45th International Conference on Acoustics, Speech, and Signal Processing (ICASSP), May 4th-8th 2020

Single-Shot Real-Time Multiple-Path Time-of-Flight Depth Imaging for Multi-Aperture and Macro-Pixel Sensors

Miguel Heredia Conde[†], Keiichiro Kagawa[⋆], Tomoya Kokado[⋆],
Shoji Kawahito[⋆] and Otmar Loffeld[†]
[†]Center for Sensorsystems, University of Siegen, Paul-Bonatz-Str.9-11,57076 Siegen, Germany
[⋆]Research Institute of Electronics, Shizuoka University, Hamamatsu, Japan

Outline

1. Introduction
2. Sensing Model
3. Single-Shot ToF Cameras
4. Parametric Estimation from Fourier Samples
5. Experimental Results
6. Conclusion

1. Introduction

Depth Imaging

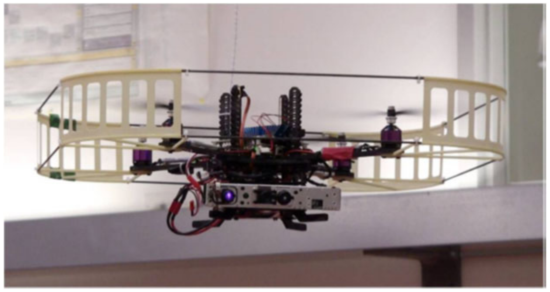
- **Depth sensing**: determining a 2D surface in a 3D space.
- **Methods** for *depth imaging*:
  - Laser Scanners
    - + High depth accuracy
    - - Mobile parts (rotary mirrors)
    - - Hard tradeoff between resolution and acquisition rate
  - Stereo Systems
    - + Passive, in presence of enough ambient light
    - - Bulkiness: at least two cameras and enough parallax
    - - Impossible to find correspondences in textureless scenes
    - - Parallax problem: hard tradeoff between large and small paralaxes
  - Light Coding Technology. Paradigmatic example: **Microsoft Kinect** (v1)

The Velodyne HDL-64E: A 64-channel LiDAR with 120m range, able to deliver up to 2.2×10^6 points per second with <2cm accuracy.

Karmin2 stereo cameras from Nerian Vision Technologies, with baselines of 10 and 25cm.

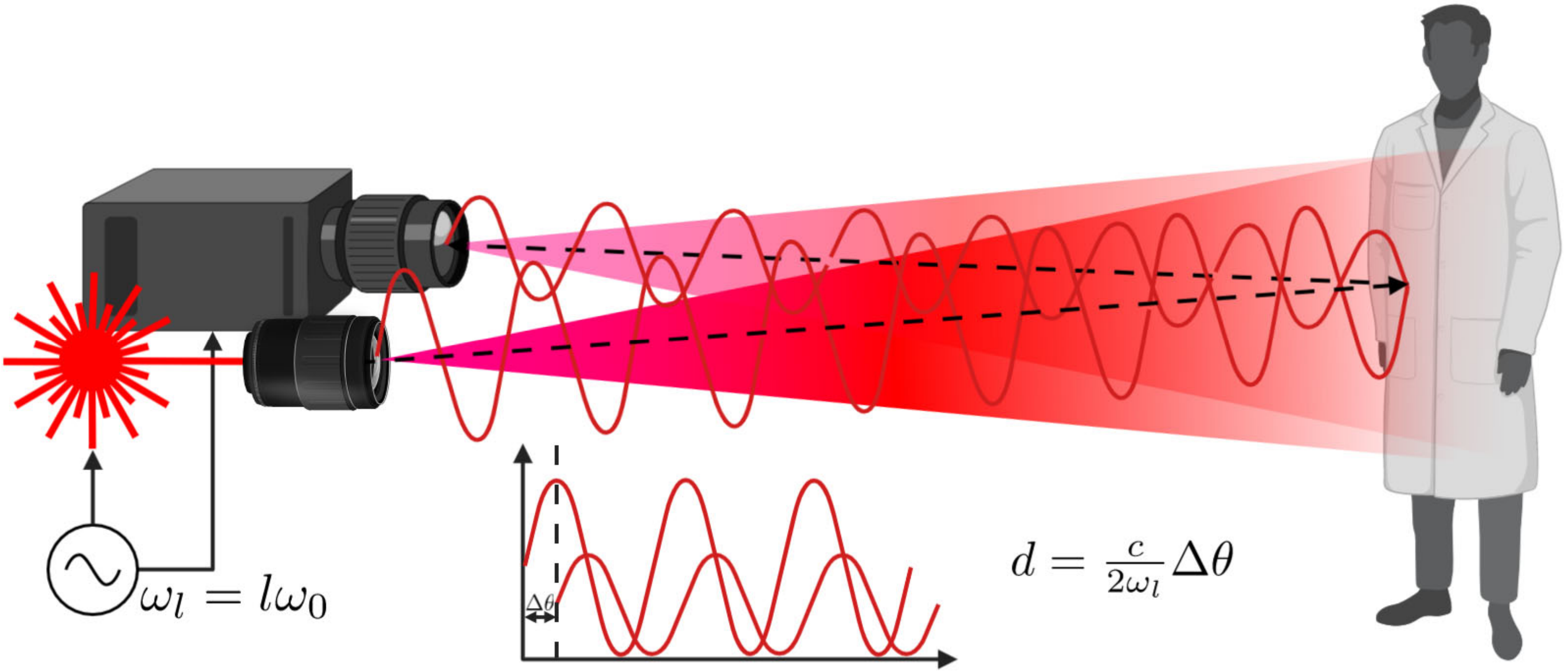
Microsoft Kinect (v1) sensor, featuring an RGB camera and a pair NIR-pattern emitter and NIR camera for depth sensing.

The PR2, from Willow Garage, features a stereo camera pair in its head. Additionally a Kinect (v1) was mounted on top. Image taken at the AIS Laboratory of the Albert-Ludwigs-Universität Freiburg.



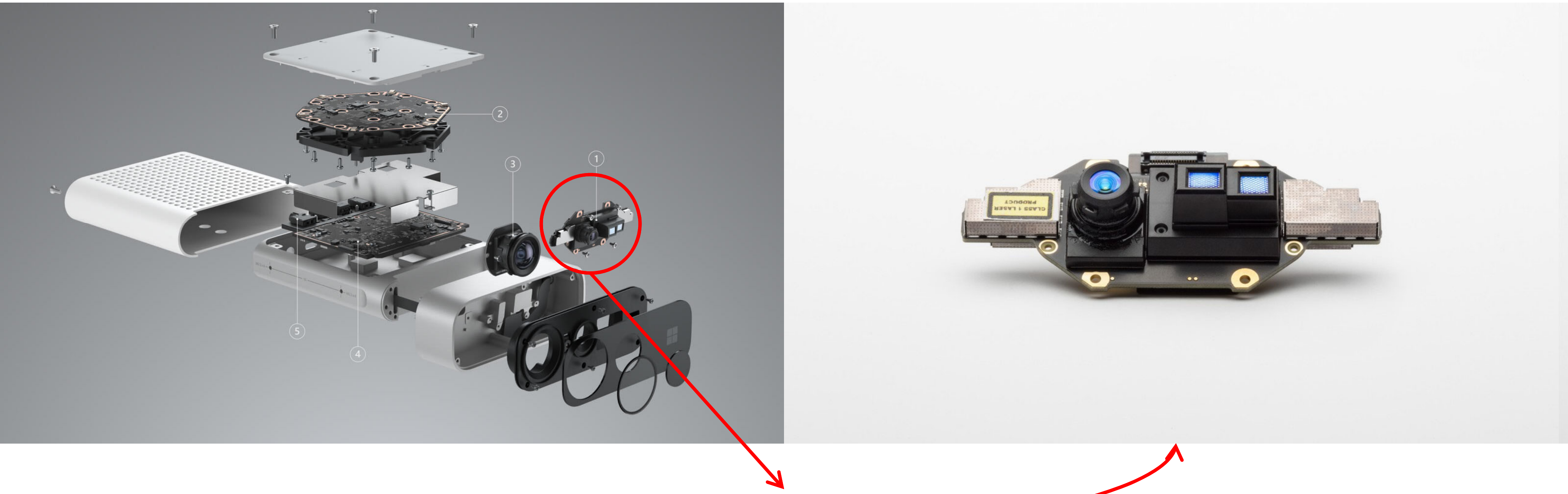
Quadcopter with a Kinect sensor mounted on it, used to perform visual odometry and mapping. Courtesy of Albert S. Huang.

