# Compressive Sensing
## *and Applications*

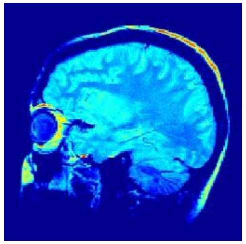
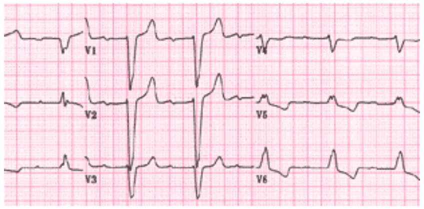
*Volkan Cevher*

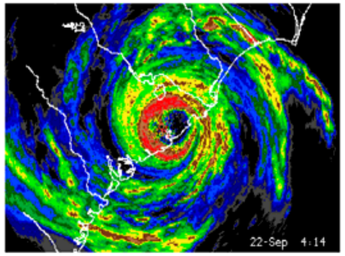
*volkan@rice.edu*

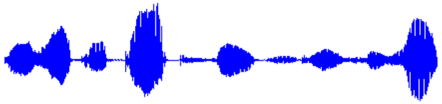
Rice University

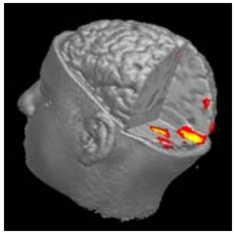
RICE UNIVERSITY
DSP







# Acknowledgements

RICE UNIVERSITY DSP

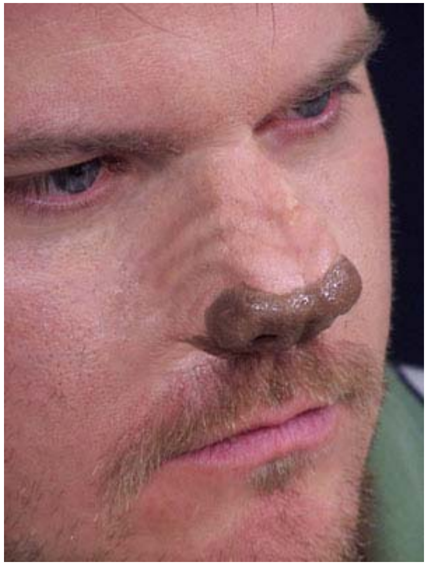
- Rice DSP Group (Slides)
  - Richard Baraniuk
    - Mark Davenport,
    - Marco Duarte,
    - Chinmay Hegde,
    - Jason Laska,
    - Shri Sarvotham,
    - Mona Sheikh
    - Stephen Schnelle...

  - Mike Wakin, Justin Romberg, 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)
  - signal recovery
- CS in Action
- Summary

# Sensing



# 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, ...

**large numbers of sensors**

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

**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

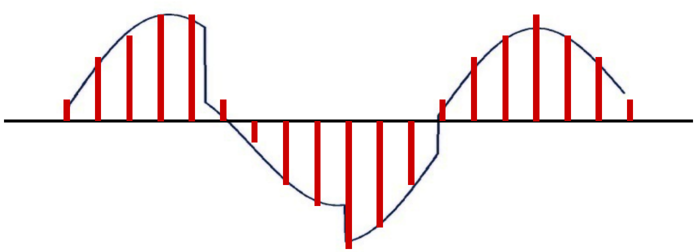
***deluge of data***

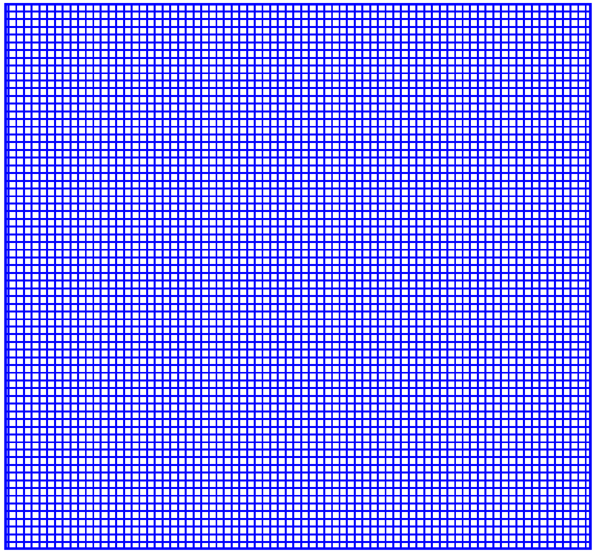
» how to **acquire**, **store**, **fuse**, **process** efficiently?

# Digital Data Acquisition

- Foundation: *Shannon/Nyquist sampling theorem*

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



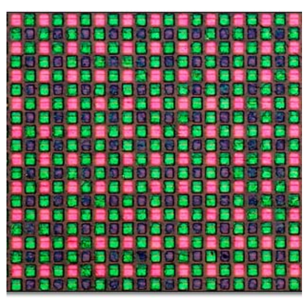


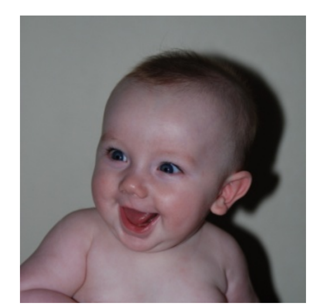
time

space

# Sensing by *Sampling*

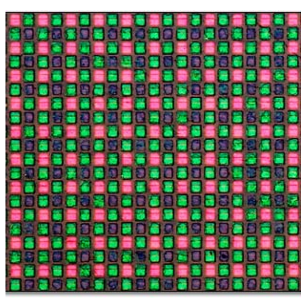
- Long-established paradigm for digital data acquisition
  - uniformly ***sample*** data at Nyquist rate (2x Fourier bandwidth)



# Sensing by *Sampling*

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

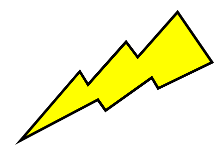
# Sensing by *Sampling*

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

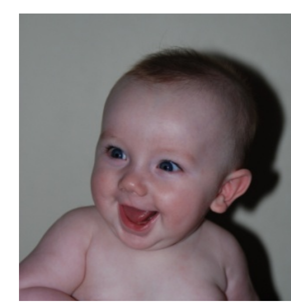
$$N \gg K$$

$x$ → sample →($N$)→ **compress** →($K$)→ transmit/store

**JPEG**
**JPEG2000**
**...**

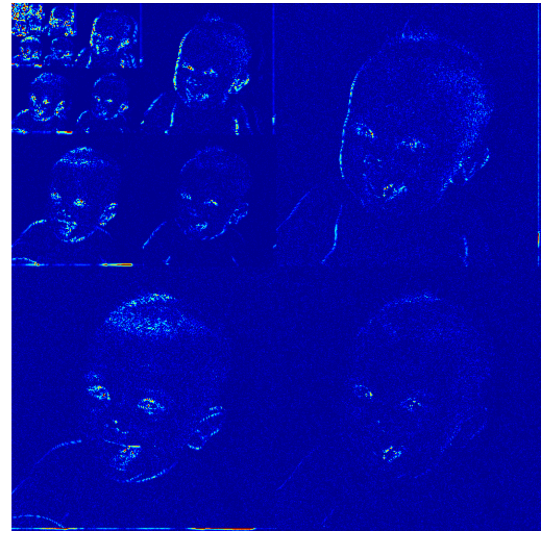


receive →($K$)→ decompress →($N$)→ $\widehat{x}$



# Sparsity / Compressibility

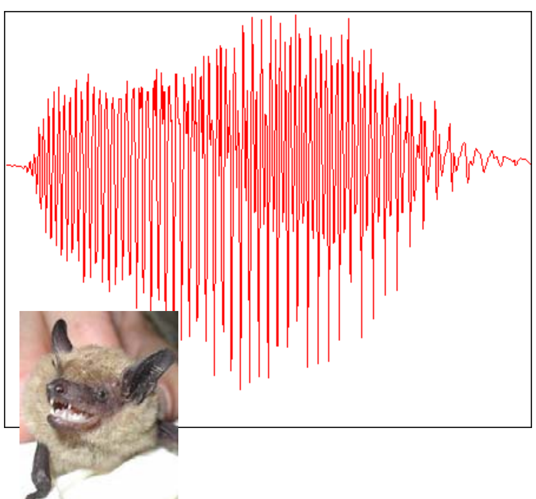
$N$
pixels

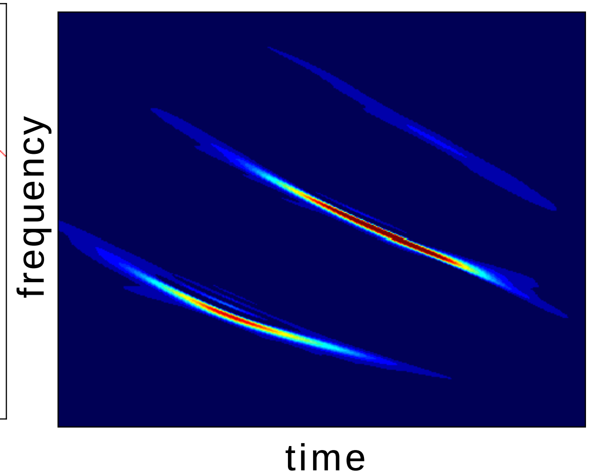


$K \ll N$
large
wavelet
coefficients

(blue = 0)

$N$
wideband
signal
samples

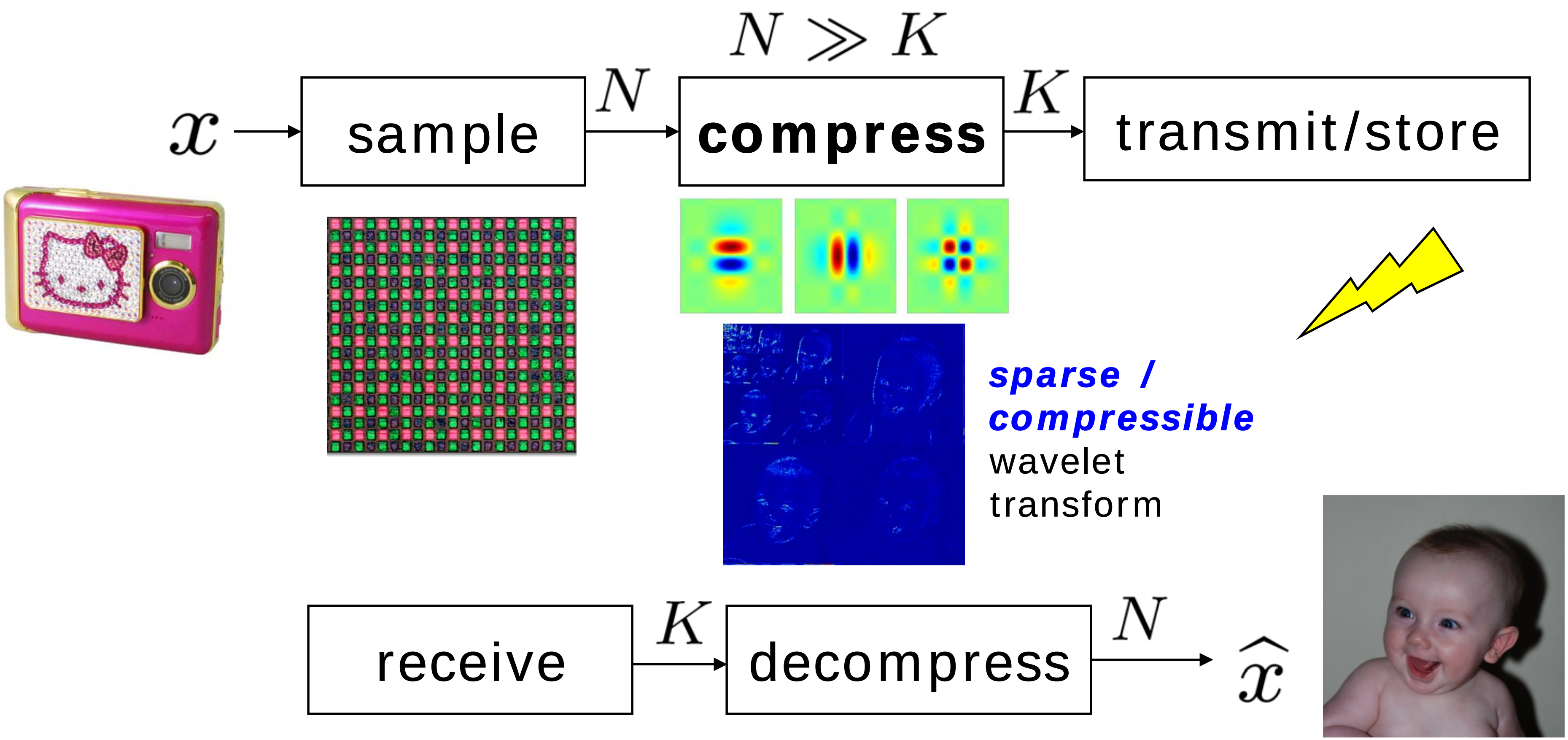





$K \ll N$
large
Gabor (TF)
coefficients

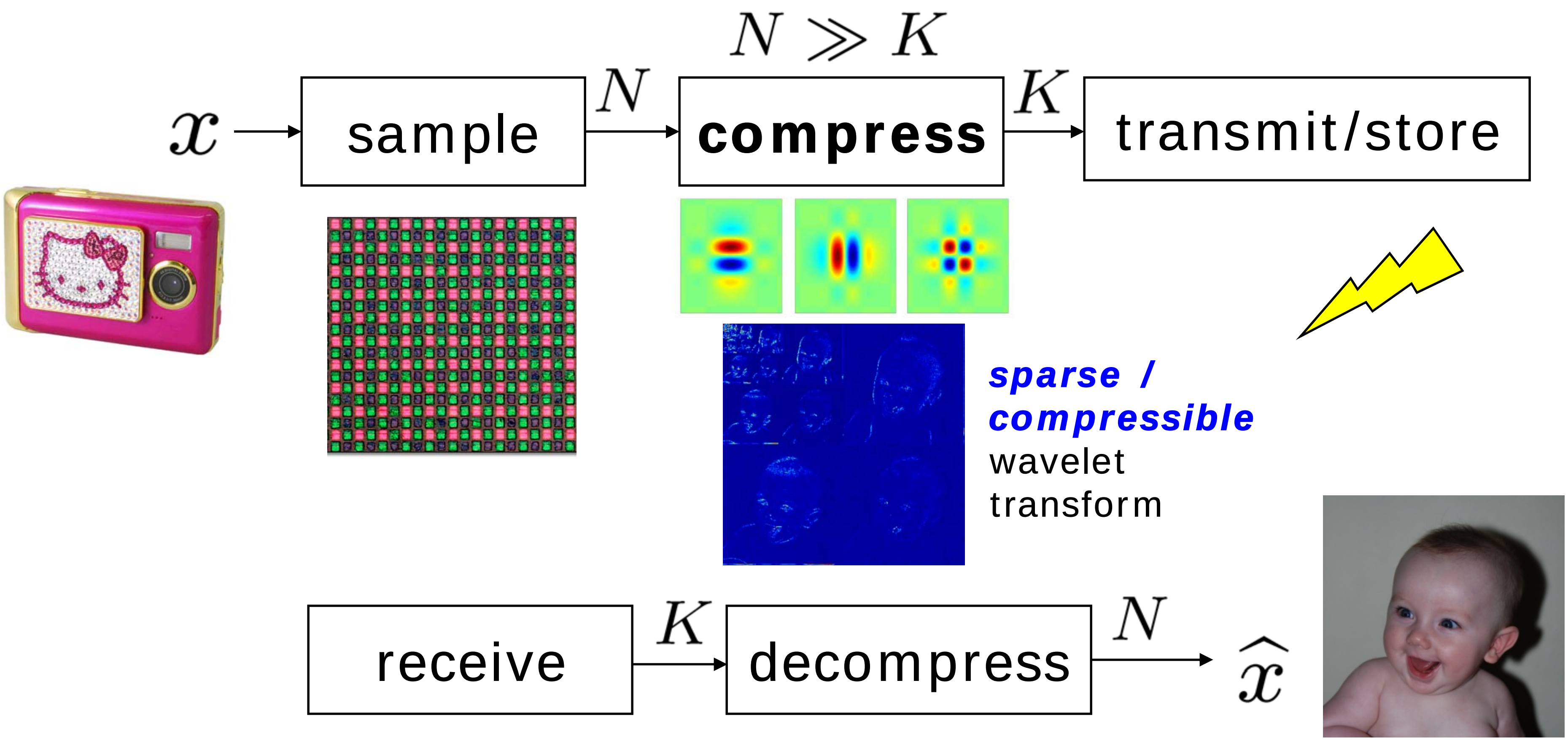
# Sample / Compress

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

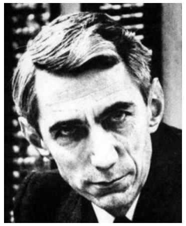


# What's Wrong with this Picture?

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



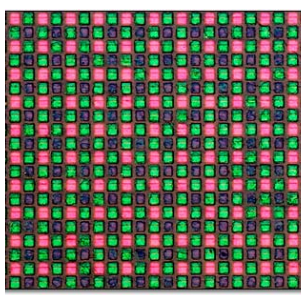
# What’s Wrong with this Picture?

