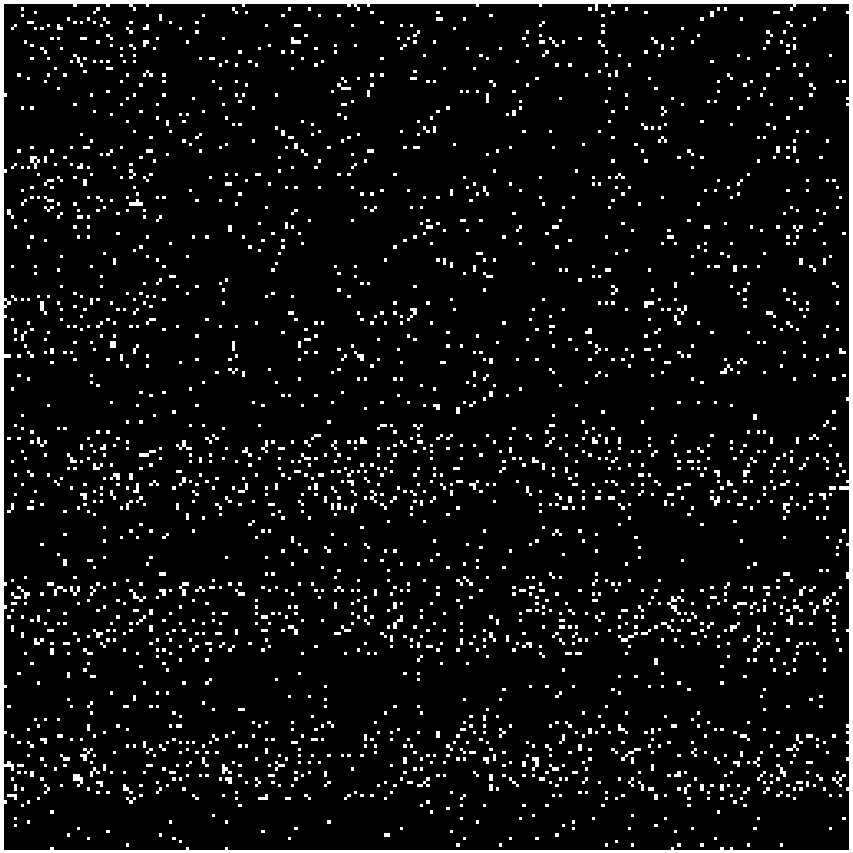
The Dark Side of Image Reconstruction:
Emerging methods for photon-limited imaging

Rebecca Willett
University of Wisconsin-Madison

Photon limitations in night vision[1]

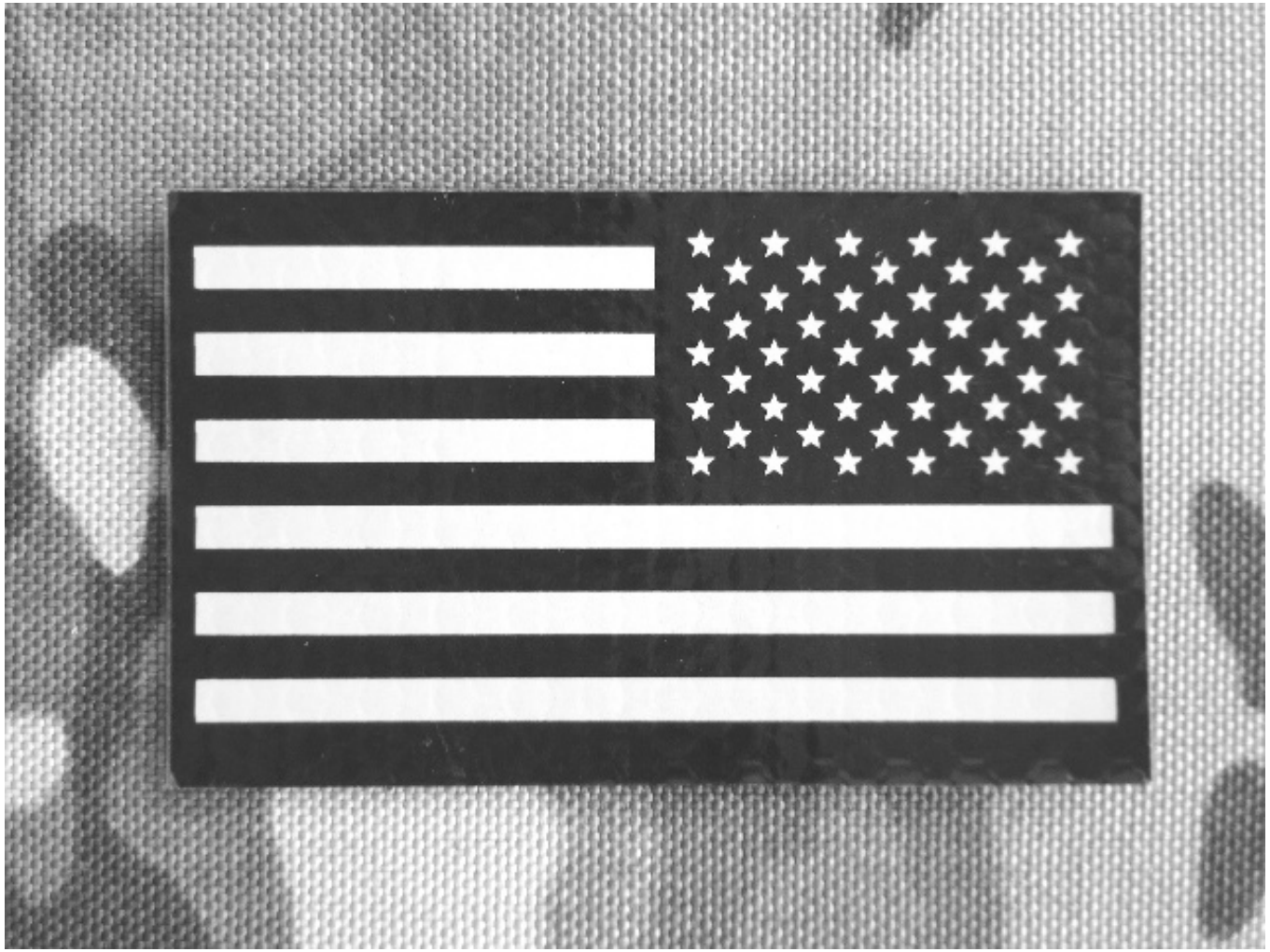
[1] BMW, 2005

Photon limitations in night vision

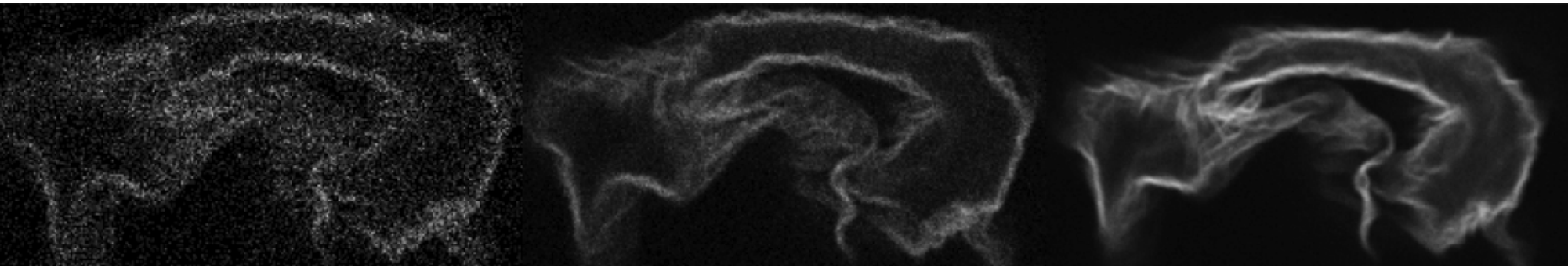


What is the distribution of light in this scene?

Photon limitations in night vision



Photon limitations in fluorescence microscopy[2]

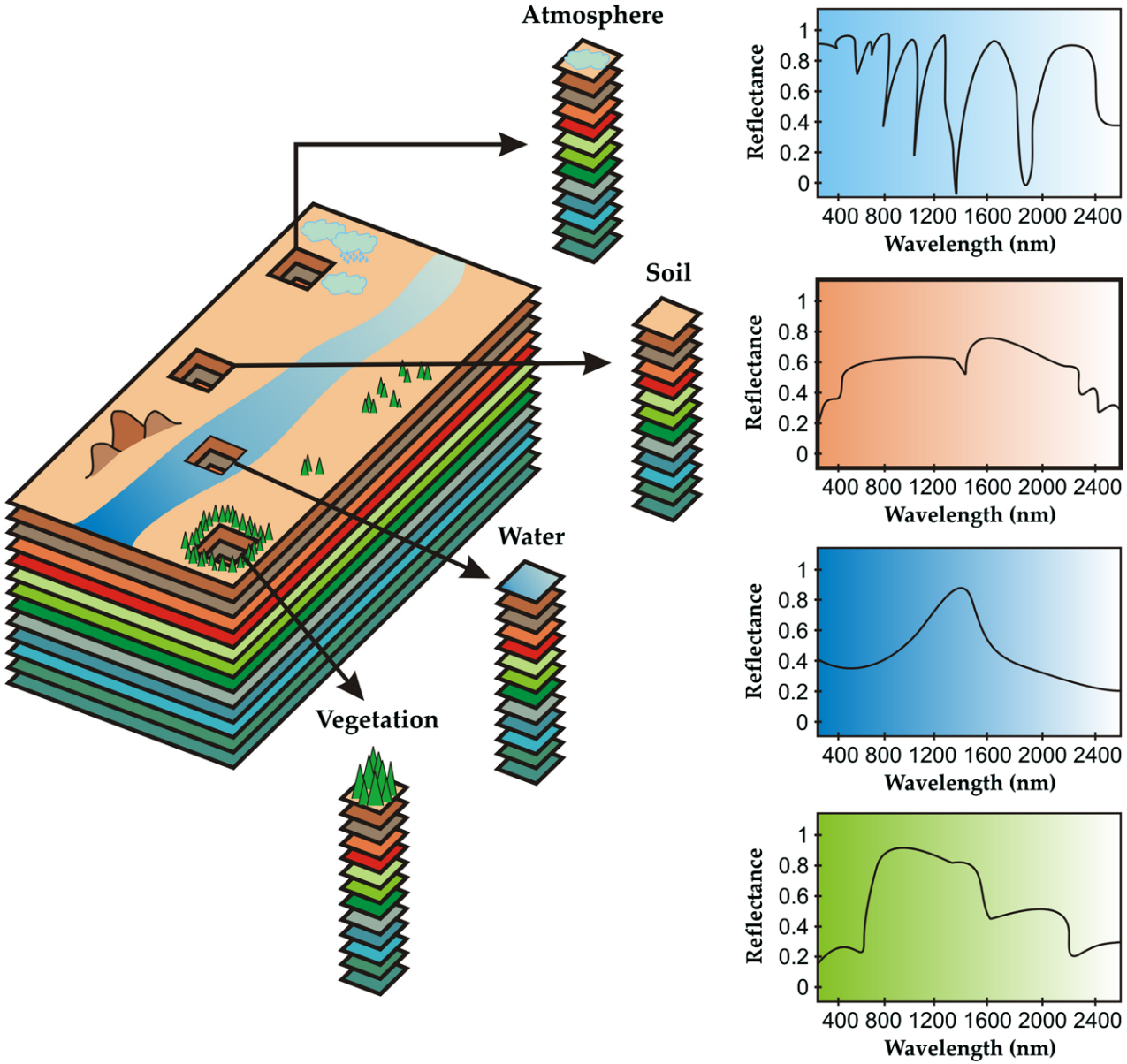


0.1s 10s 120s

- Want to limit data collection time
  - Higher throughput experiments
  - Higher temporal resolution for living cells
- Want to limit amount of dye used
  - Avoid photobleaching
  - Avoid toxic effects on living cells

[2] University of Wisconsin Laboratory for Optical and Computational Instrumentation

