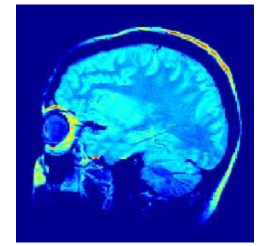
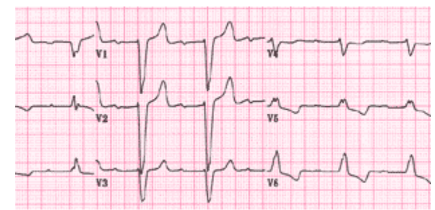
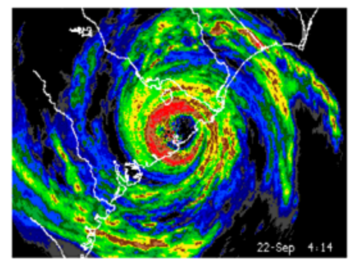
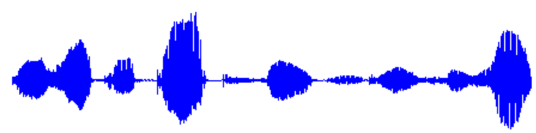
Compressive Sensing:
A New Framework for Imaging

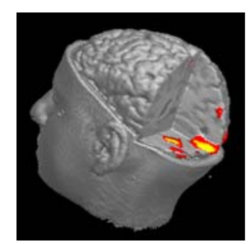
Richard Baraniuk
Justin Romberg
Robert Nowak

Rice University
Georgia Institute of Technology
University of Wisconsin-Madison





Agenda

Part I

Introduction to compressive sensing (CS) (RN)

Part II

Generalized uncertainty principles (JR)
Sparse recovery
Optimization algorithms for CS

Break

Part III

Compressive sensing in noisy environments (RN)

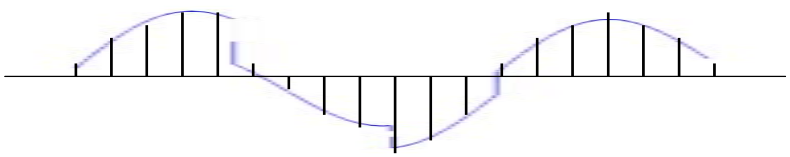
Part IV

Compressive cameras and imaging systems (JR)
Distributed compressive sensing

Part I – Introduction to Compressive Sensing

Pressure is on Digital Signal Processing

- Shannon/Nyquist sampling theorem
  - no information loss if we sample at 2x signal bandwidth
- DSP revolution:
  sample first and ask questions later



- Increasing *pressure* on DSP hardware, algorithms
  - ever faster sampling and processing rates
  - ever larger dynamic range
  - ever larger, higher-dimensional data
  - ever lower energy consumption
  - ever smaller form factors
  - multi-node, distributed, networked operation
  - radically new sensing modalities
  - communication over ever more difficult channels

Pressure is on Image Processing

- increasing pressure on signal/image processing hardware and algs to support

higher resolution / denser sampling

» ADCs, cameras, imaging systems, ...

+

large numbers of signals, images, ...

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

+

increasing numbers of modalities

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

=

deluge of data

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

Background:
Structure of Natural Images

Images

- Key ingredients of an image processing system: ***model*** *for low-level and high-level image structures*

  ex: edges, continuation

- Fairly well understood

- Wavelet models, Markov Random fields, PDEs, cartoon+texture, ...

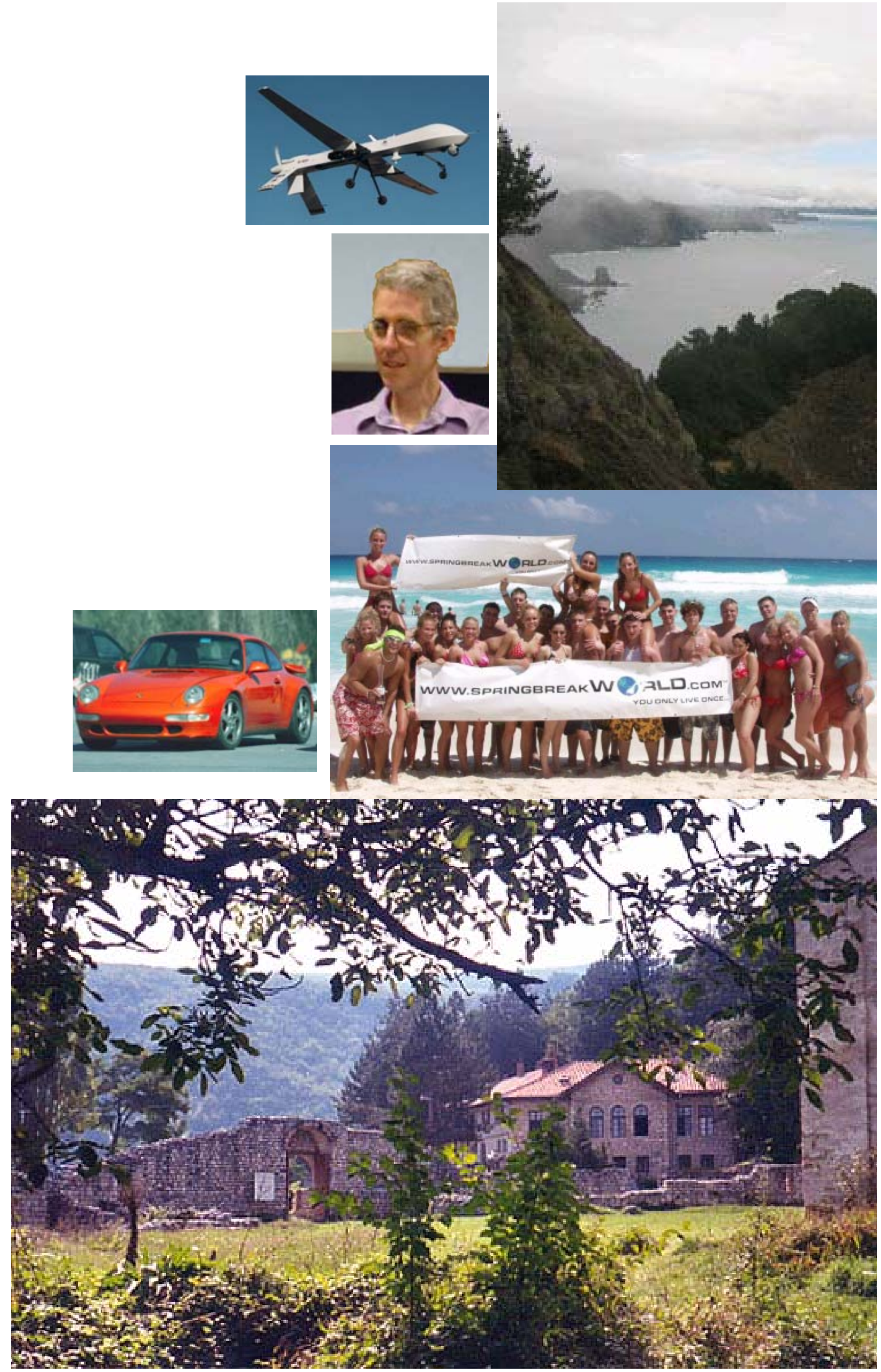
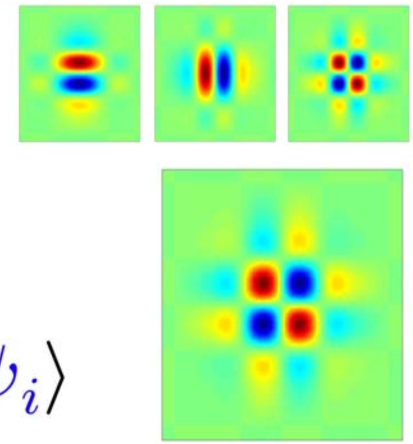


Image Representation

N
pixels

$$x = \sum_{i=1}^{N} \alpha_i \psi_i$$

$$\alpha_i = \langle x, \psi_i \rangle$$



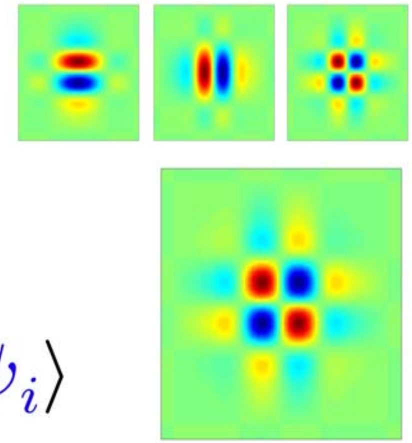
multiscale
Gabor, pyramid, wavelet, curvelet, ...
bases, frames, dictionaries

Sparse Image Representation

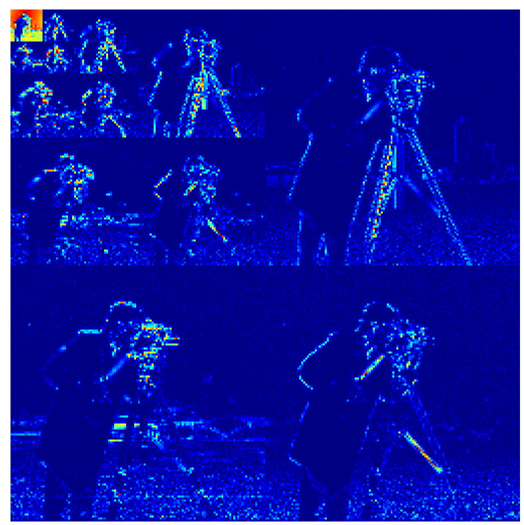
N
pixels

$$x = \sum_{i=1}^{N} \alpha_i \psi_i$$

$$\alpha_i = \langle x, \psi_i \rangle$$



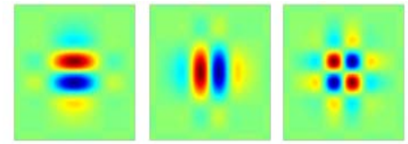
$K \ll N$
large
wavelet
coefficients



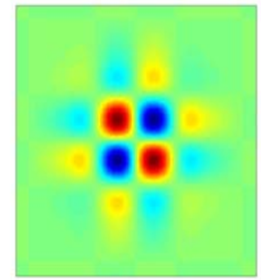
Sparse Image Representation

N
pixels

$$x = \sum_{i=1}^{N} \alpha_i \, \psi_i$$



$$\alpha_i = \langle x, \ \psi_i \rangle$$



$$x \approx \sum_{K \ll N \text{ largest terms}} \alpha_i \, \psi_i$$

compression:
JPEG, JPEG2000, MPEG, ...

Sparse Models are *Nonlinear*

+

=

Sparse Models are *Nonlinear*

N
pixels

$K \ll N$
large
wavelet
coefficients

$$x \approx \sum_{K \ll N \text{ largest terms}} \alpha_i \, \psi_i$$

model for all K-sparse
images

Sparse Models are *Nonlinear*

N pixels

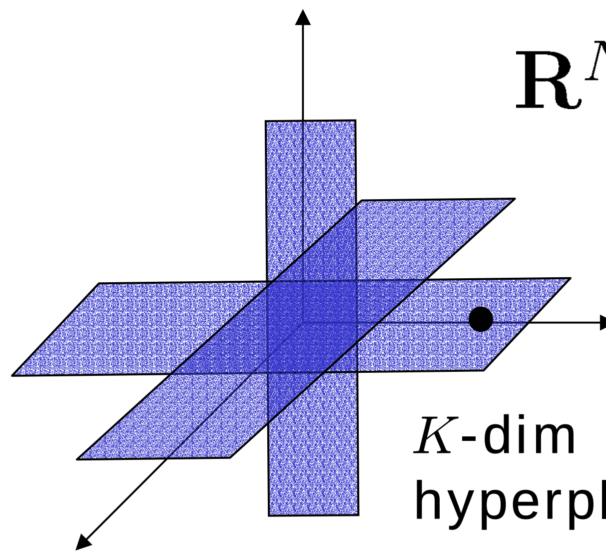
$K \ll N$ large wavelet coefficients

$$x \approx \sum_{K \ll N \text{ largest terms}} \alpha_i \psi_i$$

model for all K-sparse images:

union of subspaces

(aligned with coordinate axes)

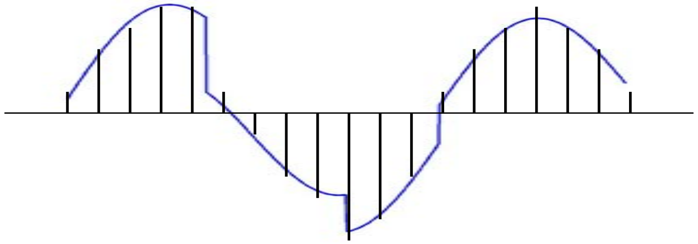


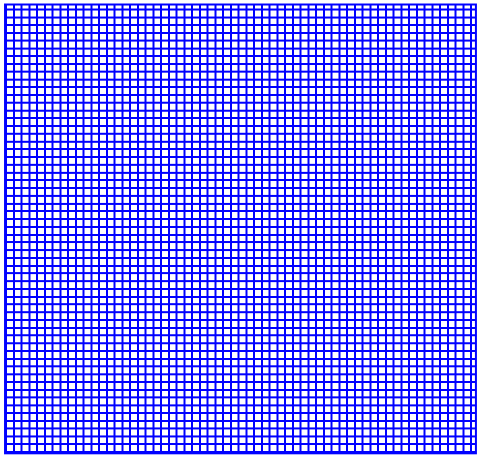
Overview of Compressive Imaging

Data Acquisition and Representation

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

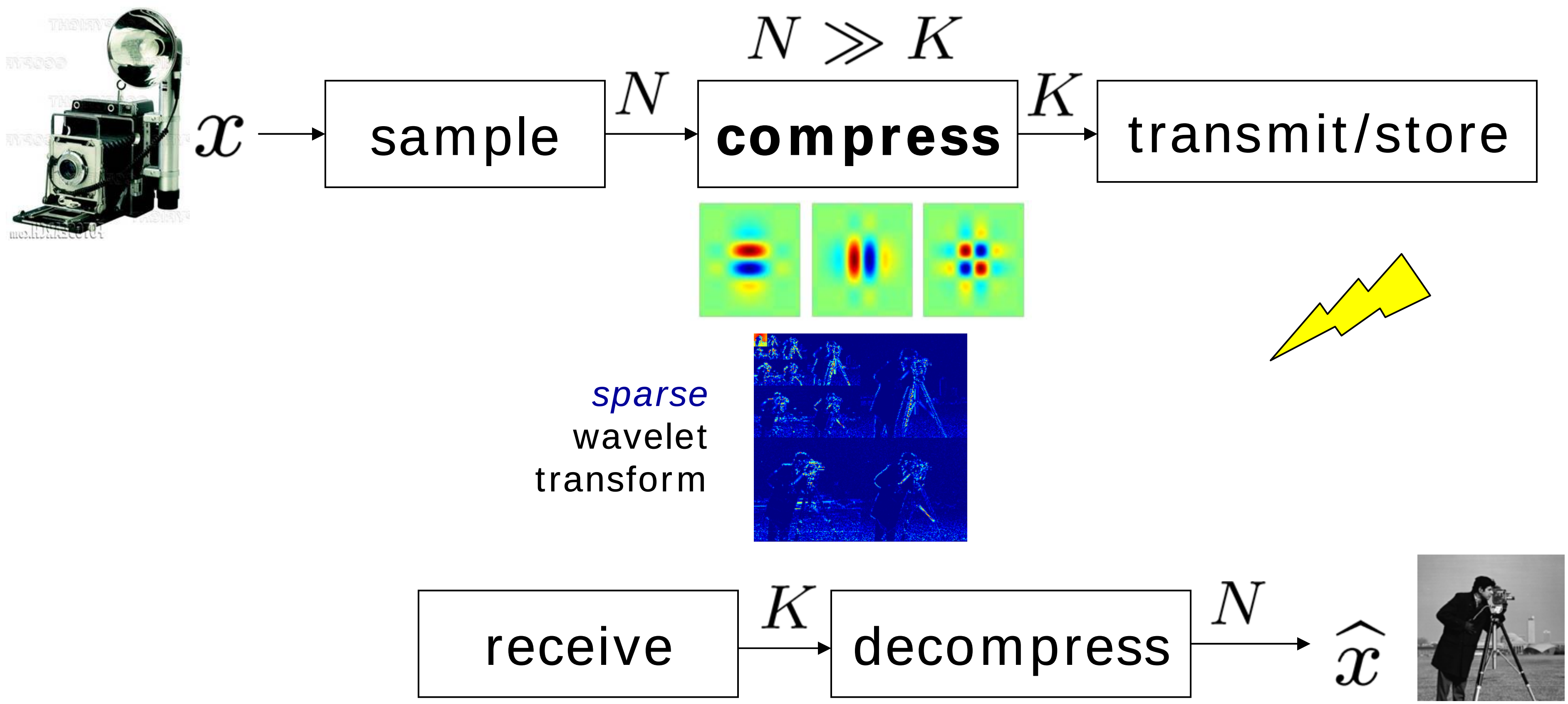
- Foundation: *Shannon sampling theorem*
  - *Nyquist rate*: must sample at 2x highest frequency in signal



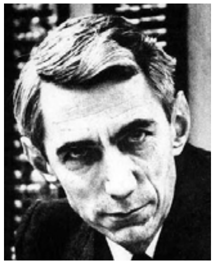


Sensing by *Sampling*

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



From Samples to *Measurements*



- Shannon was a *pessimist*
  - worst case bound for *any* bandlimited data

- ***Compressive sensing*** (CS) principle

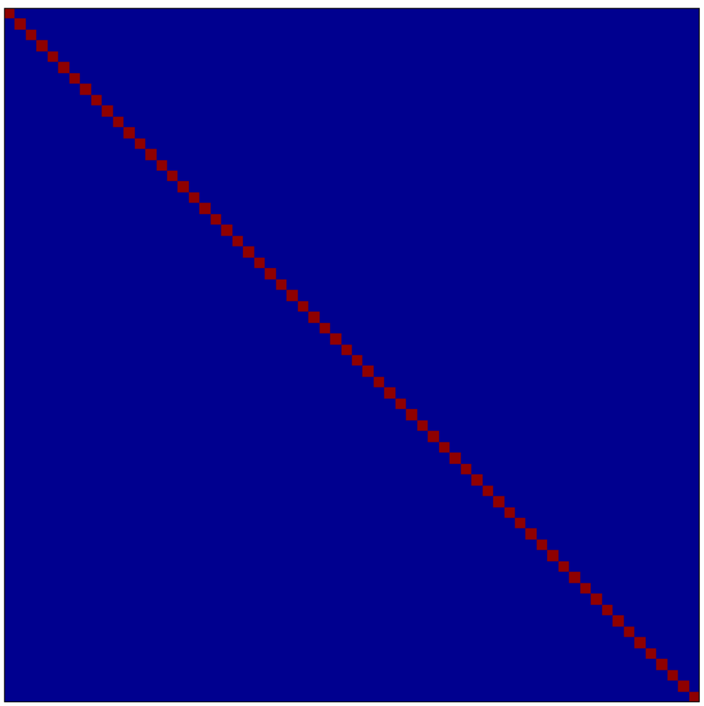
  "sparse signal statistics can be recovered from
  a small number of ***nonadaptive linear measurements***"

  - integrates sensing, compression, processing
  - based on new *uncertainty principles* and concept of *incoherency* between two bases

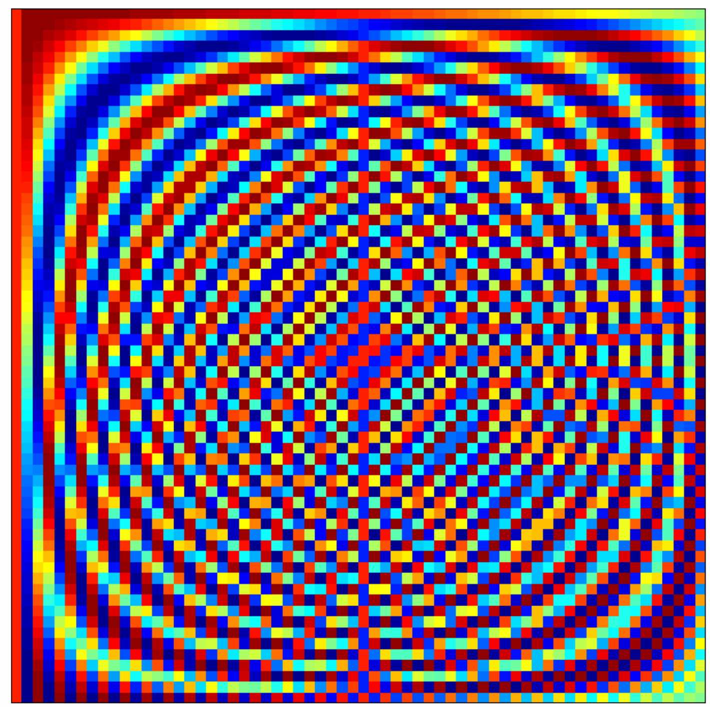
Incoherent Bases

- Spikes and sines (Fourier) (Heisenberg)

$$\Psi = I$$



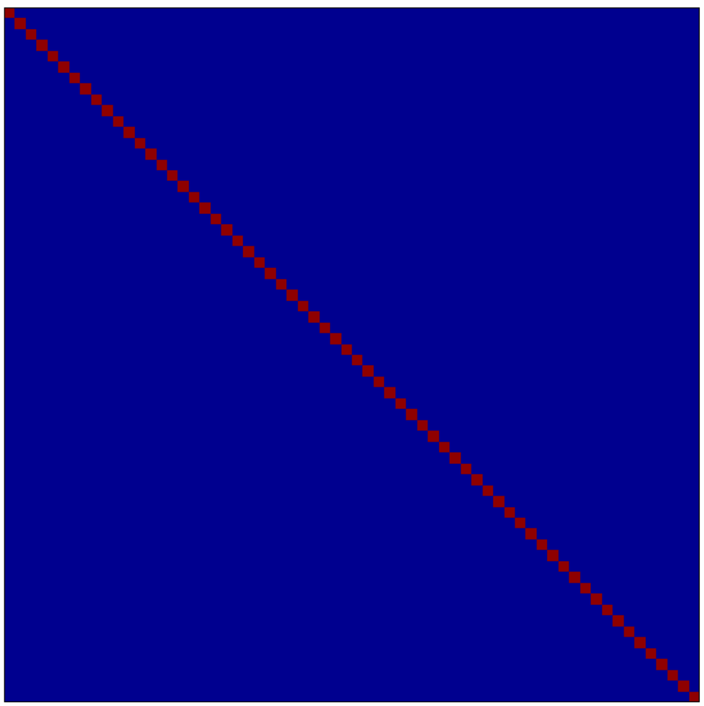
$$\Phi = \text{idct}(I)$$



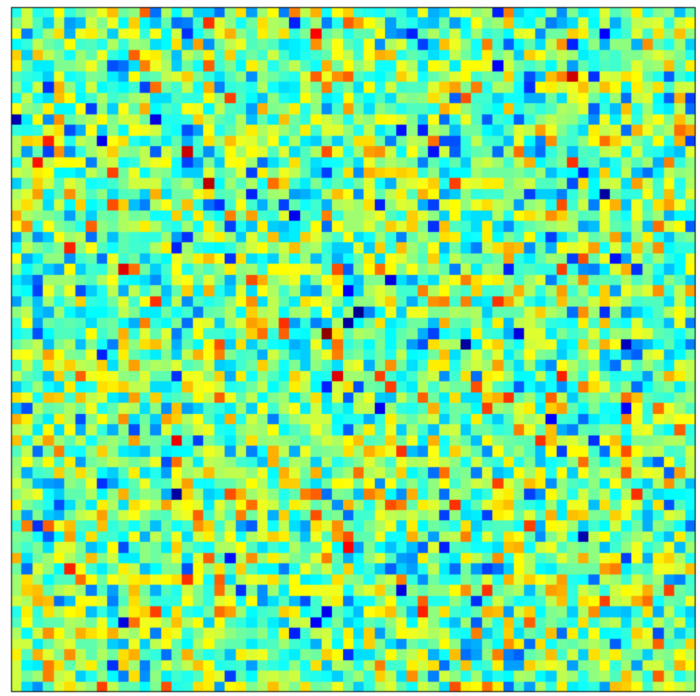
Incoherent Bases

- Spikes and “random basis”

$\Psi = I$



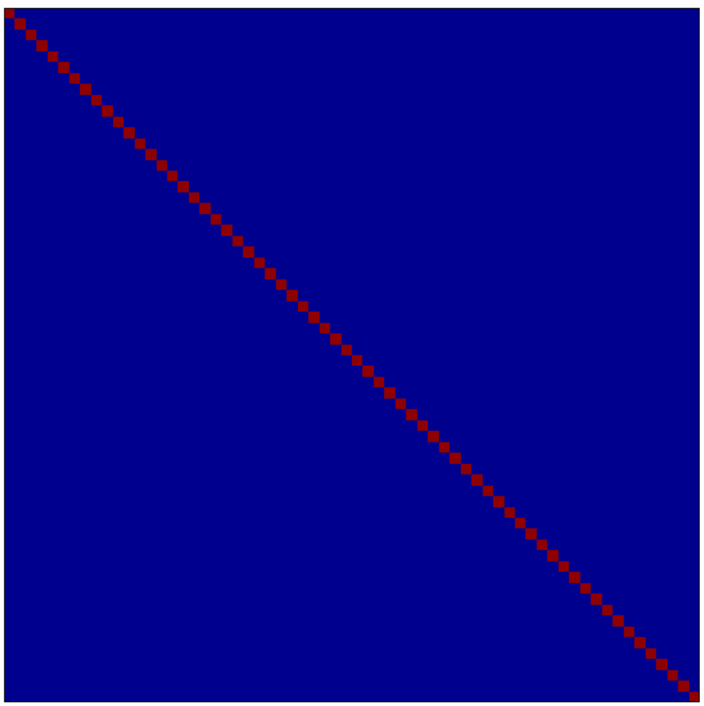
$\Phi = \mathrm{randn}(N, N)$



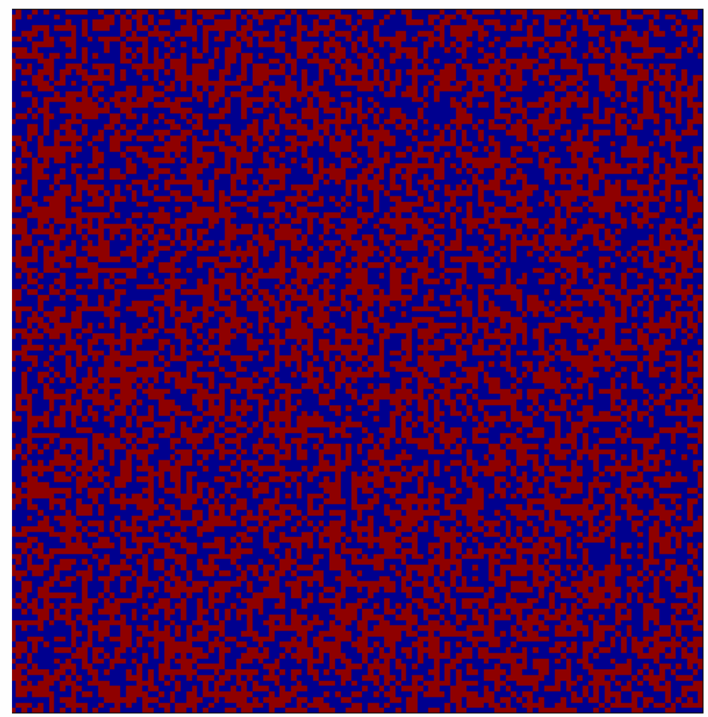
Incoherent Bases

- Spikes and “random sequences” (codes)

