tinyML® Talks

Enabling Ultra-low Power Machine Learning at the Edge

“Unlocking new frontiers in wearable technology with TinyML”

Kenneth Joel – Co-Founder & Director, Sensio Enterprises

March 21, 2024

www.tinyML.org

TINY ML TALKS webcast

Thank you, **tinyML Strategic Partners**,
for committing to take tinyML to the next Level, together


TINY ML TALKS webcast

Executive Strategic Partners

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

Edge cloud

Cloud

IoT/IIoT

Automotive

Mobile

Accelerate Your Edge Compute

SYNTIANT

Making Edge AI A Reality

www.syntiant.com

TINY ML TALKS webcast

Platinum Strategic Partners

TINY ML TALKS webcast

AITRIOS

DEPLOY VISION AI AT THE EDGE AT SCALE

SONY

TINY ML TALKS webcast

Gold Strategic Partners


Build the
Future of tinyML
on arm

TINY ML TALKS webcast

EDGE IMPULSE

The Leading Development Platform for Edge ML

edgeimpulse.com

Driving decarbonization and digitalization. Together.

Infineon serving all target markets as
Leader in Power Systems and IoT

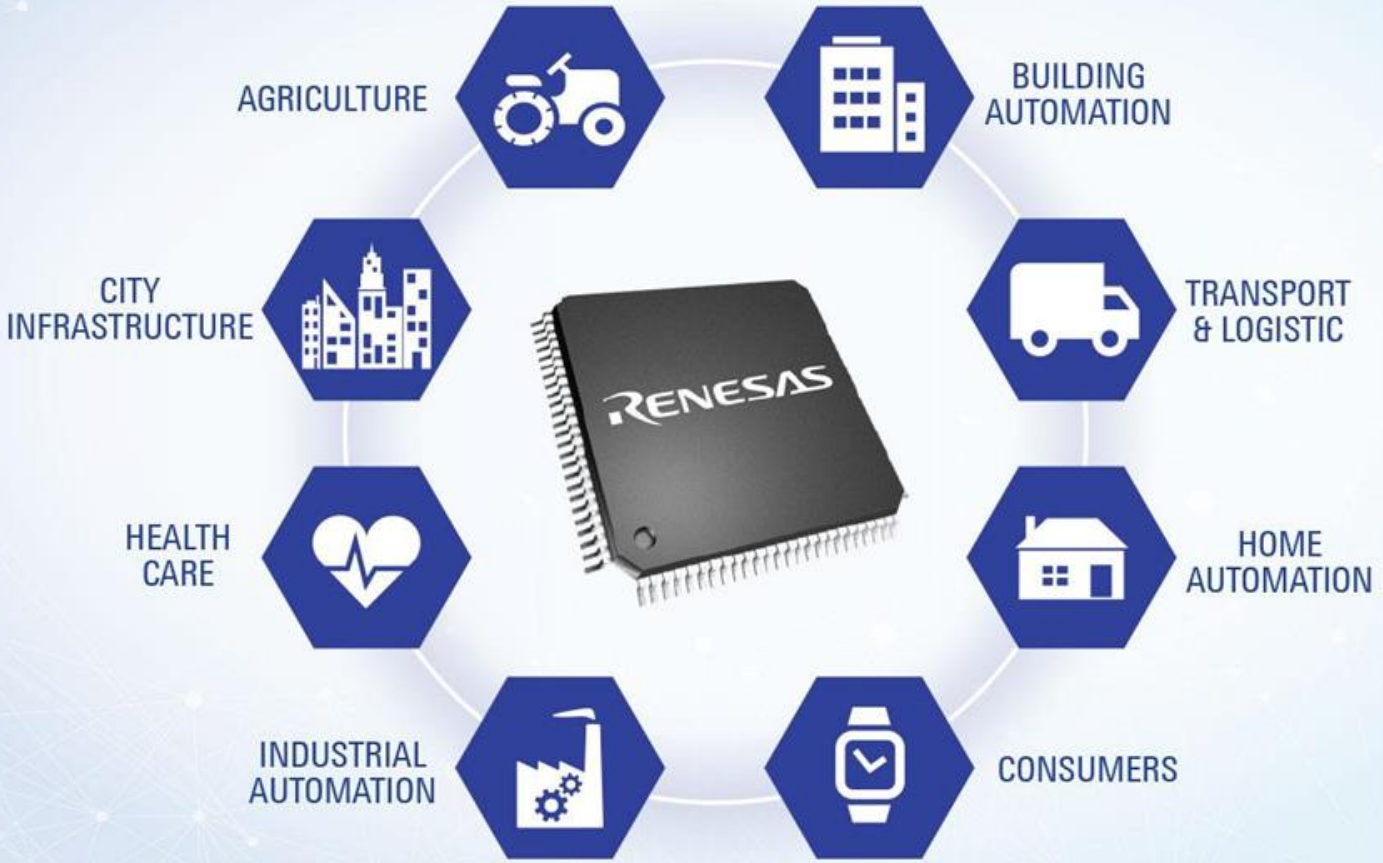
www.infineon.com

innatera

NEUROMORPHIC INTELLIGENCE FOR THE SENSOR-EDGE

www.innatera.com

Renesas is enabling the next generation of AI-powered solutions that will revolutionize every industry sector.



renesas.com

STMicroelectronics provides extensive solutions to make tiny Machine Learning easy

life.augmented

www.st.com/ai

synaptics

ENGINEERING EXCEPTIONAL EXPERIENCES

We engineer exceptional experiences for consumers in the home, at work, in the car, or on the go.

www.synaptics.com

TINY ML TALKS webcast

Silver Strategic Partners

Join Growing tinyML Communities:

