NEUROMORPHIC EVENT-BASED VISION

TECHNOLOGY AND APPLICATIONS

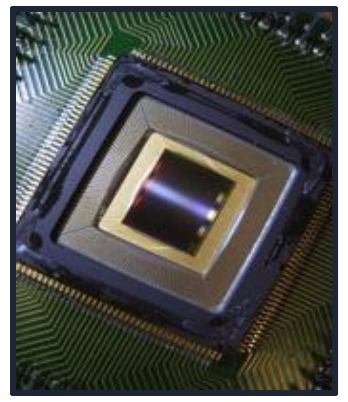
Christoph Posch, Co-founder, CTO,

PROPHESEE

THE HISTORY OF PROPHESEE

2010-2011

FIRST ATIS-1 SENSOR



FIRST PRODUCT

Pixium vision

2013-2015

$5M FUNDRAISING

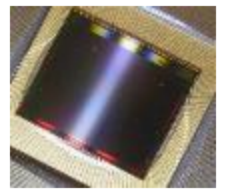
Chronocam

20+ patents in HW & SW

BOSCH Invented for life

SUPER NOVA INVEST

LAUNCH GEN 1 30µm QVGA



2016

$15M FUNDRAISING

BOSCH Invented for life

intel Capital

RENAULT NISSAN

360 CAPITAL PARTNERS

IBIONEXT

SUPER NOVA INVEST

2017

TECHNOLOGY PIONEER

WORLD ECONOMIC FORUM

TOP 100 AI STARTUPS

CB INSIGHTS

COOL VENDOR

Gartner

TOP UP & COMING IMAGE SENSOR COMPANY

IS EUROPE AWARDS

Collaboration DARPA

LAUNCH GEN 2 15µm HVGA

2018

$19M FUNDRAISING

PROPHESEE

45+ patents in HW & SW

TECHNOLOGY INNOVATION AWARD

ISSCC

MIT Technology Review

Vision Systems 2018 Innovators Awards

AutoSens Awards

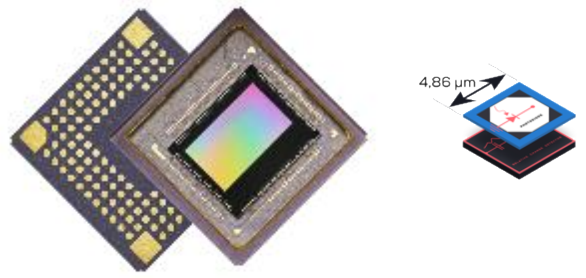
ASPENCORE EETimes

2019-2022

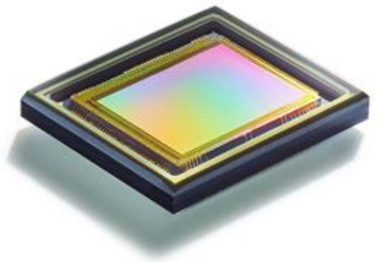
$61M FUNDRAISING

PROPHESEE | SONY SEMICONDUCTOR SOLUTIONS

LAUNCH IMX636ES 4.86µm STACKED HD SENSOR



LAUNCH GEN 3 15µm VGA PACKAGED



LAUNCH METAVISION INTELLIGENCE SOFTWARE

METAVISION INTELLIGENCE SUITE

FIRST INDUSTRIAL EMBEDDED SYSTEM

IMAGO

FIRST INDUSTRIAL USB SYSTEM

CenturyArks

ABOUT US

PROPHESEE

KEY FIGURES

2010
FIRST PRODUCT

Pixium vision

45+
PATENTS
SENSOR
SYSTEM
ALGORITHMS
APPLICATIONS

€127M
RAISED

IBIONEXT · Banque européenne d'investissement · Mi · BOSCH · intel · SINOVATION VENTURES · Prosperity7 · RENAULT NISSAN

53
INTERNATIONAL RECOGNITIONS

VISION AWARD 2021 Winner · IS Europe Awards · Tiny ML · MIT Technology Review INNOVATORS UNDER 35 EUROPE · Gartner COOL VENDOR · ISSCC · WORLD ECONOMIC FORUM

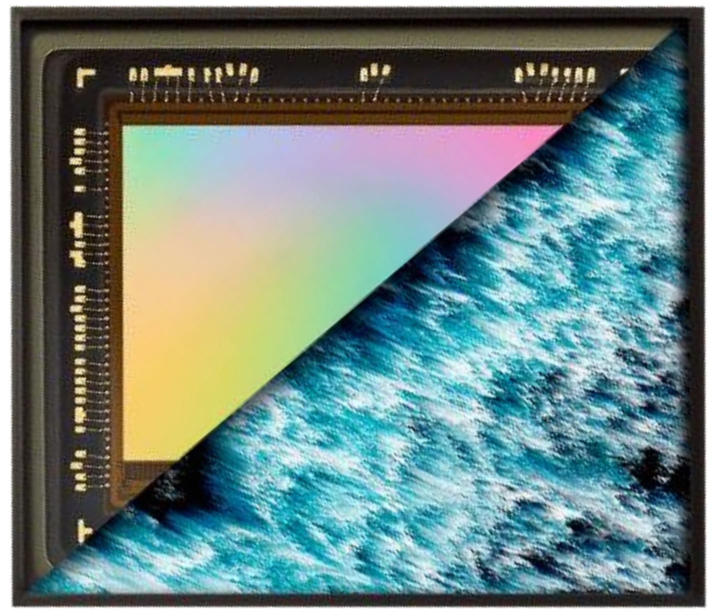
TEAM

110+
STRONG

5
OFFICES

PRODUCTS

METAVISION®
SENSORS



METAVISION®
INTELLIGENCE
SOFTWARE

DEVELOPMENT TOOLS

ECOSYSTEM

SONY · intel · Tower Semiconductor · SINOVATION VENTURES · RENAULT NISSAN · BOSCH Robert Bosch Venture Capital GmbH · XPERI · Mi · FRAMOS · OMS · iCatch · DMP · IMAGO TECHNOLOGIES · LUCID VISION LABS · CenturyArks · SynSense · cambridge consultants · terranet · Fraunhofer · cnrs · SORBONNE UNIVERSITÉ · université PARIS-SACLAY · Penn · University of Oxford · MIT Massachusetts Institute of Technology · ROBOTICS & PERCEPTION GROUP · ...

"NEUROMORPHIC" ENGINEERING (?)

CAN'T BEAT A FLY

BODY WEIGHT
< 1 GRAM

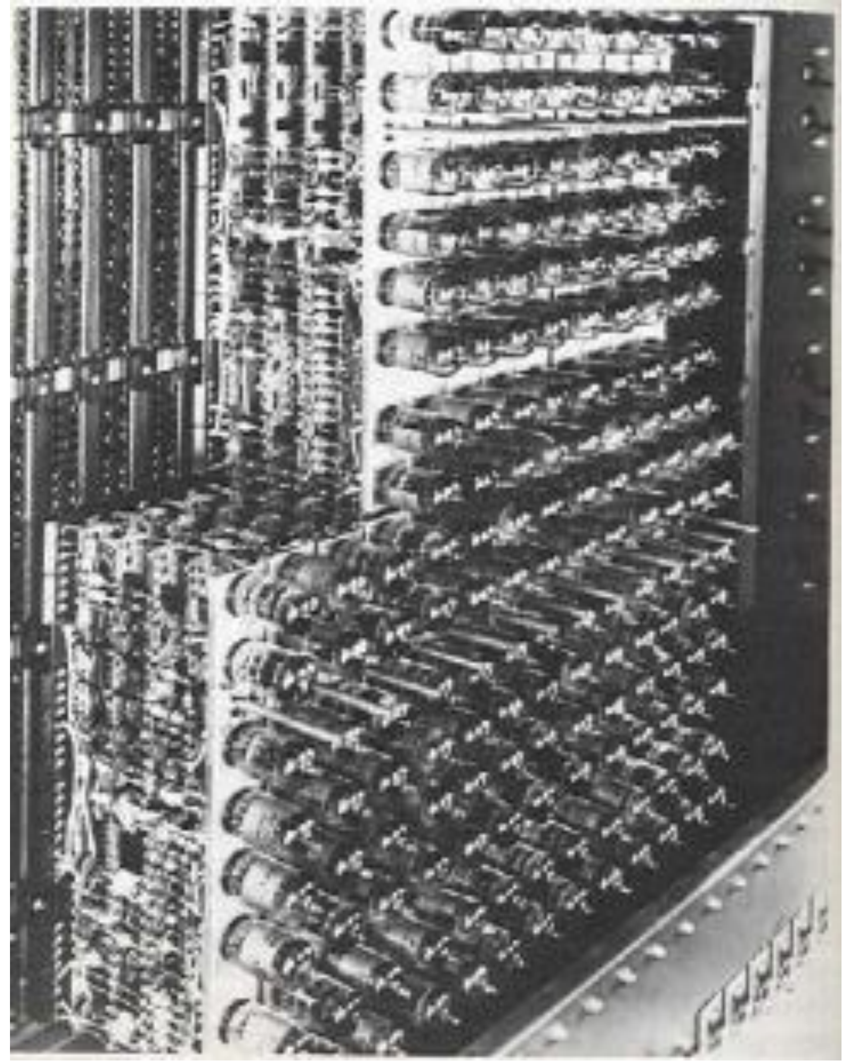
BRAIN WEIGHT
FEW MICROGRAMS

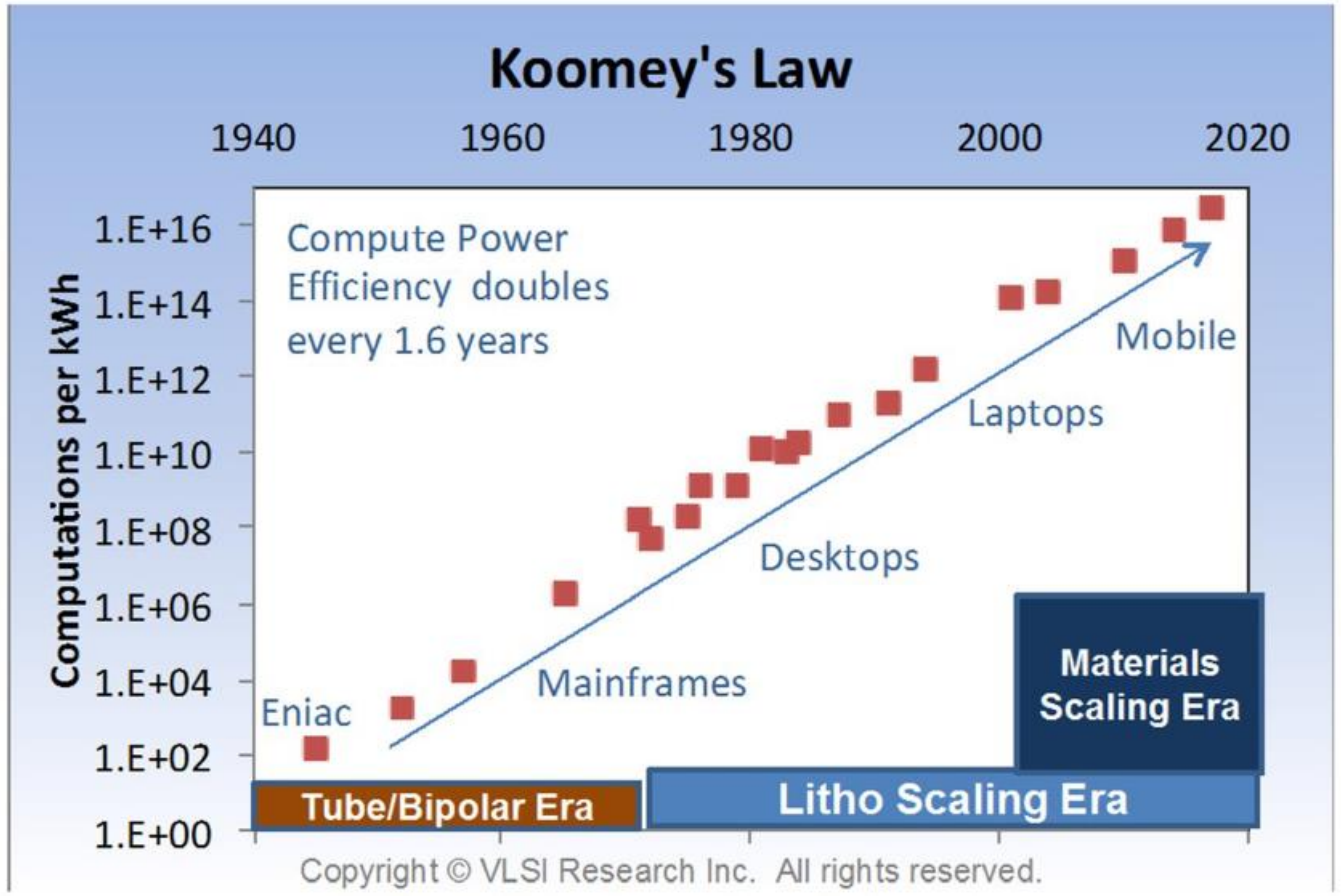
DISSIPATING POWER
~ 10μW

COMPUTING – ENERGY EFFICIENCY

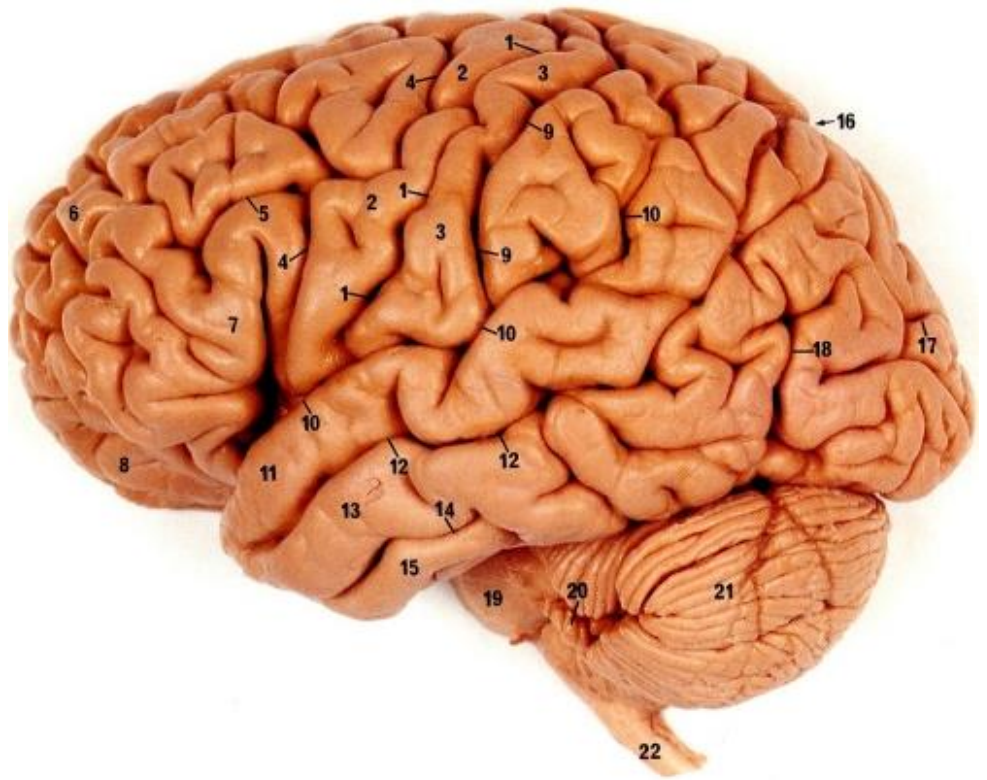
Progress of electronic information processing over past 80 years:

- from **5 Joules / operation** (vacuum tube computer, 1940s)
- to **500pJ / op** (IBM Blue Gene, 2010)
- to **5pJ / op** (GPUs, 2020)
- → 1,000,000,000,000 times better!





BUT THE HUMAN BRAIN IS STILL 1,000,000X MORE ENERGY-EFFICIENT



- Massive parallelism (10^{11} neurons)
- Massive connectivity (10^{15} synapses)
- Low-speed components (~1 – 100 Hz)

  >10^{16} complex operations / second →

	„K computer“ (RIKEN, Japan)
10 PetaFLOPS	**10 PetaFLOPS**
10 W	**10 MW**
1 PFLOP/W	**1 GFLOP/W**
1.5 kg	100s of tons

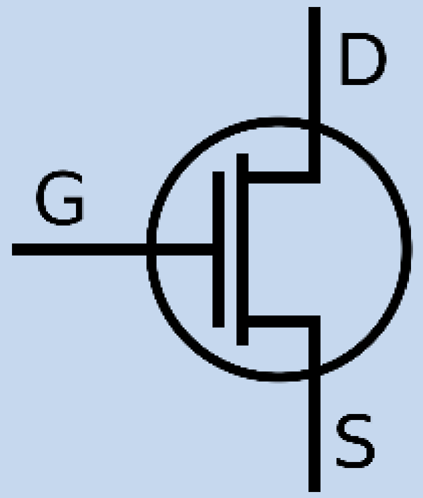
Energy efficiencies

- Computer system level: 10^{-9} J/operation
- Chip: 10^{-12} J/operation
- Brain: 10^{-15}J/operation

WHERE DOES THE ENERGY GO ?

10^{-15}J

Transistor
Activation

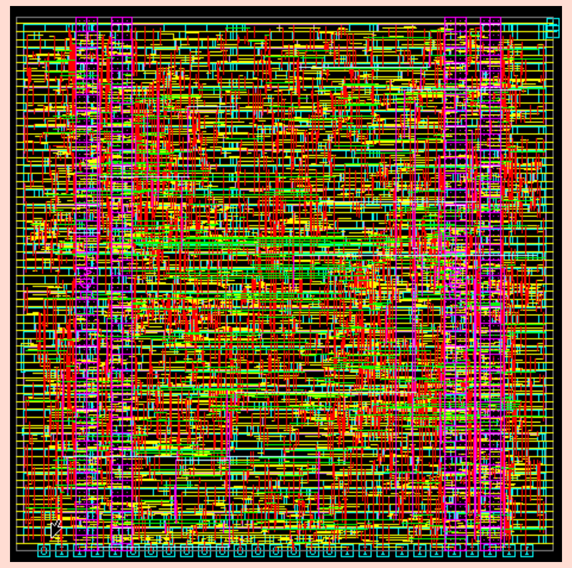


=

Synapse
Activation

-x100

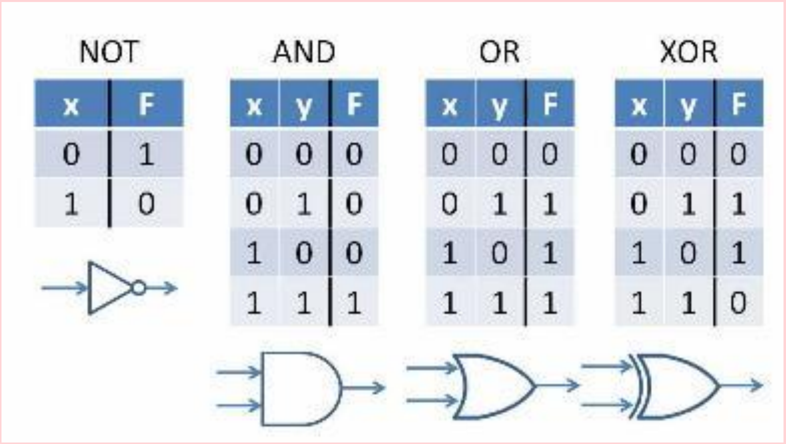
Energy wasted
in wire charging



Processing and
storage is local

-x10000

Binary Encoding
Booleans logic
operations



NOT

x	F
0	1
1	0

AND

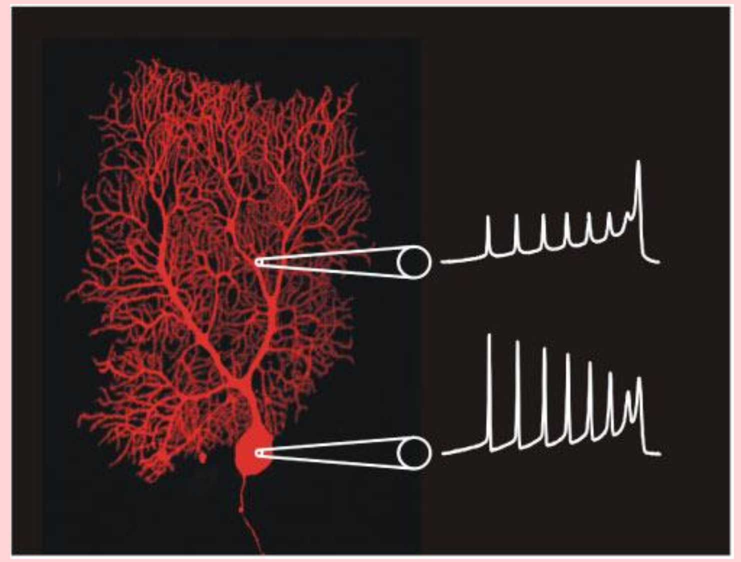
x	y	F
0	0	0
0	1	0
1	0	0
1	1	1

OR

x	y	F
0	0	0
0	1	1
1	0	1
1	1	1

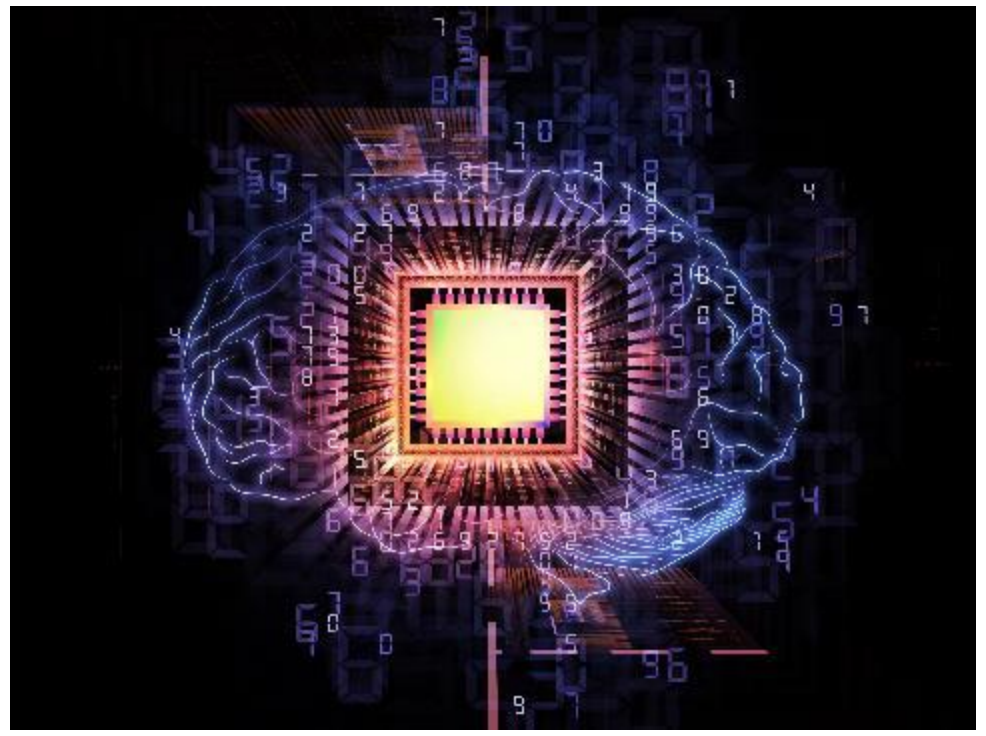
XOR

x	y	F
0	0	0
0	1	1
1	0	1
1	1	0

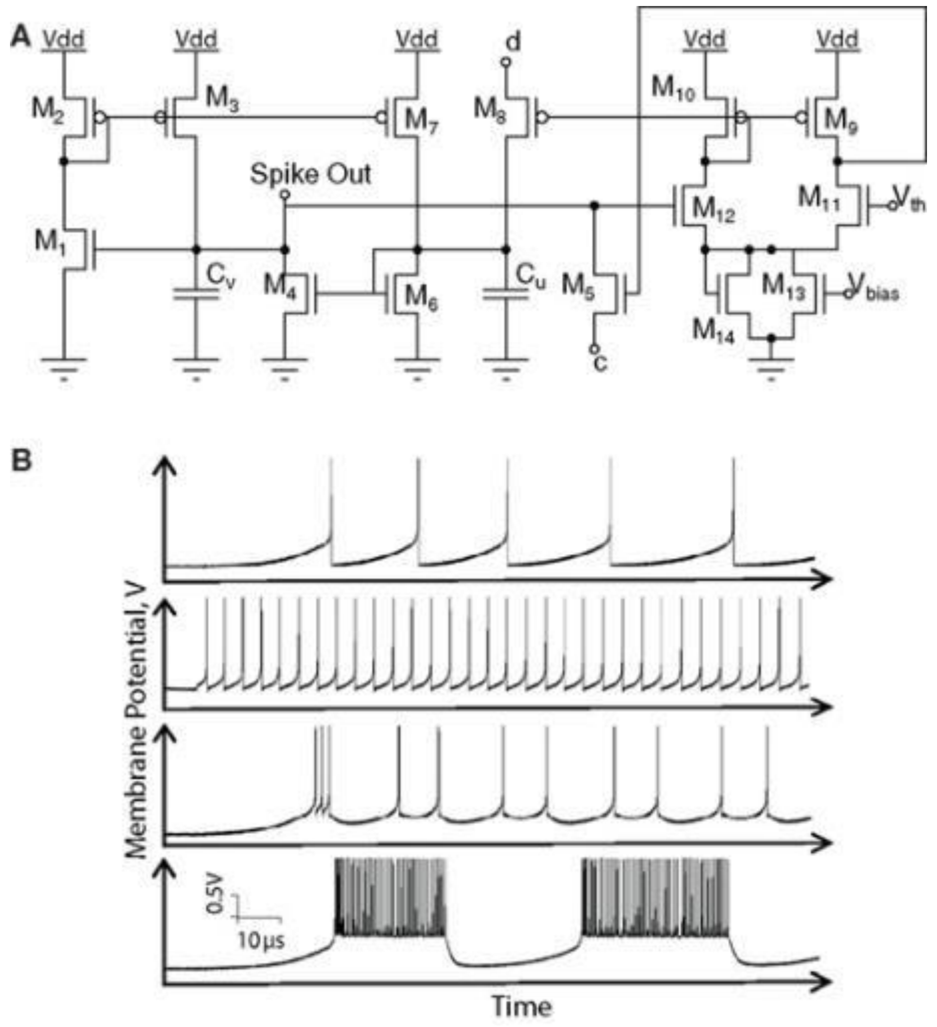


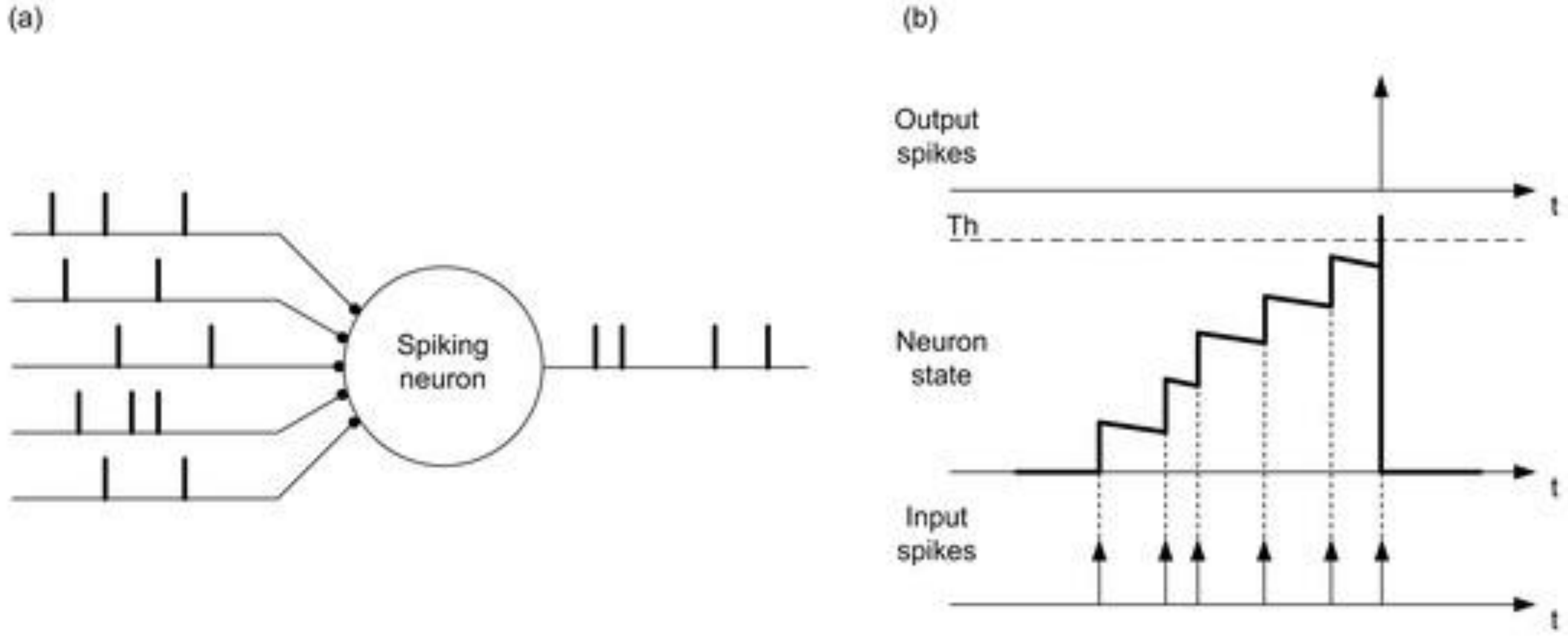
Spike-based analog
computing

„NEUROMORPHIC ENGINEERING"



