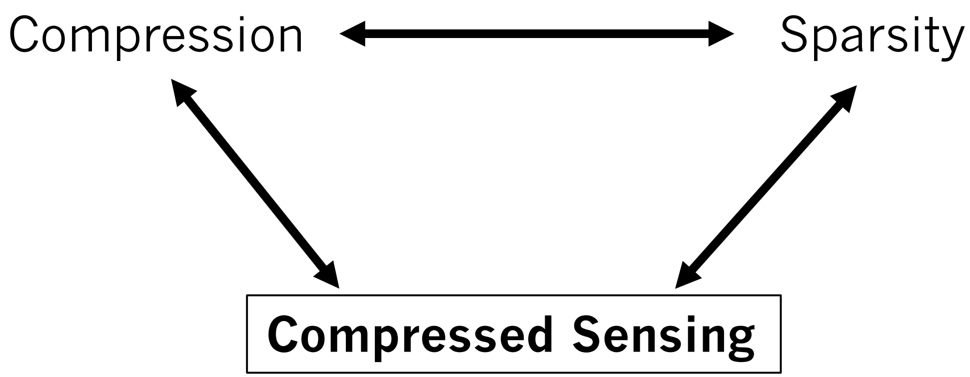
# Compressed sensing

Computational Photography

[w/ material from Richard Baraniuk / Rice University]

universitätbonn

# Compressed Sensing: Basic Idea




It’s all about encoding information (e.g images) in low dimensional representations such that recovery is perfectly possible

# Compression

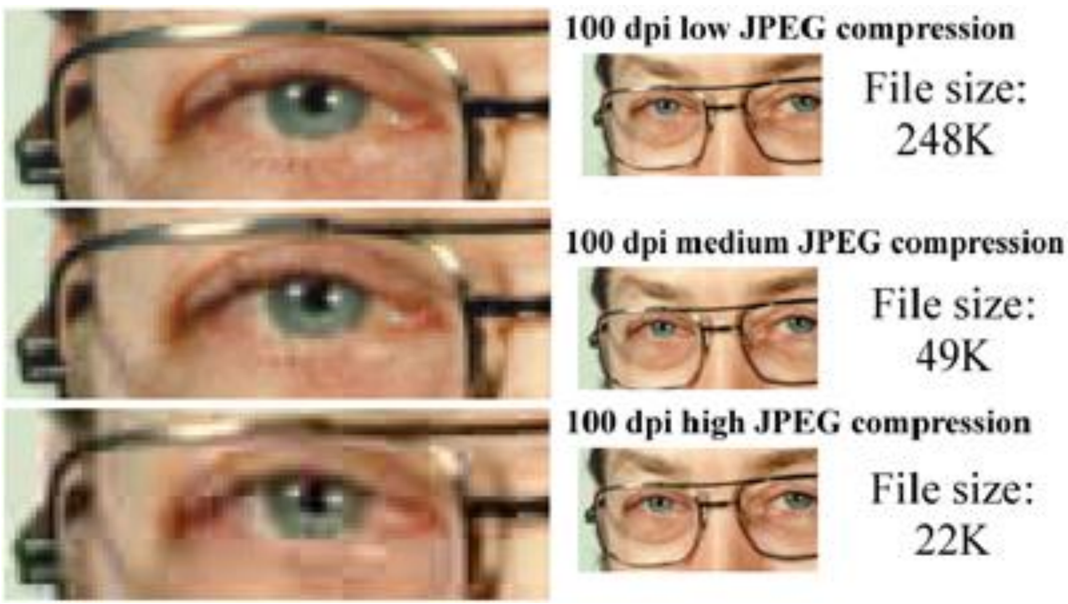
Original – 2.4 MB

Compressed 10x
257 KB

Compressed 20x
122 KB

universitätbonn

# Compression

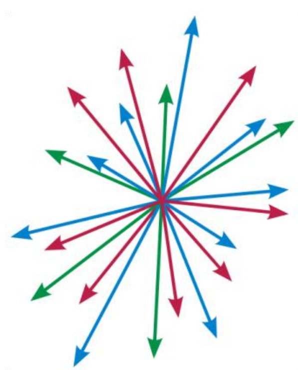


# Basics

How do we *compress* a chunk of information?

Express that chunk of information as a linear combination of some basic blocks and retain only those that are most prominent

$$x_1 \underset{V_1}{\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}} + x_2 \underset{V_2}{\begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}} + \cdots + x_n \underset{V_n}{\begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$



# Basics

- Images are information. Moreover, they are explicit functions (or signals) defined over all (i,j) pixels of its whole extension.

- Since images are signals, they can be also arranged in 1D-vectors by attaching row-after-row. This is highly convenient as it allows us to use all our linear algebra intuitions and tools

- Thus, the goal of imaging is to express an initial set of data using some vector basis

$$f(x) = \sum_{i=1}^{m} c_i \psi_i(x)$$

# Basics

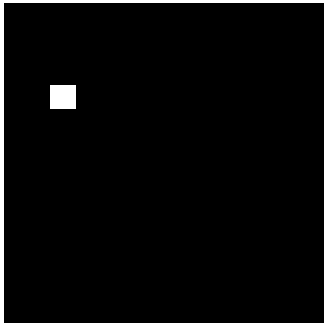
- Instead of pixels, take *linear measurements*

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

$$y = \Phi f$$

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

- Example: big pixels

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

Note: Here, the $y_i$ coefficients are the same as the $c_i$ in the previous slide

# Basics

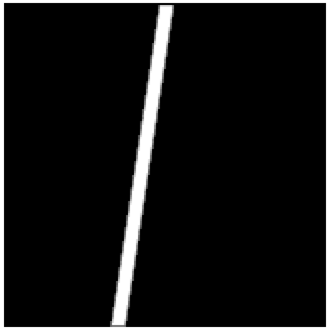
- Instead of pixels, take *linear measurements*

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

$$y = \Phi f$$

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

- Example: line integrals (tomography)

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

Note: Here, the $y_i$ coefficients are the same as the $c_i$ in the previous slide

# Which Basis Do We Choose?

$$y_k = \Big\langle$$  $$, \quad ? \Big\rangle$$

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

# Which Basis Do We Choose?

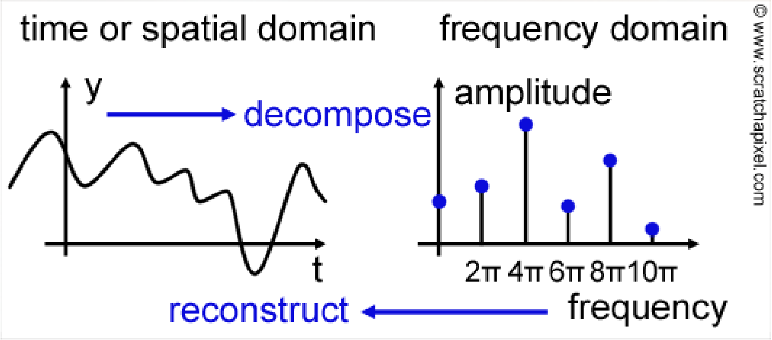
A very common basis is the discrete cosine basis (DCT), which is described by the Fourier Transform

Fourier Transform Image:

$$F(k,l) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j)\, e^{-\iota 2\pi\left(\frac{ki}{N}+\frac{lj}{N}\right)}$$

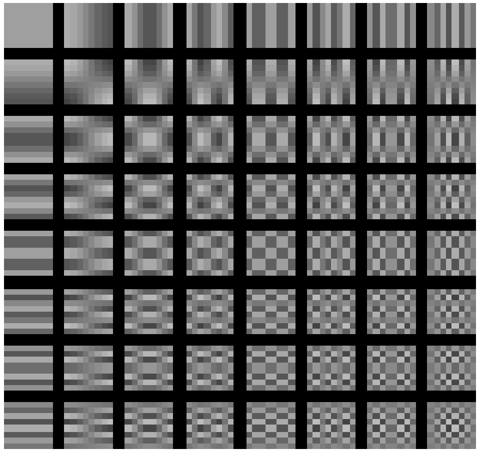
Inverse Fourier Transform Image:
(original image)

$$f(a,b) = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} F(k,l)\, e^{\iota 2\pi\left(\frac{ka}{N}+\frac{lb}{N}\right)}$$



