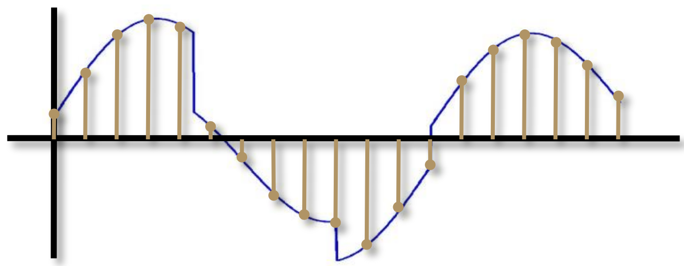
Sparsity and Structure in Imaging

Mark A. Davenport

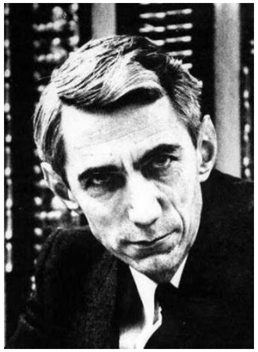
Georgia Institute of Technology
School of Electrical and Computer Engineering

Digital Revolution



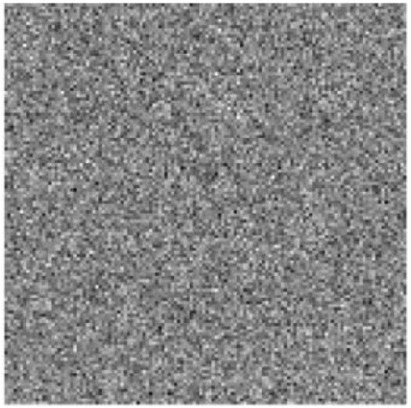
"If we sample a signal at twice its highest frequency, then we can recover it exactly."

Whittaker-Nyquist-Kotelnikov-Shannon



Dimensionality Reduction

Data with high-frequency content is often not intrinsically high-dimensional



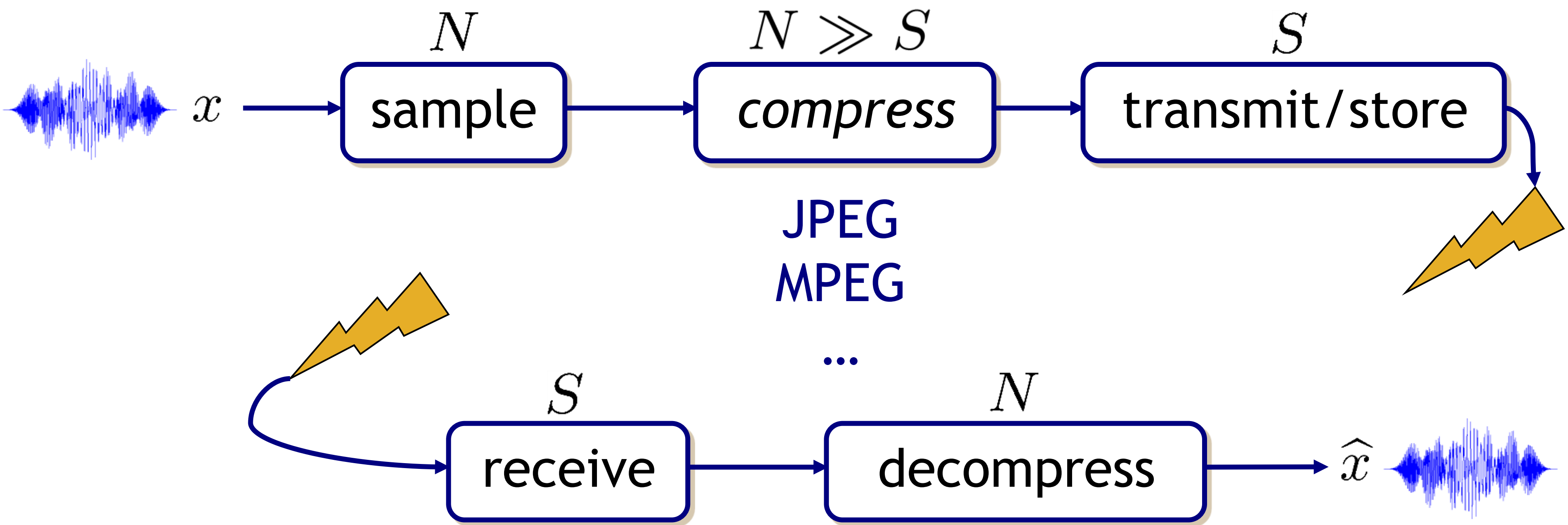
Signals often obey *low-dimensional models*

- sparsity
- manifolds
- low-rank matrices

The "intrinsic dimension" S can be much less than the "ambient dimension" N

Sample-Then-Compress Paradigm

- Standard paradigm for digital data acquisition
  - *sample* data (ADC, digital camera, ...)
  - *compress* data (signal-dependent, nonlinear)

