



Towards Tiny Transformer-Based Worker Activity Recognition for Industrial Assembly

Andrea Sbaragli* and Matteo Nardello**

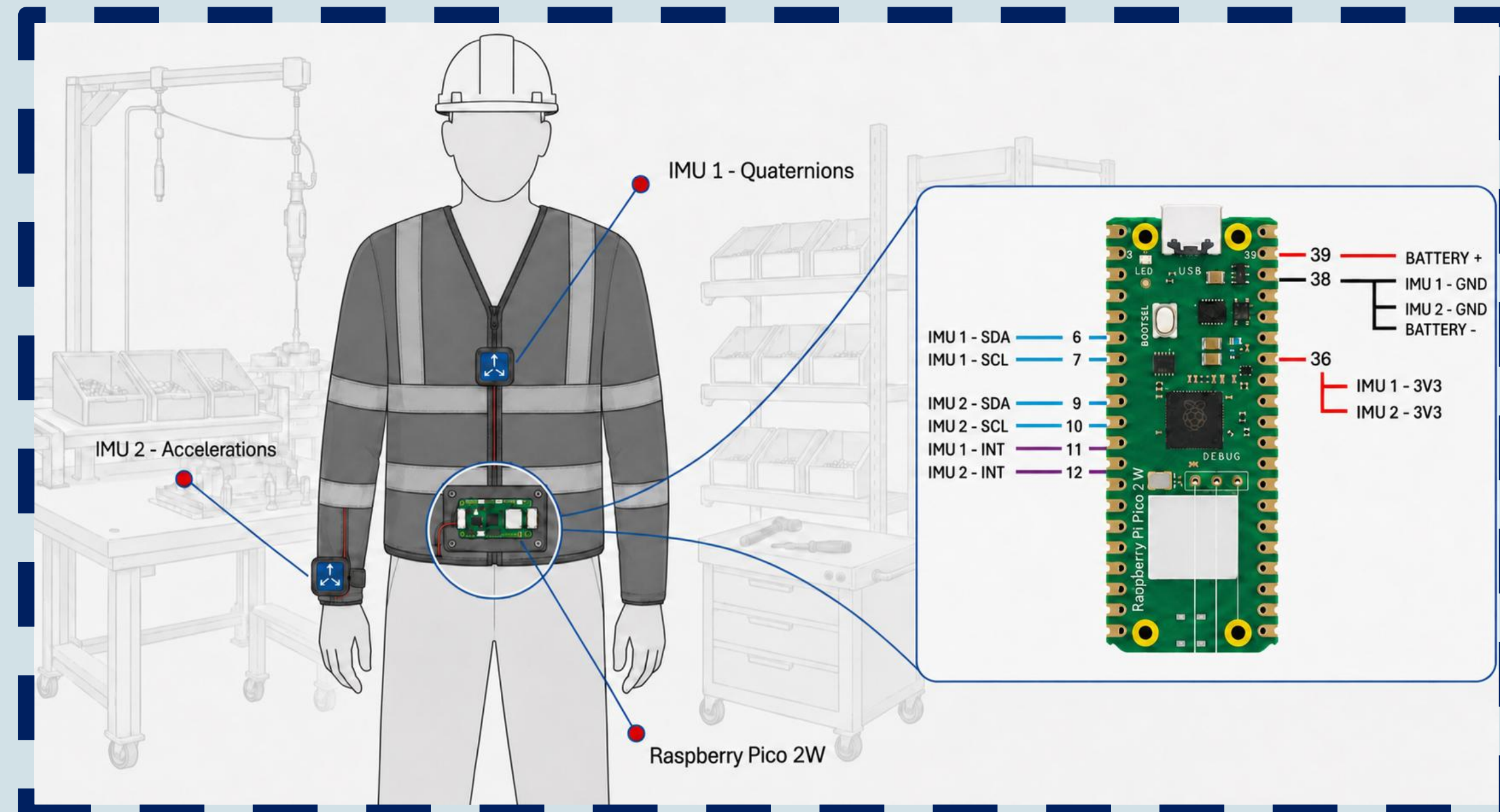
*University of Brighton, School of Architecture, Technology and Engineering, Brighton, BN2 4AT, UK

**University of Trento, Department of Industrial Engineering, Via Sommarive 9, 38123 Trento, Italy



System Implementation

Hardware



The **Raspberry Pico 2W** microcontroller is wired to two Bosch **T-BHI260AP IMUs**

- **IMU 1** provides orientation with quaternions
- **IMU 2** captures 3D acceleration of dominant hand

