BLUED: A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research

Kyle Anderson
Electrical and Computer Engineering, Carnegie Mellon University
kda@ece.cmu.edu

Adrian Filip Ocneanu
Civil and Environmental Engineering, Carnegie Mellon University
adrianocneanu@cmu.edu

Diego Benítez
Research and Technology Center, Robert Bosch, LLC, Pittsburgh, PA
Diego.Benitez@us.bosch.com

Derrick Carlson
Civil and Environmental Engineering, Carnegie Mellon University
carlsond@cmu.edu

Anthony Rowe
Electrical and Computer Engineering, Carnegie Mellon University
agr@ece.cmu.edu

Mario Bergés
Civil and Environmental Engineering, Carnegie Mellon University
marioberges@cmu.edu

ABSTRACT

The problem of estimating the electricity consumption of individual appliances in a building from a limited number of voltage and/or current measurements in the distribution system has received renewed interest from the research community in recent years. In this paper, we present a Building-Level fUlly-labeled dataset for Electricity Disaggregation (BLUED). The dataset consists of voltage and current measurements for a single-family residence in the United States, sampled at 12 kHz for a whole week. Every state transition of each appliance in the home during this time was labeled and time-stamped, providing the necessary ground truth for the evaluation of event-based algorithms. With this dataset, we aim to motivate algorithm development and testing. The paper describes the hardware and software configuration, as well as the dataset's benefits and limitations. We also present some of our detection results as a preliminary benchmark.

Keywords

Non-intrusive load monitoring, datasets, energy disaggregation, algorithm performance evaluation.

1. INTRODUCTION

In this paper, we present a one-week dataset of residential electricity usage with labels for individual appliance activity. The labels mark when appliance state transitions occur. These transitions represent changes in appliance activity accompanied by a change in power consumption level. We frequently refer to these transitions, throughout the text, as events.

Beyond detailed energy consumption information, this dataset could be mined for many other applications including, but not limited to: appliance operation patterns, security applications, occupancy detection, energy management, assisted living applications, appliance fault diagnostics, and anomaly detection. Parties interested in these problems include electricity consumers, utility and distribution companies, appliance manufacturers, and policy makers.

ACM SustKDD'12, August 12, 2012, Beijing, China Copyright 2012 ACM
978-1-4503-1558-6/12/08 ....$10.00

In general, two main energy disaggregation approaches exist for non-intrusive load monitoring: event-based and non event-based. A good review of the existing approaches can be found in [12]. The non event-based approach attempts to directly separate the sources that compose the overall signal (e.g., total power consumption) by using techniques such as latent variable models [7], blind source separation [5] or time-series motif mining [10]. Event based approaches, on the other hand, keep track of each appliance state transition by means of event detection and classification [9].

For the non-event based approach, the Reference Energy Disaggregation Data Set REDD [6] has been released. This dataset provides labels for each circuit in the electrical panel of the home and is well-suited for validating any source-separation algorithm that assumes that the sources directly correspond to these circuits. The BLUED dataset that we present in this paper is different from previous work in that it provides labels (e.g., timestamps and appliance identifiers) for each appliance state transition occurring in the dataset. This dataset is relevant to the KDD community since, to our knowledge, it will be the first publically available dataset of its kind.

The main reason for this dataset to exist is to motivate algorithm development and testing for the NILM community. Similar to what occurred in the face-detection or voice-recognition communities, performance comparisons are not meaningful unless made on common datasets using common metrics.

The BLUED dataset is publicly available for download at http://nilm.cmubi.org. Raw current and voltage files along with a list of event timestamps are provided.

The rest of this paper is organized as follows. Section 2 describes the particulars of the data collection framework used and technical challenges encountered. Section 3 describes the dataset in terms of appliances represented, power consumption, and frequency of events. Section 0 presents some preliminary experimental results on event detection, section 5 provides final thoughts and conclusions, and section 6 indicates future research plans.

2. DATA COLLECTION SET-UP

One week of voltage and current measurements was collected for a single family house in Pittsburgh, Pennsylvania. Data collection took place during October 2011. There were approximately 50 electrical appliances in the home, and our goal was to individually track the electrical operation of each device, determining when each appliance changed its operating status (e.g., turned on or off).

