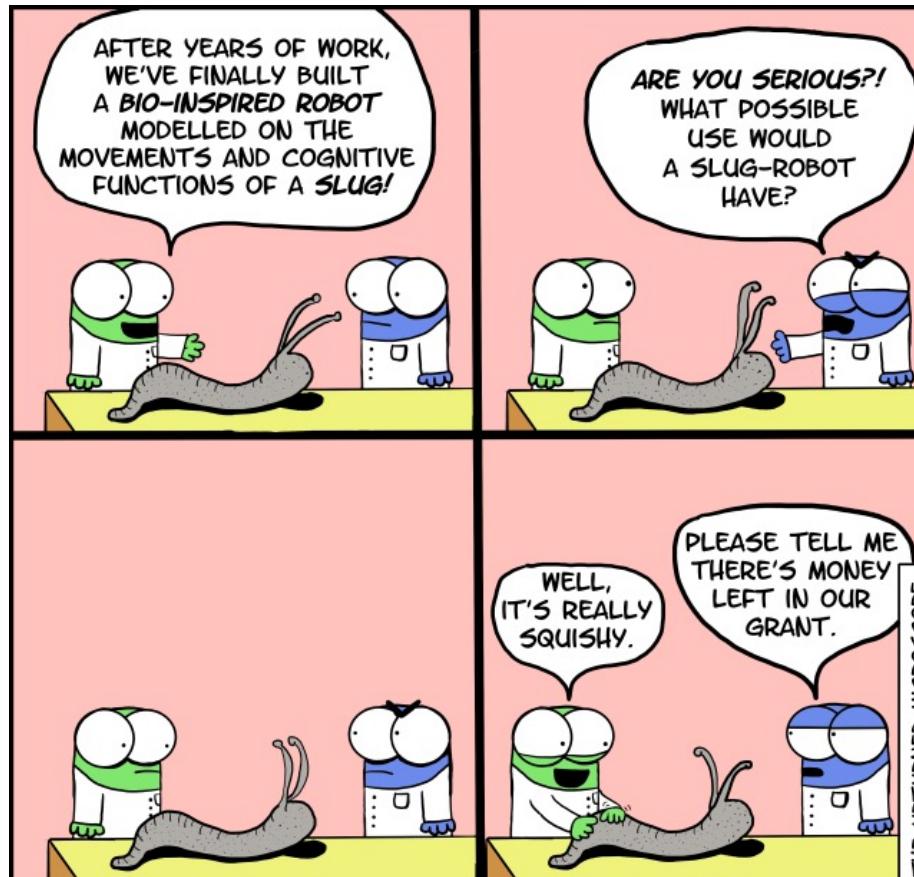




Lecture 4: “On-chip Intelligence, Bio-inspired IoT, and Bio-sensing Wearable Computing”

(the slides is partially adapted from Daniel Situnayake and Vikram Iyer's presentations)



What are ... hot?

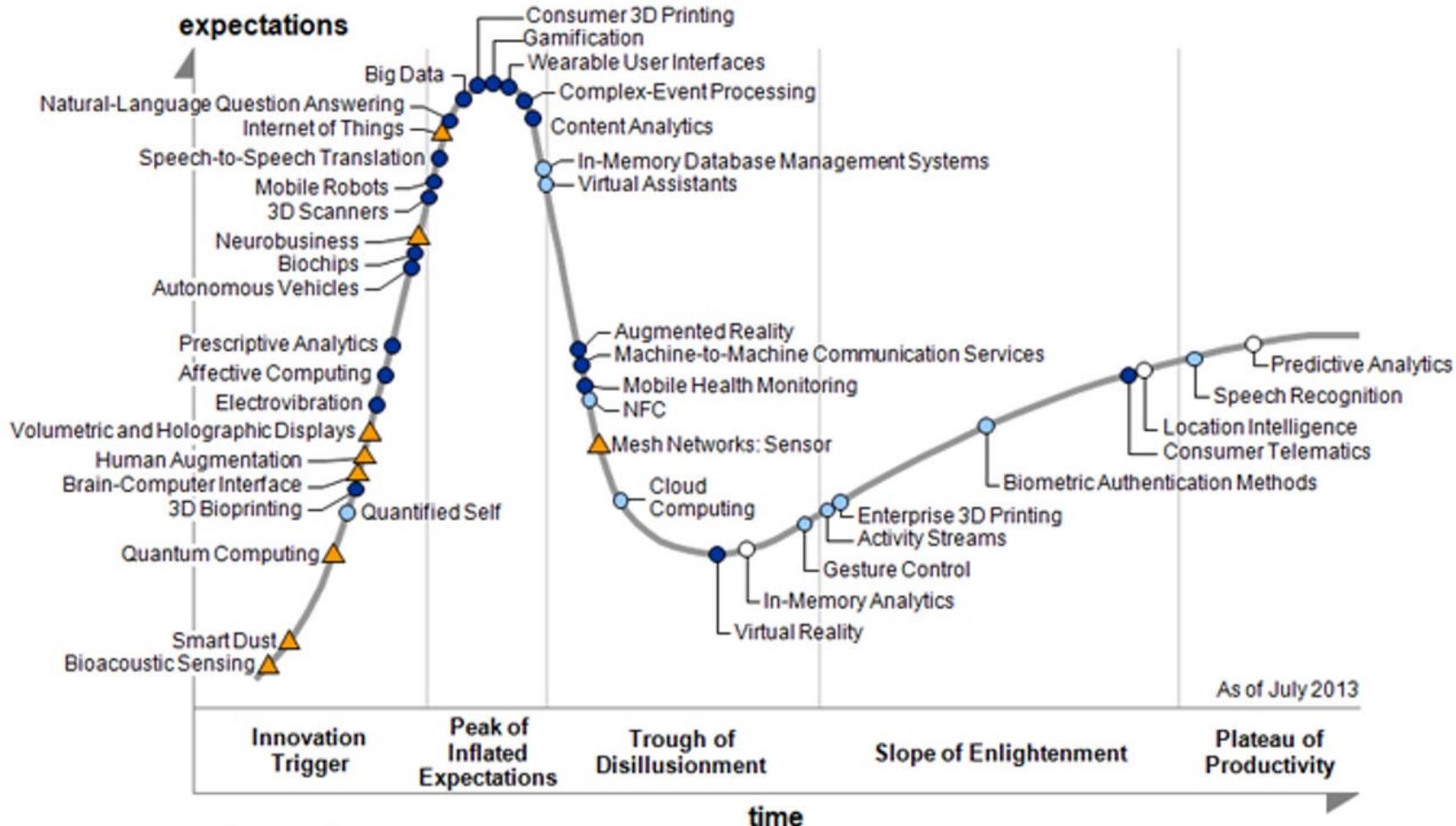
Hype Cycle for Emerging Technologies, 2020



gartner.com/SmarterWithGartner

Source: Gartner
© 2020 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner and Hype Cycle are registered trademarks of Gartner, Inc. and its affiliates in the U.S.

Gartner®



Plateau will be reached in:

○ less than 2 years

○ 2 to 5 years

● 5 to 10 years

▲ more than 10 years

obsolete
✖ before plateau

Gartner

(1) Tiny Machine Learning with TensorFlow Lite Microcontroller

TensorFlow Lite?

TensorFlow Lite is a production ready, cross-platform framework for deploying ML on mobile devices and embedded systems

Basics of Machine Learning



Basics of Machine Learning – Activity recognition



```
if(speed<4){  
    status=WALKING;  
}
```

Basics of Machine Learning – Activity recognition



```
if(speed<4){  
    status=WALKING;  
}
```



```
if(speed<4){  
    status=WALKING;  
} else {  
    status=RUNNING;  
}
```

Basics of Machine Learning – Activity recognition



```
if(speed<4){  
    status=WALKING;  
}
```



```
if(speed<4){  
    status=WALKING;  
} else {  
    status=RUNNING;  
}
```



```
if(speed<4){  
    status=WALKING;  
} else if(speed<12){  
    status=RUNNING;  
} else {  
    status=BIKING;  
}
```

Basics of Machine Learning – Activity recognition



```
if(speed<4){  
    status=WALKING;  
}
```



```
if(speed<4){  
    status=WALKING;  
} else {  
    status=RUNNING;  
}
```



```
if(speed<4){  
    status=WALKING;  
} else if(speed<12){  
    status=RUNNING;  
} else {  
    status=BIKING;  
}
```



// Oh crap

Basics of Machine Learning – Activity recognition



0101001010100101010
1001010101001011101
0100101010010101001
0101001010100101010

Label = WALKING



1010100101001010101
0101010010010010001
0010011111010101111
1010100100111101011

Label = RUNNING



1001010011111010101
1101010111010101110
1010101111010101011
1111110001111010101

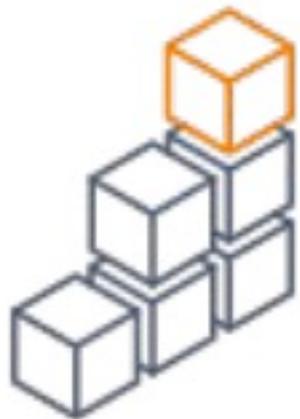
Label = BIKING



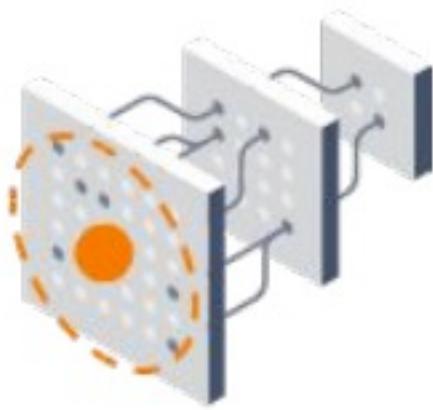
111111111010011101
0011111010111110101
0101110101010101110
1010101010100111110

Label = GOLFING

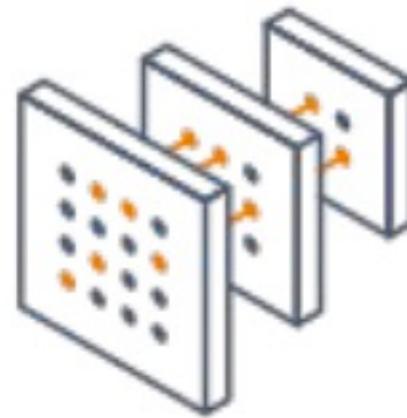
Basics of Machine Learning – Key terms



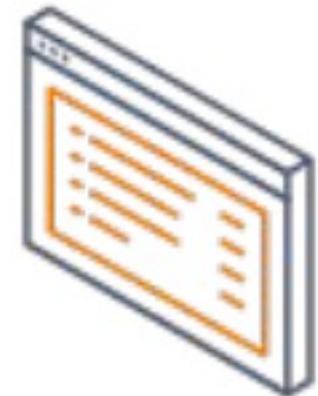
Dataset



Training

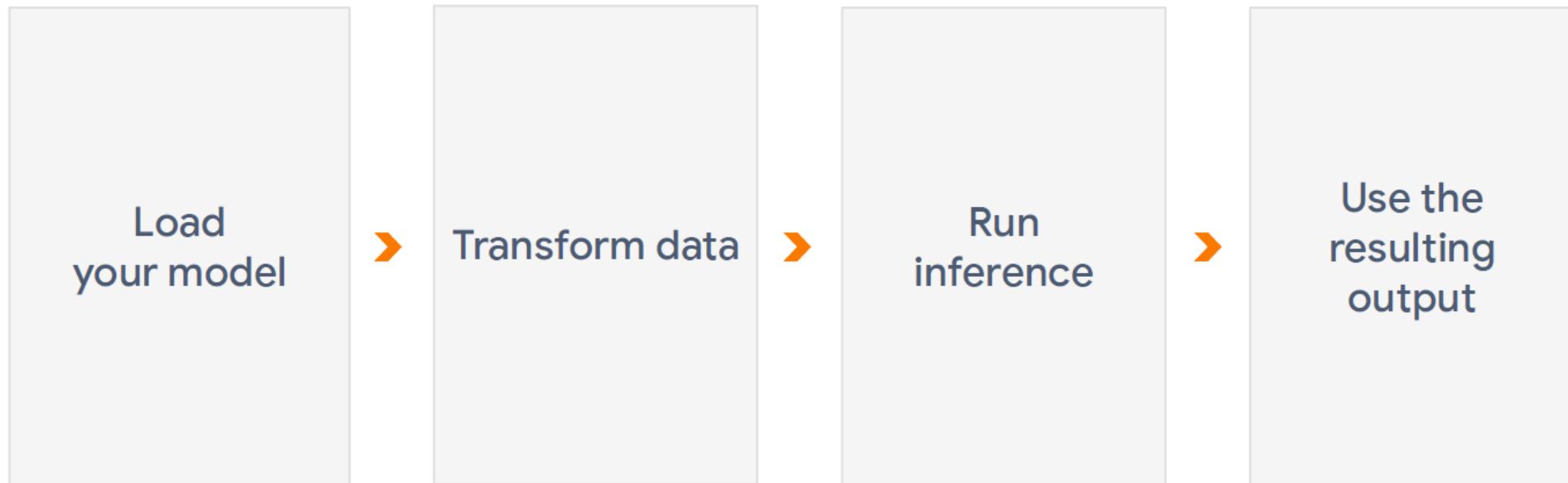


Model

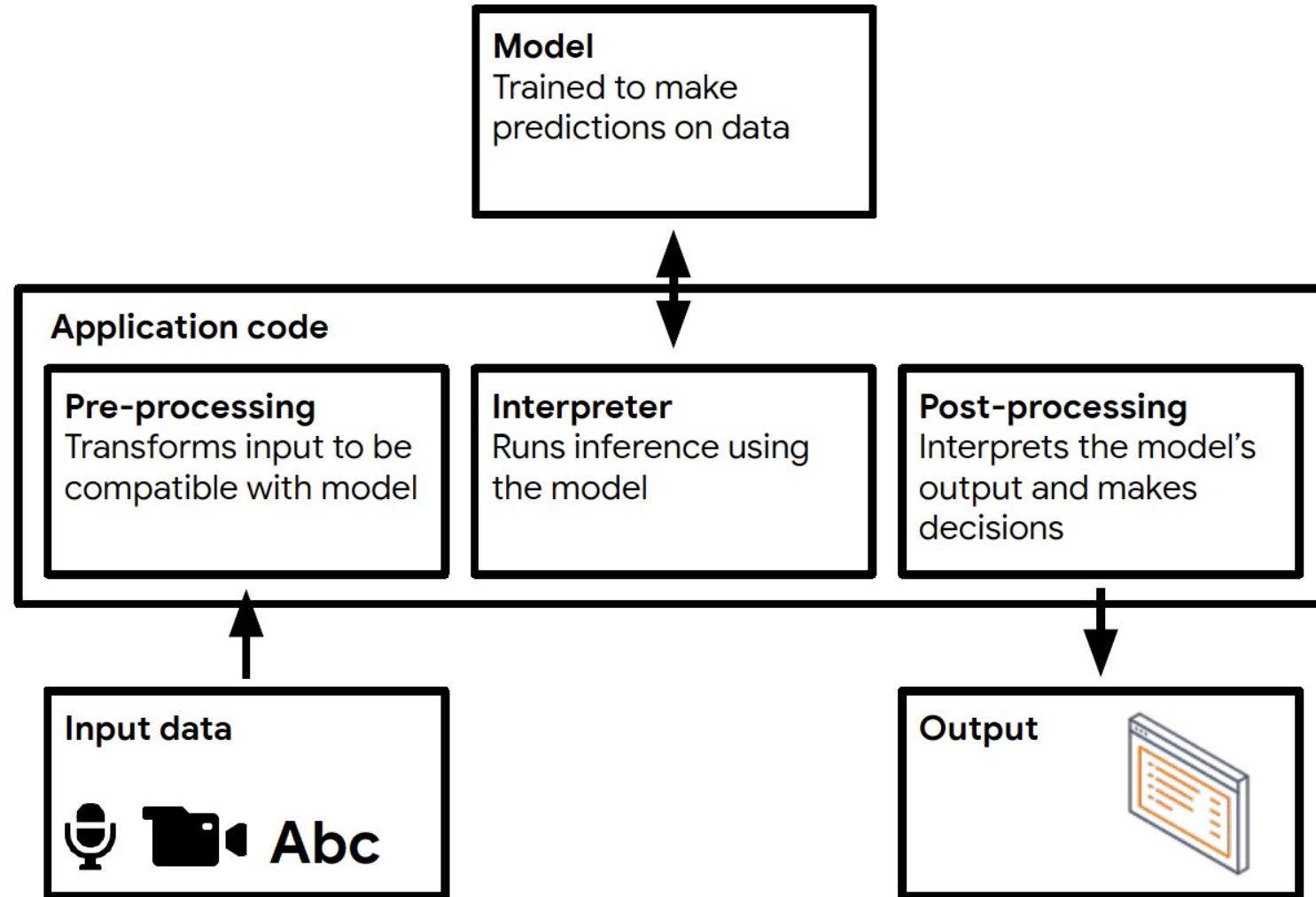


Inference

Basics of Machine Learning – Inference

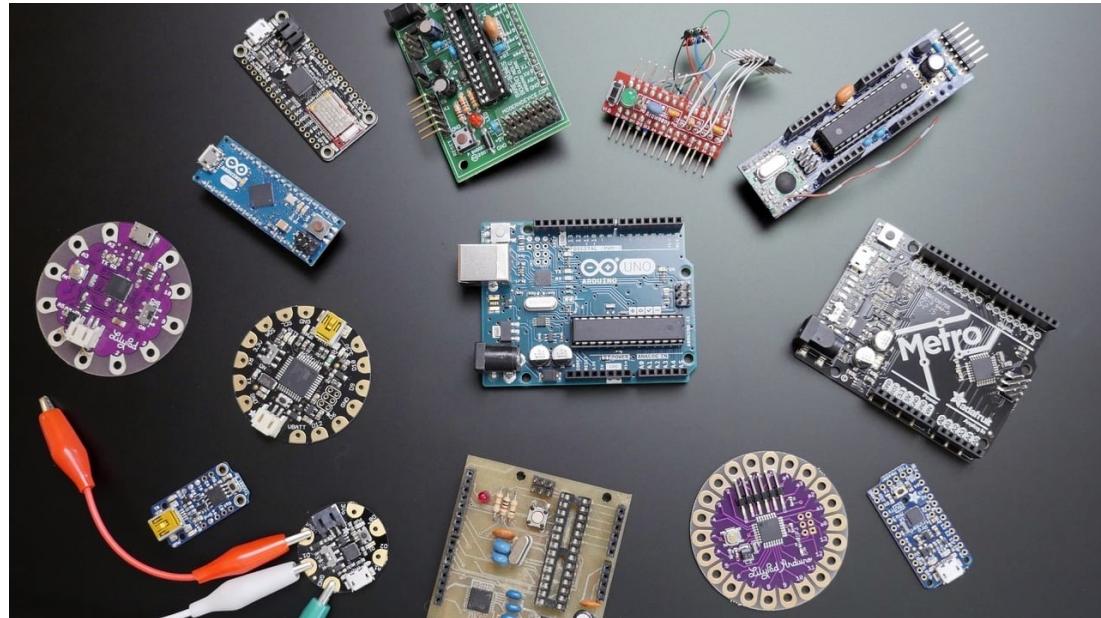


Basics of Machine Learning – TensorFlow Framework



Edge ML Explosion

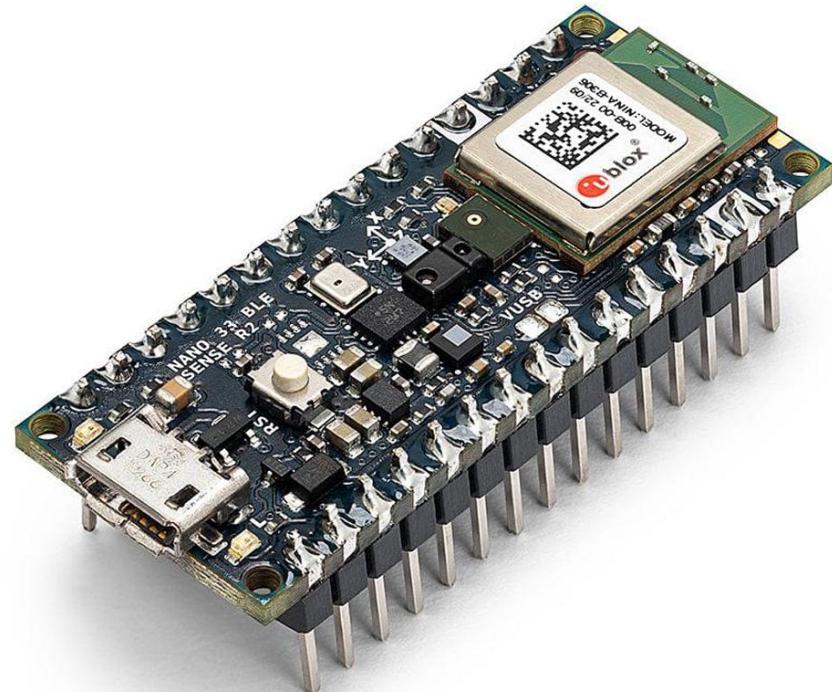
- ❑ Lower latency & close knit interactions
- ❑ Network connectivity
- ❑ Privacy preserving



Edge ML Explosion - Challenges

- ❑ Uses little compute power
- ❑ Works on limited memory platforms
- ❑ Consumes less battery

- ❑ Arduino BLE Sense
 - nRF52840 (32-bit ARM® Cortex™-M4)
 - 64MHz
 - 256 KB SRAM, 1MB flash



TensorFlow Lite

We're simplifying
on-device ML

Convert once, deploy anywhere

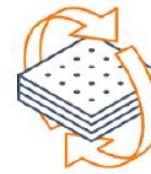


TensorFlow Lite



Pick a model

Pick a new model or retrain an existing one.



