tinyML® Talks

Enabling Ultra-low Power Machine Learning at the Edge

“Machine Learning without batteries: the case for light-powered tinyML”

Andres Gomez - University of St. Gallen

March 3, 2022

TINY ML

www.tinyML.org

tinyML Talks Strategic Partners

Sony
Semiconductor
Solutions
Corporation

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

Executive Strategic Partners

arm AI

Powering tinyML Innovation

Arm AI Virtual Tech Talks

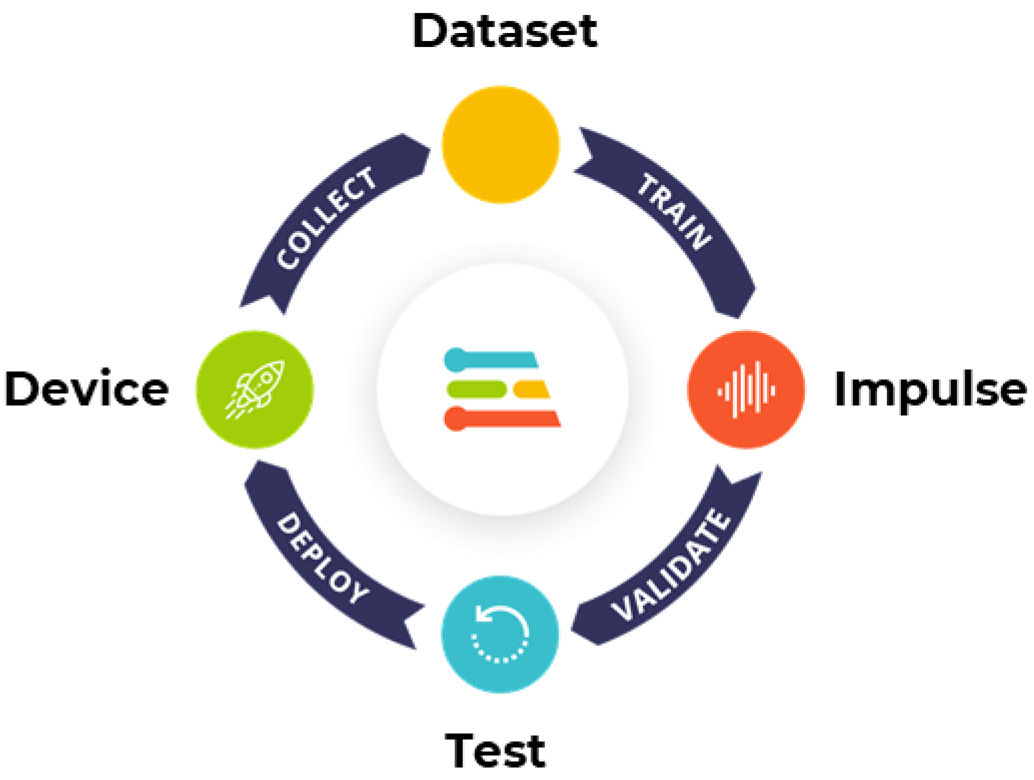
The latest in AI trends, technologies & best practices from Arm and our Ecosystem Partners.

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

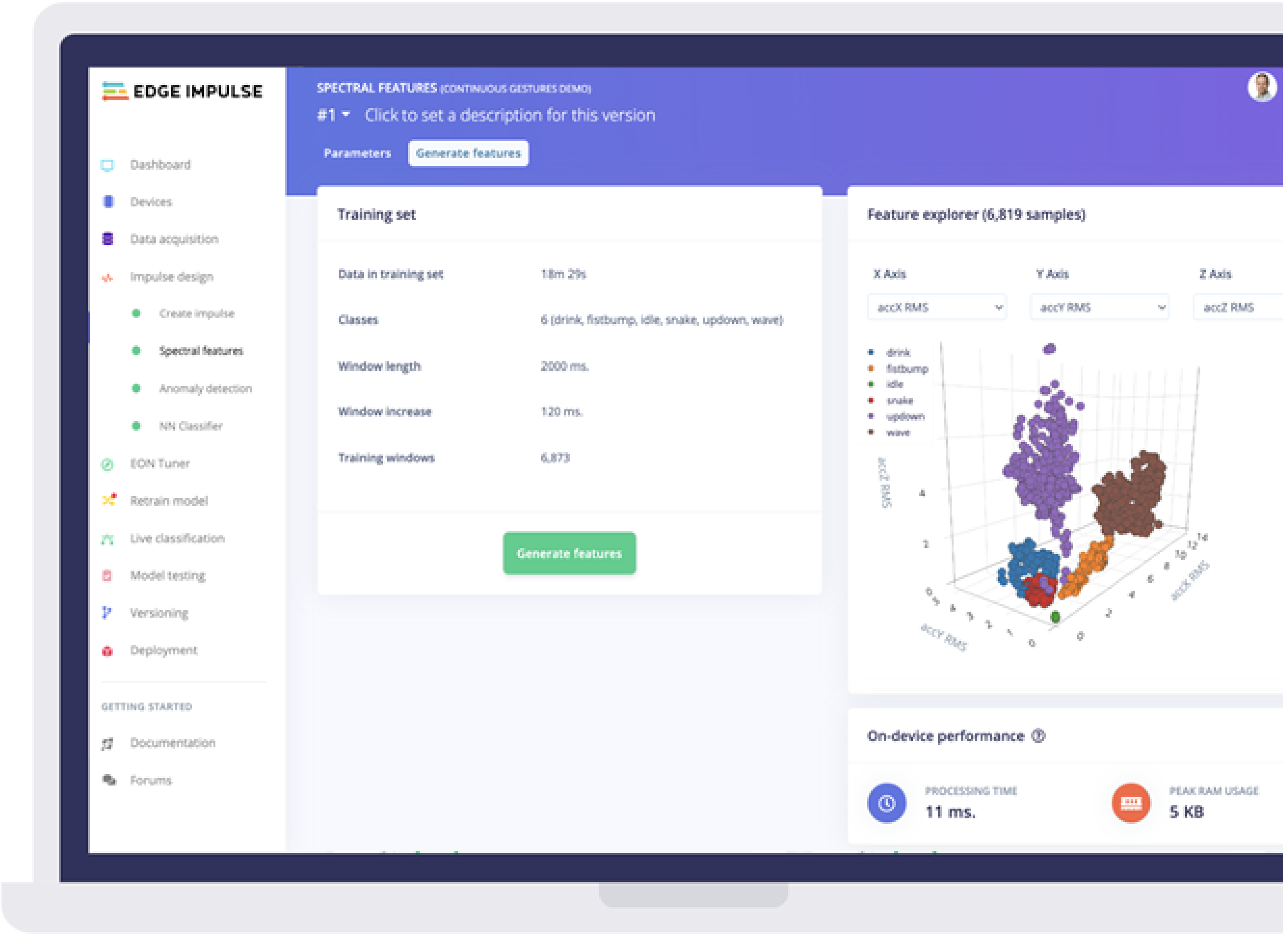
Fortnightly Tuesday @ 4pm GMT/8am PT

Find out more:
www.arm.com/techtalks

EDGE IMPULSE The leading edge ML platform



www.edgeimpulse.com



Qualcomm
AI research

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

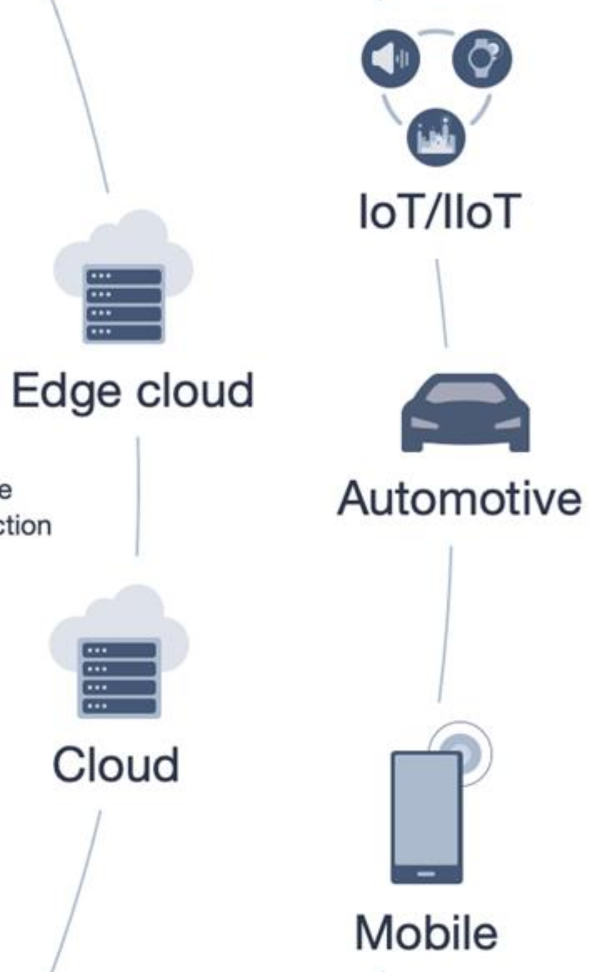
Efficient learning
Robust learning through minimal data, unsupervised learning, on-device learning