# Which Basis Do We Choose?

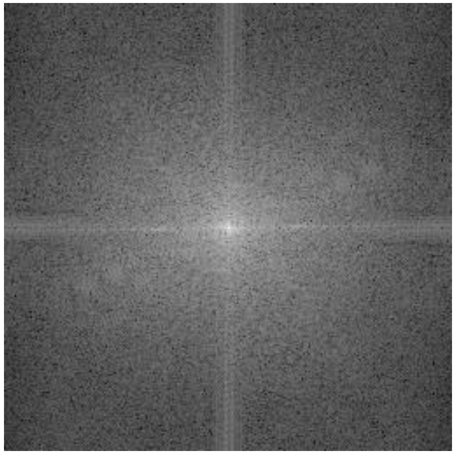
This is how the Cosine-Fourier basis looks like



Each square is a vector basis

Example

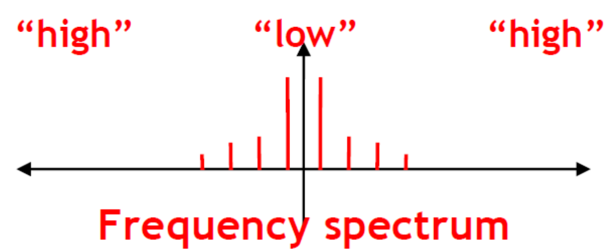
$\xrightarrow{F}$



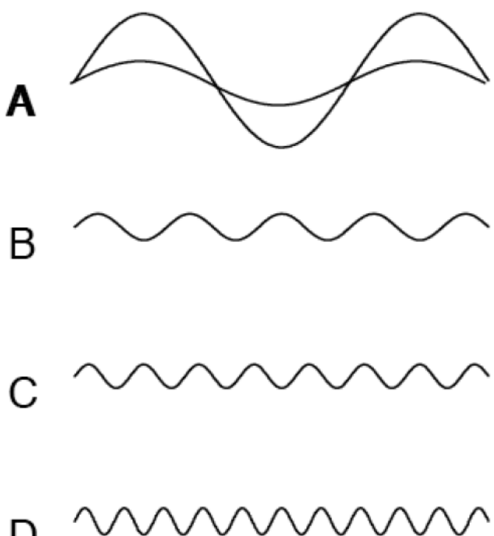
(Magnitudes)

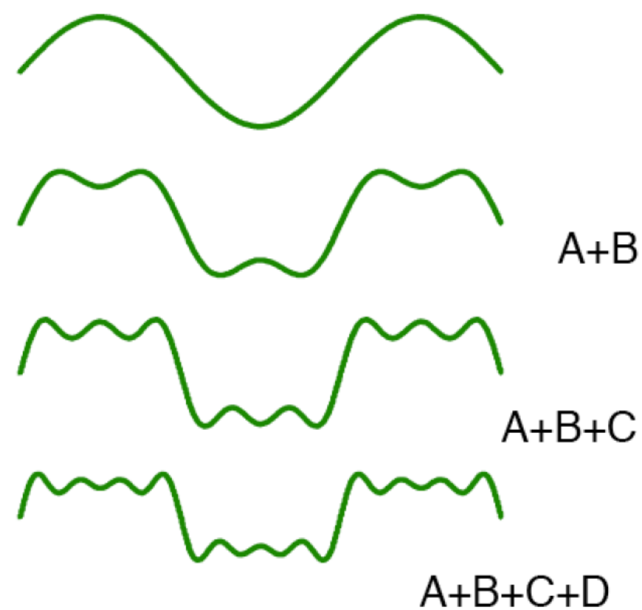
# Compression

## A small excursion into the Fourier transform to talk about spatial frequencies…




$$3\cos(x) + 1\cos(3x) + 0.8\cos(5x) + 0.4\cos(7x) + \cdots$$




Frequency coefficients

# Compression

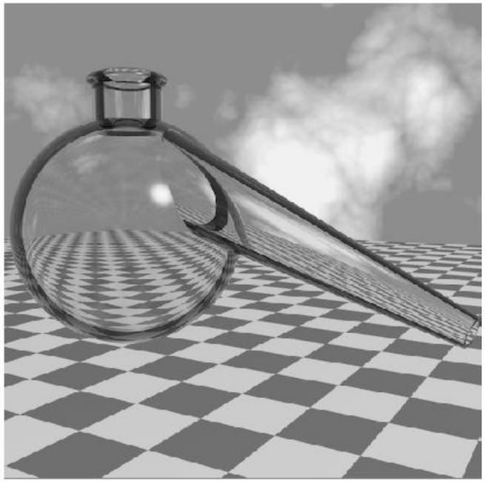
- The goal of compression is to express an initial set of data using some smaller set of data, either with or without loss of information.

$$f(x) = \sum_{i=1}^{m} c_i \psi_i(x) \longrightarrow \tilde{f}(x) = \sum_{i=1}^{\tilde{m}} \tilde{c}_i \psi_i(x)$$

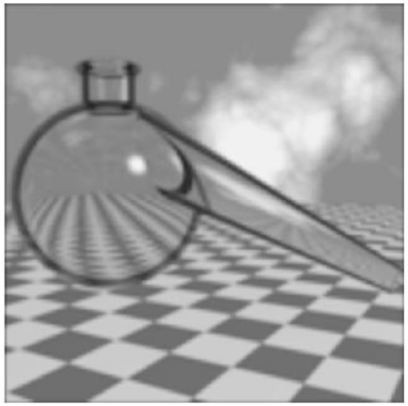
such that $\tilde{m} < m$ and $\left\| \tilde{f}(x) - f(x) \right\| < \varepsilon$ for some norm

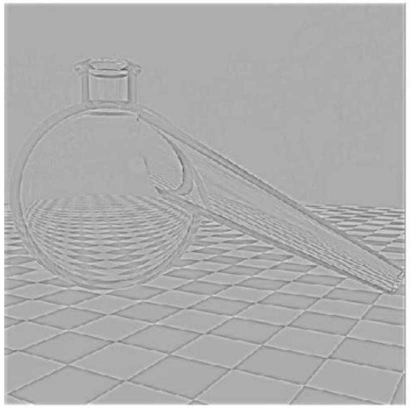
- Compression (Shrinkage)
  – set all coefficients <t to zero!

# Compression



**Original image**





A simple algorithm to compress data:

1) Take the Fourier transform of the
image

2) Remove from the expansion those
Basis vectors (frequencies) with coefficients less than a threshold

3) Compute the inverse Fourier transform

$$f(x) = \sum_{i=1}^{m} c_i \psi_i(x) \longrightarrow \underset{\tilde{m} < m}{\tilde{f}(x)} = \sum_{i=1}^{\tilde{m}} \tilde{c}_i \psi_i(x)$$

# Compression

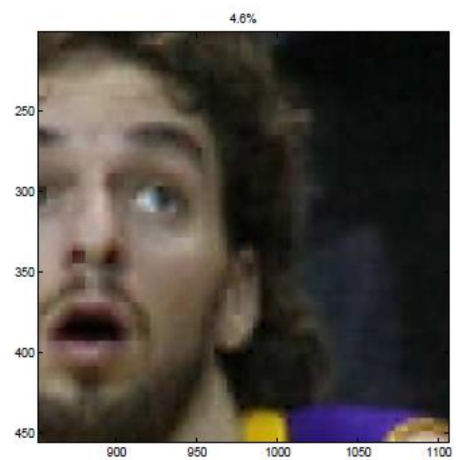
## Another Example


# Compression

Another Example: watch over 64 coefficients in the Fourier expansion

Original can be spanned using 64 coefficients since th
error |f – f’| is really small