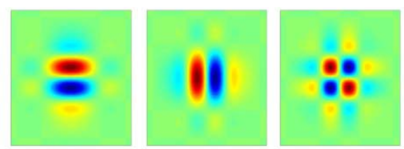


**linear** processing
**linear** signal model
(bandlimited subspace)

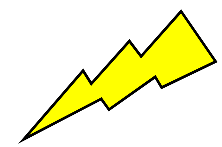
**nonlinear** processing
**nonlinear** signal model
(union of subspaces)

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

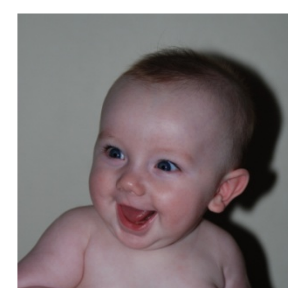




***sparse / compressible*** wavelet transform



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

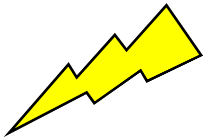


# Compressive Sensing

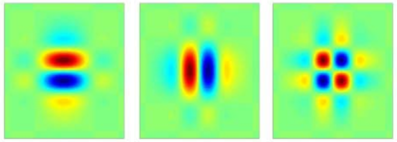
- Directly acquire "***compressed***" data
- Replace samples by more general "measurements"

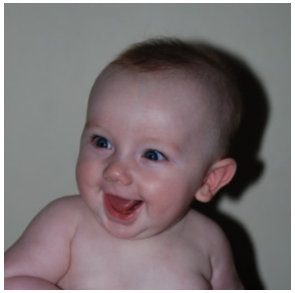
$$K \approx \underline{M} \ll N$$

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



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



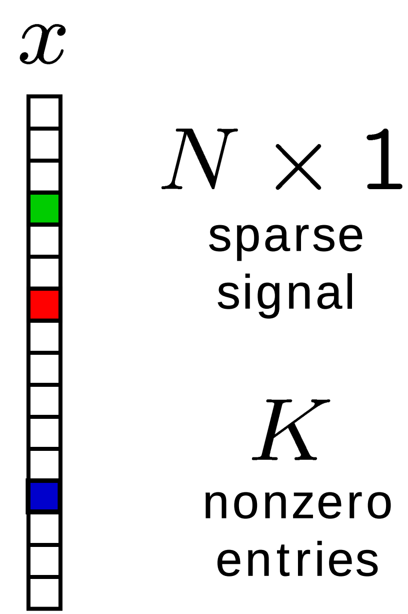


# Compressive Sensing

## Theory I
## Geometrical Perspective

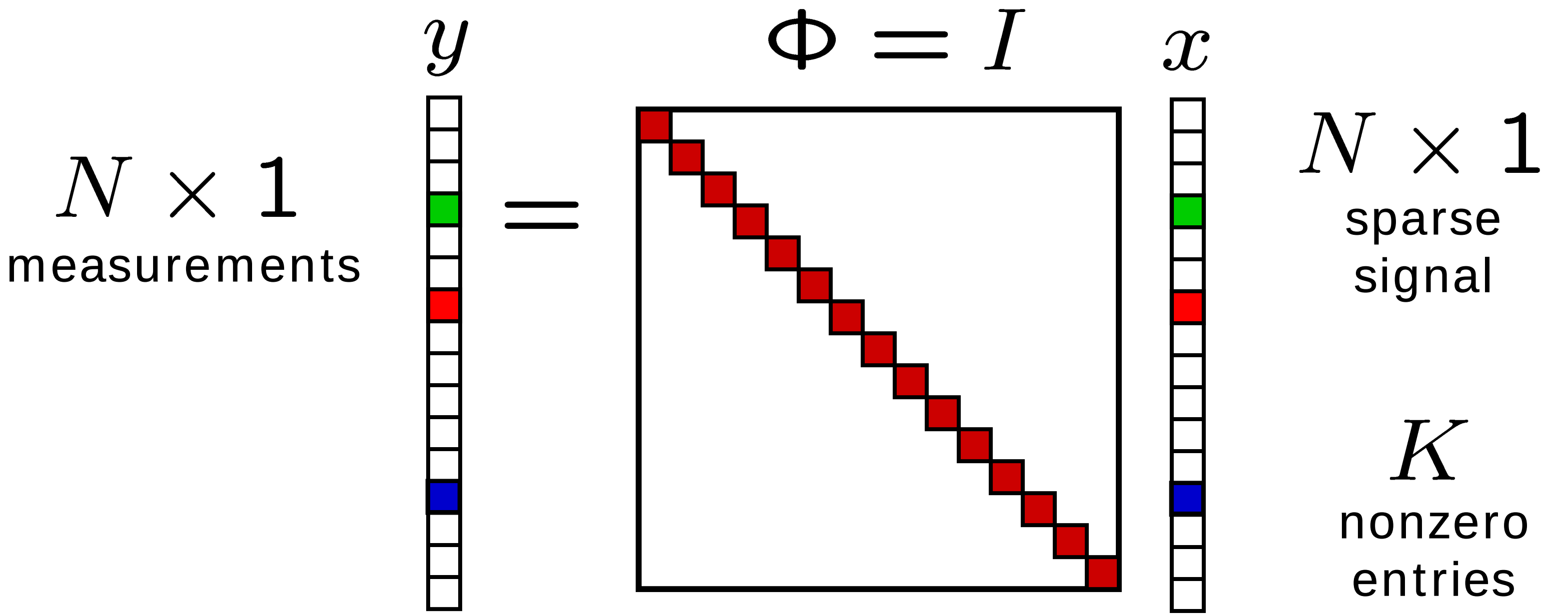
# Sampling

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



# Sampling

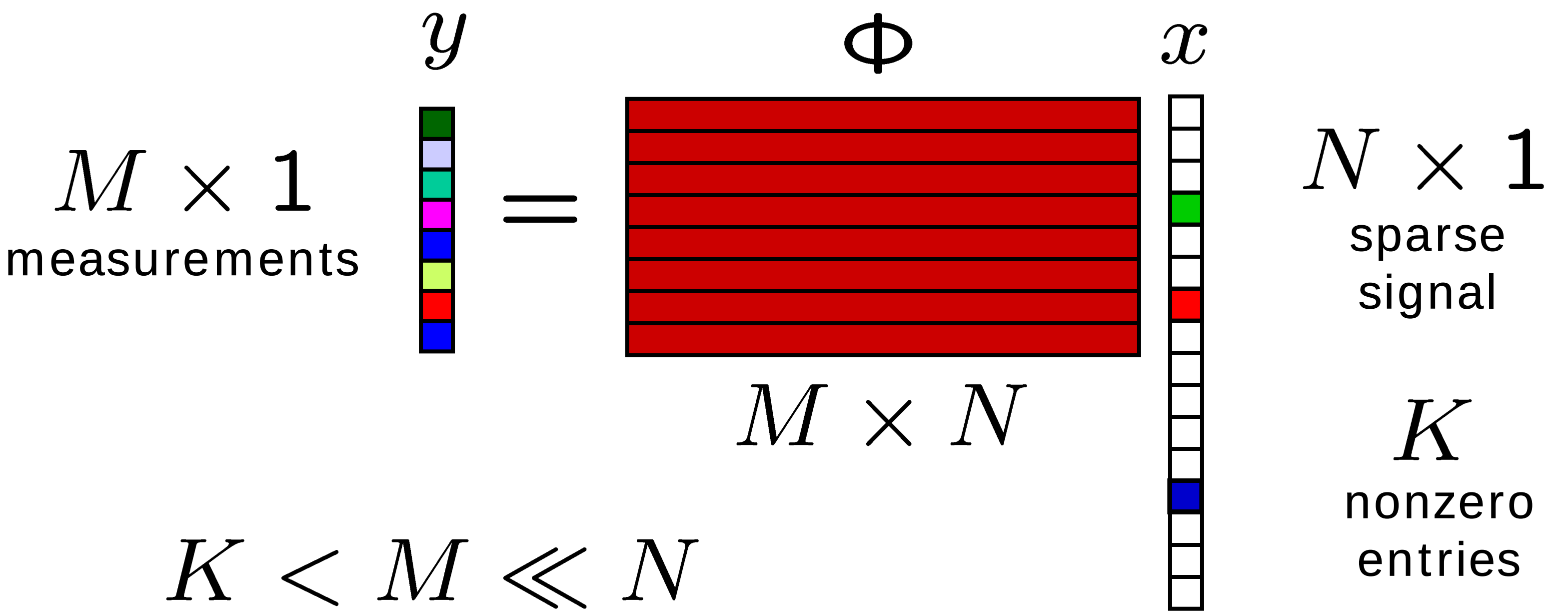
- Signal $x$ is $K$-*sparse* in basis/dictionary $\Psi$
  - WLOG assume sparse in space domain $\Psi = I$
- **Samples**



# *Compressive* Sampling

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

$$y = \Phi x$$



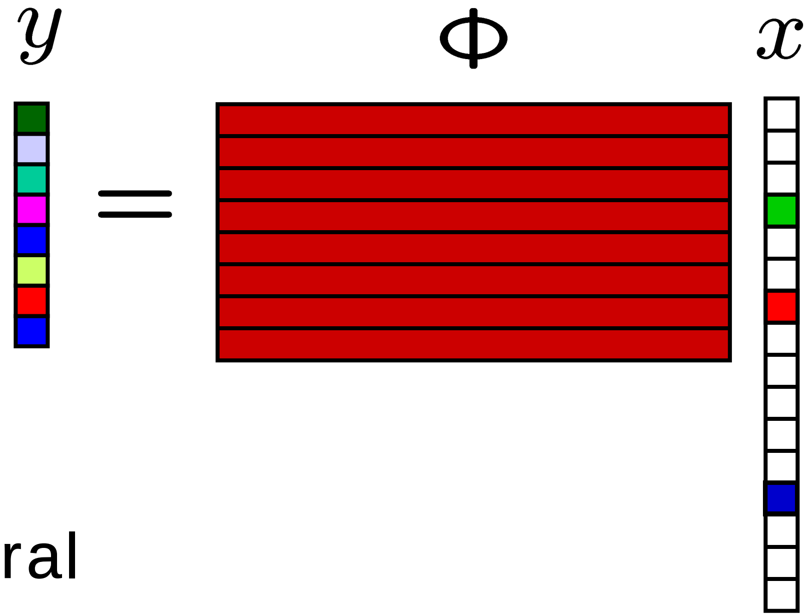

$$K < M \ll N$$

# How Can It Work?

- Projection $\Phi$
  **not full rank**...

  $M < N$

  ... and so
  **loses information** in general



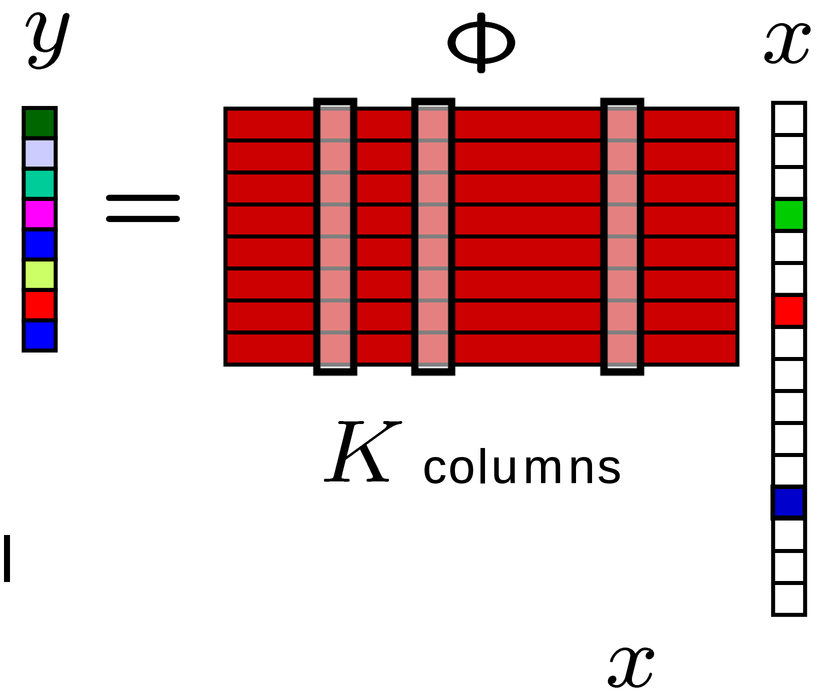

- Ex: Infinitely many $x$'s map to the same $y$

# How Can It Work?

- Projection $\Phi$ not full rank…

  $M < N$

  … and so loses information in general



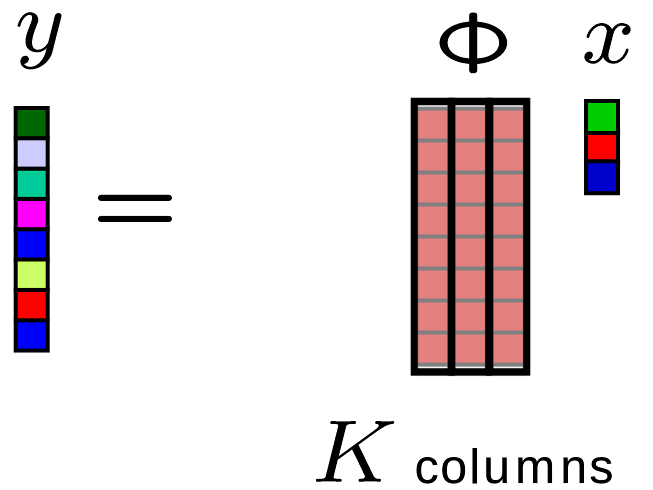

- But we are only interested in ***sparse*** vectors

# How Can It Work?

- Projection $\Phi$ not full rank…

  $M < N$

  … and so loses information in general




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

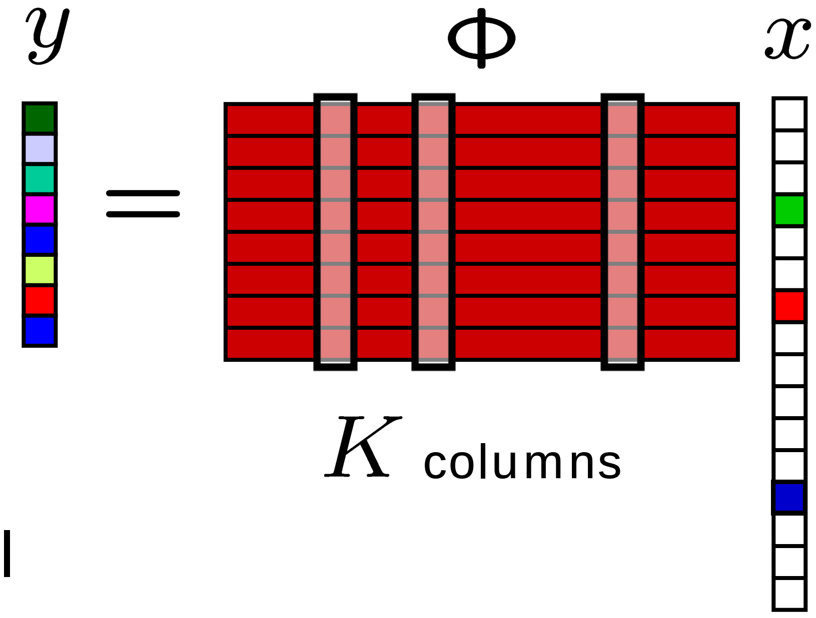
- $\Phi$ is effectively *M*x*K*

# How Can It Work?

- Projection $\Phi$
  not full rank...

  $$M < N$$

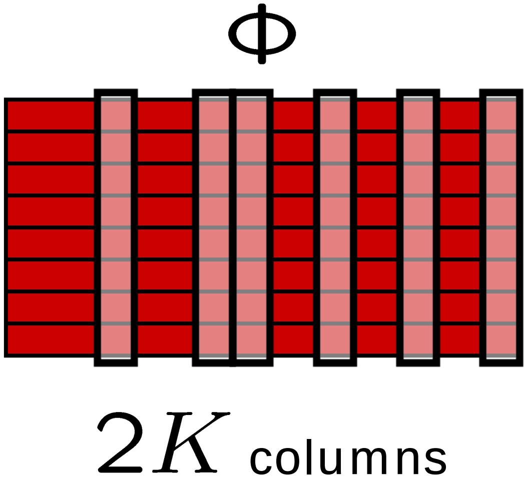
  ... and so
  loses information in general




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

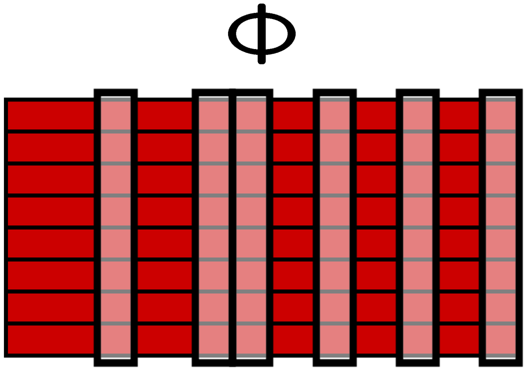
- **Design** $\Phi$ so that each of its *M*x*K* submatrices are full rank

# How Can It Work?




- **Goal:** Design $\Phi$ 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*

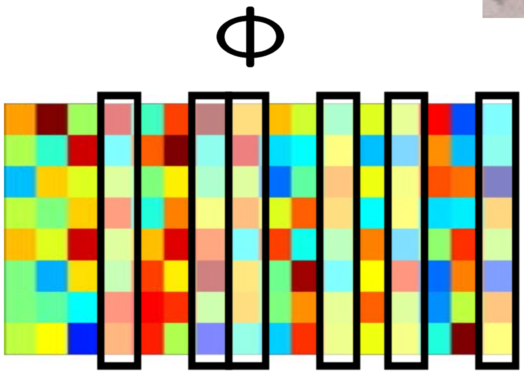
# Unfortunately...




- **Goal:** Design $\Phi$ 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]



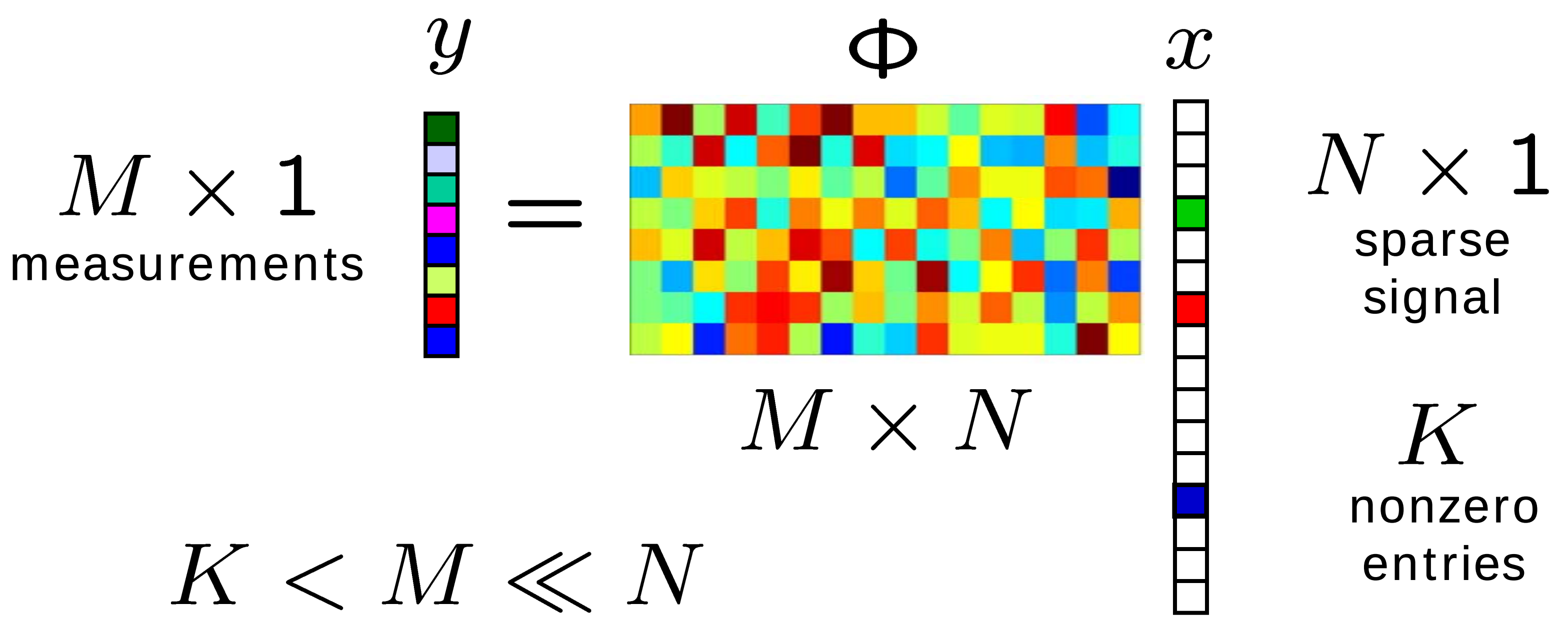
$2K$ columns

- Draw $\Phi$ at **random**
  - iid Gaussian
  - iid Bernoulli $\pm 1$
  
    ...

