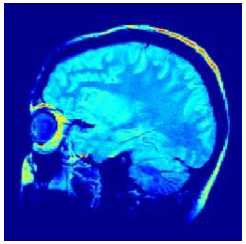
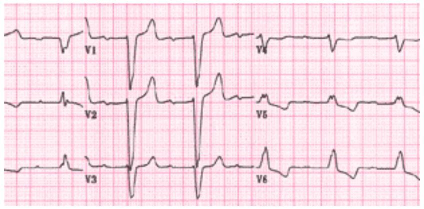
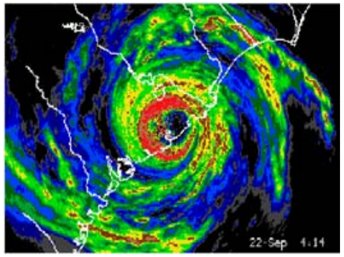
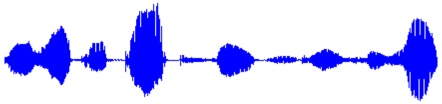
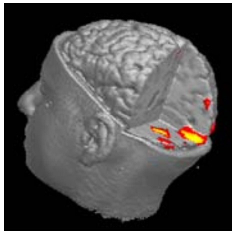
Compressive Sensing

and Applications





Volkan Cevher
volkan.cevher@epfl.ch

Justin Romberg
jrom@ece.gatech.edu

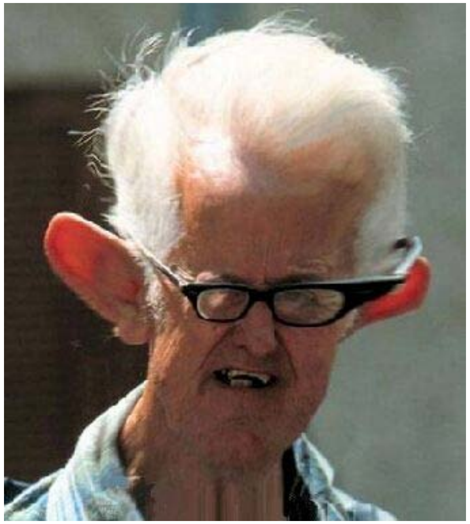
Acknowledgements

RICE UNIVERSITY DSP

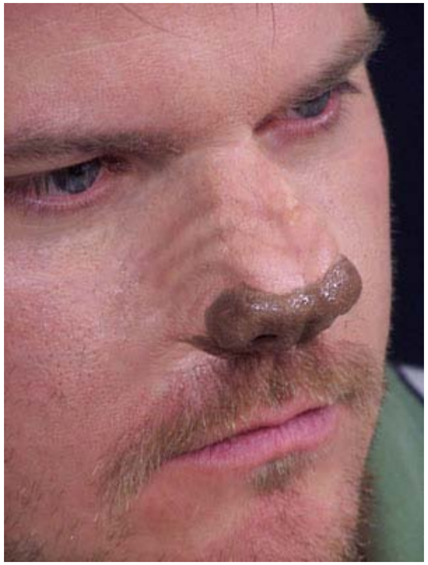
- Rice DSP Group (Slides)
  - Richard Baraniuk
    - Mark Davenport,
    - Marco Duarte,
    - Chinmay Hegde,
    - Jason Laska,
    - Shri Sarvotham,
    - Mona Sheikh
    - Stephen Schnelle...
  - Mike Wakin, Petros Boufounos, Dror Baron

Outline

- Introduction to Compressive Sensing (CS)
  - motivation
  - basic concepts
- CS Theoretical Foundation
  - geometry of sparse and compressible signals
  - coded acquisition
  - restricted isometry property (RIP)
  - structured matrices and random convolution
  - signal recovery algorithms
  - structured sparsity
- CS in Action
- Summary



Sensing



Digital Revolution

12MP

25fps/1080p

4KHz

Multi touch

Digital Revolution

1977 – 5hours

12MP

25fps/1080p

4KHz

<30mins

Major Trends

higher resolution / denser sampling

12MP

25fps/1080p

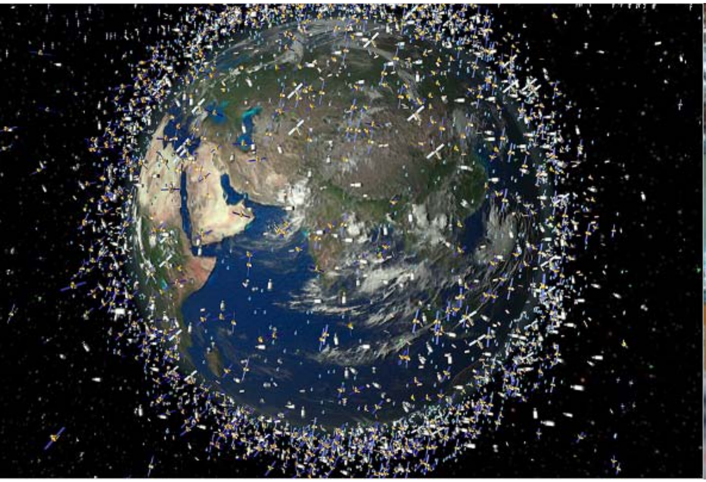
4KHz

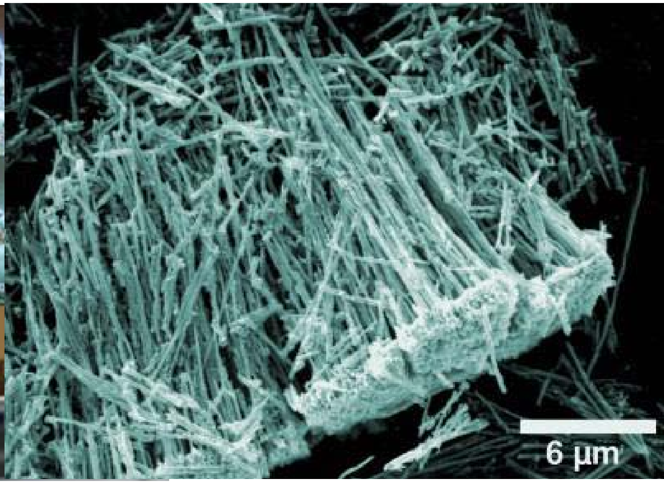
160MP

200,000fps

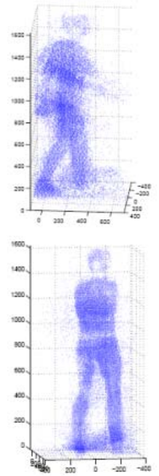
192,000Hz

Major Trends





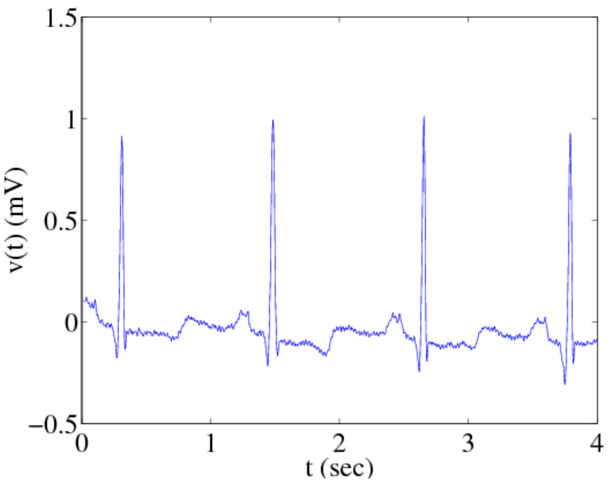
large numbers of sensors

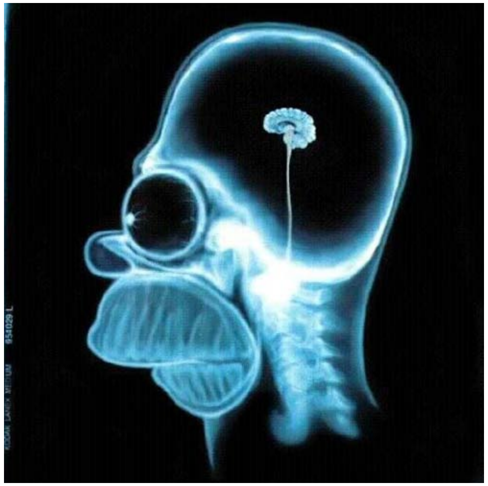


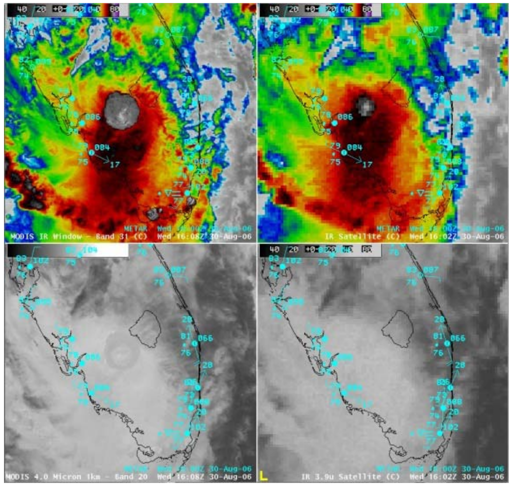
Major Trends

higher resolution / denser sampling

large numbers of sensors

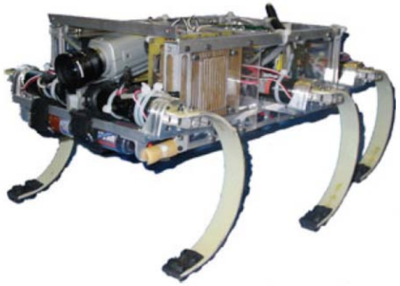






increasing # of modalities / mobility

acoustic, RF, visual, IR,
UV, x-ray, gamma ray, ...



Major Trends in Sensing

higher resolution / denser sampling

X

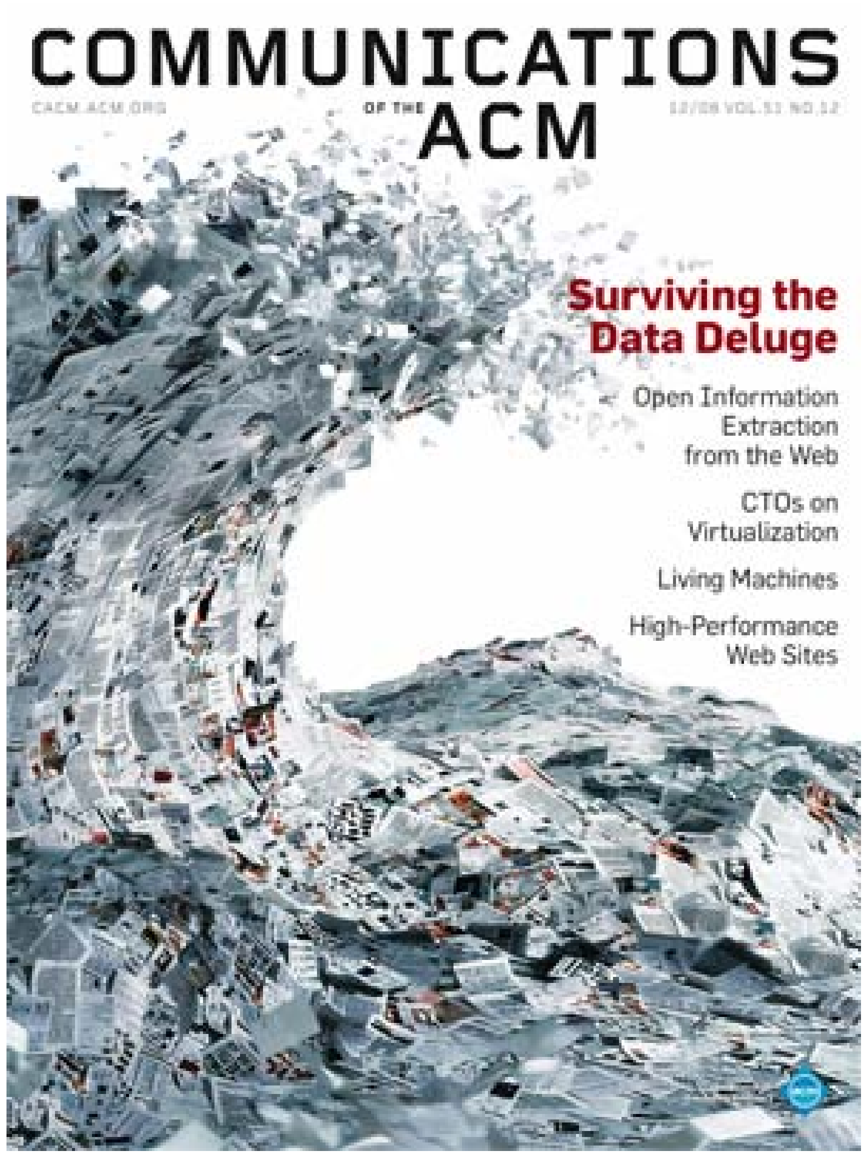
large numbers of sensors

X

increasing # of modalities / mobility

= ?

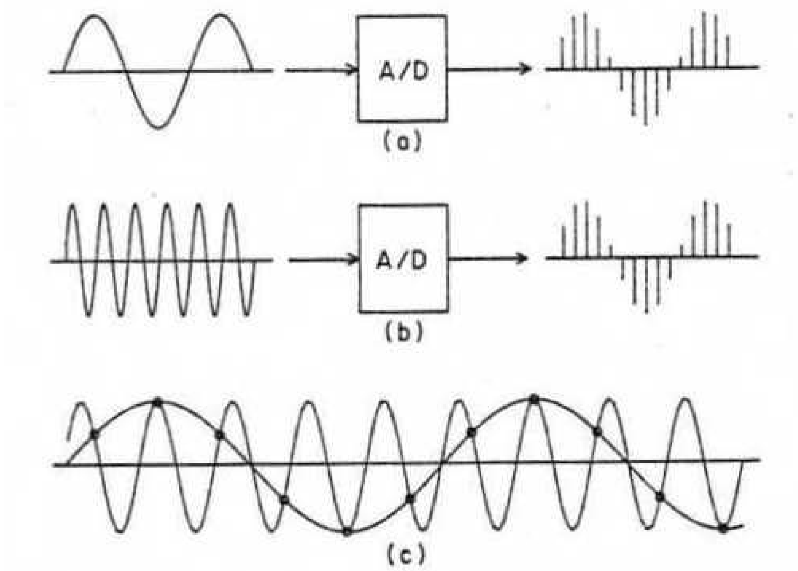
Major Trends in Sensing



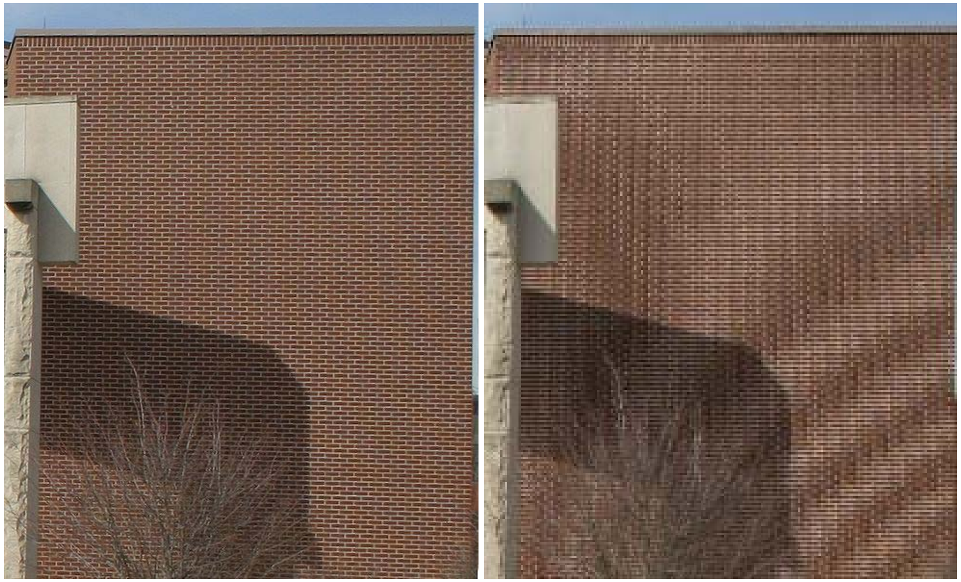
Digital Data Acquisition

Foundation: *Shannon/Nyquist sampling theorem*

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



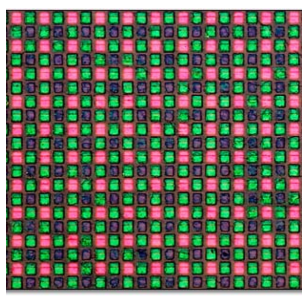
time

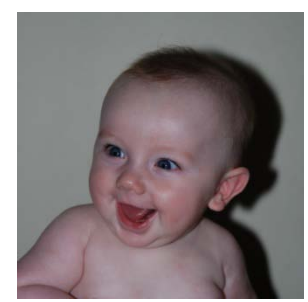


space

Sensing by *Sampling*

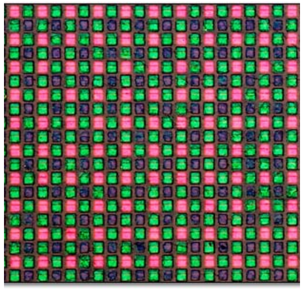
- Long-established paradigm for digital data acquisition
  - uniformly ***sample*** data at Nyquist rate (2x Fourier bandwidth)





Sensing by *Sampling*

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



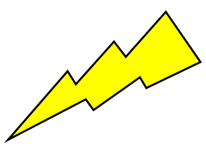
Sensing by *Sampling*

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

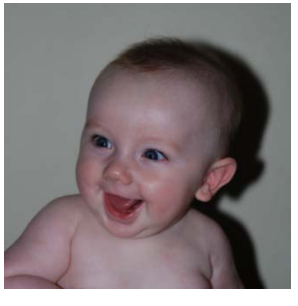
$N \gg K$

x → sample → N → **compress** → K → transmit/store

JPEG
JPEG2000
...

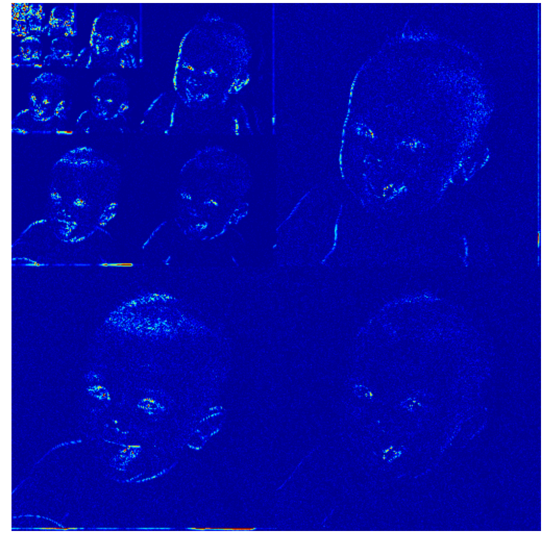


receive → K → decompress → N → $\widehat{x}$



Sparsity / Compressibility

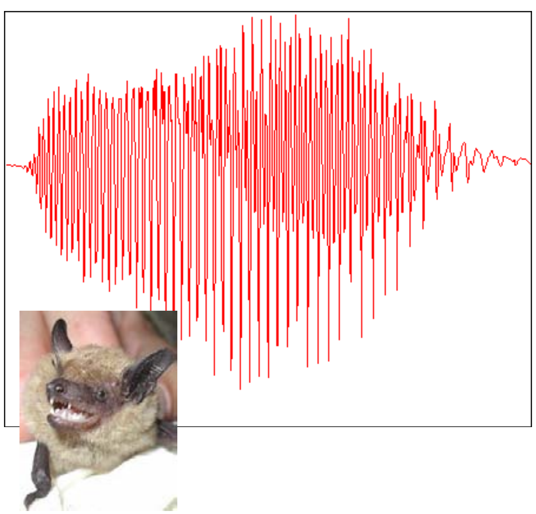
N
pixels

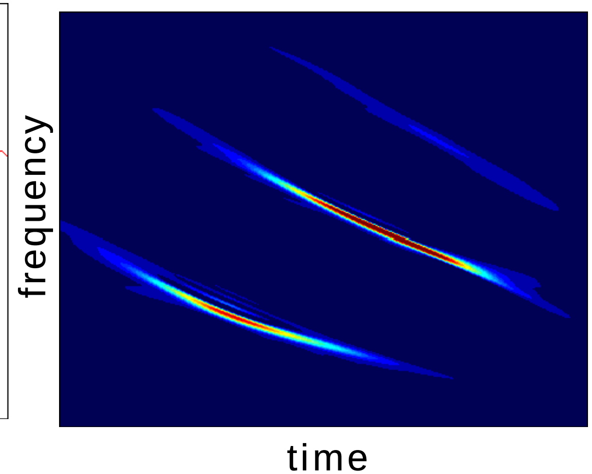


$K \ll N$
large
wavelet
coefficients

(blue = 0)

N
wideband
signal
samples

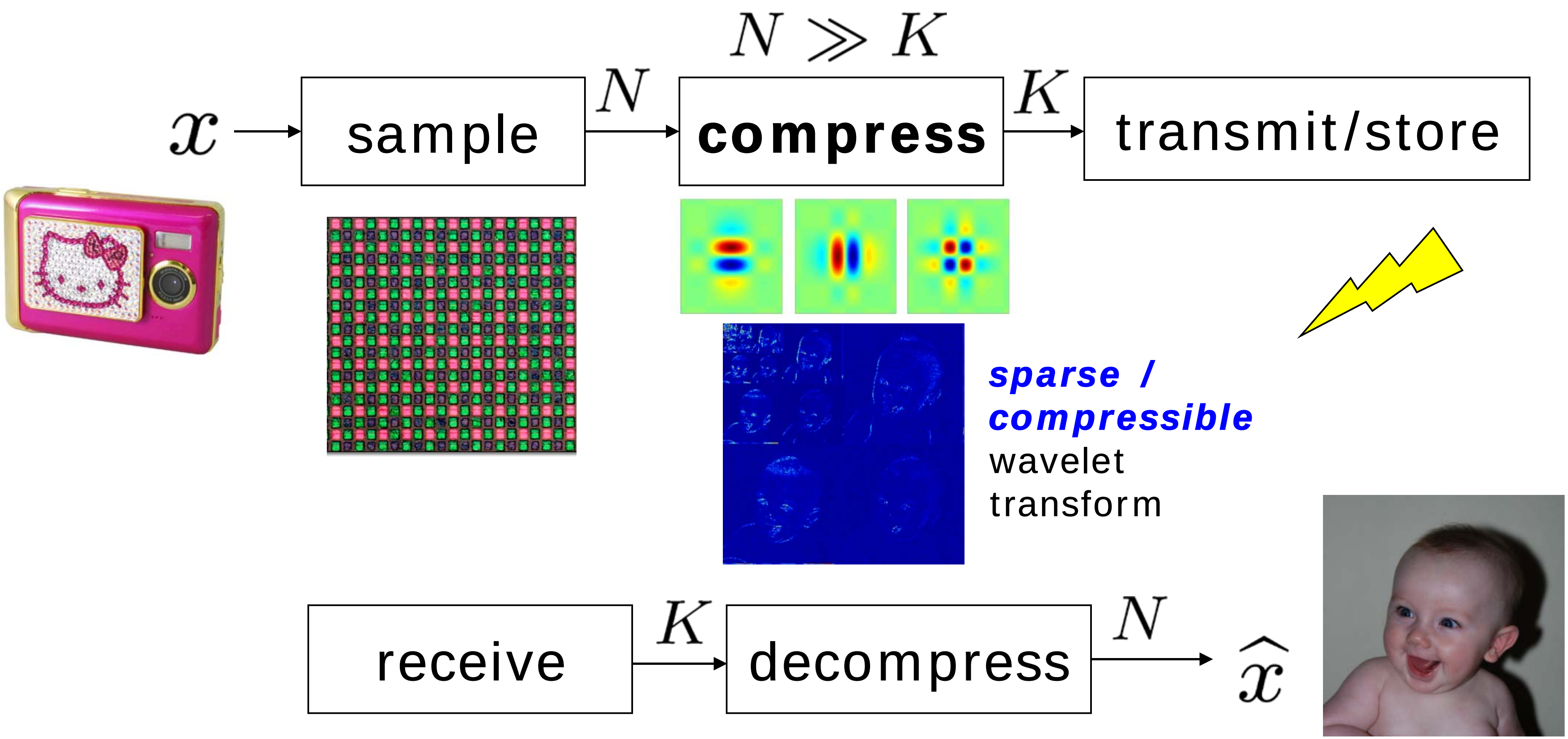




$K \ll N$
large
Gabor (TF)
coefficients

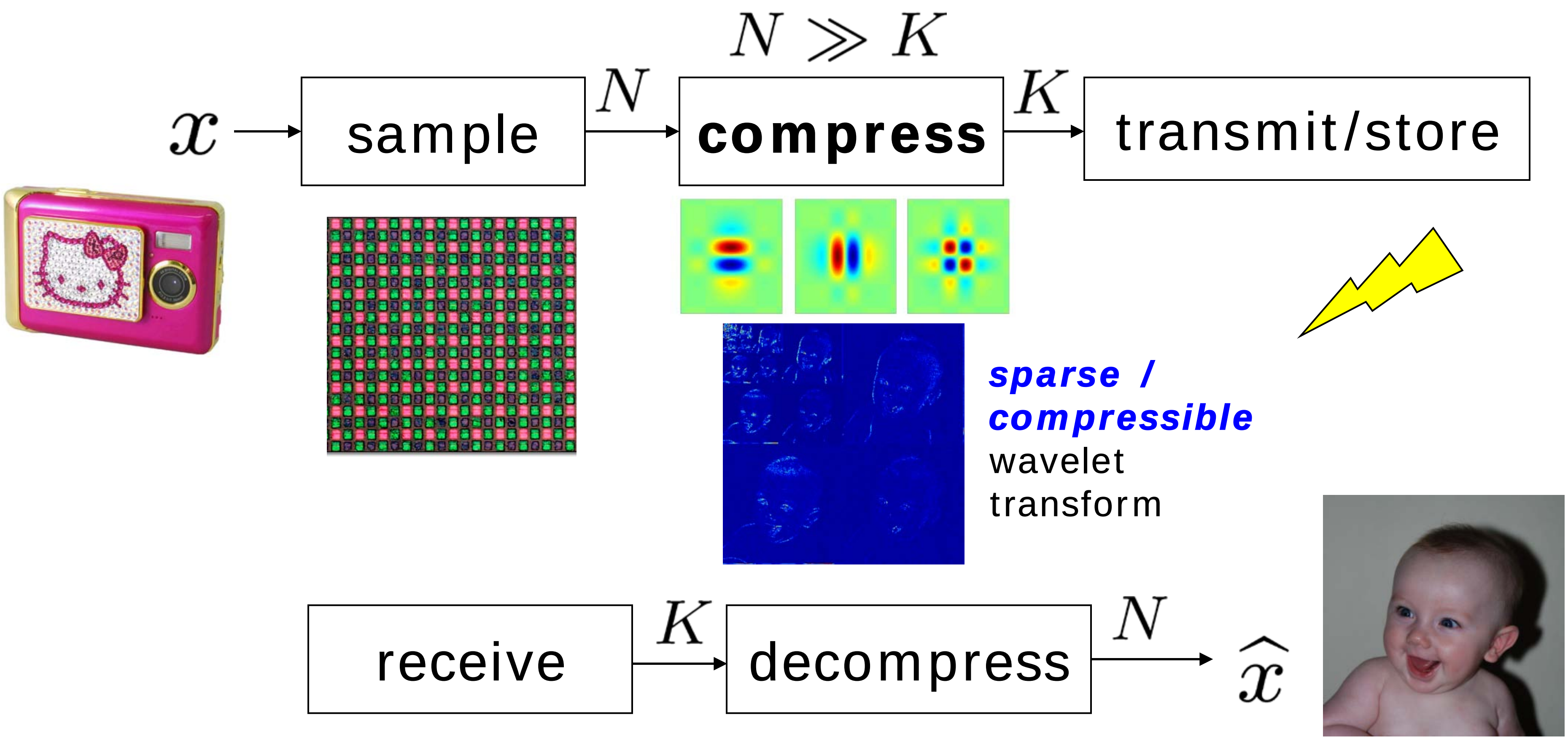
Sample / Compress

- Long-established paradigm for digital data acquisition
  - uniformly ***sample*** data at Nyquist rate
  - ***compress*** data

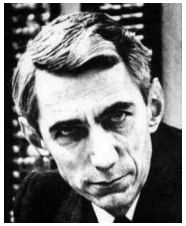


What's Wrong with this Picture?

- ***Why go to all the work to acquire* N *samples only to discard all but* K *pieces of data?***



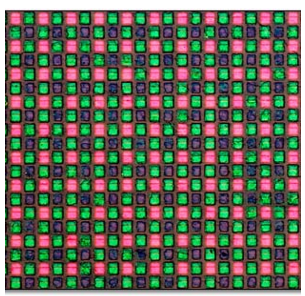
What's Wrong with this Picture?

