A New Vision Chip with SPAD Imaging and Spiking Neural Network Processing

Liyuan Liu

liuly@semi.ac.cn

Institute of Semiconductors,
Chinese Academy of Sciences,
Beijing, China

ISSW2024 @ Italy, Trento

Outline

- Background
- Chip Architecture
- Key Techniques
- Results and Comparison
- Discussion

Outline

- **Background**
- Chip Architecture
- Key Techniques
- Results and Comparison
- Discussion

Motivation

Edge machine vision applications

- Agile drone
- Intelligent robots
- Autonomous vehicles

Requirements

- Versatile (2D/3D/HDR)
- Intelligent
- Energy efficient
- Small size

High dynamic range

Complex terrain

Obstacles

Limited power supply

from bing.com

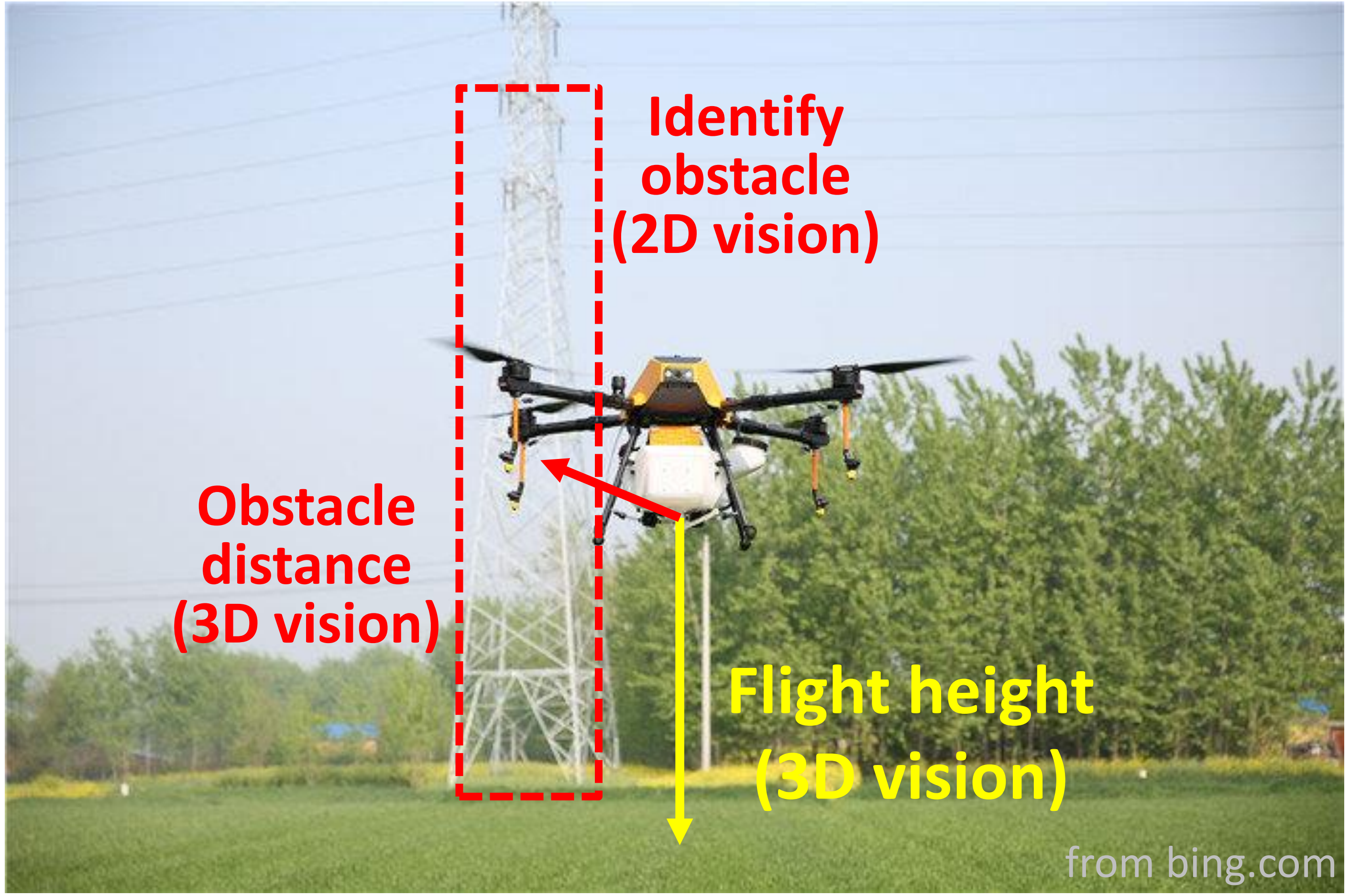
Necessity for 2D/3D Vision

Edge machine vision applications

- Agile drone
- Intelligent robots
- Autonomous vehicles

Requirements

- Versatile (2D/3D/HDR)
- Intelligent
- Energy efficient
- Small size



A single chip with 2D/3D vision

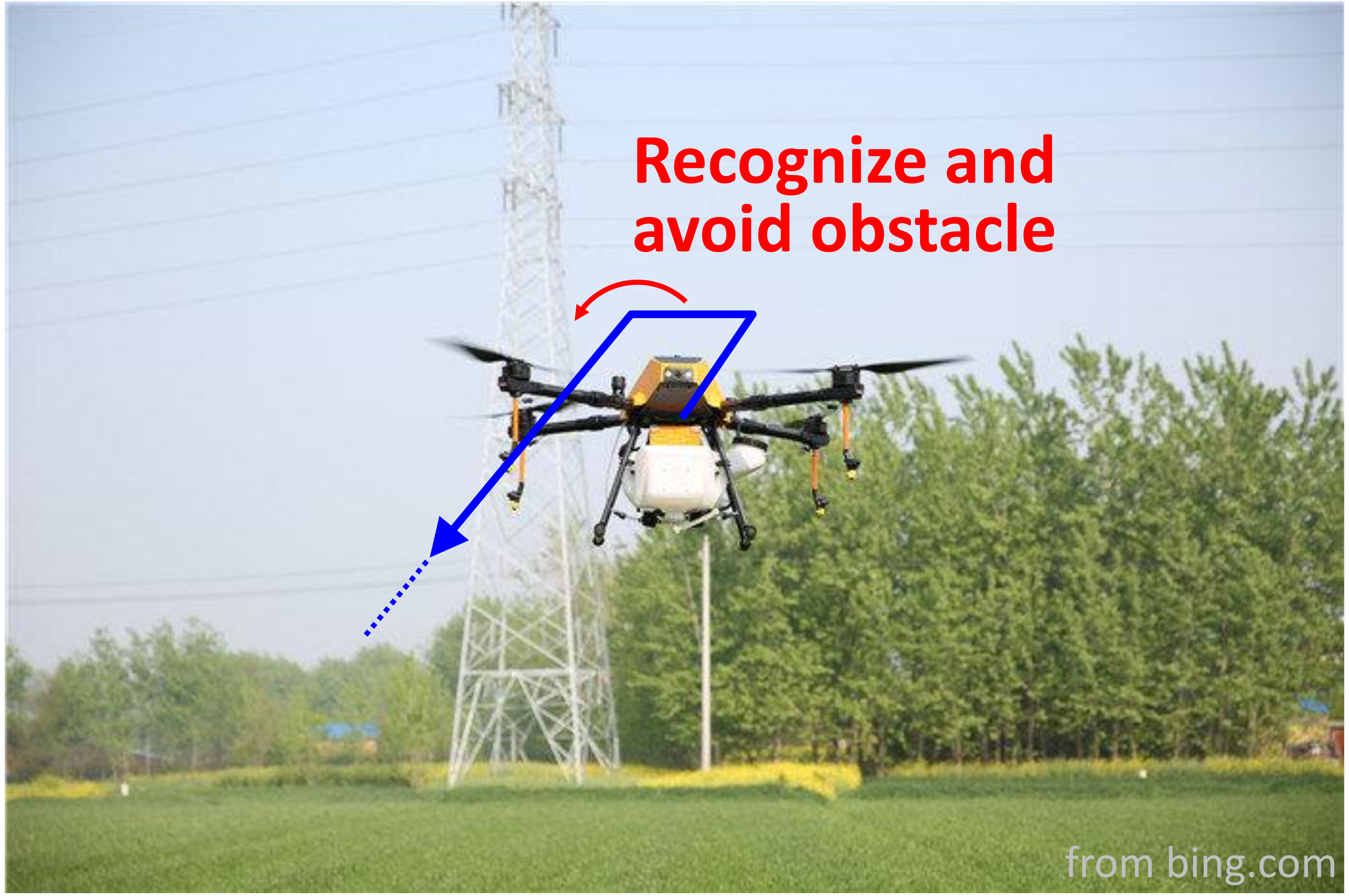
Necessity for Sensing and In-situ Processing

Edge machine vision applications

- Agile drone
- Intelligent robots
- Autonomous vehicles

Requirements

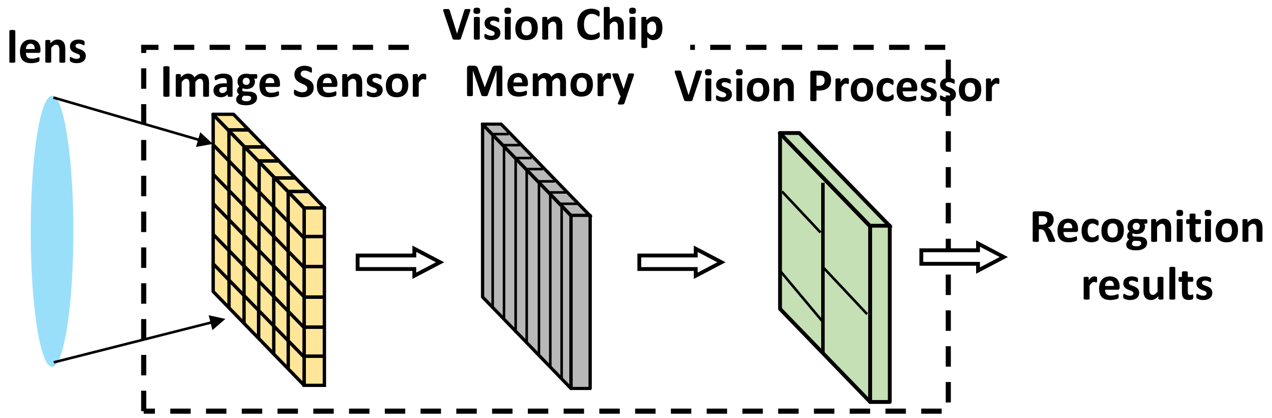
- Versatile (2D/3D/HDR)
- Intelligent
- Energy efficient
- Small size



A single chip with sensing and intelligent in-situ processing ability

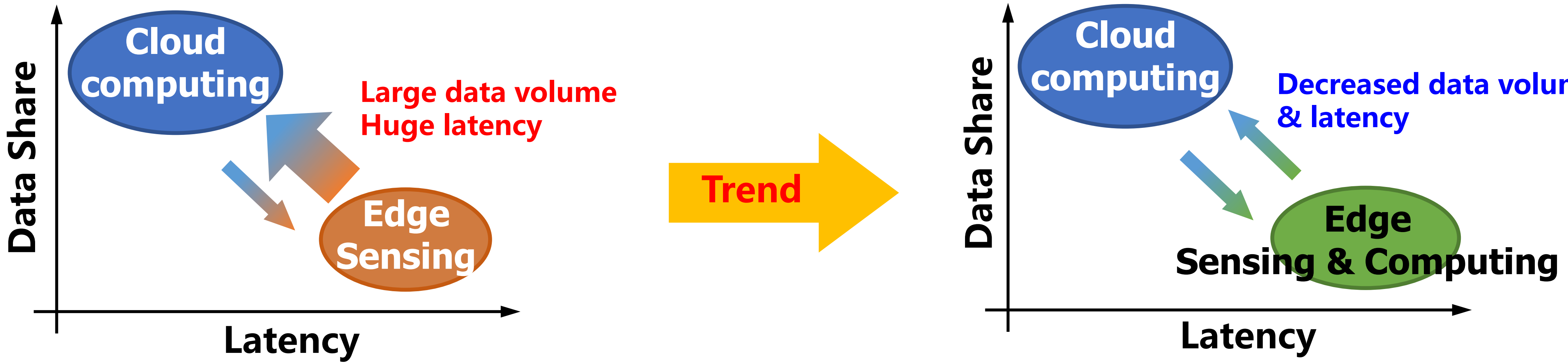
Vision Chip Concept

Vision chip integrates image sensor, memory and vision processor.
It can acquire visual information and perform in-situ processing.



Advantages:

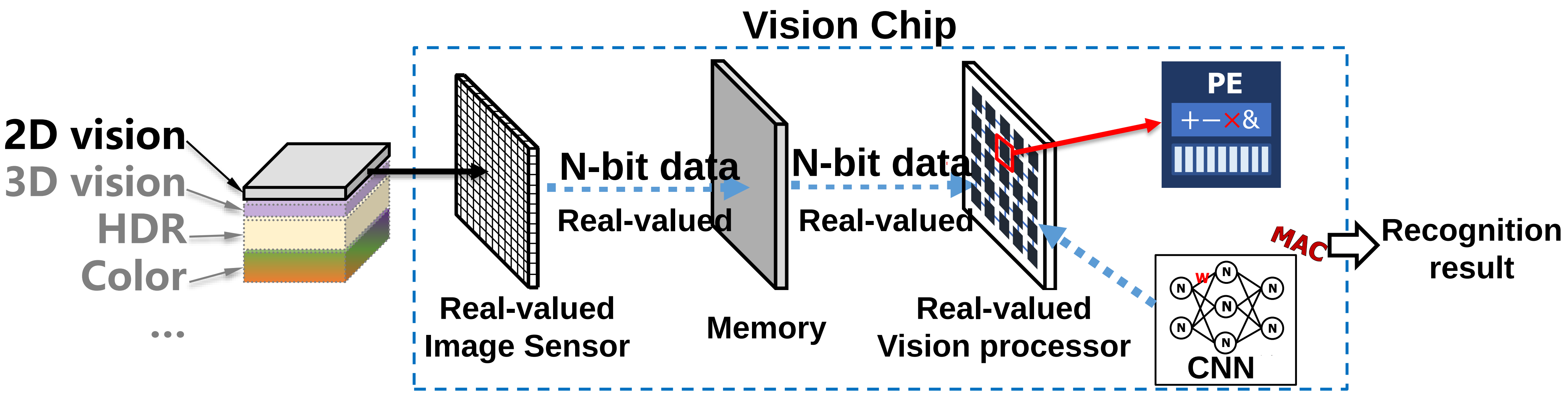
- Low volume of communication data
- High speed
- High security



Vision chips are a key technology for enabling edge vision and IoT applications

Challenge

Current vision chips perform visual acquisition, transmission, and processing in the form of multi-bit real-valued data.