Continuous Wave Time-of-Flight Imaging



Commercial CW-ToF Camera Technologies

- Microsoft Kinect (latest release: Azure)

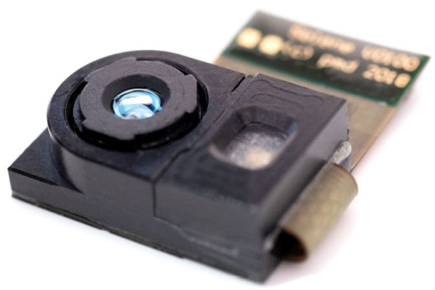


ToF Module

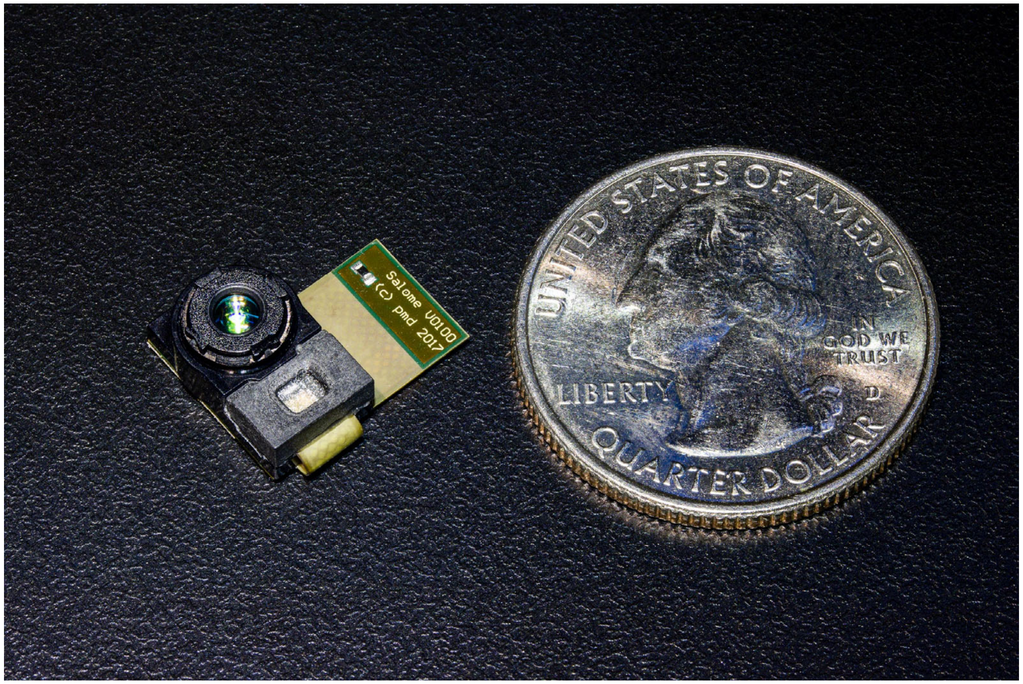
Pixel bandwidth: 320MHz

Commercial CW-ToF Cameras

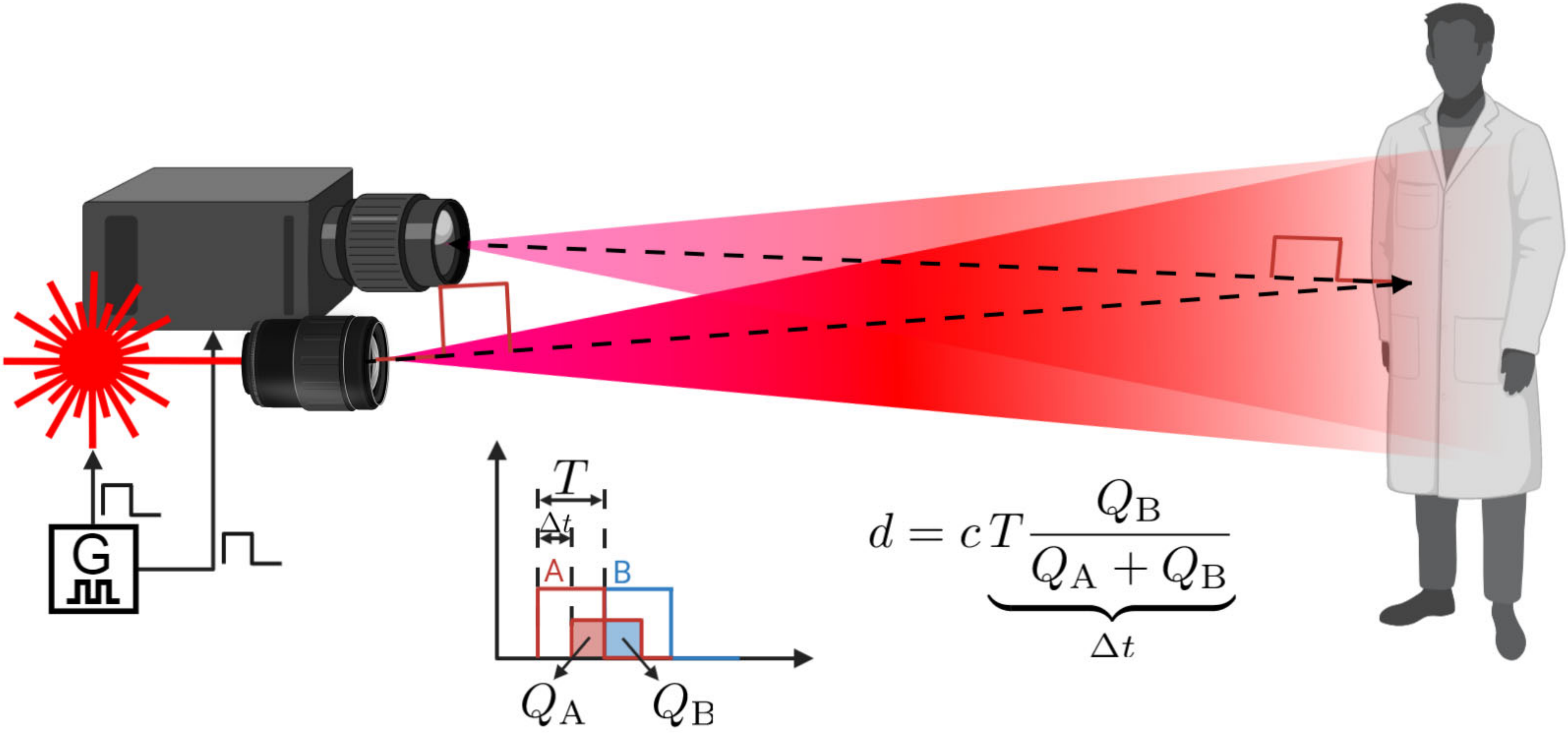
- Photonic Mixer Device (PMD). Selene module from pmdtechnologies ag:



Tested with up to 160MHz



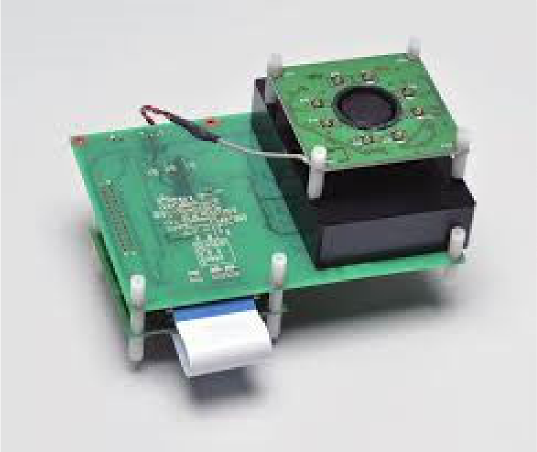
Pulsed Time-of-Flight Imaging



$$d = cT\underbrace{\frac{Q_B}{Q_A+Q_B}}_{\Delta t}$$

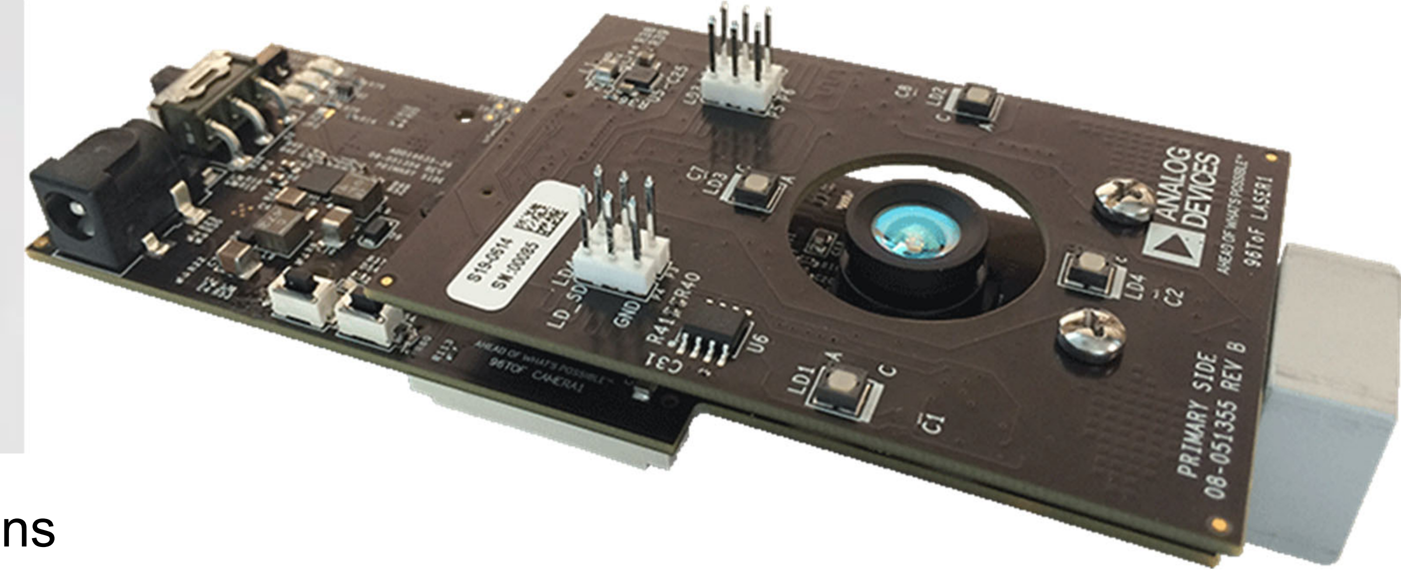
Commercial Pulsed CW-ToF Cameras

- Hamamatsu S11963-01CR:



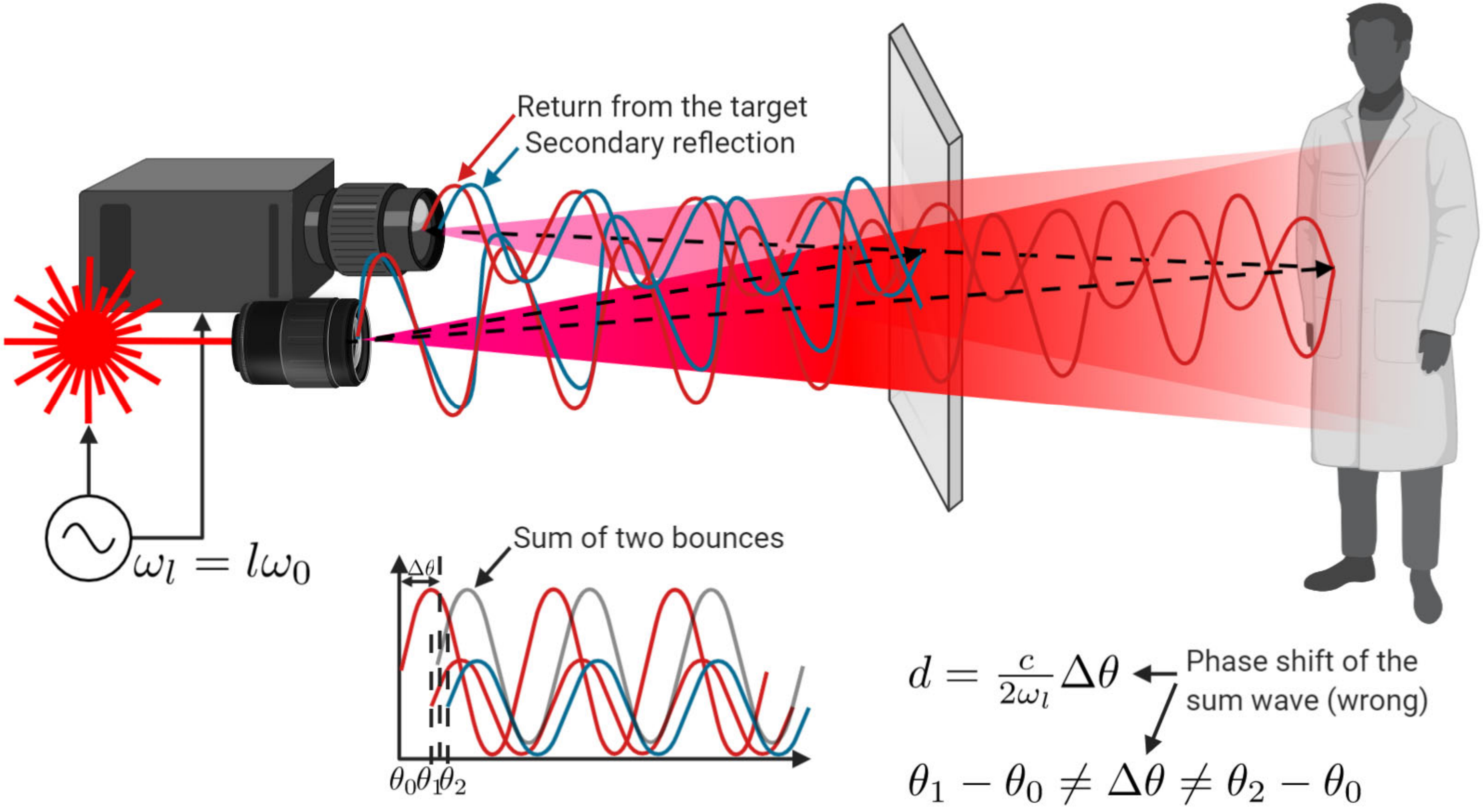
Max. Pulse width: T=50ns

- Analog Devices AD-96TOF1-EBZ:



Pulse width: T=22ns

Multiple-Path Interference (MPI):



$$\omega_l = l\omega_0$$

$$d = \frac{c}{2\omega_l}\Delta\theta$$

$$\theta_1 - \theta_0 \neq \Delta\theta \neq \theta_2 - \theta_0$$

How to Resolve Multiple Paths per Pixel?

- CW-ToF:
  - Interference of **several sinusoids** is also a **sinusoid**. Impossible with monotone CW-ToF.
  - The scene should be probed at different frequencies.
- Pulsed-ToF:
  - Acquisitions at **different time shifts** between Illumination Control Signal (ICS) and Demodulation Control Signals (DCS) are required.
- <u>The Challenge</u>: how to acquire more raw images within the same acquisition time?

2. Sensing Model

- In the case of reflective MPI, the scene response function is of the shape:

$$h(t) = \sum_{k=0}^{K-1} \Gamma_k \delta(t - t_k), \qquad t_k = \frac{2d_k}{c}$$

where t_k is the delay undergone by the k^{th} reflection, $k \in [0, K-1]$ and Γ_k is the corresponding attenuation factor.

- Let $i(t)$ be the illumination signal. Then the signal $r(t)$ received at the ToF pixel is given by the convolution:

$$r(t) = i * h(t)$$

- If $Q > 1$ raw images are to be acquired using Q different DCS $p_q(t)$, $1 \leq q \leq Q$, then the measurements are given by the cross-correlation:

$$m_q(t) = p_q \otimes\ r(t) = p_q \otimes (i * h)(t) = \left(i \otimes p_q\right) * h(t)$$

- In other words, we sample the convolution between the scene response function and several sensing functions

$$s_q(t) := \left(i \otimes p_q\right)(t)$$

- In conventional ToF, $Q = 1$ and measurements at different phase shifts are acquired. We **focus on $\boldsymbol{Q > 1}$**.

- Differently from prior work, we aim for a *single shot* camera, thus a **single measurement** per (sub-)pixel will be acquired:

$$m[q] := m_q(t_0), \qquad t_0 = 0, \qquad 1 \leq q \leq Q$$
$$= \int_{-\infty}^{\infty} s_q(t) h^*(-t) dt = \langle s_q(t), h(-t) \rangle$$

- Let $\vec{s_q}$ and $\vec{h}$ denote discrete versions of $s_q(t)$ and $h(t)$ of size n, then we have the linear model

$$\boxed{\vec{m} = \boldsymbol{S}\vec{h}}$$

where $\vec{m} := [m(q)]_{q=1}^{Q}$ and the fat matrix $\boldsymbol{S}$ of size $Q \times n$ is obtained from the vectors $\vec{s_q}$, $1 \leq q \leq Q$.

3. Single-Shot ToF Cameras

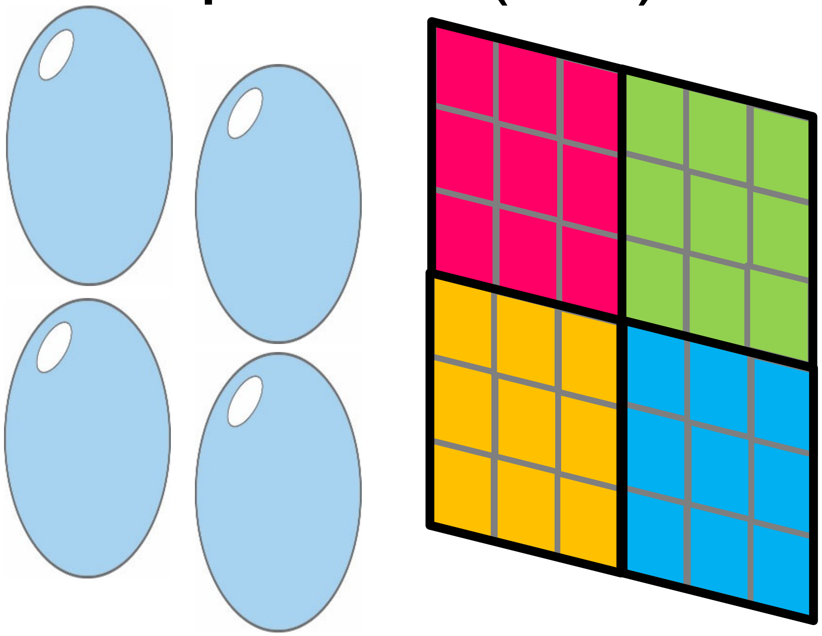
How can we measure according to Q different sensing functions?

