Tutorial on

Compressive Sensing

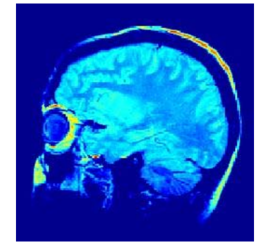
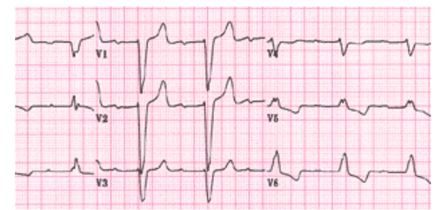
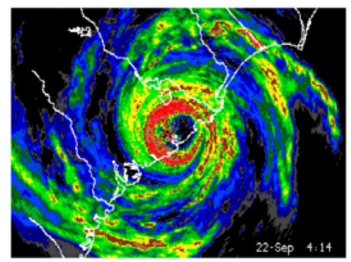
Richard Baraniuk
Rice University

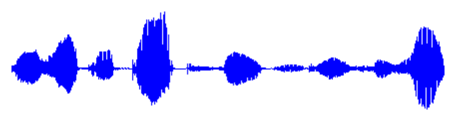
Justin Romberg
Georgia Tech

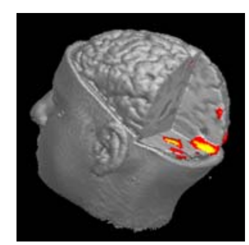
Michael Wakin
University of Michigan







Agenda

- Introduction to Compressive Sensing (CS) [richb]
  - motivation
  - basic concepts

- CS Theoretical Foundation [justin]
  - uniform uncertainty principles
  - restricted isometry principle
  - recovery algorithms

- Geometry of CS [mike]
  - *K*-sparse and compressible signals
  - manifolds

- CS Applications [richb, justin]

Compressive Sensing

Introduction and Background

Digital Revolution

Pressure is on Digital Sensors

- Success of digital data acquisition is placing increasing pressure on signal/image processing hardware and software to support

higher resolution / denser sampling

» ADCs, cameras, imaging systems, microarrays, …

x

large numbers of sensors

» image data bases, camera arrays,
distributed wireless sensor networks, …

x

increasing numbers of modalities

» acoustic, RF, visual, IR, UV, x-ray, gamma ray, …

Pressure is on Digital Sensors

- Success of digital data acquisition is placing increasing pressure on signal/image processing hardware and software to support

higher resolution / denser sampling

» ADCs, cameras, imaging systems, microarrays, …

x

large numbers of sensors

» image data bases, camera arrays, distributed wireless sensor networks, …

x

increasing numbers of modalities

» acoustic, RF, visual, IR, UV

=

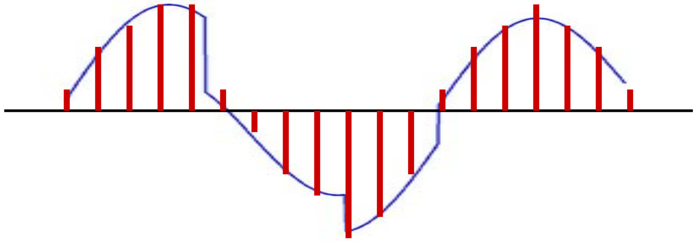
deluge of data

» how to acquire, store, fuse, process efficiently?

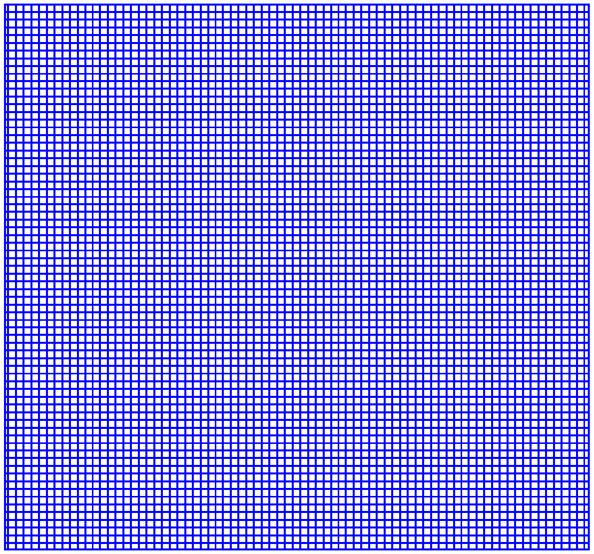
Digital Data Acquisition

- Foundation: *Shannon sampling theorem*

  “if you sample densely enough (at the Nyquist rate), you can perfectly reconstruct the original data”



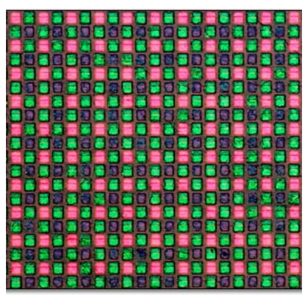
time



space

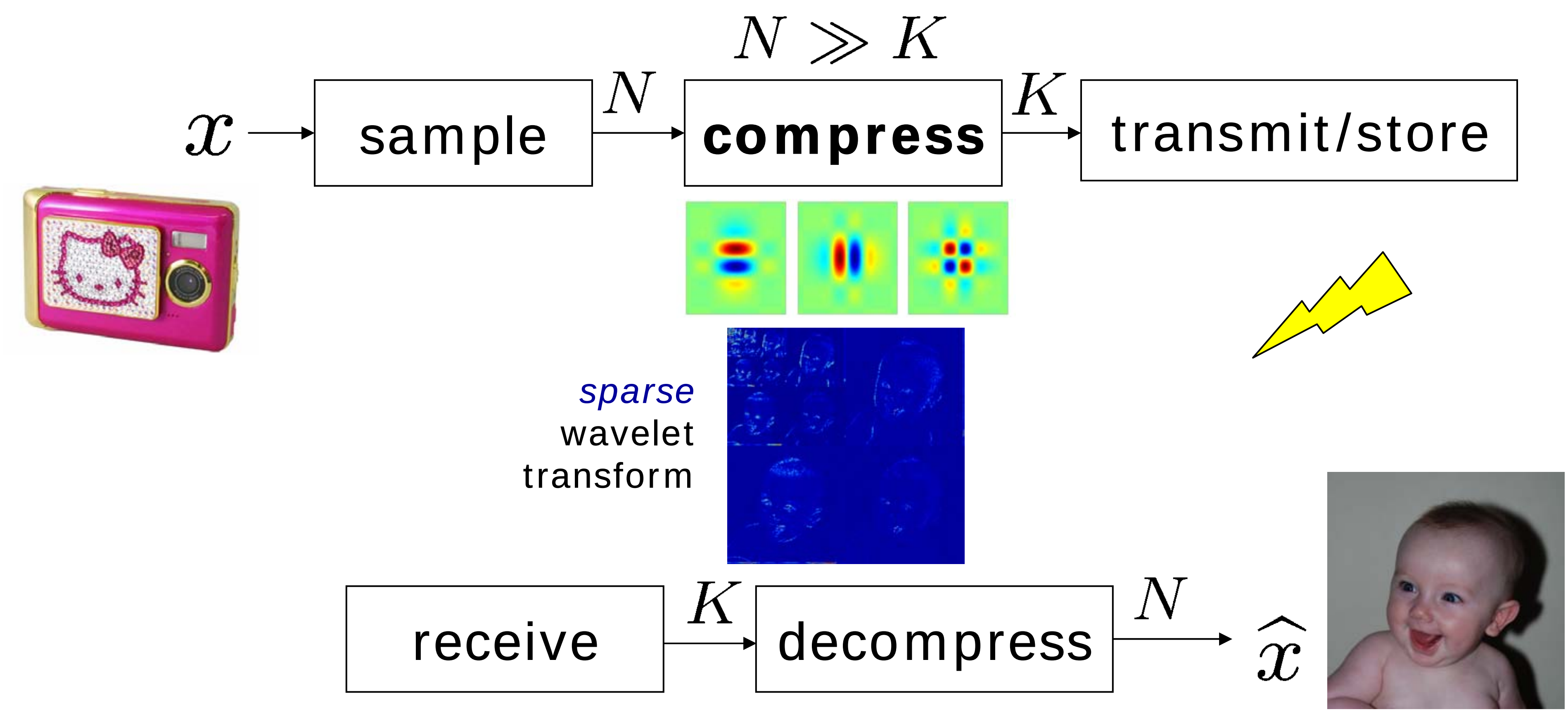
Sensing by *Sampling*

- Long-established paradigm for digital data acquisition
  - uniformly ***sample*** data at Nyquist rate (2x Fourier bandwidth)



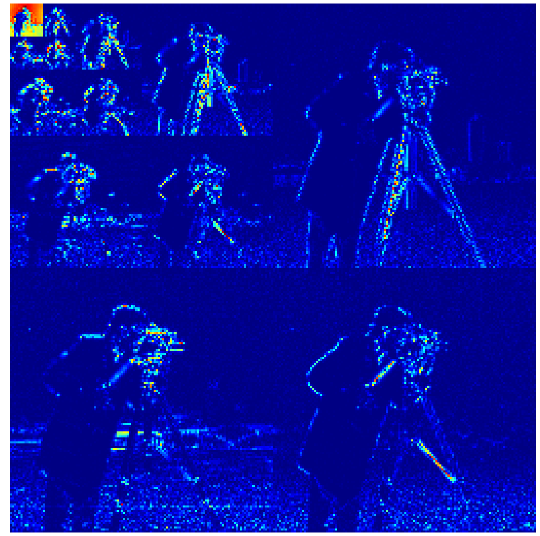
Sensing by *Sampling*

- Long-established paradigm for digital data acquisition
  - uniformly ***sample*** data at Nyquist rate (2x Fourier bandwidth)
  - ***compress*** data (signal-dependent, nonlinear)



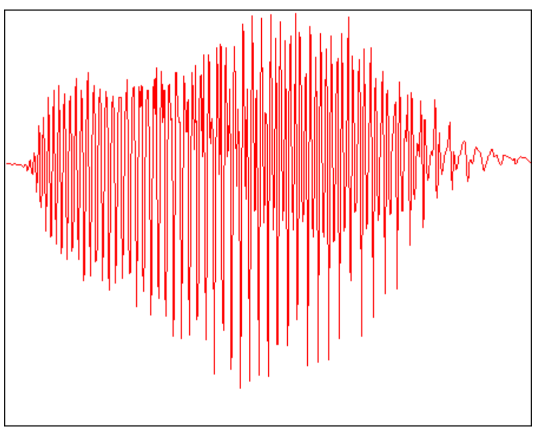
Sparsity / Compressibility

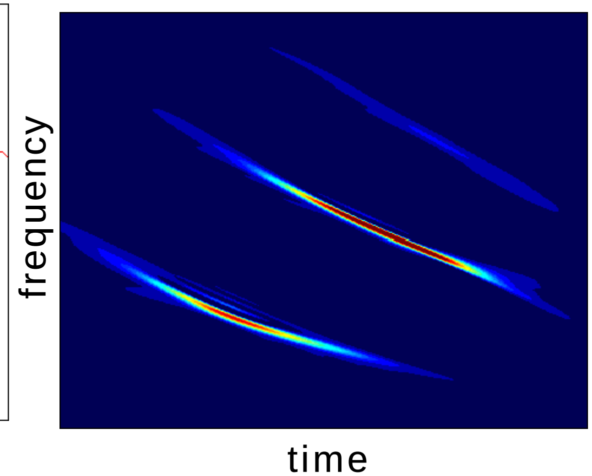
N
pixels



$K \ll N$
large
wavelet
coefficients

N
wideband
signal
samples

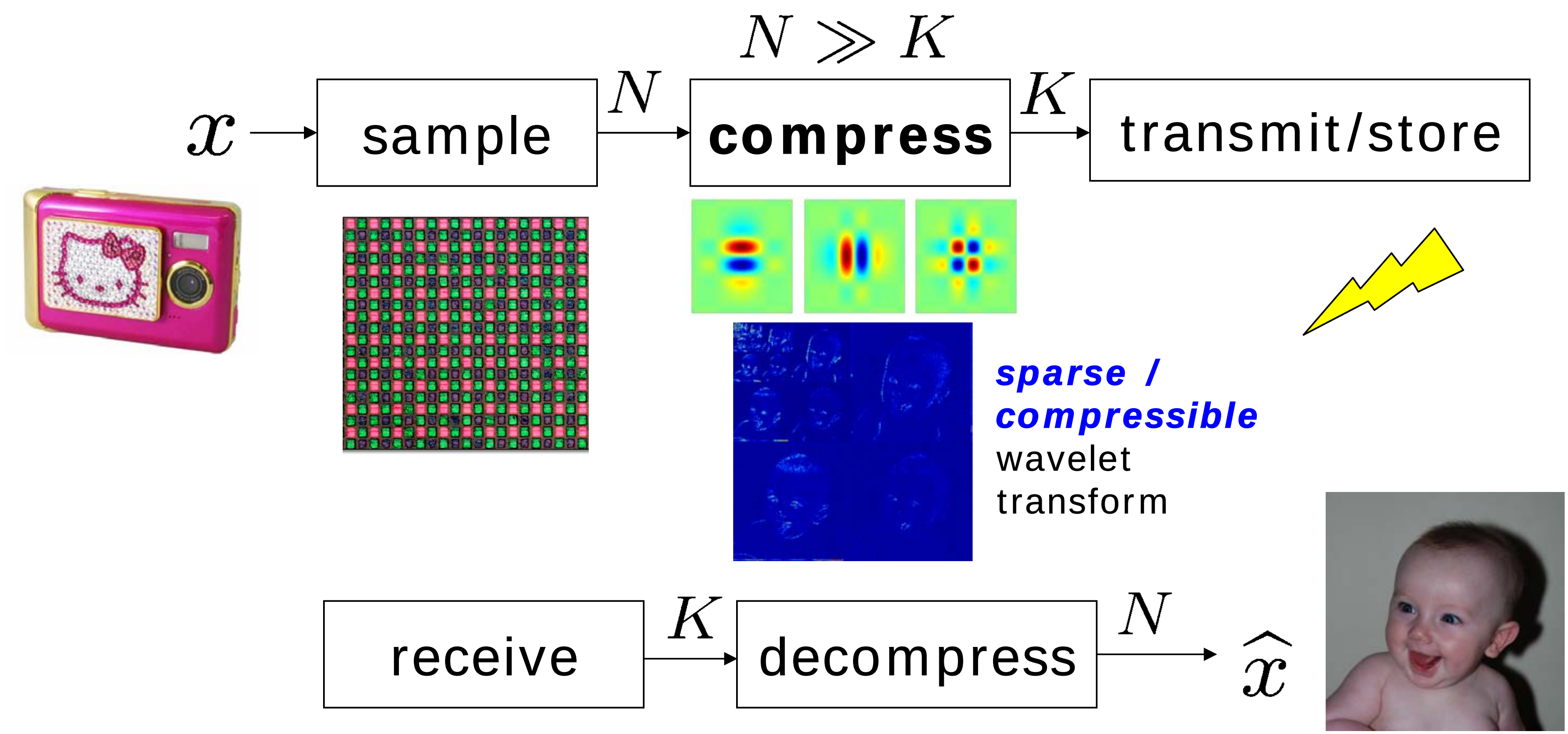




$K \ll N$
large
Gabor
coefficients

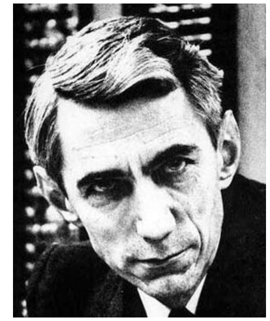
What's Wrong with this Picture?

- Long-established paradigm for digital data acquisition
  - ***sample*** data at Nyquist rate (2x bandwidth)
  - ***compress*** data (signal-dependent, nonlinear)
  - ***brick wall*** to resolution/performance



Compressive Sensing (CS)

- Recall Shannon/Nyquist theorem
  - Shannon was a *pessimist*
  - 2x oversampling Nyquist rate is a worst-case bound for *any* bandlimited data
  - sparsity/compressibility irrelevant
  - Shannon sampling is a linear process while compression is a nonlinear process

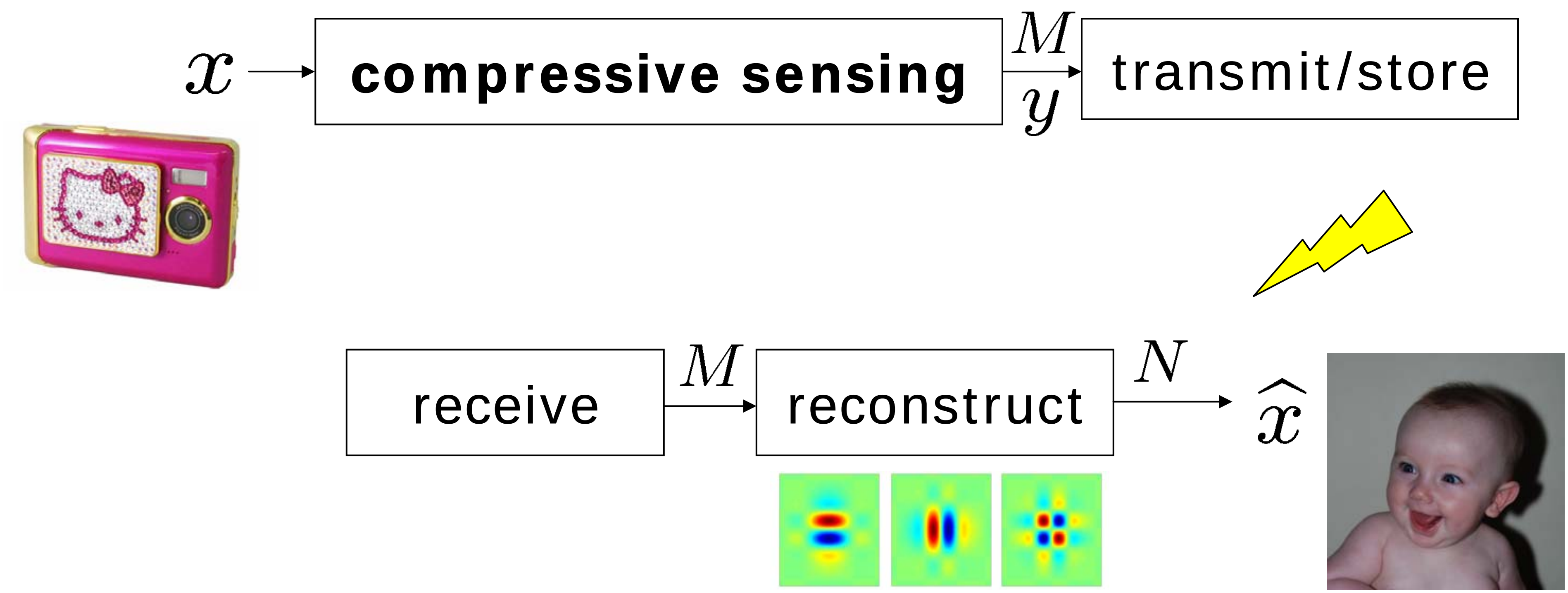


- **Compressive sensing**
  - new sampling theory that *leverages compressibility*
  - based on new *uncertainty principles*
  - *randomness* plays a key role

Compressive Sensing

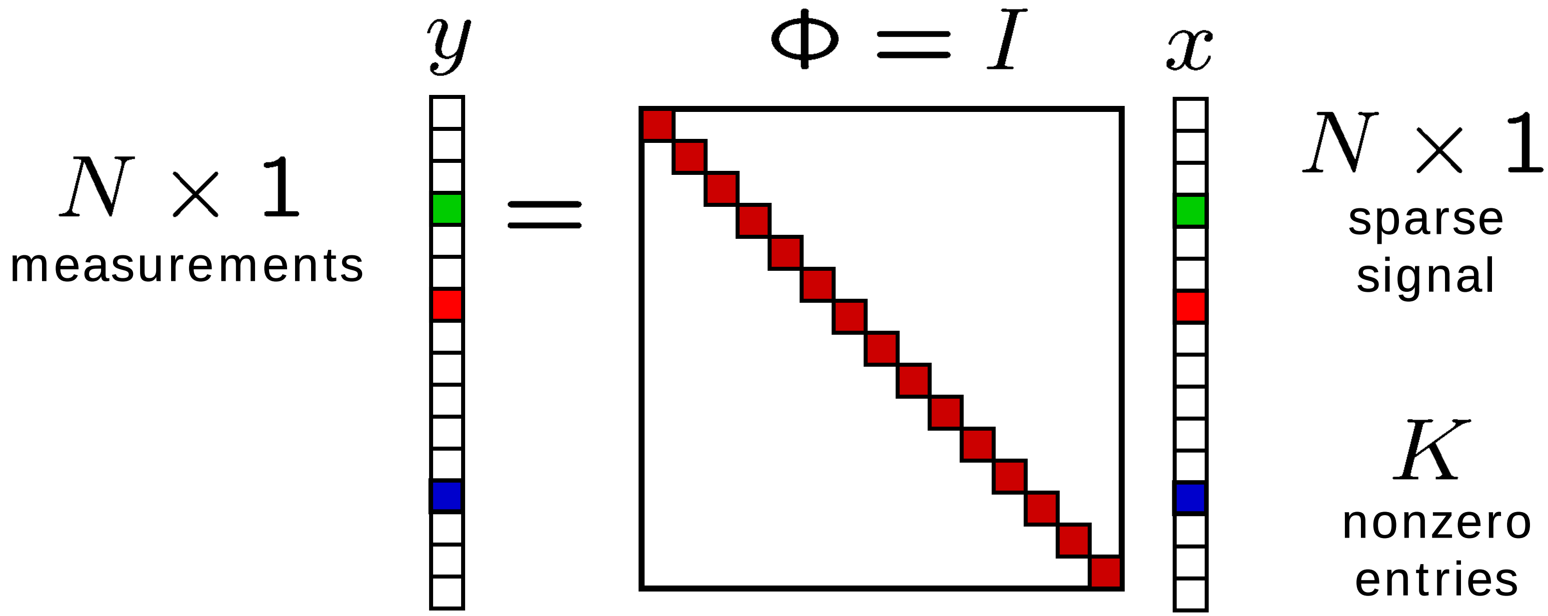
- Directly acquire "*compressed*" data
- Replace samples by more general "measurements"

$$K < \underline{M} \ll N$$



Sampling

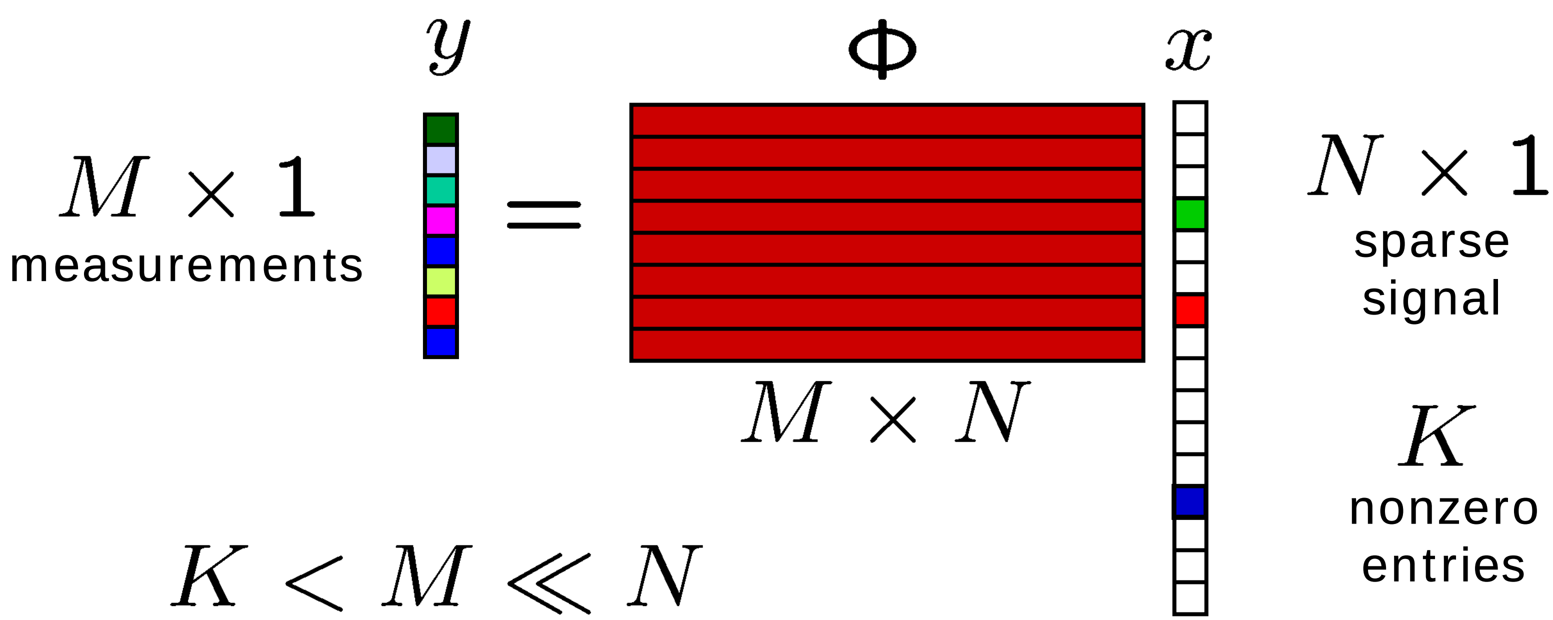
- Signal x is K-*sparse* in basis/dictionary Ψ
  - WLOG assume sparse in space domain $\Psi = I$
- **Samples**



Compressive Data Acquisition

- When data is sparse/compressible, can directly acquire a ***condensed representation*** with no/little information loss through ***dimensionality reduction***

$$y = \Phi x$$

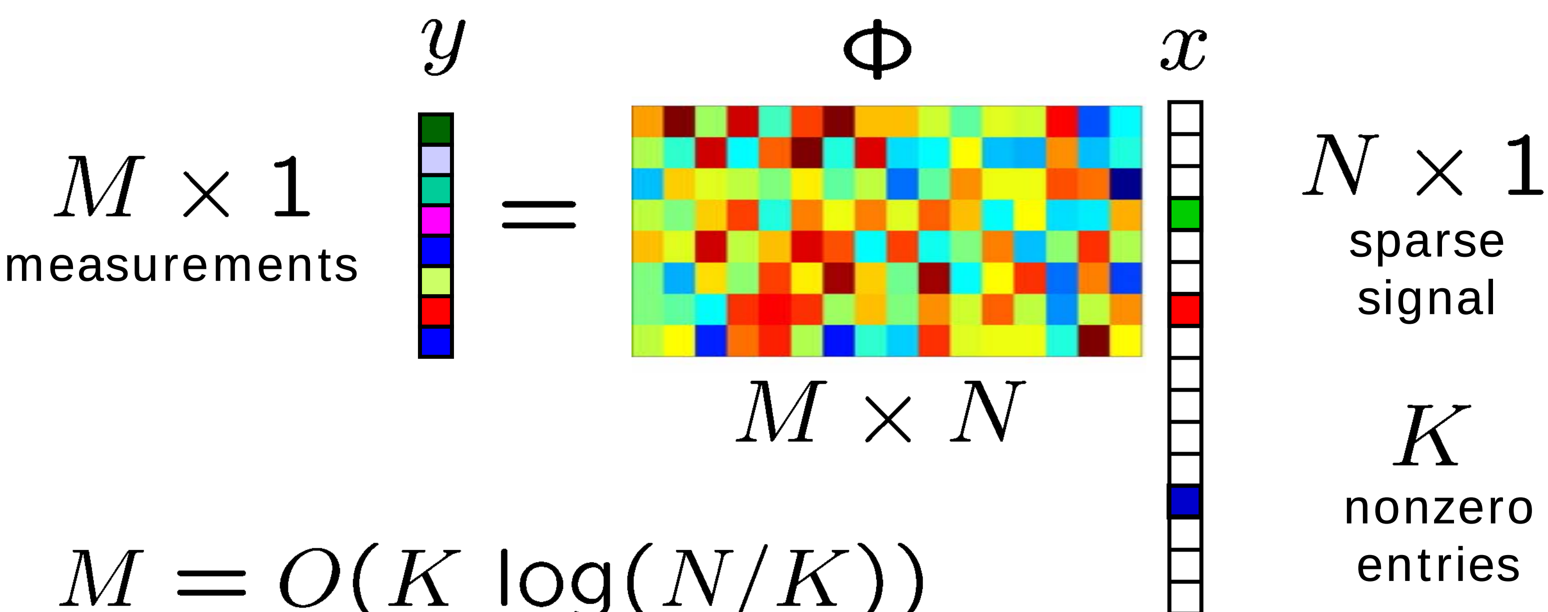


Compressive Data Acquisition

- When data is sparse/compressible, can directly acquire a ***condensed representation*** with no/little information loss

$$y = \Phi x$$

- ***Random projection*** will work



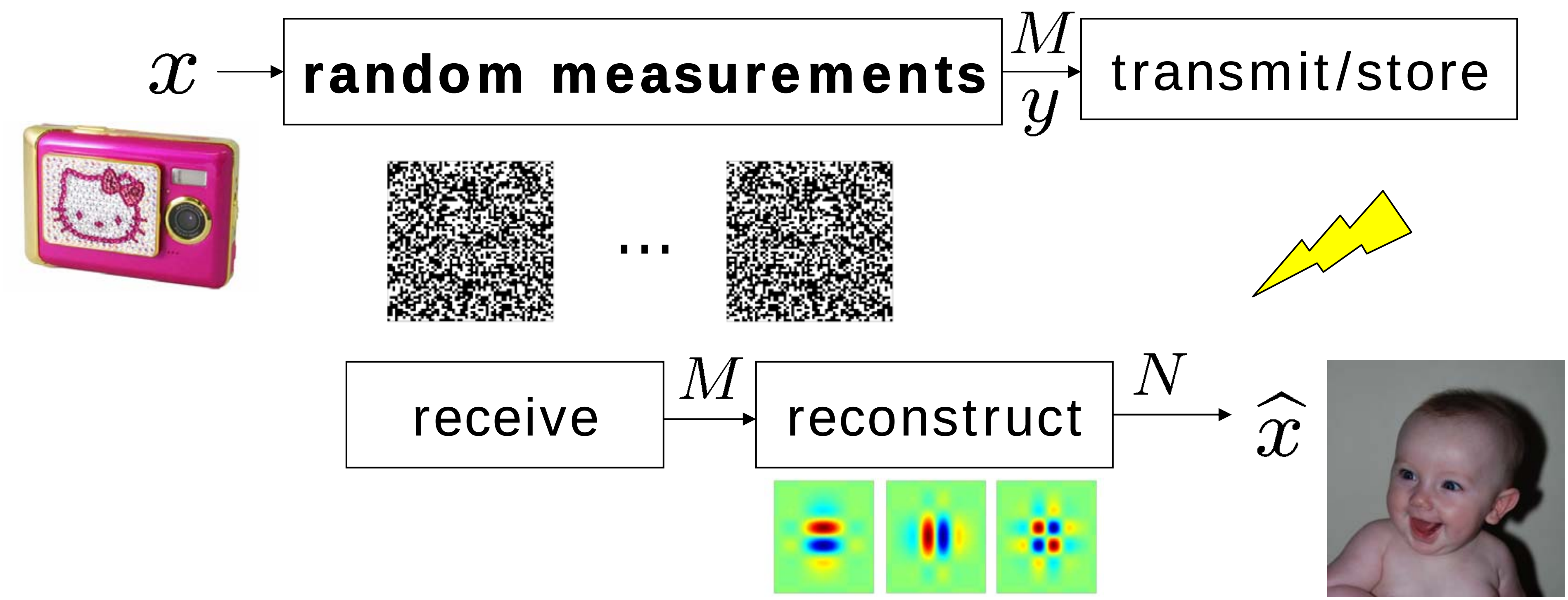
$$M = O(K \log(N/K))$$

[Candes-Romberg-Tao, Donoho, 2004]

Compressive Sensing

- Directly acquire "***compressed***" data
- Replace samples by more general "measurements"

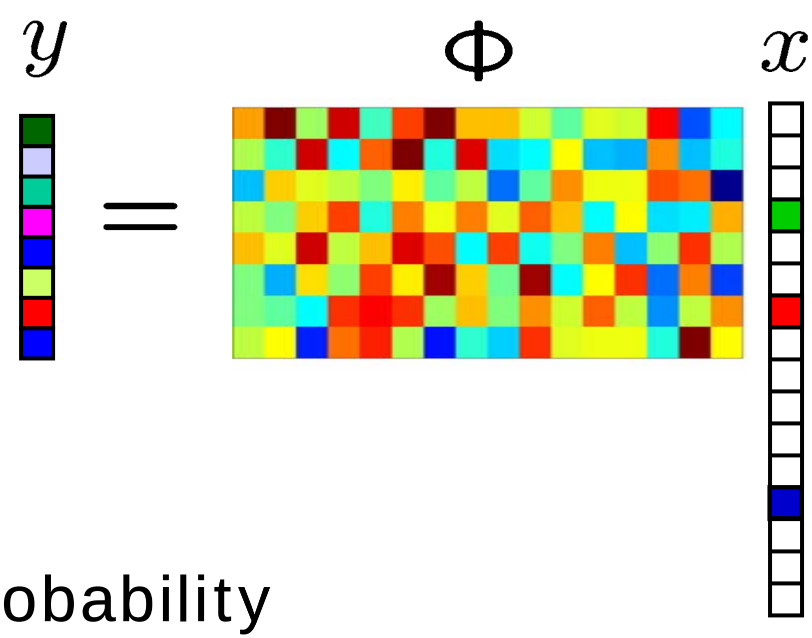
$$M = O(K \log(N/K))$$



Why Does It Work?

- Random projection Φ
  not full rank...

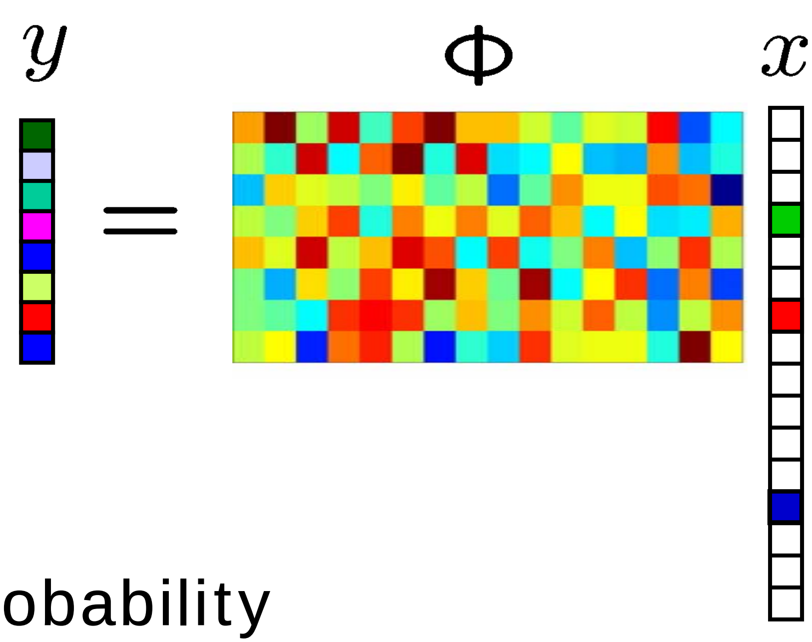
  ... but
  preserves structure
  and information
  in sparse/compressible
  signals models with high probability



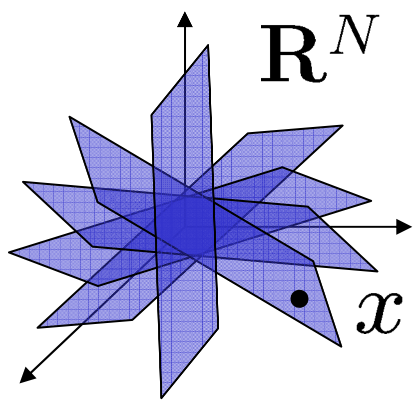
Why Does It Work?

- Random projection Φ
  not full rank...

  ... but
  preserves structure and information
  in sparse/compressible signals models with high probability



K-sparse model

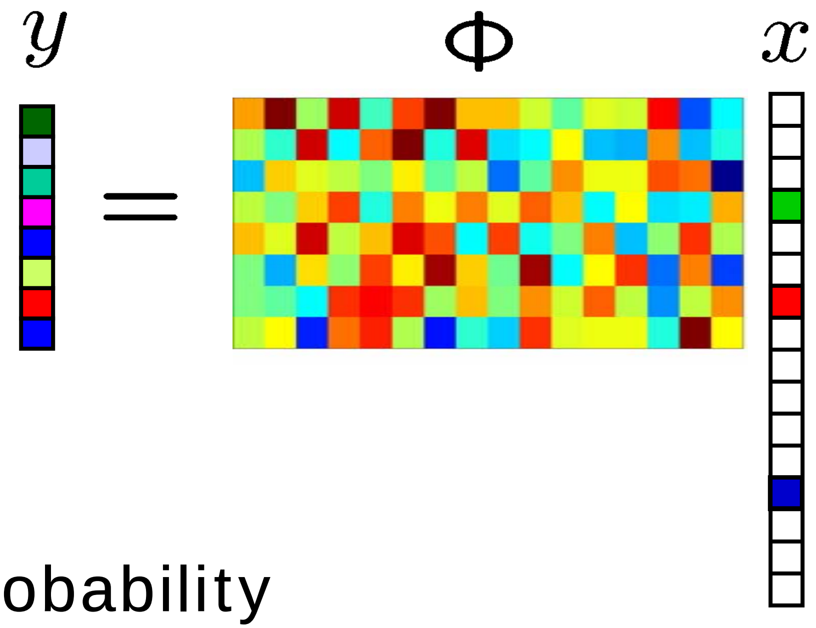


K-dim hyperplanes aligned with coordinate axes

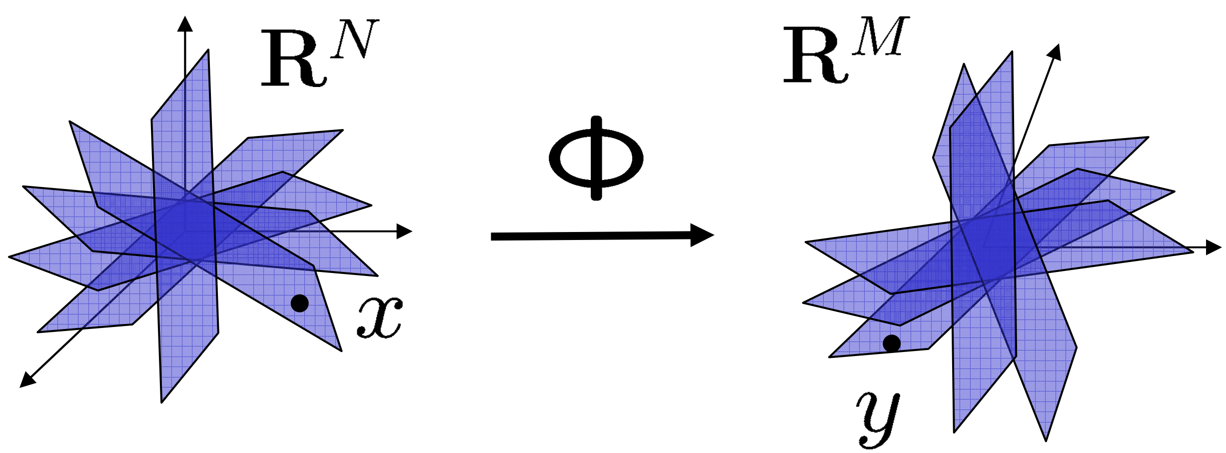
Why Does It Work?

- Random projection Φ
  not full rank...

  ... but
  preserves structure and information
  in sparse/compressible signals models with high probability



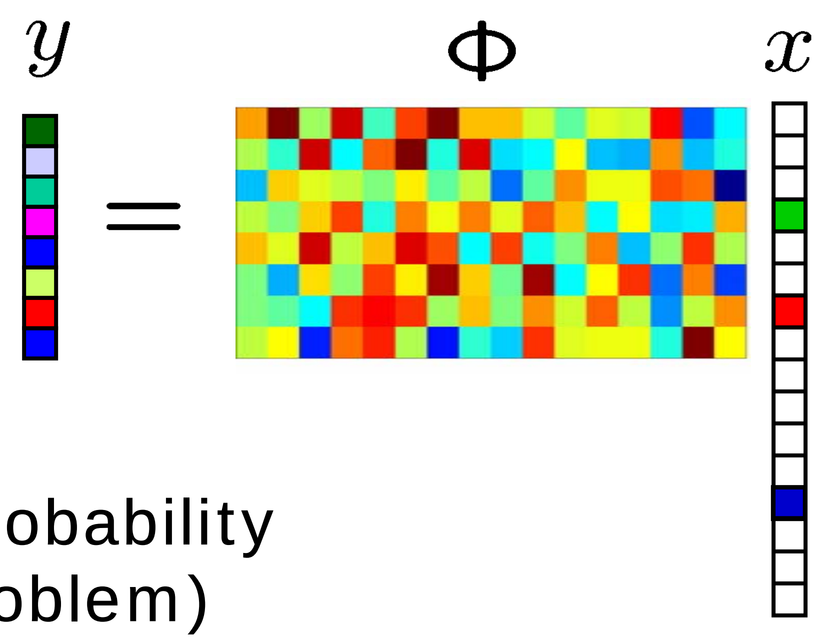
K-sparse model



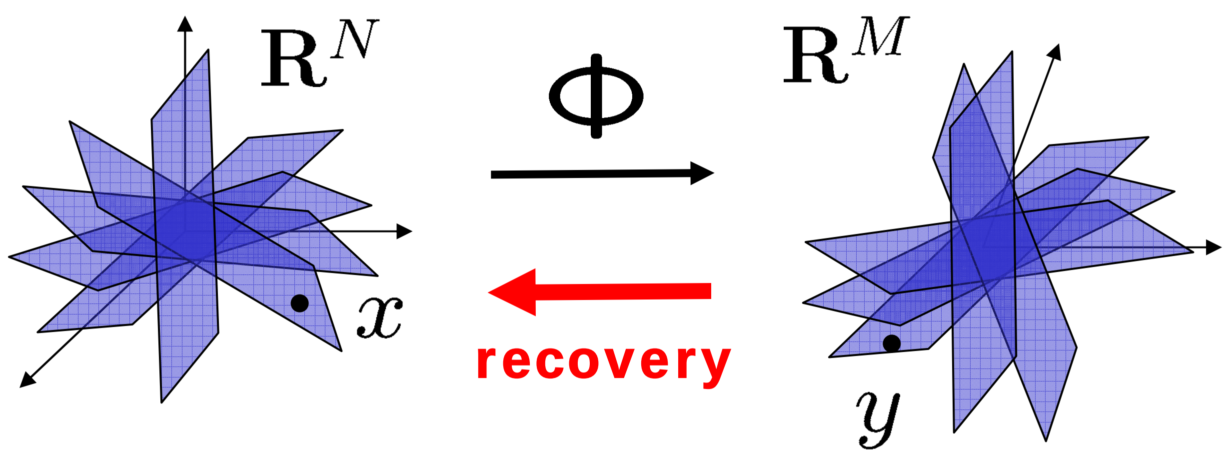
CS Signal Recovery

- Random projection Φ
  not full rank...

  ... but
  is invertible
  for sparse/compressible
  signals models with high probability
  (solves ill-posed inverse problem)



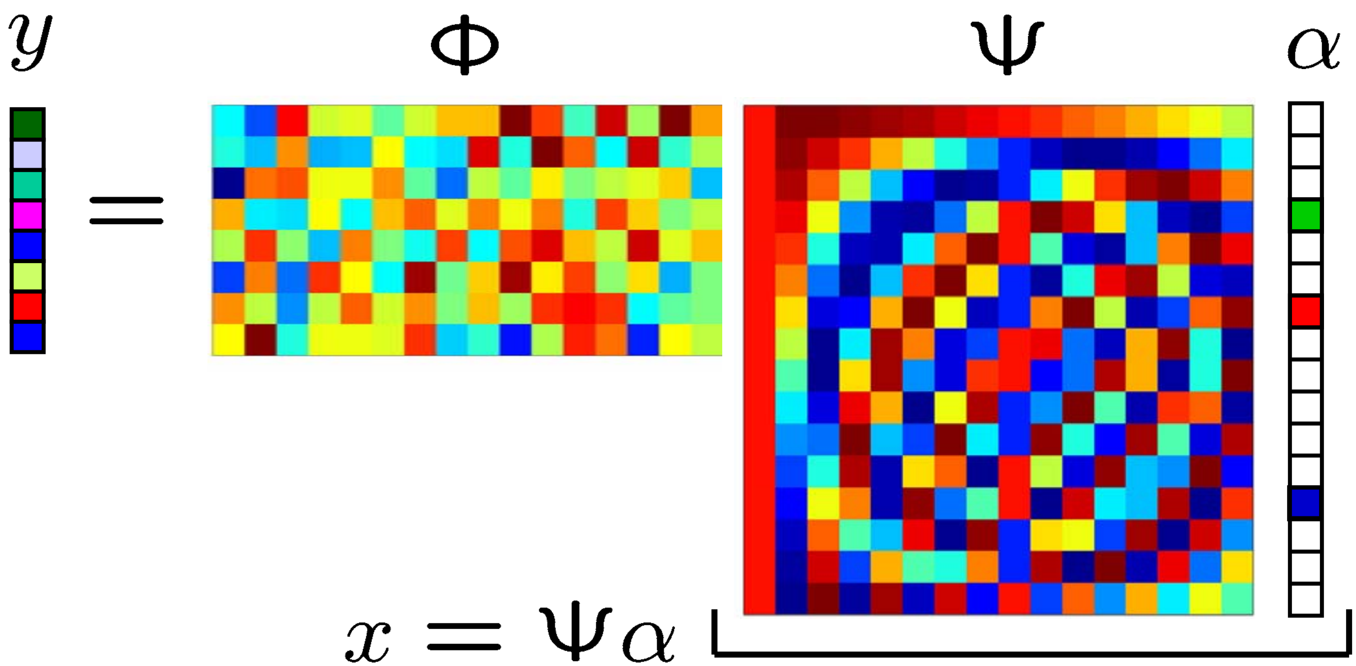
K-sparse model



Universality

- Random measurements can be used for signals sparse in *any* basis

$$y = \Phi x = \Phi \Psi \alpha$$



CS Hallmarks

- CS changes the rules of the data acquisition game
  - exploits a priori signal *sparsity* information
- **Universal**
  - same random projections / hardware can be used for *any* compressible signal class (*generic*)
- **Democratic**
  - each measurement carries the same amount of information
  - simple encoding
  - robust to measurement loss and quantization
- **Asymmetrical** (most processing at decoder)
- Random projections weakly **encrypted**

Agenda

- Introduction to Compressive Sensing (CS) [richb]
  - motivation
  - basic concepts

- **CS Theoretical Foundation** [justin]
  - uniform uncertainty principles
  - restricted isometry principle
  - recovery algorithms

- Geometry of CS [mike]
  - K-sparse and compressible signals
  - manifolds

- CS Applications [richb, justin]

Theoretical Foundations
of
Compressive Sensing

Ingredients for Compressed Sensing

- Sparse signal representation
- Coded measurements (sampling process)
- Recovery algorithms (nonlinear)

Signal and Image Representations

- Fundamental concept in DSP: *Transform-domain processing*
- Decompose f as superposition of atoms (orthobasis or tight frame)

$$f(t) = \sum_i \alpha_i \psi_i(t) \quad \text{or} \quad f = \Psi \alpha$$

  e.g. sinusoids, wavelets, curvelets, Gabor functions,...

