# Transient Dictionary Learning for Compressed Time-of-Flight Imaging

2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2023)

Miguel Heredia Conde
$4^{th}$ -$10^{th}$ June 2023

uni-siegen.de/zess

ICASSP 2023
4 - 10 JUNE, RHODES ISLAND, GREECE

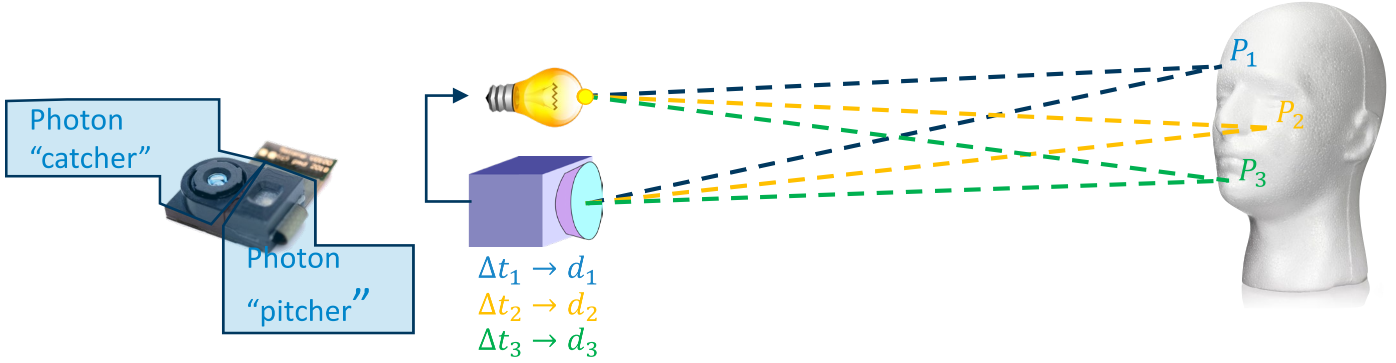

www.uni-siegen.de/zess

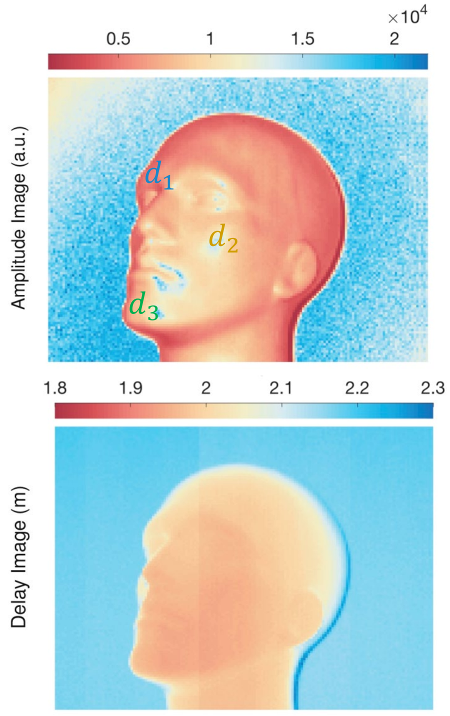
# 1 Introduction

Introduction

# Time-of-Flight 3D Imaging

## Principle of Operation:

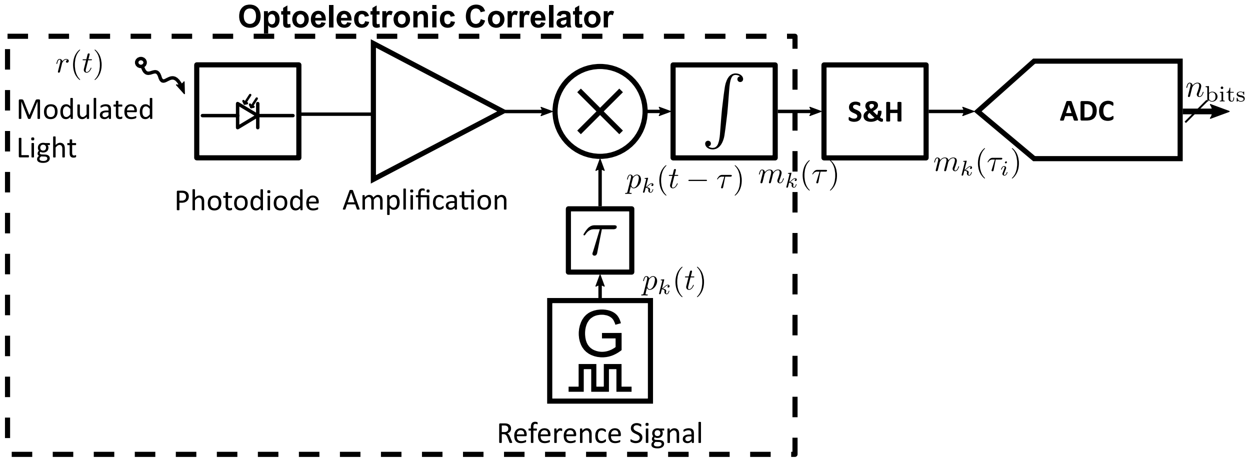



- Overarching idea: time of flight of photons encodes depth
- How to realize a time-resolved camera at low cost?
  - **Modulated illumination** → Fast NIR LED or VCSEL emitters + drivers
  - **Demodulating pixels** → Integration of photogenerated carriers controlled by custom signals
  - **Result:** electrooptical correlation sampling

# Time-of-Flight 3D Imaging

## Generic ToF Imaging Model:



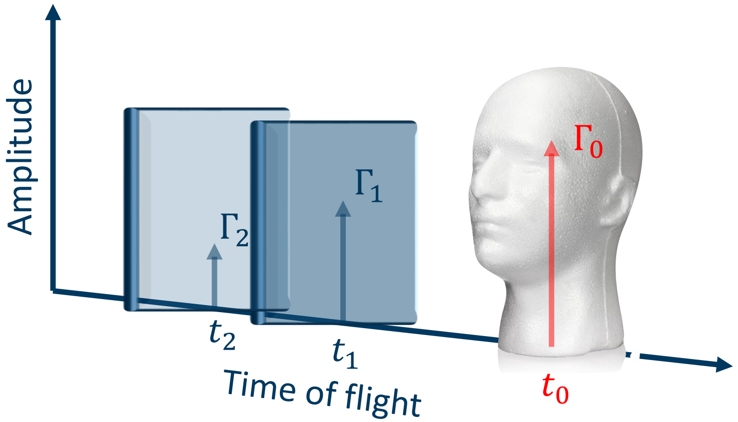

- Modulated illumination signal: $i(t)$
- Scene response function (SRF): $h(t)$
- $K \geq 1$ demodulation functions: $p_k(t)$, $1 \leq k \leq K$
- Return from the scene: $r(t) = i * h(t)$
