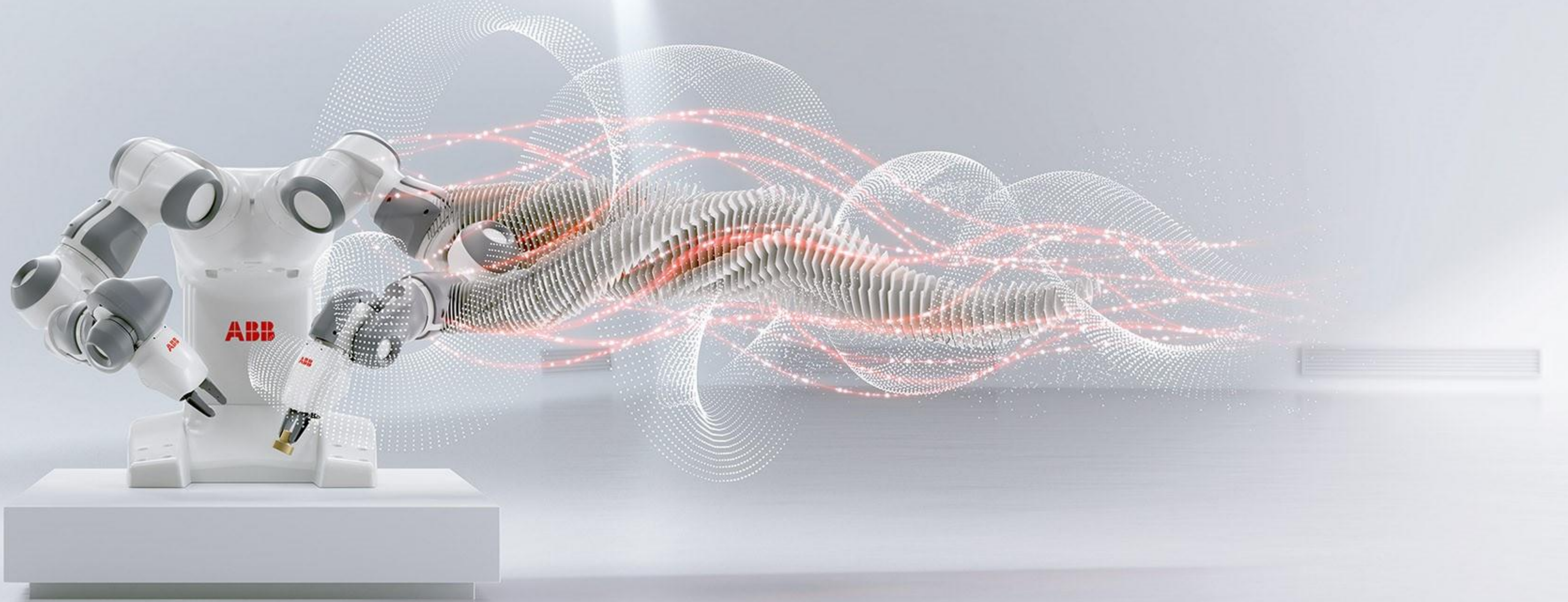


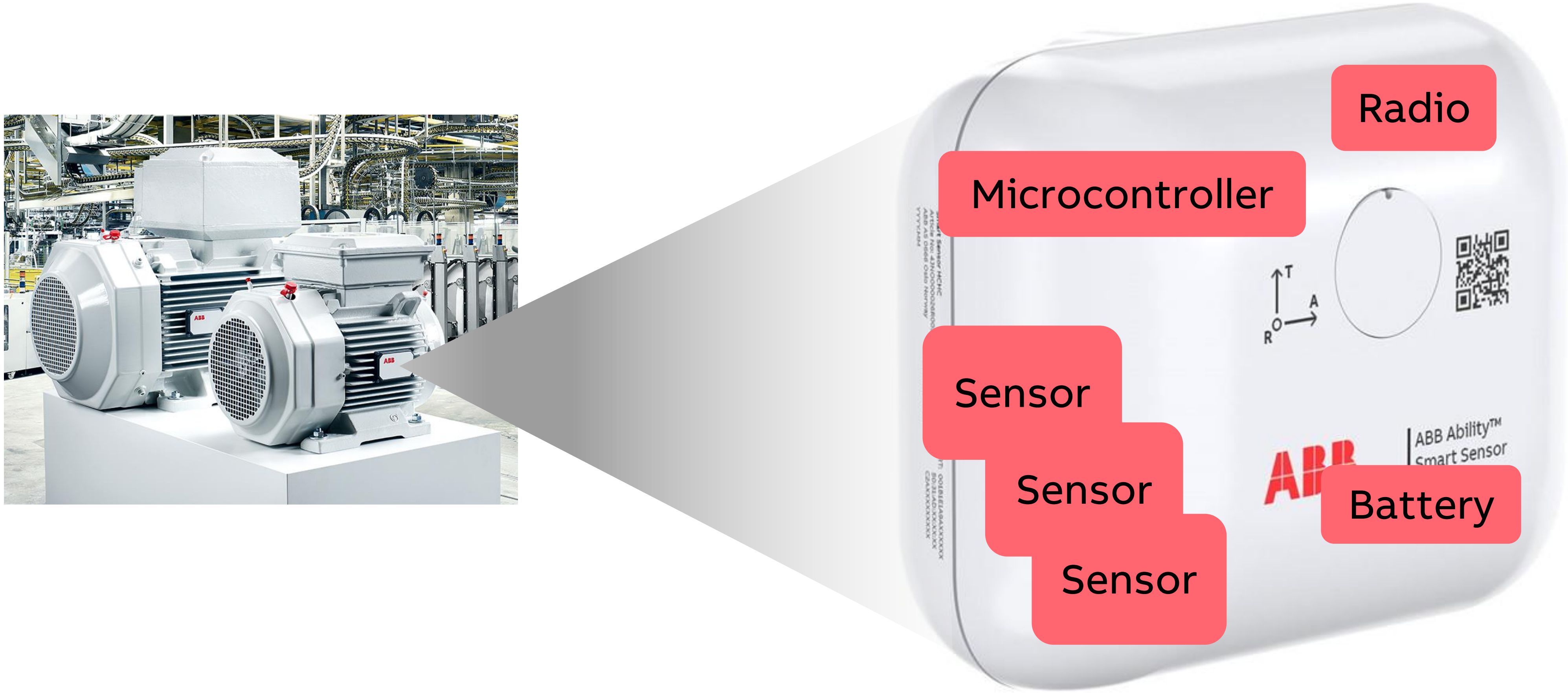
INTERNATIONAL CONFERENCE ON EDGE WIRELESS SYSTEMS AND NETWORKS (EWSN), OCTOBER 3-5, 2022

SMiLe : Automated End-to-end Sensing and Machine Learning Co-Design

Tanmay Goyal, Pengcheng Huang, Felix Sutton, Balz Maag and Philipp Sommer,
ABB Research Switzerland

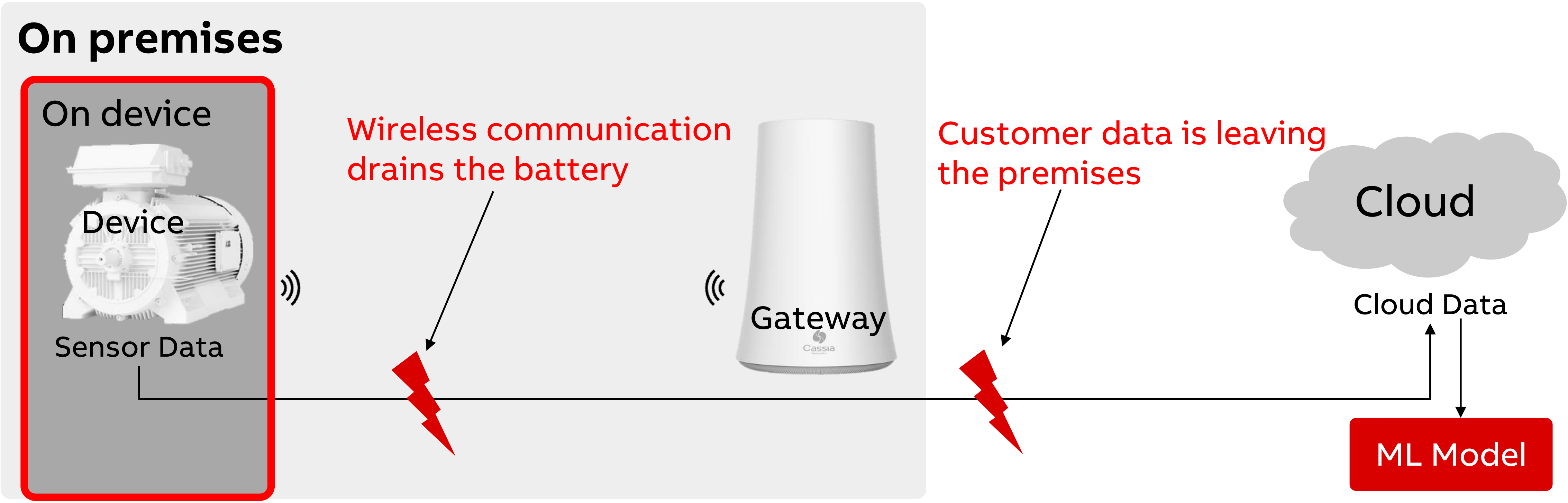
The Evolution of Tiny Machine Learning (TinyML)