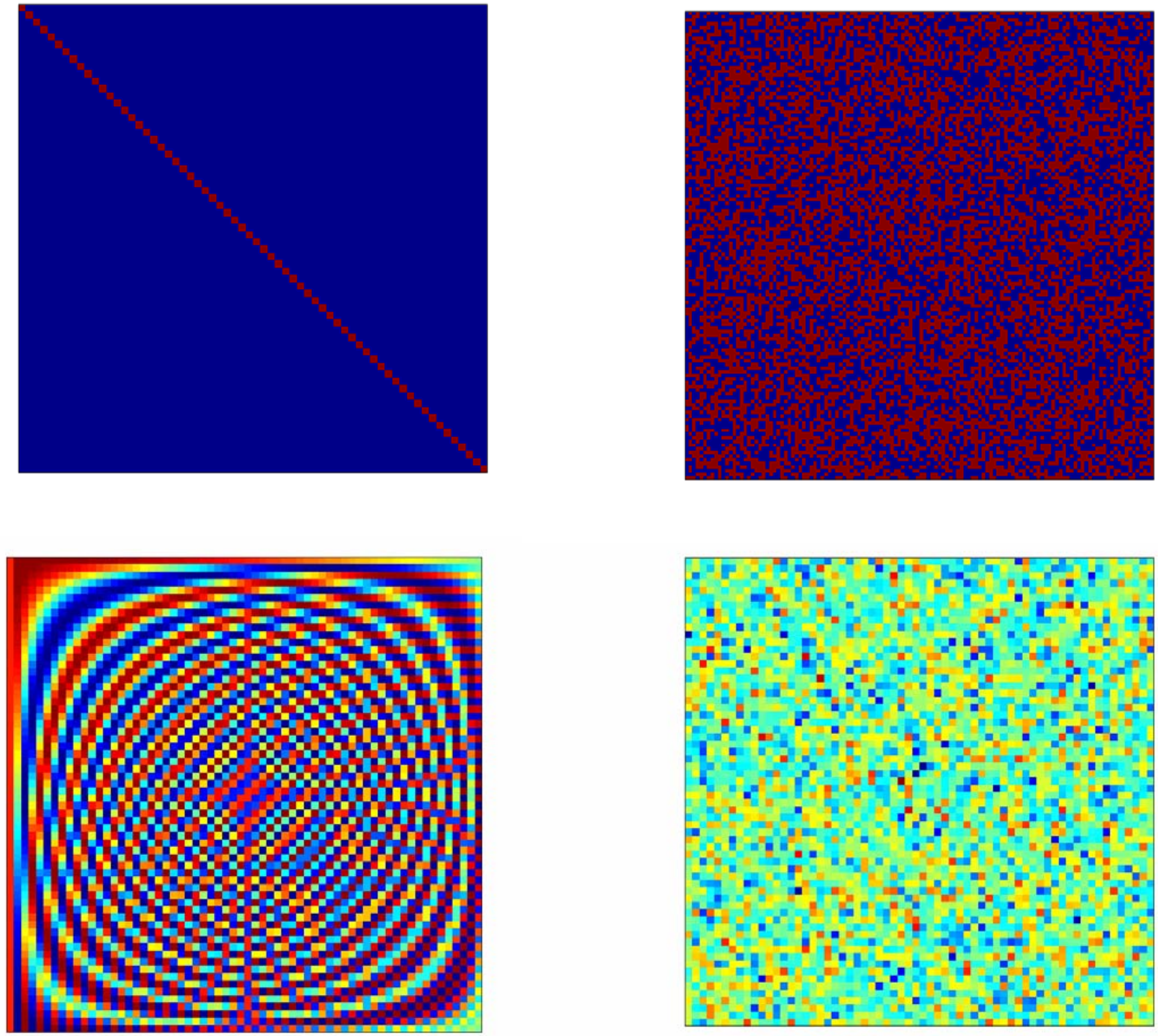
$\Psi = I$



Φ



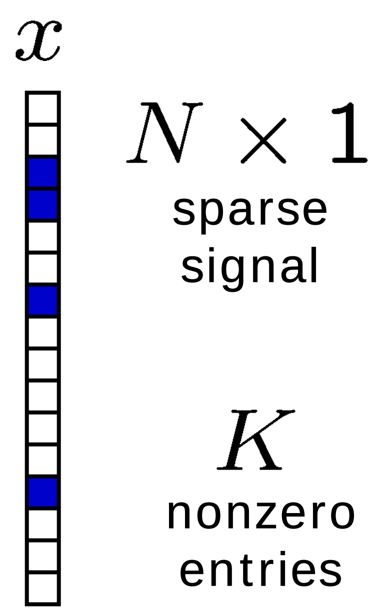
Incoherent Bases



Compressive Sensing

[Candes, Romberg, Tao; Donoho]

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

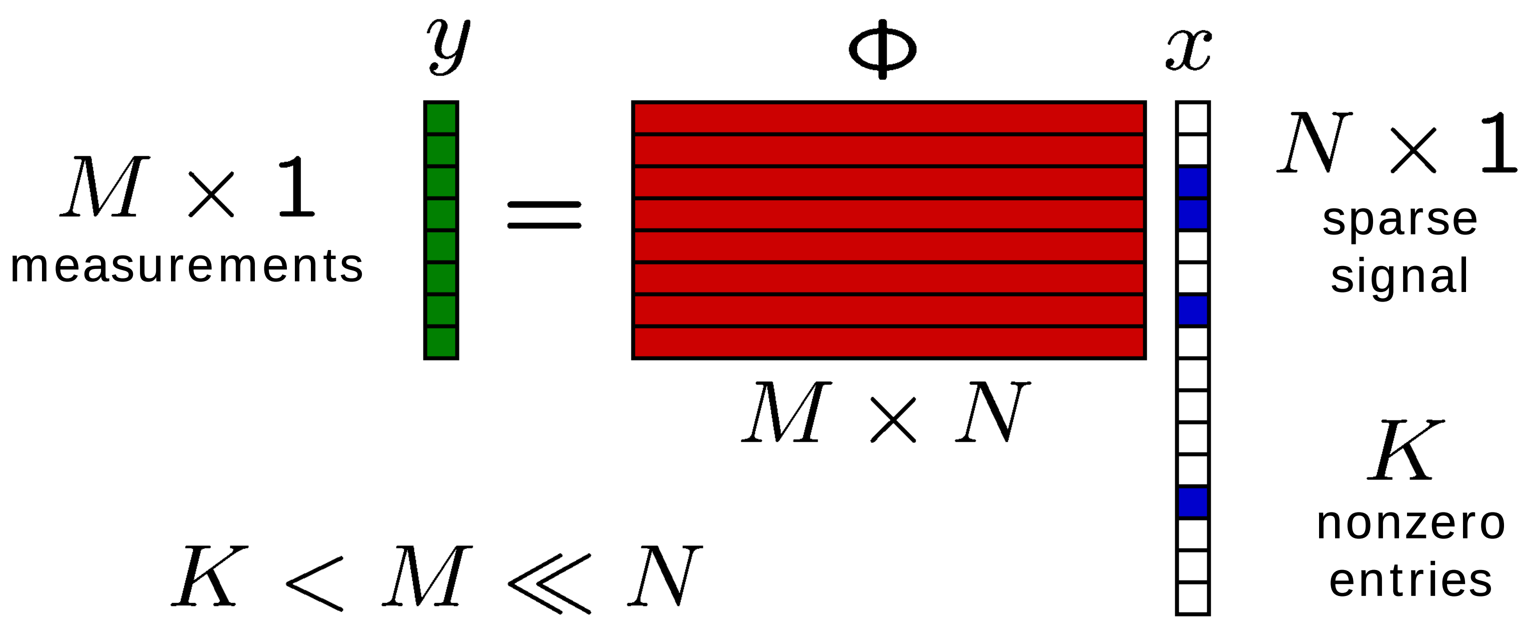


Compressive Sensing

[Candes, Romberg, Tao; Donoho]

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

- Replace samples with *few linear projections* $y = \Phi x$

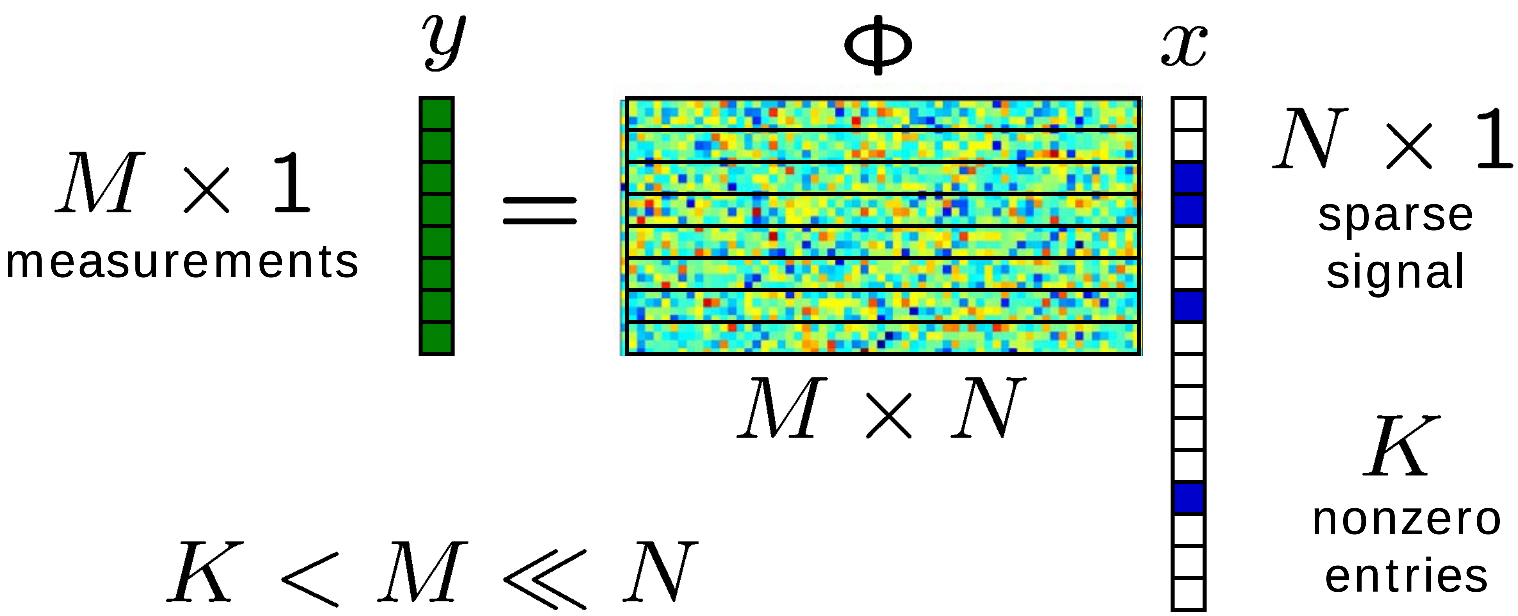


$$K < M \ll N$$

Compressive Sensing

[Candes, Romberg, Tao; Donoho]

- Signal x is K-*sparse* in basis/dictionary Ψ
  - WLOG assume sparse in space domain $\Psi = I$
- Replace samples with *few linear projections* $y = \Phi x$

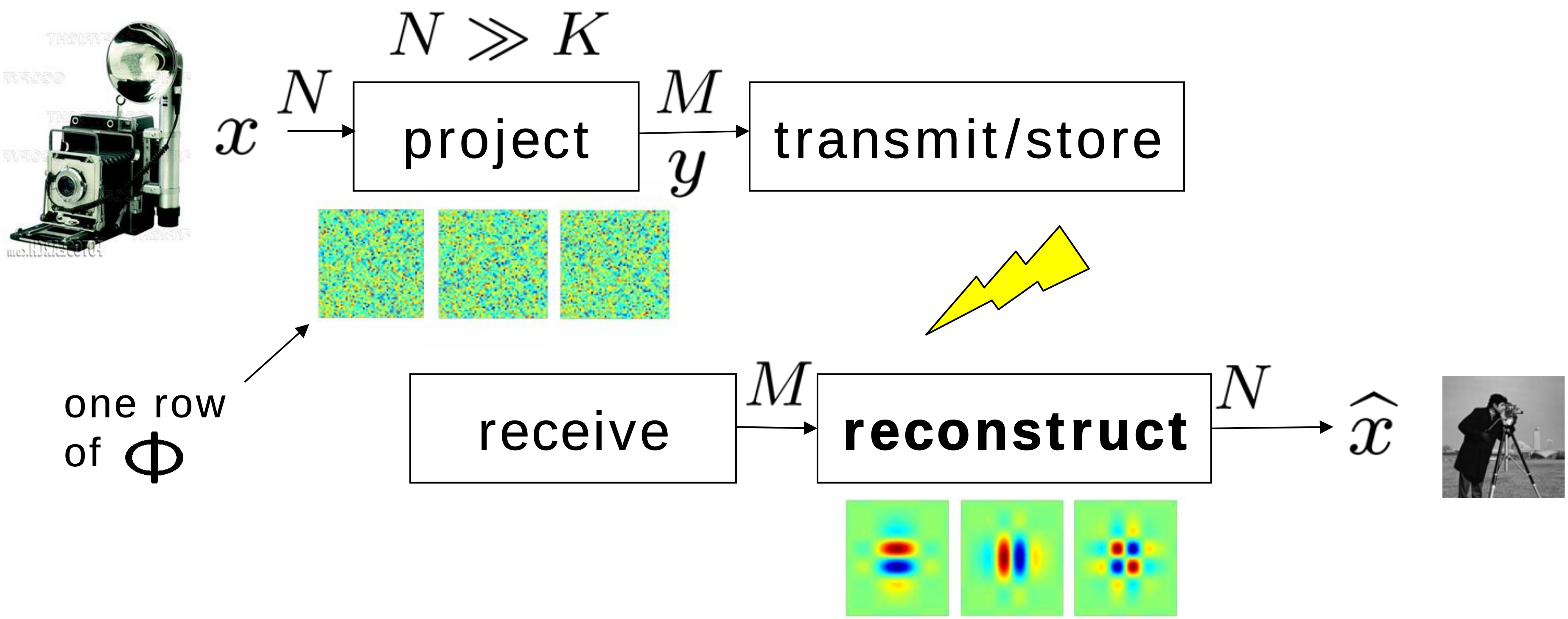


$K < M \ll N$

- *Random* measurements Φ will work!

Compressive Sensing

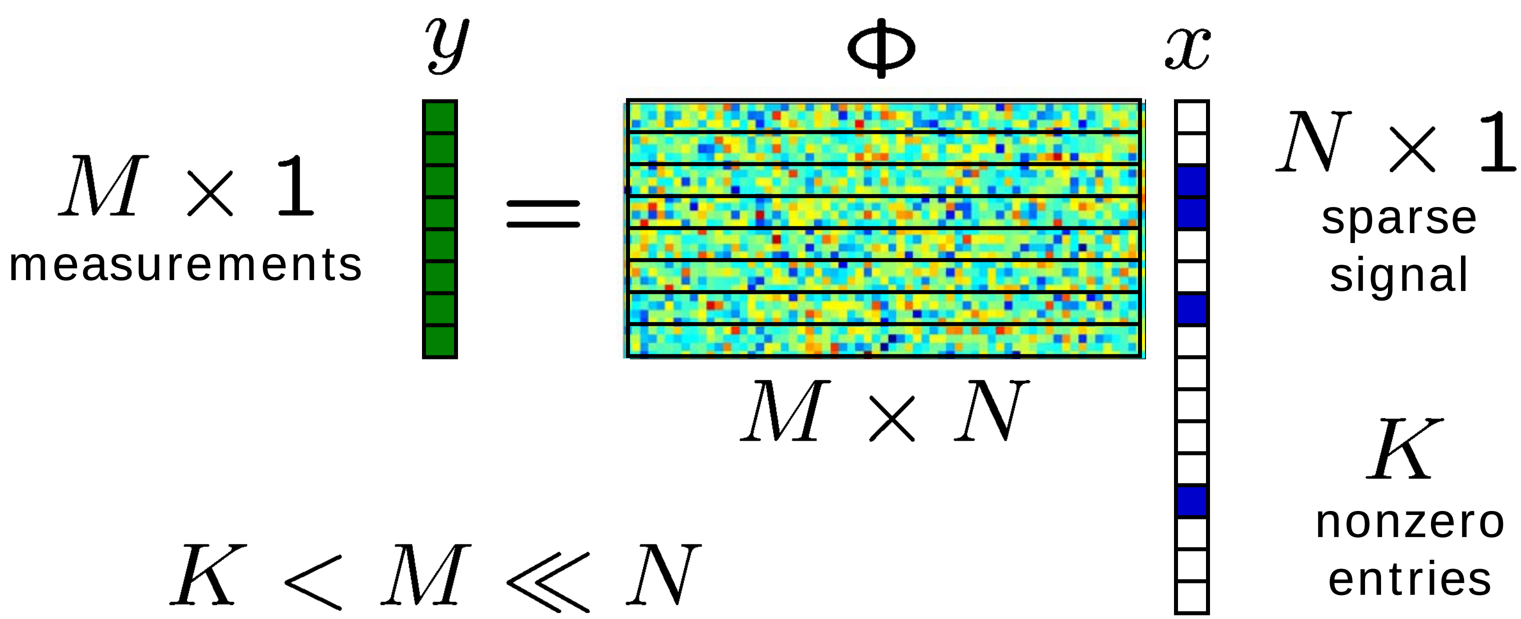
- Measure linear projections onto *incoherent* basis where data is *not sparse/compressible*



- Reconstruct via *nonlinear processing* (optimization)
  (using sparsity-inducing basis)

CS Signal Recovery

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



$$K < M \ll N$$

CS Signal Recovery

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

- ***L_2*** **fast** $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$

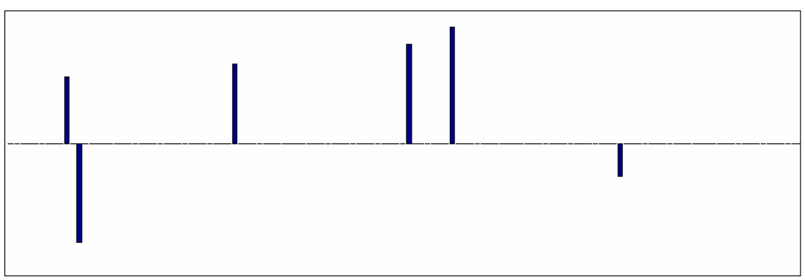
  ↑ *energy*

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

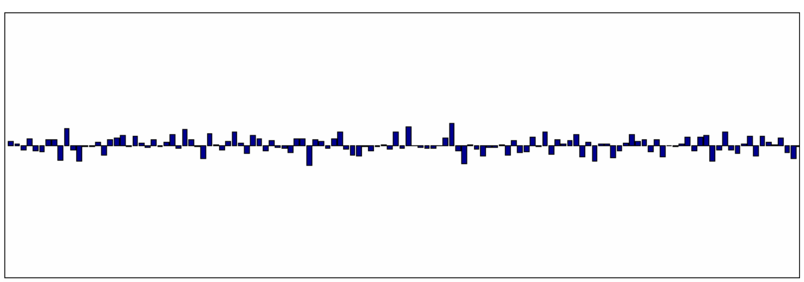
CS Signal Recovery

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

- ***L*₂** fast, **wrong** $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$



x



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

CS Signal 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, slow**
  only M=K+1 measurements required to perfectly reconstruct K-sparse signal
  [Bresler; Wakin et al]

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

number of nonzero entries

CS Signal Recovery

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

- ***L_2*** fast, wrong $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$

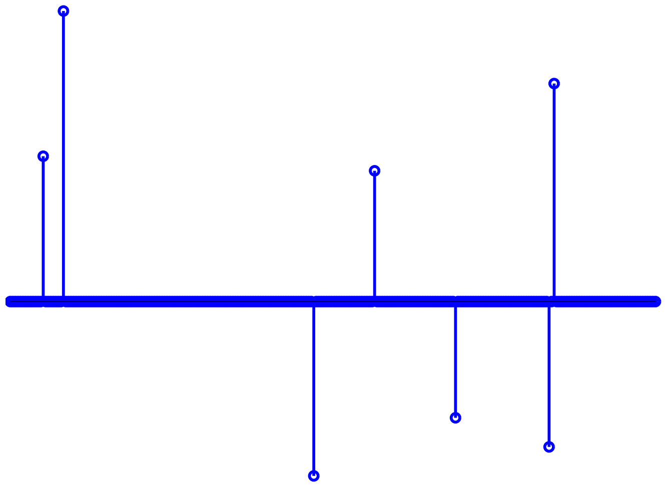
- ***L_0*** correct, slow $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_0$

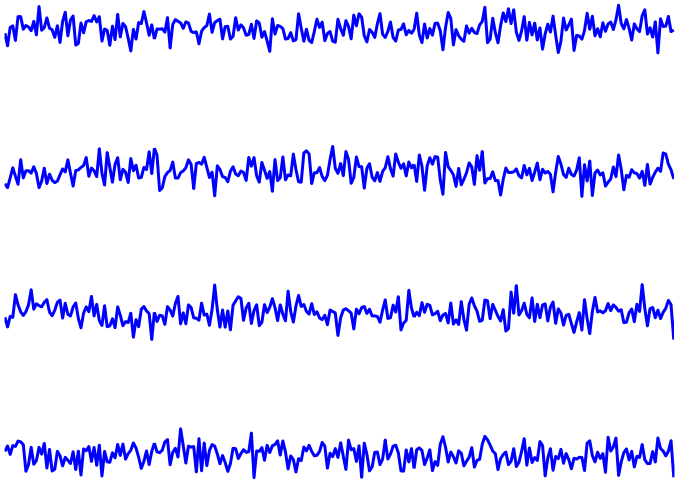
- ***L_1*** **correct,**
  mild oversampling $\widehat{x} = \arg\min_{y=\Phi x} \|x\|_1$
  [Candes et al, Donoho] *linear program*

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

Part II –
Generalized
Uncertainty Principles
and
CS Recovery Algorithms

What Makes Sparse Recovery Possible?



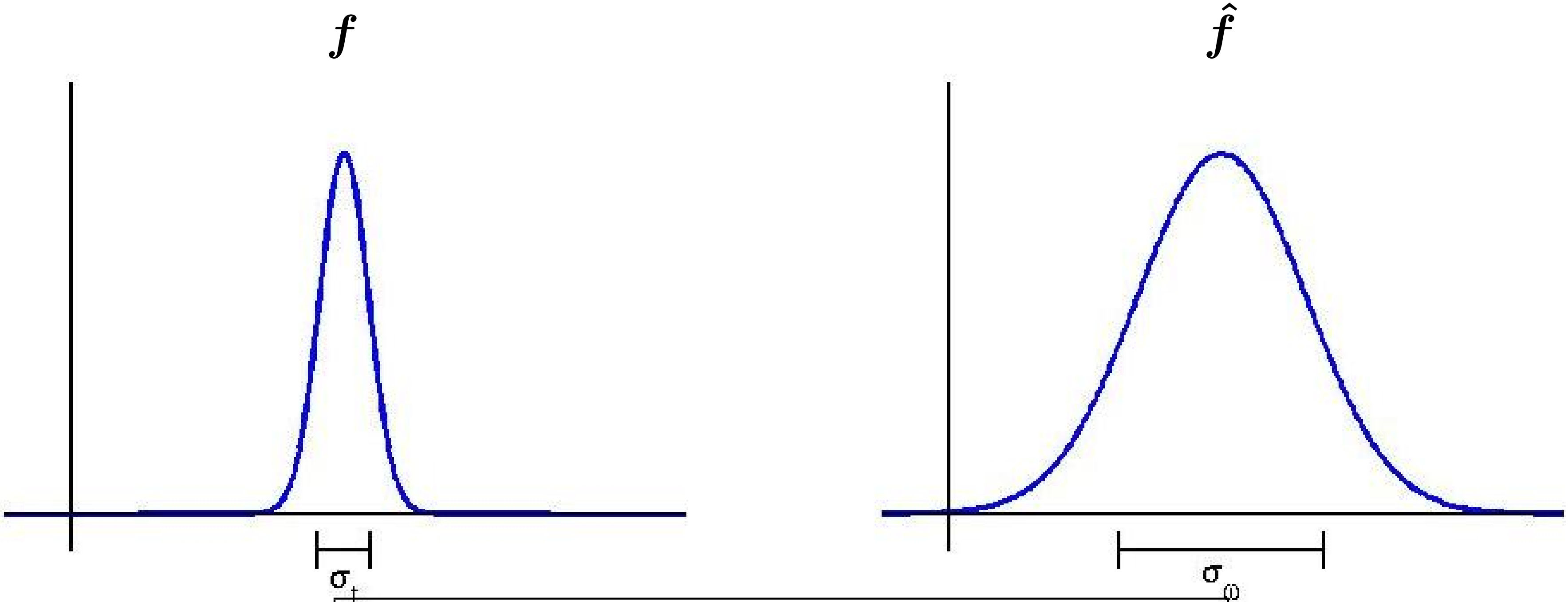


- Signal is local, measurements are global
- Each measurement picks up a little information about each component
- Triangulate significant components from measurements
- Formalization: Relies on uncertainty principles between sparsity basis and measurement system

