A Fully Unsupervised Non-intrusive Load Monitoring Framework

Ruoxi Jia, Yang Gao, Costas J. Spanos

Department of Electrical Engineering and Computer Science
University of California, Berkeley
Berkeley, CA 94720, USA
Email: {ruoxijia,yg,spanos}@berkeley.edu

***Abstract*—Monitoring an individual electrical load's energy usage is of great significance in energy-efficient buildings as it underlies the sophisticated load control and energy optimization strategies. Non-intrusive load monitoring (NILM) provides an economical tool to access per-load power consumption without deploying fine-grained, large-scale smart meters. However, existing NILM approaches require training data to be collected by sub-metering individual appliances as well as the prior knowledge about the number of appliances attached to the meter, which are expensive or unlikely to obtain in practice. In this paper, we propose a fully unsupervised NILM framework based on Nonparametric Factorial Hidden Markov Models, in which per-load power consumptions are disaggregated from the composite signal with minimum prerequisite. We develop an efficient inference algorithm to detect the number of appliances from data and disaggregate the power signal simultaneously. We also propose a criterion, Generalized State Prediction Accuracy, to properly evaluate the overall performance for methods targeting at both appliance number detection and load disaggregation. We evaluate our framework by comparing against other multi-tasking schemes, and the results show that our framework compares favorably to prior work in both disaggregation accuracy and computational overhead.**

I. Introduction

Buildings contribute to more than 40% of the total energy consumption in US [1]. The provision of per-load energy data in realtime can support "smart" buildings with improved energy-efficiency. For instance, an automated load scheduling policy can be designed to reduce a building's peak power demand by deferring one or more background loads with knowledge of energy usage and the duration of each background load. Building occupants can also learn the energy consumption of appliances and are empowered to take necessary actions to save energy. However, one major and continuing impediment of energy efficiency improvement is that deployment and maintenance of fine-grained power meters in a large scale are costly and cumbersome.

Rather than relying on expensive instrumentation, an alternative approach is non-intrusive load monitoring (NILM), also known as power disaggregation, which analyzes the aggregated power data from a single smart meter and aims to break it down to individual appliances' power usage. Generally speaking, the NILM process involves the following steps of Detection, Disaggregation and Designation (**3D** for short):

- Appliance **D**etection
- Load **D**isaggregation
- Multi-state **D**esignation

Appliance detection aims to find how many appliances (loads) are attached to the power meter of concern. Most research work has considered the number of loads as a prior knowledge and then analyzes the composite power signal using fixed-size models [2]. However, it is an eminently practical concern that the number of appliances in buildings can hardly be consistent over time. Ignorance of this fact would lead to improper choice of the models and thereby degrade disaggregation performance. Simply choosing a model sufficiently complex in order to accommodate an arbitrarily chosen large number of hypothetical loads will induce large computational overhead.

Load disaggregation consists of determining the ON/OFF state and power consumption of individual appliances given the composite signal. A common category of load disaggregation methods assumes that sub-metered ground truth data is available for training prior to performing disaggregation [3]–[6]. This assumption dramatically increases the investment required to set up such a system, since in reality deploying sub-meters may be inconvenient or time-consuming. The second category includes model-driven approaches that relax requirements for training data [7], [8]. Parson et al. [9] proposed to tune the generic appliance models to specific appliance instances using signatures extracted from the aggregate load. Tang et al. [10] formulated load disaggregation as an optimization problem and tried to minimize the variation of switching events based on the knowledge of appliance power models and the sparsity of the switching events. The third category of load disaggregation approaches obviates the need of the prior knowledge of appliances using unsupervised learning methods [11]–[13], which however often requires the number of appliances to be known.

Most of aforementioned work has made the assumption about a binary nature of appliance states. However, in practice appliances might have multiple states beyond OFF state, such as active mode, stand-by mode, etc. An appliance with multiple states may be detected as to two or more two-state appliances. The multi-state designation step is to merge states corresponding to the same appliance.

In this paper, we develop an fully unsupervised NILM framework without the need of sub-metered data or the prior knowledge about the number of appliances, and integrate **3D** process using the Bayesian network. The main contributions are summarized as follows:

• We establish a simultaneous appliance detection, load disaggregation and multi-state designation framework using nonparametric Bayesian models, which, to our knowledge, has not been used to solve the power disaggregation problem.

• We develop an efficient inference algorithm pertaining to the nonparametric Bayesian model, which takes advantage of slice sampling to adaptively truncate the size of model based on the number of appliances.

• We propose a metric that can standardize the performance evaluation procedure for power disaggregation framework without the prior knowledge about the number of appliances. We also evaluate our framework using the real-world data collected in public dataset, REDD [14]. The result shows the superiority of our framework to other appliance number detection and load disaggregation schemes based on traditional model selection.

