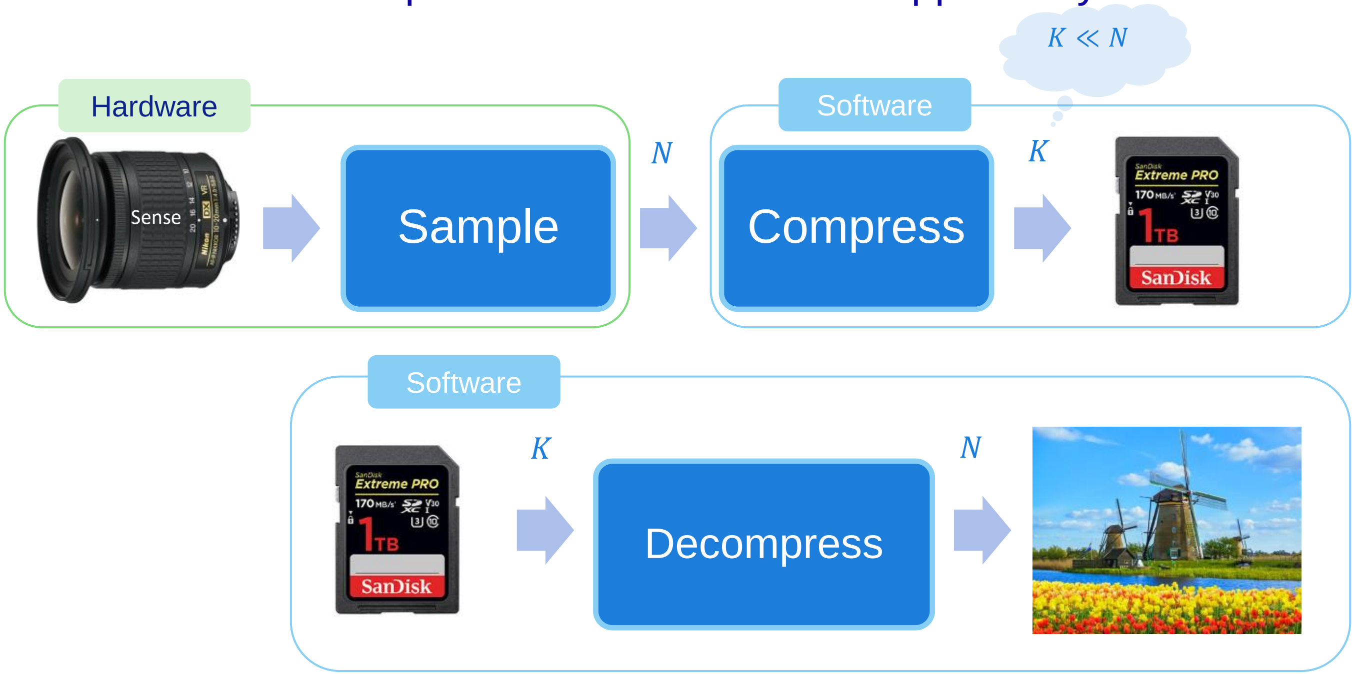
Compressive Sensing

Iman Mossavat
Senior Design Engineer, ASML Research

October 28, 2021
PyData Global 2021

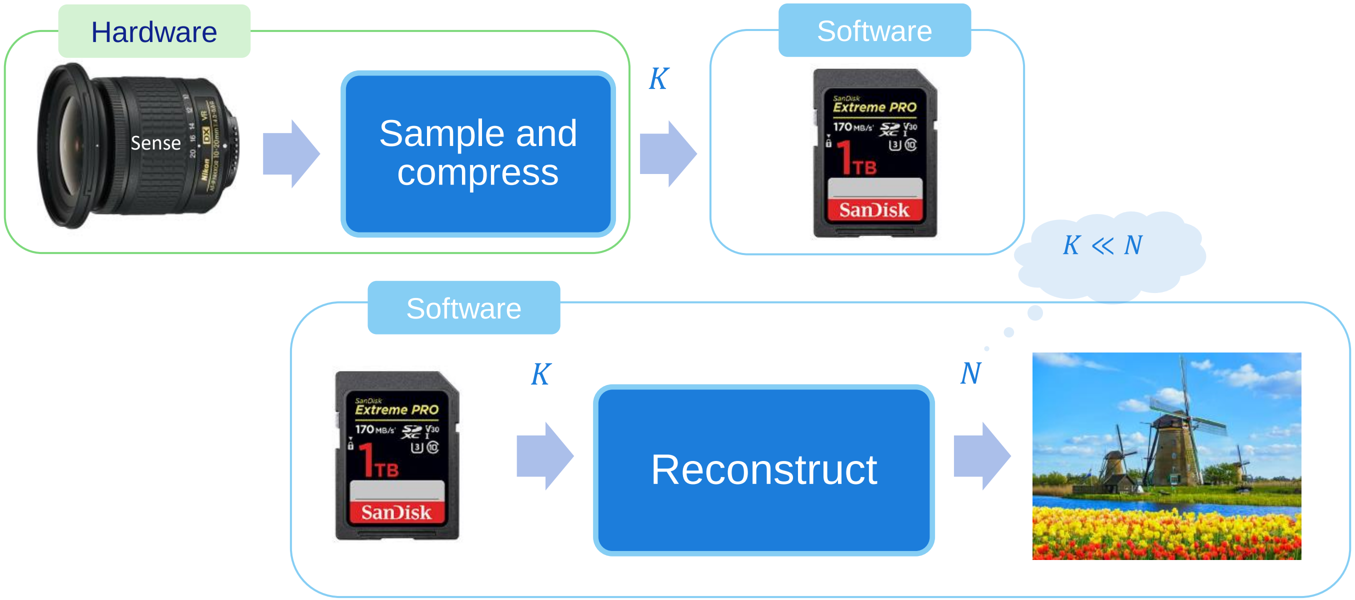
Traditional Workflow of Sensing, Storage and Usage

