Device and Time Invariant Features for Transferable Non-Intrusive Load Monitoring

PASCAL A. SCHIRMER [ID][1,2] **AND IOSIF MPORAS** [ID][1]

[1]School of Physics, Engineering and Computer Science, University of Hertfordshire, Hatfield AL10 9AB, U.K.
[2]Department of Power Electronics, BMW AG, 80809 Munich, Germany
CORRESPONDING AUTHOR: P. A. SCHIRMER (p.schirmer@herts.ac.uk)

ABSTRACT   Non-Intrusive Load Monitoring aims to extract the energy consumption of individual electrical appliances through disaggregation of the total power consumption as measured by a single smart meter in a household. Although when data from the same household are used to train a disaggregation model the device disaggregation accuracy is quite high (80% - 95%), depending on the number of devices, the use of pre-trained disaggregation models in new households in most cases results in a significant reduction of disaggregation accuracy. In this article we propose a transferability approach for Non-Intrusive Load Monitoring using fractional calculus and normalized Karhunen Loeve Expansion based spectrograms followed by a Convolutional Neural Network in order to generate device characteristic features that do not change significantly across different households. The performance of the proposed methodology was evaluated using two publicly available datasets, namely REDD and REFIT. The proposed transferability approach improves the Mean Absolute Error by 13.1% when compared to other transfer learning approaches for energy disaggregation.

INDEX TERMS   Non-intrusive load monitoring, energy disaggregation, transferability, transfer NILM, fractional calculus, Karhunen Loeve expansion.

I. INTRODUCTION

NON Intrusive Load Monitoring (NILM) aims to extract the power consumption of each appliance of a building or a household using as input only the aggregated consumption signal [1]. While NILM is intrinsically a source separation problem three different approaches have been used to solve the NILM problem. First, pattern matching (elastic matching) techniques, which are detecting device signatures in the aggregated power consumption signal have been proposed [2]–[5]. Second, source separation methods, such as matrix and tensor factorization as well as sparse coding, have been utilized separating base components and activations [6]–[9]. Third, machine learning and deep learning based models have been used to generate data driven models to estimate the power consumption of devices from the aggregated signal [10]–[13].

The latest advances of machine learning and the development of big datasets have led to successful deep learning based NILM methodologies. NILM architectures using Convolutional Neural Networks (CNNs) [14], Long-Short-Term-Memory (LSTM) [15] and Recurrent Neural

Networks (RNNs) [16] have been proposed in the literature. In detail, a causal CNN with an optimization based on gate dilation was presented in [17] and a concatenated CNN approach for high sampling frequencies was proposed in [18]. As regards LSTM, a bidirectional approach with optimization on the forward and backward path, as well as Bayesian hyper-parameter optimization was presented in [12]. RNNs have been used in combination with convolutional layers in [19] and using deep RNNs in [16]. Additionally, latest research has focused on Generative Adversarial Networks (GANs) [20], [21] and bidirectional Transformers to incorporate self-attention mechanisms and to further improve the performance of the disaggregation algorithm [22].

Specifically, the above approaches have been only evaluated on the same dataset, thus with splitting of testing and training data from the same dataset, i.e. for non-transferability setups. However, as the ground-truth appliance signals are expensive to obtain [23], the transferability capability of a NILM architecture is crucial to be implemented efficiently and cost-effective in an actual smart-meter. Therefore, the more recent evaluations have considered the transferability

capability of the NILM architecture. However, only few papers provide distinct discussions on the specific case of transfer learning in the context of NILM [23]–[27] based on the usage of real data. In detail, in [27] an approach based on voltage/current-trajectories with dedicated feature colouring and usage of image-processing deep learning models is presented. However, the approach is only evaluated on appliances signals and not on the aggregated signal. In [25] a comparison of Gated Recurrent Unit (GRU) and CNN is performed utilizing a two-branch model layout, reporting similar performances for GRU and CNN with an advantage in terms of complexity for GRUs. However, [25] reports performances only for the transferability setup, thus the performance decrease compared to a non-transferability setup cannot be evaluated. Specifically, in [23], [26] (with [23] being an extension of [26]) explicit discussions on feature invariance for NILM are provided and a sequence-to-point (s2p) architecture is proposed in order to efficiently train a CNN model for transfer learning in NILM, reporting performances and comparisons for transferability and non-transferability setups.

In this work the idea of having invariant features for the same appliances from different data, as initially discussed in [23], will be extended by utilizing a time-frequency representation based on fractional KLE features with additional post-processing. The contribution of this paper is threefold. First, a definition of feature invariance in the context of NILM is presented, with consideration to the physical nature of the appliances. Second, a methodology for transfer learning in NILM based on invariant features will be presented. Third, a discussion on the most relevant topics for transfer learning in NILM will be provided, namely amount of training data, impact of normalization and algorithm convergence.

The remainder of this paper is organized as follows: In Section II an introduction about transferability for NILM is provided. In Section III the proposed method is presented. In Section IV the experimental setup is described and in Section V the results are presented. Finally, discussion is provided in Section VI and the article is concluded in Section VII.

II. TRANSFERABILITY FOR NILM

As transferability approaches are a relatively recent direction within the area of NILM, only few approaches have been discussed in the literature [23]–[27]. Specifically, most of the proposed approaches investigate previously published architectures in terms of their transferability capability and evaluate their performance on cross domain learning [23]. However, in order to achieve high accuracies for transfer NILM systems, the architecture and input feature vectors must be specifically optimized for the NILM problem, in order to enable accurate cross domain learning. The qualitative description of such an architecture is presented below.

Let's consider two different devices, namely a fridge (FIGURE 1a) and a washing machine (FIGURE 1b), for two different manufactures (e.g. Bosch, Siemens, Samsung,

etc.) each. First, considering the time domain signal for both fridges (FIGURE 1a (i)/(ii)) it can seen, that their power consumption values are different even though they operate in the same state. In detail fridge one consumes 75 W (FIGURE 1a (i)) in steady-state while fridge two consumes 100 W (FIGURE 1a (ii)), thus a difference in scaling along the y-direction is observed. Similar observations can be made for the washing machine (FIGURE 1b (i)/(ii)). Second, there are possible shifts along the time axis, e.g. on/off transitions of the fridge or the washing machine might not be time aligned (FIGURE 1a (i)/(ii)). Third, the state probabilities are very different for the same device for each brand respectively, e.g. fridge one has by far longer off durations than fridge two (FIGURE 1a (v)/(vi)). Based on the above the following three aspects must be considered, for an accurate modelling of transfer learning in NILM:

1) Different scaling in y-direction through different power consumption values of the same device operating in the same state but being from a different manufacturer.
2) Time shifts along x-direction through different temporal patterns in different households.
3) Different state probabilities through different utilization approaches of the same device in different households.

To account for these three aspects, the following three approaches are proposed in order to efficiently model the differences of the same appliances from a different manufacturer.

