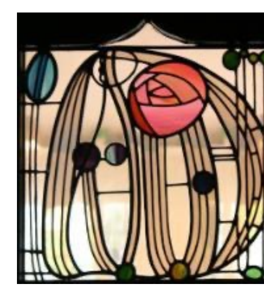
Leveraging Low-Carbon Energy for Flexible Compute Workloads in Cloud and Edge Environments

Dr. Lauritz Thamsen

Lecturer in Computer Systems

GLASS – SoCS – UofG

https://lauritzthamsen.org



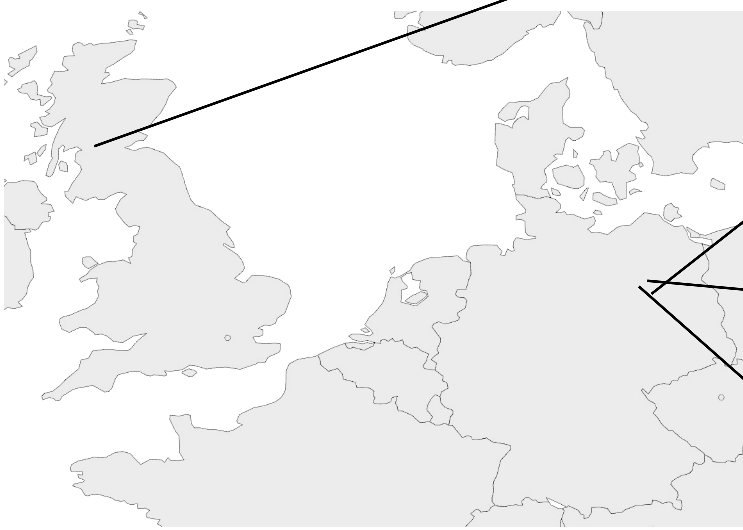
Outline

Background and Motivation

Carbon-Aware Cloud Workload Shifting

Edge Computing on Renewable Energy

Dr Lauritz Thamsen



Since ’22: Lecturer
in the Glasgow Systems Research Section

’21 – ’22: Guest Professor
in Prof. Ulf Leser’s Knowledge Management group

’20 – ’22: Senior Researcher
’18 – ’20: Postdoc Researcher
’14 – ’18: Research Assistant
in Prof. Odej Kao’s Distributed and Operating Systems group

’12 – ’14: Ph.D.-track Student
in Prof. Robert Hirschfeld’s Software Architecture group

Collaborative Work with Several Systems Groups

Distributed and Operating Systems at TU Berlin

Knowledge Management in Bioinformatics at HU Berlin

Operating Systems and Middleware at HPI, Uni Potsdam

ARM

https://adaptiveresourcemanagement.org

FONDA

https://fonda.hu-berlin.de

Funded by the DFG, BMBF, and DAAD

Adaptive Resource Management for Data-Intensive Systems

Adaptive Resource Allocation

Allocate compute resources to meet specific performance objectives and constraints

e.g. BigData’20 & '22, Cluster’21, ICFEC’21, IC2E’21 & 22, EuroPar’22, SSDBM’22

Scheduling & Dynamic Scaling

Adjust resource configurations at runtime as workloads change or components fail

e.g. CCPE journal’20, Middleware’21, ACSOS’21, SPE journal’21, CCGrid’23

Automatic System Tuning

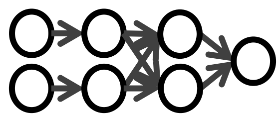
Tune system configurations using monitoring data, profiling, and performance models

e.g. BigData’19 & ’20, IC2E’22, ICWS’22

Scalable Data Processing Systems

Many data-intensive applications run on top of scalable and fault-tolerant distributed processing systems

Distributed Batch / Stream Processing

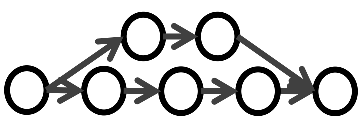


Apache Spark

Flink

Apache Storm

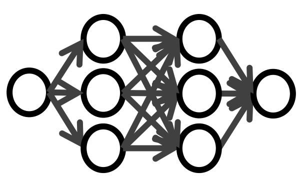
Scientific Workflows



Pegasus

nextflow

Machine Learning



TensorFlow

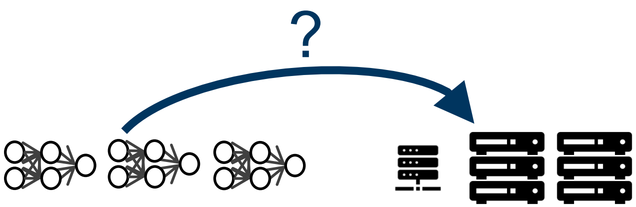
PyTorch

Research Questions

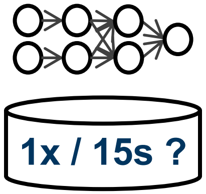
Given a job and objectives/constraints for its execution:

1. What resources to use for the job?

2. When and where to run the tasks?



3. How to set system configurations?



Computing’s Growing Footprint

- Data centers already consume > 1% of the globally produced energy, a share that is projected to rise sharply over the next decades

- More and more large-scale, long-running, resource-intensive data processing jobs (e.g. Big Data, ML/AI, and IoT)

- Emissions depend on the energy consumption, yet also the specific sources of energy

Carbon-Conscious Computing

- Objective: **Reducing the carbon footprint** of large-scale data processing applications on today's diverse distributed computing infrastructure

1. Compute when and where low-carbon energy is going to be available
2. Allocate resources for high resource utilization and highly utilize allocated resources
3. Save computation and communication through distributed and dynamic architectures