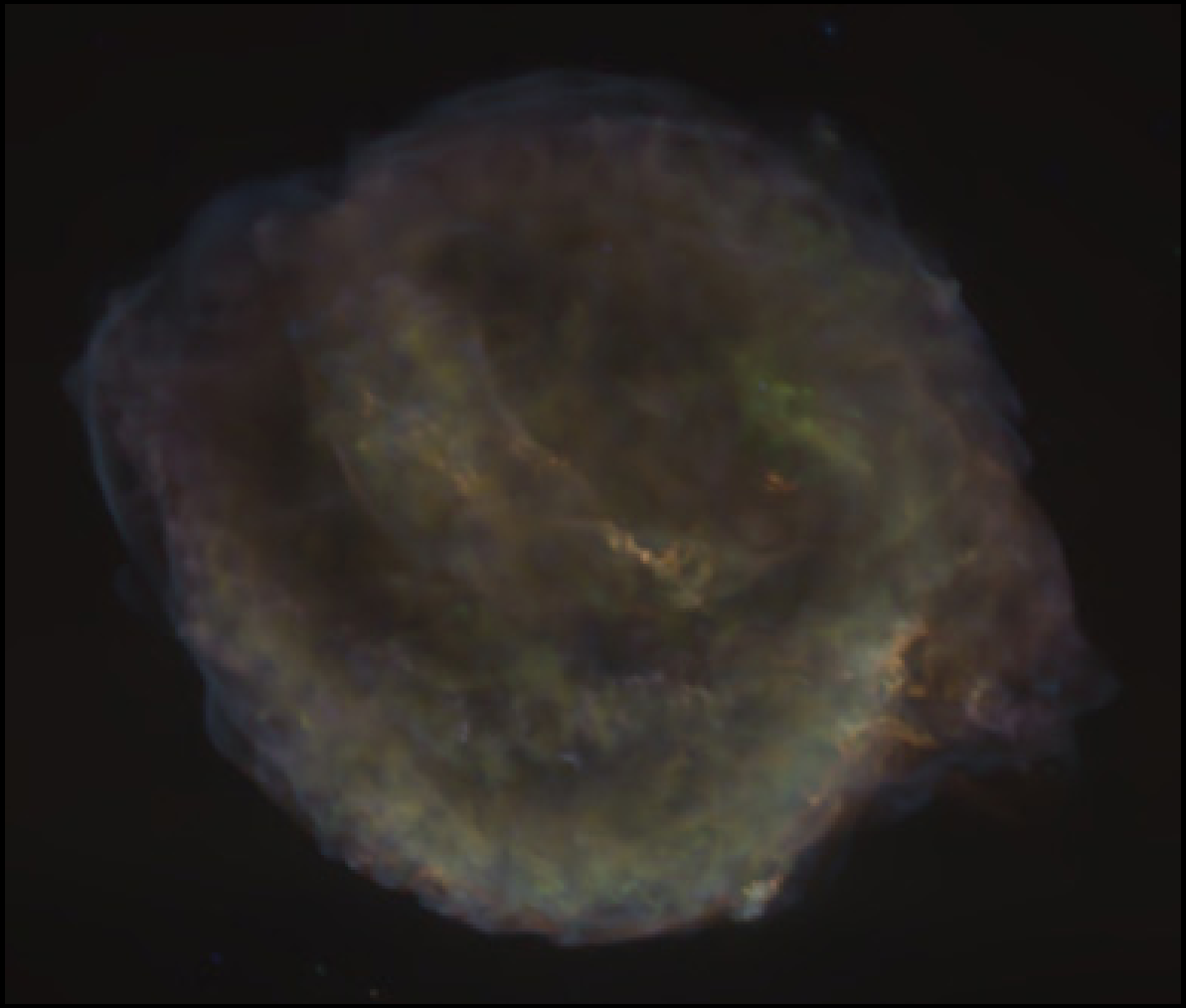
Photon limitations in spectral imaging[3]



More spectral resolution = fewer photons per spectral band

[3] Molero, Garzon, Garcia, Plaza, 2012

Photon limitations in astronomy

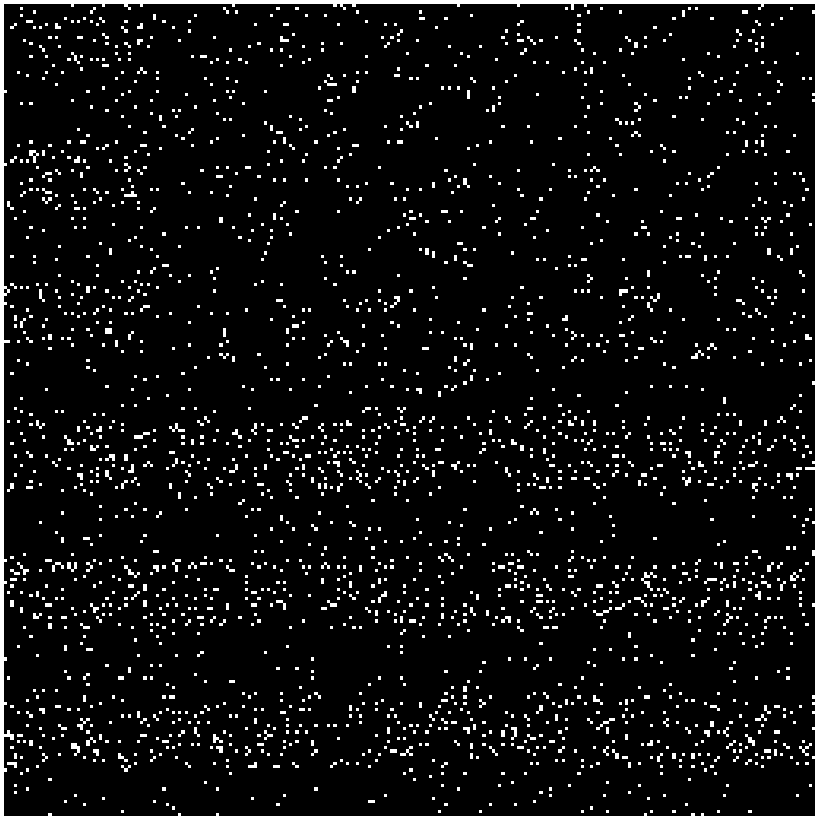


Spectral image of Kepler supernova

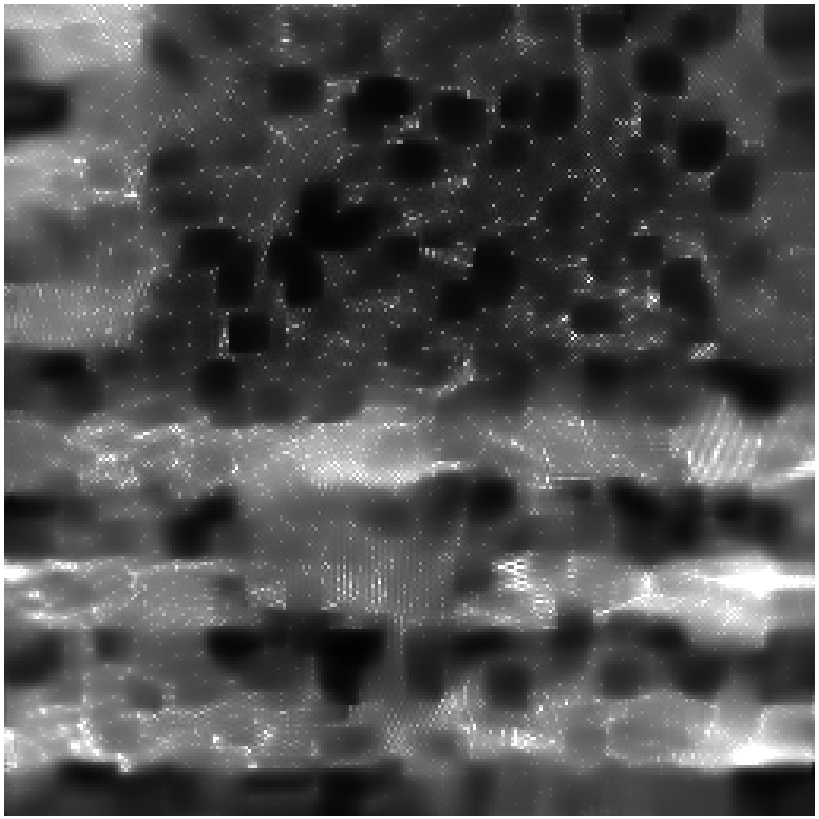
K. Borkowski, S. Reynolds, D. Green, U. Hwang, R. Petre, K. Krishnamurthy, & R. Willett, 2010

Low dose computed tomography

Off-the-shelf methods fail on photon-limited data



Noisy data



BM3D estimate

Problem formulation

We observe

$$y \sim \text{Poisson}(TAf^*)$$

$$y_i \sim \text{Poisson}\left(T\sum_j A_{i,j} f_j^*\right)$$

where

- f^* is the true image of interest,
- $T > 0$ corresponds to the sensing time or overall intensity
- A is an operator corresponding to blur, tomographic projections, compressed sensing measurements, or any other linear distortion of the image.
- Non-negativity: $f_j, y_i, A_{i,j} \geq 0$
- Signal-dependent noise: $\mathbb{E}[y_i] = \text{Var}[y_i] = (Af^*)_i$

Our goal is to estimate f^* from y.

Scarcity of photons counteracted by exploiting **sparsity** or low-dimensional structure

Finding and exploiting low-dimensional structure in photon-limited settings presents unique challenges

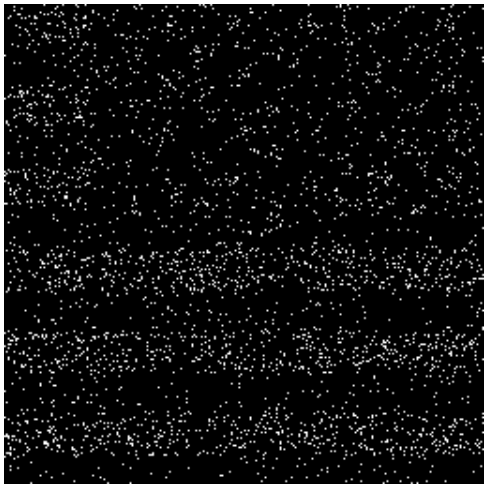
Variance Stabilizing Transforms

One challenge associated with Poisson noise is that the variance is signal-dependent. We can mitigate this via variance stabilizing transforms.

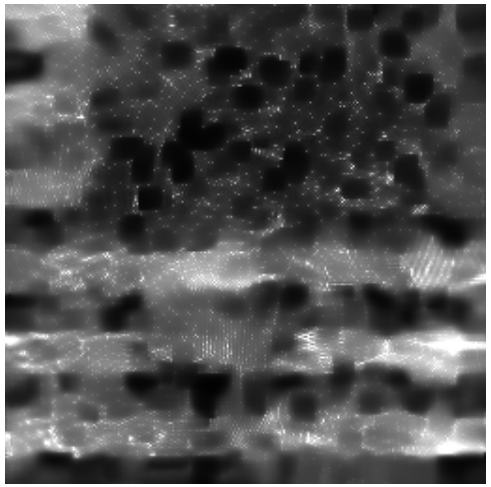
1. Assume $y \sim \text{Poisson}(f^*)$
2. Compute $z = \text{Anscombe}(y) \quad \Longrightarrow \quad z_i = 2\sqrt{y_i + \frac{3}{8}}$
3. Note $z_i \approx \mathcal{N}(2\sqrt{f_i^* + \frac{3}{8}}, 1)$
4. Denoise z using BM3D or other off-the-shelf image processing method optimized for Gaussian noise to get $\widehat{z}$
5. Compute $\widehat{f} = \text{Anscombe}^{-1}(\widehat{z})$

Extensions to inverse problems are quite challenging.

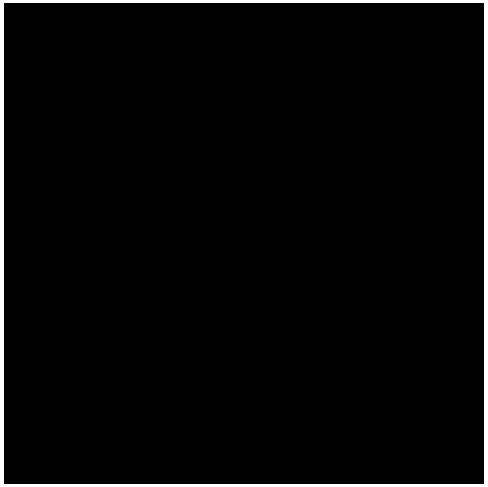
Anscombe transform in action[4]



Noisy data



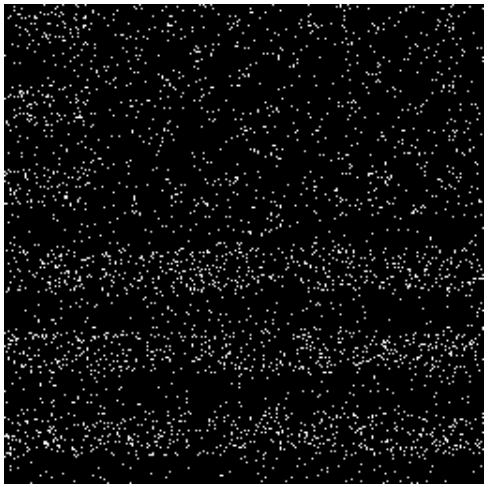
BM3D



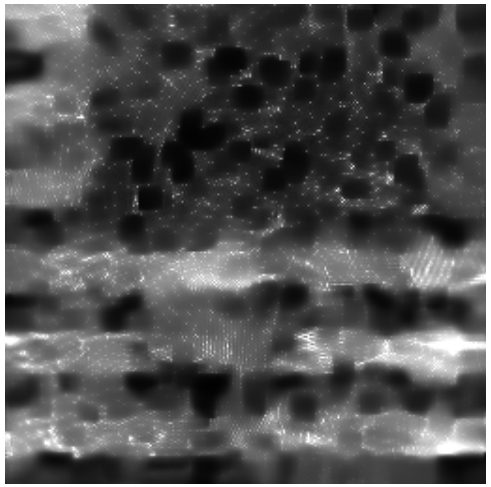
Anscombe + BM3D

[4] Mäkitalo & Foi (2011)

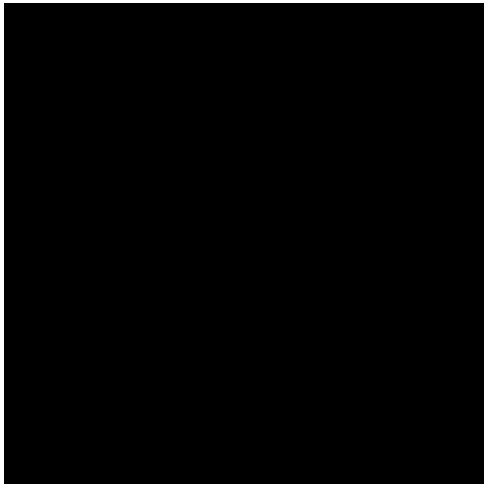
Anscombe transform in action[4]



