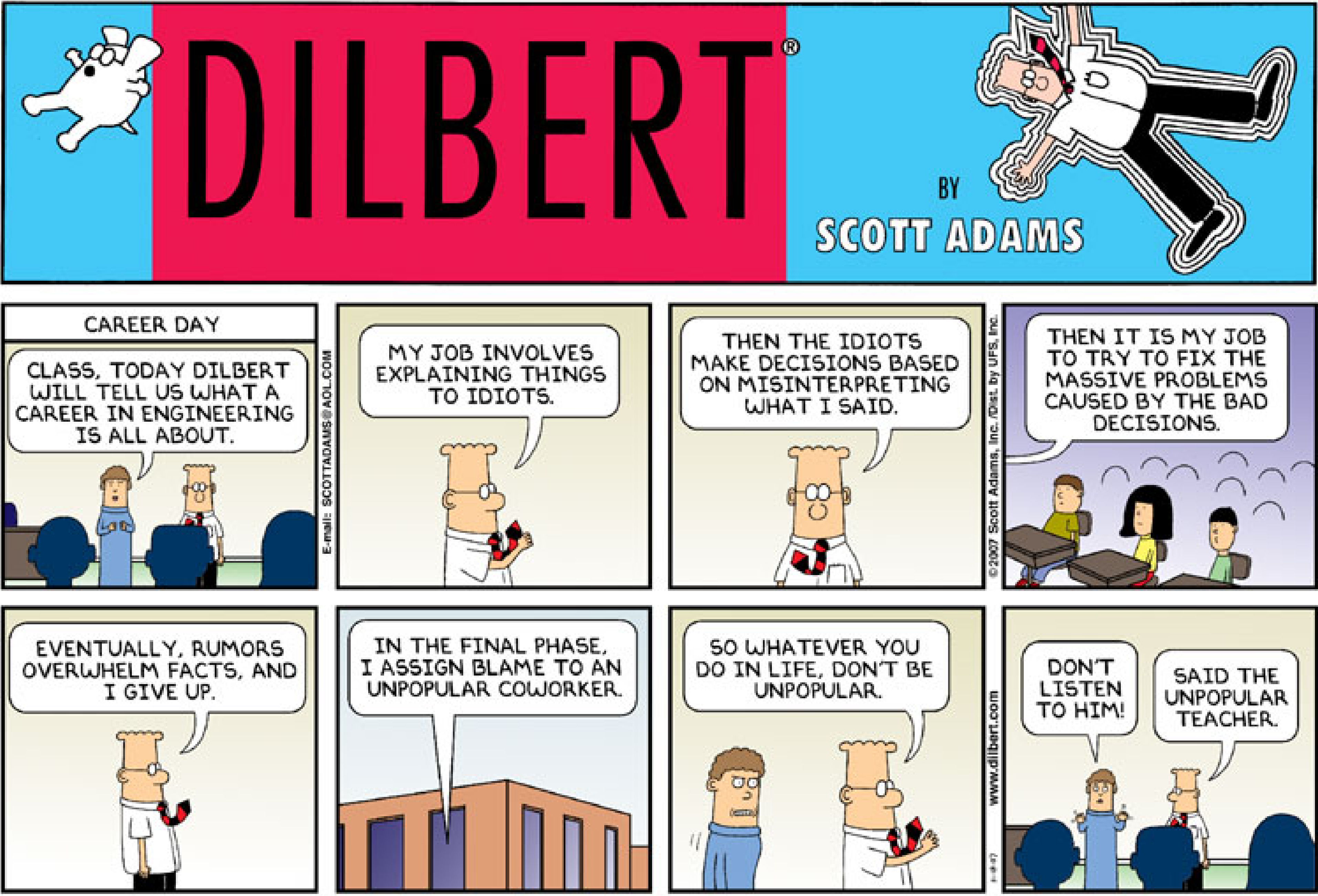

DILBERT®
BY SCOTT ADAMS
CAREER DAY
CLASS, TODAY DILBERT WILL TELL US WHAT A CAREER IN ENGINEERING IS ALL ABOUT.
E-mail: SCOTTADAMS@AOL.COM
MY JOB INVOLVES EXPLAINING THINGS TO IDIOTS.
THEN THE IDIOTS MAKE DECISIONS BASED ON MISINTERPRETING WHAT I SAID.
© 2007 Scott Adams, Inc./Dist. by UFS, Inc.
THEN IT IS MY JOB TO TRY TO FIX THE MASSIVE PROBLEMS CAUSED BY THE BAD DECISIONS.
EVENTUALLY, RUMORS OVERWHELM FACTS, AND I GIVE UP.
IN THE FINAL PHASE, I ASSIGN BLAME TO AN UNPOPULAR COWORKER.
SO WHATEVER YOU DO IN LIFE, DON'T BE UNPOPULAR.
www.dilbert.com
DON'T LISTEN TO HIM!
SAID THE UNPOPULAR TEACHER.

“Signals, Information and Sampling”

or

Some new directions in signal processing and communications

Steve McLaughlin

IDCOM, School of Engineering & Electronics

with thanks to

Mike Davies, Bernie Mulgrew, John Thompson

What are signals?

A **signal** is a time (or space) varying quantity that can carry information. The concept is broad, and hard to define precisely.

(*Wikipedia*)

What is signal processing?

Signal processing is the analysis, interpretation, and manipulation of signals.

Signals of interest include sound, images, biological signals such as ECG, radar signals, and many others. Processing of such signals includes storage and reconstruction, separation of information from noise (e.g., aircraft identification by radar), compression (e.g., image compression), and feature extraction (e.g., speech-to-text conversion).

(*Wikipedia*)

What is information?

Information theory is a branch of applied mathematics and engineering involving the quantification of information. Historically, information theory developed to find fundamental limits on compressing and reliably communicating data.

Since its inception it has broadened to find applications in statistical inference, networks other than communication networks, biology, quantum information theory, data analysis , and other areas, although it is still widely used in the study of communication.

(*Wikipedia*)

What is informatics?

Informatics includes the science of information, the practice of information processing, and the engineering of information systems. Informatics studies the structure, behaviour, and interactions of natural and artificial systems that store, process and communicate information.

Informatics is broader in scope than information theory.

(*Wikipedia*)

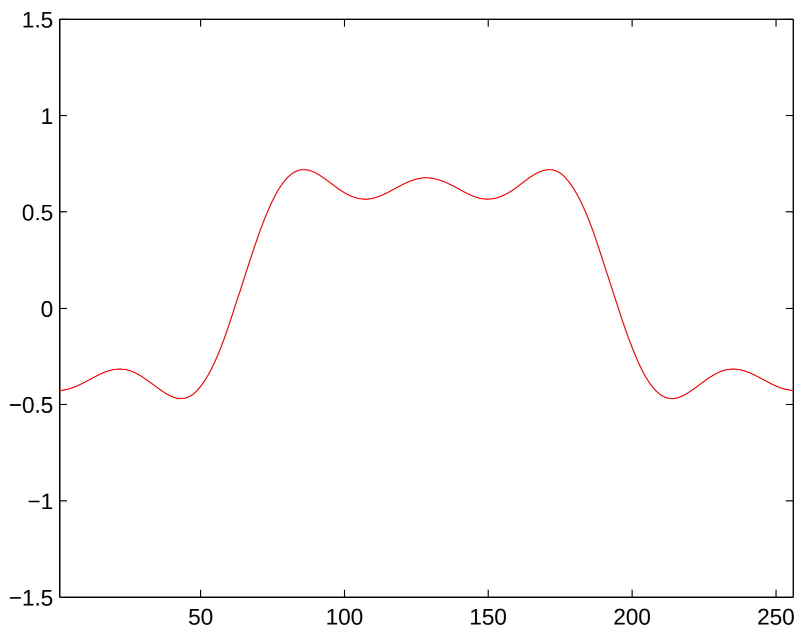
Part I: Sparse Representations and Coding

What are signals made of?