19.9k members in
49 Groups in 41 Countries

tinyML - Enabling ultra-low Power ML at the Edge

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

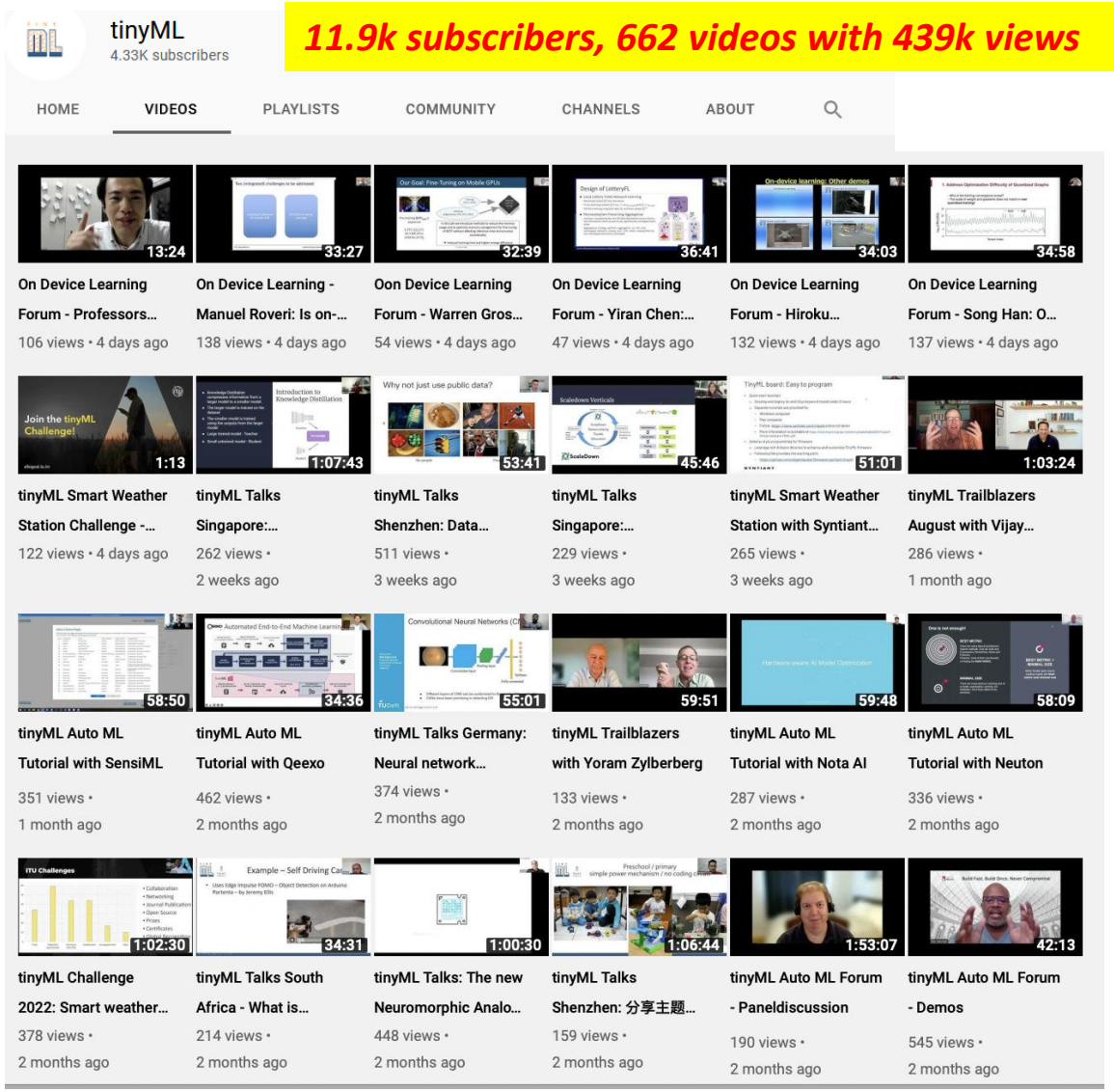
4.2k members
&
15k followers

The tinyML Community

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

TINY ML TALKS webcast

YouTube



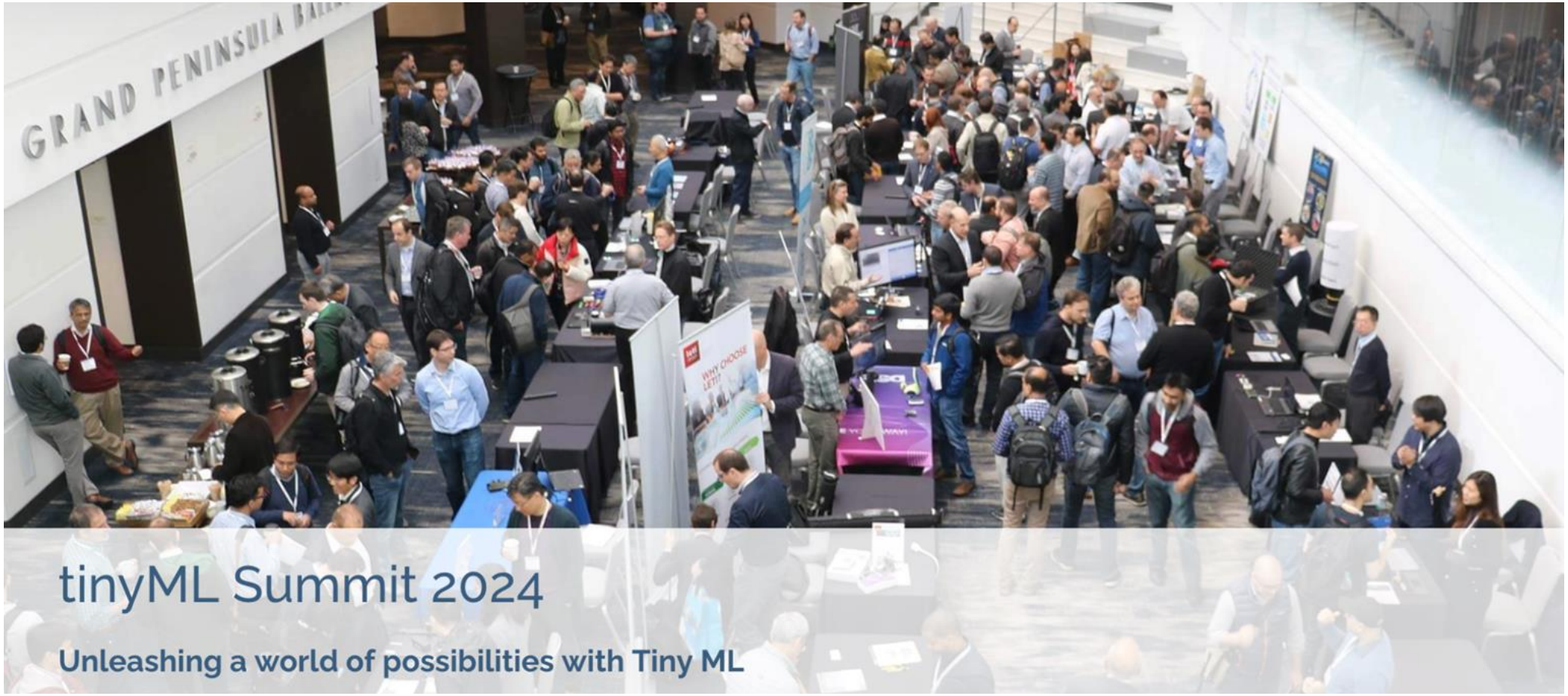
Subscribe to tinyML YouTube Channel for updates and notifications

(including this video)

www.youtube.com/tinyML

TINY ML TALKS webcast

tinyML Summit
April 22 - 24, 2024 - Register now!



tinyML Awards 2024

WEVOLVER

https://www.wevolver.com/article/2023-edge-ai-technology-report

TINY ML FOUNDATION

2023 Edge AI Technology Report

The guide to understanding the state of the art in hardware & software in Edge AI.

TINY ML TALKS webcast

Reminders

Slides & Videos will be posted tomorrow

PDF

tinyml.org/forums

youtube.com/tinyml

Please use the Q&A window for your questions

TINY ML TALKS webcast

Kenneth Joel

From creating his own sensor based shoes to creating AI-enabled software and Algorithms for Sensio, Kenneth shows excellence in his craft. He brings a fresh perspective to Sensio and the MedTech industry, thanks to his wide-ranging interests from music and fitness to technology. He is passionate about Sensors and AI. Kenneth has been an inventor from a young age, and is keen to create scalable solutions to large problems faced in the industry. At Sensio, Kenneth holds charge of all things Software, Process and HR, also contributing to Business Development. He holds a B.Tech Degree in Electrical and Electronics Engineering from PES University.

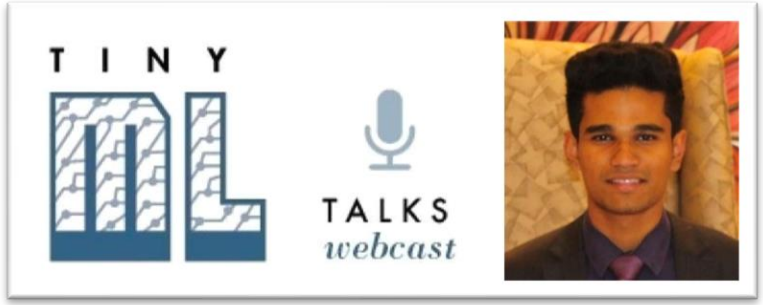
Unlocking new frontiers in wearable technology with TinyML

TinyML Use-cases and Applications

Disclaimer

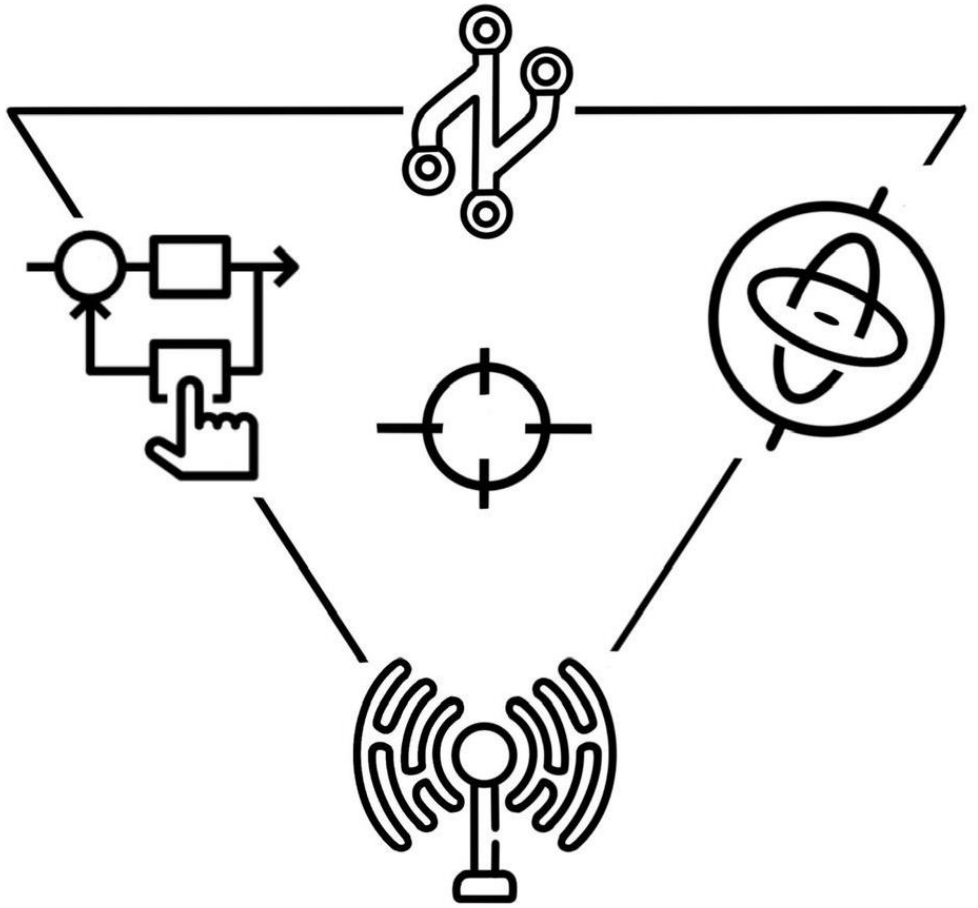
Outline

- Intro to TinyML & Myself
  - Conceptual Intro to TinyML
  - Impact of TinyML community
- About Sensio
  - What we do
  - Design Philosophy & TinyML
  - Products
- Orbyt Smart Ring – A tiny wearable powered by TinyML



