Modeling of End-Use Energy Profile: An Appliance-Data-Driven Stochastic Approach

Zhaoyi Kang, Ming Jin, Costas J Spanos
Dept. of Electrical Engineering & Computer Sciences, UC Berkeley, Berkeley, CA 94709
{kangzy, jinming, spanos}@berkeley.edu

***Abstract*—In this paper, the modeling of building end-use energy profile is comprehensively investigated. *Top-down* and *Bottom-up* approaches are discussed with a focus on the latter for better integration with occupant information. Compared to the Time-Of-Use (TOU) data used in previous *Bottom-up* models, this work utilizes high frequency sampled appliance power consumption data from wireless sensor network, and hence builds an appliance-data-driven probability based end-use energy profile model. ON/OFF probabilities of appliances are used in this model, to build a non-homogeneous Markov Chain, compared to the duration statistics based model that is widely used in other works. The simulation results show the capability of the model to capture the diversity and variability of different categories of end-use appliance energy profile, which can further help on the design of a modern robust building power system.**

I. Introduction

Buildings account for more than 40% of the total power consumption in the US, and can play a critical role in addressing the current energy and climate issues [1]. Significant effort has been invested in this topic, from benchmarking, to control and monitoring. In this paper, we will discuss the modeling of end-use energy profile of the commercial building power system.

The modeling of end-use energy profile is an important task, and is of particular interests especially in recent years because of the following reasons. Nowadays, building energy usually depends greatly on occupant behavior, especially at fine-grained metering level, such as plug-in loads, user-controlled lighting, user-adjusted HVAC, etc. [2], which brings about significant amount of diversity and fluctuation. End-use profile is believed to be able to capture, quantify and predict those variability and complicated relationships.

Moreover, as people endeavor to integrate renewable energy resources to traditional building power system, and the wide adoption of energy-efficient appliances and policies, we need accurate and robust models to understand the feasibility of such schemes and to evaluate the effects of such innovations.

Last but not least, as an important potential input of building energy & indoor climate simulation software, end-use energy profile can be widely used in early-stage building environmental design and energy system planning.

This paper is organized as follows. In Section II, a literature review is given. In Section III, the data collection and processing methods in this work is described. In Section IV, the key modules in the model are illustrated and investigated. In Section V, we run simulation and discuss the results. Finally, conclusions are drawn and discussed in Section VI.

II. Literature Review

The models of end-use energy profile can be divided into two categories, the *Top-down* approach and the *Bottom-up* approach, with reference to the hierarchical structure of data inside the whole system [3], as illustrated in Figure 1 [4].

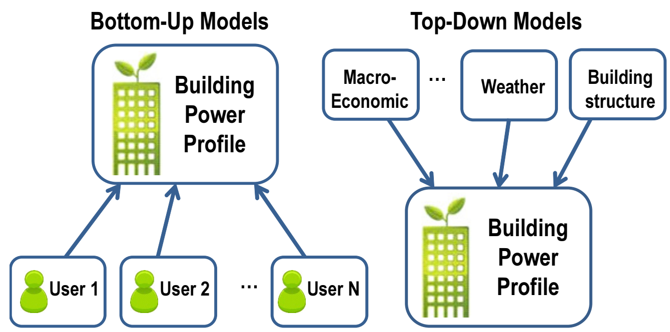

Fig. 1: Two types of approaches: Top-down and Bottom-up

The *Top-down* approach treats building as a black-box, and calculate the *collective demand* of the building. Usually, the variability of the energy profile is captured as a linear model based on *macro*-scale extraneous variables such as macroeconomic indicators (gross domestic product (GDP), income, price rate), climate, building construction, etc. [4]. The parameters of the model are estimated from training data and the energy profile of a new building can be extrapolated.

The *Bottom-up* approach, on the contrary, takes into consideration the contribution of each individual sector. Specifically, the occupant-oriented energy consumption is included, and the variability is captured either as a statistical model of the users based on *macro*-scale information, or stochastic sequences of the user patterns. The parameters of the models are estimated from a group of building energy consumption data or Time-Of-Use (TOU) survey data.

Bottom-up approaches are more recent and attracting attention, because of the following reasons:

- The building energy performance becomes more and more sensitive to occupant behaviors. The occupant-dependent variability is better captured by *Bottom-up* approach, whereas *Top-down* approach does not typically have the flexibility to model that.

- *Bottom-up* approach helps to understand the occupant-dependent demand profile, which is useful in designing building demand-response system [5].
- The *Bottom-up* approaches better adapt to the changes in the building infrastructure and new technologies & policies, while the *Top-down* approach relies a lot on historical data.

One of the earliest works of *Bottom-up* approach is by A. Capasso *et al.* [6]. They use availability probability to model presence of each member in a house, and activity probability to model presence of each activity. The probabilities are learned from TOU data. Together with duration statistics obtained from prior knowledge, power stream can be generated by Monte Carlo (MC) simulation. In [7], TOU data is also used, and nine synthetic activity patterns are defined. Non-homogeneous Markov Chain is used to model the ON/OFF of activities. Duration and ON events are sampled randomly. In [8], activity probability is trained from TOU data and other extraneous data, so that is non-homogeneous. In [9], effort is put purely in estimating activity probability patterns based on TOU survey and duration statistics.