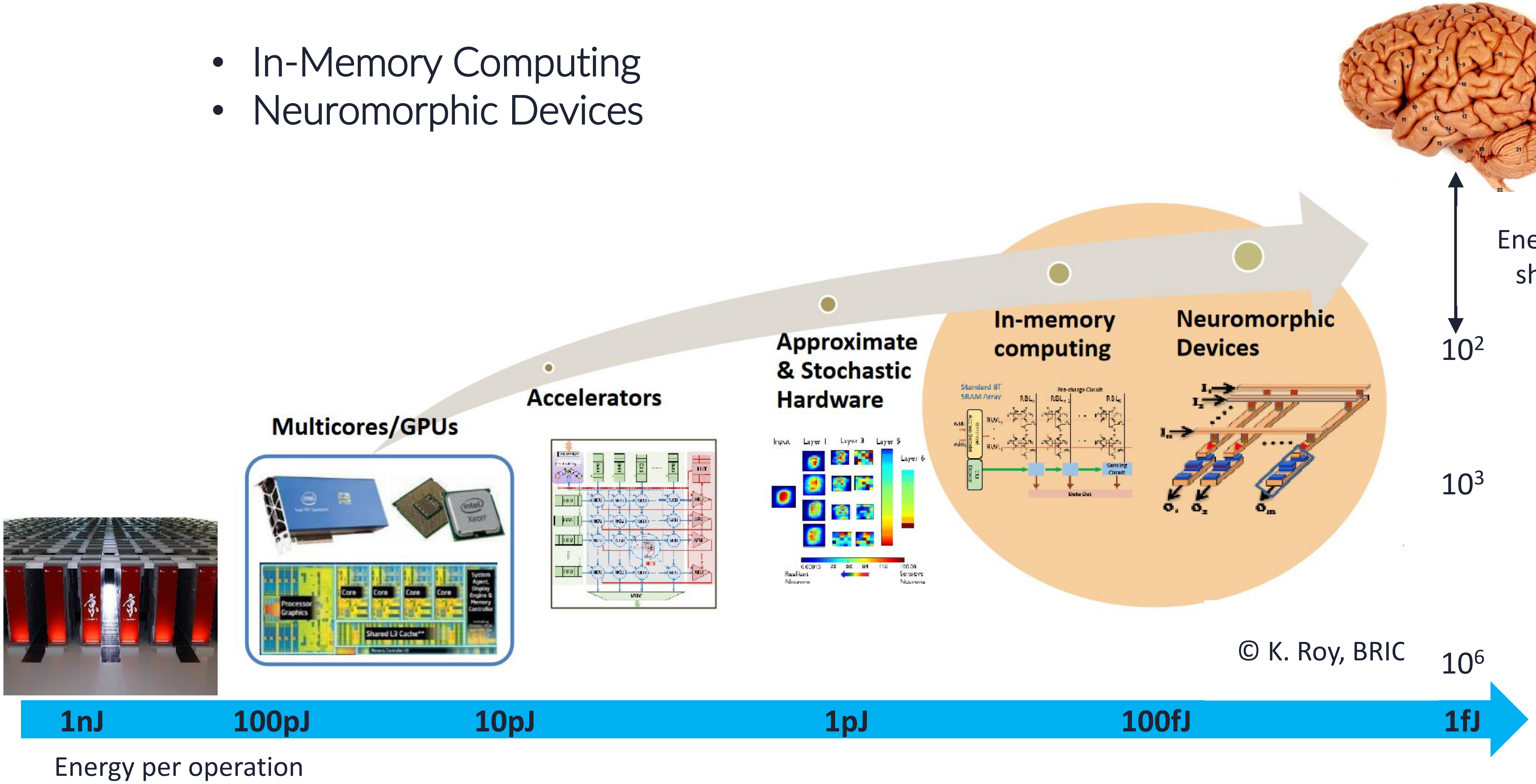
- C. Mead (CalTech, 1980`s – 90`s):
  "Neuromorphic Electronic Systems", Proc. IEEE
- Silicon VLSI technology can be used to **build circuits** that mimic **neural functions**
- Silicon primitive: **transistor** – functional **similarities** to **neurons**
- Building blocks: **neurons, axons, ganglions, photoreceptors**, …
- Biological **computational primitives**: logarithmic functions, excitation/inhibition, thresholding, winner-take-all selection …
- Encoding information in the form of "**spikes**"





NEUROMORPHIC BRAIN-INSPIRED COMPUTING

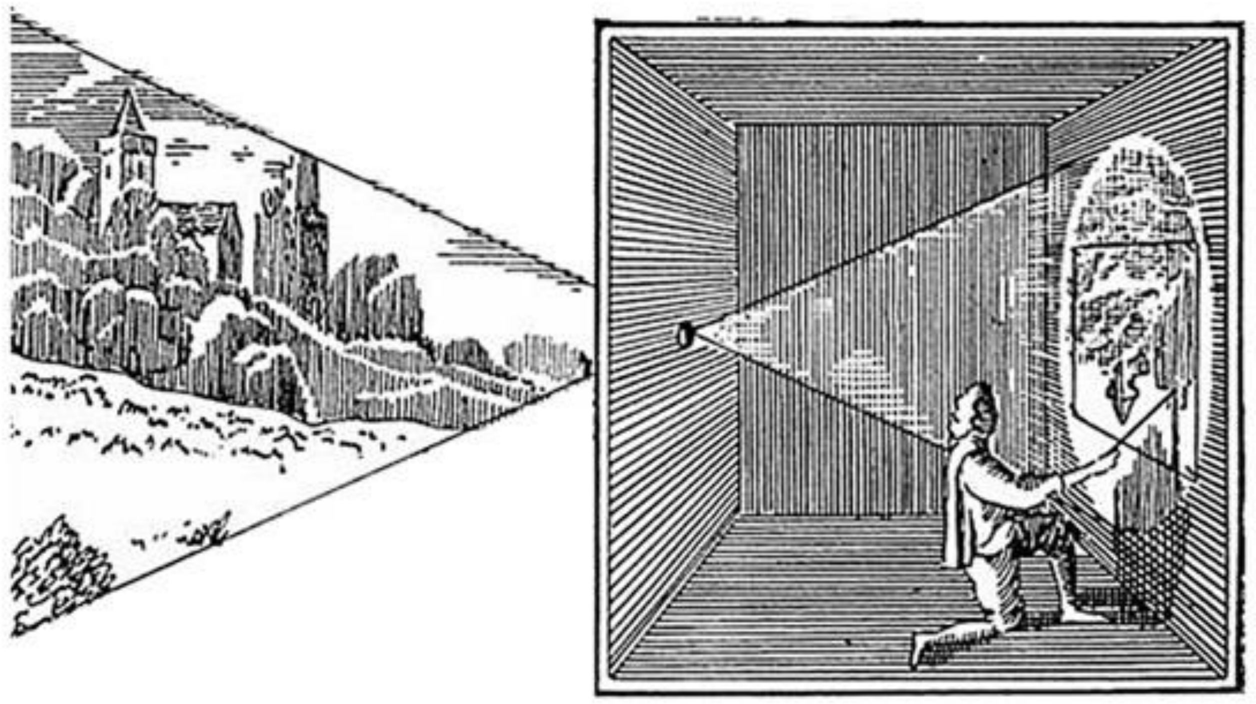
- In-Memory Computing
- Neuromorphic Devices



THE NEUROMORPHIC APPROACH TO ARTIFICIAL VISION

HUMANS CAPTURE "IMAGES" OF THE WORLD

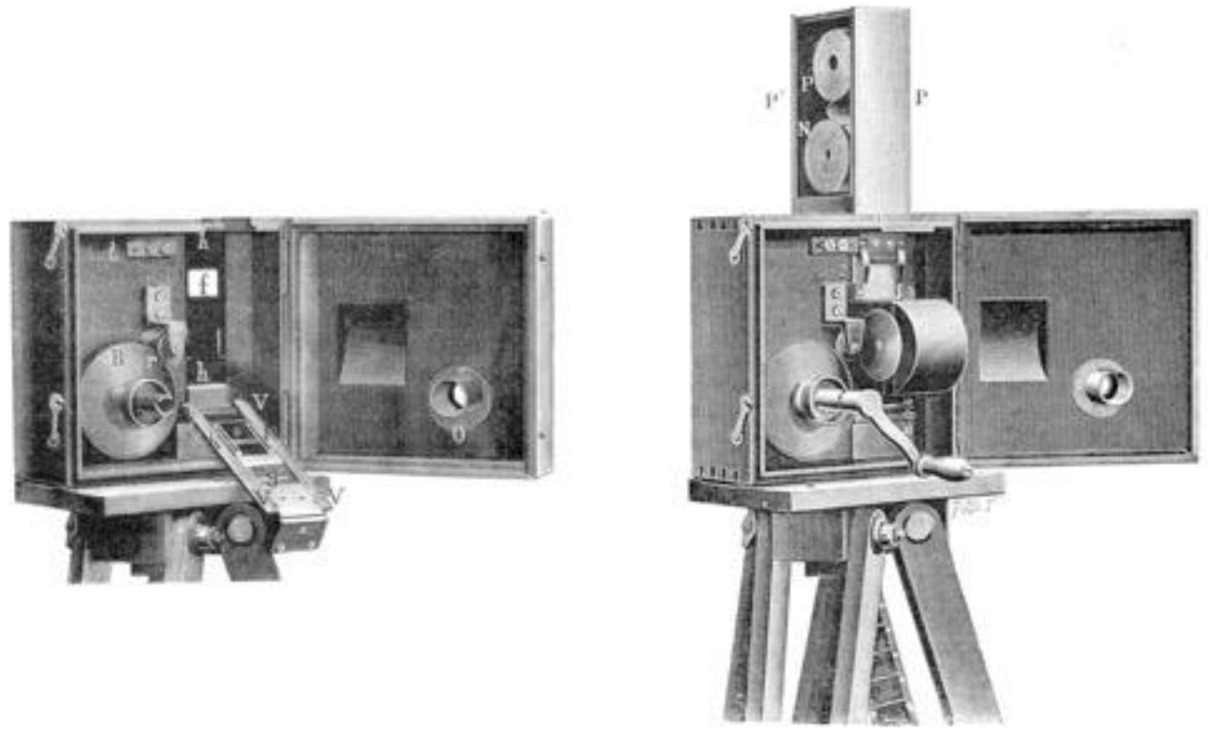
CAMERA OBSCURA



DIGITAL CAMERA

SMART PHONE

SCUOLA DI ATENE - RAFFAELLO



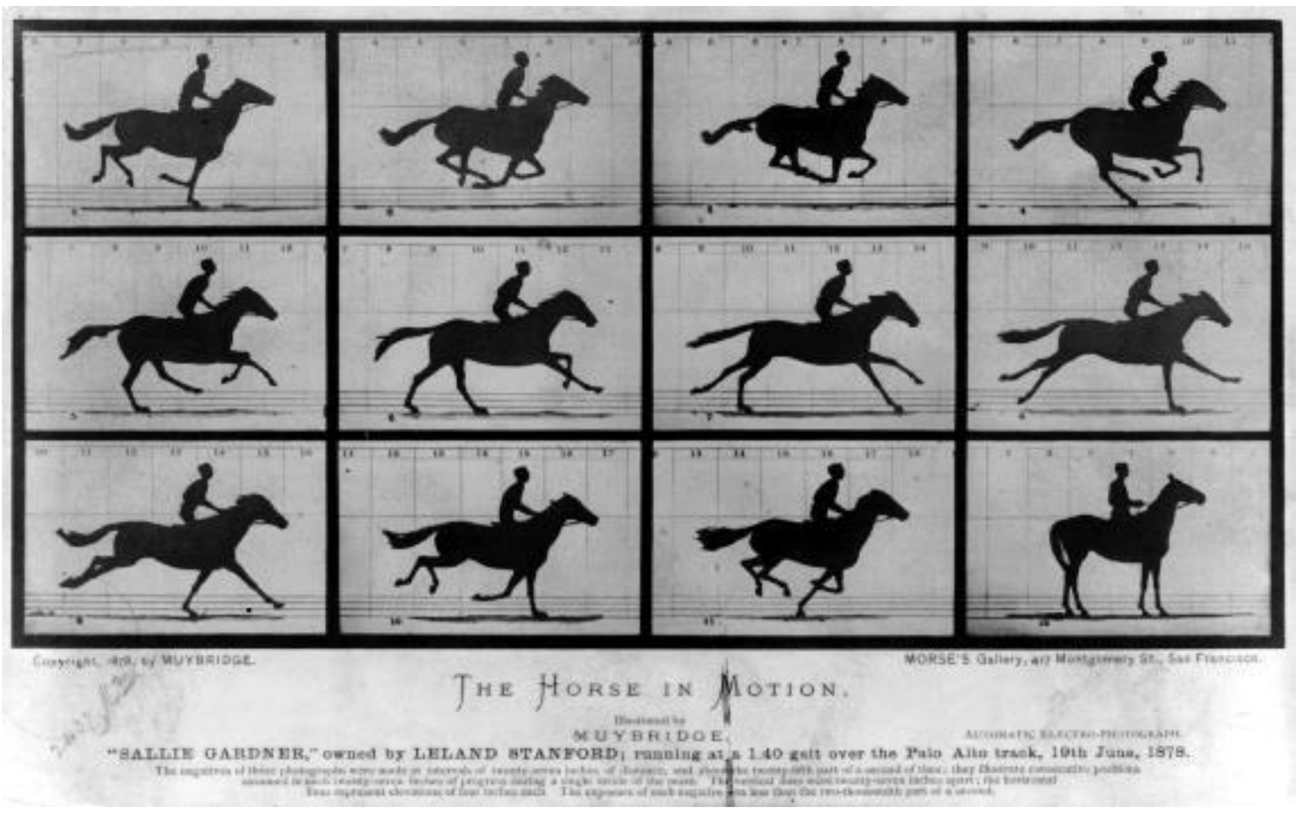
FRÈRES LUMIÈRE

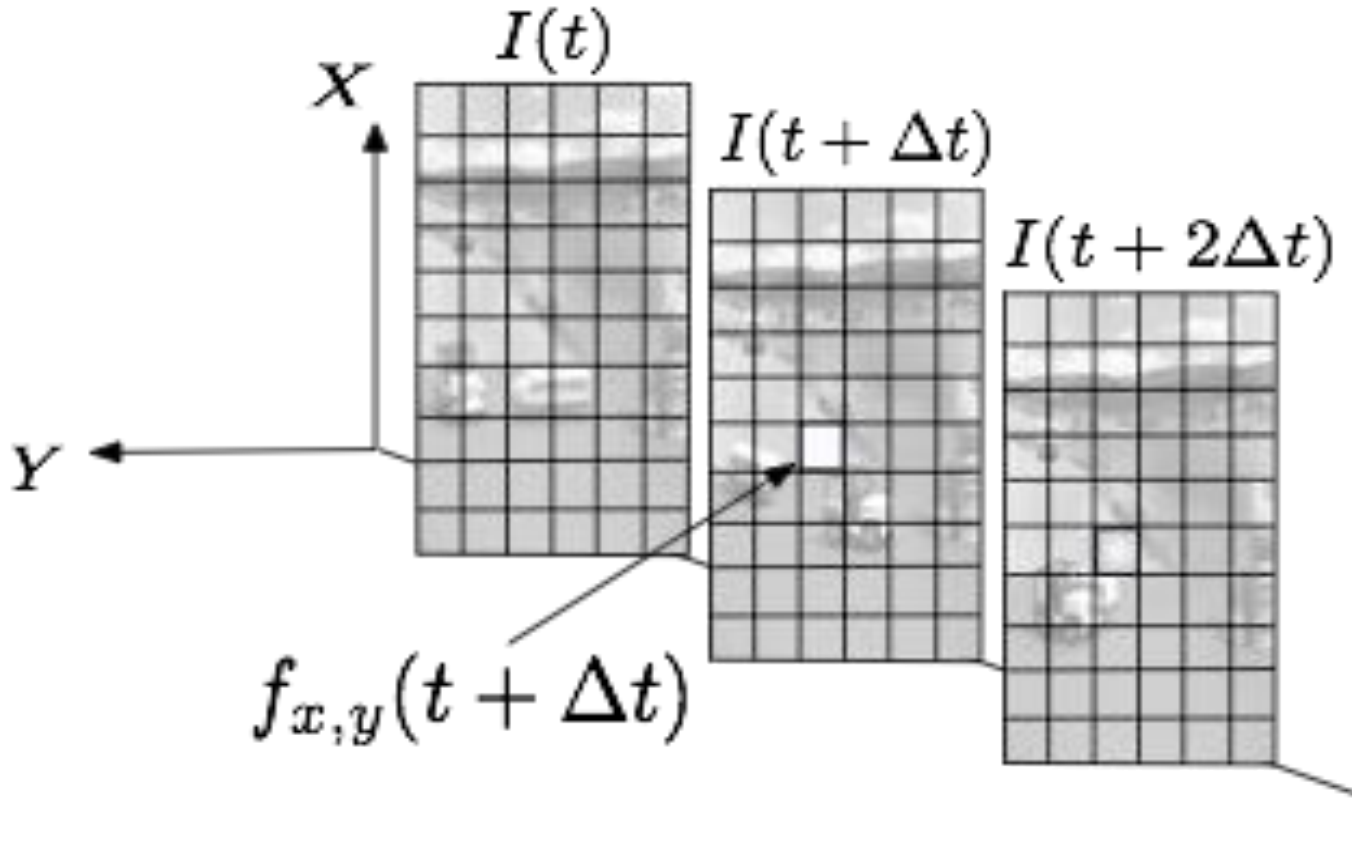
LIMITATIONS OF IMAGE SENSING IN COMPUTER VISION


- IMAGE SENSORS TAKE IMAGES (PHOTOGRAPHS)
  - great for human consumption
  - a snapshot of a scene at one point in time → static



- COMPUTER VISION / ANALYZE DYNAMIC SCENES
  - Scene dynamics (changes / motion) carry relevant information
  - i.e. object recognition, tracking, motion flow, …
- USING STATIC IMAGES TO EXTRACT DYNAMIC INFORMATION ?
  - Acquire a series of frames
  - This work-around became the universally accepted paradigm of acquiring data from dynamic scenes





Fixed Δt → "frame rate"

FRAMES AND DYNAMIC SCENES

WHAT'S THE PROBLEM WITH FRAMES?

- All pixels acquire entire scene at a same fixed rate
- But scene dynamics are different in **different parts of a scene**
- Any chosen frame rate is wrong

UNDER-SAMPLING

- Motion blur
- Displacement

OVER-SAMPLING

- Redundant, useless data
- Known from previous acquisition
- Need to acquire, transmit, store, process, …

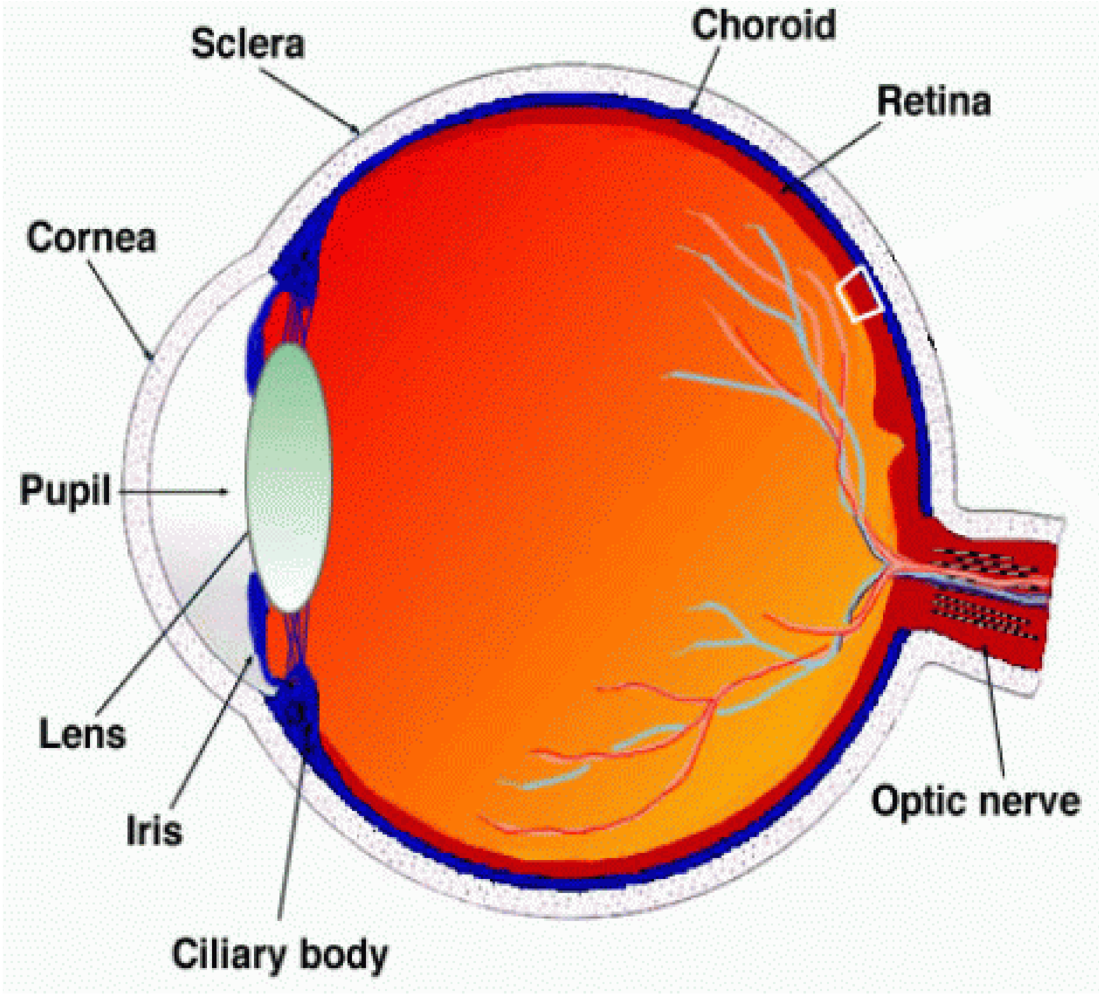
WITH DYNAMIC SCENES, THE CONVENTIONAL PARADIGM OF VISUAL ACQUISITION IS FUNDAMENTALLY FLAWED

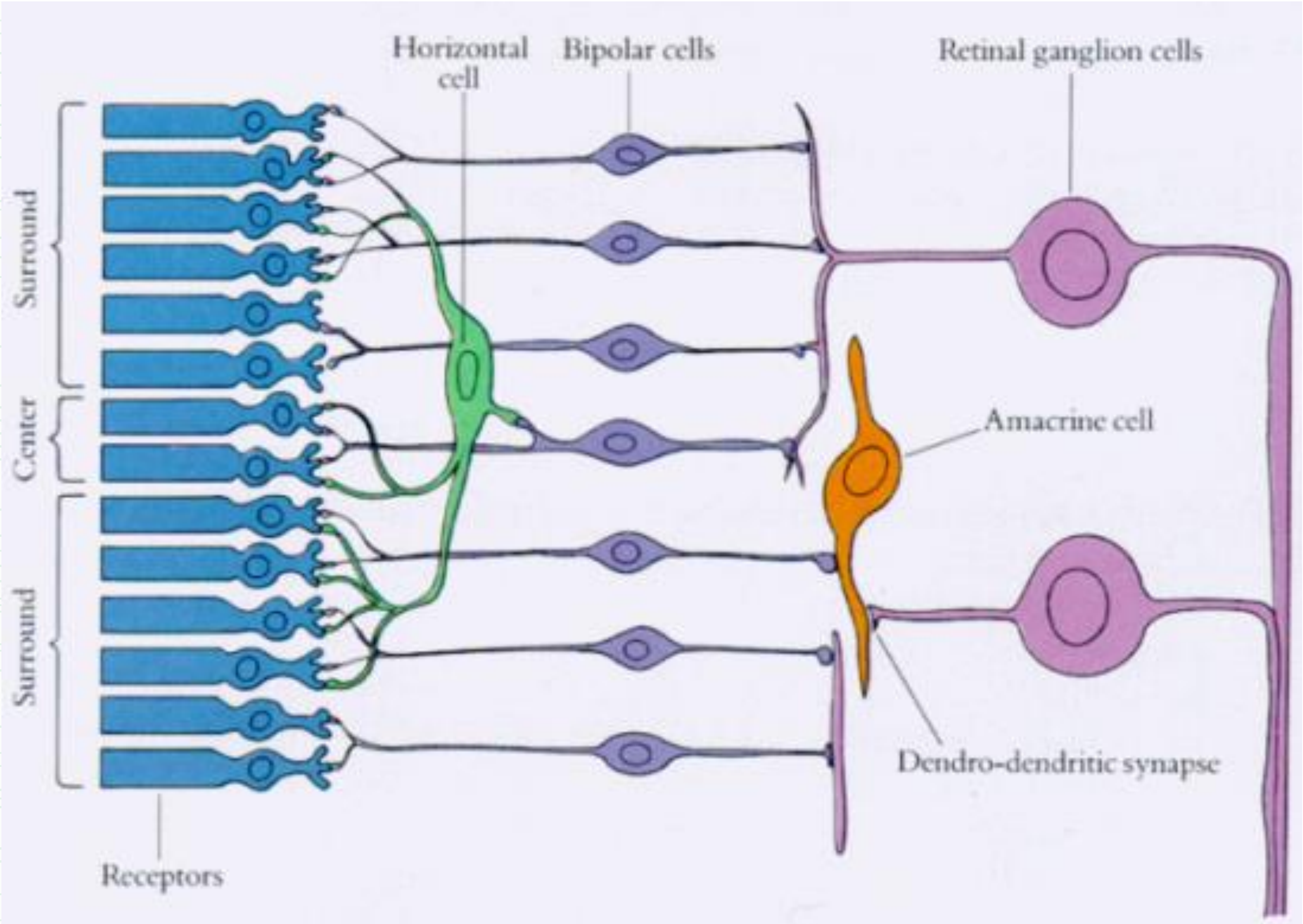
COMPUTER VISION: INSPIRED BY BIOLOGY

Biology is leading the way to a more efficient way of acquiring visual information

- Biological vision does not use **images** to see
- Biologically inspired “**event-based**“ / “**spike-based**” vision uses **autonomous pixels** to capture only **relevant** information from a dynamic scene
    - Temporal Contrast
    - Change Detection
    - Dynamic vision sensing

THE HUMAN RETINA

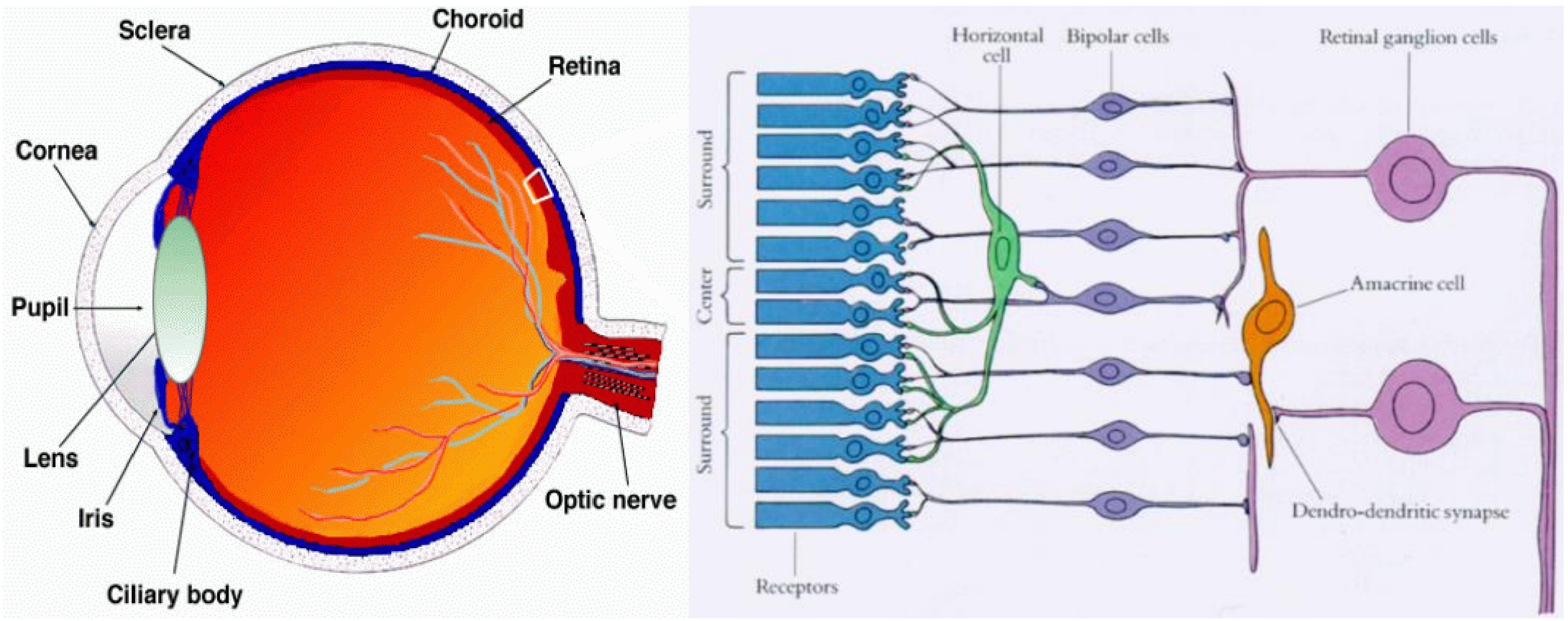




- **135 million** photoreceptors – detection threshold (rod): **1 photon**
- **1 million** ganglion cells in the retina process visual signals received from the photoreceptors.
- **Analog gain control**, spatial and temporal filtering: ~ **36 Gb/s** HDR raw image data is compressed into ~ **20 Mb/s** spiking output to the brain
- Retina encodes **useful** spatial-temporal-spectral **features** from a redundant, wide dynamic range world into a small internal signal range. Multiple “pathways” – **Transient**, Sustained
- Power consumption: **~ 3.5 mW**

NEUROMORPHIC EVENT BASED VISION

HOW TO TO <u>EFFICIENTLY</u> ACQUIRE A DYNAMIC SCENE?