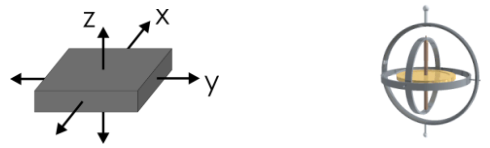
Introduction

- Electrical & Electronics Engineer
- Interested in Embedded Systems, Sensors & Algorithms
- Actively avoided ML, thought it was overhyped
- Instead focused on
  - Signal Processing
  - Control Systems
  - Analog & Digital Filter Design
  - Modern Control Theory
  - Sensor fusion

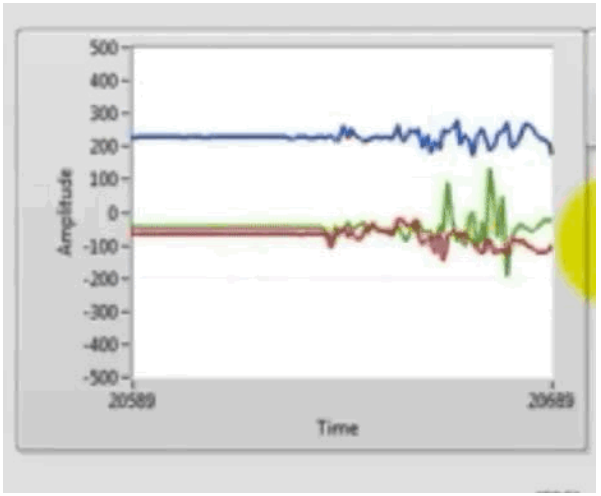


Introduction

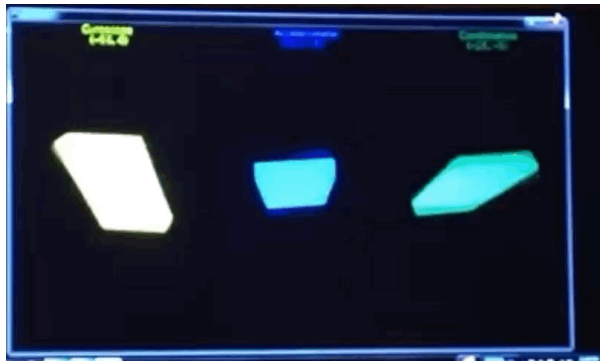
- Working with inertial sensors was a turning point



- Around 8-9 years ago you had both analog accelerometers and smart IMUs



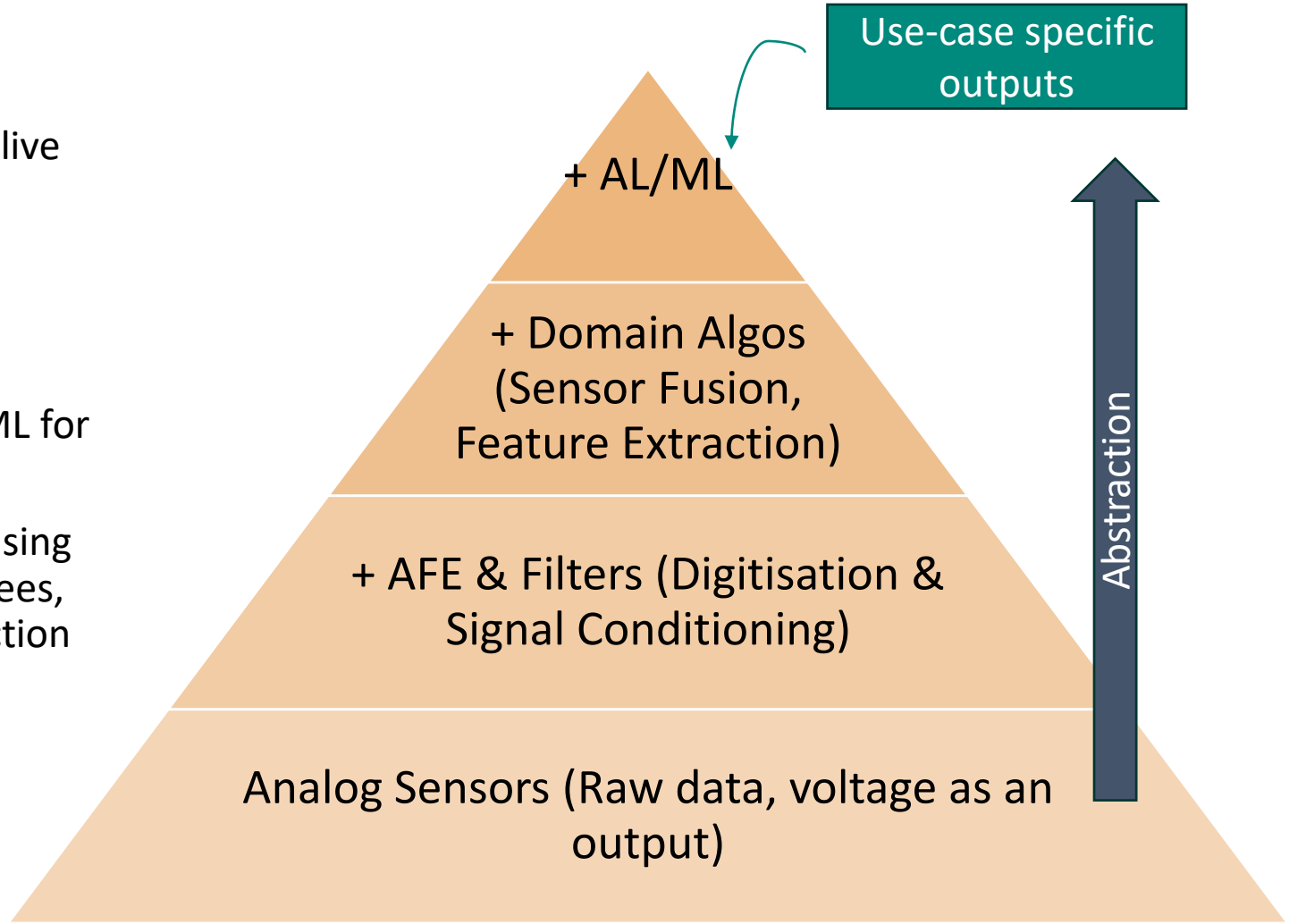
- Notice that the inferences are far away from the sensors data
- Got me thinking about sensor abstraction:
- How many layers can you abstract away while being generally useful?



- How much intelligence can you add to the sensor itself?

