Evaluation of Non-Intrusive Load Monitoring Algorithms on Industrial Load Profiles

Presented by: Sambhav Shah

17.02.202
3

Examiners

Prof. Dr. Hannah Bast
Prof. Dr. Christof Wittwer

Advisers

Matthias Hertel
Benedikt Köpfer

Contents

- Introduction
- Approach
- Evaluation

Introduction

- Rising energy prices have affected industries and households

- Load Monitoring refers to monitoring of various devices in a power network

- Real-time appliance level feedback can re energy savings of up to 12%

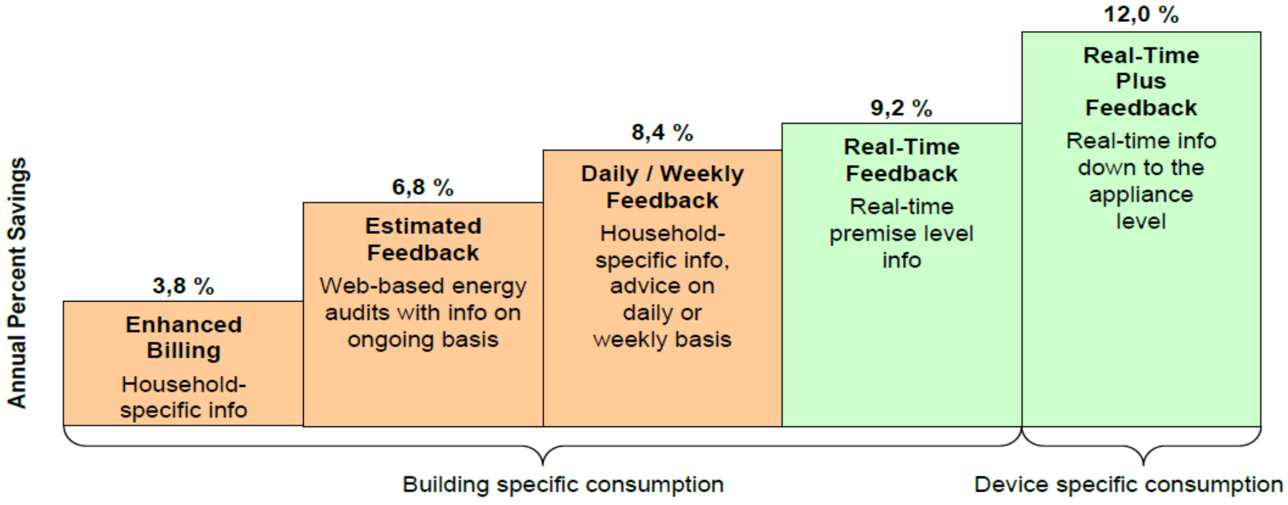


Figure. Energy savings due to advanced levels of feedback [1]

Non-Intrusive Load Monitoring (NILM)

- Main meter – aggregate consumption
- Submeter – consumption of individual devices
- **Intrusive Load Monitoring** requires installing submeters for individual appliances
- **Non-Intrusive Load Monitoring (NILM)** estimates the power readings by disaggregating the main meter readings
- The aggregated signal X_t at time *t* can be represented as the summation of the power of the constituent appliances Y_{it} at time *t* where ϵ_t is the e

$$X_t = \sum_{i=1}^{m} Y_{it} + \epsilon_t$$

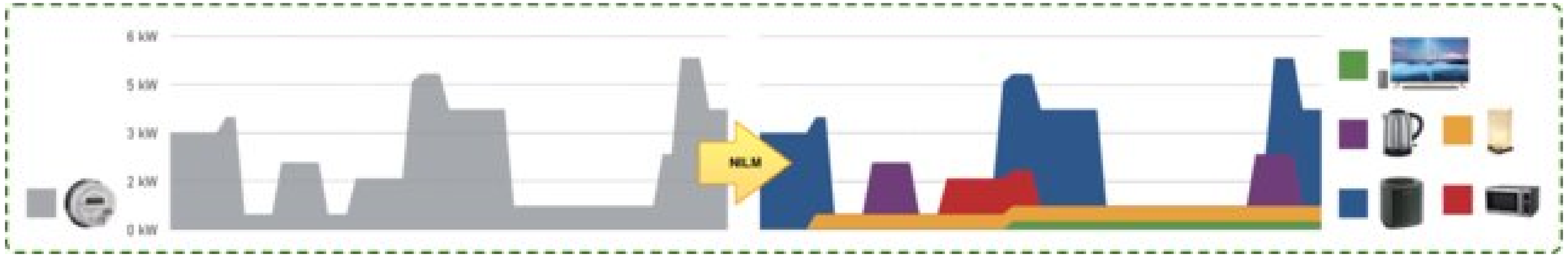


Figure. Sensing appliance power usage and consumption without sensors [2].

NILM for Industrial Data

- Most of the research work in NILM has been carried out in residential settings, fewer publicly available datasets for industrial use-case
- Households consume 26% of electrical energy, whereas the industrial sector is responsible for 44% of energy consumption[1]
- HIPE, High-resolution Industrial Production Energy data set [3] provides a comparative analysis using different NILM algorithms [4]
- Although, HIPE is smaller dataset and makes use of artificial aggregation
- Thus, we extend on the work provided in the HIPE paper

[1]https://de.statista.com/statistik/daten/studie/236757/umfrage/stromverbrauch-nach-sektorenin-deutschland/

Introduction

Questions ?

Contents

Fraunhofer ISE

Non-Intrusive Load Monitoring Toolkit (NILMTK)

- NILMTK is an open-source toolkit designed to promote research in NILM
- Allows comparative analysis with support for various publicly available datasets and well-known NILM algorithms
- Provides an end-to-end pipeline right from dataset conversion to evaluation metrics
- Easy integration of new datasets

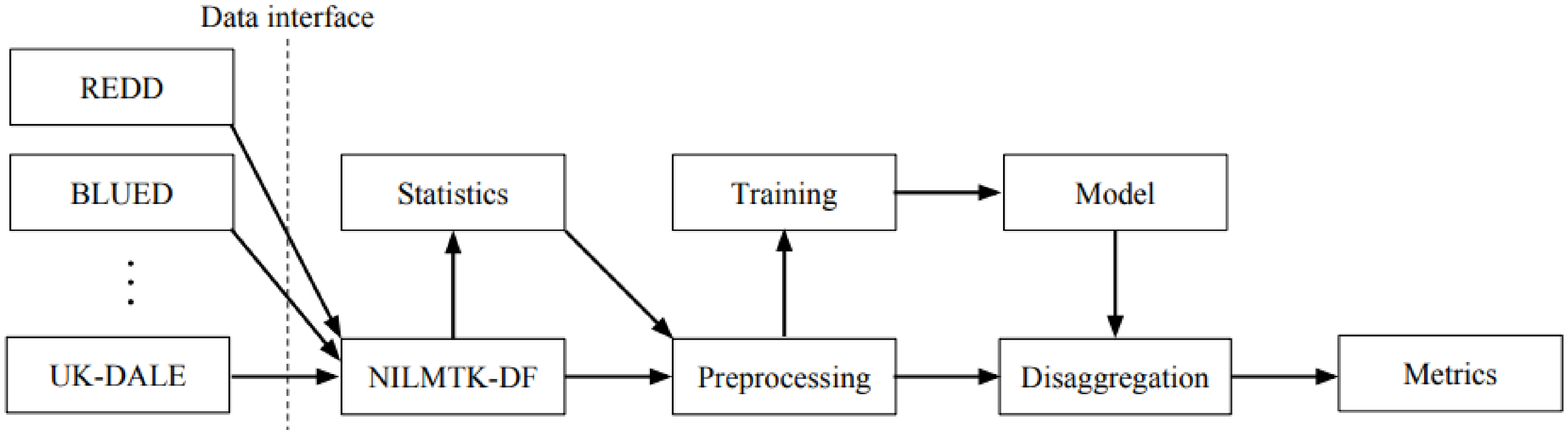


Figure. NILMTK workflow [5]