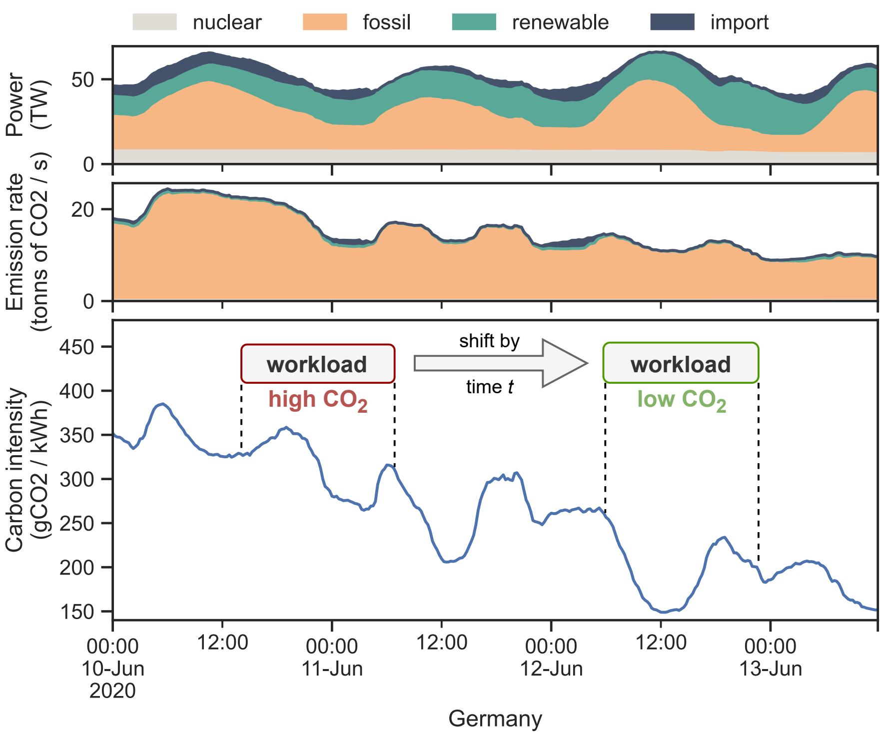
Carbon-Aware Cloud Workload Shifting

Let's Wait Awhile: How Temporal Workload Shifting Can Reduce Carbon Emissions in the Cloud.
Wiesner, Behnke, Scheinert, Gontarska, Thamsen. Middleware’21.

Motivation

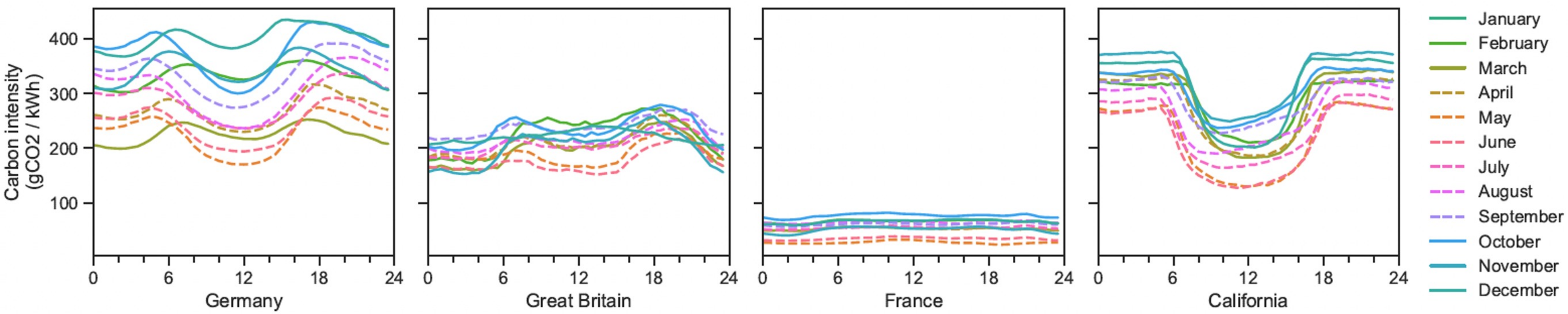
- Emissions of power grids are determined by the energy mix and demand

- Low-carbon objective: Compute when and where low-carbon energy is available



Changing Carbon Intensity (1/2)

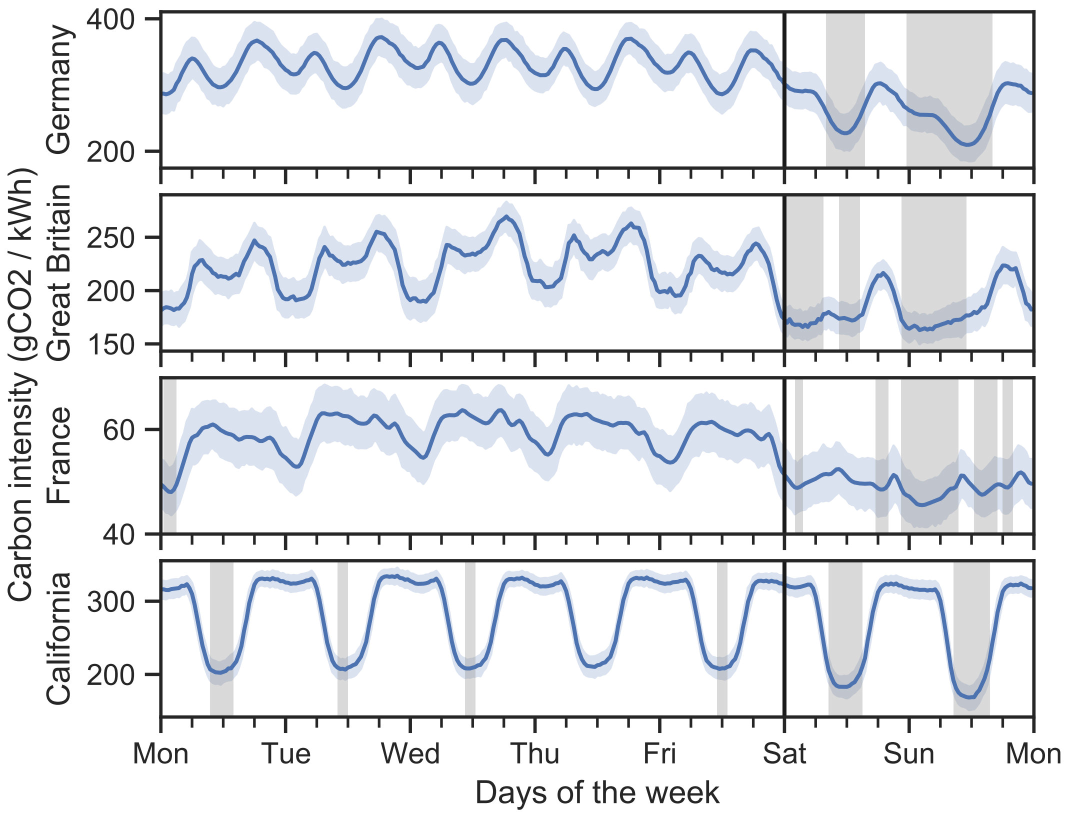
- What are the most promising times to shift work to?



