# Visual Processing with Loihi 2

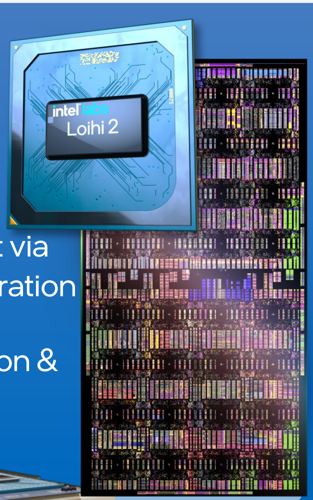
Intel Neuromorphic Computing Lab

Andreas Wild and Yulia Sandamirskaya

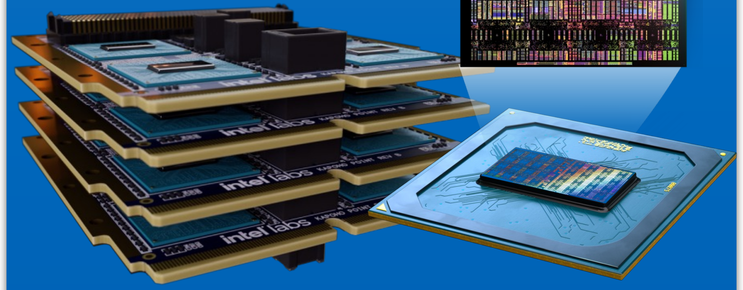
June 19, 2023, CVPR 2023 Workshop on Event-based Vision

# Loihi - Pioneering a new class of computer architecture to deliver orders of magnitude gains in energy and speed

## Loihi architecture and systems

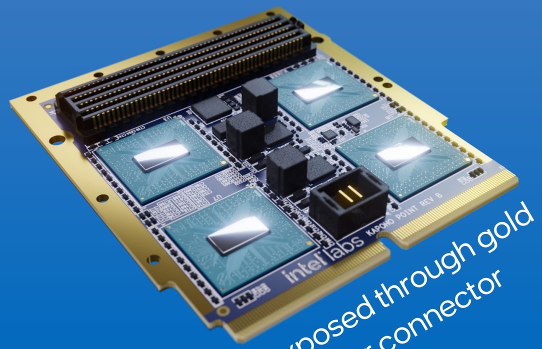



- Minimal data-movement via compute/memory integration
- Massively parallel
- Event-driven computation & communication

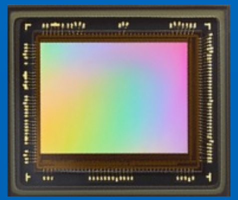


Edge ⬅ Scalable ➡ Cloud

## Event-based vision



IO exposed through gold finger connector

Coming in 2023-Q3...

## SW Products

Applications, Products, Services

| DL | Optim | VSA | ... |
|---|---|---|---|

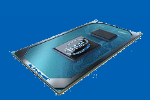
Algorithm libraries

LAVA Software Framework
github.com/lava-nc

- Event-based
- Multi-Paradigm
- Multi-Abstraction
- Multi-Platform
- Open-source