Seq2Seq and Seq2Point concepts

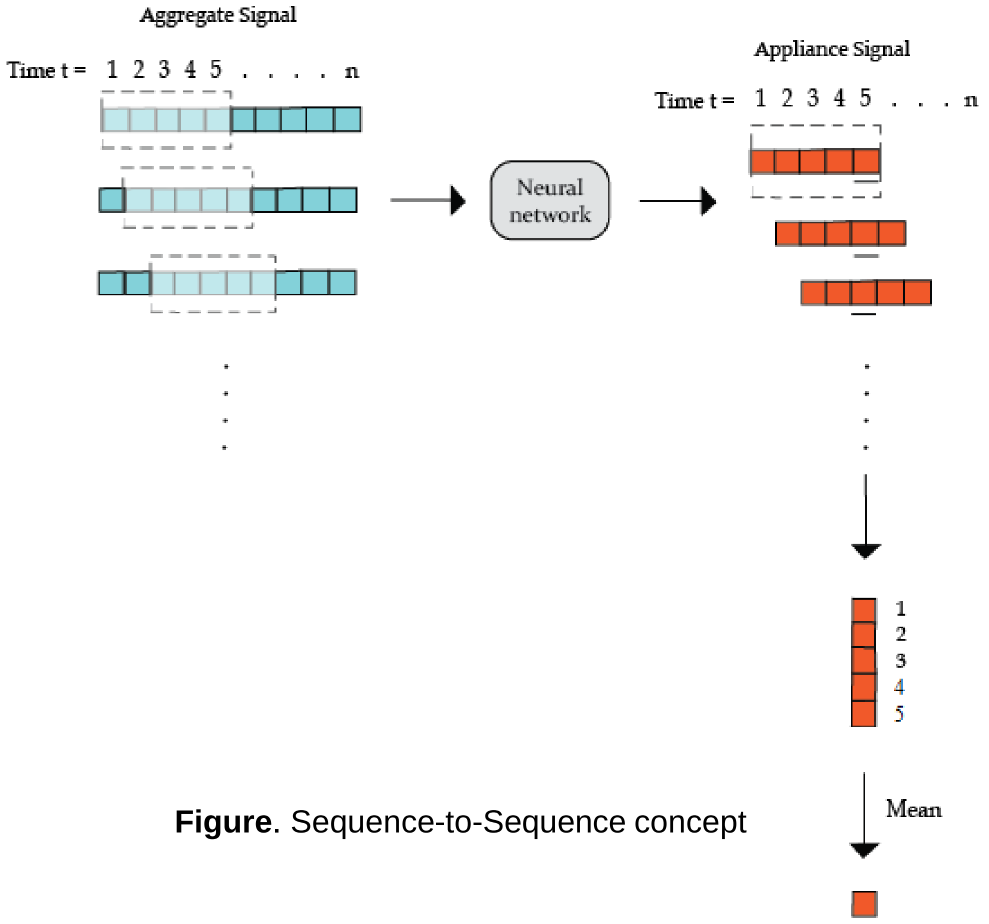


Figure. Sequence-to-Sequence concept

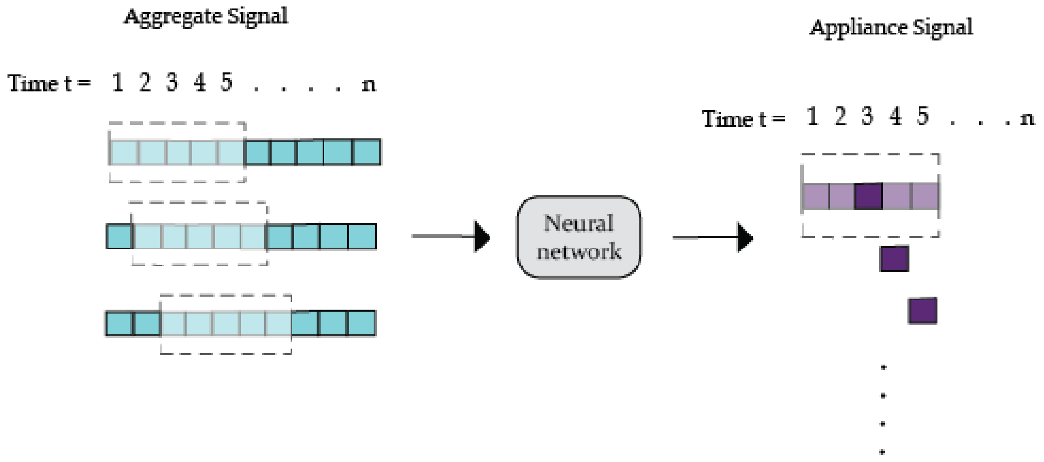
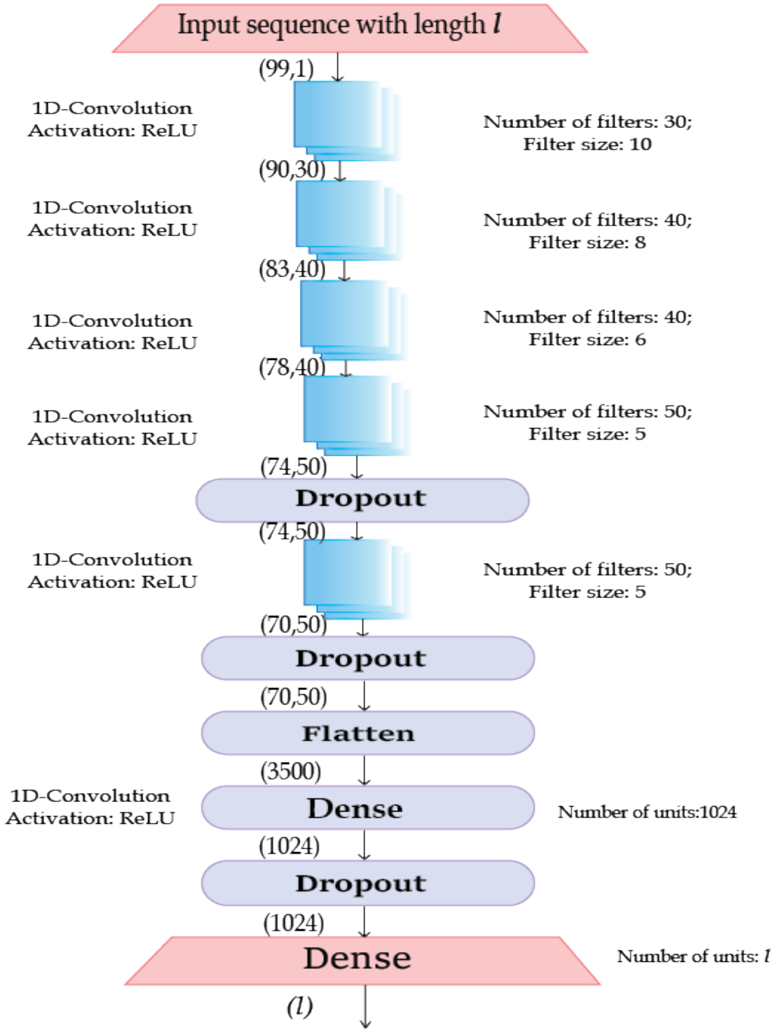