- Average carbon intensity (== CO2-equiv. greenhouse gas emissions per kilowatt hour of energy) in 2020

Changing Carbon Intensity (2/2)

- What are the most promising days to shift work to?



Simulations

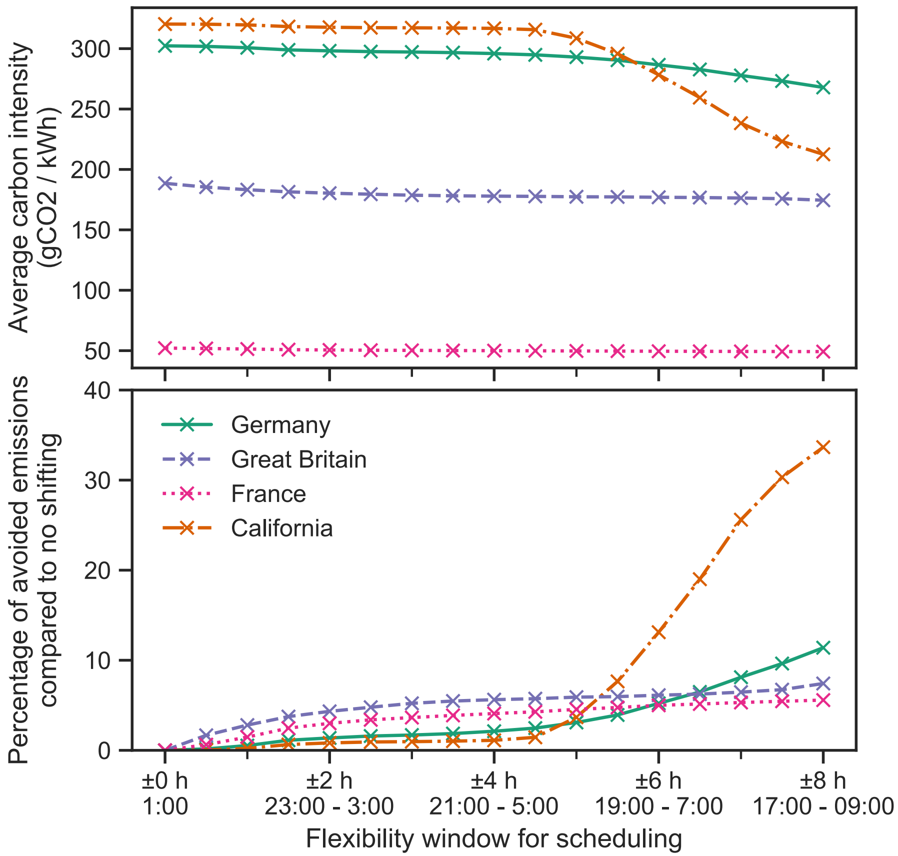
- Evaluation of two scenarios using our simulator (https://github.com/dos-group/leaf)

- Scenario 1 – Periodic Jobs: Nightly builds, integration tests, recurring generation of business reports, …

- Scenario 2 – Ad Hoc Jobs: ML training jobs, data analysis pipelines, scientific simulations, …

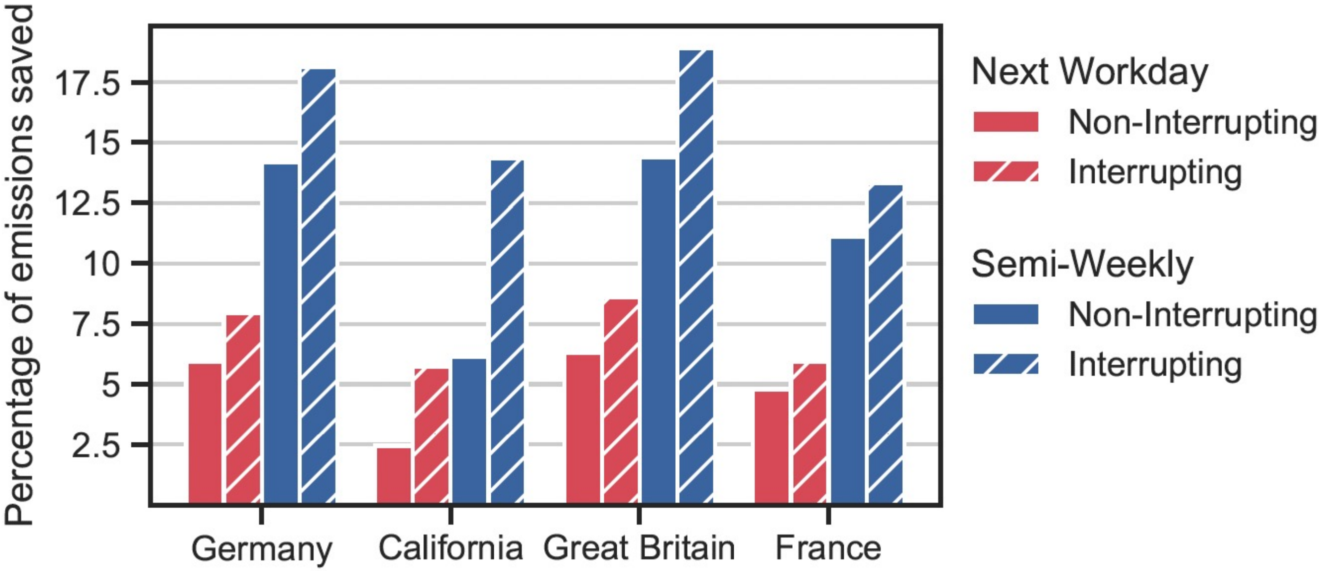
Scenario 1: Periodic Jobs

- Baseline: All jobs scheduled at 1 am in the night
- Increasing the window by +- 1h to allow scheduling between
  - 00:00 to 3:00 (+- 1h)
  - 23:00 to 4:00 (+- 2h)
  - …
  - 17:00 to 9:00 (+- 8h)



Scenario 2: Large Ad Hoc Jobs

- Based on an NVIDIA research project, which ran 3387 ML training jobs using 145.76 GPU years and 325 MWh
- Baseline: Instant scheduling of jobs that arrive randomly during working hours