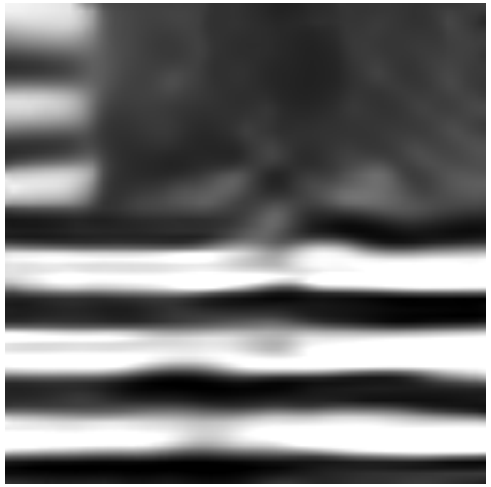
Noisy data



BM3D



Anscombe + BM3D



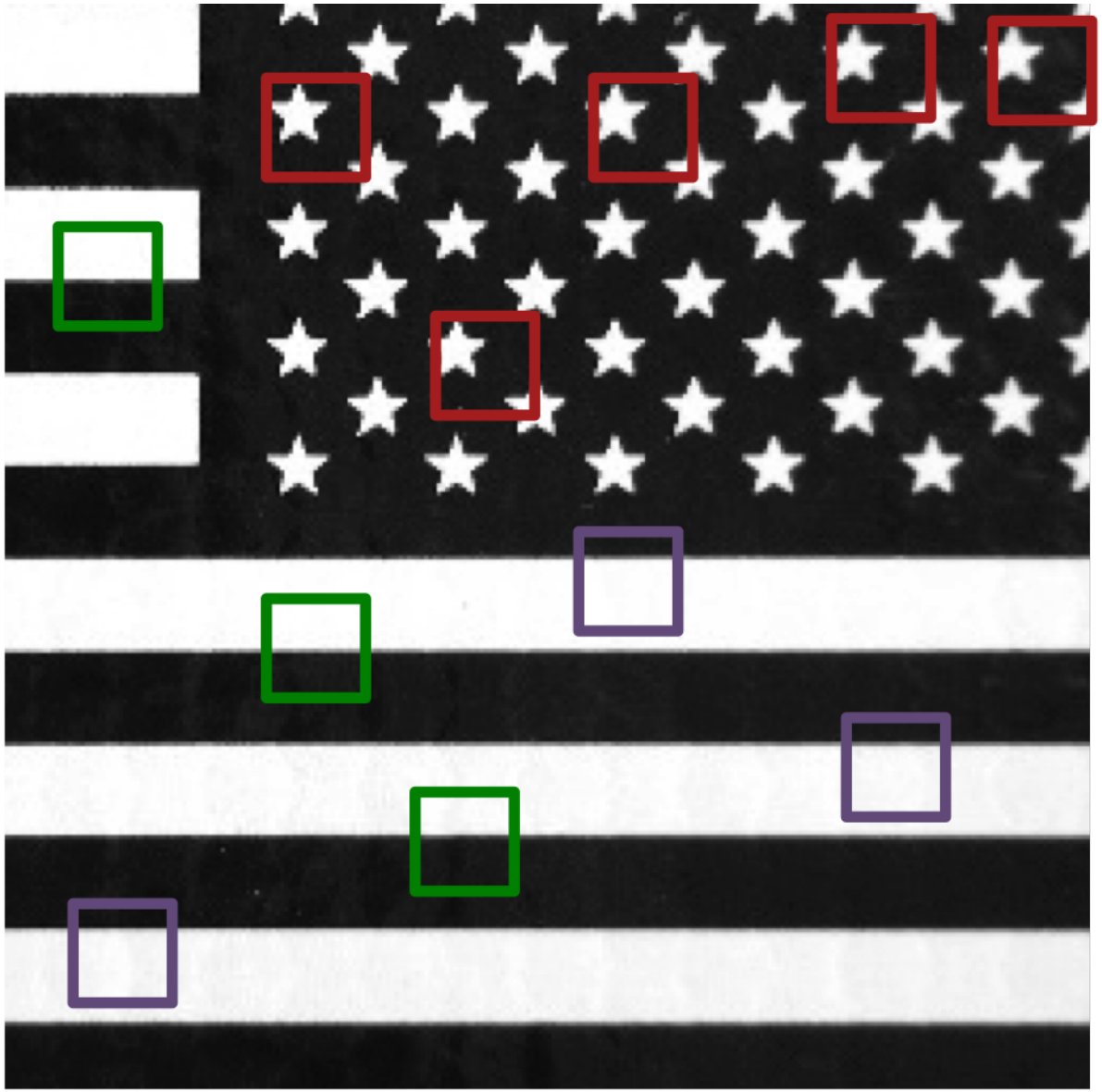
Improved Ansc. + BM3D

New methods for computing $\text{Anscombe}^{-1}(\cdot)$ make this approach more effective than ever.

Can we do better?

[4] Mäkitalo & Foi (2011)

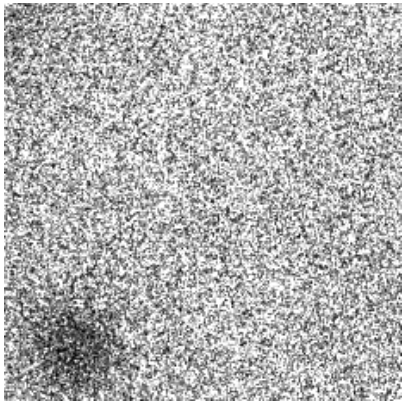
New sparsity models based on image patches



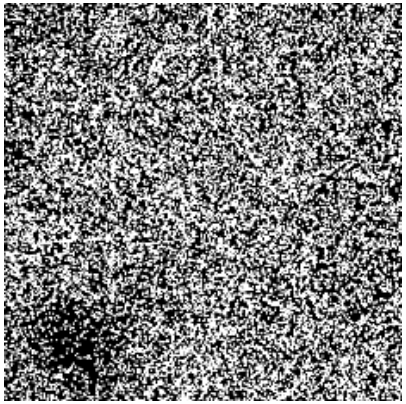
This insight underlies nonlocal means, BM3D, and dictionary learning

Case study: EDS imaging

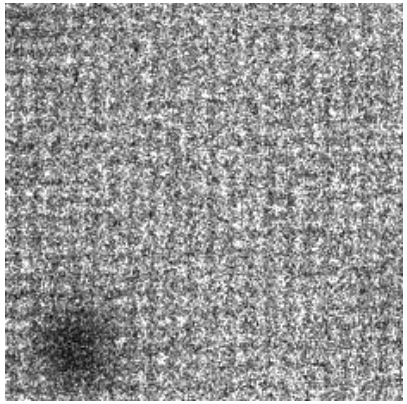
Calcium-doped neodinium titanate[5]



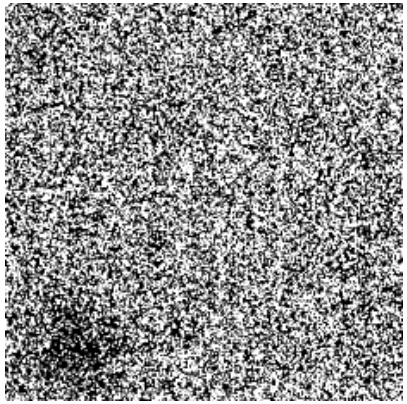
O Kα



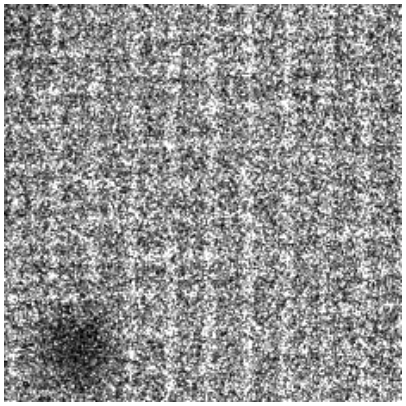
Ca Kα



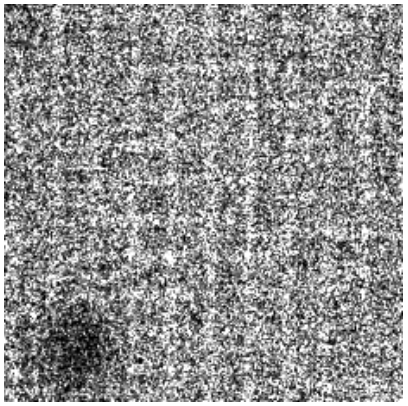
Ti Kα



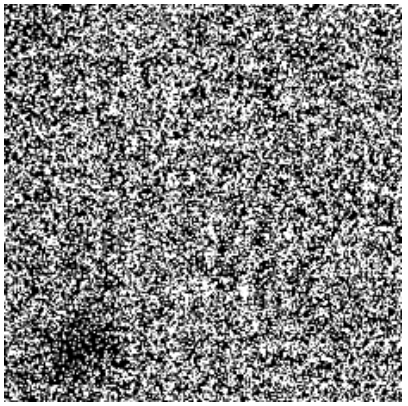
Ti Kβ



Nd Lα



Nd Lβ

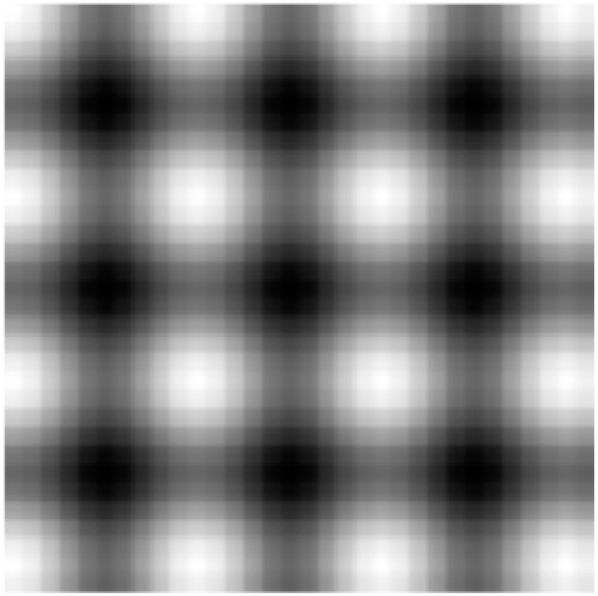


Nd Lβ4

[5] Raw data courtesy of Thomas Slater and Sarah Haigh at University of Manchester. Non-rigid alignment and averaging by Yankovich, Berkels, Dahmen, Binev, Sanchez, Bradley, Li, Szlufarska & Voyles (2014)

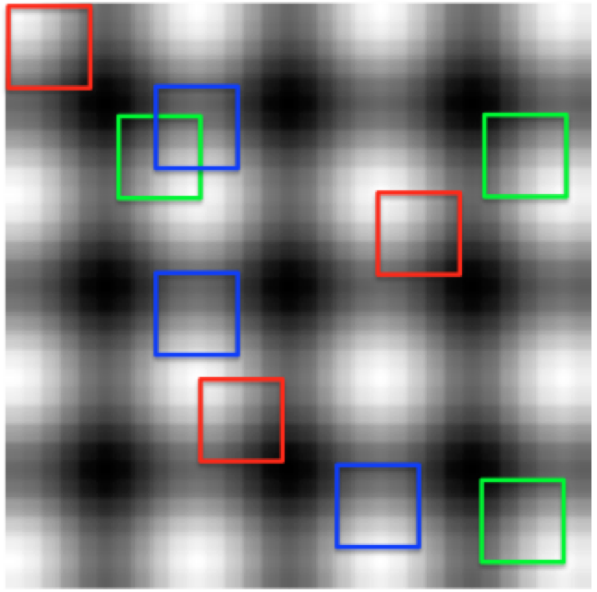
(Spectral) image model

Consider this phantom image:



(Spectral) image model

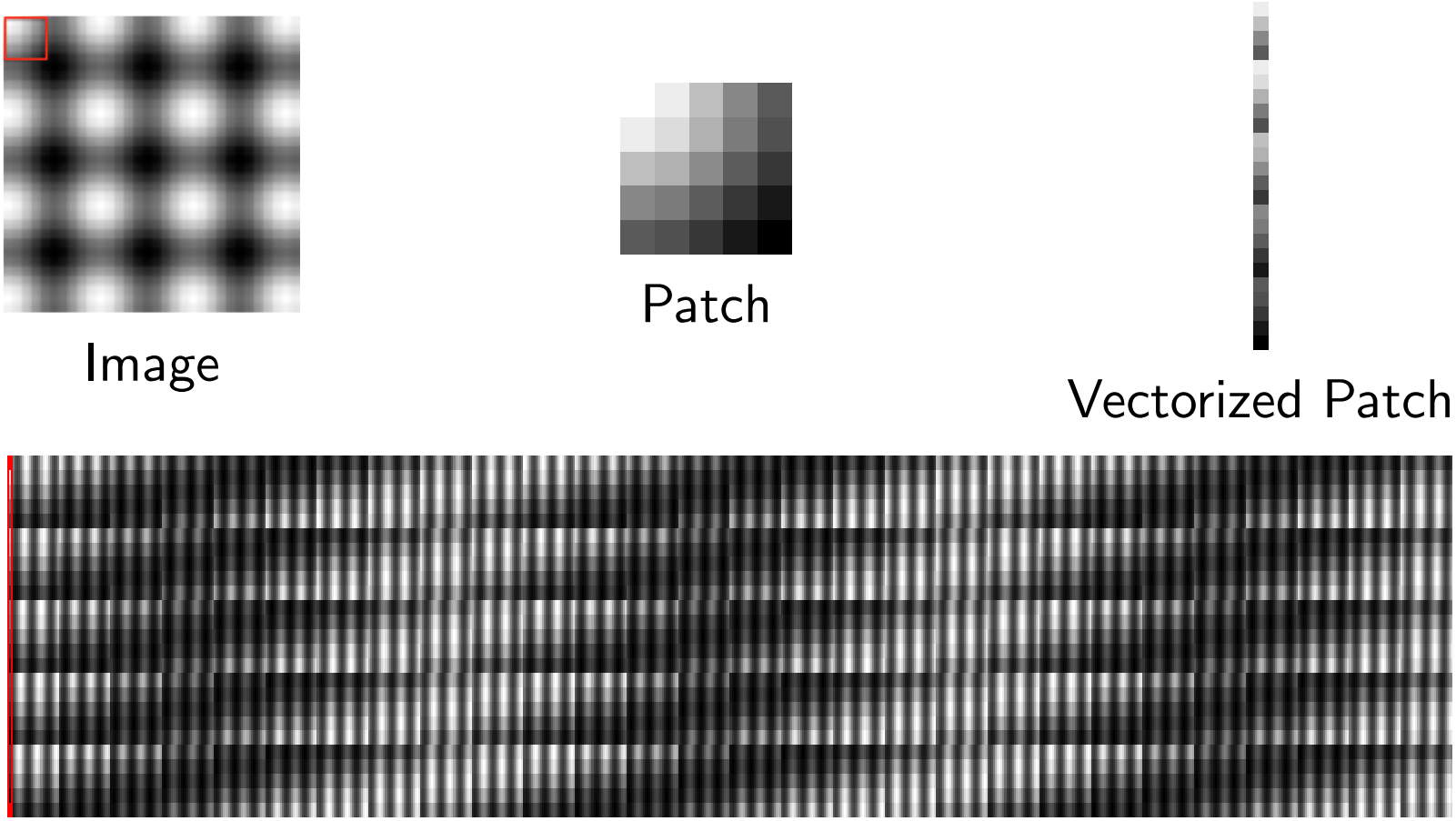
Consider this phantom image:



We want to exploit the redundancy in the image.