Uncertainty Principles

Uncertainty Principles: Classical

- Heisenberg (1927)
  Uncertainty principle for continuous-time signals

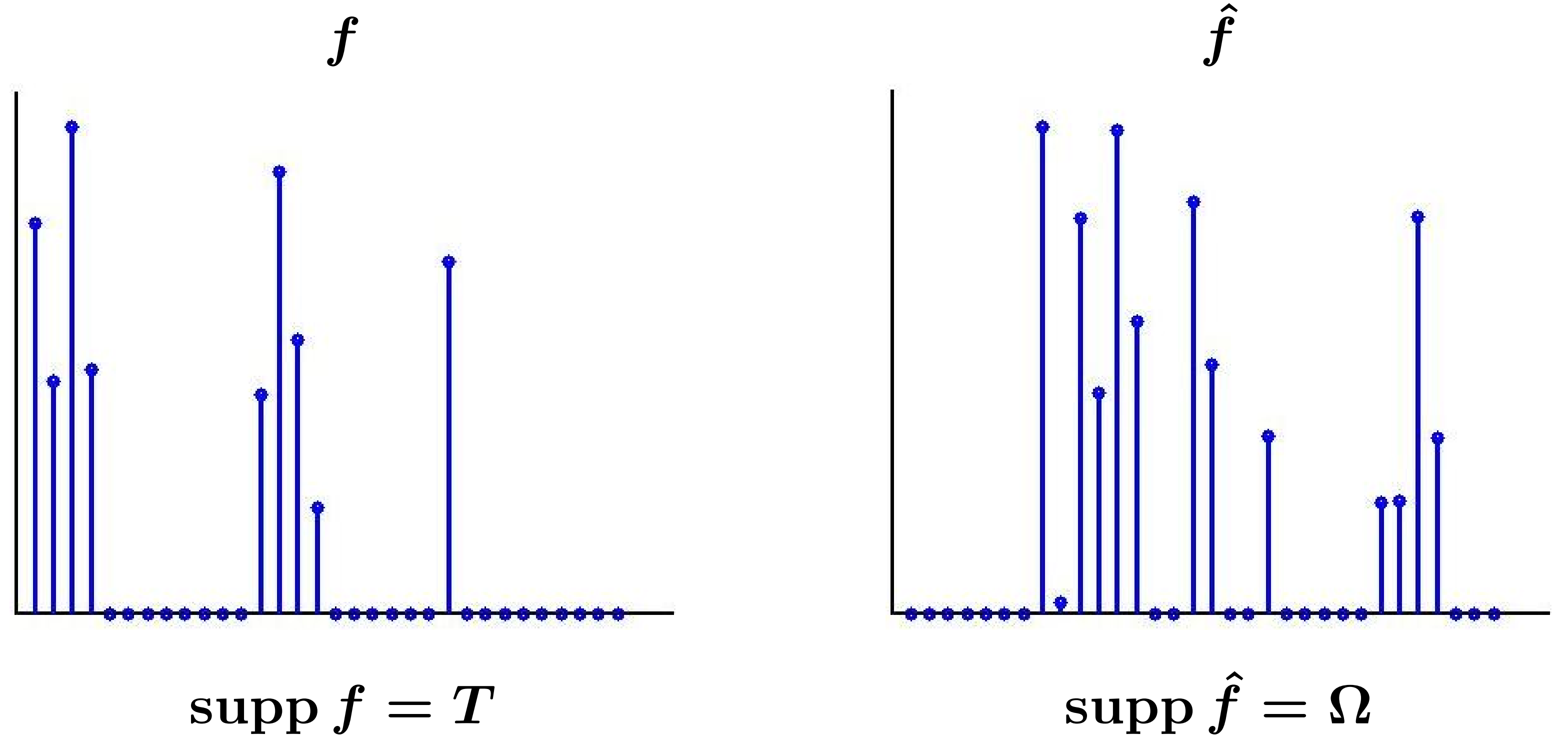


$$\boldsymbol{\sigma_t \sigma_\omega \geq \frac{1}{4\pi}} \quad \text{or} \quad \boldsymbol{\sigma_t + \sigma_\omega \geq \frac{1}{\sqrt{\pi}}}$$

- Limits *joint resolution* in time and frequency
- Relates *second moments* of continuous-time $f, \hat{f}$
- Extended by Landau, Pollack, Slepian and others in the 1960s

Uncertainty Principles: Discrete

- Donoho and Stark (1989)
  Discrete uncertainty principle for $\mathbb{C}^N$



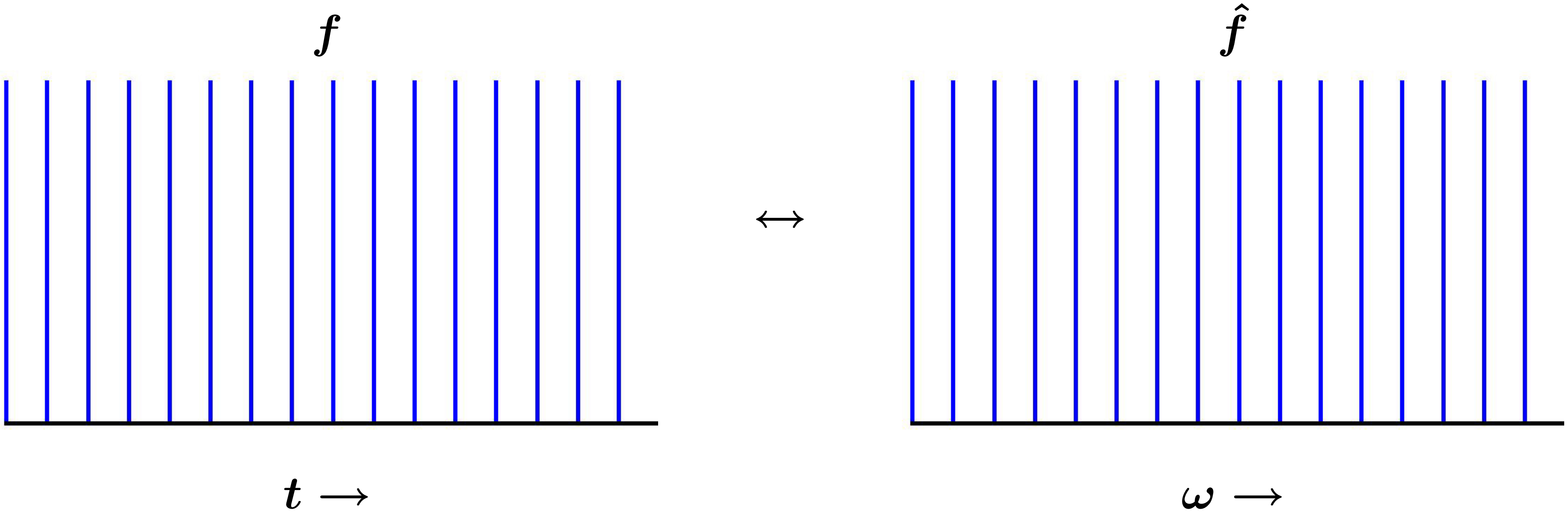
$$|T| \cdot |\Omega| \geq N \qquad \text{or} \qquad |T| + |\Omega| \geq 2\sqrt{N}$$

- Implications: recovery of extremely sparse signals,
  finding unique sparse decompositions in union of bases

- Relates *number of non-zero terms* of discrete f, $\hat{f}$

Dirac Comb

- The discrete uncertainty principle is *exact*
- Limiting case is the “Dirac comb” or “picket fence”:



- $\sqrt{N}$ spikes spaced $\sqrt{N}$ apart
- Invariant under Fourier transform ($f = \hat{f}$)
- $|T| + |\Omega| = 2\sqrt{N}$

Quantitative Uncertainty Principles

- Discrete UP: Tells us if $f, \hat{f}$ can be supported on T, Ω
- *Quantitative Uncertainty Principle*:
  Tells us how much of f_T can be concentrated on Ω in Fourier domain
  - subset of time domain T
  - signal f_T is supported on T (zero outside of T)
  - fraction of energy on Ω in frequency:

$$\begin{aligned} \|1_\Omega \cdot \hat{f}_T\|_2^2 &= \langle F_{\Omega T} f_T, F_{\Omega T} f_T \rangle \\ \Rightarrow \quad \lambda_{\min}(F_{\Omega T}^* F_{\Omega T}) \leq \|1_\Omega \cdot \hat{f}_T\|_2^2 &\leq \lambda_{\max}(F_{\Omega T}^* F_{\Omega T}) \end{aligned}$$

  - $F_{\Omega T}$: $|\Omega| \times |T|$ matrix,
    takes time-domain coefficients on T,
    returns frequency domain coefficients on Ω
  - Constructing $F_{\Omega T}$: take DFT matrix F,
    extract rows corresponding to Ω, then columns corresponding to T
- Quantitative UPs $\Leftrightarrow$ eigenvalues of minors of transform matrices

A Quantitative Robust Uncertainty Principle

(Candès, R, Tao '04)

- Choose sets T, Ω uniformly at random with size

$$|T| + |\Omega| \leq \text{Const} \cdot \frac{N}{\sqrt{\log N}}$$

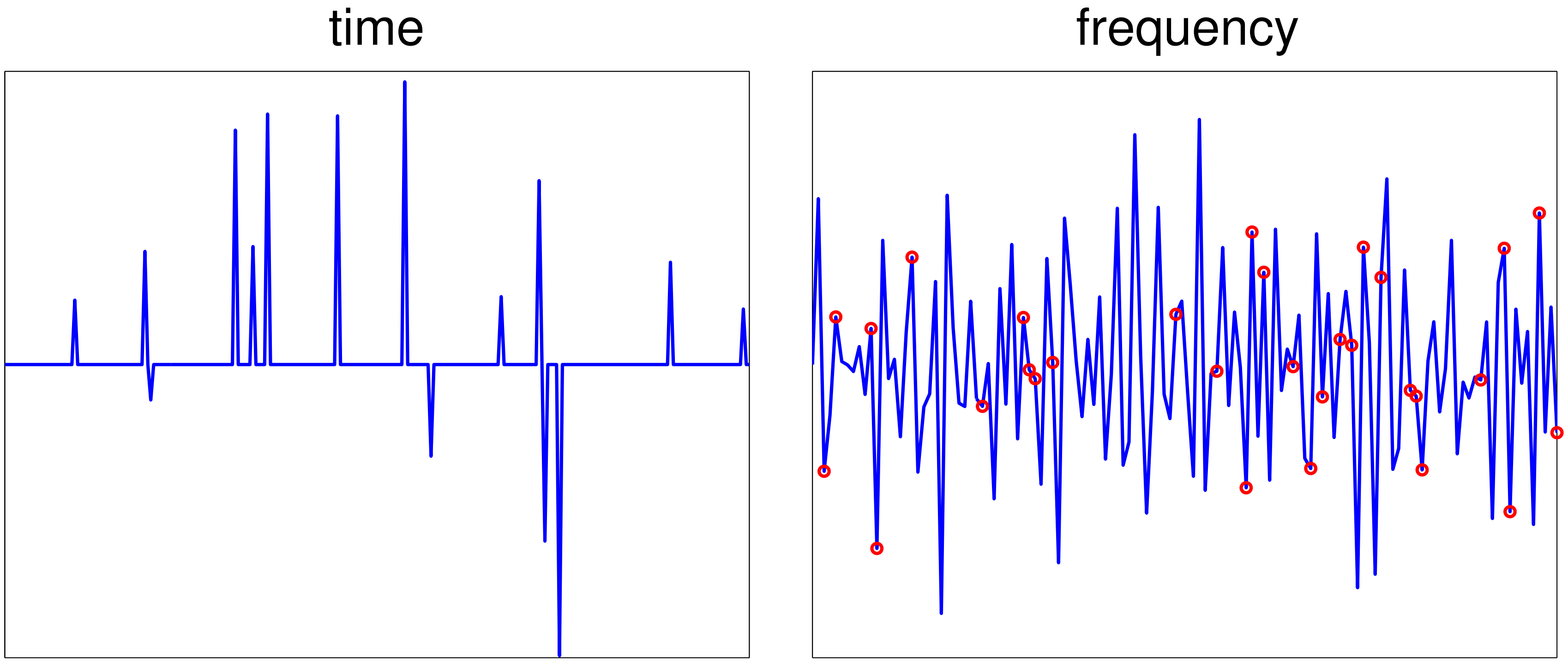
- With extremely high probability,

$$\|F_{\Omega T}\|^2 \leq 1/2$$

  - signal f supported on T: no more than $1/2$ of energy of $\hat{f}$ on Ω
  - spectrum $\hat{f}$ supported on Ω: no more than $1/2$ energy of f on T

- Randomness tells us only a very small number of T, Ω are pathological
- Factor $1/2$ is somewhat arbitrary,
  can be made smaller for large N by adjusting constant
- Compare to $|T| + |\Omega| \sim \sqrt{N}$

Uncertainty Principles: Time and Frequency



sparse signal

red circles = Ω

Slogan: *Concentrated in time, spread out in frequency*

Recovering a Spectrally Sparse Signal from a Small Number of Samples

Sampling a Superposition of Sinusoids

- Suppose f is sparse in the Fourier domain:

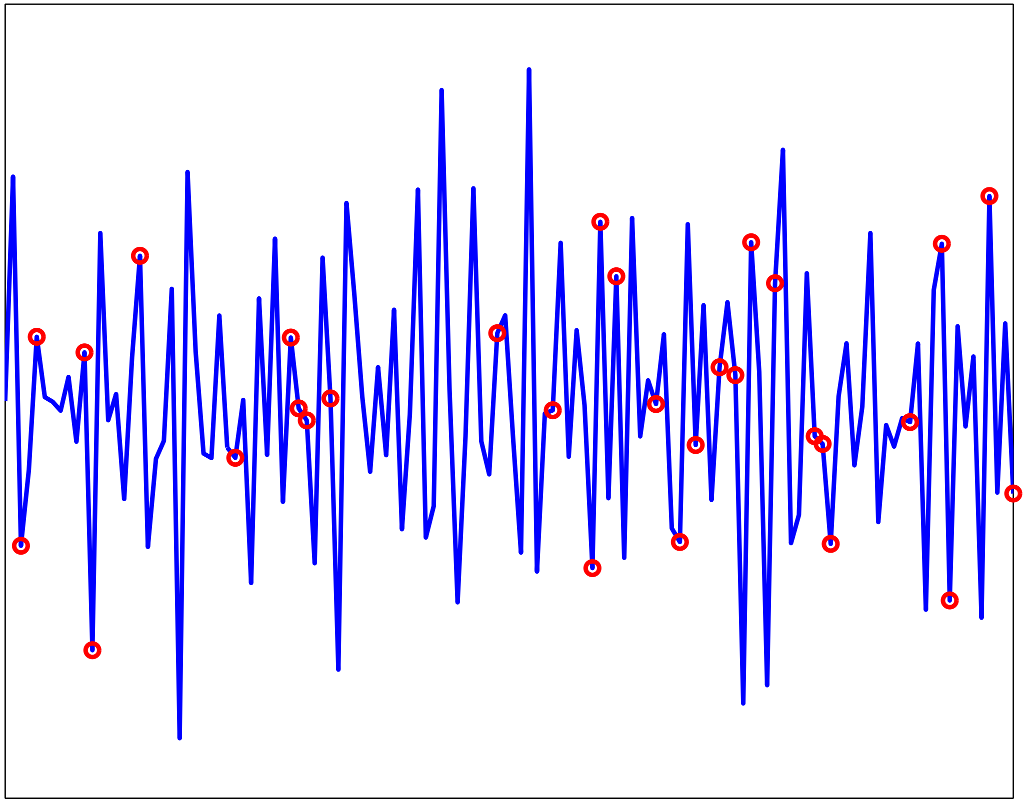
$$\hat{f}(\omega) = \sum_{b=1}^{B} \alpha_b \delta(\omega_b - \omega) \quad \Leftrightarrow \quad f(t) = \sum_{b=1}^{B} \alpha_b e^{i\omega_b t}$$

  f is a superposition of B complex sinusoids.

- Note: frequencies $\{\omega_b\}$ and amplitudes $\{\alpha_b\}$ are *unknown*.
- Take K samples of f at locations $t_1, \ldots, t_k$

Sampling Example

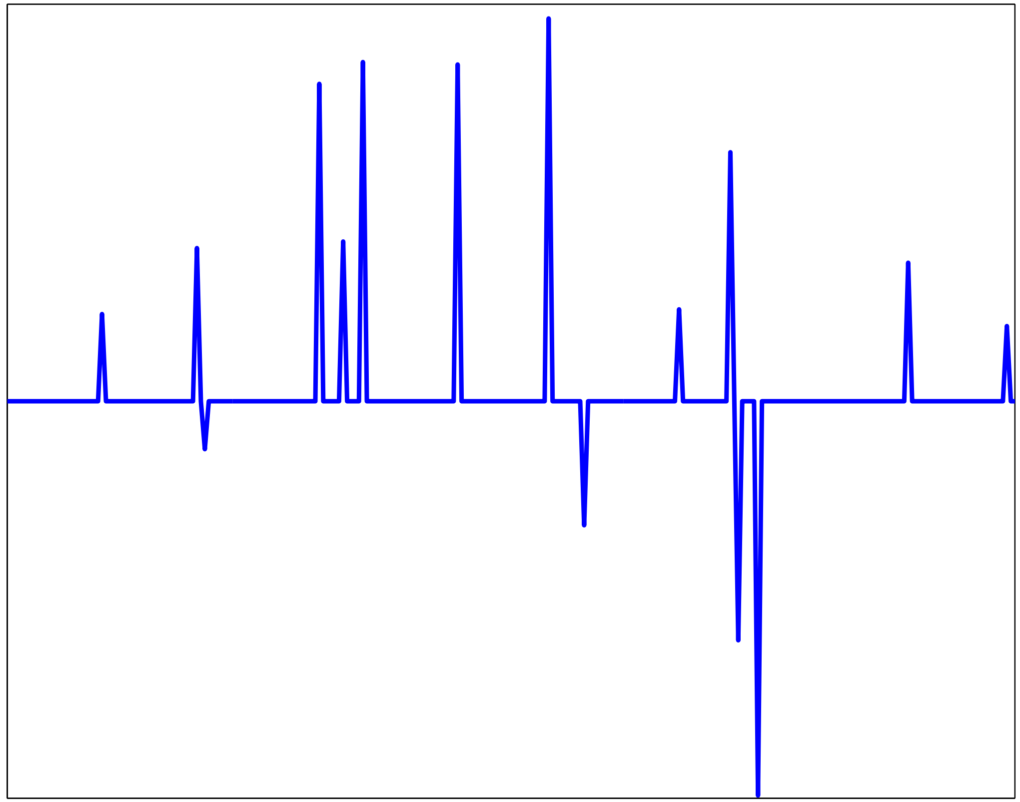
Time domain $f(t)$



Measure K samples

(red circles = samples)

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



B nonzero components

$\#\{\omega : \hat{f}(\omega) \neq 0\} := \|\hat{f}\|_{\ell_0} = B$

Sparse Recovery

- We measure K samples of f

$$y_k = f(t_k), \quad k = 1, \ldots, K$$

- Find signal with *smallest frequency domain support* that matches the measured samples

$$\min_g \ \|\hat{g}\|_{\ell_0} \quad \text{subject to} \quad g(t_k) = y_k, \ k = 1, \ldots, K$$

where $\|\hat{g}\|_{\ell_0} := \#\{\omega : \hat{g}(\omega) \neq 0\}$.

- **Theorem**: If $\|\hat{f}\|_{\ell_0} = B$, we can recover f from (almost) any set of

$$K \geq \text{Const} \cdot B \cdot \log N$$

samples.

- The program is absolutely intractable (combinatorial, NP hard).

Convex Relaxation

- Convex relaxation: use ℓ_1 norm as a proxy for sparsity

$$\|u\|_{\ell_1} := \sum_m |u(m)|$$

- Recover from samples $y_k = f(t_k)$ by solving

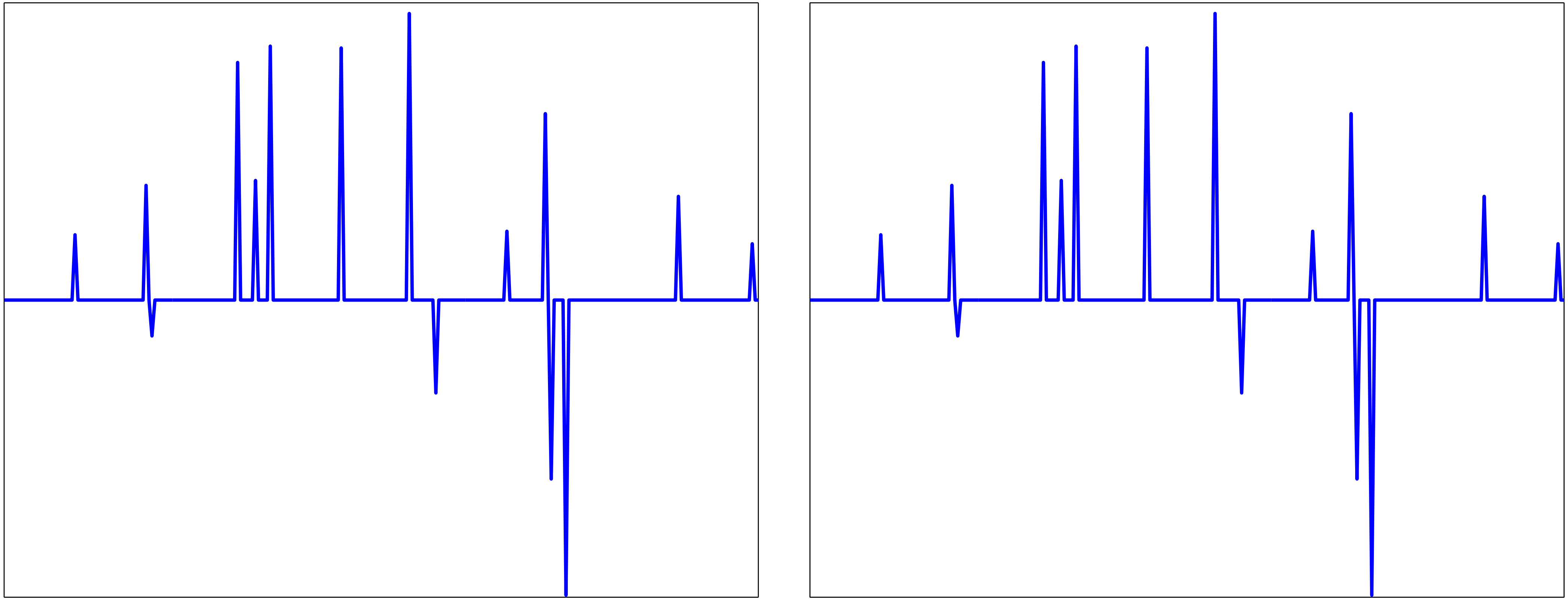
$$(P1) \quad \min_g \|\hat{g}(\omega)\|_{\ell_1} \quad \text{subject to} \quad g(t_k) = y_k, \;\; k = 1, \ldots, K$$

- Very tractable; linear or second-order cone program
- Surprise: $(P1)$ still recovers sparse signals *perfectly*.

ℓ_1 Reconstruction

Reconstruct by solving

$$\min_{g} \|\hat{g}\|_{\ell_1} := \min \sum_{\omega} |\hat{g}(\omega)| \quad \text{subject to} \quad g(t_k) = f(t_k), \quad k = 1, \ldots, K$$



original $\hat{f}$, $B = 15$ *perfect* recovery from 30 samples

A Recovery Theorem

- **Exact Recovery Theorem** (Candès, R, Tao, '04)
  - Suppose $\hat{f}$ is supported on set of size B
  - Select K sample locations $\{t_k\}$ "at random" with

$$K \geq \text{Const} \cdot B \cdot \log N$$

  - Take time-domain samples (measurements) $y_k = f(t_k)$
  - Solve

$$\min_g \|\hat{g}\|_{\ell_1} \quad \text{subject to} \quad g(t_k) = y_k, \quad k = 1, \dots, K$$

  - Solution is *exactly* f with extremely high probability.