Simplifications in our First Simulations

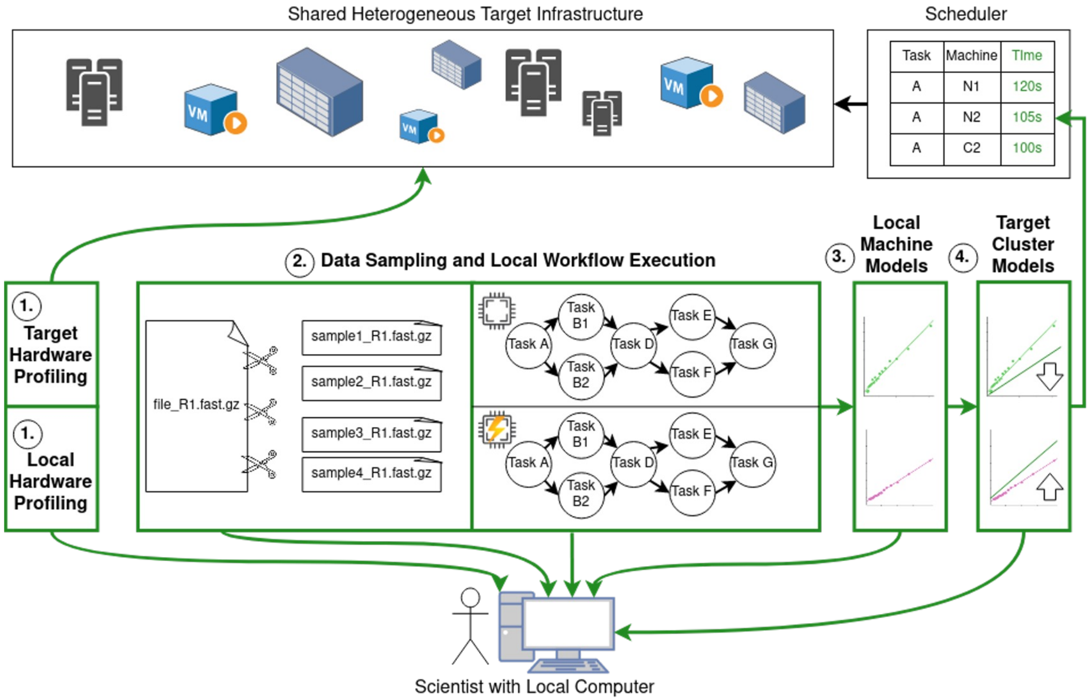
- Exact knowledge of job runtimes and no overhead for interrupting jobs repeatedly

- Given and static resource configuration for jobs

- Uniform low-error emissions forecasts and no hardware/software failures slowing down processing

Task Runtime Estimation for Scientific Workflows

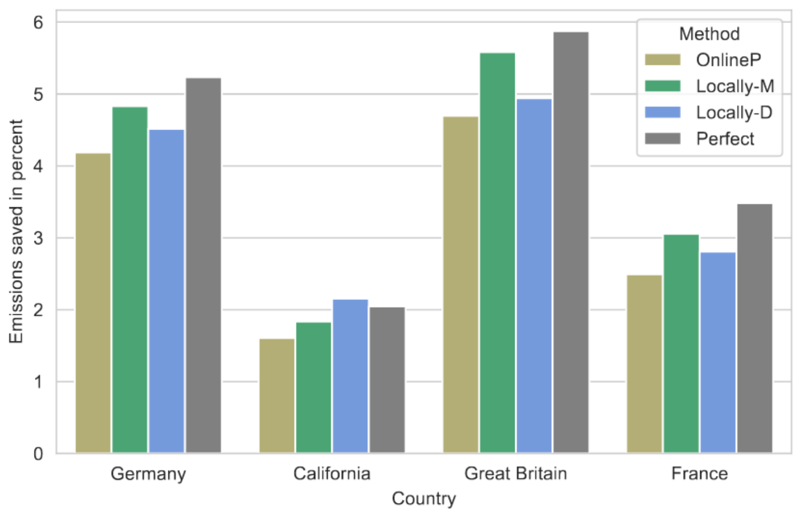
Our "Lotaru" method for estimating the task runtimes of scientific workflows on a scientist's personal machine:



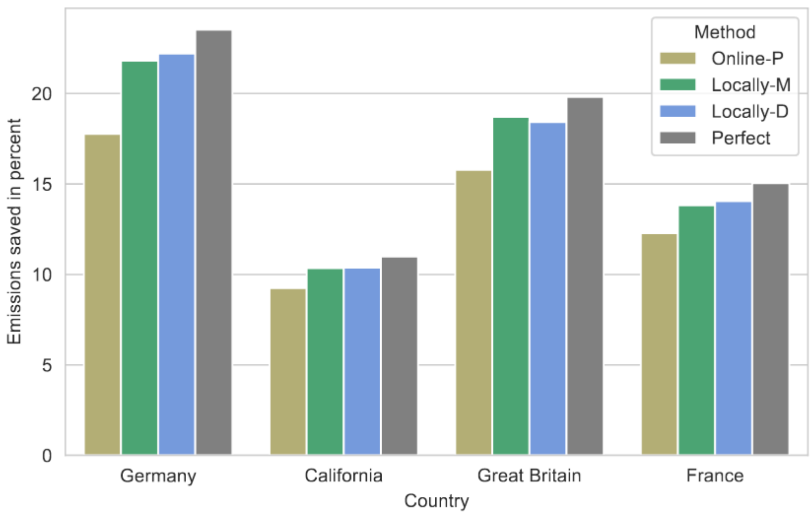
Lotaru: Locally Estimating Runtimes of Scientific Workflow Tasks in Heterogeneous Clusters. Bader, Lehmann, Thamsen, Will, Leser, and Kao. SSDBM'22.

