Disaggregation of Home Energy Display Data Using Probabilistic Approach

Michael Zeifman, *Senior Member*, IEEE

**Abstract — ***Home energy displays are emerging home energy management devices. Their energy saving potential is limited, because most display whole-home electricity consumption data. We propose a new approach to disaggregation electricity consumption by individual appliances and/or end uses that would enhance the effectiveness of home energy displays. The proposed method decomposes a system of appliance models into tuplets of appliances overlapping in power draw. Each tuplet is disaggregated using a modified Viterbi algorithm. In this way, the complexity of the disaggregation algorithm is linearly proportional to the number of appliances. The superior accuracy of the method is illustrated by a simulation example and by actual household data*[1].

Index Terms — disaggregation algorithm, power draw, appliances, signal features, energy management.

I. INTRODUCTION

Recent estimates [1] suggest that, by 2015, approximately five percent of U.S. households will have a Home Energy Management (HEM) device or system. This projected market penetration, though noticeable, is still small as compared to the penetration of some other novel consumer electronics. Installation complexity and high cost of automation-based HEM systems are examples of challenges related to the HEM mass-market penetration [2].

Home energy displays (HEDs) are inexpensive HEM devices that do not require professional installation. Usually, HEDs provide whole-house, near real-time electricity consumption information [3] with a primary goal to help homeowners save energy. In this way, HEDs are similar to the smart meters [4]. However, the information provided by either HEDs or smart meters is not actionable as the current HED energy management/saving paradigm is that occupants need to learn the energy consumption of the household devices and appliances by switching them on and off [3]. This learning process is tedious, inconvenient and may provide incorrect results; consequently, it is a significant barrier to achieving widespread deployment of disaggregated energy consumption feedback. A better alternative could be to obtain appliance-specific information by automatic disaggregation of electricity consumption, i.e., to perform nonintrusive appliance load monitoring (NIALM) [5].

The available NIALM methods, though based on different techniques, have a common principle. First, specific appliance features, or signatures, need to be selected and mathematically characterized. Then, a mathematical algorithm detects the features in the overall signal and infers the appliance presence and power draw [6]. Numerous new NIALM methods have been proposed recently in both academic and industrial research [6], [7], however, it appears that these methods are not yet applicable to the problem in question. For a NIALM method to be applicable to HEDs and/or smart meters, several requirements need to be fulfilled:

- Feature selection. The selected features must be compatible with HED data acquisition systems that are usually limited to real power sampled at 1 Hz.
- Accuracy. The algorithm accuracy shall be acceptable by the homeowner. There are indications [8] that a minimum acceptable accuracy is 80-90%.
- No training. The method may not involve significant occupant efforts for algorithm training as the need for learning is a major drawback of HEDs. The algorithm also needs to recognize new appliances and to discard appliances that are no longer in use.
- Near real-time capabilities. The algorithm needs to be computationally-efficient and robust.
- Scalability. The performance of the method shall not deteriorate drastically as the number of household appliances reasonably increases (e.g., from 10 to 20).
- Various appliance types. The method shall work with four categories of appliances [5] (i) on-off appliances (most household appliances, such as a toaster, light bulb, or water pump), (ii) finite-state appliances (a fan with several rotation speeds), (iii) variable-power devices (dimmer lights), and (iv) permanent consumer devices (a wired smoke alarm).

In this work, we propose an algorithm capable of meeting the first five requirements. Currently, the algorithm only works with on/off appliances, but potentially it can also tackle the other three appliance types. It uses stepwise power changes, power surges, time-on and time-off durations as the features, and it utilizes historical data for initial disaggregation. Whereas the selection of these features is not new, the method specifically addresses the problem of overlap in the power draw between different appliances. The uniqueness of the algorithm is that it uses approximate semi-Markov models that retain robust computationally-efficient solution.

The paper is organized as follows. In Section II, we provide a background on NIALM methods potentially suitable for HEDs. It appears that the main reason for the low accuracy of conventional NIALM algorithms is an overlap in the power draw among different appliances. Section III presents our NIALM approach. The approach includes two major

[1] M. Zeifman is with the Fraunhofer Center for Sustainable Energy Systems, Cambridge, MA 02141 USA (e-mail: mzeifman@fraunhofer.org).

Manuscript received 01/13/12
Current version published 03/21/12
Electronic version published 03/21/12.

components, clustering of historical data for initial learning and a semi-Markov model for disaggregation of the overlapping appliances. This paper mainly concentrates on the latter component. Unlike other Markov-based algorithms that process the entire time series of data, our model only uses changes of power that fall within a predetermined range. The time-on/time-off statistics are used to calculate the transition probabilities, and a modified Viterbi algorithm is implemented to optimally reconstruct the appliance states.

A simulation example is discussed in Section IV. We show that the novel way of use of the time-on/time-off statistics significantly improves the disaggregation accuracy. In Section V, we apply our algorithm to real data collected in a household. We conclude that the algorithm meets the requirements for a NIALM method in Section VI.

II. Background: NIALM Methods suitable to HEDs

We have recently reviewed NIALM methods and refer the interested reader to this work [6] and the references quoted therein for details. We will briefly discuss here the general principles involved in the algorithms suitable to the HED data and review additional most recent publications in this field.

The first NIALM method was developed at MIT in 1980s. This algorithm has been successfully commercialized [7]. This basic NIALM algorithm [5], [9] includes five steps. First, an edge detector identifies changes in steady-state power-draw levels (see Fig. 1). Second, a cluster analysis algorithm locates these changes in a two-dimensional signature space of real and reactive power (see Fig. 2). Third, positive and negative clusters of similar magnitude are paired or matched. In the fourth step, known as anomaly resolution, unmatched clusters and events are paired or associated with existing or new clusters according to a best likelihood algorithm. In the final step, pairs of clusters are associated with known power draw levels for different loads to determine their operating schedule. The reported accuracy is about 70-80% [7].