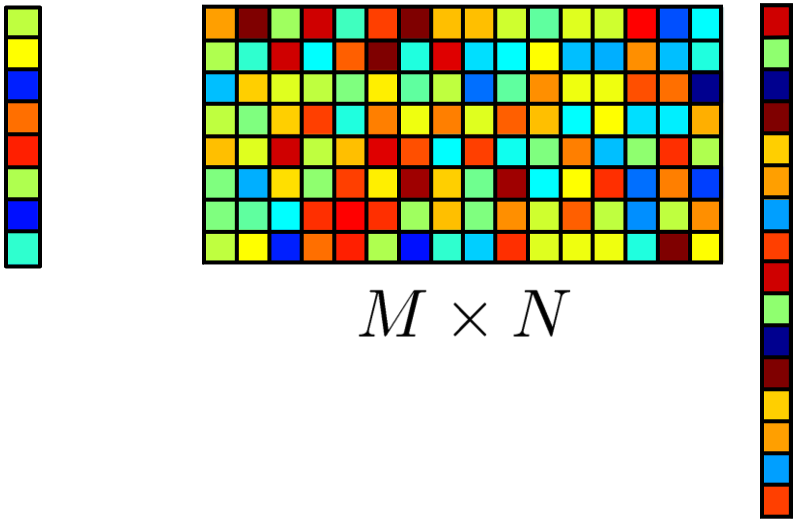


- Sample-and-compress paradigm is *wasteful*
  - samples cost $$$ and/or time

Compressive Sensing

Replace samples with general *linear measurements*

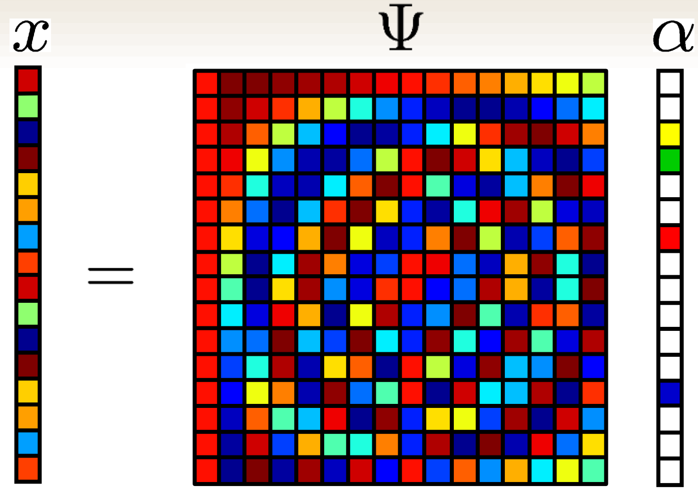
$$y = \Phi x$$



[Donoho; Candès, Romberg, Tao - 2004]

Sparsity

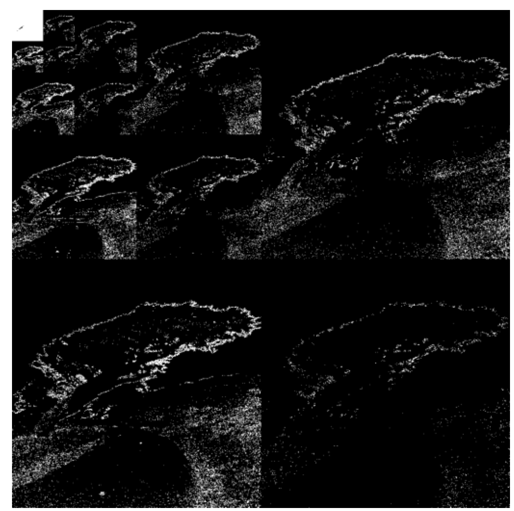
$$x = \sum_{j=1}^{N} \alpha_j \psi_j$$
$$= \Psi \alpha$$



