ELECTRICAL ENGINEERING & COMPUTER SCIENCE
COLORADO SCHOOL OF MINES

# Compressive Sensing

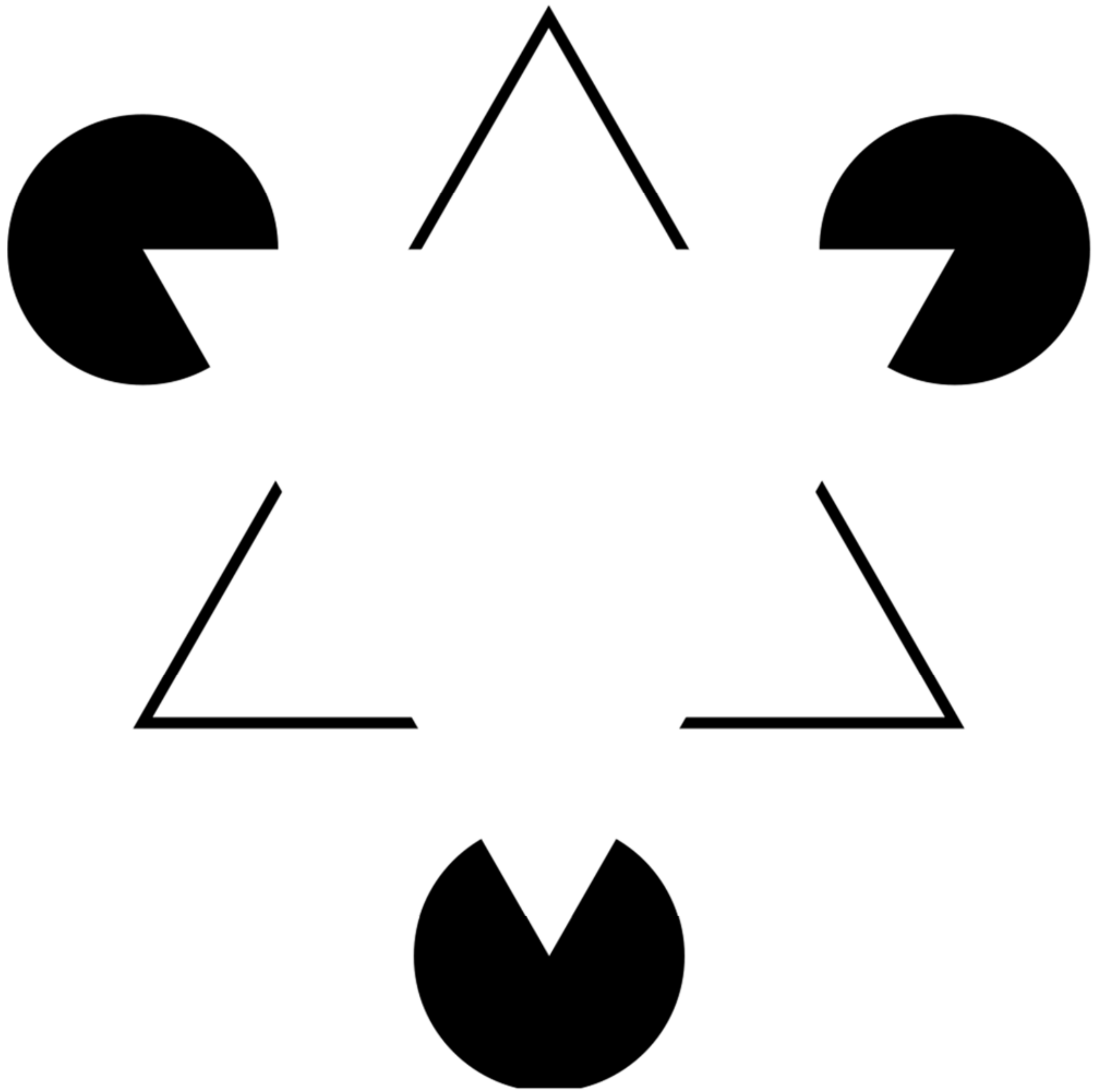
*Mike Wakin*

# Introduction:
# Filling in the blanks

H E L _ O

# H E L L O

# H E L a O



| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5 | 3 | | | 7 | | | | |
| 6 | | | 1 | 9 | 5 | | | |
| | 9 | 8 | | | | | 6 | |
| 8 | | | | 6 | | | | 3 |
| 4 | | | 8 | | 3 | | | 1 |
| 7 | | | | 2 | | | | 6 |
| | 6 | | | | | 2 | 8 | |
| | | | 4 | 1 | 9 | | | 5 |
| | | | | 8 | | | 7 | 9 |

# Two Key Ingredients

1. Model:

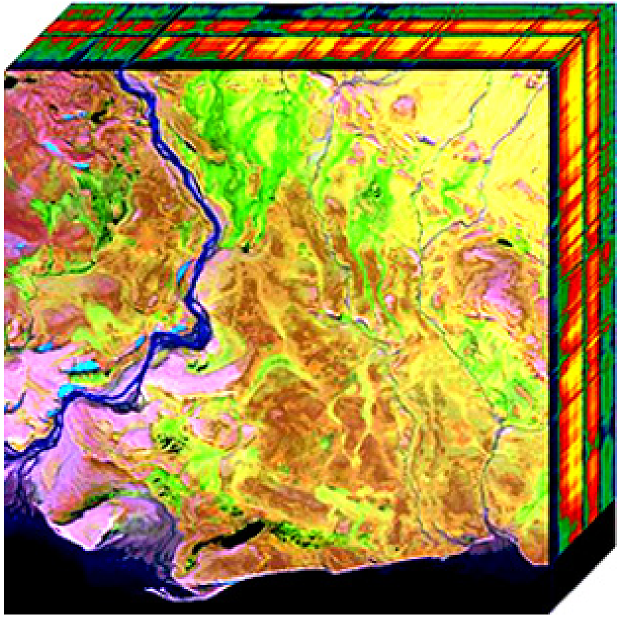
   "How is the signal supposed to behave?"

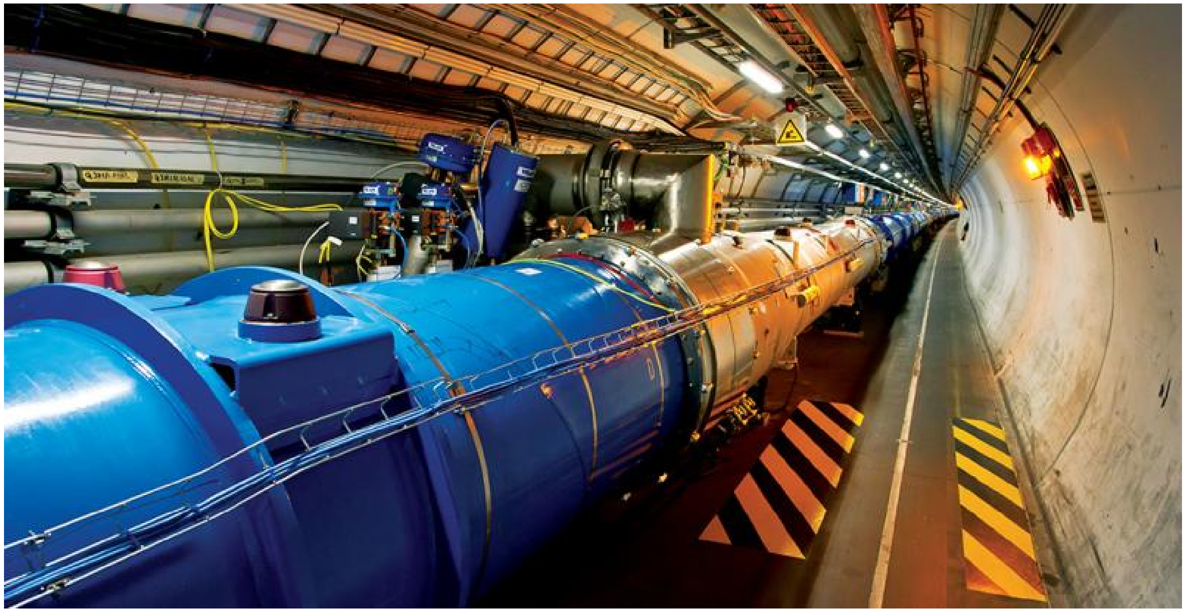
2. Algorithm:

   "How can I fill in the blanks so that the signal obeys the model?"

# Data-Rich Environments

- The Big Data Era
  - Social networks
  - Scientific instrumentation
  - Hyperspectral imaging
  - Video





# Challenges

- Storage, transmission, processing

- Conventional solution:

  collect all of the data then compress it

- New solution:

  collect less data

# Data-Poor Environments

- Limitations:
  - size, power, weight
    - embedded systems
  - cost
    - sensing hardware
    - seismic surveys
  - time
    - magnetic resonance imaging
  - communications bandwidth
    - wireless sensor networks
- Make the most of the data we can collect

# Compressive Sensing (CS) in a Nutshell

***Using signal models to fill in the blanks***

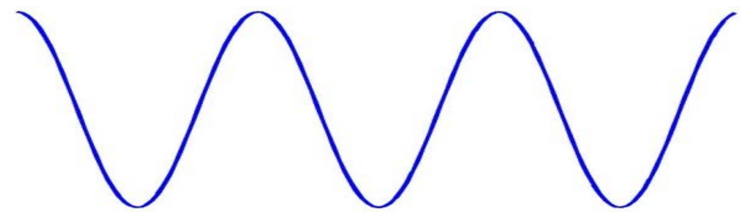
Topics:

1. What are signal models?
2. What do we mean by "blanks"?
3. How can we fill in the blanks?
4. How can we understand this process?
5. What (else) can we do with these ideas?

# 1. What are signal models?

# Sinusoids

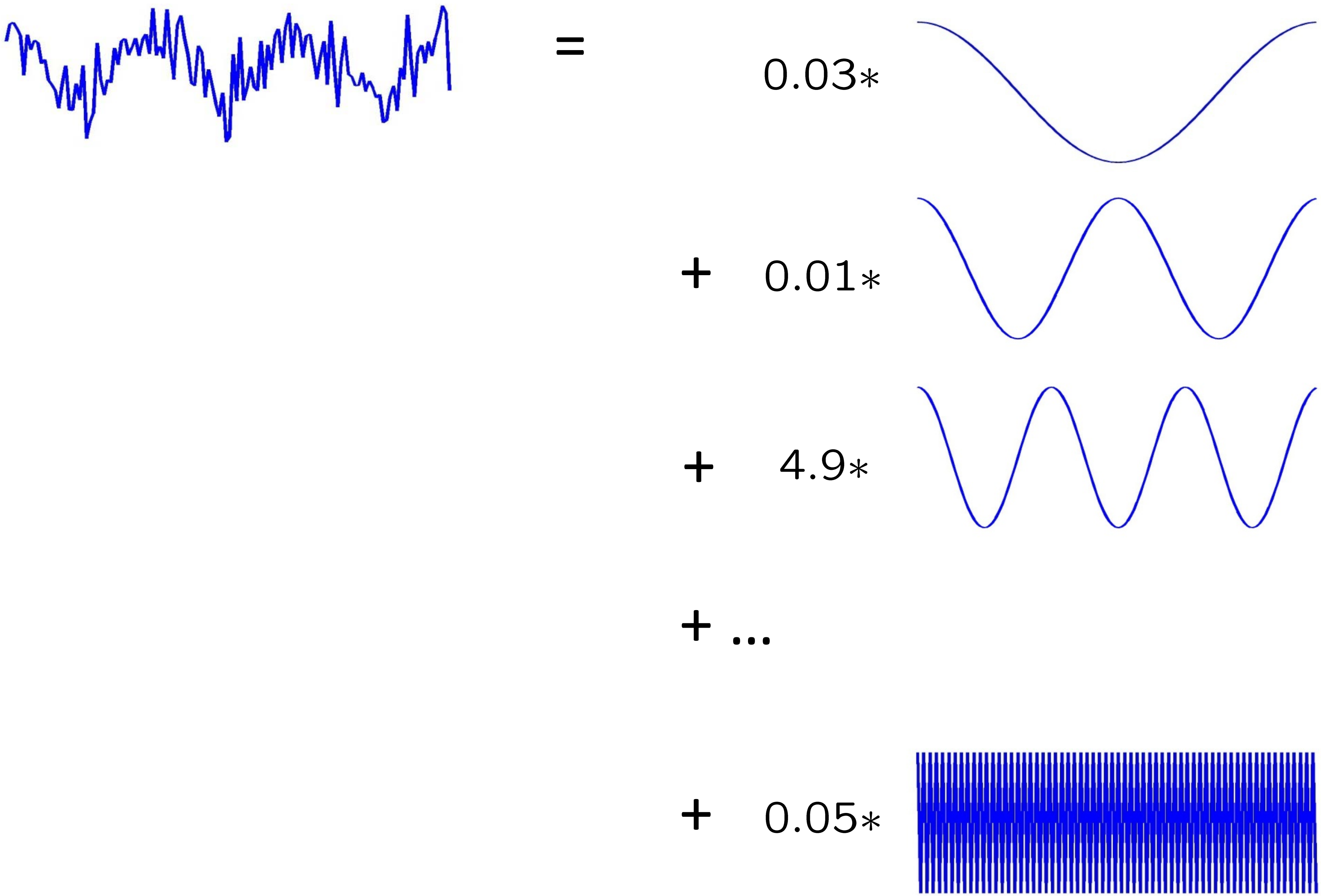
- Sinusoids are fundamental to signal processing.



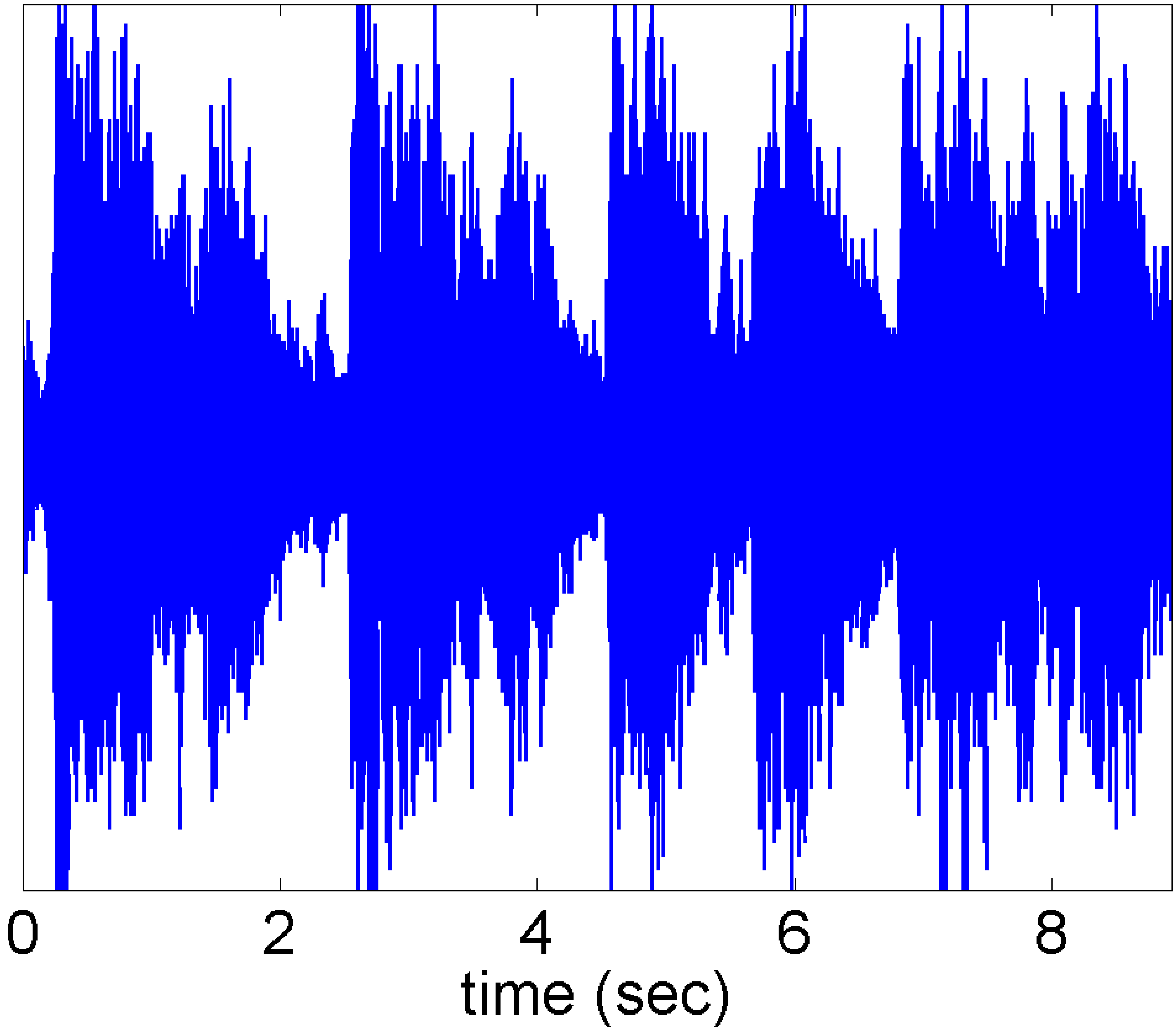
- Many real-world phenomena travel in waves or are generated by harmonic motion.
  - electromagnetic waves
  - acoustic waves
  - seismic waves

- Many real-world systems propagate sinusoids in a natural way: a sinusoidal input produces a sinusoidal output of exactly the same frequency.

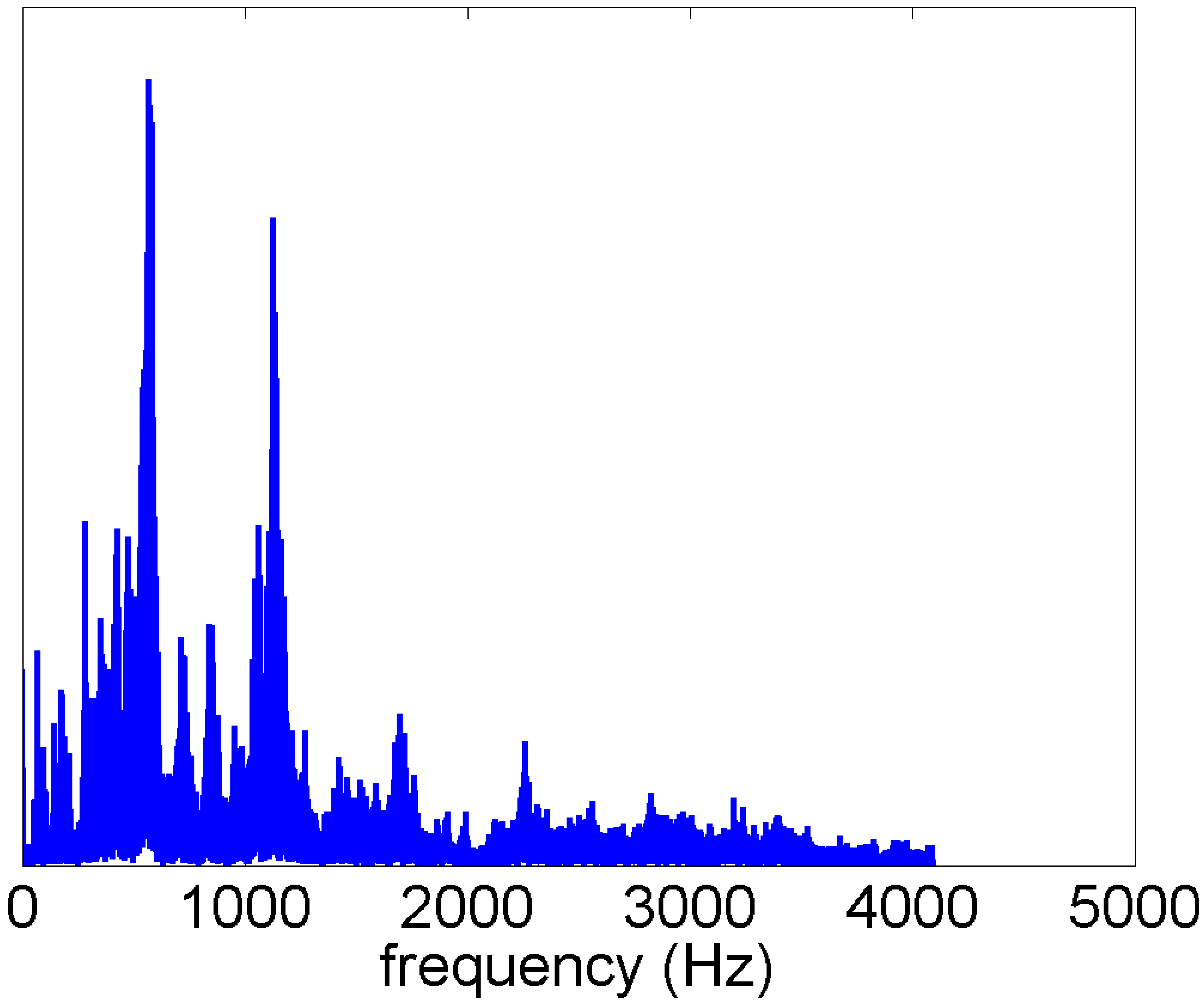
# Fourier Transform (1807)



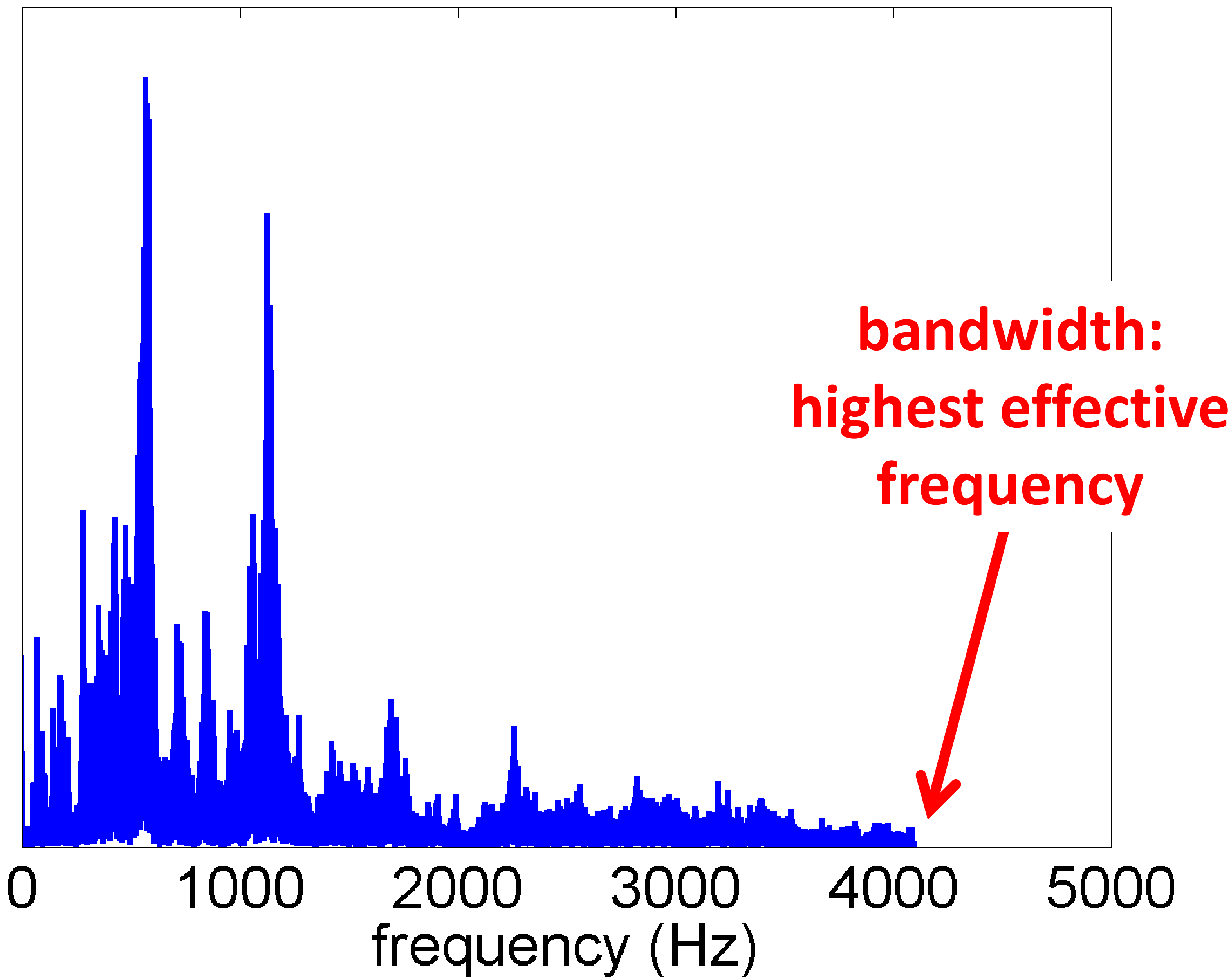
# Example Audio Signal



# Frequency Spectrum



# Bandlimited Signal Model



# Features of the Model

- Expressed in terms of a transform:

  Fourier transform

- Variable complexity:

  higher bandwidths allow for broader classes of signals but at the expense of greater complexity

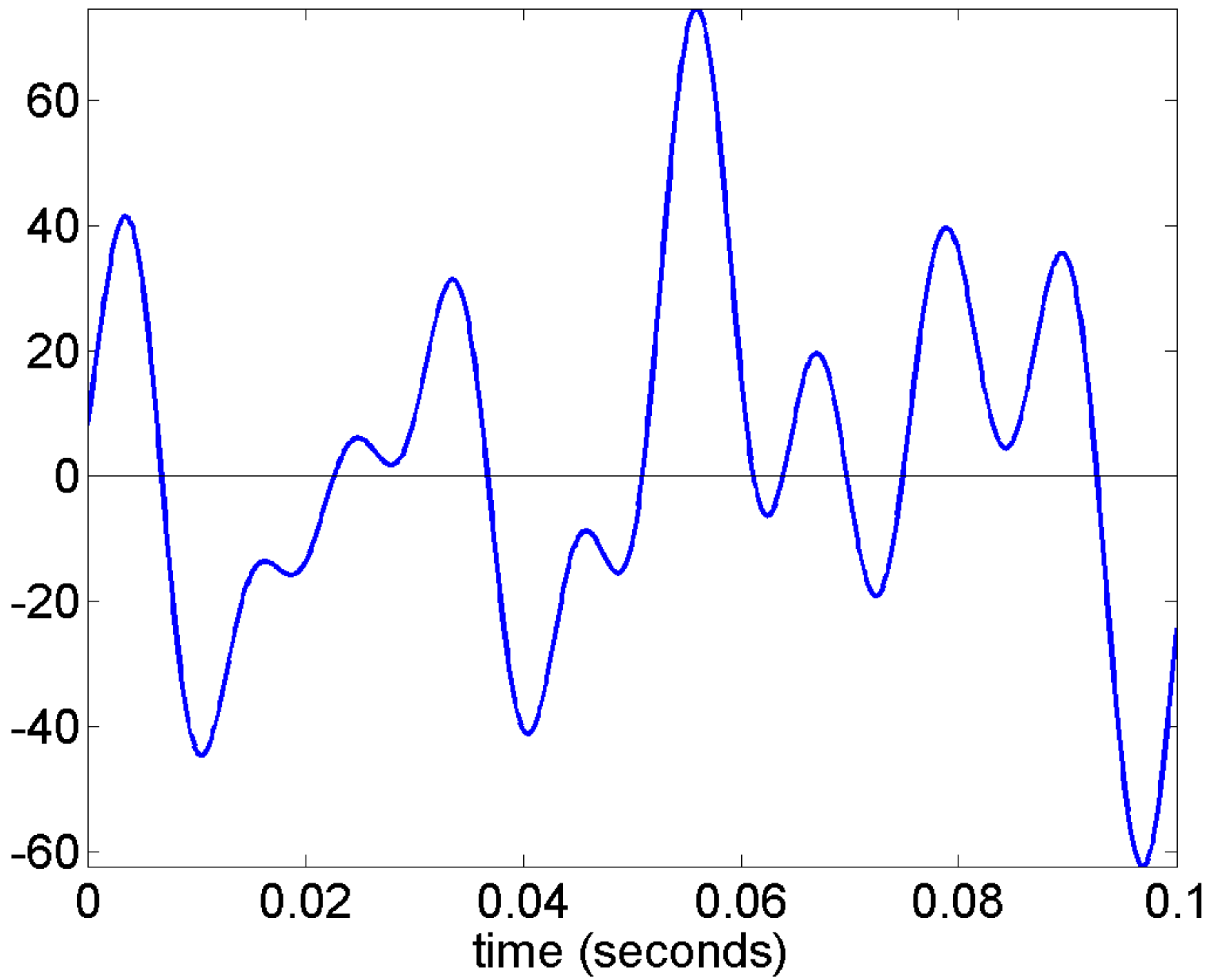
# Shannon/Nyquist Sampling Theorem (1920s-1940s)

- Theorem:

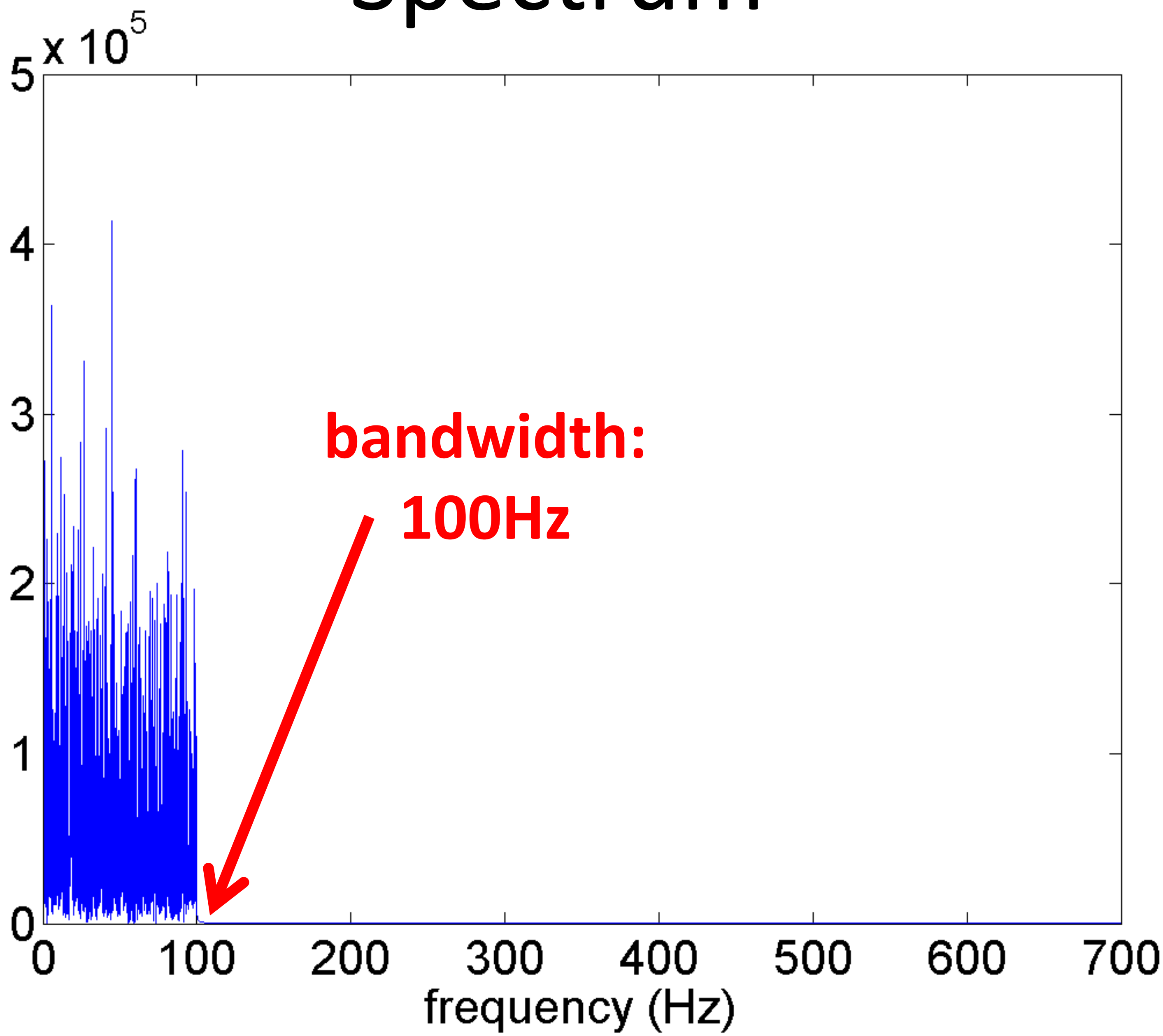
  Any bandlimited signal can be perfectly reconstructed from its samples, provided that the sampling frequency is at least twice the signal's bandwidth.

- Sampling at the information level:
  - sampling frequency is proportional to bandwidth
  - as model complexity increases, must take more samples (per second)

# Example Signal

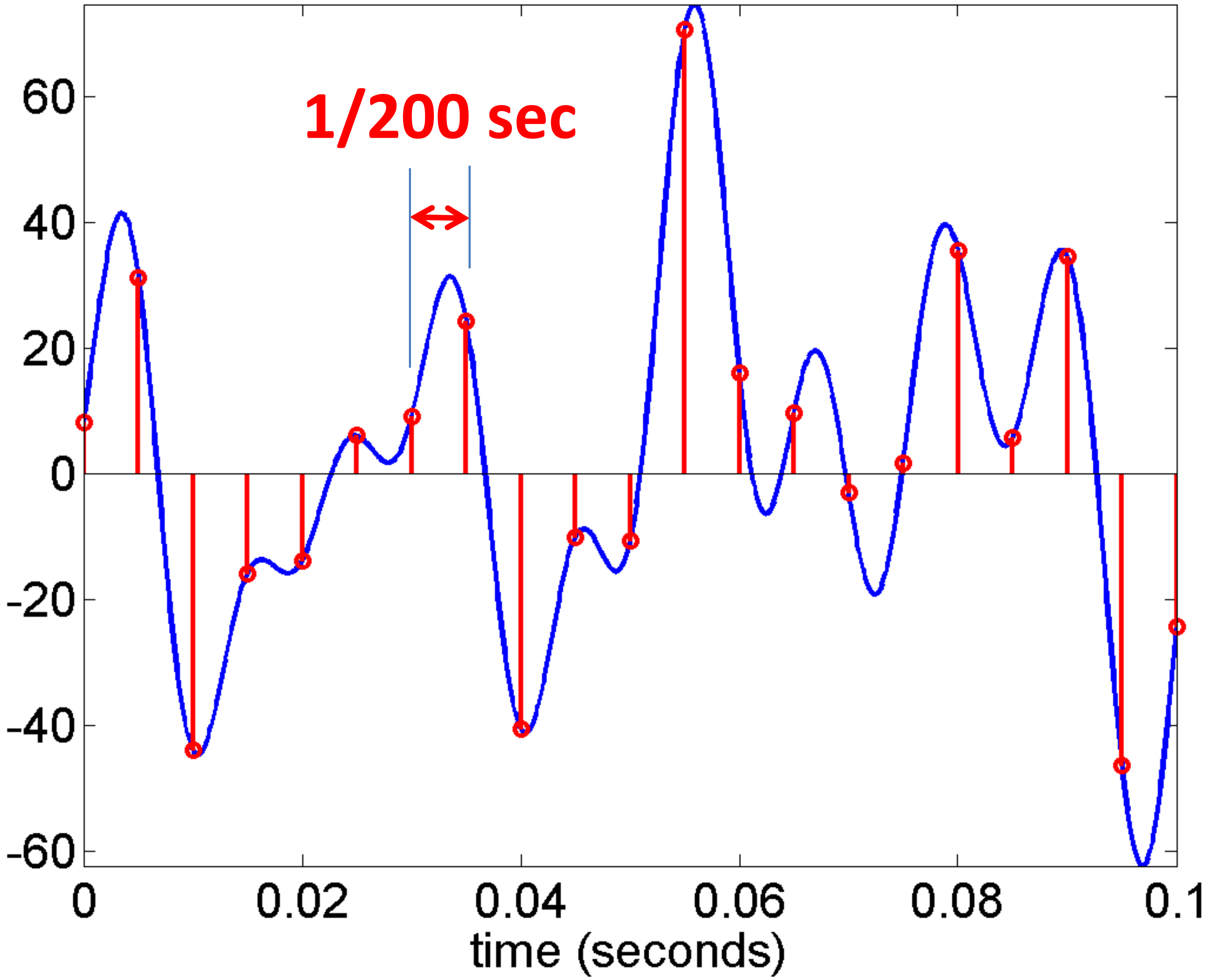


# Spectrum

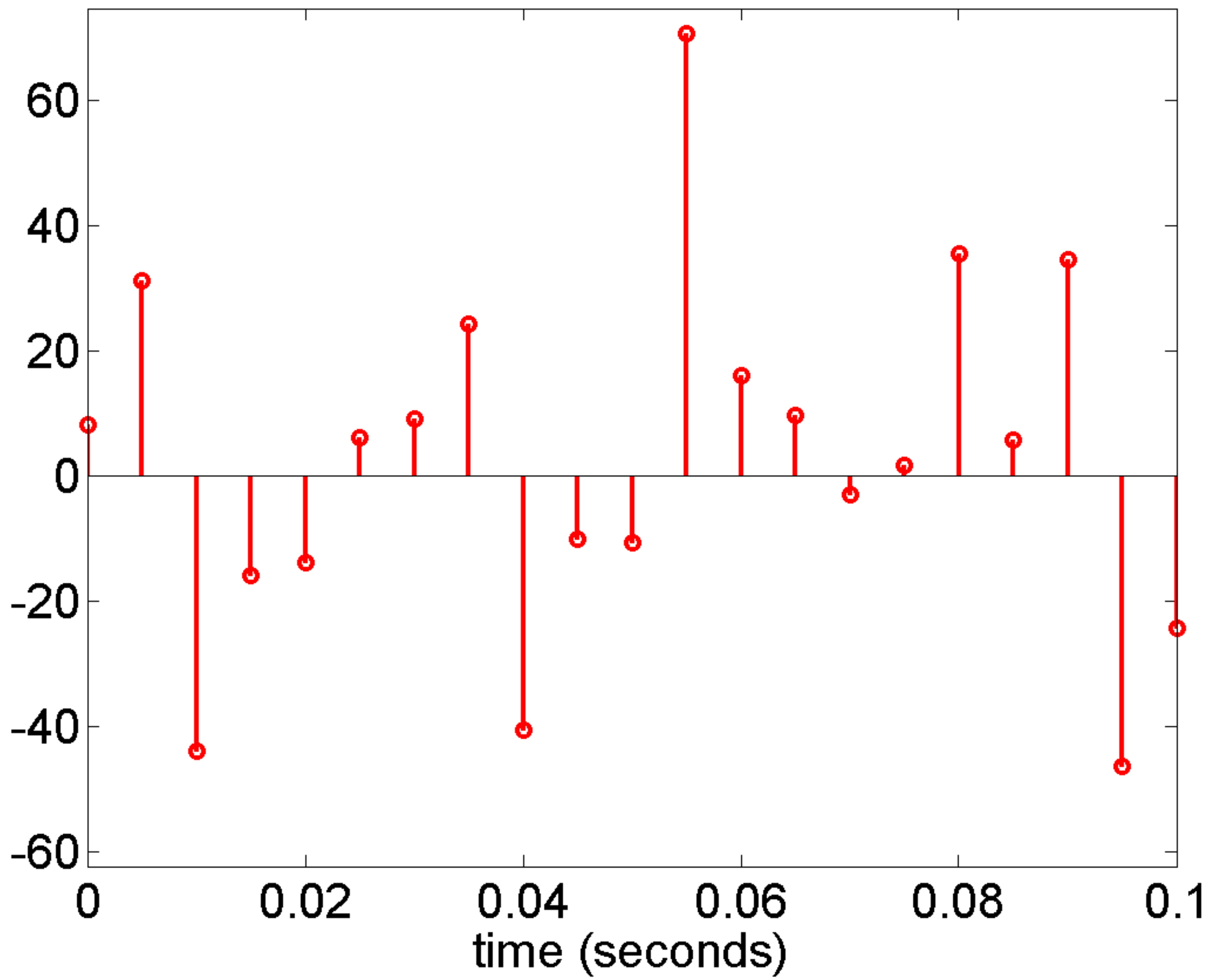


5 x 10^5
4
3
2
1
0
0
100
200
300
400
500
600
700
frequency (Hz)
bandwidth:
100Hz

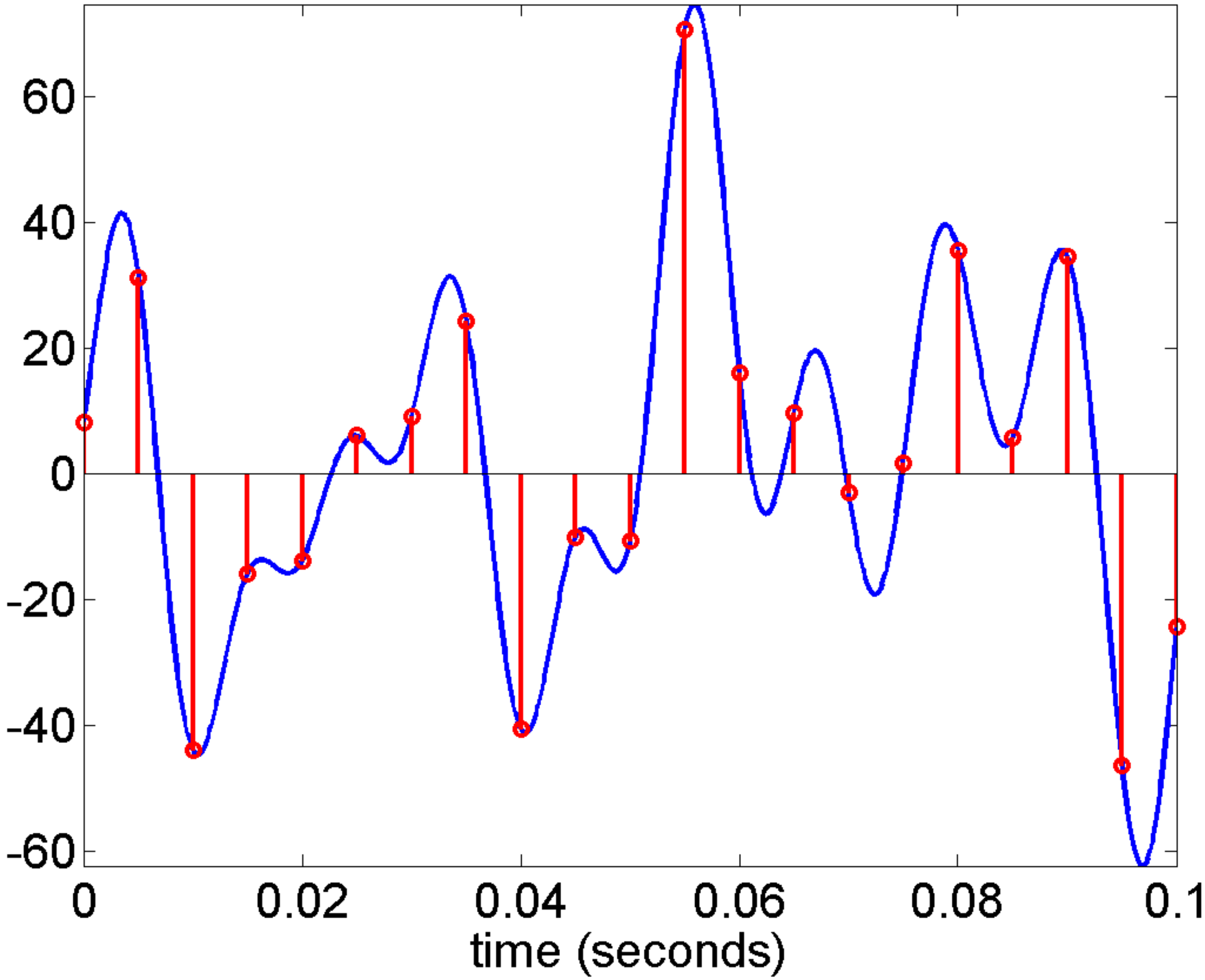
# “Nyquist-Rate” Samples



# Filling in the Blanks



# Smooth Interpolation



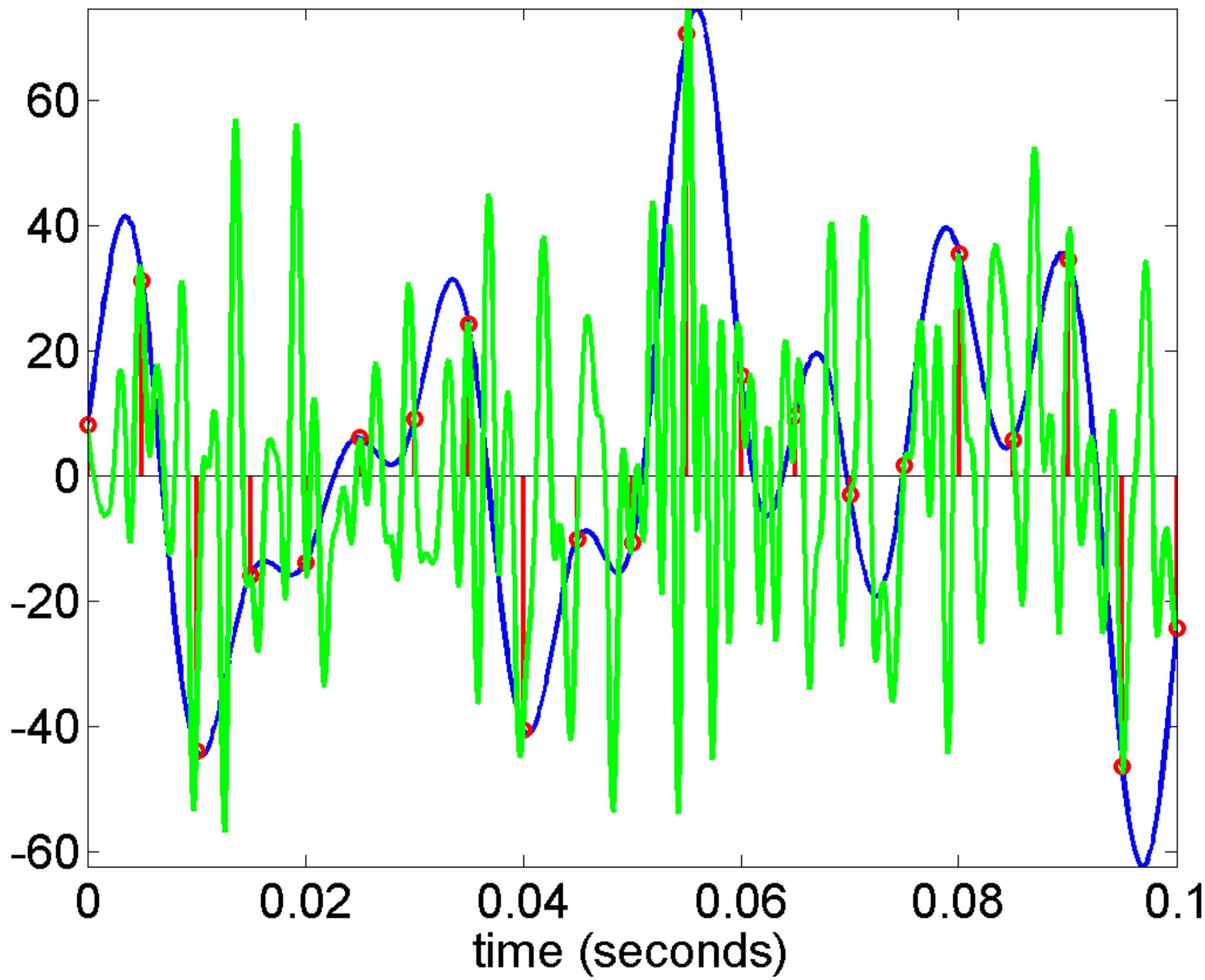
# Implications

- For bandlimited signals it is possible to ***perfectly*** fill in the blanks between adjacent samples.

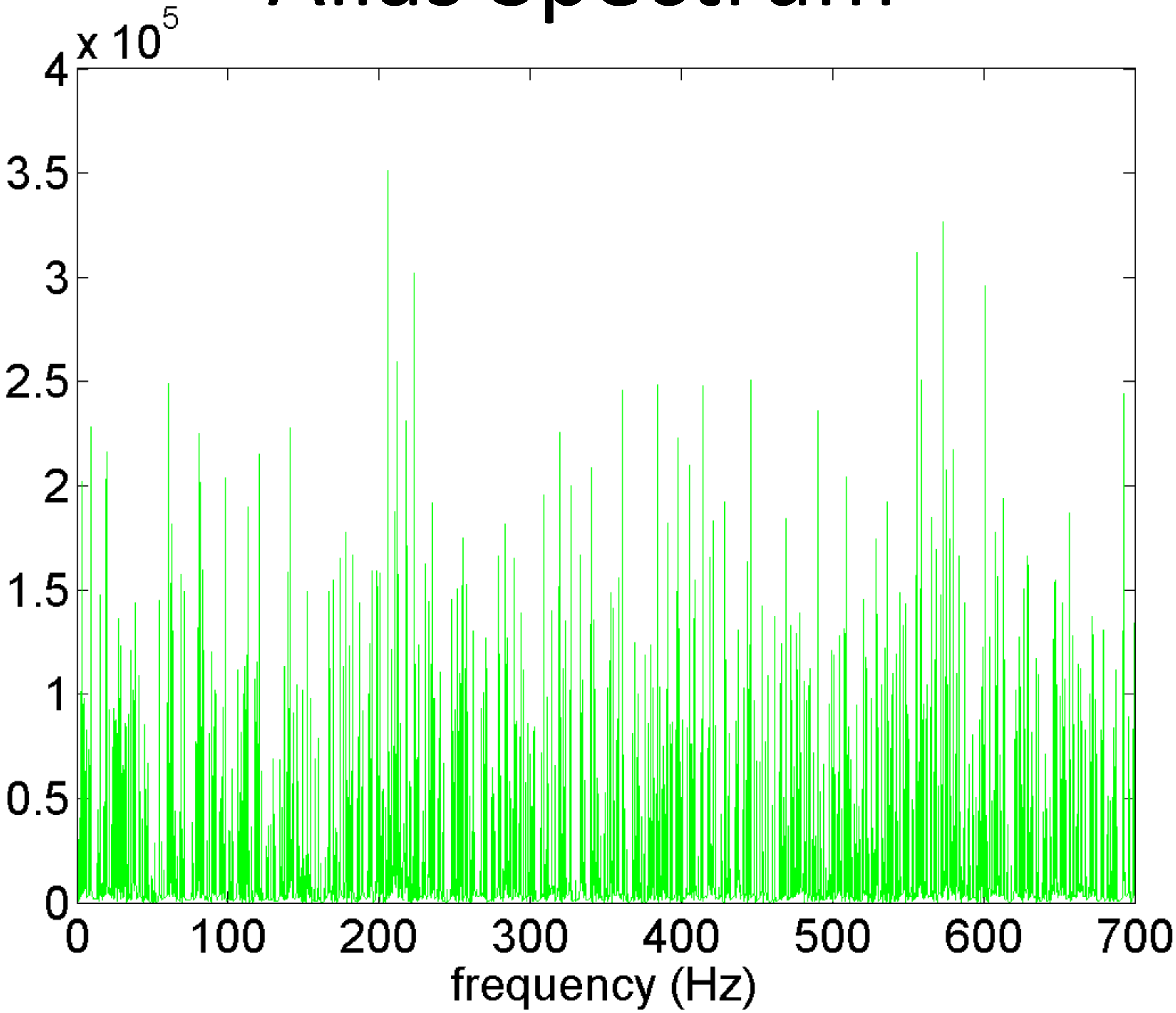
- There are many potential ways to “connect the dots” but only one of these will have the correct bandwidth.

- All of the others are known as ***aliases*** – they will all have higher bandwidths.