A. SCALING POWER CONSUMPTION
First, let's assume that similar devices from different manufactures are based on very similar electrical circuits. This assumption is reasonable as most devices, e.g. fridges or washing machines, have the same electrical components, e.g. single-phase electrical motor in case of a fridge, and these components only vary in size, e.g. according to the volume of the fridge or washing machine. From power electronics theory it is know, that the output waveforms in the frequency domain only depends on the electrical architecture and scales with the fundamental component of the current [29]. Therefore, in order to accurately capture different scaling along the y-direction the appliances' power consumption should be transfer into the frequency domain and then be normalized to its fundamental component. The effect of normalization in the frequency domain can be seen in FIGURE 1a (iii)/(iv) and FIGURE 1b (iii)/(iv) for the fridge and washing machine respectively. FIGURE 1a (iii)/(iv) and FIGURE 1b (iii)/(iv) illustrate that the harmonics of two different brands of the same device are much closer after normalization. In this paper normalized KLE representation will be used to calculate frequency transforms, since it works well especially for low-frequency signals [11], [30].

B. TIME-SHIFTS
Second, time-shifts along the x-direction should be accounted through incorporating temporal information in
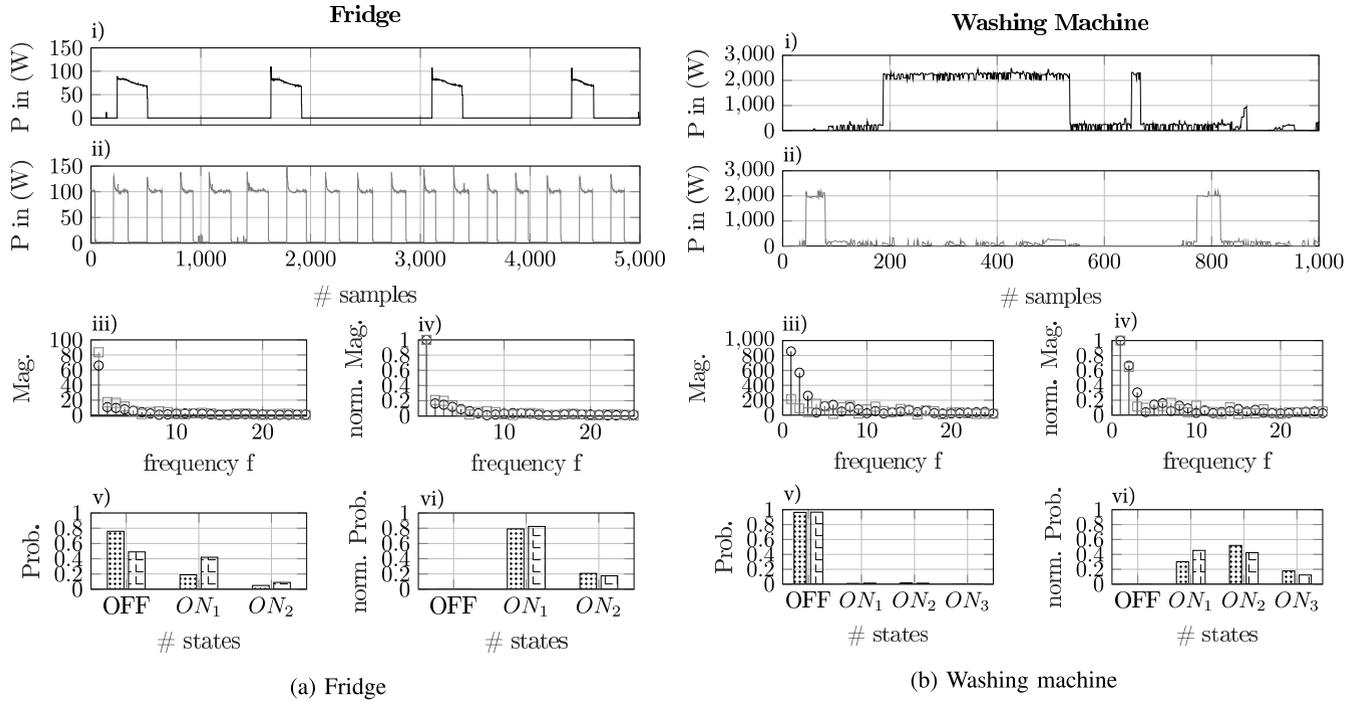
FIGURE 1. Comparison of two different appliances for two different brands respectively. a) Fridge and b) washing machine [28].

the architecture. Several different approaches have been proposed in literature to incorporate temporal information, including LSTM architectures [12], temporal concatenation [31], gate dilated CNNs [17], as well as fractional calculus [32]. In this implementation fractional calculus as proposed in [32] will be utilized to account for temporal shifts as it has been proven to work well with CNN architectures [32].

C. STATE PROBABILITIES

Third, as discussed before, a similar device from a different manufacture might show different state probabilities. This is illustrated in FIGURE 1a (v) and FIGURE 1b (v) for the fridges and washing machines respectively. However, these differences are mostly caused by the user, who is defining the ratio of on/off states, e.g. how often a washing machine is used per week. Conversely, once a device is started the internal active states only depend on the device itself, e.g. a washing machines runs through a cycle of rinse, wash and spin [10]. Therefore, state probabilities should only consider active states as they are device dependent and not user dependent. An example of the effect of not considering inactive states is illustrated in FIGURE 1a (v) and FIGURE 1b (v), showing that active states are much closer when not considering inactive states. In order to account for this behavior, the post-processing of the proposed architecture utilizes active device states only.

III. PROPOSED ARCHITECTURE

Let's consider a set of M-1 known devices each consuming power p_m with $1 \leq m \leq M$, the aggregated power p_{agg}

measured by the sensor will be:

$$p_{agg} = f(p_1, p_2, \ldots, p_{M-1}, e) = \sum_{m=1}^{M-1} p_m + e = \sum_{m=1}^{M} p_m \tag{1}$$

where $e = p_M$ is noise generated by one or more unknown devices and $f(\cdot)$ is the aggregation function. In NILM the goal is to find precise estimations $\hat{p}_m$ of the power consumption of each device m using an estimation method $f^{-1}(\cdot)$, i.e.,

$$\hat{P} = \{\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_{M-1}, \hat{e}\} = f^{-1}(p_{agg}) \tag{2}$$

Based on the discussion in Section II, the proposed architecture includes fractional calculus features to account for time shifts, normalized KLE features to account for scaling in y-direction and state correction considering only active states in the post-processing. Therefore, Eq. 2 can be reformulated in terms of feature vectors describing the frequency content in terms of magnitudes $\boldsymbol{A}$, and phase angles $\boldsymbol{\Phi}$.

$$\hat{P} = \{\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_{M-1}, \hat{e}\} = f^{-1}(\boldsymbol{A}, \boldsymbol{\Phi}) \tag{3}$$