II. Problem Formulation

In this section, we formalize the problem of power disaggregation. Suppose we are given an aggregated power signal $y(t)$ from K appliances with t from 1 to T. Here, we do not assume any prior knowledge of the number of appliances, i.e. K is unknown. Let $\mathbf{Z}$ be a $T \times K$ matrix and (t, k)-entry $z_{t,k}$ indicates the state for appliance k at time t. The goal is to determine state matrix $\mathbf{Z}$ from the aggregated power signal $\mathbf{y}$. We assume that each appliance has two states (ON/OFF) and its power consumption follows the Gaussian distribution when the appliance is ON. As can be seen from Fig. 1, the assumption is valid for most appliances, while for some appliances, e.g. laptop, the multimodality of the distribution is evident. This is because appliances might have multiple states with distinctive power consumptions when they are ON, such as active-mode and standby-mode. Even though they might have multiple states, they can be considered as comprising two or more binary-state appliances. We will first define $\mathbf{Z}$ as a binary matrix in which $0/1$ stand for OFF/ON respectively and further introduce a state designation algorithm to merge ON states corresponding to the same appliance.

III. Proposed Framework

From the description of energy disaggregation problem in the preceding section, it is evident that the number of appliances determines the dimension of state space to be estimated and thereby is directly related to the model complexity and computational overhead. Previous work determines the appropriate model by first learning a set of model candidates and select one using comparison metrics [15]. Rather than comparing models that vary in complexity, Bayesian nonparametric models are featured in fitting a single model that can adapt its complexity to the data by imposing a prior on model parameters. Inspired by this, in this section we will present our framework, a Nonparametric Factorial Hidden Markov Model (NFHMM), which can automate appliance detection process to load disaggregation. We will start by providing a brief background on Factorial Hidden Markov Model and Indian Buffet Process, based upon which we will describe our blind

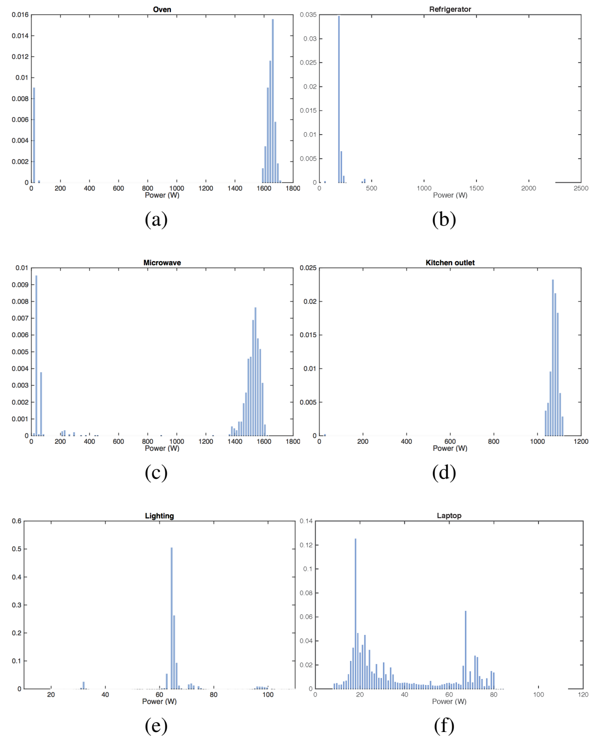


Fig. 1: Normalized histogram of (a) oven (b) microwave (c) kitchen outlet (d) lighting (e) laptop. The power distribution of some appliances can be well approximated by Gaussian distribution, while other appliances appear to be multi-modal.

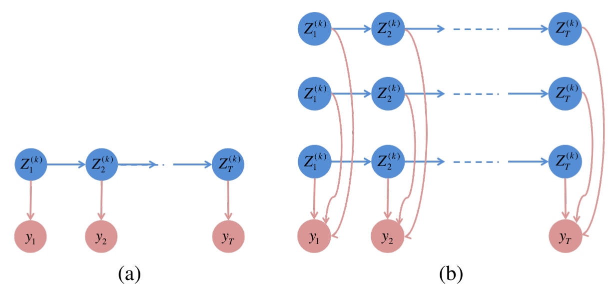


Fig. 2: (a) HMM (b) FHMM.

source disaggregation framework, NFHMM, for binary state appliance. Finally, we extend the binary state case to multiple states by introducing our multi-state designation algorithm.

A. Factorial Hidden Markov Model (FHMM)

FHMM is originally proposed in [16] as a distributed state representation of HMM (Fig. 2). Due to its distributed nature, FHMM is useful to model multiple loosely-coupled processes and has been considered as a method for power disaggregation [2]. In the FHMM, the state for each of K appliances evolves via a HMM independently and the states for all appliances at time t jointly determine the observation at time t.

B. Indian Buffet Process (IBP)

IBP has been studied by machine learning community as a prior in probabilistic models that represent a potentially infinite array of f

eatures. IBP is defined as a limit of the distribution over binary matrices with a finite number of rows and unbounded number of columns and is suitable to derive the probability of state matrix $\mathbf{Z}$ in our setting. In the original IBP, each entry of $\mathbf{Z}$ is drawn from a Bernoulli distribution parameterized by a Beta prior. Since the columns K are unbounded, the probability of $\mathbf{Z}$ is generally not well defined. Note that each appliance is exchangeable in the sense that we only care about the partitioning of the composite signal rather than the ordering of each appliance, therefore we can compute the probability of the equivalence classes of $\mathbf{Z}$ which correspond to the identical form via permutation of columns. In this way, an analytic formula for infinite-dimensional state matrix $\mathbf{Z}$ can be found and the derivation is elaborated in [17].

