# Ternarized TCN for µJ/Inference Gesture Recognition from DVS Event Frames

**Georg Rutishauser,** Moritz Scherer, Tim Fischer, Luca Benini
Integrated Systems Lab (IIS), ETH Zurich

DATE 2022
16.-23.03.2022

# Edge AI, TinyML, Smart Sensing: Energy Efficiency is Everything

- Insatiable hunger for data – number of IoT (sensor) nodes in use exploding:
  **83 Bn. devices by 2024!** [1]
- On-device processing of collected data: **Edge AI**
- This scaling can only be sustained if nodes become even more:
  - Cheap → MCU-class systems
  - Powerful → Acceleration of core algorithms
  - Versatile → battery-powered → **energy efficiency is key!**

**Efficient sensor nodes:**
**Efficient sensing + efficient processing!**

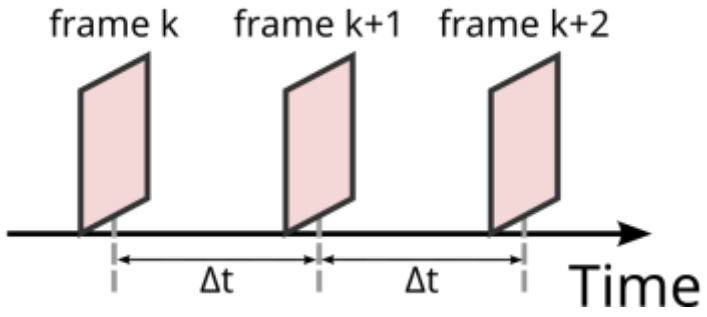


[1]: https://www.juniperresearch.com/press/iot-connections-to-reach-83-bn-by-2024

# Energy-Proportional Visual Sensing with DVS Cameras

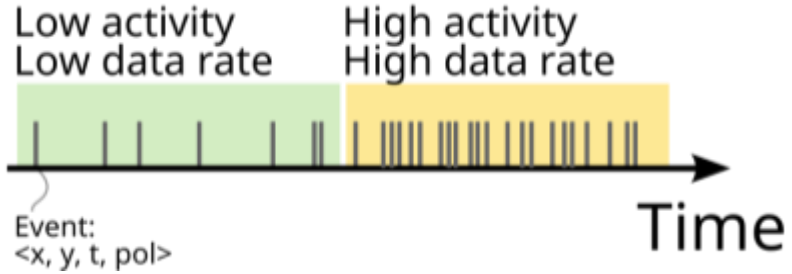
- Energy-efficient sensing → Energy-proportional sensing:
  Only transmit information about changes in Scene

## Conventional Camera: Fixed Data Rate




Output:
- **Fixed-size frames**
- **Fixed frame rate**
  **-> fixed data rate**

## Dynamic Vision Sensor: Dynamic Data Rate



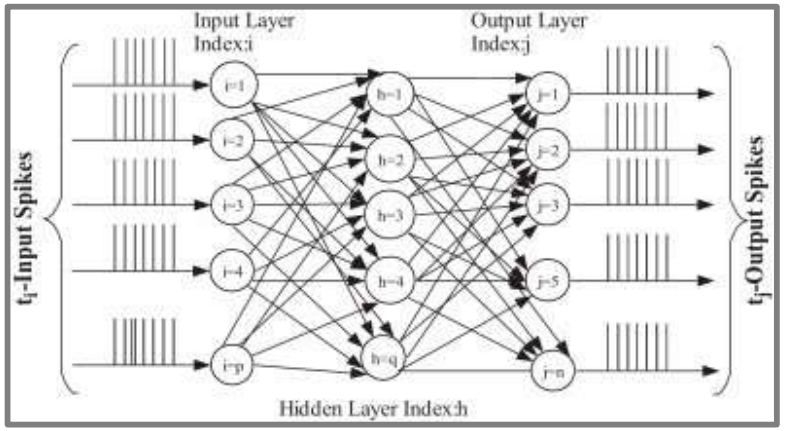

Output:
- **Binary events describing polarity of brightness change**
- **Asynchronous events**
  **→ activity-proportional event rate**

# Processing DVS Event Data: Challenging the Event-Based Paradigm

- Conventional wisdom: Event-based processing for event data
  → **Spiking Neural Networks (SNN)**
- Big-name SNN accelerators: **Large** & **Expensive**!
  - Research is working on smaller, more efficient architectures
- **Classical DNNs** have become extremely efficient:
  - Aggressive quantization: 1-bit, 2-bit
  - Ultra-efficient acceleration: **100s TOp/s/W**

**Can we use highly quantized DNNs on DVS data to combine:**
- Efficiency of low-precision DNNs
- Energy-proportional sensing of DVS?