In detail, the architecture illustrated in FIGURE 2 consists of framing, calculation of fractional power values, frequency transformation using KLE with according normalization, CNN regression for each target device m to estimate the corresponding power consumption $\hat{p}_m$, and post-processing using state correction. Detailed mathematical description of each stage is given below.
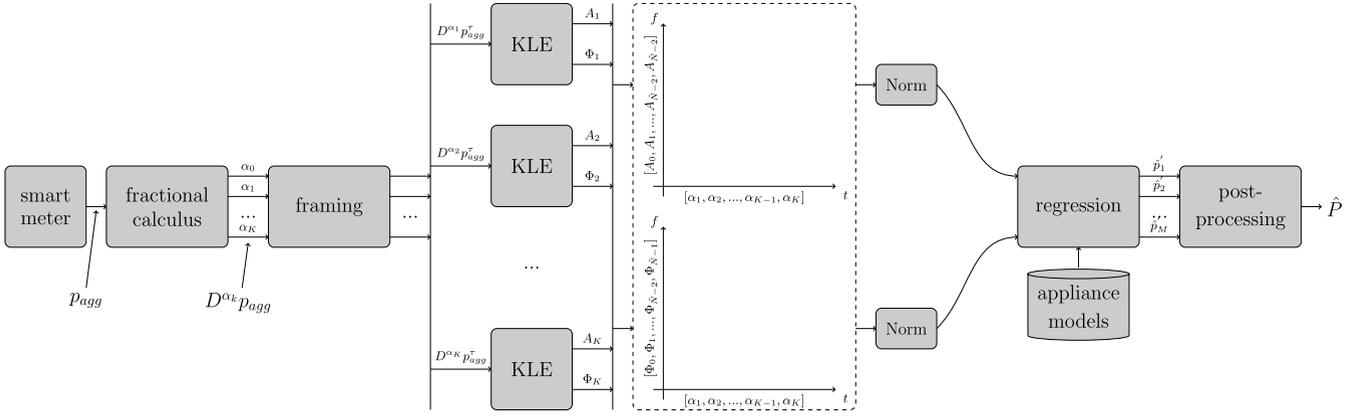
FIGURE 2. Block diagram of the proposed transferability NILM setup [28].

A. FRACTIONAL CALCULUS

Considering the aggregated signal $p_{agg}(t) \forall t : t \in \{1, \ldots, T\}$, extending the derivation to a non-integer order is given by the fundamental operator $_{t_0}D_t^\alpha$ where $[t_0, t]$ is the time frame considered and $\alpha \in \mathbb{R}$ is the fractional order [33]. Based on the above and according to Gruenwald-Letnikov [34] the fractional derivative of p_{agg} can be written as:

$$_{t_0}D_t^\alpha = \lim_{h \to 0} \frac{1}{h^\alpha} \sum_{j=0}^{\lfloor k \rfloor} (-1)^j \binom{\alpha}{j} p(t - jh) \tag{4}$$

where $k = \frac{t - t_0}{h}$ and $\lfloor k \rfloor$ is the integer part of k, h is the step width and $\binom{\alpha}{j}$ are the binomial coefficients defined by the factorial expansion of the Gamma function, $\Gamma(x)$, i.e.

$$\binom{\alpha}{j} = \frac{\alpha!}{j!(\alpha - j)!} = \frac{\Gamma(\alpha + 1)}{\Gamma(j + 1)\Gamma(\alpha - j + 1)} \tag{5}$$

B. NORMALIZED KLE

Considering K fractional components α_k with $k \in \{1, \ldots, K\}$ the fractional power signal can be written as $D^{\alpha_k} p_{agg}$. For transforming each fractional signal $D^{\alpha_k} p_{agg}$ to the frequency domain the KLE transform was used similar as in [35]. Therefore, let P_α denote one frame τ of the fractional signal $D^{\alpha_k} p_{agg}$ with frame length L. Specifically, let $\tilde{N}$ with $(\tilde{N} < L)$ be the order of the ACM used to separate each frame of the fractional signal P_α into its Subspace Components (SCs). The ACM Ψ_{PP} of signal P_α can be written as in [35]:

$$\Psi_{PP} = \begin{bmatrix} R_{PP}(0) & \ldots & R_{PP}(\tilde{N} - 1) \\ \vdots & \ddots & \vdots \\ R_{PP}(\tilde{N} - 1) & \ldots & R_{PP}(0) \end{bmatrix} \tag{6}$$

where $R_{PP}(n)$ with $0 < n < (\tilde{N} - 1)$ is the auto-correlation function of the signal P_α and n is a positive integer indicating the sample. Through eigenvector decomposition Ψ_{PP} can then be decomposed into $\tilde{N}$ mutually orthonormal eigenvectors $Q = [q_0, q_1, \ldots, q_{\tilde{N}-1}]$, where Q is unitary ($Q^T Q =$

$QQ^T = I$). The KLE transform and its inverse can be written as in Eq. 7 and Eq. 8 for each fractional component α.

$$\tilde{P}_\alpha = Q^T P_\alpha \tag{7}$$

$$P_\alpha = Q\tilde{P}_\alpha = \sum_{i=0}^{\tilde{N}-1} q_i^T P_\alpha q_i \tag{8}$$

where $\tilde{P}_\alpha \in \mathbb{R}^{\tilde{N}}$ is the KLE-transformed signal of P_α and the uncorrelated *SCs* of $\tilde{P}_\alpha$ are defined as $p_i = q_i^T P_\alpha q_i$, where p_i can be approximated by the coefficients of FIR filter [36]. Sinusoidal shape is assumed for each SC [35], thus $\tilde{P}_\alpha$ can be written in terms of magnitudes $A_\alpha \in \mathbb{R}^{\tilde{N}}$, and phase angles $\Phi_\alpha \in \mathbb{R}^{\tilde{N}}$. Furthermore, for each fractional component α a KLE transform was calculated resulting in a time-frequency representation of K time-slices and $\tilde{N}$ frequency components, i.e. $A \in \mathbb{R}^{\tilde{N} \times K}$ and $\Phi \in \mathbb{R}^{\tilde{N} \times K}$. Batch normalization was applied to KLE spectrum magnitude and phase as discussed in Section II.

C. POST-PROCESSING

To consider that the same appliance type can have different state probabilities, which might depend on outer parameters, e.g. user behavior, only the 'on' states of the appliance estimations are post-processed. An appliance is considered as being 'on', if the estimation of its active power consumption $\hat{p}_m$ is above a threshold θ. To determine the active device states, fuzzy c-means were used similar as in [37].