S nonzero entries

$\|\alpha\|_0 = S$

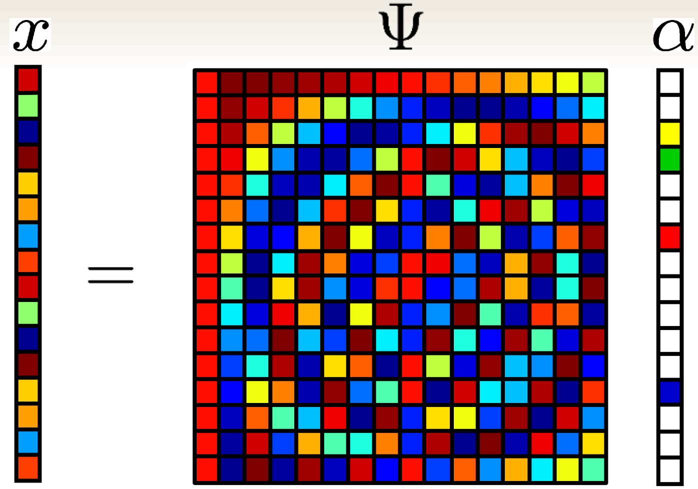
N pixels



$S \ll N$ large wavelet coefficients

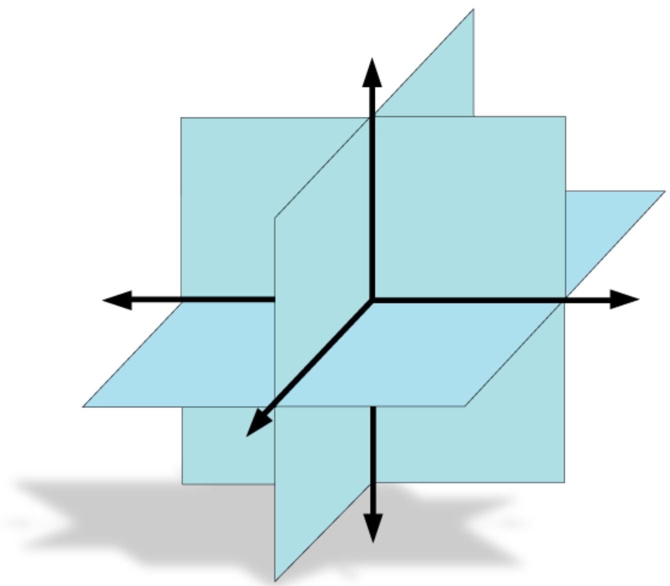
Sparsity

$$x = \sum_{j=1}^{N} \alpha_j \psi_j$$
$$= \Psi\alpha$$



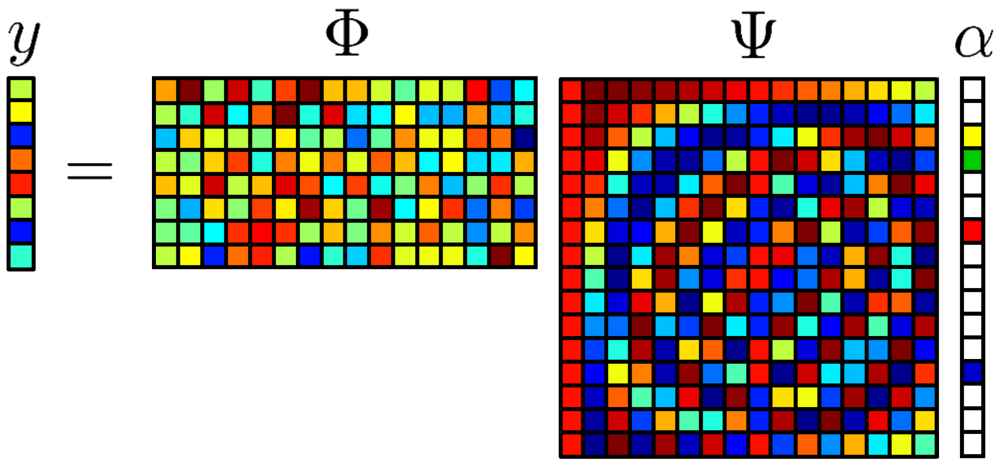
S nonzero entries

$$\|\alpha\|_0 = S$$



Core Theoretical Challenges

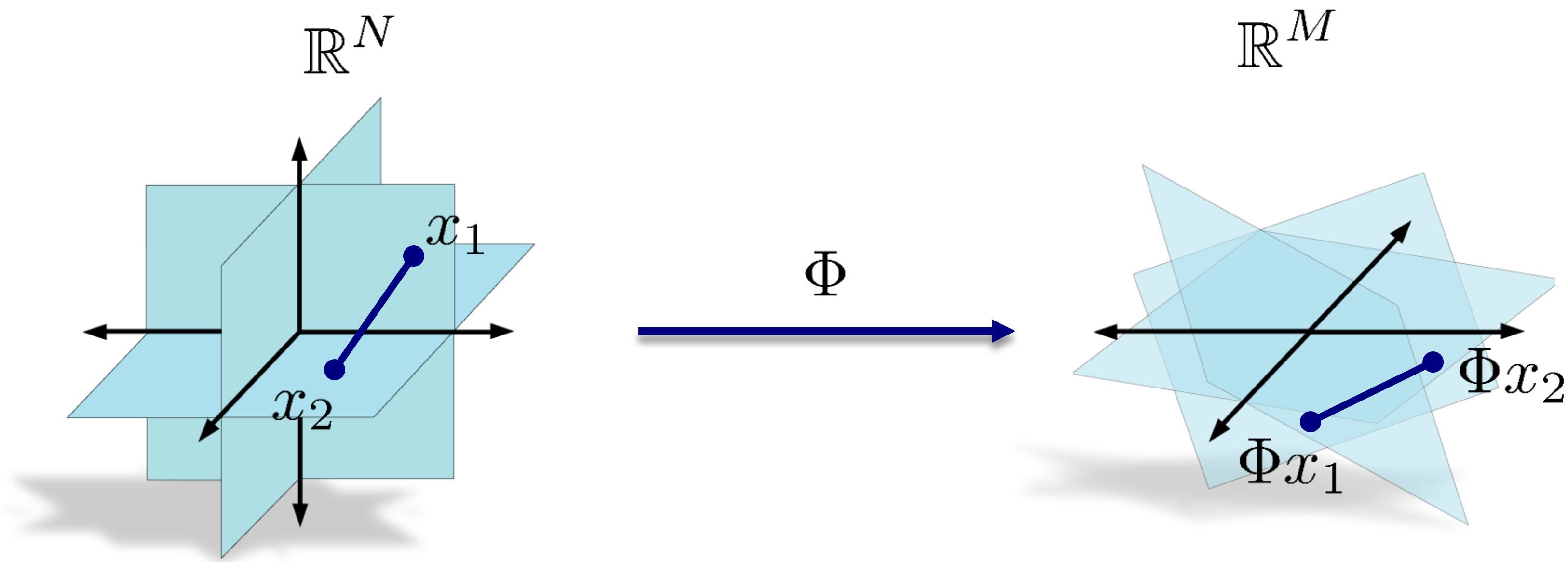
- How should we design the matrix Φ so that M is as small as possible?



- How can we recover x from the measurements y ?

Restricted Isometry Property (RIP)

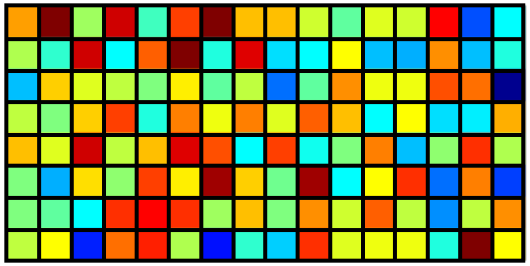
$$1 - \delta \leq \frac{\|\Phi x_1 - \Phi x_2\|_2^2}{\|x_1 - x_2\|_2^2} \leq 1 + \delta \qquad \|x_1\|_0, \|x_2\|_0 \leq S$$



$$1 - \delta \leq \frac{\|\Phi x\|_2^2}{\|x\|_2^2} \leq 1 + \delta \qquad \|x\|_0 \leq 2S$$

RIP Matrix: Option 1

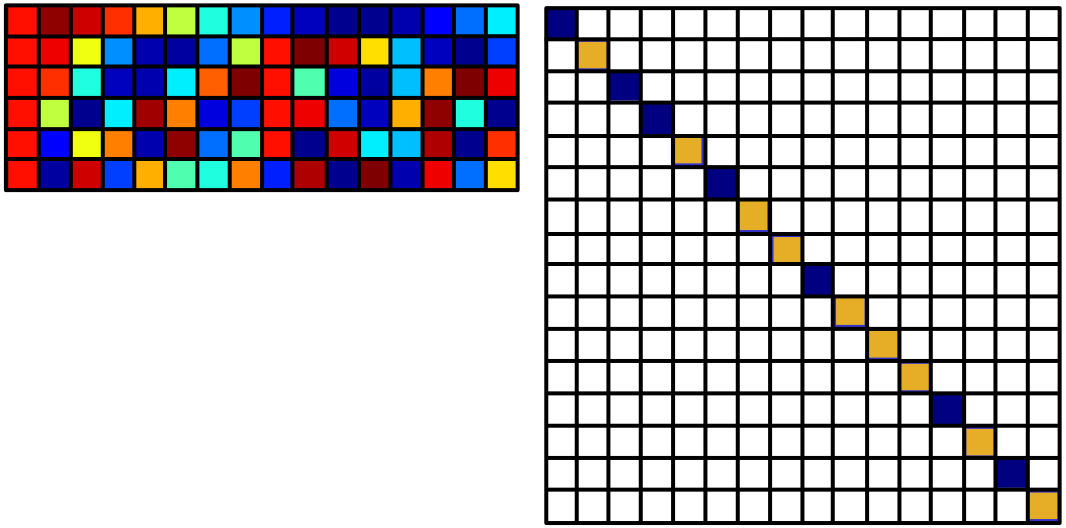
- Choose a *random matrix*
  - fill out the entries of Φ with i.i.d. samples from a sub-Gaussian distribution
  - project onto a "random subspace"



$$M = O(S \log(N/S)) \ll N$$

[Baraniuk, Davenport, DeVore, Wakin –2008]

RIP Matrix: Option 2
“Fast Johson-Lindenstrauss Transform”



- By first multiplying by random signs, a random Fourier/Hadamard submatrix can be used for efficient Johnson-Lindenstrauss (good) embeddings
- If you multiply the columns of *any* RIP matrix by random signs, you get a JL embedding!