- Process the coefficient sequence α

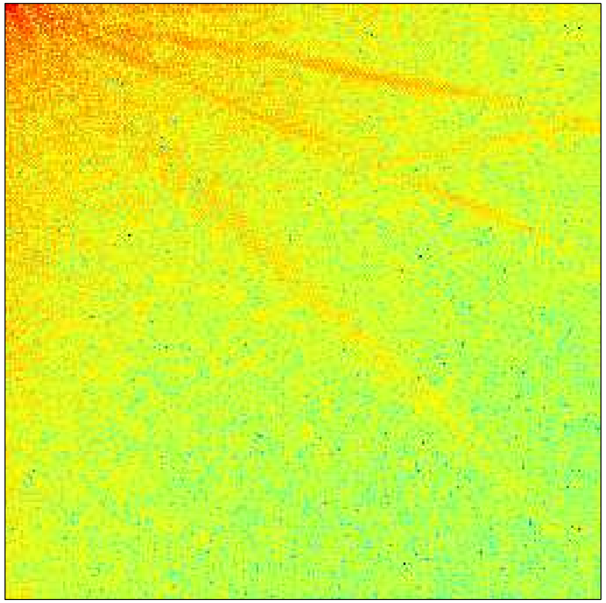
$$\alpha_i = \langle f, \psi_i \rangle, \quad \text{or} \quad \alpha = \Psi^T f$$

- Why do this?
  If we choose Ψ wisely, $\{\alpha_i\}$ will be "simpler" than $f(t)$
- (Good) transform coefficients typically *decay*

$$|\alpha|_{(k)} \sim k^{-r}, \quad \text{for some} \quad r > 1.$$

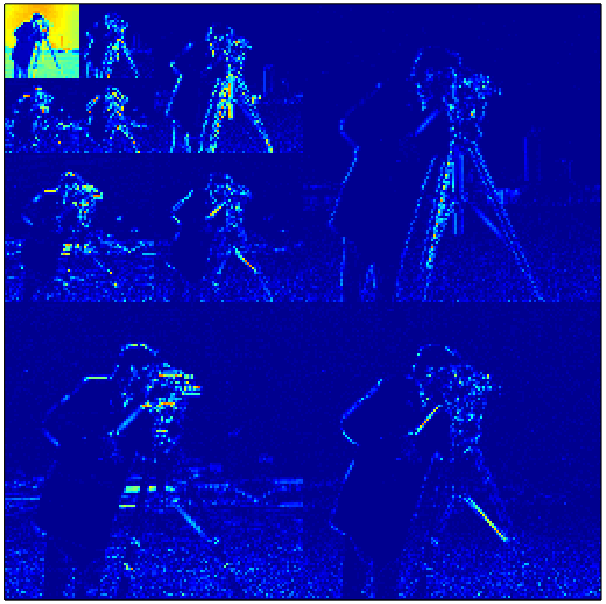
Classical Image Representation: DCT

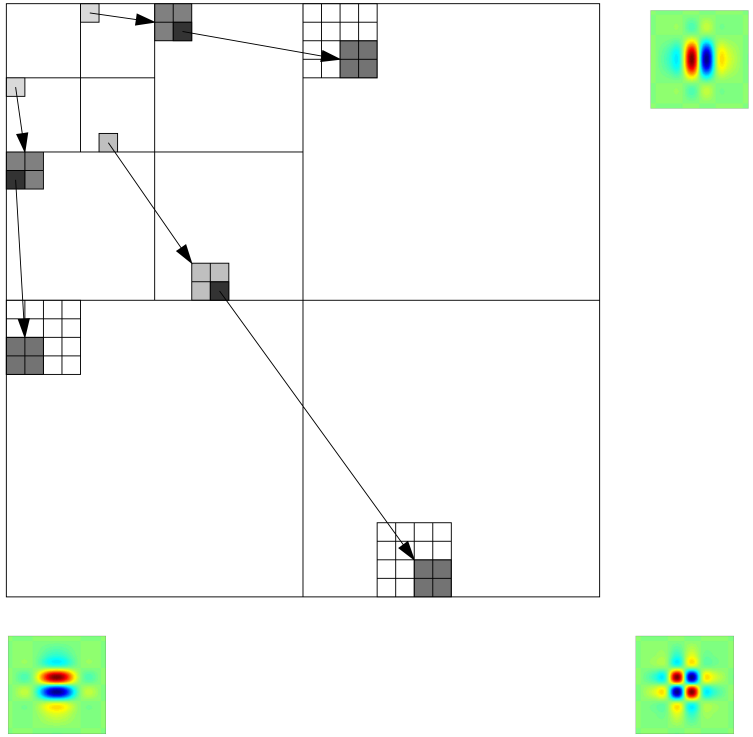
- Discrete Cosine Transform (DCT)
  Basically a real-valued Fourier transform (sinusoids)
- Model: most of the energy is at low frequencies



- Basis for JPEG image compression standard
- DCT approximations: smooth regions great, edges blurred/ringing

Modern Image Representation: 2D Wavelets



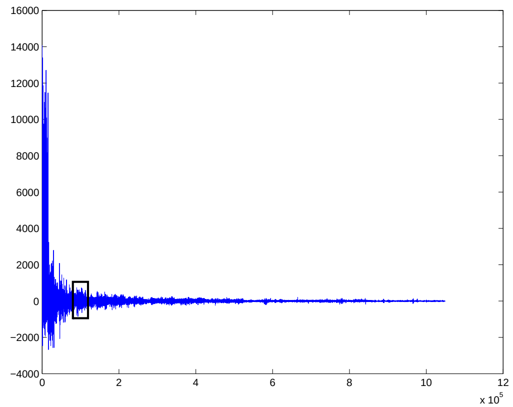


- Sparse structure: few large coeffs, many small coeffs
- Basis for JPEG2000 image compression standard
- Wavelet approximations: smooths regions great, edges much sharper
- *Fundamentally better than DCT for images with edges*

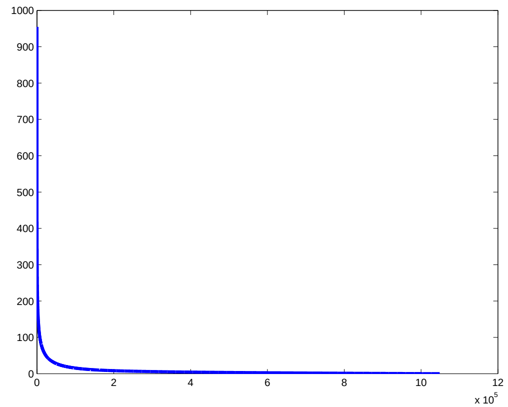
Wavelets and Images

1 megapixel image

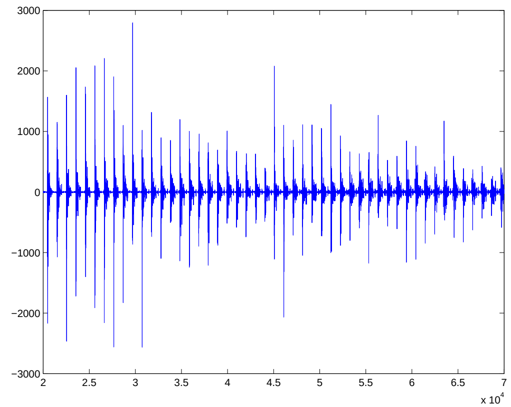
wavelet coeffs



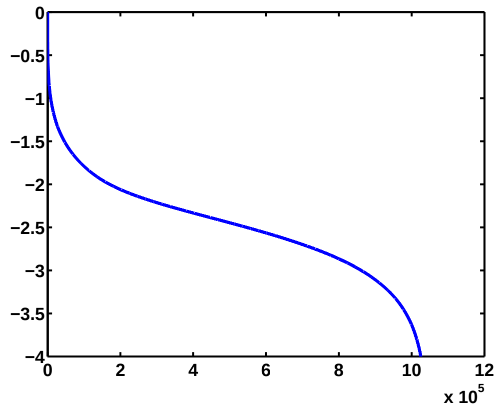
(sorted)



⇓



zoom in

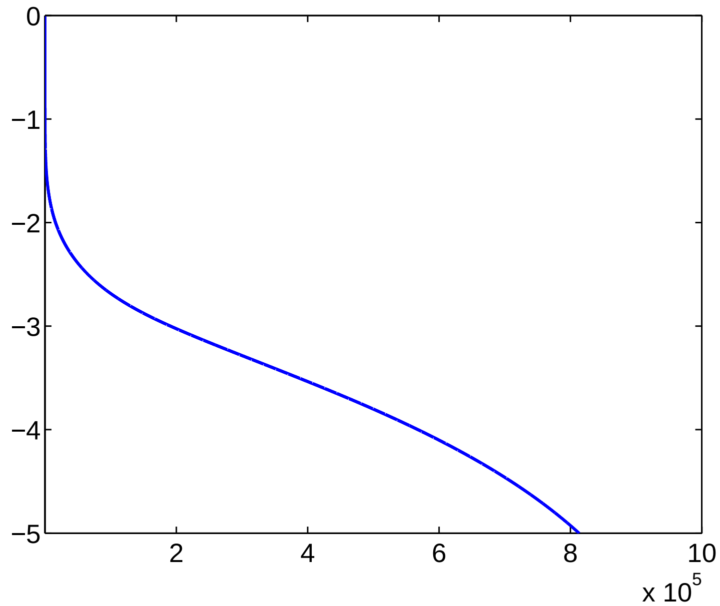


($\log_{10}$ sorted)

Wavelet Approximation

1 megapixel image

25k term approx



B-term approx error

- Within 2 digits (in MSE) with $\approx 2.5\%$ of coeffs
- Original image = f, K-term approximation = f_K

$$\|f - f_K\|_2 \approx .01 \cdot \|f\|_2$$

Computational Harmonic Analysis

- Sparsity plays a *fundamental role* in how well we can:
  - Estimate signals in the presence of noise (shrinkage, soft-thresholding)
  - Compress (transform coding)
  - Solve inverse problems (restoration and imaging)
- Dimensionality reduction facilitates modeling:
  simple models/algorithms are effective
- *Compressed sensing:*
  Sparsity also determines how quickly we can acquire signals
  non-adaptively

Coded Measurements

Coded Acquisition

- Instead of pixels, take *linear measurements*

$$y_1 = \langle f, \phi_1 \rangle, \quad y_2 = \langle f, \phi_2 \rangle, \quad \ldots, y_M = \langle f, \phi_M \rangle$$

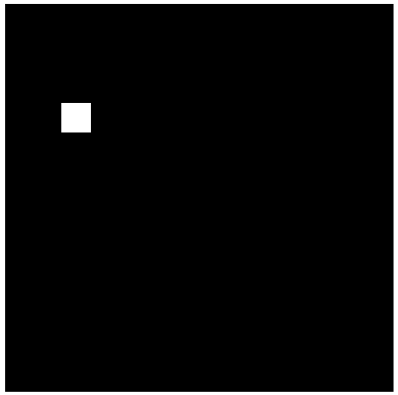
$$y = \Phi f$$

- Equivalent to transform domain sampling,
  $\{\phi_m\}$ = basis functions

- Example: big pixels

$$y_m = \Big\langle$$

,



$$\Big\rangle$$

Coded Acquisition

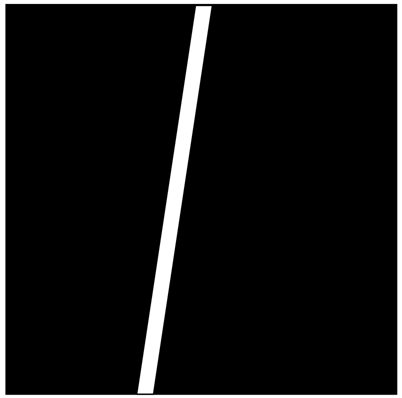
- Instead of pixels, take *linear measurements*

$$y_1 = \langle f, \phi_1 \rangle, \quad y_2 = \langle f, \phi_2 \rangle, \quad \ldots, y_M = \langle f, \phi_M \rangle$$

$$y = \Phi f$$

- Equivalent to transform domain sampling,
  $\{\phi_m\}$ = basis functions

- Example: line integrals (tomography)

$y_m = \Big\langle$  ,  $\Big\rangle$

Coded Acquisition

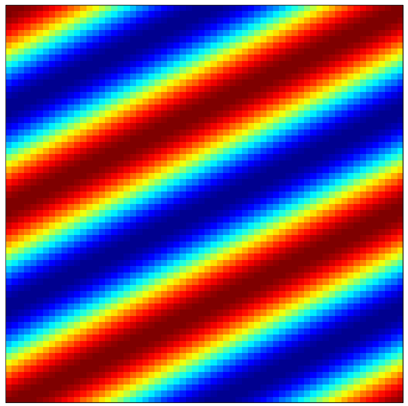
- Instead of pixels, take *linear measurements*

$$y_1 = \langle f, \phi_1 \rangle, \quad y_2 = \langle f, \phi_2 \rangle, \quad \ldots, y_M = \langle f, \phi_M \rangle$$

$$y = \Phi f$$

- Equivalent to transform domain sampling,
  $\{\phi_m\}$ = basis functions

- Example: sinusoids (MRI)

$y_m = \Big\langle$  ,  $\Big\rangle$

Sampling Domain

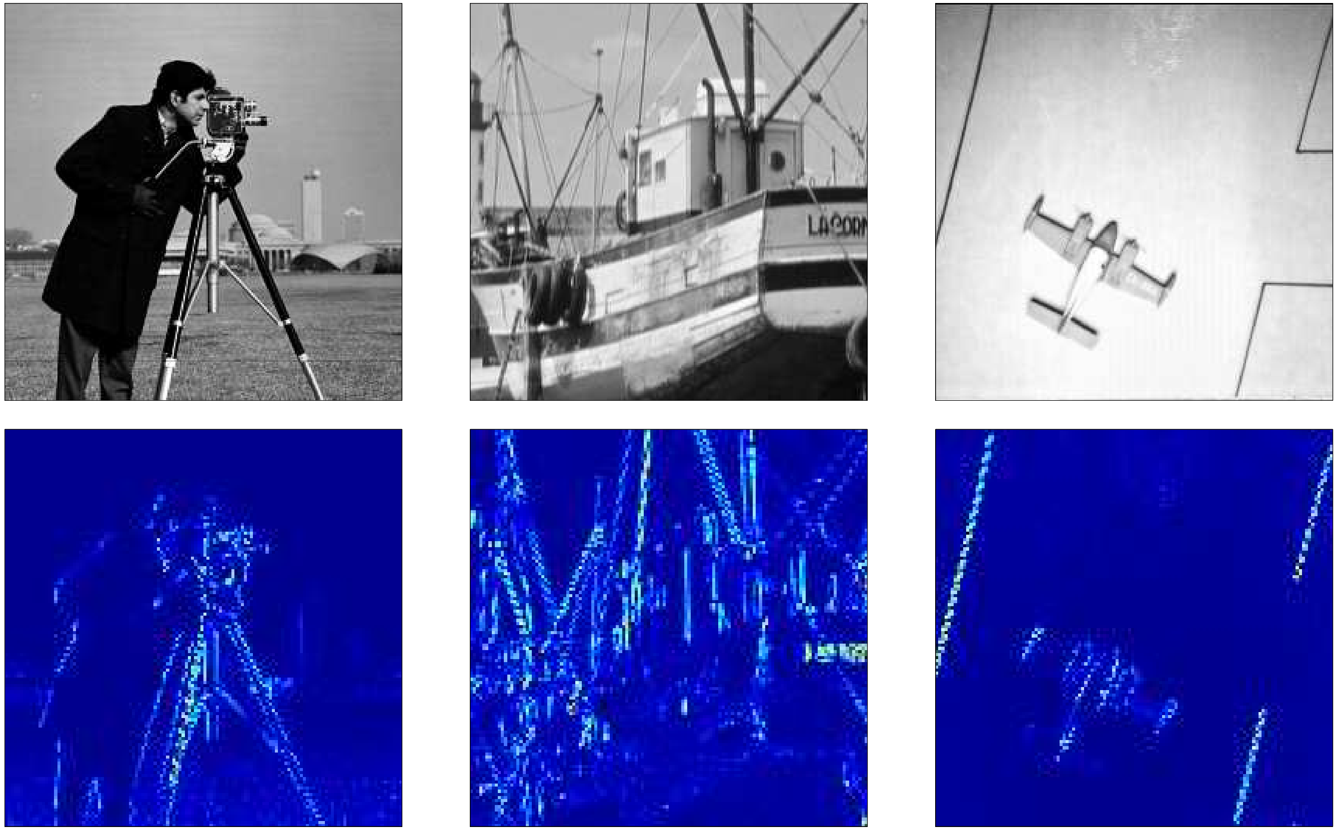
$$y_k = \Big\langle$$

$$, \quad ? \Big\rangle$$

- Which ϕ_m should we use to minimize the number of samples?
- Say we use a sparsity basis for the ϕ_m:
  M measurements = M-term approximation
- So, should we measure wavelets?

Wavelet Imaging?

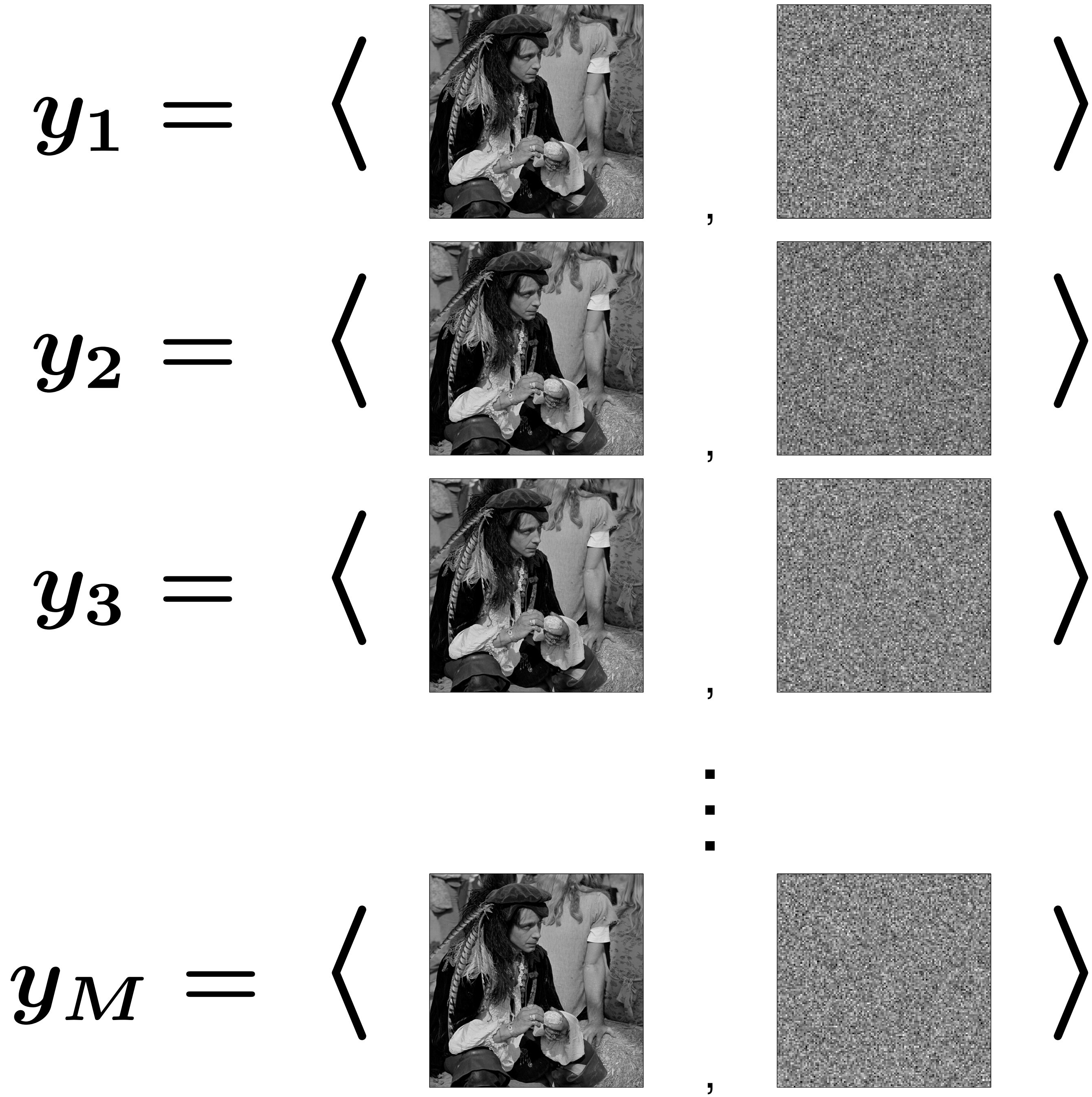
- Want to measure wavelets, but which ones?



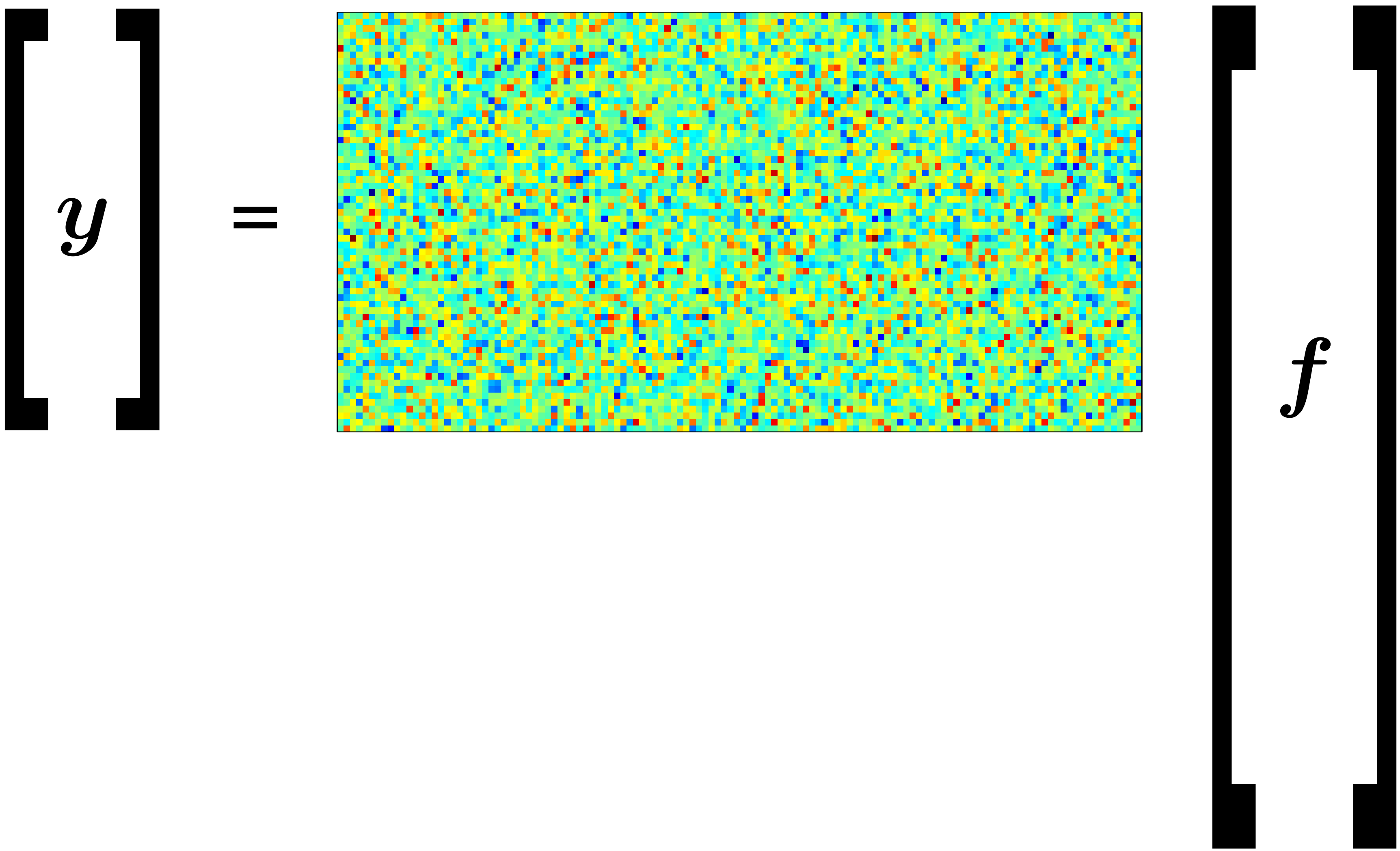
The Big Question

Can we get adaptive approximation performance from a fixed set of measurements?

- Surprisingly: yes.
- More surprising: measurements should *not* match image structure at all
- The measurements should look like random noise


$y_1 =$
$y_2 =$
$y_3 =$
$y_M =$

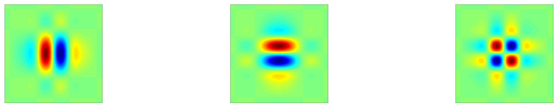
$$\left[\, y \,\right] = \left[\qquad\qquad \Phi \qquad\qquad \right] \left[\begin{array}{c} \\ f \\ \\ \\ \end{array} \right]$$

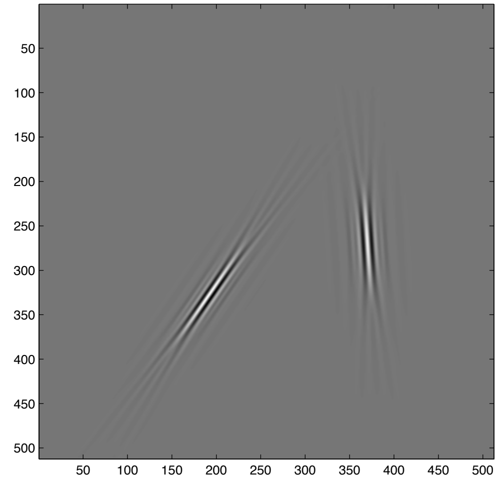

y
=
f

Representation vs. Measurements

- Image structure: *local, coherent*

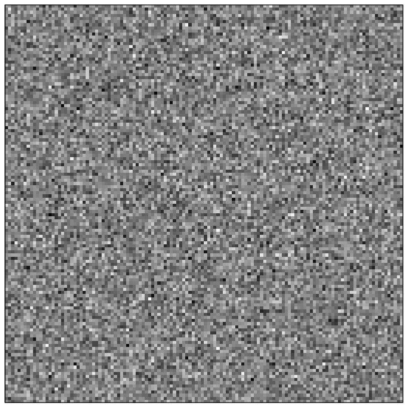
  Good basis functions:

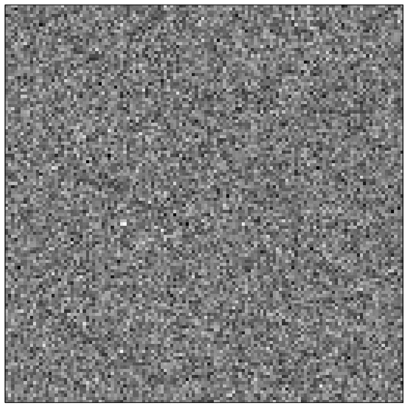


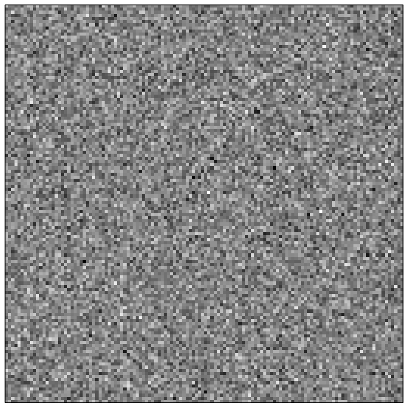


- Measurements: *global, incoherent*

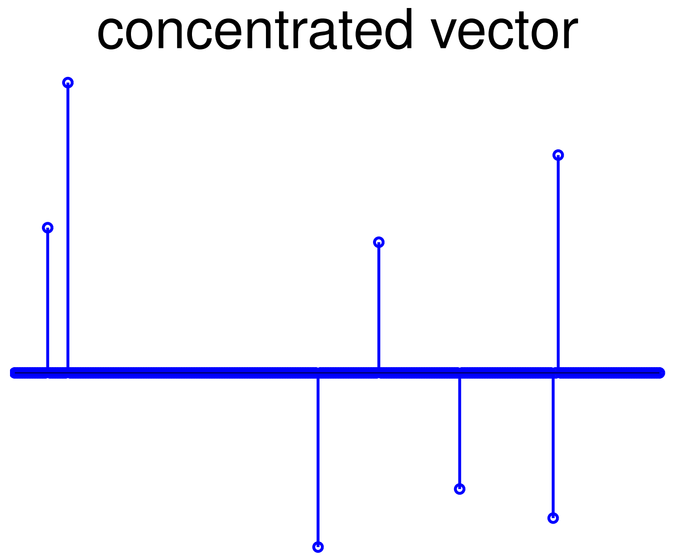
  Good test functions:

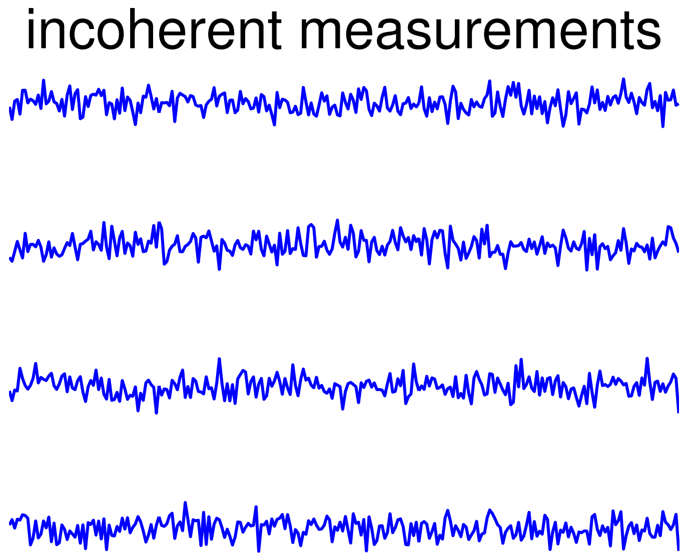






Motivation: Sampling Sparse Coefficients





- Signal is local, measurements are global
- Each measurement picks up a little information about each component
- Triangulate significant components from measurements
- Formalization: Relies on uncertainty principles between sparsity basis and measurement system

Sparse Recovery

The Restricted Isometry Property (RIP)

- $\boldsymbol{\Phi}$ obeys a RIP for sets of size $\boldsymbol{K}$ if

$$0.8 \cdot \frac{M}{N} \cdot \|f\|_2^2 \leq \|\Phi f\|_2^2 \leq 1.2 \cdot \frac{M}{N} \cdot \|f\|_2^2$$

  for every $\boldsymbol{K}$-sparse vector $\boldsymbol{f}$

- Can be interpreted as an *uncertainty principle*
- Examples: $\boldsymbol{\Phi}$ obeys RIP for $K \lesssim M/\log N$ when
  - ϕ_m = random Gaussian
  - ϕ_m = random binary
  - ϕ_m = randomly selected Fourier samples (extra $\log$ factors apply)
- We call these types of measurements *incoherent*

RIP and Sparse Recovery

- RIP for sets of size $2K$ $\Rightarrow$ there is only one K-sparse explanation for y (almost automatic)

- Say f_0 is K-sparse, and we measure $y = \Phi f_0$
  If we search for the sparsest vector that explains y, we will find f_0:

$$\min_{f} \#\{t : f(t) \neq 0\} \quad \text{subject to} \quad \Phi f = y$$

- This is nice, but impossible (combinatorial)

- But, we can use the ℓ_1 norm as a *proxy* for sparsity

Sparse Recovery via ℓ_1 Minimization

- Say f_0 is K-sparse, Φ obeys RIP for sets of size $4K$
- Measure $y = \Phi f_0$
- Then solving

$$\min_f \|f\|_{\ell_1} \quad \text{subject to} \quad \Phi f = y$$

  will recover f_0 exactly

- We can recover f_0 from

$$M \gtrsim K \cdot \log N$$

  incoherent measurements by solving a *tractable* program

- *Number of measurements* $\approx$ *number of active components*

Example: Sampling a Superposition of Sinusoids

- Sparsity basis = Fourier domain, Sampling basis = time domain:

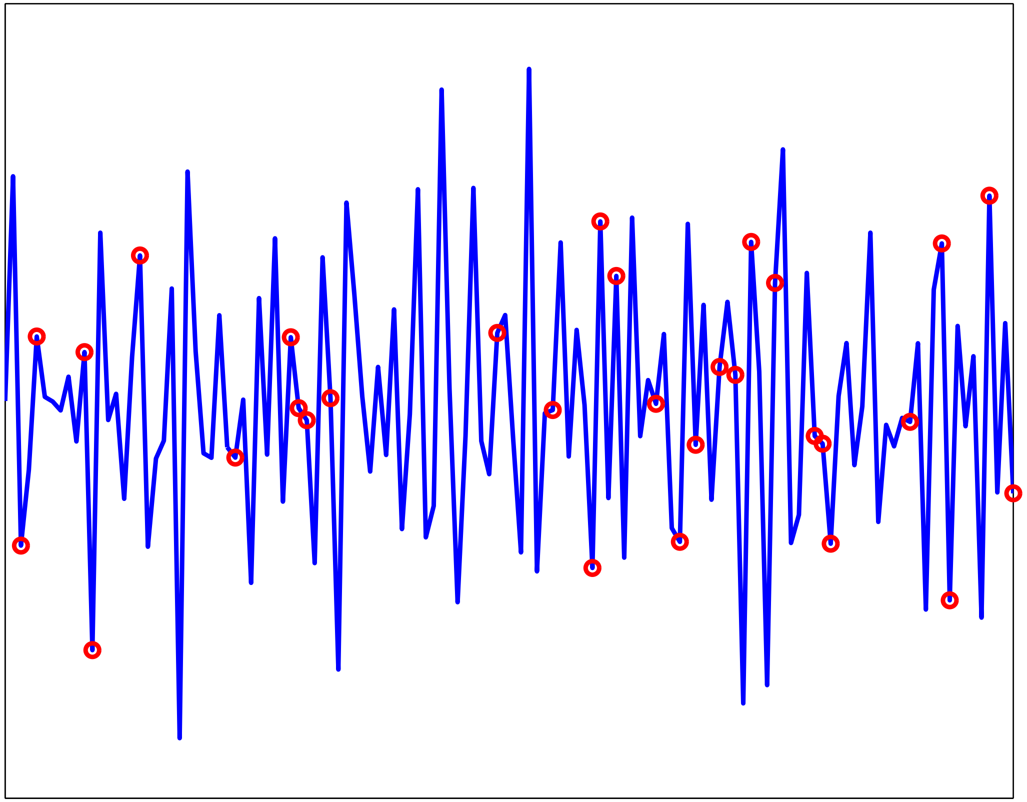
$$\hat{f}(\omega) = \sum_{i=1}^{K} \alpha_i \delta(\omega_i - \omega) \quad \Leftrightarrow \quad f(t) = \sum_{i=1}^{K} \alpha_i e^{i\omega_i t}$$

  f is a superposition of K complex sinusoids

- Recall: frequencies $\{\omega_i\}$ and amplitudes $\{\alpha_i\}$ are *unknown*.
- Take M samples of f at locations $t_1, \ldots, t_M$

Sampling Example

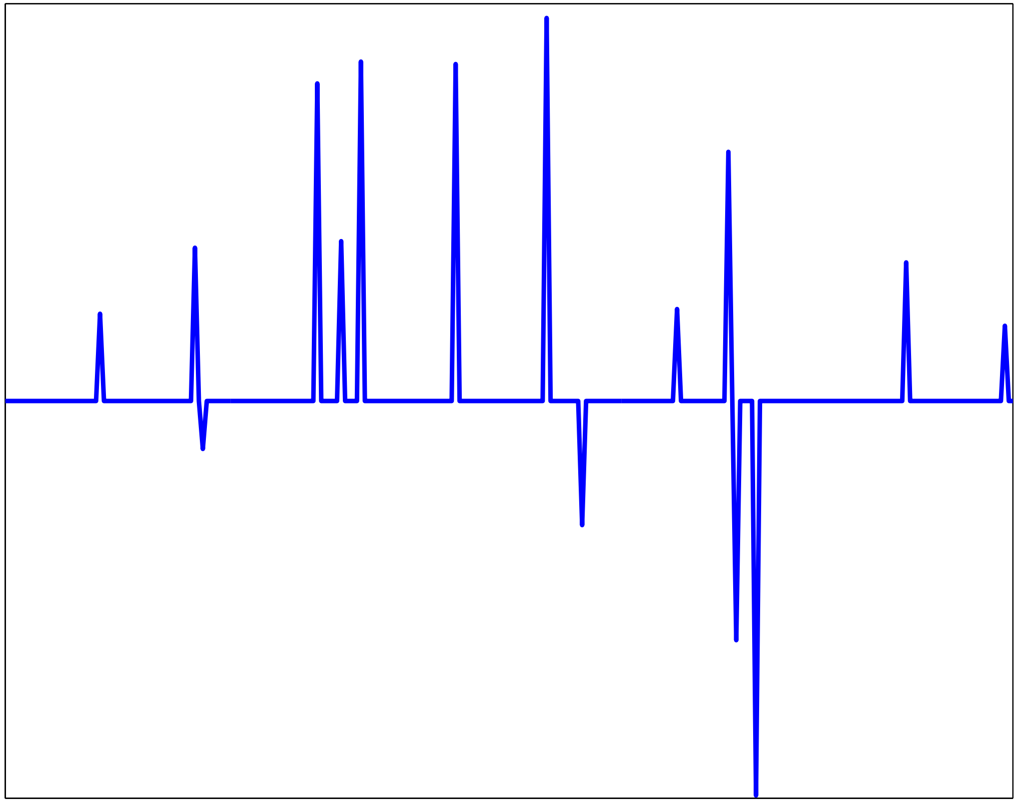
Time domain $f(t)$



Measure M samples

(red circles = samples)

Frequency domain $\hat{f}(\omega)$



K nonzero components

$\#\{\omega : \hat{f}(\omega) \neq 0\} = K$

A Nonlinear Sampling Theorem

- Suppose $\hat{f} \in \mathbb{C}^N$ is supported on set of size K
- Sample at m locations $t_1, \ldots, t_M$ in time-domain
- For the vast majority of sample sets of size

$$M \gtrsim K \cdot \log N$$

  solving

$$\min_g \|\hat{g}\|_{\ell_1} \quad \text{subject to} \quad g(t_m) = y_m, \quad m = 1, \ldots, M$$

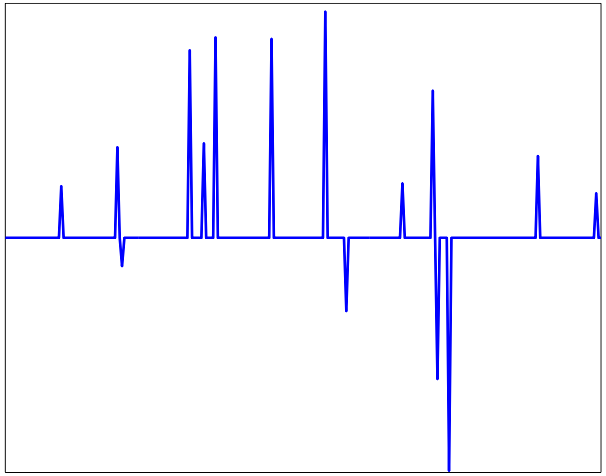
  recovers f *exactly*

- In theory, $\mathrm{Const} \approx 20$
- In practice, perfect recovery occurs when $M \approx 2K$ for $N \approx 1000$.
- *# samples required* $\approx$ *# active components*
- Important frequencies are "discovered" during the recovery

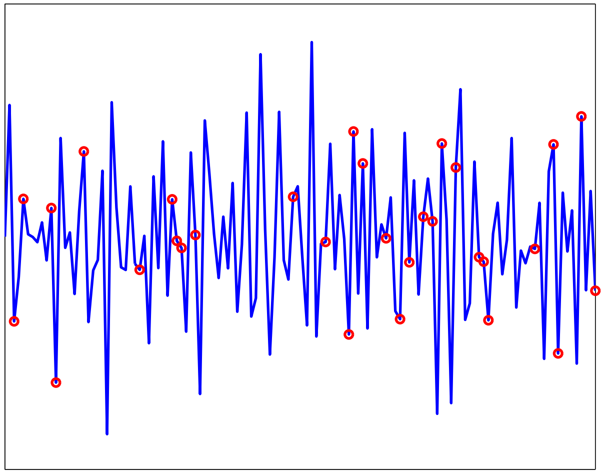
ℓ_1 Reconstruction

Reconstruct by solving

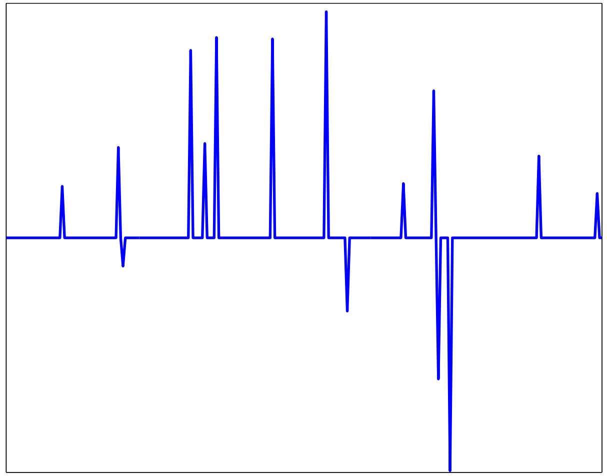
$$\min_{g} \|\hat{g}\|_{\ell_1} := \min \sum_{\omega} |\hat{g}(\omega)| \quad \text{subject to} \quad g(t_m) = f(t_m), \; m = 1, \ldots, M$$



original $\hat{f}$, $K = 15$



given $M = 30$ time-dom. samples



perfect recovery

Nonlinear sampling theorem

- $\hat{f} \in \mathbb{C}^N$ supported on set Ω in Fourier domain
- Shannon sampling theorem:
  - Ω is a known connected set of size K
  - exact recovery from K equally spaced time-domain samples
  - linear reconstruction by sinc interpolation
- Nonlinear sampling theorem:
  - Ω is an *arbitrary and unknown* set of size K
  - exact recovery from $\sim K \log N$ (almost) arbitrarily placed samples
  - nonlinear reconstruction by convex programming

Transform Domain Recovery

- Sparsity basis $\boldsymbol{\Psi}$ (e.g. wavelets)
- Reconstruct by solving

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha} = \boldsymbol{y}$$

- Need measurement to be incoherent in the $\boldsymbol{\Psi}$ domain
  - Random Gaussian: still incoherent (exactly the same)
  - Random binary: still incoherent
  - General rule: just make $\boldsymbol{\Phi}$ unstructured wrt $\boldsymbol{\Psi}$

Random Sensing "Acquisition Theorem"

- Signal/image $f \in \mathbb{C}^N$ is K-sparse in Ψ domain
- Take

$$M \gtrsim K \cdot \log N$$

measurements

$$y_1 = \langle f, \phi_1 \rangle, \ \ldots, \ y_M = \langle f, \phi_M \rangle$$

$\phi_m =$ random waveform

- Then solving

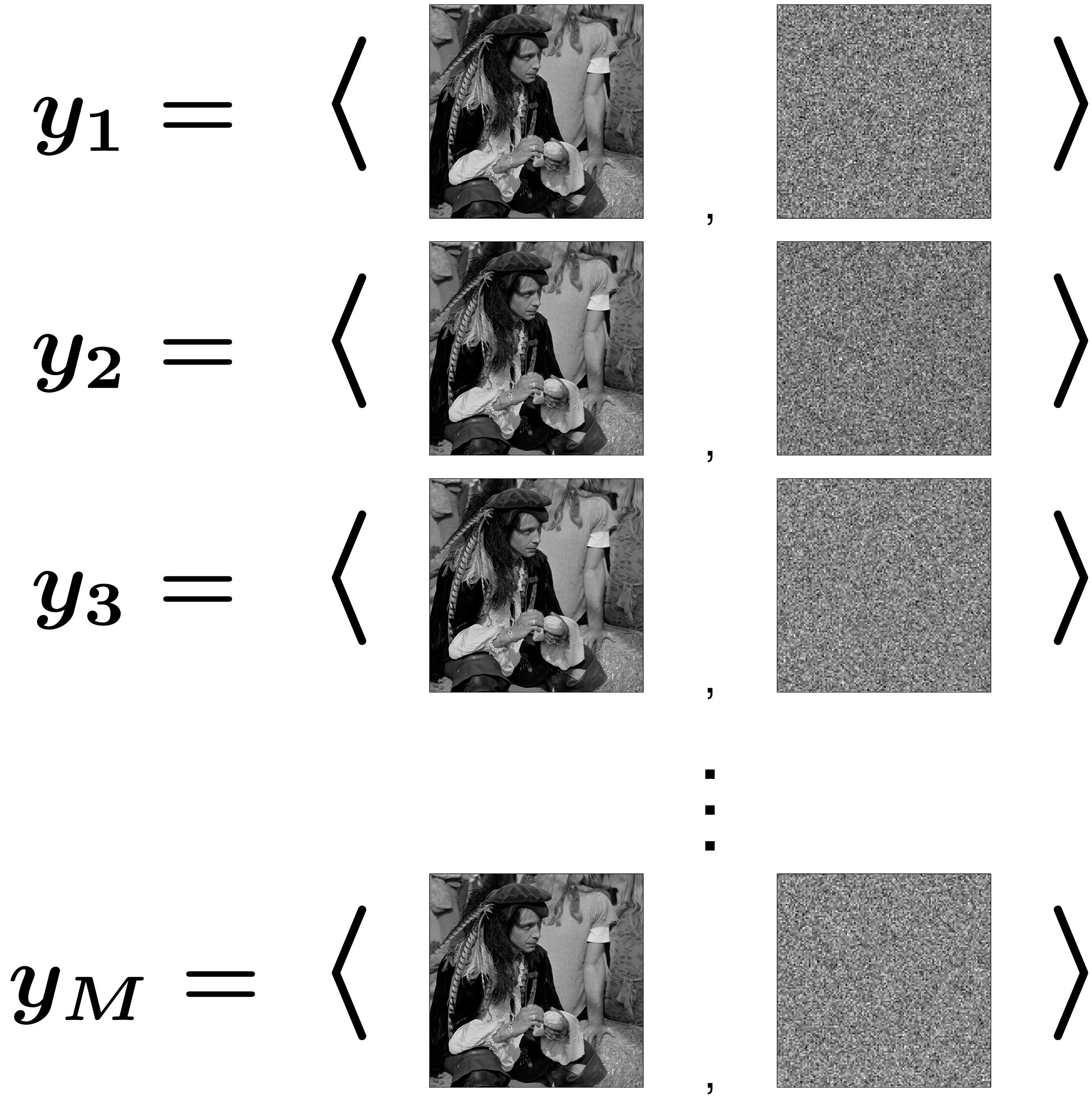
$$\min_{\alpha} \|\alpha\|_{\ell_1} \quad \text{subject to} \quad \Phi\Psi\alpha = y$$

will recover (the transform coefficients) of f exactly

- In practice, it seems that

$$M \approx 5K$$

measurements are sufficient


$y_1 =$
$y_2 =$
$y_3 =$
$y_M =$

Example: Sparse Image

- Take $M = 100,000$ incoherent measurements $y = \Phi f_a$
- f_a = wavelet approximation (perfectly sparse)
- Solve

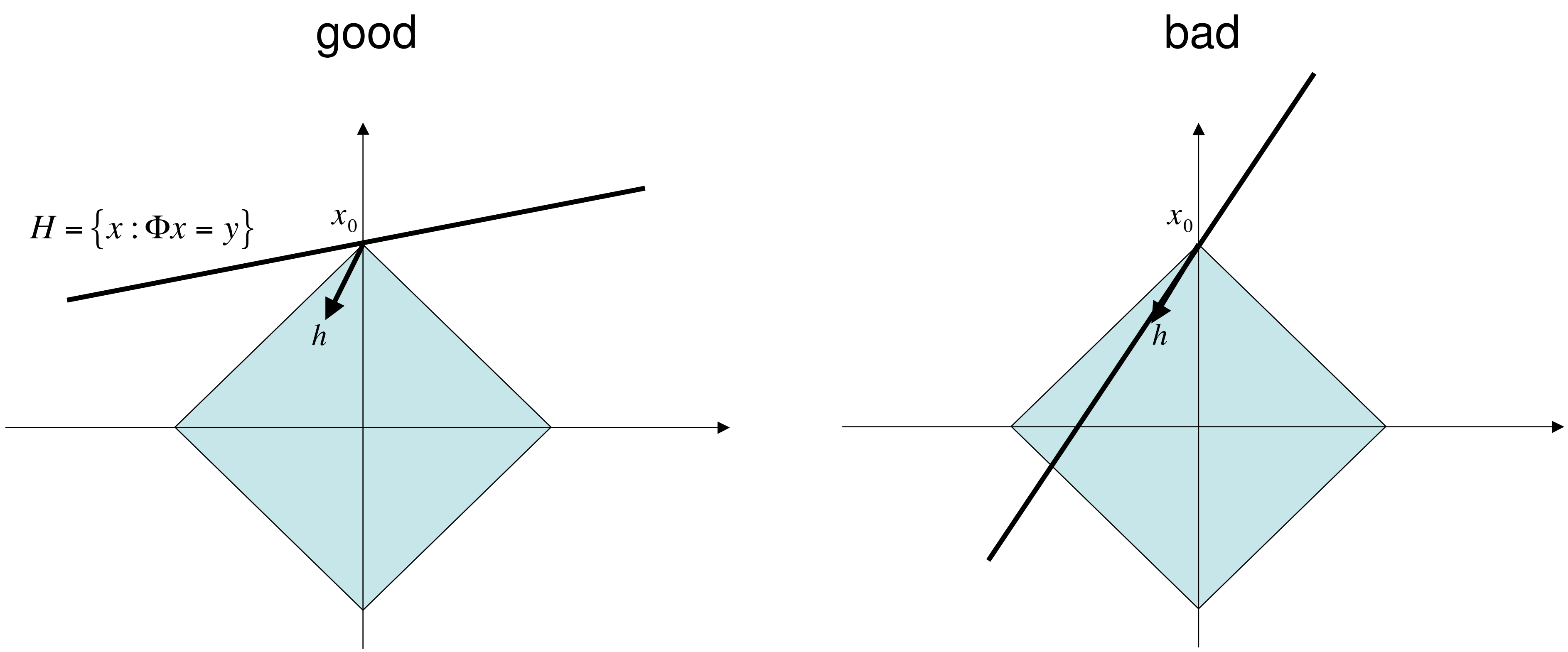
$$\min \|\alpha\|_{\ell_1} \quad \text{subject to} \quad \Phi\Psi\alpha = y$$

Ψ = wavelet transform

original (25k wavelets)

perfect recovery

Geometrical Viewpoint



- Consider and "ℓ_1-descent vectors" h for feasible f_0:

$$\|f_0 + h\|_{\ell_1} \; < \; \|f_0\|_{\ell_1}$$

- f_0 is the solution if

$$\Phi h \neq 0$$

  for all such descent vectors

Stability

- Real images are not exactly sparse
- For Φ' obeying UUP for sets of size $4K$, and *general* α, recovery obeys

$$\|\alpha_0 - \alpha^*\|_2 \lesssim \frac{\|\alpha_0 - \alpha_{0,K}\|_{\ell_1}}{\sqrt{K}}$$