API
Compiler
Runtime

Heterogenous hardware interface

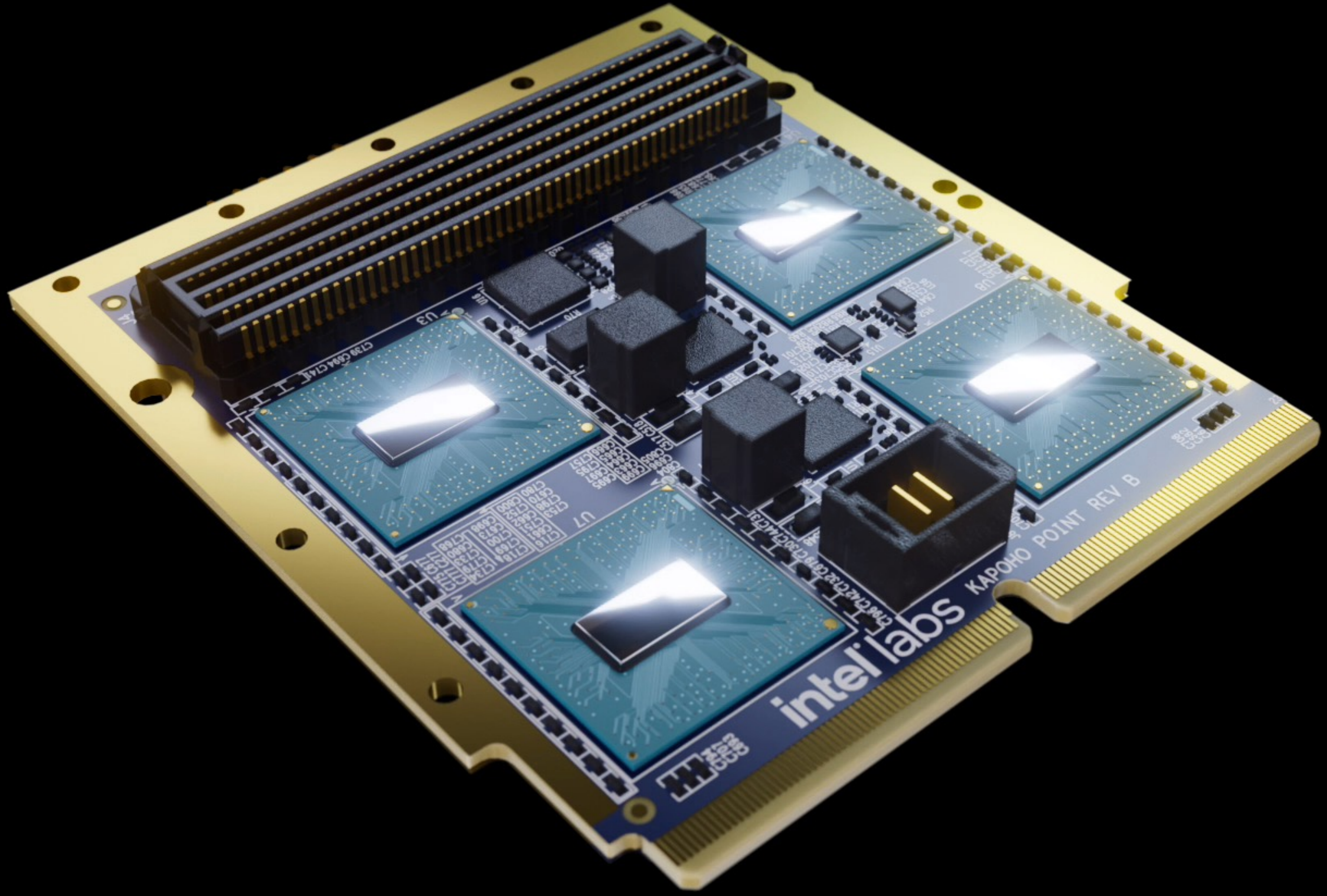

CPU


Loihi


GPU

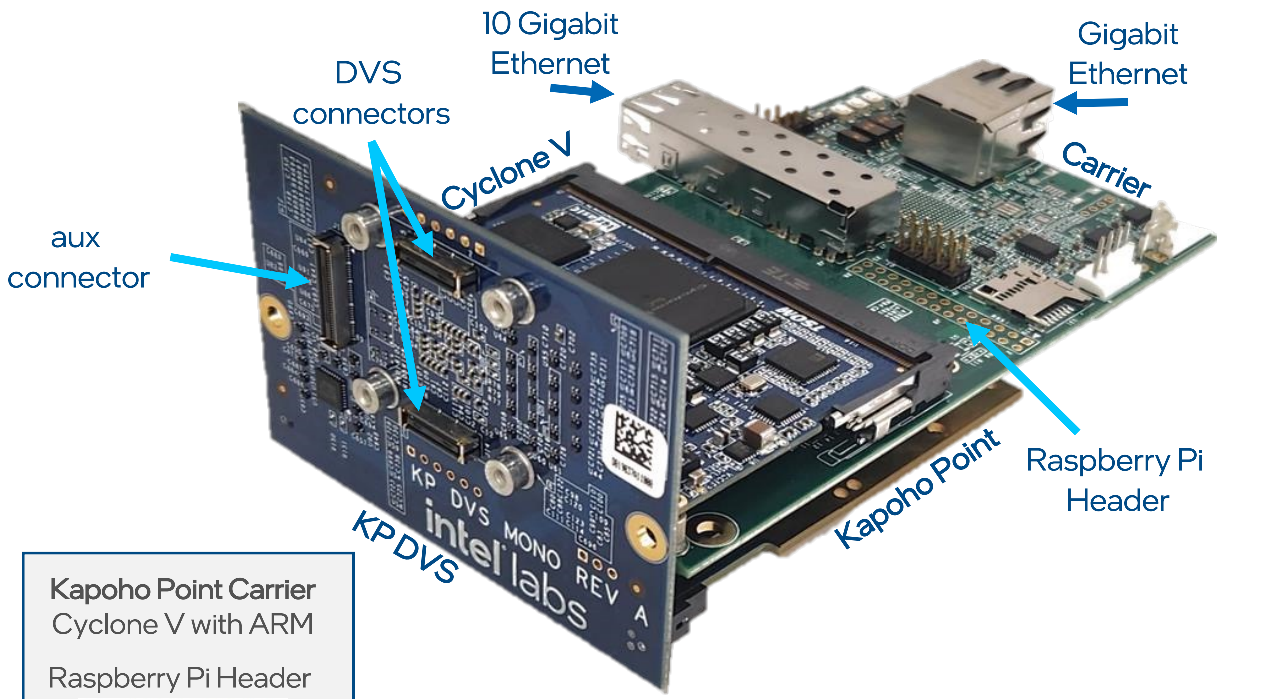
# Kapoho Point

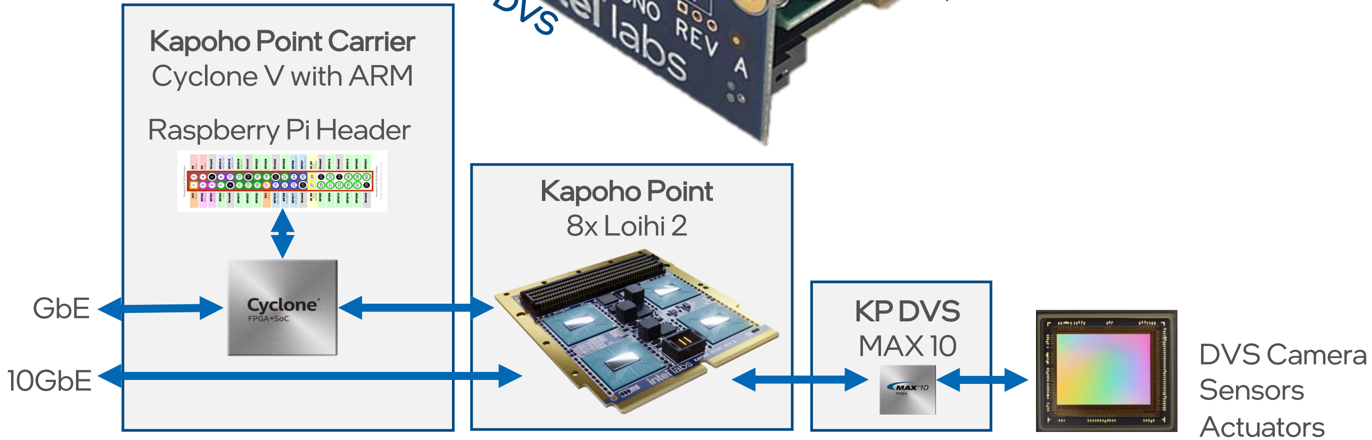
## Stackable 8-chip board



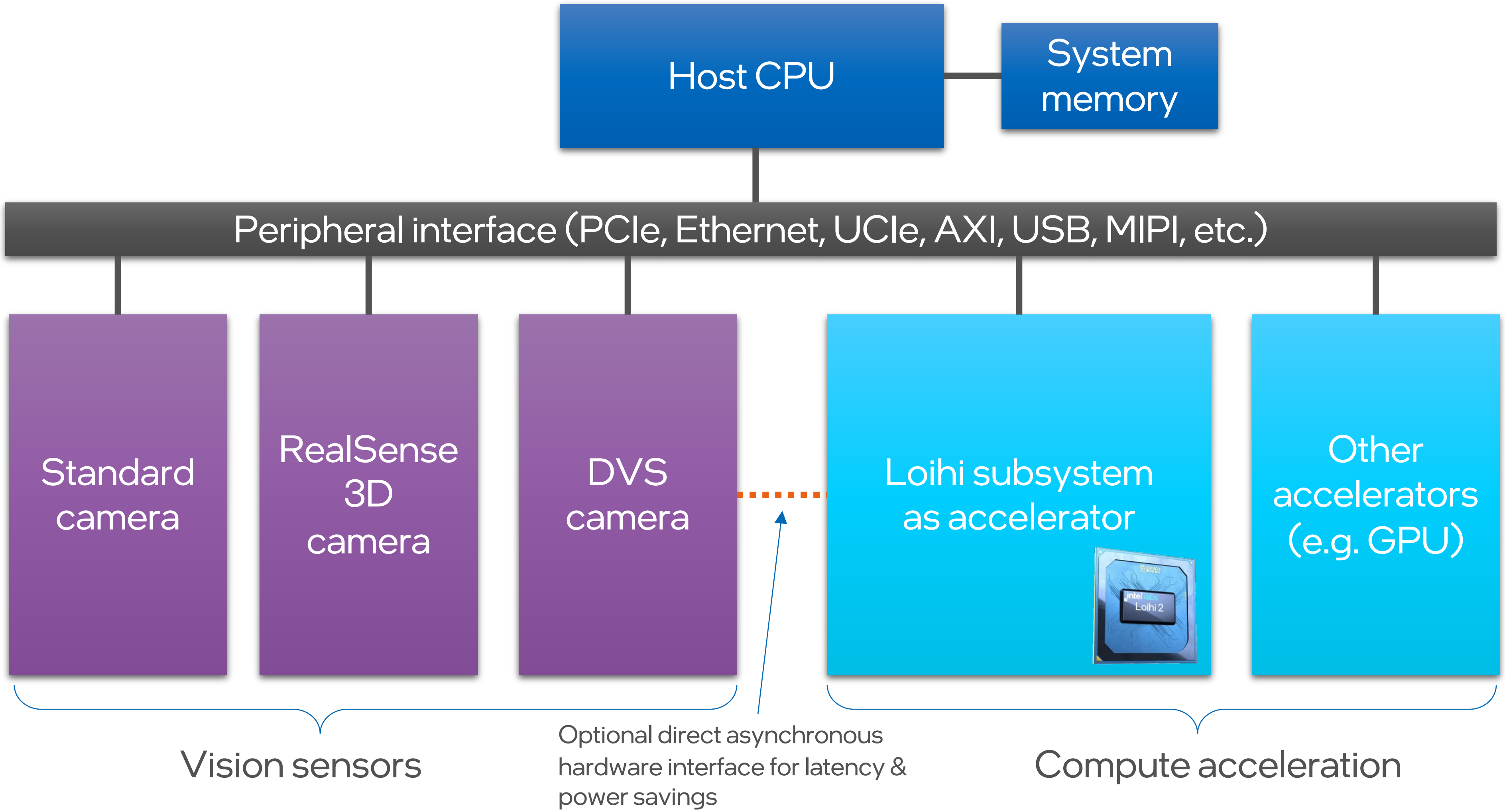

| Kapoho Point Spec per Board | |
|---|---|
| Number of chips | 8 |
| Max neurons | 8.1 M |
| Max synapses | 960 M |
| Interfaces | GbE via host board<br>10 GbE direct to Loihi<br>MIPI, GPIO, AER, SHS via interface board |
| Dimensions | 79 mm x 69 mm x 15 mm |
| Weight | 108g |
| Power supply | 12 V |