It is the success of data compression that is the ~~issue~~ opportunity.



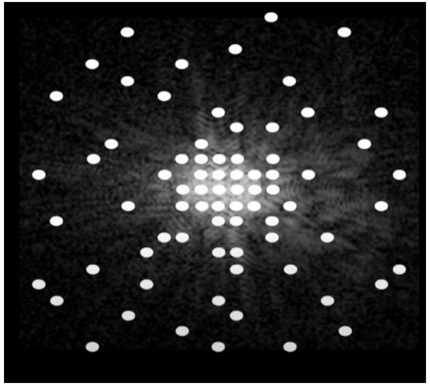
One can regard the possibility of digital compression as a failure of sensor design. If it is possible to compress measured data, one might argue that too many measurements were taken.
David Brady, professor or photonics at Duke university

Compressive sensing workflow



Most of the heavy lifting is done by the SW in the reconstruction phase.
The compression is often a very simple mechanism as you will see in the next slides.

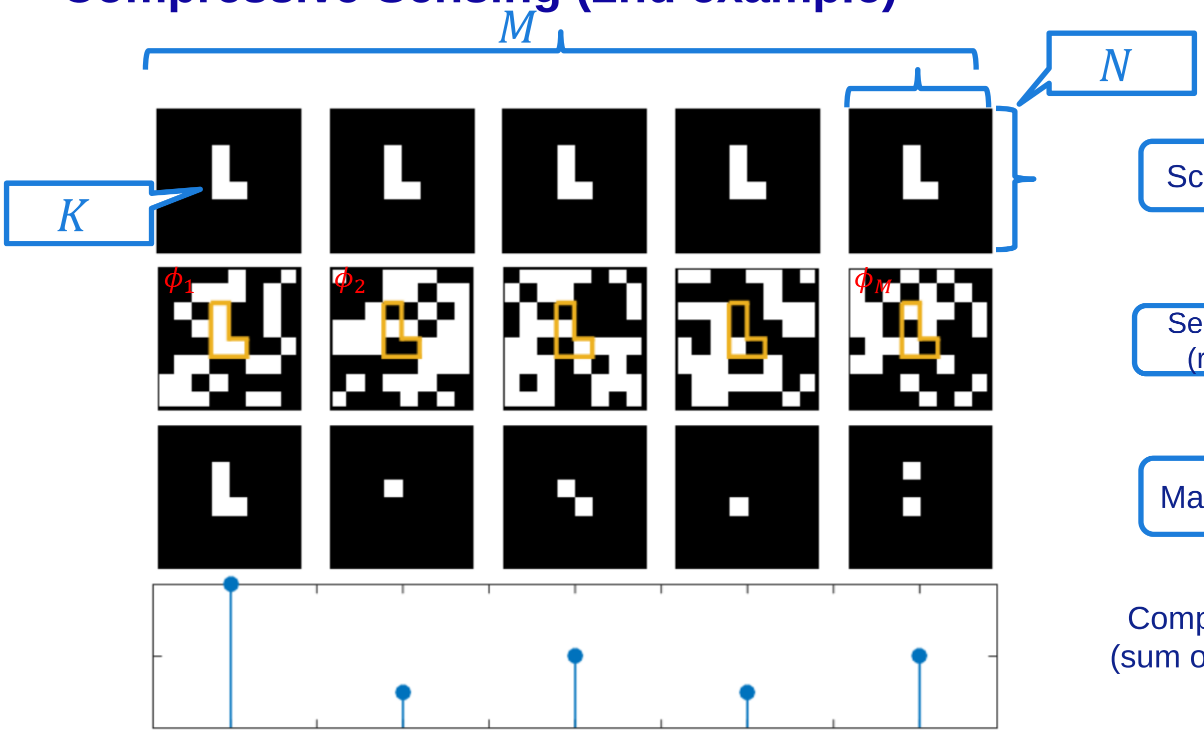
Compressive Sensing in Magnetic Resonance Imaging



MRI measures the Fourier transform of the image.