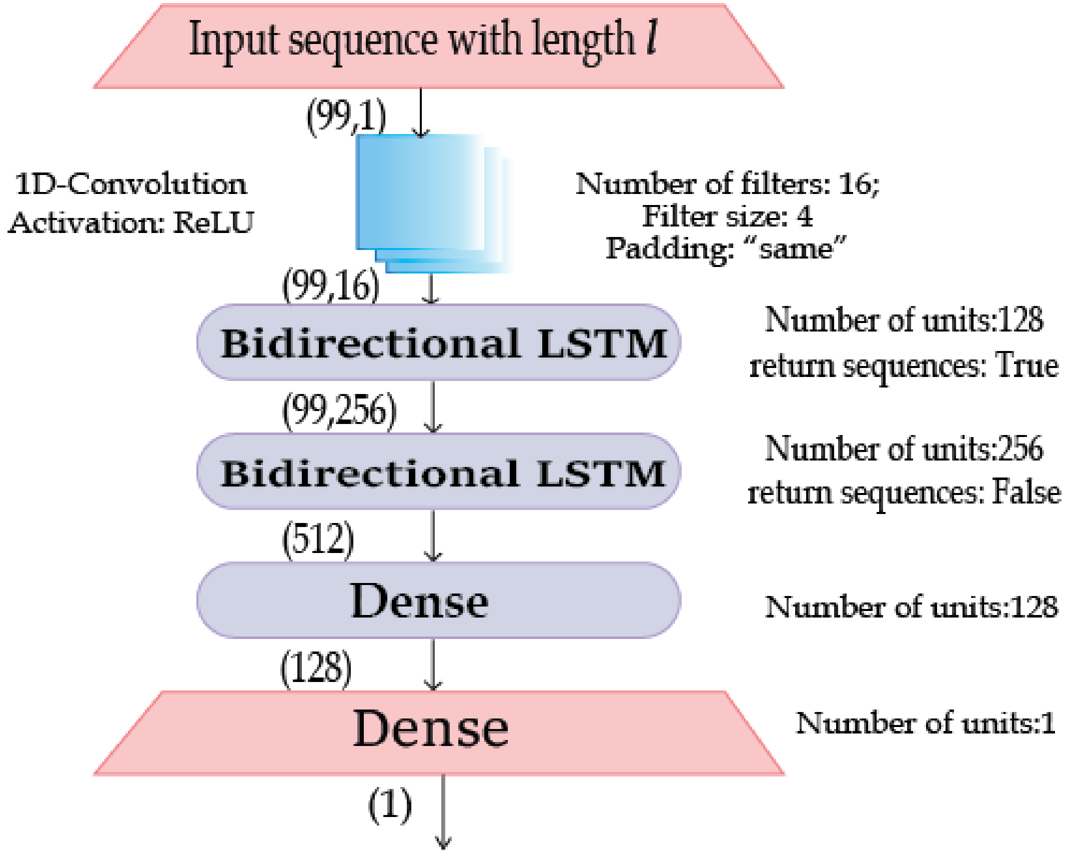
Figure. Sequence-to-Point concept

Fraunhofer ISE

CNN-Based Seq2Seq Architecture

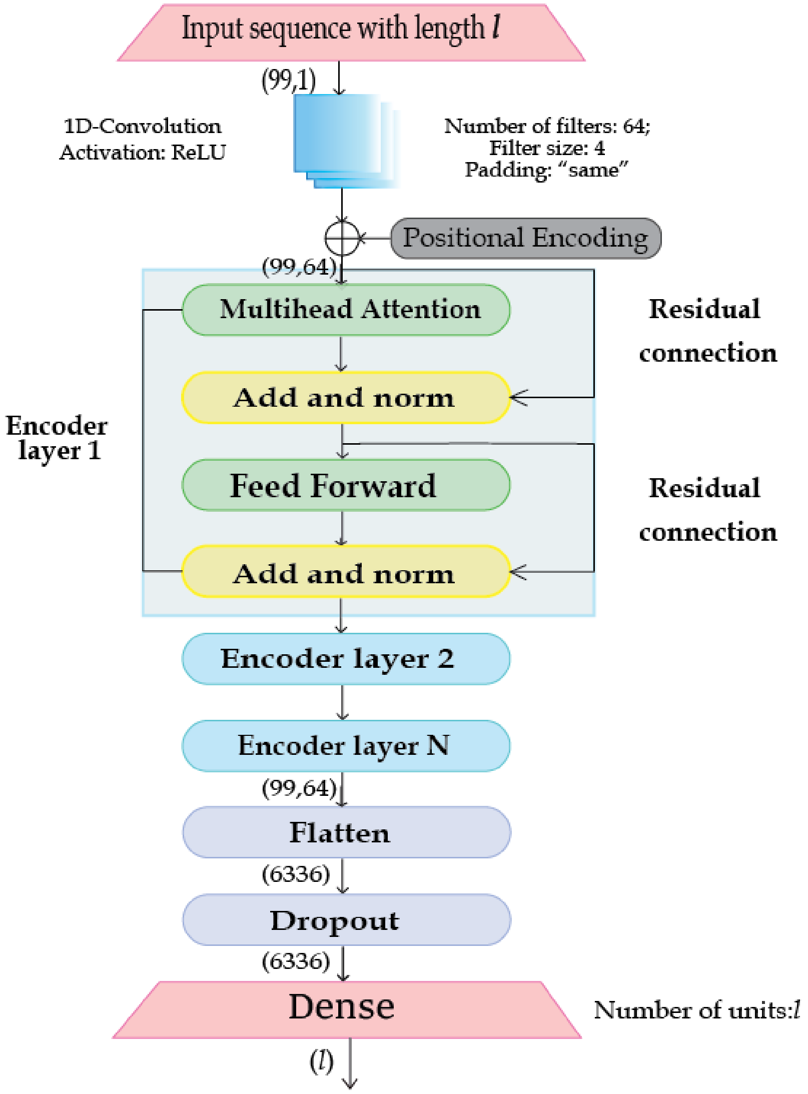


LSTM architecture



Fraunhofer ISE

BERT architecture



Fraunhofer ISE

Approach

Questions ?

Contents

Fraunhofer ISE

Datasets

- **Company A**, which is a plastic recycling company

- **Company B**, which produces micromechanical parts

Company A

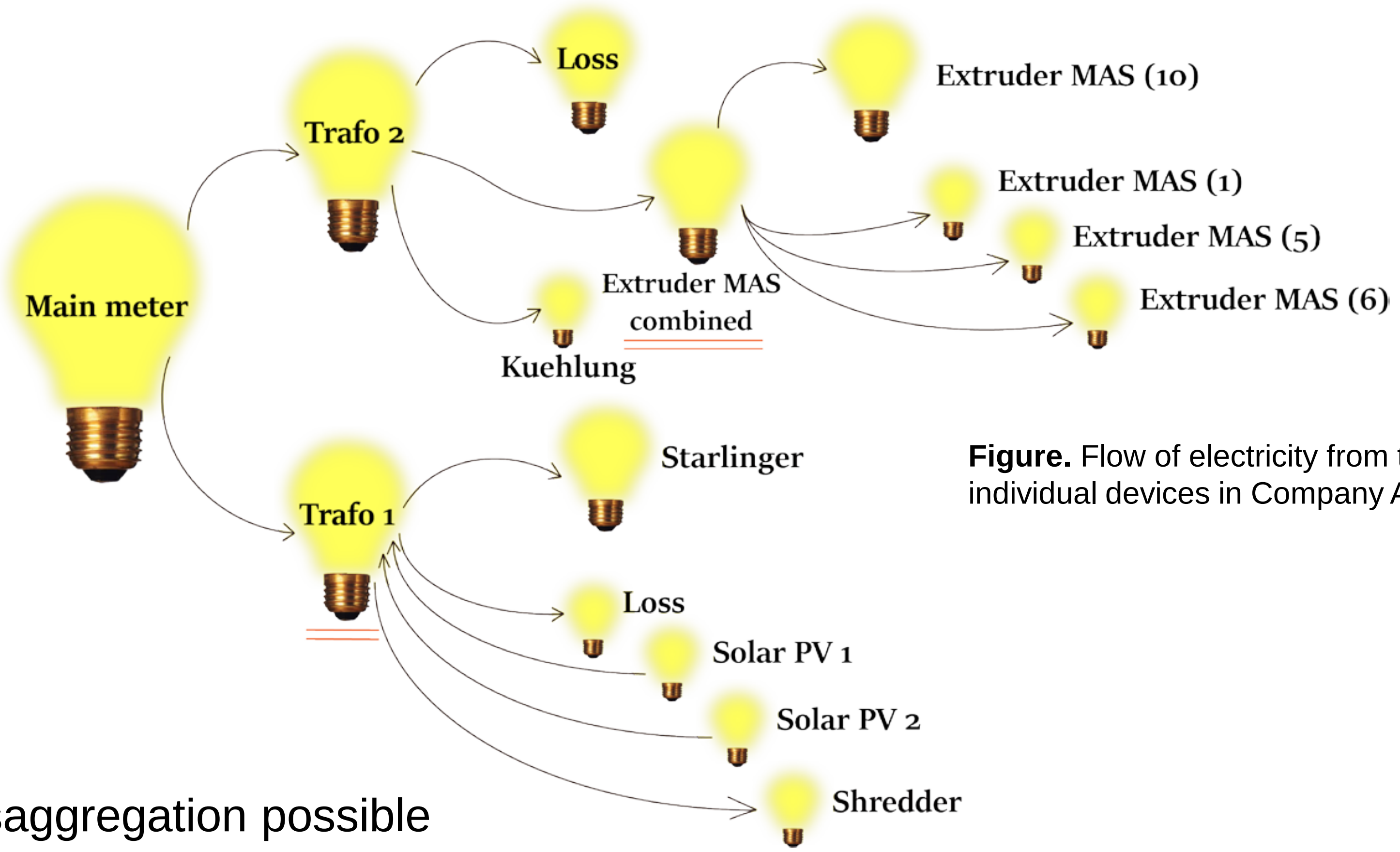


Figure. Flow of electricity from the main meter to the individual devices in Company A

- Various levels of disaggregation possible
- Underlined bulbs indicate the levels where disaggregation tasks are performed in the thesis

Fraunhofer ISE

Company A – Trafo1

- Disaggregation performed for Schredder and Starlinger
- Solar PV systems generate power
- Data sampled at a rate of 60 seconds per sample

- Starlinger consumes around 80% of energy
- Schredder only consumes around 5% of energy
- There is also a loss component

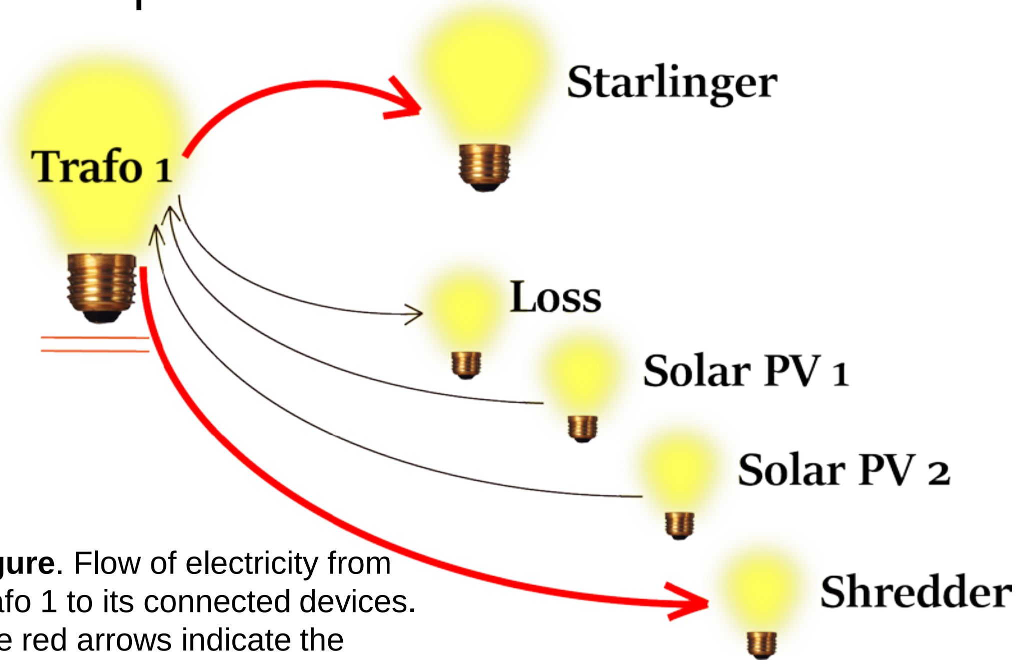


Figure. Flow of electricity from Trafo 1 to its connected devices. The red arrows indicate the target devices for the disaggregation task.