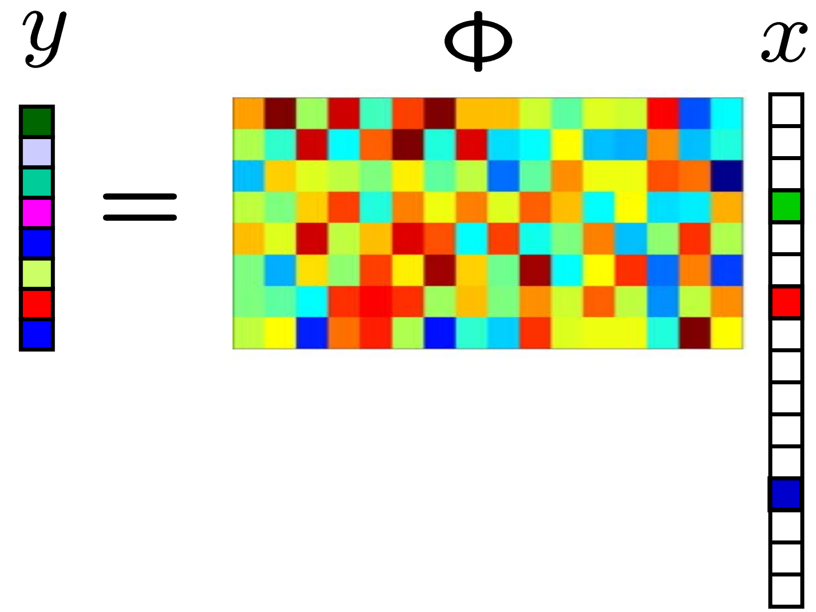
- Then $\Phi$ has the RIP with high probability as long as $M = O(K \log(N/K)) \ll N$
  - *M*x2*K* submatrices are full rank
  - stable embedding for sparse signals
  - extends to compressible signals in $\ell_p$ balls

# Compressive Data Acquisition

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



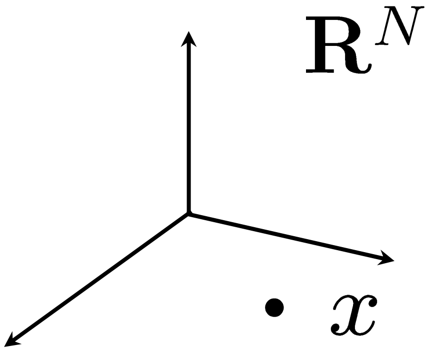
# CS Signal Recovery

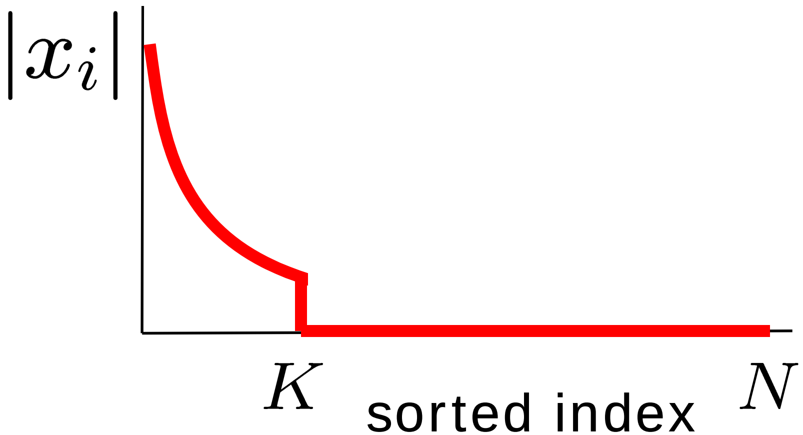



- **Goal**: Recover signal $x$ from measurements $y$
- **Challenge**: Random projection $\Phi$ not full rank (ill-posed inverse problem)
- **Solution**: Exploit the sparse/compressible ***geometry*** of acquired signal $x$

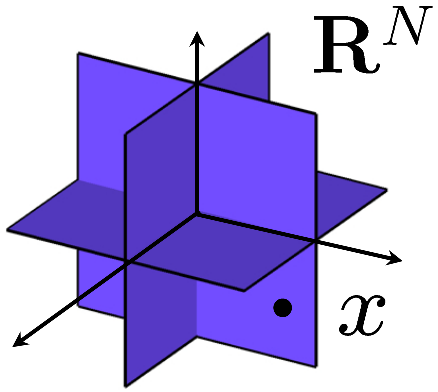
# Concise Signal Structure

- **Sparse** signal: only $K$ out of $N$ coordinates nonzero

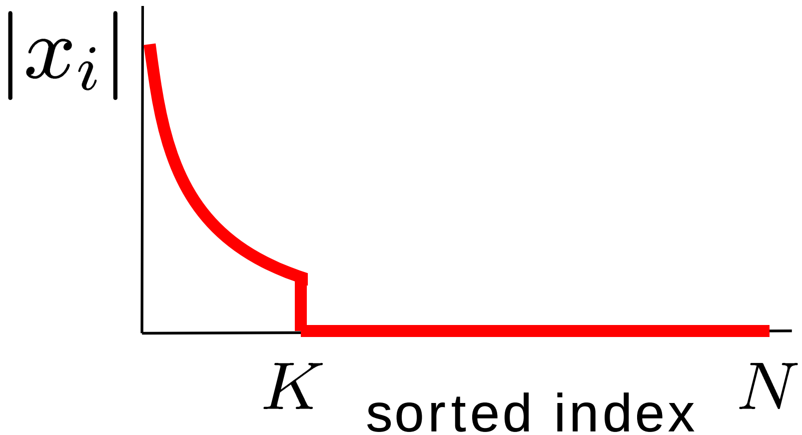


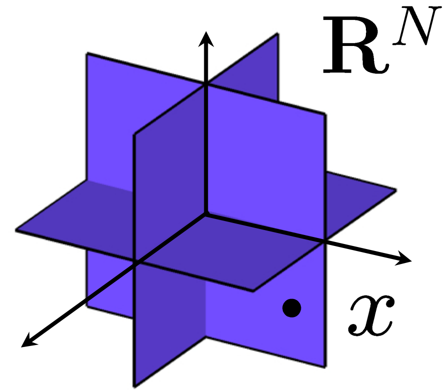
# Concise Signal Structure




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

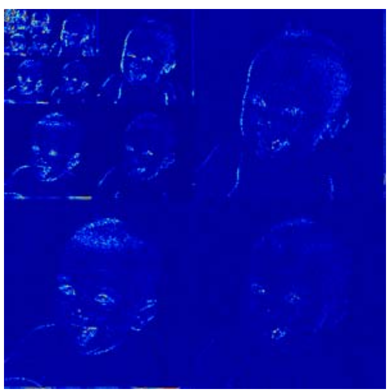


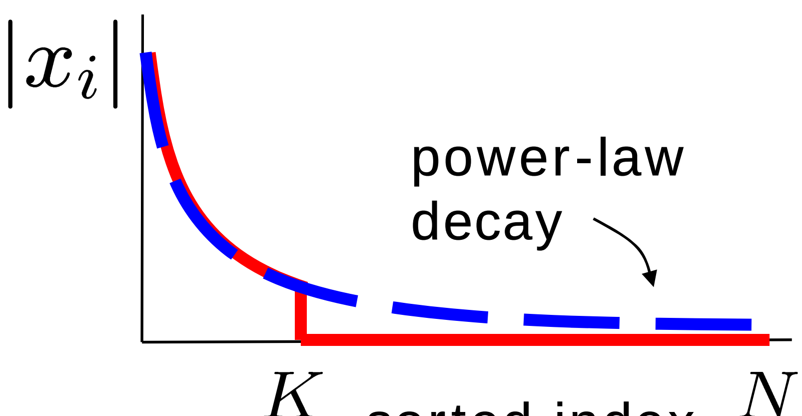
# Concise Signal Structure



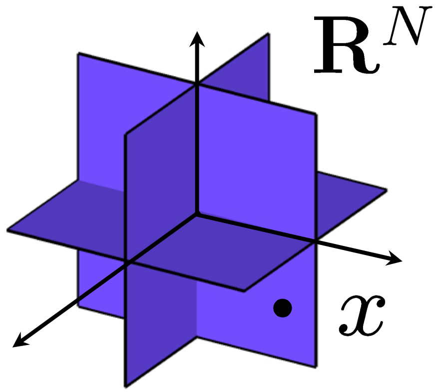

- **Sparse** signal: only $K$ out of $N$ coordinates nonzero
  - model: union of $K$-dimensional subspaces

- **Compressible** signal: sorted coordinates decay rapidly to zero



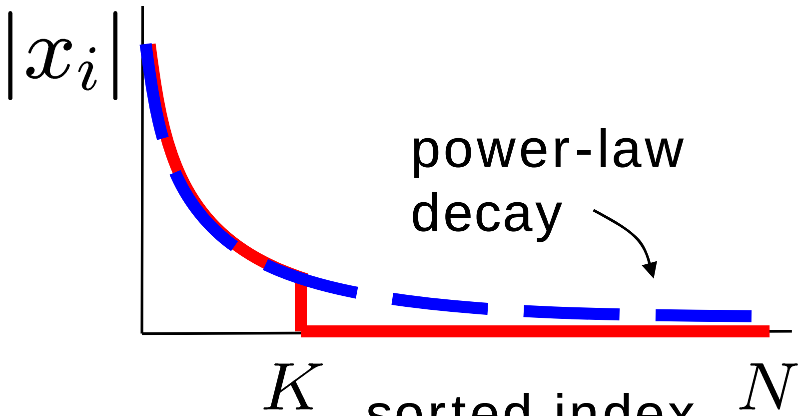


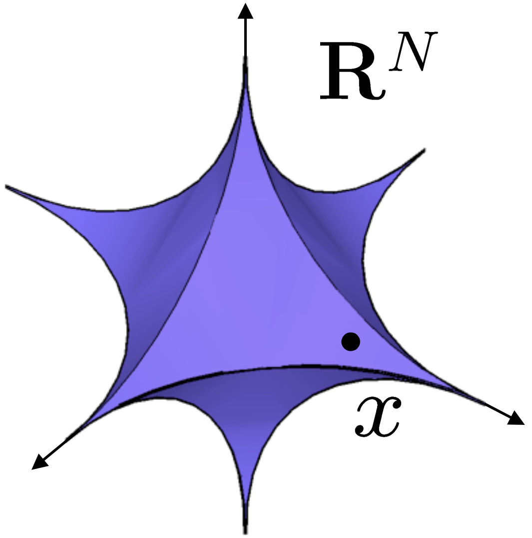
# Concise Signal Structure




- **Sparse** signal: only $K$ out of $N$ coordinates nonzero
  - model: union of $K$-dimensional subspaces

- **Compressible** signal: sorted coordinates decay rapidly to zero
  - model: $\ell_p$ ball: $\|x\|_p^p = \sum_i |x_i|^p \leq 1, \; p \leq 1$



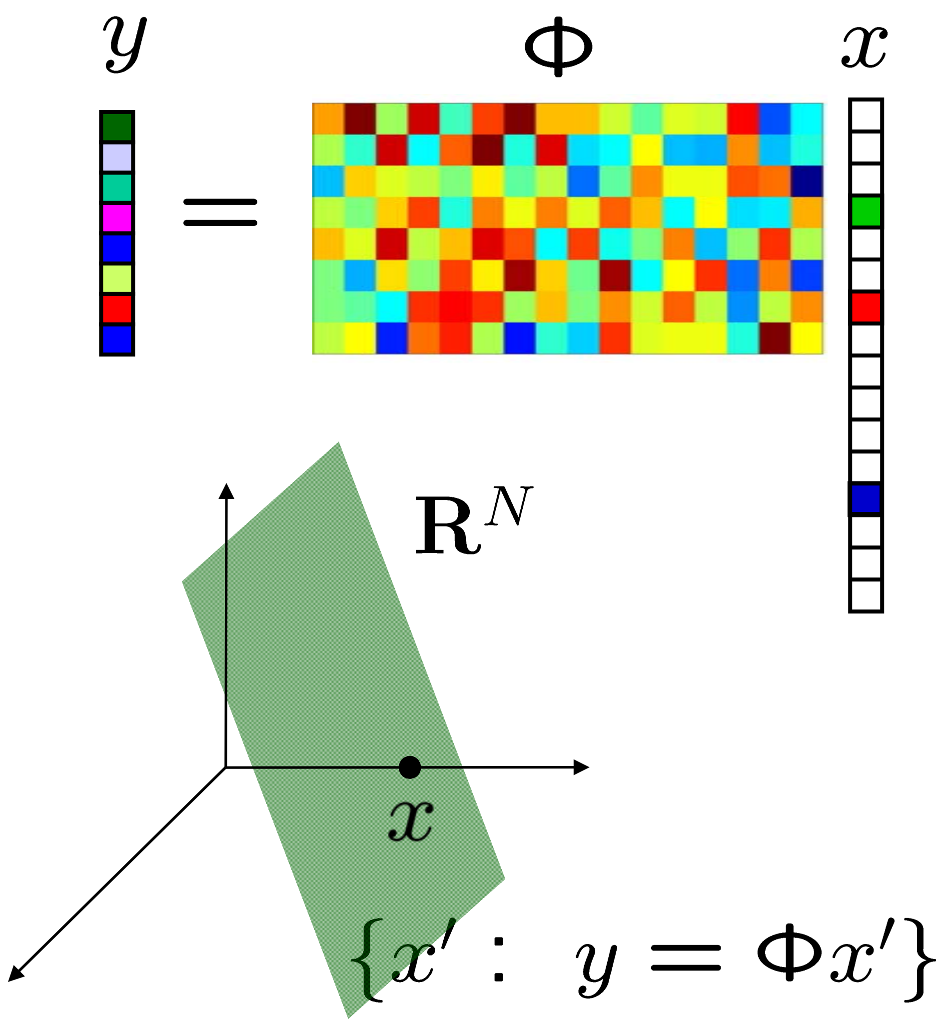
# CS Signal Recovery

- Random projection $\Phi$ not full rank

- Recovery problem:
  given $y = \Phi x$
  find $x$

- **Null space**

- So search in null space for the "best" $x$ according to some criterion
  - ex: least squares



# CS Signal Recovery

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

- $\ell_2$ **fast** $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$

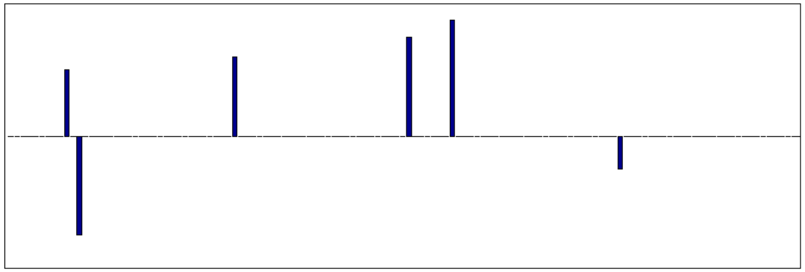
$$\widehat{x} = (\Phi^T \Phi)^{-1} \Phi^T y$$

pseudoinverse

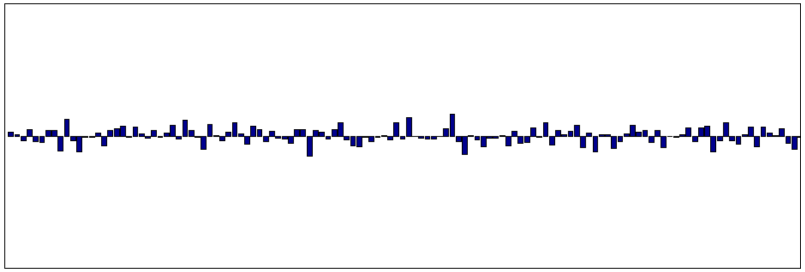
# CS Signal Recovery

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

- $\ell_2$ **fast, wrong** $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$



$x$



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

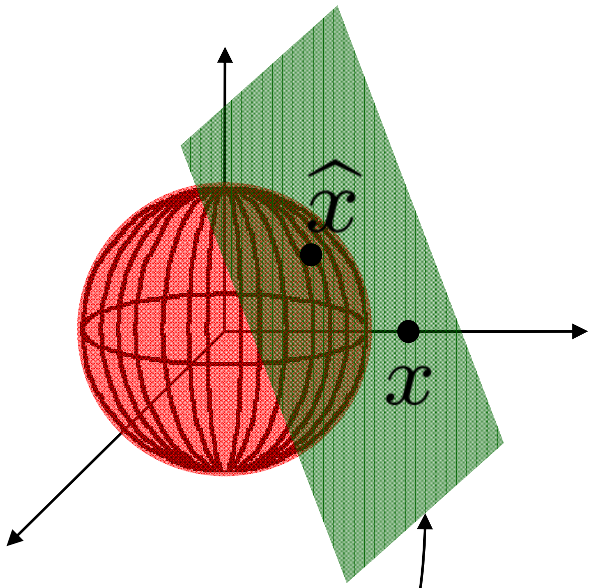
pseudoinverse

# Why $\ell_2$ Doesn't Work

for signals sparse in the
**space / time domain**

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

least squares,
minimum $\ell_2$ solution
is almost **never sparse**




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

*null space of* $\Phi$

*translated to* $x$

(random angle)

# CS Signal Recovery

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

- $\ell_2$ fast, wrong $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$

- $\ell_0$ $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_0$

↑ ***number of nonzero entries***

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

# CS Signal Recovery

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

- $\ell_2$ fast, wrong $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$

- $\ell_0$ **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*

# CS Signal Recovery

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

- $\ell_2$ fast, wrong $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$

- $\ell_0$ **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*

  **slow:** NP-complete algorithm

# CS Signal Recovery

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

- $\ell_2$ fast, wrong $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$

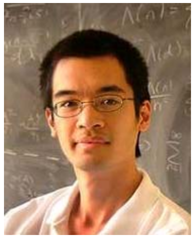
- $\ell_0$ correct, slow $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_0$