Realistic Estimations to Shift Scientific Workflows

- Shifting bioinformatics workflows (nf-core) on Google's cloud based on locally estimated runtimes using Lotaru



Results by next morning



Results by next Monday

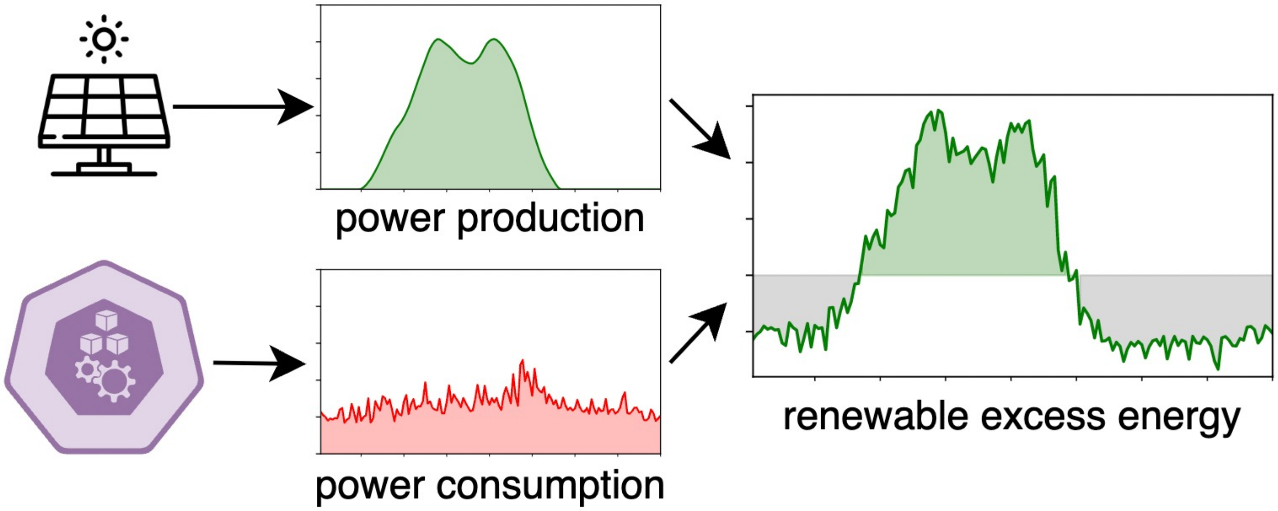
[Work in progress – for an extended TPDS article on the estimation method]

Edge Computing on Renewable Energy

Cucumber: Renewable-Aware Admission Control for Delay-Tolerant Cloud and Edge Workloads.
Wiesner, Scheinert, Wittkopp, Thamsen, Kao. EuroPar’22.

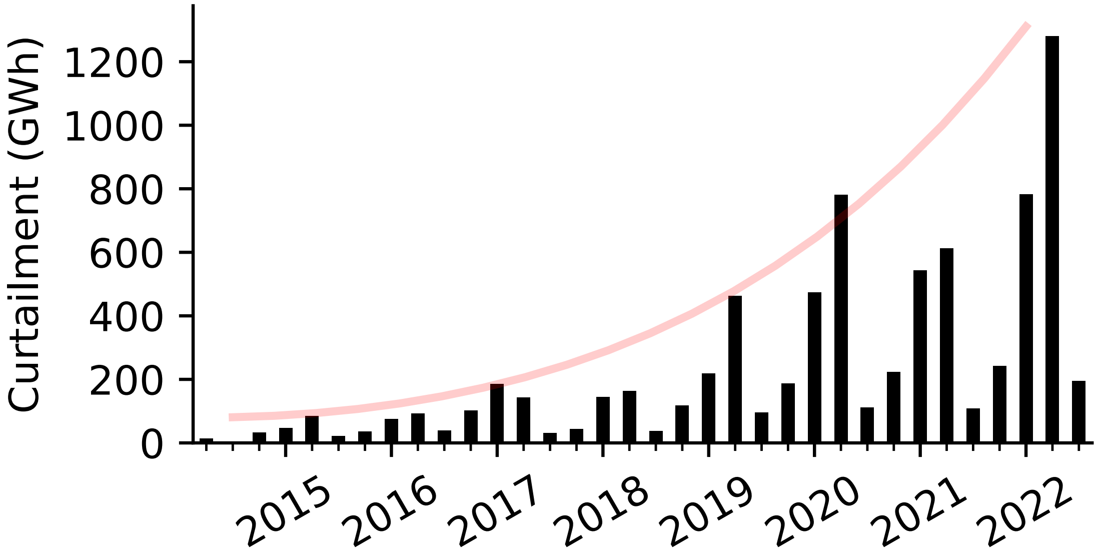
Renewable Excess Energy

The output from renewables such as solar and wind varies, and there can be more energy than demand