A platform to scale AI across the industry

Perception
Object detection, speech recognition, contextual fusion

Reasoning
Scene understanding, language understanding, behavior prediction

Action
Reinforcement learning for decision making



Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

SYNTIANT

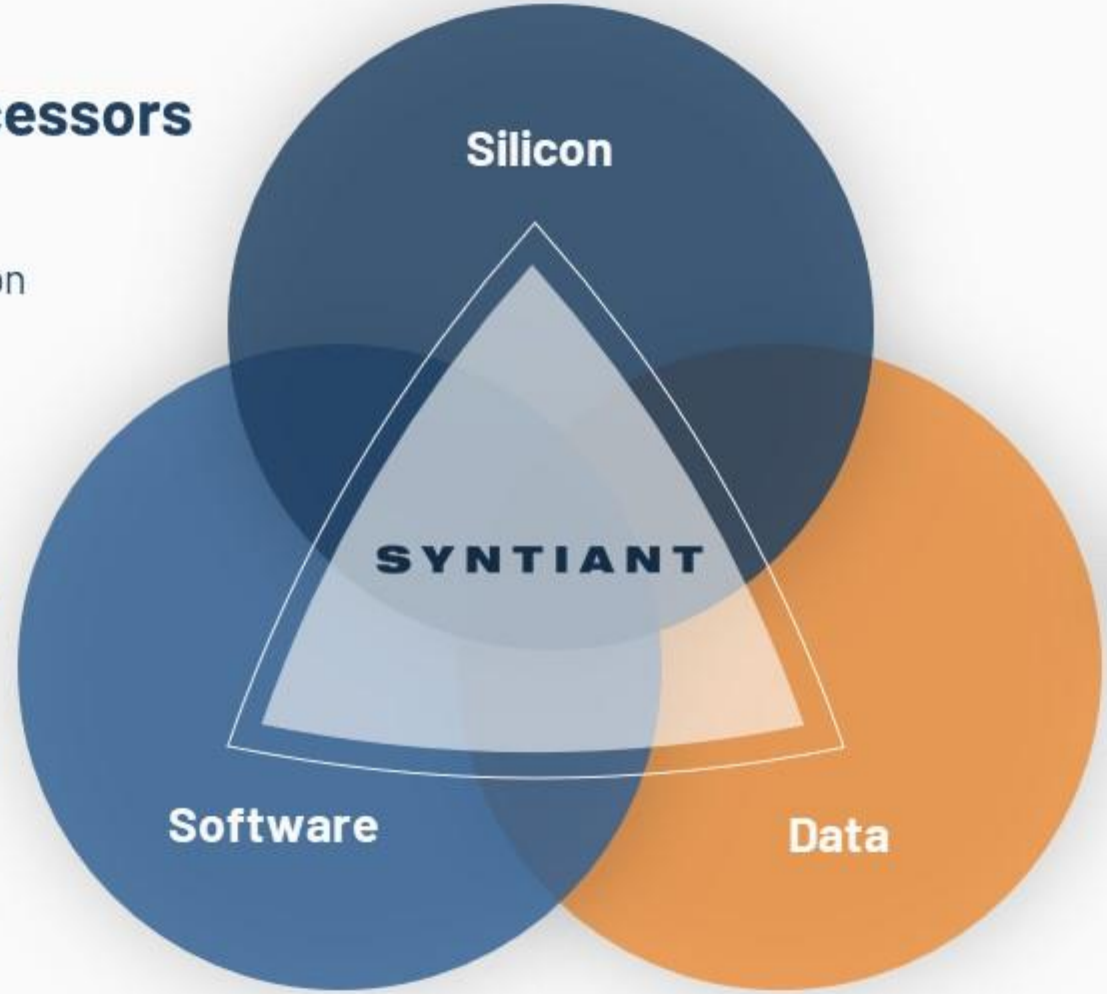
End-to-End Deep Learning Solutions

for

TinyML & Edge AI

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

partners@syntiant.com

www.syntiant.com

Platinum Strategic Partners

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**

- APPLY NOW **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 CB INSIGHTS

TINY ML TALKS webcast

KLIKA·TECH

GLOBAL IOT SOLUTIONS

Reality AI®

Add Advanced Sensing to your Product with Edge AI / TinyML

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

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

BROAD AND SCALABLE EDGE COMPUTING PORTFOLIO

Microcontrollers & Microprocessors

Arm® Core

RENESAS RA — Arm® Cortex®-M 32-bit MCUs
Arm ecosystem, Advanced security, Intelligent IoT

RENESAS RZ — Arm®-based High-end 32 & 64-bit MPUs
High-resolution HMI, Industrial network & real-time control

RENESAS RE — Arm® Cortex®-M0+ Ultra-low Power 32-bit MCUs
Innovative process tech (SOTB), Energy harvesting

Renesas Synergy™ — Arm®-based 32-bit MCUs for Qualified Platform
Qualified software and tools

Renesas Core

RENESAS RL78 — Ultra-low Energy 8 & 16-bit MCUs
Bluetooth® Low Energy, SubGHz, LoRa®-based Solutions

RENESAS RX — High Power Efficiently 32-bit MCUs
Motor control, Capacitive touch, Functional safety, GUI

RENESAS RH850 — 40nm/28nm process Automotive 32-bit MCUs
Rich functional safety and embedded security features

Core technologies

AI

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

Digital & Analog & Power Solution

Winning Combinations that combine our complementary product portfolios

Security & Safety

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

Cloud Native

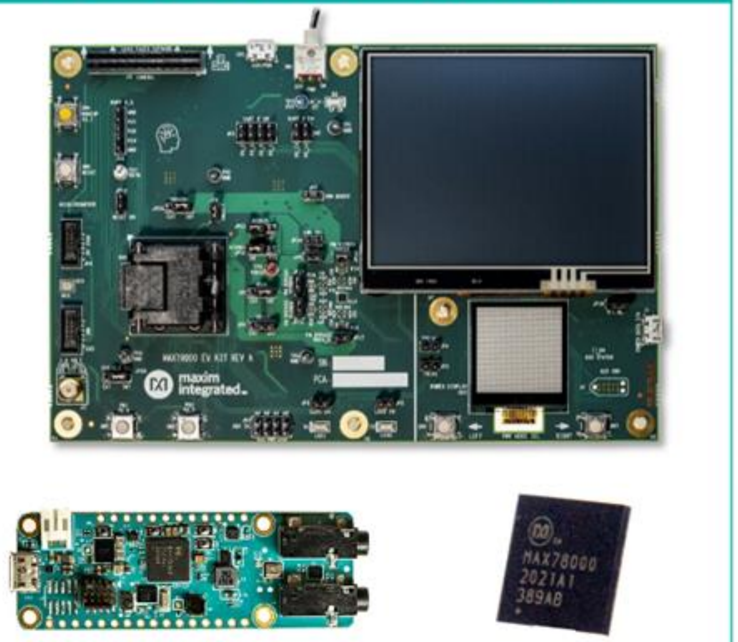
Cross-platforms working with partners in different verticals and organizations

Gold Strategic Partners

FotaHub

Maxim Integrated: Enabling Edge Intelligence

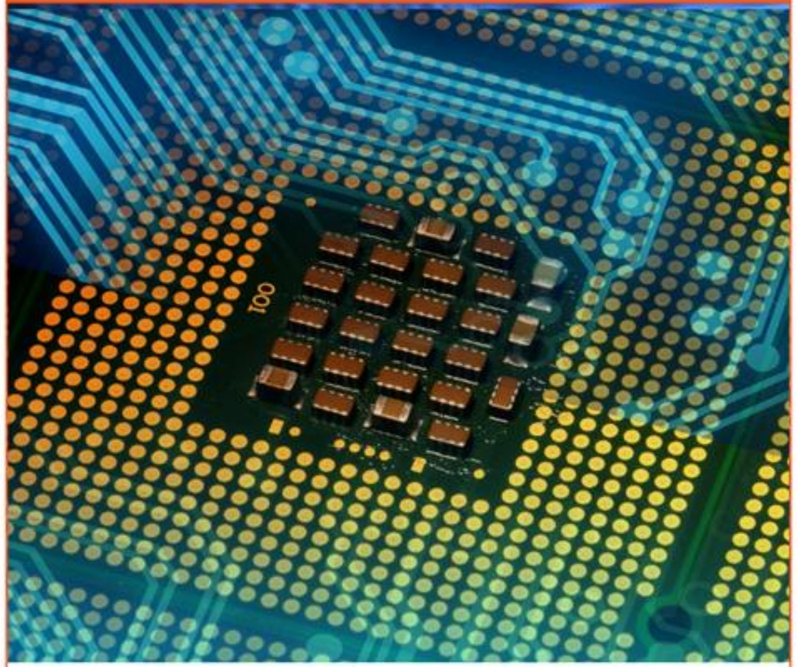
Advanced AI Acceleration IC



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

TINY ML TALKS webcast

Micro.ai

seeed studio

The IoT Hardware Enabler

SensiML™

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

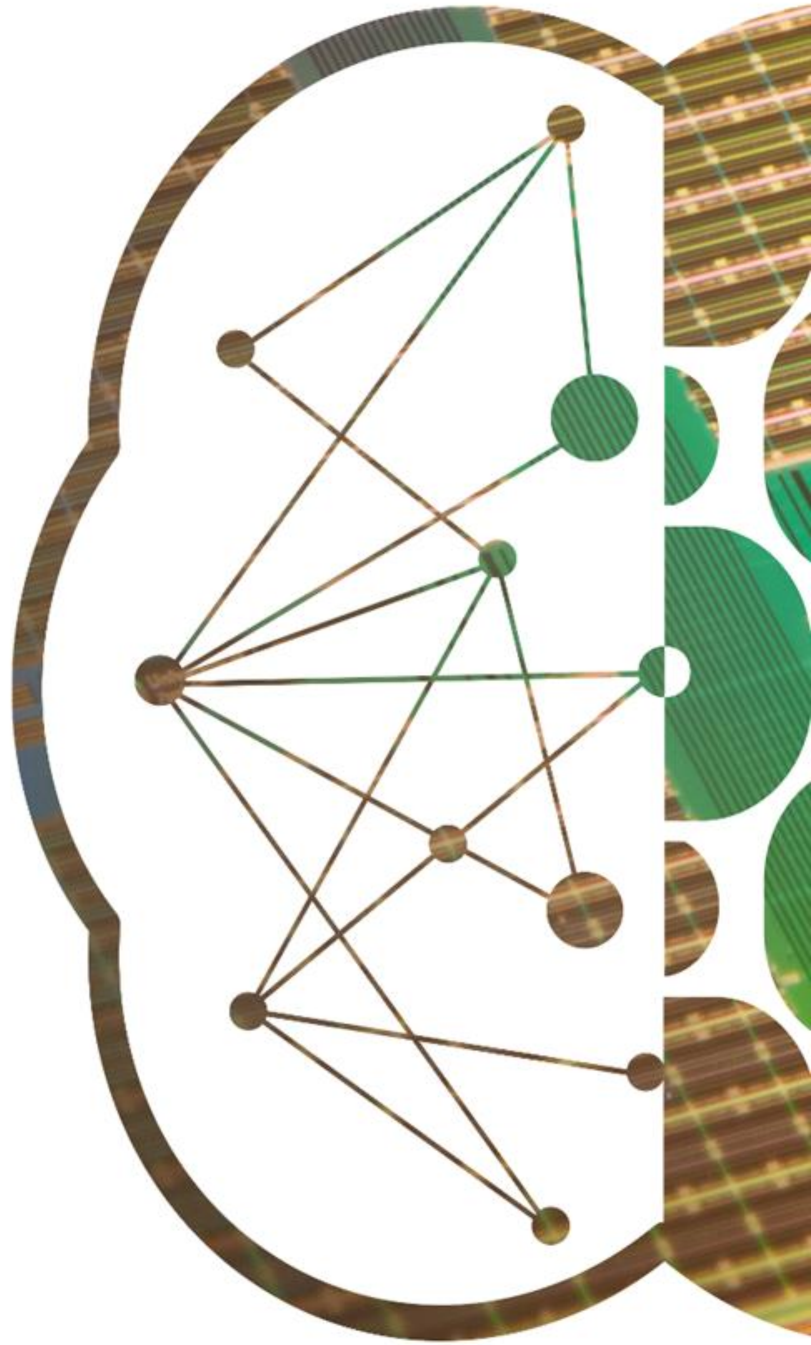
TINY ML TALKS webcast

life.augmented

SynSense

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



TINY ML TALKS webcast

Silver Strategic Partners

AONdevices

emza visual sense

GREENWAVES TECHNOLOGIES

Grovety Inc.

HOTG

imagimob

itemis

PROPHESEE METAVISION FOR MACHINES

Qeexo

REEXEN technology

SAP

SA STREAM ANALYZE

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

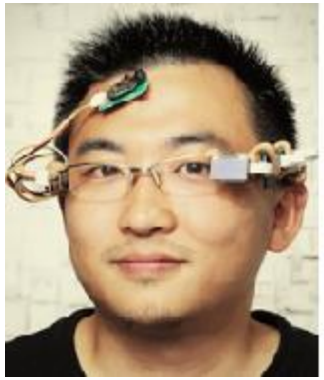
https://www.tinyml.org/event/research-symposium-2022

TINY ML TALKS webcast

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**



TINY ML TRAILBLAZERS

Register now!