- To avoid blur the patient need not move (hold breath) during the measurement for 2 minutes.
- Pediatric MRI investigations were children move frequently depend on CS in order to measure in only 10 to 20 seconds.

Compressive Sensing (2nd example)



Scene ($f = x$)

$A = \Phi$

Sensing Mask
(rows of Φ)

Masked Scene

Compressed sample
(sum of masked pixels)
Y

M: number of measurement
K: number of non-zero elements (~ problem complexity/ minimal number of measurements needed)
N: number of pixels (problem dimensionality)

In-painting: filling in missing measurements

Input with missing pixels

In-painted Image

Summary and Outlook

Generic

Compressive sensing (**CS**) shifts the complexity from sensors (Hardware) to algorithms (Software) and computing infrastructure.

1. CS trades the reduction of the *number of measurements* with higher SW complexity and more extensive calculations.

2. Risk of potential competition: simpler sensors get the possibility to compete with more complex ones over throughput and cost of goods

3. Depending on the application, gains in terms of the number of measurements can vary from 2x to 10x.

Applications

MRI (Magnetic Resonance Imaging): CS achieves ~10x reduction in measurement time *enabling entirely new MRI applications*.

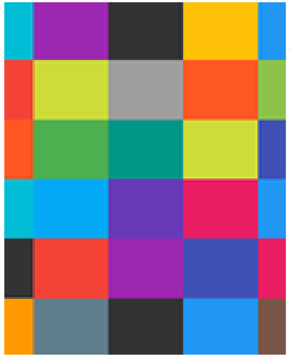
IoT (Internet-of-Things): CS enables the use of sensors with limited bandwidth and battery power and offers a form of data encryption.

Super-resolution: Endo-microscopy beyond the Abbe and Nyquist limits (spatial resolution more than 2 times better than the diffraction limit), ARCNL

The pillars of compressive sensing

Prior knowledge

- What kind of signals are we reconstructing:
  - Sparsity
  - Generative Models (data driven)



Incoherent (*) Measurements

* **Incoherence** is Not meant in the optical sense.

Sparsity

In the right representation, only a few elements carry most of the information.

Signal f is represented in basis Ψ as

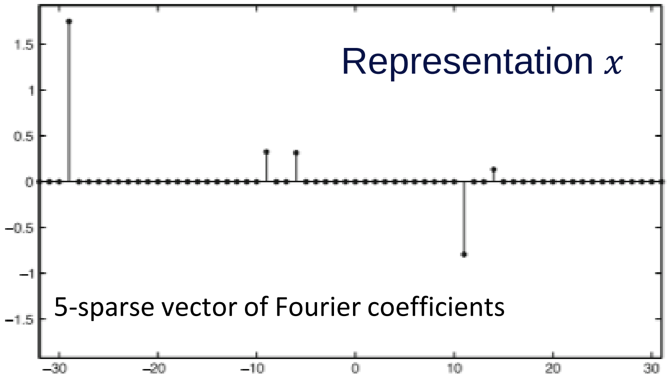
$$f = \Psi x$$

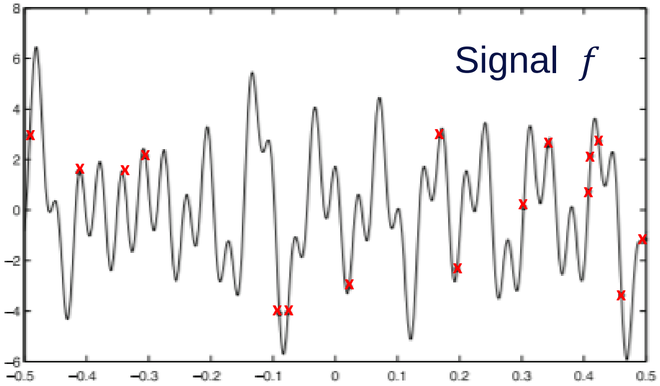
The signal f is "k" - sparse in basis Ψ if only k elements of the representation x are non-zero."