[Ailon and Chazelle – 2007; Krahmer and Ward - 2010]

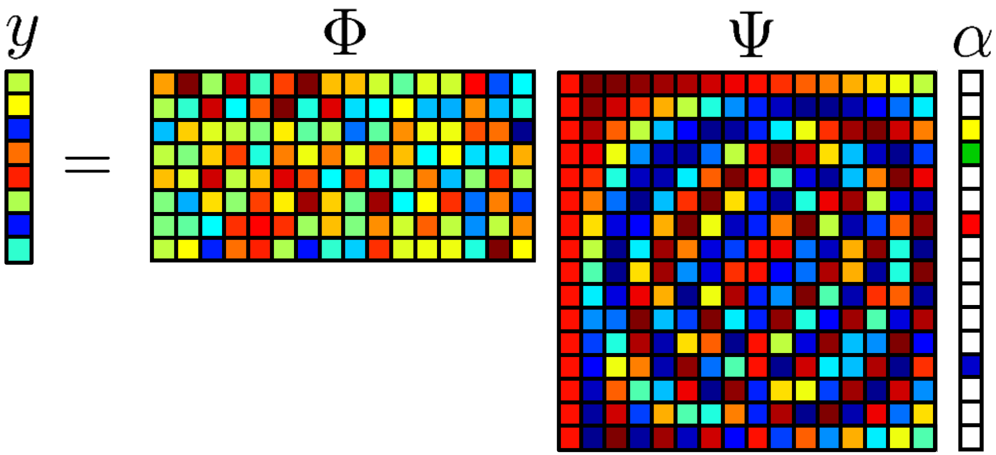
Hallmarks of Random Measurements

Stable

With high probability, Φ will preserve information, be robust to noise

Universal

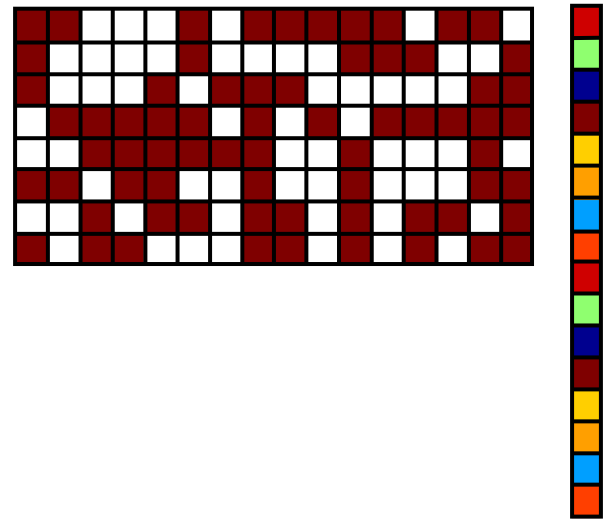
Φ will work with ***any*** fixed orthonormal basis (w.h.p.)



Democratic

Each measurement has "equal weight"

"Single-Pixel Camera"

