Proficiency of Power Values for Load Disaggregation

Manfred Pöchacker, *Member, IEEE*, Dominik Egarter, *Student Member, IEEE*, and Wilfried Elmenreich, *Member, IEEE*

***Abstract*— Load disaggregation techniques infer the operation of different power-consuming devices from a single measurement point that records the total power drawn over time. Thus, a device consuming power at the moment can be understood as information encoded in the power draw. However, similar power draws or similar combinations of power draws limit the ability to detect the currently active device set. We present an information coding perspective of load disaggregation to enable a better understanding of this process and support its future improvement. In typical cases of quantity and types of devices and their respective power consumption, not all possible device configurations can be mapped to distinguishable power values. We introduce the term proficiency to describe the suitability of a device set for load disaggregation. We provide the notion and calculation of entropy of the initial device states, mutual information of power values, and the resulting uncertainty coefficient or proficiency. We show that the proficiency is highly dependent on the device running probability, especially for devices with multiple states of power consumption. The application of the concept is demonstrated by artificial data as well as with actual power consumption data from real-world power draw data sets.**

***Index Terms*— Information theory, load disaggregation, smart metering.**

I. Introduction

THERE are several reasons why it is beneficial for a power grid to get as much information as possible in order to accomplish monitoring and controlling purposes, like giving consumption feedback or detecting devices with high energy consumption. To avoid additional costs on hardware, installation, and operation, it is highly valuable to derive this information from a few, if not a single, measurement point(s). Load disaggregation or nonintrusive load monitoring (NILM) is a technique used for reasoning about the operation of power-consuming devices from a single measurement point recording the total power draw. One of its promising applications is the field of metering within smart homes [1], where information about single appliance usage is of high interest, but monitoring with many sensors is not an option. Low-cost power monitoring on device level is one step in the integration of residential buildings into the future smart grid, which is considered to be a key technology for carbon dioxide reduction. NILM works based on information about the involved devices and permissible assumptions on usage scenarios. This is replicable as power-consuming devices unintentionally encode information into the total power draw. Load disaggregation algorithms identify attributes within the measured data and draw meaningful conclusions about the overall consumption scenario.

Load disaggregation that exclusively works based on (active) power values is of high interest because active power is simple to measure and existing metering infrastructures usually provide the necessary values. With the upcoming smart meters, accessing the data gets even easier. However, a main drawback is that devices with similar consumption characteristics are hard to distinguish and simultaneously running devices add up in power values. As a consequence, the search space of possible power values at least doubles in size with each additional device. Devices with multiple values of power consumption additionally complicate the task. A single power value can be caused either by different devices with similar characteristics or by aggregation of multiple less consuming devices, which is why the distinction of all possible scenarios using exclusively power values is difficult.

Within this paper, we discuss the problem of indistinguishable power values caused by different device configurations. We use concepts of information theory to quantize the problem for a given device set by introducing the concept of *proficiency* for load disaggregation. It allows us to compare the extent of the problem for different device sets more objectively. We do so using real data of different measurement campaigns and houses with multistate devices. Although we exclusively use power values, the introduced concepts can be extended to capture as well other device attributes. We further investigate how proficiency is influenced by statistical operation probabilities of single devices and outline how the insights are useful for improvements of future NILM algorithms.

The goal of this paper is to better understand the mapping of device configuration scenarios to power values. We identify this as a coding procedure for information communication. This knowledge is helpful for further improvement of load disaggregation, which is decoding in that context. The problem treated in this paper is related to measurement accuracy, but it has a very different root, and the two problems can be

Manuscript received April 1, 2015; revised July 14, 2015; accepted July 23, 2015. This work was supported in part by the European Regional Development Fund under Grant 20214/23743/35470 (Project MONERGY) and in part by the Carinthian Economic Promotion Fund, and the state of Austria under Grant 20214/22935/34445 (Smart Microgrid Lab). The Associate Editor coordinating the review process was Dr. Amitava Chatterjee.

M. Pöchacker and W. Elmenreich are with the Lakeside Labs, Institute of Networked and Embedded Systems, Alpen-Adria-Universität Klagenfurt, Klagenfurt 9020, Austria (e-mail: manfred.poechacker@aau.at; wilfried.elmenreich@aau.at).

D. Egarter is with the Lakeside Labs, Institute of Networked and Embedded Systems, Alpen-Adria-Universität Klagenfurt, Klagenfurt 9020, Austria, and also with Kelag, Klagenfurt 9020, Austria (e-mail: dominik.egarter@kelag.at).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2015.2477156

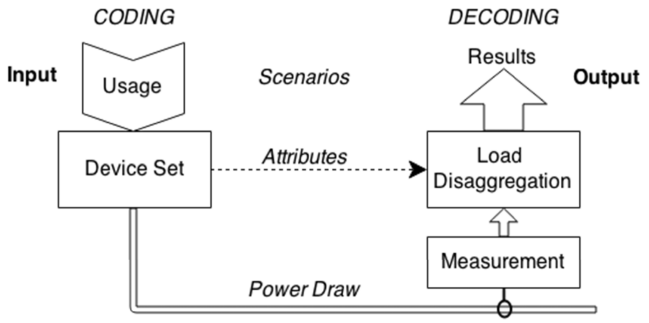

Fig. 1. Load disaggregation is the decoding procedure in an information communication process.

clearly separated. We exclusively use information theory for discrete sources and combinatorics. In that sense, we complement other work on limits of NILM by measurement accuracy as well as on quantification of disaggregation complexity for appliance sets.

Section II is dedicated to explain NILM as an information communication problem. Within Section III, the concepts of information theory are applied to the case of aggregated power values. Two artificial device sets are introduced as examples that contain solely ON–OFF devices. In Section IV, these concepts are extended to the more general case of devices with multiple power consumption values. In Section V, we apply the concepts to nine different appliance sets and compare them among themselves before we discuss the results and sketch possible follow-up work in Section VI. In Section VII, we summarize the results.

II. Load Disaggregation Within Information Communication

Fig. 1 shows a scheme of communication applied to load disaggregation, as it was introduced in [2]. It identifies load disaggregation as a decoding problem in the context of information communication theory. Load monitoring benefits from being nonintrusive, which means that any installation or device marking system is avoided. The primary source of information is the appliance usage, which causes power consumption. As the main purpose of a power cable is power supply, the utilized information content is produced unintentionally. The meaningful decoding of a signal stream on the power cable is the challenge of load disaggregation. The code, which is the mapping of the usage scenarios to the power line signals, is exclusively defined by the devices and their attributes. There are various attributes that enable identification of a specific device by its fingerprint on the power cable. Frequency, for instance, and nonharmonic device feedback on the input current is rich on information, but the required high-resolution measurement is usually costly and transmission functions of the power line circuits and their influences are barely known. To overcome such difficulties and for other arguments, such as those provided in [2], or costs, it is beneficial to use the so-called steady-state attributes, like power values, for device detection.

