Compressed Sensing and Bayesian Experimental Design

or

Optimal Sensing and Reconstruction of *N*-Dimensional Signals

by

Matthias Seeger & Hannes Nickisch

Presenter: Pete Trautman

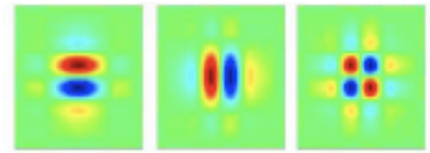
Outline

- Intro to compressive sensing
- Paper presentation

Sensing by sampling:

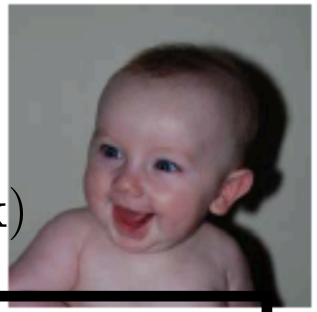
$$N \gg K$$

$$f(\mathbf{x}) \longrightarrow \boxed{\text{sample}} \xrightarrow{N} \boxed{\textbf{compress}} \xrightarrow{K} \boxed{\text{transmit/store}}$$



sparse
wavelet
transform

$$\boxed{\text{receive}} \xrightarrow{K} \boxed{\text{decompress}} \xrightarrow{N} f_N(\mathbf{x})$$



Pixel basis

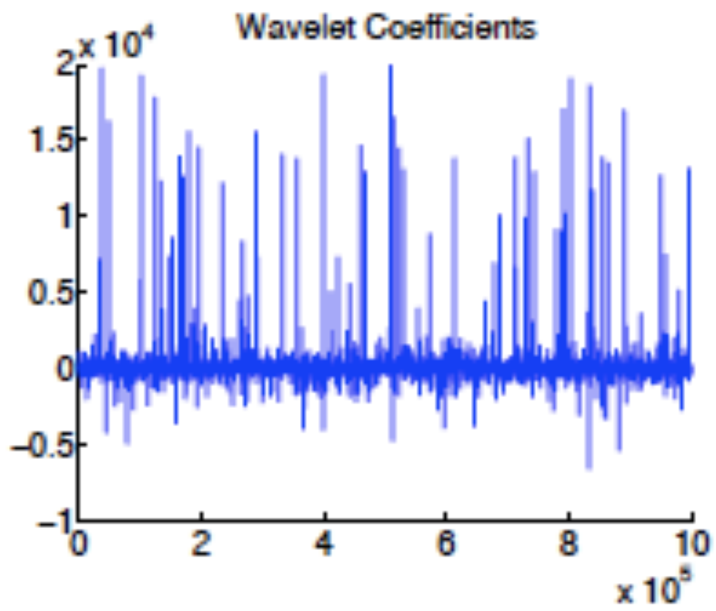
$$f(\mathbf{x}) \approx f_N(\mathbf{x}) = \sum_{i=1}^{N} f(\mathbf{x}^i)\delta(\mathbf{x} - \mathbf{x}^i)$$

Wavelet basis

$$\longrightarrow \hat{f} = \sum_{i=1}^{K} < f_N, \psi_i > \psi_i = \sum_{i=1}^{K} c^i \psi_i$$

Introduction to Compressive Sensing

Original image $f_N(\mathbf{x})$



Wavelet coefficients c^i

Image reconstruction: threshold all but 25000 largest coefficients

Begs the following: Can we measure the "compressive" measurement set directly?

A: yes.

Introduction to Compressive Sensing

- Traditional (Nyquist) sampling is highly pessimistic
  - Doesn't consider any structure of signal
- Compressive sensing is optimistic
  - *leverages compressibility*
  - => only need $K \ll N$ measurements to reconstruct an N-dim signal
- Intuition:
  - CS encodes sparsity as information
  - Allows for tradeoff between sparsity and # of measurements

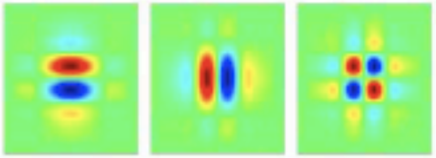
Compressive Sensing:

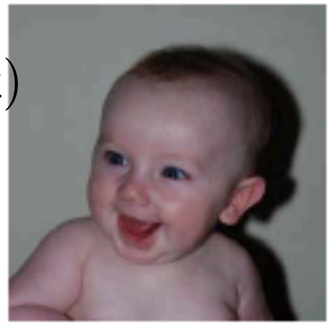
- Directly acquire "*compressed*" data
- Replace samples by more general "measurements"

