even be distinguished by a careful analysis of their time sequence (e.g., during the morning hours, the operation of the coffee machine is usually followed by that of the toaster).

Smart Meters (AMR/AMI)

Given the recent surge of activity in *smart grid* technology and standards, it bears some discussion as to how this would benefit consumers' ability to understand and manage their electric energy consumption. It is first necessary to clarify that the term *smart meter* has been used, traditionally, to refer to the ability of these devices to establish communication with the electric utilities and exchange information about the building's energy demand, pricing, and so on, as opposed to obtaining the relevant information from building owners. To some extent it means making the grid—rather than the consumer—smart. The term does not generally imply providing feedback to the users or any other direct interaction with users, despite consumer expectations for the technology.

Much of the forthcoming investment toward the smart grid will be focused on infrastructure in the transmission and distribution systems, and many homes will be outfitted with advanced metering infrastructure (AMI) equipment, which facilitates two-way communication between the meter and the utility. This will save on meter-reading costs, much as many one-way communicating automated meter reading (AMR) systems already have. It will also facilitate the implementation of demand response (DR), an idea that relies on this two-way communication. However, while AMI systems will allow utilities to send real-time pricing signals and demand-response requests to homes, they will not necessarily provide real-time power level readings to consumers or otherwise help people to understand and reduce their consumption. On the other hand, the AMI hardware and variable pricing plans can potentially alter both the available information about energy use as well as the motivation to alter those usage patterns, so this trend is well worth following. For a review of the potential benefits of smart meters and example applications enabled by the smart grid, the reader may refer to work by Chuang and colleagues (2008). Similarly, for information on DR, see the publications by the Demand Response Research Center (Lawrence Berkeley National Laboratory 2009). The applicability of such AMI systems to NILM will be discussed later in more detail.

Plug-Load Meters

Plug-load meters are designed to measure a single appliance: the meter is plugged into any electric outlet, and then one or more appliances can be plugged into the meter. Almost all have a display of instantaneous watts, accumulated watt-hours, and other metrics, and some also transmit readings via a serial, Ethernet, or wireless connection, or even store historical data for later retrieval. These meters are a useful tool for compiling an energy audit, allowing users to take power readings for individual appliances, which they can then multiply by an estimated usage rate (e.g., hours/day) to estimate cost. It should be noted that, even when measuring an individual appliance with dedicated hardware, it may take up to two days of dedicated sampling before the typical energy consumption rate can be accurately measured given variations and gaps in the loads (Cavallo and Mapp 2000).

These snapshot measurements can be used to inform equipment upgrade decisions, such as replacing a refrigerator, or coarse behavioral changes, such as unplugging a television rather than leaving it on standby. However, the process of measuring individual appliances—sometimes over an extended period in order to average the cyclical behavior, like that of a refrigerator—is time consuming. Further, capturing trends with this method, such as the variation in energy use of the television over longer time periods (e.g., weeks or months), is even more difficult.

Packaged Solutions: Smart Homes

Packaged solutions, sometimes referred to as *smart home* solutions, typically leverage multiple metering technologies to provide an all-encompassing solution with a higher price tag. For example, while wirelessly networked plug-level electricity meters can be attached to individual appliances, this strategy is not cost-effective for most residential applications. Average household expenditures for electricity in the United States are less than $100[2] per month (EIA 2005). While few plug-load monitoring and control systems have progressed from prototypes and hype to purchasable products, systems such as number 6 in table 1 are priced at the equivalent

of $40 per wireless appliance metering point, plus $125 for the gateway to connect the in-home network to the Internet and $160/year for the required reporting service (AlertMe.com 2009). Connecting the top 12 appliances that make up the majority of U.S. residential electricity consumption, presented in figure 1, would cost over $600 initially, plus the annual fee. If such a system reduced the $1,130 annual electricity consumption of the average U.S. household (EIA 2001) by the upper bound (15%) that has been empirically demonstrated to be achieved with whole-house feedback (Parker et al. 2006; Fischer 2008), then the savings would barely pay for the annual service fee. While this example is an order of magnitude less expensive than the alternatives of just a few years ago, it still represents a longer payback period than most consumers are willing to accept. In order to achieve a simple (nondiscounted) payback period of 15 or 30 years, for instance, the corresponding savings would need to be of 35% or 30%, respectively.