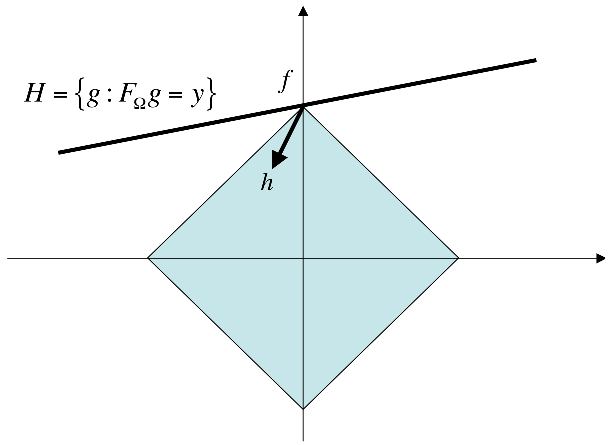
- In theory, $\text{Const} \approx 22$
- In practice, perfect recovery occurs when $K \approx 2B$ for $N \approx 1000$.
- *In general, minimizing ℓ_1 finds f from $K \sim B \log N$ samples*

Nonlinear Sampling Theorem

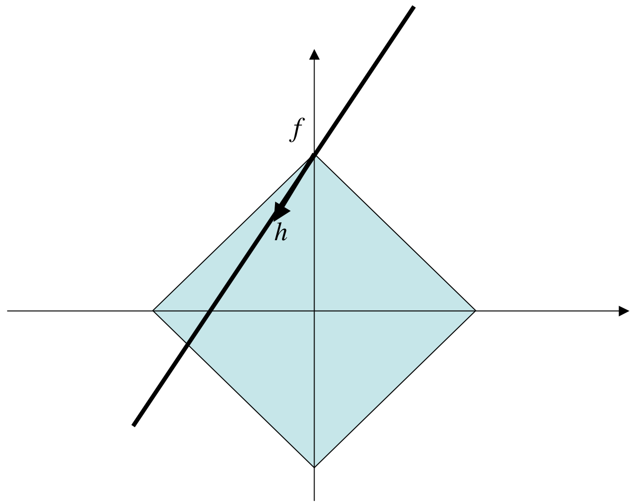
- $\hat{f} \in \mathbb{C}^N$ supported on set Ω in Fourier domain
- Shannon sampling theorem:
  - Ω is a known connected set of size B
  - exact recovery from B equally spaced time-domain samples
  - linear reconstruction by sinc interpolation
- Nonlinear sampling theorem:
  - Ω is an *arbitrary and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

Geometrical Viewpoint

good



bad



- UP for T, Ω: for $\hat{f}$ supported on Ω,

$$\frac{1}{2} \cdot \frac{|T|}{N} \leq \|F_T^* \hat{f}\|_2^2 \leq \frac{3}{2} \cdot \frac{|T|}{N}$$

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

$$\|\hat{f} + h\|_{\ell_1} \; < \; \|\hat{f}\|_{\ell_1}$$

- f is the solution if

$$F_T^* h \neq 0$$

for all such descent vectors (implied by UP)

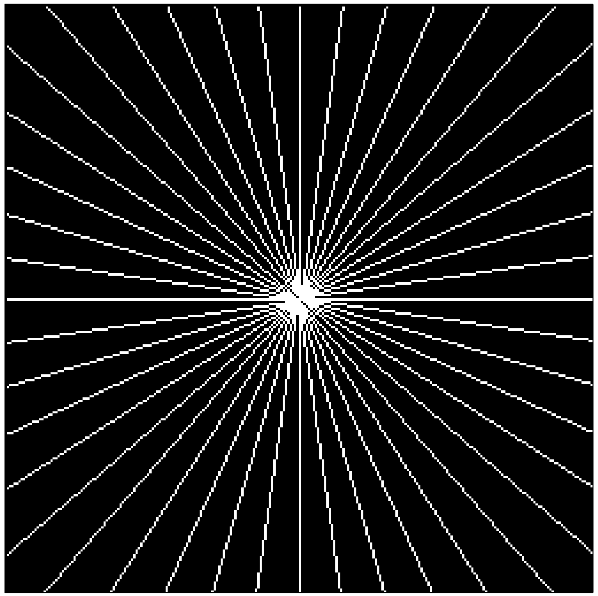
Extensions

Extension: Total-variation reconstruction

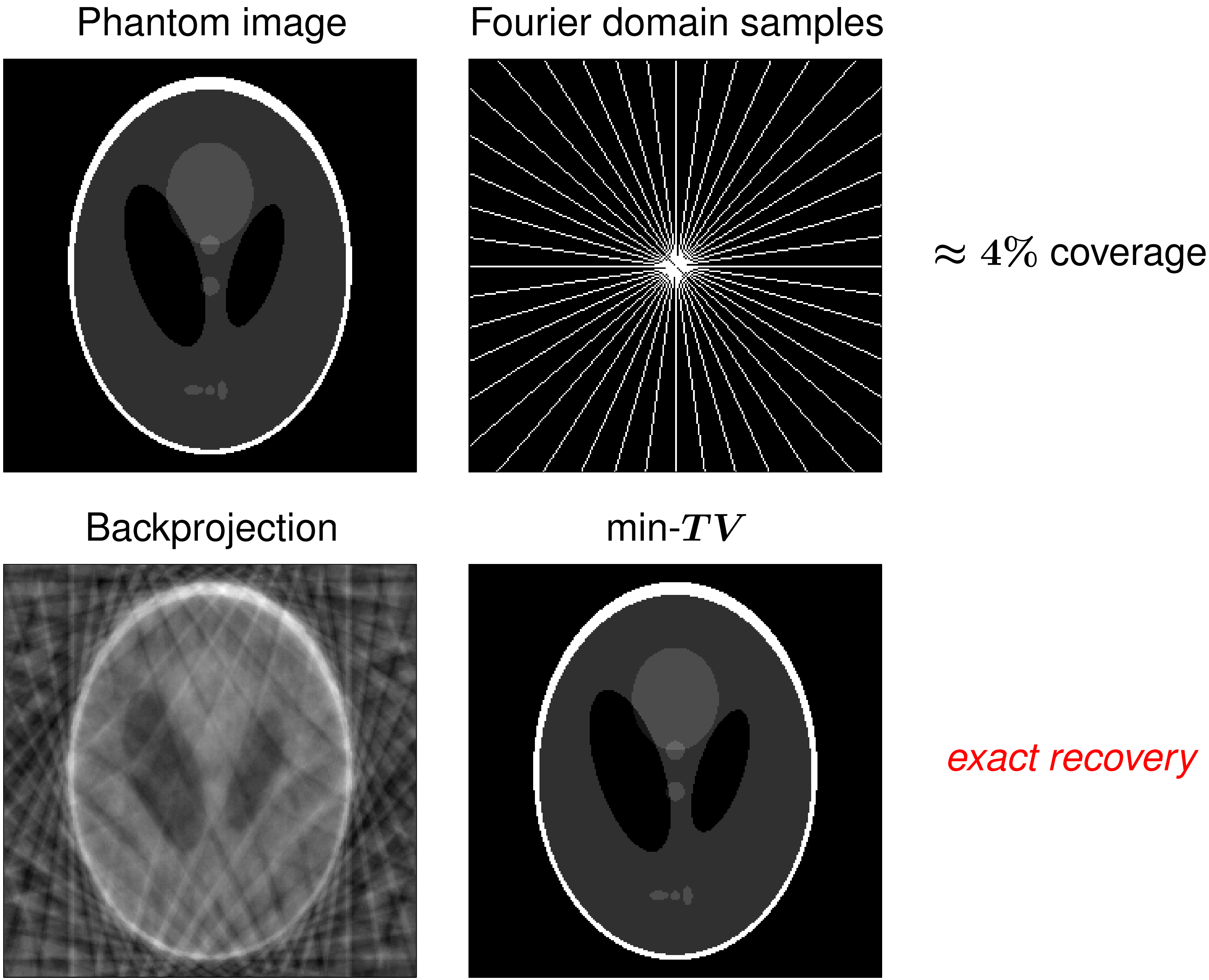
- Tomographic imaging: sample the 2D Fourier transform
- Alternate notion of sparsity: images with sparse gradients
- Given samples $y_k = \hat{f}(\omega_k),\ \omega_k \in \Omega$, solve

$$\min_g \|g\|_{TV} \quad \text{subject to} \quad \hat{g}(\omega_k) = y_k \ \ \omega_k \in \Omega$$

$$\|g\|_{TV} = \sum_{t_1,t_2} \sqrt{(g_{t_1+1,t_2} - g_{t_1,t_2})^2 + (g_{t_1,t_2+1} - g_{t_1,t_2})^2} \approx \|\nabla g\|_{\ell_1}$$



Ω is $\approx 4\%$ of Fourier coefficients



Phantom image
Fourier domain samples
≈ 4% coverage
Backprojection
min-TV
exact recovery

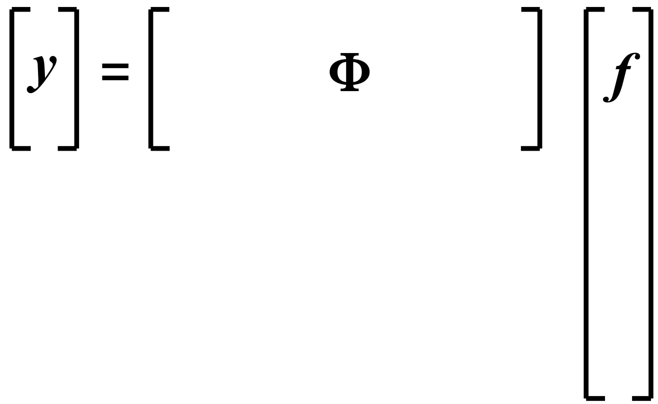
Recovery from Incomplete Measurements

- Unknown discrete signal $x_0 \in \mathbb{R}^N$
- Observe K linear measurements

$$y_k = \langle x_0, \phi_k \rangle, \quad k = 1, \dots, K \quad \text{or} \quad y = \Phi x_0$$

ϕ_k = "test functions"

- Far fewer measurements than degrees of freedom: $K \ll N$



- Examples of ϕ_k:
  - Delta functions, y_k are *samples* of $f(t)$
  - Complex sinusoids, y_k are *Fourier coefficients* of $f(t)$
  - Line "integrals", chirps, projections, . . .

General Measurement/Sparsity pairs

- f is sparse in a known orthogonal system Ψ:
  the Ψ-transform is supported on a set of size B,

$$\alpha = \Psi^T f, \quad \#\{\omega : \alpha(\omega) \neq 0\} = B$$

- Linear measurements using “test functions” $\phi_k(t)$

$$y_k = \langle f, \phi_k \rangle \qquad \text{or} \qquad y = \Phi f$$

  Measurement matrix Φ is formed by stacking rows ϕ_k^T

- To recover, solve

$$\min_{\beta} \ \|\beta\|_{\ell_1} \quad \text{such that} \quad \Phi\Psi\beta = y$$

- Exact reconstruction when (Candès and R, ’06)

$$K \geq \text{Const} \cdot \mu^2 \cdot B \cdot \log N$$

- μ is the *coherence* (similarity) between the Φ and Ψ systems
- Results tied to *generalized uncertainty principles* between “sparsity domain” Ψ and “measurement domain Φ

Generalized measurements and sparsity

- f is sparse in a known orthogonal system Ψ:
  the Ψ-transform is supported on a set of size B,

$$\alpha = \Psi^T f, \quad \#\{\omega : \alpha(\omega) \neq 0\} = B$$

- Linear measurements using "test functions" $\phi_k(t)$

$$y_k = \langle f, \phi_k \rangle, \quad \text{or} \quad y = \Phi f, \quad \Phi : K \times N$$

  Measurement matrix Φ is formed by stacking rows ϕ_k^T

- To recover, solve

$$\min_{\beta} \ \|\beta\|_{\ell_1} \quad \text{such that} \quad \Phi\Psi\beta = y$$

- Exact recovery if basis Ψ and measurement system Φ obey
  an uncertainty principle (are *incoherent*)

“Random” measurements

- Gaussian random matrix ($K \times N$):

$$\Phi_{k,n} \sim \mathrm{Normal}(0, 1)$$

- Measure $y = \Phi f$

- **Theorem** (Candès and Tao, ’04; Donoho, ’04):
  If $f = \Psi\alpha$ is B-sparse in a known orthobasis Ψ, solving

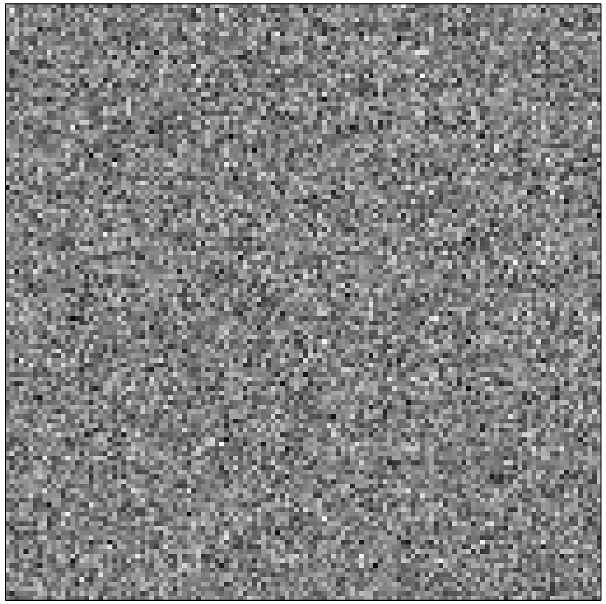
$$\min_{\beta} \|\beta\|_{\ell_1} \quad \text{subject to} \quad \Phi\Psi\beta = y$$

  recover f *exactly* when

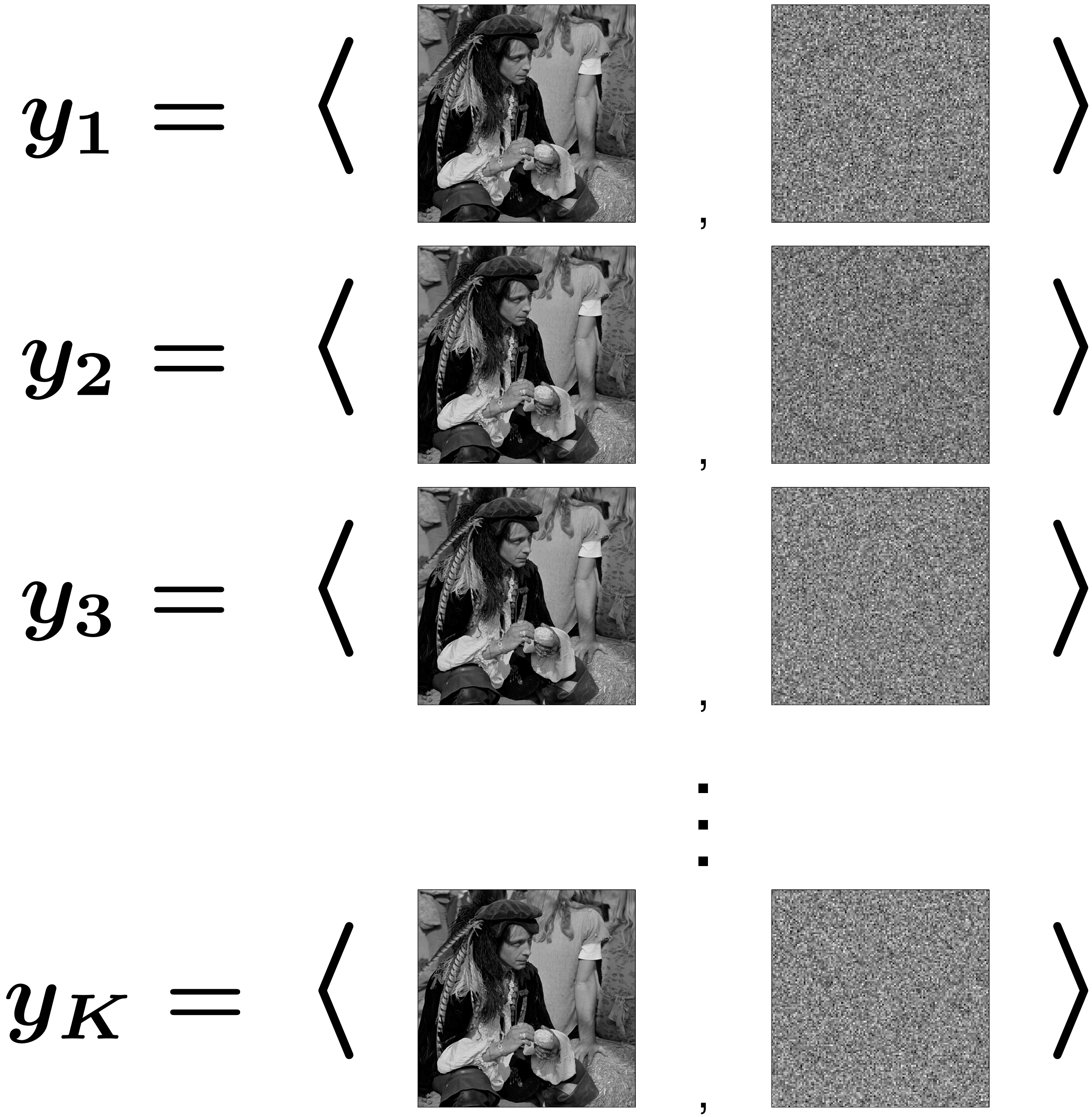
$$K \geq \mathrm{Const} \cdot B \cdot \log N.$$

- Once chosen, the same Φ can be used to recover all sparse f

- Finding incoherent measurement matrices is easy!

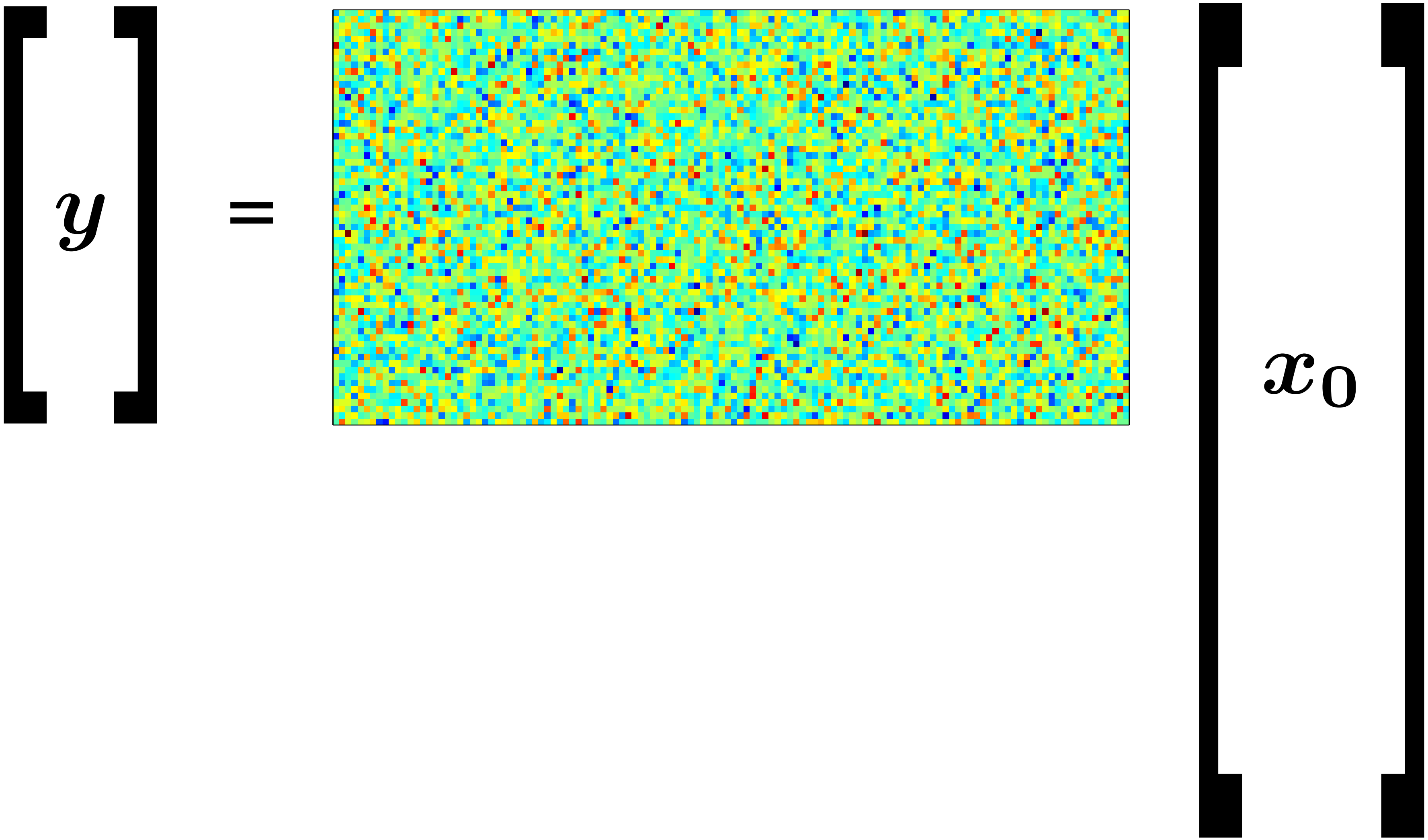
$y_k = \Big\langle$  ,  $\Big\rangle$

- Each measurement touches every part of the underlying signal/image



$y_1 =$
$y_2 =$
$y_3 =$
$y_K =$

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} & \mathbf{\Phi} & \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{x_0} \\ \\ \end{bmatrix}$$


y
=
x_0

Example: Sparse image

- Take $K = 96000$ incoherent measurements $y = \Phi f_a$
- f_a = wavelet approximation (perfectly sparse)
- Solve

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

Ψ = wavelet transform

original (25k wavelets)

perfect recovery

Stability

Stable recovery

- What happens if the measurements are not accurate?

$$y = \Phi f + e, \quad \text{with} \quad \|e\|_2 \leq \epsilon$$

$f = B$-sparse vector, $\quad e =$ perturbation

- Recover: ℓ_1 minimization with relaxed constraints

$$\min \ \|f\|_{\ell_1} \quad \text{such that} \quad \|\Phi f - y\|_2 \leq \epsilon$$

- **Stability Theorem:** (Candès, R, Tao '05)
  If the measurements Φ are incoherent, then

$$\|f^\sharp - f\|_2 \leq \text{Const} \cdot \epsilon$$

- *Recovery error is on the same order as the observation error*

Stable recovery

- What happens if f is not exactly sparse?
- Suppose f is *compressible*: for $s \geq 1$

$$|f|_{(n)} \leq C \cdot n^{-s}$$

$|f|_{(n)}$ = coefficients sorted by magnitude

- Nonlinear approximation error

$$\|f - f_B\|_2 \;\leq\; \mathrm{Const} \cdot B^{-s+1/2}$$

f_B = approximation using B largest components

- ℓ_1 recovery (from incoherent measurements) $f^\sharp$ obeys

$$\|f^\sharp - f\|_2 \quad \leq \quad \mathrm{Const} \cdot \left(\epsilon \;+\; B^{-s+1/2} \right)$$

$$\mathrm{Const} \cdot (\; \textit{measurement error} \;+\; \textit{approximation error}\;)$$

Recovery via Convex Optimization

ℓ_1 minimization

- $\boldsymbol{\ell_1}$ with equality constraints ("Basis Pursuit") can be recast as a linear program (Chen, Donoho, Saunders '95)

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

$$\Updownarrow$$

$$\min_{u,g} \ \sum_t u(t) \quad \text{subject to} \quad \begin{aligned} -u \ \leq \ g \ \leq \ u \\ \Phi g = \ y \end{aligned}$$

Total-Variation Minimization

- The Total Variation functional is a “sum of norms”

$$\begin{aligned} \mathbf{TV}(g) &= \sum_{i,j} \sqrt{(g_{i+1,j} - g_{i,j})^2 + (g_{i,j+1} - g_{i,j})^2} \\ &= \sum_{i,j} \|D_{i,j} g\|_2 \qquad D_{i,j} g = \begin{bmatrix} g_{i+1,j} - g_{i,j} \\ g_{i,j+1} - g_{i,j} \end{bmatrix} \end{aligned}$$

- Total variation minimization can be written as a second-order cone program (Boyd et. al, 1997, and others)

$$\min_g \mathbf{TV}(g) := \sum_{i,j} \|D_{i,j} g\|_2 \quad \text{subject to} \quad \|\Phi g - y\|_2 \leq \epsilon$$

$$\Updownarrow$$

$$\min_{u,g} \sum_{i,j} u_{i,j} \quad \text{subject to} \quad \begin{aligned} \|D_{i,j} g\|_2 &\leq u_{i,j}, \quad \forall i, j \\ \|\Phi g - y\|_2 &\leq \epsilon \end{aligned}$$

Primal-Dual Algorithms for LP

- Standard LP:

$$\min_{x} \langle c, x \rangle \quad \text{subject to} \quad Ax = b, \quad x \leq 0$$

- Karush-Kuhn-Tucker (KKT) conditions for optimality:
  Find $x^\star, \lambda^\star, \nu^\star$ such that

$$\begin{array}{ccc} Ax^\star = b & c + A^*\nu^\star + \lambda^\star = 0 & x_i^\star \lambda_i^\star = 0, \quad \forall i \\ x^\star \leq 0 & \lambda^\star \geq 0 & \end{array}$$

- Primal-dual algorithm:
  - Relax: use $x_i \lambda_i = 1/\tau$, increasing τ at each iteration
  - Linearize system

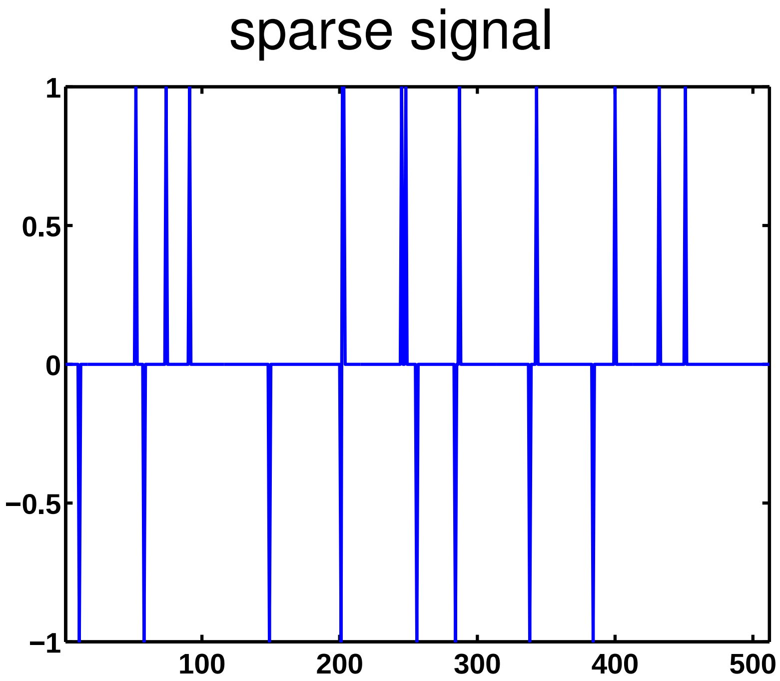
$$\begin{pmatrix} Ax - b \\ c + A^*\nu + \lambda \\ x_i \lambda_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1/\tau \end{pmatrix}$$

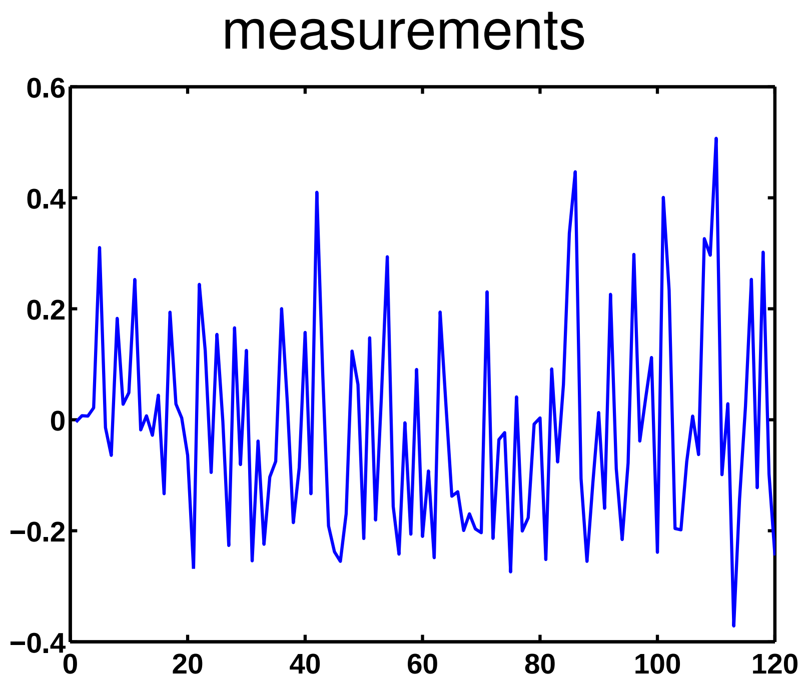
  - Solve for step direction, adjust length to stay in interior ($x \leq 0, \ \lambda \geq 0$)

Newton Iterations

- Newton: solve $f(x) = 0$ iteratively by solving a series of linear problems
  - At x_k,
  $$f(x_k + \Delta x_k) \approx f(x_k) + \Delta x_k f'(x_k)$$
  - Solve for Δx_k such that $f(x_k) + \Delta x_k f'(x_k) = 0$
  - Set $x_{k+1} = x_k + \Delta x_k$
  - Repeat
- Each Newton iteration requires solving a linear system of equations
- Bottleneck of the entire procedure:
  We need to solve a series of $K \times K$ systems (K = # constraints)
- Each step is expensive, but we do not need many steps
  - Theory: need $O(\sqrt{N})$ steps
  - Practice: need $\approx$ 10–40 steps

Example





- $N = 512$, $K = 120$
- Recover using ℓ_1 minimization with equality constraints
- Requires 12 iterations to get within 10^{-4} (4 digits)
- Takes about 0.33 seconds on high-end desktop Mac (Matlab code)

Large-Scale Systems of Equations

- The system we need to solve looks like

$$\boldsymbol{A\Sigma A^* \Delta x = w}$$

$\boldsymbol{A : K \times N}$
$\boldsymbol{\Sigma : N \times N}$ diagonal matrix; changes at each iteration