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

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

receive $\xrightarrow{M}$ reconstruct $\xrightarrow{N}$ $f_N(\mathbf{x})$

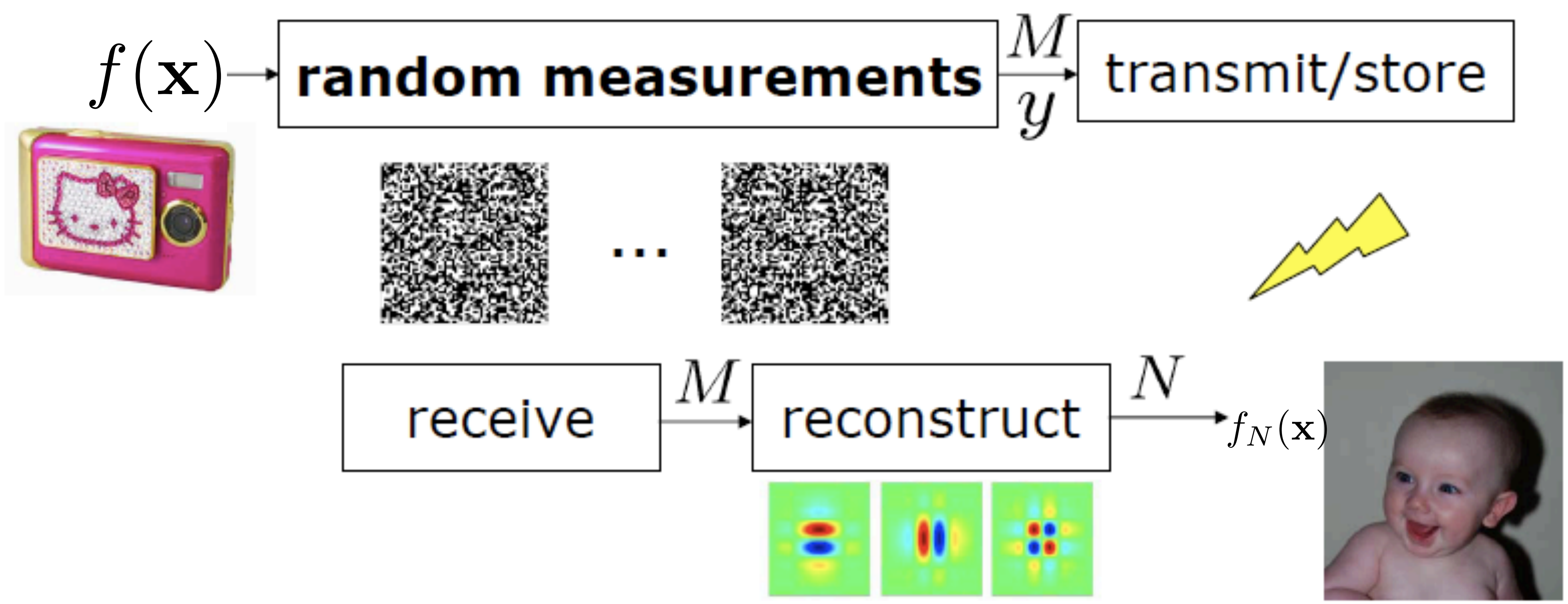




Compressive Sensing:

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

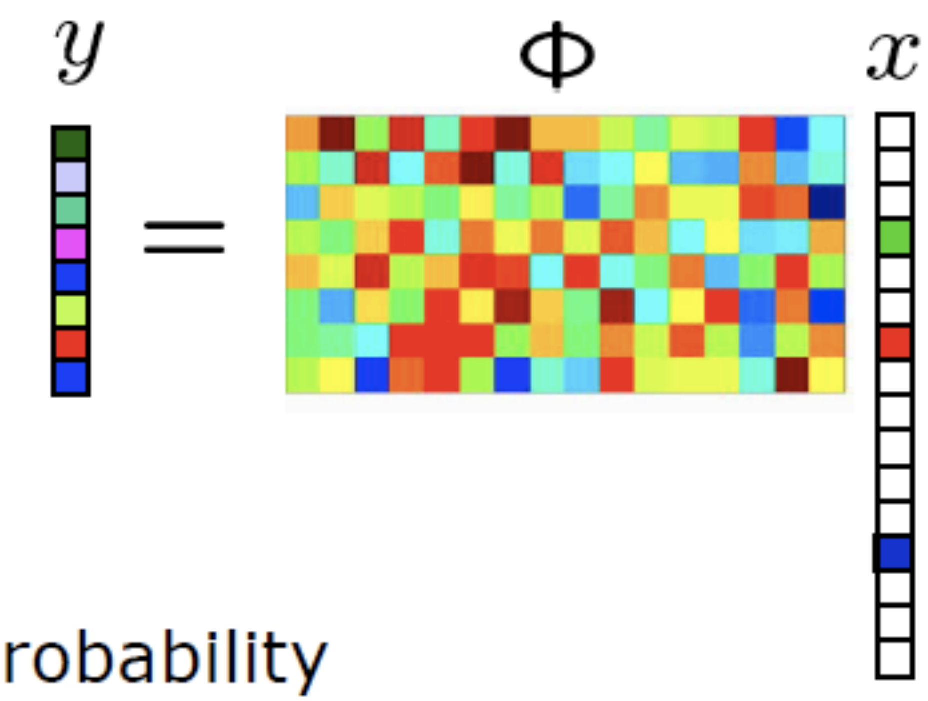
$$M = O(K \log(N/K))$$



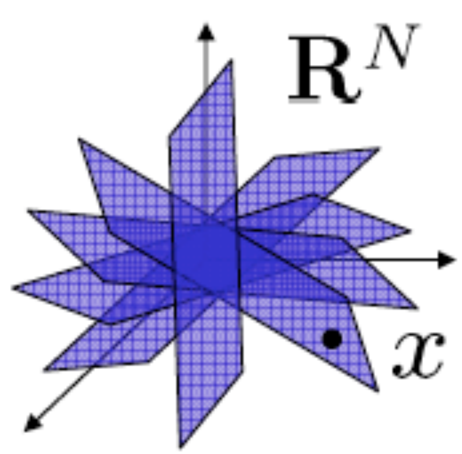
Why Does It Work?

- Random projection Φ
  not full rank...