# An Alias



# Alias Spectrum



# Digital Signal Processing (DSP) Revolution

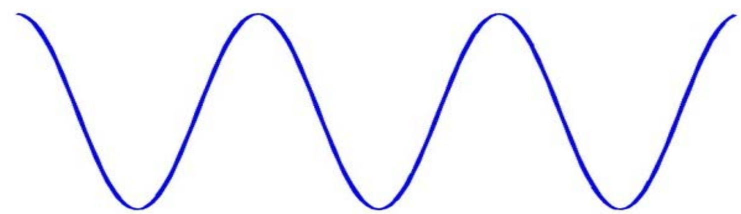
- Sampling theorem
  - sample signals without loss of information and process them on a computer
- Advances in computing
  - Moore's law
- Algorithms such as the Fast Fourier Transform
  - first discovered by Gauss (1805)
  - rediscovered by Cooley and Tukey (1965)

# From Bandlimitedness to Sparsity

- Time-frequency analysis

  frequency content may change over time

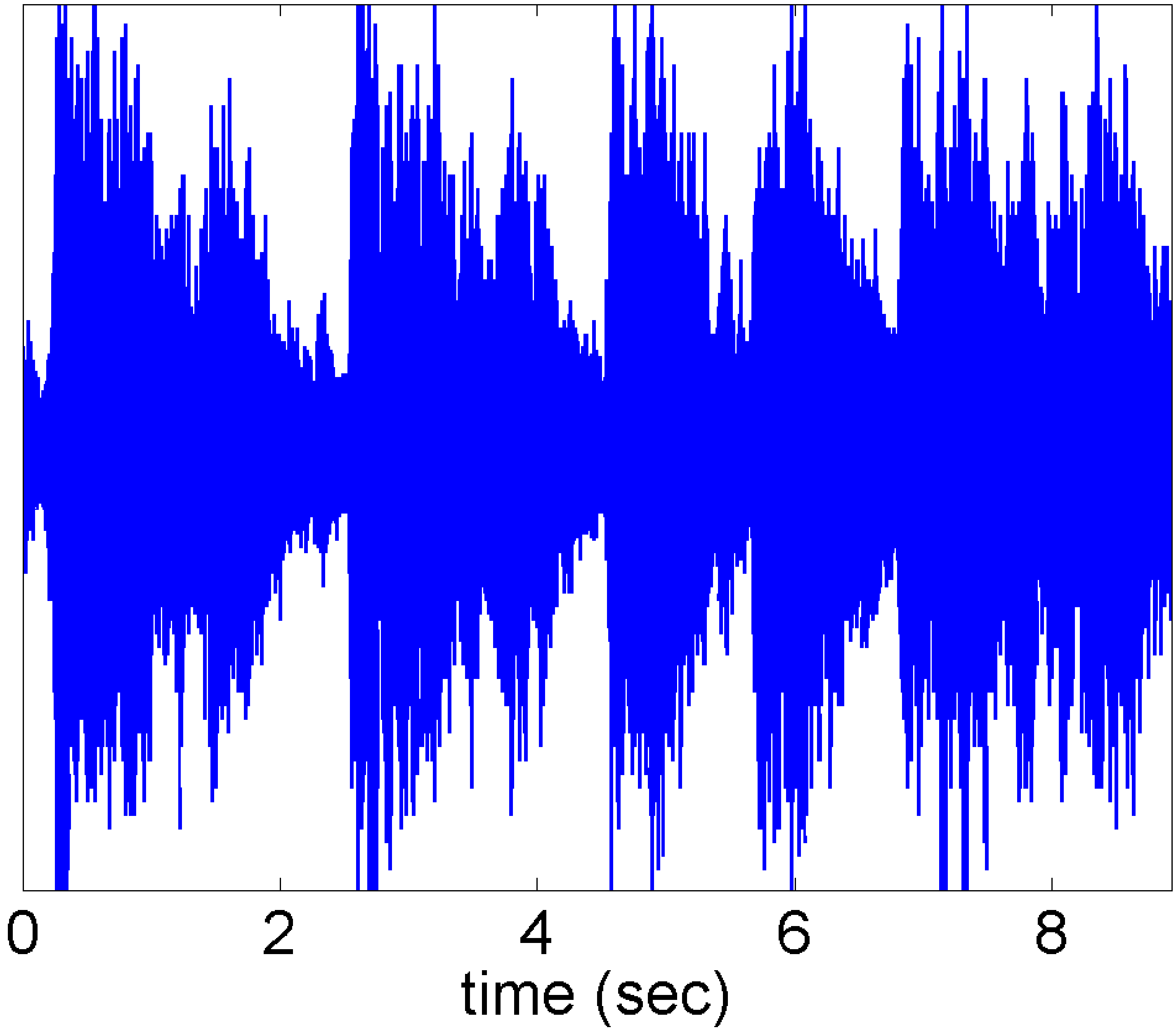
- Sinusoids go on forever …



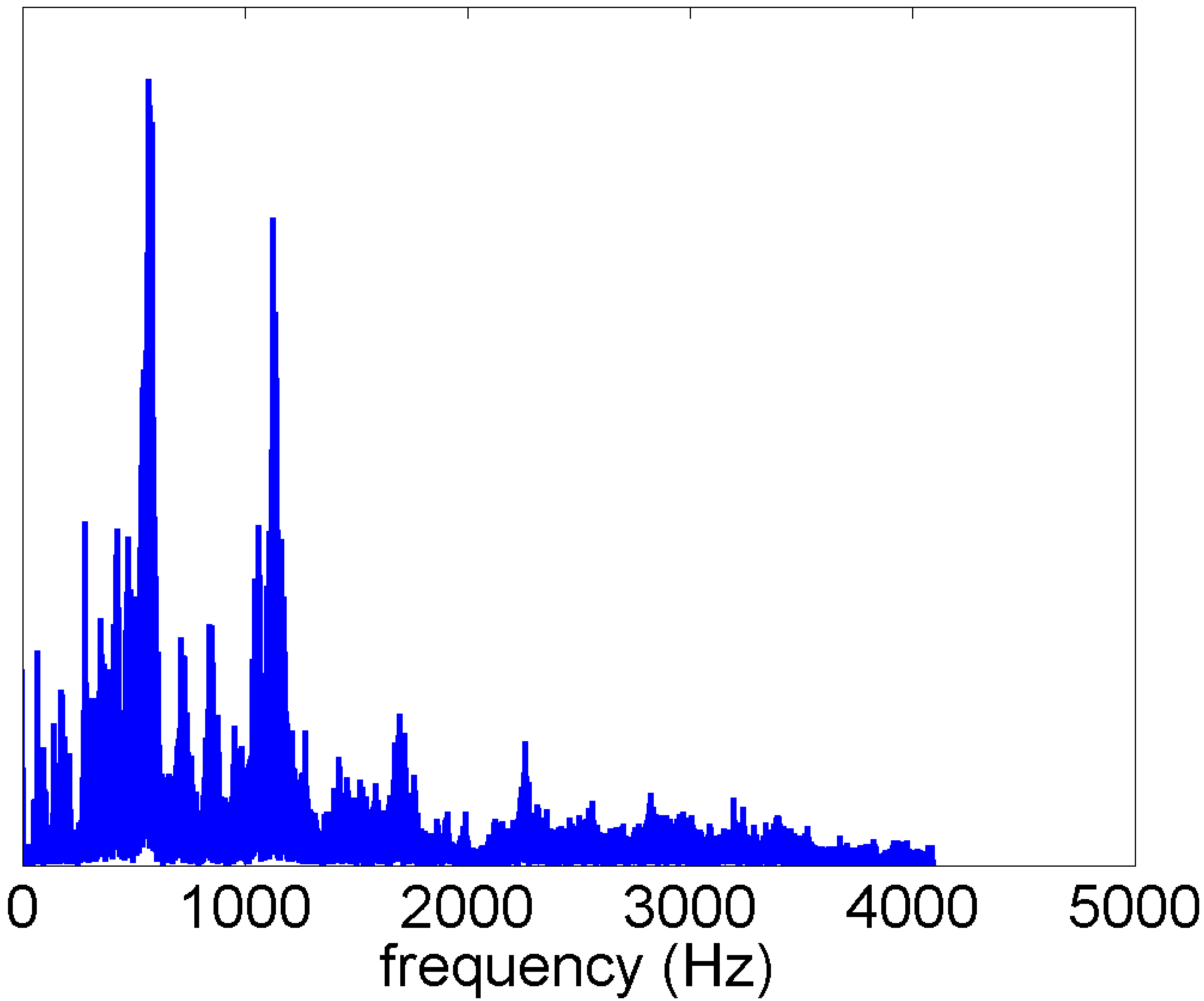
…

- Short-time Fourier transform
- Wigner–Ville distribution (1932)
- Gabor transform (1946)

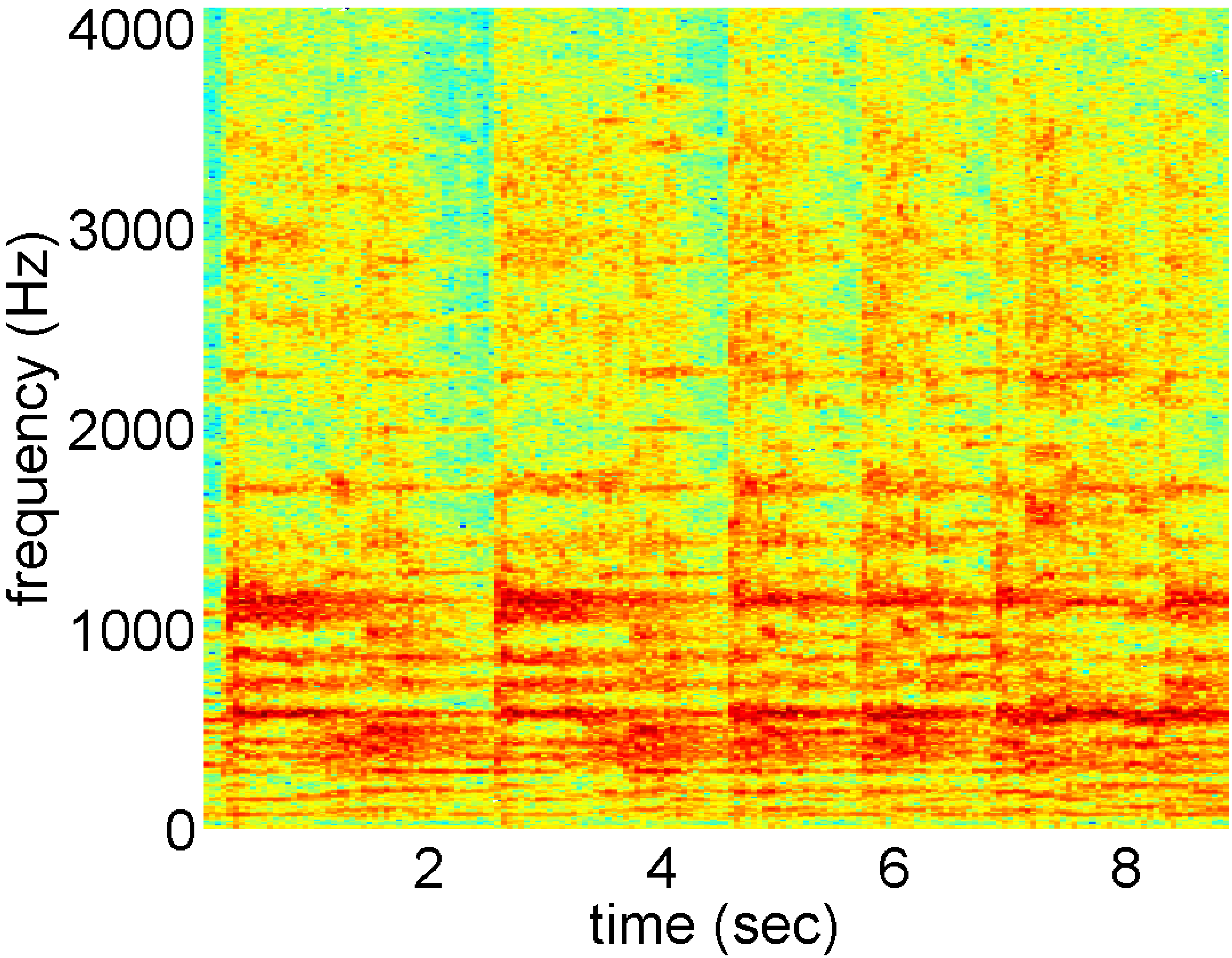
# Recall: Audio Signal



# Frequency Spectrum

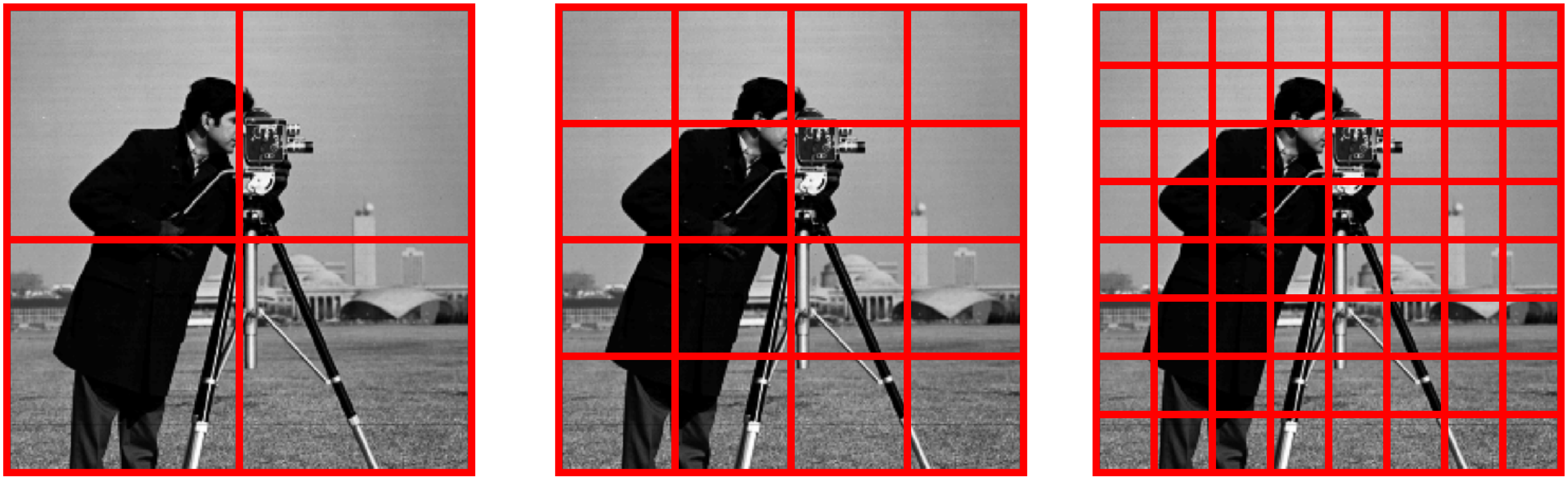


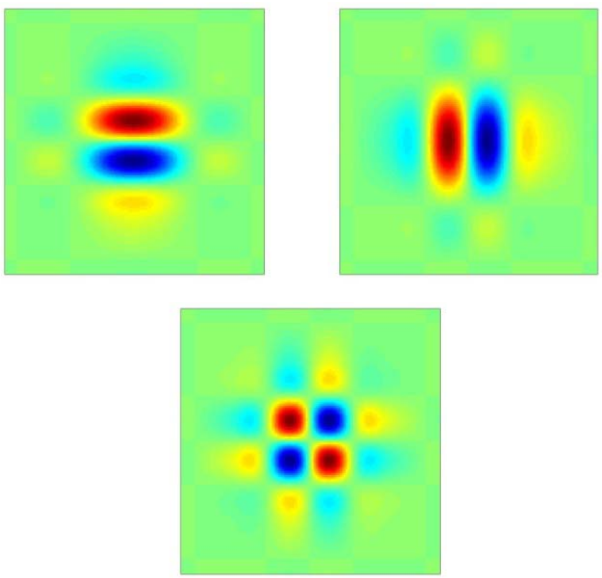
# Spectrogram

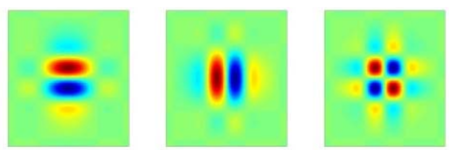


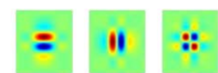
# Wavelet Analysis (1980s-1990s)

[Morlet, Grossman, Meyer, Daubechies, Mallat, …]









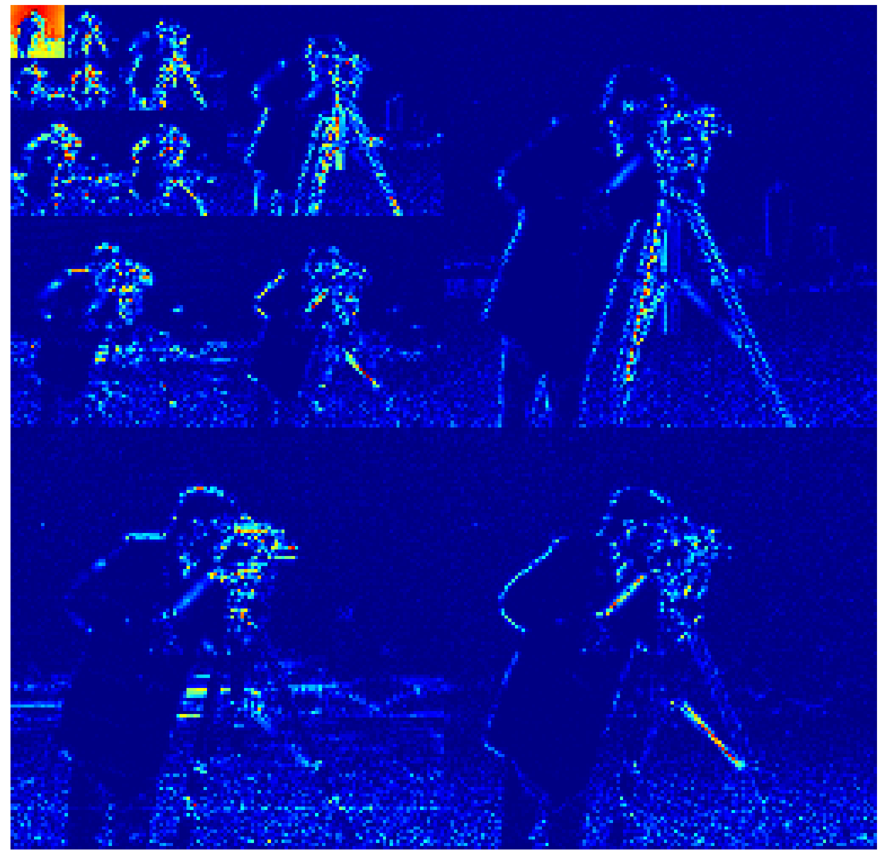
Wavelet analysis

- –multiscale
- –local
- –computationally efficient

# Wavelet Coefficients

- Which ***parts*** of a signal have high frequency behavior?

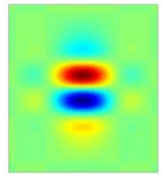
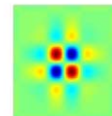
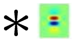
$N$-pixel image

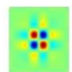
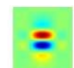
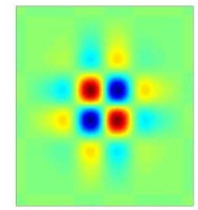


$N$ wavelet coefficients

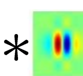
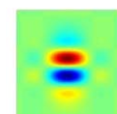
# Wavelets as Edge Detectors

# Wavelets as Building Blocks

= 65.3*  + 32.7*  + 17.4* 

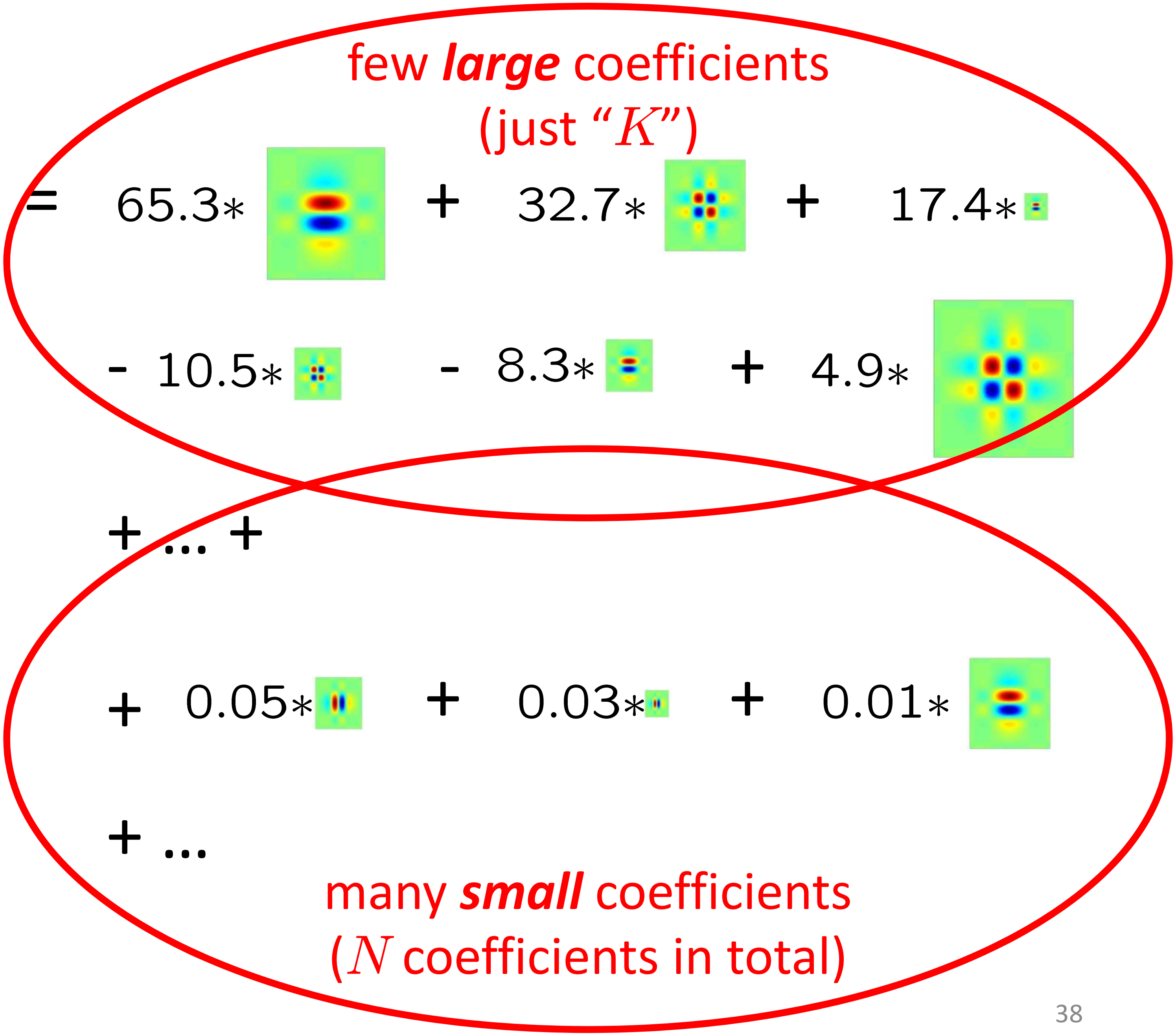
- 10.5*  - 8.3*  + 4.9* 

+ ... +

+ 0.05*  + 0.03*  + 0.01* 

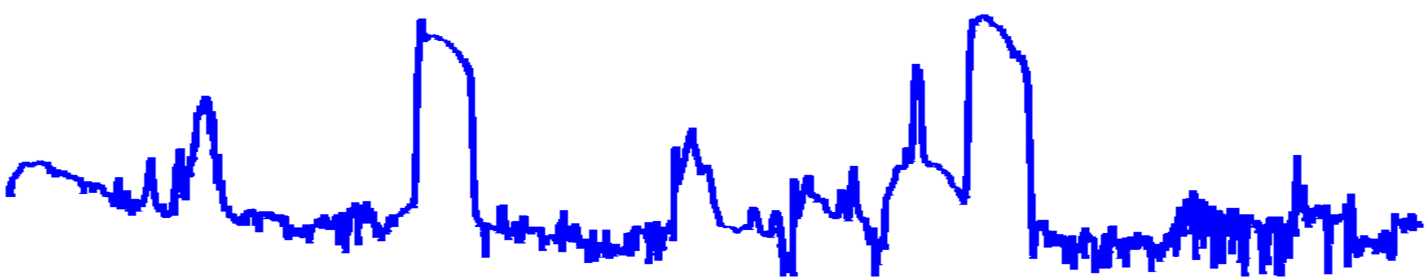
+ ...

# *Sparsity* in Wavelet Coefficients

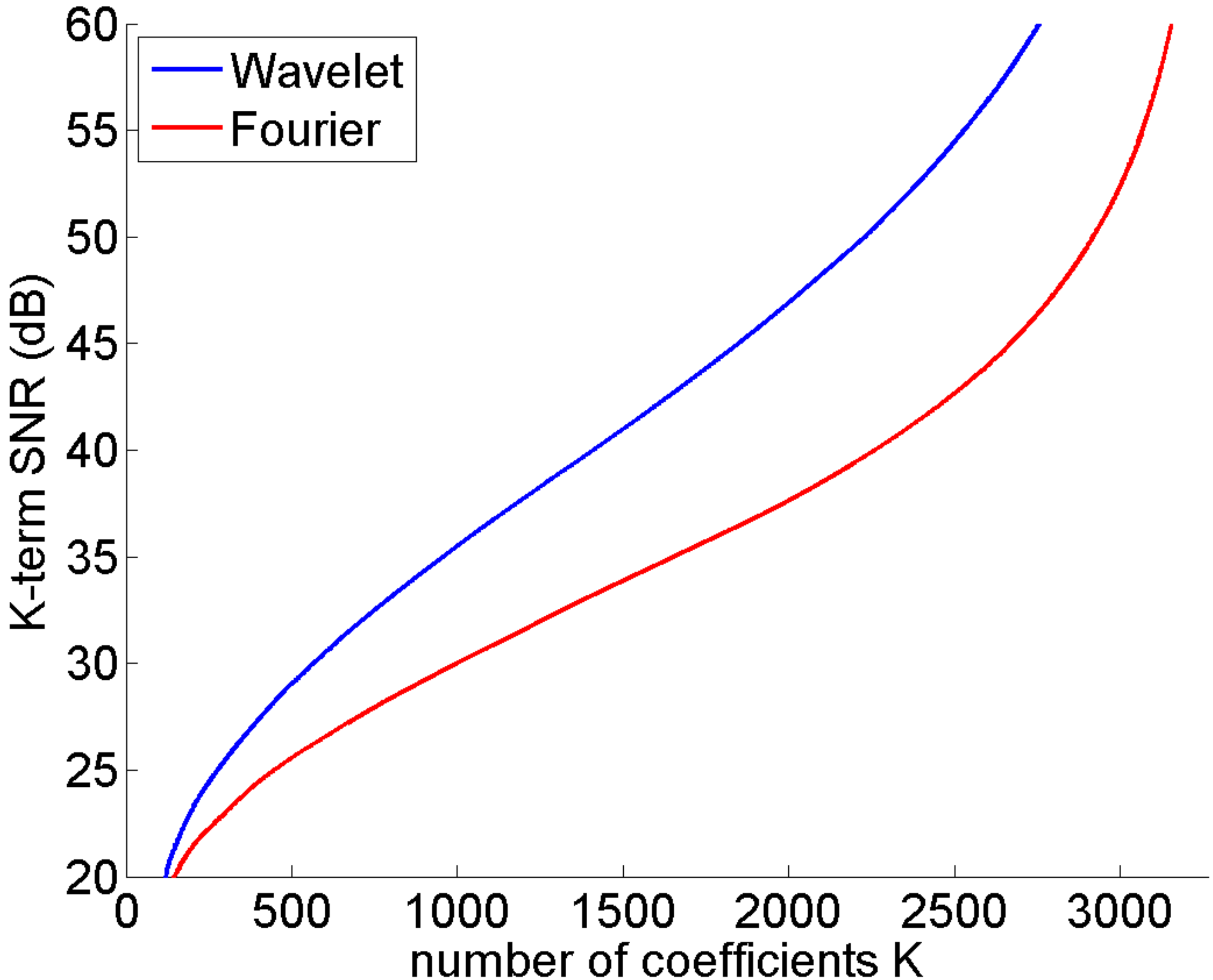


# Wavelet Advantage

- Wavelets capture the energy of many natural signals more economically than sinusoids
- Example signal:



# Compression: Wavelet vs. Fourier



# Sparsity in General

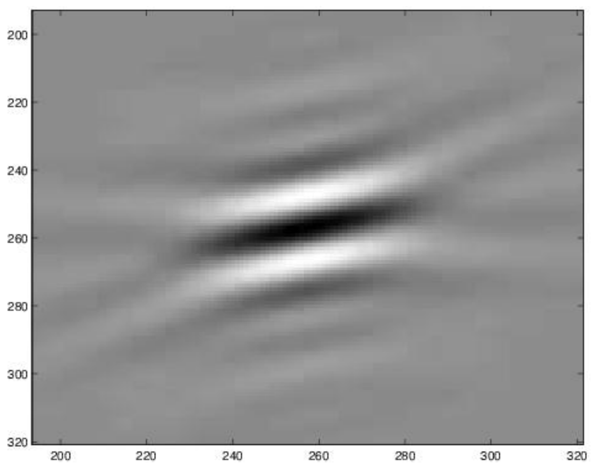
- Sparsity can exist in many domains
  - time (spike trains)
  - frequency (sparse spectrum)
  - wavelets (piecewise smooth signals)
  - curvelets (piecewise smooth images)

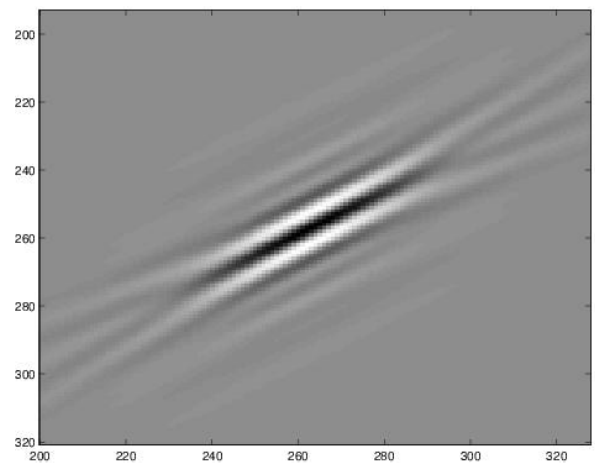
- The 1990s and 2000s witnessed many new Xlets
  - shearlets, bandelets, vaguelettes, contourlets, chirplets, ridgelets, beamlets, brushlets, wedgelets, platelets, surflets, seislets, …

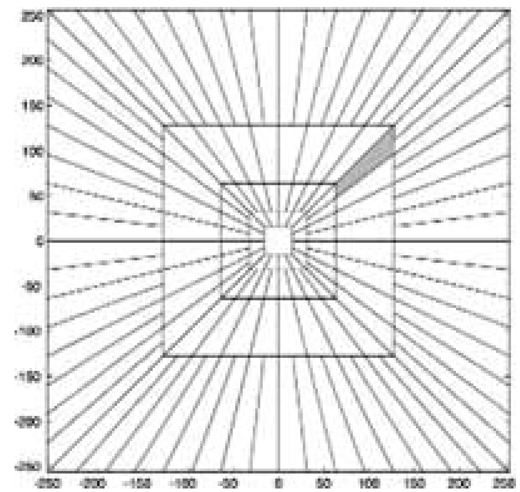
# The Curvelet Transform (1999)

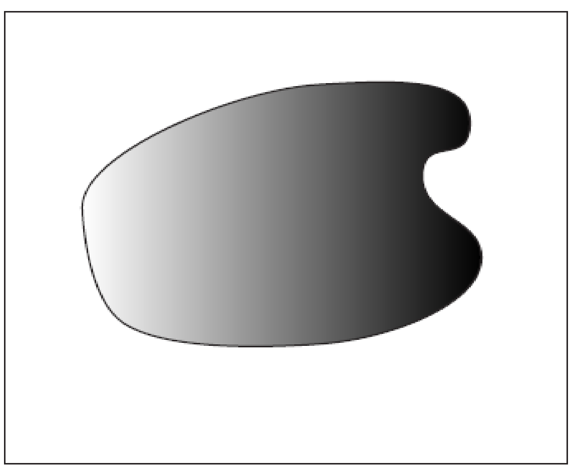
[Candès and Donoho]

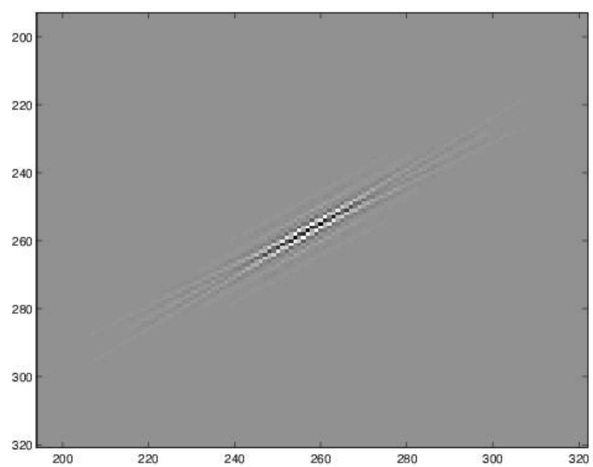
- Sparse representation for piecewise smooth images with smooth discontinuities

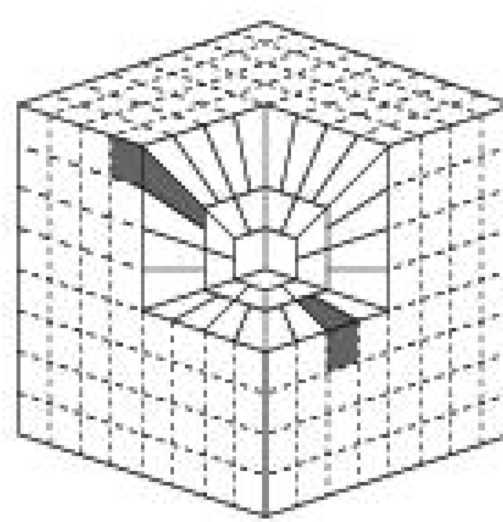


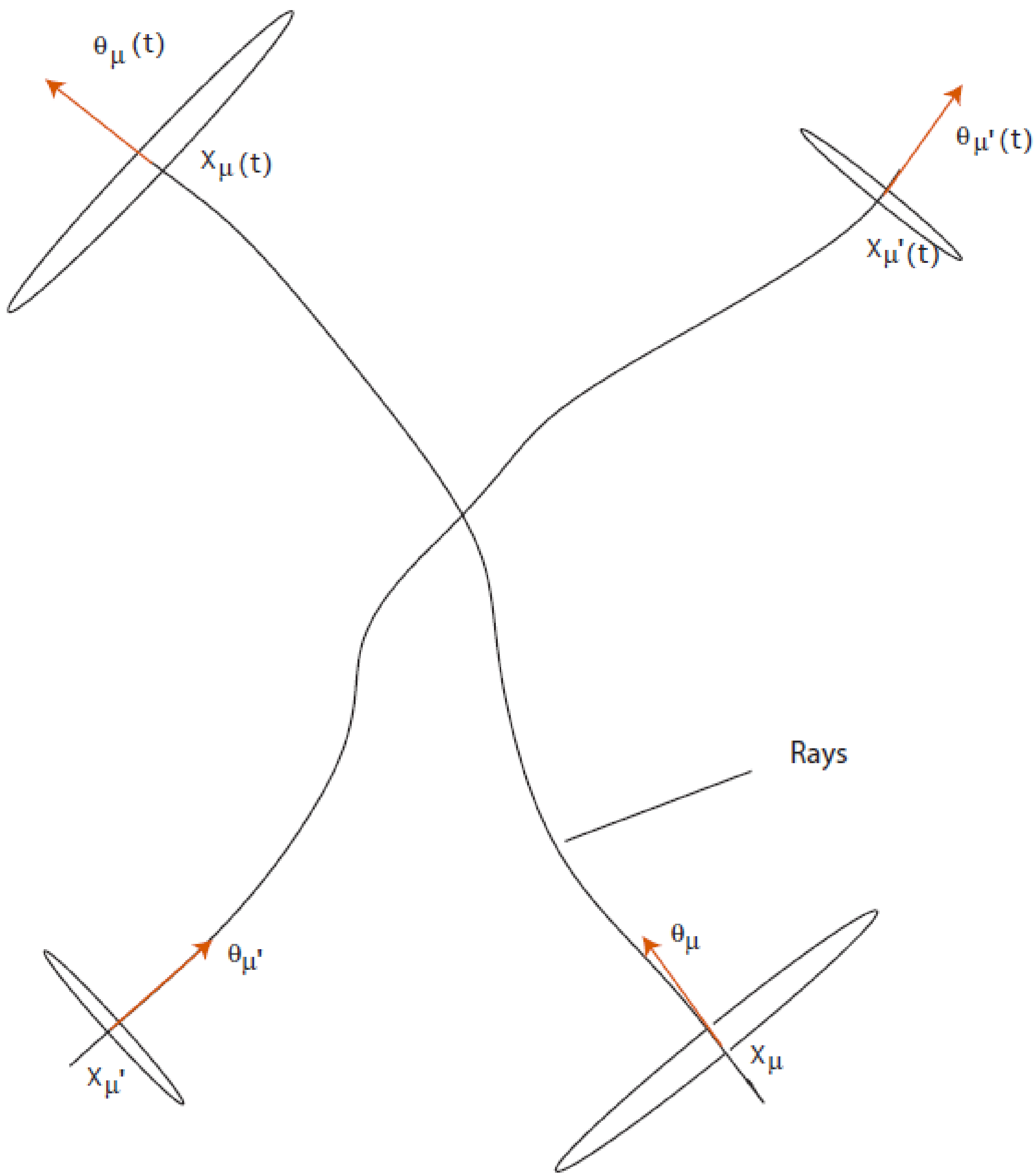




# Curvelets and the Wave Equation

[Candès and Demanet]



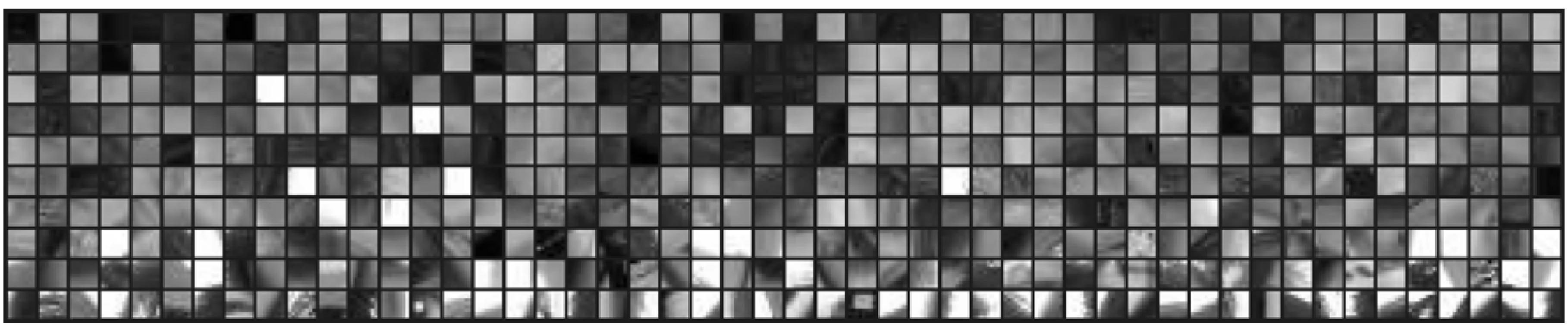

propagation of a curvelet

||

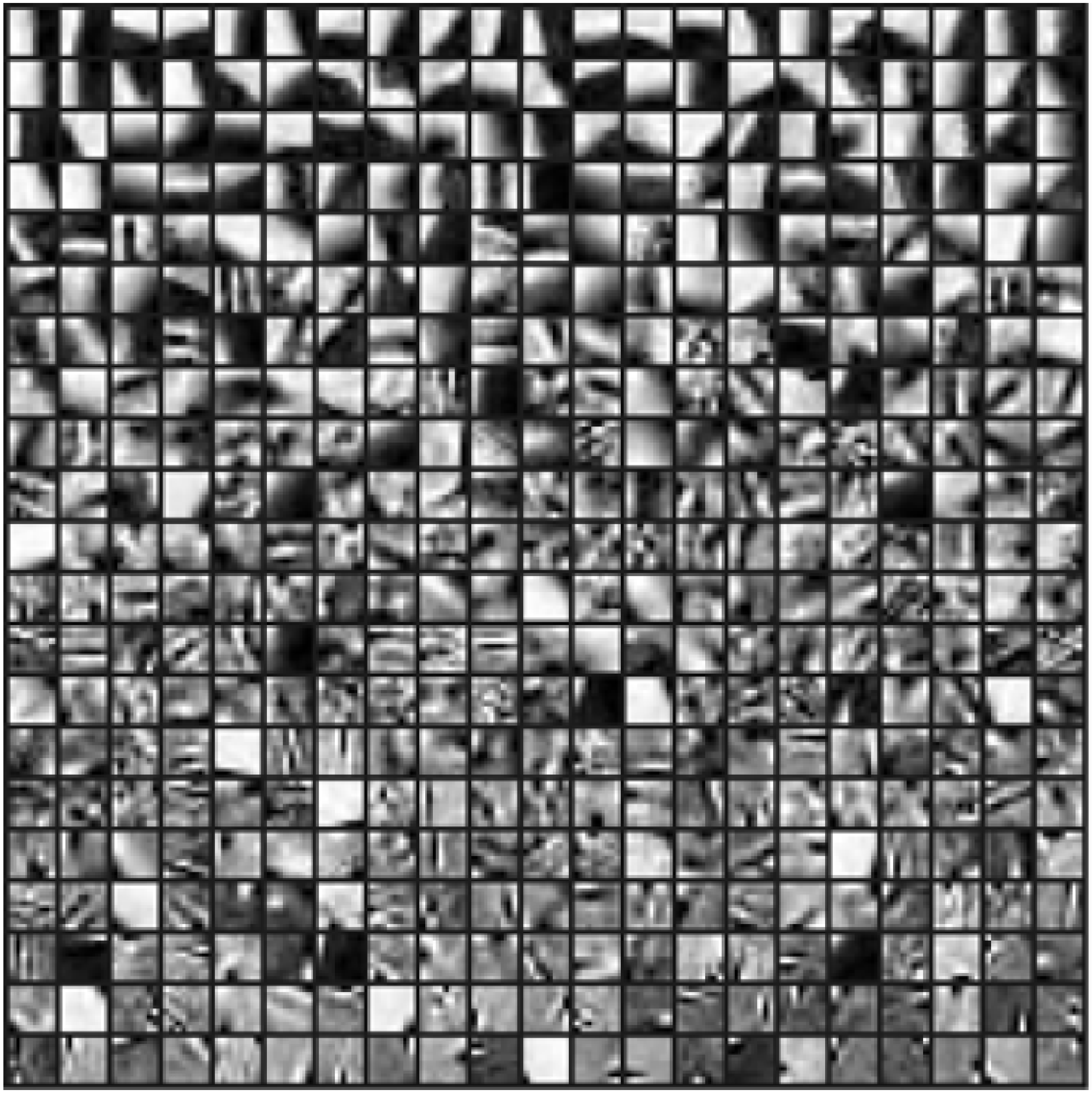
sparse sum of curvelets
at nearby scales,
orientations, locations

# Dictionary Learning

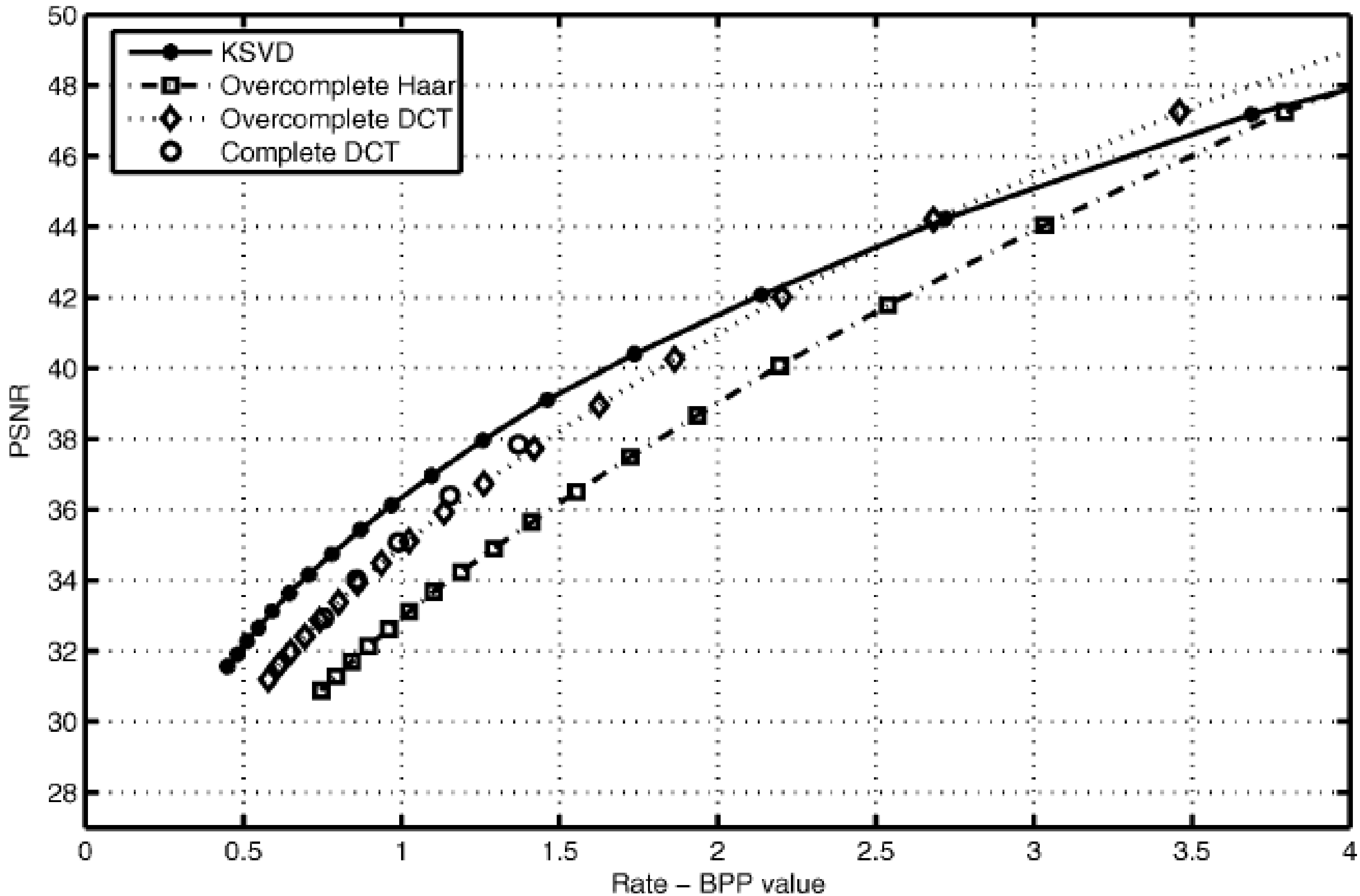
- Sparse transforms can also be ***learned*** from collections of training data
- Example algorithm: K-SVD (2006) [Aharon et al.]
- Demonstration: image patches



# Learned Dictionary



# Compression Performance



# Compression Performance

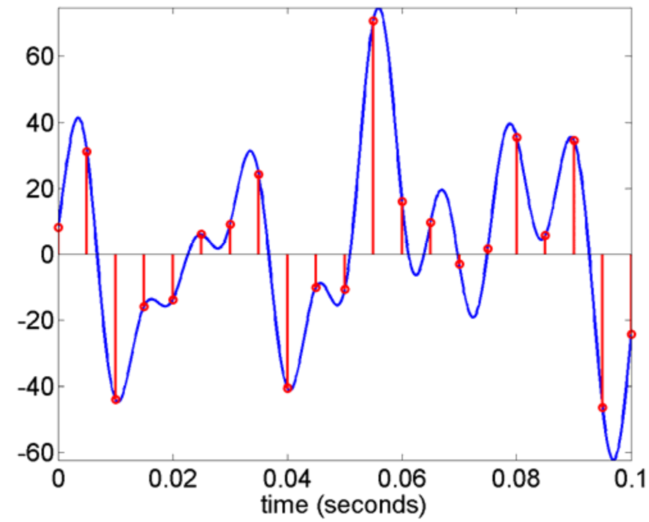
K−SVD dictionary
8 bits per coefficients
PSNR =34.1564
Rate = 0.70651 BPP

Overcomplete DCT dictionary
8 bits per coefficients
PSNR =32.4021
Rate = 0.69419 BPP

Complete DCT dictionary
8 bits per coefficients
PSNR =32.3917
Rate = 0.70302 BPP

# Implications of Sparsity

- When we sample a signal (say to obtain $N$ samples or pixels), there may be some residual ***structure*** in those samples.




- This could permit us to further compress the samples.

- Or… we could reduce the number of samples we collect, and just fill in the blanks later.

# 2. What do we mean by “blanks”?

# New Ways of Sampling

- Recall the two ingredients:
  1. Model: “How is the signal supposed to behave?”
  2. Algorithm: “How can I fill in the blanks so that the signal obeys the model?”

- Now, we will suppose that our signal is sparse in some particular transform.

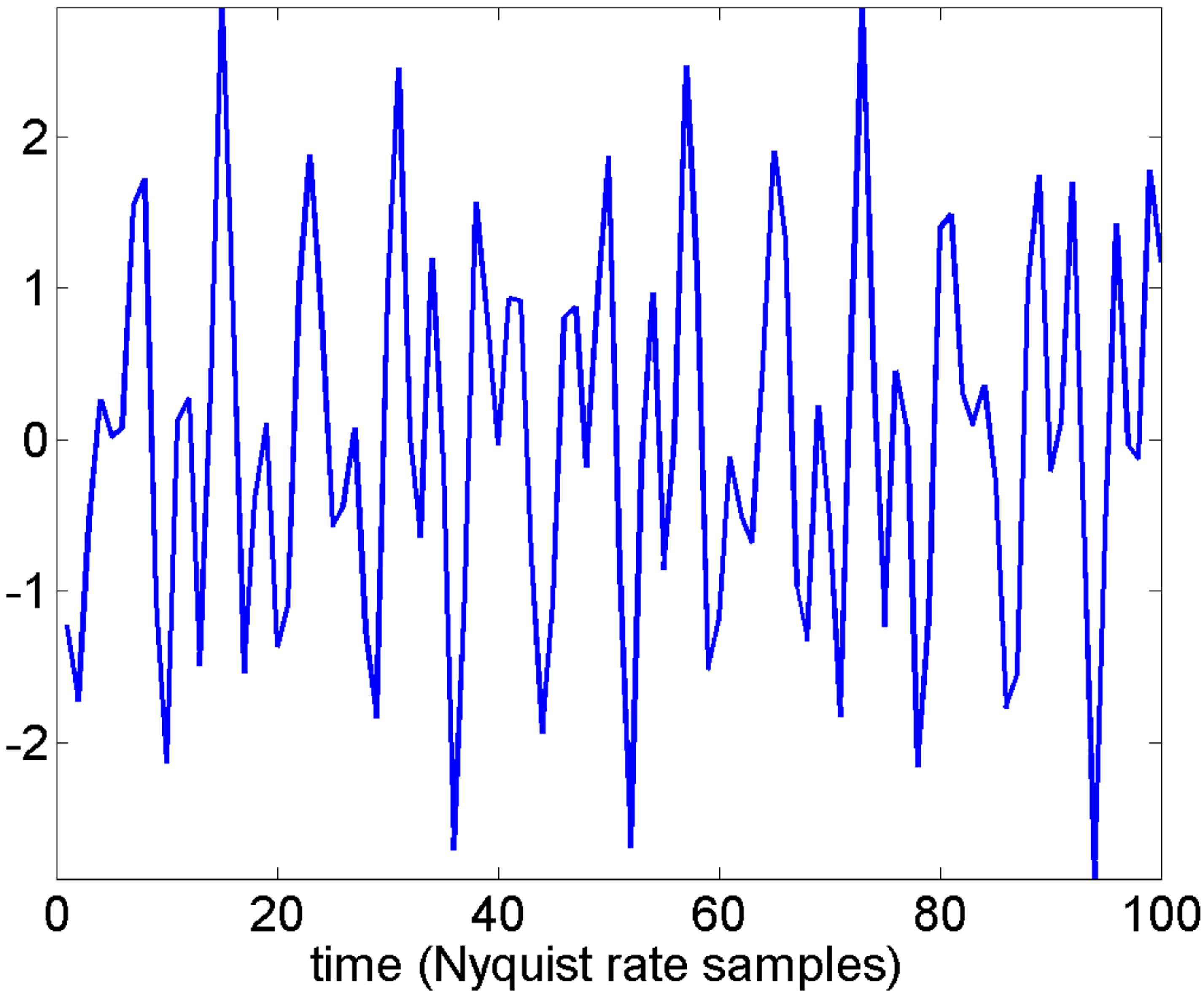
- We must be careful about how we “measure” the signal so that we don’t miss anything important.

# Don’t

- Don’t just take uniform samples more slowly.

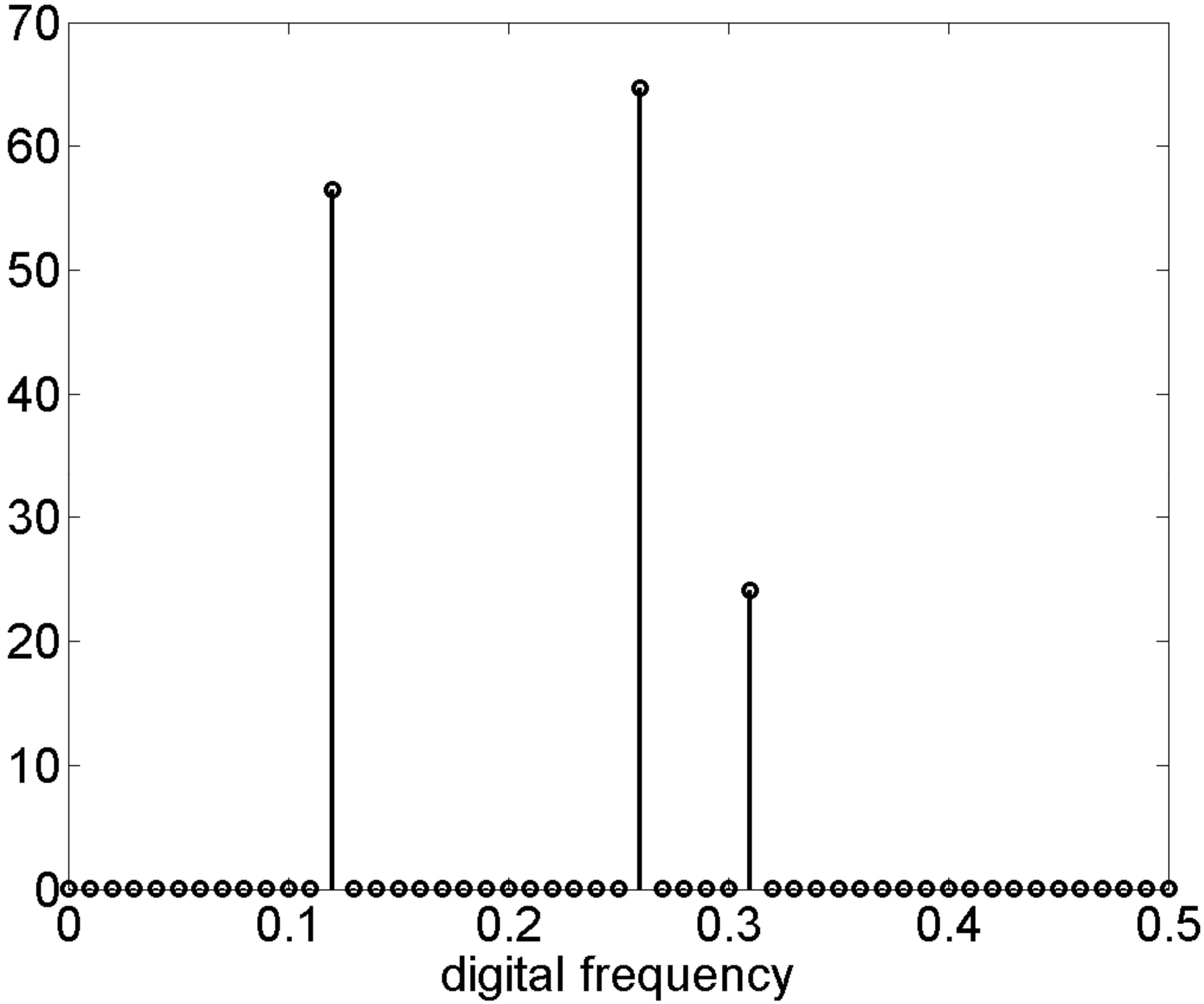
- Aliasing becomes a risk – there will be more than one way to fill in the blanks using a sparse signal.