In the future, manufacturers may integrate wirelessly networked energy meters into every major appliance, lowering production costs with efficiencies of scale and eliminating installation costs. However, while some companies, like General Electric, have recently begun testing appliances with such features, it is unknown if or when they will appear on the market and whether the live data will be available to consumers (instead of just the manufacturer). Further, it will be many years until all the legacy appliances disappear, and it may never be cost effective to build wirelessly networked meters into low-cost or small-load appliances. In short, there are many possible future outcomes related to smart homes and appliances, involving many possible information and communication technologies (ICTs).

Nonintrusive Load Monitoring: A Supporting Tool for Audits

The currently available technologies for residential electricity monitoring are either too expensive for most U.S. households or do not provide granular enough information to fully support energy audits. We now focus our attention on a technique that is less expensive and, if used prop-

erly, can provide a detailed continuous audit of a residential building.

NILM provides appliance-level energy metering using only a single whole-house meter and software running on an embedded device or a full-fledged computer. The typical methodology recognizes changes to the home's power level by monitoring the signals on the electrical wires, and then uses signal processing and/or machine-learning algorithms to identify which device caused the change by matching against a library of known *signatures* from different devices. While there are still some obstacles to completely replacing hardware submetering with NILM, it is effective at detecting the large appliances that are the subject of residential energy audits. The cost for this type of solution can only be estimated; based on the marginal hardware costs in our current test beds and based on an assumption of some central Web-based software services, it is conceivable that the price will be similar to that of the whole-house meters currently available on the market (approximately $200).

Background

Research on NILM, or nonintrusive appliance load monitoring system (NIALMS) as it is sometimes called, has been underway for over 20 years, beginning with George Hart in the 1980s (Hart 1989), who utilized changes in the total real (P) and reactive (Q) power of a building as signatures for each appliance state transition (Hart 1992). While Hart's research focused primarily on residential buildings, and dealt with appliances that were regularly present in American homes of the late 1980s and early 1990s, a number of researchers have been refining the techniques and extending the approach to other contexts and newer appliances (Shaw et al. 2008; Berges et al. 2009). It also bears mentioning that there are a small number of commercially available systems that implement NILM, mostly marketed for utilities as a tool for performing load research. For example, see work by Enetics, Inc. (2009) for details about one such system, and a report by EPRI (1997) analyzing the market value of the technology.

The main objective of NILM is to automatically identify appliance-specific characteristics

from the aggregate power-metrics of a building by careful inspection of the overall current and voltage at the main feed. More recently, some researchers have also investigated applying the technique to voltage distortions as measured in any outlet within the building (Cox et al. 2006), as well as to the electric and magnetic fields around the main panel (Robert Cox 2004).

To help illustrate the idea, figure 2 shows the total real power for a residential building. Even after a simple visual inspection of the signal, some appliance state changes can be easily identified. For example, the television start-up (turn on) has a characteristic power spike due to the necessary warm-up of its internal components. The objective of NILM is to automate and refine this process to accurately identify the appliances and their consumption.

In general terms, the solution consists of four basic steps: (1) data acquisition and preprocessing, (2) event detection, (3) feature extraction, and (4) classification. During the first step, voltage and current measurements are obtained and processed to produce power metrics (e.g., real and reactive power). Figure 2 shows one such power metric. Then, steady-state or transient changes in these preprocessed signals, corresponding to the operation of individual appliances, are detected. A set of features is then extracted from the samples surrounding the detected *event* in order to characterize it. Finally, these features (also called the *signature* of an appliance state transition) are presented to a classification or pattern-matching algorithm that will attempt to assign each an appropriate label, where the labels are all the possible appliance state transitions.

There are certain basic requirements on which most NILM implementations rely. For example, certain loads in a building may have nearly identical profiles in the real power domain, but may be distinguished by analyzing the way they affect the total reactive power of the building or their effect on the harmonics of the power system. Thus, NILM algorithms have been shown to benefit from the measurement of additional metrics, such as reactive power (Hart 1992) and power harmonics (Laughman et al. 2003).

Furthermore, in order to accurately detect individual appliance state changes, it is helpful to reduce the chances of more than one appliance state change occurring during a single sampling period by using higher sampling rates. Since some transients have distinctive characteristics, such as large start-up power spikes (e.g., the television in figure 2), a raw measurement of the power is preferable to a time-averaged output provided by some commercial meters. In addition, when the authors compared a number of commercial meters by measuring the same load with each meter, it was found that they differed by an unacceptable amount and in inconsistent ways (Matthews et al. 2008).