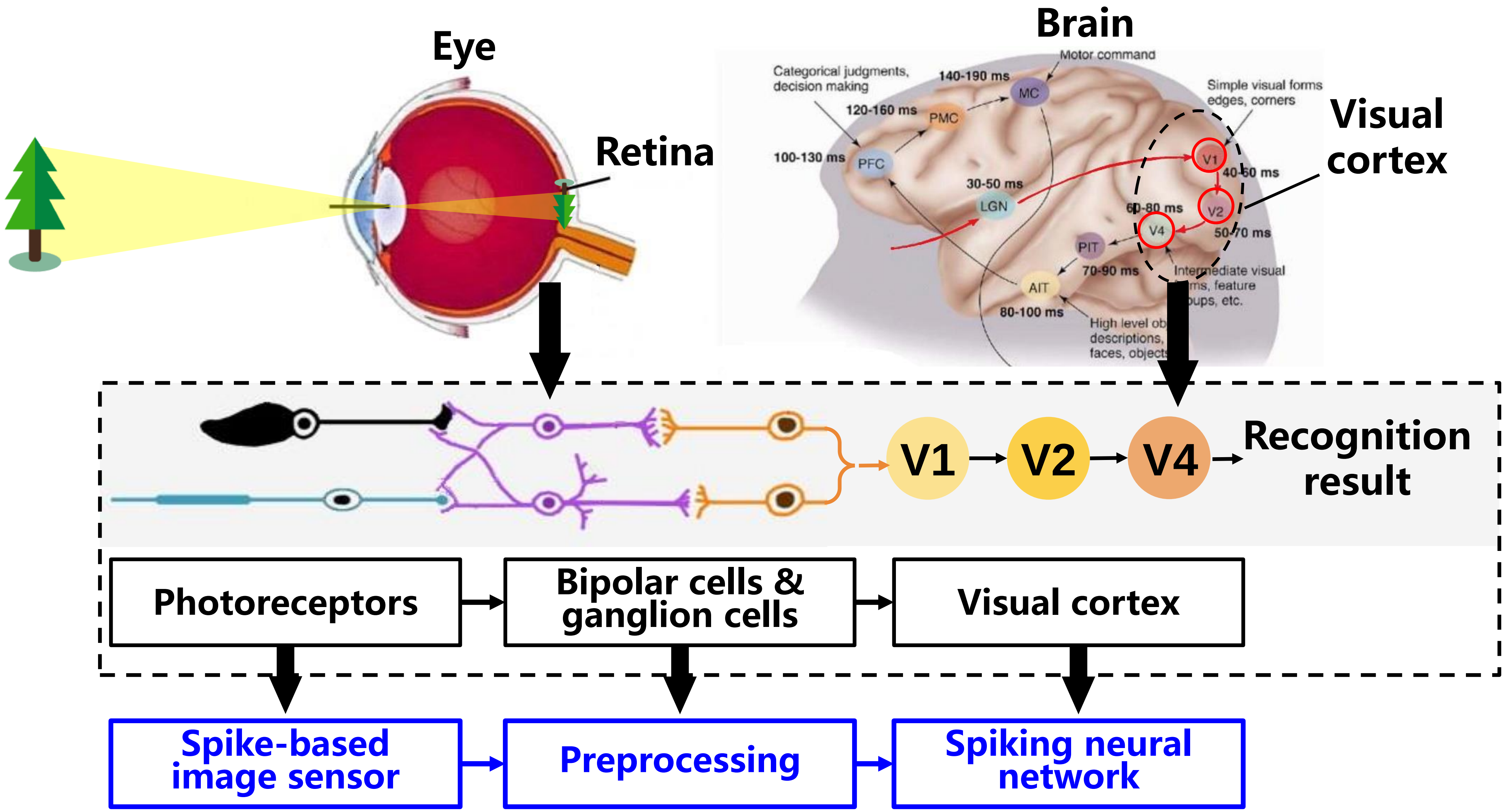
- High-speed 2D/3D imaging
- Real-time intelligent processing
- Limitations of power supply

Challenges →

- Large data volume
- Complex computation
- High cost for 2D/3D vision
- High latency and power

A new paradigm for future vision chips is urgently needed as performance gains from the process are not enough to meet demand.

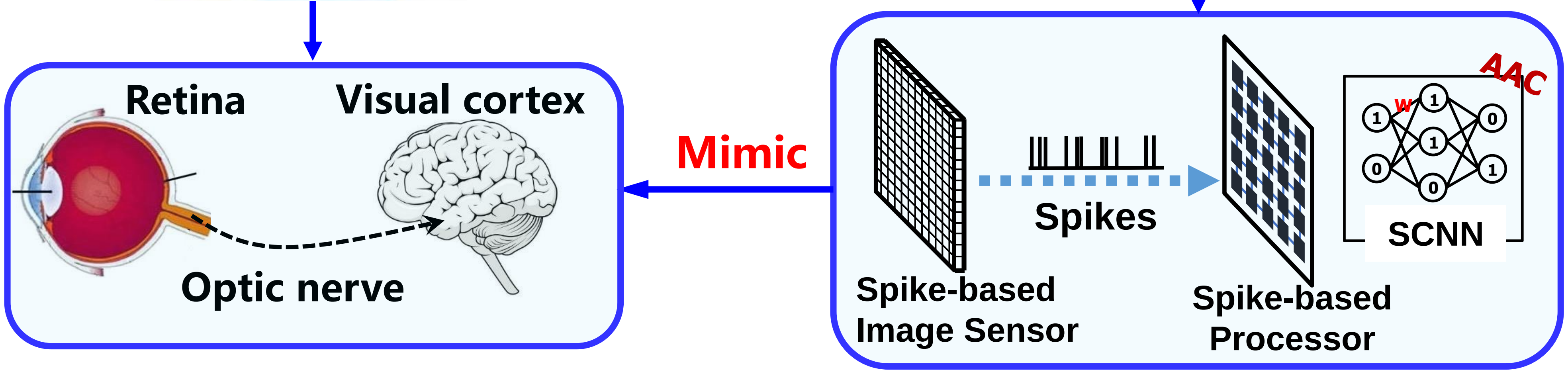
The Human Visual System



A vision chip adopts full spiking visual flow to mimic the human visual flow.

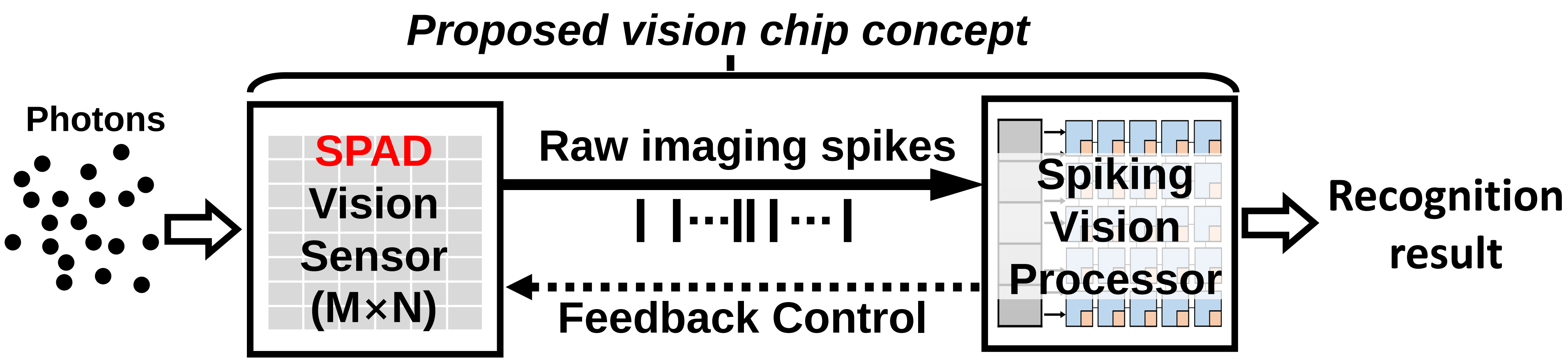
Proposed Bio-inspired Spiking Vision Chip

- Integration of image sensor and processor
- A mimicking of human vision
- Full spiking visual flow



The bio-inspired spiking vision chip integrates a spike-based image sensor and a processor to mimic the human visual system and realize a full spiking visual flow.

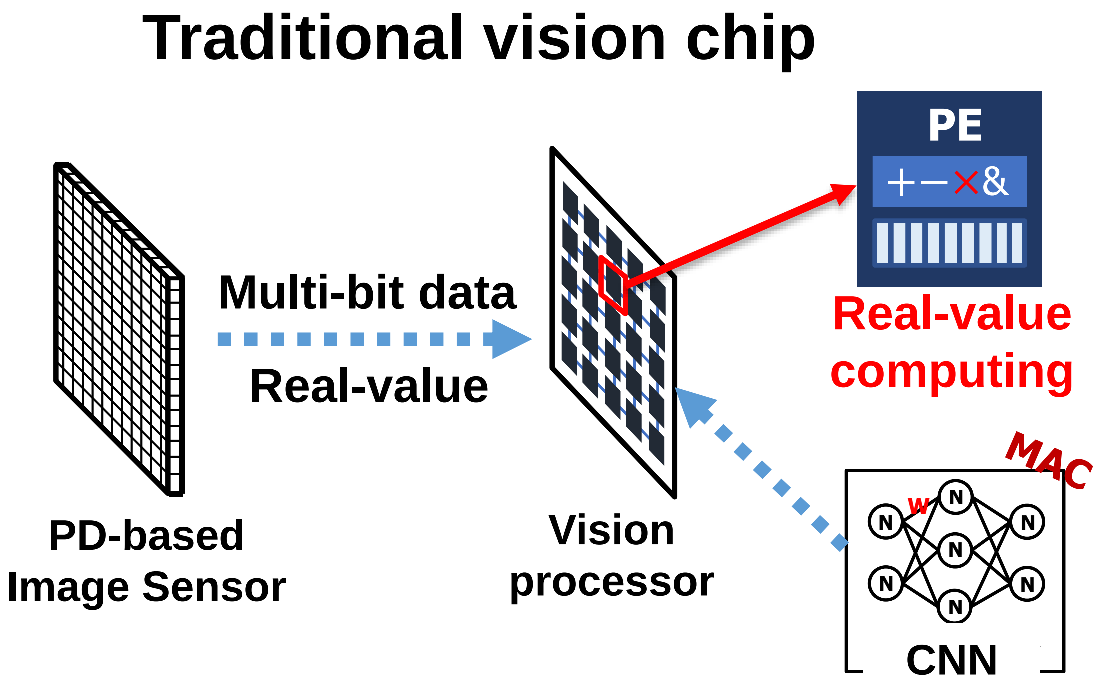
Our Approach

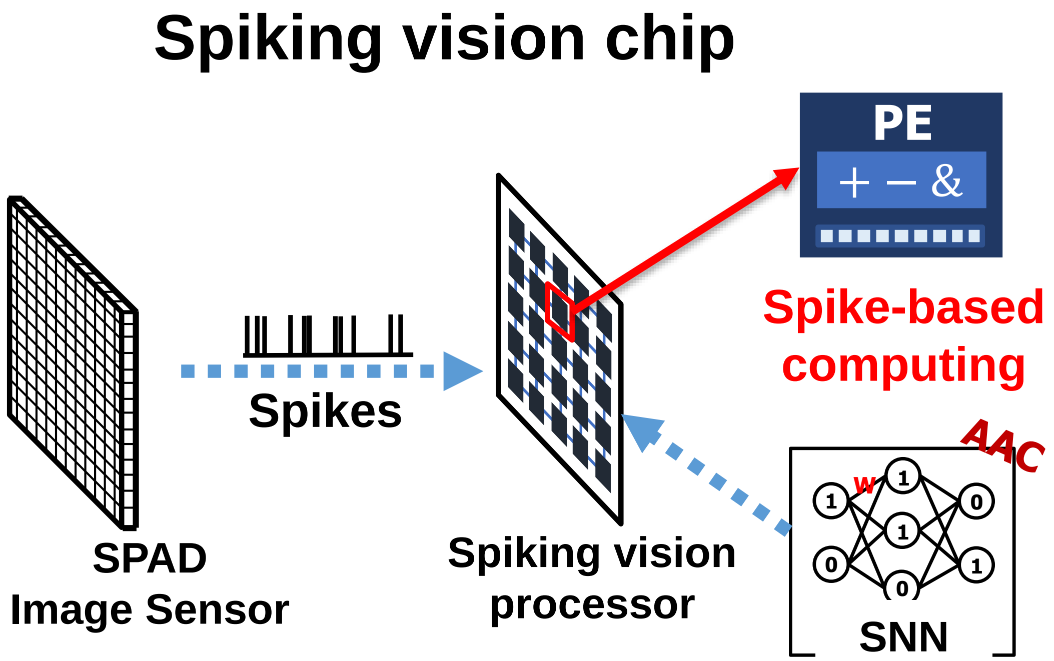


- Versatile spike-based Imaging and Processing
- Decrease Data Volume and Computation Load
- Spiking Vision Processor
  →Spiking ISP and SNN-based intelligent recognition
- Adaptive Imaging Adjustment

Versatile vision ability with low latency!

Advantage of Spiking Vision Chip





☹ large data volume

☹ complex ALU (MAC)

☹ large latency and high cost

VS.

☺ low data volume

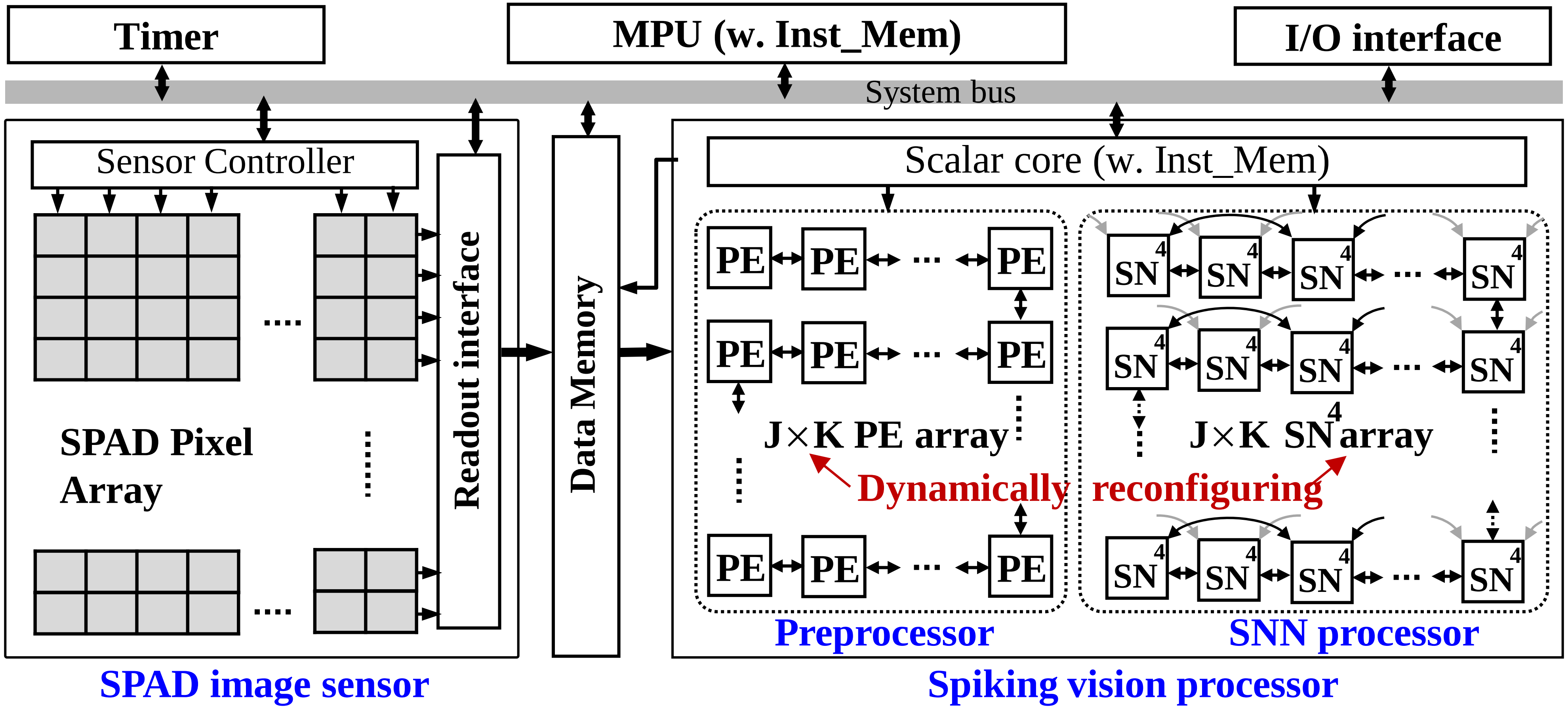
☺ simplified ALU (ACC)