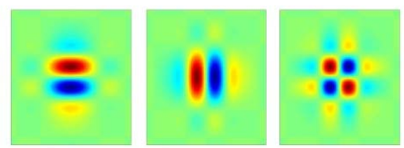


linear processing
linear signal model
(bandlimited subspace)

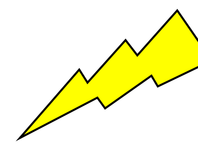
nonlinear processing
nonlinear signal model
(union of subspaces)

x → sample →N→ **compress** →K→ transmit/store

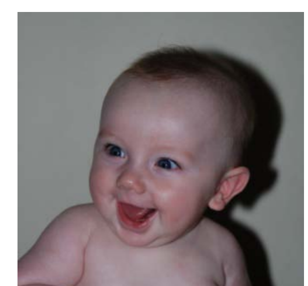




sparse / compressible
wavelet
transform



receive →K→ decompress →N→ $\widehat{x}$

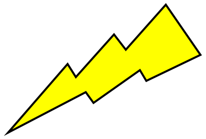


Compressive Sensing

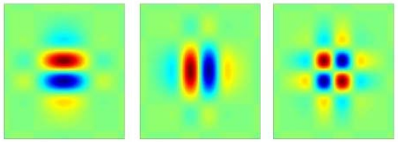
- Directly acquire "***compressed***" data
- Replace samples by more general "measurements"

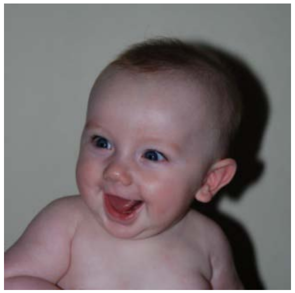
$$K \approx \underline{M} \ll N$$

$x \rightarrow$ **compressive sensing** $\xrightarrow[y]{M}$ transmit/store



receive $\xrightarrow{M}$ reconstruct $\xrightarrow{N}$ $\widehat{x}$



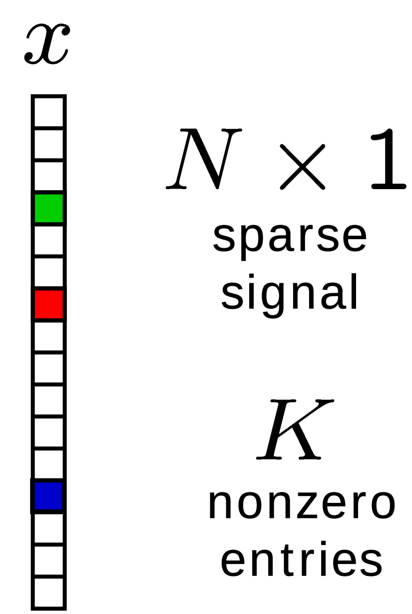


Compressive Sensing

Theory I
Geometrical Perspective

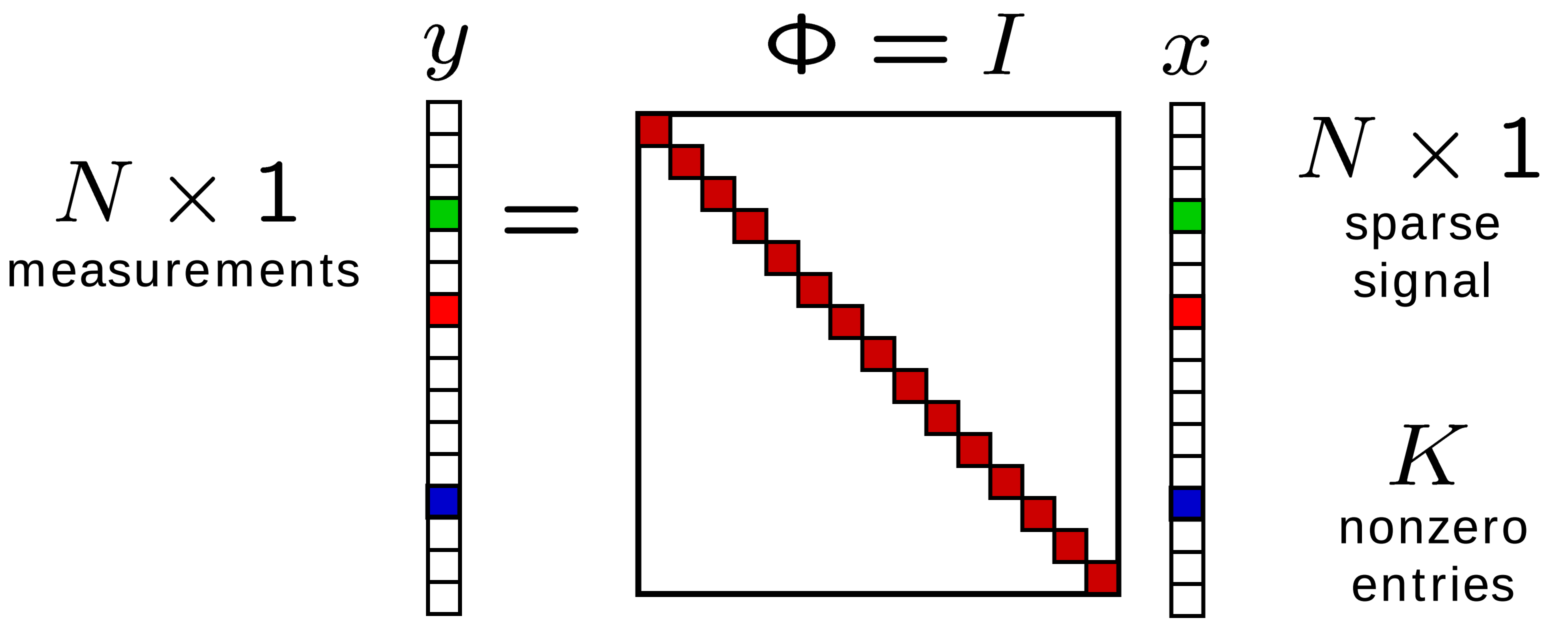
Sampling

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



Sampling

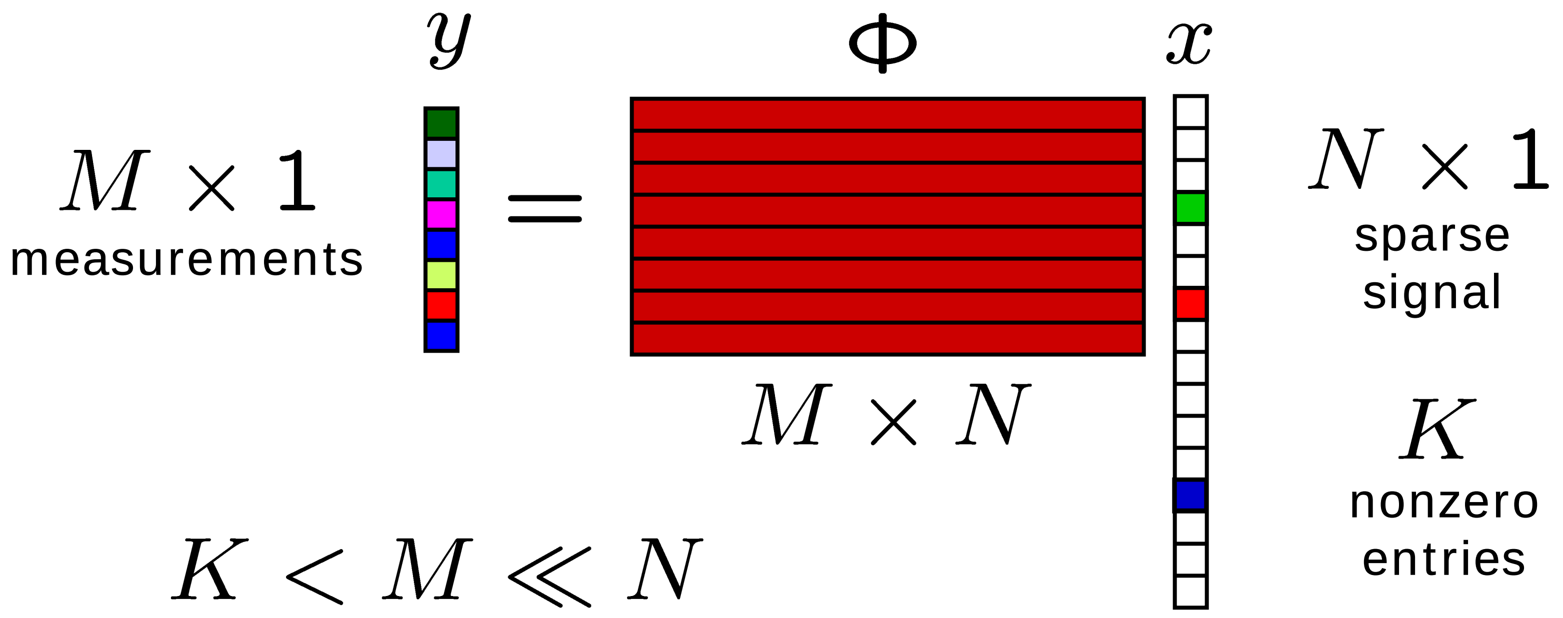
- Signal x is K-*sparse* in basis/dictionary Ψ
  - WLOG assume sparse in space domain $\Psi = I$
- **Samples**



Compressive Sampling

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

$$y = \Phi x$$



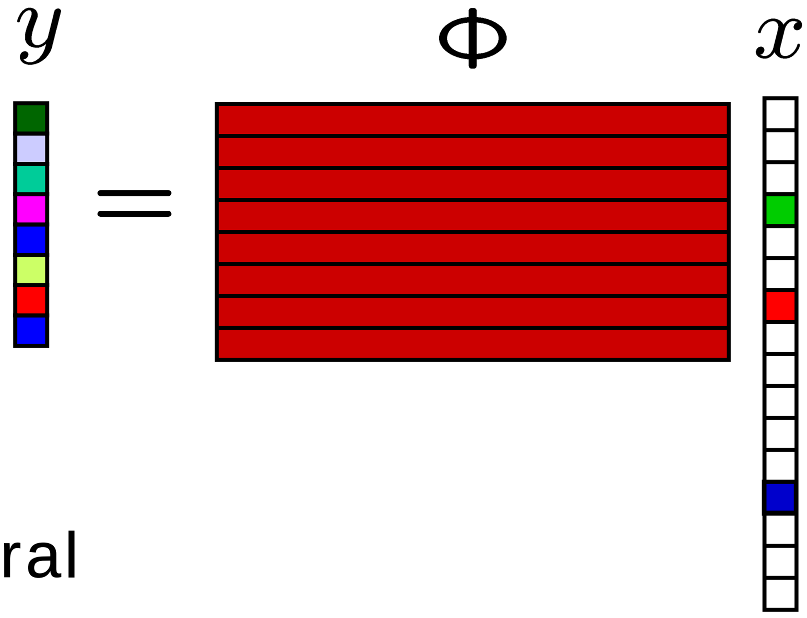
$$K < M \ll N$$

How Can It Work?

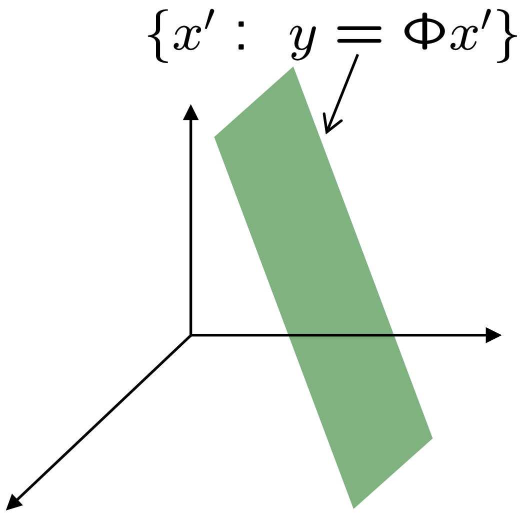
- Projection Φ
  not full rank...

  $M < N$

  ... and so
  loses information in general



- Ex: Infinitely many x's
  map to the same y

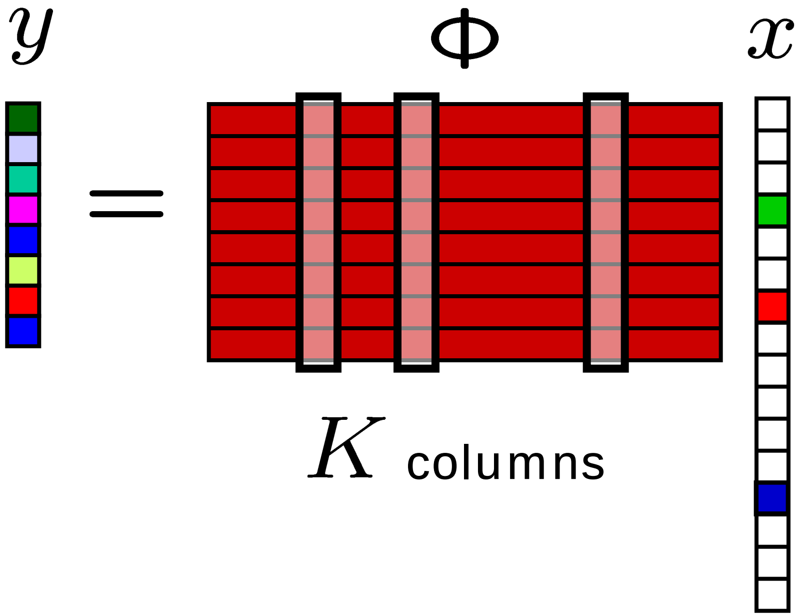


How Can It Work?

- Projection Φ
  not full rank...

  $M < N$

  ... and so
  loses information in general



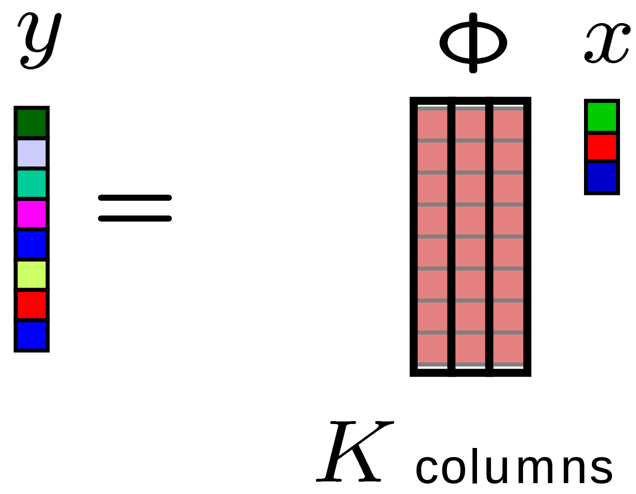
- But we are only interested in ***sparse*** vectors x

How Can It Work?

- Projection Φ
  not full rank...

  $M < N$

  ... and so
  loses information in general



- But we are only interested in ***sparse*** vectors

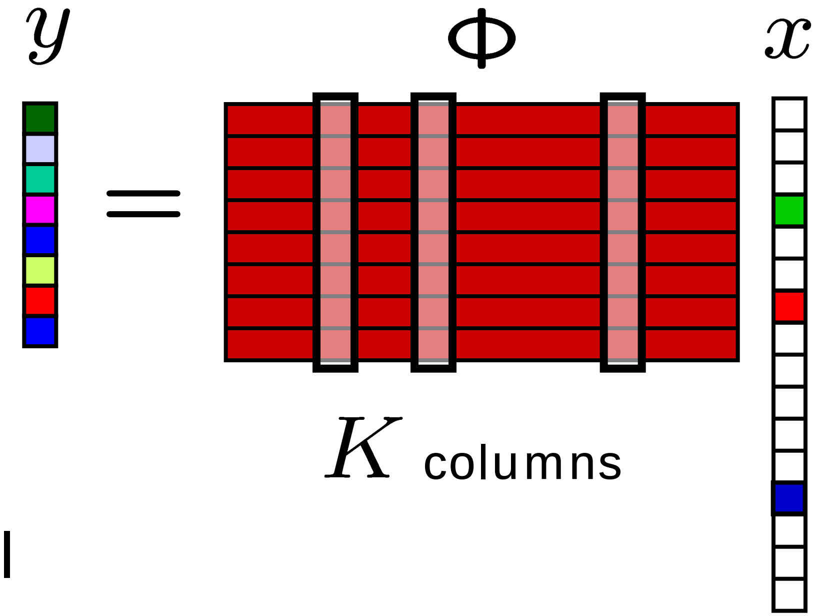
- Φ is effectively *M*x*K*

How Can It Work?

- Projection Φ
  not full rank...

  $M < N$

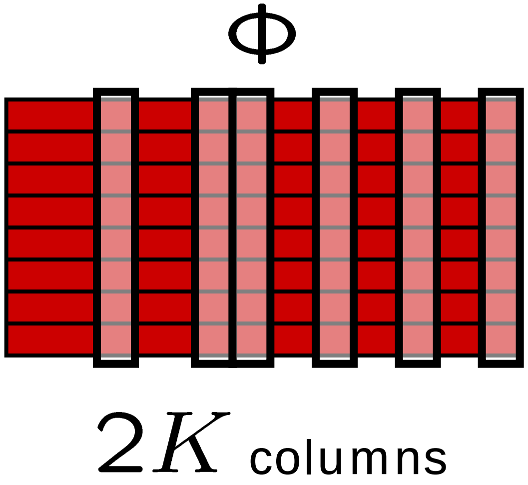
  ... and so
  loses information in general



- But we are only interested in ***sparse*** vectors

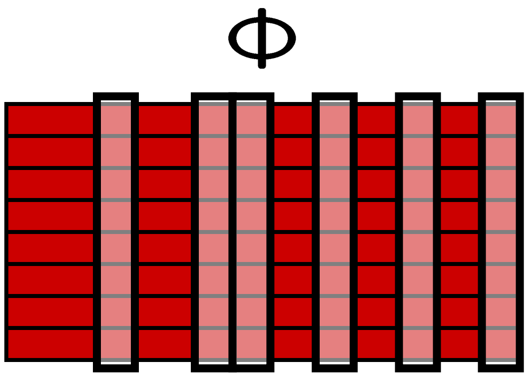
- **Design** Φ so that each of its *M*x*K* submatrices are full rank

How Can It Work?



- **Goal:** Design Φ so that its *M*x2*K* submatrices are full rank
  - difference $x_1 - x_2$ between two *K*-sparse vectors is 2*K* sparse in general
  - preserve information in *K*-sparse signals
  - **Restricted Isometry Property** (RIP) of order 2*K*

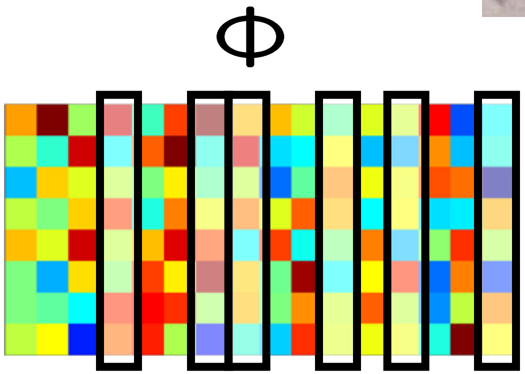
Unfortunately...



- **Goal:** Design Φ so that its *M*x2*K* submatrices are full rank
  (Restricted Isometry Property – RIP)
- Unfortunately, a combinatorial,
  NP-complete design problem

Insight from the 80's [Kashin, Gluskin]

- Draw Φ at **random**
  - iid Gaussian
  - iid Bernoulli ± 1
  - ...

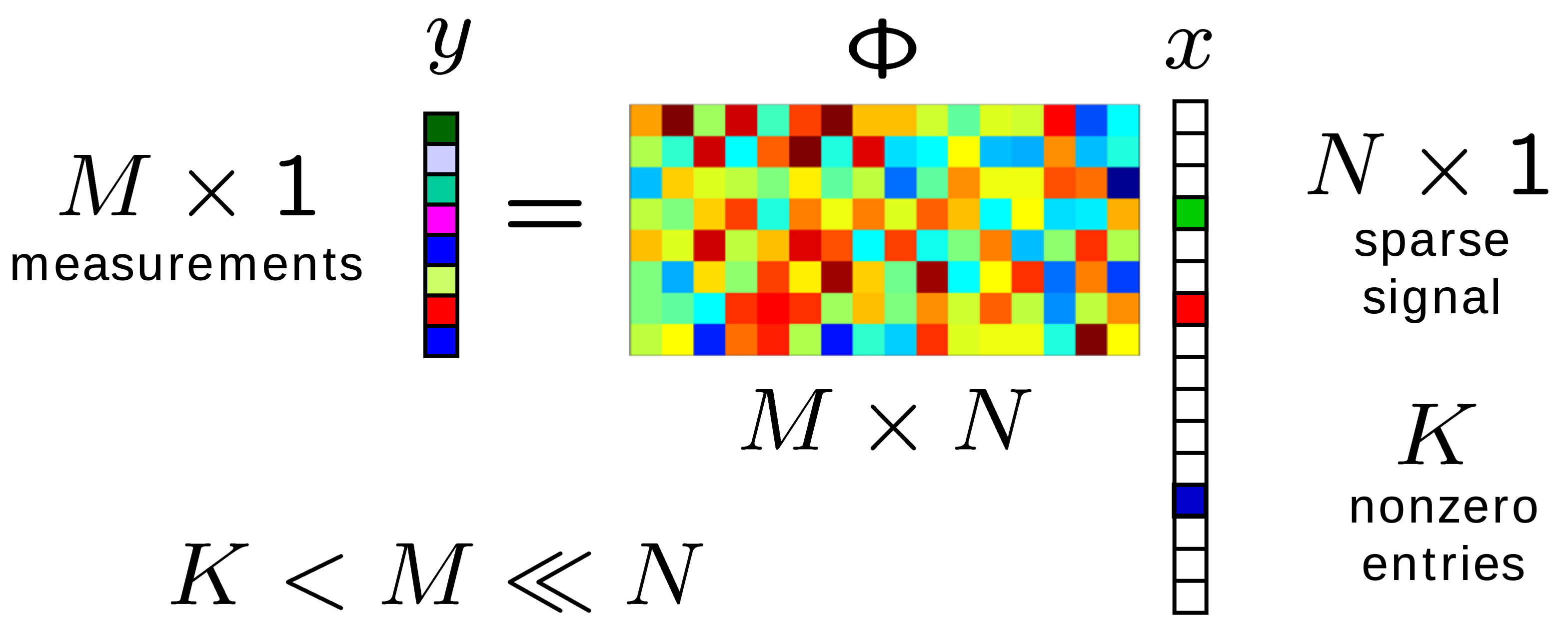


$2K$ columns

- Then Φ has the RIP with high probability as long as $M = O(K \log(N/K)) \ll N$
  - *M*x2*K* submatrices are full rank
  - stable embedding for sparse signals
  - extends to compressible signals

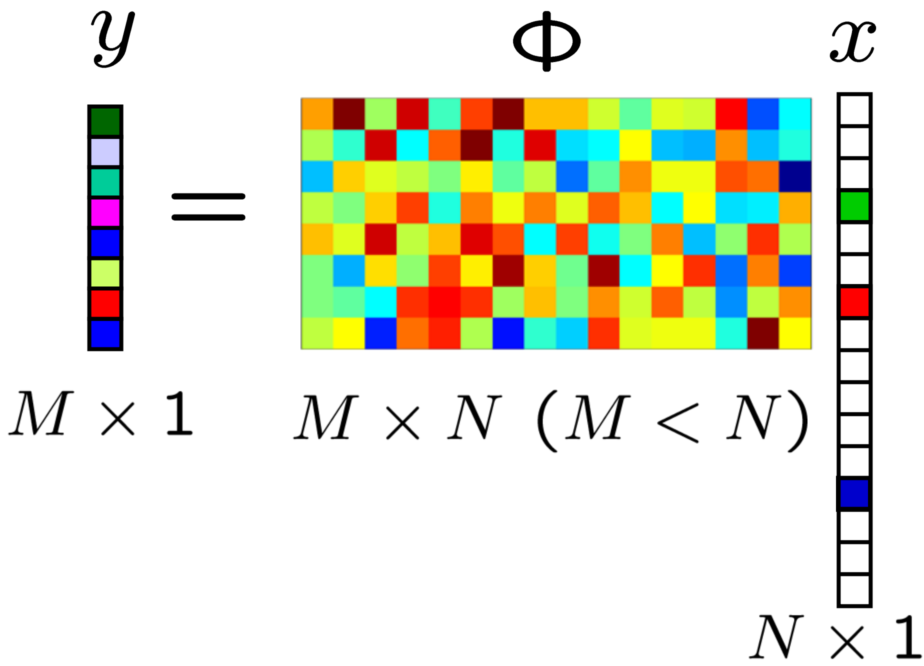
Compressive Data Acquisition

- Measurements y = **random linear combinations** of the entries of x
- WHP does not distort structure of sparse signals
  - no information loss



$$K < M \ll N$$

Compressive Sensing Recovery



1. Sparse / compressible x

   not sufficient alone

2. Projection Φ

   information preserving
   (restricted isometry property - RIP)

3. Decoding algorithms

   tractable

Compressive Sensing Recovery

- Recovery: given $y = \Phi x$
  (ill-posed inverse problem) find x (sparse)

- ℓ_2 **fast** $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$

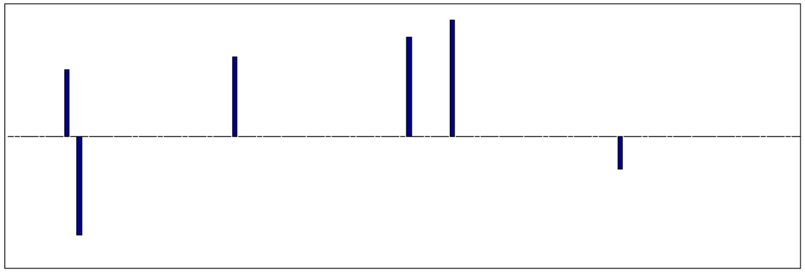
$$\widehat{x} = (\Phi^T \Phi)^{-1} \Phi^T y$$

pseudoinverse

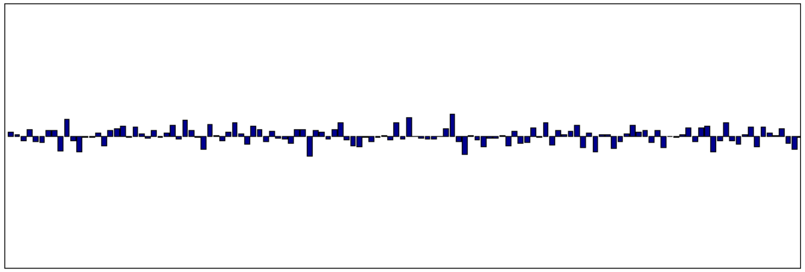
Compressive Sensing Recovery

- Recovery: given $y = \Phi x$
  (ill-posed inverse problem) find x (sparse)

- ℓ_2 **fast, wrong** $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$



x



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

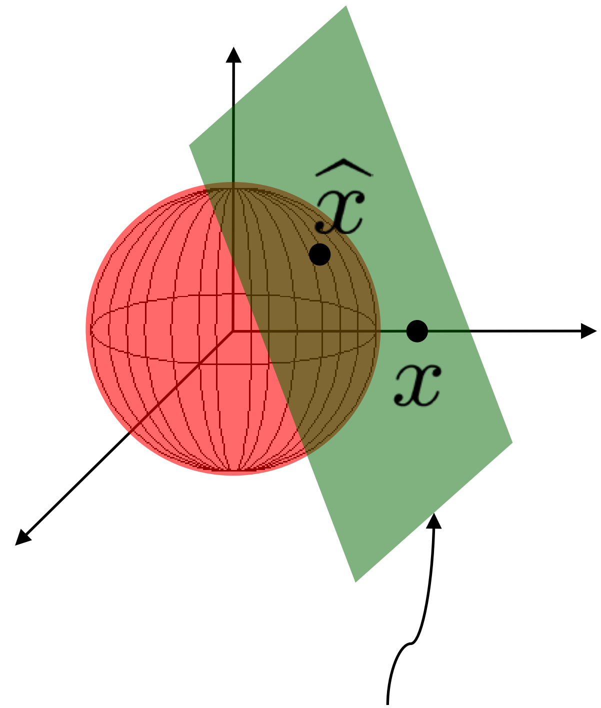
pseudoinverse

Why ℓ_2 Doesn't Work

for signals sparse in the
space / time domain

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

least squares,
minimum ℓ_2 solution
is almost **never sparse**



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

null space of Φ
translated to x
(random angle)

Compressive Sensing Recovery

- Reconstruction/decoding: given $y = \Phi x$
  (ill-posed inverse problem) find x

- ℓ_2 fast, wrong $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$

- ℓ_0 $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_0$

↑ ***number of nonzero entries***

"find ***sparsest*** x *in translated nullspace"*

Compressive Sensing Recovery

- Reconstruction/decoding: given $y = \Phi x$
  (ill-posed inverse problem) find x

- ℓ_2 fast, wrong $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$

- ℓ_0 **correct:**
  only ***M*=2*K***
  measurements
  required to
  reconstruct
  K-sparse signal

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

  ↑ *number of nonzero entries*

Compressive Sensing Recovery

- Reconstruction/decoding: given $y = \Phi x$
  (ill-posed inverse problem) find x

- ℓ_2 fast, wrong $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$

- ℓ_0 **correct:**
  only M=2K measurements required to reconstruct K-sparse signal

  slow: NP-hard algorithm

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

  ↑ *number of nonzero entries*

Compressive Sensing Recovery

- Recovery: given $y = \Phi x$
  (ill-posed inverse problem) find x (sparse)

- ℓ_2 fast, wrong $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$

- ℓ_0 correct, slow $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_0$

- ℓ_1 **correct, efficient**
  mild oversampling
  [Candes, Romberg, Tao; Donoho]
  $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_1$
  linear program