Several research has been done related to the process shown in Fig. 1, such as [3], in which limits for scenario detection due to measurement accuracy are analyzed. A successful and efficient detection of the desired scenario requests the different parts to be well coordinated. The applications of load disaggregation significantly differ concerning the acceptable effort on accuracy, maintenance, installation, computing power, measurement, and finally costs. There are examples that very high measurement rates enable distinguishing quite specific scenarios, e.g., which channel is watched on TV. Thanks to increasing performance of computation, different device attributes and techniques are successfully utilized for load disaggregation. But for many applications, especially for low-cost ones, a very high data volume causes more burden than benefit.

Current approaches to solve the load disaggregation problem can be divided into supervised and unsupervised approaches. A good overview on supervised approaches is given in [4] and [5]. The supervised approaches require a labeled data set to train a classifier and use either optimization- and/or pattern-recognition-based algorithms [6]. In the optimization-based approaches, the total power consumption and appliance power profiles are given. The composition of the known power profiles with minimal deviation to the total power consumption is selected as the current one [7], [8]. The pattern recognition approaches can be classified into clustering approaches [2], neural network algorithms [9]–[11], and algorithms based on support vector machines [12], [13]. The disadvantage of the all supervised classification approaches is the necessity of *a priori* information.

Accordingly, the recent research is more concerned with unsupervised algorithms, which are using unlabeled data. Unsupervised algorithms do not require any training data and, therefore, no *a priori* information of the system. Current approaches are based on dynamic time warping [14], clustering with blind source separation [15], hidden Markov models (HMMs) [16]–[18], factorial HMM [19], other variations of HMM [20], [21], temporal motif mining [22], and blind source separation [15]. For all these approaches, the distinction between appliances is unsupervised, whereas the labeling of a model with the corresponding appliance is not done automatically. Approaches performing automatic labeling are conducted based on Bayesian inference [23] and semisupervised classification [24].

The device states (specifically their values for power consumption) define the set of all possible device configurations, i.e., the state space. The usage is unknown and generates the aggregated power draw. Usage is dependent from the device operators and the build-in programs that make devices change their power consumption. The usage maps the possible device states to power values. Load disaggregation is reversing what usage does: the power profile constitutes input, and the current device states can be derived. The mapping of device states to power values is a coding process. The code depends on the power values of the device states exclusively. There is no guarantee that this code is uniquely decodable and possibilities to modify that code are limited.

Additional difficulties arise in the practice of load disaggregation, e.g., measurement resolution or noise, which are not considered within this paper. The theoretical constraints

demonstrated within this paper arise for an idealized case where integer power values characterize the device states. The explicit inclusion of measurement accuracy is on the one hand not necessary to demonstrate what we aim for and on the other hand offers no solution to the problem. The basic concepts are elucidated with ON–OFF devices and extended to multistate devices subsequently. Correlation between different states and time durations are not taken into account.

III. State Space of an Appliance Set

Within this section, only ON–OFF devices with only one single value of power consumption P_d are considered. The index d denotes the device number, so $d = 1, \ldots, N$ and N is the total number of devices. The set of devices and hence the set of power values

$$\mathbf{P}^D : \{P_1, \ldots, P_N\}$$

are known. We define the order of the device set in a way that $P_d \leq P_{d+1}$. Two or more devices with exactly equal power values cannot be ordered in a unique way and are indistinguishable in the assumed setting. In practice, the devices must be distinguished by another attribute. In this section, device set examples with distinct power values are used, but the formalism is not limited to this case. Examples with equal power values are used in Section IV and beyond.

Without any additional knowledge, it is possible to calculate all the possible power values P_k by the aggregation of consumption values of the device set. The state number k specifies the subset of devices, which is turned ON, and the complementary subset, which is turned OFF. The first state is defined as the power value $P_1 = 0$, and the last state's power value is the sum of all the single devices

$$P_M = P_{\text{total}} = \sum_{d=1}^{N} P_d \tag{1}$$

which can be used to characterize the device set. In between these particular cases, there are always $\binom{N}{z}$ cases where z out of the N devices are turned ON. The total number of possible states results in

$$M = \sum_{z=0}^{N} \binom{N}{z} = 2^N \tag{2}$$

which is equal to the possible states of a binary word of length N. In the context of load disaggregation, some of those states are very unlikely, even practicable impossible, to occur. But *a priori*, without knowing anything about the source and the emitted load profile, it is impossible to detect which ones are more likely to occur.

How these M states map to power values depends on the properties of the device set, i.e., the single device power values. The power value P_k of a specific state k is calculated by

$$P_k = \sum_{d=1}^{N} S_{kd} P_d \tag{3}$$

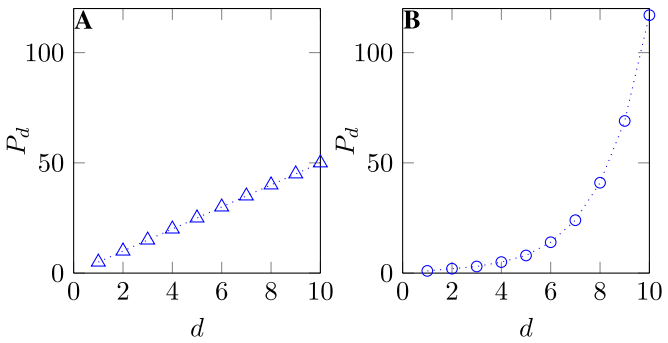

Fig. 2. Power values of device set A follow (4) and has $P_{\text{total}} = 275\ W$. Set B has $P_{\text{total}} = 284\ W$ and single device values according to (5).

where S_{kd} is the state matrix that contains a vector for each state k that holds a 1 for turned-ON and a 0 for turned-OFF devices. Repetition of all the states leads to the set of possible aggregated power values.

Furthermore, we refer to two examples of device sets containing ten ON-OFF devices of power values like shown in Fig. 2.

$$P_d = P_{d-1} + P_\Delta \tag{4}$$

where we use $P_1 = P_\Delta = 5$ W. Device set B contains the power values $\mathbf{P}^D : \{1, 2, 3, 5, 8, 14, 24, 41, 69, 117\}$. That power spectrum can be approximated by

$$P_d \approx \alpha P_{d-1} \tag{5}$$

for $\alpha = 1.69$ and $P_1 = 1$ W, and therefore is of power law type. In addition, these two sets have comparable total powers of 275 and 284 W, which are the same magnitude.

A load profile is a stream of power values P_i of length n. The total consumed energy is

$$E = \sum_{i=1}^{n} P_i \Delta t \tag{6}$$

where Δt is the sampling time and the power values P_i are averaged within a sampling duration. The average power of a load profile is

$$\hat{P} = \frac{E}{n \Delta t}. \tag{7}$$

The power values P_i result from the aggregation of power values of turned-ON devices at time step i so that

$$P(i) = \sum_{s=1}^{N} P_s S_s(i) \tag{8}$$

where N is the number of all the devices and $S_s^{\text{ON}}(i)$ is a Boolean state function, which is 1 when the device s is operated at time i and 0 otherwise.