Transformer

Operator	Configuration
Input	$64 \times 30 \times 128$ (batch, channels, sequence)
1D convolution	$30 \rightarrow 100$, $k = 3$, $s = 1$, $p = 1$, bias
Transpose	$(64, 100, 128) \rightarrow (64, 128, 100)$
Positional encoding	Sinusoidal, $d_{\text{model}} = 100$, $L_{\text{max}} = 5000$, dropout $p = 0.1$
Multi-head self-attention	$h = 2$, $d_h = 50$; $Q, K, V : 100 \rightarrow 50$ per head; scale $1/\sqrt{50}$; output projection $100 \rightarrow 100$
Feed-forward network	$100 \rightarrow 128 \rightarrow 100$, ReLU, dropout $p = 0.1$
Residual and normalization	$\text{LayerNorm}(\text{MHA}(x) + \text{Dropout}(\text{FFN}(\text{MHA}(x))))$, $d = 100$, $\epsilon = 10^{-5}$
Encoder repetition	$N = 1$ transformer encoder layer
Temporal mean pooling	$(64, 128, 100) \rightarrow (64, 100)$; mean over sequence dimension
Linear classifier	$100 \rightarrow 5$, bias; output is unnormalized logits

Tri-axial data streams are **feature extracted**

- **Time:** Mean, Variance and Root Mean Square
- **Frequency:** Mean and Median

Dataset

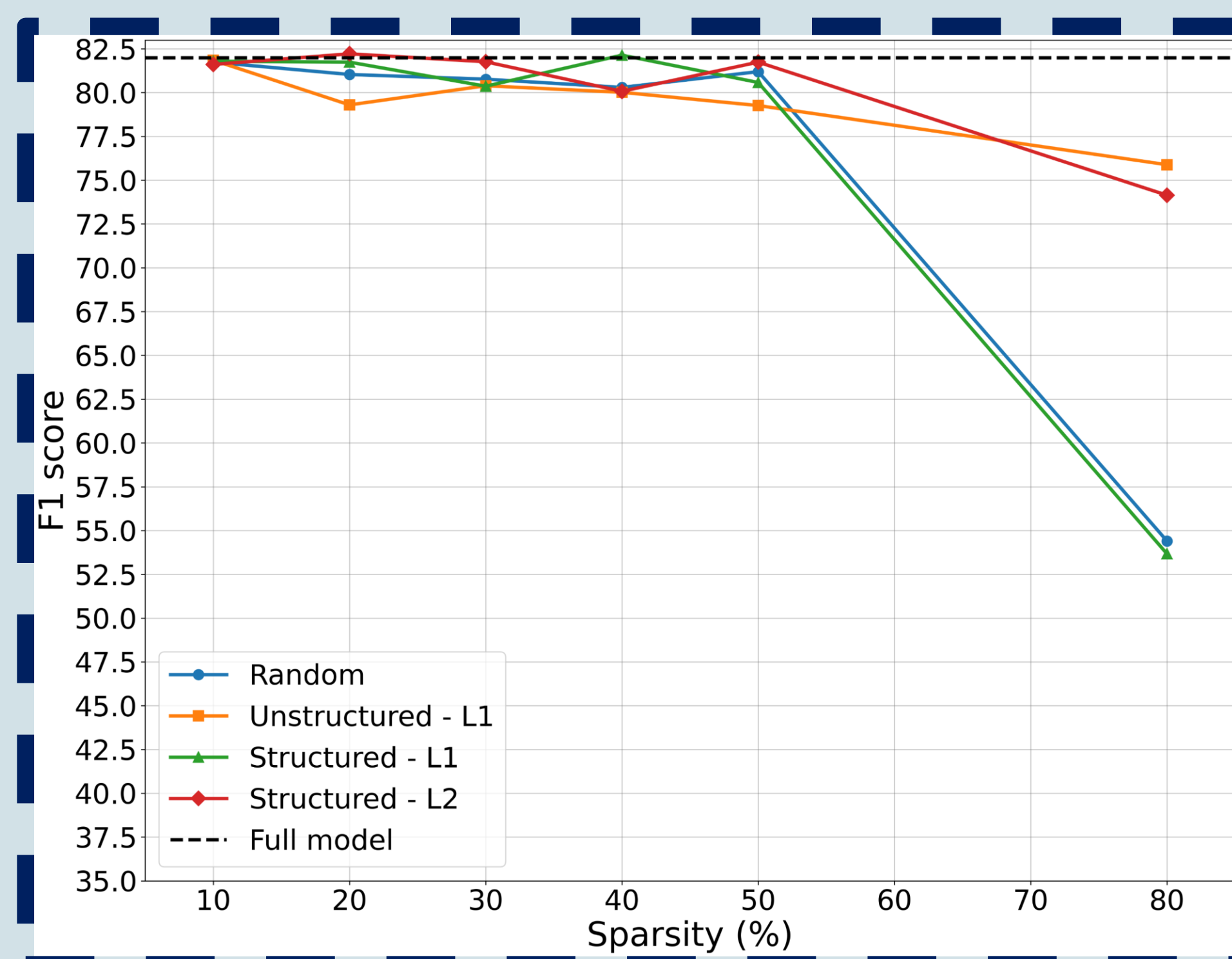
- Usage of **Philipps** Screwdriver
- Usage of **Hexagonal** Screwdriver
- Usage of **Jack** Screwdriver
- **Manual** activities and task transitions
- **Picking** and **Depositing (P/D)** from supermarket