C. Nonparametric Factorial Hidden Markov Model (NFHMM)

The main idea of our power disaggregation model is to leverage FHMM to model aggregate power signal and apply the technique analogous to IBP to resolve the problem that there are potentially infinite number of appliances, namely infinite Markov chains, in FHMM. In essence, a model of an appropriate dynamic structure is implicitly fitted to aggregate power observations. The NFHMM is described as follows: Assume the state for each device evolves according to the transition matrix

$$\mathbf{W}^{(k)} = \begin{pmatrix} 1-\mu_k & \mu_k \\ 1-b_k & b_k \end{pmatrix}$$

where $W_{ij}^{(k)} = p(z_{t+1,k} = j | z_{t,k} = i)$. The state of kth appliance at time t, $z_{t,k}$, follows $Bernoulli(\mu_k^{1-z_{t-1,k}} b_k^{z_{t-1,k}})$. We place a prior distribution on $\mu_k \sim Beta(\alpha/K, 1)$, $b_k \sim Beta(\gamma, \delta)$. Let Θ be the power profile in which kth entry of the vector denotes the Gaussian distributed power of kth appliance. The observed mixture signal y is thereby generated by the emission model $y = \mathbf{Z}\Theta + \epsilon$, where ϵ is the measurement noise drawn from $\mathcal{N}(0, \sigma_\epsilon^2)$. Fig. 3 shows the diagram of our model.

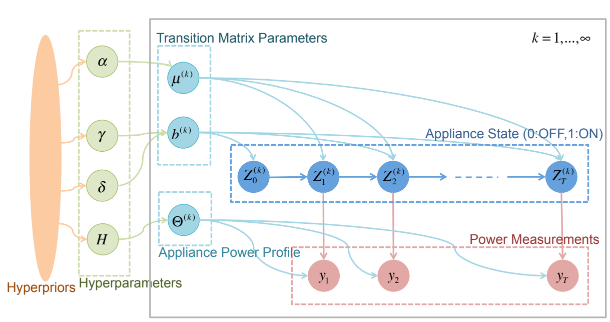


Fig. 3: Graphical model representation of NFHMM.

Since we are only interested in decomposing the mixed power signal, Markov chains associated with different appliances are exchangeable. Analogous to IBP, we can compute a meaningful limit of the probability of the equivalence class of $\mathbf{Z}$ as $K \to \infty$,

$$\lim_{K\to\infty} p([\mathbf{Z}]) = \frac{\alpha^{K^*} e^{-\alpha H_T}}{\prod_{h=0}^{2^T-1} K_h!} \cdot \prod_{k=1}^{K^*} \frac{(c_k^{01}-1)! c_k^{00}! \Gamma(\gamma+\delta)\Gamma(\delta+c_k^{10})\Gamma(\gamma+c_k^{11})}{(c_k^{00}+c_k^{01})!\Gamma(\gamma)\Gamma(\delta)\Gamma(\gamma+\delta+c_k^{10}+c_k^{11})} \quad (1)$$

where H_T is the Tth harmonic number, $H_T = \sum_{t=1}^T \frac{1}{t}$, $c_k^{ij}, i, j = 0, 1$ is the number of transitions from state i to state j in the column k and K^* gives the number of columns that contain at least one nonzero entry, which also indicates the effective dimension of the model. This representation is convenient for analysis while it is not practical for inference. We adopt an "stick breaking" representation [18] to NFHMM, which provides a distribution law for the order statistics of the appliance state transition parameter at time k, i.e., μ_k. Let $\mu_{(1)} > \mu_{(2)} > \cdots > \mu_{(K)} > \cdots$ be a decrease ordering of $\mu_{1:\infty} = \{\mu_1, \cdots, \mu_k, \cdots\}$, then $\mu_{(k)}$ obeys the following law,

$$p(\mu_{(1)}) = \alpha \mu_{(1)}^{\alpha-1} \mathbb{1}(0 \le \mu_{(1)} \le 1) \quad (2)$$

$$p(\mu_{(k+1)} | \mu_{(1:k)}) = \alpha \mu_{(k)}^{-\alpha} \mu_{(k+1)}^{\alpha-1} \mathbb{1}(0 \le \mu_{(k+1)} \le \mu_{(k)}) \quad (3)$$

This allows us to use a combination of slice sampling and Gibbs sampling to do inference.

