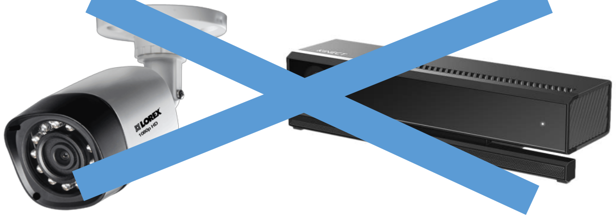
EAR: Exploiting Uncontrollable Ambient RF Signals in Heterogeneous Networks for Gesture Recognition

Zicheng Chi, Yao Yao, Tiantian Xie, Xin Liu, Zhichuan Huang, Wei Wang, and Ting Zhu

University of Maryland, Baltimore County

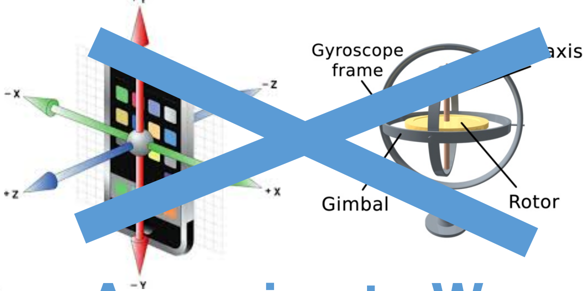
Human Activity Recognition System in the Market

Camera-based



Privacy Leakage

Sensor-based



Annoying to Wear

Voice-based

Can't be used:

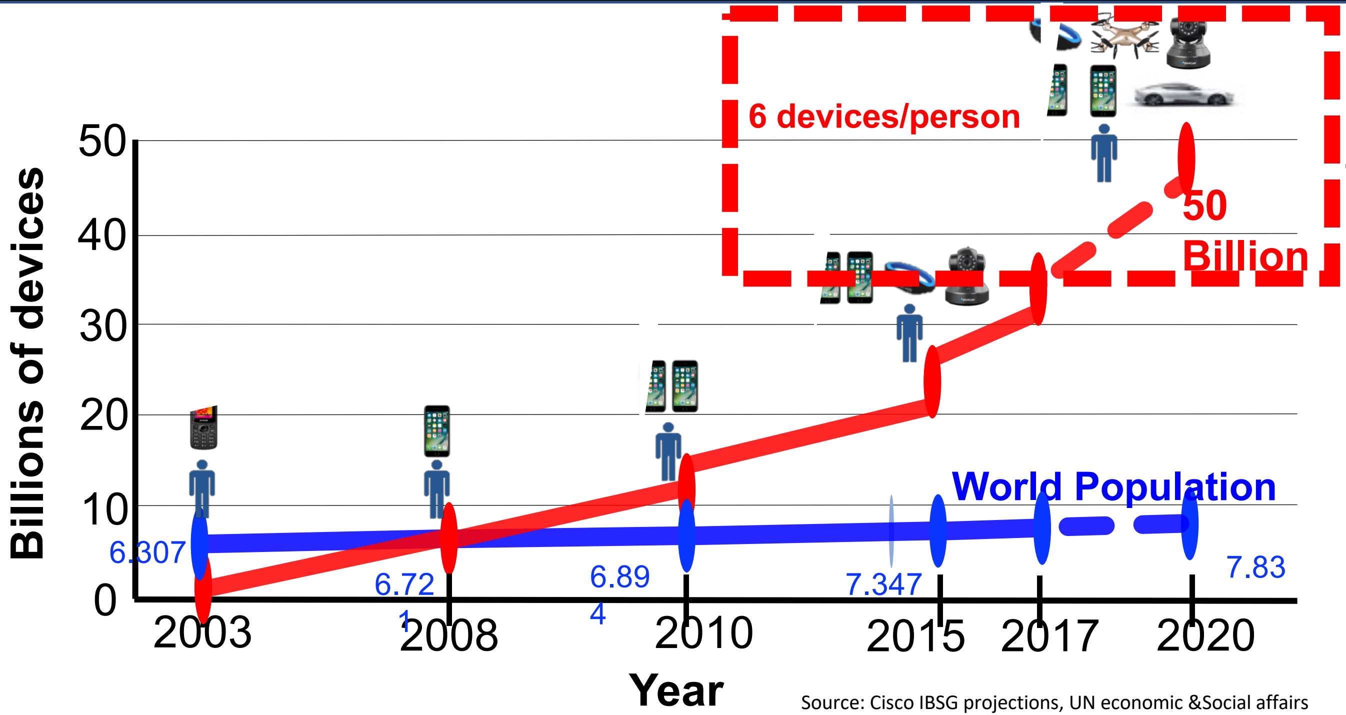
- baby room
- noisy environment

Human Activity Recognition System in Academia

- CARM Mobicom'15
- Mudra CoNEXT'16
- E-eyes MobiCom'14
- Vital-Radio CHI'15
- Etc.

- Need designated sender and traffic.
- Only work within the same wireless protocol.

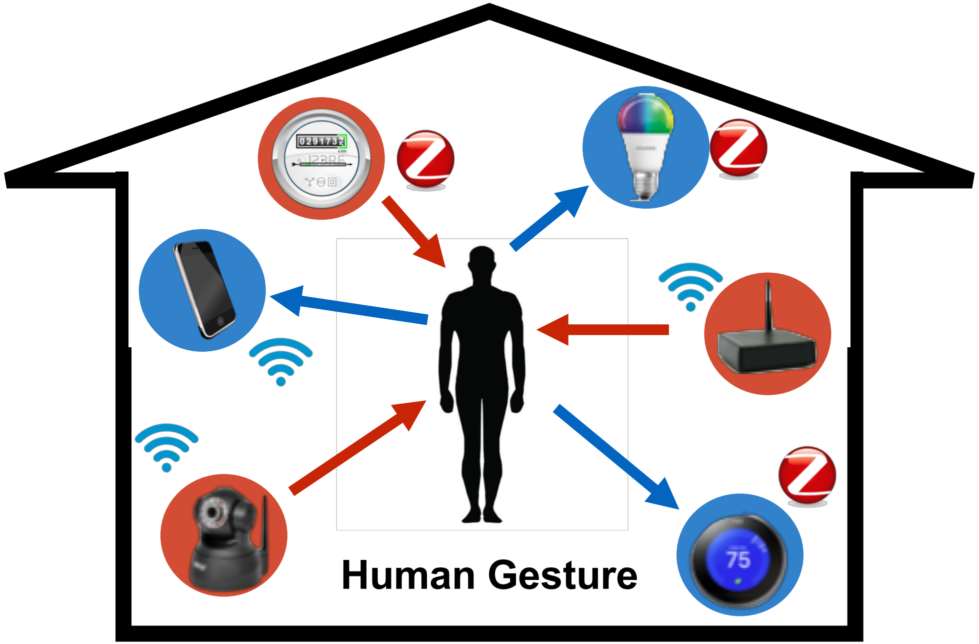
Exponentially Increasing Number of IoT Devices



