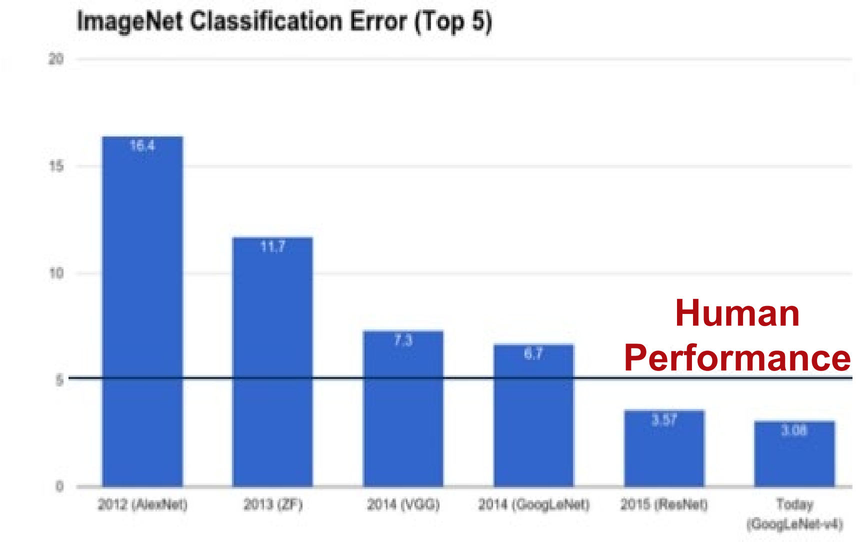
Thinking About the Technology Platform for Next-generation AI

Naveen Verma (nverma@princeton.edu),
Prakhar Kumar, Yoni Mehlman, Yue Ma, Murat Ozatay, Akash Pattnaik, Can Wu,
Prof. Sigurd Wagner, Prof. James Sturm

San Diego IEEE SSCS
April 8, 2021

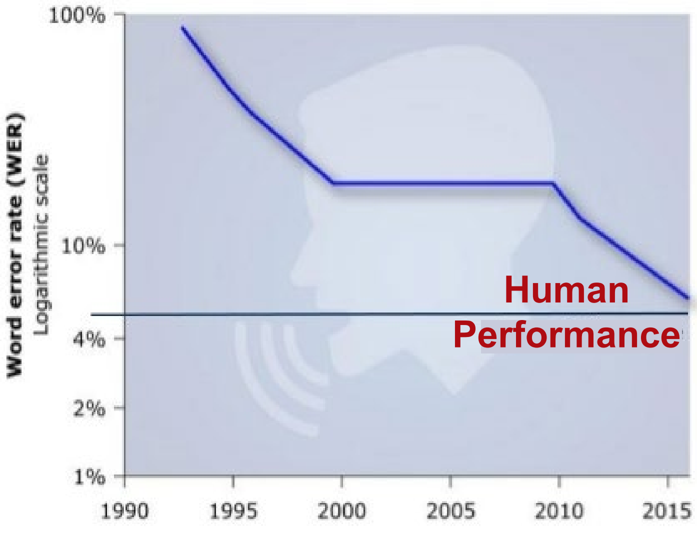
Machine capability is at an inflection point

Vision



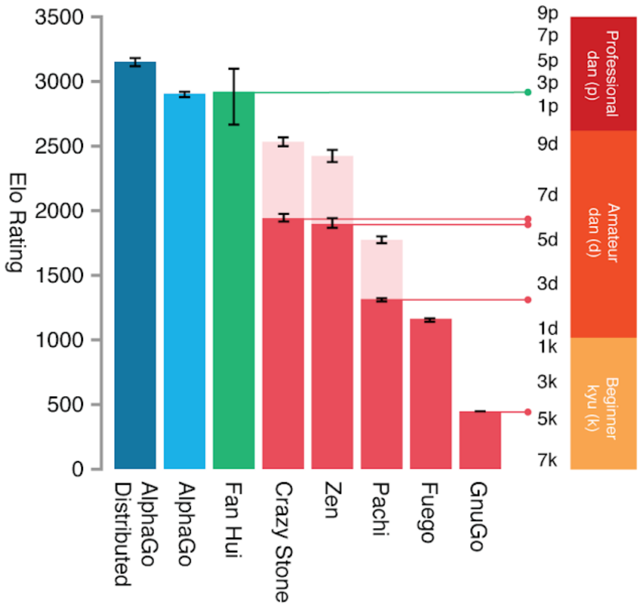
ImageNet: The "computer vision World Cup"

Speech



Deep Learning in Speech Recognition

Game Play



But, real-world deployment is challenging

A young boy is holding a baseball bat

You go first.
No, you go first.

DARPA Robotics Challenge

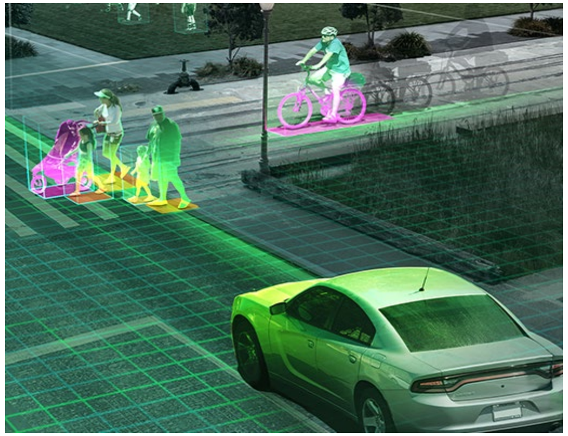
What will going forward require?
What can technology developers do?

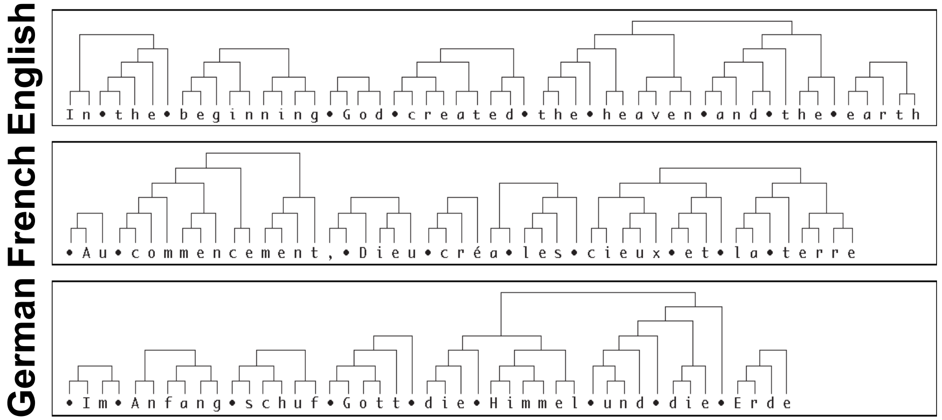
ML is about modeling how semantics are encoded in data

Images (SPATIAL structure)

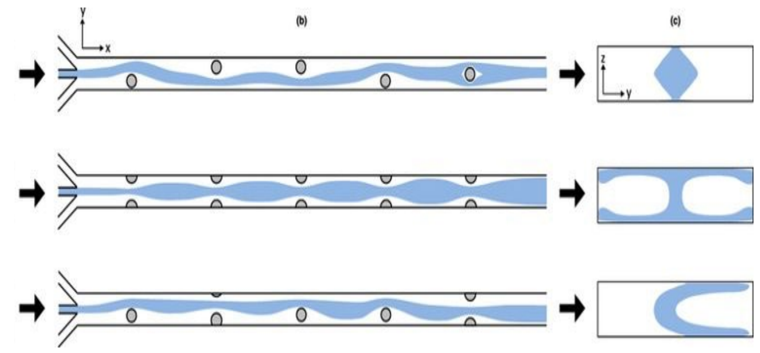
Language (SEQUENTIAL structure)

Inverse problems (PHYSICS)

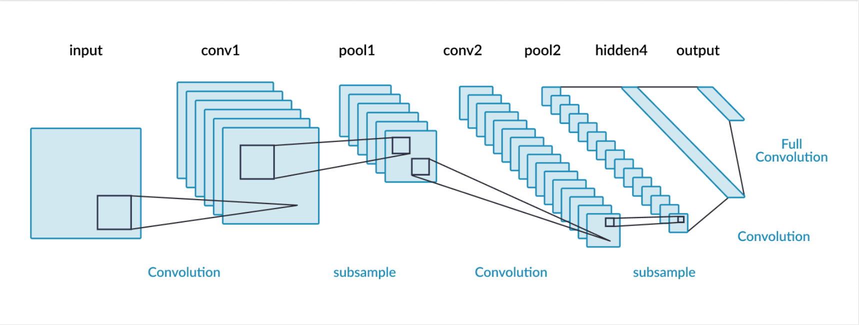




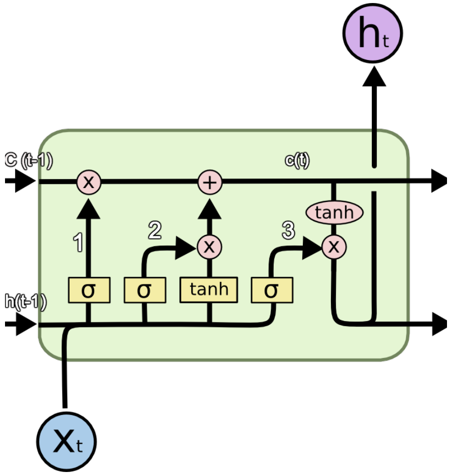
Ex. Fluid flow:



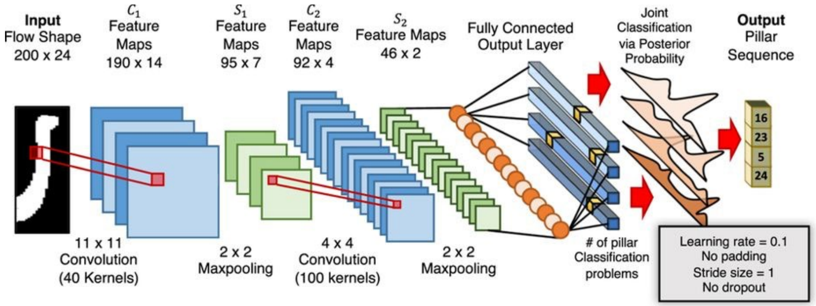
E.g., Convolutional Neural Net. (CNN)



E.g., Long-Short-Term Memory (LSTM)



E.g., Flow sculpting [D. Stoecklein, *Nature*'17]



STRUCTURE in data helps us build 'better' models

Introducing structure & multiple-modalities in sensor data

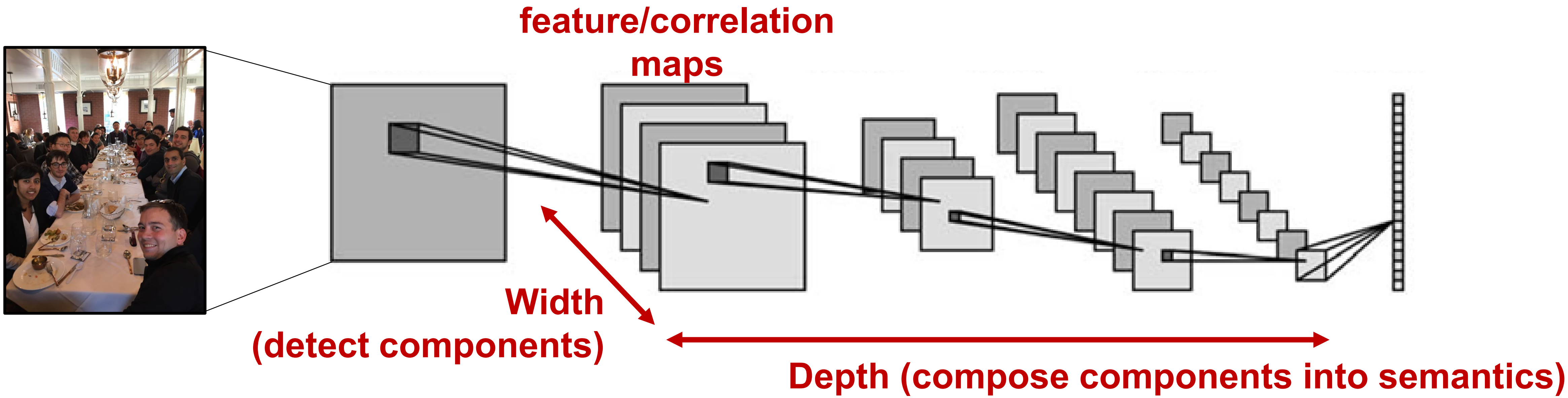
PHYSICALLY-INTEGRATED (PI) SENSING: the state and (inter)actions of physical objects says something about the activities and underlying intentions

→ structure data around states and interactions of ‘things’

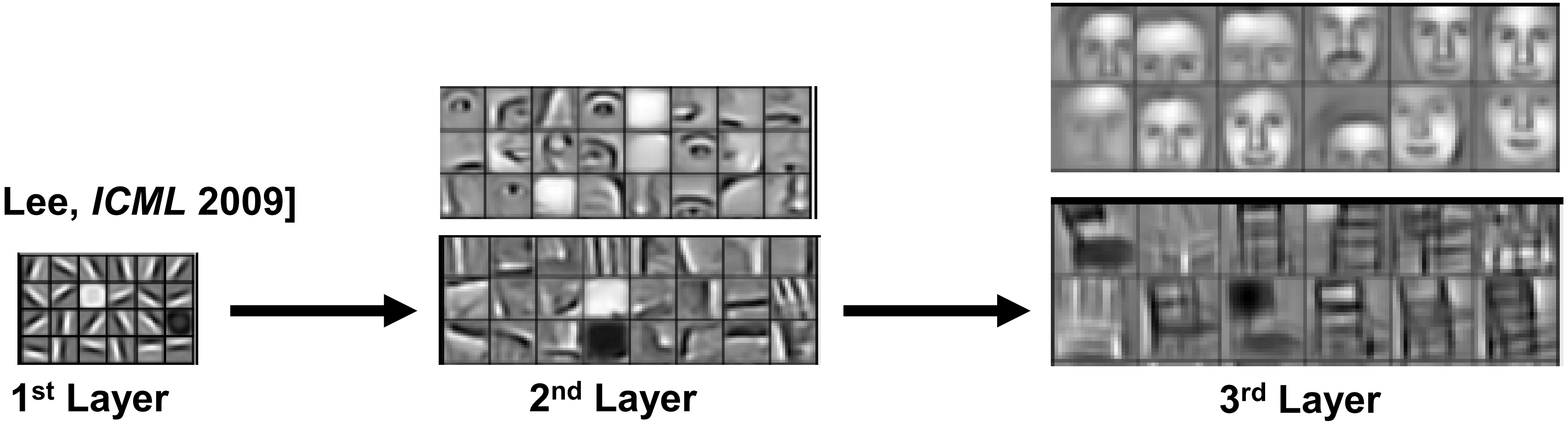
Directly associating sensor data with embedded signals enables invariant semantic structure & access to diverse modalities

Contrast: PI vs. remote sensing (vision)

Today's deep learning:



[H. Lee, *ICML* 2009]

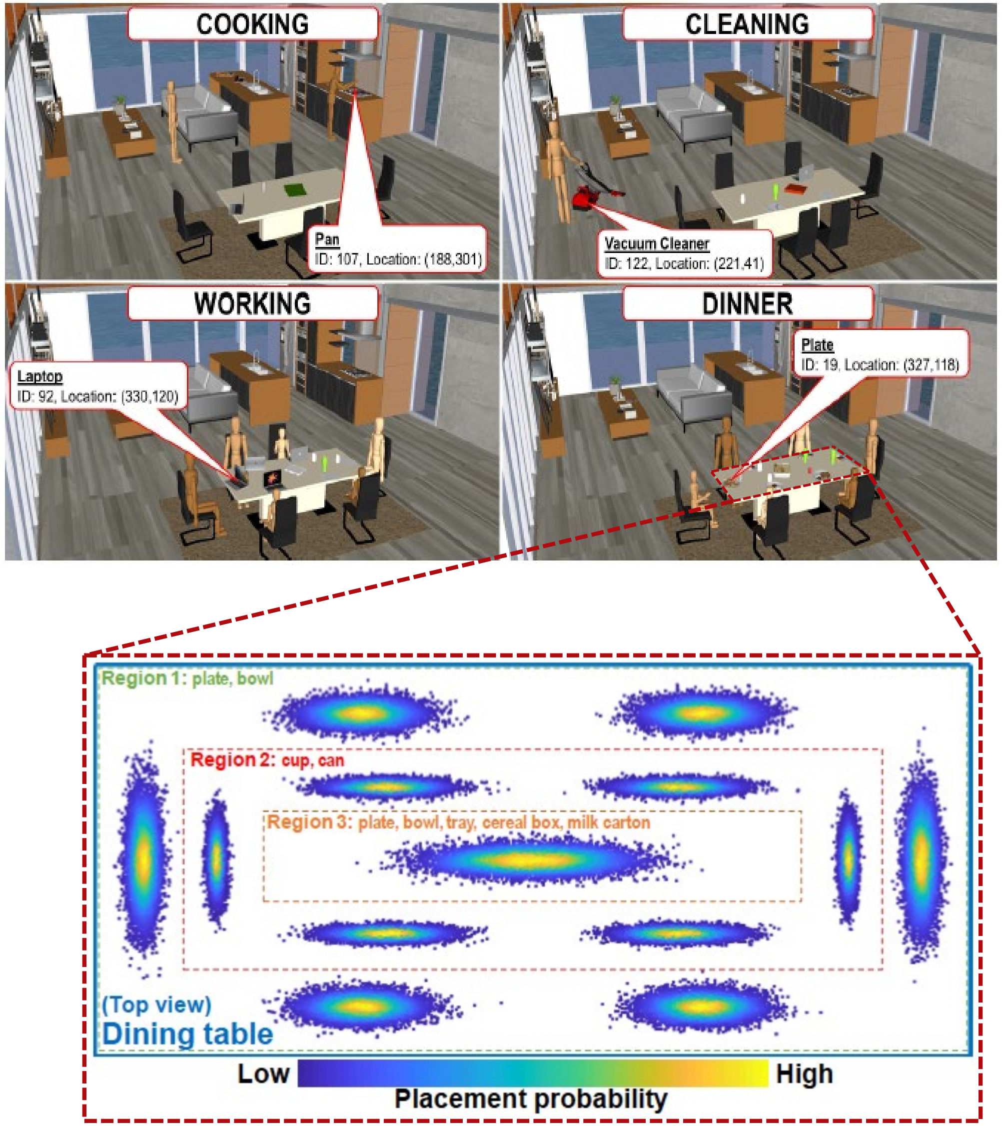


Some questions

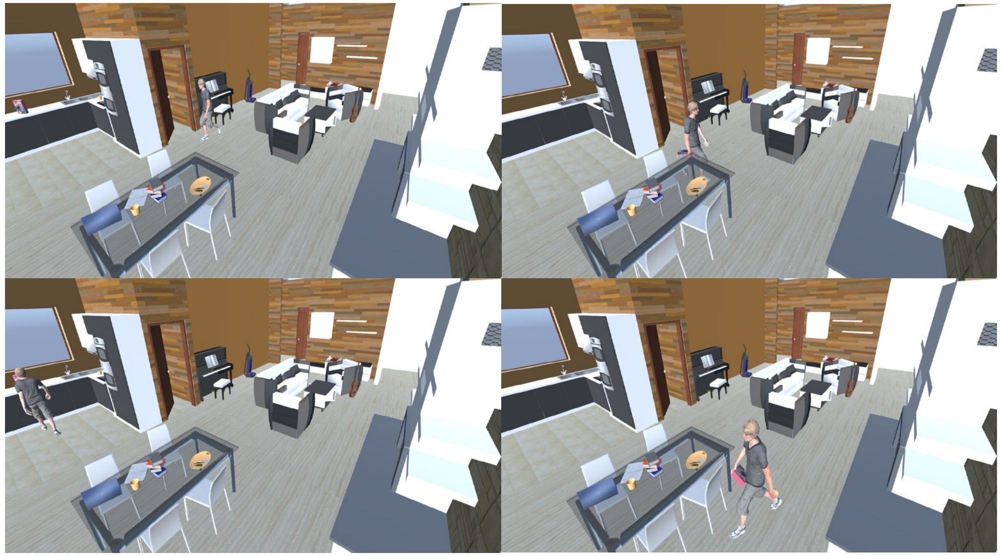
1. What kind of structure is relevant and how much is needed?
   - Synthesized human-activity scenes & actions emulate different forms of sensing
   - Enable studies on generalization, sample efficiency, transferability
2. What models (and training algorithms) exploit that structure?
   - Evaluate sensor-specific features and embeddings for efficiency
   - Explore models for sensor fusion (including with remote sensing)
3. What sensing technologies preserve/provide such structure
   - Develop large-scale, form-fitting (wireless) sensing based on large-area electronics
   - Develop architectures for in-sensor computing of features/embeddings
4. What computational architectures do these require?
   - Structure in data → structure in models/computations → architectural specialization
   - In-memory computing architectures (won't go into this today)