Proposition 1: Let s be an auxiliary variable drawn from

$$s | \mathbf{Z}, \mu_{(1:\infty)} \sim Uniform[0, \mu^*] \quad (4)$$

where μ^* is chosen to be the length of the stick for the last active appliance, $\mu^* = \min\{1, \min_{k:\exists t, z_{t,k}=1} \mu_{(k)}\}$. By imposing appropriate conjugate prior to hyperparameters σ_ϵ^2, α, γ, δ, the approximate inference of NFHMM proceeds by updating all variables in turn conditioning on the rest of variables:

$$\bullet p(s|rest) = \frac{1}{\mu^*} \mathbb{1}\{0 \le s \le \mu^*\} \quad (5)$$

•$p(z_{:,k}|rest) \propto p(y|\mathbf{Z}, \Theta) p(s|\mathbf{Z}, \mu) p(z_{:,k}|\mu, b)$, where $z_{:,k}$ is the kth column of Z (6)

$$\bullet p(\theta_k|rest) \sim \mathcal{N}(\sigma_{\theta;p}^2 (\frac{\mu_\theta}{\sigma_\theta^2} + \frac{\sum_t z_{t,k}(y_t - \sum_{i\neq k} z_{t,i}\theta_i)}{\sigma_\epsilon^2}), \sigma_{\theta;p}^2),$$

$$\text{where } \sigma_{\theta;p}^2 = (\frac{1}{\sigma_\theta^2} + \sum_t z_{t,k} \frac{1}{\sigma_\epsilon^2})^{-1} \quad (7)$$

$$\bullet p(\mu_{(k)}|rest) \propto (1-\mu_{(k)})^{c_{(k)}^{00}} \mu_{(k)}^{c_{(k)}^{01}-1} \mathbb{1}(\mu_{(k+1)} \le \mu_{(k)} \le \mu_{(k-1)}) \quad (8)$$

$$\bullet p(b_{(k)}|rest) \propto (1-b_{(k)})^{c_{(k)}^{10}+\delta-1} b_{(k)}^{c_{(k)}^{11}+\gamma-1} \quad (9)$$

Remark The inference is initialized by sampling the auxiliary slice variable s, which determines the actual complexity of the model. Since it is possible to increase the size of model according to the observations, a representation padding procedure is involved before sampling the next variable. Following the derivation in [18], $\mu_{(k)}$'s to be padded can be iteratively drawn from the distribution $p(\mu_{(k)}|\mu_{(k-1)}, z_{:,\ge k} = 0) \propto \exp(\alpha \sum_{i=1}^N \frac{1}{i}(1-\mu_{(k)})^i)(1-\mu_{(k)})^T \mu_{(k)}^{\alpha-1} \mathbb{1}(0 \le \mu_{(k)} \le \mu_{(k-1)})$ via adaptive rejection sampling, and the new columns of Z are padded to be all zeros. Other parameters can be expanded via sampling the corresponding prior distribution. Sampling of the second quantity is done by a combination of block Gibbs sampler and a standard Forward Filtering Backward Sampling procedure. The posterior distributions of the remaining variables θ_k, $\mu_{(k)}$ and $b_{(k)}$ are proven to be simple Gaussian or Beta distribution. Moreover, by imposing appropriate conjugate prior to hyperparameters, the posterior distributions of hyperparameters are all of simple analytic forms. Algorithm 1 summarizes the inference algorithm of NFHMM.

Algorithm 1 NFHMM Inference Algorithm

Input: y: Aggregate power signal
Initialization :
Z, μ, b, θ
$IterNum \leftarrow$ *Maximum iteration number*
Iterative Sampling :
1: **while** $IterNum > 0$ **do**
2: Sample slice variable s (Eqn. (5))
3: Expand representation of Z, μ, b, θ (Remark)
4: Sample Z with a blocked Gibbs sampler and run FFBS on each column of Z (Eqn. (6))
5: Sample θ, μ, b (Eqn. (7), (8), (9))
6: Sample hyperparameters $\alpha, \gamma, \delta, \mu_\theta, \sigma_\theta^2, \sigma_\epsilon^2$ (Remark)
$IterNum \leftarrow IterNum - 1$
7: **end while**
Output: Z, θ

D. Resolving Multiple States of an Appliance

In the discussion above, we have assumed that each appliance has binary states, ON/OFF, while in practice multiple states with different power consumption levels might be associated with different usage conditions of one single appliance. Fig. 4 shows real power traces of a laptop and a refrigerator, which is a superposition of spikes and a relative constant base power level. The baseline power presents the power consumption level in normal cases while spikes can be an indicator of overuse, charging process, etc. The key observation that supports our multi-state designation algorithm is that the "abnormal" mode of appliances always appears to be temporary and concurs with the baseline power. We can combine the "abnormal" mode with normal one by noticing the time slot of transient abnormalities is a subset of normal usage time. Let's formalize this in the following proposition:

Proposition 2: Let S_k to be the set of time slots within which the appliance k is ON, i.e.

$$S_k = \{[t_s, t_e] : t_s \le t \le t_e, z_{kt} = 1, z_{kt_s} = 0, z_{kt_e} = 0\} \tag{10}$$

Denote the cardinality of the set S_k to be $|S_k|$. Assume the composite power signal is aggregated from some small number of appliances, then for some appliance i and j, if the criterion

$$\frac{|S_i \cap S_j|}{|S_i|} > 1 - \tau \tag{11}$$

is satisfied, then we claim the two columns of $\mathbf{Z}$, $z_{i:}$ and $z_{j:}$, are the states corresponding to the identical appliance. τ is some small positive integer indicating the fuzziness of matching.

