Compressed Sensing

Some experimental promises **?**

Pierre Vandergheynst

Signal Processing Lab, EPFL

Journée du Traitement du Signal et de l'Image pour le Biomedical

Marseille, 10 septembre 2010

Some parts jointly with: G. Puy, Y. Wiaux (MRI) and N. Khaled, D. Atienza, H. Mamaghanian (ECG)

Some notations

$\mathbf{\Phi} \in \mathbb{R}^{d \times K}$ dictionary represented by a matrix with columns $\{\phi_1, ..., \phi_K\}$

$s = \mathbf{\Phi} b$ signal synthesized by coefficients b

$y = \mathbf{\Phi}^* s \in \mathbb{R}^K$ scalar products between s and elements of the dictionary (atoms)

The dictionary can be used to concatenate good signal components (ex: curvelets or shearlets + Gabor)

The dictionary can be learned from data

Sparsity Constrained Inverse Problems

Sparsity constrained recovery and inverse problems:

$$\tilde{b} = \arg\min_b \frac{1}{2}\|s - \mathbf{\Phi} b\|^2 + \mu\|b\|_1$$

$$\tilde{b} = \arg\min_b \frac{1}{2}\|y - \mathbf{U}\mathbf{\Phi} b\|^2 + \mu\|b\|_1 \Longrightarrow \tilde{s} = \mathbf{\Phi}\tilde{b}$$

observed signal (y)   **degrading operator** ($\mathbf{U}$)

$\tilde{y} = \mathbf{U}s$

$$\tilde{s} = \arg\min_s \frac{1}{2}\|y - \mathbf{U}s\|^2 + \mu\|s\|_{\mathrm{TV}} \qquad \|s\|_{\mathrm{TV}} = \sum_n \sqrt{|D_1 h[n]|^2 + |D_2 h[n]|^2}$$

For $\mathbf{U} = \mathbf{I}$: Rudin-Osher-Fatemi model

Fast algorithms: $\mathbf{U}$ ortho projector [Chambolle]

$\mathbf{U}$ general - proximal iterations [Combettes et al, Fadili, ...]

Take Home Messages So Far

- Many signals are sparse on some basis or dictionary
  - *zoology of fixed "optimal" bases*
  - *bases/dictionary learning*
- Sparsity offers a lot of flexibility
  - *dimensionality reduction*
  - *compression*
- Algorithms to handle sparsity (provably correct)
  - *greedy, convex relaxation ...*
- Applications !

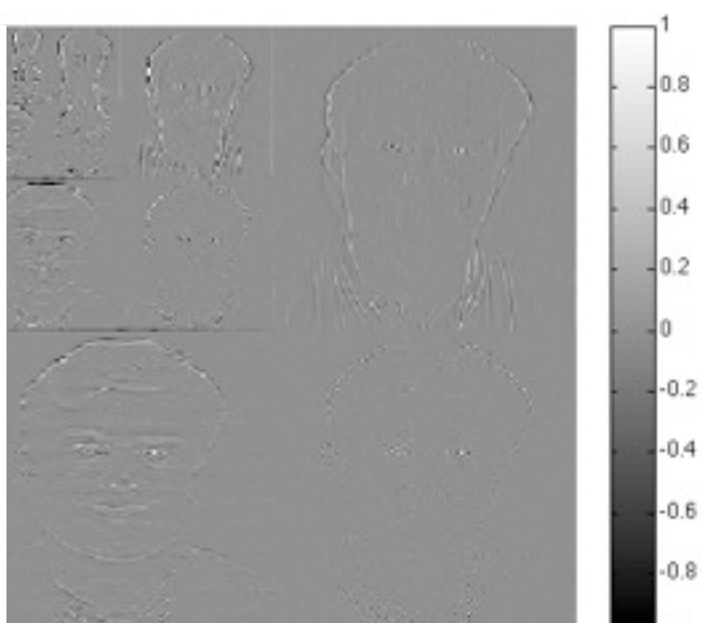
Sad Realization and Hopeful Wish

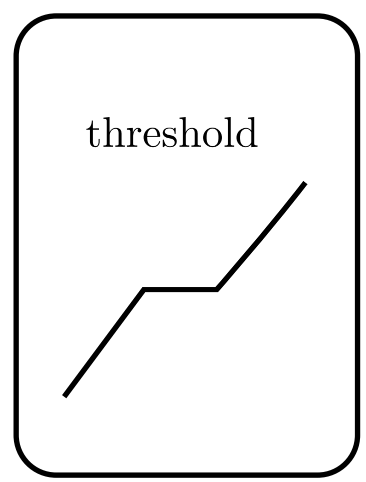
Sparse recovery techniques are great for processing data but ...

... you acquire the whole signal, i.e dimension N and then ...

you trash most of it because you know it is sparse on some good basis !

$$x = \mathbf{\Psi}\alpha$$





Sad Realization and Hopeful Wish

Sparse recovery techniques are great for processing data but ...

... you acquire the whole signal, i.e dimension d and then ...

you trash most of it because you know it is sparse on some good basis !

$$x = \mathbf{\Psi}\alpha$$

Would it be possible to acquire only those important components ???

$$y = \mathbf{\Phi}x = \mathbf{\Phi}\mathbf{\Psi}\alpha$$

$$\mathbf{\Phi} \in \mathbb{R}^{M \times N} \text{ with } M << N \text{ and } M \sim K$$

Sparse Recovery: principle

$$\alpha^* = \arg\min_{\alpha \in \mathbb{R}^N} \|\alpha\|_{\ell_1} \text{ subject to } \|\Phi\Psi\alpha - y\|_2 \leq \epsilon$$

Sparsity constrained inverse problem

$$(1-\delta_K)\|\alpha_K\|_2^2 \leq \|\Phi\alpha_K\|_2^2 \leq (1+\delta_K)\|\alpha_K\|_2^2$$

For all K-sparse vectors

Restricted Isometry Property (RIP)

RIP 2K

$$\|\alpha^* - \alpha\|_{\ell_2} \leq C_0\|\alpha - \alpha_K\|_{\ell_1}/\sqrt{K} + C_1 \cdot \epsilon$$

best K-term approximation — noise

Randomness and Incoherence

Suppose the signal is sparse on an ortho basis: $x = \mathbf{\Psi}\alpha$

Intuitively you may want to sample in an incoherent basis:

$$\mu(\mathbf{\Phi}, \mathbf{\Psi}) = \sqrt{N} \cdot \max_{k,j} |\langle \varphi_k, \psi_j \rangle|$$

$$\min_{\alpha \in \mathbb{R}^N} \|\alpha\|_{\ell_1} \text{ subject to } y = (R_M \Phi^*)\Psi\alpha$$

$$M \geq C \cdot \mu^2(\mathbf{\Phi}, \mathbf{\Psi}) \cdot K \cdot \log N$$

Bring Home Key Concepts

- Sparsity / Compressibility
  - large dimension but few degrees of freedom
- Linear (non adaptive !) measurements
  - $M = \mathcal{O}(K \log N/K)$
- Incoherence / Randomness
  - each measurement counts !
  - universality, robustness, scalability
- Recovery
  - provably correct algoS to solve inverse problem

Compressed Sensing in MRI

- Principles:
  - Image magnetization of tissues
  - RF pulses “read” Fourier transform of data

$$\nu(\boldsymbol{k}) = \int_{\mathbb{R}^2} \rho(\boldsymbol{x})\, \mathrm{e}^{-2\mathrm{i}\pi \boldsymbol{k}\cdot\boldsymbol{x}}\, \mathrm{d}^2\boldsymbol{x}.$$

- Problem
  - Long acquisition time for complete reading
- Solution
  - Reduce number of measurement

MRI model

- In the perspective of signal reconstruction, an ill-posed inverse problem has to be solved:

$$\boldsymbol{\nu} \equiv \Phi \boldsymbol{\rho} + \boldsymbol{n} \text{ with } \Phi \equiv \mathsf{MF}$$