Join Growing tinyML Communities:

8.3k members in
43 Groups in 34 Countries

tinyML - Enabling ultra-low Power ML at the Edge

https://www.meetup.com/tinyML-Enabling-ultra-low-Power-ML-at-the-Edge/

2.6k members
&
5.1k followers

The tinyML Community

https://www.linkedin.com/groups/13694488/

TINY ML TALKS webcast

Subscribe to
tinyML YouTube Channel
for updates and notifications
(including this video)

www.youtube.com/tinyML

YouTube

Next tinyML Talks

Date	Presenter	Topic / Title
Tuesday, March 8	Konstantin Meshcheriakov, Solution Architect, Klika Tech	On-device model fine-tuning for industrial anomaly detection applications

Webcast start time is 8:00 am Pacific time

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

Reminders

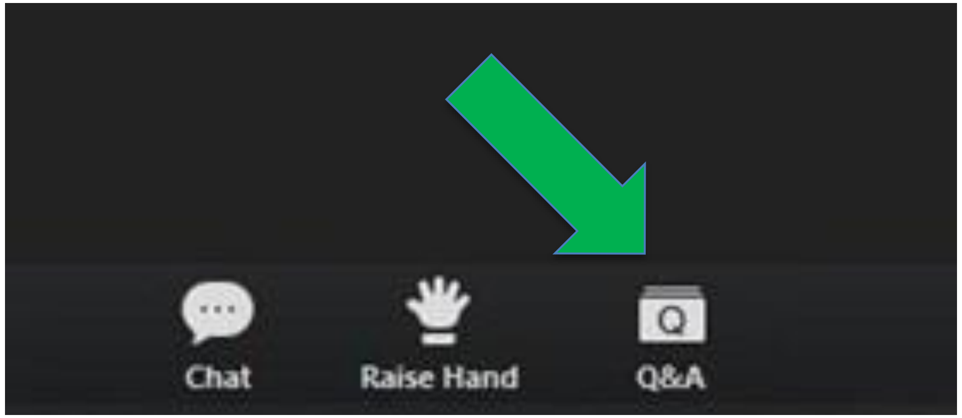
Slides & Videos will be posted tomorrow

PDF

tinyml.org/forums

youtube.com/tinyml

Please use the Q&A window for your questions



Andres Gomez

Andres Gomez received a dual degree in electronics engineering and computer engineering from the Universidad de Los Andes, Colombia, an M.Sc. degree from the ALaRI Institute (Università della Svizzera Italiana), Switzerland, and a Ph.D. from ETH Zurich, Switzerland. He has over ten years of experience with embedded systems and has worked in multiple research laboratories in Colombia, Italy, and Switzerland. More recently, he has worked as an R&D engineer at Miromico AG. He has co-authored over 30 scientific articles and has contributed to multiple open-source projects. He is currently a Postdoctoral Fellow at the University of St. Gallen, Switzerland. His current research interests include batteryless system design, the Internet of Things, and the Web of Things.

Institute of Computer Science

University of St.Gallen

Interactions

interactions.ics.unisg.ch

Machine Learning without batteries:

the case for light-powered tinyML

03.03.2022

Did you ever own one of these?

Source: GSM arena

Nokia 1110 (2004)

Average use: several days

Battery: **850 mAh**

Do you ever own one of these?

Source: GSM arena

Nokia 1110 (2004)

Average use: several days

Battery: **850 mAh**

Source: GSM arena

iPhone 13 Pro Max (2021)

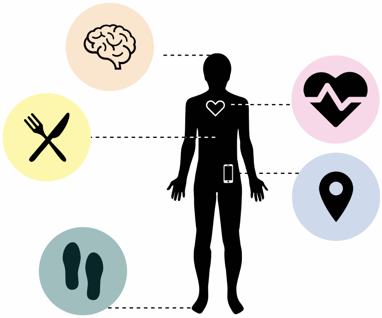
Internet use: ~12 h

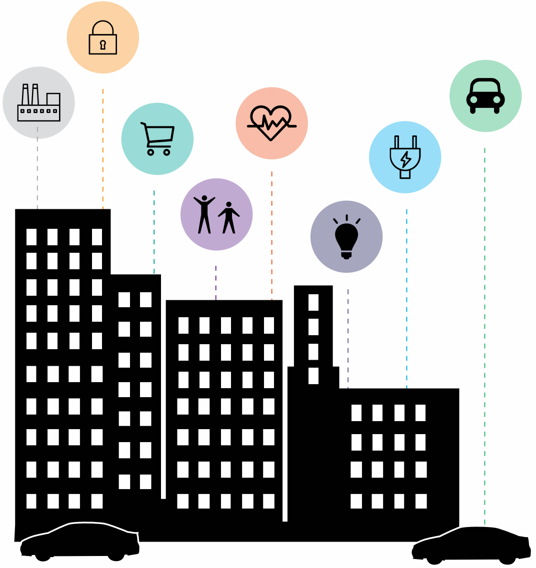
Battery: **4352 mAh**

Trend: as **functionality** increases so does the **cost**
(battery) size
environmental impact

Current Internet of Things

Making environments and devices **smarter**





People spend **90%** of their time indoors[1]

Buildings account for **28%** of CO_2 emissions[2]

Lighting and HVAC consume **64%** of energy in buildings[3]

[1] https://bit.ly/3gVWp43 [2] https://bit.ly/3sKUYe0 [3] https://bit.ly/3sQe02M

How can we improve?

Current sensing devices:

electronics

batteries

>90% of volume

~50% of cost

My research aims to **novel design paradigms:**

cost-effective	—	low-cost (organic) components
maintenance-free	—	taping into light infrastructure
environmentally friendly	—	saves energy by design

Andrés Gómez

Finished PhD studies at ETHZ in 2018

Design and Specification of Batteryless Sensing Systems

Part-time in two research institutes: 75% TIK, 75% IIS

ETH zürich

Research engineer at Miromico AG in Zurich, Switzerland (2018-2020)

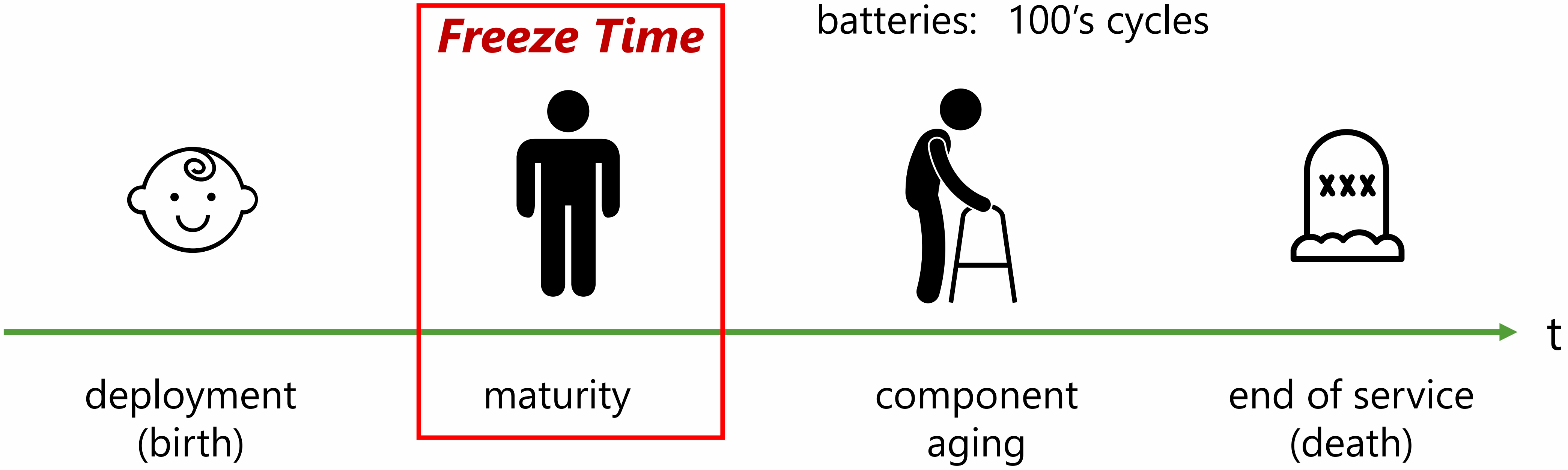
miromico

Post-doctoral fellow at University of St.Gallen since Jan. 2021

University of St.Gallen

Traditional Device Lifecycle

silicon: millennia*
solar panel: decades*
batteries: 100's cycles



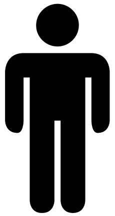
Are **long lifetimes** the *only* desirable design goal?

Predictability vs resilience

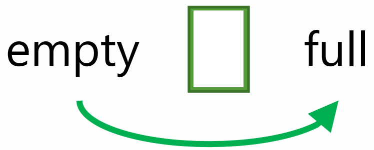
System should "live" as **long** as possible — **large** energy cycle

charging is **expensive**

Predictable
Life



System should "revive" as **quickly** as possible — **small** energy cycle



charging is **cheap**

Resilient
"Life"

A few words about energy storage

Main tradeoffs:

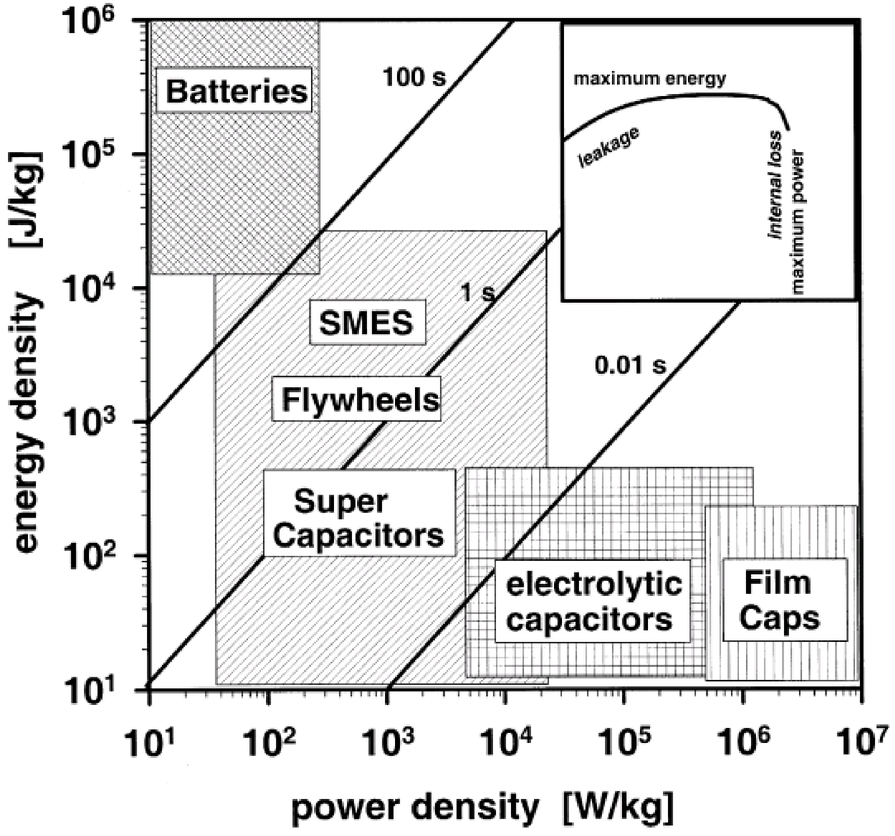
- recharge cycles
- power density
- energy density
- leakage

My objectives: **high power density**
high recharge cycles

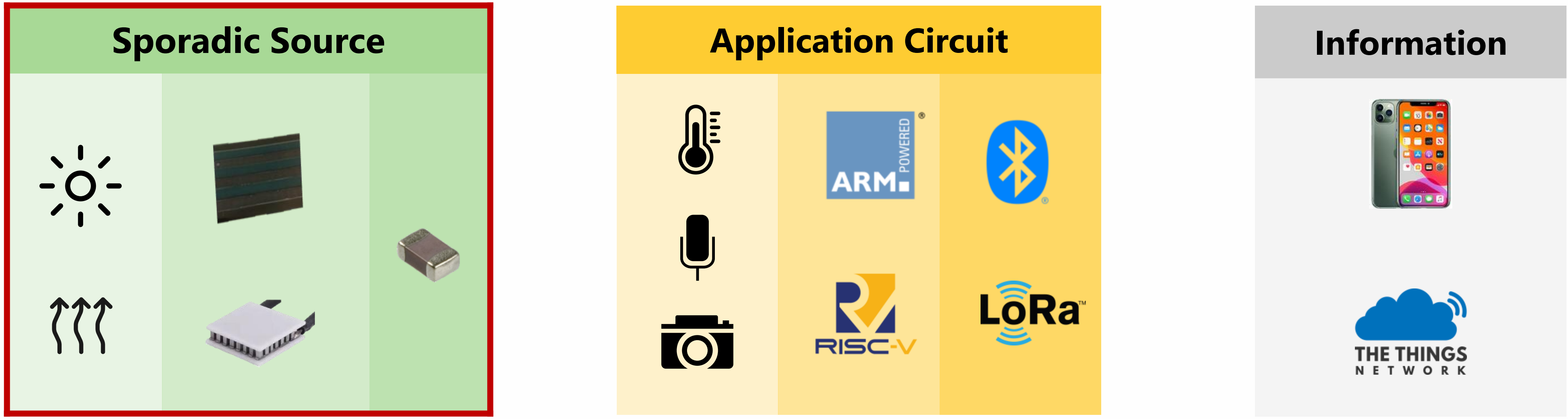
I mostly use ceramic capacitors

Ragone Plot

source: T. Christen et al.