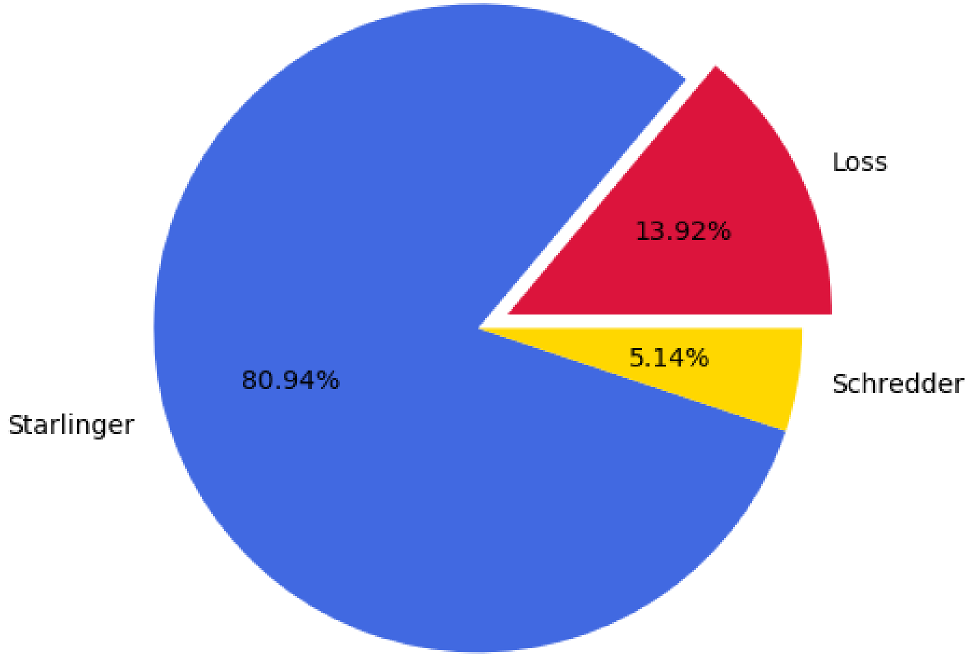


Figure. Energy composition of Trafo1 in Company A

Company A - Extruder

- Sample rate of 900 seconds
- Extruder MAS (10) is the highest consumer
- Starlinger is also an extruder, therefore this disaggregation task has been used for transfer learning

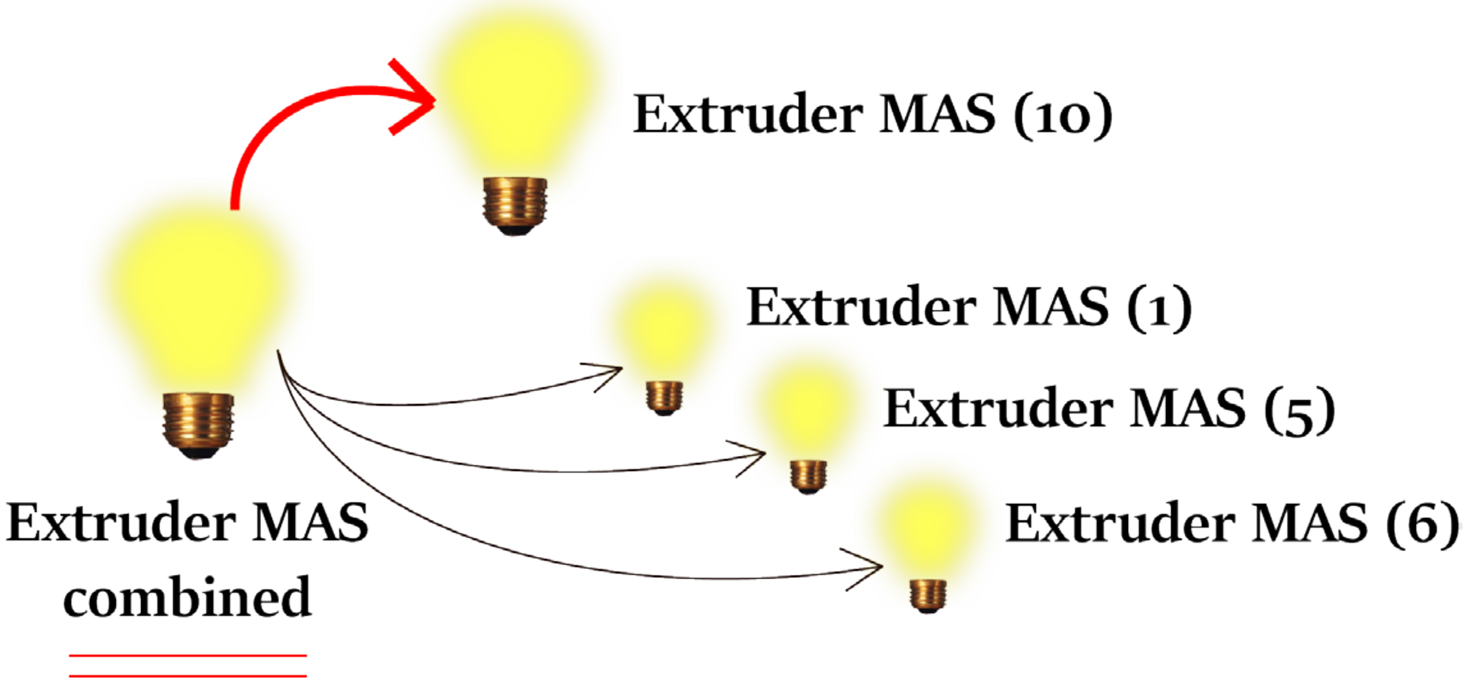


Figure. Flow of electricity from extruder to its connected devices. The red arrows indicate the target devices for the disaggregation task.

Fraunhofer ISE

Company B

- Primary focus on disaggregation of trafo1
- Sample rate of 60 seconds

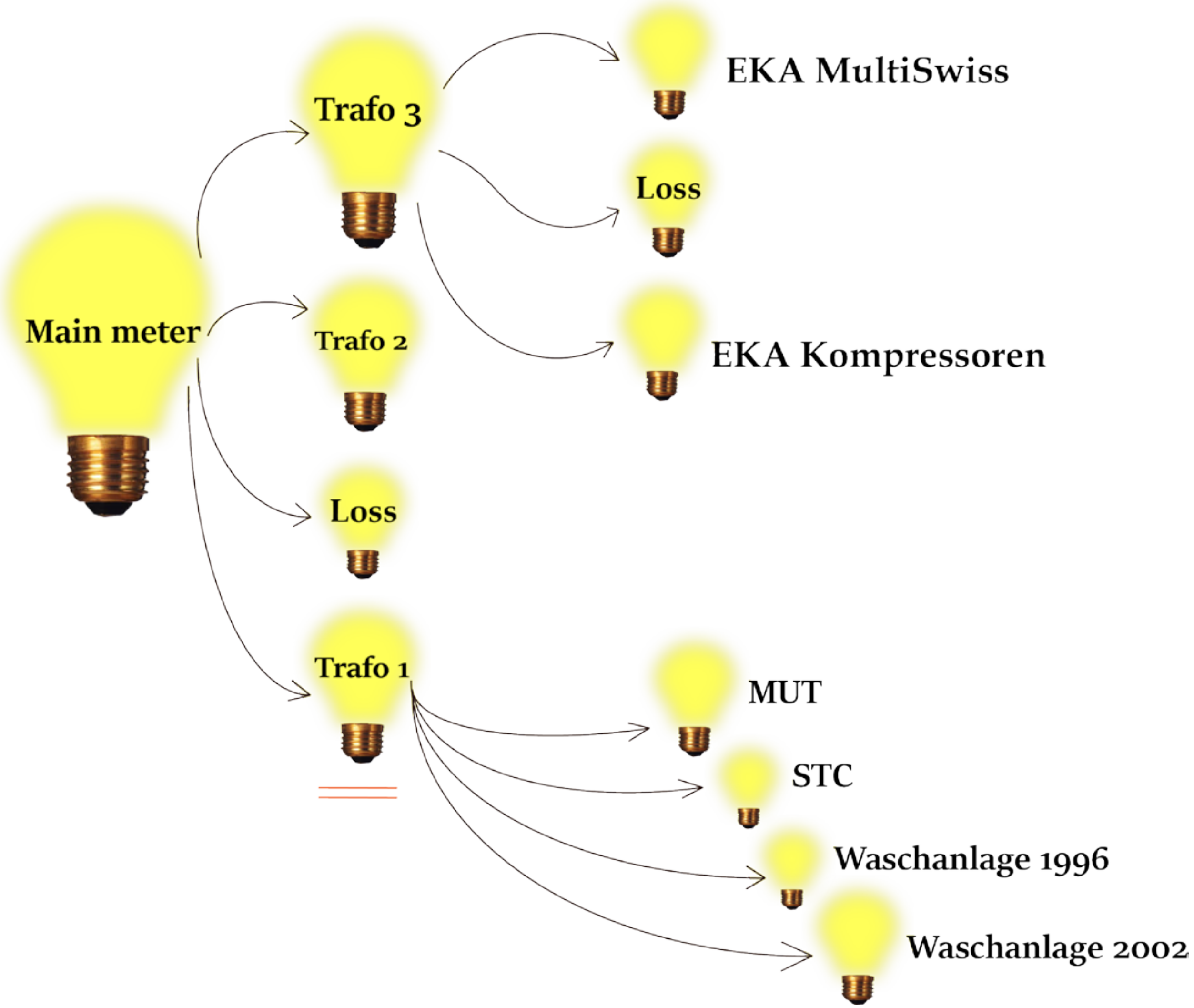


Figure. Flow of electricity from the main meter to the individual devices in Company B

Fraunhofer ISE

Company B – Trafo1

- Disaggregation performed for MUT and Waschanlage 2002 devices
- **MUT** and **WA 2002** consume around 85% of the energy

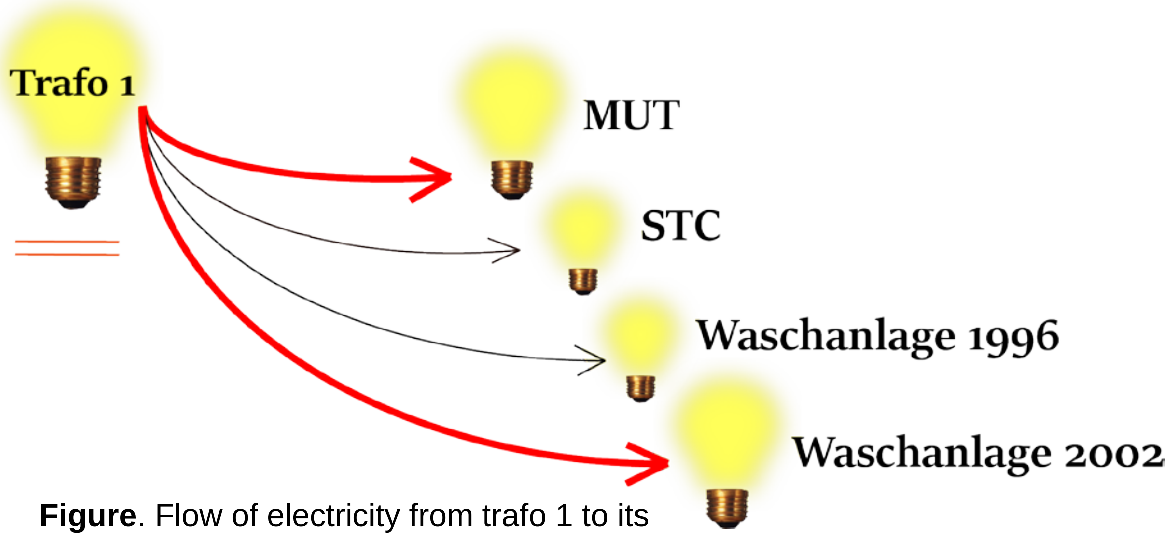


Figure. Flow of electricity from trafo 1 to its connected devices. The red arrows indicate the target devices for the disaggregation task.