# Example Signal with Sparse Spectrum

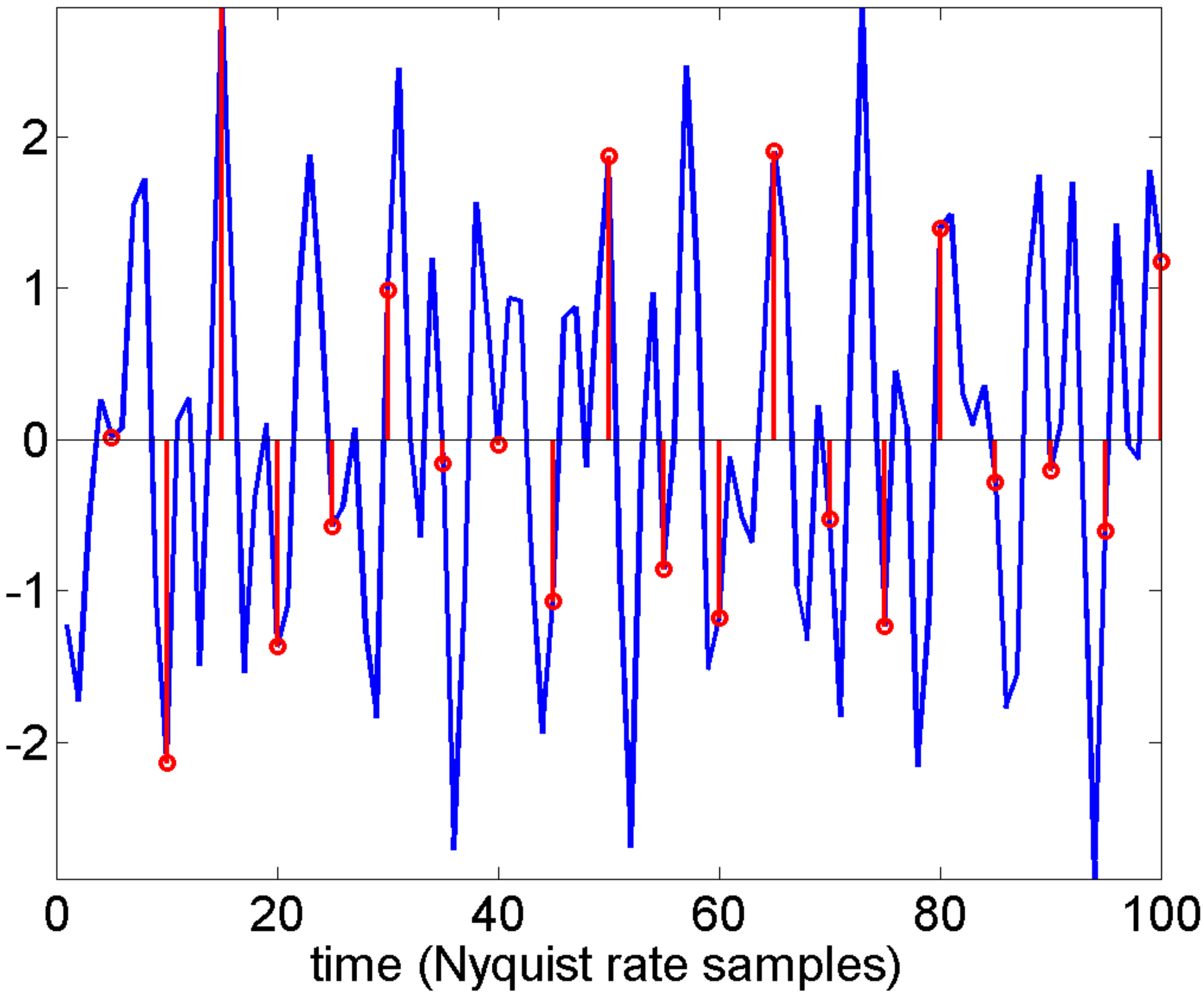



$N = 100$ Nyquist-rate samples

# Sparse Spectrum

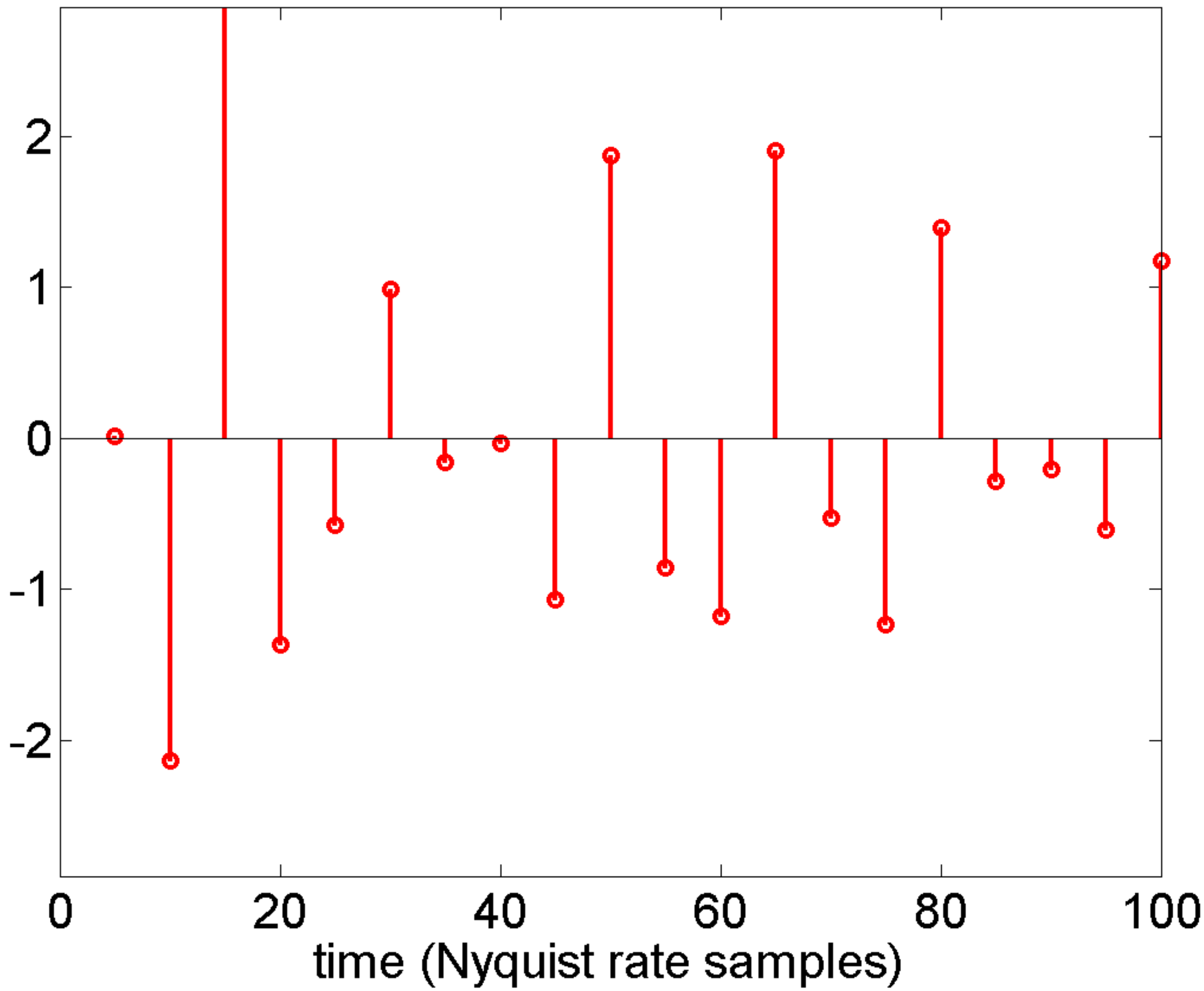


# Uniform Samples (5x Sub-Nyquist)



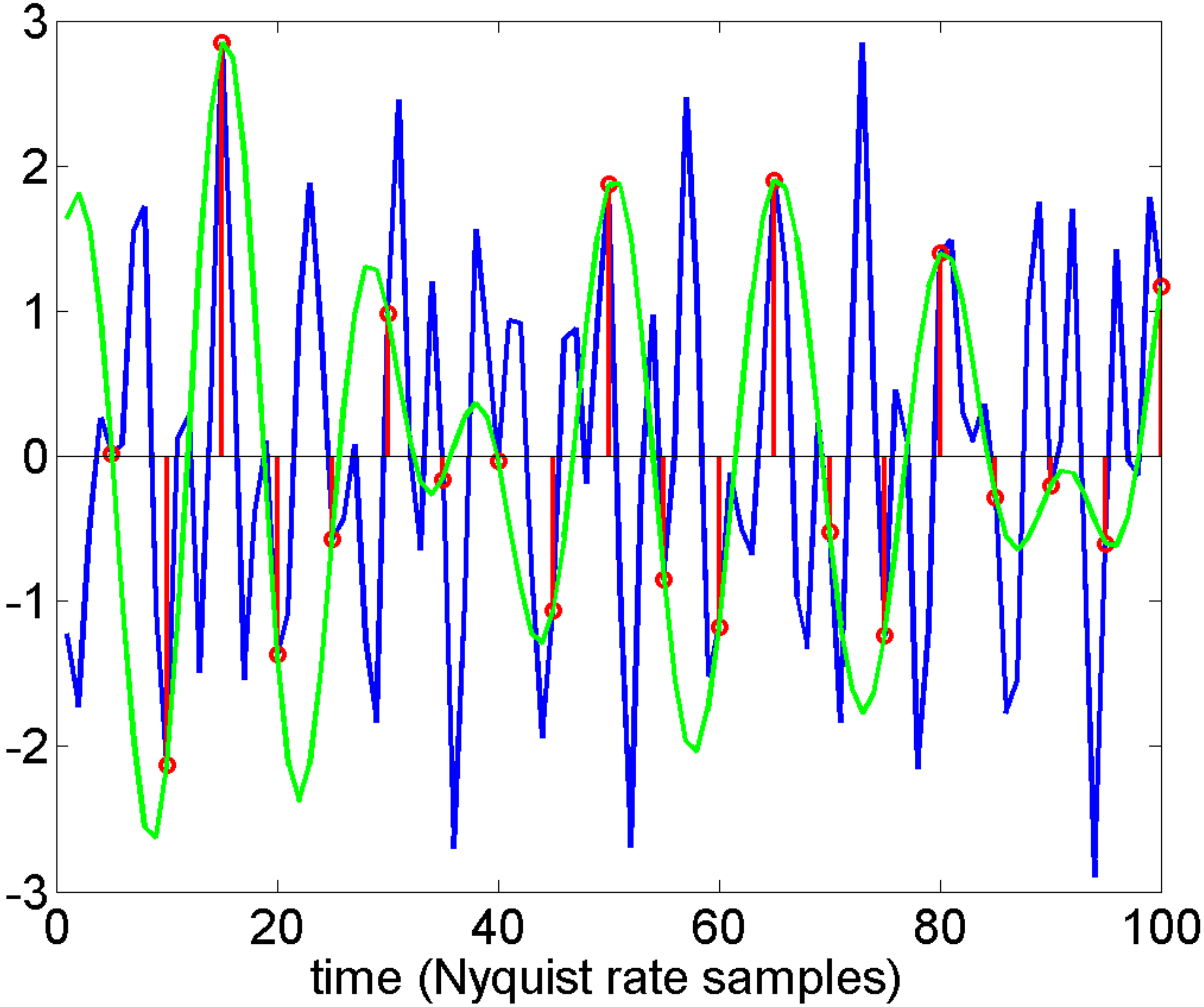

$20$ uniform sub-Nyquist samples

# Uniform Samples (5x Sub-Nyquist)

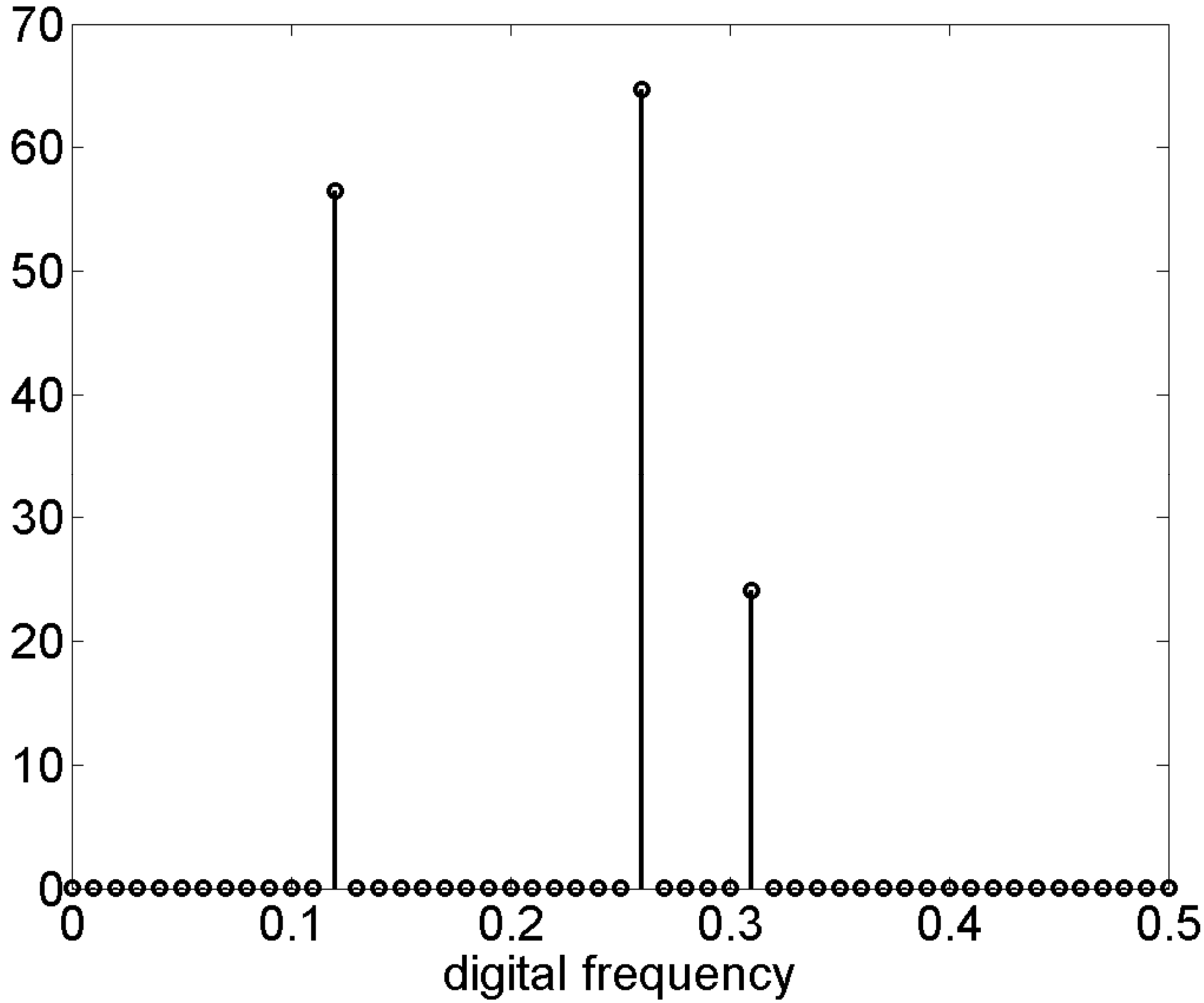



$20$ uniform sub-Nyquist samples

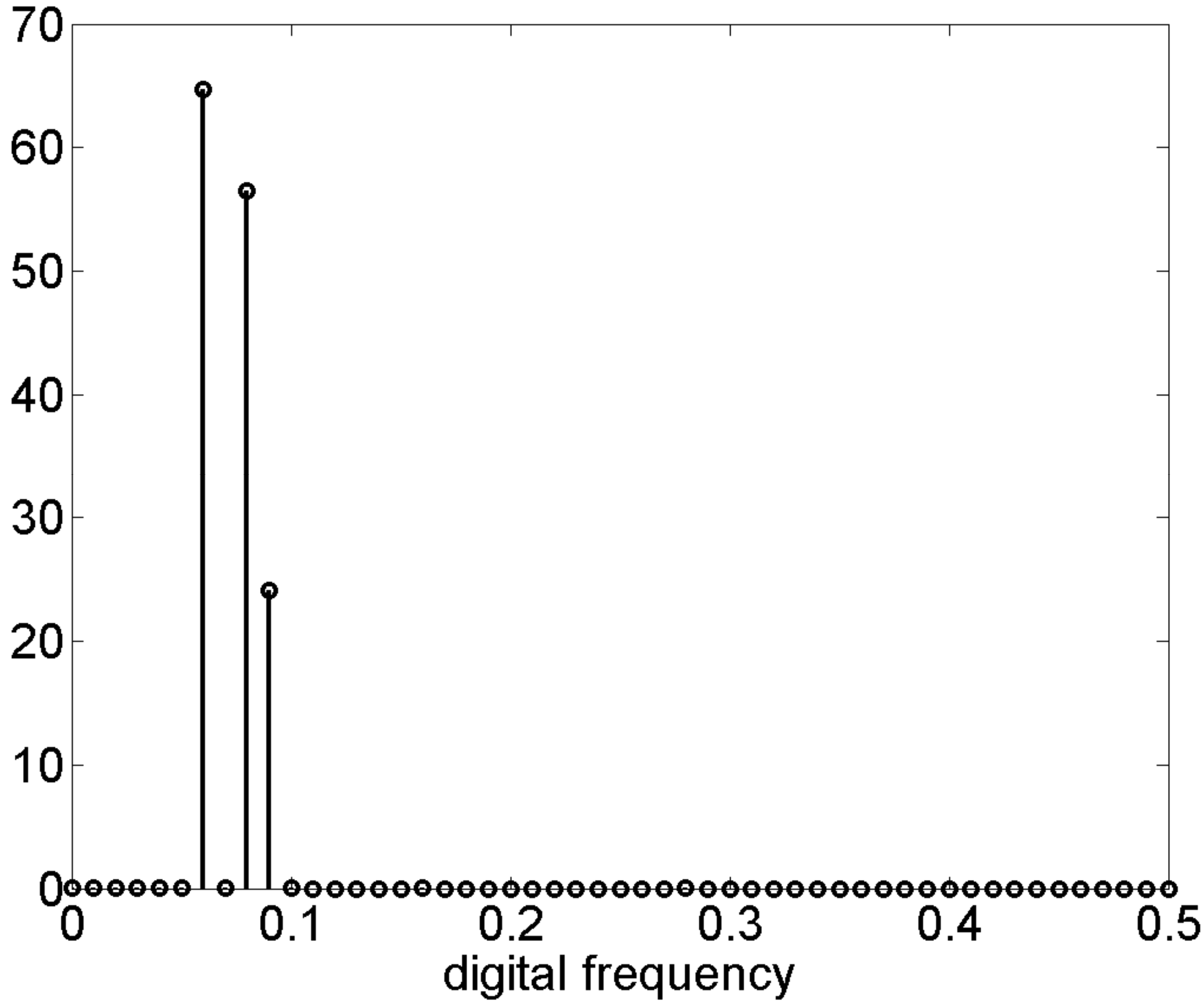
# Alias with Sparse Spectrum



# Original Sparse Spectrum



# Sparse Spectrum of the Alias



# Do

- Do take measurements that are “incoherent” with the sparse structure (we will explain).

- This ensures that there is only one way to fill in the blanks using a signal that is sparse.

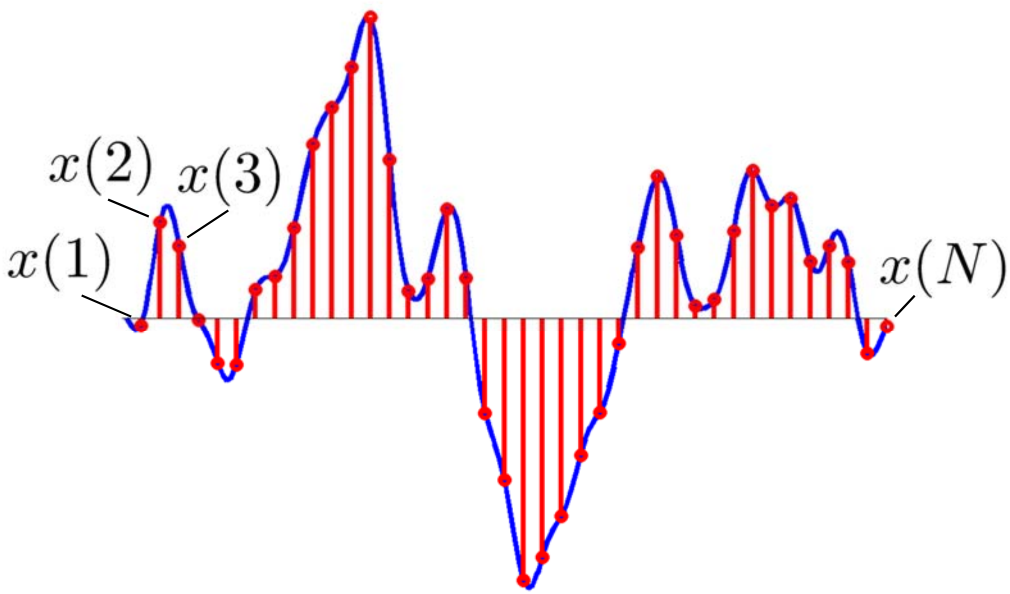
- Incoherence is often achieved using:

  (1) non-uniform samples or

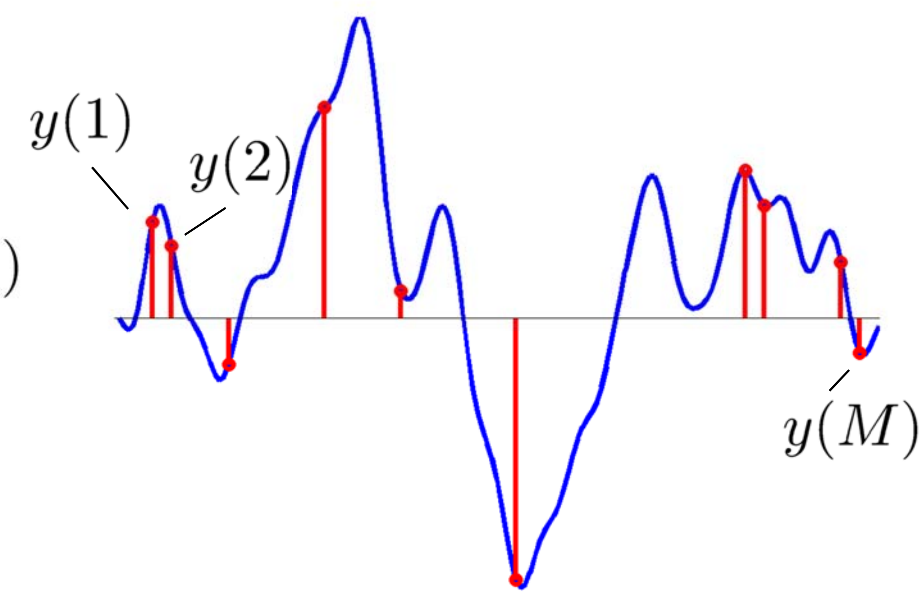
  (2) generic linear measurements

  along with some amount of ***randomness***.

# Non-uniform Sampling



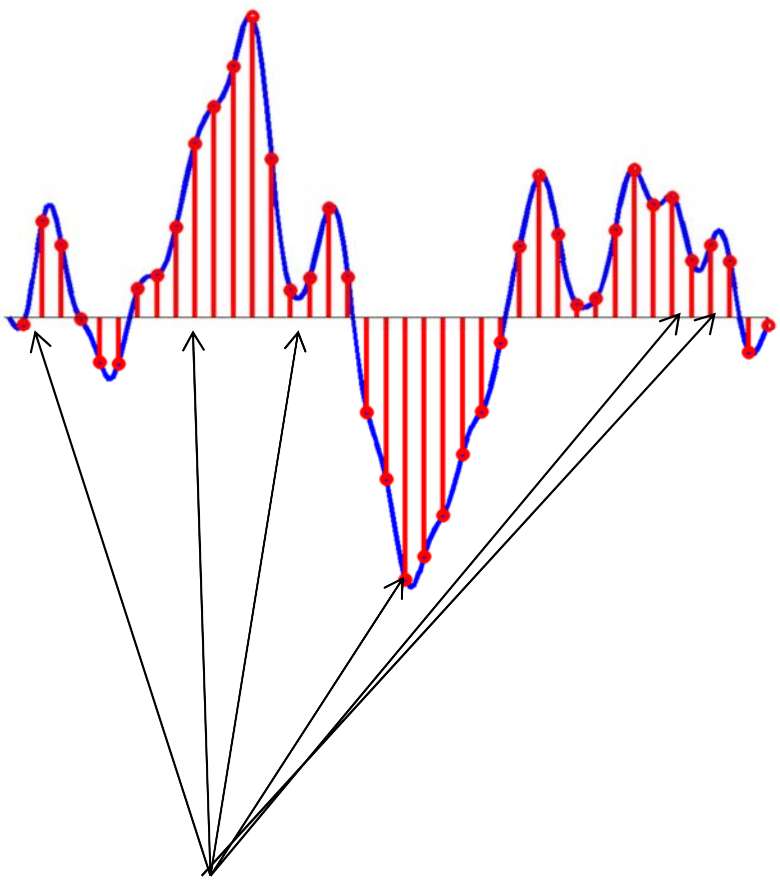

$N$ Nyquist-rate samples




$M$ non-uniform samples

$$M < N$$

# Common Approach: Sub-sample Nyquist Grid

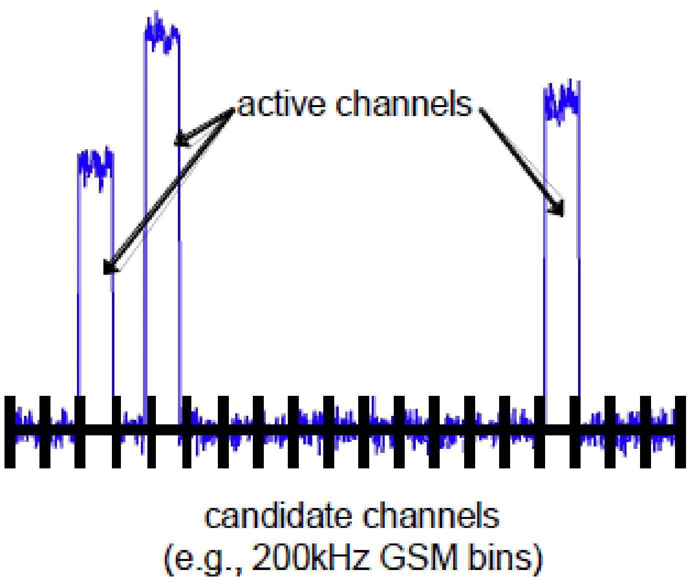


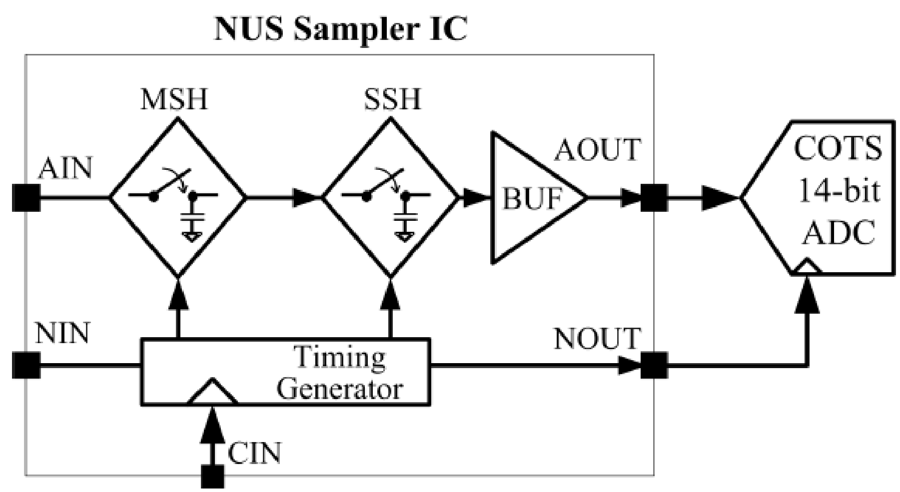
Sample/retain only a random subset of the Nyquist samples

# Example: Non-uniform Sampler

[with Northrop Grumman, Caltech, Georgia Tech]

- Underclock a standard ADC to capture spectrally sparse signals



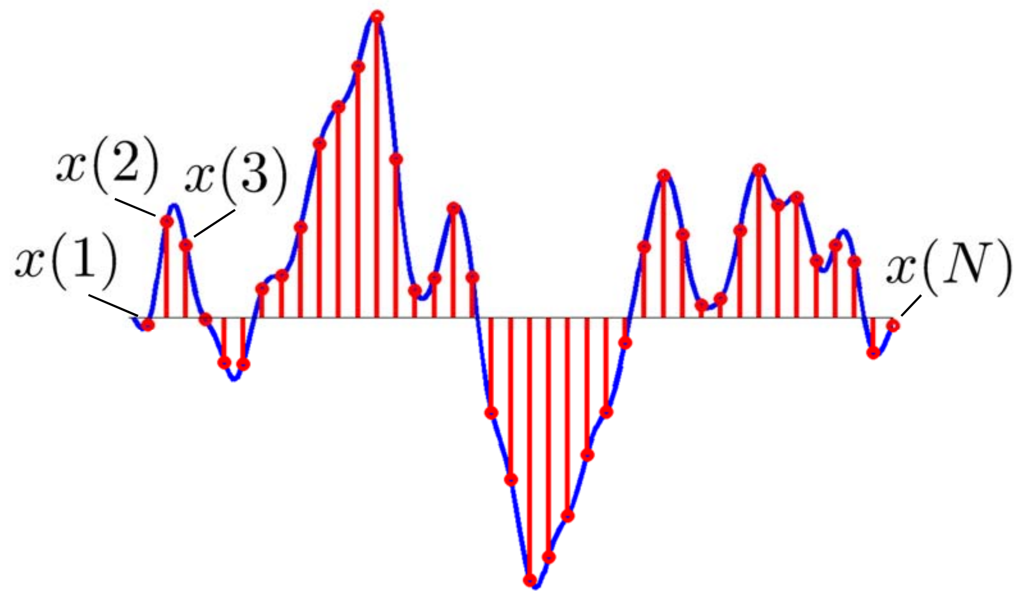




- Prototype system:
  - captures 1.2GHz+ bandwidth (800MHz to 2GHz) with 400MHz ADC underclocked to 236MHz (10x sub-Nyquist)
  - up to 100MHz total occupied spectrum

# Generic Linear Measurements

Nyquist-rate samples

(don't record these directly)




Generic linear measurements

$$
\begin{aligned}
y(1) &= 3.2 * x(1) - 2.7 * x(2) + \cdots + 8.1 * x(N) \\
y(2) &= -8.1 * x(1) + 0.3 * x(2) + \cdots - 0.4 * x(N) \\
&\vdots \\
y(M) &= 4.2 * x(1) + 1.7 * x(2) + \cdots + 0.7 * x(N)
\end{aligned}
$$

# Generic Linear Measurements

Nyquist-rate samples

(don't record these directly)


Generic linear measurements

$$
\begin{aligned}
y(1) &= 3.2 * (\text{pixel } 1) - 2.7 * (\text{pixel } 2) + \cdots \\
y(2) &= -8.1 * (\text{pixel } 1) + 0.3 * (\text{pixel } 2) + \cdots \\
&\vdots \\
y(M) &= 4.2 * (\text{pixel } 1) + 1.7 * (\text{pixel } 2) + \cdots
\end{aligned}
$$

# Then What?

- Filling in the blanks means ***reconstructing*** the original $N$ Nyquist-rate samples from the measurements.

$$\begin{aligned} y(1) &= 3.2 * (\text{pixel } 1) - 2.7 * (\text{pixel } 2) + \cdots \\ y(2) &= -8.1 * (\text{pixel } 1) + 0.3 * (\text{pixel } 2) + \cdots \\ &\vdots \\ y(M) &= 4.2 * (\text{pixel } 1) + 1.7 * (\text{pixel } 2) + \cdots \end{aligned}$$

# How Many Measurements?

- $N$: number of conventional Nyquist-rate samples
- $K$: sparsity level in a particular transform domain
- Sufficient number of incoherent measurements

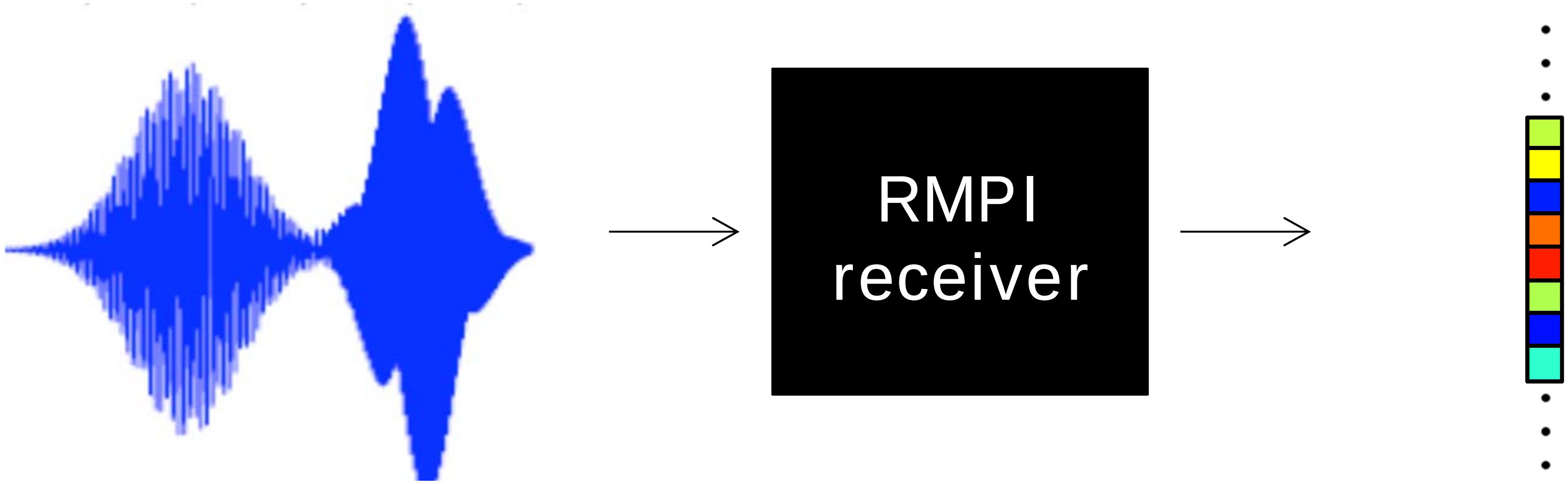
$$\boxed{M \approx K \log N}$$

- Sampling at the information level:
  - sampling frequency is proportional to sparsity
  - as model complexity increases, must take more measurements

# Random Modulation Pre-Integrator (RMPI)

[with Northrop Grumman, Caltech, Georgia Tech]

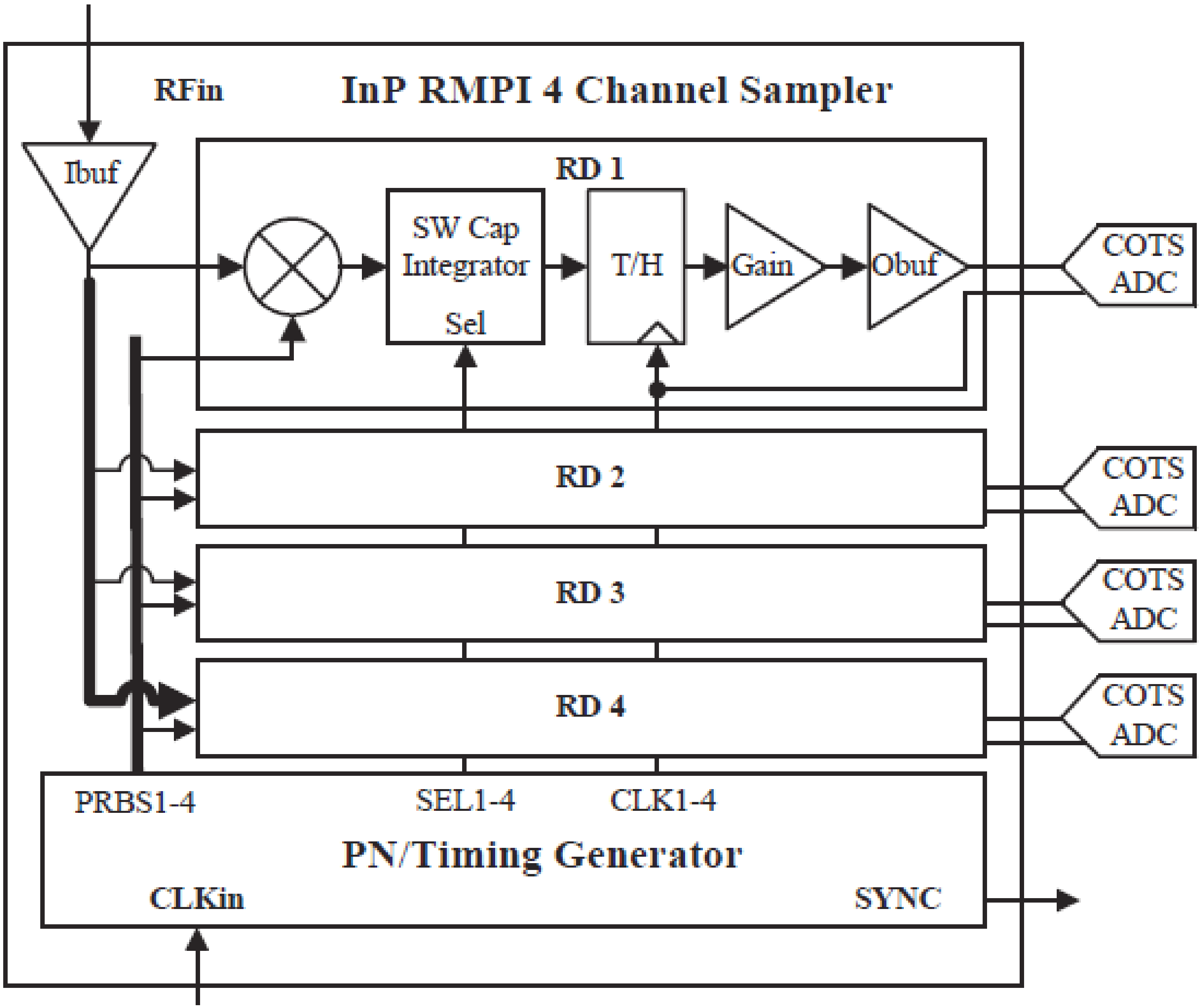
- RMPI receiver
  - four parallel "random demodulator" channels
  - effective instantaneous bandwidth spanning 100MHz—2.5GHz
  - 385 MHz measurement rate (13x sub-Nyquist)
- Goal: identify radar pulse descriptor words (PDWs)




radar pulses (analog)

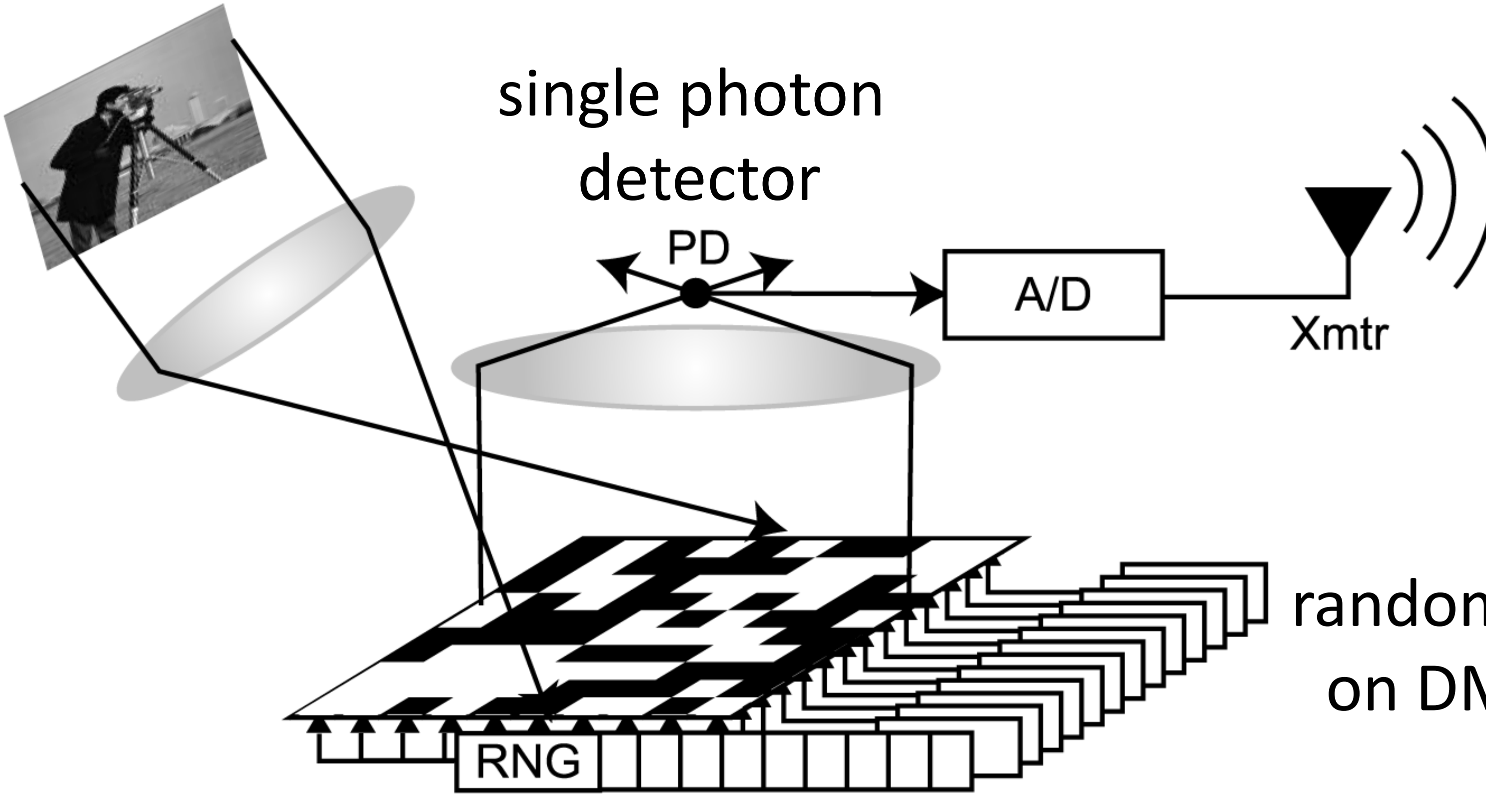
discrete, low-rate, information-carrying measurements

# RMPI Architecture



# Single-Pixel Camera

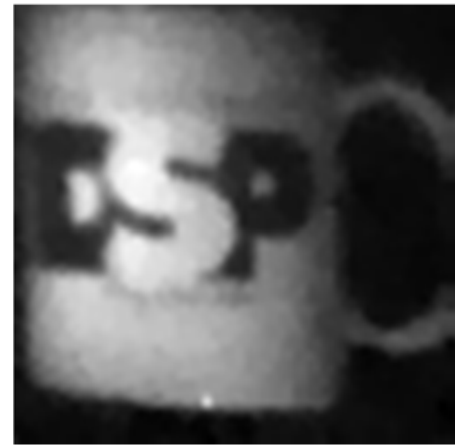
[Baraniuk and Kelly, et al.]




random pattern
on DMD array

# Single-Pixel Camera – Results

$N$ = 4096 pixels
$M$ = 1600 measurements
(40%)




true color low-light imaging

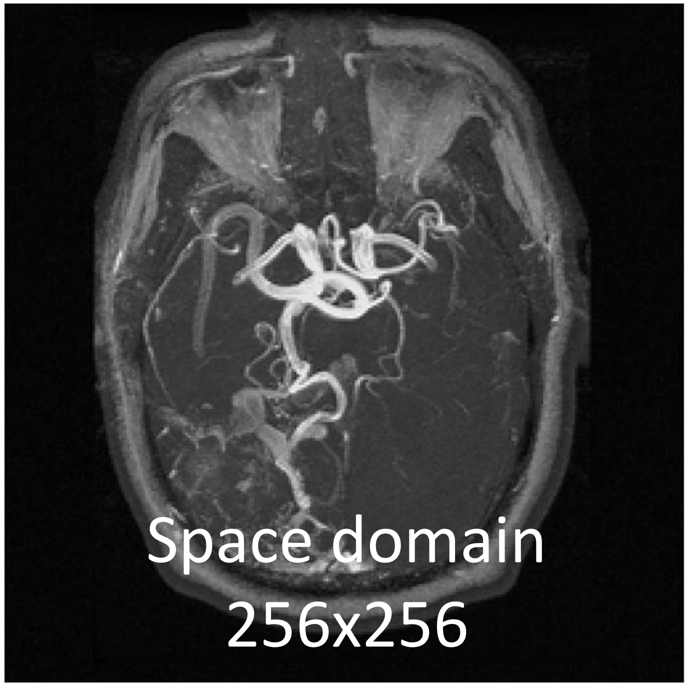
256 x 256 image with 10:1 compression

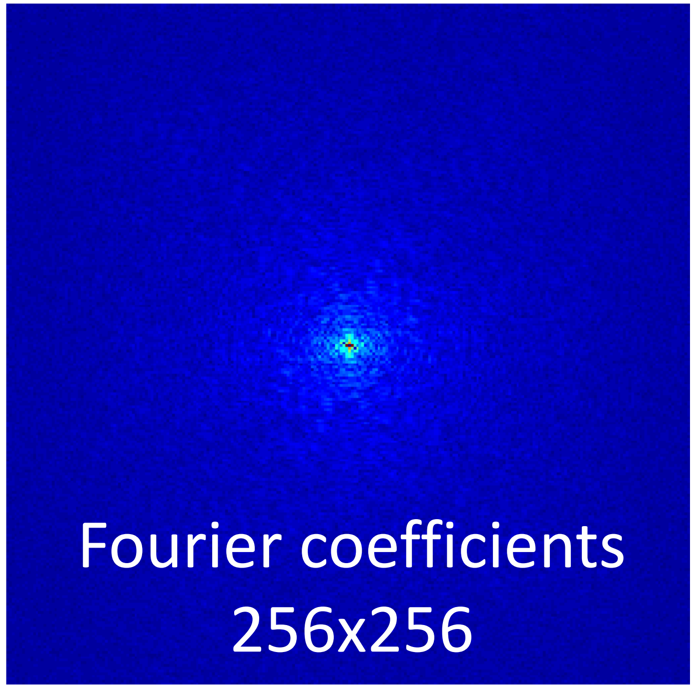
# Data-Poor Situations in the Wild

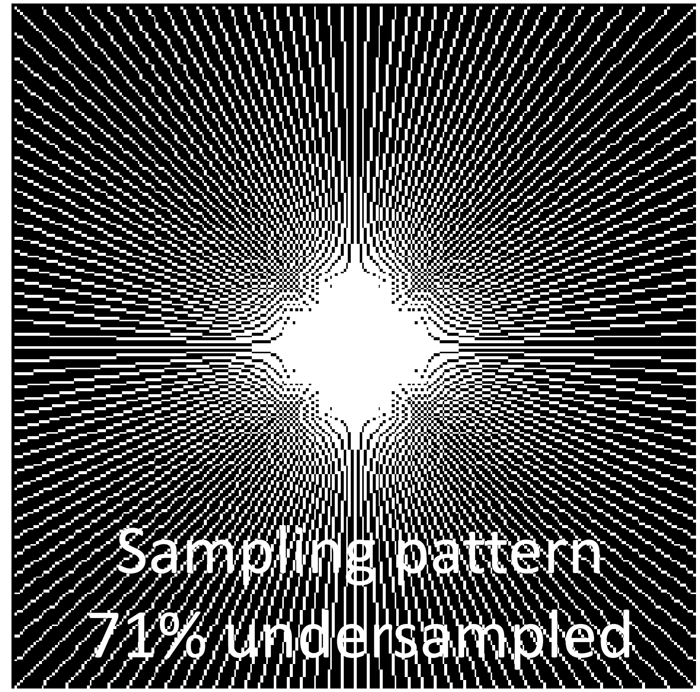
- Incomplete measurements arise naturally
  - missing/corrupted sample streams
  - missing seismic traces
  - medical imaging

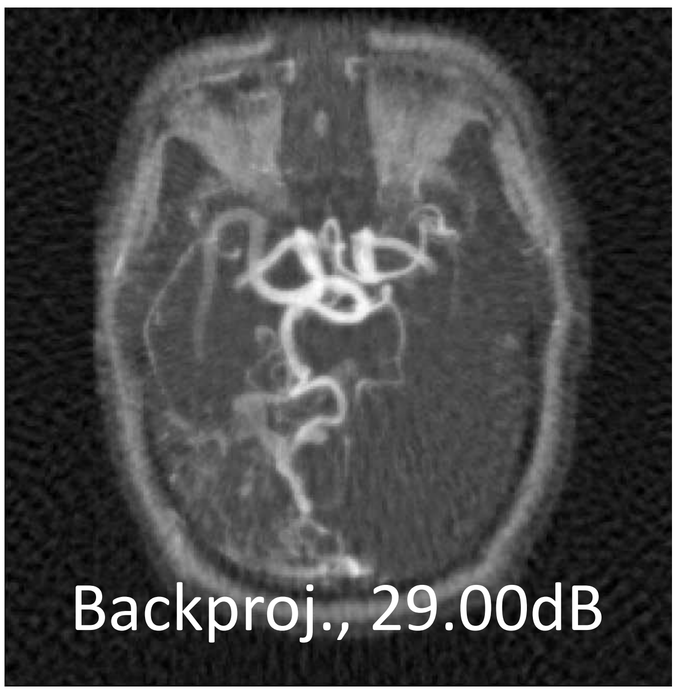
- Can try to use sparse models to fill in the blanks, even if there was nothing random about the sampling/measurement process.

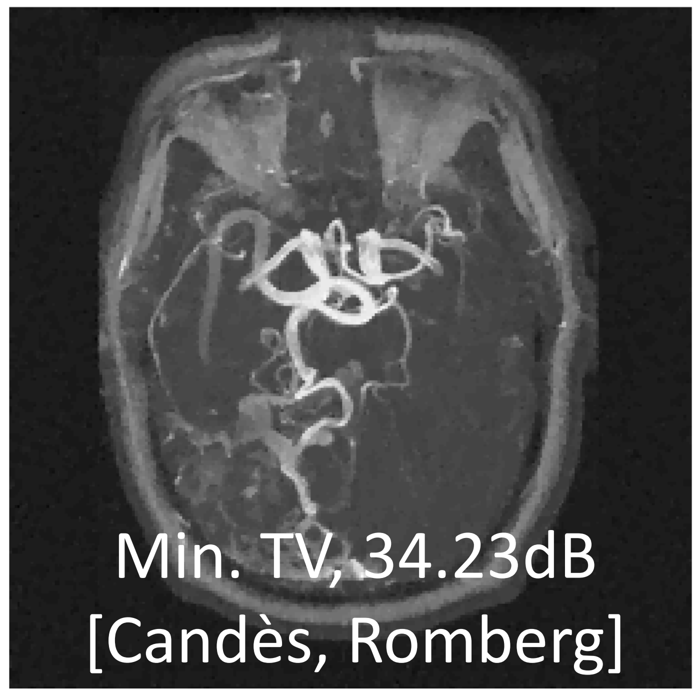
# Medical Imaging

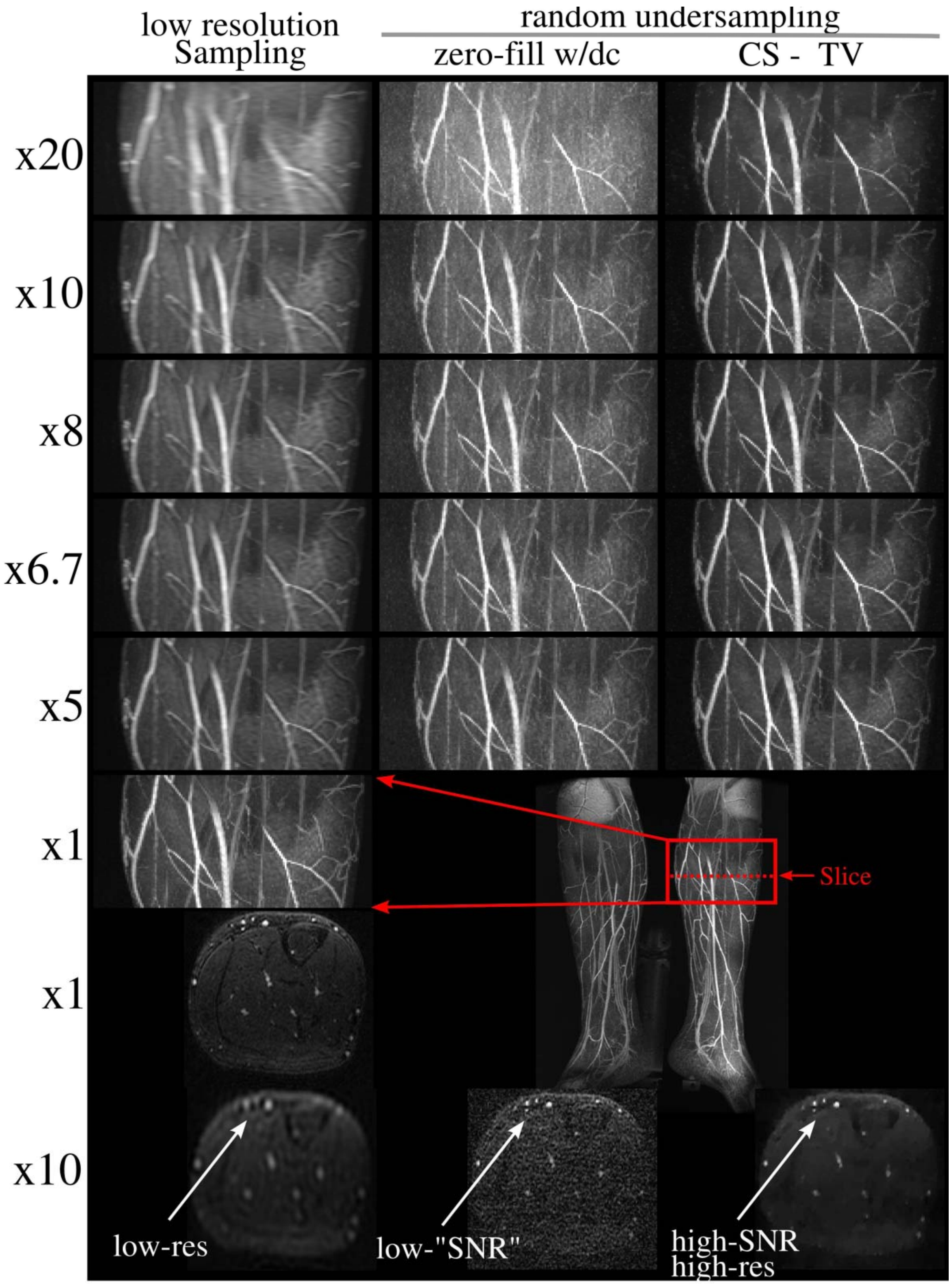





# Contrast-enhanced 3D angiography

[Lustig, Donoho, Pauly]

# 3. How can we fill in the blanks?

# New Ways of Processing Samples

- Recall the two ingredients:
  1. Model: “How is the signal supposed to behave?”
  2. Algorithm: “How can I fill in the blanks so that the signal obeys the model?”

- We suppose that our signal is sparse in some particular transform (e.g., Fourier or wavelets).

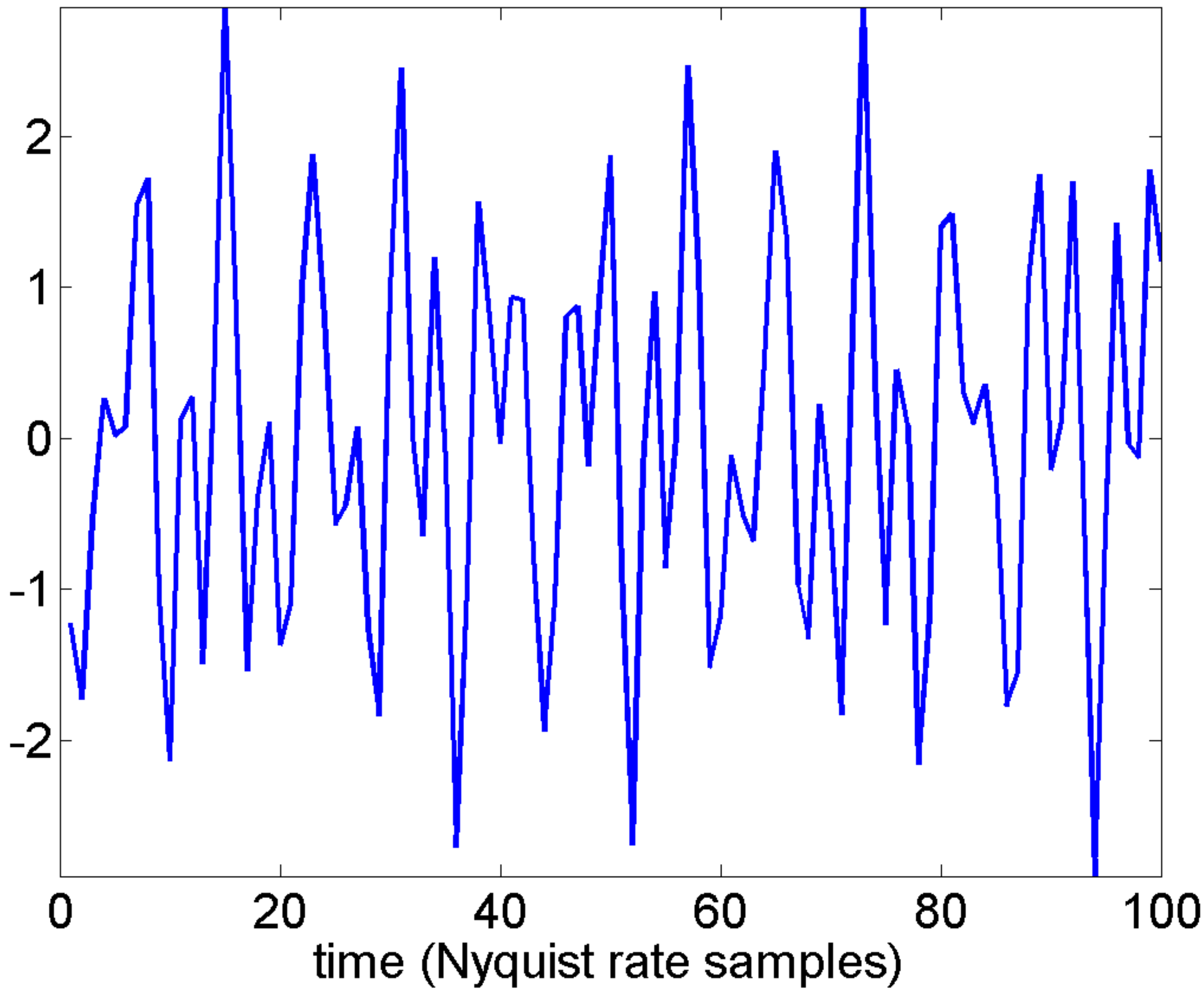
- We must fill in the blanks so that the signal obeys this sparse model.

# Don’t

- Don’t use ***smooth interpolation*** to fill in the blanks between samples.

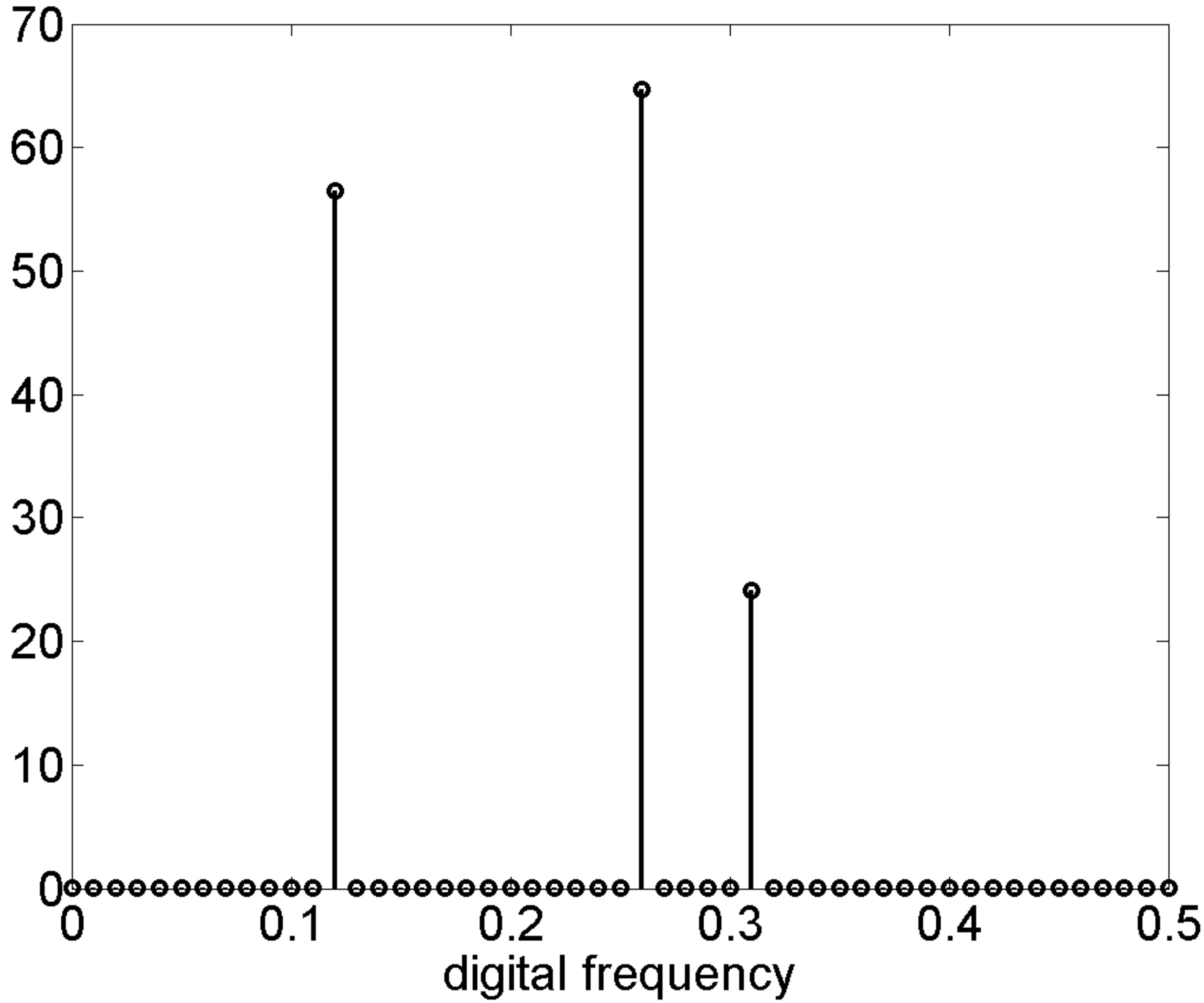
- This will fail to recover high-frequency features.