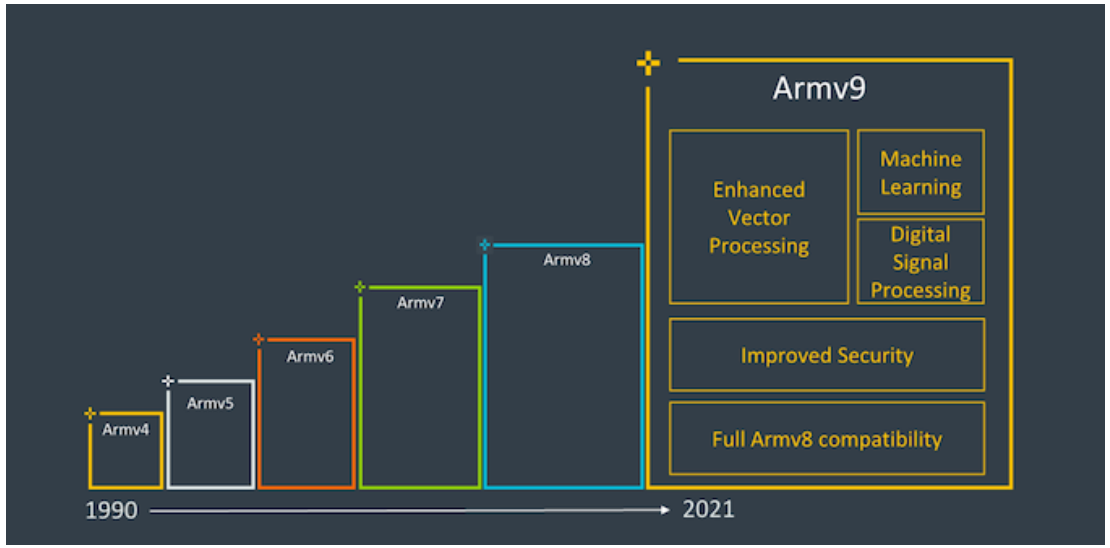
Abstraction

- TinyML → raising the abstraction level of simple sensors using low power (1mW) algorithms that live very close to the sensors themselves.
- 2017 → that one of the leading MEMS sensor companies announced a hackathon on AI/ML algorithms for an IMU
- The industry was already dropping hints about ML for Sensors
- I started *reluctantly* solving sensor problems using as little ML as possible (simple SVMs, decision trees, clustering) still investing heavily in feature extraction and domain rules
- Saw the usefulness of an ML based approach.
- Edge deployment wasn't a reality just let



Present State

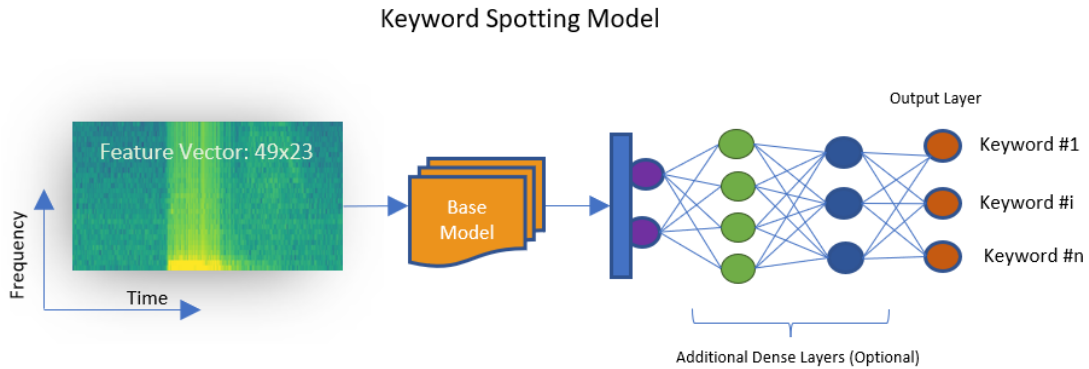
- In 2024 you have inertial sensors that you can buy for use-case specific algorithms like step counting or gesture recognition.
- Some even come with ML cores and Finite State machines
- MCU Ecosystem: ARM v9 came out in 2021 with a focus on DSP & ML
- Advantages were clear:
  - Offload the main processor even for application specific tasks, saving power.
  - Use sensors to detect use-case specific events to wake up the application processor, saving power.



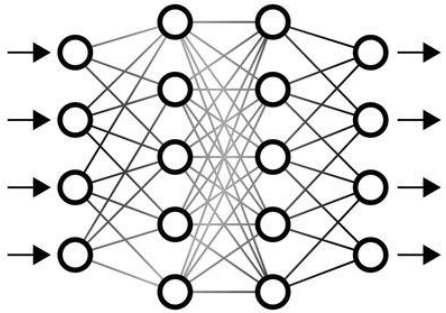
Source: AnandTech

Limitations & Future Scope

- DL Specific hardware acceleration has made it's way to embedded MCUs.
- Can run a 50-100 layer, 2-4 million weight CNN on an MCU sipping just 2-3mW
- Lot's of TinyDL demos in Audio & Video Domain
- Sensor based applications in wearables still use mostly TinyML, which is deep domain rules + ML
- This may be because
  - Wearables are both form & BoM sensitve no space or budget for DL acceleration
  - No need to force DL – Problems can be solved well with ML + Custom DSP
  - No large labelled datasets like audio and video
- Next Steps
  - New Problems that actually benefit from DL
  - Physics/Bio informed DL as a middle ground