Towards Machine Learning at the Edge



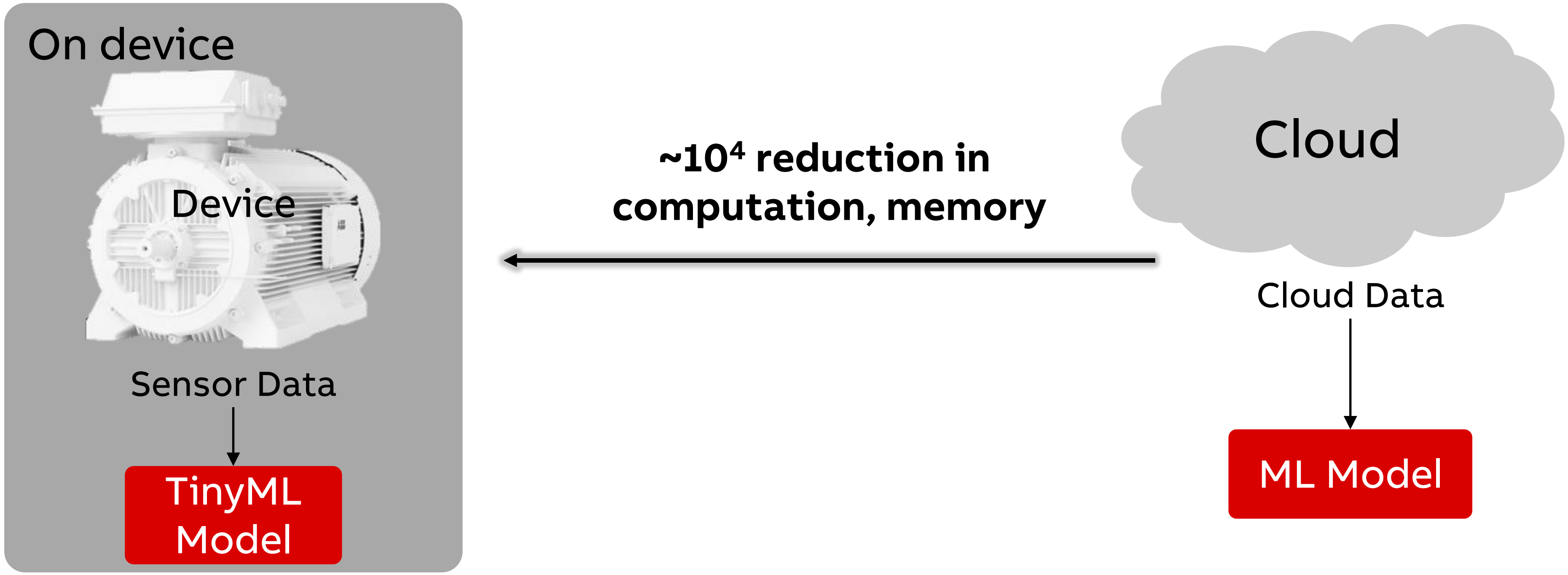
The Evolution of Tiny Machine Learning (TinyML)

Towards Machine Learning at the Edge



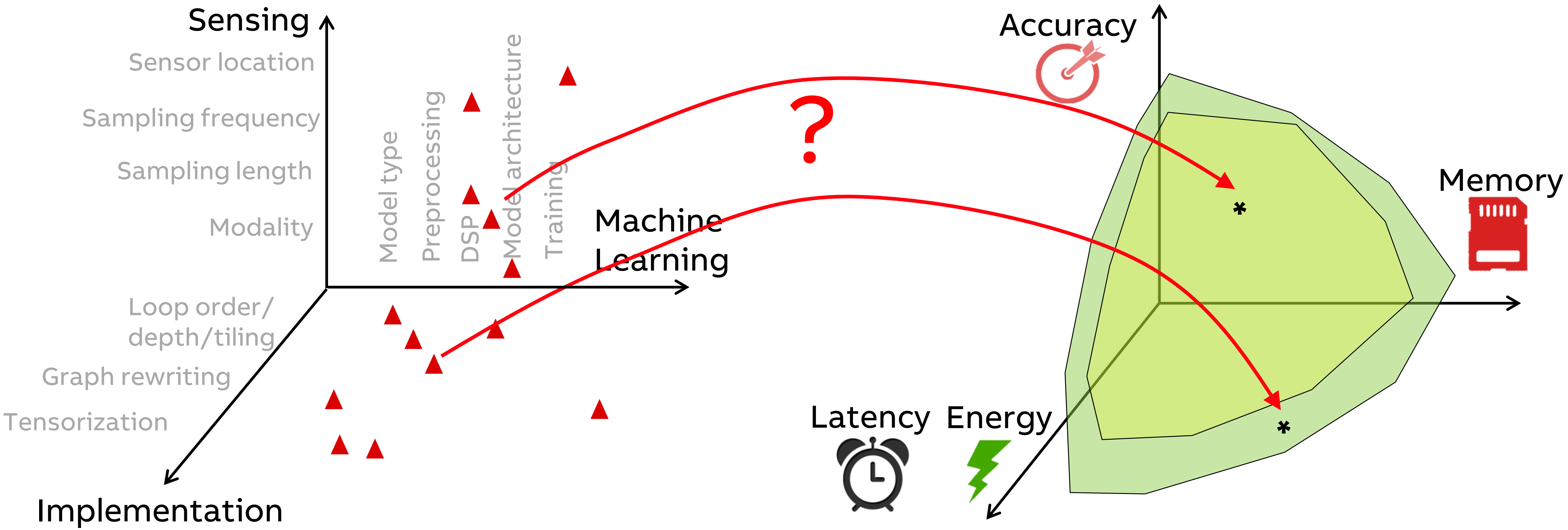
The Evolution of Tiny Machine Learning (TinyML)

Challenges



The Evolution of Tiny Machine Learning (TinyML)

Challenges



Enormous Design Space **Complex Trade-offs**

Migration from Cloud Analytics to Edge Analytics

Our 3-fold Approach

Optimization

- Make clever decisions to co-design Sensing and Machine Learning

Hardware-in-the-loop

- Measure performance directly on the edge system

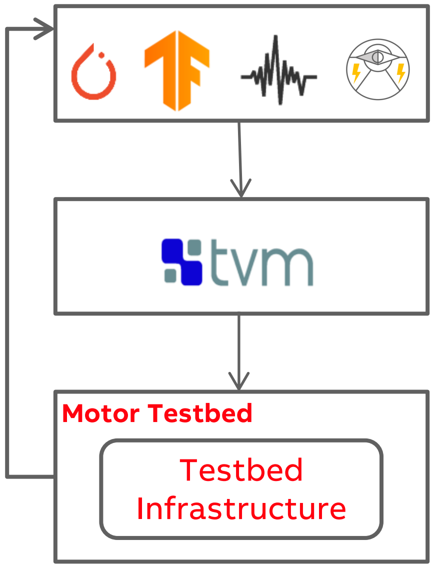
Automation

- Mitigate time consuming manual tuning

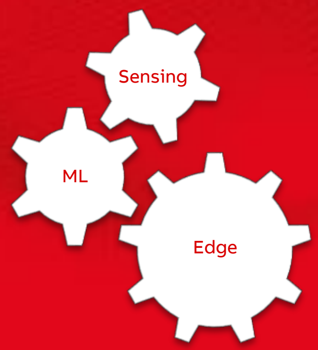
In our paper, we deliver an automated framework focussing on optimizing the Sensing and Machine Learning Co-Design using feedback from Hardware-in-the-loop