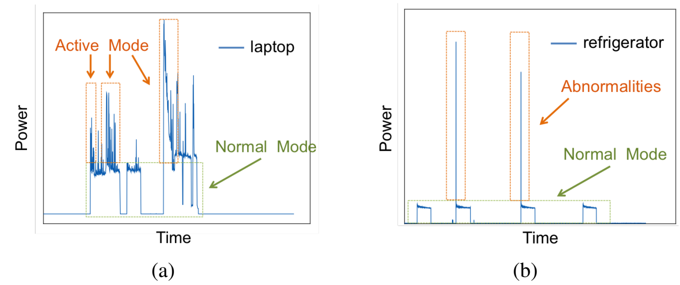


Fig. 4: Real power traces of (a) the laptop and (b) the refrigerator. Single appliance can generate multiple states with different power levels.

IV. Experimental Evaluation

Our proposed framework has been evaluated using Reference Energy Disaggregation Dataset (REDD) (http://redd.csail.mit.edu/) described in [14]. This dataset was chosen as it is benchmark dataset specifically collected for testing NILM algorithms. The REDD comprises six houses, for which both aggregate and circuit-level power consumption data were monitored. [14] and [11] provide two benchmarked disaggregation results using supervised and unsupervised method respectively based on FHMM. Our framework differs from previous methods since we target not only at disaggregating the composite power signal without labeled data but also deducing the number of appliances from observed signals, therefore a direct performance comparison is not possible. However, we will benchmark our method against the combination of unsupervised disaggregation and traditional statistical model selection procedure. To be specific, we come up with several FHMMs with fixed sizes and experiment these models with the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), which are commonly adopted in model selection [19].

A. Evaluation Metrics

State prediction accuracy is a commonly used evaluation metric if the number of appliances is provided as prior knowledge. However, nonparametric disaggregation methods aim at integrating the appliance detection and load disaggregation process, and thereby other metrics are required in order to evaluate the capability of predicting correct number of appliances. We propose a *Generalized State Prediction Accuracy* (GSPA) that standardizes the evaluation process for nonparametric disaggregation methods.

This metric is inspired by the maximum weight bipartite matching problem in graph theory. Nonparametric disaggregation methods are likely to produce different estimates of the appliance number. Therefore, an appropriate evaluation metric would take both the number of appliances and the ON/OFF state of each appliance into account when comparing the ground truth with some algorithmic prediction. Calculation of GSPA involves two steps: Firstly, find the correspondence between predicted appliances and groundtruth. Then we calculate a numeric score based on the matching quality and appliance relative importance which is indicated by the ON time duration.

For the first step, in order to obtain the correspondence between ground truth appliances and predicted appliances, we calculate a pairwise accuracy score between the predicted appliance i and ground truth appliance j, defined as

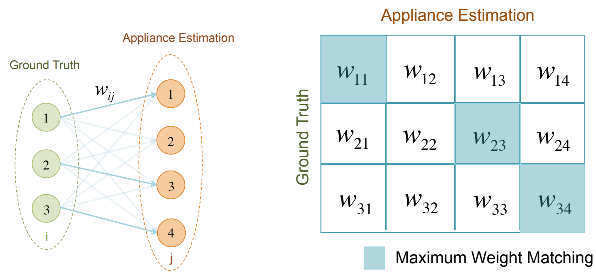
$$w_{ij} = \frac{|\hat{\Psi}_i \cap \Psi_j|}{|\hat{\Psi}_i \cup \Psi_j|} \tag{12}$$

where $\hat{\Psi}_i$ is the time slots when appliance i is estimated to be ON and Ψ_j is the period of time when appliance j in ground truth is ON. If i and j correspond to the identical appliance and the state estimation is perfect, then $w_{ij} = 1$. In this way, the correspondence finding problem is converted to the maximum weight matching problem if considering estimated appliance set and ground truth appliance set as a bipartite (Fig. 5). We use Hungarian algorithm [20] to calculate the best matching for weight matrix w_{ij}. This will be our best guess for the correspondence between predicted appliances and the ground truth appliances, and the accuracy of the best matching correspondence is considered to be the state estimation accuracy. Since we would like to produce a metric that also takes into consideration the correctness of appliance number prediction and relative importance of appliances, we will define GSPA score in the following way: Assume that we have matched predicted appliance i_k to ground truth appliance j_k for $k = 1 \cdots min(K_{pre}, K_{gt})$, where K_{pre} and K_{gt} are the predicted and actual number of appliances respectively. The GSPA score is defined to be the appliance-importance weighted sum of accuracy associated with a penalty of predicting the wrong number of appliances.