number of measurements required

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

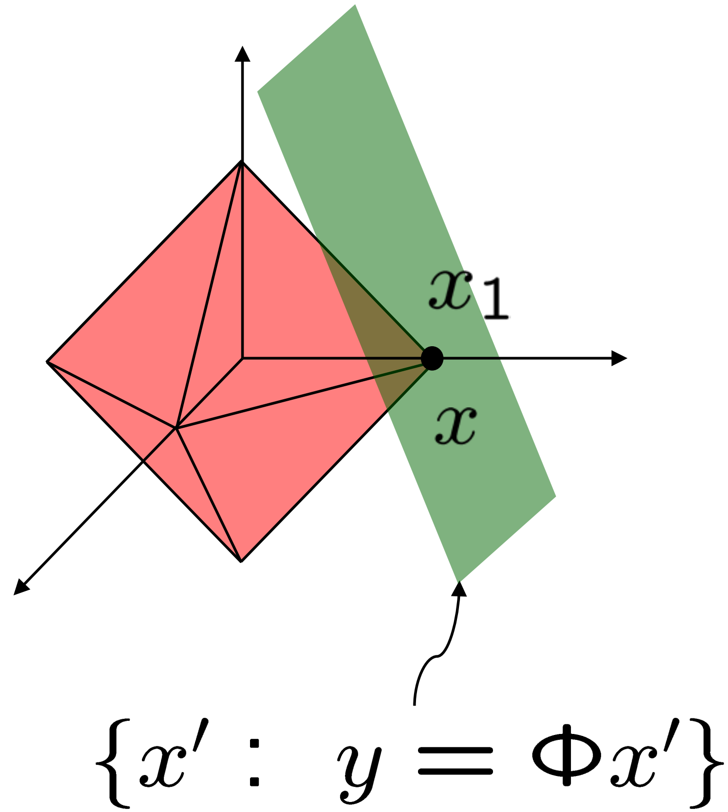
Why ℓ_1 Works

for signals sparse in the
space / time domain

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

minimum ℓ_1 solution
= sparsest solution
(with high probability) if

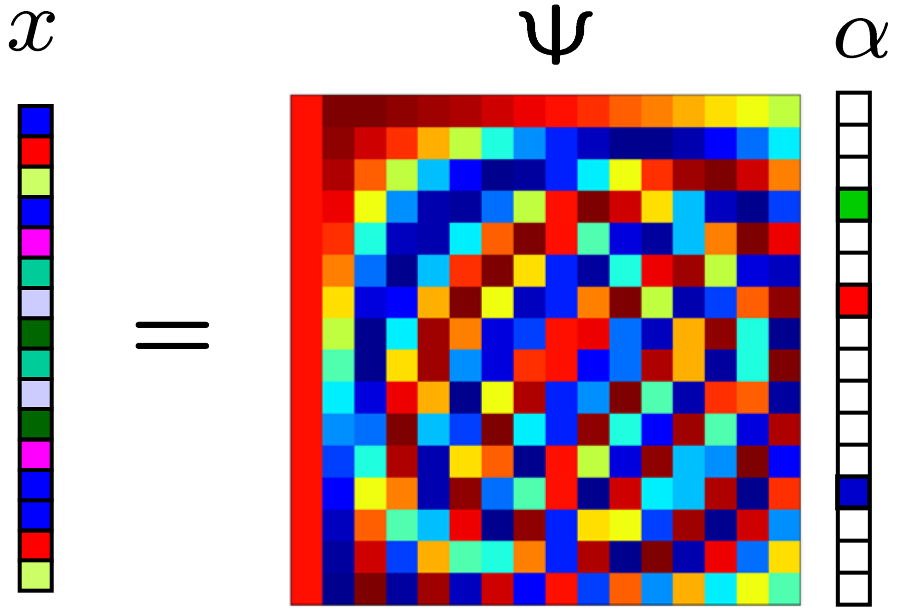
$$M = O(K \log(N/K)) \ll N$$



Universality

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

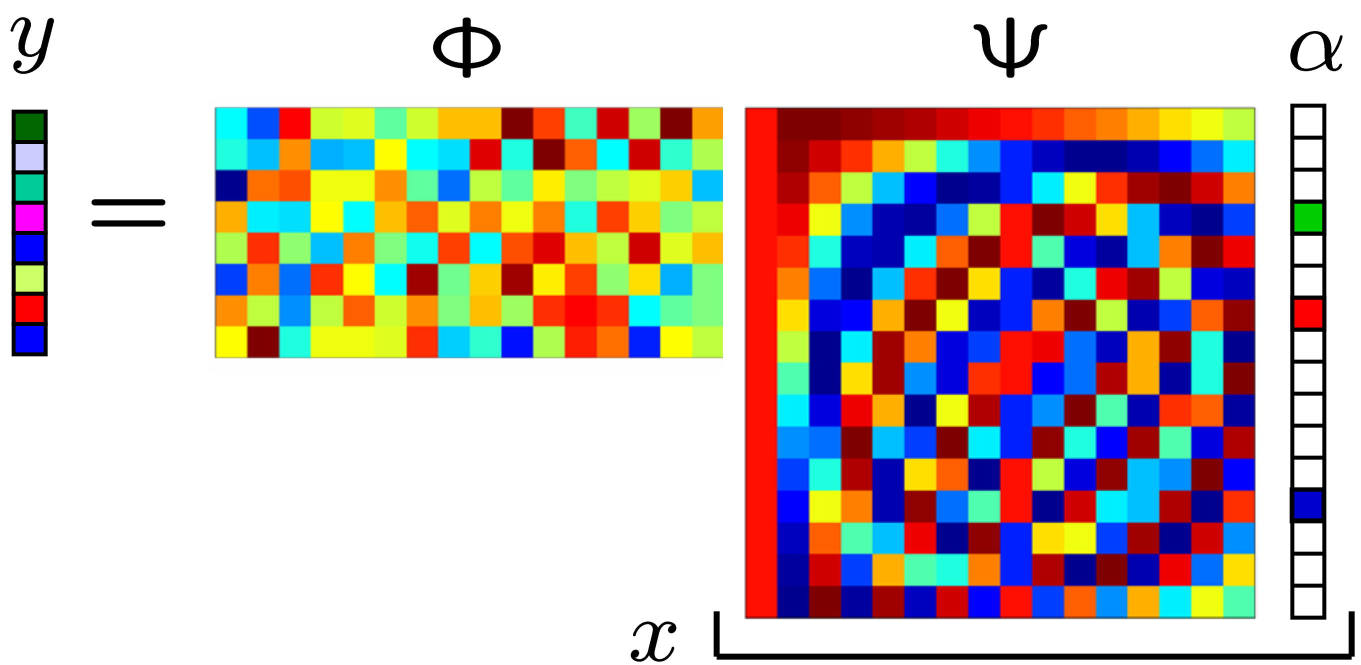
$$x = \Psi \alpha$$



Universality

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

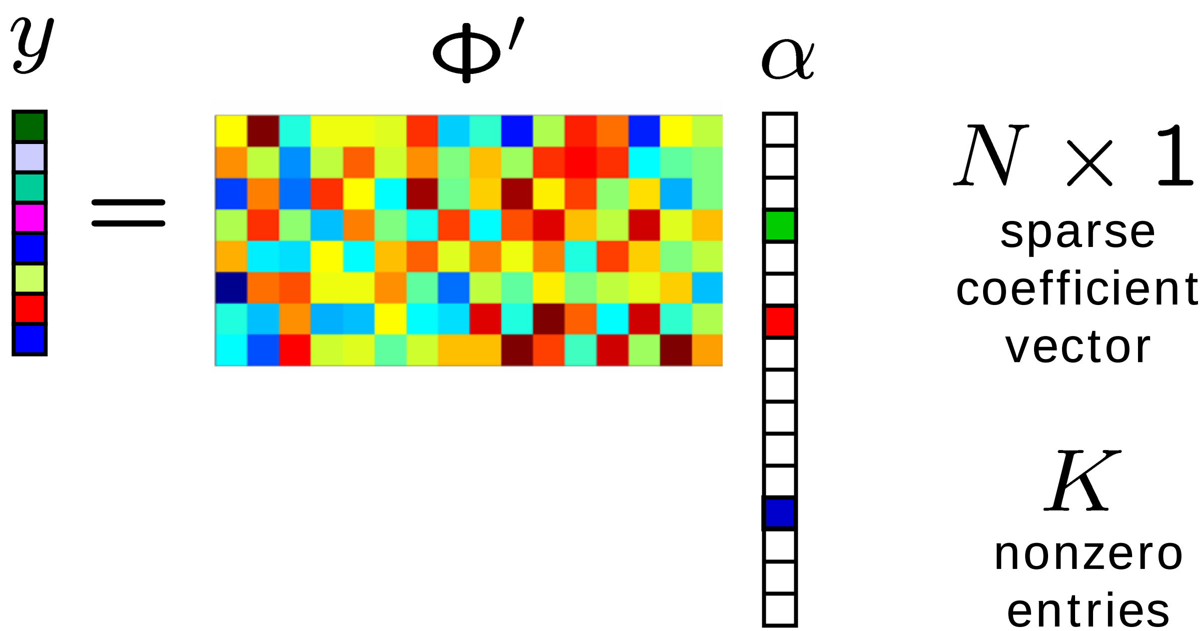
$$y = \Phi x = \Phi \Psi \alpha$$



Universality

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

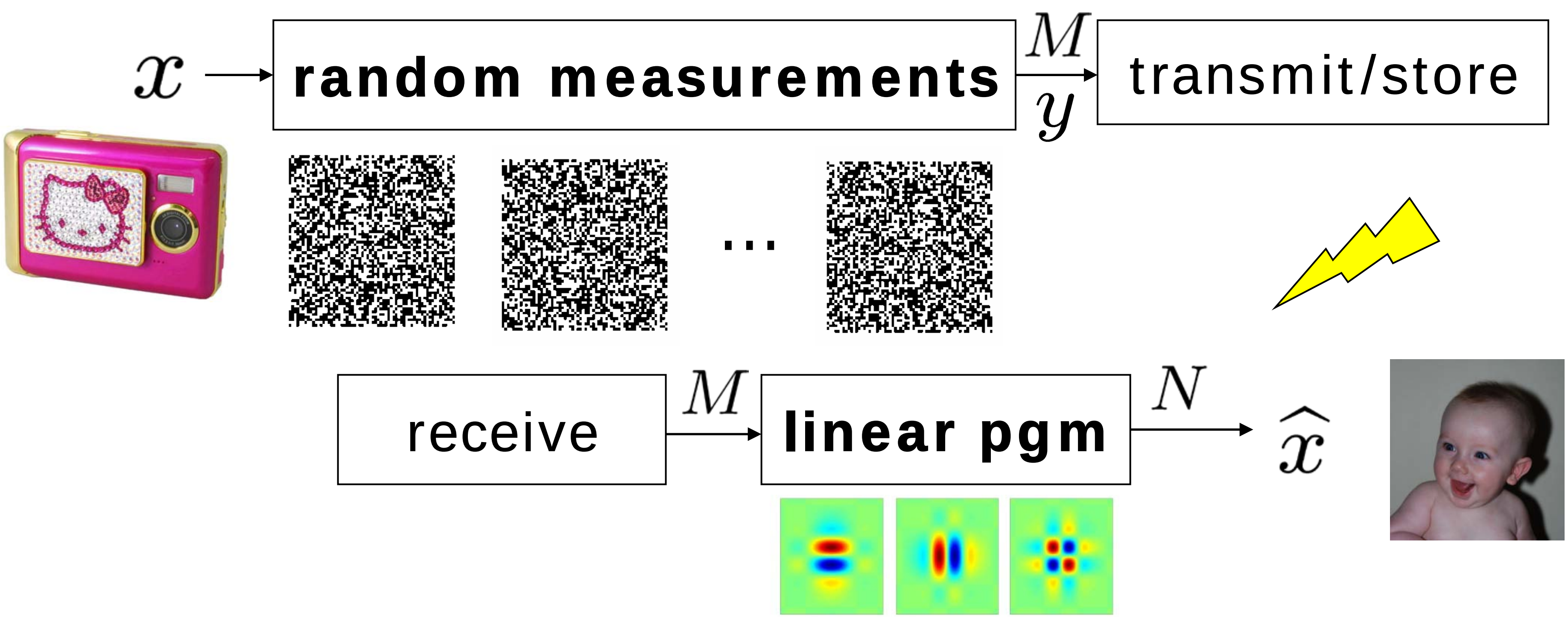
$$y = \Phi x = \Phi \Psi \alpha = \Phi' \alpha$$



Compressive Sensing

- Directly acquire "***compressed***" data
- Replace *N* samples by *M* random projections

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

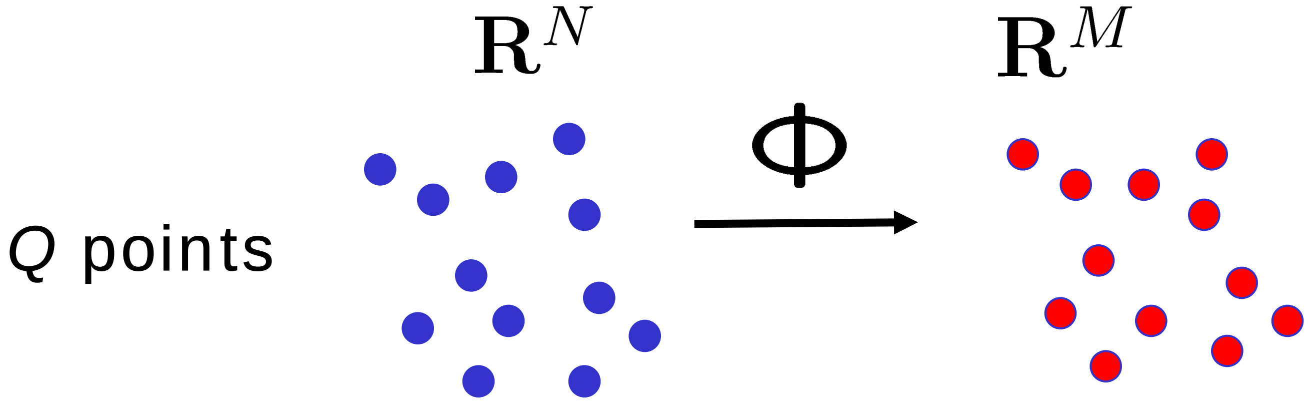


Compressive Sensing

Theory II
Stable Embedding

Johnson-Lindenstrauss Lemma

- JL Lemma: random projection stably embeds a cloud of *Q* points whp provided $M = O(\log Q)$



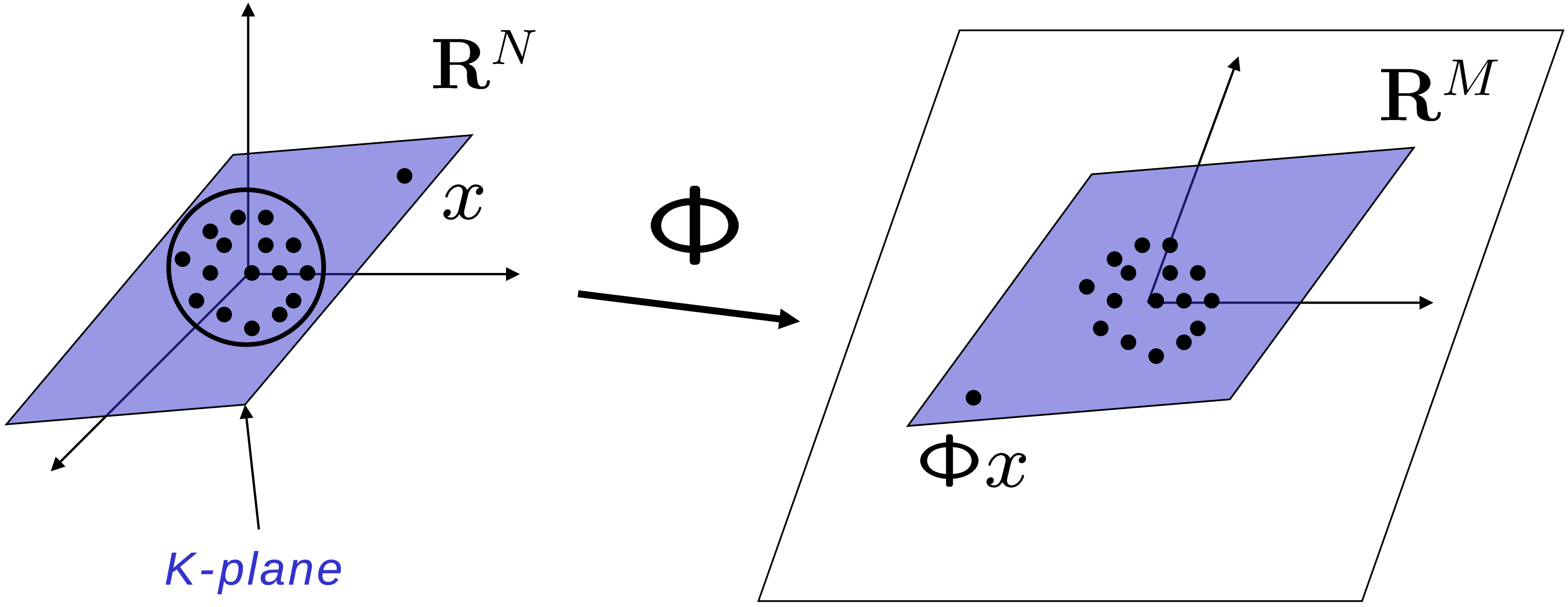
- Proved via concentration inequality

- Same techniques link JLL to RIP
  [Baraniuk, Davenport, DeVore, Wakin, *Constructive Approximation*, 2008]

Connecting JL to RIP

Consider effect of random JL Φ on each K-plane

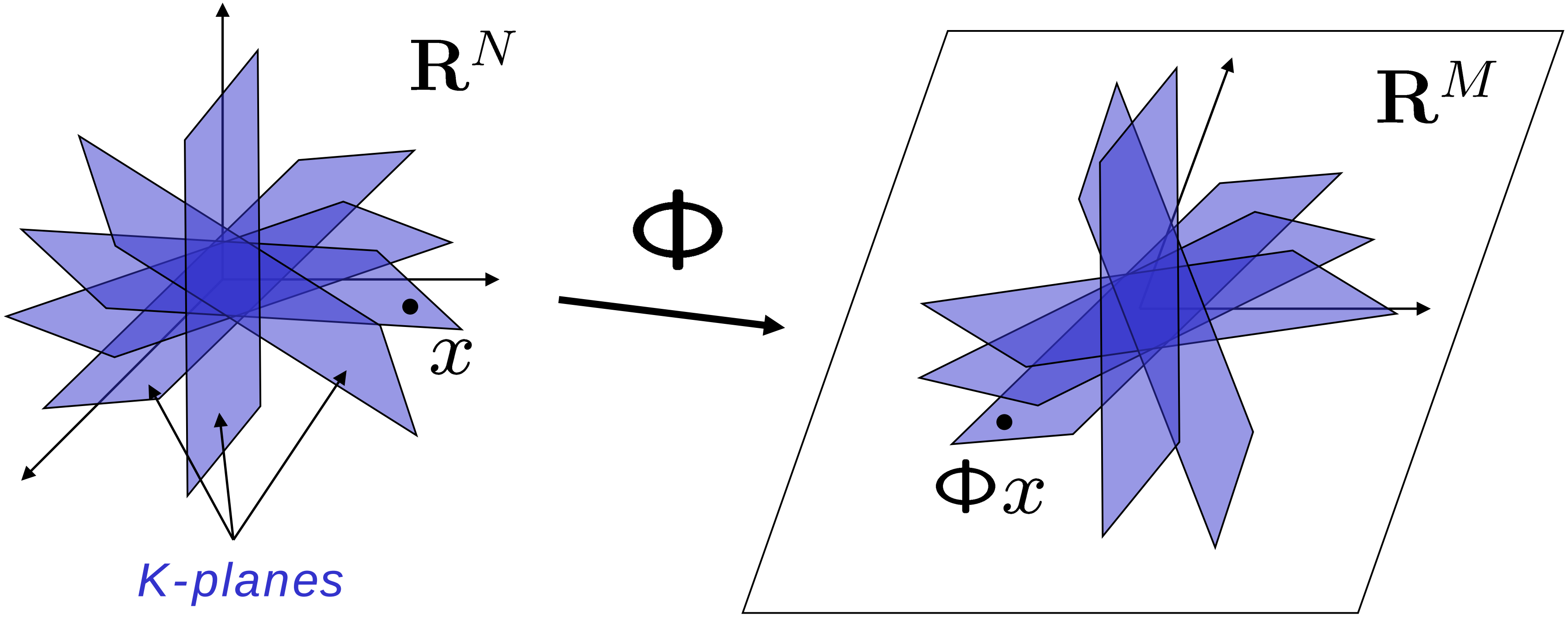
- construct covering of points Q on unit sphere
- JL: isometry for each point with high probability
- union bound ➔ isometry for all points q in Q
- extend to isometry for all x in K-plane



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 points q in Q
- extend to isometry for all x in K-plane
- union bound ➔ isometry for all K-planes



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]

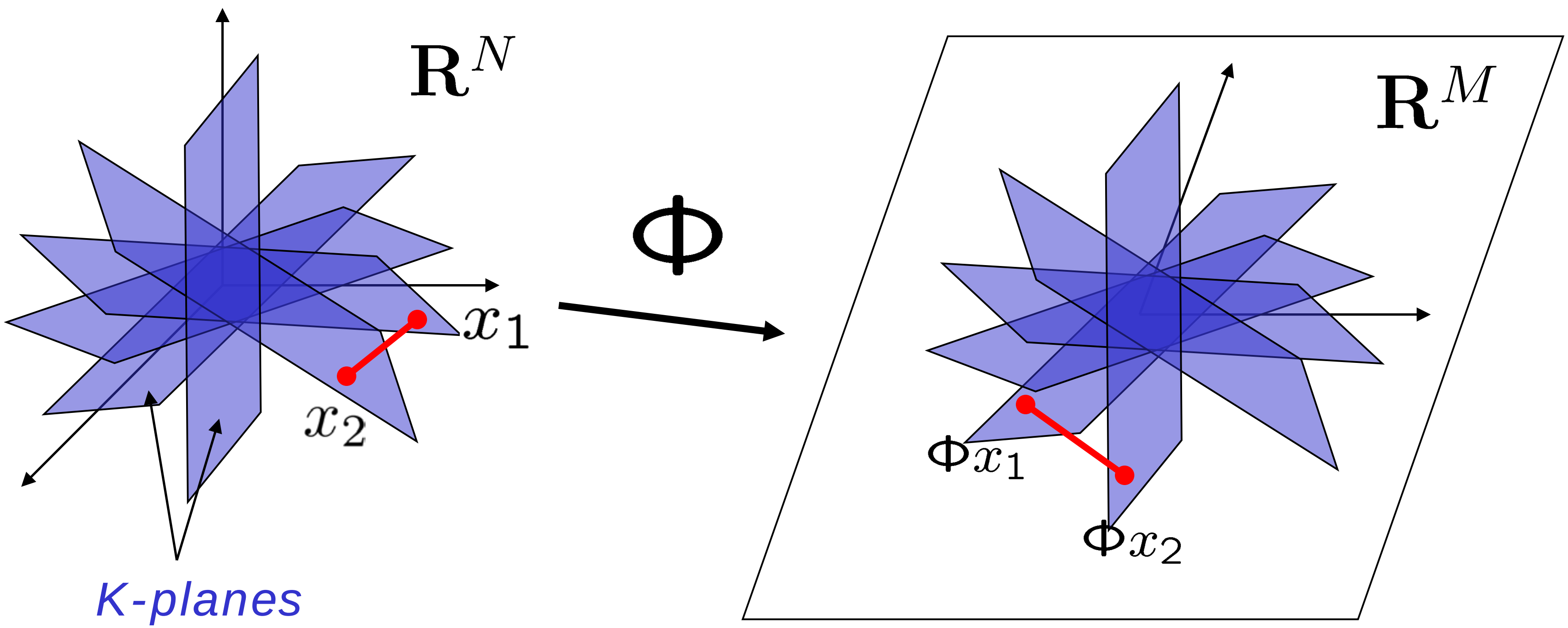
$$\phi_{i,j} := \begin{cases} +\sqrt{\frac{3}{M}} & \text{with probability } \frac{1}{6}, \\ 0 & \text{with probability } \frac{2}{3}, \\ -\sqrt{\frac{3}{M}} & \text{with probability } \frac{1}{6} \end{cases}$$

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

RIP as a "Stable" Embedding

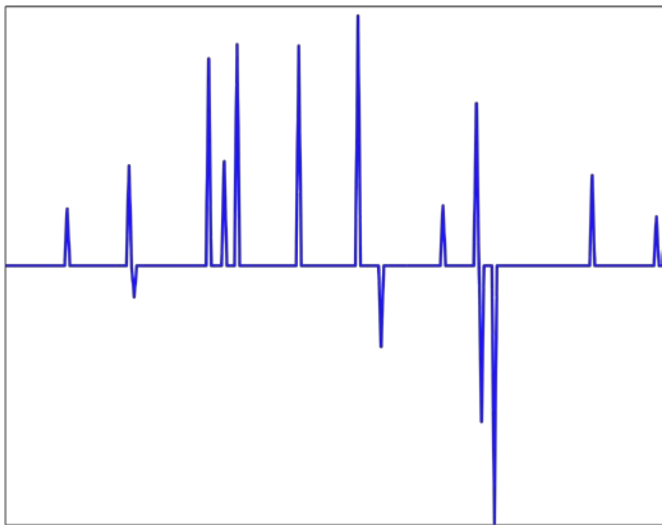
- RIP of order $2K$ implies: for all K-sparse x_1 and x_2,

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

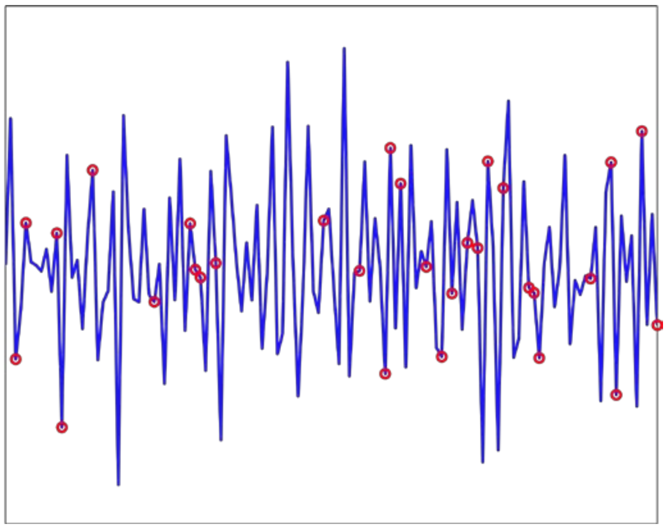


Structured Random Matrices

- There are more structured (but still random) compressed sensing matrices
- We can randomly sample in a domain whose basis vectors are *incoherent* with the sparsity basis
- Example: sparse in time, sample in frequency



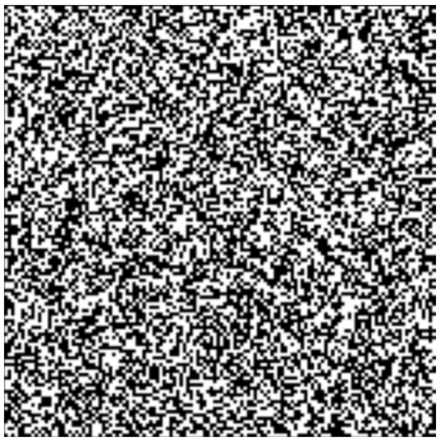
time



frequency

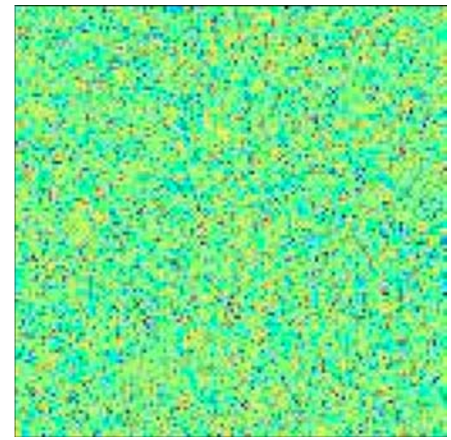
Structured Random Matrices

- Signal is sparse in the wavelet domain, measured with *noiselets* (Coifman et al. '01)



2D noiselet

wavelet domain



noiselet domain

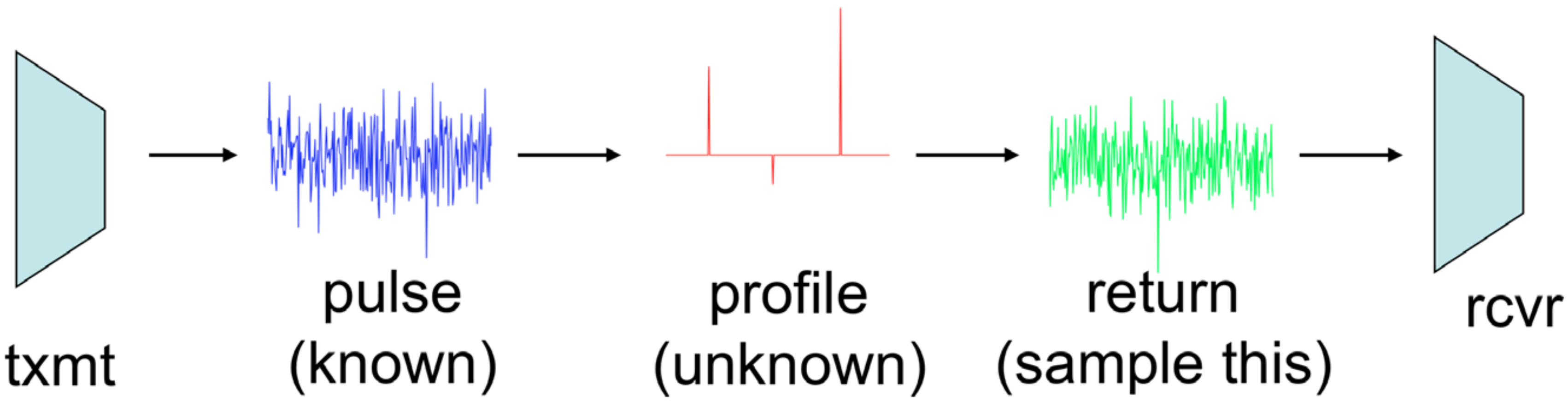
- Stable recovery from

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

measurements

Random Convolution

- A natural way to implement compressed sensing is through random convolution



- Applications include active imaging (radar, sonar,...)

- Many recent theoretical results

(R 08, Bajwa, Haupt et al 08, Rauhut 09)

Random Convolution Theory

- Convolution with a random pulse, then subsample

$$\Phi = R_\Omega F^* \Sigma F, \quad \Sigma = \mathrm{diag}(\{\sigma_\omega\})$$

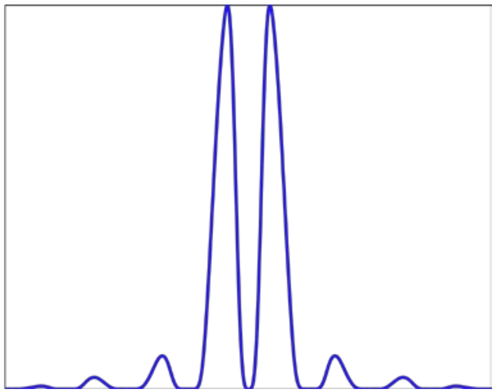
(each σ_ω has unit magnitude and random phase)

- Stable recovery from

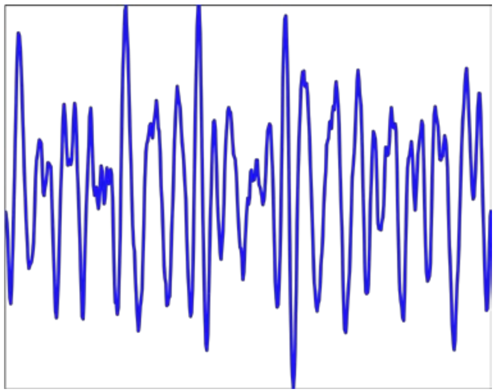
$$M = O(K \log^5 N)$$

measurements

time



frequency

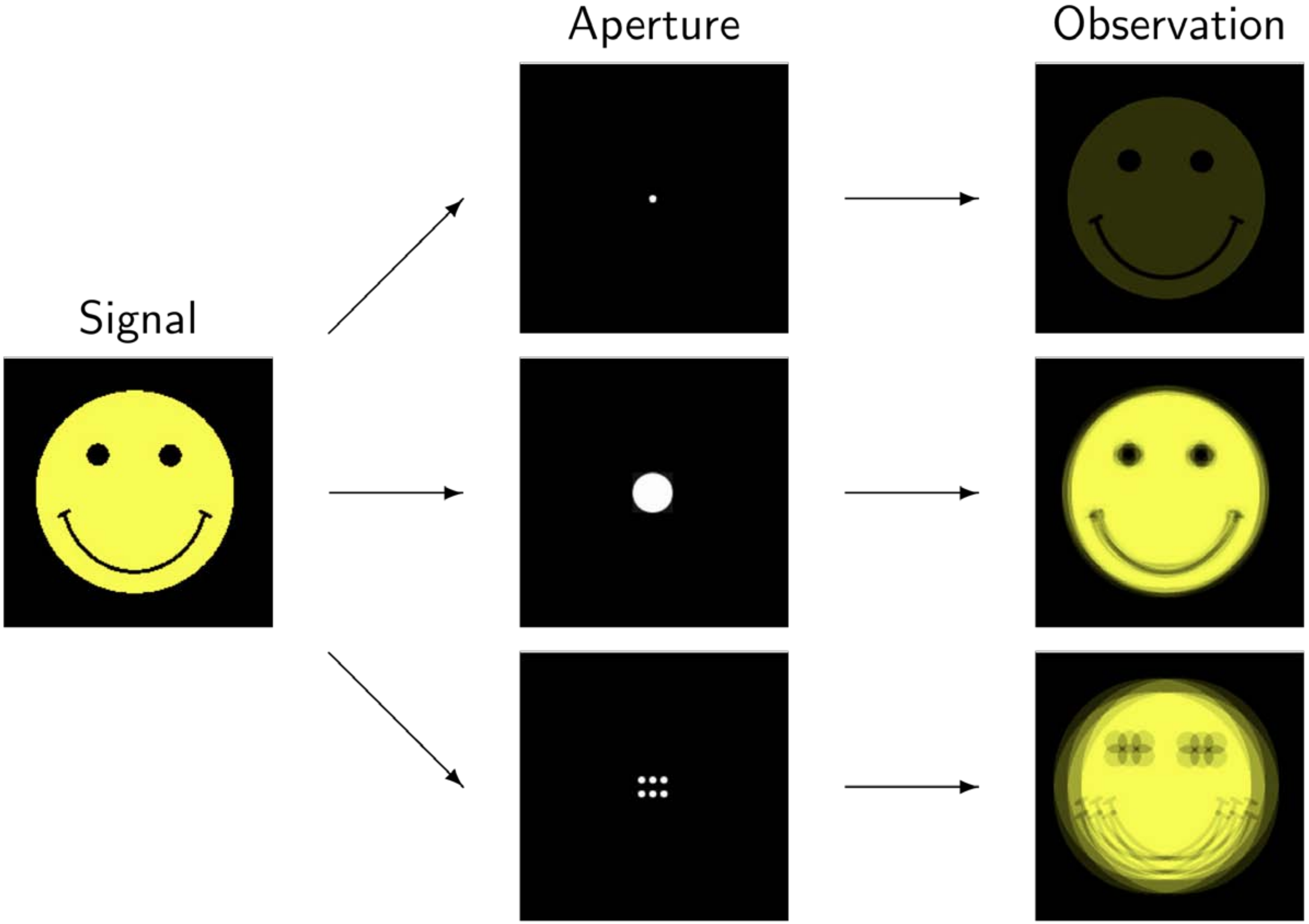


after convolution

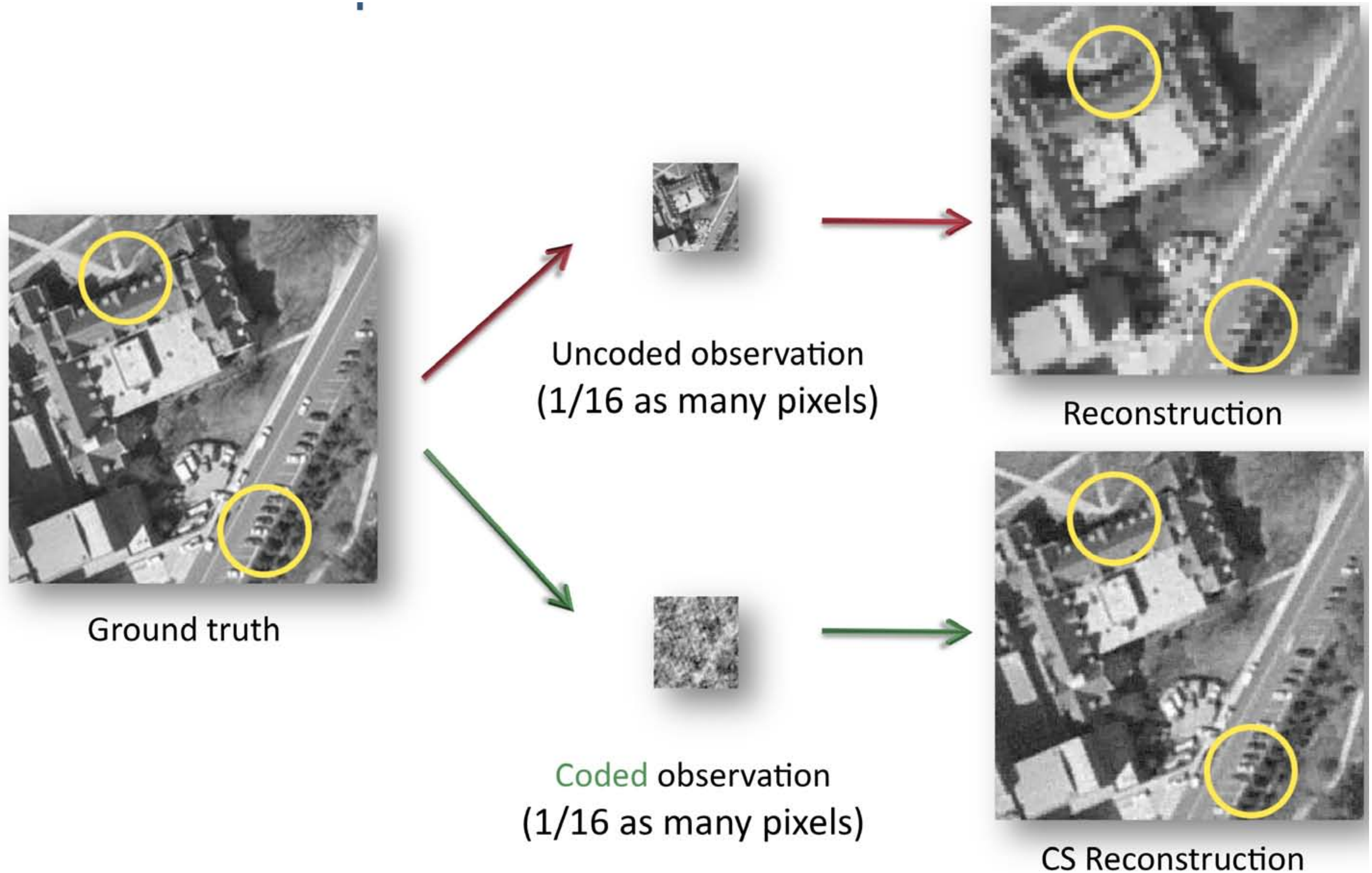
Coded Aperture Imaging

- Allows high-levels of light and high resolution

(Marcia and Willett 08, also Brady, Portnoy and others)