**We decided to find out!**




[2]

Deploy


[3]

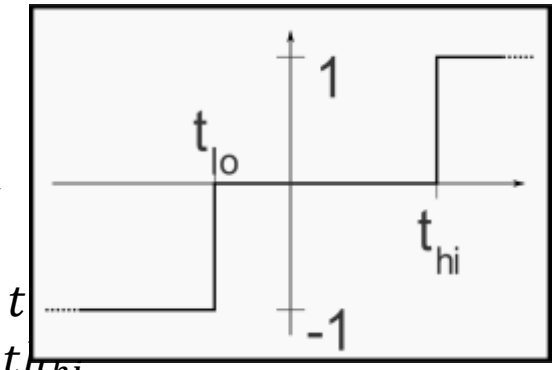

[1]

[1]: https://www.intel.com/
[2]: https://towardsdatascience.com/spiking-neural-networks-the-next-generation-of-machine-learning-84e167f4eb2b
[3]: https://www.top500.org/

# Ternary Neural Networks: The Big Brother of BNNs

- All parameters and activations $\in \{-1, 0, 1\}$
- $W \in \{-1, 0, 1\}^l, x \in \{-1, 0, 1\}^m \rightarrow y = W * x \in \mathbb{Z}^n$
- Activation function: Thresholding

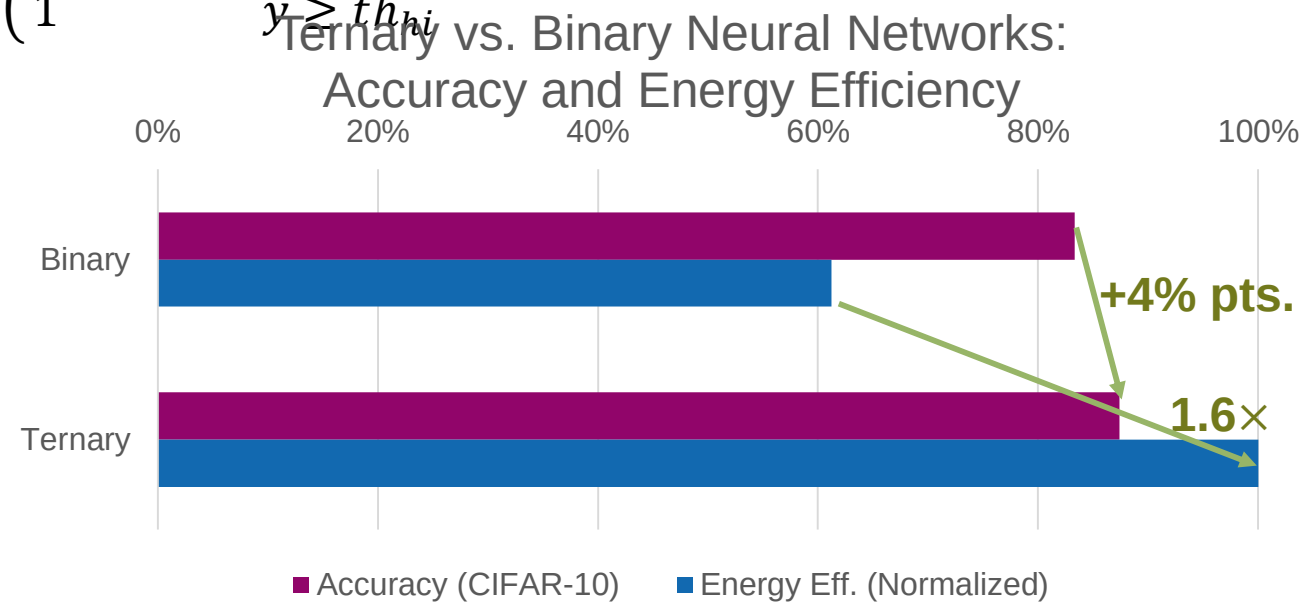
$$A(y) = \begin{cases} -1 & y < th_{lo} \\ 0 & th_{lo} \leq y < th_{hi} \\ 1 & y \geq th_{hi} \end{cases}$$




TNNs vs. BNNs:

- Better accuracy...
- ...and better energy efficiency!

**...IF you have the right accelerator!**




Data from:
Scherer et al., "CUTIE: Beyond PetaOp/s/W Ternary DNN Inference Acceleration with Better-than-Binary Energy Efficiency"
IEEE TCAD/ArXiV 2011.01713

