tinyML® Talks

Enabling Ultra-low Power Machine Learning at the Edge

"Inference with Raspberry Pi Pico and RP2040"

Eben Upton - Raspberry Pi Foundation

March 4, 2021

www.tinyML.org

tinyML Talks Sponsors

tinyML Strategic Partner

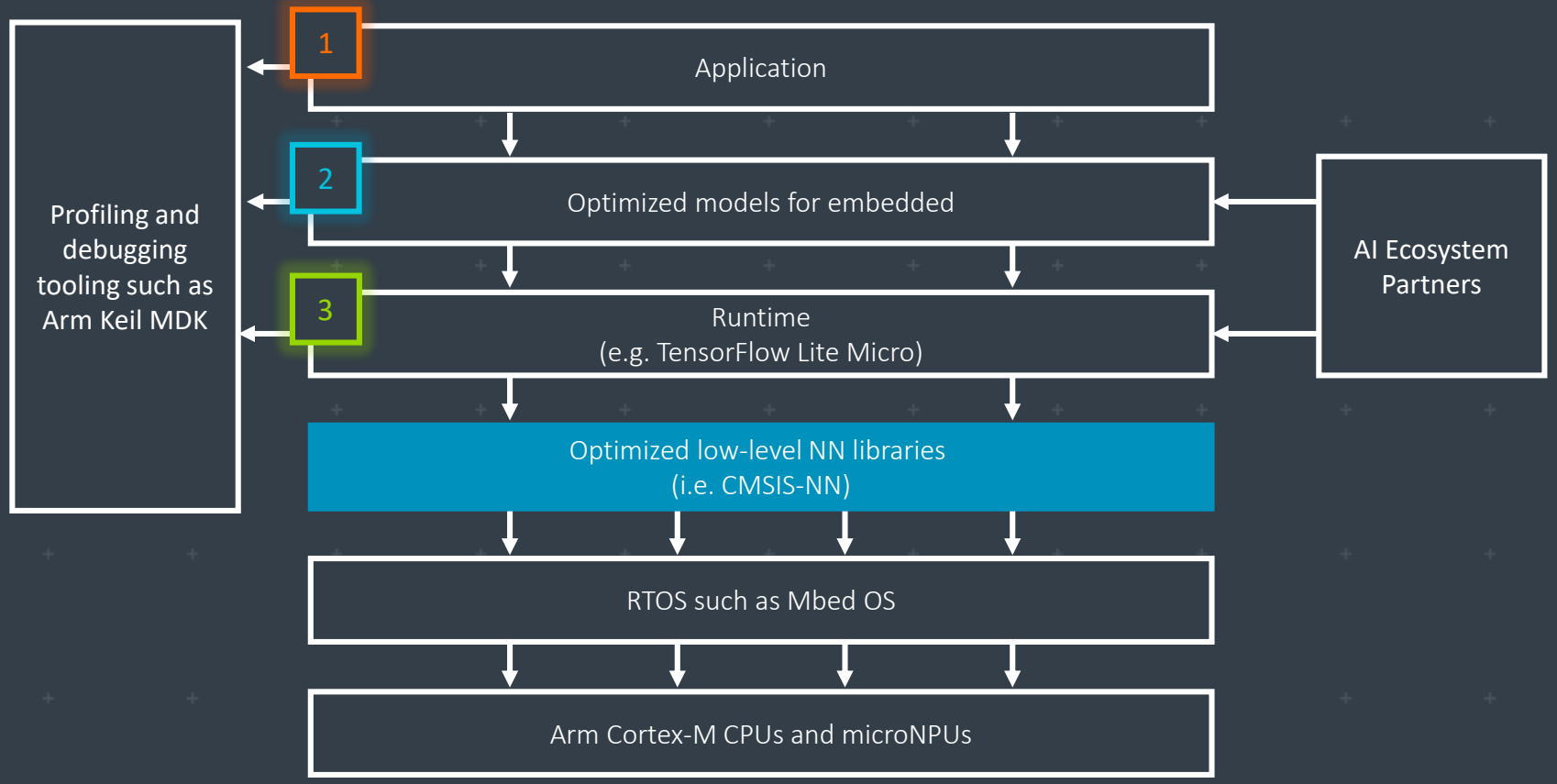
Additional Sponsorships available – contact sponsorships@tinyML.org for info

Arm: The Software and Hardware Foundation for tinyML

1 Connect to high-level frameworks

2 Supported by end-to-end tooling

3 Connect to Runtime



Stay Connected

@ArmSoftwareDevelopers

@ArmSoftwareDev

Resources: developer.arm.com/solutions/machine-learning-on-arm

arm

Deeplite

WE USE AI TO MAKE OTHER AI FASTER, SMALLER AND MORE POWER EFFICIENT

Automatically compress SOTA models like MobileNet to <200KB with **little to no drop in accuracy** for inference on resource-limited MCUs

Reduce model optimization trial & error from weeks to days using Deeplite's **design space exploration**

Deploy more models to your device without sacrificing performance or battery life with our **easy-to-use software**