If the initial prediction of the regression model is too far from any cluster center of the c-means algorithm, i.e.:

$$\min_{1 \le n \le N} \left\| \hat{p}' - s_m^n \right\| \tag{9}$$

where $\hat{p}'$ is the initial estimation of the regression model, ϵ an appliance specific error margin and s_m^n is the cluster-center of nth state of the mth appliance calculated by the fuzzy c-means, the estimation was updated as follows:

$$\hat{p}_m = \begin{cases} \hat{p}', & \text{if } \hat{p}' \le \epsilon \\ s_m^{n_{min}}, & \text{if } \hat{p}' > \epsilon \end{cases} \tag{10}$$

TABLE 1. Short description of the REDD and REFIT dataset.

Name	Country	#-houses	#-devices	features	T_s
REDD	US	6	9-24	P	3s
REFIT	UK	20	9	P	8s

TABLE 2. Data splits for REDD and REFIT dataset.

Dataset	Training	Validation	Testing
REDD	3,4,6	1	2
REFIT	5,9,11	3	2

where $s_m^{n_{min}}$ is the nth state of the mth appliance fulfilling the minimum condition in Eq. 9. In Eq. 10 only active device states are post-processed according to the discussions in Section II.

IV. EXPERIMENTAL SETUP

The NILM architecture based on fractional calculus and KLE as described in Section III was evaluated using the datasets and regression algorithm presented below.

A. DATASETS

The proposed architecture was evaluated using two different datasets, namely REDD and REFIT [38], [39]. These datasets were chosen for two reasons. First, REDD was chosen as it is most commonly used in the energy disaggregation task, thus the proposed architecture is evaluated on the REDD dataset to show its performance on a classical (non-transferability) approach, and compare it with the existing literature. The REFIT datasets was chosen as it is ideal for training a transferability approach as it contains 20 different houses with similar appliances, but from different suppliers (e.g. fridges from Samsung, Bosch, Beko, etc.). Short description of the datasets can be found in Table 1.

In the previously published literature mostly five appliances are considered for disaggregation [10], namely kettle (KT), microwave (MW), dish washer (DW), fridge (FR) and washing machine (WM), the study was limited to these five appliances. Furthermore, the data split of the transferability setup was based on these five appliances, in order for all these appliances to appear in the training, validation and testing data respectively. The splits are tabulated in Table 2.

It must be noted that the data was not further modified, e.g. larger gaps in the data were not removed, in order to provide a realistic scenario for the transferability setup. Furthermore, the data was normalized using means-std normalization using the same values as in [23].

B. CNN-STRUCTURE AND MODEL PARAMETRIZATION

For the regression stage a two-dimensional CNN was used, similar as in [23]. Similarly to [23] one model per device was trained using relu activations for all intermediate layers and a linear activation in the last layer. Moreover, the one-dimensional kernels were replaced by two-dimensional

TABLE 3. Selected CNN structure with convolutional layers of the form Conv2d (Filters, Kernel, Padding, Strides).

Nr.	Layer	Nr.	Layer
1	Input	10	Relu
2	Conv2d(30,10,'same',1)	11	Conv2d(50,5,'same',1)
3	BatchNormalization	12	BatchNormalization
4	Relu	13	Relu
5	Conv2d(30,8,'same',1)	14	Conv2d(50,5,'same',1)
6	BatchNormalization	15	Flatten
7	Relu	16	Dense(1024)
8	Conv2d(40,6,'same',1)	17	Dense(1)
9	BatchNormalization	18	Linear activation

TABLE 4. Parameter optimization of the CNN model with different numbers of SCs ($\tilde{N}$) and fractional components K. Best performances are shown in bold.

SCs $\tilde{N}$	Number of fractional components K				
	2	4	8	16	32
1	88.0%	88.5%	88.7%	87.8%	86.1%
2	88.0%	88.9%	88.6%	87.0%	86.7%
3	87.8%	87.8%	**89.0%**	88.4%	86.2%
4	88.9%	88.5%	87.9%	87.5%	85.7%
5	87.8%	88.4%	87.6%	88.1%	84.3%

TABLE 5. Hyper-parameters of the CNN model and parameters of the Adam solver, similar as proposed in [23].

Parameter	Value	Parameter	Value
input size (non-transfer)	64x8x2	learning rate	0.001
input size (transfer)	256x8x2	beta-1	0.9
batch size	1000	beta-2	0.999
epochs	50	epsilon	1e-8

ones, to account for the two-dimensional inputs (time-frequency representations) of the proposed method. Therefore, the notation of the kernel-size 'x' refers to the two-dimensional kernel (x,x). The free parameters are shown in Table 3.

The number of fractional signals K and the number of subspace components $\tilde{N}$, which are defining the input size, were optimized using grid search on a bootstrap dataset for both the conventional and the transferability setup. The results for the conventional setup are tabulated in Table 4.

As can be seen in Table 4 the optimized parameters were found to be $K=8$ and $\tilde{N}=64$. Conversely, for the transferability setup the optimal framelength, and thus number of SC components was found to be $\tilde{N}=256$. This is in line with the work in [23] indicating that transferability approaches need a large framelength of input samples to account for the local differences in appliances signals. TensorFlow was used to train the models. In detail, the Keras backend was used utilizing the Adam optimizer for the training of each model. The hyper-parameters of the model, as well as the parametrization of the Adam optimizer, are tabulated in Table 5.

TABLE 6. Six evaluated experimental protocols including their features and dimensionality D (dimensionality for the transferability setup in brackets).

Protocol	Features	D
#1	raw active power	64 (256)
#2	magnitudes frac-KLE	32x8 (128x8)
#3	phases frac-KLE	32x8 (128x8)
#4	magnitudes/phases frac-KLE	64x8 (256x8)
#5	magnitudes/phases/raw frac-KLE	64x8x2 (256x8x2)
#6	#5 + post-processing	64x8x2 (256x8x2)

C. EXPERIMENTAL PROTOCOLS

A total of six experimental protocols were designed, one with respect to the use of raw active power samples similar as in [23], serving as a baseline protocol, and five additional ones with respect to the discussions in Section II and Section III. In detail, the second and third protocol use the magnitudes and phases after the fractional KLE transformation. The fourth protocol uses both magnitudes and phases of the fractional KLE, while protocol five uses additionally the raw samples after the fractional calculation. Protocol six applies additional post-processing as discussed in Section III-C. The six protocols, including their dimensionality, are tabulated in Table 6.