Example .1:

Real part of time-domain signal with 16 samples (marked by red x)

Can we get to reconstruct? (1/2)





Images from “A Mathematical Introduction to Compressive Sensing”, Foucart & Rauhut

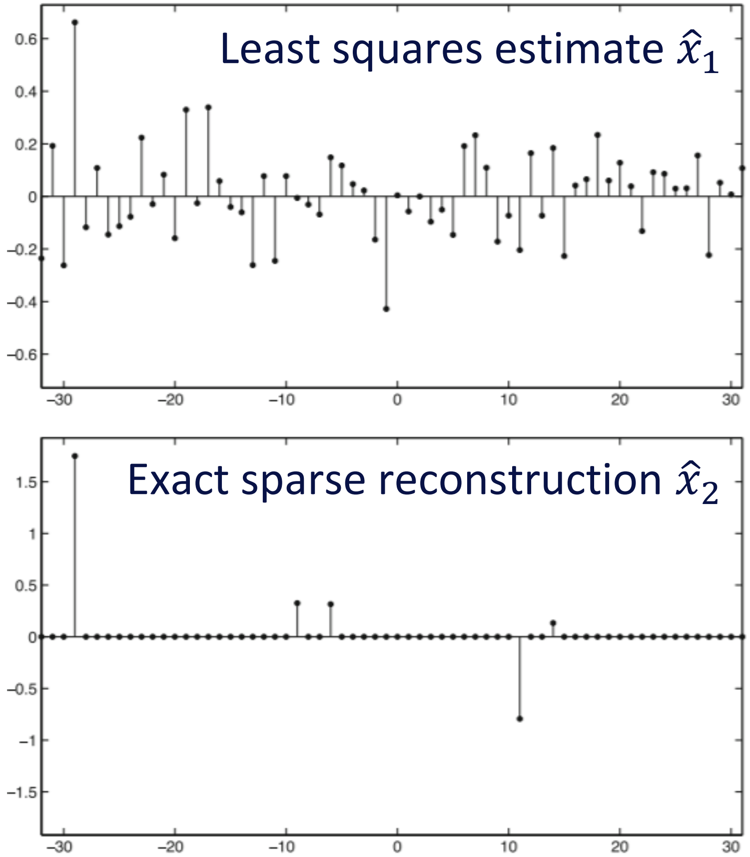
Sparsity

In the right representation, only a few elements carry most of the information.

Example .1 Cont. (2/2) :

Signal is reconstructed based on the 16 measurements (x) with the standard least squares (*Top panel*) and sparse reconstruction algorithm (*Bottom panel*).

From *A Mathematical Introduction to Compressive Sensing*
Simon Foucart Holger Rauhut



Signals are not really sparse! More advanced priors are needed

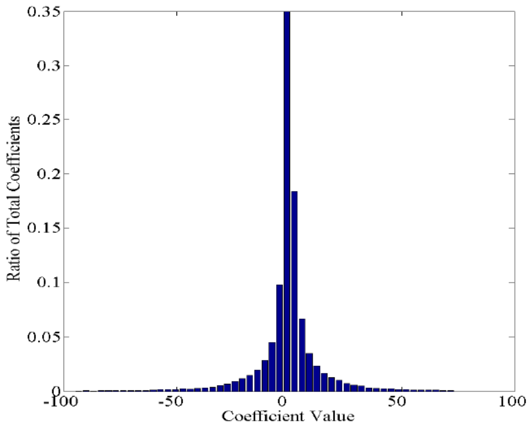
More advanced algorithm successfully deal with non-sparse signals

In practice natural signals are not sparse (they are not spiked at zero) but they are ***heavy-tail*** following a ***power-law*** (lots of energy in few elements)

$$g(x) \propto x^{-k}$$

Histogram of the Daubechies 4 wavelet coefficients of the Barbara test image. Notice the non-sparse distribution of the coefficients. Sparsity-based compressed sensing algorithms fail because of this distribution.

From Denoising to Compressed Sensing, Mezler et al



Hardware

The sensing problem

The signal f is measured via the waveforms ϕ the measured values Y are linear projections:

$$Y = \Phi f$$

The measurement

The sensing mechanism (HW)

The signal of interest

Examples:

1. If ϕ is sinusoids then we are sensing the *Fourier coefficients* such as in MRI
2. If ϕ is integrating the signal over the area of pixels, then we get *pixelated* signals as in CCD
   - Super-pixels are another example

The Full System

The signal f is measured via the waveforms ϕ the measured values Y are linear projections:

$$Y = \Phi f$$

We represent the signal f in some basis Ψ where it is k-sparse

$$f = \Psi x$$

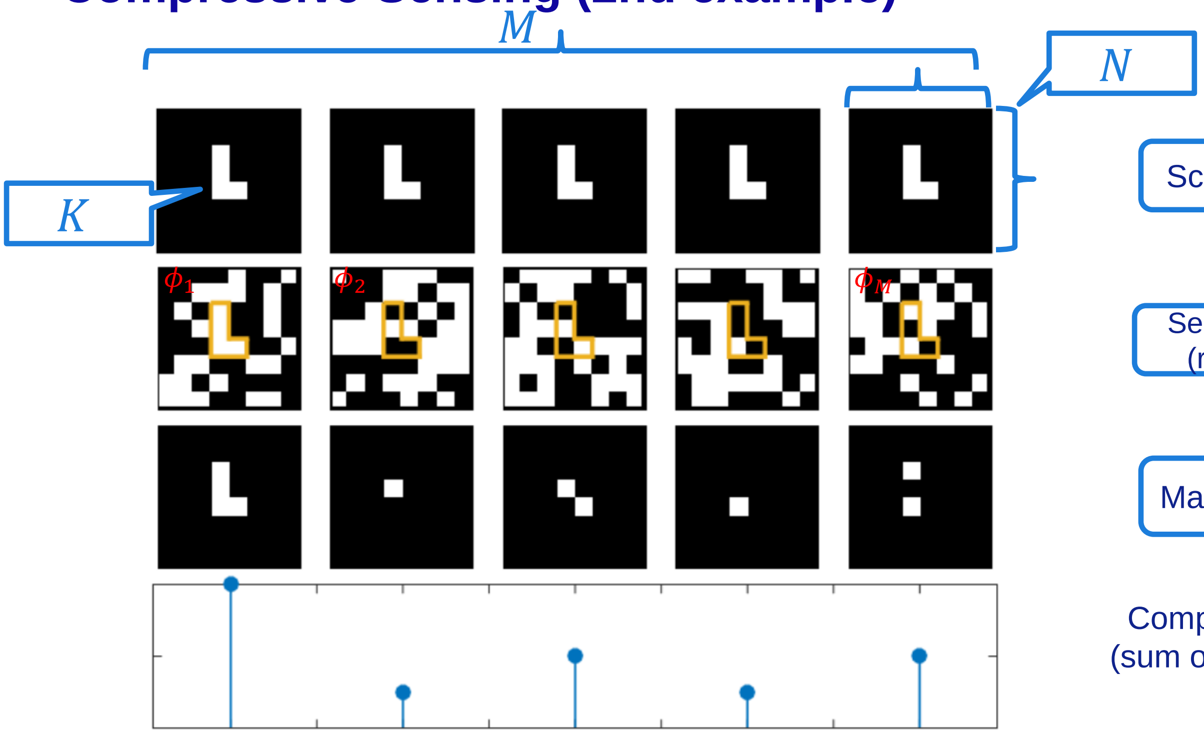
Overall we get:

$$Y = Ax$$

where

$$A = \Phi\Psi$$

Compressive Sensing (2nd example)



Scene ($f = x$)

Sensing Mask
(rows of Φ)

Masked Scene

Compressed sample
(sum of masked pixels)
Y

$f = x$
$(\Psi = I_{N\times N})$

$A = \Phi\Psi = \Phi$

M: number of measurement
K: number of non-zero elements (~ problem complexity/ minimal number of measurements needed)
N: number of pixels (problem dimensionality)

Image Source: U.S. Army Combat Capabilities Development Command Army Research Laboratory

Compressive sensing in a nutshell

Compressive sensing involves solving a system of Equations :

$$Y_{M\times 1} = A_{M\times N}x_{N\times 1}$$

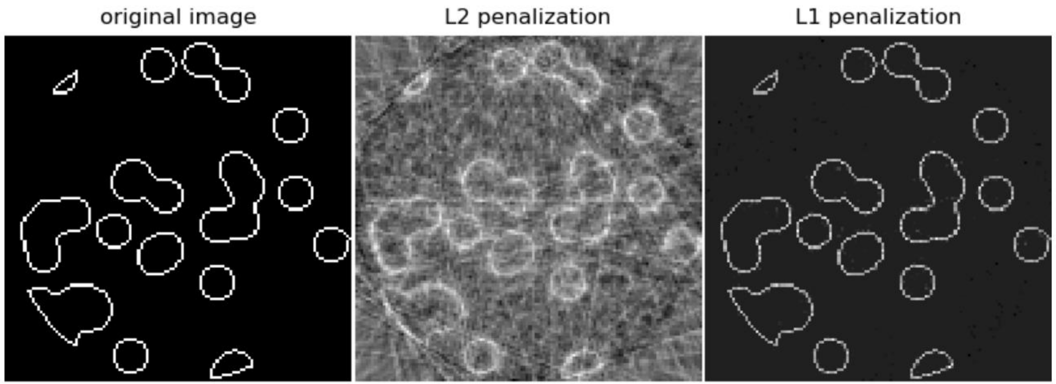
1. **The system of equations has more unknowns than equations: $N > M$.**
   - M is number of measurements and N number of unknowns
   - To solve we need extra assumptions on x, for example x having only K non-zero elements where $M \geq K$

2. **Linear system A is constrained by the hardware $\mathrm{A} = \Phi\Psi$, where Φ is hardware dependent, Ψ is a transformation that ensures the sparsity of x.**

3. **Not every A is a candidate: A should keep K-sparse signals separate from each other so that we can recover them: restricted isometry property (RIP)**
   - For full i.i.d. random matrices A, RIP requires $M \geq CK \log\frac{N}{K}$ for some constant C.