Convert

Convert a TensorFlow model into a compressed flat buffer with the TensorFlow Lite Converter.



Deploy

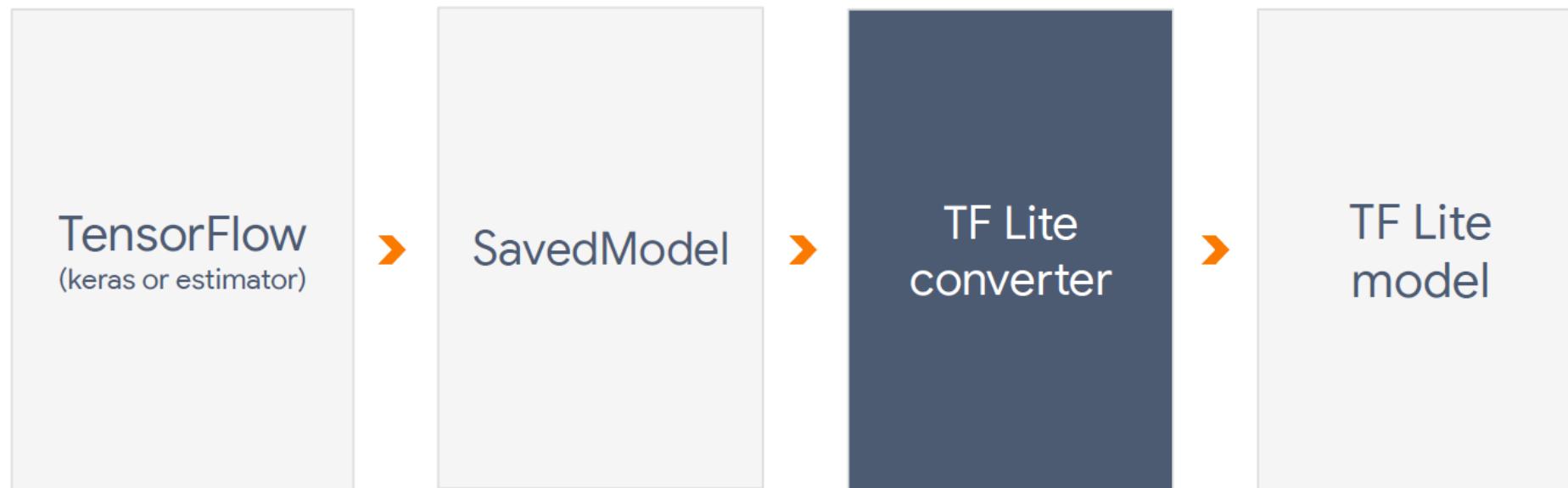
Take the compressed .tflite file and load it into a mobile or embedded device.



Optimize

Quantize by converting 32-bit floats to more efficient 8-bit integers or run on GPU.

TensorFlow Lite – Converting



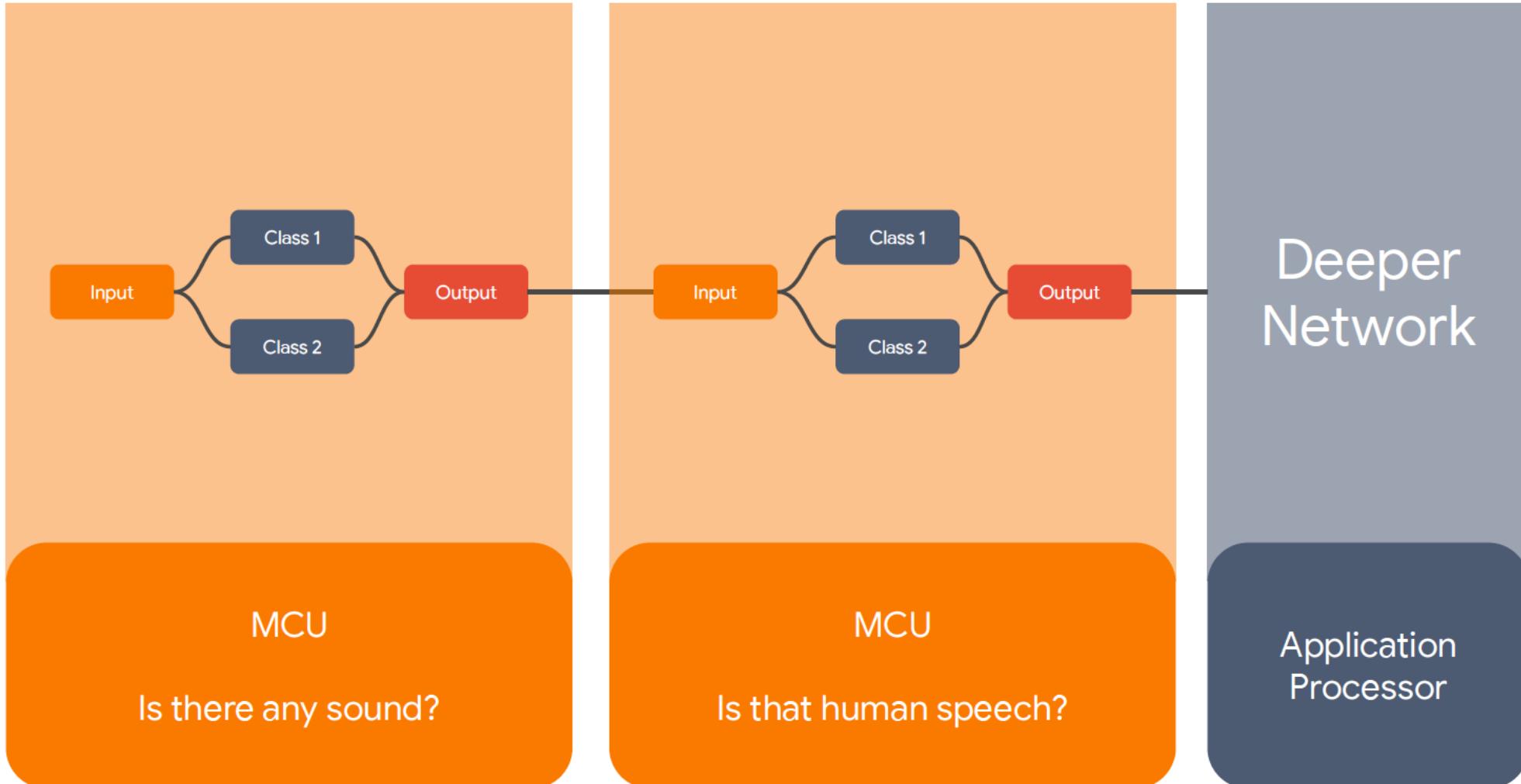
TensorFlow Lite – Running the models



Examples – Speech detection

Demo time

Examples – Speech detection



Key takeaways

- ❑ TensorFlow Lite is a framework to help with **resource-constrained embedded devices**.
- ❑ The model needs to be **trained and optimized before converting**.
- ❑ Not all neural network operations are currently supported.
 - Here is the list of supported operations.
 - https://github.com/tensorflow/tflite-micro/blob/main/tensorflow/lite/micro/all_ops_resolver.cc

(2) Living IoT: A Flying Wireless Platform on Live Insects (ACM MobiCom'19)

Vikram Iyer, Rajalakshmi Nandakumar, Anran Wang, Sawyer Fuller, Shyam Gollakota
University of Washington.

What if Internet of Things (IoT)
devices could fly?

Living IoT: A Flying Wireless Platform on Live Insects

Vikram Iyer, Rajalakshmi Nandakumar, Anca Mwang,
Sawyer B. Fuller, Shyam Gollakota



WT

Drones as flying sensor platforms

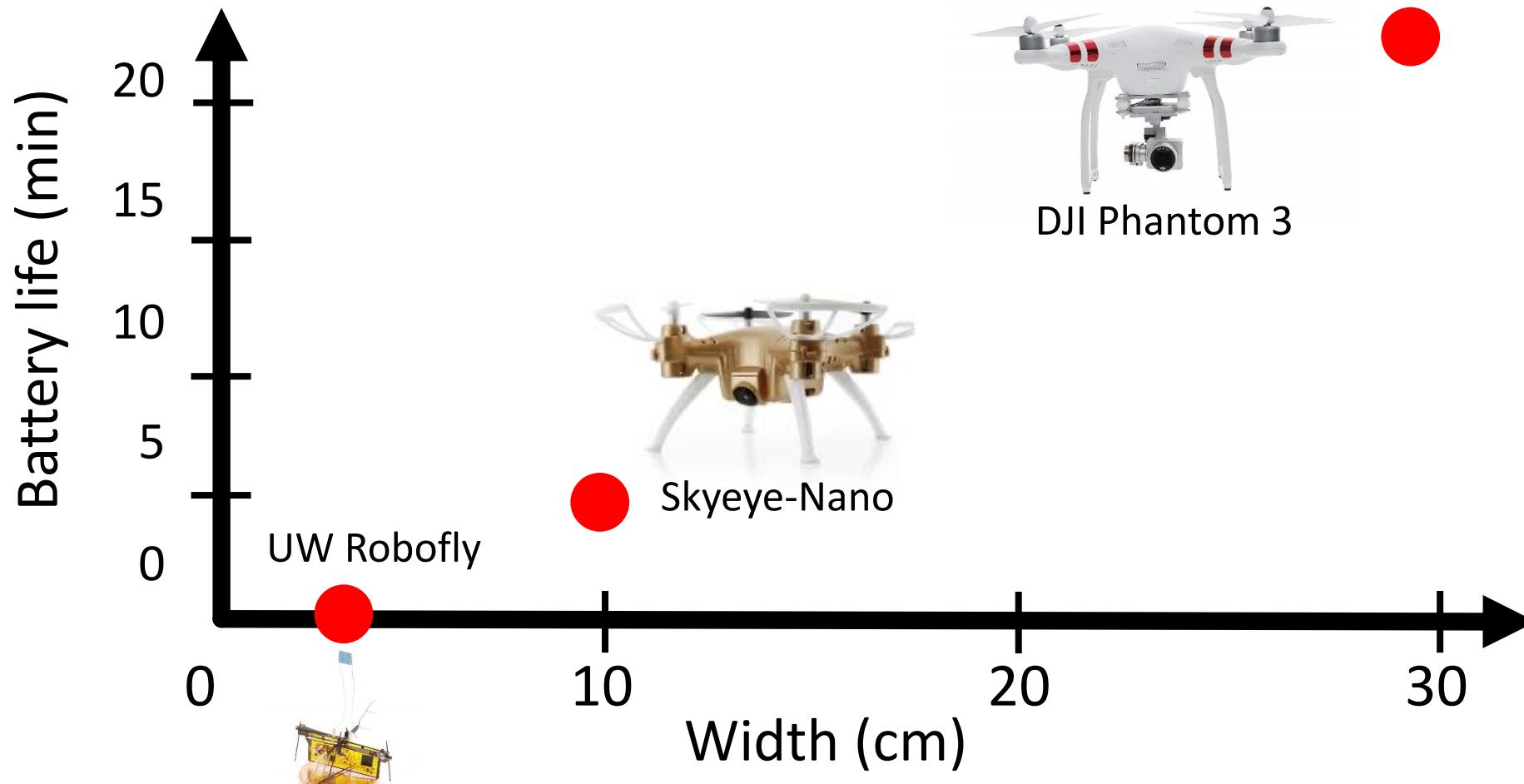


Drones consume lots of power



Need recharging every 10-20 min

Drones consume lots of power

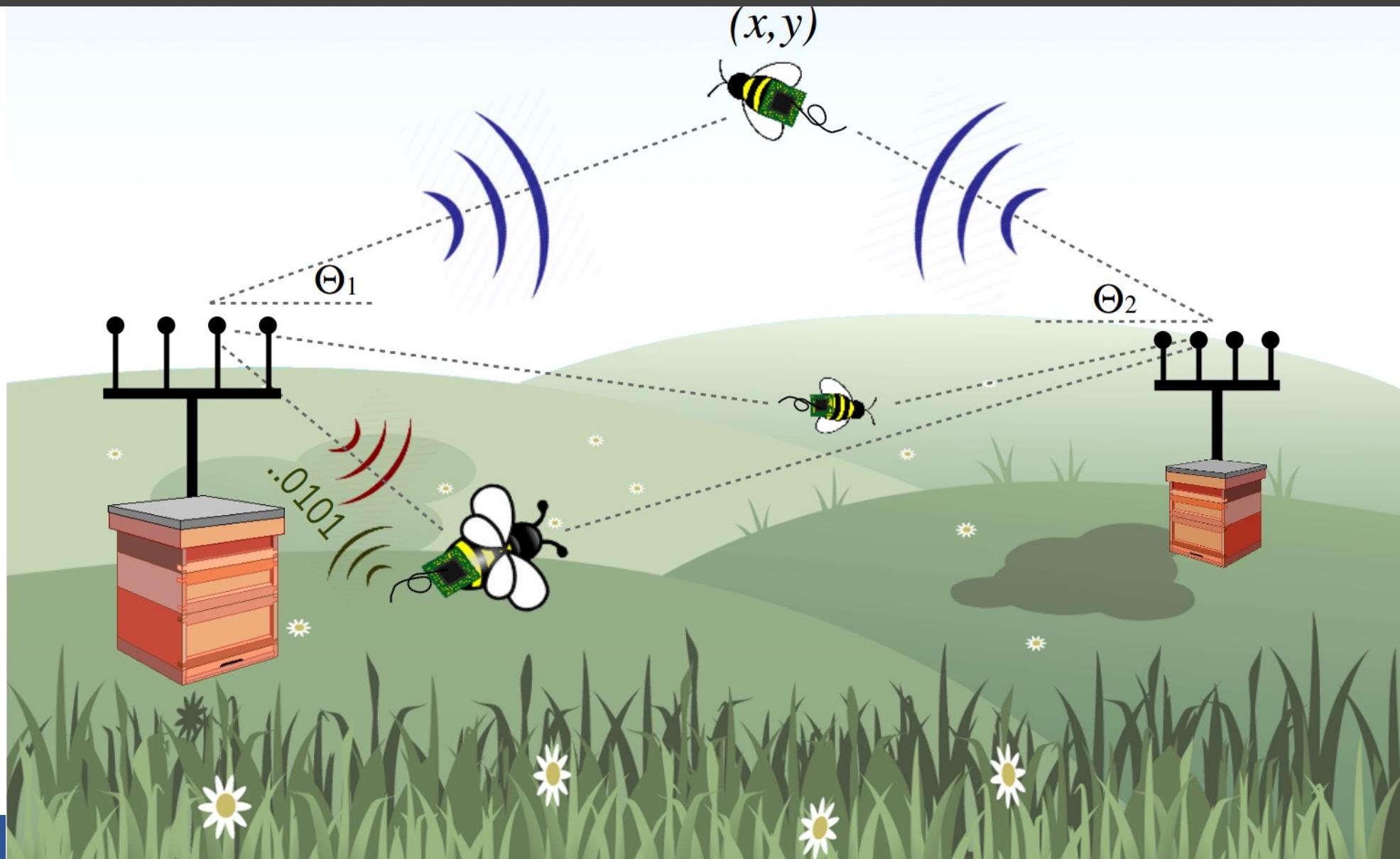


Fundamentally hard to get long battery life

Look to nature for solutions



Idea: Use live insects to carry wireless sensors



- ❑ Bio-based solution
- ❑ No need for mechanical propulsion
- ❑ Bees are introduced on farms for pollination
- ❑ Bees sense their environment



A close-up photograph of a bumblebee in flight. The bee's body is covered in a fuzzy, reddish-brown and black pattern. Along its back, there are several small, rectangular electronic components and wires attached, which are likely sensors used for scientific research. The background is a soft-focus green and blue, suggesting an outdoor environment.