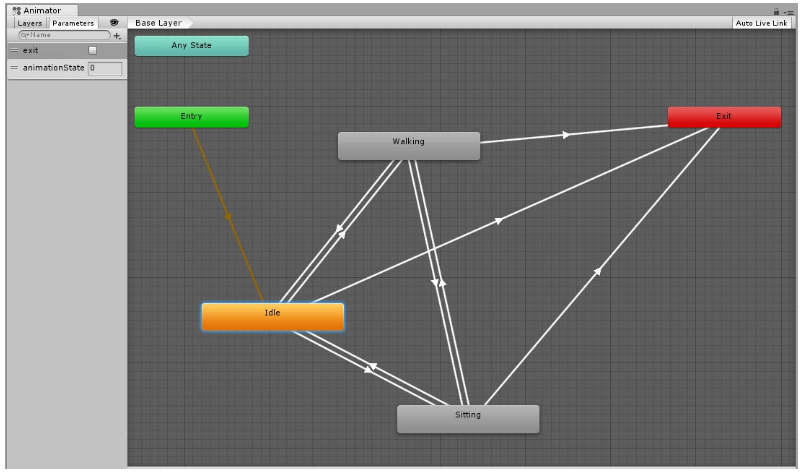
Monte Carlo synthesis of human activities

SketchUp (scenes)

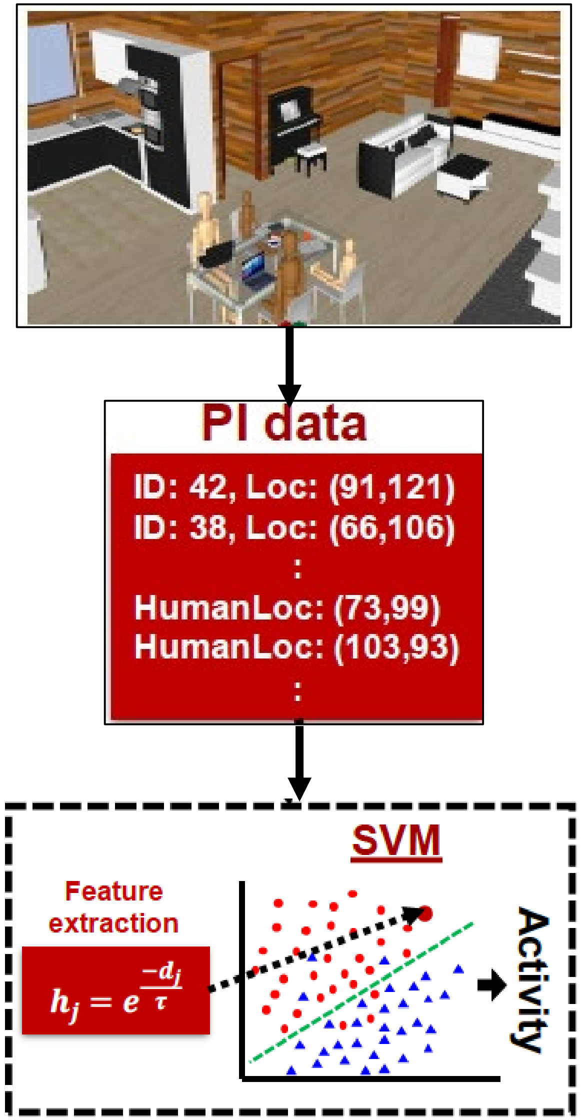


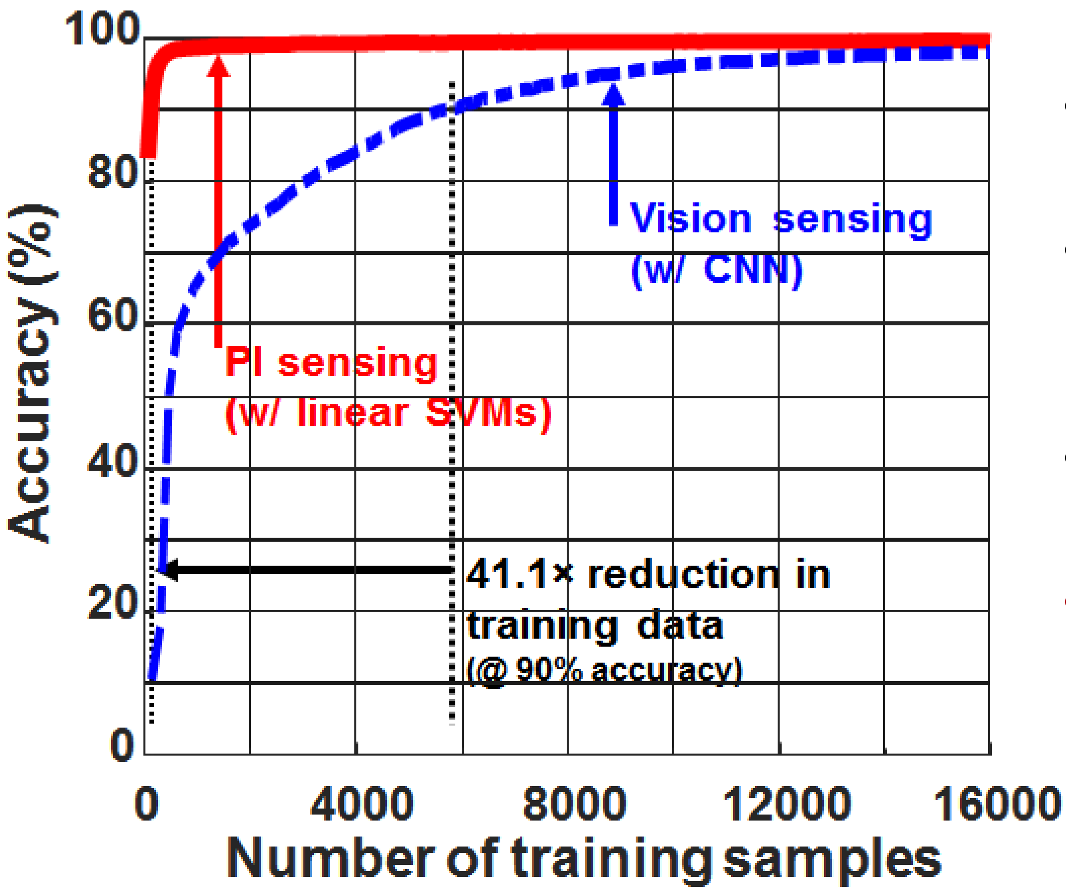
Unity3D (actions)





How does PI-sensing affect sample efficiency?

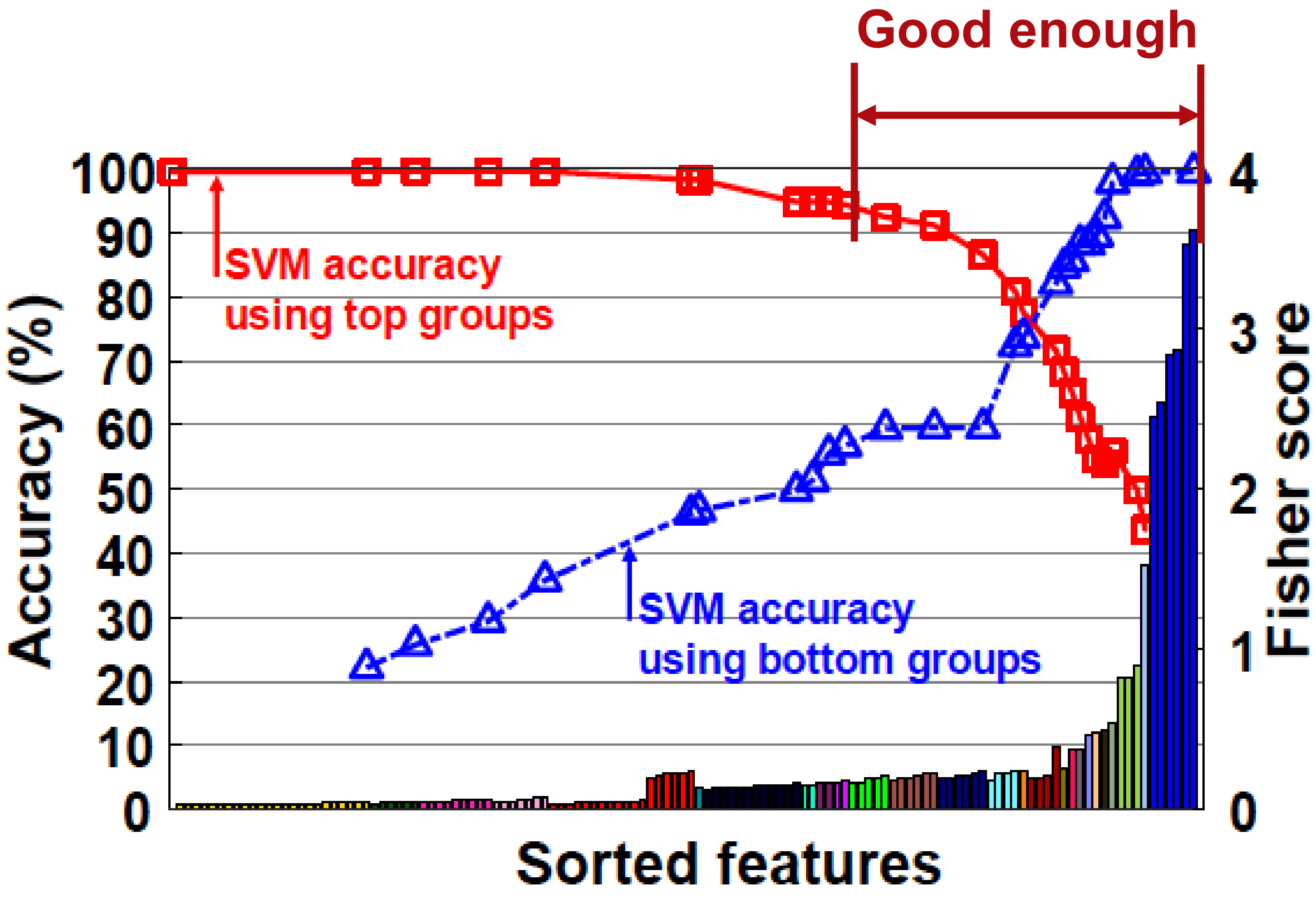




[M. Ozatay, IEEE J-IoT 2018]

- PI feature extraction of human interactions
- Simple (hand-crafted) features, simple ensemble classifier
- Enhanced sample efficiency & accuracy
- Higher cost of sensor deployment

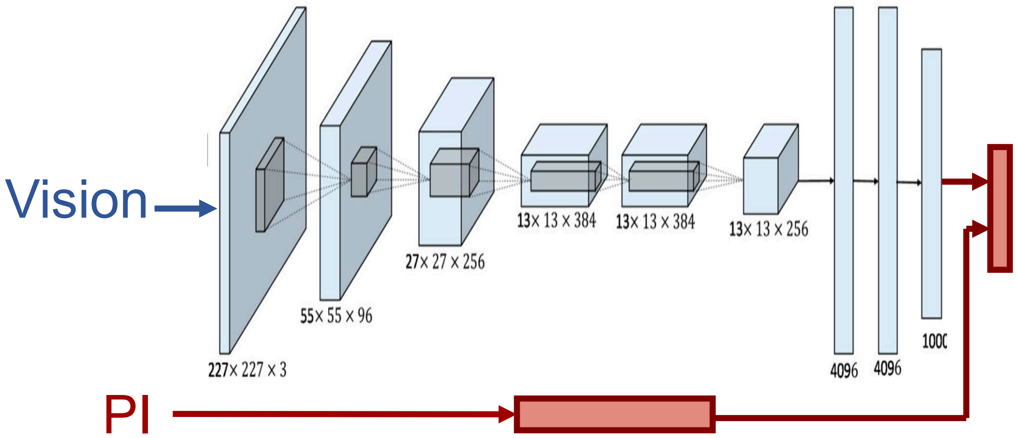
How much PI-sensing structure is needed?



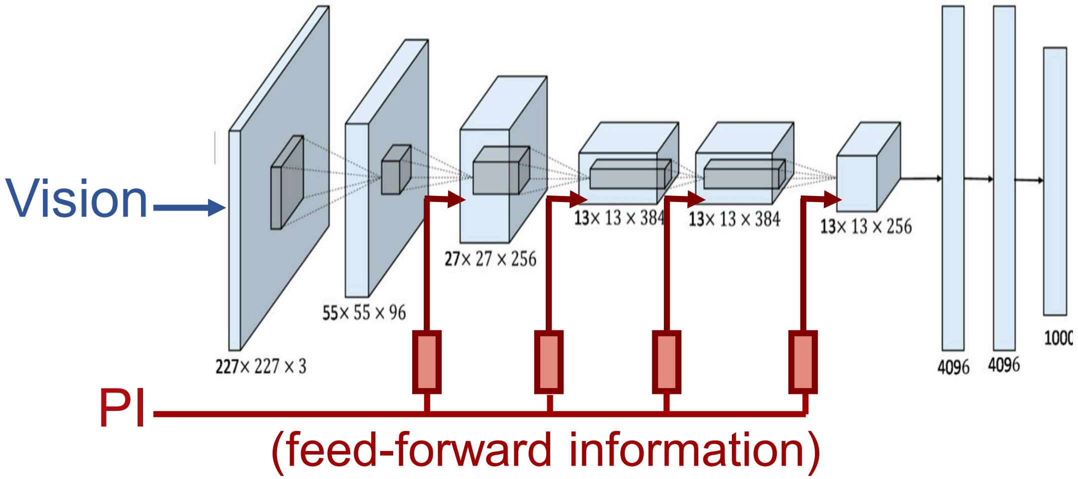
- Small number of PI sensors accounts for most of gains
- PI sensing with specific categories of objects improves perception
- Perception value of categories transfers well
- Selective deployment is feasible (w/ remote sensing…)

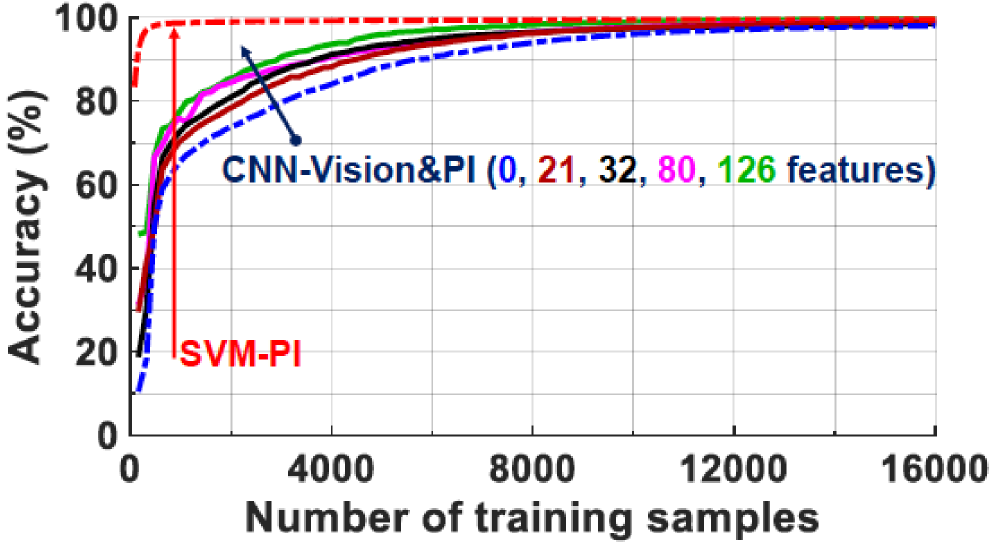
How to leverage PI & remote sensing together (fusion)?