Opportunities for RF sensing

Source: Cisco IBSG projections, UN economic &Social affairs

EAR: Leveraging existing IoT devices for human gesture recognition



***Uncontrollable RF signal sources**: we have no access to the RF signal source, which transmits on its own schedule.*

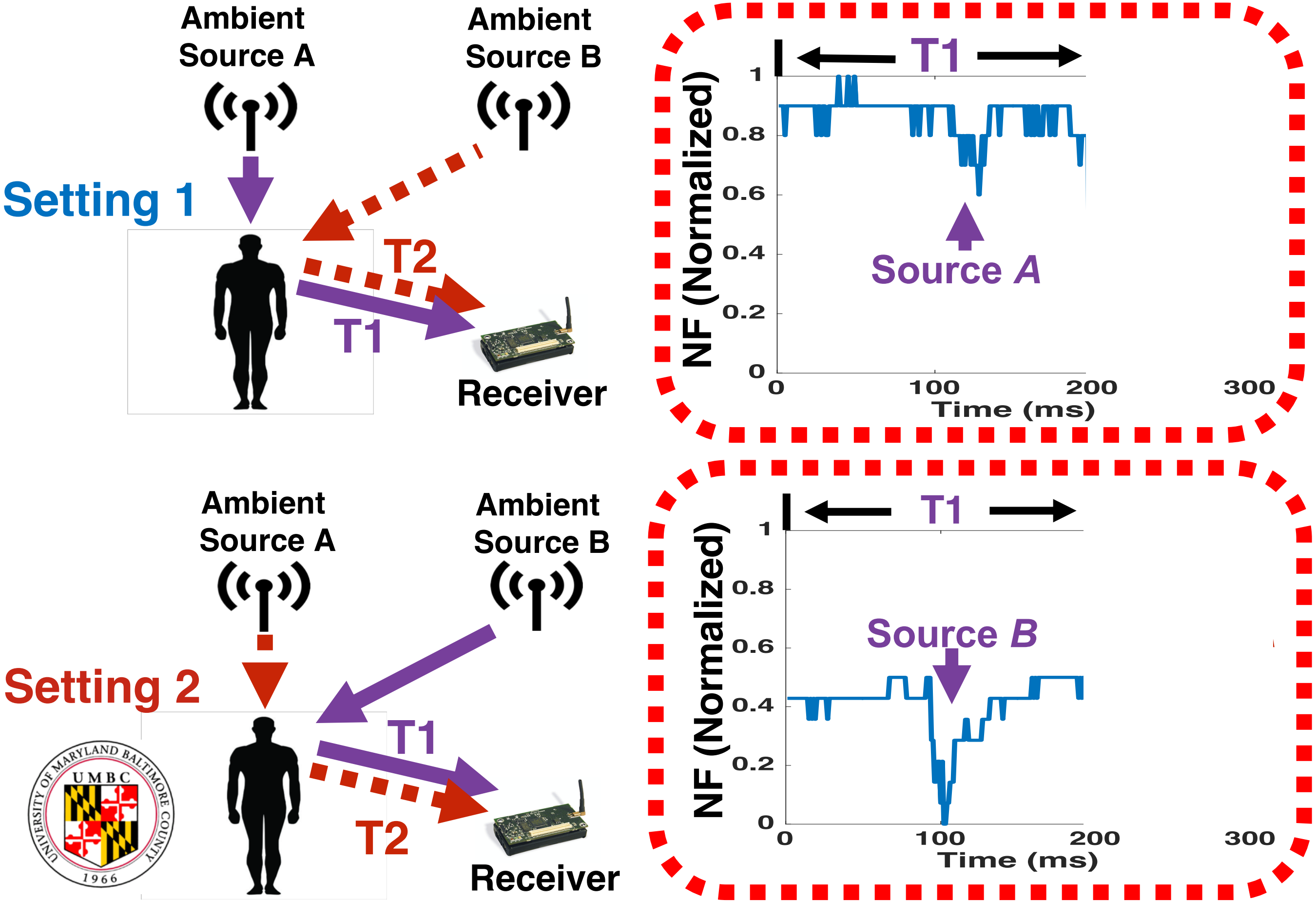
To achieve this goal, we have three challenges:

1. How to distinguish signals from heterogenous IoT devices?

2. How to deal with Intermittent RF traffics?

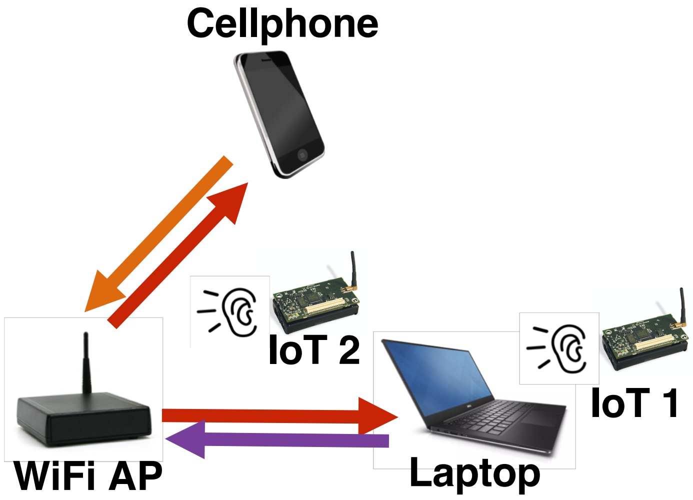
3. How to align the measurements from distributed devices without introducing extra traffic?

Why we need to distinguish signal sources?

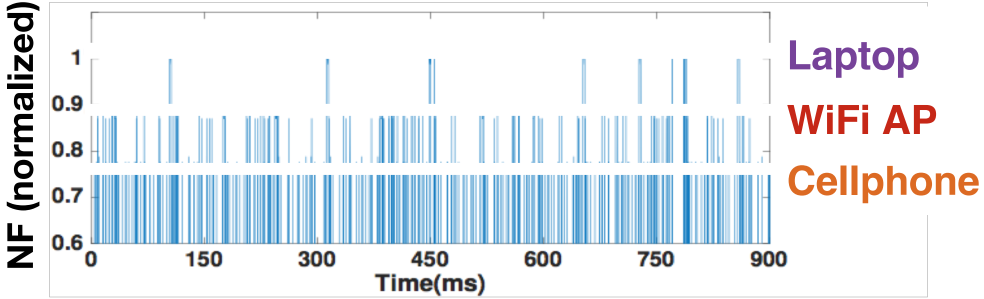


These two pieces of signals ***cannot be*** characterized by the same model.