$1/64$

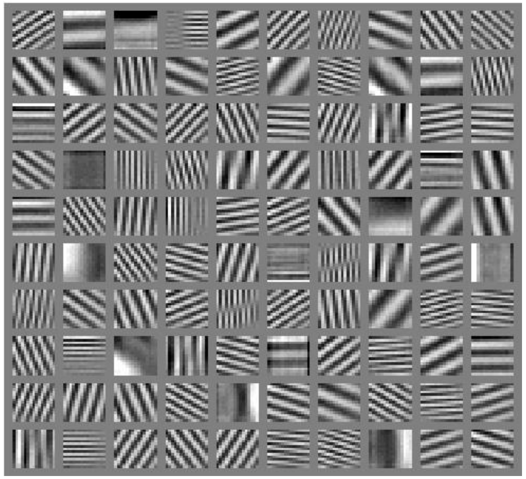


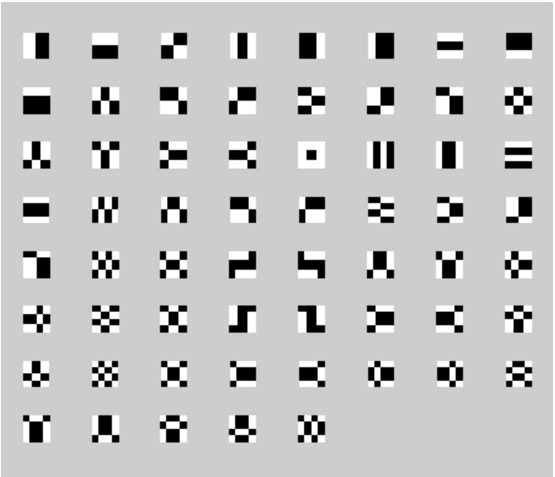

$3/64$

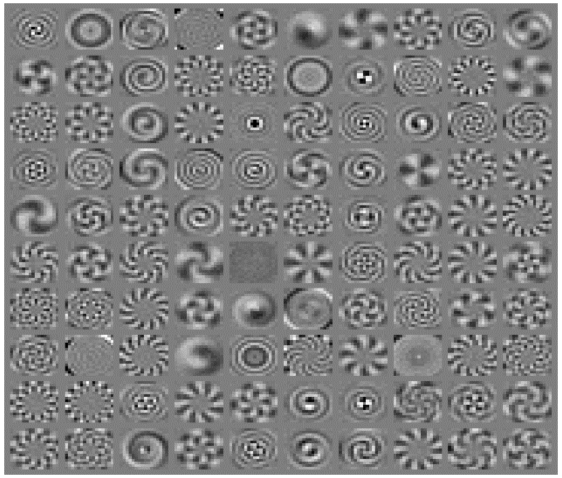

$10/64$

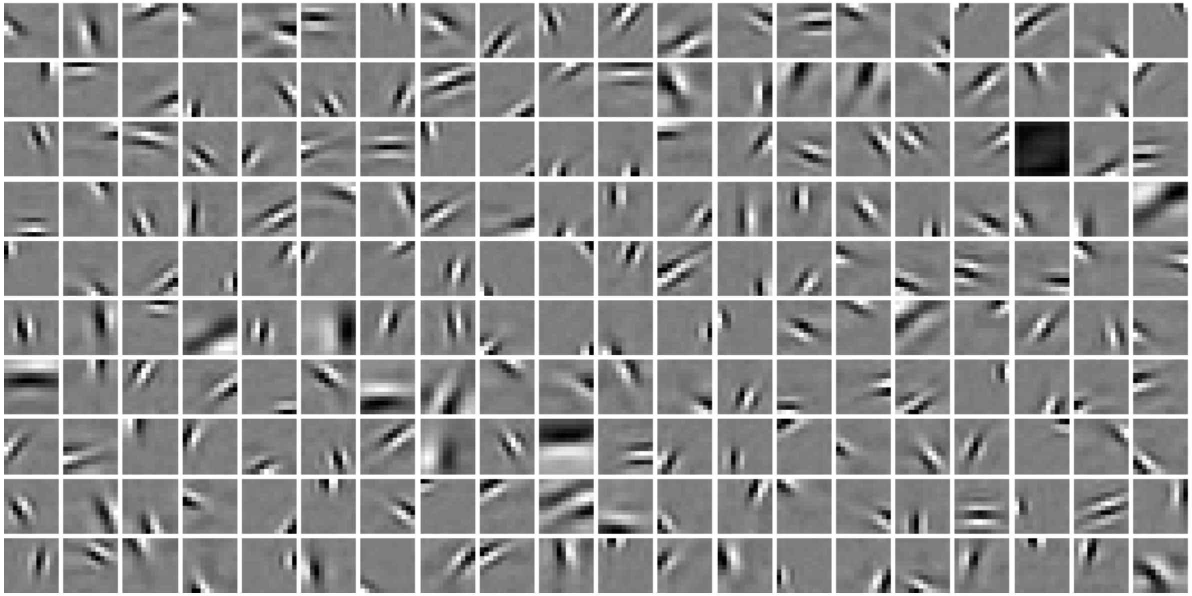
# Other Basis Functions

Analytical Bases









Learned Bases

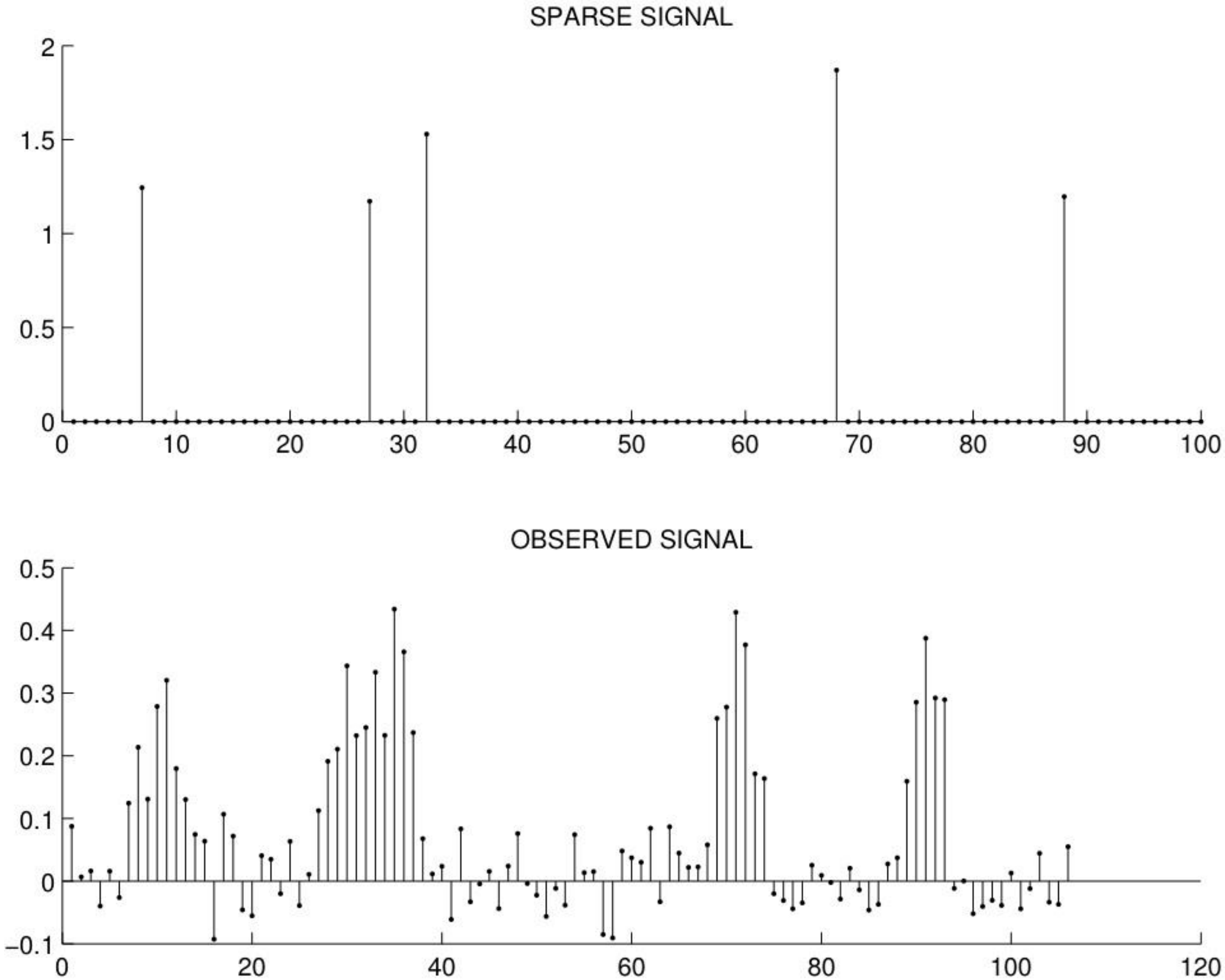
# Sparsity

The fact the information content can be simplified raises interesting questions:

1) Are some of those bases better than others?

2) Is bandwidth compression the only way to simplify information content?

# Sparsity


SPARSE SIGNAL
2
1.5
1
0.5
0
0 10 20 30 40 50 60 70 80 90 100
OBSERVED SIGNAL
0.5
0.4
0.3
0.2
0.1
0
−0.1
0 20 40 60 80 100 120

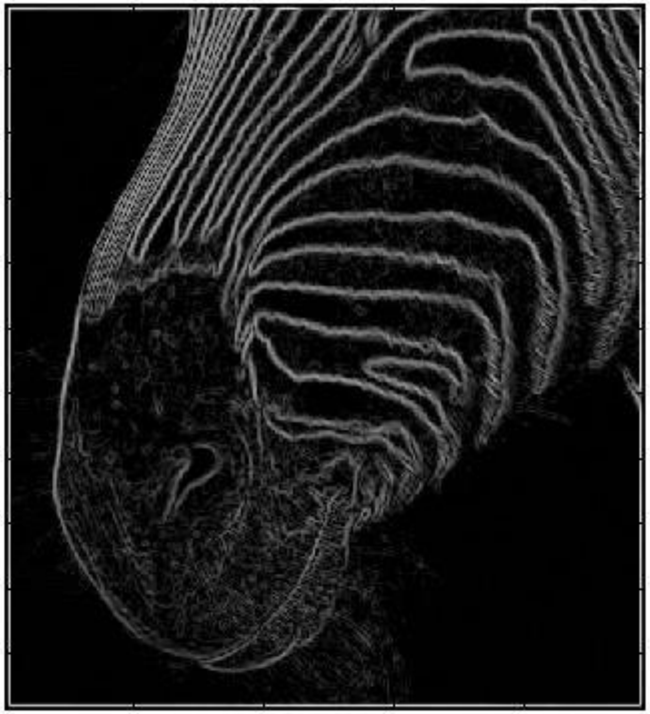
# Sparsity

**The Sparsity Assumption**: There are bases in which signals (images) manifest as sparse intensity distributions, which allows us to encode information content in the smallest amount of effort and resources

# One Type of Sparseness

- Images are sparse in the gradient domain.

$I$

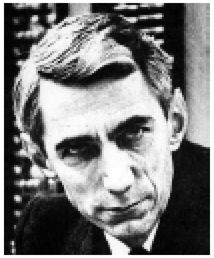


$|\nabla I|$

universitätbonn

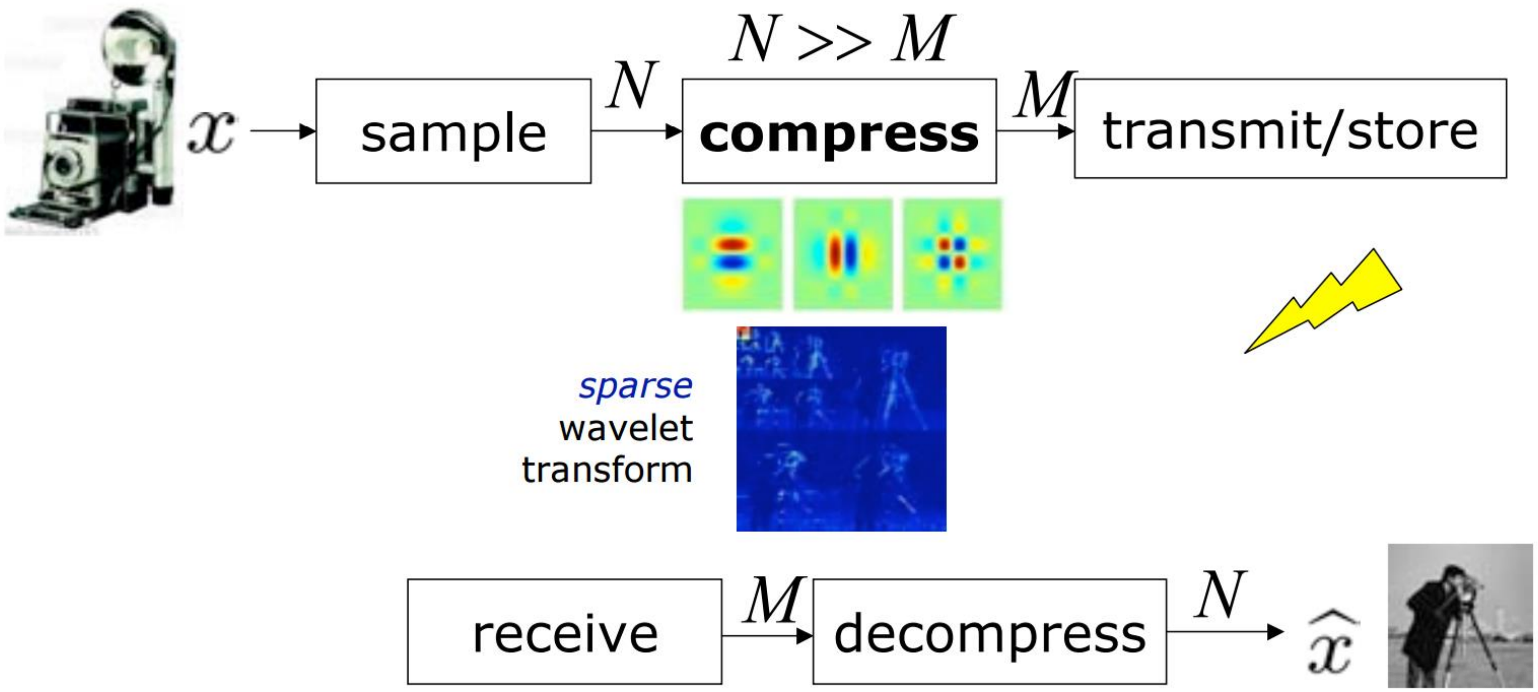
# Compressive Sensing (CS)