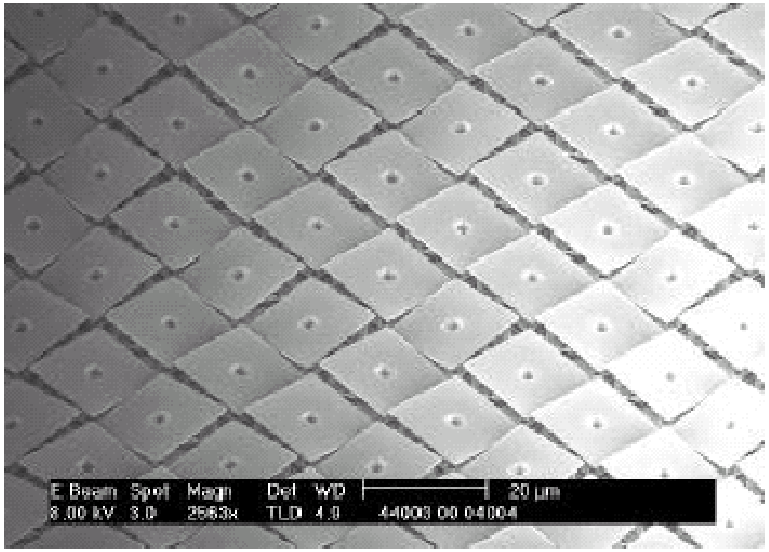


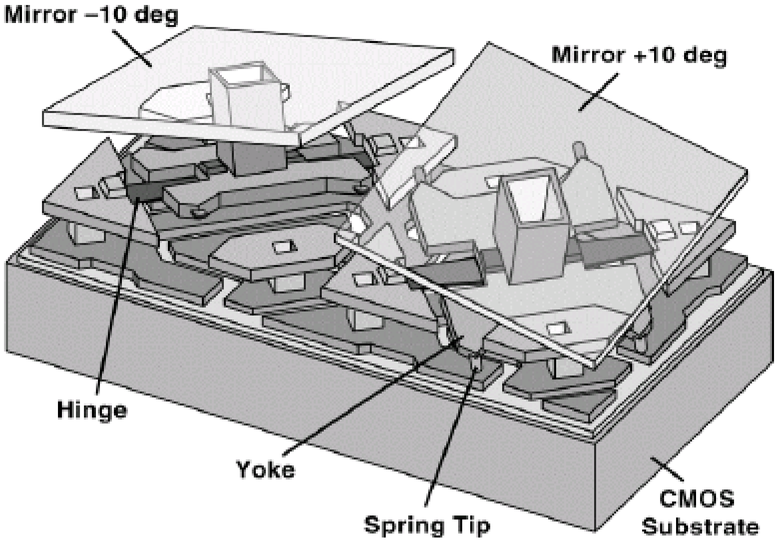
$$y[m] = \sum_{n \in I_m} x[n]$$

$$x[n] = \int\!\!\int_{\text{pixel } n} x(t_1, t_2)\, dt_1\, dt_2$$

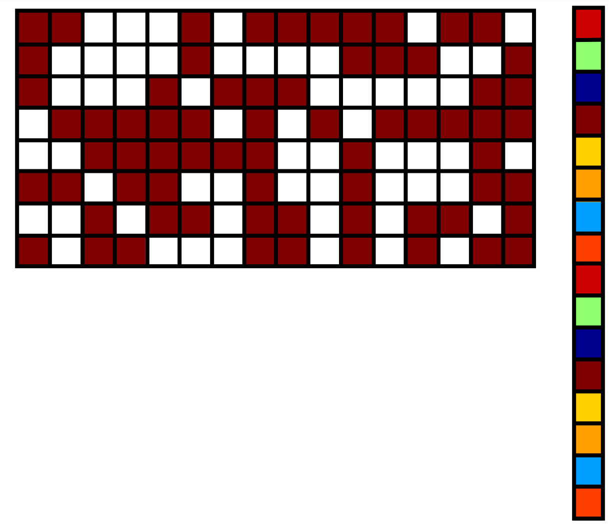
[Duarte, Davenport, Takhar, Laska, Sun, Kelly, Baraniuk - 2008]

TI Digital Micromirror Device





“Single-Pixel Camera”

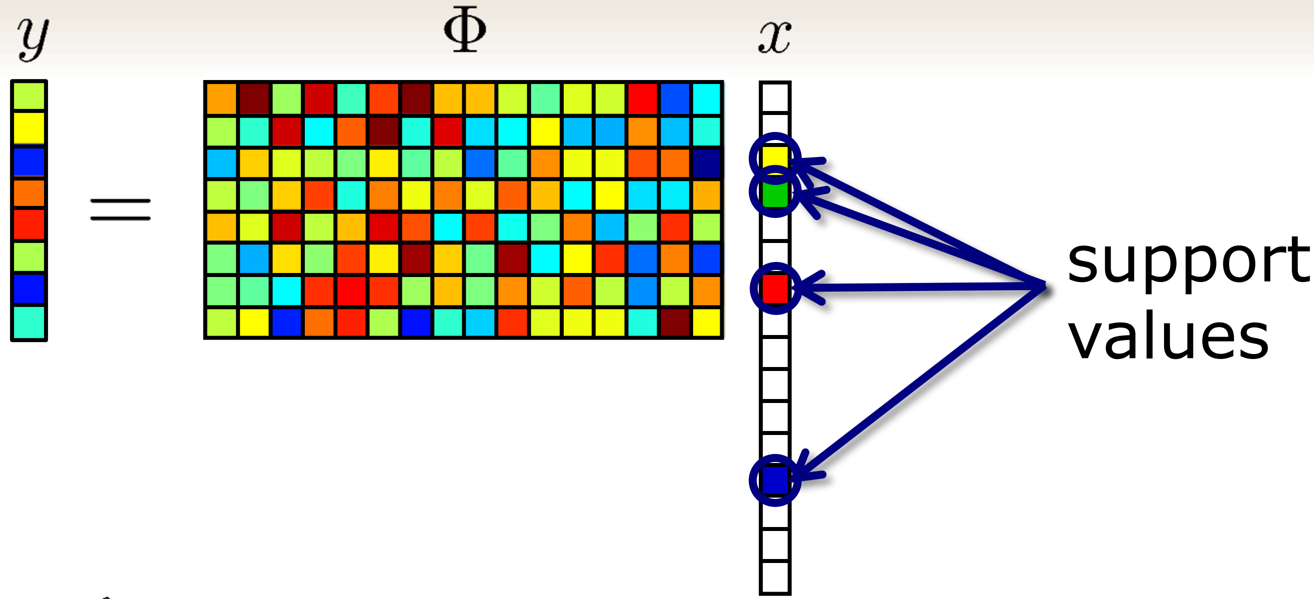


$$y[m] = \sum_{n \in I_m} x[n]$$

$$x[n] = \int\!\!\int_{\text{pixel } n} x(t_1, t_2)\, dt_1\, dt_2$$

[Duarte, Davenport, Takhar, Laska, Sun, Kelly, Baraniuk - 2008]

Sparse Signal Recovery



- Optimization / ℓ_1 -minimization
- Greedy algorithms
  - matching pursuit
  - orthogonal matching pursuit (OMP)
  - Stagewise OMP (StOMP), regularized OMP (ROMP)
  - CoSaMP, Subspace Pursuit, IHT, …

Sparse Recovery: Noiseless Case

given $y = \Phi x$
find x

- ℓ_0-minimization: $$\widehat{x} = \arg\min_{x \in \mathbb{R}^N} \|x\|_0 \quad \text{s.t.} \quad y = \Phi x$$ ← *nonconvex* *NP-Hard*

- ℓ_1-minimization: $$\widehat{x} = \arg\min_{x \in \mathbb{R}^N} \|x\|_1 \quad \text{s.t.} \quad y = \Phi x$$ ← *convex* *linear program*

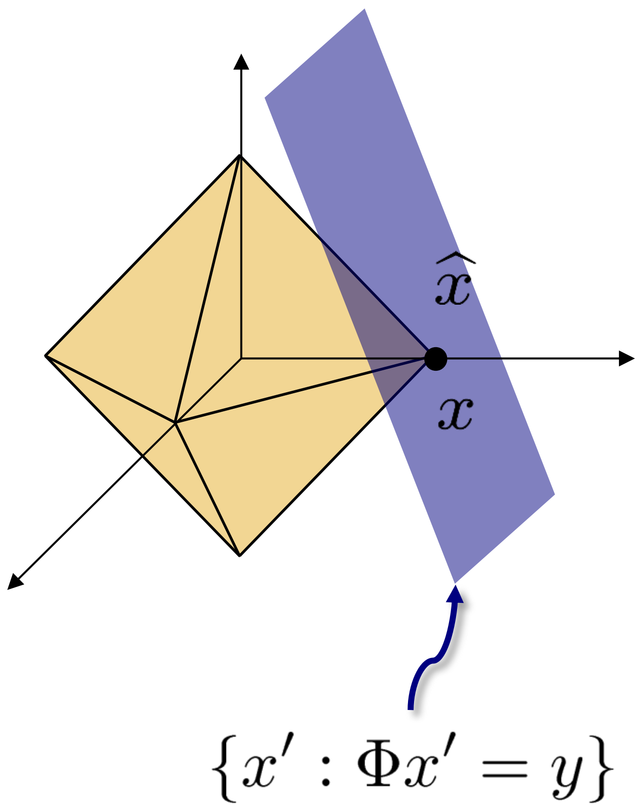
- If Φ satisfies the RIP, then ℓ_0 and ℓ_1 are equivalent!

