VIVIENNE SZE

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.

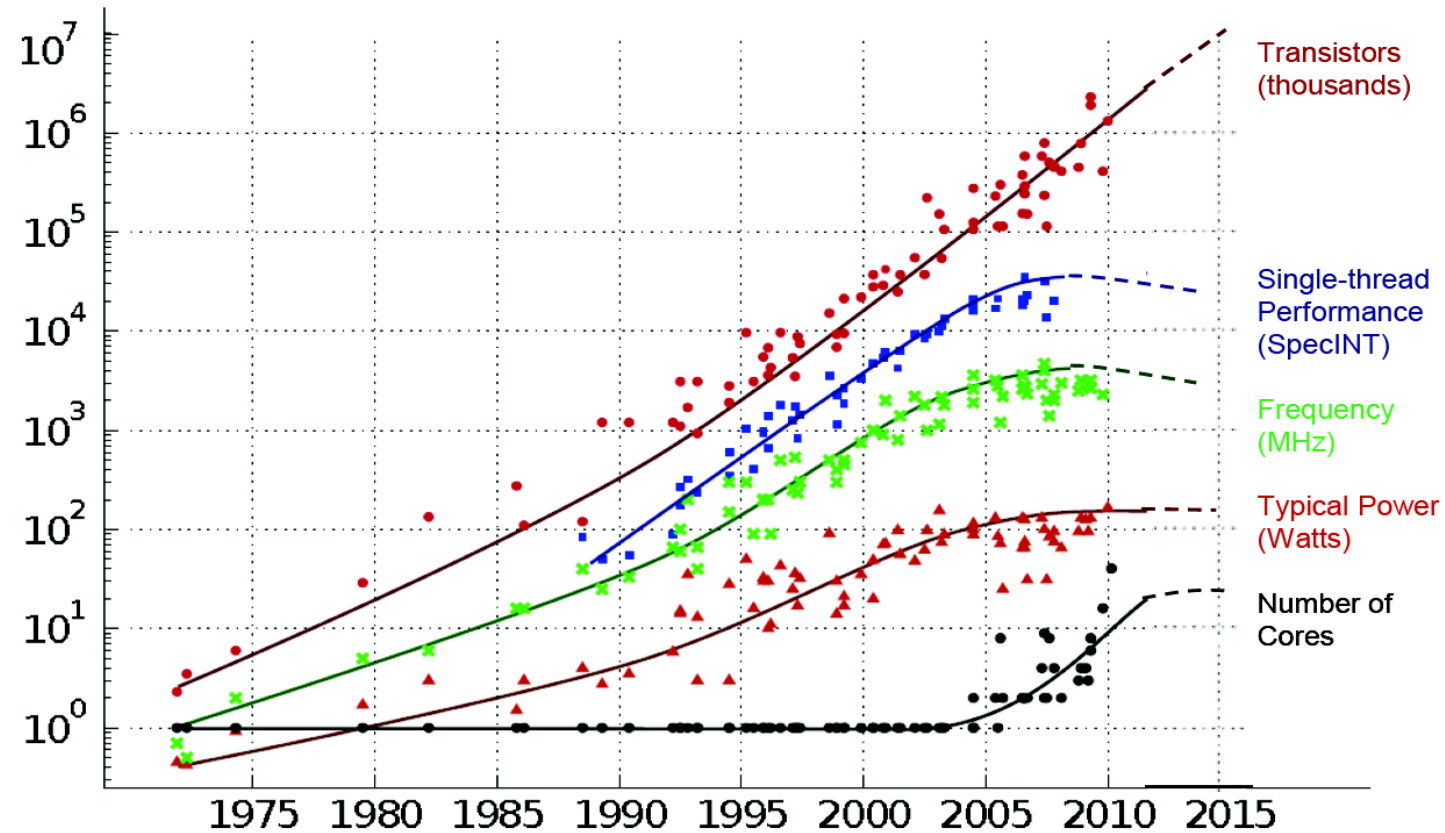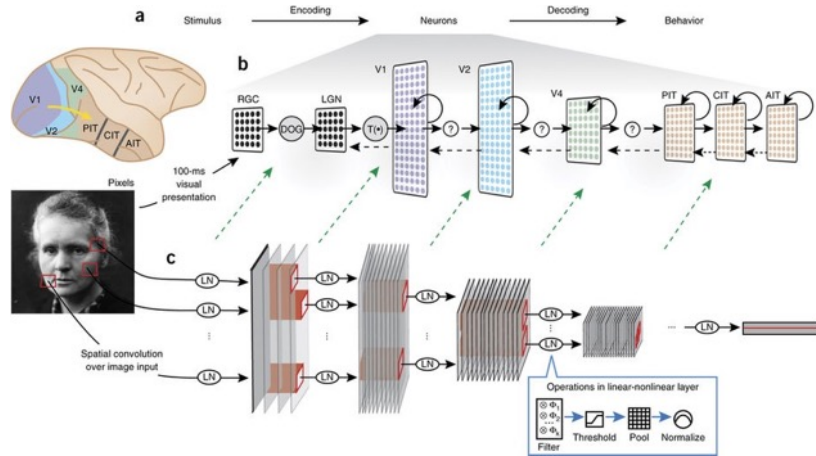# Existing Processors Consume Too Much Power