The hardware used for creating the dataset can be grouped into two basic categories: a system used for collecting the aggregate voltage and current measurements at the main distribution panel and another system used for obtaining the ground truth (i.e., the time-stamps for each appliance state transition). Figure 1 shows the overall system architecture. The data streams from these two systems were then post-processed to correlate the appliance activity with the aggregate voltage and current signals.

In this section, we describe the setup of these two systems and the post-processing steps that were performed. We also include a brief discussion of some of the technical challenges encountered during the data collection process.

2.1 Measurement

In the United States, residential buildings commonly have a 3-wire, single-phase 240V/120V power distribution system in which a single (240V) primary single phase is center-tapped at the transformer to create two 60Hz, 120V sources and a neutral. The sinusoidal voltage signals on these two "live" wires have a phase difference of 180 degrees. In this paper, we refer to them as phase A and B.

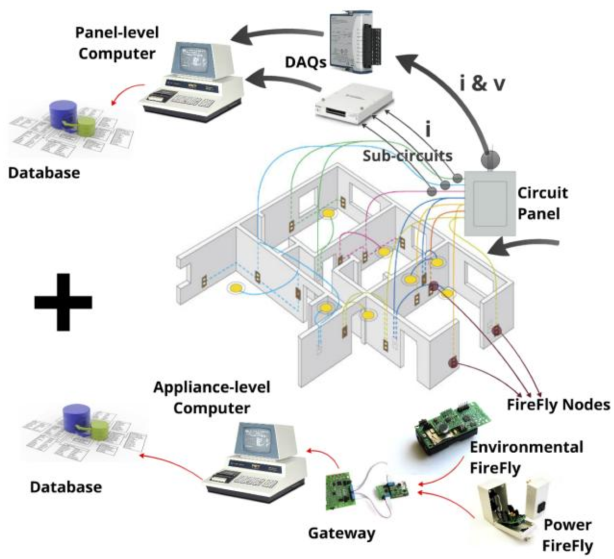

Figure 1 System architecture for data and ground truth collection.

Voltage and current measurements were collected using a 16-Bit data acquisition device from National Instruments (NI USB-9215A). We assumed that the voltage signals would be phase-shifted copies of each other (by a half-cycle) and only sampled one voltage and two current signals at 12 kHz, simultaneously. The power consumption for the entire house was then computed based on these current and voltage measurements.

To measure the electrical current, we used two QX 201-CT split-core current transformers from The Energy Detective[1]. The current clamps were placed around the two incoming power mains, as shown in the left of Figure 2. For voltage measurements, a voltage transformer from Pico Technology[2] (PICO PROBE TA041) was used to step down the 120V AC voltage (to +/- 7 V AC). The sampled signals from these sensors were stored locally on a computer. In Figure 1 this can be seen starting from the top of the circuit panel where the measured current and voltage signals (i & v) are sampled by the DAQ and saved by the panel-level computer.

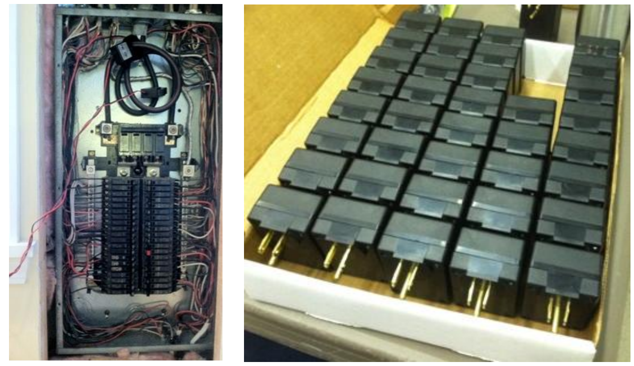

Figure 2 Left: Current clamps on the electric mains. Right: 37 Plug-level FireFly sensors ready for deployment.

2.2 Ground Truth Collection

The ground truth collection is split among three sensing modalities (1) plug-level meters, (2) environmental sensors and (3) circuit panel meters.

Plug-level and environmental data were collected using the FireFly wireless sensor networking platform [4]. Each sensor node used a custom TDMA-based collection tree networking protocol designed for 802.15.4 radios that reported sensor values every 640ms. Each node was statically assigned a unique communication slot within the TDMA frame to facilitate high-speed collision-free communication in order to minimize timing jitter and packet loss. A central gateway within the home timestamped each incoming message using the Network Time Protocol (NTP) and then locally stored the data, this can be seen in the lower half of Figure 1.