- $\ell_1$ **correct, efficient**
  **mild oversampling** $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_1$
  [Candes, Romberg, Tao; Donoho] ***linear program***

number of measurements required

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





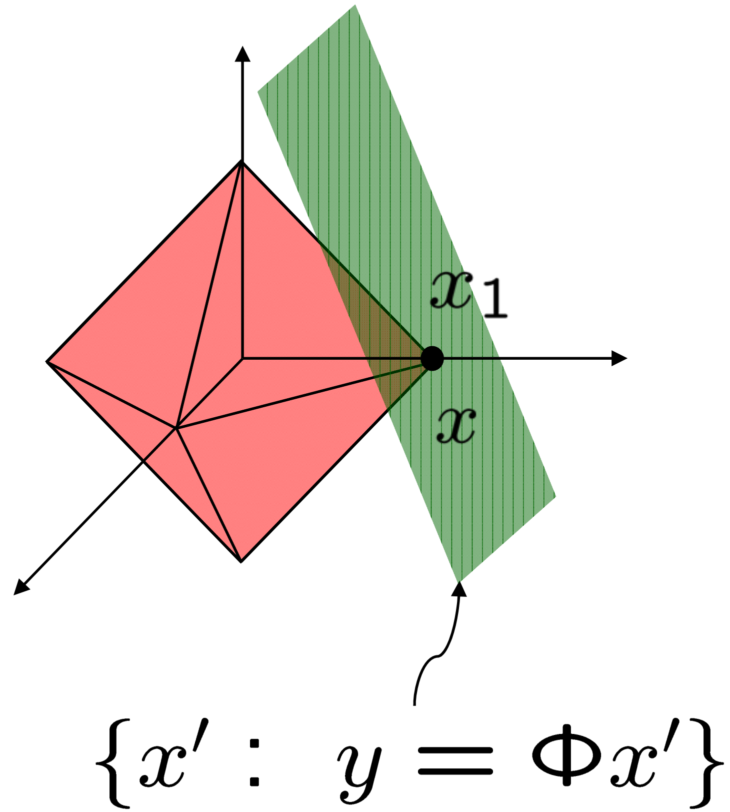

# Why $\ell_1$ Works

for signals sparse in the
**space/time domain**

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

minimum $\ell_1$ solution
= sparsest solution
(with high probability) if

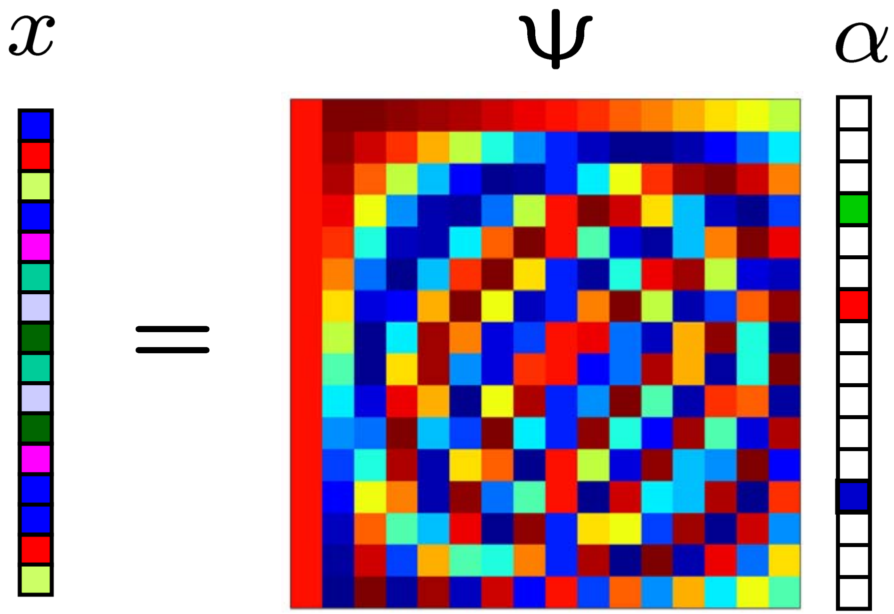
$$M = O(K \log(N/K)) \ll N$$



# Universality

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

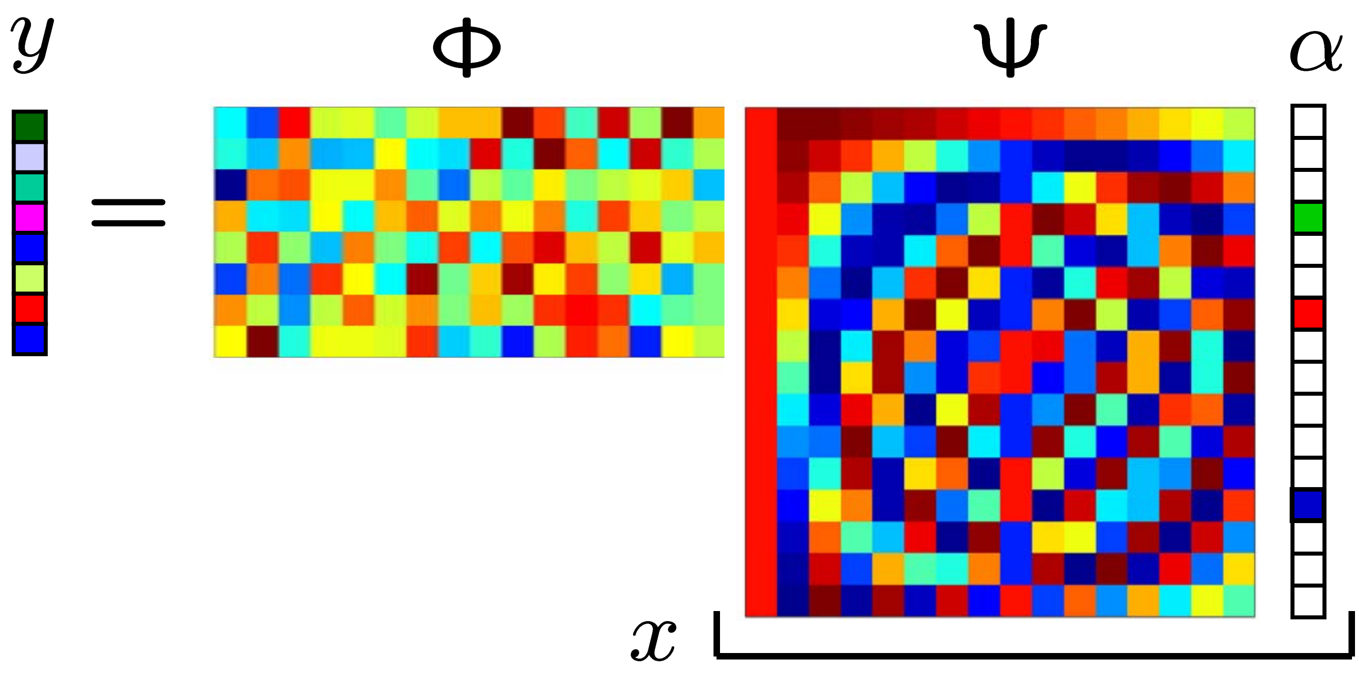
$$x = \Psi\alpha$$



# Universality

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

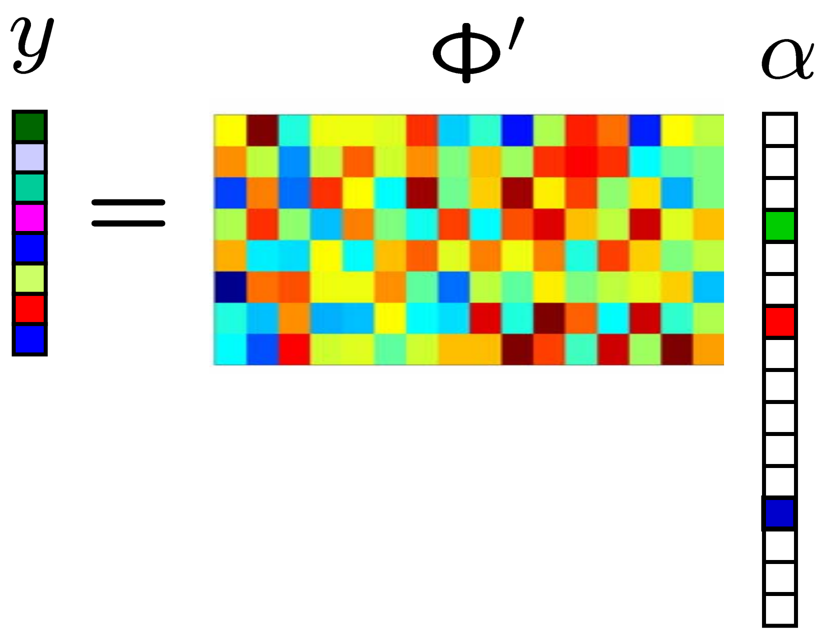
$$y = \Phi x = \Phi \Psi \alpha$$



# Universality

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

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



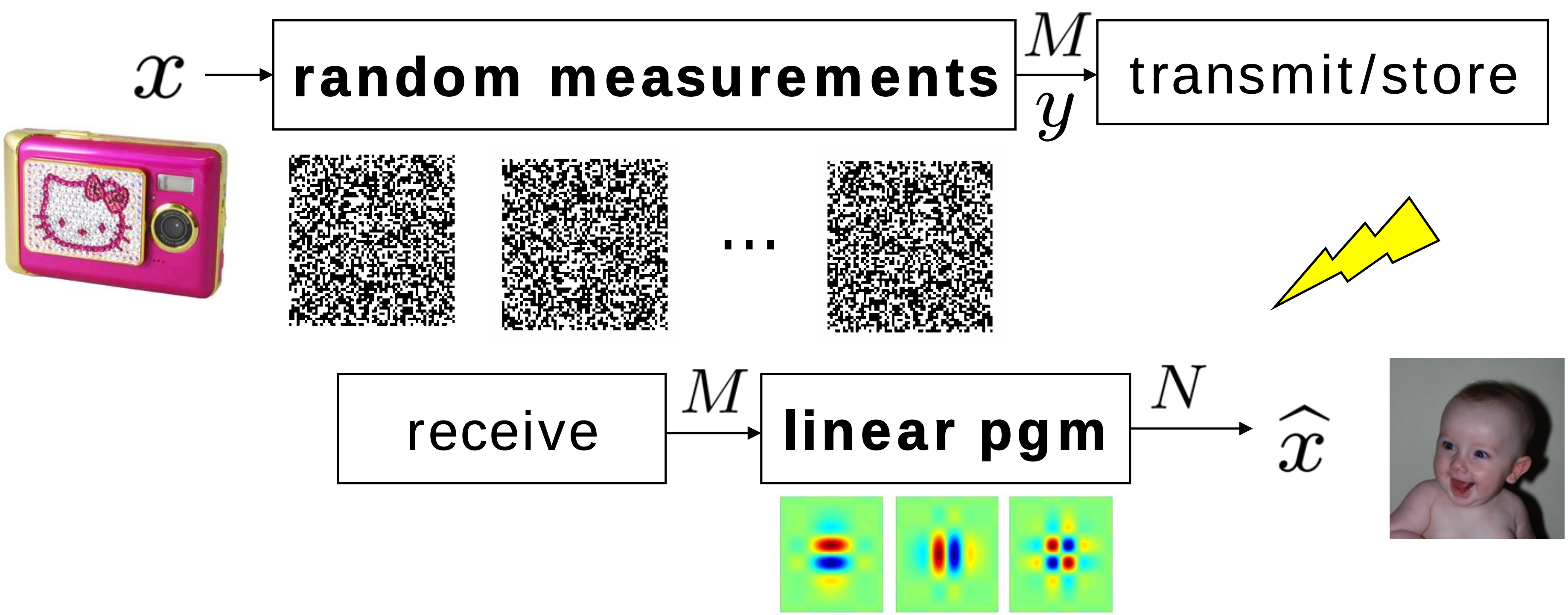

$N \times 1$
sparse coefficient vector

$K$
nonzero entries

# Compressive Sensing

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

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

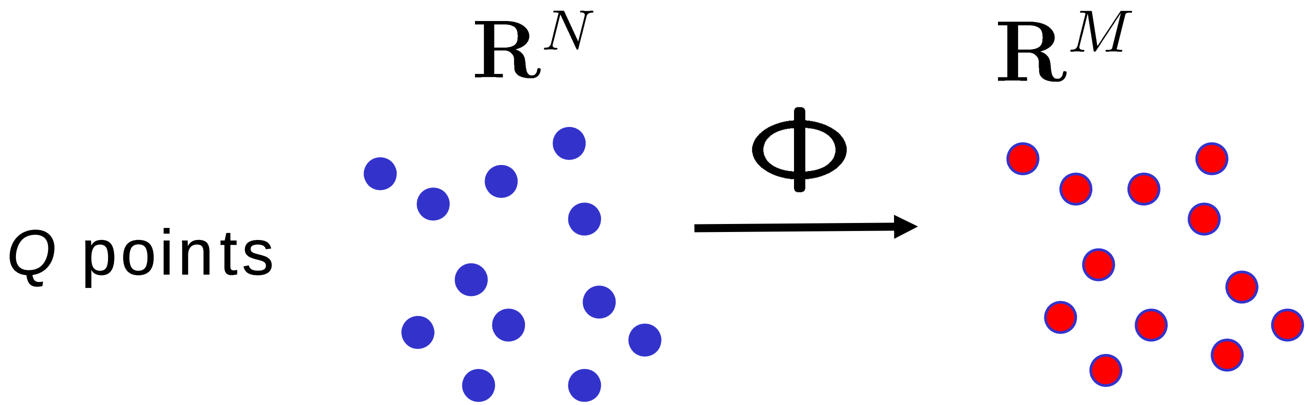


# Compressive Sensing

# Theory II
# Stable Embedding

# Johnson-Lindenstrauss Lemma

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



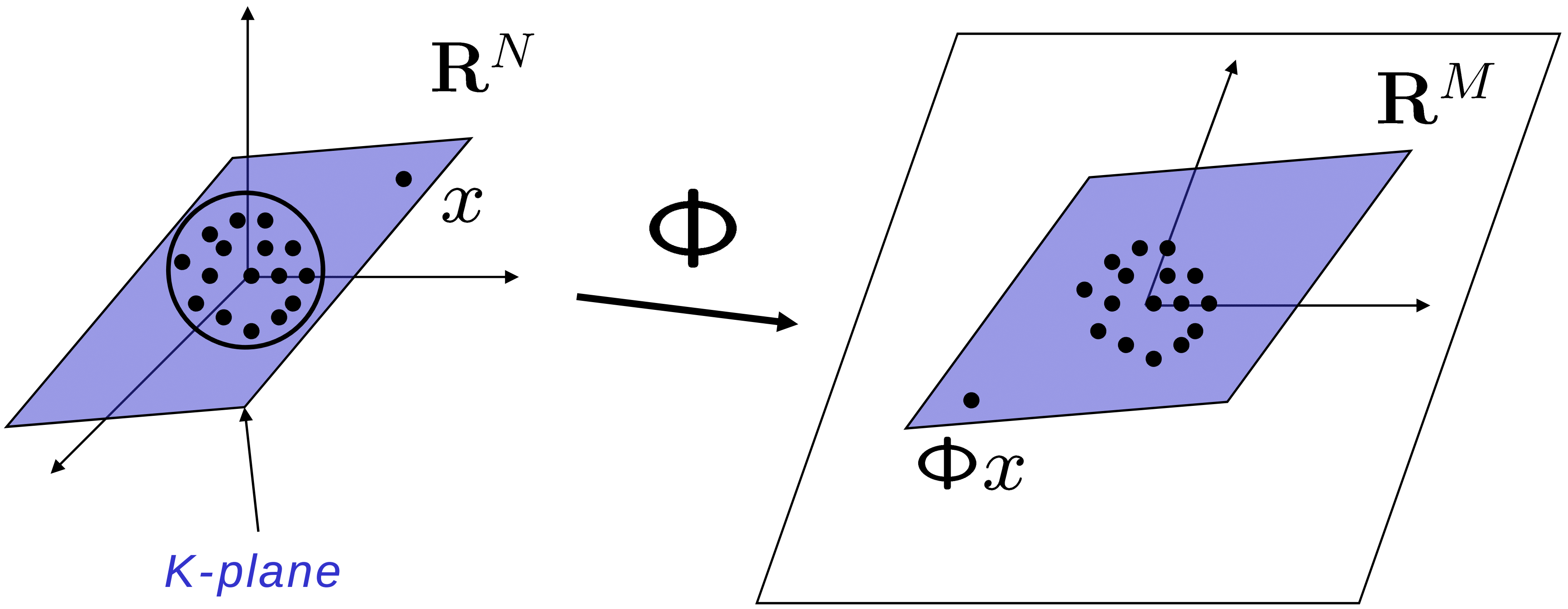

- Proved via concentration inequality

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

# Connecting JL to RIP

Consider effect of random JL $\Phi$ on each K-plane

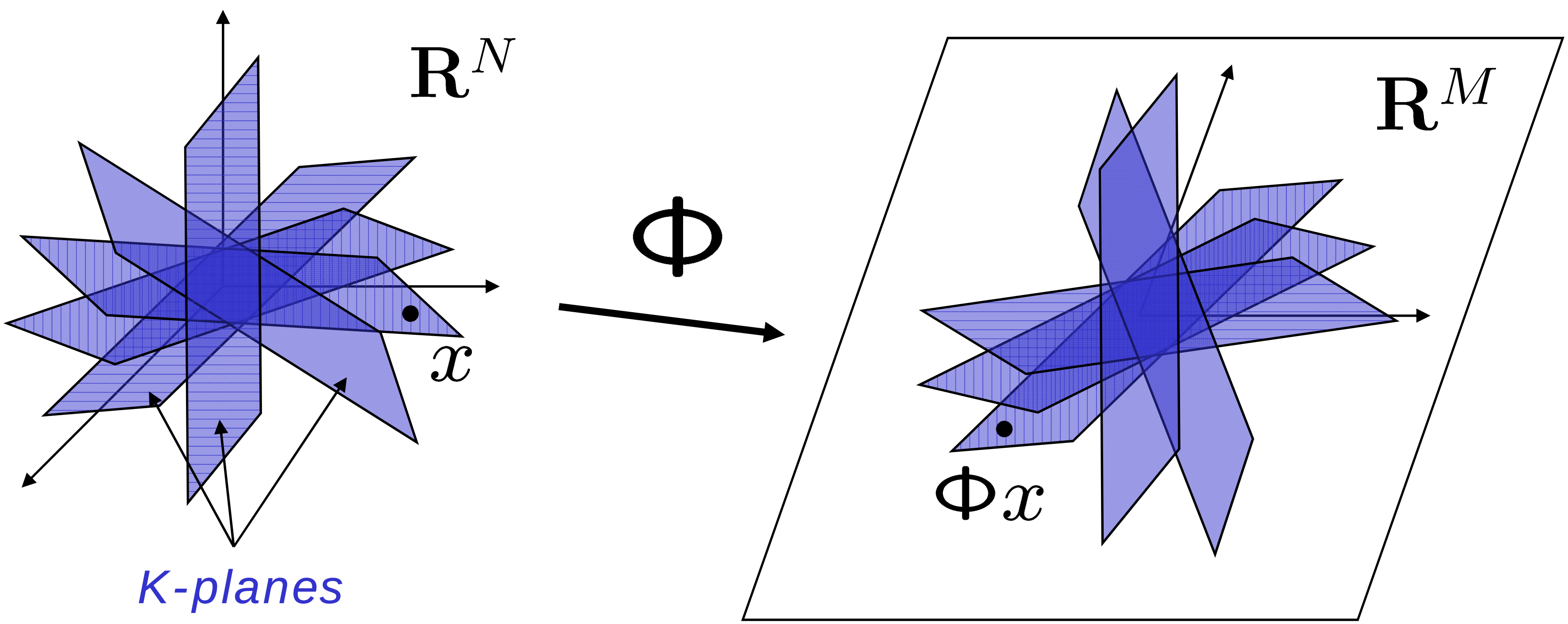
- construct covering of points Q on unit sphere
- JL: isometry for each point with high probability
- union bound ➔ isometry for all points q in Q
- extend to isometry for all x in K-plane