- Multiplex in time domain (sequential acquisition)
  - **Problem:** linear growth of acquisition time
- Multiplex in spatial domain. Our alternatives:

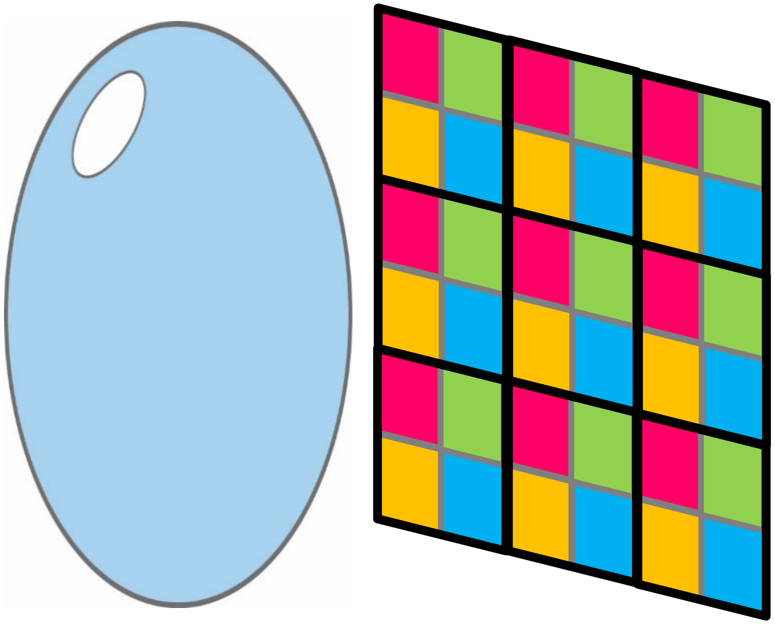
1. **Multi-Aperture** Ultra-High-Speed (MAUHS) CMOS Image Sensor (CIS)

2. Multi-tap **Macro-Pixel**-based Ultra-High-Speed CIS

Multi-Aperture (MA)

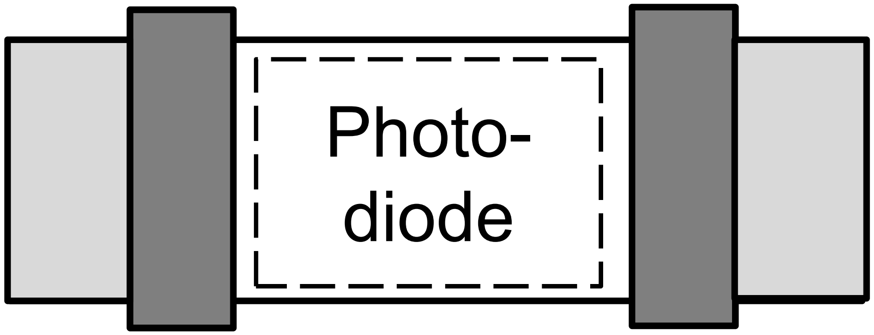


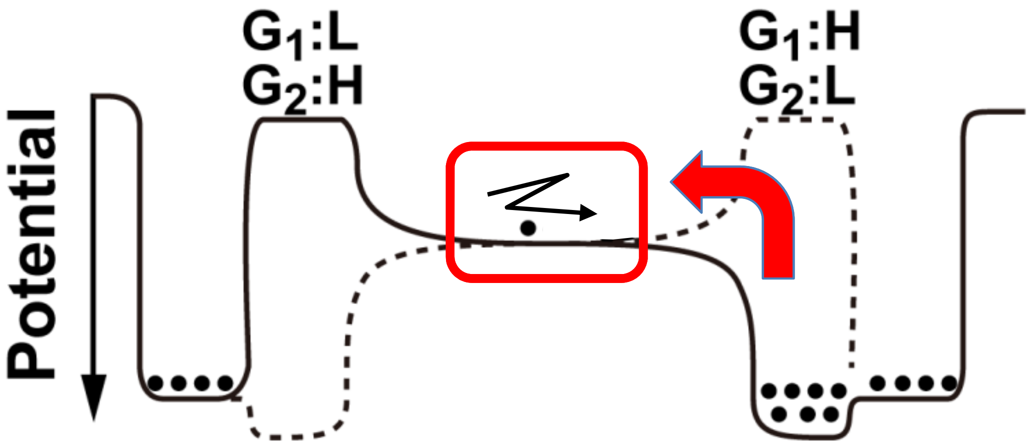
Macro-Pixel (MP)



	Multi-aperture	Macro-pixel
Shutter	Per aperture	Per subpixel
Disparity	Exists	-
Lens	Special lens array	Ordinary lenses