- Computation: $\boldsymbol{O(NK^2)}$ to construct, $\boldsymbol{O(K^3)}$ to solve
- Large scale: we must use implicit algorithms (e.g. Conjugate Gradients)
  - iterative
  - requires an application of $\boldsymbol{A}$ and $\boldsymbol{A^*}$ at each iteration
  - number of iterations depends on condition number
- $\boldsymbol{A = \Phi\Psi^*}$
  $\boldsymbol{\Phi = K \times N}$ measurement matrix
  $\boldsymbol{\Psi = N \times N}$ sparsity basis
- For large-scale Compressive Sampling to be feasible,
  we must be able to apply $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ (and $\boldsymbol{\Phi^*}$, $\boldsymbol{\Psi^*}$) quickly
  ($\boldsymbol{O(N)}$ or $\boldsymbol{O(N \log N)}$)

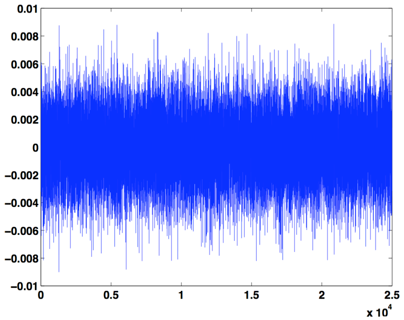
Fast Measurements

- Say we want to take $\mathbf{20,000}$ measurements of a $\mathbf{512 \times 512}$ image ($\boldsymbol{N = 262,144}$)
- If $\boldsymbol{\Phi}$ is Gaussian, with each entry a float, it would take more than an entire DVD just to hold $\boldsymbol{\Phi}$
- Need fast, implicit, noise-like measurement systems to make recovery feasible
- Partial Fourier ensemble is $\boldsymbol{O(N \log N)}$ (FFT and subsample)
- Tomography: many fast unequispaced Fourier transforms, Dutt and Rohklin, Pseudopolar FFT of Averbuch et. al
- Noiselet system of Coifman and Meyer
  - perfectly incoherent with Haar system
  - performs the same as Gaussian (in numerical experiments) for recovering spikes and sparse Haar signals
  - $\boldsymbol{O(N)}$

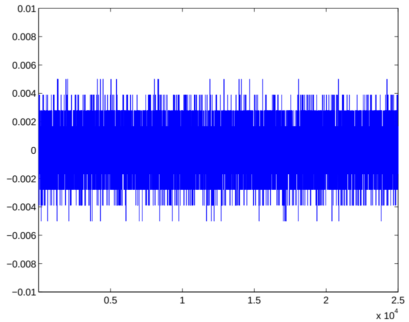
Large Scale Example

image

measure



quantize



recover

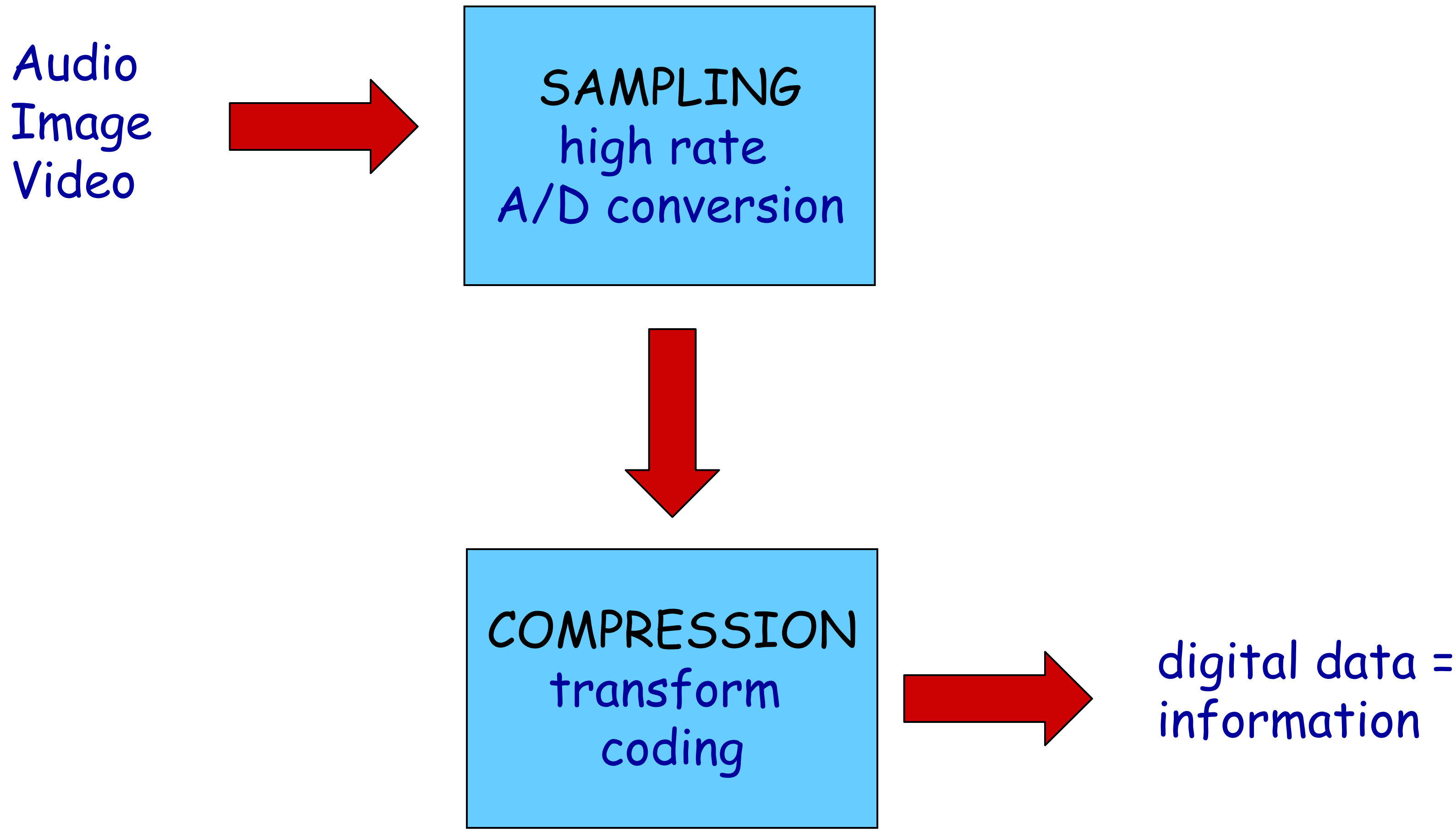
- $N = 256 \times 256$, $K = 25000$
- Measure using “scrambled Fourier ensemble” (randomly permute the columns of the FFT)
- Recover using $\mathbf{TV}$-minimization with relaxed constraints
- Recovery takes ≈ 5 minutes on high-end desktop Mac
- Vanilla log barrier SOCP solver (in Matlab)
- Note: Noise and approximate sparsity help us here

Compressive Sampling in Noisy Environments

ICIP Tutorial, 2006

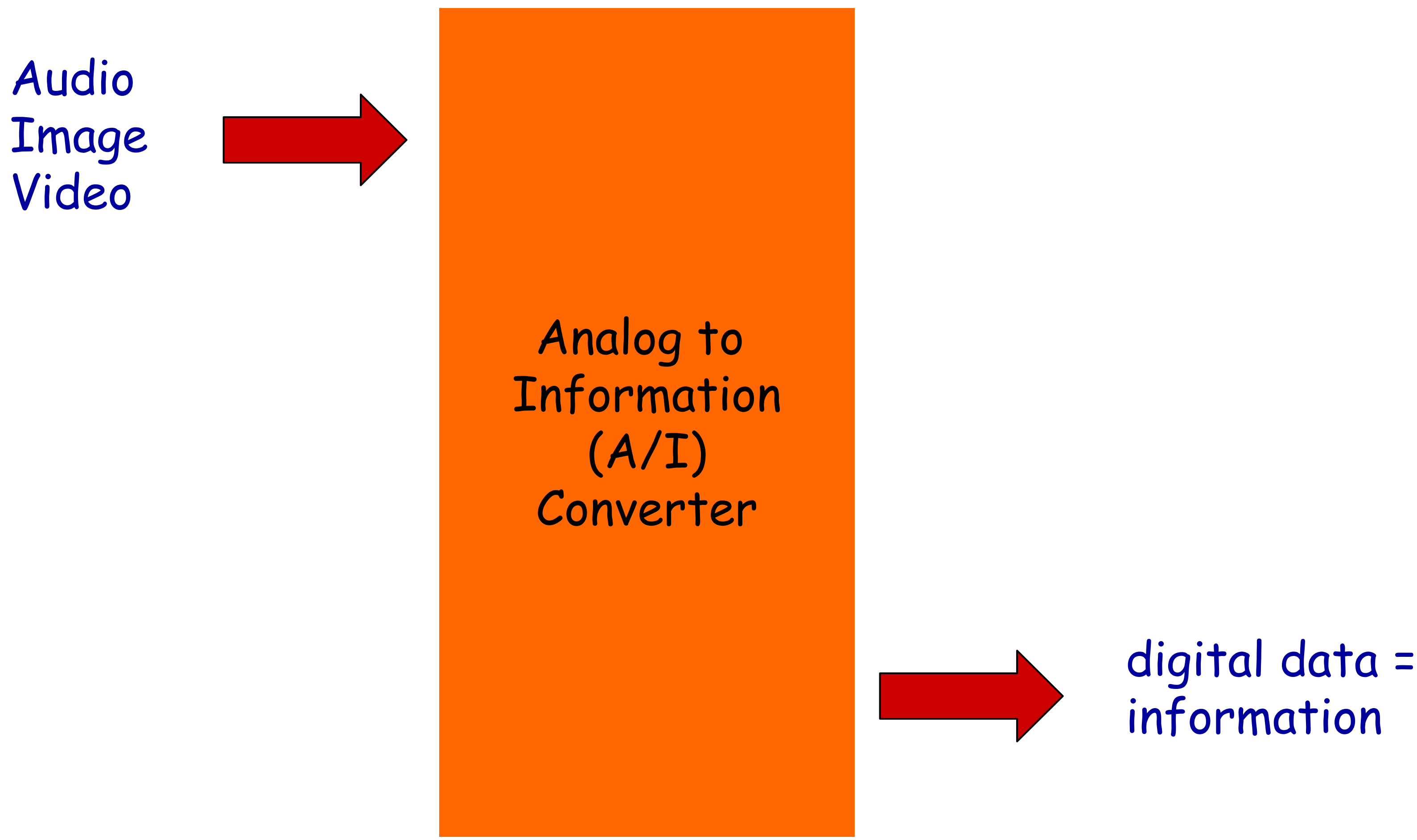
Robert Nowak, www.ece.wisc.edu/~nowak

Rethinking Sampling and Compression



Convention:
Oversample and then *remove redundancy* to extract information

Rethinking Sampling and Compression



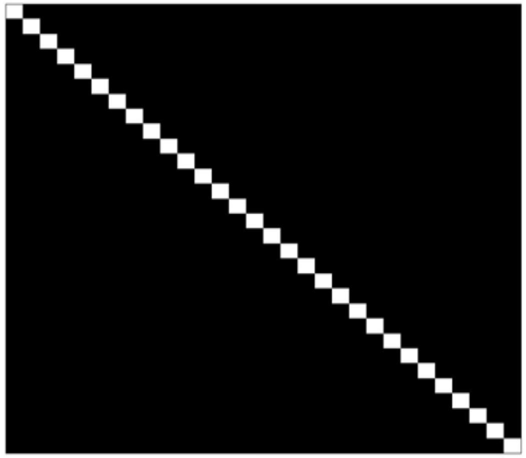
Compressive Sampling:

"*Smart*" sampling to acquire only the *important* information

Traditional vs. Compressive Sampling

Traditional Sampling:

$$\begin{bmatrix} Y_1 \\ \vdots \\ \\ \vdots \\ Y_n \end{bmatrix} =$$



$$\begin{bmatrix} f_1^* \\ f_2^* \\ \vdots \\ \vdots \\ f_n^* \end{bmatrix}$$

Sample uniformly at a very high rate … collect many, many samples

Compressive Sampling:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_k \end{bmatrix} =$$

$$\begin{bmatrix} f_1^* \\ f_2^* \\ \vdots \\ \vdots \\ f_n^* \end{bmatrix}$$

Sample randomly, using a "white noise" pattern, and take very few samples

Related Work

Peng & Bresler '96, Gastpar & Bresler '00:

Reconstruction of signals with sparse Fourier spectra from non-uniform time samples (including noisy cases)

Gilbert, Guha, Indyk, Muthukrishnan, Strauss '02,
Zou, Gilbert, Strauss, & Daubechies '05:

Reconstruction of signals with sparse Fourier spectra from non-uniform time samples (computational complexity)

Vetterli, Marziliano and Blu '02, Maravic & Vetterli '05

Reconstruction of signals with finite degrees of freedom using unconventional samples (e.g., sparse time signals)

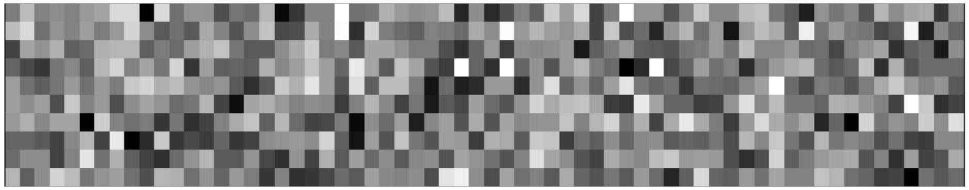
Baraniuk, Devore, Lustig, Romberg, Tropp, Wakin, others '06:

Extensions/applications of compressive sampling

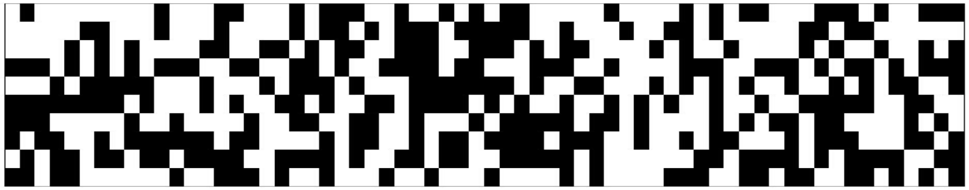
CS Websites: www.dsp.ece.rice.edu/cs/
www.acm.caltech.edu/l1magic/

Compressed Sensing Matrices

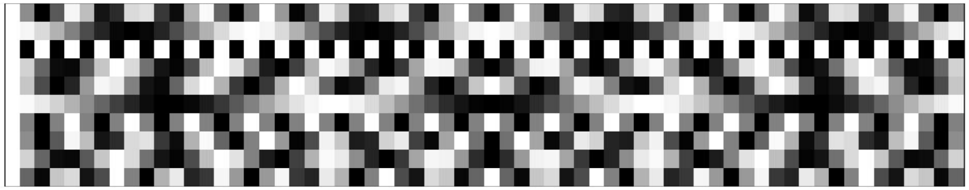
Gaussian Ensemble (iid GWN entries)



Rademacher Ensemble (iid +1, -1 entries)

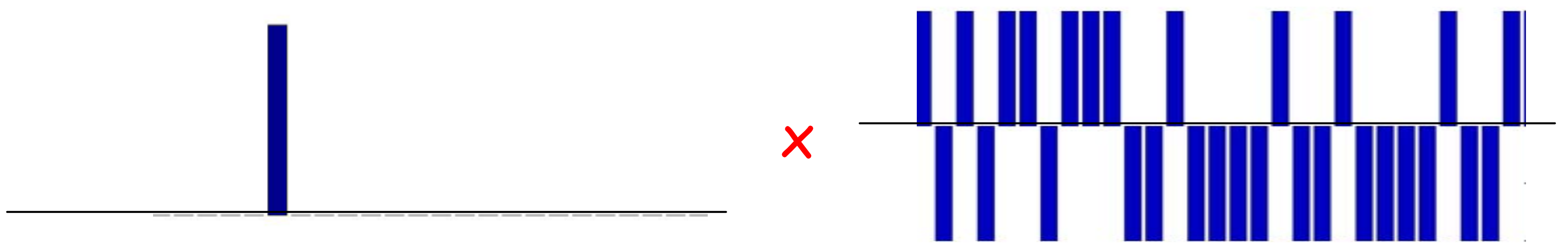


Fourier Ensemble (rows are randomly chosen DFT vectors)



A Simple Illustrative Example

Suppose that $f^* \in R^n$ has a single non-zero entry that is strictly greater than zero. How many random projections are required to perfectly reconstruct f^*?



sparse signal f*   random vector ϕ

If $\phi' f^* > 0$, then non-zero element is located at one of the $+1$ locations in ϕ, otherwise it must be at one of the -1 location. Repeat $O(\log n)$ times.

Why Does CS Works in General ?

Signal Transform

$$\begin{bmatrix} T_{1,1} & \cdots & T_{1,n} \\ T_{2,1} & \cdots & T_{2,n} \\ \vdots & \cdots & \vdots \\ \vdots & \cdots & \vdots \\ T_{n,1} & \cdots & T_{n,n} \end{bmatrix} \begin{bmatrix} f_1^* \\ f_2^* \\ \vdots \\ \vdots \\ f_n^* \end{bmatrix} = \begin{bmatrix} 0 \\ \theta_1 \\ 0 \\ \vdots \\ 0 \\ \theta_m \\ 0 \end{bmatrix}$$

Sparse vector with m << n non-zero coefficients

Transform × Signal

signal f*

transform by T

sparse coefficients θ

Why Does CS Works in General ?

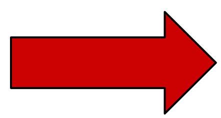
Inverse Transform

$$\begin{bmatrix} T_{1,1} & \cdots & T_{1,n} \\ T_{2,1} & \cdots & T_{2,n} \\ \vdots & \cdots & \vdots \\ \vdots & \cdots & \vdots \\ T_{n,1} & \cdots & T_{n,n} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \theta_1 \\ 0 \\ \vdots \\ 0 \\ \theta_m \\ 0 \end{bmatrix} = \begin{bmatrix} f_1^* \\ f_2^* \\ \vdots \\ \vdots \\ f_n^* \end{bmatrix}$$

Signal

Inverse Transform x coefficients

sparse coefficents θ



inverse transform by T^{-1}

signal f*

Why Does CS Works in General ?

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_k \end{bmatrix} =$$

$$\begin{bmatrix} T_{1,1} & \cdots & T_{1,n} \\ T_{2,1} & \cdots & T_{2,n} \\ \vdots & \cdots & \vdots \\ \vdots & \cdots & \vdots \\ T_{n,1} & \cdots & T_{n,n} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \theta_1 \\ 0 \\ \vdots \\ 0 \\ \theta_m \\ 0 \end{bmatrix}$$

random sampling matrix (k x n)

signal representation in terms of m << n coefficients

- The samples $Y_1, \ldots, Y_k$ give us k equations
- We have 2m unknowns (m coefficient values and m coefficient locations)
- If $k \geq 2m$, then we may have a unique solution

KEY: **Randomness** of sampling guarantees that we have a linearly independent system of equations and hence a unique solution

Example (courtesy of J. Romberg, Georgia Tech.)

Original 1 Megapixel image (approx. 25K non-zero wavelet coefficients)

Exact reconstruction from 96K random projection samples

We require less than four times as many samples as the number of significant wavelet coefficients !

Conventional Undersampling

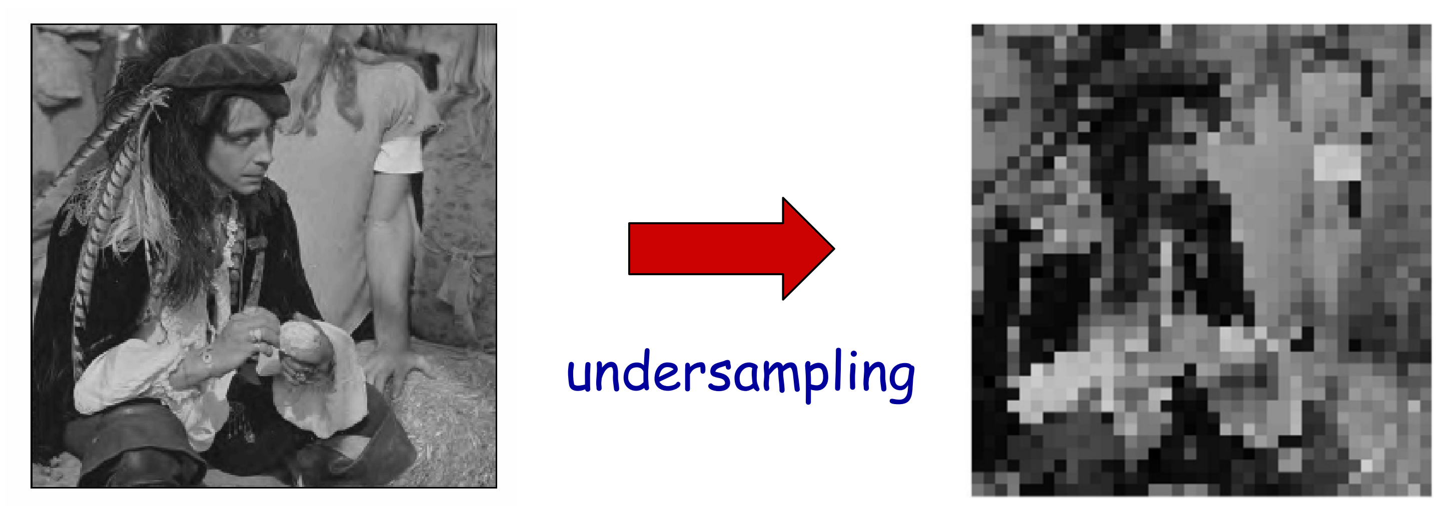
$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_k \end{bmatrix} =$$

$$\begin{bmatrix} f_1^* \\ f_2^* \\ \vdots \\ f_m^* \end{bmatrix}$$

conventional "undersampling"

signal

No matter how we decide to subsample, we will probably ***miss*** important details of the signal



Reconstruction from Compressive Samples

Compressive Samples:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_k \end{bmatrix} =$$

$$\begin{bmatrix} T_{1,1} & \cdots & T_{1,n} \\ T_{2,1} & \cdots & T_{2,n} \\ \vdots & \cdots & \vdots \\ \vdots & \cdots & \vdots \\ T_{n,1} & \cdots & T_{n,n} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \theta_1 \\ 0 \\ \vdots \\ 0 \\ \theta_m \\ 0 \end{bmatrix}$$

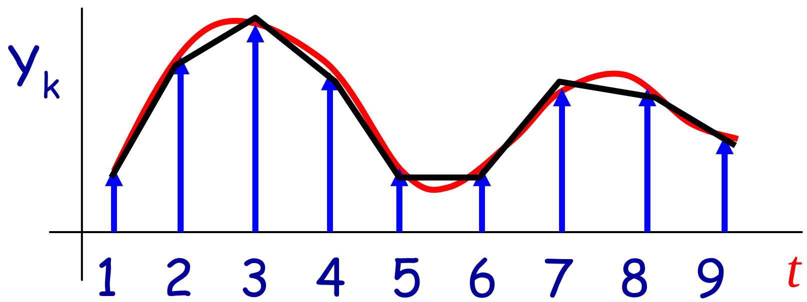
Need to solve this system of equations for the coefficients. This is very challenging since we must determine the values and the "*locations*" of the m significant coefficients !

Requires *highly nonlinear optimization* procedures in order to find the compressible signal that best fits samples

One approach is brute-force (i.e., try out all possible locations).

Another approach (*that works !)* is based on convex optimization.