$$GSPA = \frac{\sum_{k=1}^{min(K_{pre},K_{gt})} \pi_{j_k} w_{i_k,j_k}}{\sum_{k=1}^{min(K_{pre},K_{gt})} \pi_{j_k} + penalty} \tag{13}$$

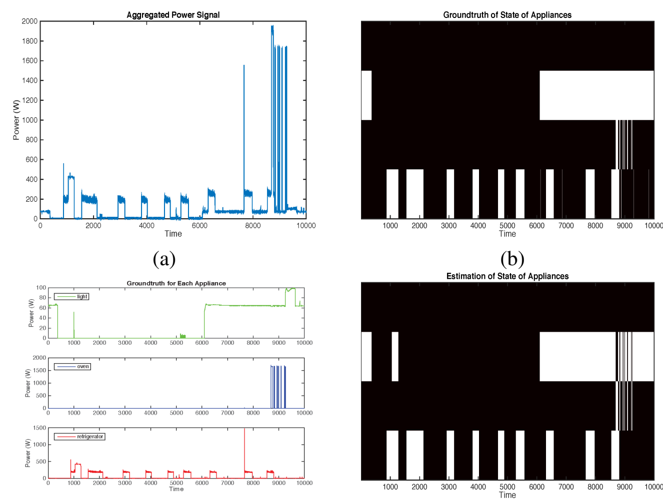
where π_{j_k} is the ON time ratio of the ground truth appliance j_k. The penalty term is the sum of the ON time ratio of the appliances which do not have correspondence. We refer to these appliances as ghost appliances, which may be created due to random noise or abnormalities of appliance operations. The GSPA will penalize prediction error in the frequently-used appliances more than in the seldom-used ones, and will also penalize missing real appliances and predicting ghost appliances.

In addition to GSPA, we also use running time as a metric to measure the computational overhead.



(a) (b)

Fig. 5: GSPA metric: (a) Compute accuracy for all possible pairs between ground truth and appliance estimation (The ground truth contains 3 appliances while 4 appliances are estimated by nonparamteric disaggregation methods in the example); (b) Calculate the maximum weight matching.



(a) (b)

(c) (d)

Fig. 6: NFHMM can detect the state of each appliance without prior knowledge of the number of appliances: (a) Aggregated power signal we try to decompose; (b) Ground truth of the state of each appliance: rows are appliances, columns are white if the appliance is ON and black otherwise; (c) Ground truth of power traces of appliances, summing up which we obtain the mixed signal in (a); (d) Estimate of the appliance state.

B. Framework Validation

We first demonstrate the validity of our framework by decomposing the aggregated signal from three appliances: oven, refrigerator and light in a randomly picked time interval. The mixed and original power consumption for each appliance are shown in Fig. 6 (a) and 6 (c), respectively. It can be seen that different electrical loads exhibit distinctive power levels. For these three appliances, the power signals tend to fluctuate around a fixed value during ON state, in which case our Gaussian assumption for the appliance power is held and thereby we expect a good separation for these devices. Fig. 6 (d) shows the state estimation of each appliance given by our algorithm and Fig. 6 (b) gives corresponding ground truth. Here, we use heatmaps to demonstrate appliance's state at a certain timestamp, where rows are appliances, columns are white if the appliance is ON and black otherwise. Visual inspection of two heatmaps shows that our algorithm achieves almost a perfect detection of appliances' states with the accuracy of 98.5%, and it can also give a correct estimation of the number of appliances simultaneously. Fig. 7 illustrates the convergence performance by showing how disaggregation accuracy varies as iteration numbers increase. As can be seen, our algorithm exhibits different convergence rate for different runs and there always exists a steep increase of accuracy instead of growing steadily. It is worth noting that the accuracy keeps above 75% even with only one step iteration.

Since our framework disaggregates power signal using only low frequency real power feature and assumes binary states for appliances, the single appliance with multiple power usage

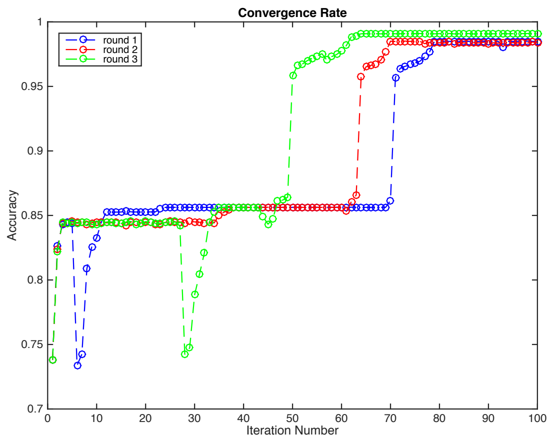


Fig. 7: Convergence of NFHMM: The accuracy grows as iteration steps increase. Even with only one iteration, the accuracy is above 75 %