SMiLe: Automated End-to-end Sensing and Machine Learning Co-Design

SMiLe Overview



Sensing & ML Co-Design



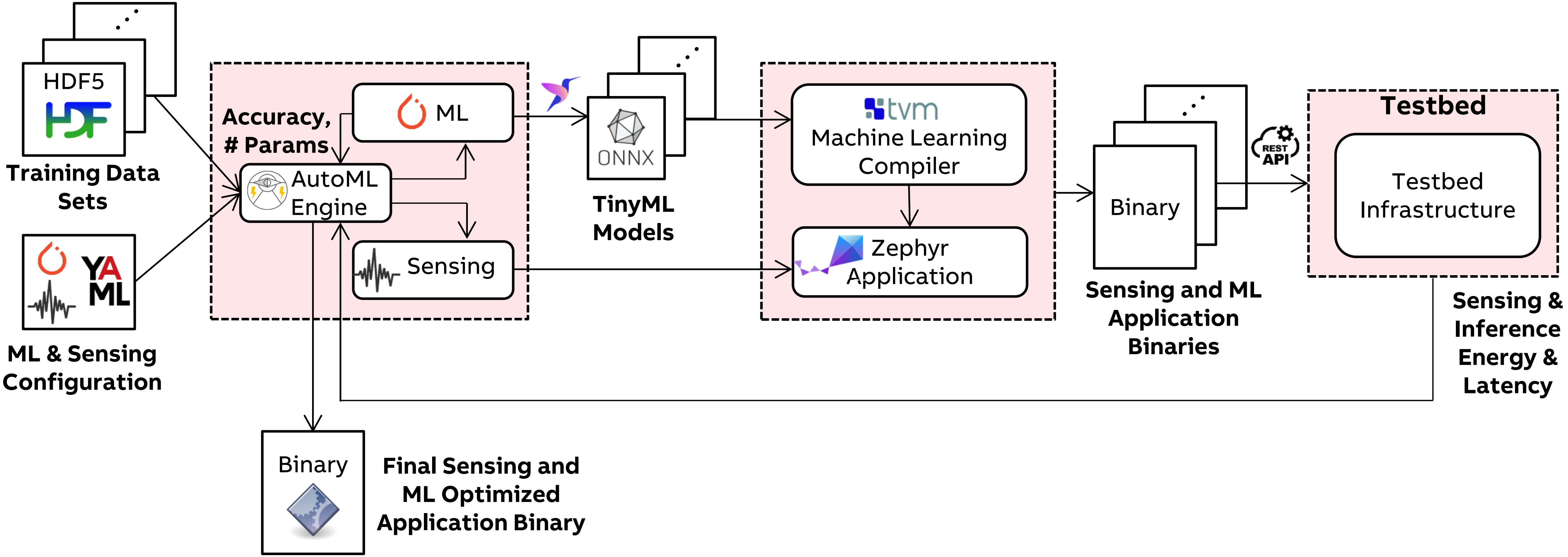
Experimental Evaluation

Smart Sensor

Nordic nRF5340

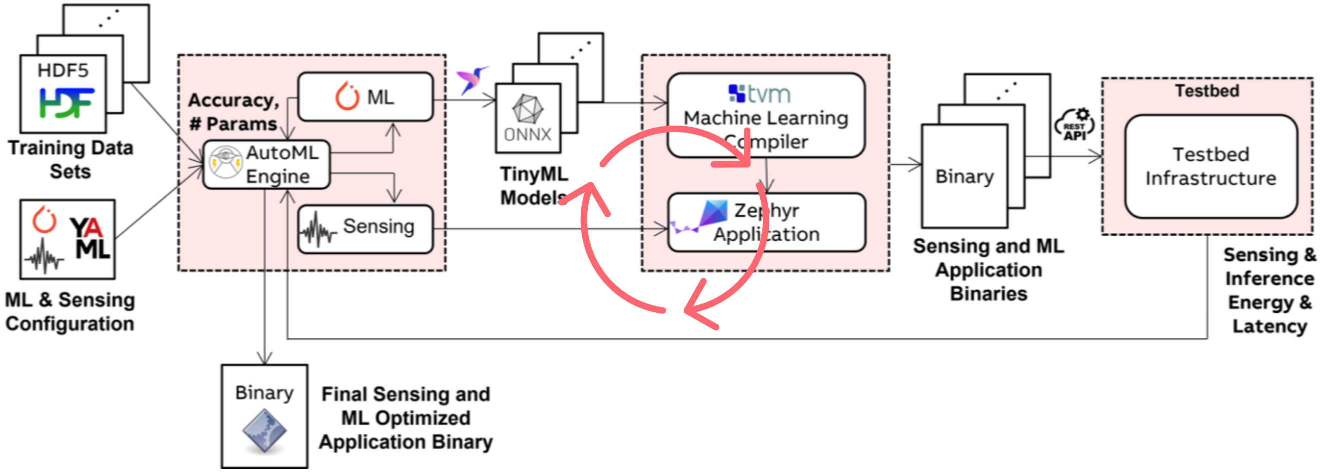
SMiLe framework

Automated end-to-end edge analytics

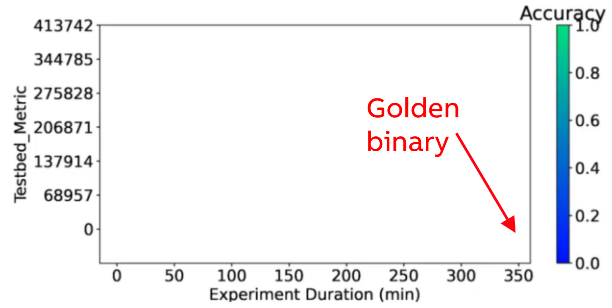


SMiLe – ABB Solution to Edge Analytics

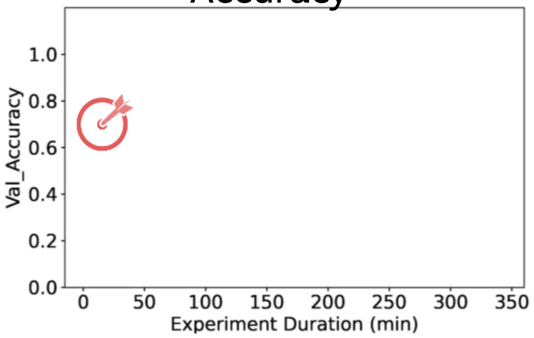
Automated end-to-end edge analytics



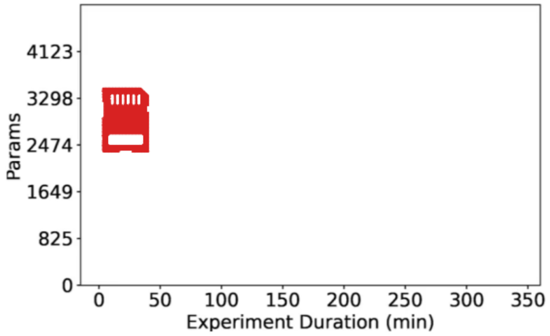
Combined global metric



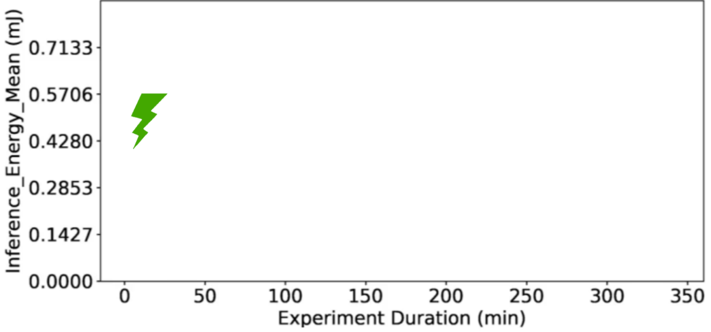
Accuracy



Mem

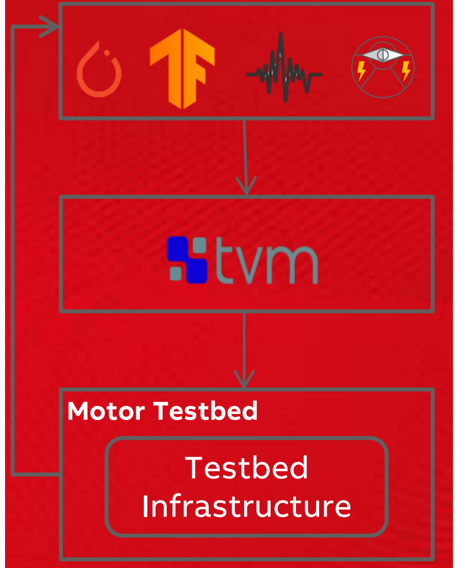


Energy

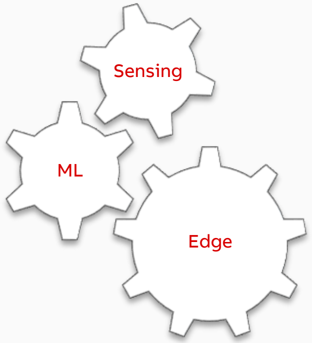


SMiLe: Automated End-to-end Sensing and Machine Learning Co-Design

SMiLe Overview



Sensing & ML Co-Design