Super-resolved Imaging



(Marcia and Willet 08)

Stability

- Recovery is robust against noise and modeling error

- Suppose we observe

$$y = \Phi x_0 + e, \quad \|e\|_2 \le \epsilon$$

- Relax the recovery algorithm, solve

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

- The recovery error obeys

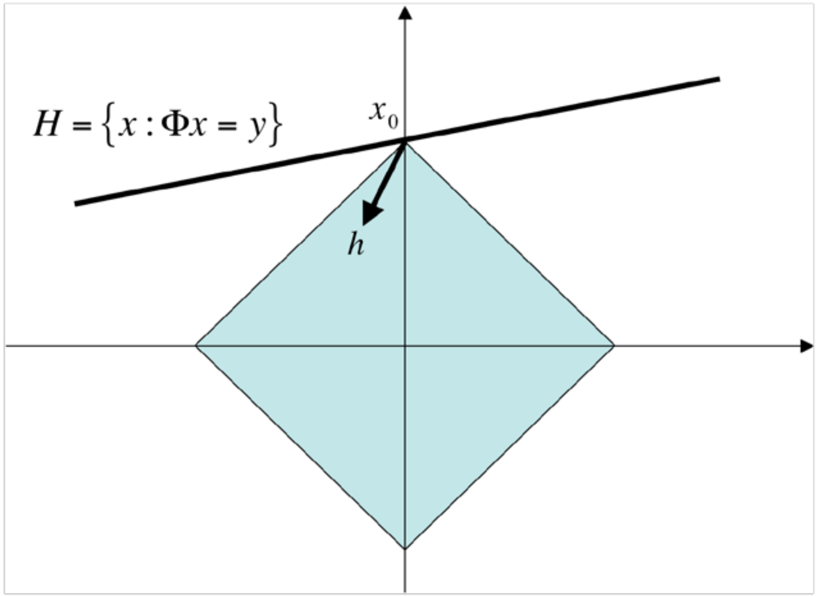
$$\|x^* - x_0\|_2 \ \lesssim \ \epsilon + \frac{\|x_{0,K} - x_0\|_{\ell_1}}{\sqrt{K}}$$

measurement error + *approximation error*

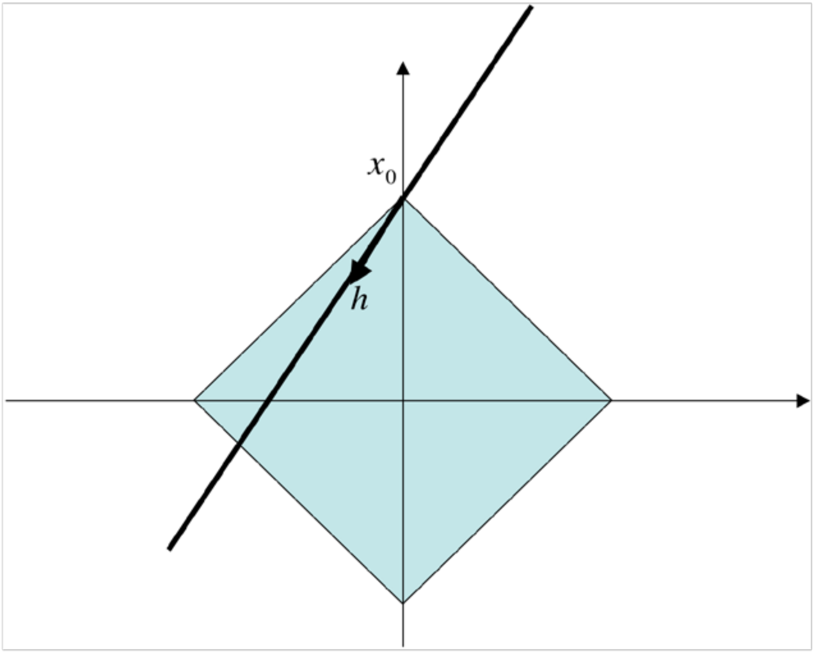
$x_{0,K} =$ best K-term approximation

Geometrical Viewpoint, Noiseless

good



bad



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

$$\|x_0 + h\|_{\ell_1} < \|x_0\|_{\ell_1}$$

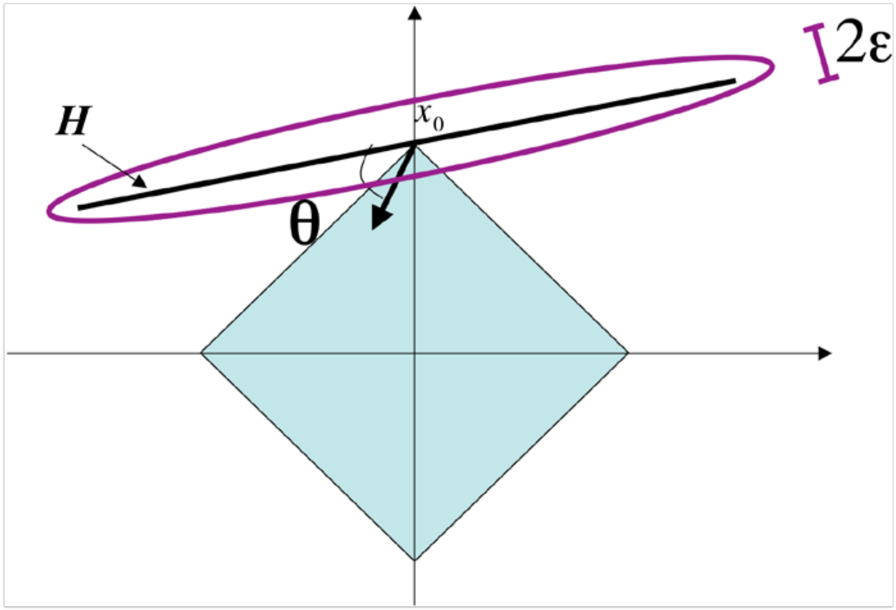
- x_0 is the solution if

$$\Phi h \neq 0$$

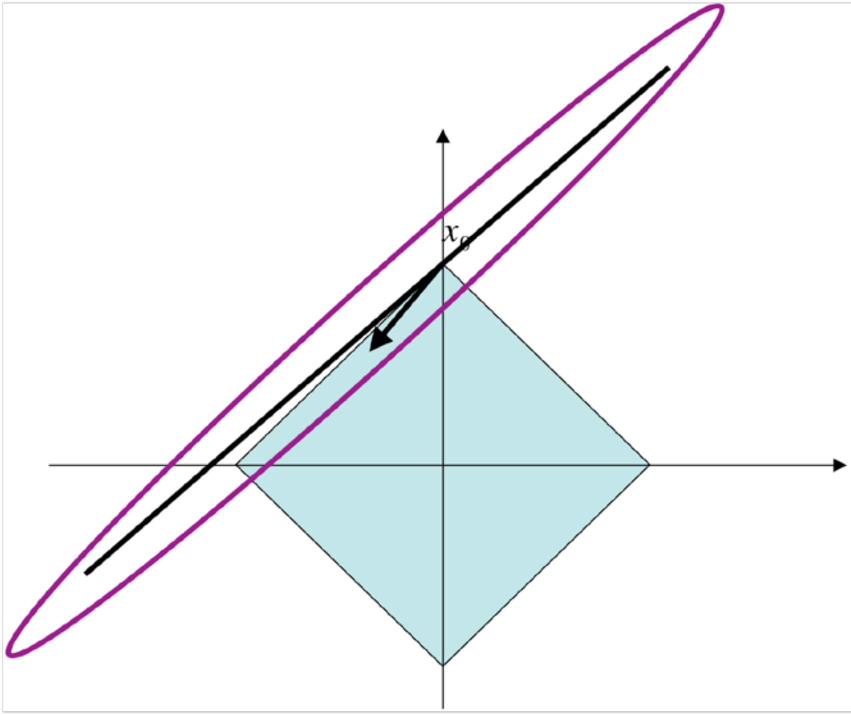
for all such descent vectors

Geometrical Viewpoint, Noise

good



bad



- Solution will be within ϵ of $\boldsymbol{H}$
- Need that not too much of the ℓ_1 ball near x_0 is feasible

Compressive Sensing

Recovery Algorithms

CS Recovery Algorithms

- Convex optimization:
  - noise-free signals
    - Linear programming (Basis pursuit)
    - FPC
    - Bregman iteration, ...
  - noisy signals
    - Basis Pursuit De-Noising (BPDN)
    - Second-Order Cone Programming (SOCP)
    - Dantzig selector
    - GPSR, ...

- Iterative greedy algorithms
  - Matching Pursuit (MP)
  - Orthogonal Matching Pursuit (OMP)
  - StOMP
  - CoSaMP
  - Iterative Hard Thresholding (IHT), ...

software @
dsp.rice.edu/cs

L1 with equality constraints = linear programming

The standard L1 recovery program

$$\min_x \; \|x\|_{\ell_1} \quad \text{s.t.} \quad y = \Phi x$$

is equivalent to the linear program

$$\min_{x,t} \sum_i t_i \quad \text{s.t.} \quad -t_i \le x_i \le t_i, \quad \Phi x = y$$

There has been a tremendous amount of progress in solving linear programs in the last 15 years

SOCP

- Standard LP recovery

$$\min \|x\|_1 \quad \text{subject to } y = \Phi x$$

- Noisy measurements

$$y = \Phi x + n$$

- Second-Order Cone Program

$$\min \|x\|_1 \quad \text{subject to } \|y - \Phi x\|_2 \leq \epsilon$$

- Convex, **quadratic program**

Other Flavors of L1

- Quadratic relaxation (called LASSO in statistics)

$$\min_x \ \|x\|_{\ell_1} \ + \ \lambda\|y - \Phi x\|_2^2$$

- Dantzig selector (residual correlation constraints)

$$\min_x \ \|x\|_{\ell_1} \quad \text{s.t.} \quad \|\Phi^T(y - \Phi x)\|_\infty$$

- L1 Analysis (Ψ is an overcomplete frame)

$$\min_x \ \|\Psi^T x\|_{\ell_1} \quad \text{s.t.} \quad \|y - \Phi x\|_2 \leq \epsilon$$

Solving L1

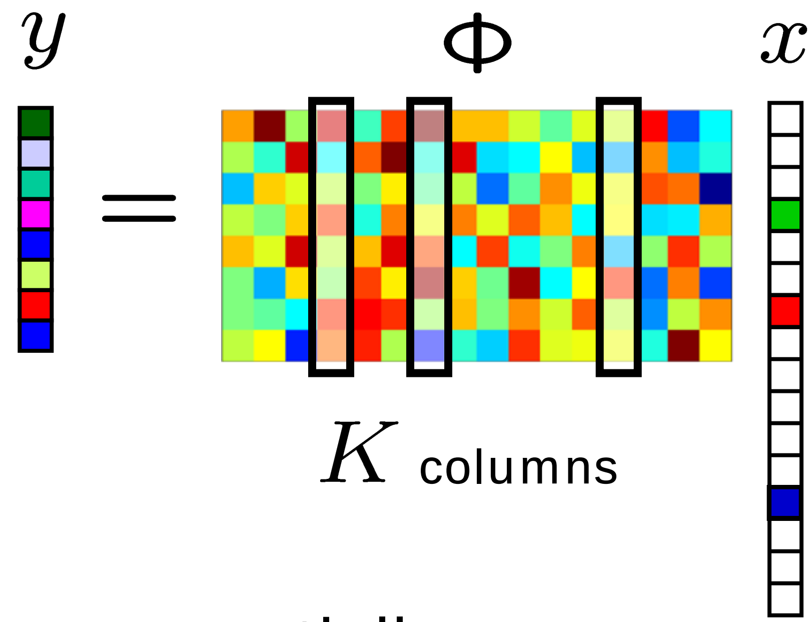
- "Classical" (mid-90s) interior point methods
  - main building blocks due to Nemirovski
  - second-order, series of local quadratic approximations
  - boils down to a series of linear systems of equations
  - formulation is very general (and hence adaptable)
- Modern progress (last 5 years) has been on "first order" methods
  - Main building blocks due to Nesterov (mid 80s)
  - iterative, require applications of Φ and Φ^T at each iteration
  - convergence in 10s-100s of iterations typically
- Many software packages available
  - Fixed-point continuation (Rice)
  - Bregman iteration-based methods (UCLA)
  - NESTA (Caltech)
  - GPSR (Wisconsin)
  - SPGL1 (UBC)…..

Matching Pursuit

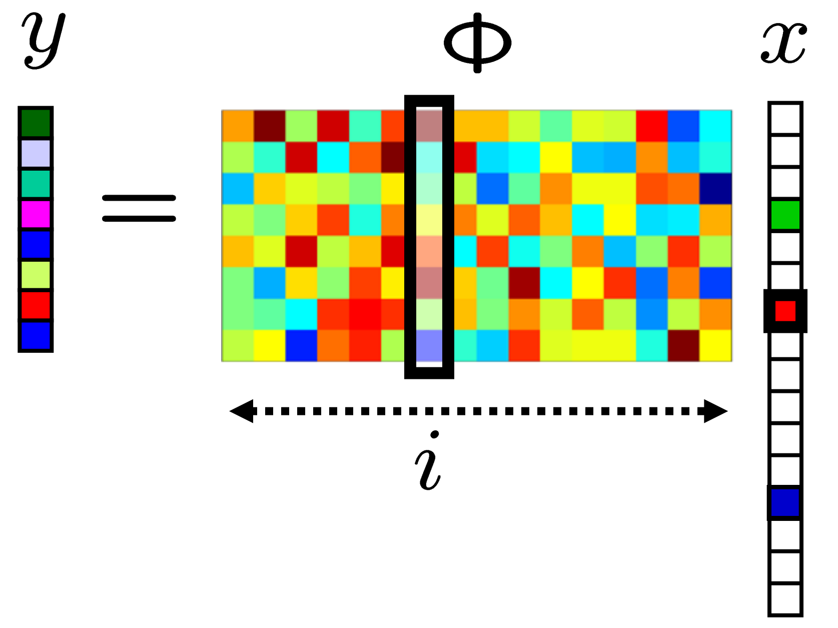
- Greedy algorithm

- **Key ideas:**
(1) measurements y
composed of sum
of K columns of Φ

(2) identify which K columns sequentially
according to size of contribution to y



Matching Pursuit



- For each column ϕ_i compute

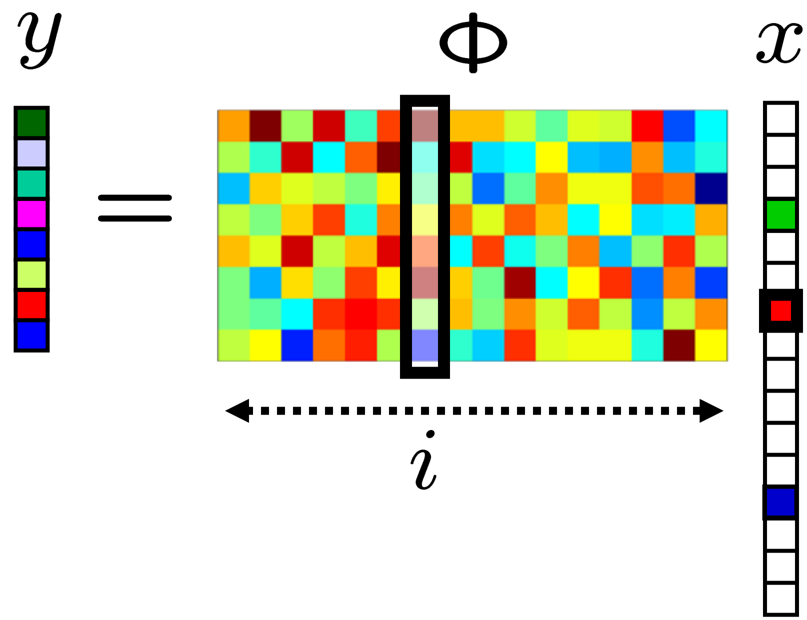
$$\widehat{x}_i = \langle y, \phi_i \rangle$$

- Choose largest $|\widehat{x}_i|$ (greedy)

- Update estimate $\widehat{x}$ by adding in $\widehat{x}_i$

- Form residual measurement and iterate until convergence $y' = y - x_i \phi_i$

Orthogonal Matching Pursuit



- Same procedure as Matching Pursuit
- Except at each iteration:
  - remove selected column ϕ_i
  - re-orthogonalize the remaining columns of Φ
- Converges in K iterations

CoSAMP

- Needell and Tropp, 2008
- Very simple greedy algorithm, provably effective

Algorithm 2.1: CoSaMP Recovery Algorithm

CoSaMP($\mathbf{\Phi}$, $\boldsymbol{u}$, s)
Input: Sampling matrix $\mathbf{\Phi}$, noisy sample vector $\boldsymbol{u}$, sparsity level s
Output: An s-sparse approximation $\boldsymbol{a}$ of the target signal

---

$\boldsymbol{a}^0 \leftarrow \mathbf{0}$ { Trivial initial approximation }
$\boldsymbol{v} \leftarrow \boldsymbol{u}$ { Current samples = input samples }
$k \leftarrow 0$

repeat
$\quad k \leftarrow k + 1$

$\quad \boldsymbol{y} \leftarrow \mathbf{\Phi}^* \boldsymbol{v}$ { Form signal proxy }
$\quad \Omega \leftarrow \mathrm{supp}(\boldsymbol{y}_{2s})$ { Identify large components }
$\quad T \leftarrow \Omega \cup \mathrm{supp}(\boldsymbol{a}^{k-1})$ { Merge supports }

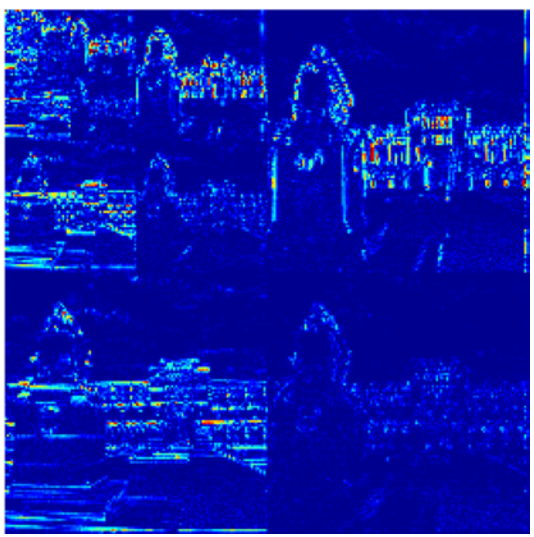
$\quad \boldsymbol{b}|_T \leftarrow \mathbf{\Phi}_T^\dagger \boldsymbol{u}$ { Signal estimation by least-squares }
$\quad \boldsymbol{b}|_{T^c} \leftarrow \mathbf{0}$

$\quad \boldsymbol{a}^k \leftarrow \boldsymbol{b}_s$ { Prune to obtain next approximation }
$\quad \boldsymbol{v} \leftarrow \boldsymbol{u} - \mathbf{\Phi} \boldsymbol{a}^k$ { Update current samples }

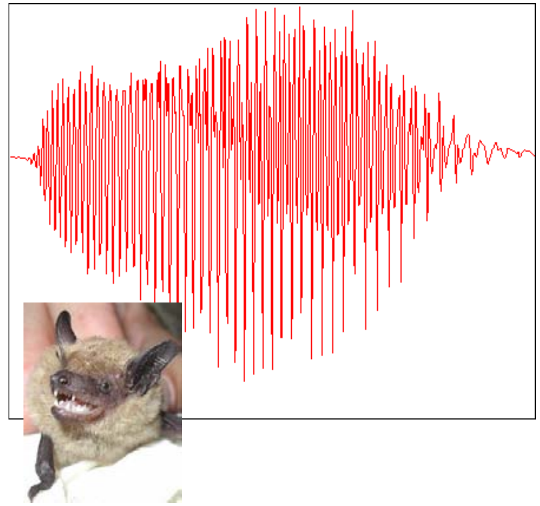
until halting criterion *true*

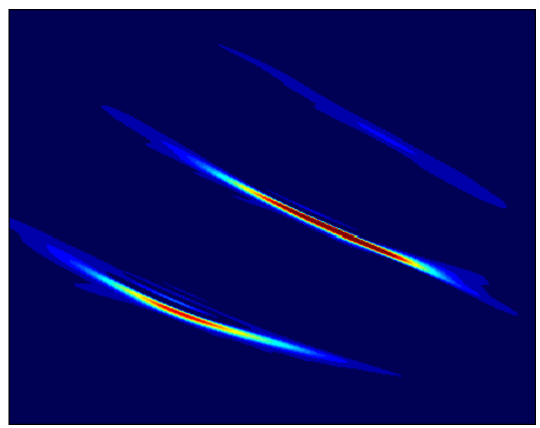
From Sparsity
to
Model-based (*structured*) Sparsity

Sparse Models

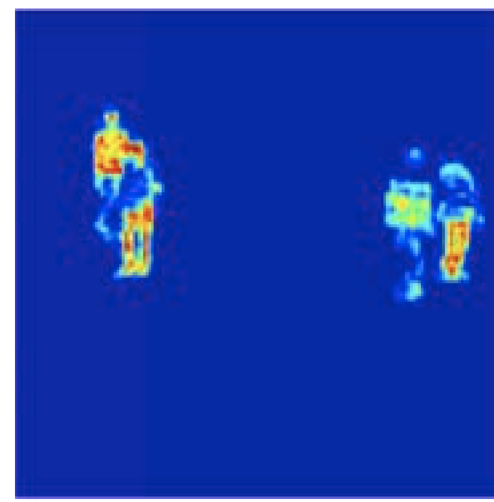


wavelets:
natural images





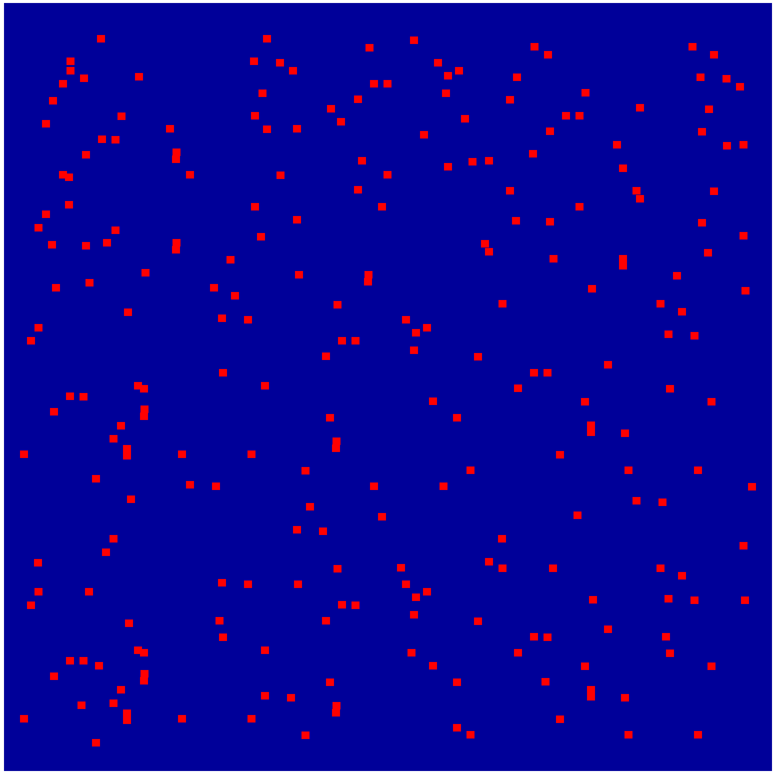
Gabor atoms:
chirps/tones



pixels:
background subtracted
images

Sparse Models

- Sparse/compressible signal model captures **simplistic primary structure**

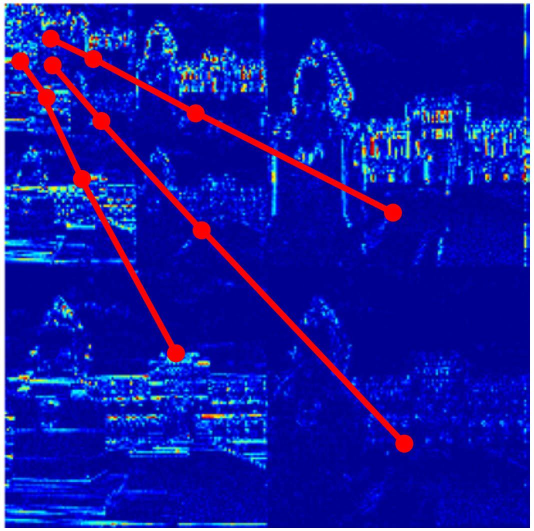


sparse image

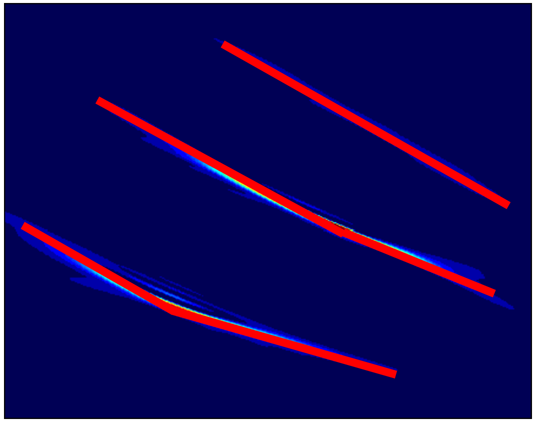
Beyond Sparse Models

- Sparse/compressible signal model captures
  simplistic primary structure

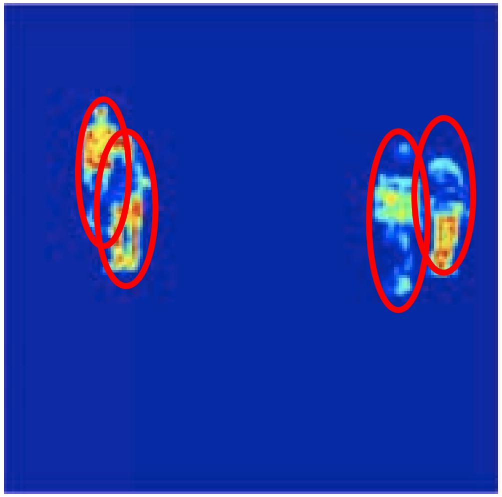
- Modern compression/processing algorithms capture
  richer secondary coefficient structure



wavelets:
natural images



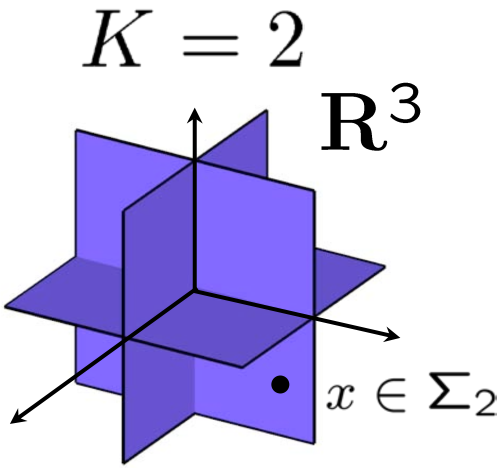
Gabor atoms:
chirps/tones

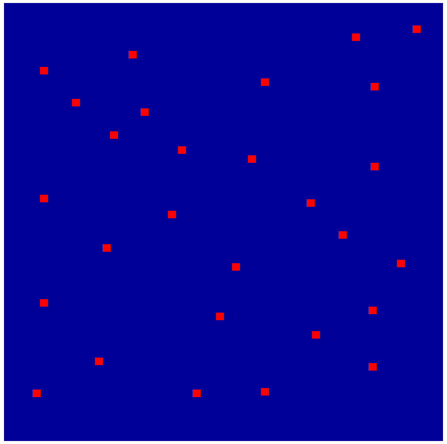


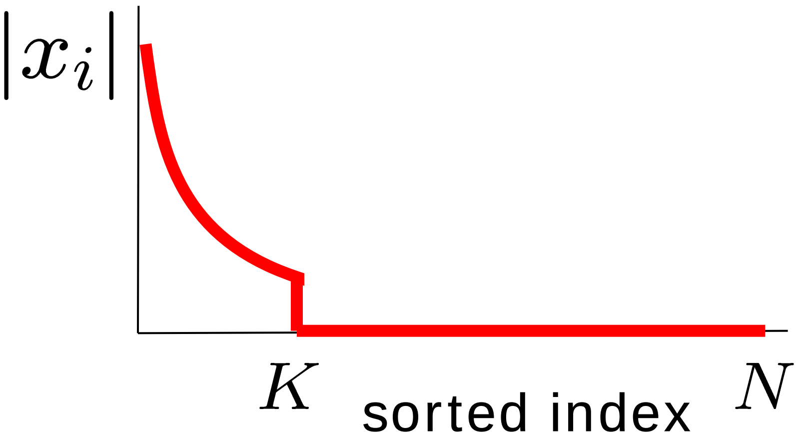
pixels:
background subtracted images

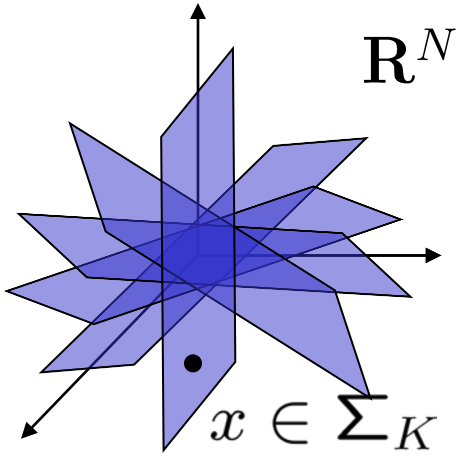
Signal Priors

- **Sparse** signal: only K out of N coordinates nonzero
  - model: union of all K-dimensional subspaces aligned w/ coordinate axes