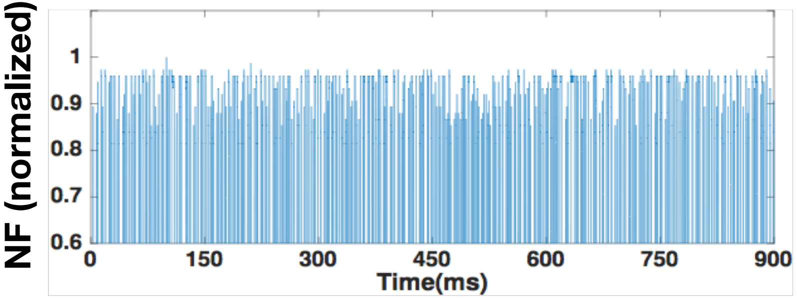
The basic idea to distinguish signal sources



(a) Deployment



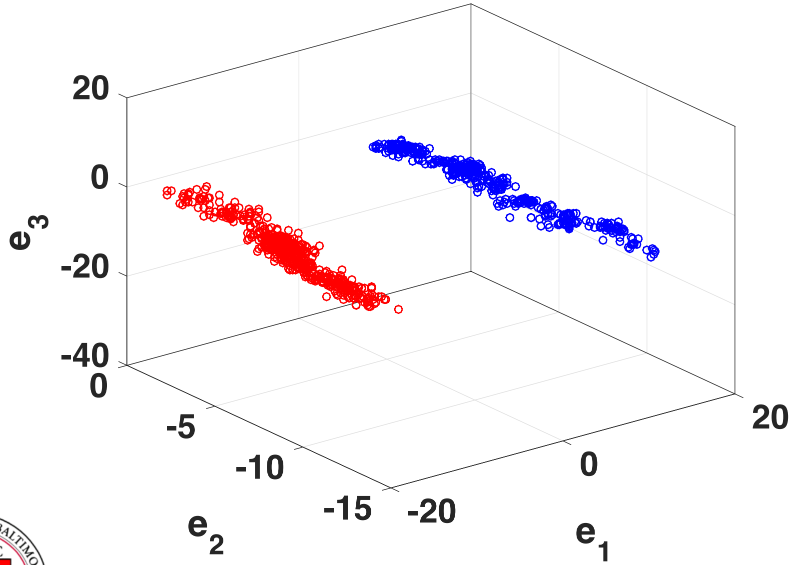
(b) NF from IoT 1

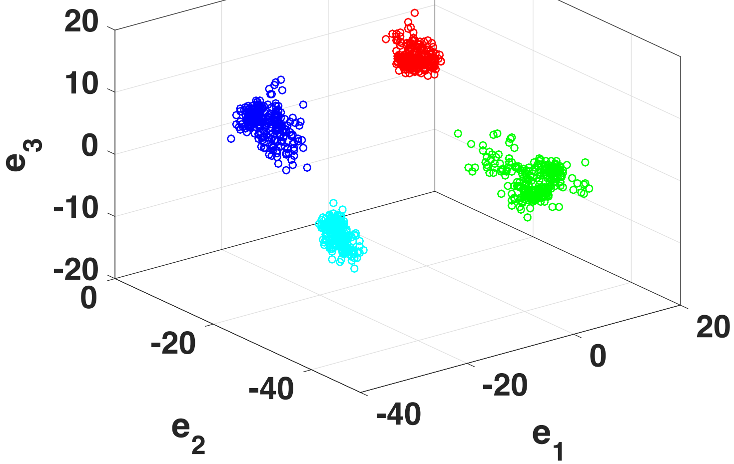


(c) NF from IoT 2

Continuous Signal Source Distinguishing (CSSD) Algorithm

Ref. Algorithm 1 in the paper. CSSD distinguishes the sources based on the diversities among the NF measurements across different IoT receivers.





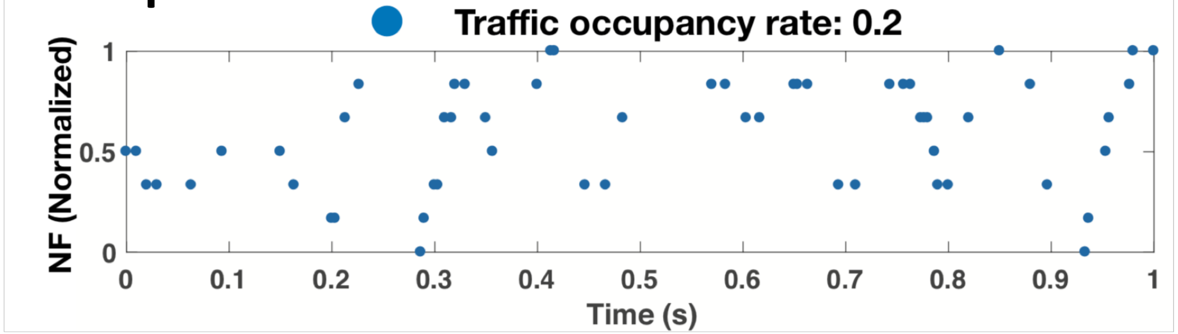
To achieve this goal, we have three challenges:

1. How to distinguish signals from heterogenous IoT devices?
2. How to deal with intermittent RF traffics?
3. How to align the measurements from distributed devices without introducing extra traffic?

UNIVERSITY OF MARYLAND BALTIMORE COUNTY UMBC 1966

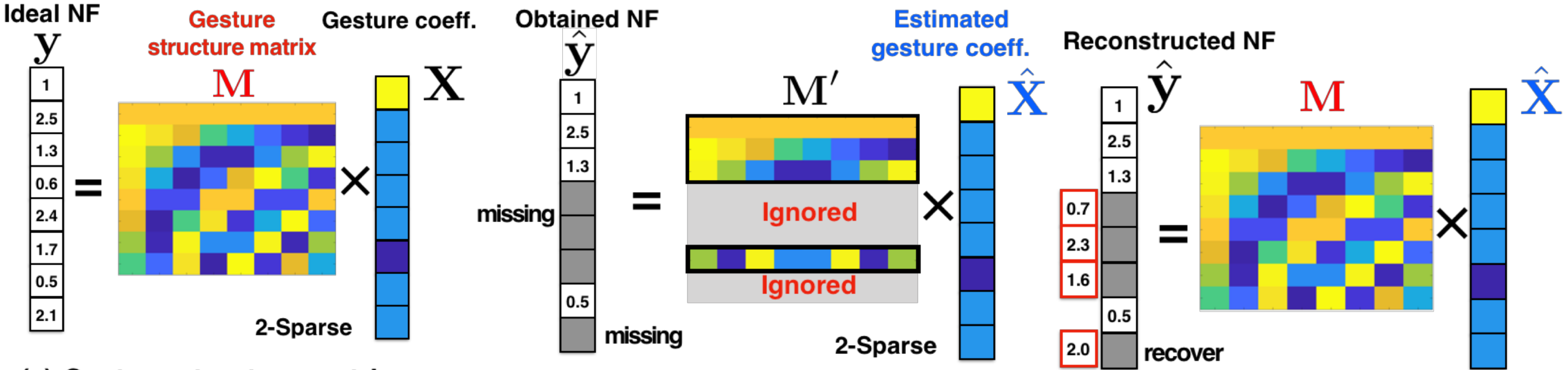
Intermittent traffic reconstruction

- **The gesture recognition requires a continuous signal measurements to characterize the human motion.**
- **However, the uncontrollable IoT devices have intermittent traffic patterns.**



- **To address this challenge, we reconstruct the signal by using compressive sensing technique.**