This description suggests that the method may not detect multi-state and variable-load appliances. This issue was partially addressed in Refs. [10], [11]. However, the most severe deficiency of MIT method is that the appliances consuming similar power may not be separated. Our Fig. 2, originally shown in Ref. [5] to illustrate typical appliance clusters, can also be used to illustrate this deficiency. It is seen in the Figure that, if only real power measurements are available, the dehumidifier cluster will totally overlap with the water pump cluster. The clusters located in the 150-400 W range will also overlap. These heavy overlaps badly affect the accuracy of any algorithm that uses changes of steady-state power draw as the only appliance feature.

Apparently, the need to overcome the overlap problem has been the motivation for the following research in NIALM, even though this problem is often not acknowledged specifically. Whereas the main current direction in NIALM research involves the use of high-frequency features (e.g., harmonics or raw waveforms [6]), several works were devoted to the HED-s type of data. For example, Albicki and Cole [12], [13] used "edges" and "slopes" as appliance features, in addition to the steady-state power draw. These two new features were defined, respectively, as the initial upward spike in power and the slower changing variation that occurs during the turn-on events of appliances with motors (such as refrigerators and heat pumps). Ref. [14] suggested using appliance-specific decision rules to detect such large appliances as water heaters and refrigerators. Although some of the proposed decision-rules were non-intuitive and complicated [6], this work implicitly used appliance time-on statistics via a scoring system. This work was further extended in Ref. [15], which made an explicit use of time-on statistics (average duration time). Similarly to the majority of other NIALM techniques, this method requires excessive training, i.e., the electric load on each appliance of interest must be continuously monitored for about a week.

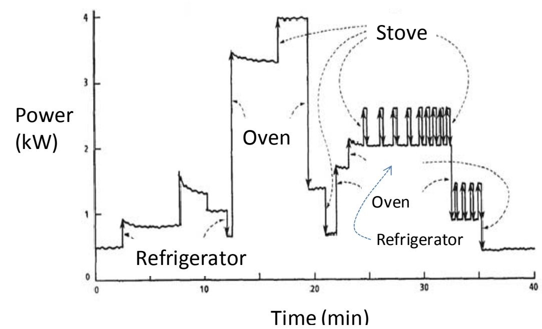

Fig. 1. Power vs. time plot shows step changes due to individual appliance events. From [5]. © 1992 IEEE

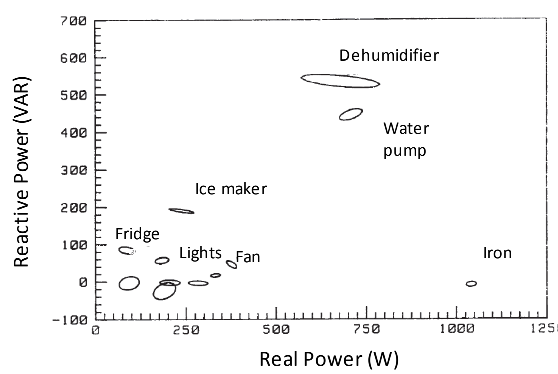

Fig. 2. Complex power space and appliance clusters. From [5]. © 1992 IEEE

Unlike the above-mentioned methods, the method of Baranski and Voss [16], [17], [18] does not require training. Instead, it creates a frequency analysis (histogram) based on the historical data, and only frequent power changes are considered further. Whereas the MIT method tries to find and match turn-on and turn-off events one at a time, Baranski's method uses an optimization algorithm to optimally match a large set of detected on/off events to appliance presences in time. The use of the simultaneous matching apparently improves the detection and monitoring accuracy of this method. The reported accuracy of Baranski's method (80-

85%) is comparable to the accuracy of the MIT method, even though the latter is based on both real and reactive power features. However, the genetic algorithm proposed as the optimization algorithm is a simulation method that may not provide the optimal solution. It is also unclear how the overlap problem can be addressed with Baranski's method.

A more recent work that incorporates both appliance electrical features and time-on/off duration is Ref. [19]. This work integrates the appliance features and the expected durations, conditional on the time of day via a special membership network. The network, originated from a Bayesian Belief Network, calculates scores for various possible scenarios and selects the scenario with the highest score. With too few technical details available in Ref. [19], it is difficult to assess how exactly the time duration features are incorporated. Also, similarly to References [14] and [15], this method [19] uses excessive training to characterize appliance-specific features.

The references considered so far do not use a probabilistic framework to infer the appliance state and power draw. Such a probabilistic framework is proposed in Ref. [20], in which the on/off household appliances are modeled by a factorial hidden Markov model (FHMM). FHMM can be considered as a mixture of independent hidden Markov models that are coupled by the observations [21]. Since most of household appliances change their states independently, FHMM can be a suitable probabilistic model for NIALM. Kim et al. [20] have extended FHMM by modeling coupling between dependent appliances and by incorporating a non-geometric (i.e., non-exponential in case of continuous time) parametric distribution (Gamma distribution) of discrete time-on duration. The reported accuracy is up to 78% for households with eight "active" appliances [20]. The method [20] reportedly does not require training, as historical data are used for estimation of model parameters. However, the number of appliances must be known *a priori*.

Ref. [20] provides significant research insights, among which is a demonstration on how the incorporation of time-on statistics alleviates the overlap problem. Nonetheless, the proposed disaggregation approach is of limited practical use. The most severe limitations are computational complexity and heavy reliance on the iterative optimization methods for both estimation and disaggregation. These methods are prone to errors and to low convergence rates [22], [23], [24]. Kim et al. [20] have acknowledged the overfitting problem and the exponential dependence of algorithm complexity on the number of appliances. It is not surprising that the disaggregation accuracy quickly decreases below an acceptable level with the number of appliances in the household [20].