  $\alpha_{0,K}$ = best K-term approximation

- Compressible: if transform coefficients decay

$$|\alpha_0|_{(m)} \lesssim m^{-r}, \quad r > 1$$

  $|\alpha_0|_{(m)} = m$th largest coefficient, then

$$\begin{aligned} \|\alpha_0 - \alpha_{0,K}\|_2 &\lesssim K^{-r+1/2} \\ \|\alpha_0 - \alpha^*\|_2 &\lesssim K^{-r+1/2} \end{aligned}$$

- *Recovery error* $\sim$ *adaptive approximation error*

Stability

- What if the measurements are noisy?

$$y = \Phi' \alpha_0 + e, \qquad \|e\|_2 \leq \epsilon$$

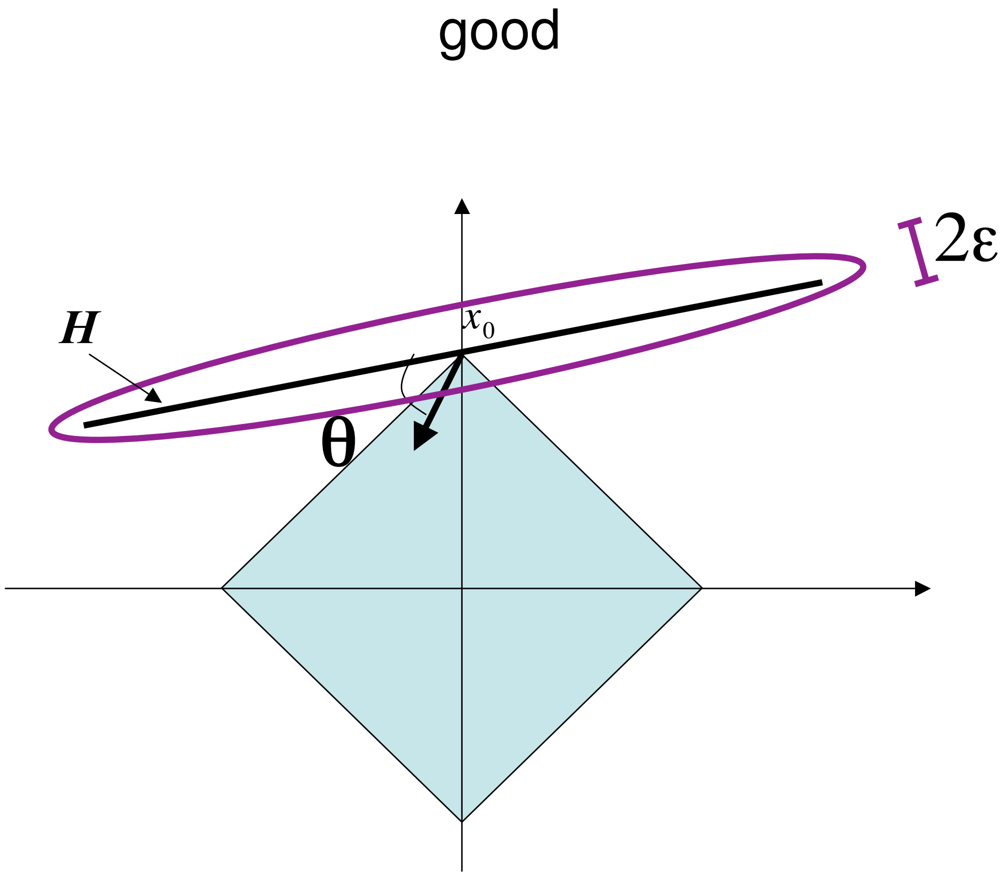
- *Relax* the recovery program; solve

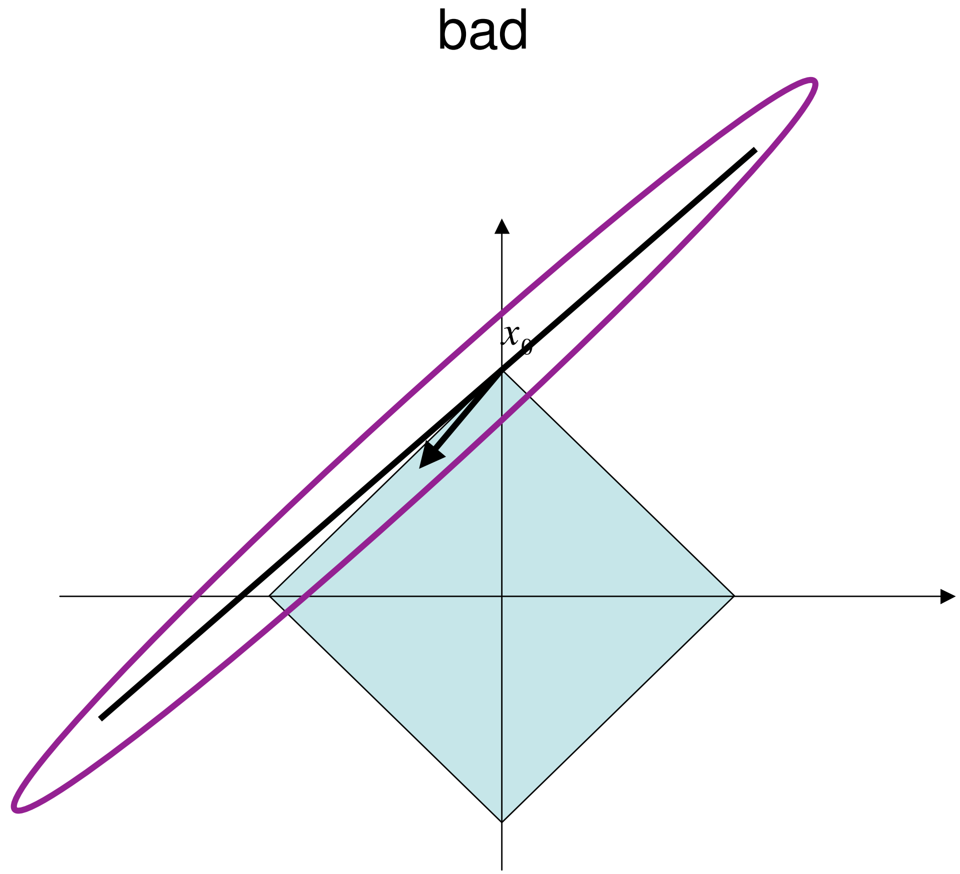
$$\min_{\alpha} \; \|\alpha\|_{\ell_1} \quad \text{subject to} \quad \|\Phi' \alpha - y\|_2 \leq \epsilon$$

- The recovery error obeys

$$\|\alpha_0 - \alpha^*\|_2 \;\; \lesssim \;\; \sqrt{\frac{N}{M}} \cdot \epsilon \;\; + \;\; \frac{\|\alpha_0 - \alpha_{0,K}\|_{\ell_1}}{\sqrt{K}}$$

measurement error + *approximation error*





- Solution will be within ϵ of $\boldsymbol{H}$
- Need that not too much of the ℓ_1 ball near f_0 is feasible

Compressed Sensing

- As # measurements increases, error decreases at near-optimal rate

$$\text{best } M\text{-term approximation}: \quad \|\alpha_0 - \alpha_{0,M}\|_2 \lesssim M^{-r}$$

$$\Rightarrow \quad \text{CS recovery}: \quad \|\alpha_0 - \alpha_M^*\|_2 \lesssim (M/\log N)^{-r}$$

- The sensing is *not adaptive*, and is simple
- Compression "built in" to the measurements
- Taking random measurements $=$ universal, analog coding scheme for sparse signals

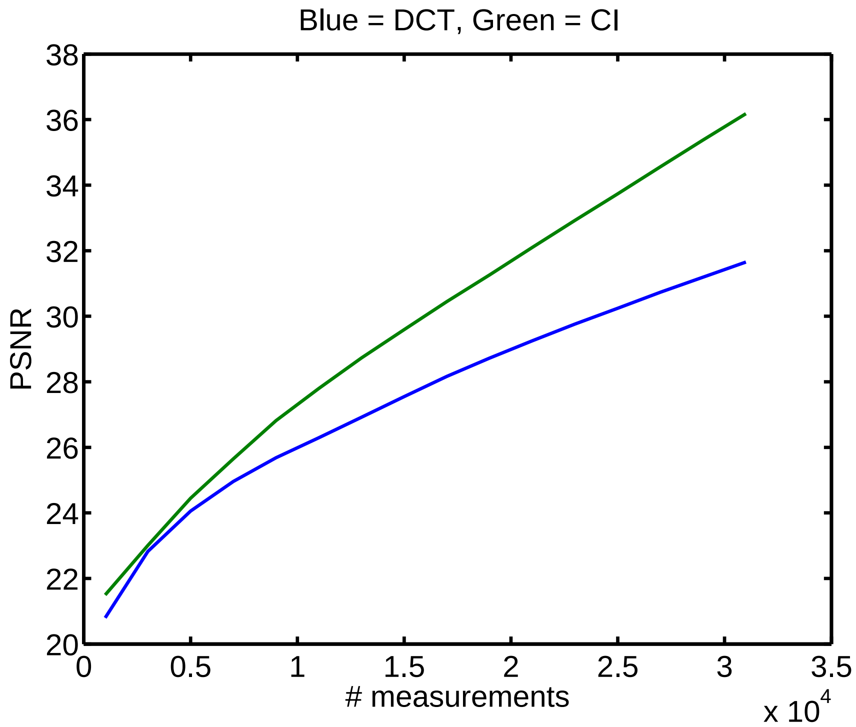
Compressed Sensing

- As # measurements increases, error decreases at near-optimal rate
- Democratic and robust:
  - all measurement are equally (un)important
  - losing a few does not hurt
- The recovery is flexible, and independent of acquisition

$$\min \ \|\boldsymbol{\alpha}\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{\Phi\Psi\alpha} = \boldsymbol{y}$$

  Different $\boldsymbol{\Psi}$ yield different recoveries from same measurements
- Use *a posteriori* computing power to reduce *a priori* sampling complexity

Compressive Imaging: Stylized Performance



- Blue line: DCT measurements in JPEG order
  Green line: lowpass + random measurements
- Not only is Compressive Imaging better, it is getting better faster!
- CI is much better at “filling in the details”

Recovery Techniques for CS

- ℓ_1 minimization
- Matching Pursuit
- Iterative thresholding
- Total-variation minimization

Recovery Techniques for CS: ℓ_1 minimization

$$\min_{x} \ \|x\|_{\ell_1} \quad \text{subject to} \quad \Phi x = y$$

- Recovery program is convex (linear or second-order cone program)
- Related to the LASSO from statistics
- Optimization algorithms:
  - interior point methods (slow, but extremely accurate)
  - first-order "gradient projection" (fast)
  - homotopy methods (fast and accurate for small-scale problems)
- Accurate and robust for recovering sparse signals;
  needs to be "tweaked" for things like images

Other flavors of ℓ_1

- Quadratic relaxation (LASSO)

$$\min_x \ \|x\|_{\ell_1} \quad \text{subject to} \quad \|\Phi x - y\|_2 \le \epsilon$$

- Dantzig selector (residual correlation constraints)

$$\min_x \ \|x\|_{\ell_1} \quad \text{subject to} \quad \|\Phi^T(\Phi x - y)\|_\infty \le \epsilon$$

- ℓ_1 analysis (Ψ is an overcomplete frame)

$$\min_x \ \|\Psi^T x\|_{\ell_1} \quad \text{subject to} \quad \|\Phi x - y\|_2 \le \epsilon$$

Recovery Techniques for CS: Matching Pursuit

- Essential algorithm:
  1. Choose the first "active" component by seeing which column of $\boldsymbol{\Phi}$ is most correlated with $\boldsymbol{y}$
  2. Subtract off of $\boldsymbol{y}$ to form the residual $\boldsymbol{y'}$
  3. Repeat with $\boldsymbol{y'}$
- Help to orthogonalize the "active set" between iterations
- Very fast for small scale problems
- Not as accurate/robust for large signals in the presence of noise

Recovery Techniques for CS: Iterative thresholding

- Essential algorithm:
  1. From guess x_k, backproject to get $w_k \approx \Phi^T \Phi x_k$
  2. Threshold/prune w_k to get x_{k+1}.
- Works since for sparse signals w_k will be big on the active set, and small elsewhere
- Analogous theory to ℓ_1 minimization
- Very fast; works very well for recovering sparse signals, works OK for recovering approximately sparse signals

Recovery Techniques for CS: TV minimization

$$\min_{\boldsymbol{x}} \ \mathbf{TV}(\boldsymbol{x}) \approx \|\boldsymbol{\nabla x}\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{\Phi x} = \boldsymbol{y}$$

- Popular formulation from variational image processing
- Sparsity = number of “jumps” in the image
- Convex program, so can be solved with interior point methods, or some kind of first-order gradient projection
- Accurate and robust (but can be slow) for recovering images

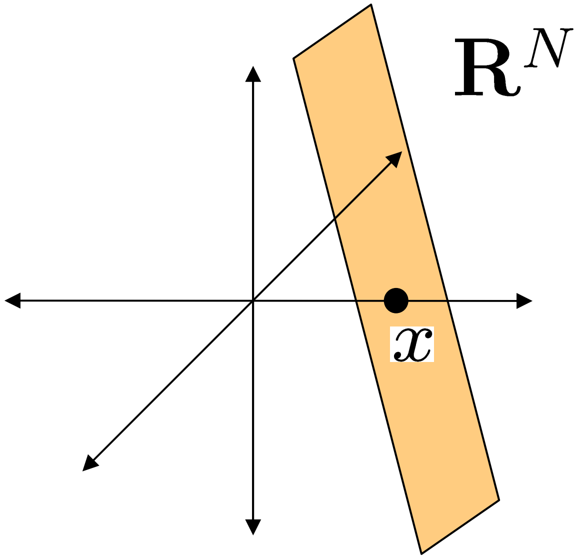
Agenda

- Introduction to Compressive Sensing (CS) [richb]
  - motivation
  - basic concepts
- CS Theoretical Foundation [justin]
  - uniform uncertainty principles
  - restricted isometry principle
  - recovery algorithms
- **Geometry of CS** [mike]
  - *K*-sparse and compressible signals
  - manifolds
- CS Applications [richb, justin]

Geometry
of
Compressive Sensing

Geometry in CS

- Major *geometric themes*:
  - where signals live in ambient space
    - before and after projection
  - implications of sparse models
  - mechanics of l_1 recovery

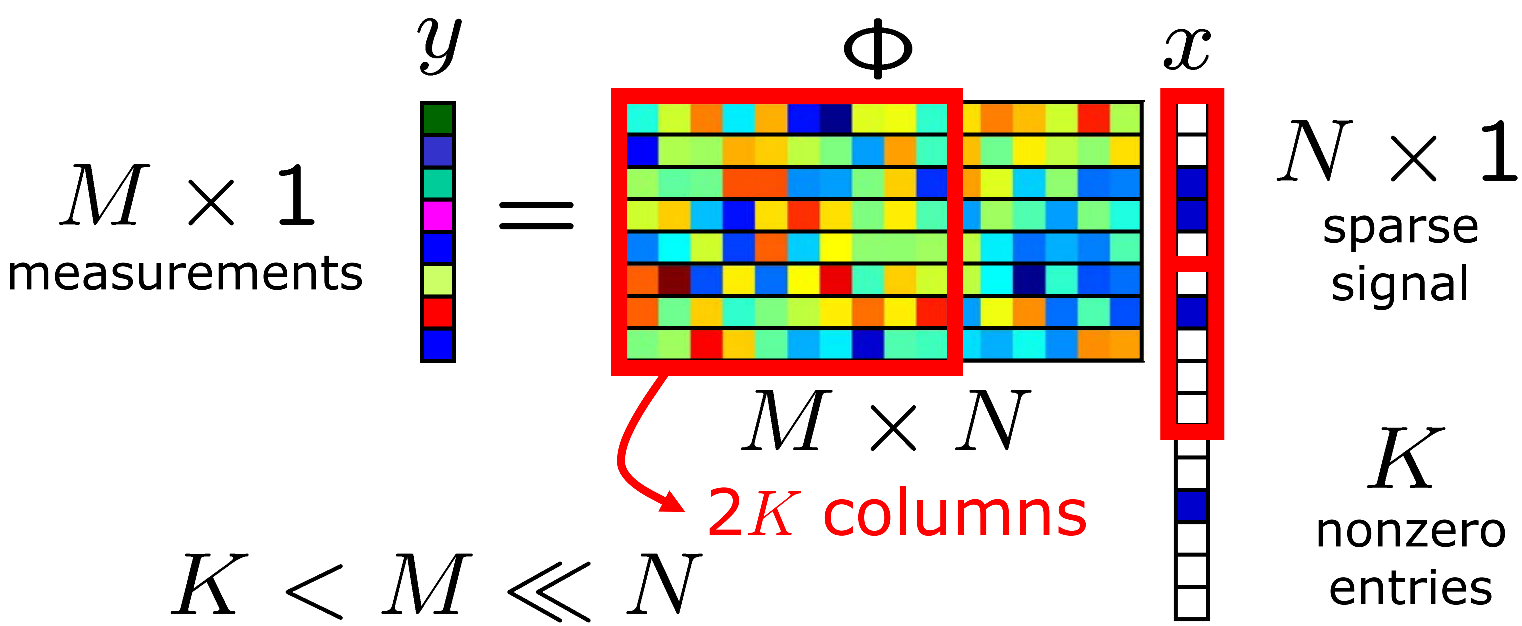


- Important questions:
  - how and why can signals be recovered?
  - how many measurements are really needed?
  - how can all this be extended to other signal models?

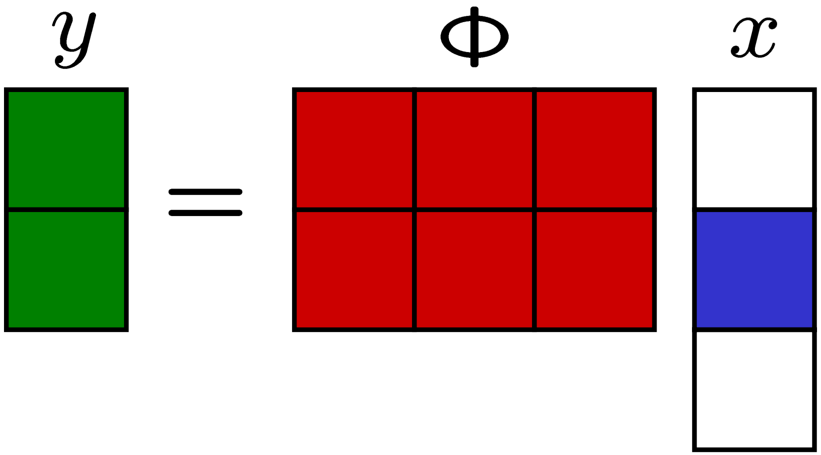
Sparsity-Based CS

One Simple Question

- When is it possible to distinguish K-sparse signals?
  - require $\Phi x_1 \neq \Phi x_2$ for all K-sparse $x_1 \neq x_2$
- Necessary: Φ must have at least 2K rows
  - otherwise there exist K-sparse x_1, x_2 s.t. $\Phi(x_1 - x_2) = 0$
- Sufficient: Gaussian Φ with 2K rows

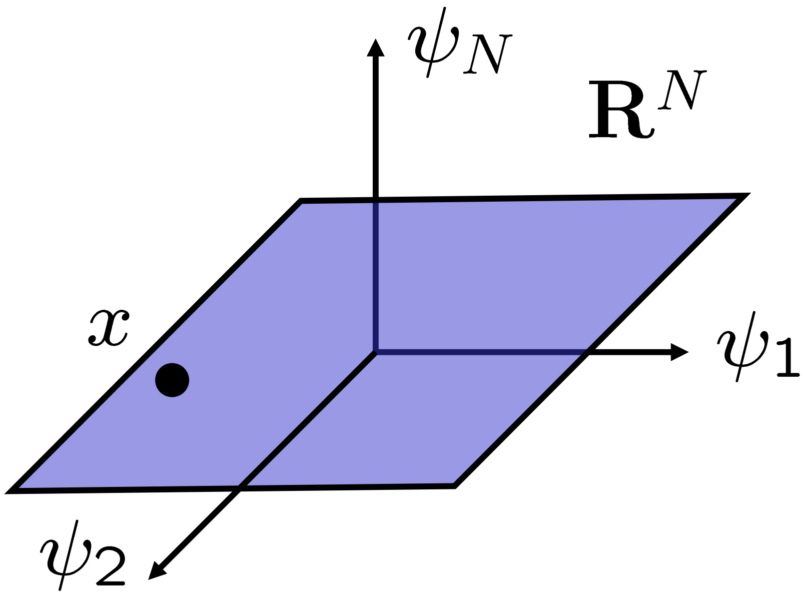


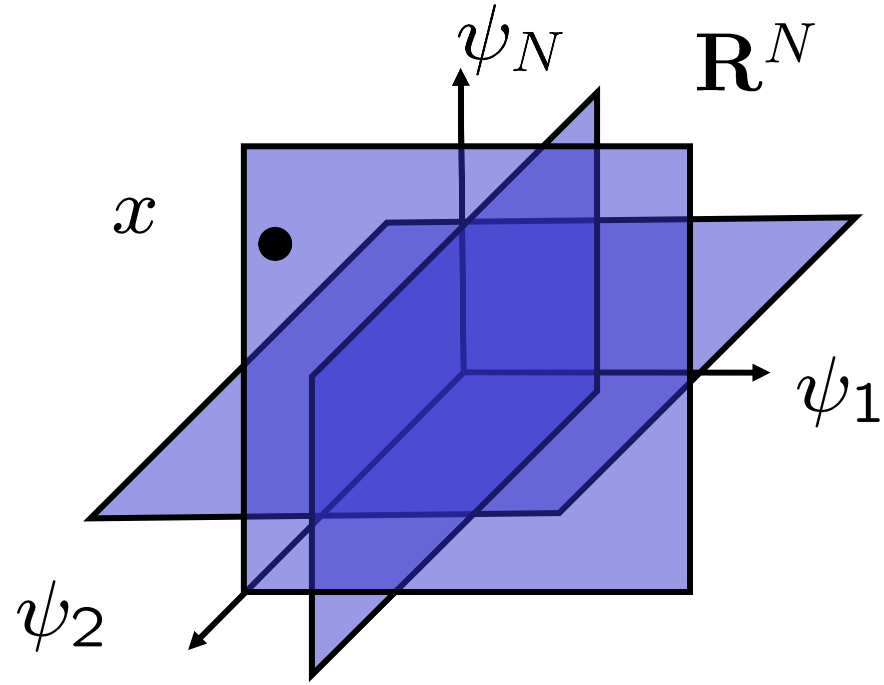
Illustrative Example



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

Geometry of Sparse Signal Sets





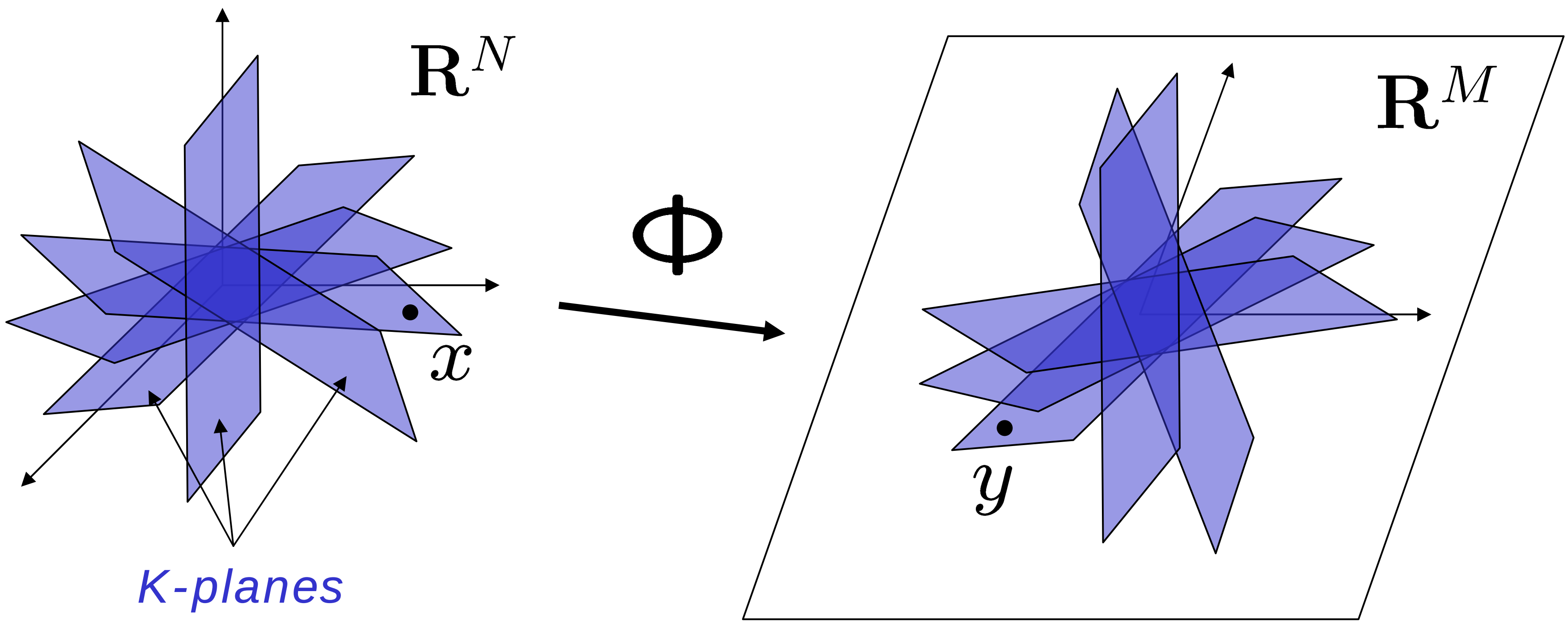
Linear

K-plane

Sparse, Nonlinear

Union of K-planes

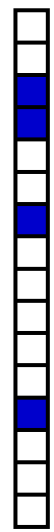
Geometry: Embedding in R^M



- Φ(K-plane) = K-plane in general
- M ≥ 2K measurements
  - necessary for injectivity
  - sufficient for injectivity when Φ Gaussian
  - but not enough for efficient, robust recovery
- See also FROI [Vetterli et al., Lu and Do]

The Geometry of L_1 Recovery

x

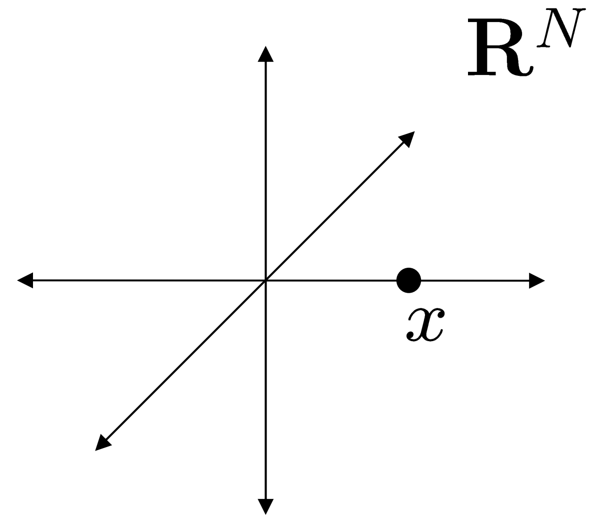


$N \times 1$

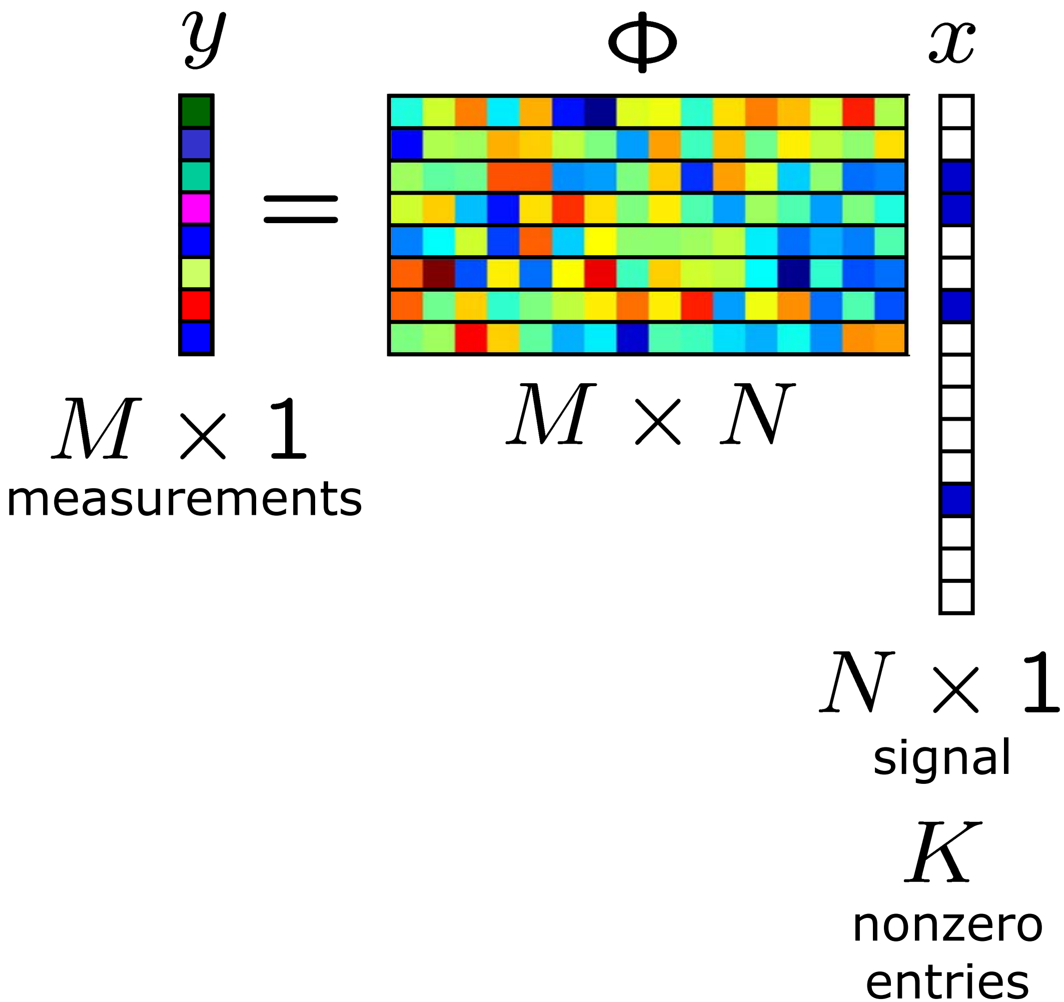
signal

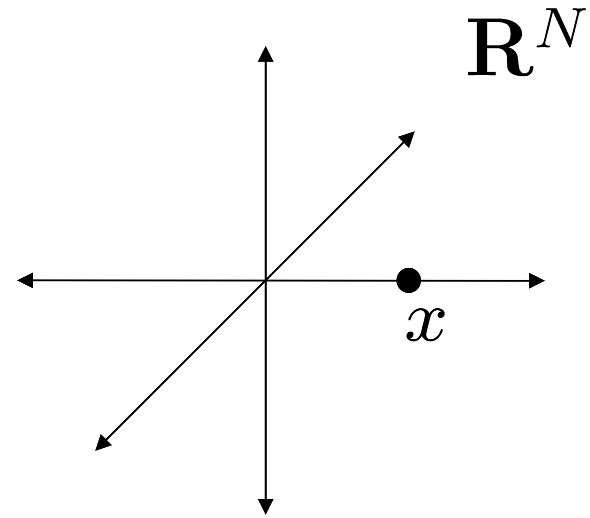
K

nonzero
entries

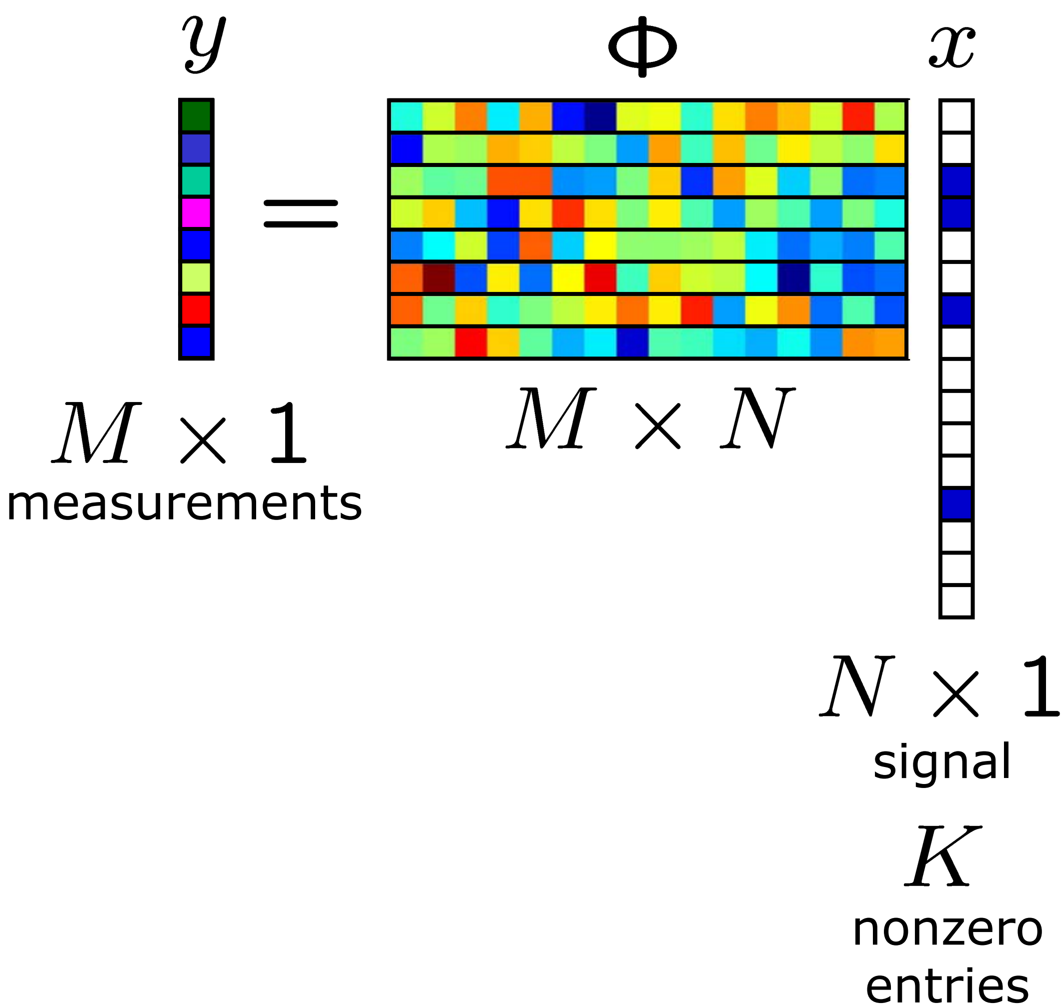


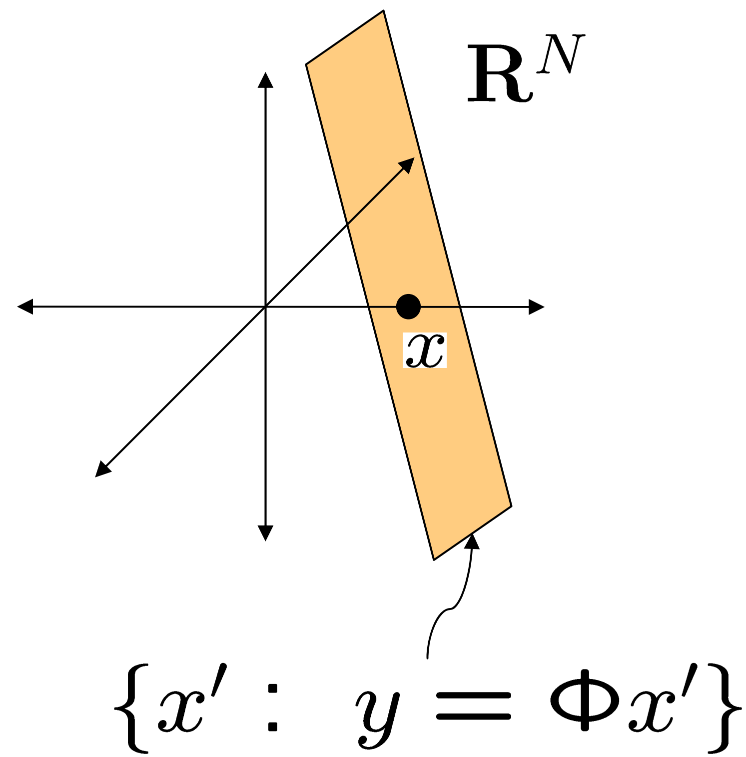
The Geometry of L_1 Recovery



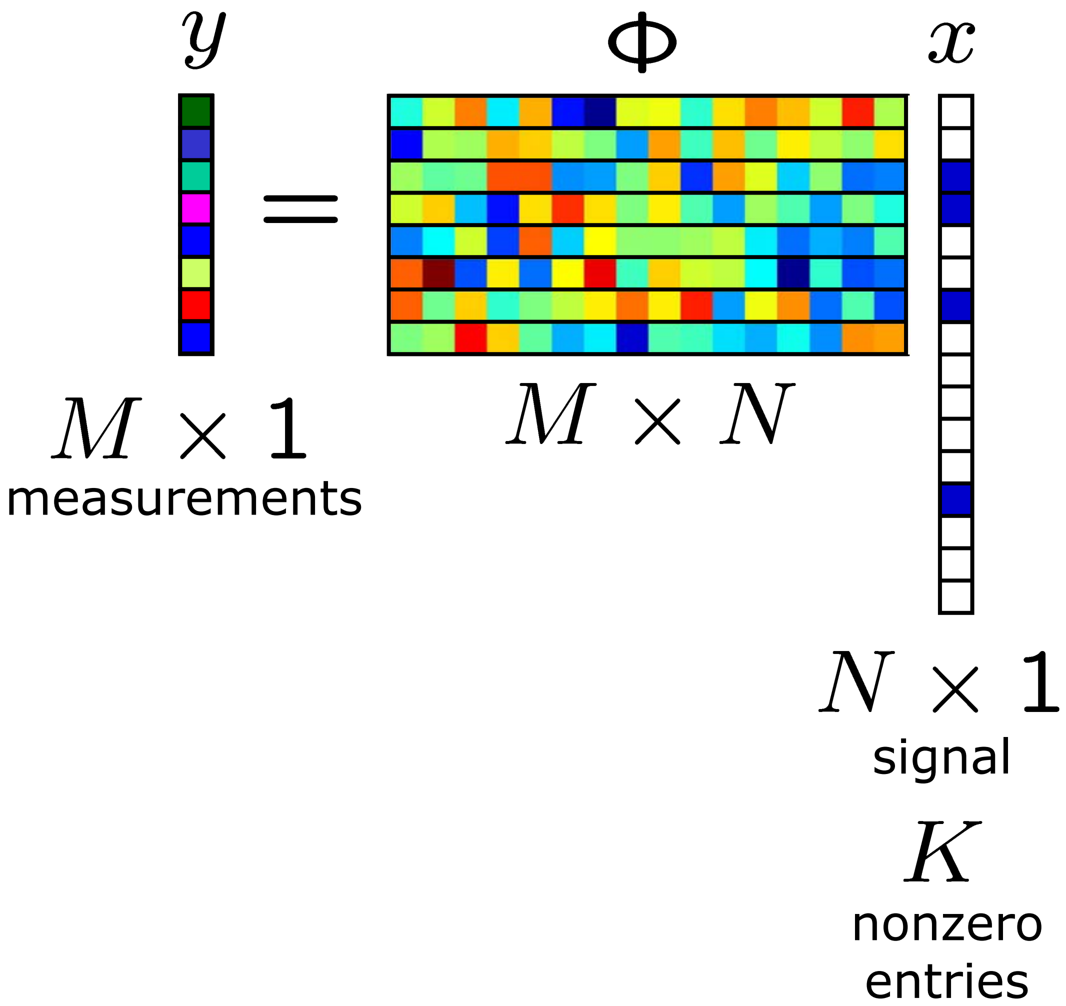


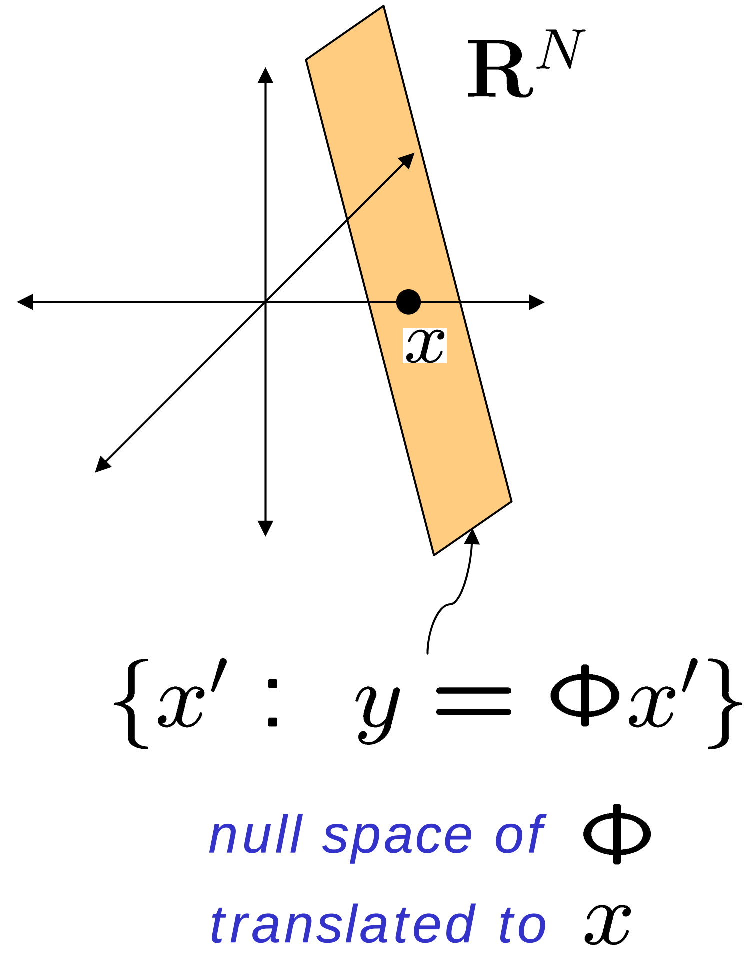
The Geometry of L_1 Recovery



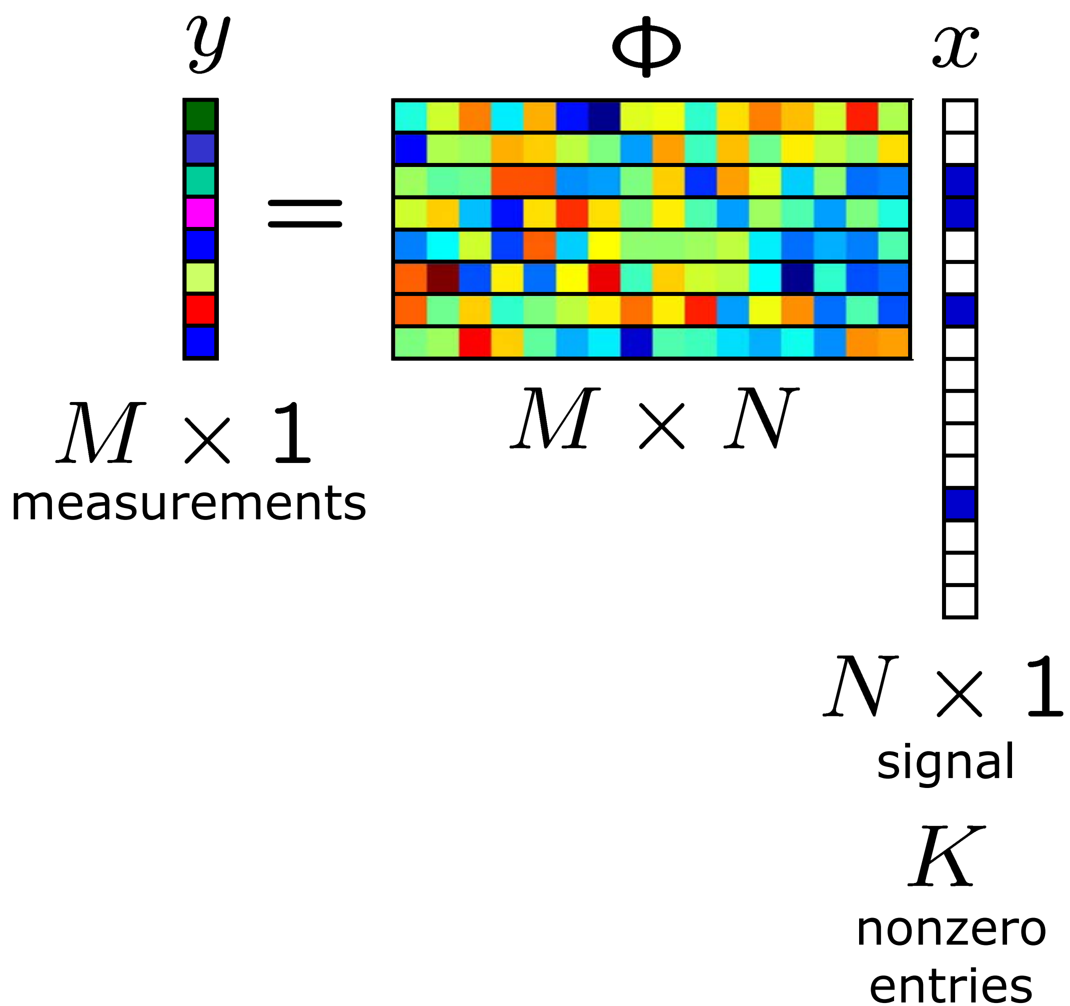


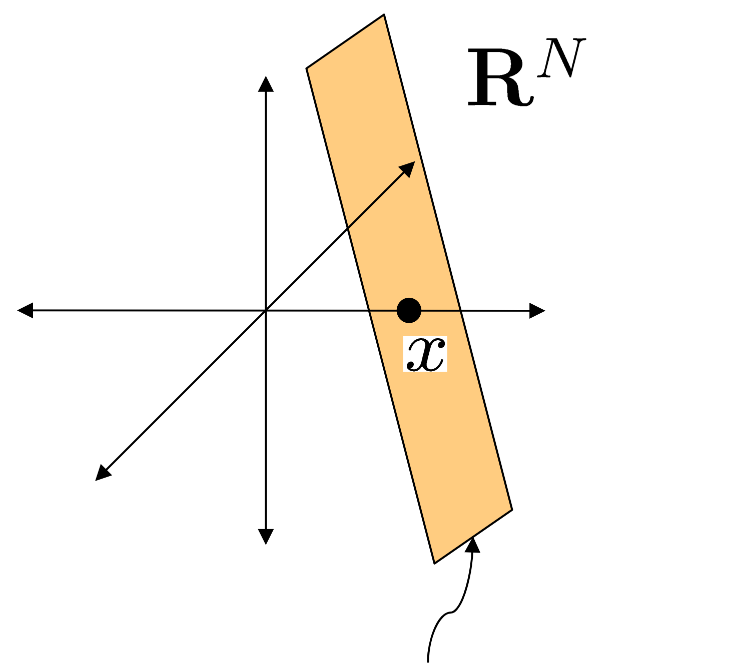
The Geometry of L_1 Recovery





The Geometry of L_1 Recovery





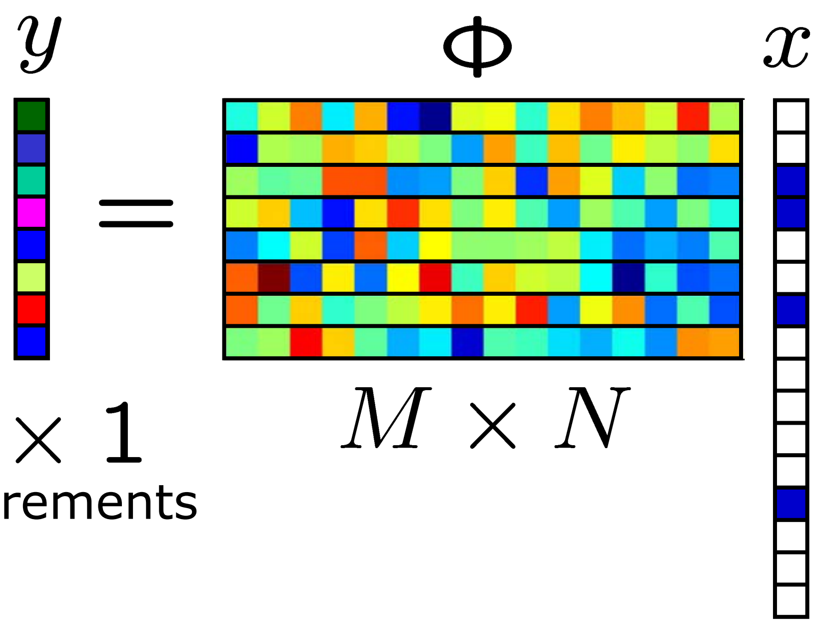
$$\{x' : \ y = \Phi x'\}$$

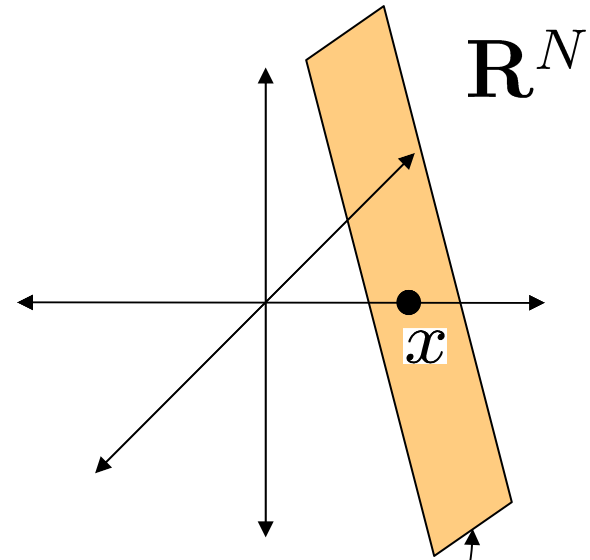
null space of Φ

translated to x

dimension N-M

The Geometry of L_1 Recovery





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

null space of Φ

translated to x

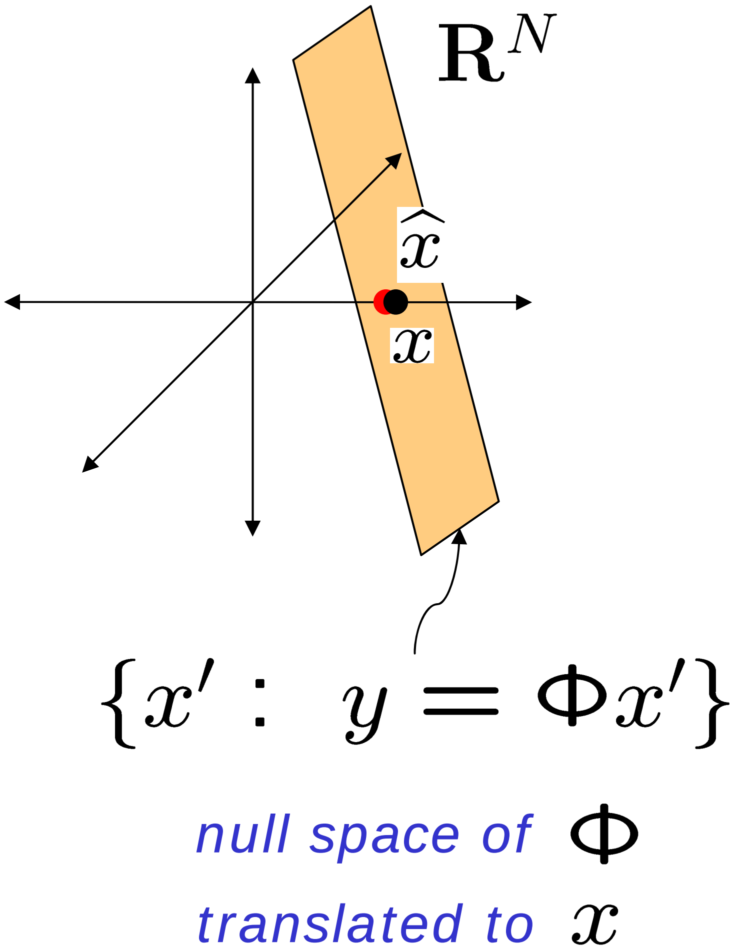
dimension N-M
random orientation

L_0 Recovery Works

$$\widehat{x} = \arg\min_{y=\Phi x'} \|x'\|_0$$

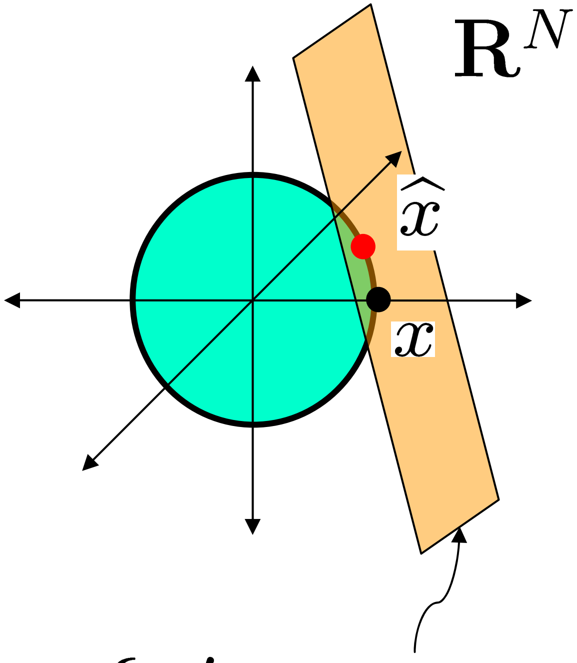
minimum L_0 solution correct
if $M \geq 2K$

(w.p. 1 for Gaussian Φ)



Why L_2 Doesn't Work

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



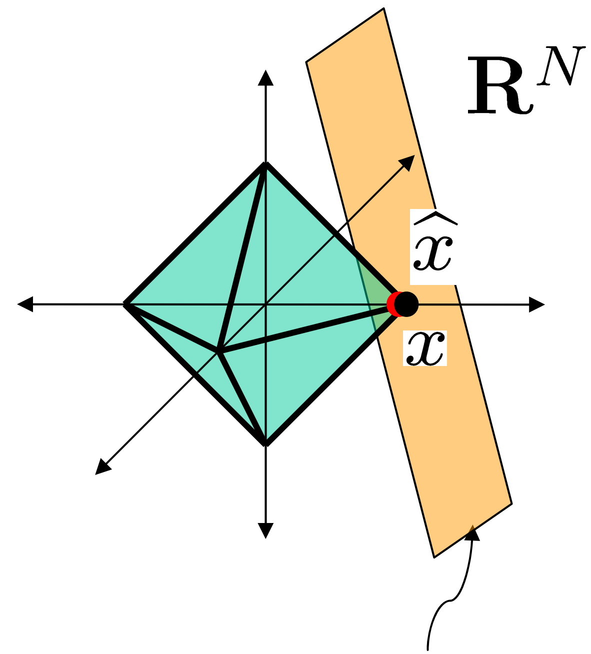
least squares,
minimum L_2 solution
is almost **never sparse**

Why L_1 Works

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

minimum L_1 solution
= L_0 sparsest solution if

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

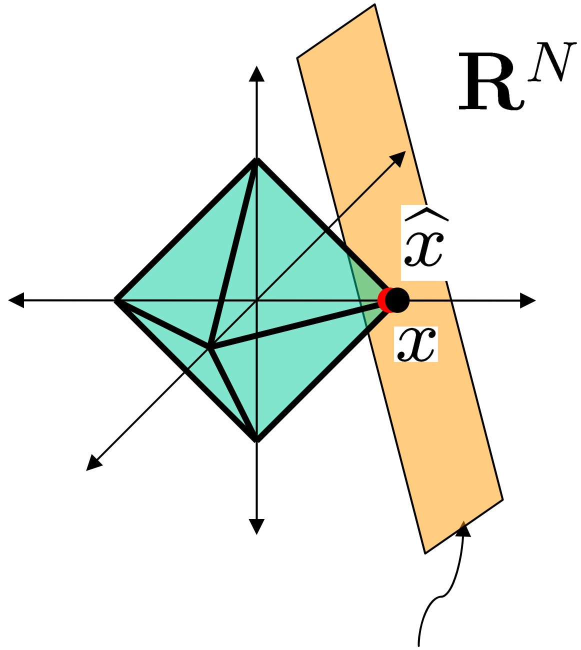


$$\{x' : \ y = \Phi x'\}$$

random orientation
dimension N-M

Why L_1 Works

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



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

random orientation
dimension N-M

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.