☺ low latency and compact size

Outline

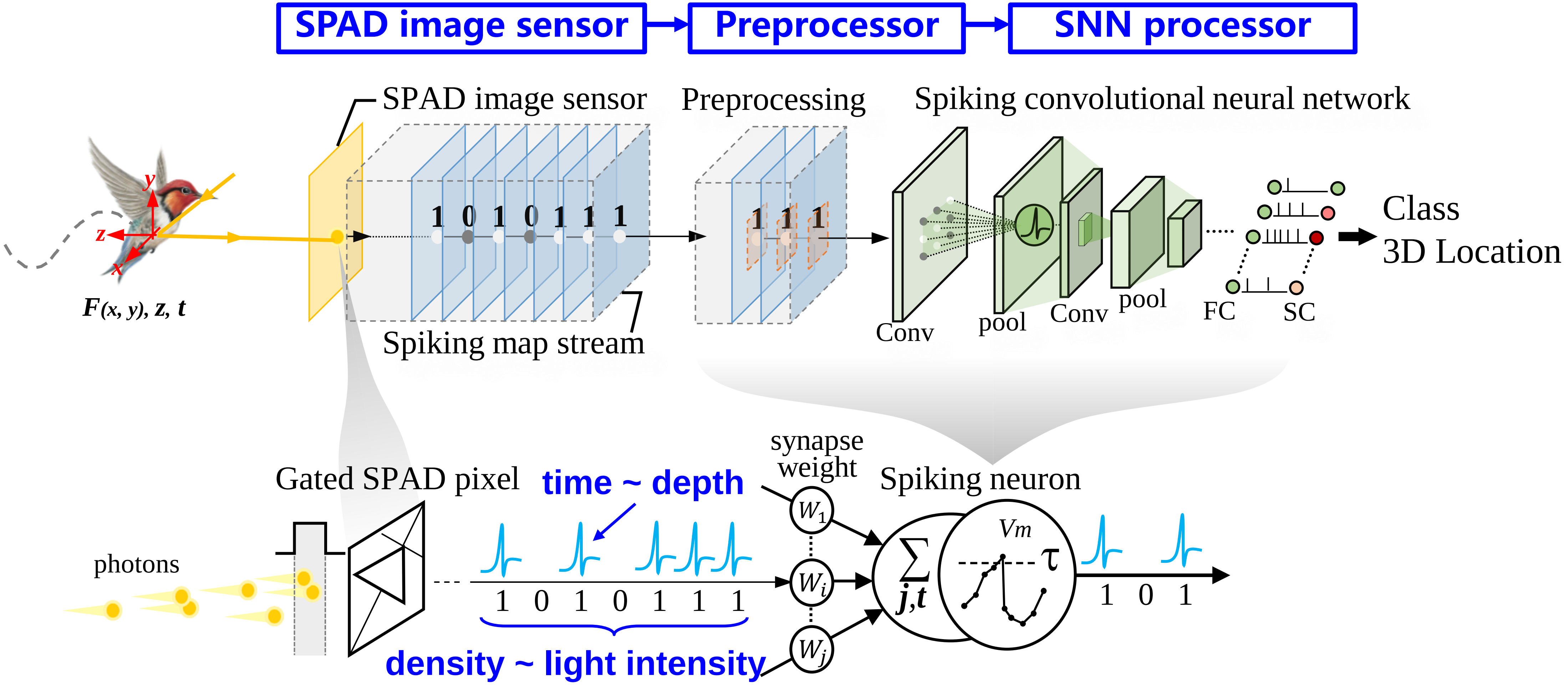
- Background
- **Chip Architecture**
- Key Techniques
- Results and Comparison
- Discussion

Chip Architecture



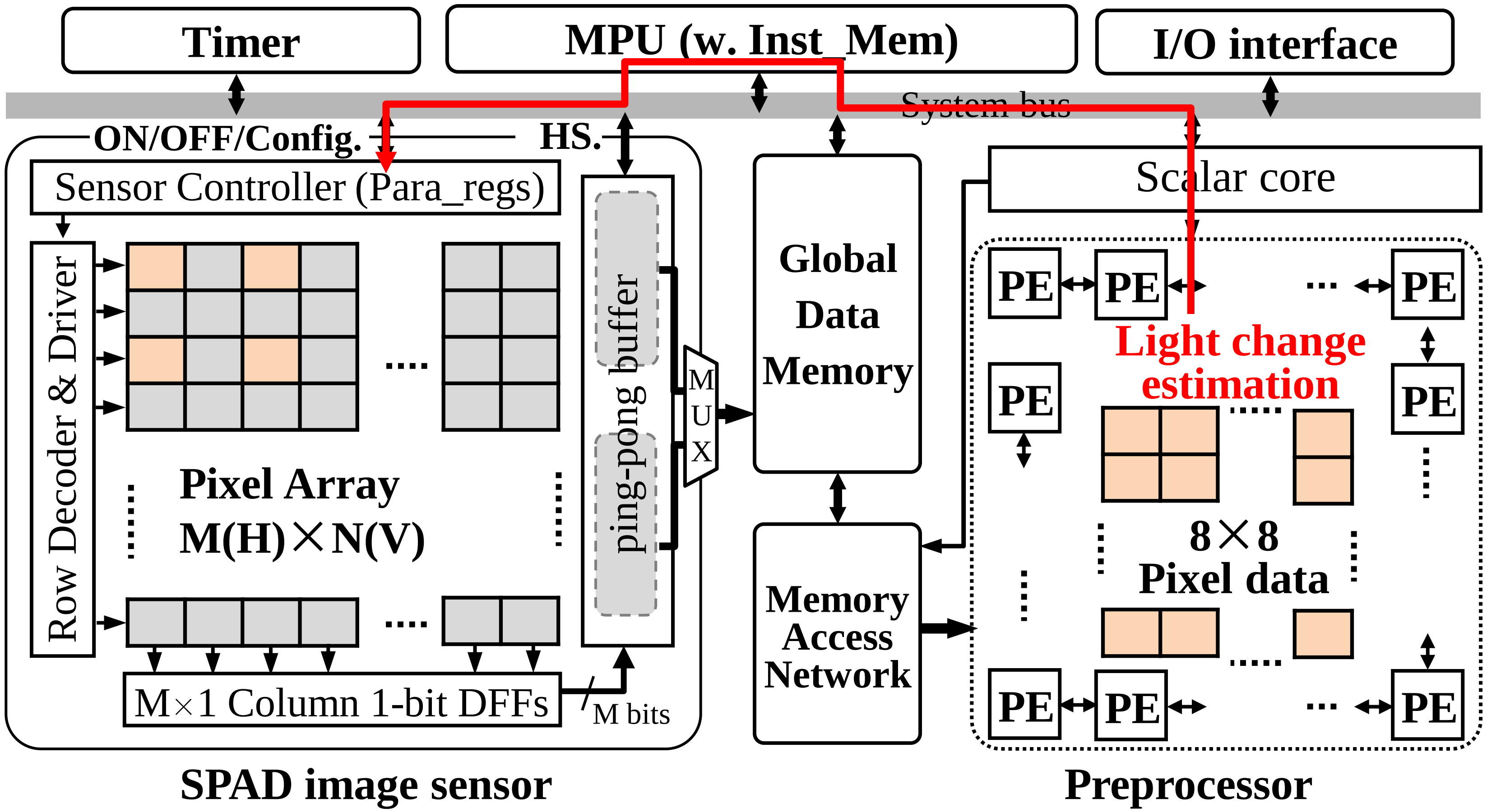
- SPAD image sensor → Naturally generate spike-based 2D/3D imaging data
- Spiking vision processor → Reconfigurable for preprocessor and SNN processor
- Processor-MPU-Configurable SPAD image sensor → System-level feedback adjustment

Spiking Visual Flow



- Bio-inspired full spiking visual flow → low end-to-end latency → light-adaptation
- SPAD imaging data and spike-based computing → versatile intelligent 2D/3D vision
- Spiking map stream → regular data flow for dynamic reconfigurable design

On-chip Feedback Adjustment



- Light change estimation
  - → Subsampled 8×8 pixel data
  - → detect CNT changes
  - → Programmable threshold setting

SPAD image sensor ← MPU ← Preprocessor

Advantage of Spiking Vision Chip

SPAD-based spiking vision chip

Feature:

1) Low data volume

2) Spike-based computing

3) Low latency

4) Structured spiking map

Advantages:

☺ Decrease hardware cost

☺ Simplified ALU (ACC)

☺ Instant feedback adjustment

☺ Time-divided multiplexing reconfigurable design

Key Techniques to Spiking Vision Chip

imaging-while-processing

SPAD image sensor

- Gated SPAD pixel
- Configurable imaging mode

Processing circuits

- Reconfigurable processing element
- Parallel computing

Modeling analysis

Processing algorithm

- Image enhancement
- Spiking neural network

Spike-computing oriented design

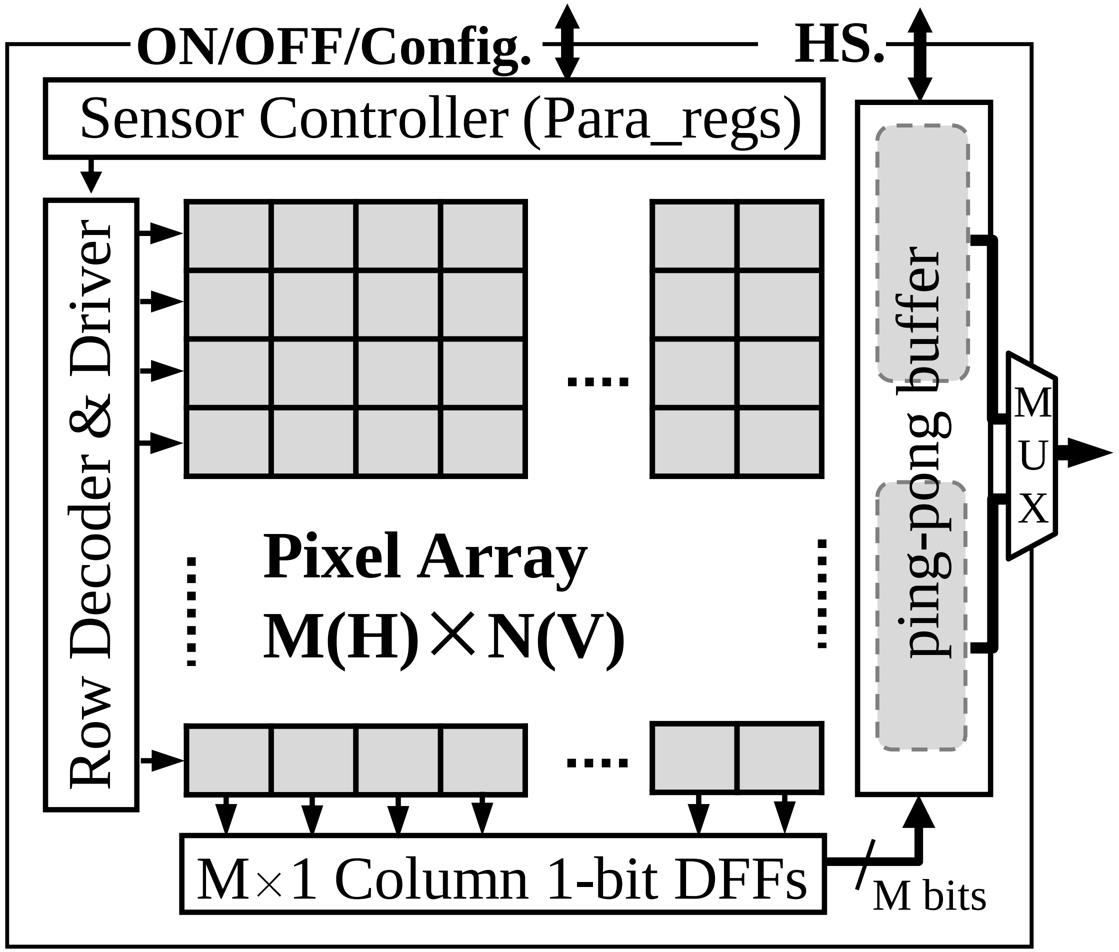
Outline

- Background
- Chip Architecture
- **Key Techniques**
  - SPAD image sensor
  - Reconfigurable spiking vision chip
  - Spike-based processing algorithm
- Results and Comparison
- Discussion

Outline

- Background
- Chip Architecture
- **Key Techniques**
  - SPAD image sensor
  - Reconfigurable spiking vision chip
  - Spike-based processing algorithm
- Results and Comparison
- Discussion