Isolated Features



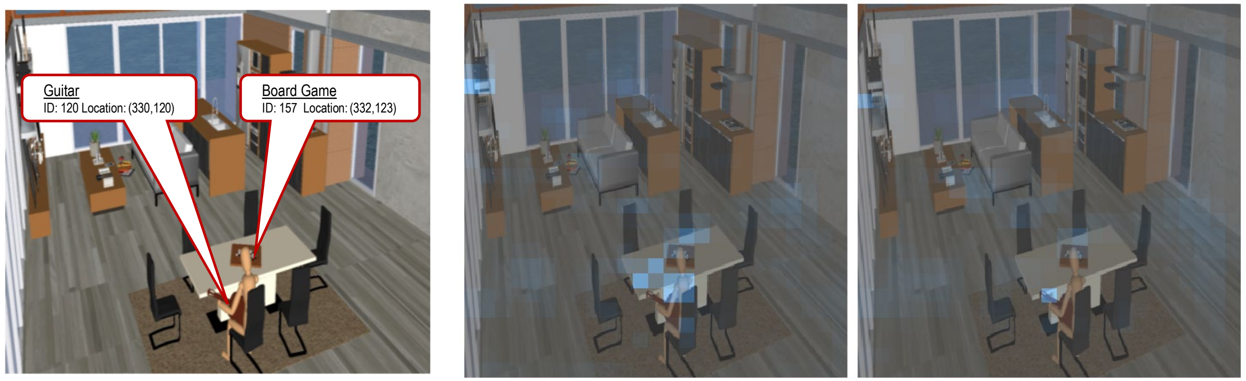
Integrated Features





[M. Ozatay, IEEE J-IoT 2018]

Vision Saliency Maps

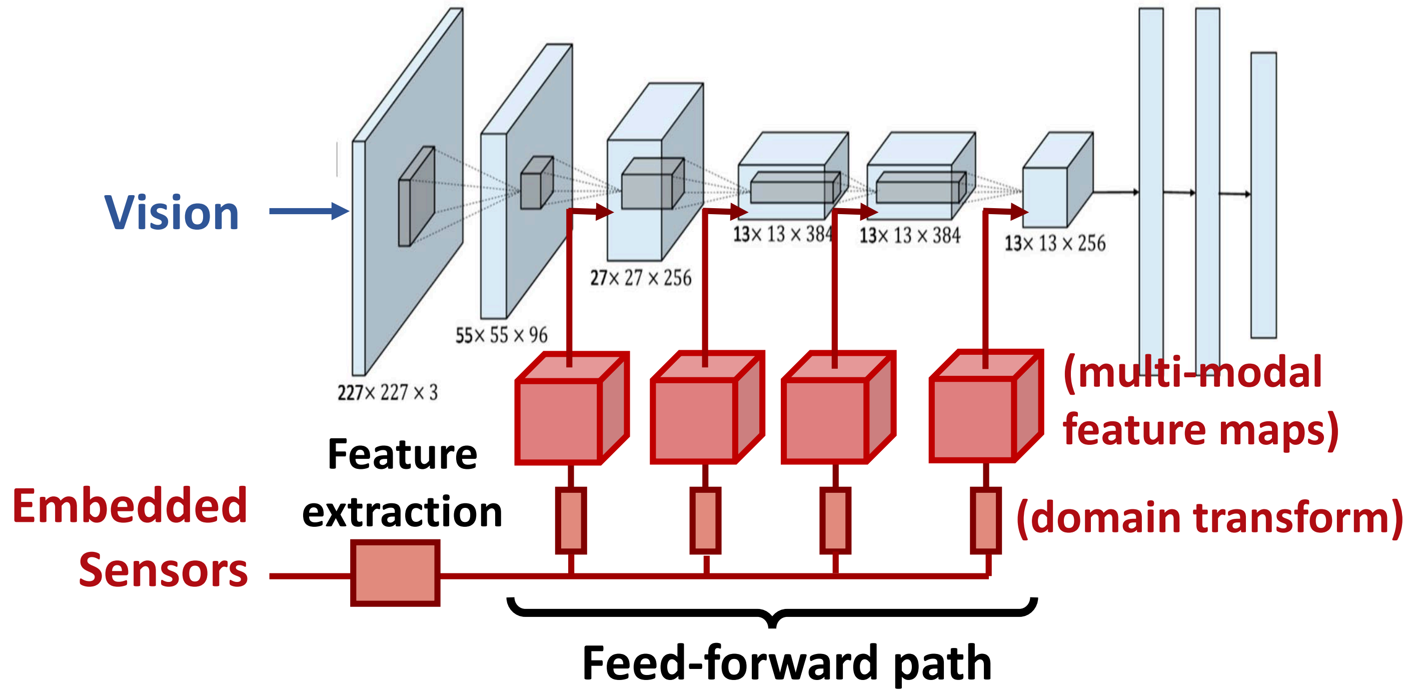


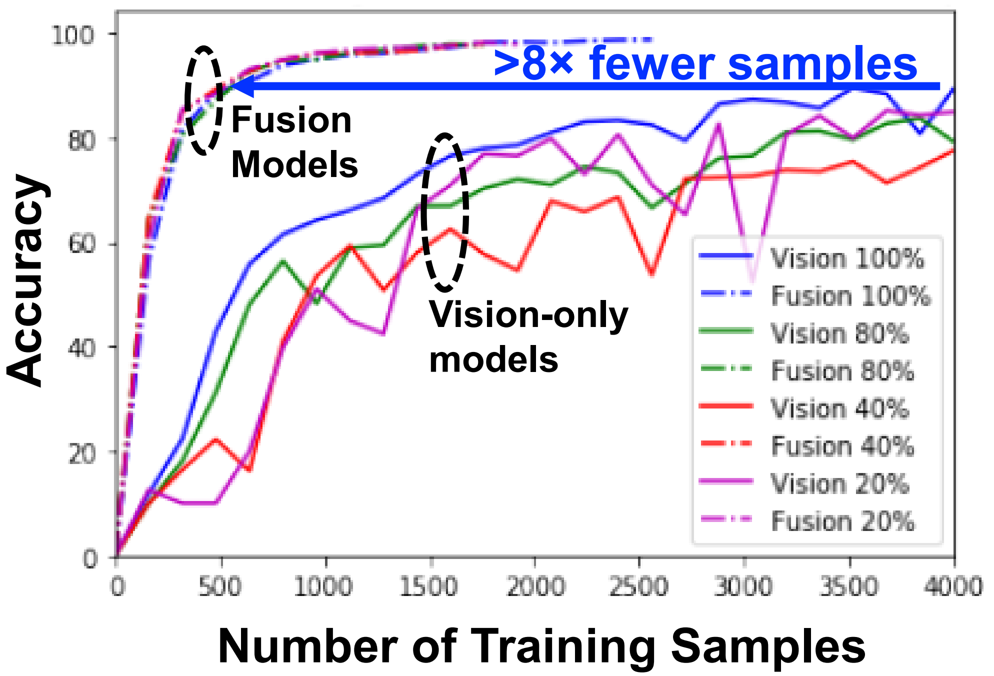
Conv. Layer 4 (post activation)

Conv. Layer 5 (post activation)

Generalized model for PI-vision fusion

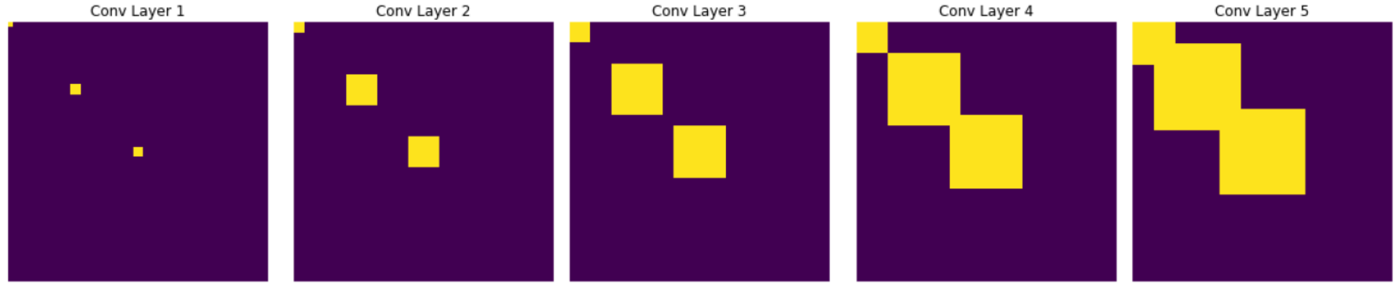
- Shared representations (feature maps) based on spatial association of PI sensing





Ex. Multi-modal Feature Map

(concatenated with conv-net output channels in each layer)

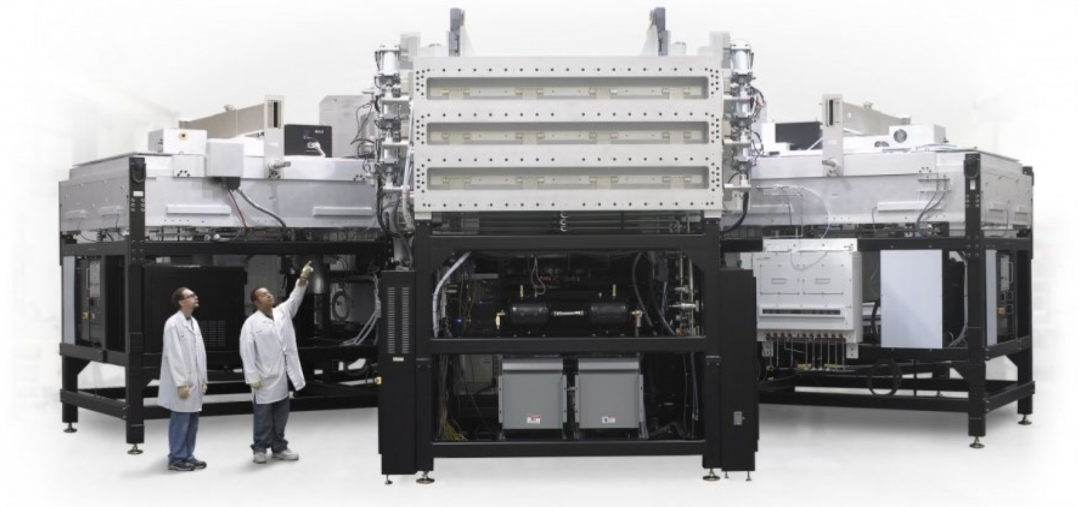


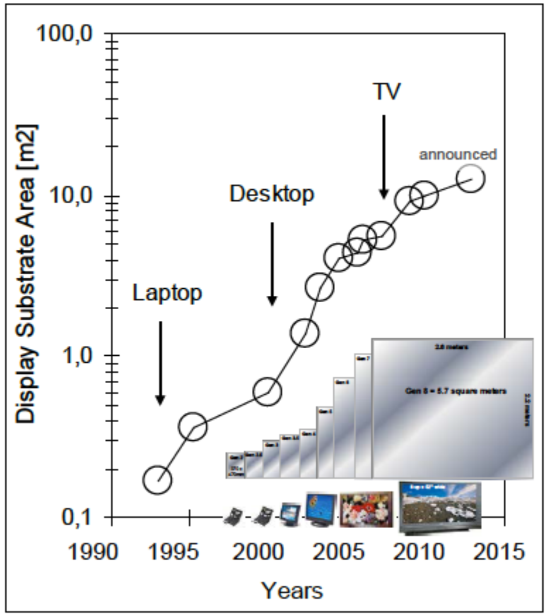
Let me pause

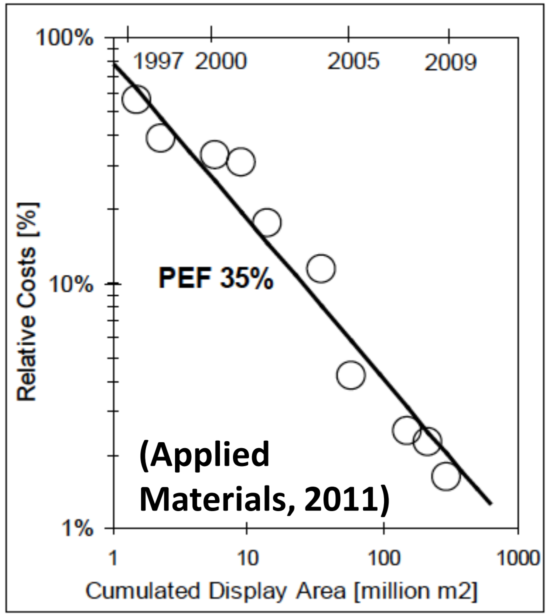
Please ask questions you think might be helpful

Multi-modal, form-fitting sensing to preserve structure

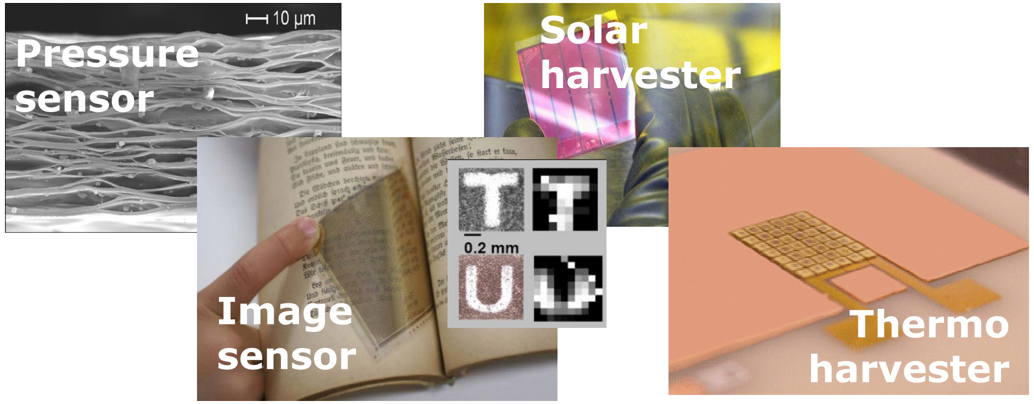
Ex.: large-area electronics (LAE):

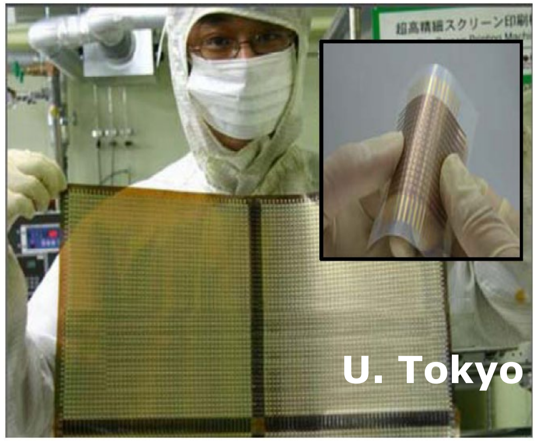






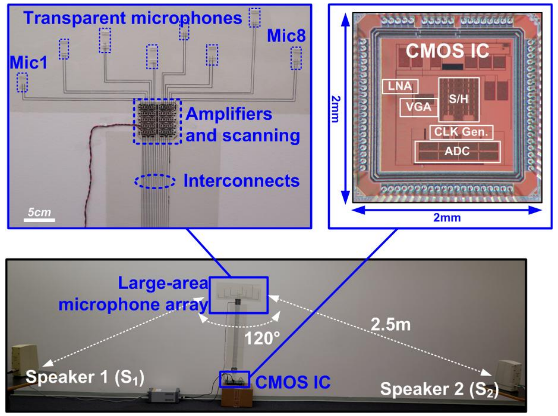
→ From displays to large-scale, form-fitting transducers



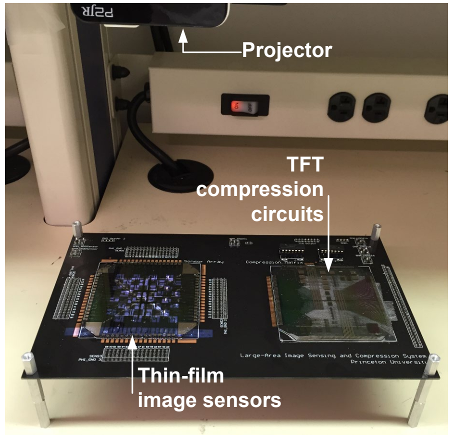


LAE sensing systems

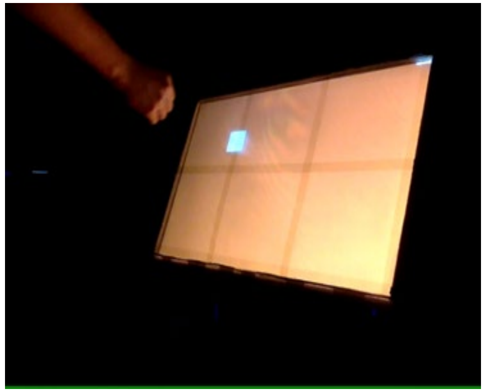
Microphone Array
[L. Huang, VLSI Symp. 2015]



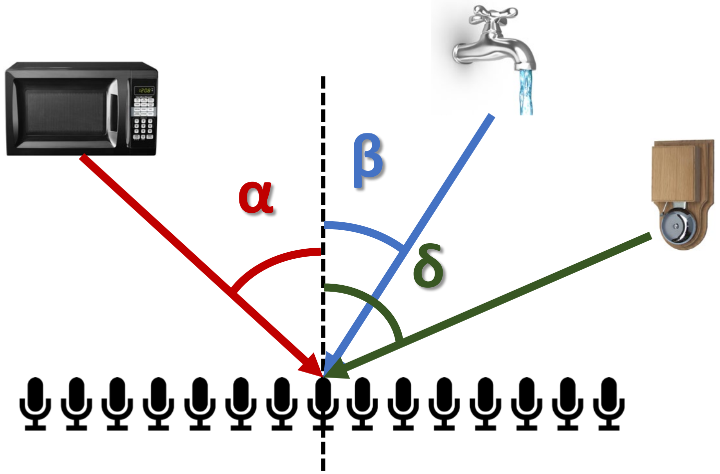
Imaging & classification
[W. Rieutort, ISSCC 2015]



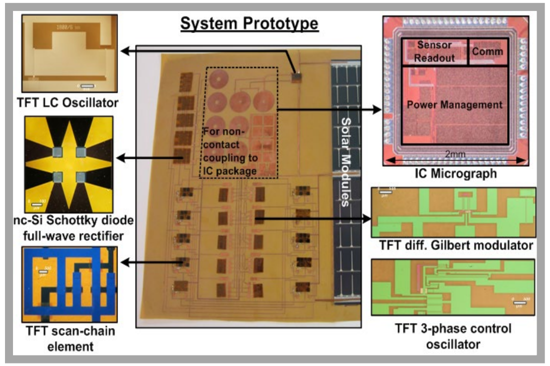
3D Gesture Sensing
[Y. Hu, ISSCC 2014]



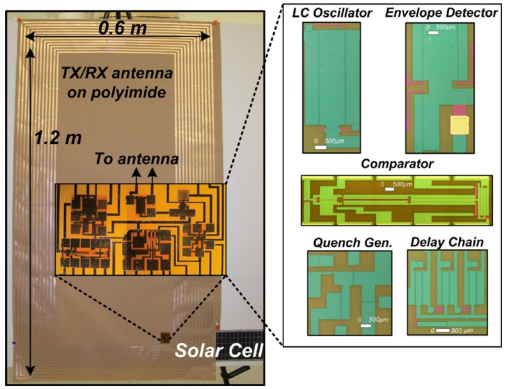
E.g., sound-based activity detection



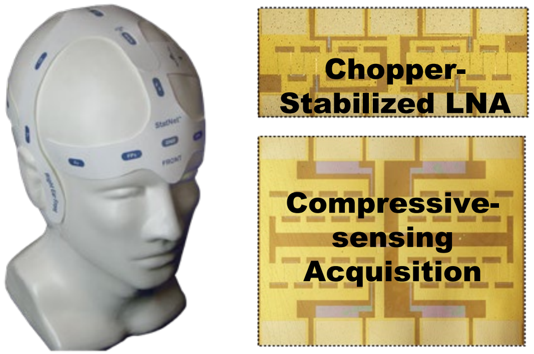
High-density, self-powered Strain Sensing
[Y. Hu, VLSI Symp. 2013]