BECOME BETA USER **bit.ly/testdeeplite**

mobilityXlab arm AI 100 CBINSIGHTS

TinyML for all developers

C++ library

Arduino library

WebAssembly

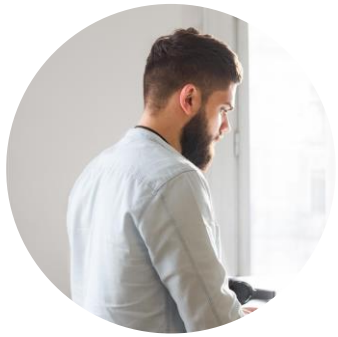
Dataset

Acquire valuable training data securely

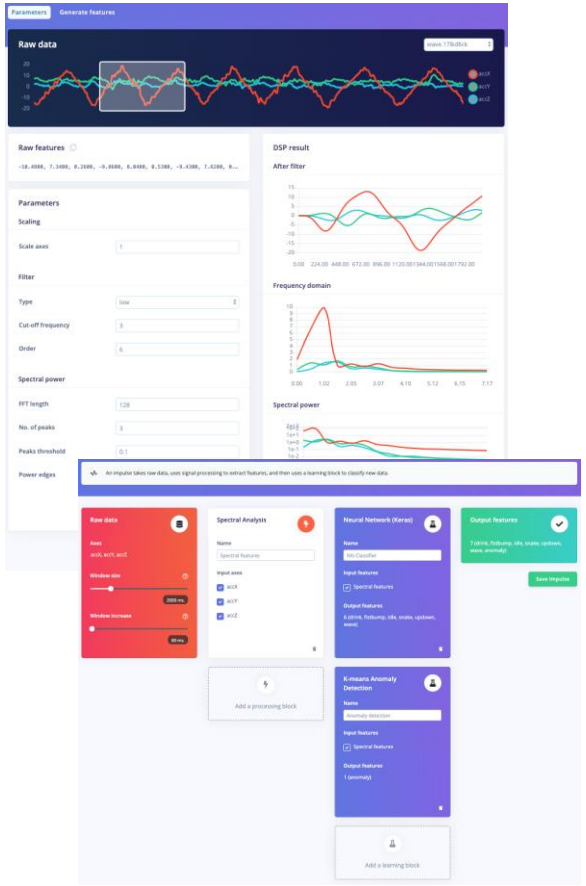
Enrich data and train ML algorithms

Edge Device

Real sensors in real time
Open source SDK



Impulse



Embedded and edge compute deployment options

Test impulse with real-time device data flows

Test

Get your free account at http://edgeimpulse.com

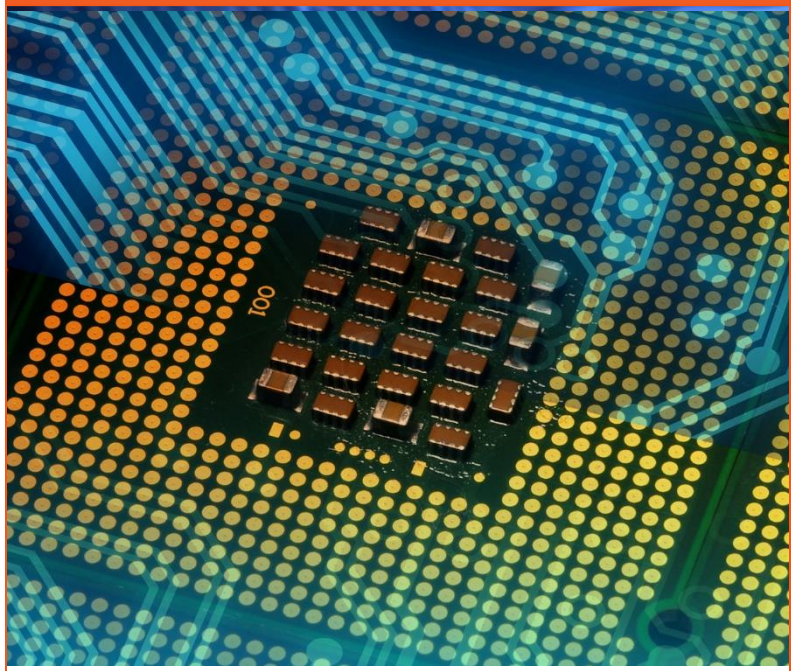
Maxim Integrated: Enabling Edge Intelligence

www.maximintegrated.com/ai

Sensors and Signal Conditioning

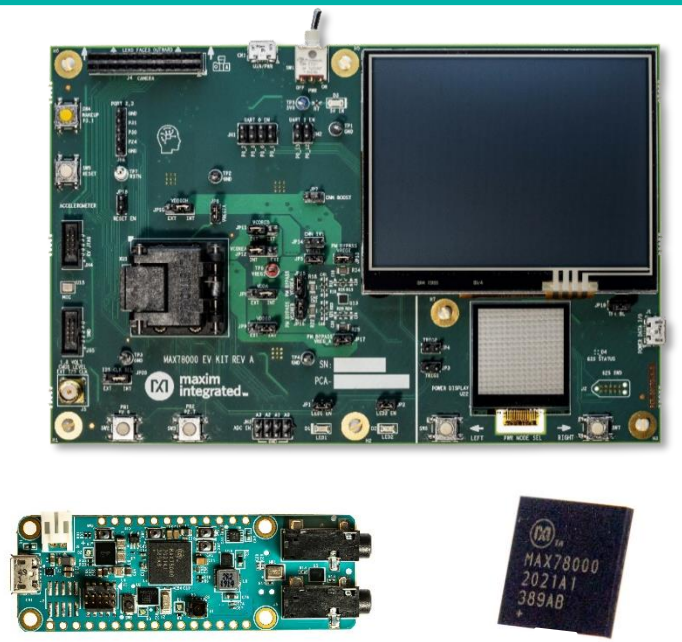
Health sensors measure PPG and ECG signals critical to understanding vital signs. Signal chain products enable measuring even the most sensitive signals.