# Connecting JL to RIP

Consider effect of random JL $\Phi$ 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]

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

- “Database-friendly” [Achlioptas]

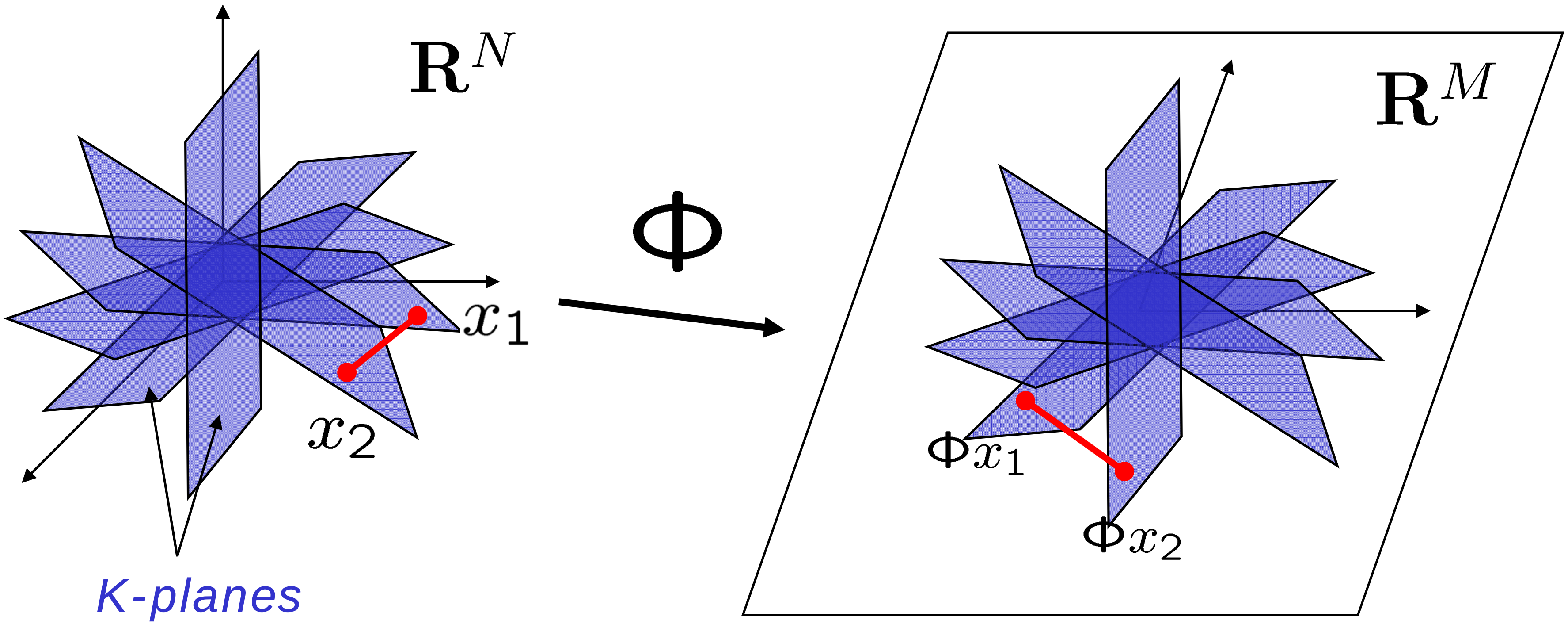
$$\phi_{i,j} := \begin{cases} +\sqrt{\frac{3}{M}} & \text{with probability} \quad \frac{1}{6}, \\ 0 & \text{with probability} \quad \frac{2}{3}, \\ -\sqrt{\frac{3}{M}} & \text{with probability} \quad \frac{1}{6} \end{cases}$$

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

# RIP as a "Stable" Embedding

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

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



# 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

# 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**

# BPDN

- Standard LP recovery

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

- Noisy measurements

$$y = \Phi x + n$$

- Basis Pursuit De-Noising

$$\min \frac{1}{2}\|y - \Phi x\|_2 + \lambda \|x\|_1$$

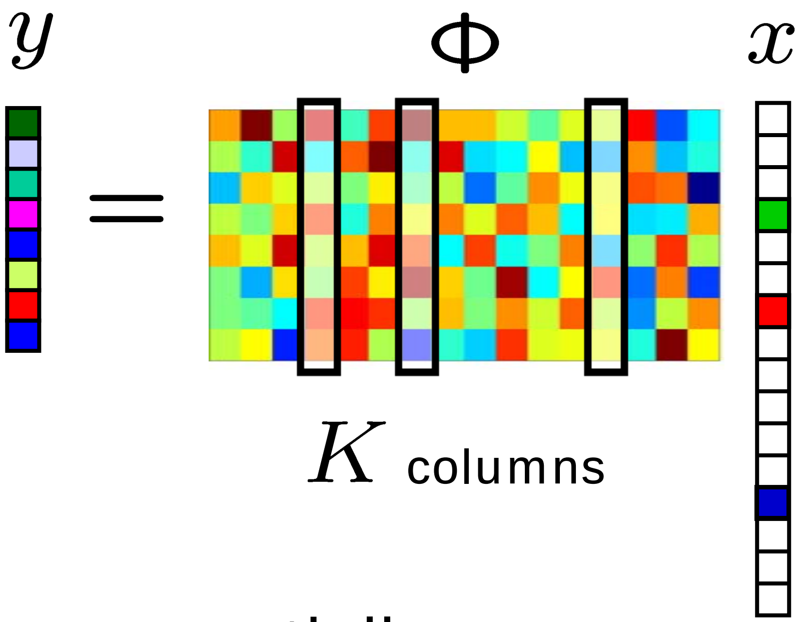
- Convex, **quadratic program**

# Matching Pursuit

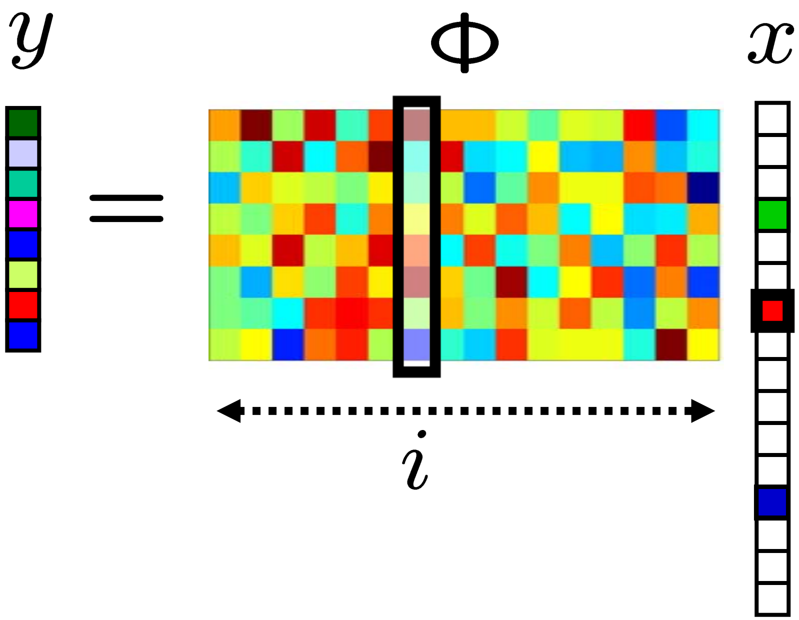
- Greedy algorithm

- **Key ideas:**
  (1) measurements $y$ composed of sum of *K* columns of $\Phi$

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



# Matching Pursuit




- For each column $\phi_i$ compute

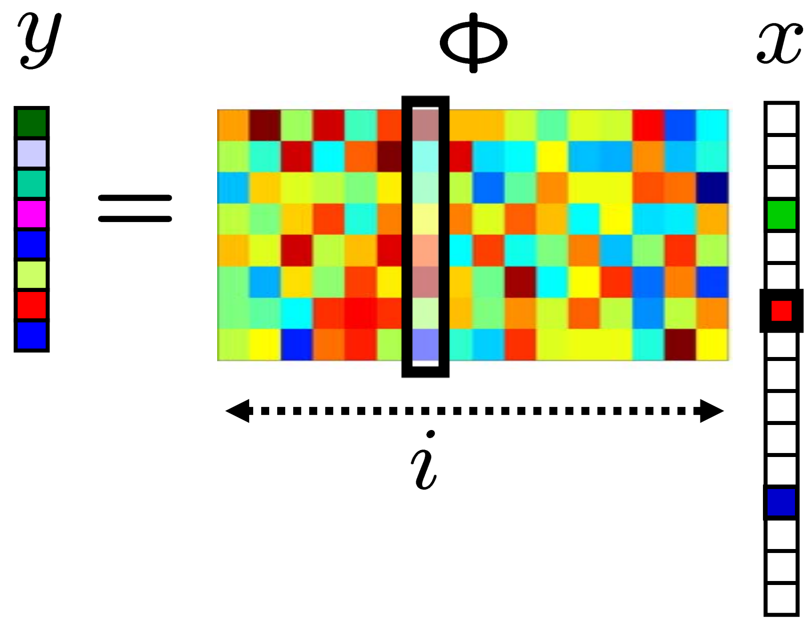
$$\widehat{x}_i = \langle y, \phi_i \rangle$$

- 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

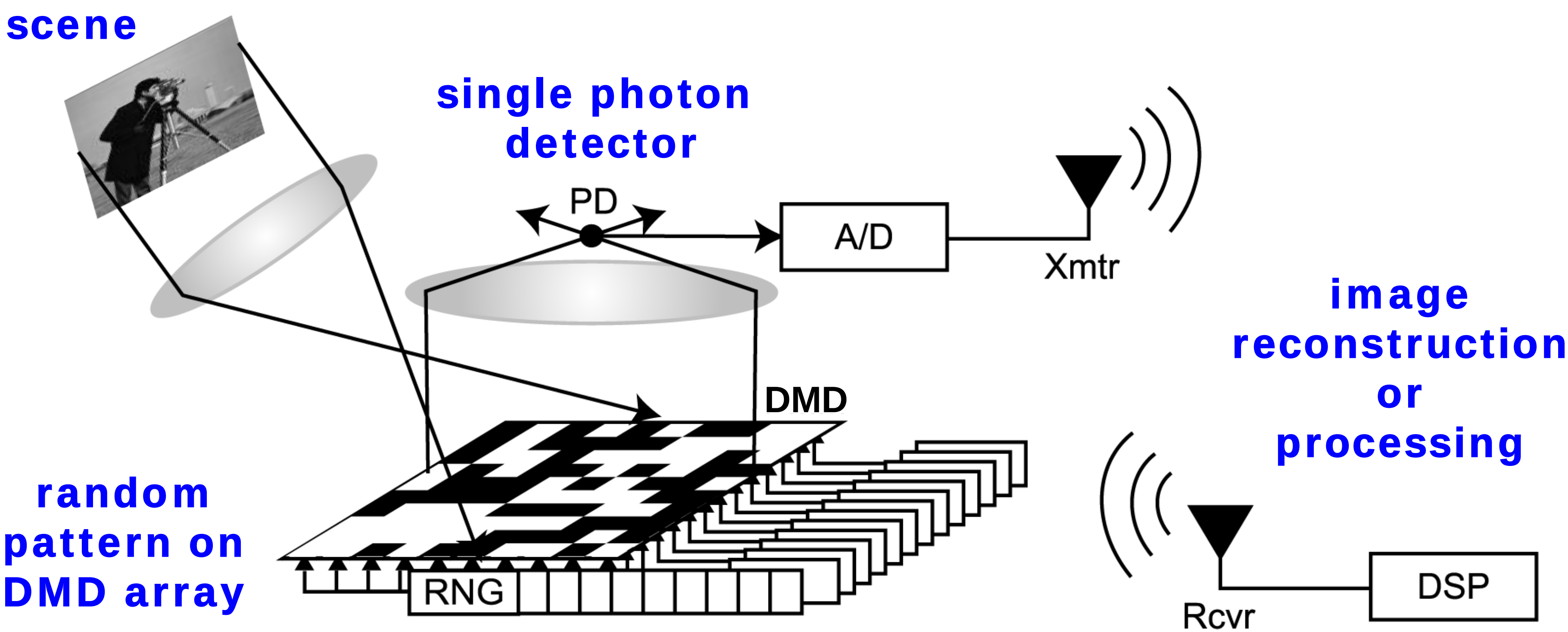
- Except at each iteration:
  - remove selected column $\phi_i$
  - re-orthogonalize the remaining columns of $\Phi$