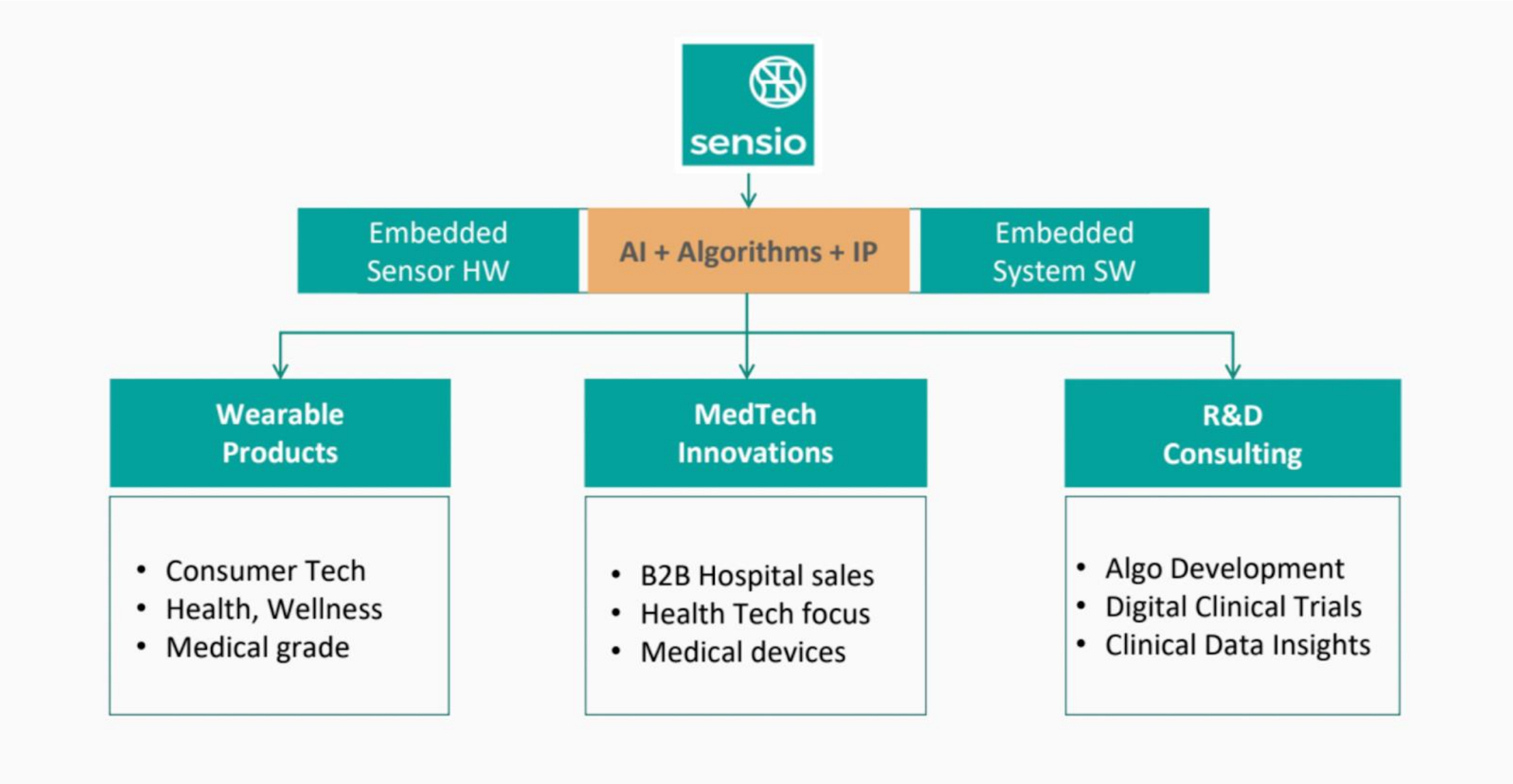


Source: SensiML



About Sensio

Sensio → Abstracting Health Sensors

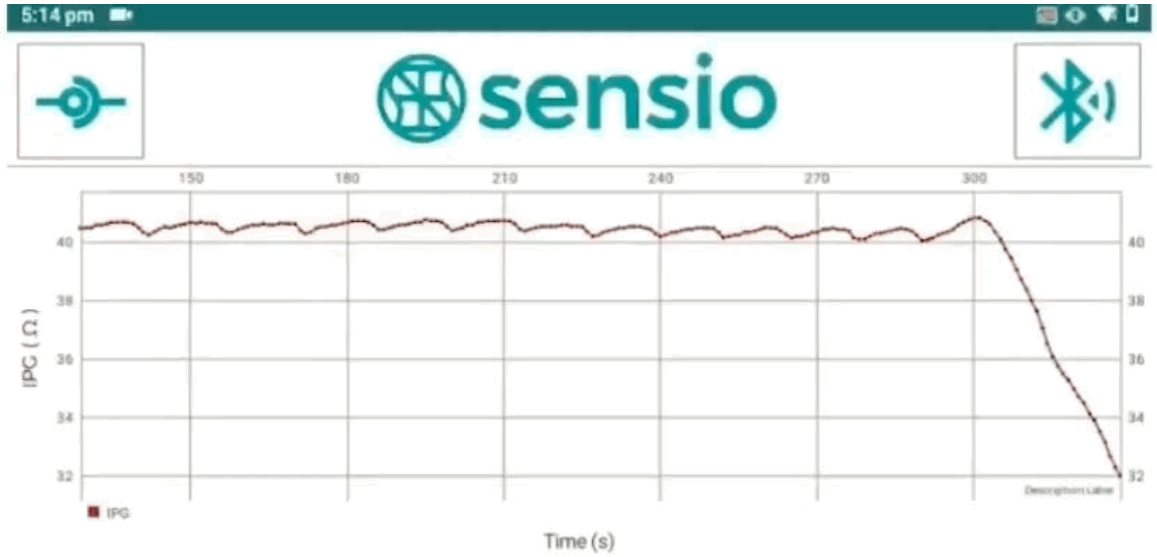


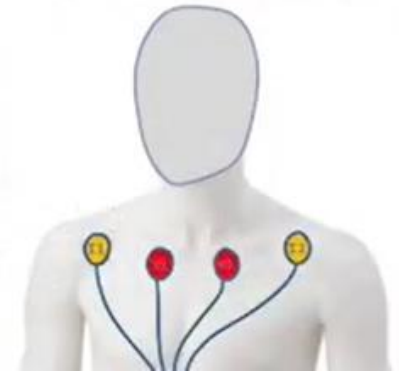
Abstracting Health Sensors

- IMUs behaving more like step counters and gesture detectors, sometimes with just an interrupt pin as the interface.
- Wearables made it necessary for more health sensors to be abstracted
- A system integrator can now purchase an integrated AFE + Sensorhub solutions for HR sensing or SpO2 sensing
- At Sensio we work on more niche sensors and sensor combinations

Respiration Sensor

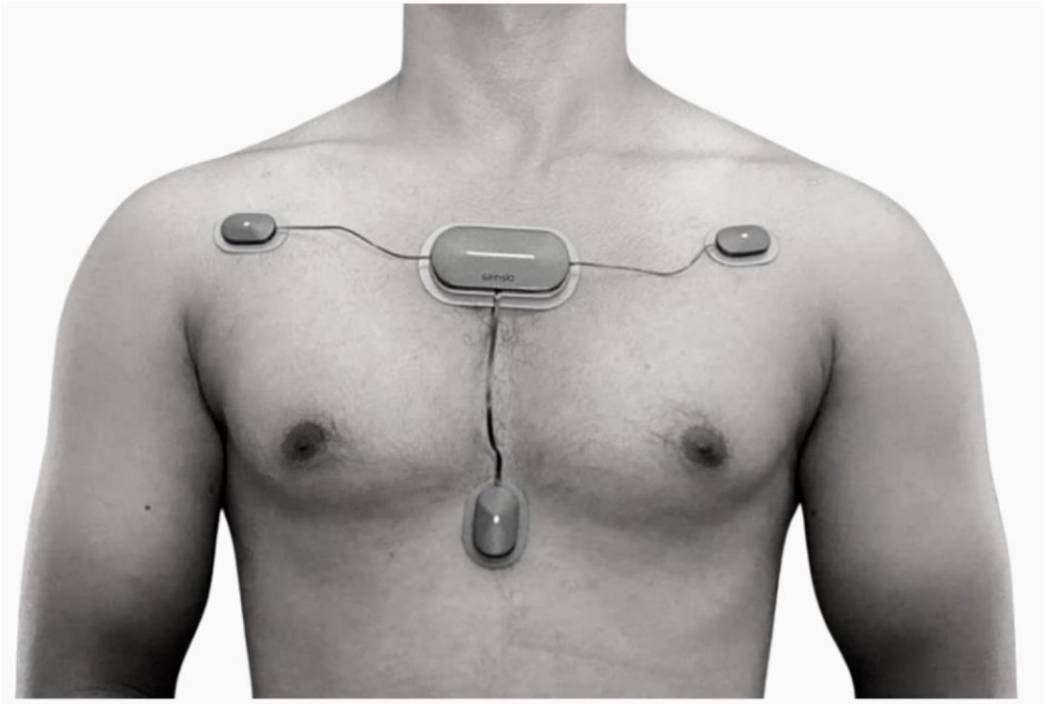
- IPG/BioZ is the base signal
- We have our own AFE, better low power AFEs are readily available
- Our abstraction gives you a module that directly gives
  - Respiration rate
  - Inspiration/Expiration Ratio
  - Respiration Depth
  - Heart Rate*
- No need to work directly with the AFE and raw signal

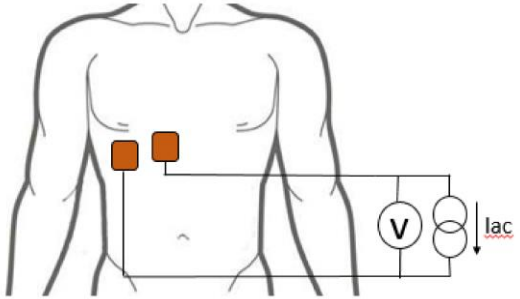


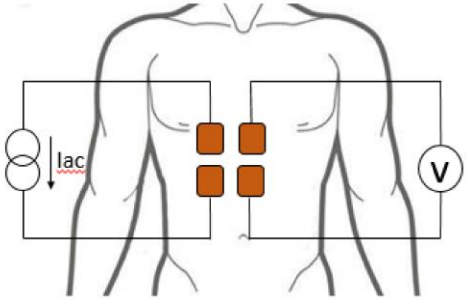


Chest Patch

- Example Deployment – Chest Patch
- Combines Sensio Resp Module with other 3rd party AFEs and Sensor Hubs
- Chest patch includes
  - Resp - IPG
  - HR / HRV – ECG
  - Actigraphy – IMU
  - SpO2 - PPG







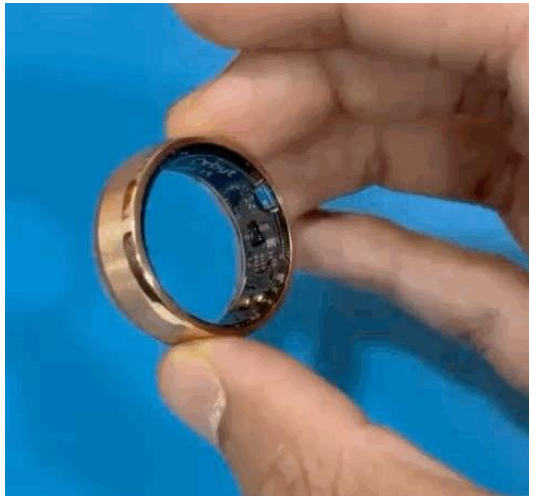
IPG Electrode Placement
Image Source: Wireless Impedance Pneumography System for Unobtrusive Sensing of Respiration, Pable et al

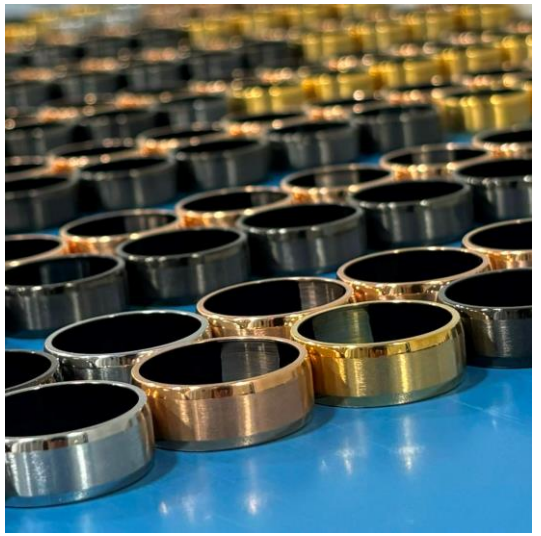
Orbyt Smart Ring

The Tiniest of Wearables Powered by TinyML

Tiny Wearables

- Credits to the TinyML community for inspiration, a talk in 2021 about sensor miniaturisation and power efficiency trends.
- WLP packaging means embedded MCUs, AFEs, PMICs are much smaller now
- It’s been a ~3 year long R&D journey, getting ready for our Private Beta now.
- Visit → Orbytring.com to sign up





Just how tiny?