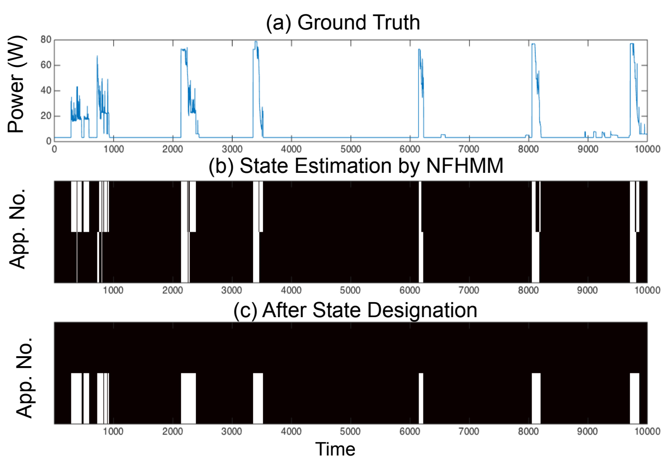


Fig. 8: Demonstration of the state designation algorithm: (a) The ground truth of the power consumption signals of a laptop in which multiple power usage modes are evident; (b) The state estimation result given by NFHMM inference where an extra appliance is produced for the high power usage mode; (c) The final state estimation result after applying the state designation algorithm where the extra appliance is eliminated.

modes is possible to be recognized as multiple appliances. Fig. 8 shows a snapshot of the ambiguity. As aforementioned, the laptop can have distinctive power consumption levels depending on the different usage conditions. Our algorithm tends to add appliances to explain the higher power usage, shown in Fig. 8(b). Nevertheless, following up the cue that the "abnormalities" always appear to be transient spikes superposed upon the baseline power, we apply our state designation algorithm to the preliminary result given by NFHMM and the "ghost" device that represents the active mode of the laptop can be merged (Fig. 8(c)).

C. Comparative Study

We scale up our experiments by aggregating all regularly-used appliances in one house and make a comparison of results

TABLE I: Performance comparison between NFHMM and FHMMs selected by AIC and BIC.

	GSPA score	Time (s)	Appliance Number
NFHMM	0.2510	12	7
FHMM+AIC	0.2350	304	4
FHMM+BIC	0.1223	304	13

produced by our framework against FHMM models selected by AIC [21] and BIC [22].

The composite signal is aggregated from oven, refrigerator, microwave, kitchen outlet and lighting in the REDD dataset. This set of appliances is representative in the sense that all common household appliances are included. The appliance set investigated here presents various complex patterns. The refrigerator has a low base power level combined with high spikes. Oven is used with irregular patterns depending on the habit of the household. Lightings can be switched on and off quit often, and its working power consumption is small which makes it hard to detect.

The methods to be compared against is to fit several FHMMs to the mixed signal and select the best model with criteria of AIC or BIC. AIC and BIC are defined as follows,

$$AIC = 2K_{param} - 2ln(L) \tag{14}$$
$$BIC = log(n)K_{param} - 2ln(L) \tag{15}$$

where K_{param} is the number of parameters, n is the number of training instances and L is the likelihood of the model. We select the best model of FHMM, i.e. the number of potential appliances K, with AIC and BIC respectively and then use the GSPA to assess the selected model. Because the NFHMM is possible to produce results of infinite appliances, we have to set an upper bound of K to enable the model selection to be finished in finite time. This bound is a free parameter that can be chosen according to the potential complexity of aggregated power signals. In our experiment, the number of components that could be generated by NFHMM is bounded above by 20, and thereby we also set the searching space of AIC and BIC upper bounded by 20.

Table I shows our comparison result. We can see that NFHMM achieves a higher GSPA score than that of FHMMs selected by both AIC and BIC. Due to the nature of power disaggregation problem, the number of observations K_{param} in BIC formula is 1. BIC fails to perform model selection and thereby gets a low GSPA score. AIC is better than BIC approach and obtains a similar score as NFHMM. However both AIC and BIC are more computationally expensive. Overall, NFHMM can achieve the best prediction performance as well as computational efficiency. Fig. 9 shows the performance of different schemes regarding the accuracy of assigning correct energy to corresponding appliances. As can be seen, the energy assignments given by our NFHMM is closest to the ground truth compared with other schemes.

V. Conclusion

In this paper, we attacked the challenge of power disaggregation without labeling of power data or prior knowledge about the number of appliances. This is of great practical