- Mimicking the “transient (Magno-cellular) pathway” of the human visual system
- ➔ Pixel-individual acquisition of scene dynamics

NOT ONE SAMPLING RATE FOR ALL PIXELS (=FRAME RATE) …

- … but many (= as many sampling rates as number of pixels), **and**
- sampling rates can **vary** on the fly and **pixel-individually**

HOW? PUT THE PIXEL IN CONTROL!

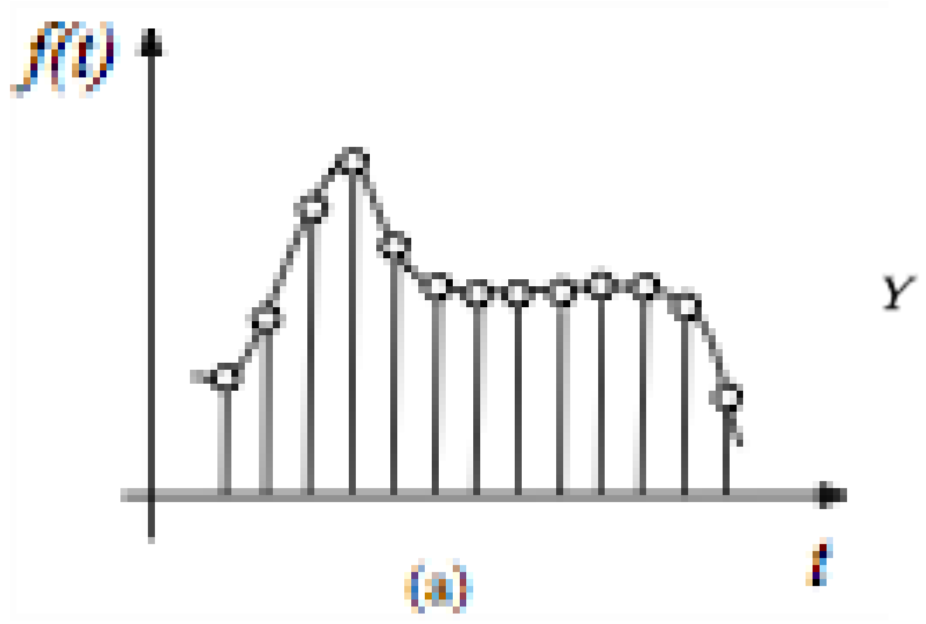
EACH PIXEL INDIVIDUALLY CONTROLS ITS OWN SAMPLING BASED ON THE INPUT SIGNAL

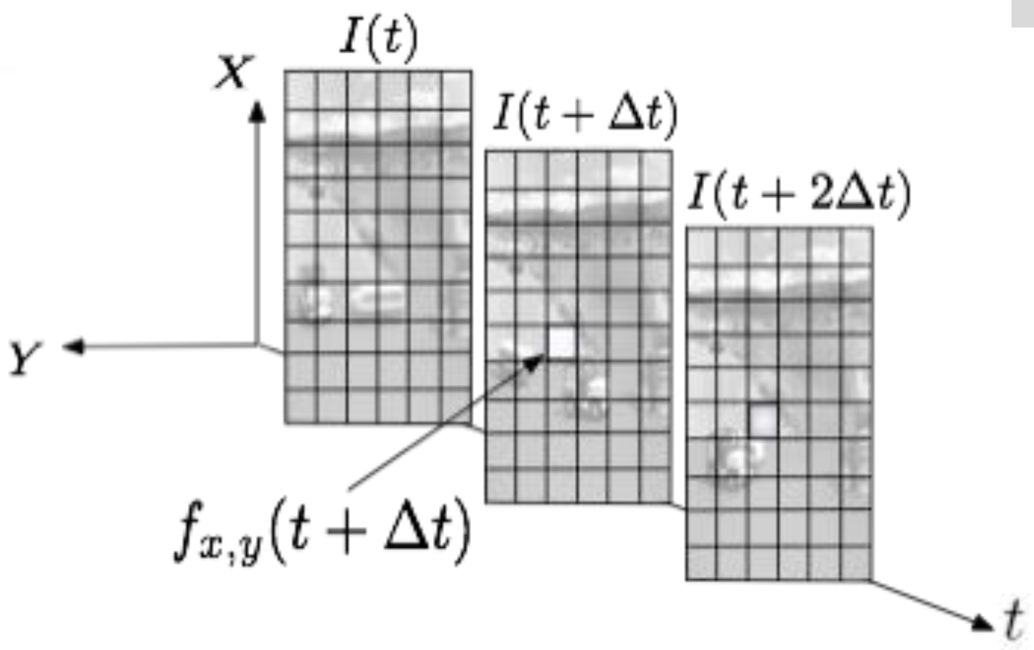
- Change sampling domain (from time to amplitude)
- Pixel does not need any external timing signals – operates autonomously
- ➔ Pixel-wise adaptive non-uniform sampling
  - Encode information in "events"
  - Pixel that is not stimulated visually does not produce output
  - Complete suppression of temporal data redundancy

PIXEL-INDIVIDUAL EVENT SENSING

from STANDARD IMAGER …

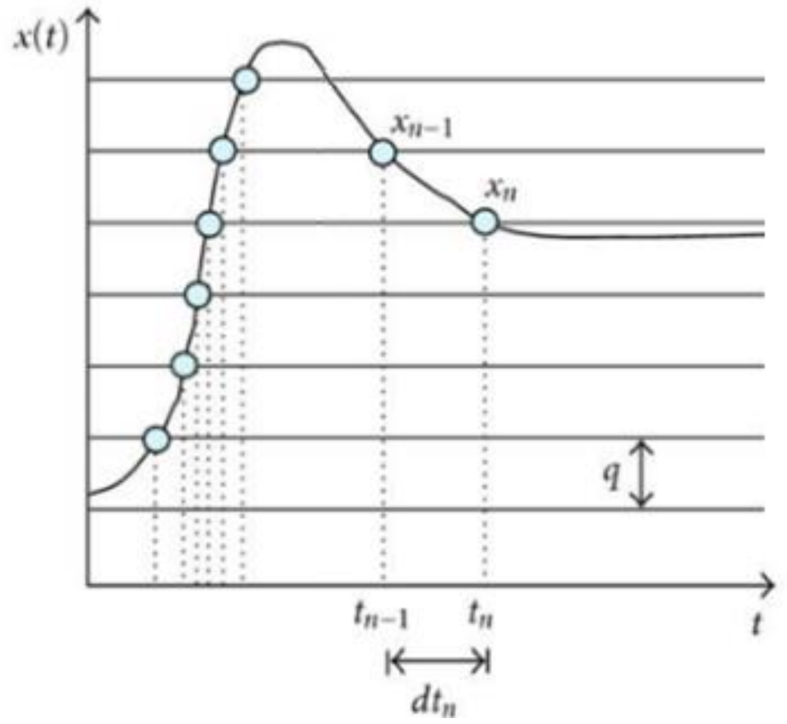
GLOBALLY CONTROLLED FRAME SAMPLING
- EXTERNALLY CONTROLLED SAMPLING PROCESS
- FIXED SAMPLING RATE
- TIME-DOMAIN

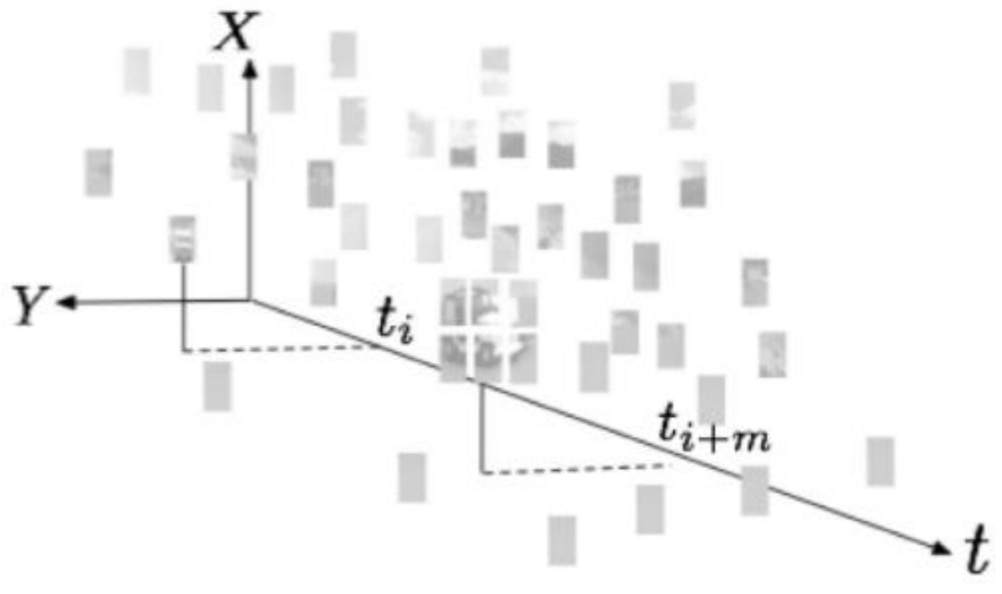


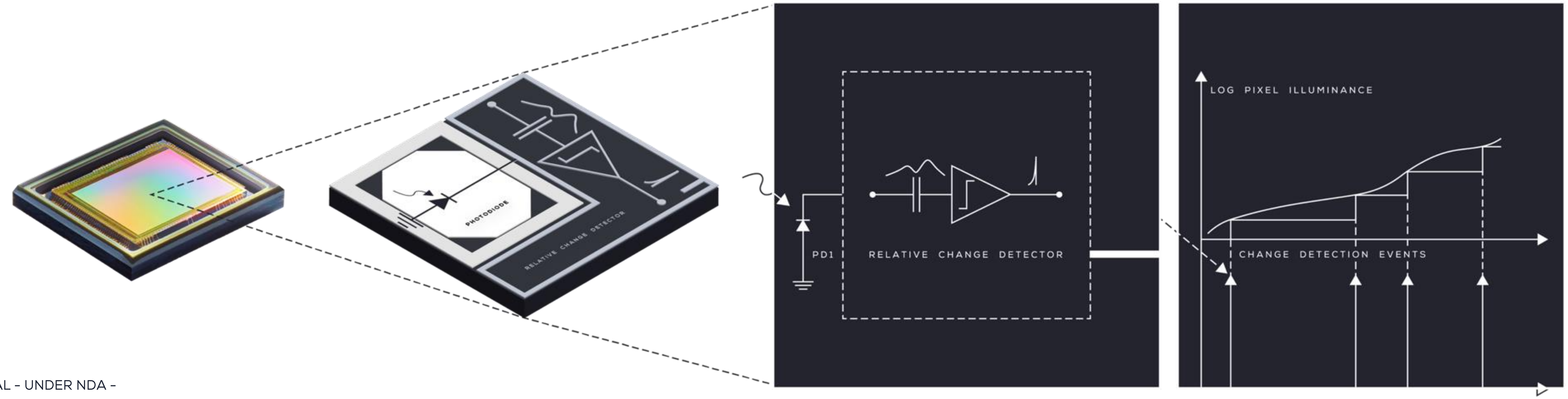


…to NEUROMORPHIC EVENT SENSOR

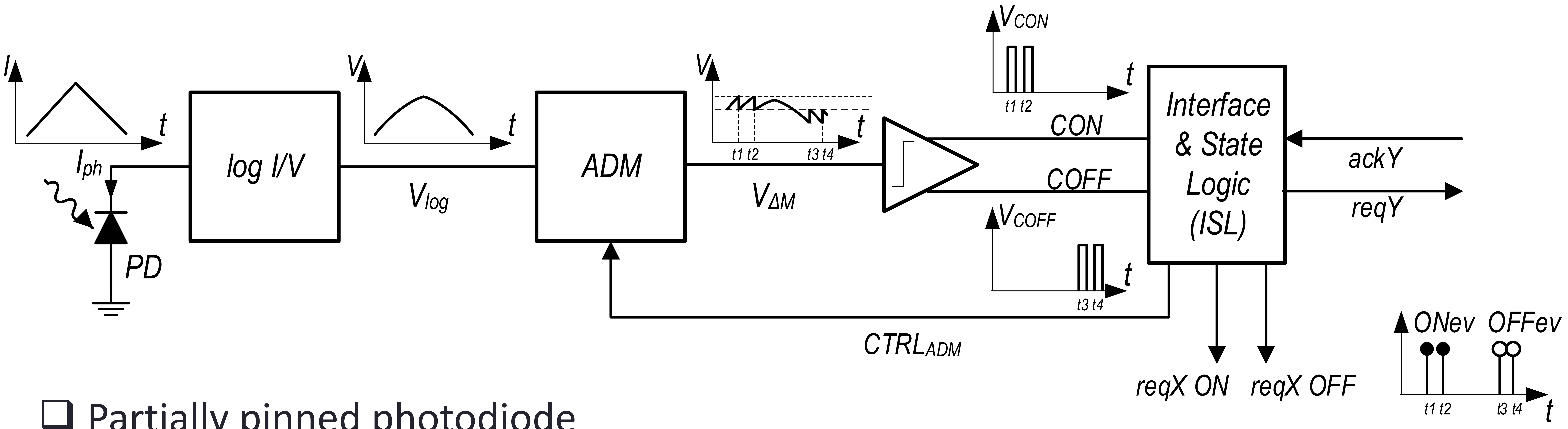
PIXEL-CONTROLLED LOCAL SAMPLING
- PIXELS CONTROL THEIR OWN SAMPLING PROCESS INDIVIDUALLY
- ADAPTIVE NON-UNIFORM SAMPLING RATE
- AMPLITUDE DOMAIN DIFFERENTIAL SAMPLING





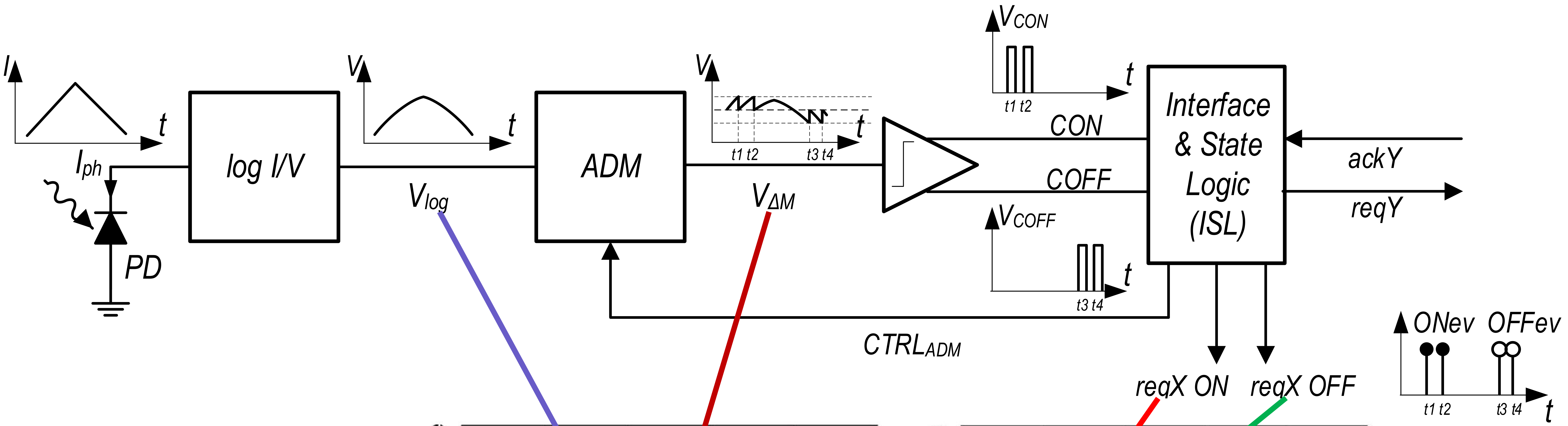


EVENT PIXEL ARCHITECTURE



- ❑ Partially pinned photodiode
- ❑ Subthreshold MOS based logarithmic photocurrent-to-voltage conversion
- ❑ ADM / level-crossing sampler
- ❑ Voltage comparators (for both polarities)
- ❑ Logic with ADM control and interface to the read-out periphery

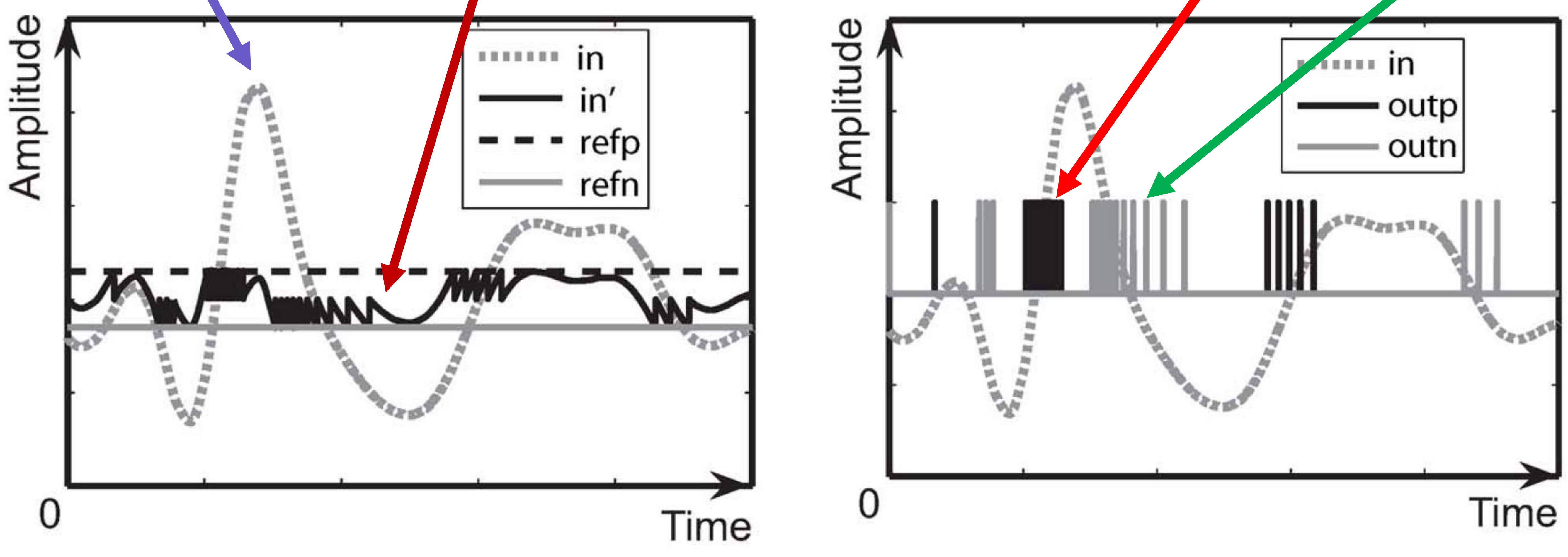
EVENT PIXEL ARCHITECTURE



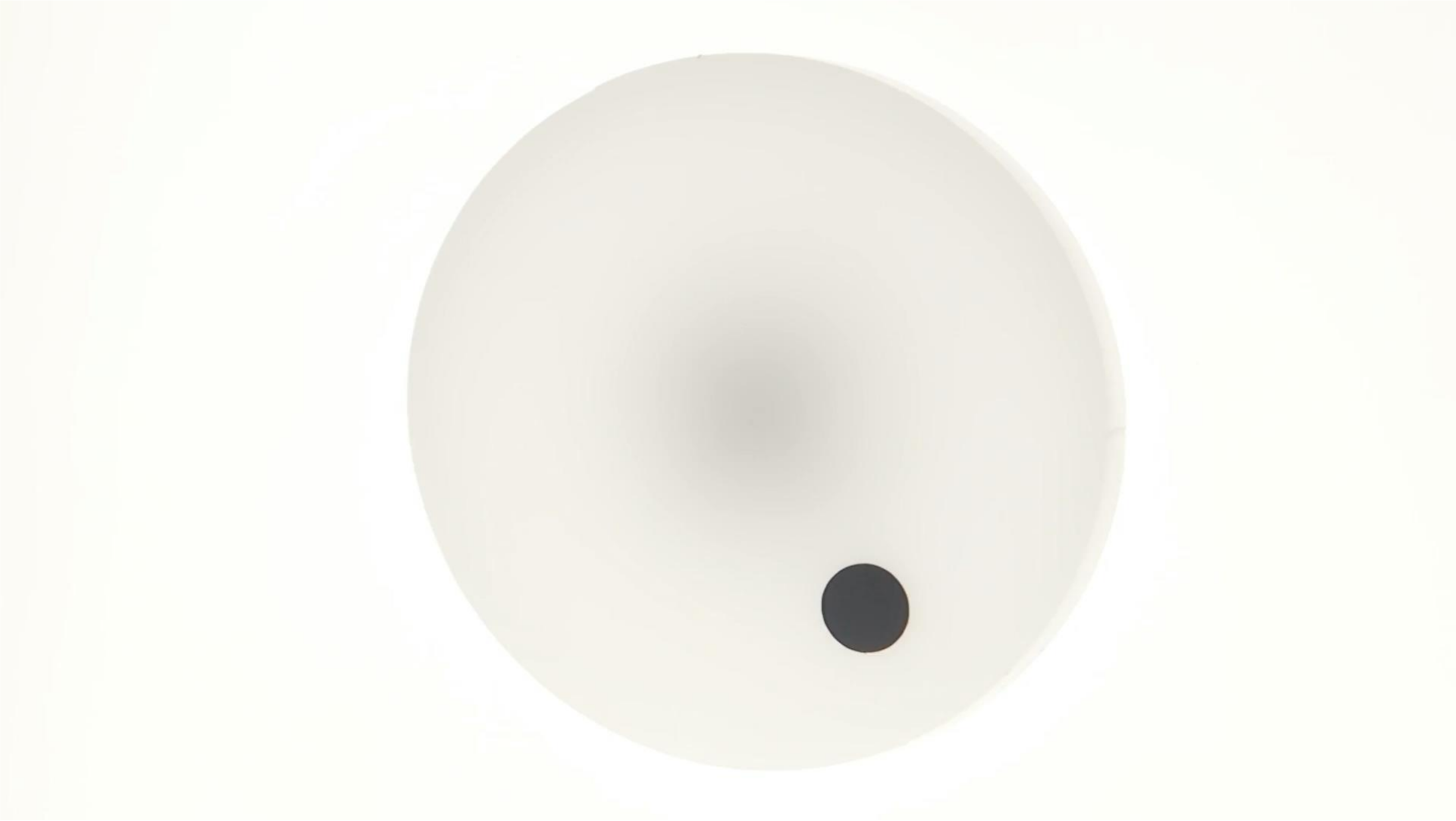
Asynchronous Delta Modulation (ADM) sampling