Energy-Driven Design



Sporadic Source	Application Circuit	Information
Inexpensive	***Sporadically Powered*** ∃t (P(t) = 0)	***Non-Deterministic*** ***possibly correlated****

Key Feature Environment sporadically supplies energy

Objective minimize wake-up time (~1 sec)

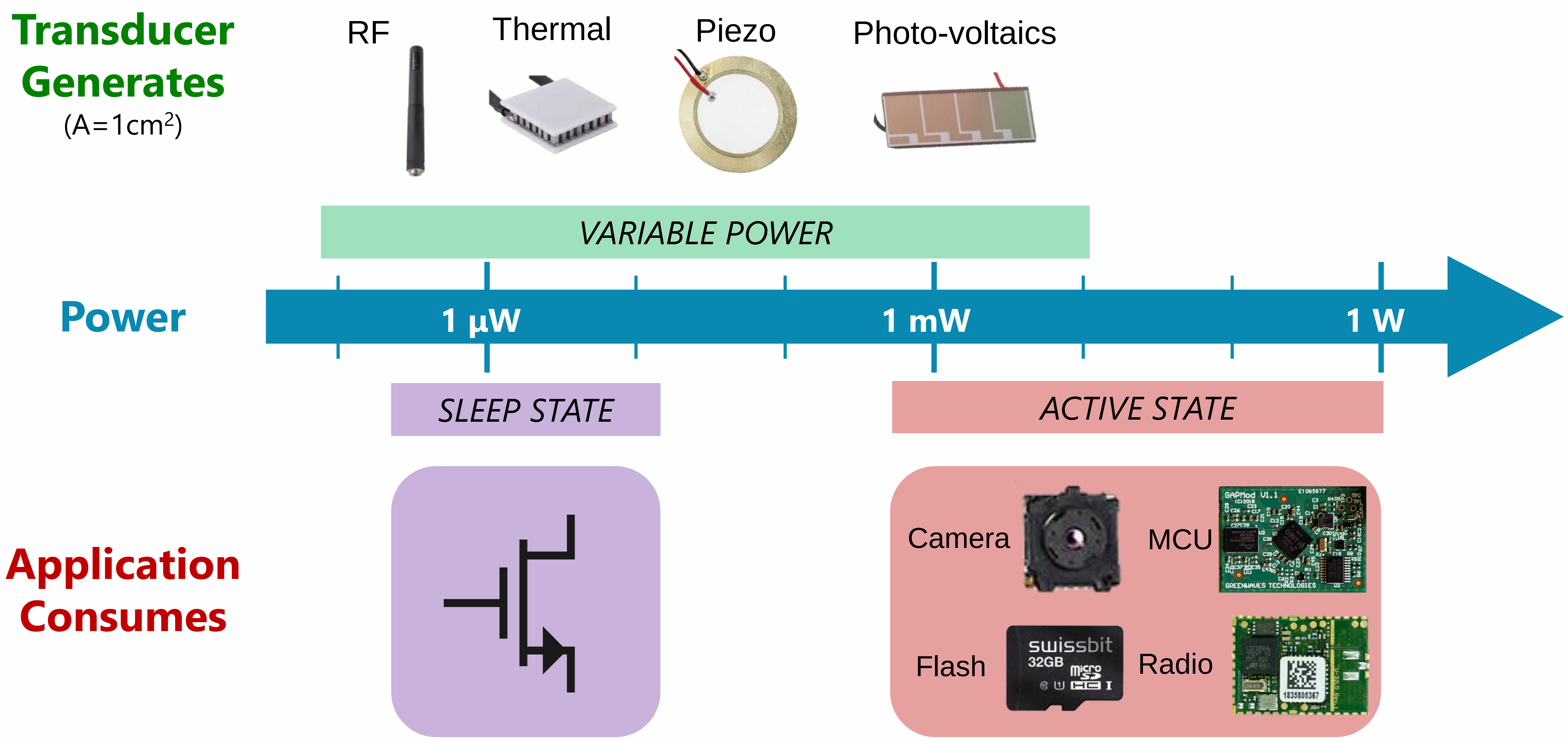
Energy Harvesting

Sporadic Source

Application Circuit

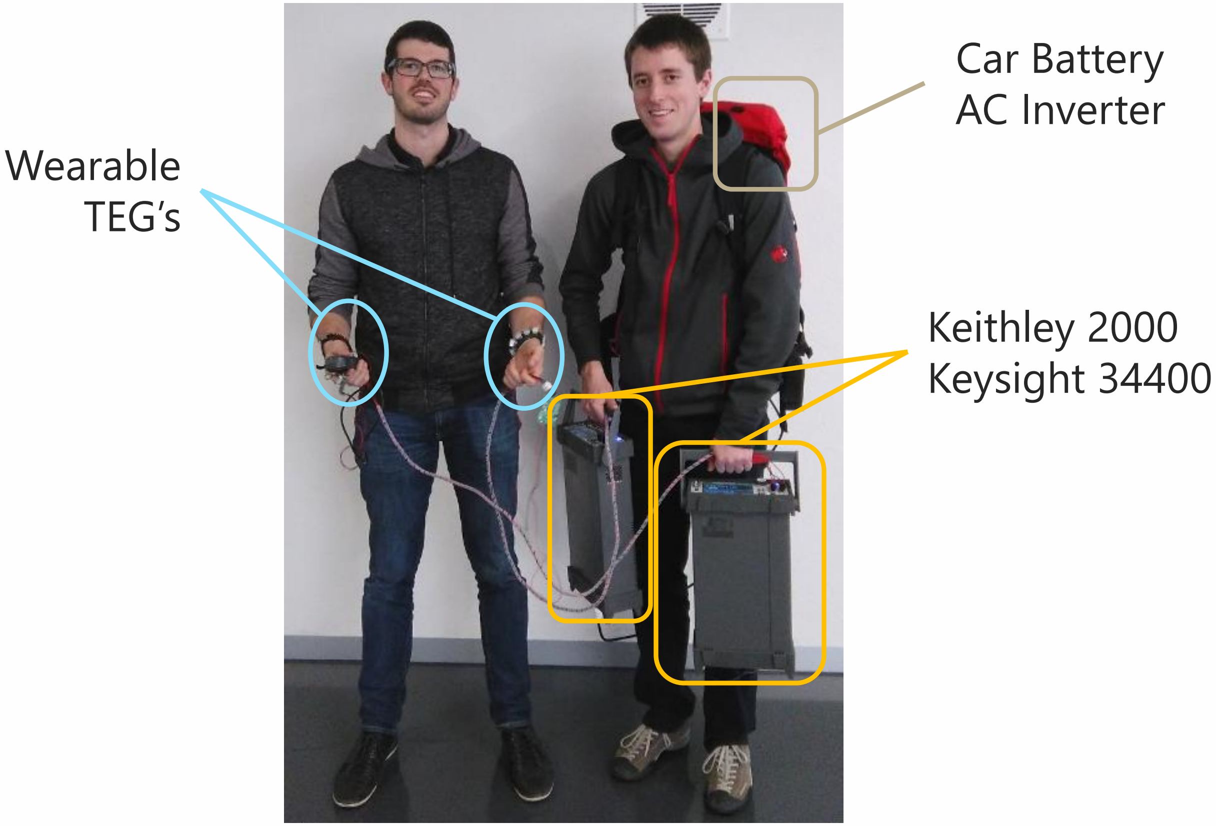
ML Model

The Great Power Imbalance



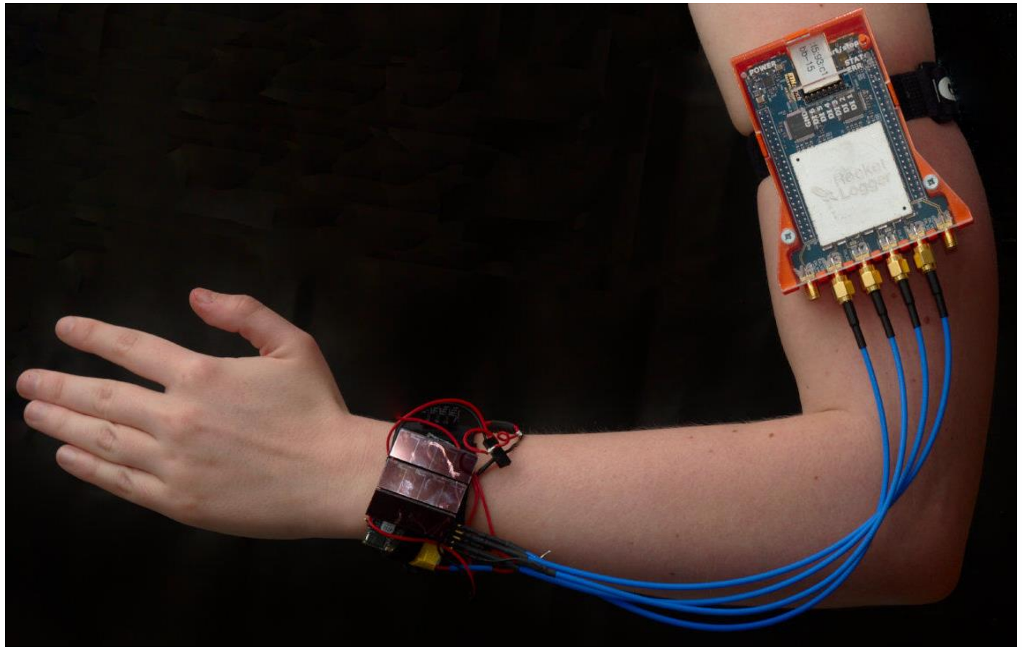
Understanding Environmental Energy

Accurate, in-situ measurements are challenging!



Understanding Environmental Energy

We designed the RocketLogger to facilitate this



2 high accuracy I channels
4 high accuracy V channels
6 digital channels

Open-Source Project:

https://rocketlogger.ethz.ch/

Used by researchers in:

How Much Energy Is Available Indoors?