- ToF **correlation measurements** (continuous):

$$m_k(t) = p_k \otimes r(t) = p_k \otimes (i * h)(t) = (i \otimes p_k) * h(t)$$

- Meaning: samples of the **convolution** between $h(t)$ and sensing functions $\phi_k(t) \coloneqq (i \otimes p_k)(t)$
- **Conventional ToF:** $K = 1$ and sampling at different $\tau_i$
  - Continuous Wave (CW) → Sinusoidal $\phi(t)$ [Heredia Conde, 2007]
  - Pulsed → Triangular $\phi(t)$
- **Coded ToF:** $K > 1$, typically only for $\tau = 0$ [Gupta *et al.*, 2018], [Lopez Paredes *et al.*, 2023]

# Ideal vs. Real Scene Responses

## Ideal Scene Response Functions:

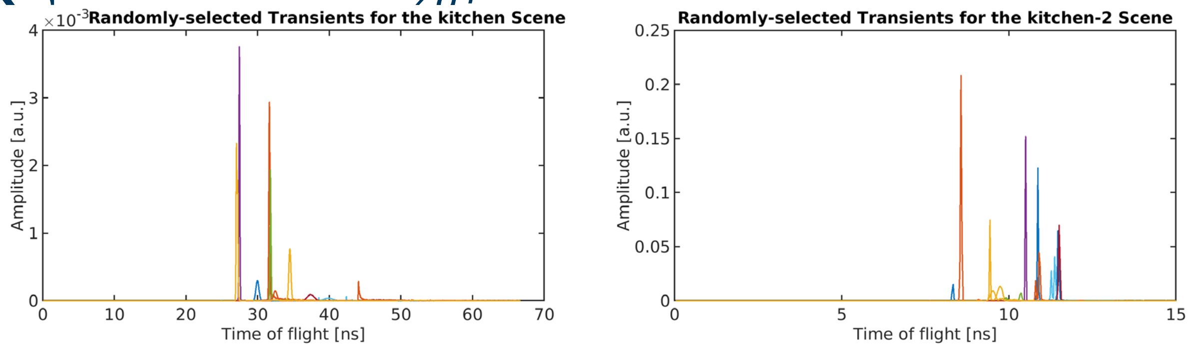



- Best case: **single bounce** per pixel
  - **SRF:** scaled and shifted **Dirac delta** function, $h(t) \coloneqq \Gamma_0 \delta(t - t_0)$
  - $\Gamma_0$ denotes the amplitude and $t_0 = 2d_0/c$ the time delay
- **Multi-path Interference (MPI):** multiple bounces per pixel
  - **SRF: weighted sum** of shifted **Dirac delta** functions:

$$h(t) \coloneqq \sum^{P-1} \ldots 2d_i \ldots$$

## Real Scene Response Functions:

- Result of global light transport effects
- Not itself sparse, but of low complexity



