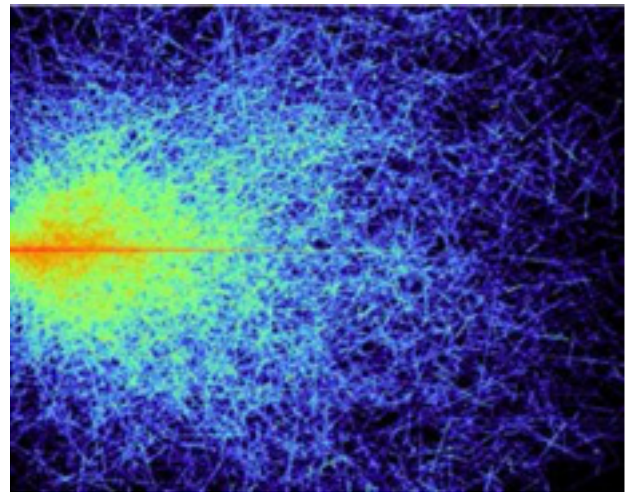
Random matrix advances in large dimensional statistics and machine learning
GdR MEGA / GdR ISIS, Paris,
November 14th, 2017



Wave propagation in strongly scattering materials : Harvesting random matrices for large-scale machine learning

Laurent Daudet, Paris Diderot University & LightOn

université PARIS DIDEROT PARIS 7

Institut Langevin ONDES ET IMAGES

LightOn

A combination of expertise from:

- optics
- signal processing
- optimization / machine learning

Sylvain Gigan

LKB (UPMC / ENS)

Florent Krzakala

LPS (UPMC / ENS)

Igor Carron

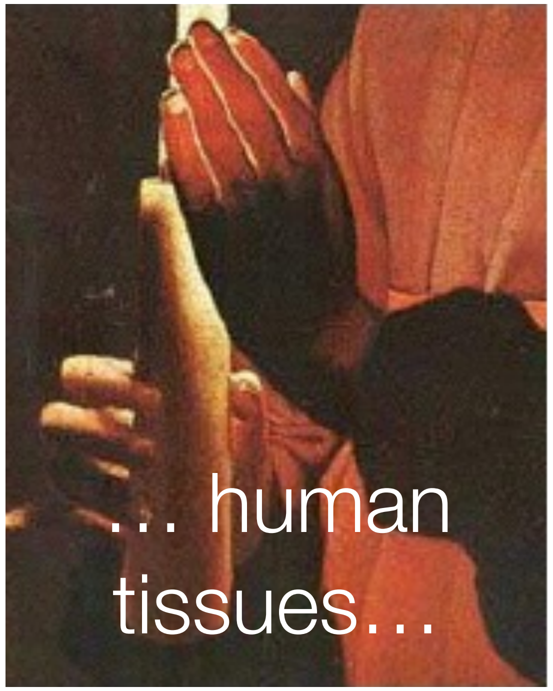
Nuit Blanche / LightOn

and Antoine Liutkus, David Martina, Sébastien Popoff, Gilles Chardon, Boshra Rajaei,

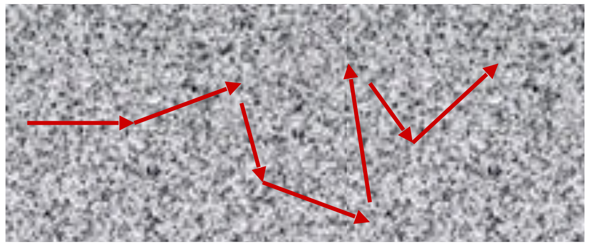
Angélique Drémeau, Ori Katz, Geoffroy Lerosey, Alaa Saade, Francesco Caltagirone …

Light scattering by diffusive materials

Is part of our everyday experience :