III. PROPOSED METHOD

The first version of our algorithm is explained in detail elsewhere [25]. Here, we provide an overview of an extended version of this algorithm. The key idea is to decompose a system of appliance models such that only appliances overlapping in power draw with a given appliance are considered at a time. In this way, the complexity of the disaggregation algorithm is linearly proportional to the number of appliances whereas the complexity of other probabilistic methods [20] exponentially grows with the number of appliances.

A. *Tuplets of Overlapping Appliances*

Suppose there are N on-off appliances ordered by their power draw. Consider appliance i, $1 < i < N$. The measured power draw of appliance i can overlap with the measured power draws of the "adjacent" appliances $(i - 1)$ and $(i + 1)$. It may also overlap with the measured power draws of more "distant" appliances, such as $(i \pm 2)$, $(i \pm 3)$, .., $(i \pm k)$, but the overlap degree becomes smaller with k. In the simplest case [25], the overlap with the distant appliances is negligible and only appliances $(i \pm 1)$ overlap with appliance i.

Whenever a change of power within the range of power draw of appliance i is observed, it can indicate a change of state of an appliance within a system comprising the three appliances, $\{i, (i - 1) \text{ and } (i + 1)\}$. Whereas these appliances can be mutually independent, the probability of a state transition for a given appliance can depend on the states of the other two appliances. For example, assume that a typical time-off duration of both appliances $(i - 1)$ and $(i + 1)$ is 1 hour and that of appliance i is 10 hours. The probability that an observed positive power change corresponds to appliance i being turned on would depend on whether the other two appliances are currently on or off, and if any of them is off, then for how long.

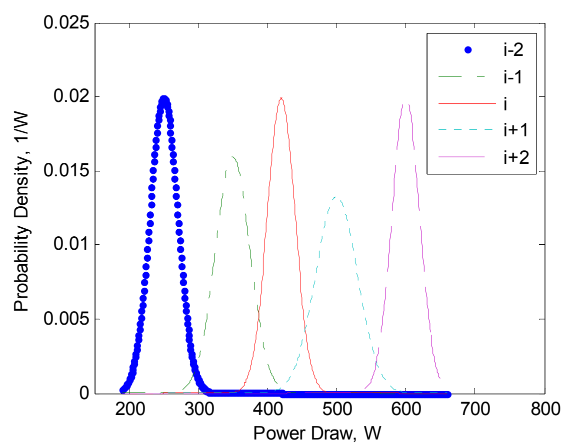

Fig. 3. Possible PDFs of power draw of appliances *i* and its neighbors.

The system of the three appliances can be modeled as a Markov Chain, and the system states can be optimally decoded using the Viterbi algorithm, a dynamic programming tool [26]. However, this system could be "non-pristine" in the sense that fractions of the power draw populations corresponding to appliances $(i - 1)$ and $(i + 1)$ could be missing and "foreign" data from other appliances could be present in the system. Fig. 3 illustrates this problem. It shows possible Probability Density Functions (PDFs) of power draws of appliance i and its neighbors. The composition of the system $\{i, (i-1), (i+1)\}$ depends on the range of power draw that we impose to define this system. For example, if we select a range [350 500] W, approximately half of $(i \pm 1)$ populations will be missing, whereas the selected range of [300 600] will bring significant fractions of the $(i \pm 2)$ populations. Moreover, the non-perfect processes of data acquisition and of detection of power changes will inevitably result in additional missing and foreign data [5] (see also Sections III.E and V.C).

The non-pristinity of the system suggests that the Markov chain needs to account for the missing and foreign data. It is difficult to characterize the time-on/off properties of the missing and foreign data. Since the "pristine" part of the system is characterized by both power draw and time-on/off features, the resultant Markov chain that includes both pristine and non-pristine components will be approximated. Also, there will be 64 transitions within the triplet system $\{i, (i-1)$ and $(i+1)\}$, which may slow down the computations.

In this work, we implement an even less "pristine" approach, in which we consider appliance pairs. To this end, we further decompose the triple $\{(i-1), i, (i+1)\}$ into two pairs, $\{(i-1), i\}$ and $\{i, (i+1)\}$. Within each pair, there are only 16 transitions. The motivation for working with pairs is threefold. First, the main purpose of this work is to demonstrate the concept of system decomposition into small groups or tuplets of overlapping appliances. The demonstration is simplest when the tuplet size is two, at the cost of highest fractions of missing/foreign data. The transition from pairs to triplets is straightforward. Second, our experience with actual household data obtained by a HED suggests that oftentimes, only two neighboring appliances have a significant overlap. Third, the obtained disaggregation results are reasonably good even with the appliance pairs (see Sections IV and V.C).

B. *System Transitions*

The states of the system comprising appliances i and $(i+1)$ are listed in Table I.

TABLE I
STATES OF TWO-APPLIANCE SYSTEM

State	Appliance i	Appliance $(i+1)$
1	Off	Off
2	On	Off
3	Off	On
4	On	On

Table II presents possible events underlying transitions that correspond to an observed negative power change (see Fig. 5, left panel for illustration of negative power change). Note that the events selected are the likeliest among other possible events. For example, the transition 2,1 can occur when the observed power change corresponds to a foreign datum and there was a missing datum between the current power change observation and the previous power change observation. Events corresponding to an observed positive power change can be listed similarly.