- $\mathbb{C}^{M/2}$ *Measurements* ($\boldsymbol{\nu}$)
- $\mathbb{C}^{(M/2)\times N}$ *Sensing matrix* (Φ)
- $\mathbb{R}^{N}$ *Image* ($\boldsymbol{\rho}$)
- $\mathbb{C}^{M/2}$ *Noise* ($\boldsymbol{n}$)
- $\mathbb{R}^{(M/2)\times N}$ *Visibility mask* (M)
- $\mathbb{C}^{N\times N}$ *Fourier transform* (F)

Compressed sensing

- $\boldsymbol{x} \in \mathbb{R}^N$ is sparse in a basis $\Psi \in \mathbb{R}^{N \times N}$: $\boldsymbol{x} \equiv \Psi \boldsymbol{\alpha}$ with $\boldsymbol{\alpha} \in \mathbb{R}^N$ containing K non-zero entries.
- The condition for accurate and stable recovery for random selection of Fourier measurements $\Phi \equiv \mathsf{MF} \in \mathbb{R}^{M \times N}$ reads as:

$$K \leq \frac{cM}{N \mu^2 (\mathsf{F}, \Psi) \ln^4 N}$$

Is it possible to optimize the setting by “pre-conditioning” the signal ?

Chirp modulation & coherence

- For signals made up of Gaussian waveforms of size t, $\Psi \equiv \Gamma^{(t)}$ the coherence takes the form:

$$\mu\left(\mathsf{FC}^{(w)}\mathsf{A}^{(t_0)}, \Gamma^{(t)}\right) = \frac{2tt_0}{t^2+t_0^2}\left[1+\left(\frac{2\pi w t^2 t_0^2}{t^2+t_0^2}\right)^2\right]^{-\frac{1}{2}}.$$

- Natural limit for signal of spikes when $t \rightarrow 0$:

$$\lim_{t\rightarrow 0}\mu\left(\mathsf{FC}^{(w)}\mathsf{A}^{(t_0)}, \Gamma^{(t)}\right) = 0 \text{ for all } w, t_0 \in \mathbb{R}_+.$$

- Incoherence lost at finite t completely recovered at high enough w:

$$\lim_{w\rightarrow \infty}\mu\left(\mathsf{FC}^{(w)}\mathsf{A}^{(t_0)}, \Gamma^{(t)}\right) = 0 \text{ for all } t, t_0 \in \mathbb{R}_+.$$

Spread spectrum universality !

Simulations

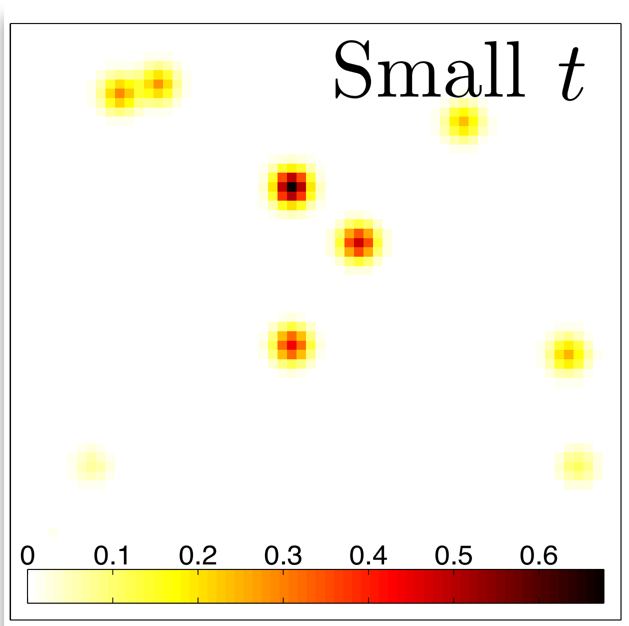
- Signals are made up of 10 waveforms in $\Gamma^{(t)}$ for 2 values of t .

- Noisy visibilities are simulated for 2 values of w (0 and 1).

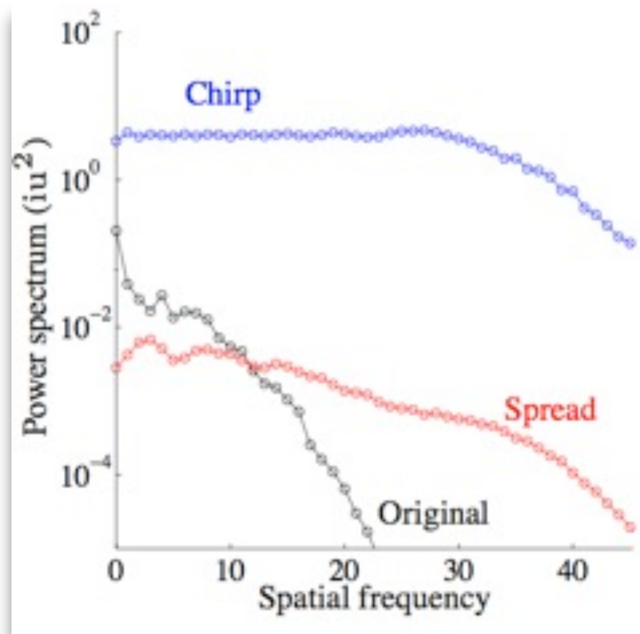
- The BP_ϵ problem is solved with 2 assumed sparsity dictionaries: for optimal sparsity $\Gamma^{(t)}$ or for optimal coherence Δ .

