“Accelerate ML development with cloud-based Arm Cortex-M models”

Matthias Hertel - Product Specialist, Arm

March 15, 2022
tinyML Talks Strategic Partners

Additional Sponsorships available – contact Olga@tinyML.org for info
Executive Strategic Partners
Arm AI Virtual Tech Talks

The latest in AI trends, technologies & best practices from Arm and our Ecosystem Partners.

Demos, code examples, workshops, panel sessions and much more!

Fortnightly Tuesday @ 4pm GMT/8am PT

Find out more: www.arm.com/techtalks
The leading edge ML platform

www.edgeimpulse.com
Advancing AI research to make efficient AI ubiquitous

**Power efficiency**
Model design, compression, quantization, algorithms, efficient hardware, software tool

**Personalization**
Continuous learning, contextual, always-on, privacy-preserved, distributed learning

**Efficient learning**
Robust learning through minimal data, unsupervised learning, on-device learning

**Perception**
Object detection, speech recognition, contextual fusion

**Reasoning**
Scene understanding, language understanding, behavior prediction

**Action**
Reinforcement learning for decision making

A platform to scale AI across the industry
Neural Decision Processors
- At-Memory Compute
- Sustained High MAC Utilization
- Native Neural Network Processing

ML Training Pipeline
- Enables Production Quality Deep Learning Deployments

Data Platform
- Reduces Data Collection Time and Cost
- Increases Model Performance

End-to-End Deep Learning Solutions for TinyML & Edge AI

partners@syntiant.com
www.syntiant.com
Platinum Strategic Partners
Fastest Video Analytics Solutions on Arm CPUs

www.deeplite.ai
Add Advanced Sensing to your Product with Edge AI / TinyML

Pre-built Edge AI sensing modules, plus tools to build your own

Reality AI solutions
- Prebuilt sound recognition models for indoor and outdoor use cases
- Solution for industrial anomaly detection
- Pre-built automotive solution that lets cars “see with sound”

Reality AI Tools® software
- Build prototypes, then turn them into real products
- Explain ML models and relate the function to the physics
- Optimize the hardware, including sensor selection and placement

https://reality.ai  info@reality.ai  @SensorAI  Reality AI
# BROAD AND SCALABLE EDGE COMPUTING PORTFOLIO

## Microcontrollers & Microprocessors

<table>
<thead>
<tr>
<th>Arm® Core</th>
<th>Renesas Core</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Arm® Cortex®-M 32-bit MCUs</td>
<td>Ultra-low Energy 8 &amp; 16-bit MCUs</td>
</tr>
<tr>
<td>Arm® ecosystem, Advanced security, Intelligent IoT</td>
<td>Bluetooth® Low Energy, SubGHz, LoRa®-based Solutions</td>
</tr>
<tr>
<td>Arm®-based High-end 32 &amp; 64-bit MPUs</td>
<td>High Power Efficiently 32-bit MCUs</td>
</tr>
<tr>
<td>High-resolution HMI, Industrial network &amp; real-time control</td>
<td>Motor control, Capacitive touch, Functional safety, GUI</td>
</tr>
<tr>
<td>Arm® Cortex®-M0+ Ultra-low Power 32-bit MCUs</td>
<td>40nm/28nm process Automotive 32-bit MCUs</td>
</tr>
<tr>
<td>Innovative process tech (SOTB), Energy harvesting</td>
<td>Rich functional safety and embedded security features</td>
</tr>
</tbody>
</table>

### Renesas Synergy™

- Arm®-based 32-bit MCUs for Qualified Platform
- Qualified software and tools

## Core technologies

### AI

A broad set of high-power and energy-efficient embedded processors

### Security & Safety

Comprehensive technology and support that meet the industry’s stringent standards

### Digital & Analog & Power Solution

Winning Combinations that combine our complementary product portfolios

### Cloud Native

Cross-platforms working with partners in different verticals and organizations
Gold Strategic Partners
Maxim Integrated: Enabling Edge Intelligence

Advanced AI Acceleration IC

The new MAX78000 implements AI inferences at low energy levels, enabling complex audio and video inferencing to run on small batteries. Now the edge can see and hear like never before.

www.maximintegrated.com/MAX78000

Low Power Cortex M4 Micros

Large (3MB flash + 1MB SRAM) and small (256KB flash + 96KB SRAM, 1.6mm x 1.6mm) Cortex M4 microcontrollers enable algorithms and neural networks to run at wearable power levels.

www.maximintegrated.com/microcontrollers

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.

www.maximintegrated.com/sensors
LatentAI

Adaptive AI for the Intelligent Edge

Latentai.com
Build Smart IoT Sensor Devices From Data

SensiML pioneered TinyML software tools that auto generate AI code for the intelligent edge.