The Frequency viewpoint (Fourier):

Joseph Fourier

Signals can be built from the sum of harmonic functions (sine waves)



$$x(t) = \sum_k c_k \phi_k(t)$$

signal

Fourier coefficients

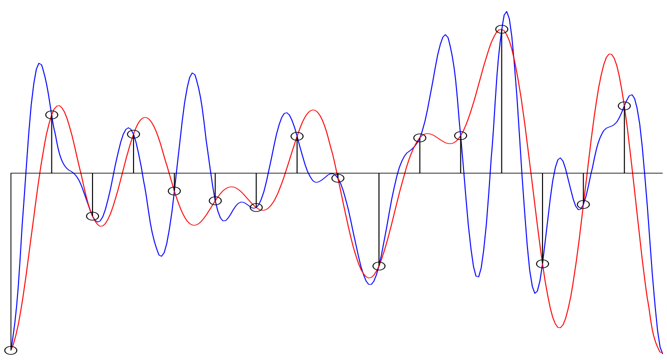
Harmonic functions

Sampling and the digital revolution

Today we are more familiar with discrete signals (e.g. audio files, digital images). This is thanks to:

Whittaker–Kotelnikov–Shannon Sampling Theorem:

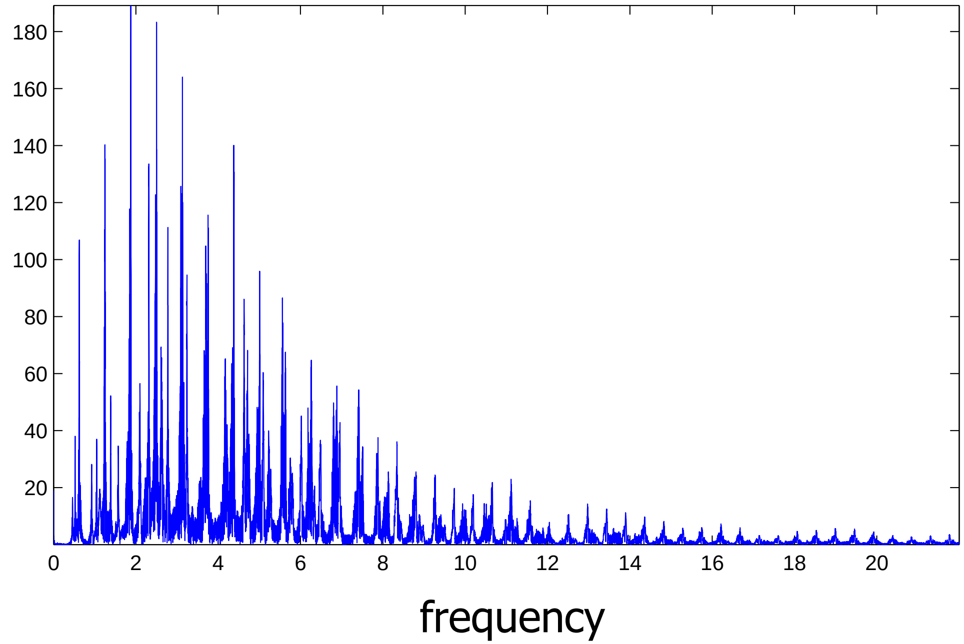
> "Exact reconstruction of a continuous-time signal from discrete samples is possible if the signal is bandlimited and the sampling frequency is greater than twice the signal bandwidth "



Sampling below this rate introduces aliasing

Audio representations

Which representation is best: time or frequency?

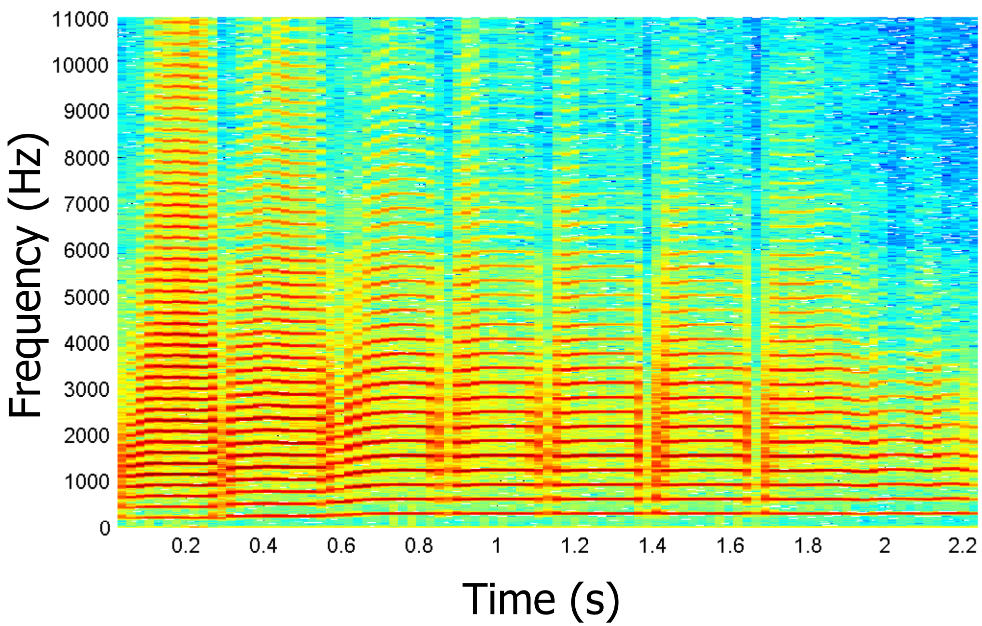


Audio representations

Time <u>and</u> Frequency (Gabor)

"Theory of Communication," J. IEE (London) , 1946

"… a new method of analysing signals is presented in which time and frequency play symmetrical parts…"



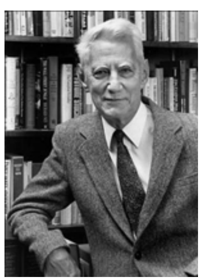
Gabor and audio coding

Time and Frequency (Gabor)

"Theory of Communication," J. IEE (London) , 1946

"…In Part 3, suggestions are discussed for compressed transmission and reproduction of speech or music…"

Modern audio coders owe as much to Gabor's notion of time-frequency analysis as it does to Shannon's paper of a similar title, two years later, that heralded the birth of information and coding theory.



"A Mathematical Theory of Communication," Bell System Technical Journal, 1948.

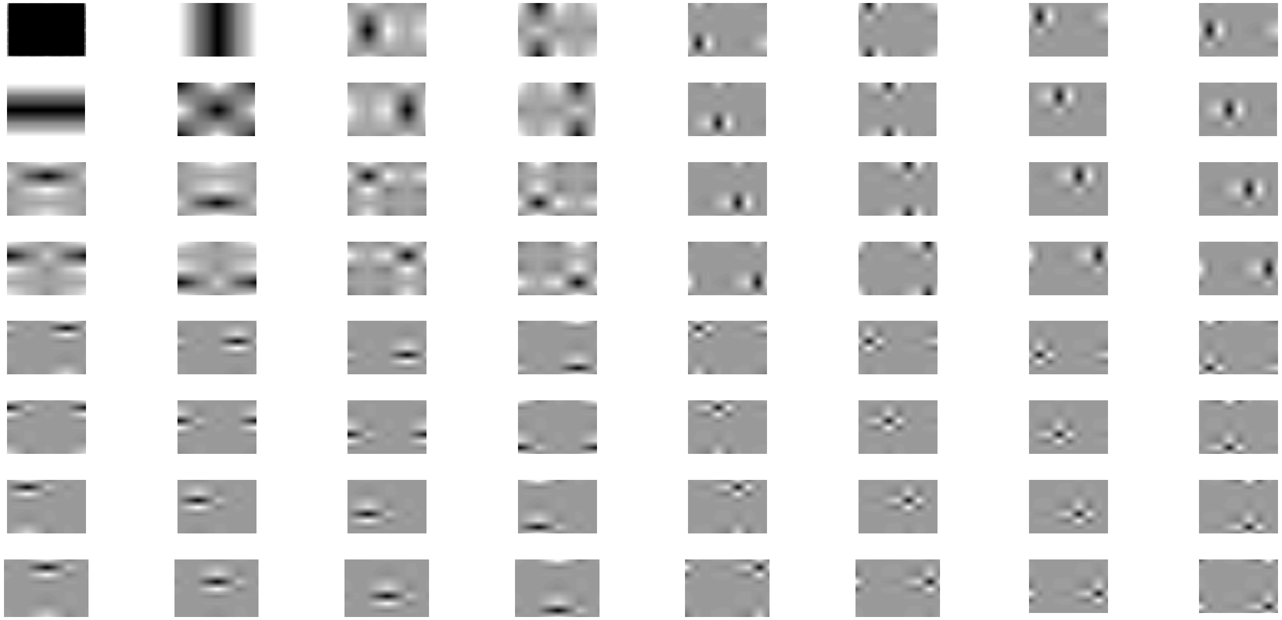
C. E. Shannon

Image representations

… Space and Scale: the wavelet viewpoint:

"Daubechies, Ten Lectures on Wavelets," **SIAM 1992**

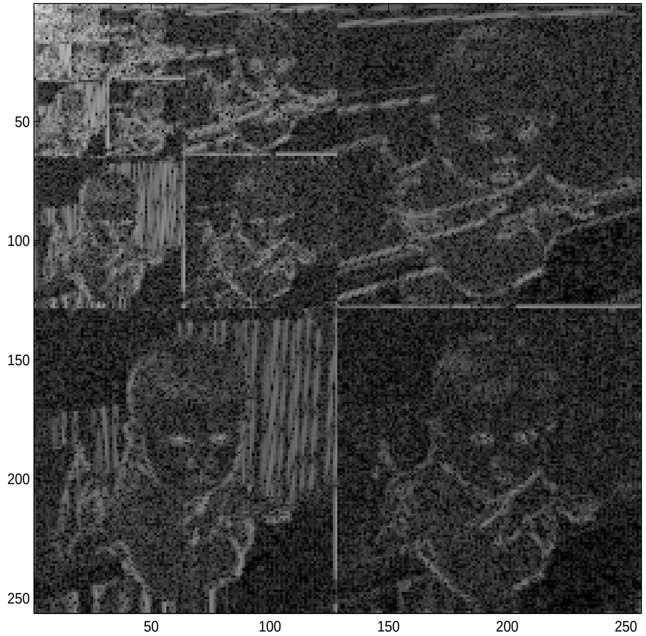
Images can be built of sums of *wavelets*. These are multi-resolution edge-like (image) functions.



