Energy - Efficient AI

Processing at the Edge Instead of in the Cloud

Communication

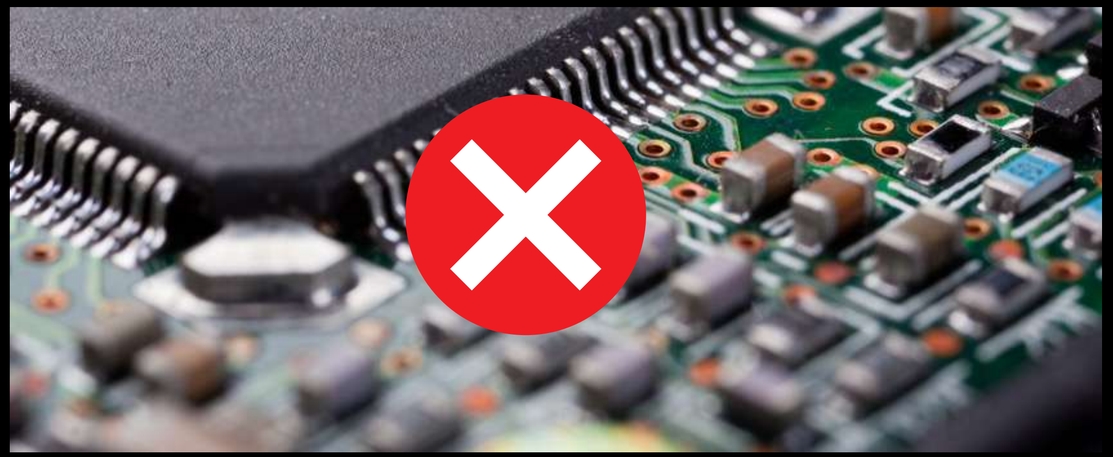
Privacy

Latency

Self-driving car prototypes use approximately 2,500 Watts of computing power.

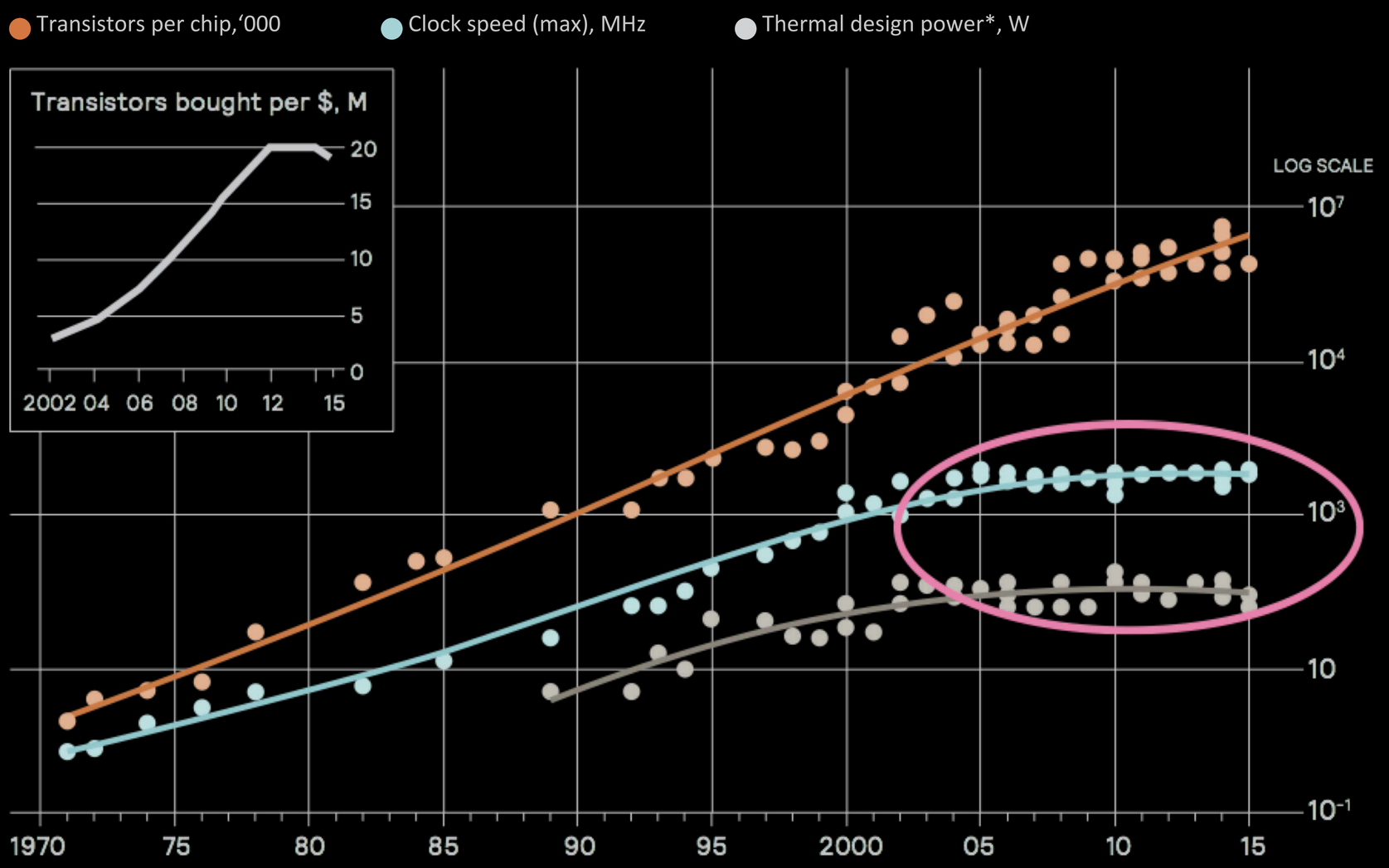
Existing Processors Consume Too Much Power