# Recall: Signal with Sparse Spectrum

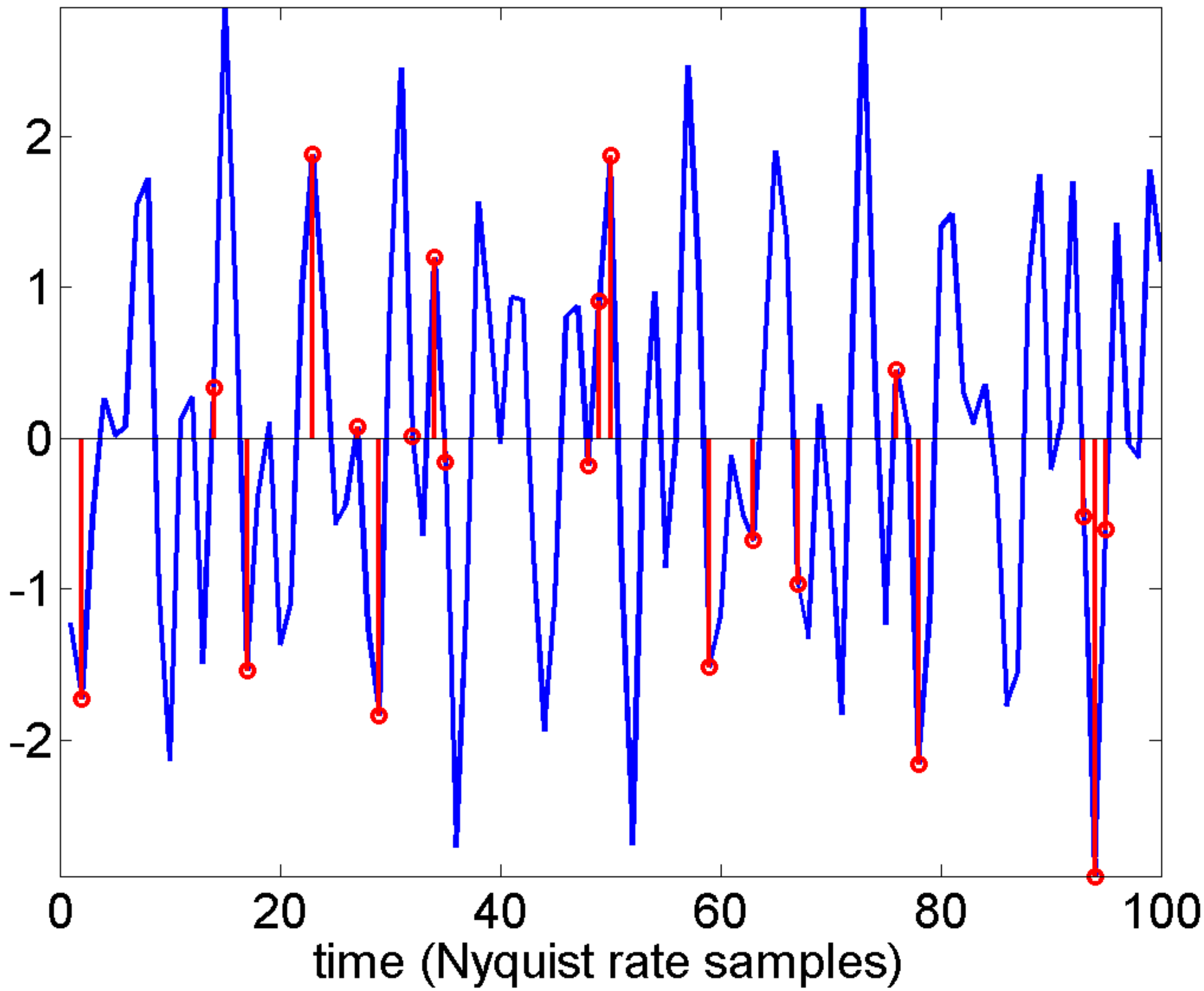



$N = 100$ Nyquist-rate samples

# Sparse Spectrum

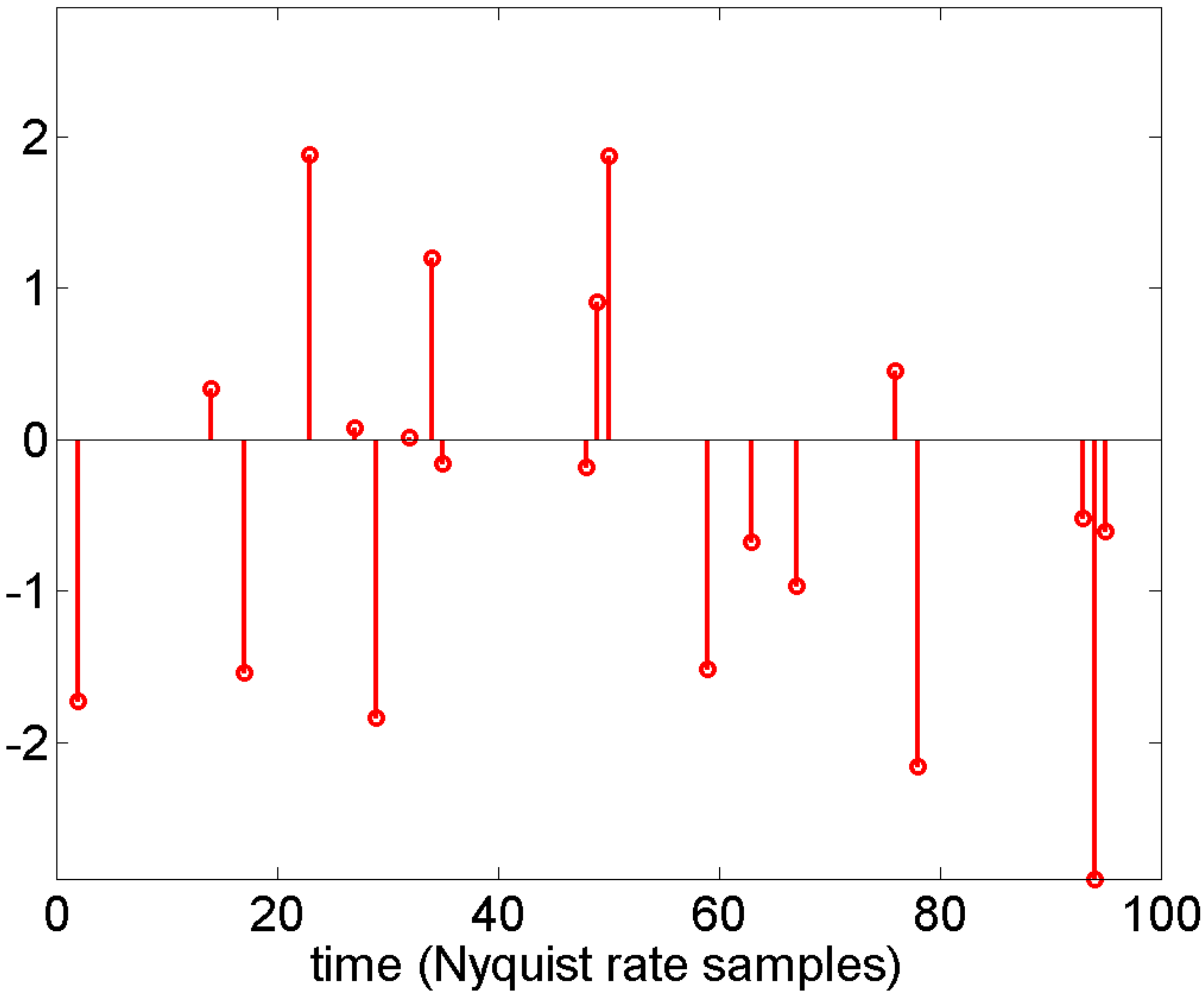


# Non-Uniform Samples (5x Sub-Nyquist)



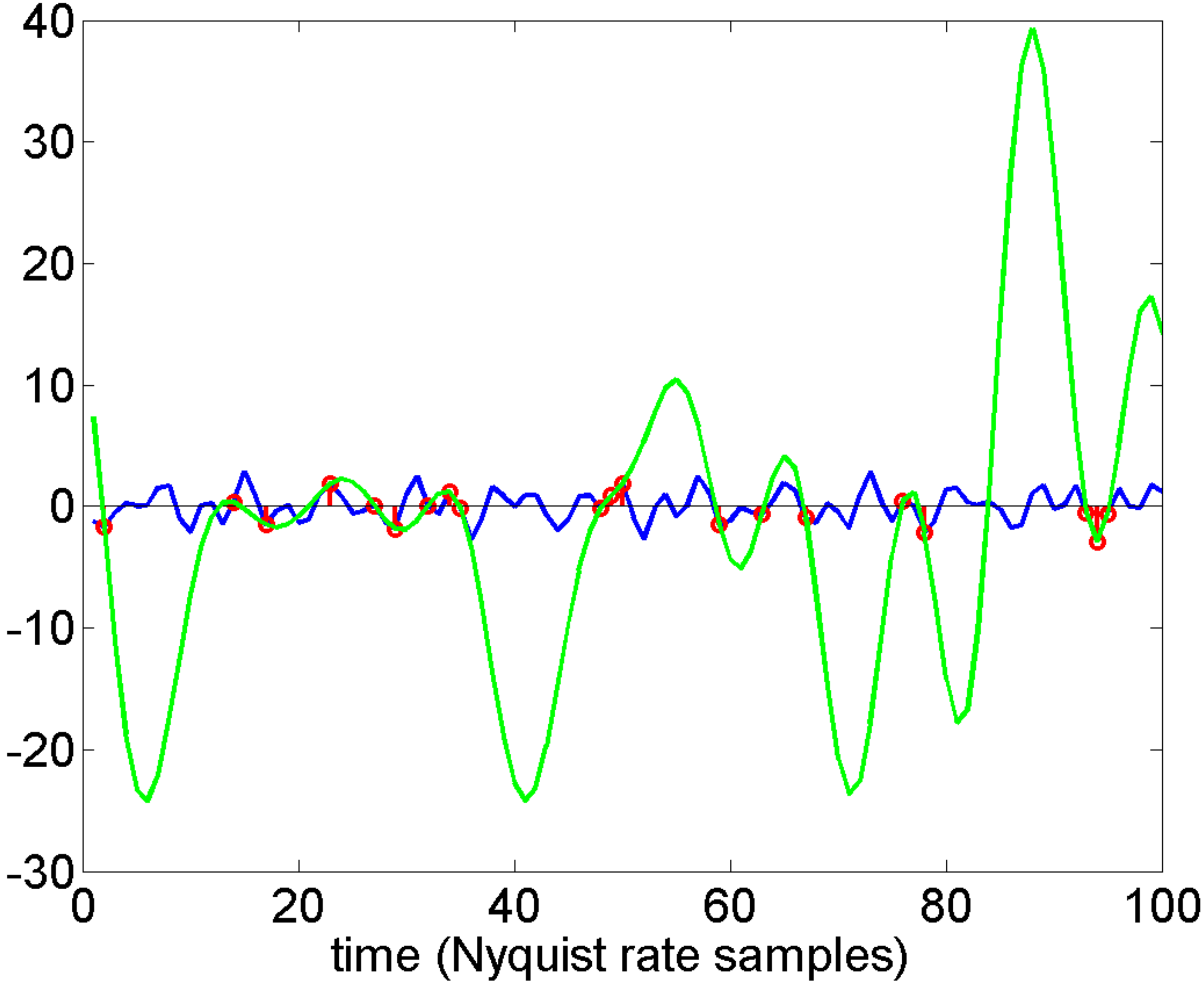

$M = 20$ non-uniform samples

# Non-Uniform Samples (5x Sub-Nyquist)

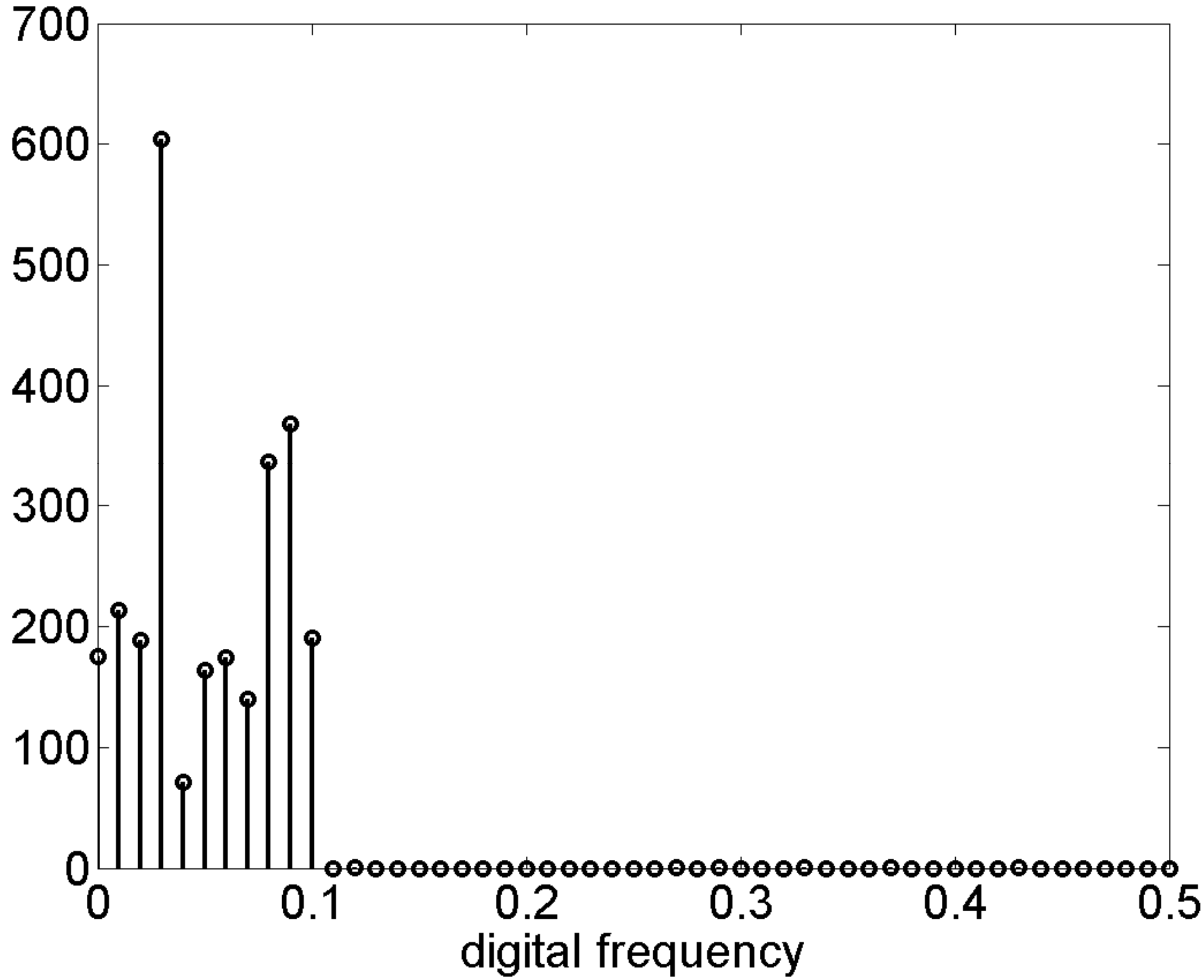



$M = 20$ non-uniform samples

# Smooth Interpolation



# Spectrum of Smooth Interpolation

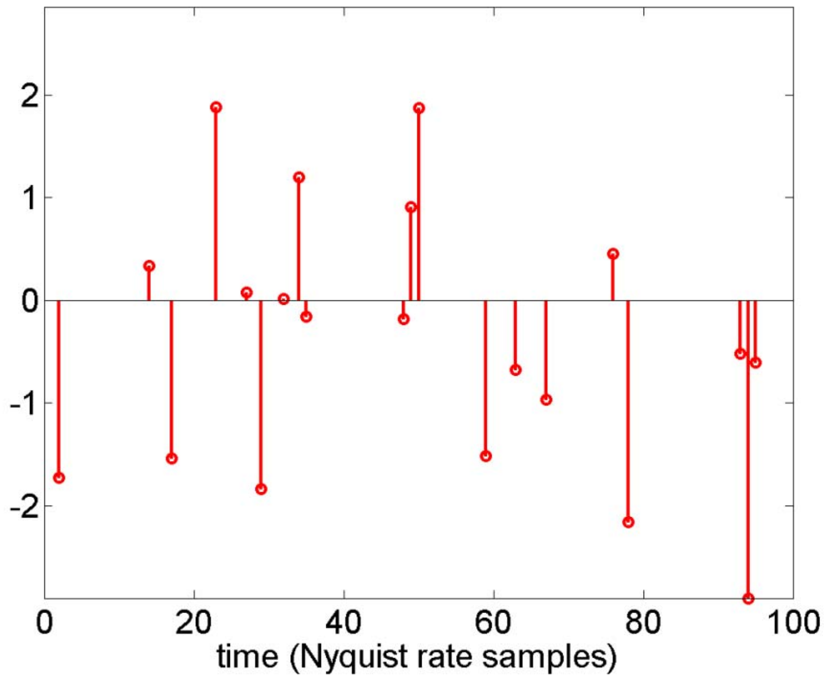


# Do

- Do ***search*** for a signal that

  (1) agrees with the measurements that have been collected, and

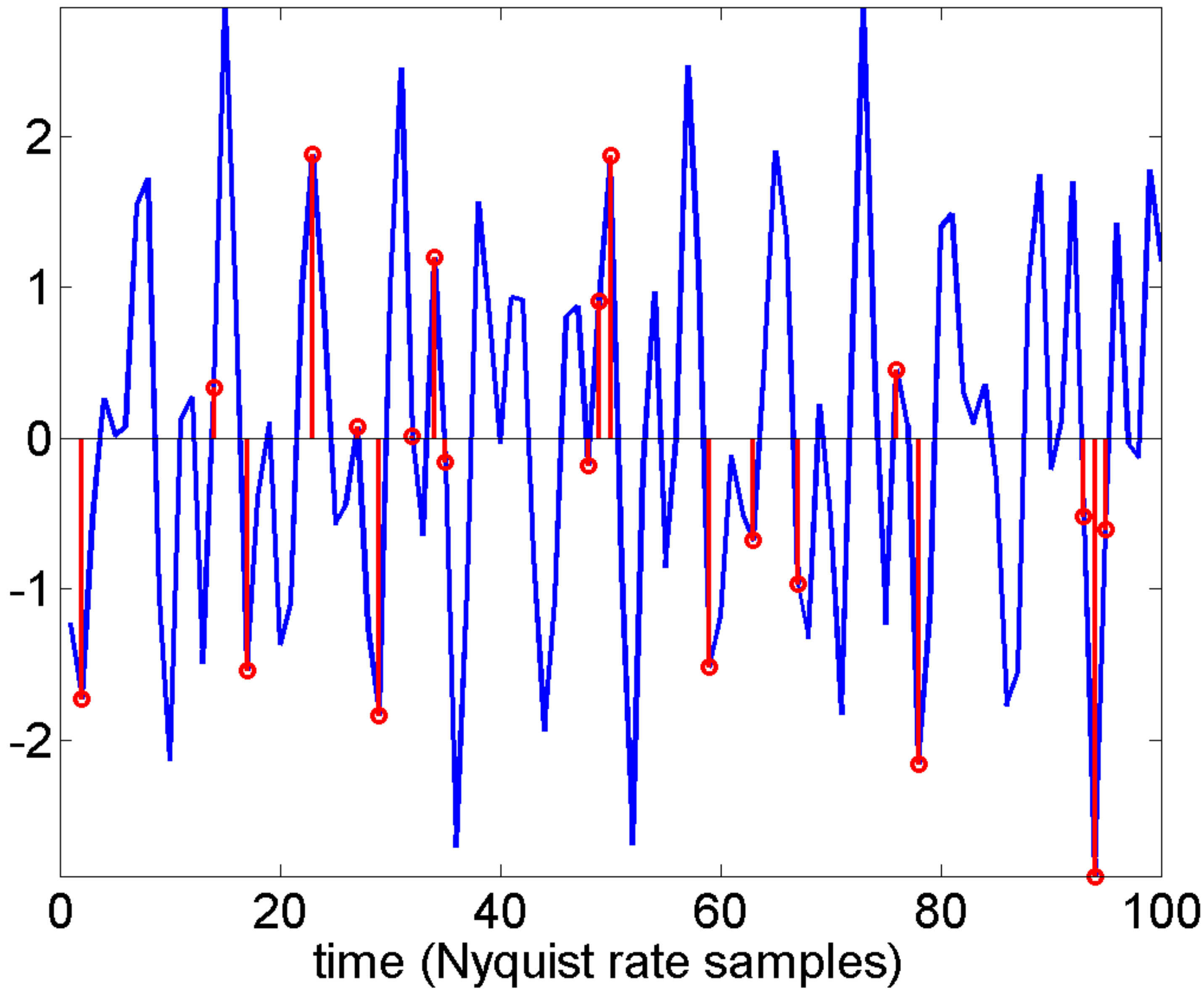
  (2) is as sparse as possible.




In this example:

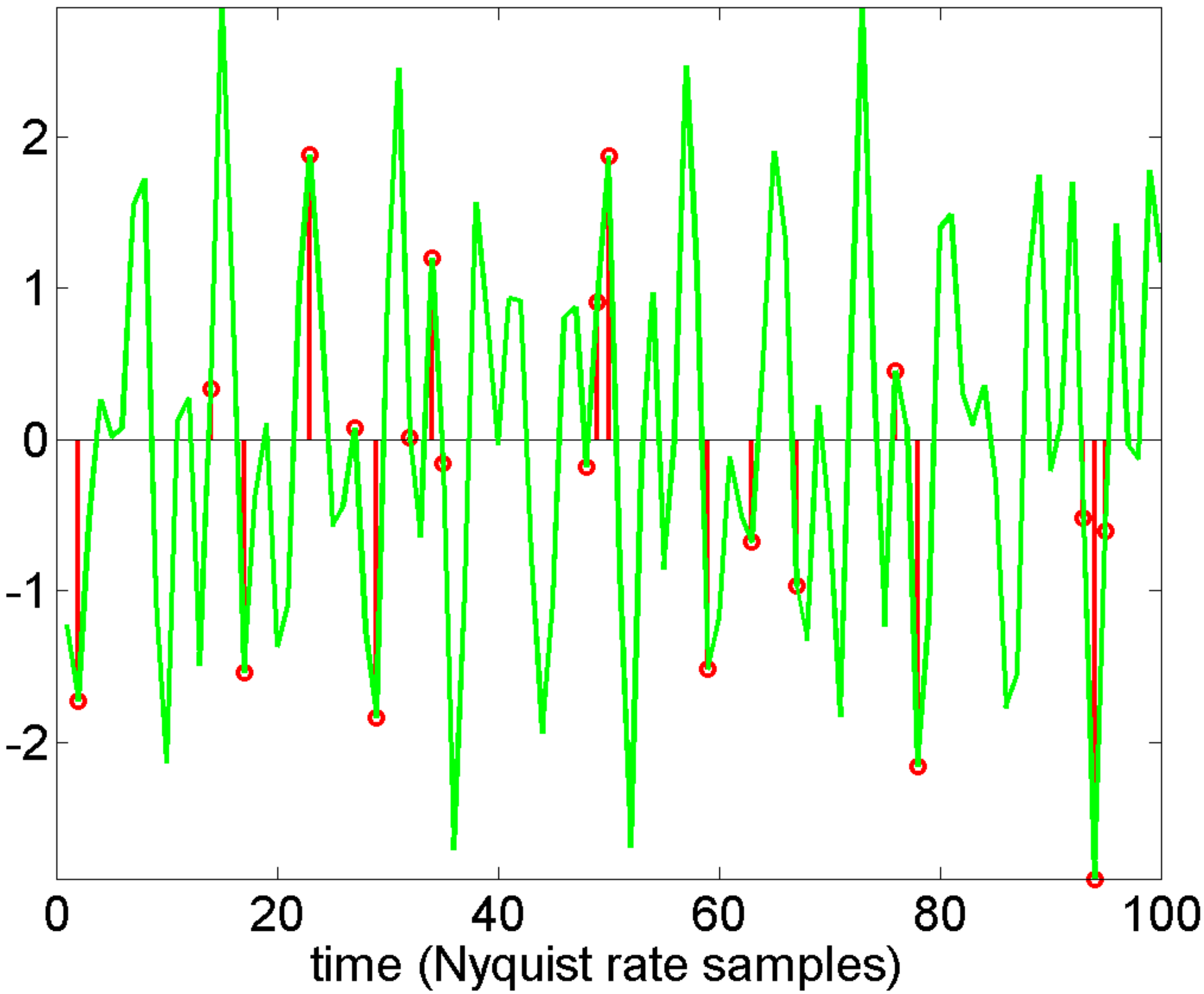
Find the signal that passes through the red samples and has the sparsest possible frequency spectrum.

# Non-Uniform Samples (5x Sub-Nyquist)

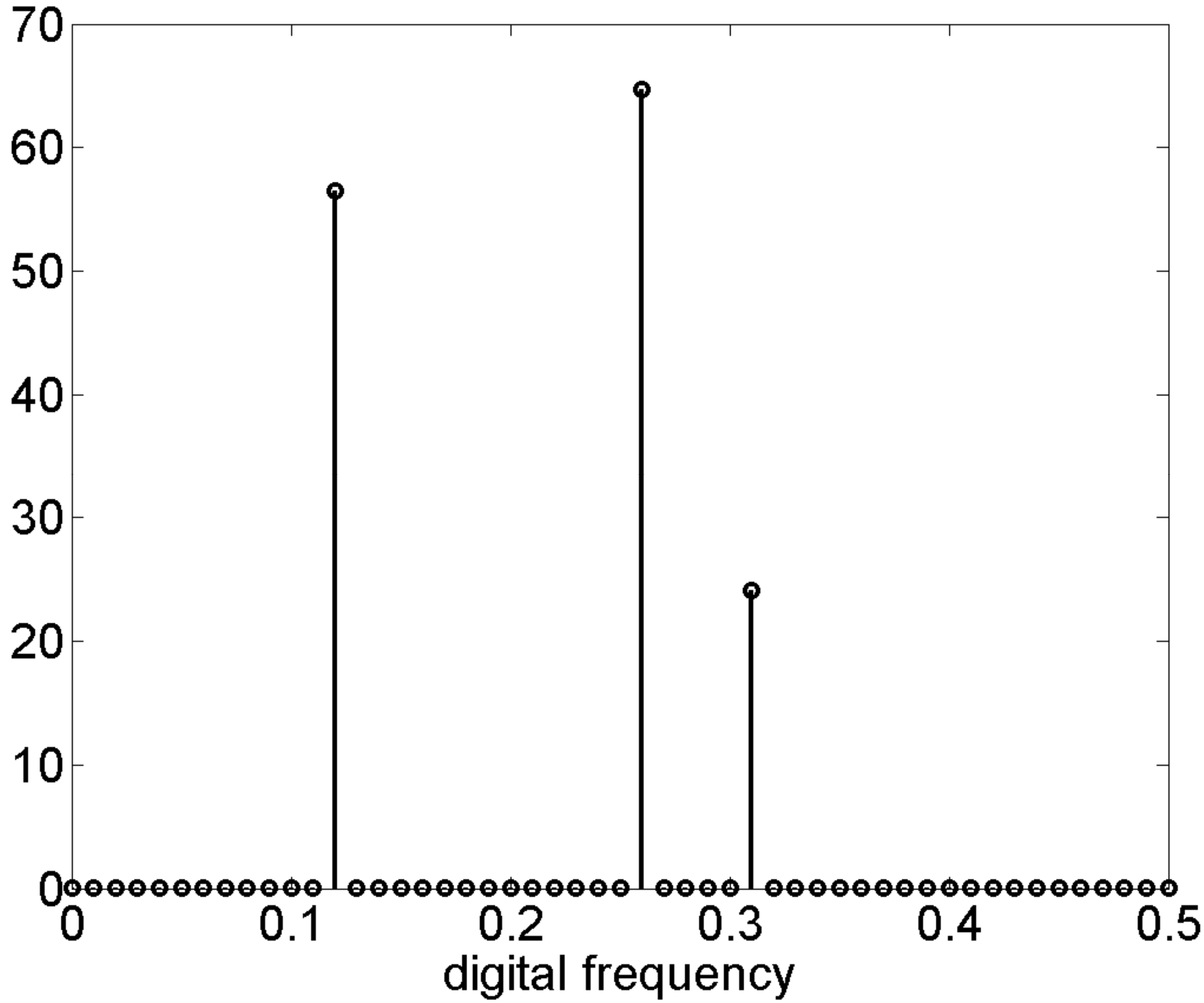



$M = 20$ non-uniform samples

# Perfect Recovery via Sparse Model



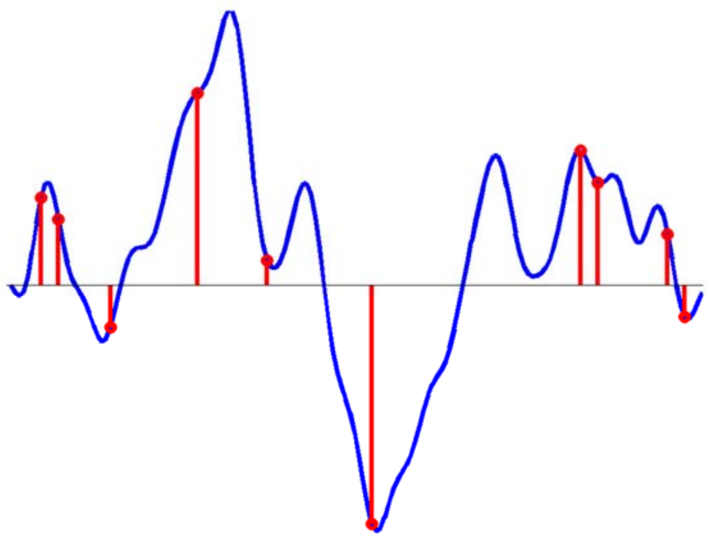
# Recovered Sparse Spectrum



# What About Aliasing?

- When sampling below Nyquist, aliasing is ***always*** a risk.

- There are infinitely many bandlimited signals that will interpolate a set of sub-Nyquist samples.



- However, there may only be ***one*** way to interpolate random samples that has a ***sparse*** frequency spectrum.

# Recovery from Generic Linear Measurements


$$
\begin{aligned}
y(1) &= 3.2 * (\text{pixel } 1) - 2.7 * (\text{pixel } 2) + \cdots \\
y(2) &= -8.1 * (\text{pixel } 1) + 0.3 * (\text{pixel } 2) + \cdots \\
&\vdots \\
y(M) &= 4.2 * (\text{pixel } 1) + 1.7 * (\text{pixel } 2) + \cdots
\end{aligned}
$$

generic linear measurements

Find the image that, if measured, would produce the same set of measurements

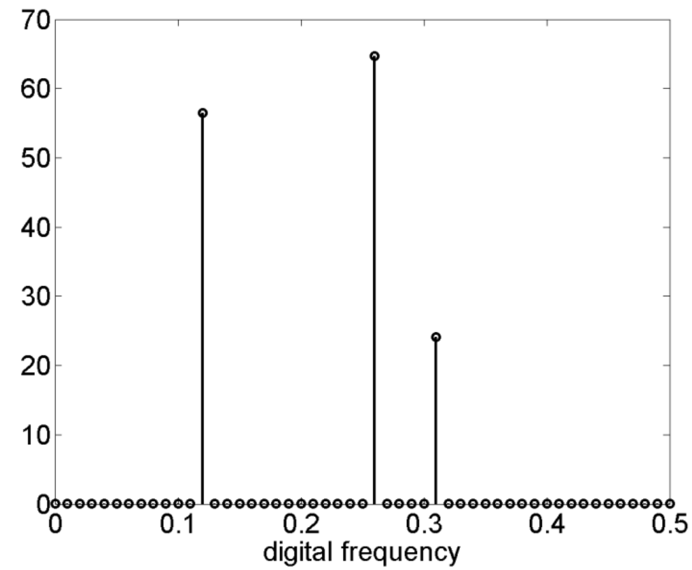
$$y(1), y(2), \ldots, y(M)$$

but has the smallest possible number of nonzero wavelet coefficients.

# Difficulty

- Finding the solution with the smallest possible number of nonzero wavelet coefficients is NP-hard in general.

- The difficulty is that we don't know in advance ***where*** to put the nonzero coefficients.

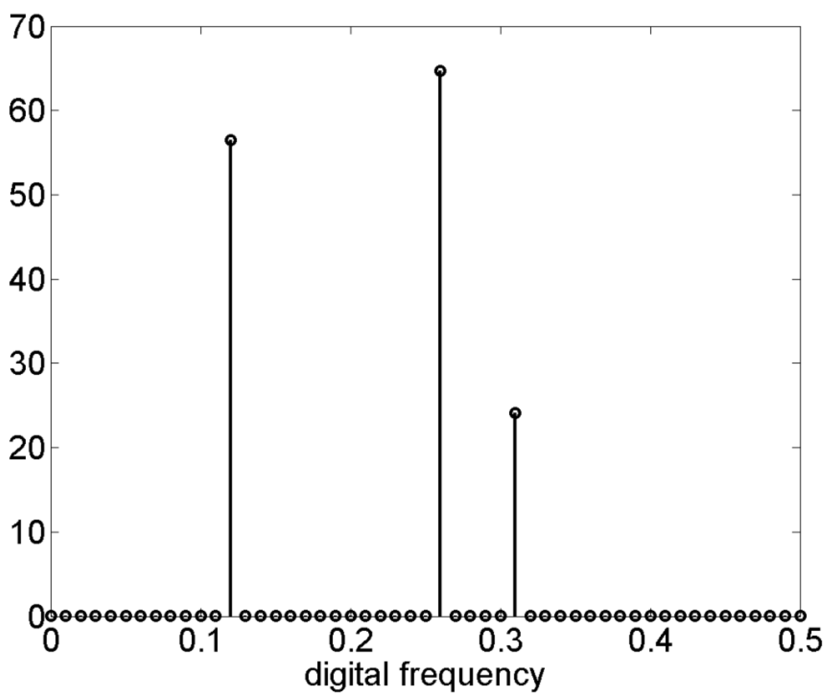



- Searching over all possible sparse coefficient patterns would be prohibitive.

# Enter the $L_1$ Norm

- Idea: use the ***$L_1$* norm** to measure the "sparsity" of the signal's coefficients

$$\|\alpha\|_1 = |\alpha(1)| + |\alpha(2)| + \cdots + |\alpha(N)|$$




In this example:

$L_1$ norm of spectrum
= 57 + 65 + 24
= 146

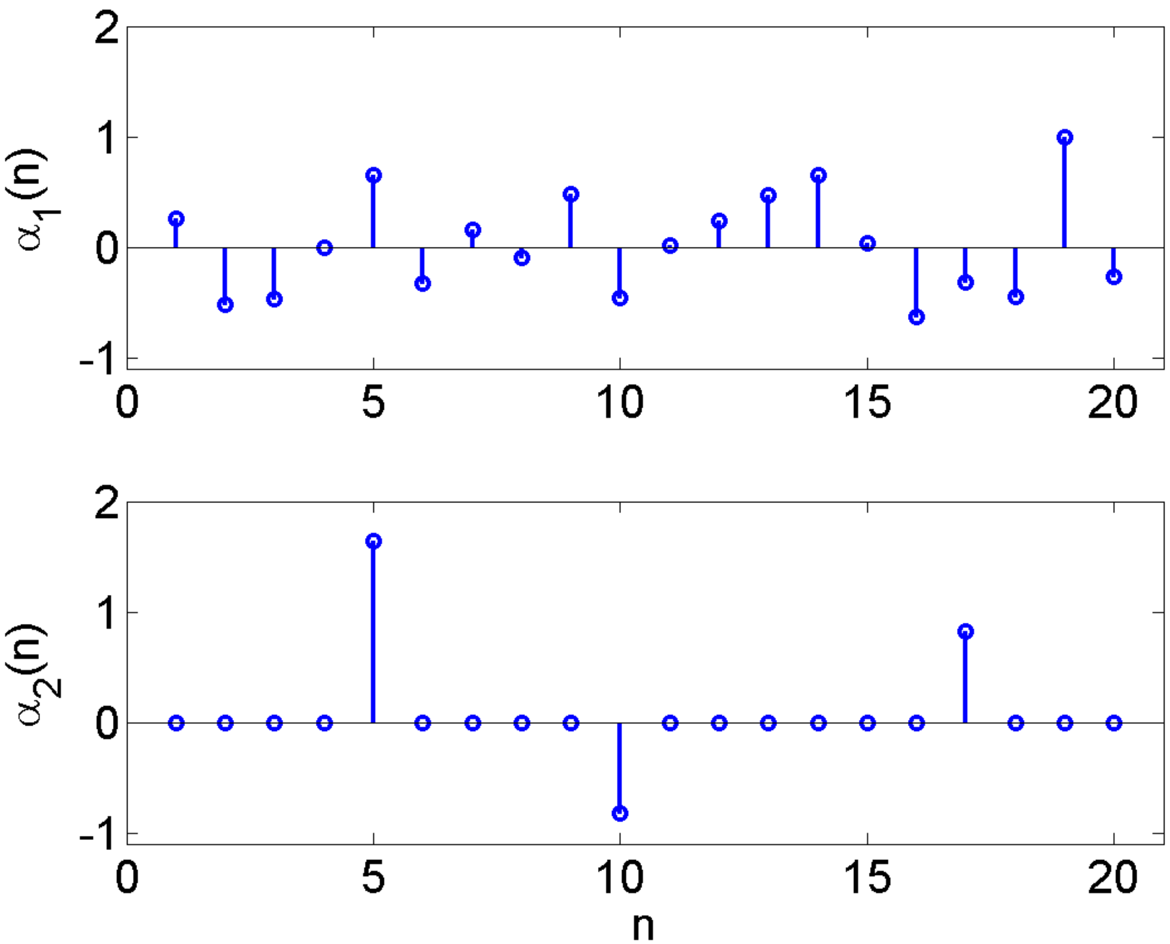
# A Rich History

- Sparse regularization in reflection seismology
  - Claerbout and Muir (1973)
  - Taylor, Banks, McCoy (1979)
  - Santosa and Symes (1986)
  - Donoho and Stark (1989)

- Other application areas
  - engineering, signal processing, control, imaging, portfolio optimization, convex optimization, …

# $L_1$ Norm Promotes Sparsity

two coefficient vectors: $\alpha_1$ and $\alpha_2$




same energy:

$$\|\alpha_1\|_2 = \sqrt{\sum_{n=1}^{N} |\alpha_1(n)|^2} = 2.012$$

$$\|\alpha_2\|_2 = \sqrt{\sum_{n=1}^{N} |\alpha_2(n)|^2} = 2.012$$

different $L_1$ norms:

$$\|\alpha_1\|_1 = \sum_{n=1}^{N} |\alpha_1(n)| = 7.496$$

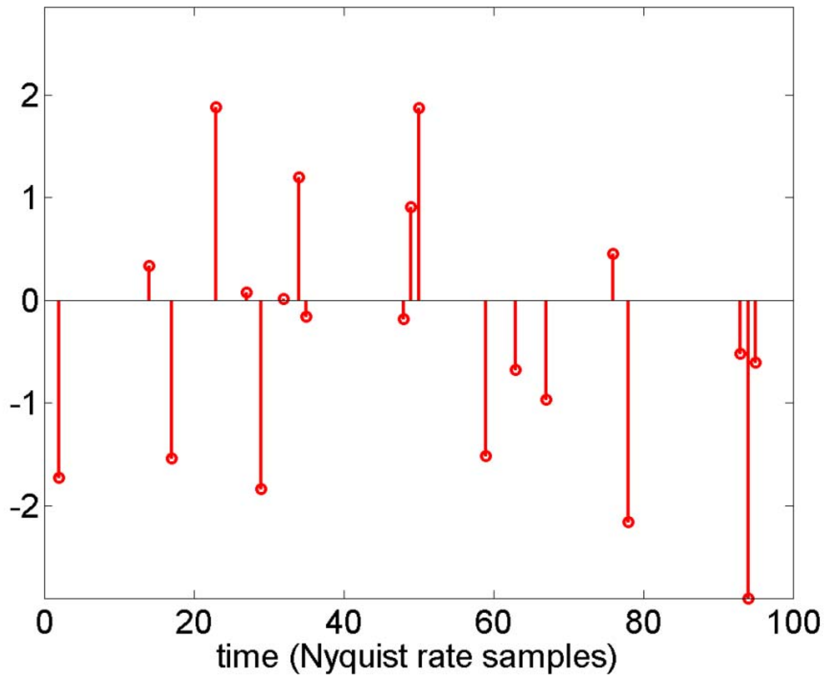
$$\|\alpha_2\|_1 = \sum_{n=1}^{N} |\alpha_2(n)| = 3.285$$

# $L_1$ Minimization for Sparse Recovery

- Do ***search*** for a signal that

  (1) agrees with the measurements that have been collected, and

  (2) is as sparse as possible. **(using the $L_1$ norm as a proxy for sparsity)**




In this example:

Find the signal that passes through the red samples and has the ~~sparsest possible~~ frequency spectrum. **with the smallest possible $L_1$ norm**

# Recovery from Generic Linear Measurements

$$
\begin{aligned}
y(1) &\quad=\quad 3.2 * (\text{pixel } 1) - 2.7 * (\text{pixel } 2) + \cdots \\
y(2) &\quad=\quad -8.1 * (\text{pixel } 1) + 0.3 * (\text{pixel } 2) + \cdots \\
&\quad\vdots \\
y(M) &\quad=\quad 4.2 * (\text{pixel } 1) + 1.7 * (\text{pixel } 2) + \cdots
\end{aligned}
$$

generic linear measurements

Find the image that, if measured, would produce the same set of measurements

$$y(1), y(2), \ldots, y(M)$$

but has the smallest possible ~~number of nonzero~~ **$L_1$ norm of its** wavelet coefficients.

# $L_1$ Minimization Algorithms

- $L_1$ minimization is a ***convex*** optimization problem

- Many general purpose solvers available
  - CVX, SPGL1, NESTA, TFOCS, FISTA, YALL1, GPSR, FPC
  - generally fast when the sparsifying transform is fast
    - FFT, wavelets, curvelets
  - complexity roughly “a few hundred” applications of the forward/adjoint sparsifying transform and measurement operator

- Can be extended to account for measurement noise

# Iterative Hard Thresholding

[Blumensath and Davies]

scaling factor

adjoint operators

measurement operator

sparse transform

$$\widehat{\alpha}^{n+1} = \mathcal{H}_K\left(\widehat{\alpha}^n + \gamma\Psi^*\Phi^*(y - \Phi\Psi\widehat{\alpha}^n)\right)$$

hard thresholding
(keep $K$ largest coefficients)

measurements

previous coefficient estimate

# 4. How can we understand this process?

# The Miracle of CS

- Recall:

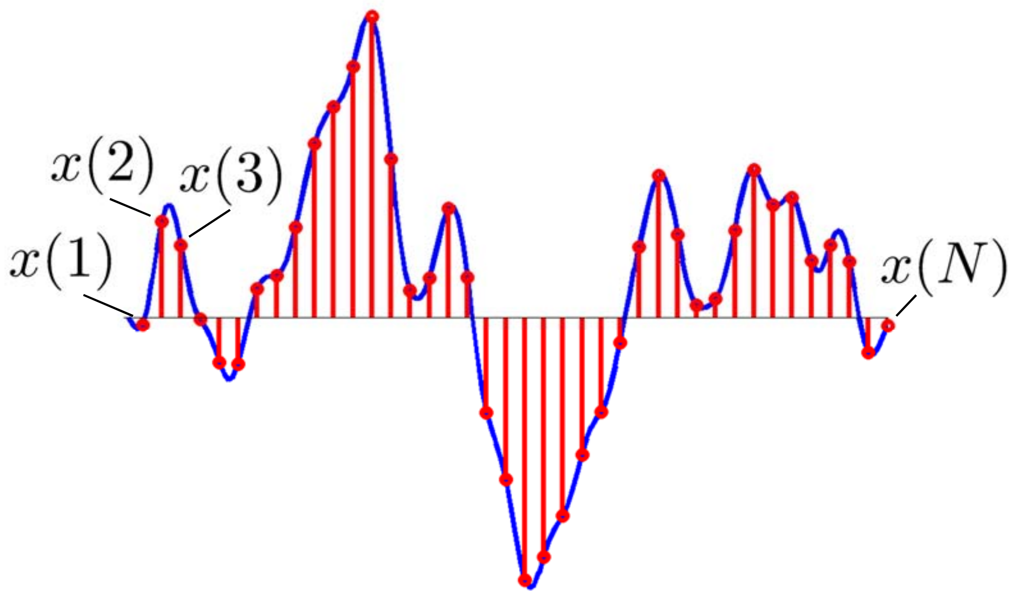
  With incoherent measurements, for a signal of sparsity level $K$, instead of collecting all $N$ Nyquist-rate samples, it suffices to take roughly $K \log N$ measurements.

- Why is this possible?
- What are ***incoherent*** measurements?

# Recall: Generic Linear Measurements

Nyquist-rate samples

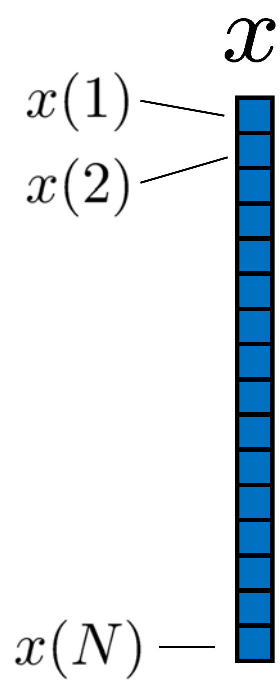
(don't record these directly)




Generic linear measurements

$$
\begin{aligned}
y(1) &= 3.2 * x(1) - 2.7 * x(2) + \cdots + 8.1 * x(N) \\
y(2) &= -8.1 * x(1) + 0.3 * x(2) + \cdots - 0.4 * x(N) \\
&\vdots \\
y(M) &= 4.2 * x(1) + 1.7 * x(2) + \cdots + 0.7 * x(N)
\end{aligned}
$$

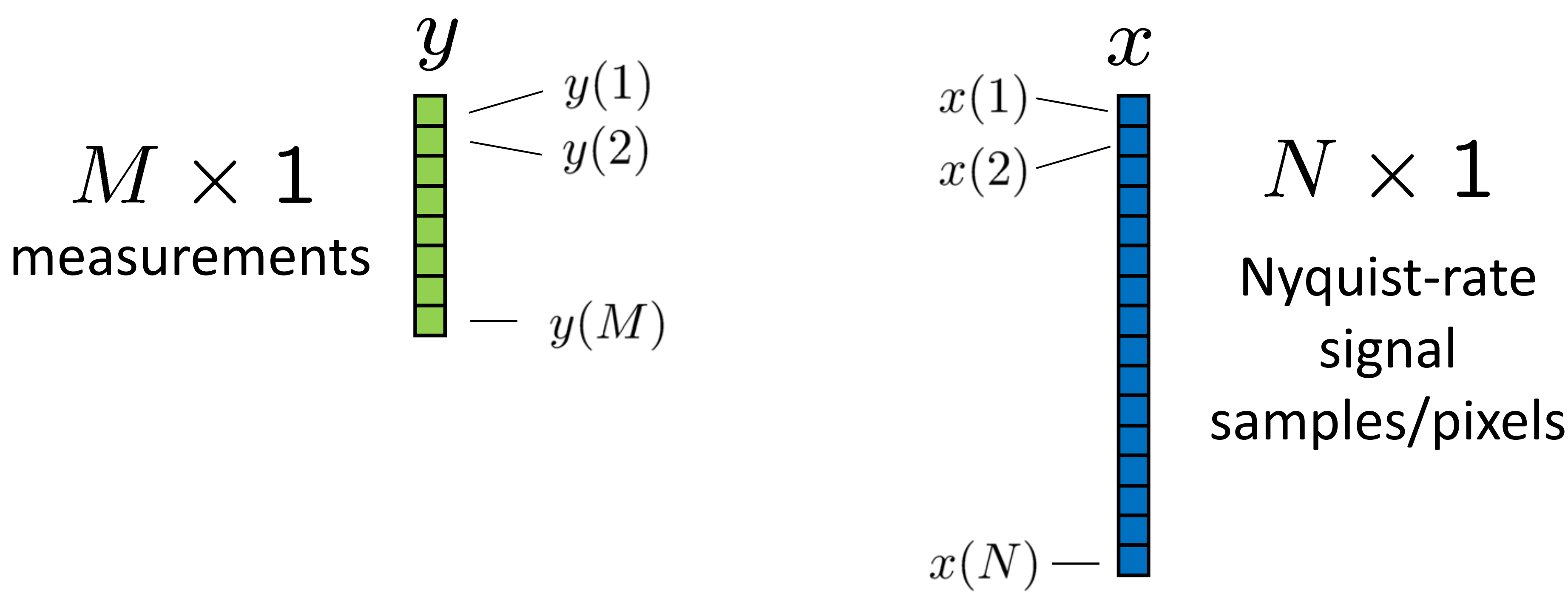
# Introducing Matrix-Vector Notation



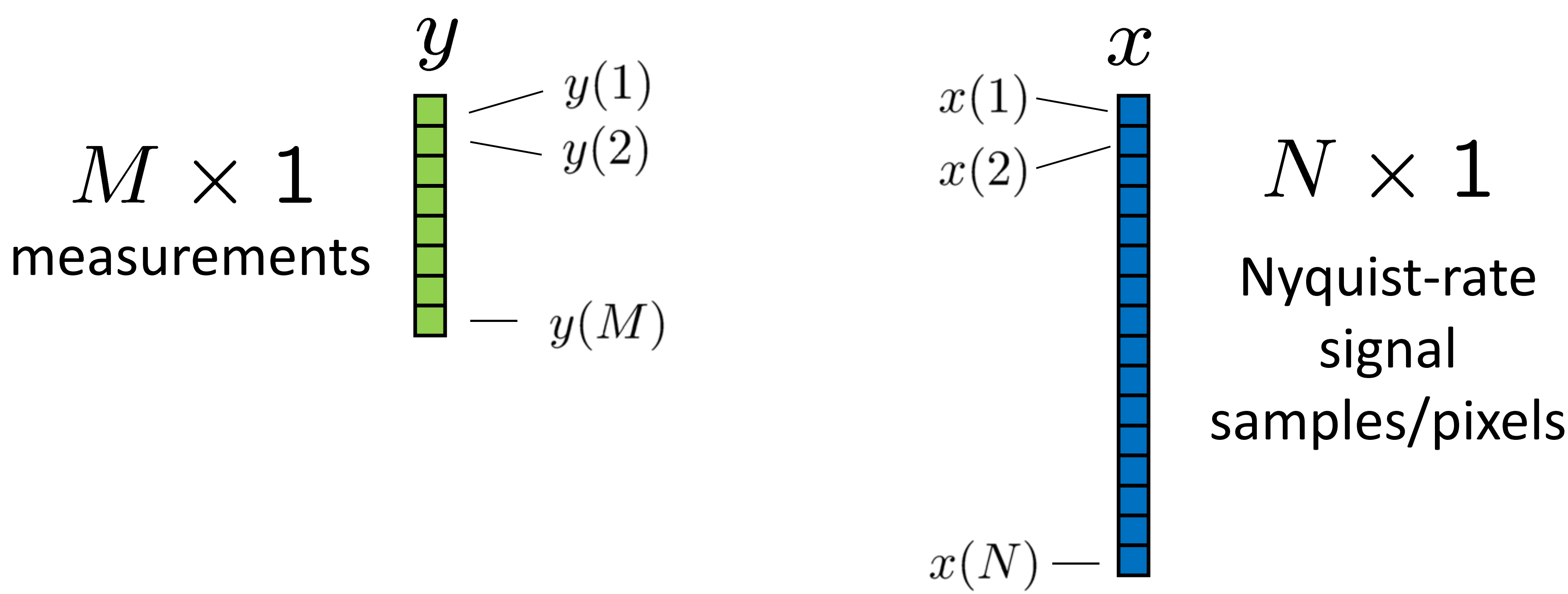

$N \times 1$

Nyquist-rate
signal
samples/pixels

# Introducing Matrix-Vector Notation

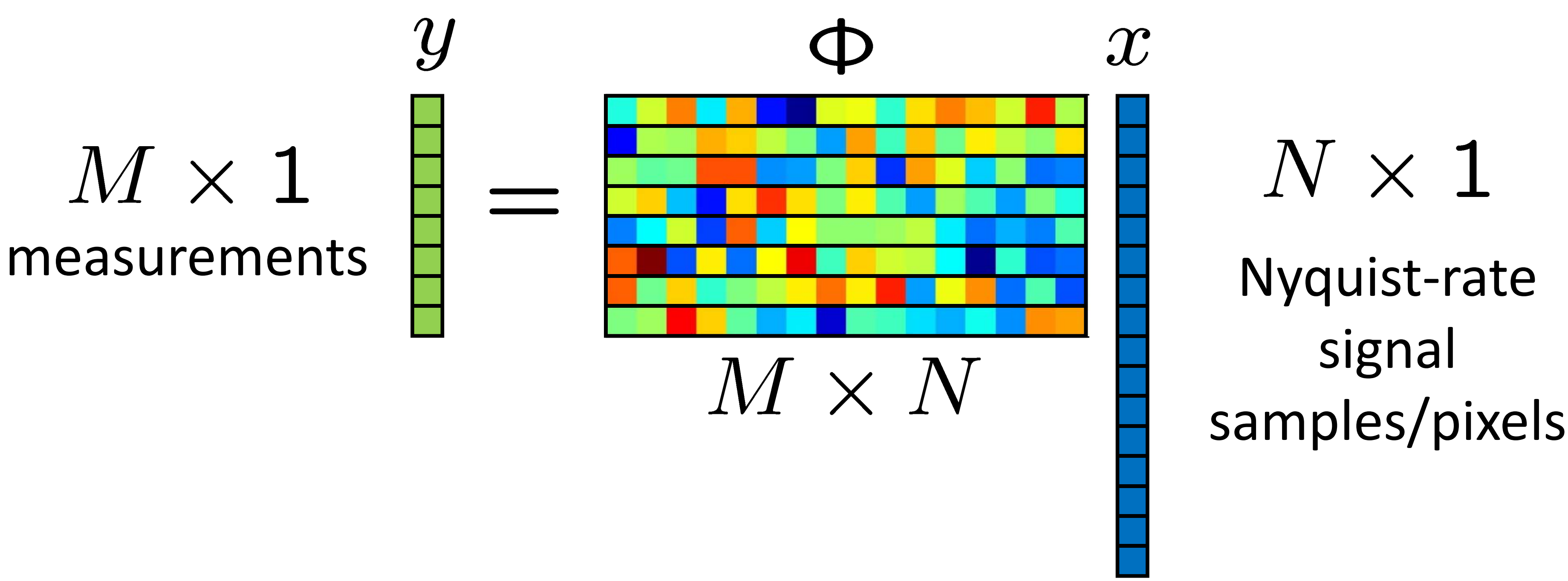


# Introducing Matrix-Vector Notation




$$
\begin{aligned}
y(1) &= 3.2 * x(1) - 2.7 * x(2) + \cdots + 8.1 * x(N) \\
y(2) &= -8.1 * x(1) + 0.3 * x(2) + \cdots - 0.4 * x(N) \\
&\vdots \\
y(M) &= 4.2 * x(1) + 1.7 * x(2) + \cdots + 0.7 * x(N)
\end{aligned}
$$

# Introducing Matrix-Vector Notation

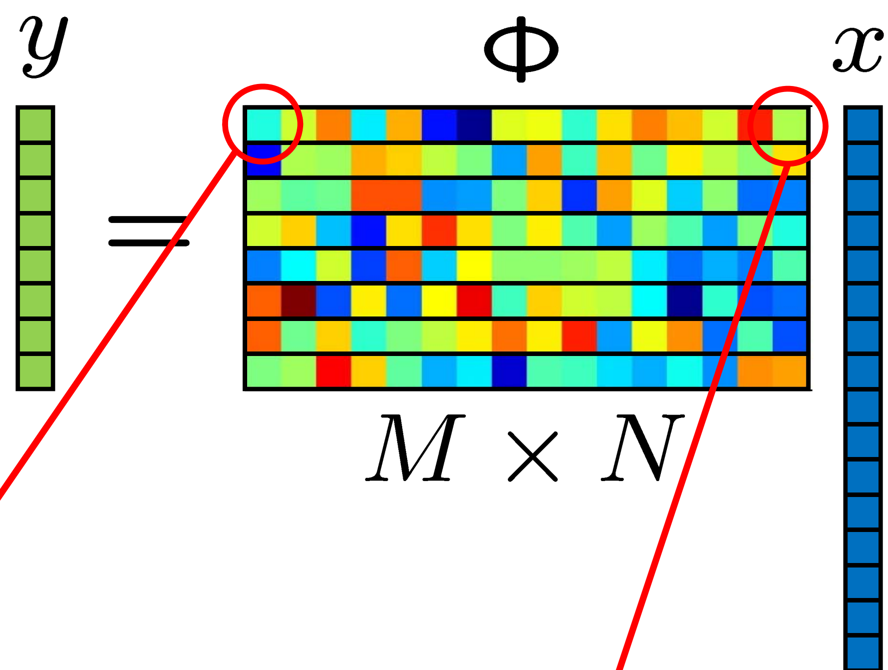



$$
\begin{aligned}
y(1) &= 3.2 * x(1) - 2.7 * x(2) + \cdots + 8.1 * x(N) \\
y(2) &= -8.1 * x(1) + 0.3 * x(2) + \cdots - 0.4 * x(N) \\
&\vdots \\
y(M) &= 4.2 * x(1) + 1.7 * x(2) + \cdots + 0.7 * x(N)
\end{aligned}
$$

# Introducing Matrix-Vector Notation

$M \times 1$

measurements




$N \times 1$

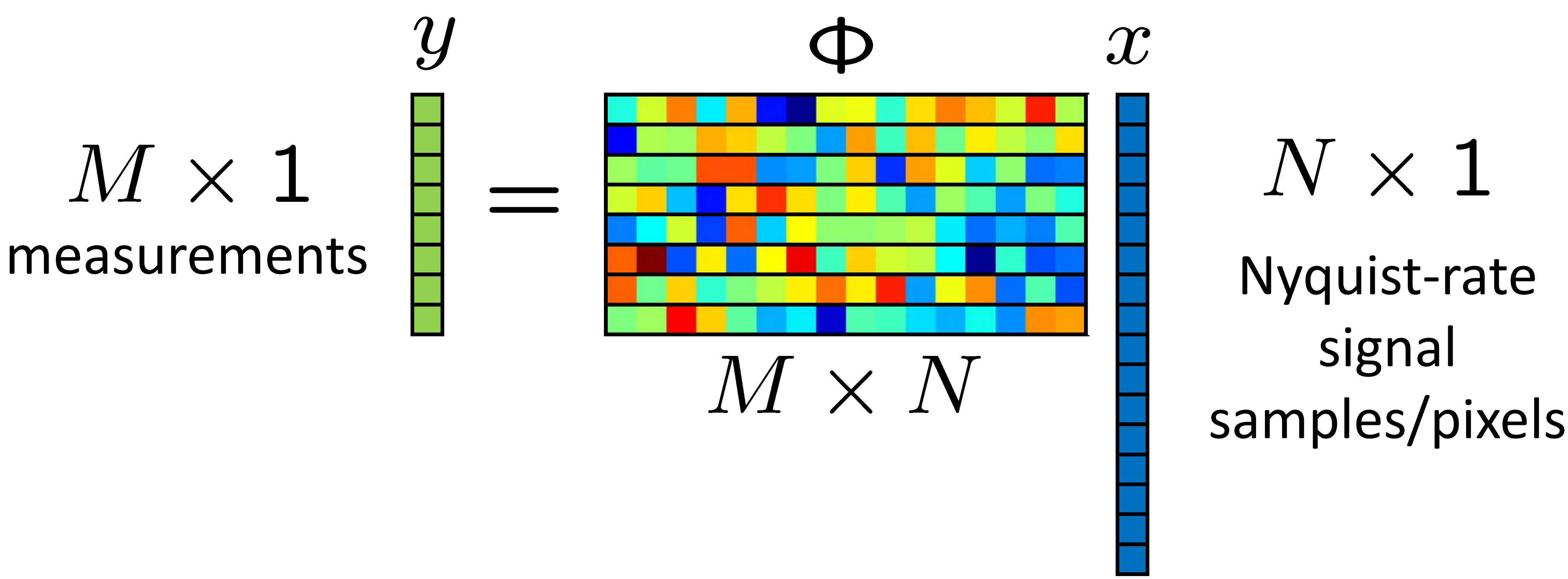
Nyquist-rate signal samples/pixels

$$
\begin{aligned}
y(1) &= 3.2 * x(1) - 2.7 * x(2) + \cdots + 8.1 * x(N) \\
y(2) &= -8.1 * x(1) + 0.3 * x(2) + \cdots - 0.4 * x(N) \\
&\vdots \\
y(M) &= 4.2 * x(1) + 1.7 * x(2) + \cdots + 0.7 * x(N)
\end{aligned}
$$

# CS Recovery: An Inverse Problem

- Given $y$, have infinitely many candidates for $x$
  - Since $M < N$, underdetermined set of linear equations $y = \Phi x$

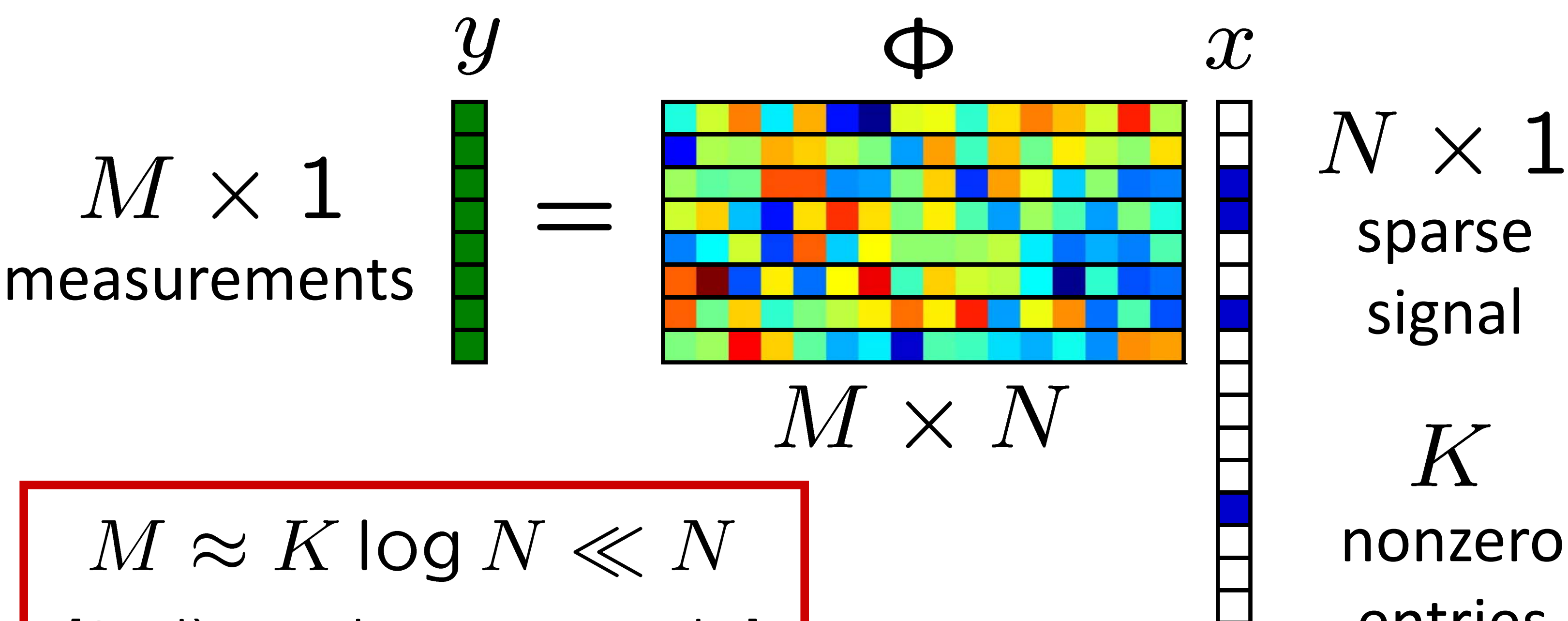
- Search among these for the one that most closely agrees with our model for how $x$ should behave