A. Equal State Probability

When there is no knowledge of a source available, it is common in information theory to assume the maximum entropy case, which means equal likelihood for all possible source symbols. All the state probabilities p_k have the same

TABLE I

M POWER STATES, THE NUMBER OF TURNED-ON APPLIANCES z, AND n_z, THE NUMBER OF DIFFERENT STATES WITH THE SAME z

k	1	2...N+1	...	...	...	M-N...M-1	M
z	0	1	2	...	N-2	N-1	N
n_z	1	N	$\binom{N}{2}$	...	$\binom{N}{N-2}$	N	1

value of $1/M$, and the entropy of the source, which is defined as

$$H = -\sum_{k=1}^{M} p_k \log_2(p_k) \tag{9}$$

has its highest possible value of $H_{\max} = \log_2(M)$. $H_{\max}$ is an upper bound for the entropy of a discrete memoryless time-invariant source. The entropy $H_{\max}$ of a load source depends on the number of states or devices. As a first step, it would allow to compare the difficulty of load disaggregation problems with different numbers of devices. Furthermore, it is an upper bound for the entropy of any load profile from this source. Table I shows that there are as many states with one device ON as there are with one device OFF. This leads to the conclusion that if all the M states are hypothetically visited one time, each device is running exactly $M/2$ times, which is half of the total duration. The equal distribution of power state probabilities results in equal average runtime for each single device. Therefore, we get the correlation

$$\frac{1}{M}\sum_{k=1}^{M} P_k = \frac{1}{2} P_{\text{total}} \tag{10}$$

between the average state power and the aggregated power of the device set P_{total}.

Counting the number of states with a specific power value P_k gives a power state occupation number c. It can be written using the Dirac delta function as

$$c(P) = \sum_{k=1}^{M} \int_0^{\infty} \delta(P_k - P) dP. \tag{11}$$

An occupation number above one reflects the challenge of distinguishing between different states consuming the same power. States with this power value are not uniquely distinguishable. Fig. 3 shows the occupation numbers for the device set examples that have the same total number of states. For set A, there are up to 40 states that map to the same power value, while there are up to 8 in set B. In set A, a majority of power values (gray color) are not used at all. Load disaggregation is therefore expected to be more difficult for set A than for set B.

The power values in Fig. 3 represent all the states of a space between zero and P_{total}, which are available to encode the primary information. In that sense, power values are the channel within a theoretical communication setup where the device states are communicated. Otherwise, as for the classical coding problem in communication theory, the coding scheme is fixed and cannot be designed according to the channel transmission function. The power values can be seen as the information source of the receiver side. In this context,

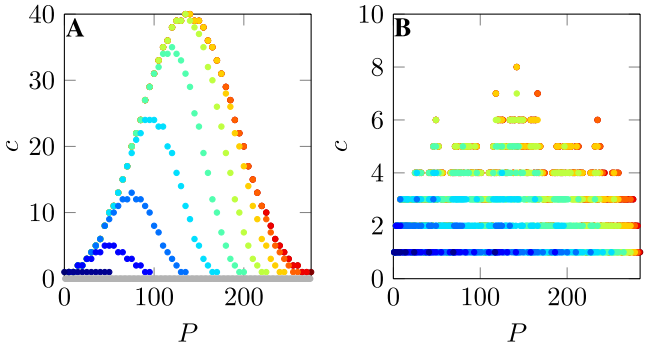

Fig. 3. Occupation numbers for sets A and B for all power values. The colors stand for the number of involved devices starting from $z = 1$ (blue) to $z = 10$ (dark red).

its entropy is the mutual information of the power value set I^P and is calculated by

$$I^P = -\sum_{P_j=0}^{P_{\text{total}}} p(P_j) \log_2(p(P_j)). \tag{12}$$

We assume the power values P_j to be a discrete set between 0 and P_{total}, but the definition can be extended to continuous probability density functions. For equal state likelihood, the power value probability is calculated by $c(P_j)$ so that we can define

$$I^P_{\max} = -\sum_{P_j=0}^{P_{\text{total}}} \frac{c(P_j)}{M} \log_2\left(\frac{c(P_j)}{M}\right) \tag{13}$$

which is the transported information by power values for the maximal entropy case. Note that it is not the theoretical maximum of transportable information by these power values. Therefore, it needs the averaged power state occupation number

$$\hat{c} = \langle c(P_j) \rangle. \tag{14}$$

On average, each of the $(M/\hat{c})$ power states is occurring with a probability of $(\hat{c}/M)$, which can be used further to approximate the mutual information by

$$I^P_{\max} \leq H_{\max} - \log_2(\hat{c}).$$

When the mutual information is smaller than the entropy of the source, it means that not all information can be transmitted, and therefore the stream cannot be decoded completely. As a measure for that loss of information, we suggest the uncertainty coefficient or proficiency, which is defined in information theory [25] as

$$C = \frac{I^P}{H} \tag{15}$$

and is shown to be a meaningful performance matrix in [26]. We name the proficiency for the maximal entropy case $C_{\max}$

$$C_{\max} = \frac{I^P_{\max}}{H_{\max}} \leq 1 - \frac{\log_2(\hat{c})}{H_{\max}}$$

which is restricted by an upper bound using the average occupation number.

TABLE II
DEVELOPED MEASURES OF AVERAGE INFORMATION FOR DEVICE SETS A AND B

	I^P_{max}	C_{max}	$\hat{c}$
Set A	5.33	0.53	18.3
Set B	8.04	0.80	3.6

Table II shows the developed information measures for device sets A and B.

For another hypothetical device set B2, which is similar to B with $\alpha = 2$ and $N = 10$, the occupation number is 1 for all the power values as shown in Fig. 6 (just like the binary representation of natural numbers). An equal probability in state space maps to equal distribution of power values with $c = 1$, which makes the mutual information reach the value of $H_{\max}$. It requires the power values space to be at least as big as the device state space to enable unique decoding. It means that only in the case $H_{\max} = I^P$, full load disaggregation by exclusive use of power values is possible. The proficiency and the averaged occupation number are both one in this case.

B. Equal Device Probability

From the point of load disaggregation, it is more suitable to deal with device probabilities than with probabilities for the combined power states. It is easier to relate devices to different user scenarios than to power states. The user-dependent devices follow behavioral patterns, e.g., starting coffee machine after getting up. Automatic devices (like a fridge) are turned ON regularly and therefore form a major part of the base load in a power draw. For many types of devices, characteristic operation probabilities can be estimated [21]. Even though their occurrence can vary, most of them are more likely to be switched OFF. For sizing of power lines (in a household), utilization factors are standard in engineering. The reasonable assumption is that not all the devices (or plugs) are used simultaneously, which allows installation of power lines with smaller cross section, which is more economic. The power factors are around 0.5 for households and little higher for industry or commercial installations, and they are expected to contain a safety buffer.

