An Ultra-Low Energy Human Activity Recognition Accelerator for Wearable Health Applications

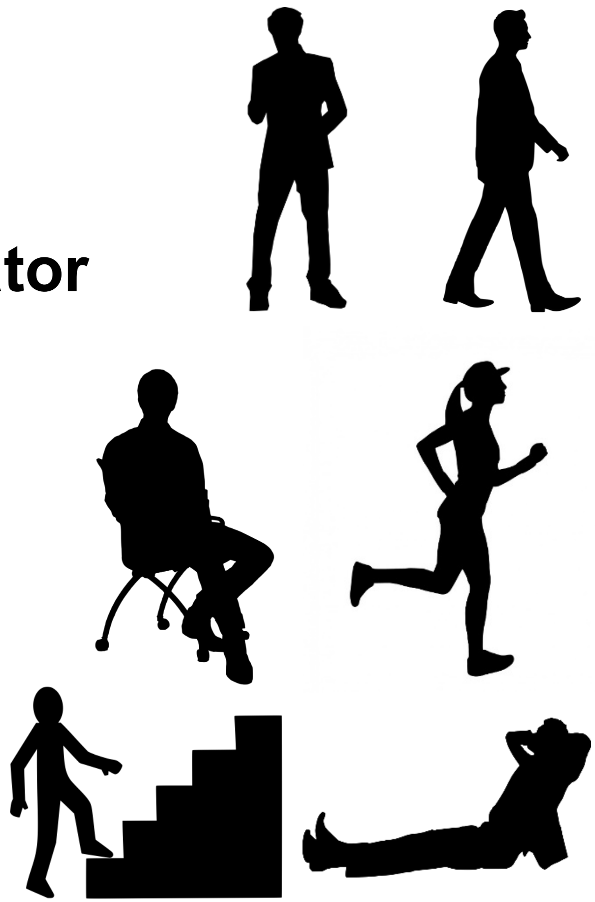
Ganapati Bhat*, Yigit Tuncel
Sizhe An, Umit Y. Ogras
Arizona State University

Hyung Gyu Lee
Daegu University

CASES, October 14, 2019

Outline

- **Motivation**
- **Related Work**
- **Human Activity Recognition Accelerator**
  - Baseline HAR Engine
  - Activity-Aware 2-Level HAR Engine
- **Low Power Optimizations**
- **Experimental Results**
- **Conclusion**

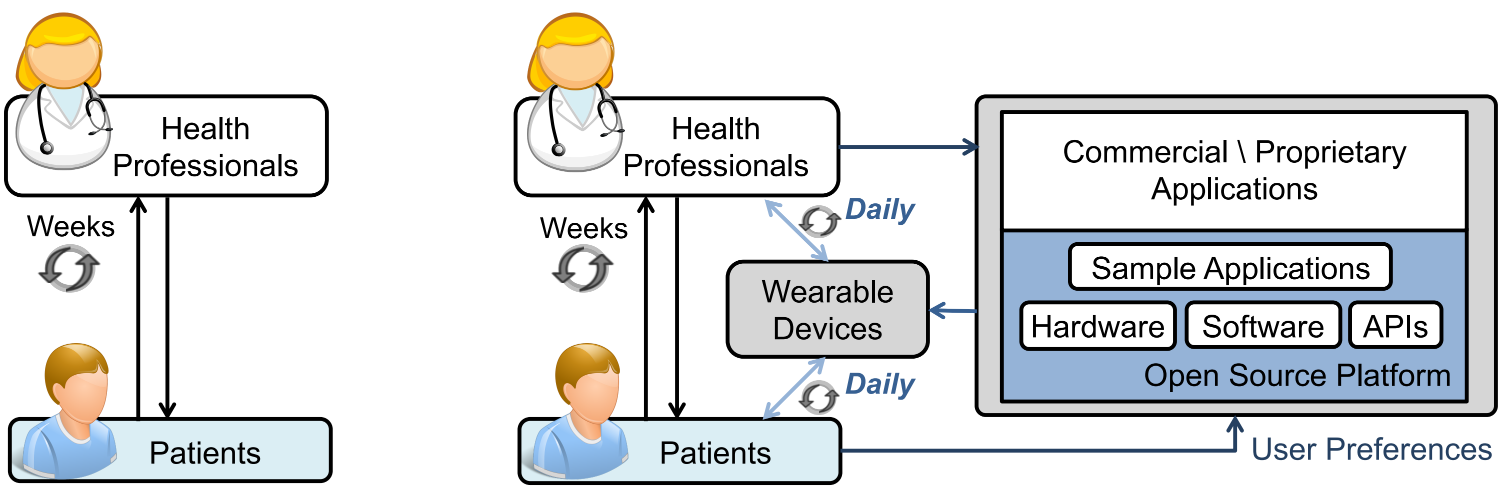


Health Monitoring using Wearables

- **15% of the world's population lives with a disability***
- **110-190 million people face difficulties in functioning***
- ***Intl. Parkinson and Movement Disorders Society Task Force on Technology:***
  - Low-cost and small form-factor wearable devices offer great potential
  - Enabled by advances in low power sensors and processors

Current Health Practice

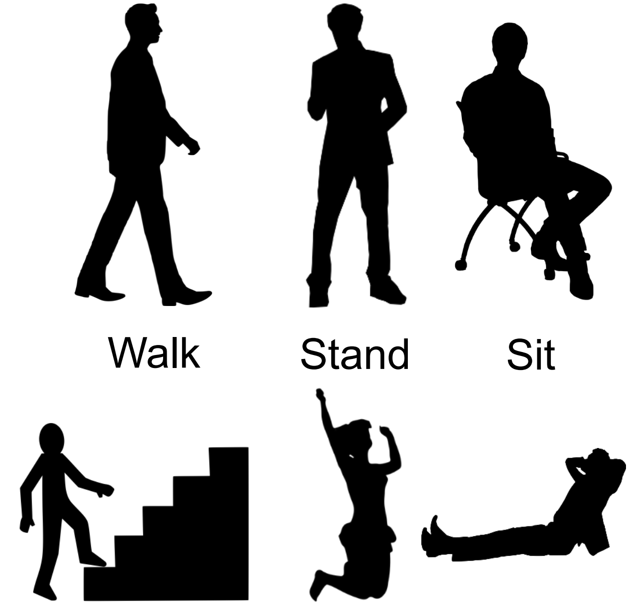
OpenHealth Wearable Vision



*World Report on Disability: http://www.who.int/disabilities/world report/2011/report/en/.

Why Human Activity Recognition (HAR)?

- **Identify activities, such as walking, sitting, driving, jogging**
- **First step to solutions for movement disorders**



Walk Stand Sit

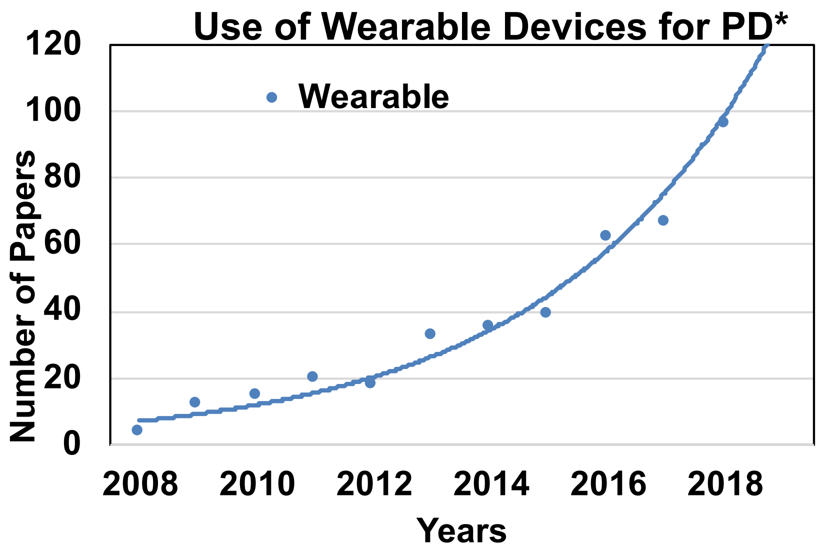
Up/down stairs Jump Lie Down

We have to know what the patient is doing to reach a conclusion

- **HAR can provide valuable insight**
- **Applications of HAR**
  - Patient rehabilitation
  - Fall detection
  - Physical activity promotion

Challenges of Wearable Health Technology

- *Adaptation* & *technology* **challenges hinder widespread adoption**
  - **Comfort:** Awkward to wear or carry a device
  - **Compliance:** Stop using technology due to maintenance
  - **Applications:** No killer applications
- **27% users give up due to charging reqs [1]**
  - Practical solutions must minimize energy



*Ranadeep Deb, MS Thesis, 2019

- **Flexible energy harvesting devices can address these problems**
- **However,**
  - Ambient power is still lower than 10 to 30 mW requirement
  - Mere 40 hrs with 130 mAh battery

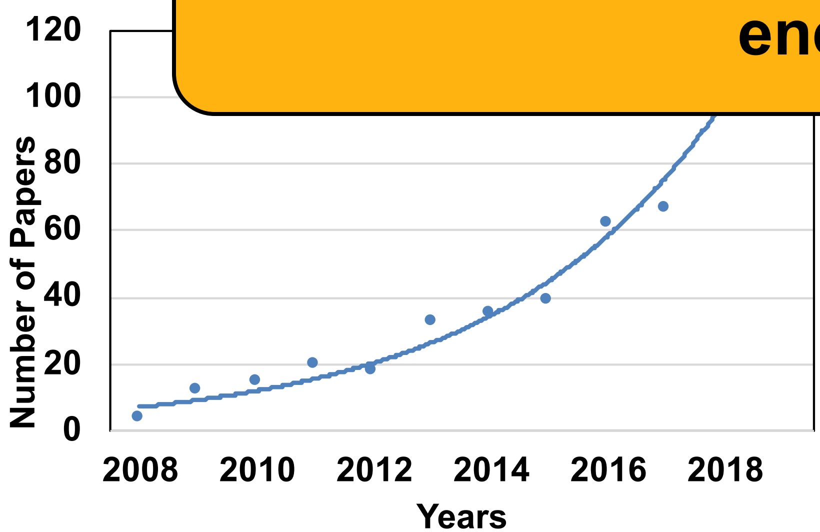
Flexible PV-cell

[1] Ana Lígia Silva de Lima et al.. *Feasibility of Large-Scale Deployment of Multiple Wearable Sensors in Parkinson's Disease*. PLOS One 12, 12 (2017), e0189161