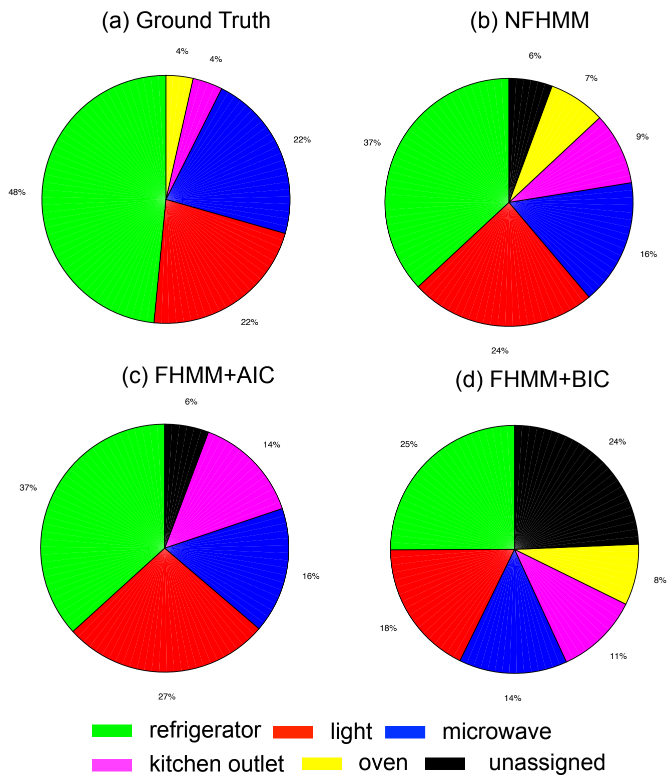


Fig. 9: Pie chart showing predicted and actual energy assignments of the aggregated power signal. NFHMM outperforms fixed-size models selected using AIC or BIC.

value as an effective method of this type would facilitate building energy conservation efforts. For this purpose, we proposed a framework based on nonparametric factorial hidden Markov model that simultaneously detects the number of appliances, decomposes the aggregated power signal and merges the multiple states corresponding to the identical appliance. Using low frequency power measurements from real world, we have showed that our framework outperforms the other "blind" disaggregation framework and is very computationally efficient.

For future work, we plan to extend this work in several ways. Firstly, since only low-frequency real power measurements are utilized as features in our framework, this will limit the scalability of our method. We need to design more features to enhance our method to deal with more appliances. Secondly, we intend to apply our disaggregation results to lighting control systems in buildings to promote energy savings.

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] H. Kim, M. Marwah, M. F. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements." in *SDM*, vol. 11. SIAM, 2011, pp. 747–758.

[3] M. Weiss, A. Helfenstein, F. Mattern, and T. Staake, "Leveraging smart meter data to recognize home appliances," in *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*. IEEE, 2012, pp. 190–197.

[4] S. Gupta, M. S. Reynolds, and S. N. Patel, "Electrisense: single-point sensing using emi for electrical event detection and classification in the home," in *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, 2010, pp. 139–148.

[5] M. Berges, E. Goldman, H. S. Matthews, L. Soibelman, and K. Anderson, "User-centered nonintrusive electricity load monitoring for residential buildings," *Journal of Computing in Civil Engineering*, vol. 25, no. 6, pp. 471–480, 2011.

[6] 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.

[7] S. Barker, S. Kalra, D. Irwin, and P. Shenoy, "Powerplay: creating virtual power meters through online load tracking," in *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*. ACM, 2014, pp. 60–69.

[8] R. Dong, L. J. Ratliff, H. Ohlsson, and S. S. Sastry, "Energy disaggregation via adaptive filtering," *arXiv preprint arXiv:1307.4132*, 2013.

[9] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," 2012.

[10] G. Tang, K. Wu, J. Lei, and J. Tang, "A simple model-driven approach to energy disaggregation," in *Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on*. IEEE, 2014, pp. 566–571.

[11] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial hmms with application to energy disaggregation," in *International conference on artificial intelligence and statistics*, 2012, pp. 1472–1482.

[12] S. Pattem, "Unsupervised disaggregation for non-intrusive load monitoring," in *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, vol. 2. IEEE, 2012, pp. 515–520.

[13] D. Egarter, V. P. Bhuvana, and W. Elmenreich, "Paldi: Online load disaggregation via particle filtering," 2014.

[14] J. Z. Kolter and M. J. Johnson, "Redd: A public data set for energy disaggregation research," in *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*, vol. 25. Citeseer, 2011, pp. 59–62.

[15] G. Claeskens and N. L. Hjort, *Model selection and model averaging*. Cambridge University Press Cambridge, 2008, vol. 330.

[16] Z. Ghahramani and M. I. Jordan, "Factorial hidden markov models," *Machine learning*, vol. 29, no. 2-3, pp. 245–273, 1997.

[17] T. Griffiths and Z. Ghahramani, "Infinite latent feature models and the indian buffet process," 2005.

[18] Y. W. Teh, D. Görür, and Z. Ghahramani, "Stick-breaking construction for the indian buffet process," in *International Conference on Artificial Intelligence and Statistics*, 2007, pp. 556–563.

[19] K. P. Burnham and D. R. Anderson, "Multimodel inference understanding aic and bic in model selection," *Sociological methods & research*, vol. 33, no. 2, pp. 261–304, 2004.

[20] R. Jonker and T. Volgenant, "Improving the hungarian assignment algorithm," *Operations Research Letters*, vol. 5, no. 4, pp. 171–175, 1986.

[21] H. Akaike, "A new look at the statistical model identification," *Automatic Control, IEEE Transactions on*, vol. 19, no. 6, pp. 716–723, 1974.

[22] G. Schwarz *et al.*, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.