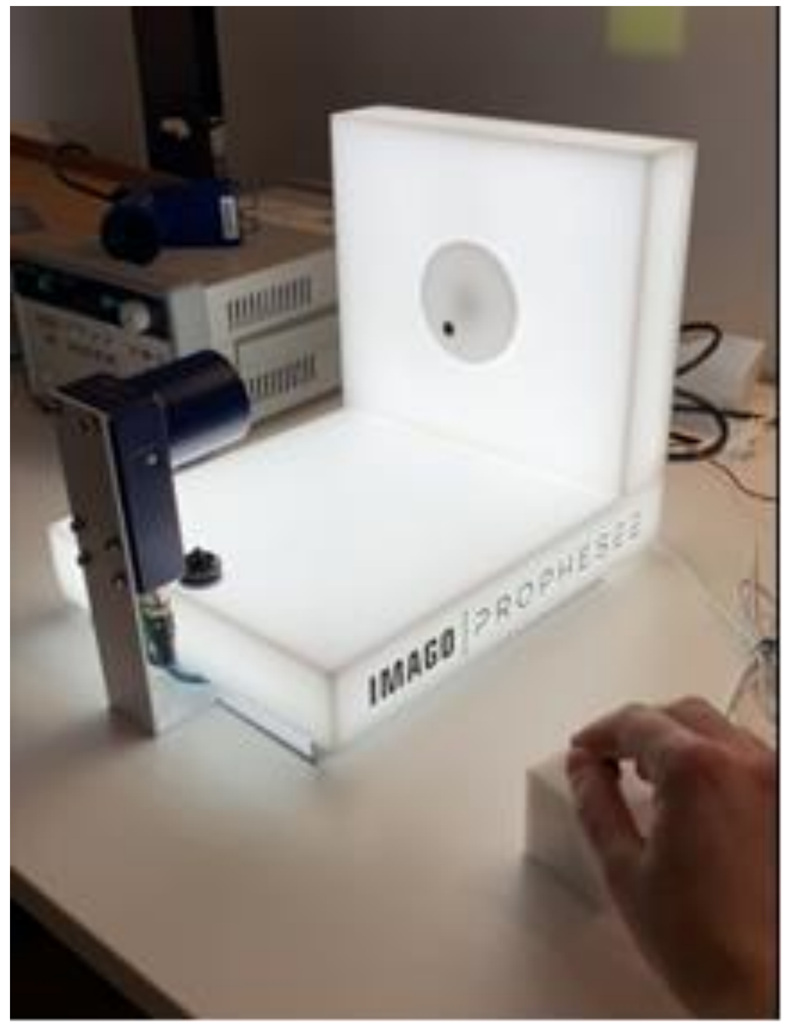
Programmable contrast threshold

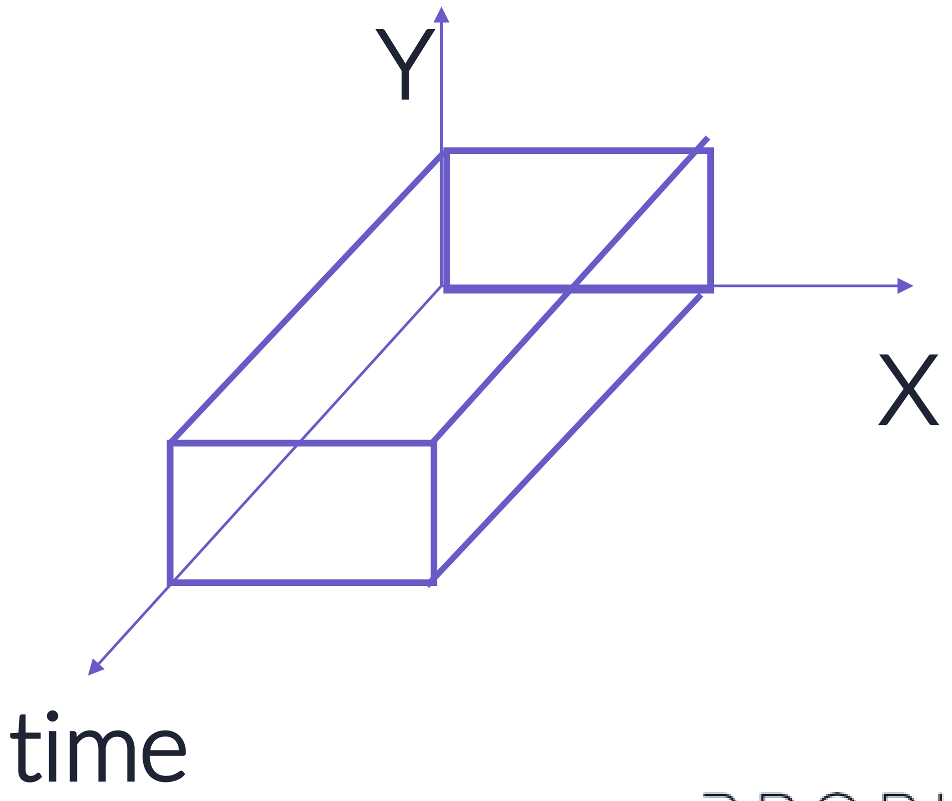
Change polarity captured



SPARSE DATA – HIGH TEMPORAL PRECISION



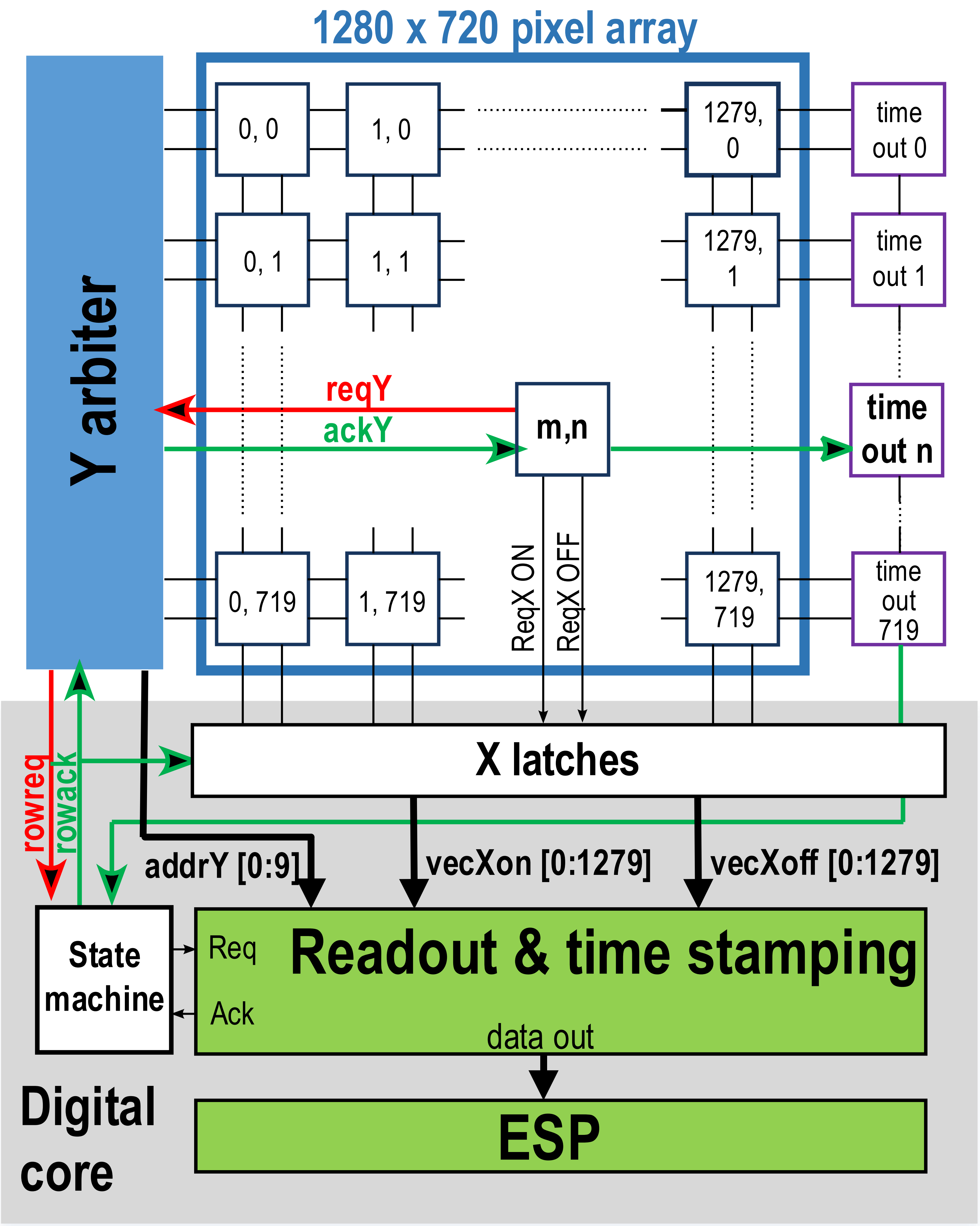




PIXEL ARRAY READOUT

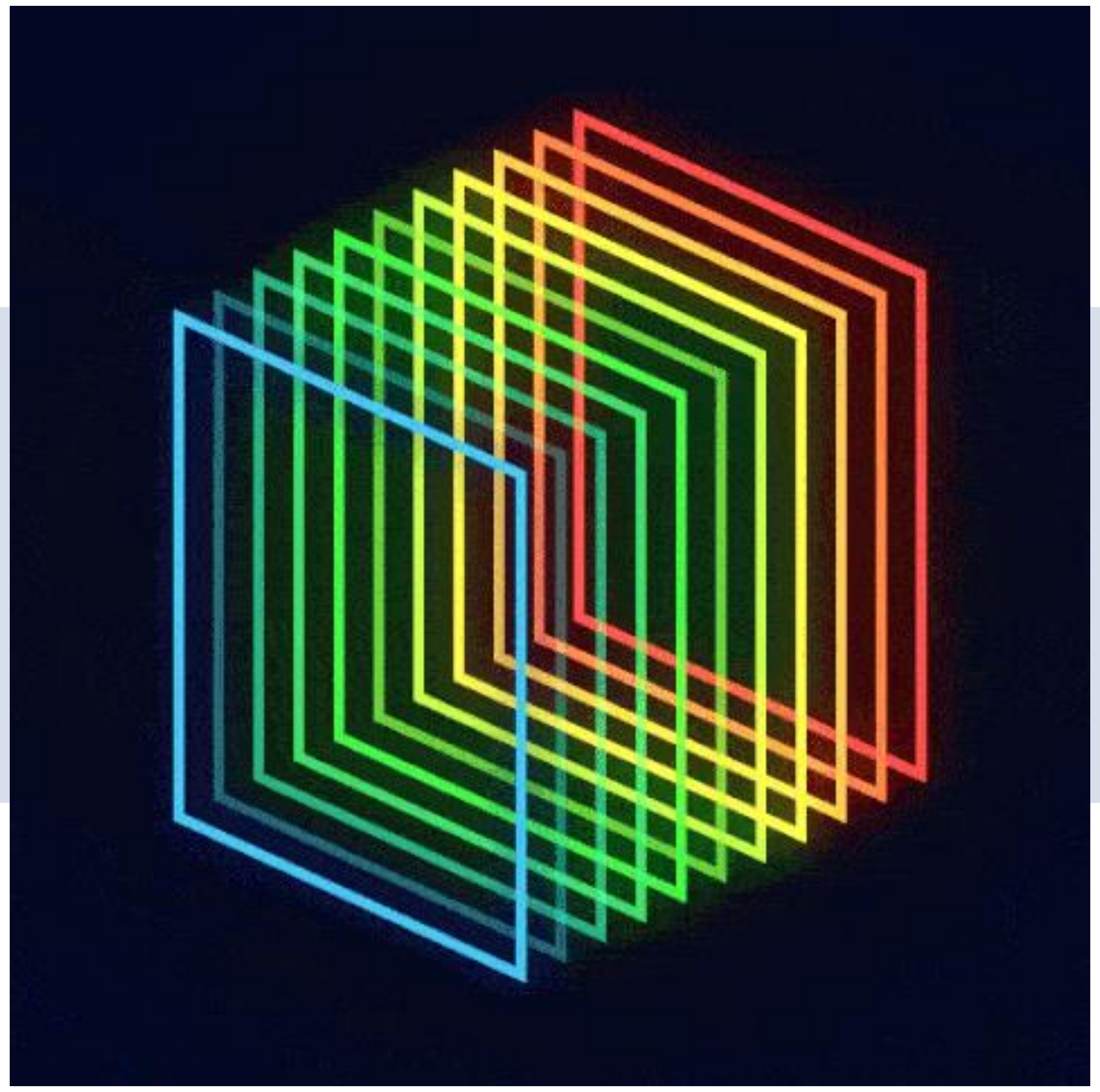
READOUT ON DEMAND

- Readout is <u>not</u> a "scan", asking pixels for data values like in conventional imagers
- Individual autonomous pixels spontaneously and asynchronously request readout via an arbitrated readout system when they have information to convey
- Asynchronous digital handshake protocol out of the pixel array
- Readout system needs to handle up to giga-events per second (GEPS) for large array (>1MP) sensors

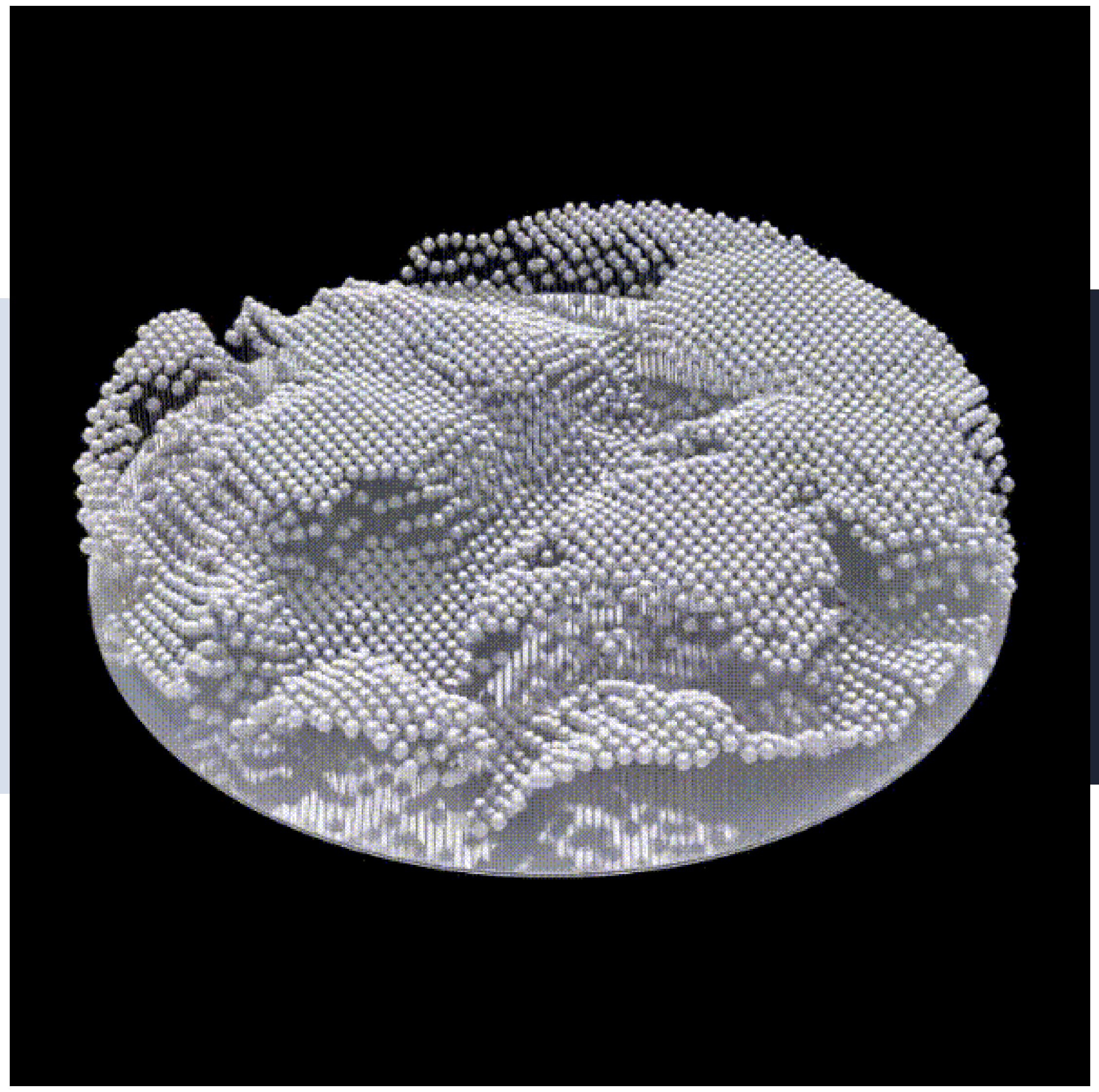


y-address+x-address+polarity+time-stamp

FRAMES VS EVENTS



FRAME-BASED



EVENT-BASED

1/ Generates (sequence of) **static** pictures
2/ **Clock-driven** (pre-defined frame rate)
3/ Has **exposure time**
4/ **Fixed amounts** of data
5/ Beautiful pictures for **human consumption** (MP resolution, color...)

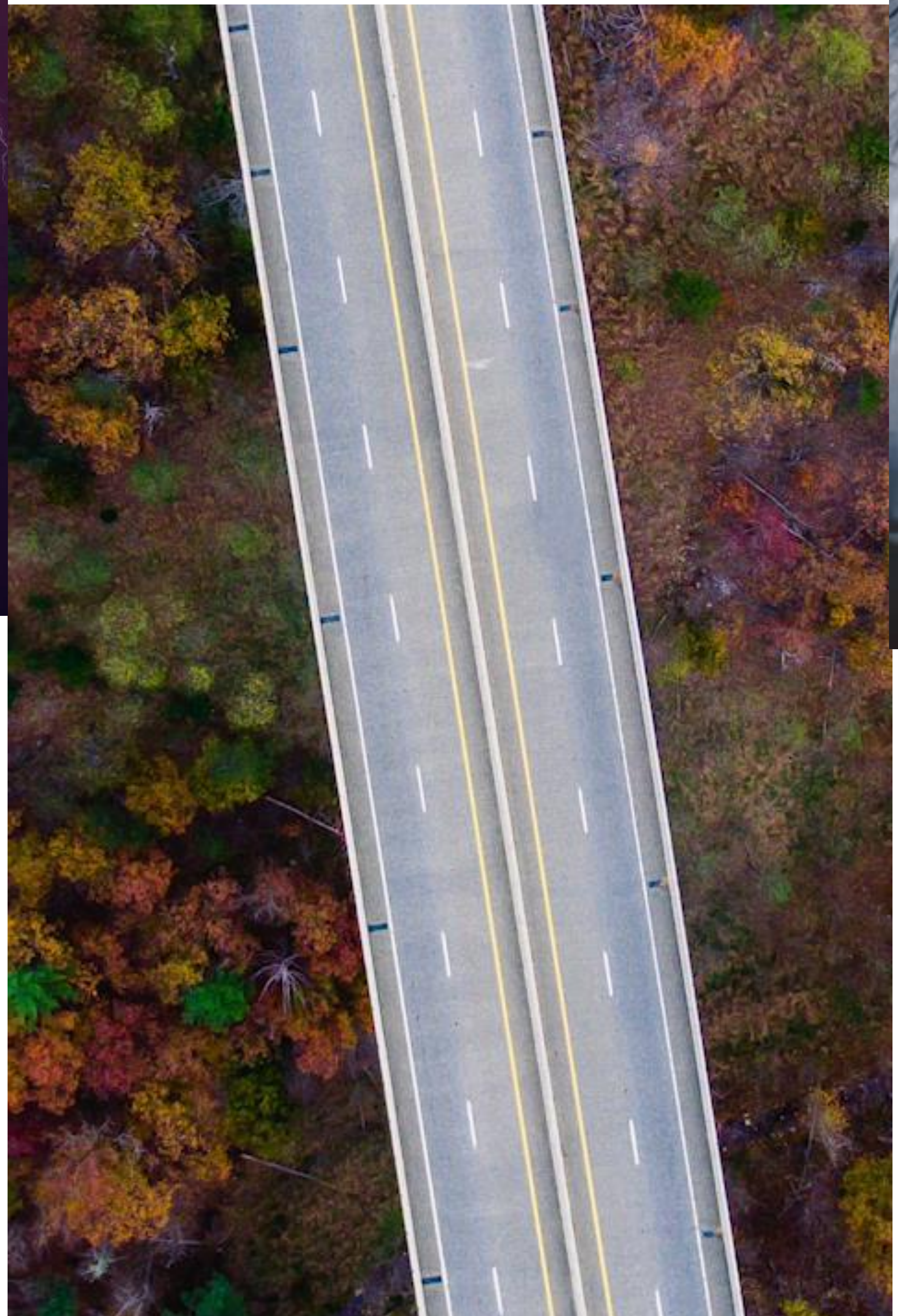
1/ Generates **continuous** events (asynchronous intelligent pixels)
2/ **Data-driven** (sub-ms time resolution - 10,000 fps equivalent)
3/ **No exposure time** (>120dB dynamic range)
4/ Amounts of data **vary with scene dynamics** (10x to 1000x less)
5/ Efficient data for **computer vision** (pre-sorted at pixel level, fast, robust to challenging lighting conditions, native motion-understanding).