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



K-sparse model

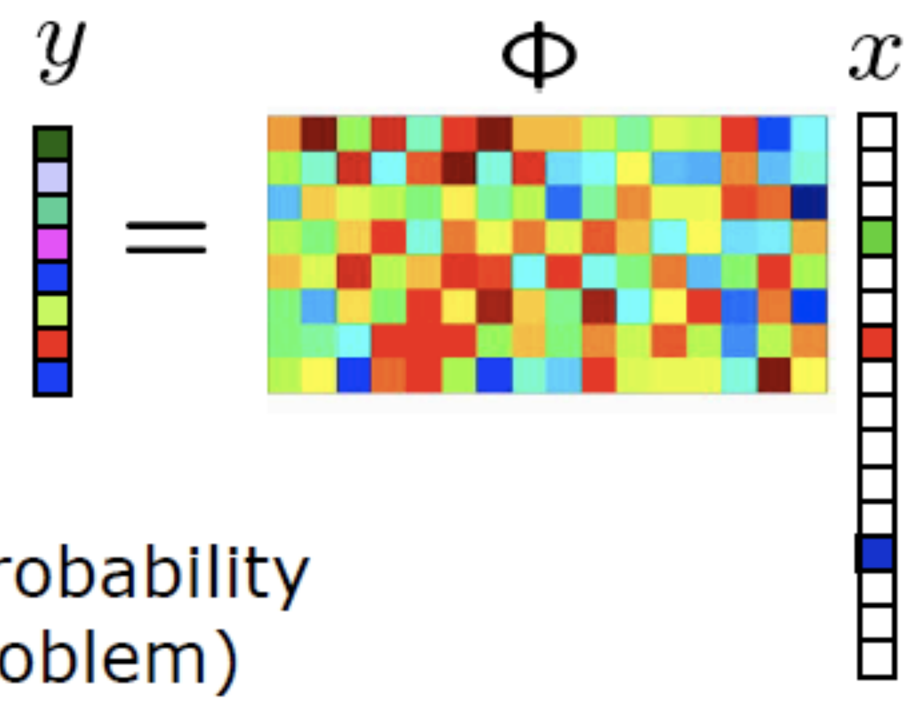


K-dim hyperplanes aligned with coordinate axes

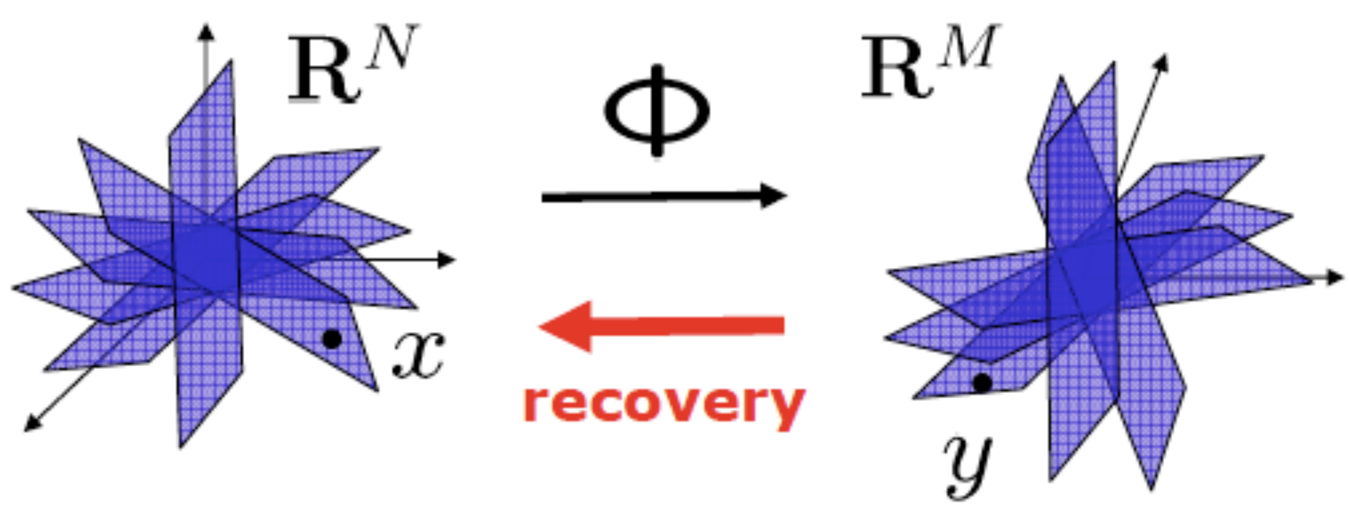
CS Signal Recovery

- Random projection Φ
  not full rank...

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



K-sparse
model



Sparse Recovery via ℓ_1 Minimization

- Say f_0 is K-sparse, Φ obeys RIP for sets of size $4K$
- Measure $y = \Phi f_0$
- Then solving
$$\min_f \ \|f\|_{\ell_1} \quad \text{subject to} \quad \Phi f = y$$
will recover f_0 exactly
- We can recover f_0 from
$$M \ \gtrsim \ K \cdot \log N$$
incoherent measurements by solving a *tractable* program
- *Number of measurements* $\approx$ *number of active components*

Sequential CS algorithm (segue)

Given seed measurement matrix $\mathbf{X} \implies \mathbf{y} = \mathbf{X}f$

1. Choose new row $\mathbf{x}_\star$ randomly
2. Form: $\mathbf{X}' = [\mathbf{X}\ \mathbf{x}_\star]^T$
3. Measure: $\mathbf{y}' = \mathbf{X}'f$
4. Reconstruct: $\hat{\mathbf{c}} = \arg\min_{\mathbf{c}}\{||\mathbf{c}||_{\ell_1} \,|\, \mathbf{y}' = \mathbf{X}'\Psi^T\mathbf{c}\}$, where $f_N = \sum_{i=1}^N \hat{c}^i \psi^i$
5. Repeat, starting with $\mathbf{X}'$

Goal of "CS and Bayesian Experimental Design": Improve Sequential CS by

- Optimizing step (1) above for general distributions
- Optimizing step (4) above for natural images

CS and BED: how to optimize

How to make these optimizations:

- let f be the signal of interest, f_N the reconstruction
- let $\mathbf{y}$ be the measurements, $\mathbf{X}$ the measurement matrix
- We seek $p(f_N|\mathbf{y})$

$$p(f_N|\mathbf{y}) \propto p(\mathbf{y}|f_N)p(f_N) \approx \mathcal{N}(\mathbf{y} = \mathbf{X}f|\mathbf{X}f_N, \sigma^2\mathbf{I})p(f_N)$$