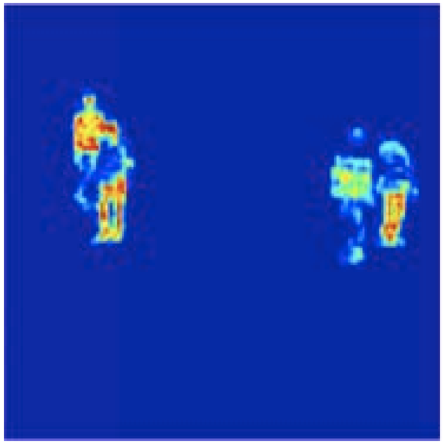


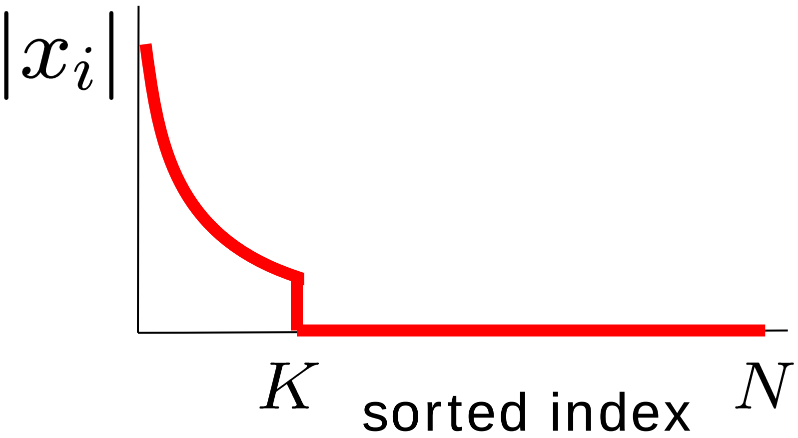


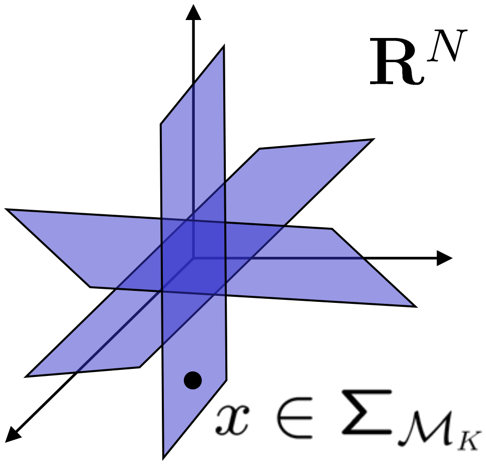
Signal Priors

- **Sparse** signal: only K out of N coordinates nonzero
  - model: union of all K-dimensional subspaces aligned w/ coordinate axes

- **Structured sparse** signal: reduced set of subspaces **(or model-sparse)**
  - model: a particular union of subspaces ex: clustered or dispersed sparse patterns







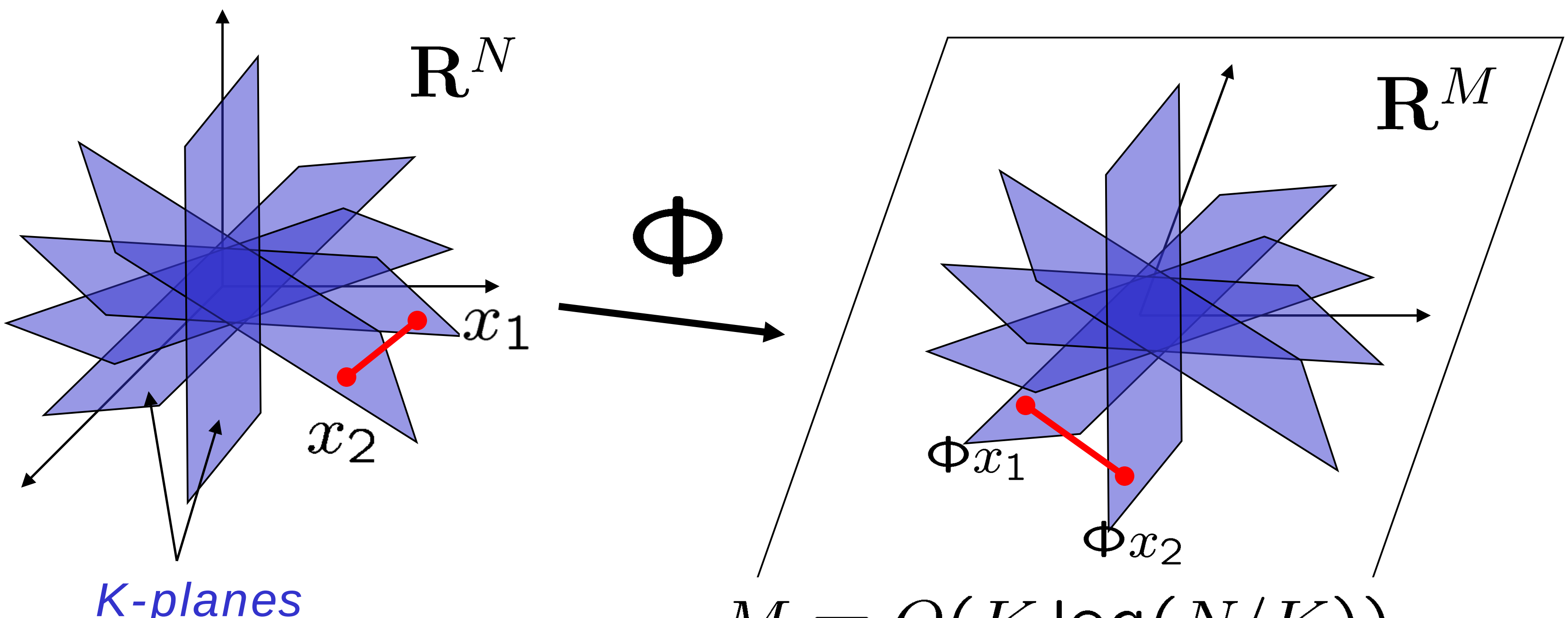
Restricted Isometry Property (RIP)

- **Model:** *K*-sparse

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

- **RIP:** stable embedding

Random subGaussian (iid Gaussian, Bernoulli) matrix **< >** RIP w.h.p.



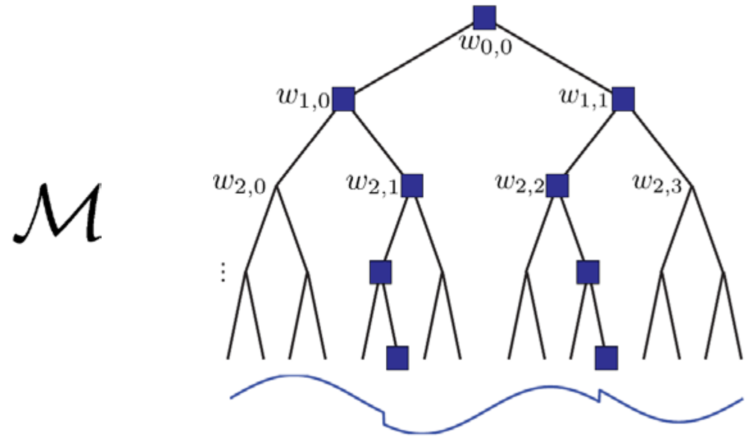
$$M = O(K \log(N/K))$$

Restricted Isometry Property (RIP)

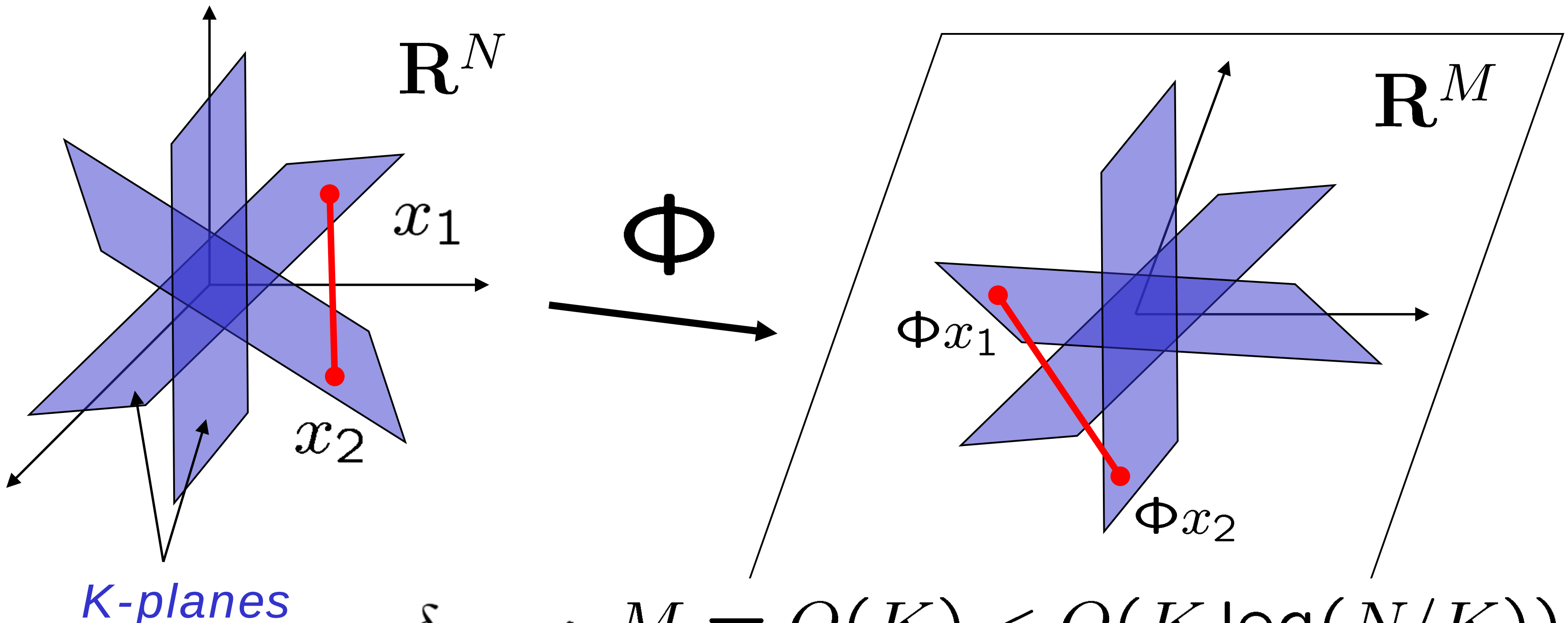
- **Model:** *K*-sparse

  \+ significant coefficients lie on a rooted subtree

  (a known model for piecewise smooth signals)



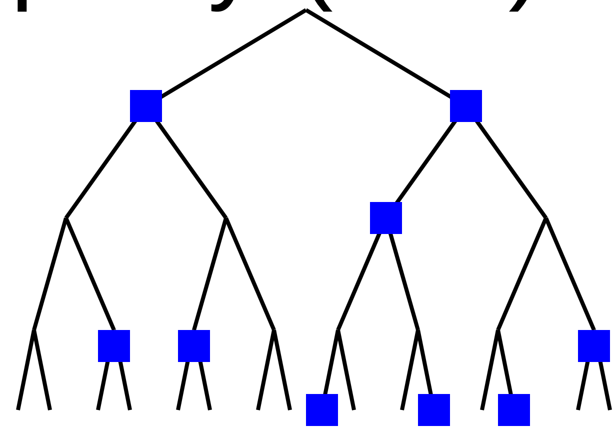
- **Tree-RIP:** stable embedding



$$\delta_{\mathcal{M}_{2K}}: \; \underline{M = O(K)} < O(K \log(N/K))$$

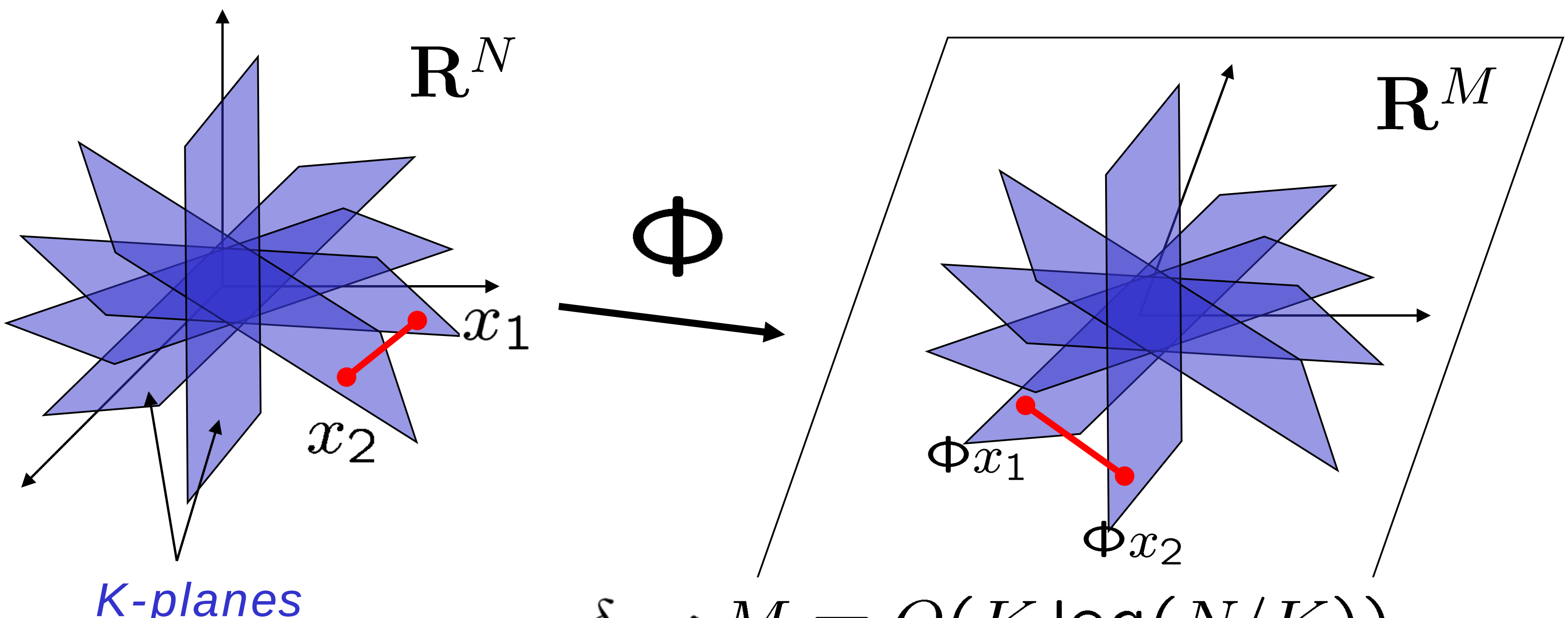
Restricted Isometry Property (RIP)

- **Model:** *K*-sparse



Note the difference:

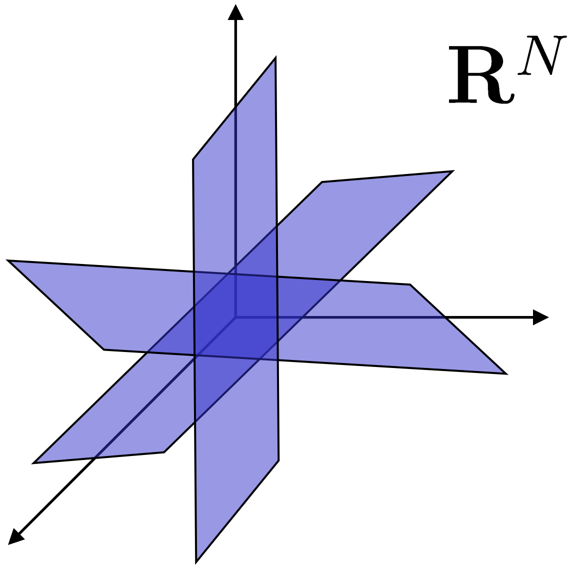
- **RIP:** stable embedding



$$\delta_{2K}\colon M = O(K \log(N/K))$$

Model-Sparse Signals

- Defn: A ***K*-sparse signal *model*** comprises a particular (*reduced*) set of *K*-dim canonical subspaces

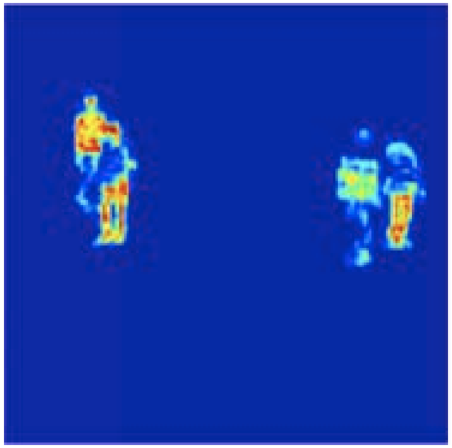


- **Structured subspaces**

< > ***fewer measurements***

< > ***improved recovery perf.***

< > ***faster recovery***



CS Recovery

- **Iterative Hard Thresholding (IHT)**

[Nowak, Figueiredo; Kingsbury, Reeves; Daubechies, Defrise, De Mol; Blumensath, Davies; ...]

Given $y = \Phi x$, recover a sparse x

initialize: $\widehat{x}_0 = 0$, $r = y$, $i = 0$

iteration:

- $i \leftarrow i + 1$
- $b \leftarrow \widehat{x}_{i-1} + \Phi^T r$ — **update signal estimate**
- $\widehat{x}_i \leftarrow \text{thresh}(b, K)$ — **prune signal estimate** (best *K*-term approx)
- $r \leftarrow y - \Phi \widehat{x}_i$ — **update residual**

return: $\widehat{x} \leftarrow \widehat{x}_i$

Model-based CS Recovery

- **Iterative Model Thresholding**

[**VC**, Duarte, Hegde, Baraniuk; Baraniuk, **VC**, Duarte, Hegde]

Given $y = \Phi x$, recover a model sparse $x \in \mathcal{M}$

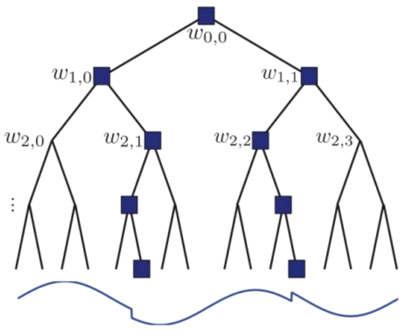
initialize: $\widehat{x}_0 = 0$, $r = y$, $i = 0$

iteration:

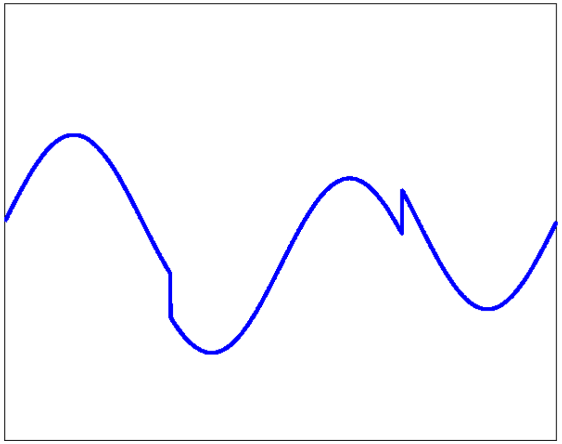
- $i \leftarrow i + 1$
- $b \leftarrow \widehat{x}_{i-1} + \Phi^T r$ — **update signal estimate**
- $\widehat{x}_i \leftarrow \mathcal{M}(b, K)$ — **prune signal estimate** (best *K*-term **model** approx)
- $r \leftarrow y - \Phi \widehat{x}_i$ — **update residual**

return: $\widehat{x} \leftarrow \widehat{x}_i$

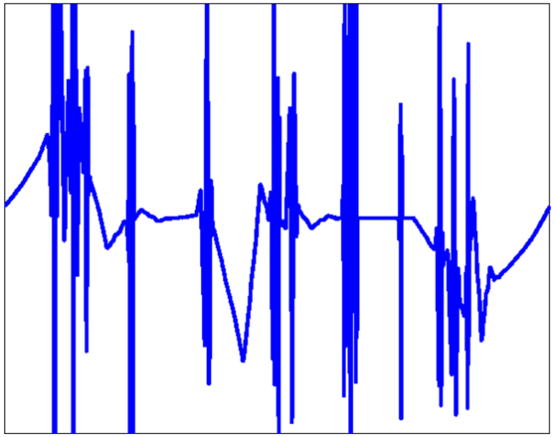
Tree-Sparse Signal Recovery



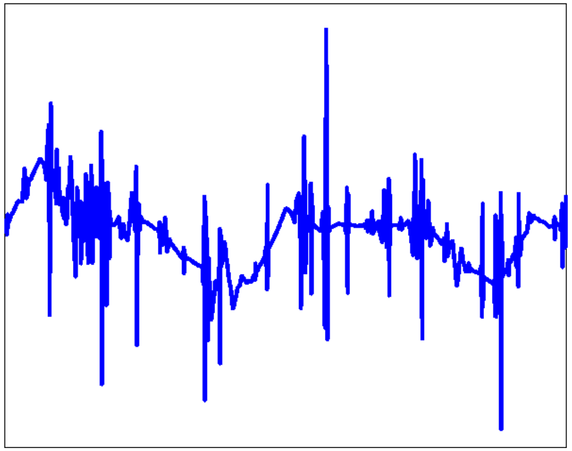
$N = 1024$
$M = 80$



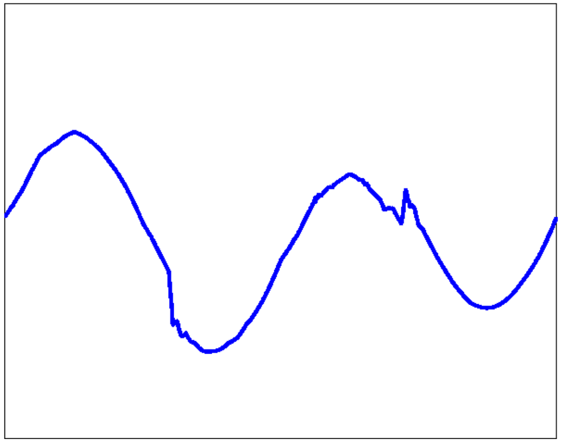
target signal



CoSaMP,
(MSE=1.12)



L1-minimization
(MSE=0.751)



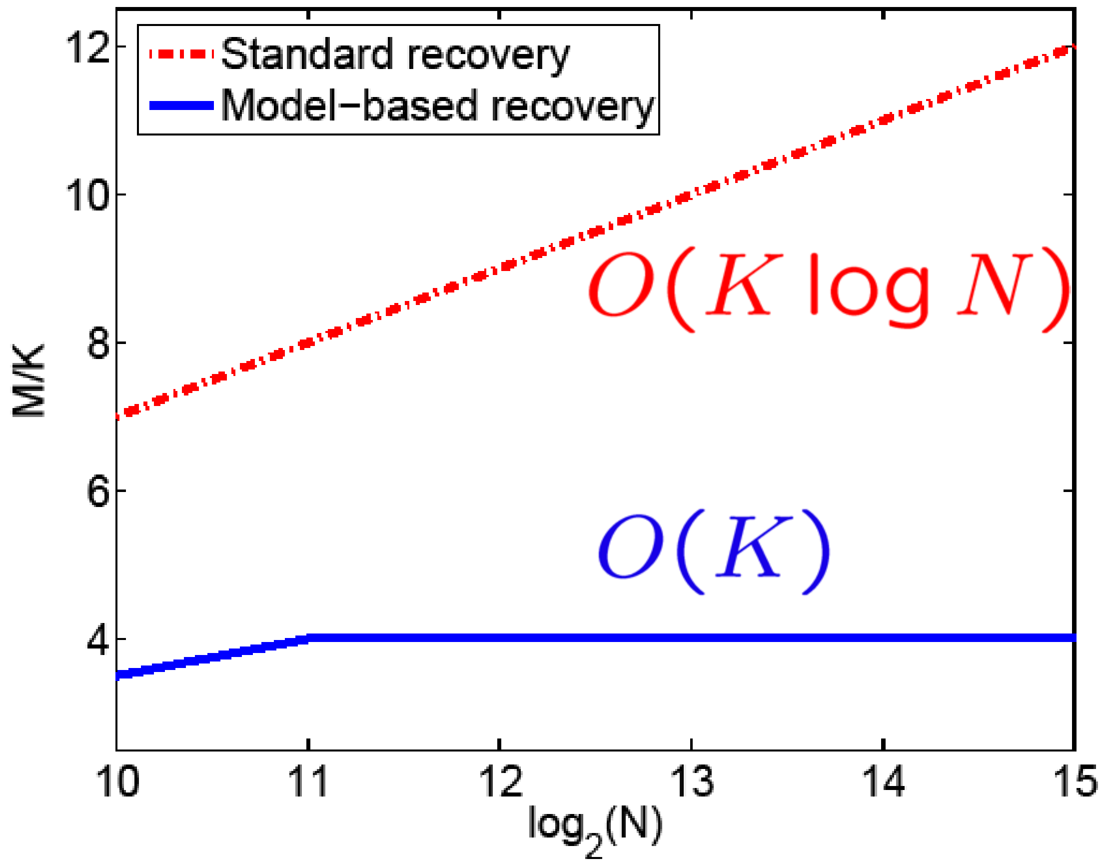
Tree-sparse CoSaMP
(MSE=0.037)

Tree-Sparse Signal Recovery

- Number samples for correct recovery with noise

- Piecewise cubic signals + wavelets

- Plot the number of samples to reach the noise level

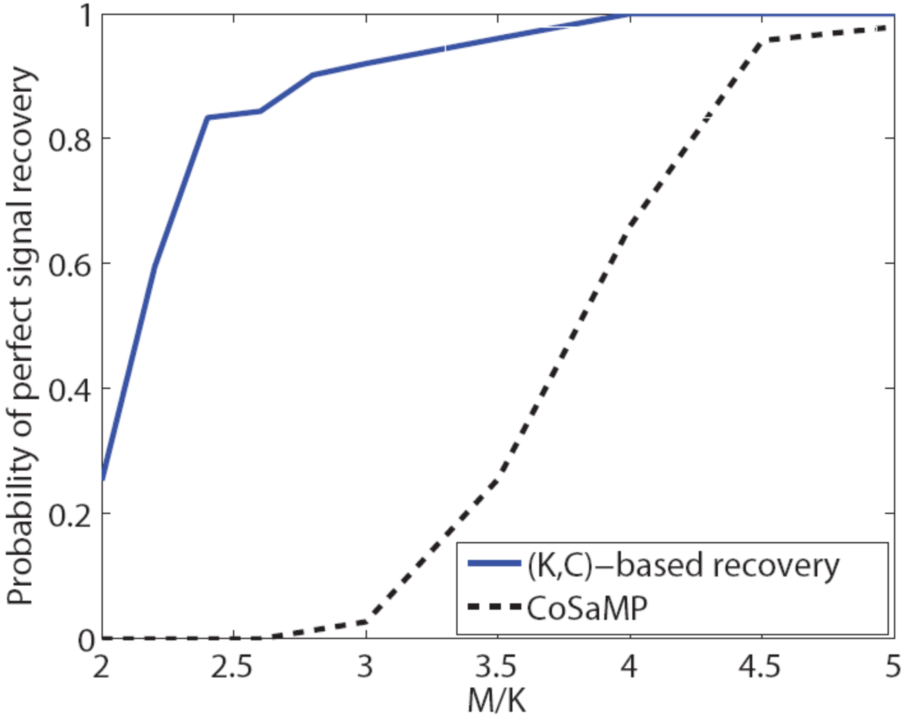


Clustered Sparsity

- **(*K*,*C*) sparse signals** (1-D)
  - *K*-sparse within at most *C* clusters

- For stable recovery $M = \mathcal{O}\left(K + C\log(N/C)\right)$

[**VC**, Indyk, Hedge, Baraniuk]



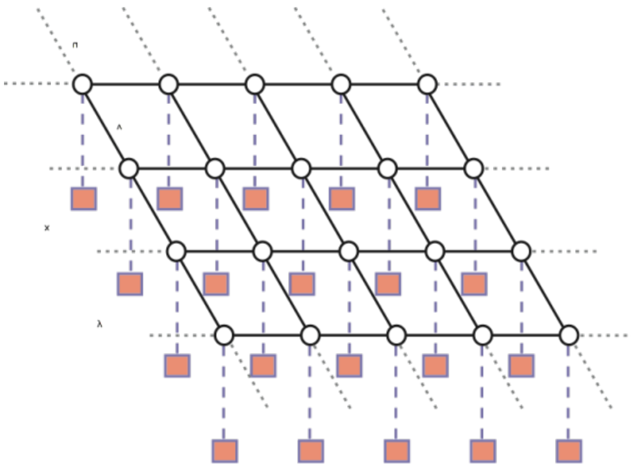
- Model approximation using **dynamic programming**

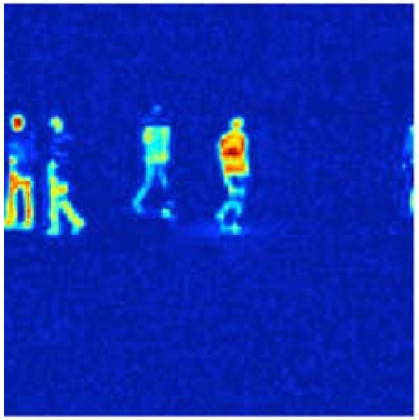
- Includes **block sparsity** as a special case

Clustered Sparsity

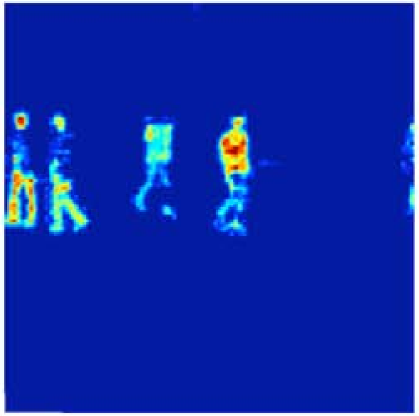
- Model clustering of significant pixels in space domain using **graphical model** (MRF)
- Ising model approximation via **graph cuts**

[**VC**, Duarte, Hedge, Baraniuk]