Intermittent traffic reconstruction



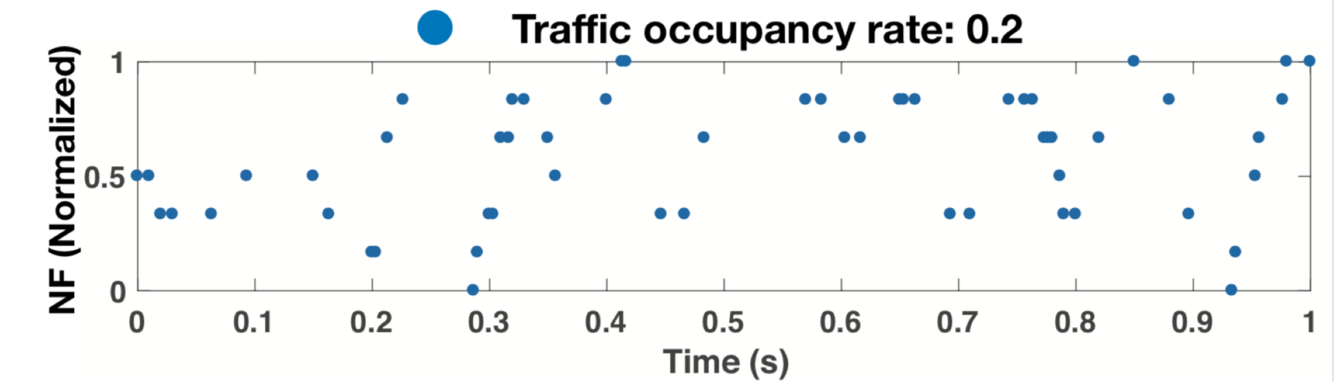
(a) Gesture structure matrix decomposition

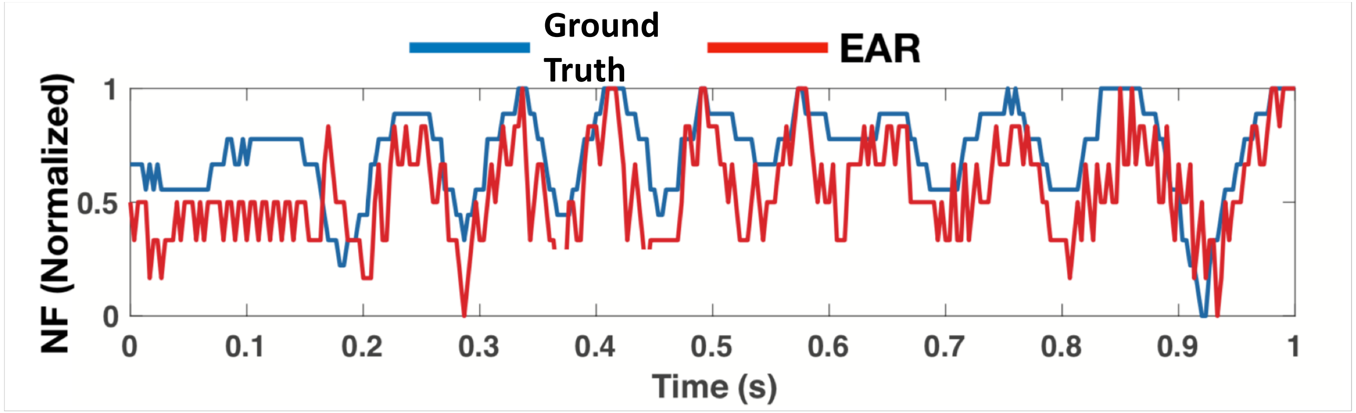
(b) Gesture coeff. estimation

(c) Missing NF samples reconstruction

$$\min_{\hat{\mathrm{x}} \in R^N} ||\hat{\mathrm{x}}||_1 \quad s.t. \; \hat{\mathrm{y}} = \mathrm{M}'\hat{\mathrm{x}}$$

Intermittent traffic reconstruction

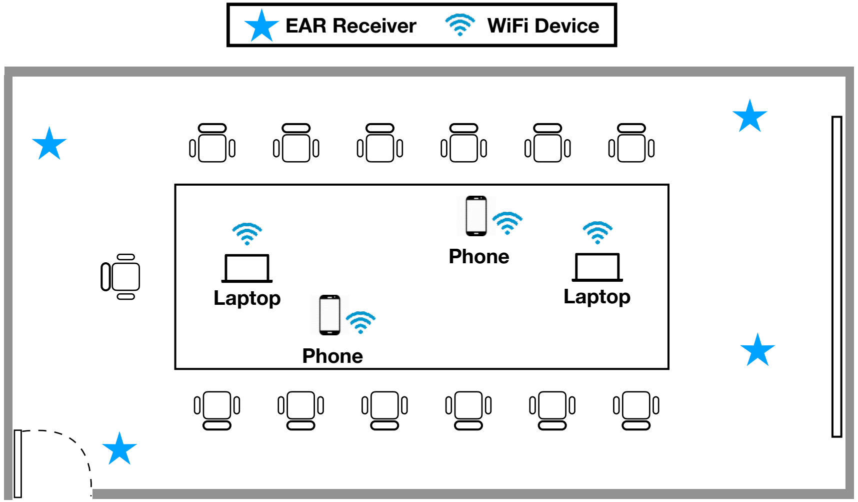




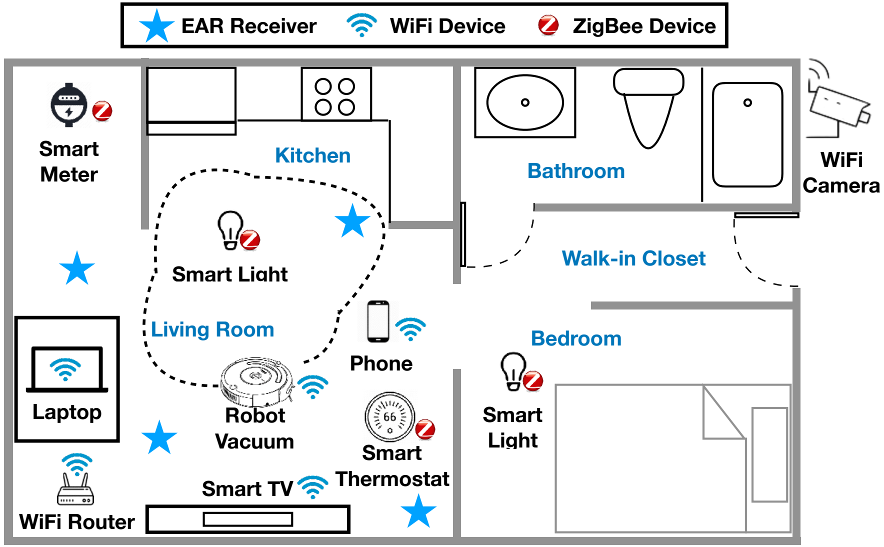
To achieve this goal, we have three challenges:

1. How to distinguish signals from heterogenous IoT devices?
2. How to deal with Intermittent RF traffics?
3. How to align the measurements from distributed devices without introducing extra traffic?