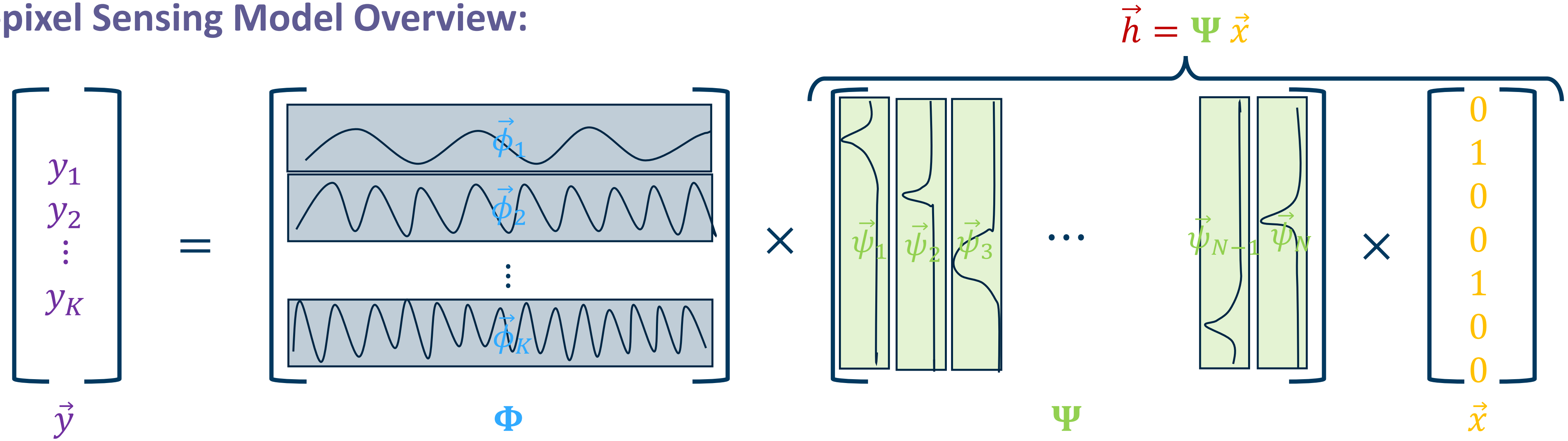

Ten randomly-selected transient profiles from the "kitchen" and "kitchen-2" scenes (IDs 12 and 13) of the iToF2dToF dataset [Gutierrez-Barragan *et al.*, 2021].

# 2

# Methodology

# A Compressed Sensing (CS) View of ToF 3D Imaging

## Per-pixel Sensing Model Overview:




**Measurement vector:** each element is obtained from one or several ToF raw data acquisitions

**Sensing matrix:** rows are discrete versions of the *sensing functions* implemented by the ToF hardware

**Dictionary:** columns are elements of a basis or a frame that allows representing the discrete SRFs in a *sparse* fashion

**Sparse vector:** encodes how few elements of $\Psi$ can be combined to obtain the SRF, $\vec{h}$

Aggregated **measurement matrix**, $A \coloneqq \Phi\,\Psi$

# A Compressed Sensing (CS) View of ToF 3D Imaging

## Consequences of the CS Model:

- Fully linear **sensing model**: $\vec{y} = \boldsymbol{A}\,\vec{x}$, with $\boldsymbol{A} \coloneqq \boldsymbol{\Phi}\,\boldsymbol{\Psi}$
- Incoherence requirement between $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$. $\vec{\psi}_i$ narrowly supported → $\vec{\phi}_i$ widely spread, $\forall\, i$
- The SRF can be readily obtained from $\vec{x}$: $\vec{h} = \boldsymbol{\Psi}\,\vec{x}$
- In turn, $\vec{x}$ can be obtained solving a linearly-constrained **sparse reconstruction** problem:

$$\hat{\vec{x}} = \underset{\vec{x}}{\operatorname{argmin}} \|\vec{x}\|_0 \;\; \text{subject to } \vec{y} = \boldsymbol{A}\,\vec{x}$$

- Or its convex relaxation:

$$\hat{\vec{x}} = \underset{\vec{x}}{\operatorname{argmin}} \|\vec{x}\|_1 \;\; \text{subject to } \vec{y} = \boldsymbol{A}\,\vec{x}$$