TABLE II
TRANSITION SCENARIOS FOR OBSERVED NEGATIVE CHANGE OF POWER

Transition between states	Underlying Event(s)
1,1	Is foreign datum
1,2	Was missing datum from state 2 and is foreign datum
1,3	Was missing datum from state 3 and is foreign datum
1,4	Were missing data from states 2 and 3 and is foreign datum
2,1	Appliance i is turning off but appliance $(i+1)$ has not turned on
2,2	Is foreign datum
2,3	Was missing datum from state 4 and is 4,3 transition
2,4	Was missing datum from state 4 and is external datum
3,1	Appliance $(i+1)$ is turning off but appliance i has not turned on
3,2	Was missing datum from state 4 and is 4,2 transition
3,3	Is foreign datum
3,4	Was missing datum from state 4 and is foreign datum
4,1	Was missing datum from either state 2 or 3 and is 2,1 or 3,1 transition
4,2	Appliance $(i+1)$ is turning off but appliance i has not turned off
4,3	Appliance i is turning off but appliance $(i+1)$ has not turned off
4,4	Is foreign datum

The probabilities of the events listed in Table II can be calculated using the statistical distributions of power- and time features. Let the power draw at turn-off of appliance i be characterized by a PDF of negative change of power $p_i(\Delta P)$. Let PDFs of time-on and time-off durations be $f_i(\tau)$ and $g_i(\tau)$, respectively, and their Cumulative Distribution Functions (CDFs) – $F_i(\tau)$ and $G_i(\tau)$. The probability of, *e.g.*, 4,2 transition is then

$$P_{4,2} \propto p_{i+1}(\Delta P) f_{i+1}(\tau_{i+1,on})[1 - F_i(\tau_{i,on})], \quad (1)$$

where $\tau_{i,\mathrm{on}}$ and $\tau_{i+1,\mathrm{on}}$ are durations of time-on of appliances i and $(i+1)$. Note that, for the positive power change probabilities, the PDFs of additional power-related features (power surge and duration – see Section III.D) shall be incorporated when applicable.

In light of the discussion on the missing and foreign data, we assign a constant probability P_F of a foreign datum and a constant probability of P_M of a missing datum. Their dependence on time is modeled through a constant probability P_T. Therefore, the proposed method is approximated in two ways (i) only likeliest events underlying each system transition are considered and (ii) constants are used to model missing and foreign data.

C. *Modified Viterbi Algorithm*

Table II and Eq. (1) suggest that the Markov chain, which approximately models the system of appliances i and $(i+1)$ is of the first order for each appliance, but is of a higher order for the entire system. If, for example, frequency of use of appliance i is ten time greater than that of appliance $(i+1)$, the current system state may depend on a system state ten steps back in time. This implies that the standard Viterbi algorithm that optimally decodes the states of a first-order Markov chain, needs to be modified.

The standard Viterbi algorithm obtains the maximum likelihood estimation $\{\hat{s}_t\}$ of the state sequence $\{s_t\}$, given the transition observations $\{\omega_t)\}$, where t is the series of discrete transition times [27]:

$$\{\hat{s}_t\} = \underset{\{s(t)\}}{\arg\max}[\{s_t\} \mid \{\omega_t)\}] \tag{2}$$

The algorithm calculates the maximum probability of observing a state after each transition step. In the last step, the state with maximum probability is selected and traced back to recover the most likely transition sequence [28].

Modifications of Viterbi algorithm for a second-order Markov chain were considered, e.g., in Ref. [29]. The key idea is to keep in computer memory not only the likelihood calculation results from the previous time step, but also from a pre-previous step. In the case of two on-off appliances comprising the system, we need to keep in memory the likelihood calculated at one step prior to either appliance status change. Even though the number of such previous steps is not known *a priori*, the computer implementation is straightforward.

D. *Disaggregation by Triplet Processing*

The algorithm implementation under the strategy of working with the pairs can take several forms. For example, for appliance i, we define a power draw range within which we consider a triplet $\{(i-1), i, (i+1)\}$. The positive and negative changes of power that fall within this range are selected. These selected power changes ΔP_1, ΔP_2, … are recorded at known time points t_1, t_2, … Then, the appliance pairs $\{i, (i-1)\}$ and $\{i, (i+1)\}$ are firstly independently processed by application of the modified Viterbi algorithm to the time series of the selected power changes. Then, the time points of no state change are identified. Such time points found in the $\{i, (i-1)\}$ pair, if match those identified in the $\{i, (i+1)\}$ pair, are excluded from consideration as the foreign points, and the Viterbi algorithm is reapplied to the remaining time points in the both pairs.

After this two-stage processing, the decoded information on the state of appliance i is fused from the two pairs using the maximum likelihood principle. Suppose that a positive power change ΔP_k, obtained at t_k was decoded as belonging to appliance i in one pair, and to appliance j ($j = i$, i -1 or $i+1$) in the other pair. The corresponding negative power change ΔP_{mi} ($m_i > k$) is identified as belonging to appliance i in the first pair, whereas for the second pair, the corresponding to appliance j negative power change is decoded to be ΔP_{mj} ($m_j > k$). The likelihood of obtaining the sequence $\Delta P_k, \Delta P_{mi}$ is then compared with the likelihood of obtaining the alternative sequence $\Delta P_k, \Delta P_{mj}$, and the sequence with the highest likelihood is chosen. For the single changes of power that cannot be matched to the opposite changes of power (in case of missing data), the likelihoods of these single data points are processed similarly.

Once power changes of all the appliances are processed in this way, we need to update the results, because treating the appliances separately may lead to overlaps in decoding. We fuse the information decoded for the individual appliances using the maximum likelihood approach as described above.

E. *Estimation of Power and Time Distributions*

To implement the algorithm, we need to estimate the distributions of power and time features. To this end, we use an approach similar to that of Baranski [16], [17], [18]. The estimation procedure starts with analyzing and modeling of historical data, i.e., the data collected by a HED over, e.g., a two-week period. The raw data are pre-processed to merge the incremental power changes (see Fig. 4) and to filter out the noise.

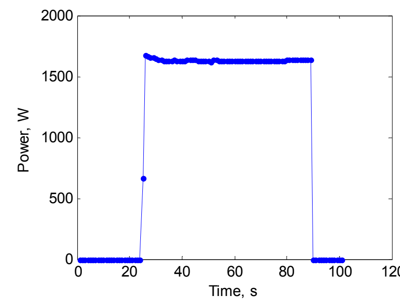

Fig. 4. Appliance takes two time increments to fully draw power at turn on. These increments need to be merged.

Stepwise negative changes of power are then identified (see Fig. 5, left panel). The identified negative changes of power are grouped into clusters. We use a simple ISODATA algorithm [30] for this purpose. Each cluster may correspond to a separate appliance being turned off.