Reconstruction from Samples



Reconstruction in traditional sampling is a simple linear interpolation problem (i.e., connect the dots)

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_k \end{bmatrix} =$$

$$\begin{bmatrix} T_{1,1} & \cdots & T_{1,n} \\ T_{2,1} & \cdots & T_{2,n} \\ \vdots & \cdots & \vdots \\ \vdots & \cdots & \vdots \\ T_{n,1} & \cdots & T_{n,n} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \theta_1 \\ 0 \\ \vdots \\ 0 \\ \theta_m \\ 0 \end{bmatrix}$$

Reconstruction from compressive samples requires the solution of a nonlinear, convex optimization problem

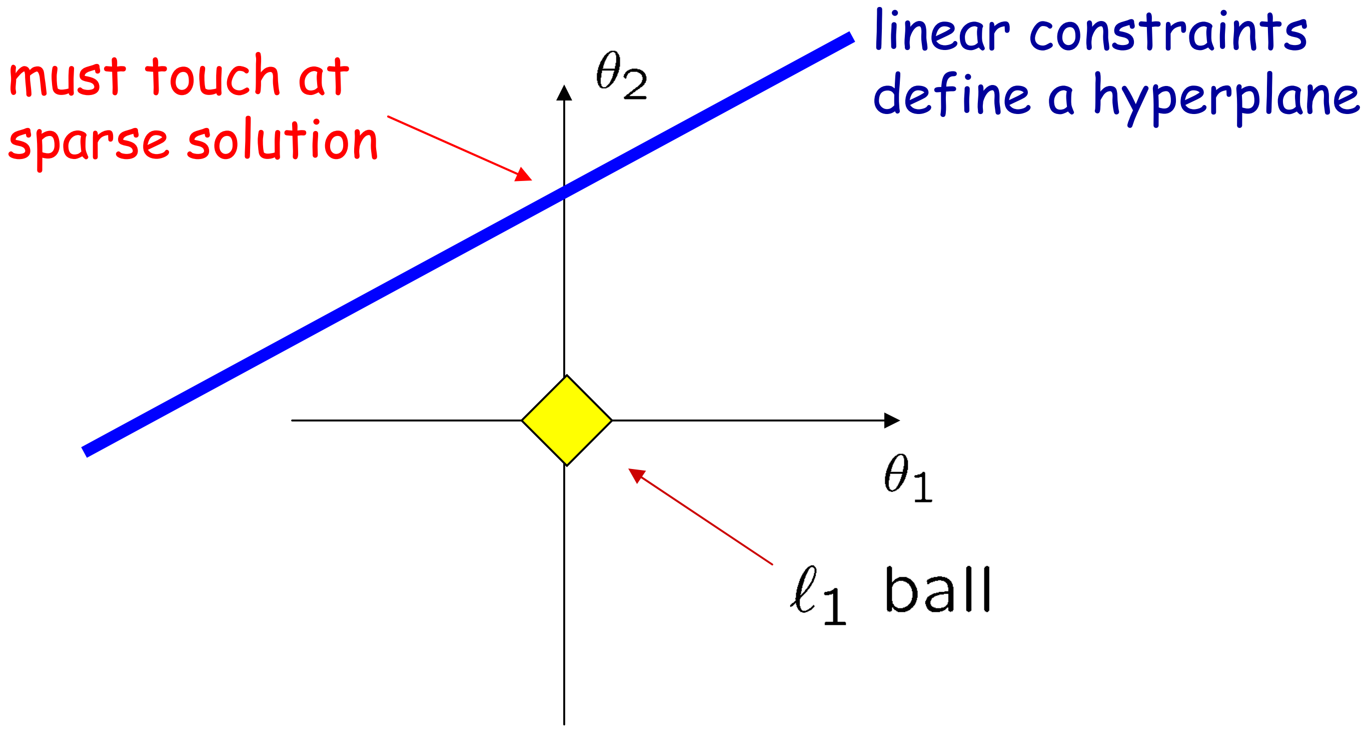
$$\min \sum_{i=1}^{n} |\theta_i| \quad \text{subject to} \quad \Phi T^{-1} \theta = Y$$

Sparsity of L1 Solution

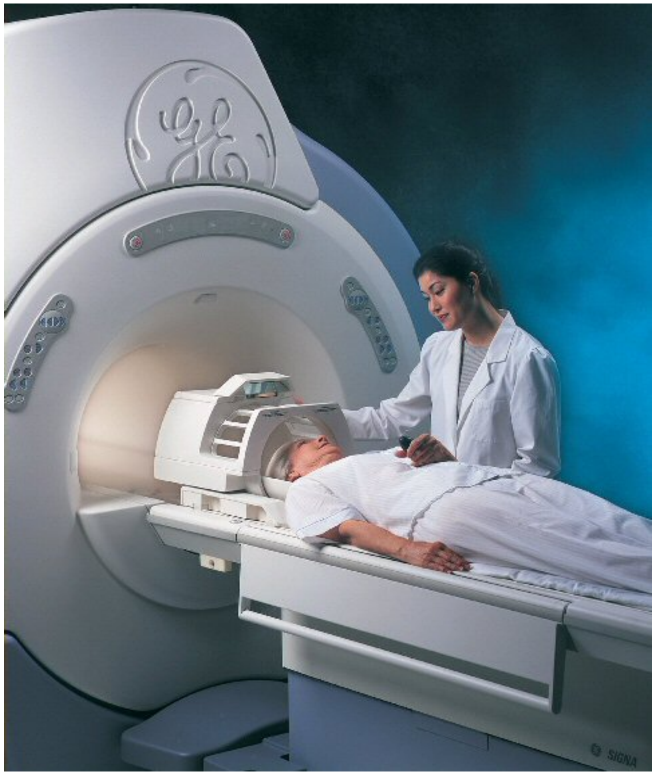
$$\min \sum_{i=1}^{n} |\theta_i| \quad \text{subject to} \quad \Phi T^{-1}\theta = Y$$

L1 norm of coefficients

constraints

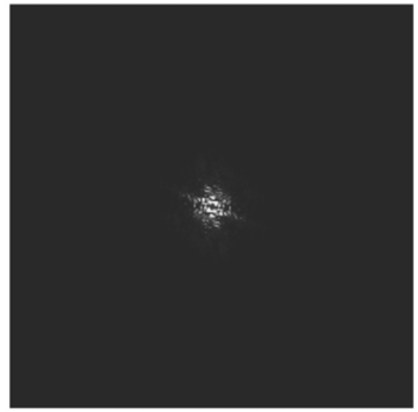


Ex. Magnetic Resonance Imaging



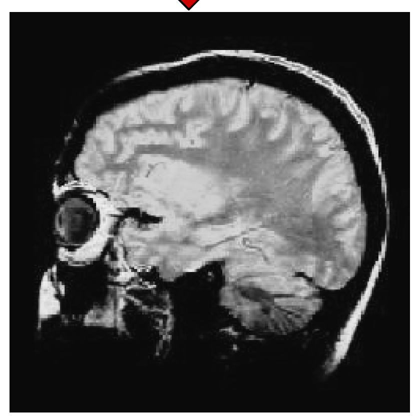
MRI "scans" the patient by collecting coefficients in the frequency domain

SCAN



coefficients in frequency (Fourier transform of brain; very sparse)

RECONSTRUCTION



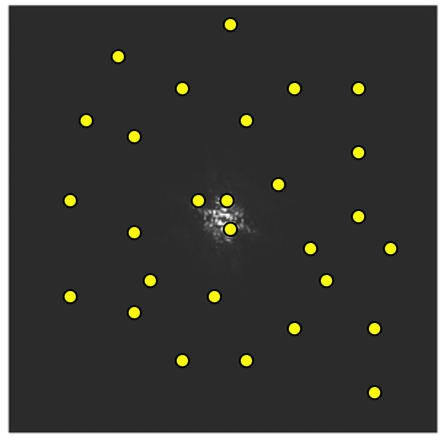
Inverse Fourier transform produces brain image

Sampling in the frequency domain is a form of compressive sampling !

Compressive Sampling in MRI

MRI angiography in lower leg (high-res sampling

"k-space data"

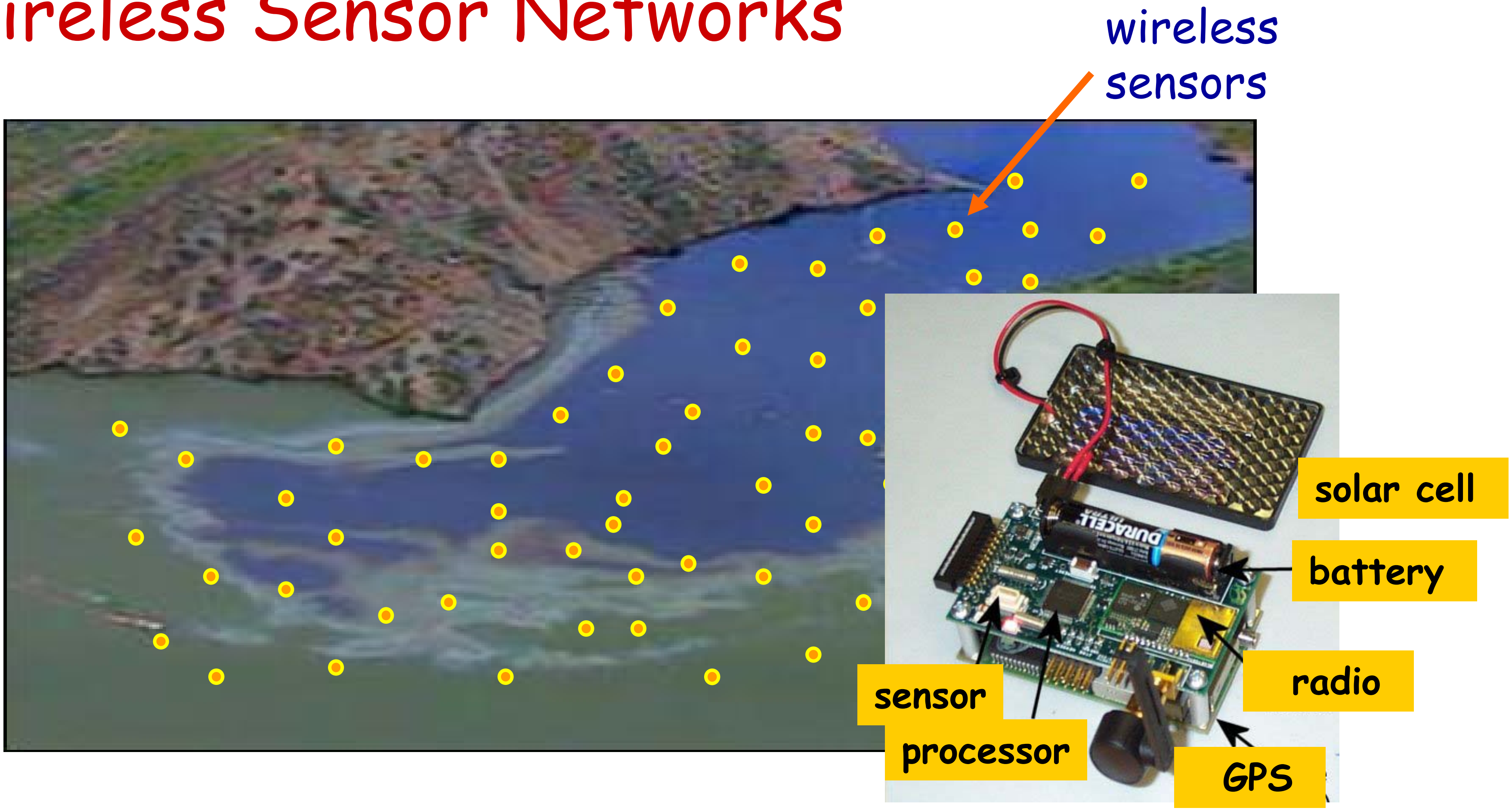


undersubsampling in FFT domain (undersampled by a factor of 10)

Reconstruction from undersampled data
Standard (left)
CS (right)

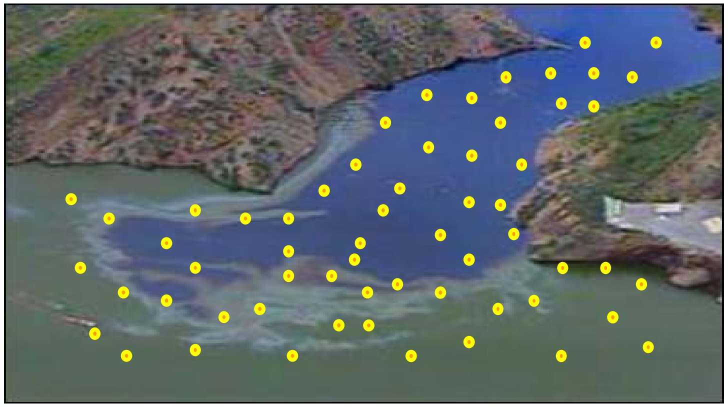
M. Lustig, D.L. Donoho, J.M Pauly (Stanford), '06

Surveillance and Monitoring with Wireless Sensor Networks

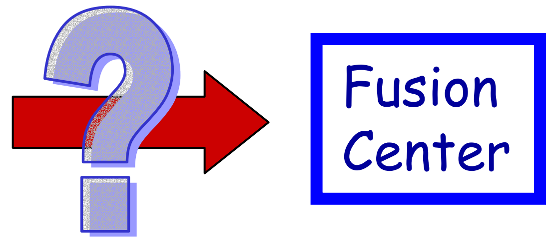


Goal: Reconstruct an accurate map of contamination while keeping number of samples/communications to a bare minimum

Compressive Sampling in Sensor Nets



compressible sensor data



Two popular approaches:

- Exhaustive sampling and point-to-point comm: WASTEFUL
- In-network processing and collaborative comm: COMPLICATED

A New Alternative:

Compressively sample sensor nodes to minimize number of queries and transmissions while maintaining high-fidelity

Haupt, Bajwa, Sayeed, Rabbat, & Nowak '05 (Madison)

Compressive Sampling

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_k \end{bmatrix} = \Phi \begin{bmatrix} f_1^* \\ f_2^* \\ \vdots \\ \vdots \\ f_n^* \end{bmatrix} + \begin{bmatrix} W_1 \\ \vdots \\ W_k \end{bmatrix}$$

random projection matrix
k << n

signal

noise

Key Questions:

1. When and why does CS work ?
2. How does performance degrade with noise ?
3. What, if anything, do we lose with CS ?
4. What basis should be used for reconstruction ?

Theory of Compressive Sampling

Assume that f^* is compressible; i.e., if f^* is compressed into m bytes, then

$$\text{compression error} \;=\; Const. \times m^{-\alpha} \;, \quad \alpha > 1$$

α measures the degree of compressibility

CS Reconstruction Theorem:

If we take k random projection samples, then we can construct a reconstruction $\widehat{f}$ satisfying:

noise-free samples

$$E\left[\frac{\|f^* - \widehat{f}\|^2}{n}\right] \;=\; O\left(k^{-\alpha}\right)$$

Candes, Romberg, Tao '04
Donoho '04

noisy samples

$$E\left[\frac{\|f^* - \widehat{f}\|^2}{n}\right] \;=\; O\left(k^{-\alpha/(\alpha+1)}\right)$$

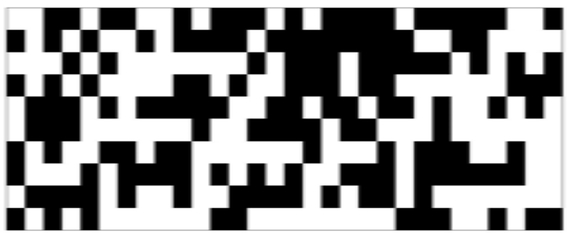
Haupt & RN '05,
E. Candes & T. Tao '05

Noiseless Compressive Sampling

Let $\Phi = \{\phi_{i,j}\}$ be an $n \times k$ array of iid zero-mean rvs

Ex. Rademacher sampling

$\Phi' =$



$$= \underbrace{\begin{pmatrix} \pm\frac{1}{\sqrt{n}} & \cdots & \pm\frac{1}{\sqrt{n}} \\ \vdots & & \vdots \\ \pm\frac{1}{\sqrt{n}} & \cdots & \pm\frac{1}{\sqrt{n}} \end{pmatrix}}_{n} \left.\vphantom{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}\right\} k$$

k samples:

$$Y = \Phi' f^*$$

Noiseless Compressive Sampling

Noiseless samples:

$$Y = \Phi' f^*$$

Let f denote a candidate reconstruction of f^*.
The reconstruction error is measured by

$$\begin{aligned}\|Y - \Phi' f\|^2_{\ell_2(k)} &= \|\Phi'(f^* - f)\|^2_{\ell_2(k)} \\ &= (f^* - f)'\Phi\Phi'(f^* - f)\end{aligned}$$

$$\begin{aligned}E\,\|Y - \Phi' f\|^2_{\ell_2(k)} &= (f^* - f)'\,E[\Phi\Phi']\,(f^* - f) \\ &= (f^* - f)'\,\frac{k}{n} I_{n\times n}\,(f^* - f) \\ &= \frac{k}{n}\|f^* - f\|^2_{\ell_2(n)}\end{aligned}$$

Risk Functions

Ideal risk function :

$$R(f) \;=\; \frac{1}{n}\|f^* - f\|^2_{\ell_2(n)}$$

Empirical risk function :

$$\widehat{R}(f) \;=\; \frac{1}{k}\|Y - \Phi' f\|^2_{\ell_2(k)}$$

Since $E\,\widehat{R}(f) = R(f)$, minimizing the empirical risk seems to be a sensible thing

Empirical Risk Minimization

Let $f^* \in \mathcal{F}$ and our reconstruction algorithm will be to select the signal in $\mathcal{F}$ that minimizes empirical risk:

$$\widehat{f} \;=\; \arg\min_{f \in \mathcal{F}} \widehat{R}(f)$$

Ideally we would like to solve

$$f^* \;=\; \arg\min_{f \in \mathcal{F}} R(f)$$

$\widehat{f}$ and f^* will *probably* agree if $R \approx \widehat{R}$ with high probability, uniformly over $\mathcal{F}$

i.e., if for all $f \in \mathcal{F}$

$$P\left(\widehat{R}(f) \approx R(f)\right) \approx 1$$

then $P(\widehat{f} \neq f^*) \approx 0$

Concentration

We know that $E\,\widehat{R}(f) = R(f)$, but does $\widehat{R}(f) \approx R(f)$ with high probability for all $f \in \mathcal{F}$?

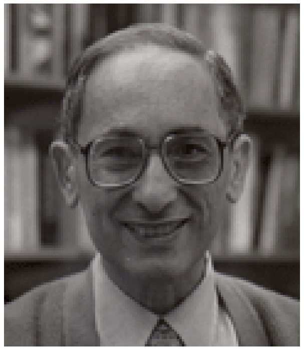
in many interesting cases, YES !

"A random variable that depends (in a smooth way) on the influence of many independent variables (but not too much on any of them) is essentially a constant." – Michel Talagrand

Ex. Chernoff's bound

If $\{X_i\}$ are i.i.d. Bernoulli random variables with $P(X_i = 1) = p$, then $S_n = \sum_{i=1}^{n} X_i$ satisfies

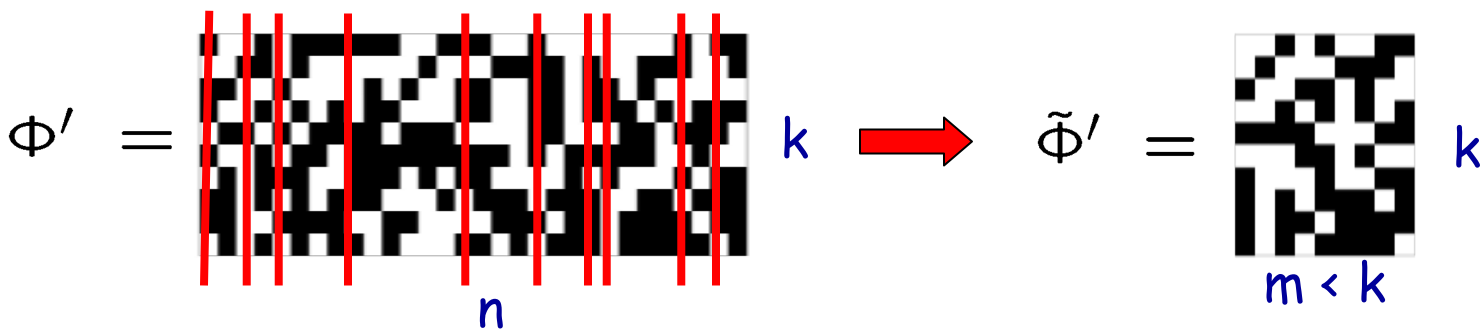
$$P(|S_n/n - p| > \epsilon) \;\leq\; 2e^{-2n\epsilon^2}$$



Empirical mean "concentrates" about true mean values exponentially quickly

A Fundamental Concentration Inequality

Recall that ϕ' is $k \times n$ and $E[\Phi\Phi'] = \frac{k}{n} I_{n\times n}$



Thm: For all $m < \frac{k}{\log n}$ and every $k \times m$ submatrix $\tilde{\Phi}$ of Φ, with very high probability

$$\frac{k}{2n} \leq \lambda_{min}(\tilde{\Phi}\tilde{\Phi}') \leq \lambda_{max}(\tilde{\Phi}\tilde{\Phi}') \leq \frac{3k}{2n}$$

A. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, Smallest singular value of random matrices and geometry of random polytopes, *Adv. Math.* **195** (2005), no. 2, 491--523.

Relevance to Compressed Sensing

Suppose that $f^* - f$ has no more than m non-zero entries

i.e., f^* and f are **sparse**

Ex. n=6, m=2, k=3

$$\Phi'(f - f^*) = \begin{pmatrix} \pm\frac{1}{\sqrt{n}} & \pm\frac{1}{\sqrt{n}} & \pm\frac{1}{\sqrt{n}} & \pm\frac{1}{\sqrt{n}} & \pm\frac{1}{\sqrt{n}} & \pm\frac{1}{\sqrt{n}} \\ \pm\frac{1}{\sqrt{n}} & \pm\frac{1}{\sqrt{n}} & \pm\frac{1}{\sqrt{n}} & \pm\frac{1}{\sqrt{n}} & \pm\frac{1}{\sqrt{n}} & \pm\frac{1}{\sqrt{n}} \\ \pm\frac{1}{\sqrt{n}} & \pm\frac{1}{\sqrt{n}} & \pm\frac{1}{\sqrt{n}} & \pm\frac{1}{\sqrt{n}} & \pm\frac{1}{\sqrt{n}} & \pm\frac{1}{\sqrt{n}} \end{pmatrix} \begin{pmatrix} f_1^* - f_1 \\ 0 \\ 0 \\ f_4^* - f_4 \\ 0 \\ 0 \end{pmatrix}$$

3x6 underdetermined system

$$= \widetilde{\Phi}'(f - f^*) = \begin{pmatrix} \pm\frac{1}{\sqrt{n}} & 0 & 0 & \pm\frac{1}{\sqrt{n}} & 0 & 0 \\ \pm\frac{1}{\sqrt{n}} & 0 & 0 & \pm\frac{1}{\sqrt{n}} & 0 & 0 \\ \pm\frac{1}{\sqrt{n}} & 0 & 0 & \pm\frac{1}{\sqrt{n}} & 0 & 0 \end{pmatrix} \begin{pmatrix} f_1^* - f_1 \\ 0 \\ 0 \\ f_4^* - f_4 \\ 0 \\ 0 \end{pmatrix}$$

effectively a 3x2 submatrix, overdetermined system !

Relevance to Compressed Sensing

$$\begin{aligned}\widehat{R}(f) &= \|\Phi'(f^* - f)\|^2_{\ell_2(k)} \\ &= \|\widetilde{\Phi}'(f^* - f)\|^2_{\ell_2(k)} = (f^* - f)\widetilde{\Phi}\widetilde{\Phi}'(f^* - f)\end{aligned}$$

If $f^* - f$ has no more than $k/\log n$ non-zero entries, then with very high probability

$$\Rightarrow \quad \frac{1}{2n}\|f^* - f\|^2_{\ell_2(n)} \leq \widehat{R}(f) \leq \frac{3}{2n}\|f^* - f\|^2_{\ell_2(n)}$$

i.e., $\Phi\Phi'$ "acts like" the identity matrix

Recall that $R(f) = \frac{1}{n}\|f^* - f\|^2_{\ell_2(n)}$

Main Result - Noiseless Case

Suppose $f^* \in \ell_2(n)$ is sparse, with no more than $k/\log n$ non-zero elements

Let $\mathcal{F} \subset \ell_2(n)$ consist of all n-vectors with no more than $k/\log n$ non-zero elements and define

$$\begin{aligned} \widehat{f} &= \arg\min_{f \in \mathcal{F}} \|Y - \Phi' f\|^2_{\ell_2(k)} \\ f^* &= \arg\min_{f \in \mathcal{F}} \|f^* - f\|^2_{\ell_2(n)} \end{aligned}$$

Then with high probability

$$\widehat{f} = f^*$$

Reconstruction Algorithms

$$\widehat{f} \;=\; \arg\min_{f \in \mathcal{F}} \|Y - \Phi' f\|^2_{\ell_2(k)}$$

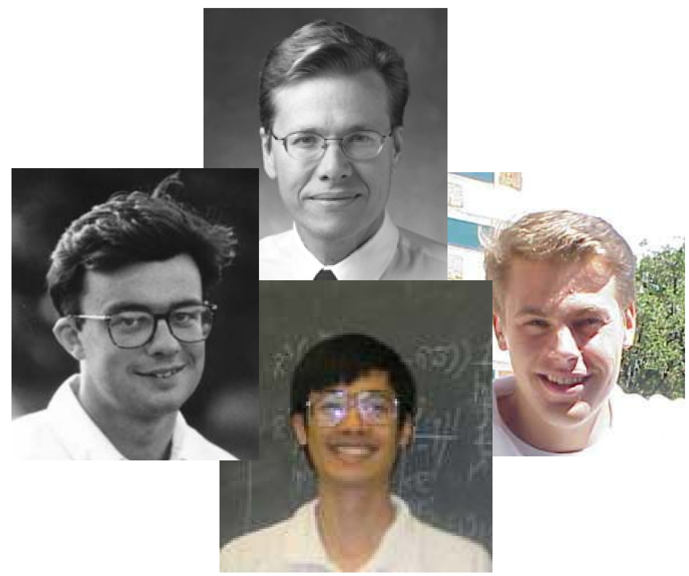
restriction to "sparse" n-vectors leads to combinatorial optimization problem !

Key result of Candes, Romberg & Tau '04, Candes & Tao '04, Donoho '04:

It suffices to solve the linear program

$$\widehat{f} \;=\; \arg\min_{f \in \ell_2(n)} \|f\|_1 \text{ s.t. } Y = \Phi' f$$

This is quite practical, even for large n



Noisy Compressive Sampling

Samples:

$$Y = \Phi' f^* + w$$

where w is a $k \times 1$ vector of Gaussian white noise

Let f denote a candidate reconstruction of f*

$$\begin{aligned} \|Y - \Phi' f\|^2_{\ell_2(k)} &= \|\Phi'(f^* - f) + \Phi' w\|^2_{\ell_2(k)} \\ &= (f^* - f)'\Phi\Phi'(f^* - f) + 2(f^* - f)'\Phi\Phi' w + w'\Phi\Phi' w \end{aligned}$$

Take expectations:

$$\begin{aligned} E\,\|Y - \Phi' f\|^2_{\ell_2(k)} &= (f^* - f)'\,E[\Phi\Phi']\,(f^* - f) + E[w'\Phi\Phi' w] \\ &= \frac{k}{n}\|f^* - f\|^2_{\ell_2(n)} + k\sigma^2 \end{aligned}$$

Risk Functions and Noise

Empirical Risk function:

$$\widehat{R}(f) \;=\; \frac{1}{k}\|Y - \Phi' f\|^2_{\ell_2(k)}$$

Ideal Risk Function:

$$R(f) \;=\; E\,\widehat{R}(f) \;=\; \frac{1}{n}\|f^* - f\|^2_{\ell_2(n)} + \sigma^2$$

Unlike the noiseless case, the risks are non-zero at f^*:

$$\widehat{R}(f^*) = \frac{1}{k}\|\Phi' w\|^2_{\ell_2(k)} \quad \text{and} \quad R(f^*) = \sigma^2$$

Some amount of averaging is required to control errors due to noise; **the risk cannot decay faster than the parametric rate 1/k**

Craig-Bernstein Inequality

Theorem (Craig, 1933): Let $\{U_i\}_{i=1}^k$ be iid random variables satisfying the moment condition

$$E\left[|U_j - E[U_j]|^k\right] \;\leq\; \frac{k!\, var(U_j)\, h^{k-2}}{2}$$

for some $h > 0$ and all integers $k \geq 2$. Then

$$P\left(\frac{1}{k}\sum_{j=1}^{k}(U_j - E\left[U_j\right]) \geq \frac{t}{k\epsilon} + \frac{\epsilon\, k\, var\left(\frac{1}{k}\sum_{j=1}^{k} U_j\right)}{2(1-\zeta)}\right) \;\leq\; e^{-t}$$

Tells us that a sum of iid random variables concentrates about the mean at an exponential rate; if variance is small then

$$\boxed{P\left(\frac{1}{k}\sum_{j=1}^{k}(U_j - E\left[U_j\right]) \geq \tau\right) \;\approx\; e^{-\tau k}}$$

Craig-Bernstein Inequality in CS

The **moment condition** is satisfied for noisy CS, but its verification involves a fair amount of non-trivial analysis

C-B implies that with very high probability and for all f

$$C_1\,(R(f) - R(f^*)) \;\; \leq \;\; \widehat{R}(f) - \widehat{R}(f^*) \; + \; C_2 \frac{\|f\|_0 \log n}{k}$$

where $\|f\|_0 \; = \;$ number of non-zero elements in f

This gives us an upper bound on the ideal risk R(f). Minimizing this upper bound over f is a good reconstruction procedure !

The constants C_1 and C_2 depend on σ^2 and $\|f^*\|^2_{\ell_2(n)}/n$

i.e., the bound depends on the SNR

Reconstruction and Oracle Bound

Select $\widehat{f}$ as the minimizer of the upper bound on the excess risk:

$$\widehat{f} \;=\; \arg\min_{f} \left\{ \widehat{R}(f) \;+\; C_2 \frac{\|f\|_0 \log n}{k} \right\}$$

With very high probability, $\widehat{f}$ approximately minimizes the ideal risk.