Challenges of Wearable Health Technology

- ***Adaptation* & *technology* challenges hinder widespread adoption**
  - **Comfort:** Awkward to wear or carry a device
  - **Compliance:** Stop using technology due to maintenance
  - **Applications:** No killer applications
- **27% users give up due to charging reqs [1]**
  -

Low-power accelerators needed to meet energy budget



*Ranadeep Deb, MS Thesis, 2019

ces
can address these problems

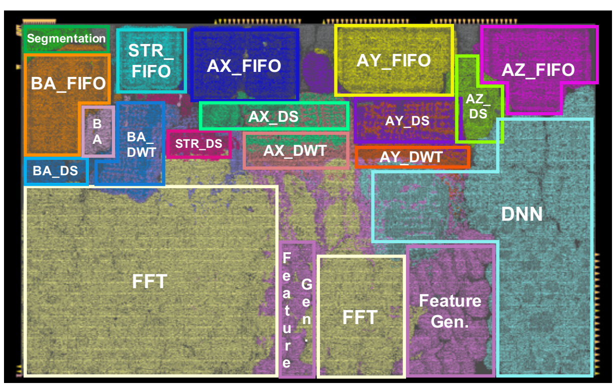
- **However,**
  - Ambient power is still lower than 10 to 30 mW requirement
  - Mere 40 hrs with 130 mAh battery

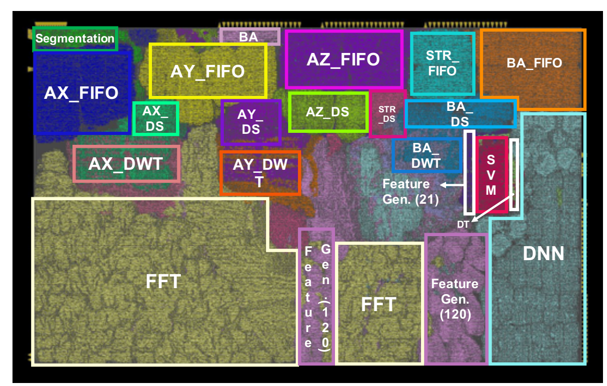
Flexible PV-cell

[1] Ana Lígia Silva de Lima et al.. *Feasibility of Large-Scale Deployment of Multiple Wearable Sensors in Parkinson's Disease*. PLOS One 12, 12 (2017), e0189161

Our Novel Contributions

- **The first integrated full hardware accelerator for HAR**
  - Sensor reading to activity classification
- **Novel activity-aware design to minimize energy consumption**
  - 22.4 µJ per activity (>17 days with 130 mAh battery)
- **Post layout evaluation using TSMC 65 nm LP**
- **Extensive experimental evaluation with 22 users**
  - Dataset released to public (https://github.com/gmbhat/human-activity-recognition)





A critical step towards *self-powered* health monitoring devices

Related Work

Ref	[1]	[2]	[3]	[4]	[5]	Proposed
Target App.	Vital signal monitoring	Vital signal monitoring	Signal acquisition	Signal acquisition	Sensor AFE for physical act.	HAR
Technology	130 nm	130 nm	180 nm	180 nm	500 nm	65 nm
Frequency	32 kHz or 16 MHz	1-20 MHz	1 MHz	Up to 2 kHz	120 Hz	100 kHz
Voltage	1.0 V	0.9 V	1.2 V	1.1 V	2.7 V - 3.3 V	1.0 V
Power	530 μW	93-322 μW	191 μW	88.6 μW	120 μW	45 – 51 μW
Area	16 mm^2	6.25 mm^2	49 mm^2	5.45 mm^2	196 mm^2	1.35 mm^2

[1] Alan CW Wong et al. *A 1 V, Micropower System-On-Chip For Vital-Sign Monitoring In Wireless Body Sensor Networks.* ISSCC 2018

[2] Yuxuan Luo et al. *A 93µW 11Mbps Wireless Vital Signs Monitoring Soc With 3-Lead ECG, Bio-Impedance, And Body Temperature*. In Proc. IEEE Asian Solid-State Circuits Conf, 2017

[3] Nick Van Helleputte et al. *18.3 A Multi-Parameter Signal-Acquisition Soc For Connected Personal Health Applications.* ISSCC 2014

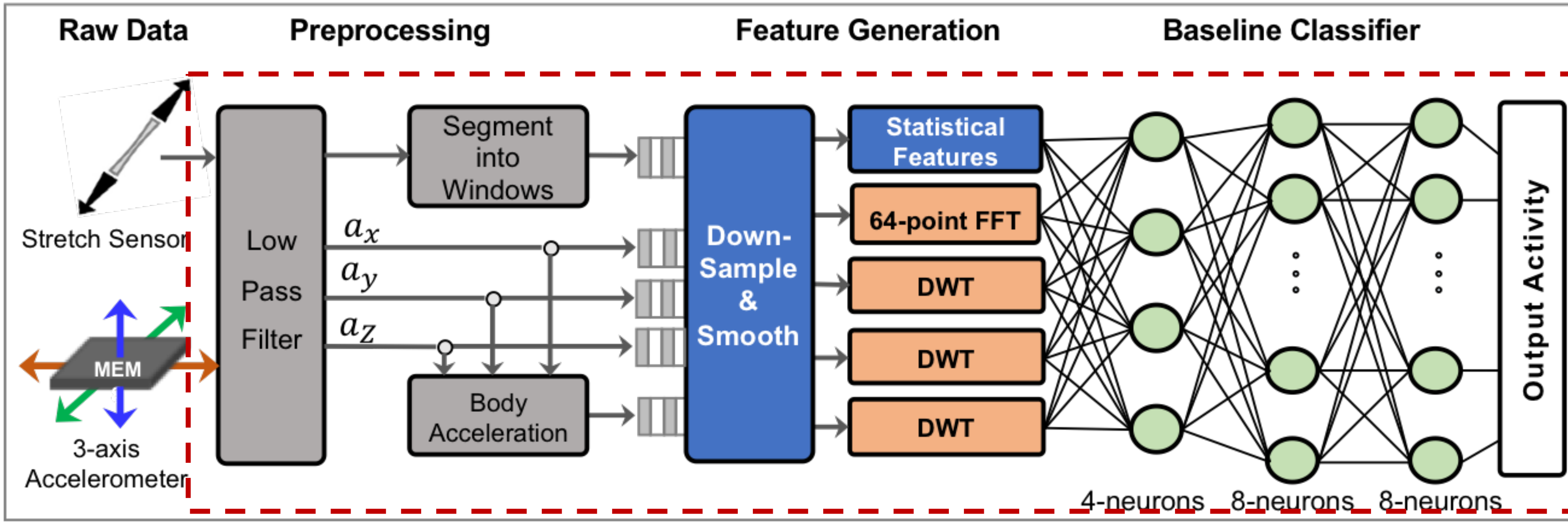
[4] Xin Liu et al. *An Ultra-Low Power ECG Acquisition And Monitoring ASIC System For WBAN Applications. IEEE J. on Emerg. and Sel. Topics in Circuits Syst.* 2, 2012.

[5] Wouter Bracke et al.. *A 1 cm^3 Modular Autonomous Sensor Node For Physical Activity Monitoring*. Ph.D. Research in Microelectronics and Electronics, 2006.

Outline

- Motivation
- Related Work
- **Human Activity Recognition Accelerator**
  - Baseline HAR Engine
  - Activity-Aware 2-Level HAR Engine
- Low Power Optimizations
- Experimental Results
- Conclusion

Baseline HAR Engine Overview

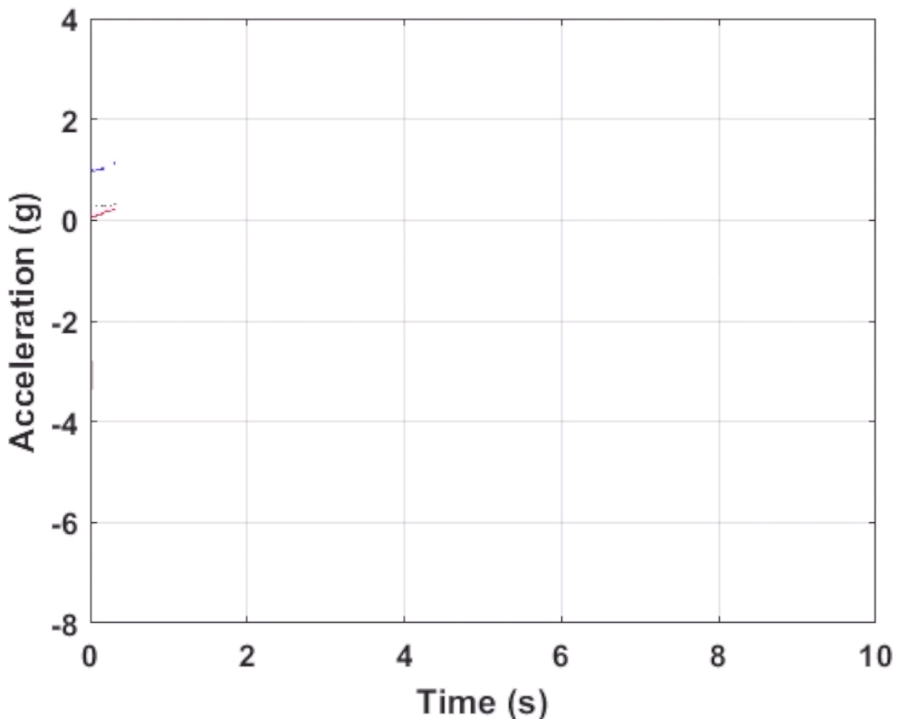


- **Stretch sensor input:** *Measures bending of the knee*
- **Accelerometer input:** *Measures acceleration at ankle*
- **Activities**