V. EXPERIMENTAL RESULTS

The architecture presented in Section III was evaluated according to the experimental setup described in Section IV. In order to provide accurate comparison with non-transferability setups, the performance was evaluated in terms of estimation accuracy (E_{ACC}), as proposed in [38].

$$E_{ACC} = 1 - \frac{\sum_{t=1}^{T} \sum_{m=1}^{M} |\hat{p}_t^m - p_t^m|}{2 \sum_{t=1}^{T} \sum_{m=1}^{M} |p_t^m|} \quad (11)$$

where $\hat{p}_m$ is the estimated power, T is the number of disaggregated frames and M is the number of disaggregated devices. Furthermore, to compare with transferability approaches previously published in the literature, additional accuracy metrics' namely Mean Absolute Error (MAE) and normalized Signal Aggregated Error (SAE) are introduced.

$$MAE = \frac{1}{T} \sum_{t=1}^{T} |\hat{p}_m^t - p_m^t| \quad (12)$$

$$SAE = \frac{|E_m - \hat{E}_m|}{E_m} \quad (13)$$

where E_m denotes the total energy consumption of the mth appliance, i.e. $E_m = \sum_t^T p_m(t) \cdot T_s$, and $\hat{E}_m$ denotes its estimated value. The results are presented for both the conventional disaggregation setup, without consideration of transferability in Section V-A, as well as for the transferability setup in Section V-B.

TABLE 7. Disaggregation results in terms of E_{ACC} for the REDD dataset using four/five appliances and the conventional setup. Best performances are shown in bold.

Data	#1	#2	#3	#4	#5	#6
REDD-1	82.8%	86.7%	65.1%	87.8%	88.4%	**88.6%**
REDD-2	86.1%	89.2%	81.2%	89.7%	92.6%	**92.9%**
REDD-3	87.3%	81.1%	53.2%	88.0%	88.3%	**88.4%**
REDD-4	85.0%	86.1%	56.2%	86.5%	87.1%	**87.7%**
REDD-6	89.7%	90.3%	51.5%	91.1%	92.5%	**92.8%**
AVG	86.2%	86.7%	61.4%	88.6%	89.8%	**90.1%**

TABLE 8. Comparison of the disaggregation accuracies for five appliances of the REDD-2 dataset.

Method	Publication	Year	Dataset	E_{ACC}
Sparse HMM	[10]	2015	REDD-2	94.8%
SIQCP	[42]	2018	REDD-2	86.4%
F-HDP-HSMM	[43]	2013	REDD-2	84.8%
F-HDP-HMM	[43]	2013	REDD-2	70.7%
Frac-KLE	Proposed	2021	REDD-2	92.6 %

A. CONVENTIONAL SETUP

For the conventional setup, the houses 1-4 and 6 of the REDD dataset have been used. Specifically, house 5 has been removed due to its significant short monitoring duration [40]. For each house the data has been split in half, with the first half being used for training and the second half for testing respectively. The results are tabulated in Table 7.

As can be seen in Table 7 the performance increase along the different protocols, starting from 86.2% for the baseline protocol, using only a frame of active power samples, and reaching 89.8% when utilizing all features. An exception is protocol #−3 where only phase angles are used, this protocol shows significantly lower performances, which is in line with the work in [41] reporting low accuracies for NILM setups using phase angles. Additional post-processing further increase performance by 0.3%, reaching an average performance of 90.1%. In order to compare the proposed architecture with the literature considering non-transferability approaches the work is compared in Table 8 with the three best performing approaches using house 2 of the REDD dataset and the estimation accuracy as performance measure.

As can be seen in Table 8 the proposed approach outperforms all other approaches except of the Sparse HMM proposed in [10].

B. TRANSFERABILITY SETUP

In a further step, we evaluate the proposed transferability setup according to the data splits tabulated in Table 2, the results on device level are presented for REDD and REFIT in Table 9 and Table 10 respectively. To have better comparability with other transferability approaches the results are presented in terms of MAE instead of E_{ACC}.

TABLE 9. Disaggregation results in terms of MAE for the REDD database using four appliances and the transferability setup. Best performances are shown in bold.

App	#1	#2	#3	#4	#5	#6
MW	11.83	7.90	11.32	**7.51**	8.26	7.82
FR	33.59	40.37	56.04	34.98	31.84	**30.68**
DW	19.36	15.94	19.50	14.31	8.64	**8.01**
WM	2.32	2.68	2.89	2.66	2.31	**2.13**
AVG	16.78	16.72	22.44	14.87	12.76	**12.16**

TABLE 10. Disaggregation results in terms of MAE for the REFIT database using five appliances and the transferability setup. Best performances are shown in bold.

App	#1	#2	#3	#4	#5	#6
MW	6.96	5.87	6.12	4.81	4.33	**4.11**
FR	35.28	33.63	42.56	31.71	28.16	**27.32**
DW	101.45	66.75	81.23	64.01	61.68	**57.14**
WM	35.21	20.53	28.30	23.13	**18.11**	18.16
KT	24.51	23.71	25.50	20.12	17.82	**15.31**
AVG	40.68	30.10	36.74	28.76	26.02	**24.40**

TABLE 11. Disaggregation results comparison in terms of MAE and SAE for the REDD and REFIT database using four/five appliances and the transferability setup. Best performances are shown in bold.

App	REDD				REFIT			
	#5		[26]		#5		[23]	
	MAE	SAE	MAE	SAE	MAE	SAE	MAE	SAE
MW	25.47	0.09	28.20	**0.06**	**10.74**	0.19	12.66	0.17
FR	25.20	**0.07**	28.10	0.18	**18.37**	**0.10**	20.02	0.33
DW	20.28	**0.34**	20.05	0.57	**11.56**	0.24	12.26	0.26
WM	9.77	0.15	18.42	0.28	15.23	0.91	16.85	2.61
KT	-	-	-	-	5.41	0.11	6.83	0.13
AVG	20.18	0.16	23.69	0.27	**12.26**	**0.31**	13.72	0.70

As can be seen in Table 9 and Table 10 the MAE is being reduced along the experimental protocols, with exception of protocol #3, similar as for the conventional disaggregation setup. In detail, the average MAE is reduced from 16.8 to 12.8 for the REDD database, while a reduction from 40.7 to 26.0 is observed for REFIT respectively. In detail, the most significant reductions of MAE are observed for the DW and the FR in the REDD dataset (55.4% and 30.2%) and for the DW and the WM in the REFIT dataset (48.6% and 39.2%). Moreover, to assure exact comparison with the previously published literature, the following results are recalculated using the data splits from [26] for REDD and [23] for REFIT. To assure fair comparison protocol #5 is used and post-processing or state-correction is omitted as neither [26] nor [23] use a knowledge based post-processing after the regression stage. The results are tabulated in Table 11.

As can be seen in Table 11 the proposed approach outperforms the approaches from [23] on average reducing the MAE and SAE values by 3.4 and 0.11 for REDD and 1.46 and

0.39 for REFIT respectively. These reductions being equal to 13.1% and 40.7% for REDD and 10.6% and 55.7% for REFIT. Again, the most significant performance improvement can be found for the FR, WM and DW.

It must be noted that there are three instances where the proposed approach only reaches roughly equal performance for one of the performance measures, namely for the MW and DW in the REDD database and for the MW in the REFIT database. In detail, for the MW in the REDD database the MAE is improved (+2.73), while the SAE is slightly reduced (−0.03). This indicates that the proposed approach assigns less energy in total (worse SAE), but with a higher accuracy (better MAE), thus having a better false positive rate compared to [23]. A similar observation can be made for the MW setup of the REFIT database showing an improvement of MAE (+1.92) and a reduction of SAE (−0.02). Conversely, for the DW in the REDD database the MAE values are almost equal with a significantly better SAE value for the proposed approach, indicating that the approach in [26] has a higher false-negative rate.