- **3 subject** with **50 pumps assemblies**
- **7 hours** of recording with more than **1 million** raw data
- **Feature extracted** dataset of **19726** labelled samples

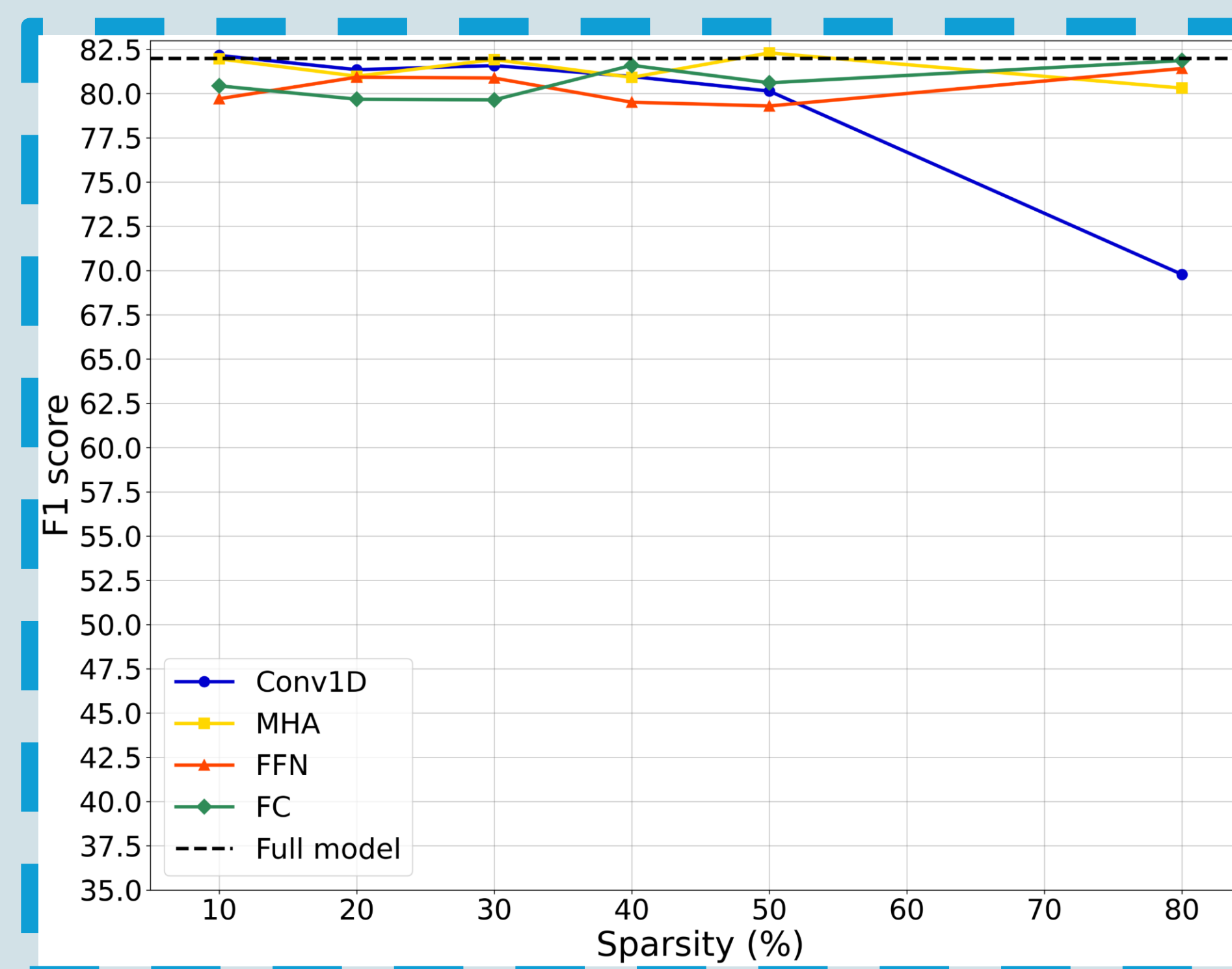
Class	P/D	Manual	Hexagonal	Jack	Phillips
Distribution [%]	14.3	26.2	19.9	19.6	20.0