Low Power Cortex M4 Micros



The biggest (3MB flash and 1MB SRAM) and the smallest (256KB flash and 96KB SRAM) Cortex M4 microcontrollers enable algorithms and neural networks to run at wearable power levels

Advanced AI Acceleration



The new MAX78000 implements AI inferences at over 100x lower energy than other embedded options. Now the edge can see and hear like never before.

maxim integrated™

Qeexo AutoML for Embedded AI

Automated Machine Learning Platform that builds tinyML solutions for the Edge using sensor data

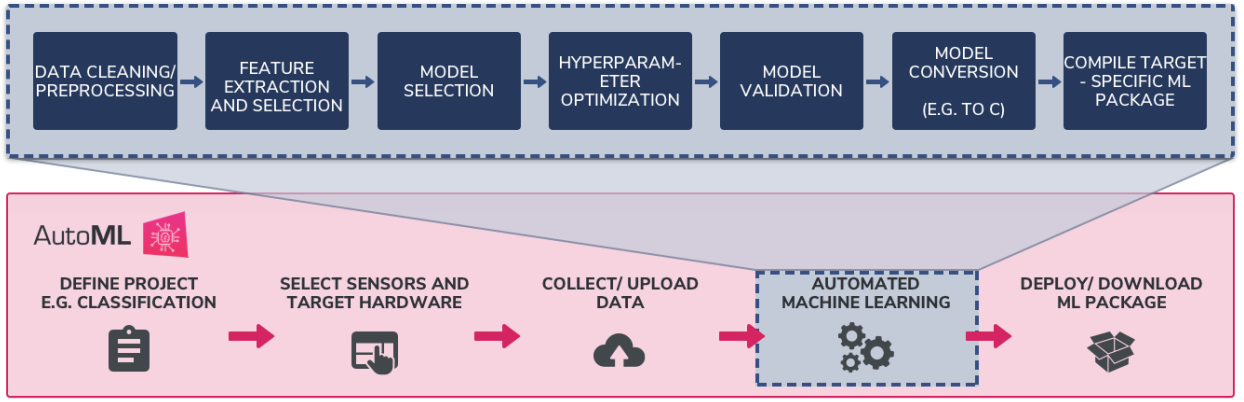
Key Features

- Wide range of ML methods: GBM, XGBoost, Random Forest, Logistic Regression, Decision Tree, SVM, CNN, RNN, CRNN, ANN, Local Outlier Factor, and Isolation Forest
- Easy-to-use interface for labeling, recording, validating, and visualizing time-series sensor data
- On-device inference optimized for low latency, low power consumption, and a small memory footprint
- Supports Arm® Cortex™- M0 to M4 class MCUs
- Automates complex and labor-intensive processes of a typical ML workflow – no coding or ML expertise required!

Target Markets/Applications

- Industrial Predictive Maintenance
- Smart Home
- Wearables
- Automotive
- Mobile
- IoT

QEEXO AUTOML: END-TO-END MACHINE LEARNING PLATFORM



For a limited time, sign up to use Qeexo AutoML at automl.qeexo.com for FREE to bring intelligence to your devices!

Reality AI®

is for

building products

https://reality.ai info@reality.ai @SensorAI Reality AI

Reality AI Tools® software

- Automated Feature Exploration and Model Generation
- Bill-of-Materials Optimization
- Automated Data Assessment
- Edge AI / TinyML code for the smallest MCUs