However, the state probabilities p_k can be easily estimated in case the single device operation probabilities p_d are known. From (3), for the state power, the calculation of the state probability p_k can be derived as

$$p_k = \prod_{d=1}^{N} (S_{kd}\, p_d + (1 - S_{kd})(1 - p_d)) \tag{16}$$

assuming that the devices are statistically independent. Specific device probabilities do not fit the maximum entropy assumption. But as there is no *a prior* knowledge on p_d, we use the expectancy value

$$\hat{p} = \langle p_d \rangle \tag{17}$$

for each device to demonstrate how the device sets' entropy is influenced. *A posteriori*, the average probability $\hat{p}$ for running any device can be calculated by

$$\hat{p} = \frac{E}{P_{\text{total}} n \Delta t} \tag{18}$$

using the energy E of a load profile of length n. In case the single devices runtimes n_d are known, even the device operation probability can be estimated by

$$p_d = \frac{n_d}{n}.$$

The averaged device probability is used to get

$$p_k(z) = \hat{p}^{z(k)}(1 - \hat{p})^{N - z(k)} \tag{19}$$

which is the state probability of a state with z turned-ON devices. It is a logarithmic function as shown in Fig. 4(left) for a set of ten devices. The state M, with all the devices ON, has the probability $\hat{p}^N$ and state 1 has $(1 - \hat{p})^N$, respectively. Fig. 4(right) shows the entropy

$$h(z) = -\binom{N}{z} p_k(z) \log_2(p_k(z))$$

which is an intermediate result when calculating the total source entropy $H = \sum_{z=1}^{N} h(z)$. In accordance with (10), the state probability is constant for the device probability of $\hat{p} = 0.5$. The total source entropy, which is shown in Table III, then reaches $H_{\max}$. The entropy function $h(z)$ is symmetric with respect to $\hat{p}$, which means that the total entropy for an operation probability of 0.1 is the same as for the probability 0.1 to be turned OFF.

The impact of device probabilities on the entropy propagates to power values, i.e., mutual information and proficiency. The calculation of the power value probabilities

$$p(P) = \sum_{k=1}^{M} p_k \int_0^{\infty} \delta(P_k - P) dP \tag{20}$$

TABLE III
TOTAL SOURCE ENTROPY H FOR DIFFERENT AVERAGED DEVICE PROBABILITIES $\hat{p}$

$\hat{p}$	0.1	0.3	0.5	0.7	0.9
H	4.69	8.81	10	8.81	4.69

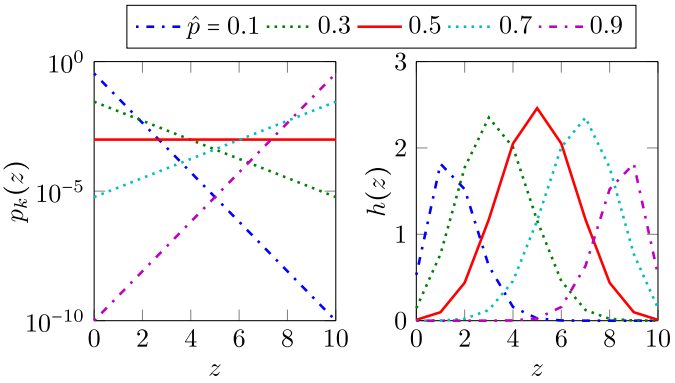

Fig. 4. Single state probabilities $p_k(z)$ (for z turned-ON devices) depend on the (averaged) device operation probabilities. The entropy $h(z)$ is additionally determined by the number of states.

TABLE IV
MUTUAL INFORMATION I^P AND PROFICIENCY C OF DEVICE SETS A AND B ACCORDING TO FIG. 5

	$\hat{p}$	0.1	0.3	0.5	$\hat{p}$	0.1	0.3	0.5
Set A	I^P	3.70	5.14	5.33	C	0.79	0.58	0.53
Set B		4.50	7.51	8.04		0.96	0.85	0.80

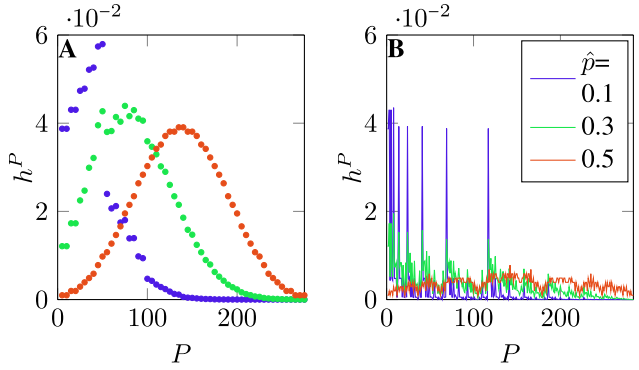

Fig. 5. Power value probabilities determine the entropy function h^P for the power states. We show it for sets A and B for three different averaged device probabilities $\hat{p}$.

requires consideration of the state probability instead of merely the occupation number $c(P)$. This is used to calculate the mutual information

$$I^P = -\sum_{P_j=0}^{P_{\text{total}}} p(P_j)\log_2(p(P_j)) = \sum_{P_j=0}^{P_{\text{total}}} h^P(P_j) \qquad (21)$$

of different single device probabilities. The function $h^P(P_j)$ is shown in Fig. 5 for device set examples A and B. The three different values of $\hat{p}$ in Fig. 3 are used for calculating the mutual information I^P and proficiency C in Table IV. Even though the mutual information for $\hat{p} = 0.5$ is higher, the proficiency and therefore the expected accuracy for disaggregation are lower than that for $\hat{p} = 0.1$. We conclude that equal probability for operation of all the single devices does not lead to equal occurrence of the states or power values.

IV. MULTISTATE DEVICES

Multistate appliances complicate the description of the power values for a set of devices. Power consumption for a device d is specified by a vector with an entry for all its $\hat{s}_d$ power values. A device set with N devices is, for instance, defined by

$$\mathbf{P}^D : \left\{\left(P_1^1, P_1^2\right); \left(P_2^1, P_2^2, P_2^3\right); (P_3); \ldots; \left(P_N^1, \ldots, P_N^{\hat{s}_N}\right)\right\}.$$

The second device has three power values $\hat{s}_2 = 3$, which means that the device has four possible states. The third device is an ON–OFF device with one power value. The total number of power values

$$S = \sum_{d=1}^{N} \hat{s}_d \qquad (22)$$

is a characteristic parameter for a device set (in the case of exclusive ON–OFF devices, $S = N$). We assume that all the power values are in increasing order to assure a unique description for a specific device set. The power values of a single device are sorted that $P_d^s < P_d^{s+1}$. The highest power value of a device $P_d^{\hat{s}_d}$ defines the order within the device set so that $P_d^{\hat{s}_d} < P_{d+1}^{\hat{s}_d+1}$.