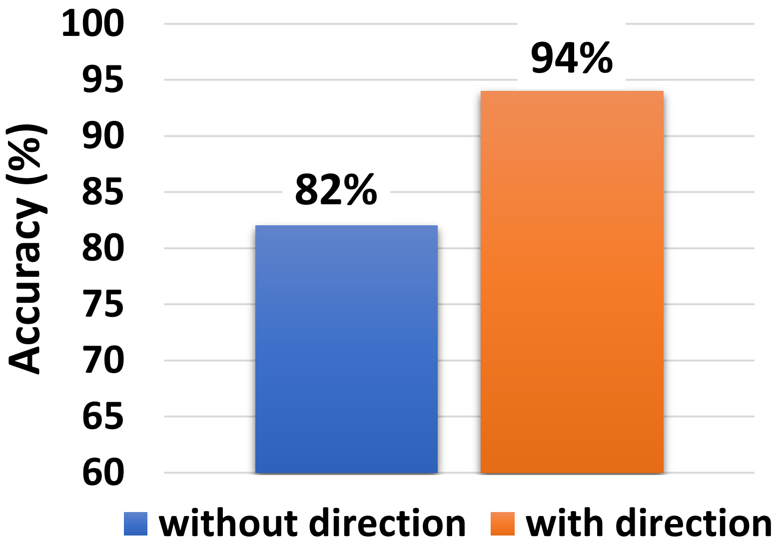


Radios on Wallpaper
[L. Huang, ISSCC 2013]



Active EEG Processing
[T. Moy, ISSCC 2016]

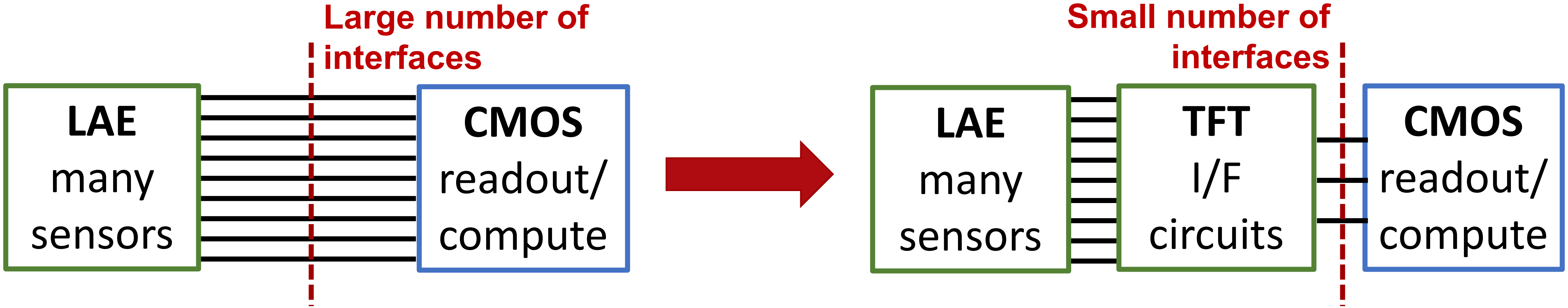




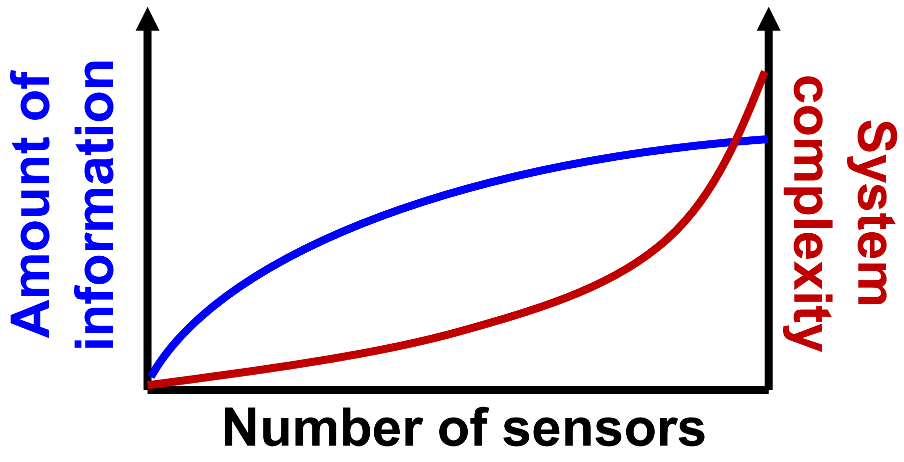
Hybrid LAE-CMOS systems

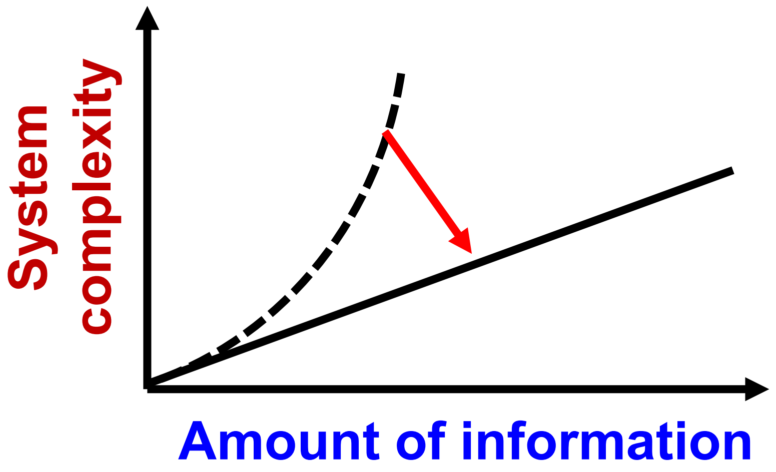
	a-Si TFT	ZnO TFT	Si CMOS (130nm)
Mobility (μ_e/μ_h)	μ_e: 2 cm^2/Vs μ_h: 0.05 cm^2/Vs	μ_e: 12 cm^2/Vs μ_h: <1 cm^2/Vs	μ_e: 1000 cm^2/Vs μ_h: 500 cm^2/Vs
$t_{Gate\text{-}oxide}$	280nm	40nm	2.2nm
V_{DD}	20V	6V	1.2 V
$C_{GD}/_{GS}$	3.3 fF/μm	9.9 fF/μm	0.34 fF/μm
f_T	1MHz	15MHz	150 GHz

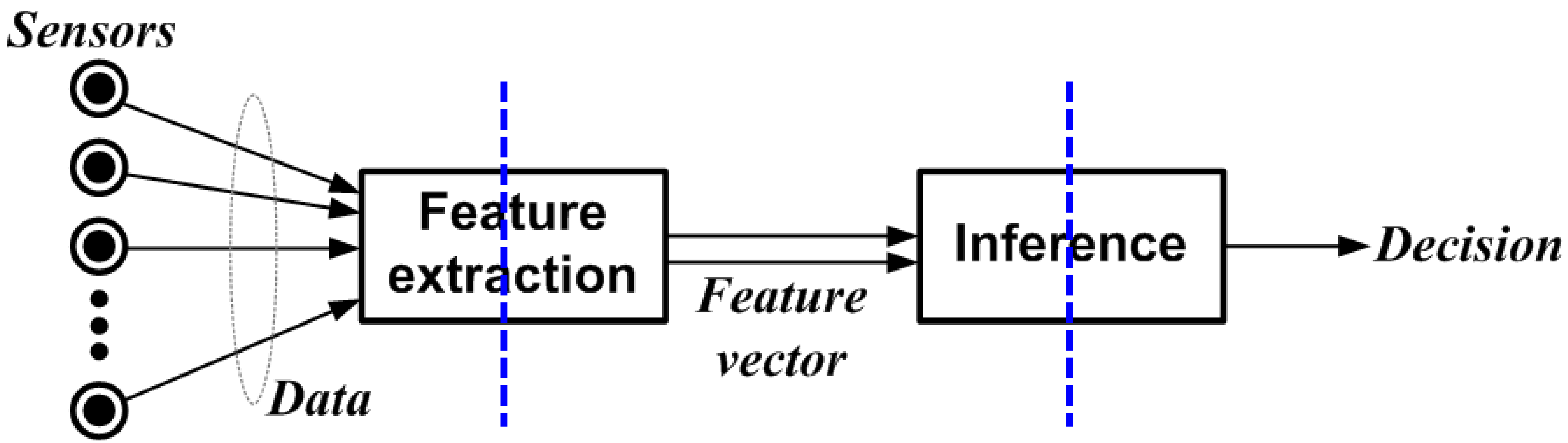
<u>System Challenge: it's the interfaces, stupid</u>



Interfacing information



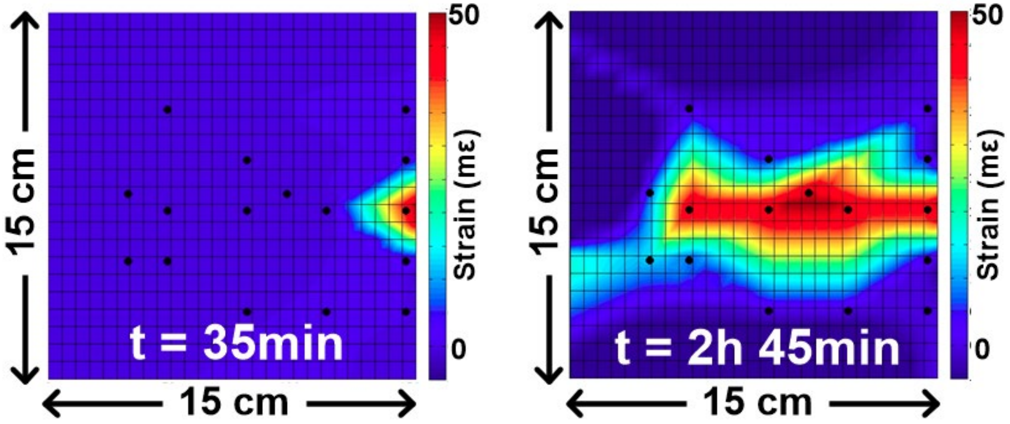


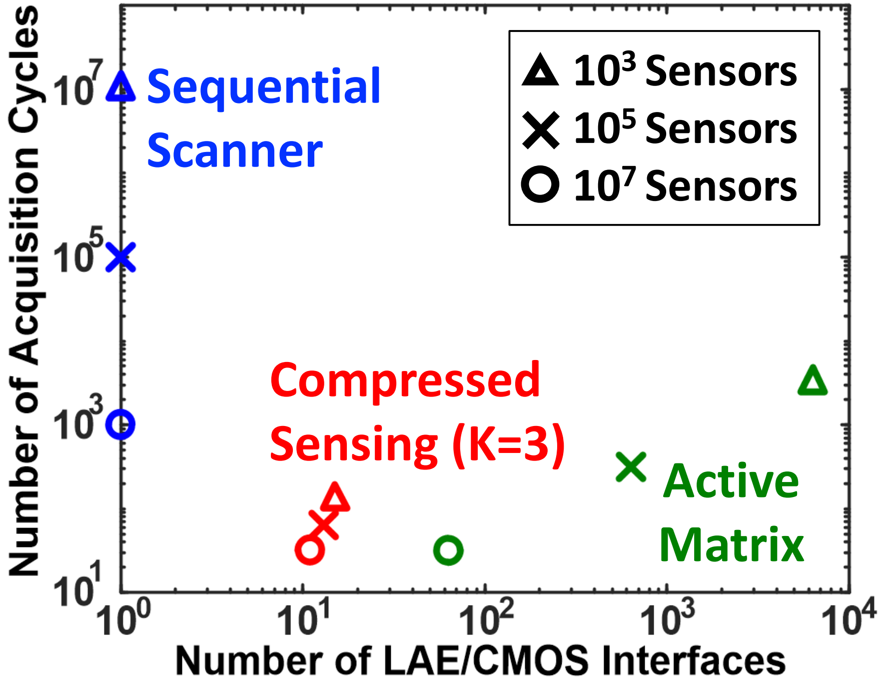


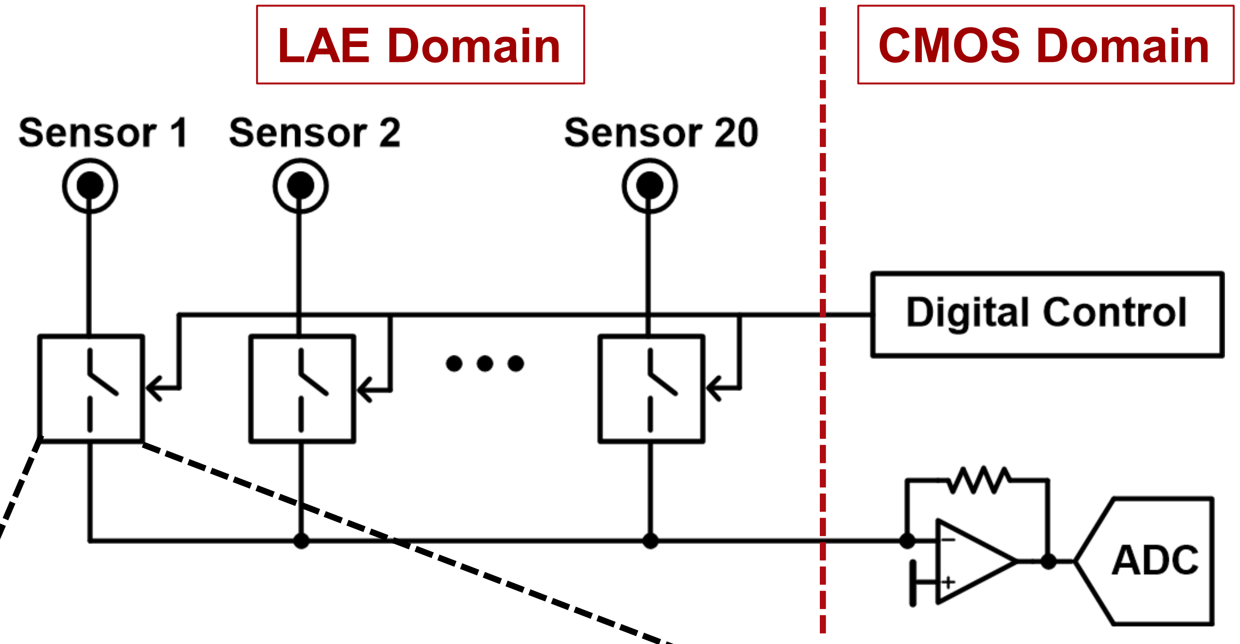
Embedded compressed sensing

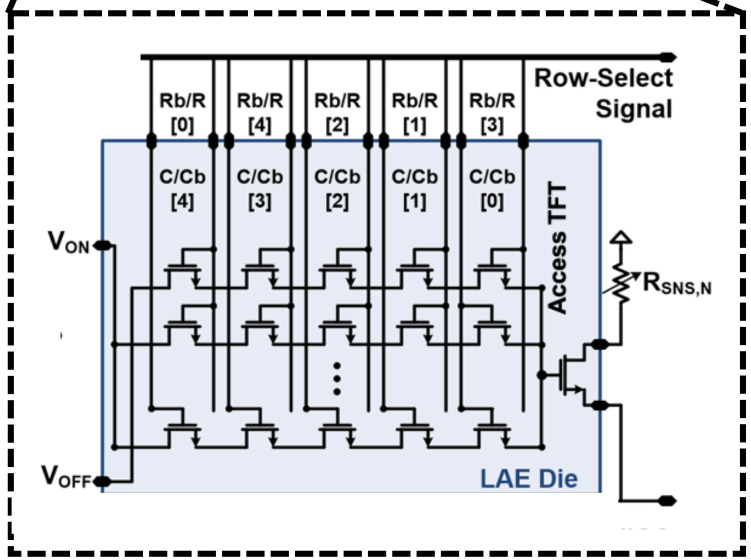
Tactile Sensing Exhibits Sparsity

(e.g., strain sensing for structural monitoring)







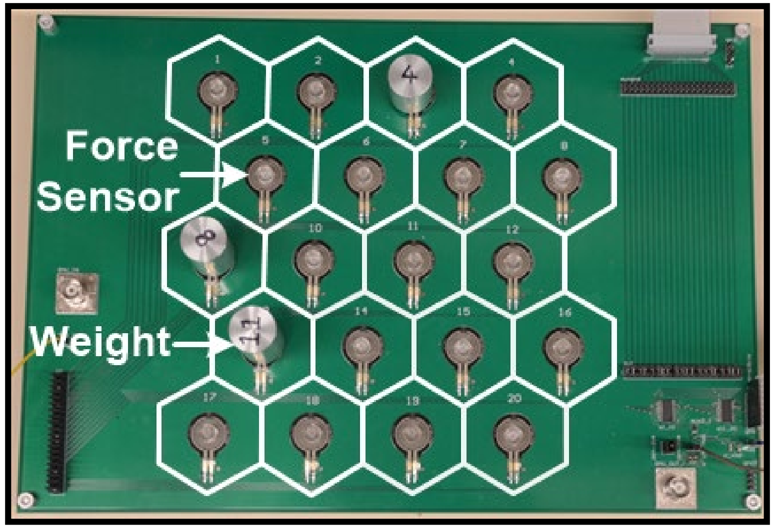


TFT switches implement sensor-current superpositions

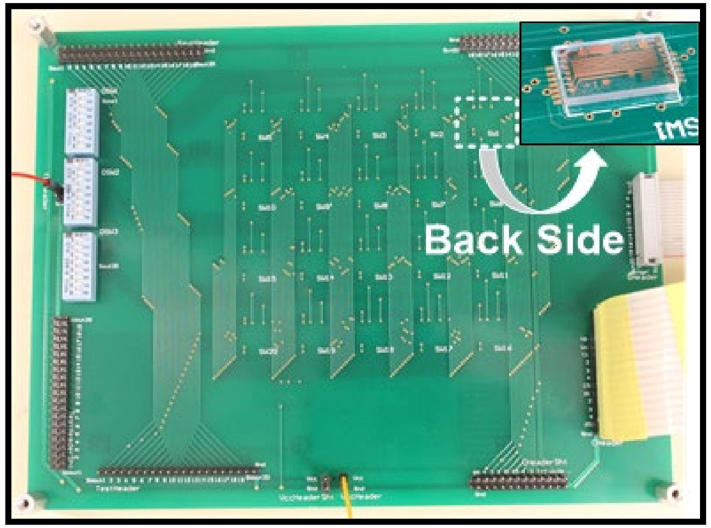
[L. Aygun, ISSCC 2019]