The existing methods that employ the *Bottom-up* approach provide great insights into the building end-use profile. However, there are some issues that need to be addressed:

- Previous works mostly used TOU data to obtain activity probability, and then convert the activity to appliance pattern. Sometimes this is problematic, since the conversion is usually not rigorously defined.
- Cross-correlation among appliances are not captured because of the conversion. A random Markov Chain model could under-estimate the demand. Moreover, most previous works mentioned about modeling *shared activities*, whereas validation of those models is difficult.
- In commercial buildings, variation among buildings is not of significant interest since their infrastructures can vary a lot. However, the variation caused by occupant fluctuation becomes especially important.

In this work, we will directly estimate probability patterns of appliances in commercial building, thanks to the large-scale wireless sensor network and distributed data storage system. Thus, we developed a model based on the ON/OFF probability to quantify the variation of building end-use energy profile.

III. Data

A. Building Profile

Our experimental space is in 406 Cory Hall at University of California Berkeley, an office shared by 25 occupants. Depending on the sets of appliances that each user owns, we can divide the users into six categories: A) 1 laptop, 1 monitor, 1 desktop; B) 1 desktop, 1 monitor; C) 1 laptop, 1 monitor; D) 1 laptop, 2 monitors; E) 2 laptops, 1 monitor; F) 1 laptop. The category can be changed if the sets of appliances change among users. With this categorization, we now have a *profile* that can describe the building's basic occupancy.

B. Data Collection

We collect appliance energy consumption through a large-scale wireless sensor network (WSN). WSN have been implemented in many different scenarios to facilitate the system estimation, conditioning, diagnosis [2] [10] [11].

DENT meter [12] is used to collect whole space real-time power consumption data. The DENT meter has 18 channels, each one monitoring a subset of appliances, e.g. plug loads, lights, kitchenware etc. ACme sensors are used to collect real-time power consumption of each occupant [13]. The data is handled using the sMAP protocol [14]. We implement one ACme sensor for each occupant to optimize the cost and experiment performance. The states of each appliance are filtered out by the power dis-aggregation algorithm from the aggregated occupant-level power consumption [15].

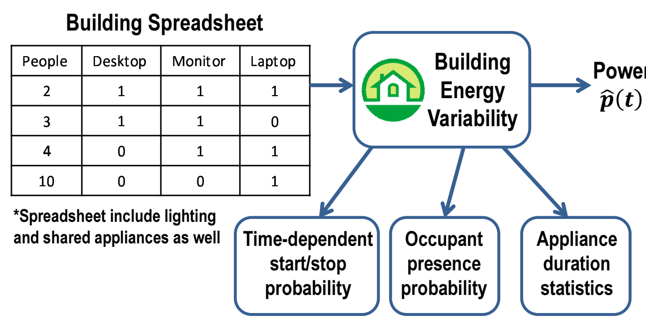

Fig. 2: Schematic of the model: Building Profile Model (BPM)

C. Power Dis-aggregation

In power dis-aggregation, we *decode* the ON/OFF state of individual appliance from an observed *aggregated* power stream. Let $p_t, \forall t = 1, \cdots, T$ be the aggregated power stream from n appliances. Let $\mathbf{S}_t$ be the state vector of the n appliances at step t. Our task is to infer $\mathbf{S}_t$ from p_t. $\mathbf{S}_t$ is a vector of n binary variables, one for each appliance, i.e. $\mathbf{S}_t \in \{0,1\}^n$, in which 1 for ON, 0 for OFF. There are in total 2^n combinations of ON/OFF states.

Several models have been used to solve this problem, including Hidden Markov Model [16], change detection [17], sparse coding [18]. In this work, we use a method based on multiple hypothesis testing [15]. It should be noted that, when an appliance has more than one states, e.g. washing machine, the appliance can be modeled in mixture model, as in [16].

IV. Model Framework

A. Big Picture

In our work, we build a *Building Profile Model* (BPM) to estimate end-use profile. BPM generates energy profile through a parametric model. The parameters include the occupants' information, the appliance categories, the ON/OFF-probabilities, user presence probability, overnight probability, and/or appliance duration statistics, generalized from historical sensor recorded data and prior knowledge. The BPM has potential of *model re-use* to a similar building which share some parameters with the building under study in this work. The schematic of the BPM is illustrated in Figure 2. In this section, we will start from the three widely used basic models, discuss their potential benefits in our scenario, and eventually arrive at our comprehensive BPM.

To facilitate the analysis, for certain appliance, given that we have M days of observations, we define $S_t^{(m)}$ as its state of m-th day, i.e. $S_t^{(m)} \in \{0, 1\}$ and 1 stands for ON.

B. Rate-of-Use Statistics

One of the basic models describing the appliance usage utilizes the Rate-Of-Use (ROU) statistics.