Similarly, positive changes of power are identified. However, since some of the appliances exhibit surge at start, the post-surge value is used for the positive power change. Fig. 5, right panel shows an example of a surge and of the post-surge power change. The surge is characterized by its magnitude (ΔP_{surge}) and duration (Δt_{surge}). The positive changes of power are not clustered at this stage.

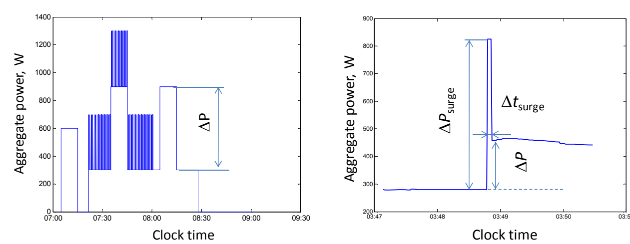

Fig. 5. Typical negative (left) and positive (right) features of power changes in aggregate electricity power. ΔP_{surge} is the power surge amplitude, ΔP is the steady-state change of power and Δt_{surge} is the surge duration in time.

The clustering procedure can result in errors. A single appliance can correspond to several identified clusters that need to be merged. Several appliances can correspond to a single cluster, which needs to be split. For cluster merging, the empirical statistics of appliance usage in time is calculated.

The statistics we use is the hourly presence or absence of the negative change of power for a given cluster. The degree of usage similarity between cluster pairs is then estimated, e.g., by calculating a fraction of hours wherein both clusters are present. Adjacent clusters with the degree of usage similarity exceeding a threshold are merged. After merging, the clusters are numbered in the order of their mean value. Note that the cluster boundaries may overlap.

Clusters of negative power change are then matched to the positive power changes [5] and preliminary time statistics of appliances being at state "on" and "off" are estimated. Because of the overlap in appliance power draw, the estimated power change distributions are usually truncated. We assume that the distributions of power changes and those of time-on (time-off) are independent, so that the time statistics estimations are not affected by the limited power changes. Finally, the clusters of positive power changes and, if applicable, the power surges, are statistically characterized. For the CDFs of time features, we use the empirical CDFs, and for the PDFs of both time and power features we use two-component Laplace mixture models. The obtained time estimates are also used to calculate the initial probabilities for the Markov models.

F. *Algorithm Realization*

Fig. 6 shows a high-level block scheme of the proposed method for historical data. The data from a HED, collected for a long period of time (about two weeks) are used for clustering and initial statistical estimation of both power (changes of power, surge amplitude) and time (time-on, time-off, surge duration) features as is explained in Section III.E. Both historical data and the estimated distributions are then used as inputs for the disaggregation algorithm (see Section III.D). The statistical distributions of the features are then updated. The real time version of our method works similarly. The matching between the identified clusters and real appliances is beyond the scope of this work (see Section VI).

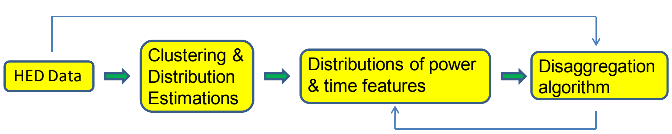

Fig. 6. High-level block scheme of the disaggregation method for historical data.

IV. SIMULATION EXAMPLE

To illustrate the work of the main algorithm, we devised a simulation example. Five on-off appliances with overlapping power draw and also overlapping time-on/time-off durations are simulated using quasi-random numbers generated by mutually independent statistical distributions. The positive and negative power changes associated with appliance turning on and off are characterized by normal (Gaussian) distributions, whereas the time durations are characterized by uniform distributions. We use different parameters (distribution means) for the positive and negative power changes of an appliance to make the simulation more realistic. We selected the uniform distribution for time-on/off to stay away from the exponential family distributions that are routinely used in Markov chain research. The distribution parameters are listed in Tables III and IV, and Fig. 7 (left panel) shows the simulated data. The total simulated time is about 17 hours.

TABLE III
PARAMETERS OF NORMAL DISTRIBUTIONS CHARACTERIZING POWER

Simulated appliance	Mean μ, positive change, W	Mean μ, negative change (absolute value), W	Standard deviation σ, W
1	110	105	10
2	130	135	13.5
3	150	160	10
4	180	190	13.5
5	210	210	10

TABLE IV
PARAMETERS OF UNIFORM DISTRIBUTIONS CHARACTERIZING TIME

Simulated appliance	Time-on, minimum, s	Time-on, maximum, s	Time-off, minimum, s	Time-off, maximum, s
1	30	100	400	600
2	70	150	500	600
3	100	180	350	500
4	150	270	200	400
5	40	140	300	700

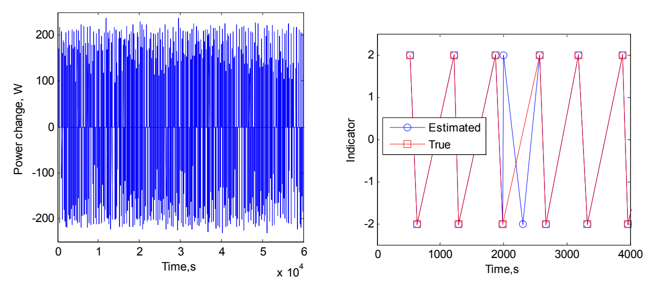

Fig. 7. Simulated data and disaggregation results. Left: aggregated power changes of five appliances. Right: Fragment of results of our algorithm for appliance 2. Overall correct reconstruction (F-measure) is 86.5%. The indicator takes value *i* if appliance *i* is turned on and value -*i* if appliance *i* is turned off.