# CUTIE: Accelerating TNNs with no Compromises

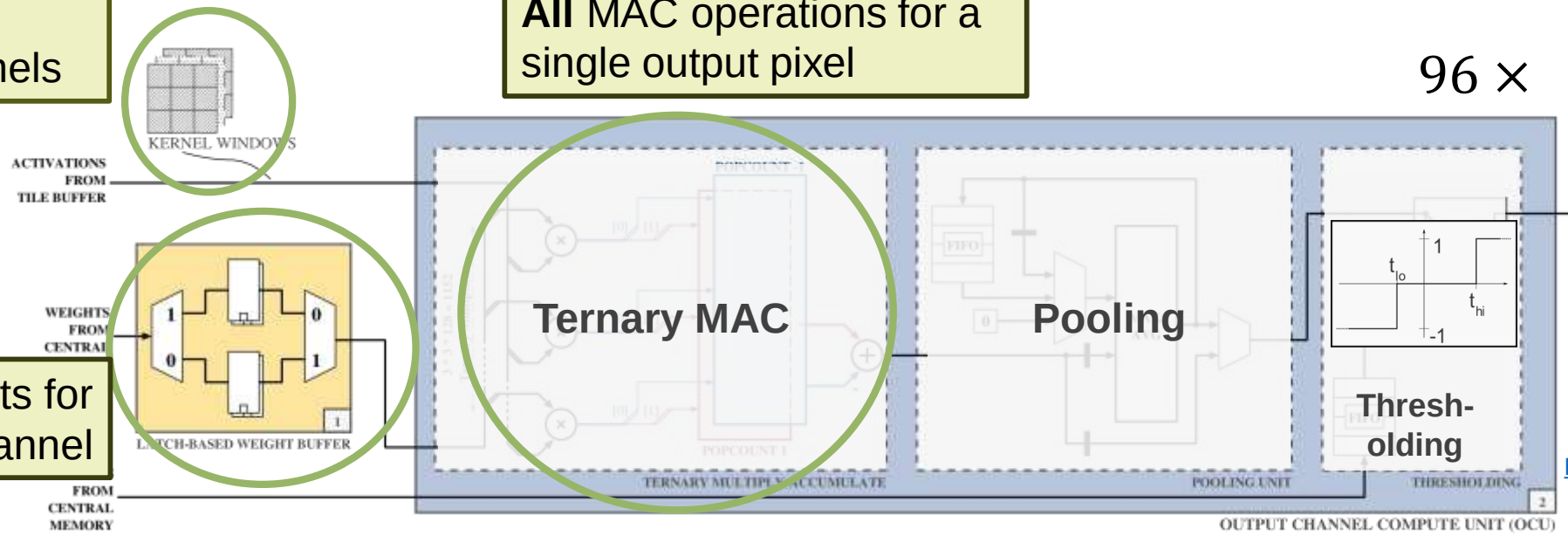
## **CUTIE**: ***C**ompletely **U**nrolled* **T**ernary **I**nference **E**ngine

- **No iterative computations** – calculate entire filter step in one go!
- ~90% of dynamic energy goes into computation – not data movement!
- SoA energy efficiency:
  - **~300 TOp/s/W** for 22nm implementation
  - **3.6 µJ/Inference** on CIFAR-10

Filter inputs of **all** input channels

**All** MAC operations for a single output pixel

96 ×




**All** weights for one output channel

Read the CUTIE paper:

https://arxiv.org/abs/2011.01713

# The Event Frame Processing Pipeline

**Energy-proportional sensing**

**Binary events → ternary event frames**

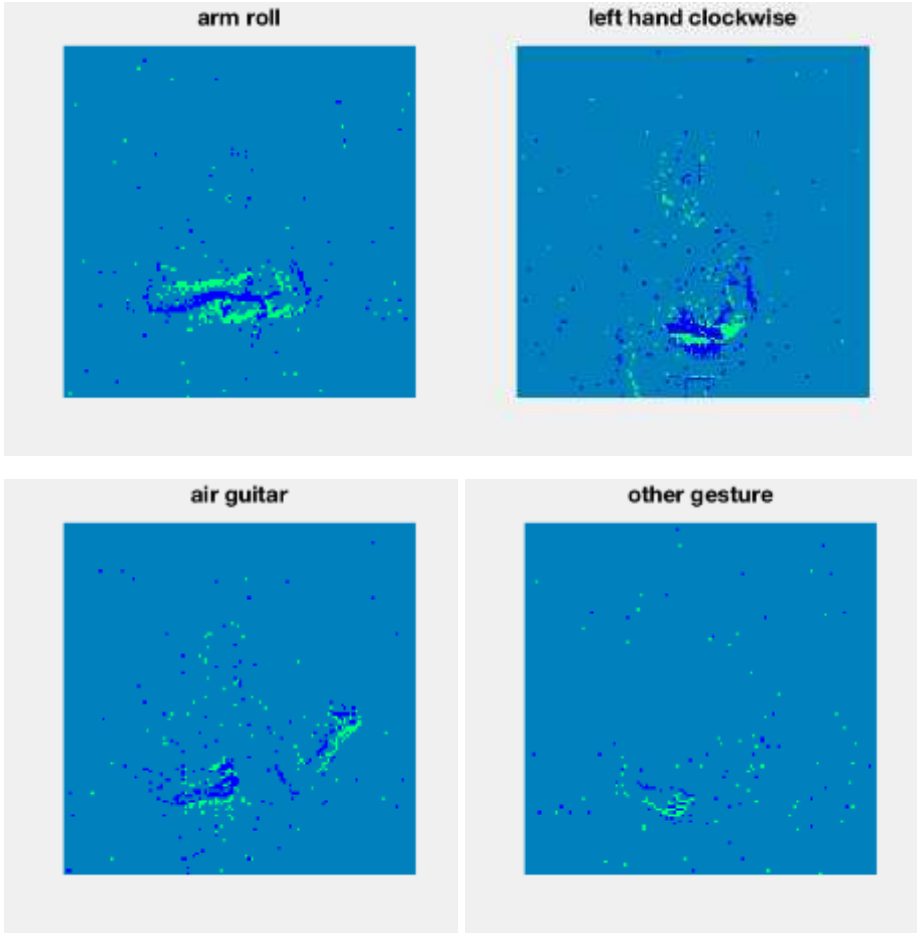
**Map to CUTIE accelerator for ultra-efficient inference!**

**Contributions:**

- Event frame-based processing pipeline for video classification of DVS event data
- Ternarized CNN/TCN architecture with **SoA accuracy (94.5%)** on DVS128 dataset (11-class)
- Hardware implementation in GF 22nm FDX, **500× lower inference energy (2.2 µJ/inf.)** vs. SoA
- Integrated and taped out as part of our latest RISC-V based MCU platform!