Training Process Produces an Energy Audit

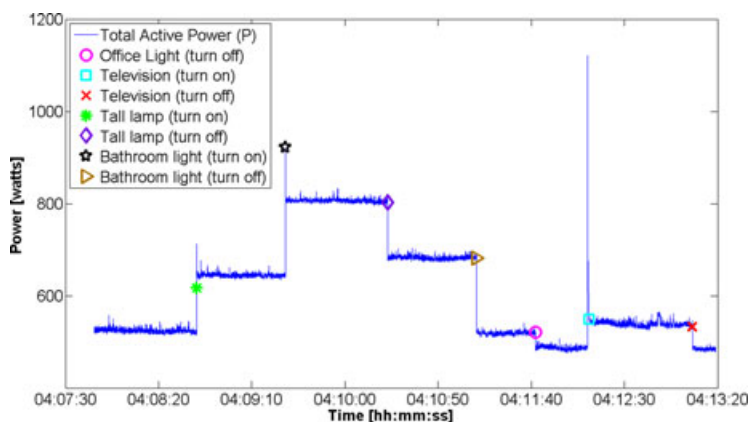One other very relevant requirement for NILM systems to function properly is a training

Figure 2 Total real power, in watts, for a residential building during a period of approximately six minutes.

process through which the characteristic signatures of the different appliances' state transitions are *learned*. The majority of research to date has favored an offline learning strategy, where the system is trained before installation or during an initial postinstallation period. We will focus on a more interactive and continuous process in which the user is constantly engaged with the system after its installation.

A first requirement for training is an enumeration of all the appliances in the home that are of interest, along with their possible state labels (e.g., on, off, high, low, etc.). Following this, the user interacts with the installed system by switching appliances on/off (changing their state), thus triggering the event detector. Finally, the system allows the user to assign one of the possible labels to the newly extracted signature.

This whole process captures the load profiles for the different appliances in the building, including the power consumption for each of the different states they can be in. A typical energy audit would use this information together with an estimate of the duty cycle for each load as a starting point for the analysis of the building's electricity consumption. As mentioned earlier, good estimates of the duty cycles are hard to obtain and typically use residential averages, which may or may not fit the audited building. However, after an NILM system has been trained, the duty cycles can be assessed continuously.

Thus, we argue that the training process can be thought of as a user-driven energy audit that does not require separately measuring the individual consumption of different appliances in the home. Furthermore, given that the NILM system will provide continuous monitoring, the effect of changes in the behavior of the residents, as reflected by the operation of the appliances, can be analyzed and presented back to the user (e.g., trends, improvements, etc.).

A Prototype NILM System

To test the feasibility of our ideas, an experimental NILM system was built using general purpose data acquisition hardware (DAQ) and slightly modified implementations of the algorithms described in the literature. The goal of our prototype was not to improve current NILM techniques, but rather to assess their effectiveness in supporting residential energy audits.

After being shown to be successful at distinguishing between several different appliances plugged into a single power strip (Berges et al. 2009), the system was installed in a residential apartment unit in Pittsburgh, PA. Using general purpose hardware allowed for the transparent calculation of metrics—including reactive power and harmonics—and permitted the rate at which these metrics are reported to be higher than what most commercially available meters provide.

Achieving similar results for a whole house to those reported when the system was installed in a controlled environment (Berges et al. 2009) was expected to be more challenging. Several loads in a home are likely to add significant noise to the signal (e.g., by continuously varying the power levels) which obscures smaller transitions, and some variable loads may also be present. In addition, with a larger number of appliances—including those that cycle independently, like the refrigerator—it is more likely that there could be overlapping events. However, initial trials demonstrated that the system was capable of correctly classifying most of the loads in the building, with higher accuracy for larger appliances. Table 2 shows the results obtained after collecting more than 200 signatures from different appliances in this single household. The F-1 measures during training are the average result of a tenfold cross-validation process where we reserved 10% of the signatures for testing and used the rest for training. The F-1 measure for validation (last column) was derived after presenting one new example of each appliance and state transition to the trained system.

Even though there is ample room to improve the prototype, we wanted to explore the idea of utilizing the system to support residential energy audits. Before we describe our experiments toward such a goal, we first summarize the hardware and software components of the prototype system.