Definition 1 (Rate-Of-Use)**.** *Rate-Of-Use (ROU) is the portion of time that the appliance is ON in each time-of-day:*

$$\mathrm{ROU}_t = \frac{1}{M} \sum_{m=1}^{M} S_t^{(m)} = \overline{S_t} \tag{1}$$

For example, in the 80 days of experiment, the monitor is ON at 12:00PM in 16 days, the ROU would be $16/80 = 0.2$ at 12:00PM. The ROU is plotted for monitor, laptop and desktop in Figure 3. Strong daily pattern is observed. ROU indicates the average energy consumption, but it doesn't indicate the usage pattern of the appliance.

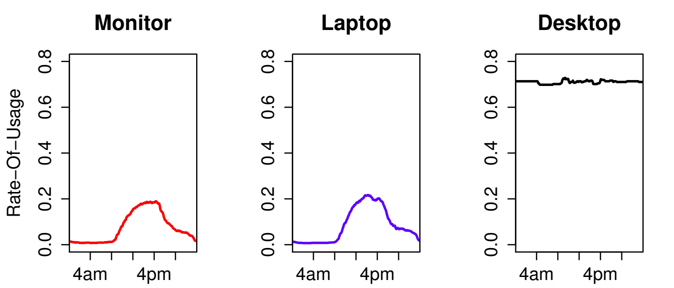

Fig. 3: Rate-Of-Use of three types of appliances: monitor (left), laptop (middle) and desktop (right)

C. ON/OFF-Probability Statistics

Another model utilizes the ON/OFF-probability [8] [9], i.e. the probability of turning-ON/OFF at each time step.

Definition 2 (ON/OFF Probability)**.** *For certain appliance at t, the empirical ON/OFF probability is defined as* $\widehat{P}_t^{\mathbf{ON/OFF}}$*:*

$$\widehat{P}_t^{\mathbf{ON}} = \frac{\sum_{m=1}^{M} S_t^{(m)}(1 - S_{t-1}^{(m)})}{\sum_{m=1}^{M}(1 - S_{t-1}^{(m)})} = \frac{\overline{S_t} - \overline{S_t S_{t-1}}}{1 - \overline{S_{t-1}}} \tag{2}$$

$$\widehat{P}_t^{\mathbf{OFF}} = \frac{\sum_{m=1}^{M} S_{t-1}^{(m)}(1 - S_t^{(m)})}{\sum_{m=1}^{M} S_{t-1}^{(m)}} = \frac{\overline{S_{t-1}} - \overline{S_{t-1} S_t}}{\overline{S_{t-1}}} \tag{3}$$

with which we can do MC simulation to obtain the state sequences of all the appliances that we are interested in.

Definition 3. *After we run J MC simulations, we defined the simulated state in the j-th MC run as $\widehat{S}_{1:T}^j$, $j = 1, \cdots, J$.*

Compared to ROU model, ON/OFF probability model can capture the usage pattern [2] [8] [9]. Previously this model is built upon some time slots, e.g. "0∼8AM", "8∼9AM", "9∼11:30AM", "11:30∼1:30PM", "1:30∼5PM", "5∼7PM", "7∼9:30PM" and "9:30PM∼0AM". The ON/OFF probability is assumed to be constant within each time slots.

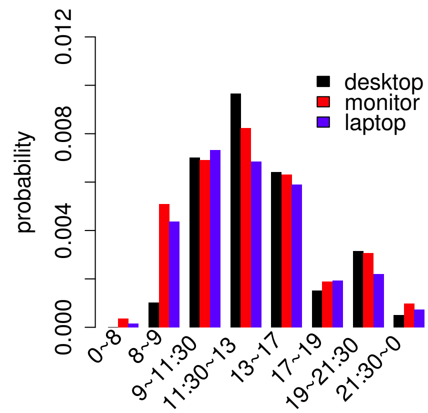

Fig. 4: Time-dependent ON probability of three types of appliances: desktop (black), monitor (red) and laptop (blue)

The time-slot-based ON-probability $\widetilde{P}_t^{\mathbf{ON}}$ is shown in Figure 4, for desktop, monitor and laptop. Note that in Figure 3 the desktop pattern seems to be at constant line, which is due to the limited number of desktops in our test space, and because some of them are kept on overnight (i.e, their $\widetilde{P}_t^{\mathbf{OFF}}$ is small once they are ON). To simulate turning-ON, we use the probability of $\widetilde{P}_t^{\mathbf{ON}}/T_{\mathbf{SLOT}}$, in which $T_{\mathbf{SLOT}}$ is the length the time slots. For example, at time interval "8∼9AM", if we use 5 min interval step, $T_{\mathbf{SLOT}} = 12$.

One concern about the time-slot-based model is that the probability inside each slot is not well captured. According to a simple Poisson model assuming independent events, within each time slot, the ON events are geometrically distributed. However, as shown in Figure 5 where we take monitor as example, most of them do not follow the model. The pattern of laptop and desktop can also demonstrate such discrepancy.

D. Duration Statistics

Previously, duration statistics are used to characterize the duration time of each activity [8] [9]. We extracted the duration statistics from sensor data after power dis-aggregation. The results are shown in Figure 6 for office appliances. A potential problem is the limited capability to model the turn-off events of the appliances.

E. Our Model

In our model, we use an appliance-data-driven high resolution ON/OFF probability model.