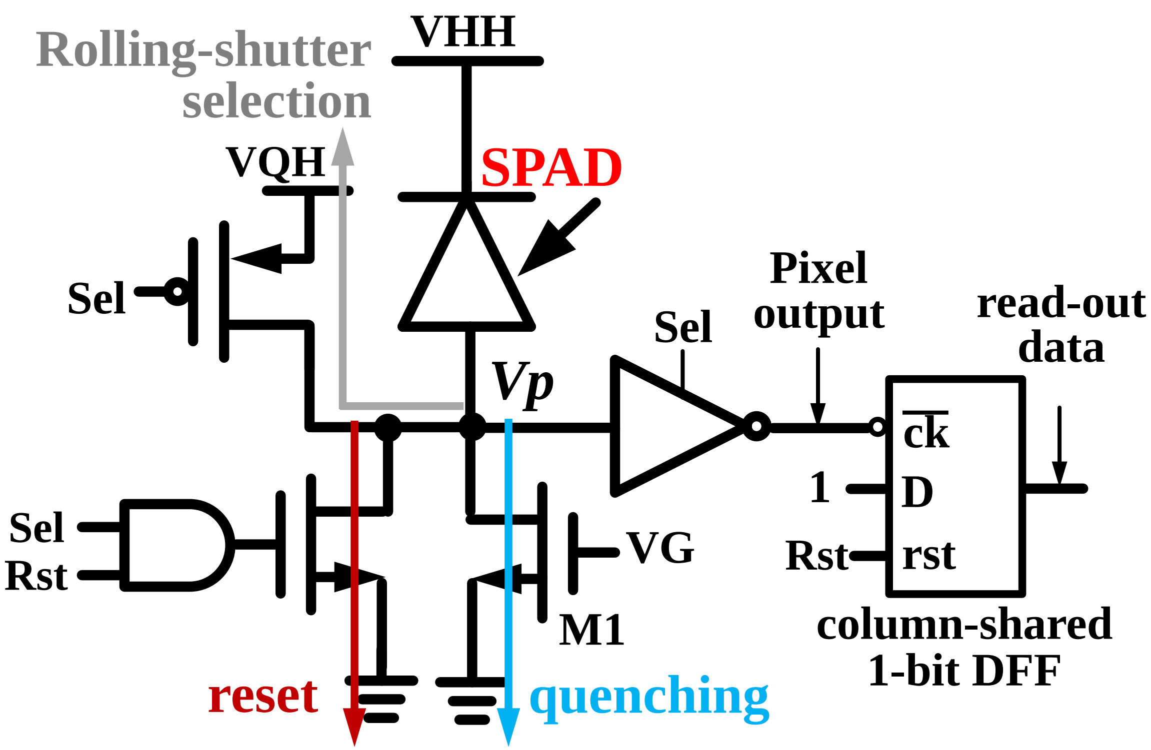
Configurable Adaptive SPAD Imaging

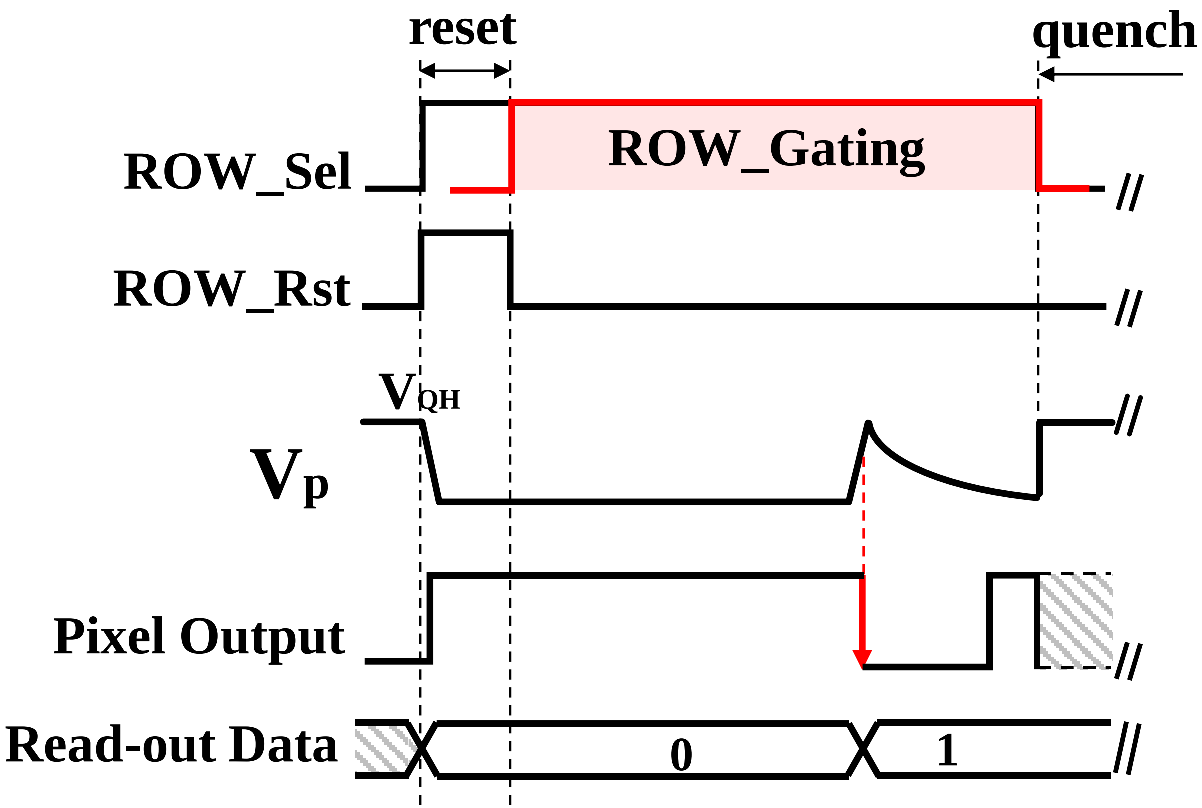


SPAD image sensor

- **Gated pixels**
  → Configurable exposure time
- **Rolling-shutter operation**
  → Stabilize SPAD array bias
- **Adaptive**
  → Adjust imaging parameters based on visual processing results

Gated SPAD Pixel

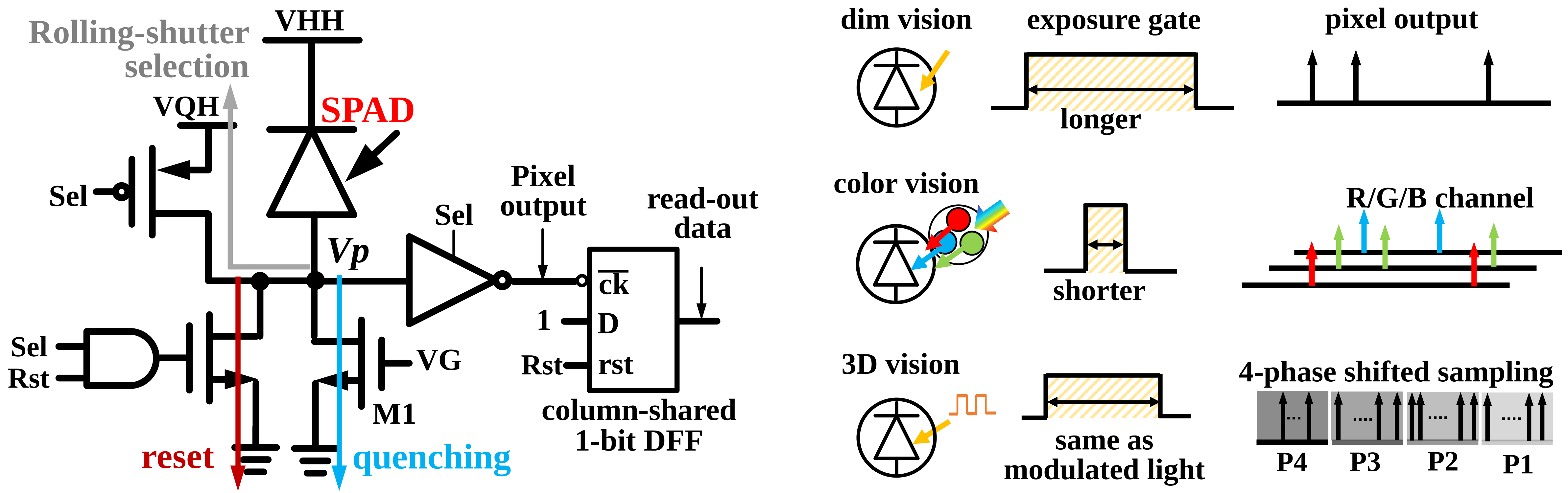




- **Gated pixel structure**
  - External reset
  - Passive quench

- **Configurable gating**
  - via *SEL* & *RST*

Configurable Imaging Mode



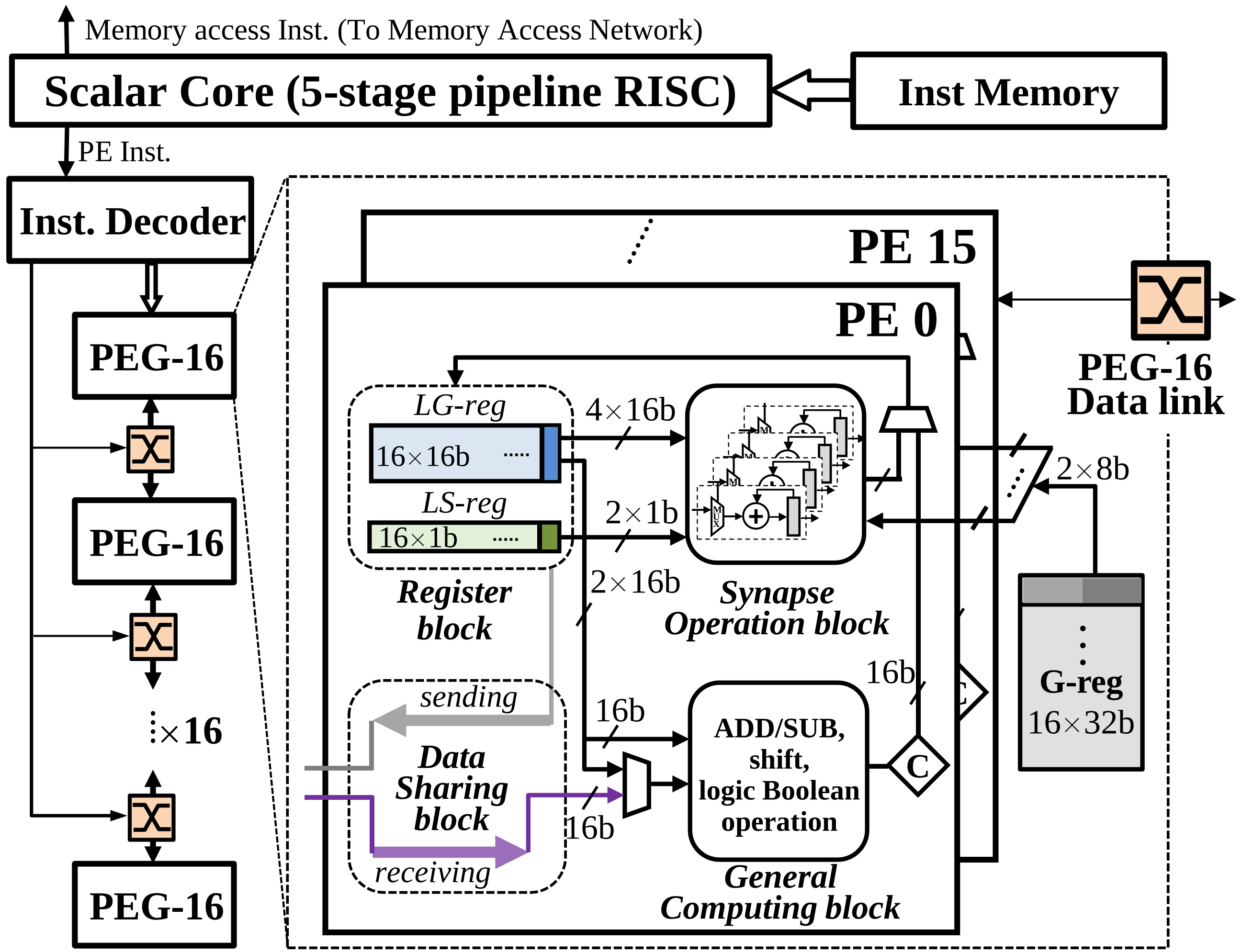
■ **Configurable gating enables**

1. **adaptive 2D imaging**
2. **iToF based 3D imaging**
3. **dim imaging ability**
4. **color imaging w. RGB color filter**

Outline

- Background
- Chip Architecture
- **Key Techniques**
  - SPAD image sensor
  - Reconfigurable spiking vision chip
  - Spike-based processing algorithm
- Results and Comparison
- Discussion

Reconfigurable Spike-based Processing



- PE array

→ Parallel computing

- Reconfigurable PE array

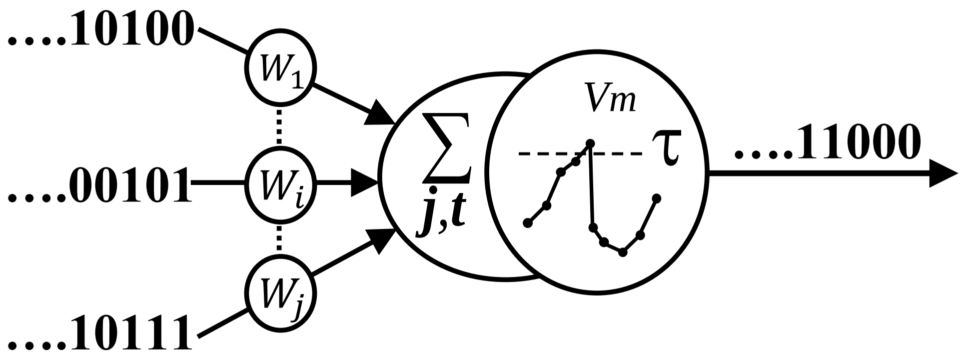
→ Hardware-efficient for preprocessing and SNN

- Instruction-level programmable

→ Support various network size

Spiking Neuron: Integrate-and-Fire (IF) Neuron

Spiking neuron model



Input/output: spikes, 1 bit

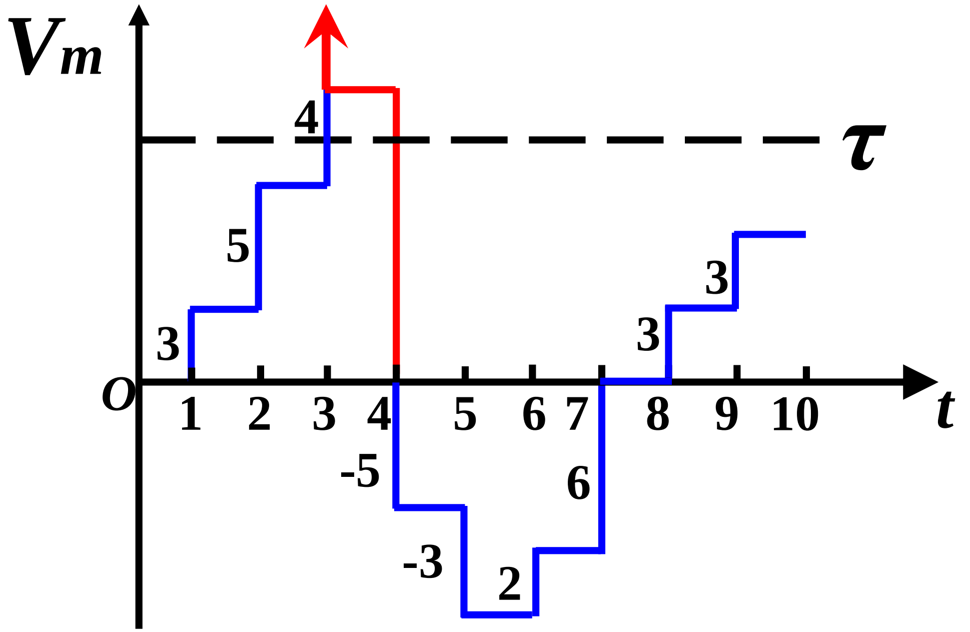
W: synapses weight, 8bit

Vm: membrane potential, 13~16bit

τ: firing threshold, 1bit

$$\text{if } V_m(t) \geq \tau, S_o(t) = 1 \ \& \ V_m(t)=0$$

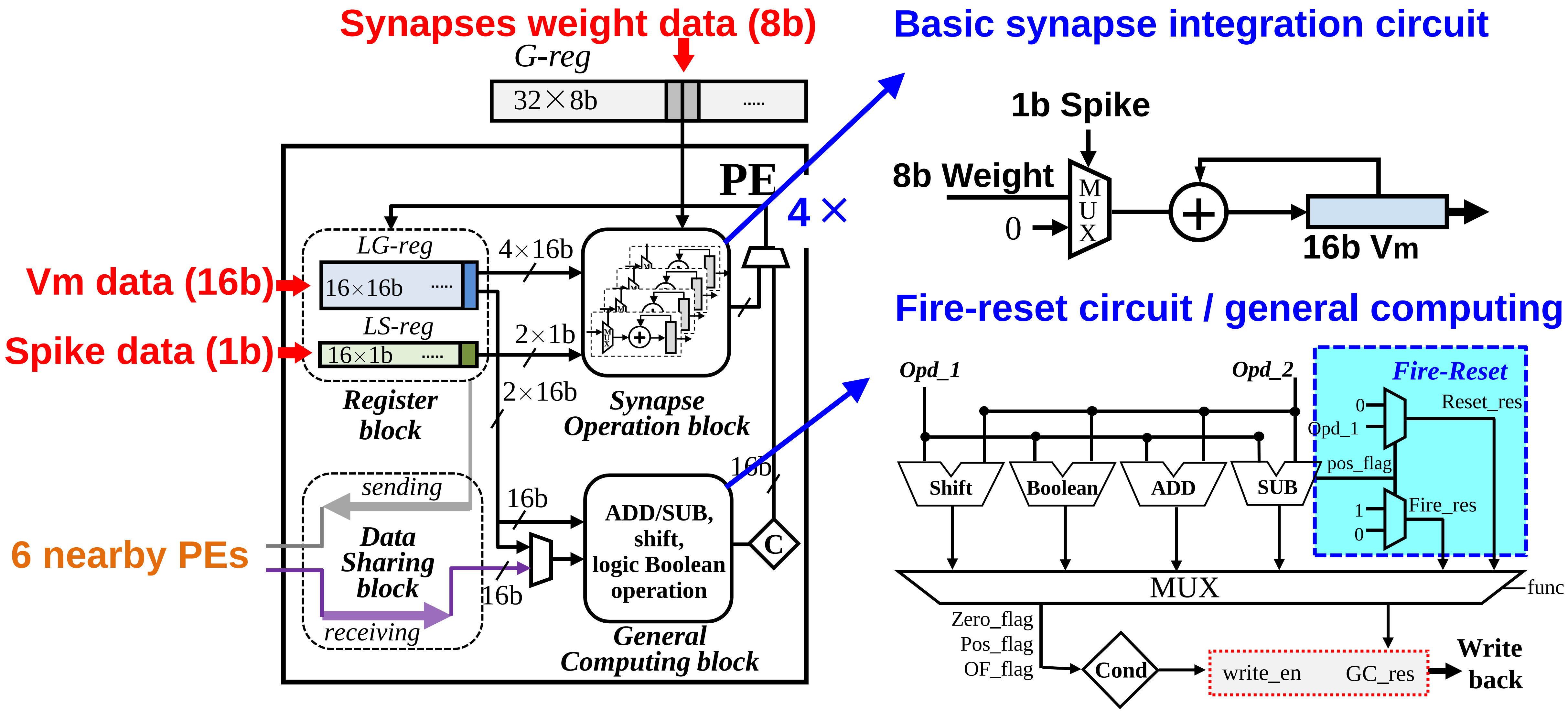
Fire-reset operation



Integration operation

$$V_m(t) = V_m(t-1) + \sum_{j \in \Gamma_j} W_{ij} \times S_i(t)$$

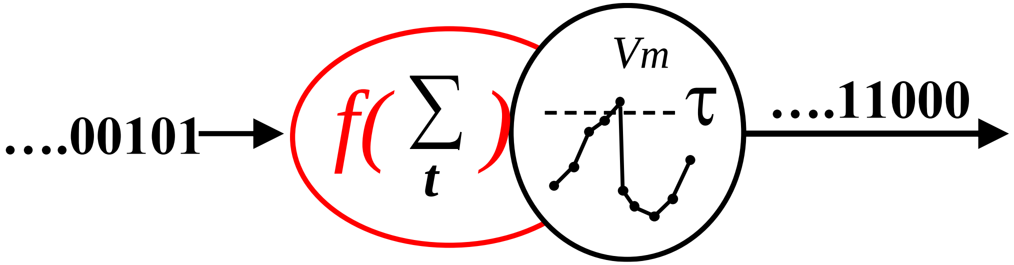
Processing Element (PE)



PE offers IF neuron computing, flexible local data access, and nearby data sharing.