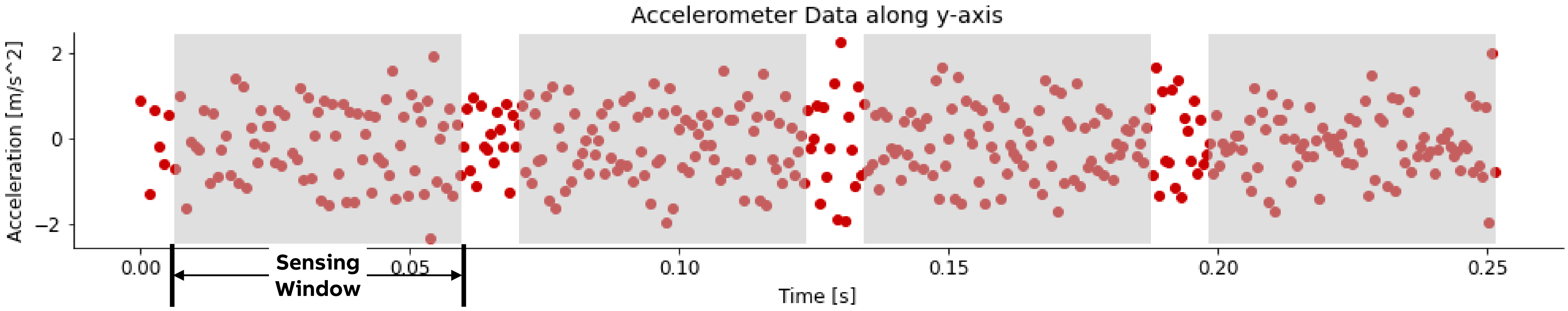
Experimental Evaluation

Smart Sensor

Nordic nRF5340

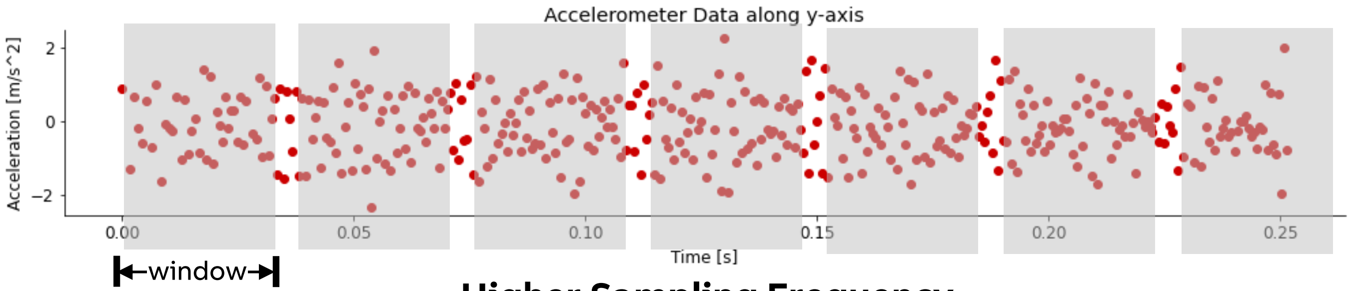
Sensing Parameters

Sampling Frequency and Sensing Window

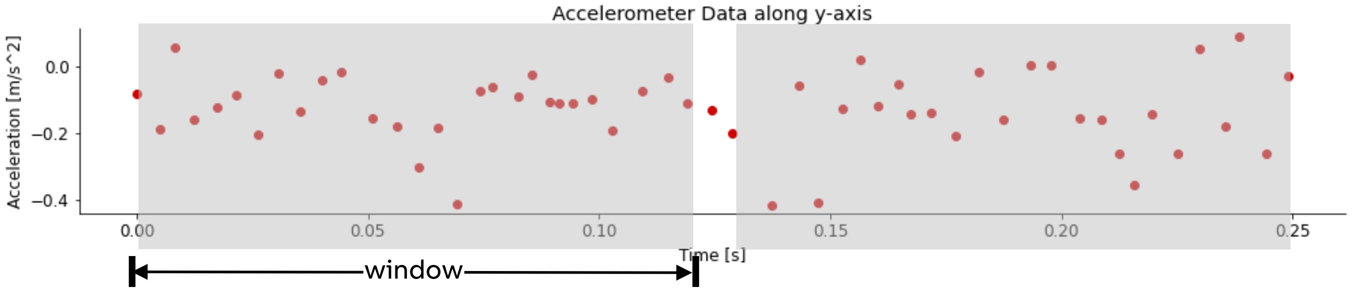


Need for Sensing and ML Co-Design

Impact of Sampling Frequency



Higher Sampling Frequency



Lower Sampling Frequency

Higher Sampling Frequency

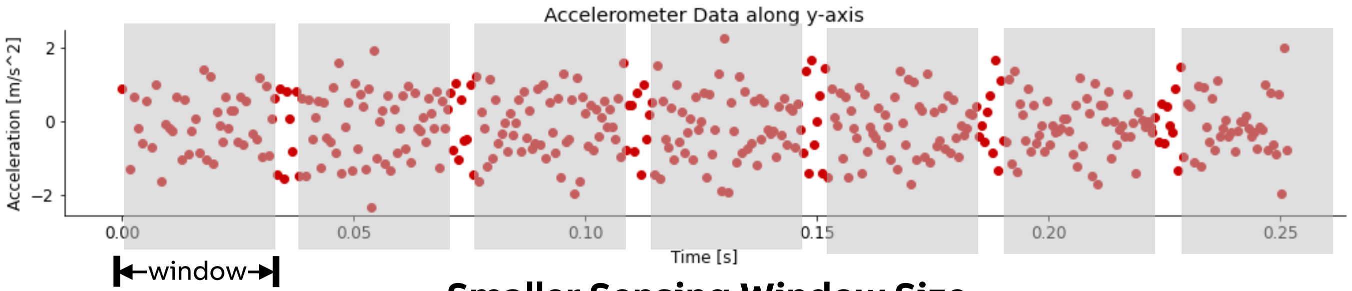
Analytics Phase	Energy
Sensing	↓
Inference	?

↓ Model Complexity ↑ Information Processed

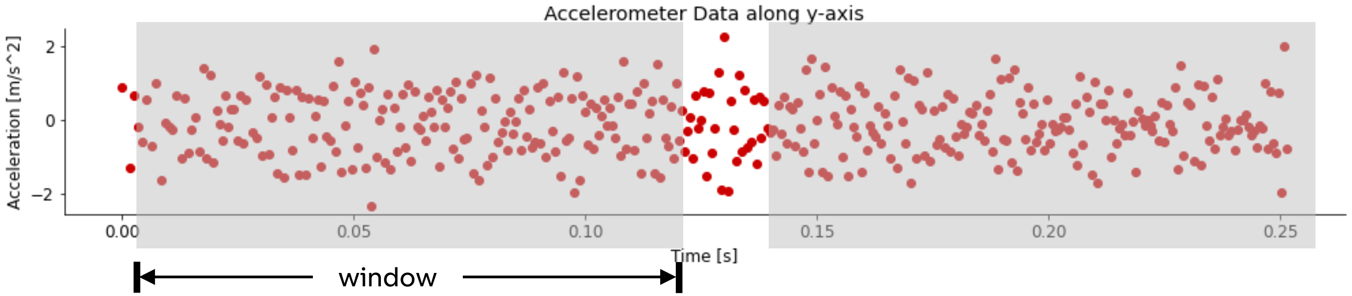
? FLOPs

Need for Sensing and ML Co-Design

Impact of Sensing Window Size



Smaller Sensing Window Size



Larger Sensing Window Size

Larger Sensing Window Size

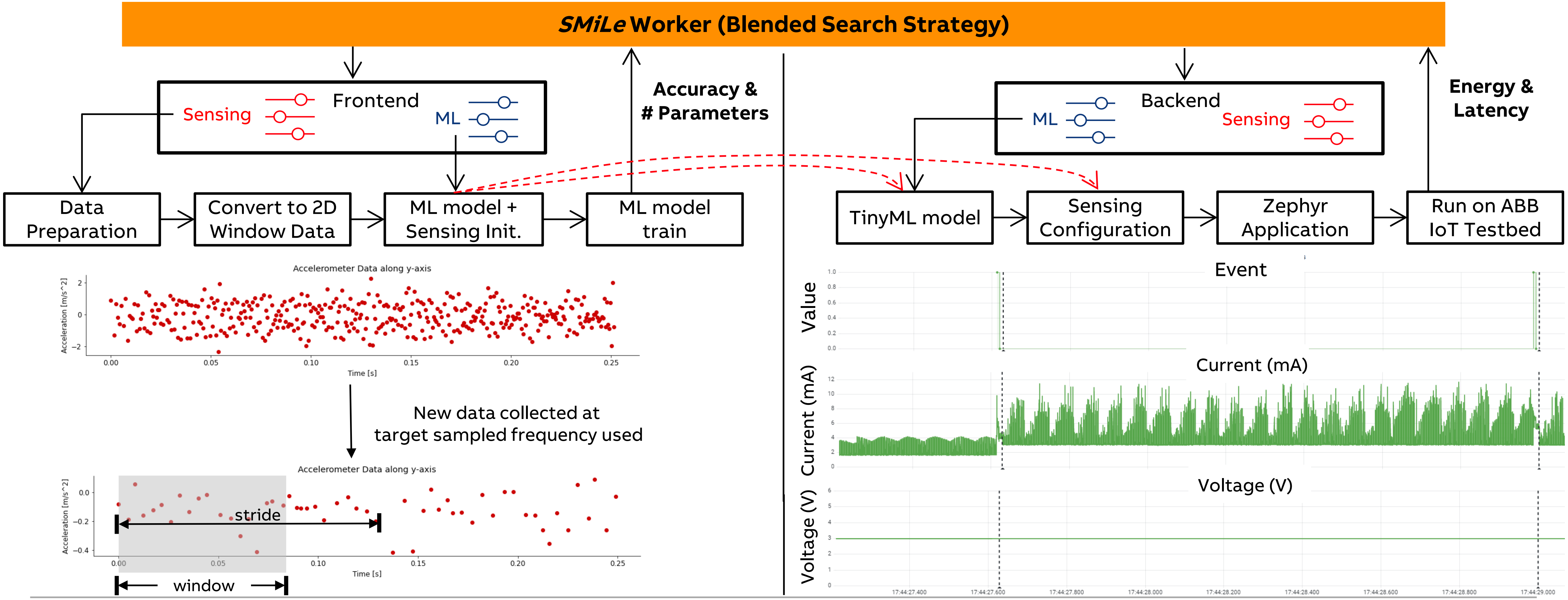
Analytics Phase	Energy
Sensing	↑
Inference	?

