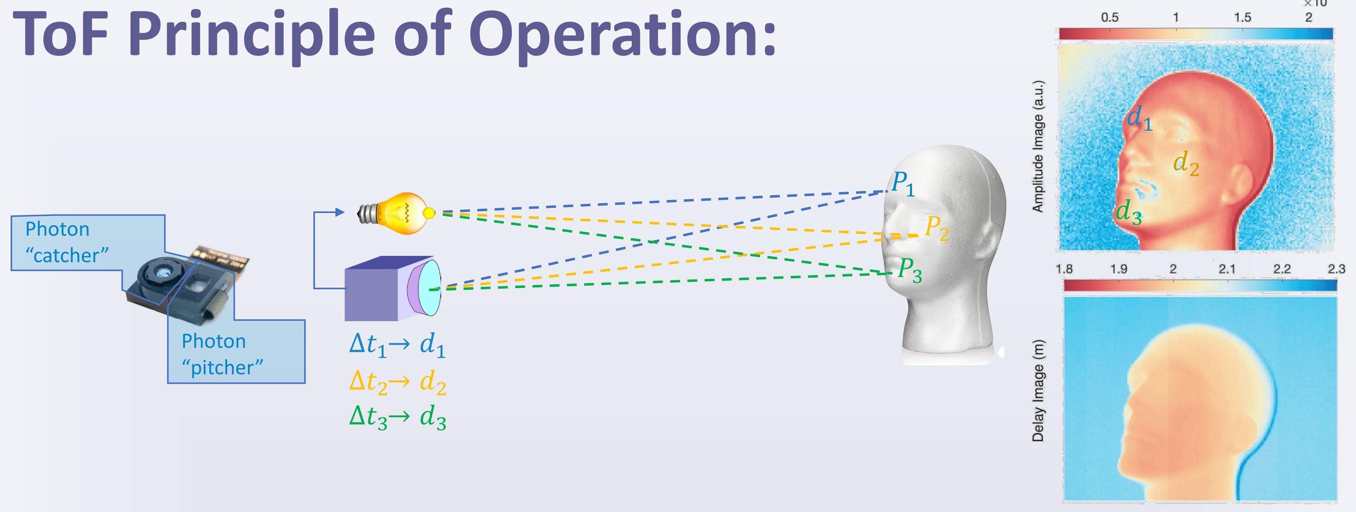
Transient Dictionary Learning for Compressed Time-of-Flight Imaging

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

Miguel Heredia Conde, Center for Sensor Systems (ZESS), University of Siegen

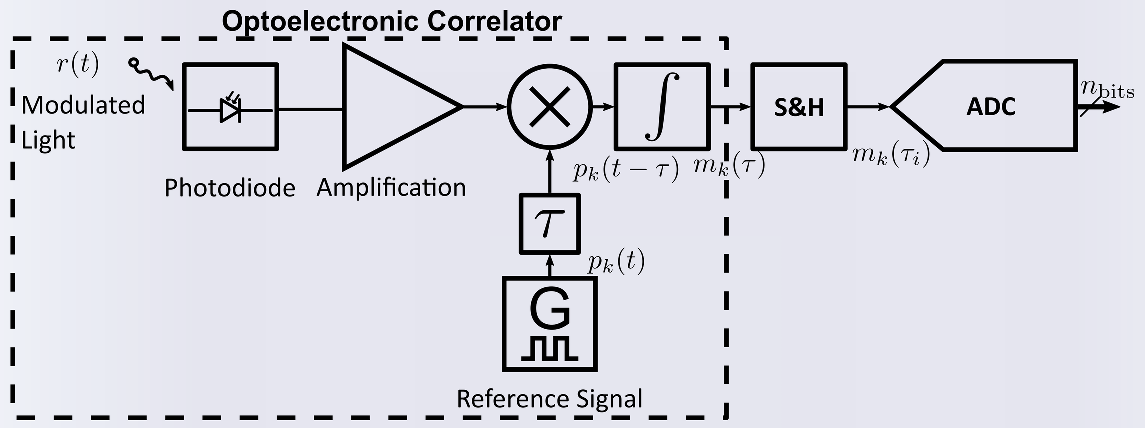
1. Introduction

ToF 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

Generic ToF Imaging Model:



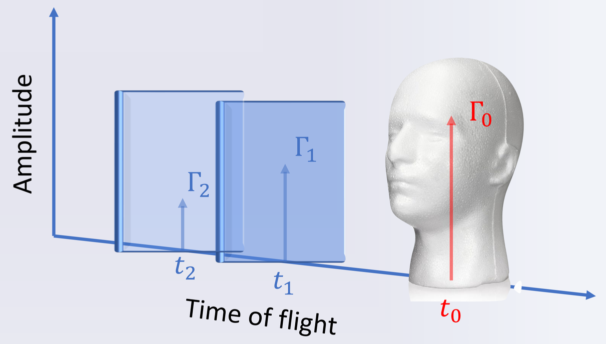
- Modulated illumination signal: $i(t)$
- Scene response function (SRF): $h(t)$
- $K \geq 1$ demodulation functions: $p_k(t)$, $1 \leq k \leq K$
- Return from the scene: $r(t) = i * h(t)$
- ToF **correlation measurements** (continuous):

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

- Meaning: samples of the **convolution** between $h(t)$ and $\phi_k(t) \coloneqq (i \otimes p_k)(t)$
- **Conventional ToF:** $K = 1$ and sampling at different τ_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]

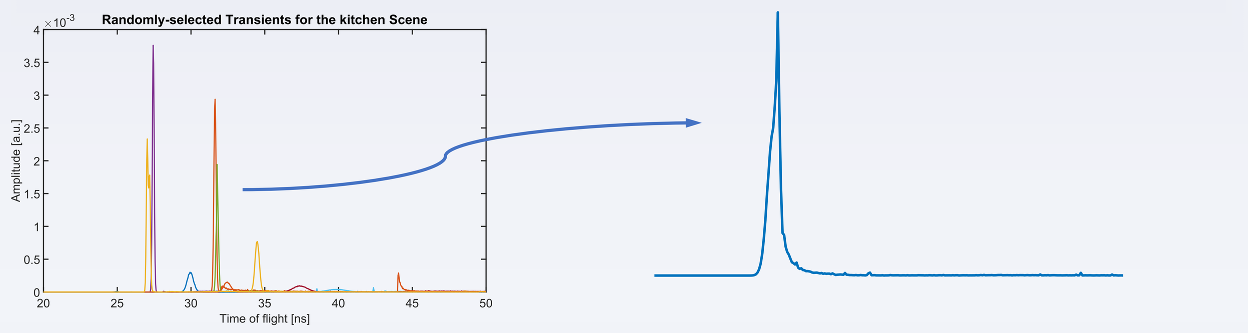
Ideal Scene Response Functions:

- Best case: **single bounce** per pixel
  - **SRF:** $h(t) \coloneqq \Gamma_0 \delta(t - t_0)$
- **Multi-path Interference (MPI):** multiple bounces per pixel
  - **SRF: weighted sum** of shifted **Dirac delta** functions: $h(t) \coloneqq \sum_{i=0}^{P-1} \Gamma_i \delta(t - t_i)$, $t_i = \frac{2d_i}{c}$



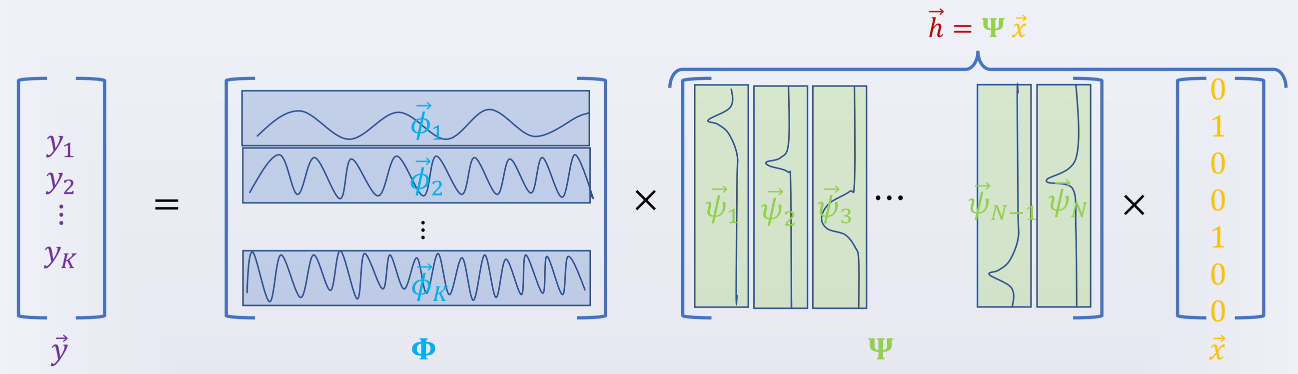
Real Scene Response Functions:

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



2. Methodology

Per-pixel Compressed Sensing (CS) Model:



Measurement vector: elements 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

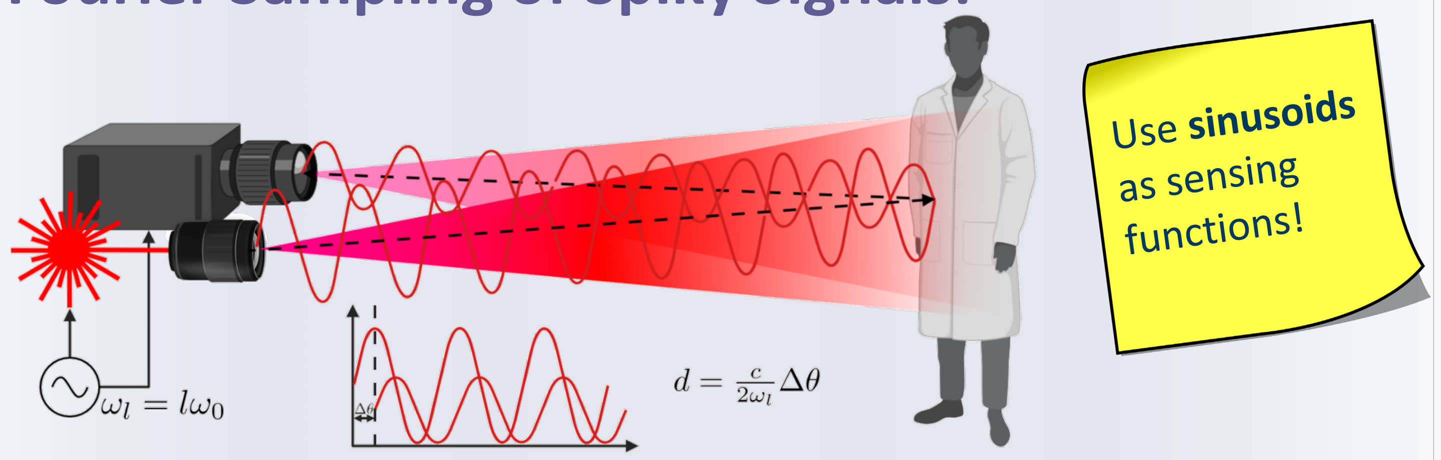
Sparse vector: encodes how few elements of Ψ can be combined to obtain the SRF, $\vec{h}$

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

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

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

Fourier Sampling of Spiky Signals:



- Complies with the incoherence requirement between Φ and Ψ
- For a given frequency, f_k, multiple raw measurements can be combined to generate a **complex** phasor:

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

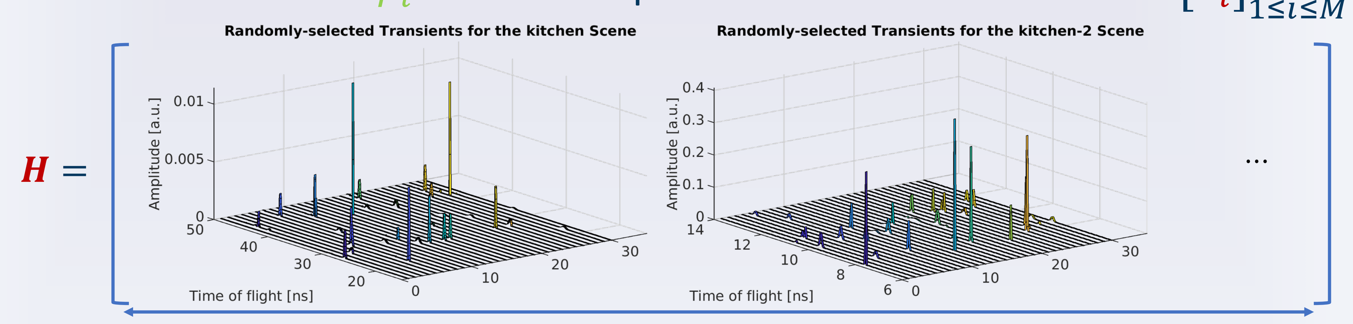
$$y_k^{\Im} = \vec{\phi}_k^{\Im\,T} \vec{h}, \quad \text{with } \vec{\phi}_k^{\Im\,T}[i] = A\sin(2\pi f_k i \Delta t)$$

where Δt denotes the discrete time step.