Each plug-level power sensor (28 in total), shown in Figure 2 on the right, measured voltage and current at a rate of 1kHz and locally computed active and apparent power along with RMS current, RMS voltage, and frequency, all averaged over one second. Environmental sensors were used to infer the activity of appliances like overhead lights and ventilation fans, which are hard to meter using plug-level sensors. Each FireFly environmental sensor (12 total) measured light level, sound intensity, vibration, humidity, barometric pressure and PIR motion. These sensors were carefully placed to specifically target certain appliances.

Appliances that were not easily metered with plug meters or environmental sensors (like two-phase appliances such as electric drying machines or appliances hard-wired to the electric panel such as garbage disposals) were monitored by measuring the current on individual sub-circuits at the distribution panel. RMS current for each sub-circuit was recorded at 20 Hz using CTs from CR Magnetics (CR 3110-3000 C1) and a 16-channel, 1.25MS/s, 16-Bit, USB-Based DAQ (National Instruments NI USB-6251). This can be seen coming out of the side of the circuit panel in Figure 1.

2.3 Post Processing

A post-processing stage was necessary to fully label the captured data. The first step was to compute the power consumption for the entire house from the current and voltage measurements (see

[1] http://www.theenergydetective.com/

[2] http://accessories.picotech.com/active-oscilloscope-probes.html

chapter 2 in [8] for technical details of this computation). We computed active power at a rate of 60Hz, and included these computed values in the published dataset.

In order to supplement the power measurements with information about when each one of the appliances in the home changed its operating state (i.e., when events occurred), there were approximately 50 separate channels of data from the ground truth collection that needed to be merged. For each ground truth channel (e.g., light intensity values from an environmental sensor placed near a ceiling light) a list of potential events was created by visually inspecting and hand-labeling transitions in activity. We defined an event to be any change in power consumption greater than 30 watts and lasting at least 5 seconds. The timestamps of these potential events were then overlaid on the power signal for the whole house. Due to small time synchronization errors between the panel-level and appliance-level computers, it was necessary to adjust the timestamps of the labeled events to match the transitions in the aggregate power signal. This was also done by visual inspection. Once all of this was completed, there was a total of 2,355 events labeled in the dataset.

A final visual inspection of the entire power signal was made to determine if there was any unlabeled activity not captured by the ground truth sensors. This check revealed an additional 127 events from unknown sources for a total of 2,482 events (904 on phase A and 1,578 on phase B) during the week of collection. The sensing infrastructure captured approximately 95% of the total number of events. These events with unknown sources are clearly labeled as such in the BLUED dataset.

Having hypothesized that these unknown events were due to appliances not sensed by the ground truth sensors, we attempted to cluster these events based on their real and reactive power consumption. The result of this clustering revealed that the 127 unknown events may be attributed to 11 distinct appliances.

2.4 Data Collection Challenges

During the entire data collection process, various unforeseen challenges were encountered. They ranged from having to perform circuit tracing of the whole house to Internet connectivity problems. We explain a few of the more prominent challenges in this section.

Following data collection, we learned that the current sensors used for measuring the electrical current in the mains had a cutoff frequency of approximately 300 Hz, which meant that the sampled signals, although sampled at 12 kHz, would only provide useful information for up to the 5^{th} harmonic of the current, approximately. This limits the types of algorithms that can be applied on this dataset. We also note that the REDD dataset, given that it uses the same current transformers, may suffer from the same limitations. To remedy this problem, higher fidelity current sensors will be deployed in future data collection efforts.

As noted above, the deployment captured 95% of all the events in the home. The 5% that were missed were likely the result of one or more of multiple factors: incorrect circuit tracing, appliances that might have been added to the house after the sensing infrastructure was deployed, or simply appliances that were moved to different, unmetered outlets, by the occupants of the house.

Also, there were no registered events for approximately 25% of the appliances in the house. For some appliances this was because they did not meet our criteria for being considered an event (30 W power consumption and 5 second duration, see section 2.3), while other appliances were not used during the week of data collection. This reflects real usage and suggests that one week may not be enough time to obtain a representative sample of all appliances.