[Donoho; Candès, Romberg, Tao - 2004]

Why ℓ_1-Minimization Works

$$\widehat{x} = \arg\min_{x \in \mathbb{R}^N} \|x\|_1$$

$$\text{s.t.} \quad y = \Phi x$$



Sparse Recovery: Noisy Case

Suppose we observe $y = \Phi x + e$, where $\|e\|_2 \leq \epsilon$

$$\widehat{x} = \arg\min_{x \in \mathbb{R}^N} \|x\|_1$$

$$\text{s.t.} \quad \|y - \Phi x\|_2 \leq \epsilon$$

$$\boxed{\|\widehat{x} - x\|_2 \leq C_0 \epsilon}$$

Similar approaches can handle Gaussian noise added to either the signal or the measurements

Sparse Recovery: Non-sparse Signals

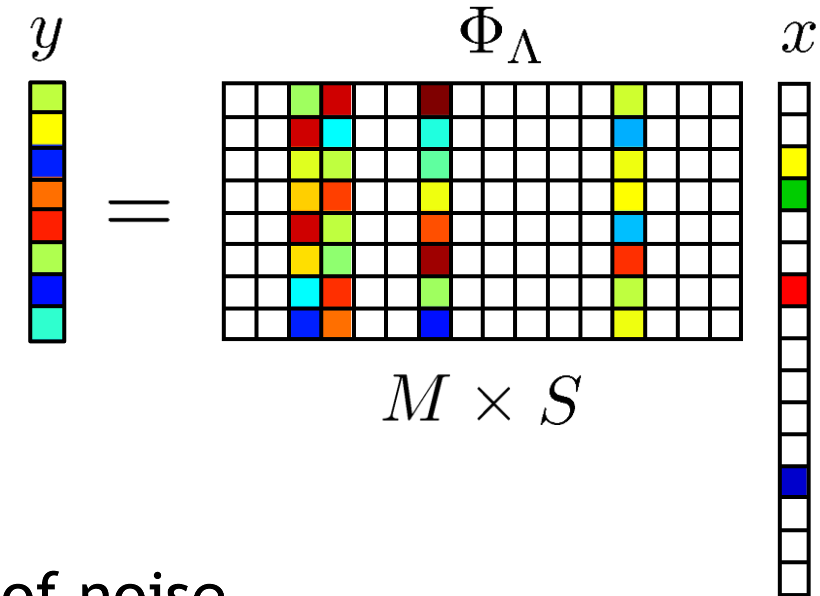
In practice, x may not be exactly S-sparse

$$\widehat{x} = \arg\min_{x \in \mathbb{R}^N} \|x\|_1 \quad \text{s.t.} \quad \|y - \Phi x\|_2 \leq \epsilon$$

$$\boxed{\|\widehat{x} - x\|_2 \leq C_0 \epsilon + C_1 \frac{\|x - x_S\|_1}{\sqrt{S}}}$$

Greedy Algorithms: Key Idea

If we can determine $\Lambda = \mathrm{supp}(x)$, then the problem becomes *over*-determined.



In the absence of noise,

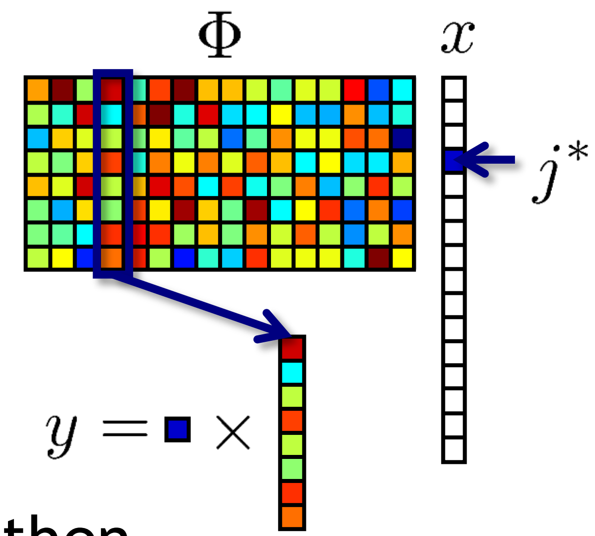
$$\begin{aligned} \Phi_\Lambda^\dagger y &= (\Phi_\Lambda^T \Phi_\Lambda)^{-1} \Phi_\Lambda^T y \\ &= (\Phi_\Lambda^T \Phi_\Lambda)^{-1} \Phi_\Lambda^T \Phi_\Lambda x \\ &= x \end{aligned}$$

Matching Pursuit

Select one index at a time using a simple *proxy* for x

$$p = \Phi^T y$$

$$\boxed{j^* = \arg\max_j |p_j|}$$



If Φ satisfies the RIP of order $\|u \pm v\|_0$, then

$$|\langle \Phi u, \Phi v \rangle - \langle u, v \rangle| \leq \delta \|u\|_2 \|v\|_2$$

Set $u = x$ and $v = e_j$

$$|p_j - x_j| \leq \delta \|x\|_2$$

Matching Pursuit

Obtain initial estimate of x

$$x^{(1)} = p_{j^*} e_{j^*}$$

Update proxy and iterate

$$p = \Phi^T (y - \Phi x^{(j-1)})$$

$$j^* = \arg\max_j |p_j|$$

$$x^{(j)} = x^{(j-1)} + p_{j^*} e_{j^*}$$

Iterative Hard Thresholding (IHT)

step size

$$x^{(j)} = \underbrace{H_S}_{\text{hard thresholding}} \left(x^{(j-1)} + \overbrace{\mu}^{\text{step size}} \underbrace{\Phi^T \left(y - \Phi x^{(j-1)} \right)}_{\text{proxy vector}} \right)$$

hard thresholding *proxy vector*

RIP guarantees convergence and accurate/stable recovery

[Blumensath and Davies – 2008]