# The Problem: DVS-Based Gesture Recognition on DVS128 Dataset

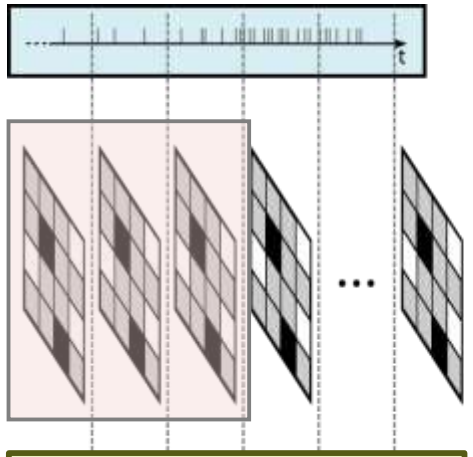
- Published in [6]
- 29 subjects recorded with DVS128 camera
  - Various lighting conditions
  - 122 samples per class
- 10+1 classes:
  - 10 pre-defined gestures
  - 1 “random gesture” class freely chosen by each subject → noise class
- **High-quality DVS video classification dataset!**




[6]: A. Amir *et al.*, “A Low Power, Fully Event-Based Gesture Recognition System”, CVPR 2017

https://research.ibm.com/interactive/dvsgesture/

# The Event Frame Processing Pipeline




Events are accumulated into ternary event frames

Ternarized CNN encodes short-term information...

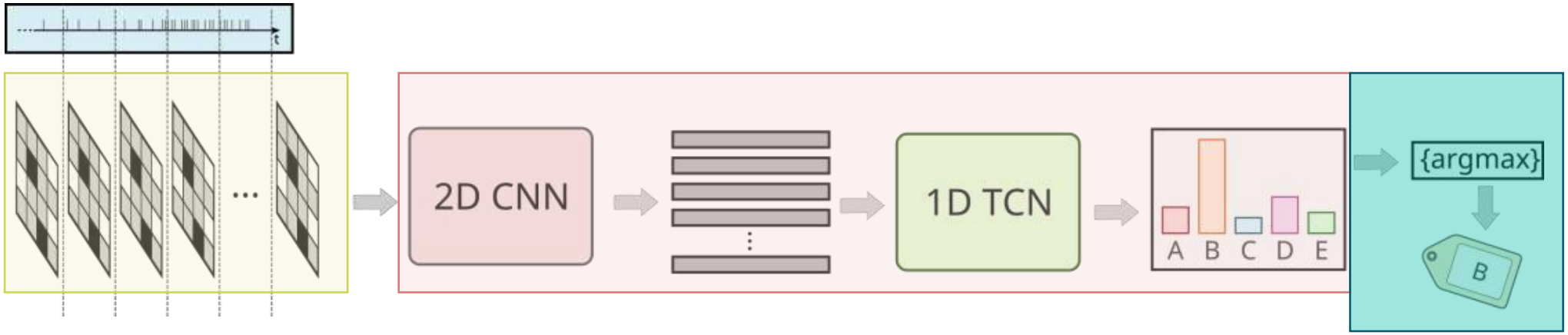
...into 1D ternary feature vectors

Ternarized TCN analyses longer-term information

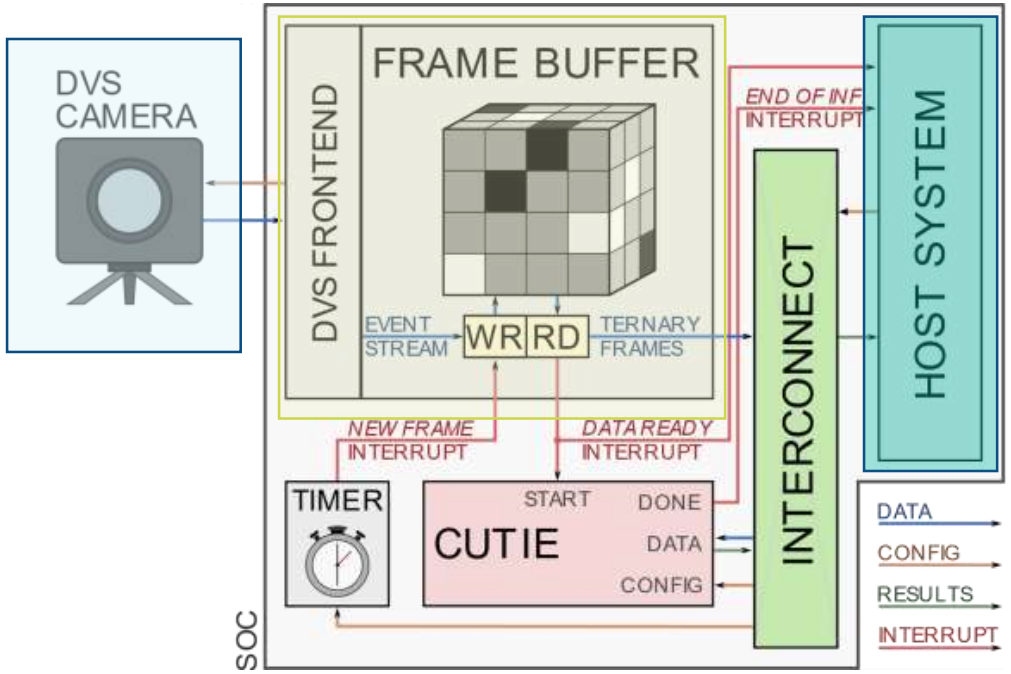
And assigns class scores

- **Hybrid architecture:** Combine fully ternarized CNN and TCN
- **CNN:** encodes **short-term information** into 1D ternary features
- **TCN: longer-term** temporal context + **classification**