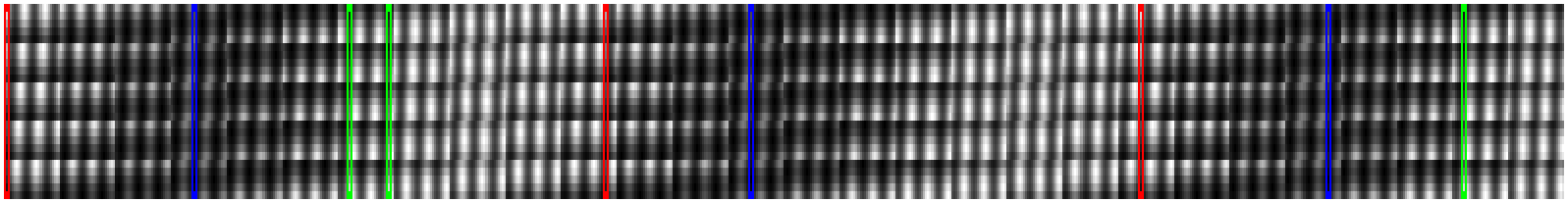
Key model idea: the (spectral) image patches lie in a union of subspaces

(Spectral) image model

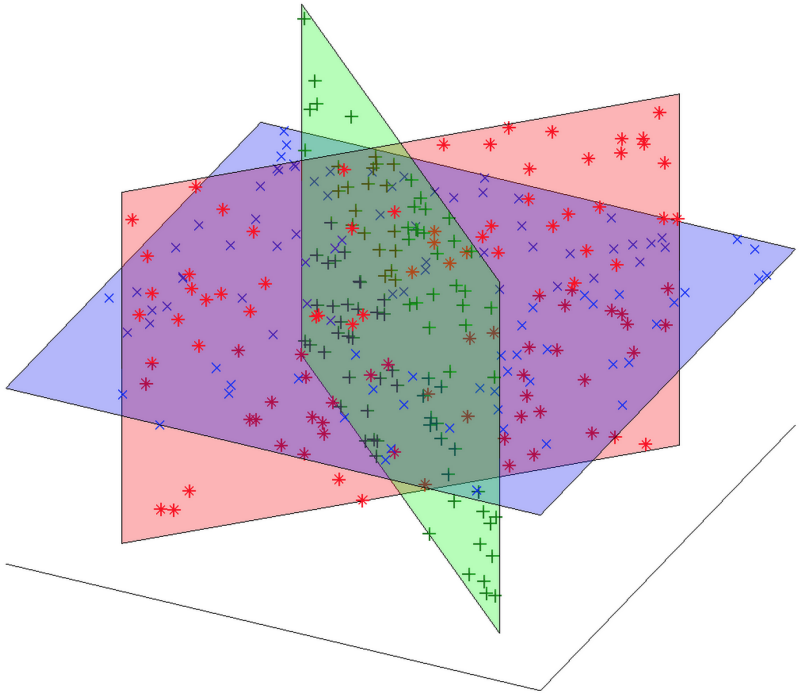


These ideas extend naturally to spectral images
(need to use 3-D patches)

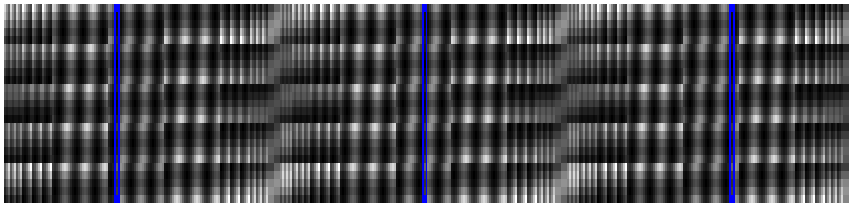
(Spectral) image model



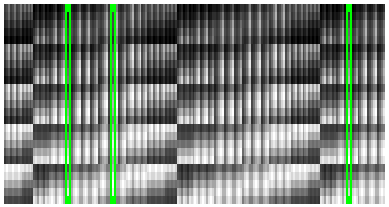
Collection of patches



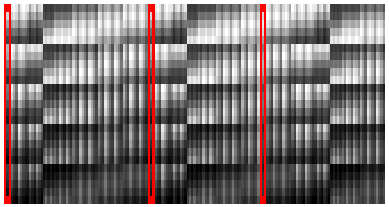
Union of subspaces



Cluster 1



Cluster 2



Cluster 3

Why a union of subspaces?

The union of subspaces model

- is a good fit to experimental data
- has limited degrees of freedom

Consider an image (spectral image) with n pixels (voxels) and patches of size p; usually $n \gg pK$:

PCA

- Estimates a single subspace of rank rK in p dimensions
- Need to estimate $p \times Kr$ principal components and $Kr \times n$ coefficients
- Total degrees of freedom is $O(rnK)$

Union of subspaces

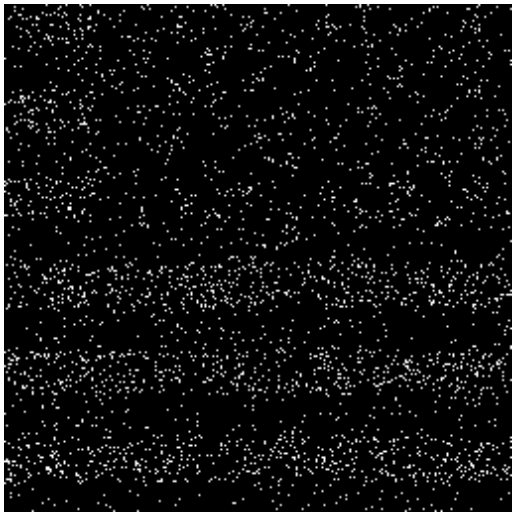
- Estimates K subspaces of rank r in p dimensions
- Need to estimate $p \times Kr$ principal components and $r \times n$ coefficients
- Total degrees of freedom is $O(rn)$

Nonlocal PCA for photon-limited imaging[6]

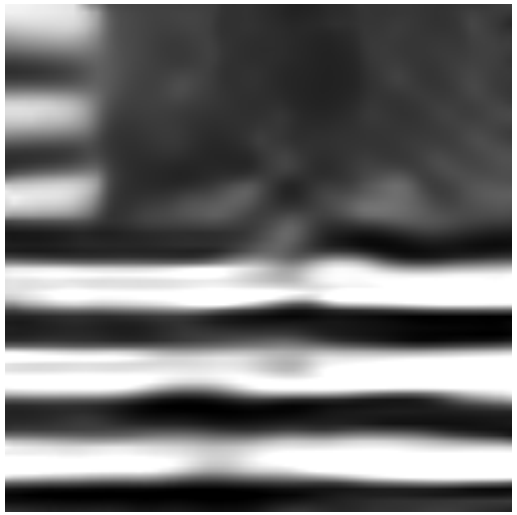
- Divide image into patches
- Cluster patches
  (using Poisson Bregman divergence to measure similarity of patches)
- Perform Poisson PCA on each cluster of patches to find low-dimensional patch subspace
  (by minimizing the negative Poisson log-likelihood with rank constraint)
- For each patch, estimate sparse PCA coefficients
  (by minimizing the negative Poisson log-likelihood + sparsity regularizer)

[6] Salmon, Deledalle, Harmany & Willett (2012)

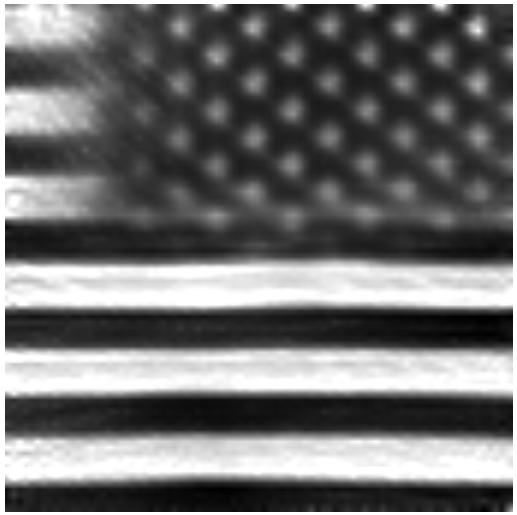
Do not underestimate the power of the dark side



Original scene

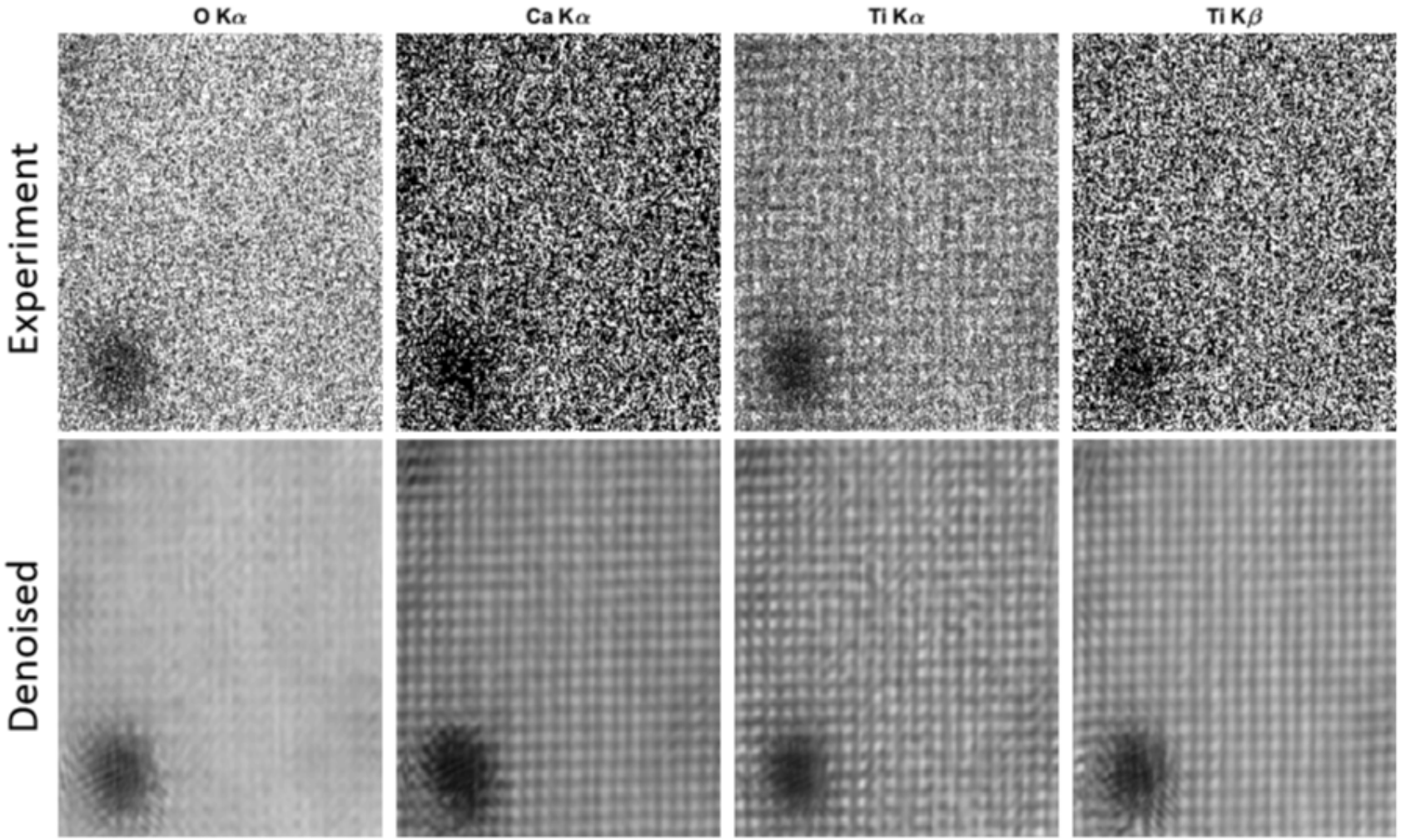


Anscombe + BM3D; PSNR = 13.91. Mäkitalo & Foi (2011)



