Compressive Sensing for High-Dimensional Data

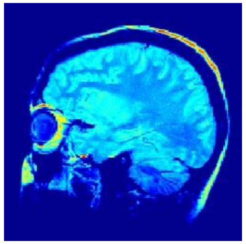
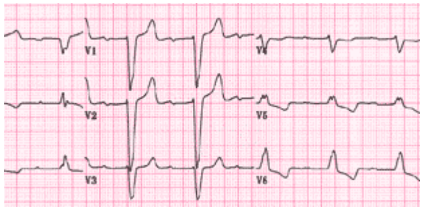
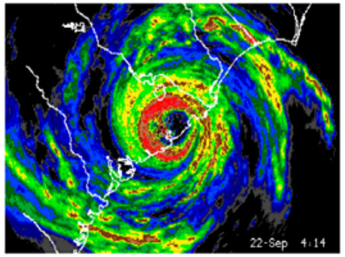
Richard Baraniuk

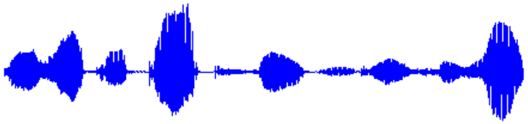
Rice University
dsp.rice.edu/cs

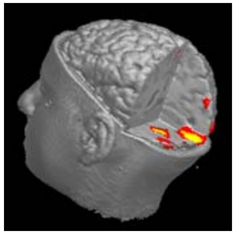
DIMACS Workshop on Recent Advances in Mathematics and Information Sciences for Analysis and Understanding of Massive and Diverse Sources of Data

RICE UNIVERSITY DSP





Pressure is on DSP

- Increasing pressure on signal/image processing hardware and algorithms to support

higher resolution / denser sampling

» ADCs, cameras, imaging systems, ...

X

large numbers of sensors

» multi-view target data bases, camera arrays and networks, pattern recognition systems,

X

increasing numbers of modalities

» acoustic, seismic, RF, visual, IR, SAR, ...

=

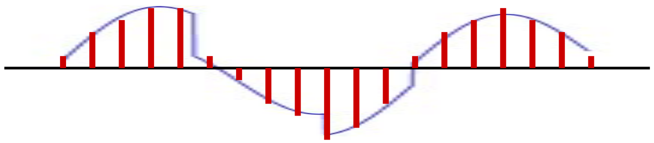
deluge of data

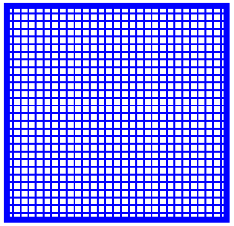
» how to acquire, store, fuse, process efficiently?

Data Acquisition

- Time: A/D converters, receivers, ...
- Space: cameras, imaging systems, ...

- Foundation: ***Shannon sampling theorem***
  - *Nyquist rate*: must sample at 2x highest frequency in signal

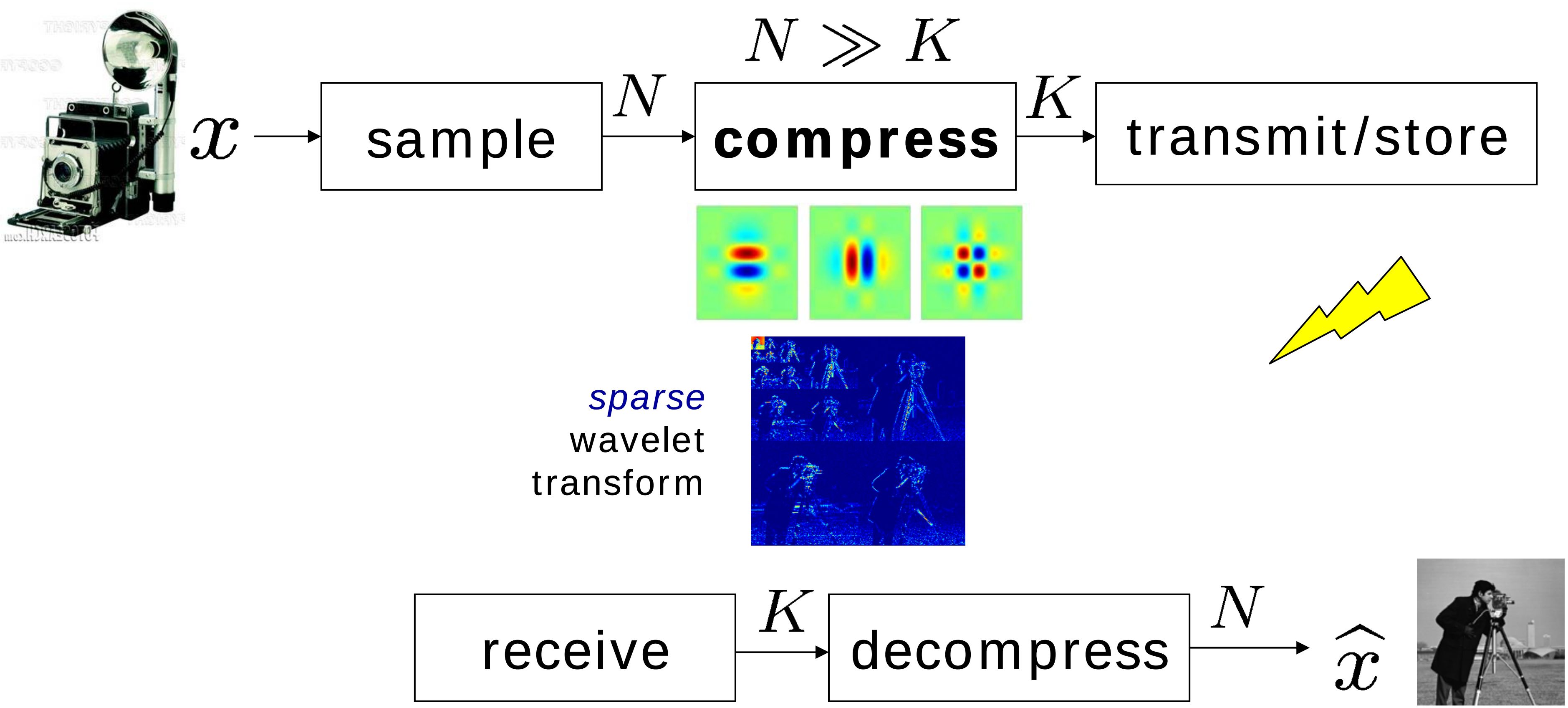




N periodic samples

Sensing by *Sampling*

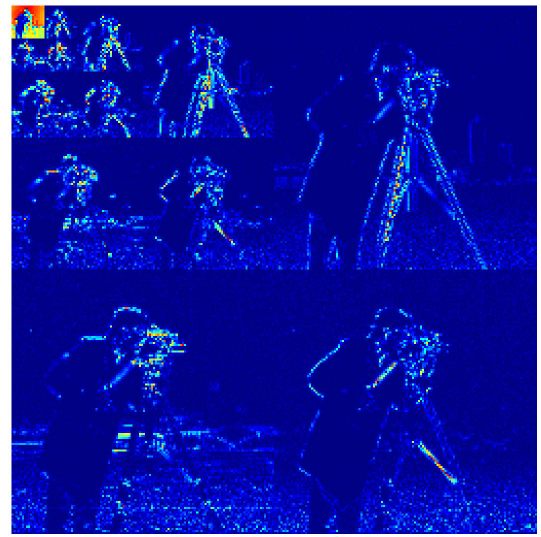
- Long-established paradigm for digital data acquisition
  - ***sample*** data (A-to-D converter, digital camera, ...)
  - ***compress*** data (signal-dependent, nonlinear)



Sparsity / Compressibility

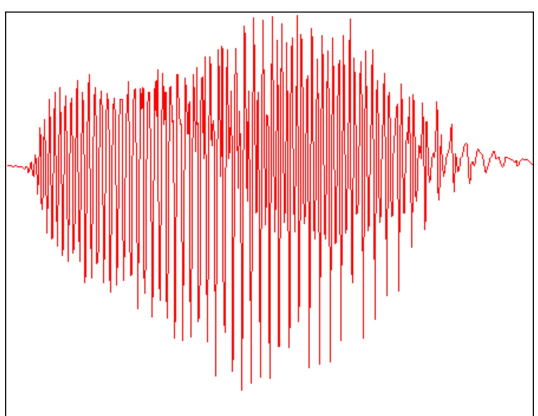
- Number of samples N often too large, so ***compress***
  - transform coding: exploit data *sparsity/compressibility* in some representation (ex: orthonormal basis)