Our algorithm (see Section III) includes estimation and disaggregation parts. For these simulated data, we only used the time estimation procedures. The disaggregation procedure was implemented as follows. For each of the appliances number i, i = 2, 3, 4, the power changes within $\mu_i \pm 4\sigma_i$ ranges are selected. Then, the appliance pairs $\{(i-1), i\}$ and $\{i, (i+1)\}$ are processed separately by the modified Viterbi algorithm using these power draws that fall within the ranges. Once the common outliers are excluded, the algorithm is reapplied and the decoding results for appliance i are fused. Finally, the decoding results for all appliances are optimized to minimize the overlap.

To assess accuracy of the disaggregation results, we utilize the F-measure [20]. The F-measure balances between the two types of errors (false negatives and false positives) by calculating a geometrical mean of precision and recall:

$$F = 2\frac{\frac{TP}{TP+FP}\cdot\frac{TP}{TP+FN}}{\frac{TP}{TP+FP}+\frac{TP}{TP+FN}} \tag{3}$$

In Eq. (3), TP stands for the number of true positives, FP – for false positives and FN – for false negatives. F-measure is not an optimal measure for disaggregation, because it does not account for errors in power change detection [6]. It is still applicable to the simulation data, because the power changes are given.

Fig. 7 (right panel) shows a fragment of disaggregation results for appliance 2. The calculated F-measure value for this appliance is 0.865.

It is interesting to compare the results of our disaggregation algorithm to the theoretical performance results of a Bayesian classifier [30], applied to the clusters of power draw with PDFs listed in Table III. If equal class probabilities are used, which is approximately the case for the simulated data, the Bayesian classifier uses the decision boundaries calculated by the intersections of the power draw PDFs. The F-measure can then be calculated.

Also of interest is the contribution of the time-off statistics to the algorithm performance. The conventional Markov models that allow non-exponential state occupancy time [20] still assume an exponential distribution of the time between state occupancies. The uniform distribution of time-off that we used in the simulation significantly differs from the exponential, so that our algorithm is expected to perform better.

TABLE V
DISAGGREGATION RESULTS

Simulated appliance	Bayesian classifier	Our algorithm, no time-off statistics used	Our algorithm
1	0.859	0.910	0.985
2	0.666	0.835	0.865
3	0.749	0.850	0.925
4	0.743	0.865	0.970
5	0.859	0.950	0.965

Table V presents the accuracy results. It is seen that the proposed algorithm dramatically improves the classification accuracy. The results of the algorithm with no time-off statistics used are somewhat intermediate. We conclude that the time-off features can be as important as the time-on features for the disaggregation.

V. REAL HOUSEHOLD EXAMPLE

Kolter and Johnson [31] have collected data in six households in Massachusetts. The data are collected using submetering at a circuit level and also by metering at power mains. Some circuits include just a single appliance, e.g., a refrigerator, whereas other circuits can include multiple appliances (lighting). The raw data are posted on the Internet and can be freely downloaded.

Unfortunately, since Kolter and Johnson measured apparent and not real power, the data collected on the mains do not match the sum of the submetered data. Also, only the data from house 1 include several common appliances (refrigerator, oven, dishwasher, microwave, stove) that were recorded for relatively long time (about 26 days). Therefore, we selected the data from house 1 for processing.

The submetering in house 1 is performed on 18 circuits, of which we selected nine. The other nine circuits were not used because: the four lighting circuits have several power modes, thus we could not reliably get the ground truth for them; a stove on a dedicated circuit was used only once; one washer-dryer circuit does not have meaningful data, and another washer-dryer circuit data are extremely noisy (we do not know if this noise is characteristic of the measurements or of the washing machine). The used circuits include (1) Oven 1, (2) Oven 2, (3) Refrigerator, (4) Dishwasher, (5) Microwave, (6) Bathroom GFI, and (7)-(9) Kitchen outlets. For the sake of convenience, we use Watts (W) for the power in Section V, even though the apparent power is measured in volt-amperes.

A. Preprocessing of submetered data

Since our model uses power changes and not the raw data, we need to process the data from each circuit in order to get the ground truth. The measured data exhibit significant surges (e.g., refrigerator) and incremental changes of power (see Fig. 4), which suggests that the estimated negative and positive power changes for each circuit may include errors. Table VI shows characteristics of the processed individual circuit data. It can be seen that significant overlap exists between circuits 1-2 and also between circuits 4-7.

TABLE VI
PARAMETERS OF POWER CHANGES

Circuit (appliance)	Positive, mean, W	Positive, standard deviation, W	Negative, mean, W	Negative, standard deviation,W
1 Refrigerator (significant power surges)	204.3	3.9	179.5	2.4
2 Dishwasher (lower mode, up to 300 W)	217.3	20.6	212.1	19.7
3 Kitchen outlet 1	1084.8	14.5	1074.6	15.1
4 Microwave	1548.5	52.4	1504.9	43.9
5 Kitchen outlet 2	1543.5	21.1	1533.5	14.5
6 Bathroom GFI (significant power surges)	1613.2	18.6	1607.3	18.4
7 Oven 1	1651.4	22.4	1640.7	21.1
8 Oven 2	2474.9	29.5	2448.4	24.2
9 Kitchen outlet 3 (significant power surges)	2767.9	51.3	2706.3	33.1

B. Aggregated Data Estimation

The raw data from individual nine submeters were aggregated together. For the dishwasher that has 2 power modes, only the lower power mode was used (< 300 W). The aggregated data were processed to merge the incremental power changes (see Fig. 4), separate the data from the surges and piecewise filter. After that, the power changes were calculated.