Measurement campaign @ ETH Zurich

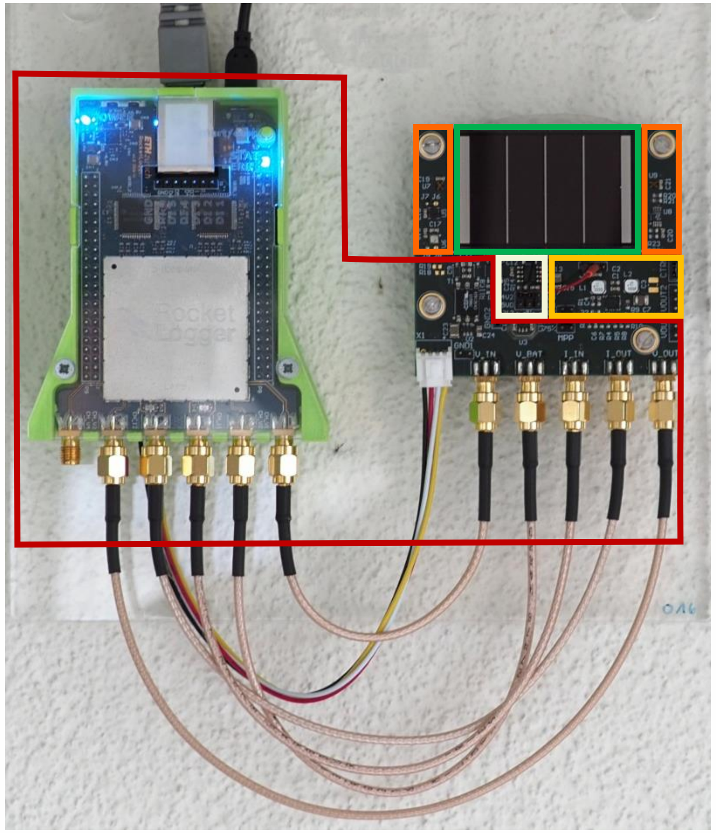
Started July 2017 with 6 indoor stations

Ambient & Electrical data:

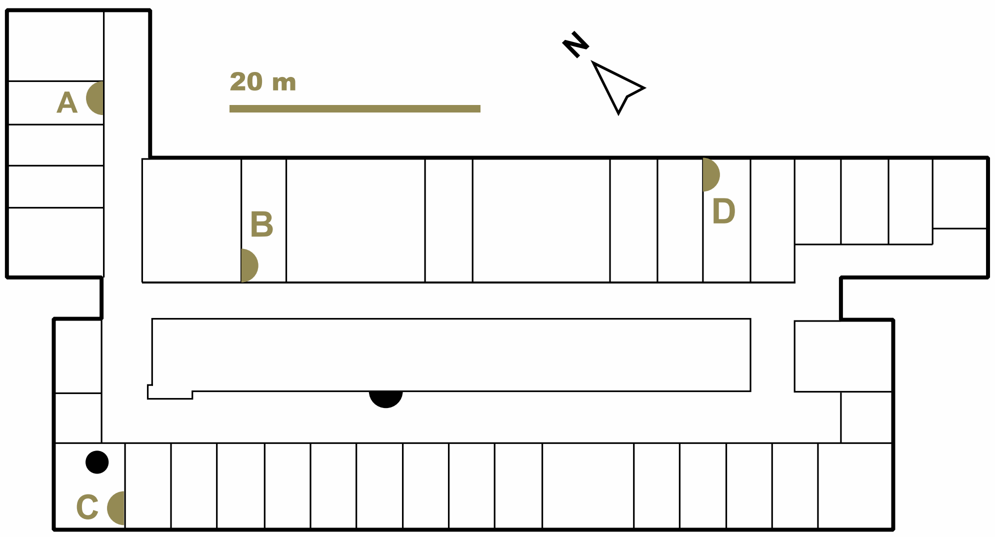
- Illuminance @ 1 Hz
- Harvested power @ 10 Hz
- 25 billion data points (139 GB of data)

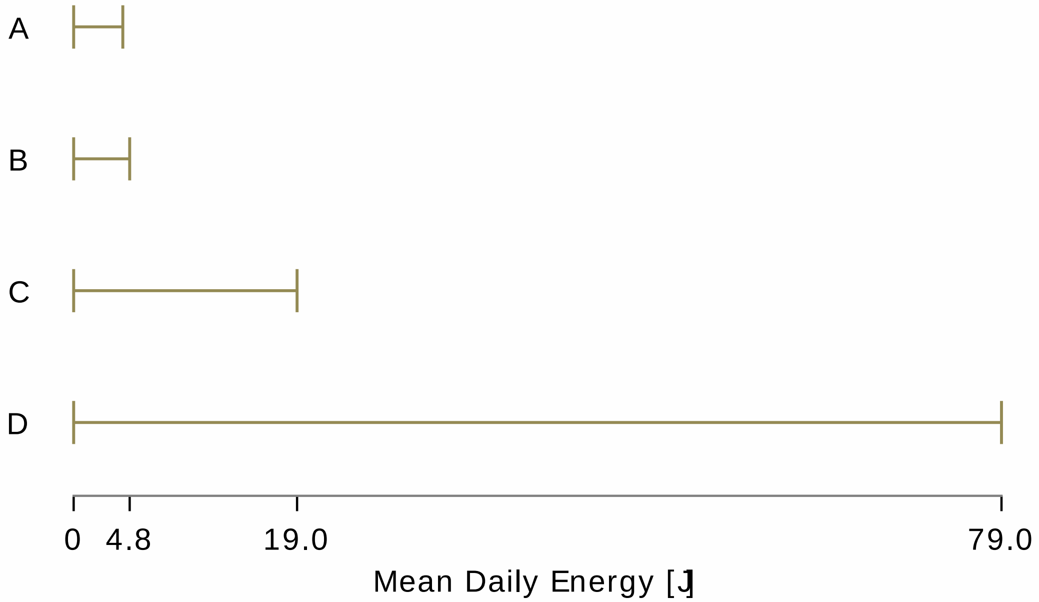
Publicly available data set:

DOI 10.5281/zenodo.3363925



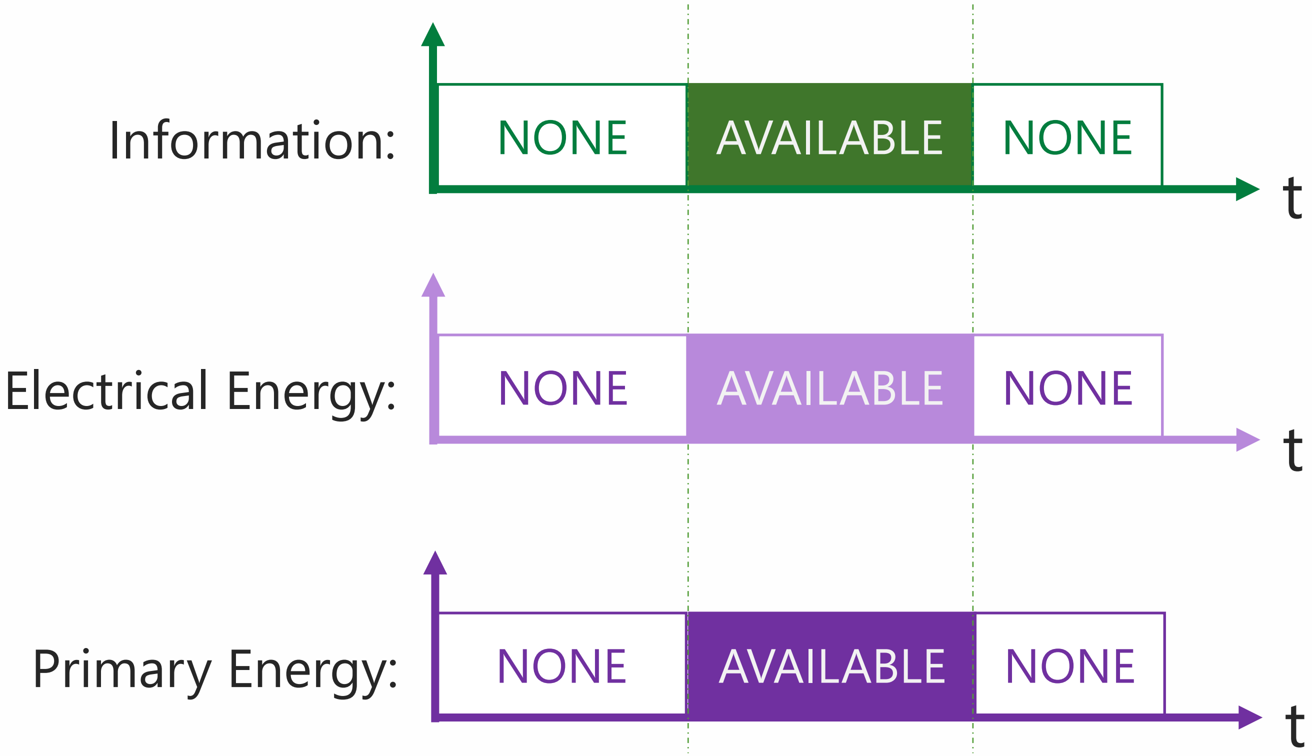
Daily Harvested Energy Distribution (2+ years)

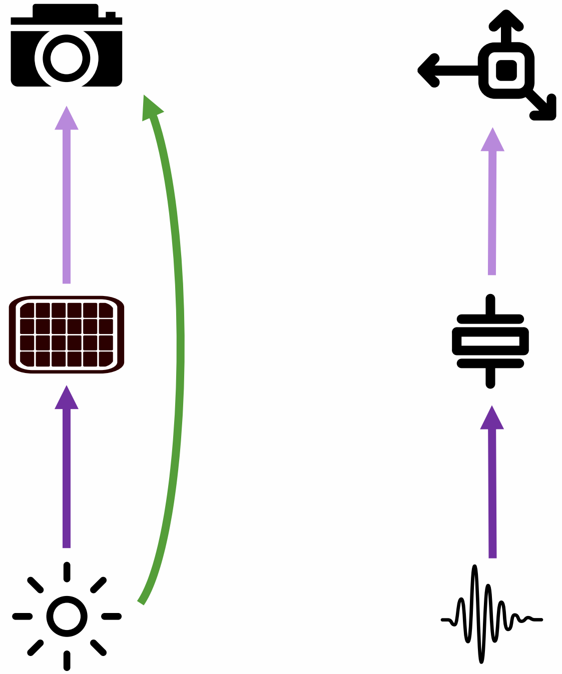




High energy variability with many variables: geography, architecture, human activity, etc

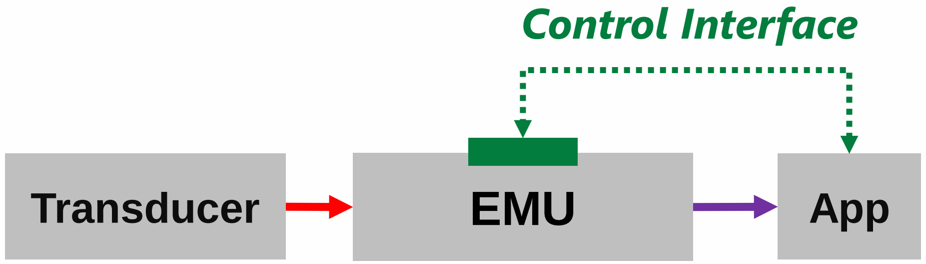
When are Energy-Driven Systems Optimal?





human activity
correlates with light

Energy Management Unit (EMU)