Methodology

# CW-ToF Sensing Model

## Fourier Sampling of Spiky Signals:

- Use *sinusoids* as sensing functions
- Complies with the incoherence requirement between $\mathbf{\Phi}$ and $\mathbf{\Psi}$
- For a given frequency, $f_k$, multiple raw measurements can be combined to generate a **complex** phasor:

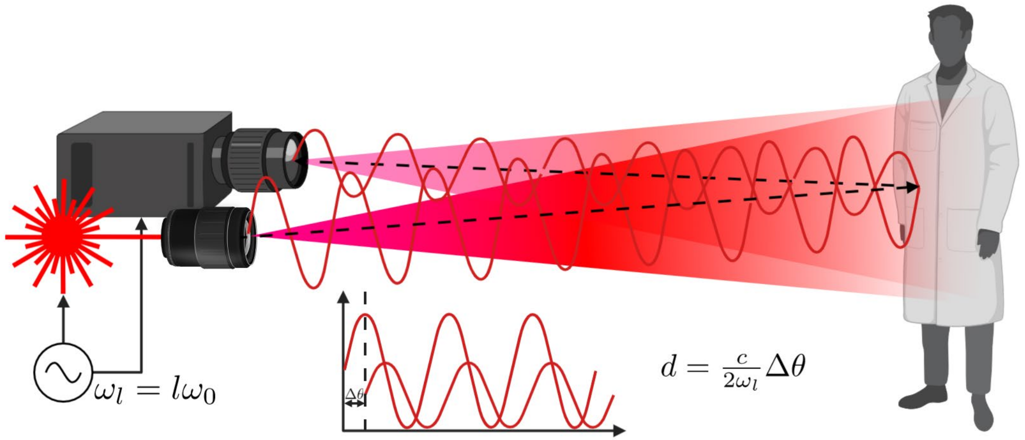
$$y_k^{\Re} = \vec{\phi}_k^{\Re\,\top}\vec{h}, \qquad \text{with } \vec{\phi}_k^{\Re\,\top}[i] = A\cos(2\pi f_k i\Delta t)$$

$$y_k^{\Im} = \vec{\phi}_k^{\Im\,\top}\vec{h}, \qquad \text{with } \vec{\phi}_k^{\Im\,\top}[i] = A\sin(2\pi f_k i\Delta t)$$

where $\Delta t$ denotes the discrete time step.

- Real SRF → **real** sensing model:

$$\vec{y} = \mathbf{\Phi}\vec{h}, \text{ with } \mathbf{\Phi} := \begin{bmatrix} \mathbf{\Phi}^{\Re} \\ \mathbf{\Phi}^{\Im} \end{bmatrix}, \quad \mathbf{\Phi}^{\Re} := \left[\vec{\phi}_k^{\Re\,\top}\right]_{k=1}^{K}, \quad \mathbf{\Phi}^{\Im} := \left[\vec{\phi}_k^{\Im\,\top}\right]_{k=1}^{K}$$

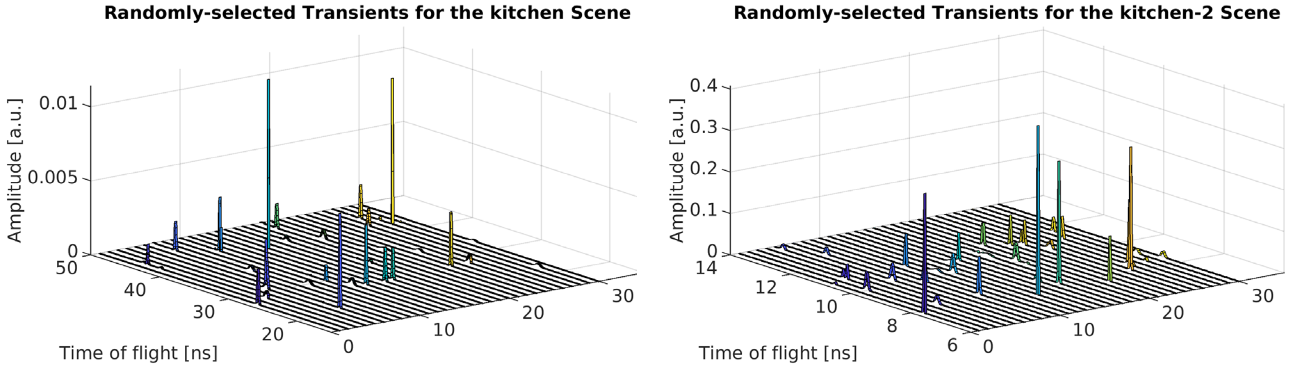


# Transient Dictionary Learning

## How to Obtain the Best Dictionary?

- **Goal:** represent any SRF, $\vec{h}$, with few $\vec{\psi}_i$, as accurately as possible
- **Idea:** find the set of $\vec{\psi}_i$ that best represent a collection of data $\boldsymbol{H} = [\vec{h}_i]_{1 \le i \le M}$