- $p(f_N)$ encodes structural information about the signal: sparsity, smoothness, etc

  —Generalizes the ℓ_1 minimization of CS

- $\mathcal{N}(\mathbf{y} = \mathbf{X}f|\mathbf{X}f_N, \sigma^2\mathbf{I})$ is the likelihood

  —Generalizes the $\mathbf{y} = \mathbf{X}f_N$ constraint

CS and BED: how to choose next measurement

How to choose the next measurement $\mathbf{y}_* = \mathbf{x}_* f$?
Maximize entropy decrease (or information gain):

$$\min_{\mathbf{y}_*} H[p(f_N|\mathbf{y})] - H[p(f_N|\mathbf{y}, \mathbf{y}_*)]$$

However, $p(f_N|\mathbf{y})$ intractable; approximate using *Expectation Propagation*

$$Q(f_N) \approx p(f_N|\mathbf{y})$$

EP provides us with the following equation for the entropy difference:

$$H[Q(\mathbf{X})] - H[Q([\mathbf{X}\ \mathbf{x}_*]^T)] = \frac{1}{2}\log(1 + \sigma^{-2}\mathbf{x}_*^T \mathrm{Cov}_{Q(\mathbf{X})}(f)\mathbf{x}_*)$$

We thus choose $\mathbf{x}_*$ along the principal eigendirection of $\mathrm{Cov}_{Q(\mathbf{x})}(f)$

CS and BED: how to encode constraints

For images, we have two types of constraints on $p(f_N)$

- Sparsity (wavelet): $B^{(sp)} \in \mathbb{R}^{n \times n}$ is a wavelet transform
- Spatial Smoothness: $B^{(sp)} \in \mathbb{R}^{2(n-\sqrt{n}) \times n}$ is an image gradient transform

We turn these constraints into a distribution by using exponentials:

$$
\begin{aligned}
p(f_N) &\propto \exp(-\tau_{sp}||B^{(sp)} f_N||_{\ell_1}) \cdot \exp(-\tau_{tv}||B^{(tv)} f_N||_{\ell_1}) \\
&= \prod_{i=1}^{q_1} \exp(-\tau_{sp}|(B^{(sp)} f_N)_i|) \prod_{j=g_1}^{q_2} \exp(-\tau_{tv}|(B^{(tv)} f_N)_j|)
\end{aligned}
$$

The exponentials favor coefficients near zero, thus enforcing sparsity in both domains