- We extract the probability that an appliance is present in some day, marked as $P_{\mathbf{PRES}}$, as well as the probability that an appliance is ON overnight, marked as $P_{\mathbf{INIT}}$, from the data. At the same time, we extract the appliance

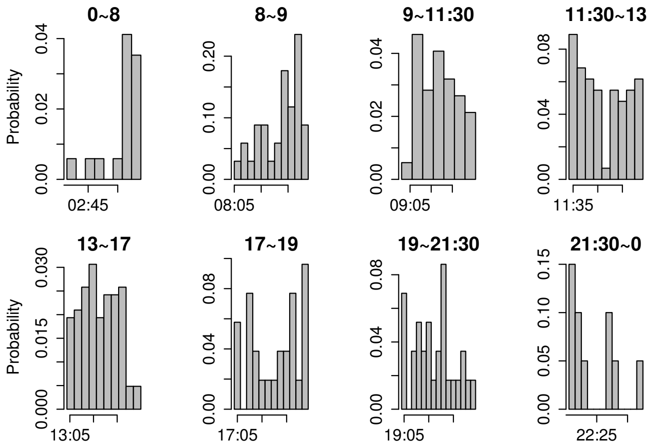

Fig. 5: ON probability inside each time slot for monitor

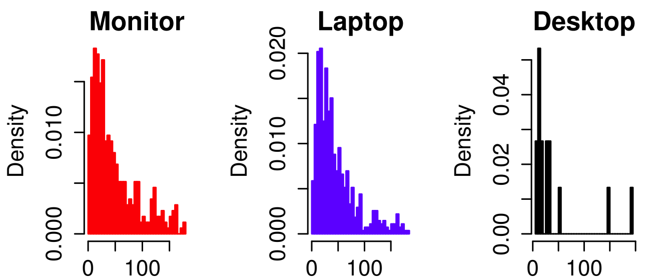

Fig. 6: Histogram of duration statistics in minutes of three types of appliances: monitor (left), laptop (middle) and desktop (right). X axis is in 5 minutes interval

ON/OFF probabilities $\hat{P}_t^{\mathbf{ON/OFF}}$ from those days that the user is present. From wireless sensor network, we collect the appliance power data, we can build the model based on appliance information, instead of on activities as in other works, in which an often problematic activity-to-appliance transformation is needed [4].

- Both ON/OFF probabilities are included and are formulated in a Markov Chain framework, whereas duration statistics are not included. Therefore, we can better model the turning-OFF events of the appliances.

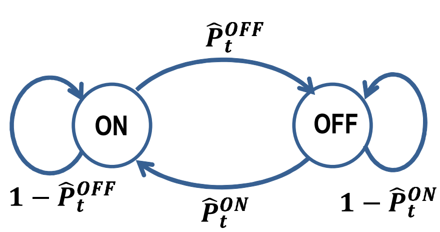

Fig. 7: FSM interpretation of the model

- Instead of the time-slot model in Figure 4, we use a non-homogeneous Markov Chain model for both ON/OFF probabilities. For each appliance, the model can be interpreted as a two-state Finite State Machine (FSM) at each step (Figure 7), with $\hat{P}_t^{\mathbf{ON/OFF}}$ as switching probabilities. Based on this FSM, we can use Monte Carlo to simulate the usage patterns of all the appliances in the space, and we can summarize statistical properties from the Monte Carlo simulations.

The MC simulated appliance ON/OFF sequences *(a)* can capture non-homogeneous stochasticity of appliance usage patterns and is easily extended to analyze new techniques & policies; *(b)* statistically converges to the ROU model in estimating states, which means this method is essentially reasonable in end-use energy profile modeling. This can be shown in the theorem below:

Theorem 1 (Convergence of MC Simulation)**.** *If $\widehat{S}^j_{1:T}$ is the j^{th} MC simulated time series from the FSM as in Figure 7 and we have J such MC simulations, then $\mathrm{E}[\frac{1}{J}\sum_j \widehat{S}^j_t] = \overline{S_t}$, in which $\overline{S_t}$ is the ROU, and $\lim_{J\to\infty}\mathrm{Var}(\frac{1}{J}\sum_j \widehat{S}^j_t) \to 0$. In other words, MC simulation converges a.s. to ROU.*

Proof. Let $\widehat{S}_1, \cdots, \widehat{S}_t$ be the states at different time steps from MC simulation. Assume that the states follows Markov Property, s.t. $\Pr(\widehat{S}_t|\widehat{S}_{t-1}, \cdots, \widehat{S}_1) = \Pr(\widehat{S}_t|\widehat{S}_{t-1})$. Then by the chain rule of expectation [19], we have:

$$\mathrm{E}[\widehat{S}_t] = \mathrm{E}[\mathrm{E}[\widehat{S}_t|\widehat{S}_{t-1}]] \tag{4}$$

Since we have:

$$\begin{aligned}\mathrm{E}[\widehat{S}_t|\widehat{S}_{t-1}] &= \Pr(\widehat{S}_t = 1|\widehat{S}_{t-1})\\ &= \widehat{P}_t^{\mathbf{ON}}(1-\widehat{S}_{t-1}) + (1-\widehat{P}_t^{\mathbf{OFF}})\widehat{S}_{t-1}\\ &= \widehat{P}_t^{\mathbf{ON}} + (1-\widehat{P}_t^{\mathbf{ON}} - \widehat{P}_t^{\mathbf{OFF}})\widehat{S}_{t-1}\end{aligned} \tag{5}$$