IF Neuron for Preprocessing

IF neuron model with temporal filtering



***f*-function:**

1) 2D visual signal enhancement

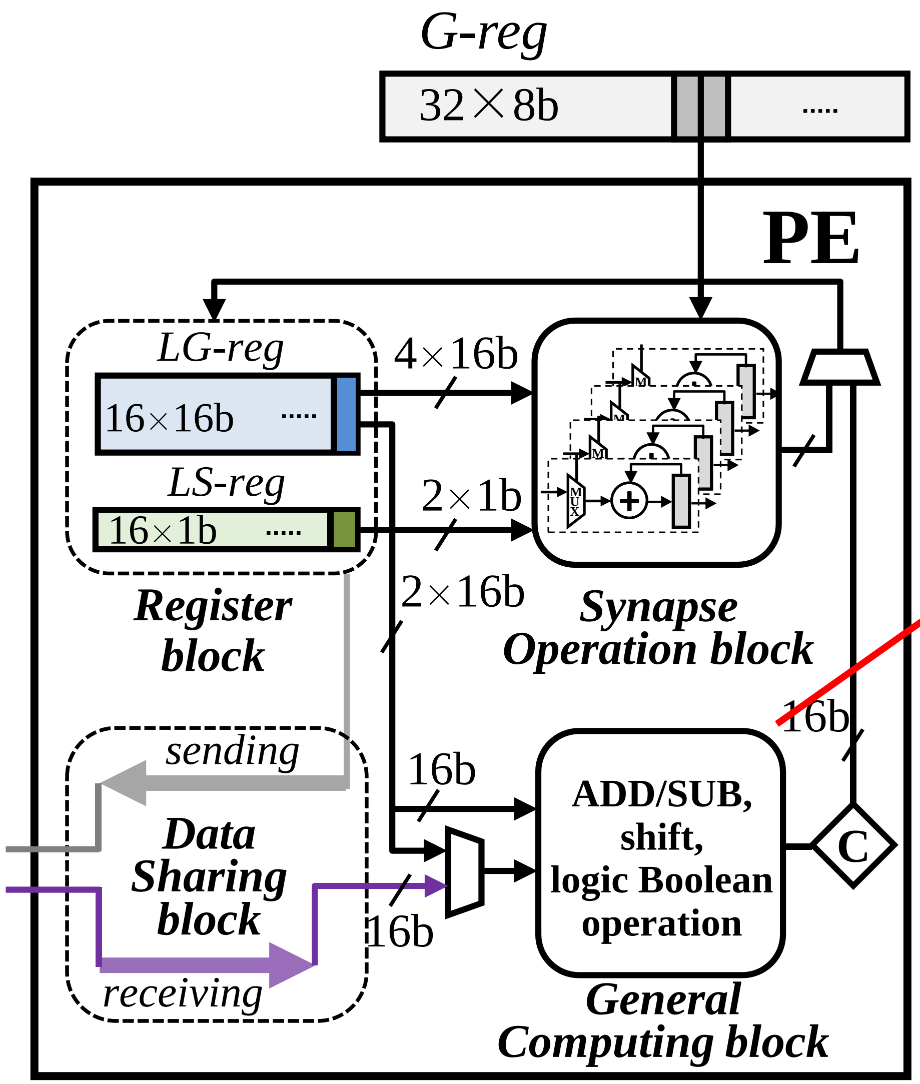
e.g. $f = log_{1-PDE}(\frac{1-R}{1-\Delta t \times DCR})$

2) 3D visual reconstruction

Feature of PE for preprocessing:

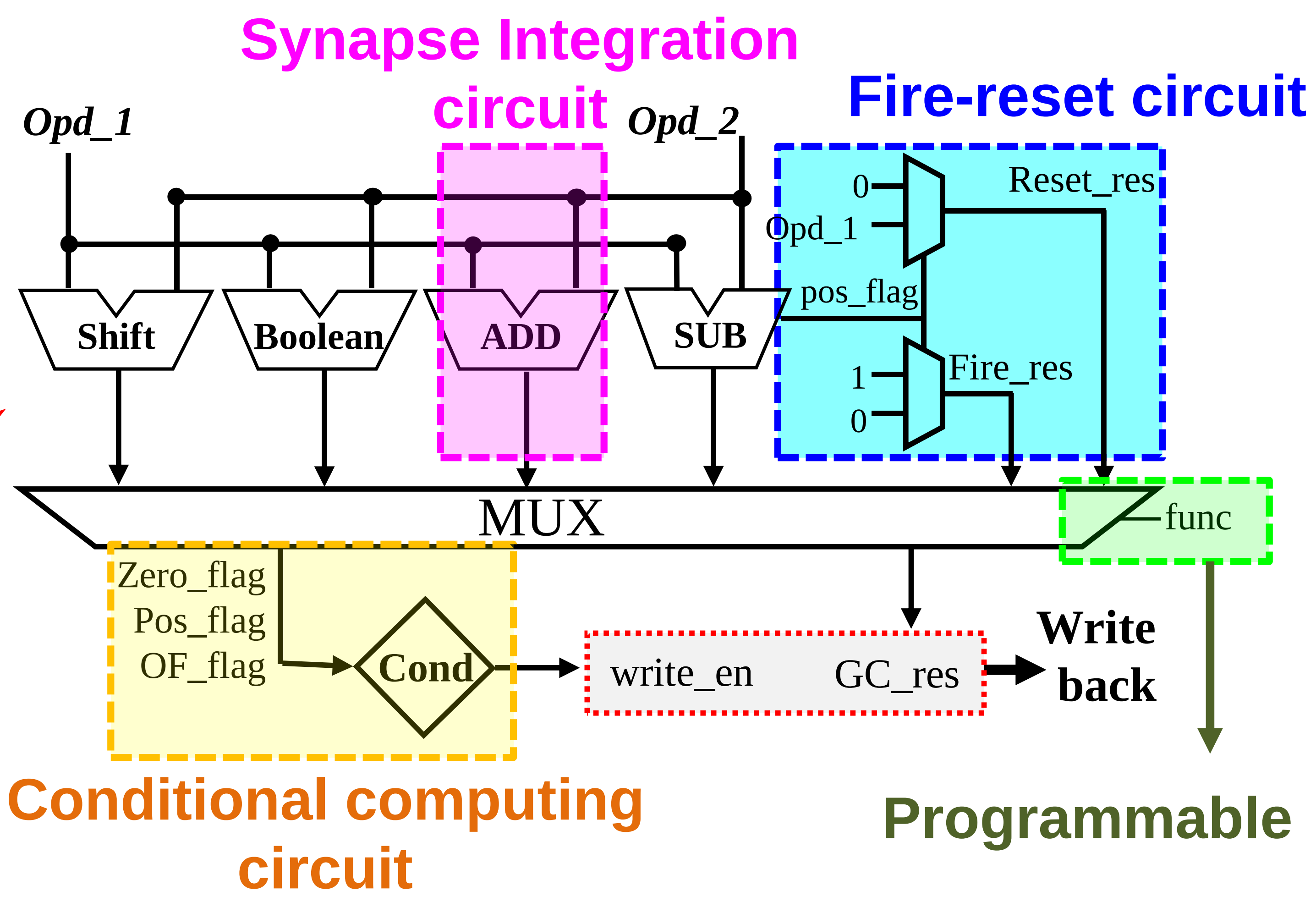
1. Pixel-wise
2. Temporal accumulation
3. Flexibly programmable ALU
4. I-F process

PE for Preprocessing



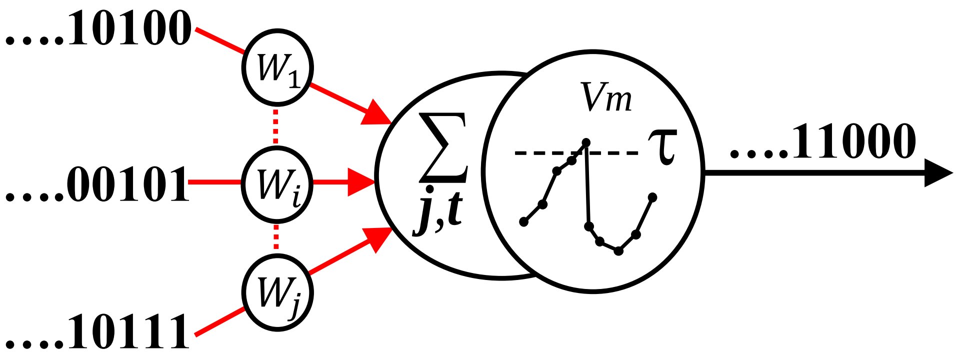
Disable: Synapse operation block
Data sharing block

Fire-reset circuit / general computing



IF Neuron for SNN

IF neuron model with dense synaptic connection



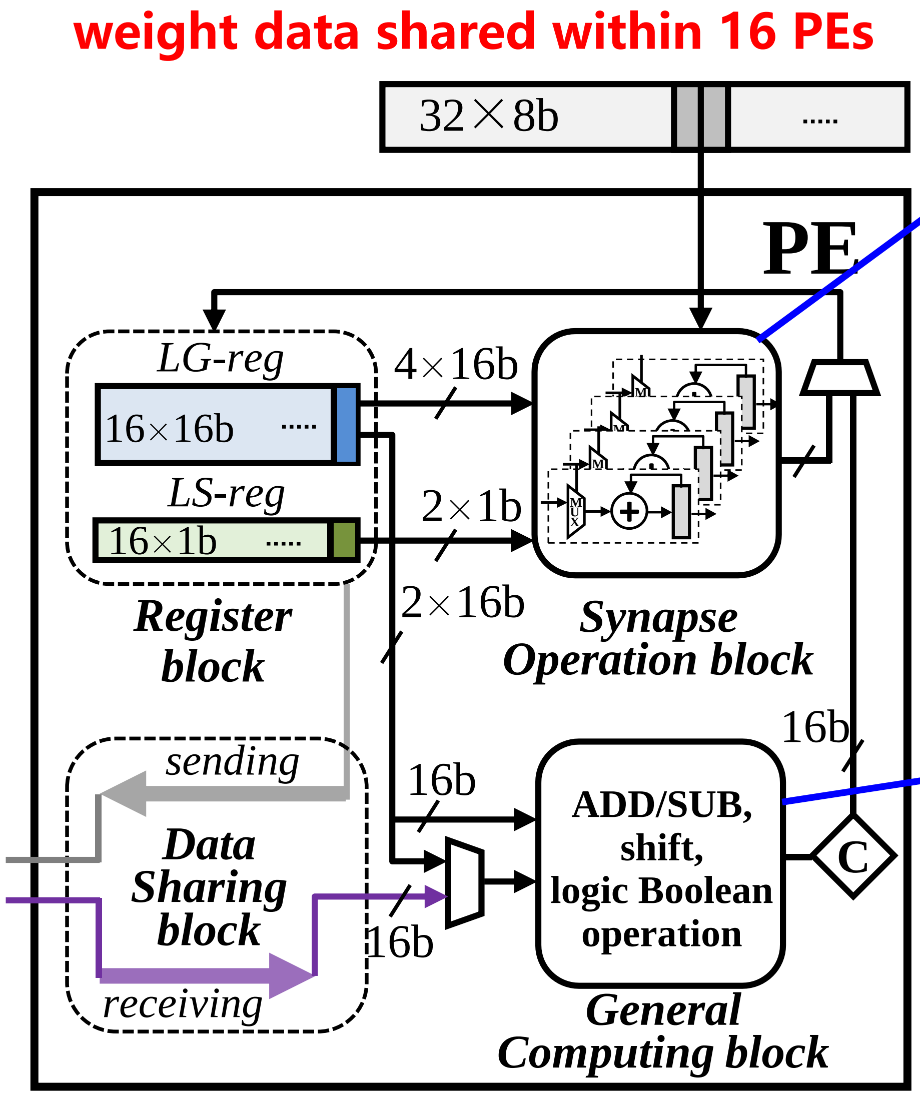
Dense connect
$k \times k \times C_{in} \times C_{out}$
(e.g. 3×3×16×64)