VI. DISCUSSION

Further, to the experimental results discussed in Section V, three topics, which are crucial for transferability approaches, are discussed. In detail, the impact on training data is discussed in Section (Section VI-A), the impact on normalization is discussed in Section (Section VI-B) and the convergence and real-time capability is discussed in Section (Section VI-C).

A. IMPACT ON TRAINING DATA

With transferability approaches the impact on training data on the performance of the model is crucial, since a transfer model must capture the appliance signatures from a different data domain [23]. In order to investigate the effect of different amounts of training data, the MAE for different amounts of data is displayed for the proposed protocol #5 as well as for the approach in [23], for a conventional and a transferability setup respectively. In detail, the data from the REDD dataset was used, utilizing REDD-1 for testing and REDD-2,3,4,5,6 for training in case of the transferability approach, while REDD-1 was split in a training and testing set for the conventional approach respectively. The results are illustrated in FIGURE 3.

As illustrated in FIGURE 3 the proposed approach outperforms the reference for both the conventional and the transfer approach, showing a total reduction of MAE loss equal to 0.02 (13.3%) and 0.03 (9.9%) respectively at the end of the training. In detail, the proposed approach shows an almost constant improvement for the non-transferability approach, while the advantage of the transfer approach is most significant above 700 k data-samples. Moreover, it can be seen that the MAE loss decreases smoothly for the conventional approach, while for the transfer approach both protocol #5 and [23] show an increase of MAE between 700-900 k data samples. This is probably due to using house five of the
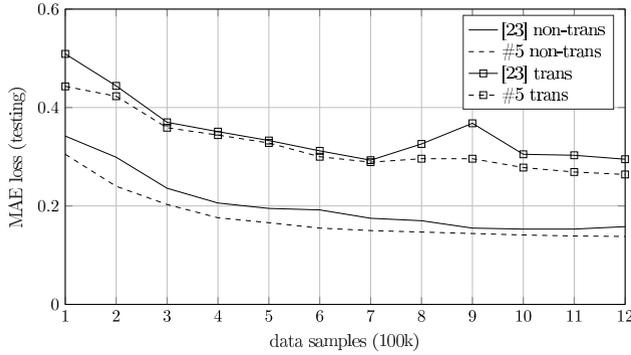
FIGURE 3. Impact of training data on two different transfer approaches (protocol 5 and [23]) for a conventional (non-trans) and a transfer (trans) approach respectively.

REDD dataset for training, which is normally excluded due to large gaps and especially short monitoring duration [40], leading to an increase of the MAE scores. However, it can also be seen that the proposed protocol #5 has a significantly smaller increase in MAE than [23], thus being more robust against the negative impact of using REDD-5 during the training process.

Furthermore, not only the amount of data samples has an influence on the accuracy and the convergence of NILM models, but also the sampling rate [44]. Specifically, as shown in [44] the optimal sampling rate of each device that is being disaggregated varies. The proposed model is utilizing data with sampling period of 3 sec and 6 sec using the KLE low-frequency features. Good performance (85% - 95%) has also been shown in [14], [17] using two-dimensional signatures for sampling rates at one sample per minute. Energy disaggregation has also been evaluated on hourly level [10]. While it was shown that excessive down-sampling does lower the performance of NILM approaches [44], good performances can be obtained for sampling rates as low as one sample per minute. Even lower sampling rates (below one sample per minute), have been evaluated in [45], but with focus on retrieving the average energy consumption over time, rather than the instantaneous power consumption.

B. IMPACT OF NORMALIZATION

As indicated in Section II, as well as in previous works for transferability in NILM [23], data normalization is crucial as appliances signatures must be invariant even though different data readings might have a different scaling (e.g. two different fridges might have a different power consumption for their on state as discussed in Section II). For the normalization to be effective, the normalization values must be constant parameters for all samples, but the normalization method might vary. Three different normalization methods will be discussed and compared to the baseline system without normalization applied. First Min/Max normalization insures that all training samples are within a range $[0,\ldots,1]$, with the same scaling value being used for validation and test data. Second,

TABLE 12. Influences of different normalization techniques in terms of MAE for the REDD dataset using four appliances and the transferability setup of [23].

App	Without	Min/Max	Mean/Std	Batch Norm
MW	28.09	26.80	25.93	25.47
FR	34.87	32.59	26.19	25.20
DW	25.56	25.43	21.47	20.28
WM	38.37	27.91	12.28	9.77
AVG	31.72	28.18	21.47	20.18
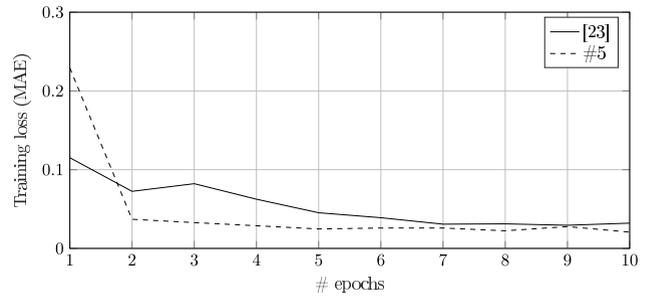
FIGURE 4. Convergence behavior for the proposed architecture and the architecture in [23] for a fixed data size (1 M samples) of the REDD database. Values are averaged across all appliances.

mean-variance scaling will be applied as in Eq. 14.

$$x = \frac{|x - \bar{x}|}{\sigma} \tag{14}$$

where $\bar{x}$ is the mean value of x and σ is the standard deviation (std) of x. Third, batch-normalizations will be used additionally to mean-std scaling for each layer of the CNN model, as a normalization for each batch might be useful especially when using frequency domain features as discussed in Section II. The results are tabulated in Table 12.

As tabulated in Table 12, the setup without normalization reports the worst performances across all appliances. This is in line with the theoretical discussion in Section II, stating a need for a constant scaling factor to level differences in power consumption from different appliances, and the observations in [23]. Conversely, Min/Max scaling as well as Mean/Std scaling improve the performance, reporting average MAE values of 28.2 and 21.5 respectively. Most significantly additional batch normalization further improves the performance, which is probably due of the equalization of the frequency patterns of the KLE as discussed in Section II.

C. CONVERGENCE AND REALTIME CAPABILITY

For sequence-to-point approaches with high dimensionality of the input data algorithm convergence and real-time capability is crucial. Therefore, the convergence behaviour for a fixed data size, as well as the execution times per sample have been calculated. The comparison of the convergence of protocol #5 of the proposed method is compared with the convergence of [23] and the results are illustrated in FIGURE 4.