HIGH SPEED VISION

Events at **sub-millisecond** time resolution

SPARSE DATA

Redundancy-free compressive sampling



ROBUST TO EXTREME LIGHTING CONDITIONS

>120dB wide dynamic range

POWER EFFICIENCY

Typical **tens-of-milliwatts** sensor power consumption

PROPHESEE


PROPHESEE

TECHNOLOGY

HDR SCENE – IMAGER VS EVENT SENSOR

EVENT SENSORS DEVELOPMENT

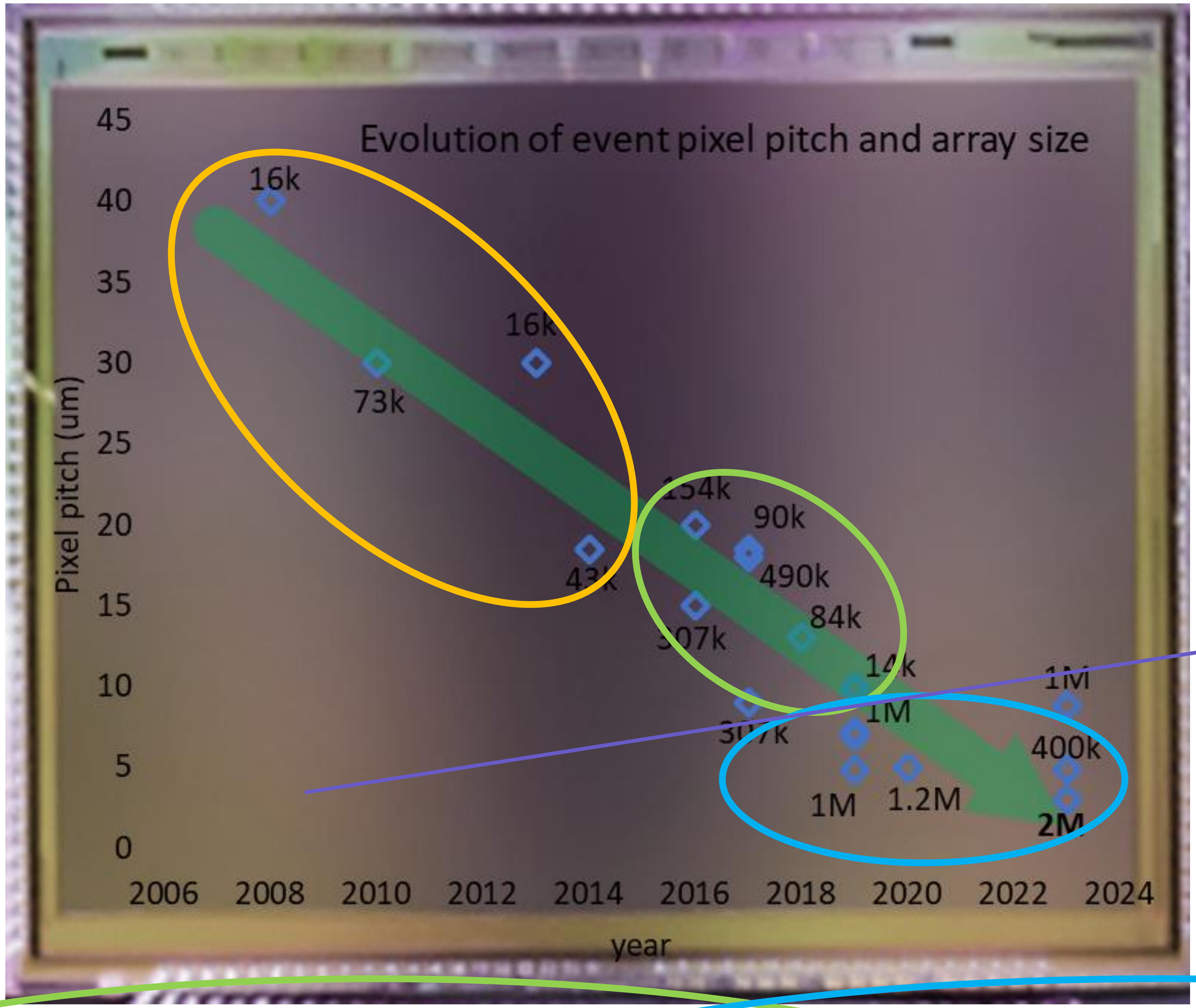
HISTORICAL PIXEL / ARRAY SIZE EVOLUTION

ACADEMIA
RESEARCH

ETH zürich

CSIC

AIT AUSTRIAN INSTITUTE OF TECHNOLOGY



FSI PLANAR

BSI 3D STACKED

START-UPS

insightness
sight for your device

inivation

PROPHESEE

INDUSTRY

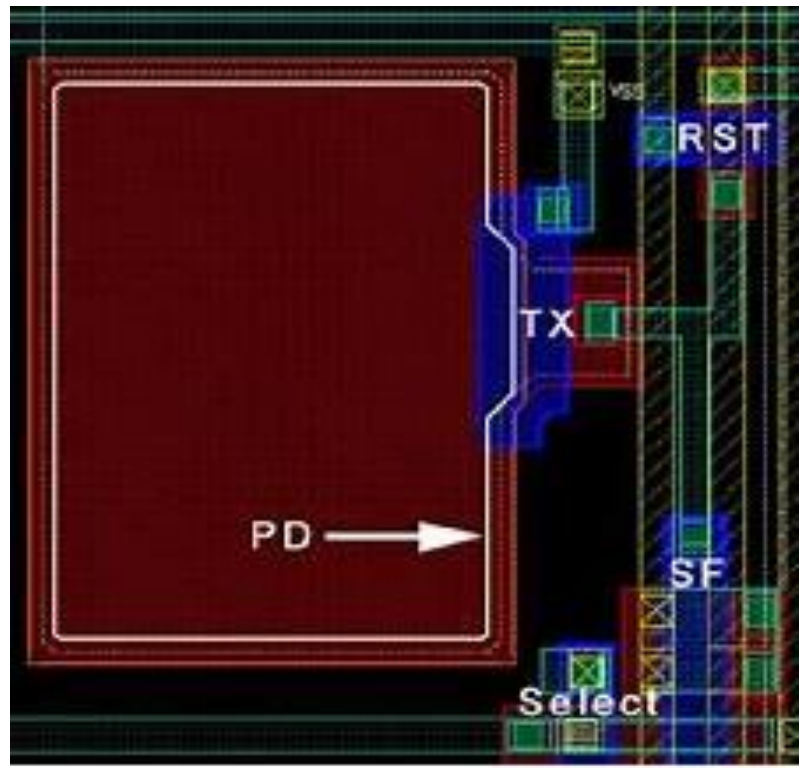
SAMSUNG

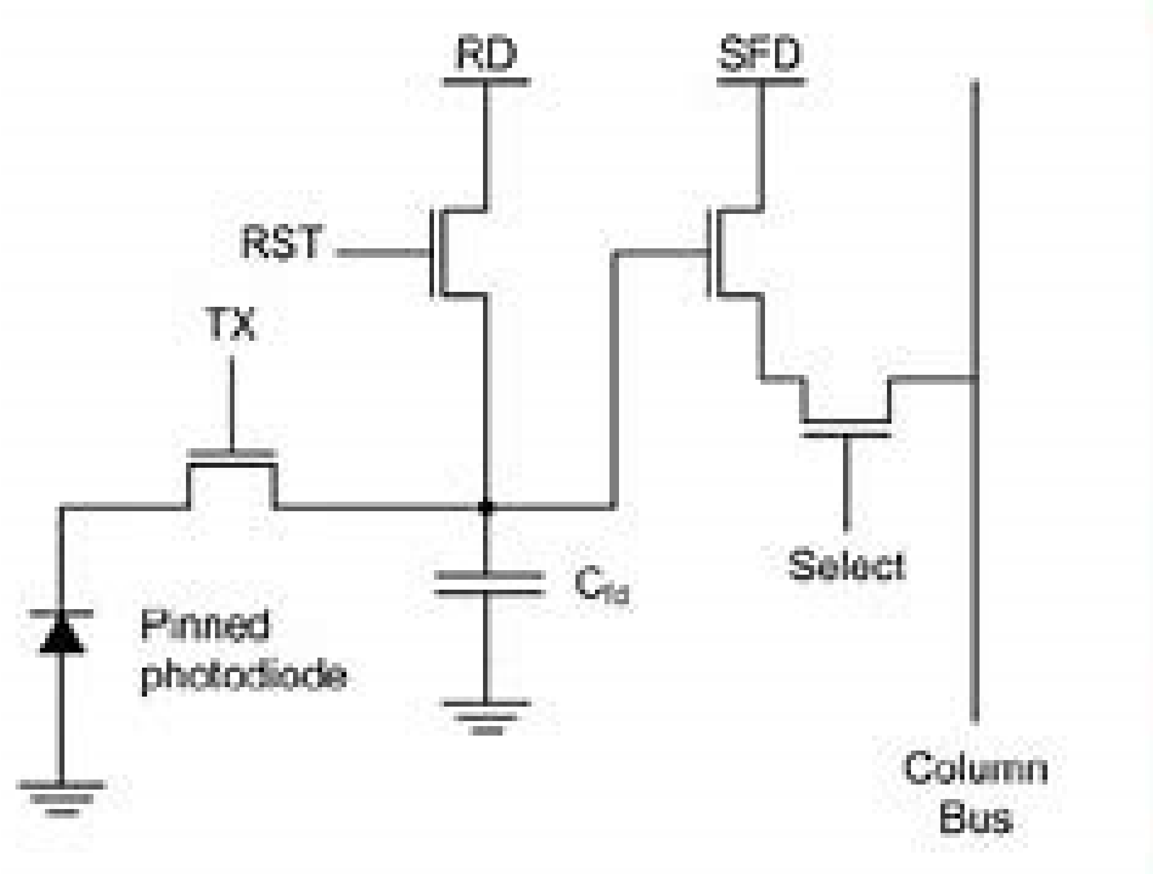
OmniVision

SONY

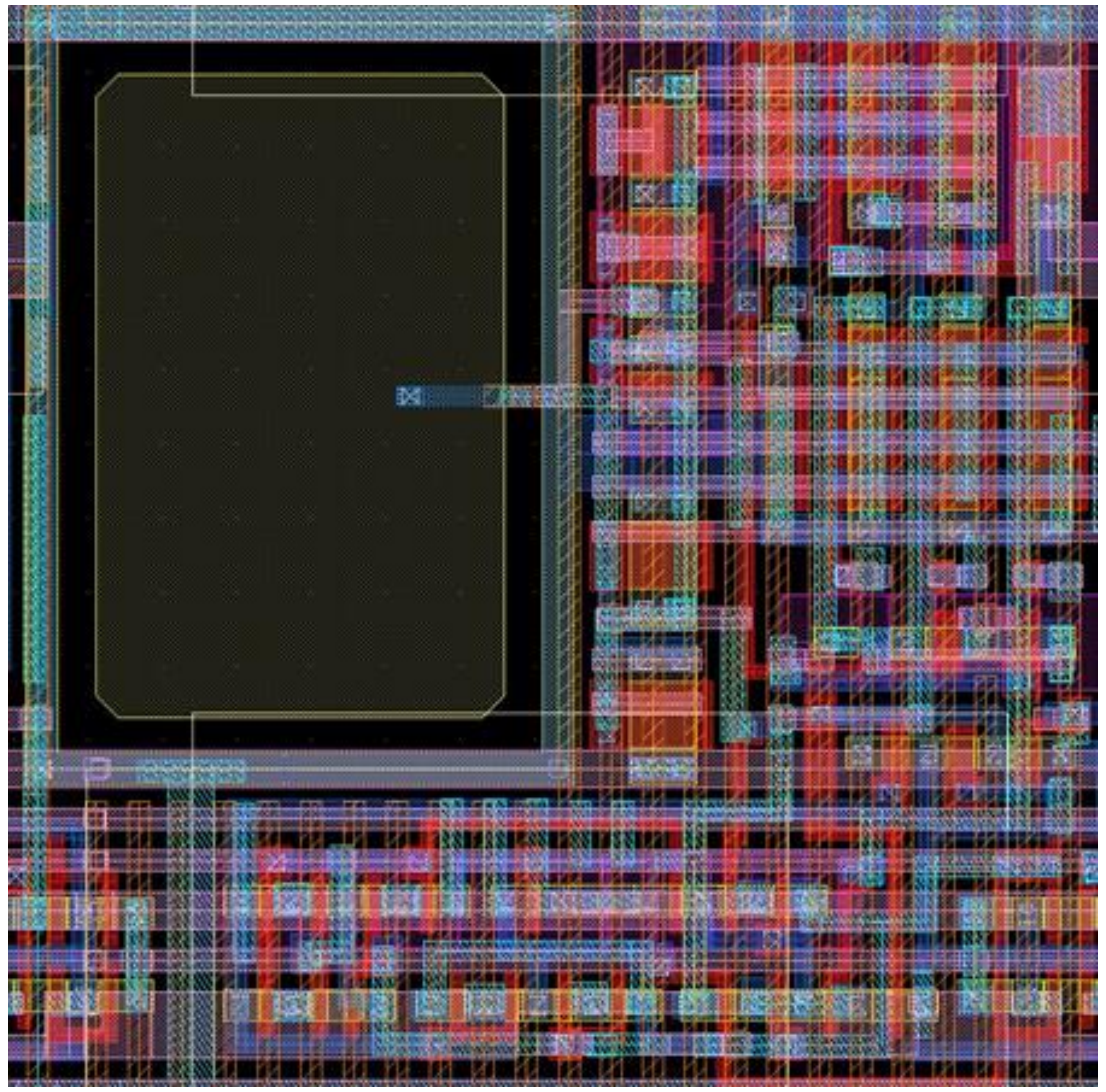
CONVENTIONAL PIXEL VS EVENT PIXEL

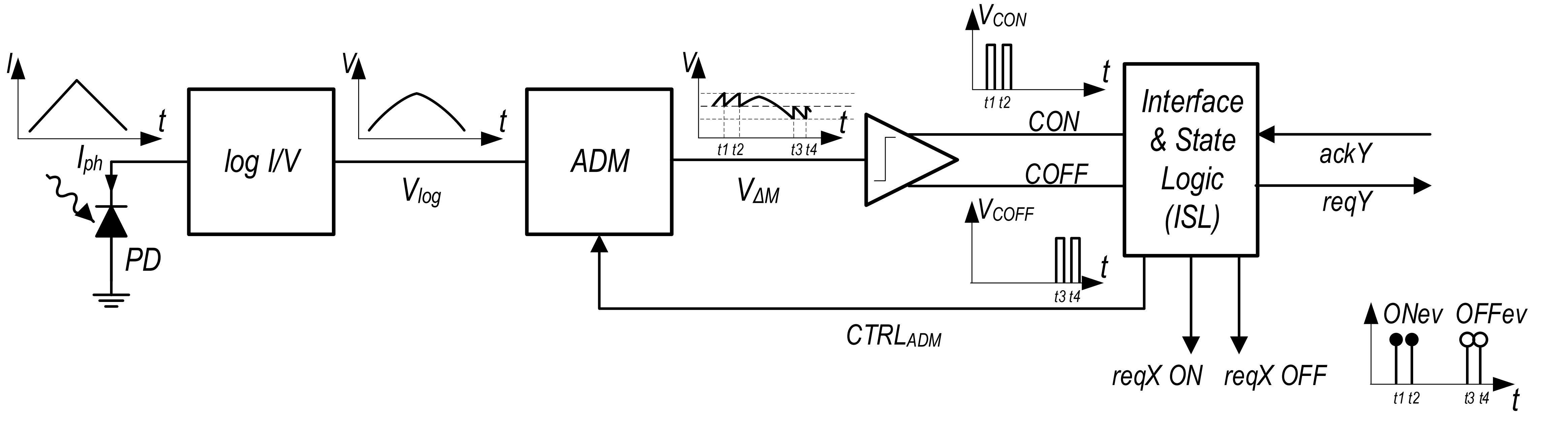
imager pixel





event pixel

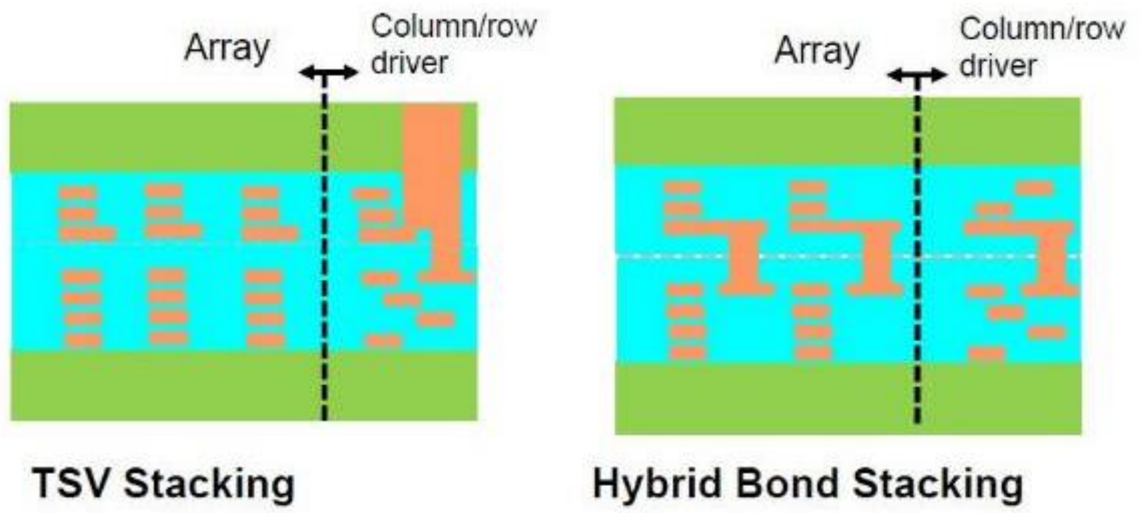


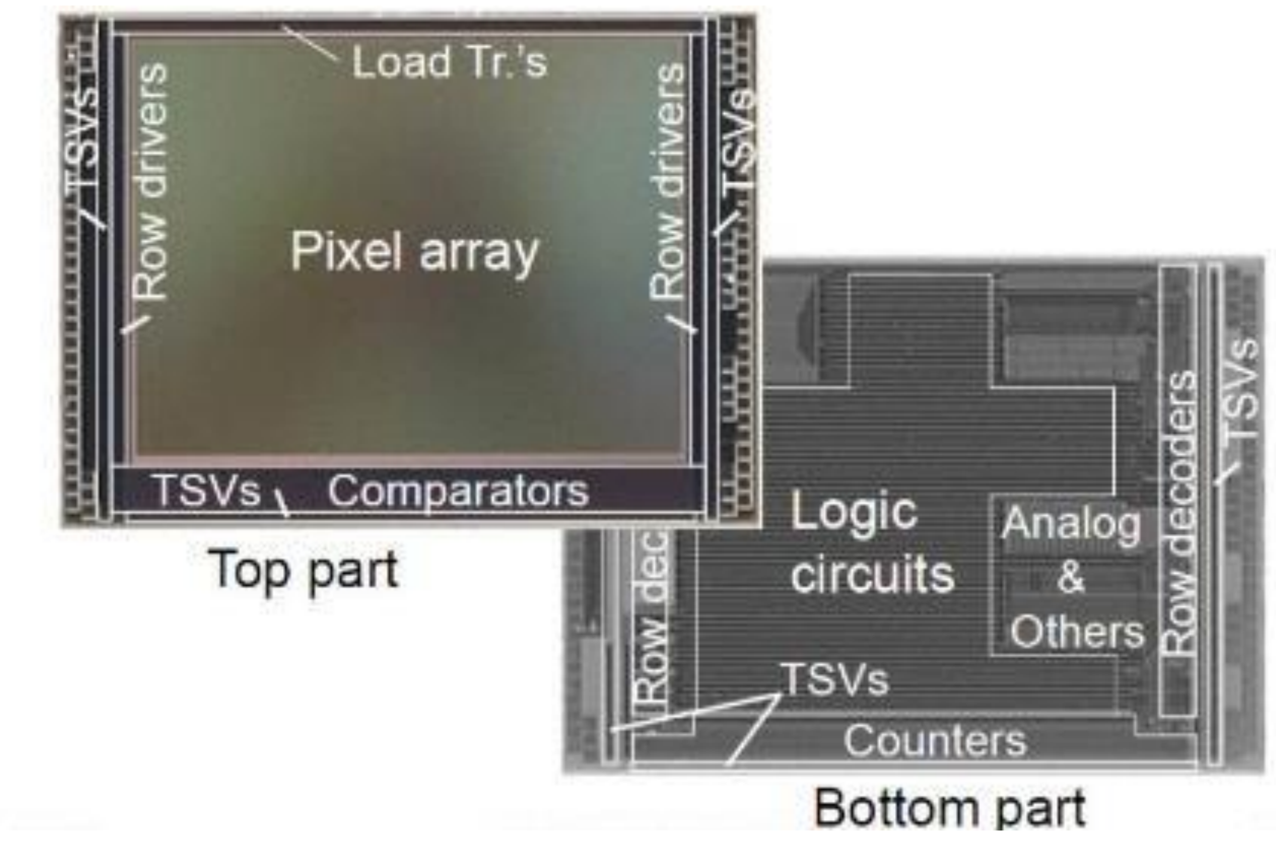


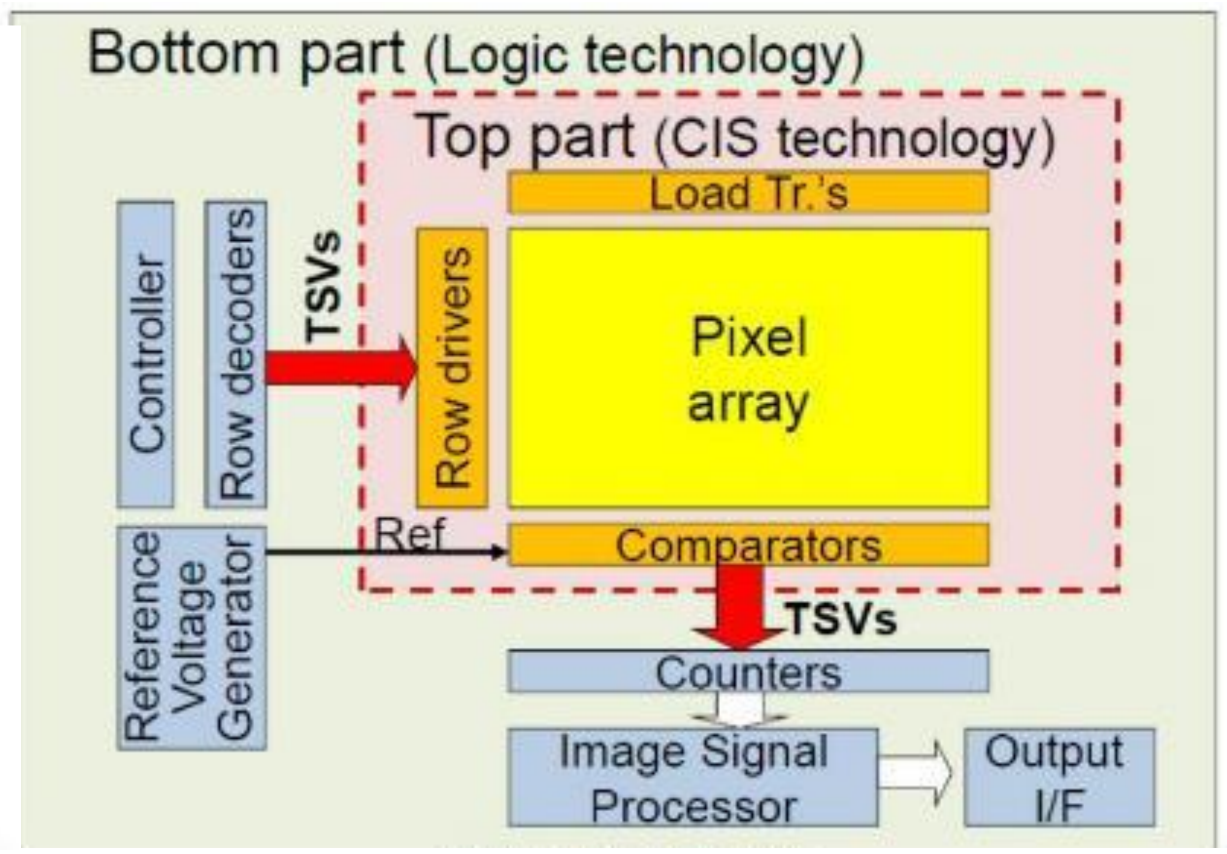
CONFIDENTIAL – UNDER NDA –

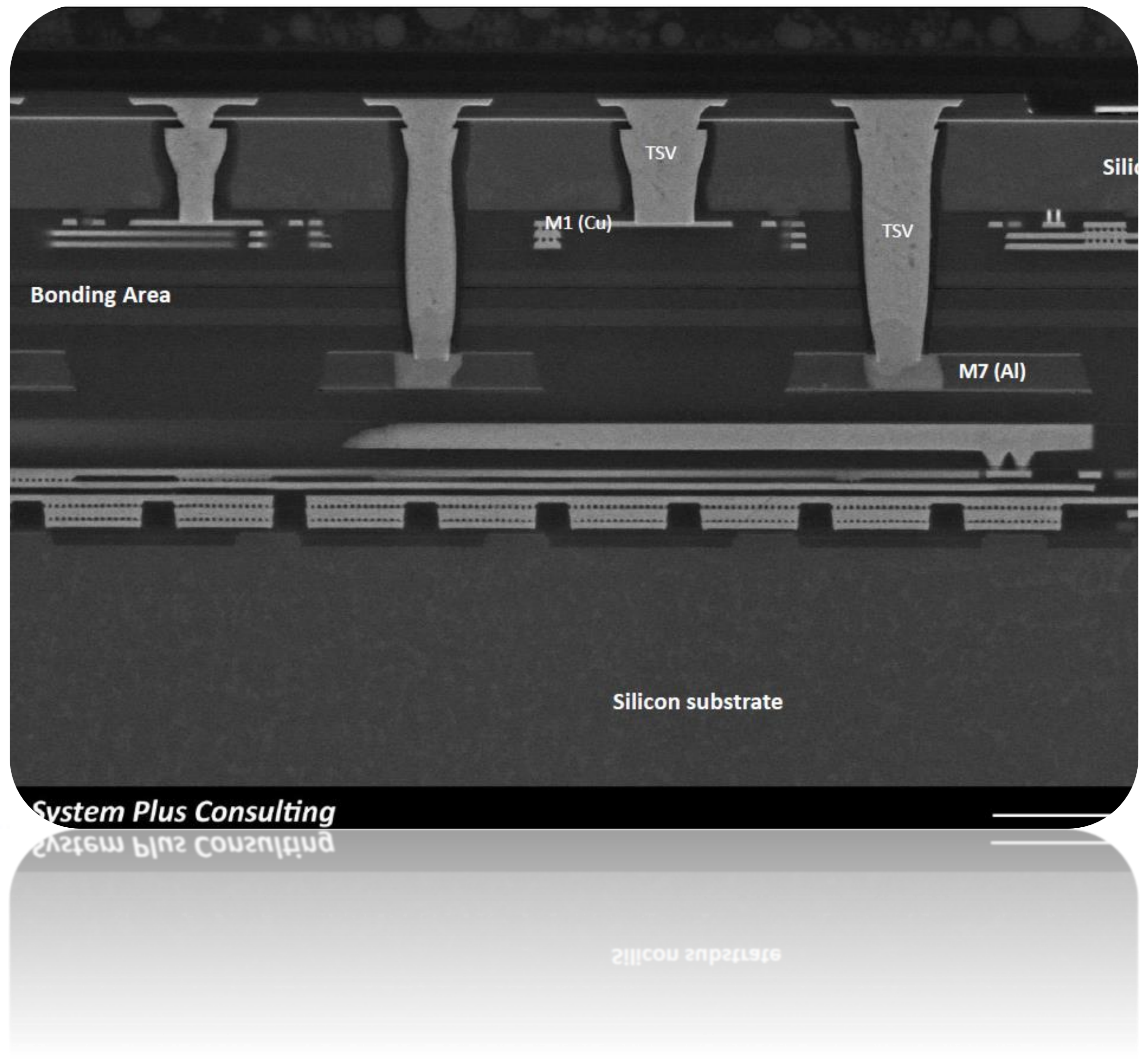
3D BSI IMAGER (CIS) PROCESS

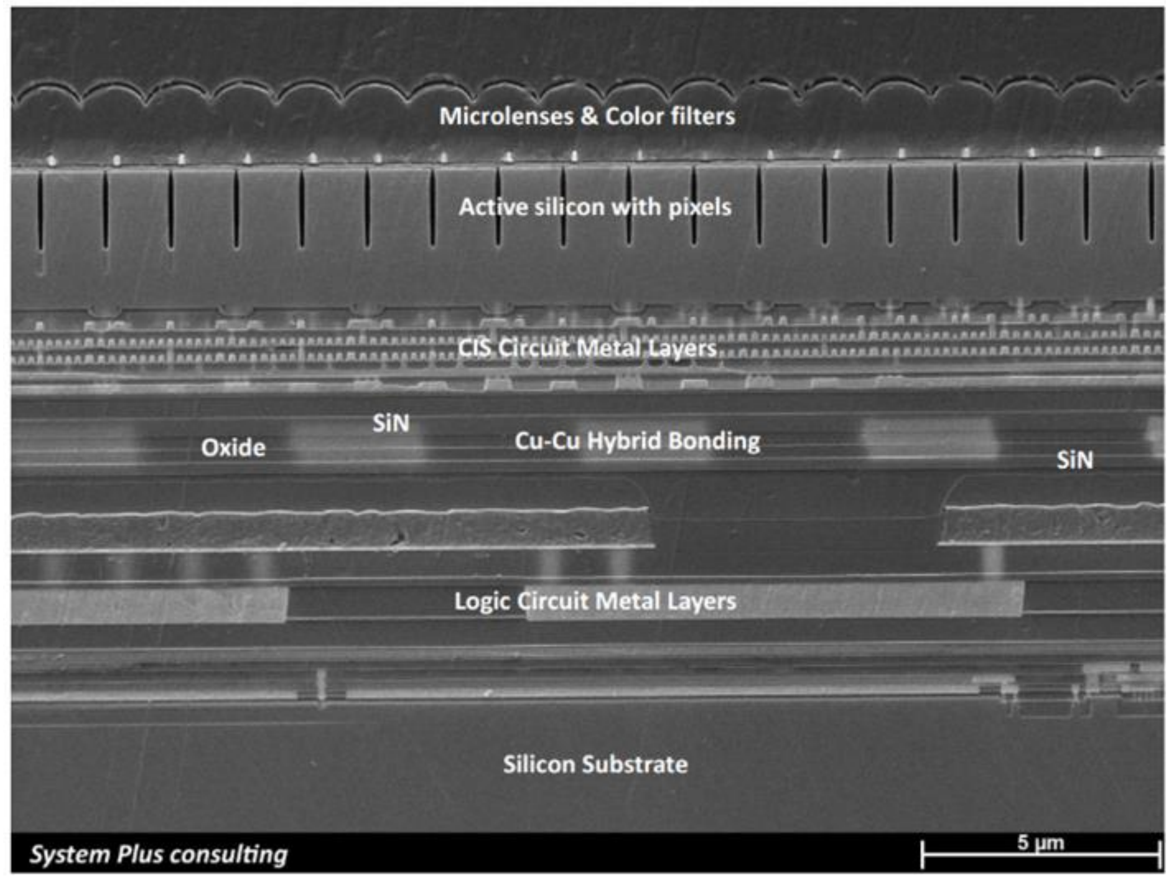
- Introduced in 2013 with 1.2µm pixel pitch, now reaching <1µm
    - Pixel array on top, readout, ADCs, logic circuits components on bottom wafer
    - Optimized processes (CIS, mixed-mode CMOS)
    - Smaller die and camera module size
- Initially array periphery TSV connections between wafers
- Now **pixel-level** connection with **direct Cu/Cu bonding**



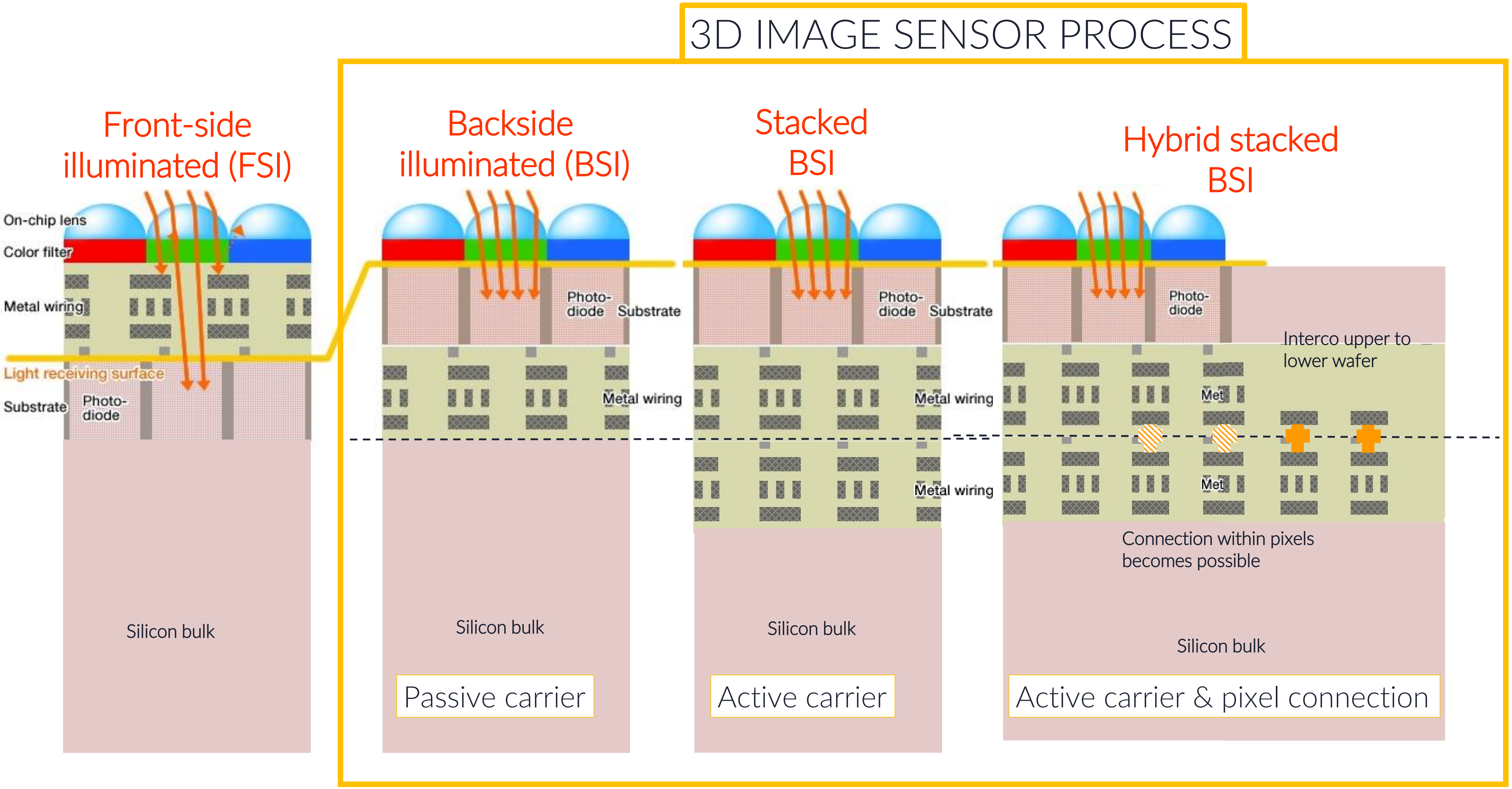






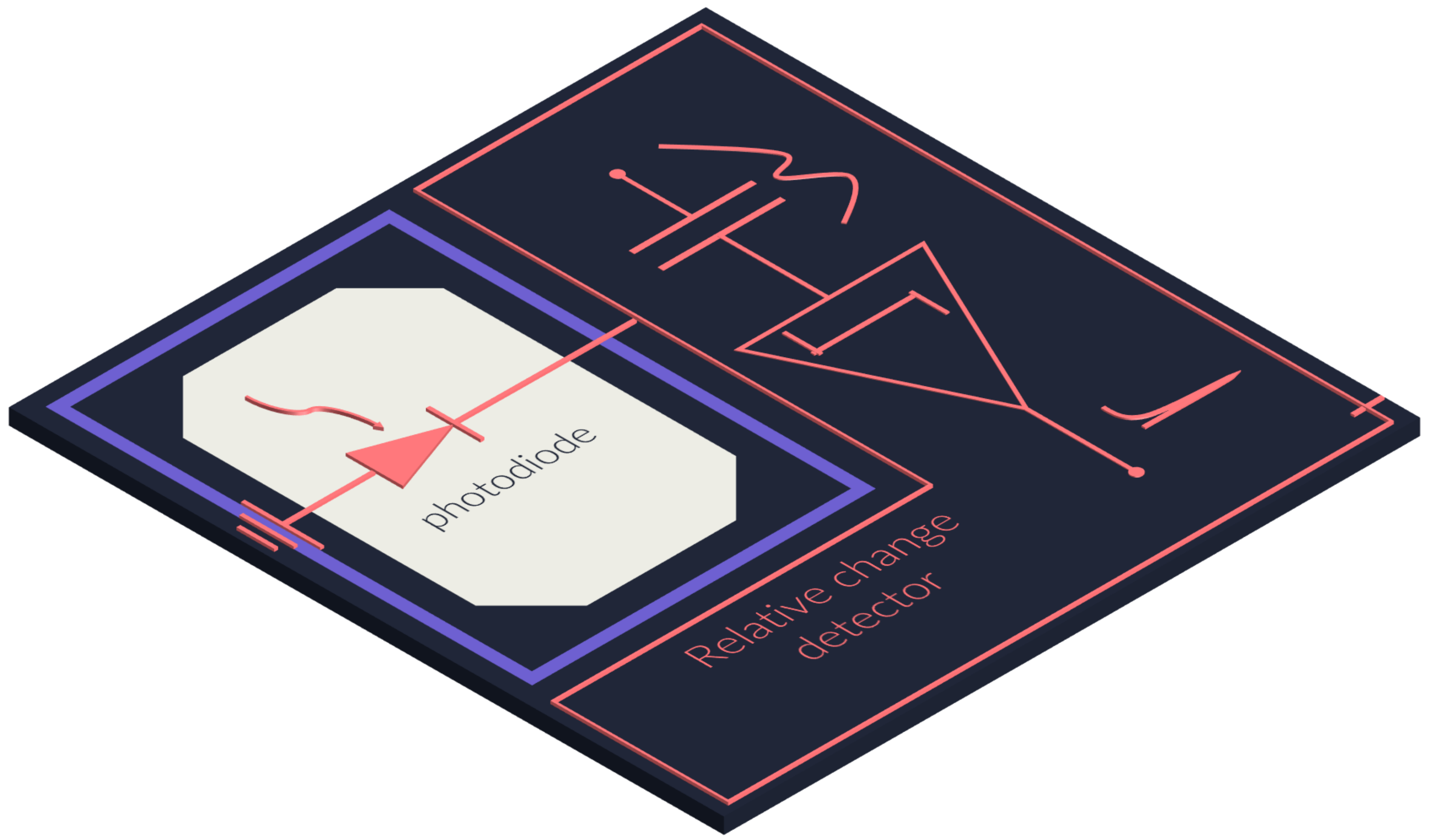


3D CMOS IMAGE SENSOR TECHNOLOGY TREND

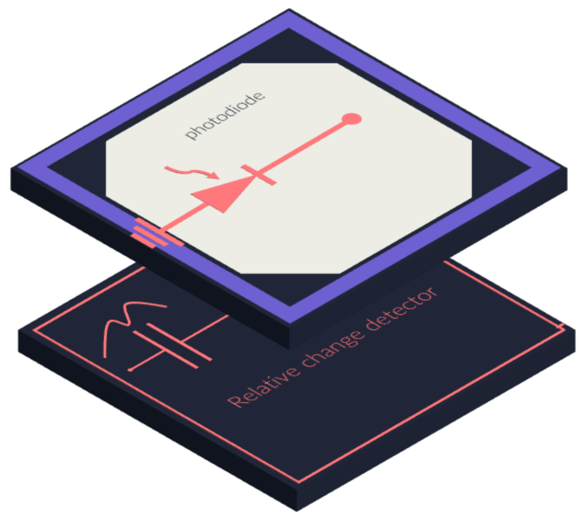


STACKED EVENT PIXEL DESIGN

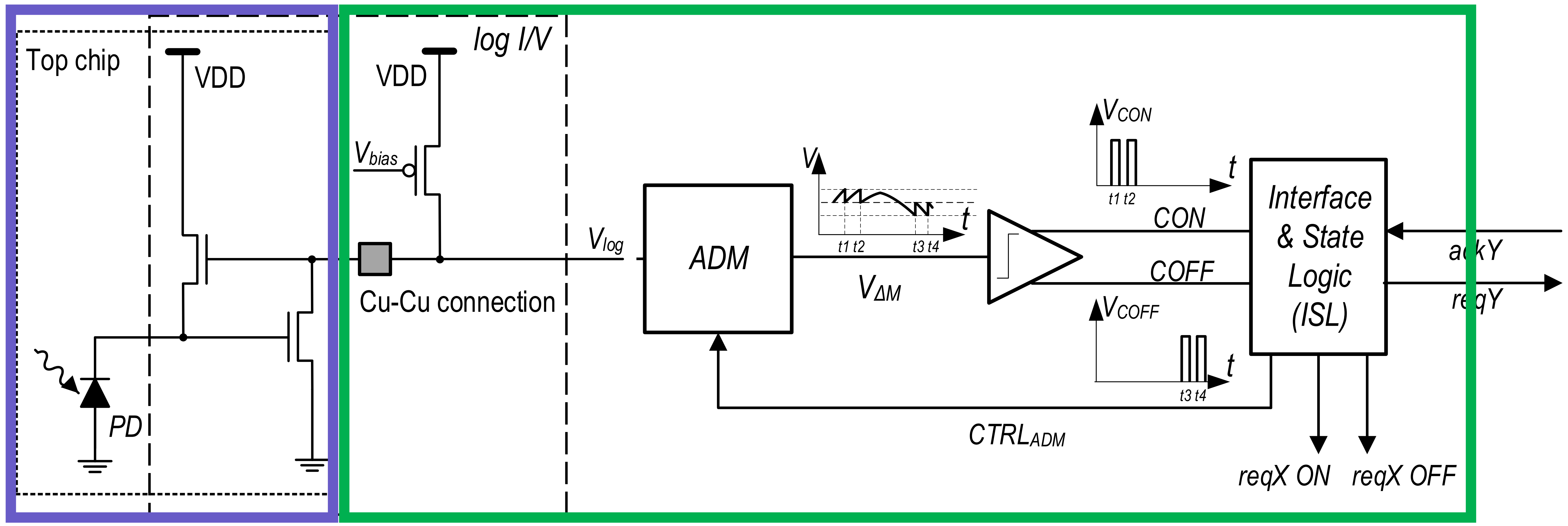
180nm FSI CIS
15µm pitch
25% fill factor