Like in the case of simple devices, the number of possible states is calculated by multiplication of the number of states for all the N devices

$$M = \prod_{d=1}^{N} (\hat{s}_d + 1). \qquad (23)$$

The M states map to the power values P_k, which can be calculated by

$$P_k = \sum_{d=1}^{N} P_d^{S_{kd}} \qquad (24)$$

using a different notation than in (3). The state matrix element S_{kd} contains the power state of device d associated with state k, which is in accordance with its earlier usage. Now, S_{kd} is used as an index not as an exponent, so $P_d^{S_{kd}}$ is the power value of device d associated with state k. This notation requires the additional definition of P^0 for all the devices in a way that

$$P_d^0 = 0 \quad \forall \quad d = \{1, \ldots, N\}.$$

The mapping of the state number k to the device power state is more difficult than for exclusive ON–OFF devices, but follows a straightforward principle. For the above example, device 2 is OFF for the first $\hat{s}_1 + 1 = 3$ states, i.e., $S_{1,\ldots,3,2} = 0$ or more generally $S_{1,\ldots,3,d\geq 2} = 0$. For the states $k = 4, \ldots, 6$, device 2 runs with its first power value, i.e., $S_{4,\ldots,6,2} = 1$ and so forth. The power value of the last state is

$$P_M = P_{\text{total}} = \sum_{d=1}^{N} P_d^{\hat{s}_d}$$

which is the highest possible one. The occupation number c is estimated from the set of power values according to (11).

The multiple possible device states require a modification of the state probabilities p_k. The device probability is written in the same way as the power values so that p_d^s is the probability that device d is running on power value s. The state probability is then gained by

$$p_k = \prod_{d=1}^{N} p_d^{S_{kd}} \qquad (25)$$

with the usage of the device state matrix. The OFF-state probability p_d^0 needs to be calculated by

$$p_d^0 = 1 - \sum_{s=1}^{\hat{s}_d} p_d^s$$

as it is required within this notation. The notation introduced for multistate devices is more general and includes

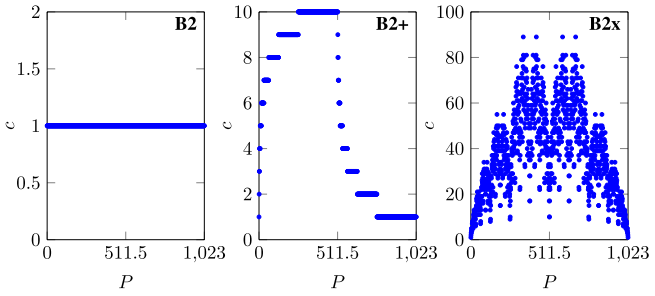

Fig. 6. Occupation numbers for power values of the three artificial device sets.

TABLE V
SEVERAL PARAMETERS FOR THE ARTIFICIAL SETS OF TEN DEVICES

Device Set	S	M	H_{max}	I^P_{max}	C_{max}	$\hat{c}$
B	10	1024	10	8.04	0.80	3.6
B2	10	1024	10	10	1	1
B2+	19	5632	12.46	9.6	0.77	5.5
B2x	19	39366	15.26	9.8	0.64	38.5

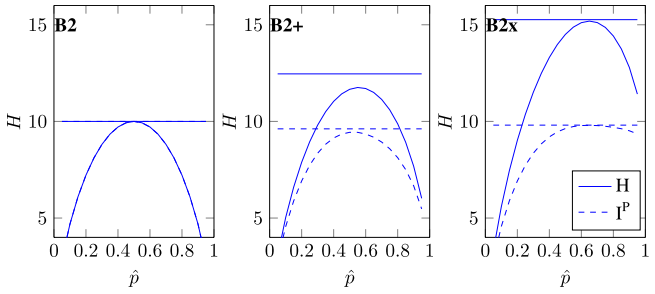

Fig. 7. Entropy H and mutual information I^P as a function of the device probability $\hat{p}$ for the three artificial device sets. The horizontal lines mark the values for the maximal entropy case.

the two state devices from Section III. The averaged device probability $\hat{p}$ is not an equal likelihood assumption. The assumption is on the likelihood of the OFF states, and the other device states are equally likely. The state probability is further used to estimate entropy (9), power values probability (20), and mutual information (21) just as in the case of ON–OFF devices.

To demonstrate the influence of multistate devices, we compare three artificial device sets. The multistate device sets are based on device set B2, which is constructed as set B but with the parameter $\alpha = 2$, i.e., all the states have the occupation number 1, which makes it trivial as visible in Fig. 6. Both derivative sets have nine additional states. For set B2+, device 10 has nine additional states with the power values of devices 1–9. For set B2×, the last nine devices have a second state with the power value of the previous device. The derivative sets have the same power values, but they are differently distributed among the devices. The reference values for several artificial device sets are listed in Table V. Fig. 6 shows the occupation numbers for the power values.

Entropy and mutual information are shown for different device probabilities in Fig. 7 including the values for the maximal entropy case. The maximum of the entropy curve for

TABLE VI
POWER VALUES OF THE DEVICE SETS ACCORDING TO [30]

Device Set	Power States
GreenD 1	[55 140 240], [1220], [60 148 470 570 1225 1265], [1790], [70 155 210 260 423 1898], [40 1900]
GreenD 2	[60], [80], [850], [1580], [80 1725], [90 173 1910]
GreenD 3	[110 235 285 360], [120 1235], [55 125 540 882 1047 1220 1630], [70 2002], [125 245 358 1998 2100], [70 160 2358 2550]
RedD 1	[200 420], [50 210 410 890 1115], [260 710 1440], [55 110 270 300 620 1405 1505], [1680 2478], [2705]
RedD 2	[123], [410], [160 420], [130 210 770], [1050], [40 1718 1850]
RedD 3	[100 400], [210 525 730], [40 365 900 1220 1520], [860 960 1285 1605], [120 540 1698], [2265]
Eco 1	[40], [72], [250 440 785], [50 1225], [1800], [90 180 250 365 2168]
Eco 2	[70], [55 175], [80 185], [50 310], [50 1840], [120 2132]
Eco 3	[100], [120], [130], [100 175 280], [40 1365 1485], [67 190 280 445 650 785 1065 1545]

set B2, which is reaching $H_{\max}$, shifts due to the additional power values in the extended sets. In set B2×, most devices (nine of ten) have two power values, which are equal to three states. For nine devices, the equal distribution of states is equivalent to the device probability of $\hat{p} = 2/3$, which is where the maximum occurs. For set B2+, the entropy function does not reach $H_{\max}$ (depicted as the horizontal line). This is due to differences in the number of power states between the devices. While device 10 reaches equal state distribution in $\hat{p} \approx 0.9$, all other two-state devices reach it at 0.5. In other words, device 10 is involved in many of the possible states, but is not operated more frequently to the same extent. The additional states in the derivative sets significantly increase the entropy while the mutual information is actually decreasing. This is caused by the constant total power P_{total} of the three device sets.