↓ Model Complexity **↑ FLOPs per layer**

? FLOPs

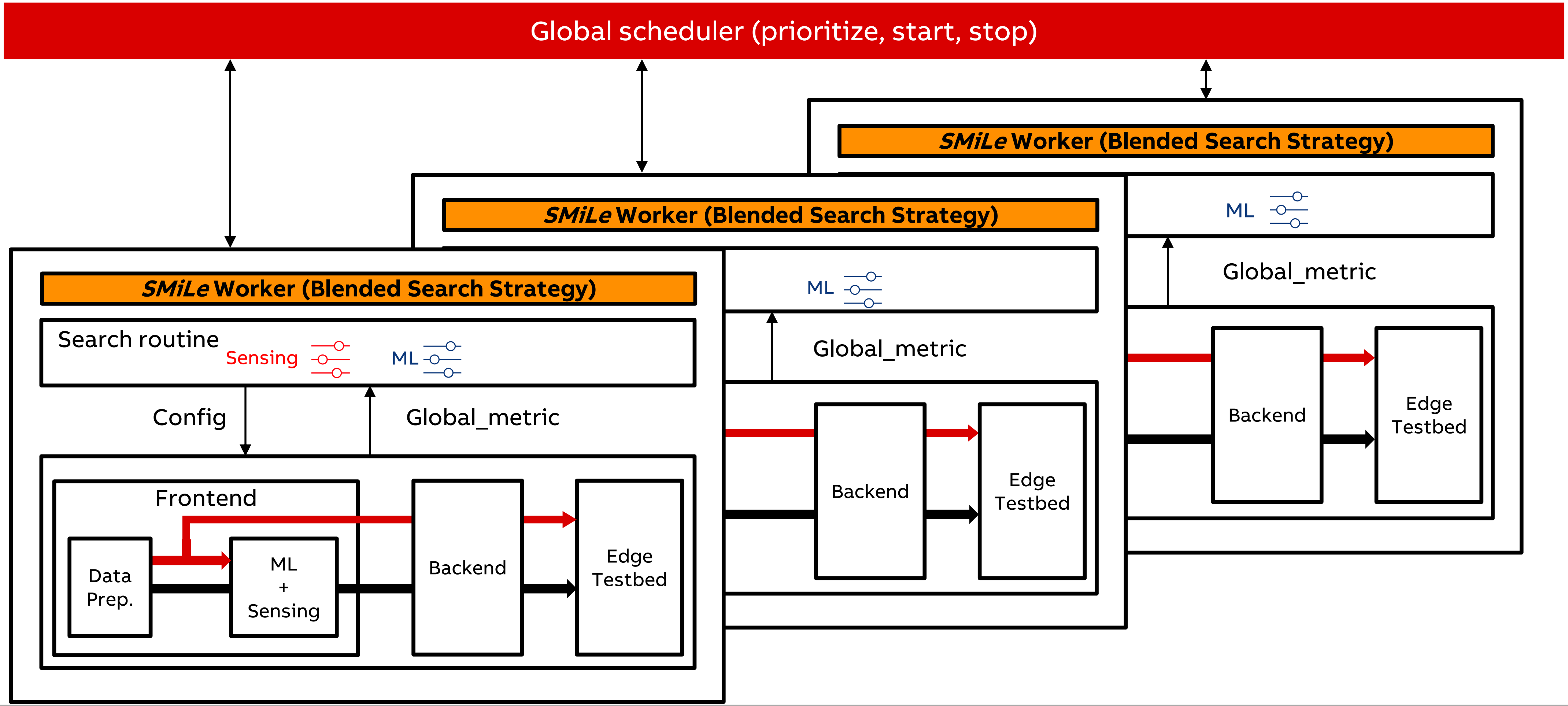
Sensing and ML Co-Optimization

Co-Optimization of sensing and ML using *SMiLe*



Parallel Co-Optimization of *SMiLe*

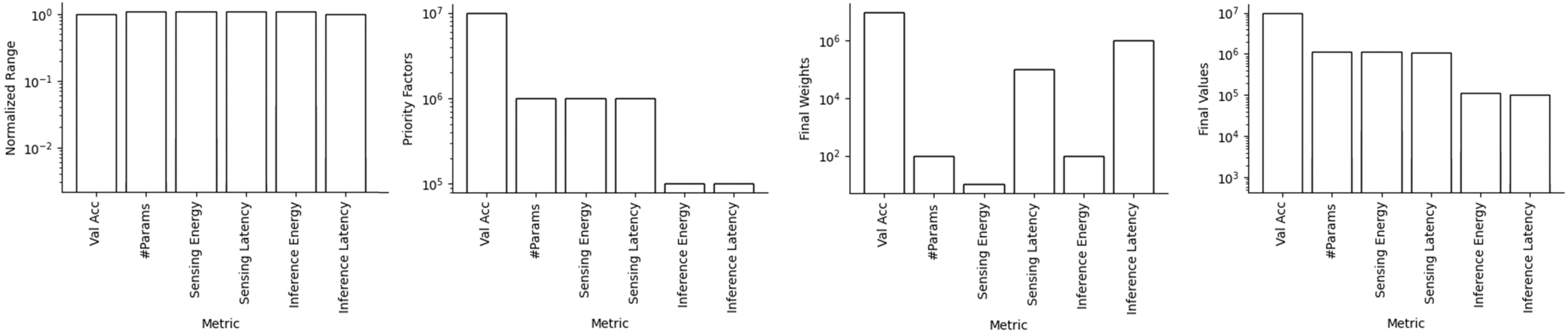
Co-Optimization of sensing and ML using *SMiLe*



Sensing and ML Co-Optimization

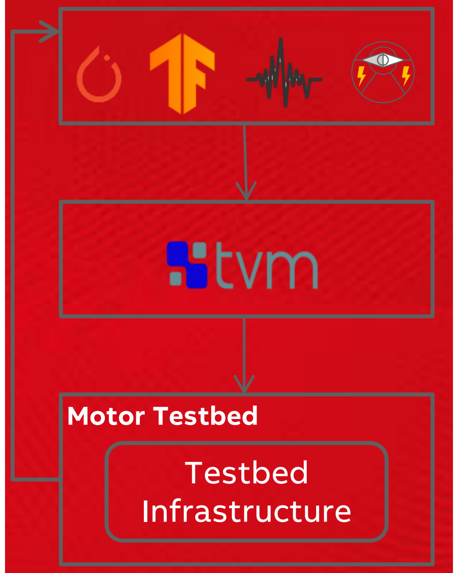
Multi-objective optimization

Multiple Objectives: Validation Accuracy, # Parameters, Sensing Energy, Sensing Latency, Inference Energy, Inference Latency → Identify ranges → Normalizing Factors (NF) → Assign Priority Factors (PF) → Weights = PF ×NF → Use Weighted Sum as Global Metric

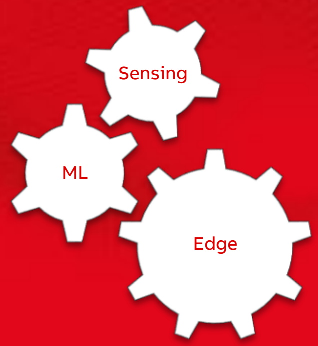


SMiLe: Automated End-to-end Sensing and Machine Learning Co-Design

SMiLe Overview



Sensing & ML Co-Design



Experimental Evaluation

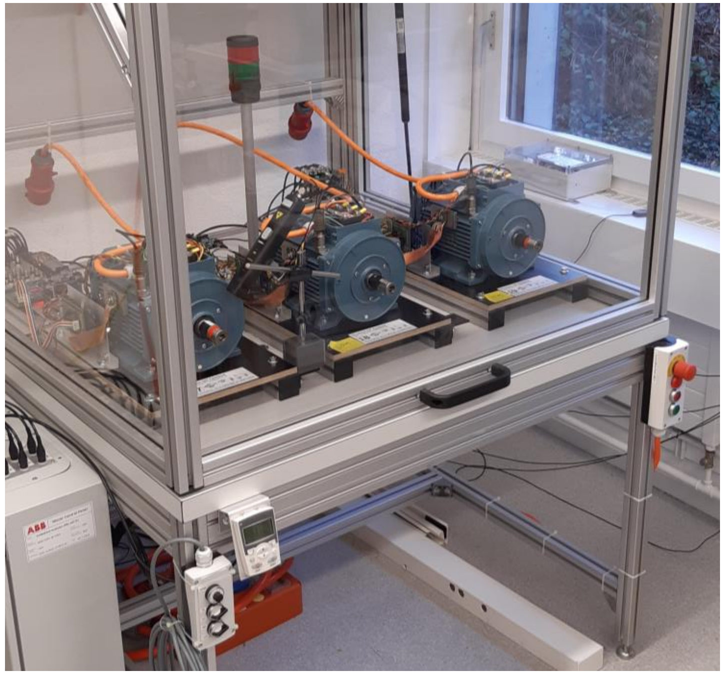
Smart Sensor

Nordic nRF5340

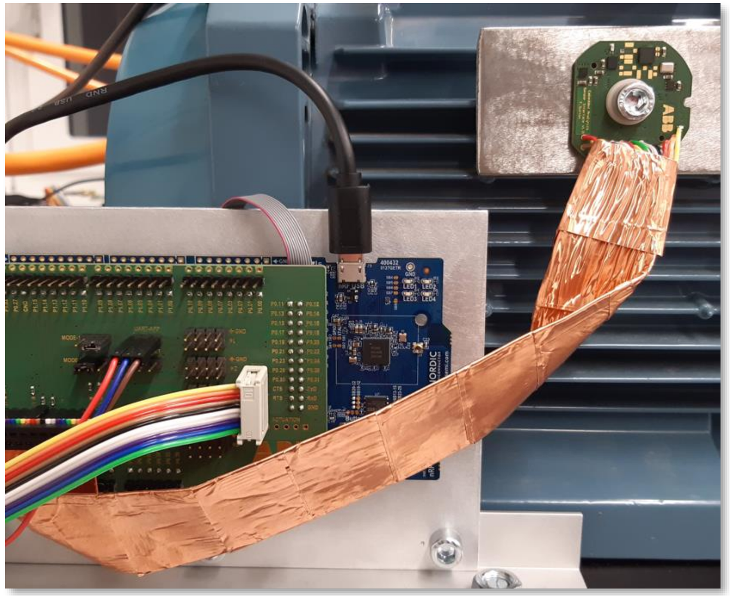
Motor Testbed @ ABB Swiss Research Center