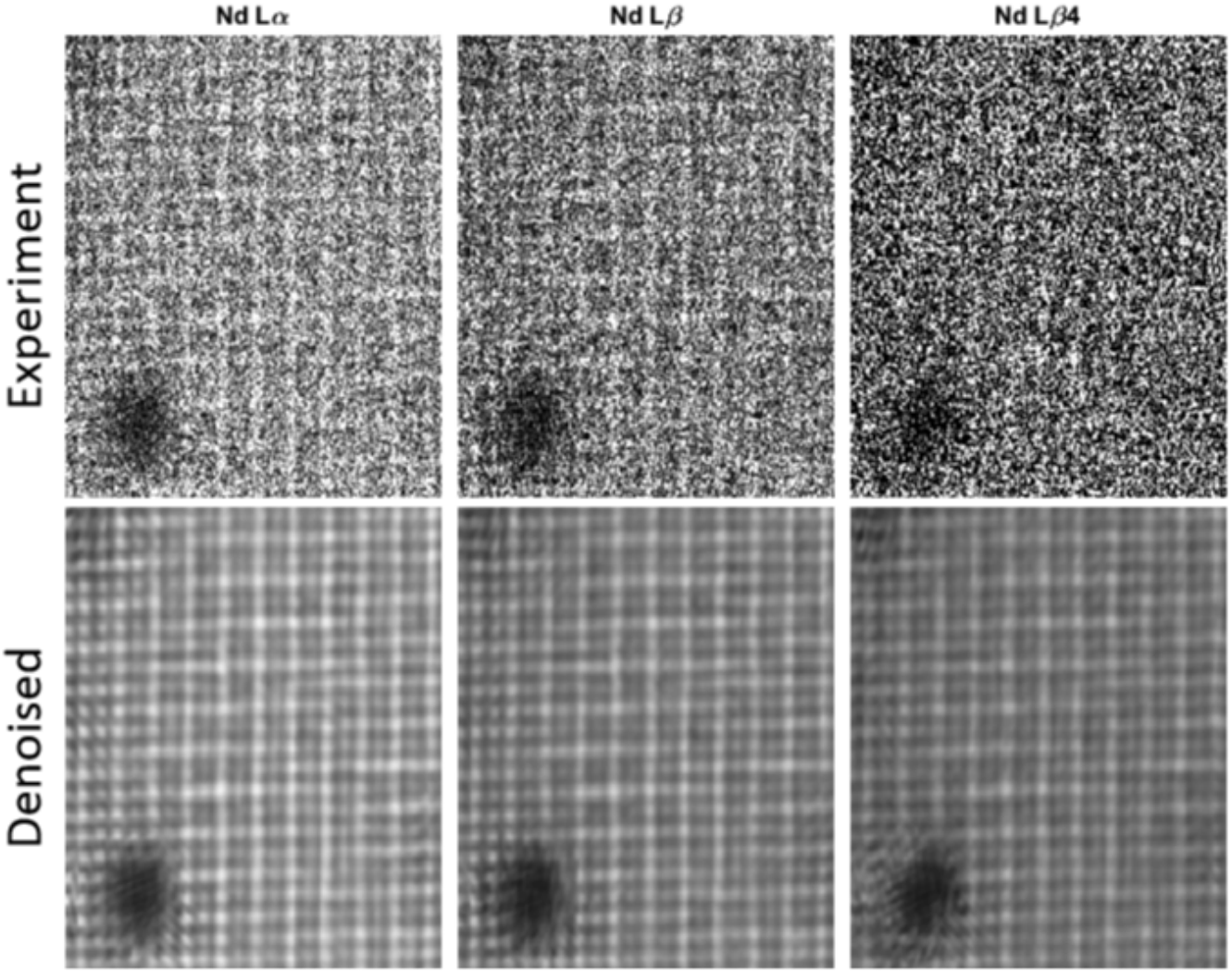
Poisson non-local PCA; PSNR = 15.99. Salmon, Deledalle, Harmany & Willett (2012)

EDS Spectral Imaging Experimental Results[7]



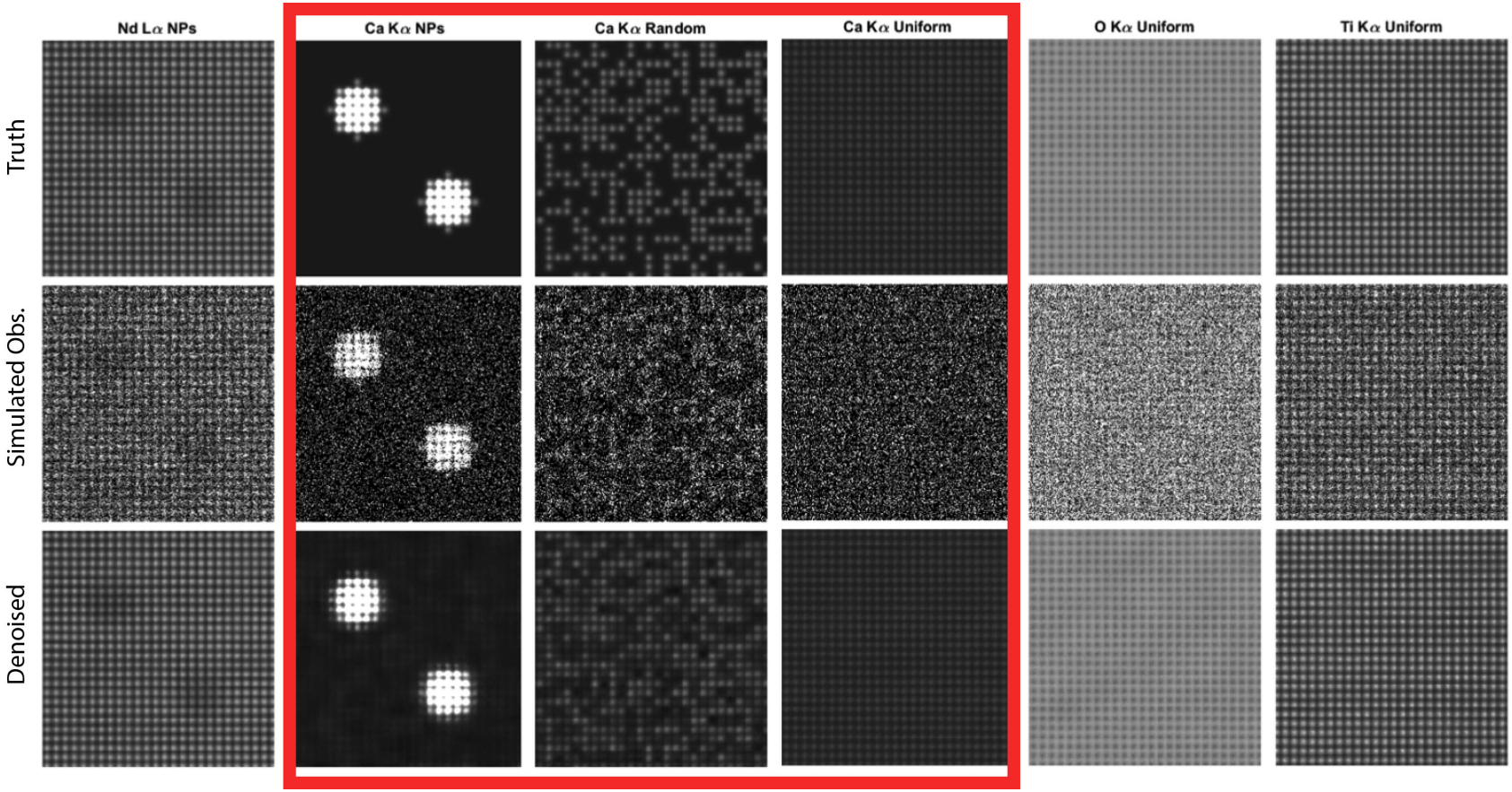
[7] Oh, Zhang, Yankovich, Willett, Voyles (2015)

EDS Spectral Imaging Experimental Results[8]



[8] Oh, Zhang, Yankovich, Willett, Voyles (2015)

EDS Spectral Imaging Simulation Results[9]

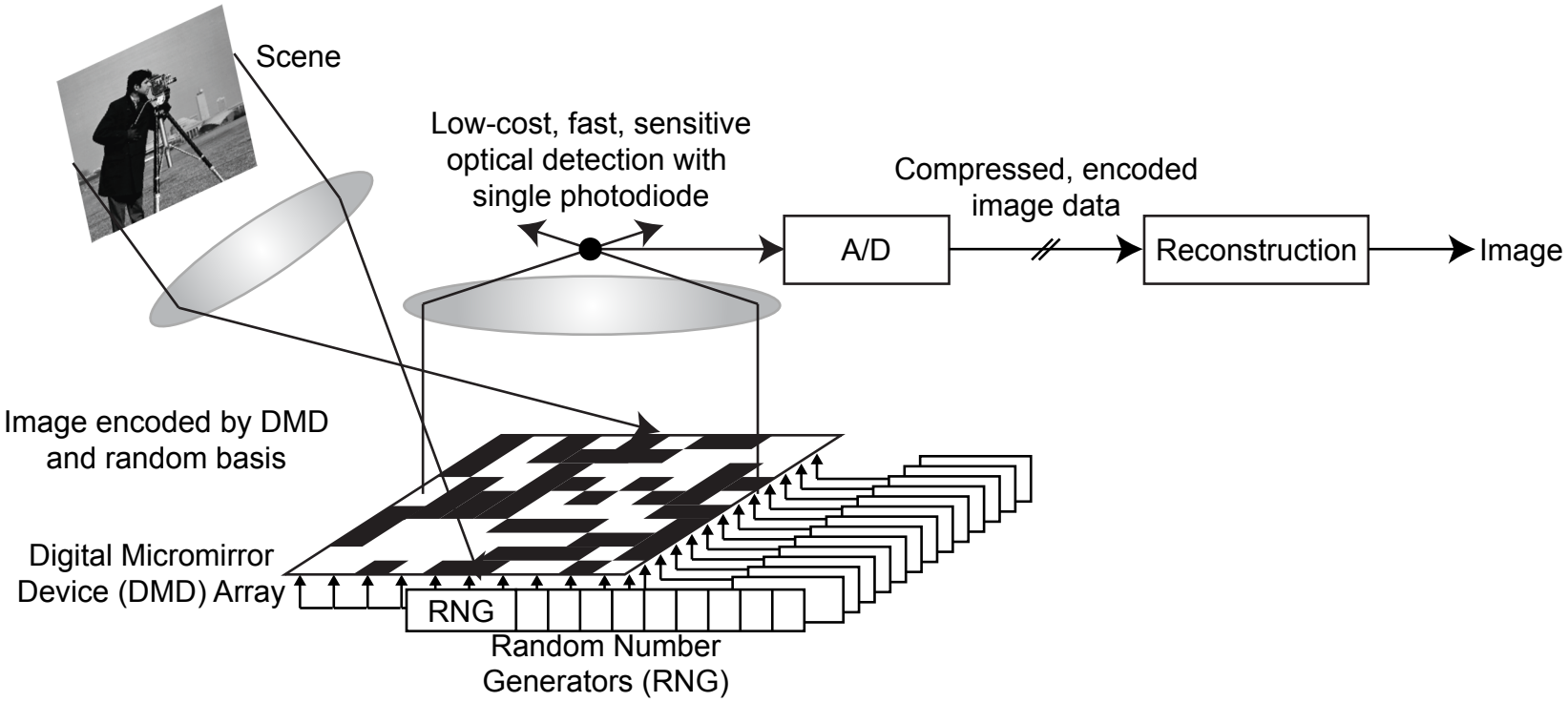


[9] Oh, Zhang, Yankovich, Willett, Voyles (2015)

Sparsity and low-dimensional structure allow us to perform accurate inference from small numbers of detected photons

How can we use this insight to solve inverse problems?

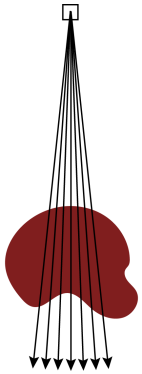
Compressive optical systems[10]



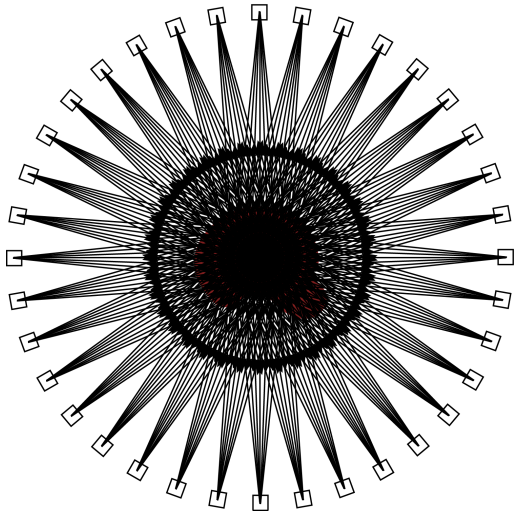
If we fix our total data acquisition time to T, then we have an explicit tradeoff between the number of projections, n, and the number of photons collected per projection, $O(T/n)$. As n increases, photon-limitations dominate errors.

[10] Duarte, Davenport, Laska, Sun, Takhar, Sarvotham, Baron, Wakin & Kelly, Baraniuk (2006)

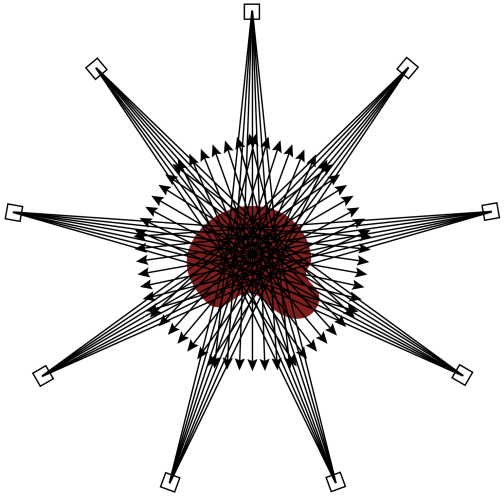
Tomography – how should we measure?