Let $G_t = 1-\widehat{P}_t^{\mathbf{ON}} - \widehat{P}_t^{\mathbf{OFF}} = \frac{\overline{S_tS_{t-1}} - \overline{S_t}\cdot\overline{S_{t-1}}}{(1-\overline{S_{t-1}})\overline{S_{t-1}}}$, combining (4) and (5) we obtain:

$$\mathrm{E}[\widehat{S}_t] = \widehat{P}_t^{\mathbf{ON}} + G_t\mathrm{E}[\widehat{S}_{t-1}] \tag{6}$$

Therefore, we can iteratively write $\mathrm{E}[\widehat{S}_t]$ as:

$$\mathbf{E}[\widehat{S}_t] = \widehat{P}_t^{\mathbf{ON}} + \sum_{\tau=3}^{t}\widehat{P}_{\tau-1}^{\mathbf{ON}}\prod_{i=\tau}^{t}G_i + \mathbf{E}[\widehat{S}_1]\prod_{i=2}^{t}G_i \tag{7}$$

The initial state at $t=1$ in MC simulation is generated from a Bernoulli process $p_1 = \mathbf{E}[\widehat{S}_1] = \overline{S_1}$. We put the expression of $\hat{P}_t^{\mathbf{ON/OFF}}$ as (2) and (3) in (7).

$$\widehat{P}_2^{\mathbf{ON}}\prod_{i=3}^{t}G_i + \overline{S_1}\prod_{i=2}^{t}G_i = \overline{S_2}\prod_{i=3}^{t}G_i \tag{8}$$

Then we have the following equation:

$$\mathbf{E}[\widehat{S}_t] = \widehat{P}_t^{\mathbf{ON}} + \sum_{\tau=4}^{t}\widehat{P}_{\tau-1}^{\mathbf{ON}}\prod_{i=\tau}^{t}G_i + \overline{S_2}\prod_{i=3}^{t}G_i$$

Therefore, we can simply equation (7) as:

$$\begin{aligned}\mathbf{E}[\widehat{S}_t] &= \widehat{P}_t^{\mathbf{ON}} + \overline{S_{t-1}}G_t\\ &= \frac{\overline{S_t} - \overline{S_tS_{t-1}}}{1-\overline{S_{t-1}}} + \frac{\overline{S_tS_{t-1}} - \overline{S_t}\cdot\overline{S_{t-1}}}{1-\overline{S_{t-1}}} = \overline{S_t}\end{aligned} \tag{9}$$

Since $\widehat{S}^j_t$s are all binary sequences, $\mathrm{Var}(\widehat{S}^j_t) = \overline{S_t}(1-\overline{S_t})$ and naturally $\lim_{J\to\infty}\mathrm{Var}(\frac{1}{J}\sum_j \widehat{S}^j_t) = \lim_{J\to\infty}\frac{1}{J}\mathrm{Var}(\widehat{S}^j_t) \to 0$.

Thus, MC simulation converges to the ROU. It should, however, be noted that Theorem 1 holds only if the ON/OFF probabilities are consistent between simulation and observation. □

F. Modeling of Cross-Correlation

In our experimental space, we have 11 monitors, 5 desktops, 14 laptops. Assuming that devices in the same category are the same, we can simulate each appliance independently and aggregate them. The mean of the aggregation, as a corollary of Theorem 1, is unbiased. The variance, however, could be underestimated. Cross-correlation among appliances needs to be addressed. In MC simulation, cross-correlation between Bernoulli sequences is difficult. Instead, we propose a way to theoretically correct the variance estimation as follow.

Let $S_{t,i}$ be the state of i-th single appliance, its variance $\mathrm{Var}(S_{t,i}) = \sigma_D^2$ we already know, D = desktop, monitor, laptop is the appliance type, then the aggregated variance of N different appliances is:

$$\mathrm{Var}\left(\sum_{i=1}^{N} S_{t,i}\right) = \sum_{i=1}^{N} \sigma_{t,a(i)}^2 + \sum_{i \neq j} \mathrm{cov}(S_{t,i}, S_{t,j}) \quad (10)$$

$a(i)$ is the type of the i-th appliance, and $\sum_{i=1}^{N} \sigma_{t,a(i)}^2 = \sum_{d \in D} \sigma_{t,d}^2 N_d$, d is the number of appliances in type d.

The term $\sum_{i=1}^{N} \sigma_{t,a(i)}^2$ is the uncorrelated variance, and the other term in RHS of (11) can be simplified as below:

$$\begin{aligned}\mathrm{cov}(S_{t,i}, S_{t,j}) = & \sum_{d \in D} \sigma_{t,d}^2 N_d (N_d - 1) \rho_d \\ & + \sum_{d_1, d_2 \in D} \sigma_{t,d_1} \sigma_{t,d_2} N_{d_1} N_{d_2} \rho_{d_1,d_2} \end{aligned} \quad (11)$$

where ρ_d is the correlation within each types of appliance, and ρ_{d_1,d_2} is the average correlation between different types of appliances, both extracted from measurement.