CS and BED: synthetic experimental results

Title = type of signal

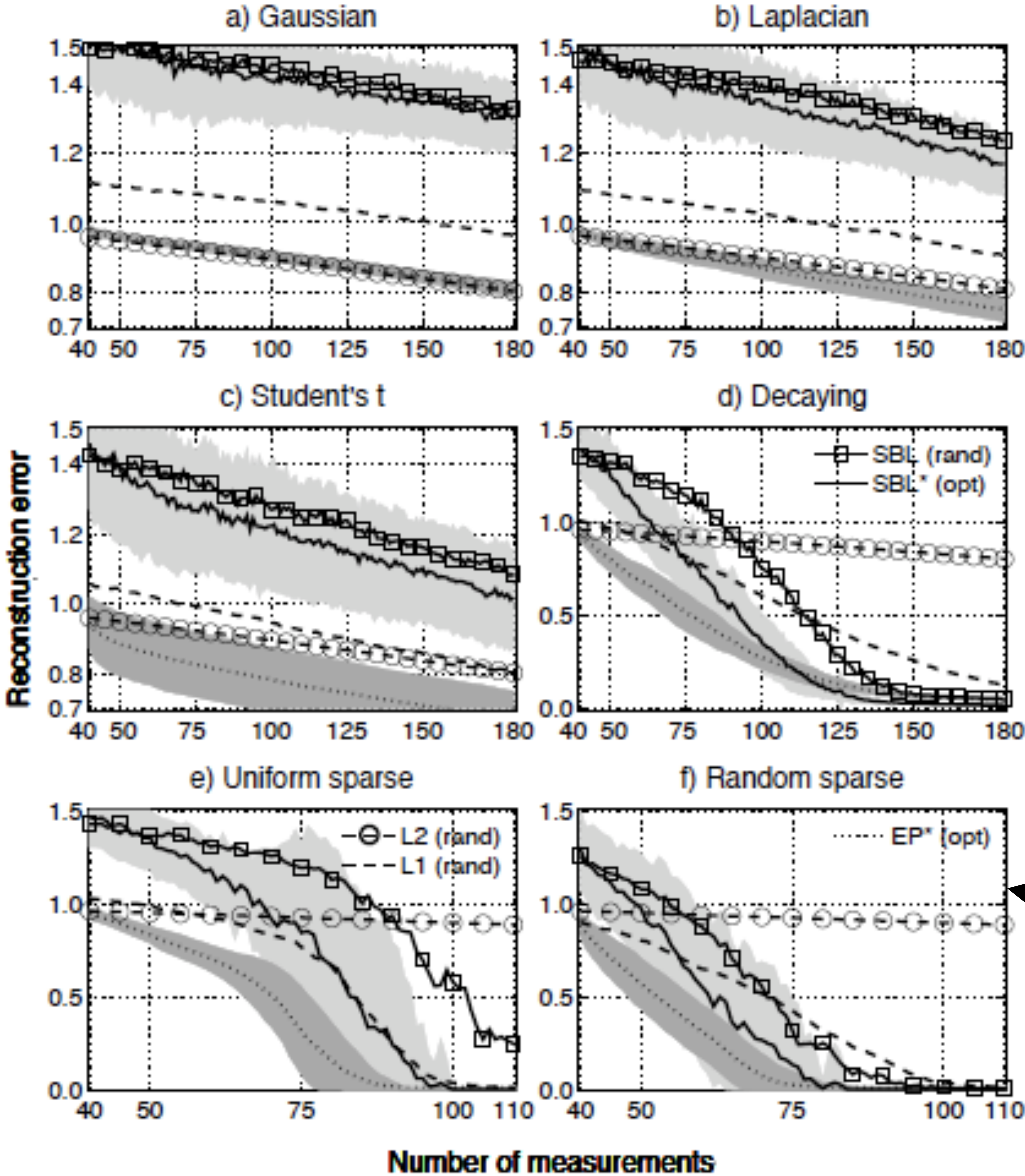


Figure 1. Comparison on 6 random synthetic signals $\boldsymbol{u} \in \mathbb{R}^{512}$. Shown are L_2-reconstruction errors (mean±std.dev. over 100 runs). All methods start with same random initial $\boldsymbol{X}$ ($m = 40$), then "(rand)" add random rows, "(opt)" optimise new rows sequentially. Noise variance $\sigma^2 = 0.005$, prior scale $\tau = 5$. SBL: (Ji & Carin, 2007), Lp: L_p reconstruction, EP: our method. (a-c): i.i.d. zero mean, unit variance Gaussian, Laplacian (Eq. 2), Student's t (3 d.o.f.). (d): $\frac{n}{2}$ of $u_i = 0$, $\frac{n}{4}$ exponential decay $1, \ldots, 0$, $\frac{n}{4}$ minus that, randomly permuted. (e-f): 20 $u_i \neq 0$ at random; (e) uniform spikes, $u_i \in \{\pm 1\}$; (f): non-uniform spikes, $u_i \sim \frac{1}{4} + |t|$, $t \sim N(0, 1)$; as in (Ji & Carin, 2007). Distributions in (d-f) normalised to unit variance.