Table 1 List of appliances in the dataset monitored by the ground truth sensing infrastructure.

Index	Name	Average Power Consumption (W)	Number of events	Phase
1	Desktop Lamp	30	26	B
2	Tall Desk Lamp	30	25	B
3	Garage Door	530	24	B
4	Washing Machine	130-700	95	B
5	Kitchen music	0	-	-
6	Kitchen Aid Chopper	1500	16	A
7	Tea Kettle	-	0	-
8	Toaster Oven	-	0	-
9	Fridge	120	616	A
10	A/V Living room	45	8	B
11	Sub-woofer Living room	0	-	-
12	Computer A	60	45	B
13	Laptop B	40	14	B
14	Dehumidifier	-	0	-
15	Vacuum Cleaner	-	0	-
16	DVR, A/V Receive Blueray Playe Basement	55	34	B
17	Sub-woofer Basement	0	-	-
18	Apple TV Basement	0	-	-
19	Air Compressor	1130	20	A
20	LCD Monitor A	35	77	B
21	TV Basement	190	54	B
22	Harddrive B	-	0	-
23	Printer	930	150	B
24	Hair Dryer	1600	8	A
25	Iron	1400	40	B
26	Empty living room socket	60	2	B
27	Empty living room socket	-	0	-
28	Monitor B	40	150	B
29	Backyard lights	60	16	A
30	Washroom light	110	6	A
31	Office Lights	30	54	B
32	Closet lights	20	22	B
33	Upstairs hallway light	25	17	B
34	Hallways Stairs lights	110	58	B
35	Kitchen Hallway light	15	6	B
36	Kitchen overhead light	65	56	B
37	Bathroom upstairs lights	65	98	A
38	Dining room overhead light	65	32	B
39	Bedroom Lights	190	19	A
40	Basement Light	35	39	B
41	Microwave	1550	70	B
42	Air Conditioner	-	0	A+B
43	Dryer	-	0	A+B

3. DATASET SUMMARY

Table 1 shows the list of appliances in the home that were monitored, along with their average power consumption (estimated from turn-on events), the number of events associated with them and the phase (A or B) that they were feeding from. It is worth noting that there is a disproportionately larger number of

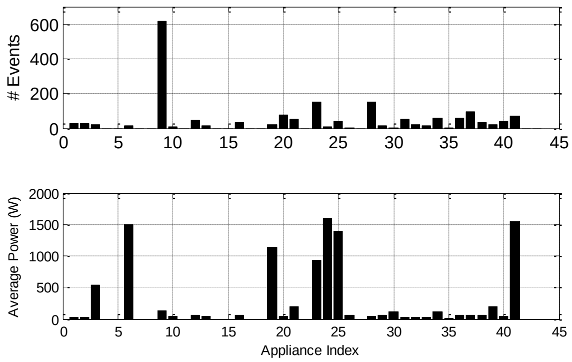

Figure 3 Top: Histogram of number of events for each appliance. Bottom: Average power consumption for each appliance.

events associated with the refrigerator and computer equipment, perhaps because of their continuous mode of operation.

In Table 1, appliances that had no events during the week have a 0 in the "Number of events" column while appliances that did not pass the event criteria mentioned above have a 0 in the "Average power" column. This information is also represented in Figure 3. The top plot shows the frequency of events for each of the appliances and the bottom plot shows the average power consumption of each appliance.

Notably absent from the list in Table 1 are the air conditioner and dryer. Neither of these was used during this week of data collection. For the air conditioner, this is not surprising since it was in October. For the dryer this may not be an uncommon experience to have a week where it is not used. This would also suggest that one week is not enough time to get a representative sampling of appliances.

To illustrate some of the characteristics of the captured data, we also describe the distribution of events over time. Figure 4 shows a histogram of the inter-event times in the dataset, separated by phase. Phases A and B have very distinct characteristics. There are much fewer appliances connected to phase A, so it has fewer overall events. The refrigerator however, is on phase A, and can clearly be seen in the histogram as the spike at a 15-minute inter-event time, which is the approximate duration of the cooling cycle. We also note that on phase A, the longest time between any two events is approximately 50 minutes (the approximate time between refrigerator cooling cycles), but on phase B there are spacings in events for up to 10 hours. This is not shown on the histogram because of scale issues, but there are only 14 instances on phase B of events that occur more than 2 hours apart during the whole week. This corresponds to twice per day: once during the night while everyone is asleep and the other during the day while no one is home.