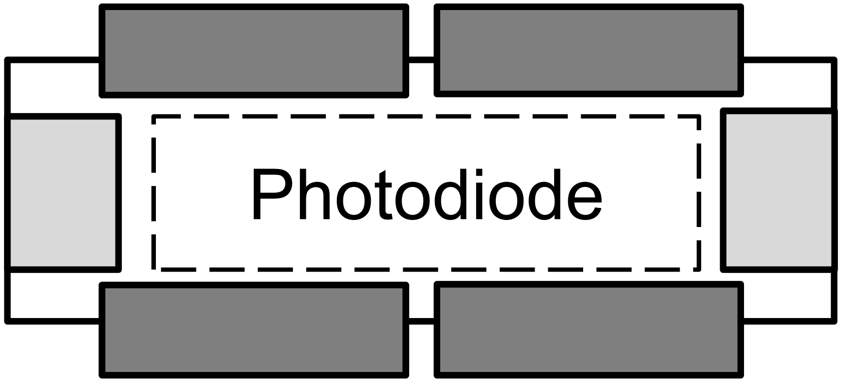
Conventional CMOS pixel

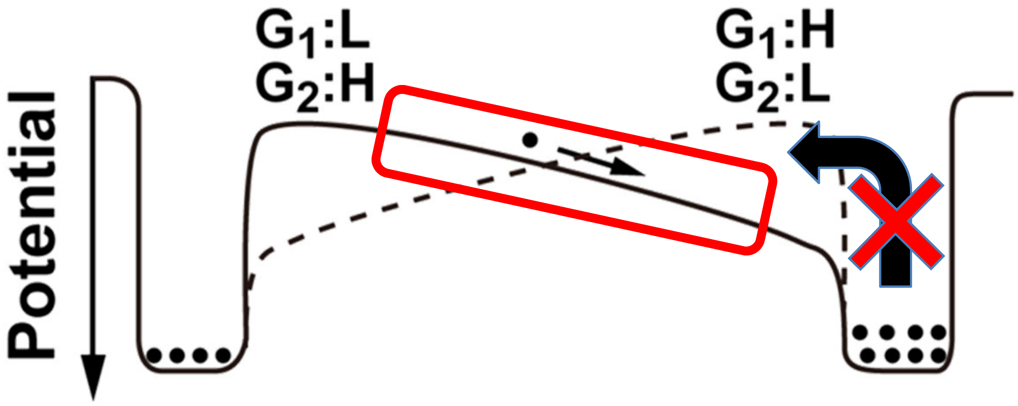




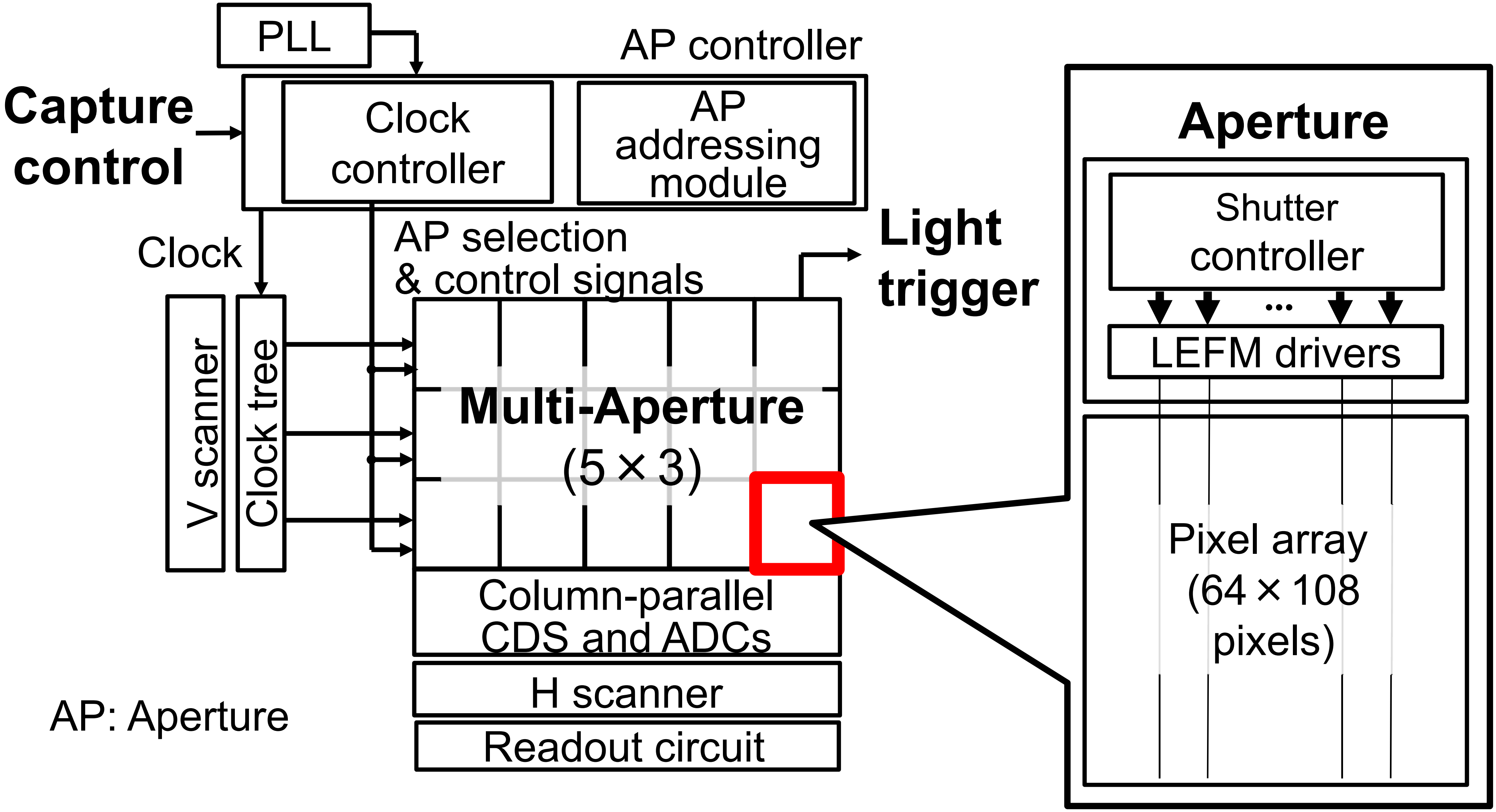
- Slow charge transfer
- Return charges

Lateral electric-field charge modulator (LEFM)





- Fast charge transfer
- No return charges



PLL

AP controller

Capture control

Clock controller

AP addressing module

Clock

AP selection & control signals

Light trigger

V scanner

Clock tree

Multi-Aperture
(5 × 3)

Column-parallel CDS and ADCs

H scanner

Readout circuit

AP: Aperture

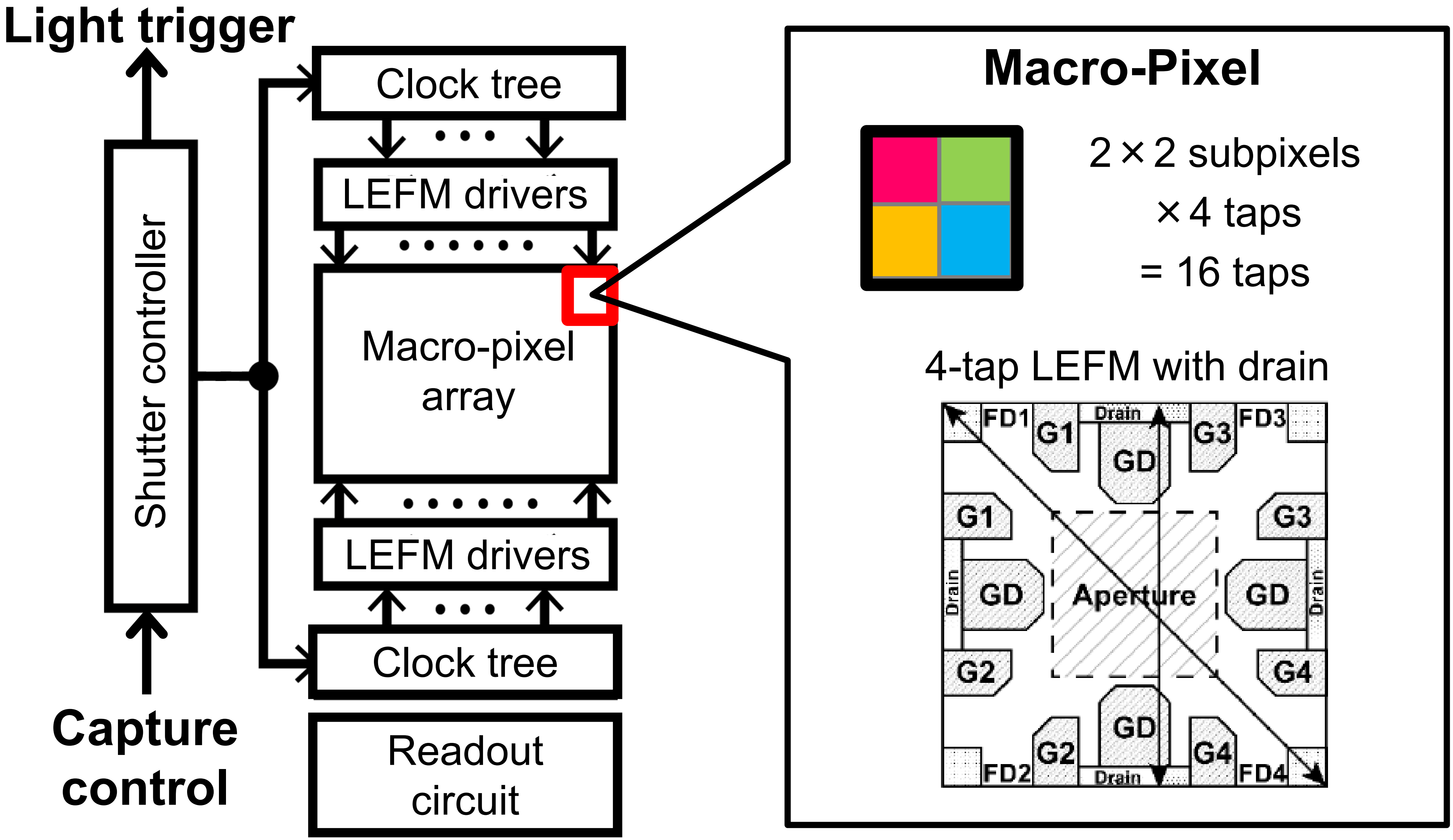
Aperture

Shutter controller

…

LEFM drivers

Pixel array
(64 × 108 pixels)