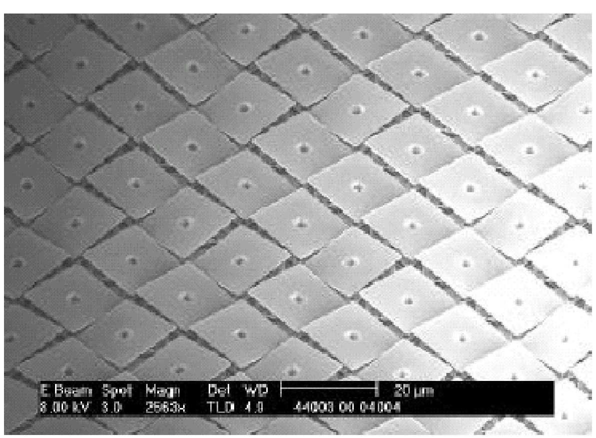
- Converges in *K* iterations

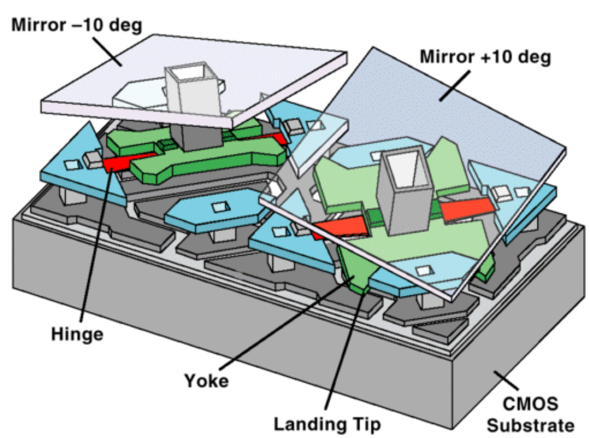
# Compressive Sensing
# *In Action*

# Cameras

# “Single-Pixel” CS Camera



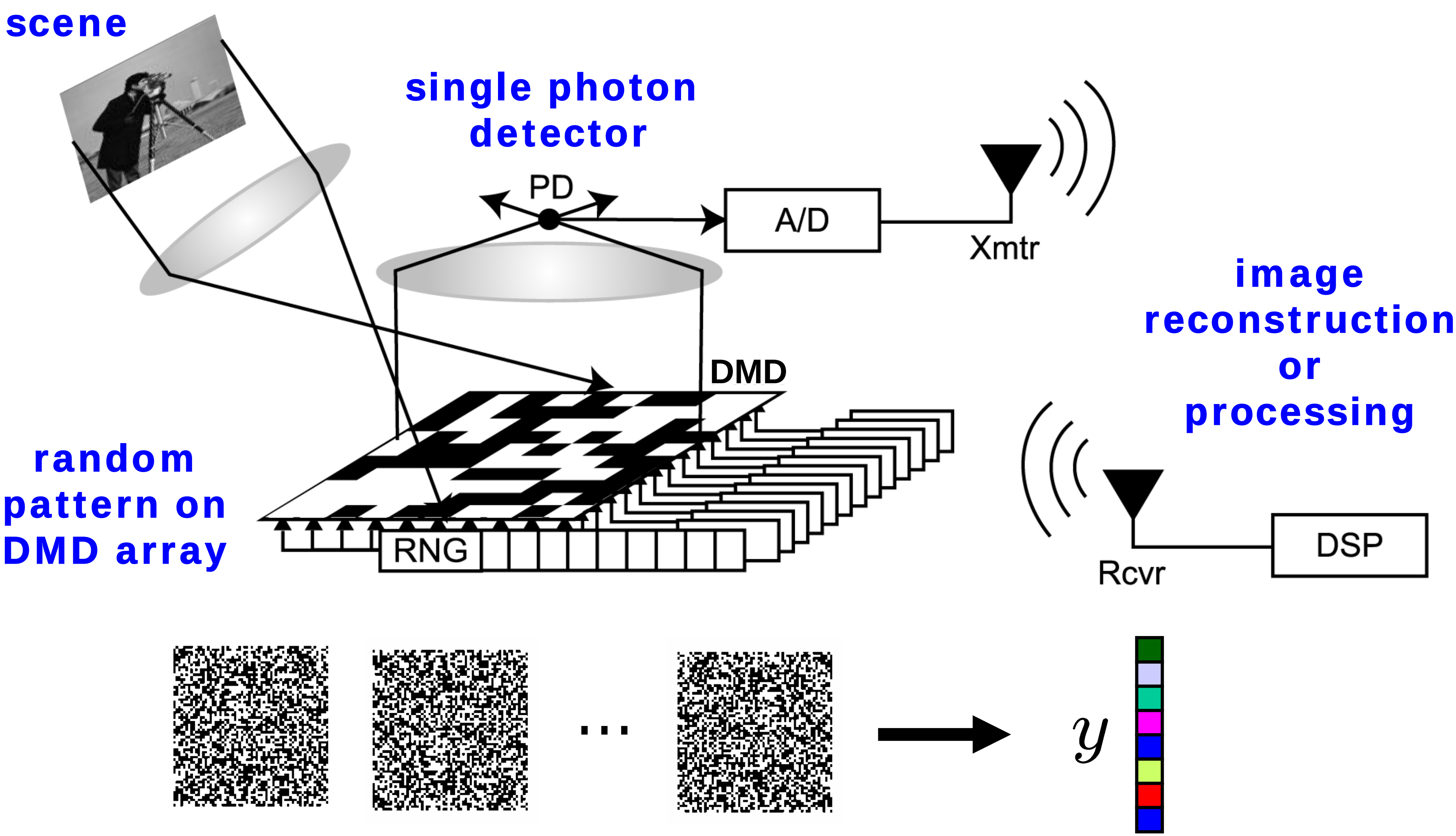






DLP
TEXAS INSTRUMENTS

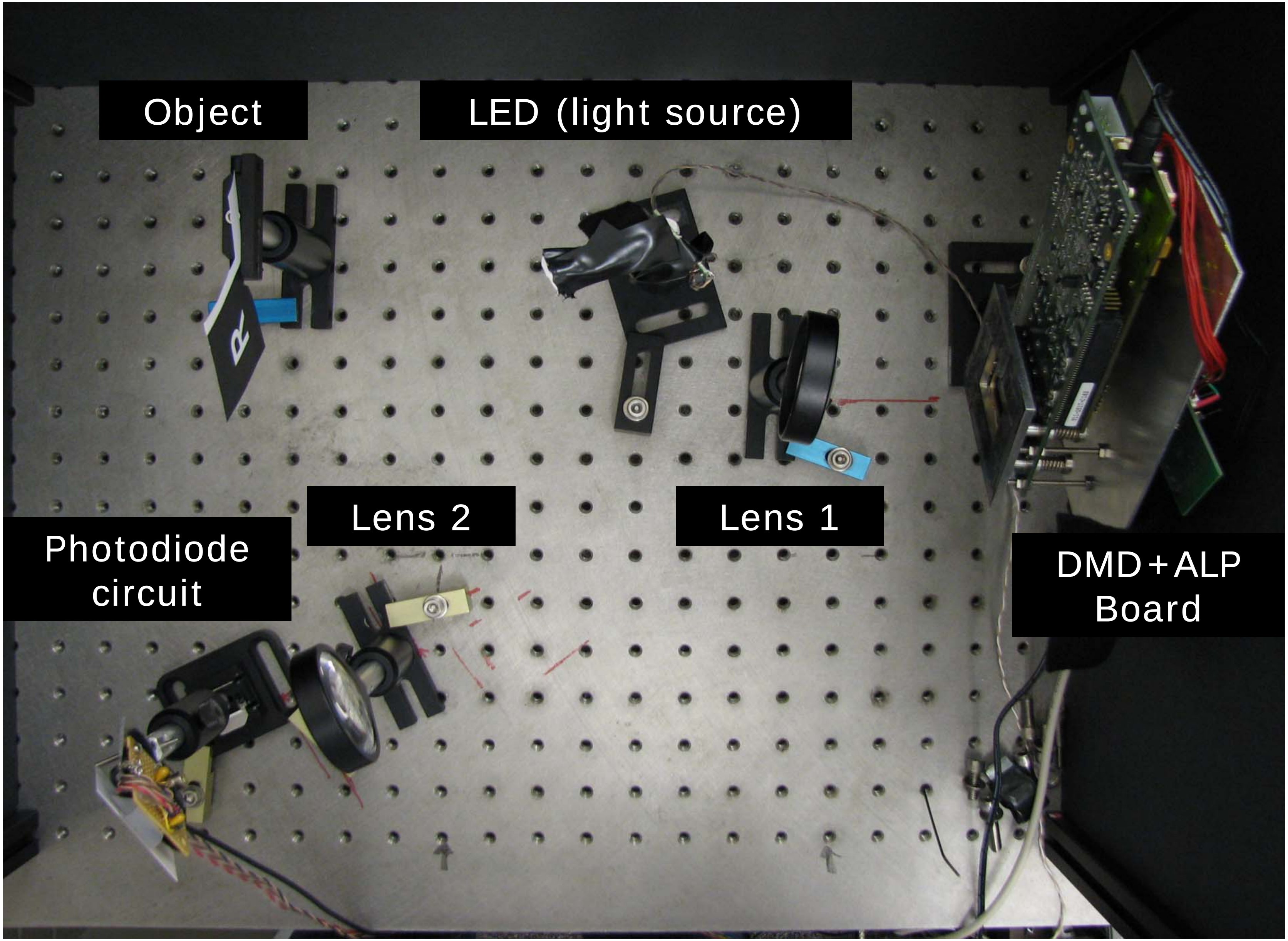
w/ Kevin Kelly

# "Single-Pixel" CS Camera

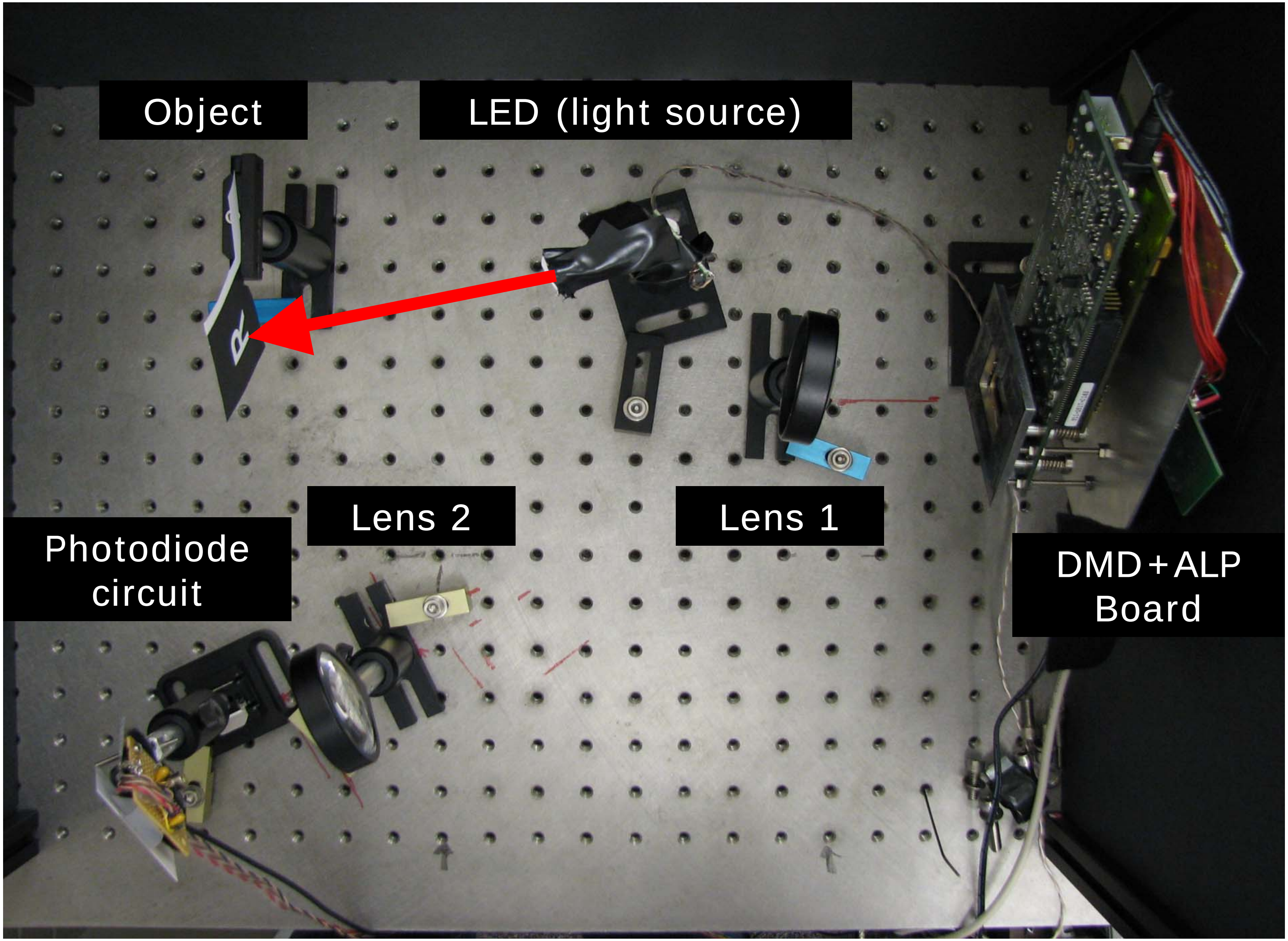



- Flip mirror array *M* times to acquire *M* measurements
- Sparsity-based (linear programming) recovery

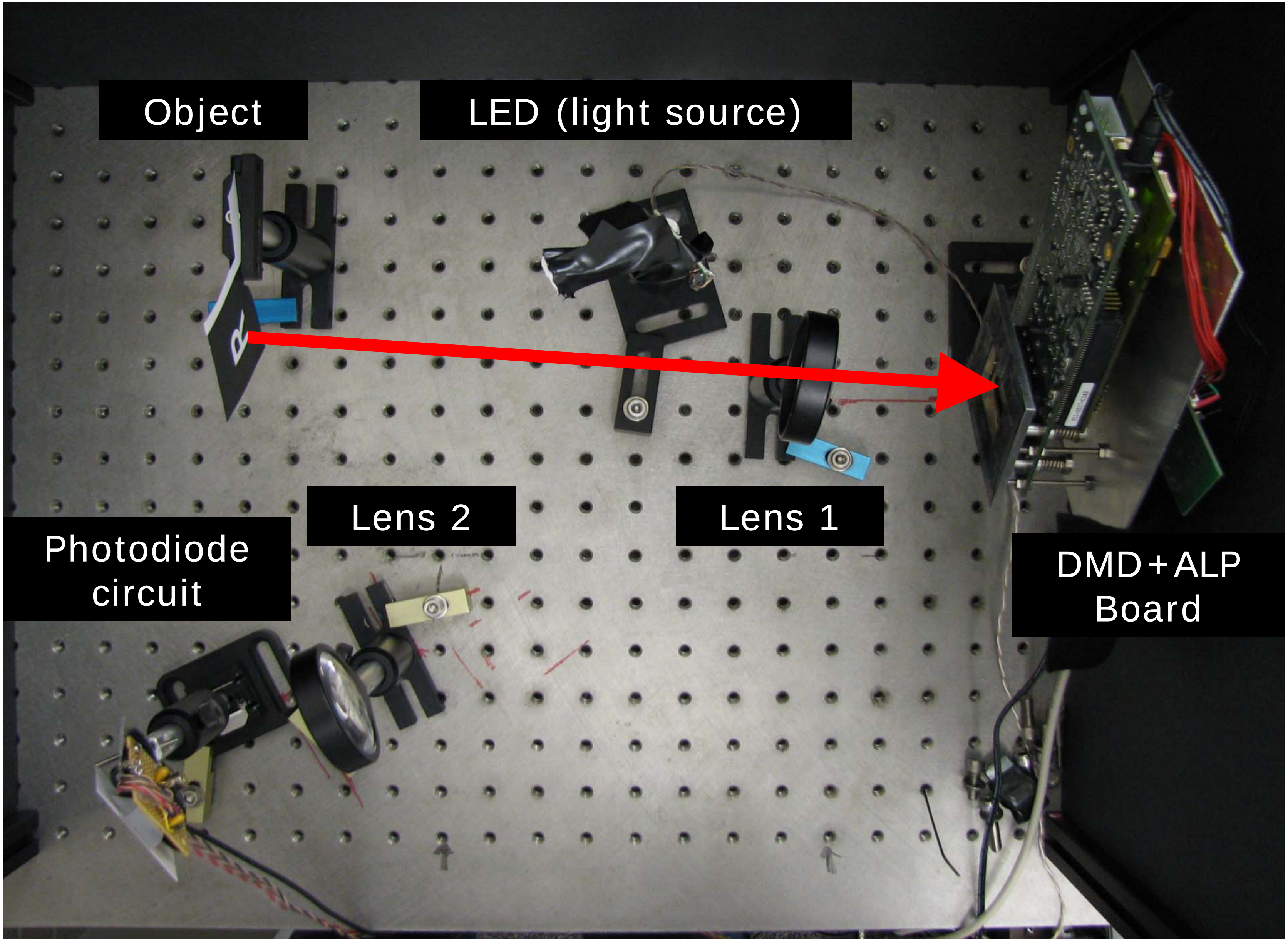
# Single Pixel Camera



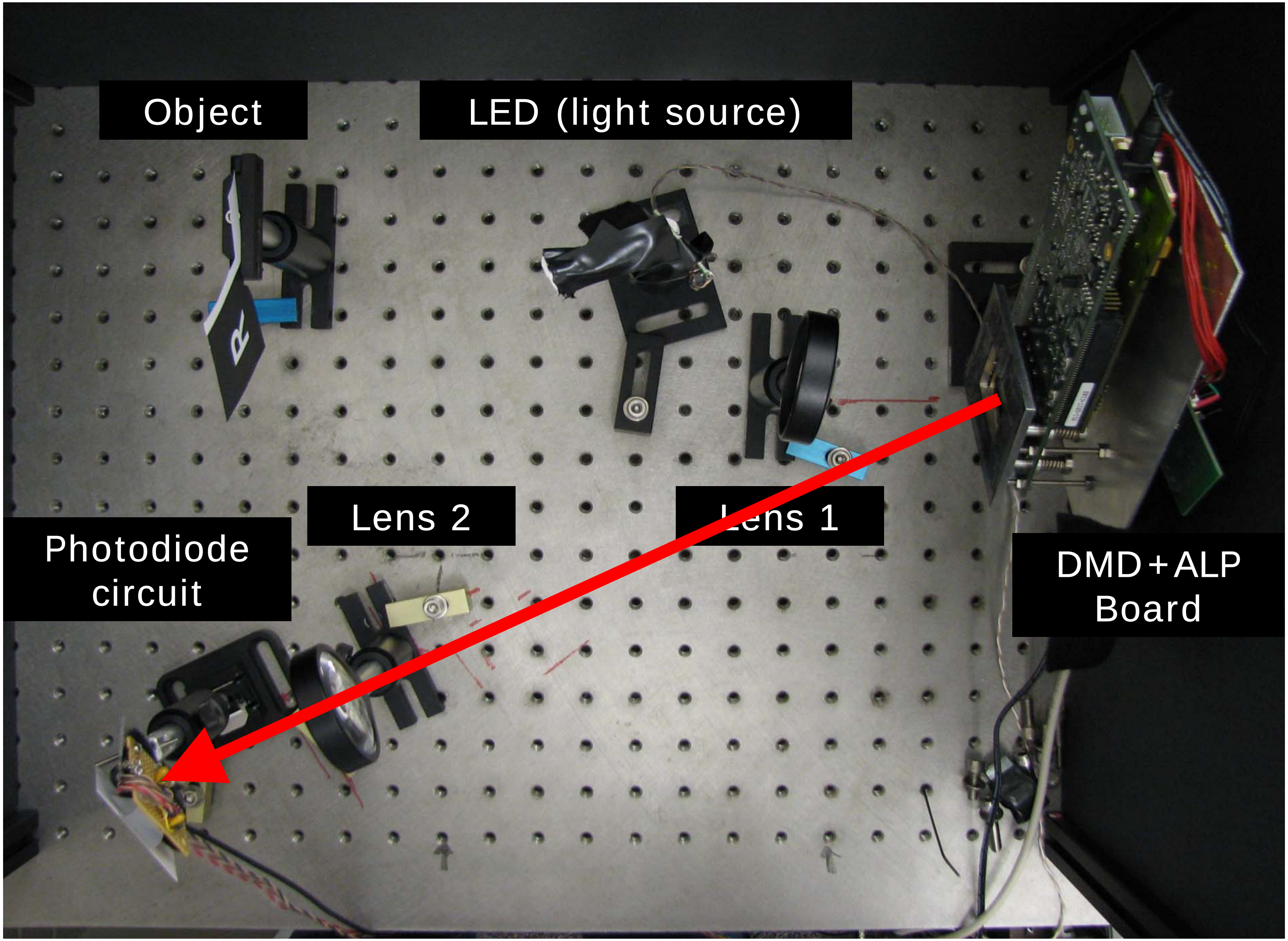
# Single Pixel Camera



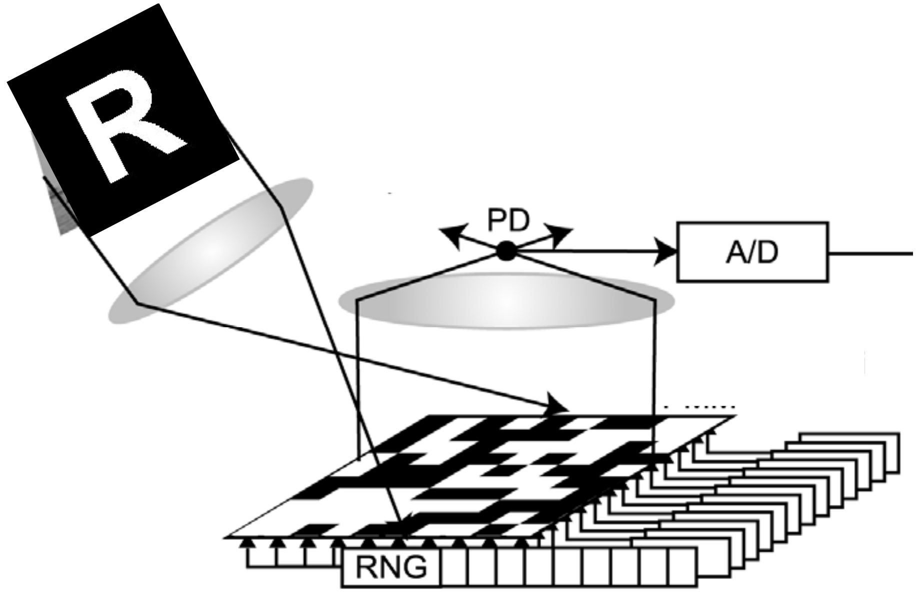
# Single Pixel Camera



# Single Pixel Camera

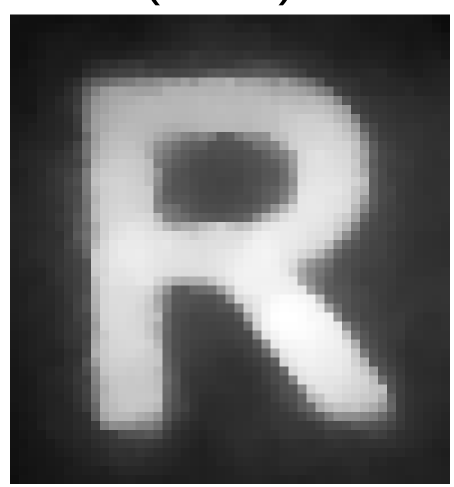


# First Image Acquisition




target
65536 pixels

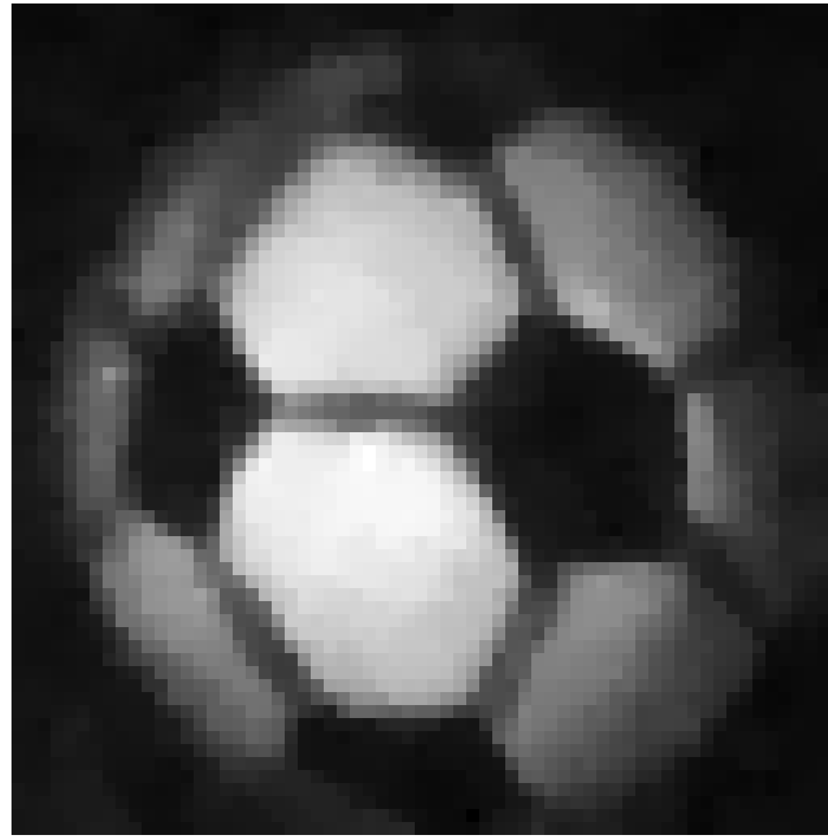
11000 measurements
(16%)