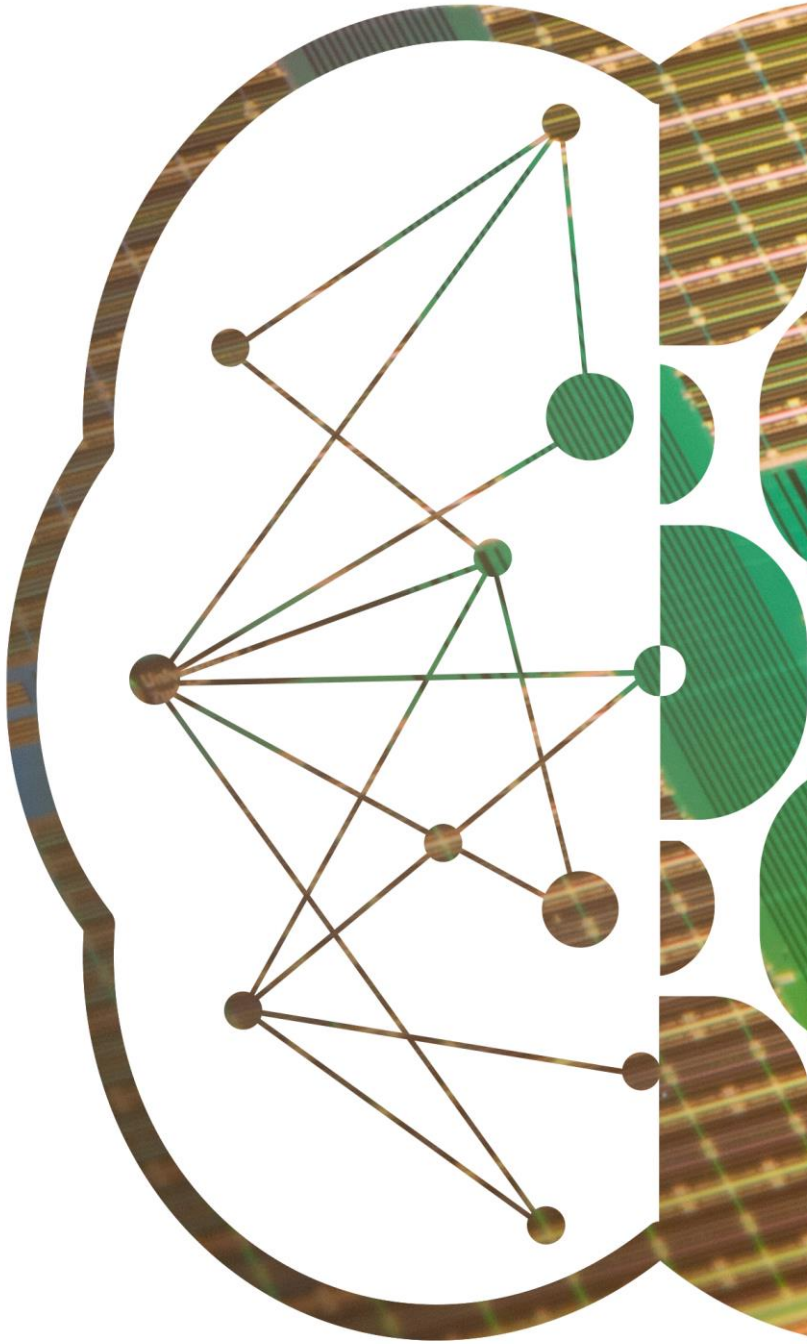
Reality AI solutions

- Automotive sound recognition & localization
- Indoor/outdoor sound event recognition
- RealityCheck™ voice anti-spoofing

SynSense

SynSense builds **ultra-low-power** (sub-mW) **sensing and inference** hardware for **embedded, mobile and edge** devices. We design systems for **real-time always-on smart sensing**, for audio, vision, IMUs, bio-signals and more.

https://SynSense.ai



Next tinyML Talks

Date	Presenter	Topic / Title
Tuesday, March 16	**Vijay Janapa Reddi** Associate Professor, Harvard University	tinyMLPerf: Deep Learning Benchmarks for Embedded Devices

Webcast start time is 8 am Pacific time

Please contact talks@tinyml.org if you are interested in presenting

Announcement

https://www.tinyml.org/event/summit-2021/

Highlights:

- Keywords: Premier Quality, Interactive, LIVE … and FREE
- 5 days, 50+ presentations
- 4 Tutorials
- 2 Panel discussions: (i) VC and (ii) tinyML toolchains
- tinyML Research Symposium
- Late Breaking News
- 3 Best tinyML Awards (Paper, Product, Innovation)
- 10+ Breakout sessions on various topics
- tinyML Partner sessions
- tinyAI for (Good) Life
- LIVE coverage, starting at 8am Pacific time