Transform Sparsity

What makes a good transform?

"TOM" image



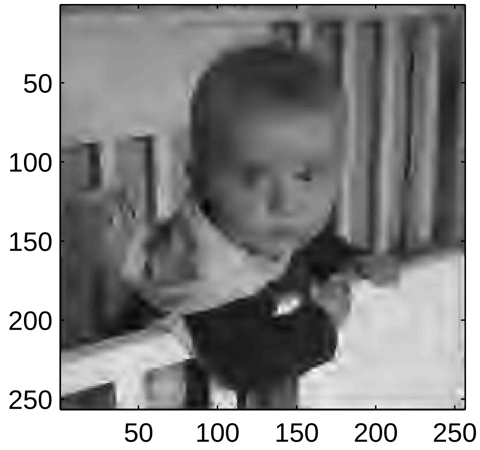
Wavelet Domain

Good representations are efficient - Sparse!

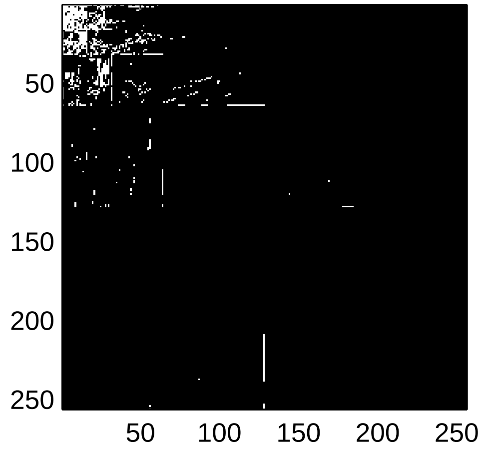
Coding signals of interest

What is the difference between quantizing a signal/image in the transform domain rather than the signal domain?

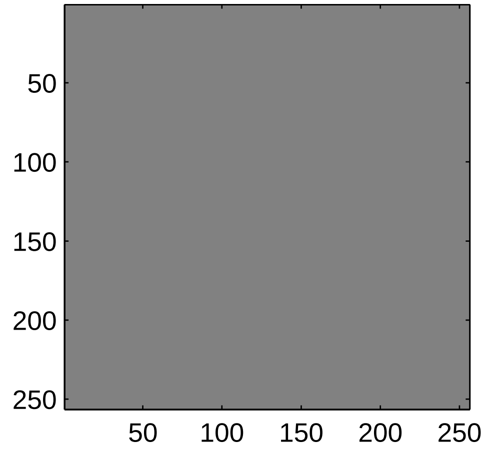
Compressed to 0.1 bits per pixel



Quantization in wavelet domain



Tom's nonzero wavelet coefficients



Quantization in pixel domain

Coding signals of interest

An important question is: what are the signals of interest?

If we digitize (via sampling) each signal is a point in a high dimensional vector space. e.g. a 5 Mega pixel camera image lives in a **5,000,000** dimensional space. What is a good signal model?

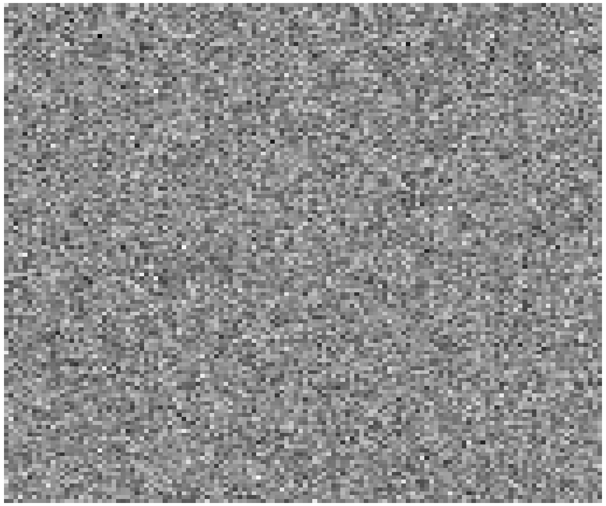
Geometric Model I

Consider the set of finite energy signals: the L_2 ball (an *n-sphere*).

Coding can be done by covering the set with ε-balls.

The L_2 ball is <u>NOT</u> a good signal model!

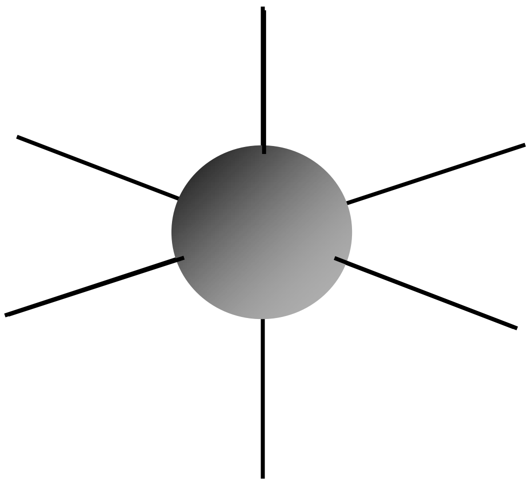
Almost all signals look like this…



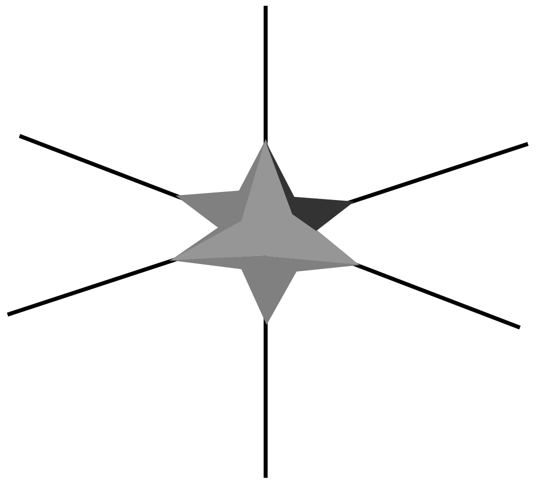
Coding signals of interest

Efficient transform domain representations implies that our signals of interest live in a much smaller set.

These sets can be covered with much fewer ε-balls and require much fewer 'bits' to approximate.



L_2 ball (not sparse)



Sparse signal model

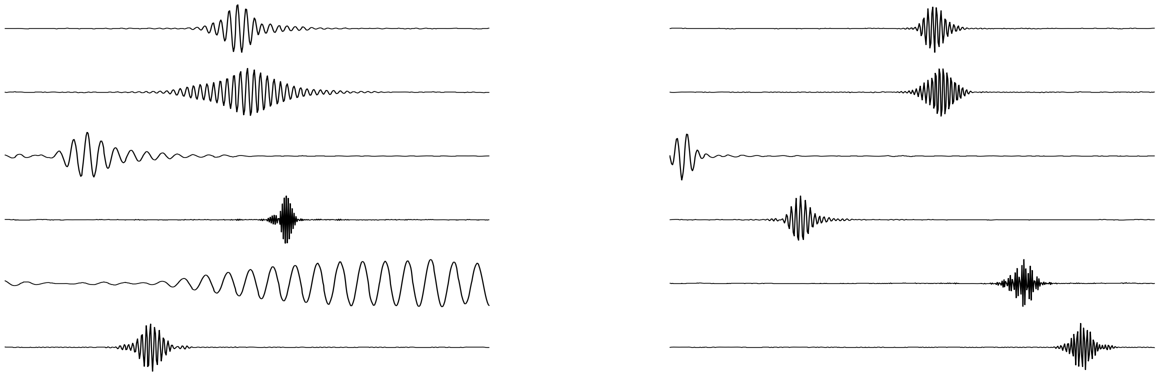
Learning better representations

Recent efforts have been targeted at *learn* better representations for a given set of signals, *x(t)*:

$$x(t) = \sum_k c_k \varphi_k(t)$$

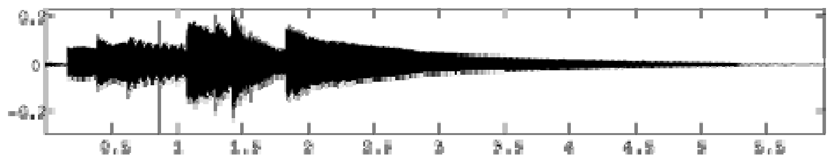
That is, learn *dictionaries* of functions $\varphi_k(t)$ that represent signals of interest with only a small number of significant coefficients, c_k.