Sit Stand Walk Jump Up/down stairs Lie Down

Input Sensor Data- Accelerometer

- **3-axis accelerometer data**
  - The most commonly used sensor for activity recognition
  - Since it is notoriously known to be noisy, preprocess using 8-point moving average filter



- **Invensense MPU-9250**
- **Low pass filter**

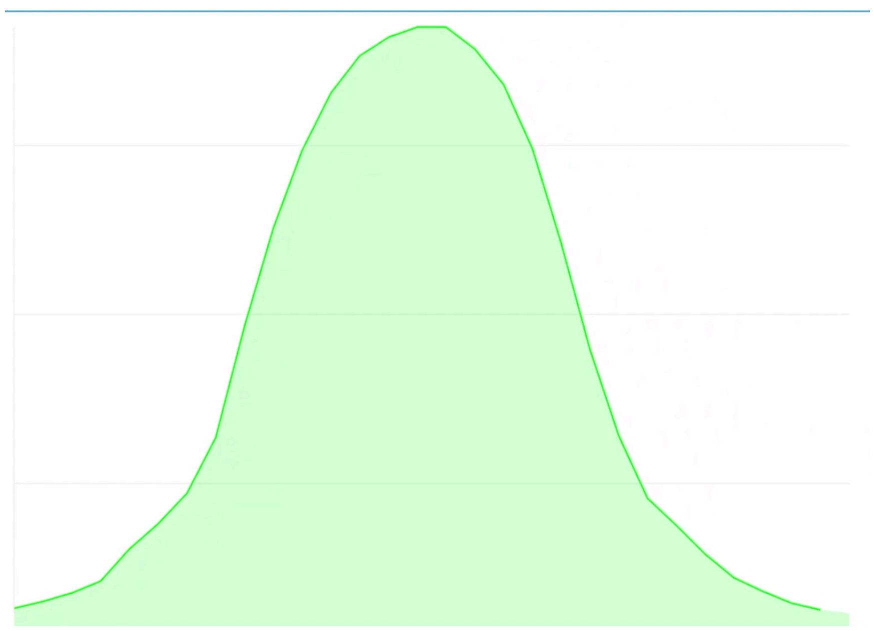
$$\bar{s}[kT_s] = \frac{1}{8}\sum_{i=-3}^{4} s[(k+i)T_s]$$

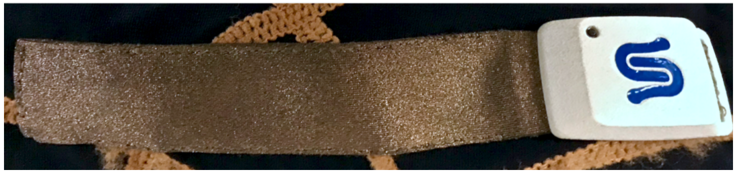
where T_s: Sampling time,
$\bar{s}[kT_s]$: Averaged sample at time kT_s
$s[kT_s]$: Raw sample at time kT_s

- **Filter applied to 3-axis data**

Input Sensor Data – Stretch Sensor

- **3-axis accelerometer data**
  - The most commonly used sensor for activity recognition
  - Since it is notoriously known to be noisy, preprocess using 8-point moving average filter
- **Use a textile-based stretch sensor (<u>first time for HAR</u>)**





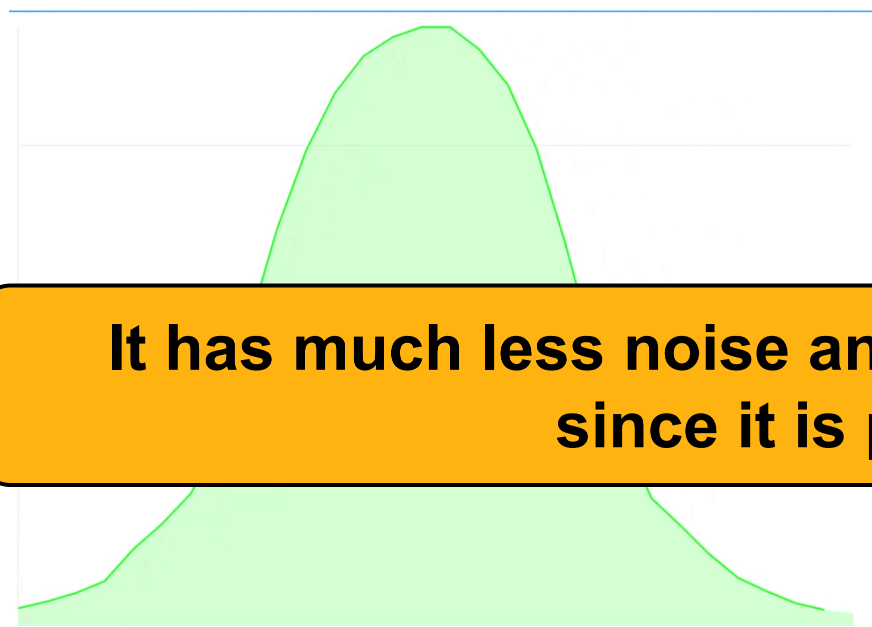
- ***Stretchsense* Stretch Sensor**
- **Low pass filter**

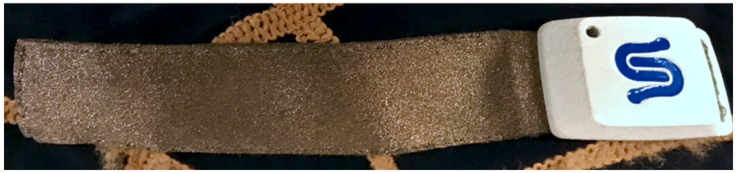
$$\bar{s}[kT_s] = \frac{1}{8}\sum_{i=-3}^{4} s[(k+i)T_s]$$

where T_s: Sampling time, $s[kT_s]$, $\bar{s}[kT_s]$: Raw, averaged sample at time kT_s

Input Sensor Data – Stretch Sensor

- **3-axis accelerometer data**
  - – The most commonly used sensor for activity recognition
  - – Since it is notoriously known to be noisy, preprocess using 8-point moving average filter
- **Use a textile-based stretch sensor (<u>first time for HAR</u>)**





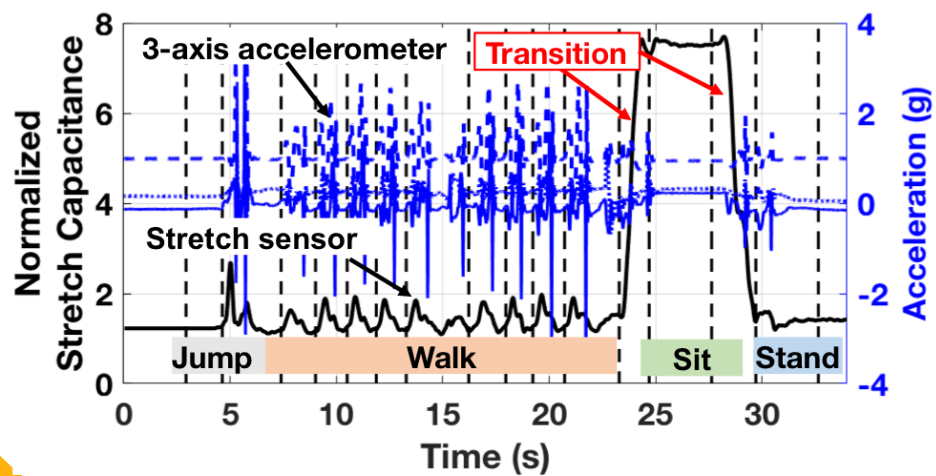
- ***Stretchsense* Stretch Sensor**
- **Low pass filter**

It has much less noise and power consumption since it is passive

where T_s: Sampling time, $s[kT_s]$, $\bar{s}[kT_s]$: Raw, averaged sample at time kT_s

Input Sensor Data – Segmentation

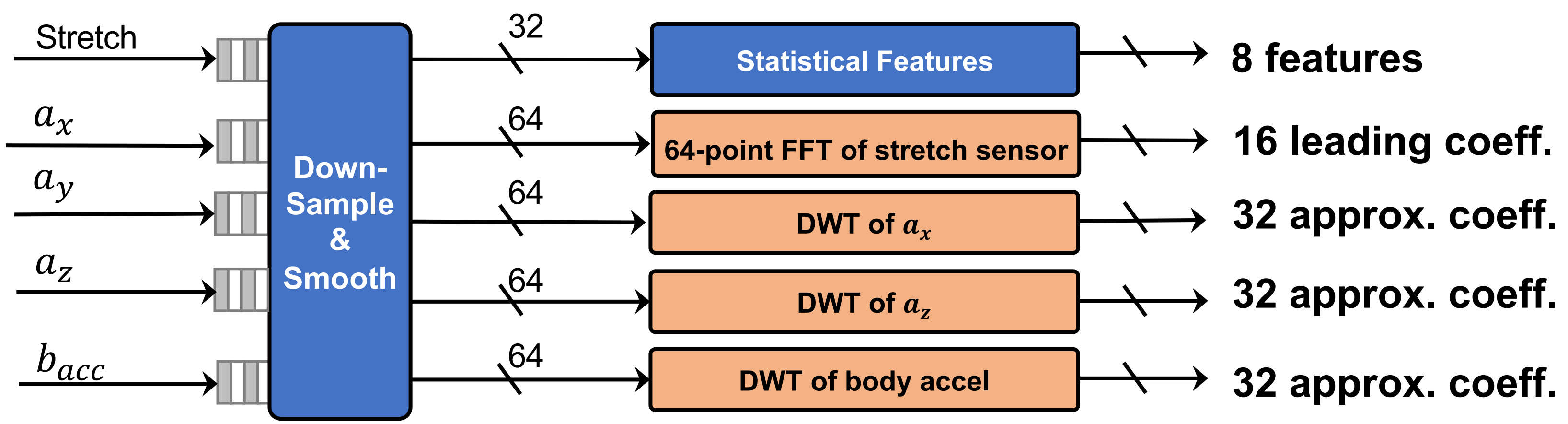
- **3-axis accelerometer data**
  - The most commonly used sensor for activity recognition
  - Since it is notoriously known to be noisy, preprocess using 8-point moving average filter
- **Use a textile-based stretch sensor (<u>first time for HAR</u>)**
- **Segment data into windows by detecting local minima in stretch sensor**