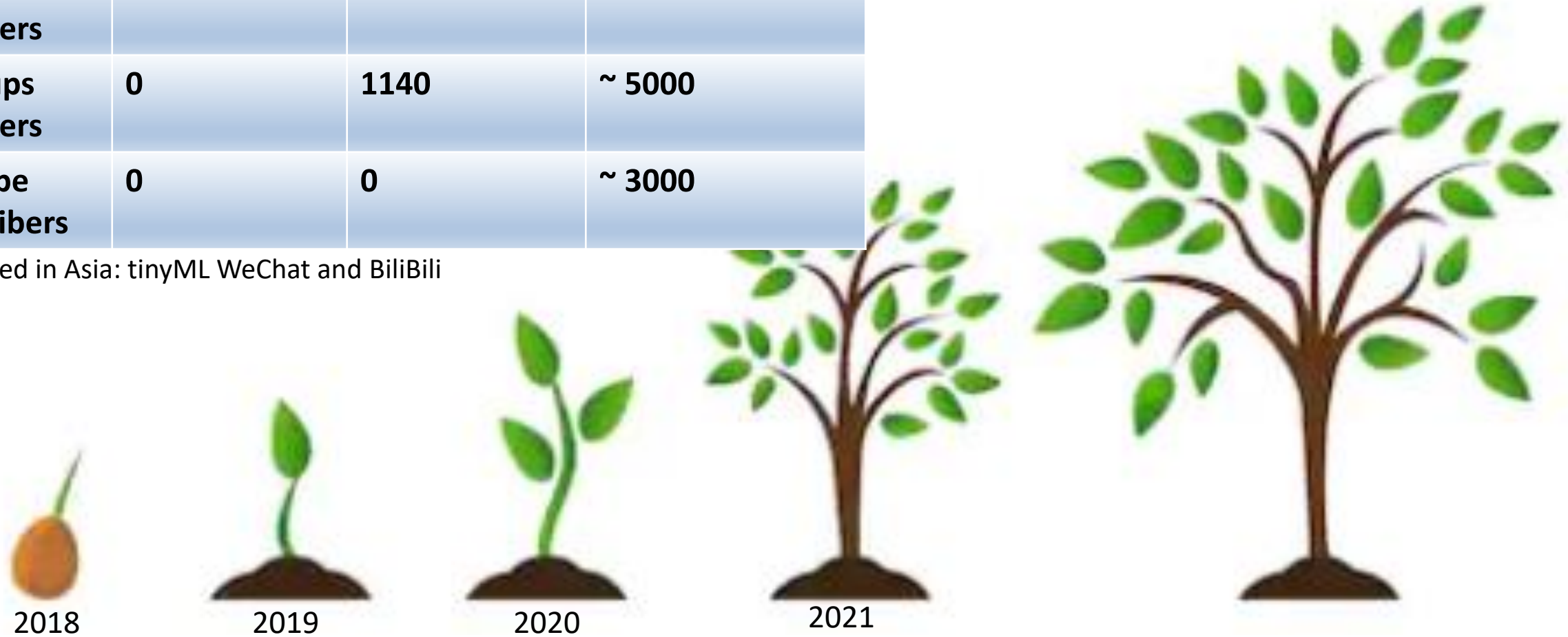
What should I do about it:

- Check out the program – you will be impressed
- Register on-line (takes 5 min)
- If interested: Submit nominations for Best Awards and/or Late News – February 28 deadline
- Block out your calendar: March 22-26
- Become a sponsor (sponsorships@tinyML.org)
- Actively participate at the Summit
- Provide your feedback – we listen !
- Don’t worry about missing some talks – all videos will be posted on YouTube.com/tinyML

tinyML is growing fast

	2019 Summit *(March 2019)*	2020 Summit *(Feb 2020)*	2021 Summit *(March 2021), expected*
Attendees	**160**	**400+**	**3000+**
Companies	**90**	**172**	**300+ (?)**
Linkedin members	**0**	**798**	**~ 2000**
Meetups members	**0**	**1140**	**~ 5000**
YouTube subscribers	**0**	**0**	**~ 3000**

also started in Asia: tinyML WeChat and BiliBili



Summit Sponsors

(as of Feb 15, 2021)

Contact: sponsorships@tinyML.org

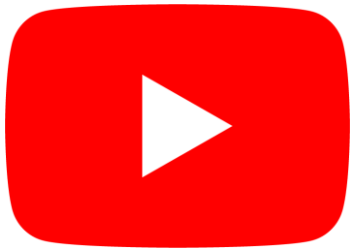
multiple levels and benefits available
(also check www.tinyML.org)