TABLE 13. **Comparison of the average execution time (AET) per sample.**

Approach	[13]	#1	#2	#3	#4	#5
AET (ms)	6ms	3ms	8ms	8ms	15ms	60ms

As illustrated in FIGURE 4, both approaches converge within the first 10 epochs, but with a significantly faster decay for the proposed approach. This is probably due to the larger input size and the two-dimensional time-frequency representation providing more distinct features to the CNN model. A similar observation has been made in [17], where the usage of higher feature dimensions leads to faster convergence of the CNN model.

The execution time per sample for the 5 protocols (protocol #6 has been omitted as post-processing does not show measurable differences in execution time) has been calculated on an Intel i7 7700k CPU with 64 GB RAM using two Nvidia GTX 1080Ti in SLI mode. The Average Execution Time (AET) per sample, when using GPU calculations, is compared to the approach of [23], which was recalculated using the same hardware. The results are shown in Table 13.

As tabulated in Table 13 the AET increasing subsequently with using higher feature dimensionality, but report AETs well below real-time (3 ms to 60 ms for disaggregating 1 sec of the aggregated data). Specifically, the AET of the proposed protocol #5 is roughly ten times slower compare to the approach of [23] which is due to the higher feature dimensionality and especially the two-dimensional kernel of the CNN. However, as illustrated in FIGURE 4 the convergence is approximately by a factor five faster, thus the effective difference of AET is in the order of a factor of two.

VII. CONCLUSION

In this article a low-frequency approach for transfer learning in NILM has been proposed. Specifically, the solution is based on a low-frequency frequency feature description utilizing fractional calculus and the Karhunen Loeve Expansion in order to capture device and time invariant signatures to accurately disaggregated NILM signals from different datasets. The proposed methodology was evaluated on the REDD and REFIT dataset showing a maximum performance improvement by 55.4% when being compare to the baseline architecture and 13.1% when being compared to the best performing approach from the literature. Detailed analysis of the device performances, as well as on the influences of data size, convergence and normalization have been presented. Based on the results the following three points should be investigated: First, due to NILM being an highly ill-posed problem transfer learning greatly benefits from approaches that are physical related to the problem statement, i.e. accurate descriptions of the current harmonics as in the proposed approach. Second, due to different datasets having different size and number of appliances, the influence of transfer learning on completely unknown appliances has to be investigated. Third, as sampling frequencies vary across different datasets

and smart meter architectures, the effect on sampling frequencies on the performance should be evaluated.

REFERENCES

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

[2] P. A. Schirmer, I. Mporas, and M. Paraskevas, "Energy disaggregation using elastic matching algorithms," *Entropy*, vol. 22, no. 1, p. 71, Jan. 2020.

[3] K. He, L. Stankovic, J. Liao, and V. Stankovic, "Non-intrusive load disaggregation using graph signal processing," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1739–1747, May 2018.

[4] B. Liu, W. Luan, and Y. Yu, "Dynamic time warping based non-intrusive load transient identification," *Appl. Energy*, vol. 195, pp. 634–645, Jun. 2017.

[5] H. Wang and W. Yang, "An iterative load disaggregation approach based on appliance consumption pattern," *Appl. Sci.*, vol. 8, no. 4, p. 542, Apr. 2018.

[6] S. Singh and A. Majumdar, "Analysis co-sparse coding for energy disaggregation," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 462–470, Jan. 2019.

[7] S. Singh and A. Majumdar, "Deep sparse coding for non–intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4669–4678, Sep. 2018.

[8] A. Rahimpour, H. Qi, D. Fugate, and T. Kuruganti, "Non-intrusive energy disaggregation using non-negative matrix factorization with sum-to-K constraint," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4430–4441, Nov. 2017.

[9] P. A. Schirmer and I. Mporas, "Multivariate non-negative matrix factorization with application to energy disaggregation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 3285–3289.

[10] S. Makonin, F. Popowich, I. V. Bajic, B. Gill, and L. Bartram, "Exploiting HMM sparsity to perform online real-time noninstrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2575–2585, Nov. 2016.

[11] S. Welikala, C. Dinesh, M. P. B. Ekanayake, R. I. Godaliyadda, and J. Ekanayake, "Incorporating appliance usage patterns for non-intrusive load monitoring and load forecasting," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 448–461, Jan. 2019.

[12] M. Kaselimi, N. Doulamis, A. Doulamis, A. Voulodimos, and E. Protopapadakis, "Bayesian-optimized bidirectional LSTM regression model for non-intrusive load monitoring," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2747–2751.

[13] 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.*, vol. 22, N. D. Lawrence and M. Girolami, Eds. La Palma, Canary Islands, 2012, pp. 1472–1482.

[14] P. A. Schirmer and I. Mporas, "Low-frequency energy disaggregation based on active and reactive power signatures," in *Proc. 29th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2021, pp. 1426–1430.

[15] H. Rafiq, H. Zhang, H. Li, and M. K. Ochani, "Regularized LSTM based deep learning model: First step towards real-time non-intrusive load monitoring," in *Proc. IEEE Int. Conf. Smart Energy Grid Eng. (SEGE)*, H. A. Gabbar, Ed. Piscataway, NJ, USA: IEEE Press, Aug. 2018, pp. 234–239.

[16] L. Mauch and B. Yang, "A new approach for supervised power disaggregation by using a deep recurrent LSTM network," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Piscataway, NJ, USA, Dec. 2015, pp. 63–67.

[17] A. Harell, S. Makonin, and I. V. Bajic, "Wavenilm: A causal neural network for power disaggregation from the complex power signal," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 8335–8339.

[18] P. A. Schirmer and I. Mporas, "Double Fourier integral analysis based convolutional neural network regression for high-frequency energy disaggregation," *IEEE Trans. Emerg. Topics Comput. Intell.*, early access, Jun. 16, 2021, doi: 10.1109/TETCI.2021.3086226.

[19] M. Kaselimi, E. Protopapadakis, A. Voulodimos, N. Doulamis, and A. Doulamis, "Multi-channel recurrent convolutional neural networks for energy disaggregation," *IEEE Access*, vol. 7, pp. 81047–81056, 2019.

[20] M. Kaselimi, N. Doulamis, A. Voulodimos, A. Doulamis, and E. Protopapadakis, "EnerGAN++: A generative adversarial gated recurrent network for robust energy disaggregation," *IEEE Open J. Signal Process.*, vol. 2, pp. 1–16, 2021.

[21] M. Kaselimi, A. Voulodimos, E. Protopapadakis, N. Doulamis, and A. Doulamis, "EnerGAN: A generative adversarial network for energy disaggregation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1578–1582.

[22] Z. Yue, C. R. Witzig, D. Jorde, and H.-A. Jacobsen, "BERT4NILM: A bidirectional transformer model for non-intrusive load monitoring," in *Proc. 5th Int. Workshop Non-Intrusive Load Monitor.*, Nov. 2020, pp. 89–93.