Extensions of Matching Pursuit

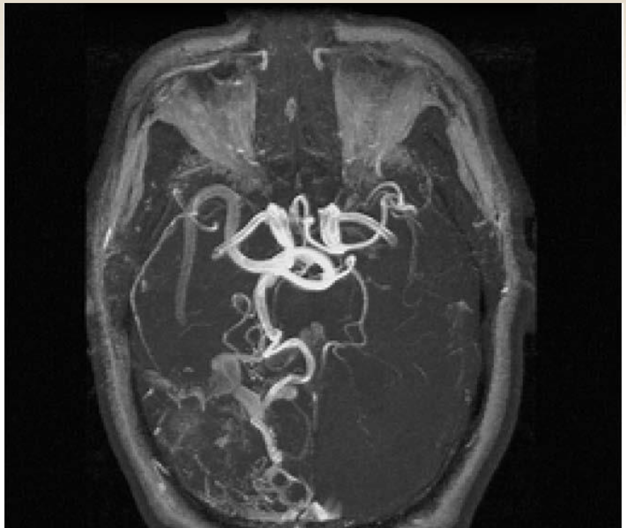
- Orthogonal matching pursuit
  - change update rule to ensure that the residual $y - \Phi x^{(j)}$ is always orthogonal to previously selected columns
  - ensures that we never pick a column twice

- StOMP, ROMP
  - select many indices in each iteration
  - picking indices for which p_j is "comparable" leads to increased stability and robustness

- CoSaMP, Subspace Pursuit, …
  - allow indices to be discarded
  - strongest guarantees, comparable to ℓ_1-minimization

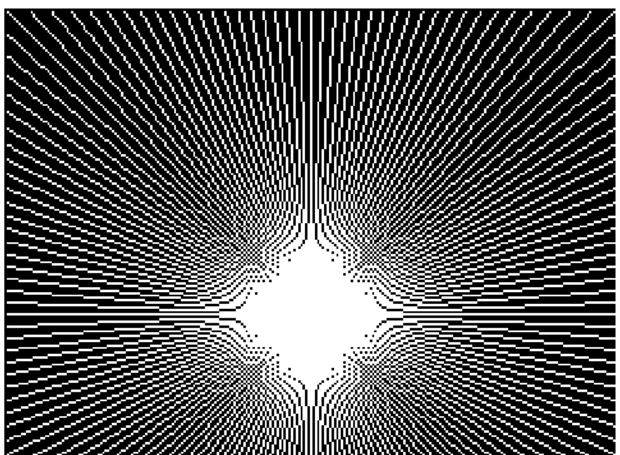
Applications of CS to Imaging

- MRI
  - Observe randomly selected Fourier coefficients
  - Exploit sparsity in wavelet basis

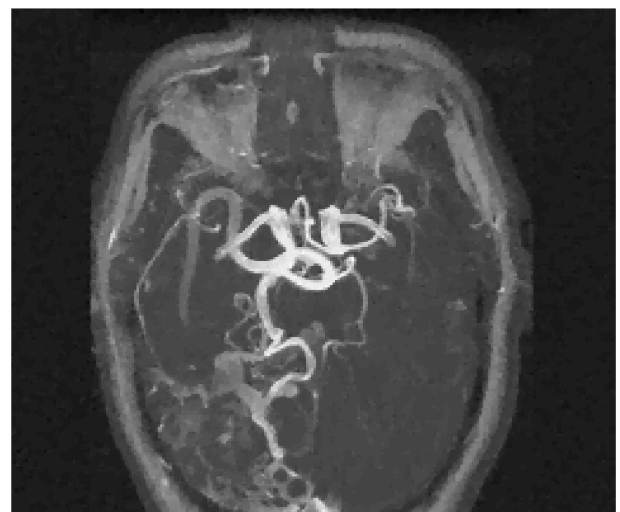


256x256 MRA

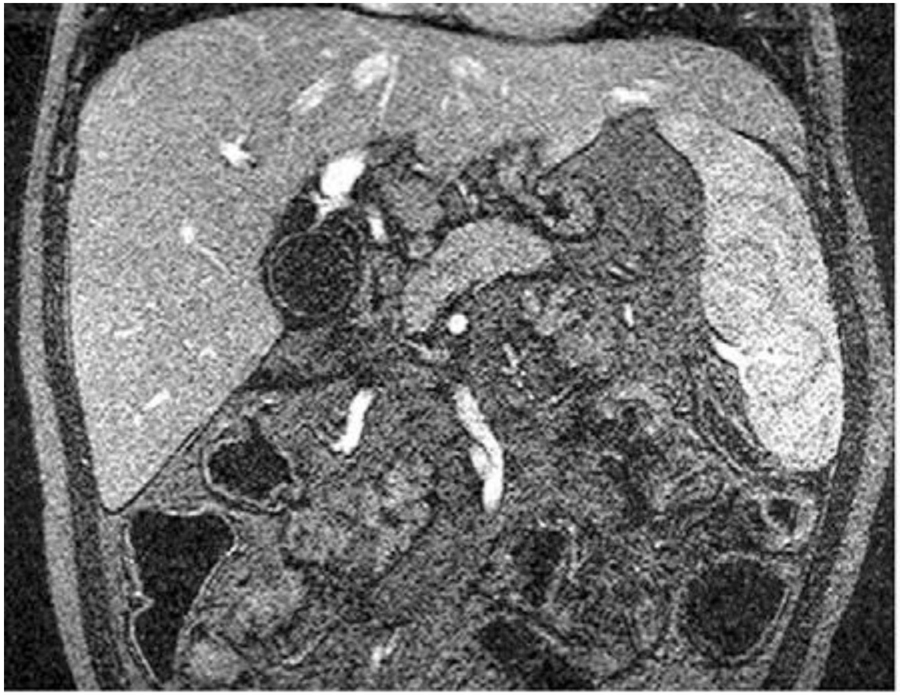
Backproj., 29.00dB



Fourier sampling
80 lines (M~0.28N)



Min TV, 34.23dB [CR]



Traditional MRI

CS MRI

4-8 x faster!