- 5-pt derivative to define trends in data
- A new segment when the trend changes from
  - *Decreasing* to *Increasing*
  - *Flat* to *Increasing*

Feature Generation

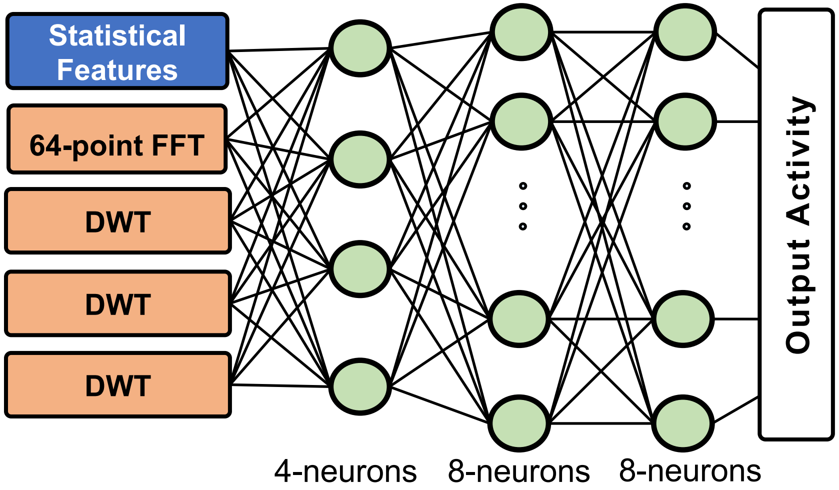
- **Non-uniform samples due to variable segment length**
- **Down sample and smooth**
  - Down sample block standardizes number of samples
  - 64 for accelerometer, 32 for stretch sensor
- **16-bit Neural Network Features**



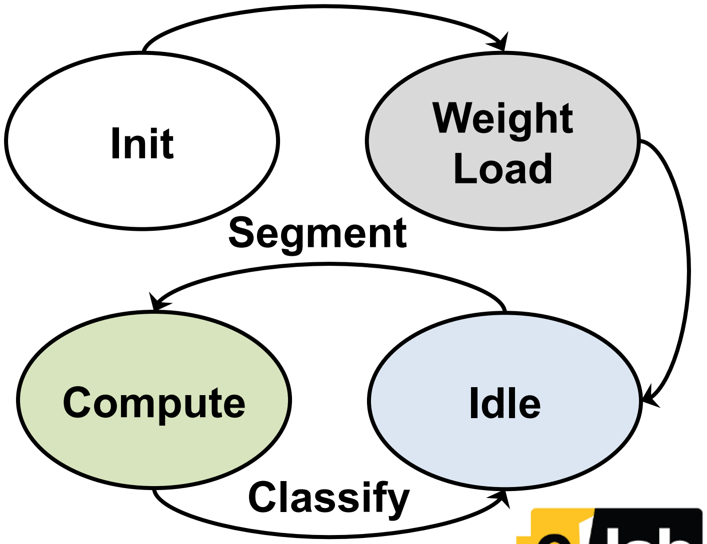
- **Statistical Features**
  - Variance of a_x, a_y, a_z, b_{acc} and mean of a_y
  - Min, max of stretch sensor and window length

Baseline DNN Classifier

- **Detailed neural architecture space exploration**
- **2 Hidden layers**
  - ReLU Activation
- **Output layer with 8 neurons**
  - Linear activation with *max*
  - More hardware-friendly compared to softmax
- **Operation and optimizations**
  - Design a parameterized module
  - Instantiate for hidden and output layers
  - Only one hour required to change from 3 layer to 2 layer network

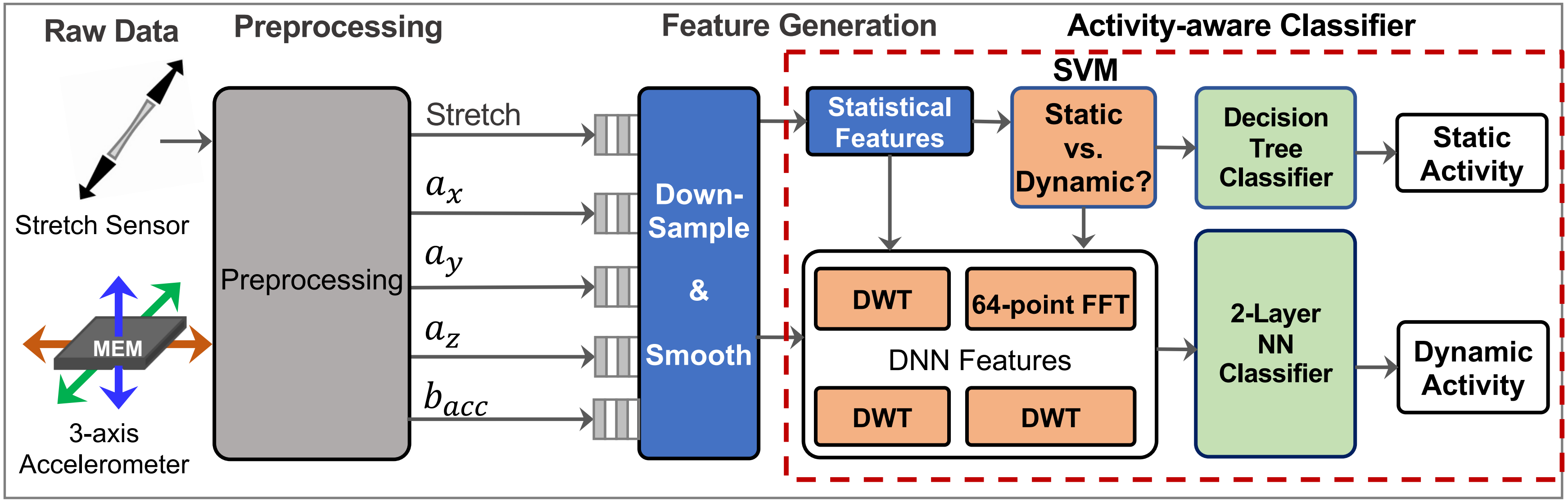


State machine for DNN



Activity-Aware 2-Level Engine

- **84% of human activities are static (e.g. sit, stand, lie down)**
  – We do not need a DNN to classify them
  – At the same time, more complex dynamic activities must be classified accurately
- **Divide the activities into two classes**
  – A simple support vector machine (SVM) to identify static vs dynamic
  – A 2-Layer NN classifier for dynamic activities

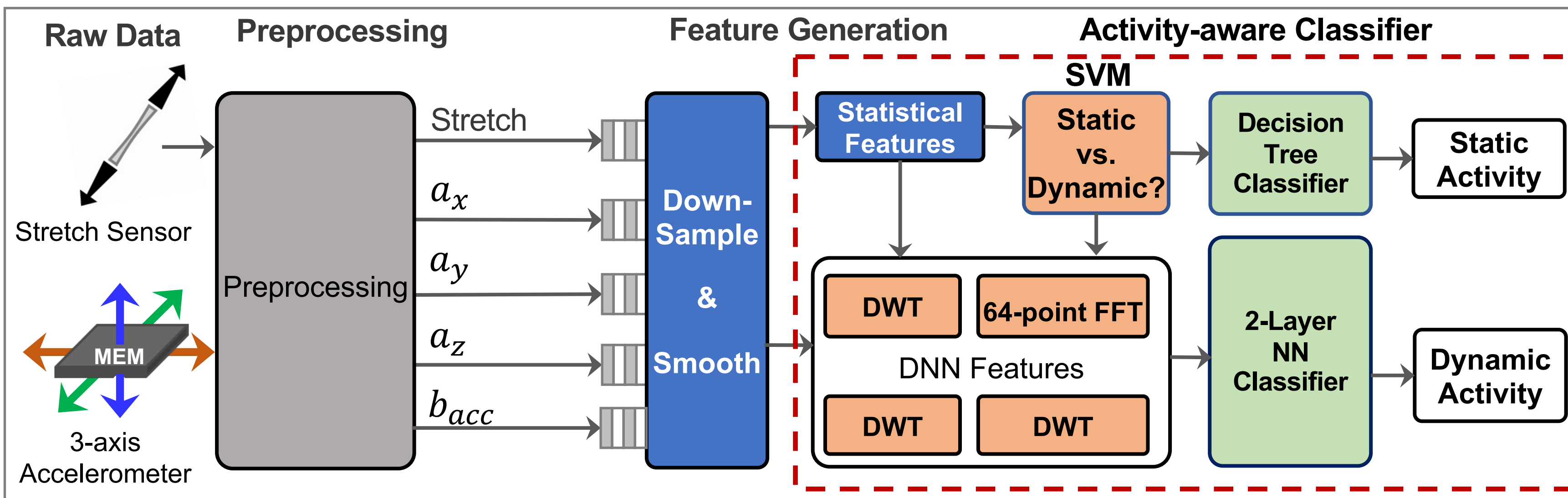


US Department of Labor. 2017. American Time Use Survey. [Online] https://www.bls.gov/tus/