TINY ML TALKS webcast

Reminders

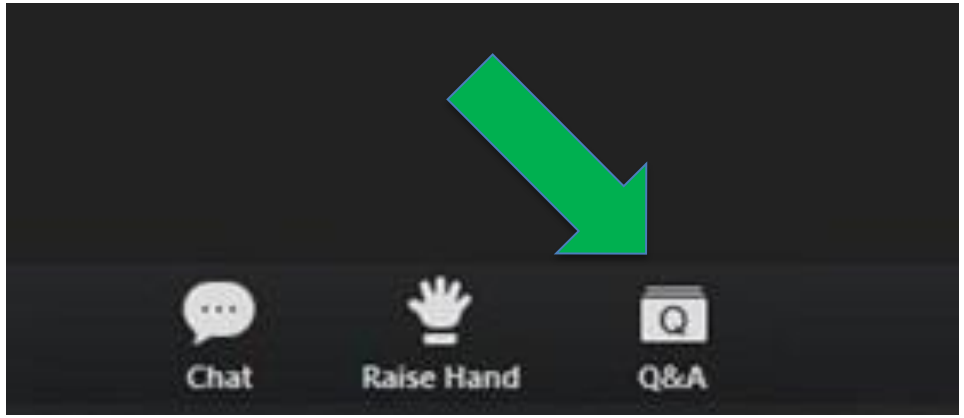
Slides & Videos will be posted tomorrow

tinyml.org/forums



youtube.com/tinyml

Please use the Q&A window for your questions



TINY ML TALKS webcast

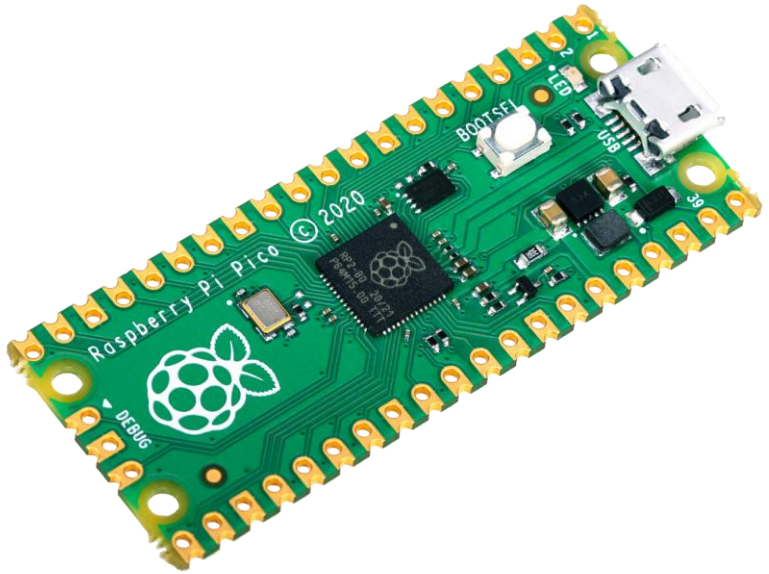
Eben Upton

Dr Eben Upton CBE FREng DFBCS HonFIET is a founder of the Raspberry Pi Foundation, a former Distinguished Engineer with fabless semiconductor manufacturer Broadcom Inc, and founder and former CTO of mobile games middleware developer Idea works 3d Ltd. He holds a BA in Physics and Engineering, a PhD in Computer Science, and an MBA, from the University of Cambridge.

TINY ML TALKS webcast

Raspberry Pi Pico

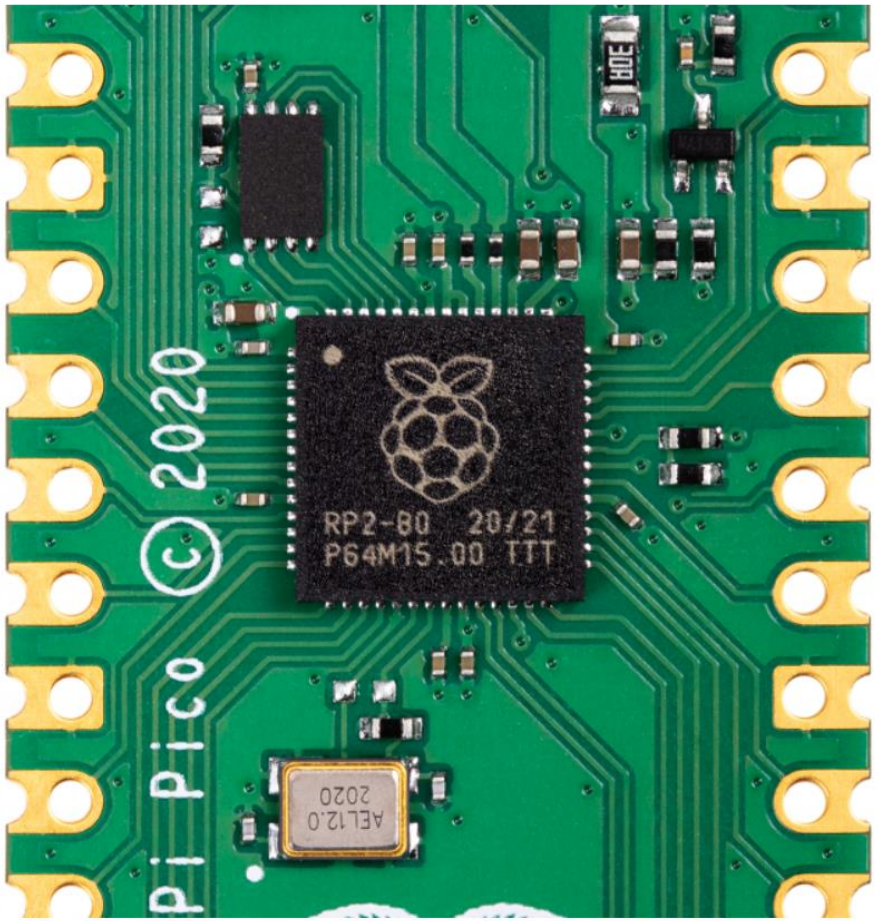
- Raspberry Pi Pico is a new $4 board
- Built on our RP2040 microcontroller
  - “Just” a break-out board
  - But with a nice power chain…
  - …and 2MB of QSPI Flash
- Showing promise as an ML platform