# System Architecture

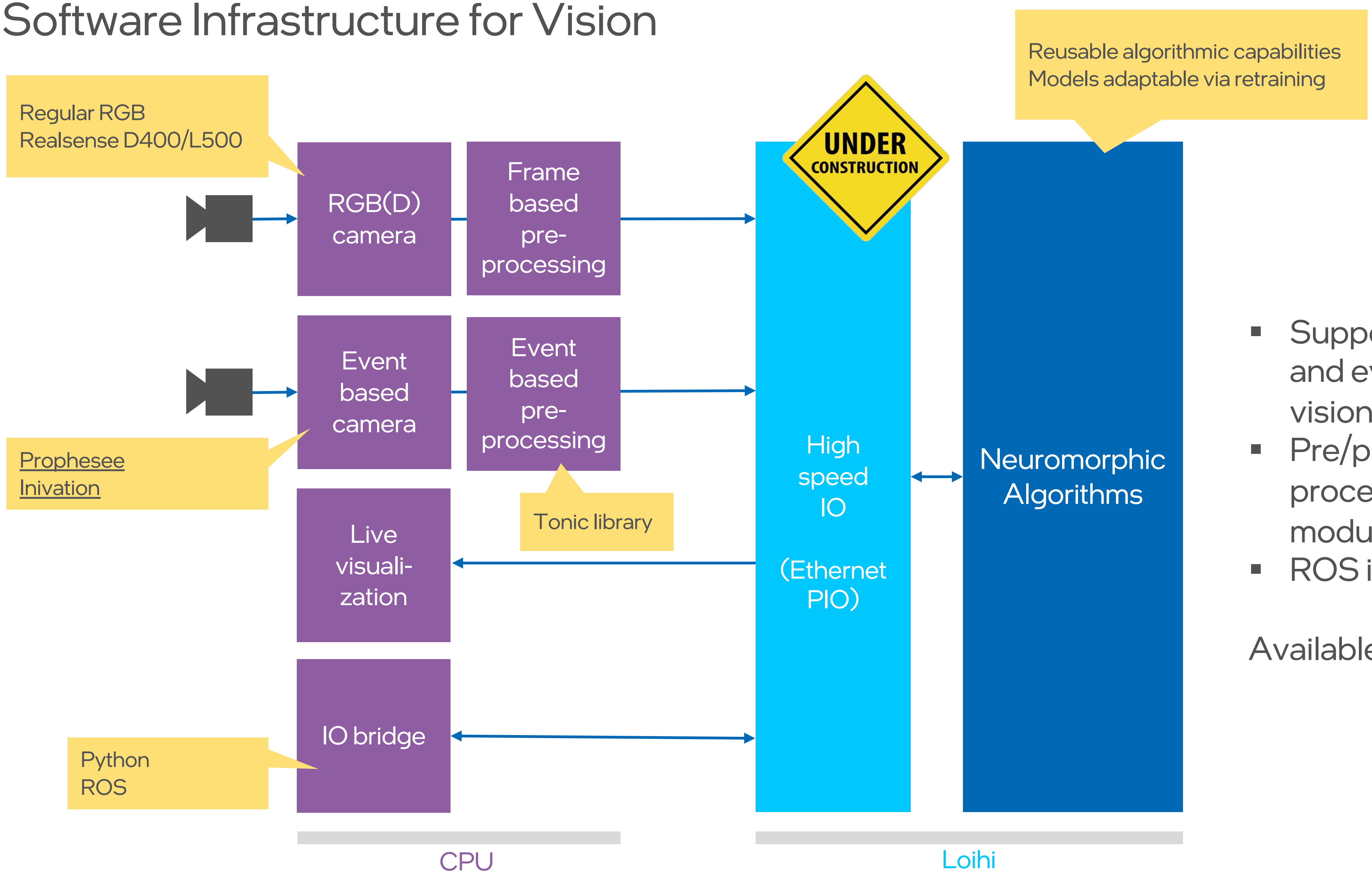


# System Architecture



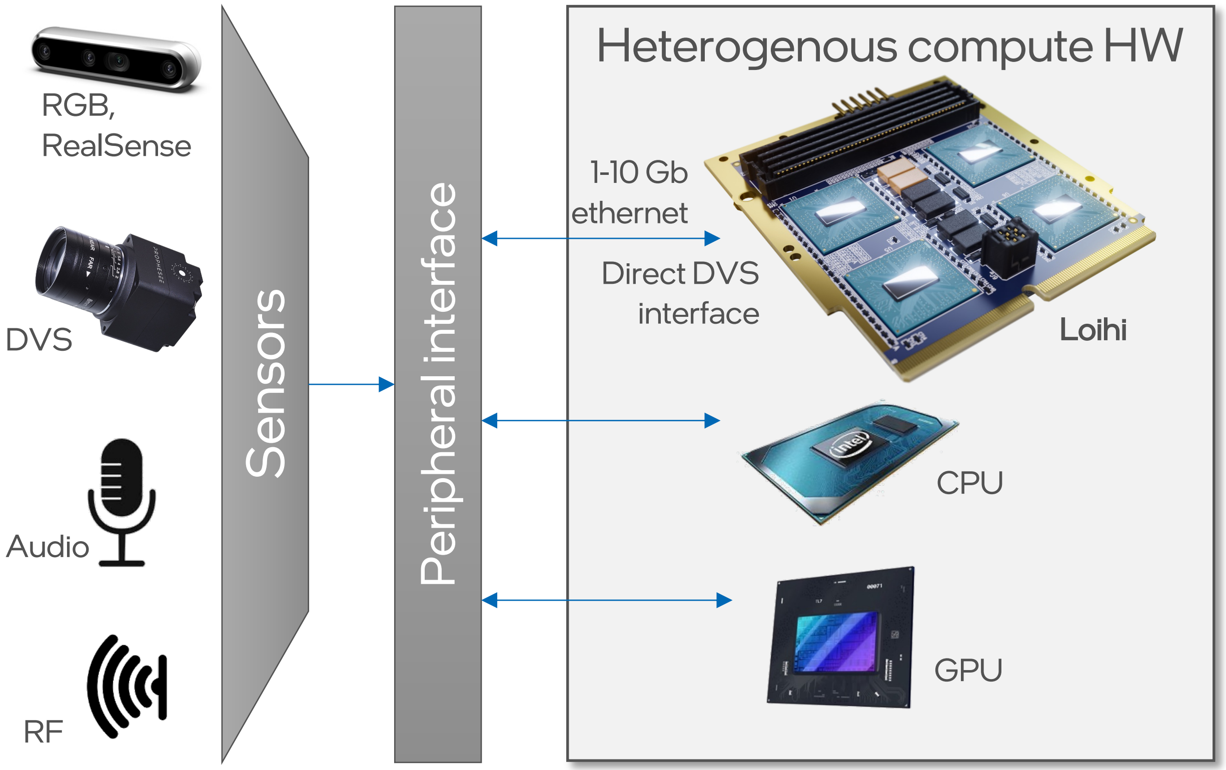
# Software Infrastructure for Vision



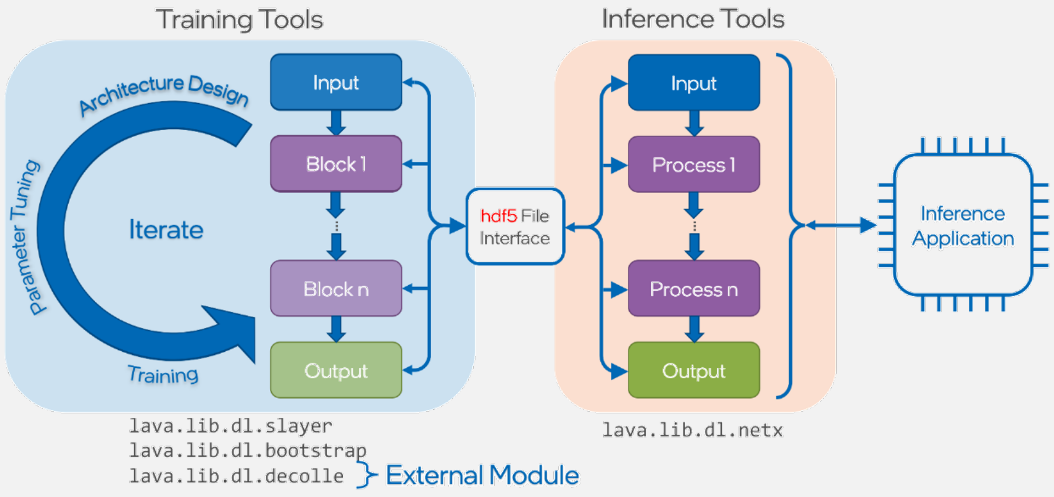

- Support for regular and event-based vision sensors
- Pre/post-processing modules
- ROS interface

Available in Q3!

# Training Vision Models with Lava-DL




## Deep Learning Tools




- Most mature way to get started with Loihi
- HW-aware/optimized DNN training
- Open-source: https://github.com/lava-nc/lava-dl

# Today's solutions – Conventional DNN architectures

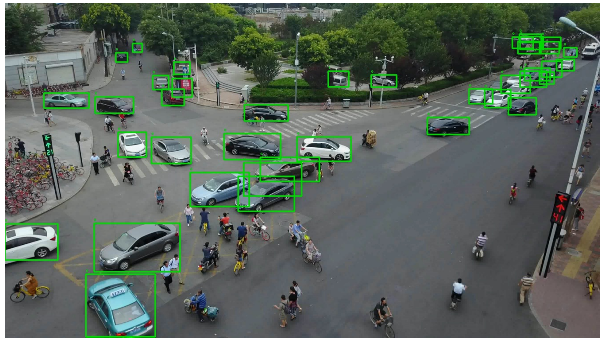
## Applied to small-scale, energy-constrained, real-time, temporal problems

### RGB/DVS visual processing

Efficiency via computing/communicating signal differences

Navigation

Object detection



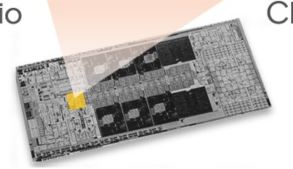
RGB frame

Sparse Delta Input

SDNN prediction

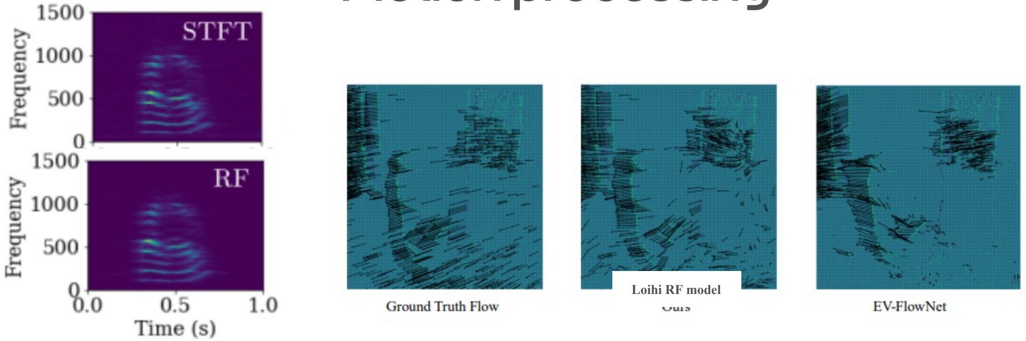

### Audio processing


- Lower dimensional than video
- Temporal processing well-suited for Loihi



Timcheck et al, Intel DNS challenge, arXiv:2303.09503

### Motion processing




- Optical flow for motion detection and analysis through temporal RF neurons
- Used in tracking, odometry, motion prediction, dynamic scene understanding

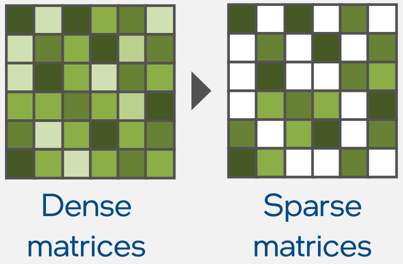
Orchard et al, Efficient neuromorphic signal processing, arXiv:2111.03746

# Next steps –Brain-inspired approaches

## Leverage Loihi's architectural differentiators for solving big/hard problems efficiently

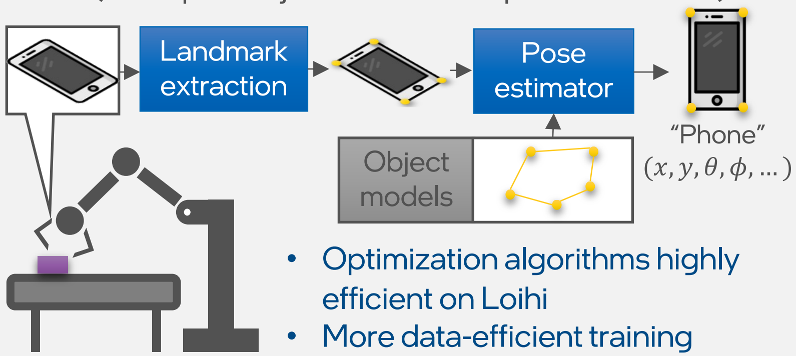
### Training (unstructured) sparse models