In clustering (see Section III.E), we used both negative power changes and power surges where applicable. More specifically, initial clustering by negative power changes resulted in seven clusters. The obtained clusters 1-3 were corresponding to actual circuits 1-3, cluster 4 was corresponding to circuits 4-7 combined, cluster 5 – to circuit 8 and clusters 6-7 to circuit 9. The use of power surge features has resulted in the separation of circuit 6, so that cluster 4 was split into three: cluster 4a (circuits 4 and 5), cluster 4b (circuit 6) and cluster 4c (circuit 7). Clusters 6 and 7 could not be merged together using our time-of-use criterion, because their underlying usage patterns were totally different. We, therefore, concluded that circuit 9 (kitchen outlet 3) actually includes two appliances. The unusually high value of the standard deviation of kitchen outlet 3 in Table VI supports this assumption. Cluster 4a is the only example of unsuccessful clustering; however, by observing Table VI we can conclude that the power draw range of circuit 5 is totally within that of circuit 4. Their time-on/time-off statistics are also similar. Therefore, we doubt that these two can be separated by a different method.

For the power feature estimates of the clusters, the mean values did not deviate significantly from the estimates listed in Table VI. The standard deviations were somewhat lower for the clusters corresponding to the overlapping circuits. For the time estimations, we have compared the empirical CDFs of time estimated from the clusters and those estimated from the submetered data. No statistically significant differences between the CDFs were found using the Kolmogorov-Smirnov criterion [32]. This coincidence supports our assumption about the independence of time and power features (see Section III.E).

However, since the ground truth data were available for the submetered data and not for the estimated clusters, we used the estimates obtained from the submetered data for disaggregation.

C. Disaggregation Results

The preprocessed aggregated data were disaggregated using the approach outlined in Sections III and IV. For the sake of comparison, the estimates of a simple Bayesian classifier were also calculated. Table VII shows the results in terms of the F-measure (see Section IV). The results for the first two circuits (refrigerator and dishwasher) and for the last two circuits (oven 2 and kitchen outlet 3) do not show any noticeable improvement by using of the proposed method. We believe that the main reason for this is that the ground truth results, estimated from the noisy data, are not perfect. Note that Kolter and Johnson [31] also used the submetered data for training of the Kim's model [20], and their results are significantly less accurate than the theoretical results of the Bayesian classifier.

TABLE VII
DISAGGREGATION RESULTS

Appliance from Table VI	Bayesian classifier	Our algorithm
1	0.859	0.831
2	0.881	0.846
3	0.989	0.936
4	0.775	0.899
5	0.409	0.840
6	0.753	0.927
7	0.800	0.908
8	1.0	0.962
9	1.0	0.971

At the same time, for the heavily overlapping circuits 4-7 the use of our method results in dramatic accuracy improvement. The accuracy level attained suggests that the proposed method can be used for disaggregation of HED data. Figs. 8-9 show examples of submetered data plotted together with the reconstructed disaggregation results. Note the problem of the incremental power merging for the microwave.

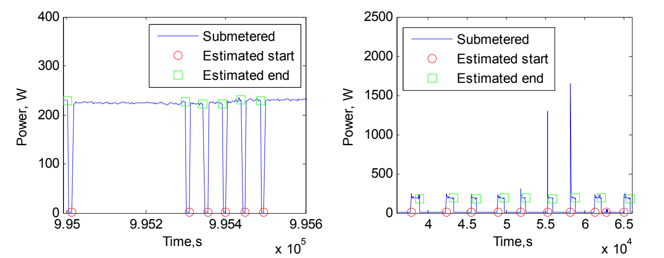

Fig. 8. Submetered data and reconstructed results for Dishwasher (left) and Refrigerator (right).

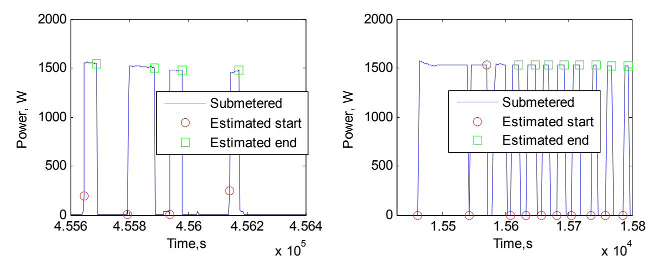

Fig. 9. Submetered data and reconstructed results for Microwave (left) and Kitchen outlet 2 (right).

VI. CONCLUSIONS

In this Section, we compare the requirements to a HED disaggregation method (see Section I) with the proposed approach.

- Feature selection: The selected features, power-related and time-related, are compatible with HEDs.

- Accuracy. The accuracy demonstrated for both simulated and real household data is of the order of at least 80-90%. Caution, however, shall be used in accuracy prediction, because the method accuracy depends on the degree of overlap in power draw of household appliances. We believe that working with triplets instead of pairs would improve the disaggregation accuracy. Other directions of future work on accuracy improvement include a better algorithm for change detection that can be based on the change-point problem [33], better algorithms for filtering and incremental merging and time-on/time-off estimates conditional on the time of the day.

- No training. The estimation component of the method uses historical data for clustering. The clusters can be matched with the actual household appliances, using both power and time features. Since the typical time-on and time-off values of many appliances are readily available [19], we believe the matching problem can be solved. Such matching, as well as more advanced methods for clustering will be considered in the future work.

- Real-time capabilities. The computational time of our MATLAB-algorithm processing of the real-household data collected over 26 days is about two minutes on a Pentium E8400, 3.00GHz machine. We believe, therefore, that the real-time implementation is feasible.

- Scalability. Since the algorithm complexity linearly depends on the number of appliances, and the accuracy is not directly related to this number, we believe the proposed method is scalable.

- Various appliance types. Currently, the method is applicable to on-off appliances. The extension to other appliance types can be based on the previous approaches [5], [16].

ACKNOWLEDGMENT

The author is grateful to Dr. Kurt Roth for stimulating discussions and insights.

REFERENCES

[1] Green Techn Media, "Smart Grid HAN Strategy Report 2011: Technologies, Market Forecast, and Leading Players," 2011.

[2] J. LaMarche, K. Cheney, S. Christian. K. Roth, "Home Energy Management Products & Trends," Fraunhofer CSE, 2011

[3] K. Ehrhardt-Martinez, K.A. Donnelly, and J.A. Laitner, "Advanced Metering Initiatives and Residential Feedback Programs: a Meta-Review for Household Electricity-Saving Opportunities," Report E105, ACEEE 2010.