# Recovery of ***Sparse*** Signals (2004)

- If $x$ is $K$-sparse, need $M \approx K \log N$ random measurements to guarantee (with high probability) stable and robust recovery

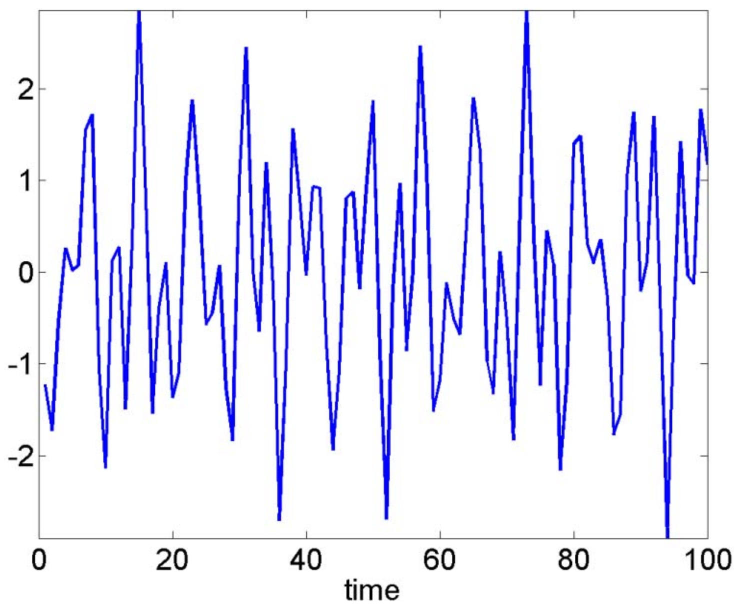
- Measurements should be incoherent with sparse dictionary



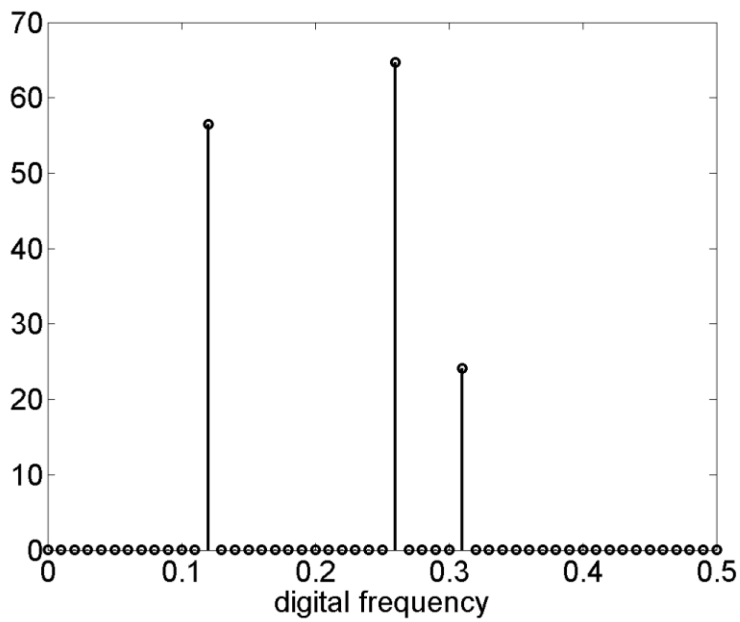

$$M \approx K \log N \ll N$$

[Candès, Romberg, Tao; Donoho]

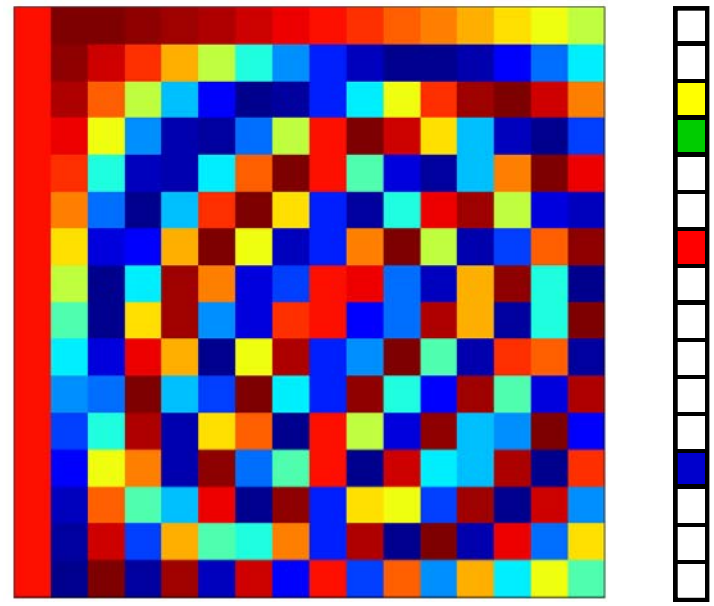
# Sparsifying Transform




$x$

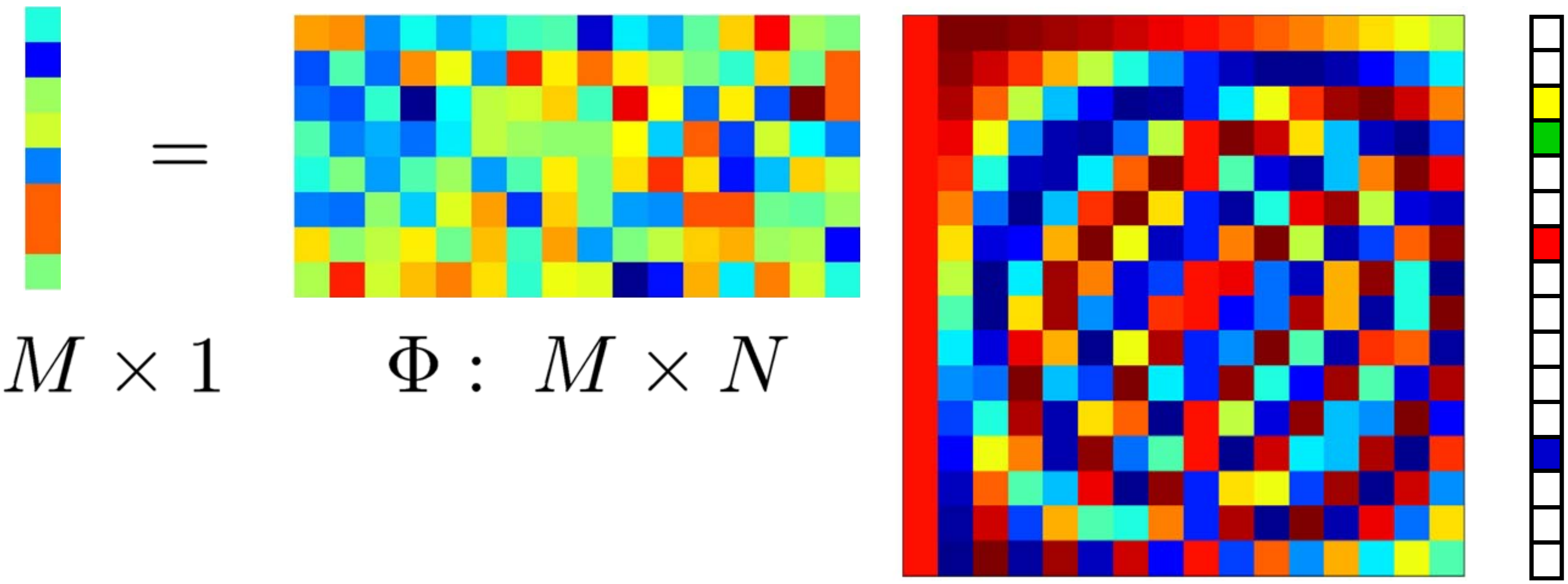



$\alpha$



$$\underbrace{\Psi : \; N \times N \qquad \alpha}_{x \,:\; N \times 1}$$

# Sparse Signal Recovery



$y : M \times 1 \qquad = \qquad \Phi : M \times N \qquad \underbrace{\Psi : N \times N \qquad \alpha}_{x \,:\, N \times 1}$

$$\widehat{\alpha} = \arg\min_{\alpha'} \|\alpha'\|_1$$
$$\text{subject to } y = \Phi\Psi\alpha'$$

then set $\widehat{x} = \Psi\widehat{\alpha}$

# $L_1$ Minimization for Sparse Recovery

$$\widehat{\alpha} = \arg\min_{\alpha'} \|\alpha'\|_1 \text{ subject to } y = \Phi\Psi\alpha'$$

$$\sum_{i=1}^{N} |\alpha'(i)|$$

$$\text{then set } \widehat{x} = \Psi\widehat{\alpha}$$

- Convex optimization problem known as ***Basis Pursuit***
  - many general purpose solvers available
- Returns a good approximation to $x$ if original signal is not perfectly sparse

# Robust $L_1$ Minimization

$$\widehat{\alpha} = \arg\min_{\alpha'} \|\alpha'\|_1$$
$$\text{subject to } \|y - \Phi\Psi\alpha'\|_2 \leq \epsilon$$

**Basis Pursuit De-Noising (BPDN)**

$$\widehat{\alpha} = \arg\min_{\alpha'} \frac{1}{2}\|y - \Phi\Psi\alpha'\|_2^2 + \lambda\|\alpha'\|_1$$

**LASSO**

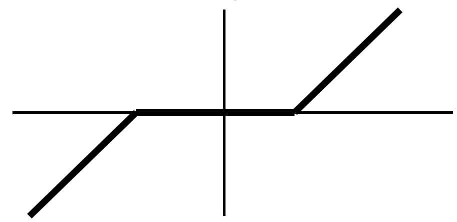
- Both are convex optimization problems
- Robust to measurement noise

# FISTA: Fast Iterative Shrinkage-Thresholding Algorithm

(2009) [Beck and Teboulle]

$$\alpha^k = \mathcal{S}_{\lambda\tau}(z^k + \tau\Psi^*\Phi^*(y - \Phi\Psi z^k))$$

soft-thresholding

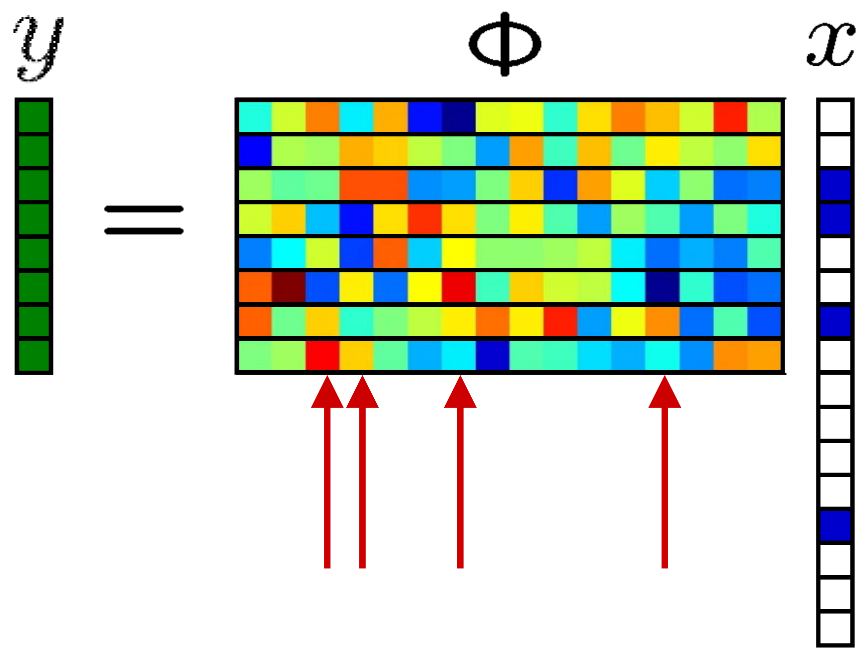



$$\gamma^{k+1} = (1 + \sqrt{1 + 4(\gamma^k)^2})/2 \qquad z^{k+1} = \alpha^k + \tfrac{\gamma^k - 1}{\gamma^{k+1}}(\alpha^k - \alpha^{k-1})$$

- Gradient descent with sparse regularization
- Objective function converges as $1/k^2$
- See also: SpaRSA [Wright et al.], FASTA [Goldstein et al.]

# Greedy Algorithms

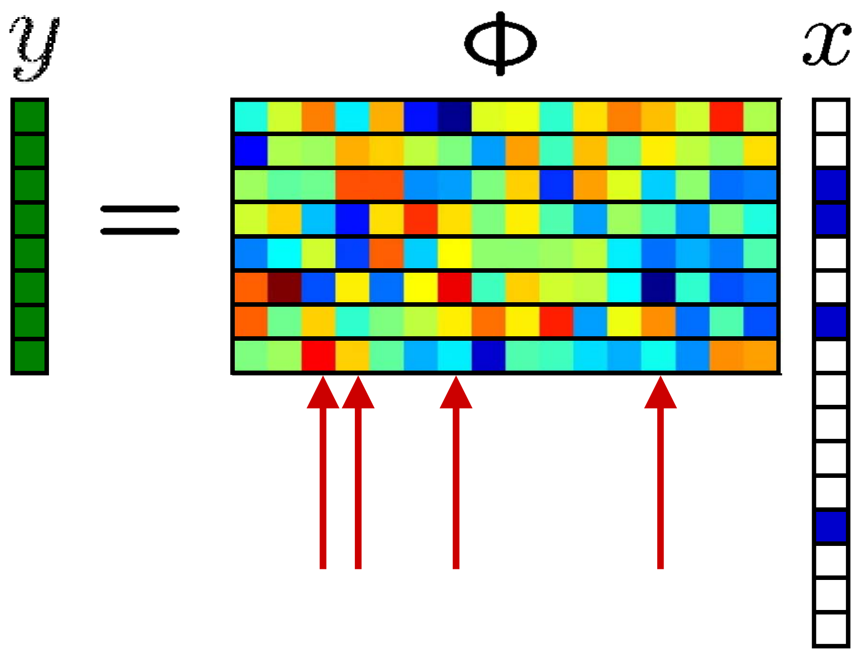
- Suppose that $x$ contains $K$ nonzero entries
  - sparsity transform $\Psi$= identity matrix




- Idea: find columns of $\Phi$ most correlated with $y$

# OMP: Orthogonal Matching Pursuit (1994)

[Davis, Mallat, and Zhang; Tropp and Gilbert]




1. Choose column most correlated with $y$.
2. Orthogonalize $y$ with respect to previously chosen columns.
3. Repeat until residual sufficiently small.
4. Recover $x$ using least-squares on chosen columns.

- See also: CoSaMP (2009) [Needell and Tropp]

# Recall: Iterative Hard Thresholding

[Blumensath and Davies]

scaling factor

adjoint operators

measurement operator

sparse transform

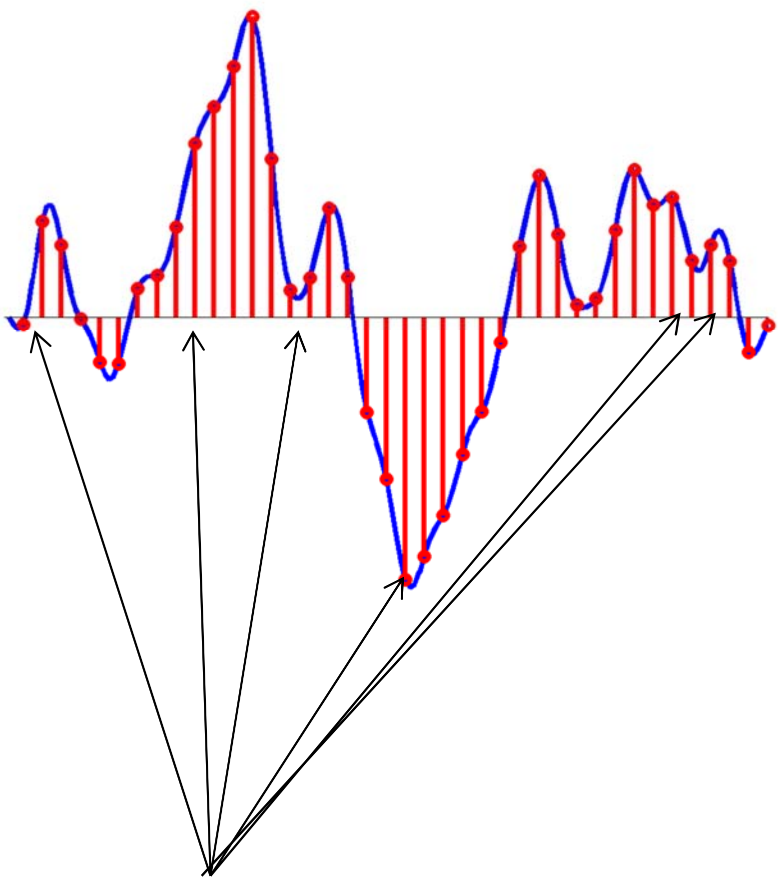
$$\widehat{\alpha}^{n+1} = \mathcal{H}_K\left(\widehat{\alpha}^n + \gamma\Psi^*\Phi^*(y - \Phi\Psi\widehat{\alpha}^n)\right)$$

hard thresholding
(keep $K$ largest coefficients)

measurements

previous coefficient estimate

# Case Study: Non-uniform Sampling

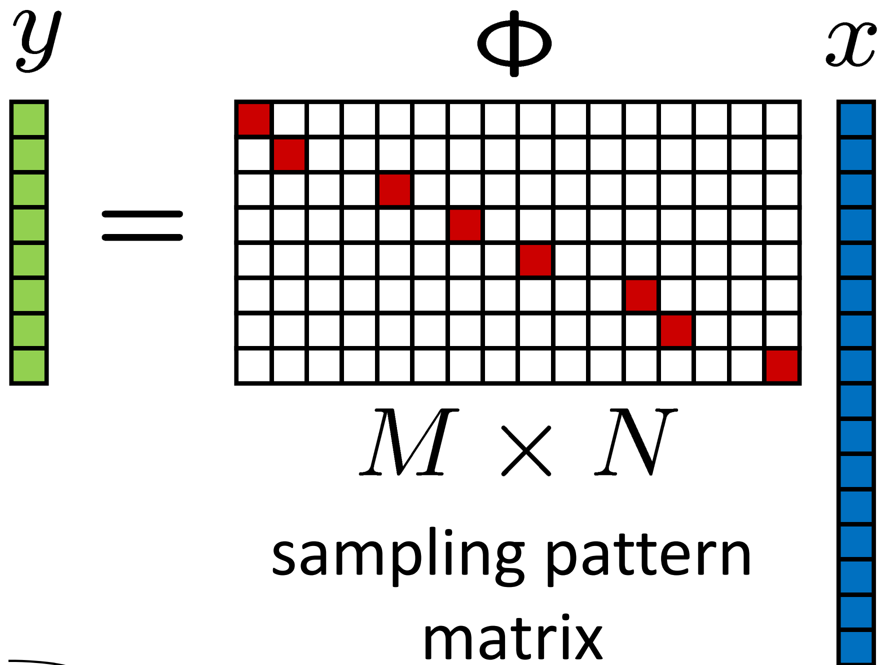


Sample/retain only a random subset of the Nyquist samples

# Matrix-vector Formulation

$M \times 1$

samples




$N \times 1$

Nyquist-rate samples

sampling pattern matrix

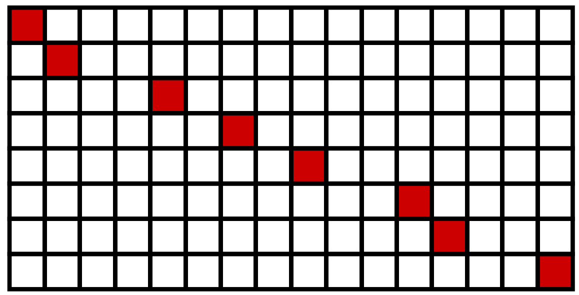
$$\left.\begin{aligned} y(1) &= x(1) \\ y(2) &= x(2) \\ y(3) &= x(5) \\ y(4) &= x(7) \\ \vdots &= \vdots \\ y(8) &= x(16) \end{aligned}\right\}$$

sampling pattern
encodes which samples to collect/keep

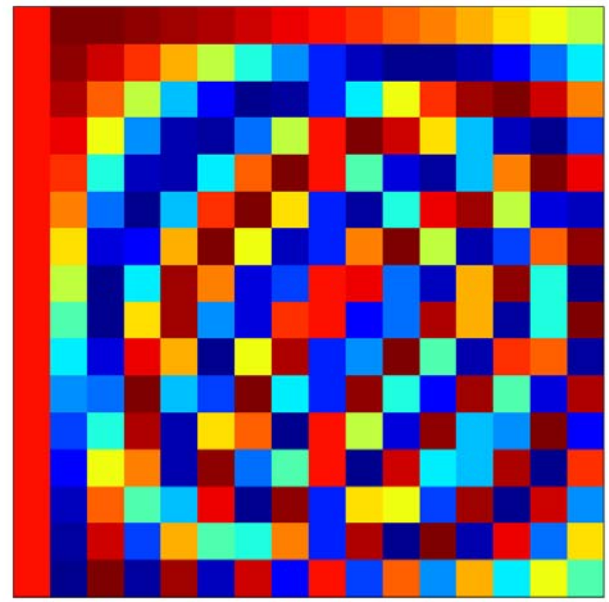
# Sparse Signal Recovery

$y : \; M \times 1$

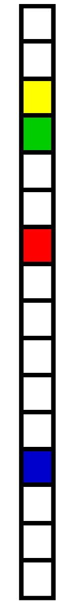
$=$



$\Phi : \; M \times N$



$\Psi : \; N \times N$



$\alpha$

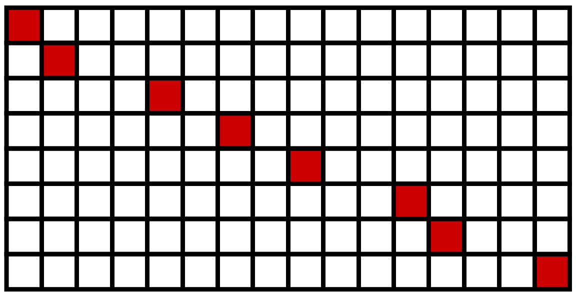
$x : \; N \times 1$

$$\widehat{\alpha} = \arg\min_{\alpha'} \|\alpha'\|_1$$
$$\text{subject to } y = \Phi\Psi\alpha'$$

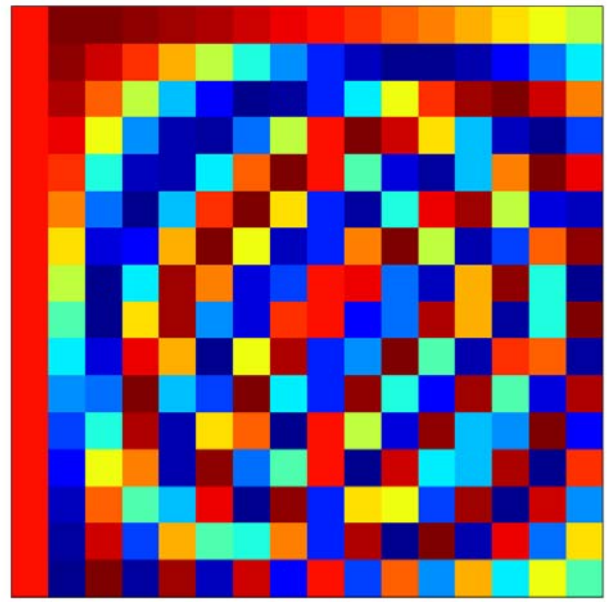
$$\text{then set } \widehat{x} = \Psi\widehat{\alpha}$$

# Incoherence

- Measurements should be *incoherent* with sparse transform
  - if sparse in time domain, cannot randomly sample in time domain



$\Phi : \; M \times N$



$\Psi : \; N \times N$

- Rows of sensing matrix $\Phi$ should have small correlation (dot product) with columns of sparsity basis $\Psi$
  - i.e., sparse basis functions cannot be "spiky"

# Incoherence: Good and Bad

- Assuming $\Phi$ = Non-uniform sampling in time

- Good: $\Psi$ = discrete Fourier transform
  – signal has a sparse spectrum
  – number of samples $M$ proportional to sparsity of spectrum
  – can use smooth windowing to reduce ringing in spectrum

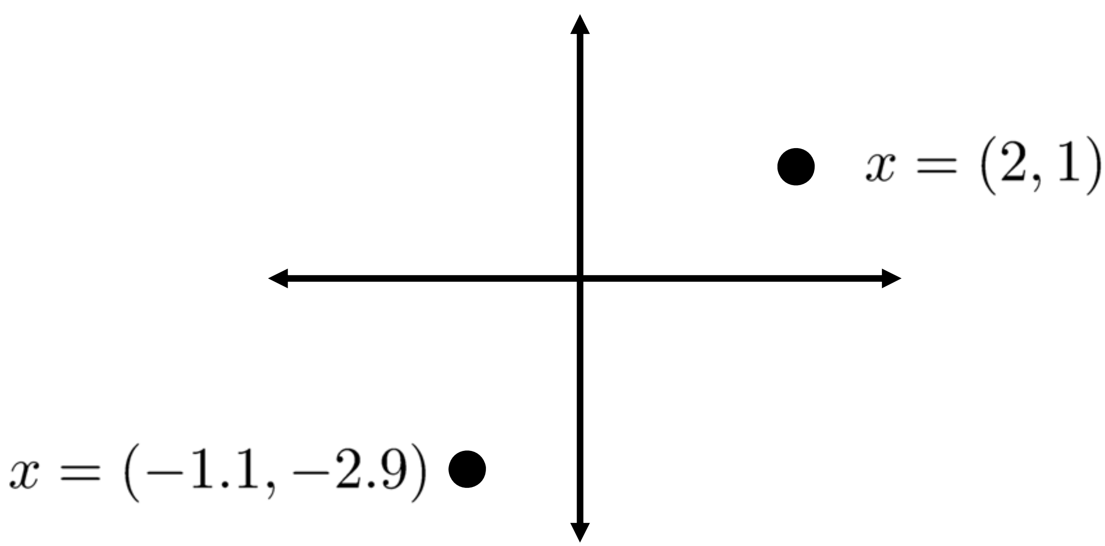
- Medium: $\Psi$ = discrete wavelet transform
  – signal is piecewise smooth
  – need to oversample compared to sparsity of wavelet transform

- Bad: $\Psi$ = identity matrix
  – signal is spiky in the time domain
  – cannot merely take # samples $\approx$ # spikes

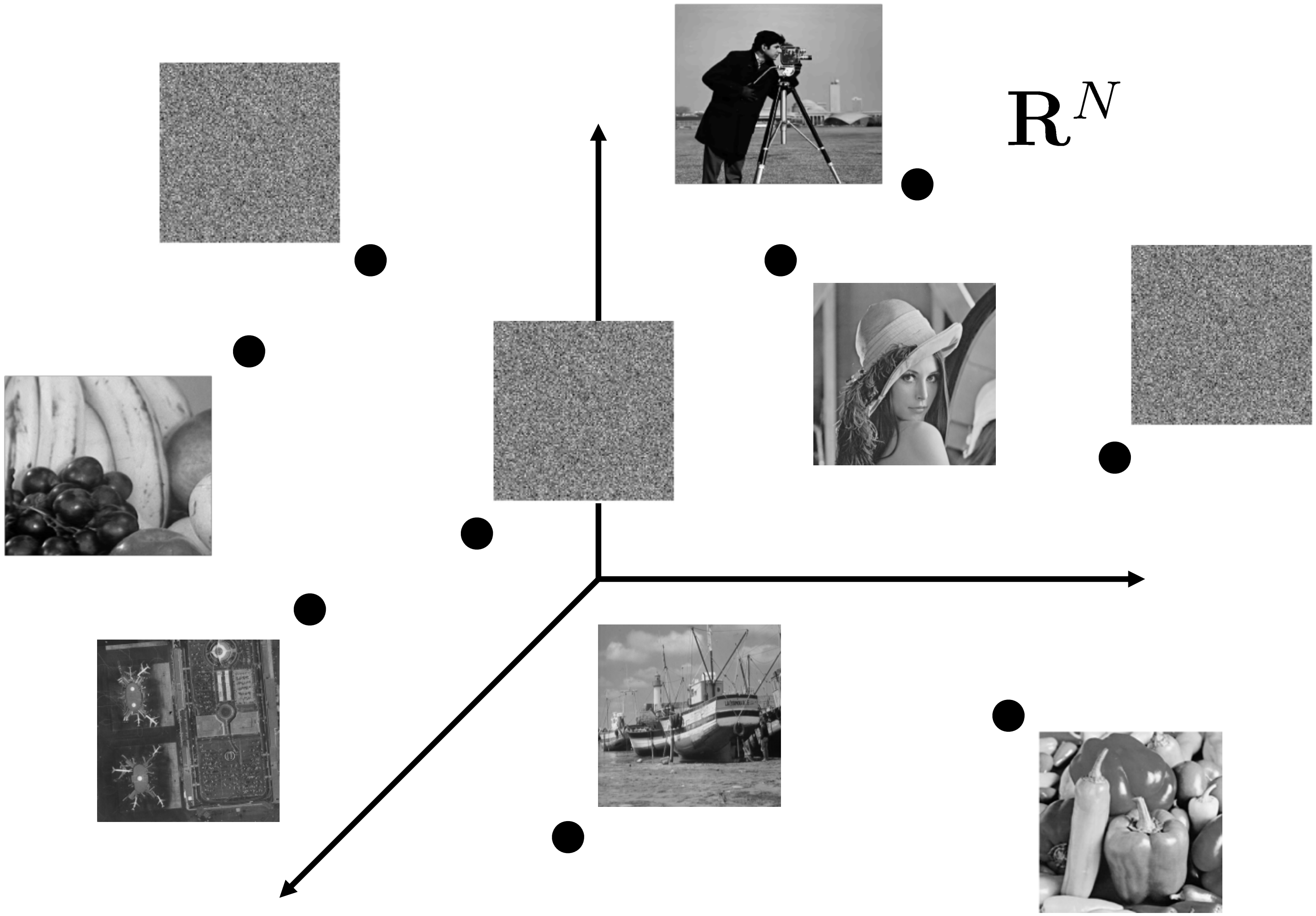
# Geometric Intuition

- Think of signals as points in some space

- Example: Two-dimensional signals

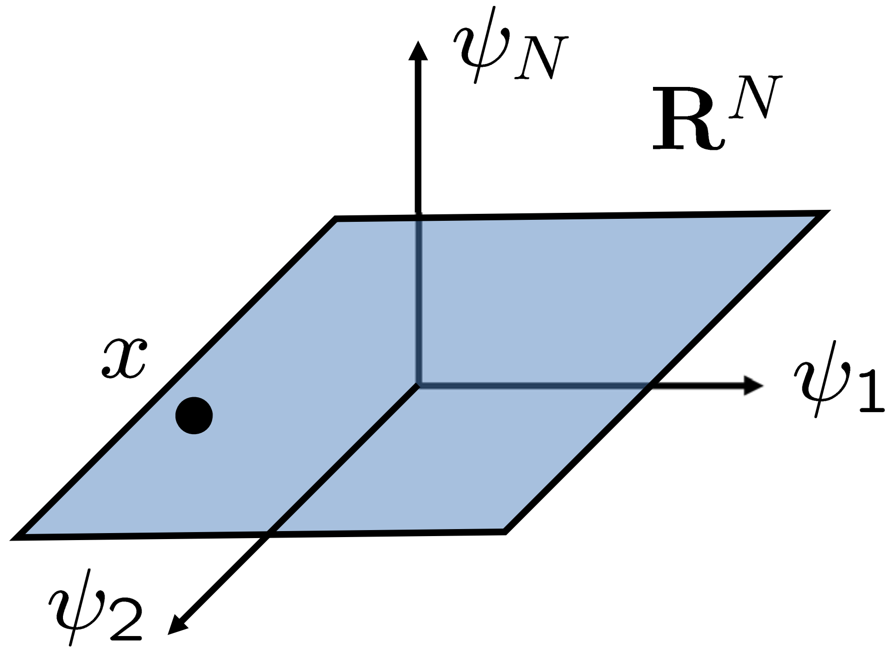


# Where are the Signals?



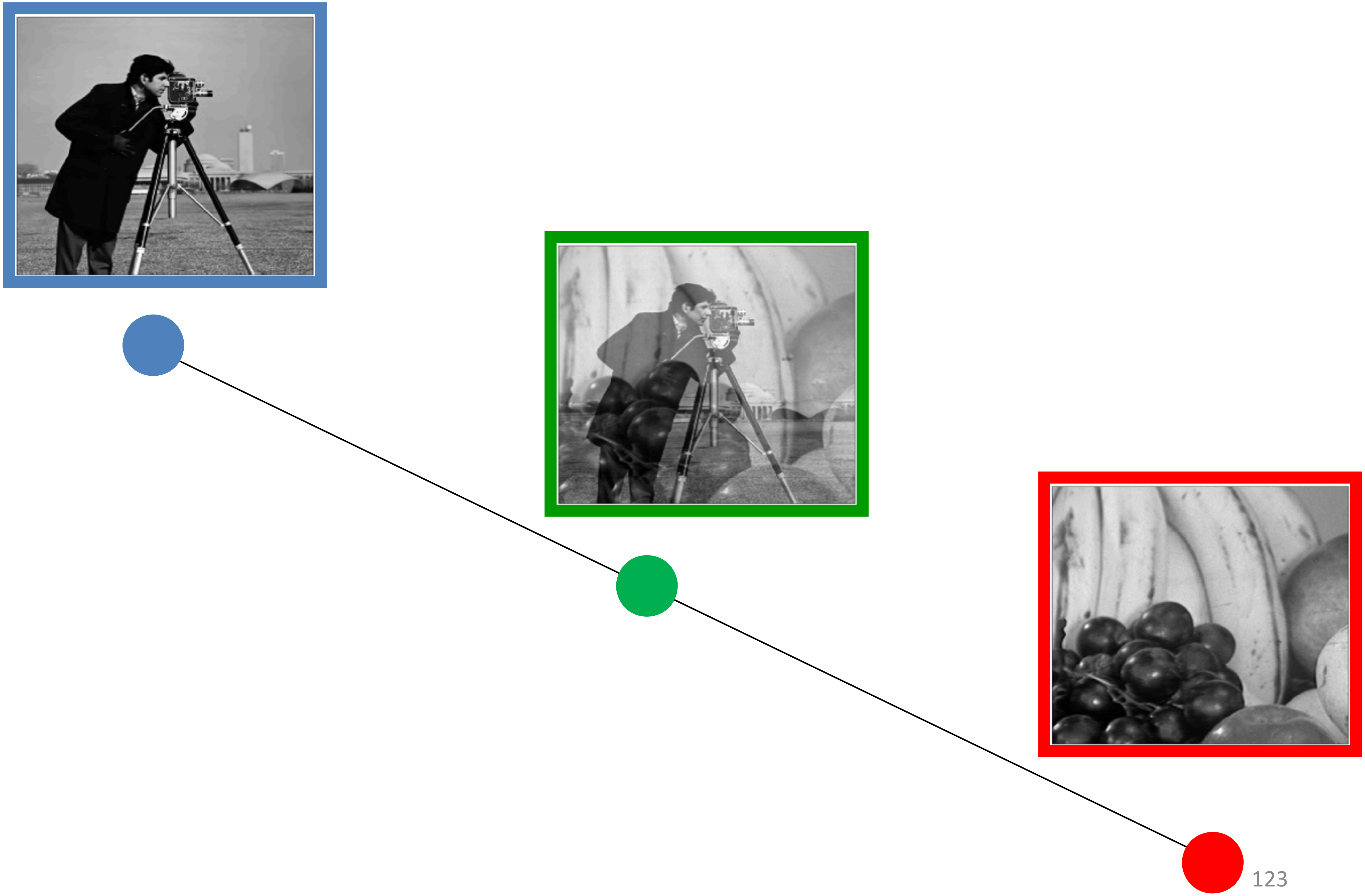

*concise* models ⇔ ***low-dimensional*** geometry

# Linear Subspace Models

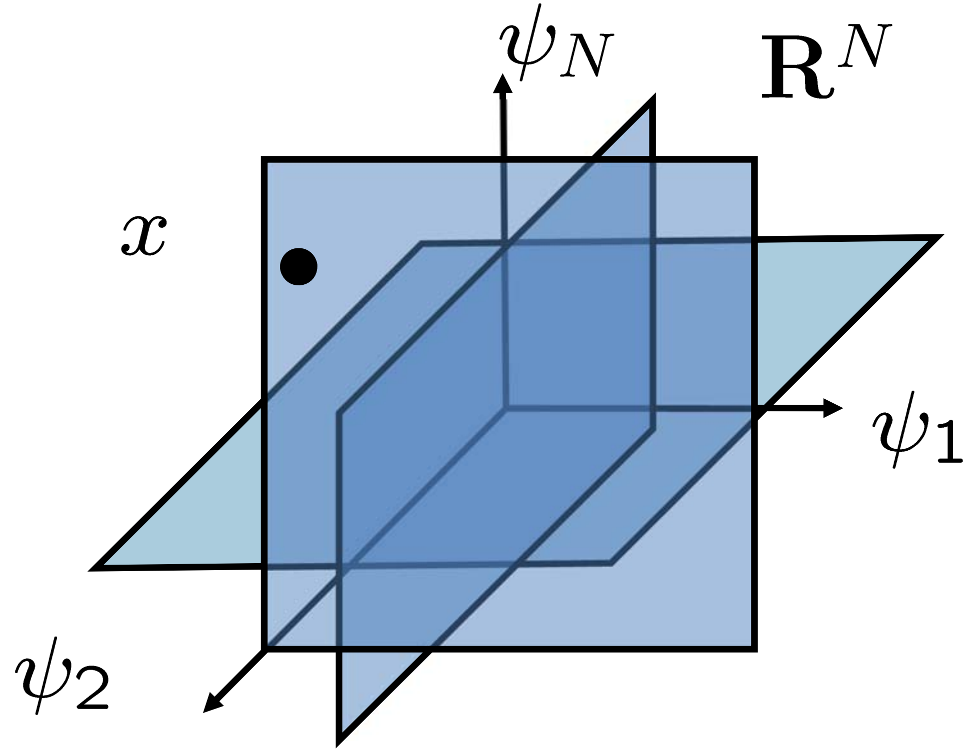



e.g., bandlimited signals in a Fourier basis

# Many Signal Families are Highly Nonlinear



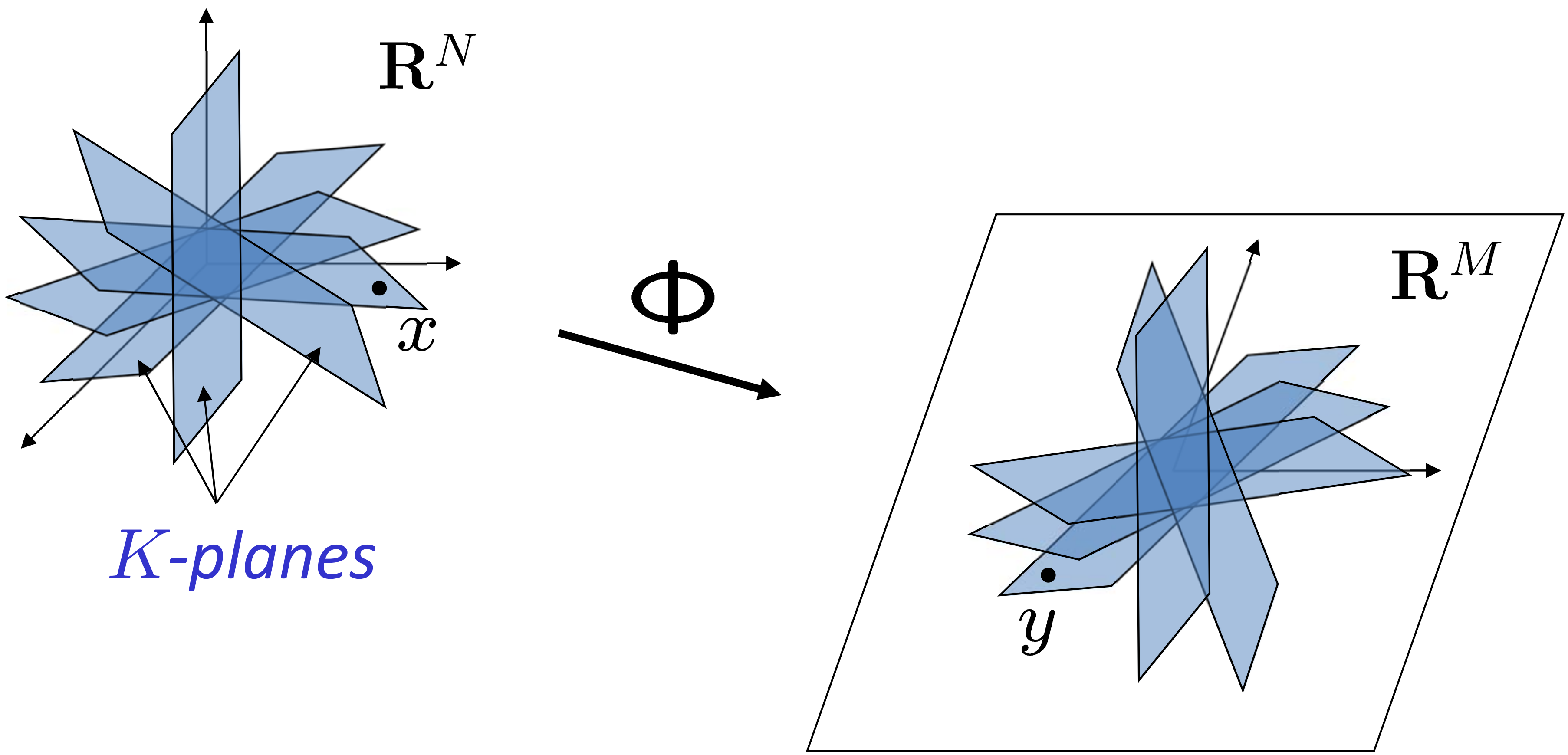
# Sparse Models: Unions of Subspaces



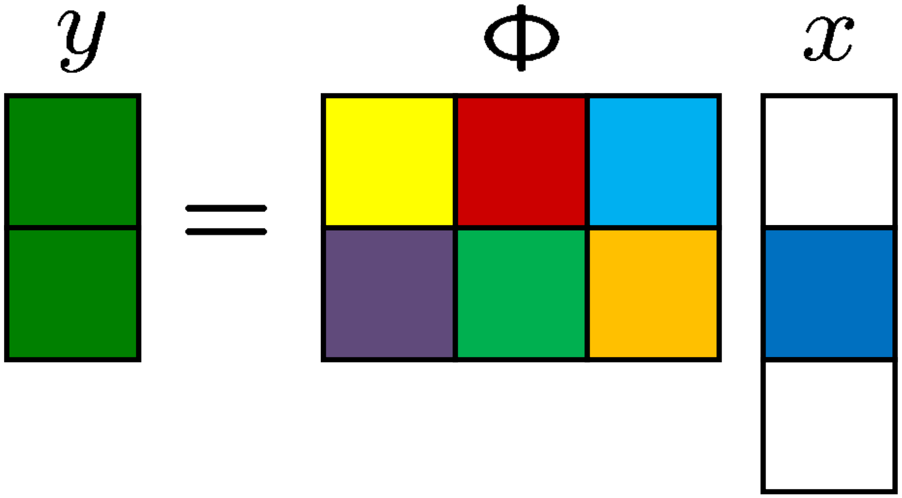

e.g., natural images in a wavelet basis

$\binom{N}{K} \approx N^K$ such $K$-dimensional subspaces in $R^N$

# Geometry: Embedding in $R^M$



# Illustrative Example

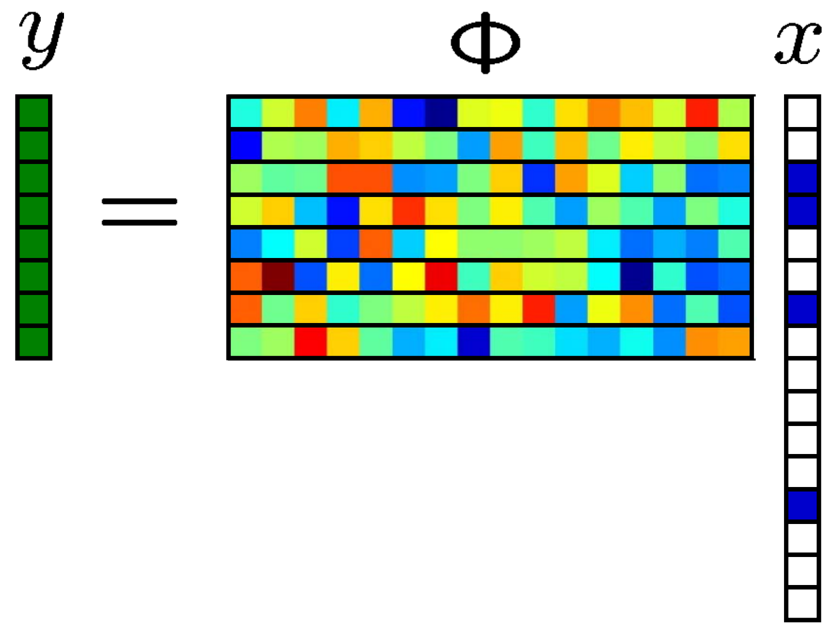



$N$ = 3: signal length
$K$ = 1: sparsity
$M$ = 2: measurements

# Restricted Isometry Property (RIP)

[Candès, Romberg, Tao]

- Sensing matrix $\Phi$ has ***RIP of order*** $K$ with isometry constant $\delta$ if




$$(1-\delta) \leq \frac{\|\Phi x\|_2^2}{\|x\|_2^2} \leq (1+\delta)$$

  holds for all $K$-sparse signals $x$.

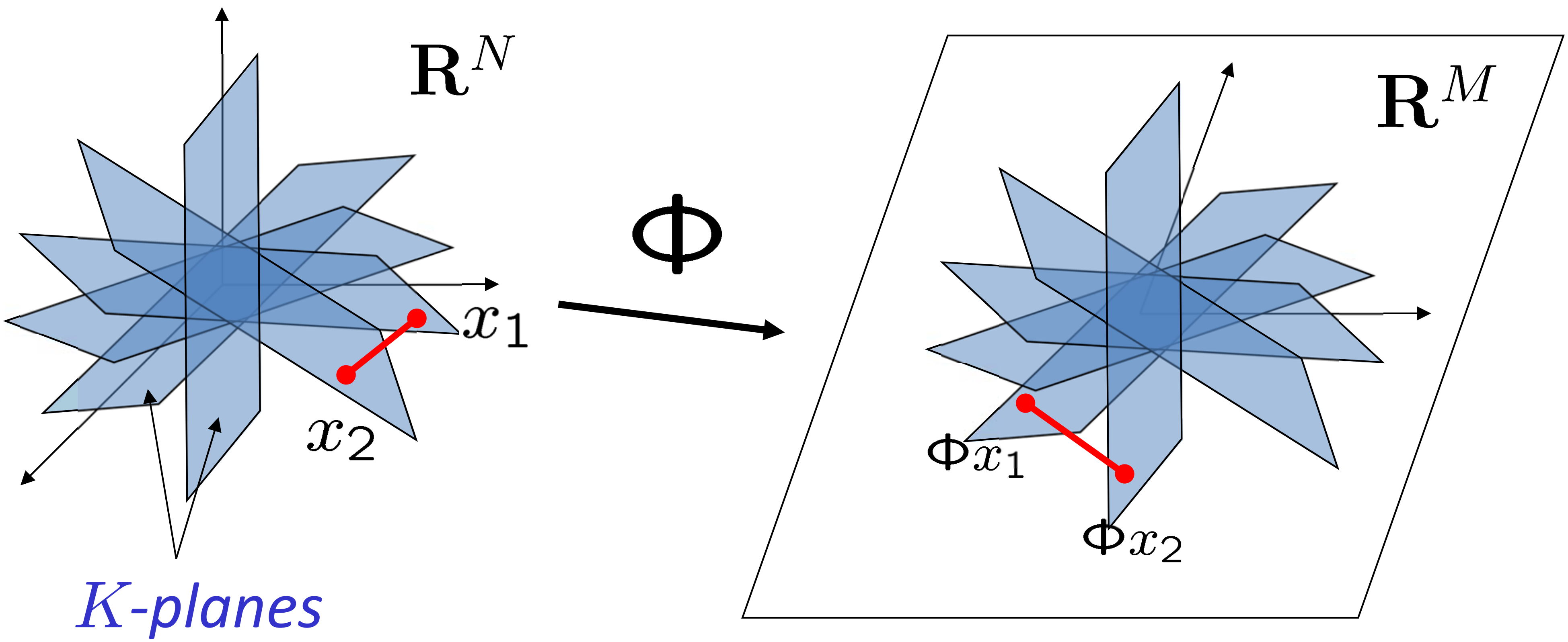
- Does *not* hold for $K$>$M$; may hold for smaller $K$.

- Implications: tractable, stable, robust recovery

# Geometric Interpretation of RIP

- RIP ***of order 2$K$*** requires: for all $K$-sparse $x_1$ and $x_2$,

$$(1-\delta) \leq \frac{\|\Phi x_1 - \Phi x_2\|_2^2}{\|x_1 - x_2\|_2^2} \leq (1+\delta)$$




- Stable embedding of the sparse signal family

# Implications of RIP

[Foucart; Candès]

RIP of order $2K$ with $\delta < 0.47$ implies:

1. *Exact recovery:*

   All $K$-*sparse* $x$ are perfectly recovered via $L_1$ minimization.

2. *Robust and stable recovery:*

   Measure $y = \Phi x + e$ with $\|e\|_2 \leq \epsilon$, and recover

   $$\widehat{x} = \arg\min \|x'\|_1 \;\; s.t. \;\; \|y - \Phi x'\|_2 \leq \epsilon.$$

   Then *for any* $x \in R^N$,

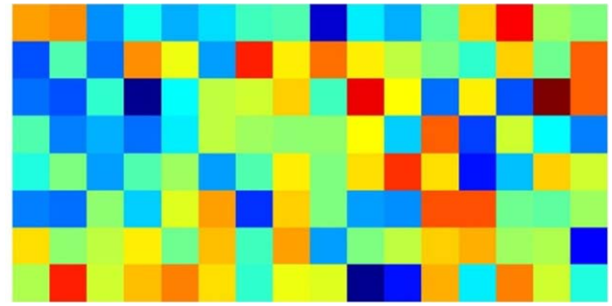
   $$\|x - \widehat{x}\|_2 \leq C_1 \frac{\|x - x_K\|_1}{\sqrt{K}} + C_2 \epsilon.$$

# Random Matrices Satisfy the RIP

[Mendelson, Pajor, Tomczak-Jaegermann; Davenport, DeVore, Baraniuk, Wakin]

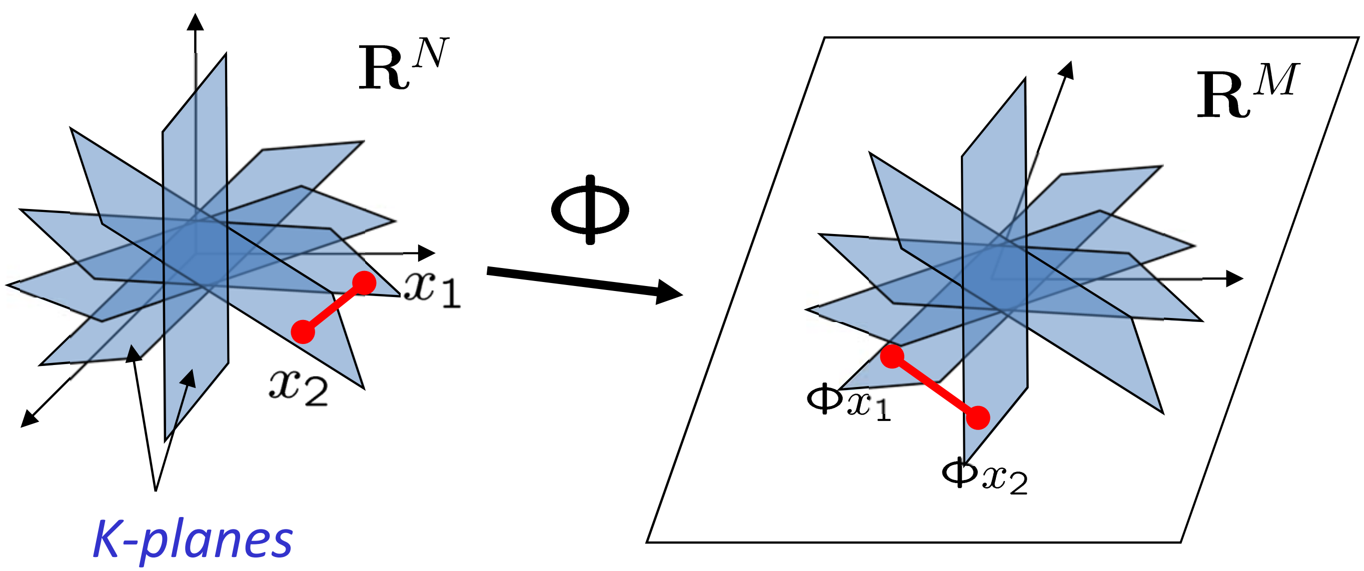
- Suppose $\Phi$ is drawn randomly with Gaussian entries and that

$$M = O(K \log N).$$



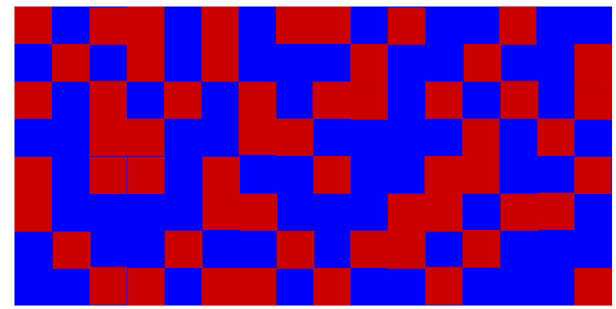
$\Phi: \ M \times N$

Then with high probability, $\Phi$ satisfies the RIP.



# Random Matrices – Other Choices

- Random Gaussian matrices
- Random Bernoulli (+/- 1) matrices
- Random subgaussian matrices



$\Phi : M \times N$

- Entries are independent and identically distributed (i.i.d.)

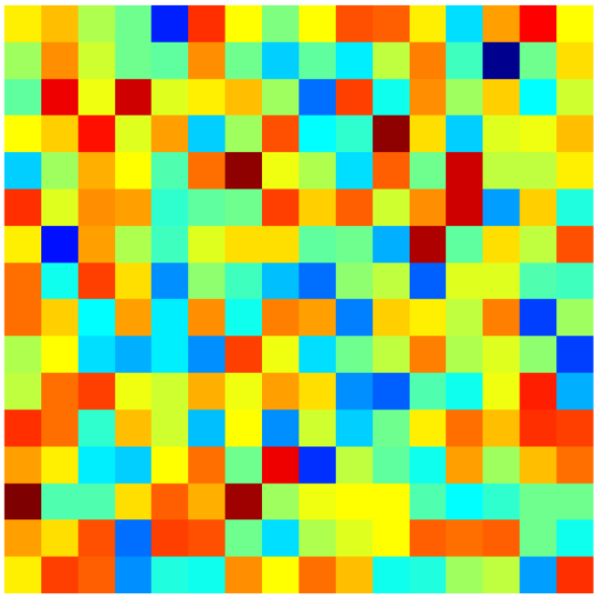
- $M$ = $O$($K$ log $N$) suffices with probability 1-$O$($e^{-CN}$)

- All of these constructions are ***universal*** in that they work for any fixed choice of the sparsifying transform $\Psi$.