Simulations

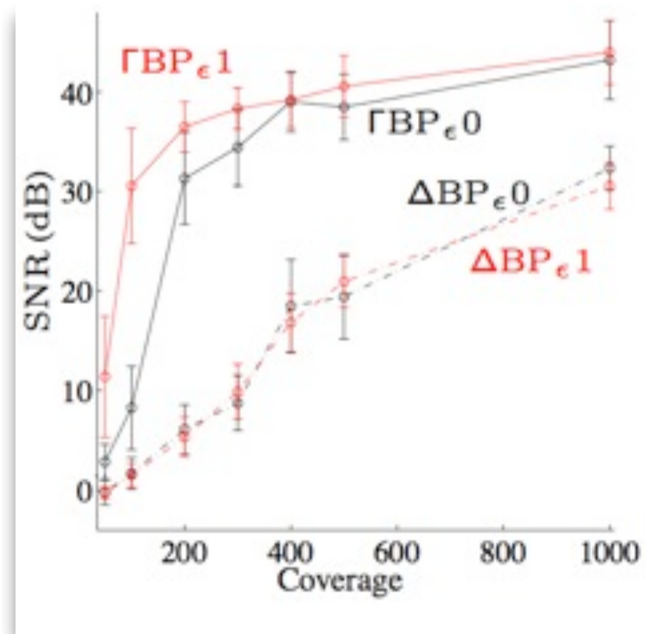
Original signal



Spread spectrum

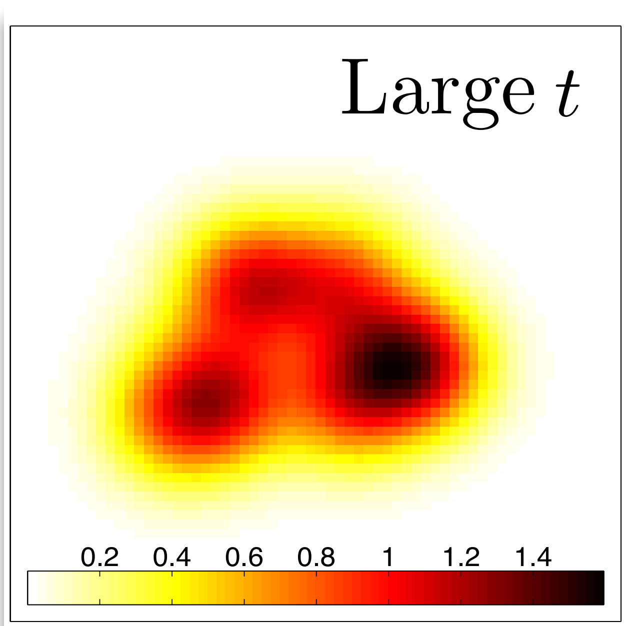


Reconstructions

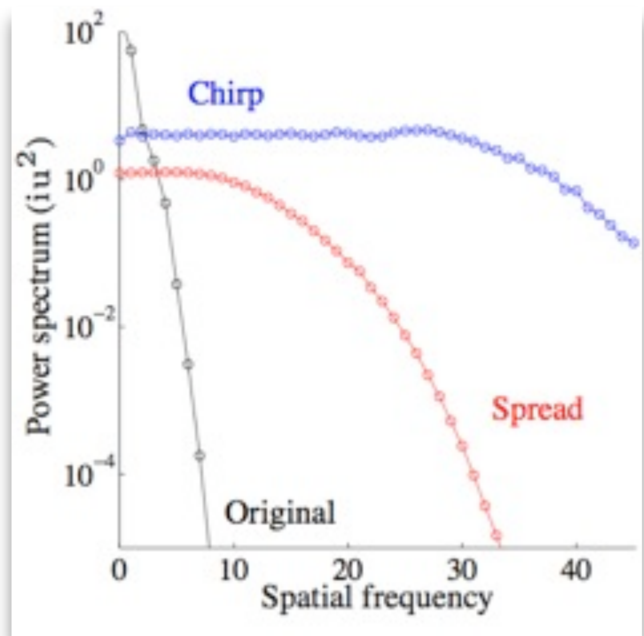


Simulations

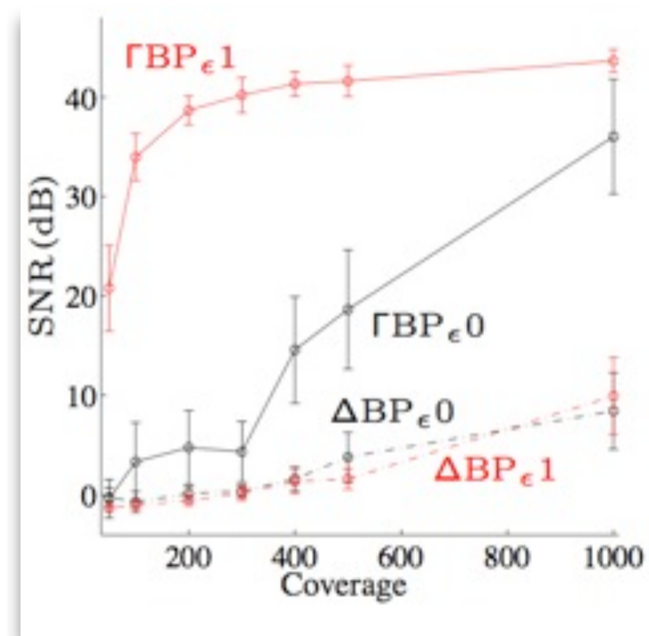
Original signal



Spread spectrum



Reconstructions



Simulations

1. $\Gamma\mathrm{BP}_\epsilon$ better than $\Delta\mathrm{BP}_\epsilon$: rather optimize sparsity than coherence!
2. $\Delta\mathrm{BP}_\epsilon 1$ equivalent to $\Delta\mathrm{BP}_\epsilon 0$ as μ is already optimal for $\Delta\mathrm{BP}_\epsilon 0$.
3. $\Gamma\mathrm{BP}_\epsilon 1$ better than $\Gamma\mathrm{BP}_\epsilon 0$ as μ is lower.
4. $\Gamma\mathrm{BP}_\epsilon 1$ independent of t: spread spectrum universality confirmed!

Theoretical model

Let $\mathbf{c}$ be a Rademacher or Steinhaus sequence

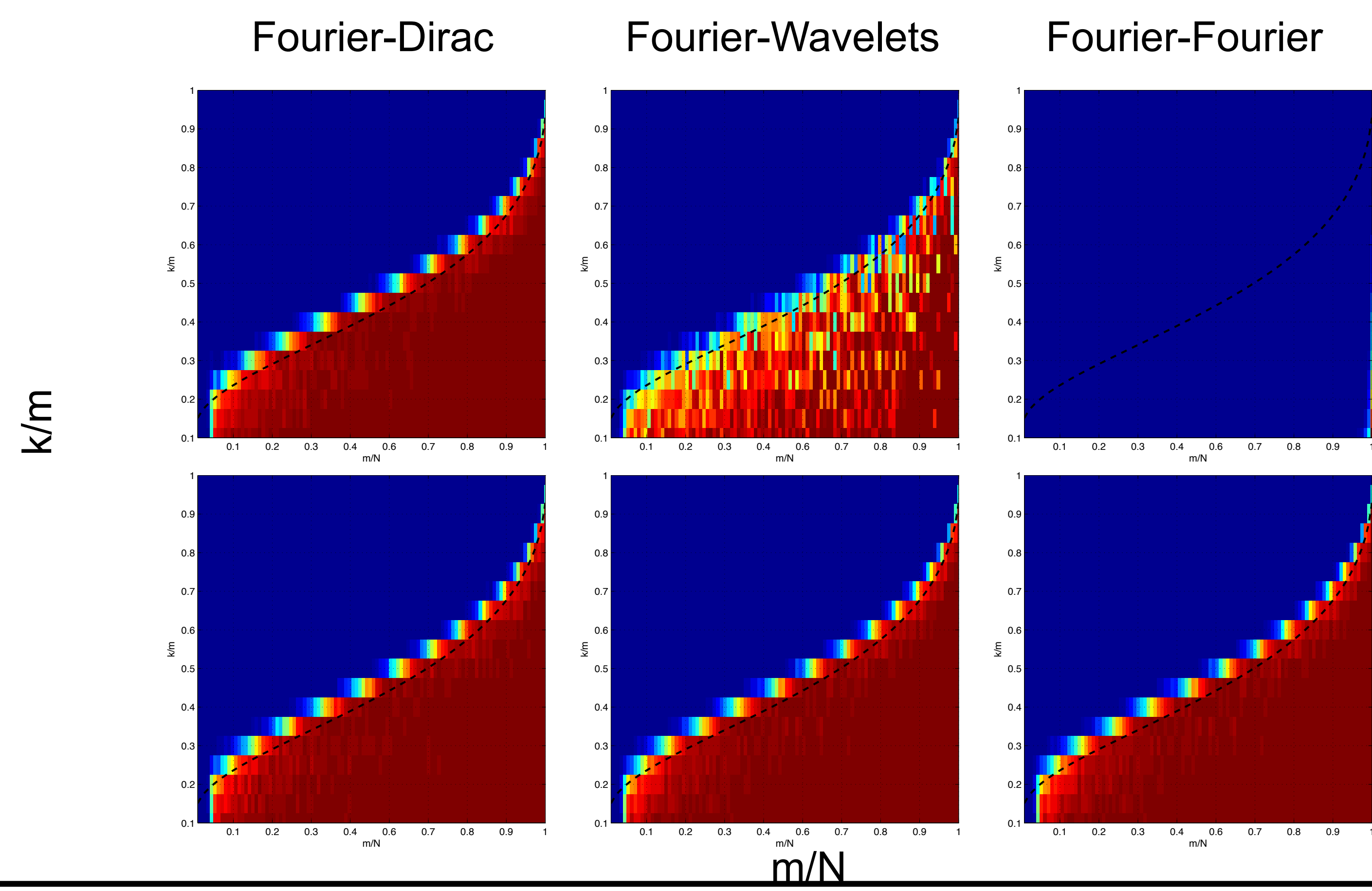
Hoeffding gives, with probability ϵ :

$$\sqrt{N}\mu\left(\mathsf{FC}, \Psi\right) > \sqrt{2\log\left(2N^2/\epsilon\right)}$$

One then shows that every K sparse vector can be recovered from:

$$M \geqslant c\, K \log^3\left(12N/\epsilon\right)$$

Independently of the sparsity basis !