Data Acquisition

To accurately measure the home's consumption and distinguish between 120-V and 240-V loads, the electric current on both legs (A and B) of the main electric supply were measured

Table 2 Tenfold Cross-Validation F-Score for Different Appliance State Transitions in a Residential Building

Appliance	State transition	Approx. power (watts)	Training F-1	Validation F-1
Fan	Medium–high	3	100%	100%
Fan	High–medium	−3	100%	100%
Fan	Low–medium	5	100%	100%
Fan	Medium–low	−5	100%	100%
Overhead light (kitchen)	On–off	10	67%	100%
Stereo (attic)	On–off	−10	100%	100%
Stereo (attic)	Off–on	10	100%	0%
Desk light (attic)	Off–on	15	50%	67%
Desk light (attic)	On–off	15	0%	0%
Ceiling light (attic)	Off–on	16	0%	0%
Ceiling light (attic)	On–off	−16	100%	0%
Overhead light (hallway)	Off–on	20	0%	0%
Overhead light (hallway)	On–off	−20	0%	0%
Electric kettle	Off–on	500	0%	0%
Electric kettle	On–off	−500	100%	100%
Stove (small burner)	Off–medium	570	100%	100%
Stove (small burner)	Medium–off	−570	100%	100%
Stove (large burner)	Medium–off	−870	67%	67%
Stove (large burner)	Off–medium	870	0%	0%
Toaster	Off–on	1,500	100%	67%
Toaster	On–off	−1,500	100%	100%
Heating system	Fan–off	−2,000	100%	100%
Heating system	Off–fan	2,000	100%	100%
Heating system	Heating–off	−14,000	100%	100%
Heating system	Off–heating	14,000	100%	100%

with split-core current transformers as shown in figure 3. To compute complex power, the voltage was also measured after attenuating it with a voltage transformer to fit the input range of the DAQ. Both current and voltage were simultaneously and continuously sampled at 10 kHz with a National Instruments PCI-6143 16-Bit DAQ. A custom LabVIEW program (National Instruments, Austin, TX) was used to drive the sampling operations and to compute complex power and other metrics for every three full periods of the signal (assuming 60-Hz signals, with our sampling rate, this corresponds to 500 samples).[3]

Event Detection

The NILM algorithm created for this research is focused on the classification of "events," or points in the time series of power measurements that correspond to abrupt changes. These events are assumed to be the result of an appliance changing its state. We also assume that the aggregate power metrics for the building quickly settle to steady-state levels after each state transition. This is a reasonable assumption for many, but not all, of the loads in modern buildings. However, it allows us to avoid having to implement a multiscale edge detector such as the one presented in work by Leeb and Kirtley (1996).

To detect events, we implemented a variation of the probabilistic event detector described in work by Luo and colleagues (2002). The algorithm was implemented using both LabVIEW and MATLAB scripts (MathWorks, Natick, MA). The main differences in our version are that (1) to reduce the number of parameters that need to be set, instead of assuming fixed values for the standard deviation, we continuously compute this metric from the samples; and (2) we
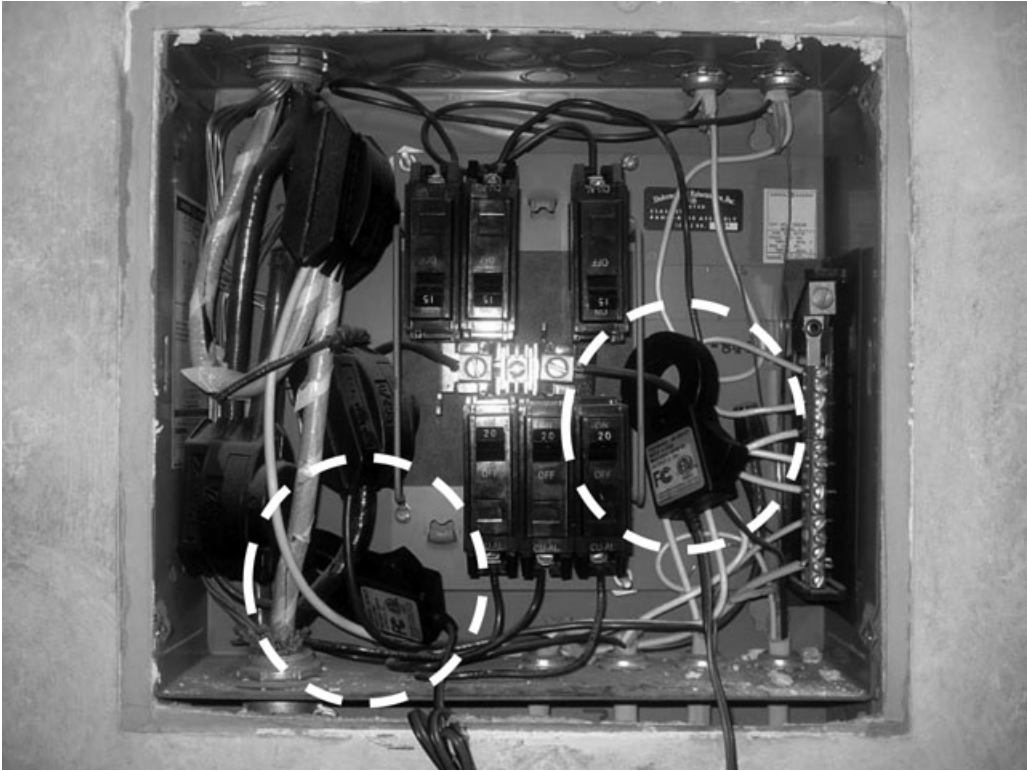
Figure 3 Current transformers (CTs), marked by dashed circles, clamped around the two main feeds of the apartment building.