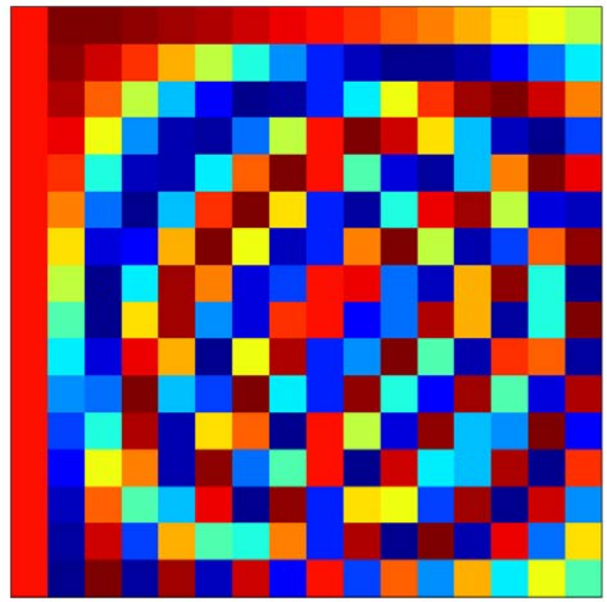
# Subsampling Incoherent Matrices

[Rudelson and Vershynin; Candès, Romberg, Tao; Rauhut]

- Start with two square matrices (each orthonormal)



$U: \ N \times N$



$\Psi: \ N \times N$

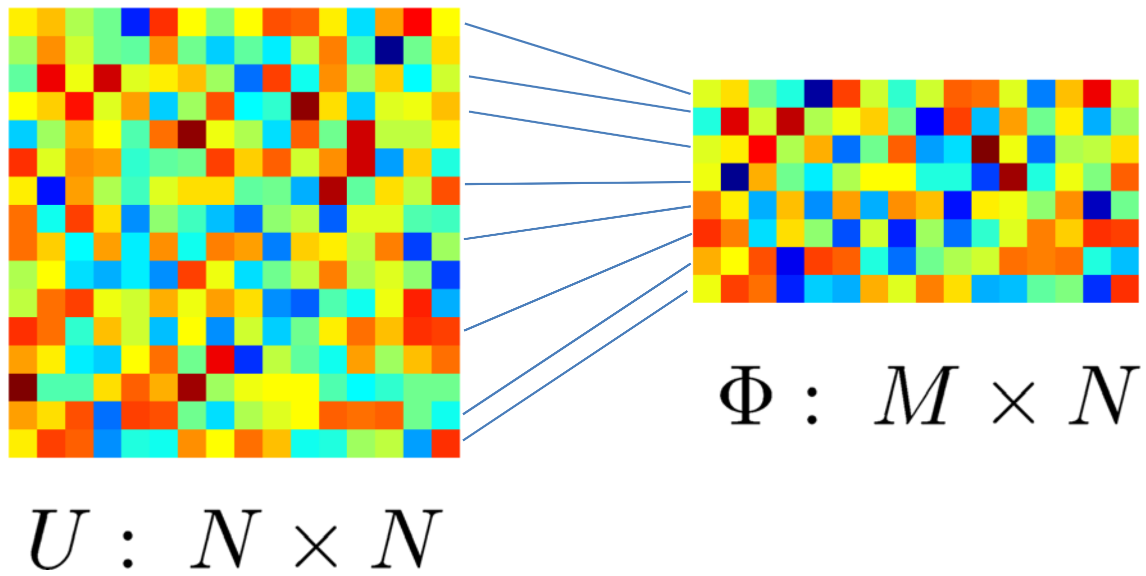
- Define coherence between matrices

$$\mu = \sqrt{N} \cdot \max_{i,j} |\langle u_i, \psi_j \rangle|$$

# Subsampling Incoherent Matrices

- Coherence between matrices

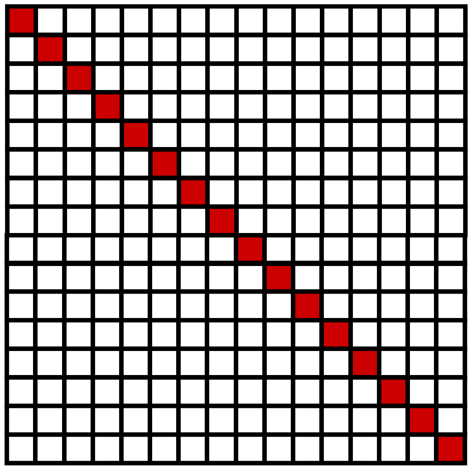
$$\mu = \sqrt{N} \cdot \max_{i,j} |\langle u_i, \psi_j \rangle|$$



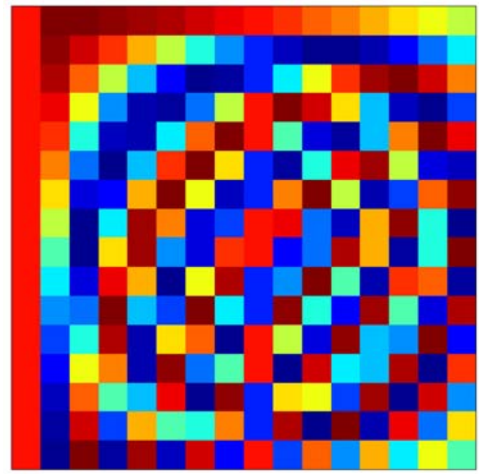

- Choose $M$ rows from $U$ to populate sensing matrix $\Phi$
- RIP satisfied with high probability if $M = O(K\mu^2 \log^4 N)$

# Incoherence: Good

- $U$ = identity matrix, $\Psi$ = discrete Fourier transform



$U : \; N \times N$



$\Psi : \; N \times N$

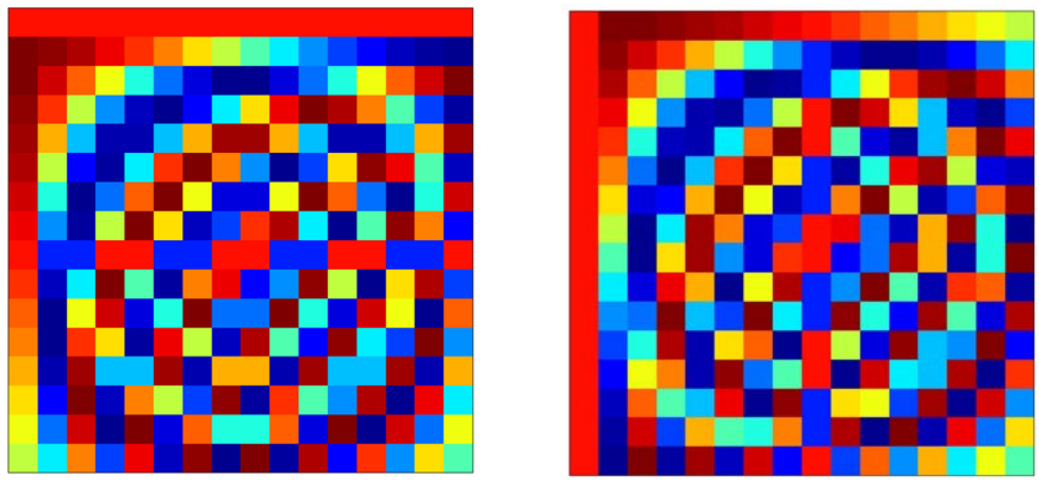
$$\mu = \sqrt{N} \cdot \max_{i,j} |\langle u_i, \psi_j \rangle| = 1$$

$$M = O(K\mu^2 \log^4 N) = O(K \log^4 N)$$

- Signals that are sparse in the frequency domain can be efficiently sampled in the time domain (and vice versa).

# Incoherence: Bad

- $U = \Psi$



$$U : \; N \times N \qquad \Psi : \; N \times N$$

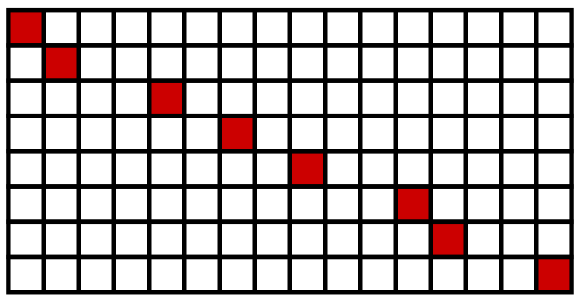
$$\mu = \sqrt{N} \cdot \max_{i,j} |\langle u_i, \psi_j \rangle| = \sqrt{N}$$

$$M = N$$

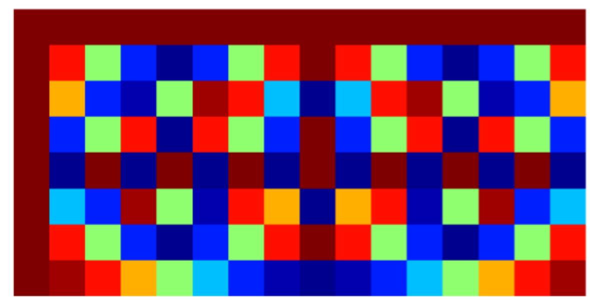
- The sampling domain must be different from the sparsity domain.

# Fast Measurement Operators

- Subsampled identity, subsampled Fourier transform



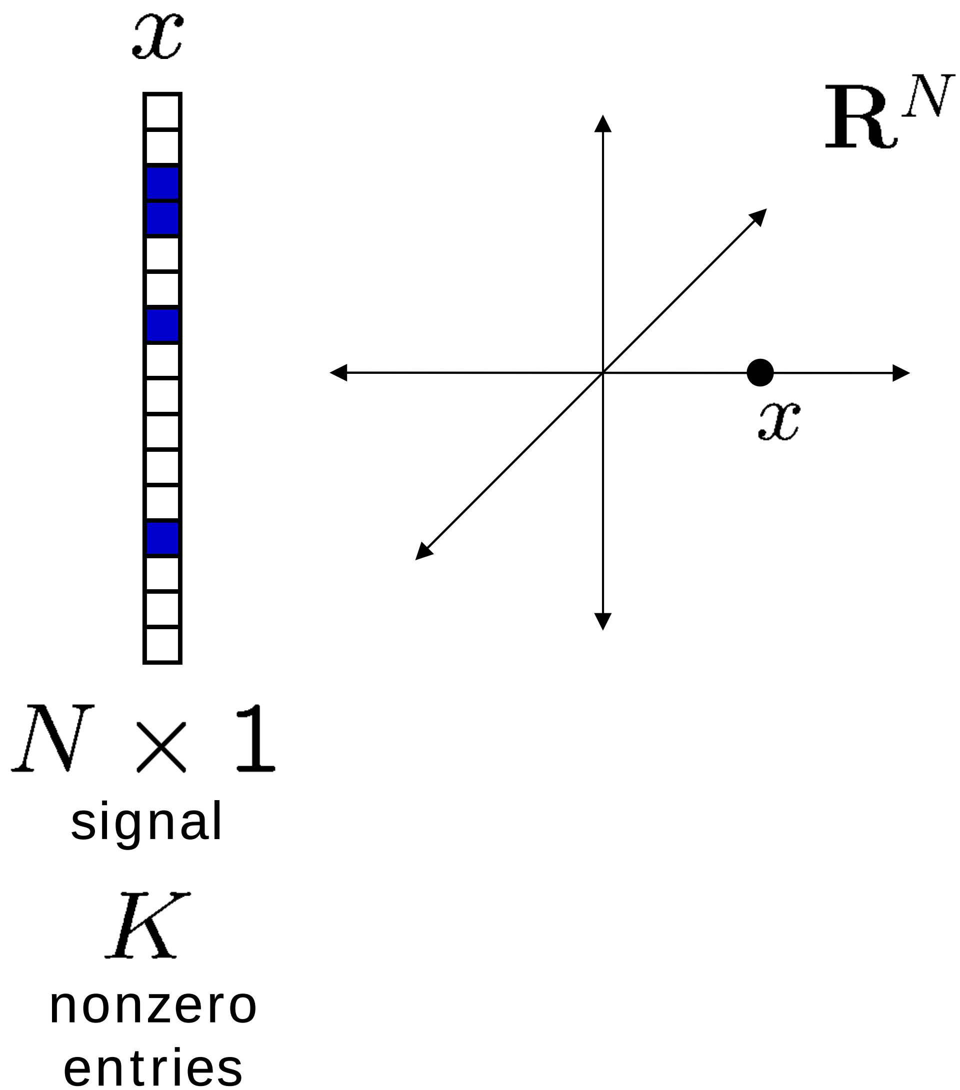
$$\Phi : \; M \times N$$



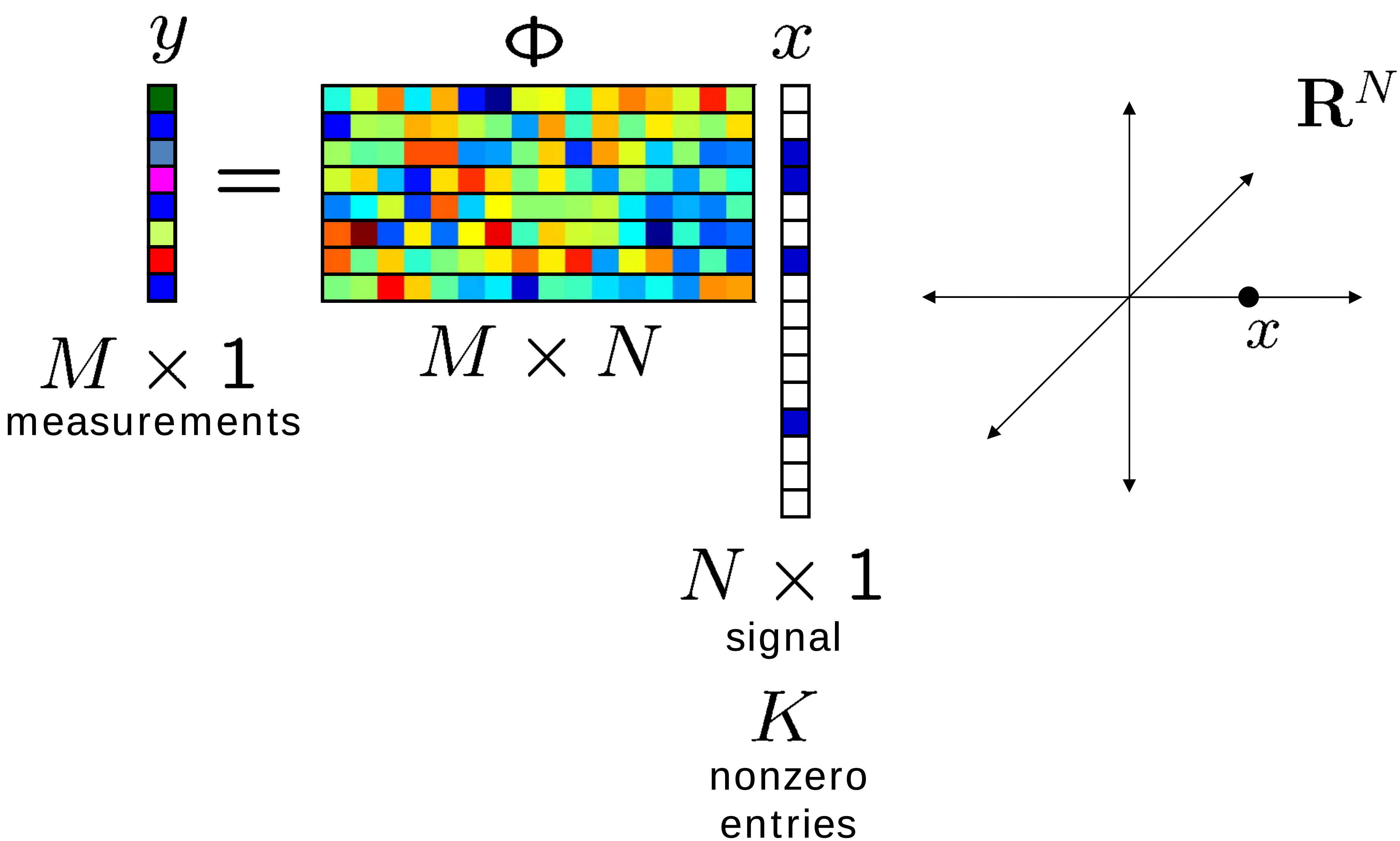
$$\Phi : \; M \times N$$

- Fast JL transforms [Ailon and Chazelle; see also Krahmer and Ward]
  – ingredients: random sign flips, Hadamard transforms, sparse Gaussian matrices, etc.
  – useful in applications where random compression is applied in software, after $x$ is sampled conventionally

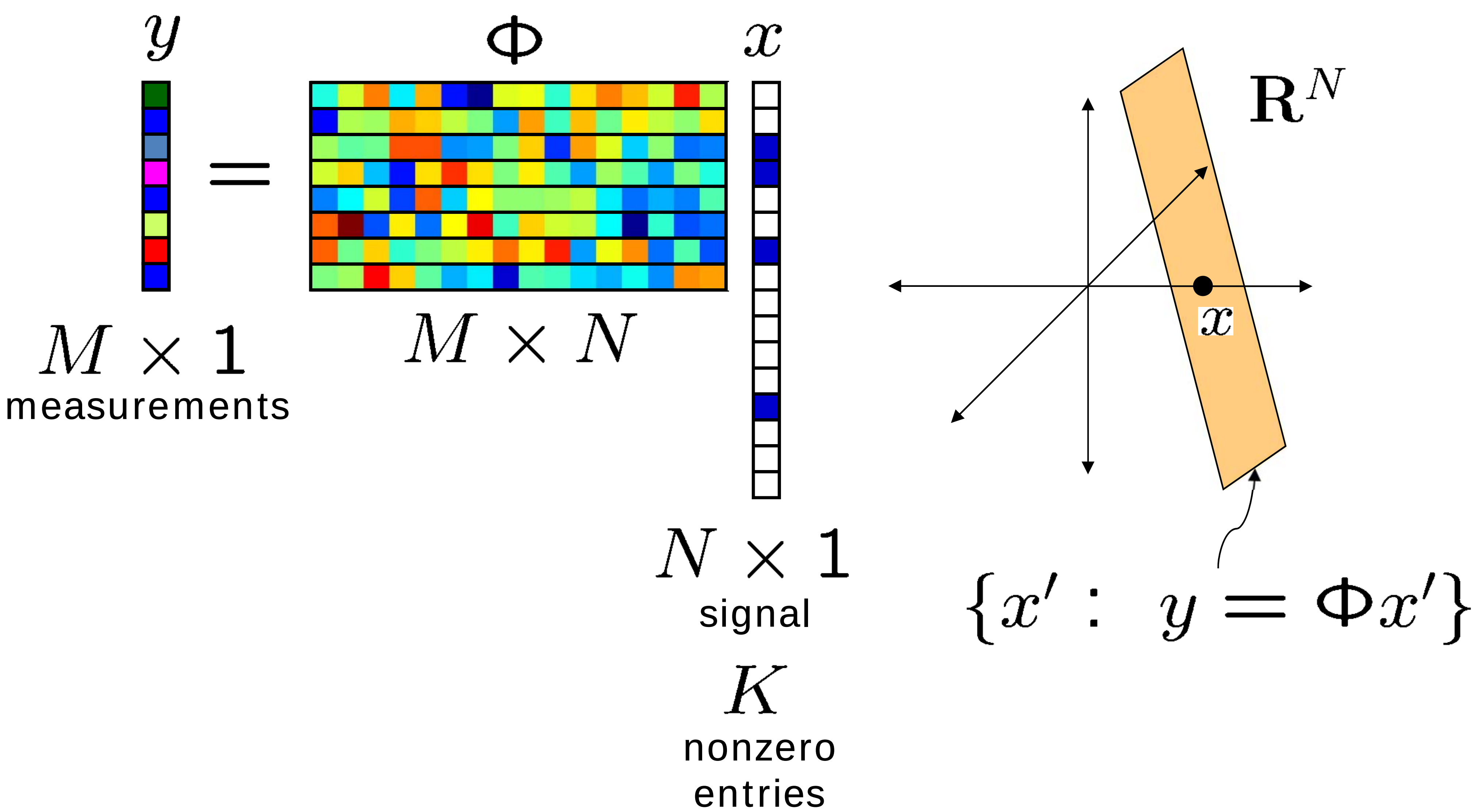
# The Geometry of $L_1$ Recovery



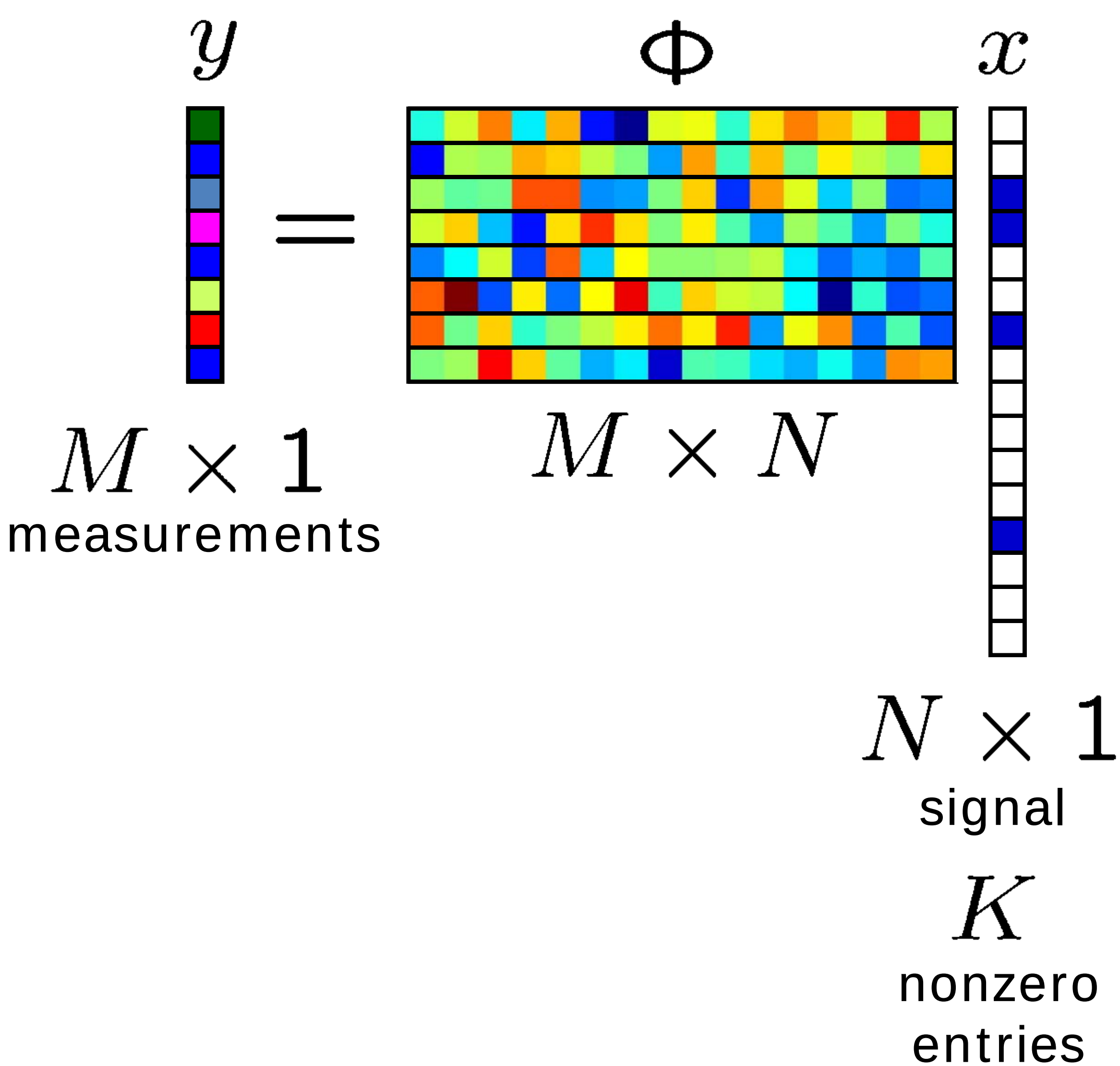
# The Geometry of $L_1$ Recovery

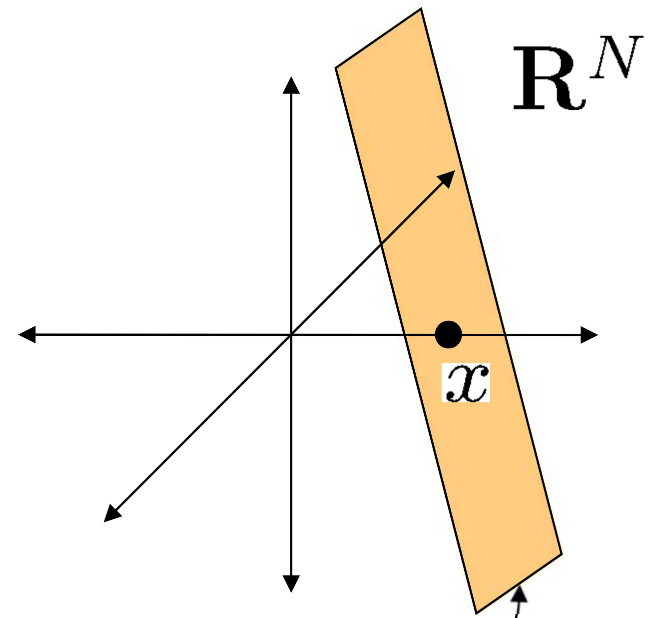


# The Geometry of $L_1$ Recovery



# The Geometry of $L_1$ Recovery



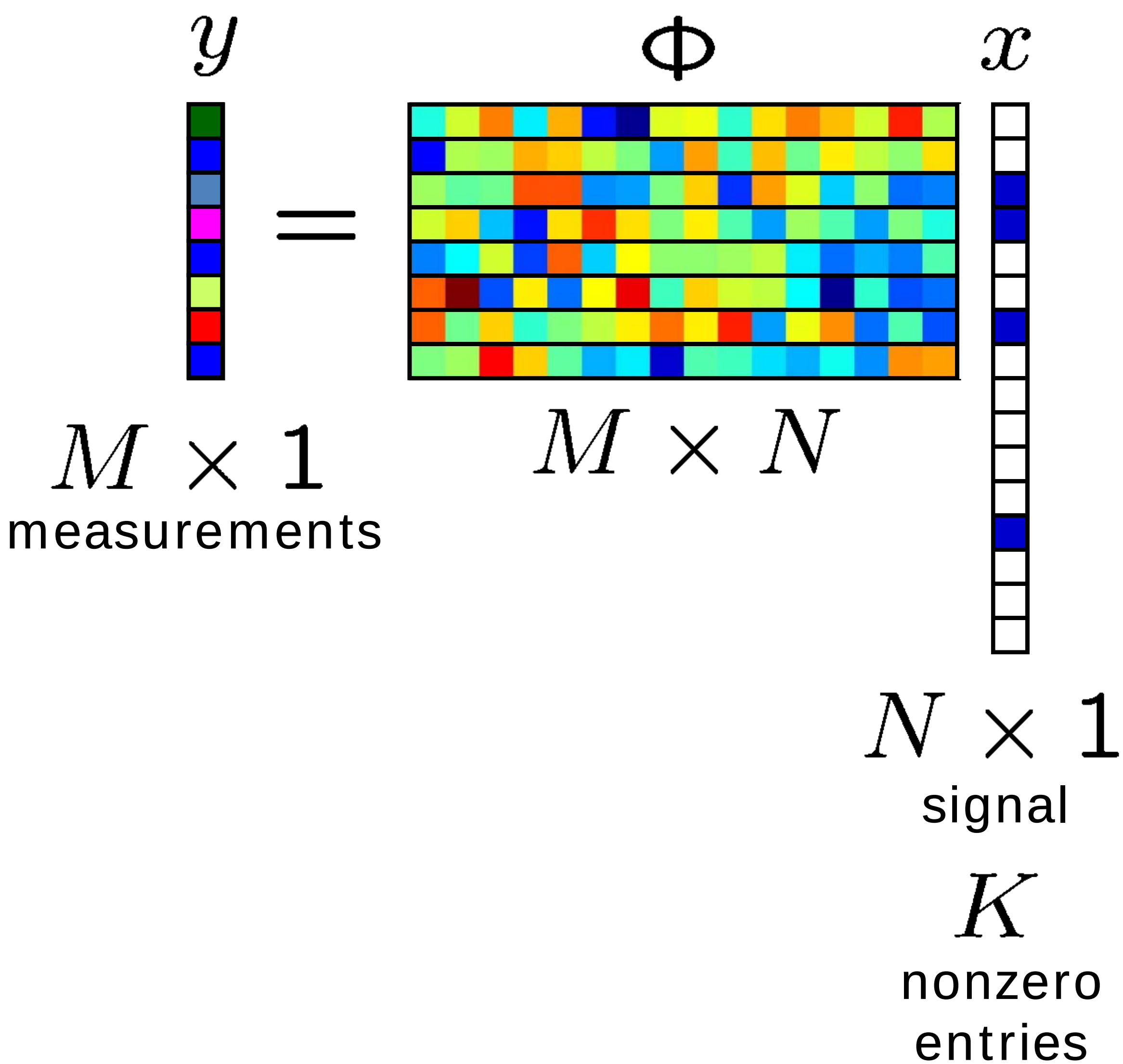



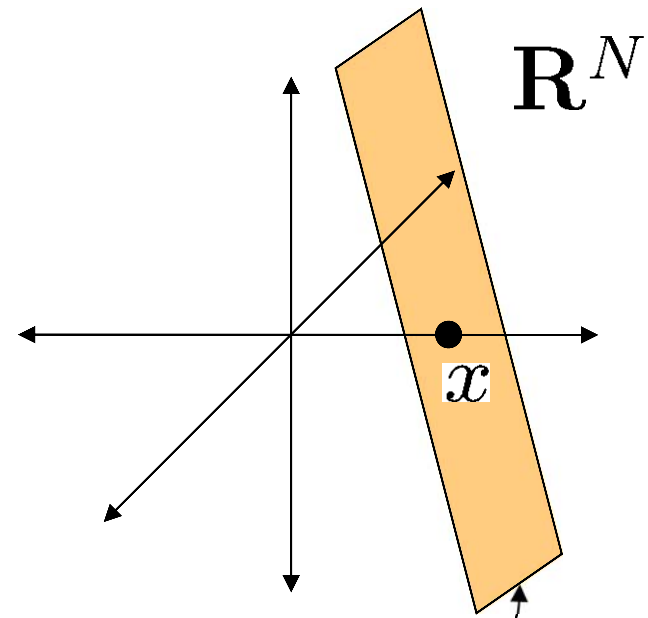

$$\{x' : \; y = \Phi x'\}$$

*null space of $\Phi$*

*translated to $x$*

# The Geometry of $L_1$ Recovery




$$\{x' : \ y = \Phi x'\}$$

*null space of $\Phi$*

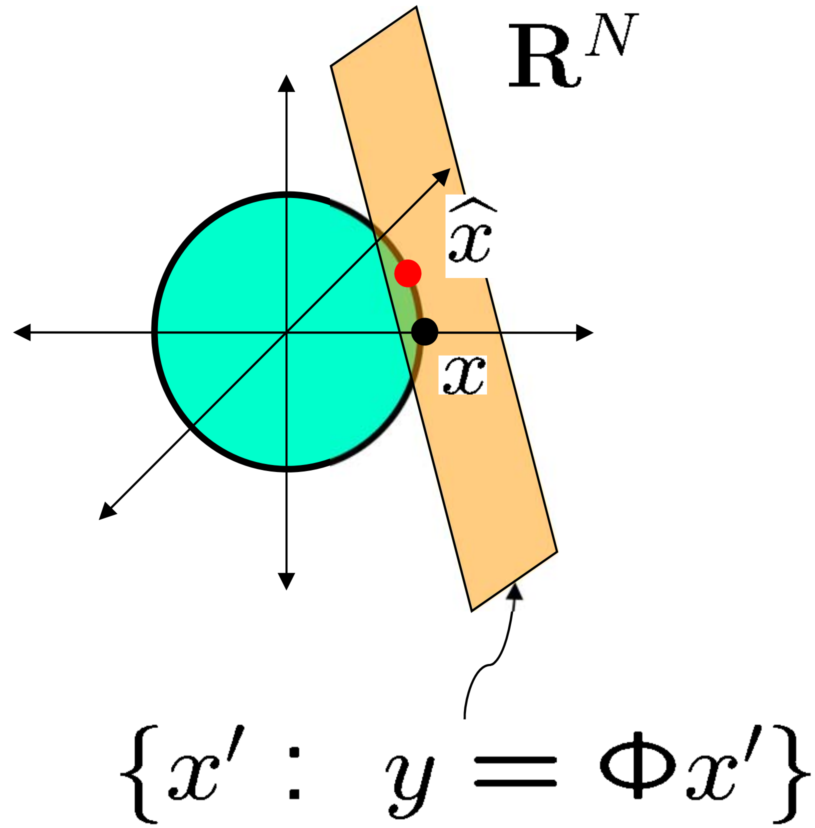
*translated to $x$*

*random orientation*

*dimension $N$-$M$*

# Why $L_2$ Doesn't Work

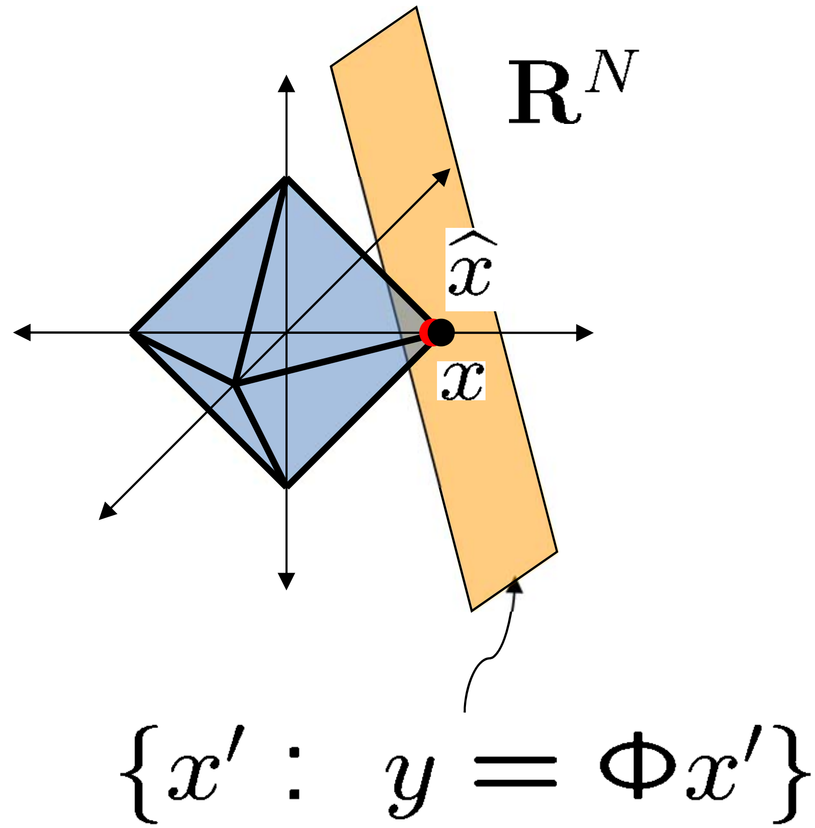
$$\widehat{x} = \arg \min_{y=\Phi x'} \|x'\|_2$$




least squares,
minimum $L_2$ solution
is almost ***never sparse***

# Why $L_1$ Works

$$\widehat{x} = \arg\min_{y=\Phi x'} \|x'\|_1$$

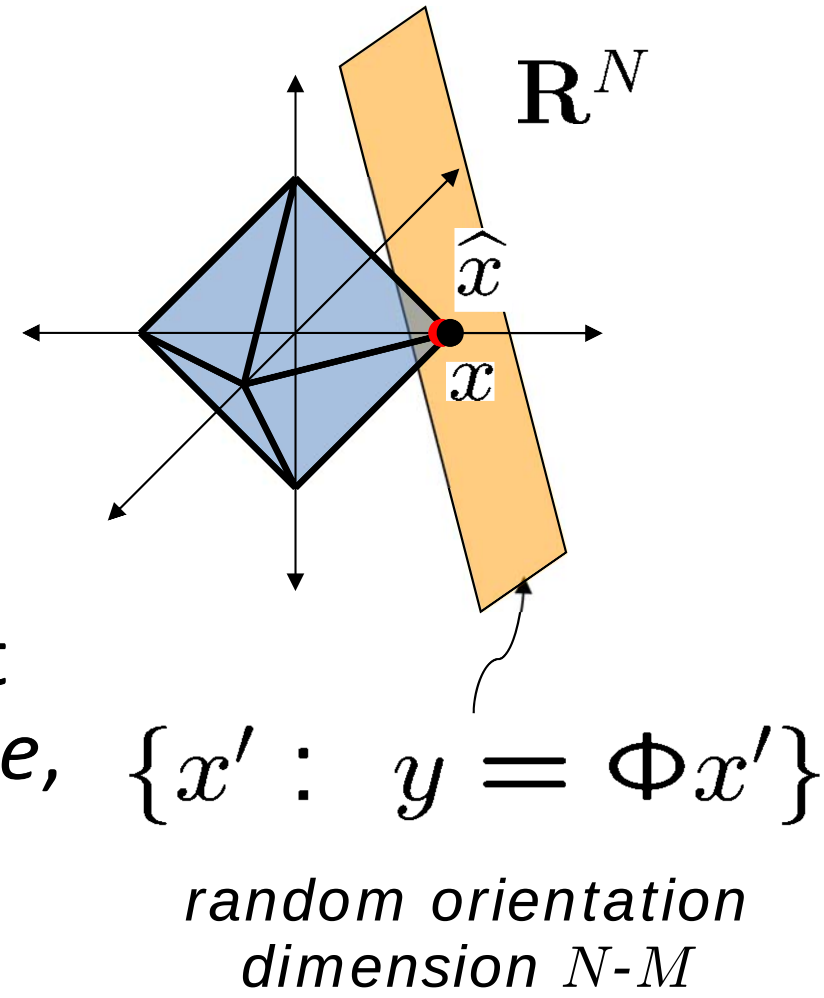



minimum $L_1$ solution is ***exactly correct*** if

$M \approx K \log N \ll N$

# Why $L_1$ Works

$$\widehat{x} = \arg\min_{y=\Phi x'} \|x'\|_1$$




**Criterion for success:**

Ensure with high probability that a *randomly oriented ($N$-$M$)-plane, anchored on a $K$-face* of the $L_1$ ball, *will not intersect* the ball.

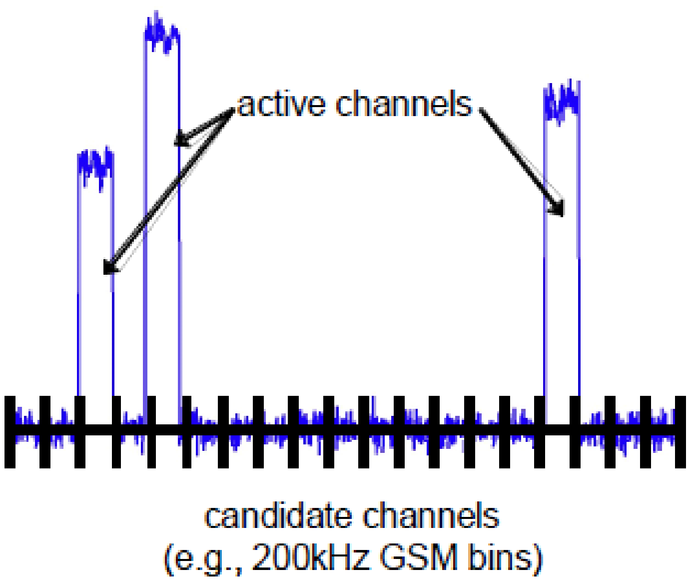
This holds when $M \approx K \log(N)$ [Donoho, Tanner]

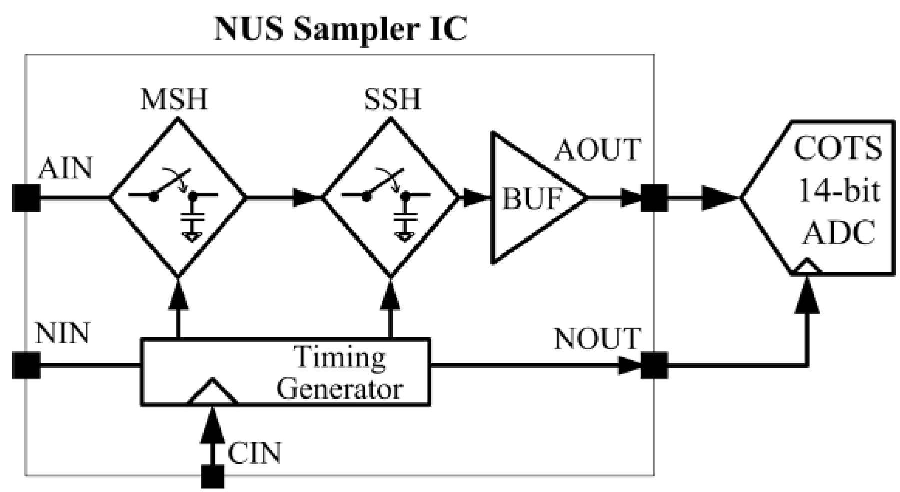
# 5. What (else) can we do with these ideas?

# Non-uniform Sampler

[with Northrop Grumman, Caltech, Georgia Tech]

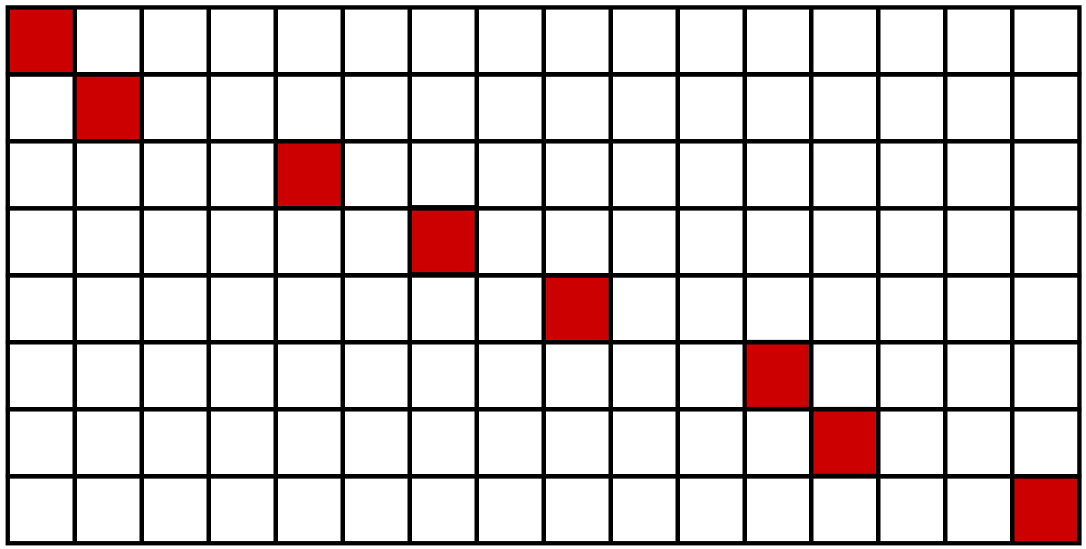
- Underclock a standard ADC to capture spectrally sparse signals




- Prototype system:
  - captures 1.2GHz+ bandwidth (800MHz to 2GHz) with 400MHz ADC underclocked to 236MHz (10x sub-Nyquist)
  - up to 100MHz total occupied spectrum

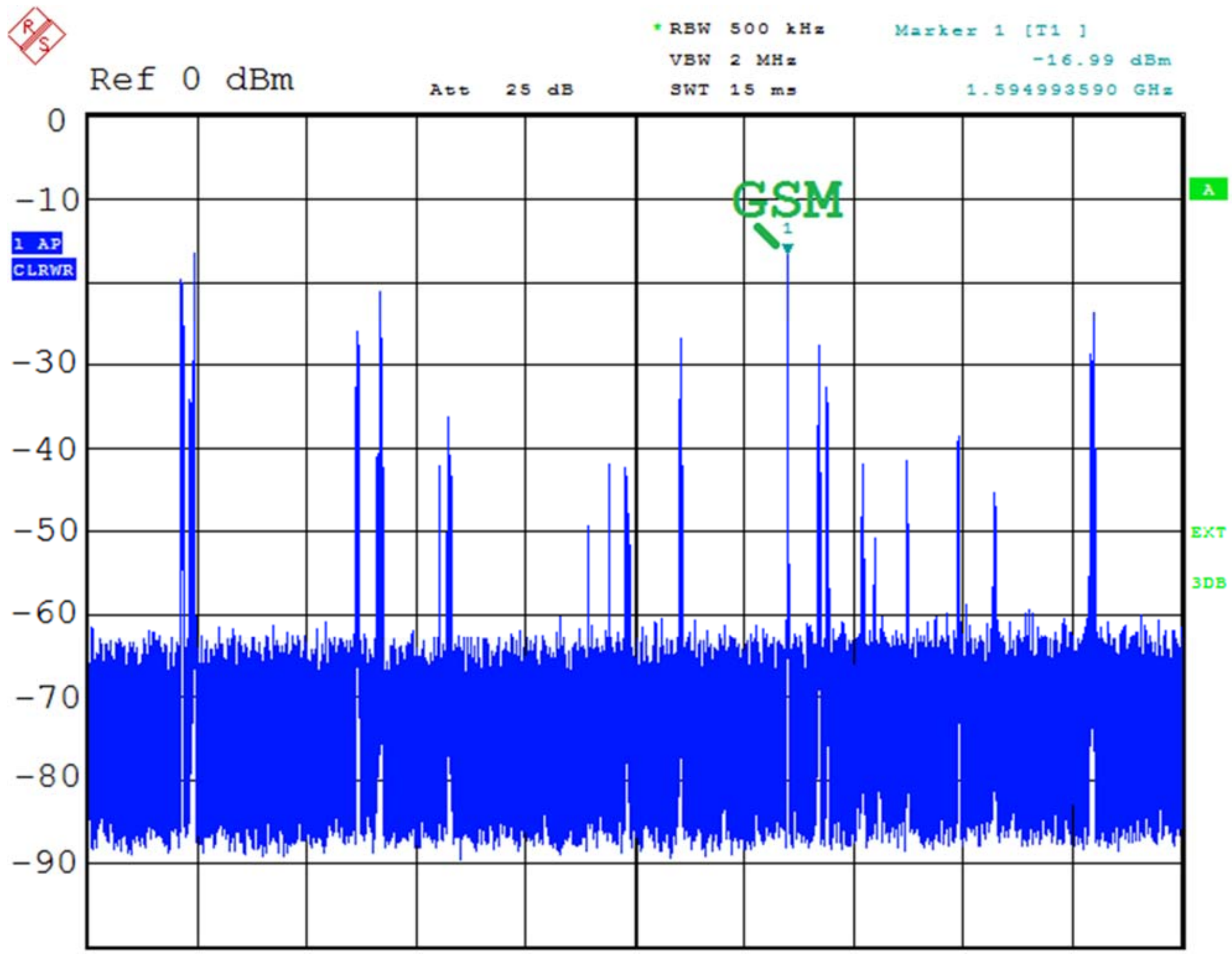
# Non-uniform Sampler – Sensing Matrix



$$\Phi : \; M \times N$$

# Non-uniform Sampler - Results

- Successful decoding of GSM signal among 100MHz of clutter



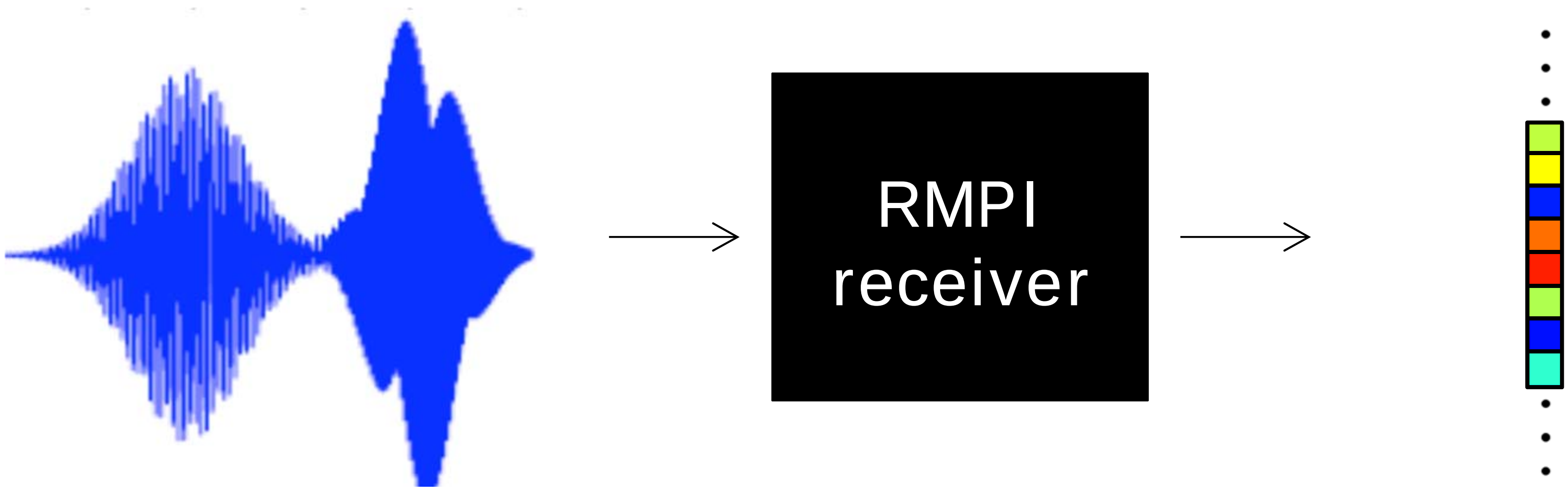

700MHz 2.1GHz

frequency

# Random Modulation Pre-Integrator (RMPI)

[with Northrop Grumman, Caltech, Georgia Tech]

- RMPI receiver
  - four parallel “random demodulator” channels
  - effective instantaneous bandwidth spanning 100MHz—2.5GHz
  - 385 MHz measurement rate (13x sub-Nyquist)
- Goal: identify radar pulse descriptor words (PDWs)

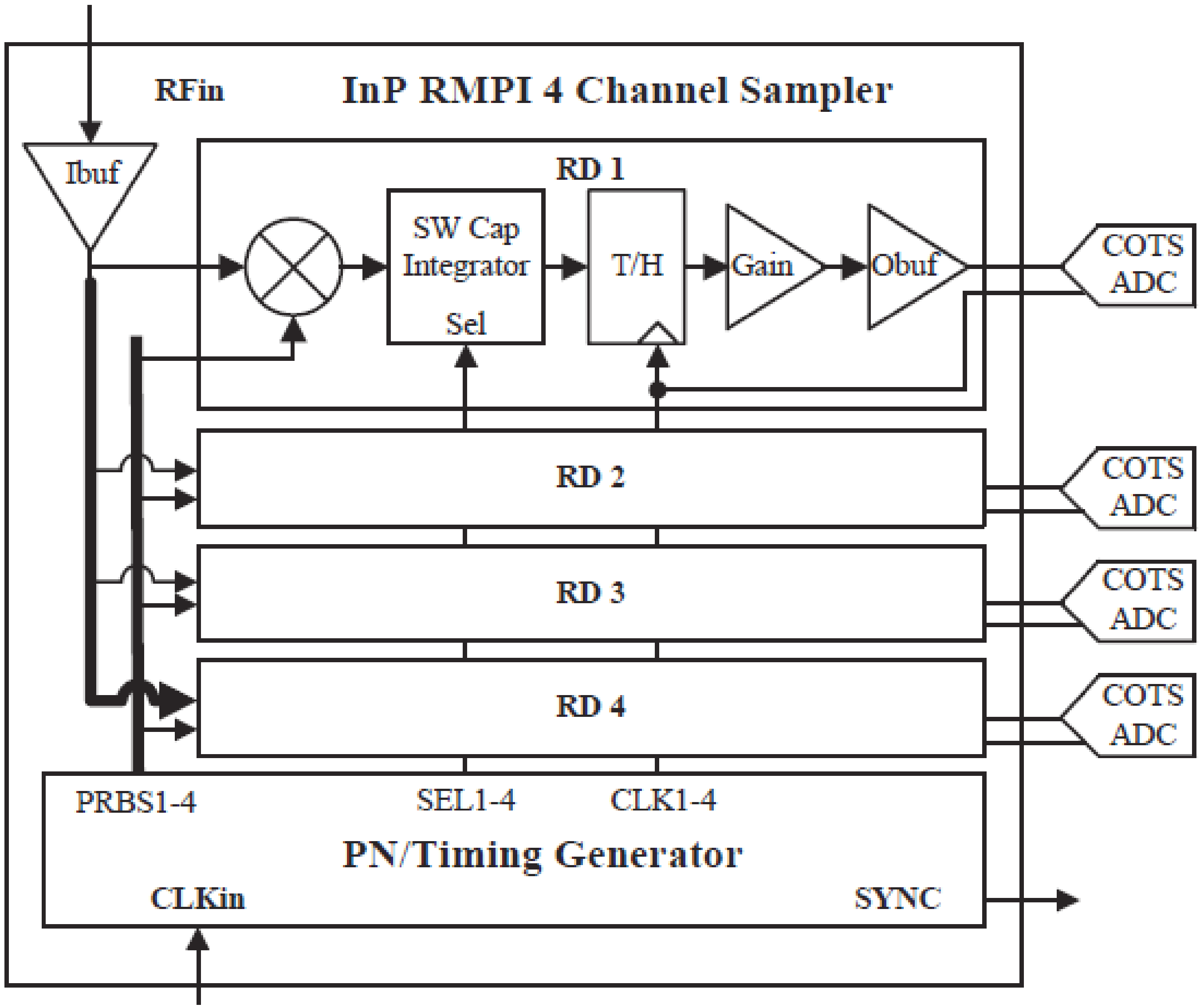



radar pulses (analog)

discrete, low-rate,
information-carrying
measurements

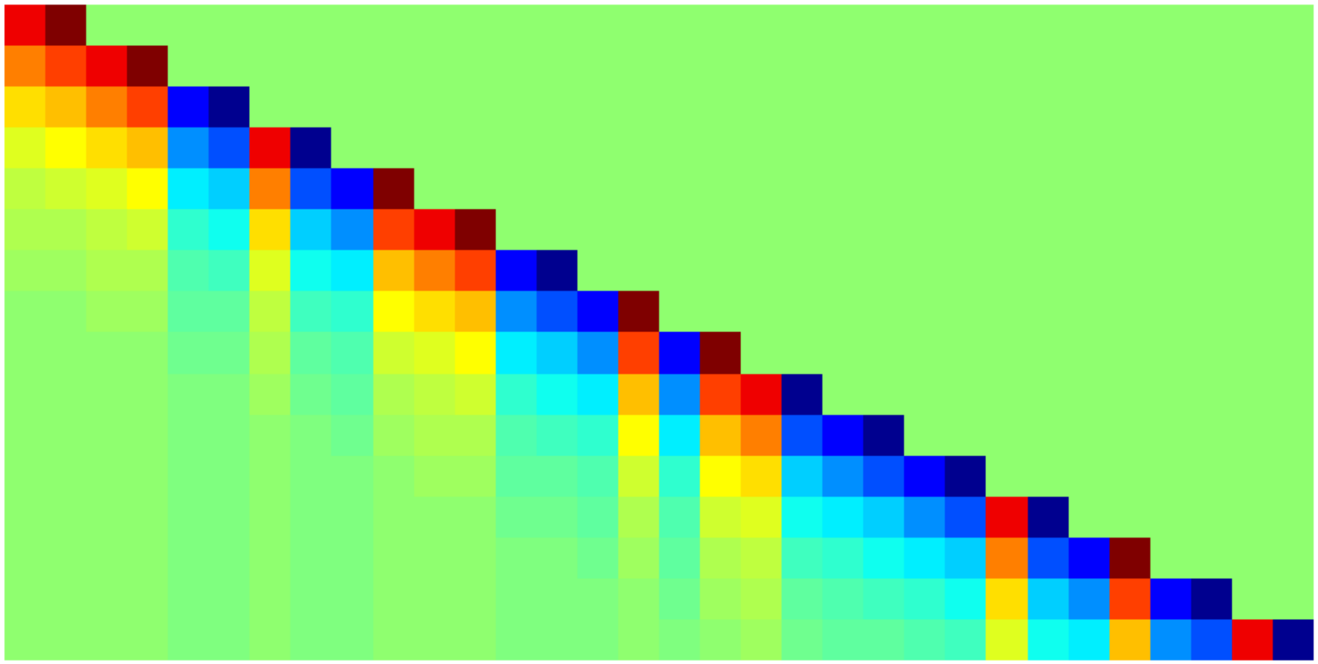
# RMPI Architecture

- Four parallel random demodulator (RD) channels



# Random Demodulator – Sensing Matrix

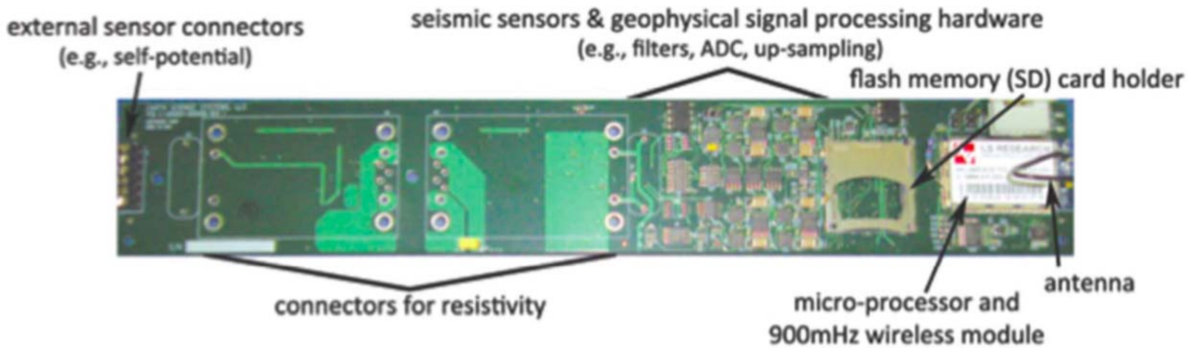
[Kirolos et al., Tropp et al.]