Scalable force-sensing system

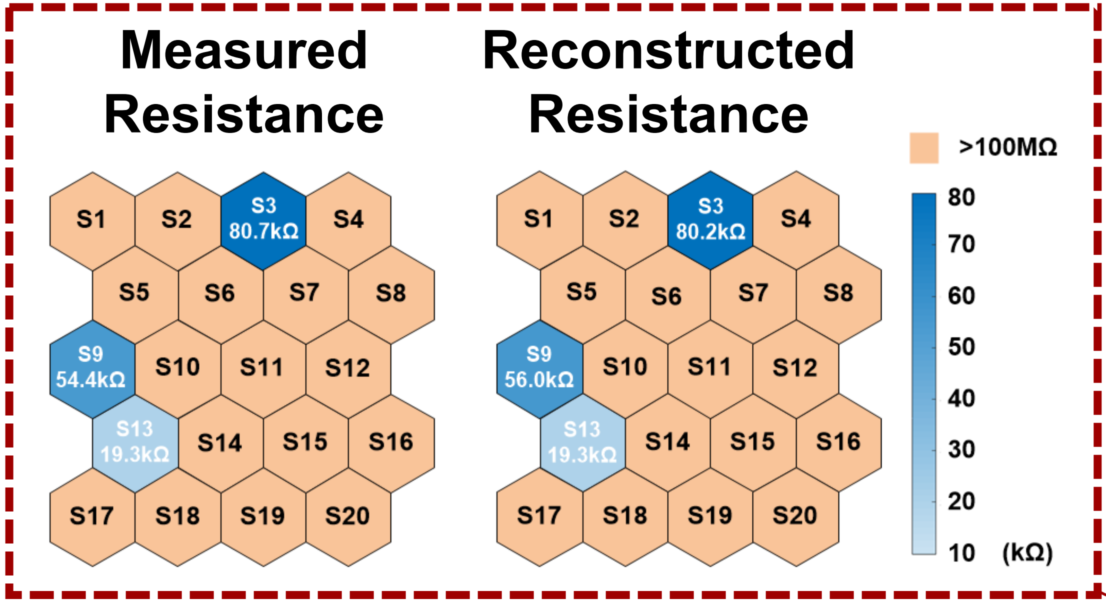
Force Sensor Board

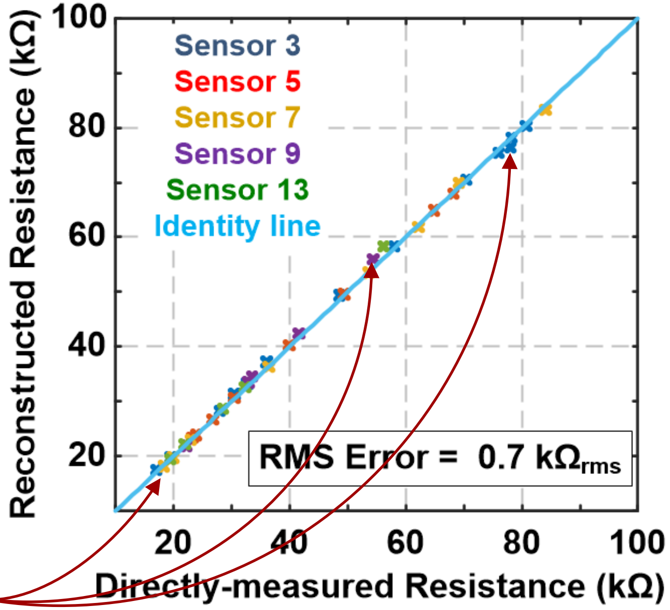


TFT Compression Board



TIA + ADC Board

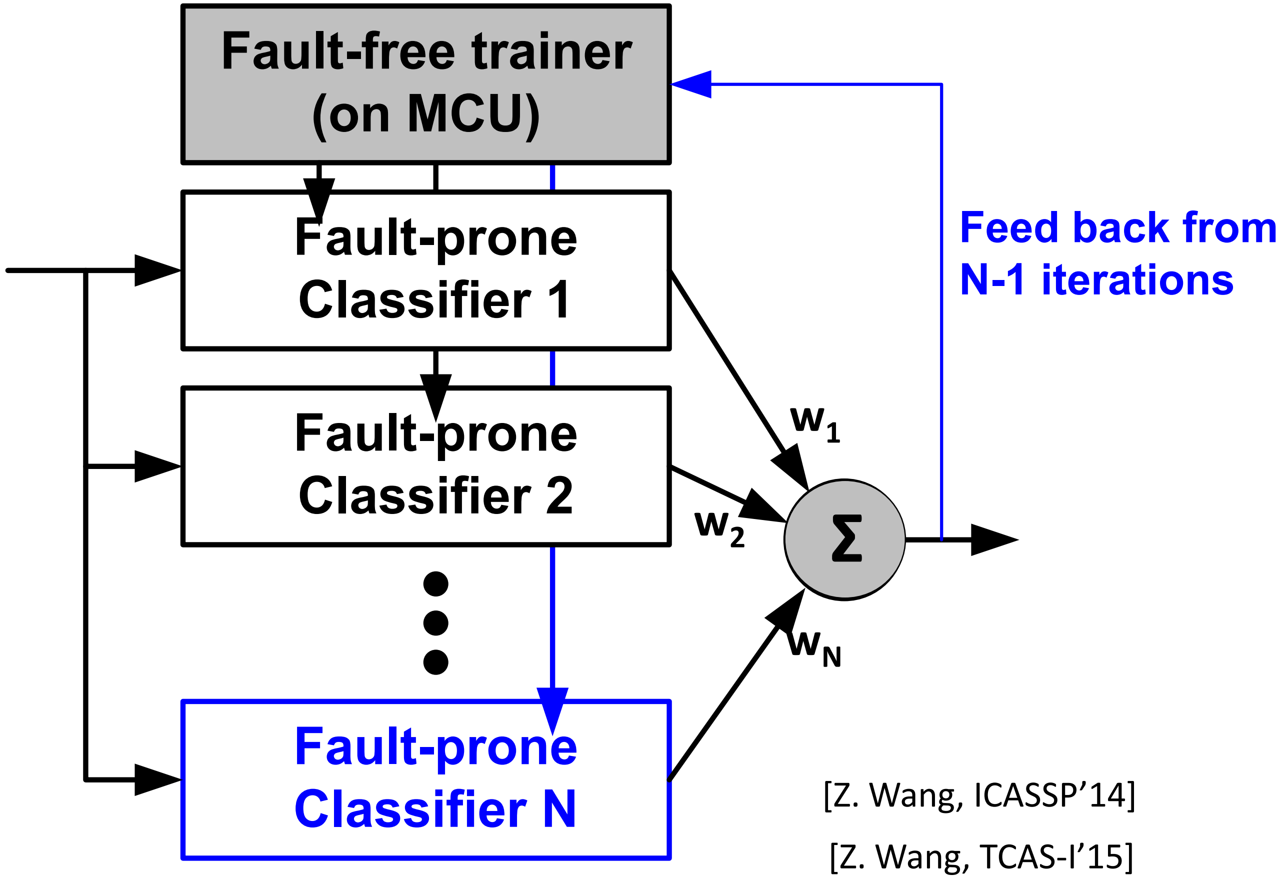




$0.7k\Omega_{RMS}$ sensor reconstruction error

[L. Aygun, ISSCC 2019]

Error-adaptive classifier boosting (EACB)

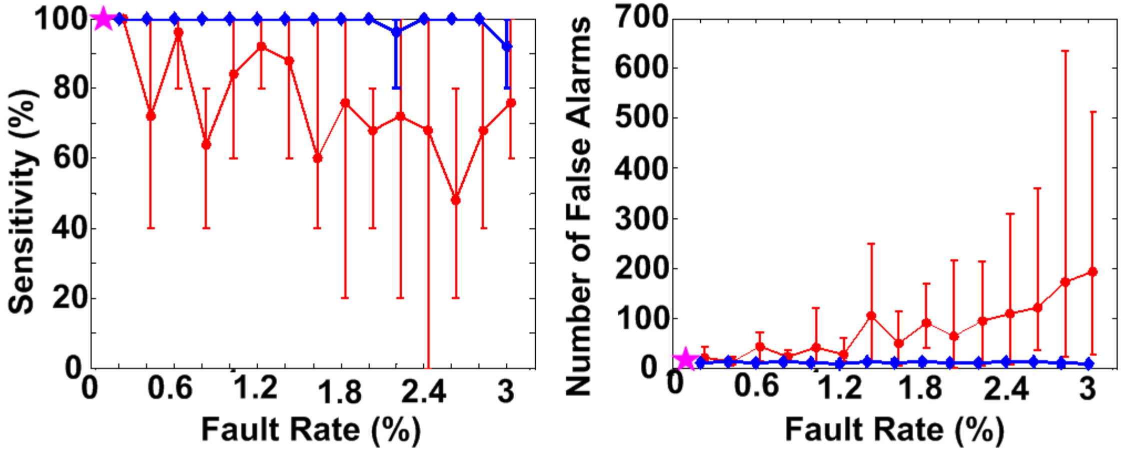


[Z. Wang, ICASSP'14]

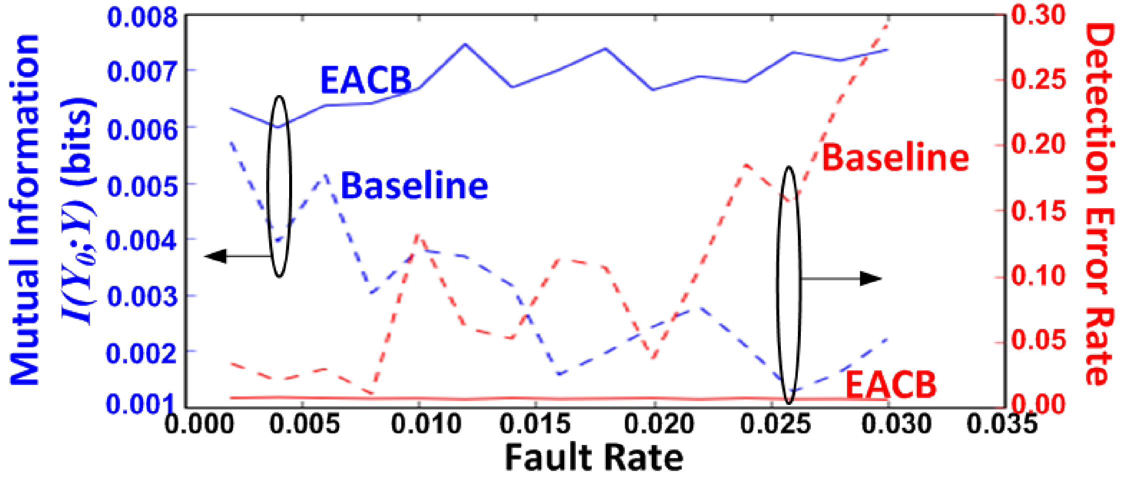
[Z. Wang, TCAS-I'15]

System Performance

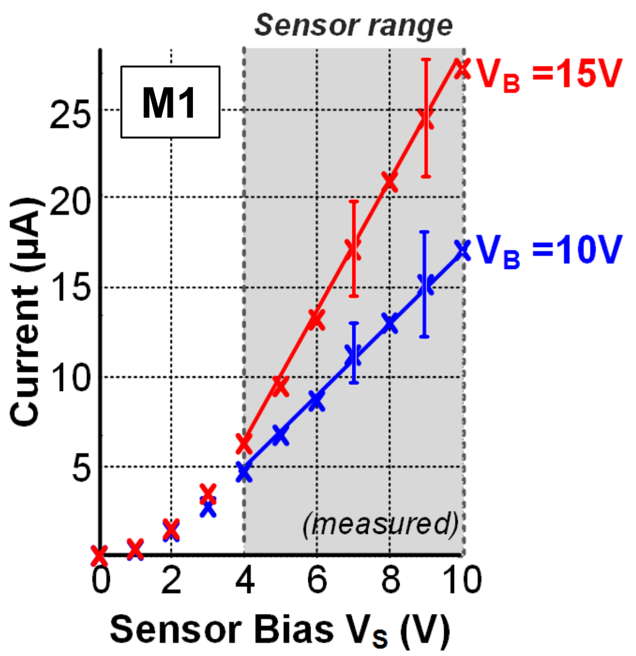
(EEG-based Seizure Detection)