- Tiny in Size:
  - Weighs 3-5gms,
  - 2.7-2.8mm thick
  - 9mm wide

- Tiny in power consumption
  - Average power <1mW

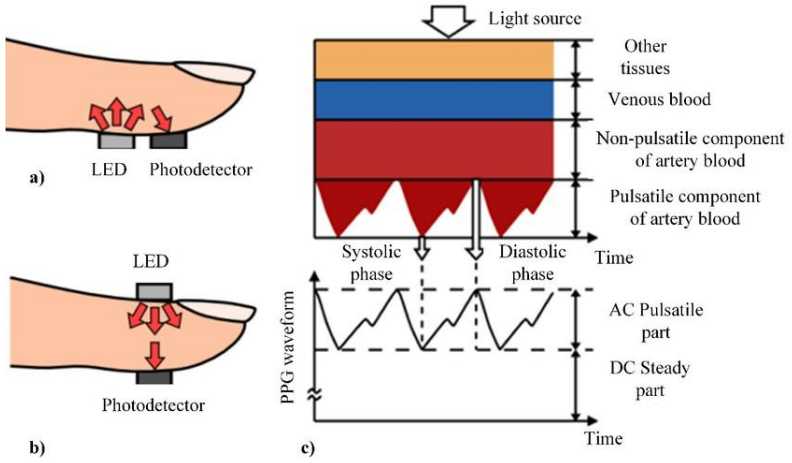
Why a Smart Ring?

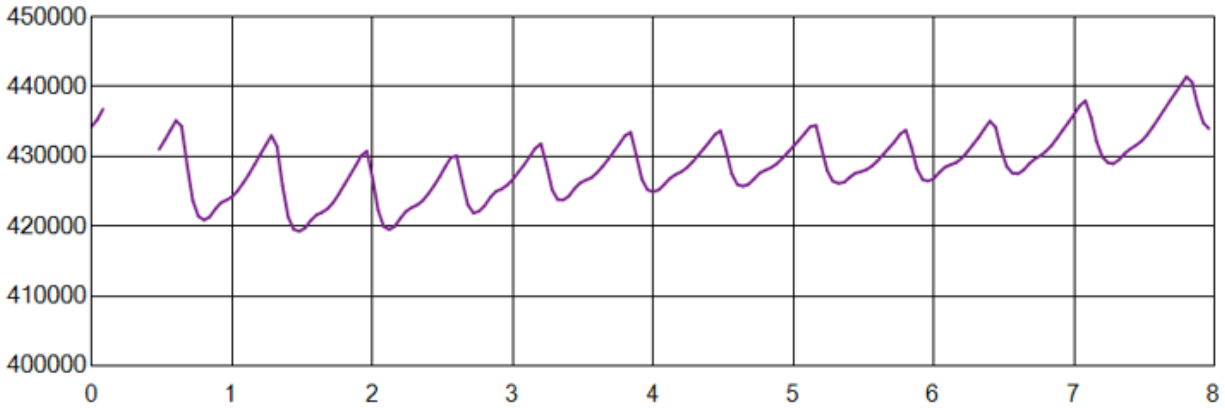
- Minimalism → Both aesthetic and digital

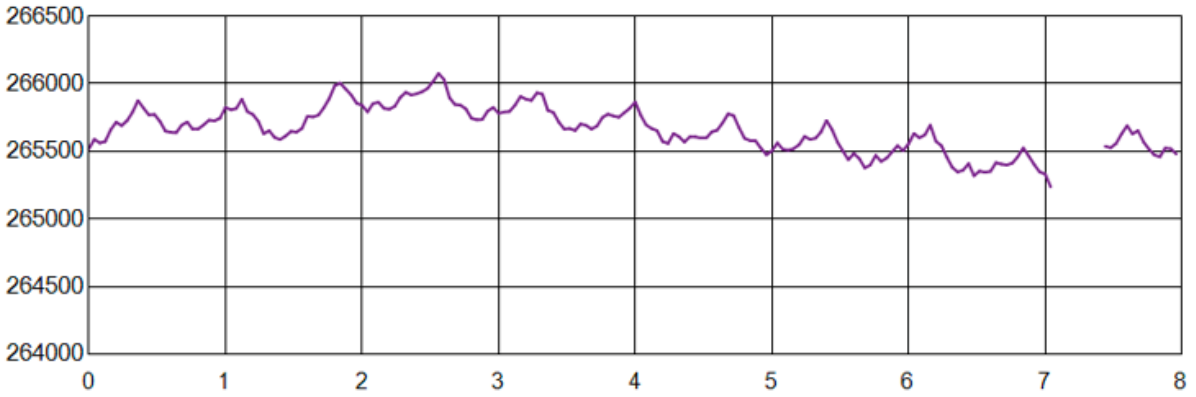
- Easy to wear 24x7
  - Sleep Friendly
  - Shower Friendly (water proof)

Why a Smart Ring?

- Photoplethysmography Signal Quality







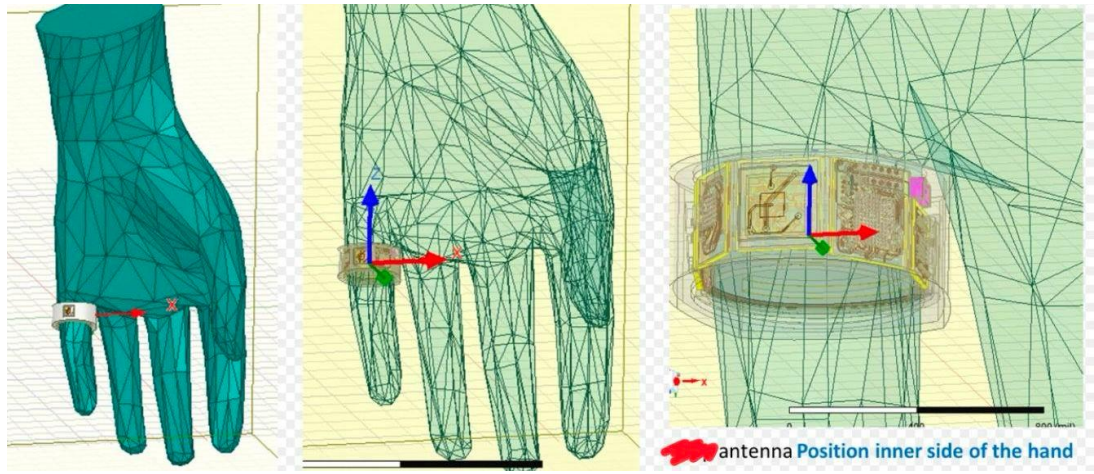
- Wrist vs Palm/Finger
  - Better Perfusion
  - More consistent contact

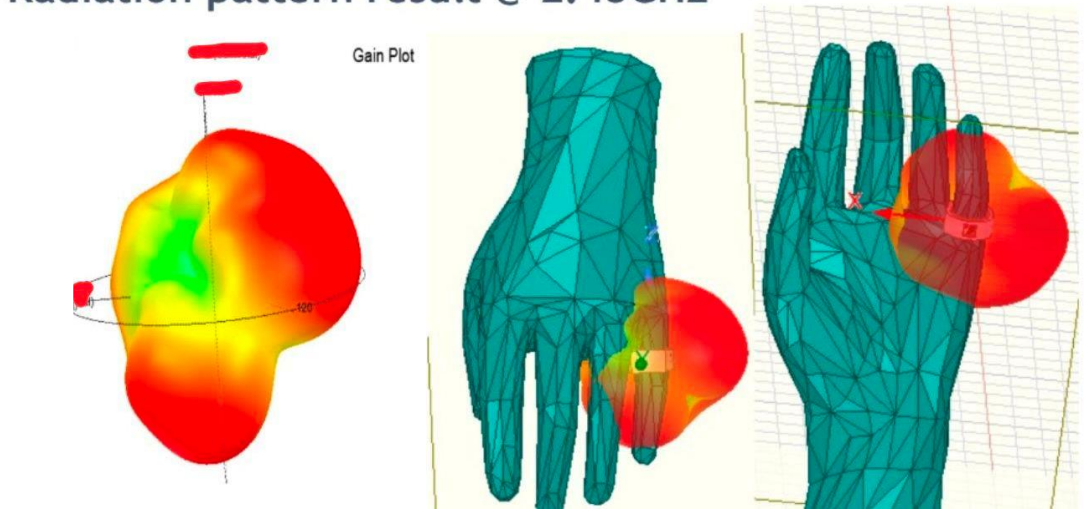
PPG Image Source: Human Emotion Regocnition, Andrius Dzedzikus et al