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



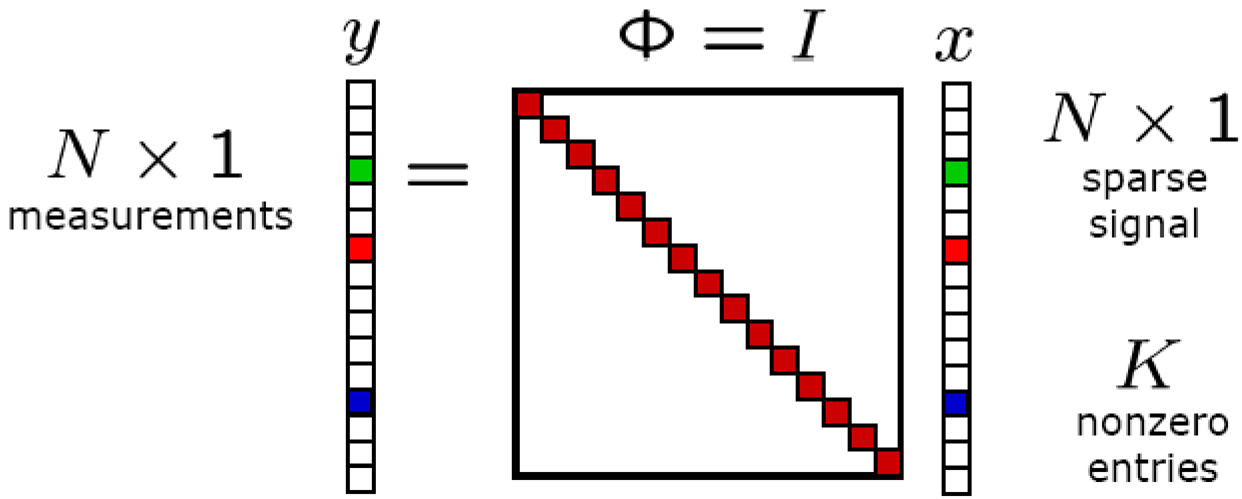
- **Compressive sensing**
  - new sampling theory that *leverages compressibility*
  - based on new *uncertainty principles*
  - *randomness* plays a key role

# Traditionally: acquire, then compress



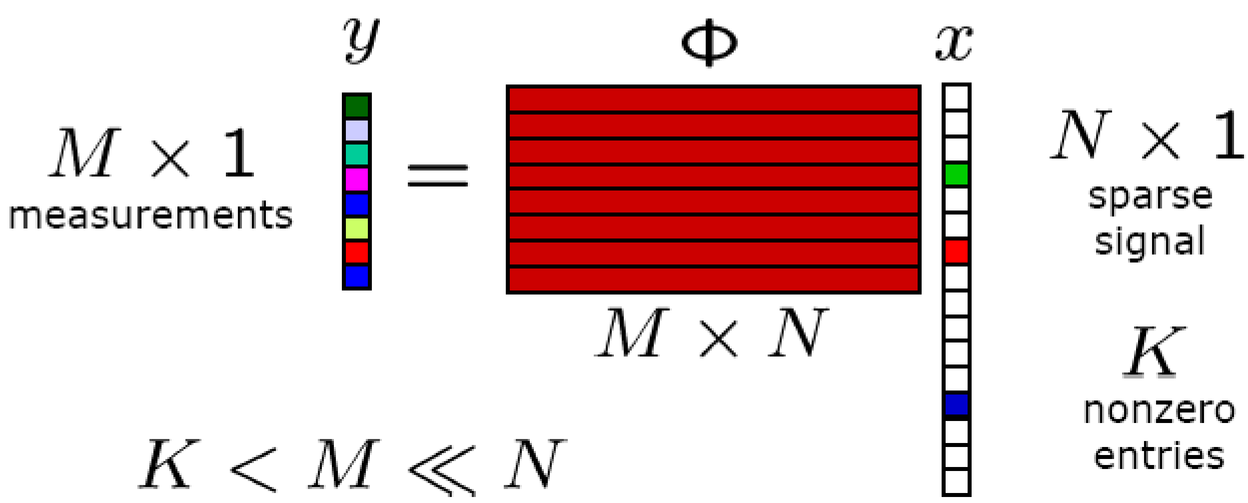
# Sampling

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



# Compressive Data Acquisition

- When data is sparse/compressible, can directly acquire a ***condensed representation*** with no/little information loss through ***dimensionality reduction*** $y = \Phi x$

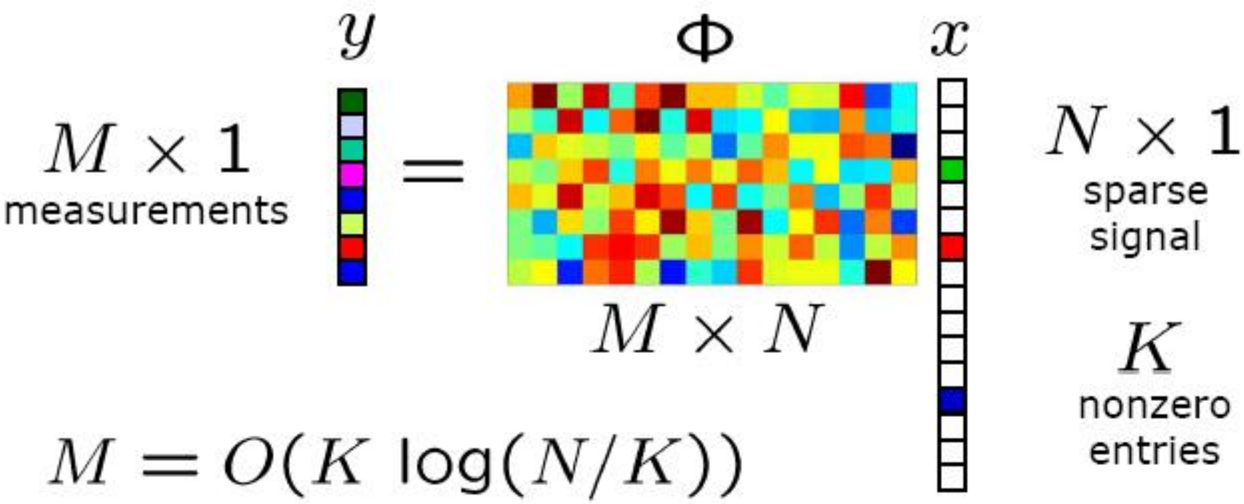


# Compressive Data Acquisition

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

$$y = \Phi x$$

- ***Random projection*** will work



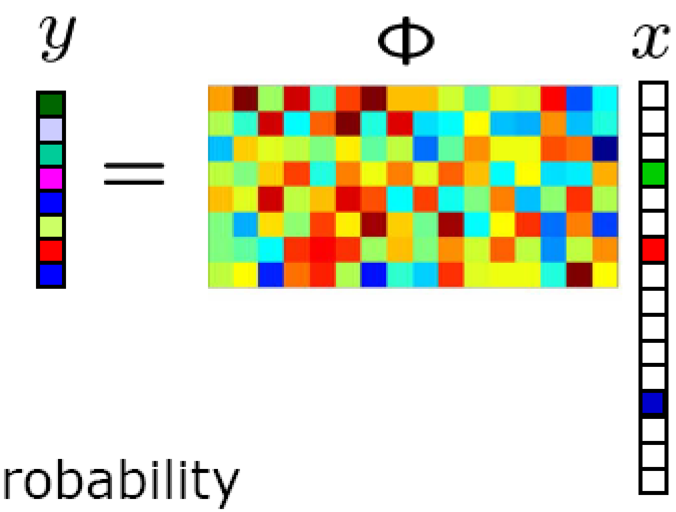

$$M = O(K \log(N/K))$$

[Candes-Romberg-Tao, Donoho, 2004]

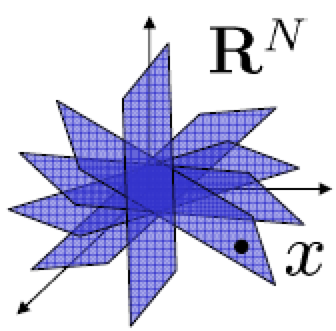
# Why does it work?

- Random projection $\Phi$
  **not full rank**...