90nm BSI CIS
on 40nm CMOS
4.86µm pitch
>77% fill factor



- ❑ Pixel-level Cu-Cu connection
- ❑ PD + NMOS on top CIS
- ❑ All other pixel circuitry (~50T) on bottom CMOS



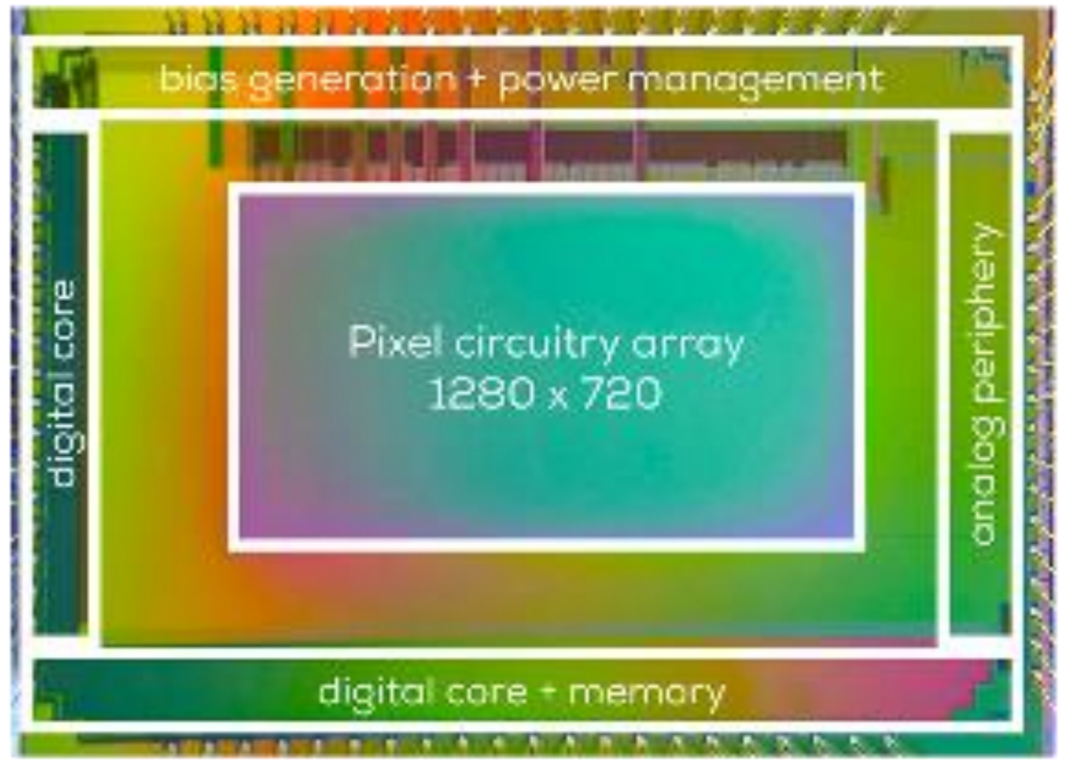
FIRST 3D-STACKED EVENT SENSOR

A 1280 x 720 Back-Illuminated Stacked Temporal Contrast Event-based Vision Sensor with 4.86µm Pixels, 1.066GEPS Readout, Programmable Event Rate Controller and Compressive Data Formatting Pipeline

Thomas Finateu[1], Atsumi Niwa[2], Daniel Matolin[1], Koya Tsuchimoto[2], Andrea Mascheroni[1], Etienne Reynaud[1], Pooria Mostafalu[3], Frederick Brady[3], Ludovic Chotard[1], Florian LeGoff[1], Hirotsugu Takahashi[2], Hayato Wakabayashi[2], Yusuke Oike[2], Christoph Posch[1]

[1] Prophesee, Paris, France
[2] Sony Semiconductor Solutions Corporation, Atsugi, Japan
[3] Sony Electronics Inc., Rochester, NY

© 2020 IEEE
International Solid-State Circuits Conference



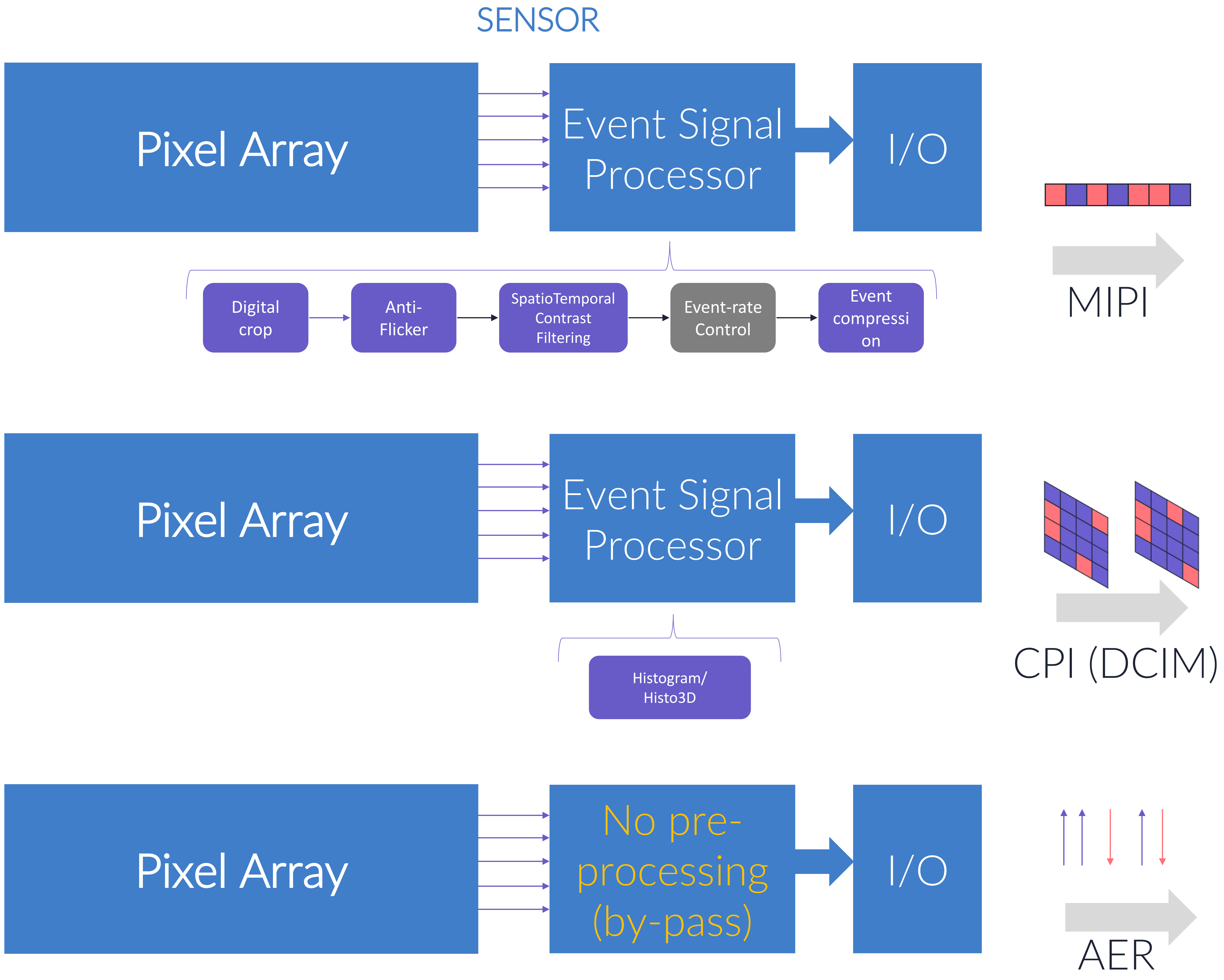
KEY FEATURES

• Resolution (px)	1280 x 720
• Optical format:	1/2.5"
• Step response latency 1kux (µs)	<100
• Dynamic Range (dB)	>120**
• Nominal contrast treshold (%)	25
• Pixel size (µm)	4.86 x 4.86
• Event Signal Processing ESP	

PROPHESEE | SONY

FUTURE EVENT SENSOR EVOLUTION

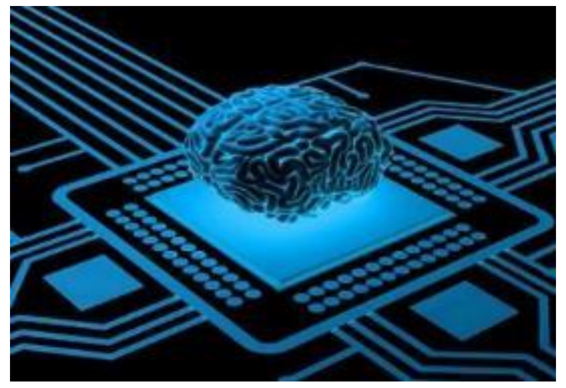
EVENT SENSOR → PROCESSOR



COMPUTE

Application Processor
Continuous EVT Stream
Embedded vision applications

Microcontroller
Direct 2D event frames with controllable rate and activity thresholds.
Smart wake-up, people monitoring, CNN edge-AI applications



Neuromorphic processor
Direct AER input (spiking NN)
AI Algos for object detection and tracking, motion analysis, optical flow, attention tracking, surveillance, gesture recognition, …

EVENT-BASED COMPUTER VISION

EVENT-BASED VISION

EVENT SENSOR

- Efficient acquisition of (dynamic) visual information
  - inherent data compression – sparse encoding
  - focus on relevant dynamic data
- High dynamic range (HDR) from individual autonomous pixel operation
- High-speed continuous-time motion capture
  - fast pixel reaction times
  - high resolution timestamping (1us)

EVENT-BASED COMPUTER VISION

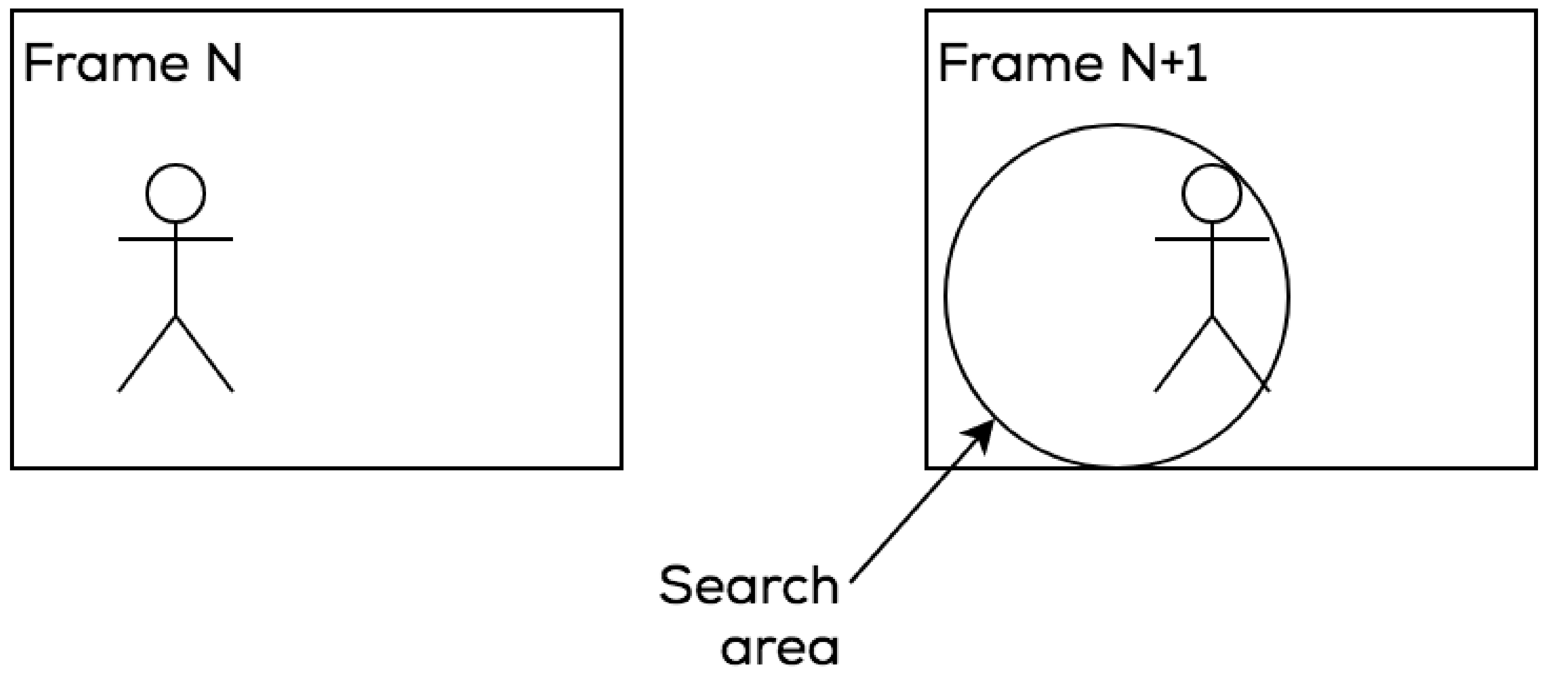
- Time-domain information processing
  - Time is another dimension of information for vision analysis tasks and efficient machine-learning models
- Sparse data benefit real-time vision:
  - Object detection, classification, tracking, …
  - Motion analysis, motion flow, stereo 3D, …
- High-temporal resolution - kHz update rates

TRACKING

STANDARD CAMERA

FRAME-BASED = SEARCH AREA THAT IS BOTH:

- SMALL ENOUGH TO GET MEANINGFUL AND FAST MATCHES
- LARGE ENOUGH TO FOLLOW FAST-MOVING OBJECTS



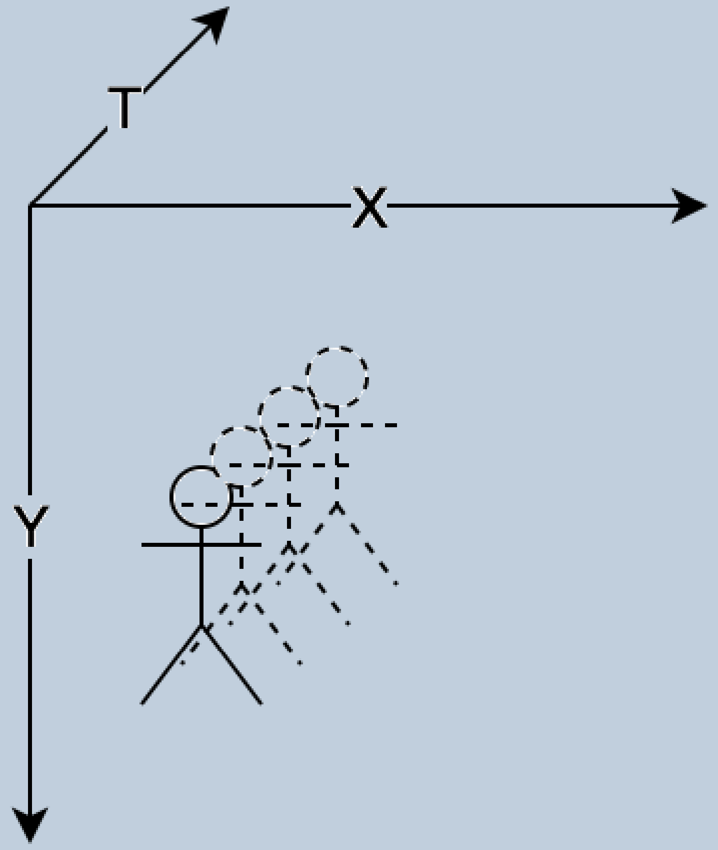
FRAME-BASED

TRACKING

EVENT CAMERA

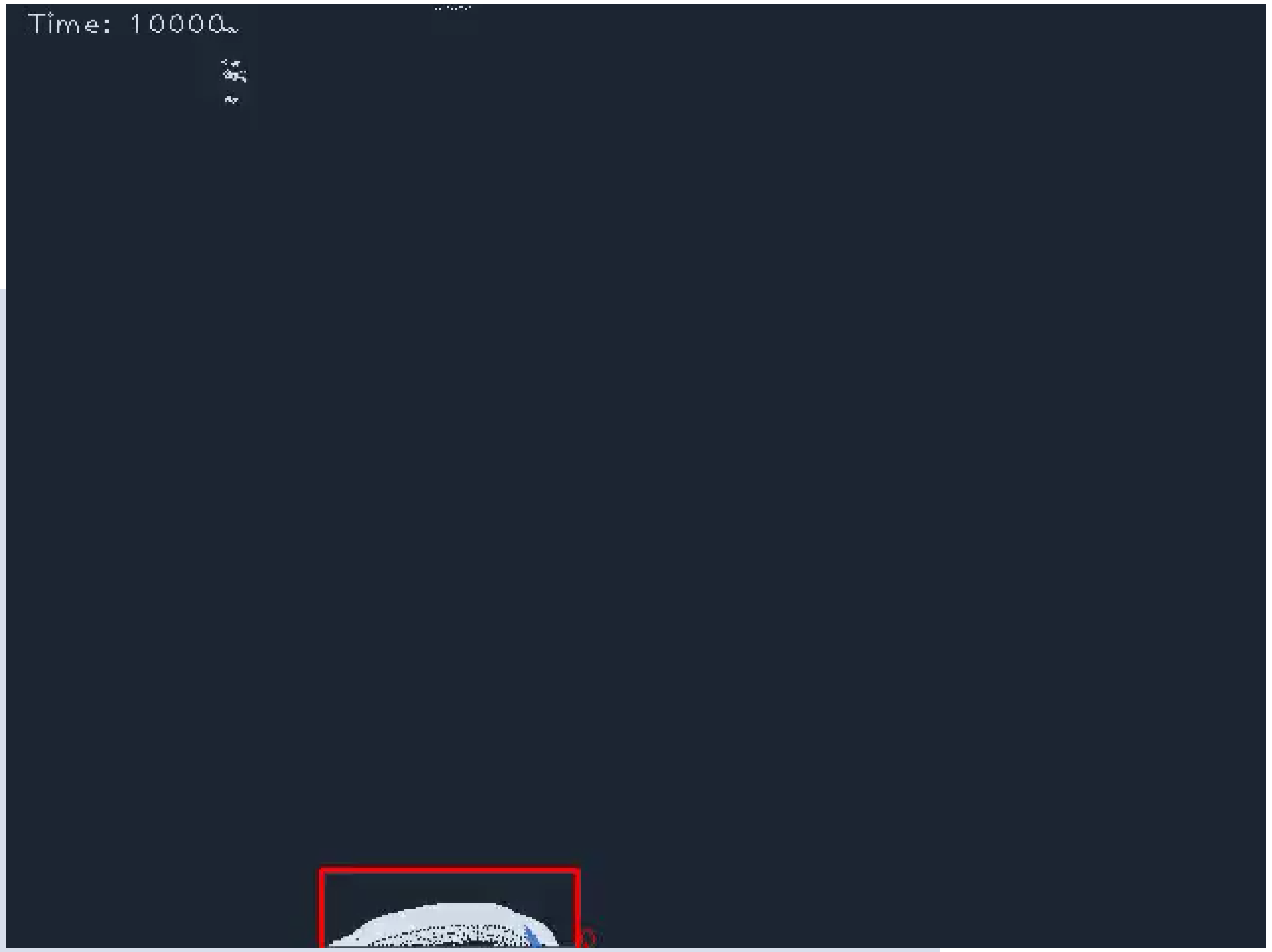
EVENT-BASED = 1 PIXEL SEARCH AREA

- THE OBJECT WILL TRIGGER EVERY PIXEL ON THE WAY



EVENT-BASED

OBJECT TRACKING



Track moving objects in the field of view. Leverage the **low data-rate and sparse information** provided by event-based sensors to track objects with **low compute power.**

Continuous tracking in time: **no more "blind spots" between frame acquisitions**
Native segmentation: **analyze only motion, ignore the static background**

OPTICAL FLOW

STANDARD CAMERA

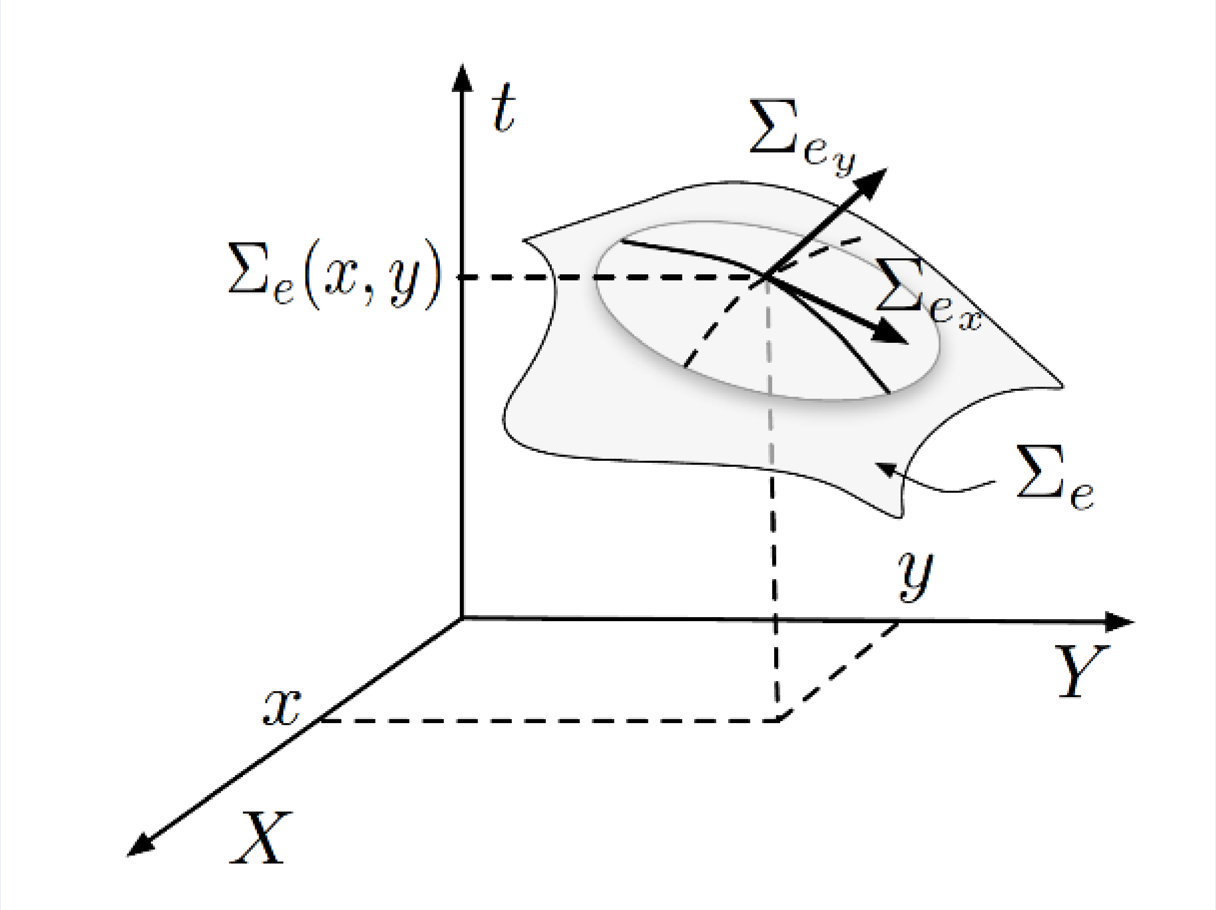
- OPTICAL FLOW IS COMPUTED FROM 2 CONSECUTIVE IMAGES
- AT EACH PIXEL, COMPUTE
  - THE LOCAL INTENSITY GRADIENT ΔI_X
  - THE INTENSITY DERIVATIVE W.R.T. PREVIOUS IMAGE ΔI_T
- FLOW IS $\Delta I_X / \Delta I_T$

FRAME-BASED

OPTICAL FLOW

EVENT CAMERA

- OPTICAL FLOW IS COMPUTED FROM THE TIME SURFACE
- AT EACH POSITION, OPTICAL FLOW IS THE TANGENT PLANE



EVENT-BASED

OPTICAL FLOW

Rediscover this fundamental computer vision building block, but with an event twist.

Understand motion much more efficiently, through **continuous pixel-by-pixel tracking and not sequential frame by frame analysis** anymore.