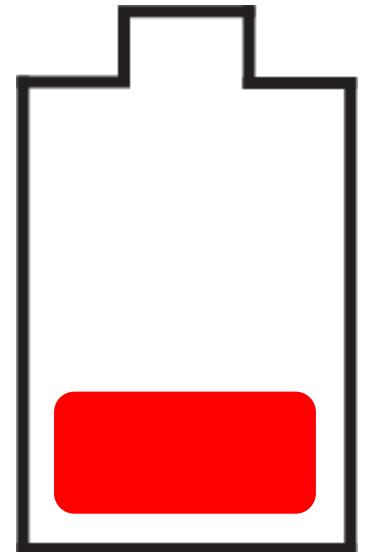
Live bumblebee carrying our sensors and electronics

Challenge 1: Small weight and low power

Size/Weight



Power



$\sim 100 \text{ mg} = 70 \text{ mg battery} + 30 \text{ mg electronics}$

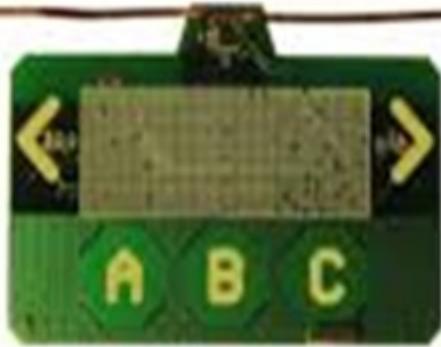
Challenge 2: Can't control bee motion



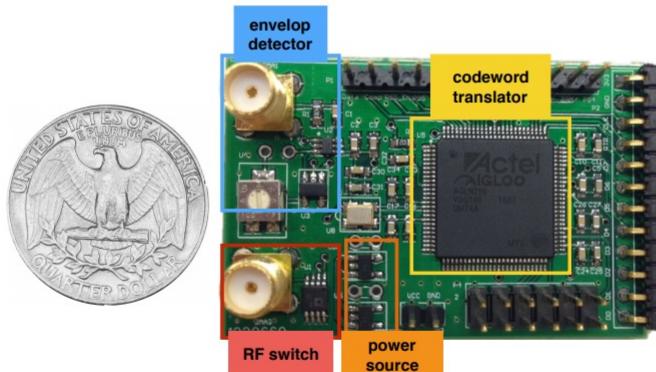
Our Solutions

- ❑ Building lightweight hardware
- ❑ 2D location tracking of bees

New problem: designing for small size and weight



Ambient backscatter



Freerider



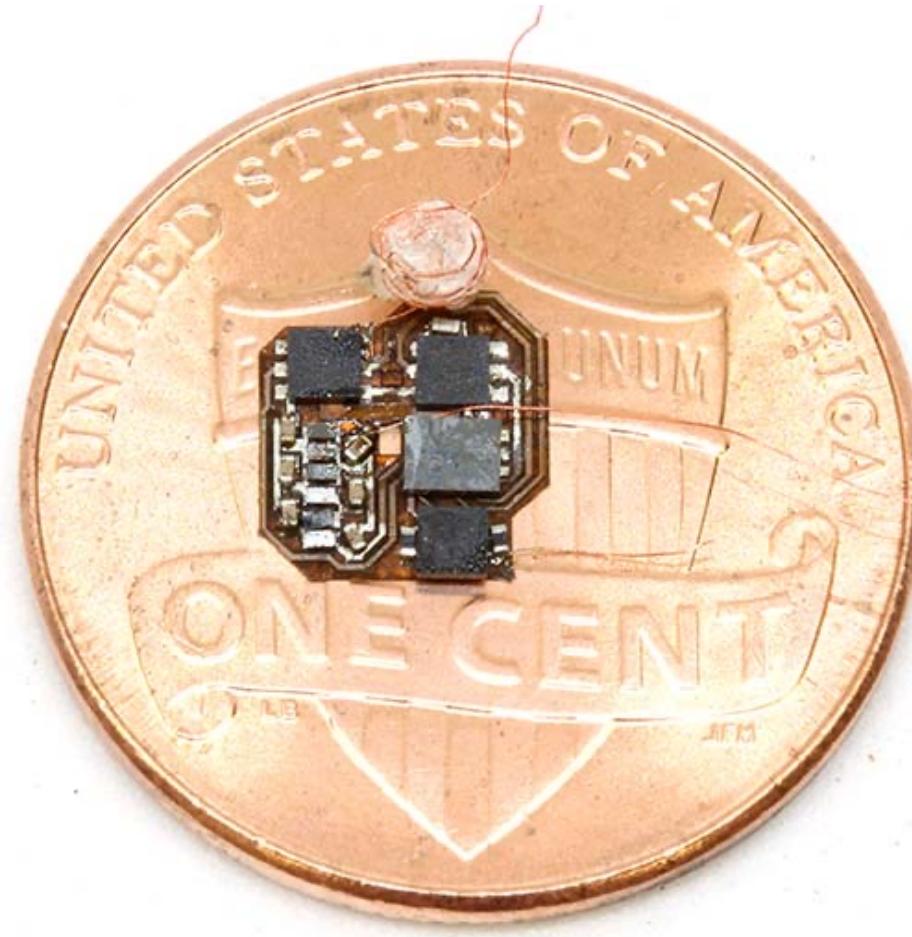
Custom ASIC

- + Lowest power and size
- Needs external components
- Long expensive process
- Not programmable

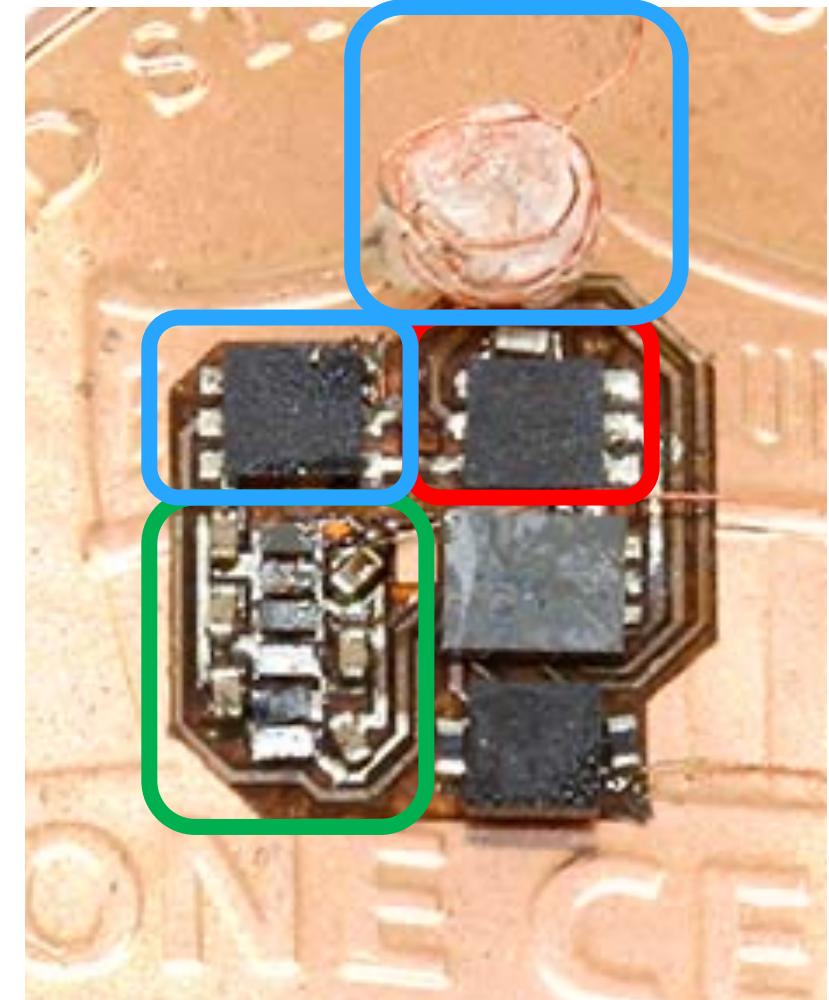
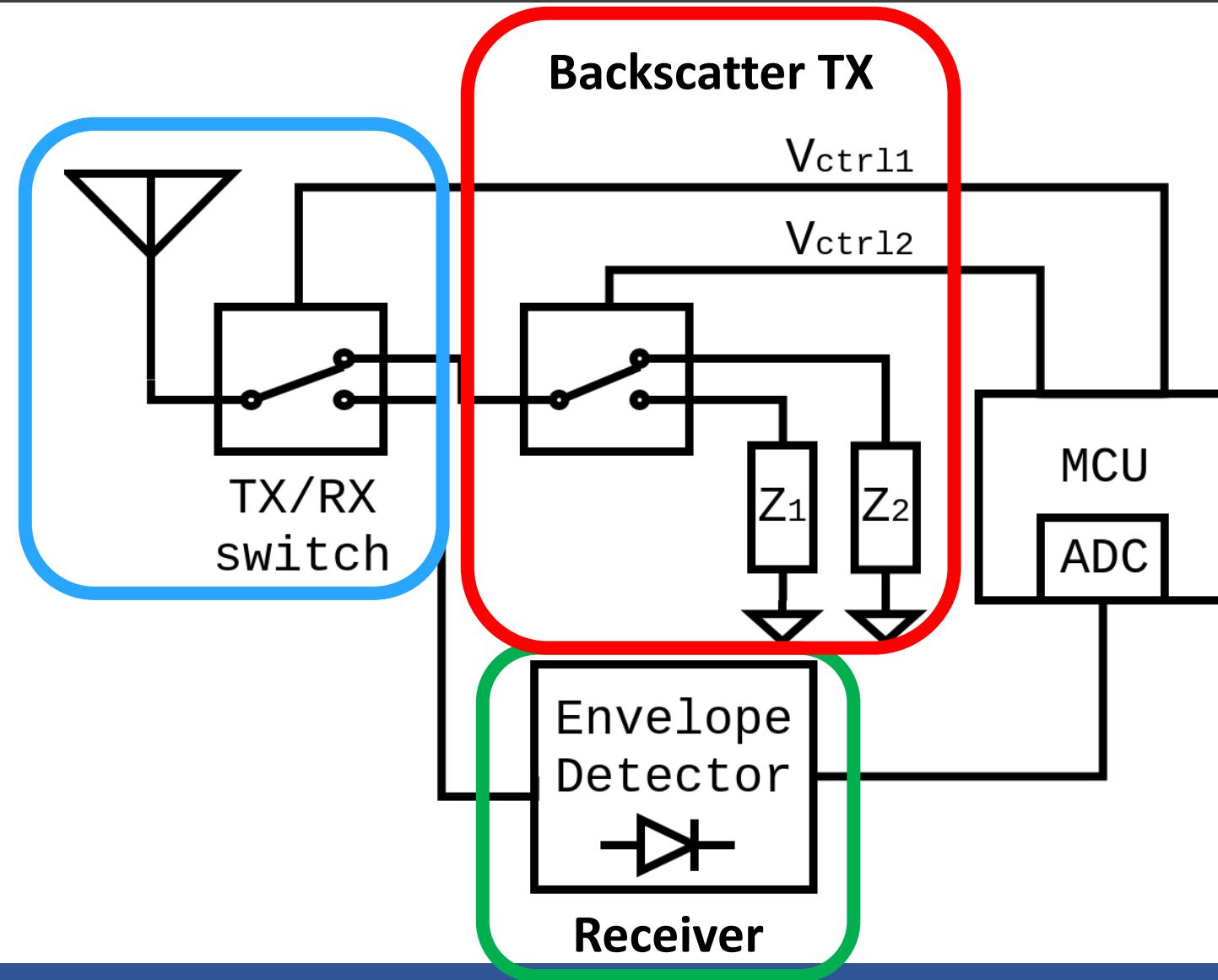
Need light-weight programmable solution

Programmable general purpose design

- Programmable microcontroller
- Interfaces with temperature and humidity sensors
- Low range backscatter communication
- Weighs < 30 mg

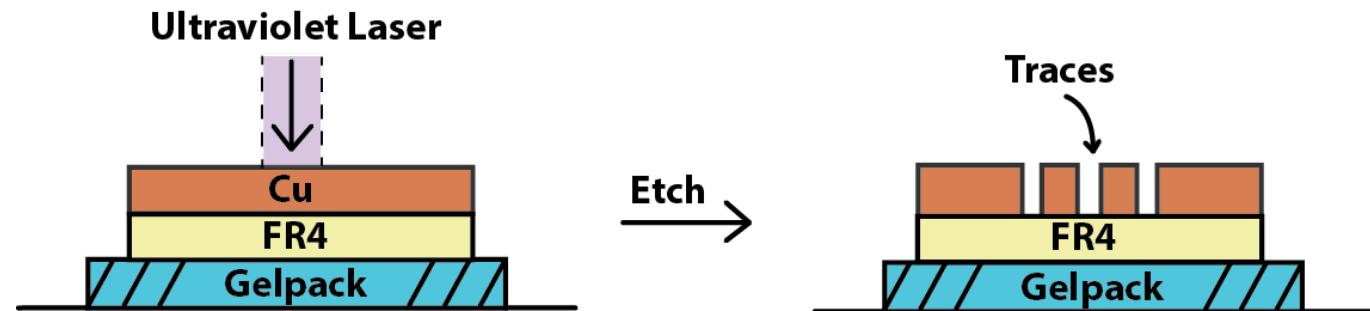


Programmable general-purpose design

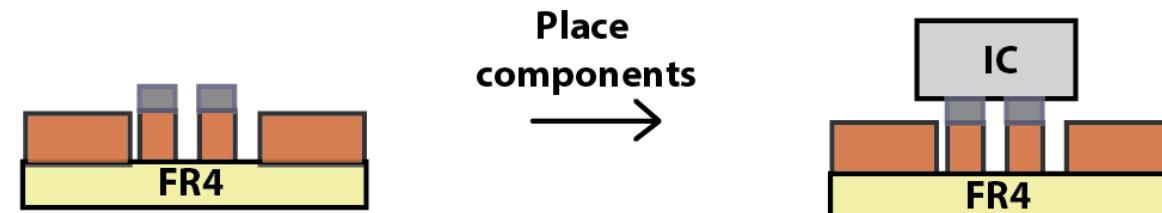


How do we fabricate it?

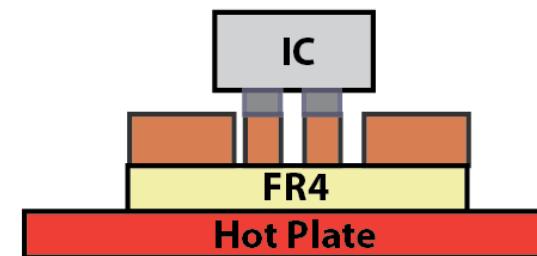
Step 1: Etch traces with laser



Step 2: Apply solder and place components



Step 3: Solder components on hot plate



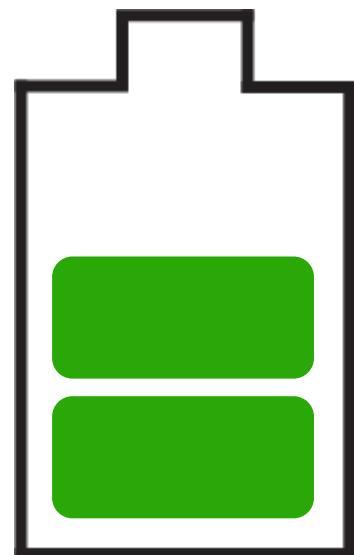
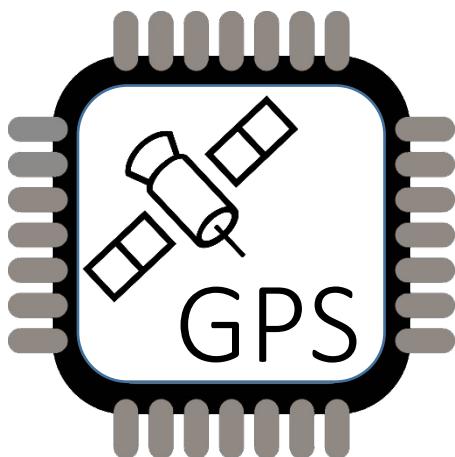
Our Solutions