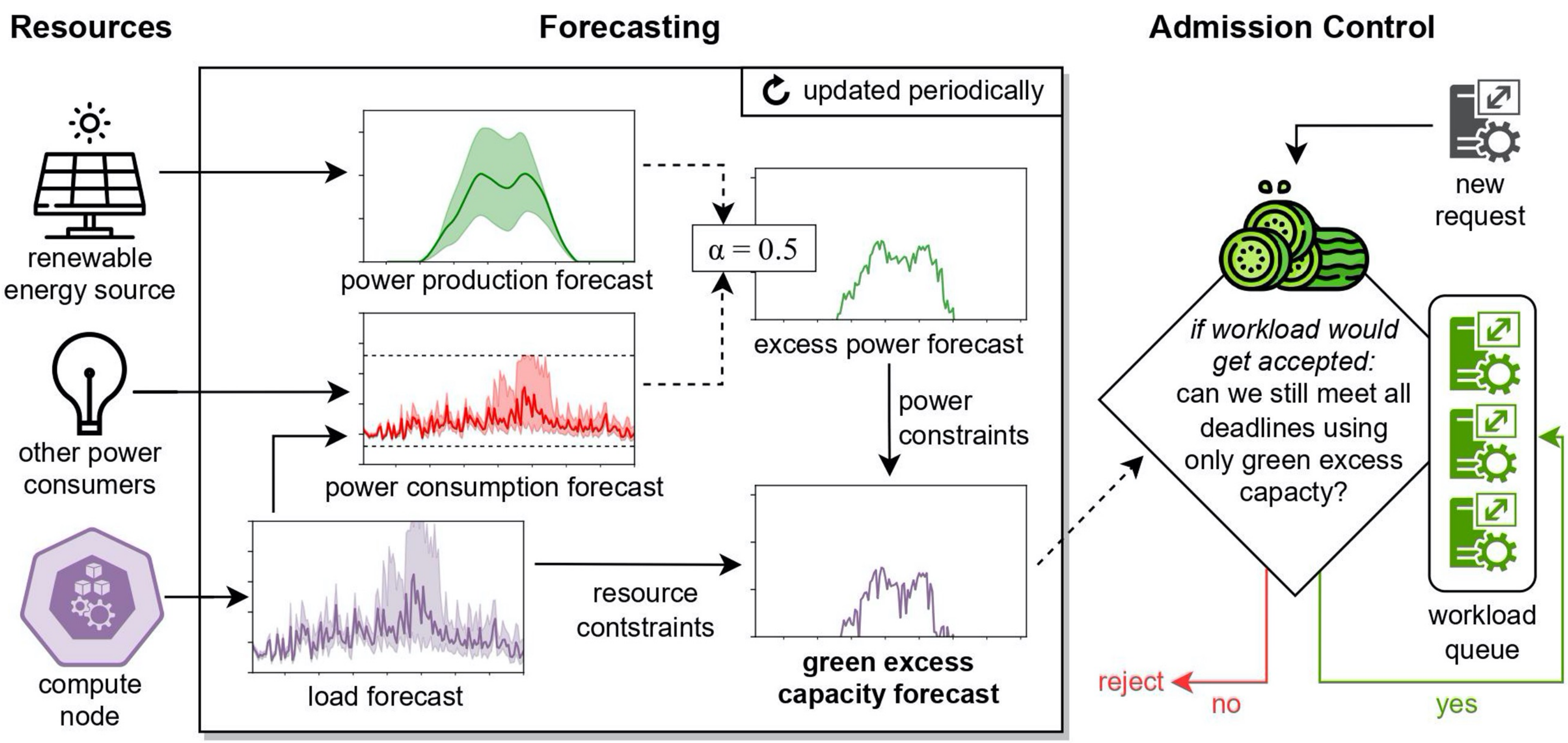
Solar Energy Curtailment in California

- Around 7% of solar power is being curtailed already



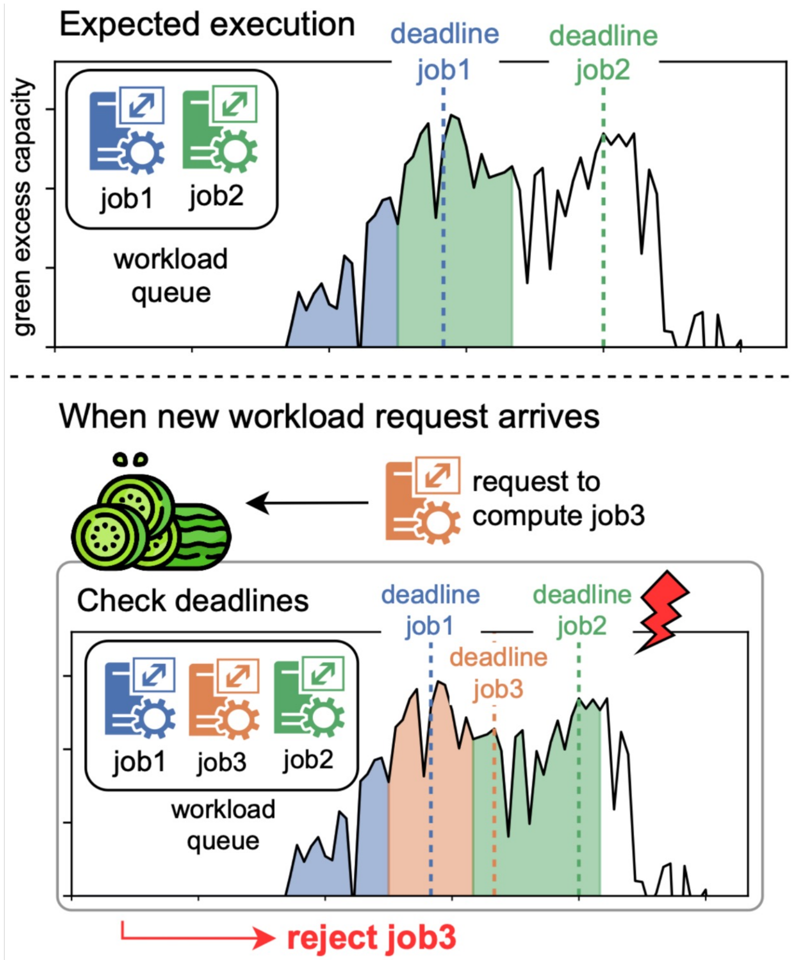
Source: California Independent System Operator (CAISO)

“Cucumber” Overview



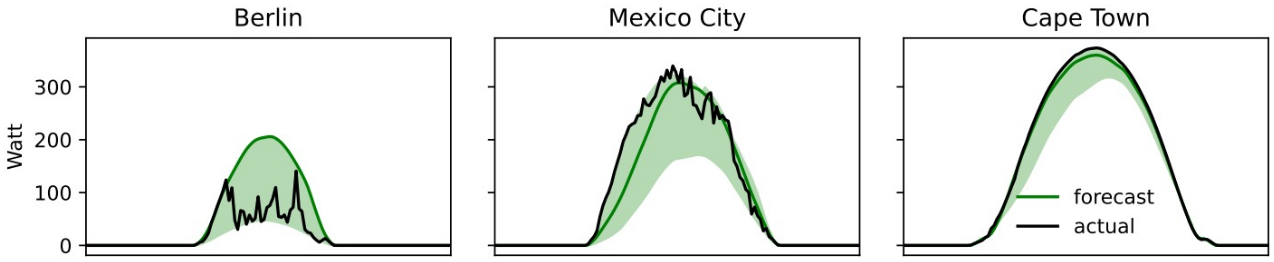
Cucumber Admission Control

- Given the size and deadlines of jobs, we admit jobs to use predicted excess energy and spare capacity
- Through probabilistic forecasts, admission can be tuned to be
  - conservative (low acceptance rate, low grid power usage) or
  - optimistic (vice versa)

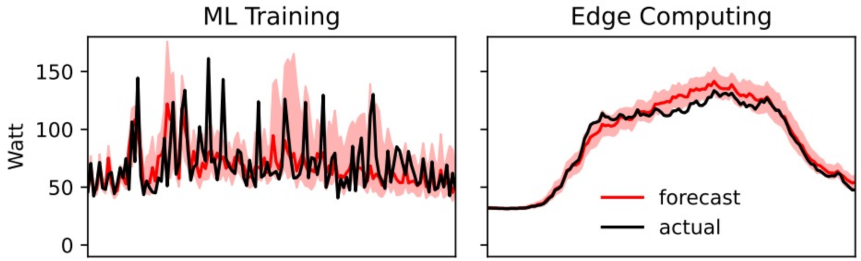


Simulation Setup

- Two weeks of solar production forecasts for 400W panels across three sites (using https://solcast.com/)