For Audio (Abdallah & Plumbley, *Proc. ICA* 2001):



Build bigger dictionaries

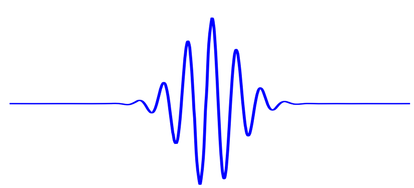
Another approach is to try to build bigger dictionaries to provide more flexible descriptions. Consider the following test signal:

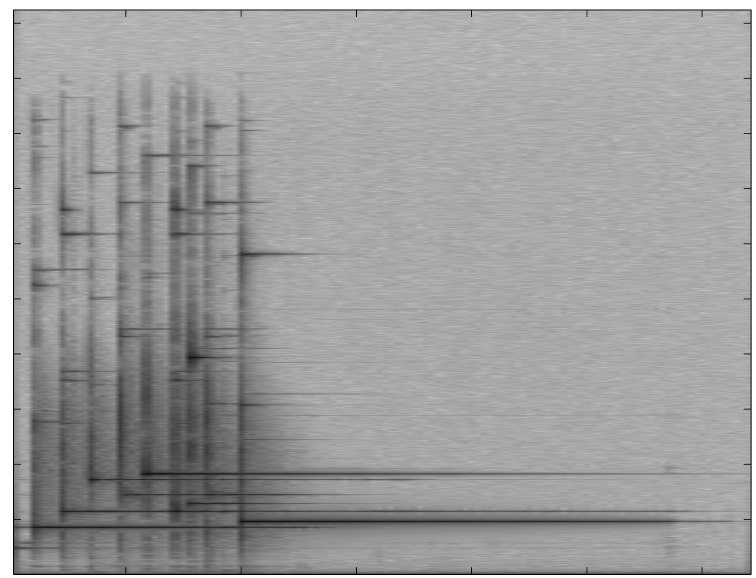


Heisenberg’s uncertainty principle implies that a Time-Frequency analysis has:

either good time resolution and poor frequency resolution

or good frequency resolution and poor time resolution

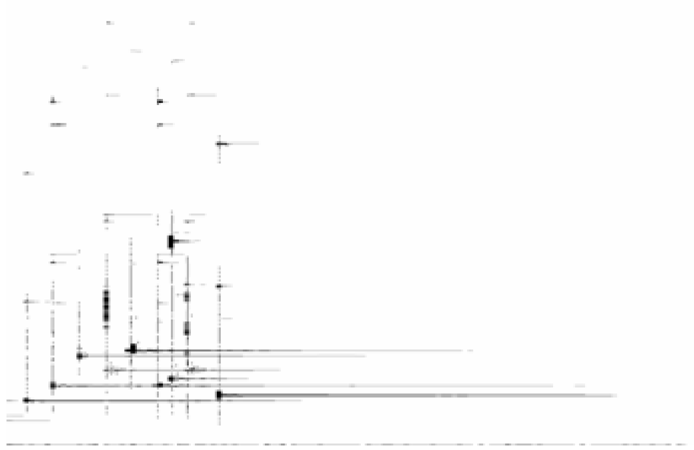




Multi-resolution representations

Heisenberg only applies to time-frequency **analysis** NOT time-frequency **synthesis**. Consider a TF synthesis representation with a combination of long (40 msec.) atoms and short (5 msec.) atoms.

Finding the sparse coefficients is now a nonlinear (and potentially expensive) operation.

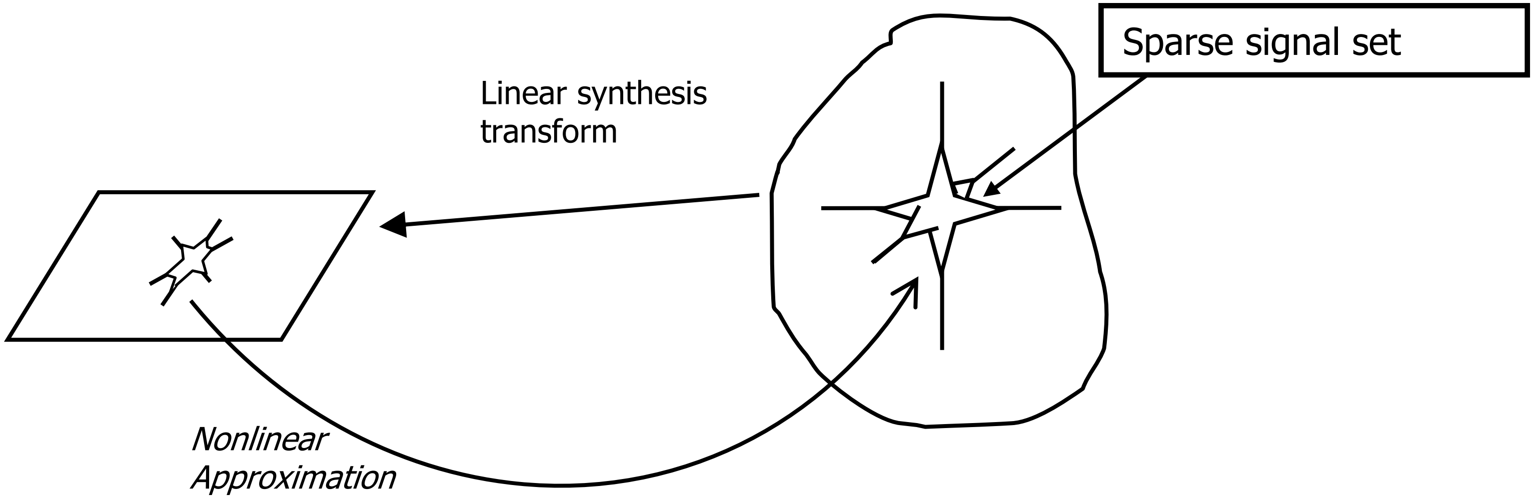


Combined representation

New uncertainty principles for sparse representations

Analysis versus synthesis

For invertible representations (e.g. wavelets) Analysis is equivalent to Synthesis. For redundant representations they are not:



Sparse approximation in redundant dictionaries requires a **nonlinear** operation.

This may require an exhaustive search of all possibilities (not practical).

So currently we use ‘greedy’ iterative methods.

Part I Review

Sparse Representations and Coding

- How we represent signals is very important;
- Sparse representations provide good compression;
- Recent efforts have targeted bigger and better representations;
- Despite the linear representations nonlinear approximations play an important role.

Part II: Sparse Representations and Sampling (compressed sensing)

Compressed sensing

Traditionally when compressing a signal (e.g. speech or images) we take lots of samples (sampling theorem) move to a transform domain and then throw most of the coefficients away!

Why can't we just sample signals at the "Information Rate"?

This is the philosophy of Compressed Sensing

E. Candès, J. Romberg, and T. Tao, **"Robust Ucertainty principles: Exact signal reconstruction from highly incomplete frequency information,"** IEEE Trans. Information Theory, 2006

D. Donoho, **"Compressed sensing,"** IEEE Trans. Information Theory, 2006

Compressed sensing

The Compressed Sensing principle:

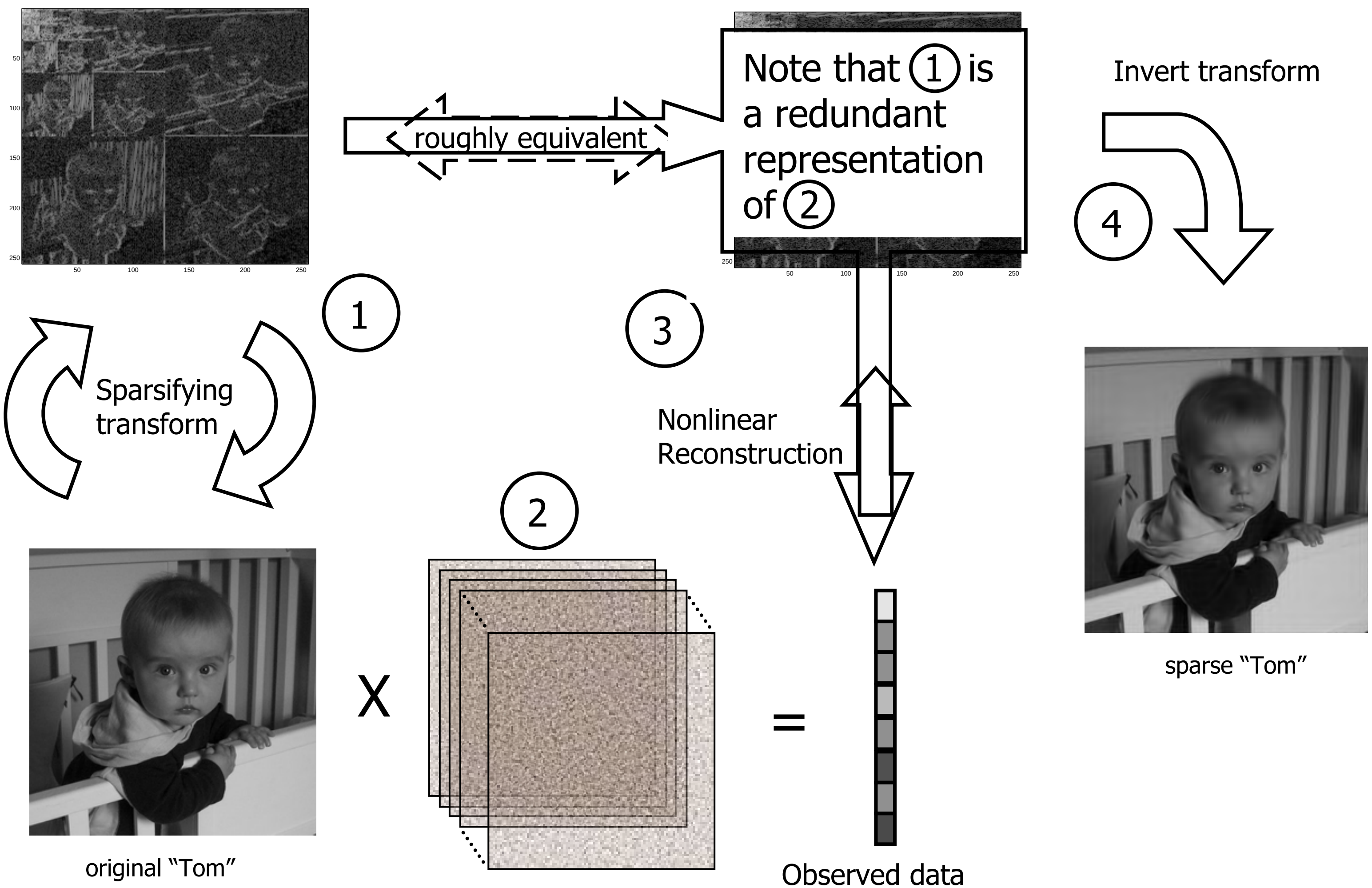
1. Take a *small* number of linear observations of a signal (number of observations << number of samples/pixels)
2. Use *nonlinear reconstruction* to estimate the signal via a transform domain in which the signal is sparse

Theoretical results

We can achieve an *equivalent approximation performance* to using the *M* most significant coefficients for an signal/image (in a sparse domain) by a fixed number of non-adaptive linear observations as long as:

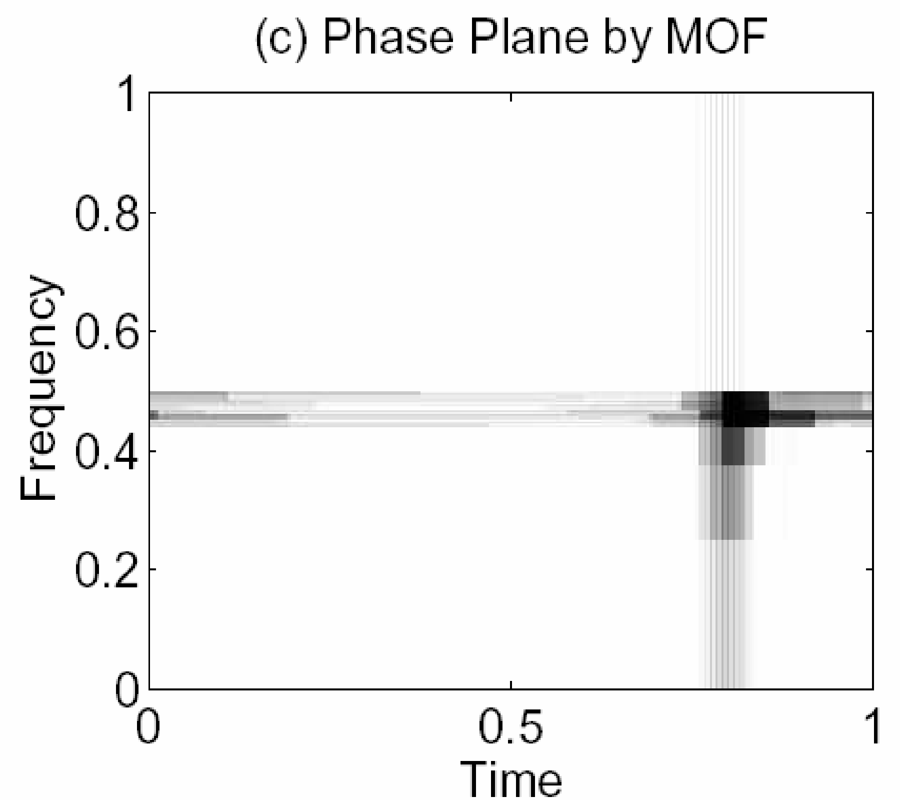
- *No. of observations ~ M x log N* (N, the size of full signal) and
- *for almost all (random) observations and*
- *can be achieved with practical reconstruction algorithms*

Compressed sensing principle



Sparse signal representations

Linear *analysis* transform coefficients (Method of Frames) are generally not sparse.

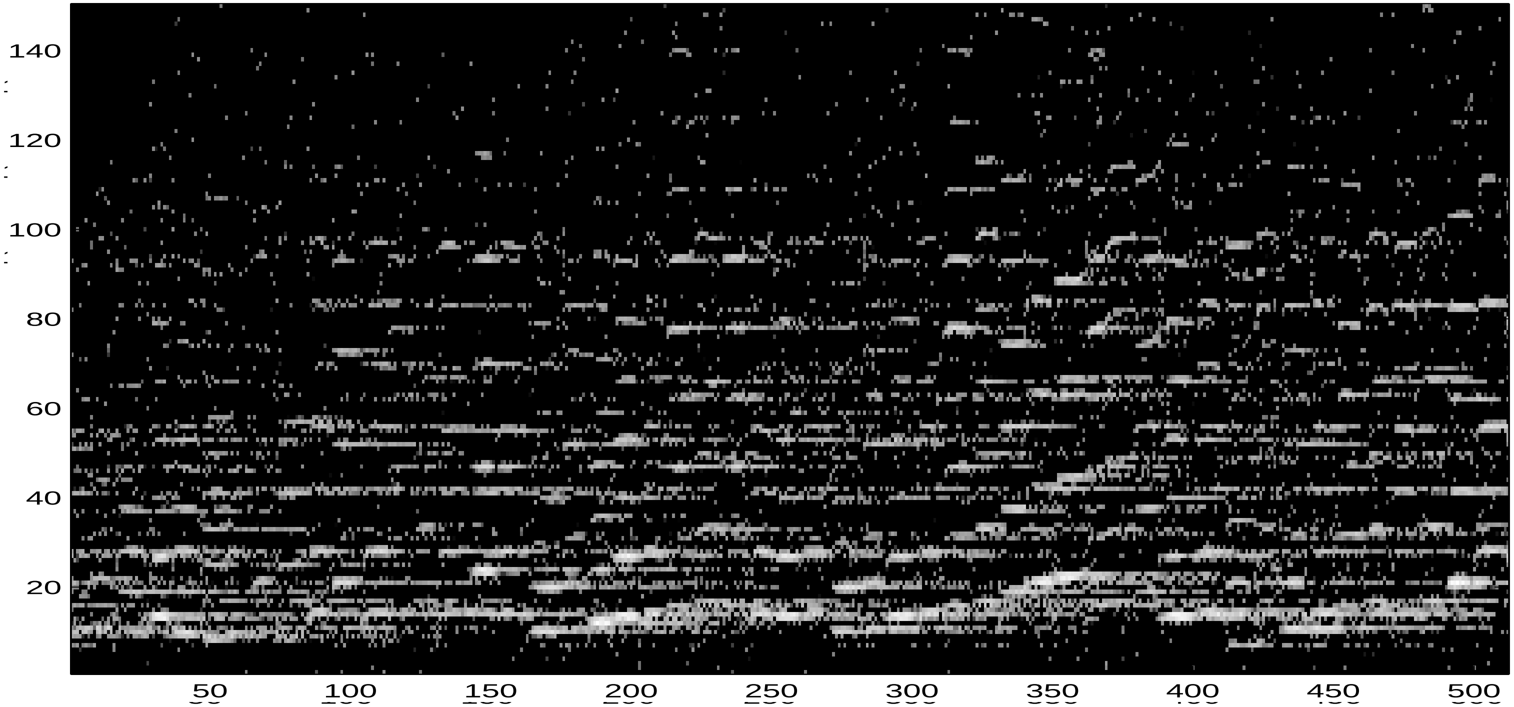


from Chen & Donoho 1995

Instead use sparse linear *synthesis* transform and invert – using various nonlinear methods

Sparse Subband representations

If we consider a oversampled subband analysis/synthesis model (e.g. STFT) then we ideally want to go from this to… something like this



Part III: Open Research Problems in Compressed Sensing

Sparse signal representations

Let $\Phi \in \mathbb{C}^{N \times M}$ define our over-complete (redundant) dictionary (M > N).