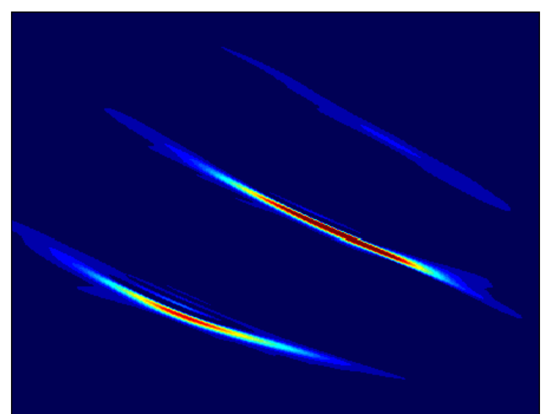
N pixels



$K \ll N$ large wavelet coefficients

N wideband signal samples



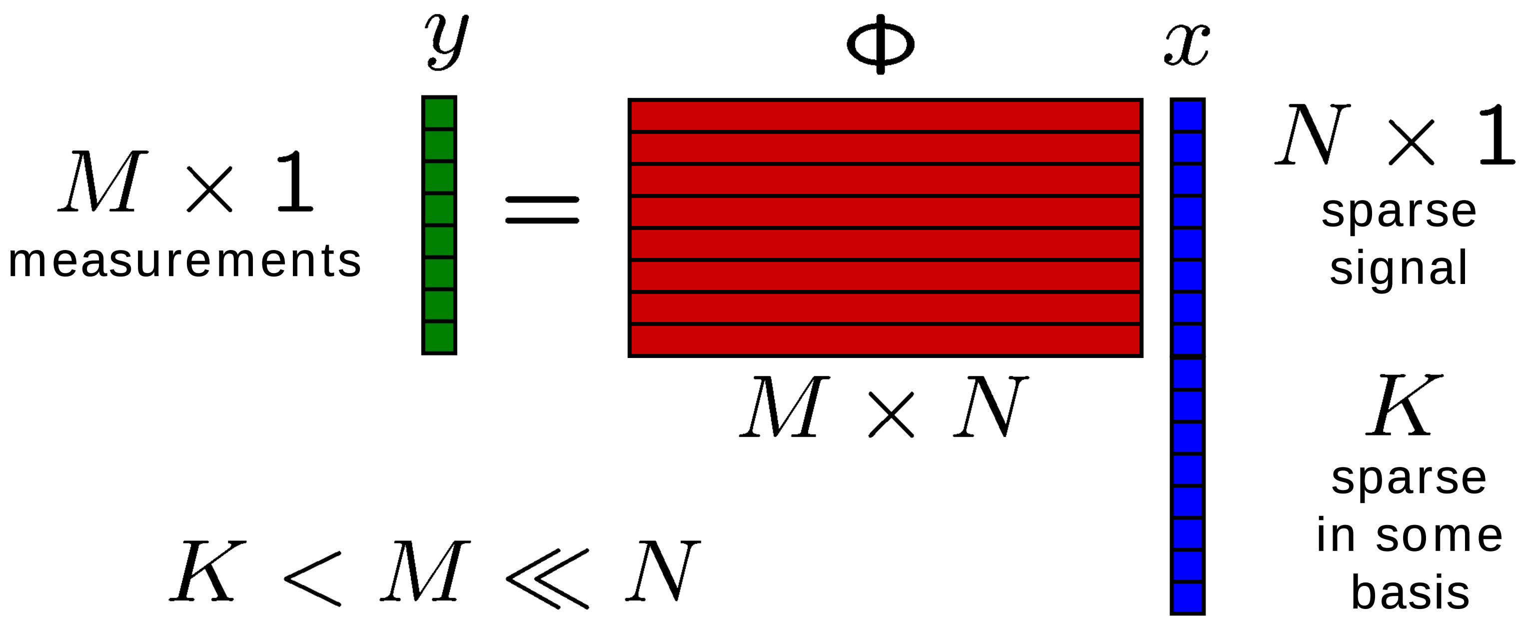


$K \ll N$ large Gabor coefficients

Compressive Data Acquisition

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

$$y = \Phi x$$

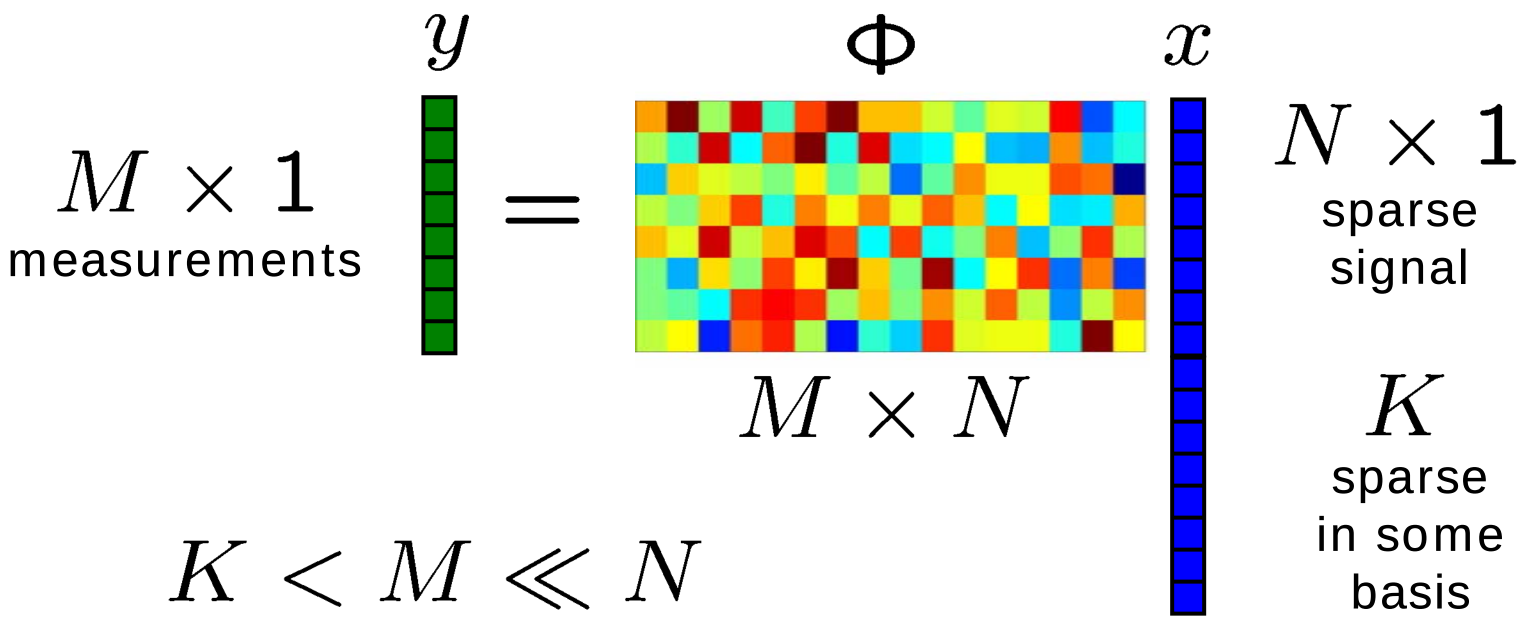


Compressive Data Acquisition

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

$$y = \Phi x$$

- ***Random projection*** will work



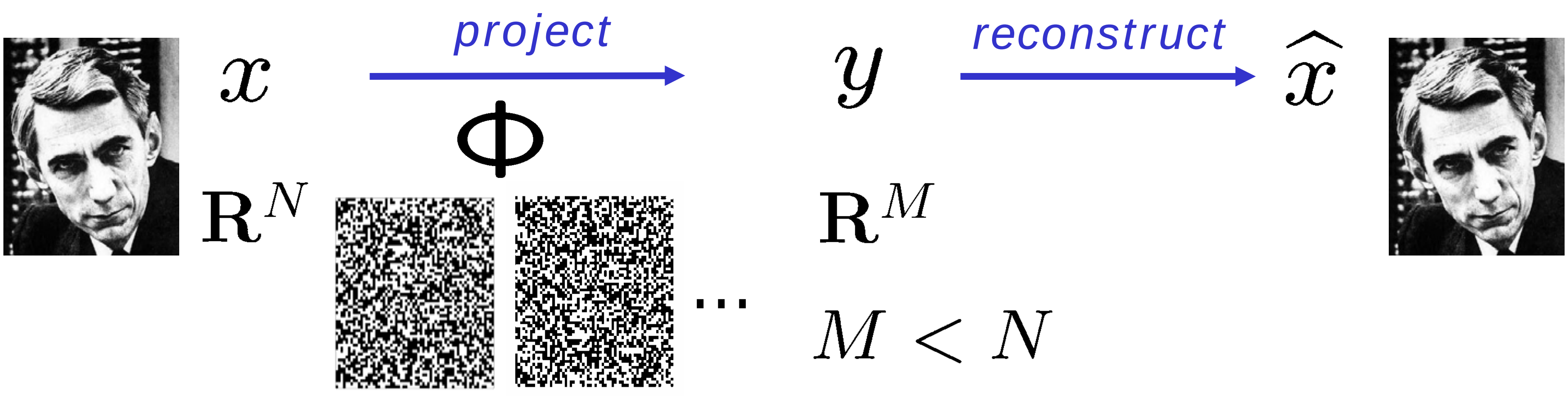
$$K < M \ll N$$

Compressive Data Acquisition

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

$$y = \Phi x$$

- ***Random projection*** *preserves information*
  - Johnson-Lindenstrauss Lemma (point clouds, 1984)
  - Compressive Sensing (CS) (sparse and compressible signals, Candes-Romberg-Tao, Donoho, 2004)