Moreover, a reverse inequality gives us the **oracle bound**

$$R(\widehat{f}) \;\leq\; C_3 \min_{f} \left\{ R(f) \;+\; C_1 \frac{\|f\|_0 \log n}{k} \right\}$$

where constant C_3 also depends only on σ^2 and $\frac{1}{n}\|f^*\|^2_{\ell_2(n)}$

We have a provably good reconstruction !

Main Result – Noisy Compressive Sampling

Theorem (J. Haupt and RN '05): Let

$$\widehat{f} \;=\; \arg\min_f \left\{ \|Y - \Phi' f\|^2_{\ell_2(k)} \;+\; C_2 \|f\|_0 \log n \right\}$$

Then

$$\frac{E\,\|f^* - \widehat{f}\|^2_{\ell_2(n)}}{n} \;\leq\; C_3 \min_f \left\{ \frac{\|f^* - f\|^2_{\ell_2(n)}}{n} \;+\; C_1 \frac{\|f\|_0 \log n}{k} \right\}$$

Ex. If f* has m* or fewer non-zero elements, then

$$\frac{E\,\|f^* - \widehat{f}\|^2_{\ell_2(n)}}{n} \;\leq\; C\left(\sigma^2, \frac{1}{n}\|f^*\|^2_{\ell_2}\right) \frac{m^* \log n}{k}$$

within a log-factor of the parametric rate 1/k

Optimization

$$\widehat{f} \;=\; \arg\min_{f \in \mathcal{F}} \left\{ \|Y - \Phi' f\|^2_{\ell_2(k)} \;+\; C_2 \|f\|_0 \log n \right\}$$

Problem: in principle this requires a combinatorial search

However, in practice iterative schemes are very effective:

M. Figueiredo and **R. Nowak,** "An EM Algorithm for Wavelet-Based Image Restoration", *IEEE Transactions on Image Processing*, 2003.

I. Daubechies, **M. Defrise** and **C. De Mol**, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math*, 2004.

Or solve the convex-relaxation (or a related form):

$$\widehat{f} \;=\; \arg\min_{f \in \mathcal{F}} \left\{ \|Y - \Phi' f\|^2_{\ell_2(k)} \;+\; C_2 \|f\|_1 \log n \right\}$$

E. Candès and T. Tao (2005), The Dantzig selector: statistical estimation when p is much larger than n.

Extensions – Compressible Signals

Sparsity of f* is crucial to basic theory of CS; similar results hold if f* is simply "compressible" in an orthobasis or in terms of more general representations (e.g., curvelets, TF dictionaries, wedgelets)

Extension to compressible signals:

Let $f^{(m)}$ the best m-term approximation of f^* and assume that

$$\frac{\|f^* - f^{(m)}\|^2}{n} \preceq m^{-\alpha}$$

Then previous bounds for sparse signals imply that

$$\sigma^2 = 0 \ : \ E\left[\frac{\|f^* - \widehat{f}\|^2}{n}\right] \preceq k^{-\alpha}$$

$$\sigma^2 > 0 \ : \ E\left[\frac{\|f^* - \widehat{f}\|^2}{n}\right] \preceq k^{-\alpha/(\alpha+1)}$$

CS vs. Pixel Sampling in Imaging

piecewise constant images
with C^2 smooth edges

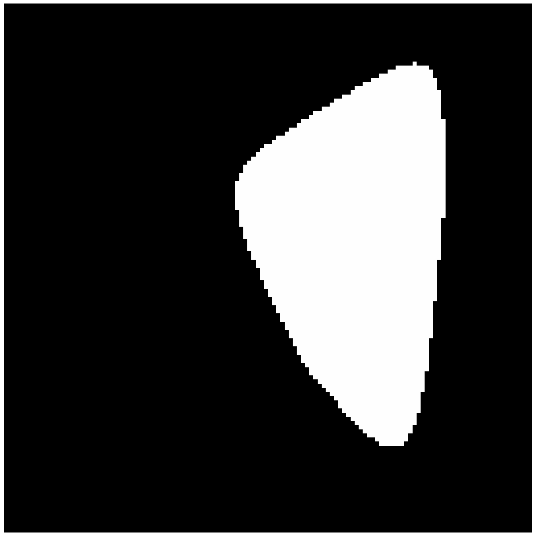


image f*

Images in this class are
very compressible

Best m-term approximation
from dictionary D:

$$f_m = \arg\min_{D_m} ||f^*-f||$$

Fourier dictionary:

$$\|f^* - f_m\|^2 \preceq m^{-1/2}$$

Wavelet dictionary:

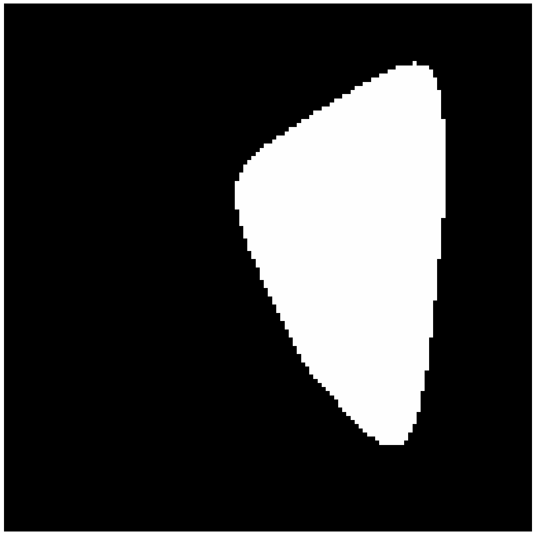
$$\|f^* - f_m\|^2 \preceq m^{-1}$$

Wedgelet/Curvelet dictionary:

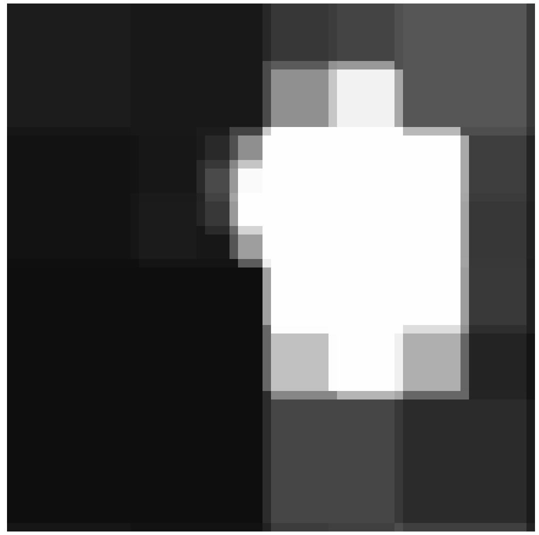
$$\|f^* - f_m\|^2 \preceq m^{-2}$$

Noisy CS, SNR=3dB

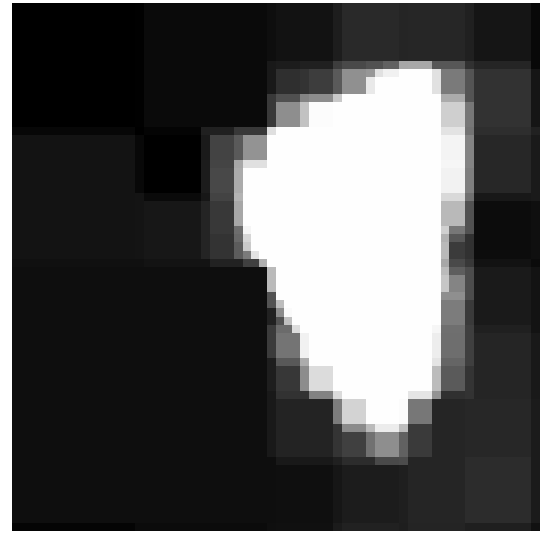
reconstruction using wedgelet dictionary



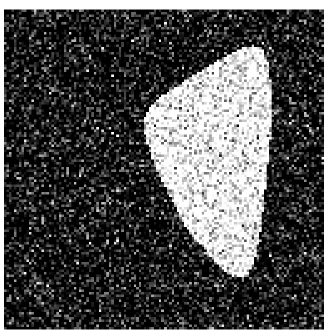
Original 128x128
(16K) Image



Reconstruction, 2K Projections



Reconstruction, 3K Projections



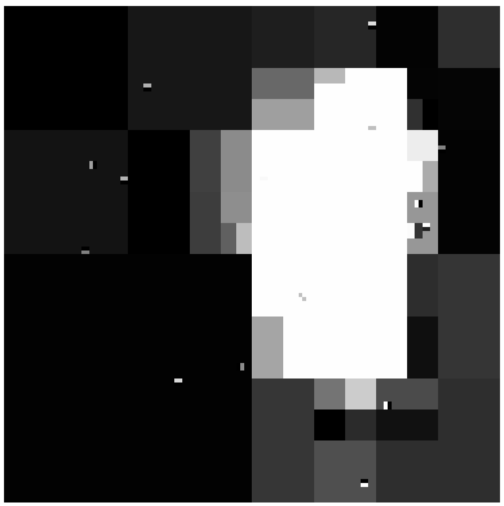
Equiv. Per-pixel Noise

How important is the representation?

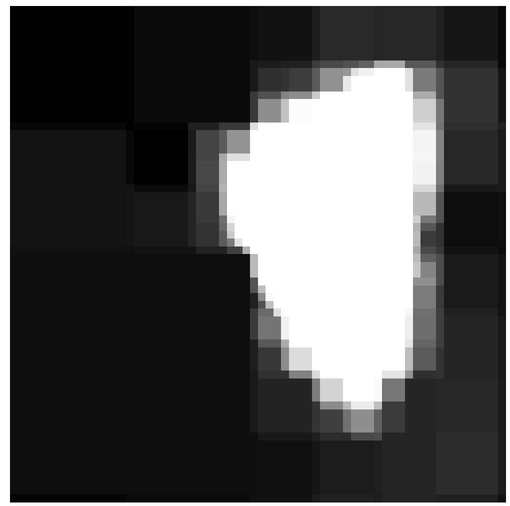
Reconstruction of piecewise constant images from noisy random projections:

wavelet err ~ $k^{-1/2}$

wedgelet/curvelet err ~ $k^{-2/3}$



3000 compressive samples + wavelet reconstruction

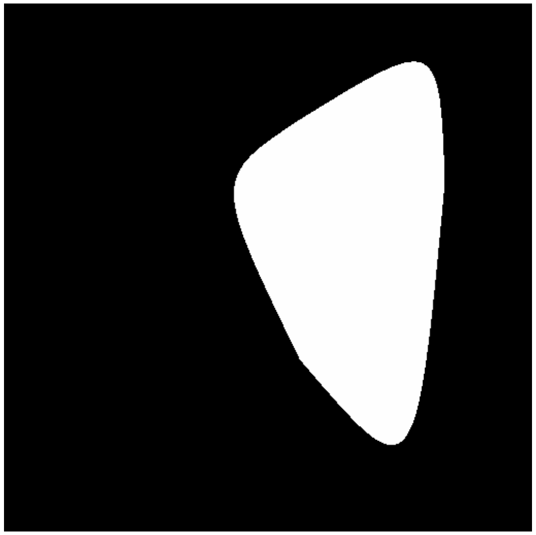


3000 compressive samples + wedgelet reconstruction

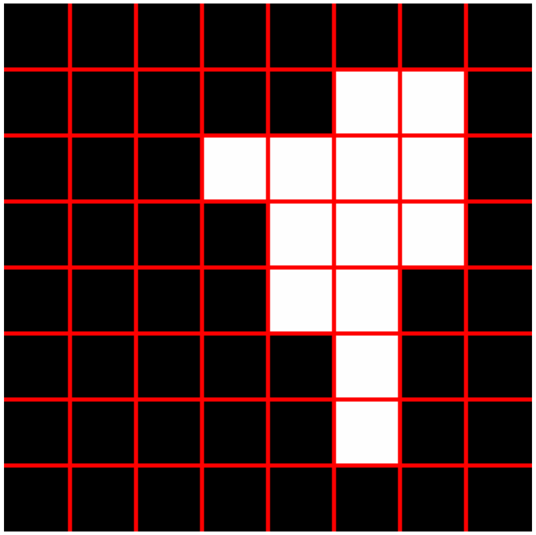
Conventional Imaging - Pixel Samping

Acquire k pixel samples

- $O(\mathbf{k}^{1/2})$ boundary pixels
- $O(\mathbf{1})$ error in each
- reconstruction error ~ $\mathbf{k}^{-1/2}$



Approximation error / bias of pixel "undersampling" limits reconstruction accuracy



Conventional Imaging vs. Compressive Sampling

	Pixel Sampling	CS(1)	CS(2)
Noiseless	$k^{-1/2}$	k^{-1}	k^{-2}
Noisy	$k^{-1/2}$	$k^{-1/2}$	$k^{-2/3}$

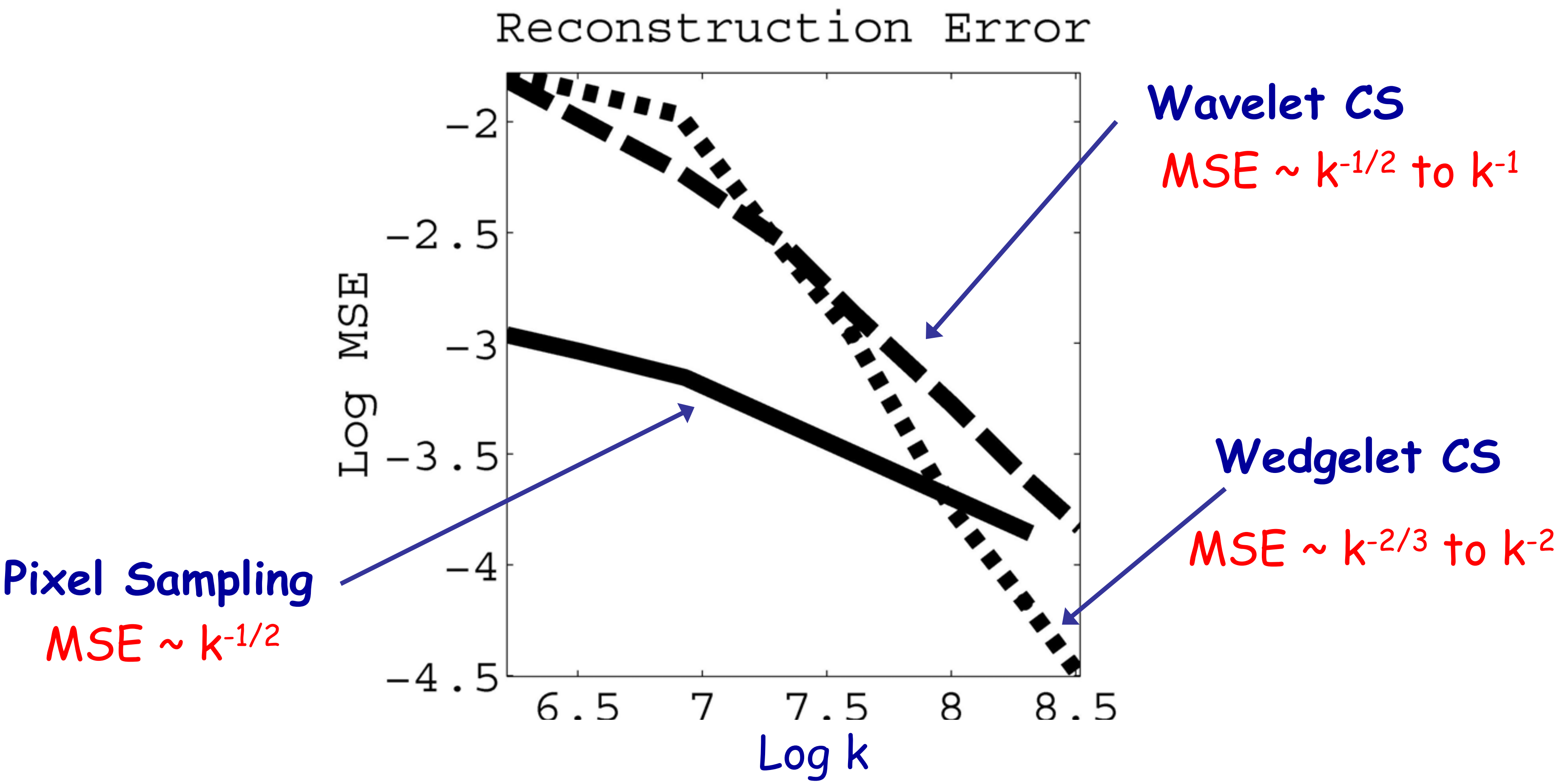
Reconstruction error bounds using **k** pixel samples or compressive samples [reconstructing with wavelets CS(1) or wedgelets CS(2)]

Compressive sampling performs (in theory) at least as well as conventional imaging

Better if SNR is high and/or target image is highly compressible

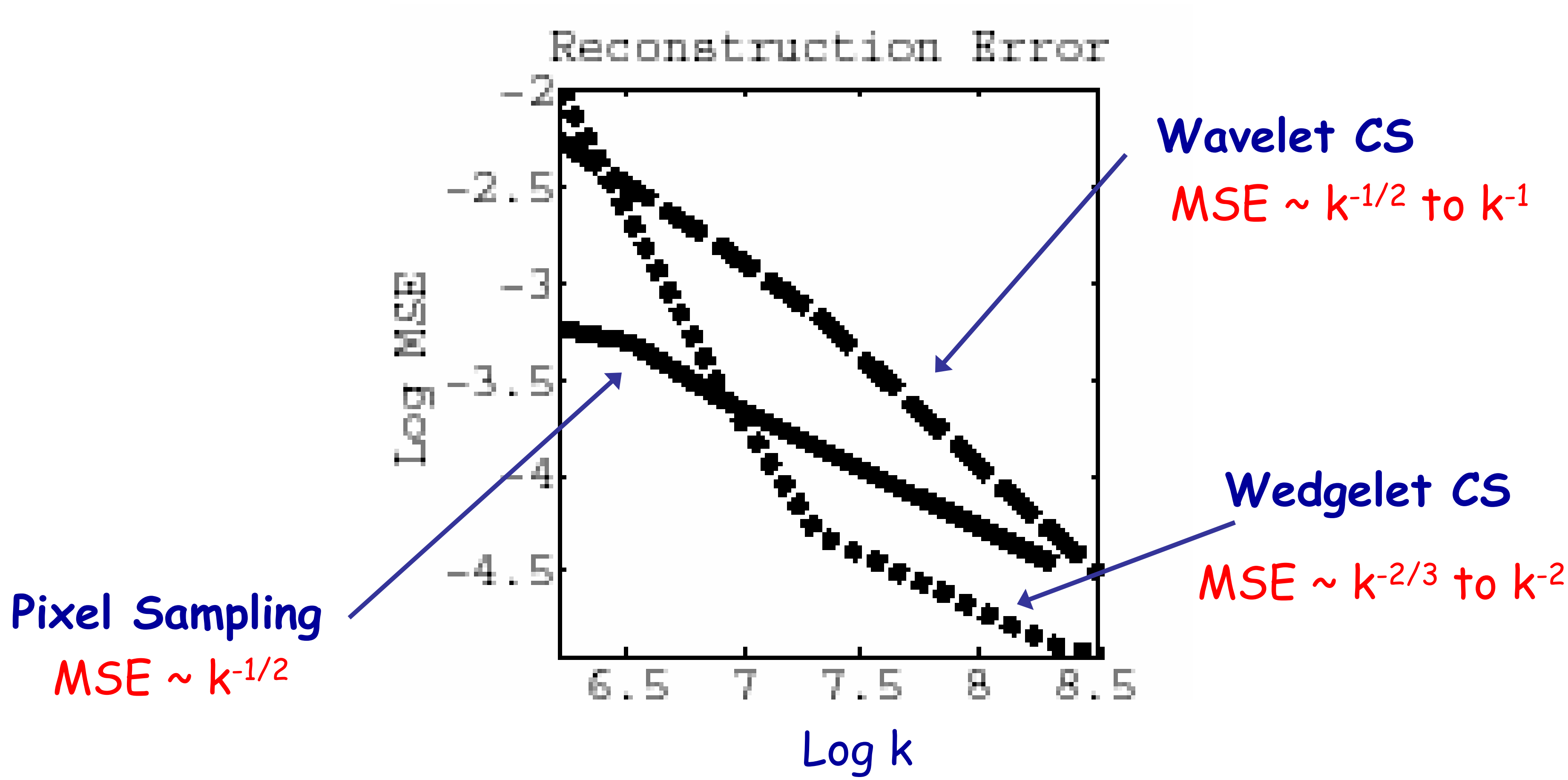
No better if SNR is not high and/or image is not highly compressible

Error Decay Rates, SNR=3dB

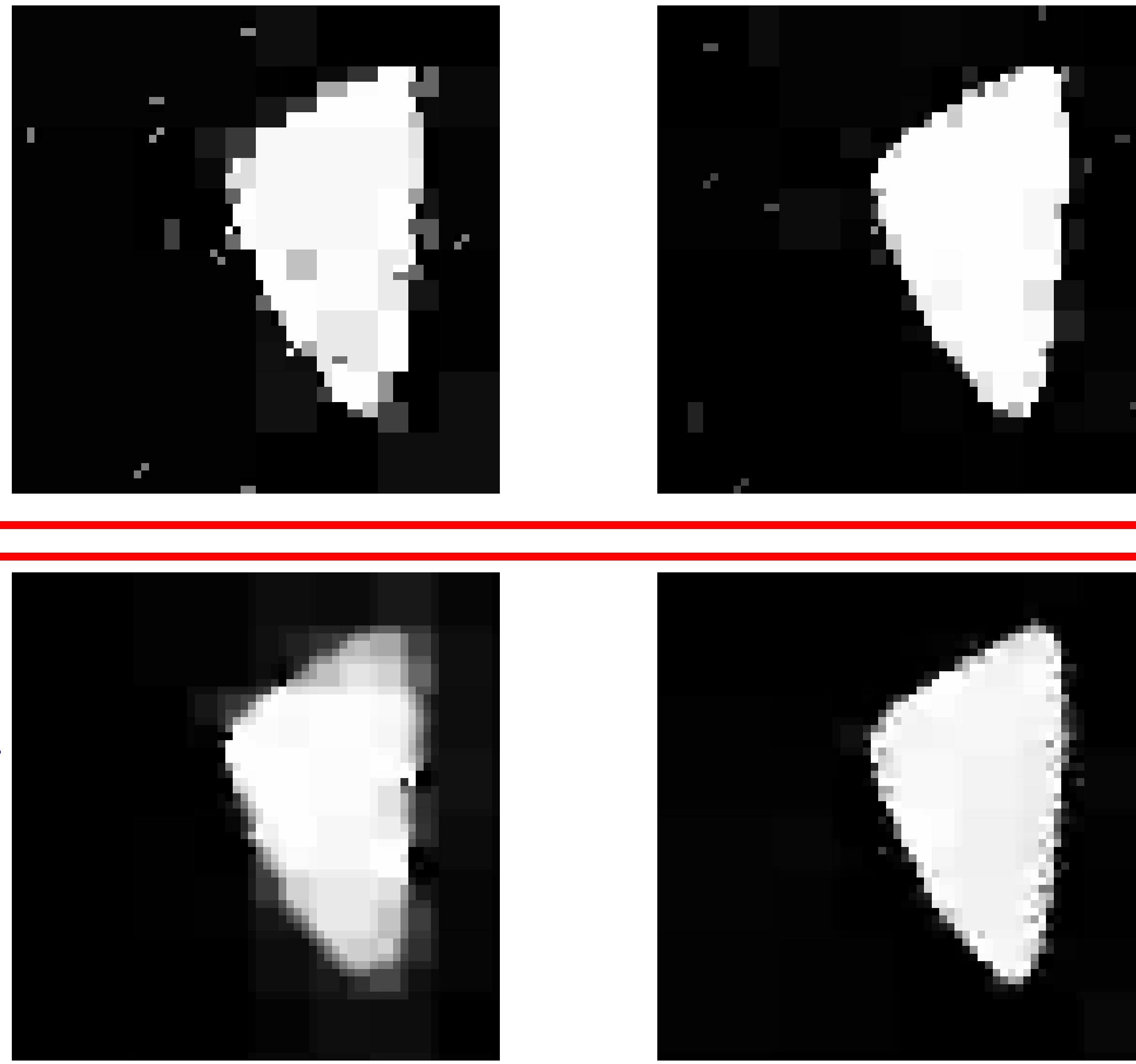


J. Haupt and R.N., "Compressive Sampling vs. Conventional Imaging," *To Appear at International Conference on Image Processing (ICIP) 2006, Atlanta, GA.*

Error Decay Rates, SNR=9dB



Conventional Imaging vs. Compressive Sampling



PIXEL SAMPLING

COMPRESSIVE SAMPLING

3dB SNR

9dB SNR

Conclusions

Compressive sampling techniques can yield accurate reconstructions even when the signal dimension greatly exceeds the number of samples, and even when the samples themselves are contaminated with significant levels of noise.