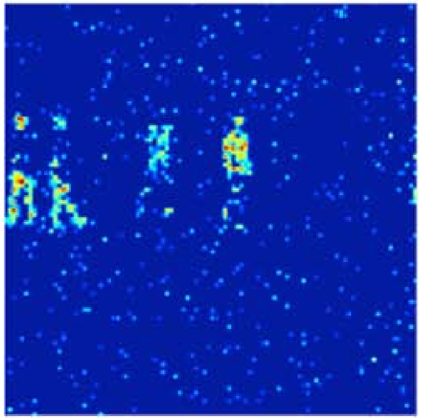




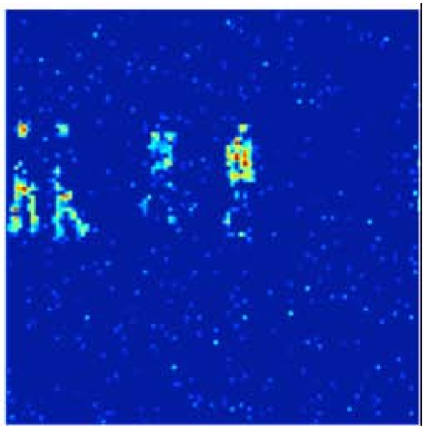
target



Ising-model recovery

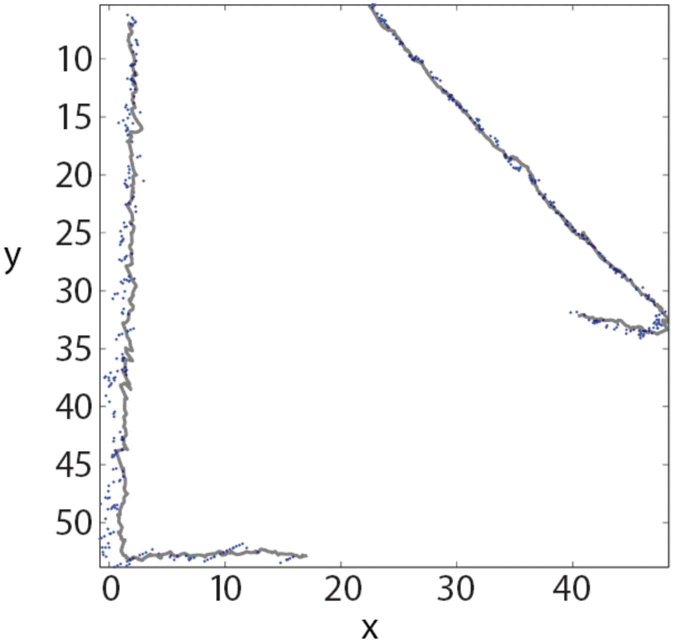


CoSaMP recovery



LP (FPC) recovery

Clustered Sparsity



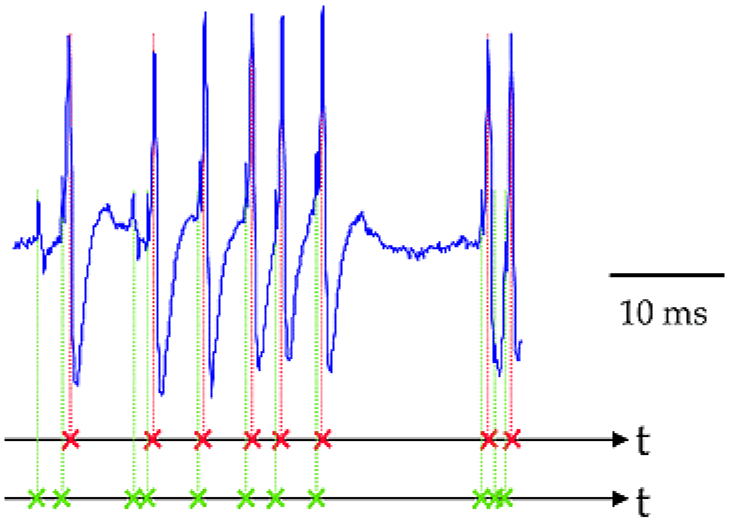
20% Compression

No performance loss in tracking

Neuronal Spike Trains

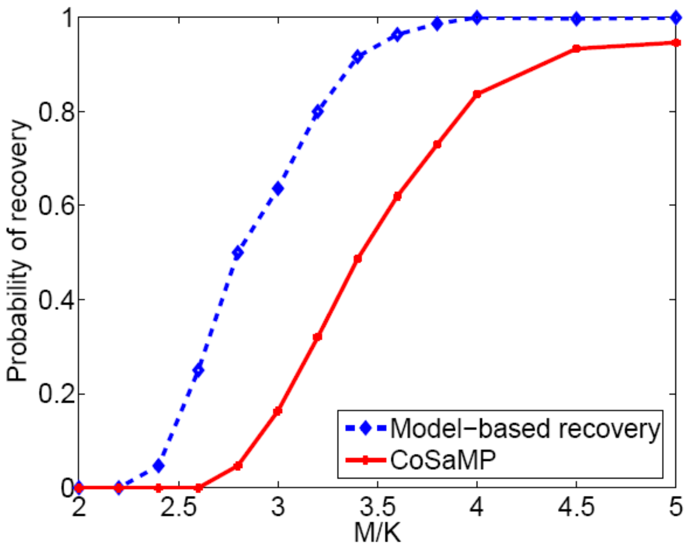
- Model the firing process of a single neuron via 1D Poisson process with spike trains
  - stable recovery

$$M = \mathcal{O}\left(K \log(N/K - \Delta)\right)$$



- Model approximation solution:
  - integer program

    efficient & provable solution
    due to
    total unimodularity of
    linear constraint
  - dynamic program



[Hegde, Duarte, **VC**] – best paper award

Performance of Recovery

- Using model-based IHT and CoSaMP

$$M = \mathcal{O}\left(\log|\mathcal{M}_K|\right)$$

$|\mathcal{M}_K|$: # of subspaces

- **Model-sparse signals**

[Baraniuk, **VC**, Duarte, Hegde]

$$\|x - \widehat{x}\|_{\ell_2} \leq C_1 \frac{\|x - x_{\mathcal{M}_K}\|_{\ell_1}}{K^{1/2}} + C_2\|n\|_2$$

CS recovery error — signal K-term model approx error — noise

- **Model-compressible signals**

w/restricted amplification property

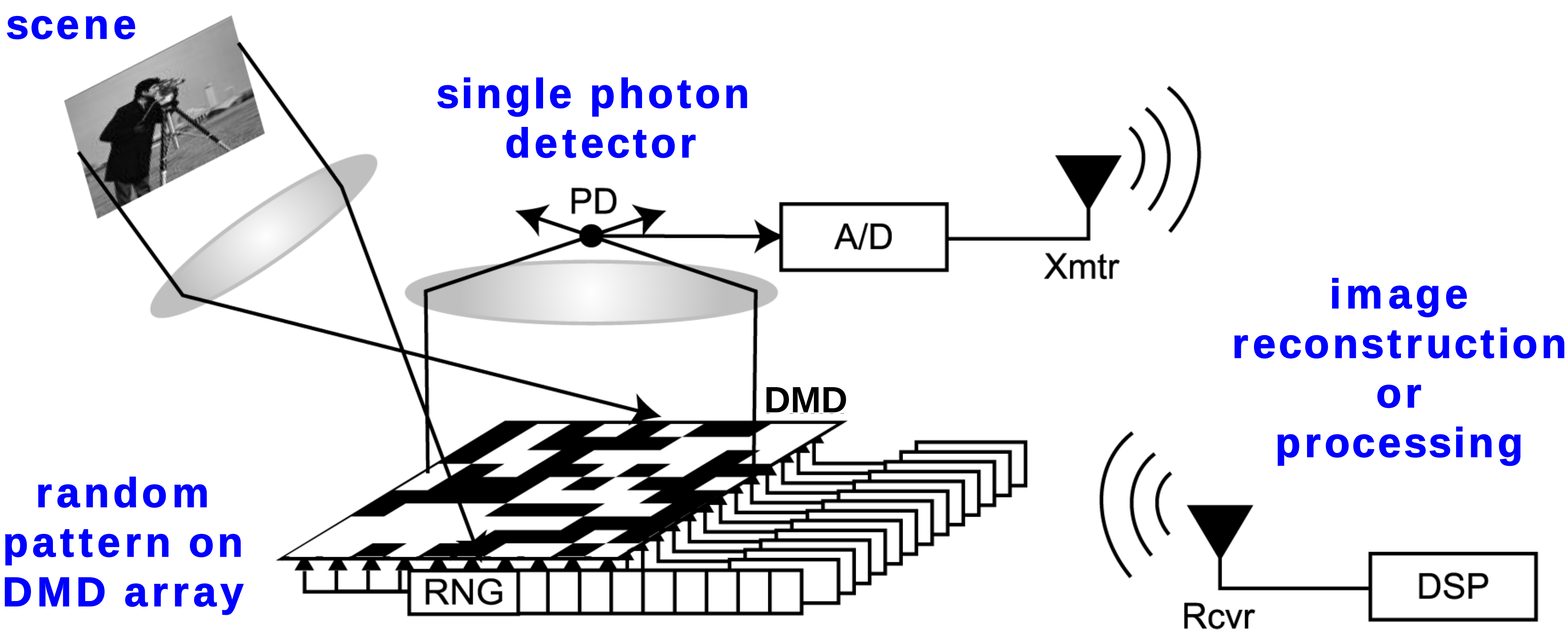
$$\|x - \widehat{x}\|_{\ell_2} \leq C_1 \log\left(\frac{N}{K}\right) \frac{\|x - x_{\mathcal{M}_K}\|_{\ell_1}}{K^{1/2}} + C_2\|n\|_2$$

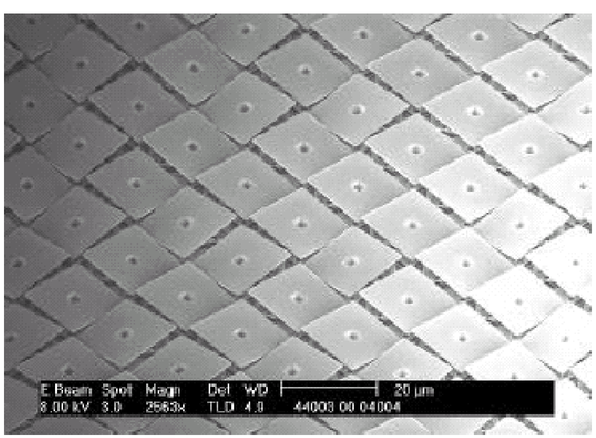
CS recovery error — signal K-term model approx error — noise

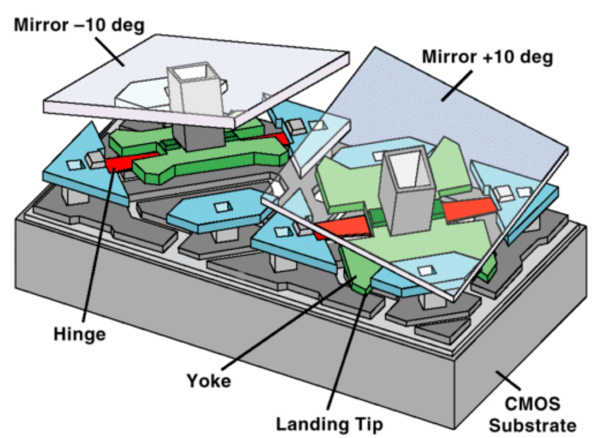
Compressive Sensing
In Action

Cameras

“Single-Pixel” CS Camera

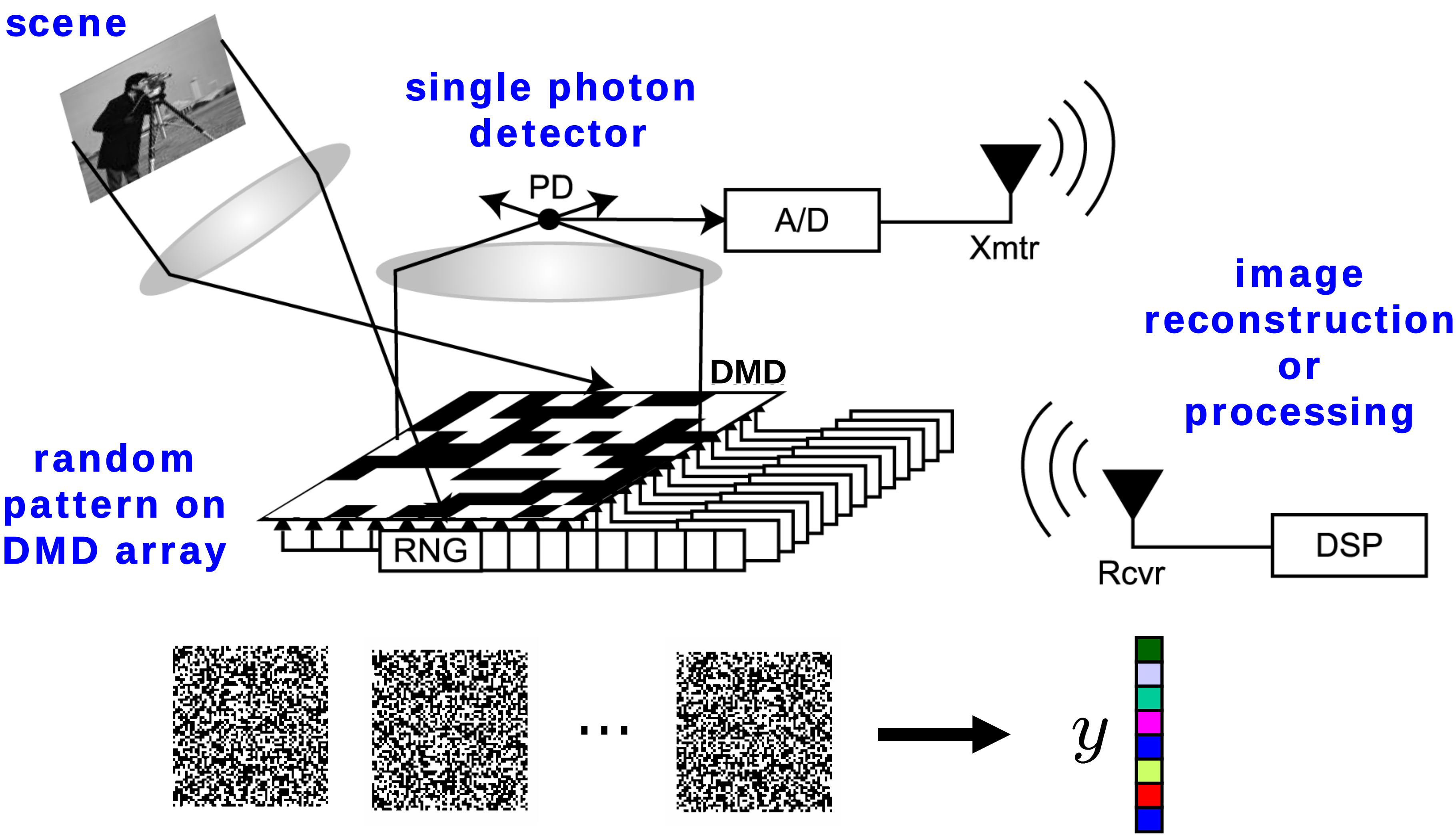






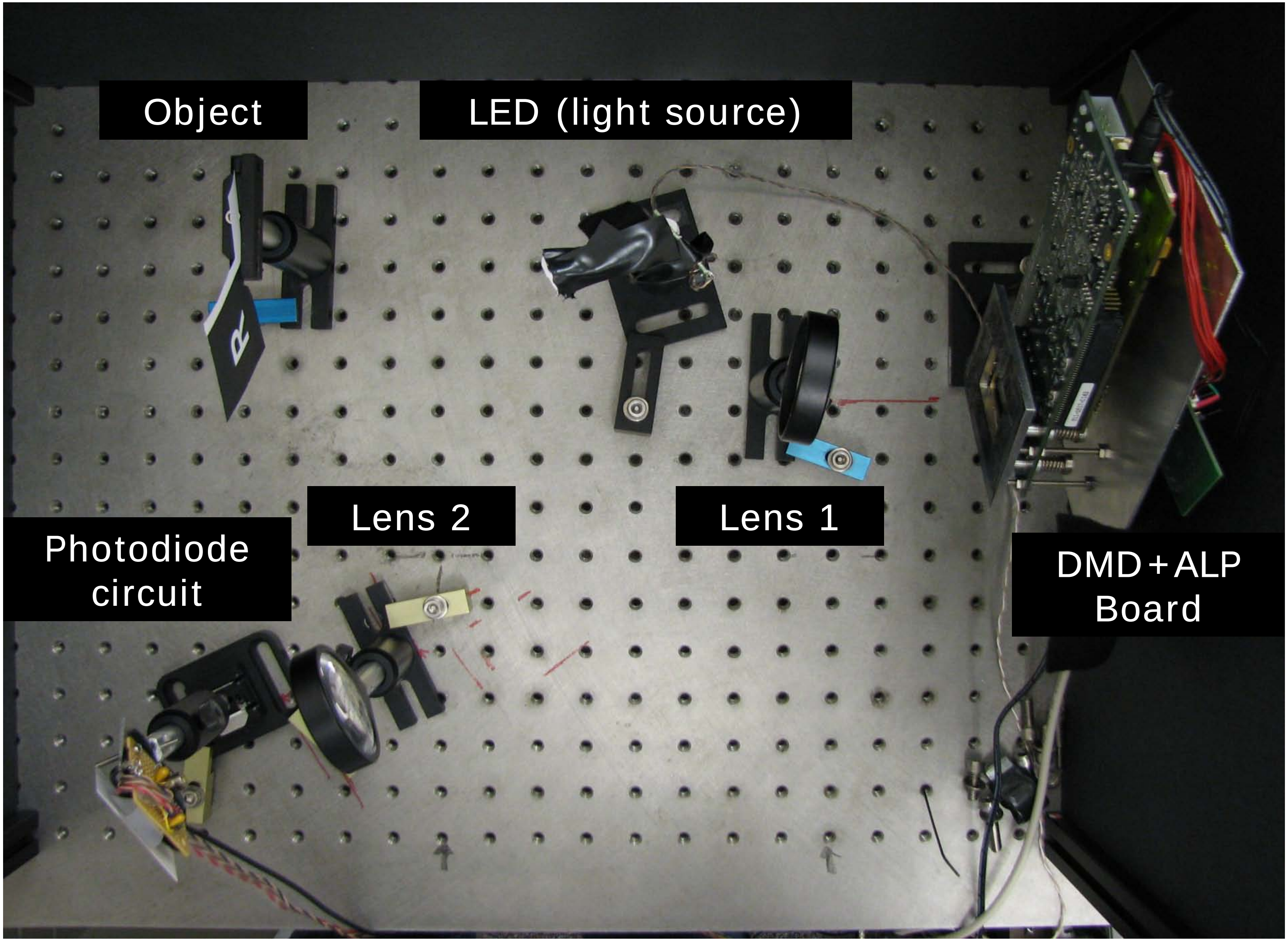
w/ Kevin Kelly

“Single-Pixel” CS Camera

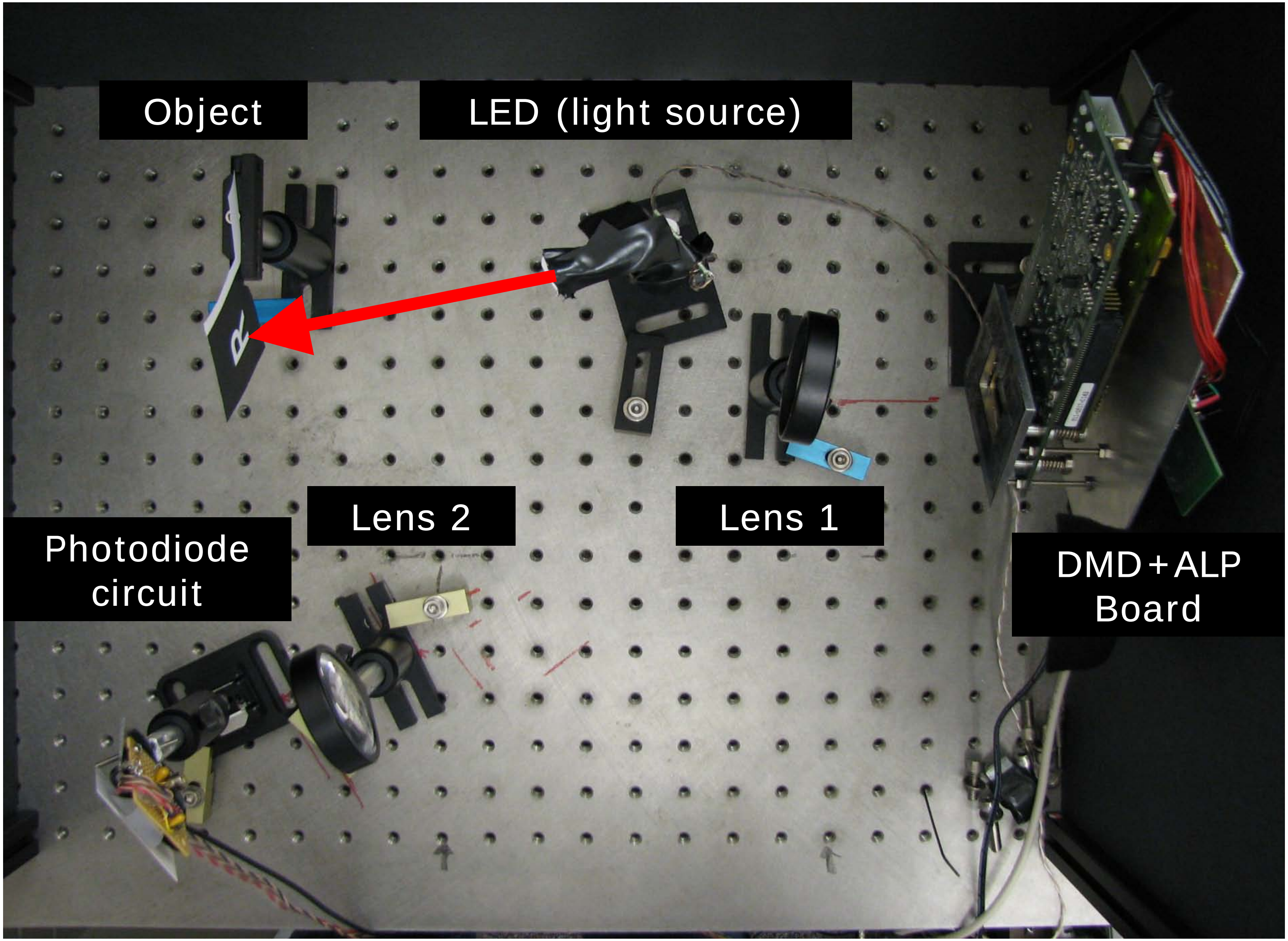


- Flip mirror array *M* times to acquire *M* measurements
- Sparsity-based (linear programming) recovery

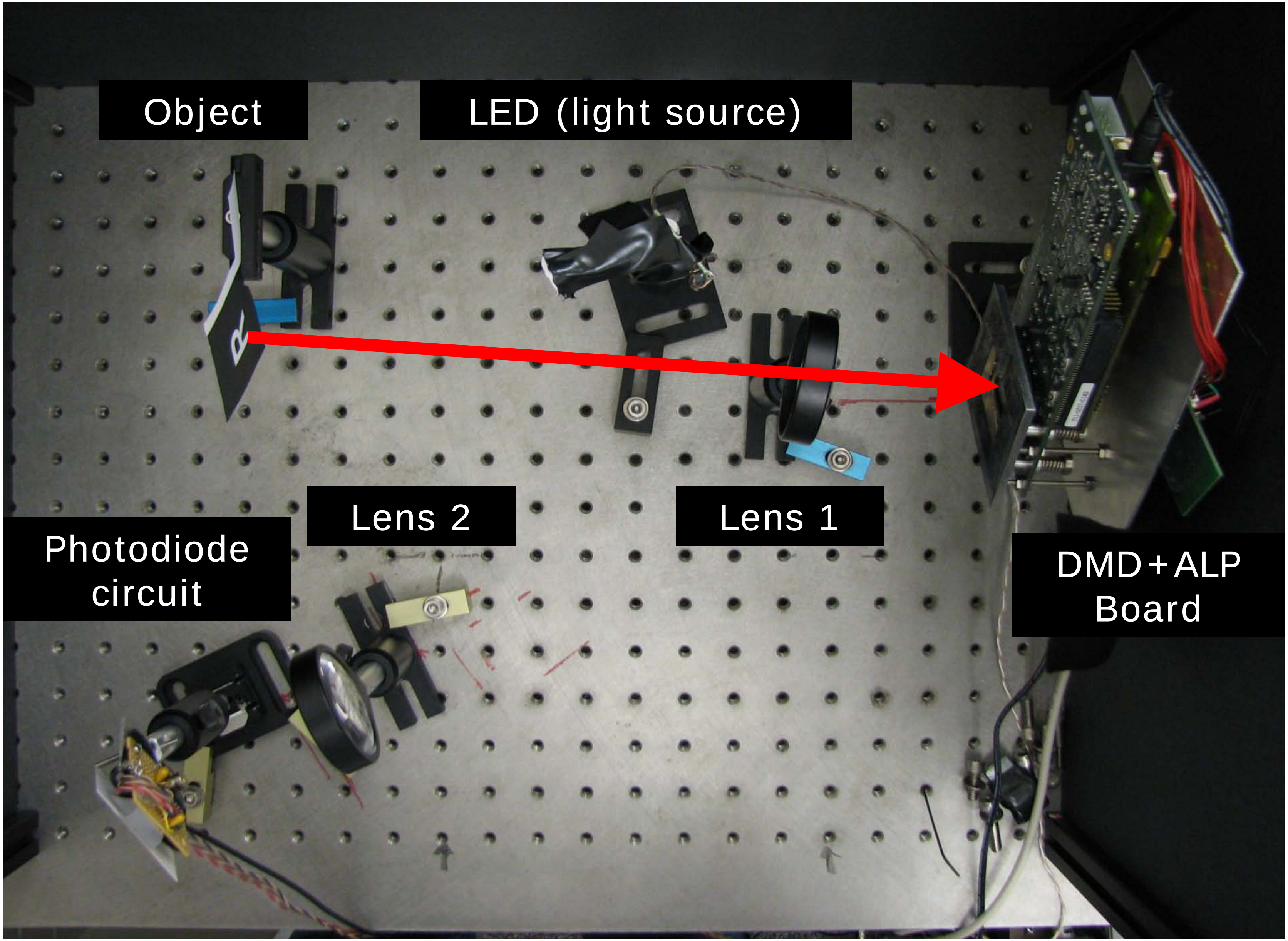
Single Pixel Camera



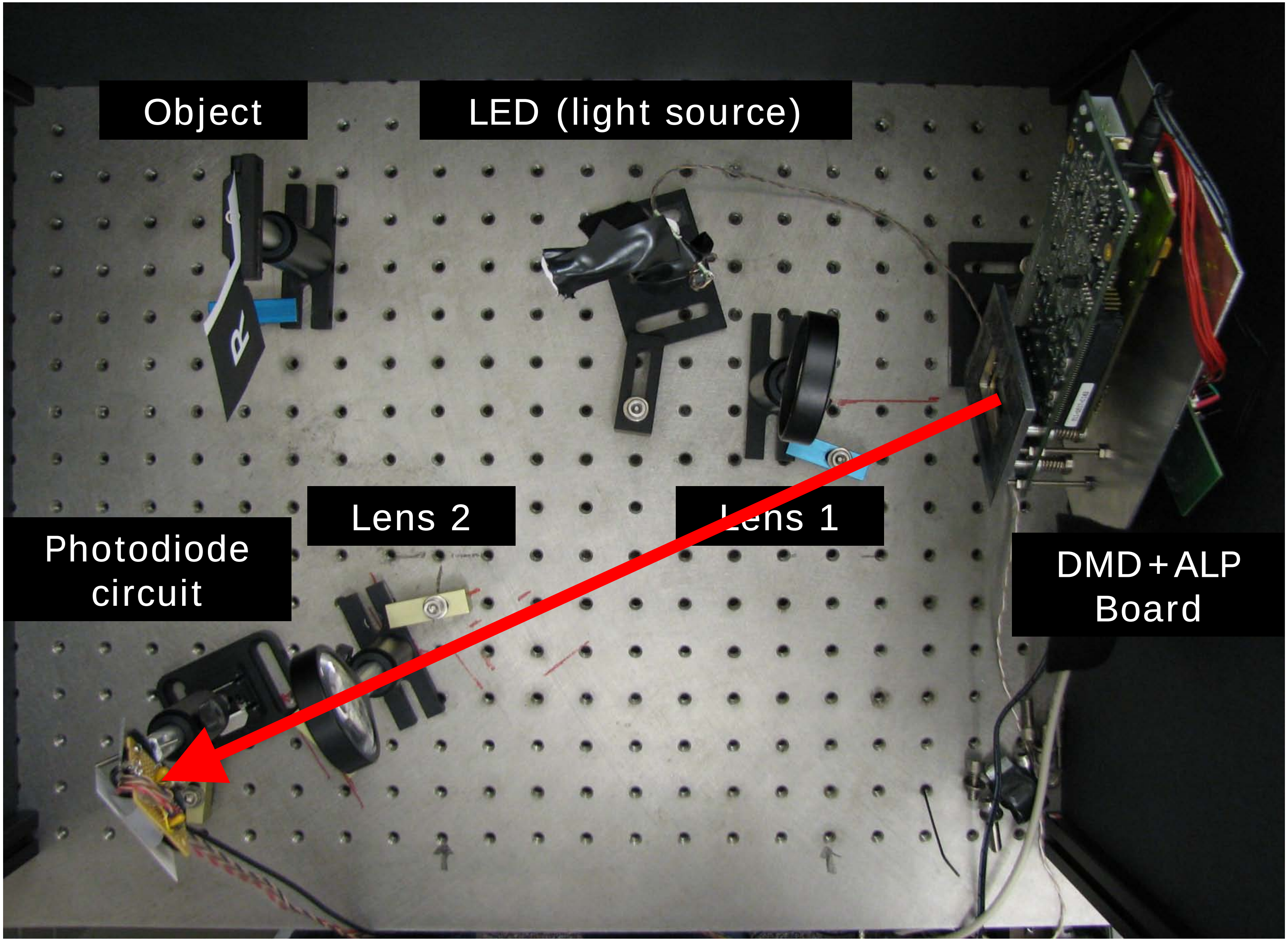
Single Pixel Camera



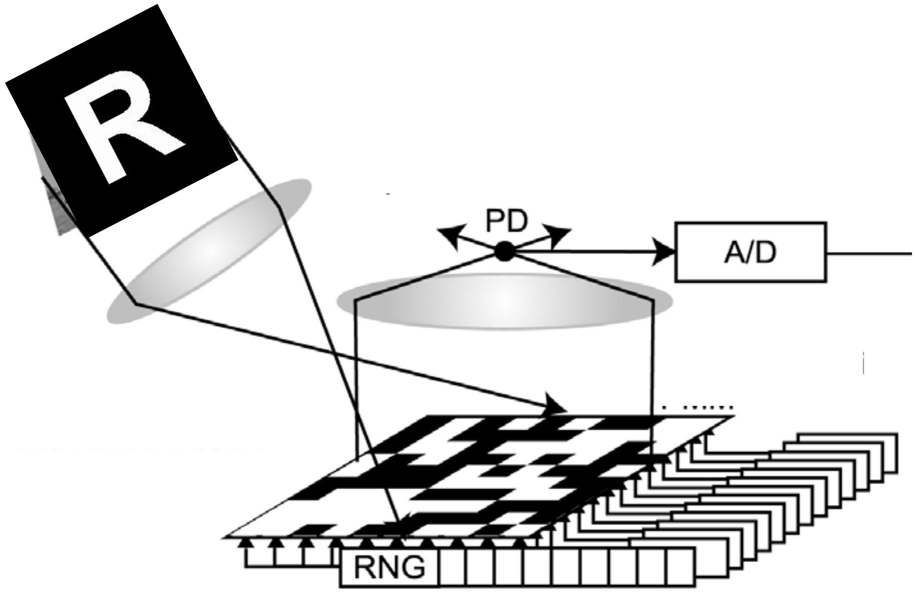
Single Pixel Camera



Single Pixel Camera

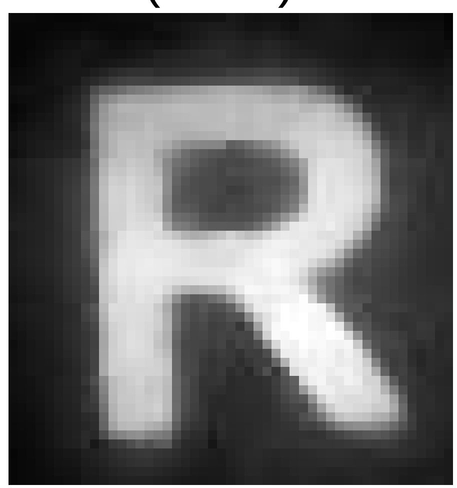


First Image Acquisition



target
65536 pixels

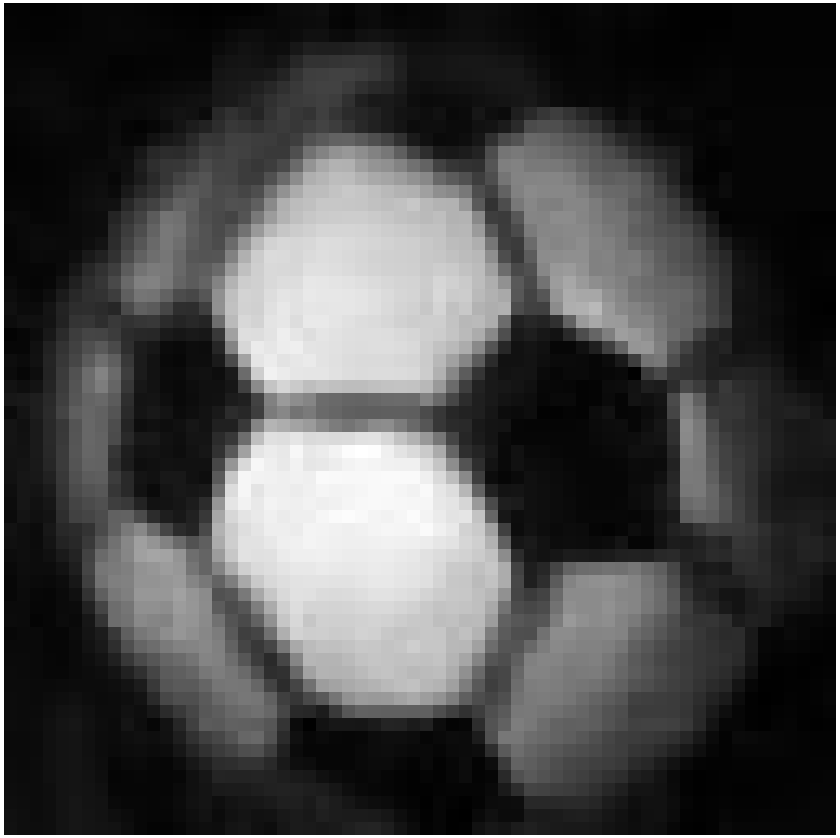
11000 measurements
(16%)