[4] H. S. Cho; T. Yamazaki, H. Minsoo, "AERO: extraction of user's activities from electric power consumption data," *IEEE Trans. Consumer Electron.,* Vol. 56, no. 3, pp. 2011-2018, 2010.

[5] G. W. Hart, "Nonintrusive Appliance Load Monitoring," *Proceedings of the IEEE*, Vol. 80, pp. 1870-1891, 1992.

[6] M. Zeifman, K. Roth, "Nonintrusive Appliance Load Monitoring: Review and Outlook," *IEEE Trans. Consumer Electron.,* Vol. 57, no. 1, pp. 76-84, 2011.

[7] M. Zeifman, K. Roth, "Nonintrusive Appliance Load Monitoring: Review and Outlook," *Proceedings of IEEE International Conference on Consumer Electronics*, pp. 239-240, 2011.

[8] Caroline Duvier, Fraunhofer CSE, personal communication, 2011.

[9] S. Drenker, A. Kader, "Nonintrusive Monitoring of Electric Loads," *IEEE Computer Applications in Power*, pp. 47-51, October, 1999.

[10] A. Bouloutas, G. W. Hart, and M. Schwartz, "Two Extensions of the Viterbi Algorithm," *IEEE Transactions on Information Theory*, Vol. 37, pp. 430-436, 1991.

[11] G. W. Hart, "Correcting Dependent Errors in Sequences Generated by Finite-State Processes," *IEEE Transactions on Information Theory*, Vol. 39, pp. 1249-1260, 1993.

[12] A. I. Albicki, and A. Cole, "Data Extraction for Effective Non-Intrusive Identification of Residential Power Loads," *IEEE Instrumentation and Measurement Technology Conference*, pp. 812-815, 1998.

[13] A. . Albicki, and A. Cole, "Algorithm for Non-Intrusive Identification of Residential Appliances," *Circuits and Systems, Proceedings of the IEEE International Symposium on*, Vol. III, pp. 338-341, 1998.

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

[15] M. L. Marceau, R. Zmeureanu., "Nonintrusive Load Disaggregation Computer Program to Estimate the Energy Consumption of Major End Uses in Residential Buildings," *Energy Conversion & Manage*ment, Vol. 41, pp. 1389-1403, 2000.

[16] M. Baranski, J. Voss, "Non-Intrusive Appliance Load Monitoring Based on an Optical Sensor," *EEE Power Tech Conference*, Bologna, 2003.

[17] M. Baranski, J. Voss, "Genetic Algorithm for Pattern Detection in NIALM Systems," *IEEE International Conference on Systems, Man and Cybernetics*, pp. 3462-3468, 2004.

[18] M. Baranski, J. Voss, "Detecting Patterns of Appliances from Total Load Data Using a Dynamic Programming Approach,"*Fourth IEEE International Conference on Data Mining (ICDM'04)*, 2004.

[19] R. Ford, "Reducing Domestic Energy Consumption Through Behaviour Modification," D. Phil. Thesis, University of Oxford, 2009.

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

[21] Z. Ghahramani and M. I. Jordan, "Factorial hidden Markov models," *Machine Learning*, vol. 29, pp. 245-273, 1997.

[22] C.K. Reddy, C. Hsiao-Dong, B. Rajaratnam, "TRUST-TECH-Based Expectation Maximization for Learning Finite Mixture Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, pp. 1146 – 1157, 2008.

[23] K. B. Petersen, O. Winther, and L. K. Hansen, "On the slow convergence of EM and TOEM in low-noise linear models," *Neural Computation*, vol. 17, pp. 1921-1926, 2005.

[24] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, "Simulated Annealing Methods," Section 10.12. *Numerical Recipes: The Art of Scientific Computing (3rd ed.)*. New York: Cambridge University Press, 2007.

[25] M. Zeifman and K. Roth, "Viterbi Algorithm with Sparse Transitions (VAST) for Nonintrusive Load Monitoring," Proceedings of IEEE Symposium Series in Computational Intelligence, pp. 1-8, 2011.

[26] G. Forney, "The Viterbi Algorithm," Proceedings of the IEEE , pp. 268-278, 1973.

[27] F. O. Hocaoglu, O. N. Gerek, M. Kurban, "A novel wind speed modeling approach using atmospheric pressure observations and hidden Markov models," *J. Wind Eng. Ind. Aerodyn*. Vol. 98, pp. 472, 2010.

[28] L. R. Rabiner, B. H. Juang, "An introduction to hidden Markov models," *IEEE ASSP Magazine* vol. 3, p. 4, 1986.

[29] Y. He, "Extended Viterbi algorithm for second order hidden Markov process," International Conference on Pattern Recognition, pp. 718-720, 1988.

[30] J. T. Tou, R. C. Gonzalez, "Pattern Recognition Principles," New York, Addison-Wesley, 1974.

[31] J. Z. Kolter, M. J. Johnson, "REDD: A public data set for energy disaggregation research," In proceedings of the SustKDD workshop on Data Mining Applications in Sustainability, 2011.

[32] E. Suhir, "Applied Probability for Engineers and Scientists," McGraw-Hill, New York, 1997.

[33] H-J Kim, D. Siegmund, "The Likelihood Ratio Test for a Change-Point in Simple Linear Regression," *Biometrika* 76, pp. 409-423, 1989.

BIOGRAPHY

Michael I. Zeifman (M'10-SM'10) received his undergraduate degree in Engineering Physics from St. Petersburg State Polytechnic University, Russia, and a M.Sc. in Quality Control and a Ph.D. in Physical Reliability from Technion – Israeli Institute of Technology. Michael has developed physics-based methods for data analysis and signal processing that significantly improved accuracy and predictive capability. He is currently leading home energy monitoring and modeling work at the Fraunhofer Center for Sustainable Energy Systems.