V. CASE STUDY ON REAL DEVICE SETS

We apply the measures developed within this paper to realistic device sets. We chose data sets frequently used for test cases within load disaggregation studies, such as the GreenD [27], the RedD [28], and the Eco [29] data set as used in [30] and [31]. To ensure comparability, we use exactly the same six appliances for each house as quoted as submetered power values in [31]. The power states of the appliance set were detected by an algorithm presented therein. For further information, e.g., how to extract appliance state information and the choice of appliances, we refer to [30].

All the parameters shown in Table VII directly result from the power values of the devices in Table VI. The values are presented in Fig. 8, which shows the houses ranked according to their number of states, and in Fig. 9 in which the set is sorted by descending proficiency. Fig. 8 shows entropy and mutual information for the maximal entropy case and characteristic power values, i.e., the total set power P_{total} and average device set power P_{av} of the sets in kilowatts. P_{av} is the expected average value when all the devices are turned ON. It is calculated by getting the average power for each device $\langle P_d \rangle = 1/\hat{s}_d \sum_s P_d^s$ and then averaging the device

TABLE VII
SEVERAL PARAMETERS FOR THE NINE SETS OF $N = 6$ DEVICES. COMPARISON IS SHOWN IN FIGS. 8 AND 9

Device Set	S	M	H_{max}	I^P_{max}	C_{max}	$\hat{c}$
GreenD1 ●	26	2352	11.2	10.21	0.91	1.23
GreenD2 ■	15	192	7.59	7.20	0.95	1.10
GreenD3 ♦	30	10800	13.4	11.69	0.87	1.76
RedD1 ●	26	3456	11.75	10.72	0.91	1.18
RedD2 ■	17	384	8.59	8.4	0.98	1.94
RedD3 ♦	24	2880	11.49	10.04	0.87	1.67
Eco1 ●	19	576	9.17	8.84	0.96	2.24
Eco2 ■	17	486	8.92	7.86	0.88	1.79
Eco3 ♦	23	1152	10.17	8.97	0.88	2.57

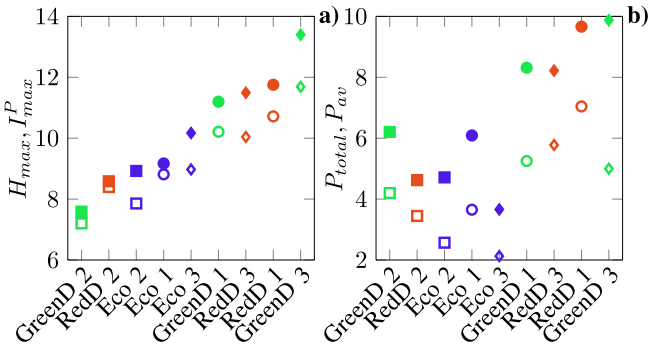

Fig. 8. (a) Maximal entropy (filled markers) and the related mutual information for all device sets. (b) P_{total} (filled) and the average of power states (empty markers) from all devices.

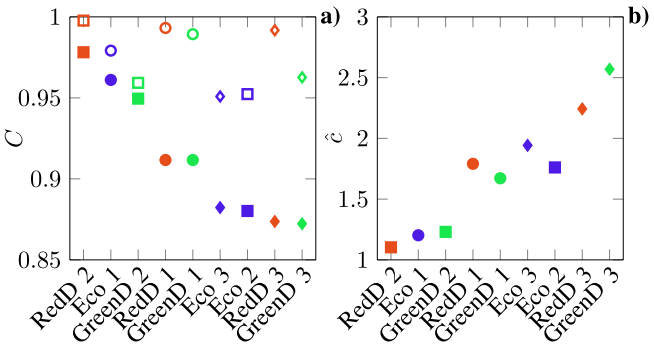

Fig. 9. (a) Proficiency C. Filled markers show the maximal entropy case, and empty ones show the values for $\hat{p} = 0.1$. (b) Average occupancy number $\hat{c}$.

set $P_{\text{av}} = 1/N \sum^d \langle P_d \rangle$. Statistically, data sets with more states are expected to have higher total power values. GreenD2 and Eco3 are exceptions, which lead to the conclusion that the bias of device selection can be significant. Conclusions about the number of states by the total power are inappropriate for individual device sets. The average set power is between 50% and 75% of the total set power.

Fig. 9(a) contains the proficiency for maximal values (filled) and the proficiency for low device probability (empty) with $\hat{p} = 0.1$. The latter one is obviously closer to one and is

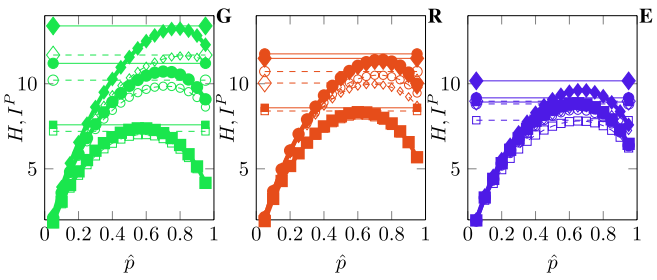

Fig. 10. Entropy and mutual information are shown for all data sets. The horizontal lines mark the respective values for the maximum entropy case.

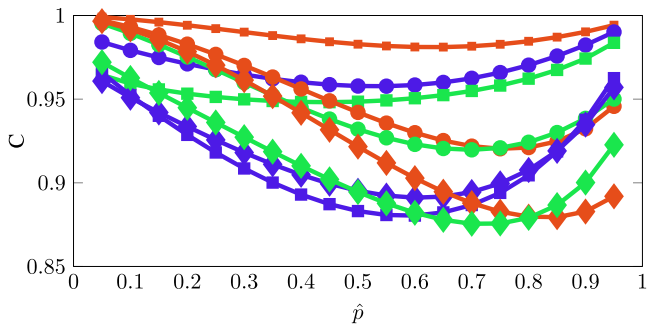

Fig. 11. Proficiency C changes with the averaged device probability. The function is mainly related to the distribution of power values among the devices.

a sample of Fig. 11. Device usage rates have a significant influence on proficiency. Fig. 9(b) shows the average occupation numbers $\hat{c}$ of power states. In general, it increases with decreasing proficiency. The real device sets all yield between 1 and 3, way below the values from the artificial sets A and B as listed in Tables II and V. The average occupation number measures the average equality of power values. The appliance set complexity (AC) from [30] is a measure for similarity of power values (without considering their likelihood), which includes similarity. If the distribution of modeling and measurement errors, which is assumed to be normal in [30], is of delta type, the AC is expected to match the value of $\hat{c}$ for a specific device set. The values for AC are therefore always above $\hat{c}$.