Spread Spectrum in MRI

CS has already been applied to MR [Lustig, see also earlier talk in this workshop]

Here: Explore potential of spread-spectrum “conditioning”

$$\nu(\mathbf{k}) = \int_{\mathbb{R}^2} \rho(\mathbf{x}) \boxed{e^{i\pi w|\mathbf{x}|^2}} e^{-2i\pi \mathbf{k}\cdot\mathbf{x}} d^2\mathbf{x}$$

Phase Scrambling

- well-known in MRI (high Dynamic, reduce aliasing)
- obtained through dedicated coils or RF pulses

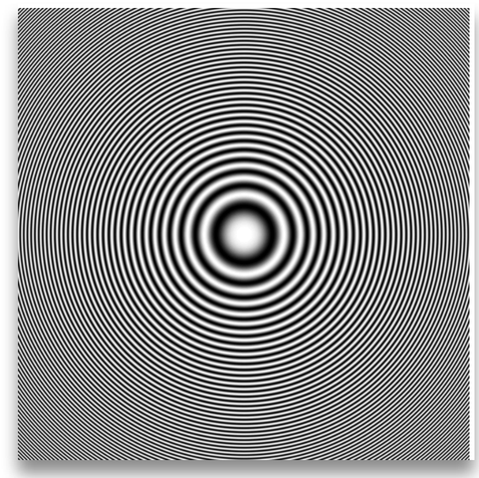
Measurement model:

$$\nu \equiv \phi^{(w)} \rho + \mathbf{n} \text{ with } \phi^{(w)} \equiv \mathbf{MFC}^{(w)}$$

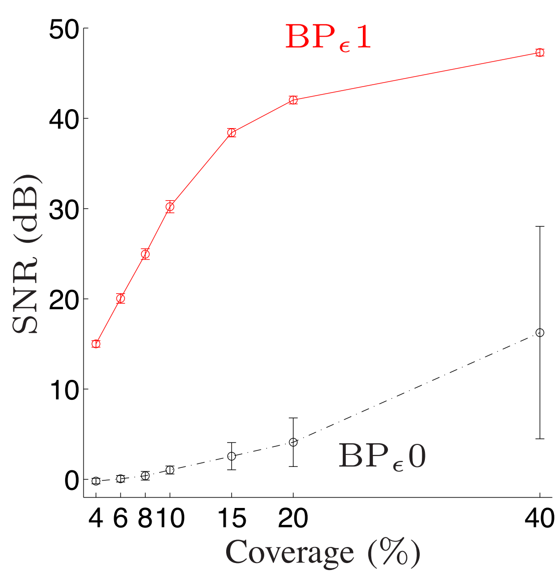
- ν: M/2 complex
- ρ: N-d real signal
- $\mathbf{M}$: Sub-sample
- $\mathbf{F}$: Fourier
- $\mathbf{C}^{(w)}$: Diagonal chirp matrix

Simulations

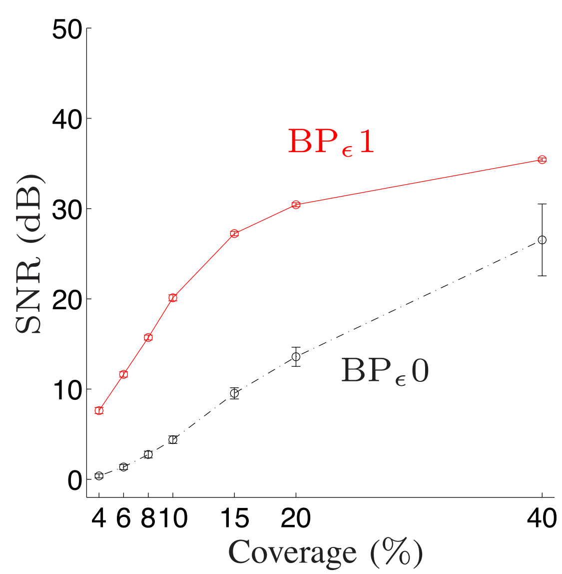
Shepp-Logan phantom and chirp



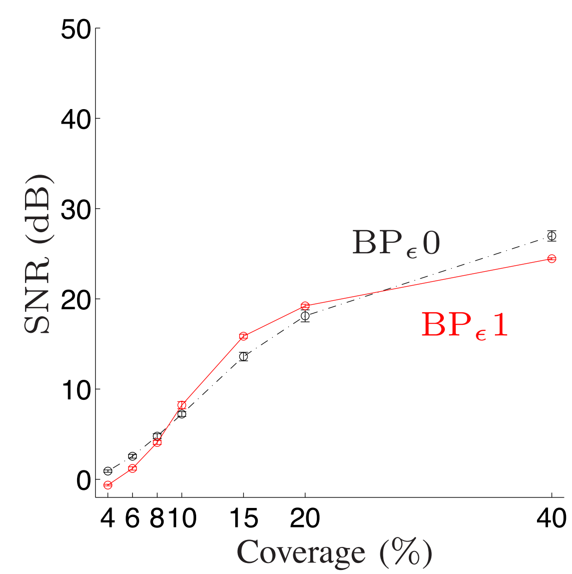
Input SNR = 30dB, sparsity basis = wavelets, 30 simulations