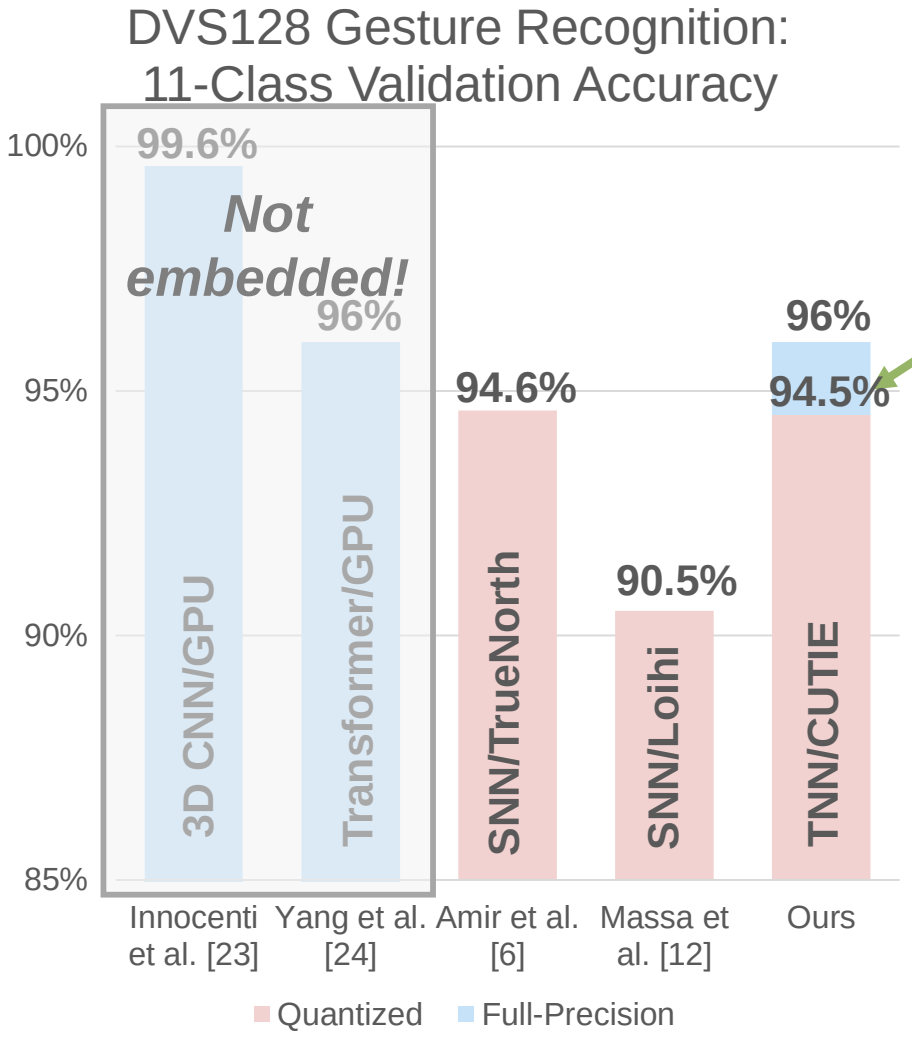
# A Low-Power System for DVS Event Frame Classification




- Frame buffer: configurable, triggers inference autonomously
- Full network runs on modified CUTIE accelerator
- Host core only needs to calculate argmax