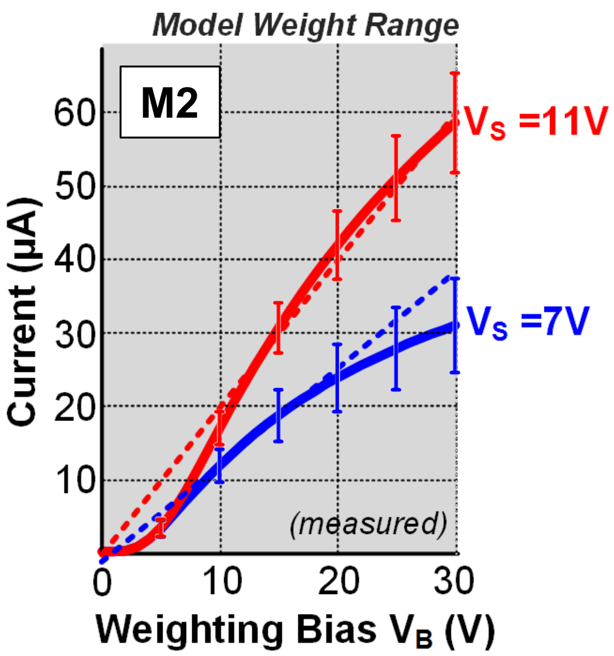


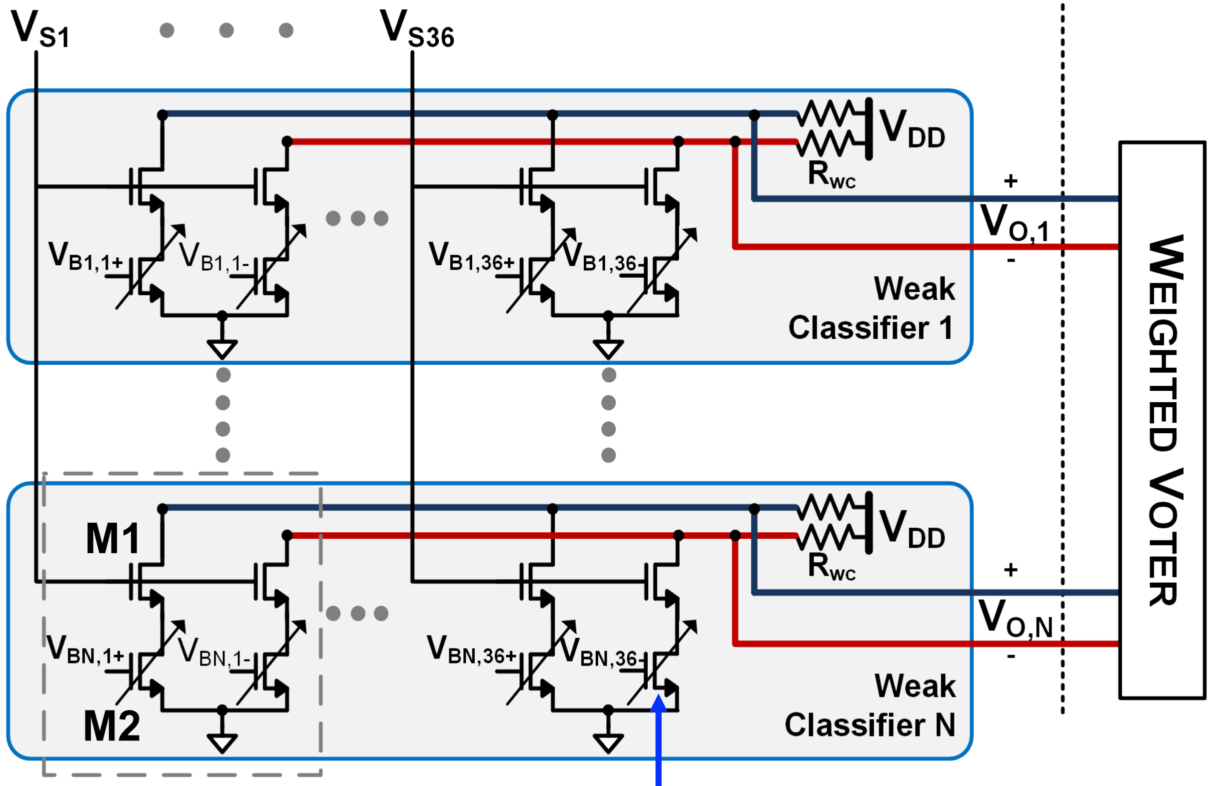
Mutual Information



Embedded weak classifiers

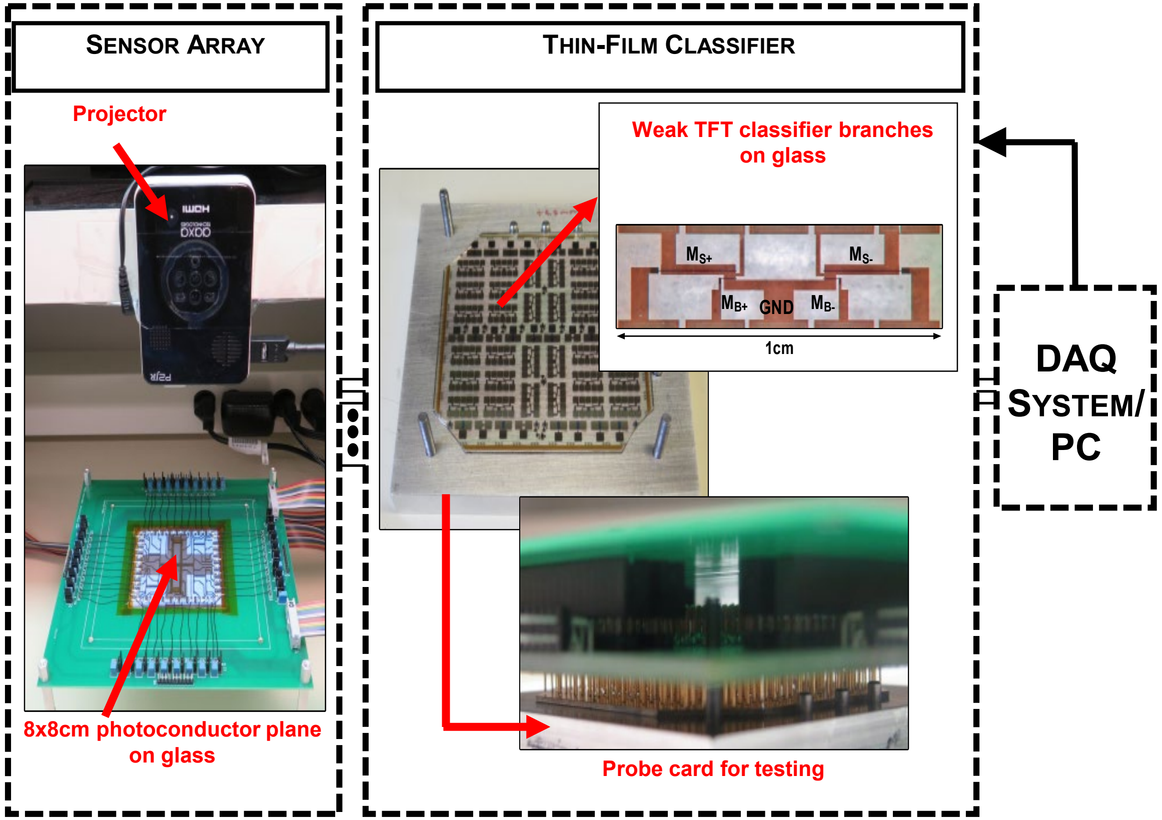






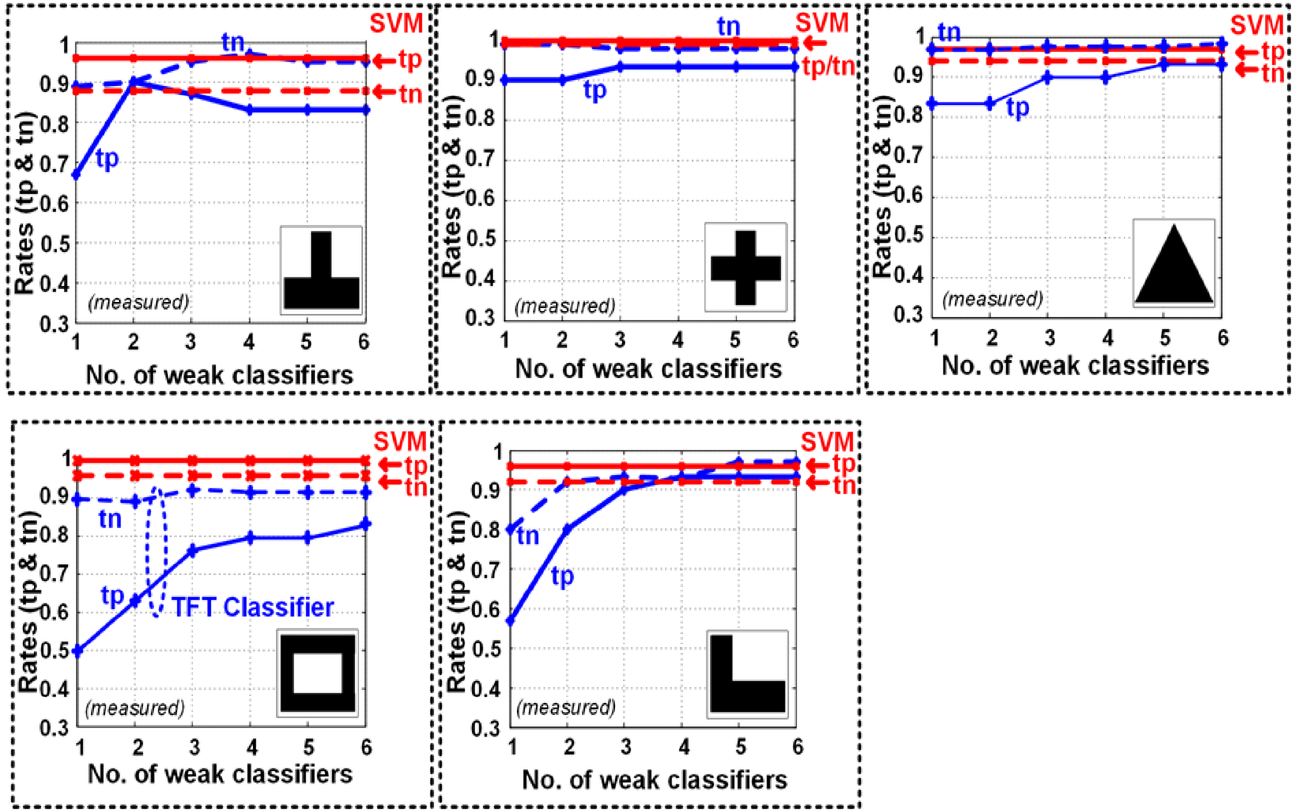
Non-volatile charge trapping stores model weights

Large-area image-detection system



[W. Rieutort-Louis, ISSCC 2015]

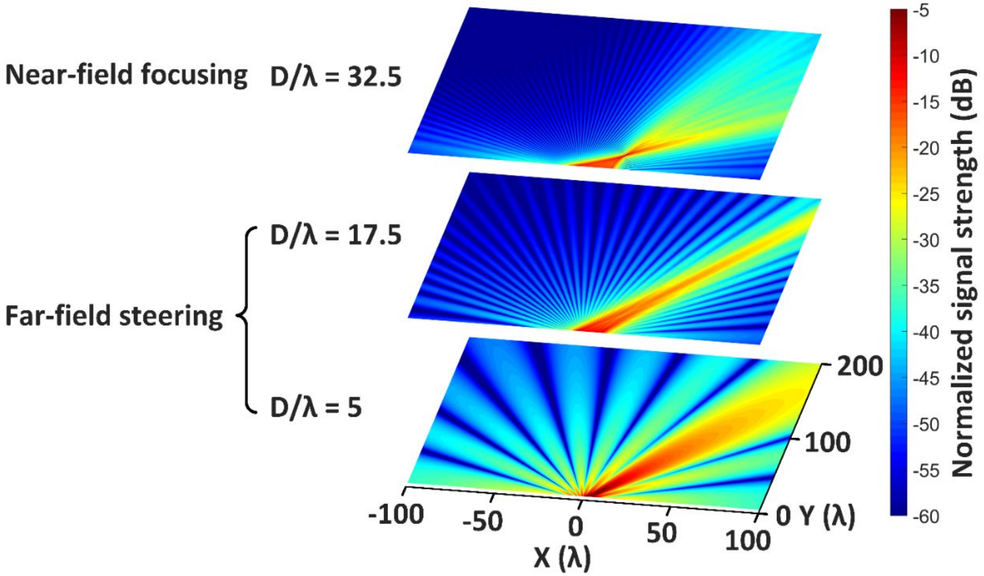
Classification Performance:

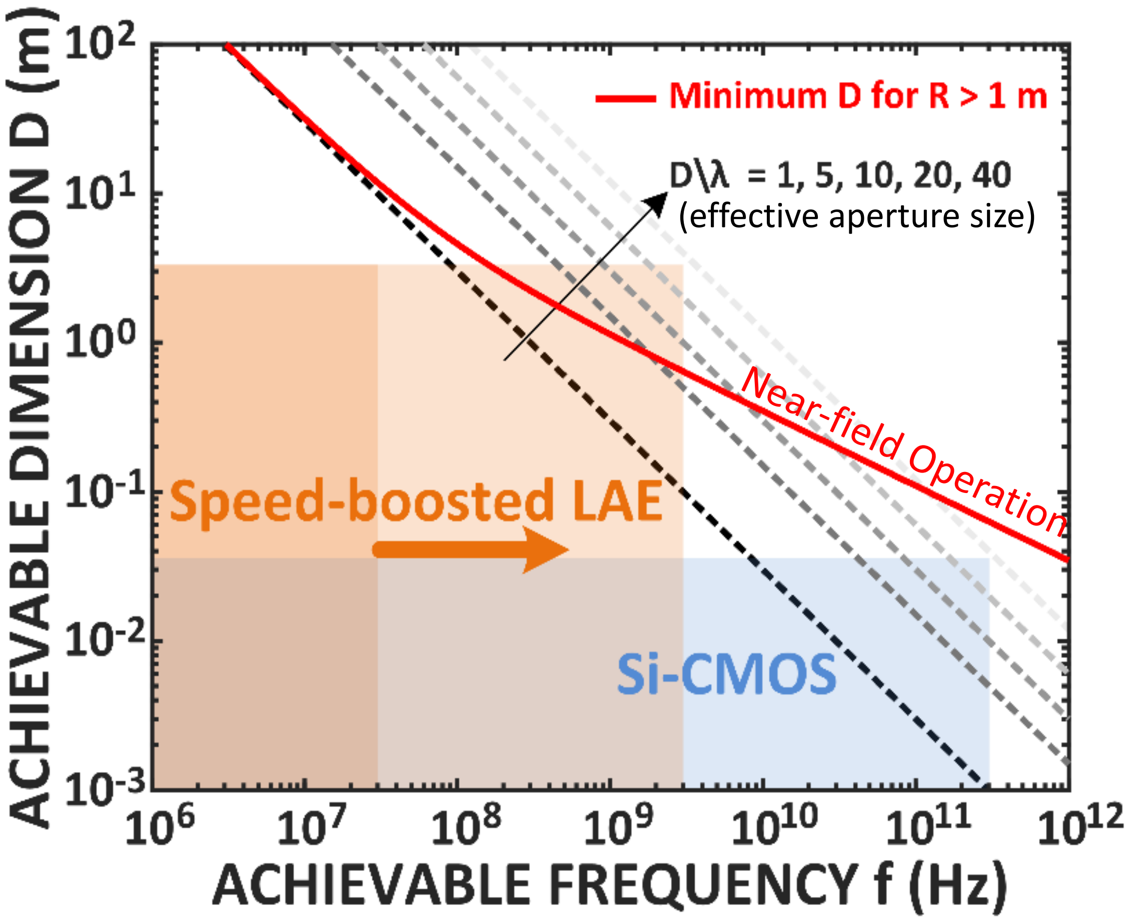


tp: true positive rate

tn: true negative rate

LAE for wireless sensing





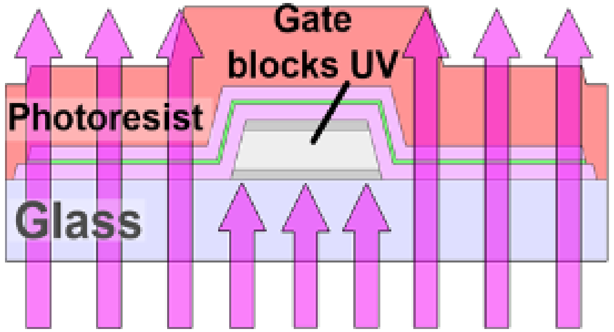
→ **Large D/λ in 1-5 GHz regime**

→ **D on the order of wireless distance**

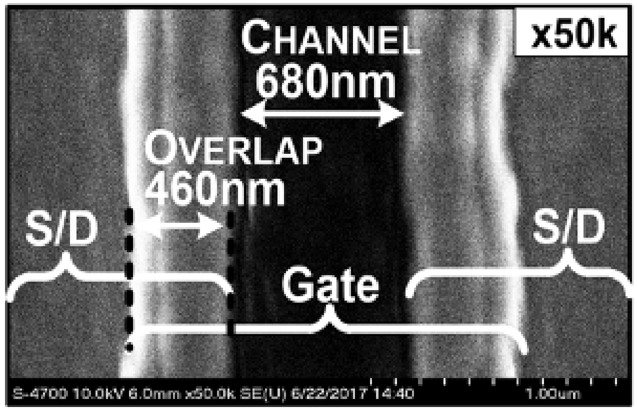
Giga-Hertz TFTs: self alignment

1. Unity Current Gain: $f_T = \dfrac{\overbrace{g_m}}{2\pi\underbrace{(C_{GS} + C_{GD})}}$

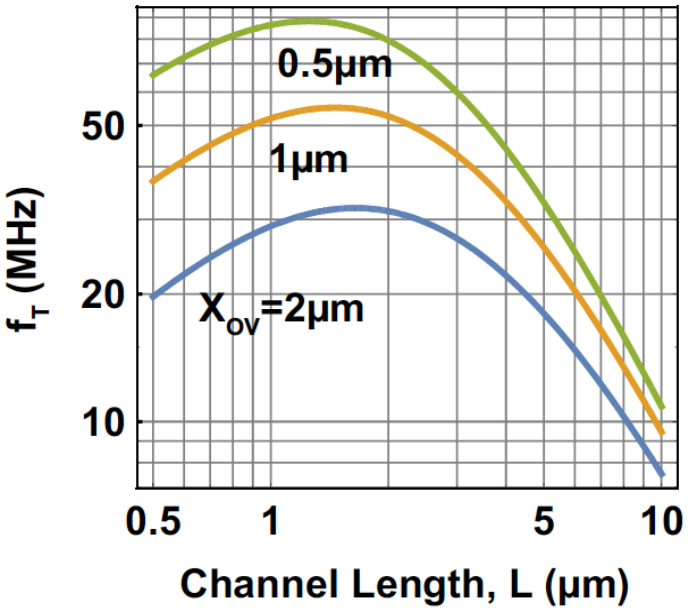
2. Unity Power Gain: $f_{MAX} = \dfrac{f_T}{2\sqrt{2\pi f_T \underbrace{C_{GD}} R_{GATE} + \frac{R_{GATE}}{r_O}}}$



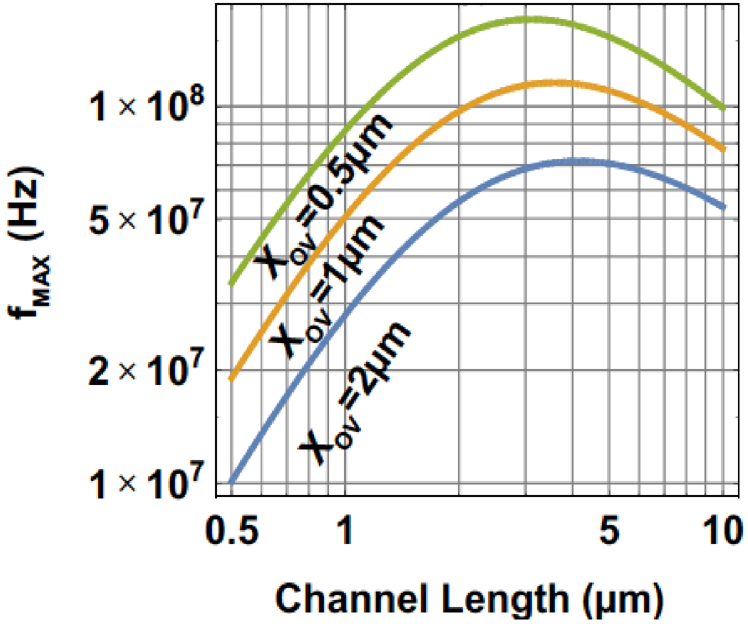
Exposure through substrate



f_T Modeling (X_{OV}, L)



f_{MAX} Modeling (X_{OV}, L)

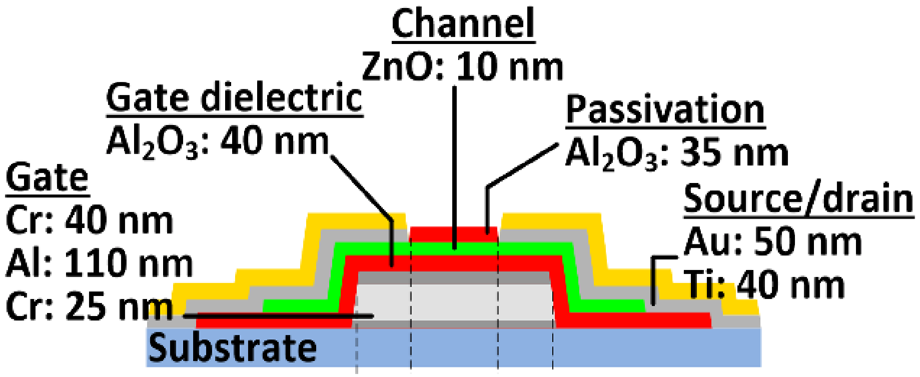


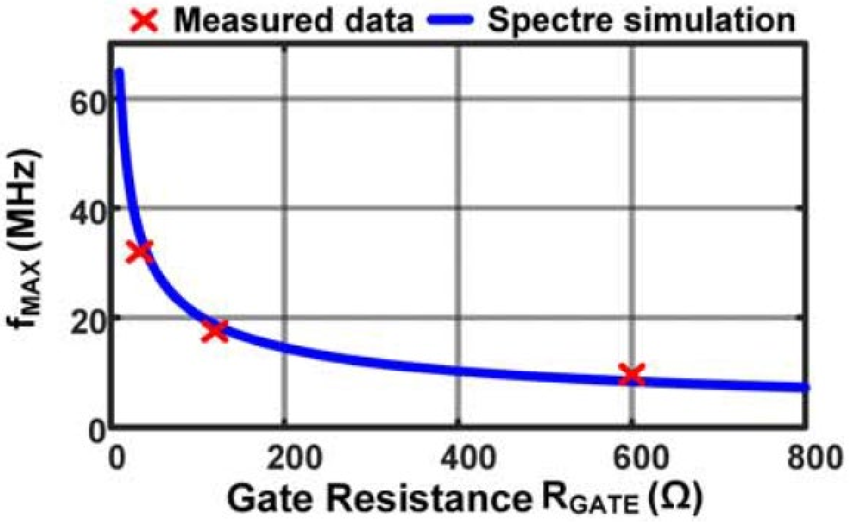
[Y. Mehlman, DRC 2019]