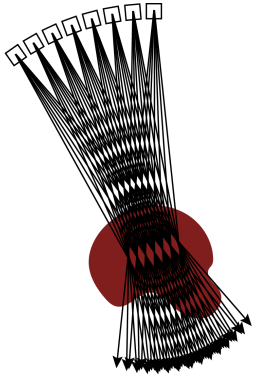
Tomographic projection



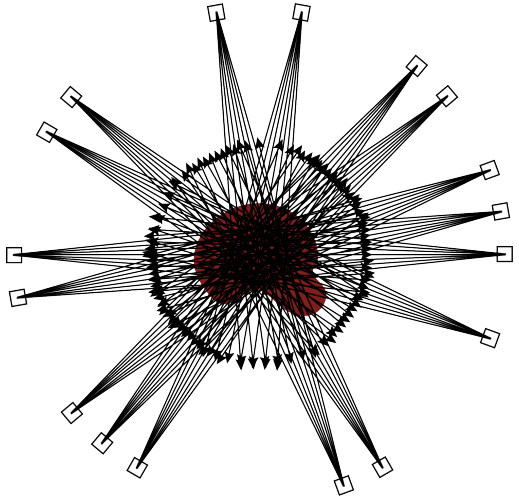
Full-angle



Sparse-angle



Limited-angle



Random-angle

Sensing model

We observe

$$y \sim \text{Poisson}(TAf^*)$$

$$y_i \sim \text{Poisson}\left(T\sum_{j=1}^{p} A_{i,j} f_j^*\right), \qquad i \in \{1,\ldots,n\},$$

where

- $y \in \mathbb{Z}_+^n$
- $T \in \mathbb{R}_+$ is the total data acquisition time
- $A \in [0,1]^{n\times p}$ is a known sensing matrix
- $f^* \in \mathcal{F}$, where

$$\mathcal{F} = \left\{ f \in \mathbb{R}_+^p \ : \ \|f\|_1 = 1, \ \|D^\top f\|_0 \le s+1 \right\}$$

for an orthonormal basis D

Our goal is to reconstruct f^* from y. How does performance depend on n, p, T, D, and A? What fundamentally limits our sensing capabilities?

1. Previous work established upper bounds; were these bounds tight?
2. Is it better to have a lot of high-noise measurements (big n), or a few low-noise measurements?
3. What is the impact of Poisson noise on inverse problems?
4. How should we compute estimates?

This is not your ordinary CS problem

Sensing matrix A has several physical constraints

Think of $A_{i,j}$ as likelihood of photon from location j in f^* hitting detector at location i:

$$
\begin{aligned}
A_{i,j} &\in [0,1] \\
\mathbb{1}^\top A &\preceq \mathbb{1} && \text{(columns sum to at most one)} \\
\|Af\|_1 &\leq \|f\|_1 && \forall f
\end{aligned}
$$

Typical CS sensing matrices do not satisfy these constraints!

Sensing matrix

Start with a sensing matrix $\widetilde{A} \in \frac{1}{2\sqrt{n}}\{-1,1\}^{n\times p}$ such that the product $\widetilde{A}D$ satisfies the RIP:

$$(1-\delta_s)\|\theta\|_2^2 \leq \|\widetilde{A}D\theta\|_2^2 \leq (1+\delta_s)\|\theta\|_2^2 \quad \forall \quad 2s - \text{sparse } \theta$$

Let

$$A \triangleq (\widetilde{A} + \frac{3}{\sqrt{n}}\mathbb{1}_{n\times p})/4\sqrt{n}.$$

Sensing matrix

Start with a sensing matrix $\widetilde{A} \in \frac{1}{2\sqrt{n}}\{-1,1\}^{n\times p}$ such that the product $\widetilde{A}D$ satisfies the RIP:

$$(1-\delta_s)\|\theta\|_2^2 \leq \|\widetilde{A}D\theta\|_2^2 \leq (1+\delta_s)\|\theta\|_2^2 \quad \forall \quad 2s-\text{sparse } \theta$$

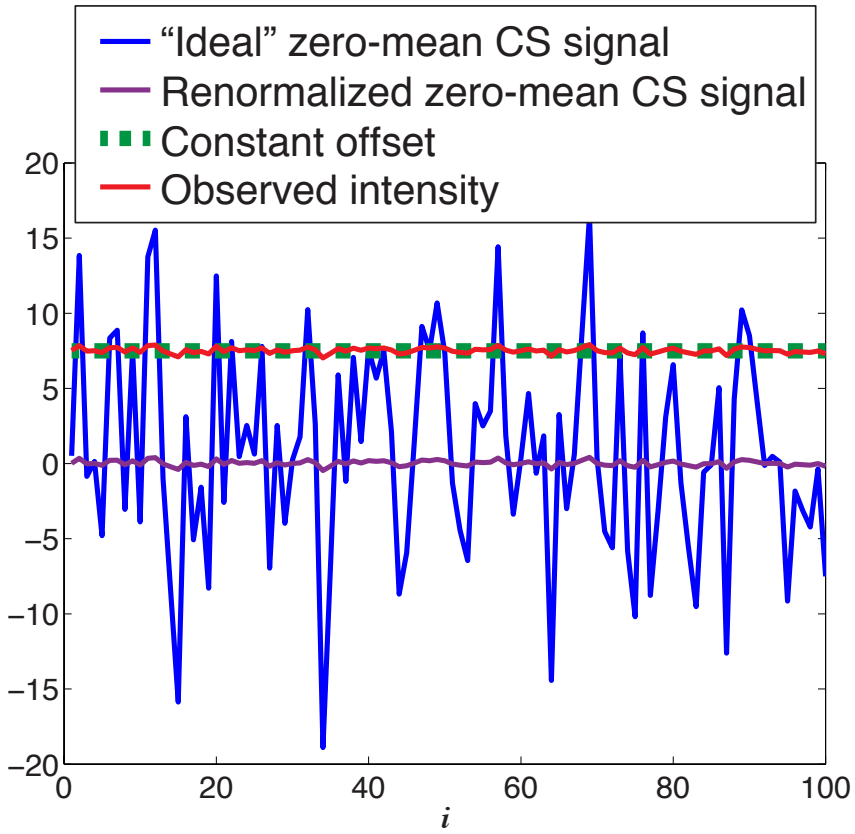
Let

$$A \triangleq (\widetilde{A} + \frac{3}{\sqrt{n}}\mathbb{1}_{n\times p})/4\sqrt{n}.$$

We observe

$$y \sim \text{Poisson}(TAf^*)$$

$$\sim \text{Poisson}\Big(\frac{T\widetilde{A}f^*}{4\sqrt{n}} + \underbrace{\frac{3T}{4n}\mathbb{1}_{n\times 1}}_{\text{determines variance}}\Big)$$



Rates for high-intensity settings (large T) [11]

Theorem:

$$\inf_{\widehat{f}} \sup_{f^* \in \mathcal{F}} \mathbb{E}[\|\widehat{f} - f^*\|_2^2] \asymp \frac{s \log p}{T}$$

where

$$\mathcal{F} = \left\{ f \in \mathbb{R}_+^p \ : \ \|f\|_1 = 1, \ \|D^T f\|_0 \leq s + 1 \right\}$$

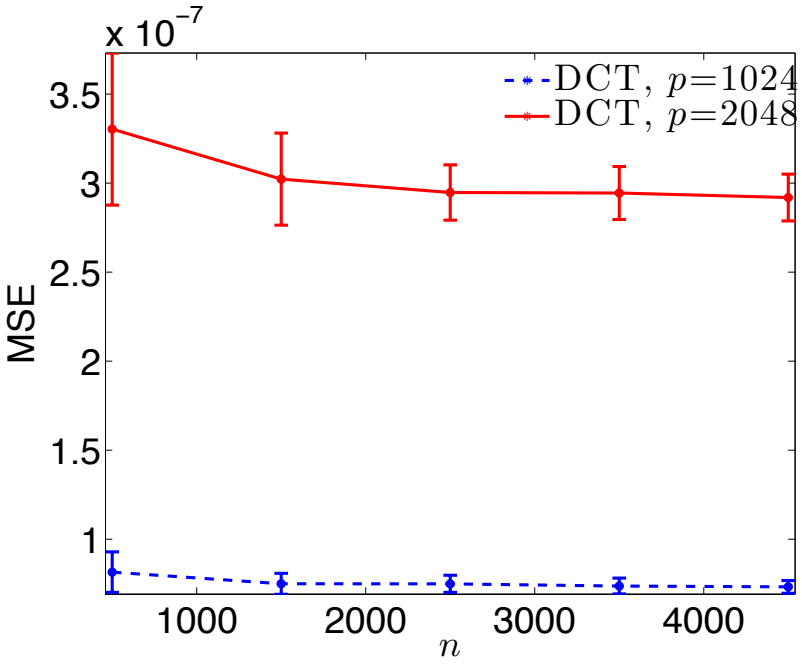
- The data acquisition time T, which reflects the signal-to-noise ratio, controls the error decay
- Once the number of measurements, n, is sufficiently large to ensure a RIP-like sensing matrix, it does not impact errors

[11] Jiang, Raskutti & Willett (2014)

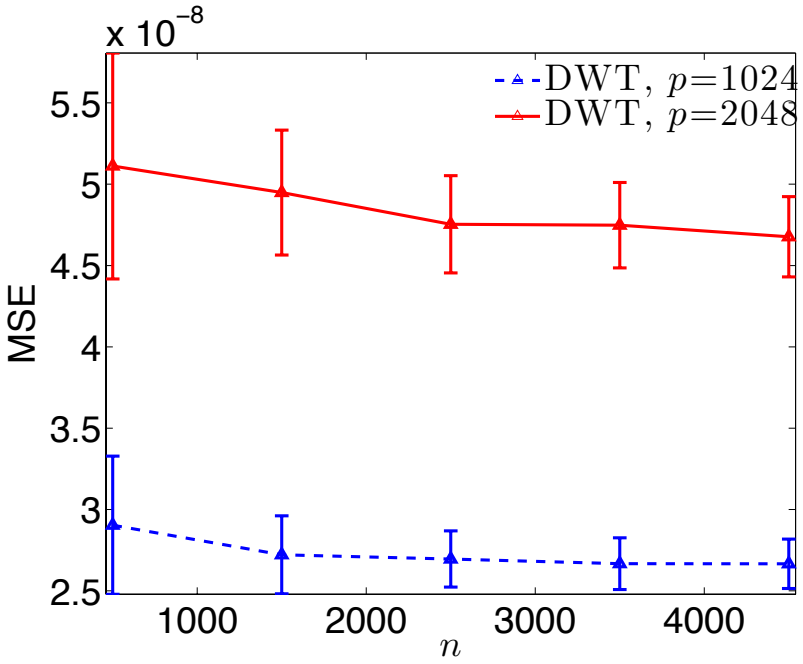
MSE vs. measurements

If we fix our total data acquisition time to T, then we have an explicit tradeoff between the number of projections, n, and the number of photons collected per projection, $O(T/n)$. As n increases, photon-limitations dominate errors.

Is it better to have a lot of high-noise measurements (big n), or a few low-noise measurements?

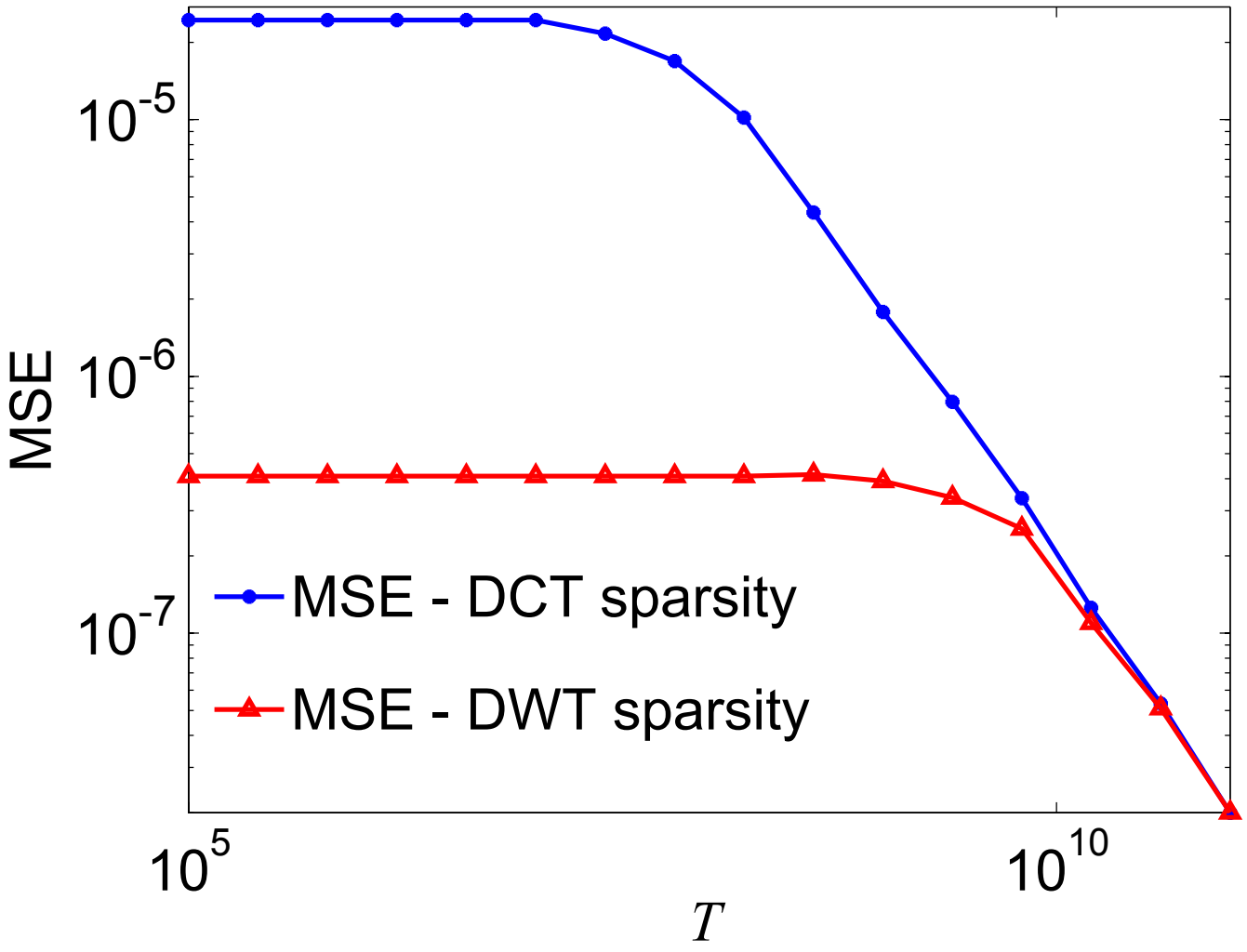


DCT sparsity



DWT sparsity

MSE vs. T: An elbow in the rates



So far we have only considered high-intensity (large T) settings. What happens in low intensities?