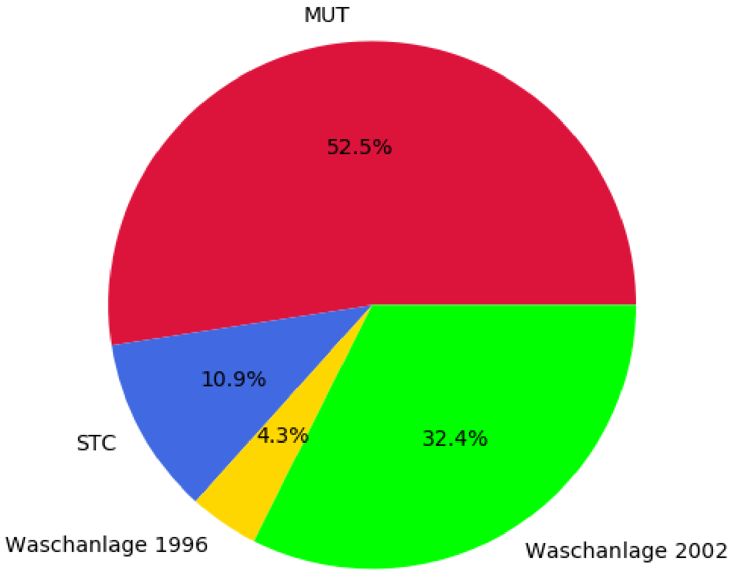


Figure. Energy consumption mix of trafo1 in Company B

Fraunhofer ISE

Setup

- Training period – 1 year
- Test period – 5 months
- Maximum number of epochs – 100, patience of 10
- Standardized input
- Implementation done using keras

Evaluation Metrics

- Root Mean Square Error (RMSE):

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \Sigma_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

- Normalized Disaggregation Error (NDE):

$$NDE(y, \hat{y}) = \sqrt{\frac{\Sigma_{i=1}^{n} (y_i - \hat{y}_i)^2}{\Sigma_{i=1}^{n} (y_i)^2}}$$

Comparing different input features

- Sample rate – 300, Sequence length – 99
- Error metric – RMSE
- *Active power, Reactive power* performs the best
- Similar trends observed for Starlinger, MUT and Waschanlage devices

Features	BERT	Seq2Point	LSTM
Active	10731	11768	15960
Active, Reactive	**7818**	**9014**	**8775**
Active, Reactive, Solar Power	8225	9420	9437
Active, Reactive, Voltage, Current	9060	10990	11322

Table. RMSE results (measured in Watts) of different NILM algorithms on the Waschanlage 2002 device from Company B compared on different input features.

Fraunhofer ISE

Comparison between different sample rates and sequence lengths

- Effective sequence length (ESL) – product of sample rate and sequence length
- BERT is the best performing algorithm for all 3 devices
- LSTM is the worst performing, especially at higher sequence lengths
- Results are worse at sample rate of 60
- Better results obtained at lower sequence lengths for higher sample rates and vice versa

Sample Rate	Sequence Length	ESL (minutes)	BERT	Seq2Point	LSTM
900	39	585	6828	7683	**7959**
	99	1485	**6752**	**7341**	9059
	159	2385	6980	7361	11733
	219	3285	7174	7767	14936

Table. Snapshot of comparing different algorithms for MUT device showing results at sample rate of 900

Fraunhofer ISE

Comparison between different sample rates and sequence lengths

- Effective sequence length (ESL) – product of sample rate and sequence length
- BERT is the best performing algorithm for all 3 devices
- LSTM is the worst performing, especially at higher sequence lengths
- Results are worse at sample rate of 60
- Better results obtained at lower sequence lengths for higher sample rates and vice versa

Sample Rate	Sequence Length	ESL (minutes	LSTM
900	39	585	**7959**
	99	1485	9059
	159	2385	11733
	219	3285	14936
300	39	195	**8829**
	99	495	8909
	159	795	10825
	219	1095	10631
180	39	117	**9594**
	99	297	9596
	159	477	10542
	219	657	11281
60	39	39	10820
	99	99	**10479**
	159	159	10989
	219	219	11450

Table. Snapshot of comparing different algorithms for MUT device showing LSTM results

Comparison between different sample rates and sequence lengths

- Effective sequence length (ESL) – product of sample rate and sequence length
- BERT is the best performing algorithm for all 3 devices
- LSTM is the worst performing, especially at higher sequence lengths
- Results are worse at sample rate of 60
- Better results obtained at lower sequence lengths for higher sample rates and vice versa

Sample Rate	Sequence Length	ESL (minutes)	BERT	Seq2Point
900	39	585	7149	8734
	99	1485	**6883**	**8290**
	159	2385	6989	8507
	219	3285	7132	8788
300	39	195	8207	9540
	99	495	7860	9101
	159	795	**7432**	**8882**
	219	1095	7802	9041
180	39	117	9126	10758
	99	297	8536	9625
	159	477	8181	9217
	219	657	**7986**	**9039**
60	39	39	9043	11624
	99	99	8666	10308
	159	159	8701	10163
	219	219	**8313**	**9855**

Table. Snapshot of comparing different algorithms for WA 2002 device showing BERT and Seq2Point results

Comparison between different devices

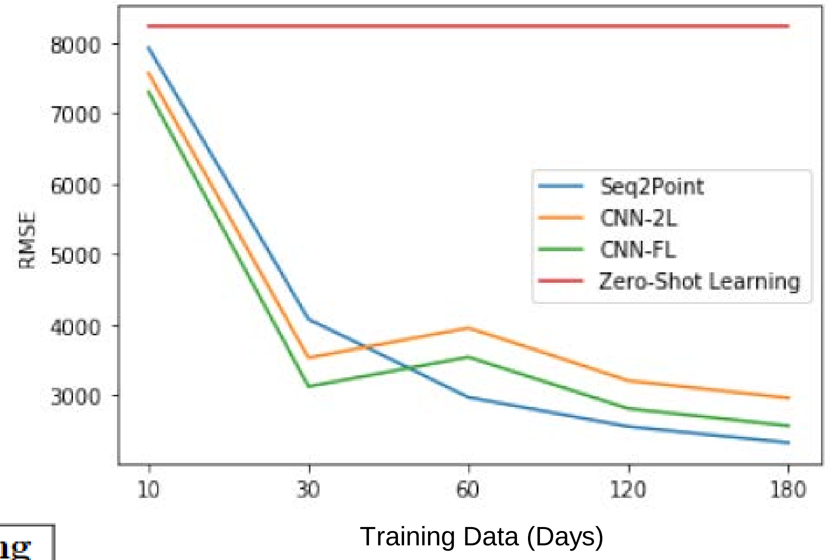
- NDE – Evaluation Metric
- Starlinger consumes ~ 80% of energy
- Schredder consumes only ~ 5% of energy

Sequence Length	Starlinger		Schredder		MUT		Waschanlage	
	BERT	Seq2Point	BERT	Seq2Point	BERT	Seq2Point	BERT	Seq2Point
39	0.056	0.071	0.812	0.954	0.154	0.173	0.168	0.201
99	0.053	0.069	0.716	0.848	**0.148**	0.168	0.159	0.188
159	**0.052**	0.066	**0.682**	0.769	0.151	0.161	**0.152**	0.180
219	0.057	0.067	0.757	0.842	0.153	0.163	0.156	0.185

Table. Results of different NILM algorithms on 4 different devices from Company A and Company B

Transfer Learning

- Transfer learning in residential setting has shown promising results [
- Pre-trained model on Starlinger used for Extruder MAS (10)
- CNN-2L – leaving last two layers unfrozen
- CNN-FL – leaving all the layers unfrozen
- Transfer learning only beneficial for when little training data available
- Results of CNN-FL, especially, indicate some negative transfer



Training Data (Days)	Seq2Point	CNN-2L	CNN-FL	Zero-Shot Learning
10	7928	7568	**7303**	8227
30	4082	3531	**3125**	8227
60	**2972**	3958	3542	8227
120	**2558**	3207	2812	8227
180	**2330**	2963	2570	8227

Table. Comparing the results of transfer learning with ‘normal’ learning

Fraunhofer ISE

Conclusion

- Better results obtained as compared to the HIPE results (evaluated using NDE)
- Using active and reactive power yields better results than simply using active power
- BERT outperforms other NILM algorithms on all the devices
- Choice of sequence length must be made in accordance with sample rate
- Additional experiments can help in finding the optimal sequence length
- More effective transfer learning strategies needed
- Disaggregation can be performed at various levels in the power network

References

1. K. Ehrhardt-Martinez, K. A. Donnelly, S. Laitner, et al., “Advanced metering initiatives and residential feedback programs: a meta-review for household electricity-saving opportunities,” American Council for an Energy-Efficient Economy Washington, DC, 2010.
2. Rodriguez, Alejandro & Makonin, Stephen. (2019). Universal Non-Intrusive Load Monitoring (UNILM) Using Filter Pipelines, Probabilistic Knapsack, and Labelled Partition Maps. 1-6. 10.1109/APPEEC45492.2019.8994618.
3. S. Bischof, H. Trittenbach, M. Vollmer, D. Werle, T. Blank, and K. Böhm, “Hipe An energy-status-data set from industrial production,” in Proceedings of th Ninth International Conference on Future Energy Systems, pp. 599–603, 2018.
4. F. Kalinke, P. Bielski, S. Singh, E. Fouché, and K. Böhm, “An evaluation of nilm approaches on industrial energy-consumption data,” in Proceedings of the Twelfth ACM International Conference on Future Energy Systems, pp. 239–243, 2021.
5. N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh and M. Srivastava, “Nilmtk: An open source toolkit for non-intrusive load monitoring,” in Proceedings of the 5th international conference on Future energy systems, pp. 265–276, 2014.
6. M. D’Incecco, S. Squartini, and M. Zhong, “Transfer learning for non-intrusive load monitoring,” IEEE Transactions on Smart Grid, vol. 11, no. 2, pp. 1419–1429, 2019.
7. L. Wang, S. Mao, B. M. Wilamowski, and R. M. Nelms, “Pre-trained models for non-intrusive appliance load monitoring,” IEEE Transactions on Green Communications and Networking, vol. 6, no. 1, pp. 56–68, 2021.