1300 measurements
(2%)

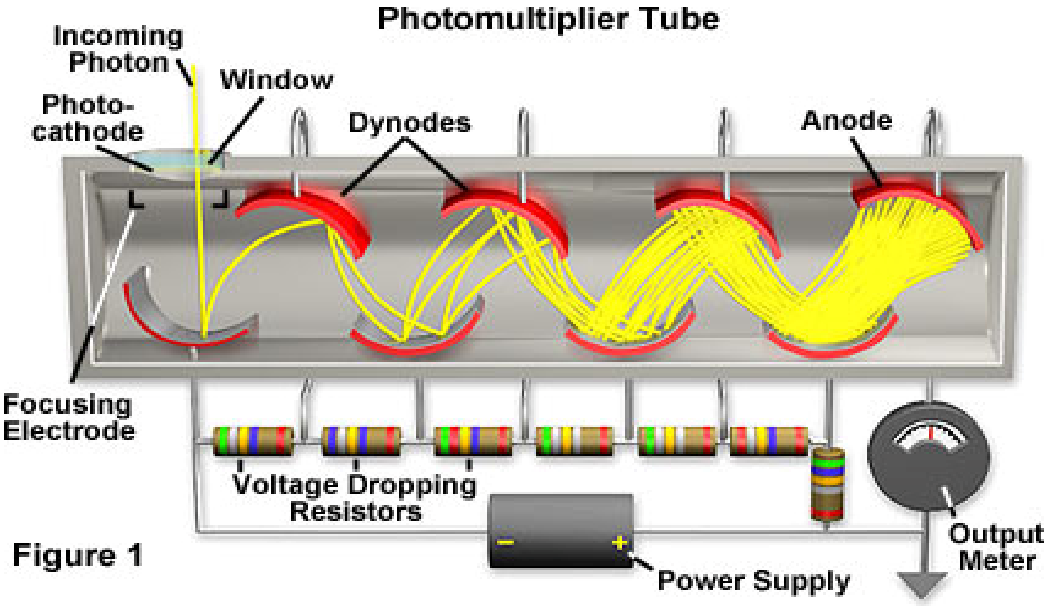
# Second Image Acquisition

4096
pixels



500
random measurements

# CS Low-Light Imaging with PMT

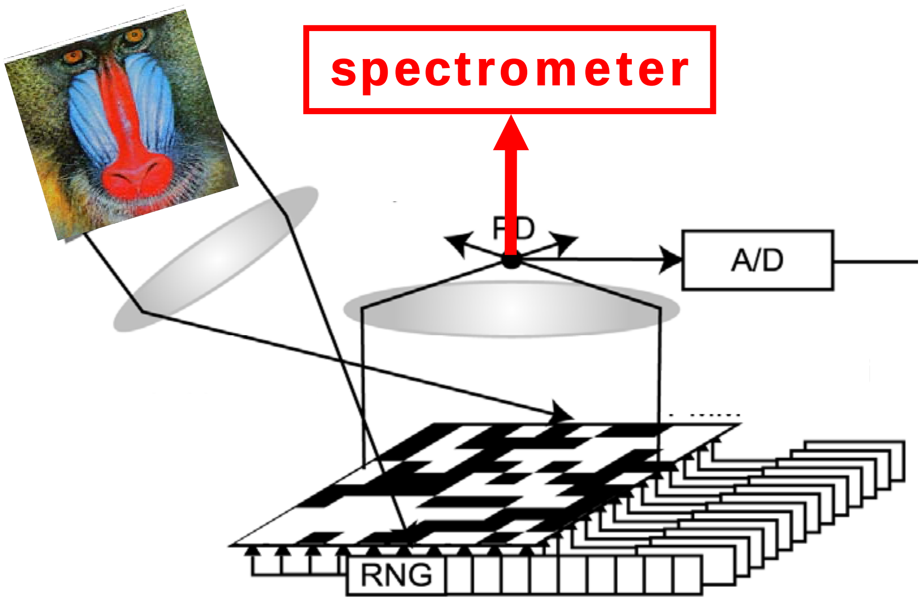



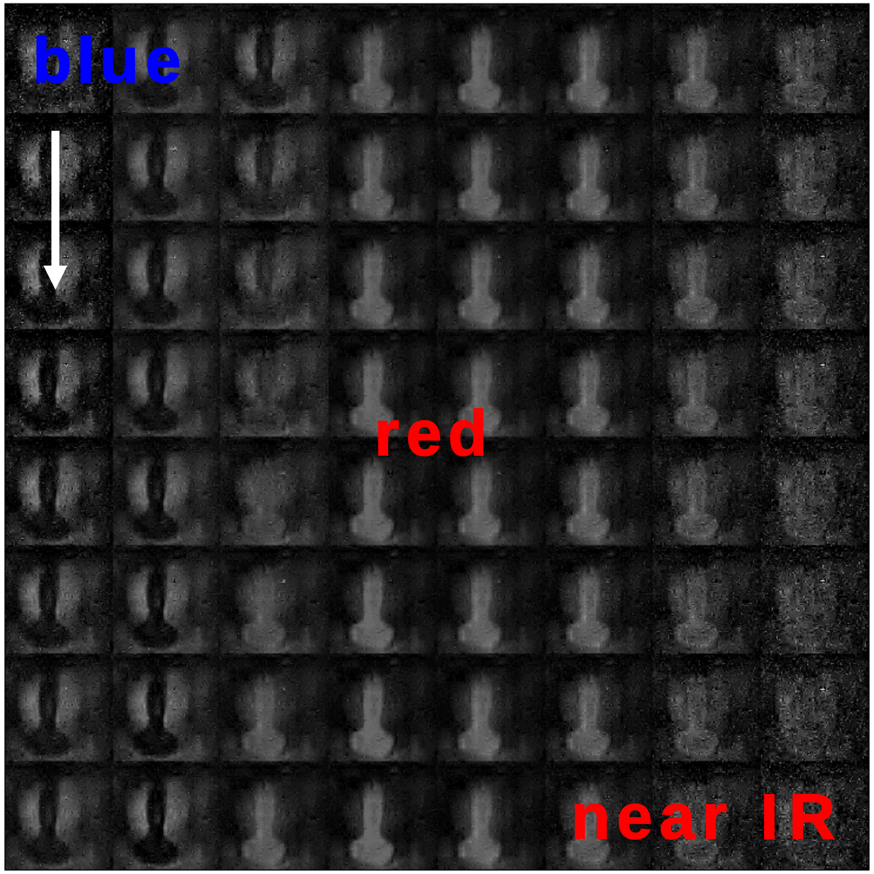

true color low-light imaging

256 x 256 image with 10:1 compression

[Nature Photonics, April 2007]

# Hyperspectral Imaging



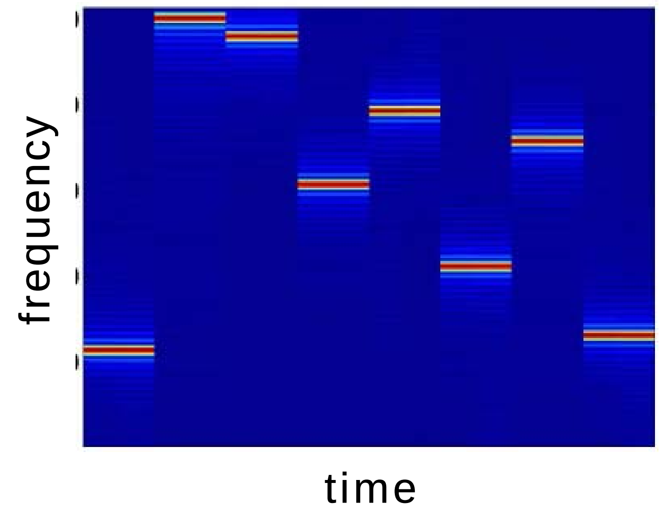
# Compressive Sensing
# *In Action*

# 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



# Analog-to-*Information* Conversion

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

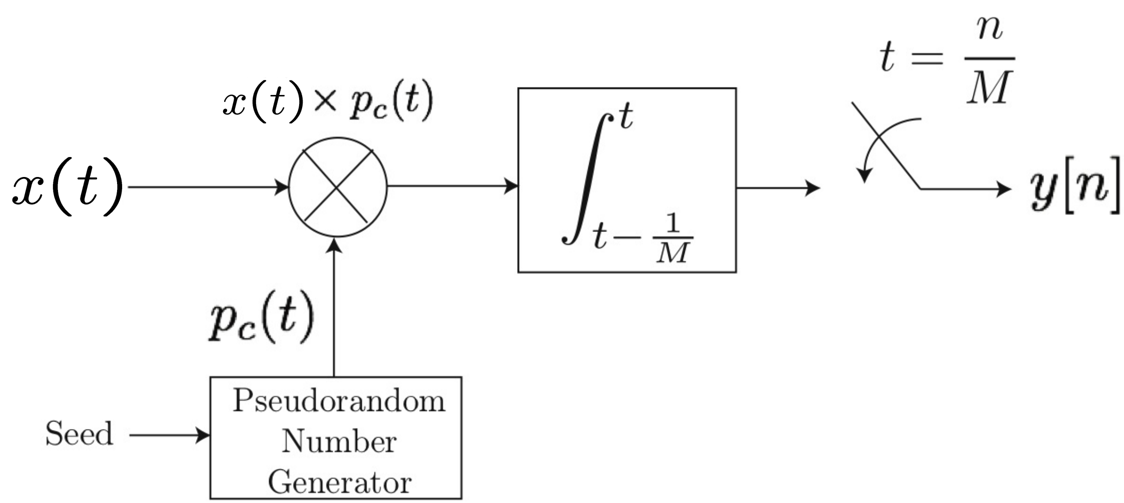
$$M = O(K \log(N/K))$$

A2I sampling rate → $M$

number of tones / window → $K$

Nyquist bandwidth → $N$

- Practical hardware: randomized demodulator (CDMA receiver)

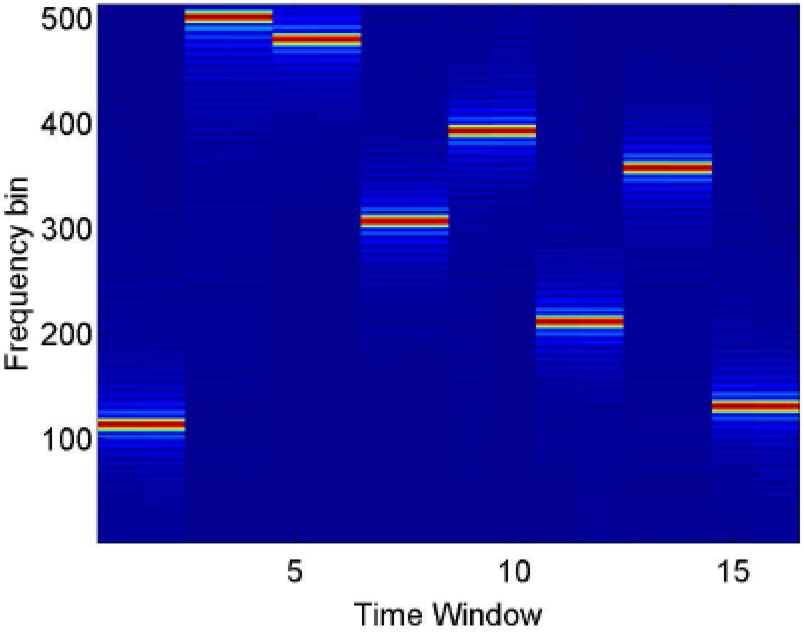


# Example: Frequency Hopper

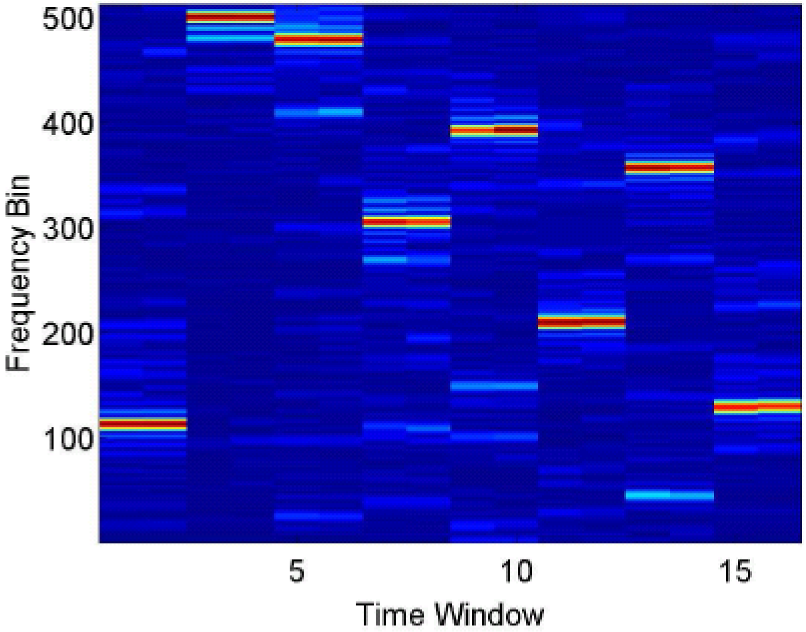
Nyquist rate sampling

20x sub-Nyquist sampling

spectrogram




sparsogram



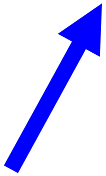
# Compressive Sensing
# *In Action*

# 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*

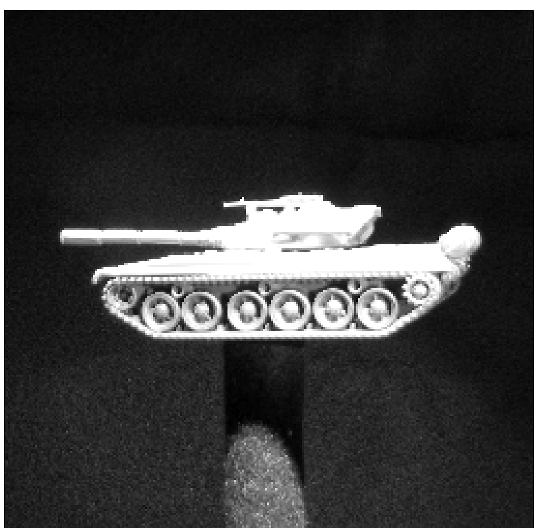
  **detection < classification < estimation < reconstruction**