Want K small, (N-M) small
(i.e., M large)

L_0/L_1 Equivalence
[Donoho, Tanner]

Theorem.

For Gaussian Φ, require

$$M \sim 2eK \log\left(\frac{N}{M\sqrt{\pi}}\right)$$

measurements to recover every K-sparse signal and

$$M \sim 2K \log\left(\frac{N}{M}\right)$$

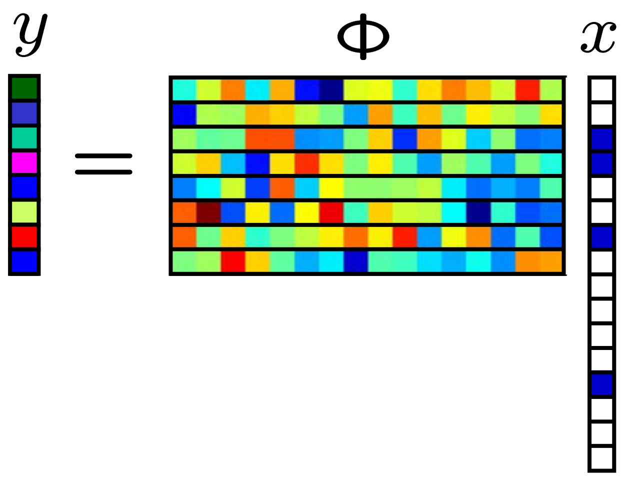
measurements to recover a large majority of K-sparse x. (These bounds are sharp asymptotically.)

Proof (geometric): Face-counting of randomly projected polytopes

Restricted Isometry Property (aka UUP)

[Candès, Romberg, Tao]

- Measurement matrix Φ has ***RIP of order K*** if



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

for all K-sparse signals x.

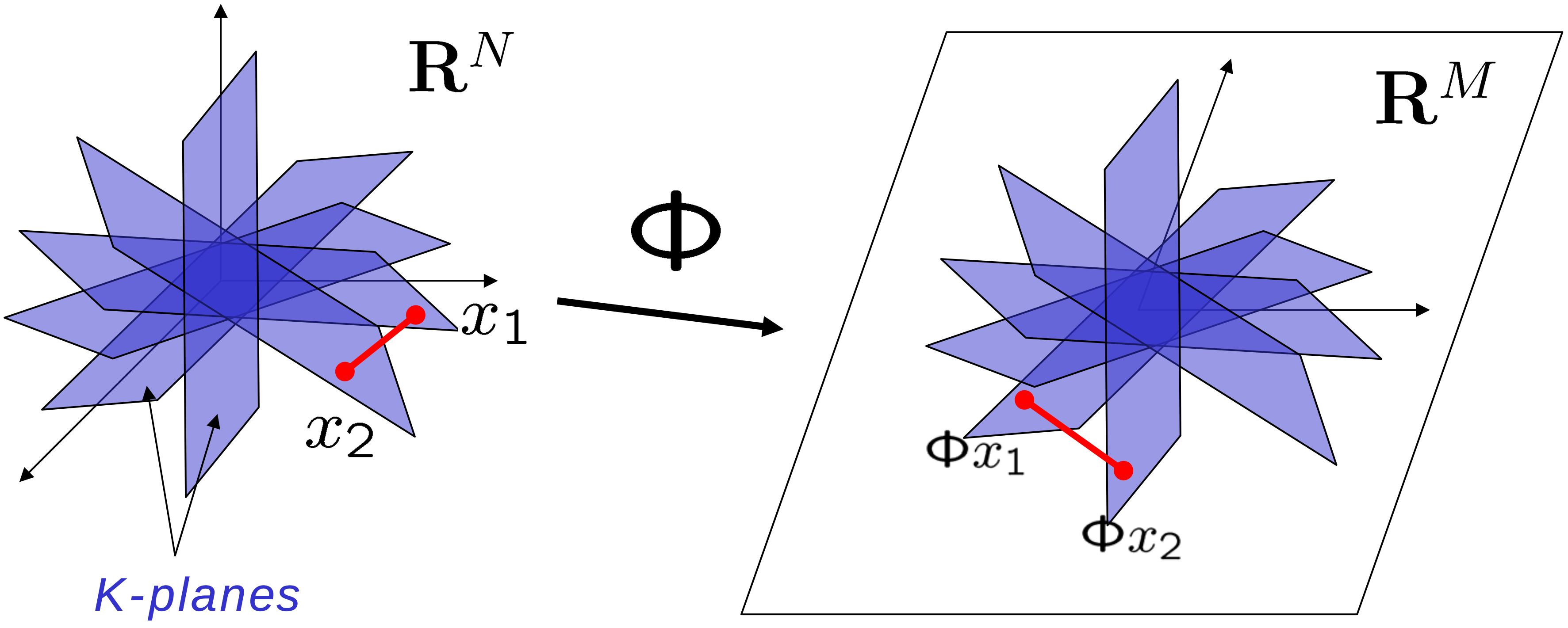
- Does *not* hold for $K > M$; may hold for smaller K.

- Implications: tractable, stable, robust recovery

RIP as a "Stable" Embedding

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

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



(if $\delta_{2K} < 1$ have injectivity; smaller δ_{2K} more stable)

Implications of RIP

[Candès (+ et al.); see also Cohen et al., Vershynin et al.]

If $\delta_{2K} < 0.41$, ensured:

1. *Tractable recovery: All K-sparse x* are perfectly recovered via ℓ_1 minimization.

2. *Robust recovery: For any $x \in R^N$,*

$$\|x - \widehat{x}\|_{\ell_1} \leq C \|x - x_K\|_{\ell_1} \text{ and } \|x - \widehat{x}\|_{\ell_2} \leq C \frac{\|x - x_K\|_{\ell_1}}{K^{1/2}}.$$

3. *Stable recovery:* Measure $y = \Phi x + e$, with $||e||_2 < \varepsilon$, and recover

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

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

$$\|x - \widehat{x}\|_{\ell_2} \leq C_1 \frac{\|x - x_K\|_{\ell_1}}{K^{1/2}} + C_2 \epsilon.$$

Verifying RIP:
How Many Measurements?

- Want RIP of order 2K (say) to hold for MxN Φ
  - difficult to verify for a given Φ
  - requires checking eigenvalues of each submatrix
- Prove ***random*** Φ will work
  - ***iid Gaussian entries***
  - ***iid Bernoulli entries (+/- 1)***
  - ***iid subgaussian entries***
  - ***random Fourier ensemble***
  - ***random subset of incoherent dictionary***
- In each case, ***M = O(K log N)*** suffices
  - with very high probability, usually $1-O(e^{-CN})$
  - slight variations on log term

Optimality
[Candès; Donoho]

- Gaussian Φ has RIP order 2K (say) with ***M = O(K log(N/M))***

- Hence, for a given M, for $x \in wl_p$ (i.e., $|x|_{(k)} \sim k^{-1/p}$), $0 < p < 1$, (or $x \in l_1$)

$$\begin{aligned} \|x - \widehat{x}\|_{\ell_2} &\leq CK^{-1/2}\|x - x_K\|_{\ell_1} \\ &\leq CK^{1/2-1/p} \\ &\leq C(M/\log(N/M))^{1/2-1/p} \end{aligned}$$

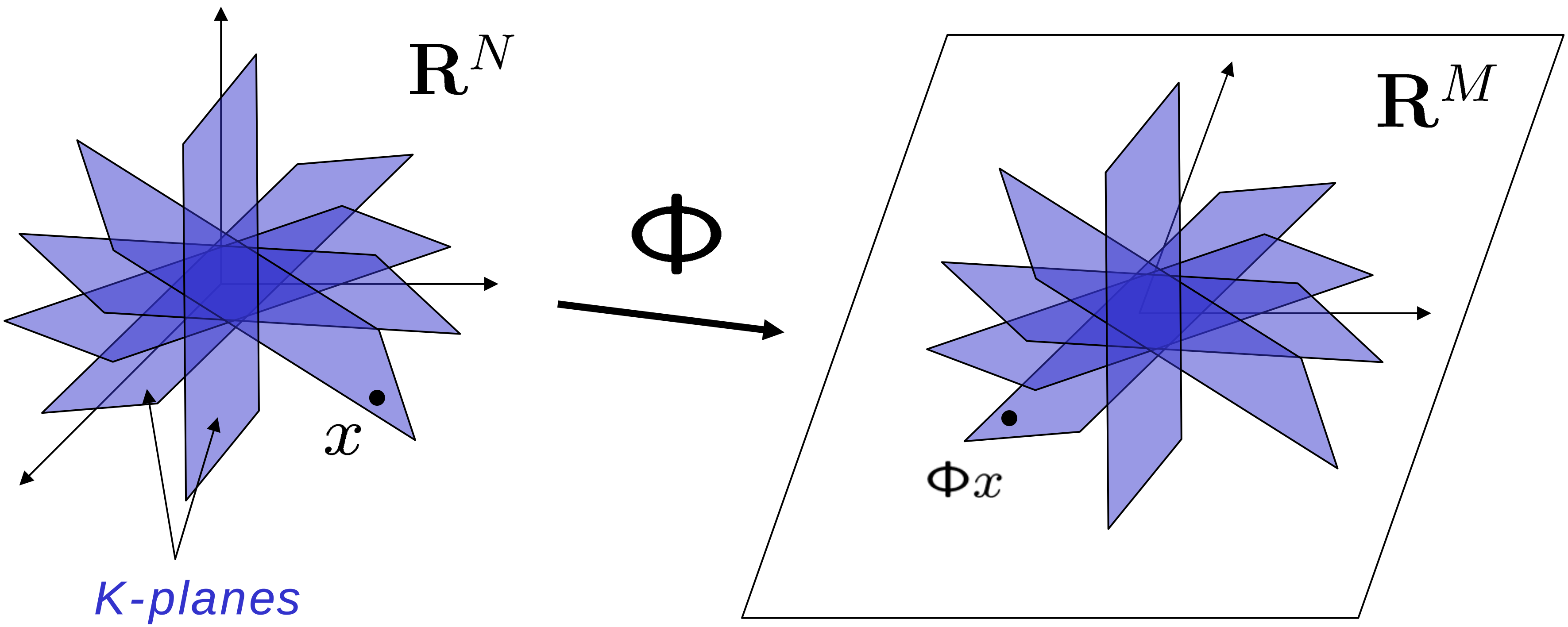
- Up to a constant, these bounds are *optimal*: no other linear mapping to R^M followed by *any* decoding method could yield lower reconstruction error over classes of compressible signals

- <u>Proof (geometric):</u> Gelfand n-widths [Kashin; Gluskin, Garnaev]

Recall: RIP

- RIP of order K requires: for all K-sparse x,

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

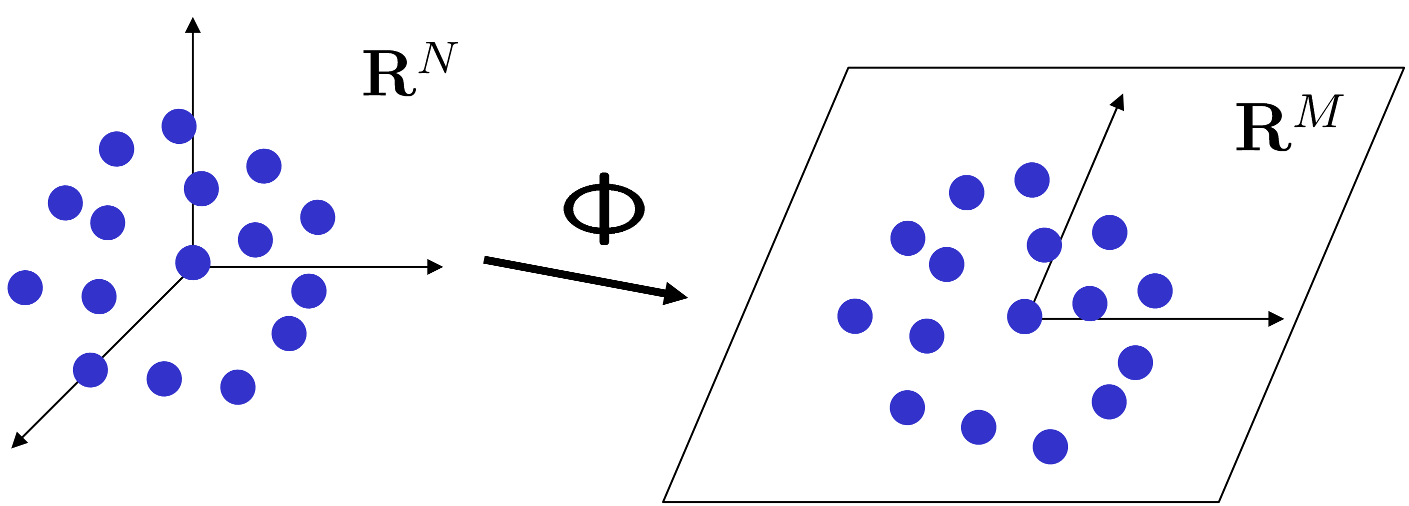


Johnson-Lindenstrauss Lemma

[see also Dasgupta, Gupta; Frankl, Maehara; Achlioptas; Indyk, Motwani]

Consider a point set $Q \subset R^N$ and random* $M \times N$ Φ with $M = O(\log(\#Q)\, \varepsilon^{-2})$. With high prob., for all $x_1, x_2 \in Q$,

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



Proof via *concentration inequality*: For any $x \in R^N$

$$\mathbf{P}(|\, \|\Phi x\|_2^2 - \|x\|_2^2| \geq \epsilon \|x\|_2^2) \leq 2e^{-\frac{M}{2}(\epsilon^2/2 - \epsilon^3/3)}.$$

Favorable JL Distributions

- Gaussian

$$\phi_{i,j} \sim \mathcal{N}\left(0, \frac{1}{M}\right)$$

- Bernoulli/Rademacher [Achlioptas]

$$\phi_{i,j} := \begin{cases} +\frac{1}{\sqrt{M}} & \text{with probability } \frac{1}{2}, \\ -\frac{1}{\sqrt{M}} & \text{with probability } \frac{1}{2} \end{cases}$$

- "Database-friendly" [Achlioptas]

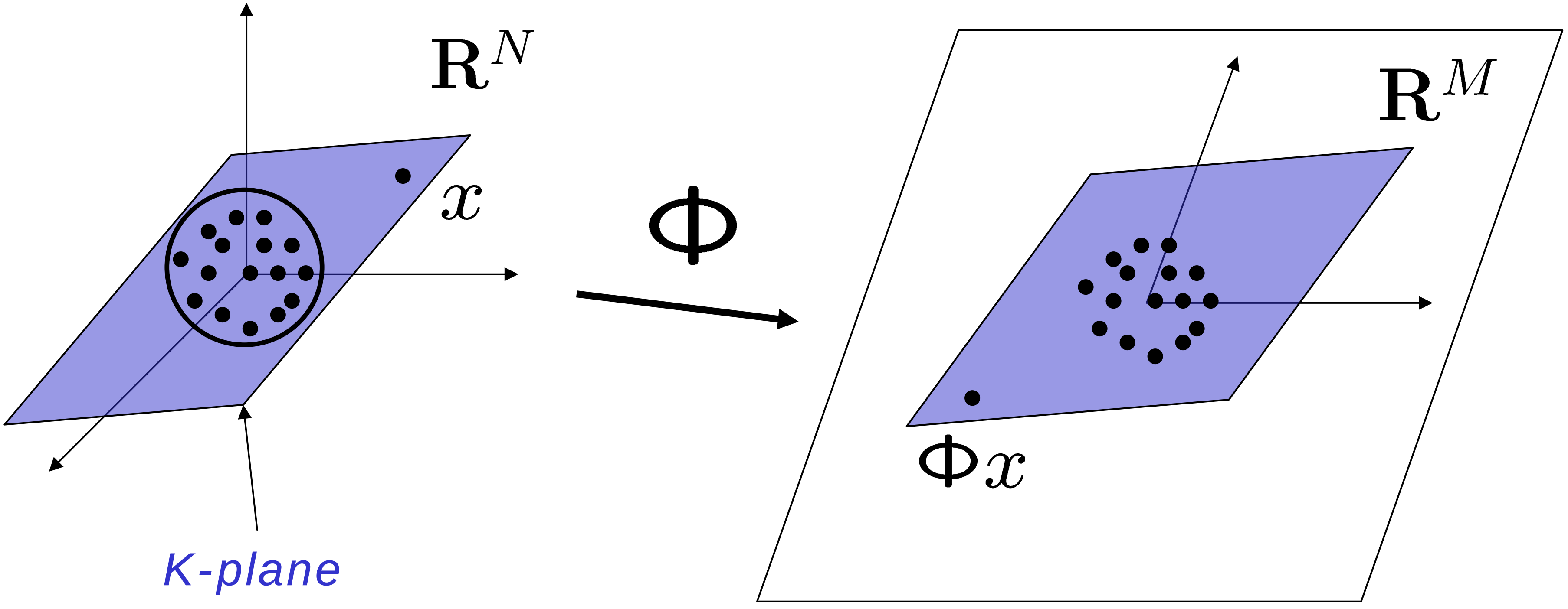
$$\phi_{i,j} := \begin{cases} +\sqrt{\frac{3}{M}} & \text{with probability } \frac{1}{6}, \\ 0 & \text{with probability } \frac{2}{3}, \\ -\sqrt{\frac{3}{M}} & \text{with probability } \frac{1}{6} \end{cases}$$

- Random Orthoprojection to R^M [Gupta, Dasgupta]

Connecting JL to RIP

Consider effect of random JL Φ on each K-plane

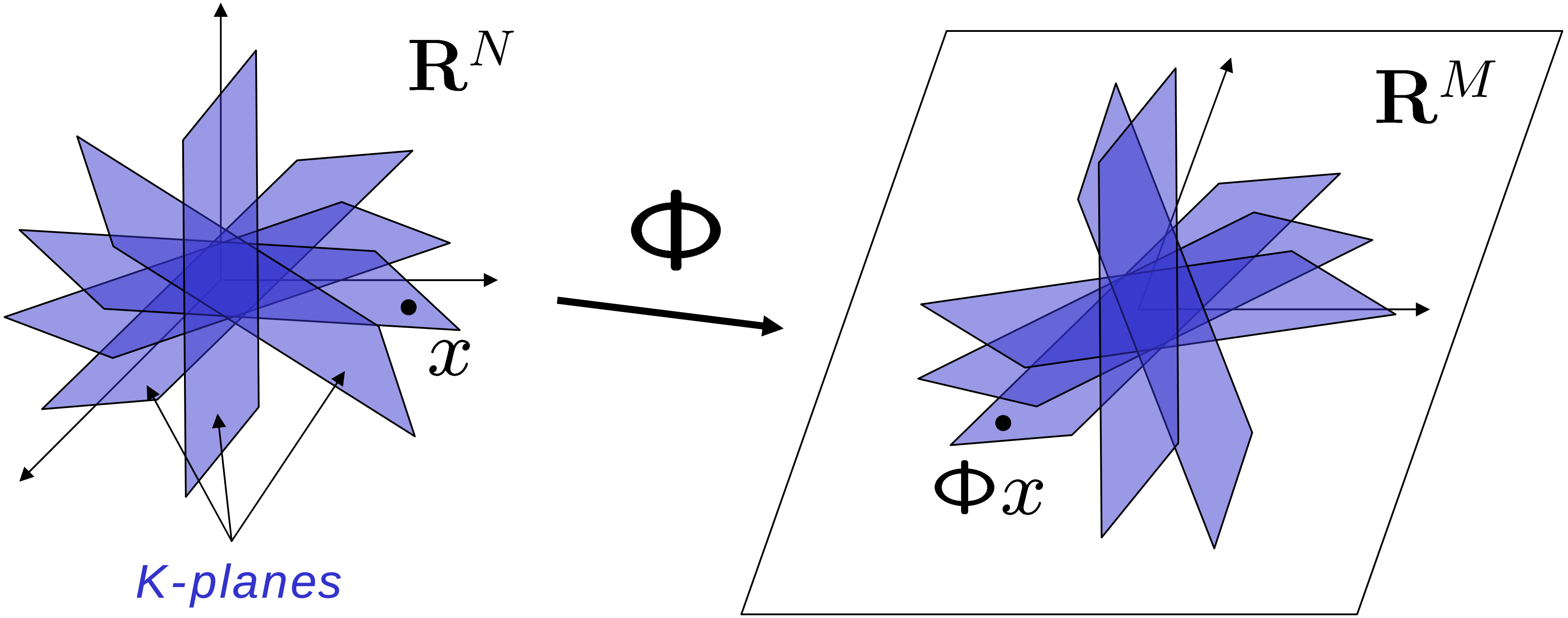
- construct covering of points Q on unit sphere
- JL: isometry for each point with high probability
- union bound ➔ isometry for all $q \in Q$
- extend to isometry for all x in K-plane



Connecting JL to RIP

Consider effect of random JL Φ on each K-plane

- construct covering of points Q on unit sphere
- JL: isometry for each point with high probability
- union bound ➔ isometry for all $q \in Q$
- extend to isometry for all x in K-plane
- union bound ➔ isometry for all K-planes



Connecting JL to RIP

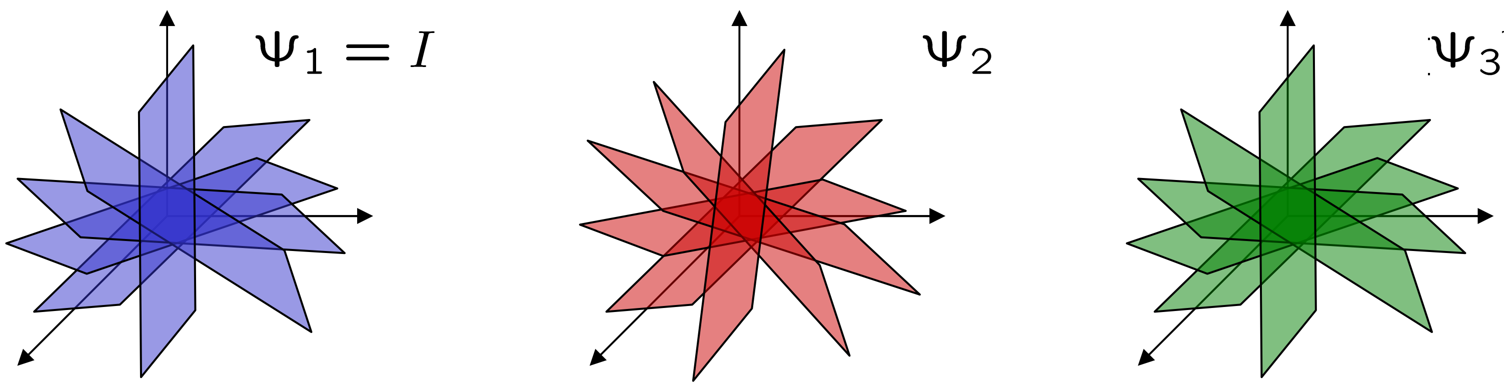
[with R. DeVore, M. Davenport, R. Baraniuk]

- **Theorem:** Supposing Φ is drawn from a JL-favorable distribution,* then with probability at least $1-e^{-C^*M}$, Φ meets the RIP with

$$K \leq C \cdot \frac{M}{\log(N/M)+1}.$$

* Gaussian/Bernoulli/database-friendly/orthoprojector

- Bonus: *universality* (repeat argument for any Ψ)



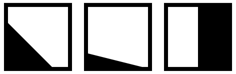
- See also Mendelson et al. concerning subgaussian ensembles

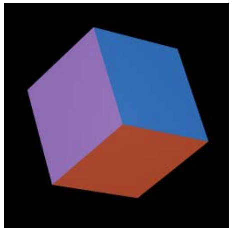
New Directions

Beyond Sparsity

- *Not all signal models fit into sparse representations*

- Other concise notions
  - constraints
  - degrees of freedom
  - parametrizations
  - articulations
  - signal families





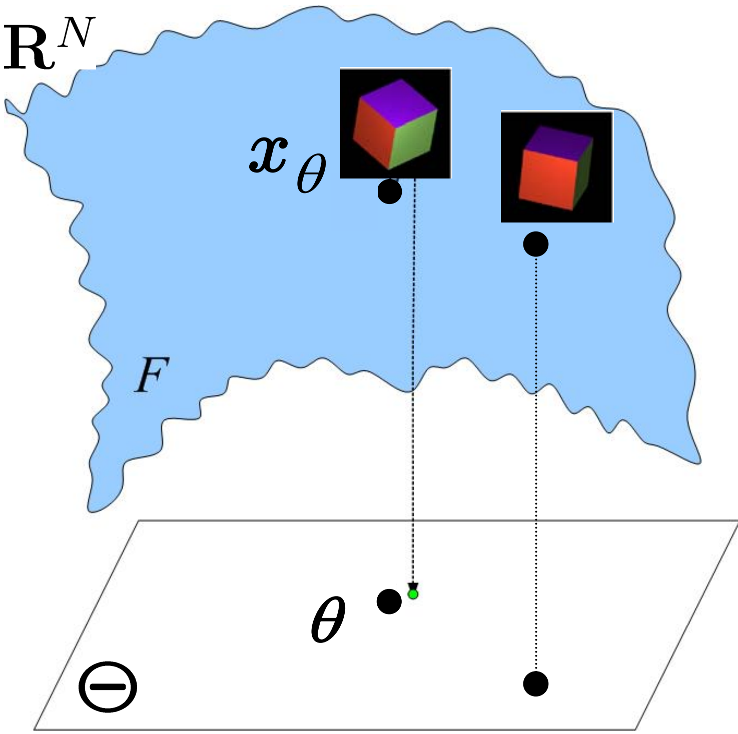
"information level" $\ll$ sparsity level $\ll N$

Challenge: How to exploit these concise models?

Manifold Models

- K-dimensional *parameter* $\theta \in \Theta$ captures degrees of freedom in signal $x_\theta \in R^N$

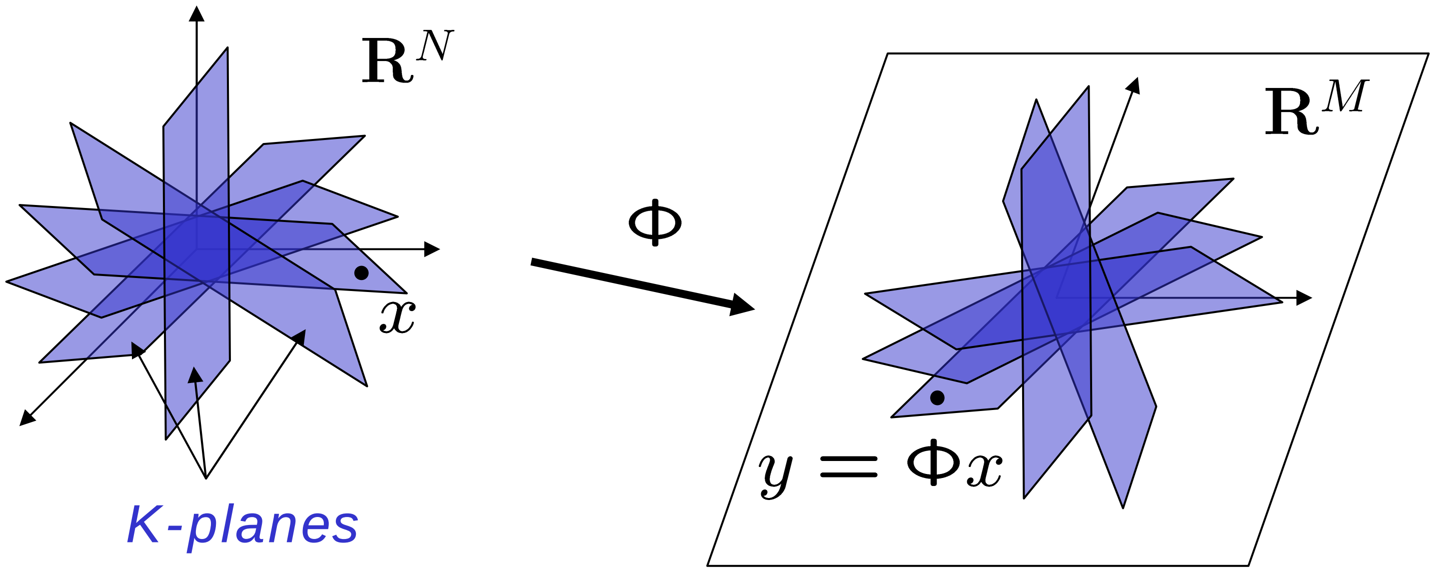
- Signal class $F = \{x_\theta : \theta \in \Theta\}$ forms a K-dimensional *manifold*
  - also nonparametric collections: faces, handwritten digits, shape spaces, etc.



- Dimensionality reduction and manifold learning
  - embeddings [ISOMAP; LLE; HLLE; …]
  - harmonic analysis [Eigenmaps; …]

Random Projections

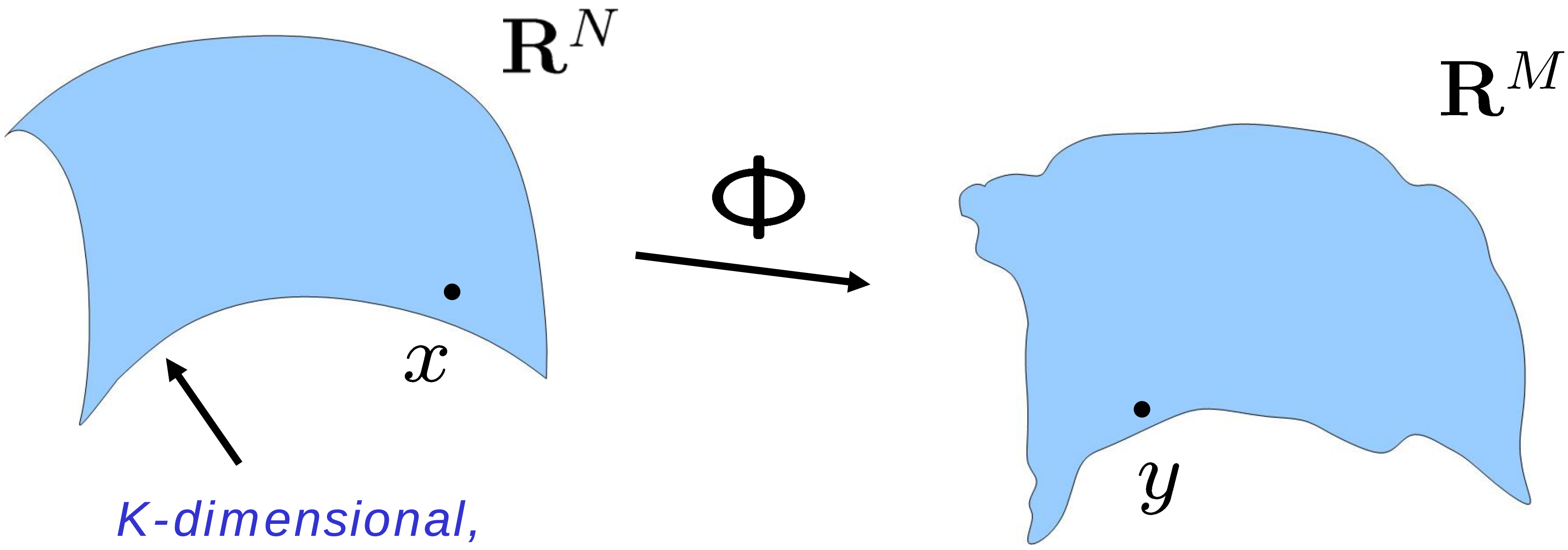
- *Random projections preserve information*
  - Compressive Sensing (sparse signal embeddings)
  - Johnson-Lindenstrauss lemma (point cloud embeddings)



- What about *manifolds*?

Manifold Embeddings

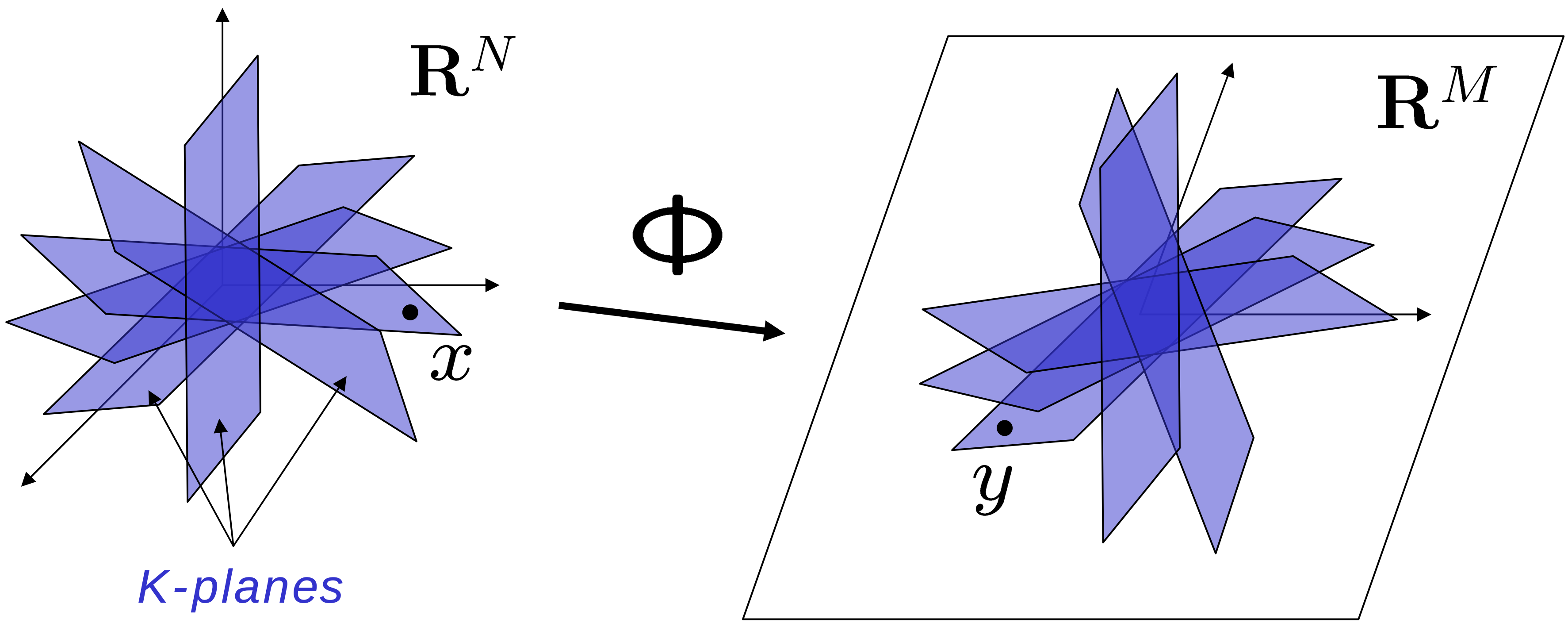
[Whitney (1936); Sauer et al. (1991)]



- M ≥ 2K+1 linear measurements
  - necessary for injectivity (in general)
  - sufficient for injectivity (w.p. 1) when Φ Gaussian

- But *not* enough for efficient, robust recovery

Recall: K-planes Embedding



- M ≥ 2K linear measurements
  - necessary for injectivity
  - sufficient for injectivity (w.p. 1) when Φ Gaussian

- But *not* enough for efficient, robust recovery

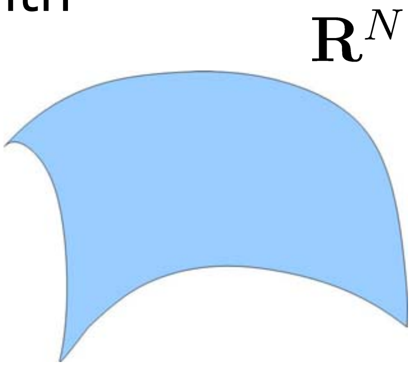
Stable Manifold Embedding

[with R. Baraniuk]

Theorem:

Let $F \subset R^N$ be a compact *K-dimensional manifold* with

- condition number $1/\tau$ (curvature, self-avoiding)
- volume V



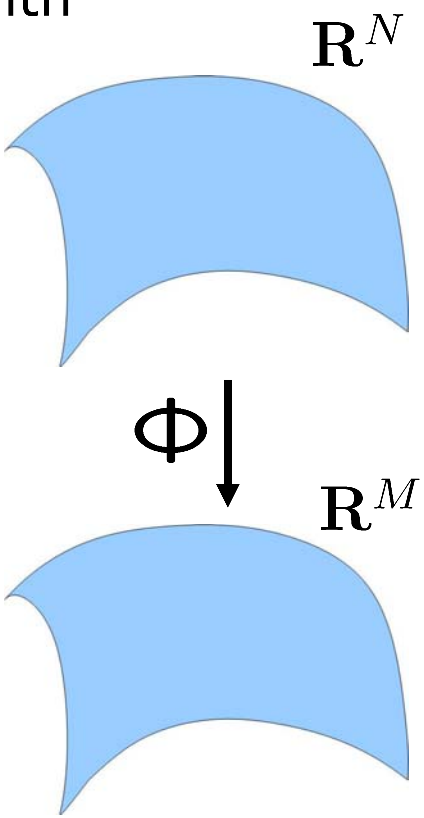
Stable Manifold Embedding

[with R. Baraniuk]

Theorem:

Let $F \subset R^N$ be a compact *K-dimensional manifold* with

- condition number $1/\tau$ (curvature, self-avoiding)
- volume V



Let Φ be a random MxN orthoprojector with

$$M = O\left(\frac{K \log(NV\tau^{-1}\epsilon^{-1}) \log(1/\rho)}{\epsilon^2}\right).$$

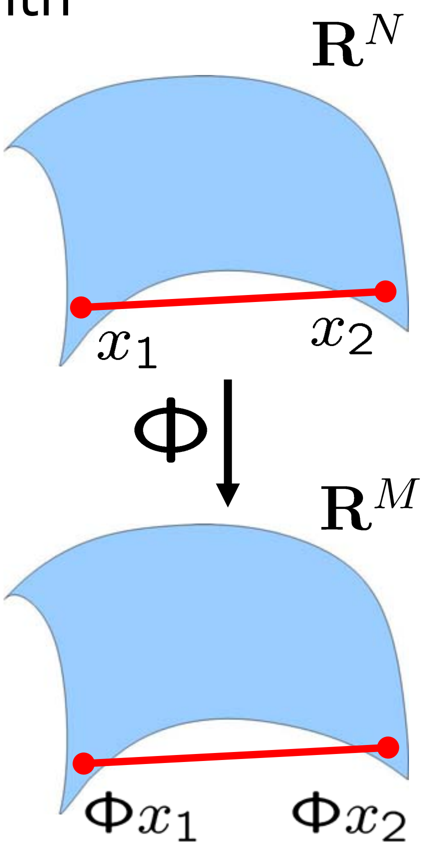
Stable Manifold Embedding

[with R. Baraniuk]

Theorem:

Let $F \subset R^N$ be a compact *K-dimensional manifold* with

- condition number $1/\tau$ (curvature, self-avoiding)
- volume V



Let Φ be a random MxN orthoprojector with

$$M = O\left(\frac{K \log(NV\tau^{-1}\epsilon^{-1}) \log(1/\rho)}{\epsilon^2}\right).$$

Then with probability at least $1-\rho$, the following statement holds: For every pair $x_1, x_2 \in F$,

$$(1-\epsilon) \le \frac{\|\Phi x_1 - \Phi x_2\|_2}{\|x_1 - x_2\|_2} \le (1+\epsilon).$$

Stable Manifold Embedding

[with R. Baraniuk]

Theorem:

Let $F \subset R^N$ be a compact *K-dimensional manifold* with

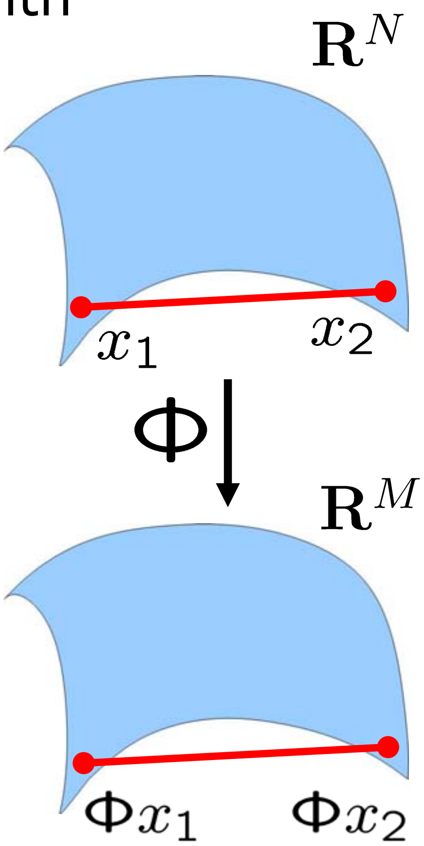
- condition number $1/\tau$ (curvature, self-avoiding)
- volume V

Let Φ be a random MxN orthoprojector with

$$M = O\left(\frac{K \log(NV\tau^{-1}\epsilon^{-1}) \log(1/\rho)}{\epsilon^2}\right).$$

Then with probability at least $1-\rho$, the following statement holds: For every pair $x_1, x_2 \in F$,

$$(1-\epsilon) \leq \frac{\|\Phi x_1 - \Phi x_2\|_2}{\|x_1 - x_2\|_2} \leq (1+\epsilon).$$



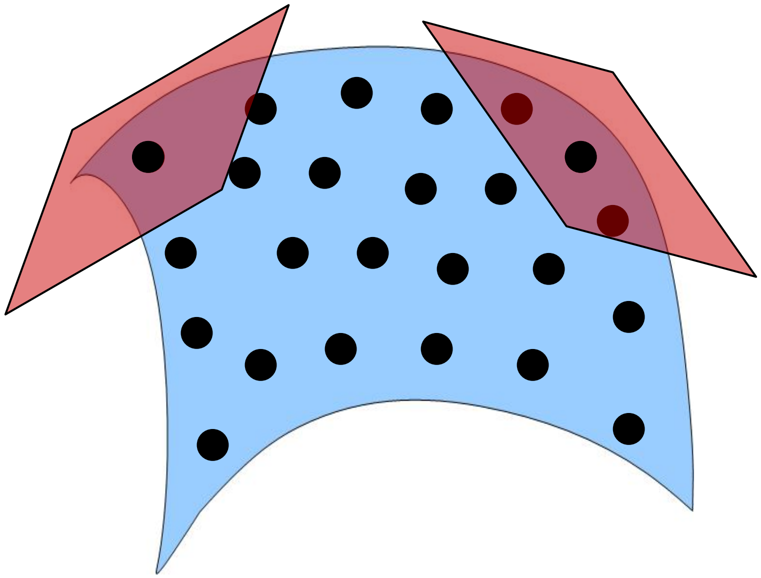
Stable Manifold Embedding

Sketch of proof:

- construct a sampling of points
  - on manifold at fine resolution
  - from local tangent spaces
- apply JL to these points

$$M = O\left(\frac{K \log(NV\tau^{-1}\epsilon^{-1}) \log(1/\rho)}{\epsilon^2}\right)$$

- extend to entire manifold

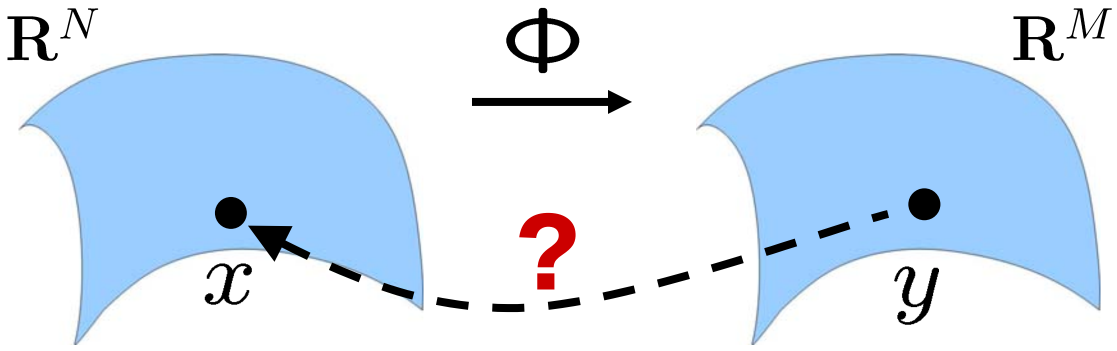


Implications:

Nonadaptive (even random) linear projections can efficiently capture & preserve structure of manifold

See also: Indyk and Naor, Agarwal et al., Dasgupta and Freund

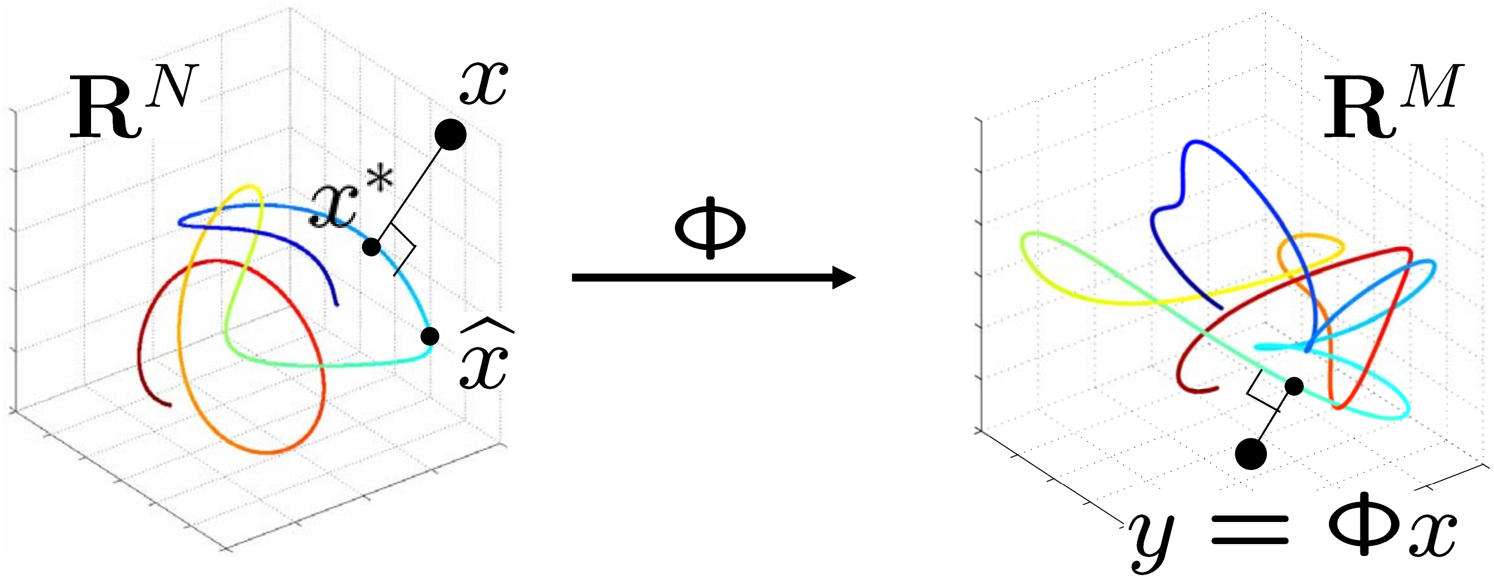
Application 1: Compressive Sensing



- Same *nonadaptive* sensing protocols/devices
- Different reconstruction models
- Measurement rate depends on *manifold dimension*
- Stable embedding guarantees robust recovery

Signal Recovery

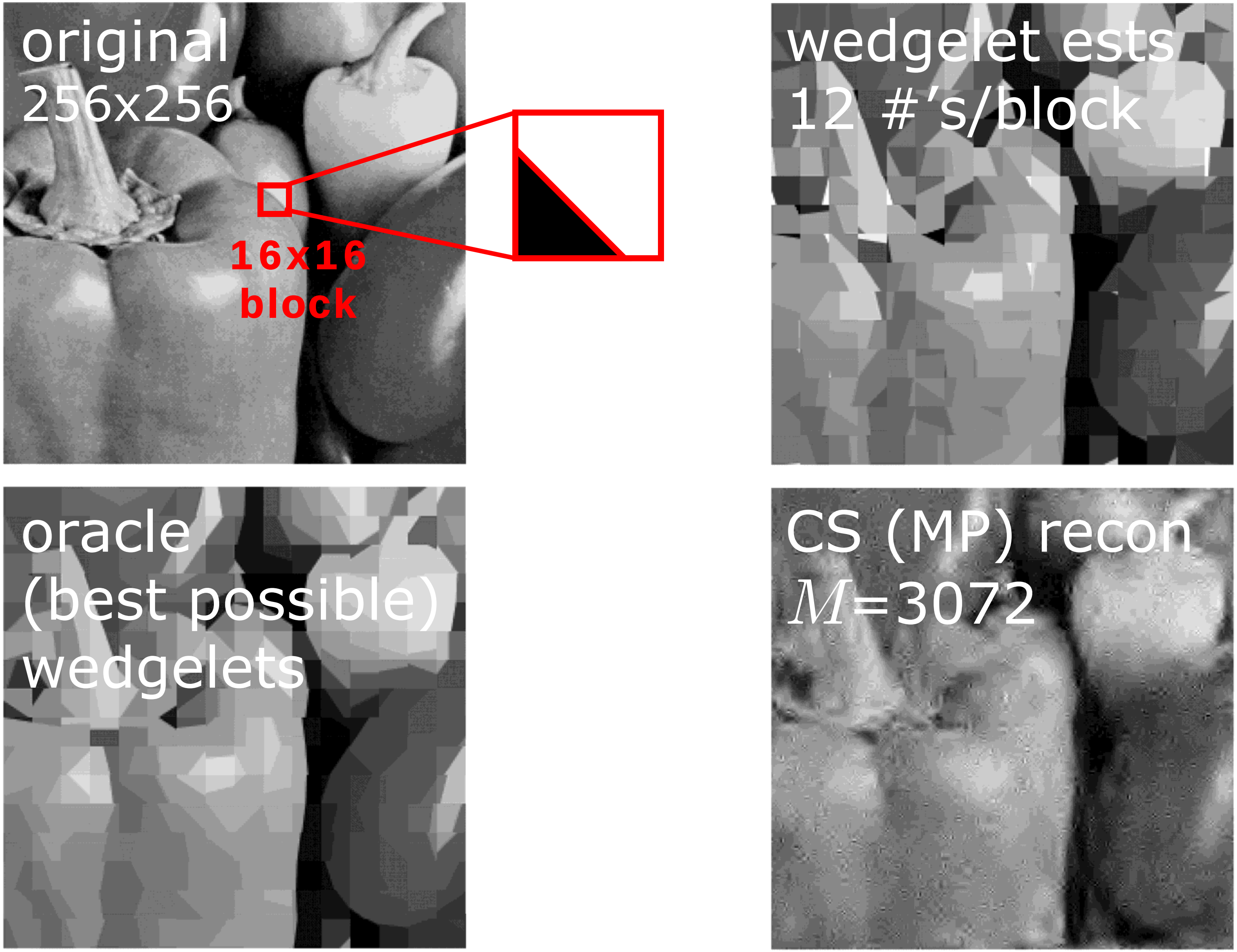
Search in projected space: $\widehat{x} = \arg\min_{x' \in \mathcal{F}} \|y - \Phi x'\|_2$



For all x: $\|x - \widehat{x}\|_2 < 2\sqrt{\frac{N}{M}}\|x - x^*\|_2$

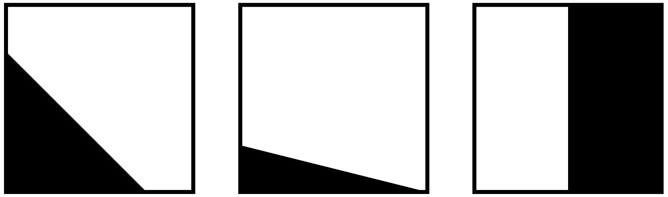
For most x: $\|x - \widehat{x}\|_2 < (1 + \epsilon)\|x - x^*\|_2 + \mathrm{C}\frac{\epsilon^2 \tau}{N}$

Example: Edges

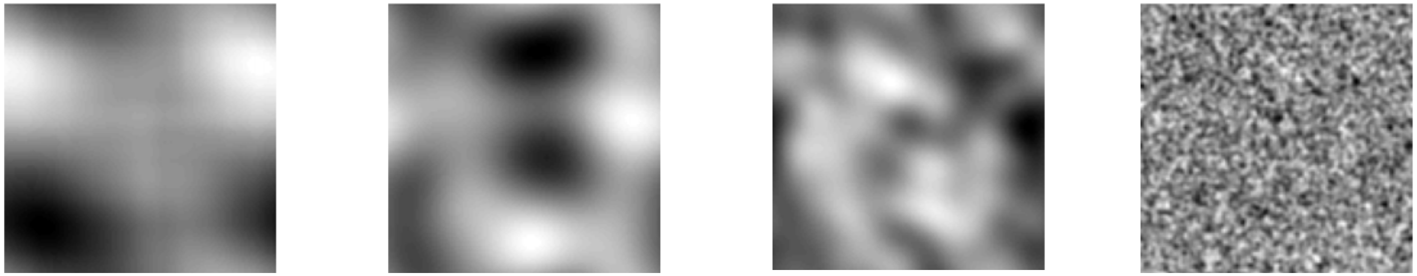


One Challenge: Non-Differentiability

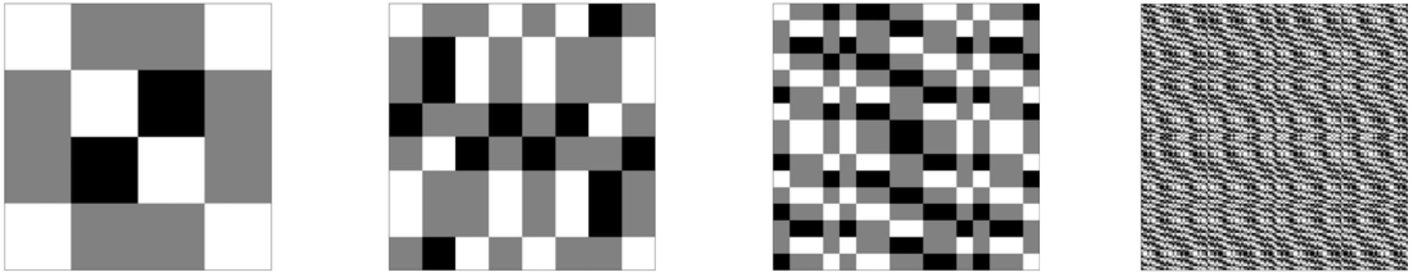
- Many image manifolds are non-differentiable
  - no embedding guarantee
  - difficult to navigate



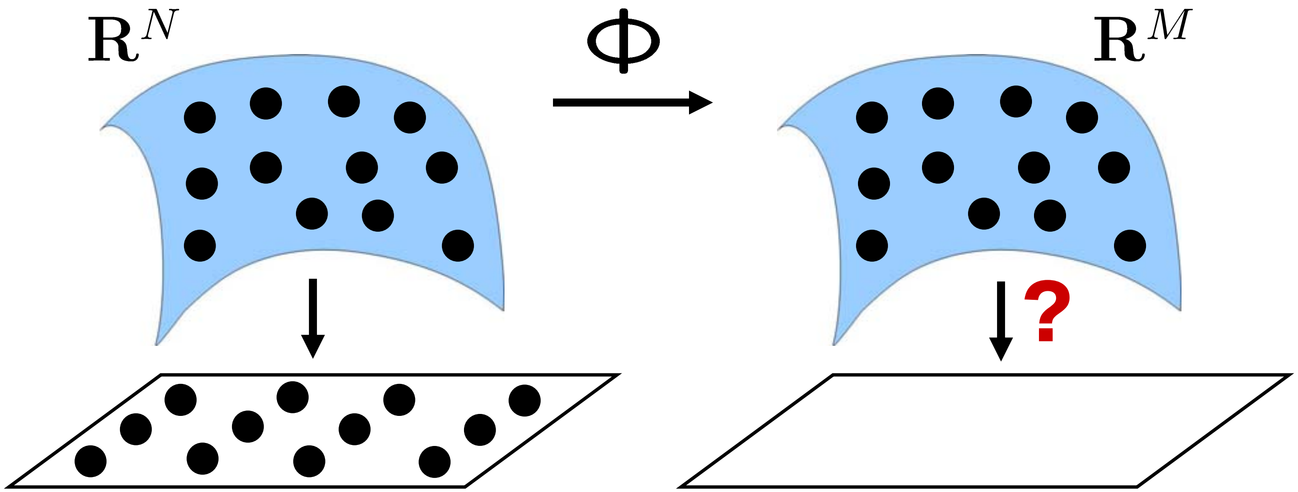
- Solution: *multiscale random projections*



- Noiselets [Coifman et al.]



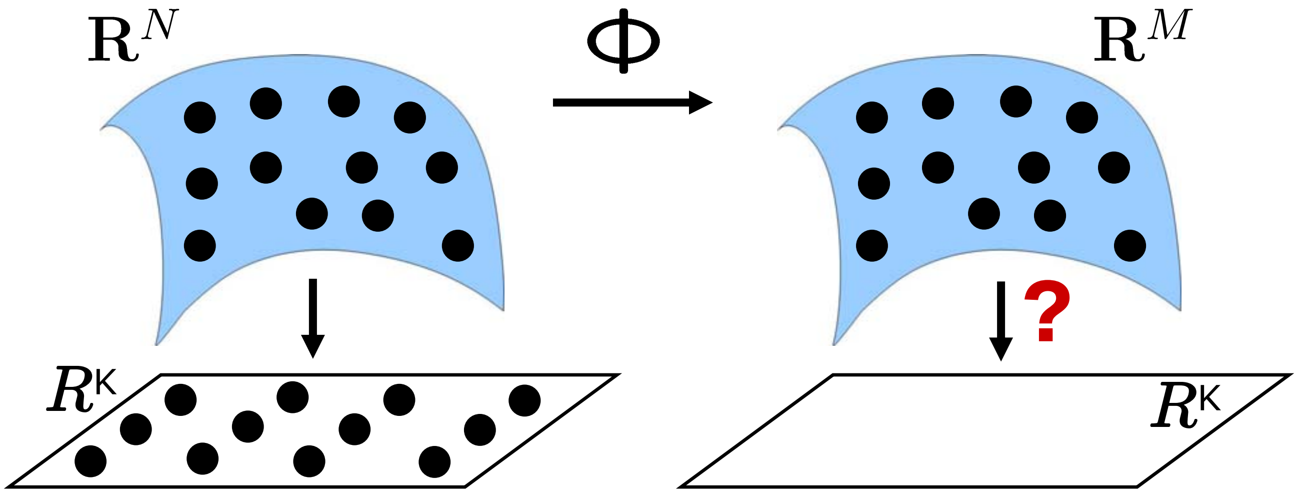
Application 2: Manifold Learning



- Many key properties preserved in R^M
  - ambient and geodesic distances
  - dimension and volume of the manifold
  - path lengths and curvature
  - topology, local neighborhoods, angles, etc…

Random Projections for Manifold Learning

[with C. Hegde, R. Baraniuk]

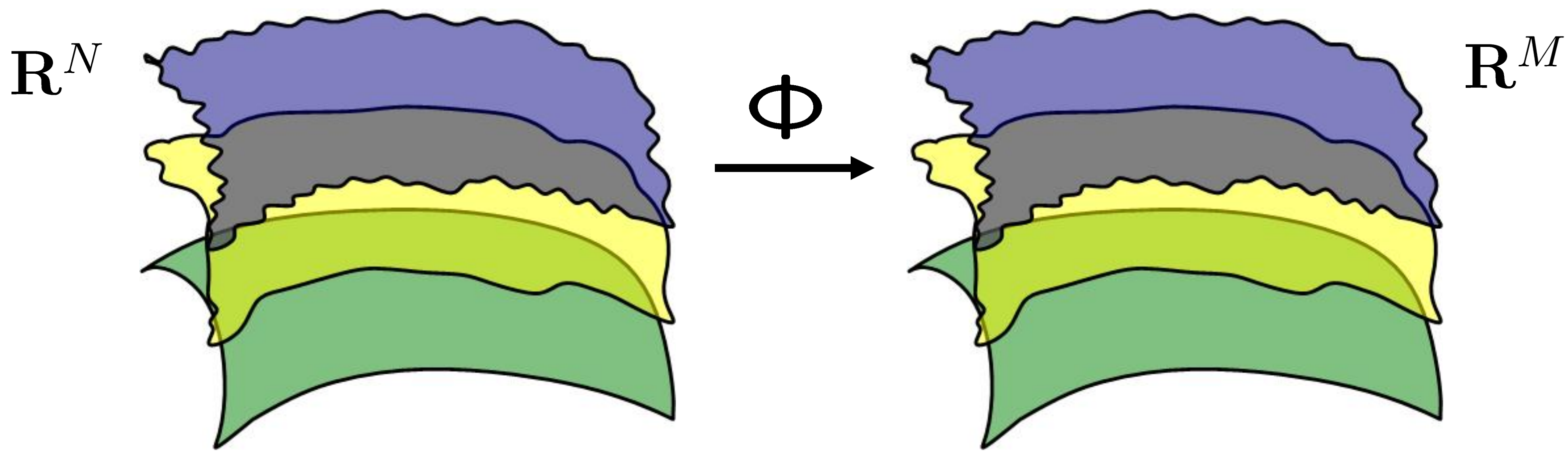


<u>Theorem:</u> With high probability, the ISOMAP algorithm embeds data in R^K with residual variance

$$R < R_{\text{orig}} + C\epsilon.$$

Application 3: Classification

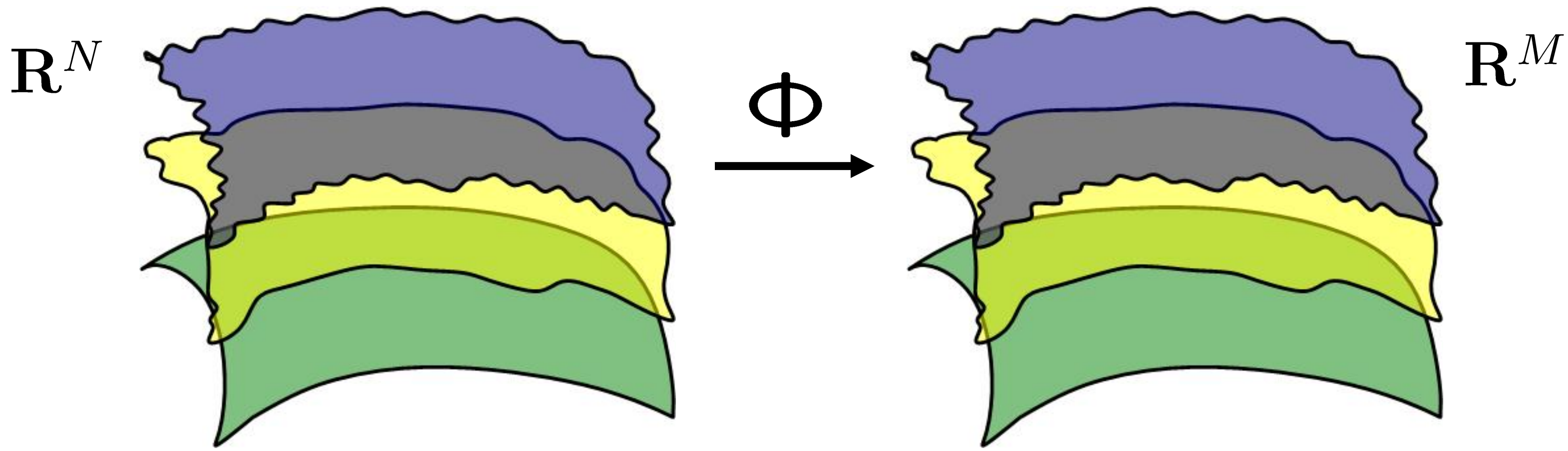
[with M. Davenport, M. Duarte, R. Baraniuk]

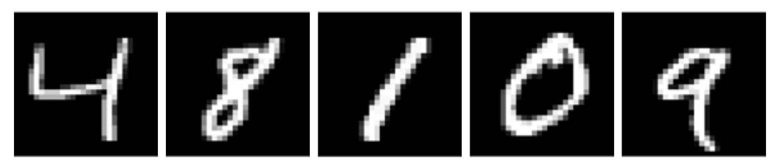


$$M = O\left(\frac{K \log(JNV\tau^{-1}\epsilon^{-1}) \log(1/\rho)}{\epsilon^2}\right)$$

Application 3: Classification

[with M. Davenport, M. Duarte, R. Baraniuk]





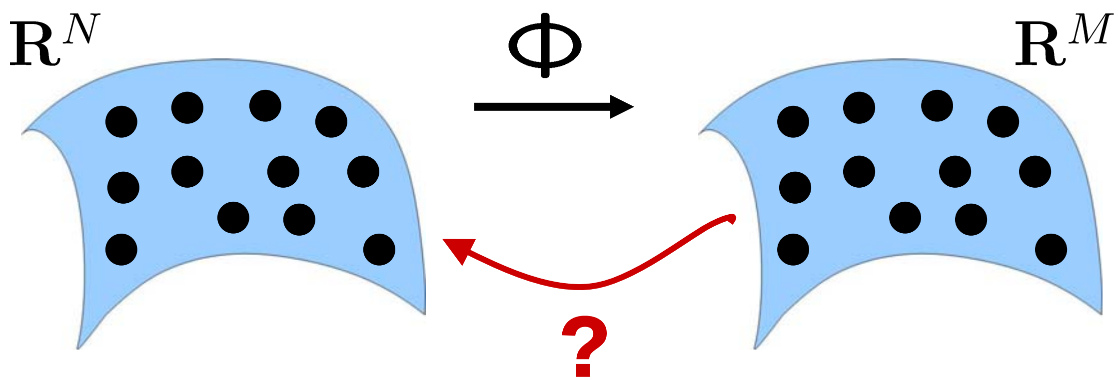
N = 16394
M = 8

100% Classification
±30° Estimation

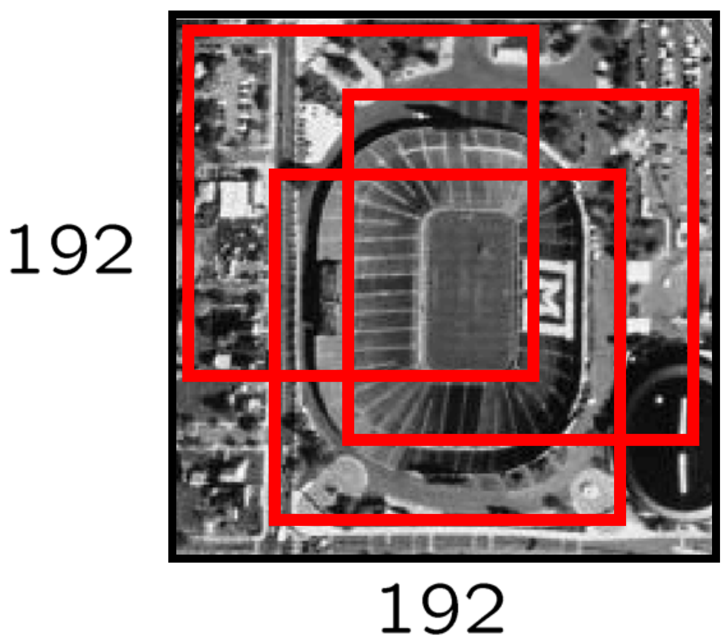
N = 784
M = 60

Laplacian Eigenmaps
90% accuracy

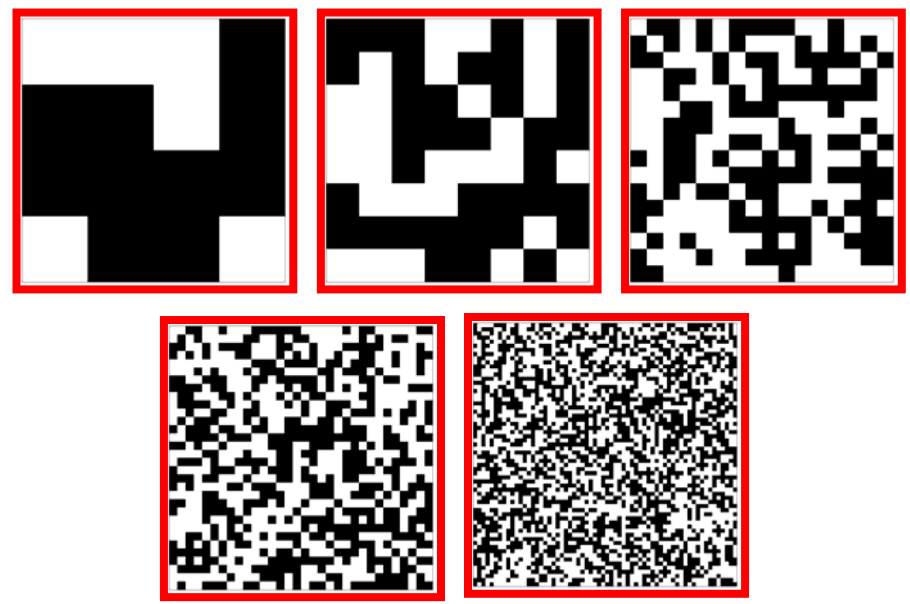
Application 4: “Manifold Lifting”



200 images
$N = 64^2 = 4096$



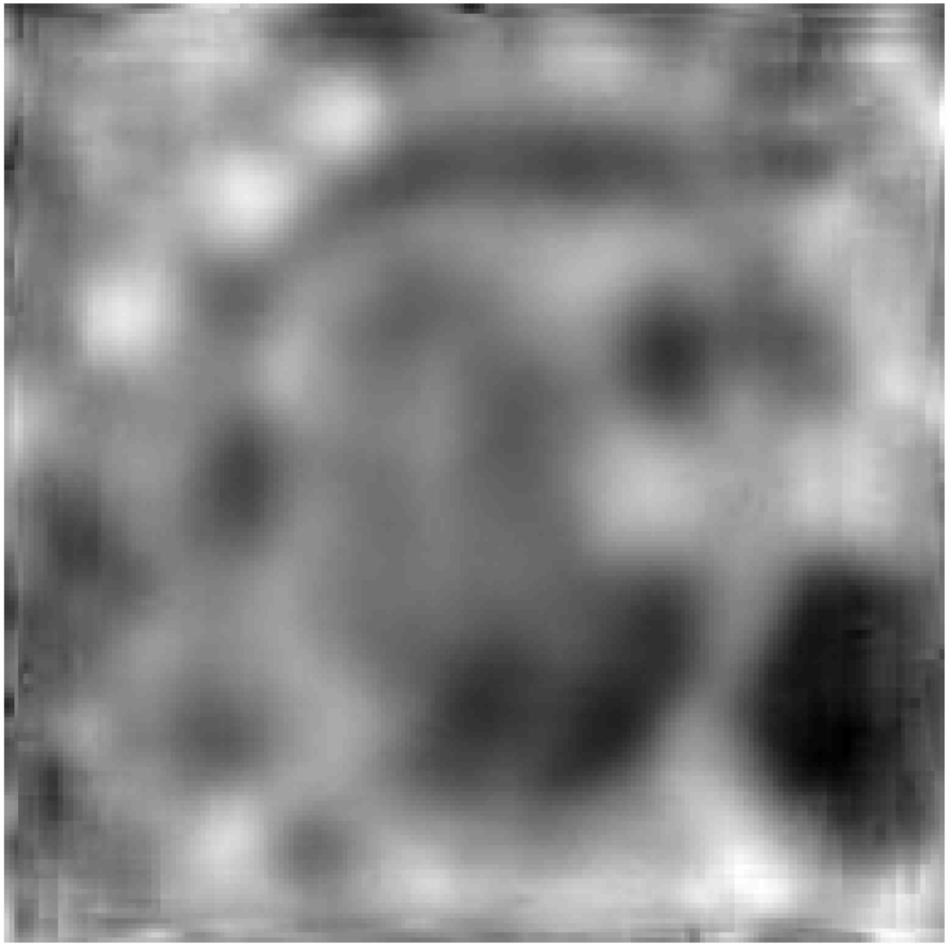
$M = 96$



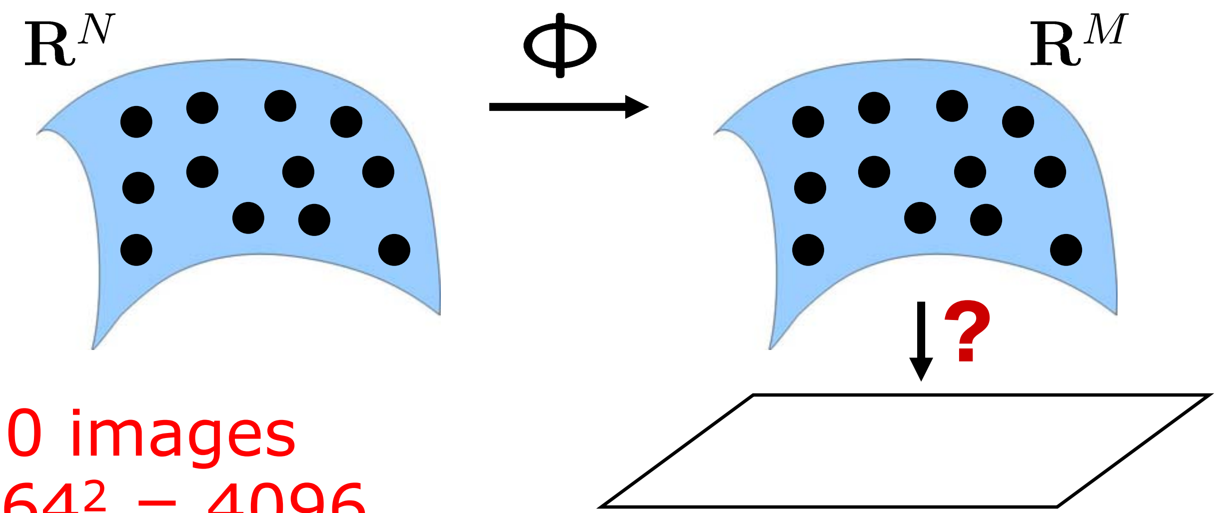
Noiselets [Coifman et al.]

Image-by-Image Reconstruction

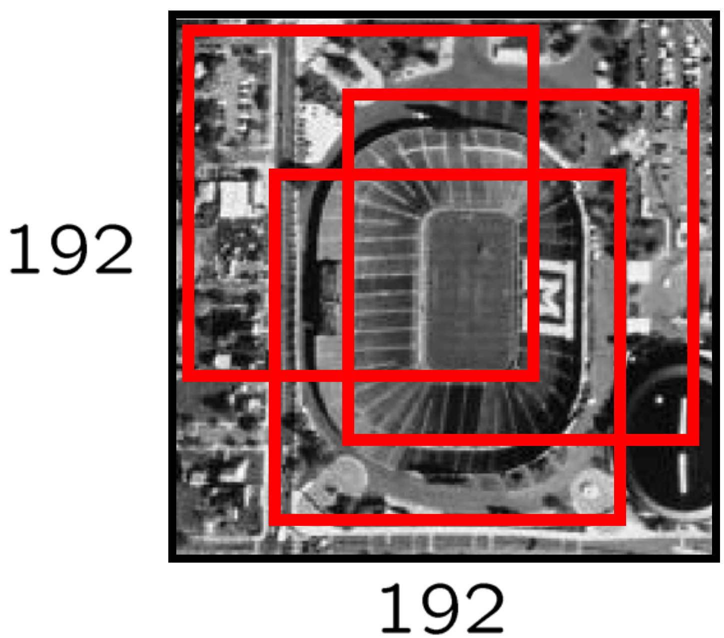
- With oracle information for camera positions



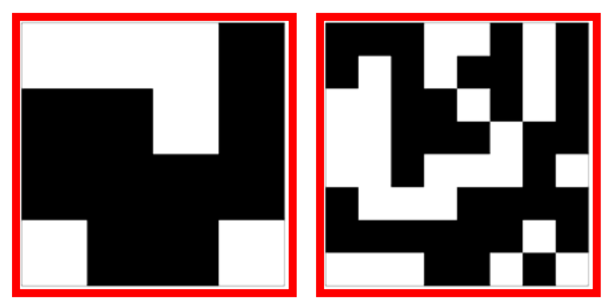
Manifold Learning for Position Estimates



200 images
$N = 64^2 = 4096$

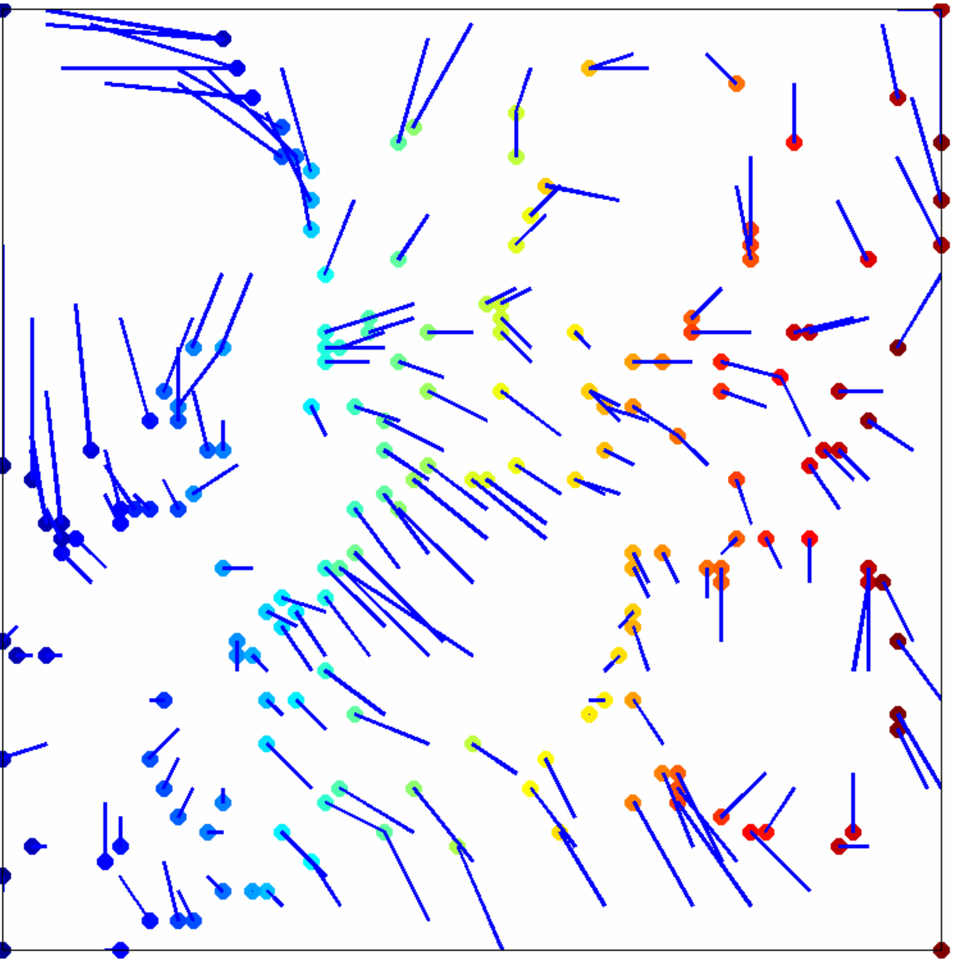


using only 2 coarse scales
(M = 36)