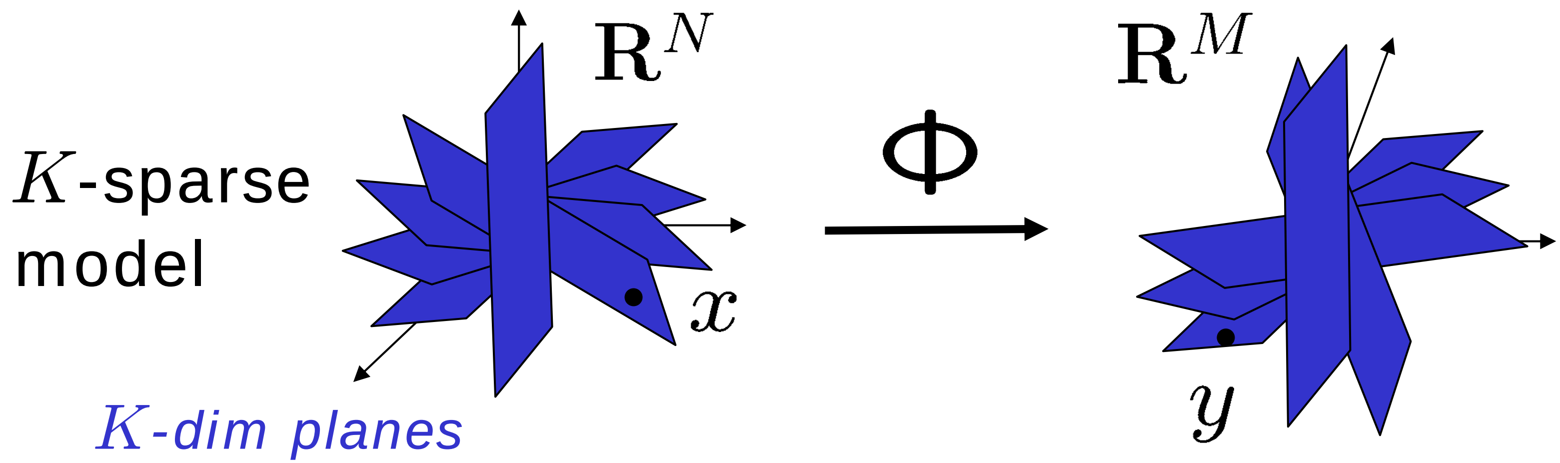
Why Does It Work (1)?

- Random projection not full rank, but ***stably embeds***
  - sparse/compressible signal models (CS)
  - point clouds (JL)

  into lower dimensional space with high probability
- Stable embedding: ***preserves structure***
  - distances between points, angles between vectors, ...

  provided M is large enough: **Compressive Sensing**

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



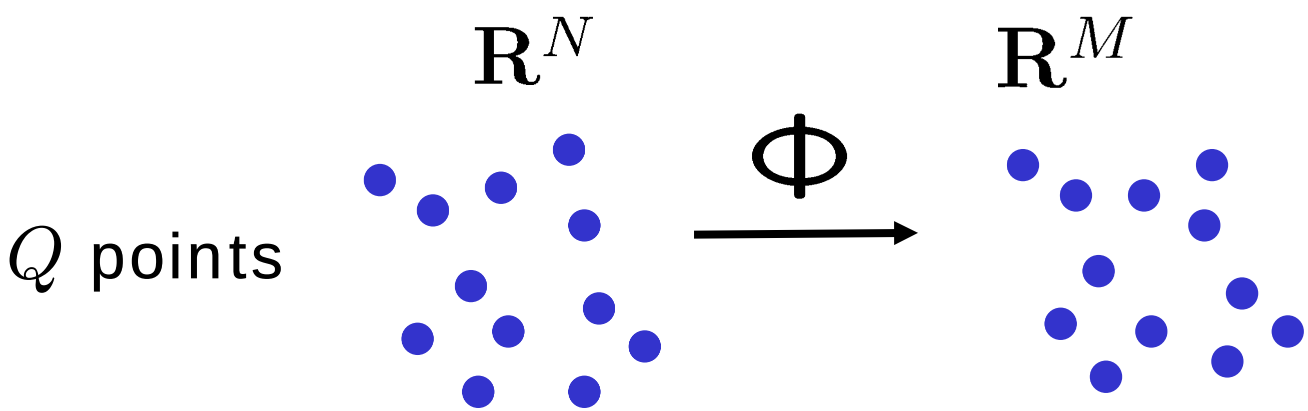
Why Does It Work (2)?

- Random projection not full rank, but ***stably embeds***
  - sparse/compressible signal models (CS)
  - point clouds (JL)

  into lower dimensional space with high probability
- Stable embedding: ***preserves structure***
  - distances between points, angles between vectors, ...