- ❑ Building lightweight hardware

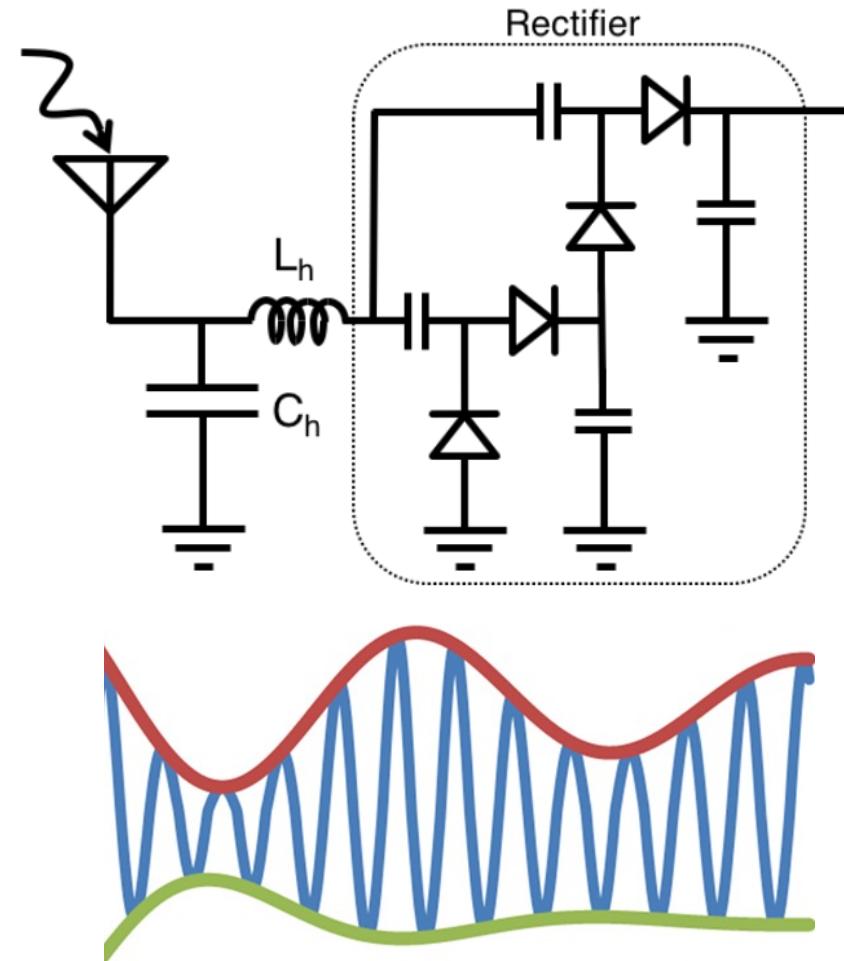
- ❑ 2D location tracking of bees

Self localization

Problem: GPS is power expensive

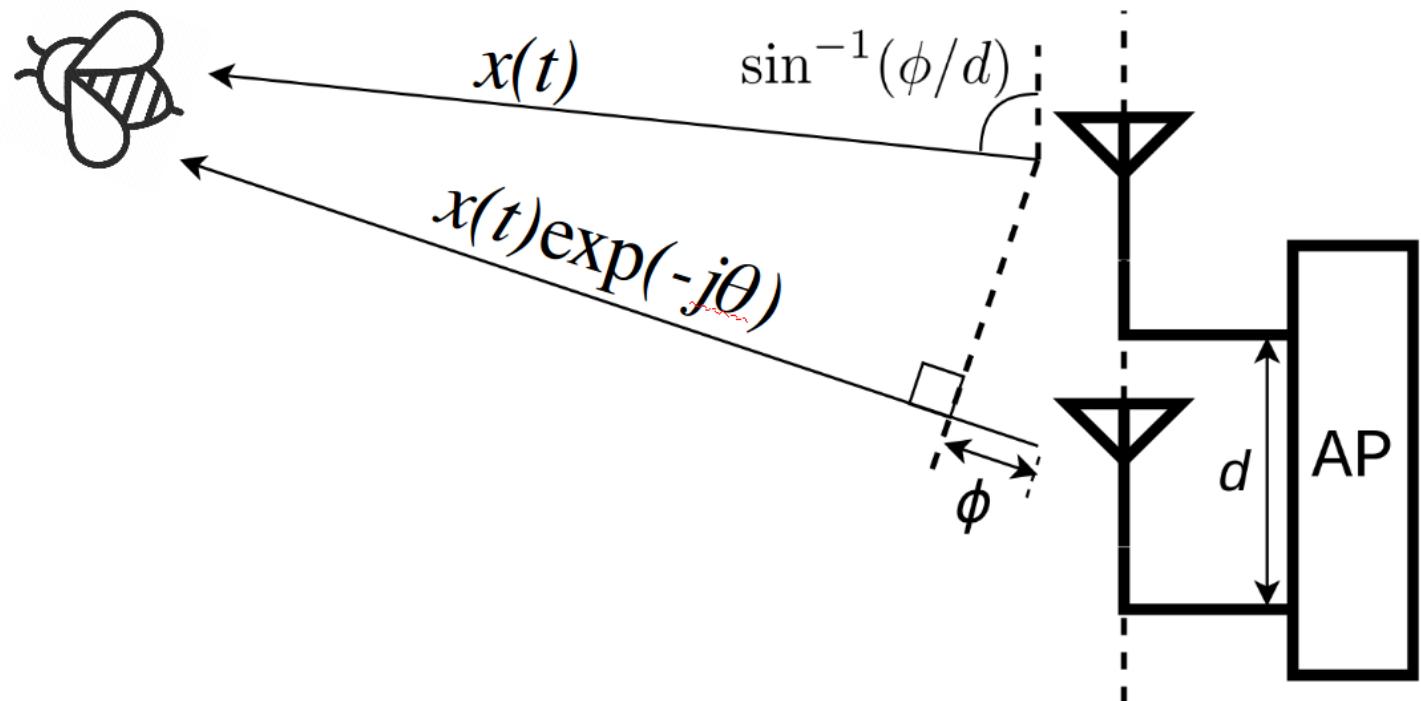
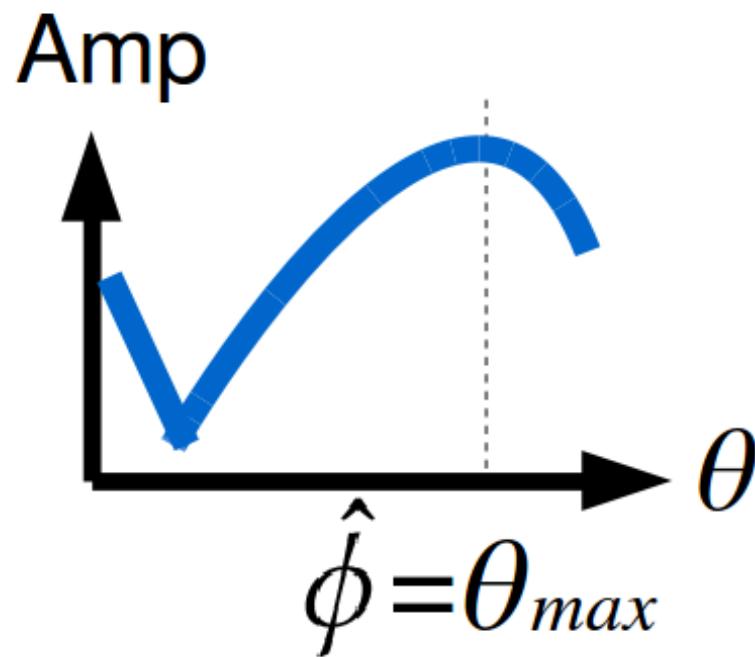


Solution: Passive receiver circuit



How can we get phase and compute
2D location with just amplitude?

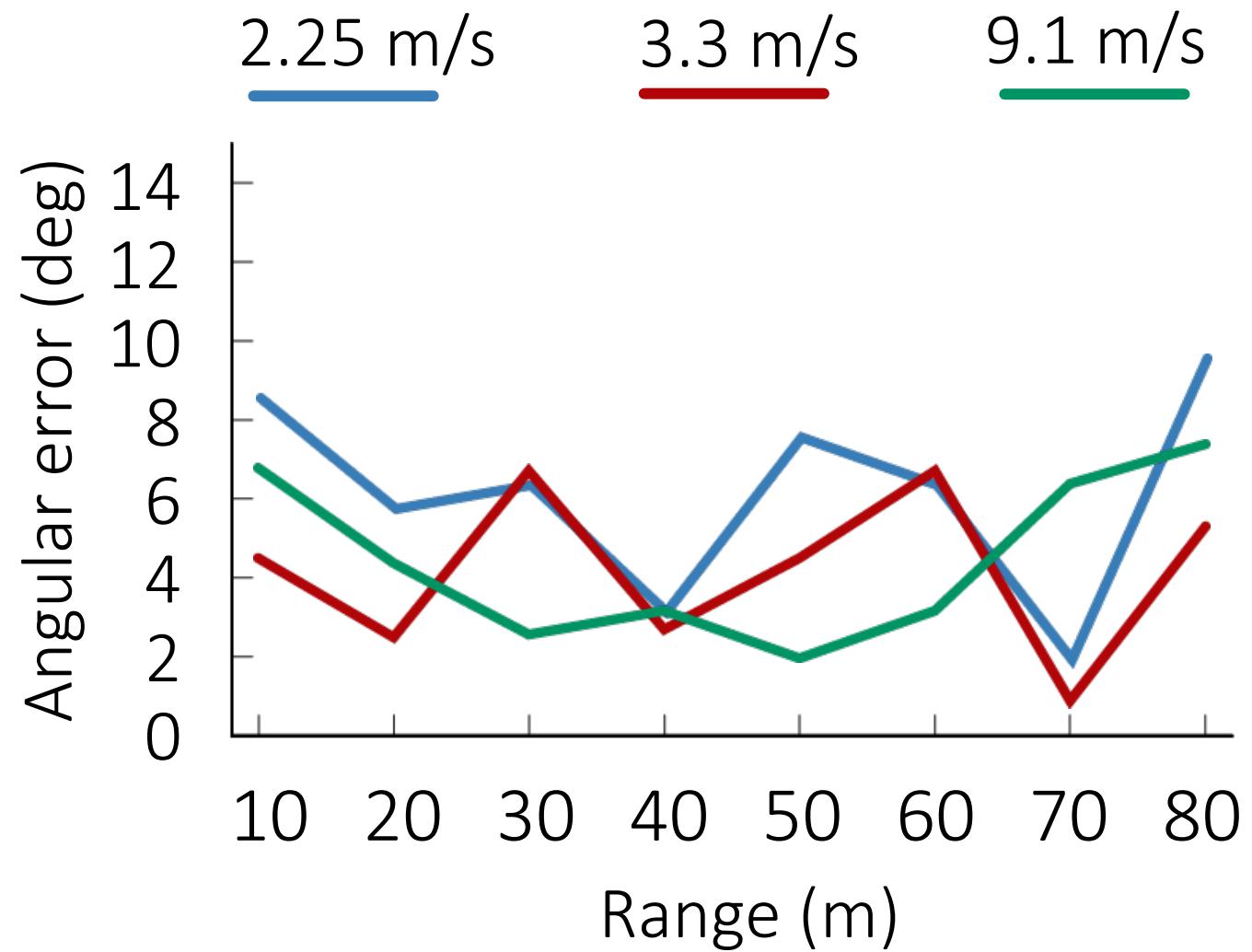
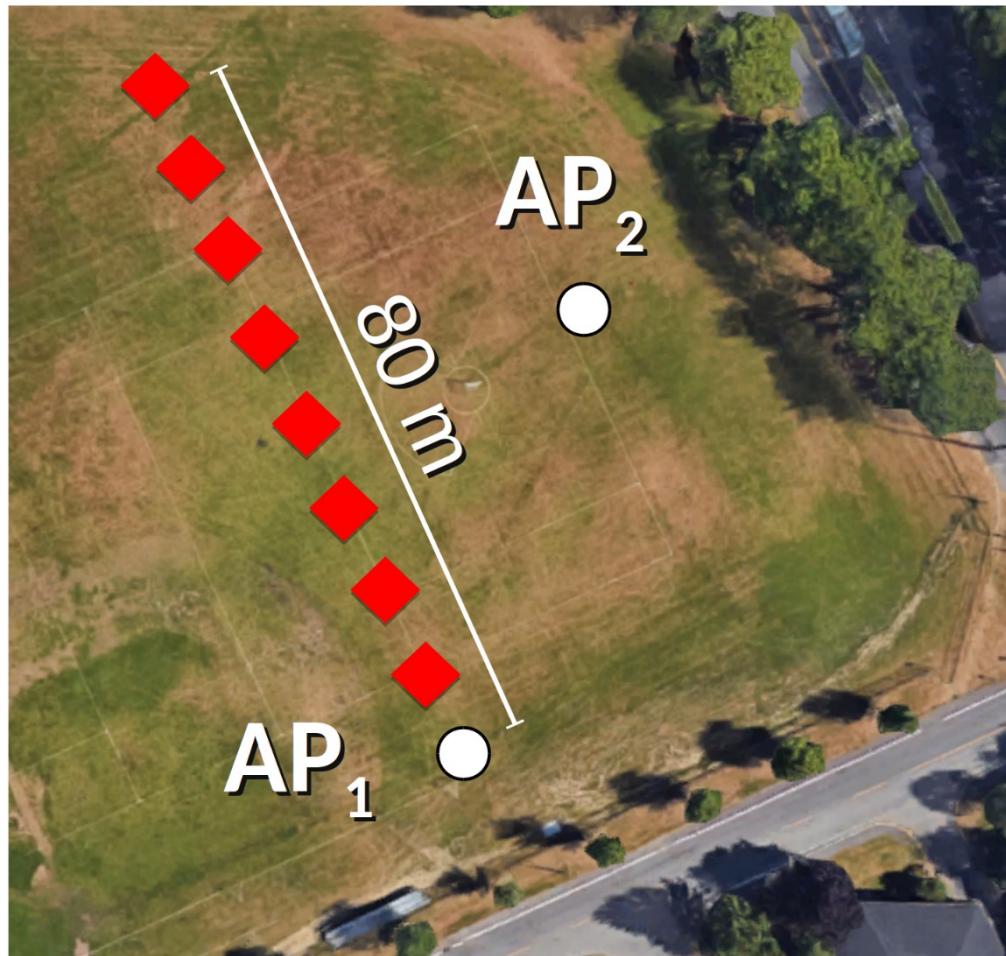
Extracting phase from amplitude



Addressing multipath:

1. On a farm line of sight path is strongest
2. Use multiple frequencies to reduce error

Evaluating localization accuracy



Evaluation



System benchmarks

- Stores up to 8000 location and humidity sensor data pairs
- Battery life of 7 hrs sampling every 4 s
- Harvester can recharge battery in a hive in 6 hrs

Contributions

- ❑ Insects can be used to carry general purpose sensors
- ❑ Novel low power self-localization technique
- ❑ Light-weight programmable platform for computing, communication and sensing

What's next?

- ❑ What other sensors can we use?
- ❑ Can we stream sensor data in real time?
- ❑ Can we use live insects to build bio-hybrid robots?
- ❑ Can we use these technologies to better study insects?



livingiot.cs.washington.edu

(3) Robust Ear-based BioSignal Sensing [IEEE TMC'21, ACM MobiSys'20, ACM MobiSys'19 Demo]

Nhat (Nick) Pham[‡], Tuan Dinh[§], Zohreh Raghebi^{††}, Taeho Kim[†], Nam Bui[†], Phuc Nguyen^{†★}, Hoang Truong[†], Farnoush Banaei-Kashani^{††}, Ann Halbower^{‡‡}, Thang Dinh^{§§}, and Tam Vu^{†‡}

[‡]*University of Oxford*, [†]*University of Colorado Boulder*, [§]*University of Wisconsin Madison*, ^{††}*University of Colorado Denver*, [★]*University of Texas at Arlington*, ^{‡‡}*Children's Hospital Colorado*, ^{§§}*Virginia Commonwealth University*



UNIVERSITY OF
OXFORD



University of Colorado
Boulder



WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON



University of Colorado
Denver



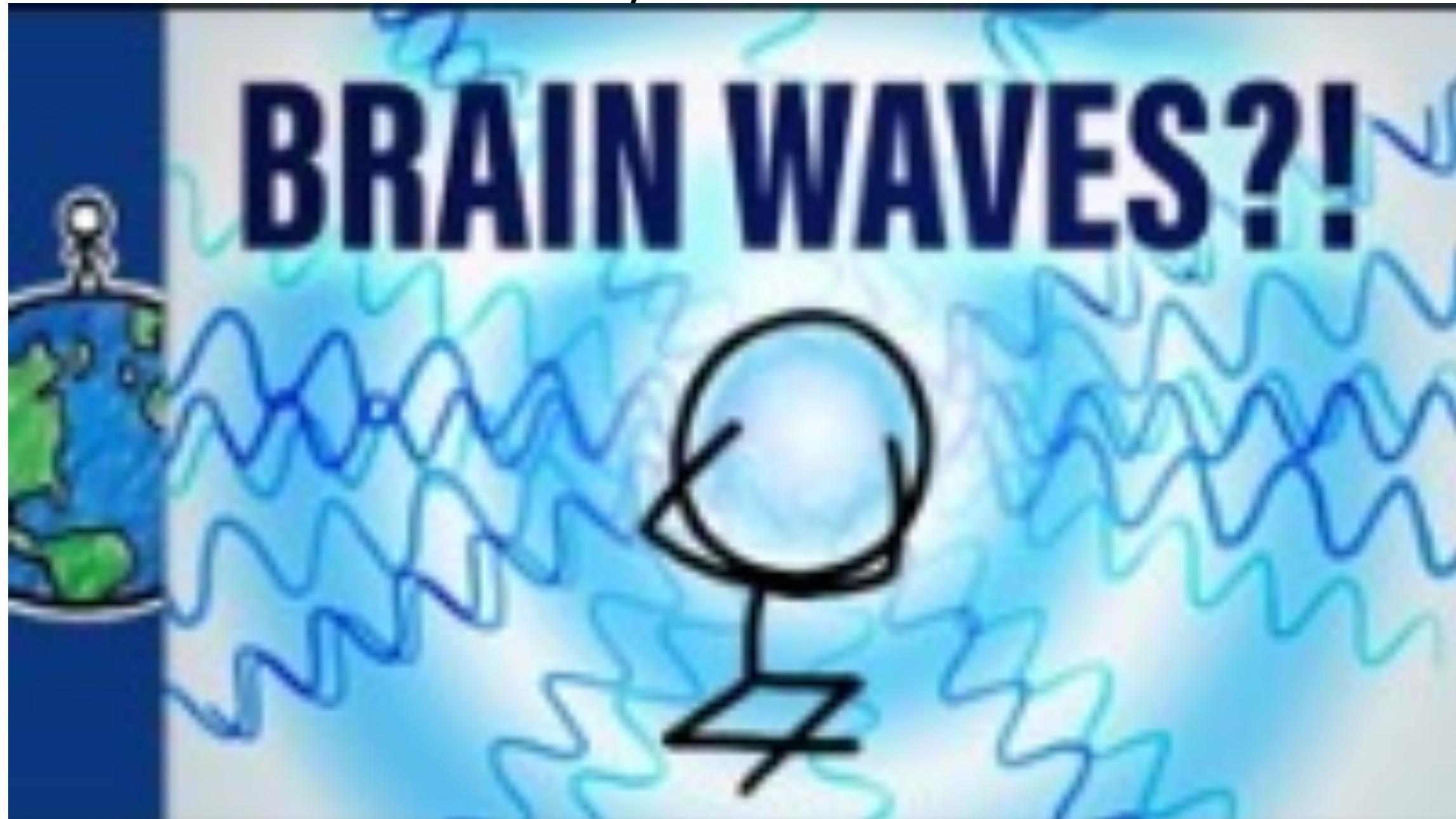
UNIVERSITY OF
TEXAS
ARLINGTON



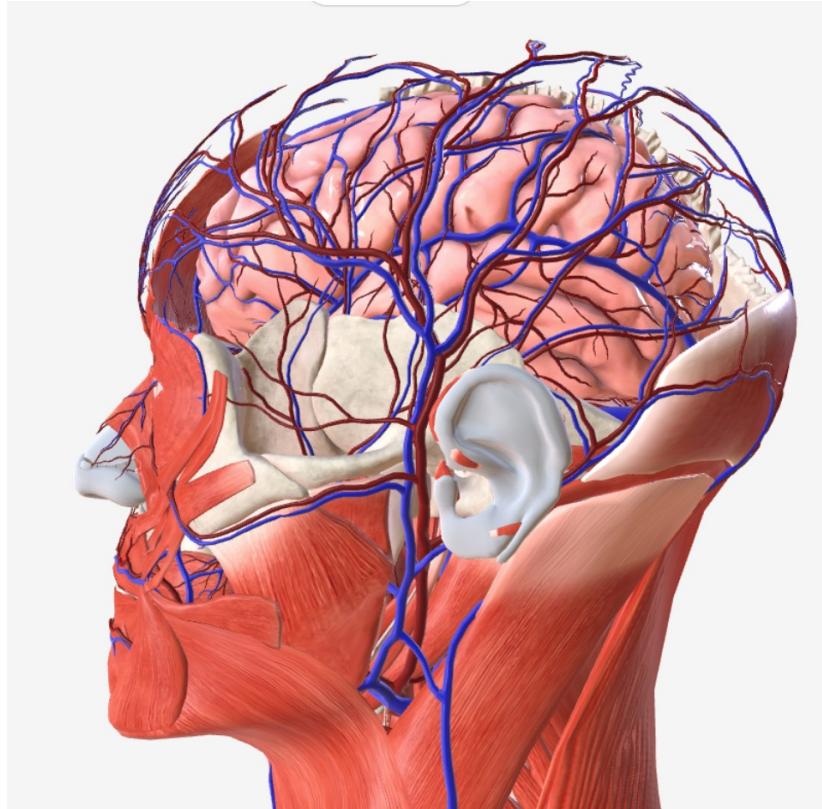
Children's Hospital Colorado



Electricity inside the brain?