V. Results and Discussions

A. ON/OFF Probability Model Estimation

As we estimate the ON/OFF probabilities, when the data points are sparse, we smooth the empirical probability function in (2) and (3) by a Kernel Smoother as below:

$$\widetilde{P}_t^{\mathbf{ON/OFF}} = \frac{\sum_{i=1}^{T} K(t,i) \widehat{P}_i^{\mathbf{ON/OFF}}}{\sum_{i=1}^{T} K(t,i)} \quad (12)$$

in which $K(t,i) = \exp\left(-\frac{(i-t)^2}{2h^2}\right)$ as Gaussian kernel, and bandwidth was chosen as the plug-in bandwidth (hpi) [20].

Remark 1. *If we use $\widetilde{P}_i^{\mathbf{ON/OFF}}$ instead of $\widehat{P}_i^{\mathbf{ON/OFF}}$, Theorem 1 no longer holds. However, under reasonably chosen bandwidth of the function $K(\cdot)$, since (6) is in closed form, and generated state will also be a* smoothed *version of ROU. It should be noted here that a strict analysis on the condition of the bandwidth would be required to fully understand the performance of smoothing, and because of the length and scope of this work, this will be a subject of future work.*

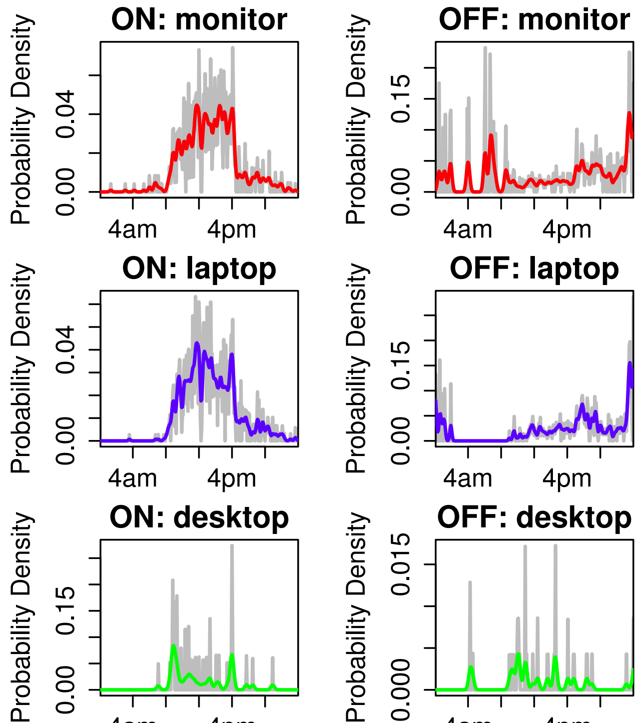

Fig. 8: ON/OFF Probability in 5 min interval for Monitor, Laptop, and Desktop. Gray lines: Measurement; Colored lines: Kernel smoothed

1) Office Appliances: The office appliances include *monitor, laptop, and desktop*. The estimated ON/OFF Probabilities for the three kinds of appliances are shown in 8. It is observed that the ON probability peaks at early morning and decreases during the day, whereas OFF probability peaks at late in the day. It should be noted that the data regarding to desktop is sparse and the ON/OFF probabilities contain more uncertainty. We only include weekdays in our study.

2) Pathway/Room Lighting: The lighting power consumption is a major contributor to building energy profile. In our test space in Cory 406 at UC Berkeley, we have pathway lighting and room lighting. Pathway lighting is shared in large working area and is more standard in schedule. Room lighting has motion sensor so that it is more adaptive to occupant behavior. The PowerScout data we collected contains the aggregated signal of lighting power in 7 rooms. For model simplicity, we assumes that the 7 rooms are the same. The result is shown in Figure 9.

The pathway lighting has little overnight activity, and the estimation has more bias, since in (3), $\overline{S_t}$ is zero for some t. We give those data point a probability of 0.5.

3) Shared Appliances: Shared appliances include a microwave, a water heater, a coffee maker, and a refrigerator. The water heater and refrigerator have strong periodic patterns, and are less dependent on occupants. The microwave and coffee maker shows spike-like patterns and the records of usage are