Low-intensity settings (small T) [12]

Let $\overline{f^*} \equiv \mathbb{1}_{p\times 1}/\sqrt{p}$ be the average of f^*. Then

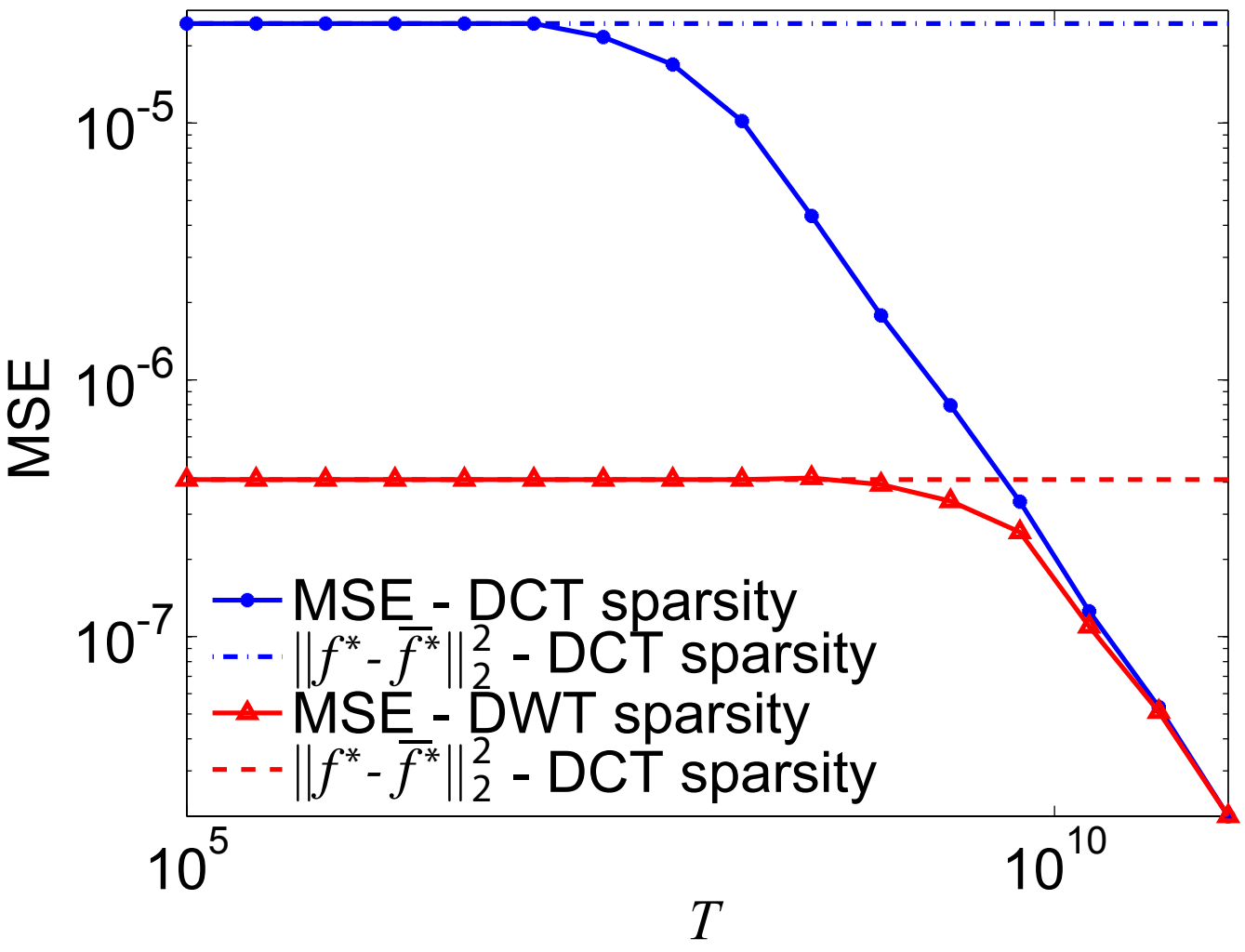
$$\mathbb{E}[\|\widehat{f} - f^*\|_2^2] \asymp \|f^* - \overline{f^*}\|_2^2$$

Rates depend on how much f^* deviates from its mean ("residual energy"), subject to the constraint that $\|f^*\|_1 = 1$ for $f^* \in \mathcal{F}$.

For different sparsifying bases D, this residual energy falls in different ranges, giving different rates.

[12] Jiang, Raskutti & Willett (2014)

MSE vs. T: An elbow in the rates

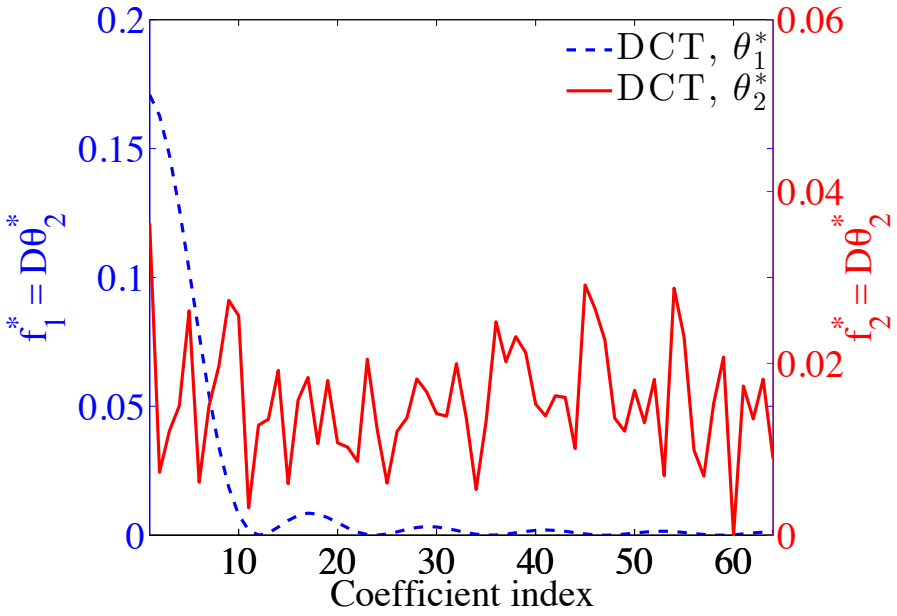


Low-intensity rates in two common bases[13]

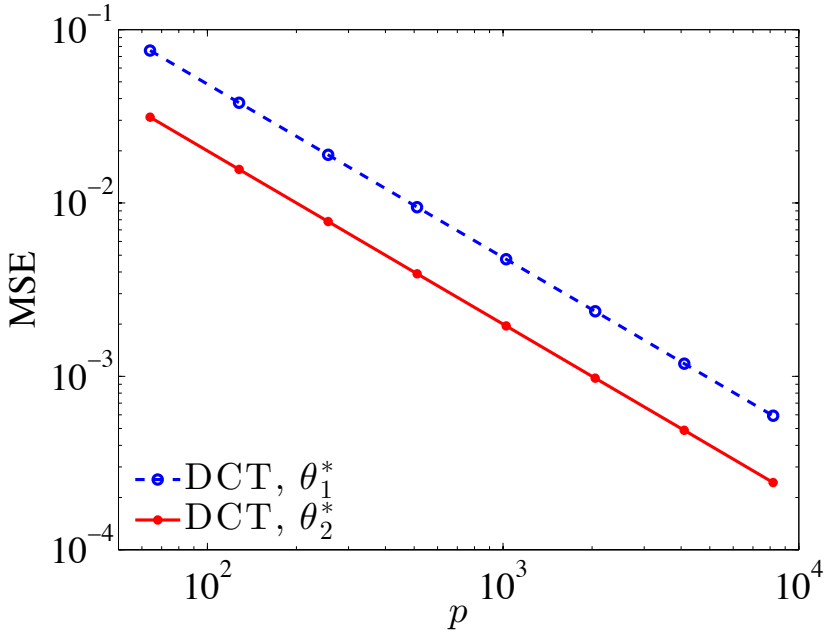
	Lower bound	Upper bound
DCT	$\frac{1}{p}$	$\frac{s}{p}$
DWT	$\frac{s}{p^2}$	1
	Dominated by flattest s-sparse signals in $\mathcal{F}$	Dominated by peakiest s-sparse signals in $\mathcal{F}$

[13] Jiang, Raskutti & Willett (2014)

Fourier sparsity



Two signals in our function class...

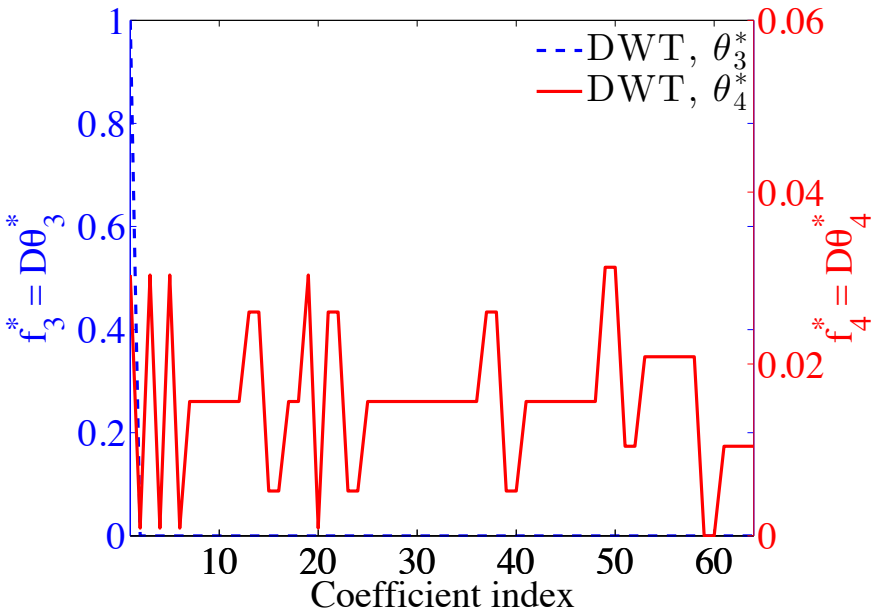


...and their rates

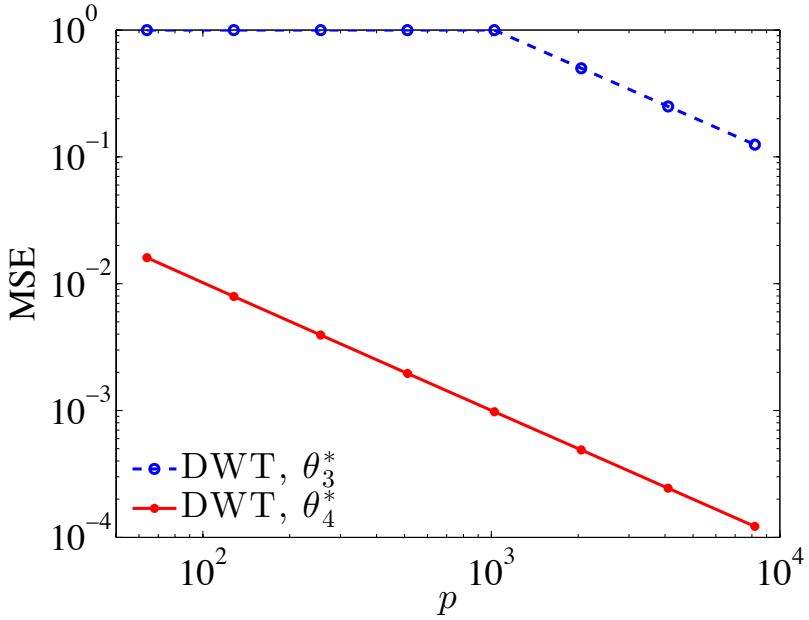
Red signal (residual energy 0.0191) controls lower bounds, blue signal controls upper bounds (residual energy 0.1151)

As predicted by theory, both signals have same scaling with p but different scaling with s

Wavelet sparsity



Two signals in our function class...



...and their rates

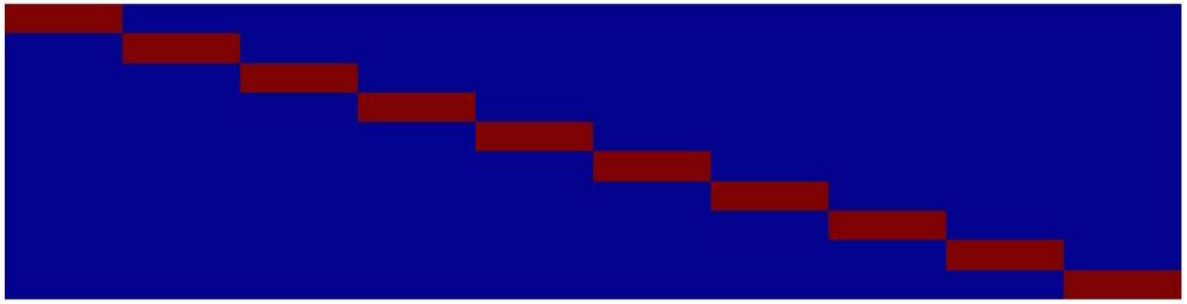
Red signal (residual energy 0.0188) controls lower bounds, blue signal controls upper bounds (residual energy 1.0000)

As predicted by theory, signal with high amplitude yields much slower rates than more diffuse signal

CS can be suboptimal at low intensities

Consider the special case where our signal is sparse in a DWT basis, and s' of the s nonzero coefficients are at coarse scales.