Head-based biosignals



Brain (EEG)

Eyes (EOG)

**Facial muscles
(EMG)**

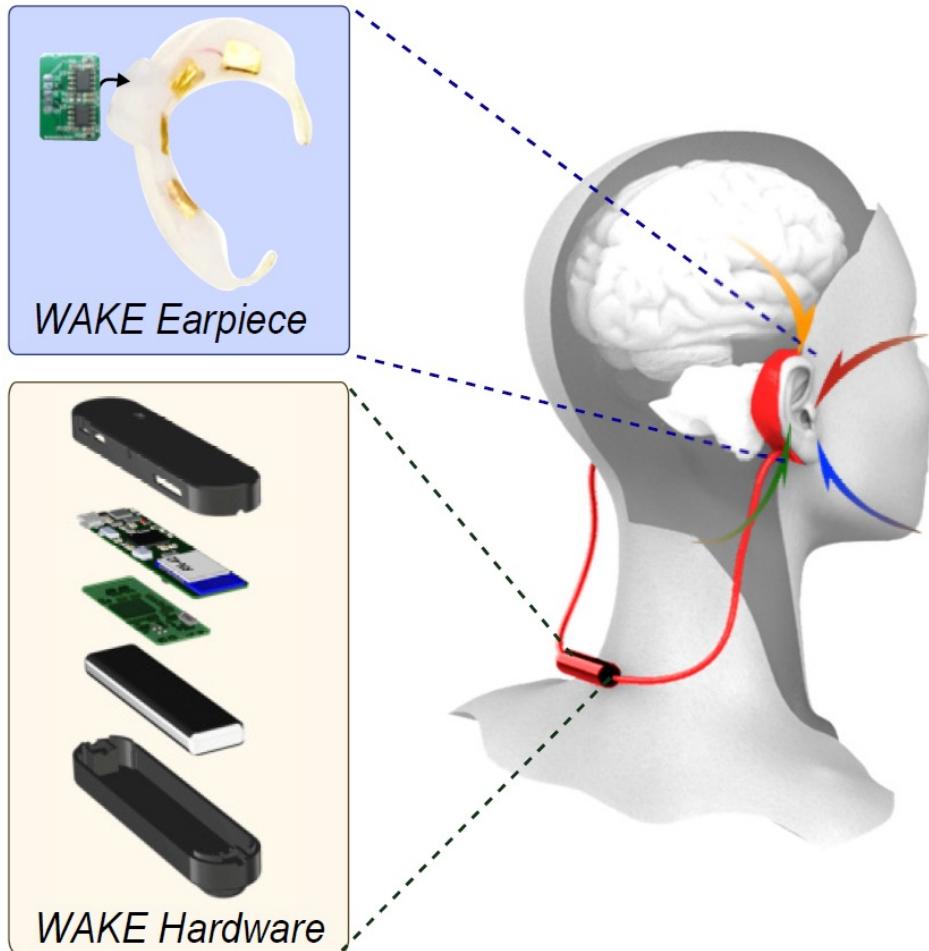
**Sweat glands
(EDA)**

**Blood vessels
(PPG)**

Current devices are cumbersome!



Our proposed Behind-the-ear wearable system



Able to capture various head-based biosignals

Compact, low cost, can be used daily

Socially acceptable



Sensing challenges

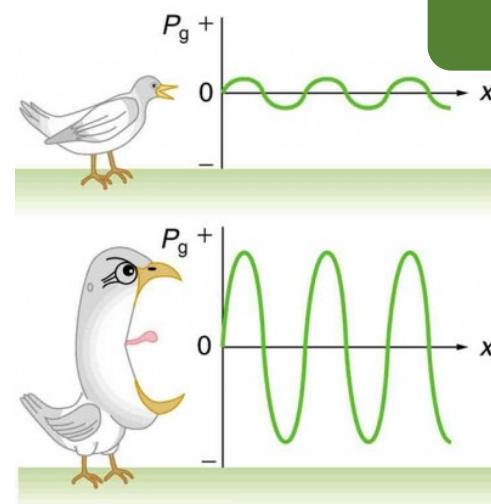


Heavy noise and motions



Unexplored sensing locations

Socially acceptable w/ minimal number of sensors



Overlap biosignals with significant (1000x) amplitude range

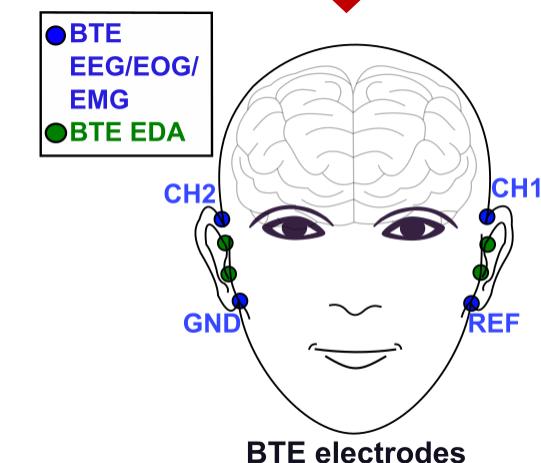
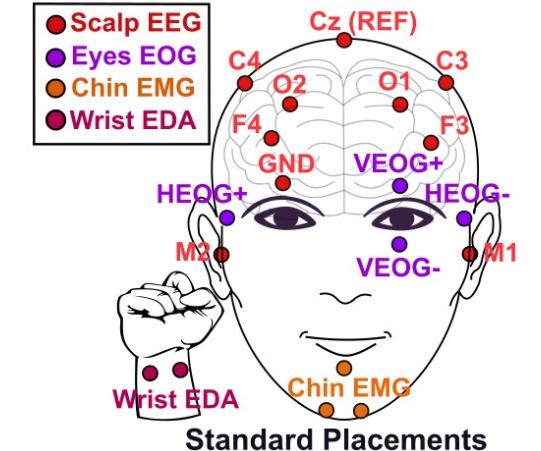
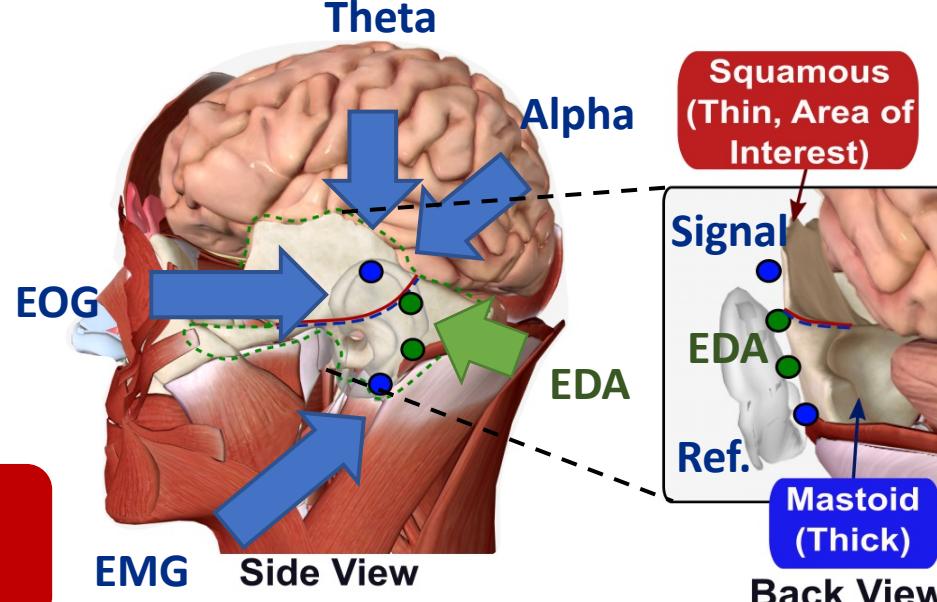
Challenge #1: Where to place the sensors? (1/2) ?



□ So that:

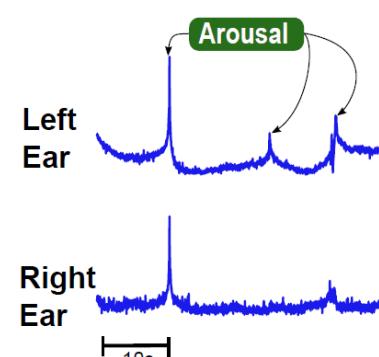
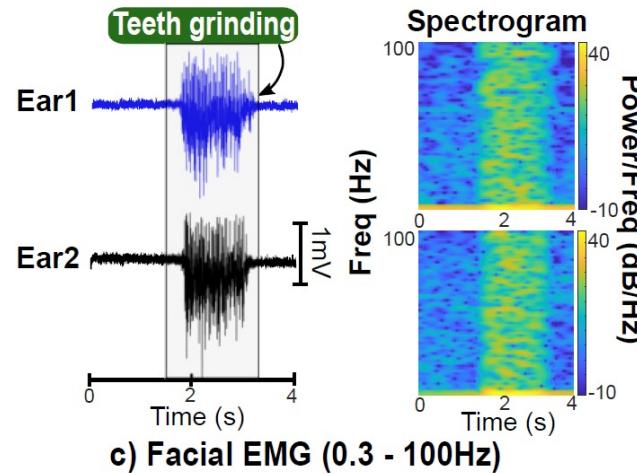
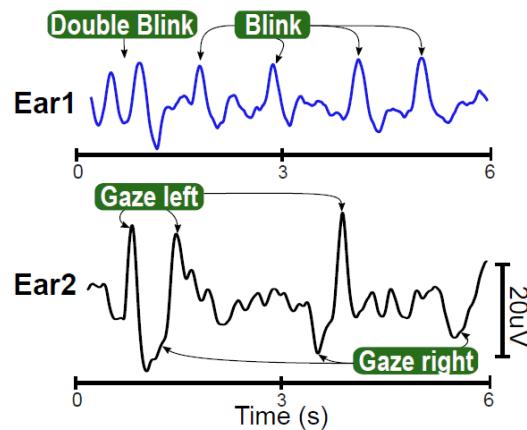
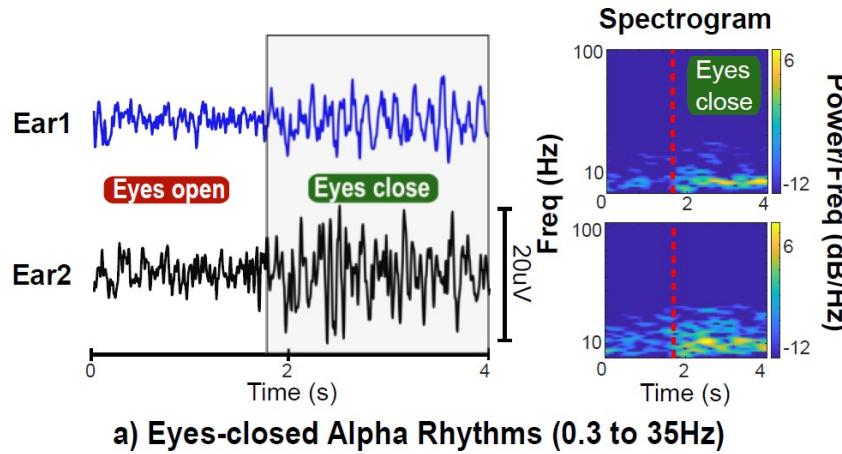
- **Wearability** and **sensing sensitivity** can be achieved.
- **Minimal number of sensors** is desirable.

The ear is the intersection of head-based biosignal!



Challenge #1: Where to place the sensors? (2/2)

Feasibility confirmation



Unique characteristics/challenges of the BTE signals?

- Low amplitude of BTE EEG/EOG. (i.e. <50uV vs. 100-500uV)
- Overlap frequency bands between BTE EEG/EOG and EMG with a significant amplitude difference (i.e. 1000x).

Challenge #2: Motion and environmental noise (1/3)



Motion and environmental noise is a long-standing challenge!

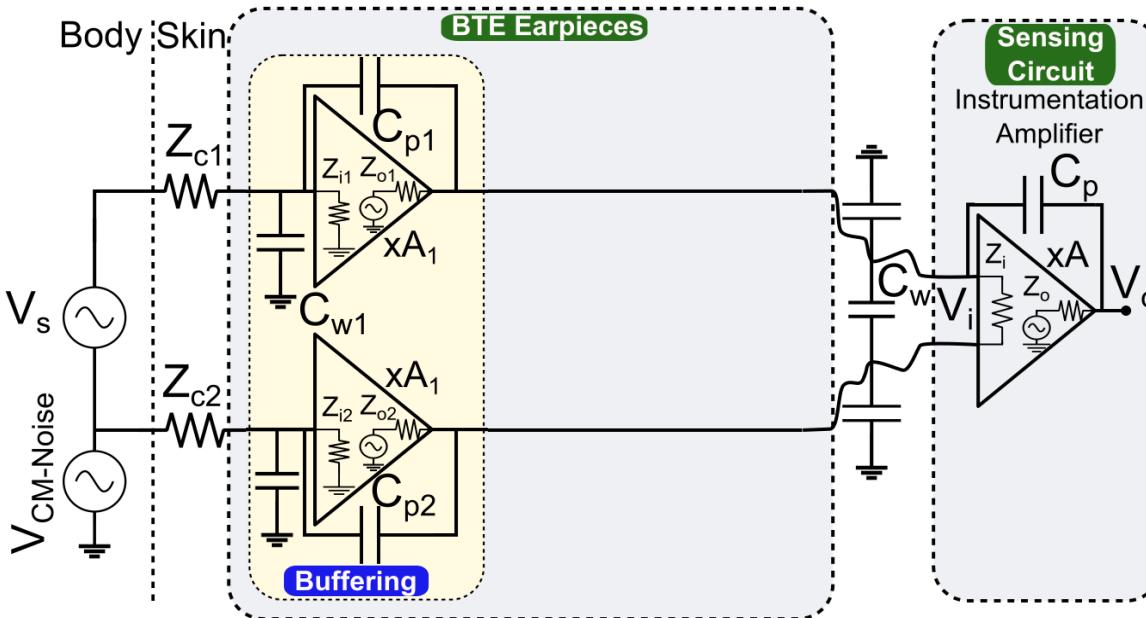
□ Motion artifacts:

- **Micro-motions** of the sensing electrodes.
- **Fluctuation (i.e. microphonic triboelectric effect)** of the signal wires.

□ Environmental noise:

- Noise coupled through the **human body** and **signal wires**.
- Noise characteristic **varies across environments**.

Challenge #2: Motion and environmental noise (2/3) Three-folds cascaded amplifying (3CA) – Motion artifacts



□ Electrical model:

$$V_o = G * V_s = \frac{A * V_s}{1 + (\mathbf{Z}_{c1} + \mathbf{Z}_{c2})(\frac{1}{R_i} + jw(\mathbf{C}_w + \mathbf{C}_i + (A - 1)\mathbf{C}_p))} (*)$$

- Movement of the wires => changes in \mathbf{C}_w
- Micro-motion of the electrode => changes in $\mathbf{Z}_{c1}, \mathbf{Z}_{c2}$
=> Fluctuations of the output signal.

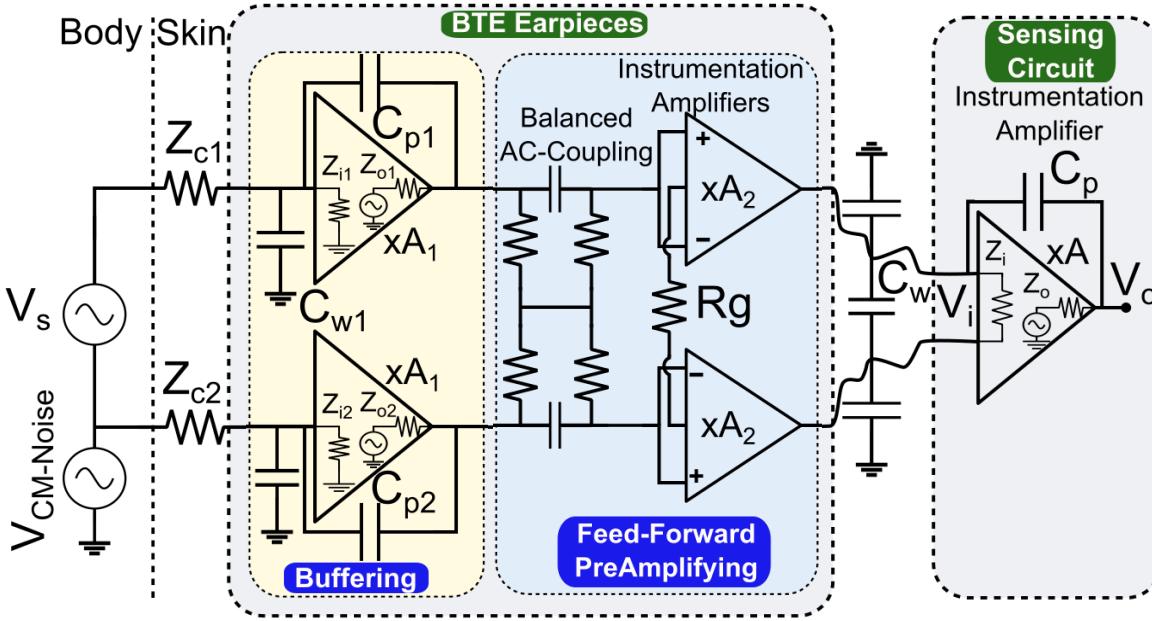
□ Introduce Stage 1 - Unity-gain amplifying:

- Z transformation: transform $\mathbf{Z}_{c1}, \mathbf{Z}_{c2}$ in (*) to $\mathbf{Z}_{o1}, \mathbf{Z}_{o2}$ (~ 0) => eliminate the effect of \mathbf{C}_w .
- Minimizing effect of \mathbf{Z}_{c1} changes: Minimize γ by using $A=1$, maximizing R_{i1} , minimizing \mathbf{C}_{i1} , \mathbf{C}_{w1} .

$$V_o = \frac{A_1 * V_s}{1 + \mathbf{Z}_{c1}(\frac{1}{R_{i1}} + jw(\mathbf{C}_{w1} + \mathbf{C}_{i1} + (A - 1)\mathbf{C}_{p1}))} = \frac{A_1 * V_s}{1 + \mathbf{Z}_{c1}\gamma}$$

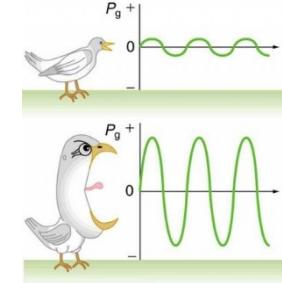
Challenge #2: Motion and environmental noise (3/3)

Three-folds cascaded amplifying (3CA) – Environmental noise

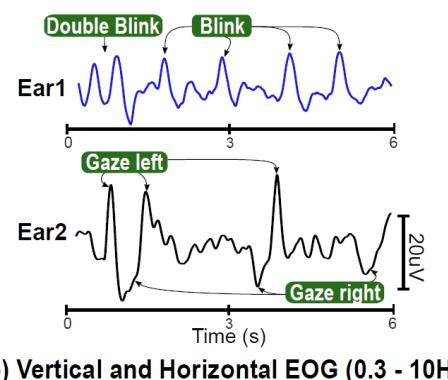
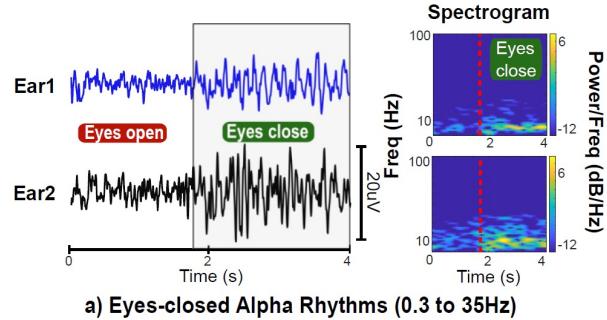


□ Introduce Stage 2 - Feed Forward Differential PreAmplifying (F2DP):

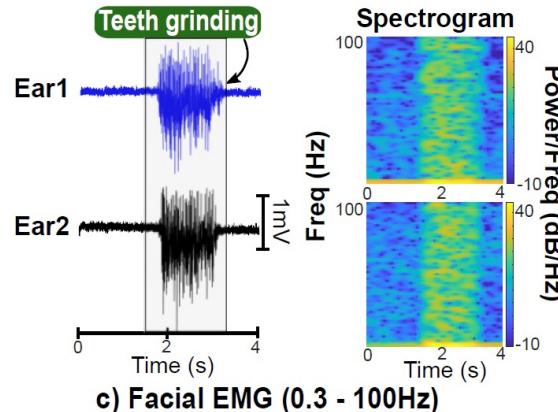
- **2 separate amplifying stages** minimize the effect of motion due to contact impedance.
- **Feed-Forward Differential Amplifying** technique with dual instrumentation amplifiers:
 - Enhance Common-mode rejection ratio (CMRR).
 - Produce **amplified, fully differential** signals => robust again environmental noises.
- **Balanced AC-coupling** topology: efficiently remove DC component while mitigating component mismatches issue.