[Vasanawala, Alley, Hargreaves, Barth, Pauly, Lustig - 2010]

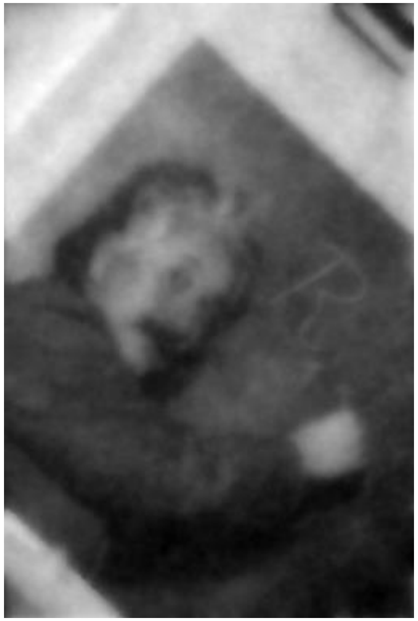
Applications of CS to Imaging

- MRI
  - Observe randomly selected Fourier coefficients
  - Exploit sparsity in wavelet basis

- Single pixel camera
  - Replace light sensor with something more sophisticated
    - SWIR sensor
    - Spectrometer
    - …

SWIR Single Pixel Camera

256×384 pixels



10%

20%

30%

40%

InView

Applications of CS to Imaging

- MRI
  - Observe randomly selected Fourier coefficients
  - Exploit sparsity in wavelet basis

- Single pixel camera
  - Replace light sensor with something more sophisticated
    - SWIR sensor
    - Spectrometer
    - …

- Many more

Challenges

Imaging challenges some of the key assumptions in much of the CS theory

- In the context of imaging, Φ tells us how light propagates through our system
  - nonnegative
  - non-standard normalization
- Gaussian noise is often not a safe assumption
  - poisson noise models are generally more difficult to exploit and analyze

Why is This a Problem?

- Standard CS theory suggests setting the entries of Φ to be $\pm 1/\sqrt{M}$
- In imaging we must shift and rescale Φ

$$\widetilde{\Phi} = \frac{\Phi + 1/\sqrt{M}}{2\sqrt{M}}$$

- Entries now are either 0 or $1/M$
- Observations given by

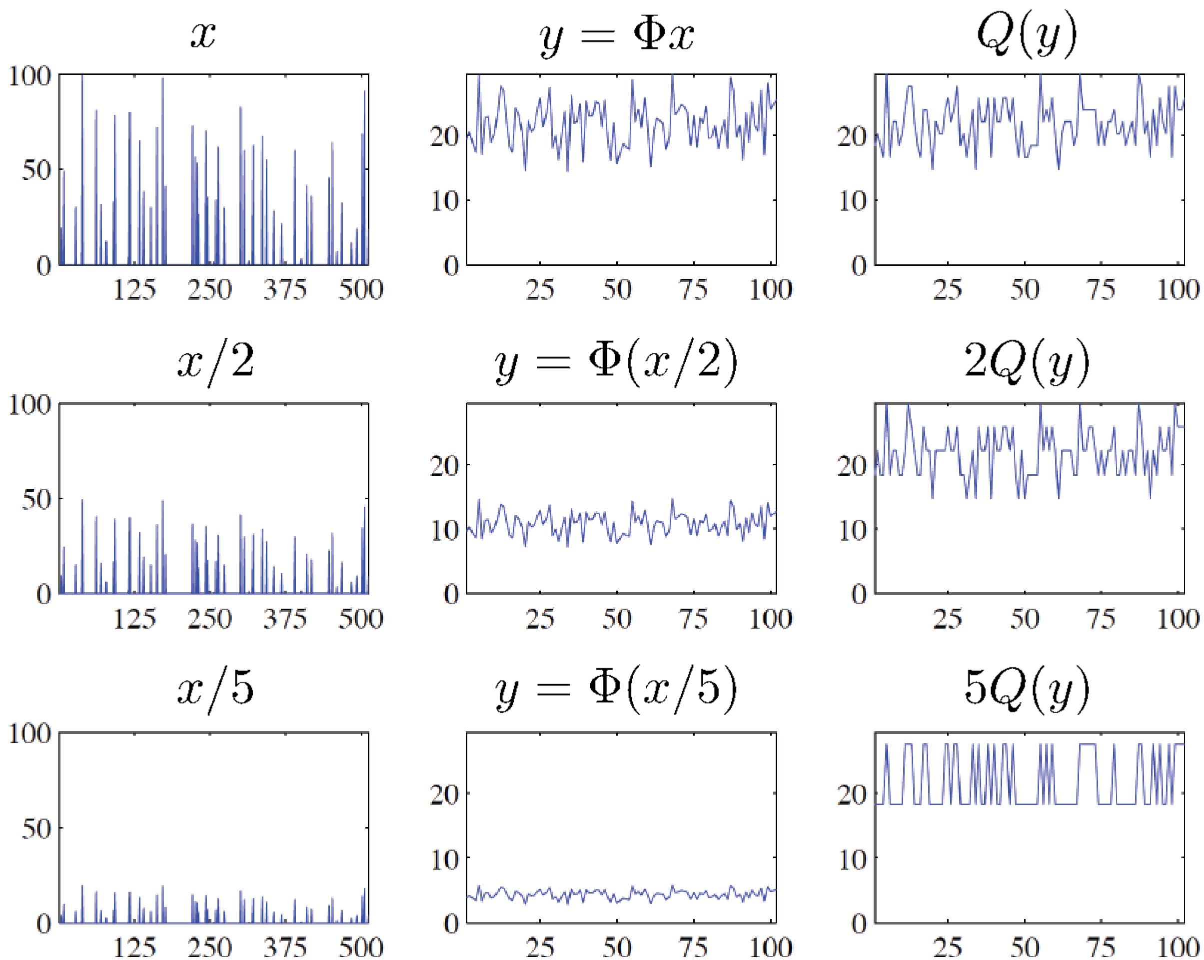
$$y = \widetilde{\Phi}x + e = \frac{\Phi x}{2\sqrt{M}} + \frac{\|x\|_1}{2M} + e$$

signal

DC offset
(photon noise)

Dynamic Range

What about the impact of quantization?



Conclusions

- The theory of compressive sensing allows for new sensor designs, but requires new techniques for signal recovery
- Compressive sensing can be applied in the context of imaging, but doing so successfully requires an awareness of the gaps between CS theory and imaging practice
- Many open questions remain
  - CS may seem more sensitive to noise, but enables the use of higher quality sensors. What is the real impact of noise?
  - How sensitive is CS to imperfect system models?
  - How does CS impact the dynamic range of our system?