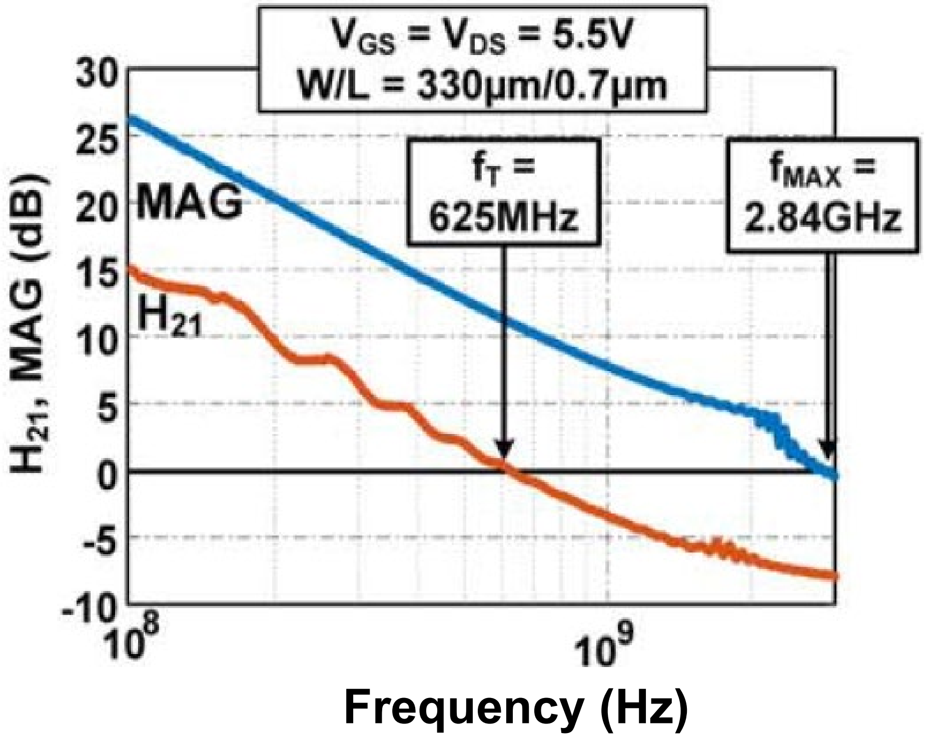
Giga-Hertz TFTs: low-resistance gate

1. Unity Current Gain: $f_T = \dfrac{g_m}{2\pi(C_{GS} + C_{GD})}$

2. Unity Power Gain: $f_{MAX} = \dfrac{f_T}{2\sqrt{2\pi f_T C_{GD} \underbrace{R_{GATE} + \frac{R_{GATE}}{r_O}}}}$



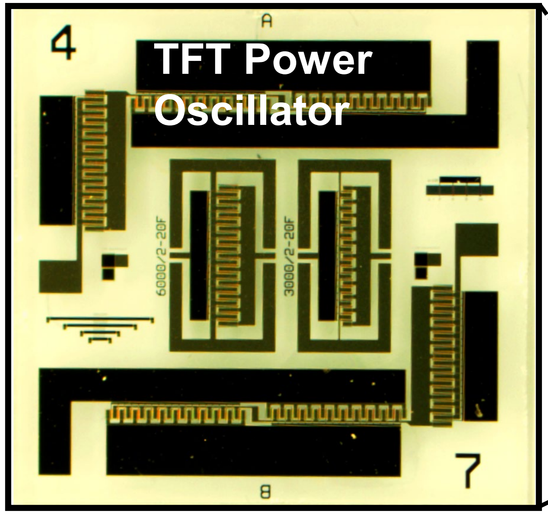


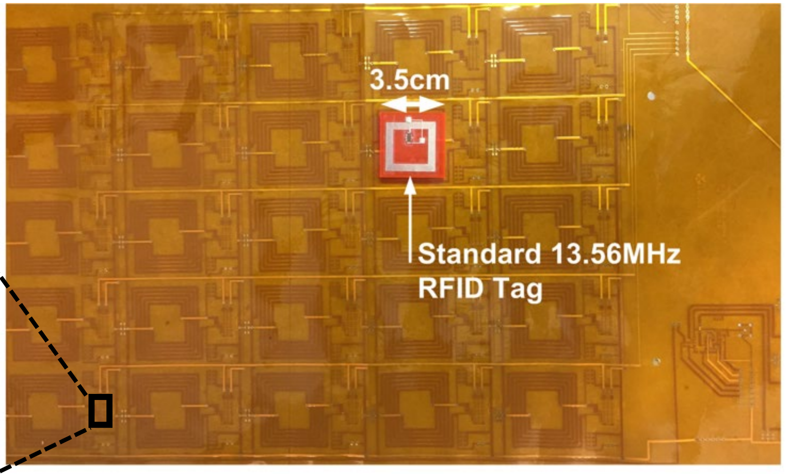


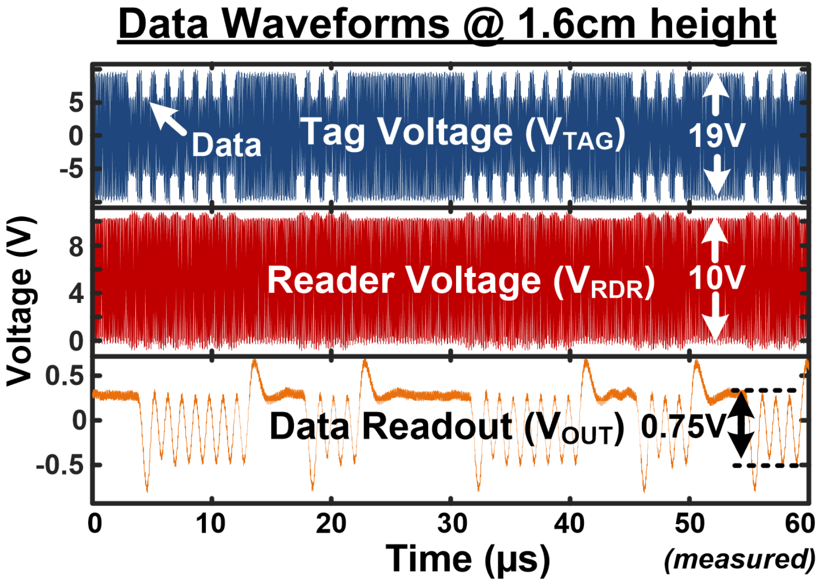
→ Focus on f_{MAX}-limited circuit topologies

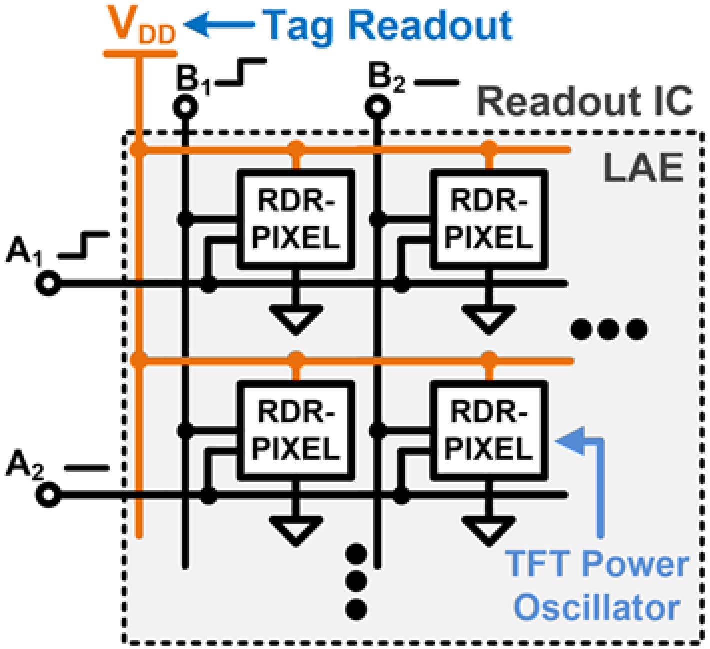
[Y. Mehlman, DRC 2019]

13.56 MHz RFID reader arrays



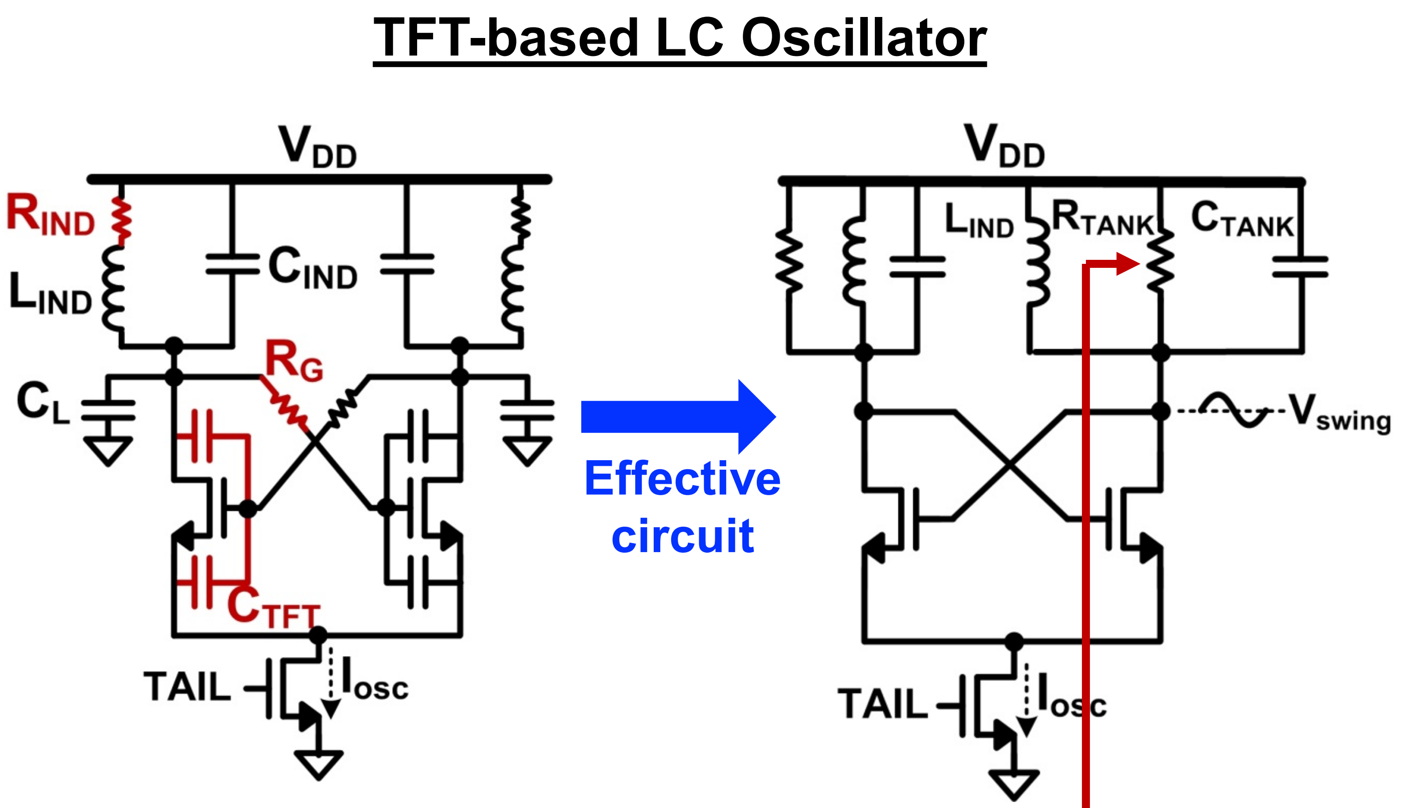




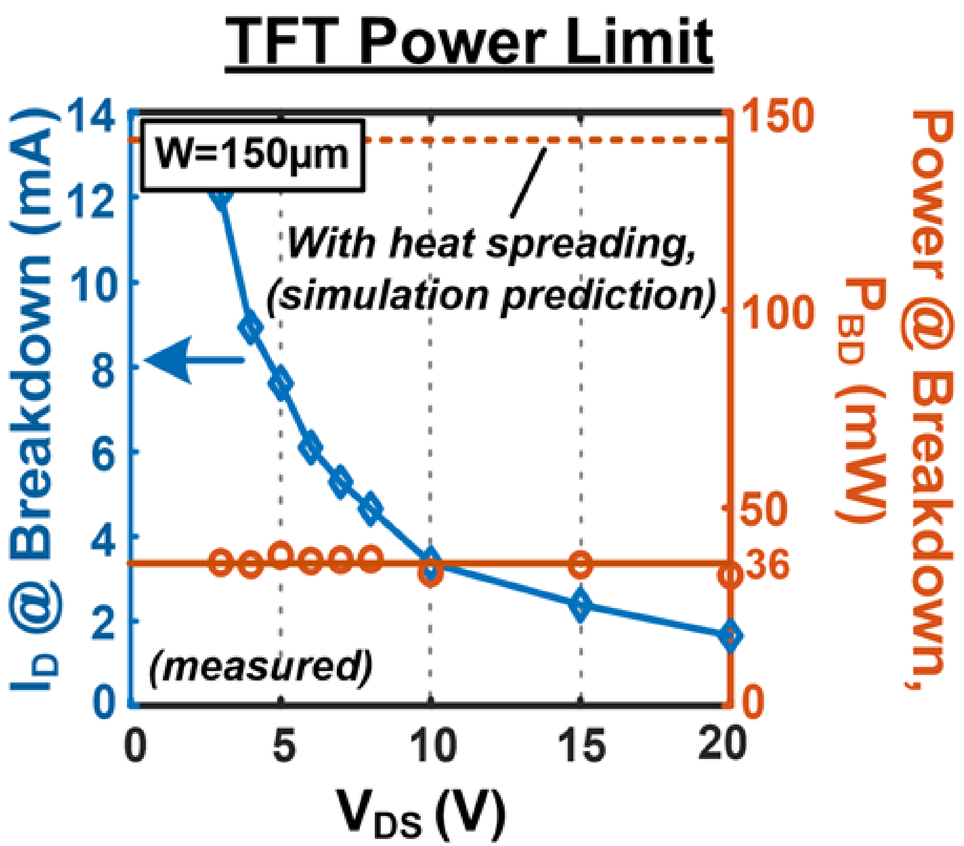


[Y. Mehlman, SSCL 2018]

13.56 MHz RFID reader arrays

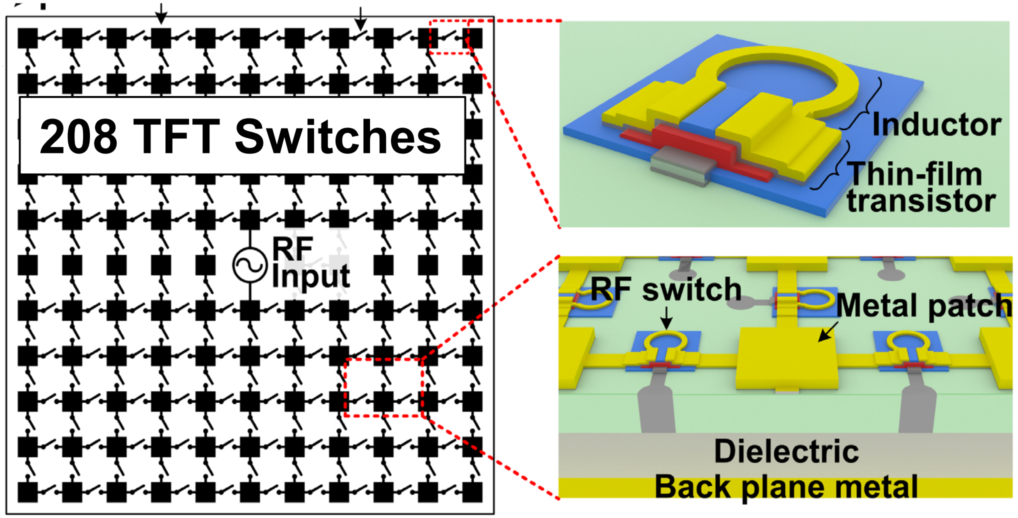


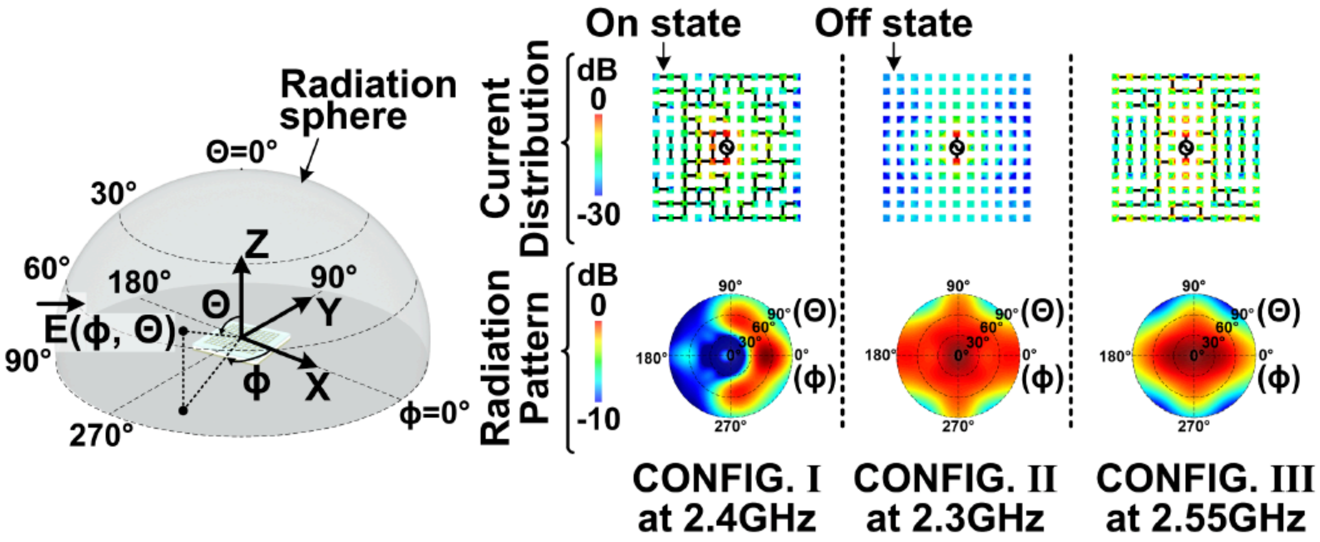
Exploits high-quality LAE passives (inductor)



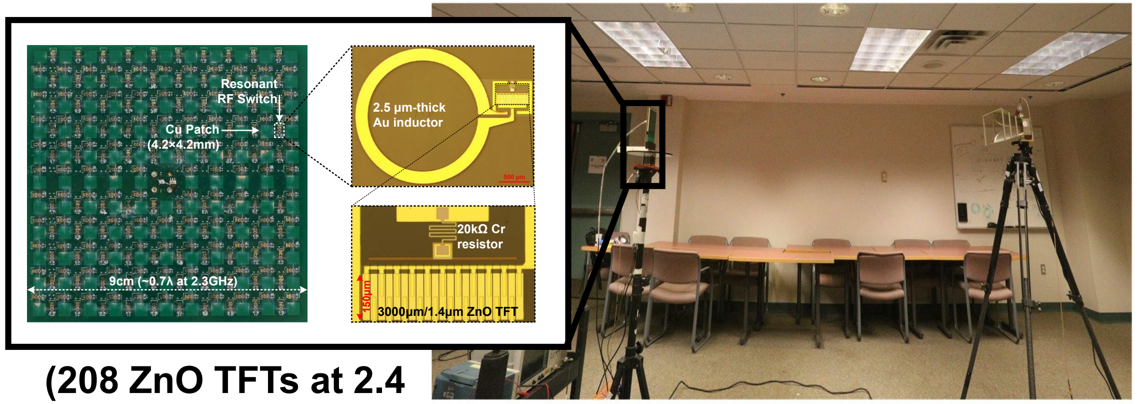
[Y. Mehlman, SSCL 2018]

2.4 GHz reconfigurable antenna



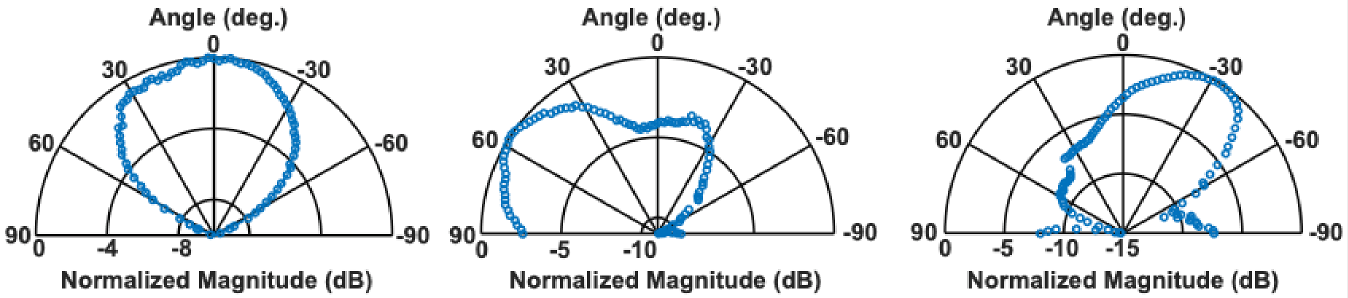


Experimental Demonstration:



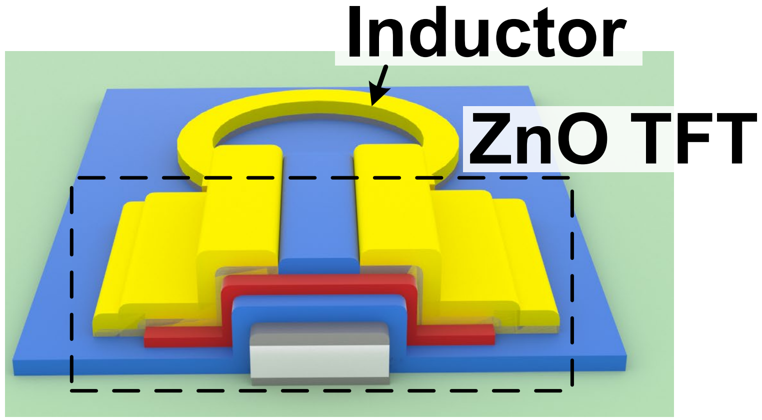
(208 ZnO TFTs at 2.4 GHz in 9cm×9cm array)

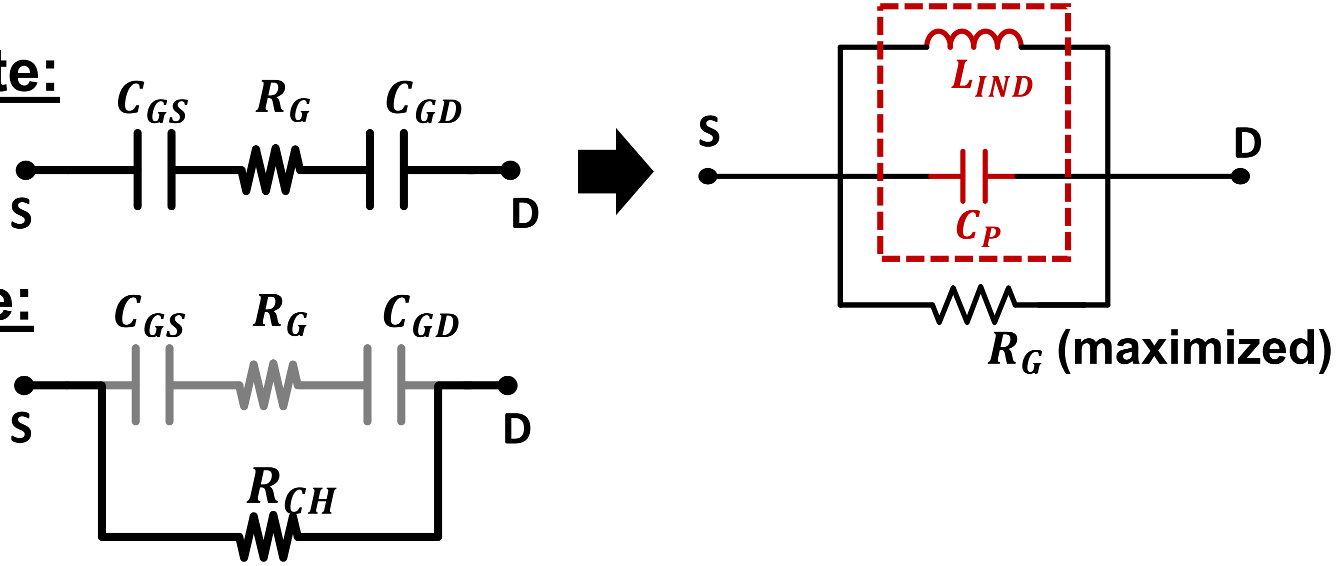
- Tunability in frequency response, polarization, radiation pattern
- Monolithically integrable to large & flexible formfactors



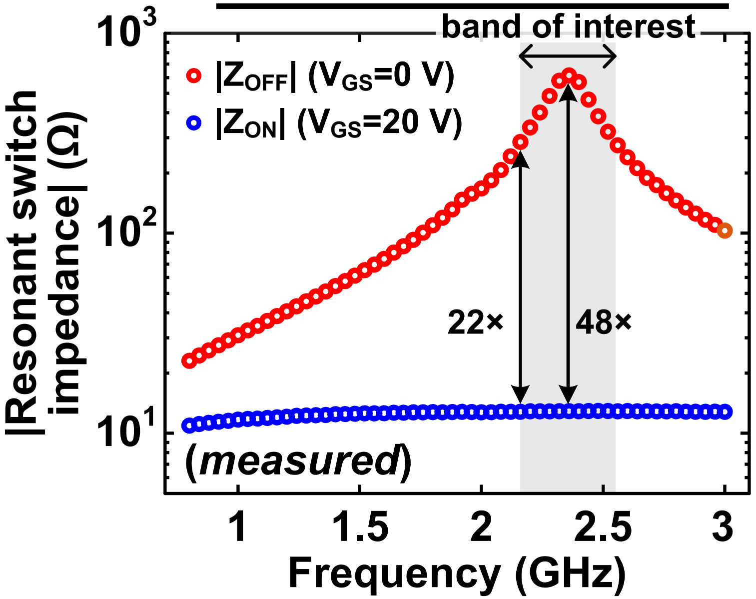
[C. Wu, IEDM 2020]

2.4 GHz reconfigurable antenna

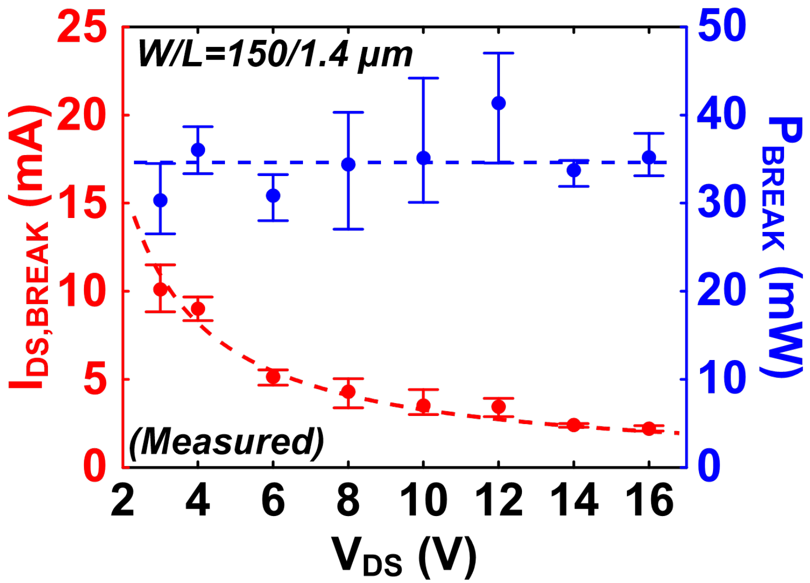


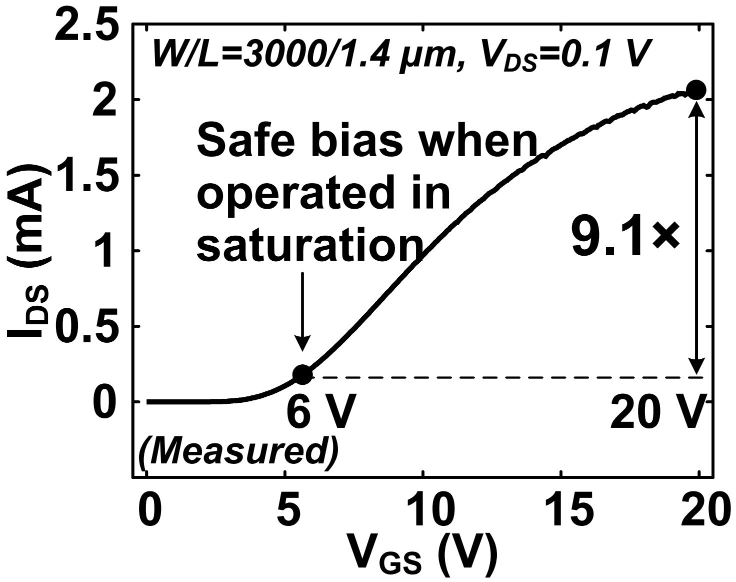


Switch Performance

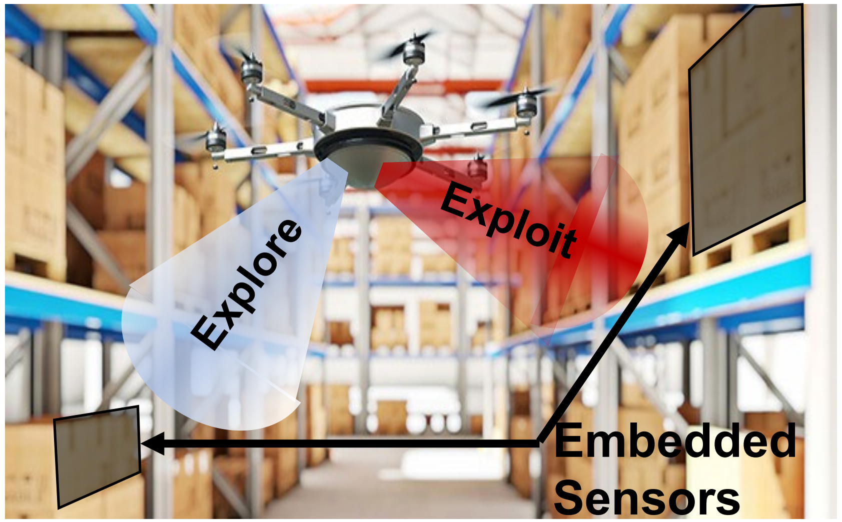


Breakdown-safe Biasing



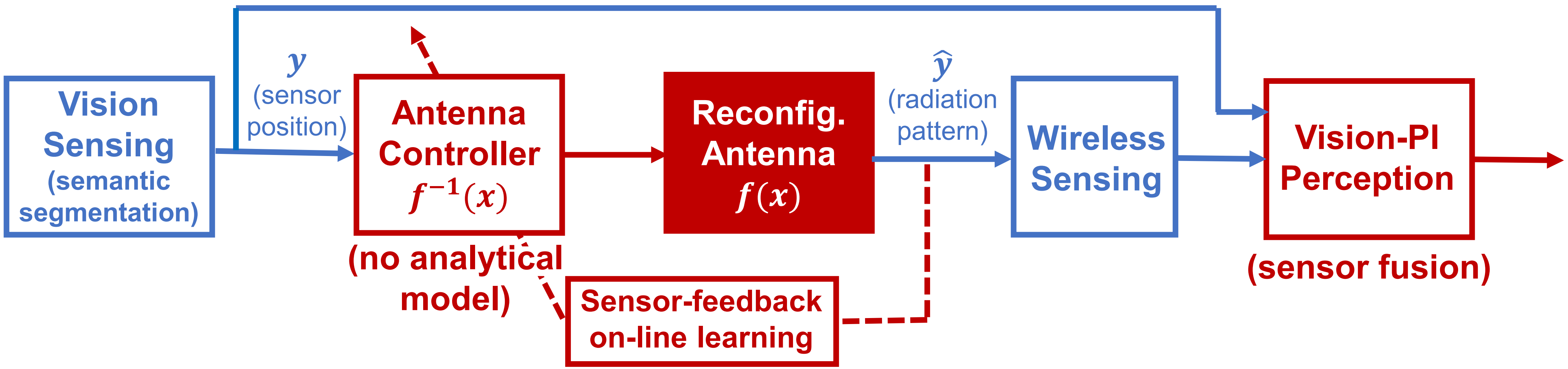


Task-driven large-scale wireless sensing



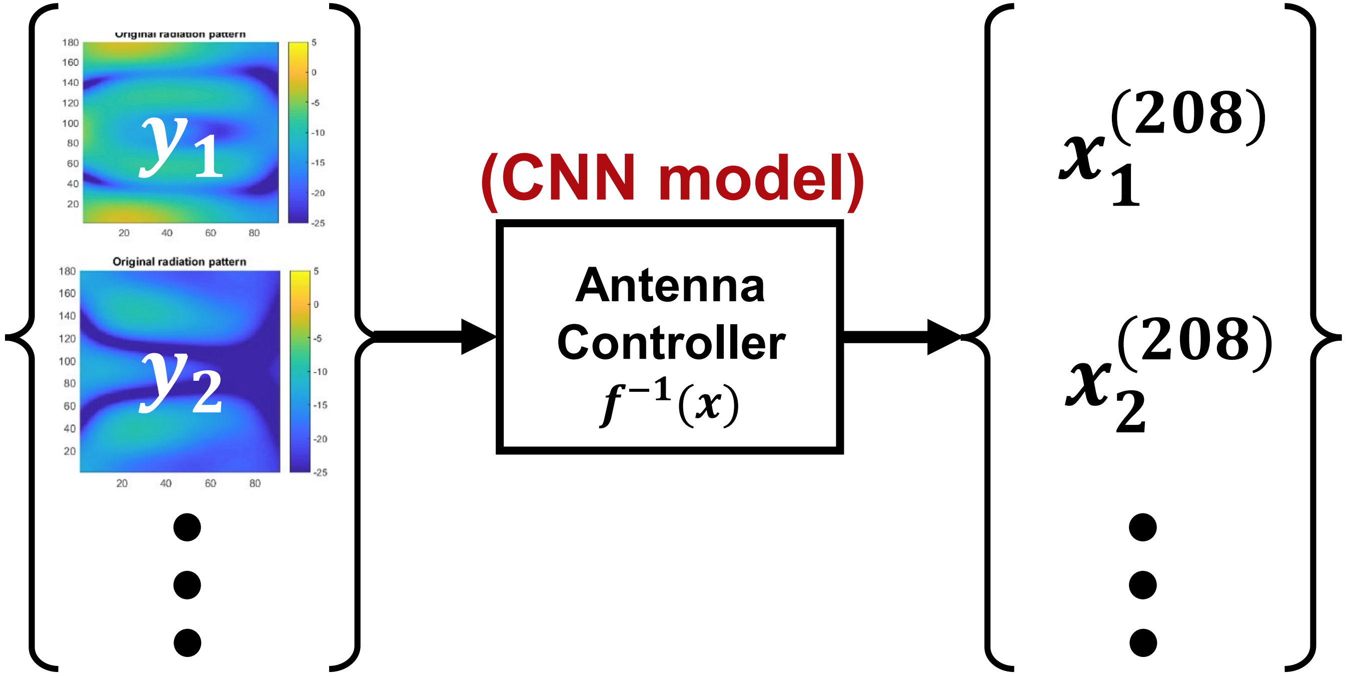
- **Embedded sensors introduce structure**
  - Signal invariance
  - Spatial invariance
- **Reconfigurable spatial accessing enables algorithmic control of sensing**
- **Exploit sensor-data structure for feature extraction (sensor fusion models)**

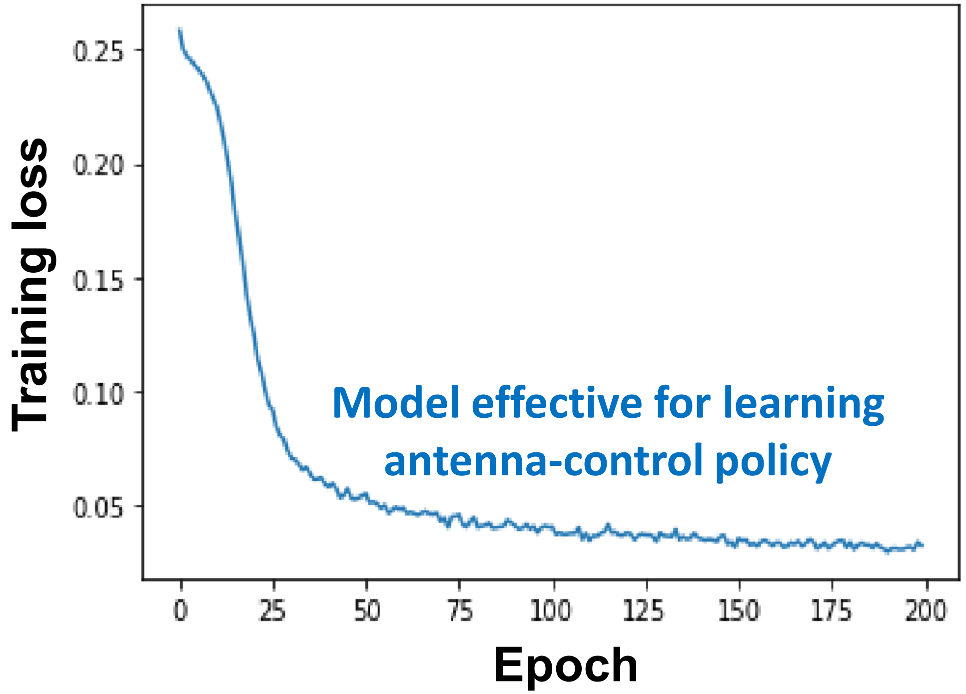
Closed-loop Wireless Sensing:



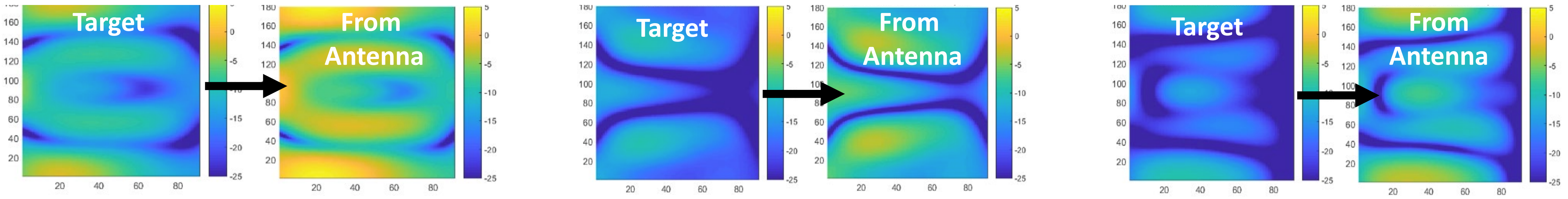
Models for high-dimensional antenna control

- **Antenna control via CNN for modeling 2-D antenna physics (spatial semantics)**





Example Antenna Radiation Patterns



Conclusions

The real world presents statistically-complex processes

⟶ structure in data is showing profound potential for enhancing learning

The real world presents rich structure

⟶ preserving real-world structure in sensor data can enhance machine perception

Preserving real-world structure requires specialized sensing technologies

⟶ large-area electronics (LAE) enables expansive, form-fitting sensing

Sensor-algorithm co-design enables structure and exploitation of structure

⟶ this will lead to specialized ML models for sensor fusion & task-driven control

Acknowledgements: funding provided by DARPA, JUMP/SRC