Challenge #3: Overlap signal with a significant range (1/2)



VS.



- **Using a fixed gain is not efficient!**
 - High gain => saturate BTE EMG signal.
 - Low gain => increase noise floor for weak BTE EEG/EOG signals.

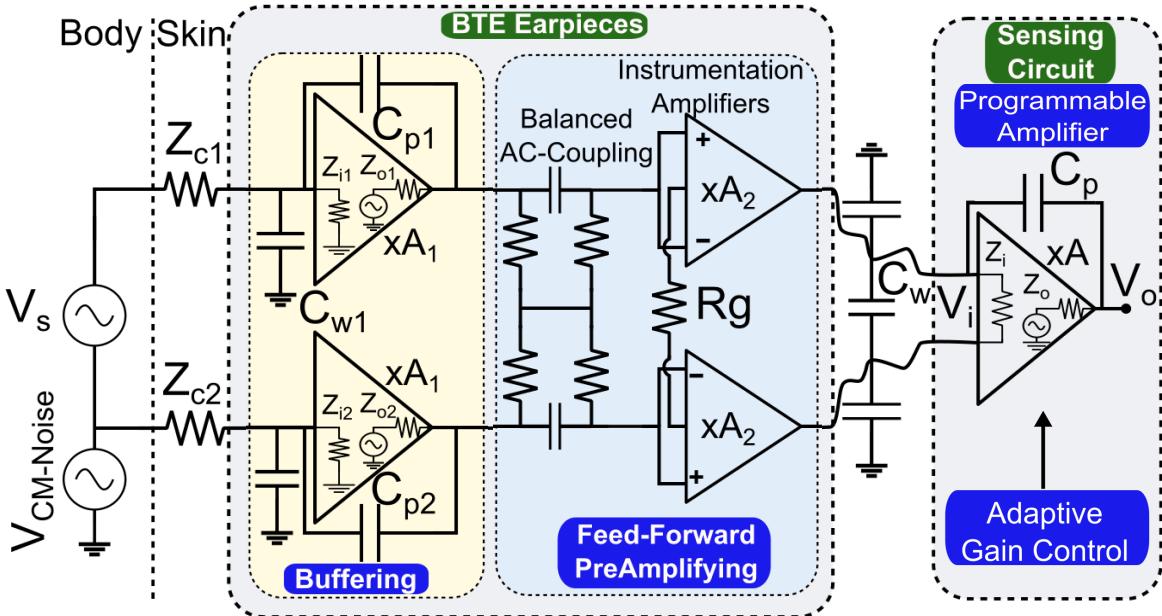
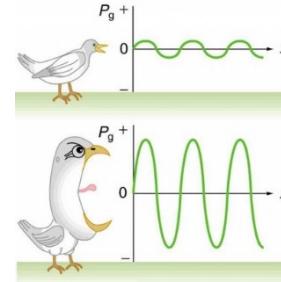
=> The amplifier gain needs to be changed on-the-fly.

- **Observations on BTE signal patterns:**
 - Strong EMG events don't happen frequently.
 - EMG events can happen abruptly.
 - EMG signal is stochastic and can vary significantly.

BTE EEG/EOG is overlap with EMG in a three-orders magnitude range!

Challenge #3: Overlap signal with a significant range (2/2)

Adaptive Amplifying and Adaptive Gain Control



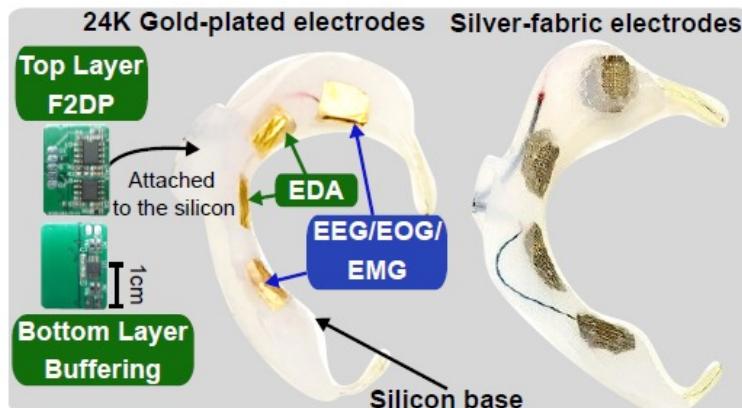
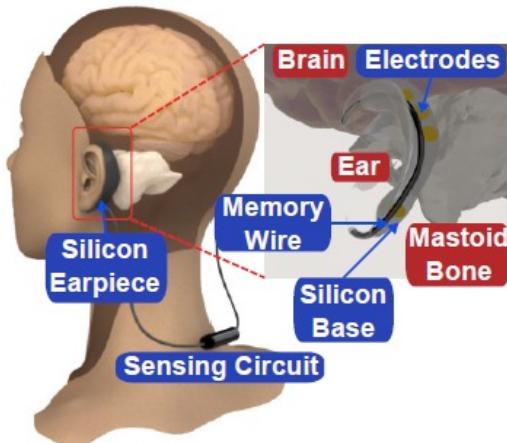
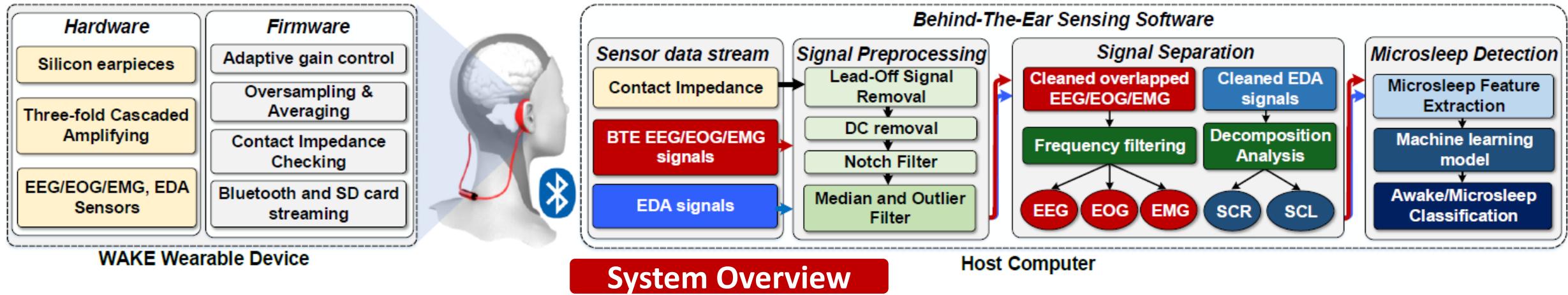
□ Introduce Stage 3 – Adaptive Amplifying with an Adaptive Gain Control algorithm:

- Initially, **keep the gain at maximum** for BTE EEG/EOG signal.
- **React quickly to abrupt increases** from the initial state => capture an EMG event quickly.
- **React slowly to abrupt decreases** while an EMG event is happening => avoid gain oscillation.

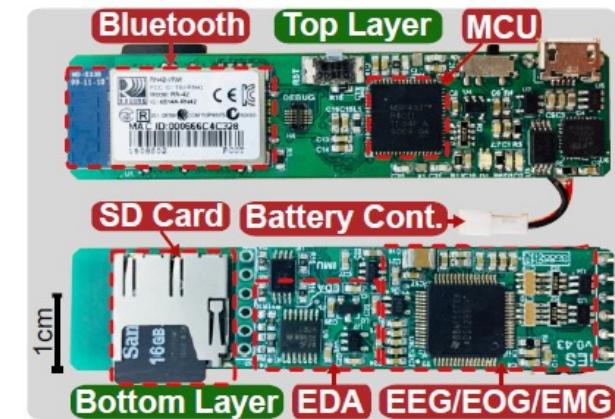
□ Square Law Detector vs. Peak Envelope Detector:

- Both can be used for AGC.
- PED with dynamic windows is used because of low complexity.

Implementation

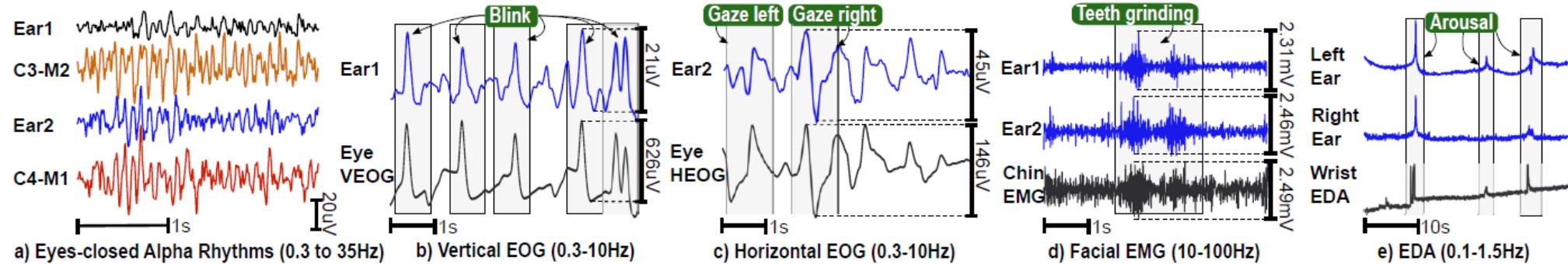


Earpieces design



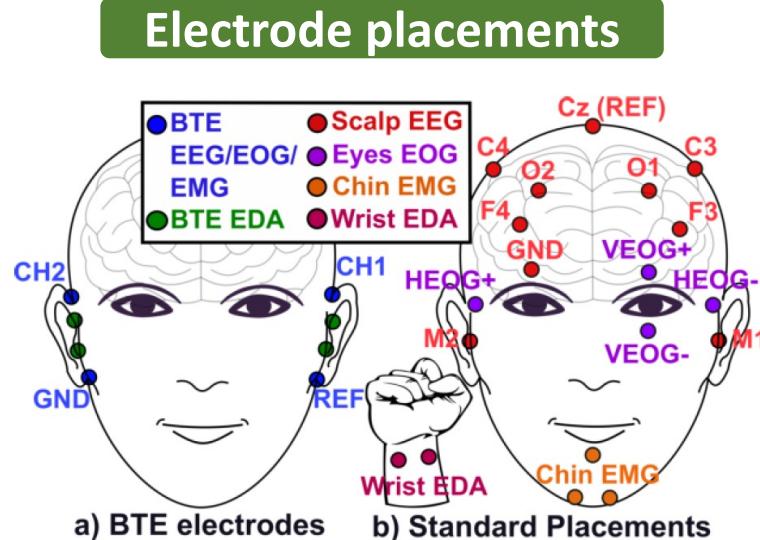
Sensing device

Evaluation #1 – Signal Sensitivity Validation



BTE vs. Ground-truth signals

Normalized Cross Correlation

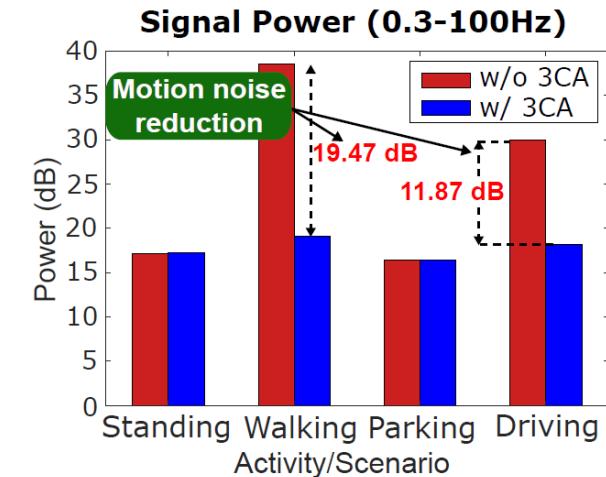
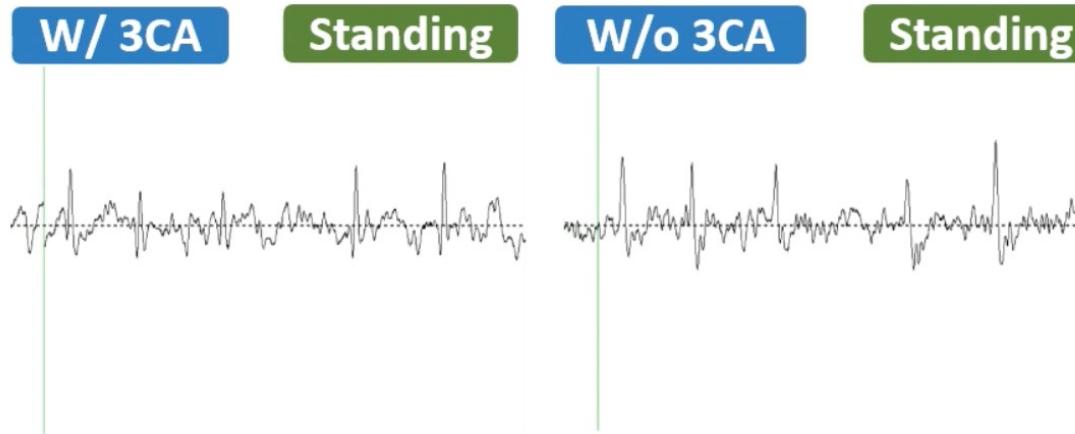


	Ear1	Ear2
C3-M2	0.35 (moderate)	
C4-M1		0.44 (moderate)
O1-M2	0.28 (weak)	
O2-M1		0.52 (moderate)
VEOG	0.47 (moderate)	
HEOG		0.59 (strong)
Chin EMG	0.62 (strong)	0.76 (strong)
Left Wrist EDA	0.37 (moderate)	

Evaluation #2 – Motion and Environmental Noise Suppression

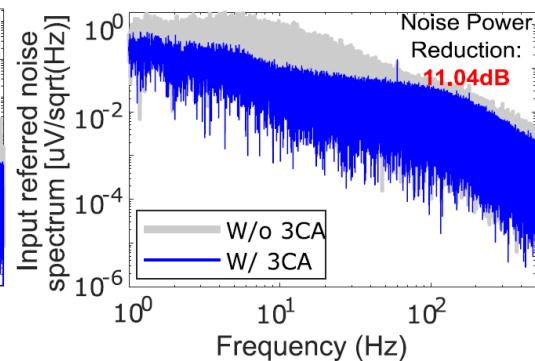
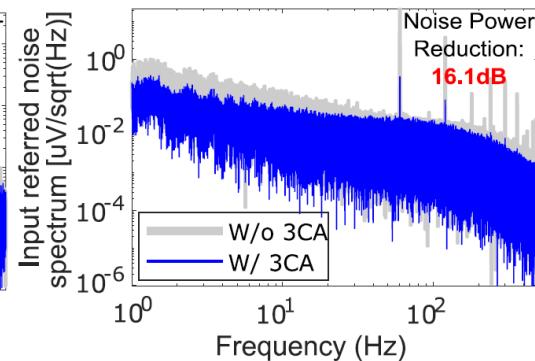
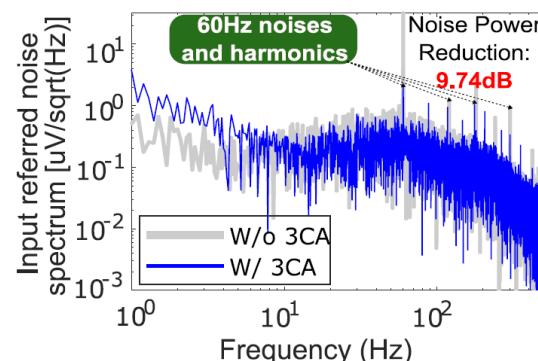
❑ Motion artifact evaluations:

- Standing vs. Walking.
- Parking (w/ a running engine) vs. Driving.
- Durations: 40-60 minutes



❑ Environmental noise evaluations:

- 3 different environments: Office, Residential area, and Inside a car.
- Durations: 60 minutes

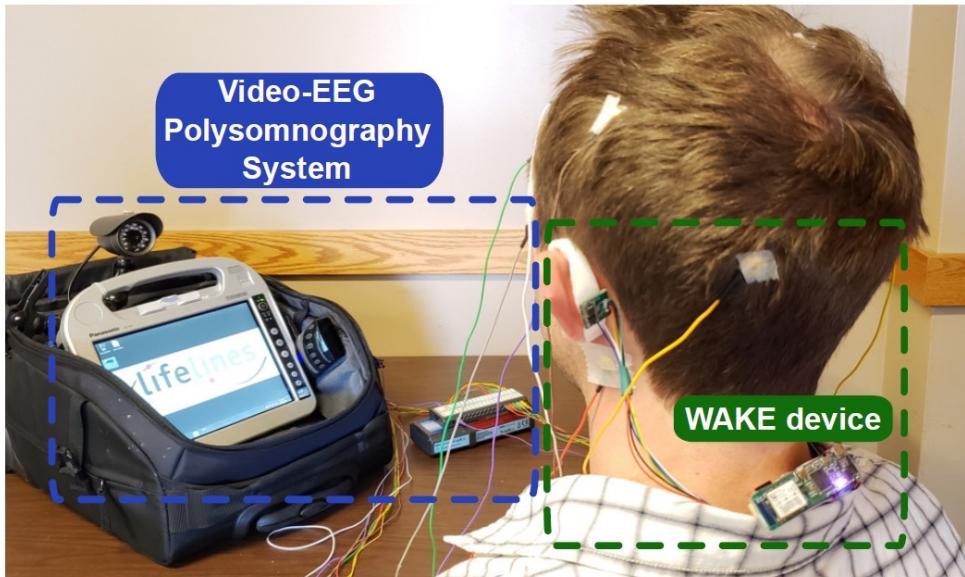


Office

Residential Area

Inside a Car

Evaluation #3 – Microsleep Detection Performance



Experiment Setup

□ Demographic:

- 19 subjects.
- Healthy: 9, Sleep deprivation: 9, Narcolepsy: 1.
- Experiment duration: maximum 2h.
- **Ground-truth:** Video-PSG system with 2 sleep experts.

□ Classification model:

- 35,558 awake and 8,845 microsleep data points.
- Epoch size: 5s, 80% overlap (i.e. slide every 1s).
- Durations: maximum 2 hours/each subject.
- **Hybrid model of a hierarchical classifier** (Random Forest, Adaboost, SVM) and **EMG-event-based heuristic rule**.

Classification Performance

Experiment	Precision	Sensitivity	Specificity
Leave-one-subject-out (Inter-subject)	0.76	0.85	0.85
Test-set (75%/25%) (Intra-subject)	0.87	0.9	0.96
Leave-one-sample-out (Intra-subject)	0.88	0.89	0.96

Key takeaways

- To get good application results, the **input data/signal needs to be clean**. We should not depend on machine learning to do all the work.
- **Software solutions might not always able to reliably solve the challenges** in Cyber-Physical systems
- Low-level hardware solutions are sometimes needed to solve the root cause.
- **Data collections on human can take time** (and can be quite messy).