$\boldsymbol{H} =$




$\cdots$

$120 \times 160$ pixels $\times$ 21 scenes

- **How?** Optimization problem:

$$\widehat{\boldsymbol{\Psi}}, \widehat{\mathbf{X}} = \underset{\boldsymbol{\Psi},\mathbf{X}}{\operatorname{argmin}} \|\boldsymbol{H} - \boldsymbol{\Psi}\mathbf{X}\|_F^2, \text{ subject to } \|\vec{x}_i\|_0 \le s_{\max}, \forall i$$

where $\mathbf{X} = [\vec{x}_i]_{1 \le i \le M}$ and $s_{\max}$ is an upper bound for the sparsity $s$.

Max. $s_{\max}$ non-zeros

$$\mathbf{X} = \begin{bmatrix} 0 & 0 & 1 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

# 3

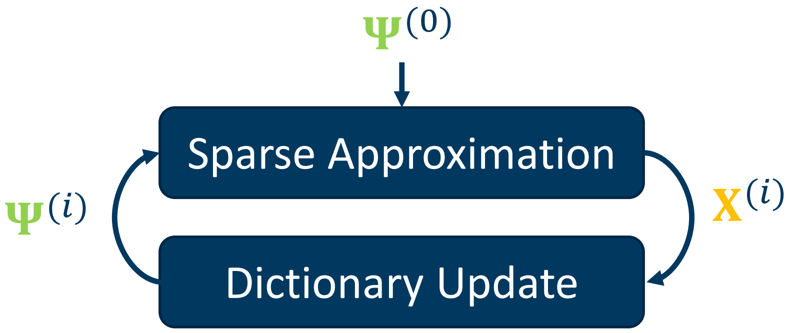
# Experimental Evaluation

# What is the Best Method for Learning Sparse Transient Dictionaries?

## Candidate Methods:

- Method of Optimal Directions (MOD) [Engan *et al.*, 1999]
- K-Singular Value Decomposition (K-SVD) [Aharon *et al.*, 2006]
- Approximate K-SVD [Rubinstein *et al.*, 2018]
- Online Dictionary Learning (ODL) [Mairal *et al.*, 2009]
- Reweighted Least Squares Dictionary Learning Algorithm (RLS-DLA) [Skretting and Engan, 2010]

## General Structure:



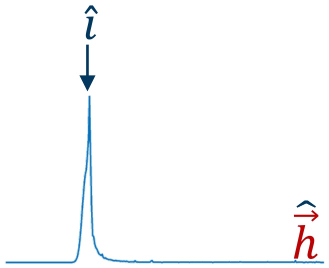

## Homogenized Conditions:

- **Same** data-agnostic **tight frame,** $\mathbf{\Psi}^{(0)}$, used as seed; $s_{\max} = 16$ for training $N = 8000$ atoms
- Orthogonal Matching Pursuit (OMP) used as sparse approximation method for speed
- Random selection of $\mathbf{10^5}$ **transients for training**, over the $> 4 \times 10^5$ available in 21/25 scenes of iToF2dToF [Gutierrez-Barragan *et al.*, 2021]
- **Four** remaining scenes for posterior **validation**

# Depth Retrieval Performance

## Depth Retrieval from Reconstructed Transient Profiles:

- Via peak detection:



$$\hat{d} = \frac{c}{2(\hat{\imath}\Delta t)}, \; \hat{\imath} = \underset{i}{\operatorname{argmax}} \, \hat{\vec{h}}[i], \text{ s.t. } \hat{\vec{h}}[i] > \epsilon,$$

with $\hat{\vec{h}} = \boldsymbol{\Psi} \, \hat{\vec{x}}$

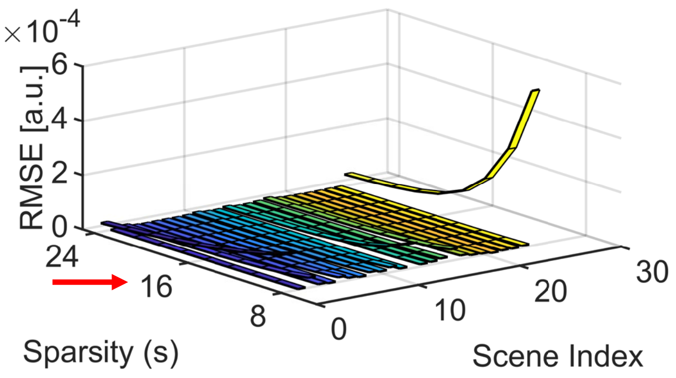
## Depth MAE per Percentile [mm] for Scene 12 ("kitchen"):