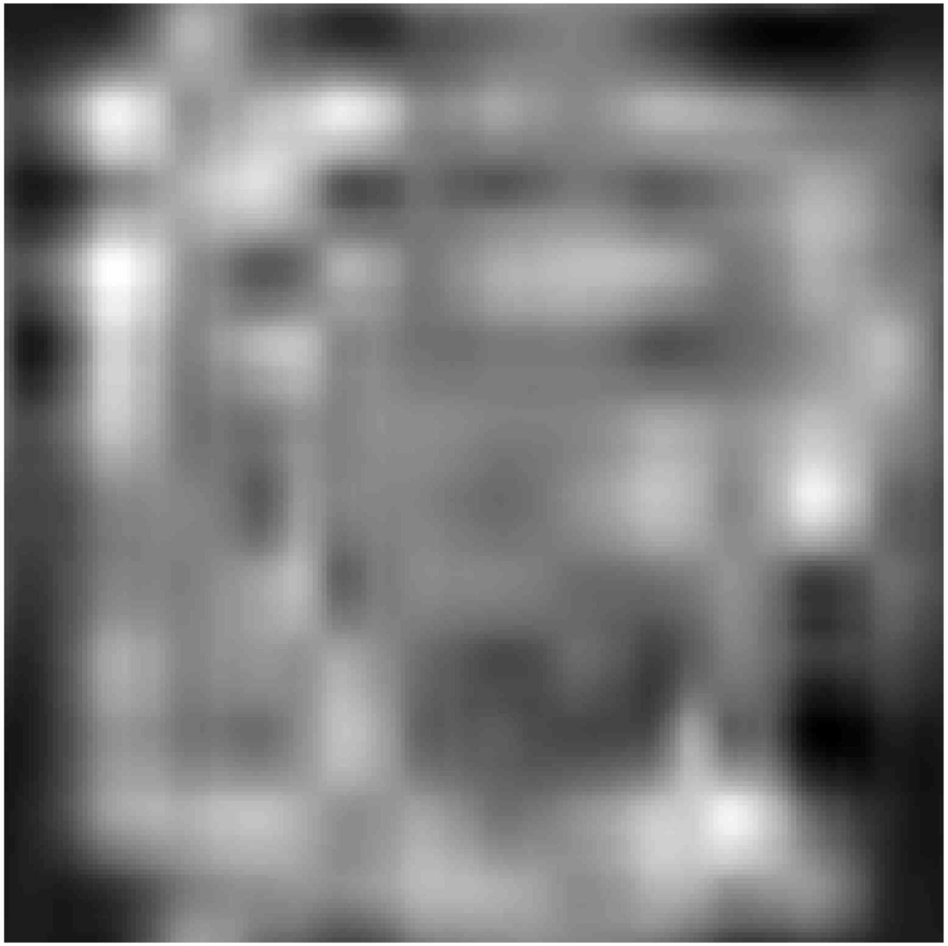


Initial Estimates

ISOMAP – 2 coarse scales

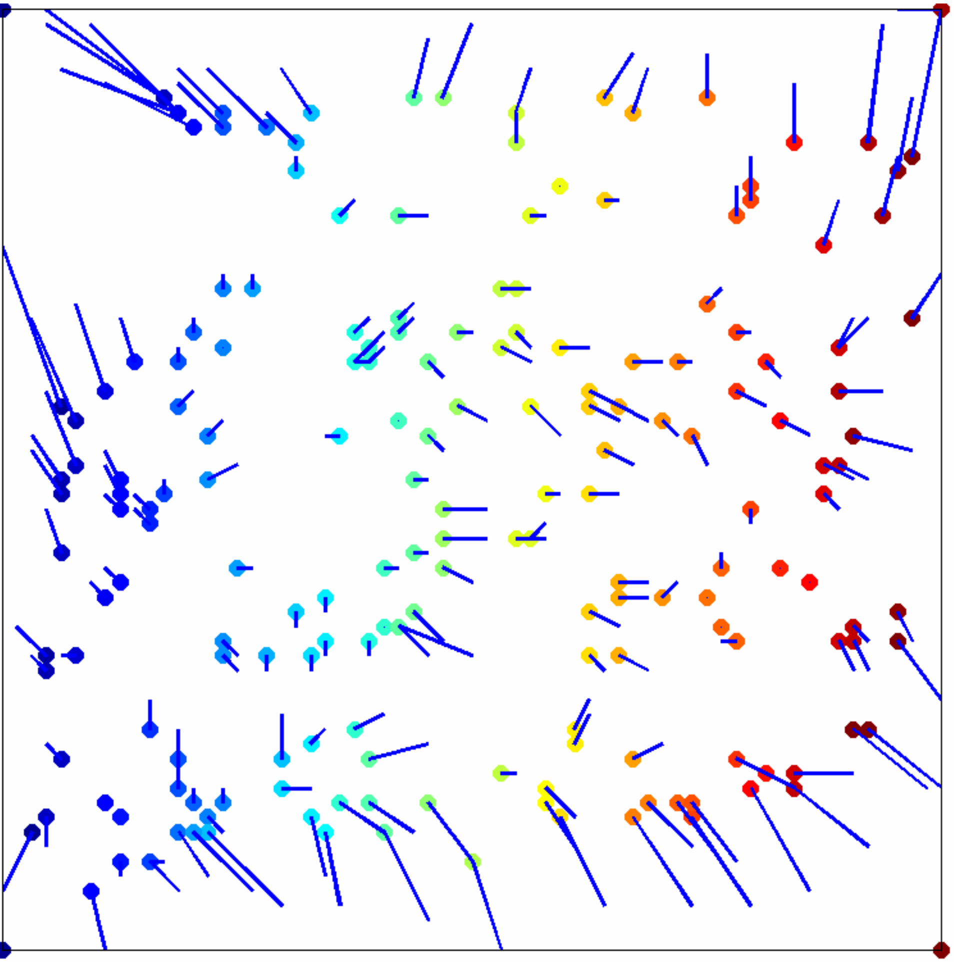


L_1 reconstruction – 2 scales

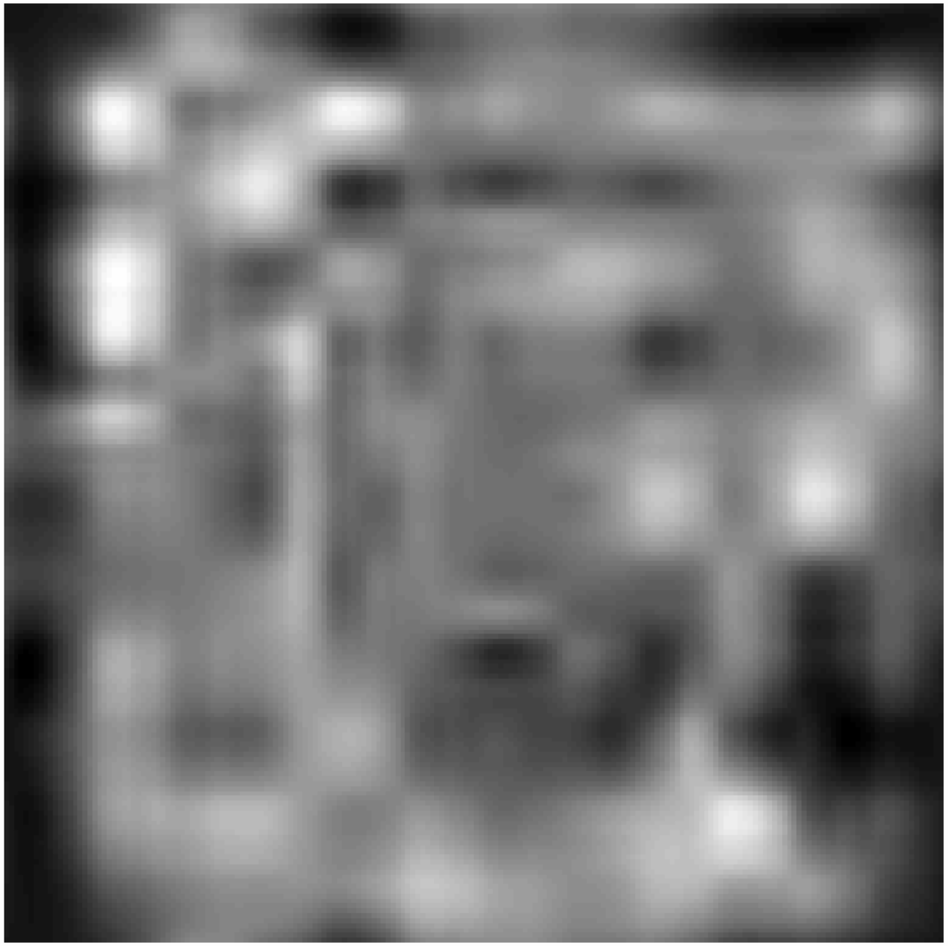


Refinements - 1

Re-registration

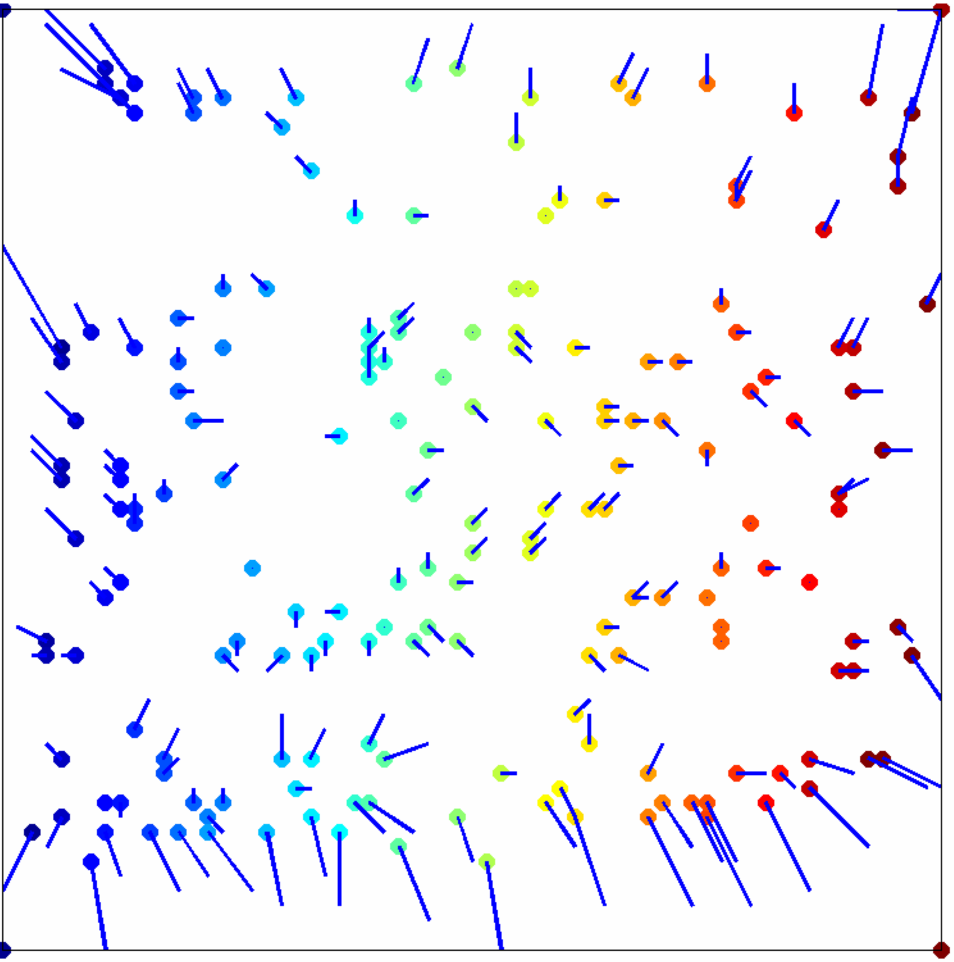


L_1 reconstruction – 2 scales



Refinements - 2

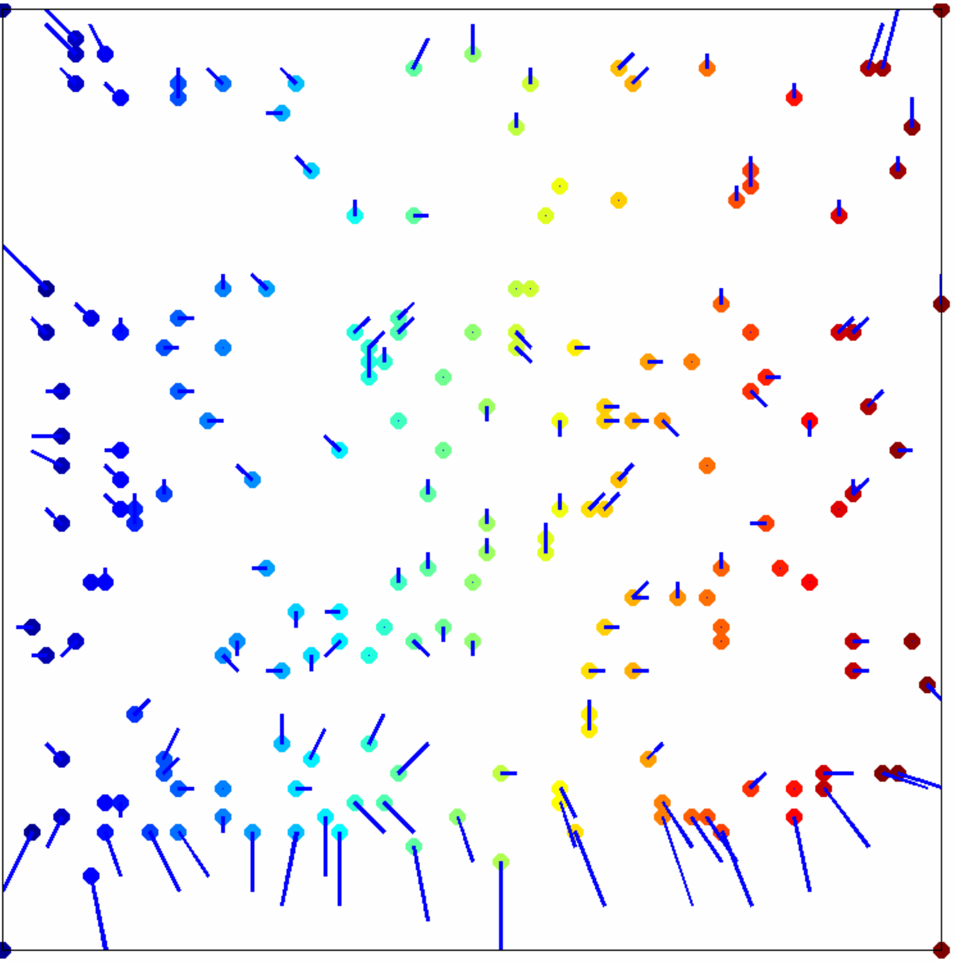
Re-registration



L_1 reconstruction – 2 scales

Refinements - 3

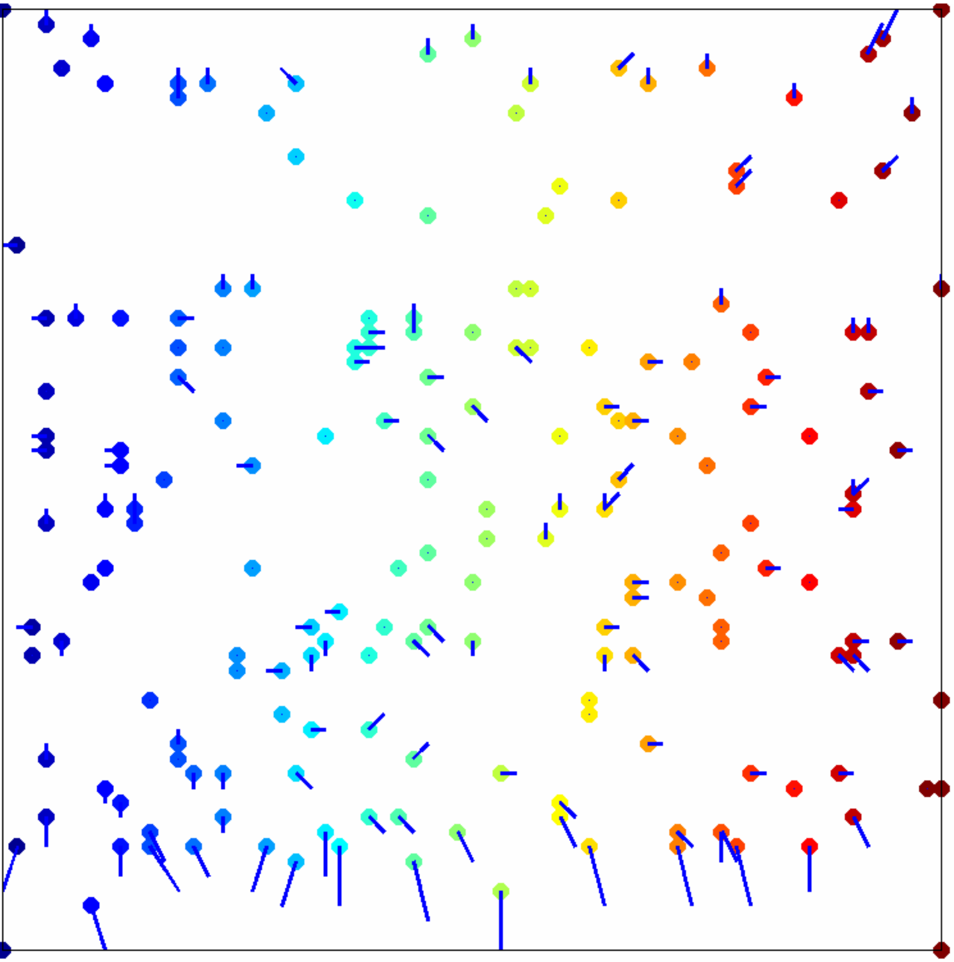
Re-registration



L_1 reconstruction – 3 scales

Refinements - 4

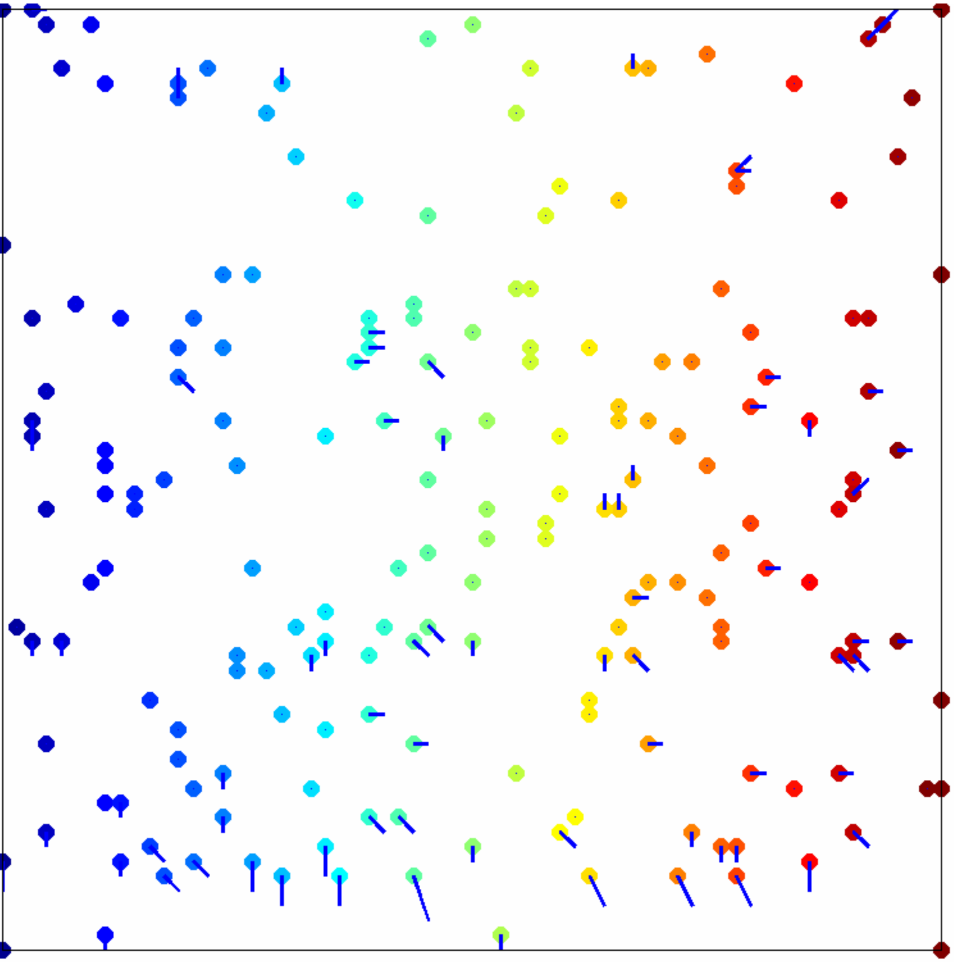
Re-registration



L_1 reconstruction – 3 scales

Refinements - 5

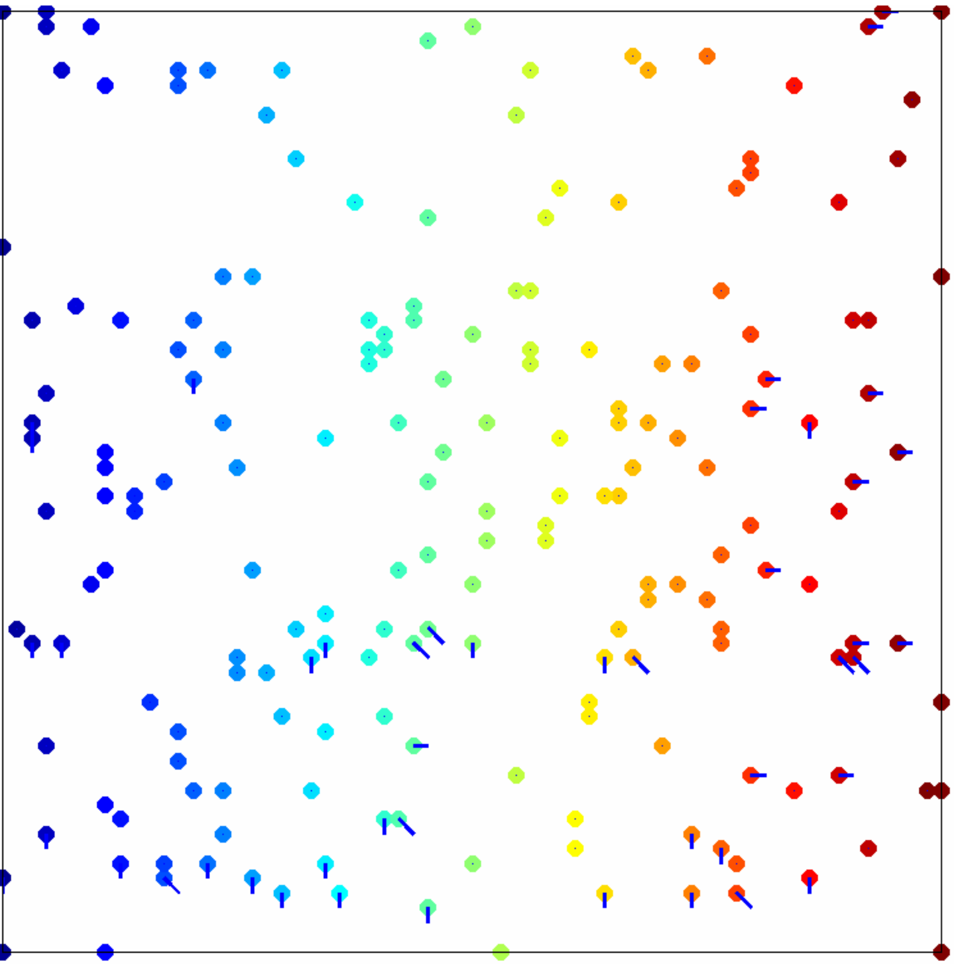
Re-registration



L_1 reconstruction – 3 scales

Refinements - 6

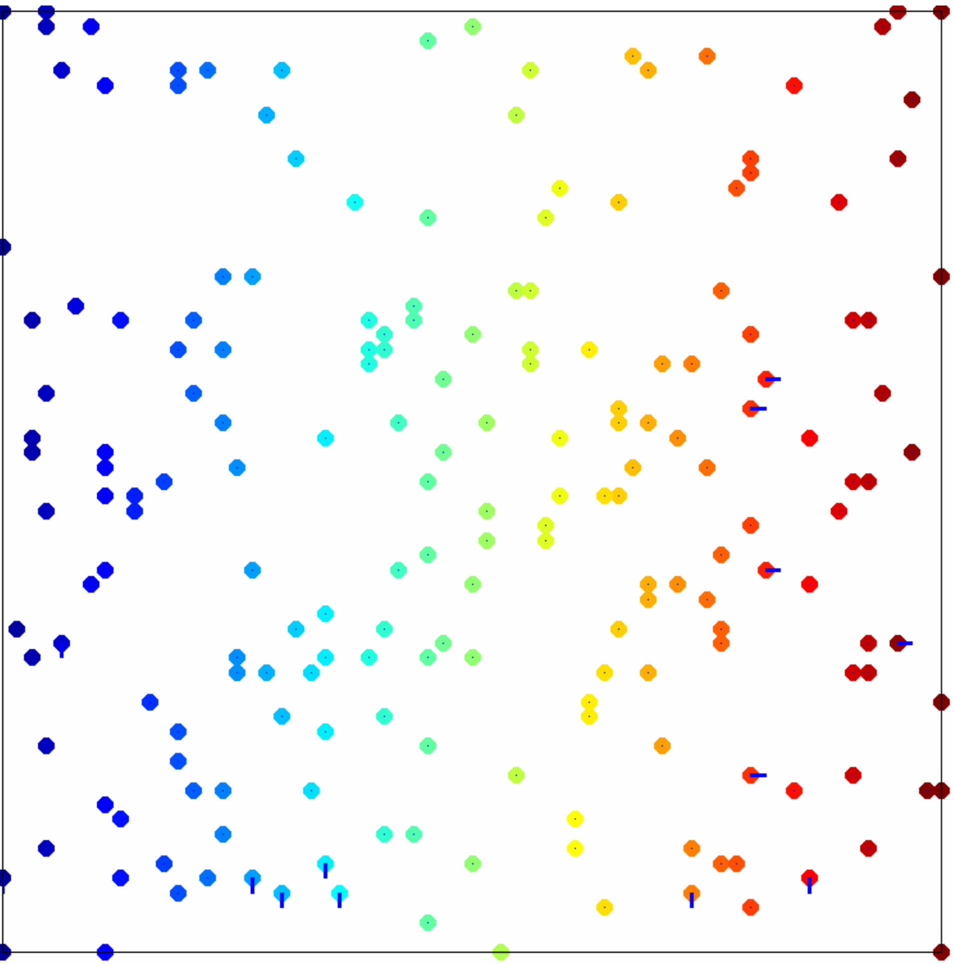
Re-registration



L_1 reconstruction – 4 scales

Refinements - 7

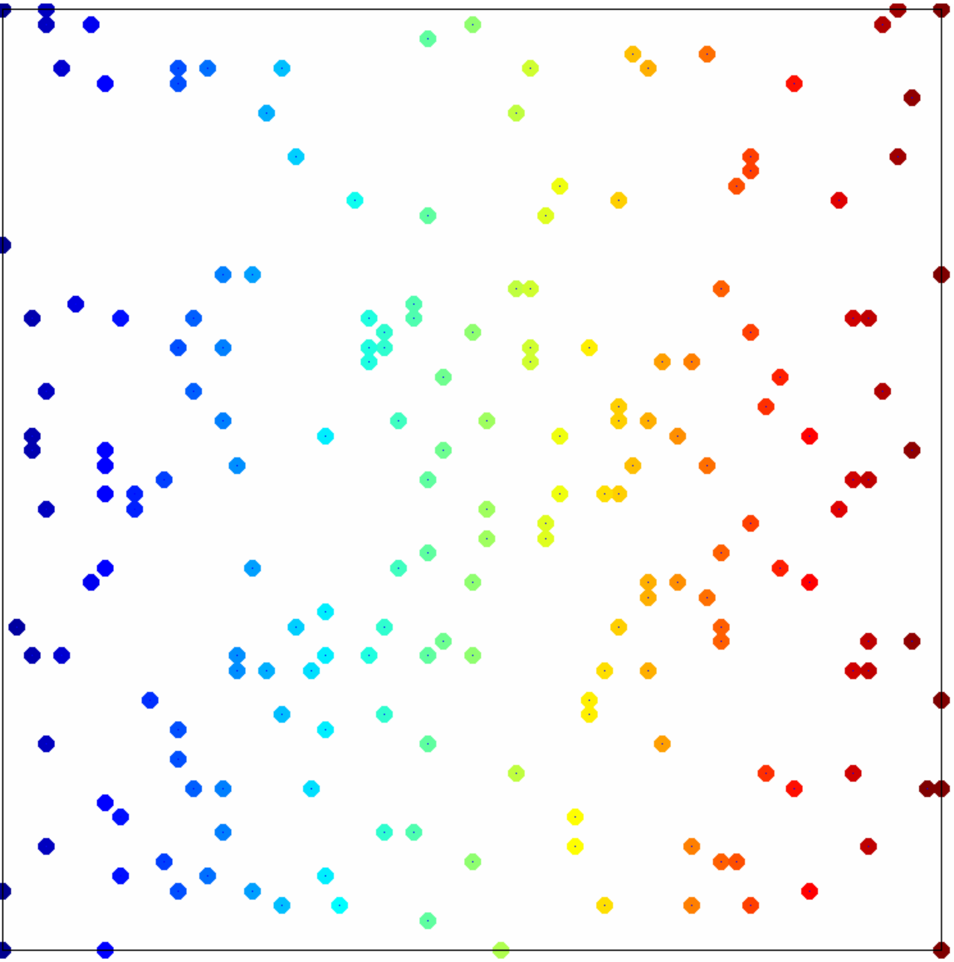
Re-registration



L_1 reconstruction – 4 scales

Refinements – 8

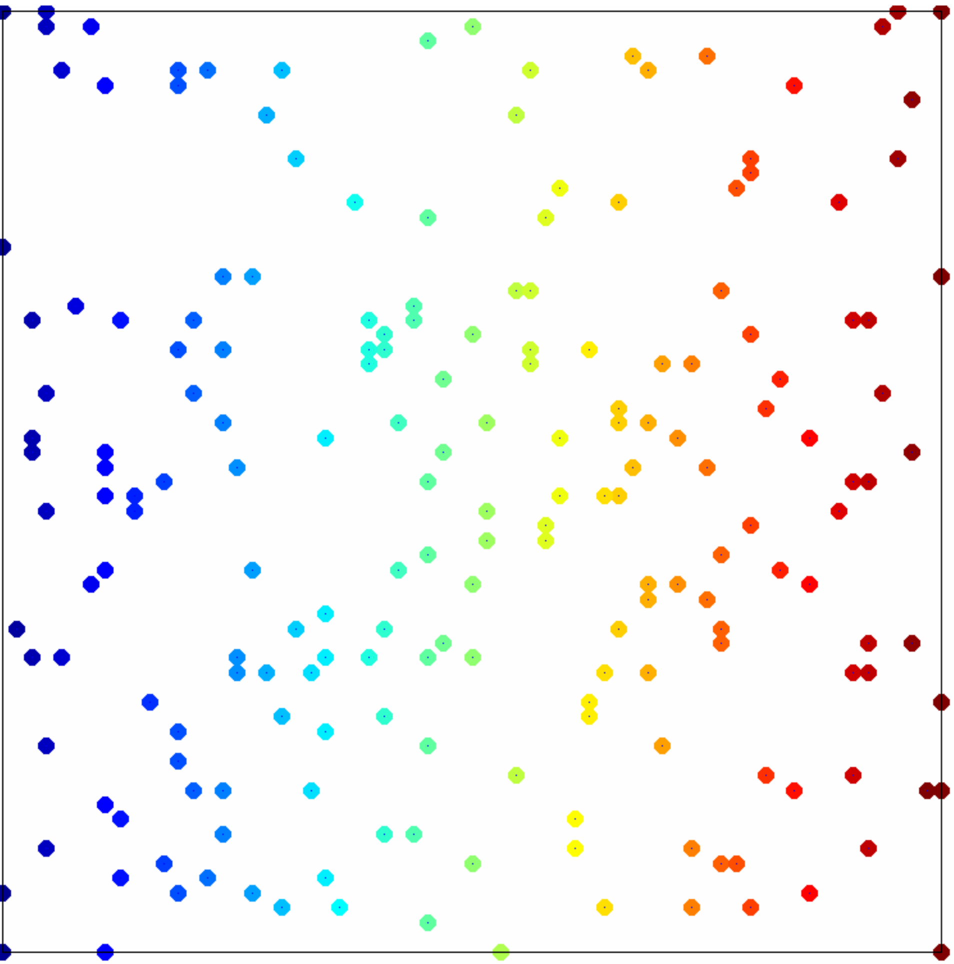
Re-registration



L_1 reconstruction – 4 scales

Refinements - 9

Re-registration



L_1 reconstruction – 5 scales

Final Reconstruction

Global fine-scale noiselet measurements

L_1 reconstruction – 5 scales

Summary – Geometry in CS

- ***Concise models → low-dimensional geometry***
  - bandlimited
  - sparse
  - manifolds

- ***Random projections***
  - stable embedding thanks to low-dimensional geometry
  - model-based recovery; use the best model available

- ***Sparsity-based Compressive Sensing***
  - powerful results for explicit, multi-purpose recovery algorithms

- ***Manifolds & other models***
  - specialized algorithms may be required; but apps beyond CS
  - work in progress

References – Geometry (1)

L_0 Recovery:

- P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc., Atlanta, GA, 1996, vol. 2, pp. 1689–1692.
- Dror Baron, Michael Wakin, Marco Duarte, Shriram Sarvotham, and Richard Baraniuk, Distributed compressed sensing. (Preprint, 2005)

Finite Rate of Innovation:

- Martin Vetterli, Pina Marziliano, and Thierry Blu, Sampling signals with finite rate of innovation. (IEEE Trans. on Signal Processing, 50(6), pp. 1417-1428, June 2002)
- Y. M. Lu and M. N. Do, Sampling signals from a union of subspaces, IEEE Signal Processing Magazine, to appear.

L_1 Recovery & Random Polytopes:

- David Donoho and Jared Tanner, Counting faces of randomly-projected polytopes when the projection radically lowers dimension. (Submitted to Journal of the AMS)

Optimality & n-widths:

- David Donoho, Compressed sensing. (IEEE Trans. on Information Theory, 52(4), pp. 1289 - 1306, April 2006)
- Emmanuel Candès and Terence Tao, Near optimal signal recovery from random projections: Universal encoding strategies? (IEEE Trans. on Information Theory, 52(12), pp. 5406 - 5425, December 2006)

References – Geometry (2)

RIP/UUP & Implications:

- Emmanuel Candès and Terence Tao, Decoding by linear programming. (IEEE Trans. on Information Theory, 51(12), December 2005)
- David Donoho, For most large underdetermined systems of linear equations, the minimal ell-1 norm solution is also the sparsest solution. (Communications on Pure and Applied Mathematics, 59(6), June 2006)
- Emmanuel Candès, Justin Romberg, and Terence Tao, Stable signal recovery from incomplete and inaccurate measurements. (Communications on Pure and Applied Mathematics, 59(8), August 2006)
- Emmanuel Candès and Terence Tao, The Dantzig Selector: Statistical estimation when p is much larger than n (To appear in Ann. Statistics)
- Rudelson, M., Vershynin, R., Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements, Preprint, 2006.
- Albert Cohen, Wolfgang Dahmen, and Ronald DeVore, Compressed sensing and best k-term approximation. (Preprint, 2006)
- Holger Rauhut, Karin Schass, and Pierre Vandergheynst, Compressed sensing and redundant dictionaries. (Preprint, 2006)
- Ronald A. DeVore, Deterministic constructions of compressed sensing matrices. (Preprint, 2007)
- Deanna Needell and Roman Vershynin, Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. (Preprint, 2007)

References – Geometry (3)

Johnson-Lindenstrausss Lemma:

- D. Achlioptas. Database-friendly random projections. In Proc. Symp. on Principles of Database Systems, pages 274–281. ACM Press, 2001.
- S. Dasgupta and A. Gupta. An elementary proof of the Johnson-Lindenstrauss lemma. Technical Report TR-99-006, Berkeley, CA, 1999.
- P. Frankl and H. Maehara, The Johnson-Lindenstrauss lemma and the sphericity of some graphs, J. Combinatorial Theory Ser. B 44 (1988), no. 3, pp. 355–362.
- P. Indyk and R. Motwani, Approximate nearest neighbors: Towards removing the curse of dimenstionality, Symp. on Theory of Computing, 1998, pp. 604–613.

Geometric proofs of RIP/UUP:

- Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin, A simple proof of the restricted isometry property for random matrices. (To appear in Constructive Approximation)
- S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, Uniform uncertainty principle for Bernoulli and subgaussian ensembles. (Preprint, 2006)

References – Geometry (4)

Manifolds and Other Embeddings:

- Hassler Whitney, Differentiable manifolds, Ann. Math. 37:645-680 (1936).
- Tim Sauer, James A. Yorke, and Martin Casdagli, Embedology, Journal of Statistical Physics, Volume 65, Numbers 3-4, November, 1991.
- Richard Baraniuk and Michael Wakin, Random projections of smooth manifolds. (To appear in Foundations of Computational Mathematics)
- P. Indyk and A. Naor. Nearest neighbor preserving embeddings. ACM Trans. Algorithms, Volume 3, Issue 3 (August 2007).
- Pankaj K. Agarwal, Sariel Har-Peled, and Hai Yu, Embeddings of Surfaces, Curves, and Moving Points in Euclidean Space, Proceedings of the twenty-third annual symposium on Computational geometry, 2007.
- S. Dasgupta and Y. Freund, Random projection trees and low dimensional manifolds, UCSD Technical Report CS2007-0890, 2007.
- Mark Davenport, Marco Duarte, Michael Wakin, Jason Laska, Dharmpal Takhar, Kevin Kelly, and Richard Baraniuk, The smashed filter for compressive classification and target recognition. (Computational Imaging V at SPIE Electronic Imaging, San Jose, California, January 2007)
- Chinmay Hegde, Michael Wakin, and Richard Baraniuk, Random projections for manifold learning. (Neural Information Processing Systems (NIPS), Vancouver, Canada, December 2007)

Agenda

- Introduction to Compressive Sensing (CS) [richb]
  - motivation
  - basic concepts

- CS Theoretical Foundation [justin]
  - uniform uncertainty principles
  - restricted isometry principle
  - recovery algorithms

- Geometry of CS [mike]
  - K-sparse and compressible signals
  - manifolds

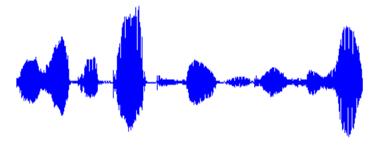
- **CS Applications** [richb]

Applications
of
Compressive Sensing

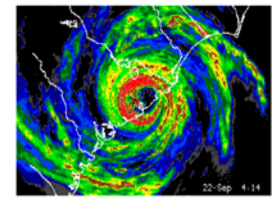
CS Hallmarks

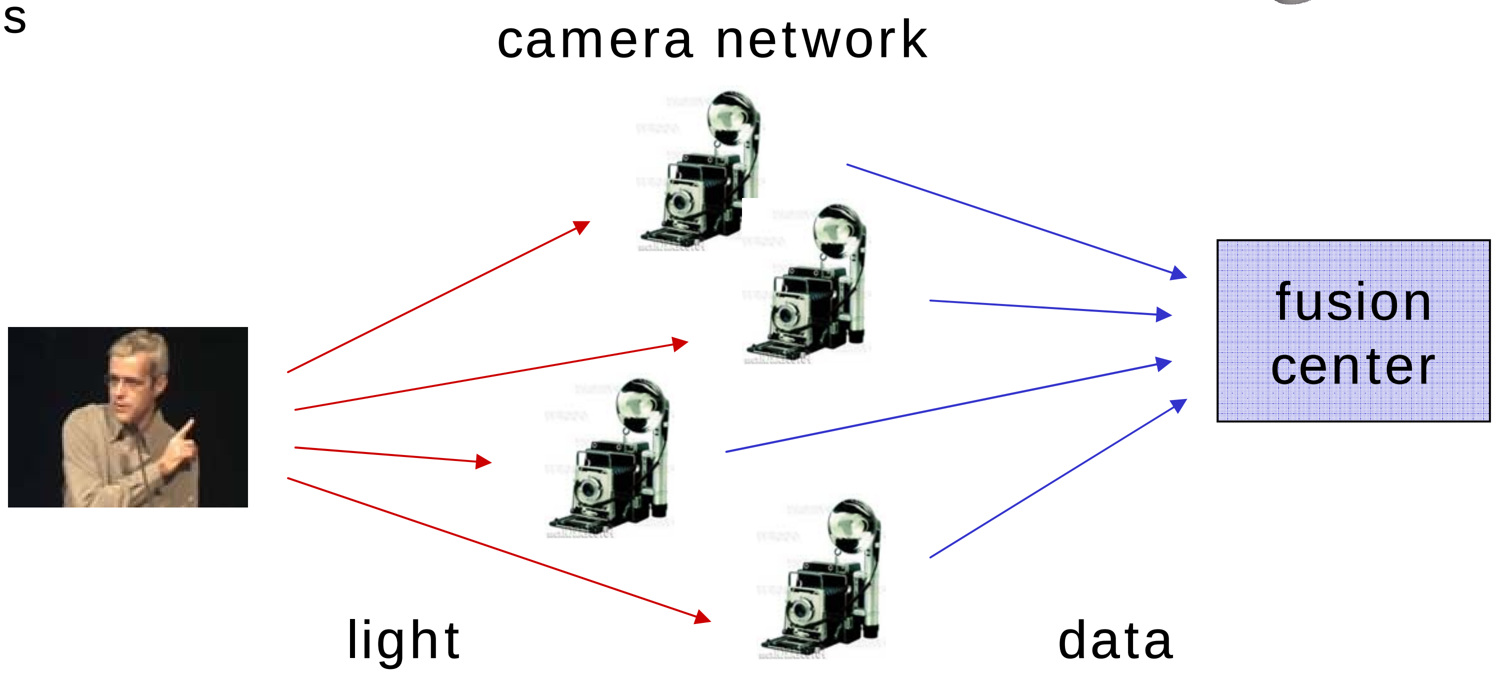
- CS changes the rules of the data acquisition game
  - exploits a priori signal *sparsity* information
- **Universal**
  - same random projections / hardware can be used for *any* compressible signal class (*generic*)
- **Democratic**
  - each measurement carries the same amount of information
  - simple encoding
  - robust to measurement loss and quantization
- **Asymmetrical** (most processing at decoder)
- Random projections weakly **encrypted**

CS for Sensor Networks



- Measurement, monitoring, tracking of *distributed physical phenomena* (“macroscope”) using wireless embedded sensors
  - environmental conditions
  - industrial monitoring
  - chemicals
  - weather
  - sounds



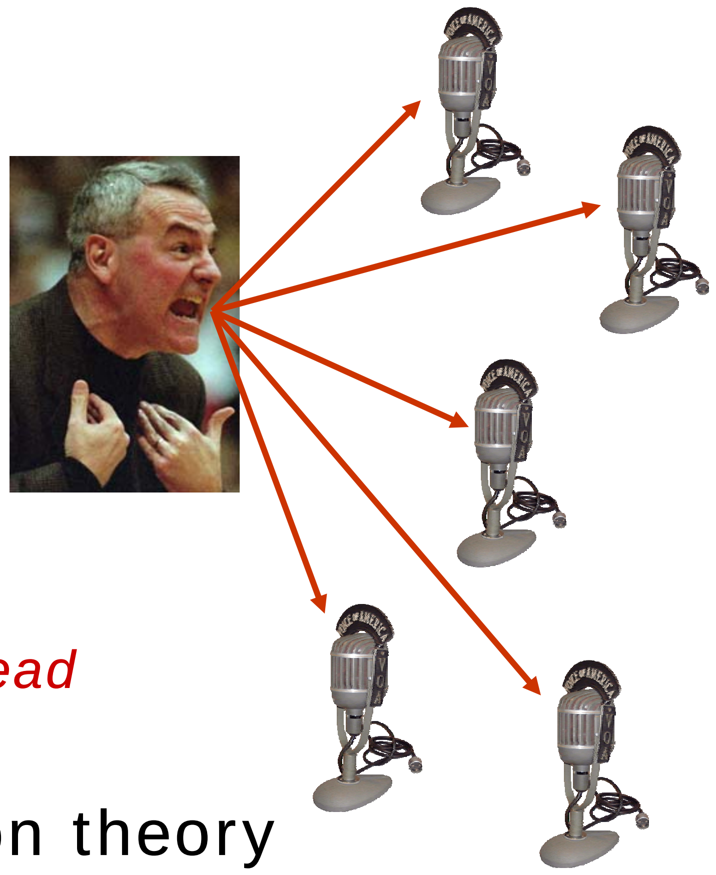


Sensor Network Challenges

- Computational/power ***asymmetry***
  - limited compute power on each sensor node
  - limited (battery) power on each sensor node
- Must be ***energy efficient***
  - minimize communication
- Hostile ***communication*** environment
  - multi-hop
  - high loss rate
- ***Deluge of data*** at collection point
  - bandwidth collapse

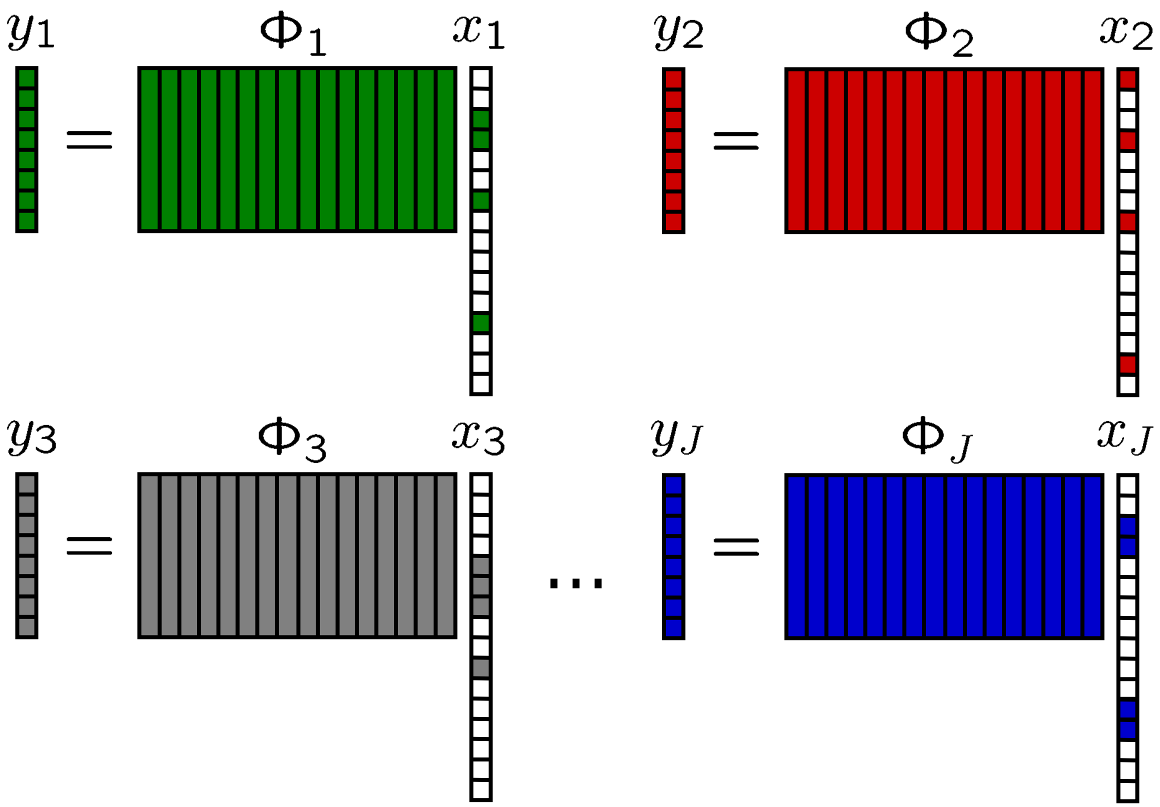
Multi-Signal CS

- Sensor networks:
  intra-sensor and
  inter-sensor correlation
- Can we exploit these to *jointly compress?*
- Popular approach: *collaboration*
  - inter-sensor *communication overhead*
- Ongoing challenge in information theory
- One solution: Compressive Sensing



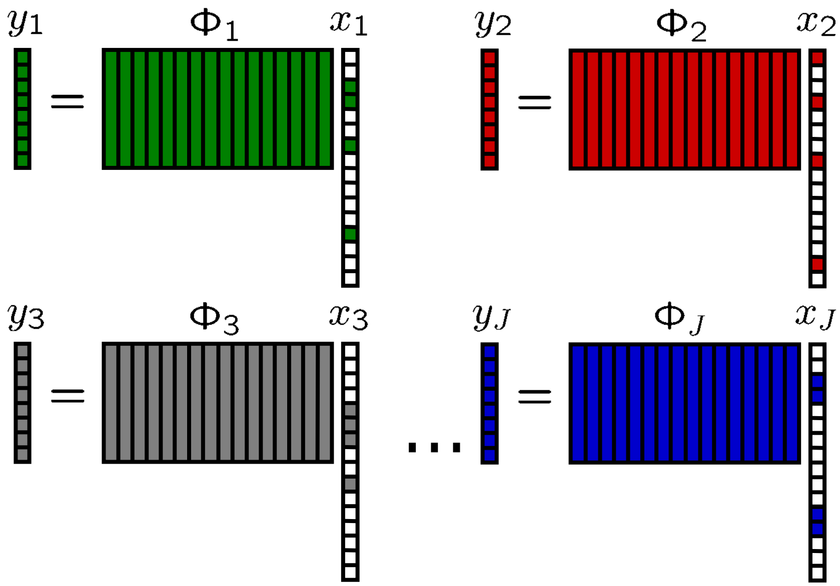
Distributed CS (DCS)

- "Measure separately, reconstruct *jointly*"
- Ingredients
  - models for **joint sparsity**
  - algorithms for joint reconstruction
  - theoretical results for measurement savings



Distributed CS (DCS)

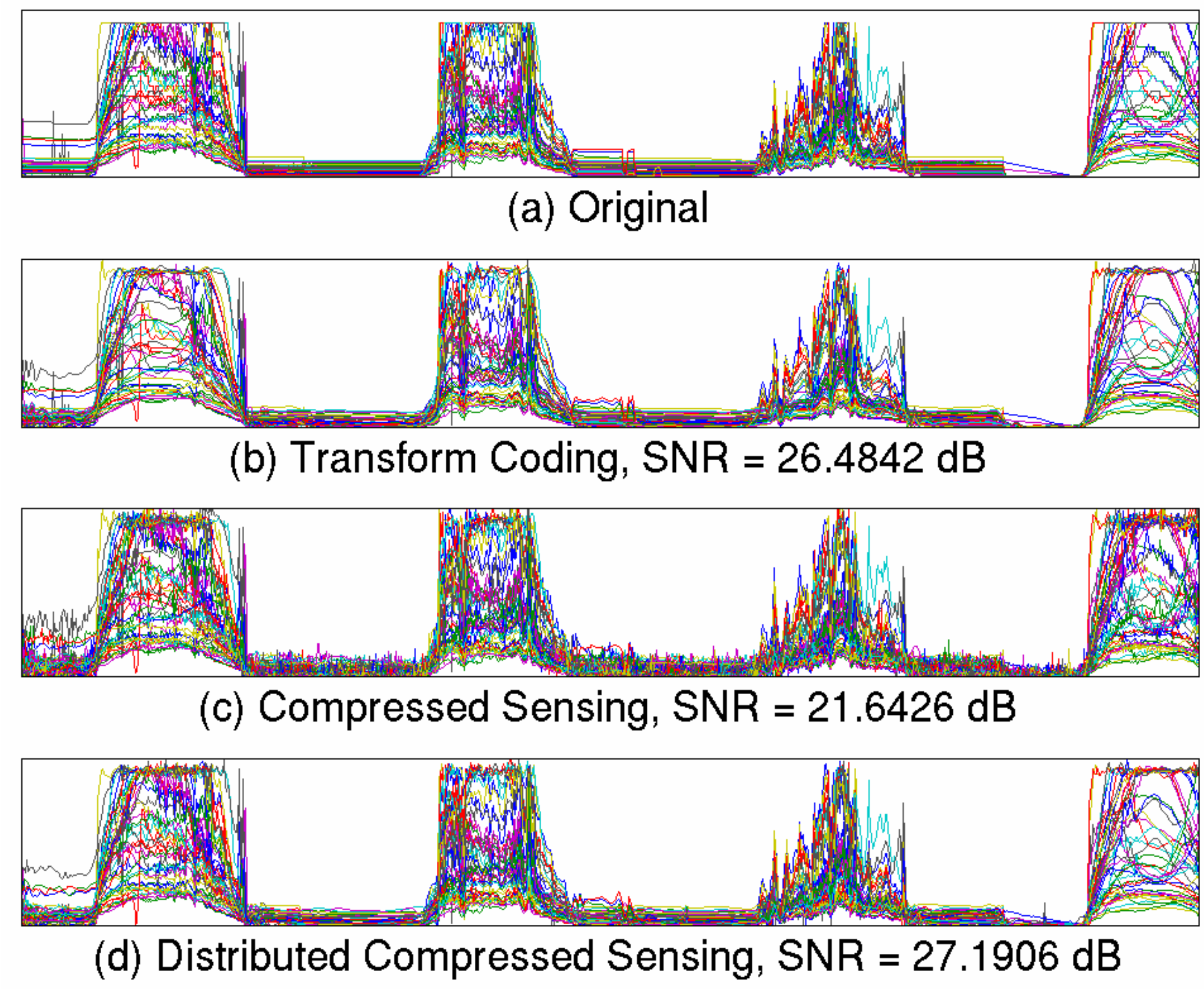
- "Measure separately, reconstruct *jointly*"
- Ingredients
  - models for joint sparsity
  - algorithms for joint reconstruction
  - theoretical results for measurement savings



- Zero collaboration, trivially scalable, robust
- Low complexity, universal encoding

Real Data Example

- Light Sensing in Intel Berkeley Lab
- 49 sensors, $N = 1024$ samples each, Ψ = wavelets



K=100

M=400

M=400

References – Data Compression (1)

Information Theoretic:

- Emmanuel Candès and Terence Tao, Near optimal signal recovery from random projections: Universal encoding strategies? (IEEE Trans. on Information Theory, 52(12), pp. 5406 - 5425, December 2006)
- David Donoho, Compressed sensing. (IEEE Trans. on Information Theory, 52(4), pp. 1289 - 1306, April 2006)
- Emmanuel Candès and Justin Romberg, Encoding the ell-p ball from limited measurements. (Proc. IEEE Data Compression Conference (DCC), Snowbird, UT, 2006)
- Shriram Sarvotham, Dror Baron, and Richard Baraniuk, Measurements vs. bits: Compressed sensing meets information theory. (Proc. Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 2006)
- Petros Boufounos and Richard Baraniuk, Quantization of sparse representations. (Rice ECE Department Technical Report TREE 0701 - Summary appears in Proc. Data Compression Conference (DCC), Snowbird, Utah, March 2007)

References – Data Compression (2)

Sensor Networks and Multi-Signal CS:

- Dror Baron, Michael Wakin, Marco Duarte, Shriram Sarvotham, and Richard Baraniuk, Distributed compressed sensing. (Preprint, 2005)
- Waheed Bajwa, Jarvis Haupt, Akbar Sayeed, and Rob Nowak, Compressive wireless sensing. (Proc. Int. Conf. on Information Processing in Sensor Networks (IPSN), Nashville, Tennessee, April 2006)
- Rémi Gribonval, Holger Rauhut, Karin Schnass, and Pierre Vandergheynst, Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms. (Preprint, 2007)
- Wei Wang, Minos Garofalakis, and Kannan Ramchandran, Distributed sparse random projections for refinable approximation. (Proc. Int. Conf. on Information Processing in Sensor Networks (IPSN), Cambridge, Massachusetts, April 2007)

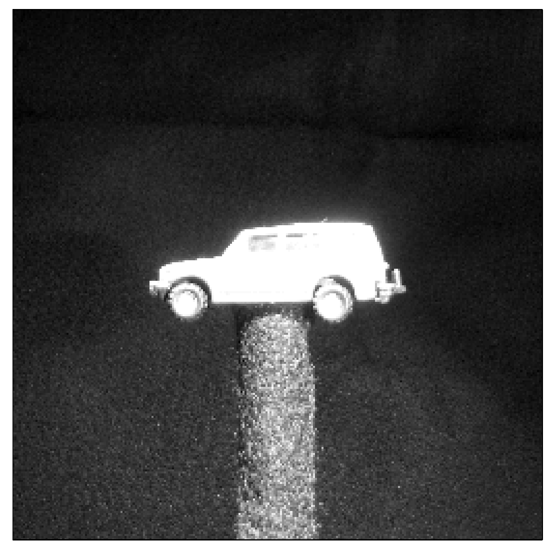
Compressive Signal Processing

- ***Information Scalability***: If we can ***reconstruct*** a signal from compressive measurements, then we should be able to perform other kinds of statistical signal processing:
  - ***detection***
  - ***classification***
  - ***estimation***
  - ***learning …***
- Number of measurements should relate to complexity of inference
- Ex: compressive matched filter >> "smashed filter"

Matched Filter

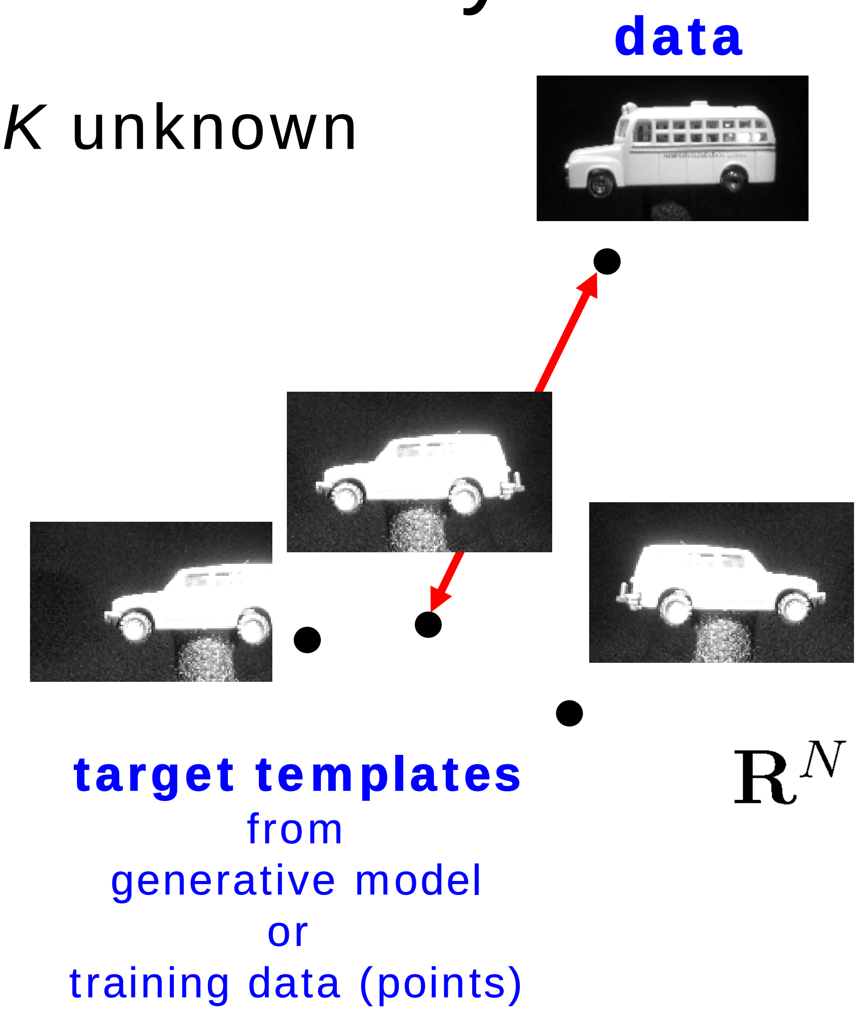
- Detection/classification with *K* unknown **articulation parameters**
  - Ex: position and pose of a vehicle in an image
  - Ex: time delay of a radar signal return
- **Matched filter**: joint parameter estimation and detection/classification
  - compute sufficient statistic for each potential target and articulation
  - compare "best" statistics to detect/classify