< 1 Watt



> 10 Watts

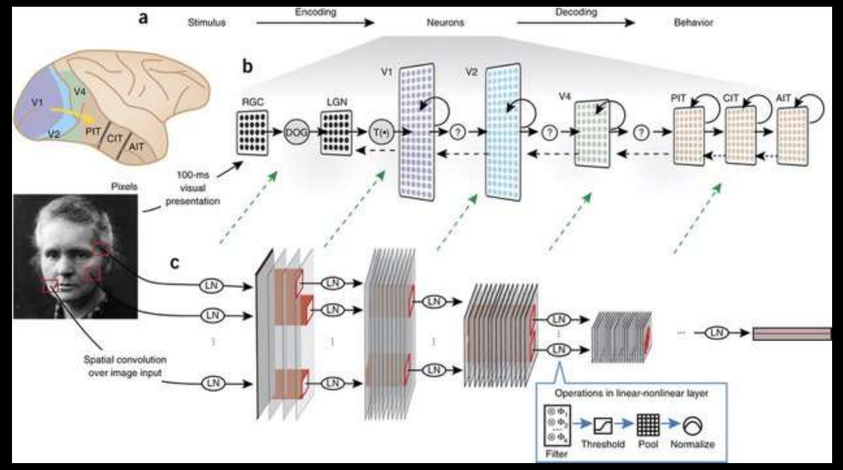
Transistors Are Not Becoming More Efficient



SOURCE: INTEL, PRESS REPORTS, BOB COLWELL, LINLEY GROUP, IB CONSULTING, *THE ECONOMIST*

*MAXIMUM SAFE POWER CONSUMPTION

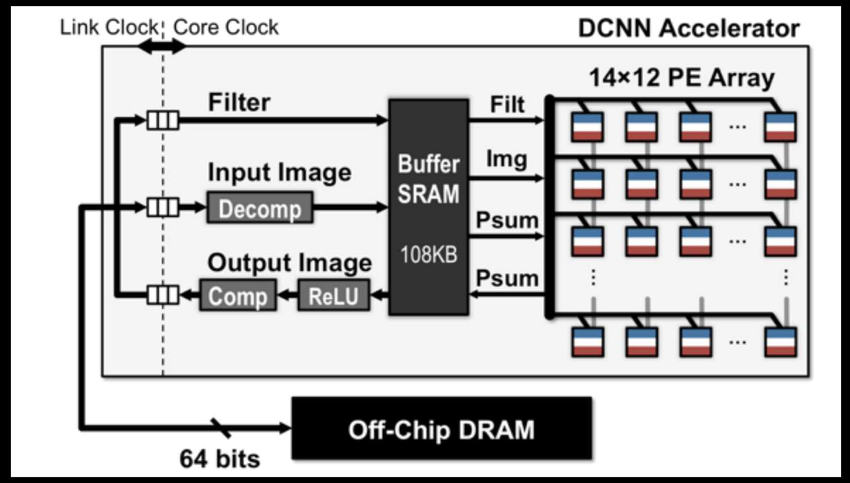
Energy-Efficient AI with Cross-Layer Design