| Percentile | 0-75% | 75-85% | 85-95% | 95-99% |
|---|---|---|---|---|
| MOD | 0 | 3.248 | 7.041 | 24.38 |
| K-SVD | 0 | 4.571 | 7.270 | 24.97 |
| Approx. K-SVD | 0.1240 | 5.000 | 8.162 | 26.29 |
| ODL | 0.3212 | 5.000 | 10.22 | 31.07 |
| RLS-DLA | 1.077 | 5.341 | 12.27 | 40.47 |
| Best of [G.-B. *et al.*, 2021] | 7.19 | 20.40 | 32.17 | 71.56 |

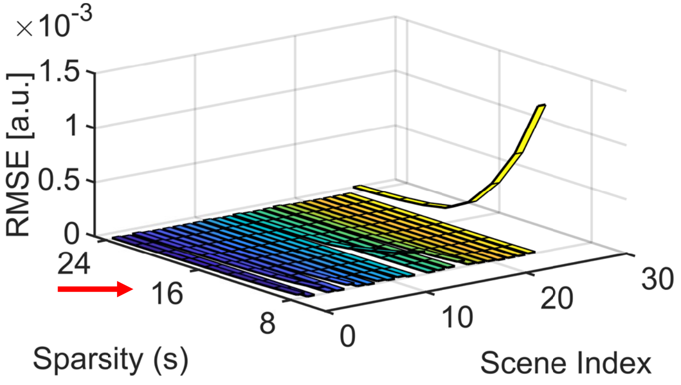
Best in all percentiles

# How Sparse are the Transient Profiles?

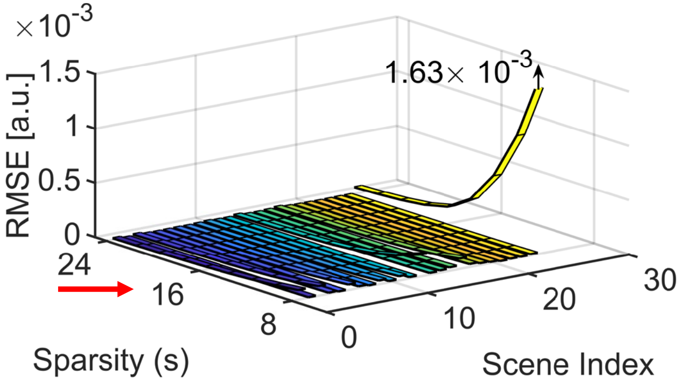
## Evolution of Normalized RMSE of the Transient Profiles vs. Sparsity, $s$:




MOD

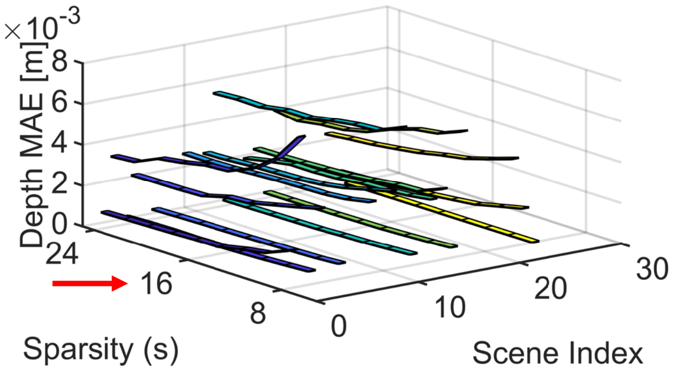



K-SVD

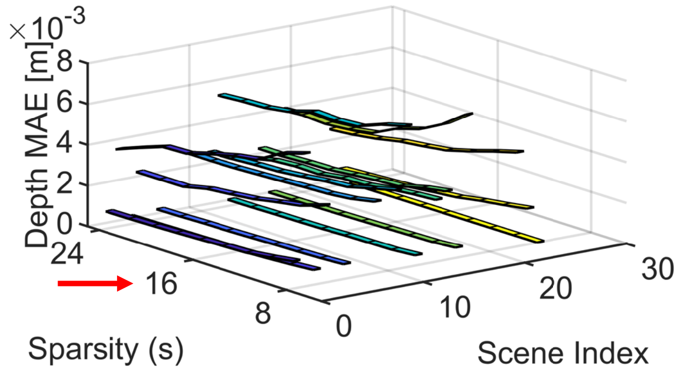



Approx. K-SVD

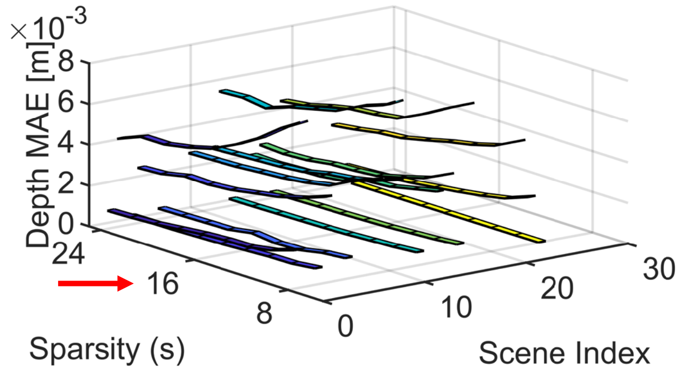
## Evolution of Depth MAE vs. Sparsity, $s$:




MOD




K-SVD




Approx. K-SVD

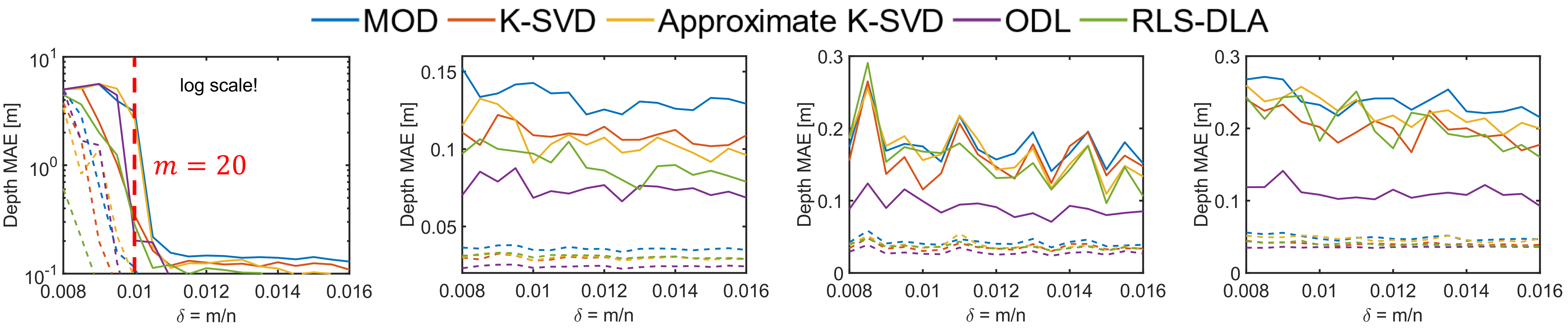
Experimental Evaluation

# How Many Measurements?

## How to Select the Frequencies, $f_k$? Four Options:

a) Uniform    b) Random within grid    c) Random    d) Sparse ruler

## Evolution of Depth MAE vs. $\delta := m/n$ (solid: 95-99%, dashed: 85-95%):

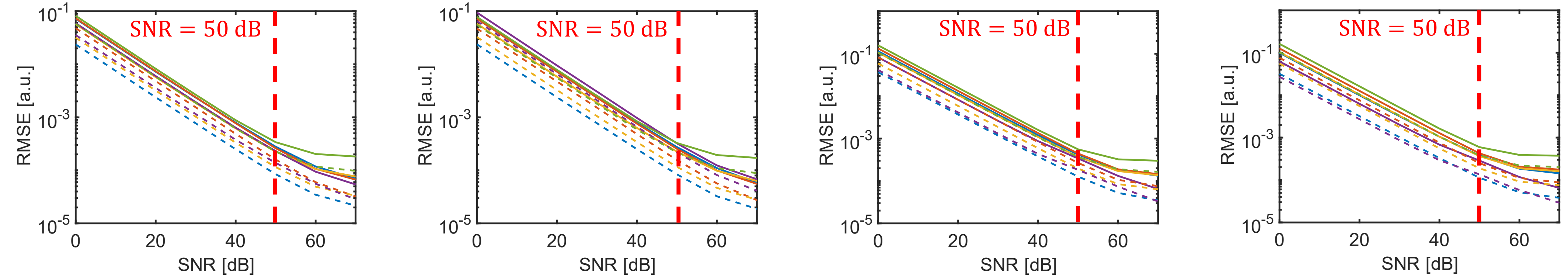



a) Uniform    b) Random within grid    c) Random    d) Sparse ruler

## Experimental Evaluation

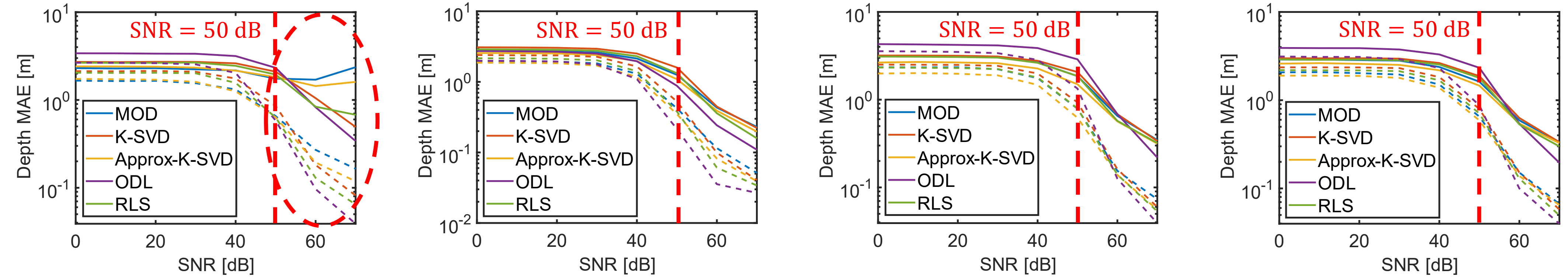
# Robustness to Noise

### How Robust is the Reconstruction to Measurement Noise? Results for $m = 20$:




a) Uniform — b) Random within grid — c) Random — d) Sparse ruler

### Evolution of Depth MAE vs. SNR:




a) Uniform — b) Random within grid — c) Random — d) Sparse ruler

# 4

# Conclusions

# Conclusions

## In a Nutshell…

- Robust **CS-based** depth estimation from *few* **MPI**-corrupted **ToF** measurements demonstrated
- **CW-ToF** sensing model leveraging **uniform** and **non-uniform** frequency sampling schemes
- Classical **sparse dictionary learning** methods used to learn a representation for transient profiles
- Learnt representations only limit transient profile reconstruction accuracy beyond $50\ \mathrm{dB}$

## Take-home Messages

- **CS + trained dictionary** as alternative to [or baseline for] deep learning models
- **Number of measurement**s decoupled from the transient ambient dimension, $m \sim \mathcal{O}(s)$
- **NUS** schemes allow for operating with fewer measurements

# References

[Heredia Conde, 2007] M. **Heredia Conde**, *Compressive Sensing for the Photonic Mixer Device - Fundamentals, Methods and Results*. Springer Vieweg, 2017.

[Gupta *et al.*, 2018] M. Gupta, A. Velten, S. K. Nayar, E. Breitbach, "What Are Optimal Coding Functions for Time-of-Flight Imaging?," in *ACM Transactions on Graphics*, vol. 37, no. 2, pp. 1-18, 2018, doi: 10.1145/3152155.

[Lopez Paredes *et al.*, 2023] A. Lopez Paredes, M. **Heredia Conde** and O. Loffeld, "Sparsity-Aware 3-D ToF Sensing," in *IEEE Sensors Journal*, vol. 23, no. 4, pp. 3973-3989, 2023, doi: 10.1109/JSEN.2023.3234533.

[Gutierrez-Barragan *et al.*, 2021] F. Gutierrez-Barragan, H. Chen, M. Gupta, A. Velten and J. Gu, "iToF2dToF: A Robust and Flexible Representation for Data-Driven Time-of-Flight Imaging," in *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1205-1214, 2021, doi: 10.1109/TCI.2021.3126533.

[Engan *et al.*, 1999] Kjersti Engan, Sven O. Aase, and John Hakon Husoy, "Method of Optimal Directions for Frame Design," in *Proceedings of the 1999 IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 1999, vol. 5, pp. 2443–2446.

[Aharon *et al.*, 2006] Michal Aharon, Michael Elad, and Alfred Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[Rubinstein *et al.*, 2018] Ron Rubinstein, Michael Zibulevsky, and Michael Elad, "Efficient Implementation of the K-SVD Algorithm Using Batch Orthogonal Matching Pursuit," Tech. Rep., Technion – Computer Science Department, 2018.

[Mairal *et al.*, 2009] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro, "Online Dictionary Learning for Sparse Coding," in *Proceedings of the 26th Annual International Conference on Machine Learning*, New York, NY, USA, 2009, ICML '09, pp. 689–696, Association for Computing Machinery.

[Skretting and Engan, 2010] Karl Skretting and Kjersti Engan, "Recursive least squares dictionary learning algorithm," *Trans. Sig. Proc.*, vol. 58, no. 4, pp. 2121–2130, apr 2010.

# Thank You for your Attention!

**Miguel Heredia Conde**
Research Group on “Compressive Sensing for the Photonic Mixer Device”
Center for Sensor Systems (ZESS)
University of Siegen

heredia@zess.uni-siegen.de
www.uni-siegen.de/zess

Looking for postdocs and PhD students in Siegen (Germany)!