Experimental Setup



A Meeting Room



An Apartment

Experimental Setup

Gesture	Notation	Gesture	Notation
Slight Kick	KK	Pull	PL
Wave	WE	Swipe	SP
Punch	PN	Raise Arms	RA
Circle	CE	Stand Still	ST
Push	PH		

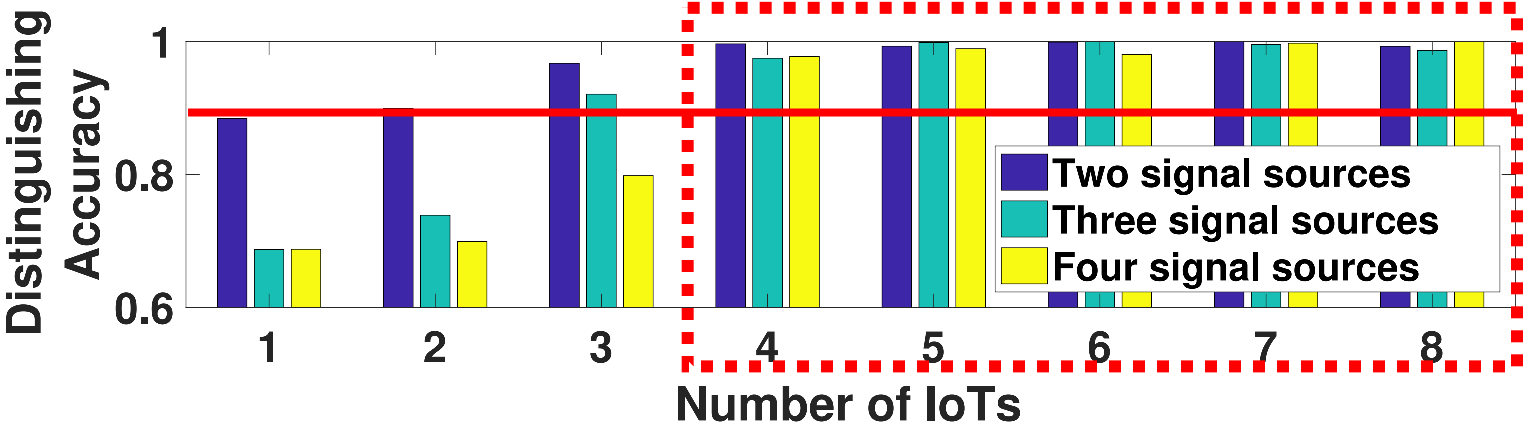
We have *10 volunteers* to perform *9 different gestures* for the evaluation. We collected *11,250 samples* in total.

Recognition Confusion Matrix

	KK	WE	PN	CE	PH	PL	SP	RA	ST
KK	88.0	0.0	0.0	0.0	1.0	11.0	0.0	0.0	0.0
WE	1.0	92.0	0.0	0.0	4.0	2.0	0.0	1.0	0.0
PN	5.9	4.0	83.2	1.0	0.0	5.0	0.0	1.0	0.0
CE	0.0	0.0	0.0	97.0	0.0	3.0	0.0	0.0	0.0
PH	0.0	1.0	0.0	1.0	96.0	2.0	0.0	0.0	0.0
PL	0.0	0.0	0.0	0.0	4.9	95.1	0.0	0.0	0.0
SP	0.0	1.0	4.9	1.9	2.9	0.0	89.3	0.0	0.0
RA	1.0	3.0	0.0	0.0	2.0	1.0	0.0	93.0	0.0
ST	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0

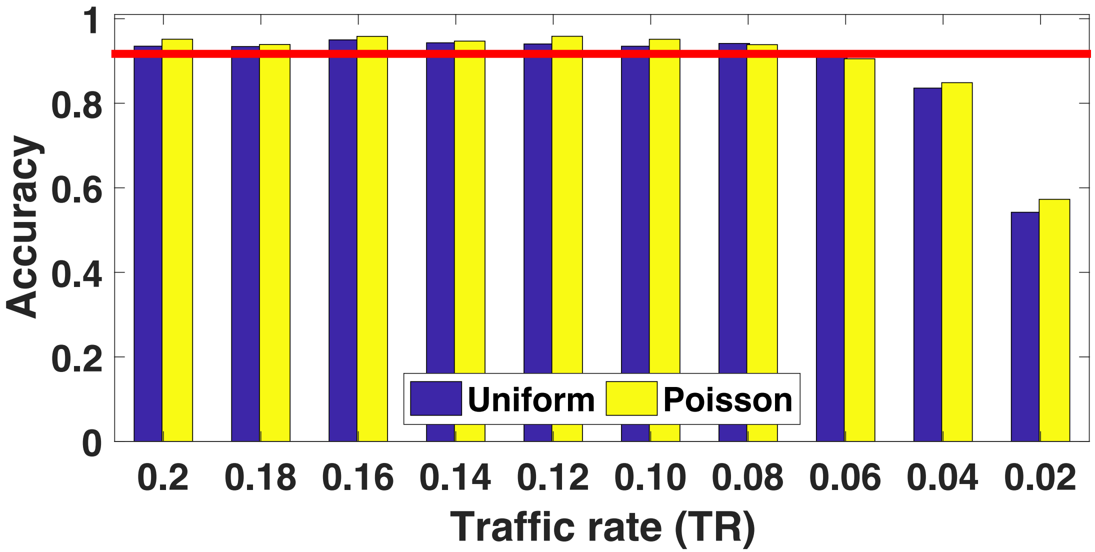
The average accuracy is 92.63%

Signal Source Distinguishing Accuracy



- Two signal sources, higher than 90% accuracy.
- With four receivers, EAR can easily distinguish all sources (up to 99.76% accuracy).

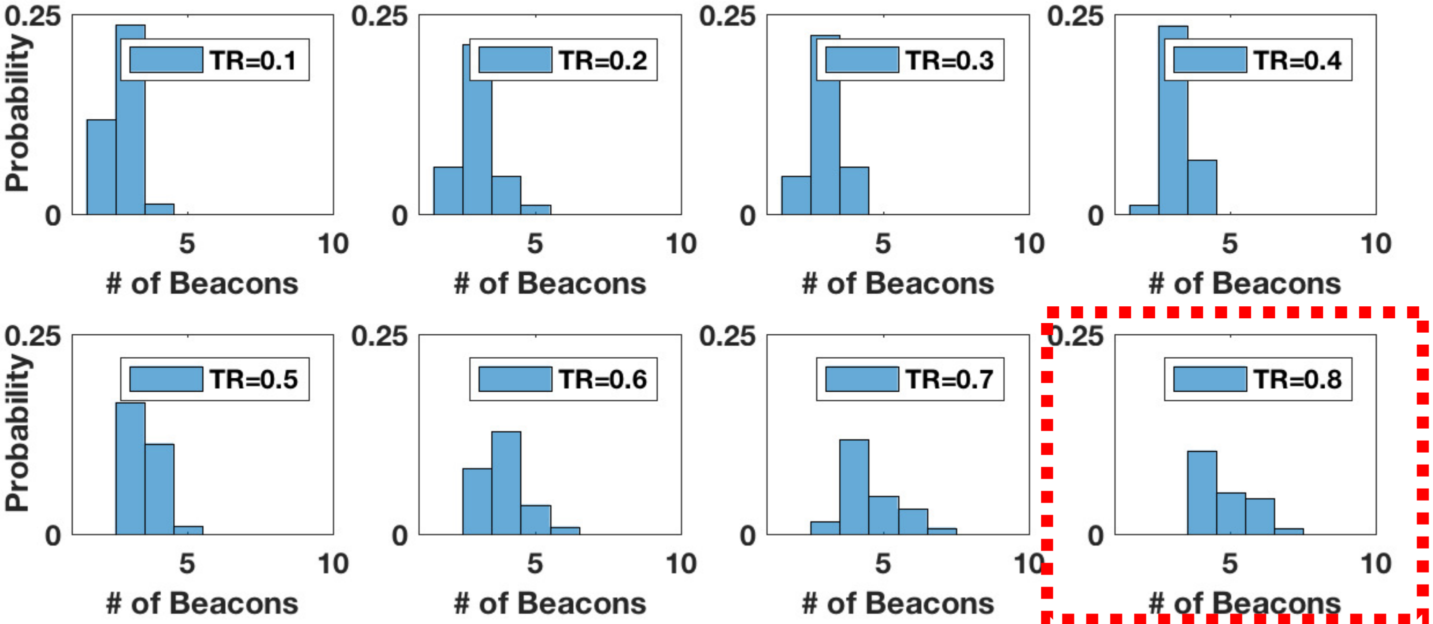
Accuracy comparison under different traffic rate