  provided M is large enough: **Johnson-Lindenstrauss**

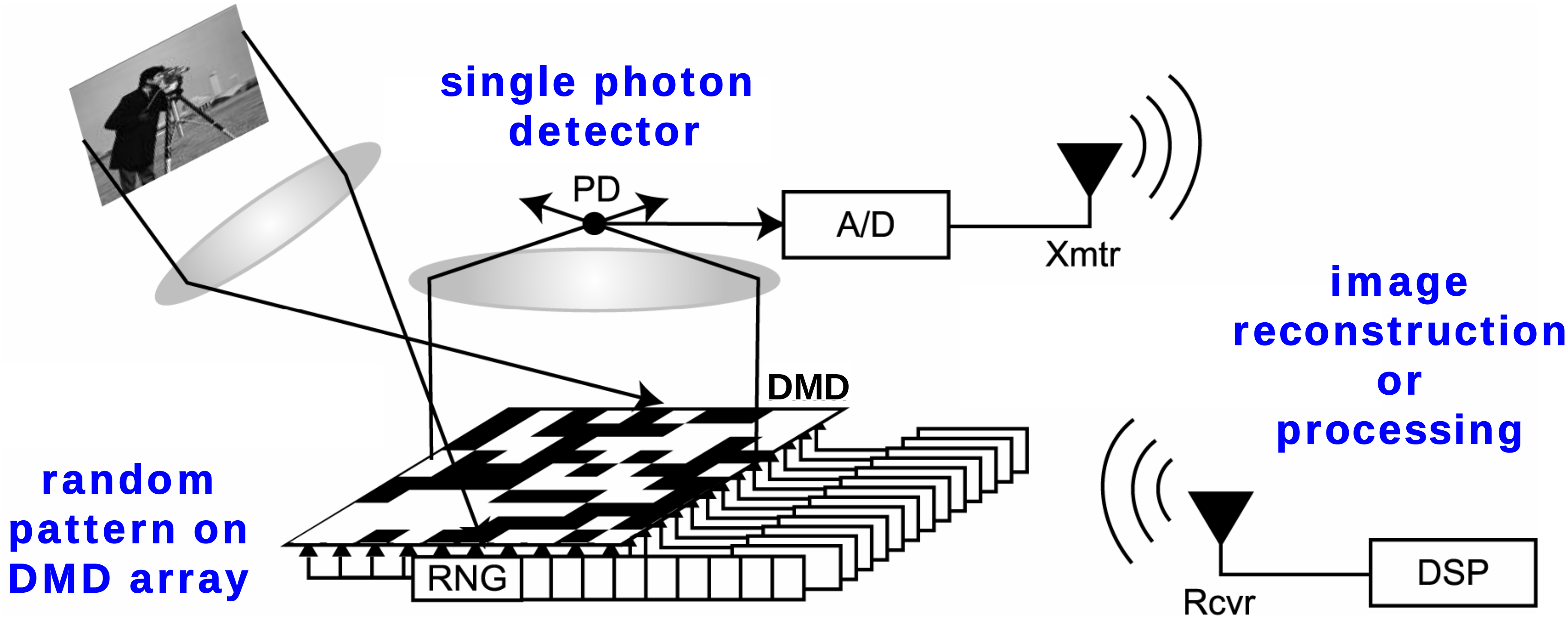
$$M = O(\log Q)$$

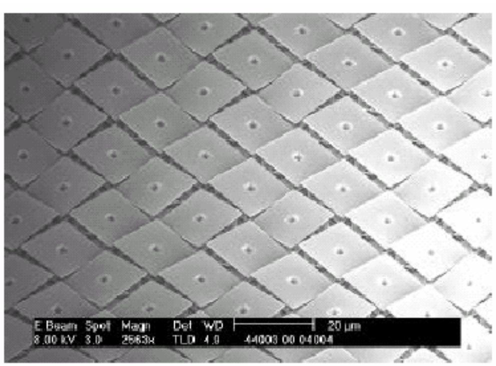


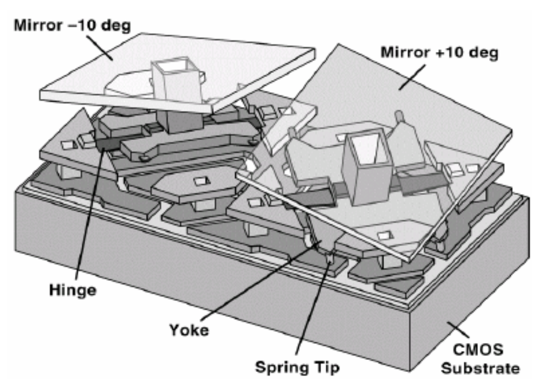
CS Hallmarks

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

Example: “Single-Pixel” CS Camera

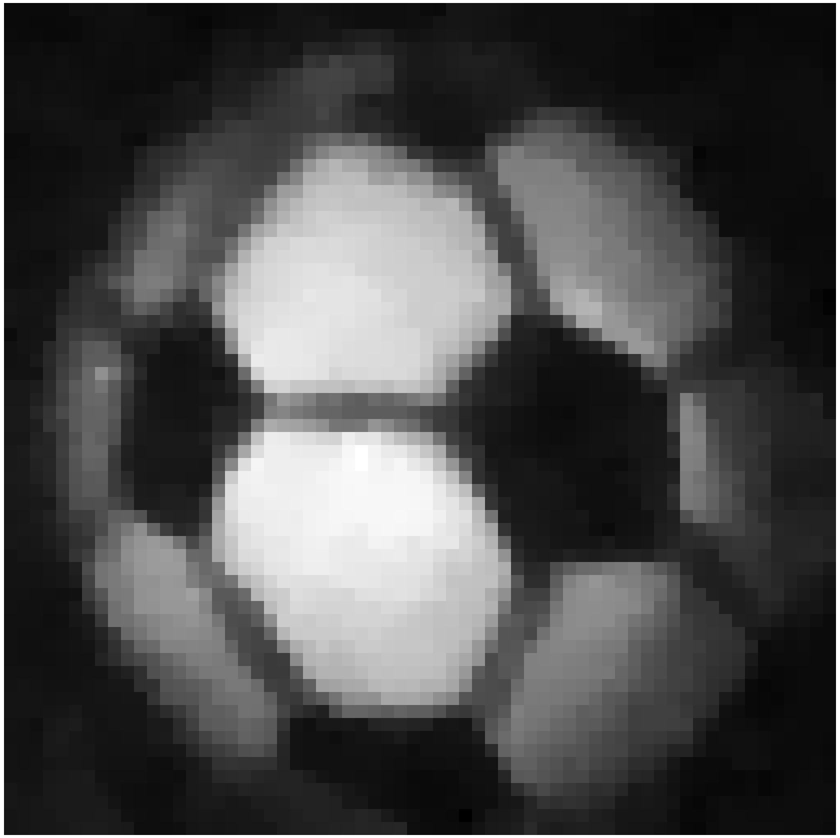






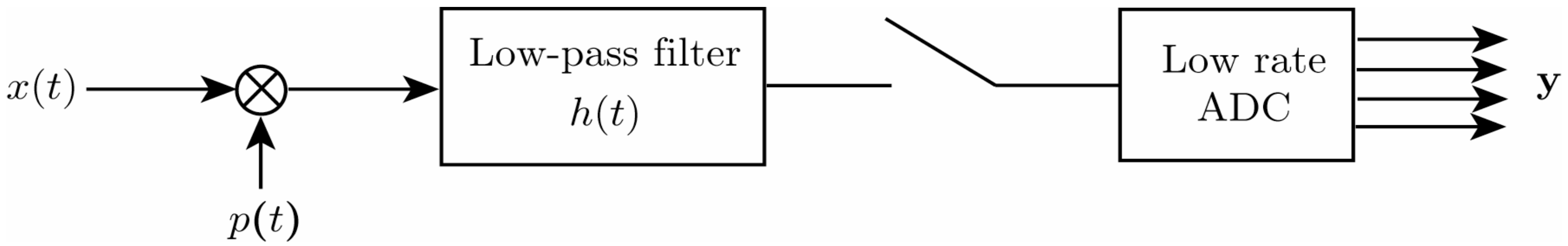
Example Image Acquisition

4096
pixels



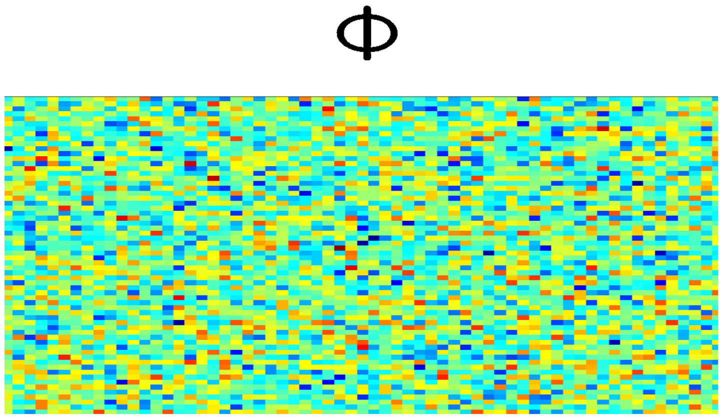
500
random measurements

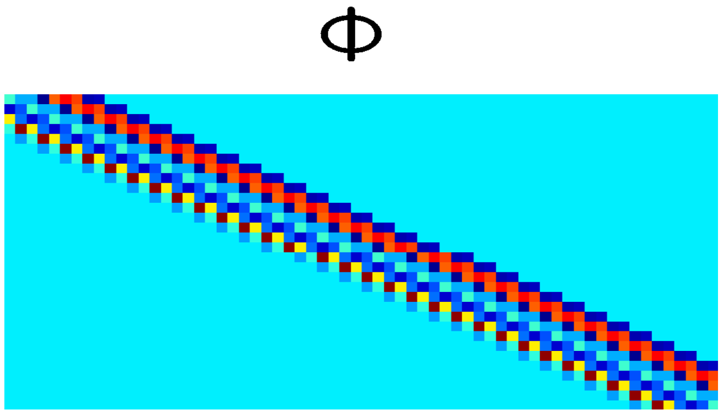
Analog-to-*Information* Conversion



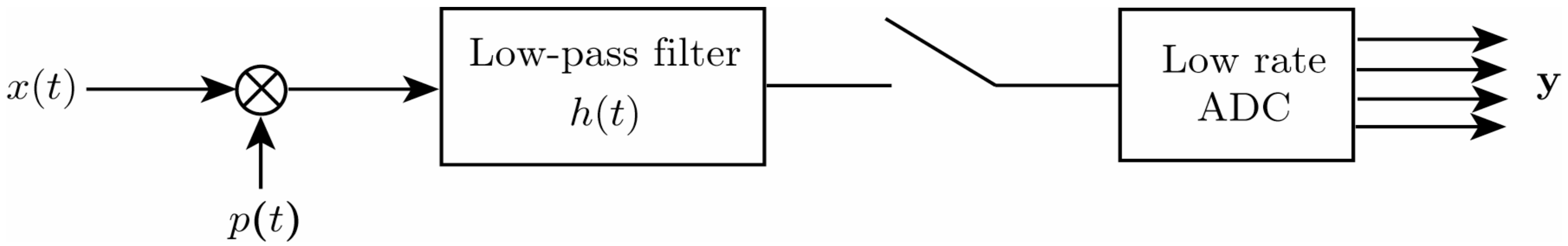
pseudo-random code

- For **real-time, streaming use**, Φ can have banded structure
- Can implement in analog hardware

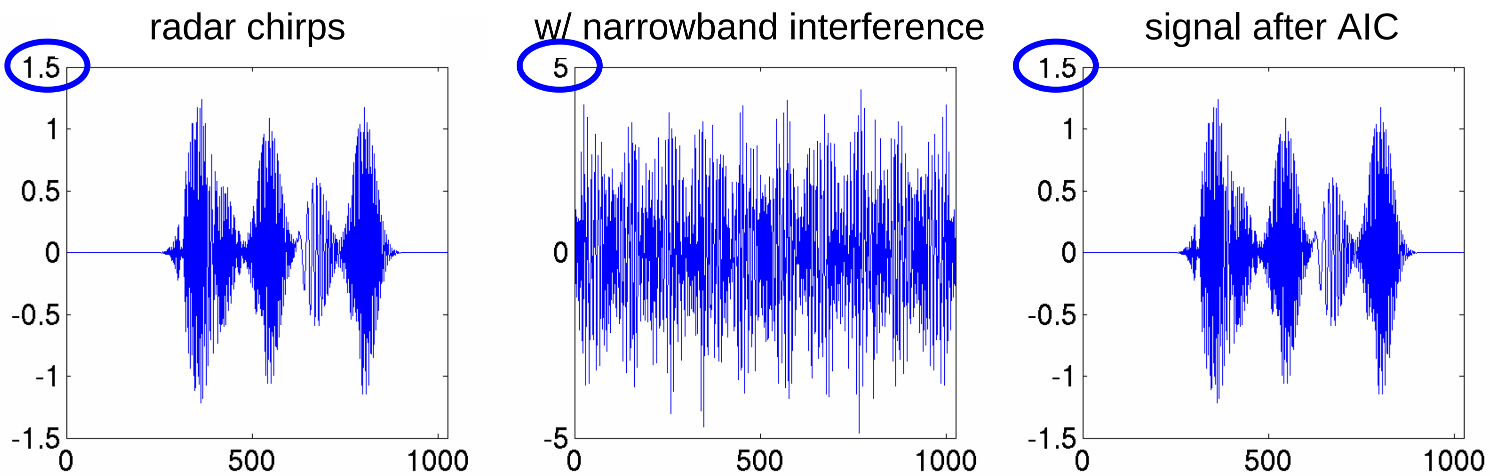




Analog-to-*Information* Conversion



- For **real-time, streaming use**, Φ can have banded structure
- Can implement in analog hardware