# Ternarized DNNs Can Achieve SoA Accuracy on DVS128




**Ternarization Drop: -1.5% pts. → Problem well-suited for TNNs!**

**94.5% Valid. accuracy: SoA for embedded implementations!**

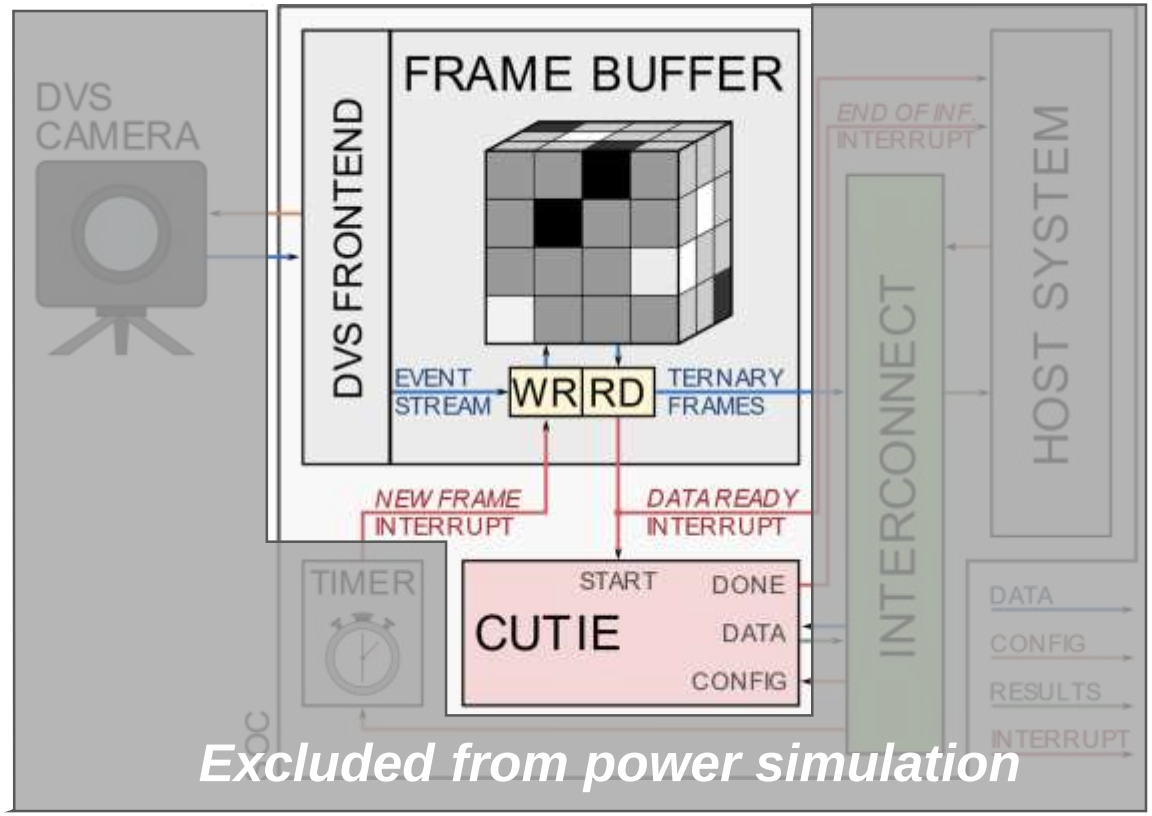
- Dataset: **11-class DVS128**
  - Training set: users 1-23
  - Validation set: users 24-29
- Network parameters:
  - 30 FPS <-> 33 ms frame time
  - 4x downsampling → input resolution 64x64
  - Receptive time window: 667 ms
  - New classification every 100 ms


[6]: A. Amir *et al.*, "A Low Power, Fully Event-Based Gesture Recognition System", CVPR 2017
[12]: R. Massa *et al.*, "An Efficient Spiking Neural Network for Recognizing Gestures with a DVS Camera on the Loihi Neuromorphic Processor", IJCNN 2020
[23]: S.U. Innocenti *et al.*, "Temporal Binary Representation for Event-Based Action Recognition", ICPR 2020
[24]: J. Yang *et al.*, "Modeling Point Clouds with Self-Attention and Gumbel Subset Sampling", CVPR 2020

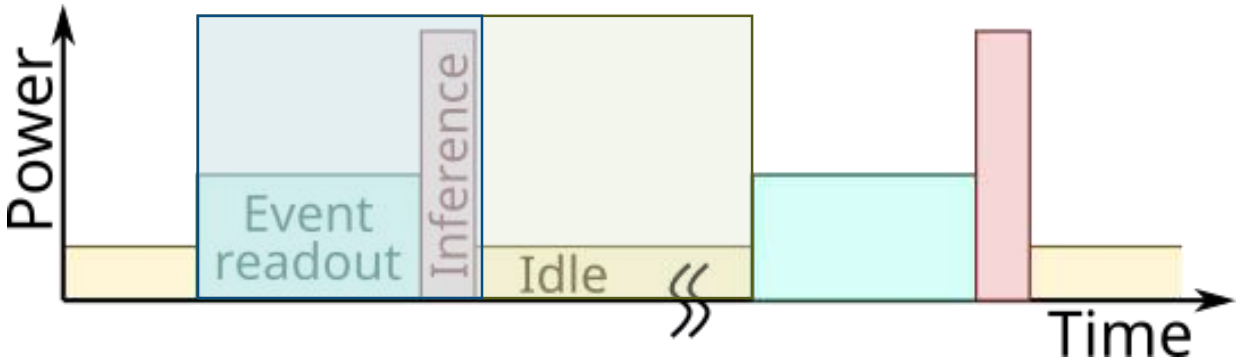
# Energy Efficiency: Measurement Setup

**Post-synthesis power simulation:**

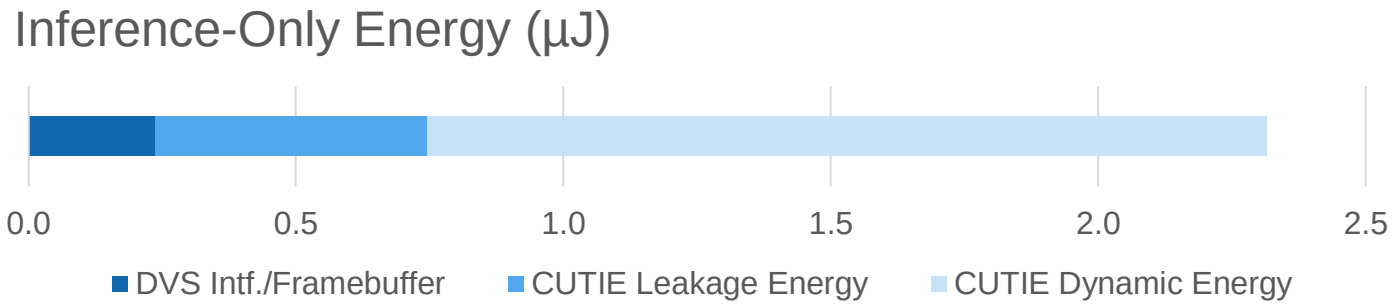
- System synthesized in GF 22nm FDX
- $V_{DD}$ = 0.8 V, scaled to 0.65 V
- TT corner
- $f_{clk}$:
  - DVS intf./frame buffer @ 50 MHz
  - CUTIE @ 17.6 MHz
- Only consider power consumed by:
  - **DVS interface**
  - **Frame buffer**
  - **CUTIE accelerator**