1300 measurements
(2%)

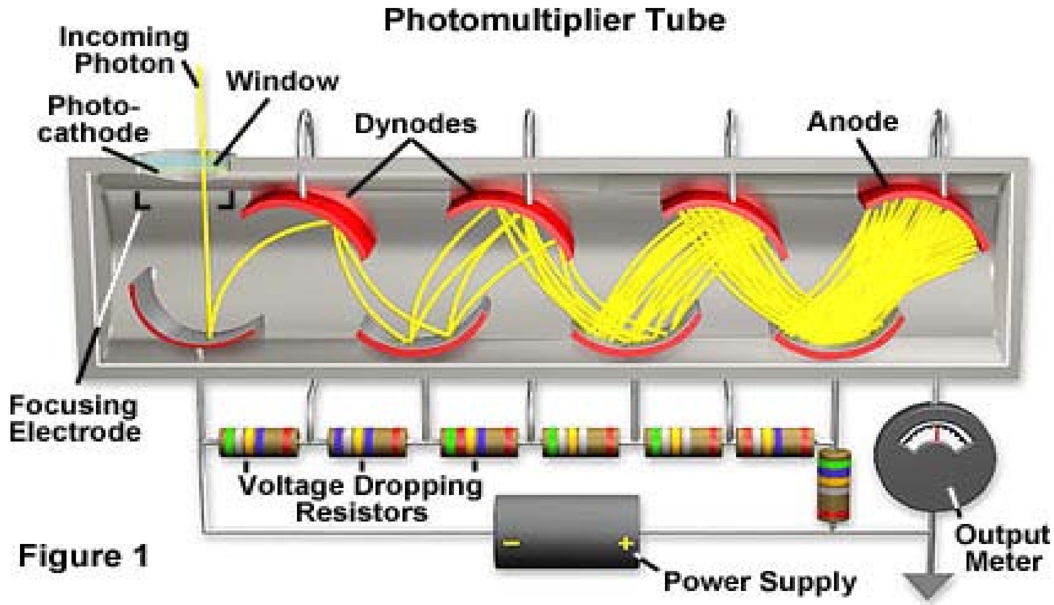
Second Image Acquisition

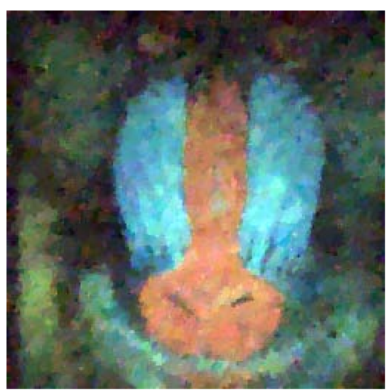
4096
pixels



500
random measurements

CS Low-Light Imaging with PMT



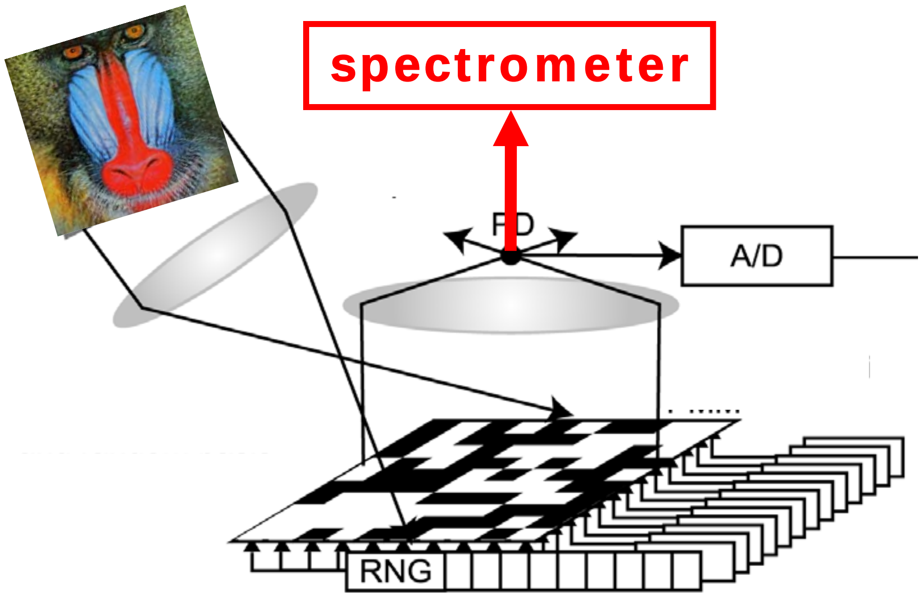


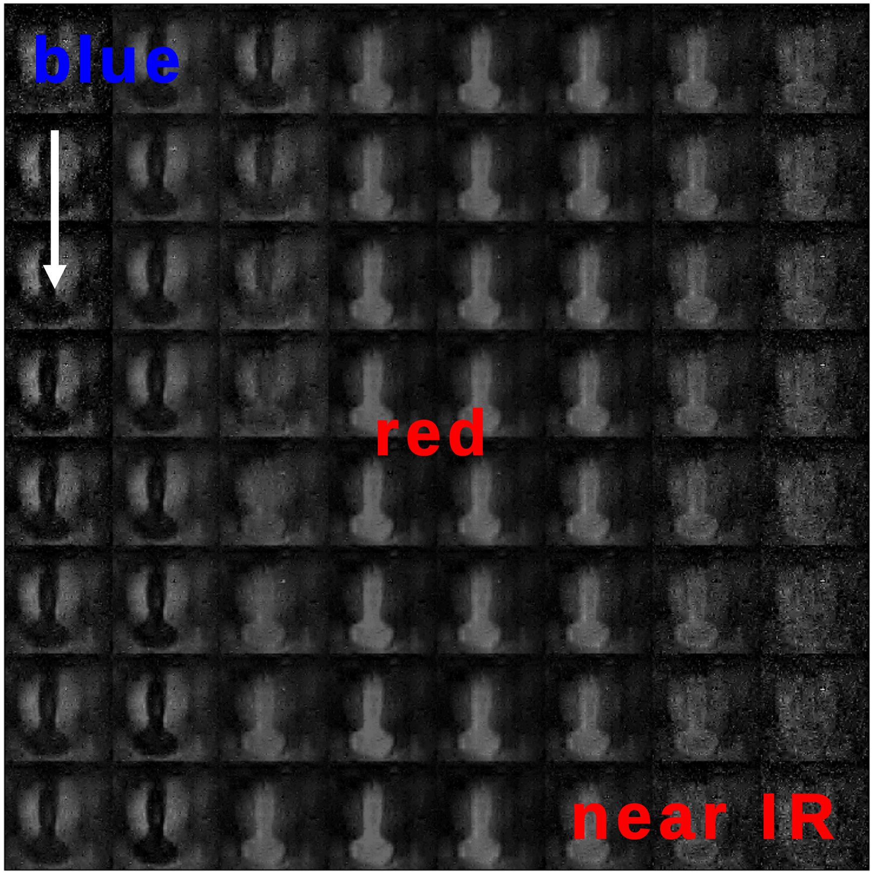
true color low-light imaging

256 x 256 image with 10:1 compression

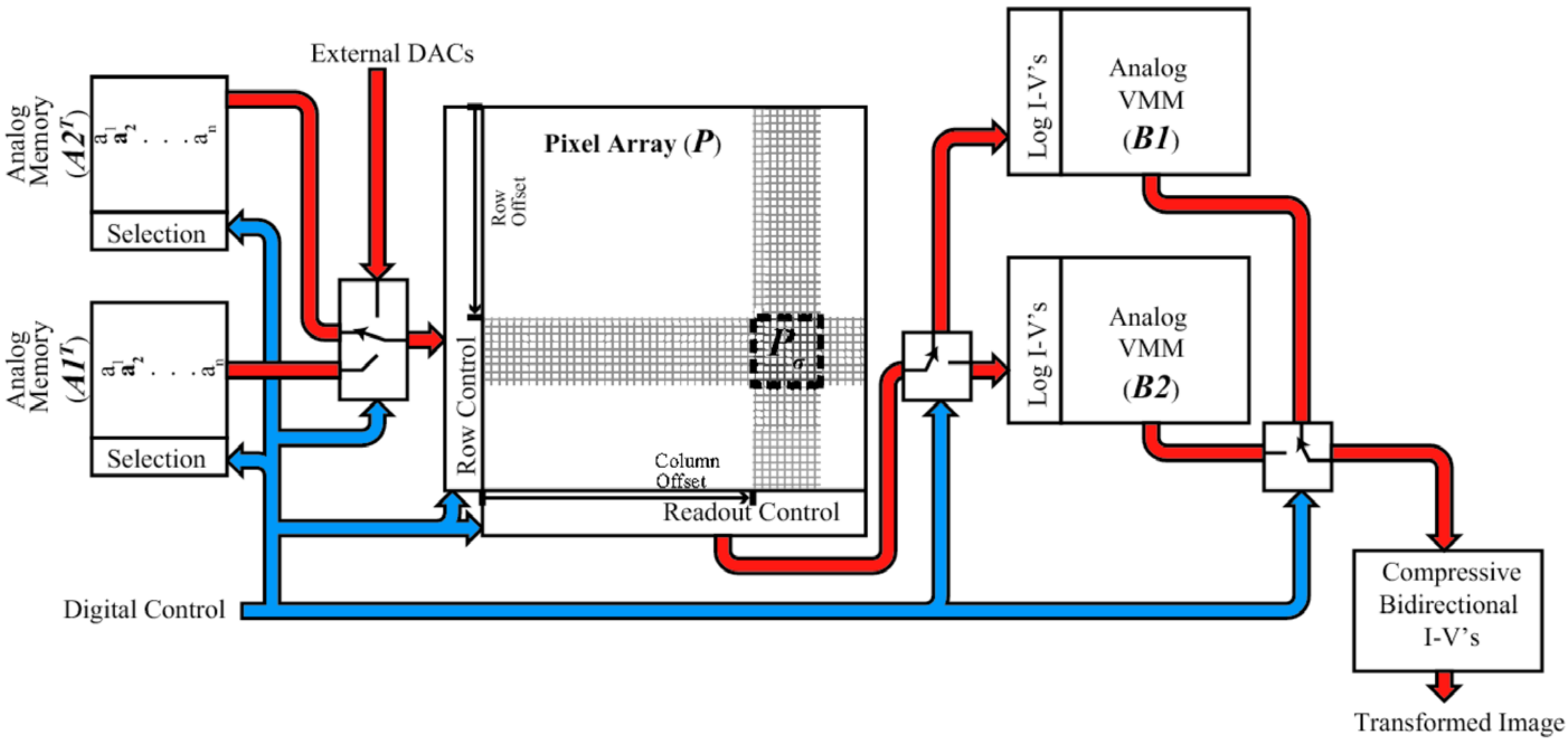
[Nature Photonics, April 2007]

Hyperspectral Imaging



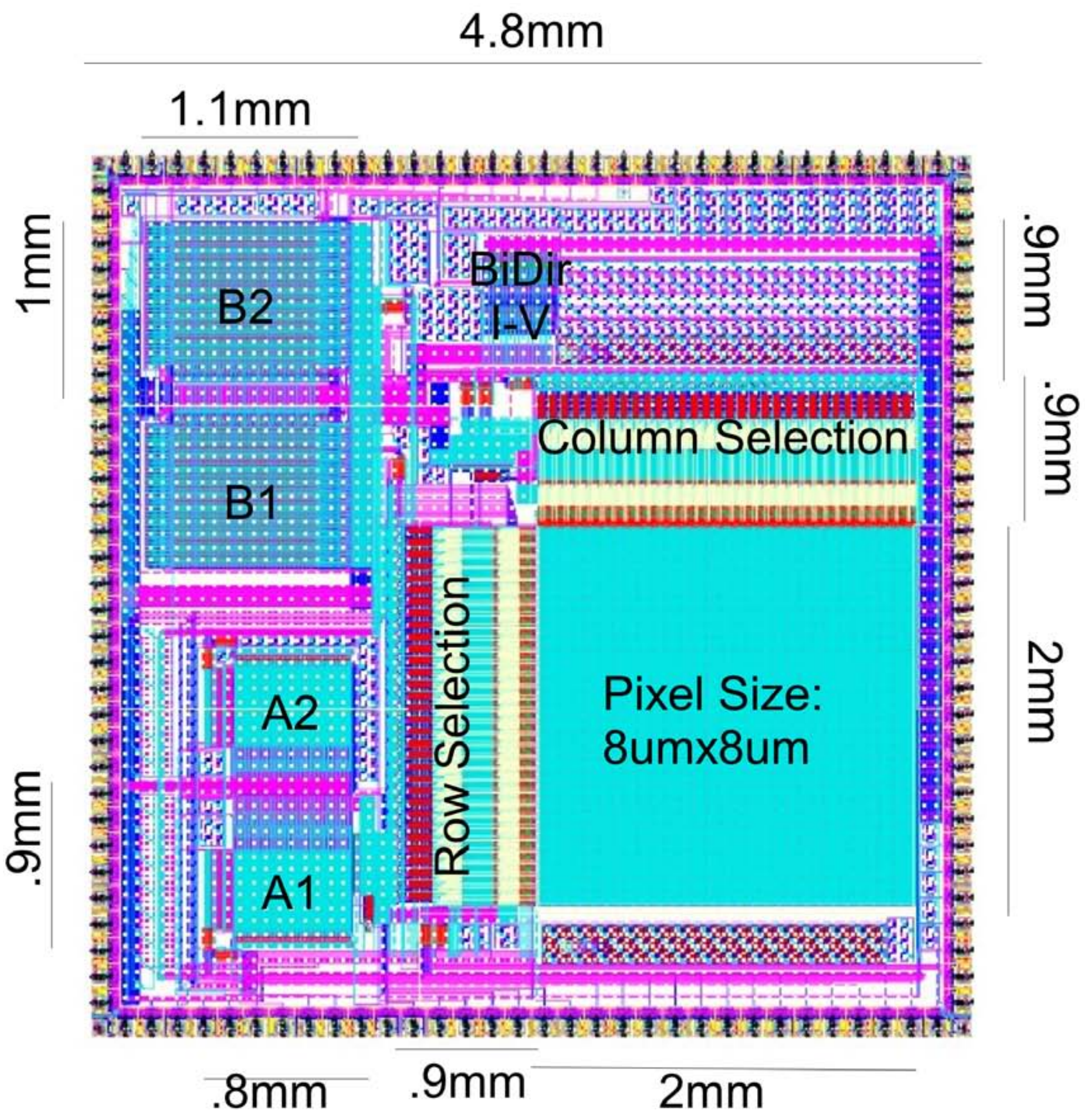


Georgia Tech Analog Imager

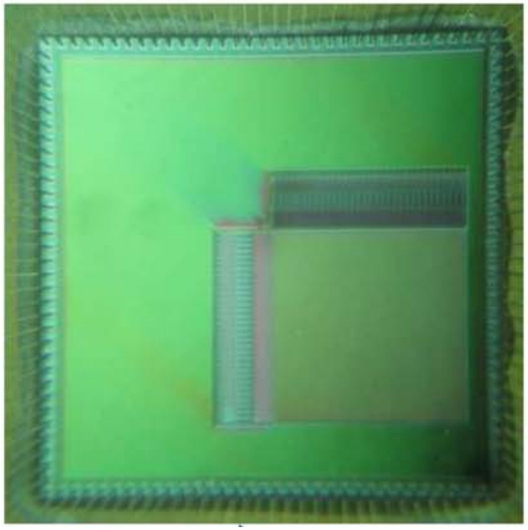


- Robucci and Hasler 07
- Transforms image *in analog*,
  reads out transform coefficients

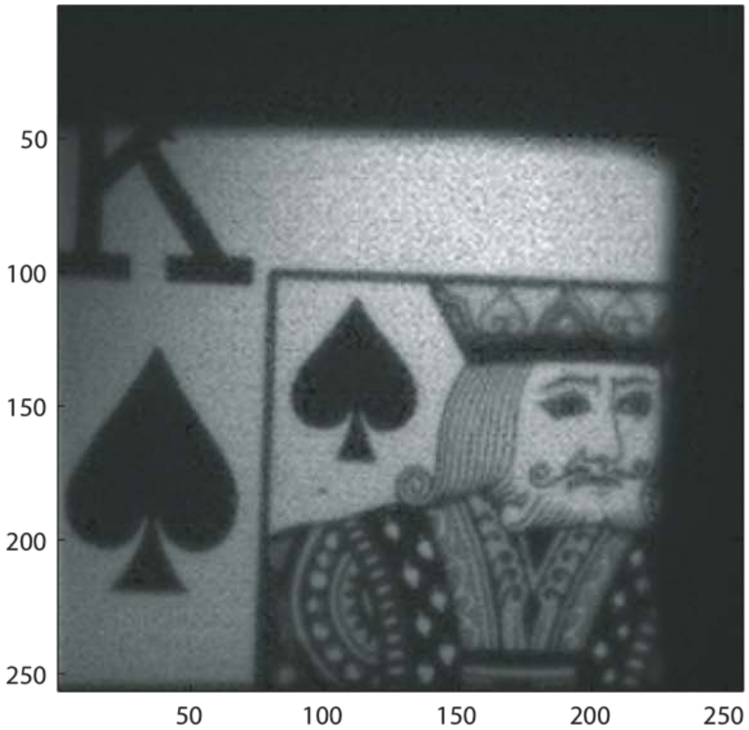
Georgia Tech Analog Imager



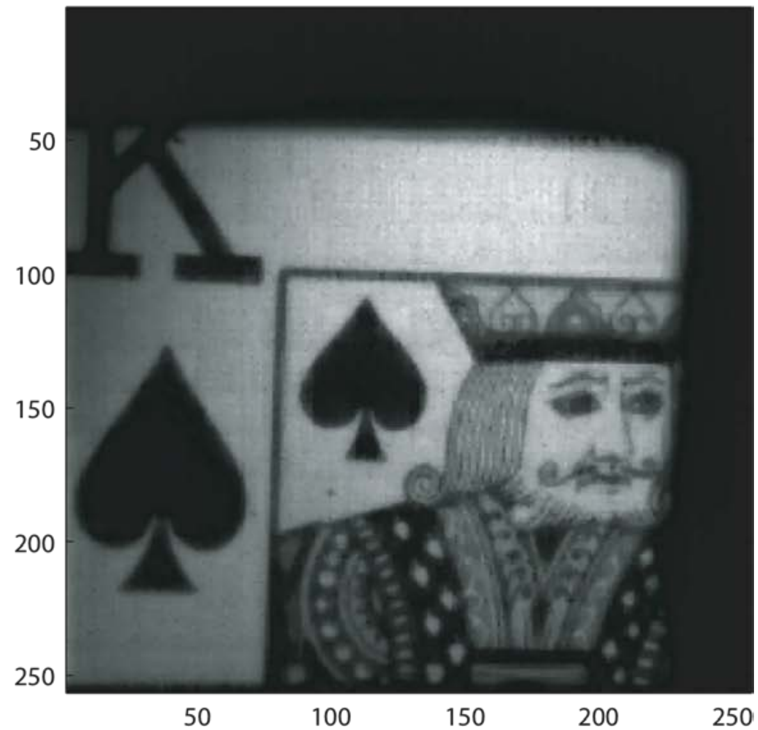
.35um CMOS process



Compressive Sensing Acquisition



10k DCT coefficients



10k random measurements

(Robucci et al 09)

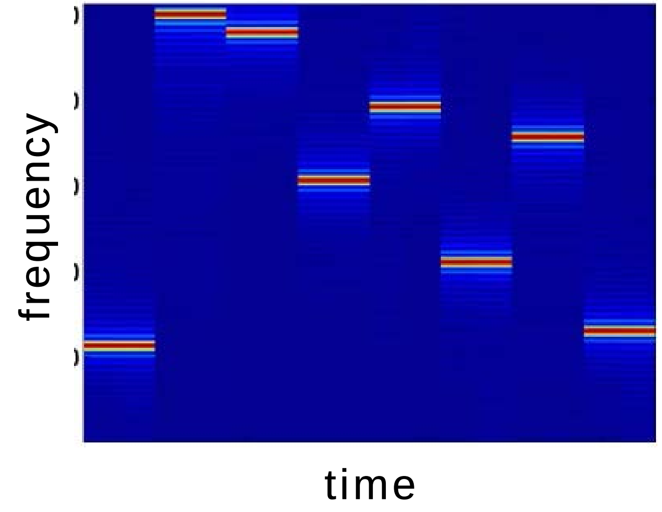
Compressive Sensing
In Action

A/D Converters

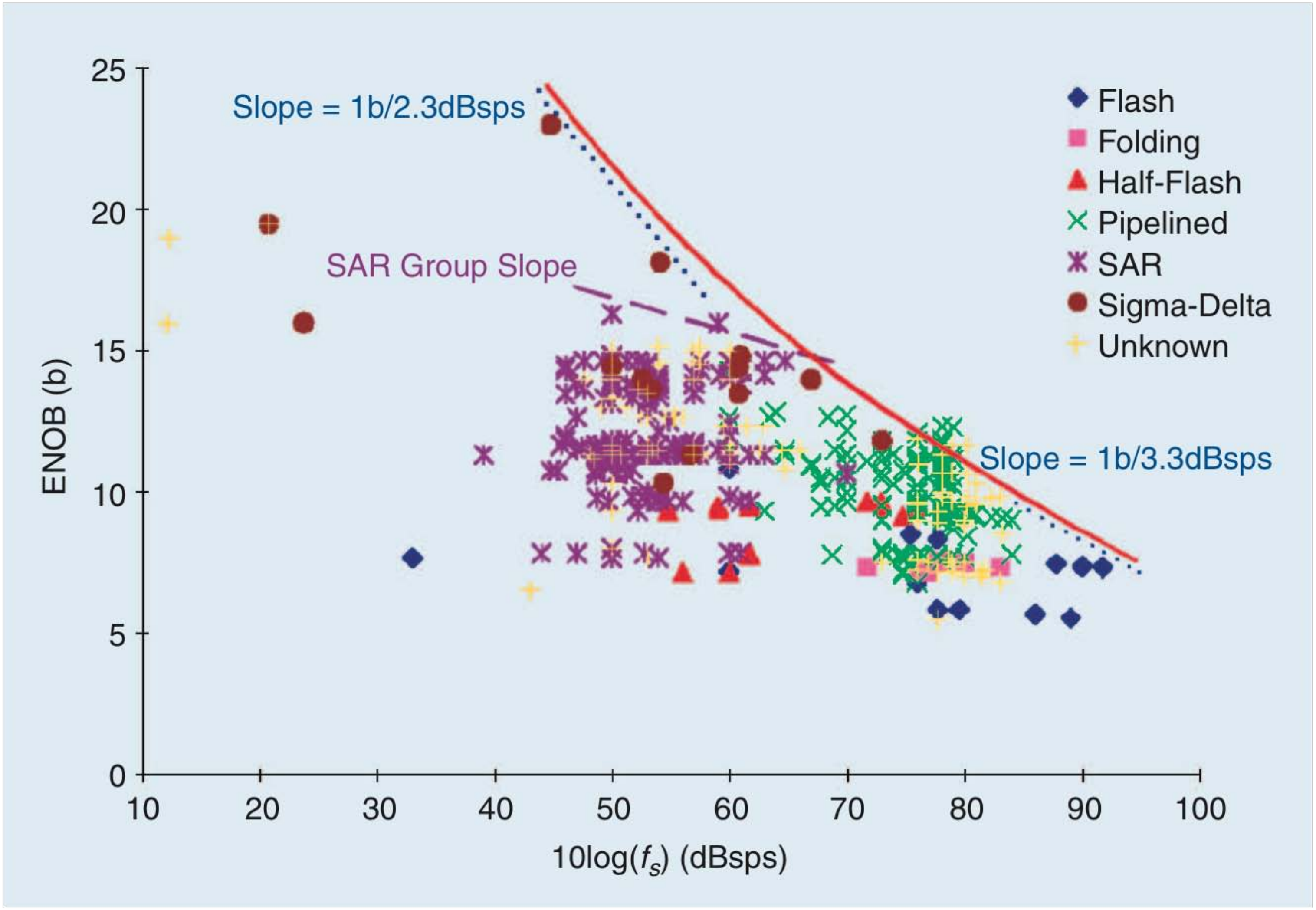
Analog-to-Digital Conversion

- Nyquist rate limits reach of today's ADCs
- "Moore's Law" for ADCs:
  - technology Figure of Merit incorporating sampling rate and dynamic range doubles every **6-8 years**
- DARPA Analog-to-Information (A2I) program
  - wideband signals have high Nyquist rate but are often sparse/compressible
  - develop new ADC technologies to exploit
  - new tradeoffs among Nyquist rate, sampling rate, dynamic range, …

frequency hopper spectrogram



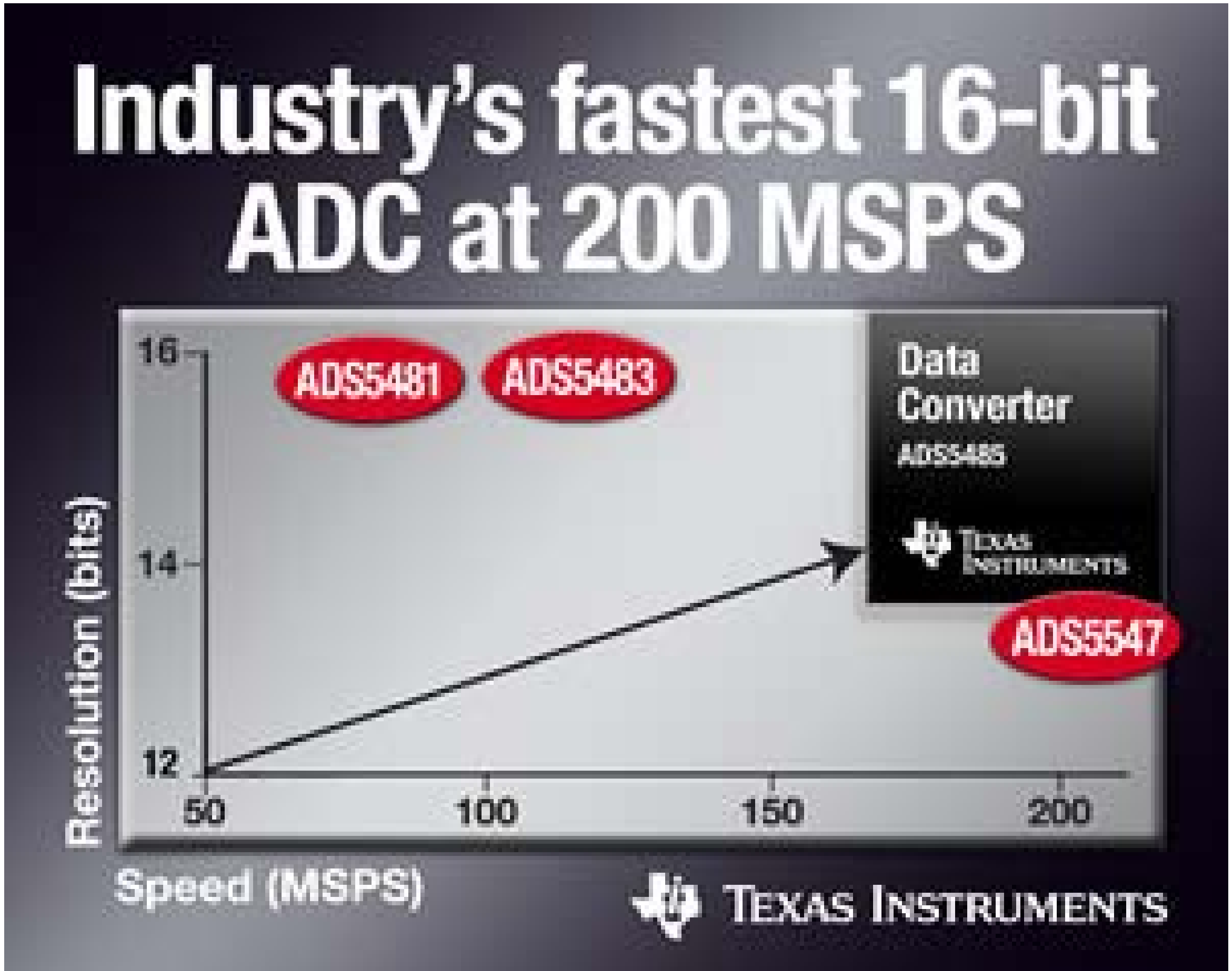
ADC State of the Art (2005)



The bad news starts at 1 GHz...