As demonstrated in Section IV, entropy and proficiency are a function of device operation probability. Fig. 10 shows entropy and mutual information for each device set grouped by the three data sets GreenD, RedD, and Eco. The maximal entropy values are depicted by the horizontal lines. In Fig. 11, the proficiency for the values presented in Fig. 10 is plotted. The device sets differently react to varying device probability. The three RedD sets get close to 1 for low $\hat{p}$, which means that the power values with view involved devices are generally distinguishable. The device sets RedD3 and GreenD3 have the lowest C-values at a comparatively high $\hat{p}$, which means that the indistinguishable power values include many devices, making them less likely to occur. The GreenD2 set is special, as proficiency is barely influenced by $\hat{p}$. It is ranked the lowest according to the number of states, i.e., M or H_{max} in Fig. 9(a), while for low $\hat{p} < 0.1$, proficiency is smaller than that of GreenD3, which is uppermost in Fig. 9(a).

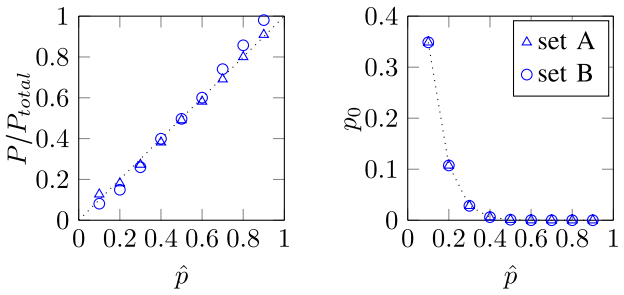

Fig. 12. Average device operation probability correlates with the average power of the resulting time series. The likelihood of the zero power value logarithmically decreases with increasing average device probability.

VI. Discussion

Load disaggregation is the decoding process within an information communication problem. The code exclusively depends on device attributes and their representation in the power draw. Entropy, as a measure of the amount of the initial states (equaling possible device configurations), has the advantage that it adds up in the case of two merging device sets. This is generally not true for the mutual information of power values, which is an entropy type measure as well. The values for the maximal entropy case are a bound for more realistic cases that include the probabilities of devices to run and the power values, respectively. Proficiency gives the fraction of information about the device states that can be reproduced from the power values. It therefore might qualify as an upper bound for detection rates of NILM algorithms. To show to which extend this is true would require the evaluation of an NILM algorithm with a considerable set of power draws. Furthermore, it is necessary to define mutual information for continuous power values with respect to the signal-to-noise ratio.

A set of power draws could be used for follow-up projects. The estimation of the single device power values by the analysis of the power draw histogram would improve unsupervised NILM. The probability for the usage of a specific type of device could be assessed by simple measures of a power draw. We analyzed random power draws based on set A and set B which are shown in Fig.12 to test the following. For a specific power draw, the total consumed average power in relation to the total power of the device set allows us to estimate the average device runtimes. The proportion of time steps without any running device indicates a similar reasoning and is easy to estimate.

The concepts developed in this paper can be extended to any parameter space with other attributes as used in [32]. Those can but do not necessarily include power values. The concepts help to decide if there are more promising attributes of the power draw to distinguish scenarios, or whether a single device can cause difficulties.

VII. Conclusion

We have modeled load disaggregation as a decoding process within an information communication problem. Description and improved understanding of the respective coding process helps in decoding. If power values are used for NILM, the coding scheme is likely to be not entirely bijective as not all possible device configurations are mapped to distinguishable power values. We have established the calculation of entropy of the initial device states, mutual information of power values, and the resulting uncertainty coefficient or proficiency. We demonstrated that the proficiency is highly dependent on the device running probability, especially for devices with multiple values of power consumption. We used artificial device sets as well as real measured values of devices that were repeatedly used for other load disaggregation studies to demonstrate the meaning of these parameters. The insights on the coding procedure from device states to aggregated power values contribute to the improvement of existing NILM algorithms.

References

[1] L. Peretto, "The role of measurements in the smart grid era," *IEEE Instrum. Meas. Mag.*, vol. 13, no. 3, pp. 22–25, Jun. 2010.

[2] G. W. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.

[3] R. Dong, L. Ratliff, H. Ohlsson, and S. S. Sastry. (Oct. 2013). "Fundamental limits of nonintrusive load monitoring." [Online]. Available: http://arxiv.org/abs/1310.7850

[4] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Trans. Consum. Electron.*, vol. 57, no. 1, pp. 76–84, Feb. 2011.

[5] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16838–16866, 2012.

[6] S. R. Shaw, S. B. Leeb, L. K. Norford, and R. W. Cox, "Nonintrusive load monitoring and diagnostics in power systems," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 7, pp. 1445–1454, Jul. 2008.

[7] J. Liang, S. K. K. Ng, G. Kendall, and J. W. M. Cheng, "Load signature study—Part I: Basic concept, structure, and methodology," *IEEE Trans. Power Del.*, vol. 25, no. 2, pp. 551–560, Apr. 2010.

[8] M. Baranski and J. Voss, "Genetic algorithm for pattern detection in NIALM systems," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2004, pp. 3462–3468.

[9] D. Srinivasan, W. S. Ng, and A. C. Liew, "Neural-network-based signature recognition for harmonic source identification," *IEEE Trans. Power Del.*, vol. 21, no. 1, pp. 398–405, Jan. 2006.

[10] T. Bier, D. O. Abdeslam, J. Merckle, and D. Benyoucef, "Smart meter systems detection & classification using artificial neural networks," in *Proc. 38th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2012, pp. 3324–3329.

[11] Y. Xu and J. V. Milanović, "Artificial-intelligence-based methodology for load disaggregation at bulk supply point," *IEEE Trans. Power Syst.*, vol. 30, no. 2, pp. 795–803, Mar. 2015.

[12] O. Kramer *et al.*, "On ensemble classifiers for nonintrusive appliance load monitoring," in *Hybrid Artificial Intelligent Systems* (Lecture Notes in Computer Science), vol. 7208, E. Corchado, V. Snášel, A. Abraham, M. Woźniak, M. Graña, and S.-B. Cho, Eds. Berlin, Germany: Springer-Verlag, 2012, pp. 322–331. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-28942-2_29

[13] H. Altrabalsi, J. Liao, L. Stankovic, and V. Stankovic, "A low-complexity energy disaggregation method: Performance and robustness," in *Proc. IEEE Symp. Comput. Intell. Appl. Smart Grid (CIASG)*, Dec. 2014, pp. 1–8.