Design Considerations Unique to Smart Rings

BLE Radio

- Most power hungry subsystem in screenless wearables
- Wireless Signals and Metal Enclosures don’t get along
- Need higher Tx power or Coded PHY to compensate
- Generally Less BLE transmission means more battery life
- With low power MCUs doing most of the number crunching on the extreme edge means less data to send

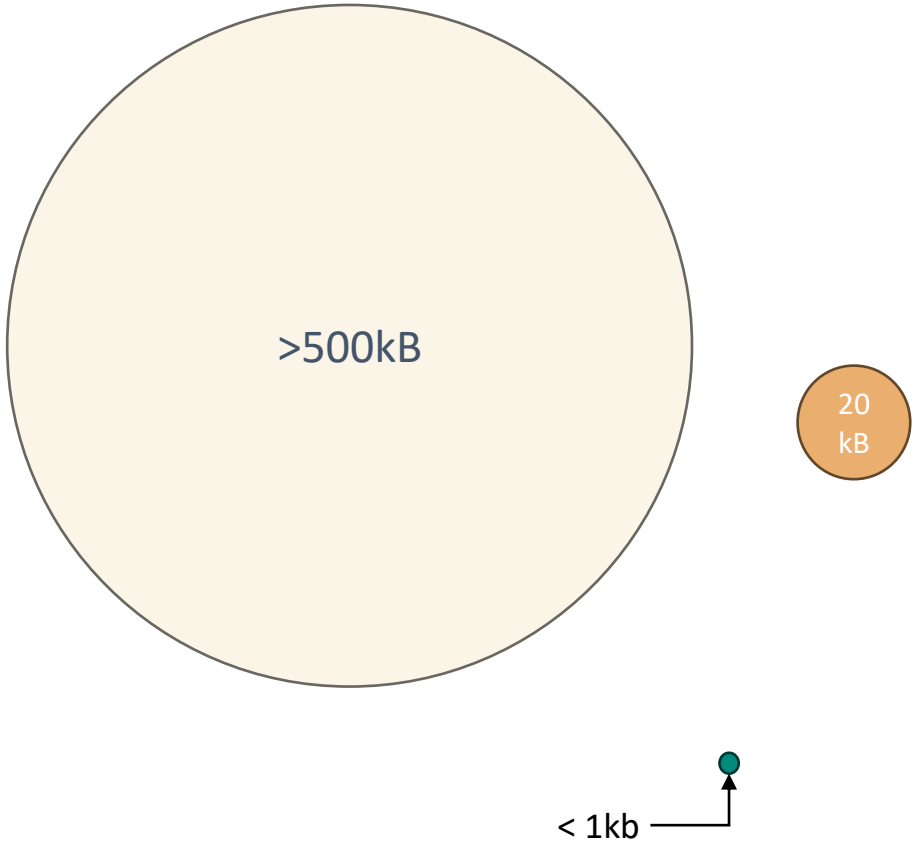




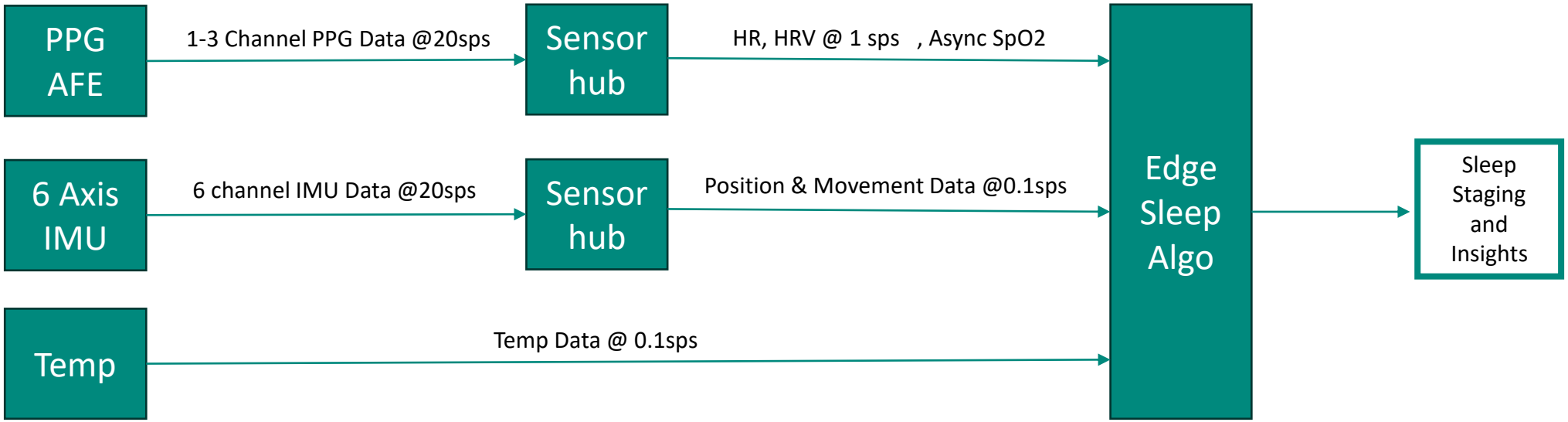
Design Considerations Unique to Smart Rings

Flash Memory

- Flash Read/Writes are expensive, especially if it's an external flash
- Use real time TinyML to store more abstract inferences rather than raw data
- Example: Sleep Algos
  - One nights sleep staging data : <1kB
  - Inferences required to stage sleep : <20kB
  - Raw data required to stage sleep: 600-900kB
  - Hybrid approach: >500kB
  - Storing raw data for a few nights of sleep tracking means you need a big flash.

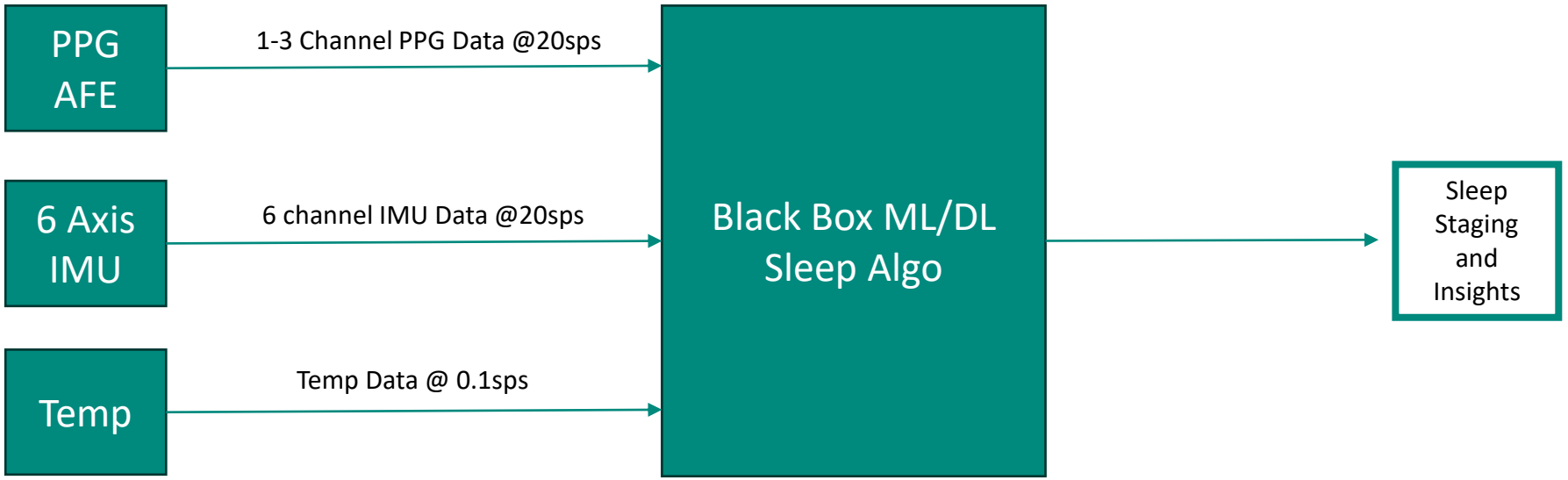


Stacking Simple ML Models



Advantages: Independently useful intermediately inferences, easily swappable intermediate IP blocks