Activity-Aware 2-Level Engine

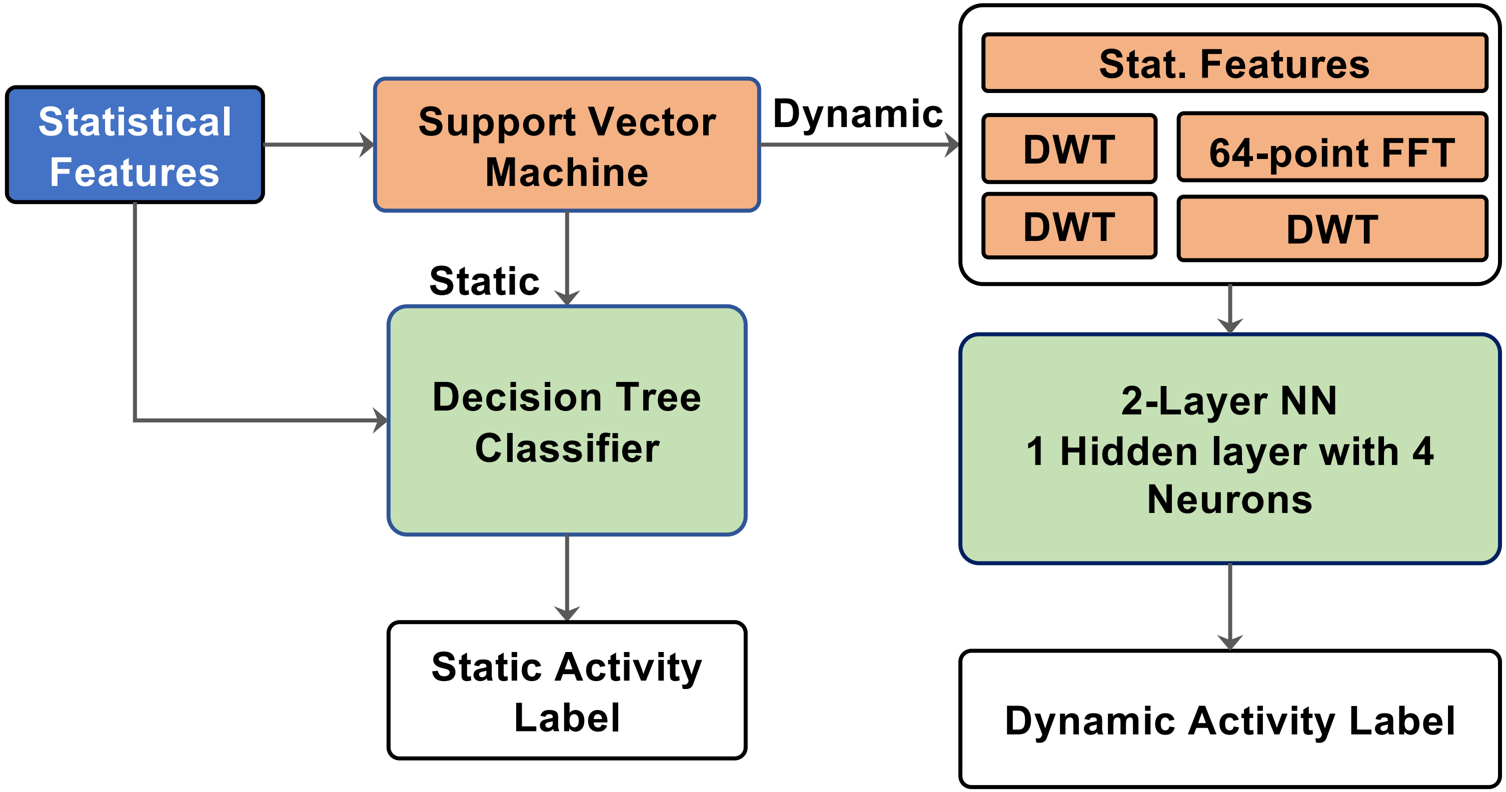
- **84% of human activities are static (e.g. sit, stand, lie down)**
  - We do not need a DNN to classify them
  - At the same time, more complex dynamic activities must be classified accurately

Avoids power hungry FFT and DNN blocks for 84% of activities



US Department of Labor. 2017. American Time Use Survey. [Online] https://www.bls.gov/tus/

Activity-Aware Classification



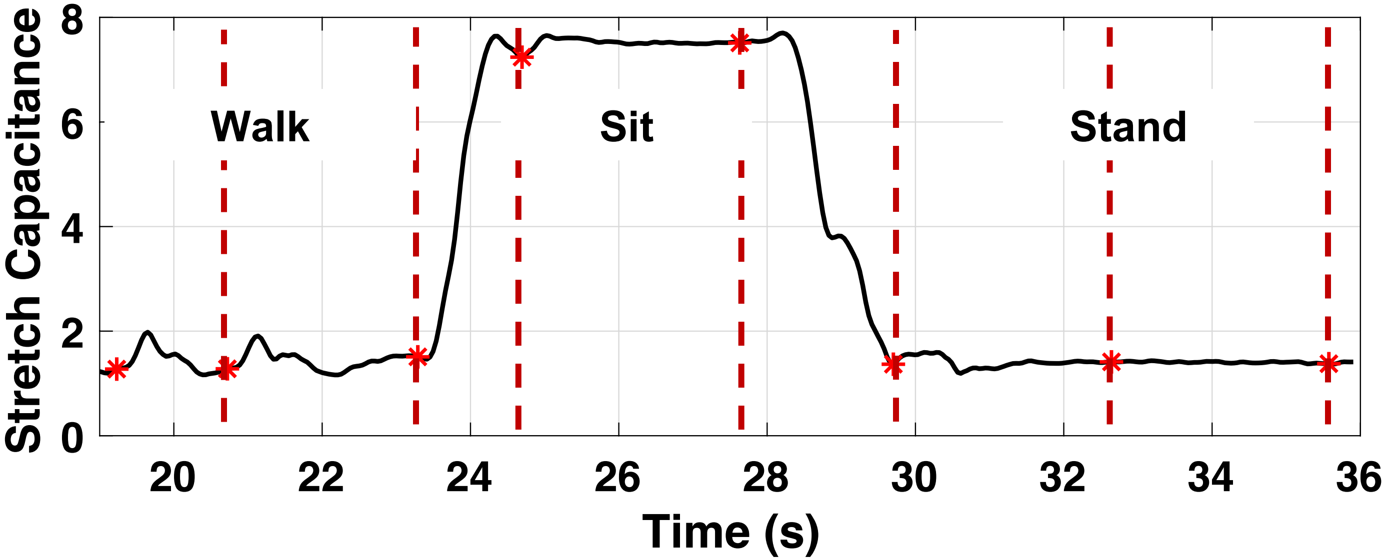
- **Features are reused between SVM and decision tree**
- **DWT and FFT calculated *only if* activity is dynamic**

Outline

- Motivation
- Related Work
- Human Activity Recognition Accelerator
  - Baseline HAR Engine
  - Activity-Aware 2-Level HAR Engine
- **Low Power Optimizations**
- Experimental Results
- Conclusion

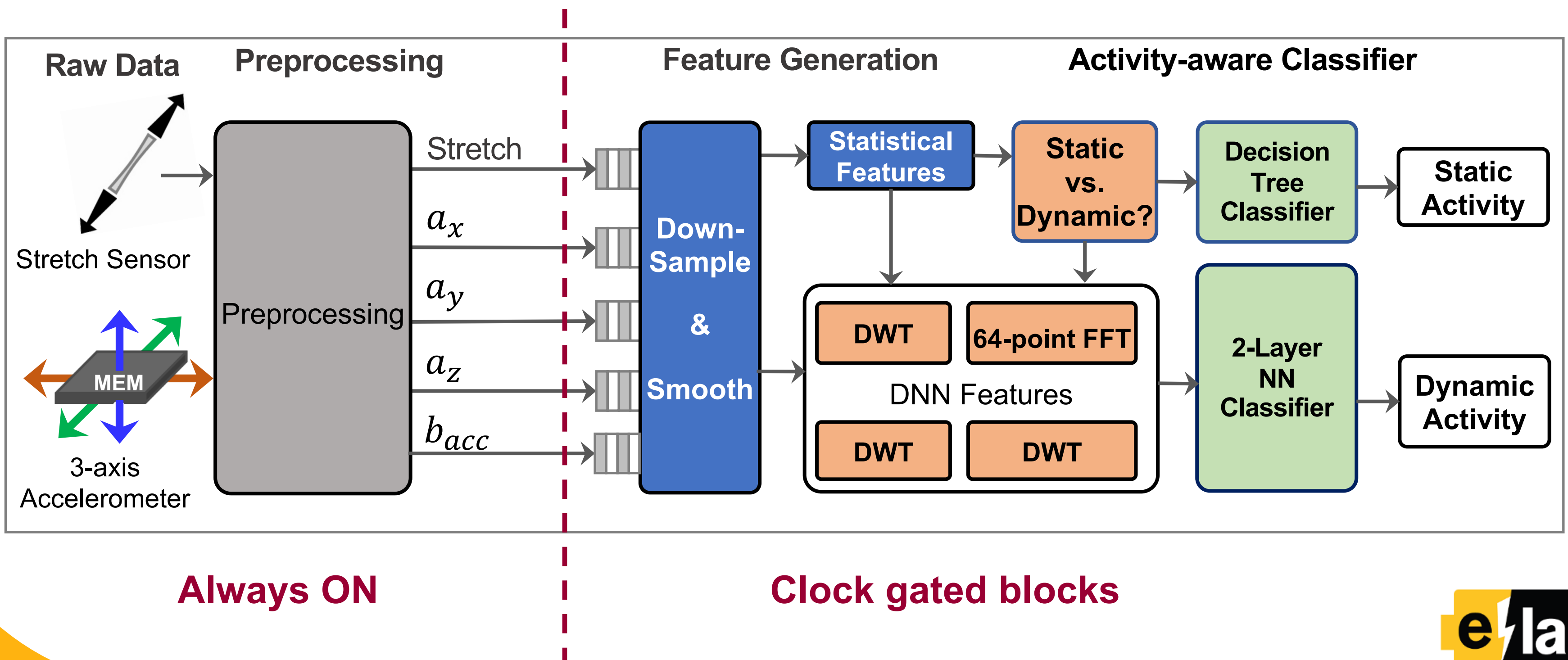
Clock and Data Gating

- **Human activities are in the order of few Hz**
  - Use this information to clock gate unused blocks