implement a voting scheme on top of the output of the maximization of the detection statistic, allowing each sample to receive, at a maximum, as many votes as the number of samples in the detection window.

Appliance Signatures

Once an event is identified, a fixed-size window of samples surrounding it (from here on referred to as the "transient") will be compared with previously labeled examples in order to classify it. Several different features can be used to characterize the signature of these transients. Similarly, there are several different approaches for comparing them. One simple signature to compare would be the difference in average real power before and after the event. This typically yields unique signatures for many of the larger appliances in a home (e.g., the 14 kW of the heating unit in table 2), but, as we previously discussed, it is likely that two

or more appliances will have power levels that are indistinguishably close, and including other metrics may be necessary. Other possible features are slope and offset, first and higher order derivatives, the whole transient itself, and so forth.

Our prototype system implements the above reasoning in a series of MATLAB scripts that extract features from the real and reactive power transients for legs A and B. To capture the shape of these transients, we decided to apply linear regression on each of the four power values (P_a, P_b, Q_a, Q_b) and use the regression coefficients as the signature. Different basis functions were tested for the regression: Fourier, polynomial, radial, and so forth. The best results were obtained with Fourier basis functions (Berges et al. 2009).

Classification

The crux of the NILM process is a machine-learning algorithm that takes the signature

of a newly detected event and automatically classifies it based on a corpus of known appliance state transitions. Classification accuracy was compared across a number of different off-the-shelf machine-learning algorithms (Berges et al. 2009); however, a simple 1-nearest neighbor (1-NN) algorithm using a Euclidian distance measure between the signature vectors produced acceptable results, as presented in table 2.

Training

The classification algorithm in the system relies on an instance-based learning approach, where a new unlabeled event is compared with existing, labeled events using a distance metric. The algorithm determines which "class" the new event belongs to based on the set of examples that are closest to this new one (in this case, a class is the transition of an appliance from one state to another).

Typically, to obtain the set of examples that will compose the training set, a human observer assigns labels to specific events in a time series of power values recorded in a building. Other disaggregation systems, like the one presented in work by Farinaccio and Zmeureanu (1999), use temporary hardware submetering to obtain this ground truth data. While this last approach provides an authoritative set of signatures with which to train the classification algorithm, it is impractical for most real-world applications because of the additional hardware/installation costs required. We instead chose to train the system by manually labeling events captured by the whole-house meter. This approach can scale to a greater number of appliances in the house and does not require that an appliance be on a dedicated circuit or otherwise make special allowances for submetering equipment.

Experiment

As a proof of concept, and in order to obtain preliminary data that would help us evaluate the feasibility of using NILM to support electricity audits in a residential building, we decided to focus our attention on one appliance from the list presented earlier in figure 1: the refrigerator. The NILM prototype system was installed in an apartment building, and plug-level power meters were used to accurately track the individual consumption of this appliance. The experiment consisted of monitoring this load for a week, using the two methodologies (NILM and plug-level meters), and then comparing the estimated energy consumption as computed by each. The measurements taken by the plug-level meters would be considered as the "ground truth."

The NILM system was first trained on the refrigerator, by providing it with two start-up and two turn-off signatures, and then on 17 other two-state appliances present in the home. Adding these additional appliances increases the chances of erroneous classifications, but it also brings the experiment closer to what would be encountered in a real-life deployment. The plug-level meters that were used are part of a custom wireless sensor network platform (Rowe et al. 2006) developed at Carnegie Mellon University.