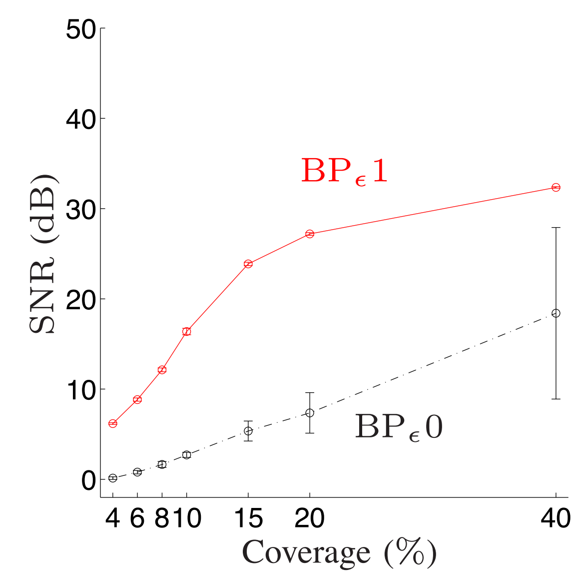
s=5



s=3



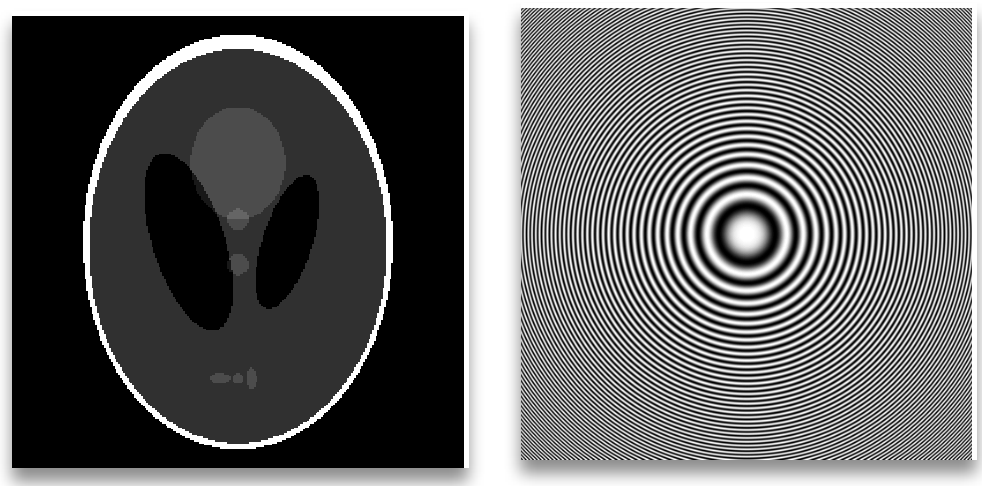
s=1



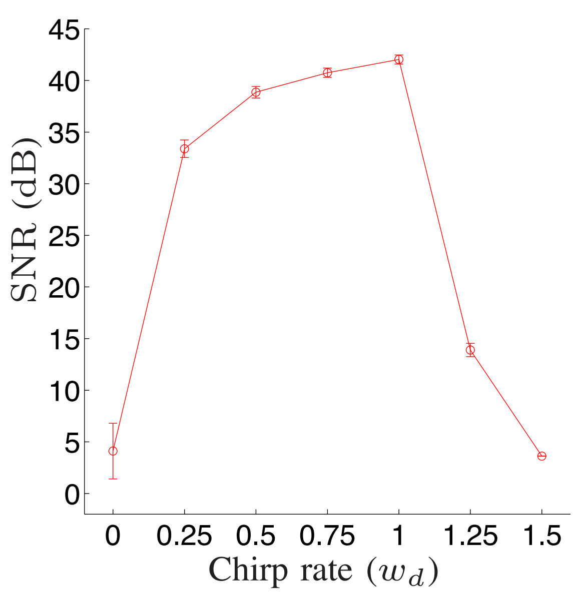
overall

Simulations - leakage

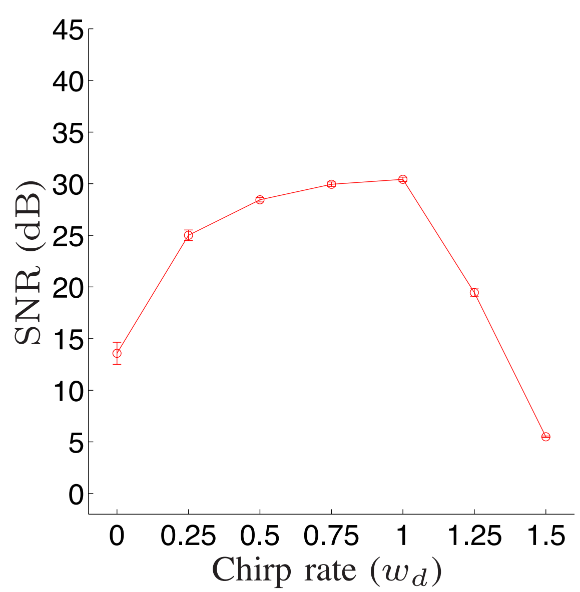
Shepp-Logan phantom and chirp



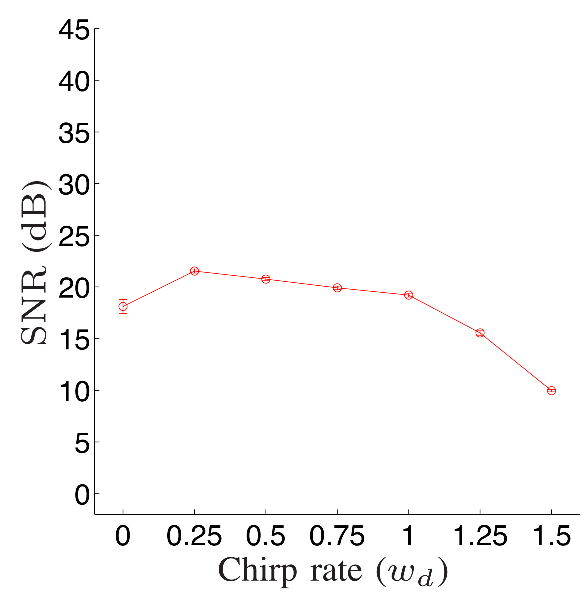
Input SNR = 30dB, coverage 20%, sparsity basis = wavelets, 30 simulations