- Higher accuracy than structured sparsity (on GPU)
- Reduced energy/latency/ area




### (Differentiable) online optimization

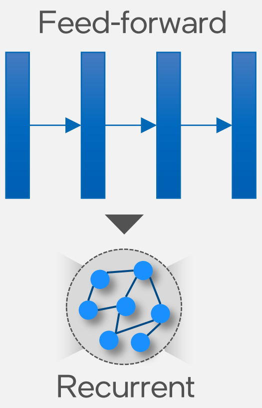
(Example: Object detection & pose extraction)




- Optimization algorithms highly efficient on Loihi
- More data-efficient training

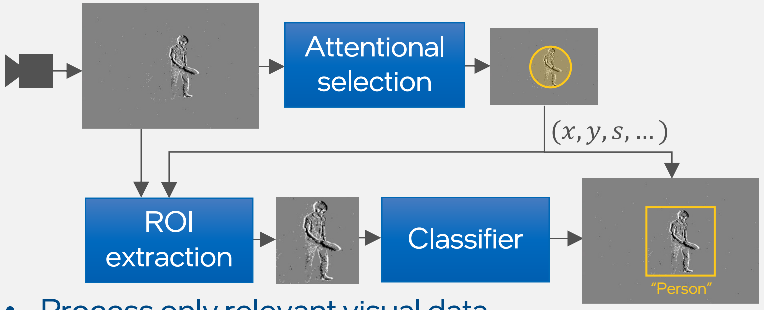
### Training recurrent models

- Processing of temporal/sequential data
- Compact through parameter sharing
- Comparatively low-memory access latency




### Attention-driven active vision

(Example: Object localization)




- Process only relevant visual data
- Reduced resource requirements

### Neuromorphic Vision Value Proposition

- Efficient & smart sensing at the edge:
  - Low-latency (~1ms)
  - Low-power (~1W)
- Improved efficiency with event cameras:
  - High dynamic range
  - No blur
- Multi-sensor support
- Online/on-chip model adaptation/learning

# Legal Information

Performance varies by use, configuration and other factors. Learn more at www.Intel.com/PerformanceIndex.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure.

Your costs and results may vary.

Results have been estimated or simulated.

Intel technologies may require enabled hardware, software or service activation.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

intel.

# Backup

# Application Development

Optional: Could also end on lava-dl

| Benefits: | ▪ Energy savings (10-100x)<br>▪ Lower latency (1-10x)<br>… for real-time signal processing | Trade offs: | Compute/memory integration:<br>▪ High efficiency<br>▪ At expense of lower capacity | Focus areas | ▪ Small, energy-constraint problems<br>▪ Brain-inspired approaches |
|---|---|---|---|---|---|

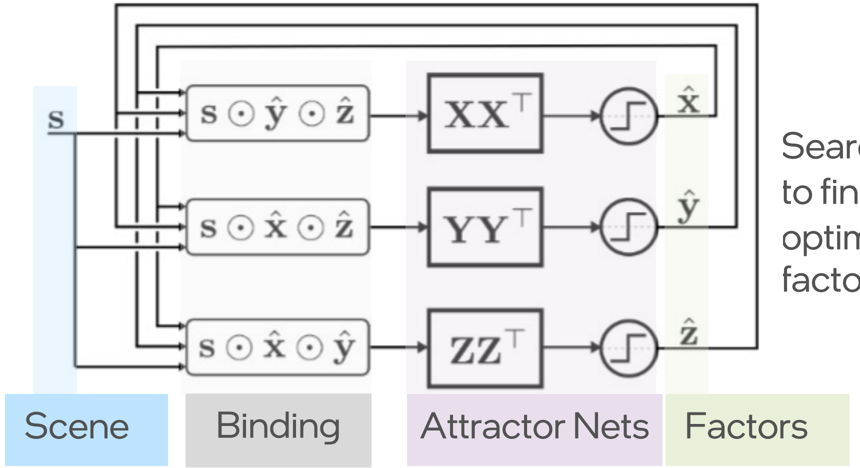
Message idea

- Loihi is fast and highly energy efficient but comes at expense of capacity without external memory
- Today's focus has been mostly on smaller models where energy-efficiency matters
- Tomorrow's focus will be more on novel (bio-inspired) approaches to solve big/hard problems with fewer resources exploiting Loihi's versatile, programmable nature

# Hyperdimensional (VSA-based) optimization for visual understanding

$$\underset{x,y,z}{\operatorname{argmax}}[\boldsymbol{s} \cdot (\boldsymbol{x} \odot \boldsymbol{y} \odot \boldsymbol{z})]$$

Hypervectors: Scene Generative factors




Searches to find optimal factors

**Resonator Networks, 1: An Efficient Solution for Factoring High-Dimensional, Distributed Representations of Data Structures**

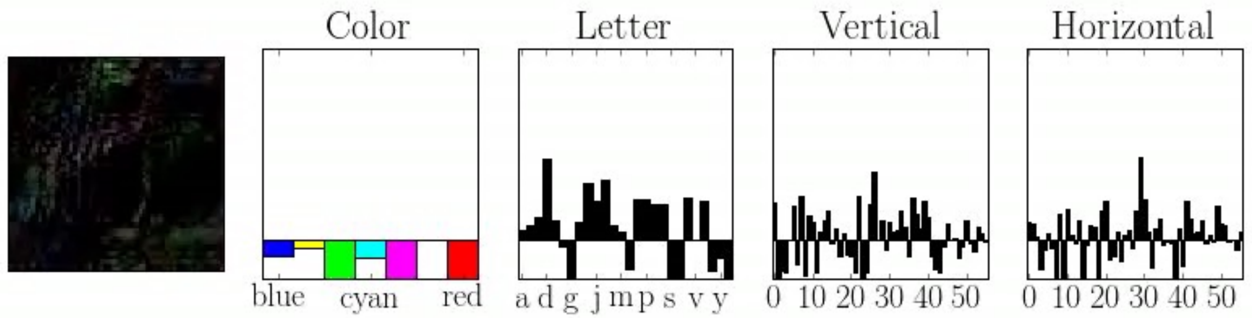
E. Paxon Frady, Spencer J. Kent, Bruno A. Olshausen, Friedrich T. Sommer

*Neural Computation* (2020) 32 (12): 2311–2331.

Example: Disentangling color, type, position of digits

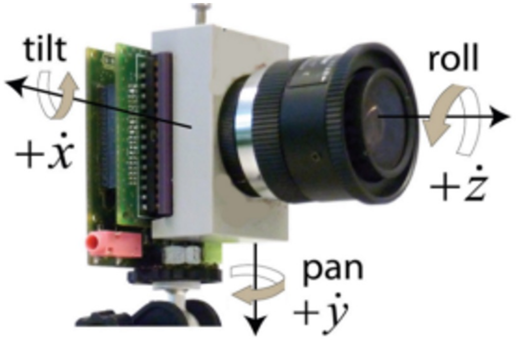
$$\mathbf{s} = \mathbf{c}_{\mathtt{cyan}} \odot \mathbf{d}_7 \odot \mathbf{v}_{\mathtt{top}} \odot \mathbf{h}_{\mathtt{left}} + \mathbf{c}_{\mathtt{pink}} \odot \mathbf{d}_3 \odot \mathbf{v}_{\mathtt{top}} \odot \mathbf{h}_{\mathtt{right}} + \mathbf{c}_{\mathtt{red}} \odot \mathbf{d}_8 \odot \mathbf{v}_{\mathtt{middle}} \odot \mathbf{h}_{\mathtt{left}}$$

Structural scene representation using high-dimensional vectors, binding and composition

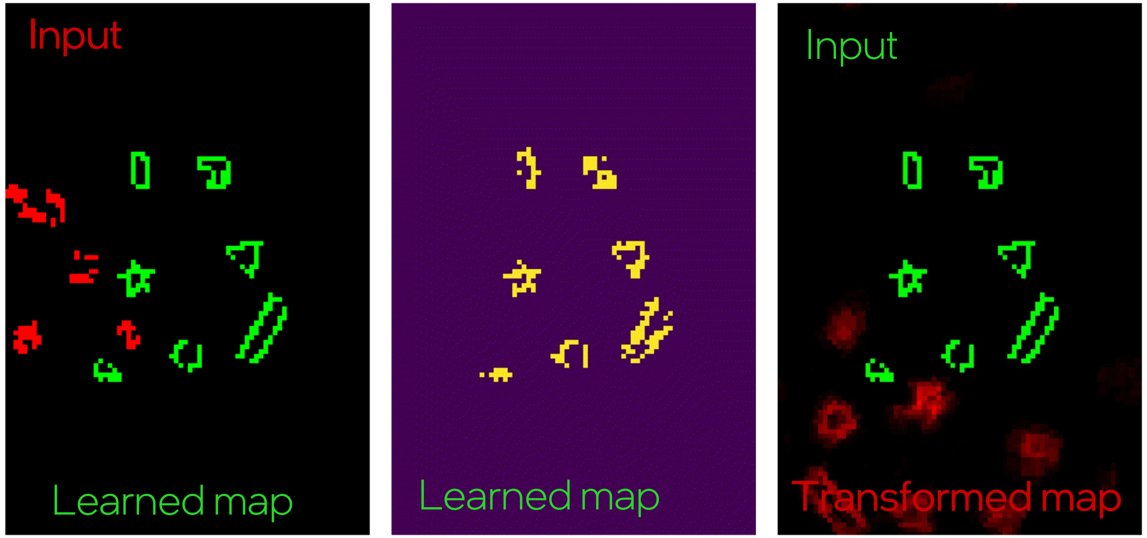



Factorization does not require model to be trained on full combinatorial input space

# Applied to Visual Odometry




Problem:
Determine pose of moving camera given input scene of scattered shapes




| Publication | Name | angle error (°) random split | angle error (°) "novel split" |
|---|---|---|---|
| [69] Kendall et al. (2015) | PoseNet | 7.39 | 12.5 |
| [70] Kendall et al. (2016) | Bayesian PoseNet | 9.56 | 12.1 |
| [71] Laskar et al. (2017) | Pairwise-CNN | 6.33 | 10.4 |
| [72] Walch et al. (2017) | LSTM-Pose | 4.44 | 7.6 |
| [77] Nguyen et al. (2019) | SP-LSTM | 2.26-2.95* | 5 |
| [73] Reinbacher et al. (2017) | Panoramic | - | 5* |
| Renner et al. (2022) (ours) | NEVO | - | $\boxed{3}$ |

Dataset: Mueggler, E., Rebecq, H., Gallego, G., Delbruck, T. & Scaramuzza, D. The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam. *The International Journal of Robotics Research.*

DNN solutions, trained on dataset
1-100M parameters

VSA Resonator solution.
No need for DNN training. ~12 free parameters.