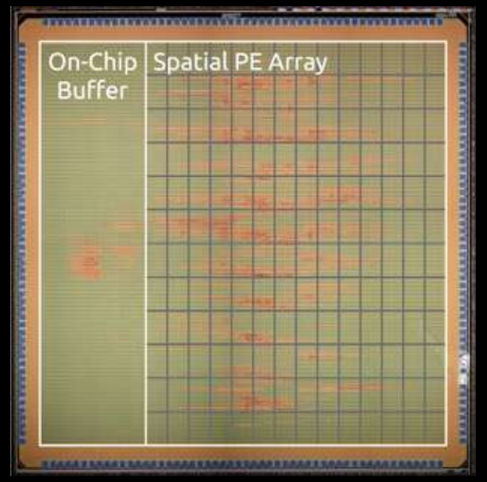
Algorithms

Systems

Specialized Compute Hardware

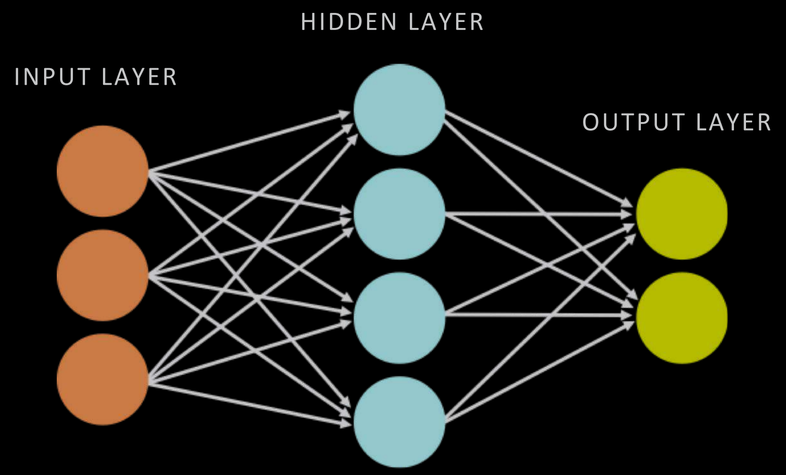


Computer Architectures



Circuits

Deep Neural Networks

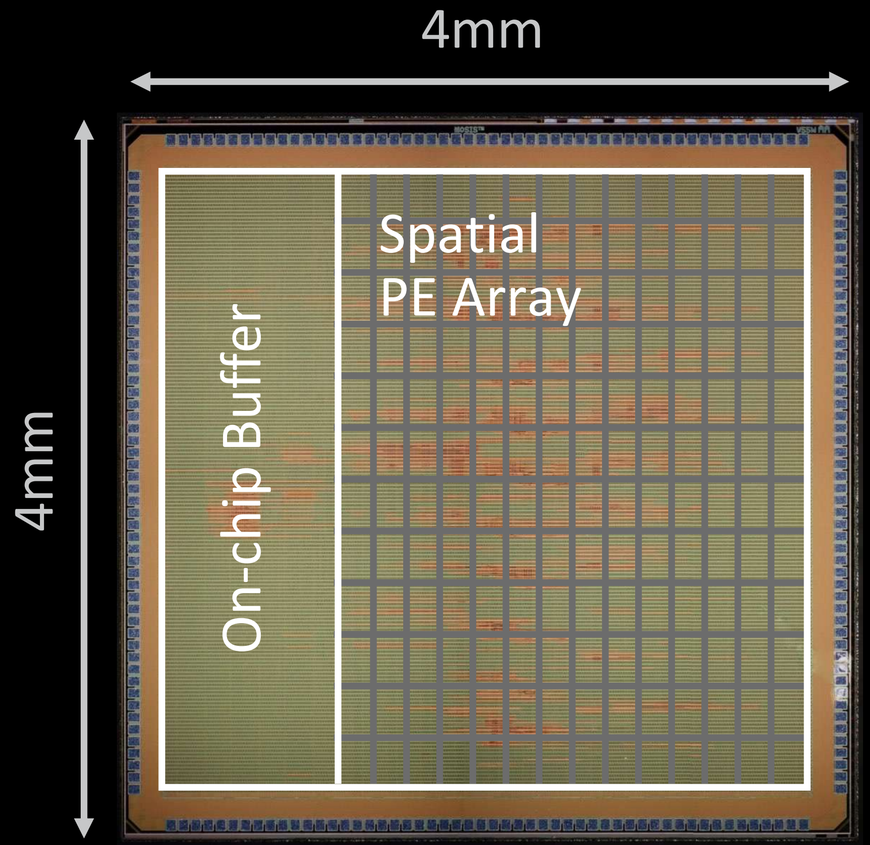


Cornerstone of AI applications