The end goal was to predict the energy consumption of the refrigerator with the NILM system. Energy values are obtained from integrating power over time. Given that our current prototype simply provides information on which appliance state transition took place and when, some extra steps (discussed in the following sections) needed to be taken to transform this into energy values. In contrast, for the plug-level power meters, the operation was simpler since the data we obtained from them were already power values.

Results

We evaluated the data obtained during a period of 5.5 days,[4] calculating energy measurements on both the predictions of the NILM and the measurements taken by the plug-level meter. Figure 4 illustrates the result of doing this by showing the actual power consumption as measured by the plug-level meter, and the prediction of the NILM system. It can be seen that when the system detects an off-on transition for the refrigerator, it assumes that the appliance will steadily draw 230 watts of power until an on-off transition is found.

The resulting difference in energy estimates for the 5.5 days of the experiment was 14.8%, with the NILM system underestimating the actual consumption by 2.29 kWh.[5] The plug-level

Figure 4 Power consumption, in watts, of one cycle of the refrigerator as estimated by nonintrusive load monitoring (NILM; dashed line) and measured by the plug-level meter (filled gray area). Note that due to the time resolution of these measurements, the turn-on and turn-off events do not appear as step changes.
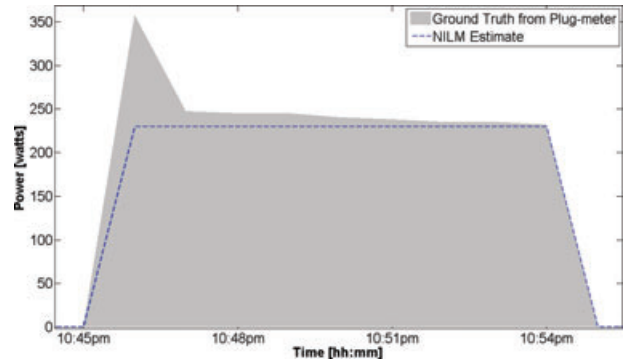
meter measured 15.48 kWh, whereas the NILM algorithms predicted 13.19 kWh. Figure 5 shows a longer time frame for the power estimates.

Discussion

The results presented above, although limited to a single appliance, show promise for the feasibility of utilizing NILM for supporting the goal of residential electricity audits. Even if not for the energy calculations, estimates of when the refrigerator cycles take place (when the events occur) can be helpful. Perhaps the most significant source of error for this experiment was the refrigerator's defrost cycle. This type of event, shown at 3:41 A.M. in figure 5, was not identified during the training process. Thus, the NILM algorithms detected an event when this occurred, but could not classify it as belonging to the refrigerator. The result is that the system believes that the refrigerator was "on" longer during that cycle. The algorithm also tries to maintain consistency by enforcing an appliance state model

that prevents two consecutive off-on or on-off events. This is why the estimates are misaligned for another two cycles.

The importance of training the system on the universe of all possible appliance state transitions in a residence becomes evident after seeing the effect that an untrained event can have on the results. While the system began correctly classifying events after only one training example, in most cases more examples were required in cases where two devices had similar signatures (e.g., two incandescent light bulbs of approximately the same wattage). Also, the training process required two people: one to turn appliances on and off, and a second to label the events on the computer with the interface running. In an early trial, we were able to record approximately 200 labeled events per hour and to accumulate several examples of every appliance in a small house in about two hours. We have developed a mobile interface to allow one-person training and to facilitate ongoing training of appliances that automatically cycle on and off, such as refrigerators. More

Figure 5 Power consumption, in watts, of the refrigerator as estimated by nonintrusive load monitoring (NILM; dashed line) and measured by the plug-level meter (filled gray area).
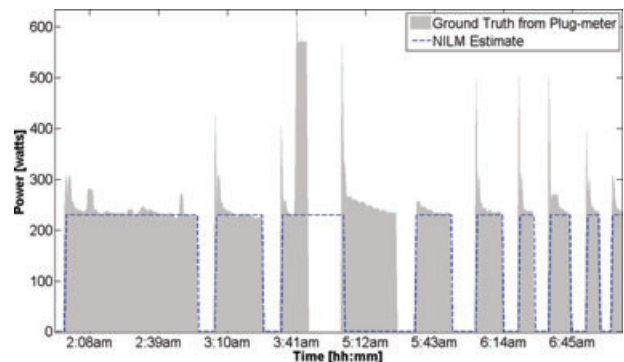
details on this interface can be found in work by Berges and colleagues (2009).