17x less power **compared to traditional image-based approaches**
Get features only on moving objects

STEREOVISION

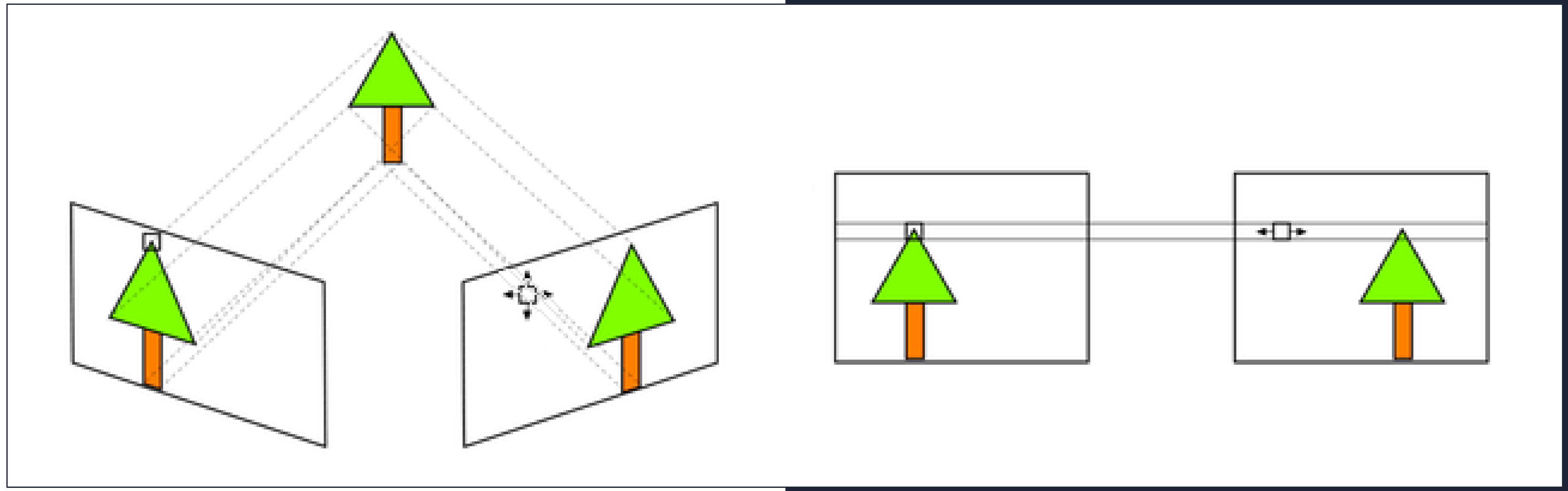
STANDARD CAMERA

- FOR EACH PATCH OF LEFT IMAGE, FIND THE MOST SIMILAR ON RIGHT IMAGE
- USE EPIPOLAR CONSTRAINT TO NARROW THE SEARCH

STEREOVISION

EVENT CAMERA

- FOR EACH EVENT OF LEFT CAMERA, FIND THE CLOSEST ONE IN TIME OF RIGHT CAMERA
- CAN ALSO BENEFIT FROM EPIPOLAR CONSTRAINT TO SOLVE AMBIGUITY



FRAME-BASED

EVENT-BASED

APPLICATIONS

A 4TH DISRUPTION IN IMAGE SENSING

MARKET

Film Photography

Digital Photography

Mobile Photography

SENSING
(not solely related to 'Pretty Picture' parameters)

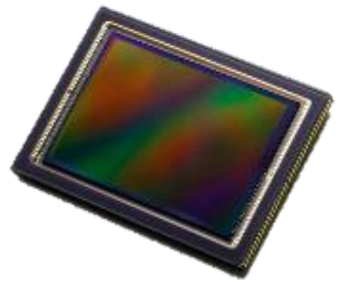
+

1865 — 1945 — 2005 — 2015 — 2020

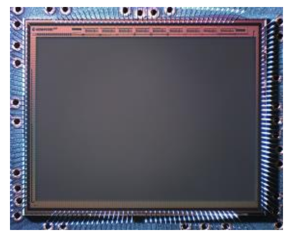
TECHNOLOGY

TUBES

CCD



CMOS



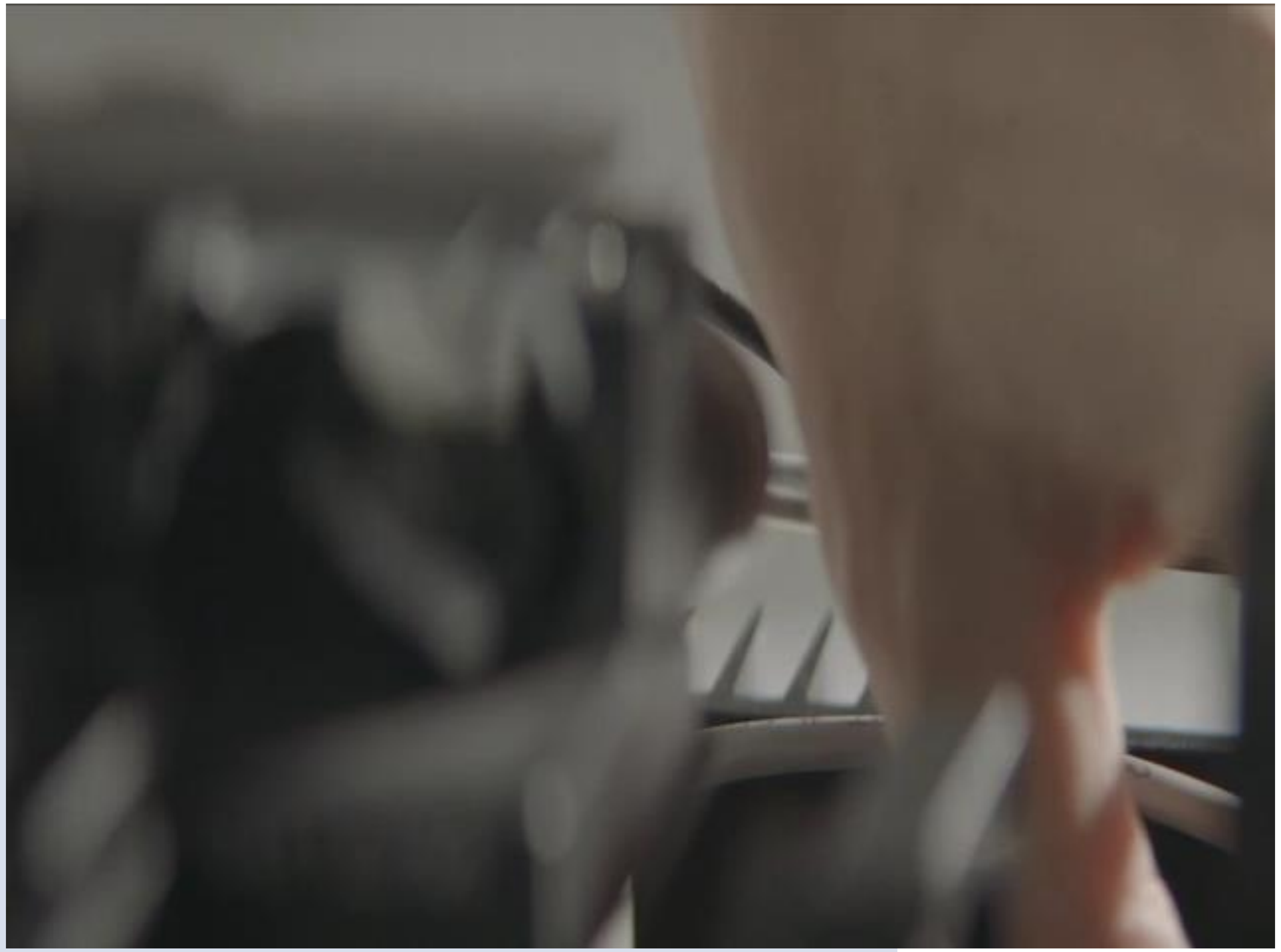
CMOS+ (Enhanced 3D Stacking focusing on full solution at edge)

PROPHESEE
METAVISION FOR MACHINES

FRAME-BASED → EVENT-BASED


MACHINE LEARNING
DETECTION INFERENCE
DETECTION TRAINING
EVENT SIMULATOR
DETECTION KPI
OPTICAL FLOW INFERENCE
OPTICAL FLOW TRAINING
ANALYTICS
PARTICLE SIZE MONITORING
Counter : 0
JET MONITORING
OBJECT TRACKING
VIBRATION MONITORING
Frequency : N/A
SPATTER MONITORING
HIGH-SPEED COUNTING
Counter : 0
CALIBRATION
CALIBRATION
COMPUTER VISION 3D
NEW
EDGELET TRACKING
COMPUTER VISION
OPTICAL FLOW
CORE
ULTRA SLOW MOTION
XYT VISUALIZATION
DATA RATE VISUALIZATION

ULTRA FAST LASER TRACKING



Delivers time-resolution equivalent to **200,000+ frames per second, live**, while generating orders of magnitude **less data** than conventional high-speed cameras.

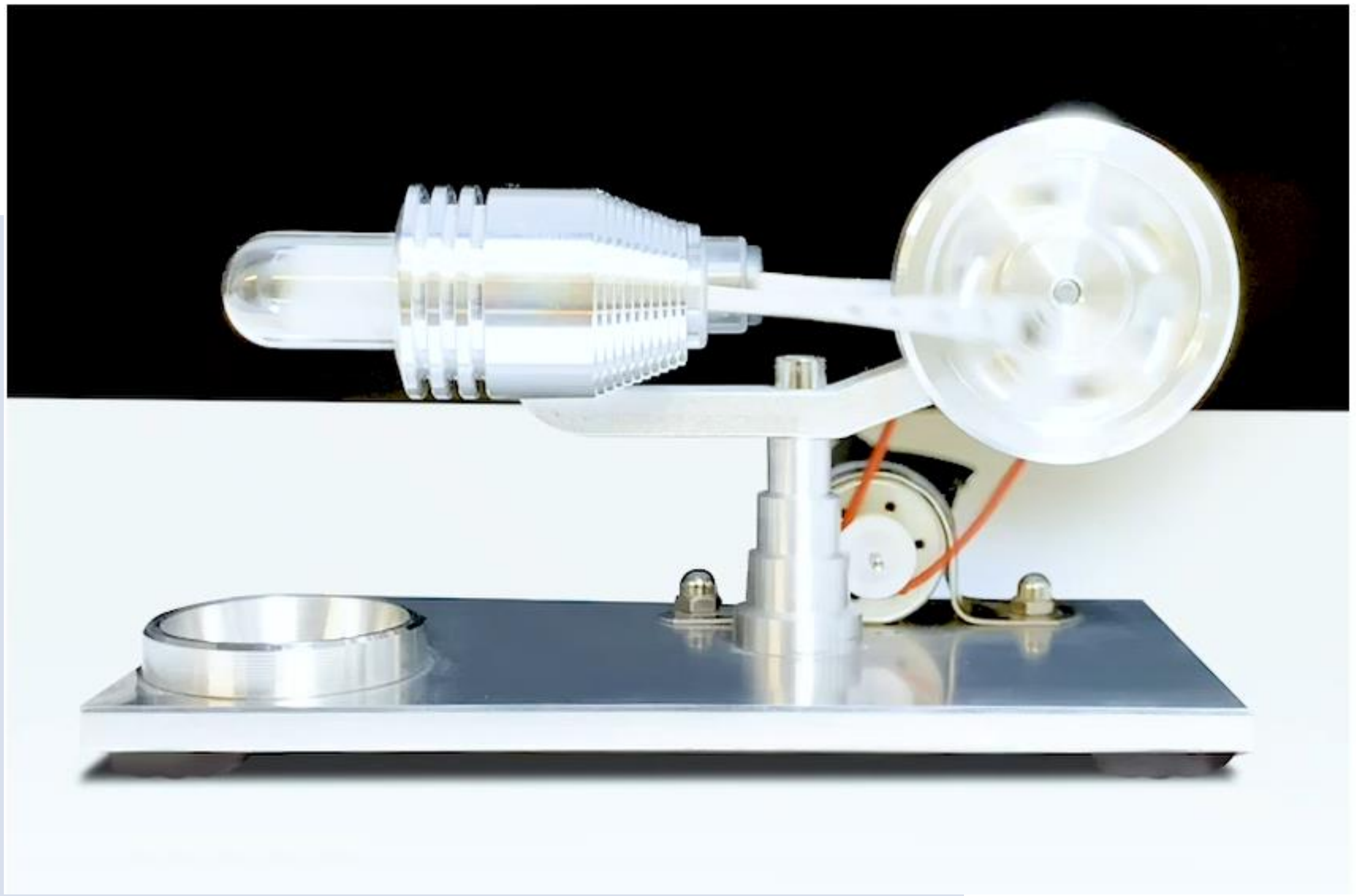
Analyze **finest motion dynamics** hiding in ultra fast processes.

5µs temporal resolution
(200,000 frames-per-second equivalent)

VIBRATION & FREQUENCY MONITORING

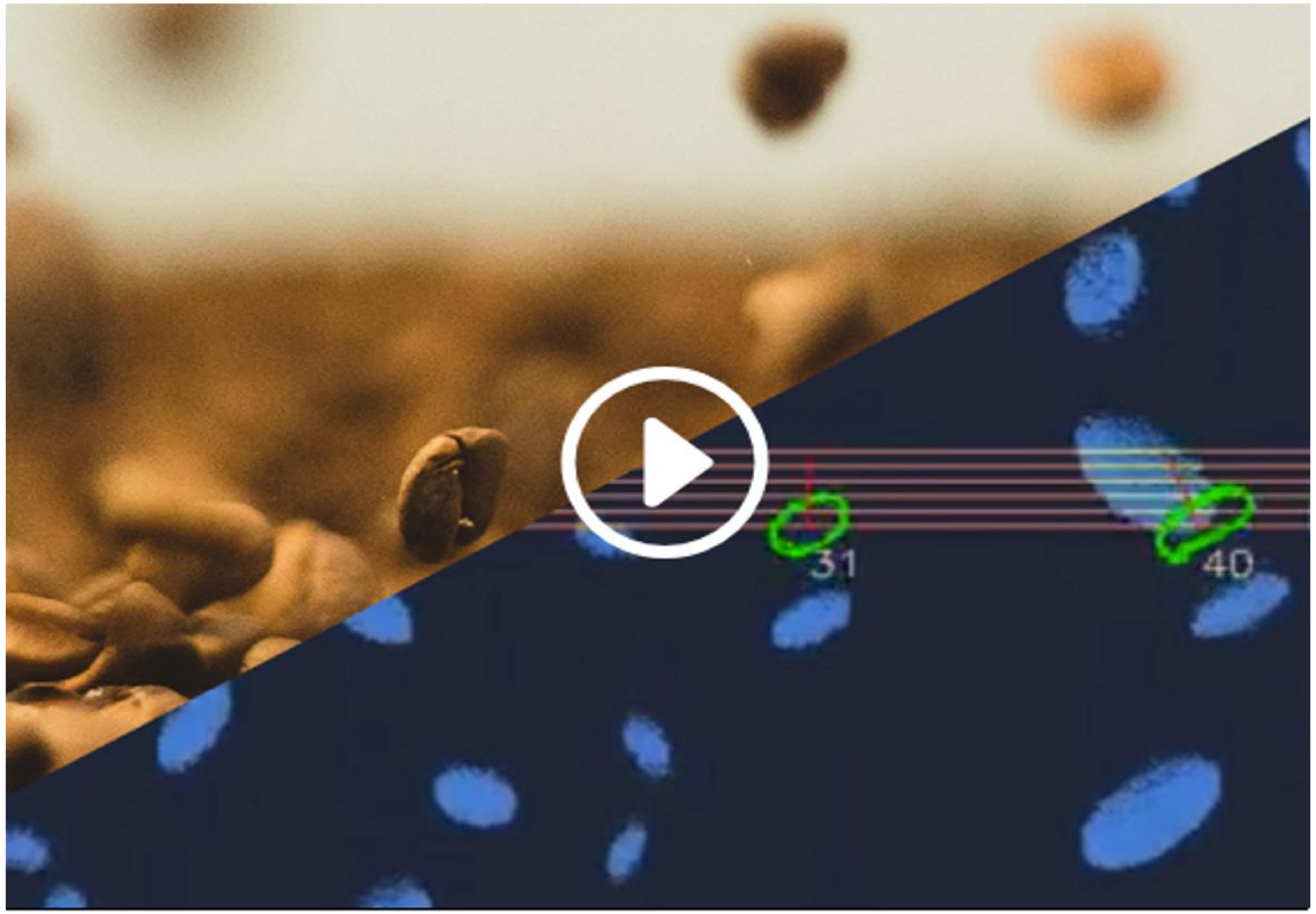
ANALYTICS

- **Remotely** – no need to access or touch the object
- **Measure multiple frequencies simultaneously** at different parts of the object / scene
- From below **1Hz to tens of kHz**
- Industrial process monitoring, predictive maintenance, …



Typical use cases: *Motion monitoring, Vibration monitoring, Frequency analysis for predictive maintenance*

HIGH-SPEED COUNTING PARTICLE SIZE MONITORING



Count and **measure size** of objects:

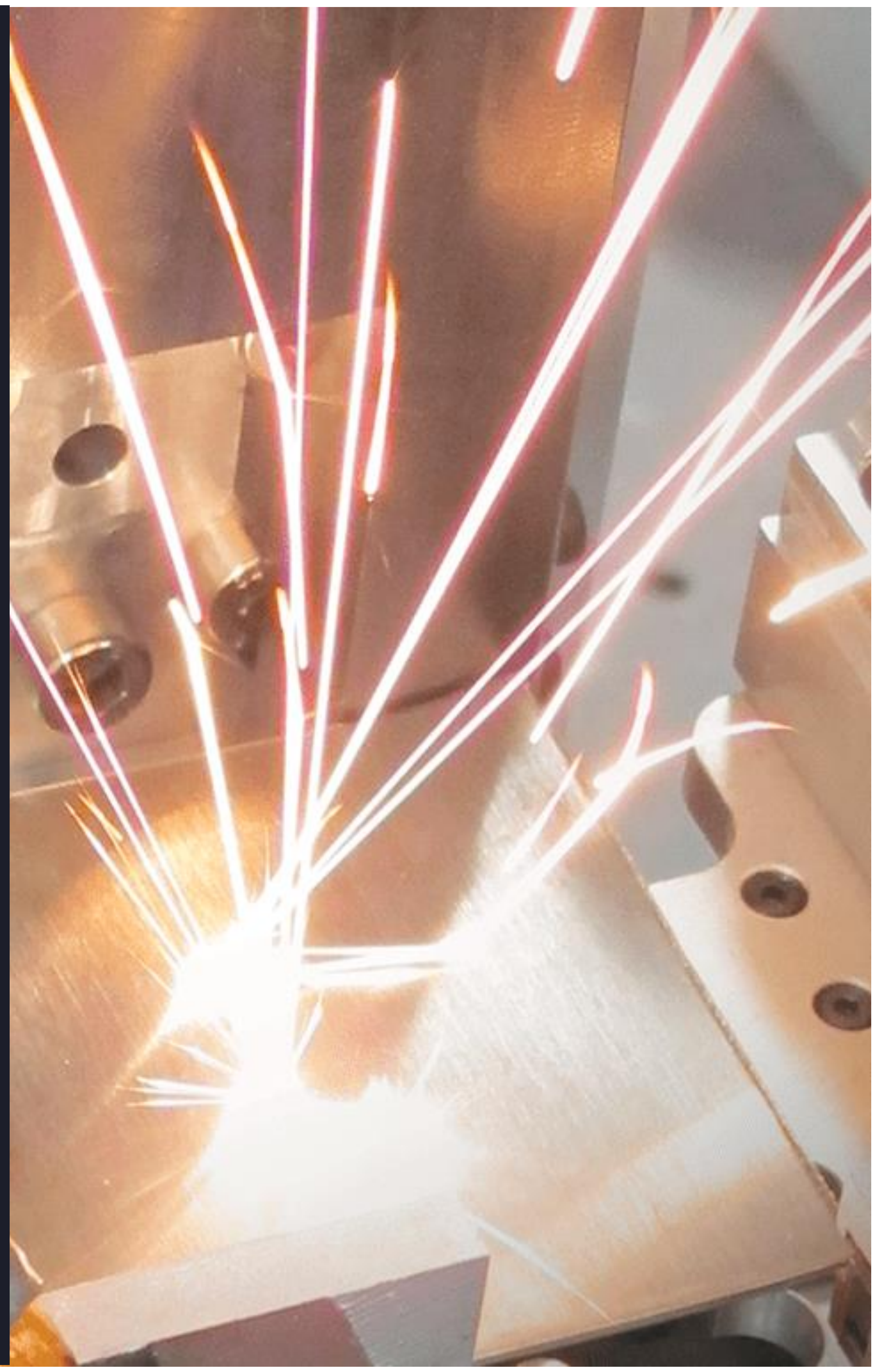
- **High speed**: Up to 500 000 pixels/second (e.g. 10m/s @ 10cm disctance)
- **High rate**: >1000 objects/second
- **Precision**: >99% **counting precision**
- Single object **tracking**
- Runs on **low-power mobile** processor

SPATTER MONITORING

Identify and track small particles (typ. size 10pixels) with fast motion in HDR environment.

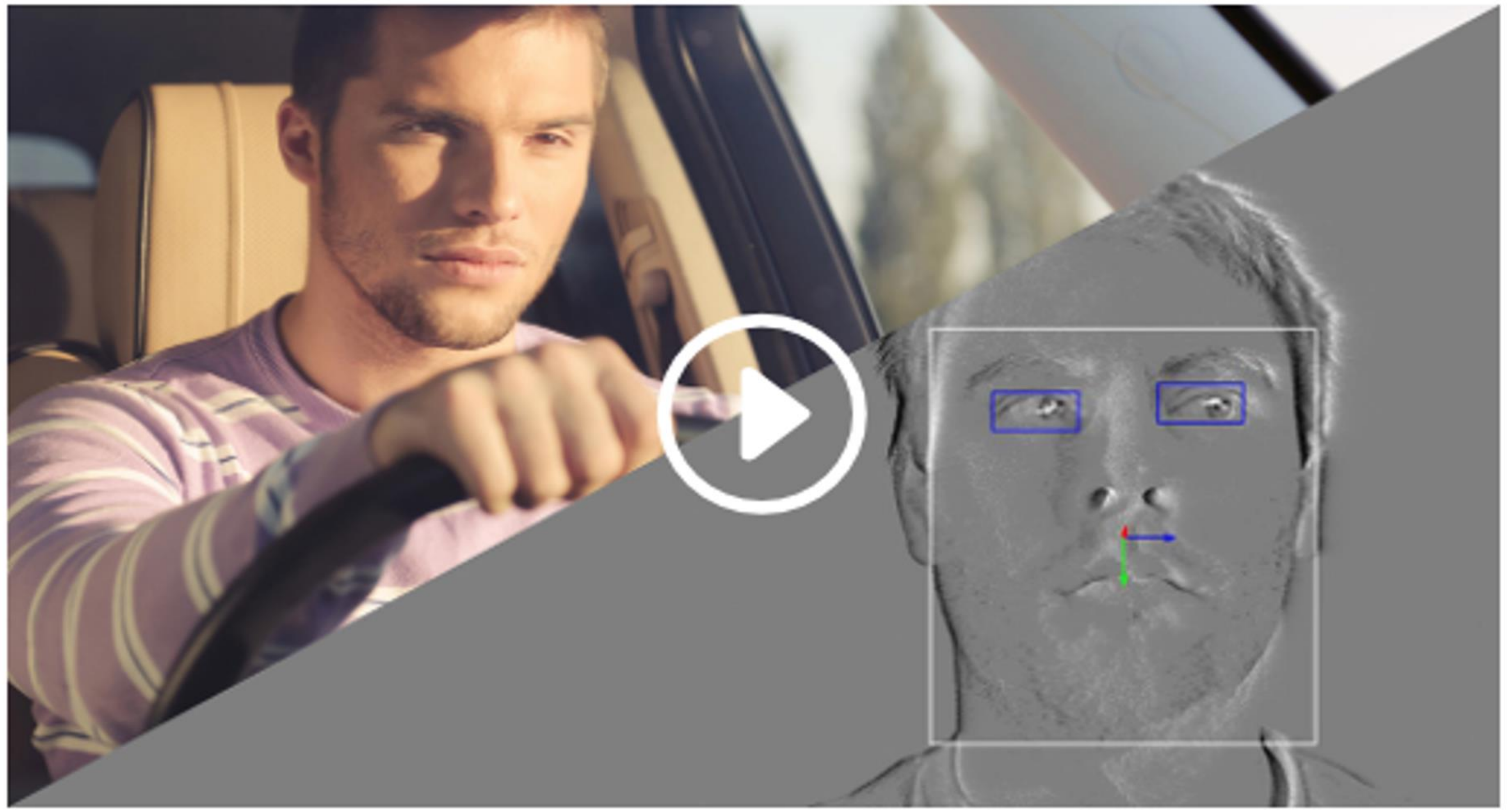
Exploits **high time resolution** and wide intrascene **dynamic range**

Up to 200k fps rendering (5µs time resolution)
Simulatenous XYT tracking of all particles



IN-CABIN / DRIVER MONITORING

XPERI | METAVISION® BY PROPHESEE



Exploiting the unique characteristics of Prophesee event sensing technology, Xperi developed the **world's first neuromorphic driver monitoring solution** (DMS)

- **Low light** – no active IR illuminator used
- **Multiple features:** Head pose, gaze, eye blink
- **Eye blink statistics:** Frequency and time signature (duration) for assessing state of driver alertness
- **Micro-expressions monitoring**

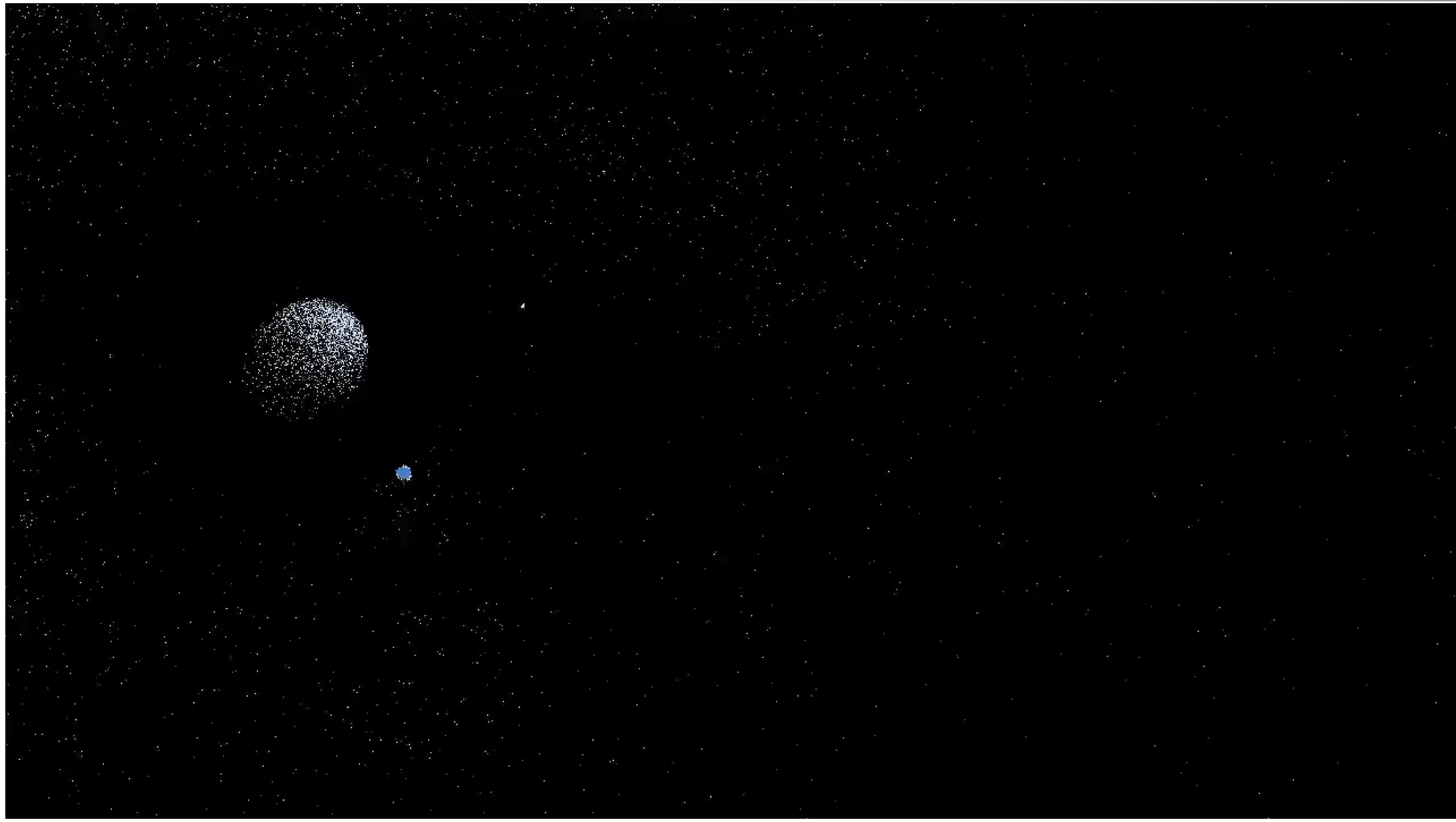
Courtesy of Xperi


EYE TRACKING
HAND TRACKING
CONSTELLATION TRACKING
METAVISION_XR
LOCALIZATION AND MAPPING
FOVEATED RENDERING
STRUCTURED LIGHT
FOVEA

EYE TRACKING

Ultra fast **eye tracking at >1kHz**

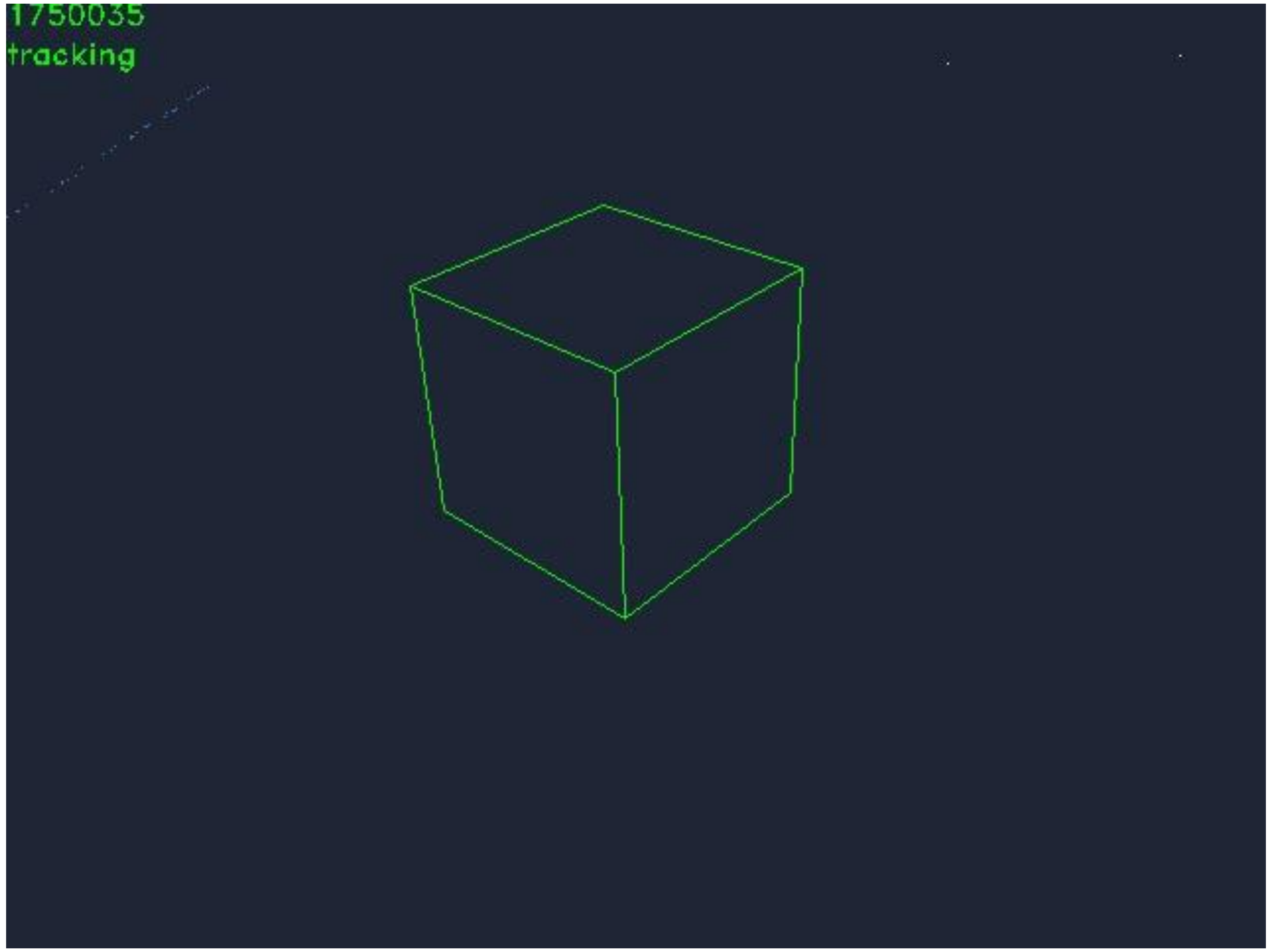
- **True eye contact** for realistic avatars
- **Non-verbal emotional communication**
- **Seamless navigation** in menus
- Foveated rendering to unlock next-gen XR system performance (resolution – fps – autonomy…)



INSIDE-OUT TRACKING

Accurate user location in space, without the need for base stations.