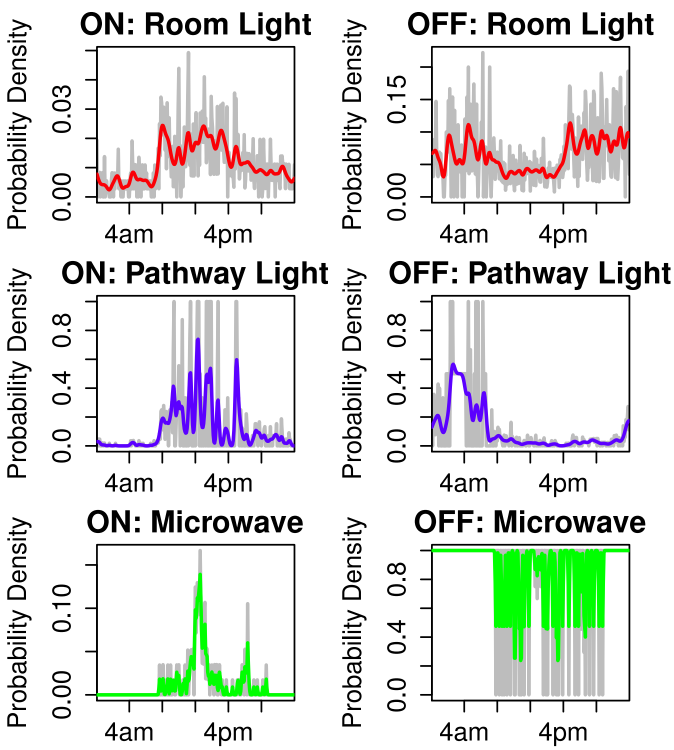

Fig. 9: ON/OFF Probability in 5 min interval for Room lighting, Pathway lighting and Microwave. Gray lines: Measurement; Colored lines: Kernel smoothed

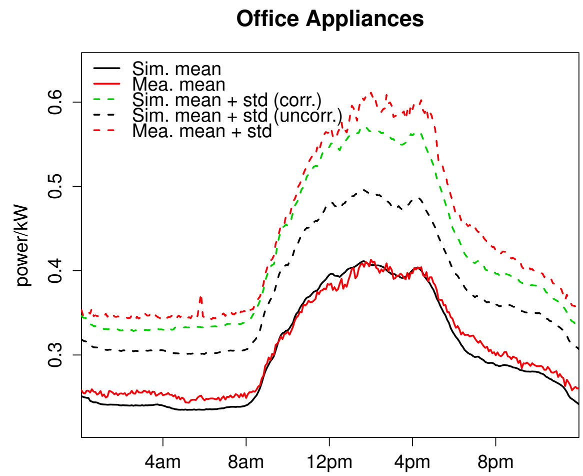

Fig. 10: MC Simulation for office appliances (11 monitors, 14 laptops and 5 desktops). *Sim.* stands for simulation; *Mea.* stands for measurement. Both mean and std are shown here, with *corr.* stands for simulation results with correction from Equation (10), *uncorr.* stands for simulation without the correction from (10).

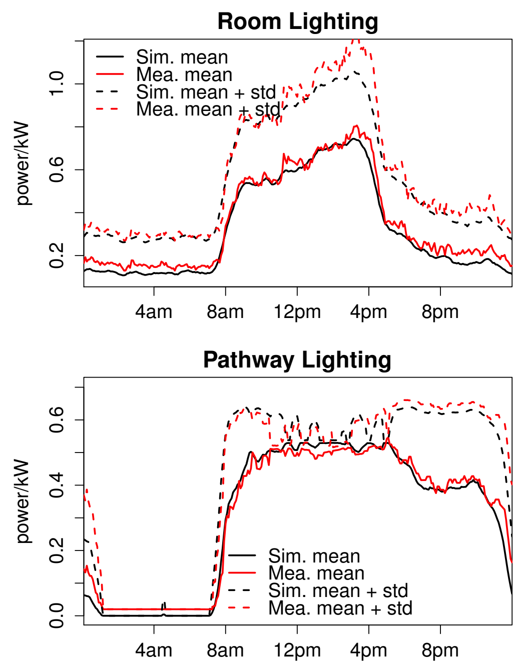

Fig. 11: MC Simulation Results for Room and Pathway Lighting. *Sim.* stands for simulation; *Mea.* stands for measurement.

sparse. The estimated probability densities for the Microwave are also shown in Figure 9. Notice that the OFF probability is very high since the duration of each ON is usually very short, compared to our 5-minute estimation interval.

B. Monte Carlo Simulation

We use 10'000 runs of MC simulations. In each run, we follow the steps described here:

Firstly, we generate random sample with probability $P_{\mathbf{PRES}}$, if the outcome is 0, the appliance is not present. If the outcome is 1, then we generate startup state $S_1 = \mathrm{Ber}(P_{\mathbf{INIT}})$.

Secondly, we simulate all the appliances of certain type and sum them up. After that, we correct the sample variance term with the cross-correlation terms as in (10) and (11).

The simulated end-use energy profiles are shown in Figure 10, Figure 11, and Figure 12, for office appliances, lighting, and representative shared appliances (we pick the microwave, other shared appliances work similarly), respectively. Both mean and standard deviation are extracted from MC simulation and only the upper bound of standard deviation is plotted since it is of more interest in early-stage demand estimation. Generally speaking, the model performs in all three categories of appliances. At the same time, we also have some interesting findings.

- Note that in Figure 10, cross correlation is shown to better capture the standard deviation level, which means that in the end-use profile modeling, the correlations among appliances have large impact on the overall variability.
- The standard deviation is poorly captured for microwave (not shown in Figure 12) and other appliances with spiking patterns, because of the sparse pattern. Variation-reduction techniques such as importance sampling or

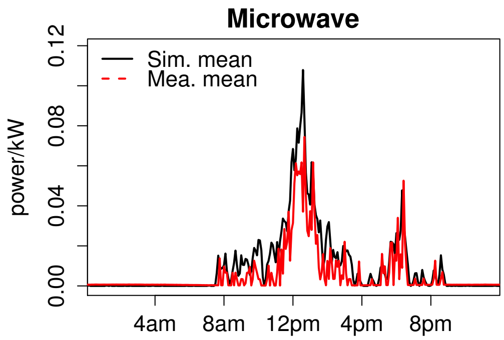

Fig. 12: MC Simulation Results for Microwave. *Sim.* stands for simulation; *Mea.* stands for measurement.