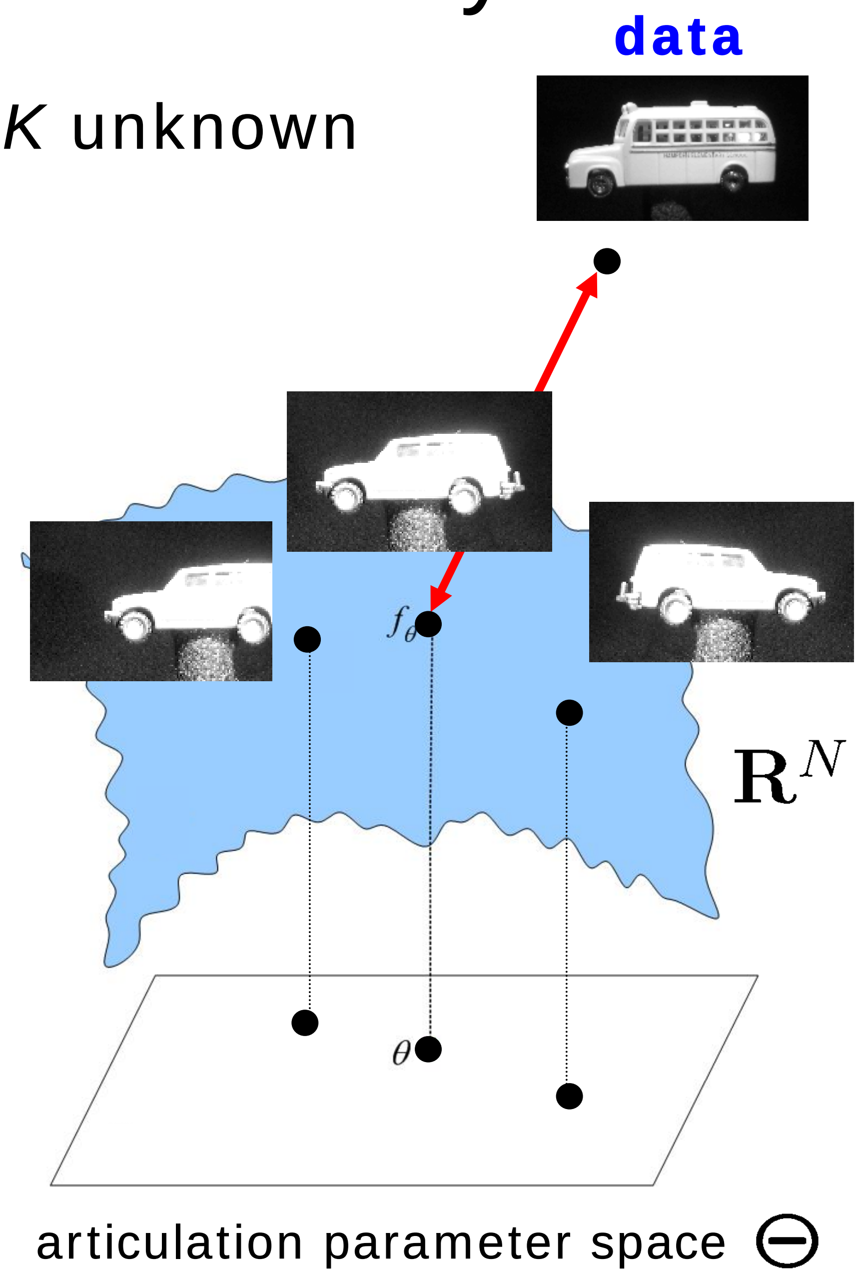
Matched Filter Geometry

- Detection/classification with *K* unknown articulation parameters
- Images are points in $\mathbf{R}^N$
- **Classify** by finding closest target template to data for each class (AWG noise)
  - distance or inner product



Matched Filter Geometry

- Detection/classification with *K* unknown articulation parameters
- Images are points in $\mathbf{R}^N$
- Classify by finding closest target template to data
- As template articulation parameter changes, points map out a *K*-dim **nonlinear manifold**
- Matched filter classification = **closest manifold search**

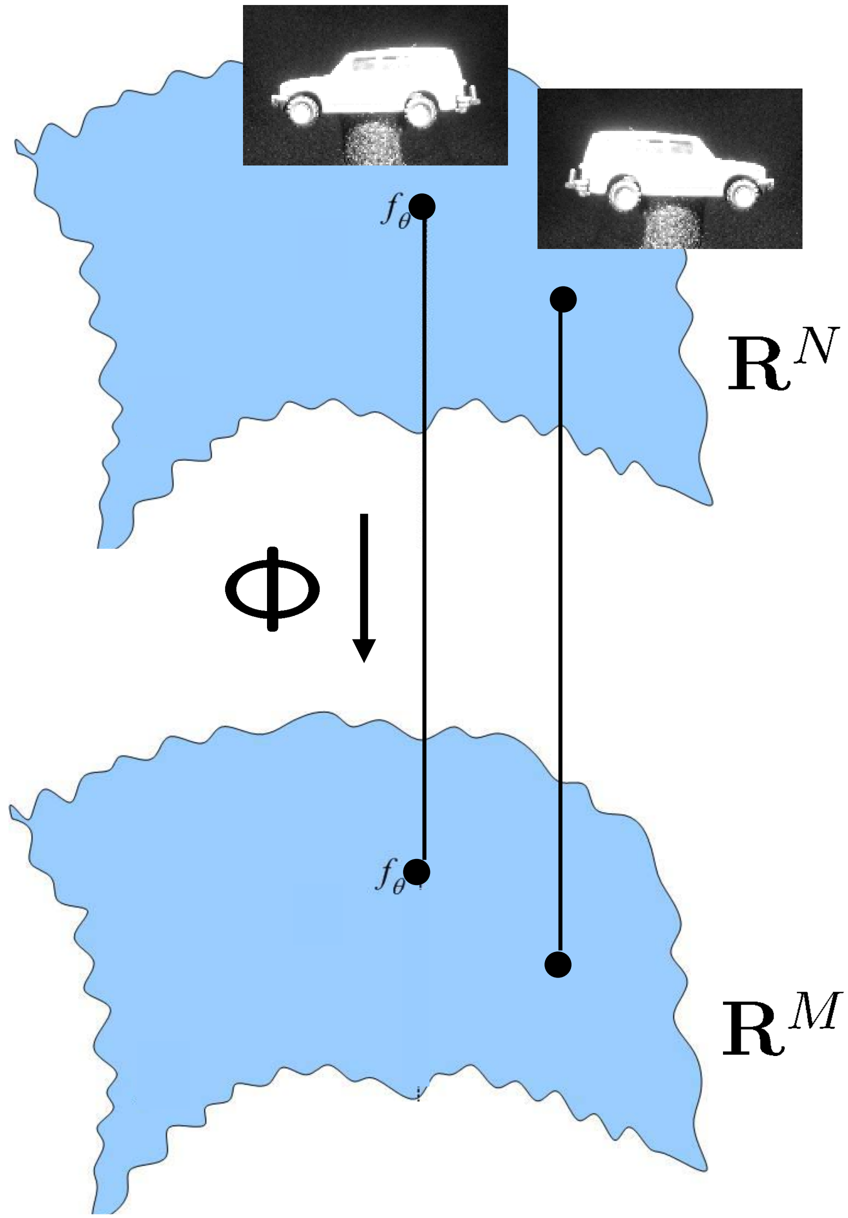


CS for Manifolds

- **Theorem:**
  $$M = O(K \log N)$$
  random measurements **preserve manifold structure**

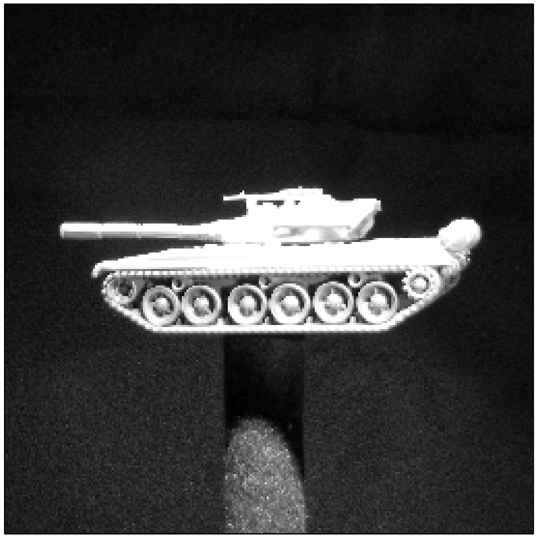
  [Wakin et al, FOCM '08]

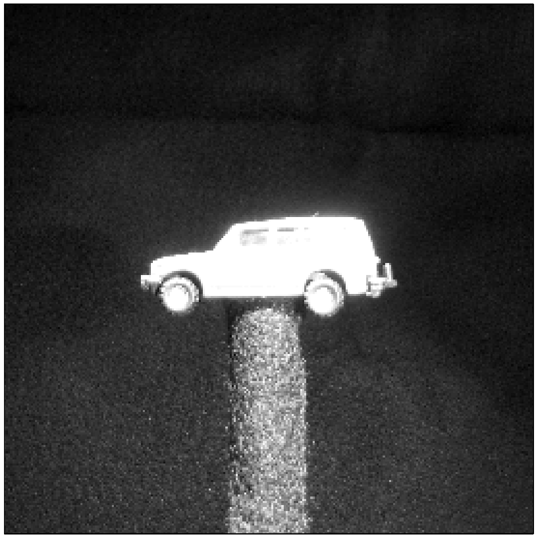
- Enables parameter estimation and MF detection/classification directly on compressive measurements
  - *K* very small in many applications



Example: Matched Filter

- Detection/classification with $K = 3$ unknown **articulation parameters**
  1. horizontal translation
  2. vertical translation
  3. rotation



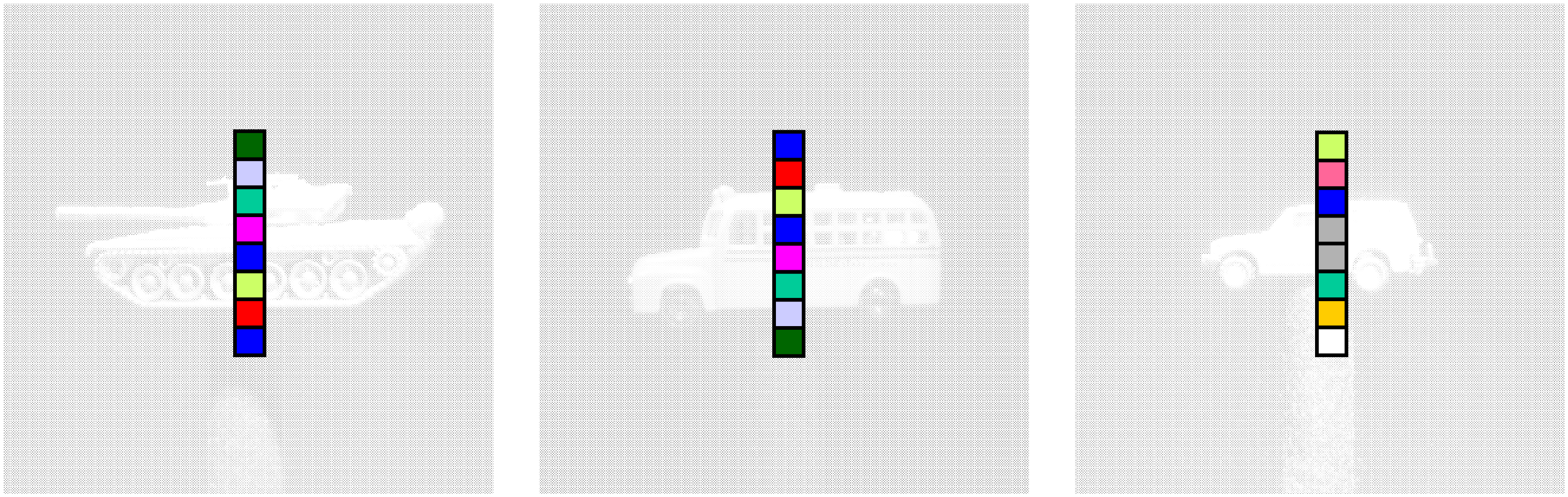


Smashed Filter

- Detection/classification with K=3 unknown articulation parameters **(manifold structure)**

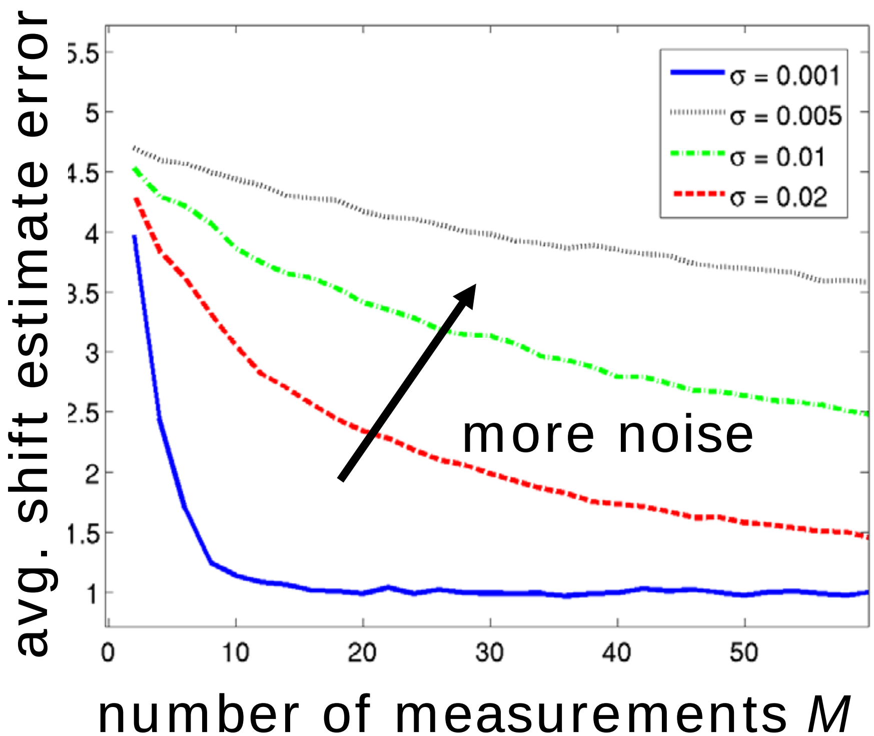
- Dimensionally reduced matched filter directly on compressive measurements

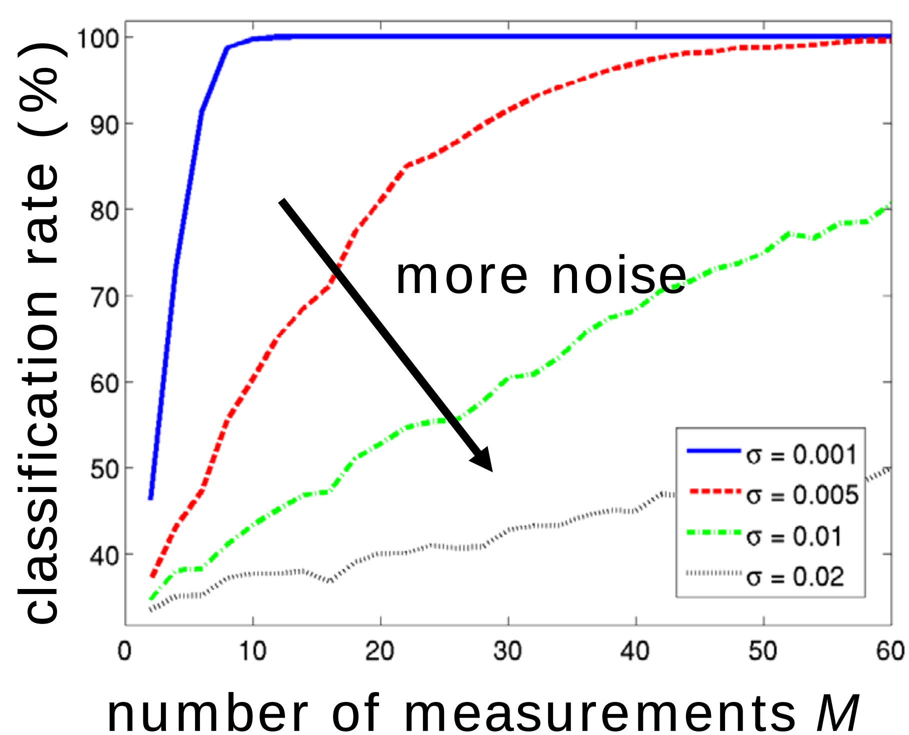
$$M = O(K \log N)$$



Smashed Filter

- Random shift and rotation (K=3 dim. manifold)
- Noise added to measurements
- Goal: identify most likely position for each image class
  identify most likely class using nearest-neighbor test





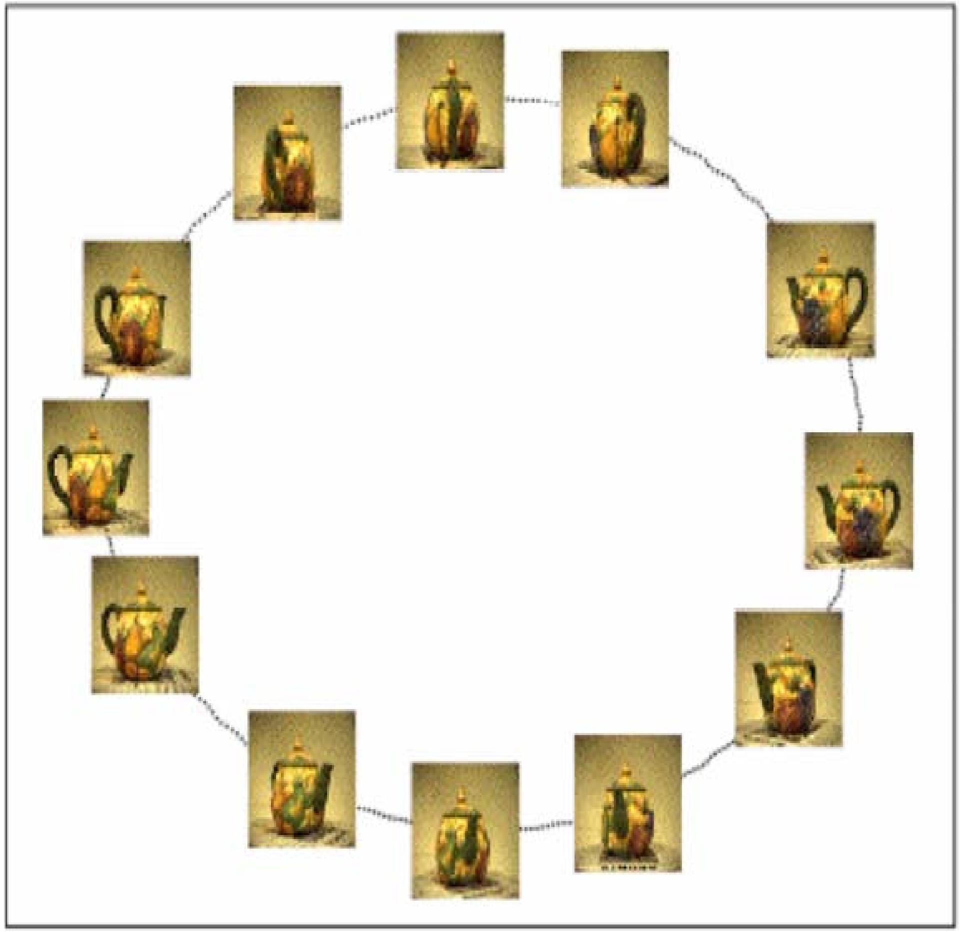
Manifold Learning

- Given training points in $\mathbf{R}^N$, learn the mapping to the underlying K-dimensional articulation manifold

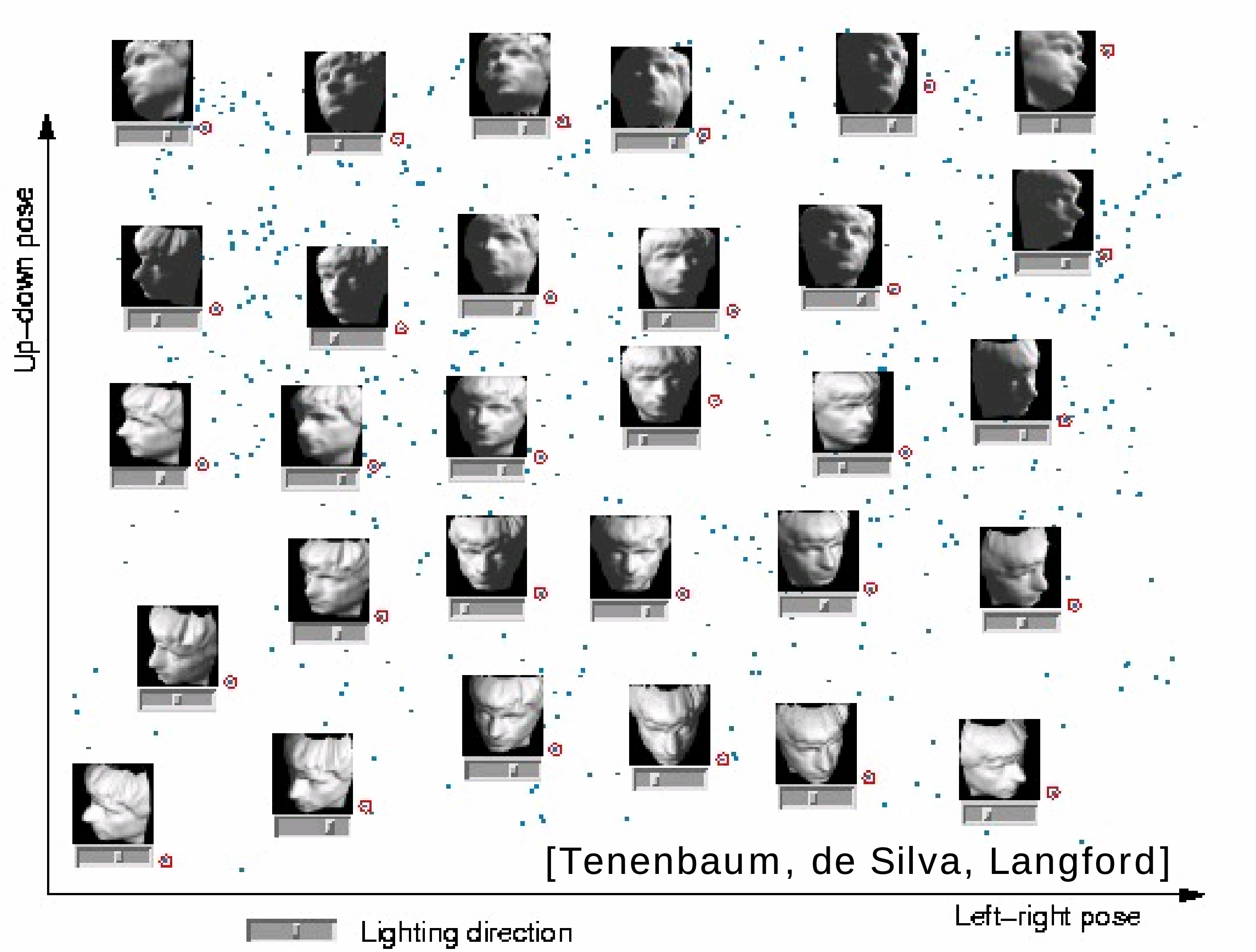
- ISOMAP, LLE, HLLE, ...

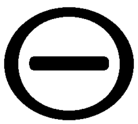
- Ex: images of
  rotating teapot

  articulation space
  = circle



Up/Down Left/Right Manifold





K = 2

Compressive Manifold Learning

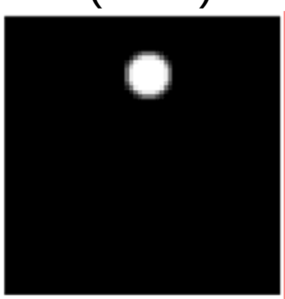
- ISOMAP algorithm based on geodesic distances between points

- Random measurements preserve these distances

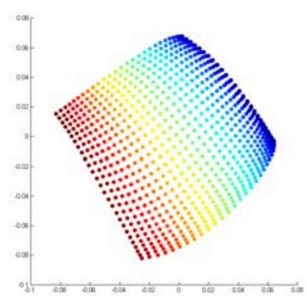
- **Theorem**: If $M = O(K \log(N)/\delta^2)$, then the ISOMAP residual variance in the projected domain is bounded by the additive error factor $R_\Phi < R + C\delta$

[Hegde et al '08]

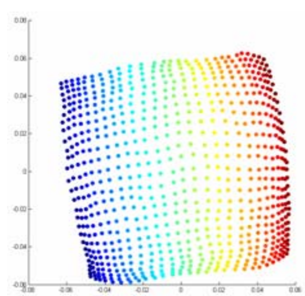
translating disk manifold (K=2)



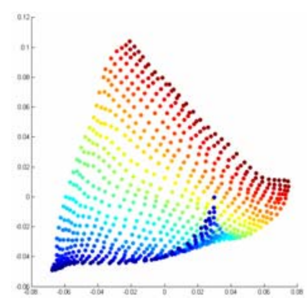
full data (N=4096)



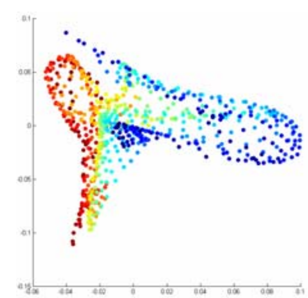
M = 100



M = 50



M = 25



References – Compressive DSP (1)

Statistical Signal Processing & Information Scalability:

- D. Waagen, N. Shah, M. Ordaz, and M. Cassabaum, "Random subspaces and SAR classification efficacy," in Proc. SPIE Algorithms for Synthetic Aperture Radar Imagery XII, May 2005.
- Marco Duarte, Mark Davenport, Michael Wakin, and Richard Baraniuk, Sparse signal detection from incoherent projections. (Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Toulouse, France, May 2006)
- Mark Davenport, Michael Wakin, and Richard Baraniuk, Detection and estimation with compressive measurements. (Rice ECE Department Technical Report TREE 0610, November 2006)
- Jarvis Haupt, Rui Castro, Robert Nowak, Gerald Fudge, and Alex Yeh, Compressive sampling for signal classification. (Proc. Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, California, October 2006)

References – Compressive DSP (2)

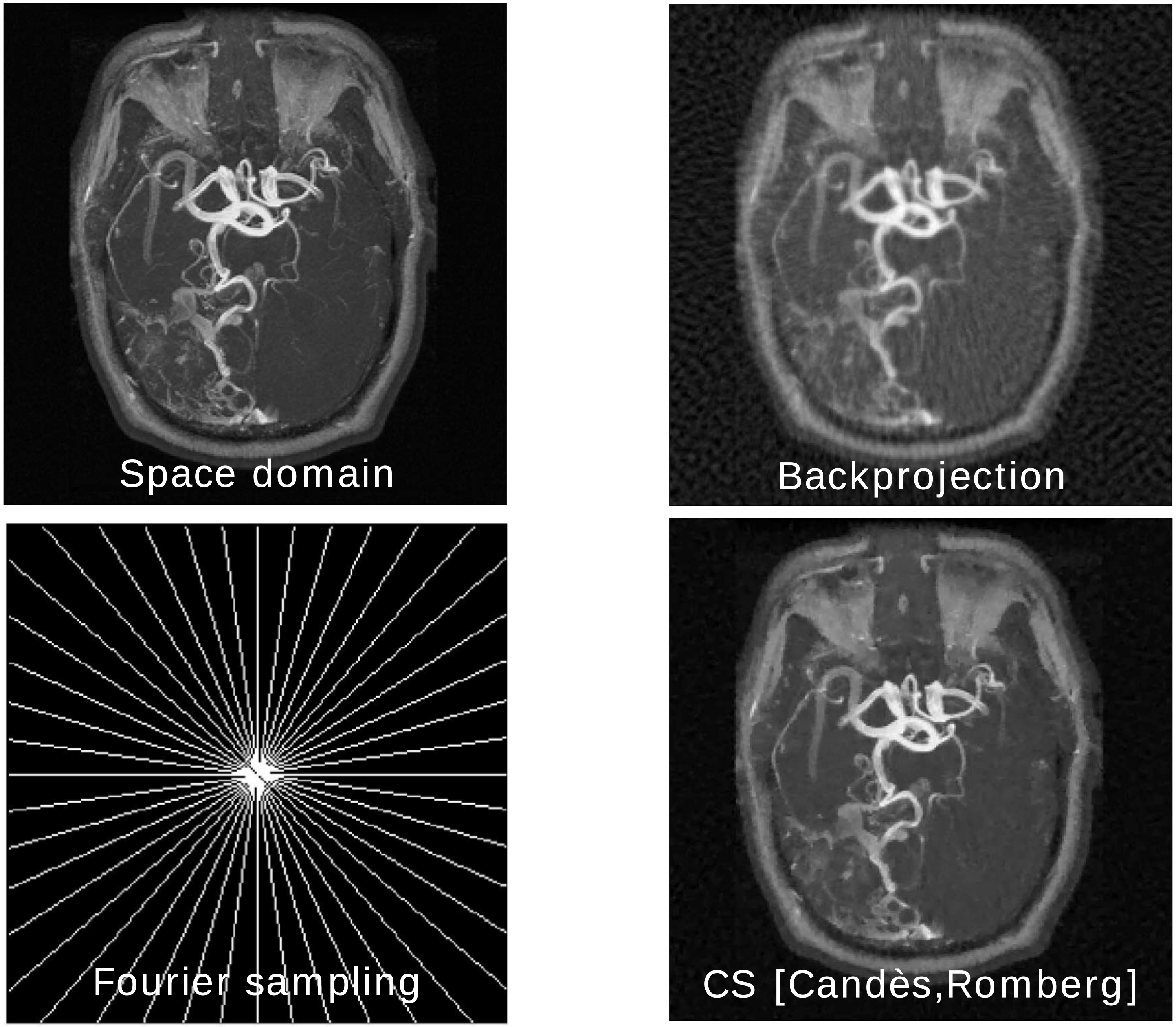
Manifolds, Manifold Learning, Smashed Filter:

- Richard Baraniuk and Michael Wakin, Random projections of smooth manifolds. (To appear in Foundations of Computational Mathematics)
- Mark Davenport, Marco Duarte, Michael Wakin, Jason Laska, Dharmpal Takhar, Kevin Kelly, and Richard Baraniuk, The smashed filter for compressive classification and target recognition. (Proc. of Computational Imaging V at SPIE Electronic Imaging, San Jose, California, January 2007)

Theoretical Computer Science & Data Streaming Algorithms:

- N. Alon, P. Gibbons, Y. Matias, and M. Szegedy, “Tracking join and self-join sizes in limited storage," in Proc. Symp. Principles of Database Systems (PODS), Philadelphia, PA, 1999.
- Nitin Thaper, Sudipto Guha, Piotr Indyk, and Nick Koudas, Dynamic multidimensional histograms. (Proc. SIGMOD 2002, Madison, Wisconson, June 2002)
- Anna Gilbert, Sudipto Guha, Piotr Indyk, Yannis Kotidis, S. Muthukrishnan, and Martin J. Strauss, Fast small-space algorithms for approximate histogram maintenance. (Proc. 34th Symposium on Theory of Computing, Montréal, Canada, May 2002)
- S. Muthukrishnan, Data Streams: Algorithms and Applications, now, 2005.

Magnetic Resonance Imaging



References – Inverse Problems

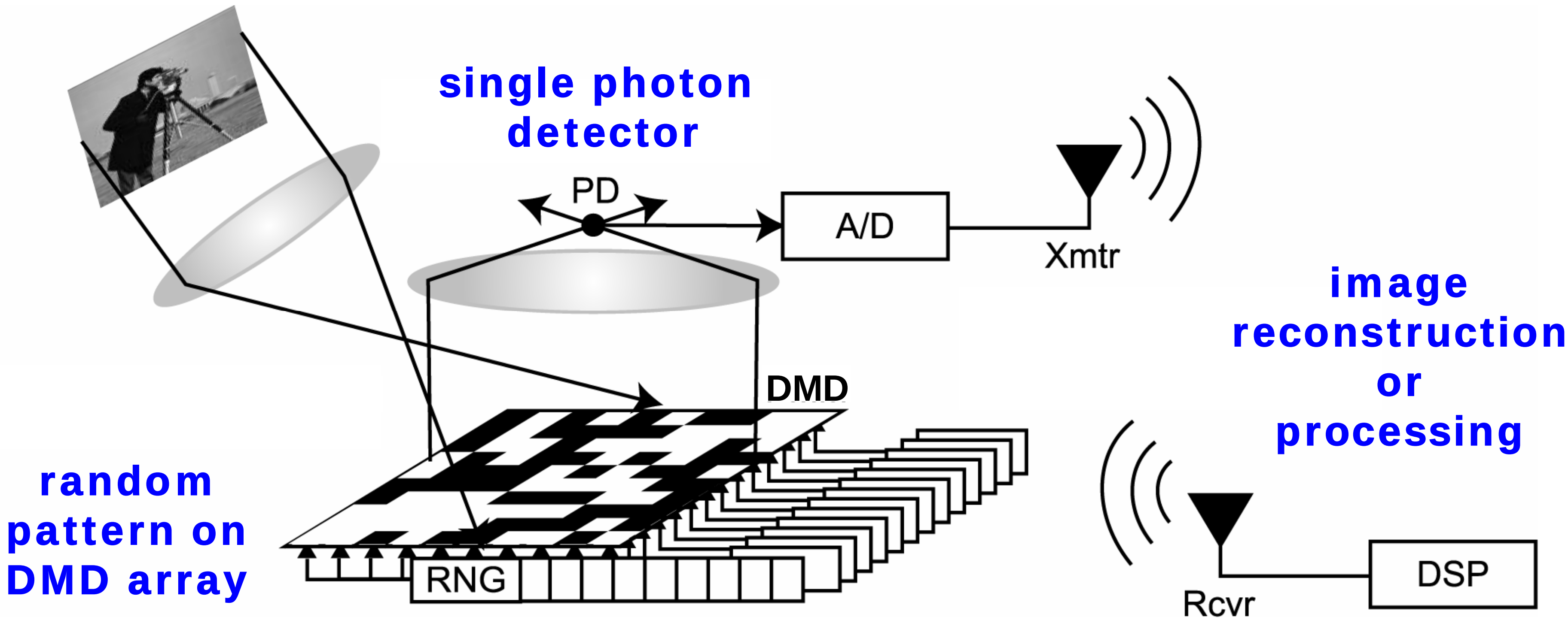
Medical Imaging:

- Emmanuel Candès, Justin Romberg, and Terence Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. (IEEE Trans. on Information Theory, 52(2) pp. 489 - 509, February 2006)
- Michael Lustig, David Donoho, and John M. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging. (Preprint, 2007)
- Jong Chul Ye, Compressed sensing shape estimation of star-shaped objects in Fourier imaging (Preprint, 2007)

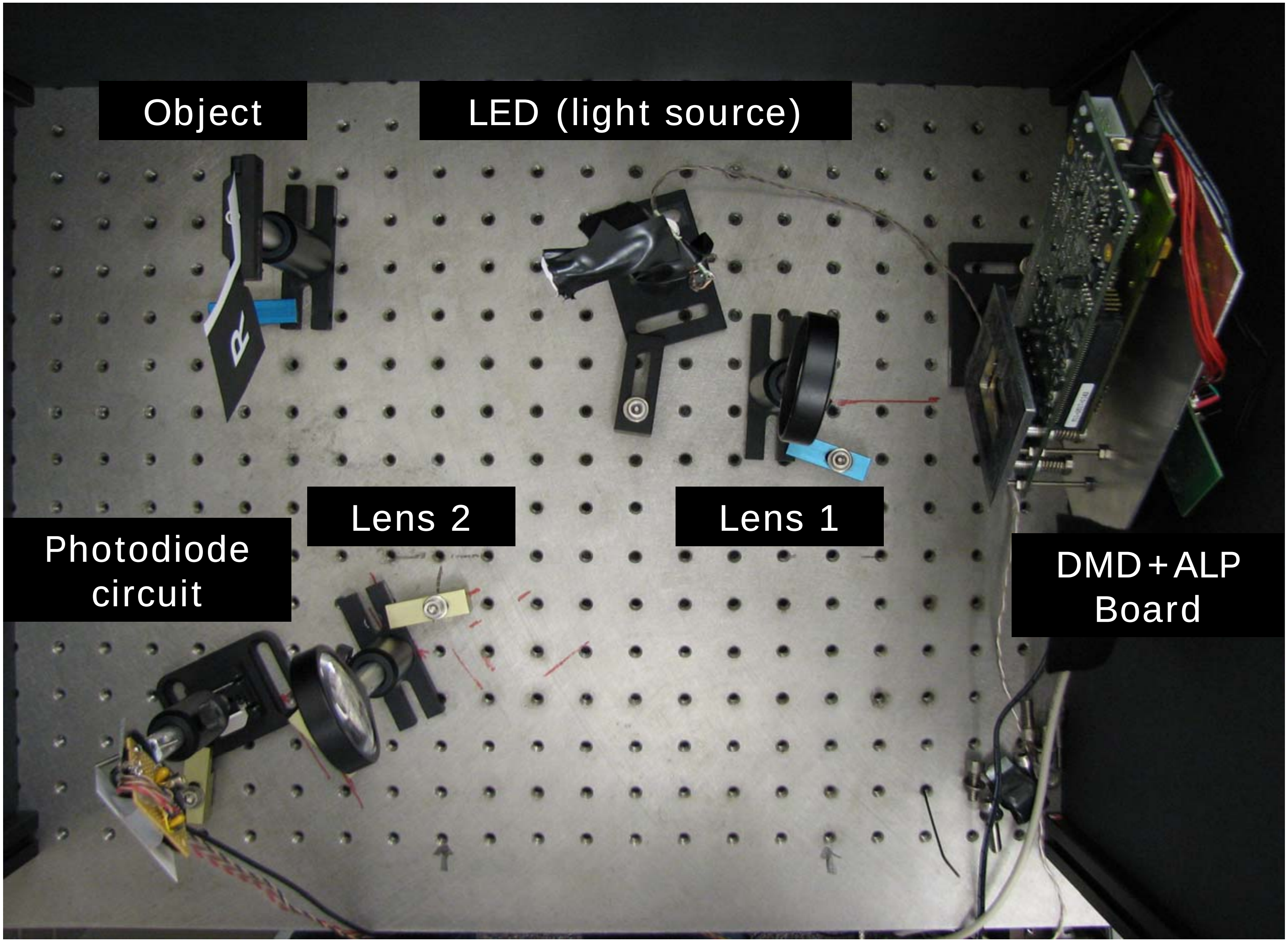
Other:

- Ingrid Daubechies, Massimo Fornasier, and Ignace Loris, Accelerated projected gradient method for linear inverse problems with sparsity constraints. (Preprint, 2007)
- Mário A. T. Figueiredo, Robert D. Nowak, and Stephen J. Wright, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. (Preprint, 2007)
- José Bioucas-Dias and Mário Figueiredo, A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration. (Preprint, 2007)
- Lawrence Carin, Dehong Liu, and Ya Xue, *In Situ* Compressive Sensing. (Preprint, 2007)

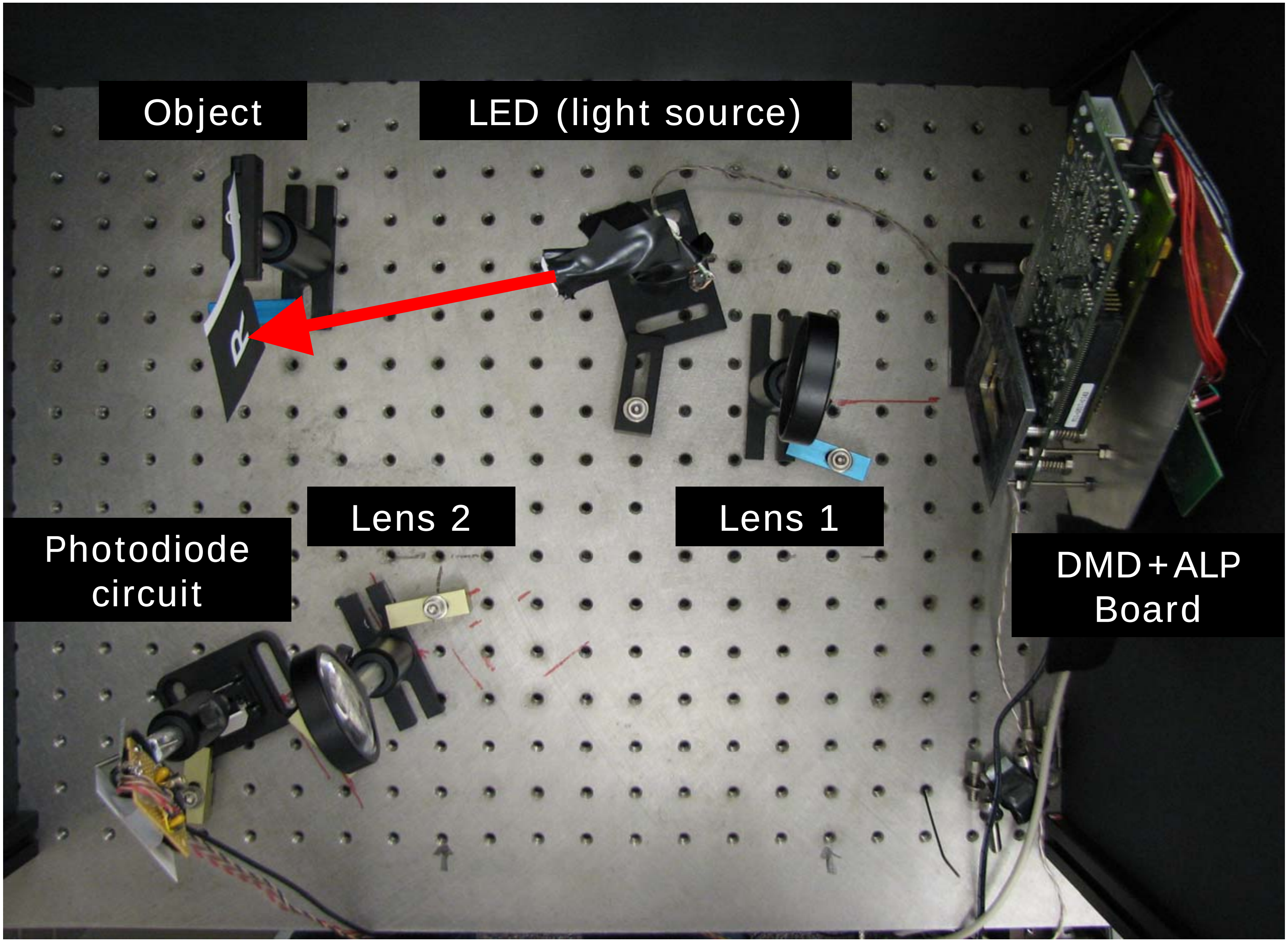
Rice Single-Pixel CS Camera



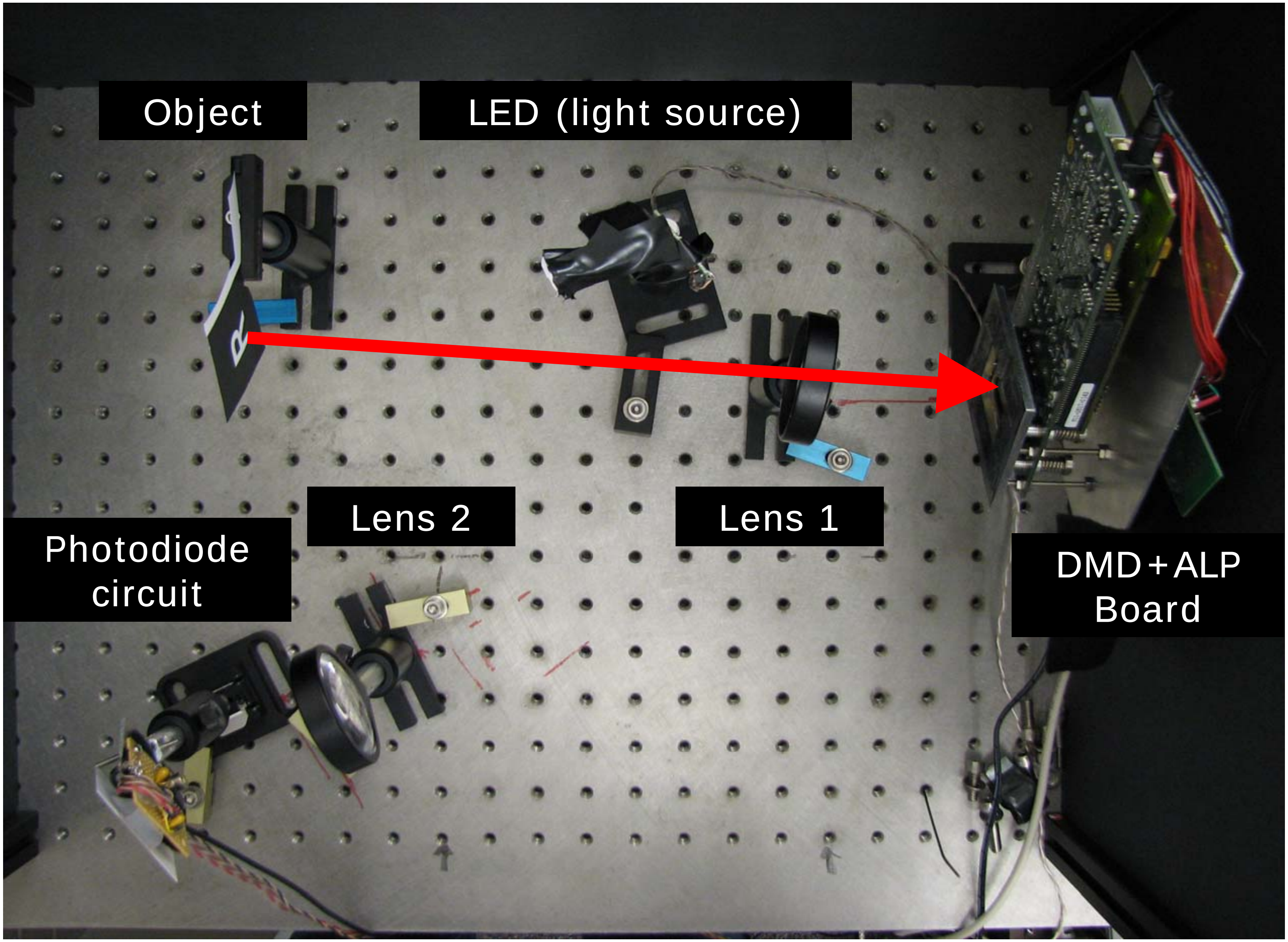
Single Pixel Camera



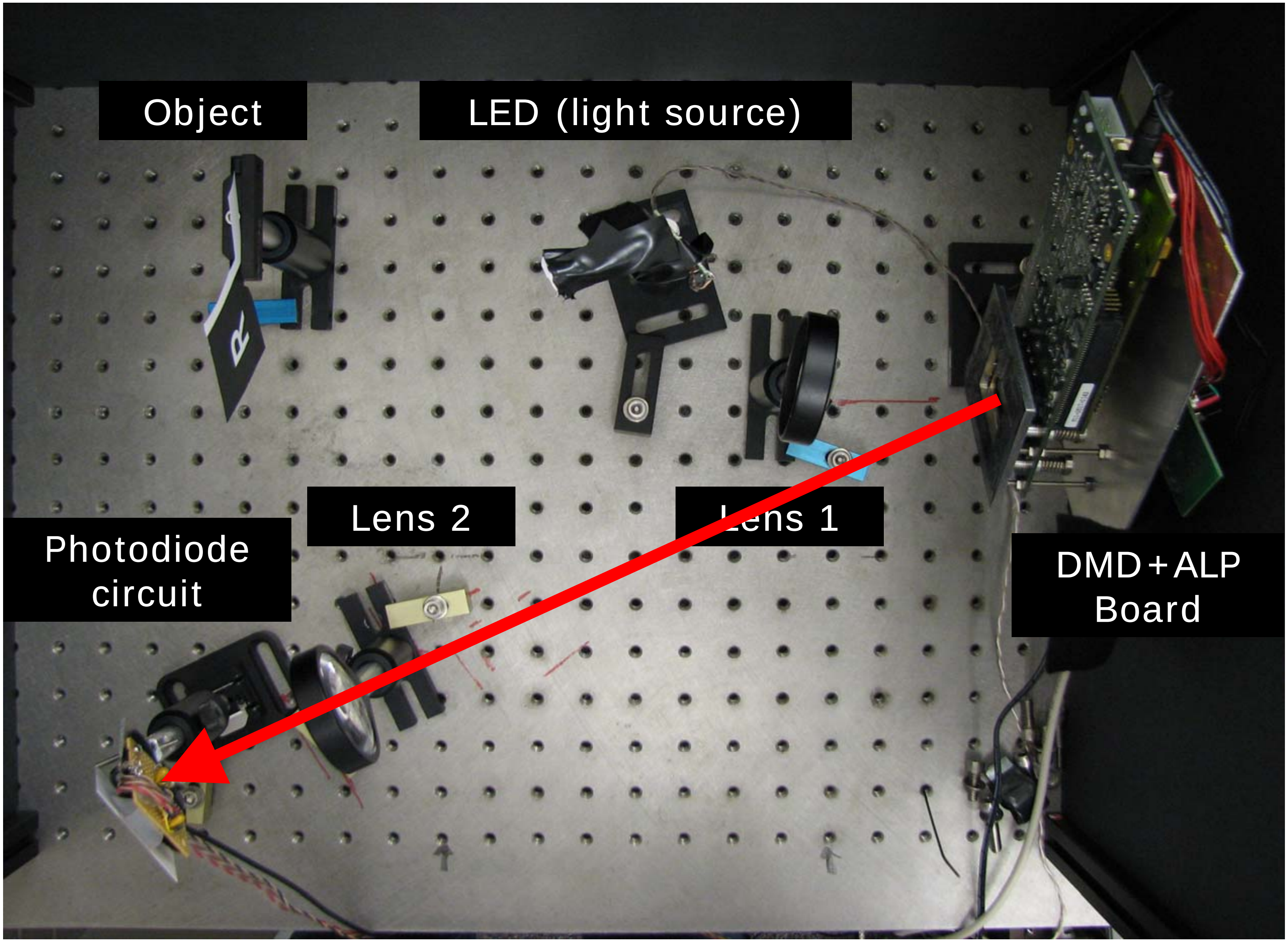
Single Pixel Camera



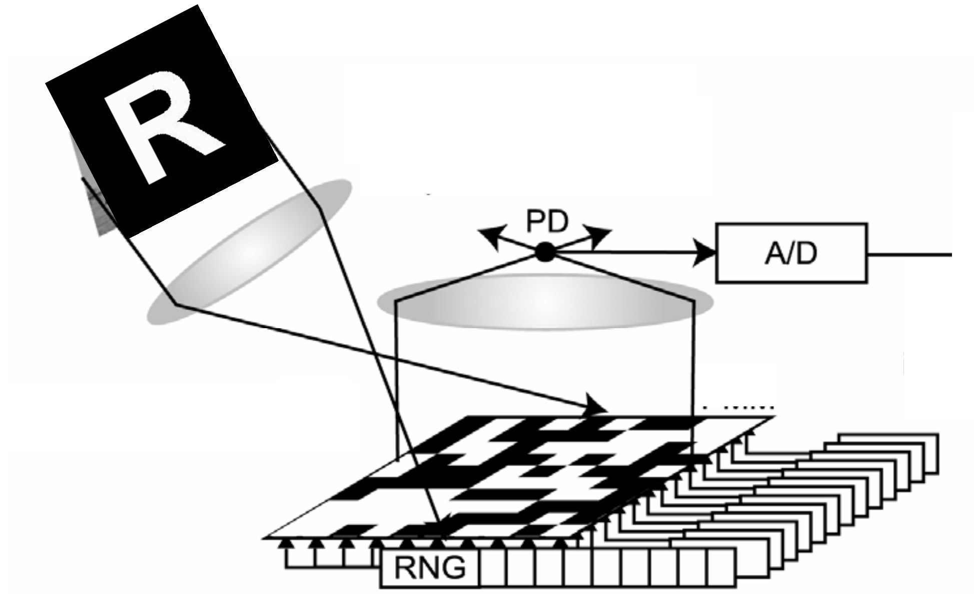
Single Pixel Camera



Single Pixel Camera

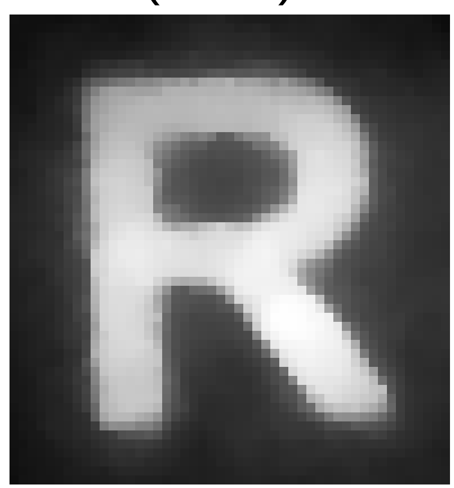


First Image Acquisition



target
65536 pixels

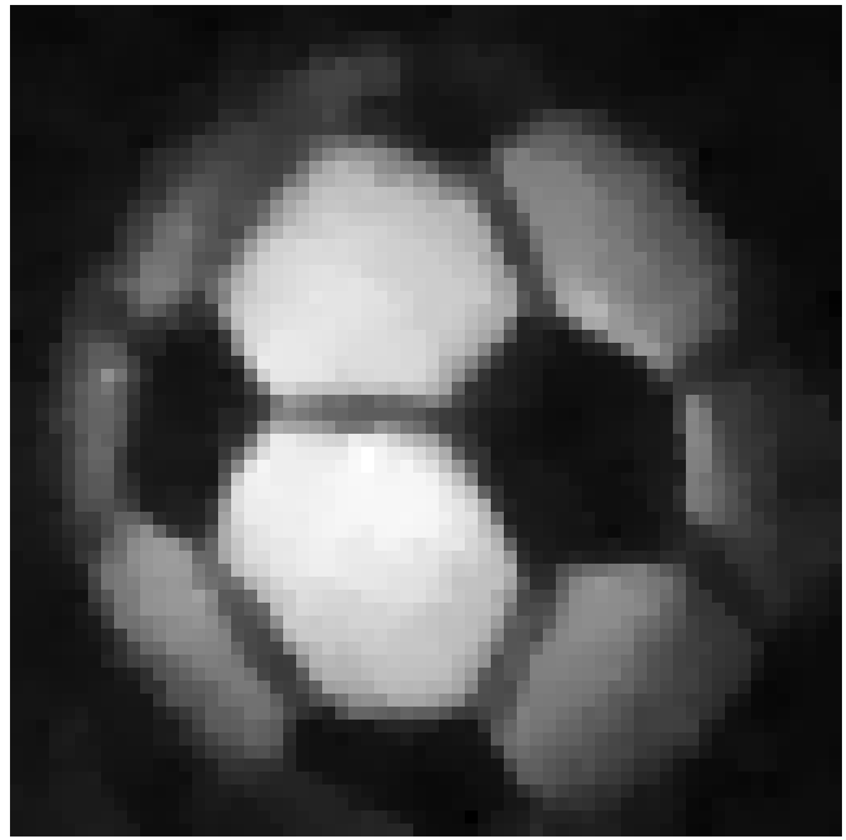
11000 measurements
(16%)