The overall gesture recognition accuracy is higher than 90.0% when the traffic rate is larger than 0.04.

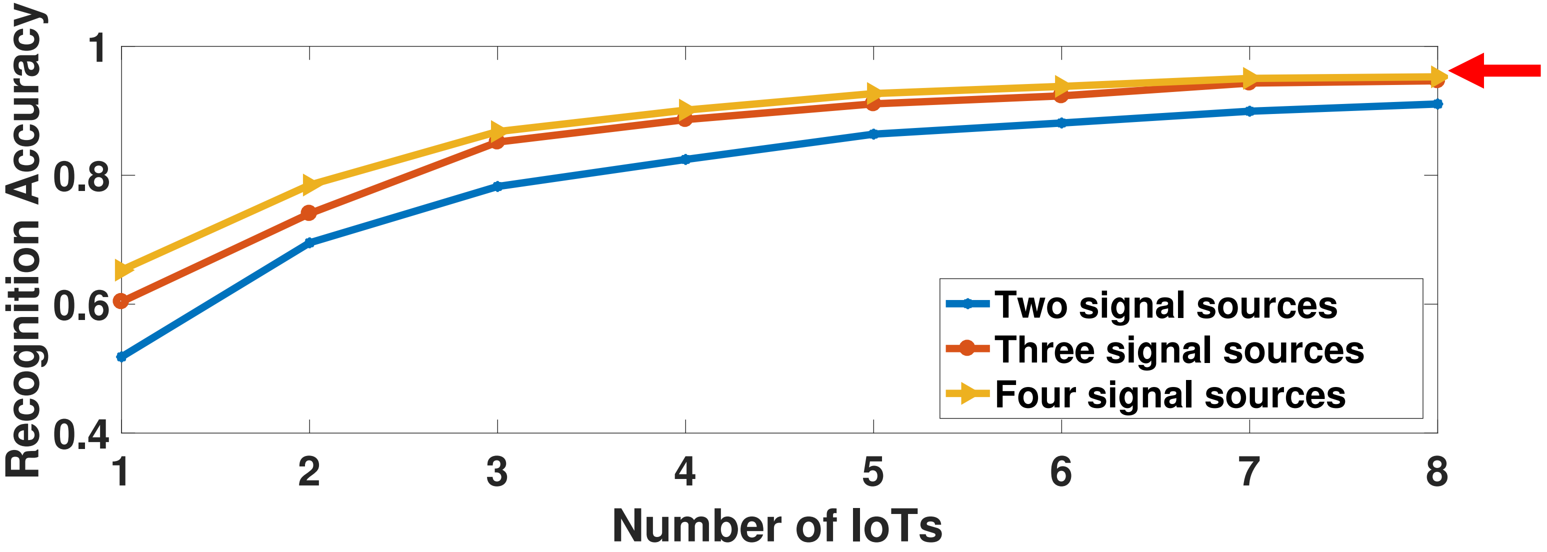
Traffic rate (TR): channel busy time divided by total time

Alignment Efficiency v.s. Traffic Rate



- The busier the traffic, the more periods of beacons is required to align distributed devices.
- However, with 0.8 TR, the mean periods are only 5.

Impact of signal sources' and receivers' number



- Recognition accuracy increases when the number of signal sources and receivers increases.

Conclusion

- EAR enables the ambient RF signals from IoT devices for human gesture recognition without introducing any designated RF sensing traffic.

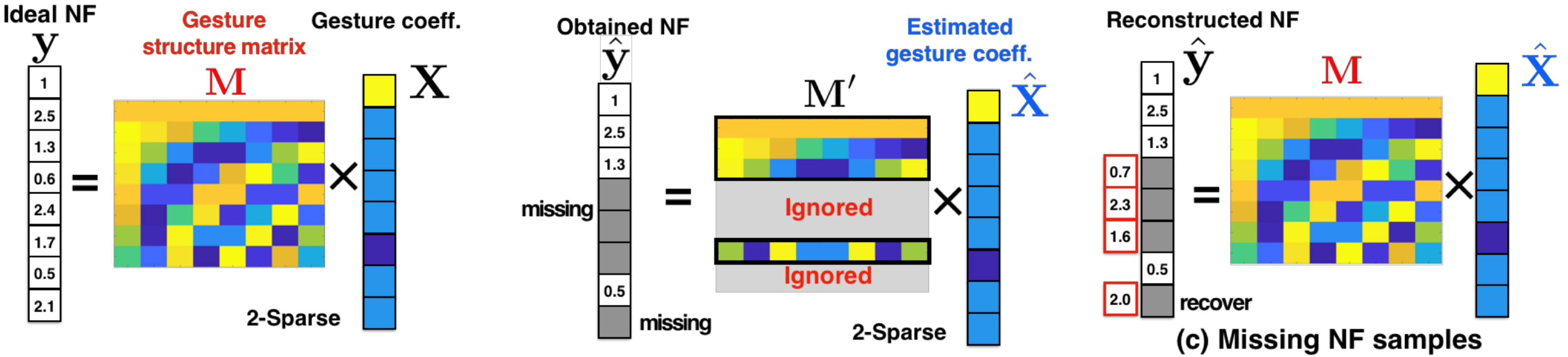
- This is also the first work that seamlessly combines multiple pieces of ambient RF signals for gesture recognition.

- The design of the signal source distinguishing and intermittent RF traffic reconstruction modules are generic. It is possible to extend EAR to be a middleware.

Q & A

Thanks!