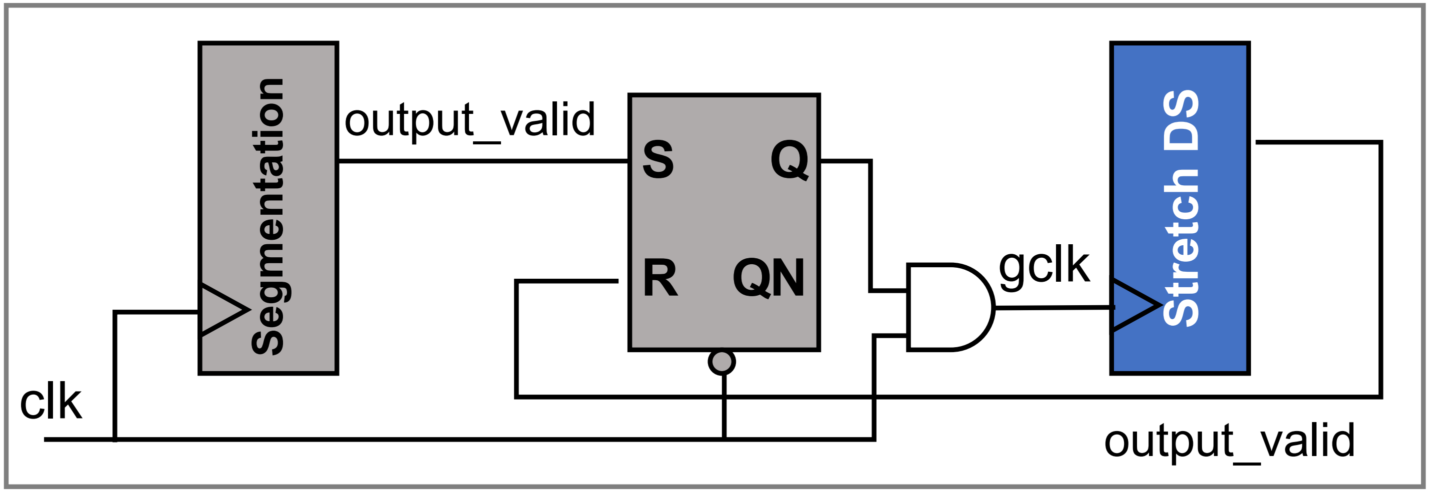
Clock and Data Gating

- **Human activities are in the order of few Hz**
  - Use this information to clock gate unused blocks
- **Data dependencies**
  - e. g., downsampling depends on segment detection



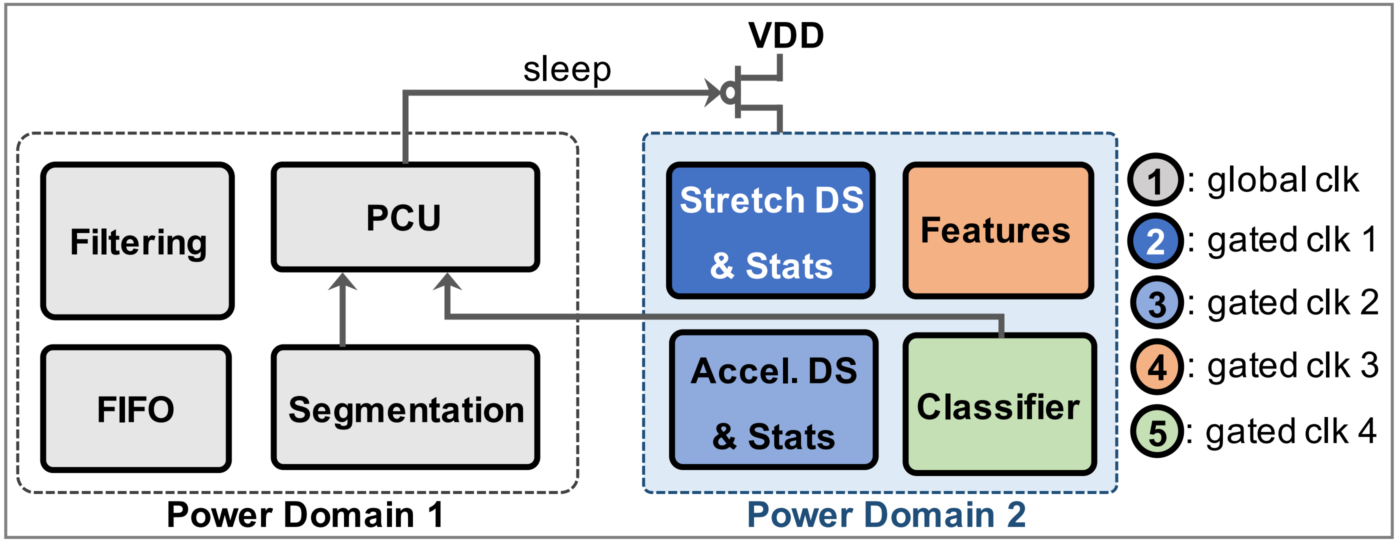
Clock and Data Gating

- **Human activities are in the order of few Hz**
  – Use this information to clock gate unused blocks
- **Data dependencies**
  – e. g., downsampling depends on segment detection



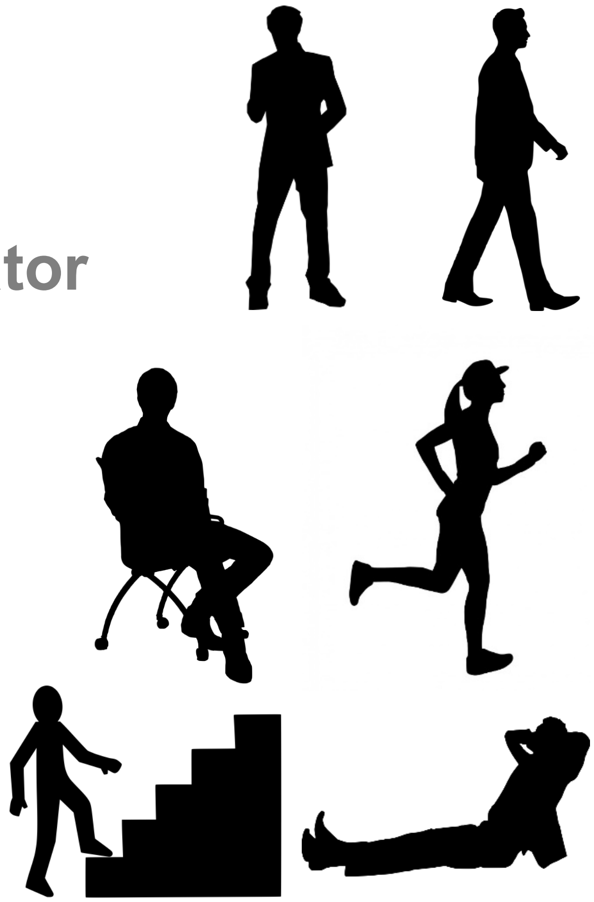
Power Gating

- **Insight from wearable applications domain**
  - Data collection and preprocessing have to be always ON
  - Processing blocks can be activated after the data is available
- **Major power savings potential by turning off processing pipeline**
- **Divide logic into two domains**
  - Segmentation, filtering, FIFO in *always-ON domain*
  - Downsample, feature generation and NN in *gated domain*
- **Use signal from segmentation to wake up**



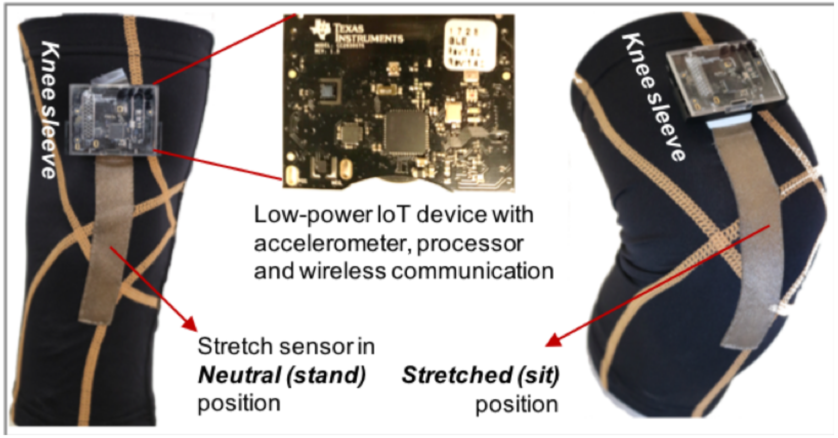
Outline

- Motivation
- Related Work
- Human Activity Recognition Accelerator
  – Baseline HAR Engine
  – Activity-Aware 2-Level HAR Engine
- Low Power Optimizations
- **Experimental Results**
- **Conclusion**



Experimental Setup

- **Design tools and hardware technology**
  - TSMC 65 nm LP
  - Cadence Innovus for APR
  - Synopsys PrimeTime for power
- **User studies**
  - Data from 22 users
  - Total of 4740 segments
- **Training data split**
  - 4 users for test
  - 18 users for training
    - 60% train, 20% cross-val, 20% test
  - 37% test data from unseen users

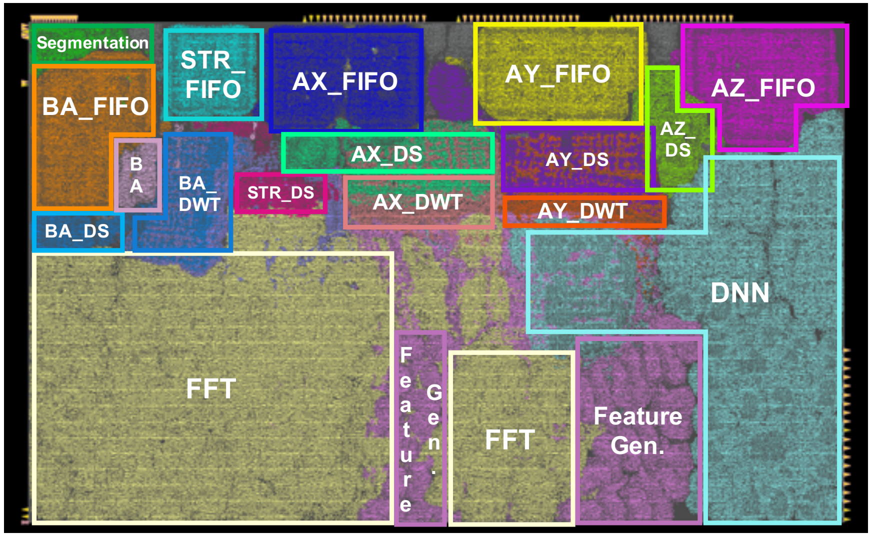


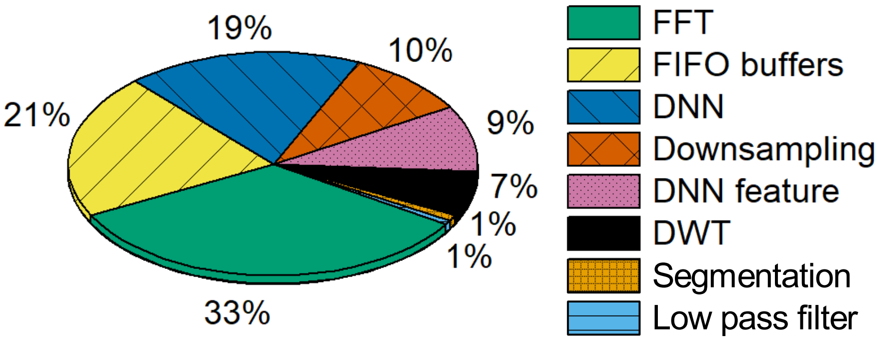
- **Data used in ESWEEK IoMT design contest**
  - 16 teams from 7 countries
- **Presentations on Tuesday 15th *12 pm to 1pm***
- **Data available open source**