- End-to-end AI workflow
- Multi-user auto-labeling of time-series data
- Code transparency and customization at each step in the pipeline

We enable the creation of production-grade smart sensor devices.

sensiml.com
SynSense builds sensing and inference hardware for ultra-low-power (sub-mW) 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
Silver Strategic Partners
tinyML Summit 2022
Miniature dreams can come true…
March 28-30, 2022
Hyatt Regency San Francisco Airport
https://www.tinyml.org/event/summit-2022/

The Best Product of the Year and the Best Innovation of the Year awards are open for nominations between November 15 and March 14.

tinyML Research Symposium 2022
March 28, 2022
https://www.tinyml.org/event/research-symposium-2022

More sponsorships are available: sponsorships@tinyML.org
Our next tinyML Trailblazers Series
Success Stories with Eric Pan
(Founder, Seeed Studio and Chaihuo Makerspace)

LIVE ONLINE April 6th, 2022 at 8 am PST

Register now!
Join Growing tinyML Communities:

tinyML - Enabling ultra-low Power ML at the Edge

The tinyML Community
https://www.linkedin.com/groups/13694488/

8.5k members in 43 Groups in 34 Countries
2.7k members & 5.4k followers
Subscribe to tinyML YouTube Channel for updates and notifications (including this video)

www.youtube.com/tinyML
Next tinyML Talks

<table>
<thead>
<tr>
<th>Date</th>
<th>Presenter</th>
<th>Topic / Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wednesday, March 16</td>
<td>Vincent Kok, Soham Chatterjee and Archana Vaidheeswaran</td>
<td>TinyML SG Meet and Greet</td>
</tr>
</tbody>
</table>

Webcast start time is 4:30 am Pacific time

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

Slides & Videos will be posted tomorrow

tinyml.org(forums)  youtube.com/tinyml

Please use the Q&A window for your questions
Matthias Hertel is Product Specialist for microcontroller tools at Arm. He brings nearly two decades of experience with software development solutions for Arm-based microcontrollers, most currently exploring how to make ML and IoT workflows more efficient with cloud-native methodologies.
Accelerate ML development with cloud-based Arm Cortex-M models

Arm Virtual Hardware in TensorFlow workflows

Matthias Hertel
2022, March 15th

© 2022 Arm
A Cortex-M (ML) project

- TensorFlow Lite Micro micro_speech example on an nxp i.MXRT1064 Cortex-M7 MCU

https://github.com/ARM-software/VHT-TFLmicrospeech
A Cortex-M (ML) project software stack

https://github.com/ARM-software/VHT-TFLmicrospeech
CMSIS-Packs: Simple and safe software integration
TensorFlow CMSIS-Pack life-cycle

- DevOps powered by Arm Virtual Hardware

Integrate in Project

gh/TensorFlow/tflite-micro

gh/mdk-packs/tensorflow-pack

CMSIS Pack Index

Open-CMSIS Pack

File System
Graphics
Machine Learning
NPU Support
TensorFlow
Kernel Utils
Kernel
Testing
Native Driver
Network

File Access on various storage devices
User Interface on graphical LCD displays
Software Components for Machine Learning
Driver for Ethos-U5
Tensorflow Lite Micro Library
Tensorflow Lite Micro Library
Tensorflow Lite Micro Library
IPv4 Networking using Ethernet or Serial protocols

Amazon EC2

arm Virtual Hardware
TensorFlow CMSIS-Pack CI using GitHub Actions

Build CMSIS-Pack from latest or specified upstream source

Create test projects

Execute test suite on Virtual Hardware at AWS

Use GitHub action plugin to build and execute test projects on Arm Virtual Hardware

https://github.com/mdk-packs/tensorflow-pack
Cloud-based test infrastructure for CI automation

REGRESSION TESTING

Arm Virtual Hardware for unit or integration tests offer several significant benefits:
- **Speed**: no flash download overhead.
- **Scale**: run many tests in parallel.
- **Reliable**: no failure in case of misuse
- **Flexible**: execute on local computer or cloud server

<table>
<thead>
<tr>
<th>Test Platform</th>
<th>Time *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware Board</td>
<td>488s</td>
</tr>
<tr>
<td>Virtual Hardware (single system)</td>
<td>259s</td>
</tr>
<tr>
<td>Virtual Hardware (two in parallel)</td>
<td>160s</td>
</tr>
</tbody>
</table>

* Execution time measured with 124 test cases

Project Environment in the Cloud

Software Development Tools
- Arm C/C++ Compiler
- CMSIS-Build, CMake, Python ...
- Arm VHT System Simulation Model

Resources
- CMSIS Software Packs
- RTOS, IoT connectors, ...
- GitHub repositories
- application code, test cases

Flexible usage during Development Process

IDE
Local software development and debugging of application code and test cases.