Intermittent traffic reconstruction



(a) Gesture structure matrix decomposition

(b) Gesture coeff. estimation

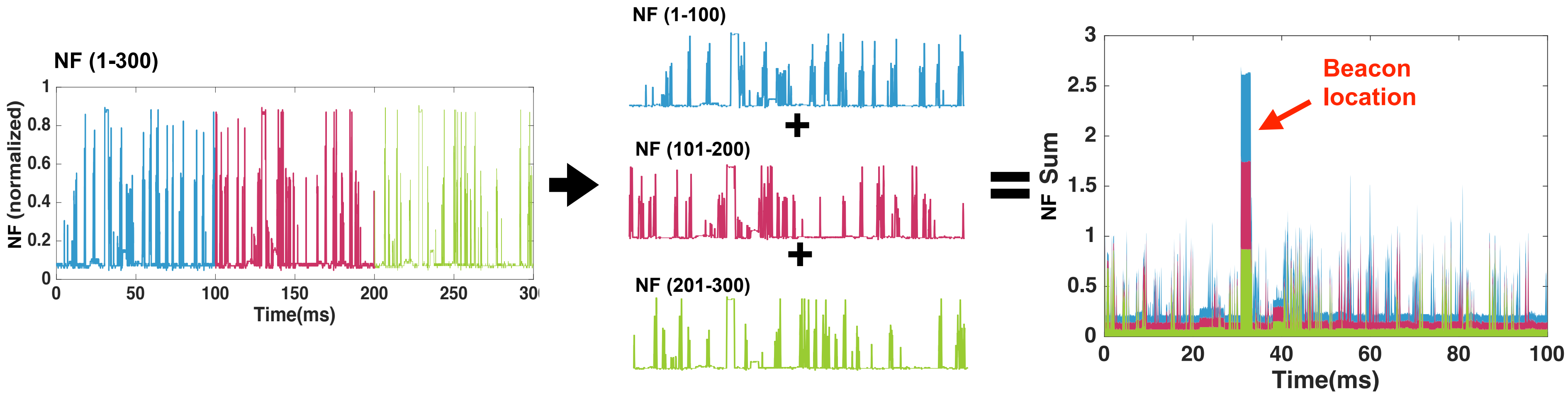
(c) Missing NF samples reconstruction

$$\min_{\hat{\mathbf{x}} \in \mathbf{R}^N} ||\hat{\mathbf{x}}||_1 \quad s.t. \; \hat{\mathbf{y}} = \mathbf{M}'\hat{\mathbf{x}}$$

M is a discrete cosine transform (DCT) matrix.

$$\mathbf{M}(i,j) = \begin{cases} \frac{1}{\sqrt{N}} & i = 0, 0 \le j \le N-1 \\ \sqrt{\frac{2}{N}} cos \frac{i \cdot (2j+1)\pi}{2N} & 1 \le i \le N-1, 0 \le j \le N-1 \end{cases}$$

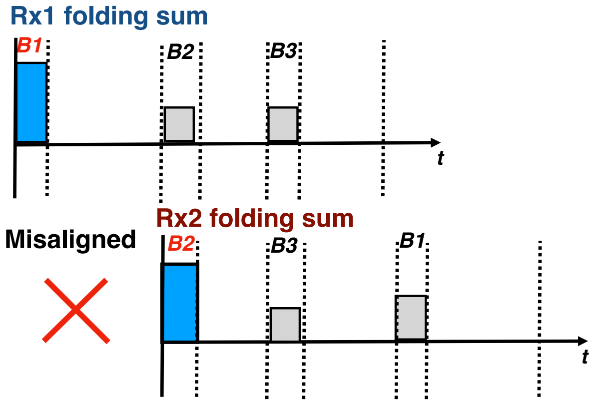
Using folding technique to align the measurements among distributed devices



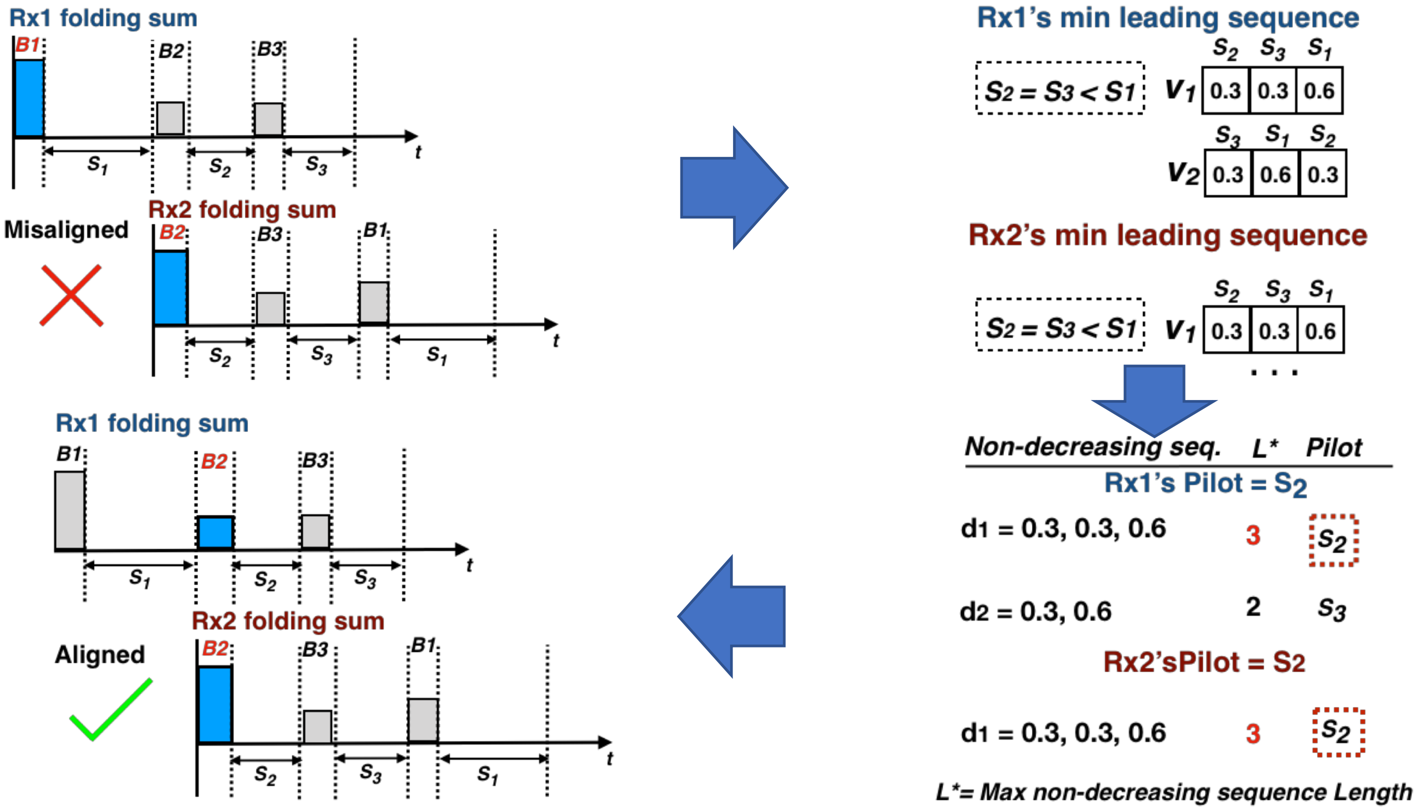
[1] R Lovelace, J Sutton, and E Salpeter. 1969. Digital Search Methods for Pulsars. In Nature.