< 1 Watt
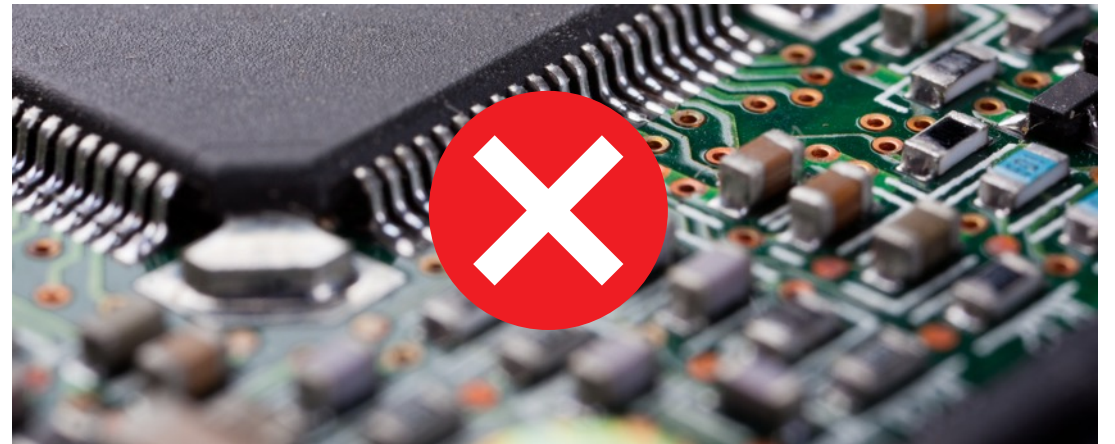
> 10 Watts

35 YEARS OF MICROPROCESSOR TREND DATA

Original data collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond and C. Batten
Dotted line extrapolations by C. Moore

Transistors Are Not Becoming More Efficient

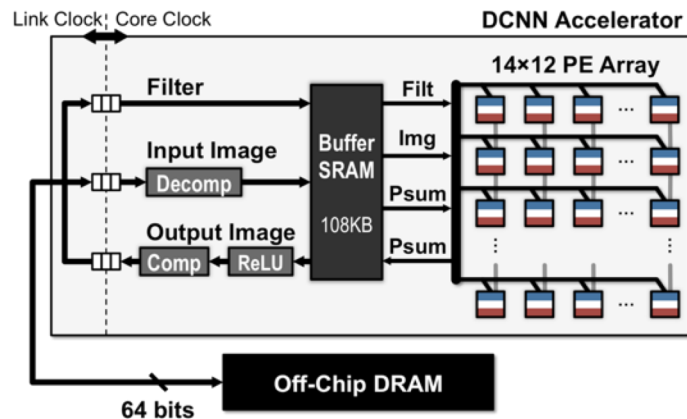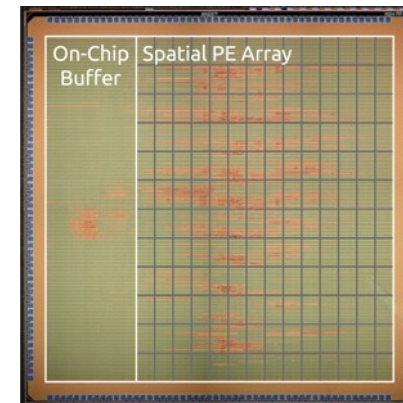# Energy-Efficient AI with Cross-Layer Design