- High **robustness to light conditions**
- Ultra fast and smooth **tracking 10kHz**
- **Hand tracking**
- **Environment reconstruction**
- **Passive / active configurations** for SLAM / Visual odometry / Stereo / Structured light



M E T A V I S I O N _ X R

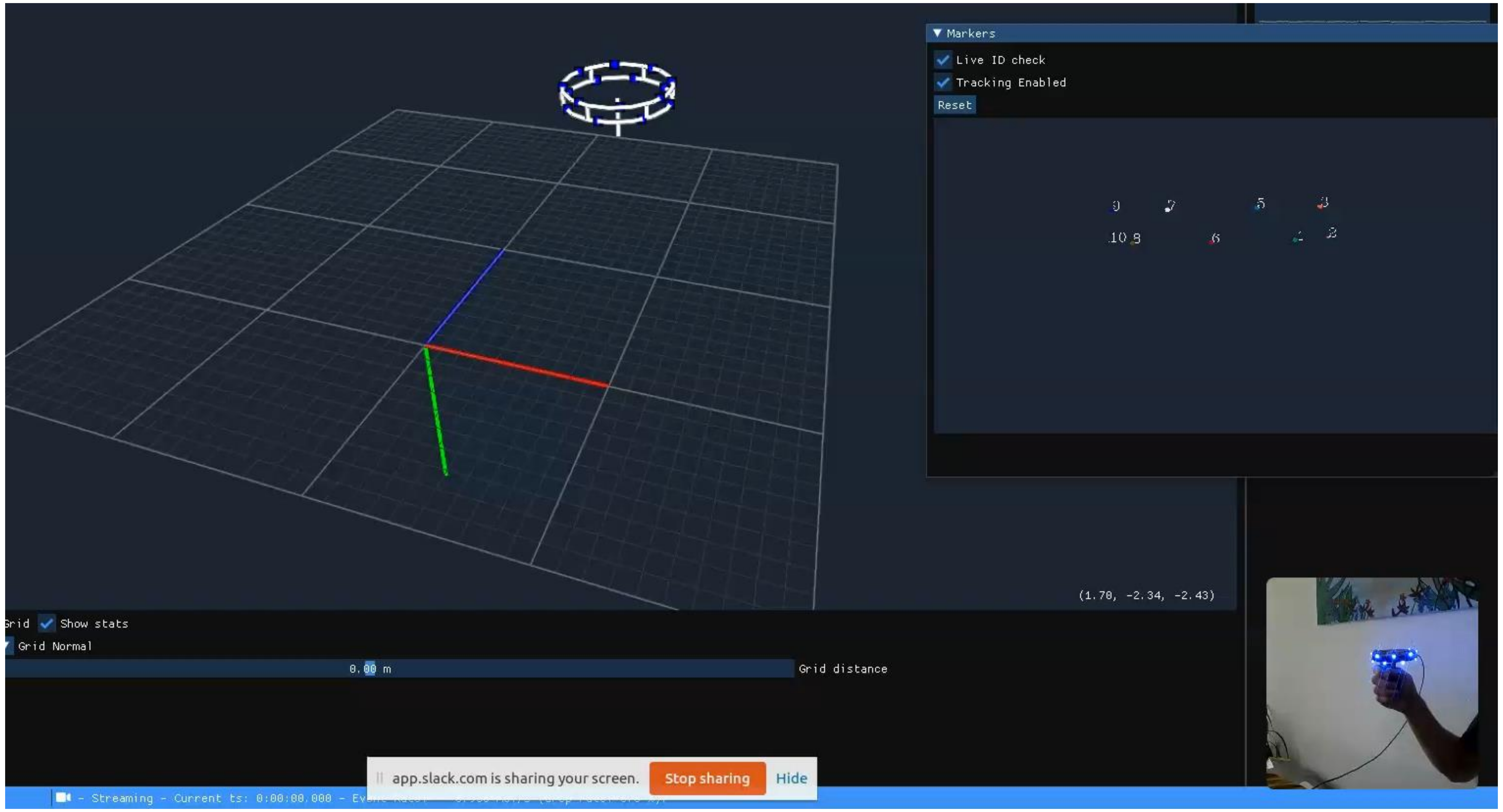
CONSTELLATION TRACKING

4 Metavision® sensors facing the scene for

Ultra-fast controler LED tracking

- **High precision LED tracking**
- Frequency analysis to **filter out parasite** flickering lights
- **High-speed frequency detection (kHz)** for high speed pose estimation

Achieve smoother controler trajectories computing and access advanced predictions capabilities.



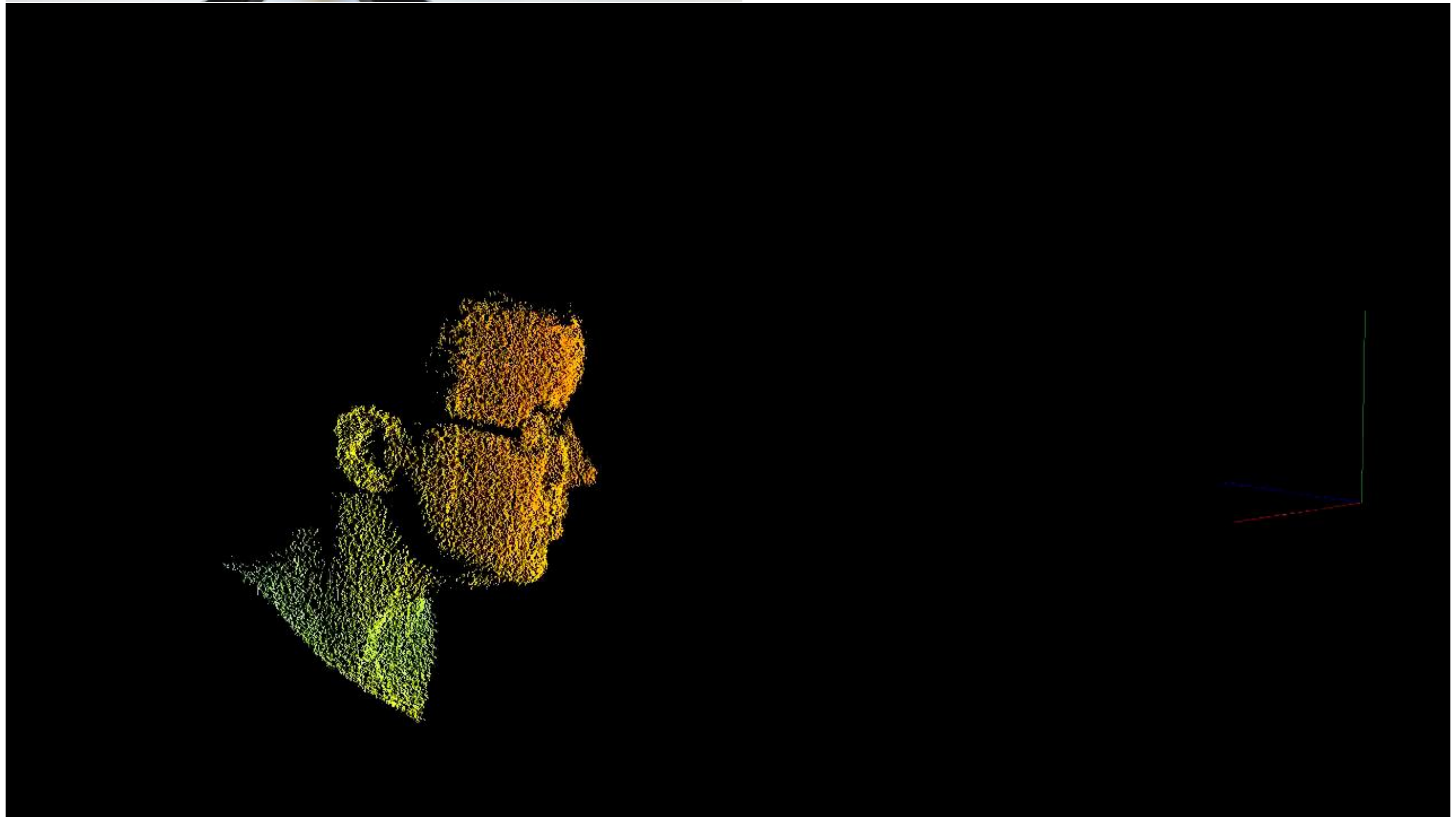
CONFIDENTIAL

EVENT-BASED STRUCTURED LIGHT

HIGH PRECISION
3D DEPTH SENSING
Event-based structured light

The high temporal resolution of event sensing allows to **uniquely encode every element** (line, point, ...) of the structured light pattern using **time-domain** modulation:

- **State of the art accuracy**
- **Up to 50x faster scanning times:** 1k point clouds/second
- **Processing complexity reduction** (matching is done not on frames but pixel by pixel)
- **Outdoor usage** (ultra-fast pulse detection enables to increase power while keeping eye-safety)



MACHINE LEARNING WITH EVENTS

MAIN BENEFITS FOR PROCESSING EVENTS WITH ML MODELS

ULTRA-LOW LATENCY

High temporal resolution allows lower latency detection
Inference at any rate is virtually possible
Only limited by computation time

EASIER GENERALIZATION

Light invariance allows for easier generalization
E.g. models trained at day light perform with night scenes

REDUCED COMPUTATION

Learn simpler patterns and features
No need to learn invariance to background (for static camera)

ULTRA-LOW LATENCY

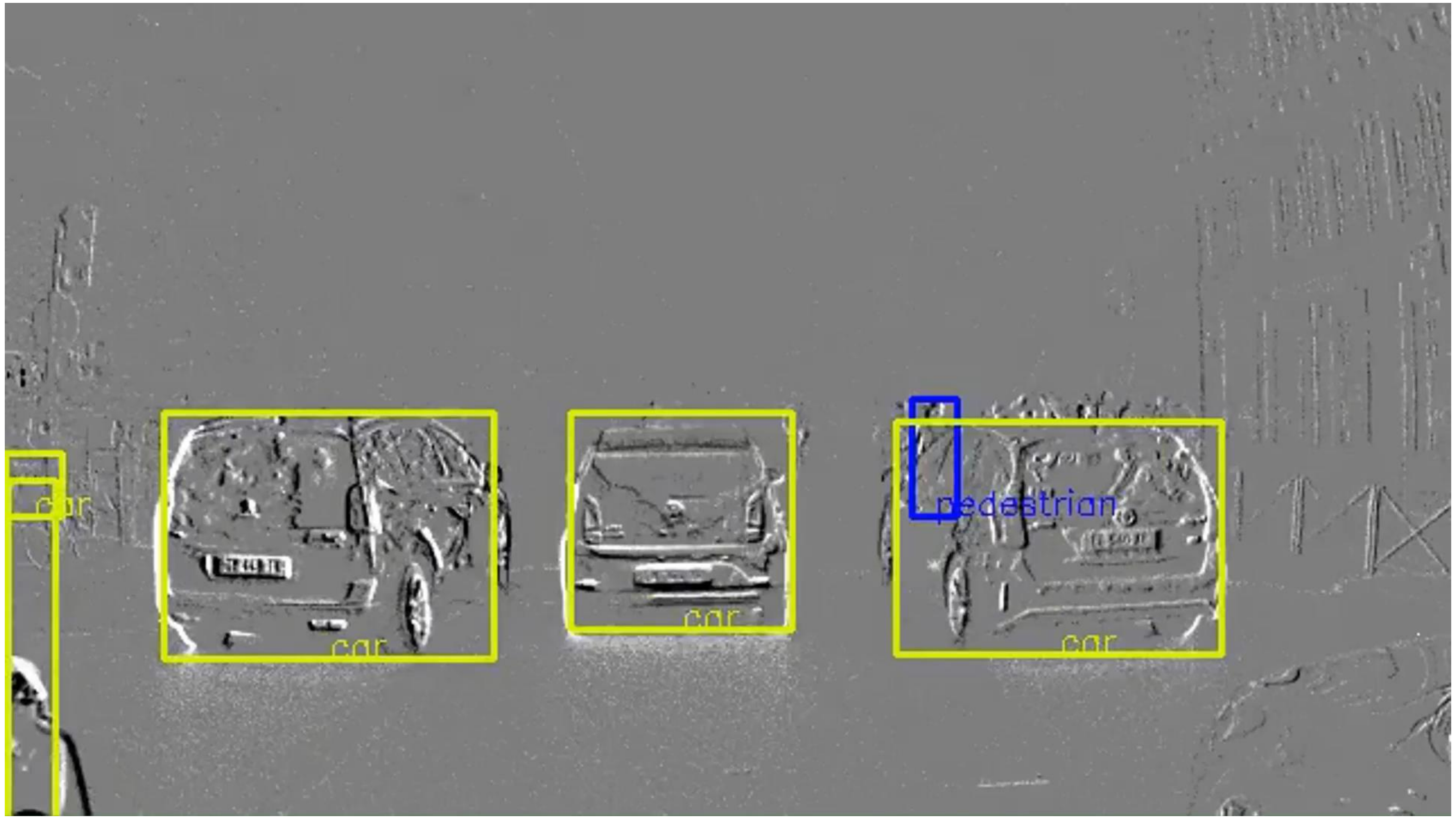
Event-based benefit #1:

Temporal Resolution

- Low-latency detection
- Inference at high rates

Event-Based automotive dataset:

- 7 classes (Car, truck, van, pedestrian, two-wheeler, …)
- 25M boxes for object detection and tracking
- HD 1280x720 event camera resolution



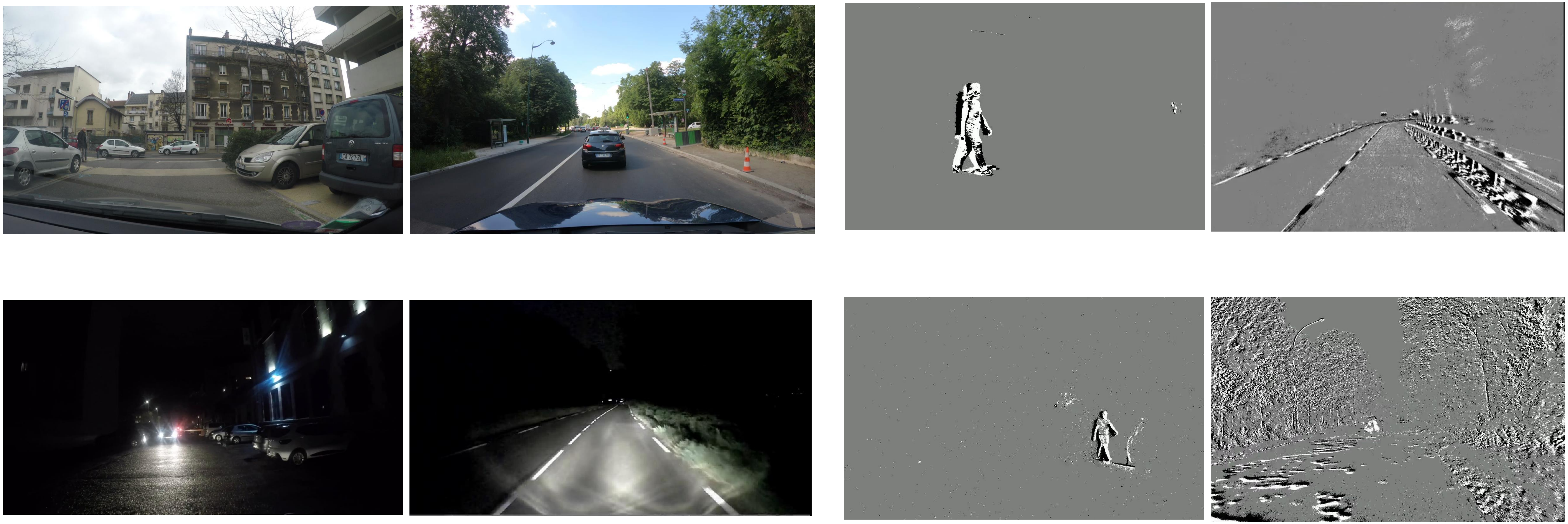
50Hz inference VGA sensor on mobile processor (cfr. Frame-based Mobilenet-v2 13Hz)

LIGHT INVARIANCE

Event-based benefit #2:

Light invariance

- Event sensors react to **relative changes**, independently of **absolute light levels**
- Light invariance allows for **easier generalization** of ML models



Day

Night

OBJECT DETECTION NIGHT

Light invariance

Inference on night data with network trained only with day data

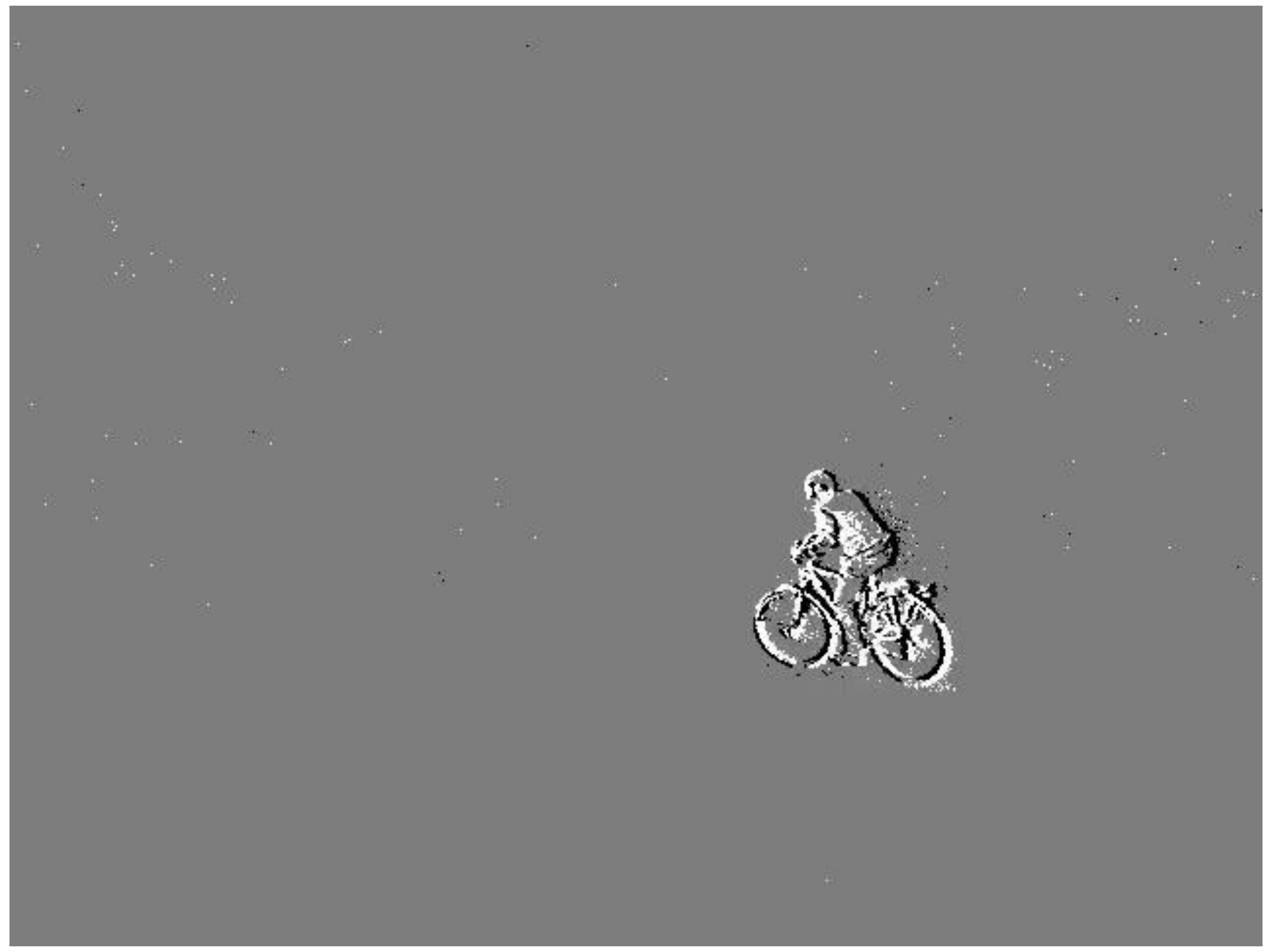
DATA SPARSITY

Event-based benefits #3:

Sparsity

- Sparse input allows for **reduced computation**
- Learn simpler patterns
- No need to learn invariance to background

Frames: Complex Background
Complex Texture



Events: Only relevant contrast features

MACHINE LEARNING WITH EVENTS - OBJECT DETECTION INFERENCE

- Pretrained automotive network model
- Trained on 15h / 25M labels automotive dataset
- Live detection and tracking @100Hz

FORWARD COLLISION WARNING

FRAME-BASED

TECH \ TRANSPORTATION \ CARS

68

Cars with high-tech safety systems are still really bad at not running people over

AAA brought receipts

By Andrew J. Hawkins | @andyjayhawk | Oct 4, 2019, 1:26pm EDT

SHARE

https://www.theverge.com/2019/10/4/20898773/aaa-study-automatic-emergency-braking-pedestrian-detection

« Vehicles struck the dummy pedestrians (…) 60 % of the time (…)in daylight hours at speeds of 20 mph. »

« With a child-sized version, the results got much, much worse: a collision occurred 89% of the time. »

« None of the cars tested were able to detect an adult pedestrian at night. »

EVENT-BASED

FRAMES + EVENTS

MOTION BLUR FREE IMAGES

BY SYNCHRONIZING A TRADITIONAL CAMERA WITH NEUROMORPHIC METAVISION® SENSOR

Pictures are sharper, clearer, more beautiful than ever, at any speed and in any lighting conditions

Your first shot is always the best one

PUSHING THE LIMITS OF PHOTOGRAPHY

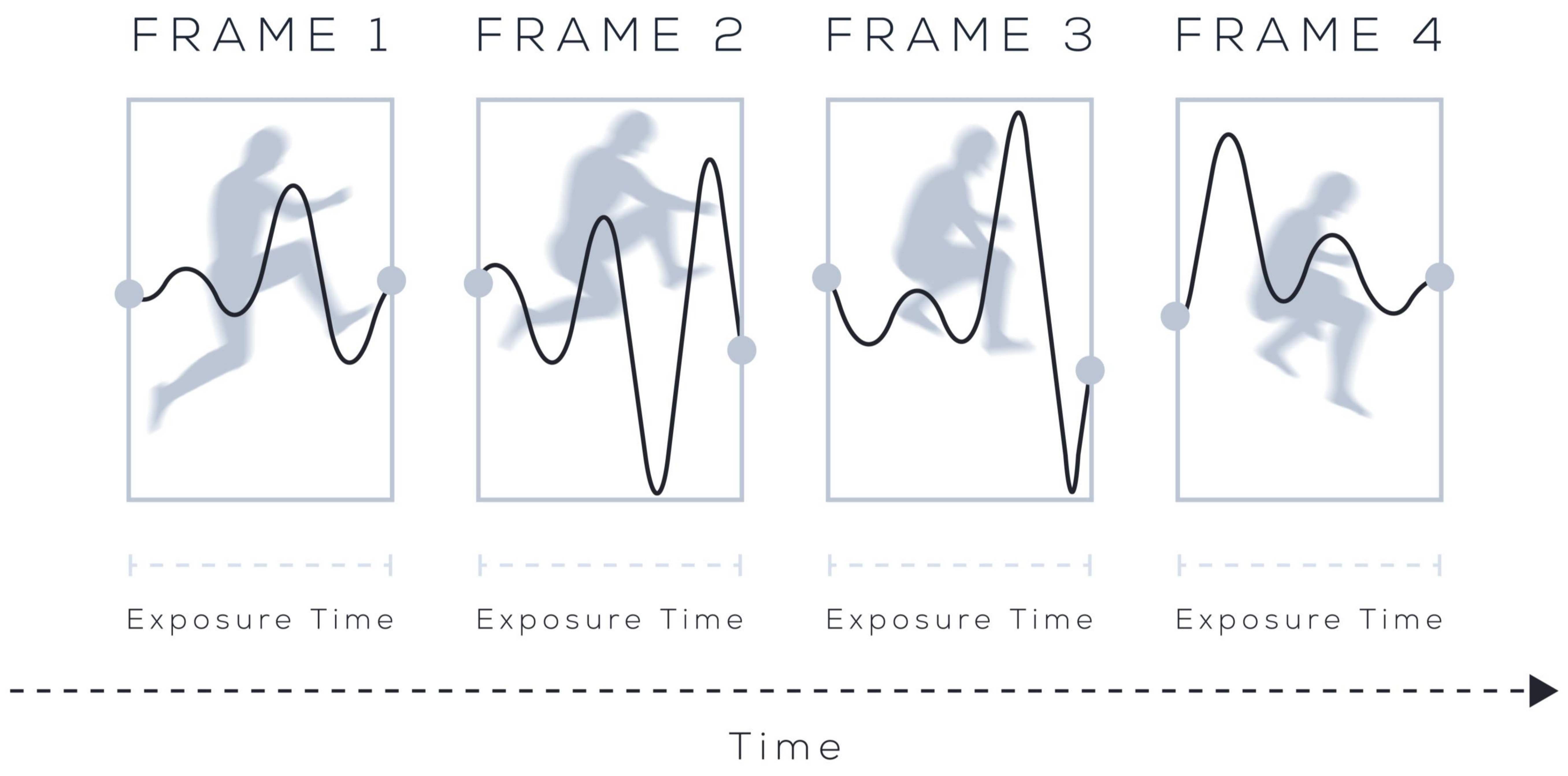
Image and video deblur

LIVE DEBLURRING

Using microsecond Events inside the frames

High-Performance Event-Based deblurring is achieved by **synchronizing a frame-based and an event-based sensor** on the same time base. This enables the system to **relate events to the exposure time of each frame.**

Results are achieved by focusing specifically on events happening **during the exposure time of each frame**. Using these events, algorithms can extract motion with **1 microsecond time resolution** as well as the **motion blur** associated to it.



VIDEO DEBLUR

MOBILE CAMERA

PROPHESEE

PUSHING THE LIMITS OF PHOTOGRAPHY

Still image deblur

Action / Sports

Low Light face + Text

Multi Depth

Night time / HDR


THANK YOU!
PROPHESEE
METAVISION FOR MACHINES