Information Scalability

- If we can ***reconstruct*** a signal from compressive measurements, then we should be able to perform other kinds of statistical signal processing:
  - ***detection***
  - ***classification***
  - ***estimation …***

Multiclass Likelihood Ratio Test

- Observe one of P known signals in noise

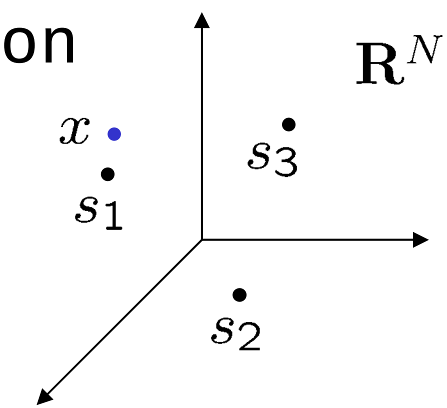
$$\begin{aligned} H_1 &: \; x = s_1 + n \\ H_2 &: \; x = s_2 + n \\ &\vdots \\ H_P &: \; x = s_P + n \end{aligned}$$

- Classify according to:

$$\arg\max_{j=1,\ldots,P} \; p(x|H_j)$$

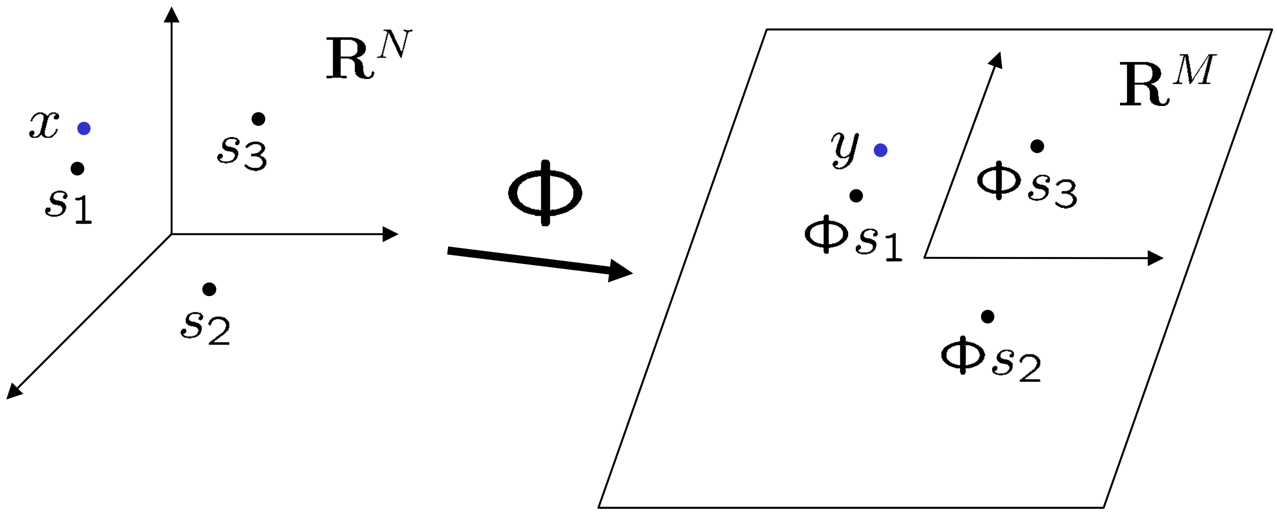
- AWGN: ***nearest-neighbor*** classification

$$\arg\min_{j=1,\ldots,P} \; \|x - s_j\|_2$$



Compressive LRT

- Compressive observations: $H_j \;:\; y = \Phi(s_j + n)$



$$t_1 = \|y - \Phi s_1\|_2$$
$$t_2 = \|y - \Phi s_2\|_2$$
$$t_3 = \|y - \Phi s_3\|_2$$

by the JL Lemma these distances are preserved (*)

[Waagen et al 05; RGB, Davenport et al 06; Haupt et al 06]

Matched Filter

- In many applications, signals are ***transformed*** with an unknown parameter; ex: translation

$$H_j \; : \; x = s_j(t - \theta_j) + n$$

- Elegant solution: ***matched filter***

  Compute

$$\langle x, s_j(t - \theta_j) \rangle \text{ for all } \theta_j$$

$$\Updownarrow$$

$$x * s_j(-t)$$

Challenge: Extend compressive LRT to accommodate ***parameterized signal transformations***

Generalized Likelihood Ratio Test

- Matched filter is a special case of the GLRT

$$\arg\max_{j=1,\ldots,P} \; p(x|\widehat{\theta}_j, H_j)$$

$$\widehat{\theta}_j = \arg\max_{\theta \in \Theta_j} \; p(x|\theta, H_j)$$

- GLRT approach extends to any case where each class can be ***parameterized*** with K parameters

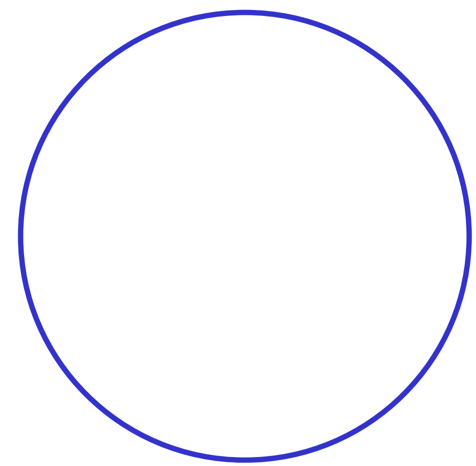
- If mapping from parameters to signal is well-behaved, then each class forms a ***manifold*** in $\mathbf{R}^N$

What is a Manifold?

“Manifolds are a bit like pornography: hard to define, but you know one when you see one.”

– S. Weinberger [Lee]

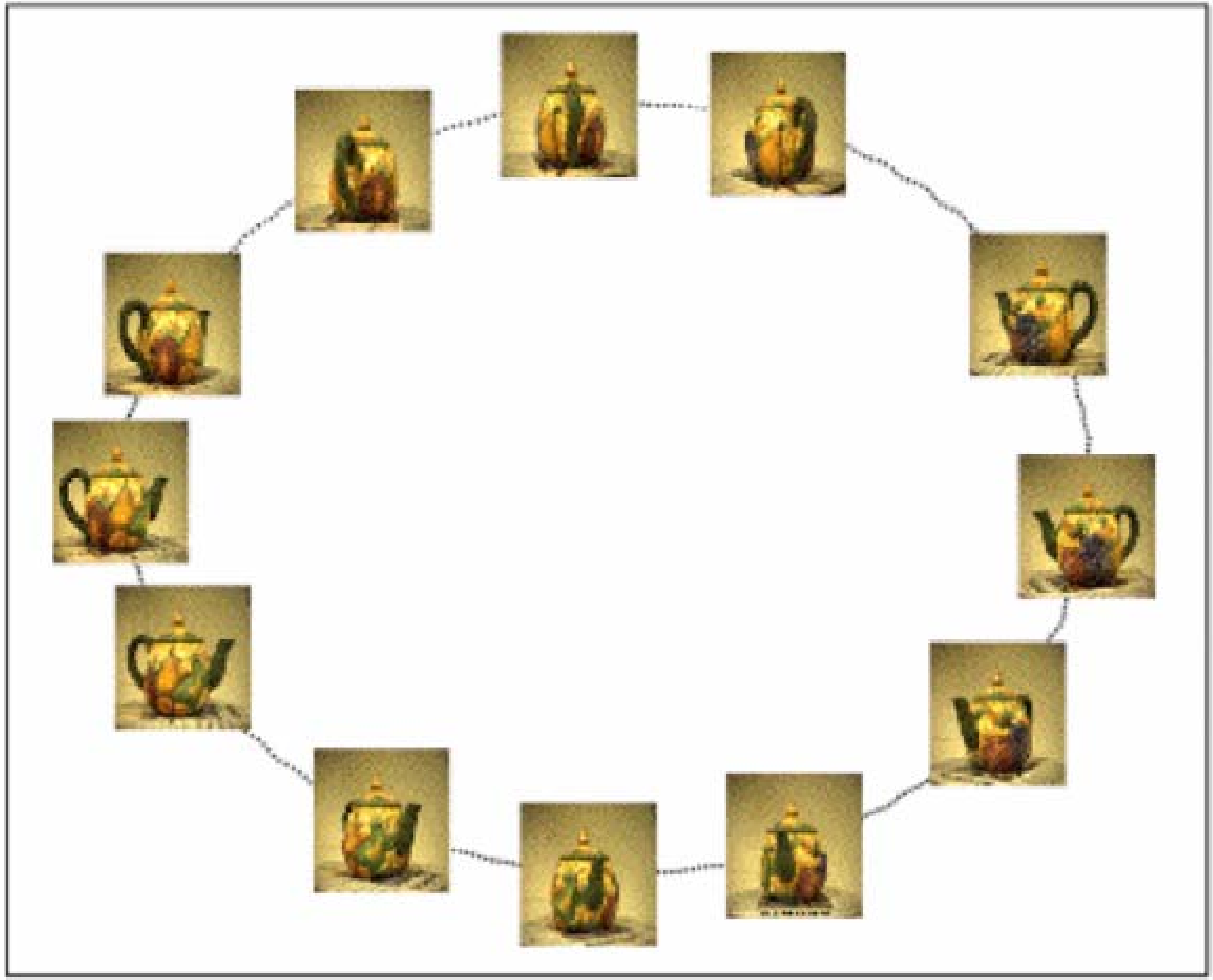
- Locally Euclidean topological space



- Roughly speaking:
  - a collection of mappings of open sets of $\boldsymbol{R}^{\boldsymbol{K}}$ glued together (“coordinate charts”)
  - can be an abstract space, not a subset of Euclidean space
    - e.g., $\boldsymbol{SO3}$, Grassmannian