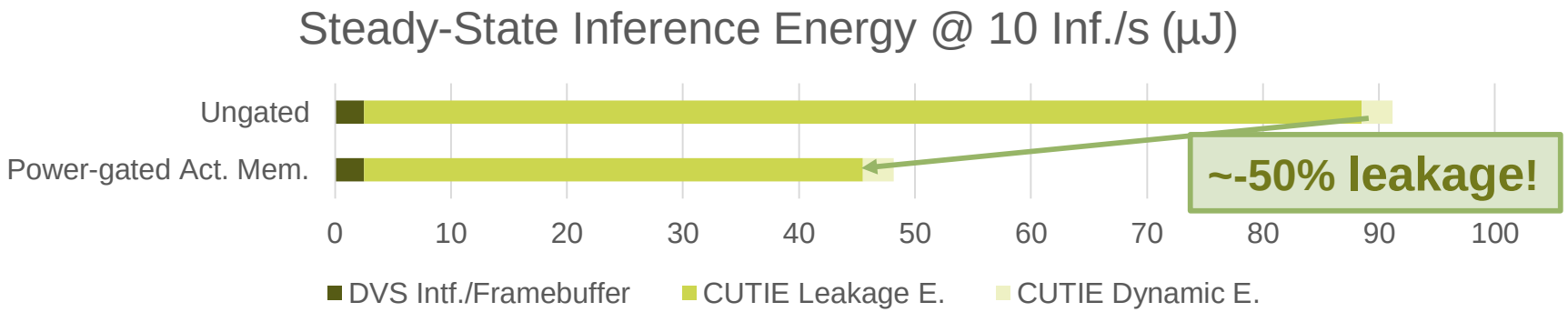
# Inference Energy: It depends how you measure!




**Inference time: ~300 µs**
**Frame interval : 33.3 ms**
**→ ~99% of time in idle!**
**Total energy/inference *including idle?***