Stacking Simple ML Models



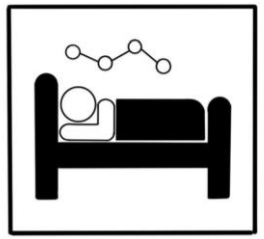
Design Trade-off: Model Size should be less than buffered inference data in previous approach

TinyML Benefits

- Running Costs
  - System Power
  - Cloud Costs
- Reduce Latency & Protect Privacy
  - Inferences happen in real time
  - No external uncertainties
  - Raw data stays on device
- Abstraction of Sensors
  - Offloading the application core
  - Faster dev by using pre built algo IP blocks
- Custom Models
  - Tailor Made models for each user
  - More cost effective to deploy to edge than host multiple instances in cloud

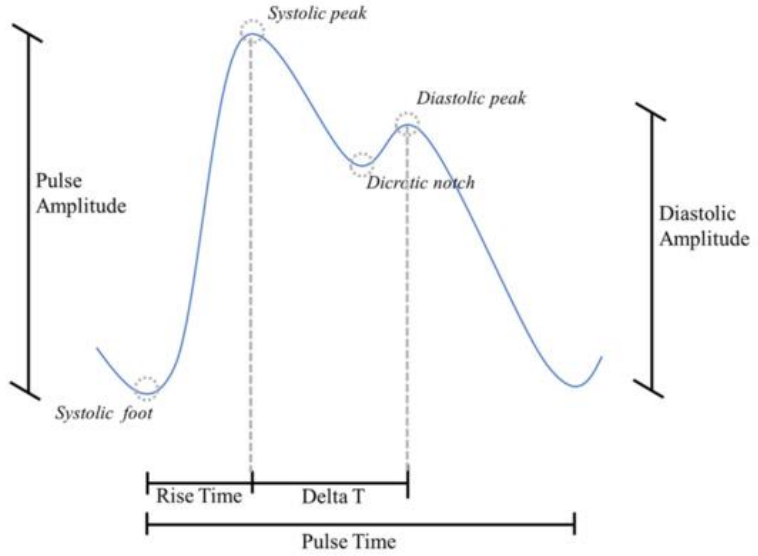
Custom Models

- A unique opportunity for Vertically Integrated Startups
- Be vertically integrated wrt. your goal
- Integrate Offline services that reinforce product based insights
  - Stress Monitoring Algos and Cortisol tests
  - Sleep Monitoring Algos and PSG tests
  - Cuffless BP Estimation and ABPM services
- Edge model can be tweaked based on agreement with results from medical tests.
- Analogy: User specific model tuning for speech recognition

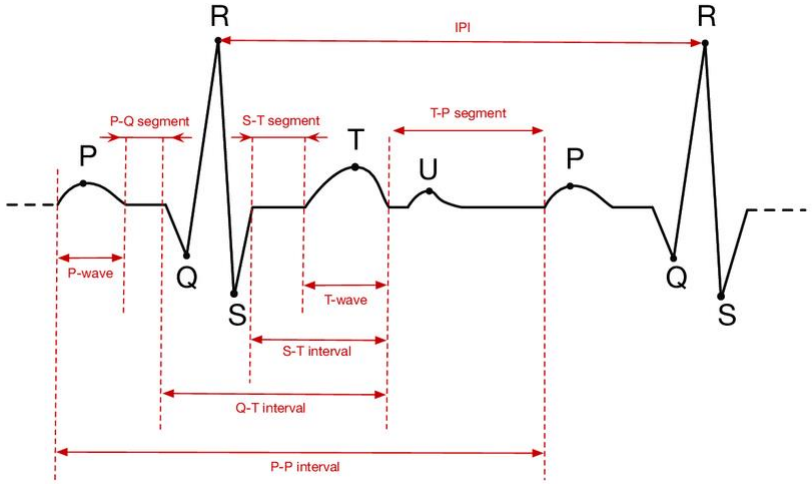


Open Problems Requiring TinyML solutions

- Pulse Morphology Analysis
  - PPG waveform contains rich information about vasculature
  - May be the answer to more accurate BPT algos

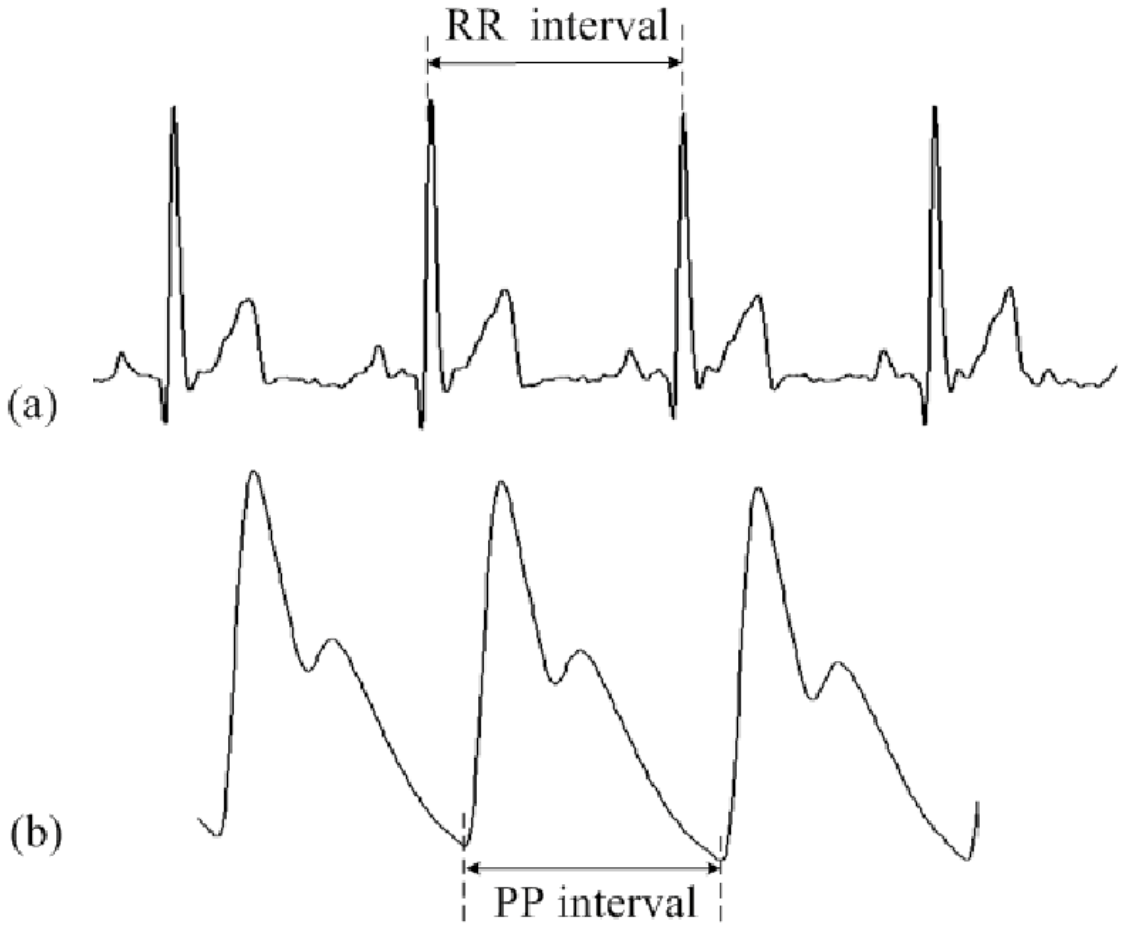


- User Auth & Identification using ECG
  - Similar to Face & Voice recognition but using your heart's unique electrical signature



Open Problems Requiring TinyML solutions

- HRV vs PRV
  - HRV contains a lot of useful information about stress and recovery. To do with nervous system and electrical behaviour of the heart, only ECG gives true PRV
  - Pulse rate variability has been sold to us as heart rate variability
  - Rationale: Low accuracy PRV over a longer time period better than instantaneous HRV
  - Relationship between the two is complex
  - Can we design user specific calibration algorithms to better reconcile the two?



Thank you

Copyright Notice

www.tinyml.org

Copyright Notice

www.tinyml.org