- *Typically for signal processing*:
  - nonlinear $\boldsymbol{K}$-dimensional “surface” in signal space $\boldsymbol{R}^{\boldsymbol{N}}$

Object Rotation Manifold



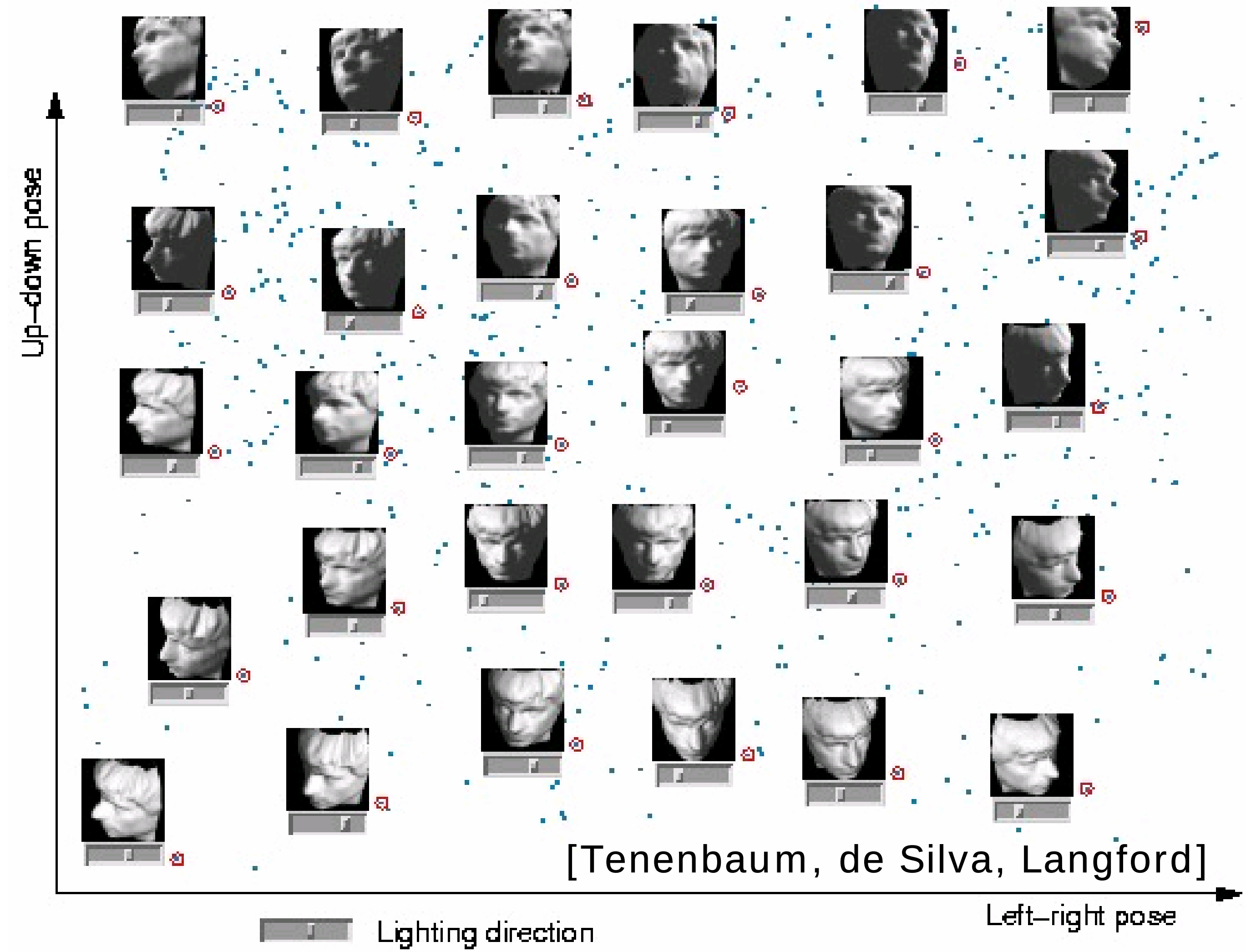
$\ominus$

$K = 1$

Up/Down Left/Right Manifold

$\ominus$

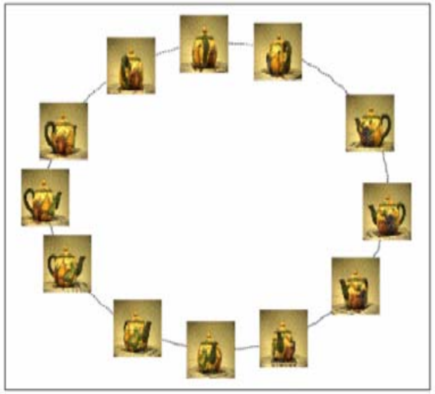
$K = 2$



[Tenenbaum, de Silva, Langford]

Manifold Classification

- Now suppose data is drawn from one of P possible manifolds:



$$H_j \;:\; x = m_j + n, \quad m_j \in M_j$$

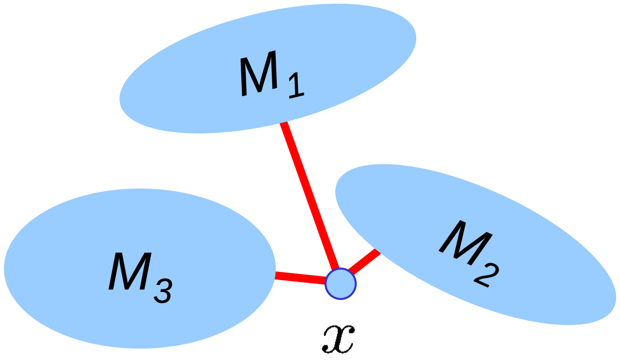
$$m_j = f_j(\theta_j)$$

- AWGN: ***nearest manifold classification***

$$\arg\max_{j=1,\ldots,P} \; p(x|\widehat{\theta}_j, H_j)$$

$$= \arg\min_{j=1,\ldots,P} \; \|x - f_j(\widehat{\theta}_j)\|_2$$

$$\widehat{\theta}_j = \arg\min_{\theta \in \Theta_j} \; \|x - f_j(\theta_j)\|_2$$

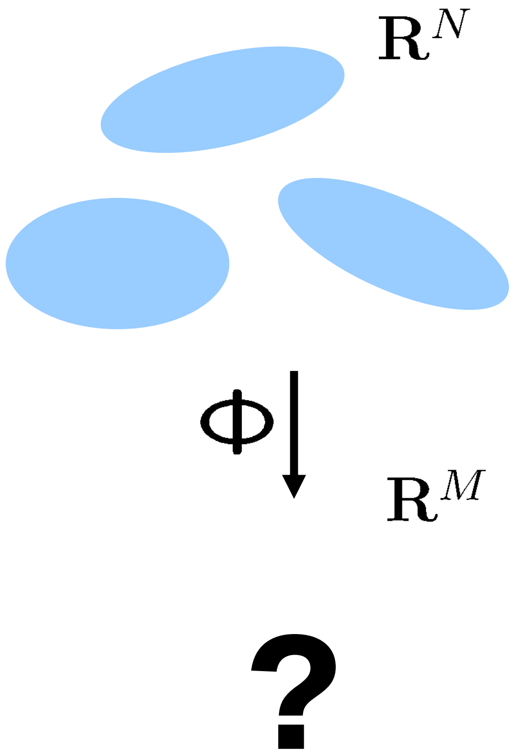


Compressive Manifold Classification

- Compressive observations:

$$H_j \;:\; x = \Phi(m_j + n)$$

$$m_j \in M_j$$

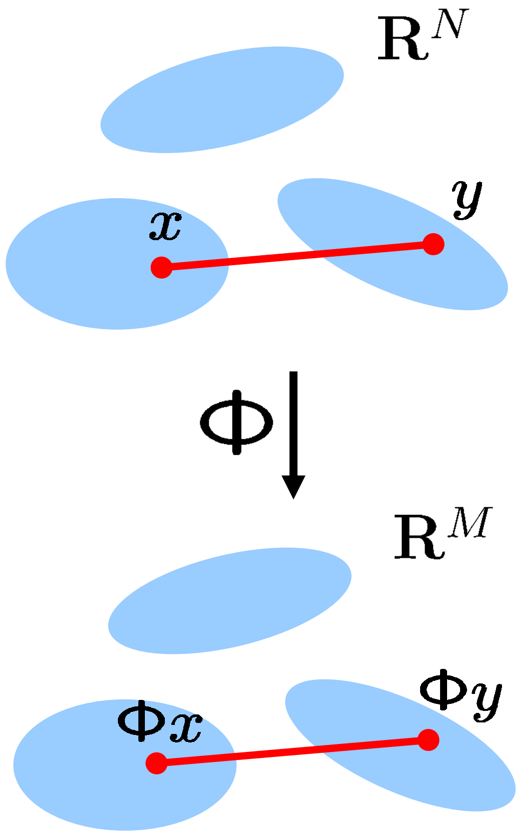


Compressive Manifold Classification

- Compressive observations:

$$H_j \;:\; x = \Phi(m_j + n)$$

$$m_j \in M_j$$



- **Good news**: structure of smooth manifolds is preserved by random projection provided

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

  – distances, geodesic distance, angles, ...