- **Real** sensing model: $\vec{y} = \Phi\vec{h}$, with $\Phi \coloneqq \begin{bmatrix}\Phi^{\Re} \\ \Phi^{\Im}\end{bmatrix}$, $\Phi^{\Re} \coloneqq \left[\vec{\phi}_k^{\Re\,T}\right]_{k=1}^{K}$, $\Phi^{\Im} \coloneqq \left[\vec{\phi}_k^{\Im\,T}\right]_{k=1}^{K}$

How to Obtain the Best Dictionary?

- **Idea:** find the set of $\vec{\psi}_i$ that best represent a data collection $H = [\vec{h}_i]_{1 \leq i \leq M}$



120×160 pixels $\times$ 21 scenes

- **How?** Optimization problem:

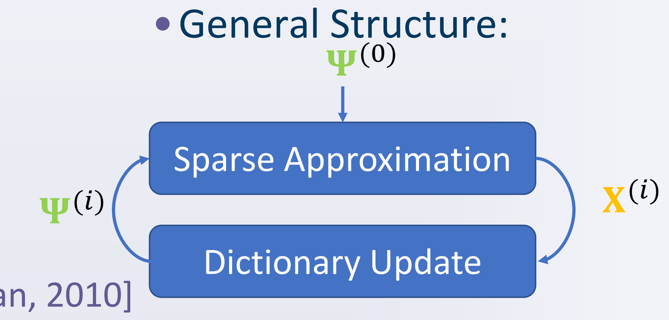
$$\hat{\Psi}, \hat{X} = \underset{\Psi, X}{\operatorname{argmin}} \|H - \Psi X\|_F^2, \text{ subject to } \|\vec{x}_i\|_0 \leq s_{\max}, \forall i$$

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

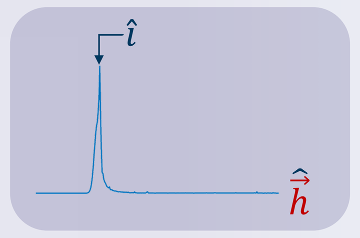
3. Experimental Evaluation

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:



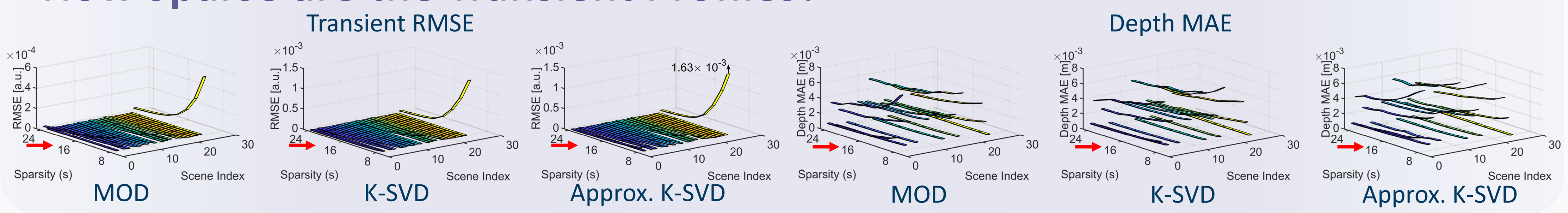
- Homogenized Test Conditions:
  - **Same** data-agnostic **tight frame**, $\Psi^{(0)}$, used as seed; $s_{\max} = 16$ for training $N = 8000$ atoms
  - Orthogonal Matching Pursuit (**OMP**) used as sparse approximation method for speed
  - Random selection of 10^5 **transients for training**, over the $> 4 \times 10^5$ available in 21/25 scenes of iToF2dToF [Gutierrez-Barragan *et al.*, 2021]. **Four** remaining scenes for posterior **validation.**
- Depth Retrieval Performance:
  - Depth Retrieval via **peak detection**: $\hat{d} = \frac{c}{2(\hat{\imath}\Delta t)}$, $\hat{\imath} = \underset{i}{\operatorname{argmax}}\, \hat{\vec{h}}[i]$, s.t. $\hat{\vec{h}}[i] > \epsilon$, with $\hat{\vec{h}} = \Psi\,\hat{\vec{x}}$