Thank You!

Questions?

Additional Slides

Electricity prices over the years

Fraunhofer ISE

Formal definition of NILM

- Observed aggregate time series $X = (X_1, X_2, \ldots , X_T)$

- m appliances, each represented by $Y_i = (Y_{i1}, Y_{i2}, \ldots , Y_{iT})$ where $1 <= i <= m$

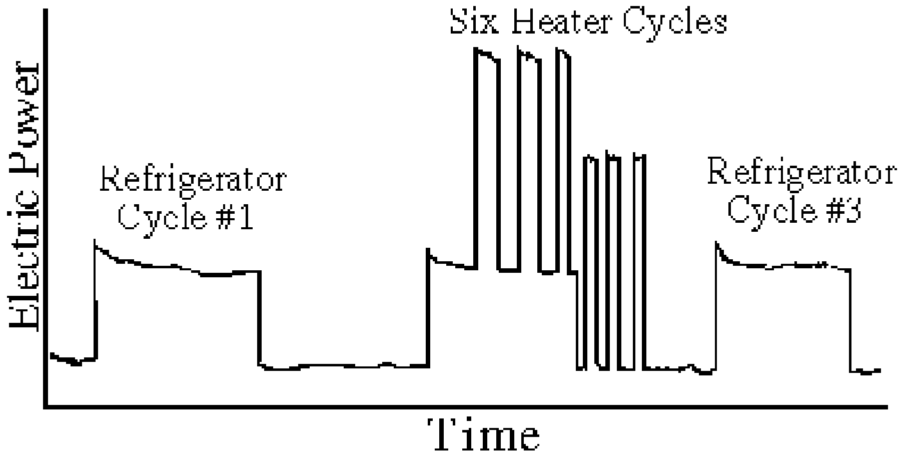
- The aggregated signal X_t at time *t* can be represented as the summation of the power measured of the constituent appliances at time *t* where ϵ_t is the error at time *t*.

$$X_t = \sum_{i=1}^{m} Y_{it} + \epsilon_t$$

- The goal is to estimate the unknown signals Y_i given the aggregate signal X

Fraunhofer ISE

HART


Six Heater Cycles
Electric Power
Refrigerator
Cycle #1
Refrigerator
Cycle #3
Time

Training Time

Model	Total training time on average (s)
Seq2Point	200
Seq2Seq	200
LSTM	1900
BERT with 6 encoder layers	2500
BERT2Point	2400

Table 1: Training time for different algorithms at sample rate of 300 and sequence length of 99.

Algorithm	Number of epochs	Average time taken per epoch	Average total training time
Seq2seq	50	1 Second	50 Seconds
Seq2point	50	1 Second	50 Seconds
RNN	50	10 Seconds	500 Seconds
GRU	50	67 Seconds	3350 Seconds
BERT with 1 encoder layer	50	24 Seconds	4800 Seconds
BERT with 4 encoder layers	50	47 Seconds	9400 Seconds
BERT with 6 encoder layers	50	71 Seconds	14200 Seconds

Table. Training time for Bert without improvements

Comparison of different input features

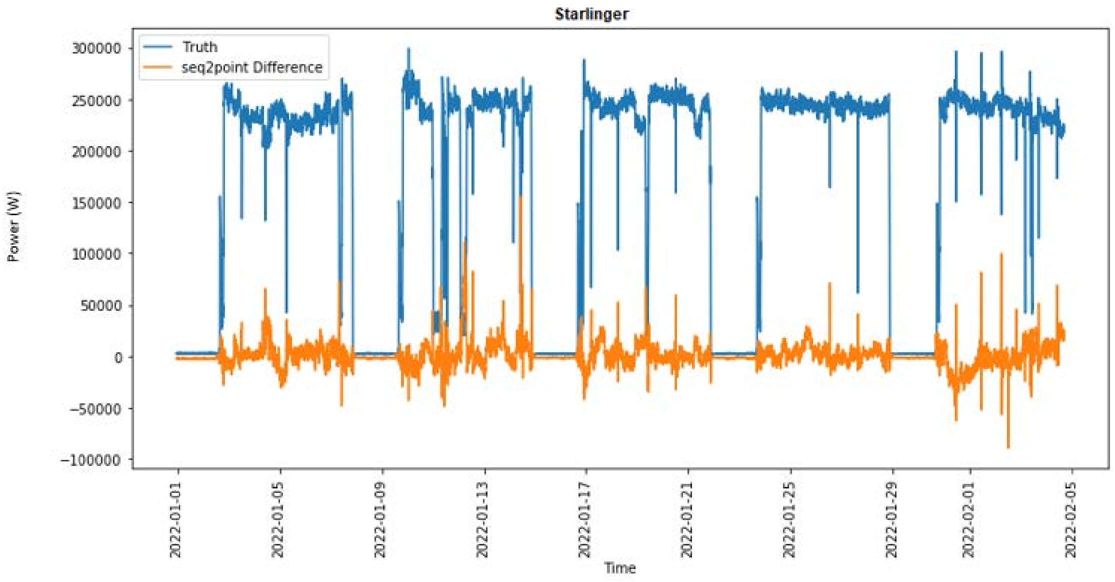
Features	BERT	Seq2Point	LSTM
Active	15123	20220	21893
Active, Reactive	**10242**	**12566**	**16140**
Active, Reactive, Solar Power	10685	15371	17189
Active, Reactive, Voltage, Current	15632	22311	25526

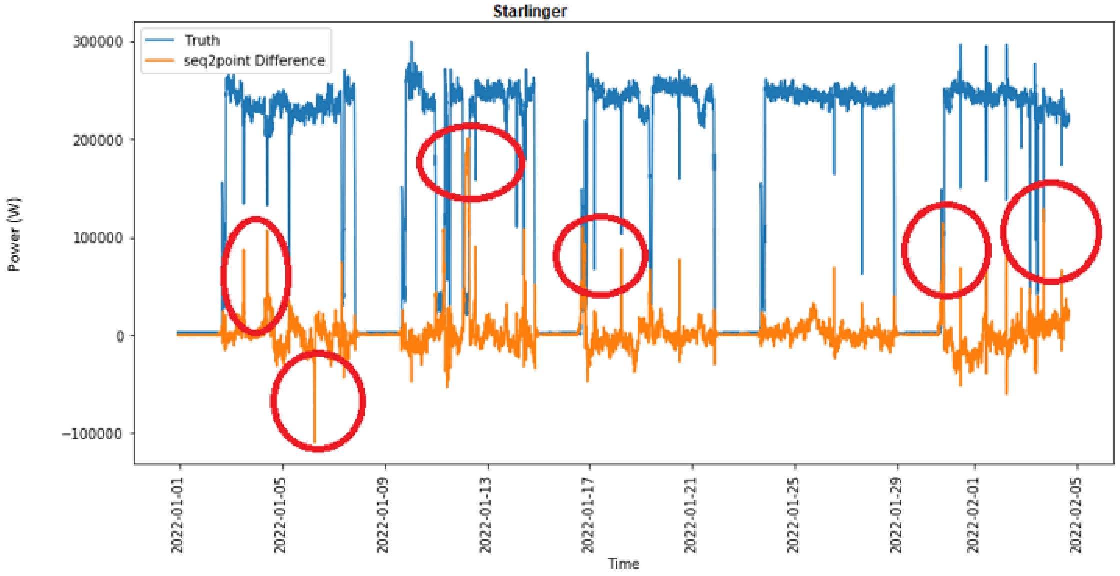
Table 2: RMSE results (measured in Watts) of different NILM algorithms on the Starlinger device from Company A compared on different input features.

Features	BERT	Seq2Point	LSTM
Active	8945	9759	11594
Active, Reactive	**7341**	**8363**	**8380**
Active, Reactive, Solar Power	7727	9183	10627
Active, Reactive, Voltage, Current	8051	10368	12068

Table 3: RMSE results (measured in Watts) of different NILM algorithms on the MUT device from Company B compared on different input features.