Specialized hardware focuses on reducing data movement and memory accesses

Image classification under a third of a Watt

> 10x energy reduction comparable to mobile GPU

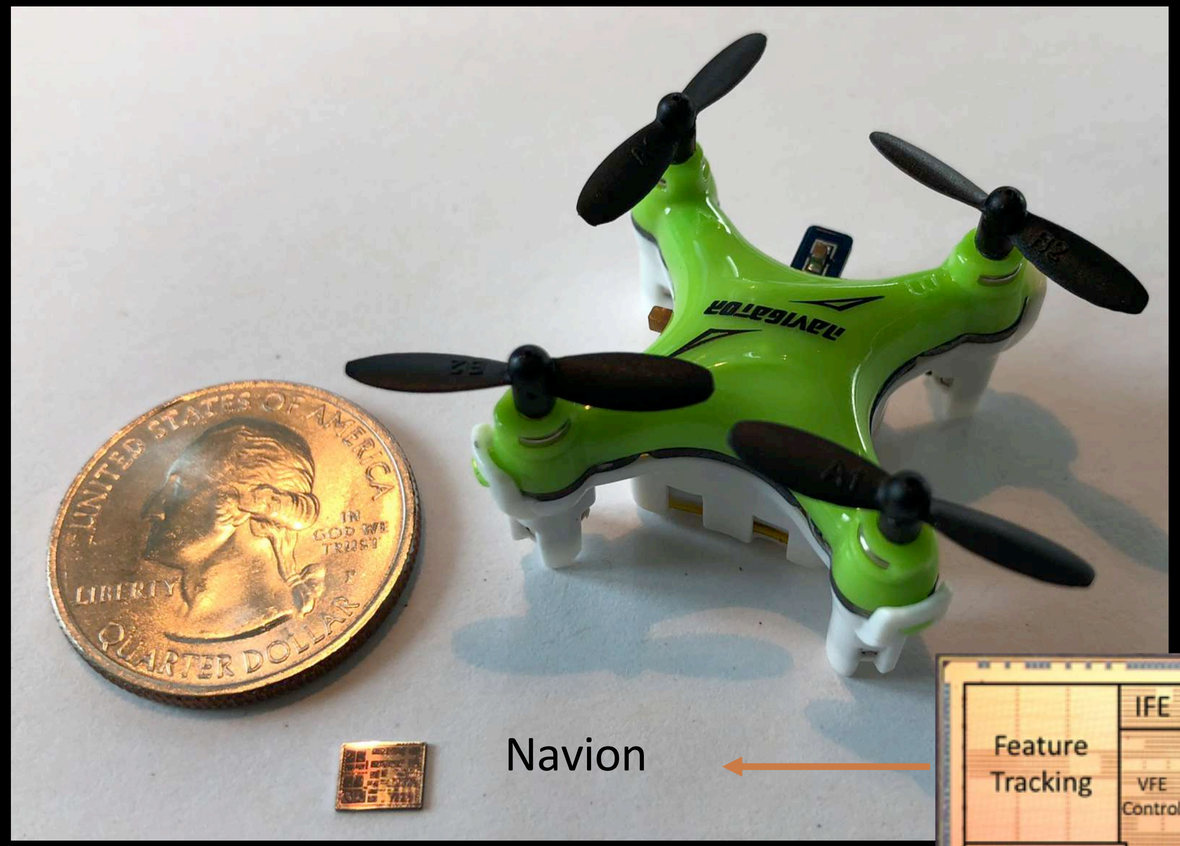


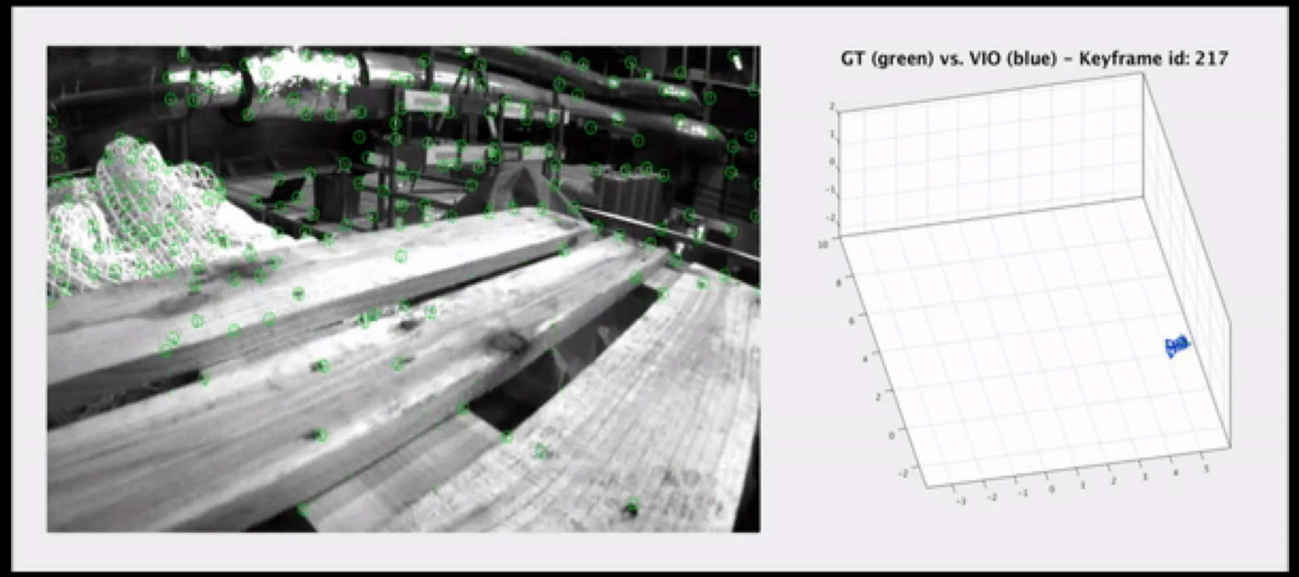
Eyeriss

JOINT WORK WITH JOEL EMER

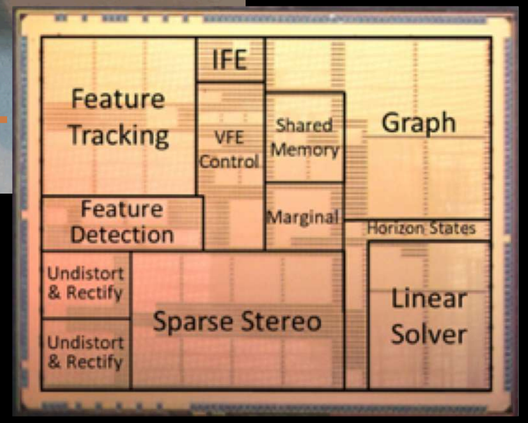
SOURCE: CHEN, ISSCC 2016

Robot Localization in under a Tenth of a Watt





EUROC DATASET



Navion: Fully integrated system — no off-chip processing or storage!