Weight kernel
Shared within layer

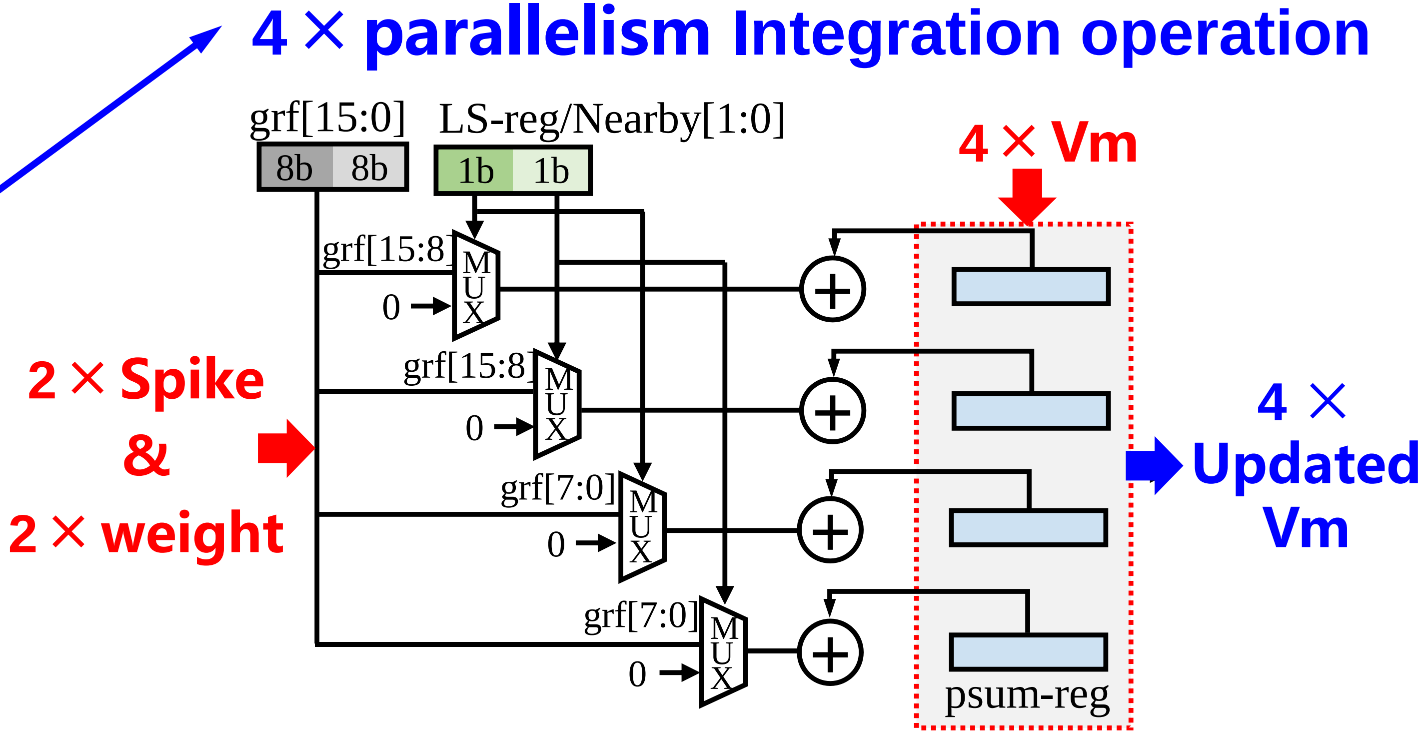
Feature of PE for preprocessing:

1. Accelerate synapse integration
2. Increase local data reuse
3. Efficient data access
4. I-F process

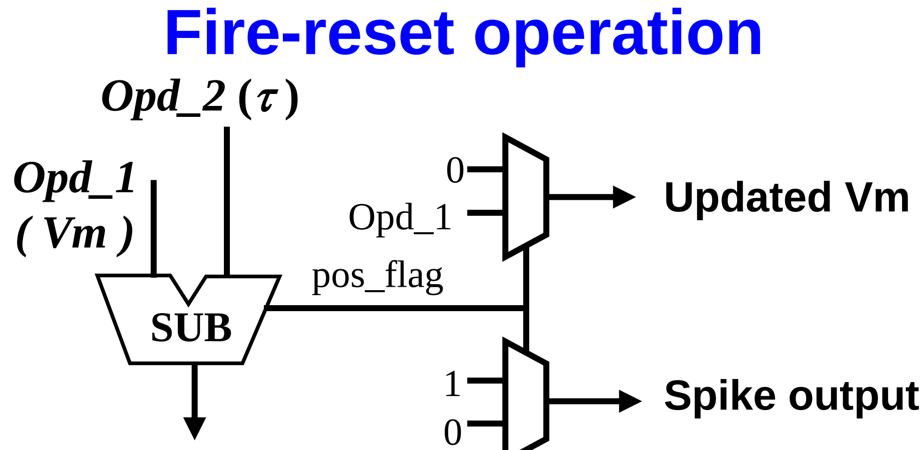
PE for SNN



weight data shared within 16 PEs
32×8b
.....
PE
LG-reg
16×16b
LS-reg
16×1b
4×16b
2×1b
2×16b
Register block
Synapse Operation block
16b
sending
Data Sharing block
receiving
ADD/SUB, shift, logic Boolean operation
C
General Computing block

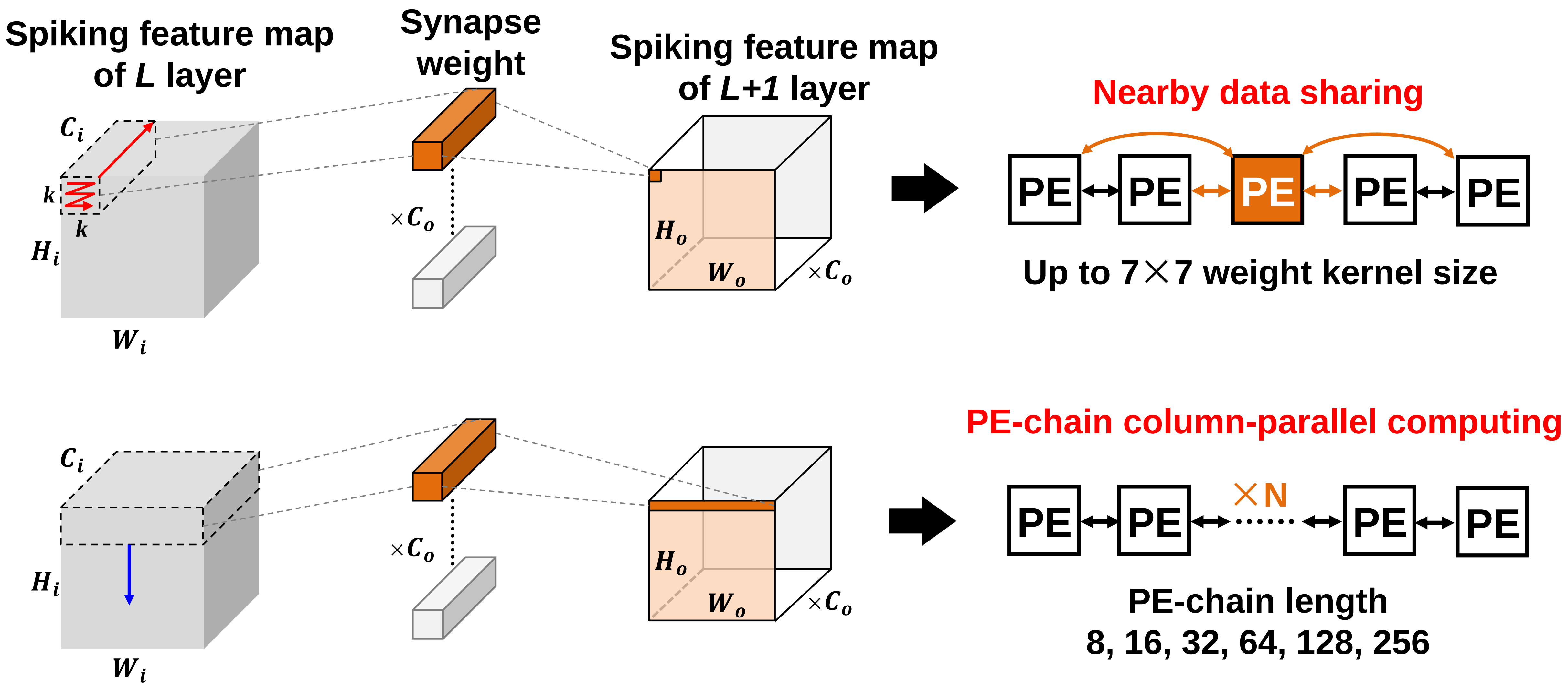


4× parallelism Integration operation
grf[15:0]
LS-reg/Nearby[1:0]
8b 8b
1b 1b
4× Vm
grf[15:8]
grf[7:0]
0
MUX
2× Spike & 2× weight
psum-reg
4× Updated Vm



Fire-reset operation
Opd_2 (τ)
Opd_1 (Vm)
SUB
0
Opd_1
pos_flag
1
Updated Vm
Spike output

PE Chain Parallel Computing

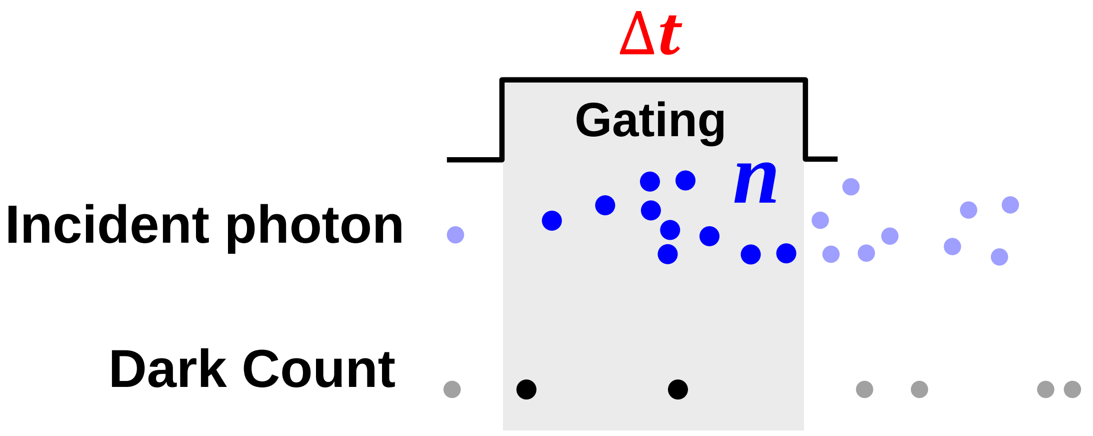


Outline

- Background
- Chip Architecture
- **Key Techniques**
  - SPAD image sensor
  - Reconfigurable spiking vision chip
  - Spike-based processing algorithm
- Results and Comparison
- Discussion

2D Visual Signal Enhancement

Denoise filtering

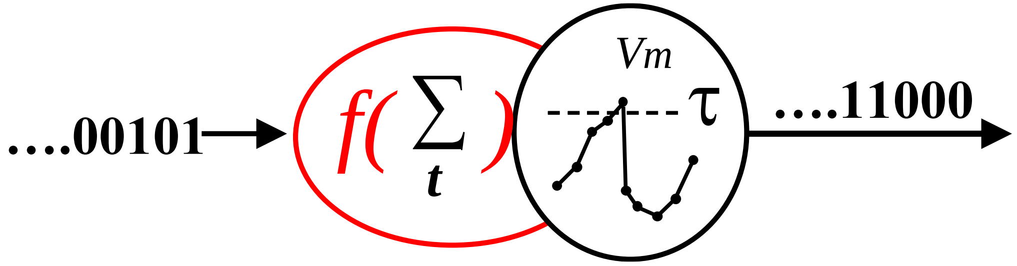


$$R \sim P_{ava} = 1 - P_{\bar{L}} \times P_{\bar{D}}$$

$$P_{ava} = 1 - (1 - PDE)^{n} \times (1 - \Delta t \times DCR)$$

Signal intensity→ $n = log_{1-PDE}\left(\frac{1-R}{1-\Delta t \times DCR}\right)$

2D depth imaging reconstruction



1) Temporal accumulation $\sum_t$
2) Denoise function f

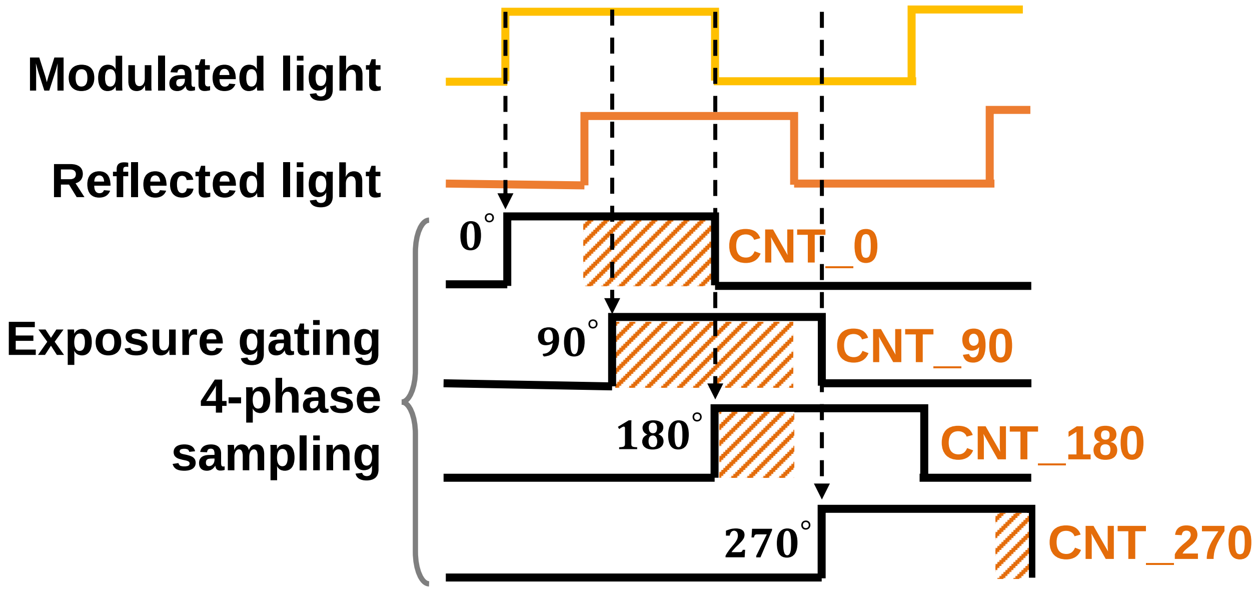
$$f(R) = log_{1-PDE}\left(\frac{1-R}{1-\Delta t \times DCR}\right)$$

Correct the effect of DCR & PDE variation

For weak lighting condition, preprocessing solves signal from noise

3D Visual Reconstruction

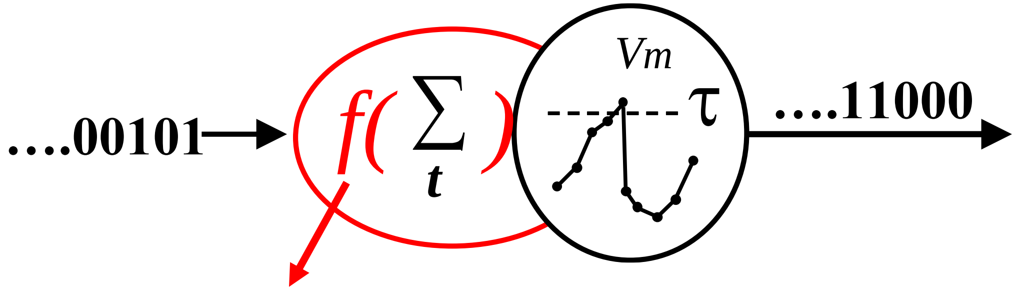
iToF-based depth imaging



- Modulate frequency *f*
- Several exposures for each phase

  → obtain light intensity (avalanche count CNT)

3D depth imaging reconstruction



Depth-solving function:

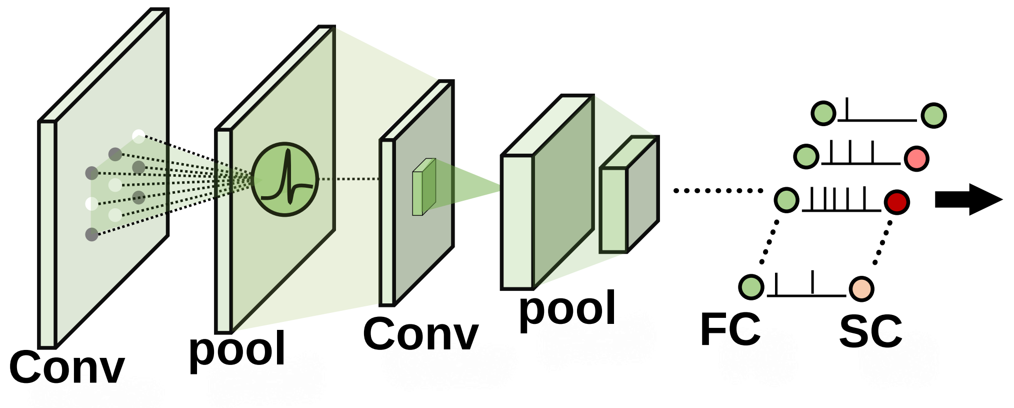
a =CNT_0 - CNT_180

b =CNT_90 - CNT_270

$$d = \frac{c}{8f}\begin{cases} \dfrac{b}{a+b} & a>0\ \&\ b>0, \\ \dfrac{-a}{b-a}+1 & a<0\ \&\ b>0, \\ \dfrac{b}{a+b}+2 & a<0\ \&\ b<0, \\ \dfrac{a}{a-b}+3 & a>0\ \&\ b<0 \end{cases}$$