We want an *approximate* over-complete representation:

$$\Phi s = x + e$$

such that standard least squares term mall approximation enforces sparsity

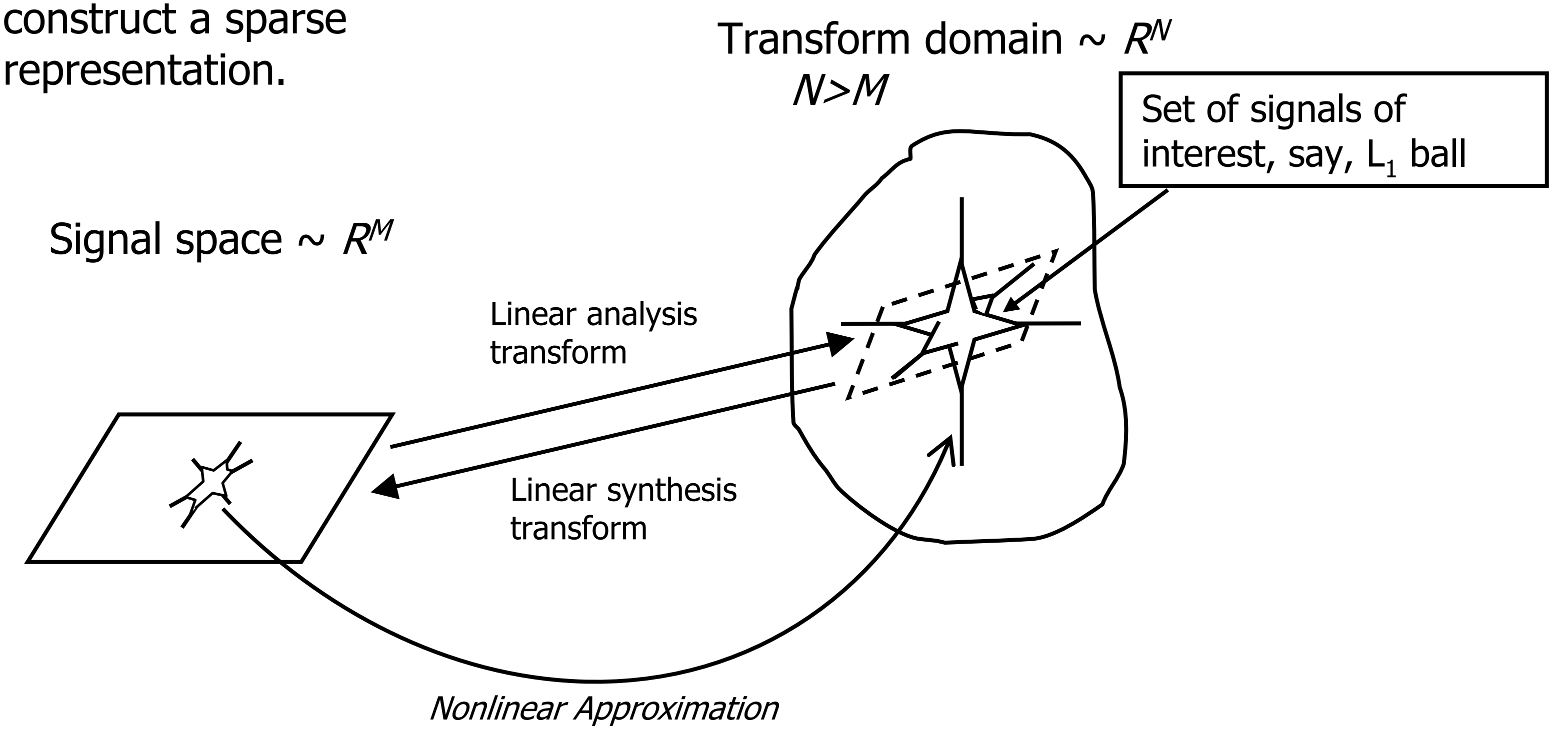
One appr zed least squares p Basis Pursuit De-Noising' – Chen *et al* 1995)

$$s = \arg\min_s ||x - \Phi s||_2^2 + \sum_n f(s_n)$$

Direct solution can be computationally expensive.

Overcomplete Sparse Approximations

Sparse Approximation uses a nonlinear Approximation to construct a sparse representation.



Problems of Interest

Sparse Approximations

1. Overcomplete dictionary design - how overcomplete should/can a dictionary be, particularly when constrained for example to represent Time-Frequency tiling?

2. How efficient can overcomplete representations be for coding?

3. Provably good algorithms for finding optimal, or near optimal, sparse representations.

   We know L_1 regularization and we know “Greedy” algorithms work under certain conditions. Also we know that the general problem is NP-hard. Is there a gap to be filled? For example, empirically stochastic search techniques work well, such as MCMC, however as far as I know there are no provable results for these within a given complexity.

4. relationship between sparse redundant representations, compressed sensing and super-resolution.

What are good observation matrices?

Current algorithms are provably good when a *Restricted Isometry Property* holds…

Candès, Romberg, and Tao [2] show that the critical property needed for the matrices Φ is a form of restricted isometry. Given a matrix Φ and any set T of column indices, we denote by Φ_T the $n \times \#(T)$ matrix composed of these columns. Similarly for a vector $x \in \mathbb{R}^N$, we denote by x_T the vector obtained by retaining only the entries in x corresponding to the column indices T. We say that the matrix Φ has the *Restricted Isometry Property* (RIP) of order k if there exists a $0 < \delta_k < 1$ such that

$$(1-\delta_k)\|x_T\|_{\ell_2}^2 \leq \|\Phi_T x_T\|_{\ell_2}^2 \leq (1+\delta_k)\|x_T\|_{\ell_2}^2, \quad \#T \leq k. \tag{3.9}$$

For $\delta_k < ½$ it has been shown that the following linear programme provides exact reconstruction (for k-exact sparse signals):

$$(\mathrm{P}_1) \qquad \min_{\tilde{x} \in \mathbb{R}^N} \|\tilde{x}\|_{\ell_1} \quad \text{subject to} \quad \Phi\tilde{x} = y.$$

and random matrices have been shown to satisfy this when:

$$S = O(K/\log(N/K)).$$

Problems of Interest

Compressed Sensing

- What are good observation matrices (currently only random is provably good). How do we go about designing good CS observation matrices, particularly when there may be constraints on the form of the observation (as in MRI).
- provably good decoding/approximation schemes for Compressed Sensing/Sparse Approximation (same as for Sparse Approximations)
- Faster good reconstruction algorithms.
- Extensions of CS theory beyond sparsity

Part IV: Applications (Measurements and Bits)

Compressed sensing applications

Compressed Sensing provides a new way of thinking about signal acquisition.

Applications areas include:

- Medical imaging
- Distributed sensing
- Remote sensing
- Very fast analogue to digital conversion (DARPA A2I research program)

Still many unanswered questions… Coding efficiency? Restricted observation domains? Etc.

Compressed Sensing Hallmarks

Compressed Sensing changes the rules of data acquisition game by exploiting a-priori signal *sparsity* information (signal is compressible)

- Hardware: *Universality*
  - same random projections / hardware for *any* compressible signal class
  - simplifies hardware and algorithm design

- Processing: *Information scalability*
  - random projections ~ sufficient statistics
  - same random projections for range of tasks
    - reconstruction > estimation > recognition > detection
  - far fewer measurements required to detect/recognize

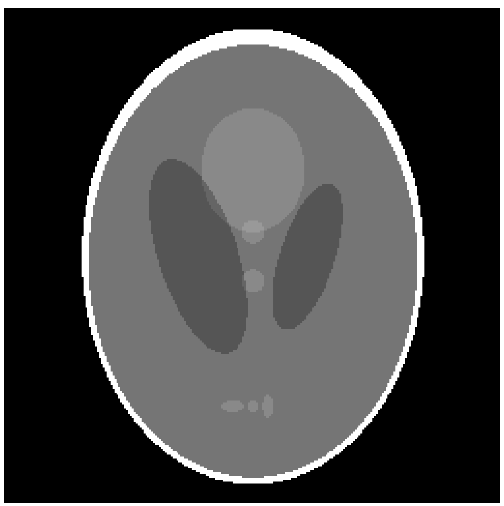
- Next generation data acquisition
  - new imaging devices and A/D converters [DARPA]
  - new reconstruction algorithms
  - new distributed source coding algorithms [Baron et al.]