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

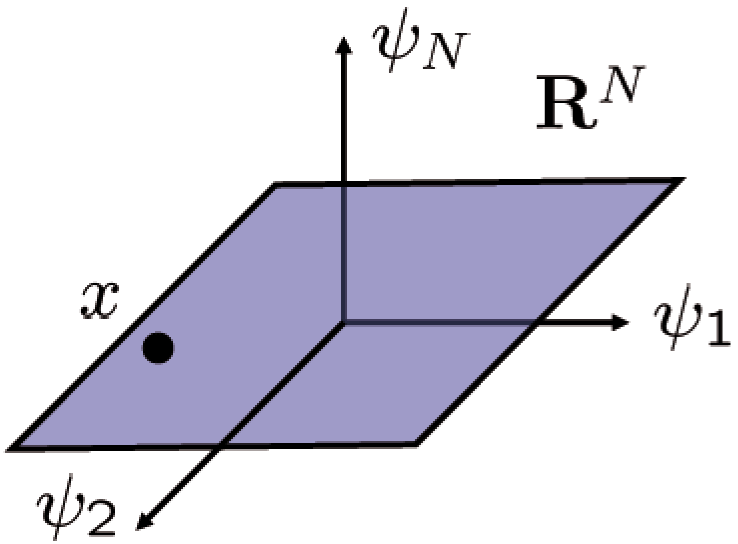


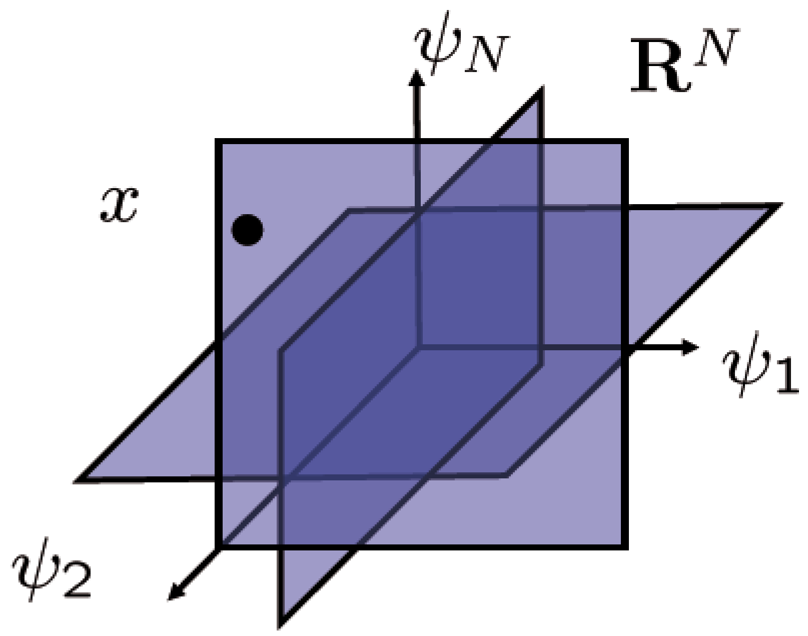

*K*-sparse model




*K-dim hyperplanes aligned with coordinate axes*

# Geometry of Sparse Signal Sets

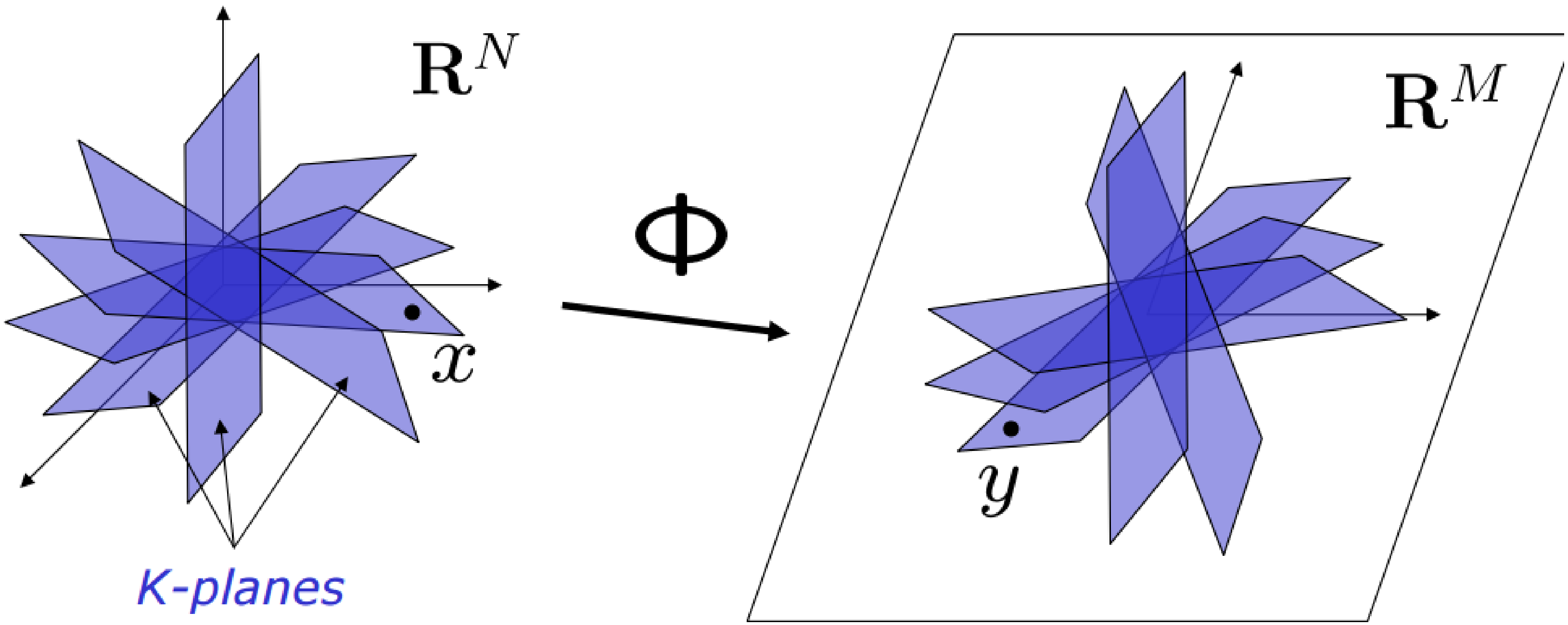



***Linear*** — ***Sparse, Nonlinear***

***K-plane*** — ***Union of K-planes***

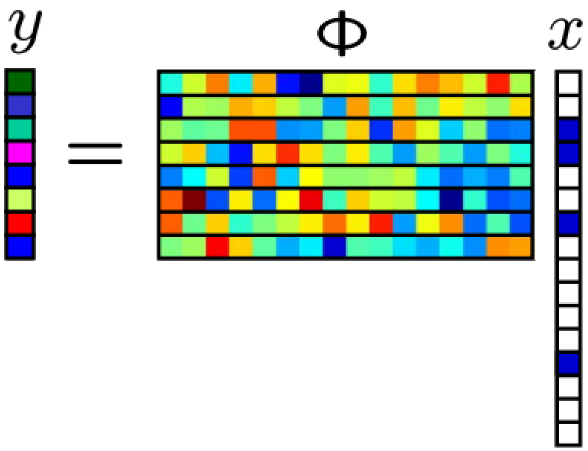
# Geometry: Embedding in $R^M$




- $\Phi$(K-plane) = K-plane in general
- M ≥ 2K measurements
  - necessary for injectivity
  - sufficient for injectivity when $\Phi$ Gaussian
  - but not enough for efficient, robust recovery
- (PS - can distinguish *most* K-sparse x with as few as M=K+1)

# Restricted Isometry Property

- Measurement matrix $\Phi$ has ***RIP of order $K$*** if




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

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

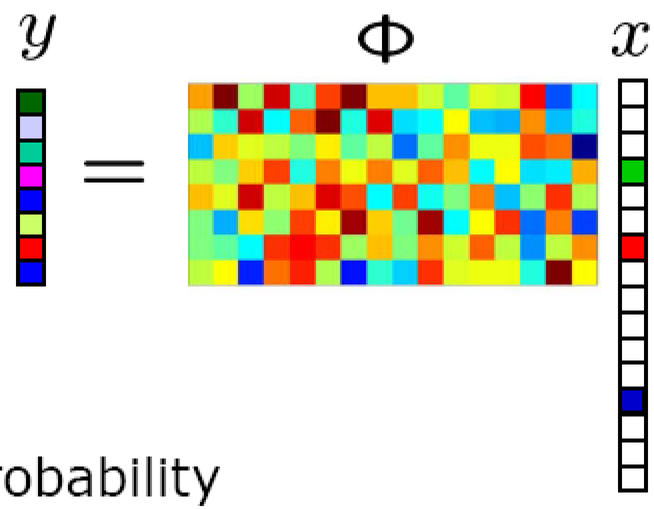
- Does *not* hold for $K > M$; may hold for smaller $K$.