[23] M. D'Incecco, S. Squartini, and M. Zhong, "Transfer learning for nonintrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1419–1429, Mar. 2020.

[24] Y. Liu, X. Wang, and W. You, "Non-intrusive load monitoring by voltage–current trajectory enabled transfer learning," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5609–5619, Sep. 2019.

[25] D. Murray, L. Stankovic, V. Stankovic, S. Lulic, and S. Sladojevic, "Transferability of neural network approaches for low-rate energy disaggregation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 8330–8334.

[26] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton, "Sequenceto-point learning with neural networks for nonintrusive load monitoring," 2016, *arXiv:1612.09106*.

[27] Y. Liu, Y. Sun, and B. Li, "A modified IP-based NILM approach using appliance characteristics extracted by 2-SAX," *IEEE Access*, vol. 7, pp. 48119–48128, 2019.

[28] P. A. Schirmer, "Modelling of electrical appliance signatures for energy disaggregation," Ph.D. dissertation, School Phys., Eng. Comput. Sci., Univ. Hertfordshire, Hatfield, U.K., 2021. [Online]. Available: https://147.197.131.104/handle/2299/24203

[29] D. G. Holmes and T. A. Lipo, *Pulse Width Modulation for Power Converters: Principles and Practice*, D. G. Holmes, T. A. Lipo, Hoboken, NJ, USA: Wiley, 2003. [Online]. Available: http://www.loc.gov/catdir/description/wiley037/2003057626.html

[30] P. A. Schirmer and I. Mporas, "Energy disaggregation from low sampling frequency measurements using multi-layer zero crossing rate," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 3777–3781.

[31] P. A. Schirmer, I. Mporas, and A. Sheikh-Akbari, "Robust energy disaggregation using appliance-specific temporal contextual information," *EURASIP J. Adv. Signal Process.*, vol. 2020, no. 1, p. 394, Dec. 2020.

[32] P. A. Schirmer and I. Mporas, "Energy disaggregation using fractional calculus," in *Proc. ICASSP—IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 3257–3261.

[33] M. Pavlickova and I. Petras, "A note on time series data analysis using a fractional calculus technique," in *Proc. 15th Int. Carpathian Control Conf. (ICCC)*, I. Petráš, Ed. Piscataway, NJ, USA: IEEE, May 2014, pp. 424–427.

[34] J. A. T. Machado, "Analysis and design of fractional-order digital control systems," *Syst. Anal. Model. Simul.*, vol. 27, pp. 107–122, Jan. 1997.

[35] C. Dinesh, B. W. Nettasinghe, R. I. Godaliyadda, M. P. B. Ekanayake, J. Ekanayake, and J. V. Wijayakulasooriya, "Residential appliance identification based on spectral information of low frequency smart meter measurements," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2781–2792, Nov. 2016.

[36] H. G. C. P. Dinesh, D. B. W. Nettasinghe, G. M. R. I. Godaliyadda, M. P. B. Ekanayake, J. V. Wijayakulasooriya, and J. B. Ekanayake, "A subspace signature based approach for residential appliances identification using less informative and low resolution smart meter data," in *Proc. 9th Int. Conf. Ind. Inf. Syst. (ICIIS)*, K. V. Arya, Ed. Piscataway NJ, USA: IEEE, Dec. 2014, pp. 1–6.

[37] T. Y. Ji, L. Liu, T. S. Wang, W. B. Lin, M. S. Li, and Q. H. Wu, "Nonintrusive load monitoring using additive factorial approximate maximum a posteriori based on iterative fuzzy C-means," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6667–6677, Apr. 2019.

[38] J. Z. Kolter and M. J. Johnson, "REDD: A public data set for energy disaggregation research," in *Proc. Workshop Data Mining Appl. Sustainability (SIGKDD)*, vol. 25. San Diego, CA, USA: Citeseer, 2011.

[39] D. Murray, L. Stankovic, and V. Stankovic, "An electrical load measurements dataset of united kingdom households from a two-year longitudinal study," *Sci. Data*, vol. 4, no. 1, pp. 1–12, Dec. 2017.

[40] V. Andrean, X.-H. Zhao, D. F. Teshome, T.-D. Huang, and K.-L. Lian, "A hybrid method of cascade-filtering and committee decision mechanism for non-intrusive load monitoring," *IEEE Access*, vol. 6, pp. 41212–41223, 2018.

[41] A. S. Bouhouras, P. A. Gkaidatzis, E. Panagiotou, N. Poulakis, and G. C. Christoforidis, "A NILM algorithm with enhanced disaggregation scheme under harmonic current vectors," *Energy Buildings*, vol. 183, pp. 392–407, Jan. 2019.

[42] W. Kong, Z. Y. Dong, J. Ma, D. J. Hill, J. Zhao, and F. Luo, "An extensible approach for non-intrusive load disaggregation with smart meter data," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3362–3372, Jul. 2018.

[43] M. J. Johnson and A. S. Willsky, "Bayesian nonparametric hidden semi-Markov models," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 673–701, 2013.

[44] P. A. Schirmer and I. Mporas, "Improving energy disaggregation performance using appliance-driven sampling rates," in *Proc. 27th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2019, pp. 1–5.

[45] B. Zhao, M. Ye, L. Stankovic, and V. Stankovic, "Non-intrusive load disaggregation solutions for very low-rate smart meter data," *Appl. Energy*, vol. 268, Jun. 2020, Art. no. 114949.

PASCAL A. SCHIRMER received the B.Eng. degree in electrical engineering from the University of Applied Sciences, Esslingen, Germany, in 2018, and the Ph.D. degree in electrical engineering from the University of Hertfordshire, U.K., in 2021. Since 2021, he has been working for the Research and Development Department for the development of power electronics at BMW AG, Munich, Germany, where he is responsible for the lifetime evaluation of power electronic systems. Furthermore, he is a Visiting Research Fellow with the University of Hertfordshire, where he is focusing on NILM (software and hardware applications) and a Visiting Lecturer for electro-mobility with TAE, Esslingen.

IOSIF MPORAS received the Diploma degree in electrical and computer engineering and the Ph.D. degree from the University of Patras, Greece, in 2004 and 2009, respectively. He has participated in more than ten EU-funded research and development projects as a Researcher, a Senior Researcher, and the principal investigator. He is a Reader in signal processing and machine learning (Associate Professor) with the University of Hertfordshire, U.K. He has authored or coauthored more than 100 articles in international journals and conferences cited more than 1,385 times (H-index: 18). His research interests include applications of signal processing and machine learning. He serves as a reviewer for grant applications and international journals and a Program Committee Member in international conferences, while he was the General Chair of the joint SPECOM/ICR 2017 Conference and the Technical Chair of the ICESF 2020 Conference.

● ● ●