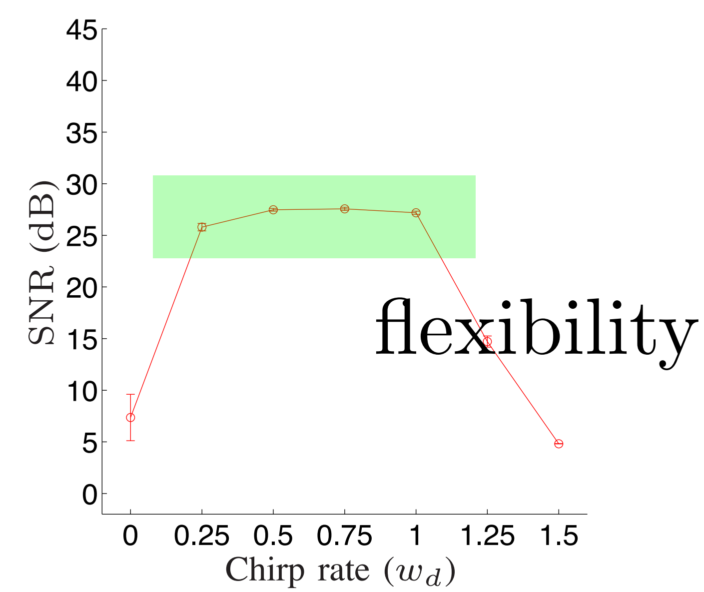
s=5



s=3



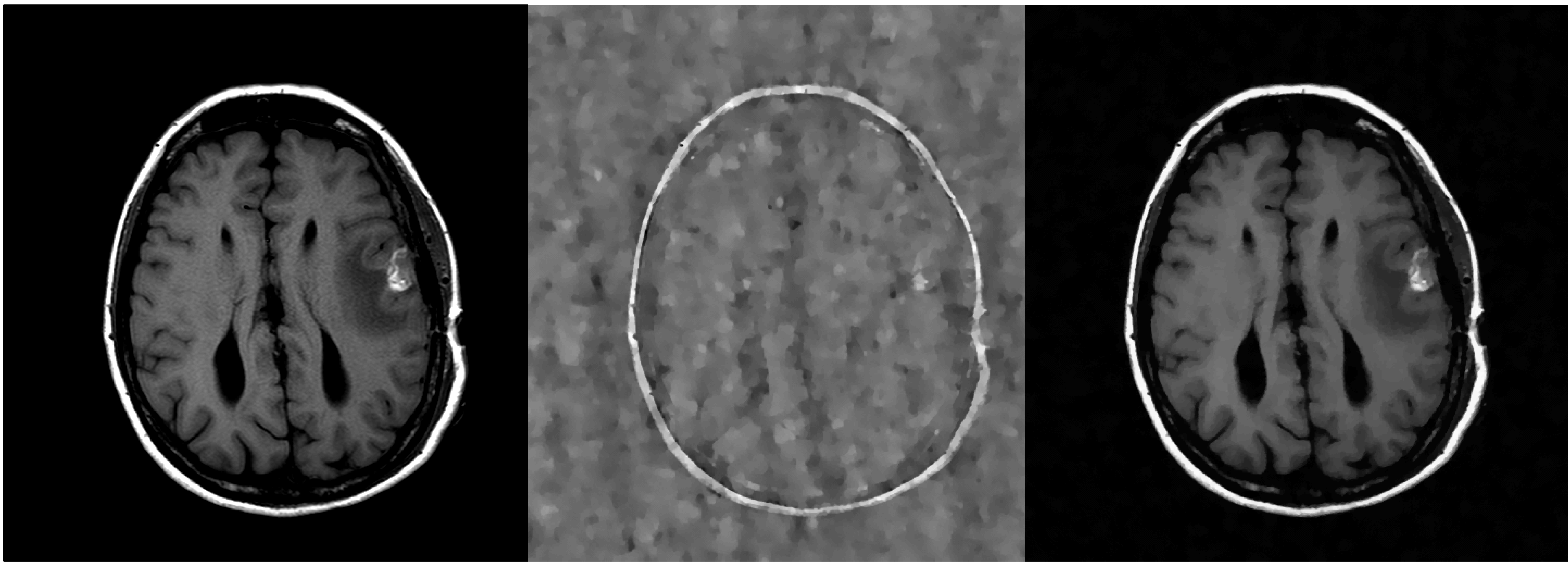
s=1



overall

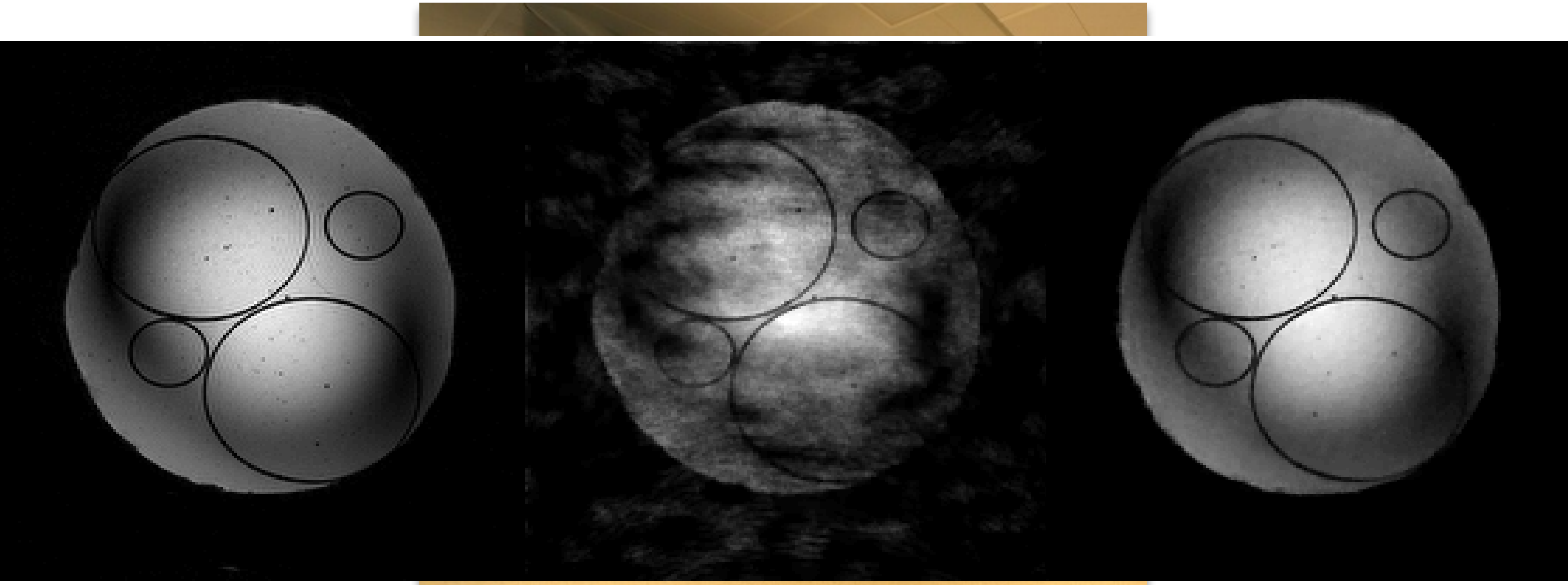
Compressed Sensing in MRI

Original Image **10% Fourier coverage**

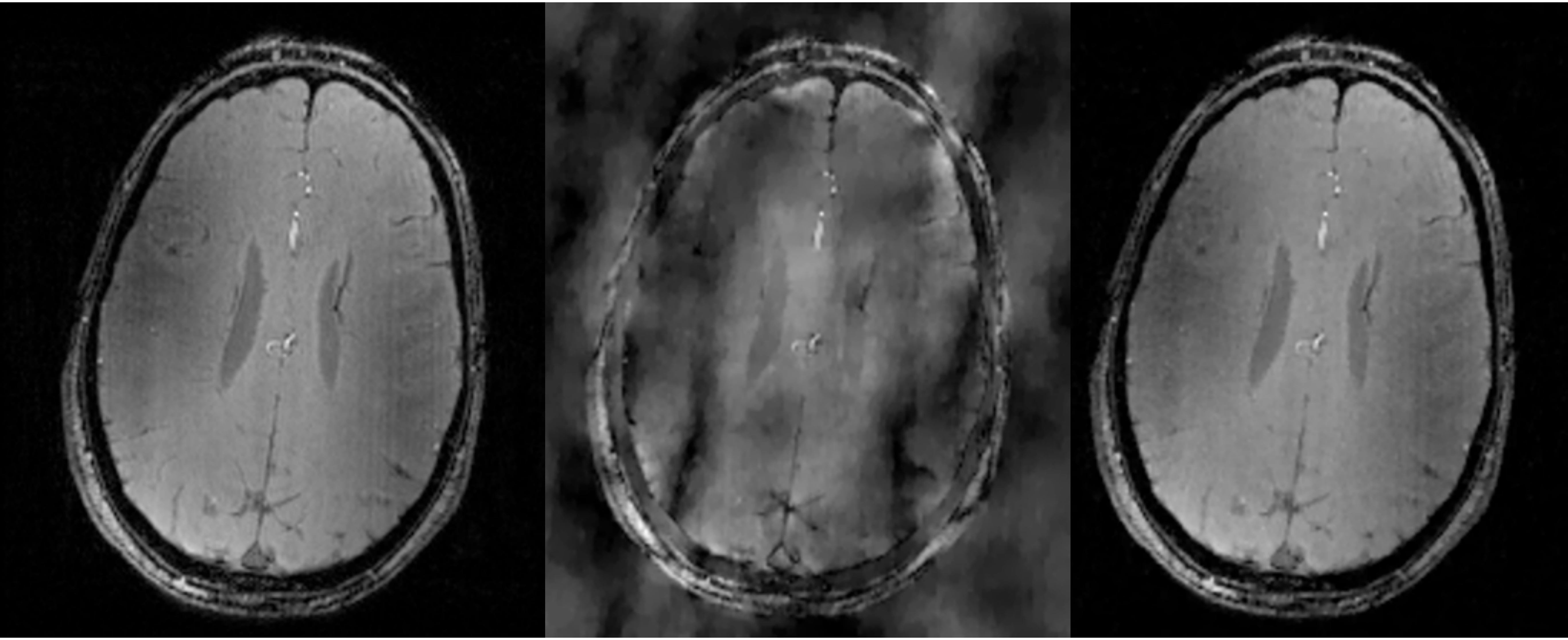


Compressed Sensing in MRI

Real data acquisition (+phantom "Marie"), 7T MRI@EPFL
chirp pre-modulation implemented with a dedicated shim coil



"Full scan" image   No pre-modulation   With chirp



Low-power ECG ambulatory system

Problem: sense and transmit ECG (possibly multi-lead) from a low power body-area network

Compression ? Surely if we transmit less, we will waste less power in communication.

Sure, but if we compress more we will waste energy using a complex encoder !

Can CS offer an interesting trade-off ?

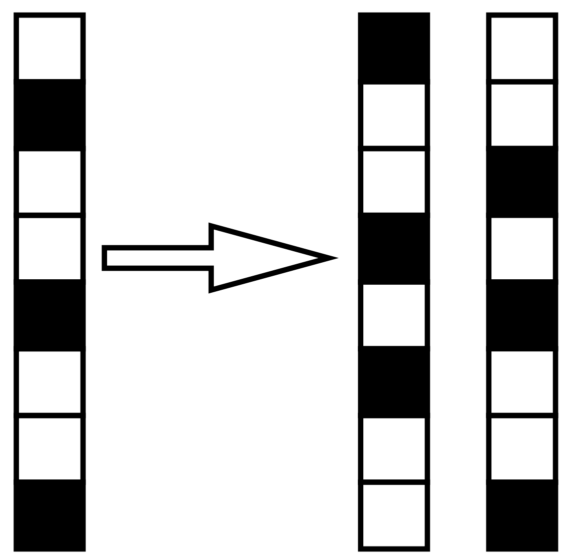
Can everything be real-time ?

- State-of-the-art
  - Wavelet transform, followed by thresholding, quantization and entropy coding
  - Pros: excellent compression results, signals nicely sparse (at least ventricular part)
  - Cons: Full wavelet transform must be implemented on the sensing node

What is a good sensing matrix for low-power sensing ?