Results

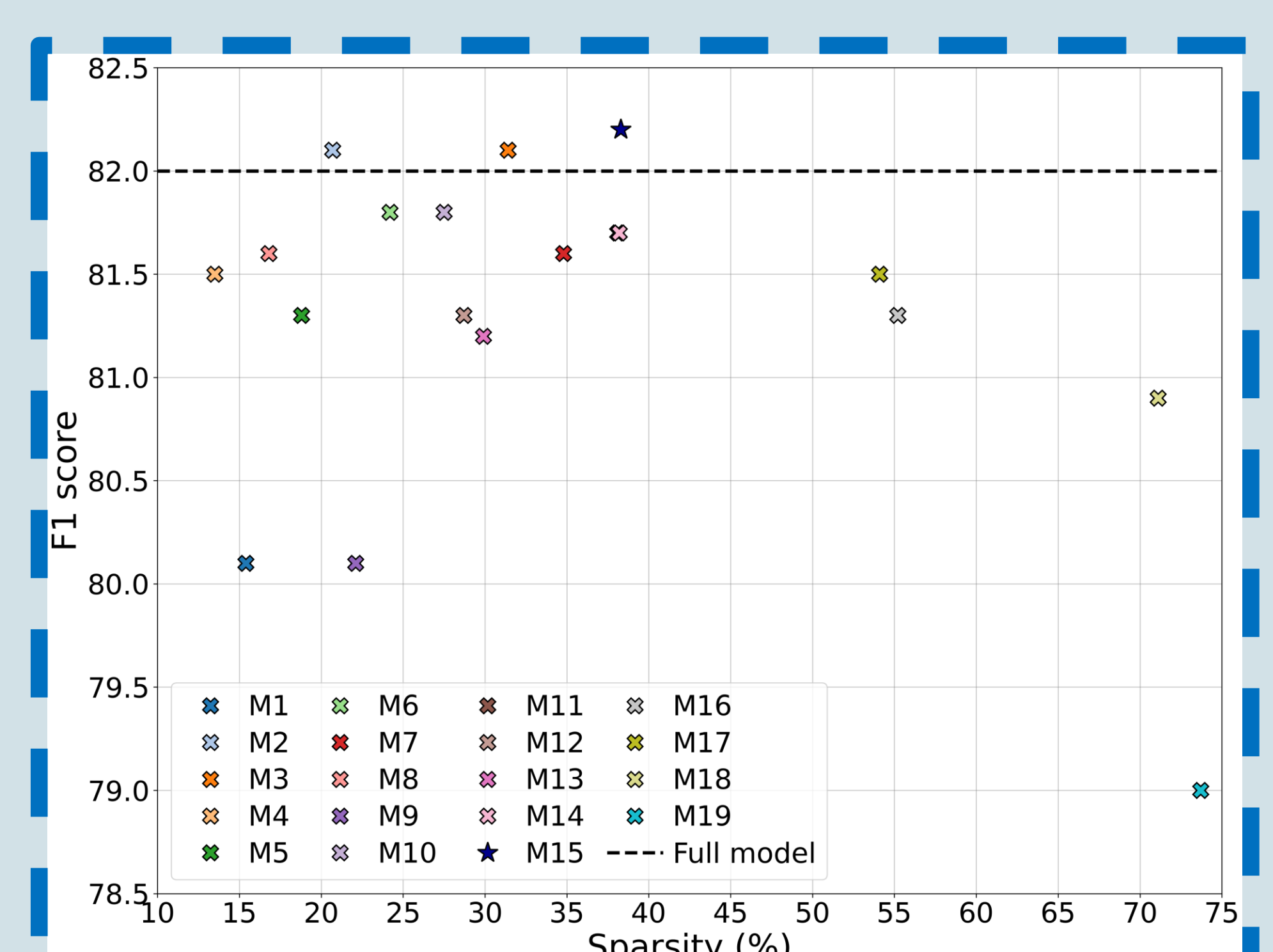
Uniform Transformer pruning



Pruning individual Transformer operators

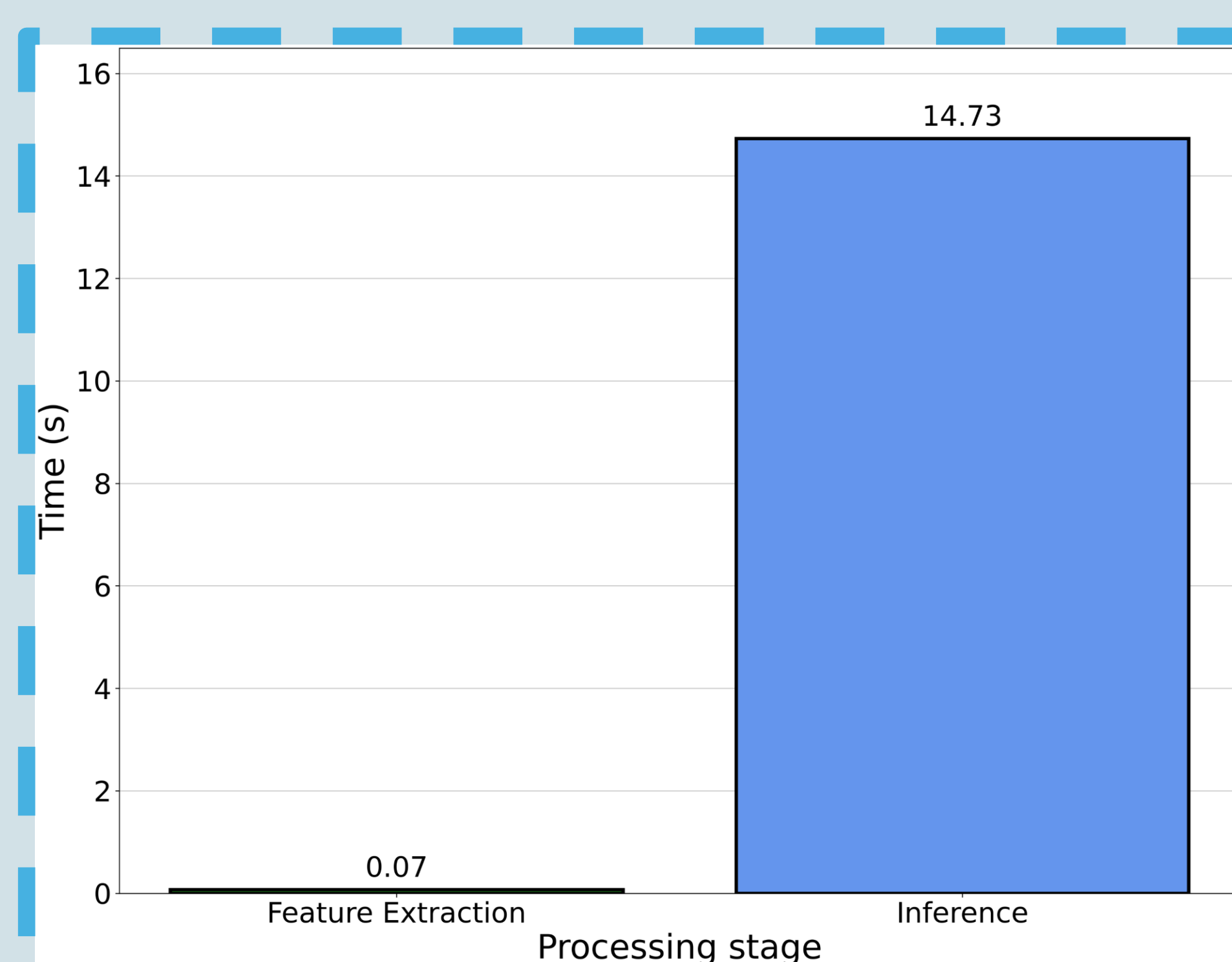


Mixed Transformer operators pruning



Edge Deployment

- **M15** mixed pruning exported
- Export with **ExecuTorch v1.2** with the **Arm Cortex-M** backend
- The graph remains with portable FP32 operators
- Artifacts have a size of **7.05 MB** for the UF2 file and **3.53 MB** for the BIN file
- Memory buffer of **272 KB** and an activation arena of **273 KB**



Performance

- **2 hours** USB serial at **52 Hz**
- Feature extraction is executed in real-time with a **mean** and **standard deviation** of **0.07 s** and **0.05 ms**
- Inference dominates the execution time with a **mean** and **max latency** of **14.73 s** and **14.77 s**
- **12 times** slower than data generation rate