Comparing active and reactive power





Fraunhofer
ISE

Comparison between different sequence lengths and sample rates

Sample Rate	Sequence Length	ESL (minutes)	BERT	Seq2Point	LSTM
900	39	585	7149	8734	**8616**
	99	1485	**6883**	**8290**	8729
	159	2385	6989	8507	14402
	219	3285	7132	8788	20151
300	39	195	8207	9540	9571
	99	495	7860	9101	**8743**
	159	795	**7432**	**8882**	14235
	219	1095	7802	9041	14121
180	39	117	9126	10758	9537
	99	297	8536	9625	**9067**
	159	477	8181	9217	9260
	219	657	**7986**	**9039**	14749
60	39	39	9043	11624	9780
	99	99	8666	10308	**9347**
	159	159	8701	10163	9721
	219	219	**8313**	**9855**	9725

Table. Results for WA 2002

Sample Rate	Sequence Length	ESL (minutes	BERT	Seq2Point	LSTM
900	39	585	6828	7683	**7959**
	99	1485	**6752**	**7341**	9059
	159	2385	6980	7361	11733
	219	3285	7174	7767	14936
300	39	195	7730	8387	**8829**
	99	495	**7474**	8203	8909
	159	795	7651	**8082**	10825
	219	1095	7712	8294	10631
180	39	117	8334	9841	**9594**
	99	297	8106	9255	9596
	159	477	**7994**	9082	10542
	219	657	8027	**8979**	11281
60	39	39	9880	11584	10820
	99	99	9310	10210	**10479**
	159	159	9253	10197	10989
	219	219	**9215**	**10032**	11450

Table. Results for MUT

Sample Rate	Sequence Length	ESL (minutes)	BERT	Seq2Point	LSTM
900	39	585	11789	13869	**17112**
	99	1485	11892	**13658**	21347
	159	2385	**11751**	14123	33645
	219	3285	12136	14105	35527
300	39	195	11347	13928	**14644**
	99	495	10615	13664	16477
	159	795	**10311**	**13021**	23316
	219	1095	11182	13299	24482
180	39	117	11184	12955	**13677**
	99	297	10598	12485	15662
	159	477	10338	12713	22176
	219	657	**10287**	**12157**	21868
60	39	39	12165	14450	13520
	99	99	12208	13324	**12880**
	159	159	11360	13105	15078
	219	219	**11140**	**12788**	17366

Table. Results for Starlinger

Trainable Parameters

```
Layer (type) Output Shape Param #
=================================================================
conv1d_77 (Conv1D) (None, 99, 16) 144
_________________________________________________________________
bidirectional_10 (Bidirectio (None, 99, 256) 148480
_________________________________________________________________
bidirectional_11 (Bidirectio (None, 512) 1050624
_________________________________________________________________
dense_190 (Dense) (None, 128) 65664
_________________________________________________________________
dense_191 (Dense) (None, 1) 129
=================================================================
Total params: 1,265,041 Trainable params: 1,265,041 Non-trainable
params: 0
```

Trainable parameters for RNN

```
Layer (type) Output Shape Param #
=================================================================
conv1d_85 (Conv1D) (None, 99, 64) 576
_________________________________________________________________
encoder_15 (Encoder) (None, 99, 64) 232320
_________________________________________________________________
flatten_27 (Flatten) (None, 6336) 0
_________________________________________________________________
dropout_243 (Dropout) (None, 6336) 0
_________________________________________________________________
dense_251 (Dense) (None, 99) 627363
=================================================================
Total params: 860,259 Trainable params: 860,259 Non-trainable
params: 0
```

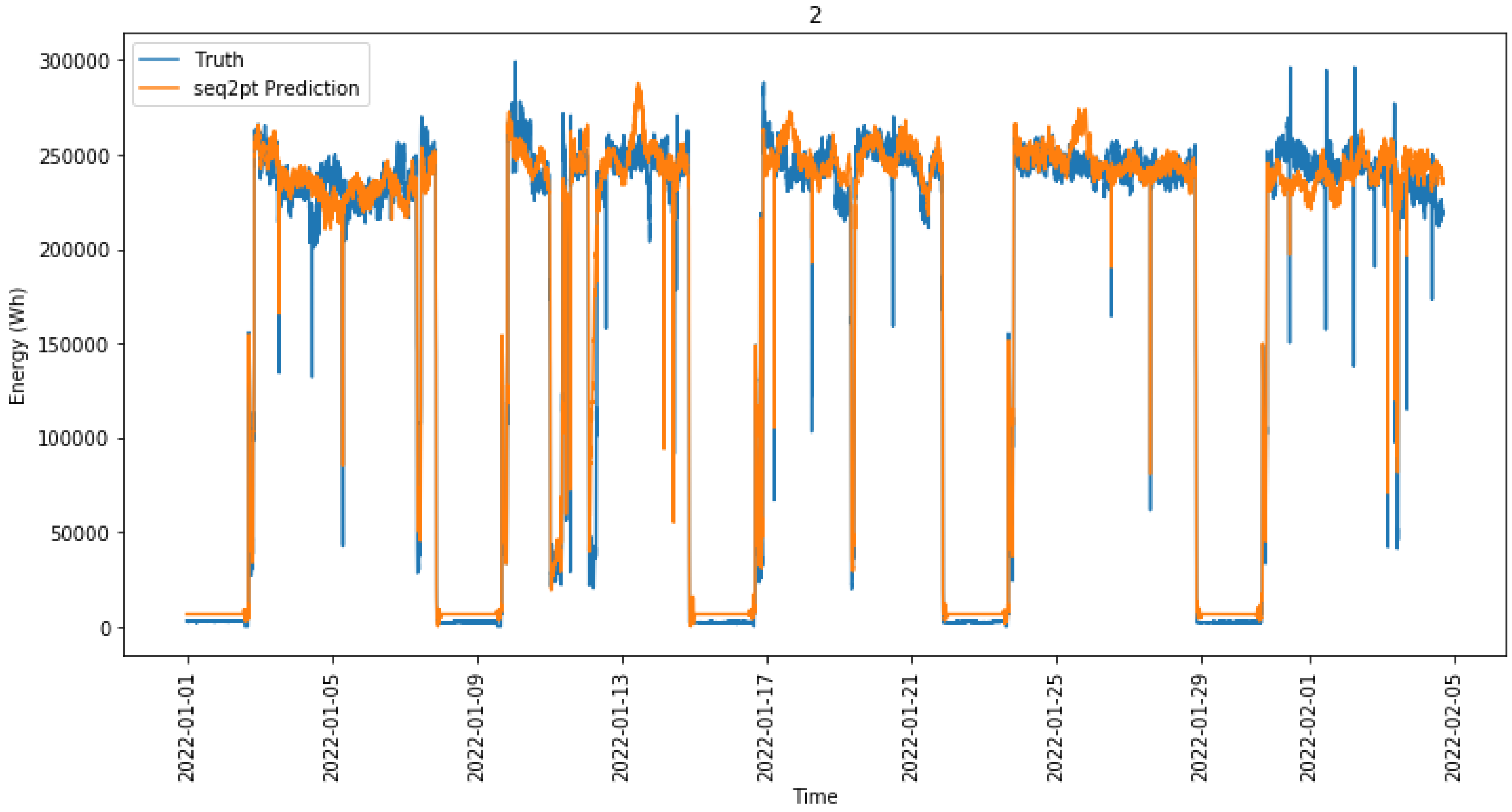
Trainable parameters for BERT

Trainable Parameters

```
Layer (type) Output Shape Param #
=================================================================
conv1d_10 (Conv1D) (None, 90, 30) 630
_________________________________________________________________
conv1d_11 (Conv1D) (None, 83, 30) 7230
_________________________________________________________________
conv1d_12 (Conv1D) (None, 78, 40) 7240
_________________________________________________________________
conv1d_13 (Conv1D) (None, 74, 50) 10050
_________________________________________________________________
dropout_6 (Dropout) (None, 74, 50) 0
_________________________________________________________________
conv1d_14 (Conv1D) (None, 70, 50) 12550
_________________________________________________________________
dropout_7 (Dropout) (None, 70, 50) 0
_________________________________________________________________
flatten_2 (Flatten) (None, 3500) 0
_________________________________________________________________
dense_4 (Dense) (None, 1024) 3585024
_________________________________________________________________
dropout_8 (Dropout) (None, 1024) 0
_________________________________________________________________
dense_5 (Dense) (None, 1) 1025
=================================================================
Total params: 3,623,749 Trainable params: 3,623,749 Non-trainable
params: 0
```

Trainable parameters for CNN-based models

Prediction plot



Prediction plot for starlinger

Comparison between Seq2Point and Seq2Seq algorithms

- Comparing the performance of BERT2Point and Seq2Point vs BERT and Seq2Seq
- Seq2Seq methods fare better on Starlinger
- Seq2Point methods work comparatively better with MUT

Sequence Length	Starlinger				MUT			
	BERT	BERT2Point	Seq2Seq	Seq2Point	BERT	BERT2Point	Seq2Seq	Seq2Point
39	0.056	0.0663	0.064	0.071	0.154	0.163	0.174	0.173
99	0.0529	0.0628	0.0612	0.069	0.148	0.155	0.171	0.168
159	**0.0523**	**0.0605**	0.0626	0.066	0.151	0.156	0.167	0.161
219	0.057	0.0609	0.0646	0.067	0.153	0.157	0.170	0.163

Table. Comparison of results of Seq2Point and Seq2Seq algorithms

Algorithms	Starlinger	MUT
BERT vs BERT2point	**14.9%**	**4.12%**
Seq2Seq vs Seq2Point (CNN)	**7.47%**	**-2.62%**