To minimize the effort required to install an NILM system, individual users should not have to train the system on every appliance in their home; adding four person-hours of training time to the hardware costs and installation labor would likely render the system impractical to all but the most dedicated users. Ideally, the system would be able to access a library of known appliances that it would be able to recognize without additional training, and only require user intervention for a small portion of the home's appliances. However, this scenario presents a couple of challenges for which there are not yet ready solutions. For example, such a library should be able to (1) handle generalized representations of appliance types as opposed to specific appliance instances; and (2) deal with the differences in power systems (e.g., U.S. vs. E.U.), sampling rates, selected features, and so forth that have a direct effect on the re-sulting signatures.

The automated classification approach to ap-pliance disaggregation that has been used here is based on the assumption that the set of appli-ances in a given home can be grouped into dis-tinct clusters in n-dimensional feature space. Us-ing between two and eight features to distinguish among the 17 appliances in the subject home (which generated 44 different transition events because of the number of the possible states for each appliance) appears to be a tractable problem. However, matching an unknown event against a library containing thousands or tens of thousands of appliance state transitions could prove much harder.

Future Work

The first branch of research that follows from the work presented here is to extend and repeat the experiments. In other words, it is necessary to include as many appliances as possible from the list of 12 presented in figure 1 and perform similar comparisons to verify that NILM systems can effectively provide good estimates of the en-ergy consumption and duty cycles. Additionally, it would be useful to evaluate the practicality of this approach by allowing experienced energy

auditors to interact with a prototype system and share their opinions.

Admittedly, better signal processing and ma-chine learning algorithms can be investigated, and this is part of our ongoing research; however, we believe that the most important hurdles to be overcome have to do with allowing these systems to be easily deployed and used. This primarily amounts to improving the way the algorithms are trained (e.g., by reducing the amount of time needed to train them, implementing an online, distributed signature library, etc.). It also seeks to improve the interfaces that the end users need to deal with in order to obtain the information they need.

Acknowledgements

Notes

1. One watt (W, SI) $\approx$ 3.412 British Thermal Units (BTU)/hour $\approx 1.341 \times 10^{-3}$ horsepower (HP).
2. All the prices listed in this document are in US dollars.
3. One hertz (Hz) = one cycle per second.
4. A power outage reduced our experimentation time from the proposed one week, to 5.5 days.
5. One kilowatt-hour (kWh) $\approx 3.6 \times 10^6$ joules (J, SI) $\approx 3.412 \times 10^3$ British Thermal Units (BTU).

References

AlertMe.com Ltd. 2009. *AlertMe—Smart energy and monitoring solutions.* www.alertme.com. Accessed December 2009.

Berges, M., L. Soibelman, H. S. Matthews, and E. Gold-man. 2010. Evaluating the electric consumption of residential buildings: Current practices and fu-ture prospects. Reston, VA: ASCE (American Society of Civil Engineers).

Berges, M., E. Goldman, H. S. Matthews, and L. Soibelman. 2009. Learning systems for electric consumption of buildings. Reston, VA: ASCE (American Society of Civil Engineers).

Blue Line Innovations. 2009. The Powercost Monitor. www.bluelineinnovations.com/powercostmonitor.php. Accessed December 2009.

Cavallo, J. and J. Mapp. 2000. Monitoring refrigerator energy usage. *Home Energy Magazine* 17: 32–36.

Chuang, A., J. Hughes, M. McGranghan, and X. Mamo. 2008. *Integrating new and emerging technologies into the California smart grid infrastructure.* Sacramento, CA: Electric Power Research Institute.

Cox, R. 2004. *A field-based approach to non-intrusive load monitoring.* Boston: Massachusetts Institute of Technology.

Cox, R., S. Leeb, S. Shaw, and L. Norford. 2006. Transient event detection for nonintrusive load monitoring and demand side management using voltage distortion. Paper presented at *Applied Power Electronics Conference and Exposition, 2006*, 19–23 March.

Darby, S. 2006. *The effectiveness of feedback on energy consumption.* Oxford, UK: Environmental Change Institute, University of Oxford.

Electric Power Research Institute. 1997. *Low-cost NIALMS technology: Market issues & product assessment.* Palo Alto, CA: Electric Power Research Institute.

Electric Power Research Institute. 2009. *Residential electricity use feedback: A research synthesis and economic framework.* Palo Alto, CA: Electric Power Research Institute.

Electronic Educational Devices. 2009. Watts up? PRO. www.wattsupmeters.com. Accessed December 2009.