What CS is made for

CS and BED: image experimental results

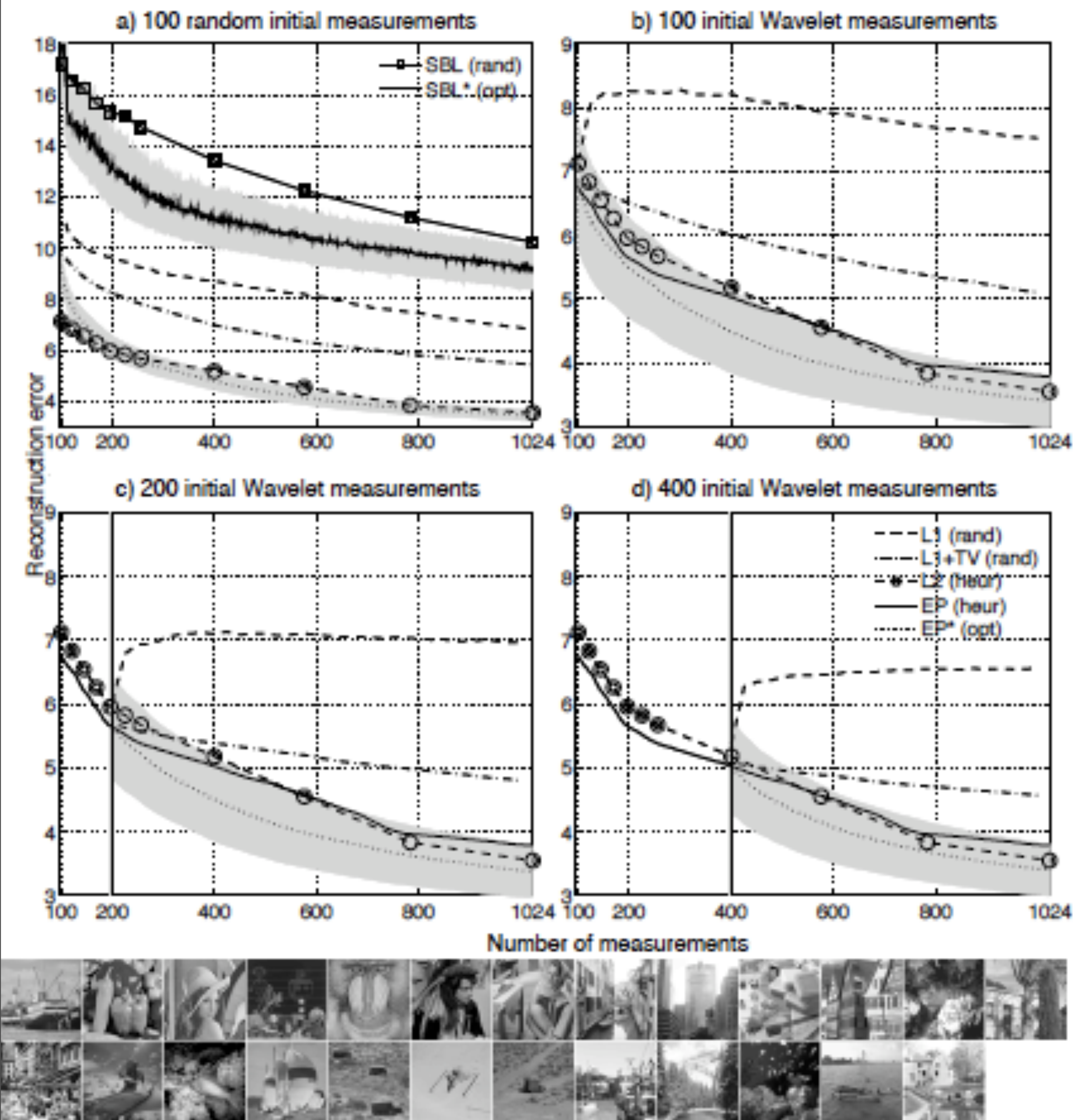


Figure 2. Experiments for measuring natural images ($64 \times 64 = 4096$ pixels). Shown are L_2-reconstruction errors averaged over 25 grayscale images typically used in computer vision research (from `decsai.ugr.es/cvg/dbimagenes/`) ($\pm\frac{1}{4}$std.dev. for "$*$"). Noise level $\sigma^2 = 0.005$. SBL: (Ji & Carin, 2007), Lp: L_p reconstruction, $L_1 + TV$: Lasso with TV/wavelet penalties, EP: our method. True σ^2 supplied, τ parameters chosen optimally for each method individually: $\tau_{sp} = \tau_{tv} = 0.075$ ($L_1 + TV$), $\tau_{sp} = 0.075$, $\tau_{tv} = 0.5$ (EP). New rows of $\boldsymbol{X}$ random unit norm (rand), actively designed (opt), acc. to wavelet heuristic (heur).
(a): Start with $m = 100$, $\boldsymbol{X}$ random unit norm. (b-d): Start with $m = 100, 200, 400$, $\boldsymbol{X}$ acc. to wavelet heuristic.

CS and BED: discussion

- Sequential Design outperforms CS protocols
  - However, measurement matrix of CS known in advance => much faster
- BED encompasses CS
- Much can be gained from the BED framework
  - enables encoding of many types of structural information
  - Optimizes information capture