TINY ML TALKS webcast

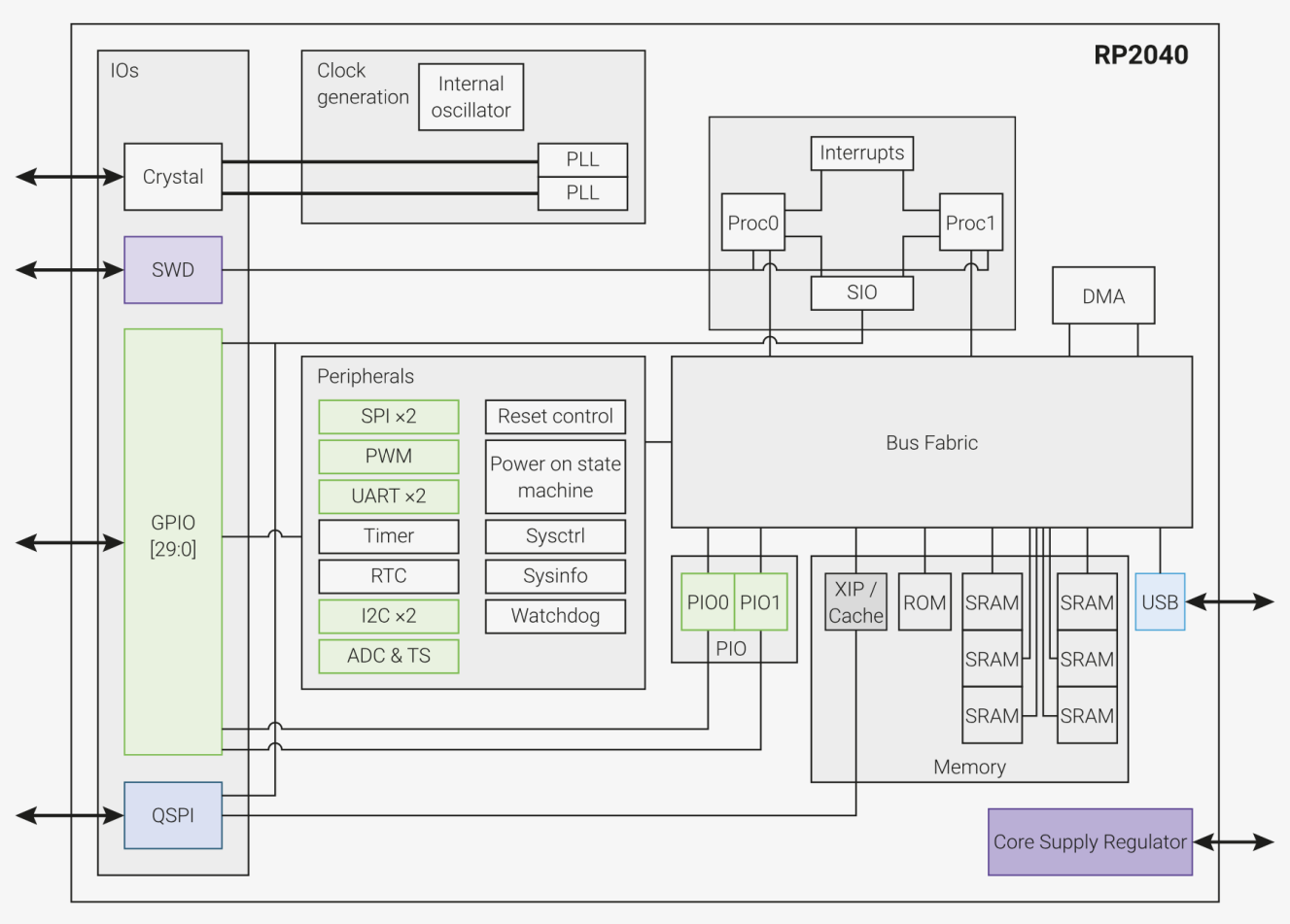
RP2040

- Our first in-house silicon design
- Dual Cortex-M0+ @ 133MHz
- 264KB on-die SRAM
- “Flashless” architecture
- Simple, deterministic bus fabric
- Rich peripheral set
  - UART, SPI, I2C
  - USB 1.1
  - Programmable I/O (PIO)
- Third-party boards also available