Light trigger

Shutter controller

Capture control

Clock tree

LEFM drivers

Macro-pixel array

LEFM drivers

Clock tree

Readout circuit

Macro-Pixel

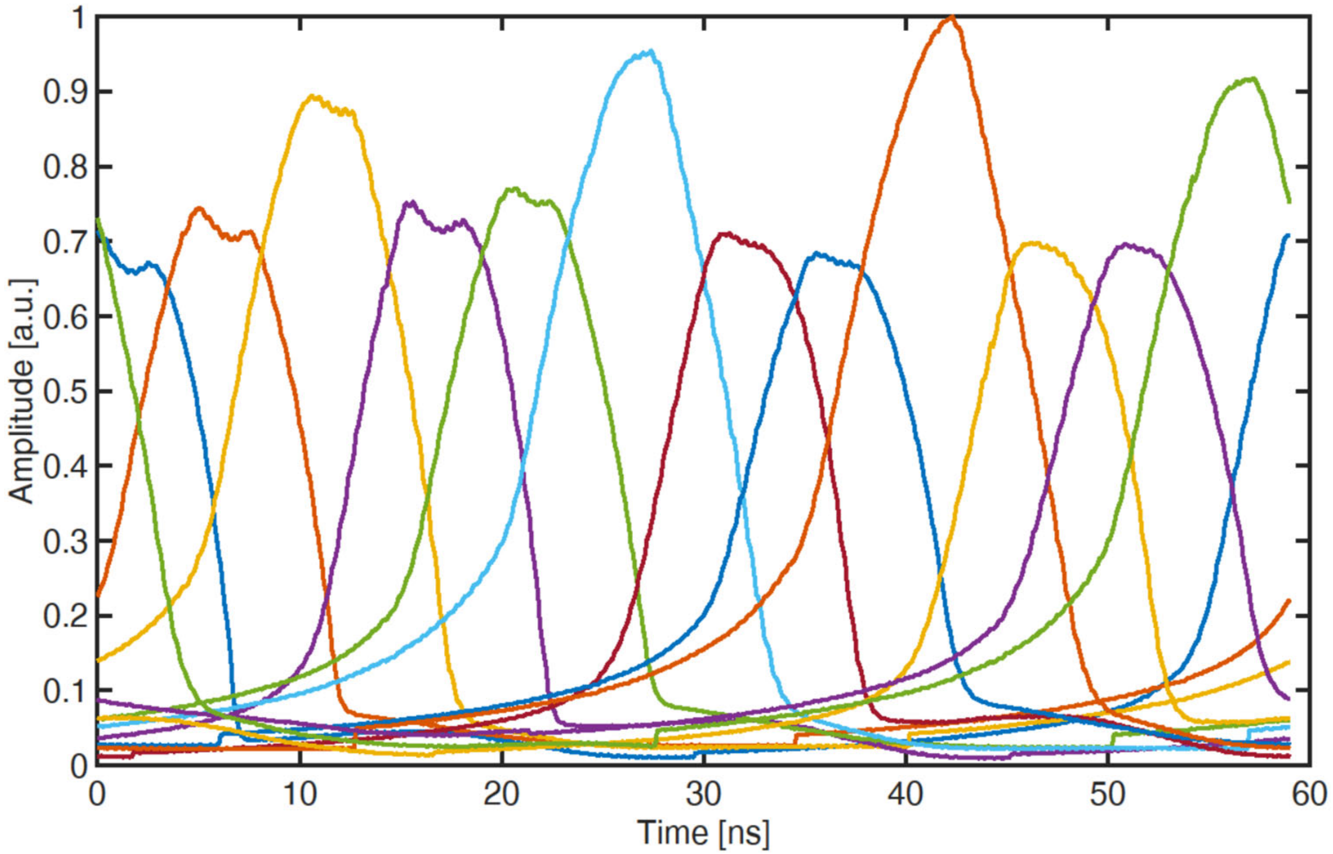
2 × 2 subpixels

× 4 taps

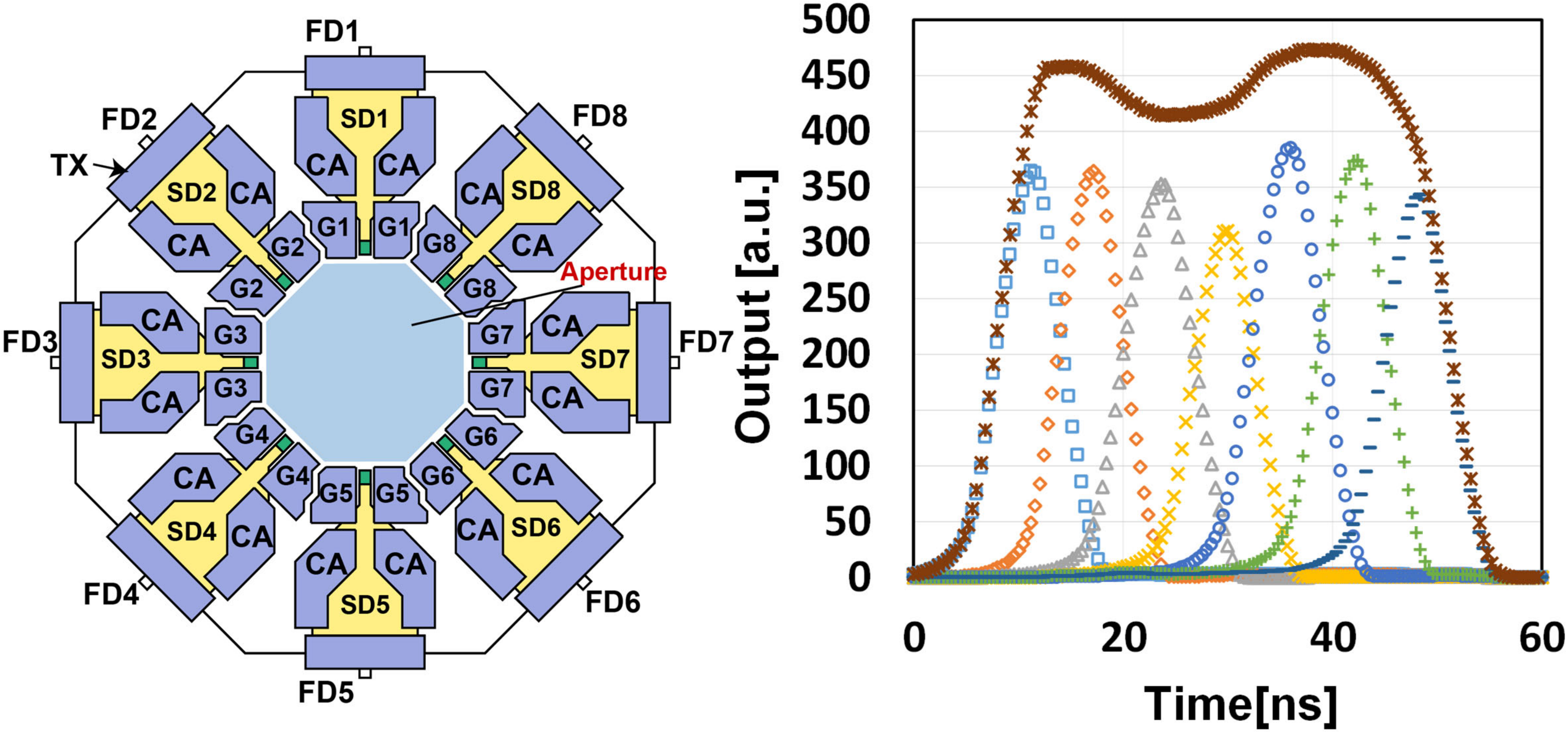
= 16 taps

4-tap LEFM with drain

How do the sensing functions of MP pixels look like?



Empirical sensing functions $s_q(t)$ of 12 of the $Q = 16$ available taps of our MP ToF sensor

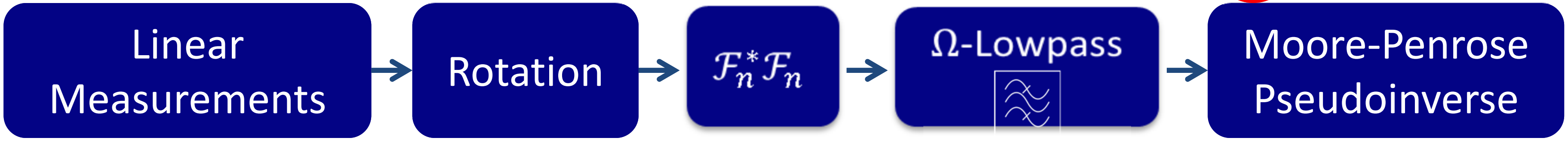


Y. Shirakawa *et al.*, MDPI Sensors **20**, Article 1040 (2020).

4. Parametric Estimation from Fourier Samples

- How to Extract the Fourier Samples?

$\enclose{circle}{\vec{m}} = \boldsymbol{S}\vec{h}$ $\quad$ $\enclose{circle}{\vec{y}} = \boldsymbol{A}\vec{h}$ $\quad$ $\vec{y} = \boldsymbol{B}^*\vec{\mathcal{H}}$ $\quad$ $\vec{y} \approx [\boldsymbol{B}^*]_{:,\Omega}\vec{\mathcal{H}}_\Omega$ $\quad$ $\enclose{circle}{\vec{\mathcal{H}}_\Omega} \approx \{[\boldsymbol{B}^*]_{:,\Omega}\}^\dagger \vec{y}$



Linear Measurements → Rotation → $\mathcal{F}_n^*\mathcal{F}_n$ → Ω-Lowpass → Moore-Penrose Pseudoinverse

$\boldsymbol{A} = \boldsymbol{RS}$ $\quad$ $\boldsymbol{B} = \mathcal{F}_n\, \boldsymbol{A}^*$ $\quad$ $|\Omega| \le Q$

$\boldsymbol{R}$ orthonormal

- Parametric estimation from Fourier samples:
  - From the sparse scene response model we have:

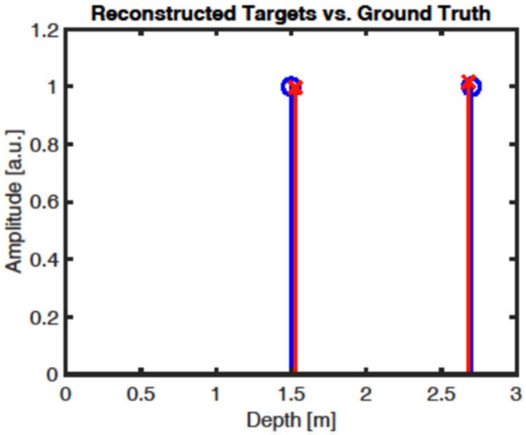
$$\mathcal{H}_l = \sum_{k=1}^{K-1} \Gamma_k e^{il\omega_0 t_k}$$

We use a robust variant of **Prony's method** to obtain $\{\Gamma_k, t_k\}_{k=1}^{K}$ from $\vec{\mathcal{H}}_\Omega$ in a closed form.

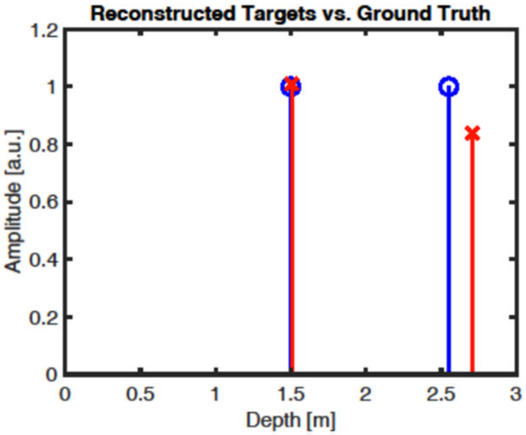
that is, the elements of $\vec{\mathcal{H}}_\Omega$ are samples of a *sum of sinusoids*, and the problem of sparse estimation in a high-dimensional domain boils down to estimating the frequencies of a sum of K sinusoids given $N \geq 2K + 1$ samples. → Classical Spectral Estimation!