ADC State of the Art

From 2008…



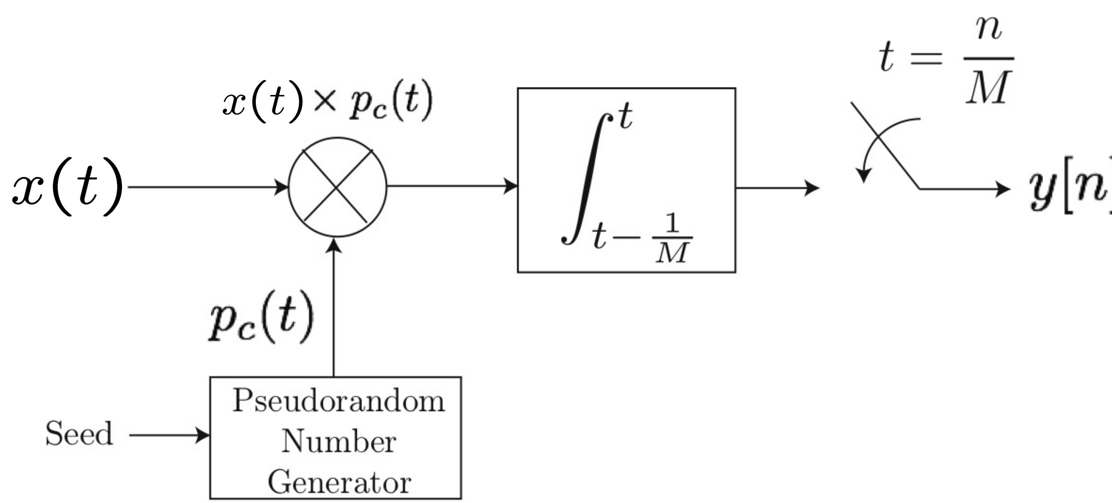
Analog-to-*Information* Conversion

- Sample near signal's (low) "information rate" rather than its (high) Nyquist rate

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

- M: A2I sampling rate
- K: number of tones / window
- N: Nyquist bandwidth

- Practical hardware: randomized demodulator (CDMA receiver)

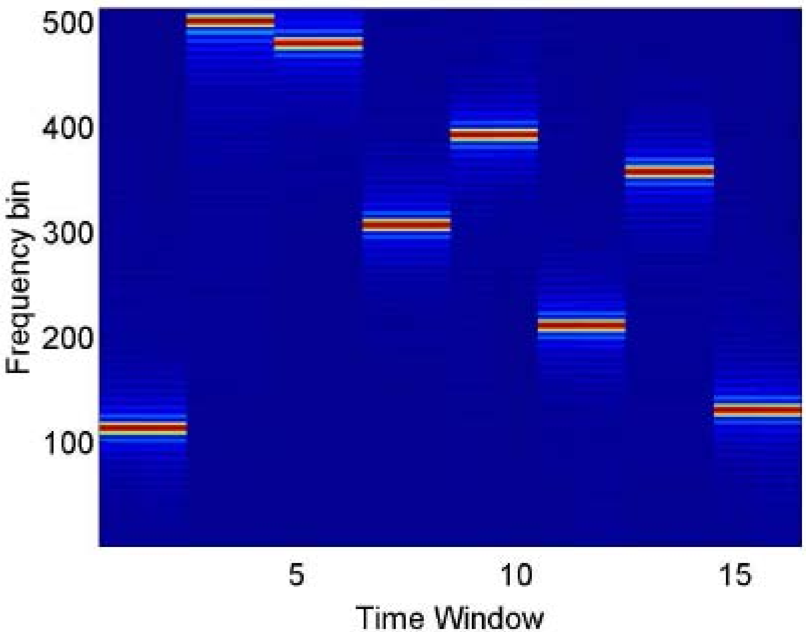


Example: Frequency Hopper

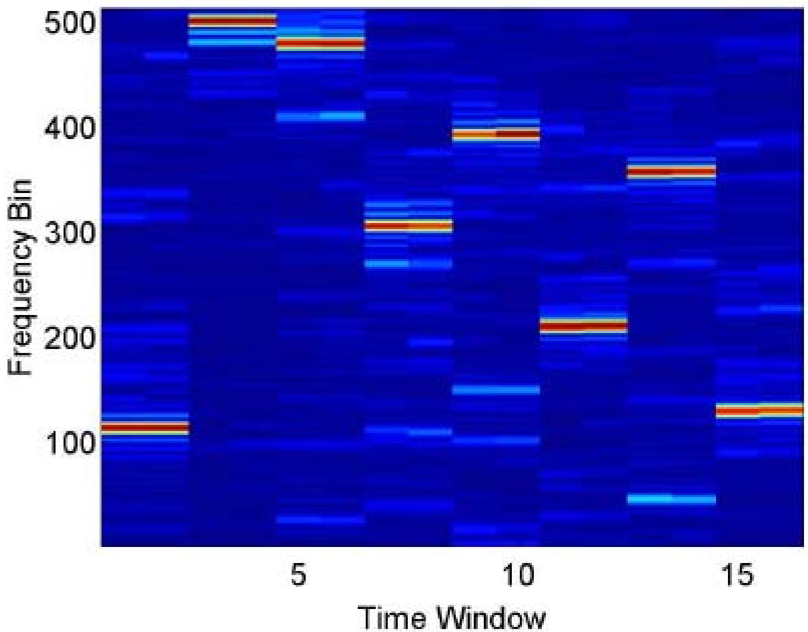
Nyquist rate sampling

20x sub-Nyquist sampling

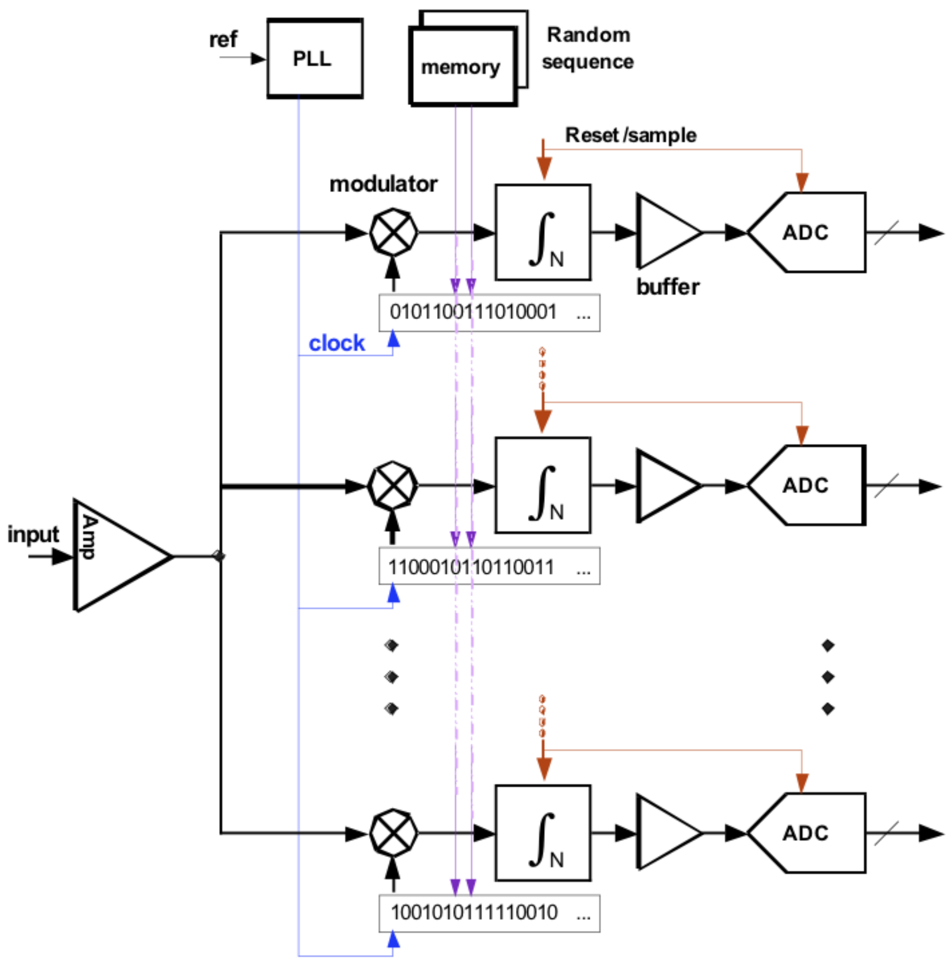
spectrogram



sparsogram



Multichannel Random Demodulation



- Random demodulator being built at part of DARPA A2I program (Emami, Hoyos, Massoud)
- Multiple (8) channels, operating with different mixing sequences
- Effective BW/chan = 2.5 GHz Sample rate/chan = 50 MHz
- Applications: radar pulse detection, communications surveillance, geolocation

Compressive Sensing
In Action

Data Processing

Information Scalability

- Many applications involve signal ***inference*** and not ***reconstruction***

detection < classification < estimation < reconstruction

fairly
computationally
intense

Information Scalability

- Many applications involve signal *inference* and not *reconstruction*

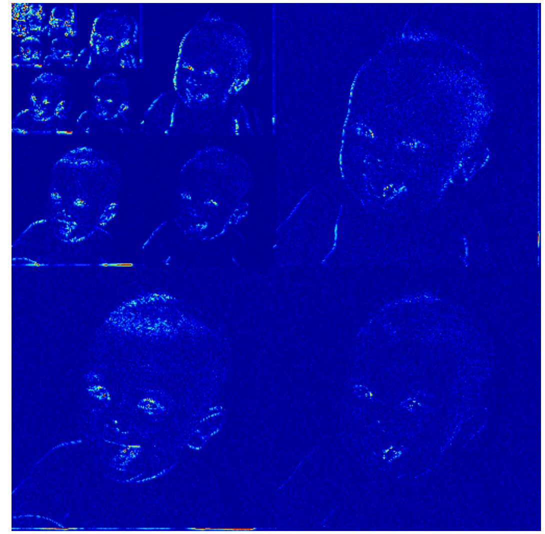
  detection < classification < estimation < reconstruction

- **Good news**: CS supports efficient learning, inference, processing directly on compressive measurements

- **Random projections ~ sufficient statistics** for signals with concise geometrical structure

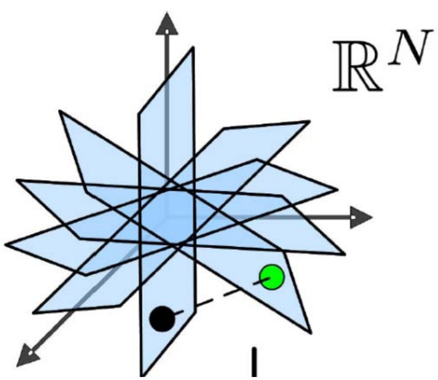
Low-dimensional signal models

N
pixels

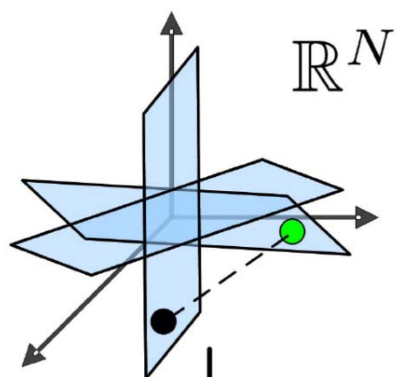


$K \ll N$
large
wavelet
coefficients

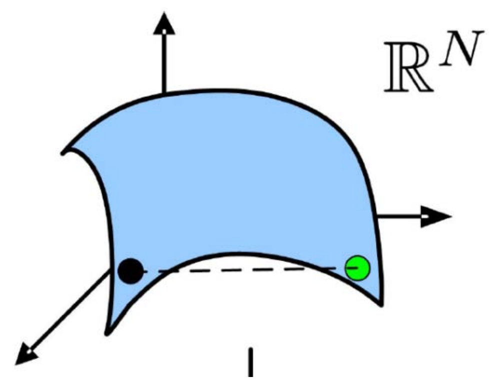
(blue = 0)



sparse
signals



structured
sparse signals

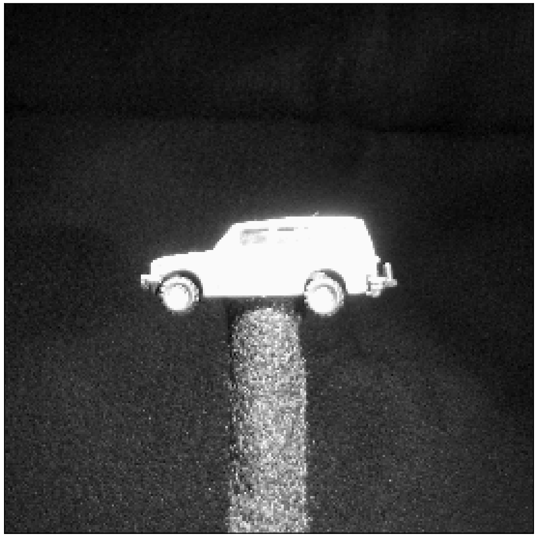


parameter
manifolds

Matched Filter

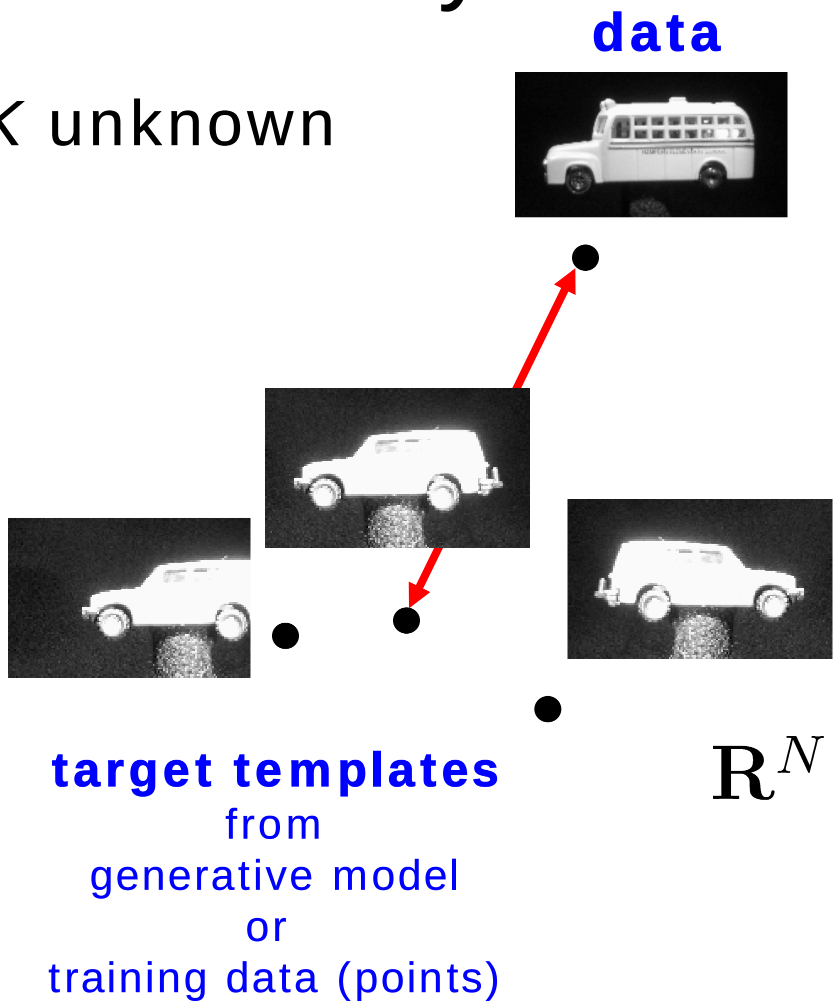
- Detection/classification with *K* unknown **articulation parameters**
  - Ex: position and pose of a vehicle in an image
  - Ex: time delay of a radar signal return
- **Matched filter**: joint parameter estimation and detection/classification
  - compute sufficient statistic for each potential target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





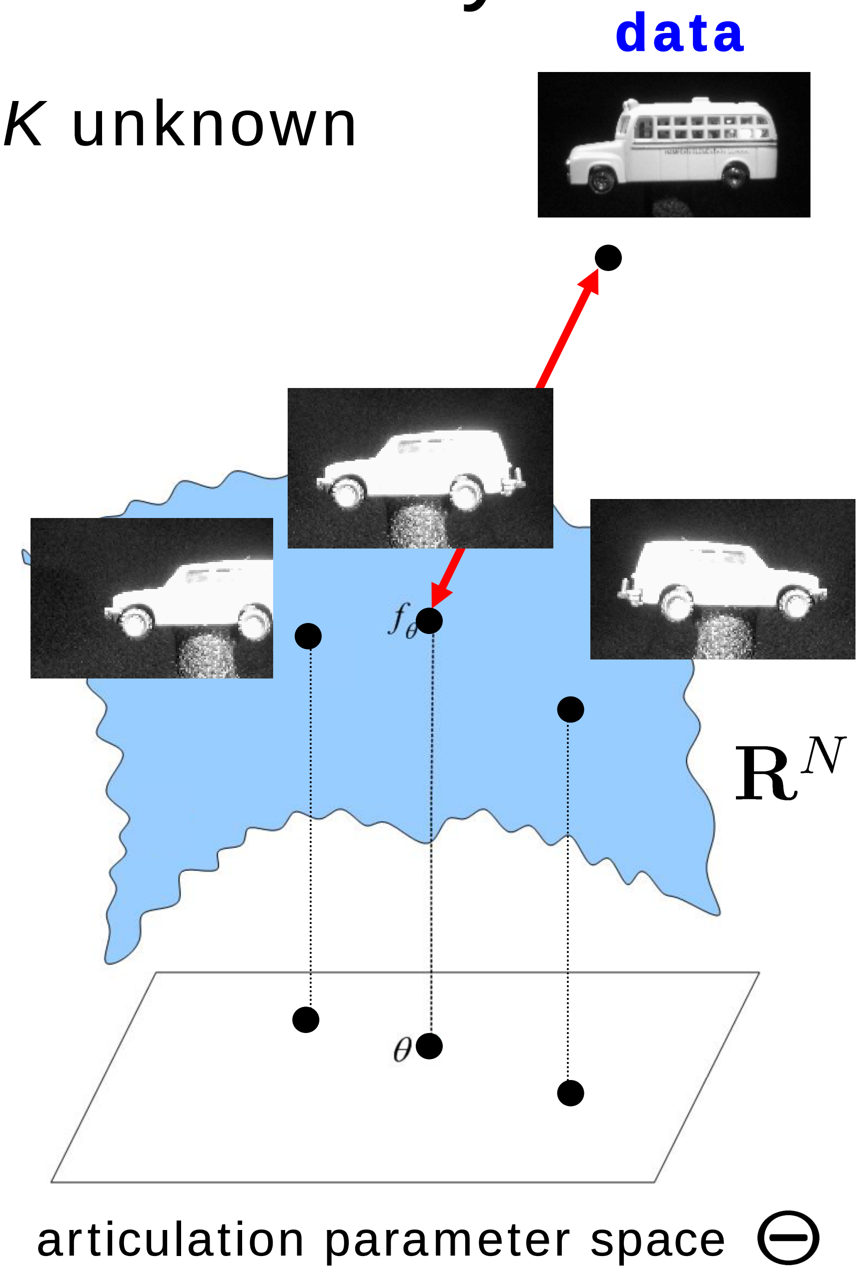
Matched Filter Geometry

- 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**



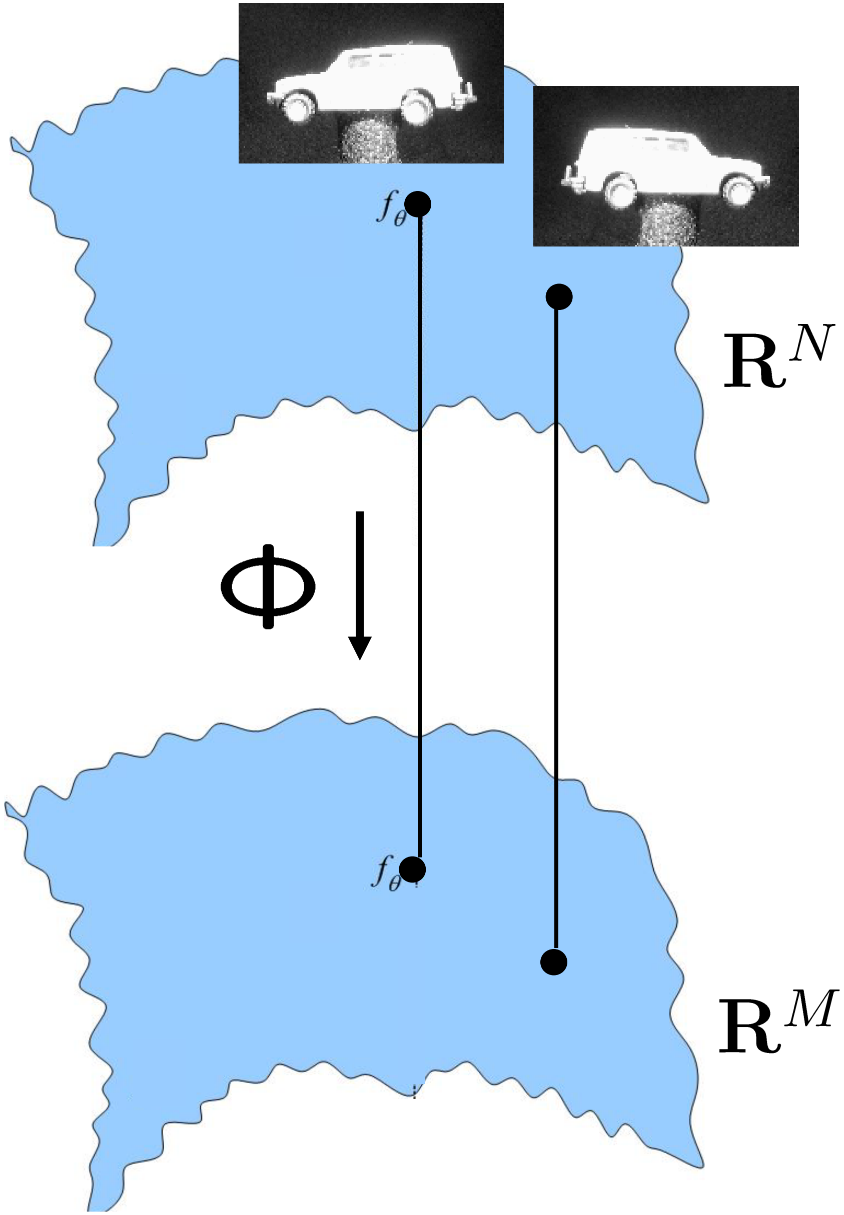
CS for Manifolds

- **Theorem:**
  $M = O(K \log N)$
  random measurements
  stably embed manifold
  whp

  [Baraniuk, Wakin, *FOCM* '08]
  related work:
  [Indyk and Naor, Agarwal et al.,
  Dasgupta and Freund]

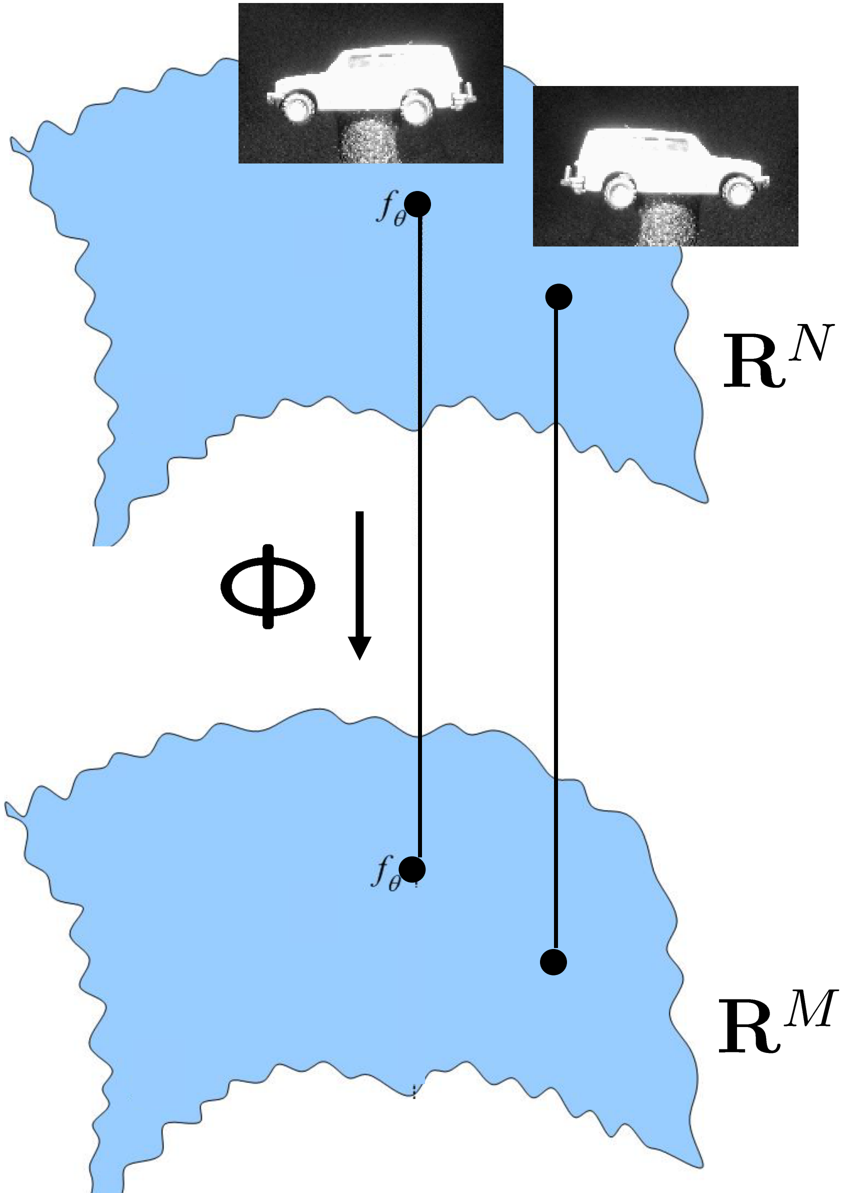
- Stable embedding

- Proved via concentration inequality arguments (JLL/CS relation)



CS for Manifolds

- **Theorem:**
  $M = O(K \log N)$
  random measurements
  stably embed manifold
  whp

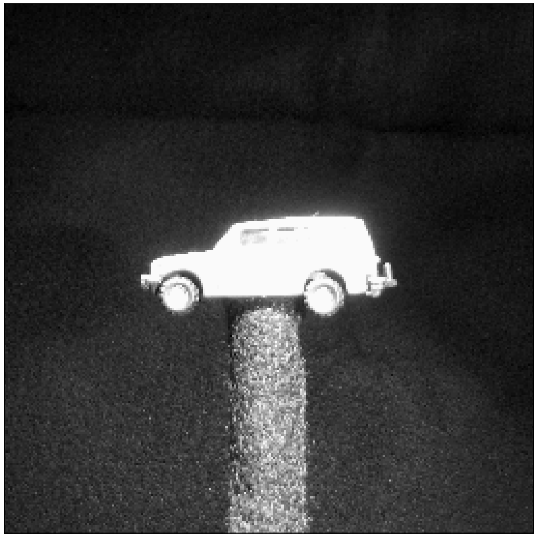


- Enables parameter estimation and MF detection/classification **directly on compressive measurements**
  – *K* very small in many applications (# articulations)

Example: Matched Filter

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



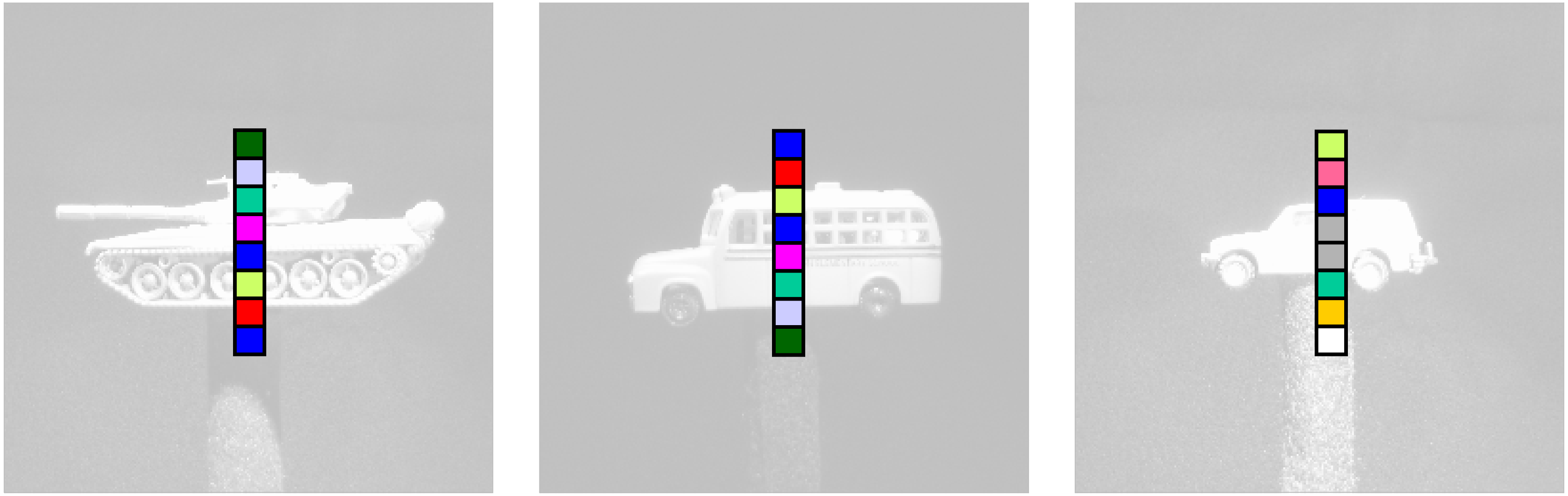


Smashed Filter

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

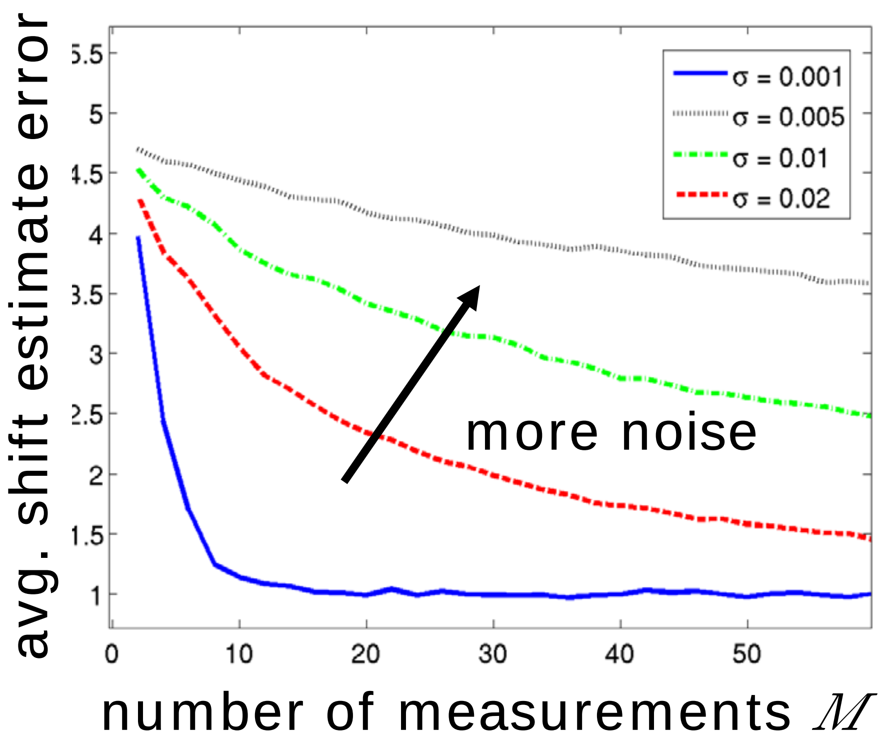
- Dimensionally reduced matched filter directly on compressive measurements

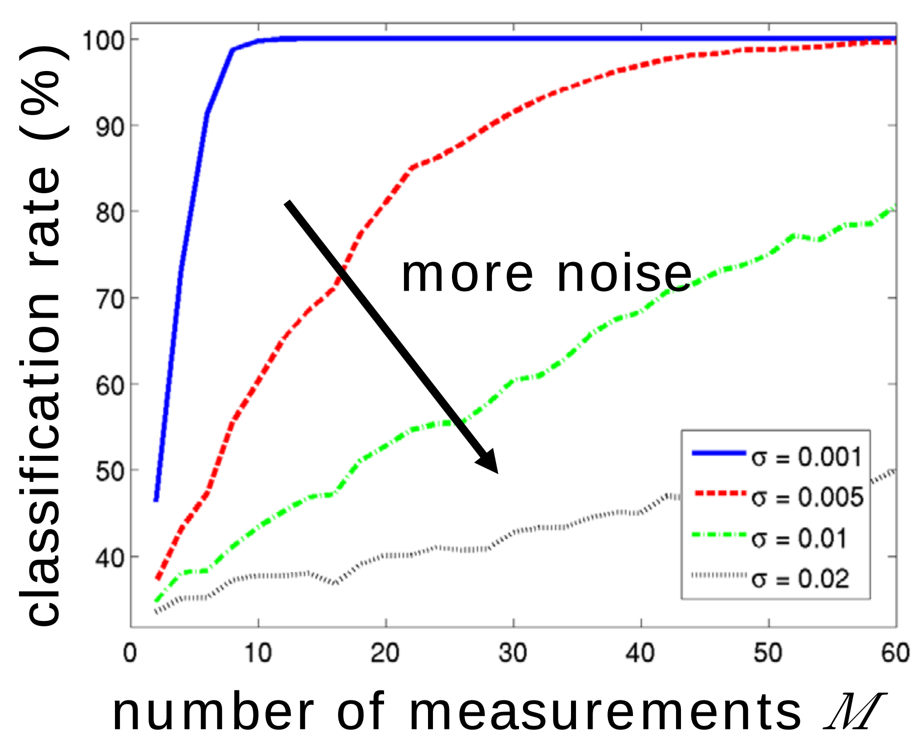
$$M = O(K \log N)$$



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





Compressive Sensing

Summary

CS Hallmarks

- CS changes the rules of the data acquisition game
  - exploits a priori signal *sparsity* information
- **Stable**
  - acquisition/recovery process is numerically stable
- **Universal**
  - same random projections / hardware can be used for *any* compressible signal class (*generic*)
- **Asymmetrical** (most processing at decoder)
  - conventional: smart encoder, dumb decoder
  - CS: dumb encoder, smart decoder
- Random projections weakly **encrypted**

CS Hallmarks

- **Democratic**
  - each measurement carries the same amount of information
  - robust to measurement loss and quantization simple encoding

- Ex: wireless streaming application with data loss

  - conventional: complicated (unequal) error protection of compressed data
    - DCT/wavelet low frequency coefficients

  - CS: merely stream additional measurements and reconstruct using those that arrive safely (fountain-like)