[14] J. Liao, G. Elafoudi, L. Stankovic, and V. Stankovic, "Non-intrusive appliance load monitoring using low-resolution smart meter data," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Venice, Italy, Nov. 2014, pp. 535–540.

[15] H. Goncalves, A. Ocneanu, M. Bergés, and R. H. Fan, "Unsupervised disaggregation of appliances using aggregated consumption data," in *Proc. KDD Workshop Data Mining Appl. Sustainability*, Aug. 2011, pp. 21–24.

[16] T. Zia, D. Bruckner, and A. Zaidi, "A hidden Markov model based procedure for identifying household electric loads," in *Proc. 37th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Nov. 2011, pp. 3218–3223.

[17] S. Pattem, "Unsupervised disaggregation for non-intrusive load monitoring," in *Proc. 11th Int. Conf. Mach. Learn. Appl. (ICMLA)*, vol. 2. Dec. 2012, pp. 515–520.

[18] D. Egarter, V. P. Bhuvana, and W. Elmenreich, "PALDi: Online load disaggregation via particle filtering," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 2, pp. 467–477, Feb. 2015.
[19] A. Zoha, A. Gluhak, M. Nati, and M. A. Imran, "Low-power appliance monitoring using factorial hidden Markov models," in *Proc. IEEE 8th Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process.*, Apr. 2013, pp. 527–532.
[20] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial HMMs with application to energy disaggregation," in *Proc. 15th Int. Conf. Artif. Intell. Statist.*, 2012, pp. 1472–1482.
[21] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements," in *Proc. 11th SIAM Int. Conf. Data Mining*, 2011, pp. 747–758.
[22] H. Shao, M. Marwah, and N. Ramakrishnan, "A temporal motif mining approach to unsupervised energy disaggregation: Applications to residential and commercial buildings," in *Proc. 27th AAAI Conf. Artif. Intell.*, Bellevue, WA, USA, Jul. 2013, pp. 1327–1333.
[23] M. J. Johnson and A. S. Willsky, "Bayesian nonparametric hidden semi-Markov models," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 673–701, Feb. 2013.
[24] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "An unsupervised training method for non-intrusive appliance load monitoring," *Artif. Intell.*, vol. 217, pp. 1–19, Dec. 2014.
[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Telecommunications and Signal Processing). New York, NY, USA: Wiley, 2005.
[26] J. V. White, S. Steingold, and C. G. Fournelle, "Performance metrics for group-detection algorithms," in *Proc. Interf. Conf. Appl. Statist., Comput. Biol. Bioinformat.*, May 2004.
[27] A. Monacchi, D. Egarter, W. Elmenreich, S. D'Alessandro, and A. M. Tonello, "GREEND: An energy consumption dataset of households in Italy and Austria," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Venice, Italy, Nov. 2014, pp. 511–516.
[28] J. Z. Kolter and M. J. Johnson, "REDD: A public data set for energy disaggregation research," in *Proc. KDD Workshop Data Mining Appl. Sustain. (SustKDD)*, 2011.
[29] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, and S. Santini, "The ECO data set and the performance of non-intrusive load monitoring algorithms," in *Proc. 1st ACM Conf. Embedded Syst. Energy-Efficient Buildings (BuildSys)*, 2014, pp. 80–89. [Online]. Available: http://doi.acm.org/10.1145/2674061.2674064
[30] D. Egarter, M. Pöchacker, and W. Elmenreich. (Jan. 2015). "Complexity of power draws for load disaggregation." [Online]. Available: http://arxiv.org/abs/1501.02954
[31] M. Figueiredo, B. Ribeiro, and A. de Almeida, "Electrical signal source separation via nonnegative tensor factorization using on site measurements in a smart home," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 2, pp. 364–373, Feb. 2014.
[32] Y.-H. Lin and M.-S. Tsai, "Non-intrusive load monitoring by novel neuro-fuzzy classification considering uncertainties," *IEEE Trans. Smart Grid*, vol. 5, no. 5, pp. 2376–2384, Sep. 2014.

Manfred Pöchacker (M'13) received the Diploma Engineer degree in applied physics from the Vienna University of Technology, Vienna, Austria, in 2010, with a thesis on the stability analyses of a random genetic network model. He is currently pursuing the Ph.D. degree with the Smart Grids Group, Institute of Networked and Embedded Systems, University of Klagenfurt, Klagenfurt, Austria.

He has been involved in the simulation of power systems, specifically smart micro grids, since 2012. He is developing free and open-source software for the Renewable Alternative Power System Simulation-RAPSim. He is currently with Lakeside Labs, Klagenfurt, a research and innovation cluster on self-organizing networked systems. His recent publications are about time series simulation for Microgrids. His current research interests include entropy-related concepts, such as information theory and its application.

Mr. Pöchacker was involved in the activities of the local student branch of the IEEE, which he chaired in 2014.

Dominik Egarter (S'13) received the M.Sc. (Hons.) degree from the Higher Technical School for Electrical Engineering, Klagenfurt, Austria, in 2007, with a focus on energy and network engineering, and the Degree (Hons.) in information technology from the University of Klagenfurt, Klagenfurt, in 2012, with a focus on networked and embedded systems, where he is currently pursuing the Ph.D. degree with the Smart Grids Group, Institute of Networked and Embedded Systems.

He is with the Lakeside Labs, Klagenfurt, a research and innovation cluster on self-organizing networked systems. He has authored several papers in the field of smart energy systems with a focus on load disaggregation. His current research interests include intelligent energy applications in smart homes and smart micro grids, design and analysis of intelligent techniques and algorithms to detect legacy nonsmart appliances in aggregated household power draws, design of architectural requirements for smart appliances, integration of smart and nonsmart appliances into smart homes, and the design of concepts and algorithms to ensure household and consumer privacy.

Mr. Egarter is a Student Member of the Association for Computing Machinery.

Wilfried Elmenreich (M'03) received the Ph.D. and Venia Docendi degrees in computer science from the Vienna University of Technology, Vienna, Austria, in 2002 and 2008, respectively.

He was a Visiting Researcher with Vanderbilt University, Nashville, TN, USA, in 2005, and the CISTER/IPP-Hurray Research Unit, Polytechnic Institute of Porto, Porto, Portugal, in 2007. In 2007, he moved to the University of Klagenfurt, Klagenfurt, Austria, as a Senior Researcher. He was an Acting Professor of Complex Systems Engineering with the University of Passau, Passau, Germany, in 2012 and 2013. Since 2013, he has been a Professor of Smart Grids with the University of Klagenfurt, where he is currently with the Institute of Networked and Embedded Systems. He is also with Lakeside Labs, Klagenfurt, a research and innovation cluster on self-organizing networked systems. He has edited five books and has authored over 150 papers in networked and embedded systems.

Dr. Elmenreich is a Senate Member of the University of Klagenfurt and a Counselor of the Klagenfurt's IEEE Student Branch.