Design Area: Baseline Engine

- **Synthesize at 100 kHz**
- **Floorplan during APR**
  - – Optimize to match logic
- **Total area = 1.353 mm^2**

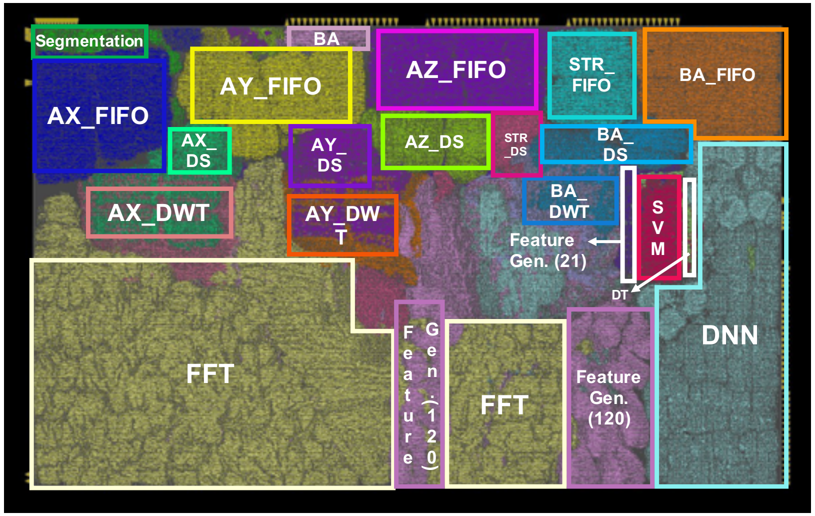
- **FFT has the highest area**
- **Blocks with memory have higher area**
  - – FIFO for storing samples
  - – Neural network

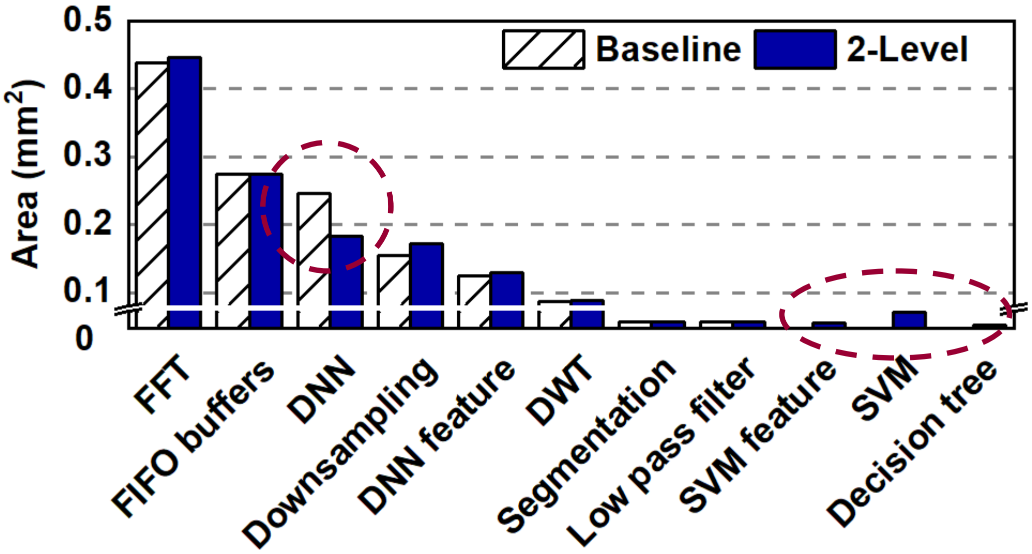


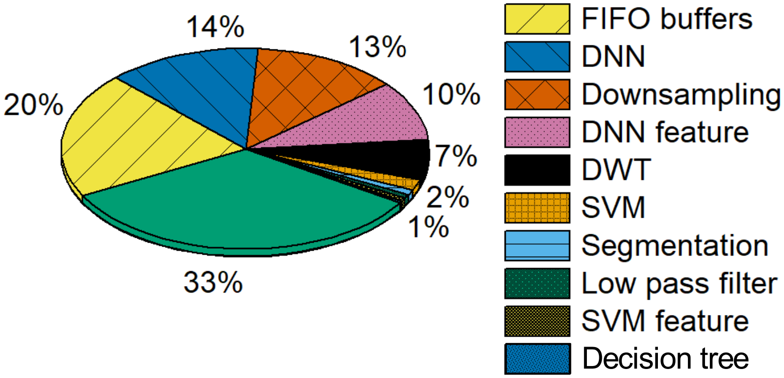


Design Area: 2-Level Engine

- **Total area = 1.357 mm²**
  - **Only 0.3% larger than the baseline design**
- **Resembles baseline design**
  - Processing blocks are common







Accuracy of the Baseline Engine

- **Weight and Activation Quantization to 16-bits**

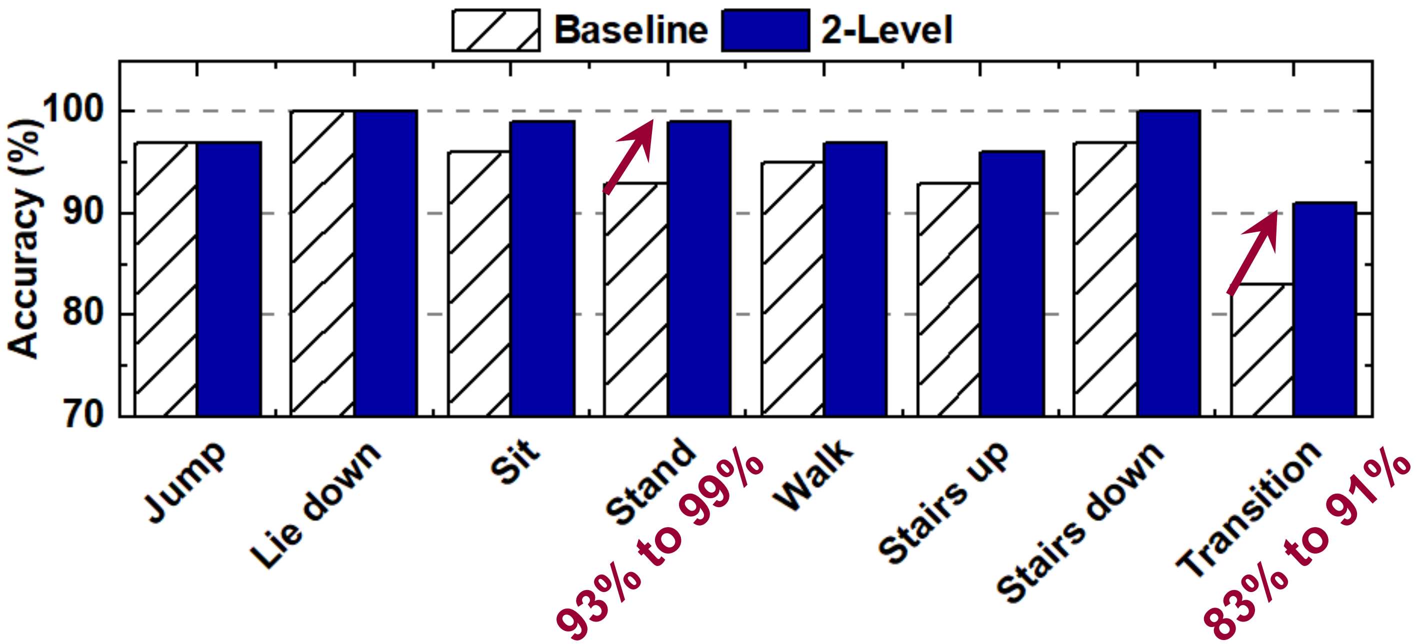
$$\Delta_q = \frac{2W_{max}}{2^{16}} \text{ where } W_{max}\text{: Largest weight}$$

- **Confusion matrix for baseline classifier**
  - Greater than 93% accuracy for all activities

	Jump	Lie Down	Sit	Stand	Walk	Stairs up	Stairs Down	Tran-sition	Accuracy (%)
Jump	**442**	0	0	0	5	0	5	6	**97**
Lie down	0	**474**	0	0	0	0	0	0	**100**
Sit	0	0	**665**	26	0	0	0	5	**93**
Stand	0	0	16	**576**	1	0	0	27	**93**
Walk	31	0	1	10	**1913**	0	10	42	**95**
Stairs up	0	0	0	0	1	**101**	6	1	**93**
Stairs down	0	0	0	0	1	1	**97**	1	**97**
Transition	7	2	7	14	14	4	0	**229**	**83**