Surely not gaussian ! (dense, complex to apply to signal and even complex to generate ...)

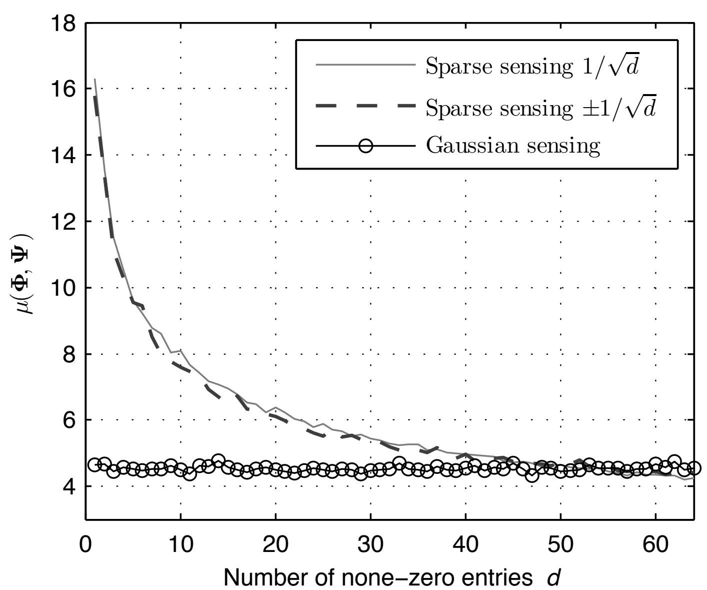
Sparse matrices, binary entries (ex: expander graphs)



generate binary vector
with d non-zero elements

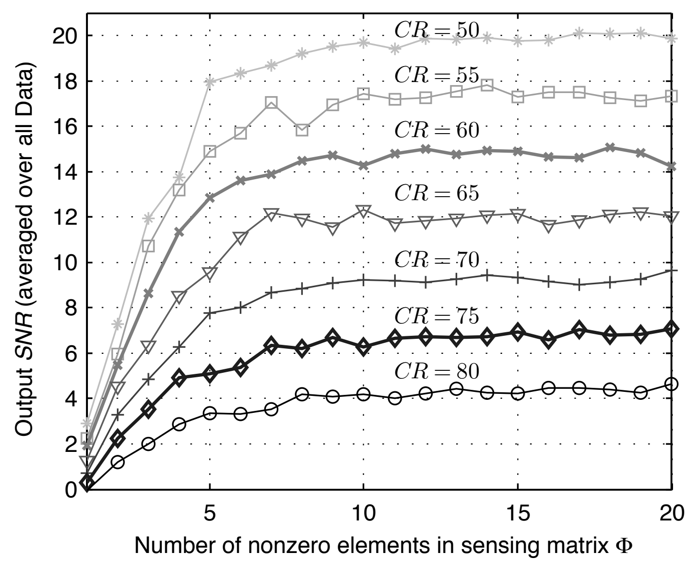
generate full matrix by
random permutations

Performance indexes for sensing mechanism

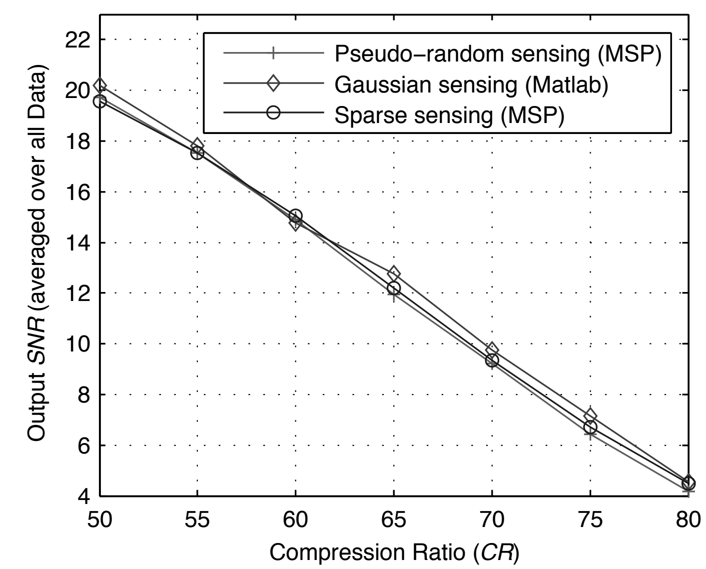


Mutual coherence $\mu(\Phi, \Psi)$ *vs.* d

coherence quickly approaches "optimal case"



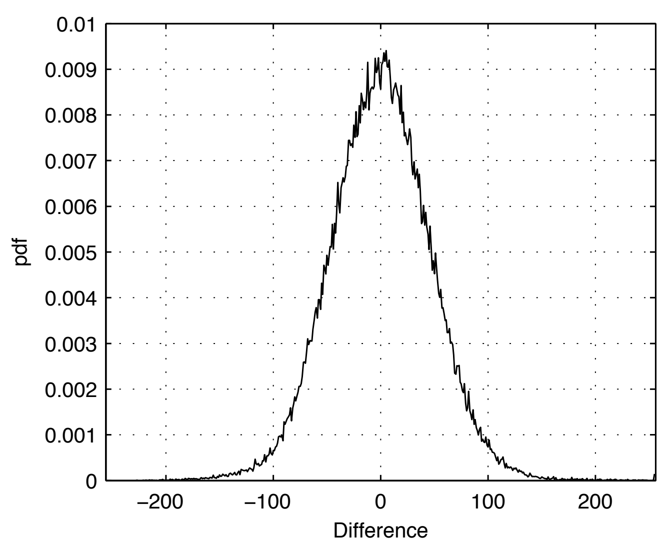
SNR saturates after d=12 non-zero elements