CS can be advantageous in noisy imaging problems if the underlying image is highly compressible or if the SNR is sufficiently large

Related Work:

Dantzig Selector: Candes & Tao '05 ℓ_1-based optimization

CompressiveSampling: Candes & Romberg, Donoho & Tanner,
Baraniuk et al, Gilbert & Tropp

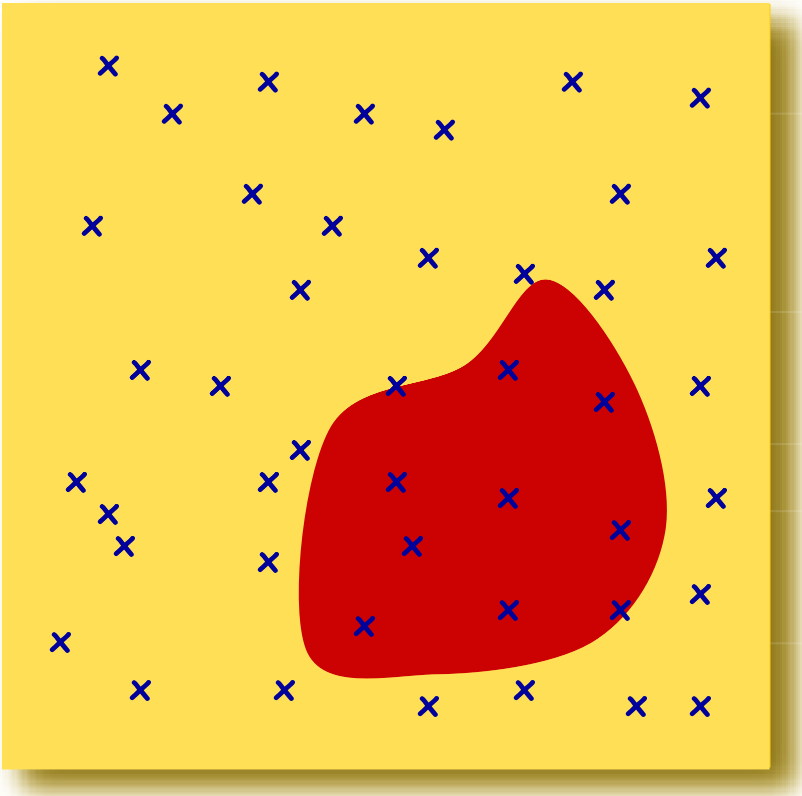
Emerging Schemes:

Adaptive Sampling: Korostelev '99, Hall & Molchanov '03,
Castro, Willett & Nowak '05
Castro, Haupt, Nowak '06

Nonlinear Sampling: Raz et al '05

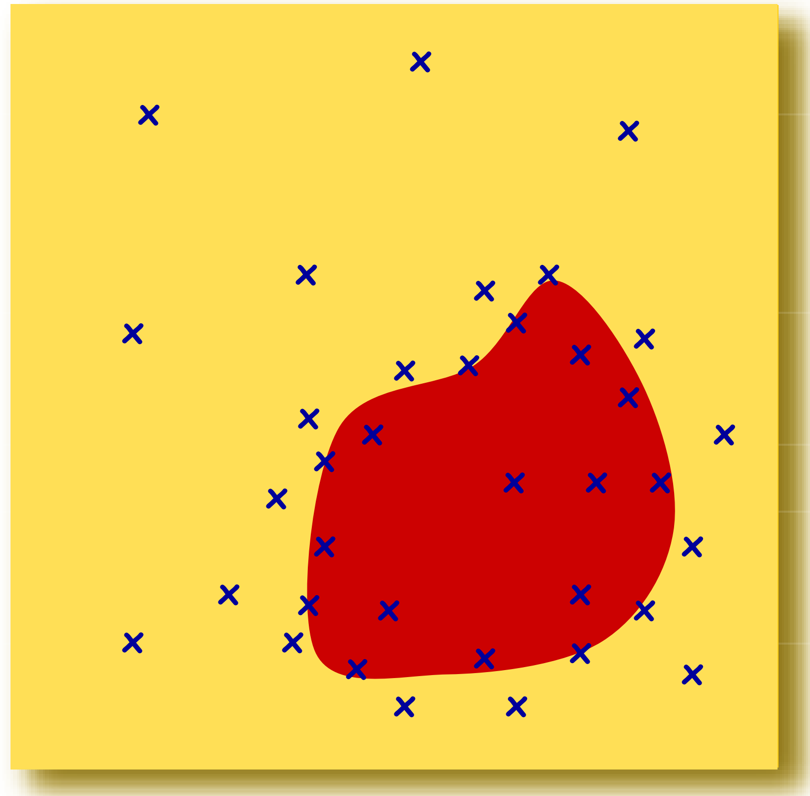
Adaptive Sampling

conventional sampling



Sample uniformly at random
(or deterministically)

adaptive sampling



Sequentially sample using
information gleaned from
previous samples

Theory of Adaptive Sampling

Key theoretical results for Adaptive Sampling

1. For certain classes of signals (e.g., piecewise smooth images) there exist adaptive sampling schemes that are optimal

→ **adaptive sampling cannot be significantly outperformed by any other strategy, including compressive sampling**

2. For piecewise constant signals, adaptive sampling is provably better than compressive sampling

→ **adaptive sampling may outperform compressive sampling by a large margin when sampling in the presence of noise**

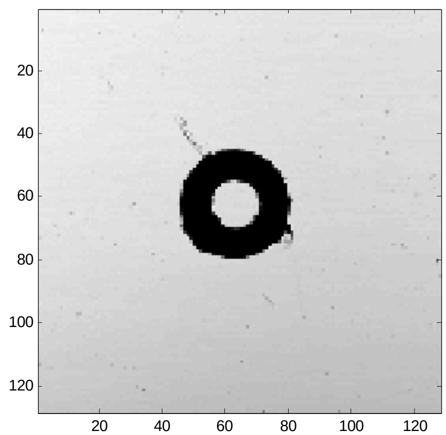
Castro, Willett and RN, "Faster Rates in Regression via Active Learning," NIPS 2005
Castro, Haupt and RN "Compressed Sensing vs. Active Learning," ICASSP 2006

Adaptive Sampling in Action

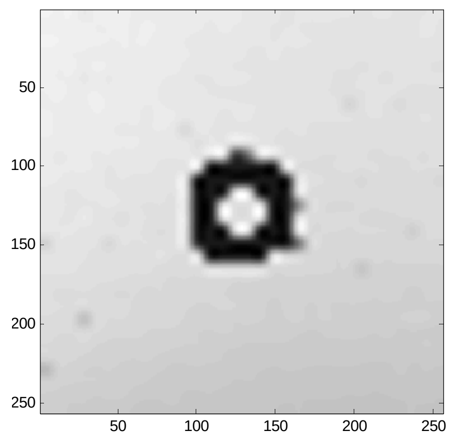
Laser imaging through highly scattering medium

Milanfar, Friedlander, Shakouri, Christofferson, Farsiu, Nowak, and Eriksson '06

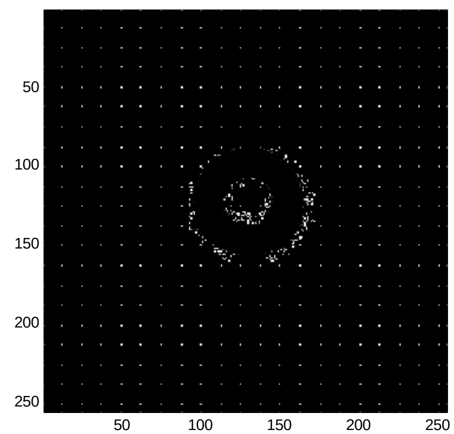
Below: Comparison of conventional and adaptive sampling of a "ring" phantom through a solid plastic diffuser



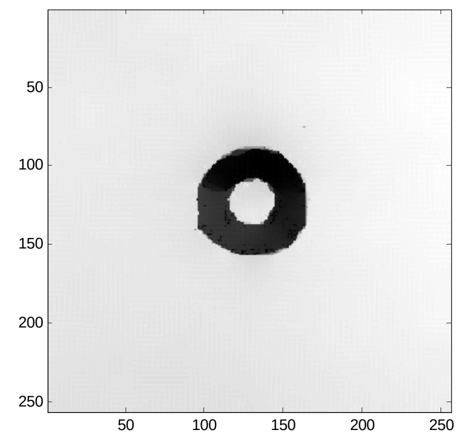
16K uniformly spaced samples



1K uniformly spaced samples



1K adaptively selected samples

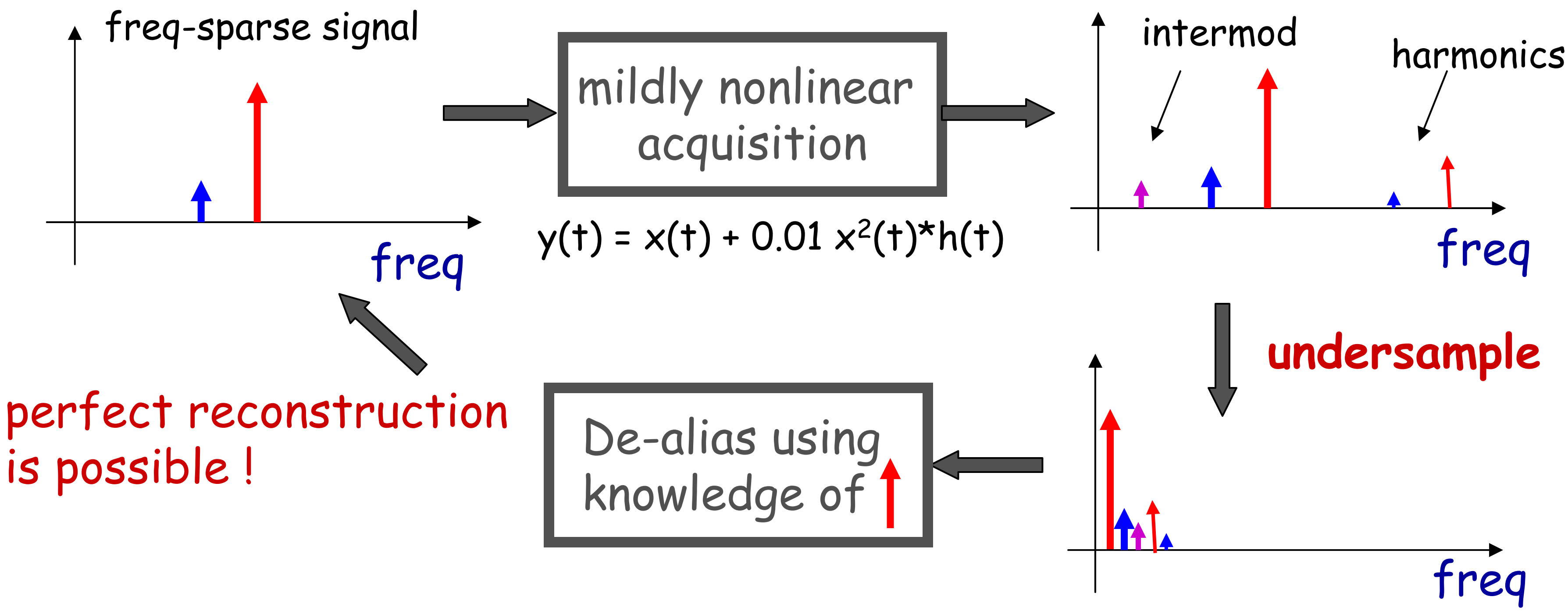


recon from adaptive samples

Nonlinear Sensing

Many sensing and sampling devices have inherent nonlinearities which can provide another source of *signal diversity*

Example:



Nonlinearities disambiguate aliasing of severly undersampled signals

G. Raz, "Method and System for Nonlinear and Affine Signal Processing," Publication number: US-2006-0132345-A1, U.S. patent pending.

2007 IEEE Statistical Signal Processing Workshop

26-29 August 2007 Madison, Wisconsin

General Chairs: R. Nowak & H. Krim
Tech Chairs: R. Baraniuk and A. Sayeed

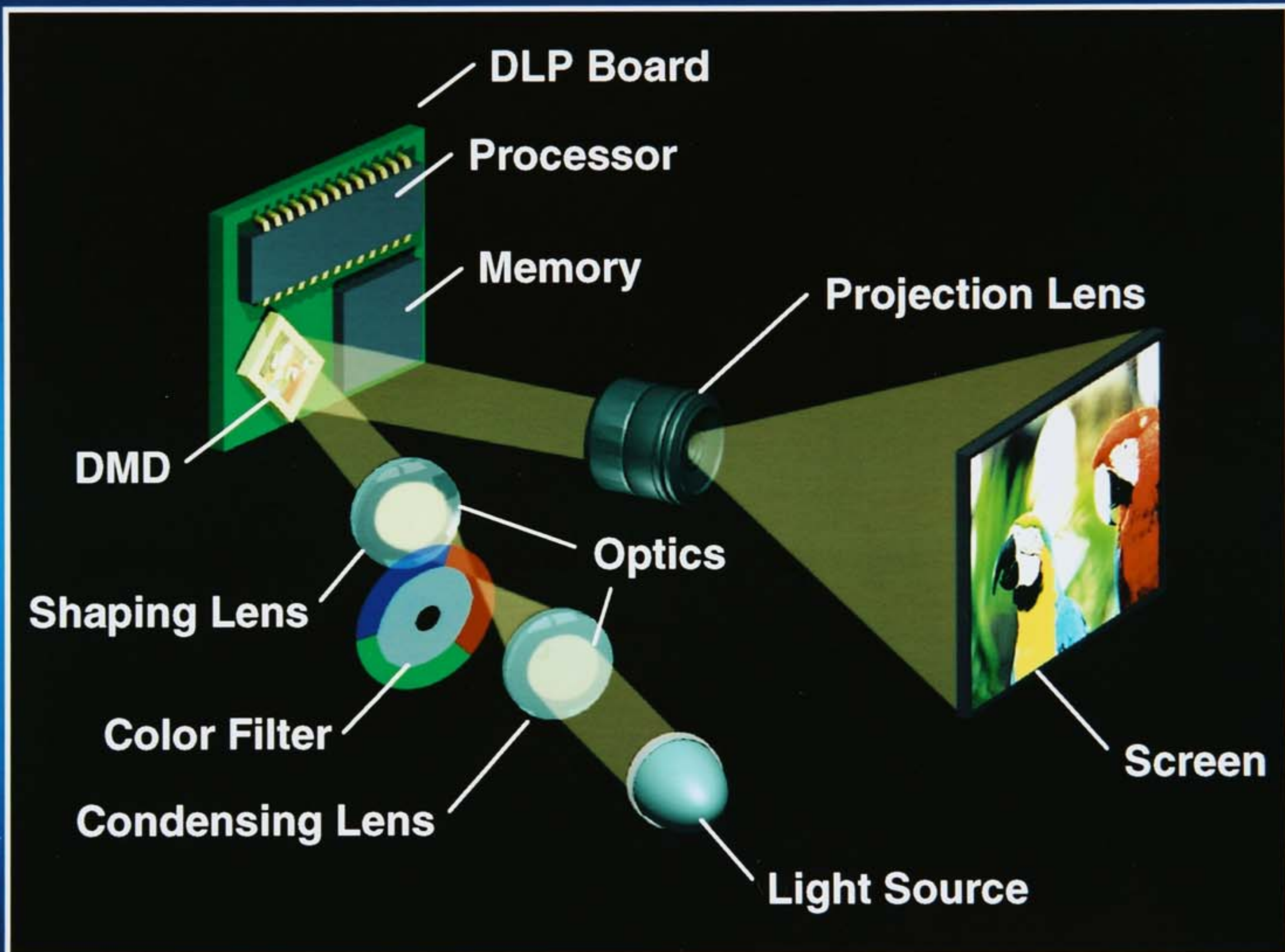
Theoretical topics	**Application areas**
Adaptive systems and signal processing	Bioinformatics and genomic signal processing
Monte Carlo methods	Automotive and industrial applications
Detection and estimation theory	Array processing, radar and sonar
Distributed signal processing	Communication systems and networks
Learning theory and pattern recognition	Sensor networks
Multivariate statistical analysis	Information forensics and security
System identification and calibration	New methods, directions and applications.
Time-frequency and time-scale analysis	Biosignal processing and medical imaging

Part IV – Applications and Wrapup

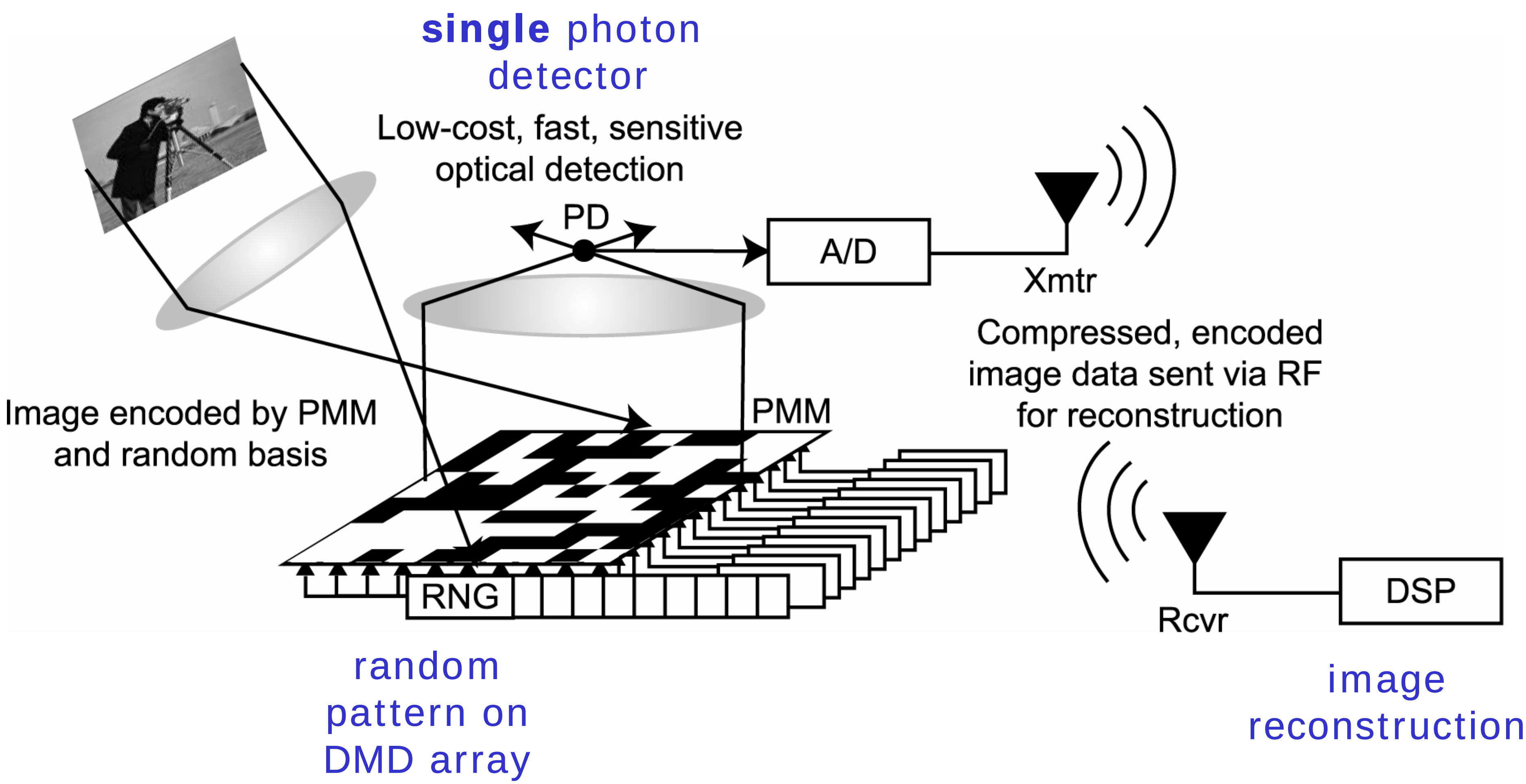
CS Hallmarks

- CS changes the rules of the data acquisition game
  - exploits a priori signal *sparsity* information
  - slogans: “sample less, compute more”
  - natural when measurement is *expensive*
- 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

1 Chip DLP™ Projection

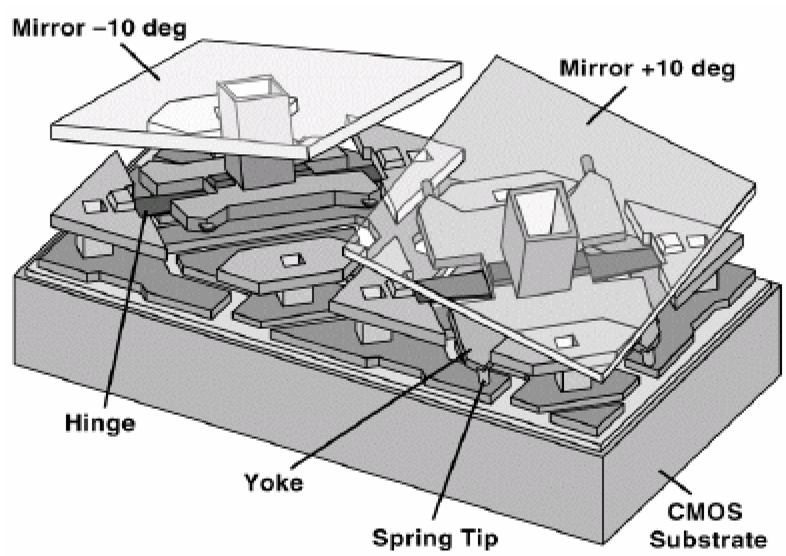


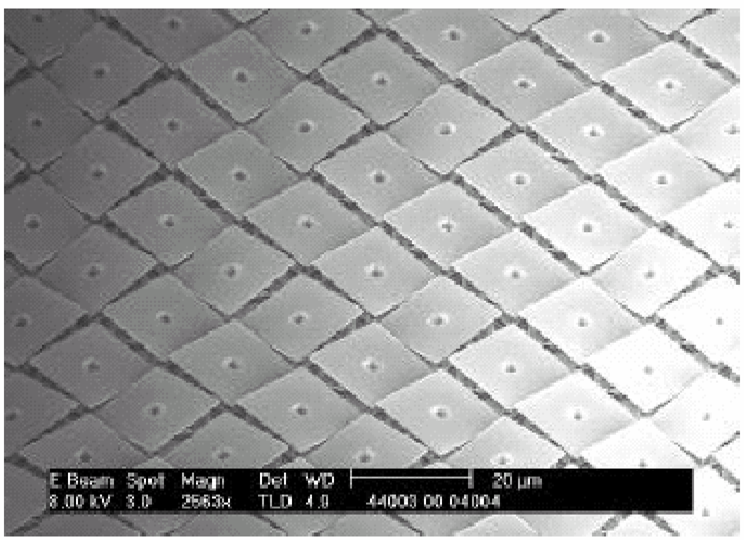
Rice CS Camera



with Kevin Kelly and students

TI Digital Micromirror Device (DMD)

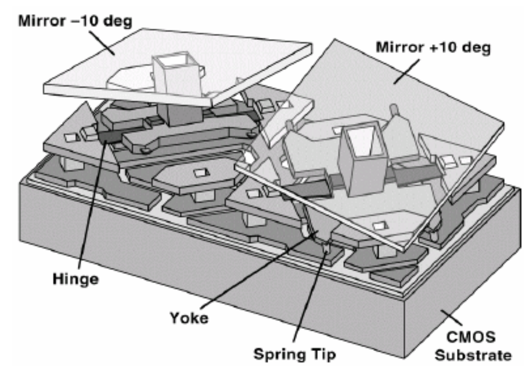




DLP 1080p --> 1920 x 1080 resolution

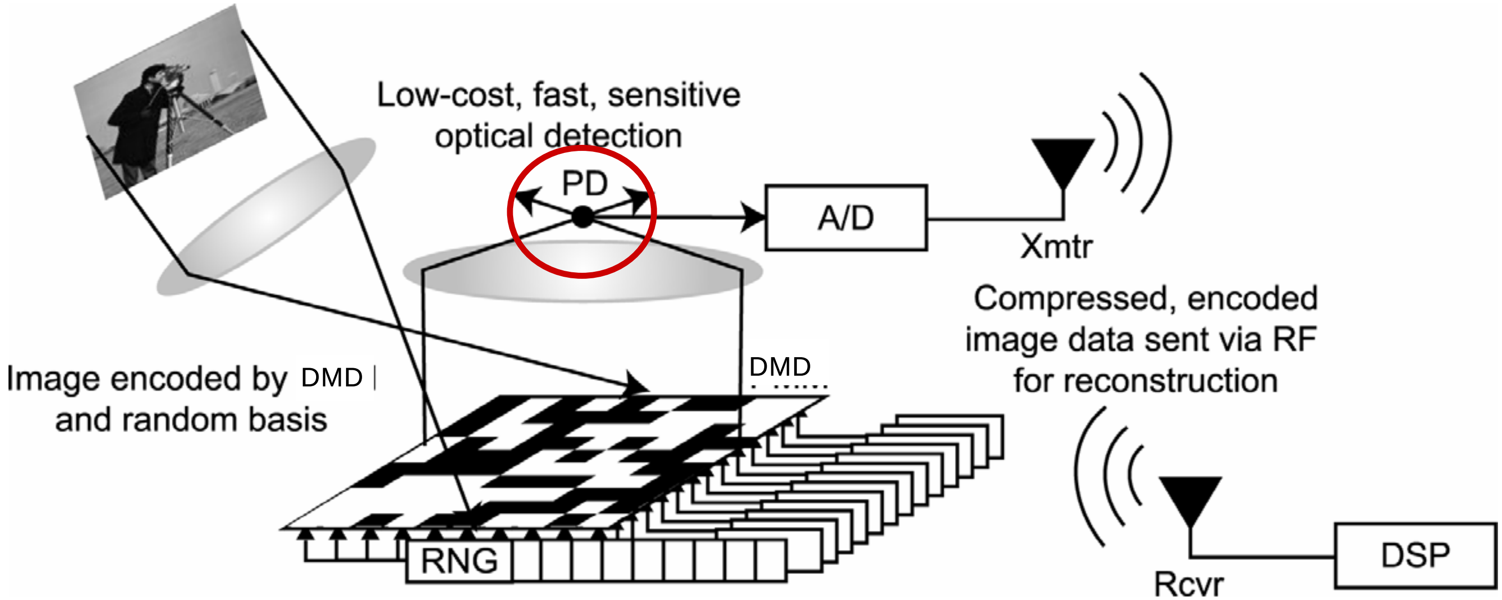
(Pseudo) Random Optical Projections

- Binary patterns are loaded into mirror array:
  - light reflected towards the lens/photodiode (1)
  - light reflected elsewhere (0)
  - pixel-wise products summed by lens



- Pseudorandom number generator outputs measurement basis vectors
- Mersenne Twister [Matsumoto/Nishimura, 1997]
  - Binary sequence (0/1)
  - Period 2^{19937}-1

Single Sensor Camera



Potential for:

- new modalities

  beyond what can be sensed by CCD or CMOS imagers
- low cost
- low power

First Image Acquisition

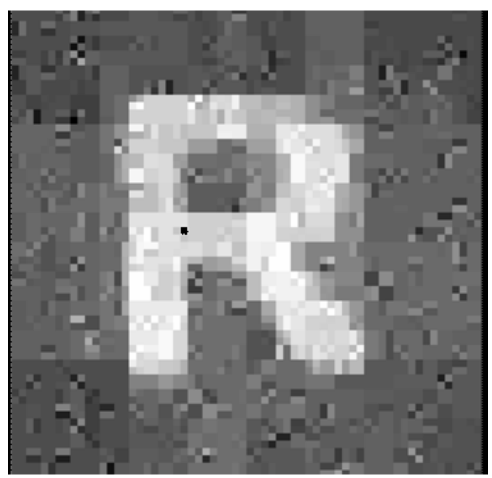
ideal 4096 pixels

205 wavelets

409 wavelets

image at
DMD array

820
random meas.



1638
random meas.

World’s First Photograph

- 1826, Joseph Niepce
- Farm buildings and sky
- 8 hour exposure
- On display at UT-Austin

Conclusions

- **Compressive sensing**
  - exploit image *sparsity* information
  - based on new *uncertainty principles*
  - "sample smarter", "universal hardware"
  - integrates sensing, compression, processing
  - matural when measurement is expensive

- Ongoing research
  - new kinds of *cameras* and *imaging* algorithms
  - new *"analog-to-information"* converters (analog CS)
  - new algs for *distributed source coding*
    - sensor nets, content distribution nets
  - *fast algorithms*
  - *R/D* analysis of CS (quantization)
  - CS vs. adaptive sampling

Contact Information

Richard Baraniuk
Rice University
richb@rice.edu
dsp.rice.edu/~richb

Robert Nowak
University of Wisconsin
nowak@engr.wisc.edu
www.ece.wisc.edu/~nowak

Justin Romberg
Georgia Institute of Technology
jromberg3@mail.gatech.edu
www.acm.caltech.edu/~jrom

CS community website
dsp.rice.edu/cs