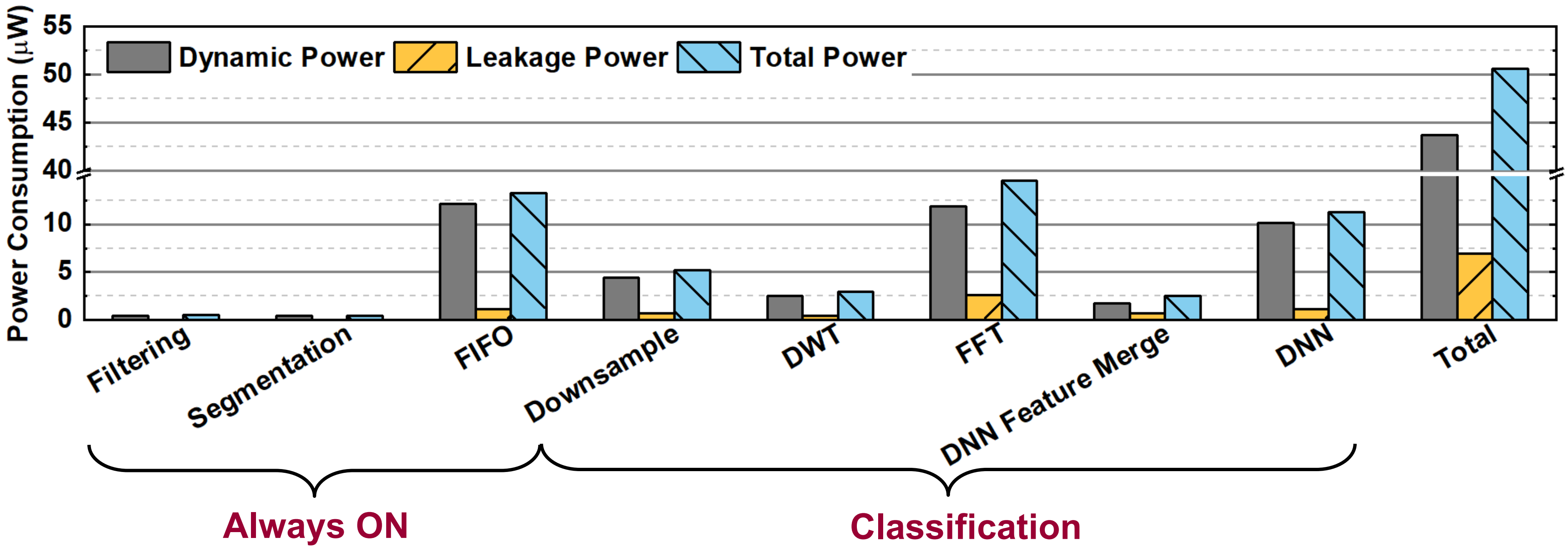
Accuracy of 2-Level Engine

- 99% accuracy in classifying static and dynamic activities
- Accuracy improvement with 2-Level engine



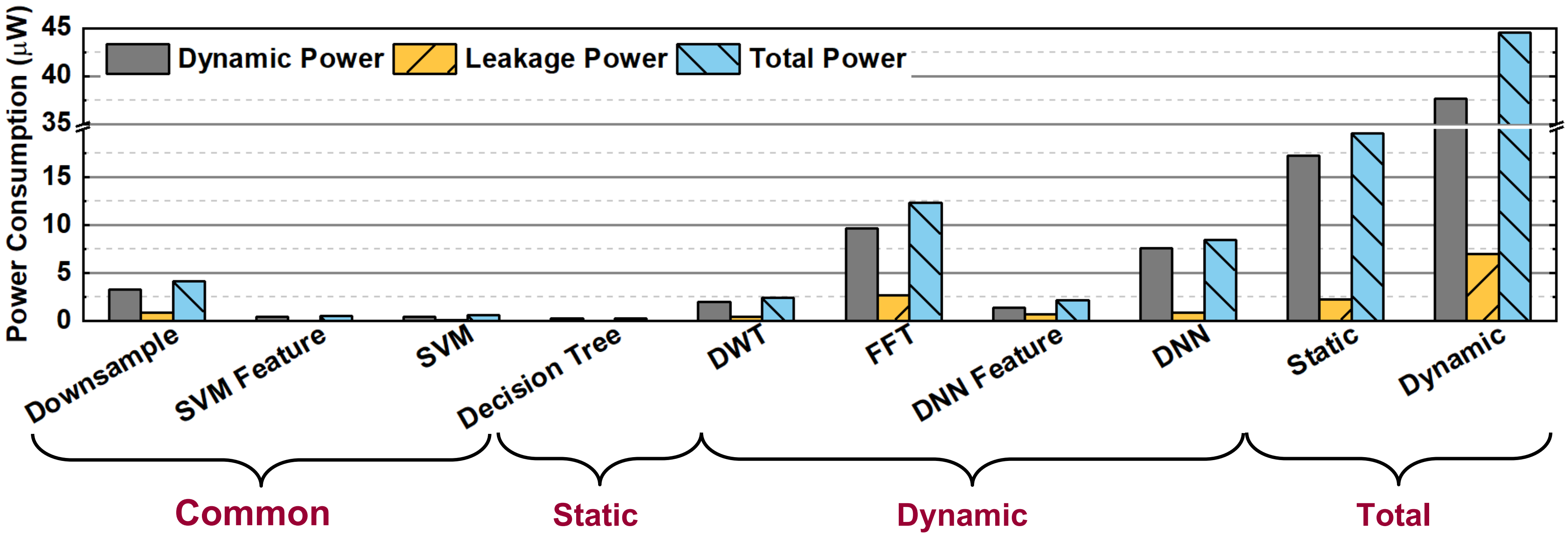
1% to 8% accuracy improvement with only 0.3% larger area

Power Consumption of Baseline Engine



- Always ON modules consume about 14 µW
- FFT has highest power among classification blocks
- Total power consumption of 51 µW

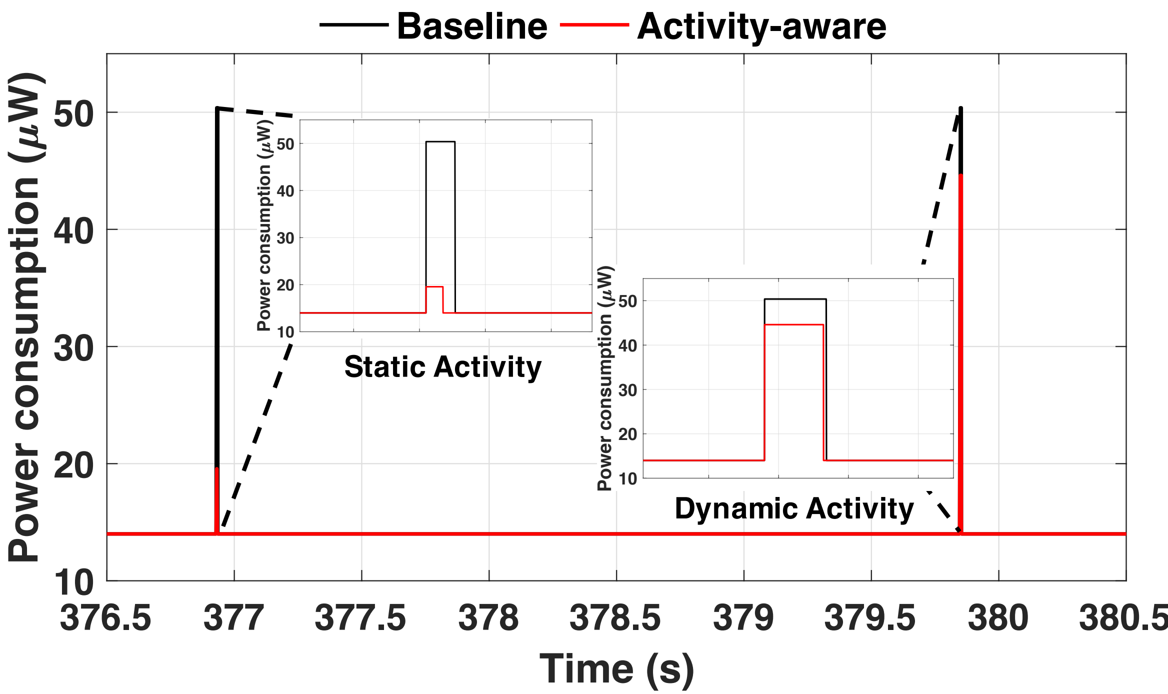
Power Consumption of 2-Level Engine



- **Static activities consume 19.5 μW (2.6× reduction)**
- **Dynamic activities consume 44.6 μW (1.14× reduction)**
- **10× improvement compared to embedded solutions**
  – Including sensor and communication energy
- **17 day operation using a 130 mAh flexible battery**

Peak Power Consumption Benefits

- **Our goal is to operate with ambient energy**
  – Peak power must be lower than energy harvesting capacity
- **More than 80% time spent in static activities**
  – Activity-aware engine provides lower peak power

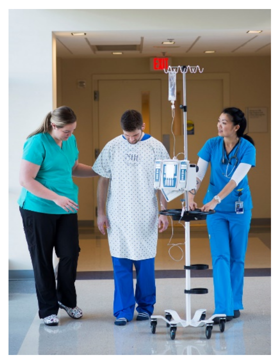


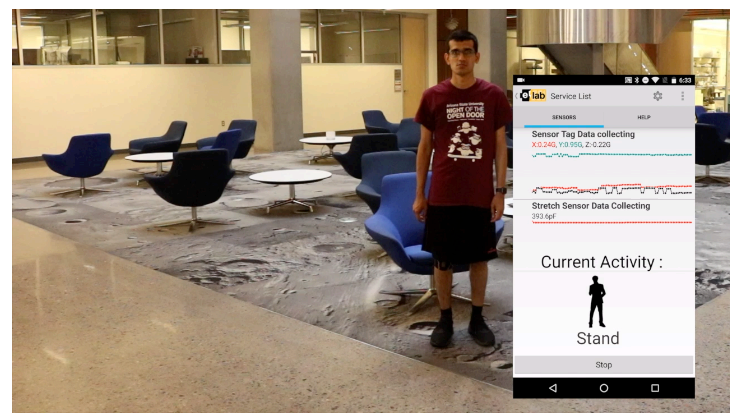
2.6× and 1.1× reduction in peak power for static and dynamic activities

Facilitates operation with ambient energy

Conclusion

- **Presented two human activity recognition engines**
  - Fully integrated solution from sensor to activity classification
  - Novel activity-aware engine
  - 22.4 µJ per activity using TSMC 65 nm LP
  - Further power savings possible with voltage scaling
- **Dataset from 22 users released to public**





A critical step towards *self-powered* healthy monitoring devices