[2] T. Hao, R. Zhou, G. Xing, and M. Mutka. 2011. WizSync: Exploiting Wi-Fi Infrastructure for Clock Synchronization in Wireless Sensor Networks. In 2011 IEEE 32nd Real-Time Systems Symposium

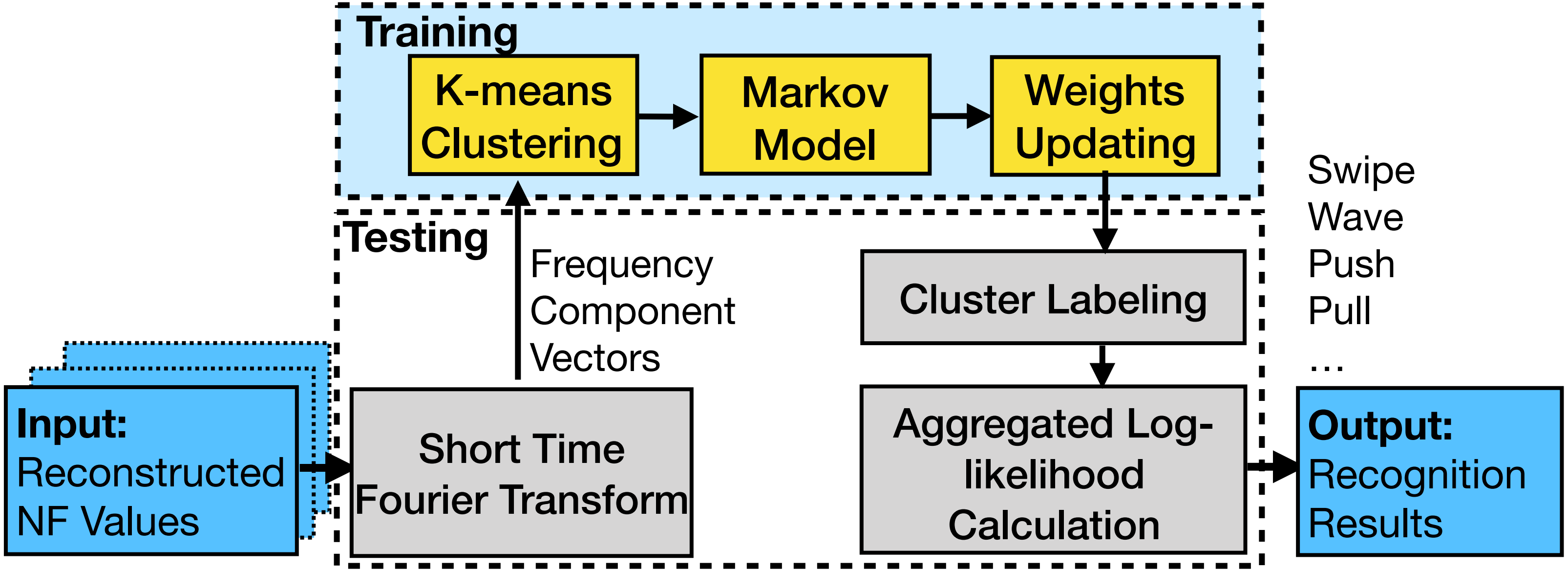
If multiple beacons have the same or integer multiple beacon periods.



Advanced Aligning Approach ?animation



Recognition Steps

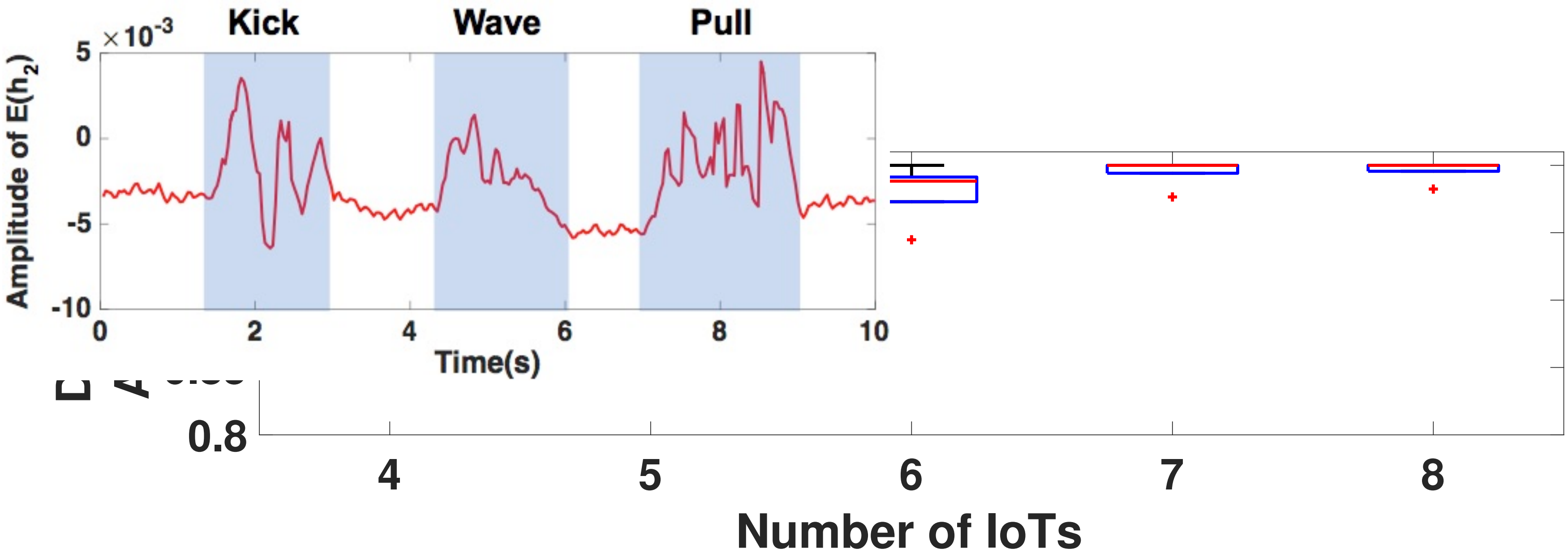


Recognition Confusion Matrix (meeting room)

	KK	WE	PN	CE	PH	PL	SP	RA	ST
KK	86.0	1.0	0.0	1.0	2.0	10.0	0.0	0.0	0.0
WE	0.0	95.0	0.0	0.0	2.0	2.0	0.0	1.0	0.0
PN	5.0	3.0	88.1	0.0	1.0	3.0	0.0	0.0	0.0
CE	0.0	1.0	1.0	97.0	0.0	1.0	0.0	0.0	0.0
PH	0.0	2.0	0.0	0.0	92.1	5.0	0.0	1.0	0.0
PL	1.0	1.0	0.0	0.0	3.9	94.1	0.0	0.0	0.0
SP	1.0	1.0	7.8	1.0	1.0	0.0	88.3	0.0	0.0
RA	0.0	3.0	1.0	0.0	5.0	1.0	0.0	90.0	0.0
ST	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0

The overall accuracy is 92.2%

Gesture Detection Accuracy



- Recognition accuracy increases when the number of signal sources and receivers increases.