Hardly any difference between proposed sensing and gaussian sensing

2 s of signal are sensed in 82 ms

Coding: simple predictive scheme

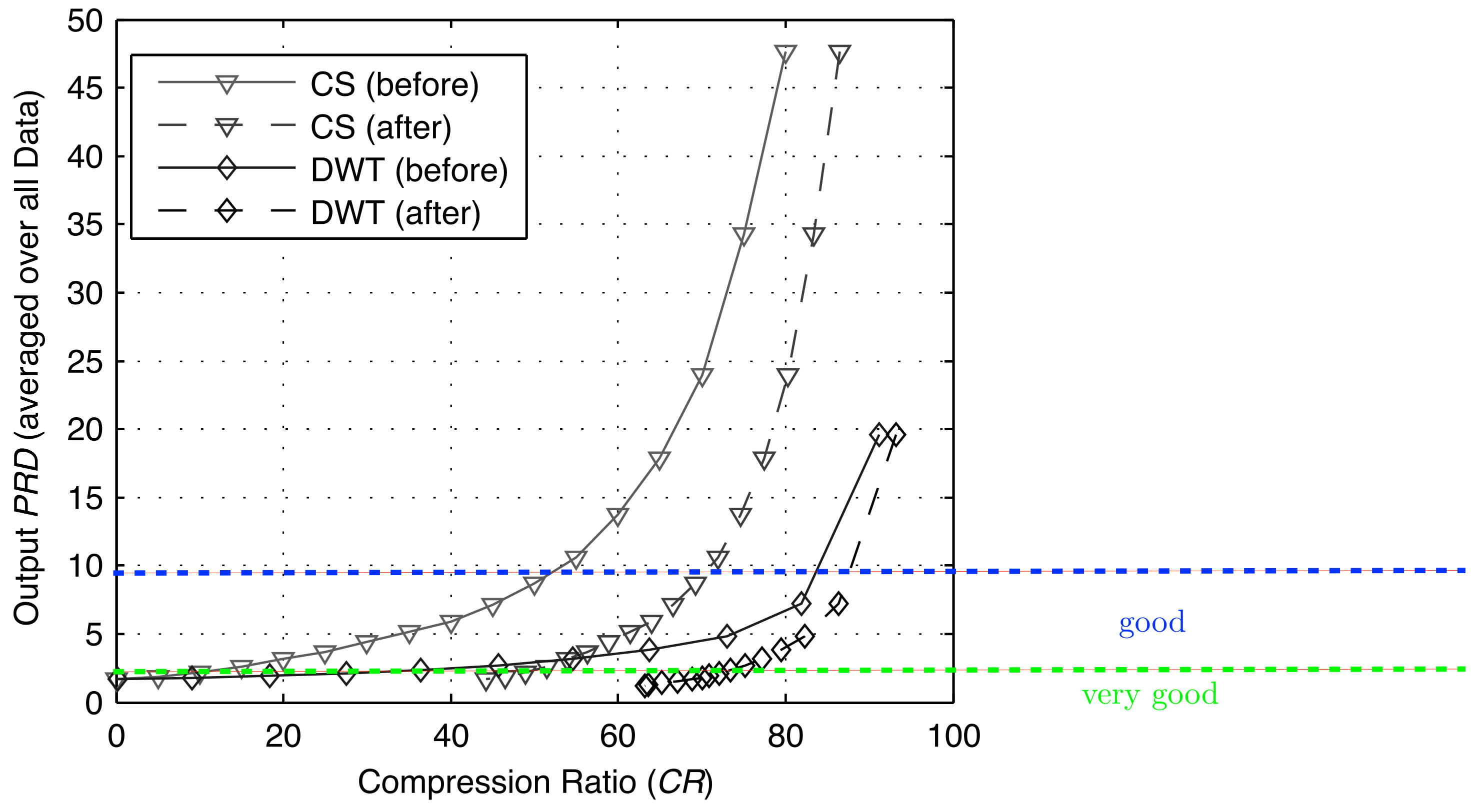


difference between successive sensing vectors
Compression Ration: 20%

Gaussian RD theory
9 bits quantizer
Huffman coding
1.5 kB codebook stored on platform

Total memory footprint of CS implementation:
6.5 kB of RAM for computations
7.5 kB of Flash

Comparisons - Quality vs Compression



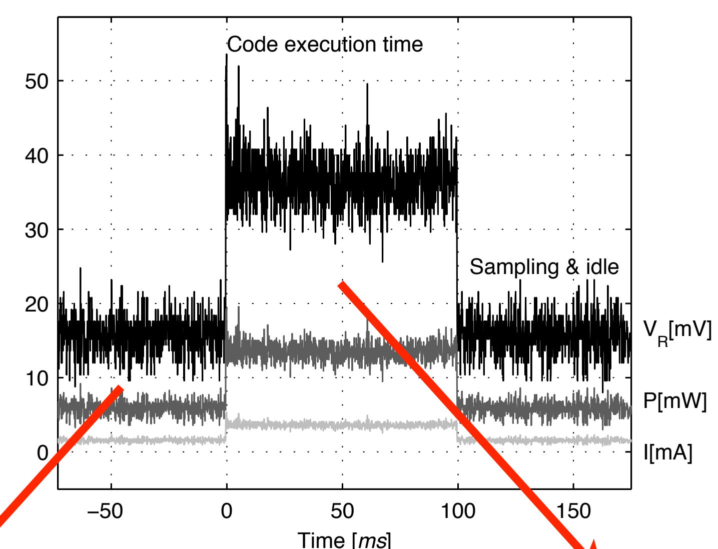
In terms of pure compression performance, an optimized DWT encoder is clearly (and obviously better) than the non-adaptive CS scheme

Comparisons - Power consumption

Consumption *measured* on the platform in real-life use

Code execution time: 95ms for CS, 580ms for DWT

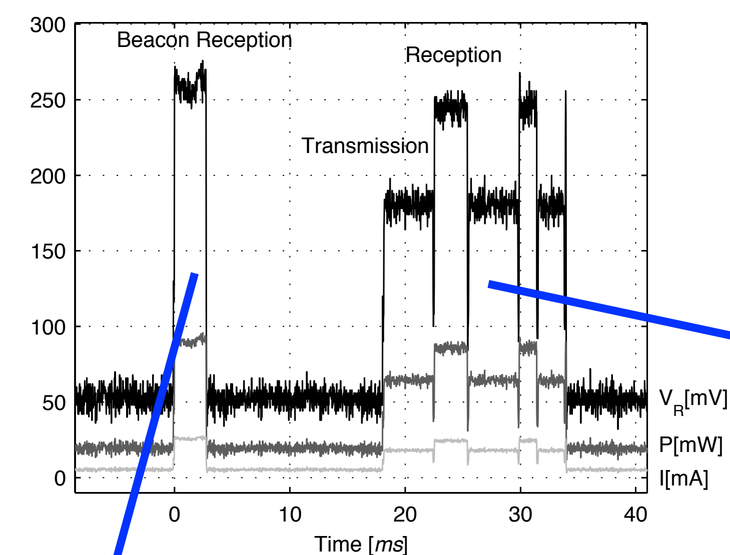
Node consumption



MSP430 idle and sampling till buffer full

CS code running

Radio consumption



periodic “ping”

transmission & reception

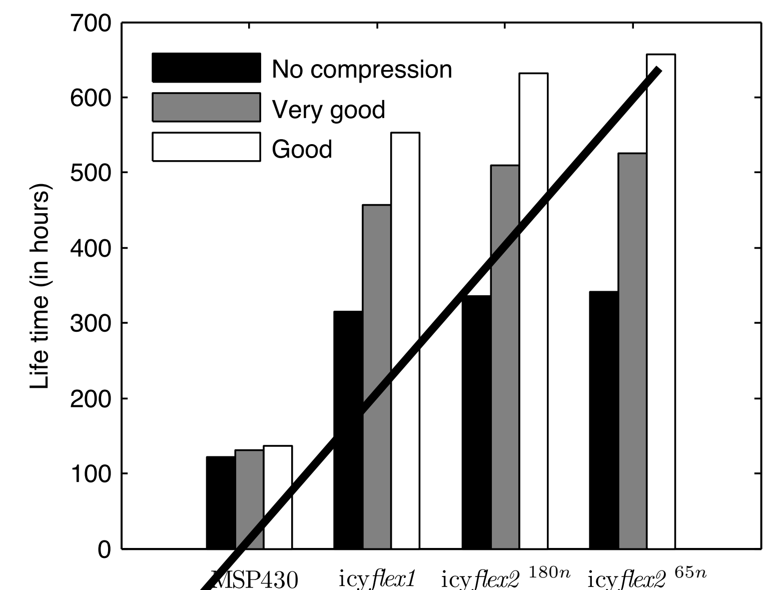
Note: contrary to what is usually assumed in litterature, optimized antennas are really low power !

Comparisons - Power consumption

Final results: it is important to know your architecture VERY well

NODE LIFE TIME FOR "VERY GOOD" RECONSTRUCTION QUALITY WITH EMBEDDED COMPRESSION

	DWT	CS	No Comp.
Compression Ratio (%)	73	51	0
Code execution time (ms)	580	99.5	0
Packet Ready every ... (ms)	1099.5	605.9	296.9
Beacon Interval (ms)	4398	2423	1187
Energy Consumption (mJ)	9.08	7.81	8.37
Life time (h) (280 mAh@3.7 V)	110	127.9	119.4



MSP430: ratio micro-controller/antenna is such that DWT compression does NOT improve on "raw" streaming ...
CS improvement rather small

State-of-the-art homegrown low-power microntrollers

92% lifetime extension, 6 times better than MSP430

Note: CS decoder runs real-time on a jailbroken iPhone 3GS

Conclusions & Outcome

- Many interesting applications in “niches”
  - *low power embedded systems (ex: in vivo bio-chips)*
  - *coded-aperture super-resolution*
  - *high-transmission cost, reduced computational power (ex: satellites)*
- Full system implementations are very sparse
  - *in theory, it’s easy but in practice one gets surprises*
  - *it takes more time to publish ...*
  - *interestingly multidisciplinar*