Infrastructure

Testbed Infrastructure



Smart Sensor Connected to Testbed



Power & Timing Profiler

Motor Health Prediction

Problem Formulation

Bearing Fault Creation

Remove load-side bearing

Bearing

Add «metallic-dust» to the bearing

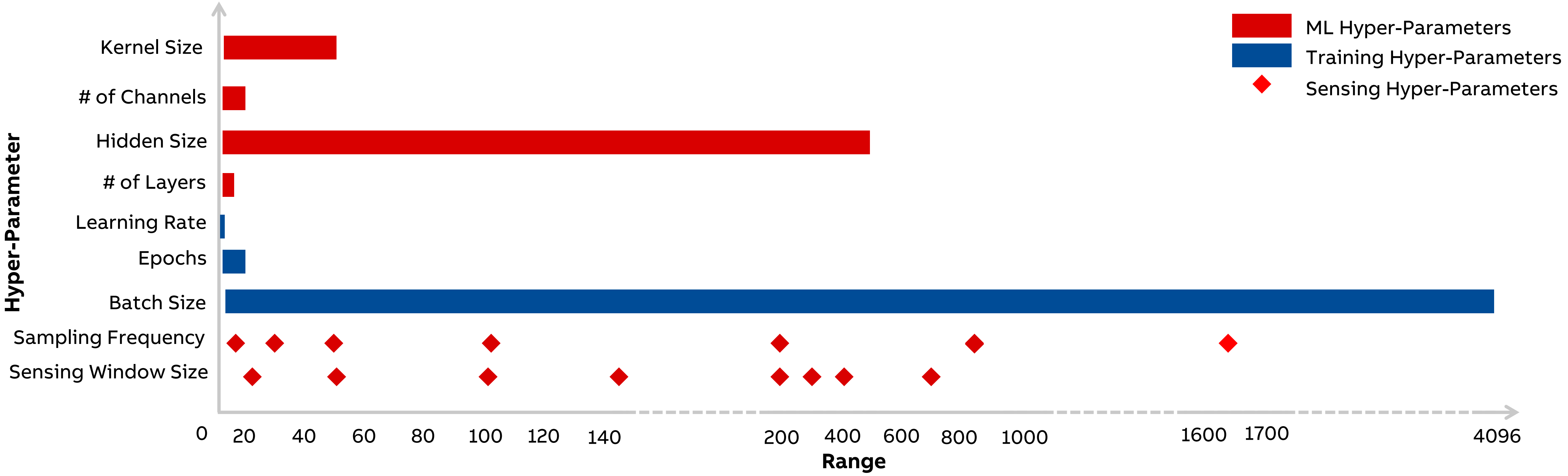
Bearings with 0g, 0.25g and 1g metallic dust

Problem

- **Type:** A 3-class classification problem (0g, 0.25g and 1g of metallic dust)
- **Input Data:** Acceleration
- **Objective:** Perform Multi-objective optimization based on Accuracy, # Parameters, Energy and Latency

Motor Health Prediction

Design Space

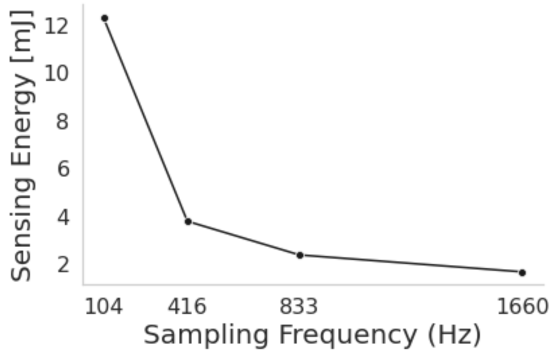


Total possible Configurations: 13,440,000,000

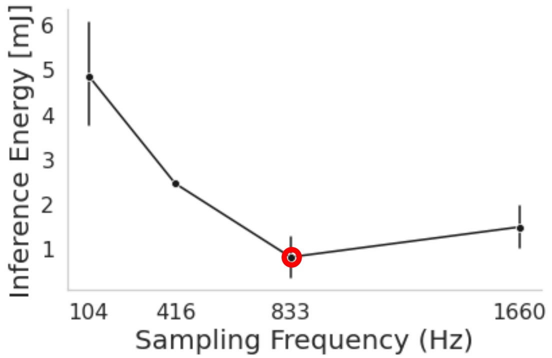
Motor Health Prediction

Sampling frequency – how often should we read the sensor?

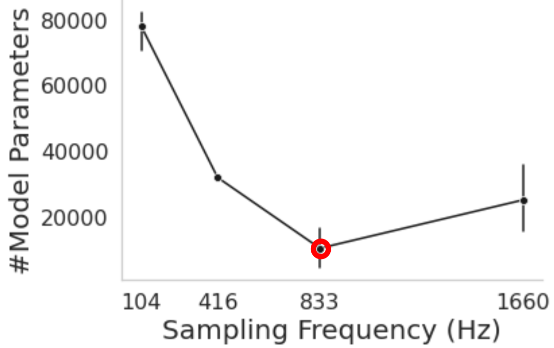
Frequency vs Sensing Energy



Frequency vs Inference Energy



Frequency vs # Parameters

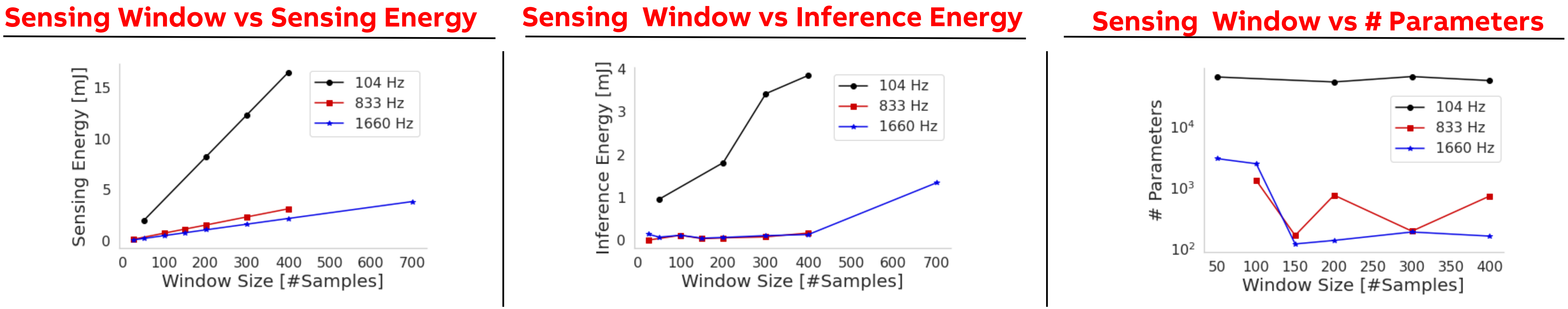


- **Sensing Energy** $\propto \frac{\mathbf{1}}{\textbf{Sampling Frequency}}$
- **Inference Energy** and **# Parameters** are highly correlated
- **Inference Energy** has a **convex** trend with Sampling Frequency; thus it has a **minima**
- Need to explore design space using ***SMiLe*** for finding the minima of **Inference Energy**

Reduce energy requirement by finding optimal sampling frequency using *SMiLe*

Motor Health Prediction

Sensing Window (# Samples) – what is a good amount of data for ML prediction



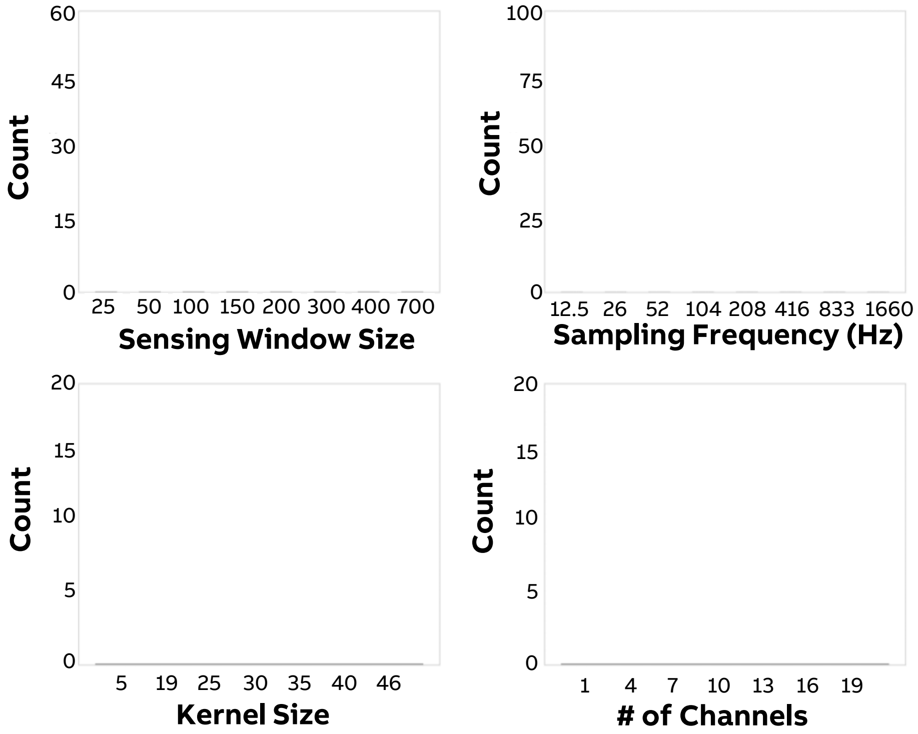
- **Sensing Energy ∝ Sensing Window Size**
- **Inference energy ∝ Sensing Window Size** (with few outliers)
- **# Parameters** follow **no clear trend** with **Sensing Window Size**
- Need to explore design space using ***SMiLe*** for finding the minimum **# Parameters**