CLOUD
Trigger CI testing on Git commits. This requires a virtual test and tool environment.
Arm Virtual Hardware

- Precise simulation models of Cortex-M device subsystems designed for complex software verification and testing
- Runs any RTOS or bare metal code
- Provides virtual peripheral interfaces for I/O simulation
- Enables test automation of diverse software workloads, including unit, integration tests, and fault injection
- Cloud service that can be integrated in CI/CD and MLOps development flows

Virtual Hardware Systems of a Cortex-M Processors
- Cortex-M
  - TrustZone
  - SIMD
  - Helium
- Ethos-U65/U55 microNPU

Memory
- Secure/Non-secure
- DMA

Peripherals
- GPIO
- UART, SPI, I²C
- Ethernet

Virtual I/O
- Data values
- Streaming
- BSD-Socket

Debug Interface
- MDK, DS
- GDB
- Event Recorder

Developer Resources
- I/O drivers
- Test scripts
- CI/CD integration
- Usage examples
- Test report tools

AWS Cloud Service
- Arm VHT Systems
- C/C++ Compiler
- Build utilities
GitHub Actions unit test results

Nightly CI using Arm Virtual Hardware Nightly CI using Arm Virtual Hardware #79

GitHub Actions / Unit Test Results
succeeded on 19 Dec 2021 in 0s

All 468 tests pass in 0s

60 files  60 suites  60s
468 tests  468 ✓  0 ✘
471 runs  471 ✓  0 ✘

Results for commit a44b4e5.

ANNOTATIONS

1 Check notice on line 0 in .github

github-actions / Unit Test Results
468 tests found

There are 468 tests, see "Raw output" for the full list of tests.

Raw output
Tensorflow CMSIS-Pack tested variants - today

- Compilers:
  - Arm Compiler
  - GCC
- Kernel variants:
  - Reference
  - CMSIS-NN
  - Ethos-U
- Cores tested:
  - Cortex-M0/Cortex-M0+
  - Cortex-M3
  - Cortex-M4
  - Cortex-M7
  - Cortex-M33 / Cortex-M55
Arm Virtual Hardware scaling

GitHub Actions Jobs launch Arm Virtual Hardware instances
Machine Learning (ML) requires real-world data

MLOps aims to deploy and maintain ML models in production reliably and efficiently

- Correct decisions can only be made in areas where training data exist
- Learning means that ML algorithms are retrained based on new “experience”
- For effective ML algorithm selection, training, and validation, a large set of representative and qualified is a pre-requisite

* Arm Virtual Hardware Targets simplify these steps.
Virtual Streaming Interface

Flexible I/O interfaces for test automation supports a wide range of use cases.

The streaming peripheral is flexible and allows to implement a wide range of use cases.

Also, the Cortex-M side implements a flexible streaming peripheral that can serve a wide range of use cases.

In the examples we provide the same Audio Peripheral Driver API implemented on a real microcontroller.
IDE development
Local installation

Keil MDK with:
• Interactive Debug and Trace Views
• Arm Compiler
• Arm Virtual Hardware

Create and Debug

Test Automation

CI development
Jenkins or GitHub

AWS Cloud Service with:
• CMSIS-Build
• Arm Compiler
• Arm Virtual Hardware

Version Control System
Code Repository

Multiple Builds
Regression Tests
Verification Results

Developer 1
Local Builds

Developer 2
Local Builds

Developer ‘n’
Local Builds

https://github.com/ARM-software/VHT-GetStarted
Get execution details with Event Annotations

```c
void loop() {
    EventStartCv(0, current_time, previous_time);
    TfLiteStatus feature_status = feature_provider->PopulateFeatureData(error_reporter, previous_time, current_time, ...);
    EventStopCv(0, feature_status, how_many_new_slices);
    EventStartCv(1, current_time, how_many_new_slices);
    TfLiteStatus invoke_status = interpreter->Invoke();
    EventStopCv(1, invoke_status, 0U);
    EventStartCv(2, current_time, 0U);
    TfLiteStatus process_status = recognizer->ProcessLatestResults(output, current_time, &found_command, &score, &is_new_command);
    EventStopCv(2, process_status, score);
```
Performance profiling using EventRecorder

Tensorflow Inference call measured

interpreter->Invoke()
Resources


- Get Keil Studio Cloud: https://www.keil.arm.com/

- Get started with Arm Virtual Hardware: https://www.arm.com/virtual-hardware

- Join the Arm Virtual Hardware Lab Series bit.ly/arm-virtual-hardware-series

- The example project: https://github.com/ARM-software/VHT-TFLmicrospeech

- The Open-CMSIS-Pack project: https://www.open-cmsis-pack.org/
Copyright Notice

This presentation in this publication was presented as a tinyML® Talks webcast. The content reflects the opinion of the author(s) and their respective companies. The inclusion of presentations in this publication does not constitute an endorsement by tinyML Foundation or the sponsors.

There is no copyright protection claimed by this publication. However, each presentation is the work of the authors and their respective companies and may contain copyrighted material. As such, it is strongly encouraged that any use reflect proper acknowledgement to the appropriate source. Any questions regarding the use of any materials presented should be directed to the author(s) or their companies.

tinyML is a registered trademark of the tinyML Foundation.

www.tinyml.org