P_{in} = 5 μW

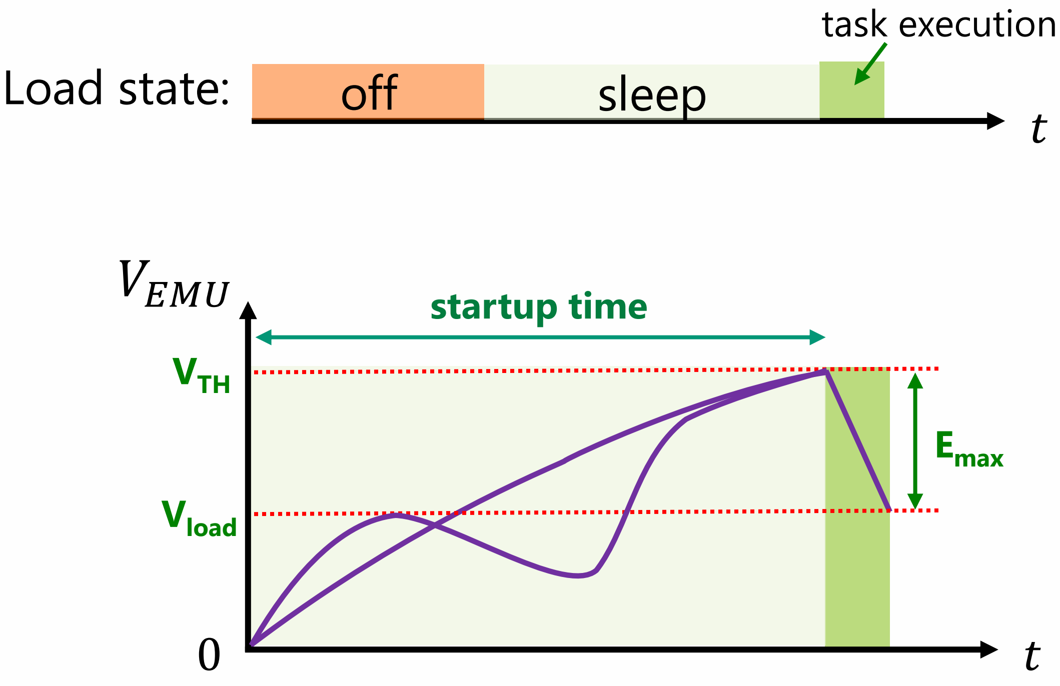
P_{app} = 100 mW

Source/Load have **decoupled** voltage and current

Energy input/output simultaneously optimized

Optimizing $C, \Delta V^2$ minimizes startup time

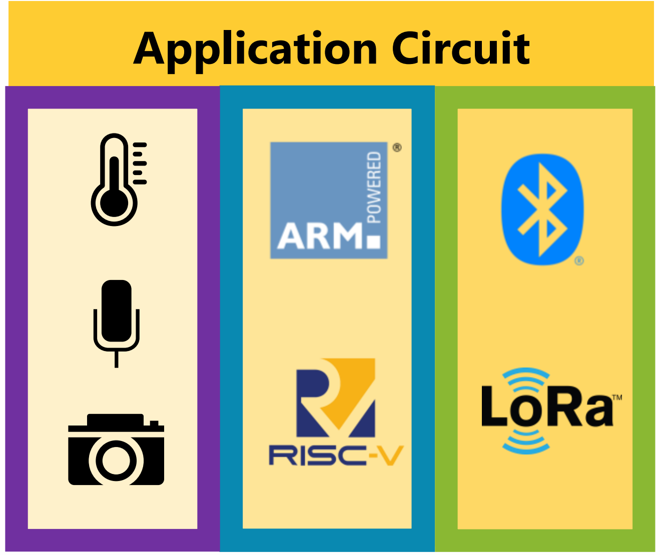
Transient behavior:



System's energy efficiency: $\eta_{sys} \leq$ **0.9**

Intermittent Operation

Energy-driven systems must be compatible with power cycling



Sensors:

Digital sensors are easy to turn on and turn off

Transcievers:

Easy to stop with asynch. comm.

Digital computation:

If small enough, no need to interrupt process

Reliable Execution

Sporadic Source

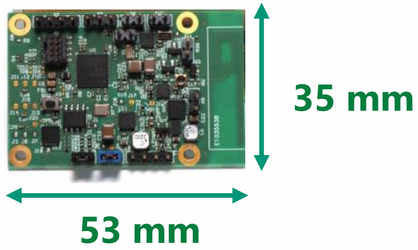
Application Circuit

ML Model

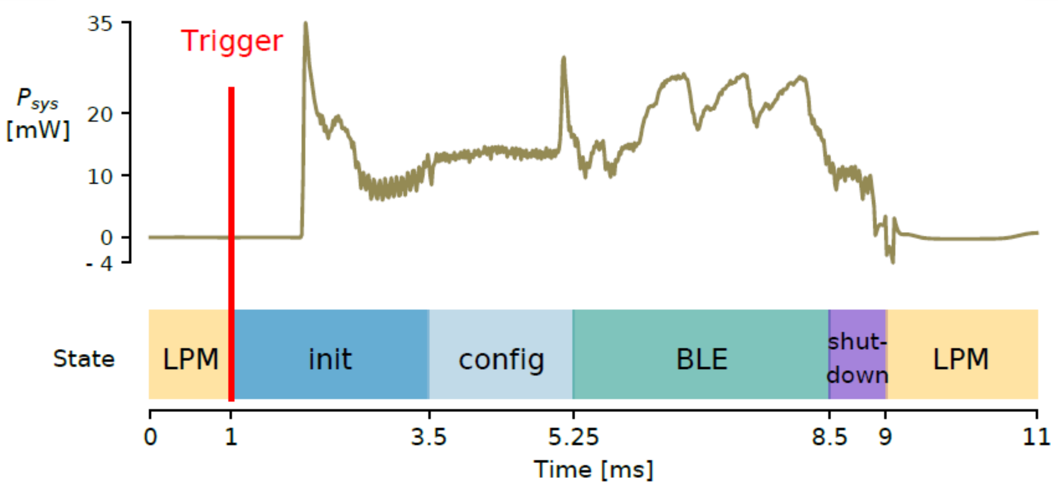
Atomic Task Execution

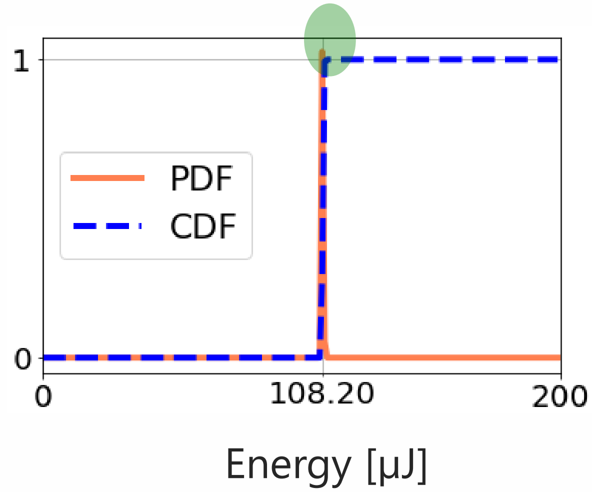
BLE Ambient Sensor

CC2650-based Platform



Power Trace for a single activation





"Batteryless" Limit

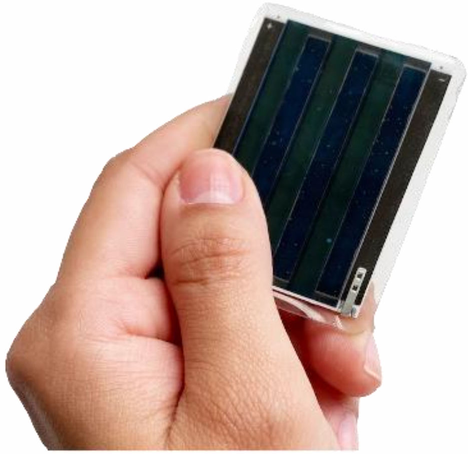
Q_{max} = 109 [μJ]

- **Reliable Operation**
- **Min. Startup Time**

In **indoor** Environments:

> 28K activations per day
> 3 pkt/s (for 8 hours)

Privacy-Enabled Edge Devices



Without Light: Full power down

impossible to gather data

With Lights: System powers up quickly

Gathers, processes and transmits data

miroCard is **powerful and private**:

It can process sensor data streams

and users can disable it by covering it

Open Source HW/SW

https://mirocard.swiss/

2021 HiPEAC
Tech Transfer Award

ML on the MiroCard

Light-Powered
Batteryless

User Agency
Controlled exposure

Sensors
Temperature, Humidity, & Acceleration

Bluetooth Low Energy
Data Transmission

Machine Learning
Local processing capability

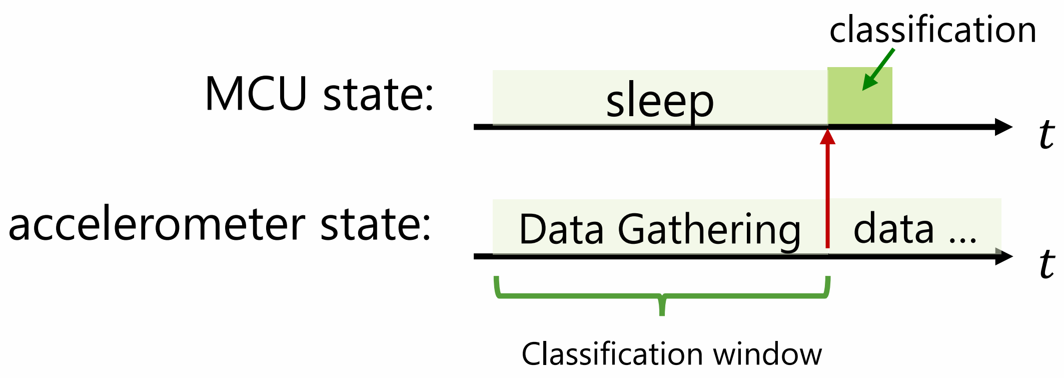
Can we do gesture detection on miroCards?

We need long time series of data

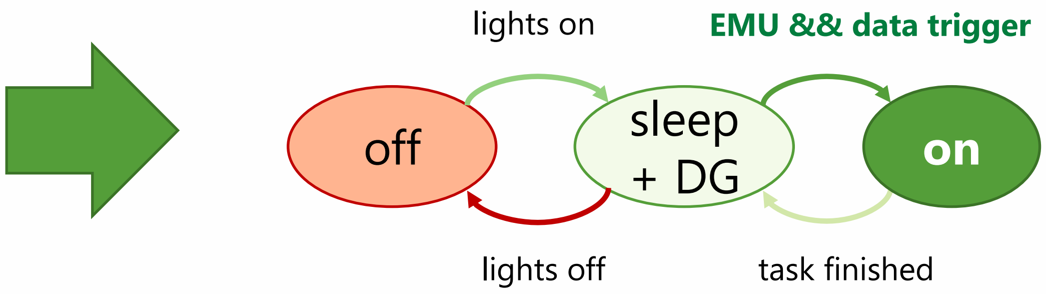
We are adding a data requirement to batteryless processing

Data streaming applications

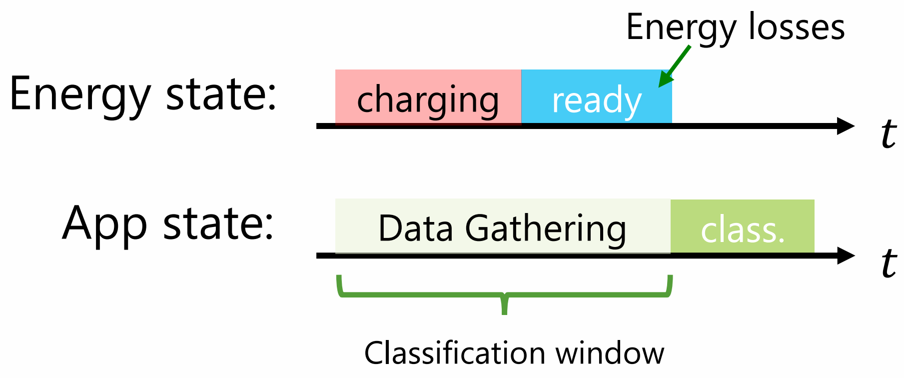
Traditional classification behavior:



Batteryless data classification:



Energy + data dependency may lead to **inefficiencies**



Gesture Detection

Sporadic Source

Application Circuit

ML Model

Data Set Acquisition

Training Data

12 Participants

Label Distribution		
Random	22 583	33.2%
Up-Down	22 357	32.8%
Sideways	23 096	33.9%
Total	68 036	100%

Set Distribution		
Training	32 865	48.3%
Validation	14 085	20.7%
Testing	21086	31%
Total	68 036	100%

ML model:

inf XL

Simple C-code

- Additions
- 8- & 16-bits integers

Low memory usage

Adaptable & Reusable

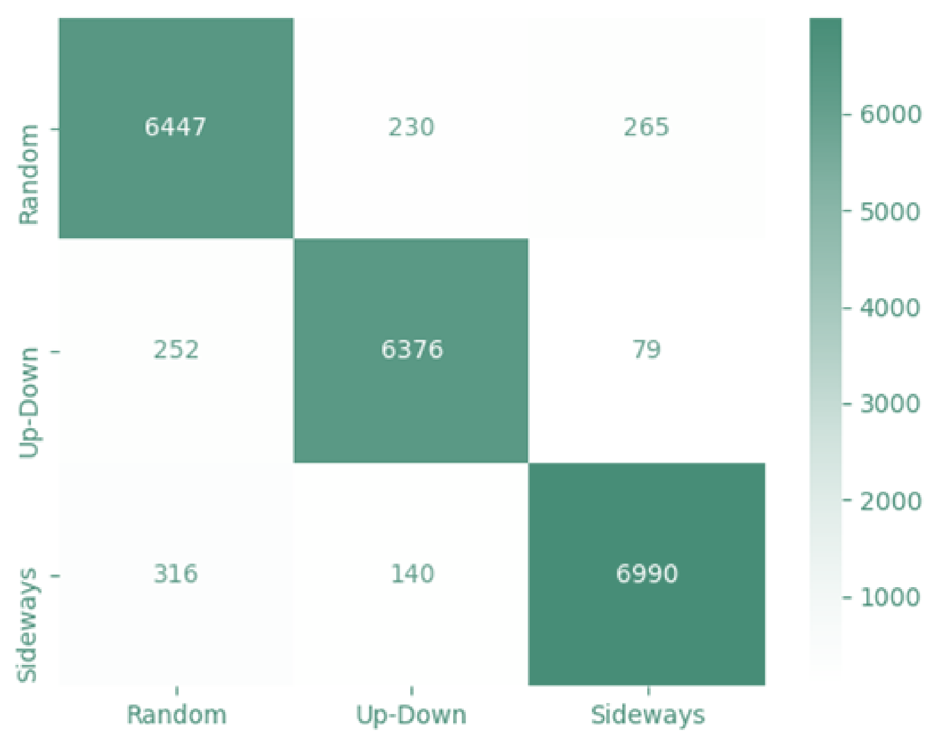
- Easy to embed in OS

Low energy consumption

- Minimal increase in additional energy demand

Trained Model Results

Confusion Matrix

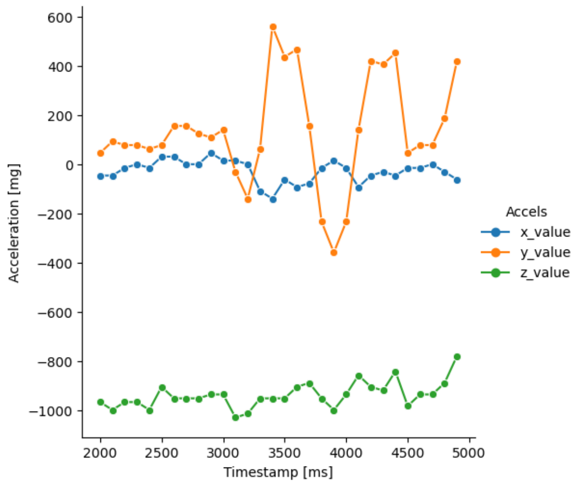


Model Evaluation

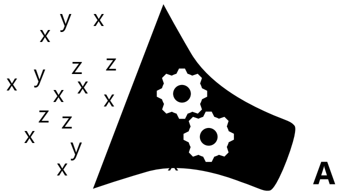
Labels	Precision	Recall	F1-score	Sample Size
Rand	0.92	0.93	0.92	6 933
Updown	0.95	0.95	0.95	6 707
side	0.95	0.94	0.95	7 446
Accuracy			0.94	21 086
Macro avg	0.94	0.94	0.94	21 086

Windowed-Data Classification

Data Exploration



Data Processing



Data Labelling

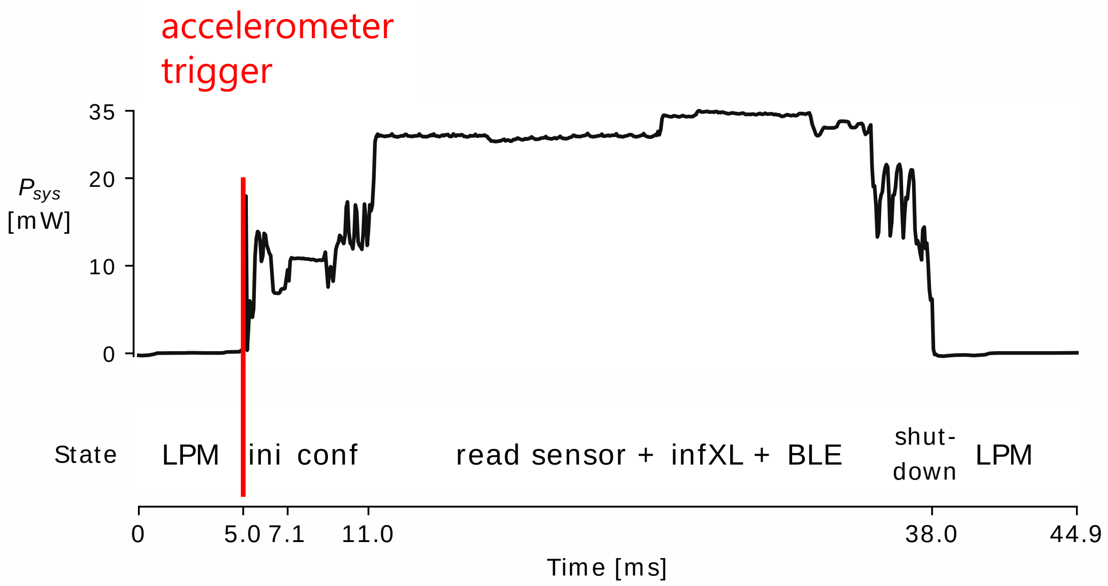

[0]: Random


[1]: Up-Down


[2]: Sideways

Embedded Implementation

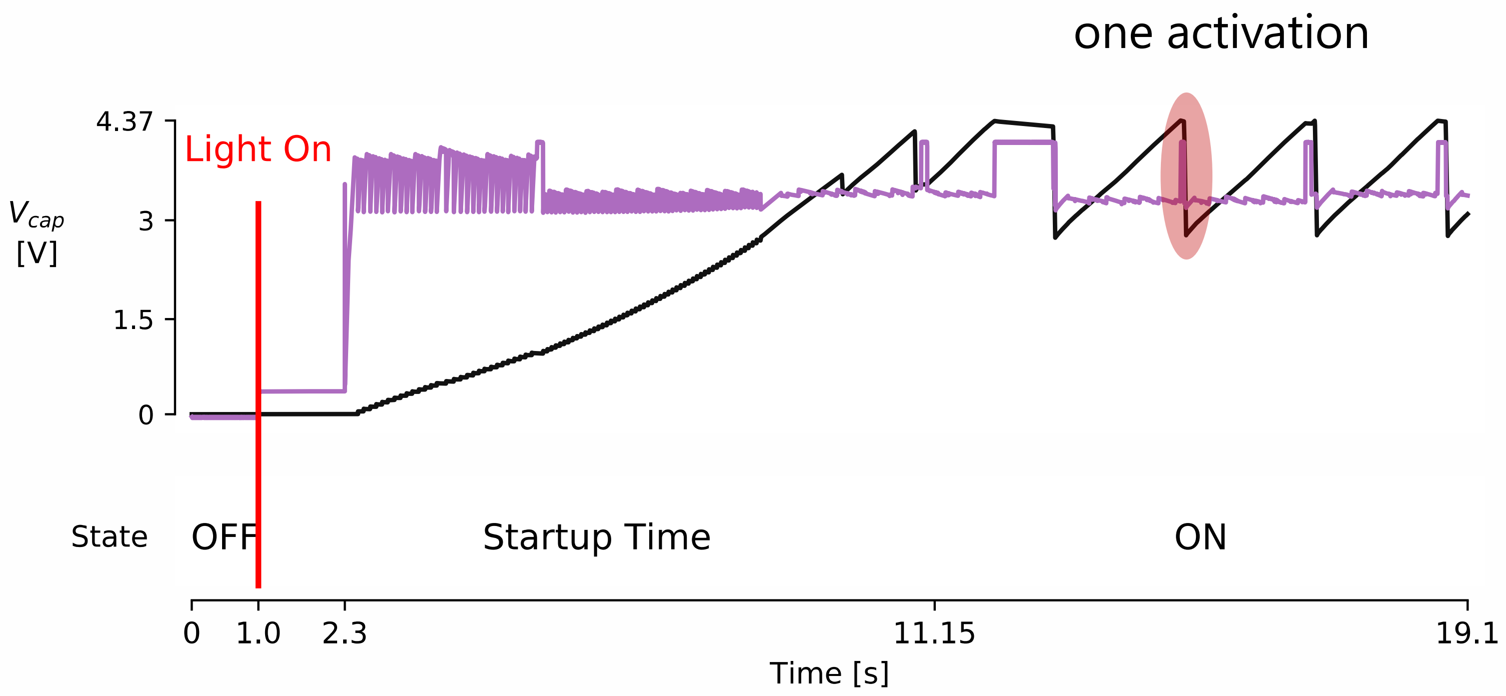
Application's power trace:



RocketLogger Measurements		
	Average	**Standard Deviation**
Execution Time	32.7 ms	0.4 ms
Energy @ 2.5V	728.01 uJ	0.8 uJ
$\mathbf{P_{sleep+dg}}$	65 uW	-