# Value proposition

### Event-based sensing

- High dynamic range
- High speed
- Event-sparsity

### Loihi processor

- Efficiency through...
  - minimal data movement
  - event-based communication/ computation
  - optimization for sparsity
- Support online learning

### Neuromorphic algorithms

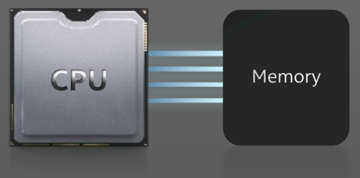
- Emphasize sparse processing
- Exploit temporal redundancy
- Extract spatio-temporal correlations
- Optimized to data and HW

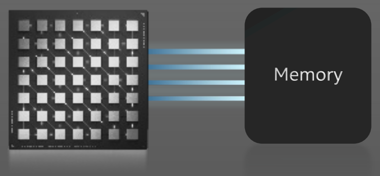
### Application scope

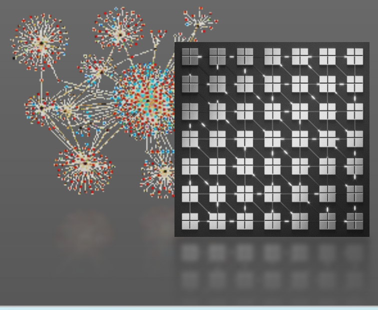
Vision problems with continuous input
High speed, varying lighting conditions
Size/weight/power constrained

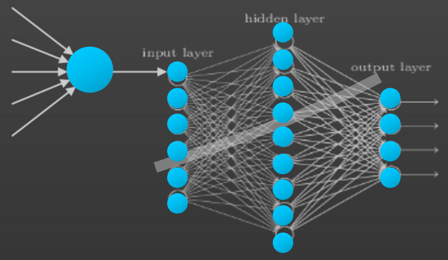
# Leads us to a new class of computer architecture

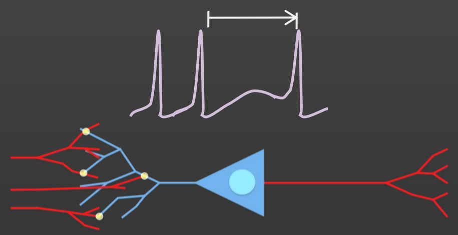
| Standard Computing | Parallel Computing | Neuromorphic Computing |
|---|---|---|
| PROGRAMMING BY ENCODING ALGORITHMS | OFFLINE TRAINING USING LABELED DATASETS | LEARN ON THE FLY THROUGH NEURON FIRING RULES |
| SYNCHRONOUS CLOCKING | SYNCHRONOUS CLOCKING | ASYNCHRONOUS EVENT-BASED SPIKES |
| SEQUENTIAL THREADS OF CONTROL | PARALLEL DENSE COMPUTE | PARALLEL SPARSE COMPUTE |



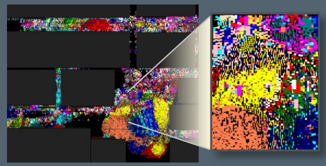
# Intel's neuromorphic research scope

## Asynchronous Design

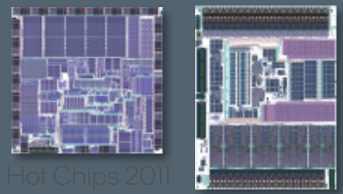
**Clockless IP providing energy savings**

Network-on-chip
SRAM
RISC-V
Wave pipelined I/O

**Standard tool flow automation**

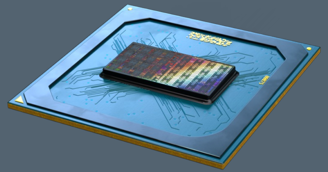


**Commercially proven** in five generations of networking products

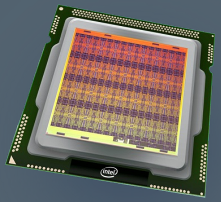



## Silicon

**Loihi 2**
Intel 4
2021

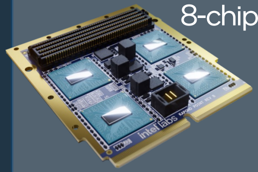


**Loihi 1**
14nm
2018

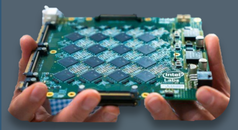


## Systems

**Kapoho Point**
8-chip stackable

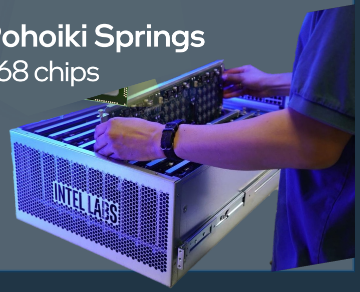


**Kapoho Bay**
USB



**Nahuku**
32 chips

**Pohoiki Springs**
768 chips




## Software

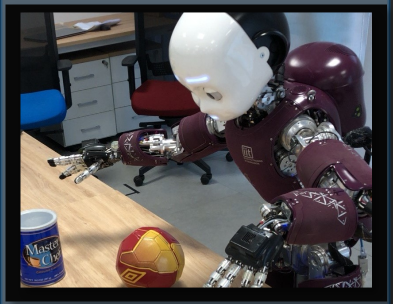
LAVA

Designed bottom-up for neuromorphic computing

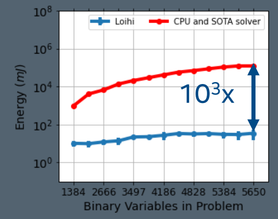
- Open source
- Composable
- Multi-paradigm
- Multi-platform

**State-of-the-art neuromorphic tools & methods accessible for all**

## Applications

**Train scheduling** DB



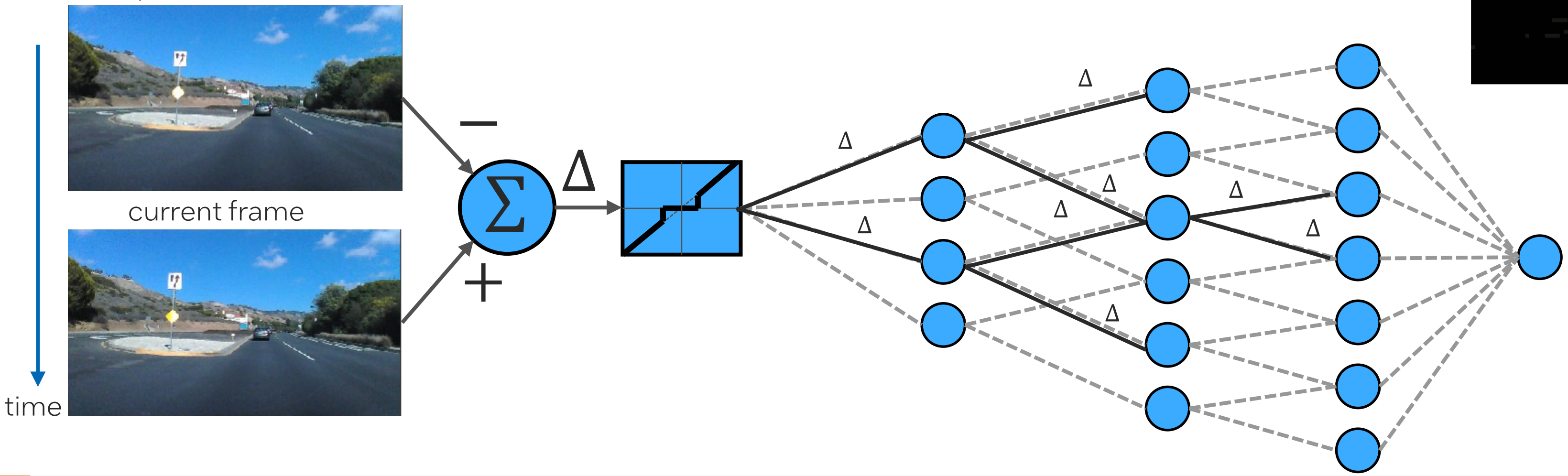
# Conventional vision (Ex: Navigation)

- Uses familiar CNN/DNN topologies
- Efficient frame-based vision via Σ/Δ coding
- Leverages temporal continuity in video
- Typical SynOp reduction >10x via sparse communication