Figure 5 shows an empirical cumulative distribution function (CDF) of the inter-event times. This graph suggests that almost half of the events require sub-minute resolution in order to allow for them to be uniquely distinguished in the dataset.

4. EXPERIMENTAL RESULTS

In this section we present preliminary experimental results for a simple event detection algorithm applied to BLUED. The event detection algorithm used was a modified generalize likelihood ratio (GLR) detector described in [2] and [3]. We present our

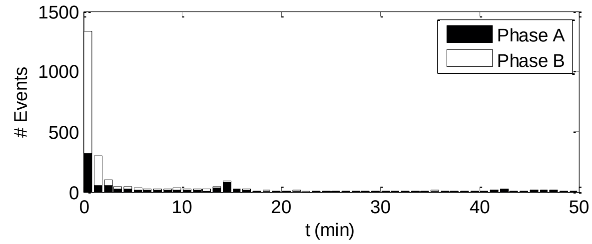

Figure 4 Stacked inter-event time histogram.

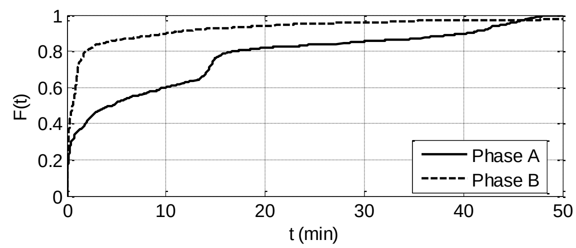

Figure 5 Empirical CDF of inter-event times

results in Table 2 in terms of absolute number of true positives, false positives, and misses (TP, FP, and M, respectively).

Note that in these numbers one can see a large difference in performance on each phase. The activity on phase A is well-spaced and thus the detection rate is very high, while phase B has a lot of electronics and appliances that frequently overlap causing the degradation in detection rate.

Table 2 GLR event detection results

	Phase A	Phase B
TP	855	1092
FP	18	159
M	16	459

We present these results as a benchmark for event detection performance on BLUED, and to facilitate comparisons with other event detection schemes. These numbers were obtained after doing an extensive parameter search for the GLR detector, applied to the entire week of data. We did not split it into training and testing sets, but in future algorithmic development this will be taken into consideration. The particular parameters selected depend on many performance criteria that are beyond the scope of this paper, but the interested reader should refer to [2].

5. CONCLUSIONS

We presented a fully labeled dataset aimed at tackling the lack of publicly available data for event-based NILM algorithm development and testing. We described the data collection system, discussed some of its limitations and suggested improvements for the collection process.

We argue that this dataset is useful for different NILM approaches because of the fine granularity of event labeling. Regarding the representativeness of this dataset, the number of appliances in this house may very well represent the value for the average US home. The 2001 Residential Energy Use Consumption Survey by the Energy Information Administration (EIA) provides ownership estimates for 42 appliances [11]. Given that the survey's appliances are at a more aggregate level (e.g., lighting vs. individual lamps) and that the survey is over ten years old, 50 appliances in a home is not unreasonable and seems likely for many homeowners.

We recognize the importance of a publicly available dataset due to several challenges. As reported by many of our peers, during the 1st International NILM Workshop [1], the inherent financial and time costs, as well as the lack of standardized sensing and labeling approaches, keep most of the data collection efforts confined in controlled lab environments. We believe that representative, whole-home datasets like the one we presented here, allow for the evaluation of NILM algorithms under more realistic scenarios. It is our hope that this dataset can enable the research community to compare the performance of different algorithms, in the same way that standardized datasets have helped in other domains.

6. FUTURE WORK

We plan to incrementally expand the dataset to other homes in order to include examples of other appliances, climate zones, seasons, and household compositions (e.g., families with children, working families, etc.). To capture a richer dataset it will also be necessary to monitor the homes for a much longer period of time.

We are also developing our own detection and classification algorithms and plan to continue this work in parallel with the expansion of the dataset.

Another major challenge for the NILM research community, besides the lack of reference datasets, is the need for standardized performance metrics. We are currently in the process of creating metrics for different parts of the NILM problem and using them to evaluate our algorithms.

7. ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge the support from the Robert Bosch LLC Research and Technology Center North America and the National Science Foundation (NSF) grant #09-30868. The opinions expressed herein are those of the authors and not of the NSF. Thanks also to Da Eun Sung for her help in processing the dataset.

8. REFERENCES

[1] 1st International Workshop on Non-Intrusive Load Monitoring, Pittsburgh, PA, May 2012. http://www.ices.cmu.edu/psii/nilm/

[2] Anderson, K.D., Berges, M.E., Ocneanu, A., Benitez, D., and Moura, J.M.F. "Event Detection for Non Intrusive Load Monitoring", *Submitted* IEEE Industrial Electronics Conference (IECON), Montreal, Canada, 2012.

[3] Berges, M., Goldman, E., Soibelman, L., Matthews, H.S., and Anderson, K. "User-centric Non-Intrusive Electricity Load Monitoring for Residential Buildings". ASCE Journal of Computing in Civil Engineering, vol. 25, no. 471, 2011.

[4] Buevich, M., Rowe, A., and Rajkumar, R. "Tracking and Visualization of Building Energy," The 1st International Workshop on Cyber-Physical Systems, Networks, and Applications (CPSNA 2011), Toyama, Japan, August 2011.

[5] Goncalves, H., Ocneanu, A.F., and Bergés, M.E. "Unsupervised Disaggregation of Appliances using Aggregated Consumption Data", Proceedings of the SustKDD Workshop on Data Mining Applications in Sustainability, 2011.

[6] Kolter, J.Z., and Johnson, M.J. "REDD: A Public Data Set for Energy Disaggregation Research", Proceedings of the SustKDD Workshop on Data Mining Applications in Sustainability, 2011.

[7] Kolter, J.Z., and Jaakkola, T. "Approximation Inference in Additive Factorial HMMs with Application to Energy Disaggregation", Proceedings of the International Conference on Artificial Intelligence and Statistics, 2012.

[8] Lee, K.D. Electric Load Information System based on Non-Intrusive Power Monitoring, Ph.D. thesis, Massachusetts Institute of Technology, 2003.

[9] Norfold, L.K., and Leeb, S.B. "Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms", Energy and Buildings, Vol. 24, pp. 51-64, 1996.

[10] Shao, H., Marwah, M., and Ramakrishnan, N. "A Temporal Motif Mining Approach to Unsupervised Energy Disaggregation" Proceedings of 1st International Non-Intrusive Load Monitoring Workshop, 2012, http://www.ices.cmu.edu/psii/nilm/abstracts/shao_VT_NILM2012_abstract.pdf

[11] U.S. Energy Information Administration, "End-Use Consumption of Electricity 2001" http://www.eia.gov/emeu/recs/recs2001/enduse2001/enduse2001.html

[12] Zeifman, M., and Roth, K. "Nonintrusive Appliance Load Monitoring: Review and Outlook", IEEE Transactions on Consumer Electronics, Vol. 57, No. 1, February 2011.

About the authors:

Kyle Anderson is a Ph.D. student with the Department of Electrical and Computer Engineering at Carnegie Mellon University. His research interests include non-intrusive load monitoring, sensor networks, and signal processing.

Adrian Ocneanu is a Ph.D. student with the Department of Civil and Environmental Engineering Carnegie Mellon University. His research interests include pattern recognition, renewable energies, non-intrusive monitoring, and signal processing.

Diego Benítez is a Senior Research Engineer at Robert Bosch LLC Research and Technology Center in Pittsburgh, PA. His research interests include intelligent instrumentation and measurement systems for energy management, security and smart buildings applications.

Derrick Carlson is a Ph.D. student with the Department of Civil and Environmental Engineering at Carnegie Mellon University. His research interests include building energy management, residential energy efficiency, energy policy, carbon footprinting, and life-cycle assessment.

Anthony Rowe is an Assistant Research Professor in the Department of Electrical and Computer Engineering at Carnegie Mellon University. His research interests include real-time systems, operating systems, light-weight computer vision, and robotics.

Mario Bergés is an Assistant Professor in the Department of Civil and Environmental Engineering at Carnegie Mellon University. His research interests include infrastructure monitoring, building energy management, smart grid, machine learning for signal processing, and sensor networks.