- Implications: tractable, stable, robust recovery

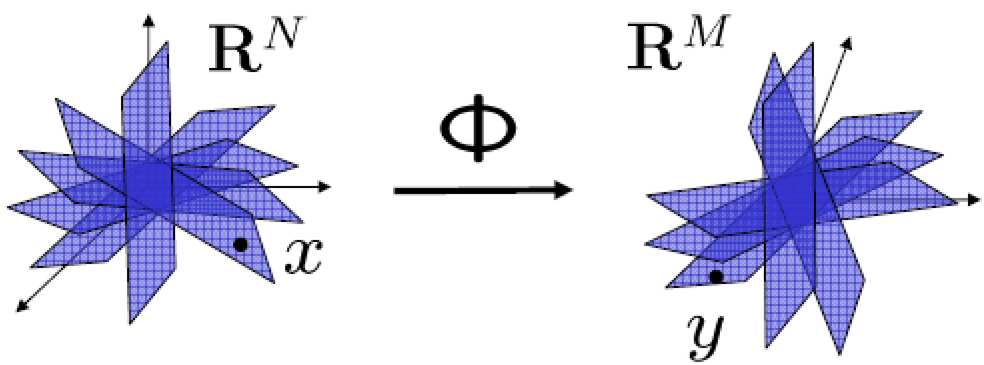
# Why does it work?

- Random projection $\Phi$
  **not full rank**...

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



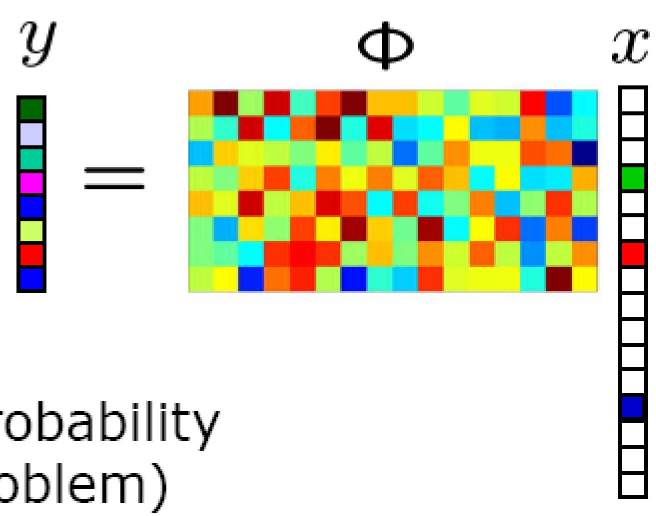

*K*-sparse model



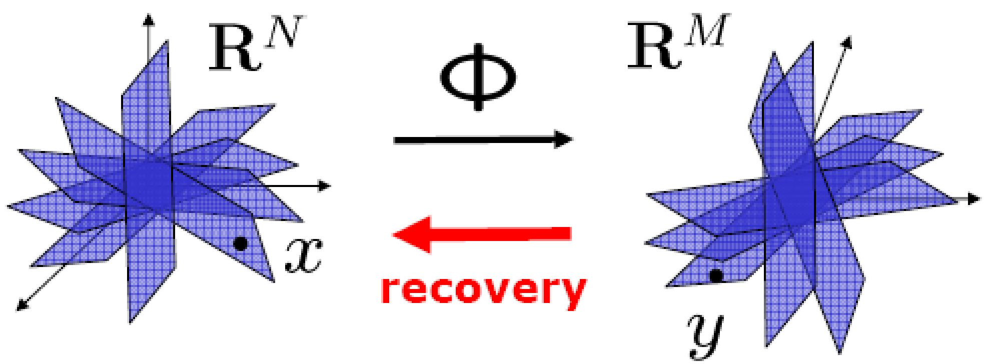
# CS Signal Recovery

- Random projection $\Phi$
  **not full rank**...

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




$K$-sparse
model



# CS Signal Recovery

What is Recovery? Essentially, it's an optimization problem

$$\vec{y} = \phi \vec{x}$$

$$\longrightarrow \boxed{argmin_x |\vec{y} - \phi \vec{x}|_p}$$

where **p** indicates the norm you want to use

**p** = 0 → $L_0$ norm

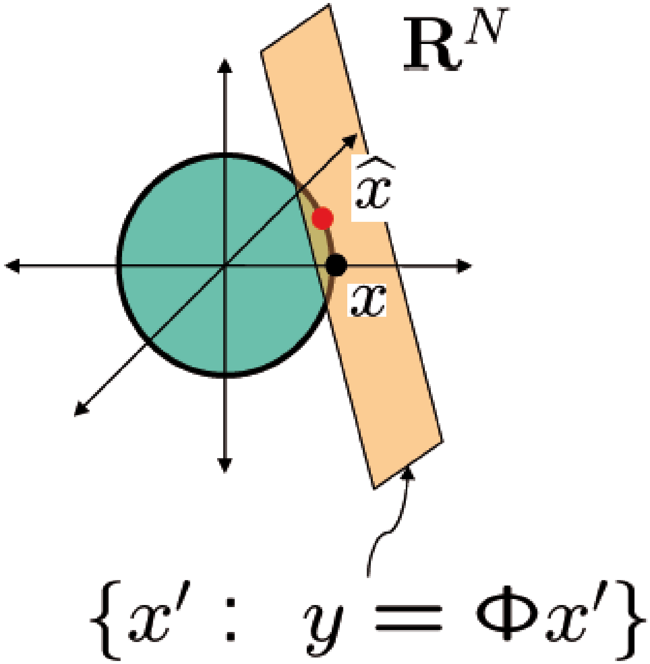
**p** = 1 → $L_1$ norm

**p** = 2 → $L_2$ norm

etc...

# Well, $l_2$ Doesn't Work

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



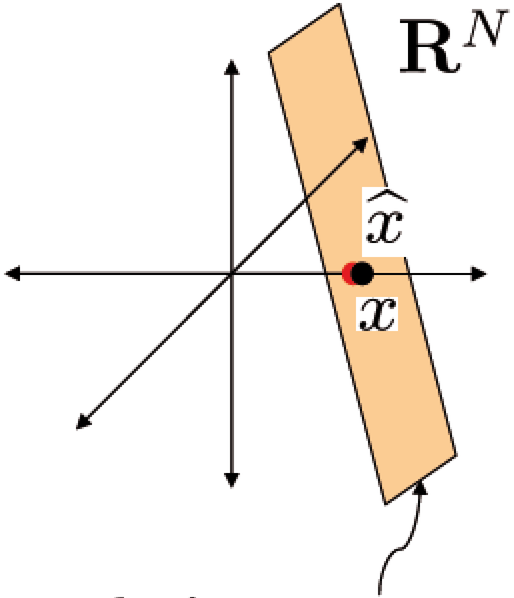

least squares,
minimum $L_2$ solution
is almost **never sparse**

# $l_0$ Recovery works but...

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

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

(w.p. 1 for Gaussian $\Phi$)