(4) Efficient computing on wearables (ACM MobiCom'22)

Nhat (Nick) Pham¹, Hong Jia², Minh Tran¹, Tuan Dinh³, Nam Bui⁴, Young Kwon², Dong Ma⁵, Phuc Nguyen⁶, Cecilia Mascolo², and Tam Vu⁴

¹*University of Oxford*, ²*University of Cambridge*, ³*University of Wisconsin Madison*, ⁴*University of Colorado Boulder*,
⁵*Csingapore Management University*, ⁶*University of Texas at Arlington*.



UNIVERSITY OF
OXFORD



UNIVERSITY OF
CAMBRIDGE



University of Colorado
Boulder



UNIVERSITY OF WISCONSIN-MADISON



UNIVERSITY OF
TEXAS
ARLINGTON



SINGAPORE
MANAGEMENT
UNIVERSITY

Motivation

What can we enable w/ 40X high-fidelity biosignal compression on wearables?



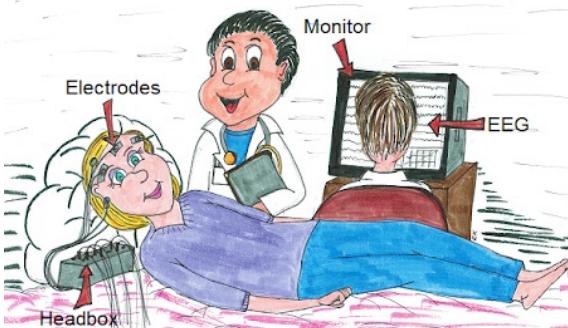
Prevent fatalities!



Enhance usability!

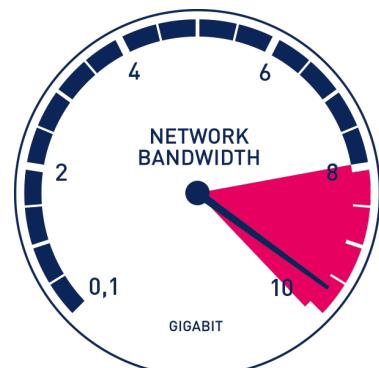
The problem

Many healthcare biosignal-based applications are not practical for wearables!



Long-term, high-fidelity monitoring

Weeks/Days not hours!



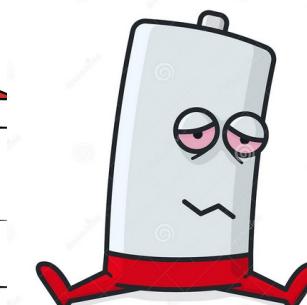
High data rate

Communication is energy heavy!



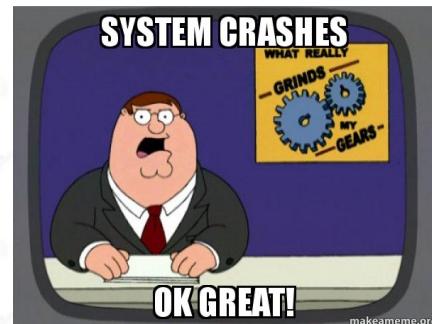
Low-latency response

Slow responses could be fatal!



Small battery, resource-constrained devices

MHz CPU, KB Memory!



makeameme.org

The conventional trade-off

Signal Quality



VS.

Battery Life



Medical Biosignal Monitoring

Brain/Eyes/Muscle: 10-20x1024Hz.

Battery life: >24h, w/ a big bag of batteries

Need constant monitoring by technicians.

Embrace2 Wearable

Battery life: 48h

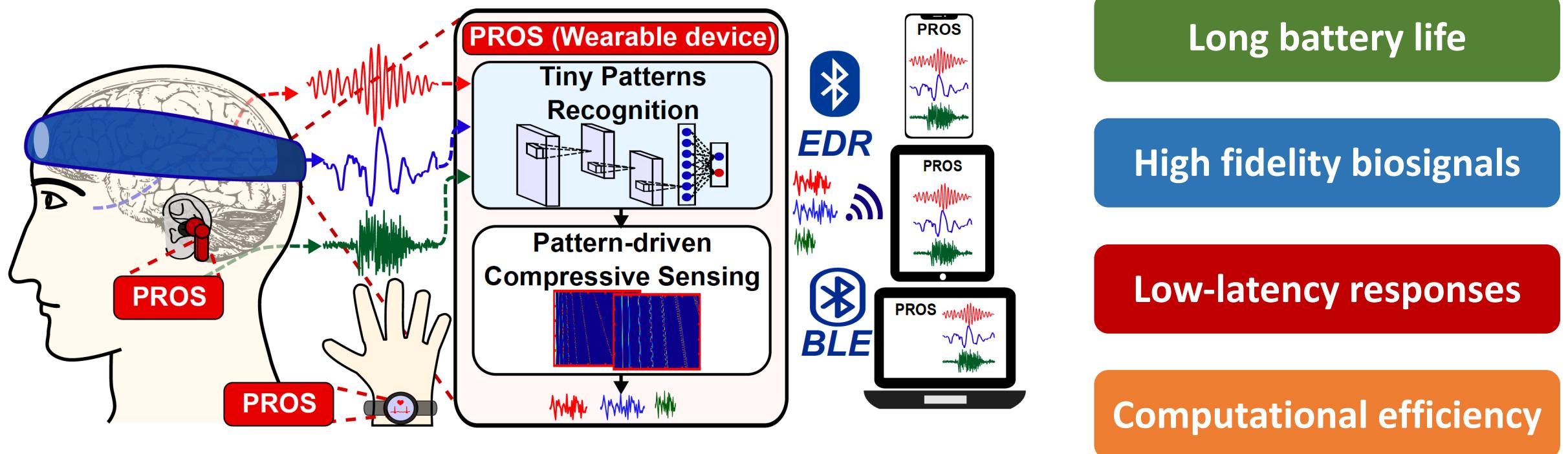
Sweat: 4Hz (vs 512Hz in diagnostic)

Motion: 3x32Hz, Temp.: 1Hz.

Real-time alerts through a mobile app.

A new solution is needed to overcome this trade-off!

Our proposed Pattern-dRiven cOmpressive Sensing (PROS) system

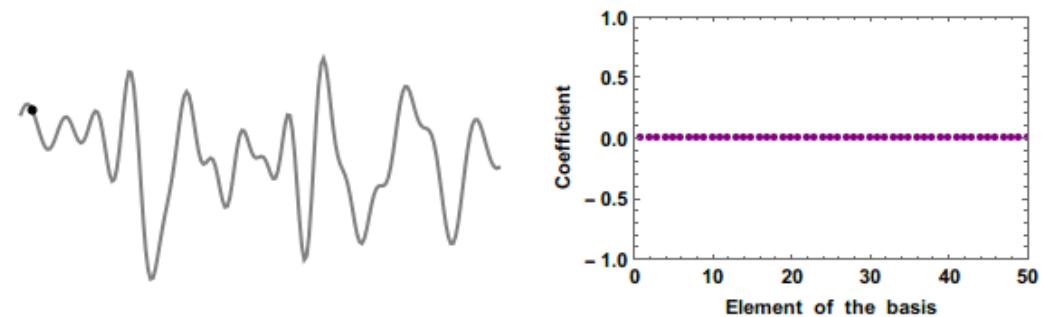


Observations

- ❑ #1: Biosignal events are sparse!



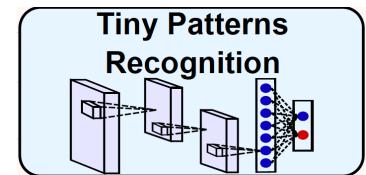
- ❑ #2: Biosignals are (potentially) sparse!



Exploiting event and signal sparsity -> significant streaming data reduction

- ❑ #3: However, developed algorithms need to be efficient!

Challenge #1: How to detect biosignal events efficiently on wearables? (1/2)

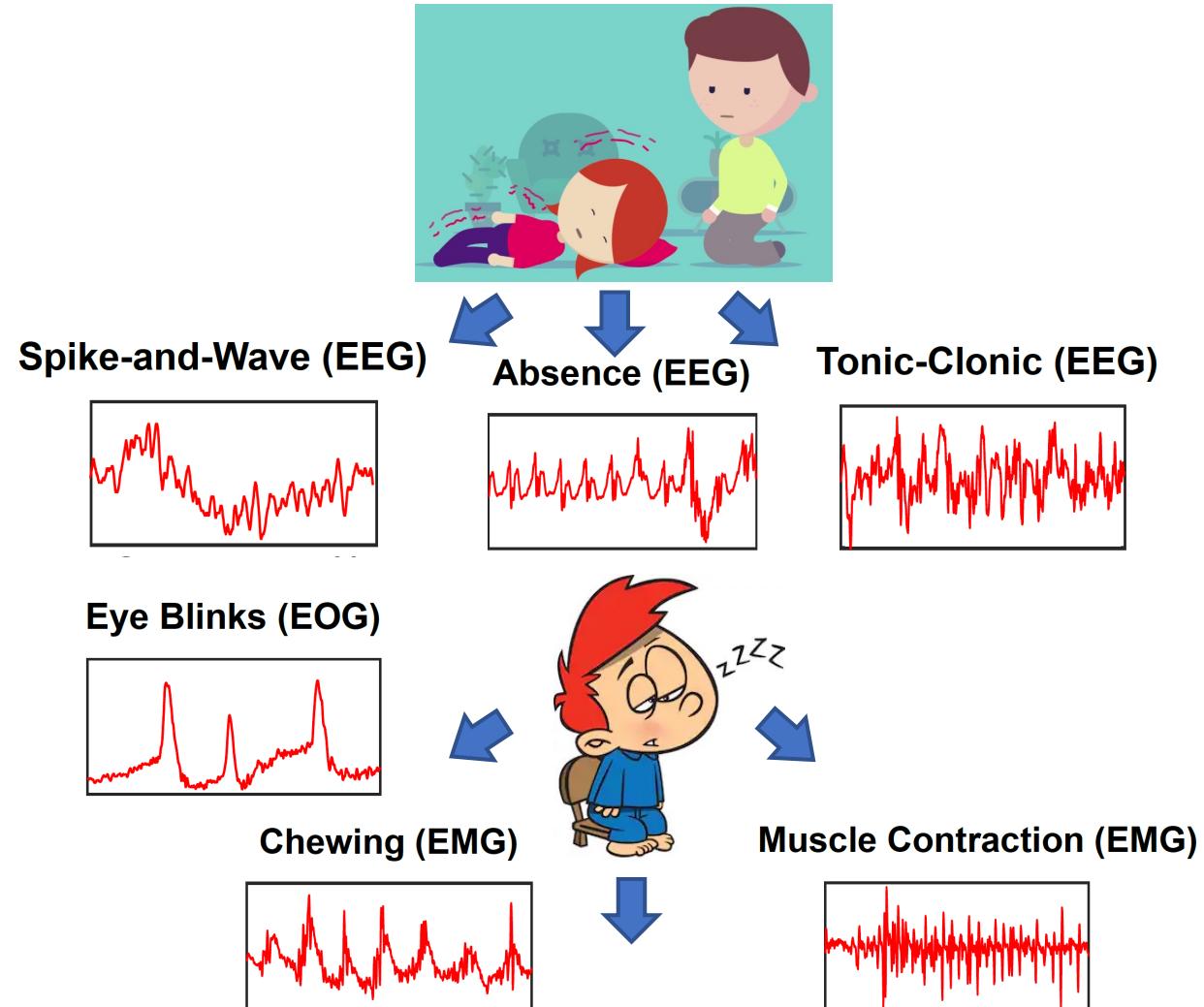


□ Detection of **biosignal events** on **wearables** is challenging!

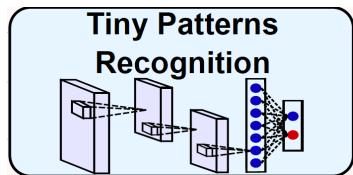
- Multiple signal modalities (e.g., EEG, EOG, EMG, etc.)
- Constrained hardware and energy resources.



Complex biosignal events can be broken down into patterns of interest (Pol's)!



Challenge #1: How to detect biosignal Pols efficiently on wearables? (2/2)



❑ Tiny Pattern Recognition models (TinyPRs)

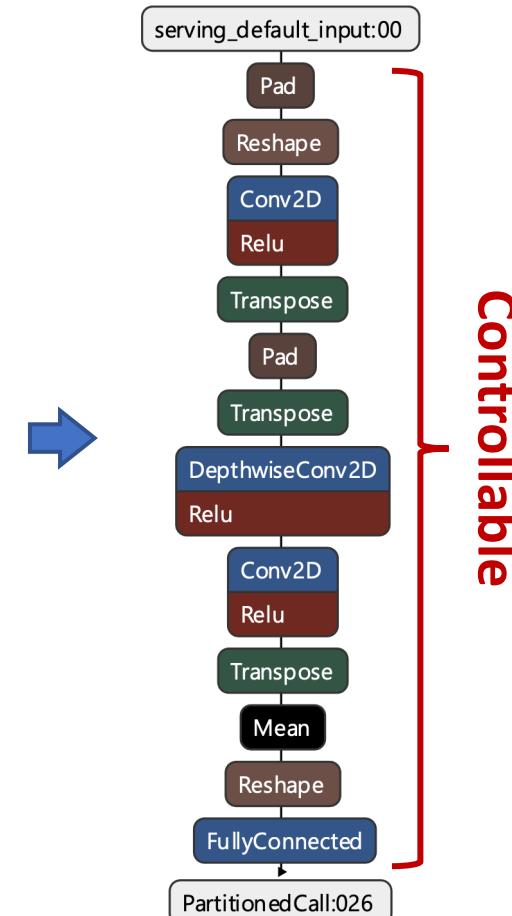
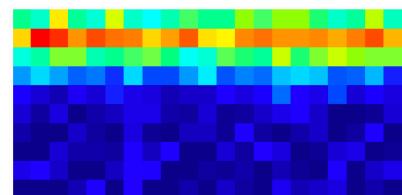
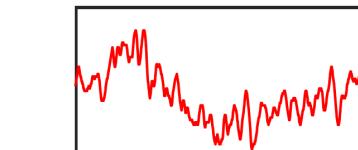
- Each TinyPR detects one pattern.
- Fine-tuned feature extraction (MFCC).
- Flexible Depth-wise and Point-wise Convolution layers.
- Accelerated by SIMD and DSP.



How can we push the data transmission reduction further?



Spike-and-Wave (EEG)



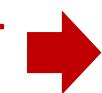
Pattern probability

$$y = \Phi x$$

Challenge #2: How to compress the detected Pol's? (1/2)

□ Biosignals are potentially sparse in time-frequency domains (wavelets, Gabor, etc.), but...

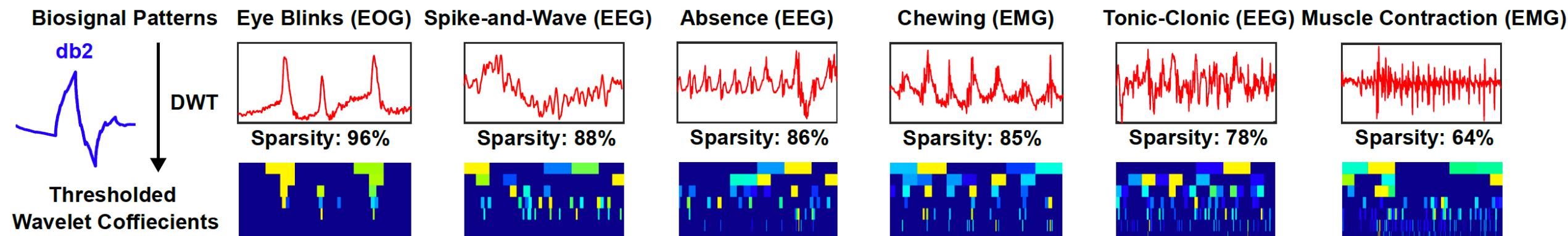
- Large variations among channels and trails. -> **sub-optimal compression!**
- A universal optimal sparse domain might not exist.



Optimal domains for individual biosignal patterns!



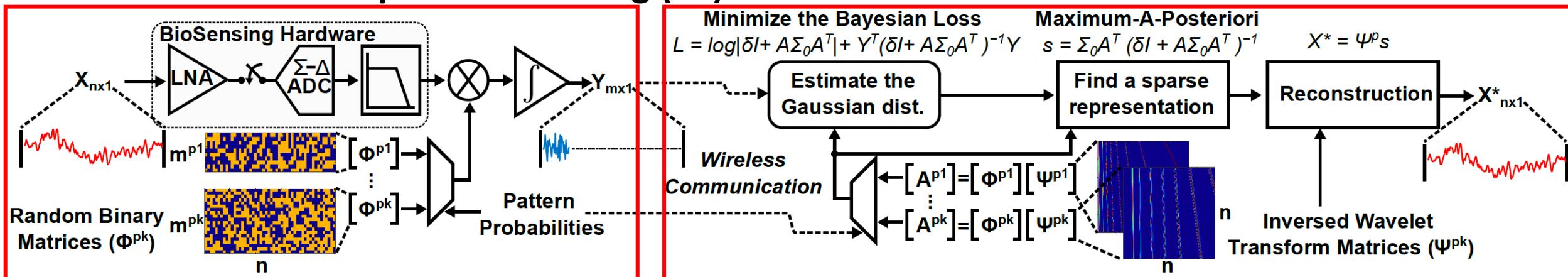
Optimal Wavelet Search Algorithm



$$y = \Phi x$$

Challenge #2: How to compress the detected Pols? (2/2)

□ Pattern-driven Compressive Sensing (CS)



○ Wearable device:

- **Pattern probability** from TinyPRs.
- **Dynamic Random Binary Matrices** compressing based on the detected patterns.

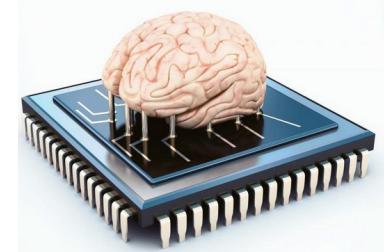


Processing efficiency?

○ Mobile device:

- Biosignal reconstruction based on **Block Sparse Bayesian Learning**.
- **Dynamic wavelet domains selection** on the fly.