Energy, Inc. 2009. The Energy Detective. www.thenergydetective.com. Accessed December 2009.

Energy Information Administration. 2001. *End-use consumption of electricity by end use and appliance.* Washington, DC: Energy Information Administration.

Energy Information Administration. 2005. *Residential energy consumption survey.* Washington, DC: Energy Information Administration.

Energy Information Administration. 2008. *Annual energy review.* Washington, DC: Energy Information Administration.

Enetics, Inc. 2009. Enetics advanced metering and analyzers. www.enetics.com. Accessed December 2009.

Farinaccio, L. and R. Zmeureanu. 1999. Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses. *Energy and Buildings* 30: 245–259.

Fischer, C. 2008. Feedback on household electricity consumption: A tool for saving energy? *Energy Efficiency* 1: 79–104.

Granderson, J. G. Ghatikar, M. A. Piette, and P. Price. 2009. *Preliminary findings from an analysis of building energy information system technologies.* Berkeley, CA: Lawrence Berkeley National Laboratory.

Hart, G. 1989. Residential energy monitoring and computerized surveillance via utility power flows. *IEEE Technology and Society Magazine* 8: 12–16.

Hart, G. 1992. Nonintrusive appliance load monitoring. *Proceedings of the IEEE* 80: 1870–1891.

Kempton, W., C. K. Harris, J. G. Keith, and J. S. Weihl. 1985. Do consumers know "what works" in energy conservation? *Marriage and Family Review* 9: 115–133.

Laughman, C., K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong. 2003. Power signature analysis. *IEEE Power and Energy Magazine* 1: 56–63.

Lawrence Berkeley National Laboratory. 2009. All publications. Demand Response Research Center. drrc.lbl.gov/drrc-pubsall.html. Accessed December 2009.

Leeb, S. and J. L. Kirtley Jr. 1996. Transient event detector for use in nonintrusive load monitoring systems. U.S. Patent 5483153, January 9, 1996.

Luo, D., L. Norford, S. Shaw, and S. Leeb. 2002. Monitoring HVAC equipment electrical loads from a centralized location—Methods and field test results. *ASHRAE Transactions* 108: 841–857.

Matthews, H. S., L. Soibelman, M. Berges, and E. Goldman. 2008. Automatically disaggregating the total electrical load in residential buildings: A profile of the required solution. *Intelligent Computing in Engineering (ICE08) Proceedings* 381–389.

Motegi, N., M. A. Piette, S. Kinney, and J. Dewey. 2003. *Case studies of energy information systems and related technology: Operational practices, costs, and benefits.* Berkeley, CA: Lawrence Berkeley National Laboratory.

P3 International, Kill A Watt. 2009. www.p3international.com. Accessed December 2009.

Parker, D., D. Hoak, A. Meier, and R. Brown. 2006. How much energy are we using? Potential of residential energy demand feedback devices. Paper presented at *Proceedings of the 2006 Summer Study on Energy Efficiency in Buildings*, 13–18 August, Pacific Grove, CA.

Powerkuff LLC. 2009. Powerkuff Monitor Systems and Wire Wrap. www.powerkuff.com. Accessed December 2009.

Rowe, A., R. Mangharam, and R. Rajkumar. 2006. FireFly: A time synchronized real-time sensor

networking platform. In *Wireless ad hoc networking: Personal-area, local-area, and the sensory-area networks*. Boca Raton, FL: CRC Press.

Shaw, S., S. Leeb, L. Norford, and R. Cox. 2008. Nonintrusive load monitoring and diagnostics in power systems. *IEEE Transactions on Instrumentation and Measurement* 57: 1445–1454.

Thumann, A. and W. J. Younger. 2003. *Handbook of energy audits*. Sixth edition. Lilburn, GA: Fairmont Press.

About the Authors

Mario E. Bergés is an assistant professor in the Department of Civil and Environmental Engineering at Carnegie Mellon University (but was a PhD student at the same institution when this was written). **Ethan Goldman** was a PhD student at Carnegie Mellon University in Pittsburgh, PA, at the time the article was written. He is currently the Measurement & Verification Specialist at Efficiency Vermont, Burlington, VT. **H. Scott Matthews** is a professor in the Department of Civil and Environmental Engineering and research director for the Green Design Institute, at Carnegie Mellon University. **Lucio Soibelman** is a professor at Carnegie Mellon University in the Department of Civil and Environmental Engineering in Pittsburgh, PA, USA.