	Normalized RMSE ($\times\ 10^{-5}$ [a.u.]) for "kitchen"				Depth MAE [mm] for scene 12 ("kitchen")			
Percentile	**0-75%**	**75-85%**	**85-95%**	**95-99%**	**0-75%**	**75-85%**	**85-95%**	**95-99%**
MOD	0.0301	0.1914	0.2486	0.6386	0	3.248	7.041	24.38
K-SVD	0.0663	0.2018	0.3634	0.8074	0	4.571	7.270	24.97
Approx. K-SVD	0.0943	0.2891	0.5366	1.215	0.1240	5.000	8.162	26.29
ODL	0.0697	0.2263	0.4628	1.251	0.3212	5.000	10.22	31.07
RLS-DLA	0.3868	1.236	2.511	5.901	1.077	5.341	12.27	40.47
Best of [G.-B. *et al.*, 2021]	0.0301	0.1914	0.2486	0.6386	7.19	20.40	32.17	71.56

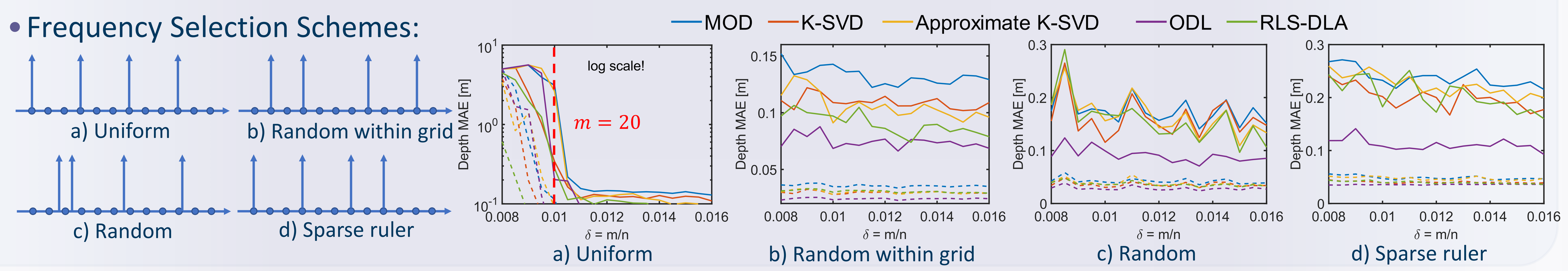
Best in all percentiles

How Sparse are the Transient Profiles?

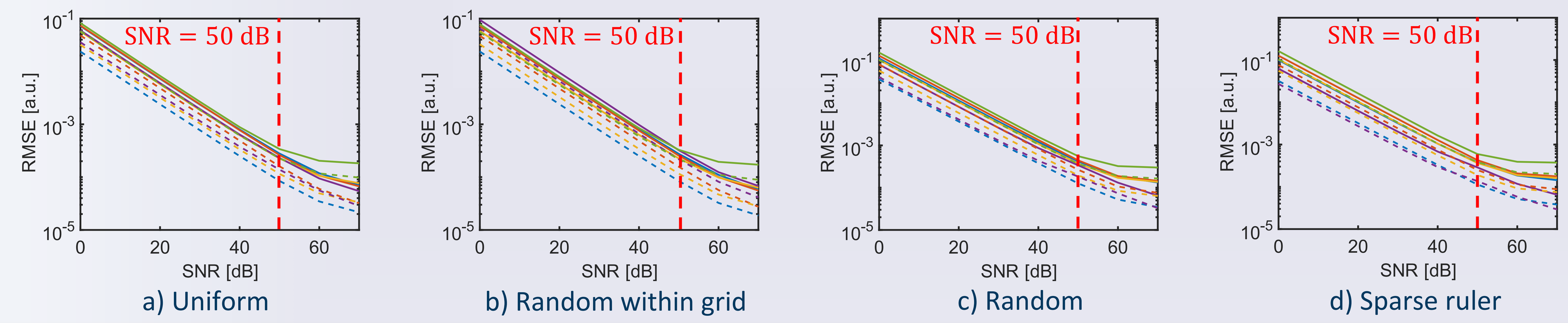


How Many Measurements Are Required?

- Frequency Selection Schemes:



How Robust is the Reconstruction to Measurement Noise?



Results for $m = 20$

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 Trans. on Graphics*, vol. 37, no. 2, pp. 1-18, 2018.

[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.

[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 Trans. on Computational Imaging*, vol. 7, pp. 1205-1214, 2021.

[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.