Solve and encode depth information into rate-coding spike output

Spiking Convolutional Neural Network



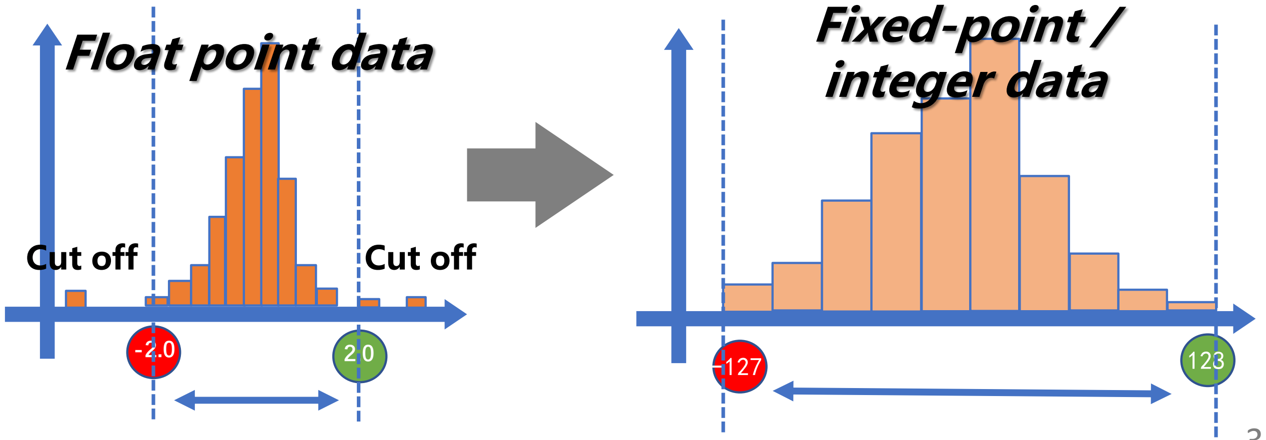
- **Hierarchical neural network**
  →Convolutional layer (Conv)
  pool layer (pool)
  full connect layer (FC)
  spike counting layer (SC)

- **Training method**
  →Converted from a well-trained real-value CNN with the same network structure

- **Network structure**

2D classification	Output
conv5-12, avg-pool/2, conv5-64, avg-pool-2, FC-10	Confidence of digits 0-9 (MNIST)
3D localization	**Output**
Conv64×1-1	Horizontal and depth position (X, Z)

- **On-chip deployment**
  →Quantized weights and cut off outlier

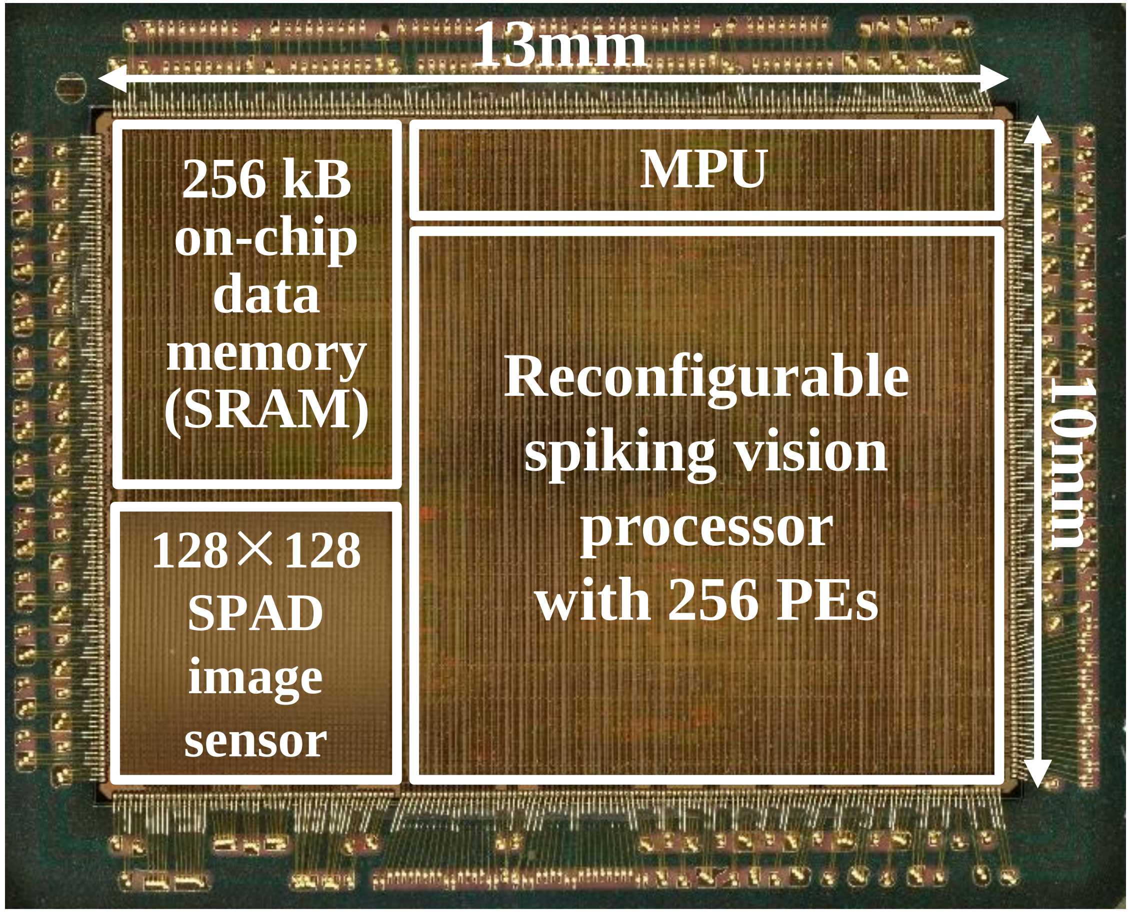


Outline

- Background
- Chip Architecture
- Key Techniques
- **Results and Comparison**
- Discussion

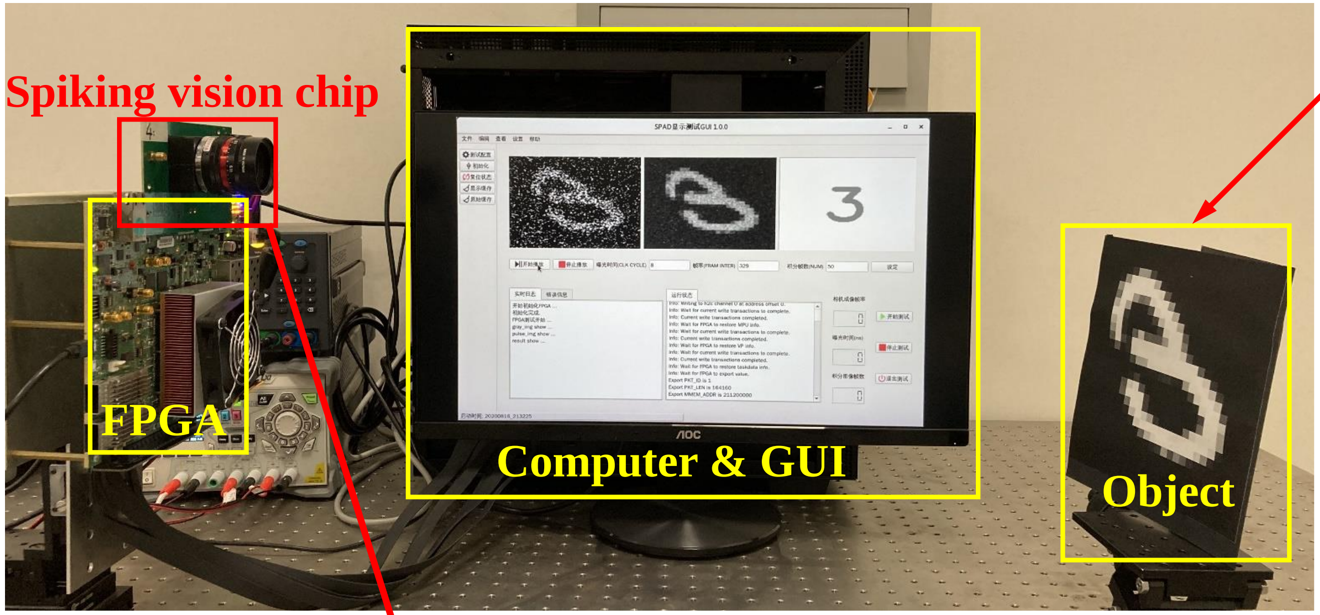
Implemented Vision Chip

Chip Microphotograph



	Specifications	
Technology	180nm CMOS	
Clock frequency	80 MHz	
Supply voltage	1.8V (Logic), 11V (VHH), 0.3V (VG)	
SRAM	256 kB (Data), 64 kB (Inst)	
Imaging rate	100,000 SMps	
PE array	256	
	Preprocessor	SNN processor
# neurons	256	1024
Peak Performance	20.48 GSOPS	81.9 GSOPS

Measurement Setup



Test board

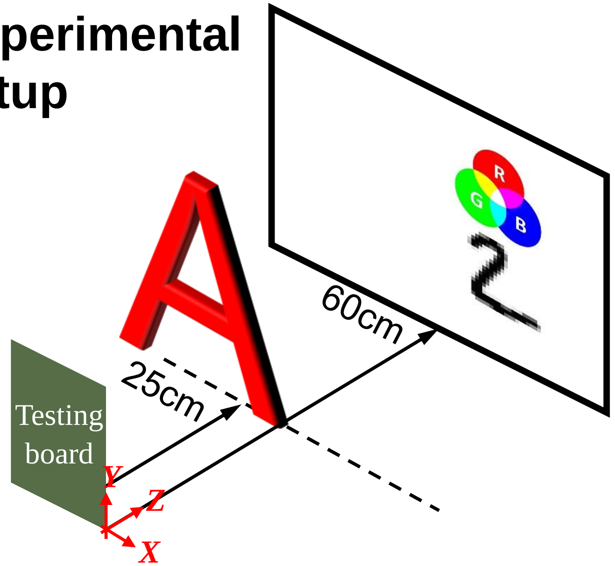
Experimental setup:
Object for imaging
LCD for classification

Different lighting conditions and noise level is simulated by setting:

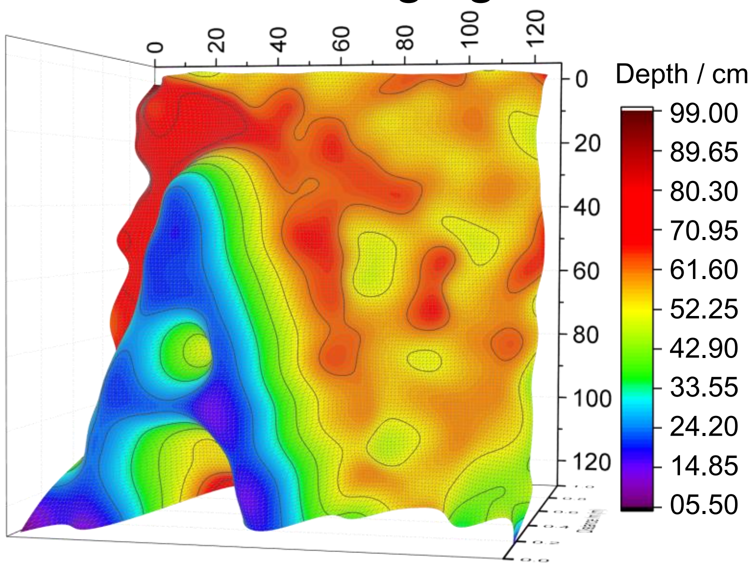
- screen brightness of LCD
- screen contrast of LCD
- dataset picture contrast

Measured Imaging Results

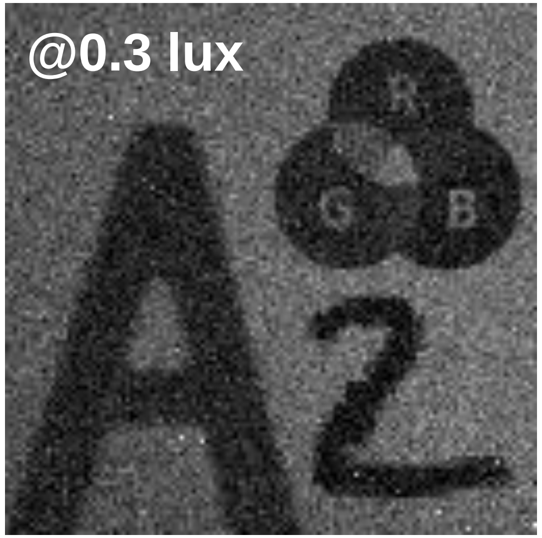
Experimental setup



3D imaging

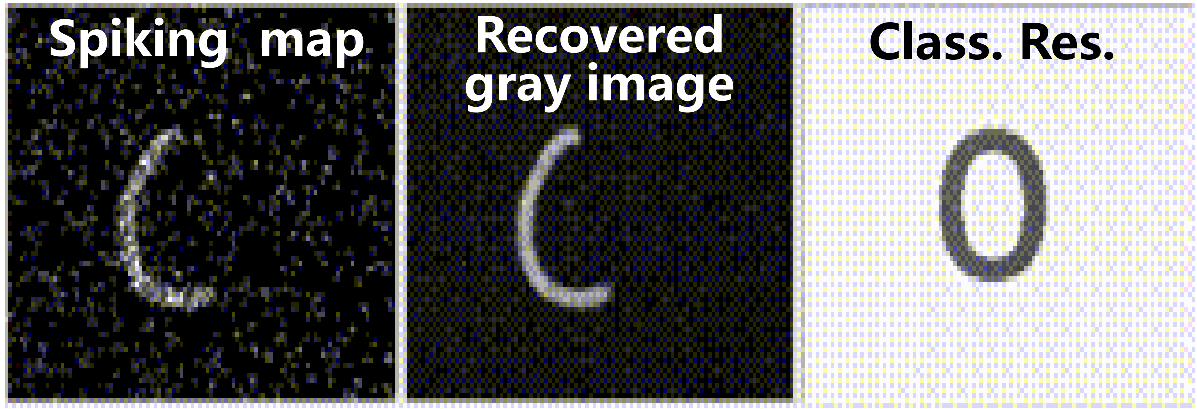


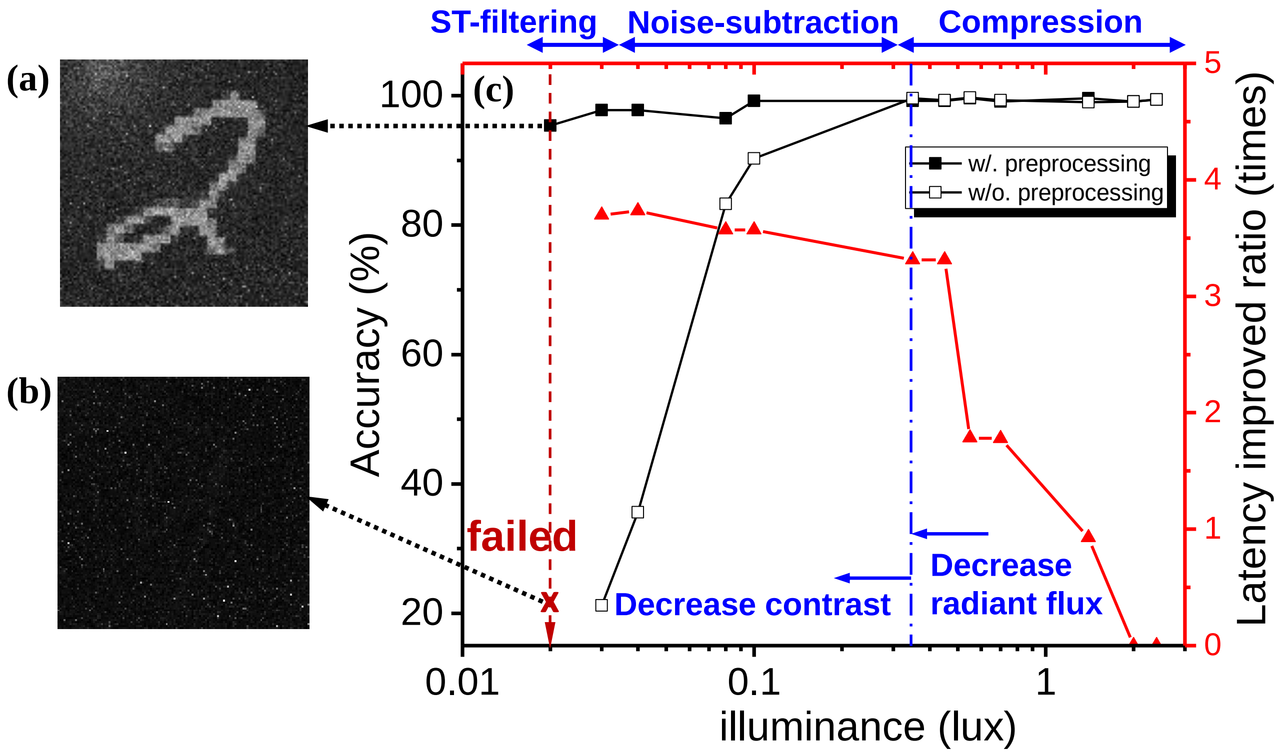
2D imaging



- Dynamic range 100 dB
- 15.75% PDE @ 503 nm
- 3D depth imaging error 2.7cm
- 2D color vision and dim vision ability
- dim-vision classification @0.02lux