RGB frame

Sparse Delta Input

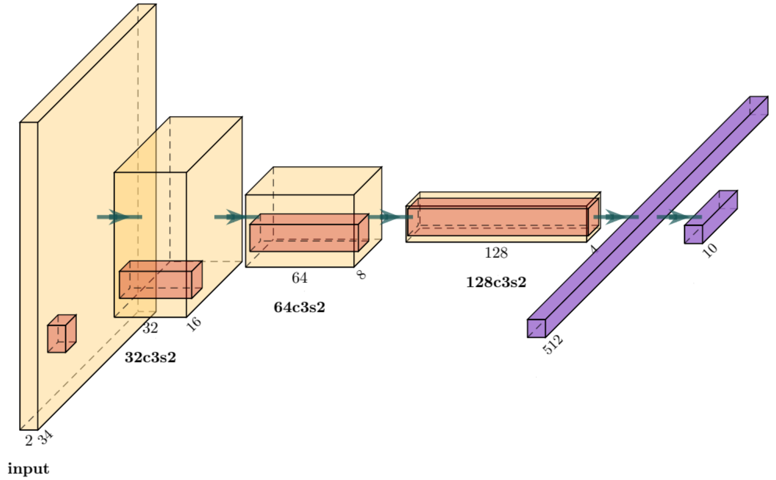
SDNN prediction

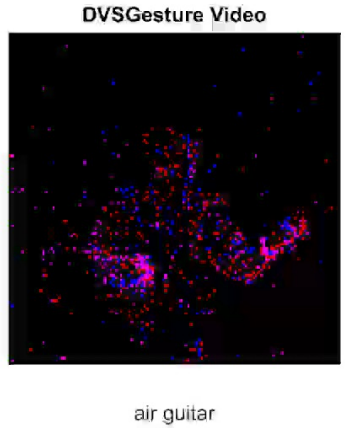


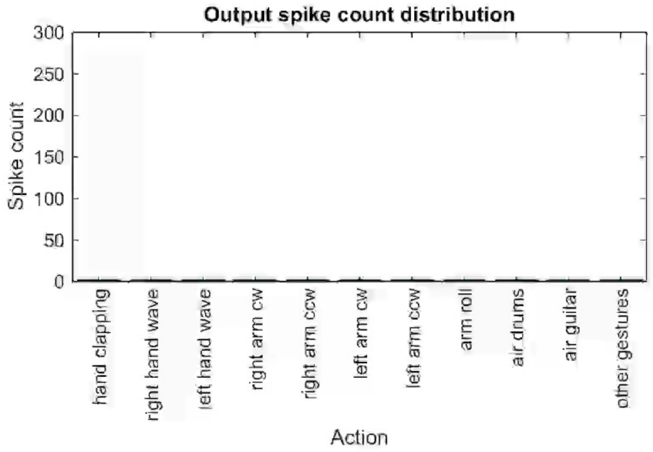
# Event-based vision (Ex: Gesture recognition)

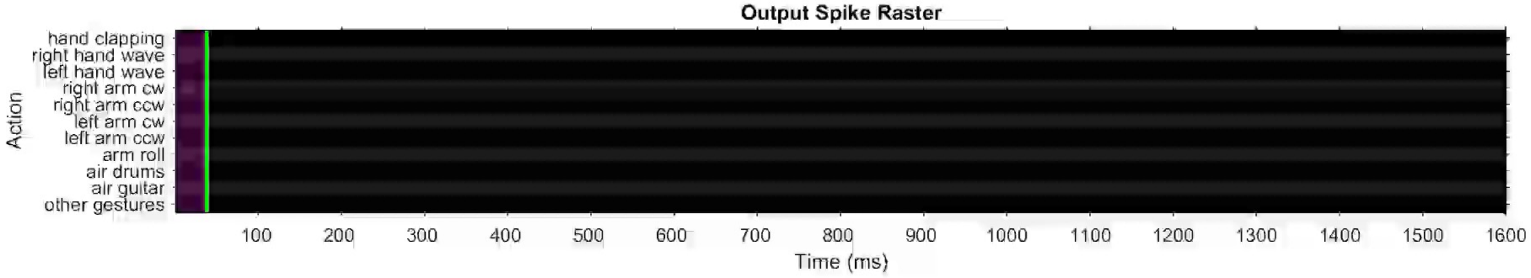
- Efficient event-based vision starting at sensor
- Extremely low $\mu s$ to $s$ latency
- High dynamic range
- Training challenge