Markov Chain Monte Carlo (MCMC) could be used in the future to reduce the fluctuation.

It should be expected that, in a larger office building, when more appliances are present, our model can be more capable to capture the overnight patterns. Moreover, it should be noted that, when the building occupancy schematic changes, the only thing that needs to be tuned is the building *profile*. As long as we have a reasonable category of users, we can evaluate the building energy performance accordingly.

VI. Conclusion

In this paper, the modeling of end-use energy profile is comprehensively investigated. The two categories *Top-down* and *Bottom-up* approaches are discussed and the latter is preferred because of the better integration with occupant-oriented information. Compared to the Time-Of-Use (TOU) data used in previous *Bottom-up* model, this work utilizes high frequency sampled data from wireless sensor network, and builds an appliance-data-driven end-use model. ON/OFF probabilities of appliances are extracted, and a theoretically unbiased FSM Markov Chain model is developed, with cross-correlation correction. The simulation results show the capability of the model to capture diversity and variability of building end-use energy profile, which can further help on the design of robust building power system.

Acknowledgment

This research is funded by the Republic of Singapore's National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. BEARS has been established by the University of California, Berkeley as a center for intellectual excellence in research and education in Singapore.

References

[1] S. Chu and A. Majumdar, "Opportunities and challenges for a sustainable energy future," *nature*, vol. 488, no. 7411, pp. 294–303, 2012.

[2] W. Kastner *et al.*, "Communication systems for building automation and control," *Proceedings of the IEEE*, vol. 93, no. 6, pp. 1178–1203, 2005.

[3] L. G. Swan and V. I. Ugursal, "Modeling of end-use energy consumption in the residential sector: A review of modeling techniques," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 8, pp. 1819–1835, 2009.

[4] A. Grandjean, J. Adnot, and G. Binet, "A review and an analysis of the residential electric load curve models," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 9, pp. 6539–6565, 2012.

[5] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *Industrial Informatics, IEEE Transactions on*, vol. 7, no. 3, pp. 381–388, 2011.

[6] A. Capasso, W. Grattieri, R. Lamedica, and A. Prudenzi, "A bottom-up approach to residential load modeling," *Power Systems, IEEE Transactions on*, vol. 9, no. 2, pp. 957–964, 1994.

[7] J. Widén and E. Wäckelgård, "A high-resolution stochastic model of domestic activity patterns and electricity demand," *Applied Energy*, vol. 87, no. 6, pp. 1880–1892, 2010.

[8] I. Richardson, M. Thomson, D. Infield, and C. Clifford, "Domestic electricity use: A high-resolution energy demand model," *Energy and Buildings*, vol. 42, no. 10, pp. 1878–1887, 2010.

[9] U. Wilke, F. Haldi, J.-L. Scartezzini, and D. Robinson, "A bottom-up stochastic model to predict building occupants' time-dependent activities," *Building and Environment*, vol. 60, pp. 254–264, 2013.

[10] S. Lanzisera *et al.*, "Data network equipment energy use and savings potential in buildings," *Energy Efficiency*, vol. 5, no. 2, pp. 149–162, 2012.

[11] H. Karl and A. Willig, *Protocols and architectures for wireless sensor networks*. Wiley-Interscience, 2007.

[12] "Dent meter," http://www.dentinstruments.com/, accessed: 2014-04-10.

[13] X. Jiang, S. Dawson-Haggerty, P. Dutta, and D. Culler, "Design and implementation of a high-fidelity ac metering network," in *Information Processing in Sensor Networks, 2009. IPSN 2009. International Conference on*. IEEE, 2009, pp. 253–264.

[14] S. Dawson-Haggerty, X. Jiang, G. Tolle, J. Ortiz, and D. Culler, "smap: a simple measurement and actuation profile for physical information," in *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2010, pp. 197–210.

[15] Z. Kang, Y. Zhou, L. Zhang, and C. J. Spanos, "Virtual power sensing based on a multiple-hypothesis sequential test," in *Smart Grid Communications (SmartGridComm), 2013 IEEE International Conference on*. IEEE, 2013, pp. 785–790.

[16] H. Kim, "Unsupervised disaggregation of low frequency power measurements," Ph.D. dissertation, University of Illinois, 2012.

[17] G. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[18] J. Z. Kolter, S. Batra, and A. Y. Ng, "Energy disaggregation via discriminative sparse coding." in *NIPS*, 2010, pp. 1153–1161.

[19] S. M. Ross, *Introduction to probability models*. Academic press, 2006.

[20] M. Wand and M. Jones, "Multivariate plug-in bandwidth selection," *Computational Statistics*, vol. 9, no. 2, pp. 97–116, 1994.