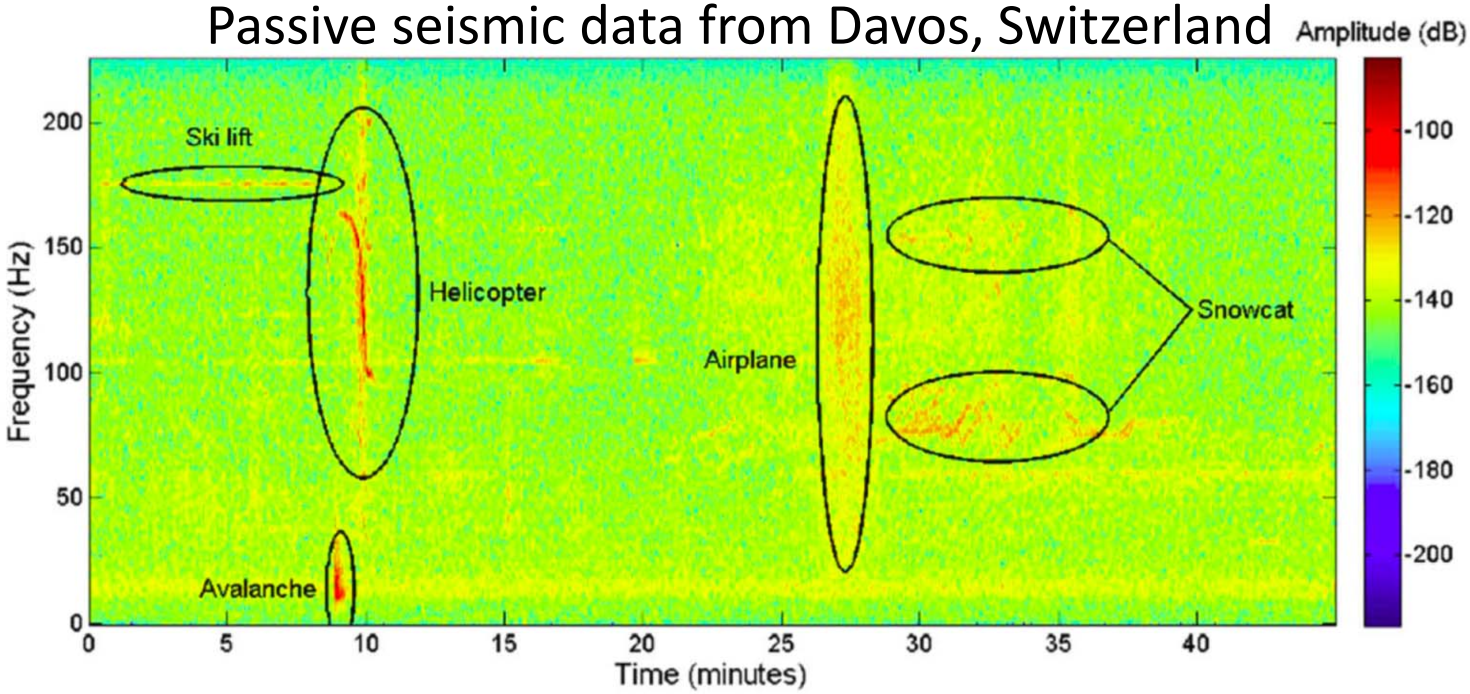
$$\Phi : \ M \times N$$

# Wireless Sensor Networks

[with Rubin and Camp]



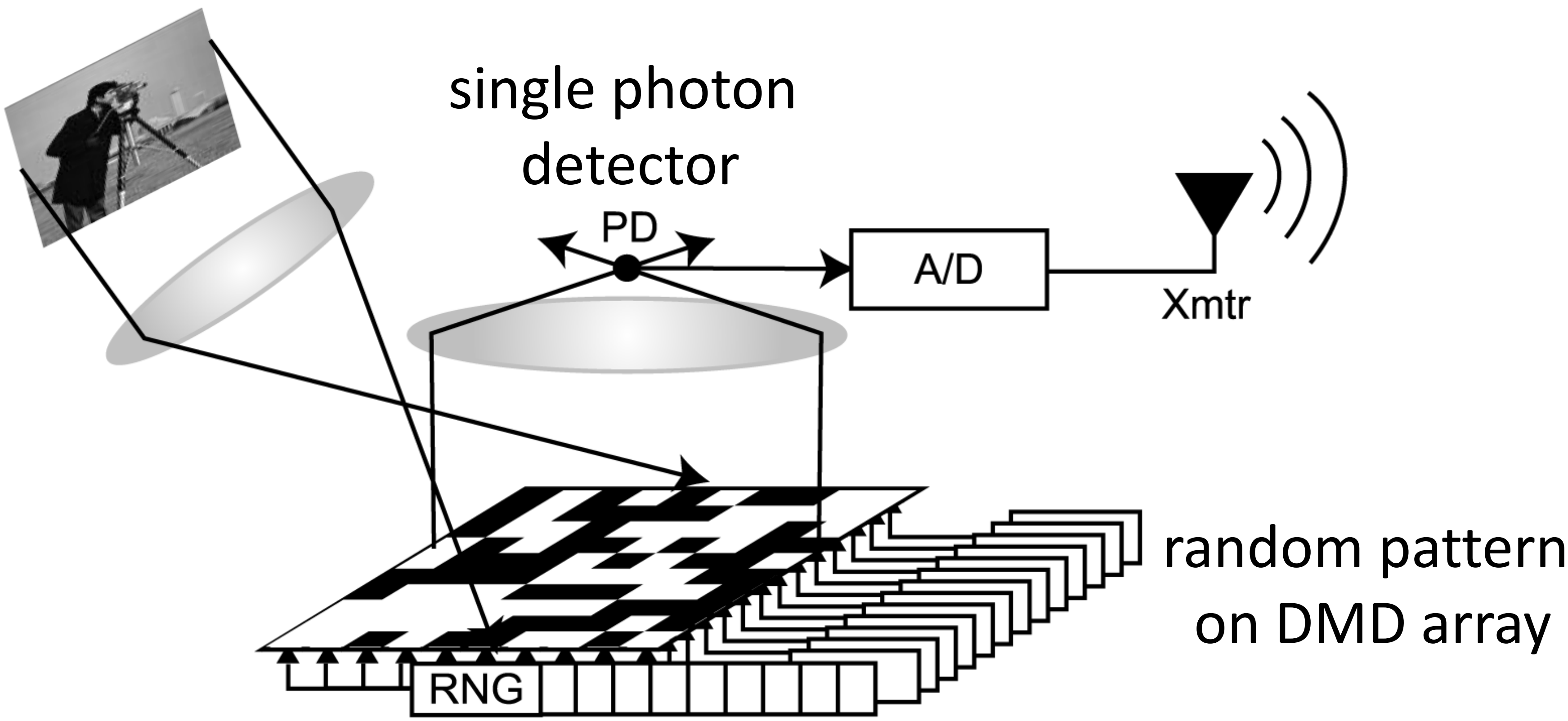

Passive seismic data from Davos, Switzerland



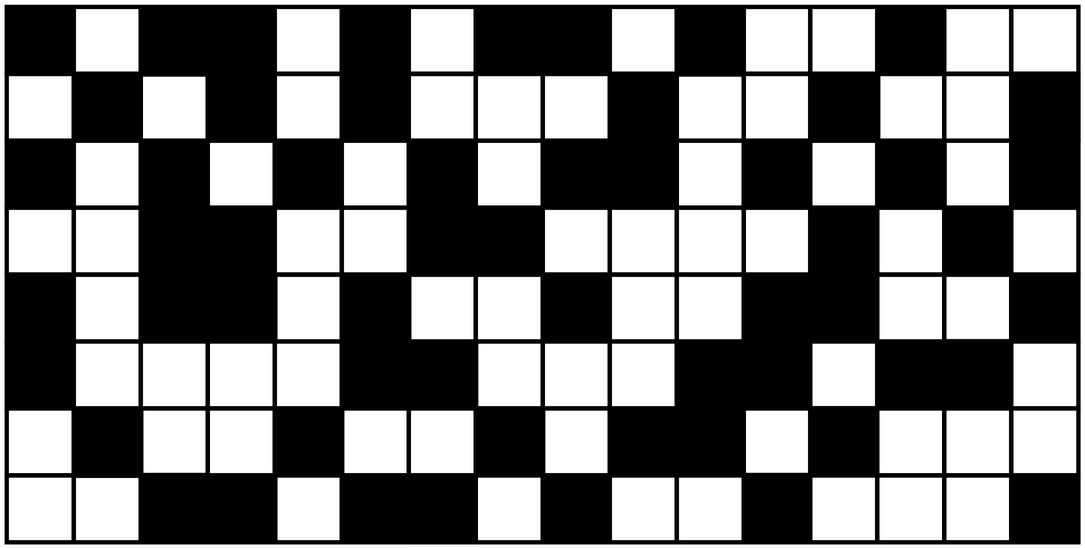

CS superior to standard compression algorithms on resource constrained motes

# Single-Pixel Camera

[Baraniuk and Kelly, et al.]



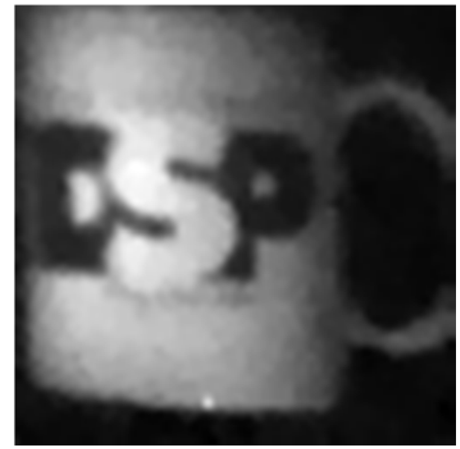
# Single-Pixel Camera – Sensing Matrix



$$\Phi : \; M \times N$$

# Single-Pixel Camera – Results

$N$ = 4096 pixels
$M$ = 1600 measurements
(40%)

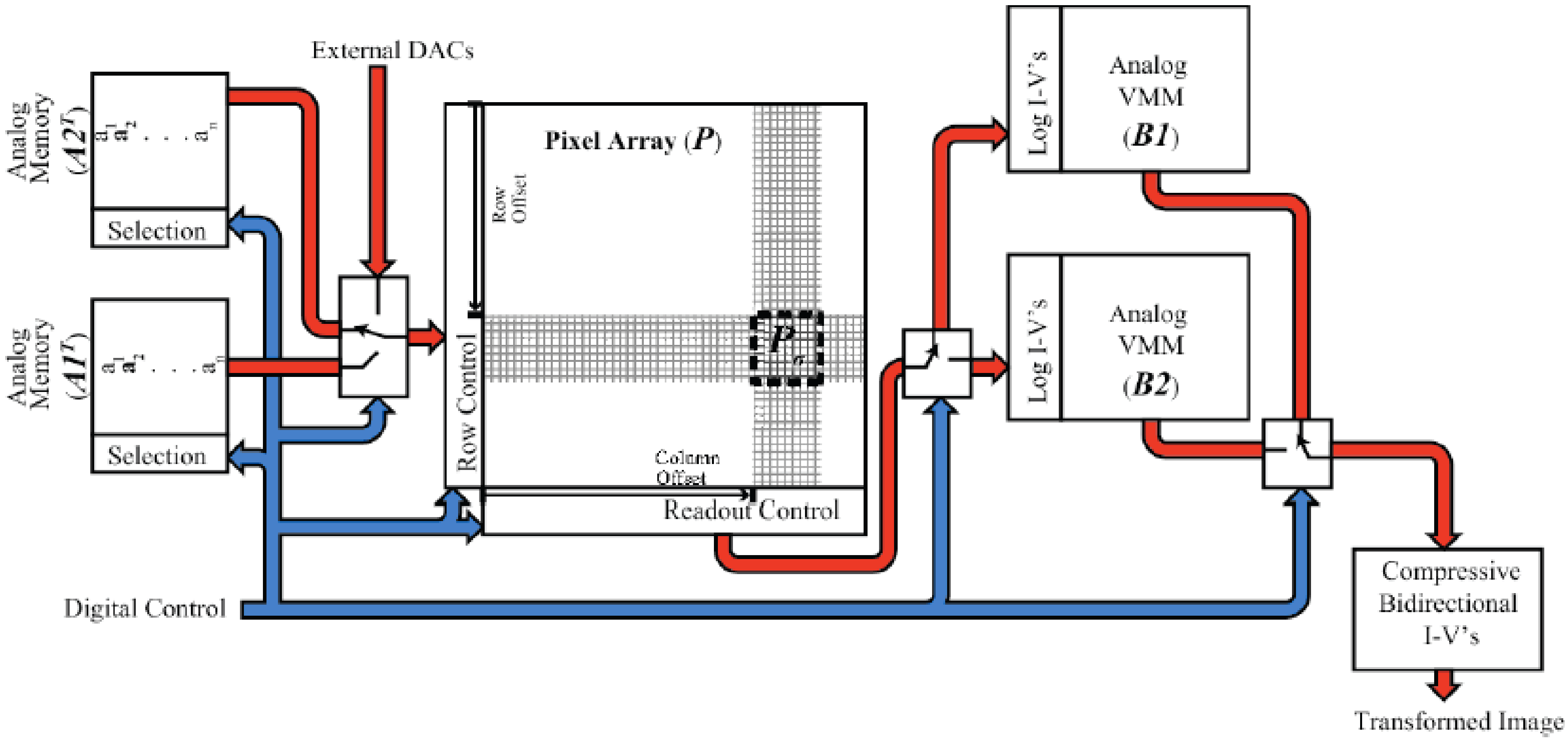


true color low-light imaging

256 x 256 image with 10:1 compression

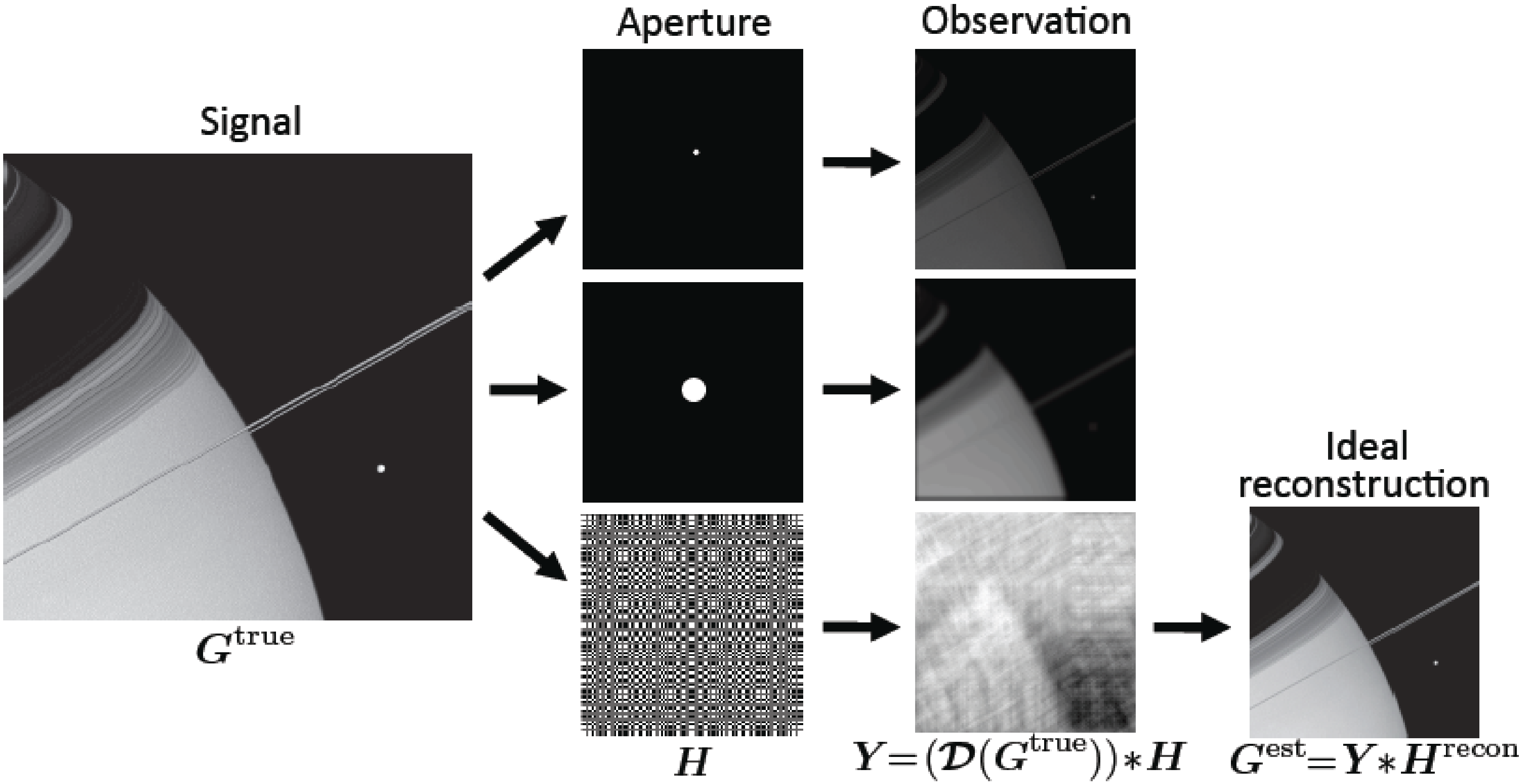
# Analog Imager

[Robucci and Hasler, et al.]



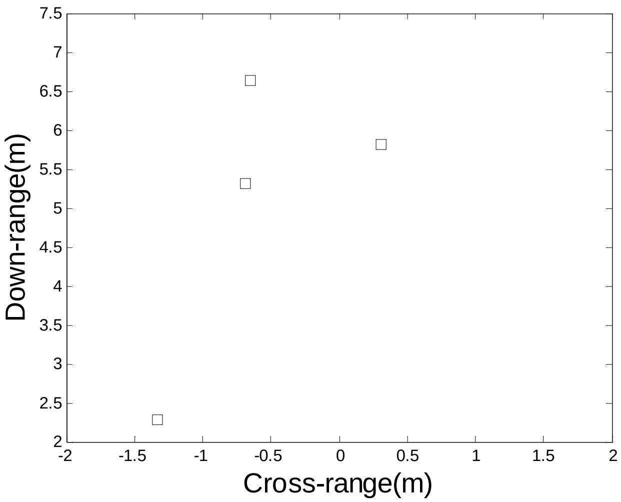
# Coded Aperture Imaging

[Marcia and Willett; Rivenson et al.; Romberg]

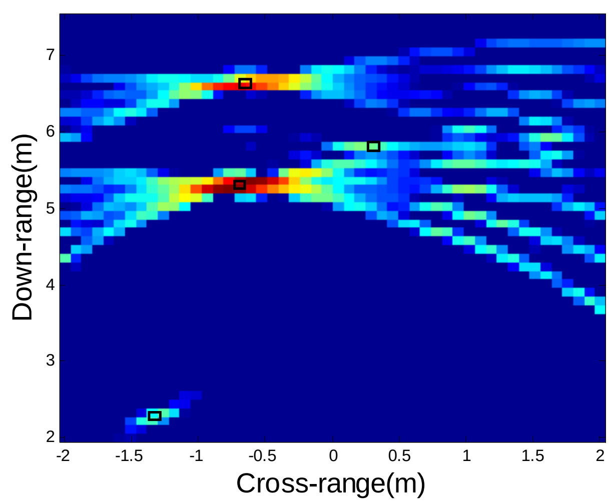


# Radar Imaging

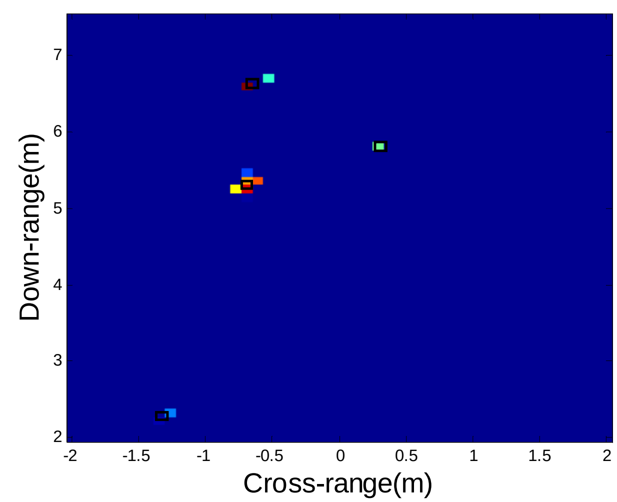
- Ground penetrating radar [Gurbuz et al.]
- Through-the-wall radar [Ahmad et al., Zhu and Wakin]




four behind-wall targets

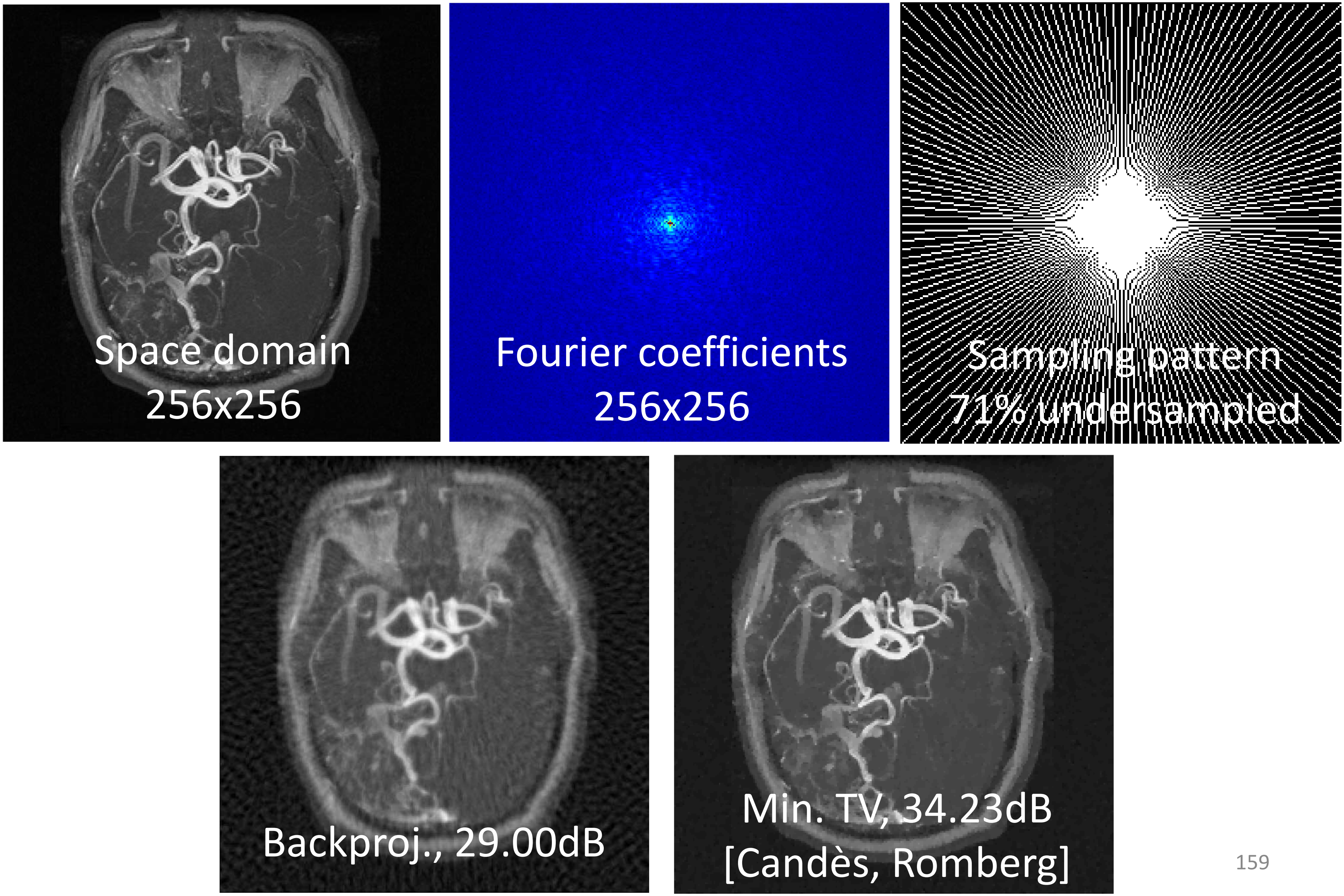



backprojection

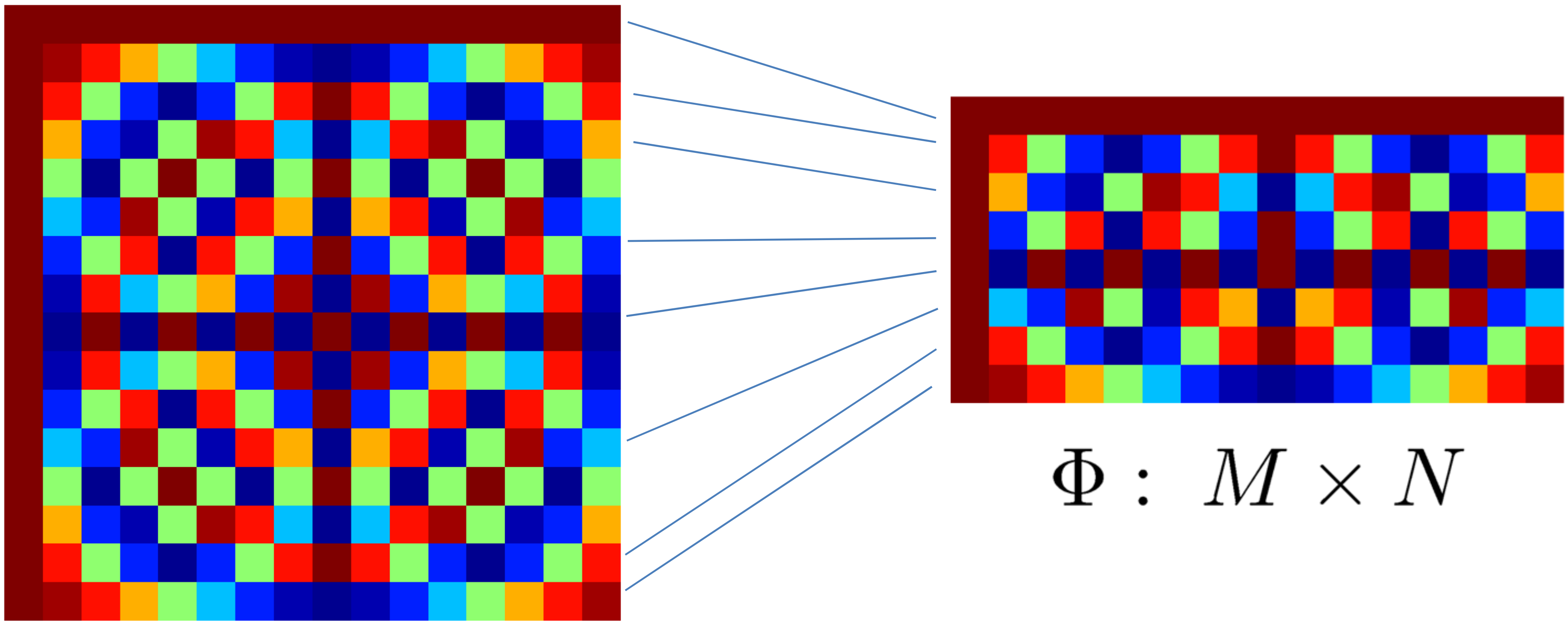



$L_1$ recovery

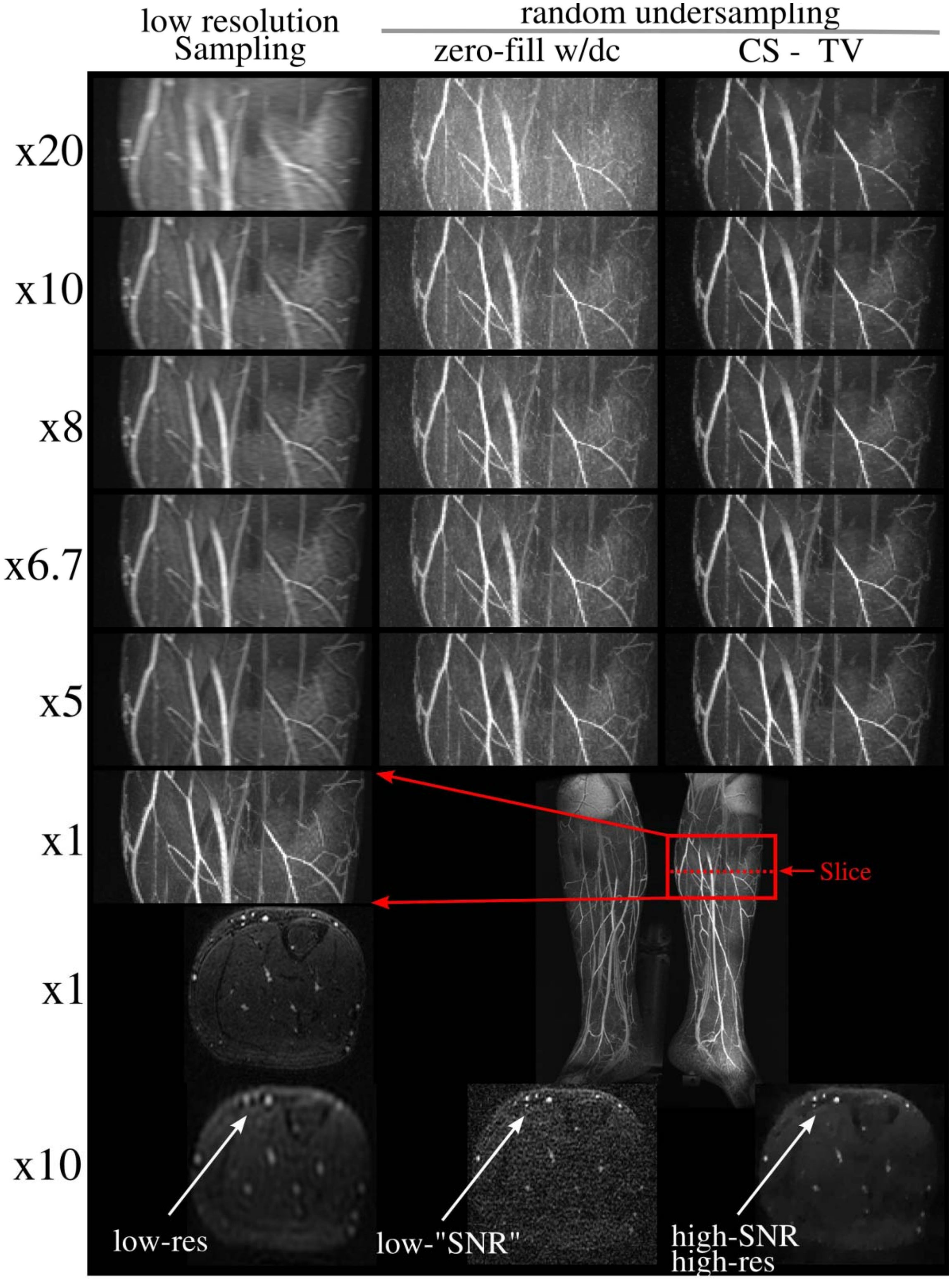
# Medical Imaging



# MRI – Sensing Matrix




$N \times N$ DFT matrix

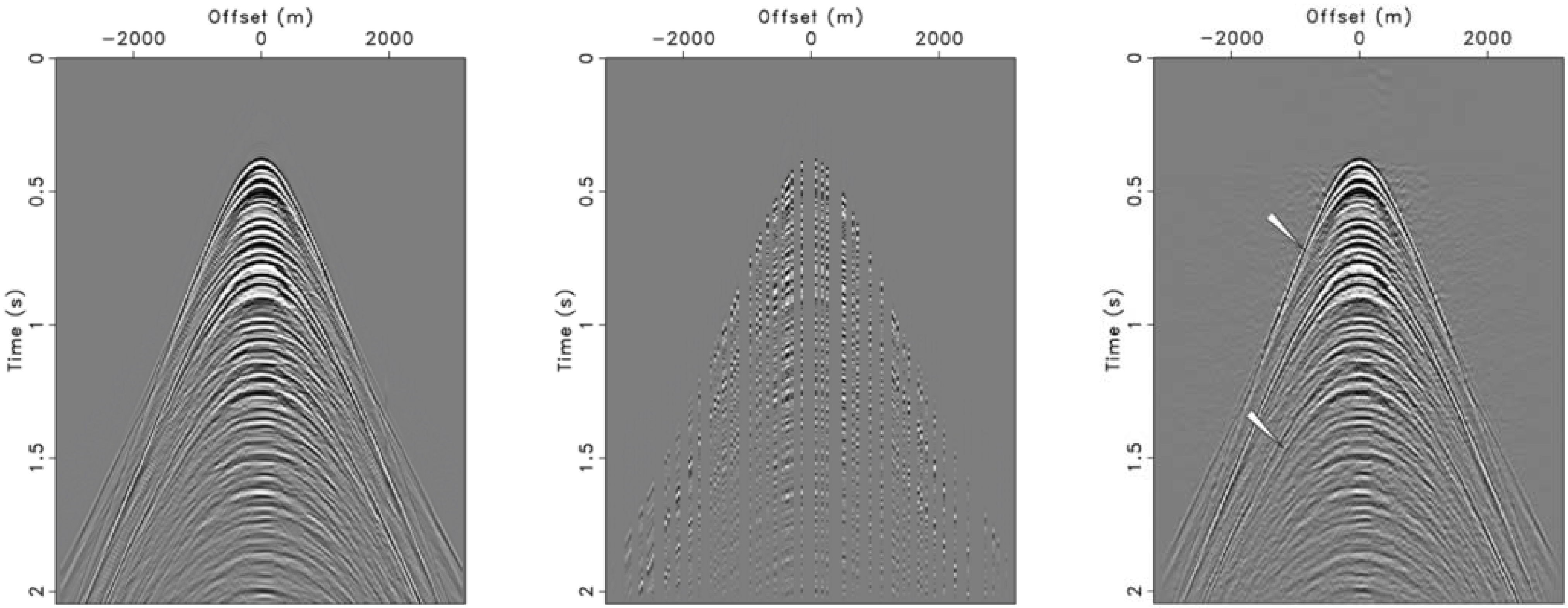



# Contrast-enhanced 3D angiography

[Lustig, Donoho, Pauly]

# Subsampling Seismic Surveys

[Herrmann et al.], [Andrade de Almeida et al.]




original

80% subsampled

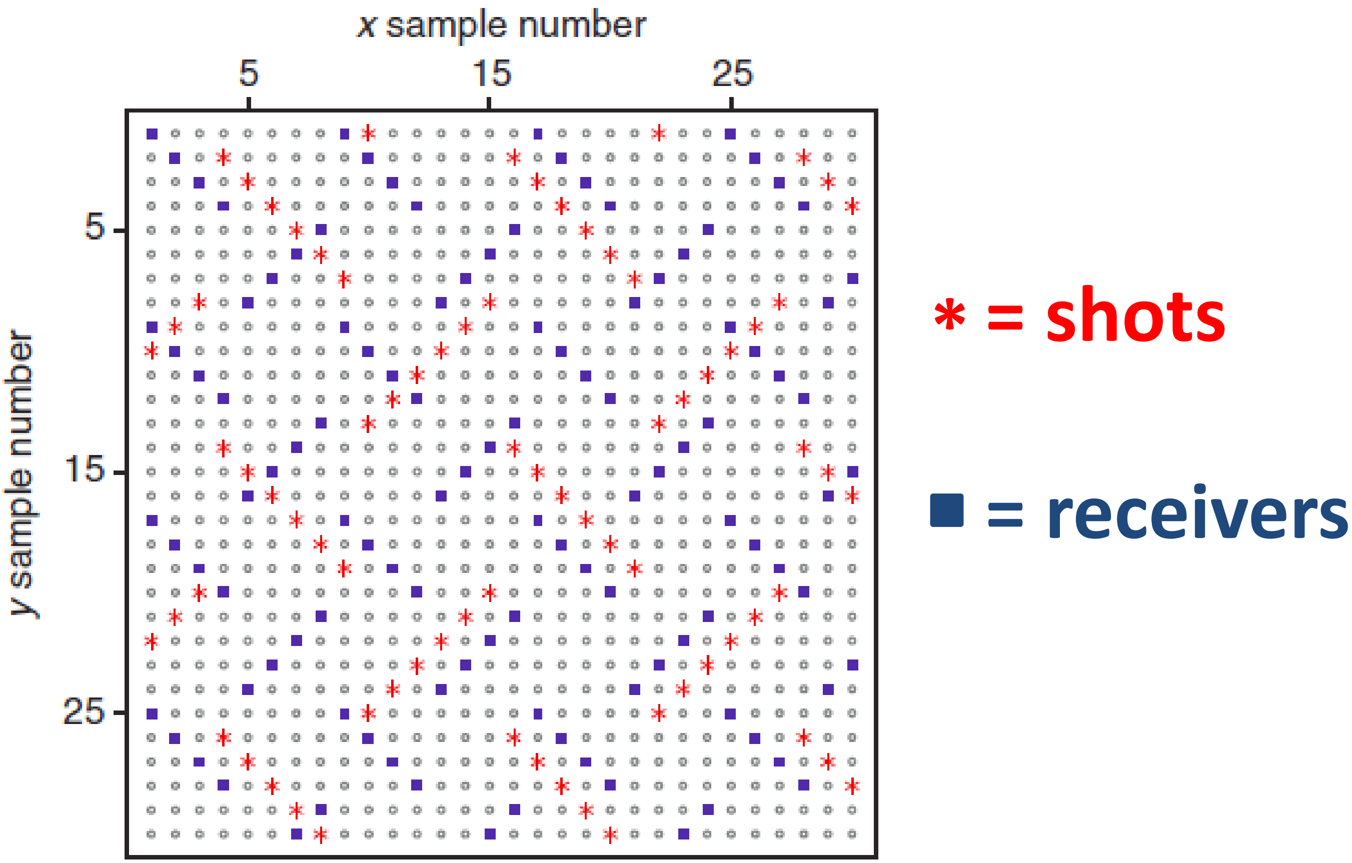
curvelet-based reconstruction

Also:

- random jittering of source locations for airgun arrays [Mansour et al.; Wason and Herrmann]
- low-rank matrix models for trace interpolation [Aravkin et al.]

# Weave Sampling

[Naghizadeh]

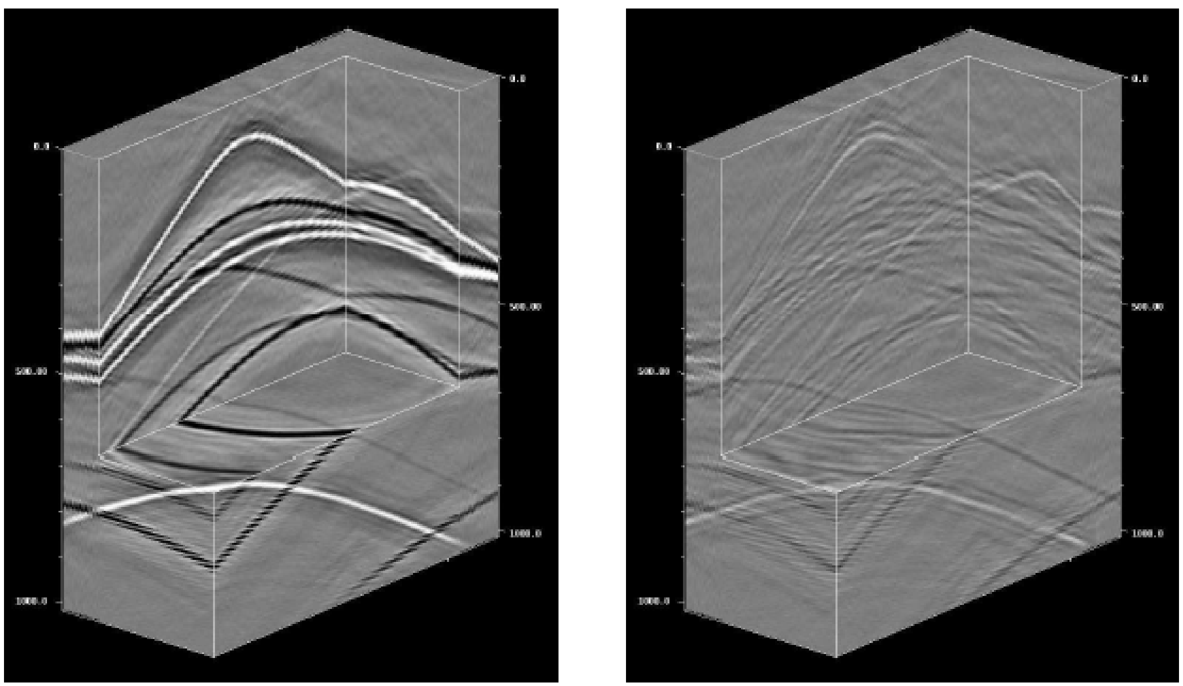



- Double-weave 3D acquisition pattern
  - structured alternative to random sampling
- Sparse reconstruction of 5D data cube

# Simultaneous Random Sources

[Neelamani et al.]

- Activate multiple sources simultaneously
  - randomly weighted impulses or noise-like waveforms
- Separate responses from random source waveforms
  - regularize using sparsity of Green's function



(a) Estimated (16x faster, SNR=11.3 dB).

(b) Estimation error (Figure 2b minus 5a)

Useful in field acquisition *or* forward modeling

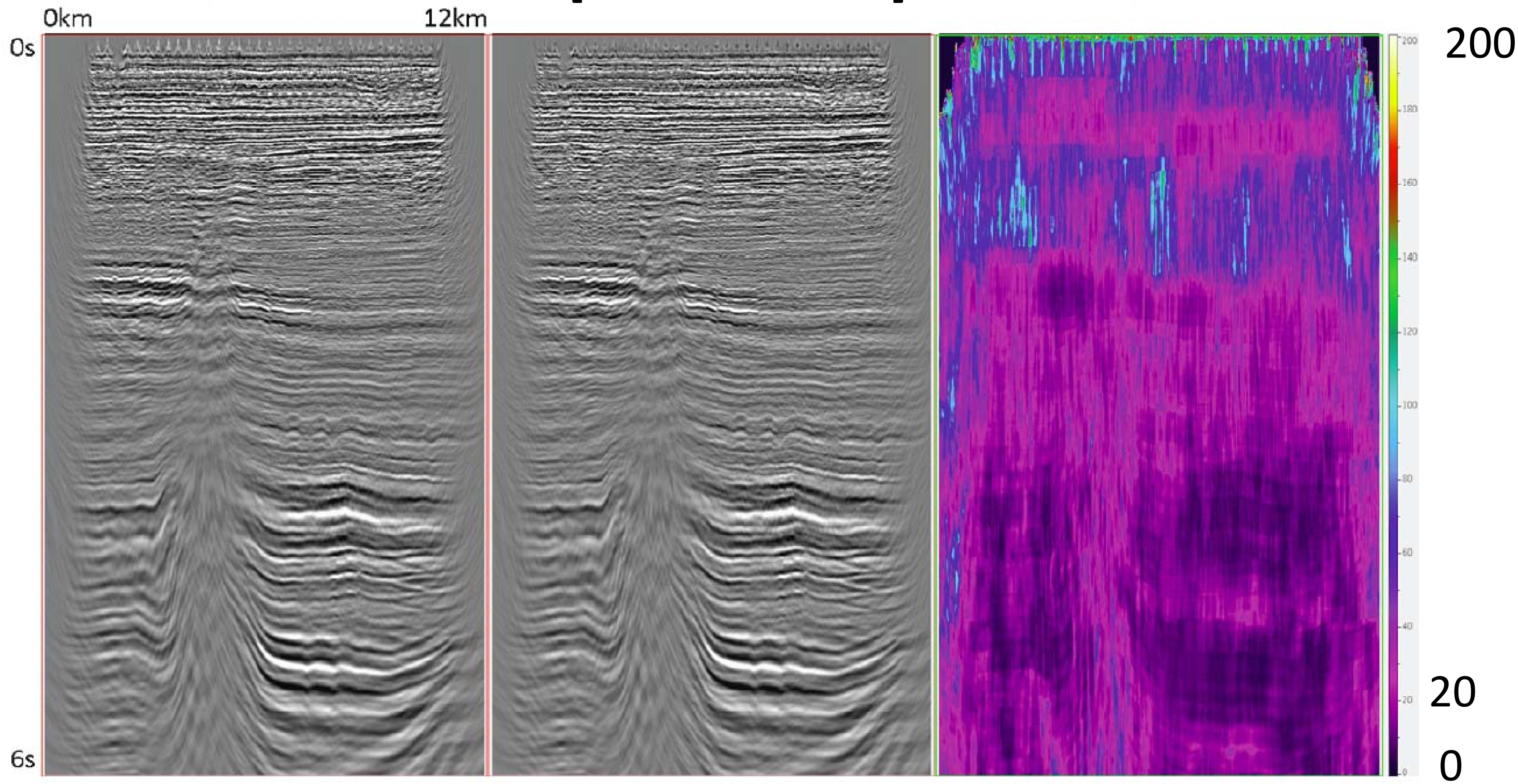
Also:

Full-waveform inversion [Herrmann et al.]

Deblending via sparsity [Abma et al.]

# Field Testing

[Mosher et al.]

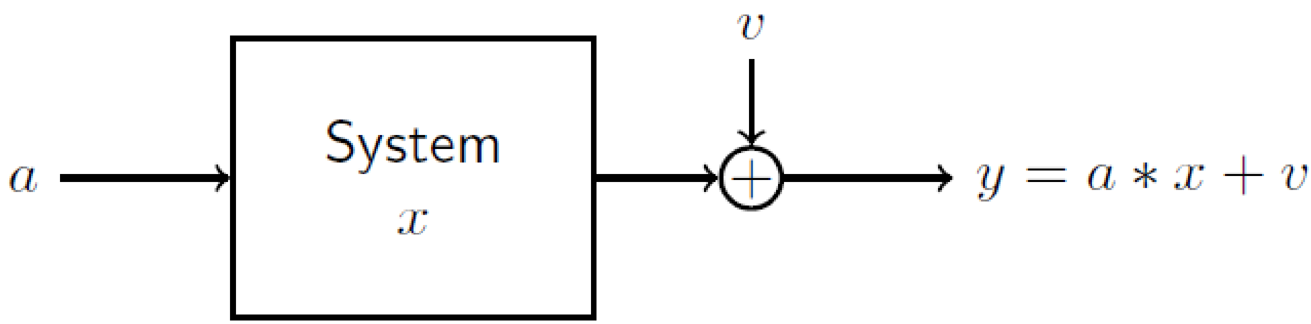



dual source non-uniform | single source uniform | NRMS difference

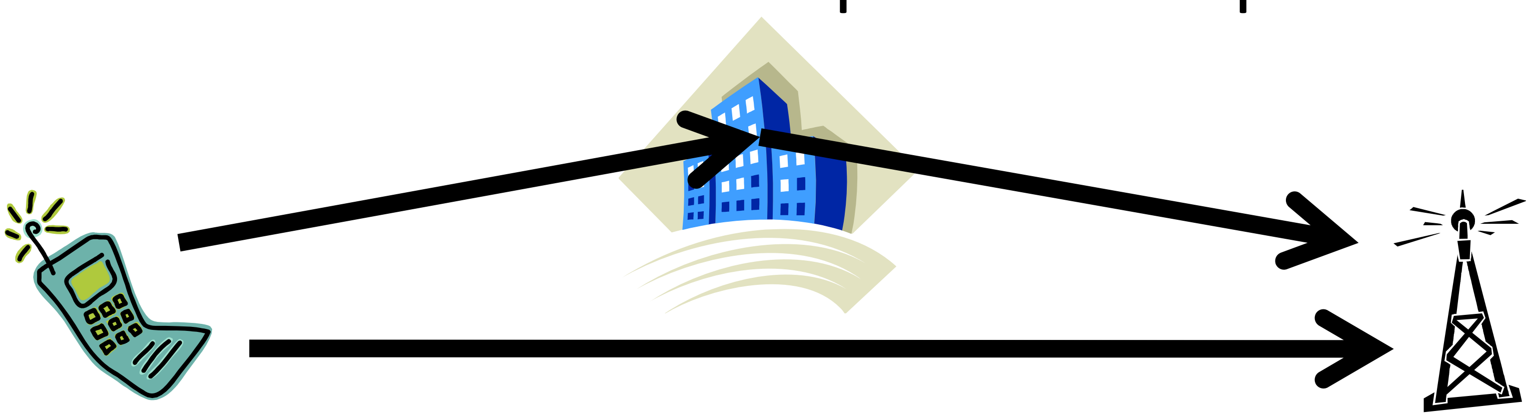
"Initial results from field trials show that shooting time can be reduced in half when two vessels are used, with data quality that meets or exceeds the quality of uniform shooting."

# System Identification

- Characterize behavior of a system by providing an input $a$ and observing the output $y$
  - control complexity by keeping $a$ and $y$ short

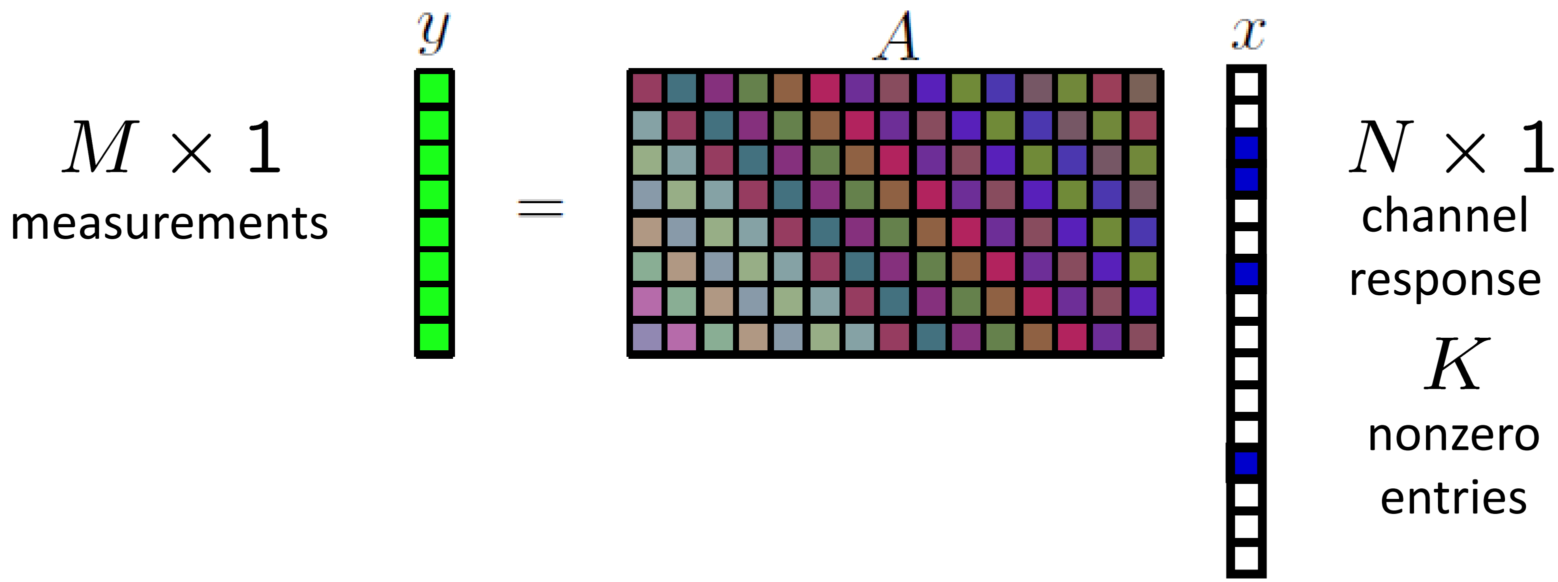



- Some systems (such as multipath wireless channels) are high-dimensional but have a sparse description



# System Identification

- Suppose we let $a$ be a random probe signal. We can write

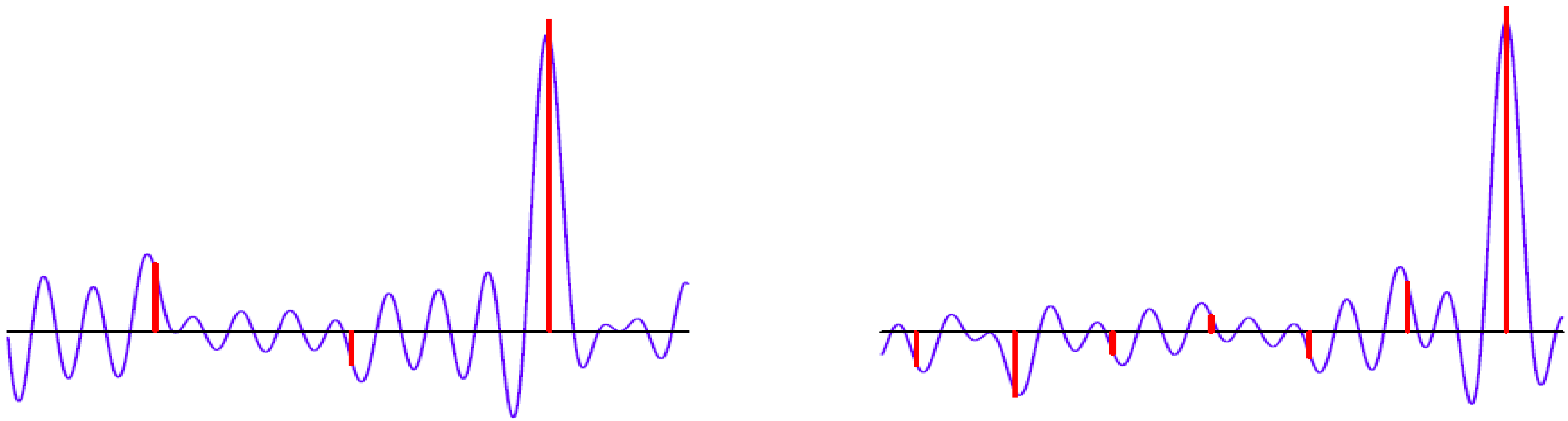



where each row of the Toeplitz matrix $A$ is a shifted copy of $a$.

- With $M$ on the order of $K$, this matrix will be favorable for CS.
  - applications in seismic imaging, wireless communications, etc.
  - [Rauhut, Romberg, Tropp; Nowak et al.; Sanandaji, Vincent, Wakin]

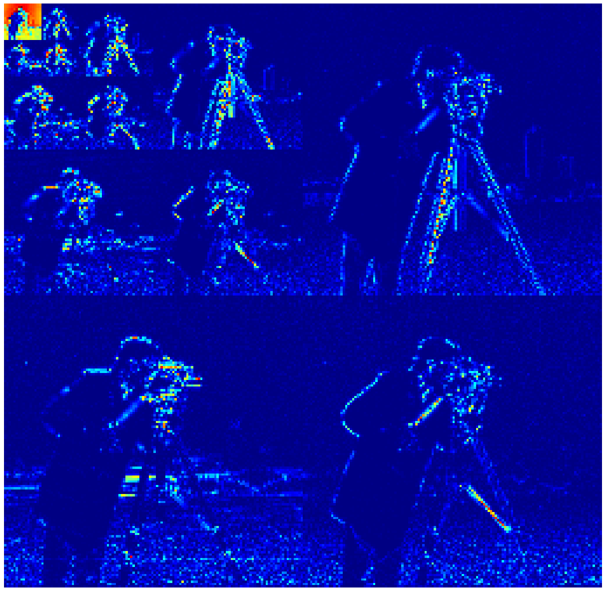
# Super-Resolution

- Goal: resolve impulses from low-frequency, non-random measurements



- Idea: solve continuous analog to $L_1$ minimization
- Guarantee: perfect resolution if no two spikes are separated by less than $2/f_c$ [Candès and Fernandez-Granda]

# Structured Sparsity

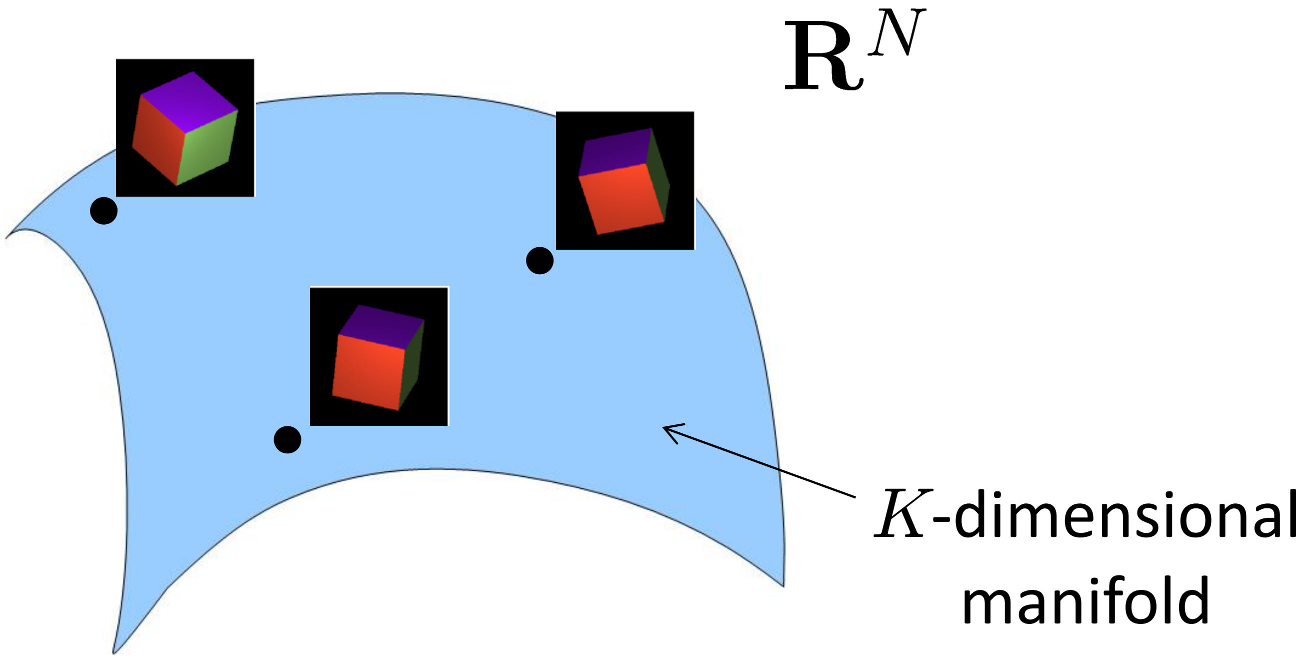


- New signal models:
  - block sparse structure
  - connectedness in wavelet tree
  - sparsity in redundant dictionaries
  - analysis vs. synthesis sparsity

- Goals:
  - capture structure in as few parameters as possible
  - develop reconstruction algorithms to exploit these models (e.g., "Model-Based CS" [Baraniuk et al.])