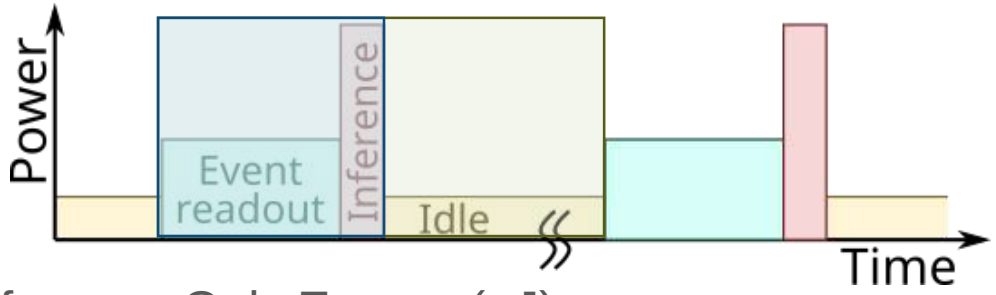

**2.2 µJ/inference* – excluding idle power**

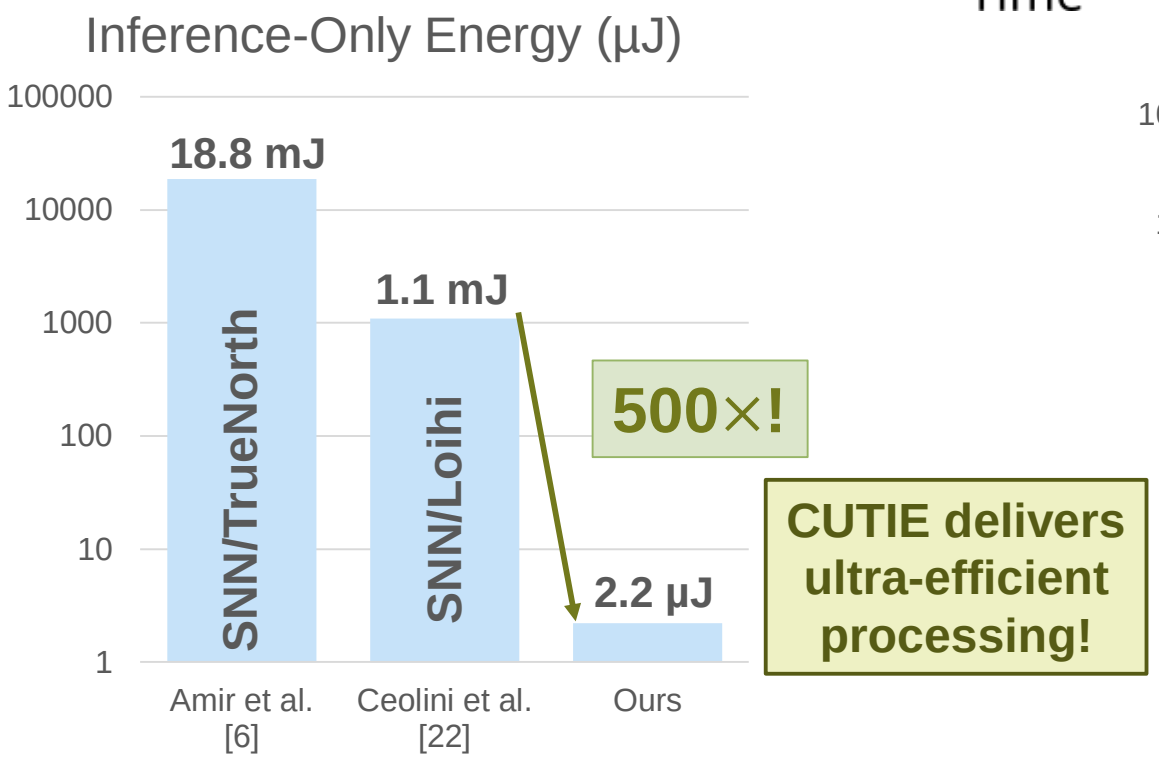



**Steady-state inference energy: 92 µJ**
***Leakage-dominated!***
**Remedy: Power-gate activation memories!**

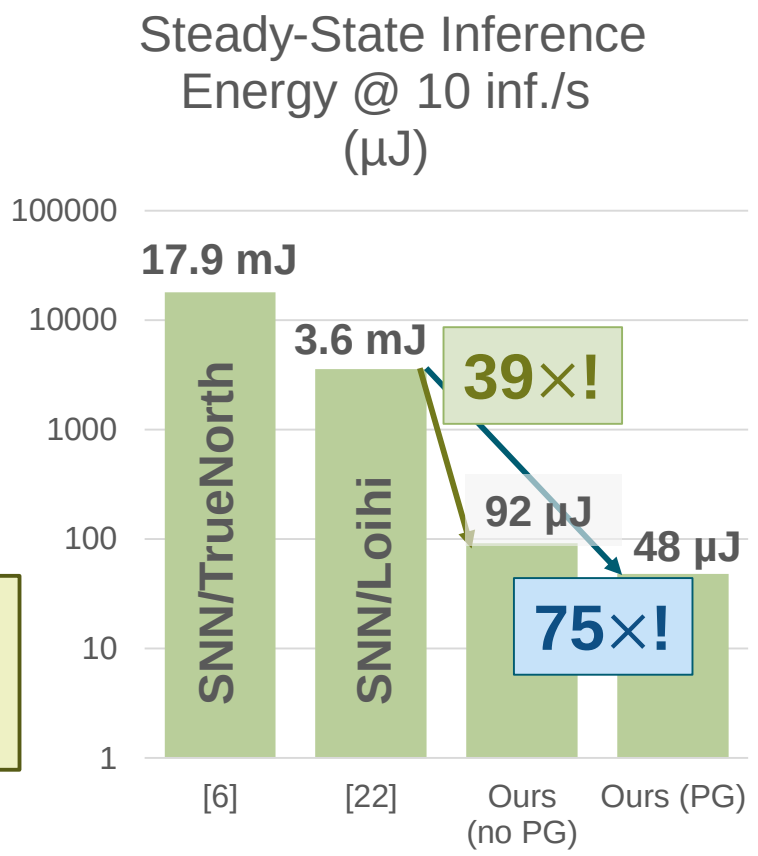
*due to a miscalculation in the submitted paper, this figure has changed from 1.7 µJ to 2.2 µJ

# Inference Energy: Comparison to State of the Art




**CUTIE delivers ultra-efficient processing!**




**Including idle power narrows the gap...**

**...but it can be improved by power-gating volatile memories**


[6]: A. Amir *et al.*, "A Low Power, Fully Event-Based Gesture Recognition System", CVPR 2017
[22]: E. Ceolini *et al.*, "Hand-Gesture Recognition Based on EMG and Event-Based Camera Sensor Fusion: A Benchmark in Neuromorphic Computing", Frontiers in Neuroscience, 2020

# Conclusion: DVS + TNN Is a Great Match!

Process DVS data as event frames with ternarized DNNs to:

- Reach SoA classification accuracy of **94.5%**
- Reduce inference-only energy by **500×** by mapping to CUTIE

Leakage dominates steady-state power at 30 FPS:

- **Power gate leaky SRAMs** to mitigate idle consumption
- Opportunity: CUTIE's high throughput allows **1000's of inferences/s**
  → High-frequency applications of DVS could amortize leakage by reducing idle time!

# Thanks for Your Attention!

Download these slides:

https://bit.ly/3qQEh1k

Download the paper:

https://bit.ly/3fLMEEW

Contact us:
georgr@iis.ee.ethz.ch
scheremo@iis.ee.ethz.ch
fischeti@iis.ee.ethz.ch
lbenini@iis.ee.ethz.ch

***Stay tuned for...***