Python Example

Basic L1 reconstruction in Scikit-Learn



Compressive sensing: tomography reconstruction with L1 prior (Lasso)

https://scikit-learn.org/stable/auto_examples/applications/plot_tomography_l1_reconstruction.html

A more formal description of the cost function

Originally the compressive sensing problem amounts to the following constrained optimization problem

$$\arg_{\mathrm{x}} \min |x|_0$$

such that

$$Y = Ax$$

$|x|_0$ is called the ℓ_0 norm, which denoted the number of non-zero elements.

This is a **nightmare** to solve when the length of x increases (NP-hard). Run-time grow exponentially with length of x (you need to figure out which subset of x elements are non-zero.)

The ℓ1 magic

Given sufficient number of measurements, $M \geq C\mu^2 K \log N$ one can replace the challenging ℓ_0 norm problem with a much simpler problem that can be solved very quickly with no loss of accuracy!

$$\arg_x \min |x|_1$$

such that

$$Y = Ax$$

$|x|_1$ is called the ℓ_1 norm, the sum of the absolute values of elements. This is a convex function and we have powerful tools for it.

https://statweb.stanford.edu/~candes/software/l1magic/

Problem gets easier with increasing number of measurements (M)

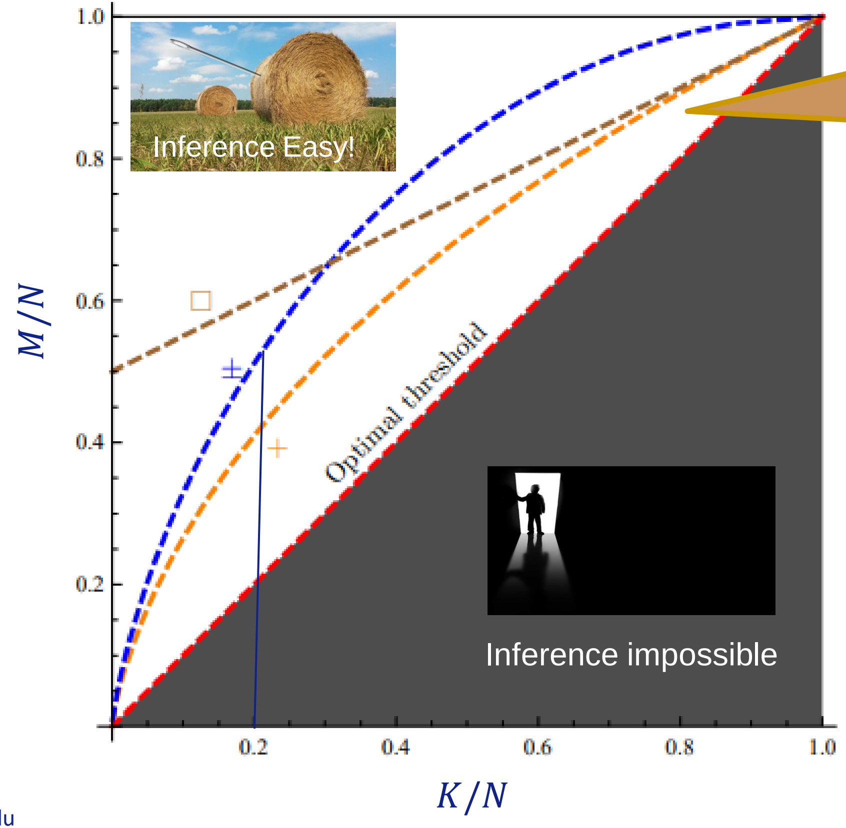
$M \leq K$	Only ℓ_0 norm works	ℓ_1 norm works
		
Find the black cat in the dark room	Find the needle in the haystack	Find the big! needle in the haystack
Insufficient Information (Impossible in information sense)	Computationally infeasible (Information is there but no polynomial time algorithm exists)	Easy! Information is there and can be retrieved in polynomial time.