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

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

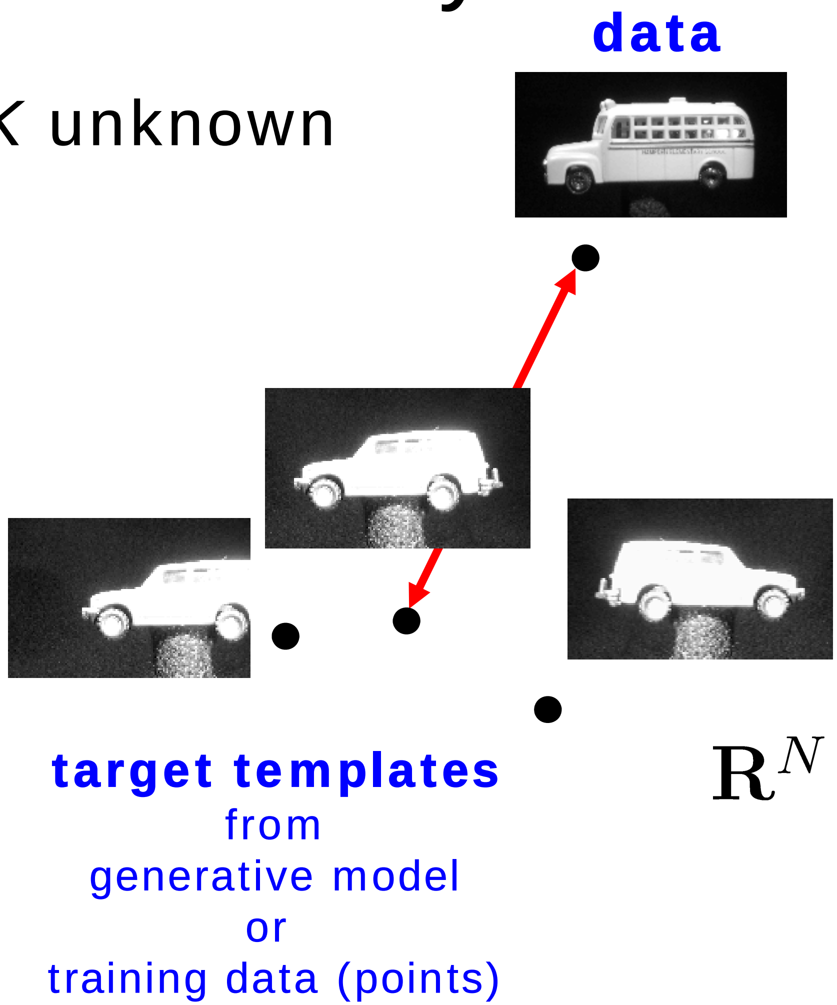
# Matched Filter

- 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



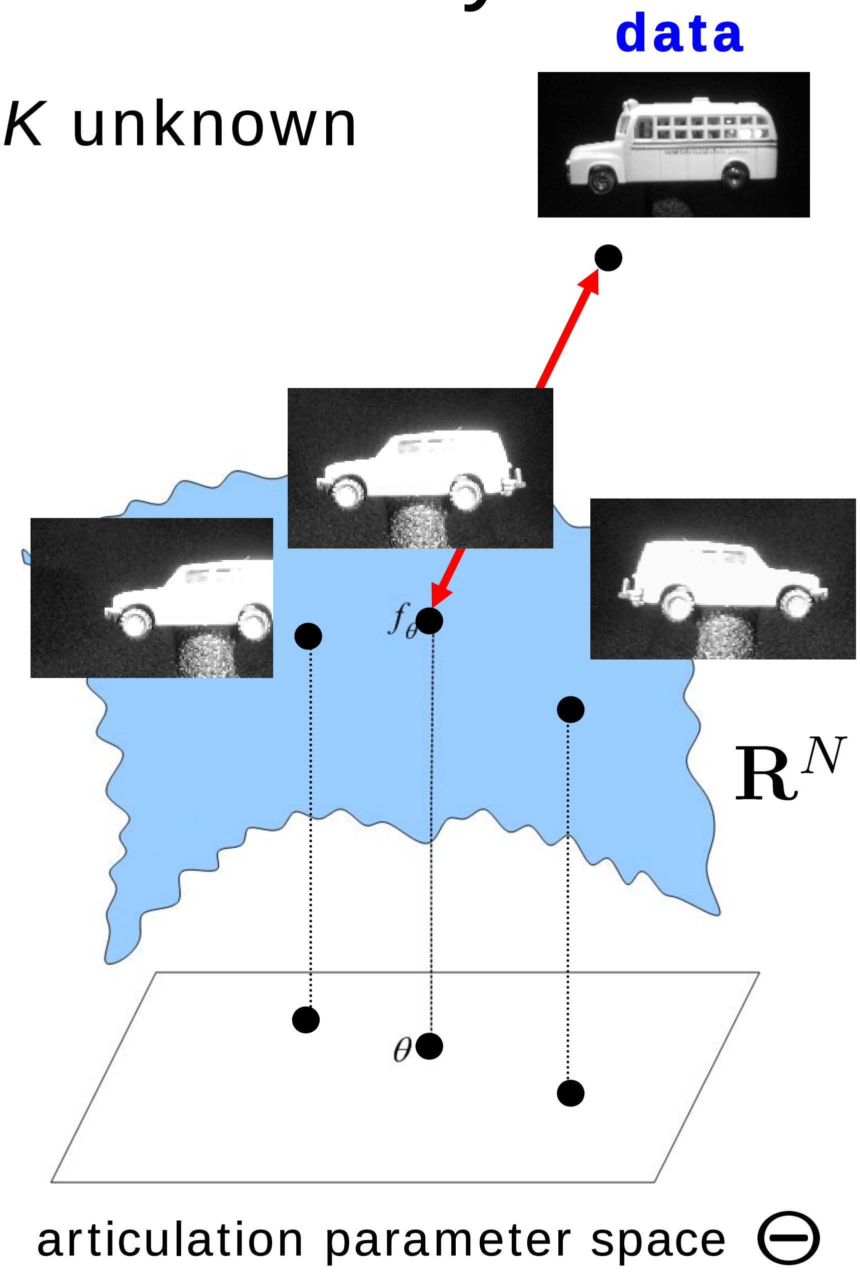
# Matched Filter Geometry

- 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**



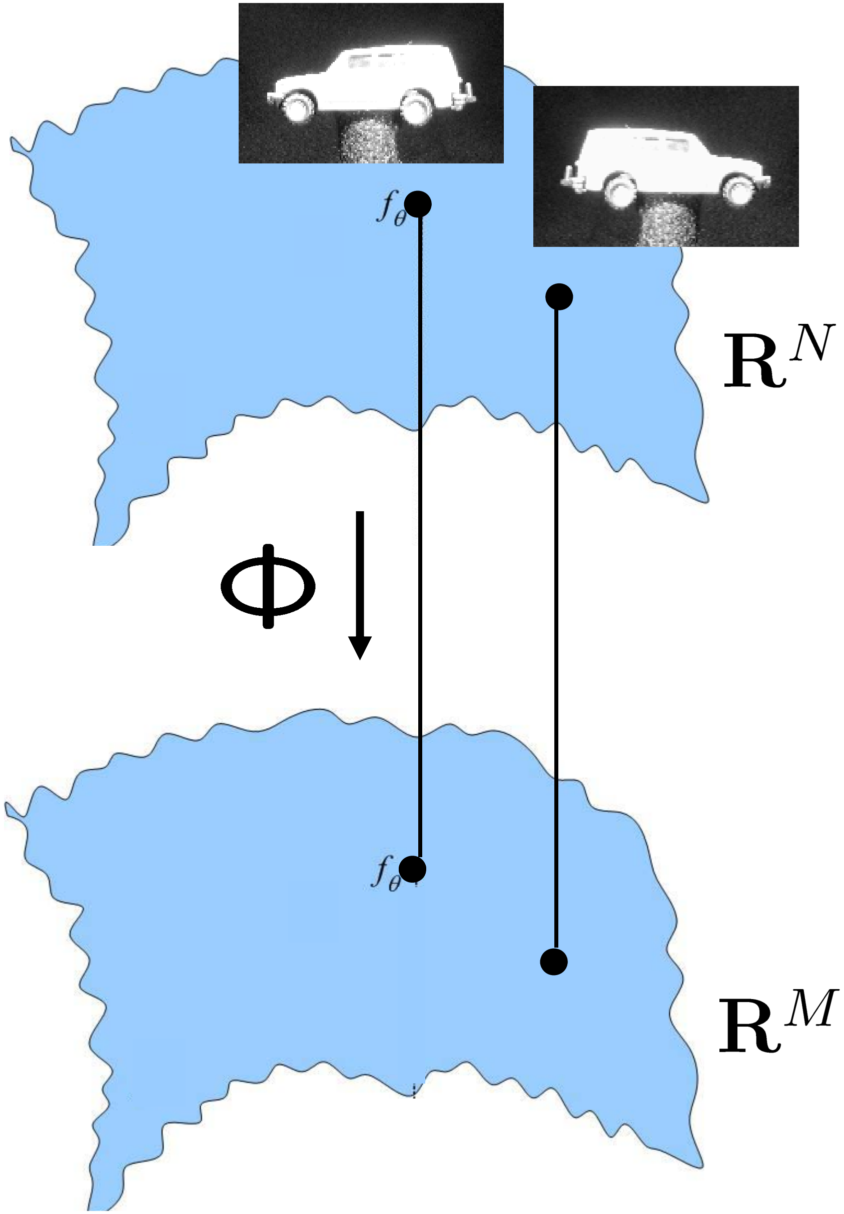
# CS for Manifolds

- **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]

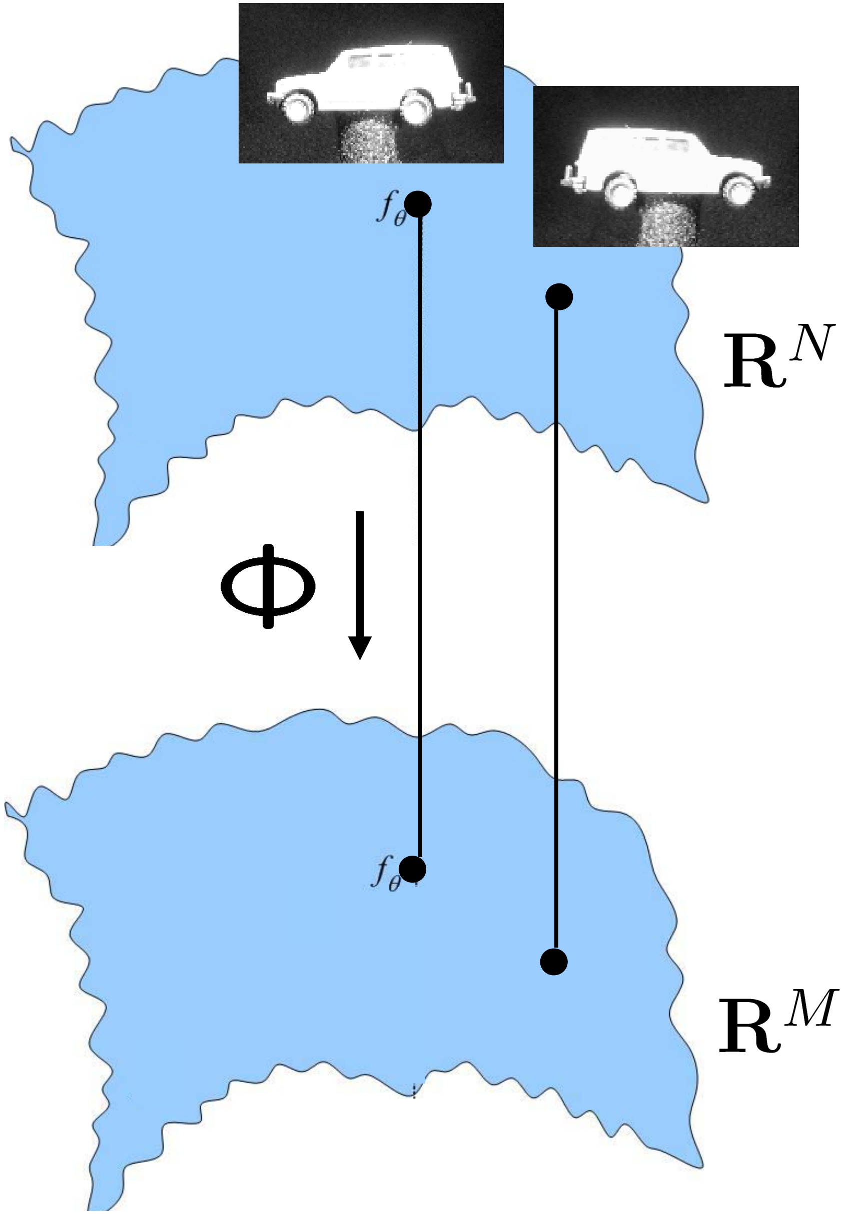
- Stable embedding

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



# CS for Manifolds

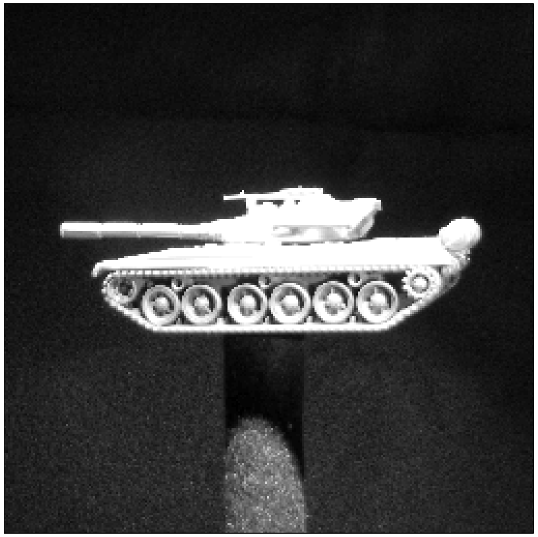
- **Theorem:**
  $$M = O(K \log N)$$
  random measurements stably embed manifold whp

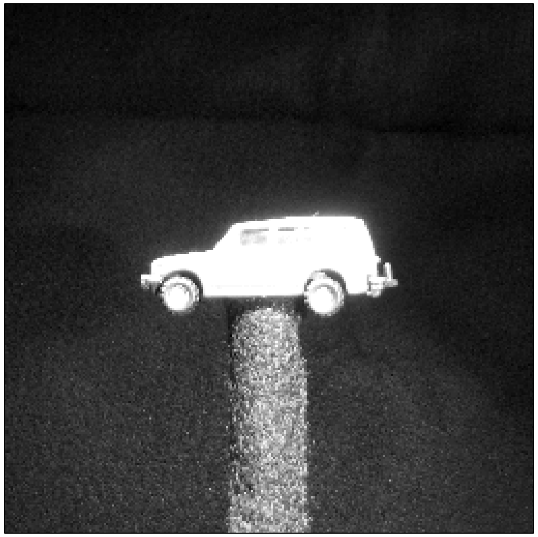



- Enables parameter estimation and MF detection/classification **directly on compressive measurements**
  - *K* very small in many applications (# articulations)

# Example: Matched Filter

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

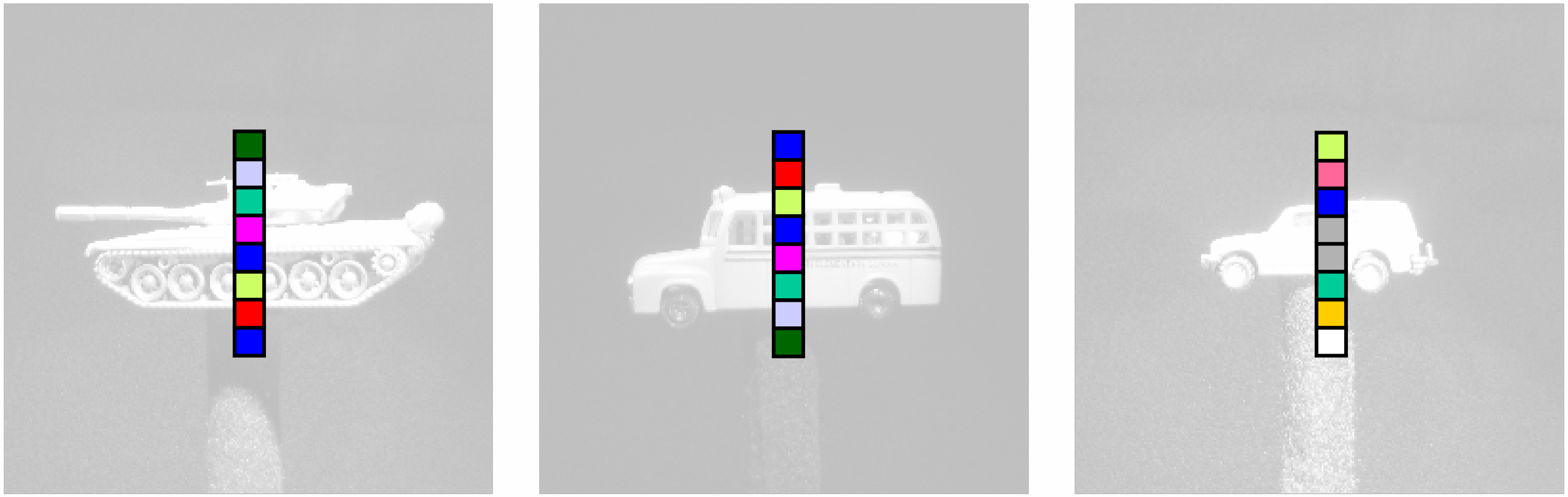




# *Smashed* Filter

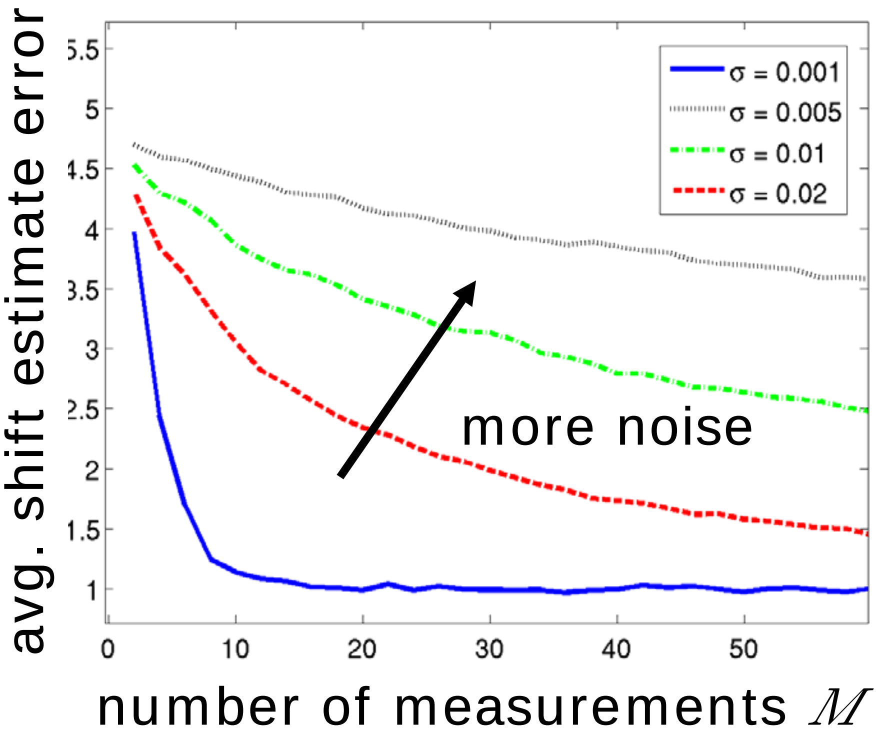
- Detection/classification with $K$=3 unknown articulation parameters **(manifold structure)**
- Dimensionally reduced matched filter directly on compressive measurements

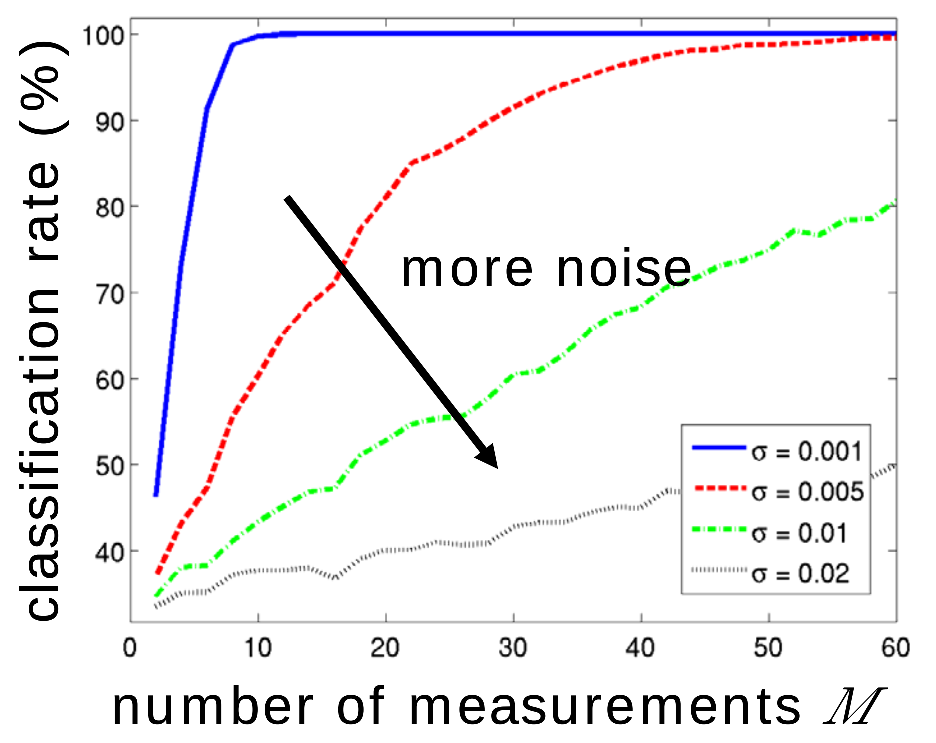
$$M = O(K \log N)$$



# 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)

# After the Break

Beyond Sparsity with structured sparsity models.

volkan@rice.edu

dsp.rice.edu/cs