Reduce energy requirement by decreasing Sensing Window size

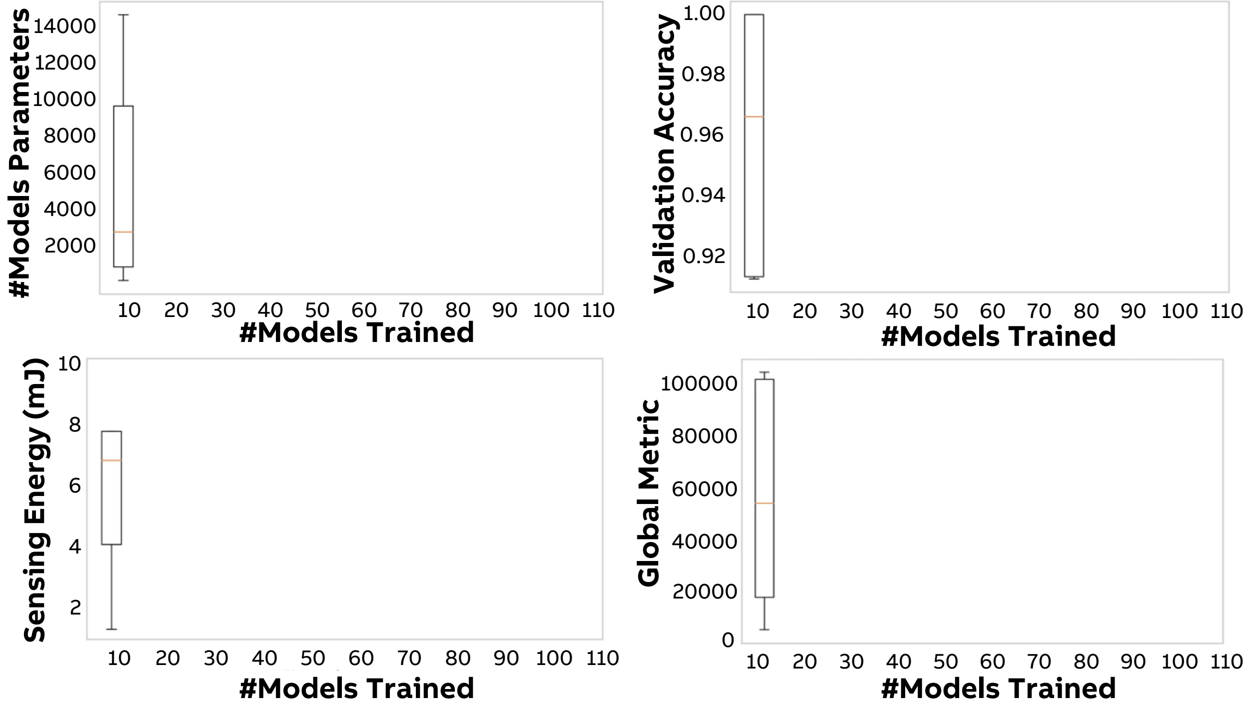
SMiLe exploration

8-hour experiment for Motor Health Prediction

Exploration of Hyper-Parameters



Evolution of various metrics

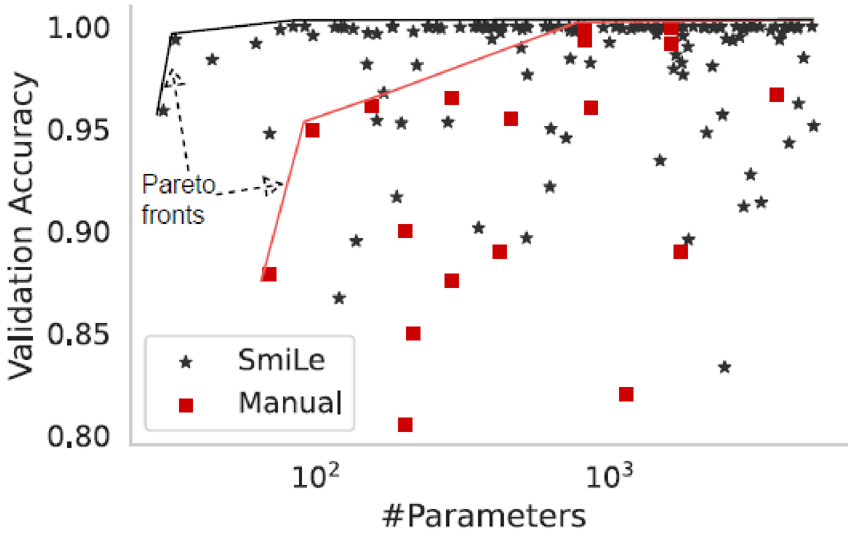


***SMiLe* design space exploration**

SMiLe for Edge Analytics

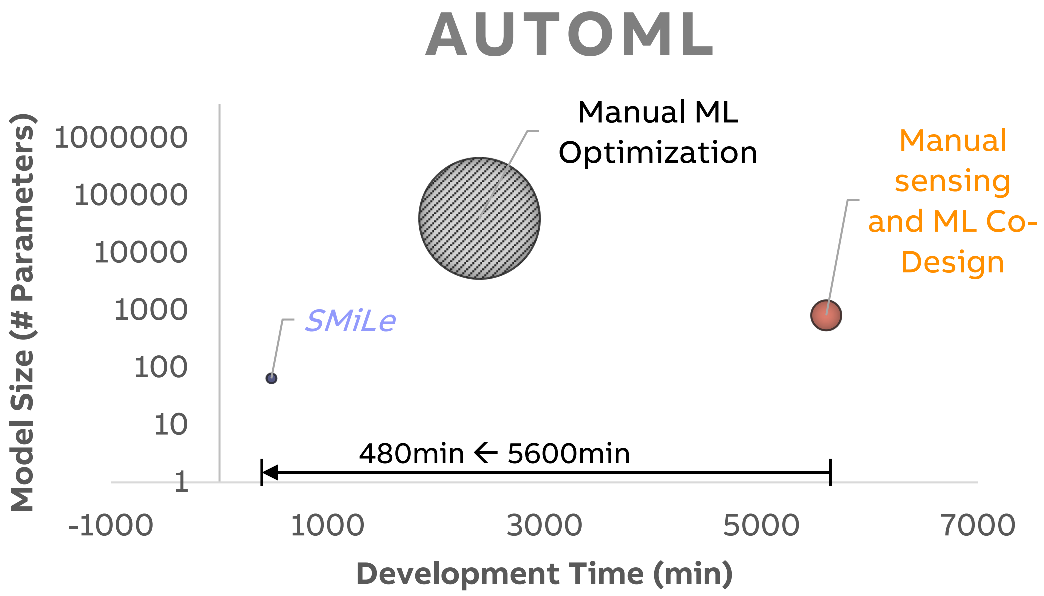
Results for Motor Health Prediction

Model improvement with *SMiLe*



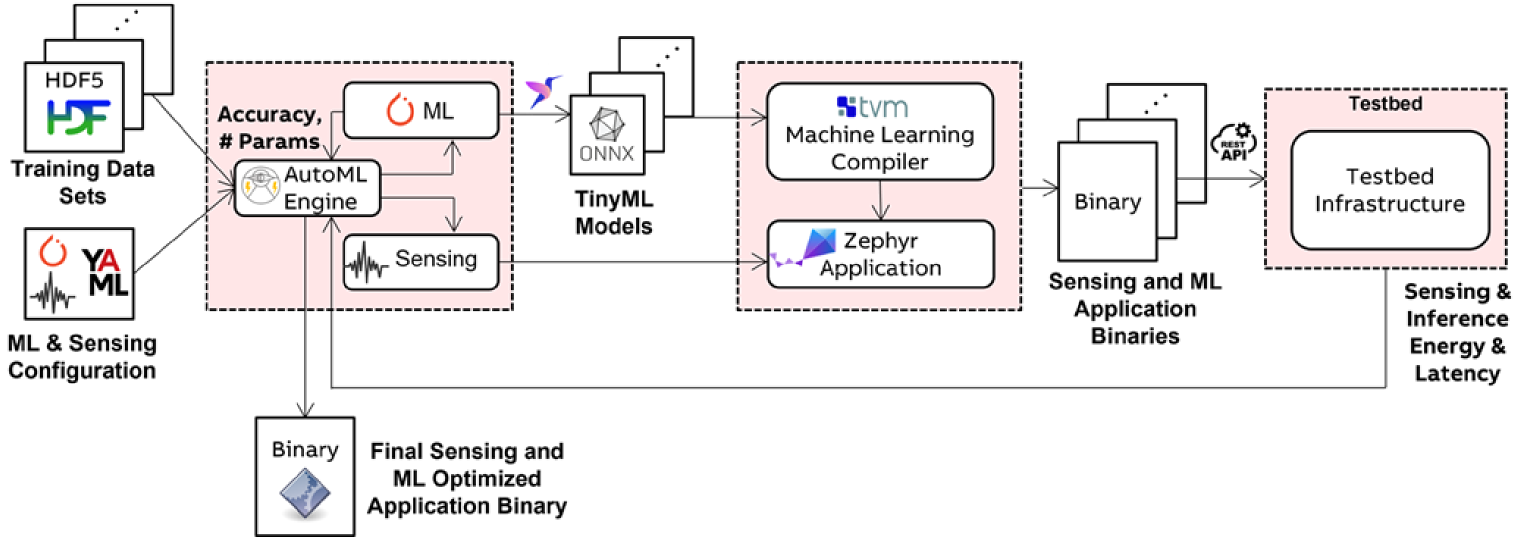
SMiLe reduces # Parameters by 95.9% for achieving similar Validation Accuracy

Development time improvement with *SMiLe*



SMiLe reduces development time by 91.5%

Conclusions



Our Contributions …

… *SMiLe* for automated optimization of sensing and machine learning Co-Design

… optimization based on real-time feedback from Hardware-in-the-loop

… validation of *SMiLe* on real-world use case – Motor Health Prediction

Our Results show that …

… sensing Hyper-Parameters are important to be optimized alongside ML

… *SMiLe* significantly reduces exploration time and improves exploration results