Axis:
M/N: Measurement rate
K/N: Probability of non-zero value

Signal Type:

$\pm$: Sparse signals
$+$: Sparse non-negative signals
□ : Simple signals

Note that these curves capture asymptotic behavior as image dimensions $N \to \infty$



Statistical Physics of Inference

Decreasing Temperature

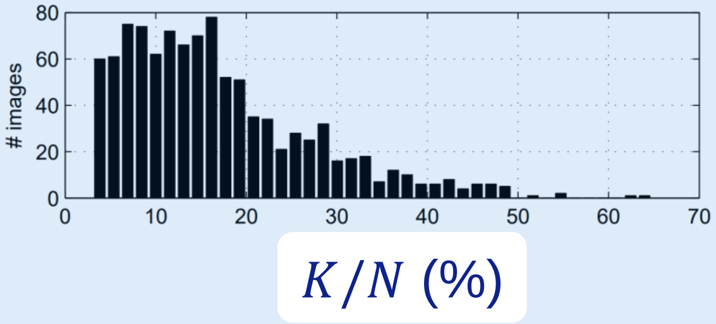
Increasing number of measurements

Free Energy= Energy – Entropy

There is a deep correspondence between physical phases such as liquid, super-cooled liquid or glass, and solid, and regions of parameters for which a given data analysis task is algorithmically impossible, hard or easy.

https://theconversation.com/cracking-big-data-with-statistical-physics-79864

Reminder: Signals are not really sparse!

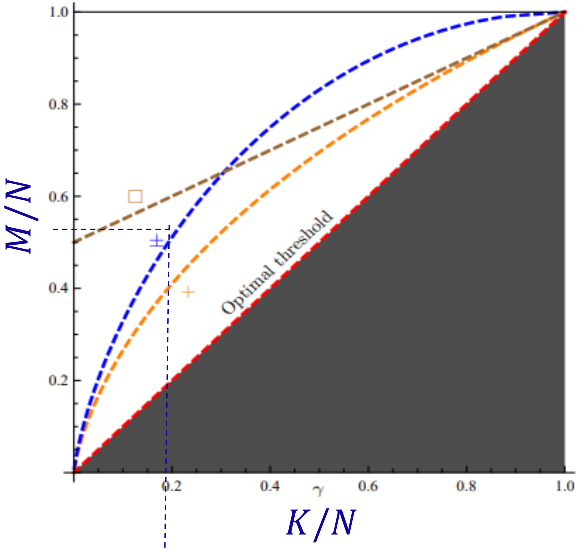


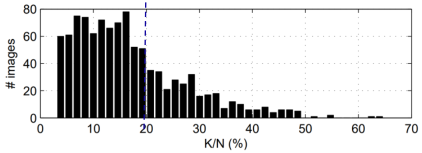
- **The average sparsity is 17%, with half of all images less than 15% sparse, and three-quarters less than 20% sparse.**

- Sparsity is measured by counting the number of block-DCT coefficients K that account for at least 99.75% of the image energy.

Sparse imaging for fast electron microscopy, Anderson et al 2012

Signals are not really sparse!



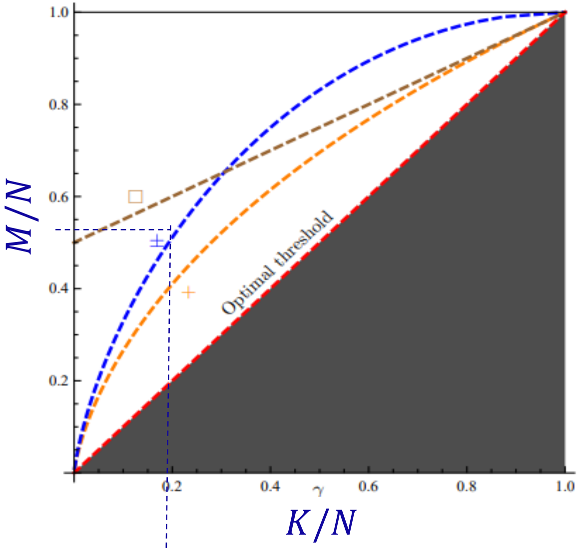


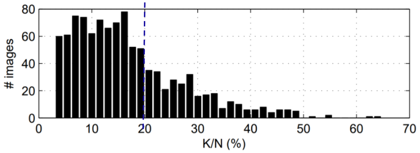
Axis:
M/N: Measurement rate
K/N: Probability of non-zero value

Optimal phase transitions in compressed sensing, Wu, Verdu

Sparse imaging for fast electron microscopy, Anderson et al 2012

Deep learning priors are key to reduce the number of measurements.

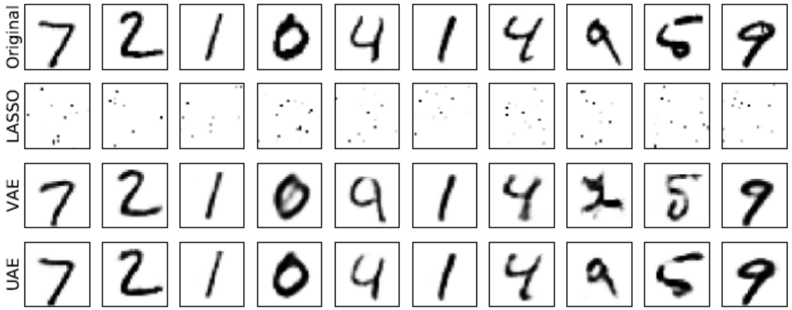




ℓ_1

NN

NN



With only 25 measurements the NN based solutions can work.