Challenge #3: How to optimise processing efficiency? (1/1)



□ Dynamic Frequency and Voltage Scaling (DVFS)

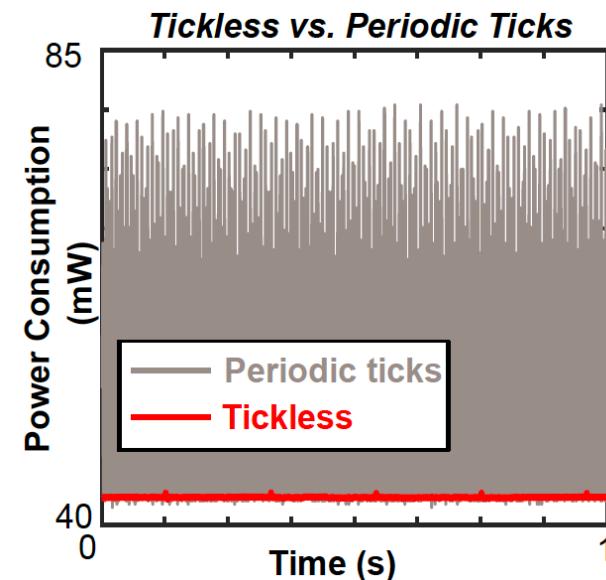
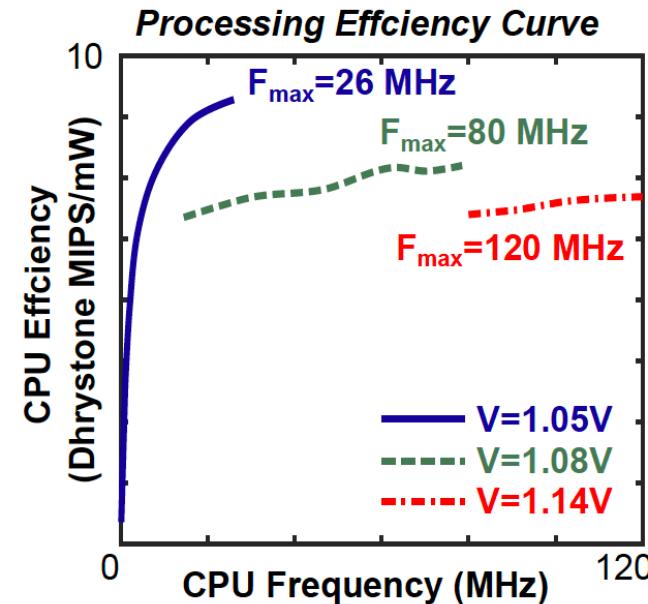
- Choose the optimal CPU efficiency point based on application's requirements.
- ↑ CPU Frequency -> ↑ Sleep time -> ↑ Voltage

□ Adaptive Energy Detector

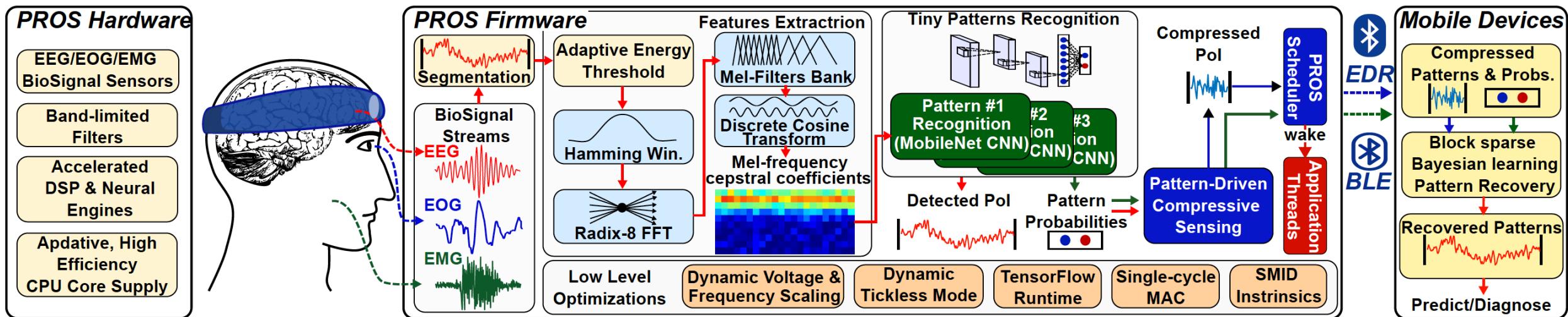
- Some background signals are too obvious for TinyPRs!

□ Tickless Kernel Mode

- Eliminate periodic CPU wake-ups!
- Ultra-low power interrupt-based timebase.



Implementation

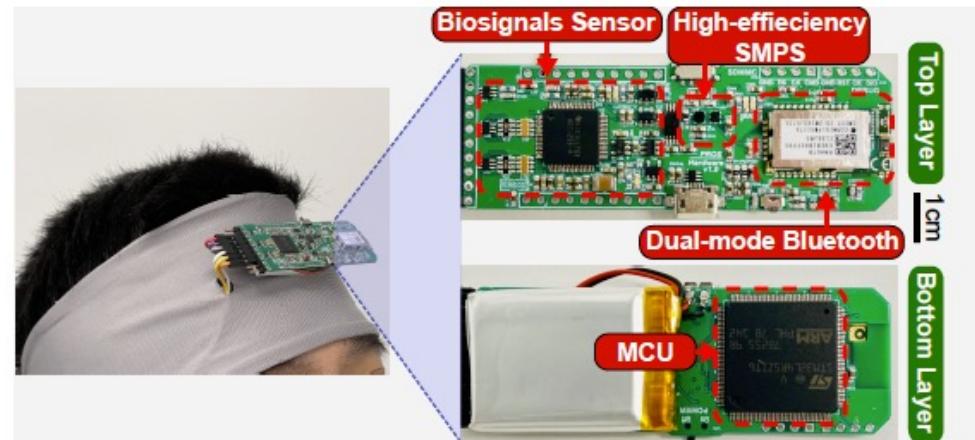


☐ Wearable

- Low-power ARM MCU, 120 MHz, 1MB FLASH, 256KB RAM,
- TinyPR pipeline,
- Pattern-driven Compressive Sensing,
- Hardware and low-level optimisations.

☐ Mobile

- Surface Go, Galaxy S20,
- Reconstruction algorithm.

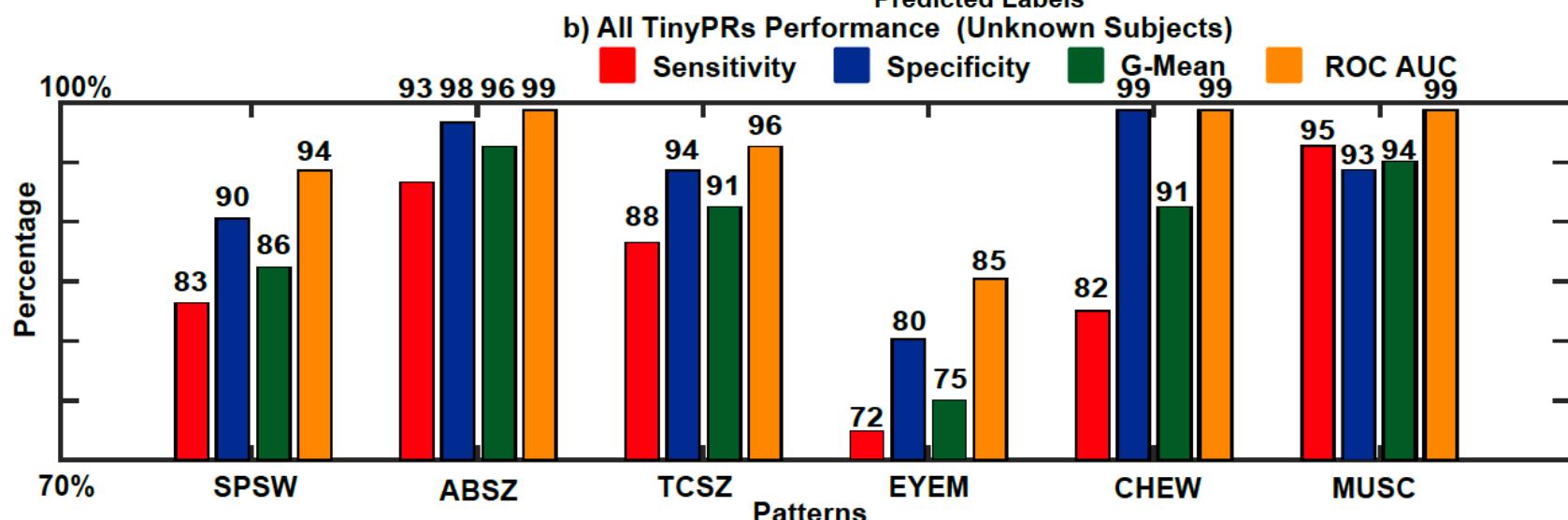
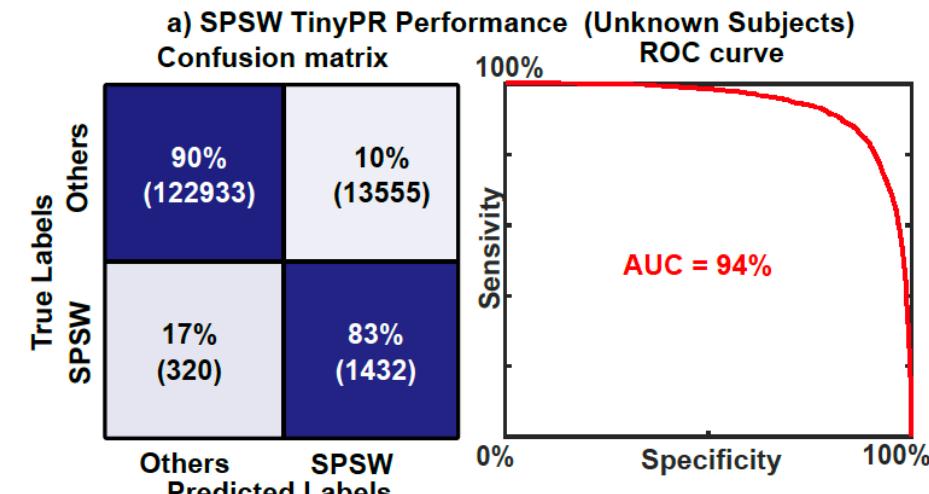


Optimised biosensing hardware

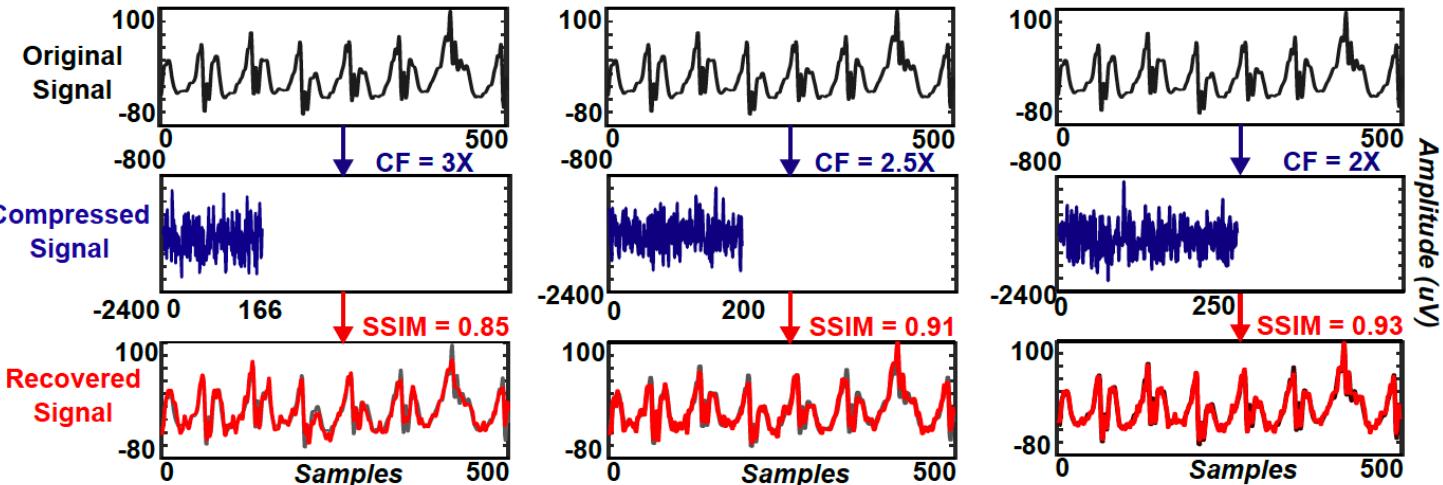
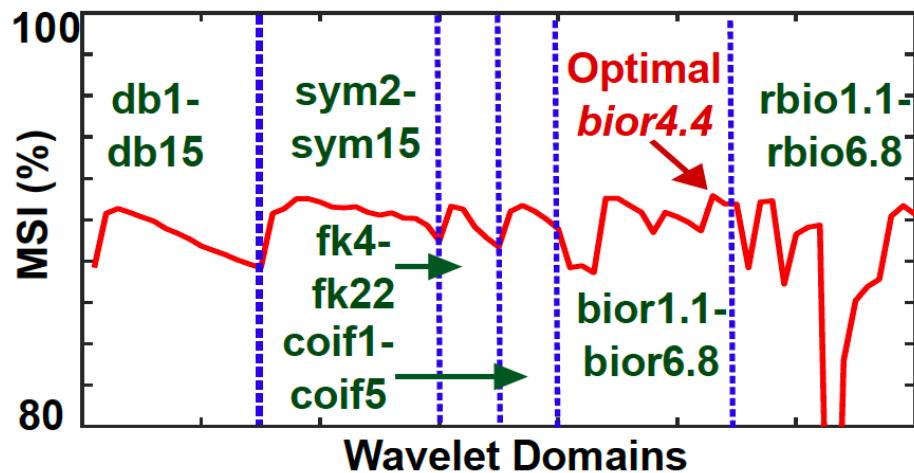
Evaluation #1 – Tiny Pattern Recognition Performance

❑ TinyPR's performance:

- 2 open datasets (hospital settings).
- 6 biosignal patterns.
- 120 subjects (100 train/20 test), 2,099,479 data points.
- **75% - 96% Sensitivity and Specificity** on unseen subjects.



Evaluation #2 – Pattern-Driven Compressive Sensing Performance



□ Optimal wavelet search

- 71% - 93% sparsity for each biosignal pattern.

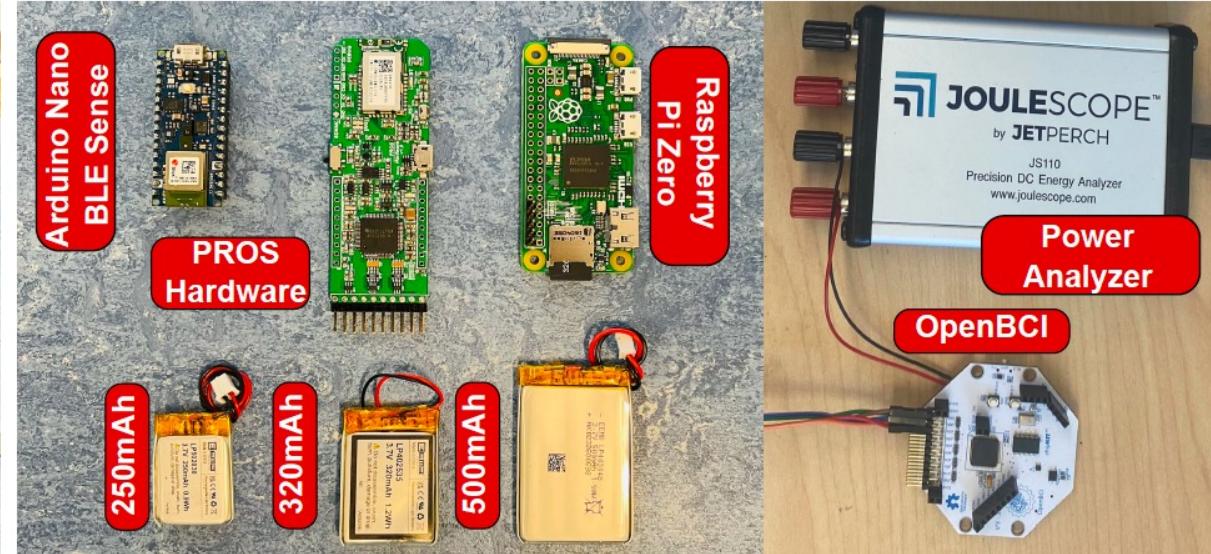
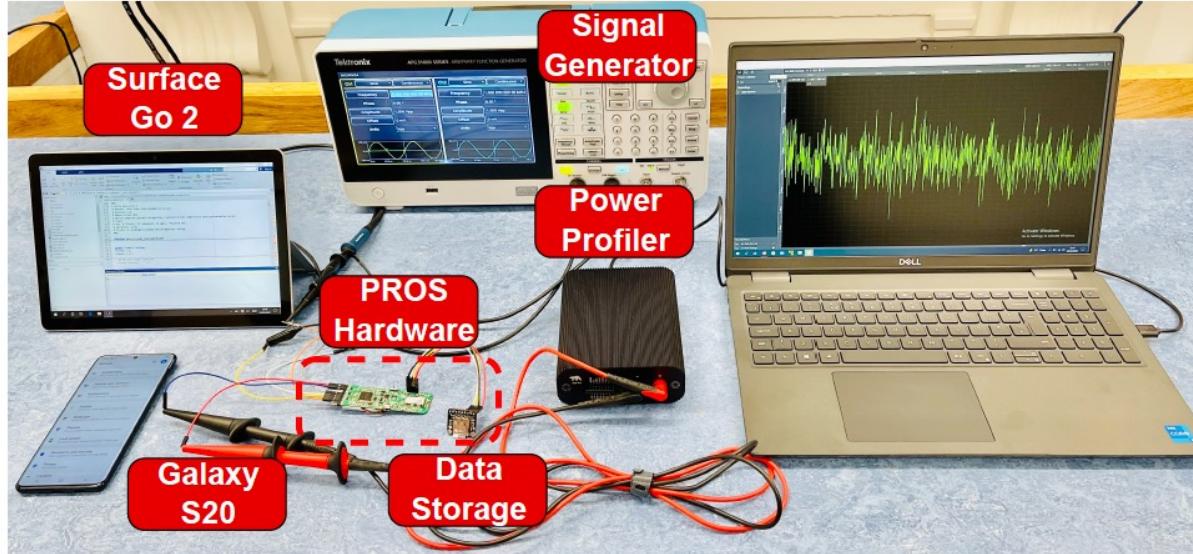
□ CS performance:

- Up to **5X compression** (CS alone)
- Up to **40X compression** (CS + TinyPR)

Table 1: Recovery quality with different CFs.

Pattern	Wavelet Domain	MSI (%)	SSIM with different CFs (w=3s)				
			1.5X	2X	3X	4X	5X
SPSW	bior6.8	92.6	0.99	0.98	0.96	0.94	0.89
ABSZ	bior4.4	91.2	0.99	0.97	0.94	0.89	0.82
TCSZ	sym14	71.4	0.84	0.81	0.57	0.39	0.31
EYEM	sym5	89.6	0.98	0.97	0.91	0.84	0.80
CHEW	bior4.4	84.0	0.93	0.93	0.84	0.78	0.71
MUSC	sym5	79.7	0.92	0.88	0.70	0.60	0.50

Evaluation #3 – Hardware runtime performance



Hardware performance

- **TinyPR:** 5KB FLASH, 30 KB RAM, 26ms (@120MHz).
- **CS:** 4-30KB FLASH, 1ms (Wearable), 50-94ms (Mobile devices)

Epileptic Seizure Monitoring use case:

- **10 unseen subjects:** 277,662 data points.
- **3 patterns:** Focal, Absence, Tonic-Clonic.
- **High-fidelity:** 85% Sensitivity, 0.92 SSIM.
- **Transmission reduction:** 24X.
- **Energy improvement:** >1200%.
- 500mAh -> last the whole week.
- Able to **detect fatal tonic-clonic seizures < 32ms.**

Key takeaways

- On-chip intelligence could **enable new optimizations on low-power devices**.
- The performance of embedded models is **always be constrained by energy**.
- **Pre-processing techniques** are important to help increasing accuracy, especially in the low-power domain.

Invisible bio-sensing