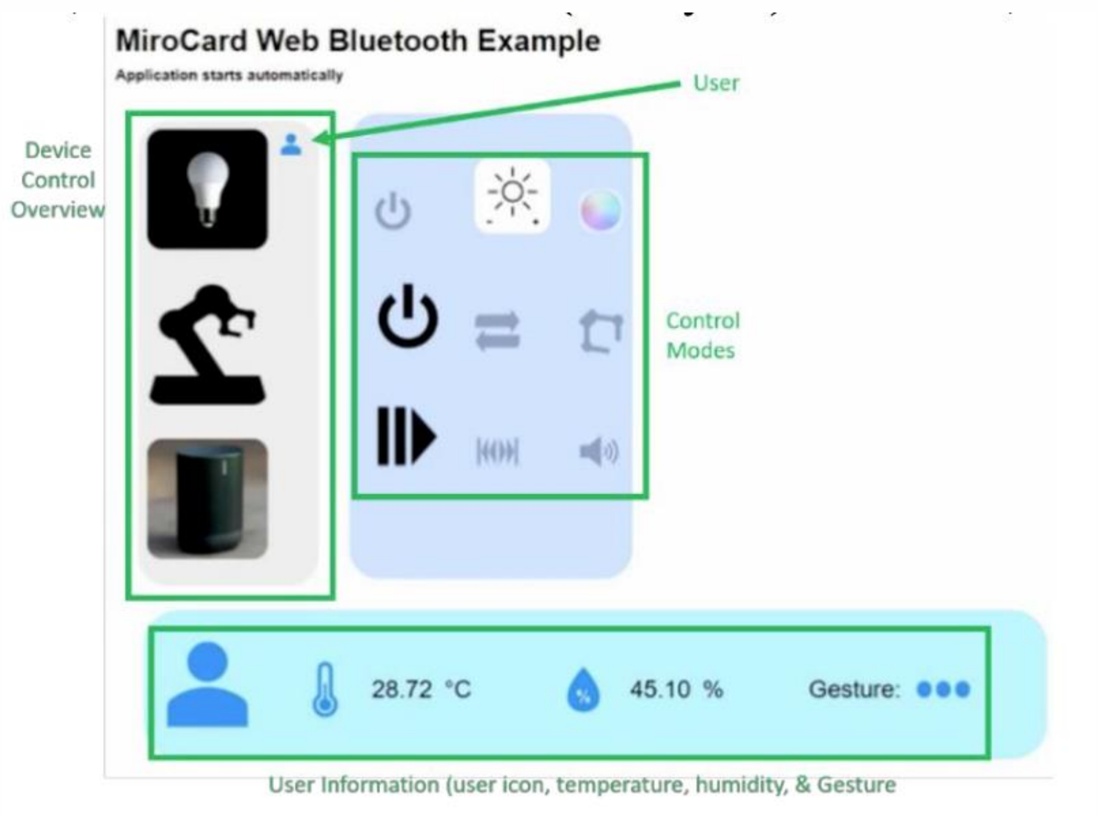
Fast Cold-Start

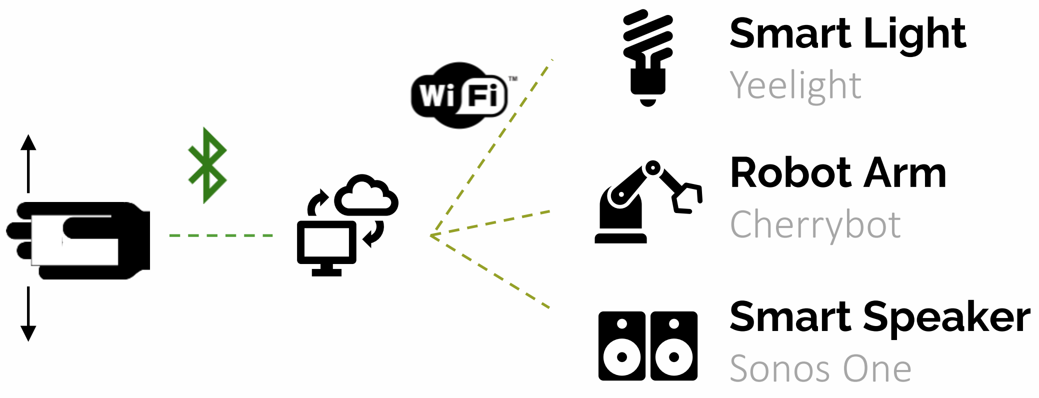
Cold-start lasts only 10.15 seconds when exposed to a constant 900 lux environment



Demo

Accelerometer-controlled interactions



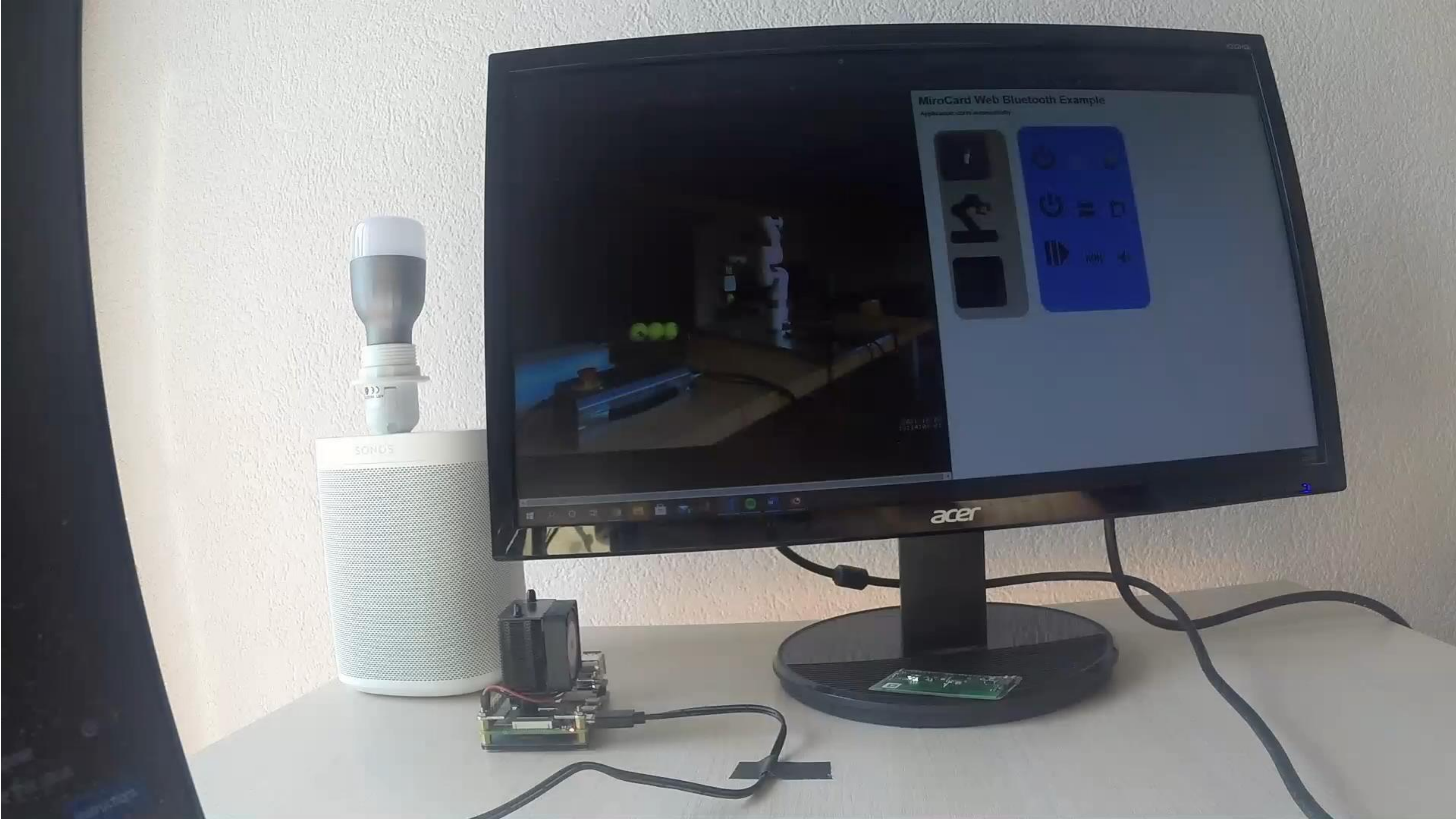


Actions:

Change light color, intensity

Control robot arm movements

Change song, volume


MiroCard Web Bluetooth Example
SONOS
acer

Acknowledgements

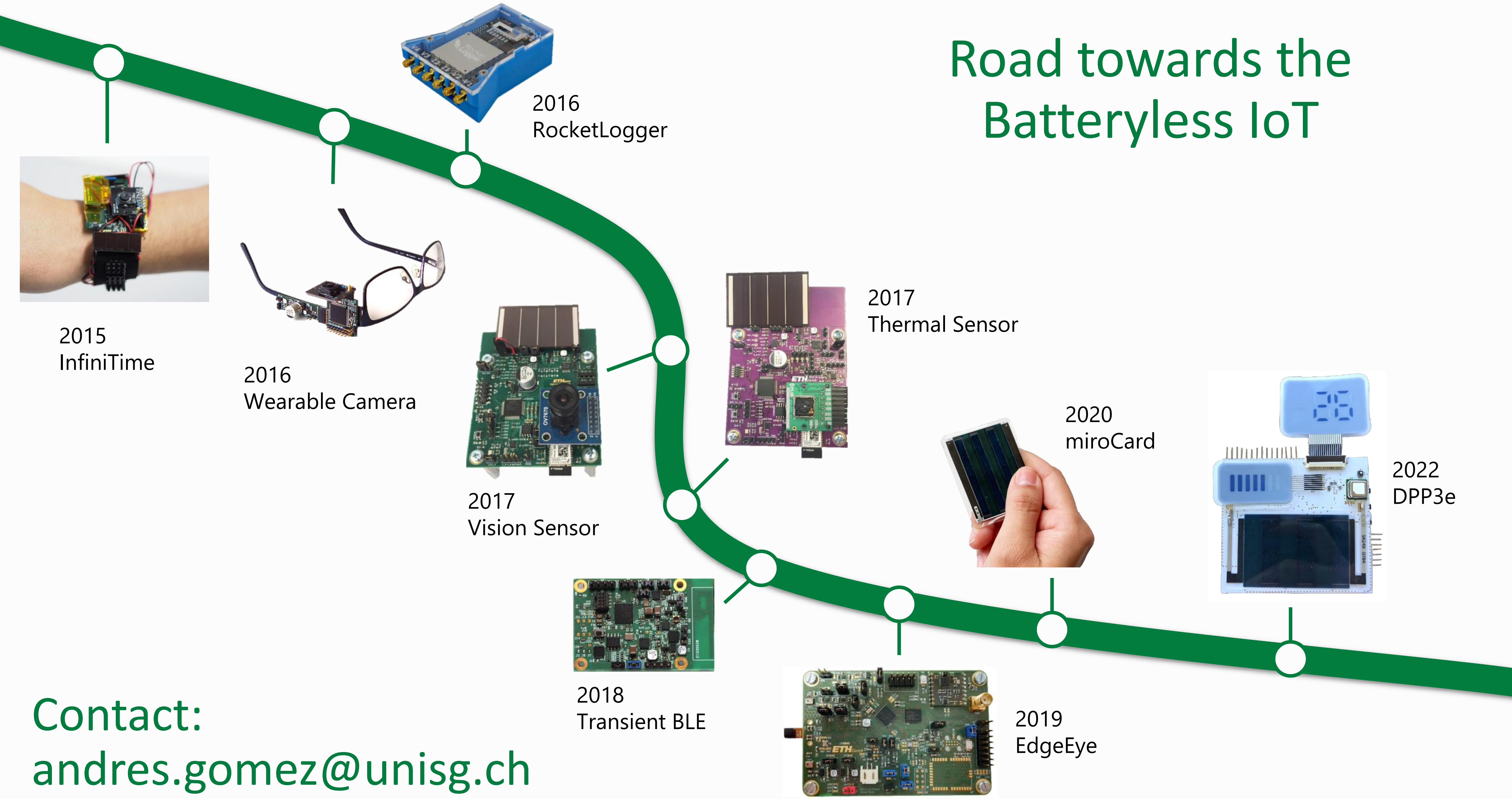
Professors Prof. Dr. S Mayer, Prof. Dr. L Thiele, Prof. Dr. L Benini

Researchers I Mizutani, Dr. L Sigrist, Dr. PA Hager

Students L Sutter, T Schalch, M Leubin, S Lippuner, P Sanmagurujah

Companies miromico, inf XL, epishine harvesting light, e-peas semiconductors, NXP

Among **many others …**



Road towards the Batteryless IoT
2015
InfiniTime
2016
Wearable Camera
2016
RocketLogger
2017
Vision Sensor
2017
Thermal Sensor
2018
Transient BLE
2019
EdgeEye
2020
miroCard
2022
DPP3e
Contact:
andres.gomez@unisg.ch

Copyright Notice

www.tinyml.org

Copyright Notice

www.tinyml.org