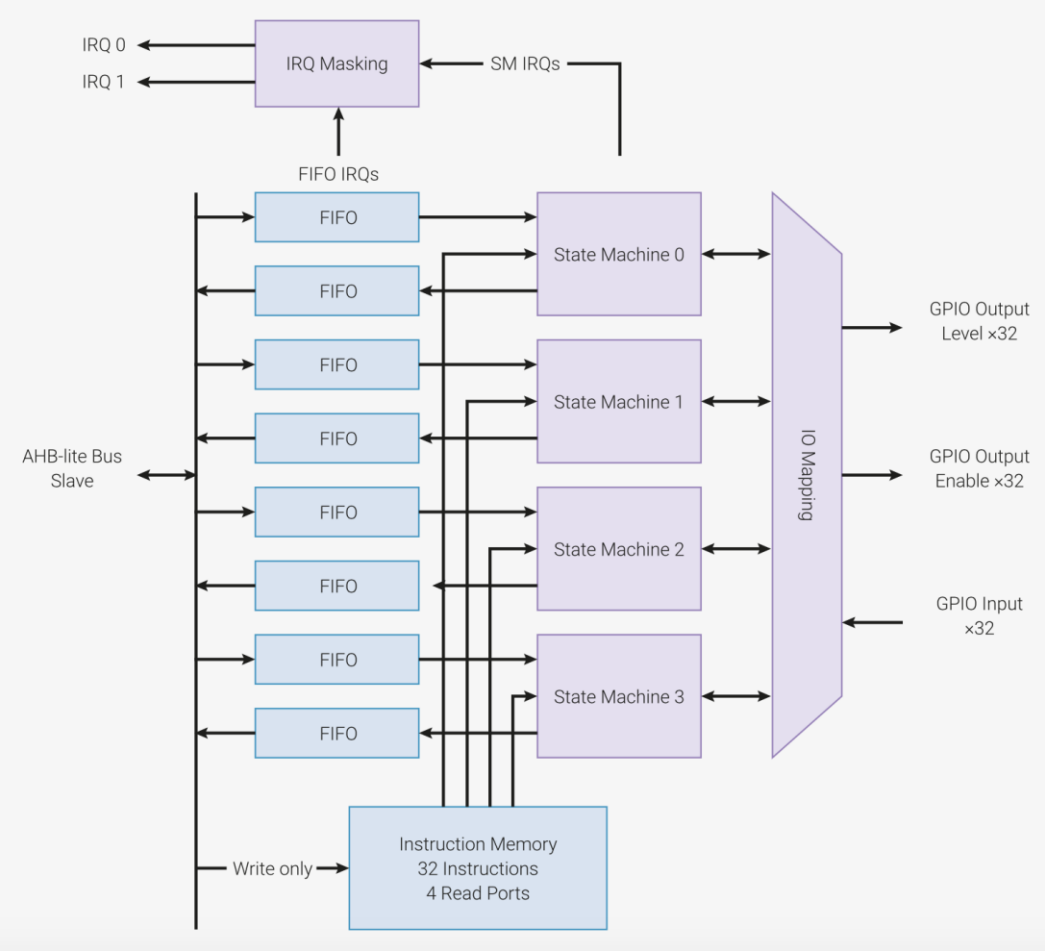
TINY ML TALKS webcast

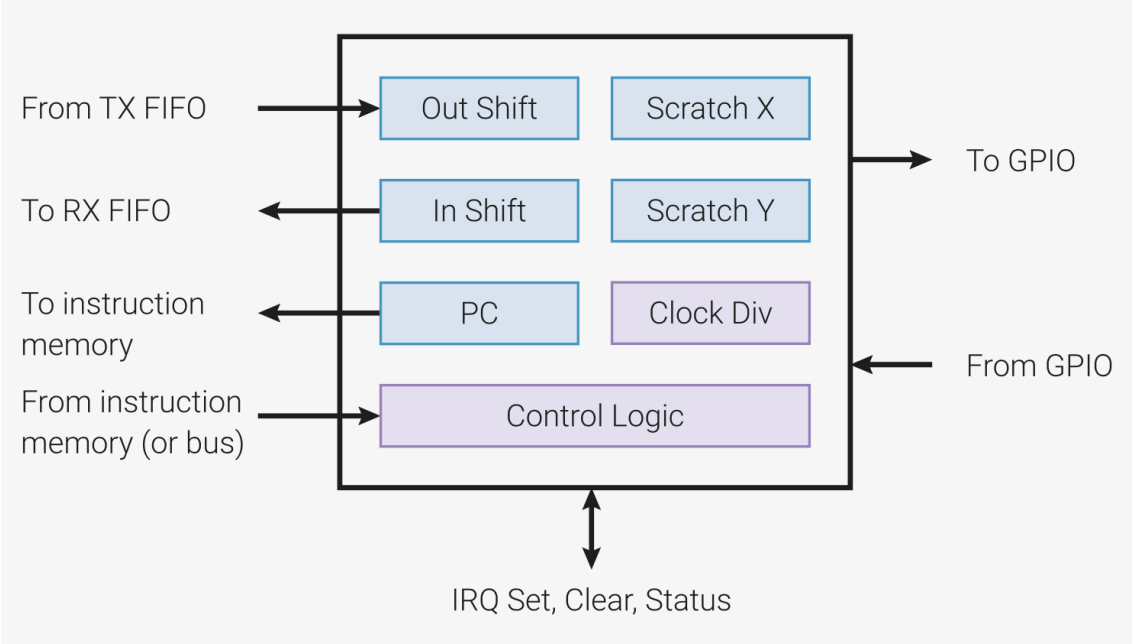
RP2040 detail



TINY ML TALKS webcast

Programmable I/O





TINY ML TALKS webcast

Early ML work

- The good
  - High clock rate
  - Dual core
  - Large on-chip SRAM
- The bad
  - No SIMD
  - No single-cycle MAC
  - Currently limited sensor choice
- Initial TensorFlow Lite port
  - Stock clocks (2×)
  - Single-core (2×)
  - Model parameters in SPI Flash (2.7×)

	keyword	person detect
SparkFun Edge (Cortex-M4 @ 48MHz)		800ms
SparkFun Edge (Cortex-M4 @ 96MHz)		400ms
Arduino BLE Sense Nano (Cortex-M4 @ 64MHz)		600ms
Raspberry Pi Pico (Cortex-M0+ @ 125MHz; model in Flash)	10.2ms	2200ms
Raspberry Pi Pico (Cortex-M0+ @ 125MHz; model in RAM)	3.8ms	

TINY ML TALKS webcast

Future directions

- ML-focused third-party boards
  - SparkFun MicroMod RP2040
  - Arduino Nano RP2040 Connect
  - ArduCam Pico4ML
- Optimised TensorFlow Lite
  - 1.2V operating point
  - Dual-core support
  - Streaming model parameters via DMA
- Other frameworks
- Future silicon
  - Lightweight (4-8MACs/clock) accelerators

Copyright Notice

www.tinyML.org