# Beyond Sparsity

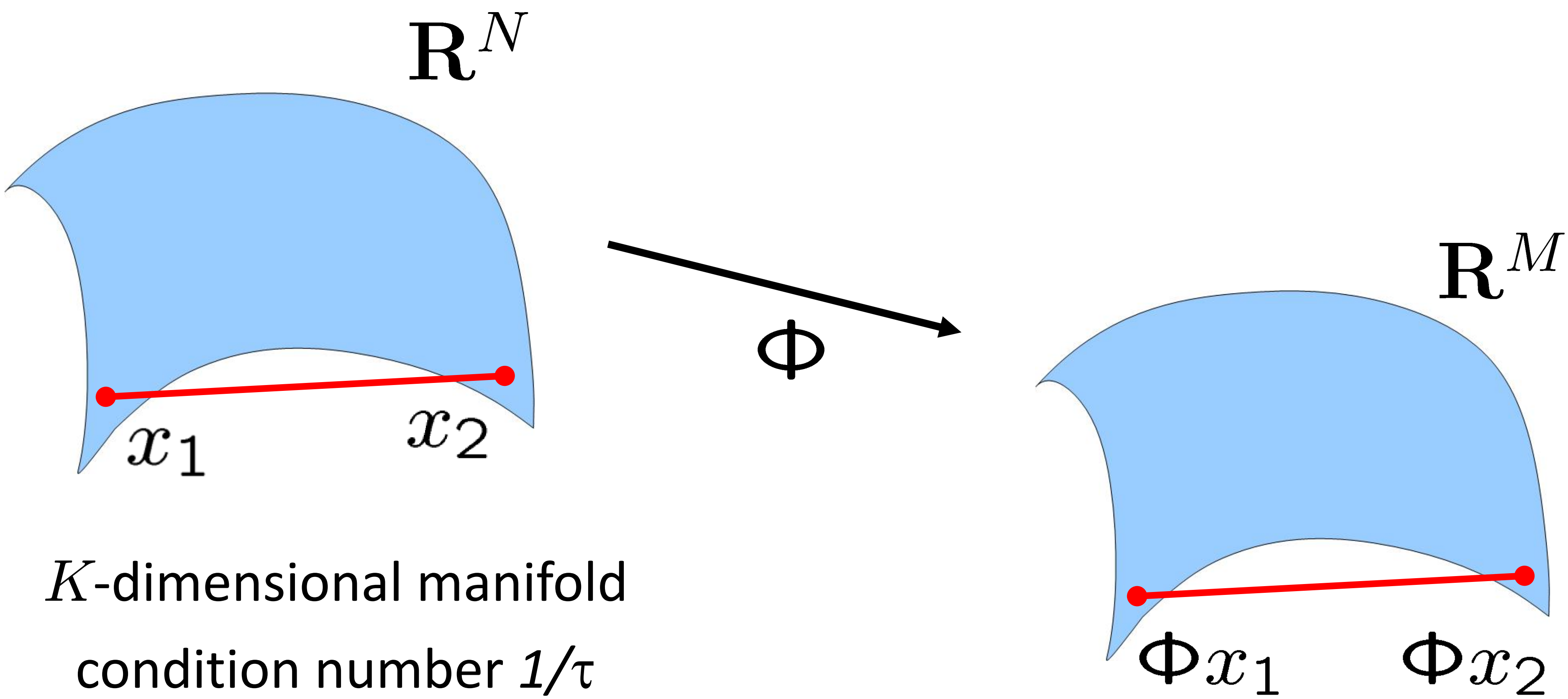
- Suppose that all uncertainty about a signal can be captured in a set of $K$ parameters




- Geometrically, the set of all possible signals under this model forms a $K$-dimensional ***manifold*** within the $N$-dimensional signal space

# Manifolds Are Stably Embedded

[with Baraniuk, Eftekhari]



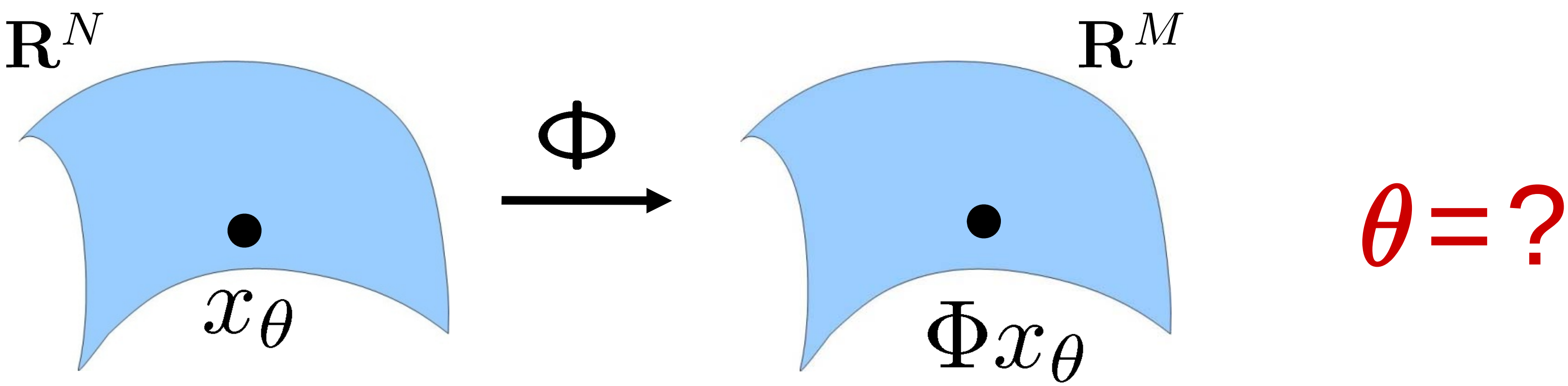

$K$-dimensional manifold

condition number *1/*$\tau$

volume $V$

$$M = O\left(K \log(NV\tau^{-1})\right)$$

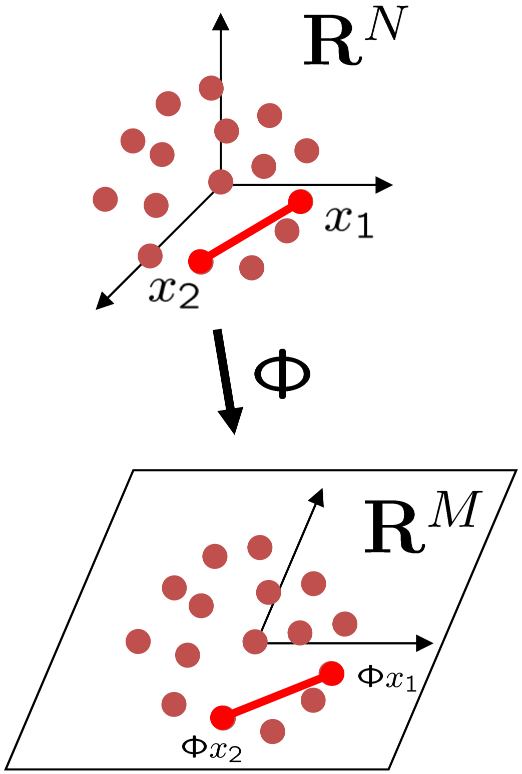
# Compressive-Domain Parameter Estimation




- Do we really need to recover the full signal?
  - Signal recovery can be demanding in high-bandwidth/high-resolution problems.
  - Is there a simpler way to extract the salient information? Perhaps with fewer measurements?
- Options for parameter estimation:
  - nearest neighbor search, grid search, iterative Newton method, Bayesian inference, specialized matched filters

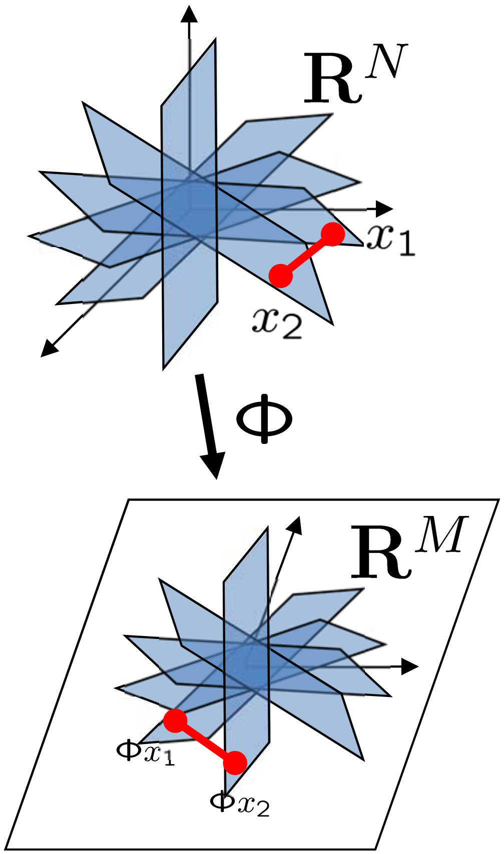
# A Variety of Stable Embeddings

*$Q$ arbitrary signals*



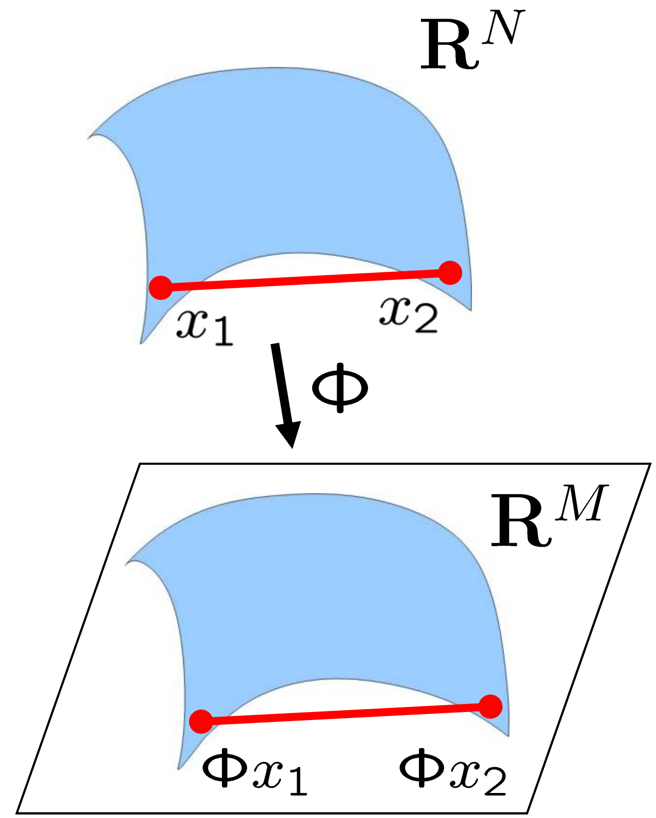

$$M = O(\log\ Q)$$

*$K$-sparse signals*




$$M = O(K \log N)$$

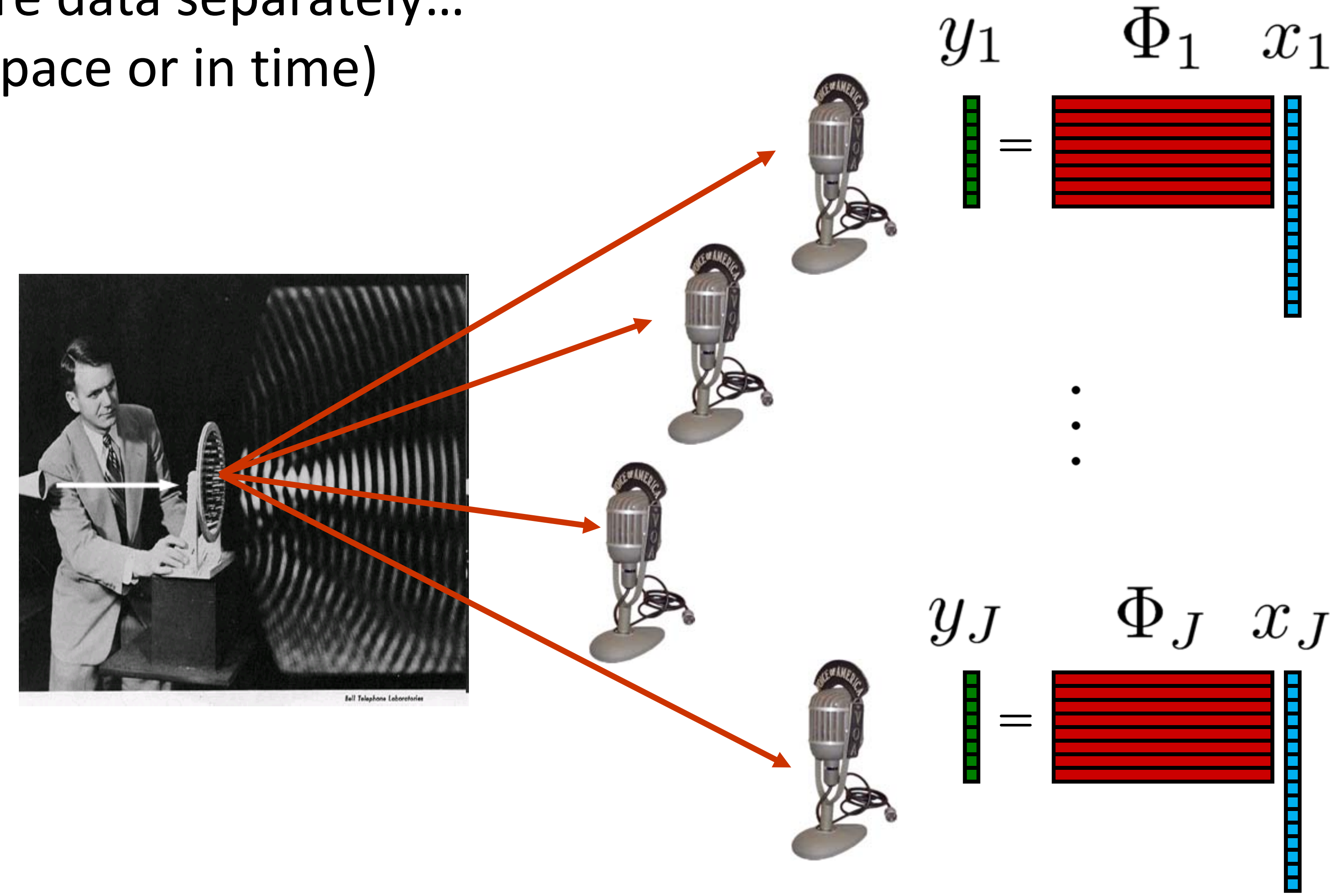
*$K$-dimensional manifold*




$$M = O(K \log N)$$

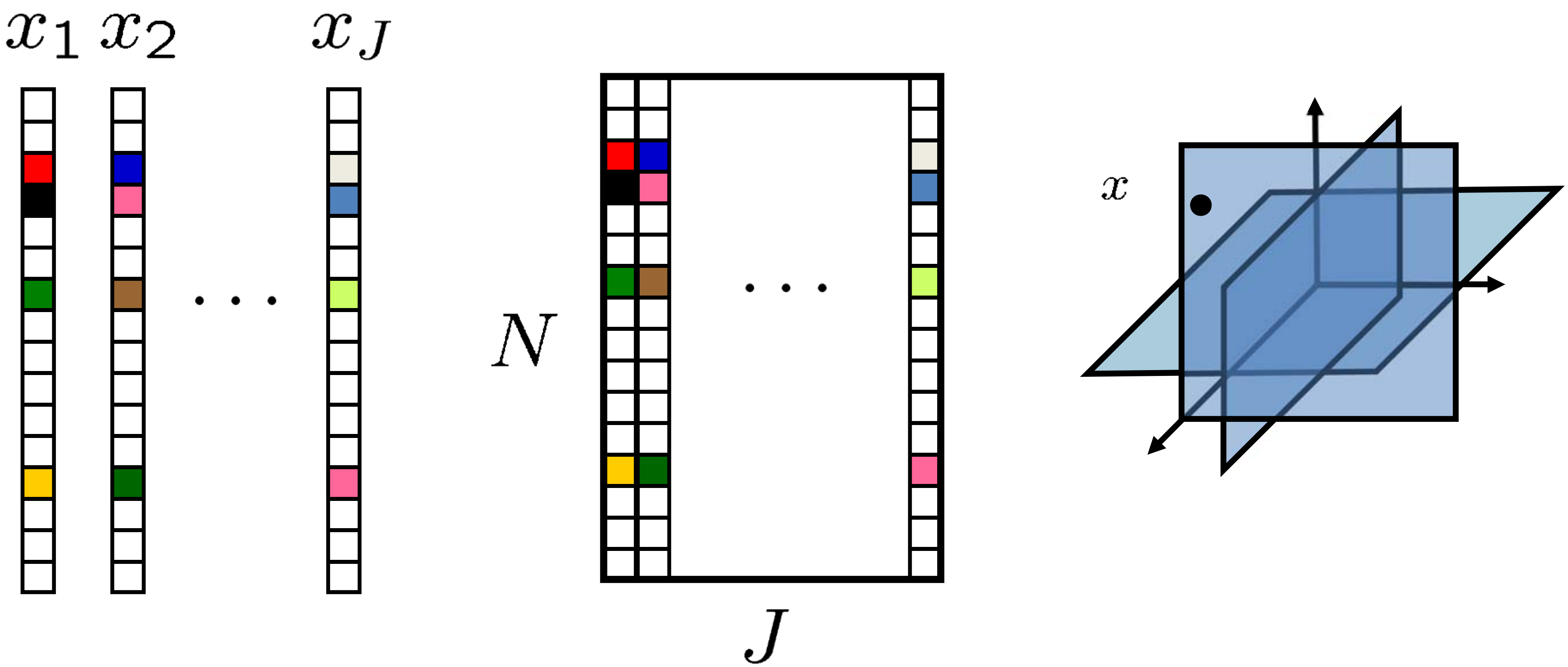
# Multi-Signal Compressive Sensing

Measure data separately…
(in space or in time)




…process jointly

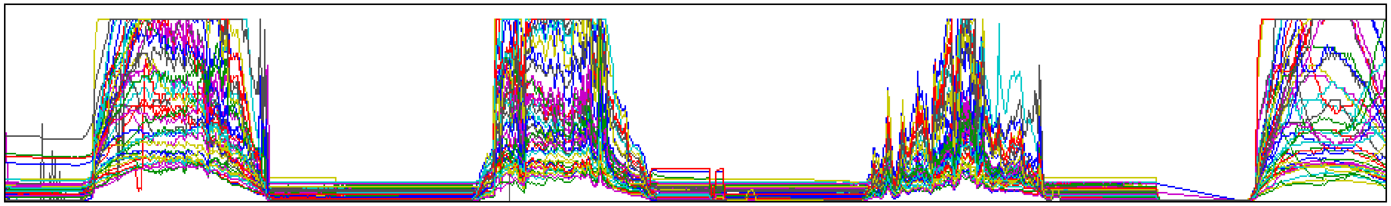
# Simultaneous Sparsity




- Greedy algorithms [Tropp et al.],[Gribonval et al.],[Duarte et al.]
- Convex optimization [Tropp],[Fornasier and Rauhut],[Eldar and Rauhut]
- Unions of subspaces and block sparsity
  [Eldar and Mishali],[Baraniuk et al.],[Blumensath and Davies]

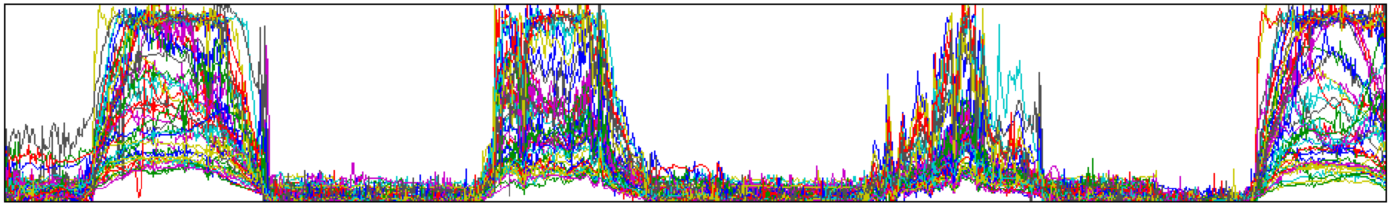
# Example

- Light sensing in Intel Berkeley Lab

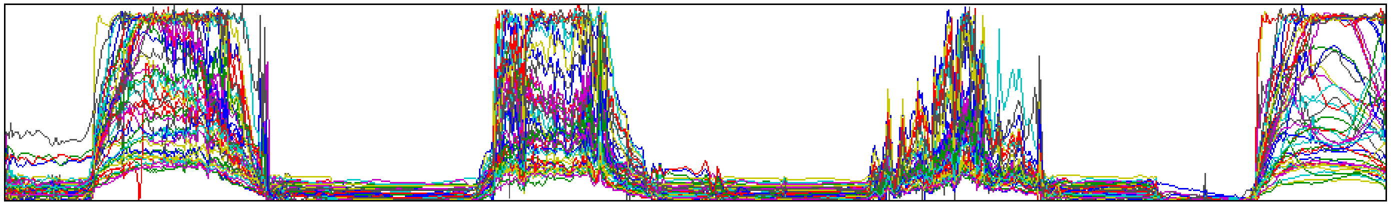


$J$ = 49 sensors
$N$ = 1024 samples

- $M$ = 400 random measurements of each signal
- Reconstruct using wavelets



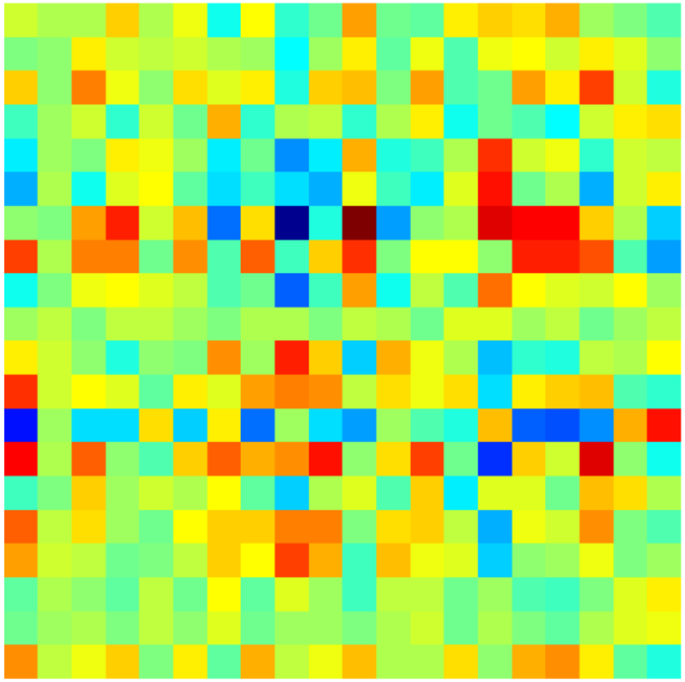
recon. separately
SNR = 21.6dB



recon. jointly
SNR = 27.2dB

# From Signals to ***Matrices***

- Many types of data naturally appear in matrix form
  - signal ensembles
  - distances between objects/sensor nodes
  - pairwise comparisons
  - user preferences ("Netflix problem")

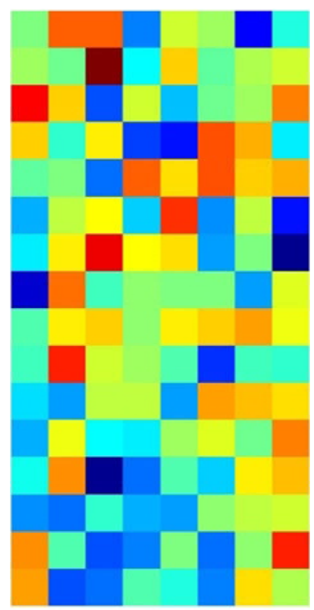
$$X =$$ 

# Sketched SVD

[with Park and Gilbert]

- Consider a data matrix $X$ of size $N \times J$ ($N \geq J$)
  - each column represents a signal/document/time series/etc.
  - recordings are distributed across $J$ nodes or sensors

$X =$

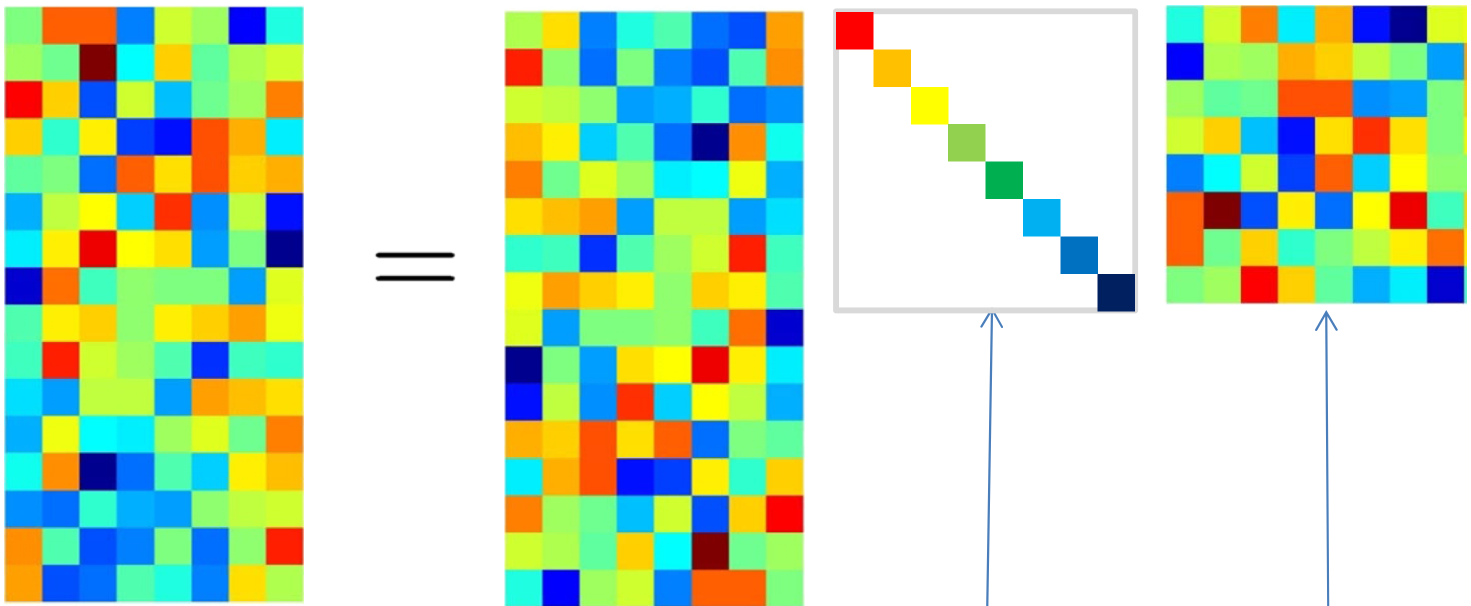


length- $N$ signals

$J$ sensors

# Singular Value Decomposition (SVD)

$$X = U_X \Sigma_X V_X^T$$



orthonormal columns

principal column directions

diagonal, positive

relative energies

orthonormal rows

principal row directions

# Spectral Analysis

- SVD of $X$:

$$X = U_X \Sigma_X V_X^T$$

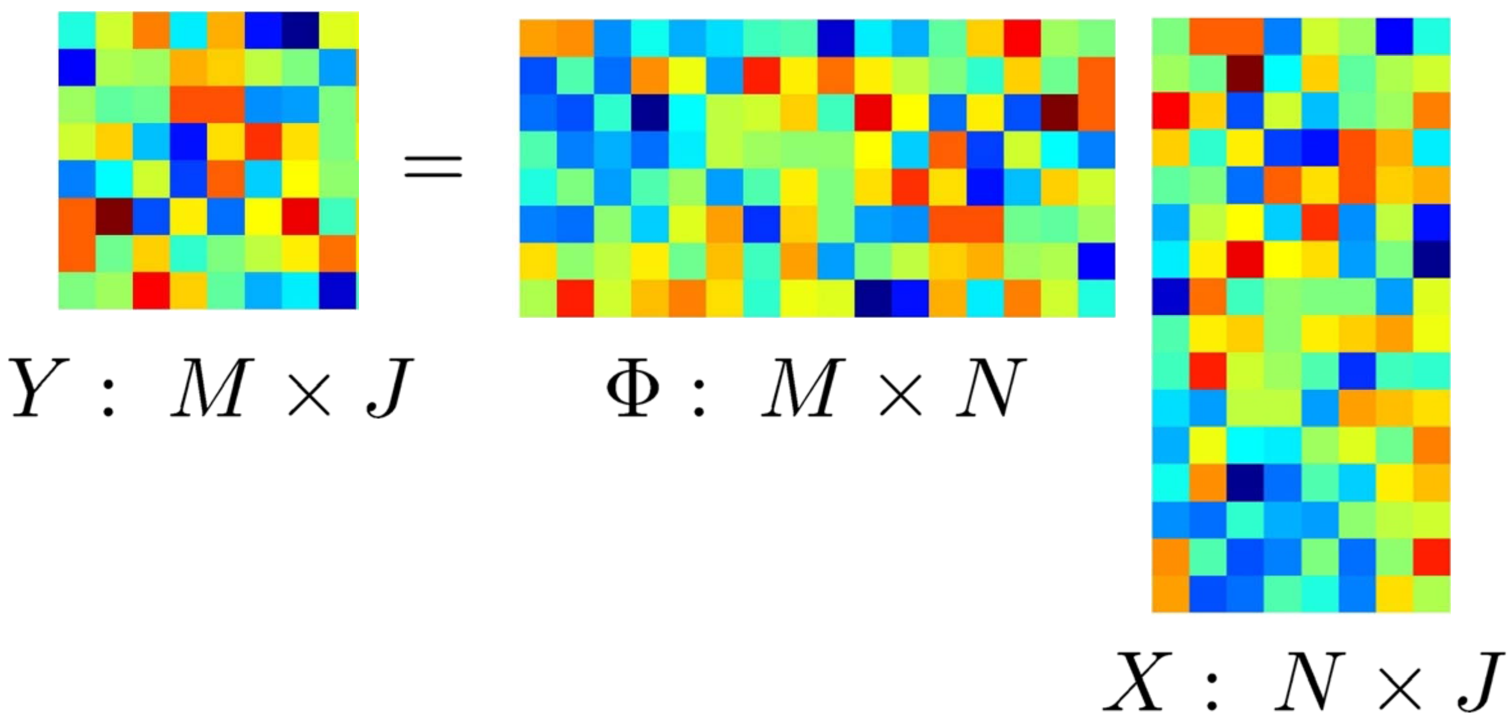
- Our interest: $\Sigma_X$ and $V_X$, from which we can obtain
  - principal directions of rows of $X$ (but not columns)
  - KL transform: inter-signal correlations (but not intra-signal)
  - stable, low-dimensional embedding of data vectors via $\Sigma_X V_X^T$

**Challenge:**

Obtaining $X$ and computing SVD($X$) when $N$ is large.

# Sketching

- Data matrix $X$ of size $N \times J$ ($N \geq J$)
- Construct random $M \times N$ sketching matrix $\Phi$
- Collect a one-sided sketch $Y = \Phi X$
  - can be obtained column-by-column (“sensor-by-sensor”)
  - easily updated dynamically if $X$ changes



# Sketched SVD

- Sketched matrix of size $M \times J$:

$$Y = \Phi X = \Phi U_X \Sigma_X V_X^T$$

- We simply compute the SVD of $Y$:

$$Y = U_Y \Sigma_Y V_Y^T$$

- Suppose $X$ is rank $R$ for some small $R$. If

$$M = O(R\epsilon^{-2})$$

then with high probability, $\Sigma_Y \approx \Sigma_X$ and $V_Y \approx V_X$.

# Sketched SVD

- More formally, for $j$ = 1,2,…, $R$,
  - singular values are preserved
    [Magen and Zouzias]

$$(1-\epsilon)^{1/2} \leq \frac{\sigma_j(Y)}{\sigma_j(X)} \leq (1+\epsilon)^{1/2}$$

  - right singular vectors are preserved

$$\|v_j(X) - v_j(Y)\|_2 \leq \frac{\epsilon\sqrt{1+\epsilon}}{\sqrt{1-\epsilon}} \underbrace{\max_{i \neq j} \frac{\sqrt{2}\sigma_i(X)\sigma_j(X)}{\min\limits_{c \in [-1,1]} \{|\sigma_i^2(X) - \sigma_j^2(X) \cdot (1+c\epsilon)|\}}}_{\text{small if } \sigma_j(X) \text{ is well separated from other singular values of } X}$$

roughly $\epsilon$

small if $\sigma_j(X)$ is well separated from other singular values of $X$

# Related Work: Randomized Linear Algebra

- Compressive PCA [Fowler], [Qi and Hughes]
  - interested in *left* singular vectors rather than right
  - different aspect ratio for data matrix
  - utilize different random projections for different columns

- Randomized techniques for low-rank matrix approximation [Rokhlin et al.], [Feldman et al.], [Halko et al.]
  - focused on *subspaces* and matrix approximations rather than individual singular vectors
  - can require multiple passes over data matrix
  - theme: randomness to accelerate ***computation***

# Structural Health Monitoring

- Automated monitoring of buildings, bridges, etc.

I-35W Mississippi River Bridge (2007)

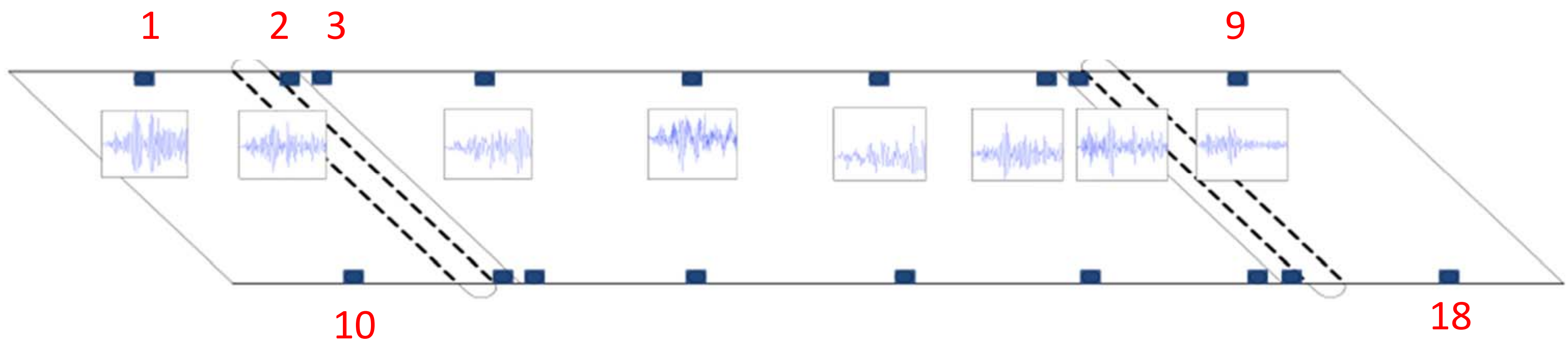
Sampoong Department Store (1995)

- Wireless sensors
  - acquire vibration data, transmit to central node
  - **goal:** maximize battery life and accurately assess health

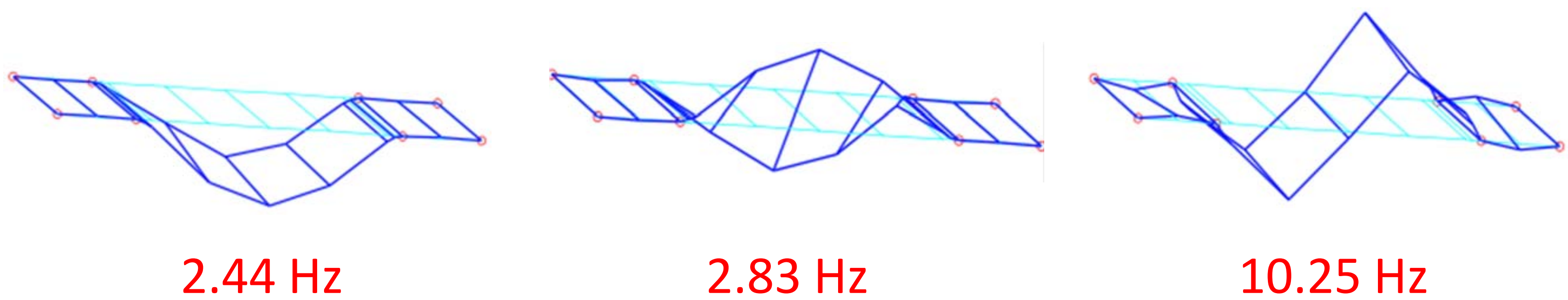
# CS for Modal Analysis

[with Park and Gilbert]

- Undersample or compress vibration recordings from sensors

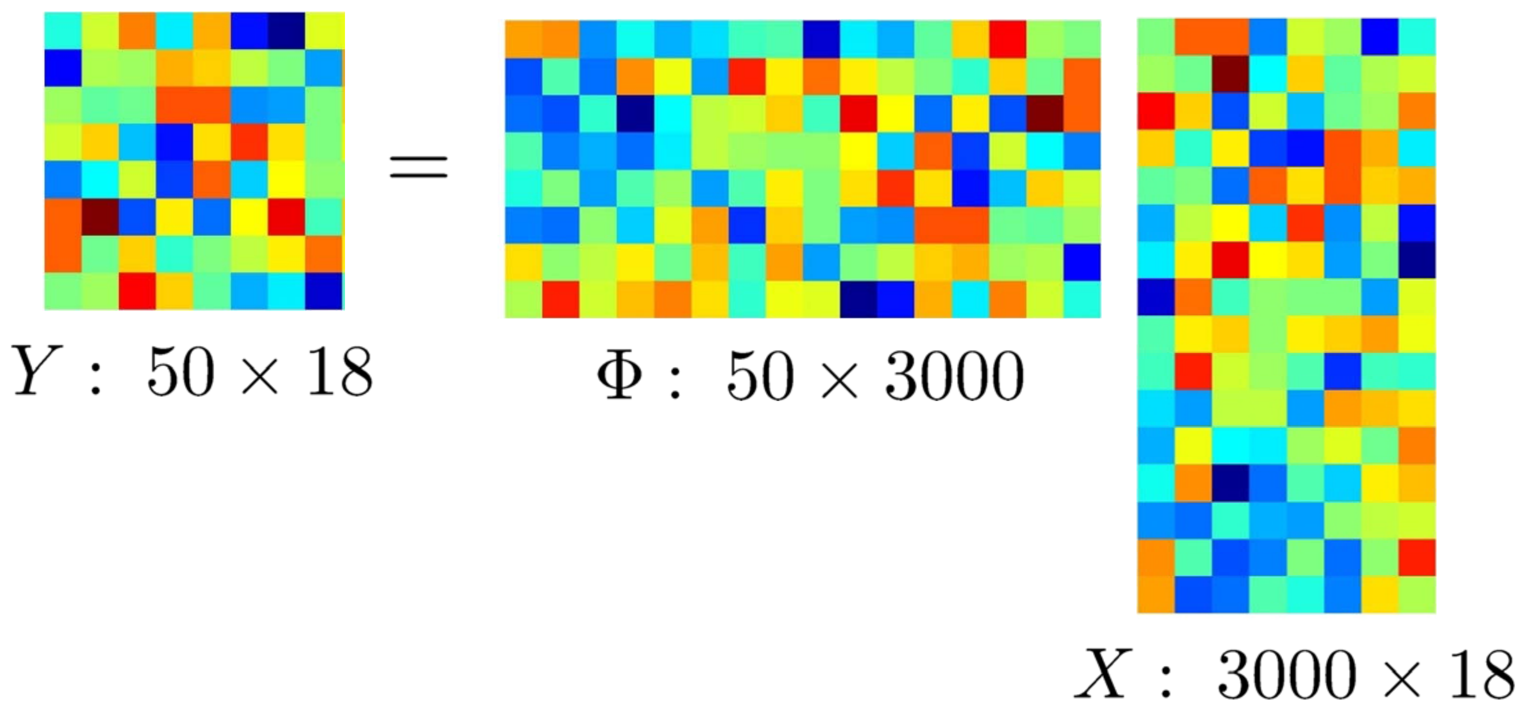



- Estimate vibrational modes directly from compressed data—without reconstruction



# CS for Modal Analysis - Results

- $J$ = 18 sensors recording vibration data
- 3 dominant mode shapes in this data set
- $N$ = 3000 samples from each sensor
- $M$ = 50 random Gaussian measurements per sensor



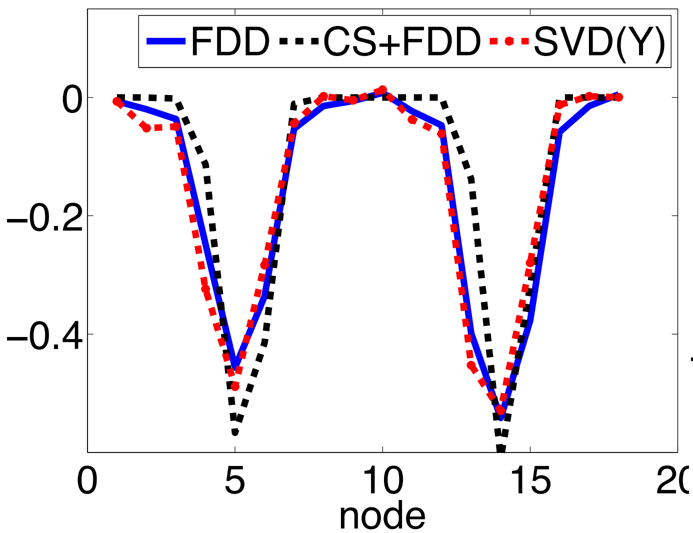

- Process measurements $Y$ centrally to estimate mode shapes

# CS for Modal Analysis - Results

FDD: popular modal analysis algorithm

CS+FDD: reconstruct each signal, then pass through FDD

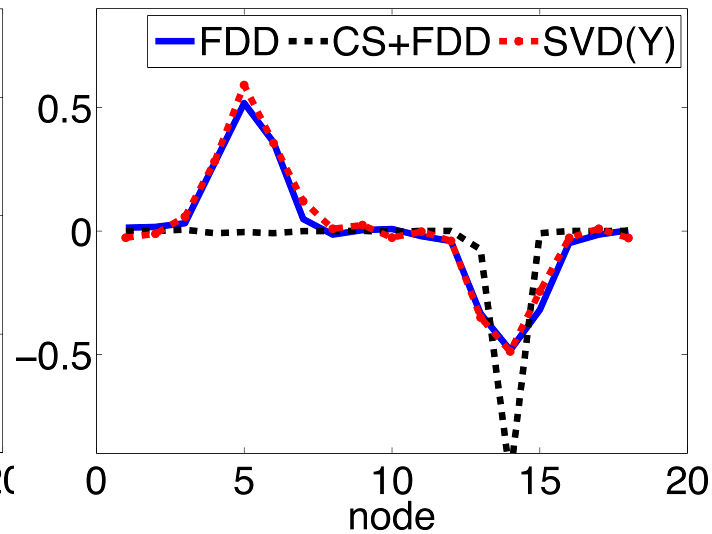
SVD($Y$): proposed method




$$\|\{\psi_1\} - \{\psi_1^{'}\}\|_2$$

CS+FDD=0.35

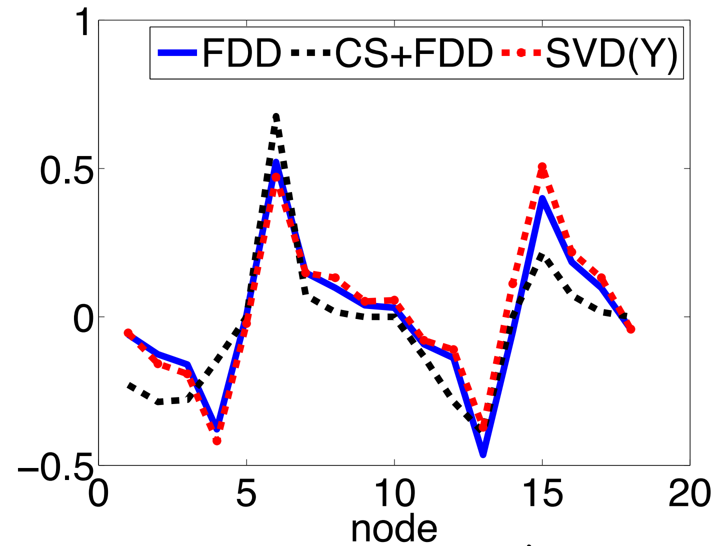
SVD($Y$)=0.16




$$\|\{\psi_2\} - \{\psi_2^{'}\}\|_2$$

CS+FDD=0.96

SVD($Y$)=0.14



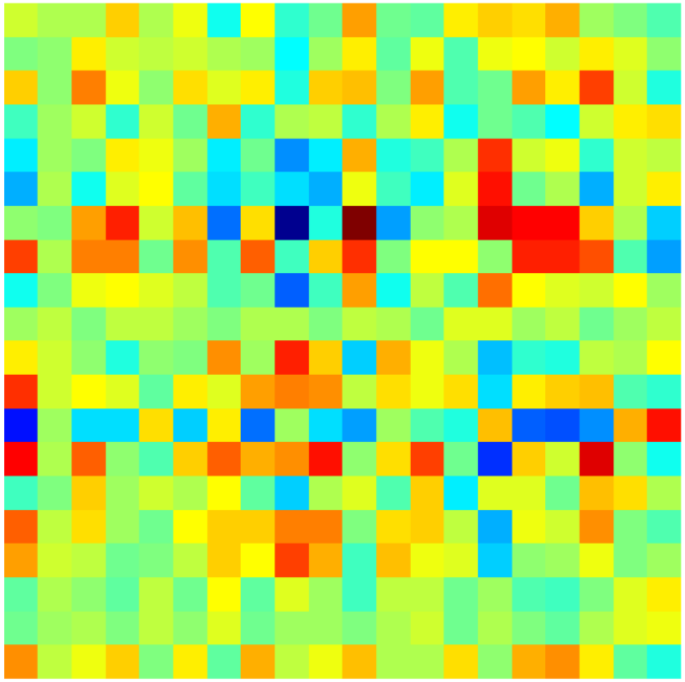

$$\|\{\psi_3\} - \{\psi_3^{'}\}\|_2$$

CS+FDD=0.50

SVD($Y$)=0.19

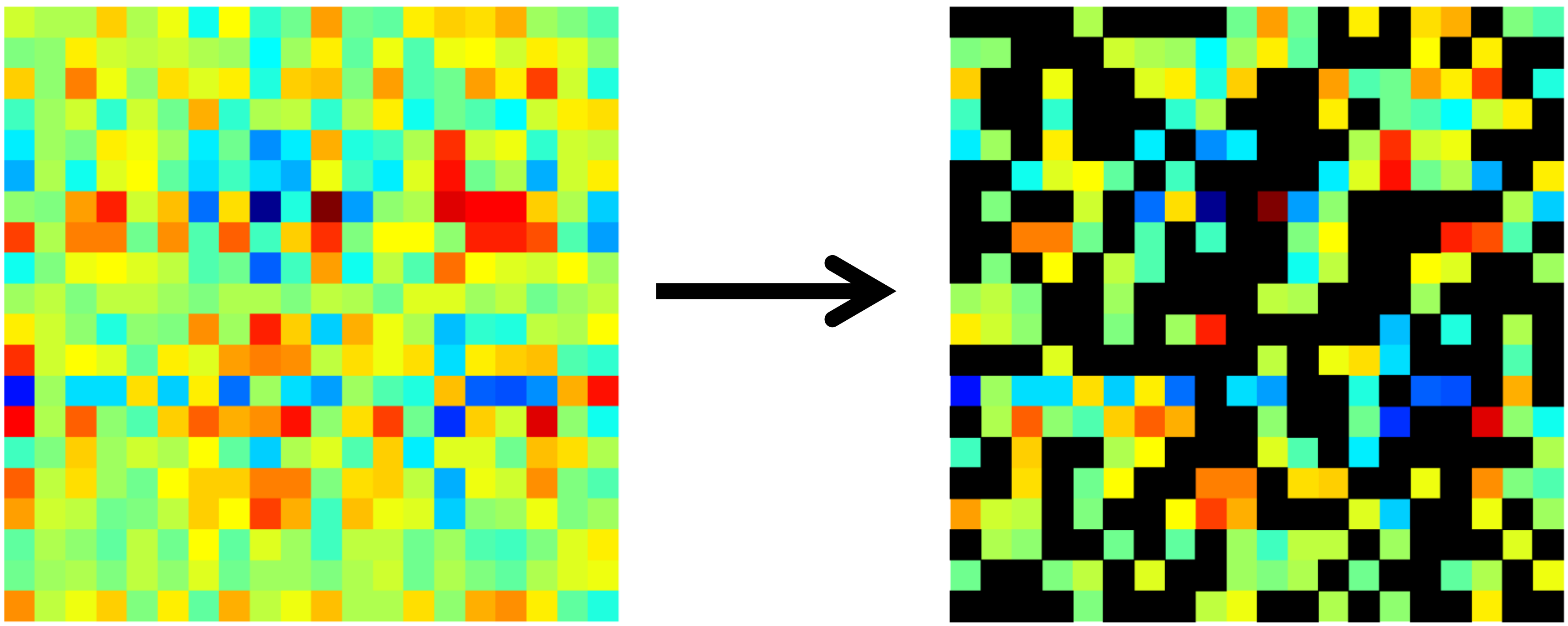
# Matrix Completion

- Many types of data naturally appear in matrix form
  - signal ensembles
  - distances between objects/sensor nodes
  - pairwise comparisons
  - user preferences ("Netflix problem")

$$X =$$ 

# Missing Data

- Often this data may be incomplete

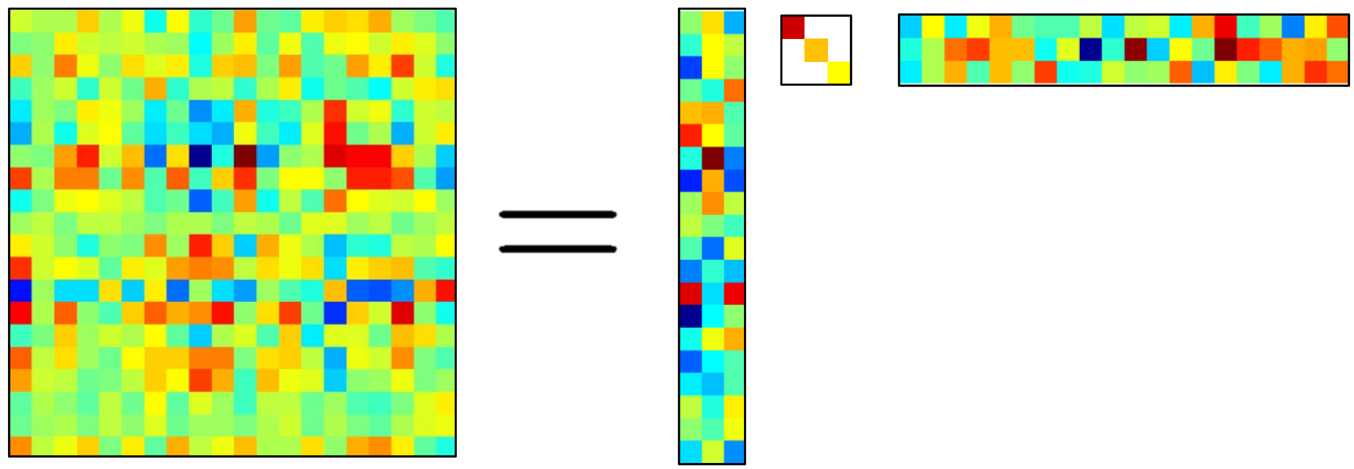


- How can we fill in the blanks?

# Low-Complexity Model

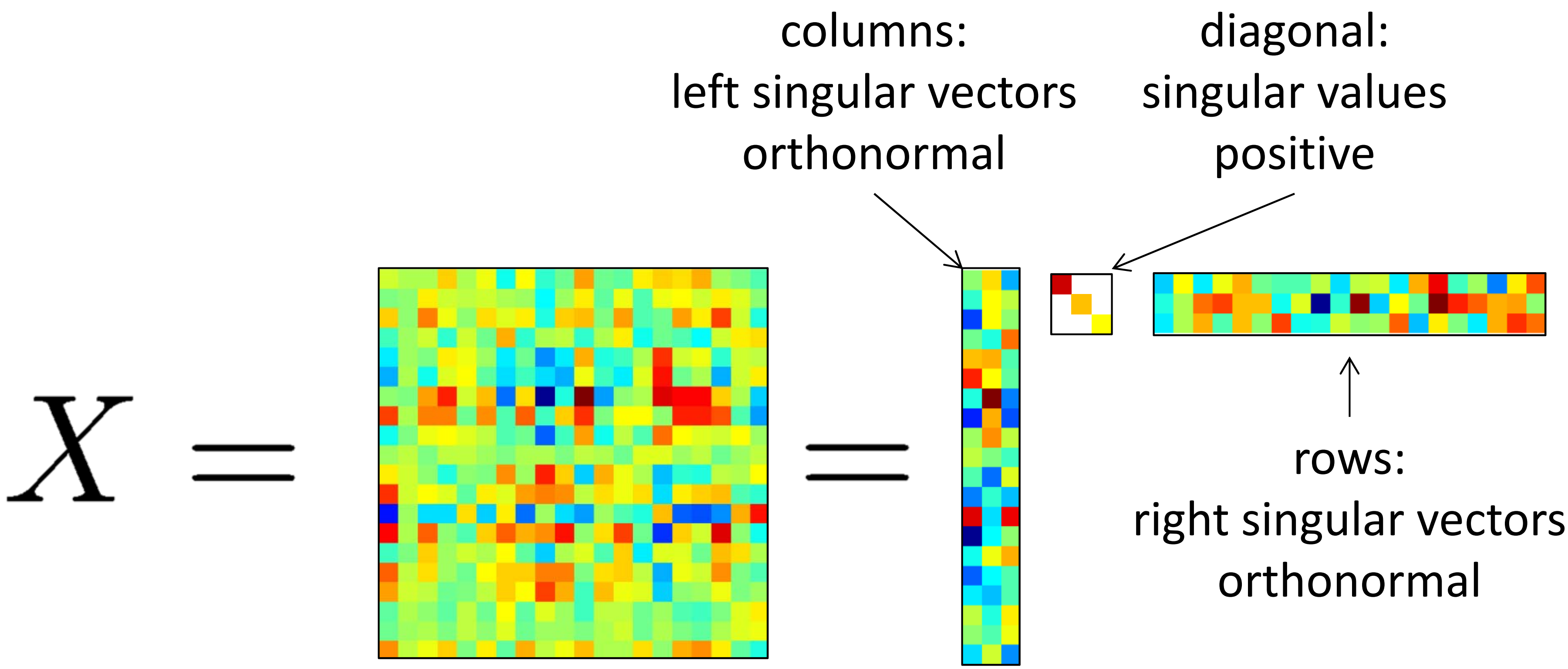
- Many large matrices obey concise models

$$X =$$



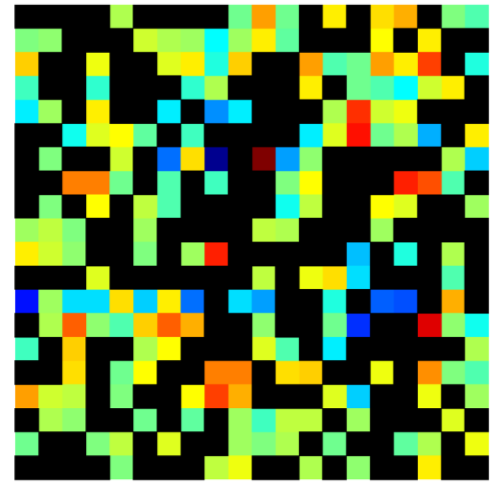
- Low-rank model:
  - columns live in a low-dimensional subspace
  - any column can be written as a weighted sum of just a few other columns
  - can be factored into a product of smaller matrices

# Low-Rank SVD




rank($X$) = # of nonzero singular values

# Low-Rank Matrix Completion

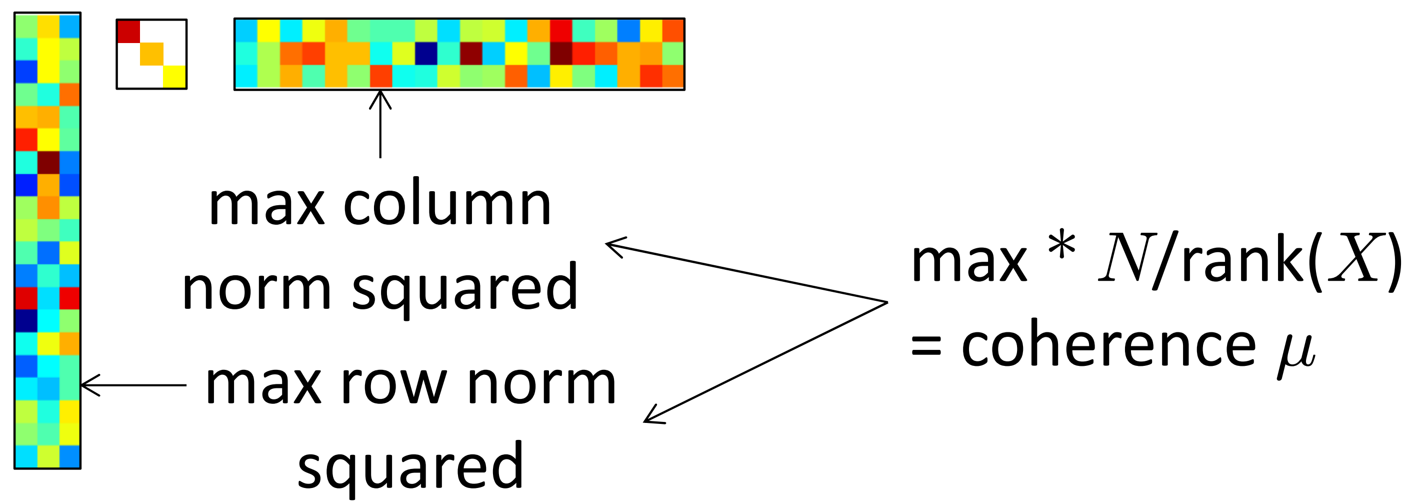


- Ultimate goal
  - find the matrix with the smallest rank that agrees with the observed entries
- Difficulty
  - this problem is NP-hard
- Solution
  - find the matrix with the smallest ***nuclear norm*** that agrees with the observed entries
  - nuclear norm = $L_1$ norm of singular values
  - convex optimization problem
  - strong theory and algorithms [Candès, Recht, Tao, Fazel, Ma, Gross, …]

# Theory for Low-Rank Matrix Completion

- Performance depends on coherence of matrix $X$

$X =$




- Low coherence: $\mu \approx 1$
  – matrix entries well distributed
- High coherence: $\mu \approx N/\text{rank}(X)$
  – matrix entries “spiky”

# Sampling Complexity

- For $N \times N$ matrix with rank $R$ and coherence $\mu$,

$$\boxed{\#\ \text{samples} \approx \mu N R \log^2(N)}$$

  - each column has $\approx R$ degrees of freedom
  - sampling requirement is $\approx \mu\, R$ samples per column

- Conclusion:
  - Low-rank matrices with low coherence can be recovered from small numbers of random samples

# Extension: Phase Retrieval

- Suppose we collect measurements of the form

$$y_m = |\langle \phi_m, x \rangle|^2, \quad m = 1, 2, \ldots, M$$

  where $x$ is an unknown length-$N$ vector.

- We observe only the magnitude of the linear measurements, not the sign (or phase if complex).
  – X-ray crystallography, diffraction imaging, microscopy

- Difficult problem: quadratic system of equations.

# Lifting Trick

- Idea: define a rank-1 matrix

$$X = xx^*$$

- Now, each quadratic measurement of $x$

$$\begin{aligned} y_m &= |\langle \phi_m, x\rangle|^2 = \langle \phi_m, x\rangle^* \langle \phi_m, x\rangle \\ &= \phi_m^* x x^* \phi_m \\ &= \phi_m^* X \phi_m = \langle X, \phi_m \phi_m^* \rangle \end{aligned}$$

is a ***linear*** measurement of $X$.

# PhaseLift (2011)

[Candès, Strohmer, Voroninski]

minimize $\mathrm{Trace}(X)$

subject to

$$y_m = \langle X, \phi_m \phi_m^* \rangle, \quad m = 1, 2, \ldots, M$$

$X$ is positive semidefinite

then factor $X$ to retrieve $x$

# Extension: Blind Deconvolution

- Recover two unknown vectors, $w$ and $x$, from their circular convolution [Ahmed, Recht, Romberg]
  - assume that each vector belongs to a known subspace

- Recover spike train convolved with unknown PSF [Chi]
  - spikes are "off-grid"
  - measurements are low-frequency
  - assume PSF belongs to a known subspace

- Extension to non-stationary PSFs [with Yang and Tang]
  - all PSFs belong to a known subspace

# Conclusions

- ***Concise models enable "filling in the blanks"***
  - collect less data and/or make the most of the available data
- ***Many types of concise models***
  - bandlimitedness, sparsity, structured sparsity, manifolds, low-rank matrices
- ***Algorithms must be tailored to model type***
  - $L_1$ minimization for sparse models
- ***Random measurements***
  - incoherence ensures we don't miss anything
  - stable embedding thanks to low-dimensional geometry
  - randomness convenient but not necessary
- ***Extensions and applications***
  - sampling, imaging, sensor networks, accelerating computations, …

**mwakin@mines.edu** **mines.edu/~mwakin**



Thanks to Dehui Yang, Lucas Andrade de Almeida, Zhihui Zhu, Shuang Li