5. Experimental Results

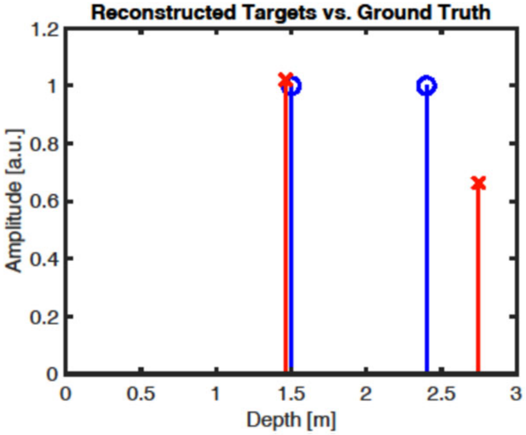
Synthetic Experiments with Real Sensing Functions:



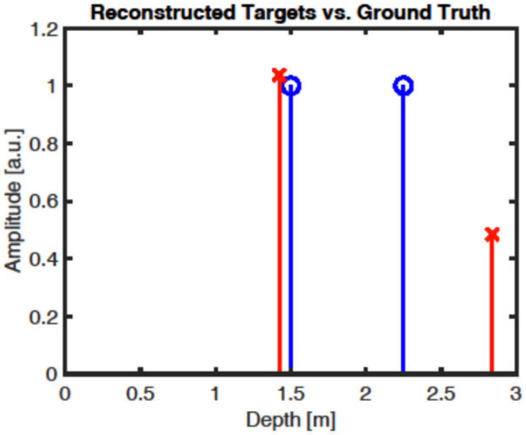
$\Delta d = 1.20\text{m}$



$\Delta d = 1.05\text{m}$



$\Delta d = 0.90\text{m}$

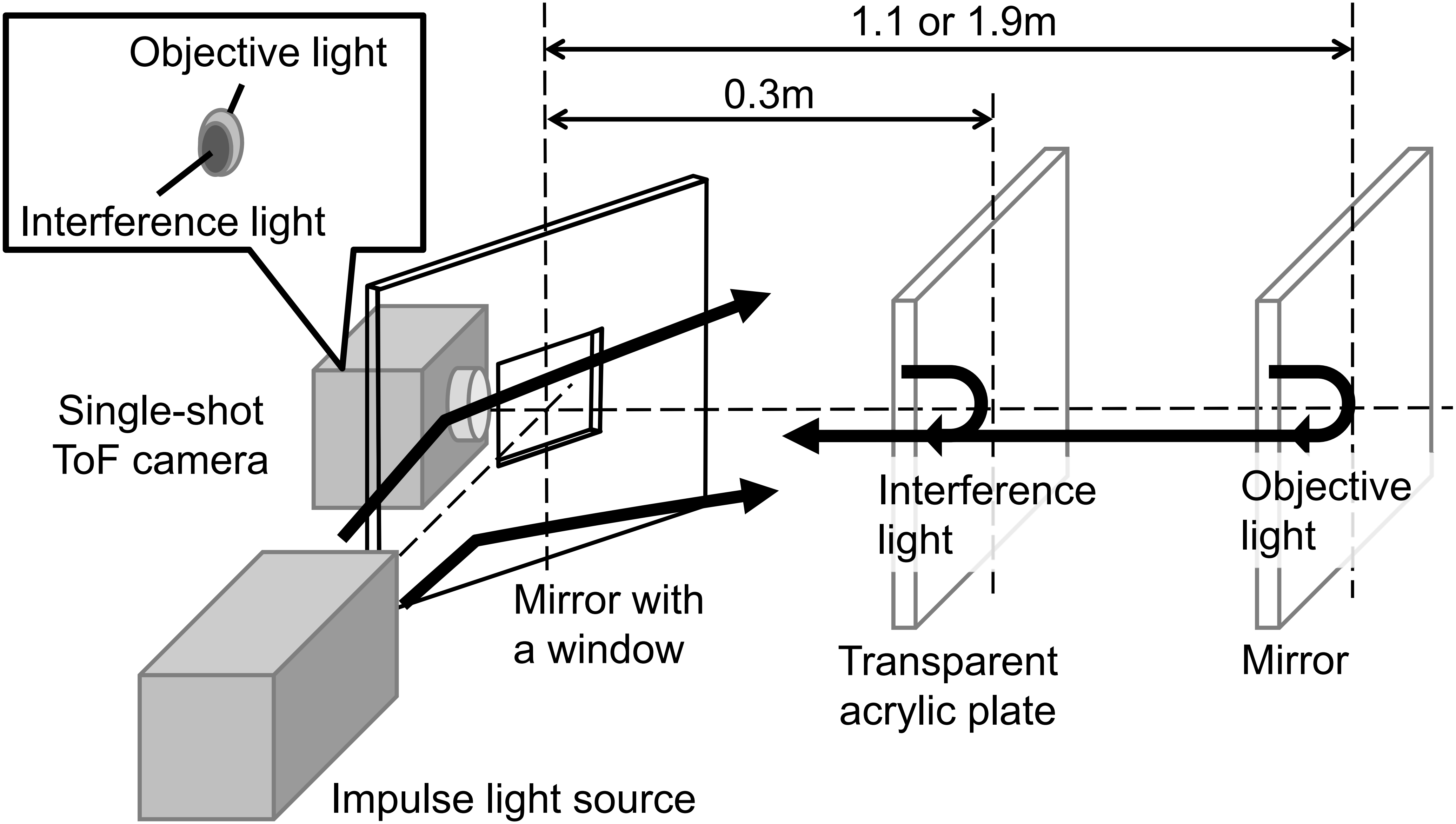


$\Delta d = 0.75\text{m}$

$Q = 16$, $T = 141.6\text{ns}$, Step size: $8.85\text{ns} \rightarrow 1.33\text{m}$ resolution.

With our parametric estimation approach, we observe target separation failure for $\Delta \leq 0.60\text{m}$.

For a single target, reconstruction is exact.



Objective light

Interference light

Single-shot ToF camera

1.1 or 1.9m

0.3m

Mirror with a window

Interference light

Objective light

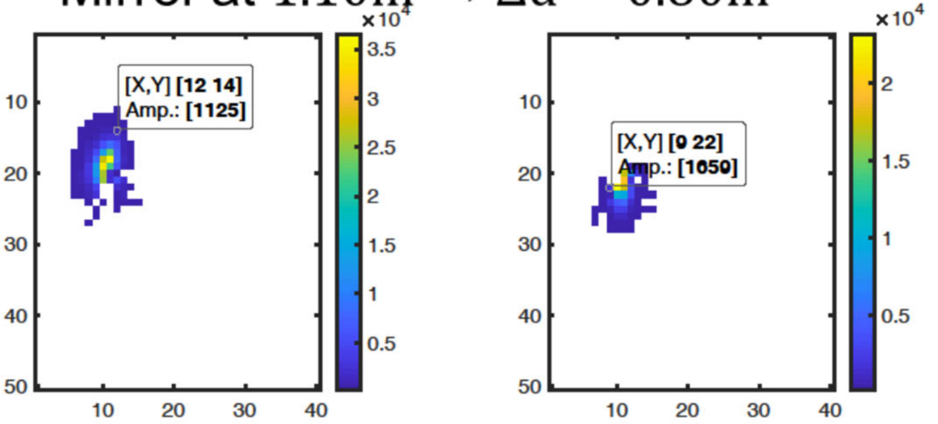
Transparent acrylic plate

Mirror

Impulse light source

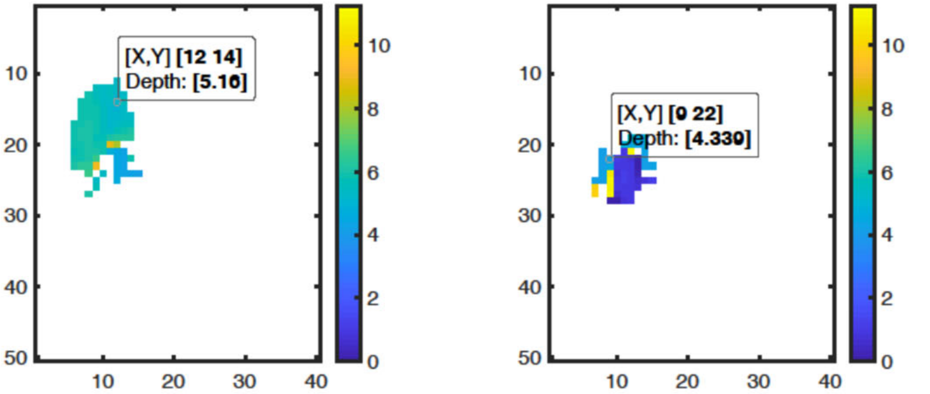
Real Experiments: Multi-Aperture (MA) Sensor, $Q = 15$:

Mirror at 1.10m → $\Delta \mathrm{d} = 0.80\mathrm{m}$



Amplitude, Path 1

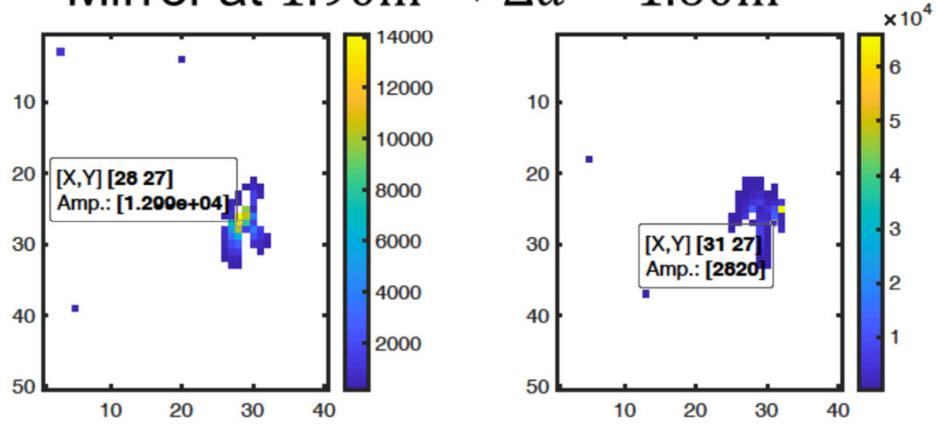
Amplitude, Path 2



Depth, Path 1

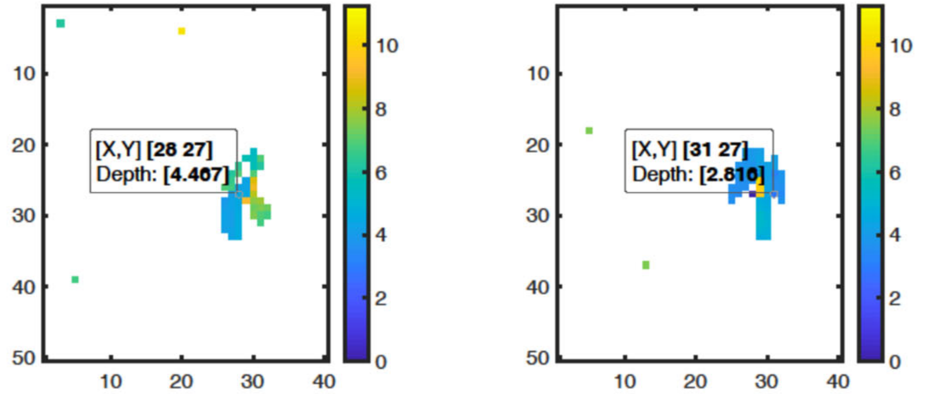
Depth, Path 2

Mirror at 1.90m → $\Delta d = 1.60\mathrm{m}$



Amplitude, Path 1

Amplitude, Path 2

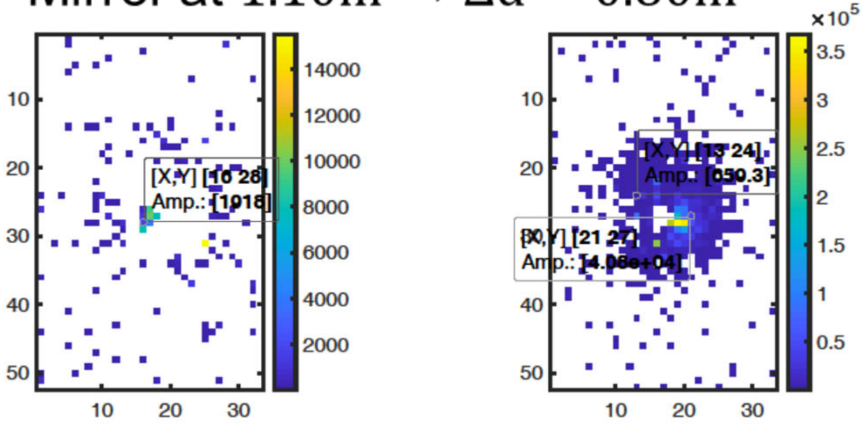


Depth, Path 1

Depth, Path 2

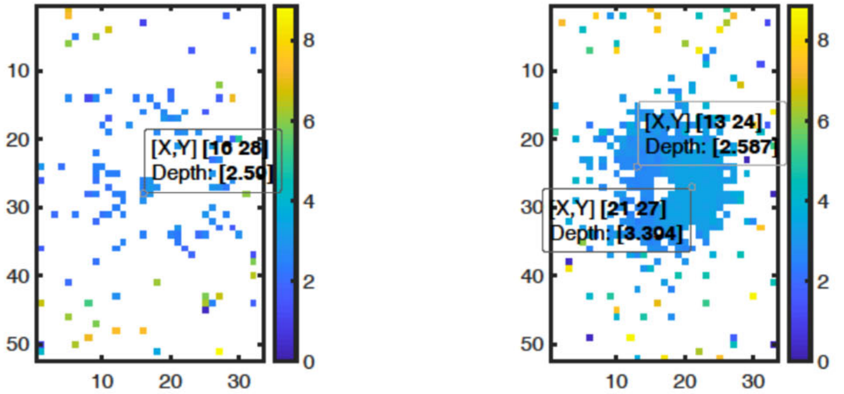
Real Experiments: Macro-Pixel (MP) Sensor, $Q = 12$:

Mirror at $1.10\mathrm{m} \rightarrow \Delta \mathrm{d} = 0.80\mathrm{m}$



Amplitude, Path 1

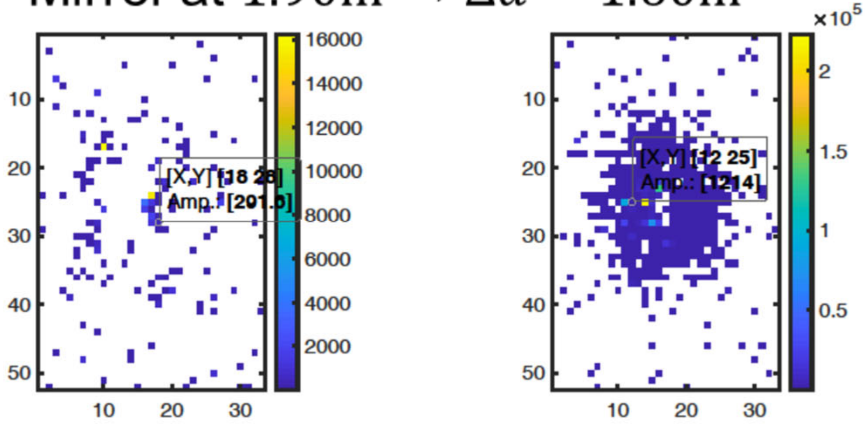
Amplitude, Path 2



Depth, Path 1

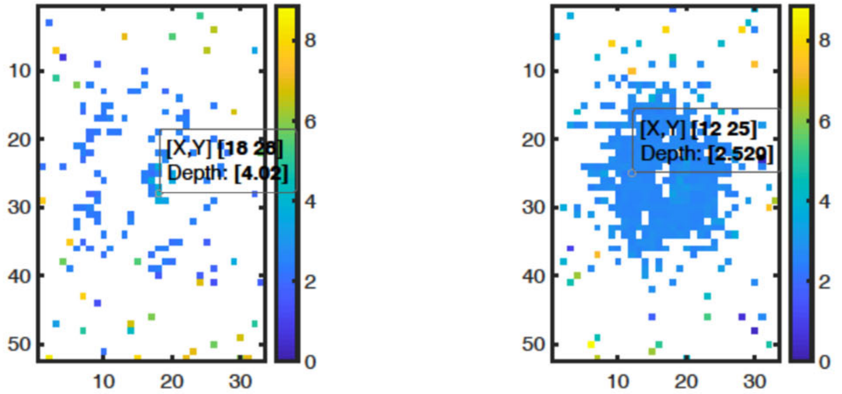
Depth, Path 2

Mirror at $1.90\mathrm{m} \rightarrow \Delta d = 1.60\mathrm{m}$



Amplitude, Path 1

Amplitude, Path 2



Depth, Path 1

Depth, Path 2

Summary of Results with Real Data:

- For single-path ToF imaging:
  - Very accurate reconstruction, down to $\sigma = 1.67\mathrm{cm} \rightarrow \sim \times 10^2$ superresolution factor
- For two-path ToF imaging:
  - The two paths can be identified and, for some pixels, properly separated.
  - For the pixels selected in the previous figures, we observe depth differences between path 1 and 2 that are similar to the ground truth:

Ground Truth	MA-ToF	MP-ToF
$\Delta\mathrm{d} = 0.80\mathrm{m}$	$\Delta d \sim 0.82\mathrm{m}$	$\Delta d \sim 0.80\mathrm{m}$
$\Delta d = 1.60\mathrm{m}$	$\Delta d \sim 1.65\mathrm{m}$	$\Delta d \sim 1.49\mathrm{m}$

5. Conclusions

- In ToF imaging, retrieving more than a single depth per pixel requires **multiple raw images** per frame.
- Time-domain multiplexing precludes real-time operation.
- We have proposed using two hardware architectures to attain **single-shot** ToF imaging, namely:
  - Multi-Aperture ToF arrays (MA)
  - Muti-tap Macro-Pixel ToF arrays (MP)
- Instead of adopting a classical time-gating formulation, which ties the temporal resolution to the number of samples, we propose a formulation in **Fourier domain** and solve the problem using a **fast and robust parametric spectral estimation method**.
- Results from both realistic simulations and experiments using real MA and MP prototypes have unveiled the potential of our method, showing **superresolution factors up to** $\sim \mathbf{10^2}$.

Thank you for your Attention!

Don’t hesitate forwarding your questions to:
heredia@zess.uni-siegen.de