1300 measurements
(2%)

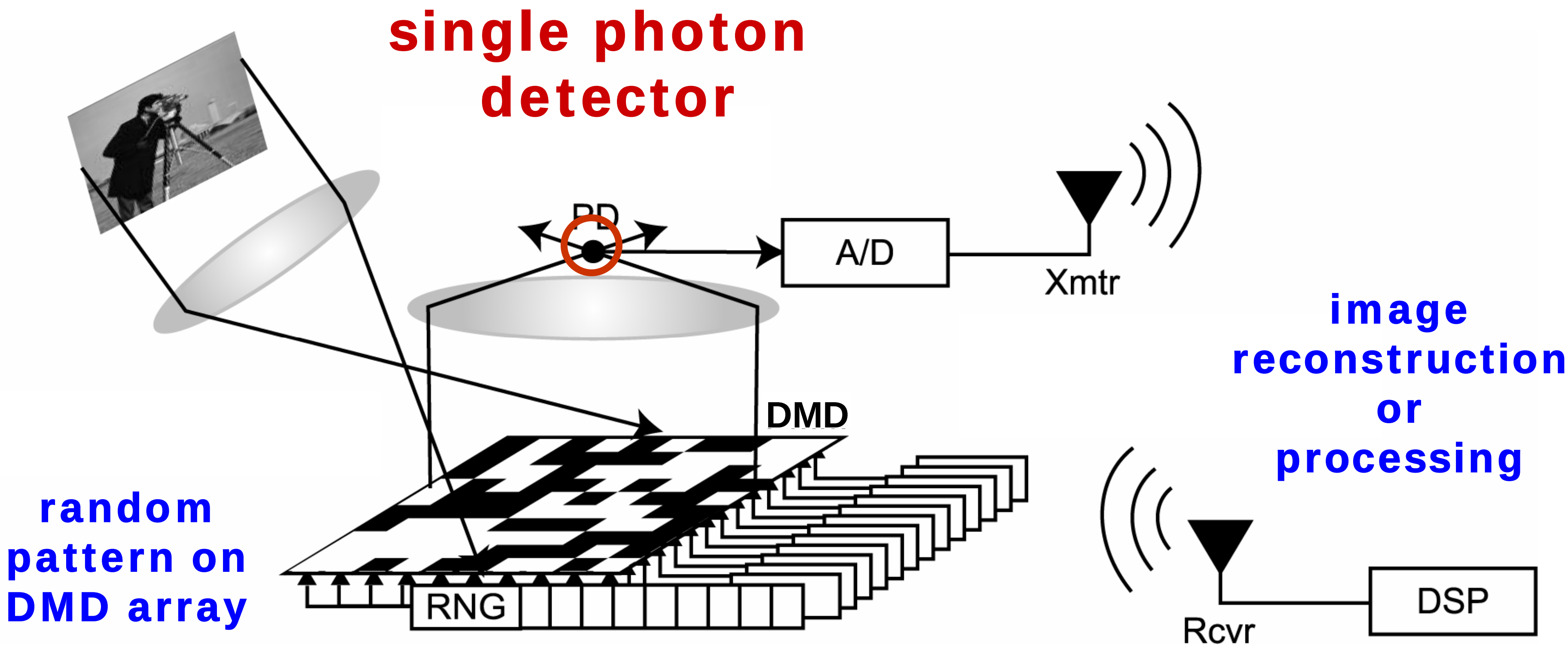
Second Image Acquisition

4096
pixels

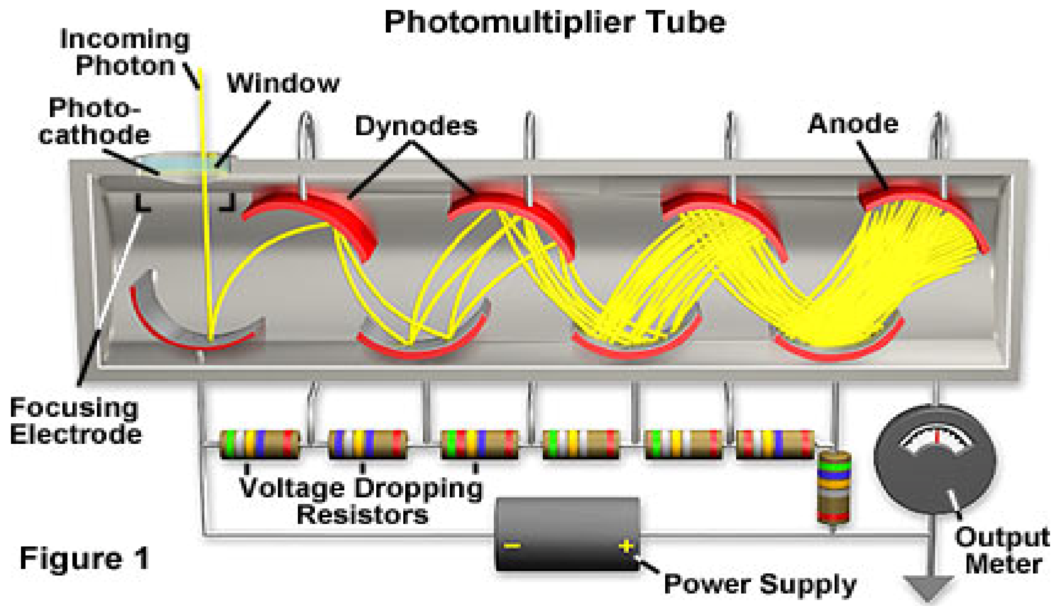


500
random measurements

Single-Pixel Camera



CS Low-Light Imaging with PMT



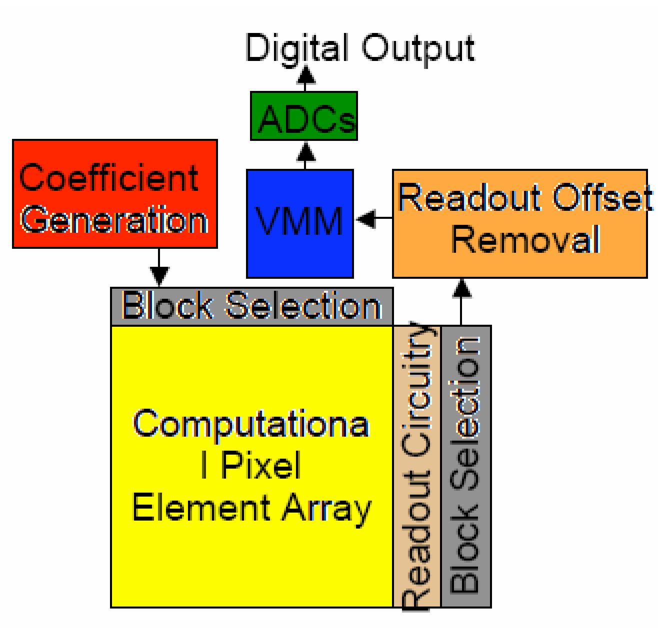
true color low-light imaging

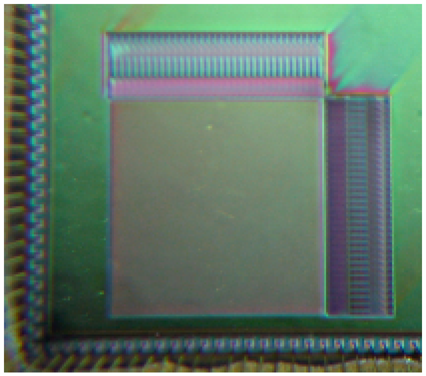
256 x 256 image with 10:1 compression

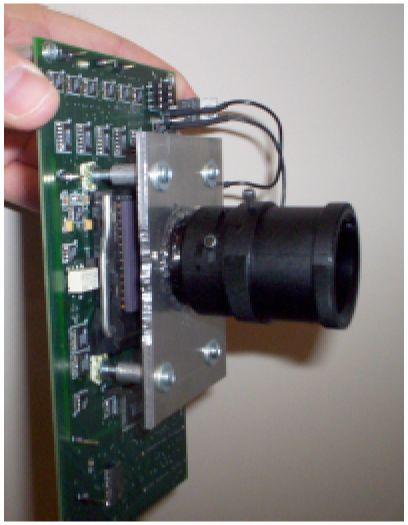
[Nature Photonics, April 2007]

Georgia Tech Analog Imager

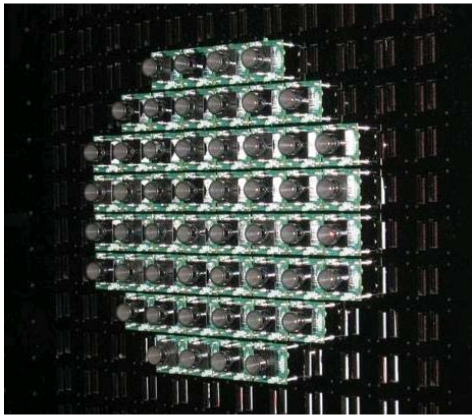
- Bottleneck in imager arrays is **data readout**
- Instead of quantizing pixel values, take CS inner products *in analog*
- Potential for tremendous (factor of 10000) power savings



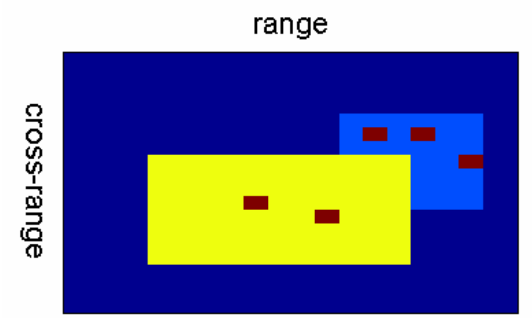




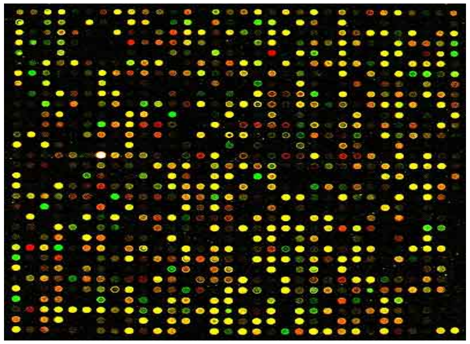
Other Imaging Apps



- Compressive **camera arrays**
  - sampling rate scales logarithmically in both number of pixels and number of cameras

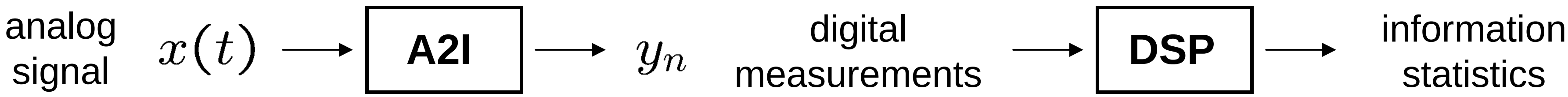


- Compressive **radar** and **sonar**
  - greatly simplified receiver
  - exploring radar/sonar networks



- Compressive **DNA microarrays**
  - smaller, more agile arrays for bio-sensing
  - exploit sparsity in presentation of organisms to array [O. Milenkovic et al]

Analog-to-*Information* Conversion



- Analog-to-information (A2I) converter takes **analog input signal** and creates **discrete (digital) measurements**

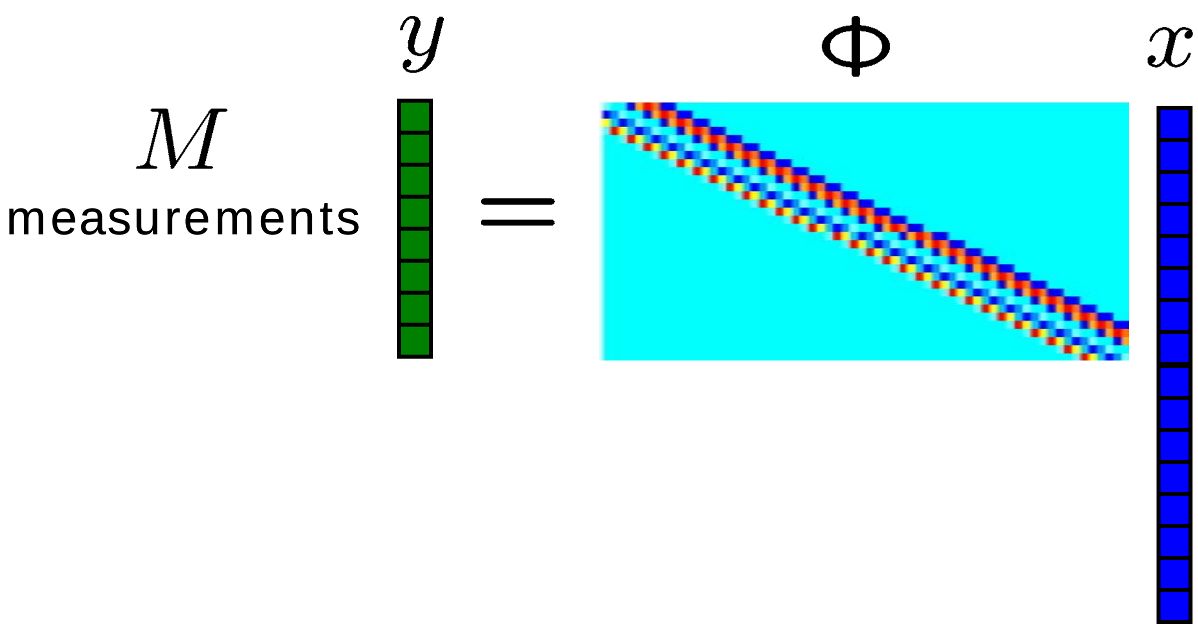
- Extension of analog-to-digital converter that samples at the signal's "information rate" rather than its Nyquist rate

Streaming Measurements

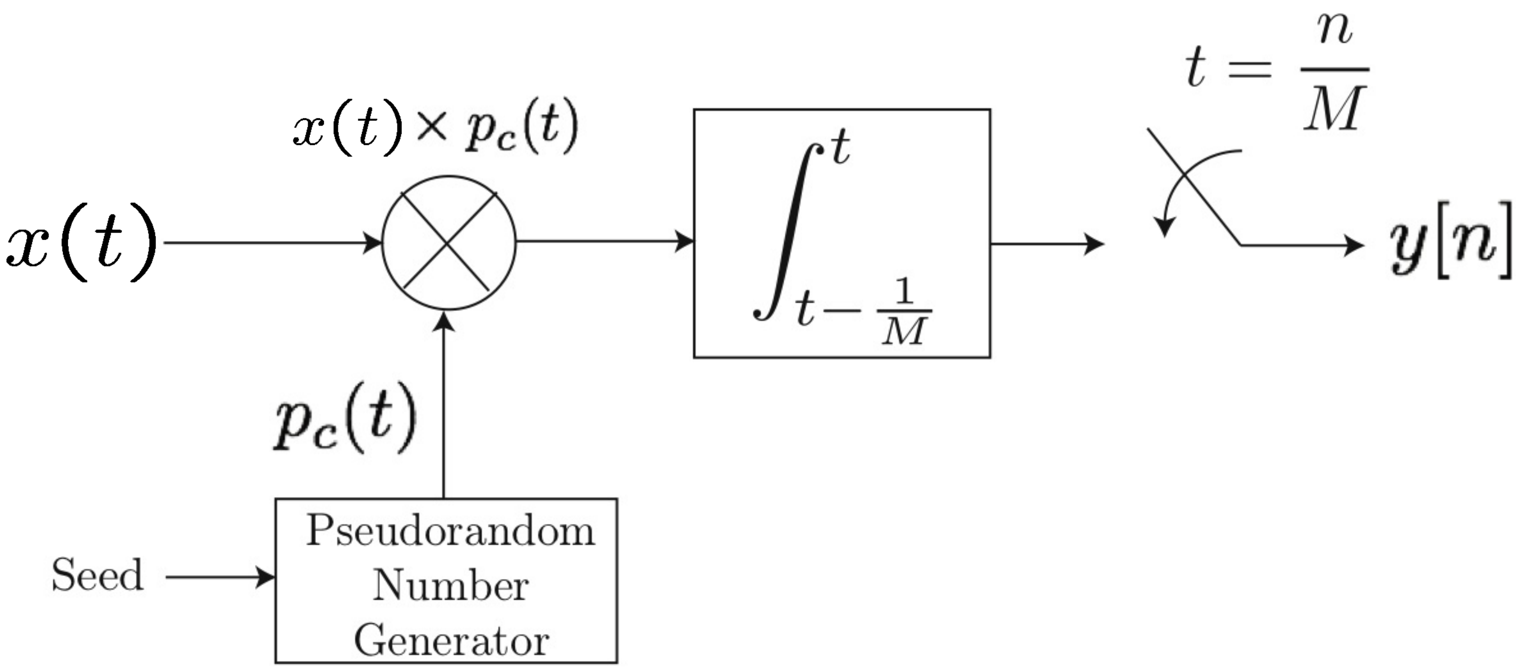
- Streaming applications: cannot fit entire signal into a processing buffer at one time

$$y = \Phi x$$

streaming requires special Φ

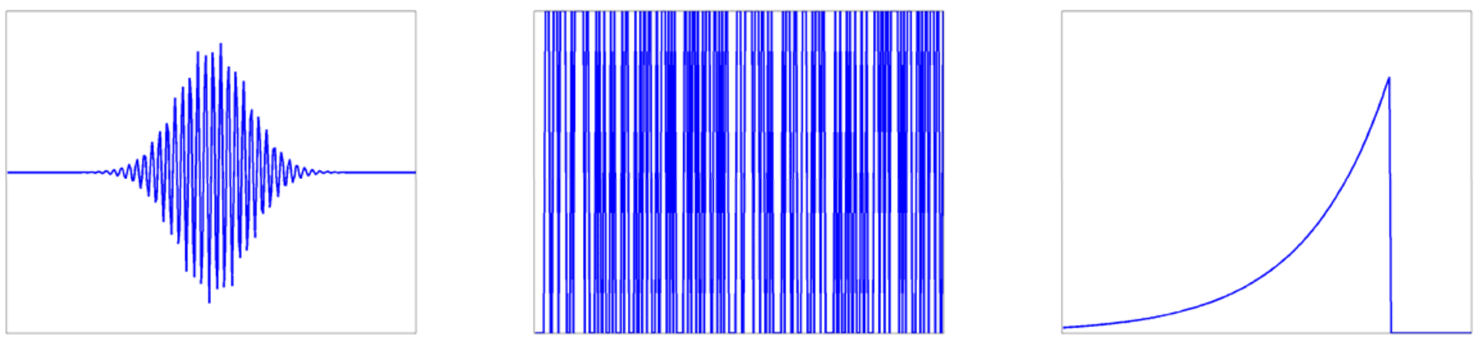


Random Demodulator



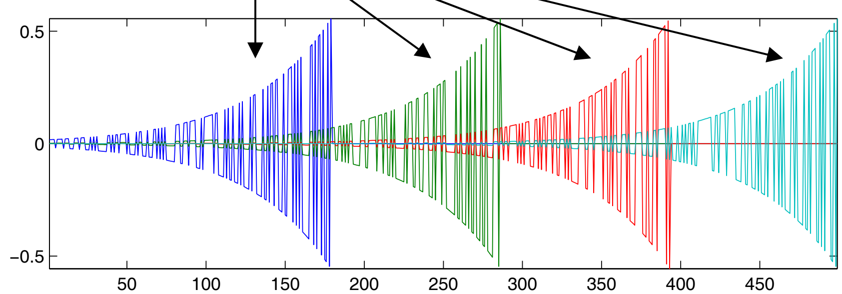
Random Demodulator

$$y[m] = ((x \times p) * h)(t)|_{t=mT}$$

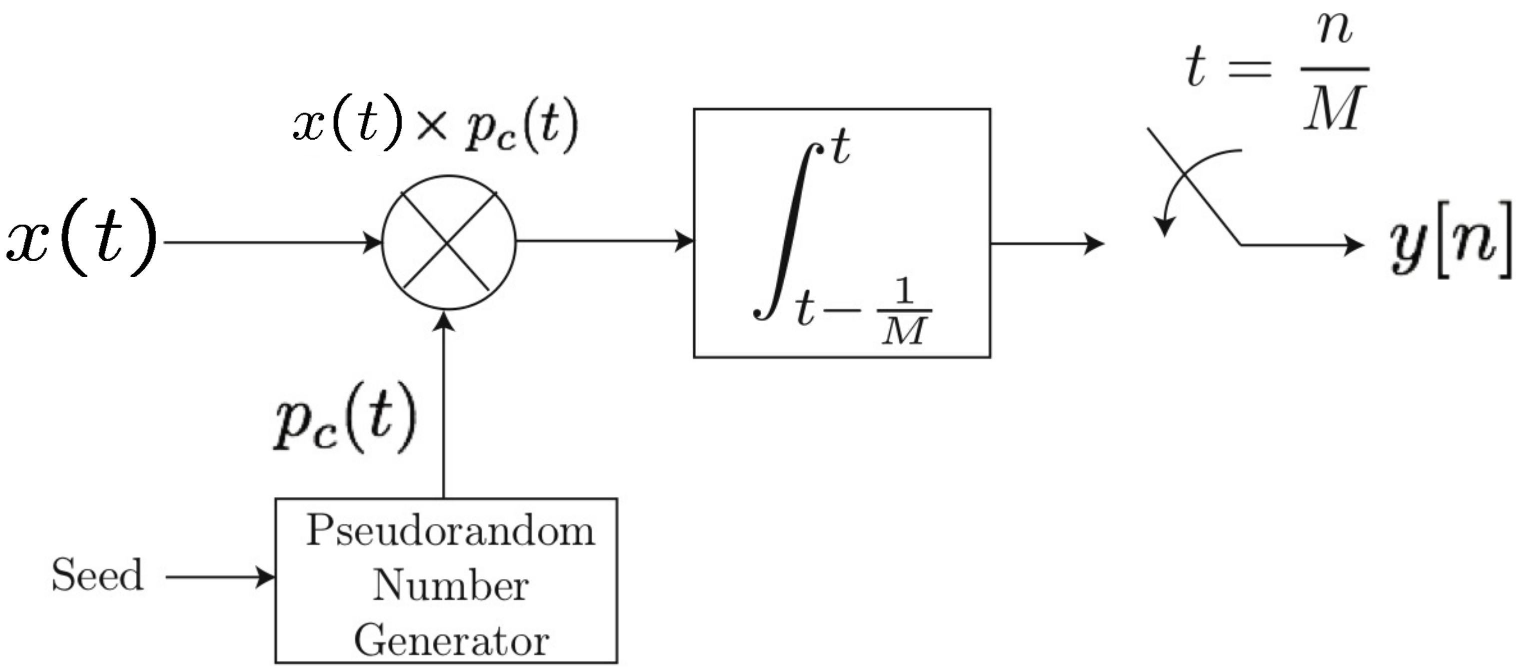


$$h(t) = e^{-at}, \quad t \geq 0$$

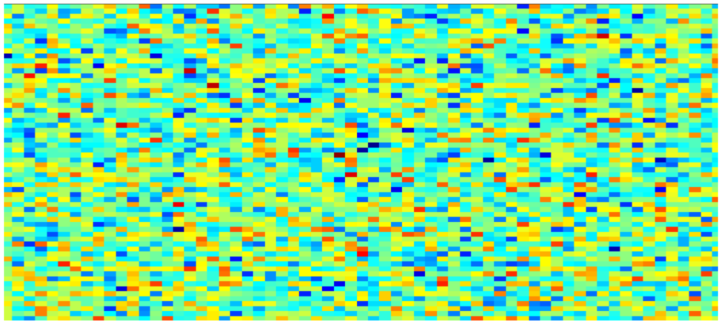
$$= \langle x, \phi_m \rangle$$



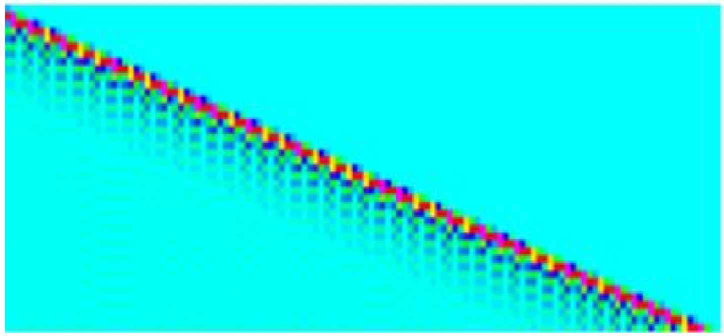
Random Demodulator



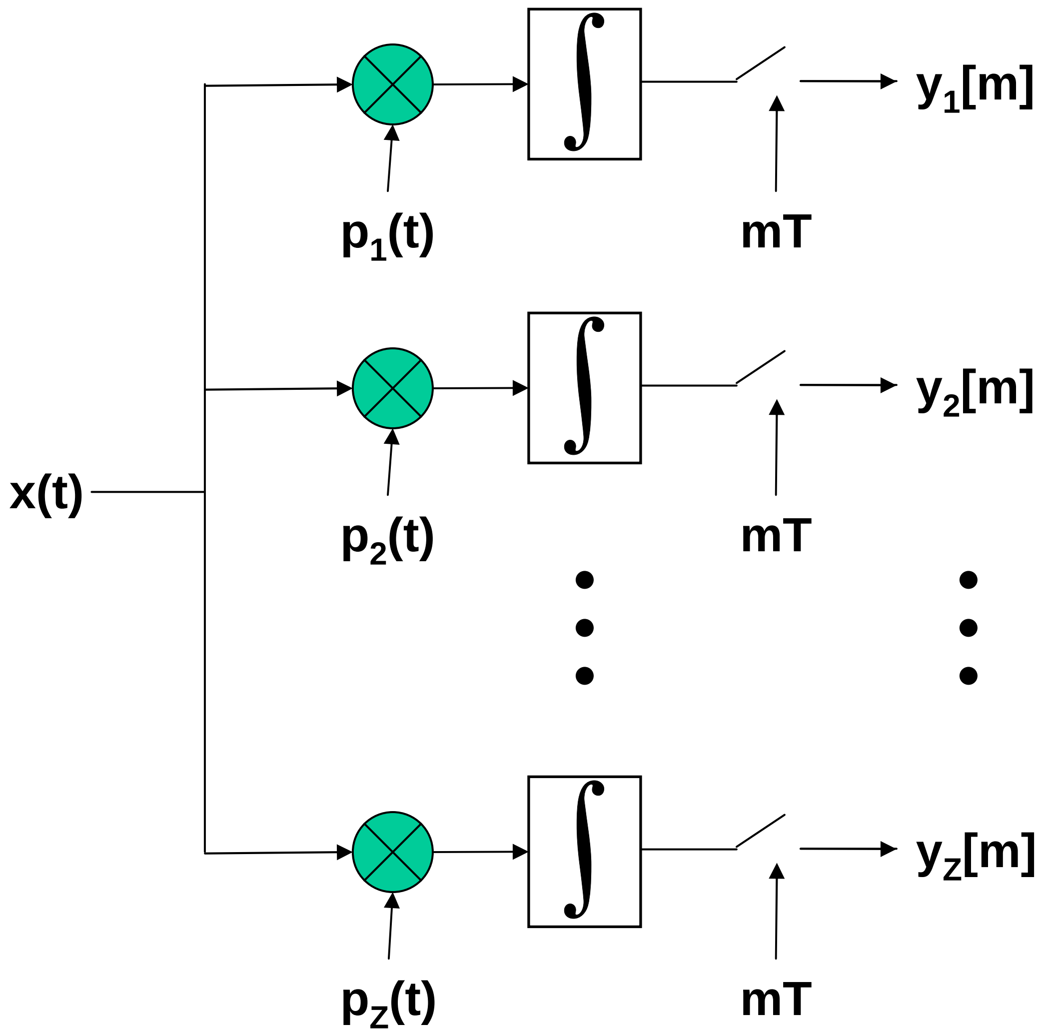
Φ



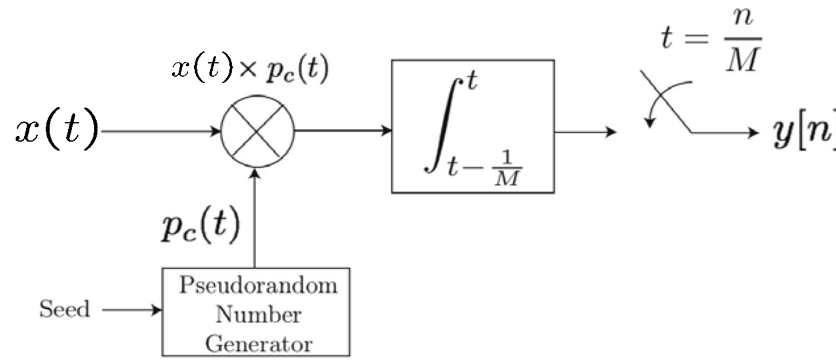
Φ



Parallel Architecture



Random Demodulator



- **Theorem** [Tropp et al 2007]

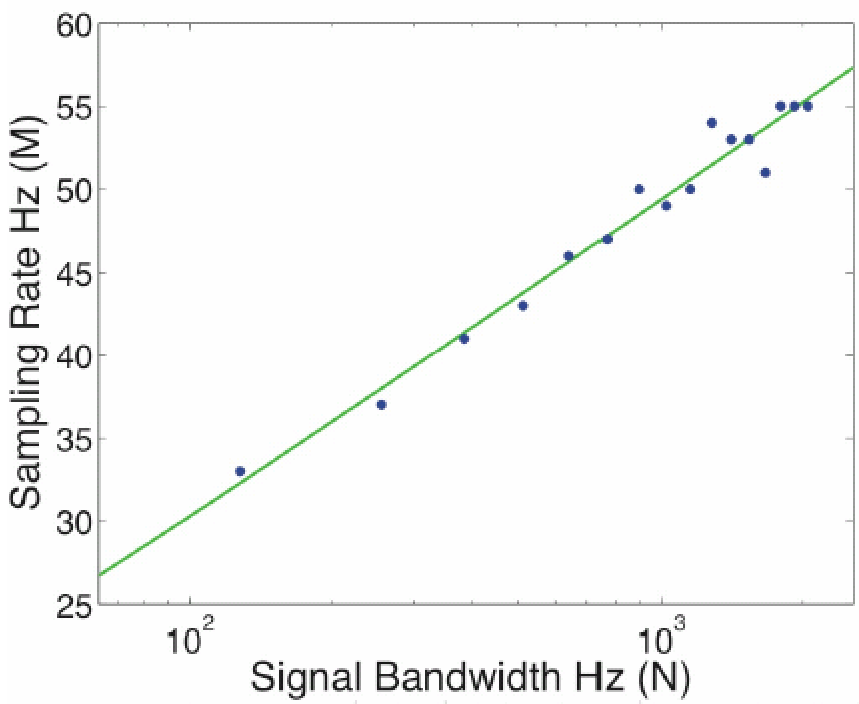
If the sampling rate satisfies

$$M > cK \log^2(N/\delta), \quad 0 < \delta < 1$$

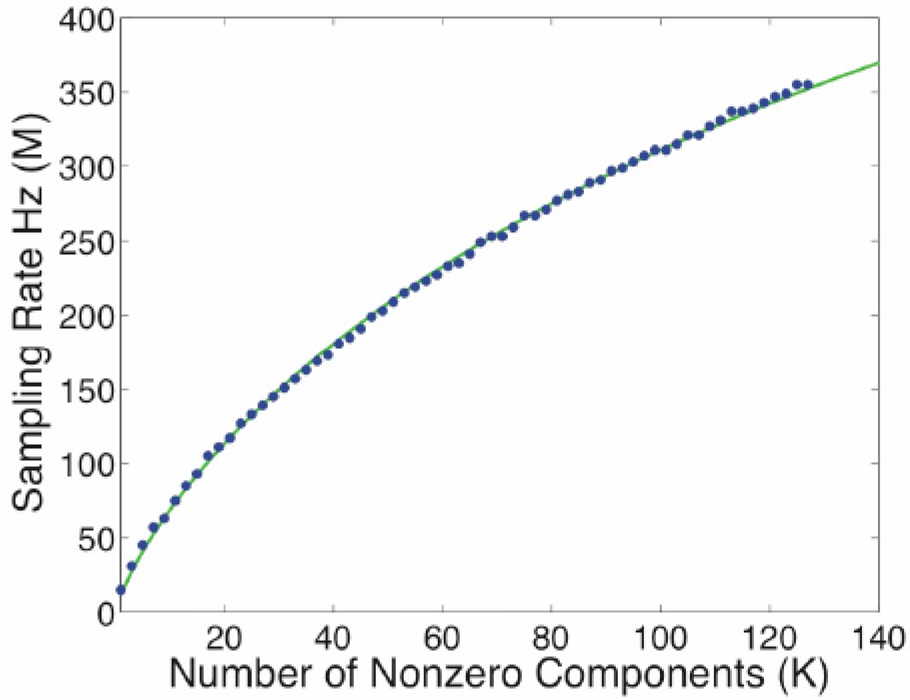
then locally Fourier K-sparse signals can be recovered exactly with probability

$$1 - \delta$$

Empirical Results



$$1.69K\log(N/K+1)+4.51$$



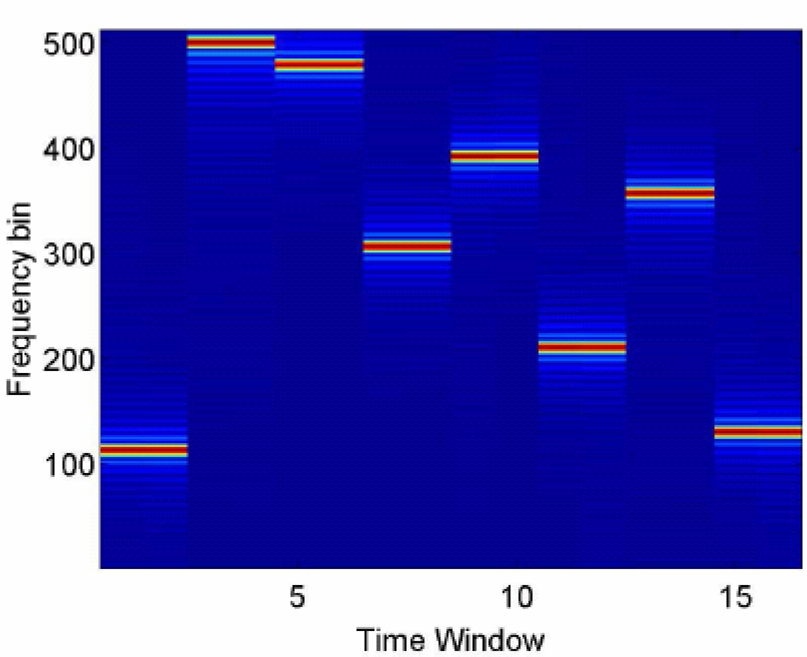
$$1.71K\log(N/K+1)+1$$

$$M \le CK\log(N/K+1)$$
$$C \sim 1.7$$

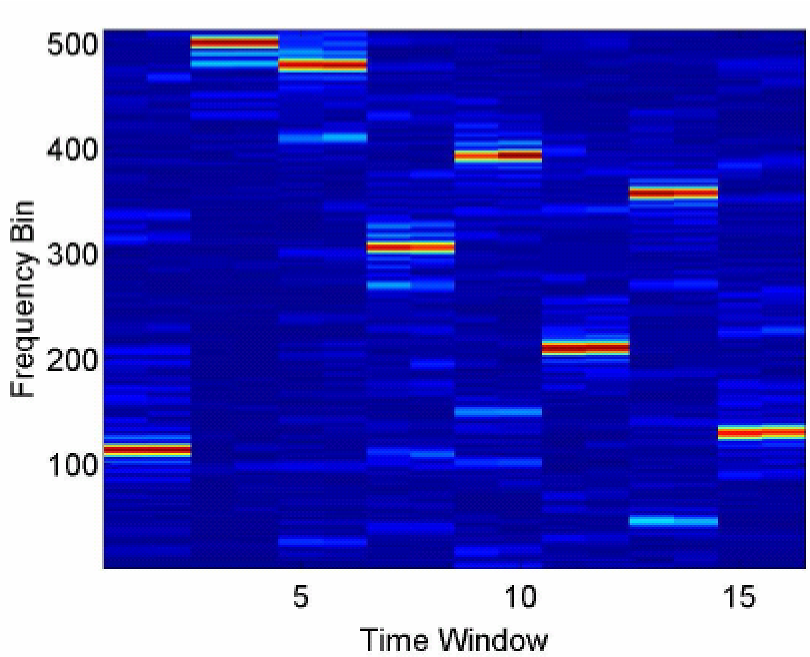
Example: Frequency Hopper

- Random demodulator AIC at 8x sub-Nyquist

spectrogram



sparsogram



References – Data Acquisition (1)

CS Camera:

- Dharmpal Takhar, Jason Laska, Michael Wakin, Marco Duarte, Dror Baron, Shriram Sarvotham, Kevin Kelly, and Richard Baraniuk, A new compressive imaging camera architecture using optical-domain compression. (Proc. of Computational Imaging IV at SPIE Electronic Imaging, San Jose, California, January 2006)
- Duncan Graham-Rowe, Digital cameras: Pixel power, Nature Photonics 1, 211 - 212 (2007).
- CS Camera Website: http://www.dsp.ece.rice.edu/cs/cscamera/

Analog-to-Information Conversion:

- Joel Tropp, Michael Wakin, Marco Duarte, Dror Baron, and Richard Baraniuk, Random filters for compressive sampling and reconstruction. (Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Toulouse, France, May 2006)
- Jason Laska, Sami Kirolos, Marco Duarte, Tamer Ragheb, Richard Baraniuk, and Yehia Massoud, Theory and implementation of an analog-to-information converter using random demodulation. (Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS), New Orleans, Louisiana, 2007)

References – Data Acquisition (2)

Analog-to-Information Conversion [cont.]:

- Tamer Ragheb, Sami Kirolos, Jason Laska, Anna Gilbert, Martin Strauss, Richard Baraniuk, and Yehia Massoud, Implementation models for analog-to-information conversion via random sampling. (To appear in Proc. Midwest Symposium on Circuits and Systems (MWSCAS), 2007)

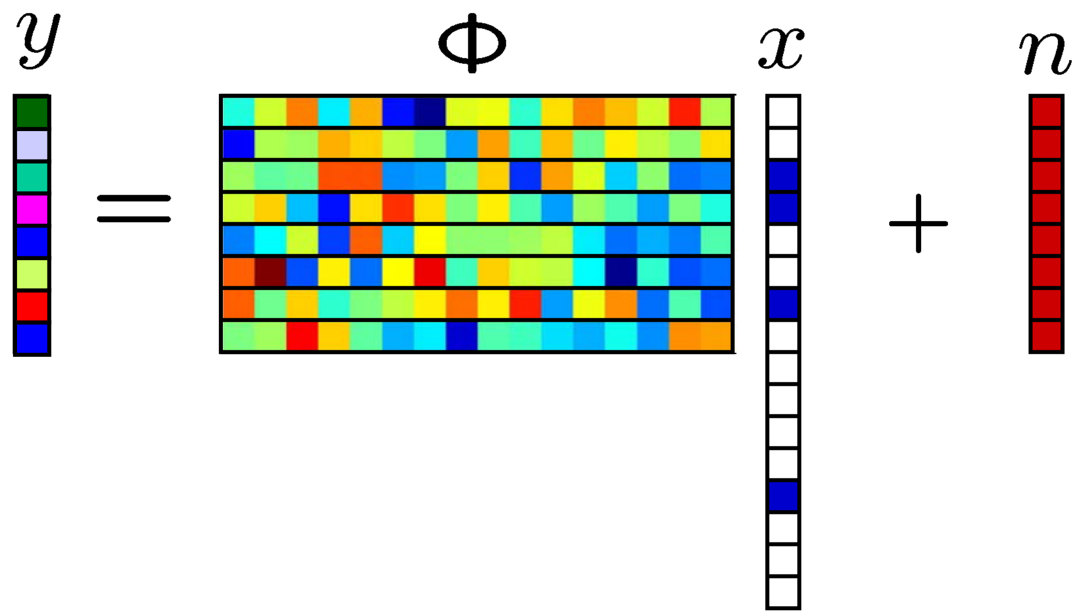
Analysis versus Synthesis in L_1 minimization:

- J.-L. Starck, M. Elad, and D. L. Donoho, “Redundant multiscale transforms and their application for morphological component analysis,” Adv. Imaging and Electron Phys., vol. 132, 2004.
- M. Elad, P. Milanfar, and R. Rubinstein, “Analysis versus synthesis in signal priors,” Inverse Problems, vol. 23, pp. 947–968, 2007.

Statistical Estimation

Idea: Model selection when #variables >> #observations

- sparse model provides simple explanation



$$\min \|x'\|_1 \text{ subject to } \|\Phi^*(y - \Phi x')\|_\infty \leq \epsilon$$

Dantzig selector [Candes and Tao]

References – Statistical Estimation

Dantzig Selector:

- Emmanuel Candès and Terence Tao, The Dantzig Selector: Statistical estimation when p is much larger than n (To appear in Annals of Statistics)

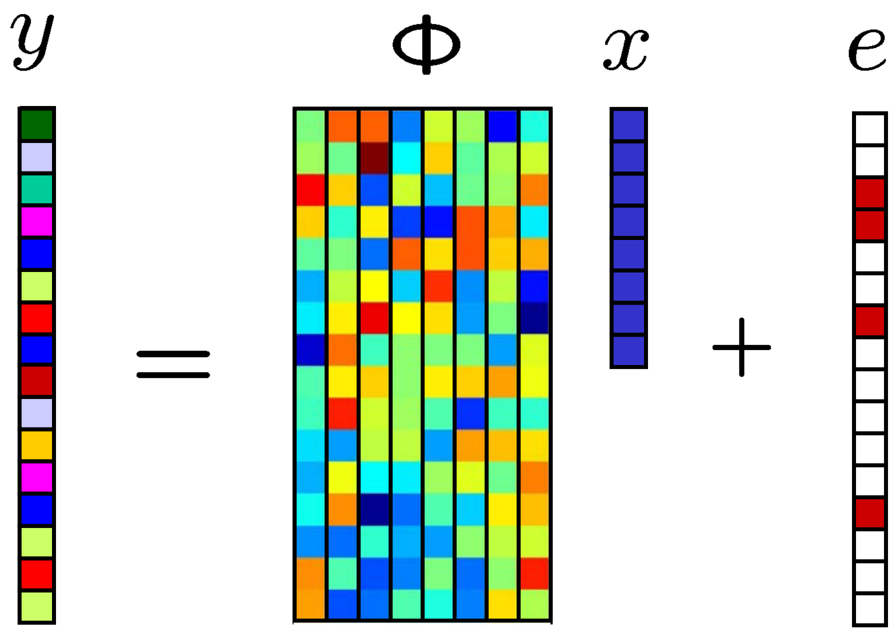
Phase Transition:

- David Donoho and Victoria Stodden, Breakdown Point of Model Selection When the Number of Variables Exceeds the Number of Observations, International Joint Conference on Neural Networks, 2006.

Error Correction

Idea: Channel coding using CS principles

- unconstrained minimization problem
- robust to some large and many small errors



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

References – Error Correction

Error Correction

- Emmanuel Candès and Terence Tao, Decoding by linear programming. (IEEE Trans. on Information Theory, 51(12), pp. 4203 - 4215, December 2005)
- Mark Rudelson and Roman Vershynin, Geometric approach to error correcting codes and reconstruction of signals. (International Mathematical Research Notices, 64, pp. 4019 - 4041, 2005)
- Emmanuel Candès and Paige Randall, Highly robust error correction by convex programming. (Preprint, 2006)
- Rick Chartrand, Nonconvex compressed sensing and error correction. (Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Honolulu, Hawaii, April 2007)
- Cynthia Dwork, Frank McSherry, and Kunal Talwar, The price of privacy and the limits of LP decoding. (Proc. Symposium on Theory of Computing (STOC), San Diego, California, June, 2007)

Additional References

Related areas:

- Martin Vetterli, Pina Marziliano, and Thierry Blu, Sampling signals with finite rate of innovation. (IEEE Trans. on Signal Processing, 50(6), pp. 1417-1428, June 2002)
- Anna Gilbert, Sudipto Guha, Piotr Indyk, S. Muthukrishnan, and Martin Strauss, Near-optimal sparse Fourier representations via sampling. (Proc. ACM Symposium on Theory of Computing (STOC), 2002)

Other CS applications:

- David Donoho and Yaakov Tsaig, Extensions of compressed sensing. (Signal Processing, 86(3), pp. 533-548, March 2006)
- Mona Sheikh, Olgica Milenkovic, and Richard Baraniuk, Compressed sensing DNA microarrays. (Rice ECE Department Technical Report TREE 0706, May 2007)

More at: **dsp.rice.edu/cs**

Summary

Summary

- Compressive Sensing (CS)
  - integrates sensing, compression, processing
  - exploits signal sparsity/compressibility information and new uncertainty principles

- Why CS works: stable embedding for signals with concise geometric structure
  - sparse signals (*K*-planes), compressible signals (Lp balls)
  - smooth manifolds

- Information scalability
  - detection< classification < estimation < reconstruction
  - compressive measurements ~ sufficient statistics

- Applications
  - enables new sensing architectures and modalities
  - most useful when measurements are expensive