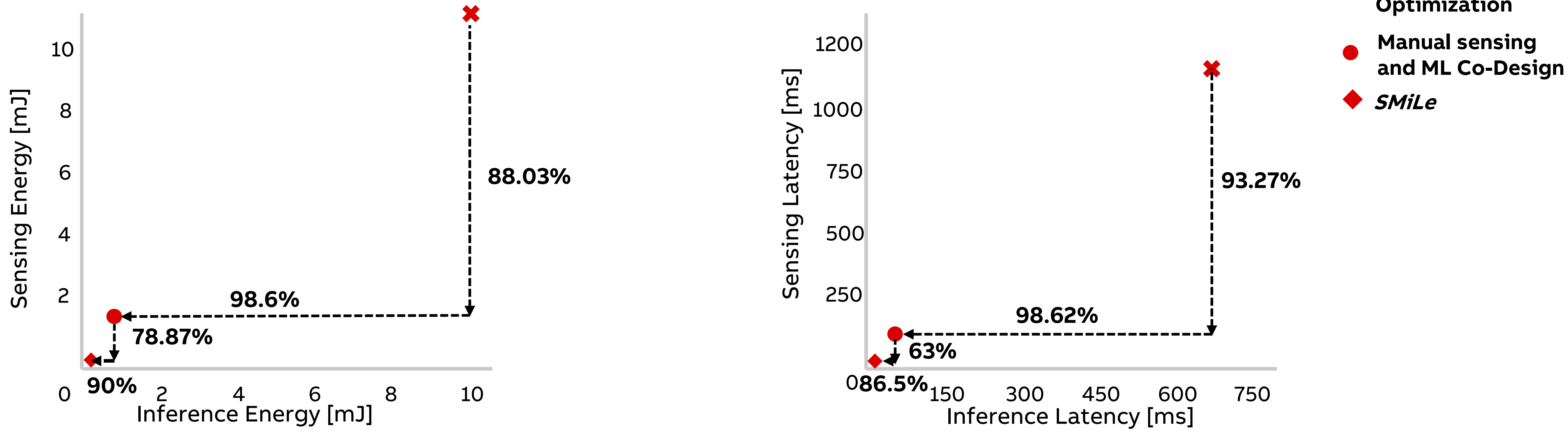
… Hardware-in-the-loop enables direct optimization of energy and latency

ABB

Motivation

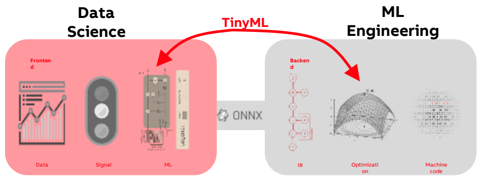
Impact of *SMiLe* on Real world use case - **Motor Health Prediction**

Training Hyper-Parameters	Batch Size, # of Epochs, Learning rate, Architecture
Model Hyper-Parameters	# of Layers, Hidden Size, # of Channels, Kernel Size
Sensing Hyper-Parameters	Sampling Frequency, Sensing Window Size



★ Real-time motor fault detection by 1-d convolutional neural networks. IEEE Transactions on Industrial Electronics, 63(11), 2016; A deep autoencoder-based CNN framework for bearing fault classification in induction motors. Sensors, 21(24), 2021; Deep Learning & its applications to machine health monitoring. Mechanical Systems & Signal Processing, 115, 2019

Energy requirements & Latency are minimum when Sensing & ML parameters are co-optimized using *SMiLe*

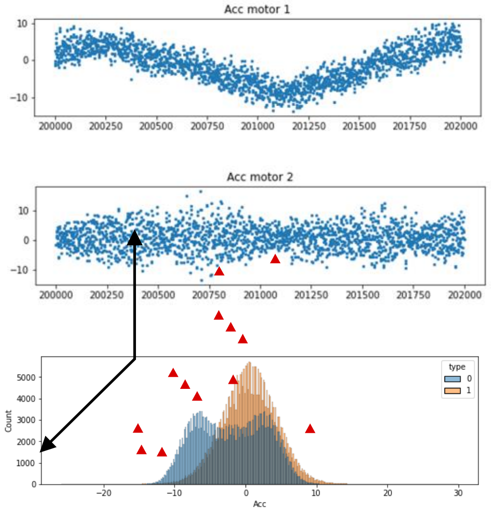


ML Data Processing Pipeline

Frontend – Analysis & Model Development

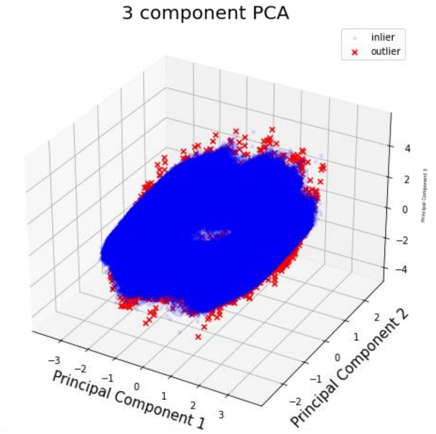
1. Data Exploration

- Visualization (e.g. histrogram/scatter/line)
- Covariance Matrix
- Dimensionality Reduction



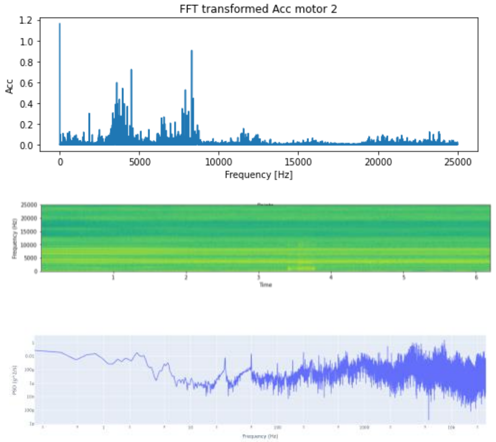
2. Data Cleaning

- Local Outlier Factor
- Isolation Forest
- Quantile



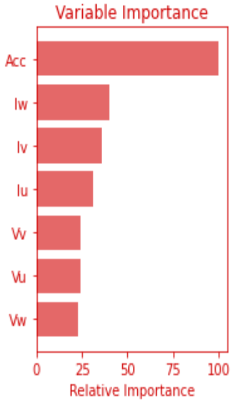
3. Spectrum Analysis

- FFT
- Spectrogram
- Power Spectrum Density



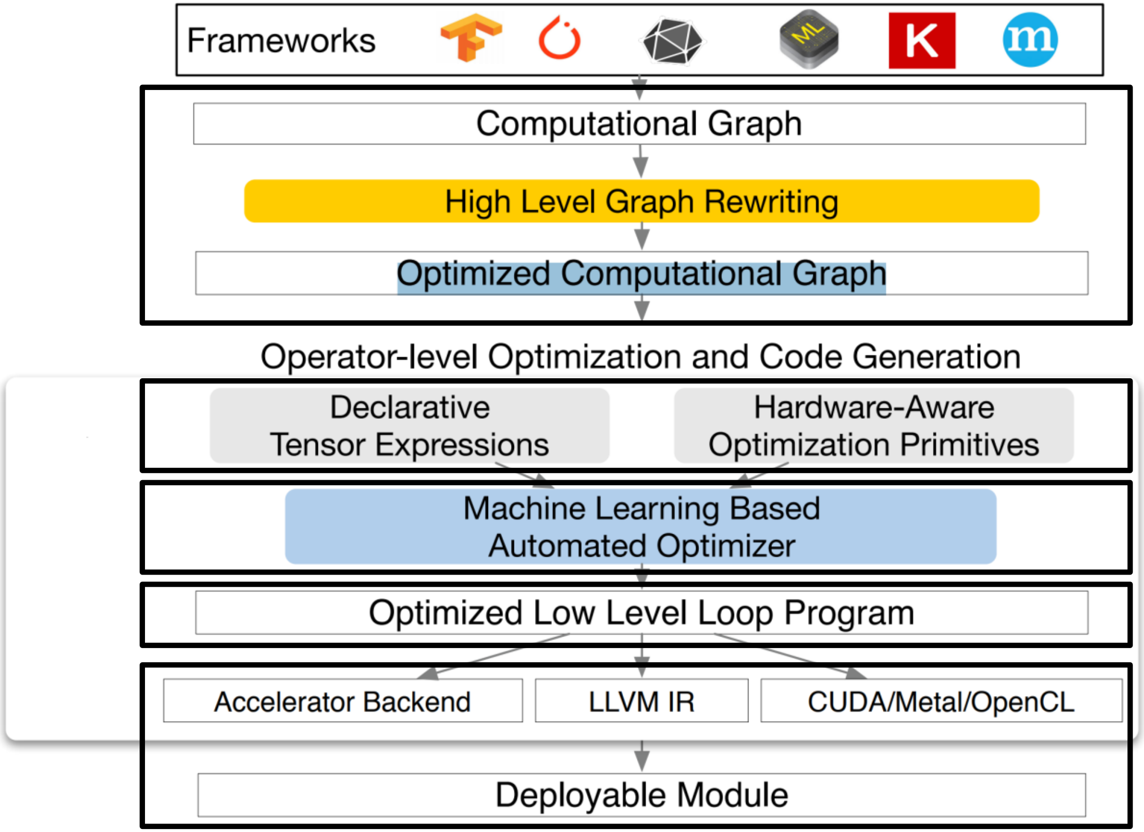
4. Machine Learning

- Logistic Regresion, LDA
- KNN, Decision Tree, Boosting
- Neural Networks



	Model	Score
0	LR	0.6242
1	LDA	0.6241
2	KNN	0.7581
3	DT	0.8055
4	AB	0.6758
5	GBM	0.7962
6	RF	0.8612
7	ET	0.8646

TVM/OctoML



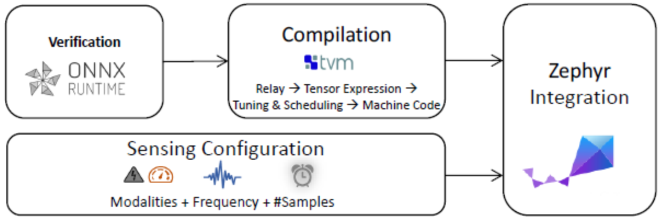


Figure 5: Overview of the *SMiLe* machine learning backend