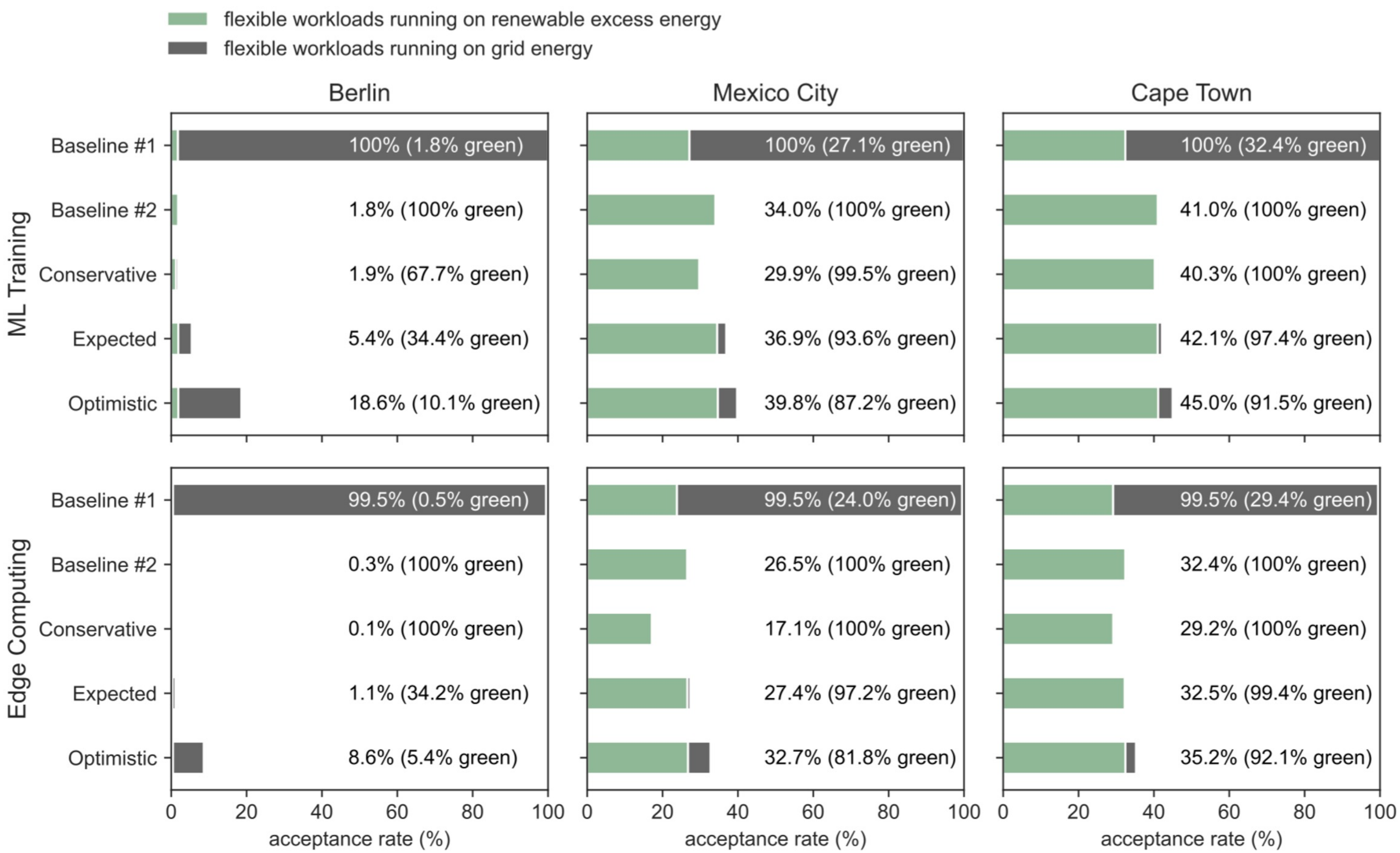


- Two workload traces



ML Training based on Alibaba GPU cluster traces (deadlines set to midnight)
Edge Computing based on a NYC taxi trip dataset (deadlines derived from trip lengths)

Simulation Results



Where are the Flexible Low-Priority Jobs Coming From?

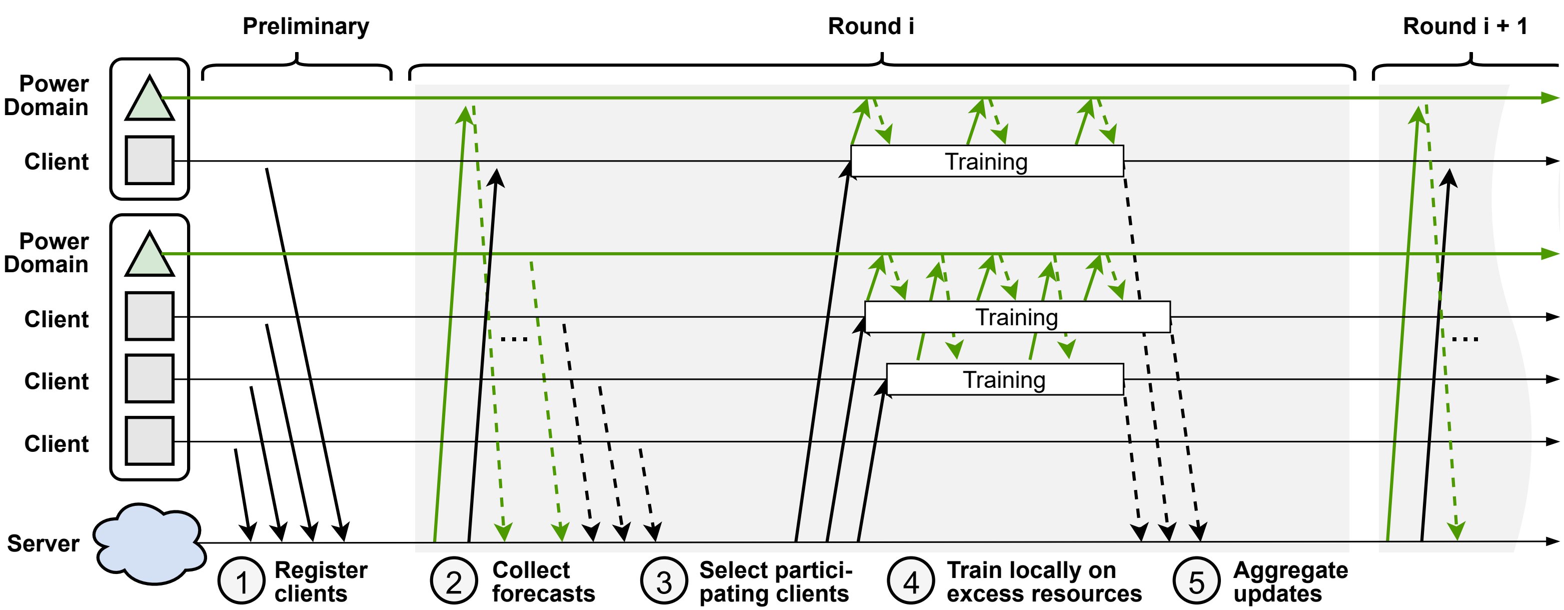
A recent trend in distributed ML is Federated Learning (FL), improving data privacy