Compressed Sensing example: Magnetic Resonance Imaging

Compressed Sensing ideas can be applied to reduced sampling in Magnetic Resonance Imaging:

- MRI samples lines of spatial frequency
- Each line takes time and Energy
- Each line heats up the patient!

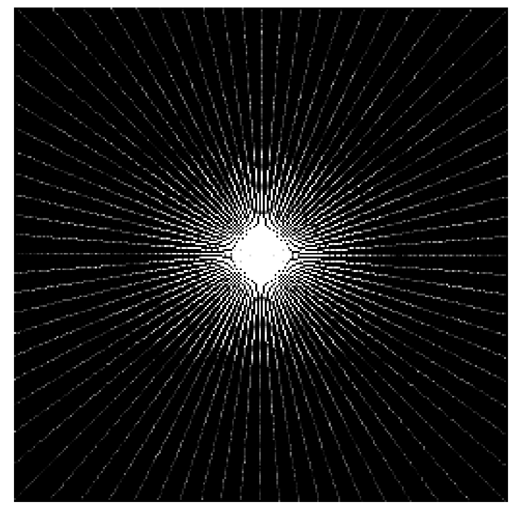
The Logan-Shepp phantom image illustrates this:



Logan-Shepp phantom

Sub-sampled Fourier Transform

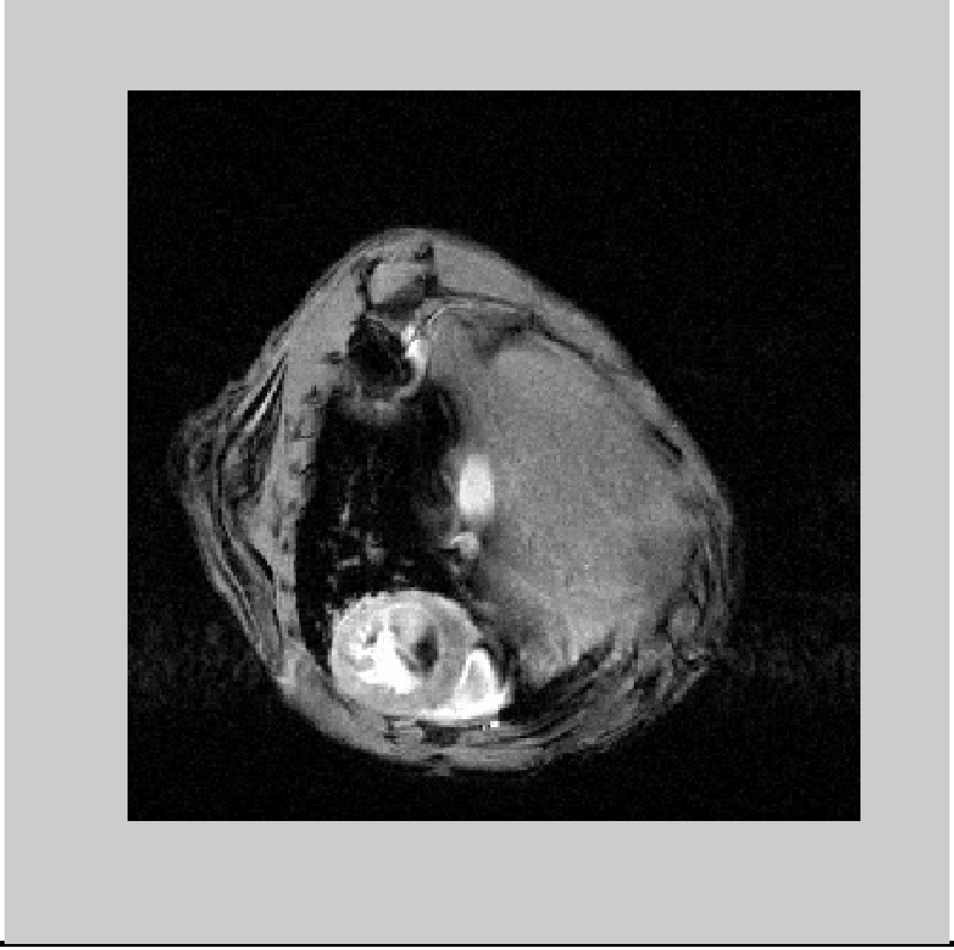
≈ 7 x down sampled (no longer invertible)



…but we wish to sample here

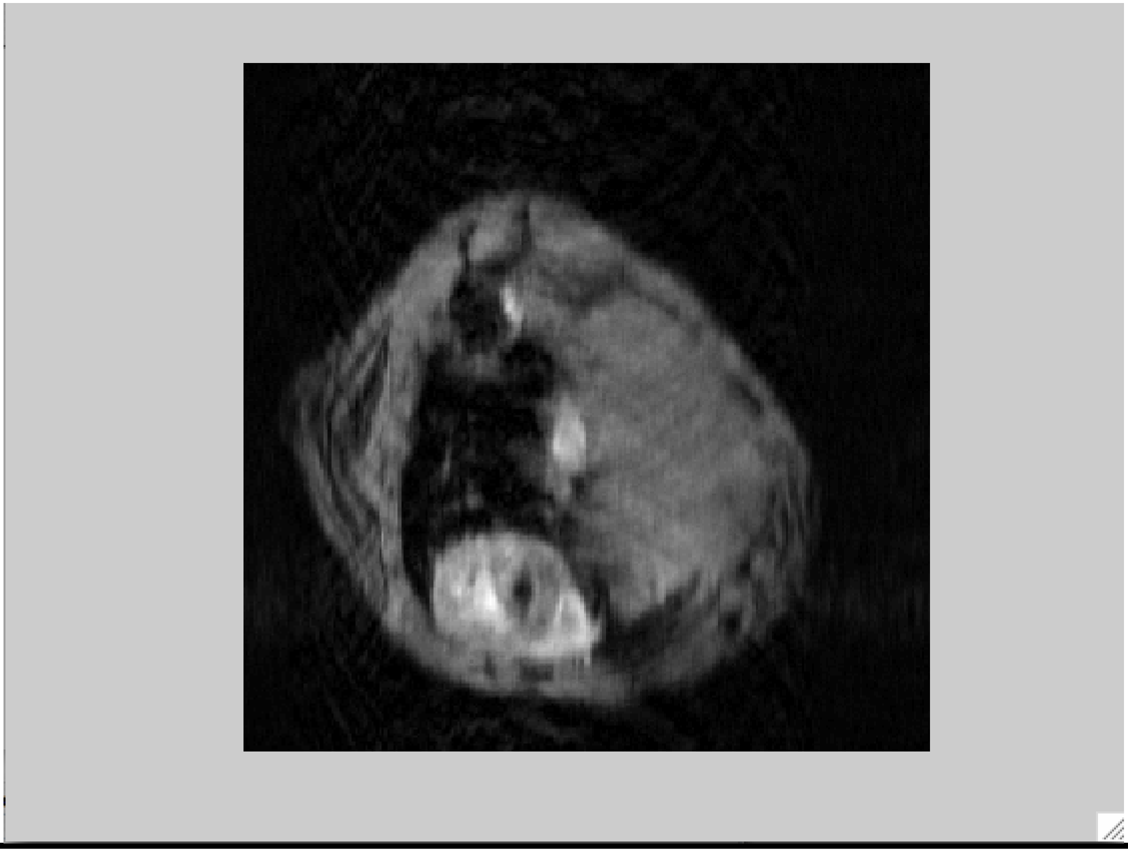
Compressed Sensing in Practice: Magnetic Resonance Imaging