MNIST Classification





Bright vision

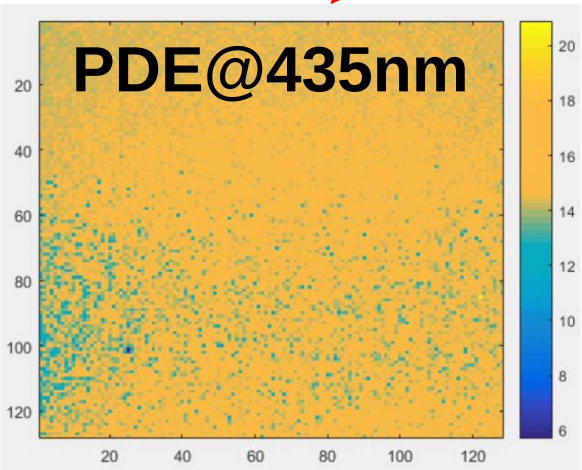
- A 5-layer SCNN
- 99.33% Acc. 300 infer/s @ MNIST

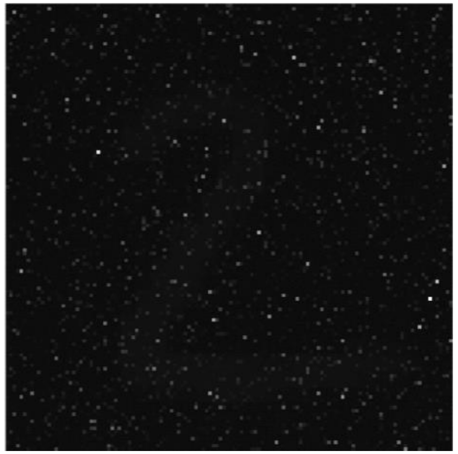
Dim vision w. preprocessing

- Efficiently improve the SNR
- Merely 3.9% Acc. loss @ 20 mlx
- ~4× latency improvement

Spike-based Imaging Signal Enhancement

$$n(x,y) = log_{(1-PDE(\lambda,x,y))}\frac{(1-P_a(x,y))}{1-DCR(x,y)}$$





@0.02lux

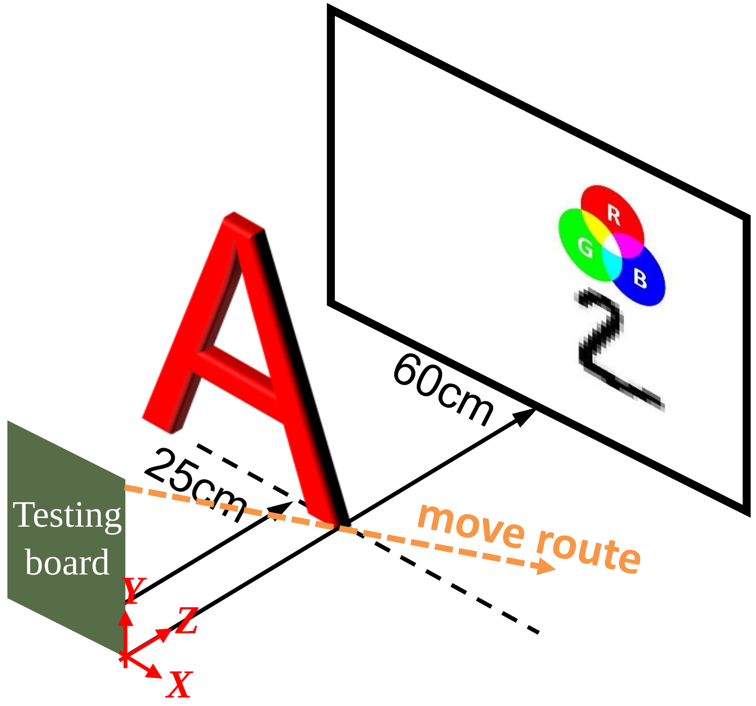
w./o Enhancement

w./o Enhancement
Float-point

w./o Enhancement
Fixed-point
approximation computation

On-chip approximation computation with fixed-point data representation can realize similar improvements on SNR.

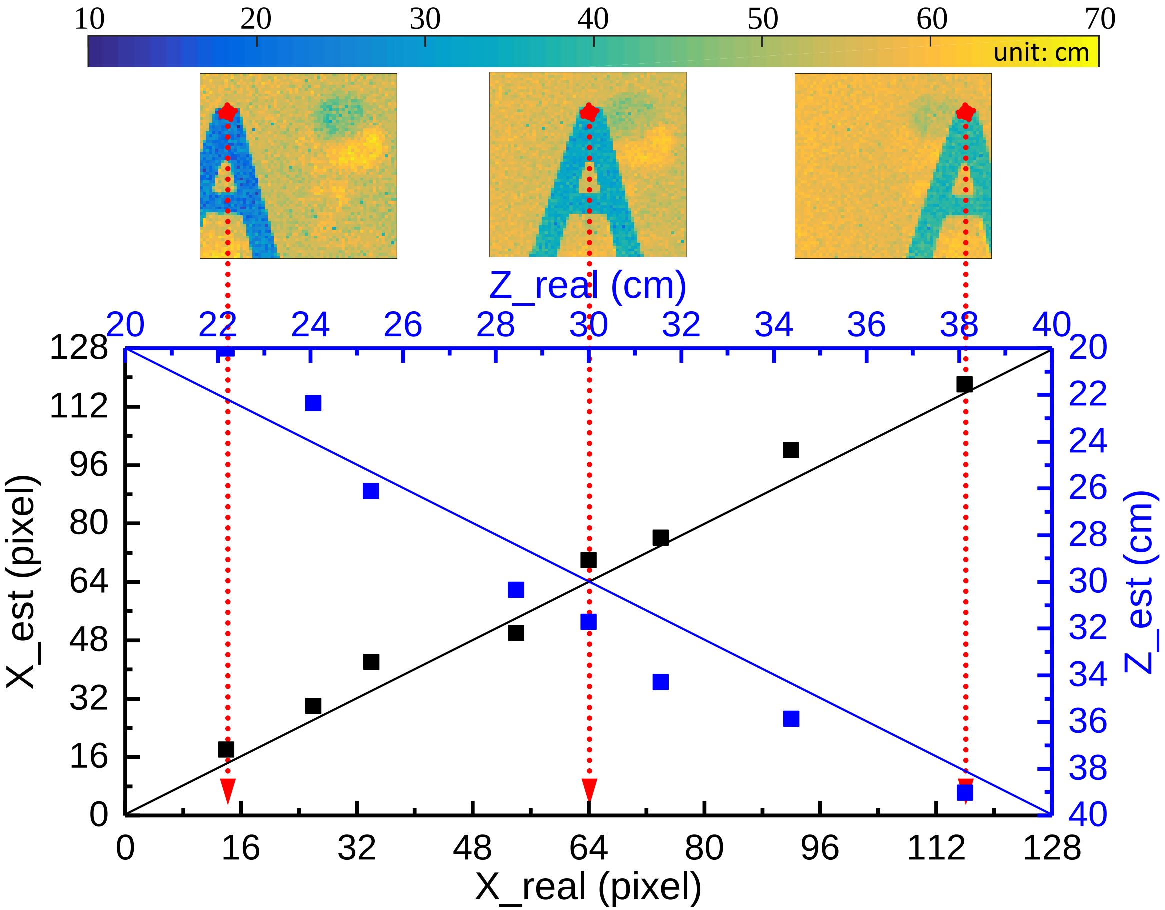
Obstacle Localization



■ **Experimental setup**

Z-direction: 20cm→ 40cm

X-direction: left → right within FoV



■ **Localization error**

1.68 cm @Z-direction

4.75 pixel @X-direction

State-of-the-Art Comparison

	ISSCC-2017(Sony)	JSSC-2019	ISSCC-2021(Sony)	Ours
Process	90 nm 1P4M/ 40nm 1P7M	130nm 1P6M/ 130nm 1P6M	65nm/ 22nm	180nm 1P8M
Integration	Stacked BSI	Stacked BSI	Stacked BSI	Single chip, FI
Photoreceptor	PD	PD	PD	SPAD
Resolution	1296×976	1024×769	4056×3040	128×128
Dynamic range	80dB	54dB	-	100dB
Frame rate	60fps	340fps	120fps	100000 fps
Temporal resolution	16.7ms	2.9ms	8.3ms	10μs
Vision mode	2D vision	2D vision	2D vision	2D vision Dim vision 3D depth vision
Processor architecture	PE array	PE array	ISP + CNN DSP	PE array + MPU
Parallelism	1034	3072	2304	1024
Bit-width	1/4	8	8/16/32	1/8/16
On-chip memory	168KB	171KB	9 MB	256KB
CV	Spatial filtering Morphology	Spatial filtering Motion detection	Signal processing	Temporal filtering
NNs	N/A	N/A	CNN	SNN
Light-adaptation	No	No	No	Yes (3.85μs)
Clock frequency	108MHz	80MHz	262.5MHz	80MHz
Peak performance	140GOPS@4bit	61GOPS@8bit	1210GOPS@8bit	81.9GOPS@1bit

Summary

■A Vision Chip is proposed:

- **Full spiking visual flow based on SPAD imaging and spike-based computing**
- **Configurable gated SPAD image sensor**
- **Reconfigurable spike-based vision processor**

■Measurement results show:

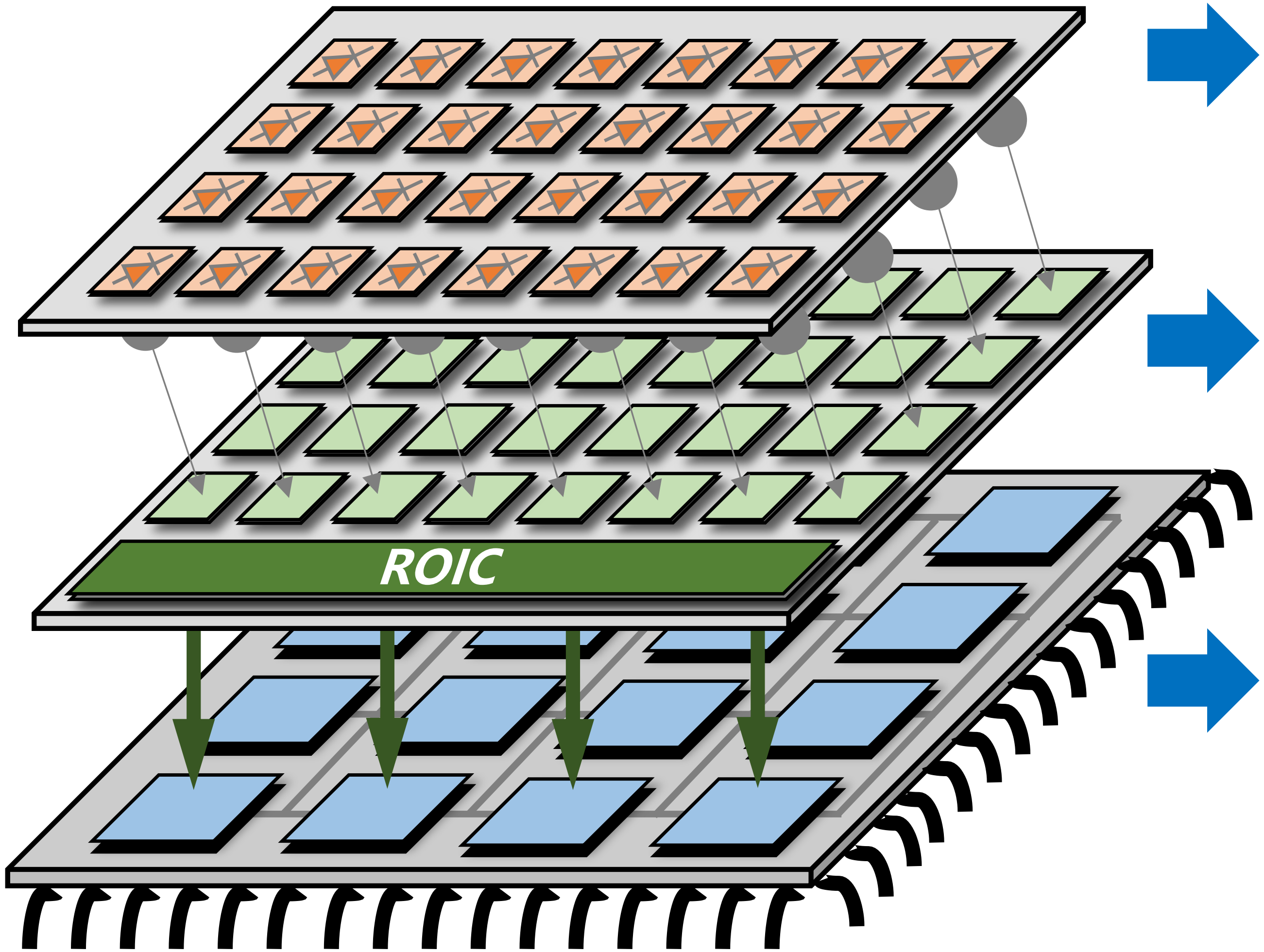
- **Versatile visual capabilities (2D/3D/dim vision)**
- **99.3% Acc. and 300 infer/s @ MNIST classification**
- **Merely 3.9% Acc. loss @ 20 mlx**
- **1.68 cm obstacle detection error**
- **3.85 μs self-adaptation for ambient light changes**

Outline

- Background
- Chip Architecture
- Key Techniques
- Results and Comparison
- **Discussion**

Discussion

Future vision chip



SPAD device array
→ High resolution
→ High fill factor

Low-level in-pixel processing circuits
→ Denoise preprocessing
→ Extract ROI

High-level intelligent processor
→ AI-based signal enhancement
→ More bio-inspired mechanisms

Thanks!

Q&A