FL seems a very promising candidate for carbon-aware computing:

- constitutes energy-intensive batch jobs
- scheduled in geo-distributed environments
- without strict runtime requirements

Newest Work: “FedZero”

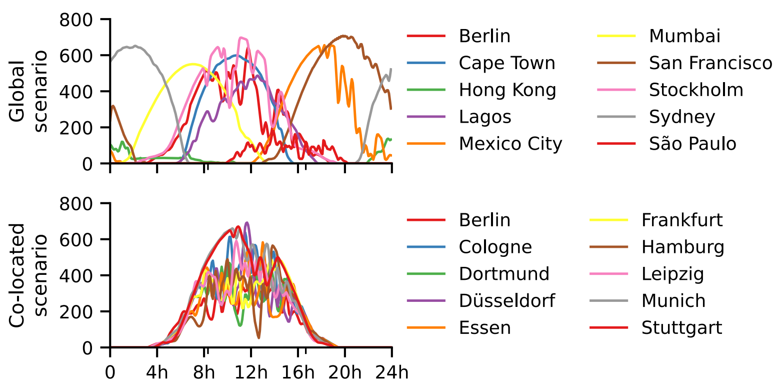
- Scalable client selection strategy for Zero-Carbon Federated Learning with fast convergence and fair client participation based on forecasts



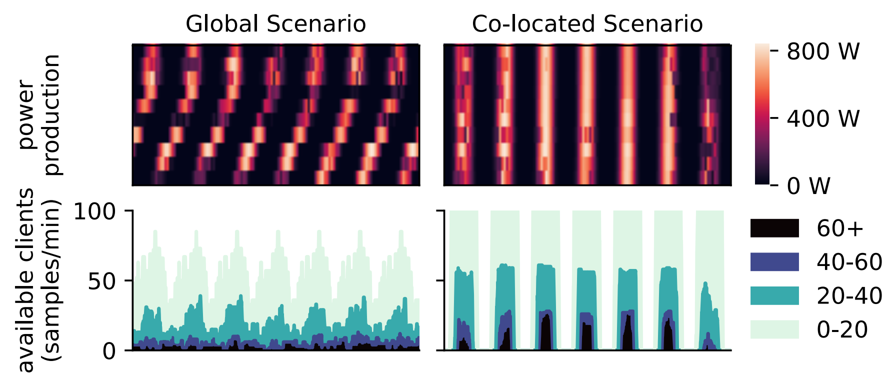
[Currently under review]

FedZero Evaluation Setup

- Discrete-event simulation on top of Flower (https://flower.dev)
- 100 clients with different Nvidia GPUs in terms of throughput and energy consumption – emulated on six actual cards
- Two scenarios with real solar forecasts and baseload derived from an Alibaba trace dataset



Day 1 solar power production (in Watt)



Energy/client availability over seven days

[Currently under review]

FedZero Overall Results

Dataset & model	Data distribution & aggregation strategy	Approach	Global			Co-located		
			Target accuracy	Time-to-accuracy	Energy-to-accuracy	Target accuracy	Time-to-accuracy	Energy-to-accuracy
CIFAR-10 ResNet-18	iid FedAvg	Constrained	83.26 %	7.0 d	72.1 kWh	84.04 %	6.6 d	101.5 kWh
		FedZero		4.6 d	74.5 kWh		5.6 d	109.7 kWh
		Unconstrained		1.5 d	85.1 kWh		2.2 d	128.1 kWh
	non-iid FedProx	Constrained	79.11 %	6.7 d	71.3 kWh	80.53 %	6.6 d	86.9 kWh
		FedZero		4.3 d	67.4 kWh		4.8 d	88.0 kWh
		Unconstrained		1.4 d	68.7 kWh		2.1 d	105.6 kWh
CIFAR-100 DenseNet-121	iid FedAvg	Constrained	57.85 %	6.7 d	78.6 kWh	58.90 %	6.7 d	101.2 kWh
		FedZero		4.4 d	82.1 kWh		4.5 d	94.3 kWh
		Unconstrained		1.5 d	89.0 kWh		2.0 d	119.6 kWh
	non-iid FedProx	Constrained	56.32 %	6.7 d	76.6 kWh	57.63 %	6.8 d	102.1 kWh
		FedZero		4.5 d	82.7 kWh		4.6 d	99.5 kWh
		Unconstrained		1.5 d	88.3 kWh		2.2 d	128.7 kWh
Shakespeare LSTM	non-iid FedProx	Constrained	52.14 %	5.7 d	79.3 kWh	52.57 %	6.7 d	77.2 kWh
		FedZero		1.4 d	27.7 kWh		2.2 d	37.8 kWh
		Unconstrained		1.4 d	46.7 kWh		1.9 d	65.0 kWh

[Currently under review]

Summary

Carbon-Aware Edge/Cloud Computing

- Fluctuations in grid carbon intensities can be leveraged to reduce the footprint of flexible workloads by 5-20%

- Renewable excess energy and spare capacity can drive down the footprint of e.g. ML and FL substantially

- Interesting research ahead of us to realize the potential carbon savings – e.g. dynamic scheduling and scaling based on performance models and forecasts

- Contact: lauritz.thamsen@glasgow.ac.uk