– Original MRI of a Mouse



Compressed Sensing in Practice: Magnetic Resonance Imaging

–MRI of Mouse with 25% Nyquist sampling and an L2 reconstruction

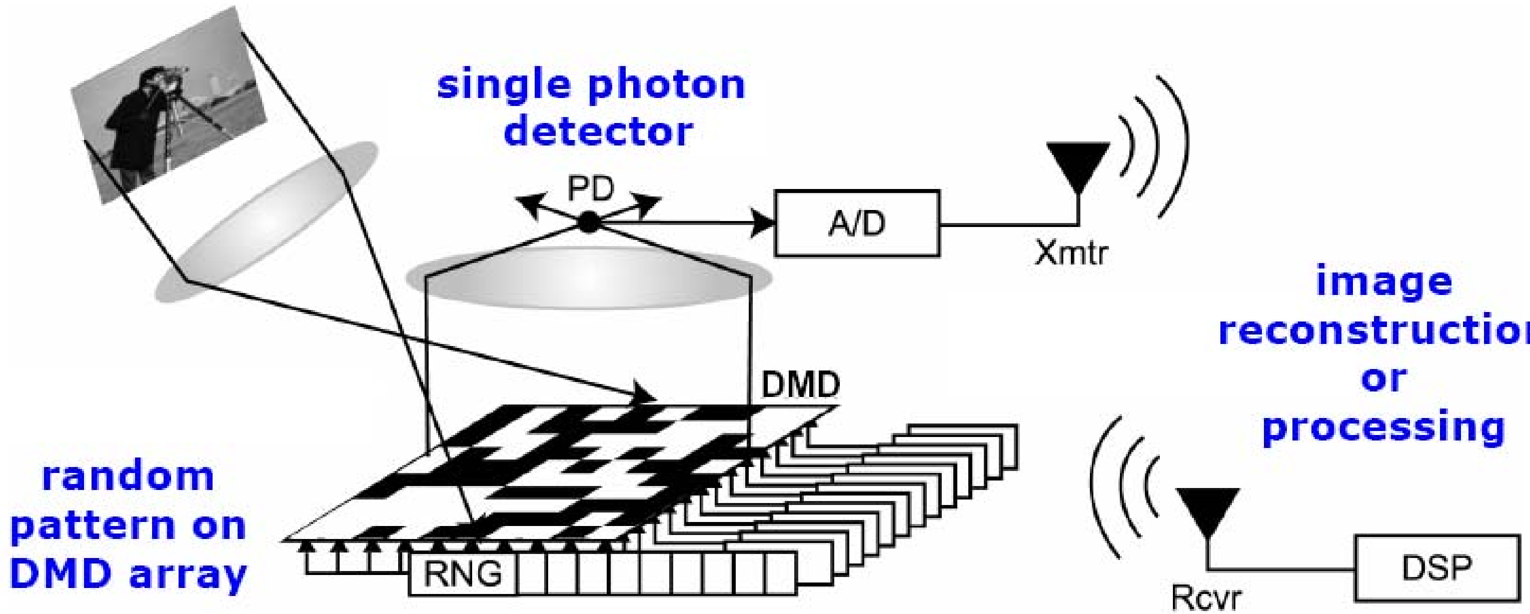


Compressed Sensing in Practice: Magnetic Resonance Imaging

–MRI of Mouse with 25% Nyquist sampling and CS reconstruction

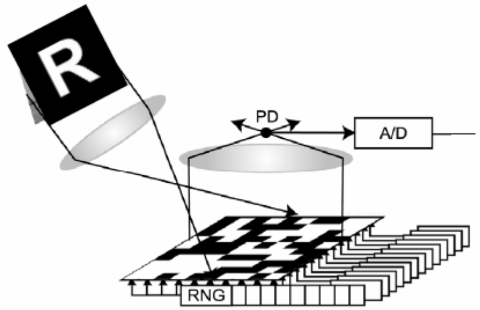
Single Pixel Camera

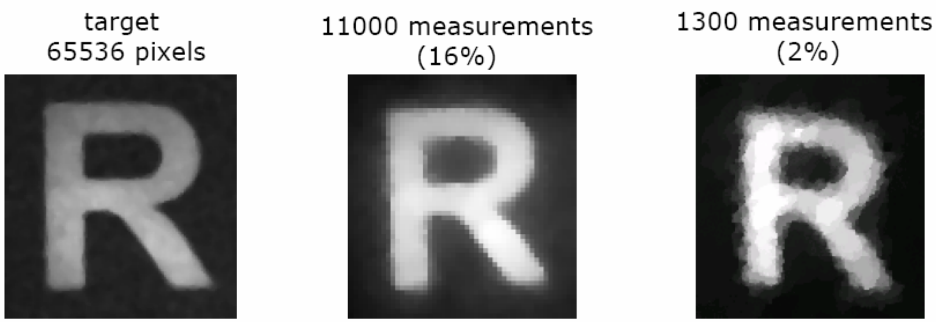
- Encoder integrates sensing, compression, processing



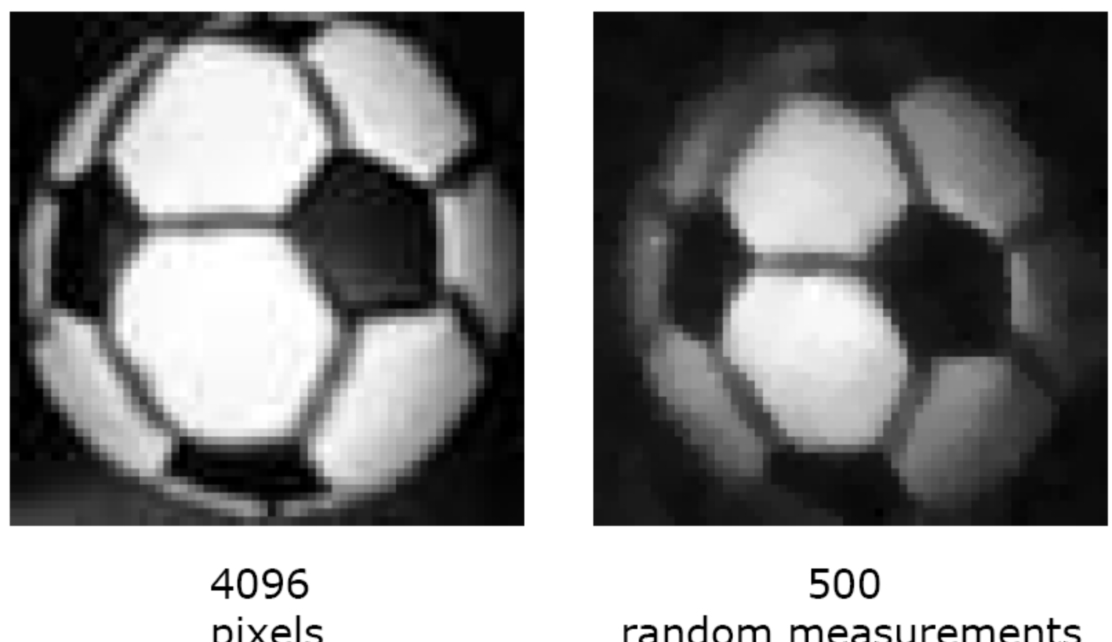
Directly acquires random projections of a scene without first collecting the pixels/voxels, employing a digital micromirror array to optically calculate linear projections of the scene onto pseudorandom binary patterns. Ability to obtain an image or video with a single detection element while measuring the scene fewer times than the number of pixels/voxels. Since the camera relies on a single photon detector, it can also be adapted to image at wavelengths where conventional CCD and CMOS imagers are blind.

First Image Acquisition





Second Image Acquisition



A/D Conversion Below Nyquist Rate

- *Challenge:*
  - wideband signals (radar, communications..)
  - currently impossible to sample at Nyquist rate
- Proposed CS-based solution:
  - sample at "information rate"
  - simple hardware components
  - good reconstruction performance reported

Slepian-Wolf Theorem (Distributed lossless coding)

- Consider a communication system with two correlated signals, X and Y Assume they come from separate sources that cannot communicate, so the signals are encoded independently or are "distributed". The receiver can see both encoded signals and can perform joint decoding.
- A sensor system composed of low-complexity spatially separated sensor nodes, sending correlated information to a central processing receiver, is an example of such system.
- What is the minimum encoding rate required such that X and Y can still be recovered perfectly?

Slepian-Wolf Theorem (Distributed lossless coding)

- The Slepian-Wolf Theorem, established in 1971, shows it is possible to transmit X and Y without any loss if all of the following hold true:

R1 > H(X1|X2) (conditional entropy)

R2 > H(X2|X1) (conditional entropy)

R1+R2 > H(X1,X2) (joint entropy)

CS Approach- Measure separately, reconstruct *jointly*

- Zero collaboration, trivially scalable, robust
- Low complexity, universal encoding
- Ingredients
  - models for joint sparsity
  - algorithms for joint reconstruction
  - theoretical results for measurement savings
- The power of random measurements
  - *single-signal:* efficiently capture structure without performing the sparse transformation
  - *multi-signal:* efficiently capture joint structure without collaborating or performing the sparse transformation

Summary

Sparse representations provide a powerful mathematical model for many natural signals in signal processing and are a basis for:

- good compression;
- good source separation and;
- efficient sampling

There is an interesting interplay between linear representations and nonlinear approximation

Compressed sensing is only in its infancy…

For more info see:

http://www.dsp.ece.rice.edu/cs/