Example: DVS gesture CNN

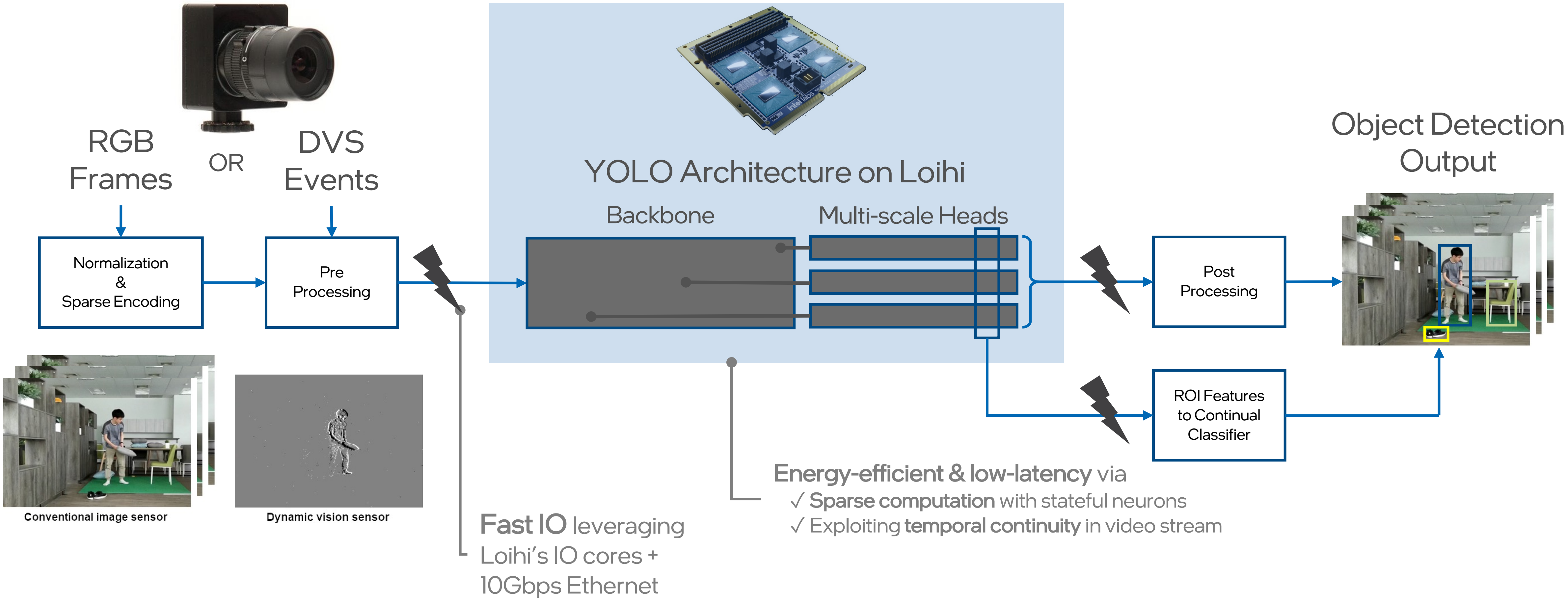


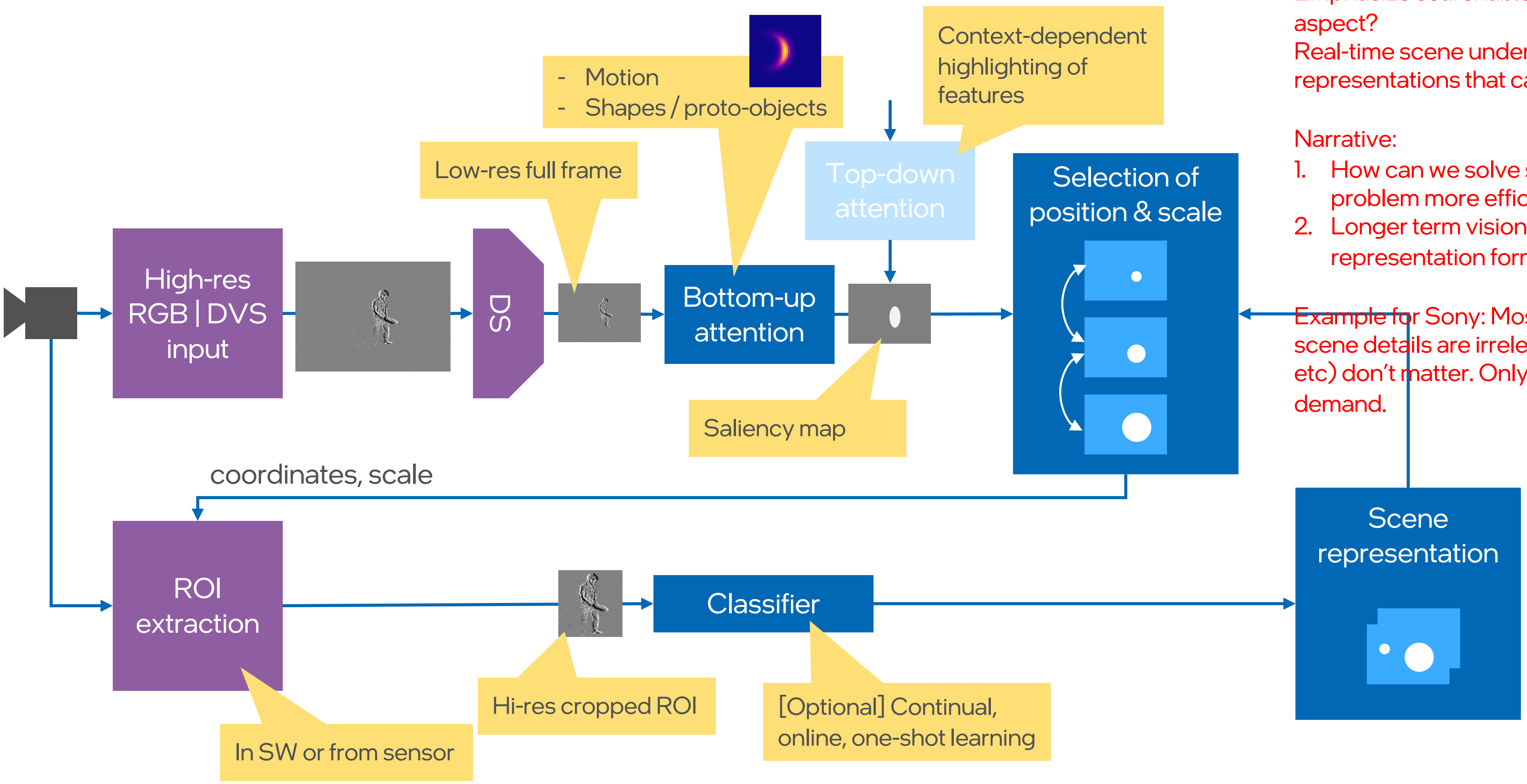



# Object detection via YOLO architecture



# Object detection via feedback-driven attention

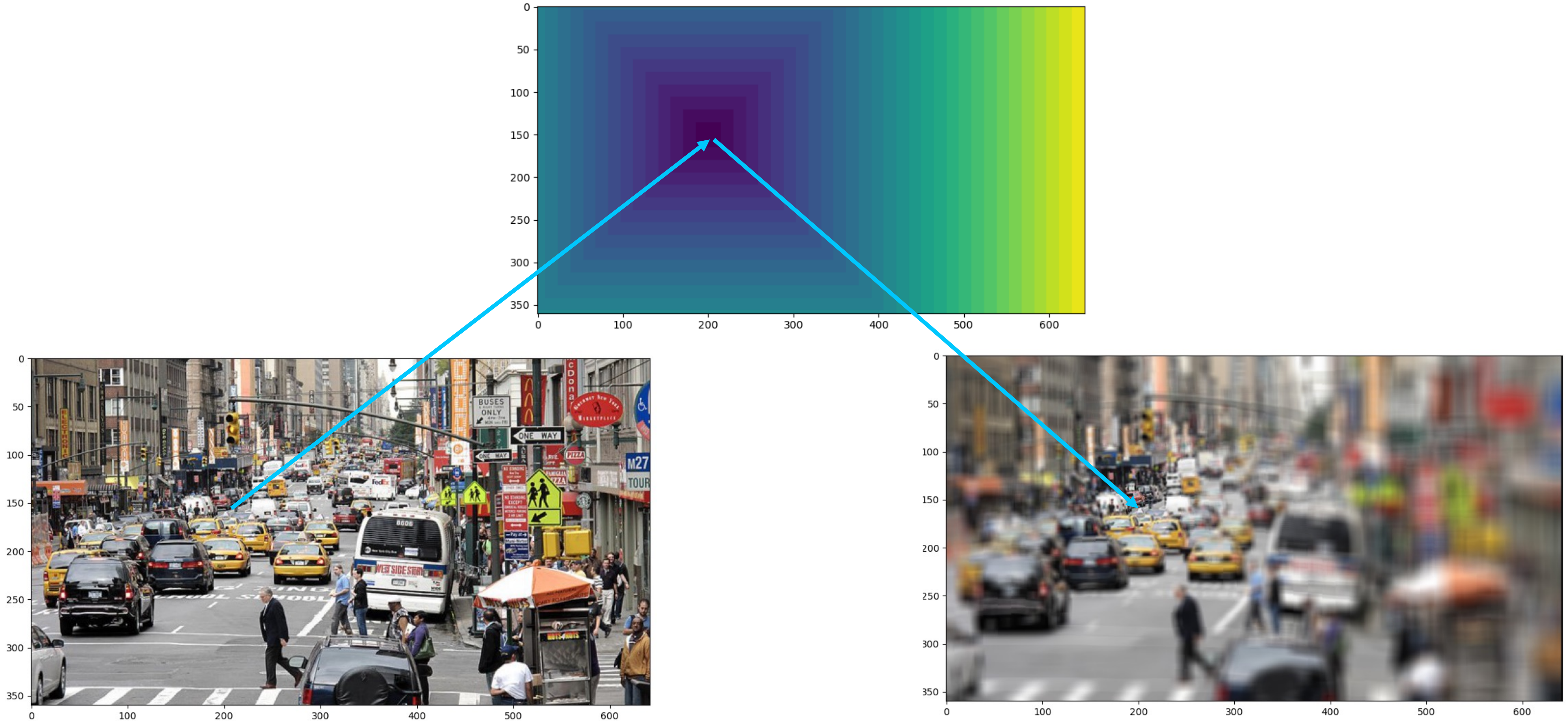



Emphasize searchable scene representation aspect?
Real-time scene understanding. Scene representations that can be queried.

Narrative:

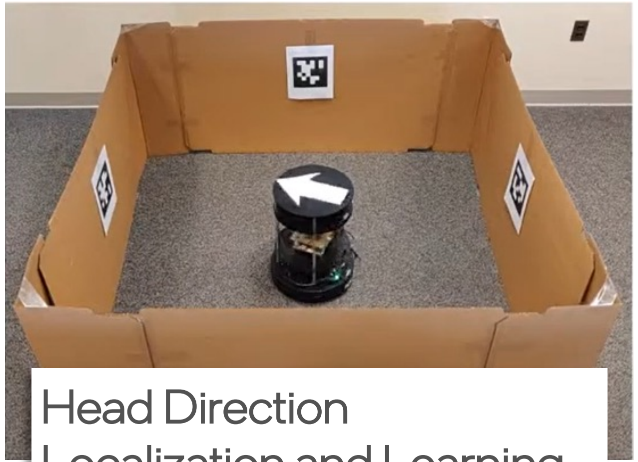
1. How can we solve same object detection problem more efficiently but with this arch?
2. Longer term vision: Real-time scene representation formation (queryable, …)

Example for Sony: Most of the time lots of scene details are irrelevant (car color, model, etc) don't matter. Only extract this info on demand.


0
50
100
150
200
250
300
350
0
100
200
300
400
500
600
BUSES ONLY
ONE WAY
M27
TOUR
PIZZA

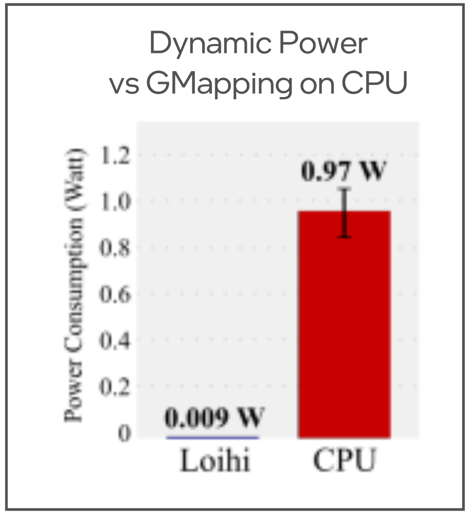
# SLAM-like capabilities

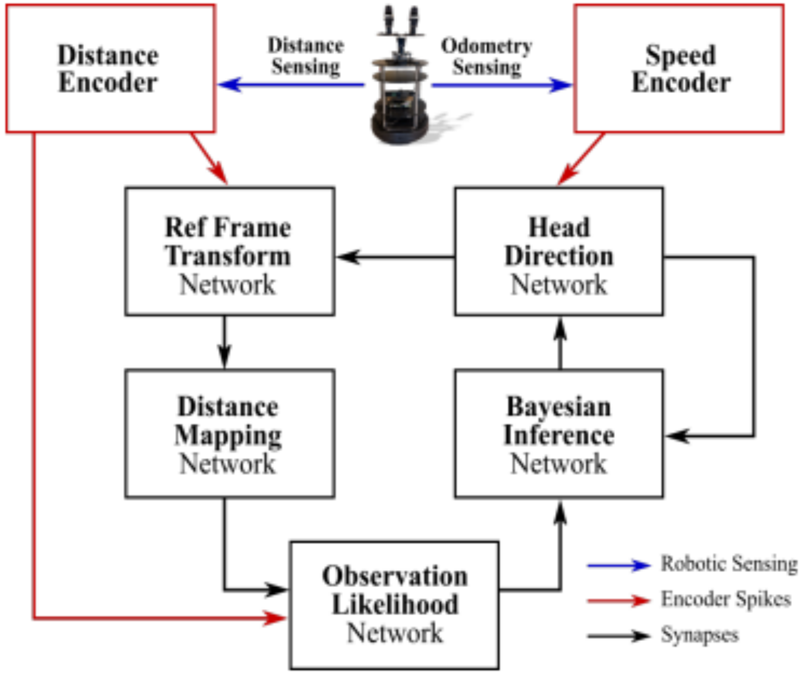
- Rutgers (Konstantinos Michmizos)