JOINT WORK WITH SERTAC KARAMAN

SOURCE: ZHANG, RSS 2017; SULEIMAN, VLSI 2018

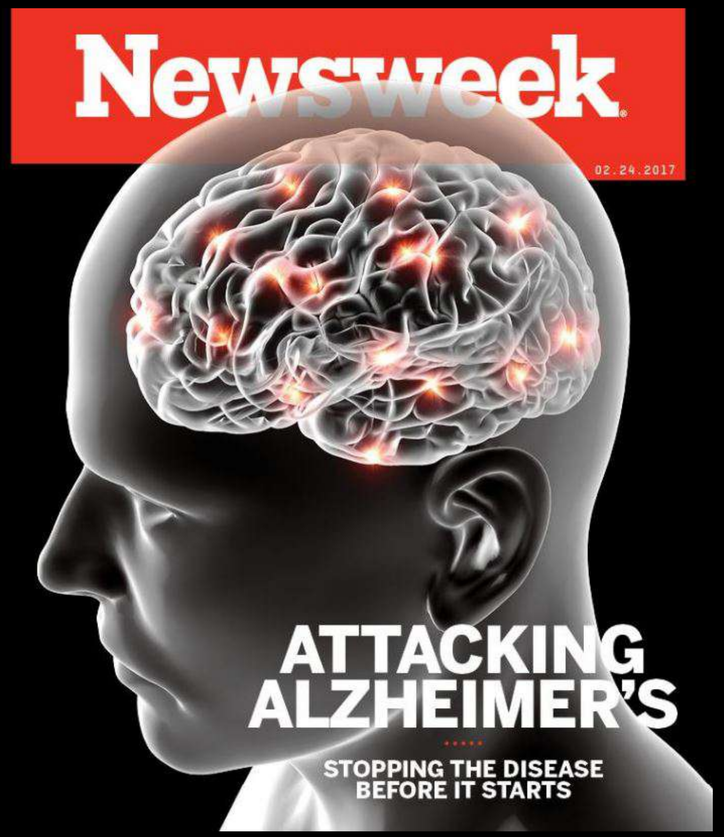
Low Energy Robotics

Lighter than Air Vehicles

Miniature Satellites

Origami Robots

Monitoring Neurodegenerative Diseases



SOURCE: NEWSWEEK

Dementia affects 50 million people worldwide today (75 million in 10 years) [World Alzheimer's Report]

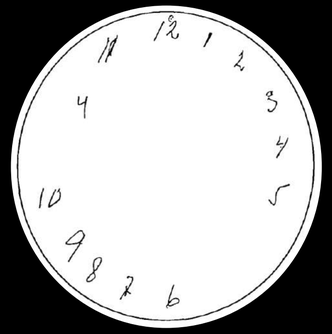
Mini-Mental State Examination (MMSE)

Q1. What is the year? Season? Date?

Q2. Where are you now? State? Floor?

Q3. Could you count backward from 100 by sevens? (93, 86, …)

Clock-drawing test



Agrell et al.
Age and Ageing, 1998.

- Neuropsychological assessments are time consuming and require a trained specialist
- Repeat medical assessments are sparse, mostly qualitative, and suffer from high retest variability

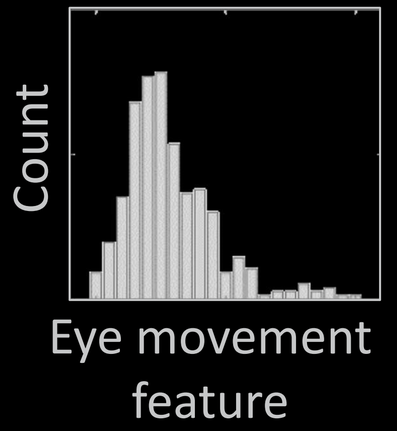
JOINT WORK WITH THOMAS HELDT

Use Eye Movements for Quantitative Evaluation

Eye movements can be used to quantitatively evaluate severity, progression or regression of neurodegenerative diseases

Eye movements

Smart phone



We are investigating how to perform eye movement tests on a smart phone in order to enable low-cost, in-home measurements

JOINT WORK WITH THOMAS HELDT

SOURCE: LAI, JHBI 2019

Summary

Energy-Efficient AI extends the reach of AI beyond the cloud
➔ Critical step to making AI ubiquitous!

A cross-layer approach from efficient algorithms to specialized hardware is required to enable energy-efficient AI
➔ Critical to the progress of AI over the next decade!