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

*null space of* $\Phi$
*translated to* $x$

**But it is extremely expensive! → NP complete**

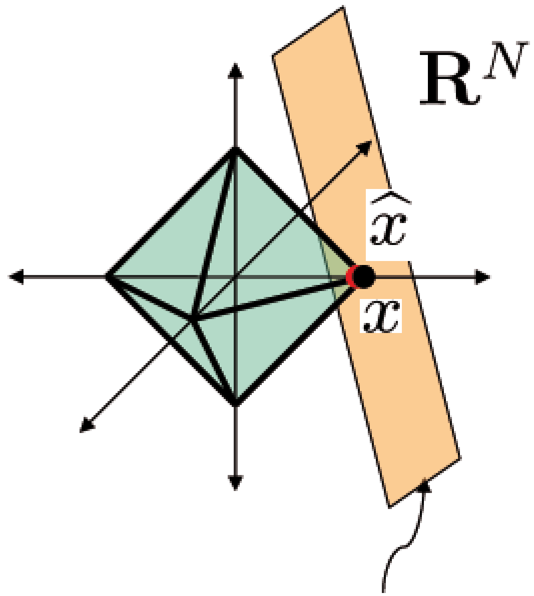
# $l_1$ is the best choice

Convex relaxation of $l_0$ problem:

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

minimum $\boldsymbol{L_1}$ solution
= $\boldsymbol{L_0}$ sparsest solution if

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




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

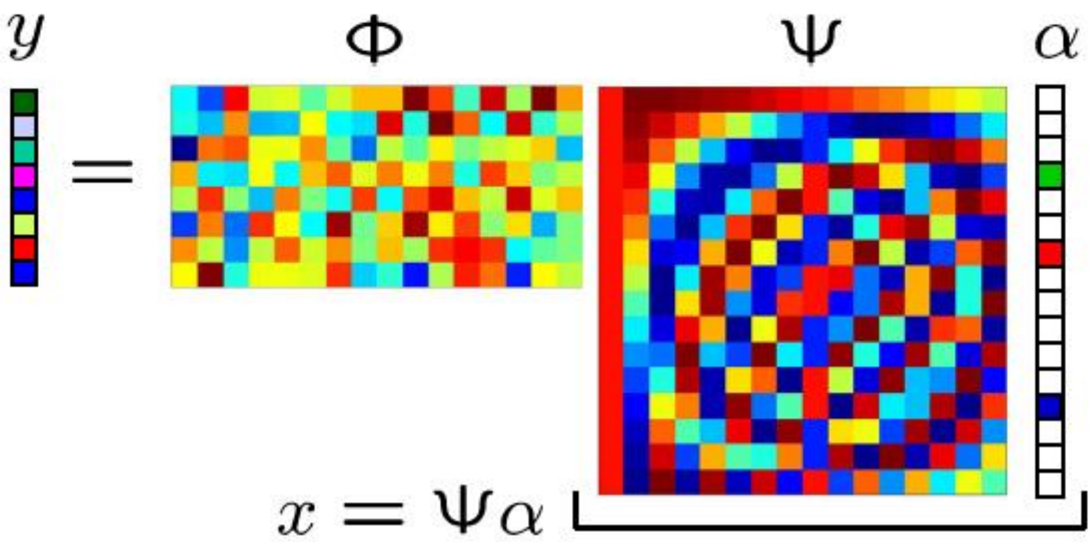
*random orientation*
*dimension N-M*

Correct solution!
Polynomial-time algorithm (linear programming)

# Universality

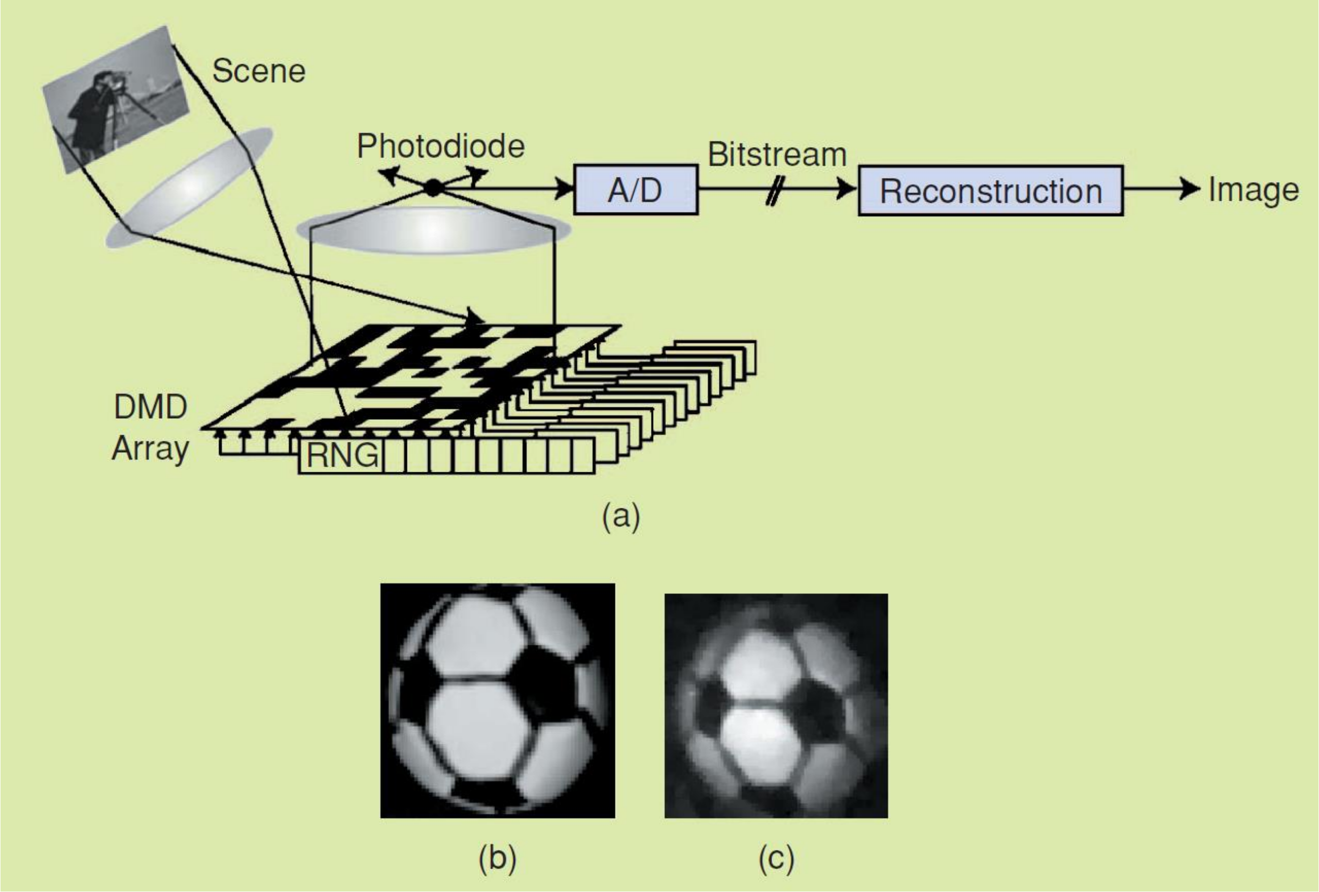
- Random measurements can be used for signals sparse in *any* basis

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



# Application

- Example: One Pixel Camera




universitätbonn

# Summary

1. Many signals in nature of engineering are sparse and hence can be acquired directly in basis (sampling matrix) where they happen to be sparse.

2. A K-sparse signal *lives* is the union of K-dimensional hyperplanes.

3. The sampling matrix could be a random matrix, since these allow to acquire/recover sparse signals.

4. The sampling matrix must obey the restrictive isometry property.

5. The recovery can be done efficiently through $L_1$ optimization, e.g, single pixel camera.

6. Compressed acquisition reduces storage and overload issues in modern technology since the sampling process occurs at rates below the Nyquist theorem

# References


- Baraniuk 2007, “Compressive Sensing”. Lecture Notes. University Of Columbia.

- Baraniuk‘s lecture at University Of Delaware:
  https://www.youtube.com/watch?v=RvMgVv-xZhQ

- Justin Romberg 2003. “Sparse Representations”. Lectures at Tsinghua:
  http://jrom.ece.gatech.edu/tsinghua-oct13/

- Davenport et. Al. 2011. “Introduction to Compressed Sensing”. Stanford University.