We will compare the CS paradigm from earlier with a simple downsampling system A^{DS}:



Measuring $A^{\mathrm{DS}} f^*$ is equivalent to measuring coarse-scale Haar wavelet coefficients of f^*.

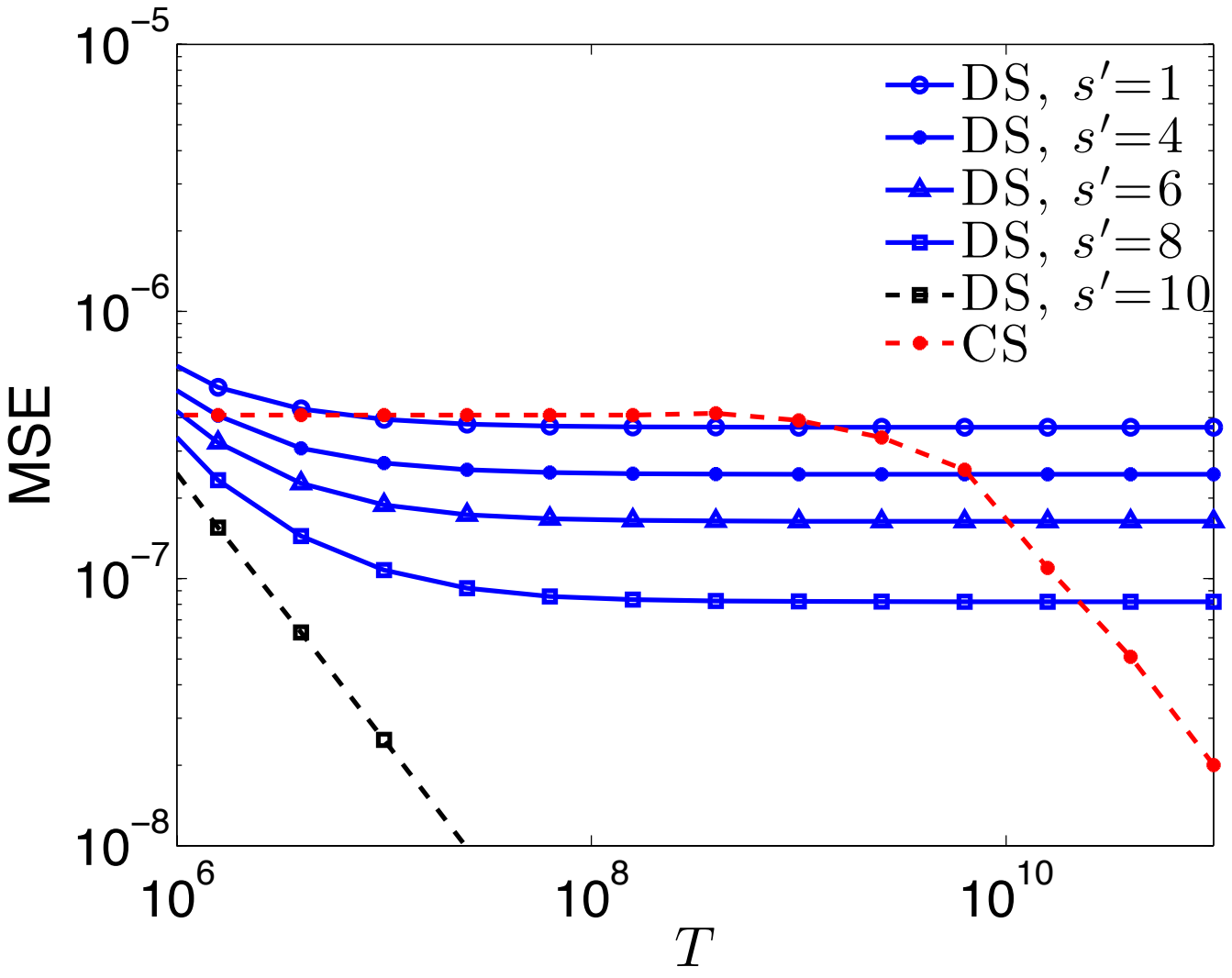
CS can be suboptimal at low intensities

Consider the special case where our signal is sparse in a Haar basis, and s' of the s nonzero coefficients are at coarse scales.

Compressive sampling	Direct measurement of coarse-scale info
$\min\left(\frac{s\log p}{T}, \frac{s}{p^2}\right),$	$\frac{s'n}{Tp^2} + \frac{s-s'}{p^2}$
independent of s'	dependent on s'

These curves cross; for small T and moderate s', downsampling can perform significantly better than compressive sampling.

Compressive vs. direct measurements



Ramifications

CS conventional wisdom (for Gaussian noise settings) tells us rates are

- Independent of sparsifying basis
- Not much worse than if we collected non-compressive measurements

In Poisson noise settings, because of the interaction between physical constraints and sparsity assumptions

- Rates are highly dependent on sparsifying basis
- Depending on the sparsity assumptions, we can do far better using non-compressive measurements

Computational methods for Poisson inverse problems

Option 0: LASSO

Recall $y \sim \text{Poisson}(TAf^*)$.

$$\hat{f}^{\mathrm{L}} = \arg\min_f \frac{1}{2}\|y - TAf\|_2^2 + \gamma\|D^\top f\|_1.$$

This approach ignores the effects of Poisson noise.

Option 1: penalized weighted least squares[14]

Recall $y \sim \text{Poisson}(TAf^*)$.

$$\hat{f}^{\text{WLS}} = \arg\min_f \frac{1}{2}\|\Sigma^{-1}(y - TAf)\|_2^2 + \gamma\|D^\top f\|_1,$$

where Σ^{-1} is a diagonal matrix with $\Sigma_{ii} \approx \text{std}(y_i)$.

Challenge is choosing Σ without knowing f^*.

- In tomography Σ is estimated by smoothing the noisy sinogram
- Does not generalize well
- No known statistical error bounds

[14] Fessler, 1994

Option 2: regularized maximum likelihood[15]

Recall $y \sim \text{Poisson}(TAf^*)$.

$$\hat{f}^{\mathrm{ML}} = \arg\min_f -\log p(y|TAf) + \gamma \|D^\top f\|_1.$$

Unfortunately, this approach yields technical challenges:

- difficult to analyze
  - most analysis based on ℓ_0 regularizer
  - critical to ensure Af bounded away from zero
- slower to compute
  - gradient is not Lipschitz
  - off-the-shelf methods slow to converge
  - need specialized tools

[15] Raginsky, Willett, Harmany, & Marcia (2010);
Raginsky, Jafarpour, Harmany, Marcia, Willett, & Calderbank (2011)
Jiang, Raskutti, & Willett (2015)

Proposition: *Weighted* LASSO[16]

Recall $y \sim \text{Poisson}(TAf^*) = \text{Poisson}(TADx^*)$ where $f^* \equiv Dx^*$.

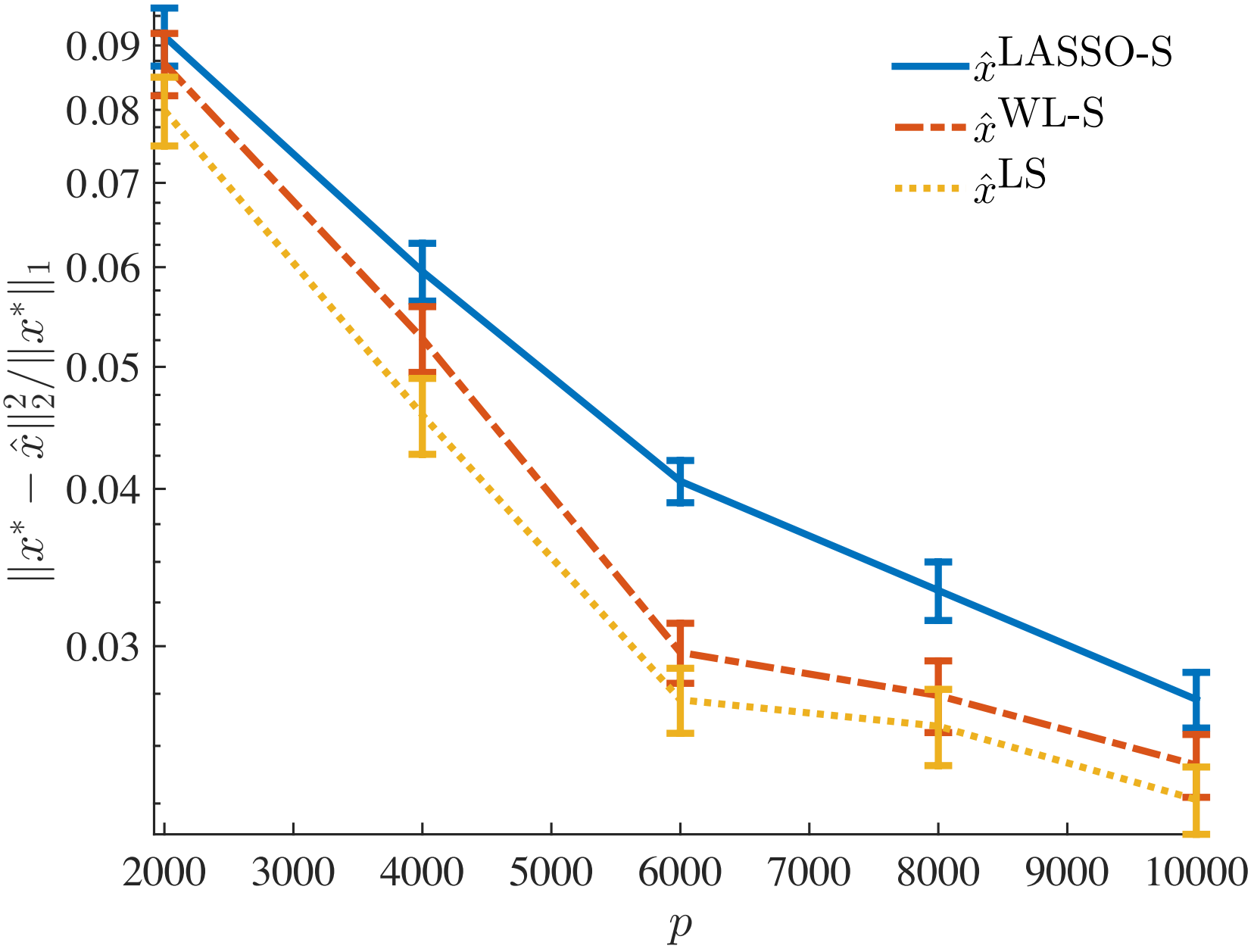
The weighted LASSO estimator has the form

$$\begin{aligned}\widehat{x}^{\mathrm{WL}} &= \arg\min_x \frac{1}{2}\|y - TADx\|^2 + \gamma \sum_{k=1}^{p} w_k |x_k| \\ \widehat{f}^{\mathrm{WL}} &= D\widehat{x}^{\mathrm{WL}}\end{aligned}$$

- Easy to compute with off-the-shelf LASSO software
- Data-dependent weights handle Poisson noise
- New statistical error bounds
  - Consistent with previous estimator performance bounds
  - Much easier to generalize

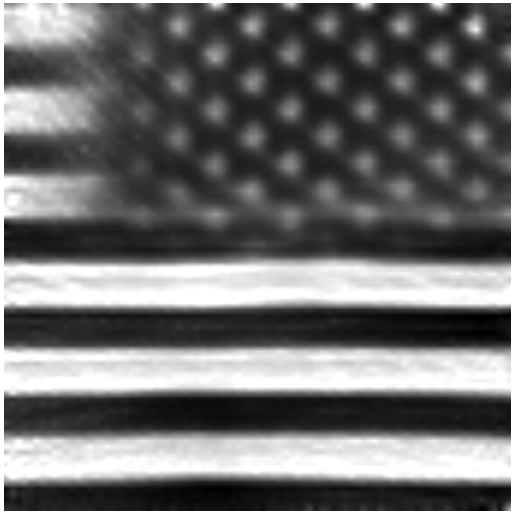
[16] van de Geer, Bühlmann, & Zhou, 2011; Bickel, Ritov, & Tsybakov, 2009

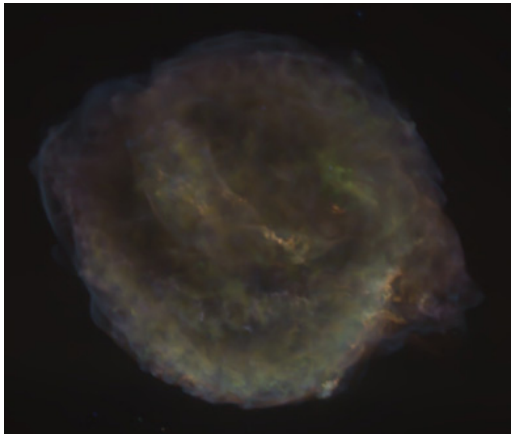
Simulation results



Photon-limited imaging

- Poisson noise important because fewer photons = less data collection time, more temporal resolution
- Generalized sparsity models allow accurate inference from limited photon count data
- Both theory and methods face myriad of tradeoffs related to
  - SNR
  - Sparsity
  - Efficiency of computation
  - Measurements system design





Become seduced by the dark side of imaging

Exciting mathematical challenges at the forefront of sensor technology

http://willett.ece.wisc.edu/