Head Direction Localization and Learning

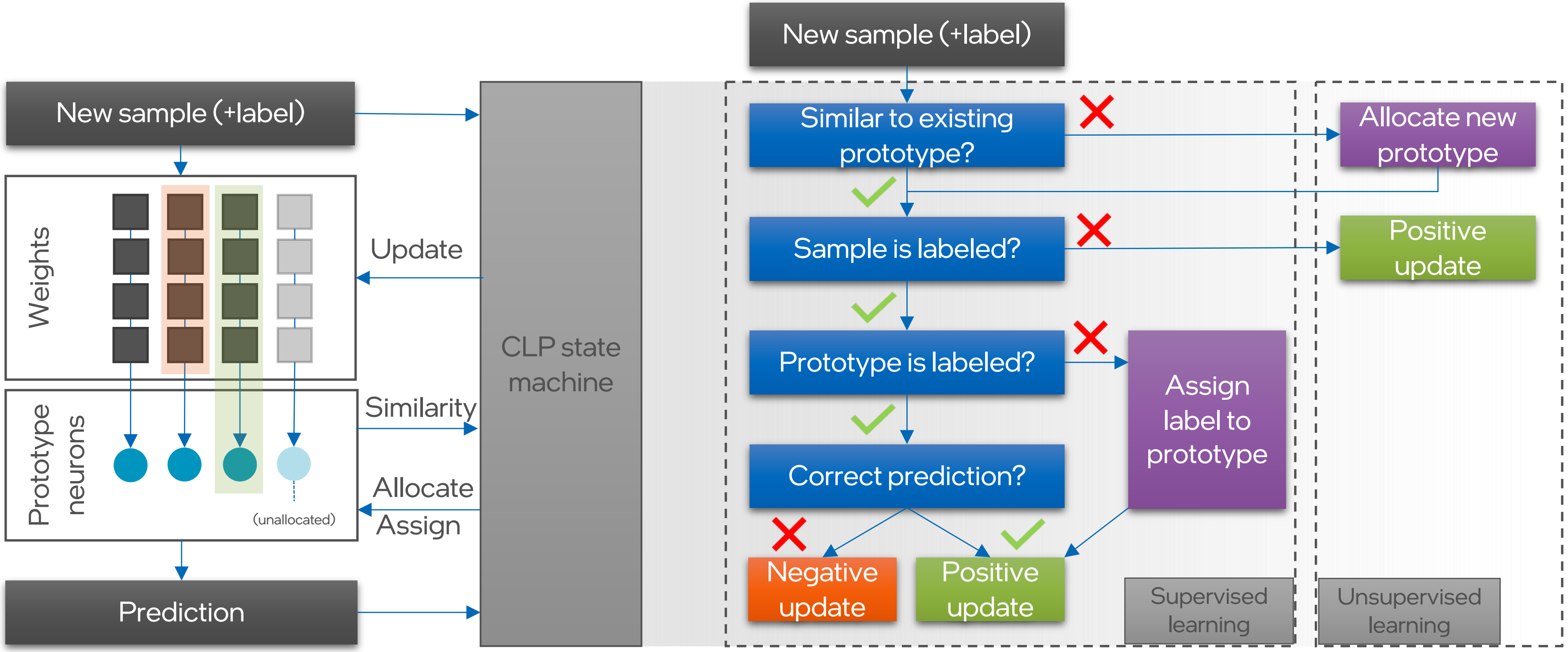
- 100x lower power vs CPU

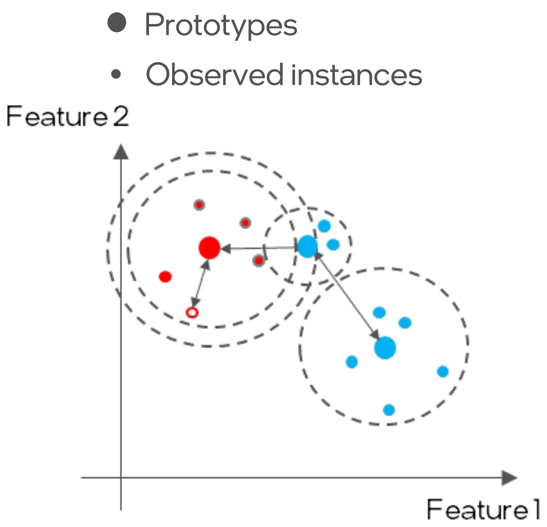



Tang, Shah, Michmizos. "Spiking Neural Network on Neuromorphic Hardware for Energy-Efficient Unidimensional SLAM," in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, Nov 4-8, 2019

# Continual Learning Prototype Classifier



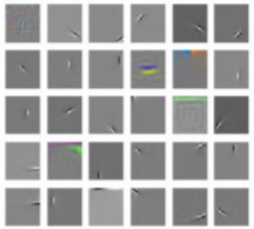



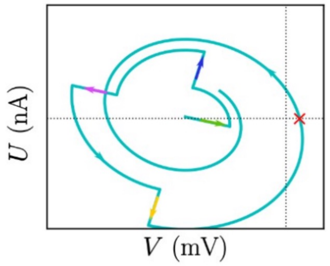

**Characteristics:**
- Single-layer, local updates
- Interpretable
- Adjustable memory capacity
- Performant & energy efficient

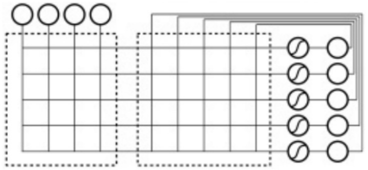
**Capabilities:**
- Novelty detection
- One-shot learning
- Continual online learning
- Open-set recognition

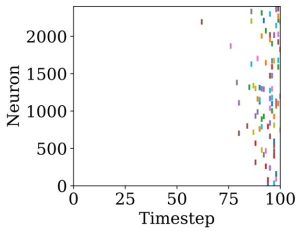
# Upcoming signal processing library

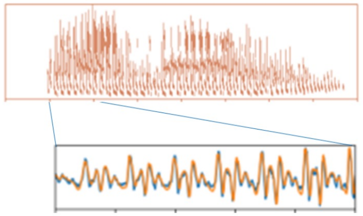
| | | | | |
|---|---|---|---|---|
| Pre-processing<br>Dim.-Reduction<br>Sparsification | Filtering<br>Feature-extraction<br>Fourier Analysis | Memory<br>High-order features<br>Signal Integration | Classification<br>Decision Making<br>Anomoly Detection | Reconstruction<br>Control<br>Denoising |





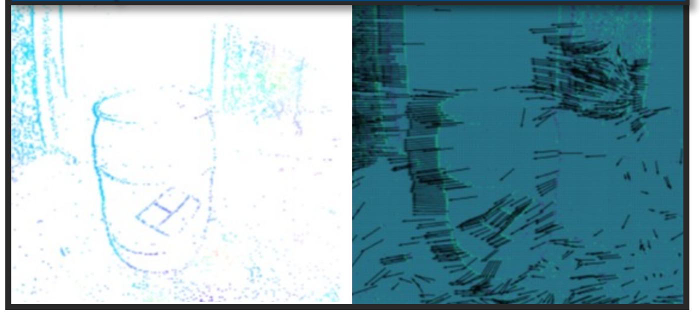


# Optical flow estimation with resonate-and-fire neurons

Resonate and Fire neurons compute optical flow for event-cameras with higher accuracy and 90x fewer ops than leading DNN solution

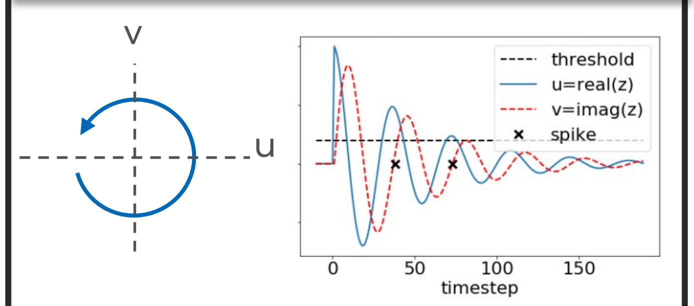
## Optical Flow for Event-Cameras



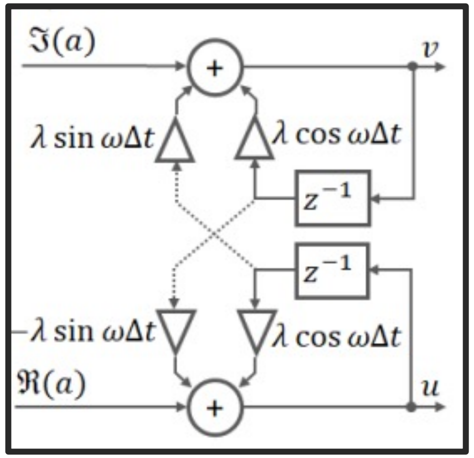
## Average Endpoint Error on MVSEC

| | Indoor Flying 1 AEE | Indoor Flying 1 % outlier | Indoor Flying 2 AEE | Indoor Flying 2 % outlier | Indoor Flying 3 AEE | Indoor Flying 3 % outlier |
|---|---|---|---|---|---|---|
| EV-FlowNet$_{2R}$ | 1.03 | 2.2 | 1.72 | 15.1 | 1.53 | 11.9 |
| Ours$_{\text{DENSE}}$ | 0.91 | **0.35** | 1.28 | 5.83 | 1.04 | 2.88 |
| Ours$_{\text{SPIKES}}$ | **0.83** | 0.68 | **1.22** | **5.42** | **0.97** | **2.65** |

## Resonate & Fire Neuron Behavior




$$z_k[t] = \lambda e^{i\omega\Delta t} z_k[t-1] + a_k[t]$$




E. P. Frady et al, "Efficient Neuromorphic Signal Processing with Resonator Neurons." Journal of Signal Processing Systems, 2022.

# Latest Tools: Loihi 2 and Lava


* specs and configuration details can be found at intel.com/neuromorphic

- Up to 10x faster processing capability*
- Up to 60x more inter-chip bandwidth*
- Up to 1 million neurons with 15x greater resource density*
- Programmable neurons
- Graded spikes
- 3D scalable
- 10G Ethernet I/F to host

## LAVA

Event-based communication

Multi-Paradigm

Multi-Abstraction

Multi-Platform

Open-Source and Community-Driven

https://github.com/lava-nc

intel labs

# Loihi 2

## 32 $mm^2$ in Intel 4