Variational Compressive Sensing using Uncertainty Autoencoders, Grover, Ermon

Optimal phase transitions in compressed sensing, Wu, Verdu

Sparse imaging for fast electron microscopy, Anderson et al 2012

Spectrum Of Methods (2017)

Data Efficient Interpretable Model Physics		Data Extensive Low bias (Flexible)		
Simple Model	**Advanced Model**	**Simple Model**	**Advanced Model**	**Model Agnostic**
ISTA AMP/VAMP CoSAMP TVAL3	Model CoSaMP NLR-CS D-AMP/D-VAMP C-GD	LISTA LAMP/VAMP RBM-AMP	ADMM-NET RIDE-CS OneNet LDAMP	SDA ReconNet DR2-Net Deep Inverse

Spectrum Of Methods (2017)

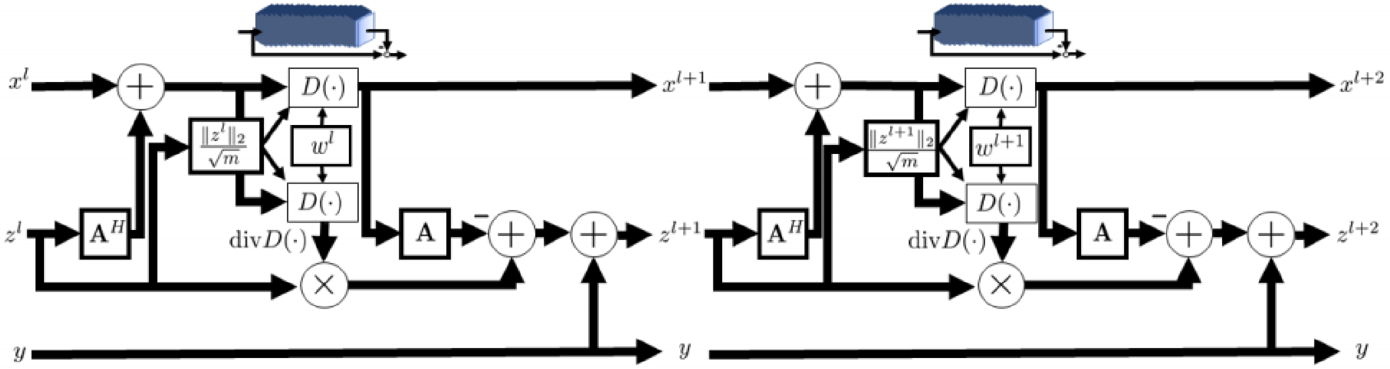
Data Efficient Interpretable Model Physics	Data Extensive Low bias (Flexible)

Simple Model	Advanced Model	Simple Model	Advanced Model	Model Agnostic
Based on Sparsity	Hand-designed Priors	Unrolled Solvers NN-Based Priors NN-Based Posteriors	Unrolled Solvers NN-Based Priors NN-Based Posteriors	Black-box

Hybrid

Learned-Denoising-AMP

AMP: Approximate Message Passing



LDAMP Neural Network

$$\begin{aligned}
b^l &= \frac{z^{l-1}\mathrm{div}D^l_{w^{l-1}(\hat{\sigma}^{l-1})}(x^{l-1} + \mathbf{A}^H z^{l-1})}{m}, \\
z^l &= y - \mathbf{A}x^l + b^l, \\
\hat{\sigma}^l &= \frac{\|z^l\|_2}{\sqrt{m}}, \\
x^{l+1} &= D^l_{w^l(\hat{\sigma}^l)}(x^l + \mathbf{A}^H z^l).
\end{aligned} \tag{5}$$

Deep Learning is much faster and more accurate

But no guarantees!

Table 1: PSNRs and run times (sec) of 128×128 reconstructions with i.i.d. Gaussian measurements and no measurement noise at various sampling rates.

Method	$\frac{m}{n} = 0.10$		$\frac{m}{n} = 0.15$		$\frac{m}{n} = 0.20$		$\frac{m}{n} = 0.25$	
	PSNR	Time	PSNR	Time	PSNR	Time	PSNR	Time
TVAL3	21.5	2.2	22.8	2.9	24.0	3.6	25.0	4.3
BM3D-AMP	23.1	4.8	25.1	4.4	26.6	4.2	27.9	4.1
DL LDIT	20.1	**0.3**	20.7	**0.4**	21.1	**0.4**	21.7	**0.5**
DL LDAMP	**23.7**	0.4	**25.7**	0.5	**27.2**	0.5	**28.5**	0.6
NLR-CS	23.2	85.9	25.2	104.0	26.8	124.4	28.2	146.3