Table. Mean Error difference

Comparison between Seq2Point and Seq2Seq algorithms

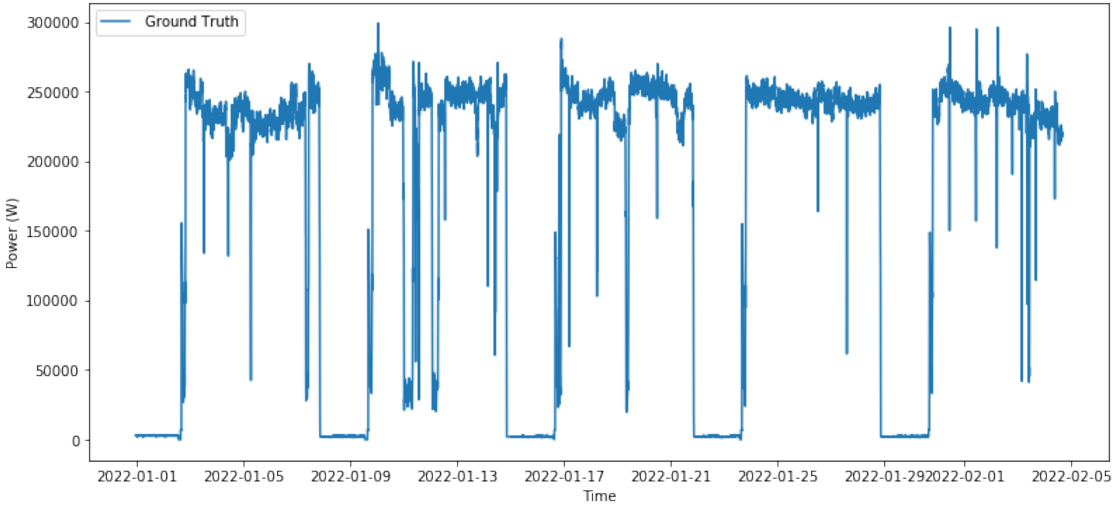


Figure. The ground truth shows less sharper peaks for Starlinger

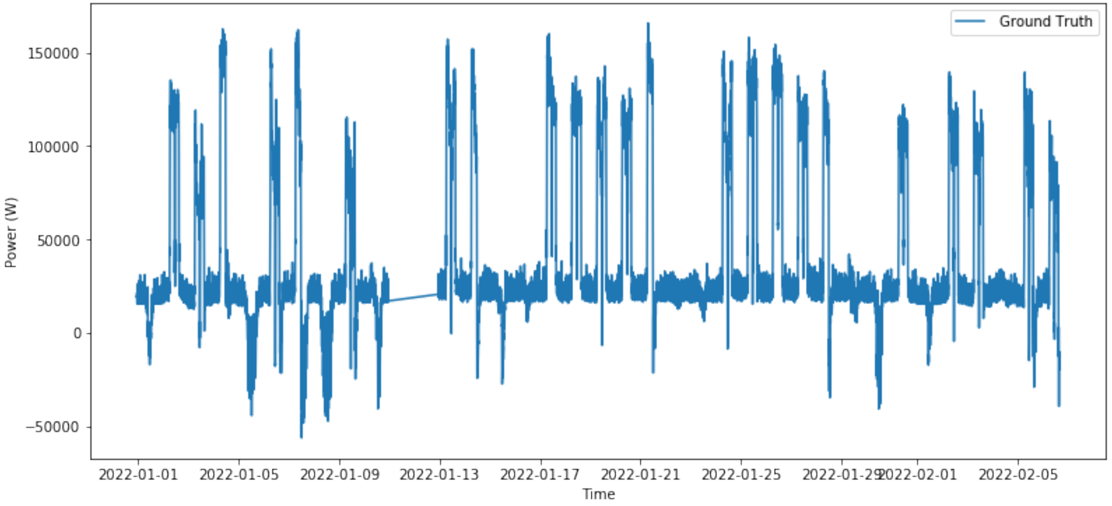


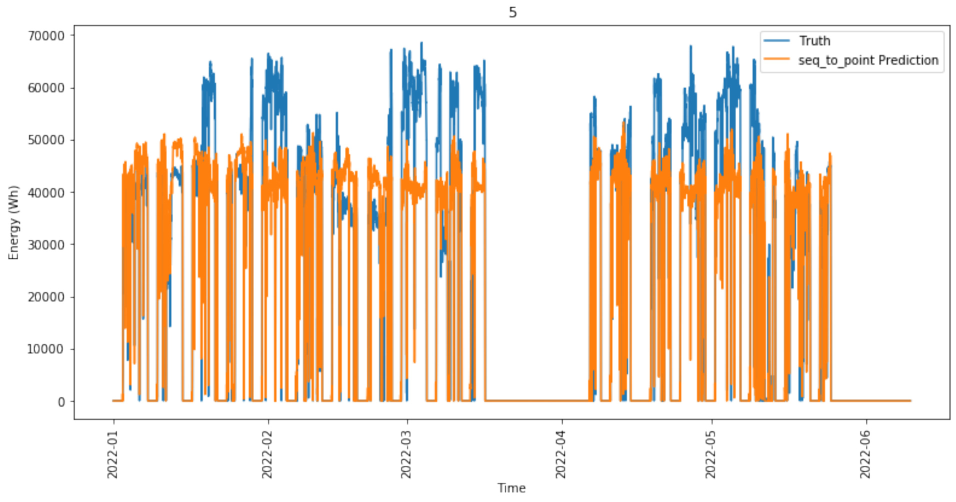
Figure. The ground truth shows sharper peaks for MUT

Hyperparameter Optimisation

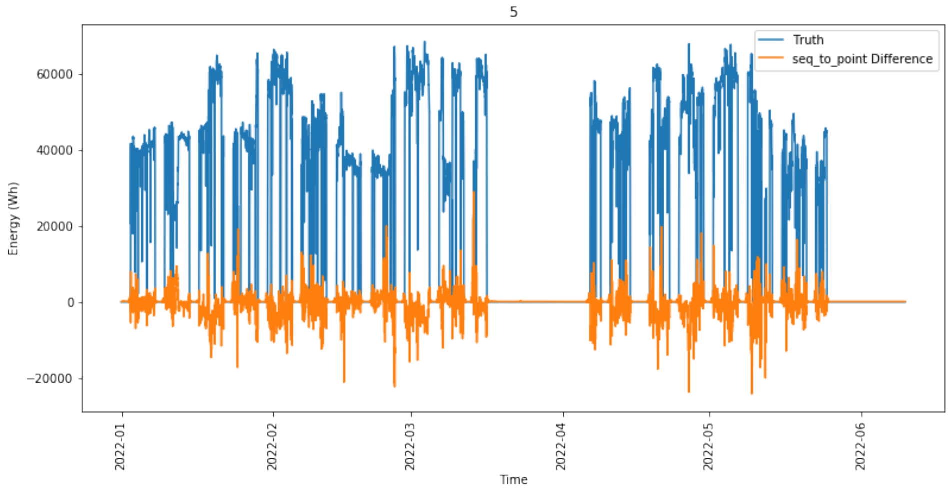
Description	Range	Best Configuration
Learning Rate	$[1 \cdot 10^{-4}, 5 \cdot 10^{-4}, 1 \cdot 10^{-3}, 5 \cdot 10^{-2}, 1 \cdot 10^{-2}]$	$1 \cdot 10^{-3}$
Number of Enc. Layers	$[1, 2, 4, 6]$	6
No. of filters	$[8, 16, 32, 64]$	64
No. of attention heads	$[1, 2]$	2
Dropout rate	$[0.1, 0.2]$	0.2

Table. Hyperparameters and their search space

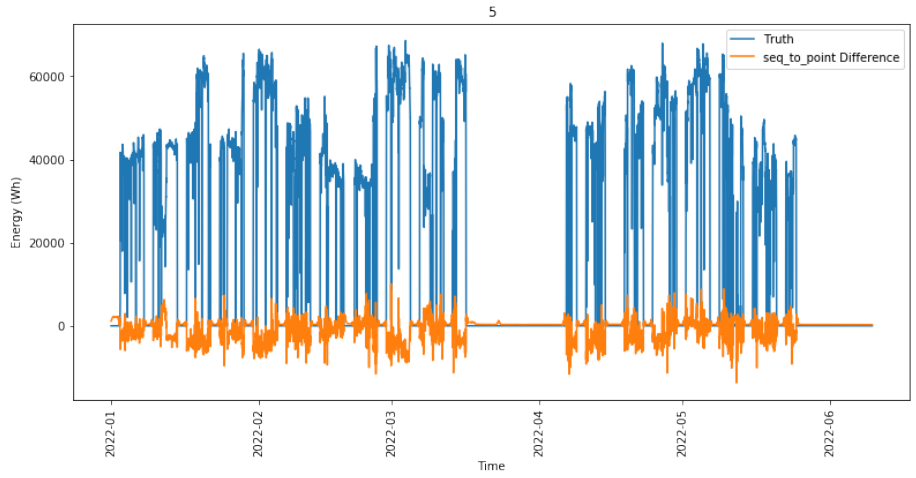
Transfer learning plots



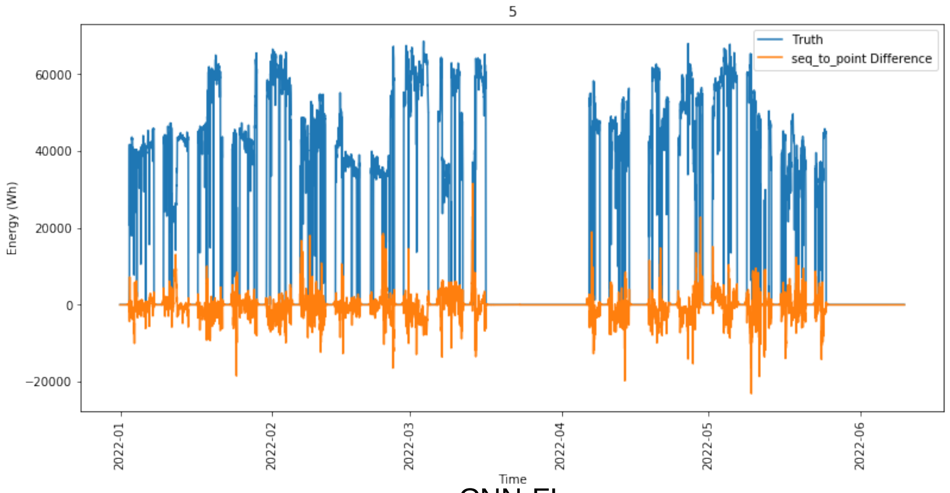
Zero-shot-Learning



CNN-2L



Normal Seq2Point



CNN-FL