Algorithms

Systems

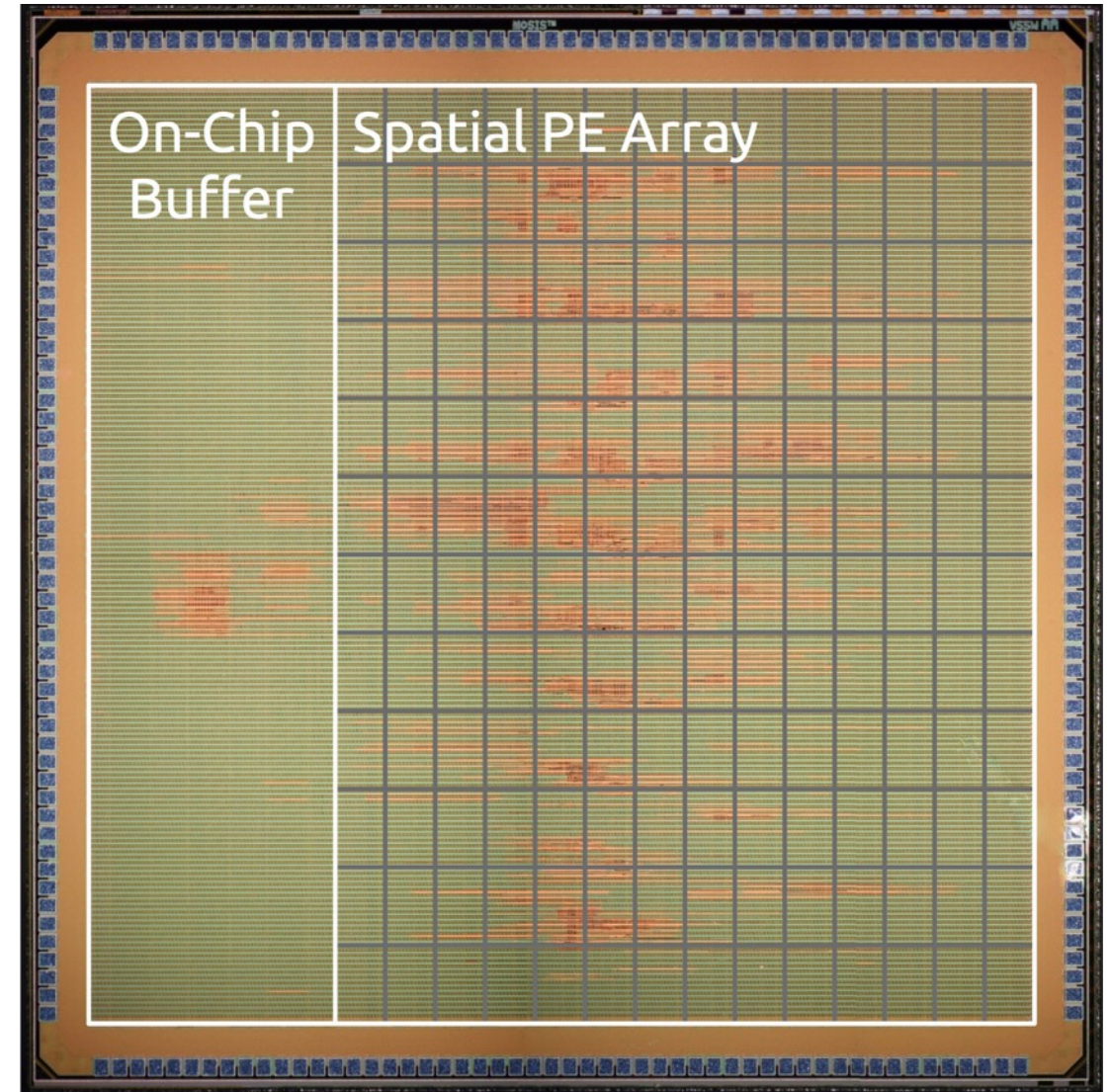## Specialized Compute Hardware

Compute 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
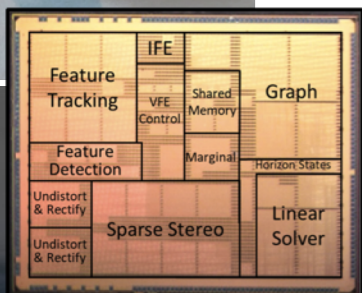
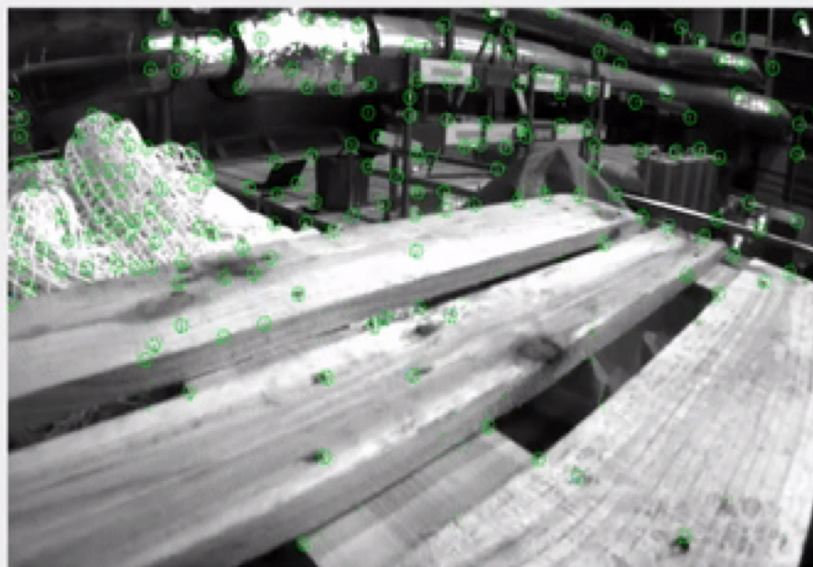- > 10x energy reduction comparable to mobile GPU

Eyeriss

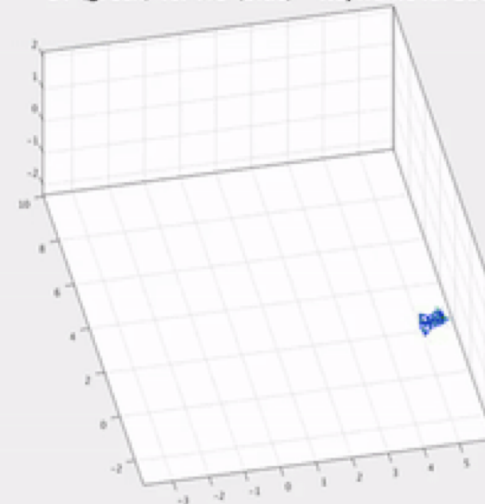# Robot Localization Under a Tenth of a Watt

Navion

Sertac Karaman

GT (green) vs. VIO (blue) – Keyframe id: 217

[EuRoC Dataset]

Low Energy Robotics

[Georgia Tech]

Lighter than Air Vehicles

Miniature Satellites

Origami Robots

Neuropsychological Testing

- Important component in treatment of:
  - traumatic brain injury
  - dementia
  - psychiatric disorders

- Assessments are:
  - pencil and paper test
  - time consuming
  - require a trained specialist

- Repeat assessments are:
  - sparse
  - mostly qualitative
  - high retest variability

Thomas Heldt

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

"Whatever it is that [chip vendors and researchers] build will influence the progress of AI over the next decade."

Yann LeCun, Facebook's chief AI scientist