[RGB and Wakin, 06]

Stable Manifold Embedding

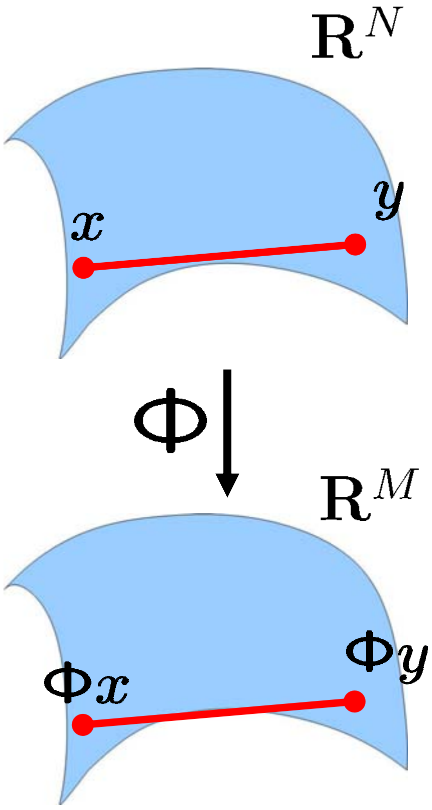
Theorem:

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

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

Let Φ be a random MxN orthoprojector with

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



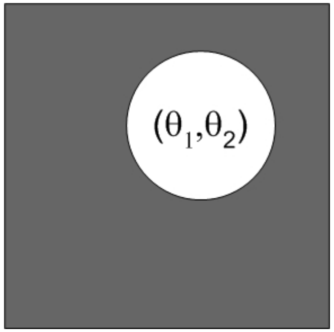
Then with probability at least $1-\rho$, the following statement holds:

For every pair $x,y \in F$,

$$(1-\epsilon)\,\|x-y\|_2 \leq \|\Phi x - \Phi y\|_2 \leq (1+\epsilon)\,\|x-y\|_2.$$

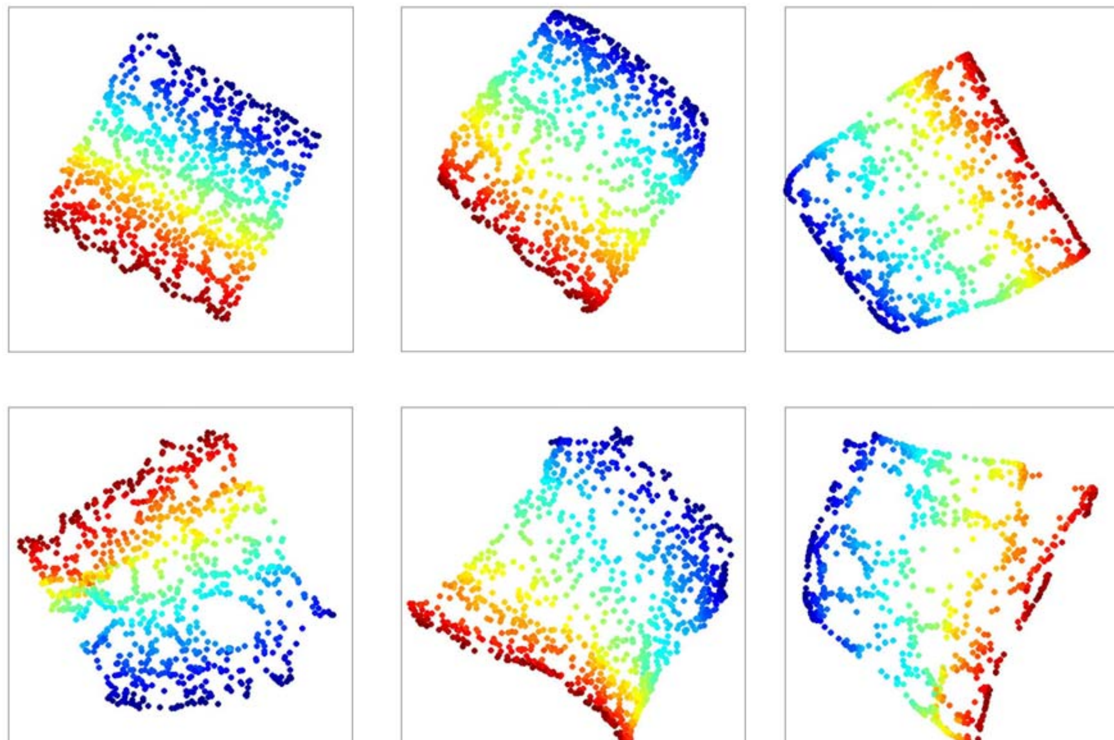
[Wakin et al 06]

Manifold Learning from Compressive Measurements



ISOMAP | **HLLE** | **Laplacian Eigenmaps**

R^{4096}



R^M

M=15 | M=20 | M=15

The *Smashed Filter*

- ***Compressive manifold classification*** with GLRT
  - nearest-manifold classifier based on manifolds

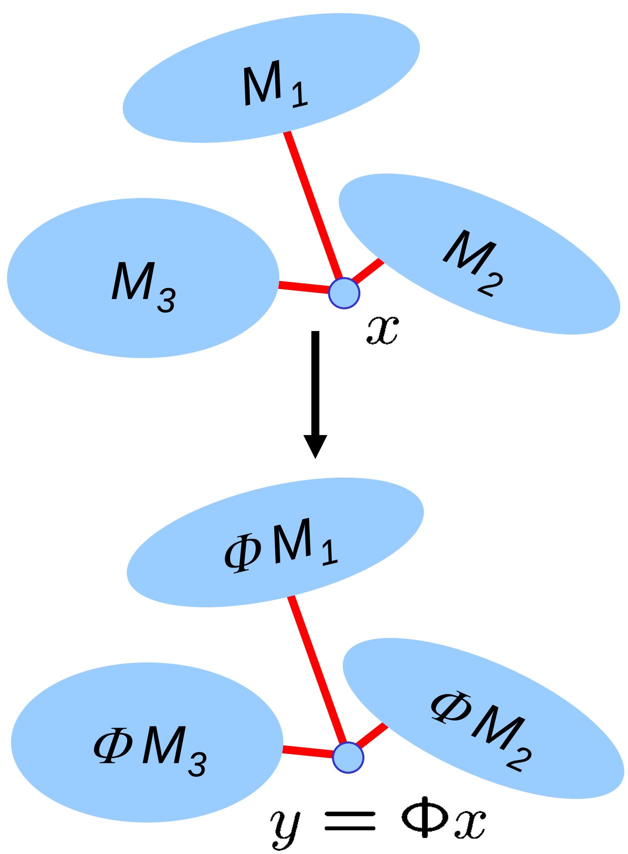
$$\Phi M_j = \{\Phi f_j(\theta_j) \ : \ \theta_j \in \Theta_j\}$$

$$H_j \ : \ y = \Phi(m_j + n), \quad m_j \in M_j$$

$$m_j = f_j(\theta_j)$$

$$\underset{j=1,\ldots,P}{\arg\min} \ \|y - \Phi f_j(\widehat{\theta}_j)\|_2$$

$$\widehat{\theta}_j = \underset{\theta\in\Theta_j}{\arg\min} \ \|y - \Phi f_j(\theta_j)\|_2$$

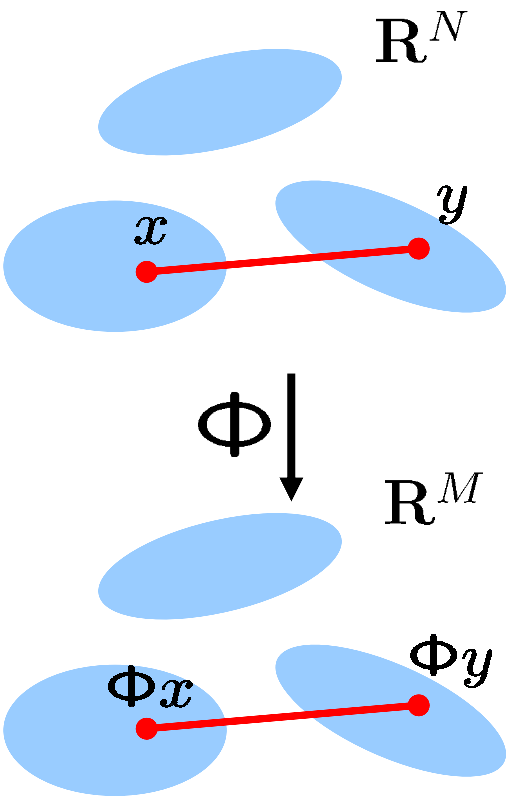


Multiple Manifold Embedding

Corollary:

Let $M_1, \ldots, M_P \subset R^N$ be compact K-dimensional manifolds with

- condition number $1/\tau$ (curvature, self-avoiding)
- volume V
- min dist(M_j,M_k) > τ (can be relaxed)



Let Φ be a random MxN orthoprojector with

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

Then with probability at least $1-\rho$, the following statement holds:

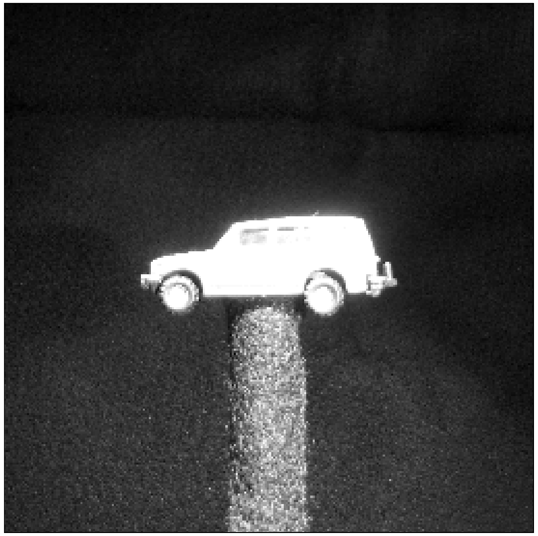
For every pair $x, y \in \bigcup M_j$,

$$(1-\epsilon)\,\|x-y\|_2 \le \|\Phi x - \Phi y\|_2 \le (1+\epsilon)\,\|x-y\|_2.$$

Smashed Filter - Experiments

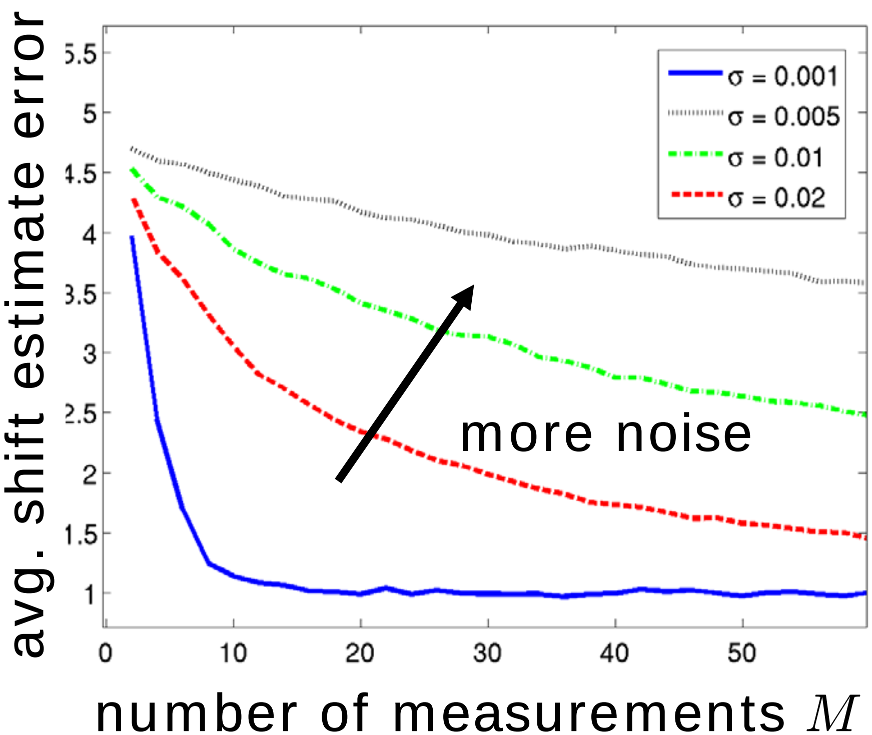
- 3 image classes: tank, school bus, SUV
- N = 64K pixels
- Imaged using single-pixel CS camera with
  - unknown shift
  - unknown rotation

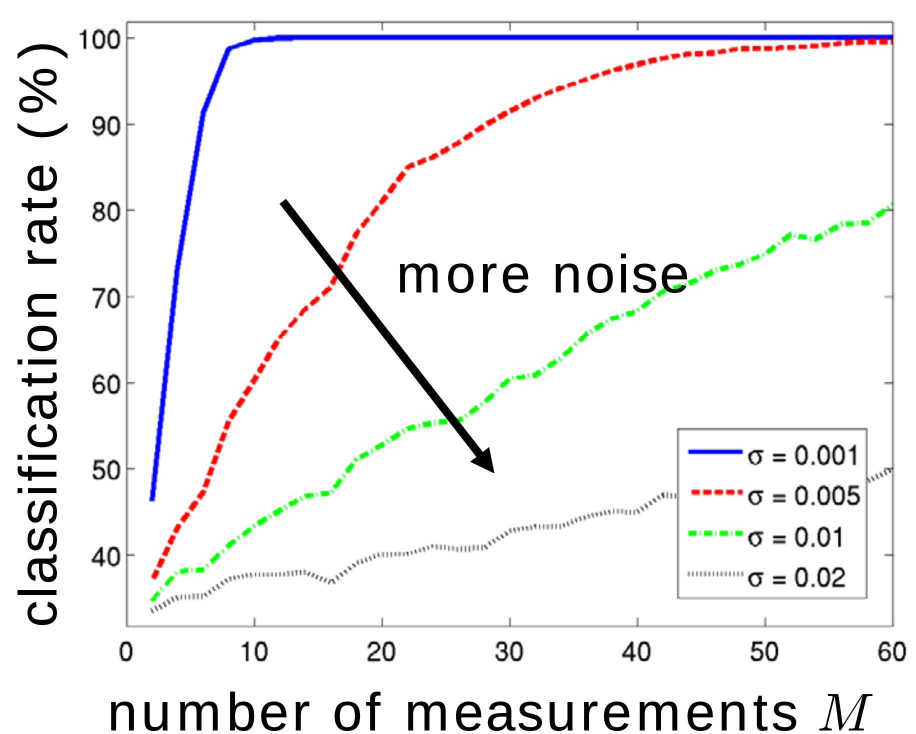




Smashed Filter – Unknown Position

- Image shifted at random (K=2 manifold)
- Noise added to measurements
  - identify most likely position for each image class
  - identify most likely class using nearest-neighbor test



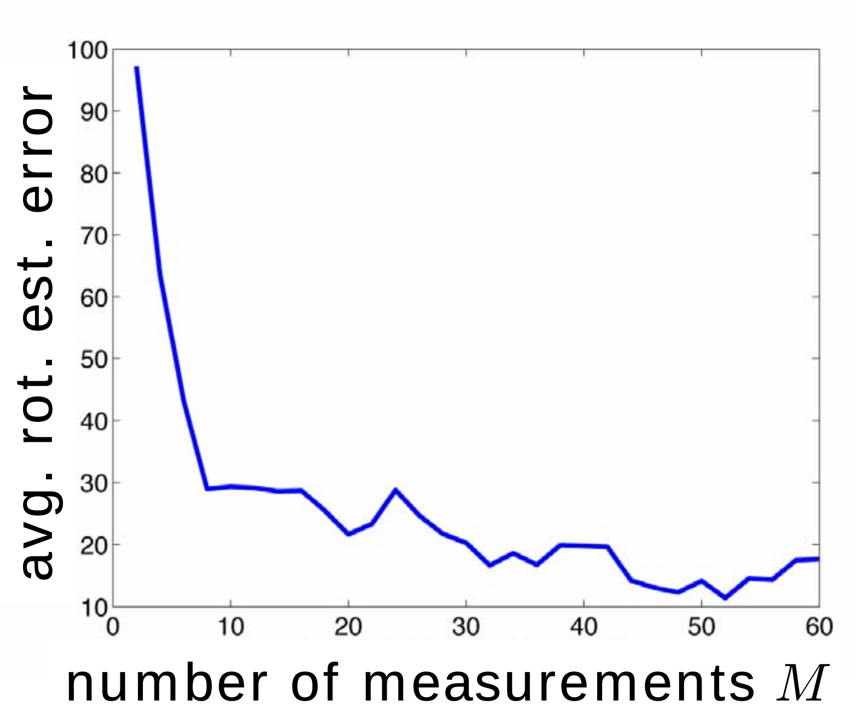


Smashed Filter – Unknown Rotation

- Training set constructed for each class with compressive measurements
  - rotations at 10°, 20°, ... , 360° (K=1 manifold)
  - identify most likely rotation for each image class
  - identify most likely class using nearest-neighbor test

- *Perfect* classification with as few as 6 measurements

- Good estimates of the viewing angle with under 10 measurements



Conclusions

- Compressive measurements are ***information scalable***

    reconstruction > estimation > classification > detection

- ***Smashed filter*:** dimension-reduced GLRT for parametrically transformed signals
    - exploits compressive measurements and manifold structure
    - broadly applicable: *targets do not have to have sparse representation in any basis*
    - effective for image classification when combined with single-pixel camera

- Current work
    - efficient parameter estimation using multiscale Newton's method [Wakin, Donoho, Choi, RGB, 05]
    - linking continuous manifold models to discrete point cloud models [Wakin, DeVore, Davenport, RGB, 05]
    - noise analysis and tradeoffs (M/N SNR penalty)
    - compressive k-NN, SVMs, ...

dsp.rice.edu/cs