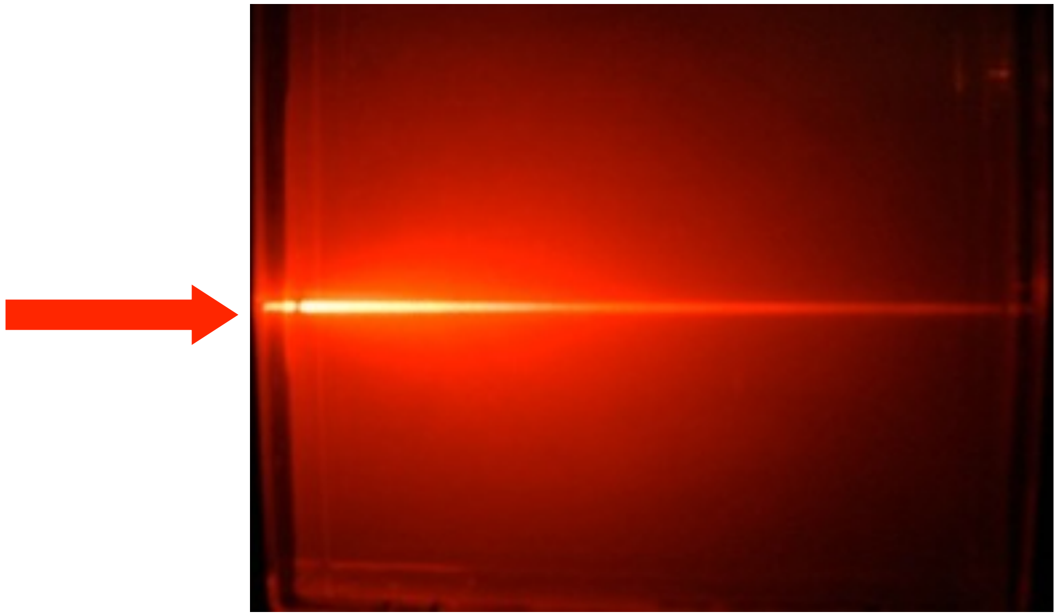


Origin: light is scattered by inhomogeneities



Imaging in scattering media

Conventionally : information from only unscattered (*'ballistic')* light



Beer-Lambert Law: Exponential decay of the ballistic light

→ No imaging beyond a few hundred microns in living tissues

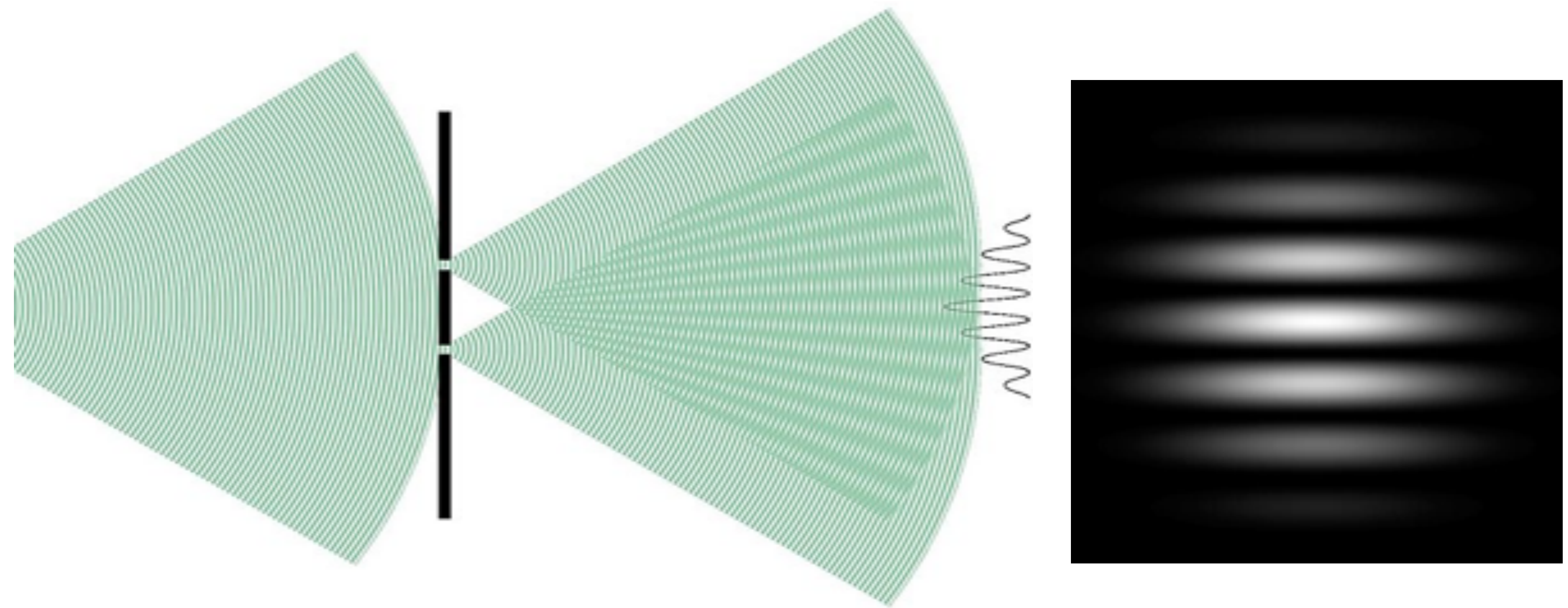
CAN WE GO DEEPER?

Scattering : a coherent process

Young's slit experiment:

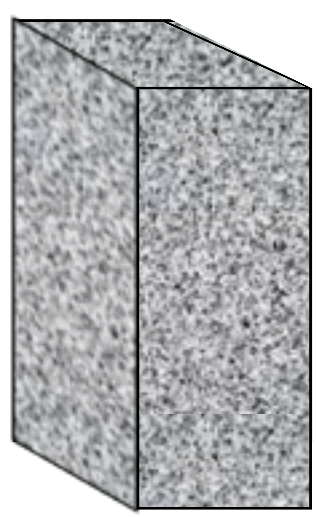
two wave interference

Fringes

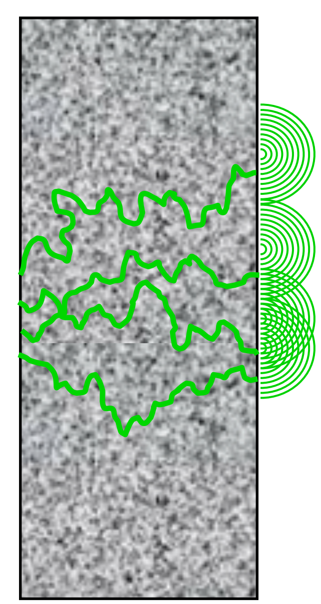


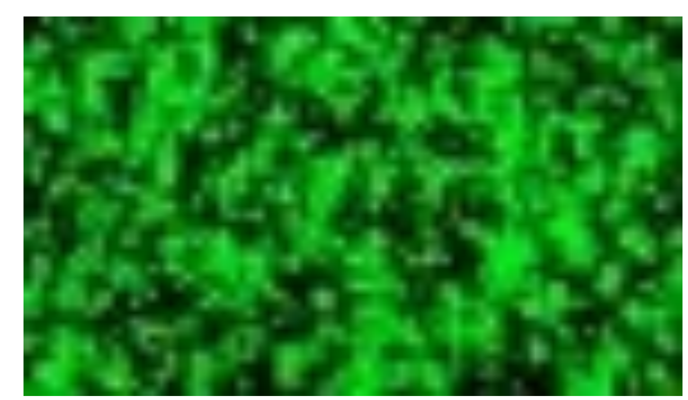
Volume scattering:

Coherent light
(laser)



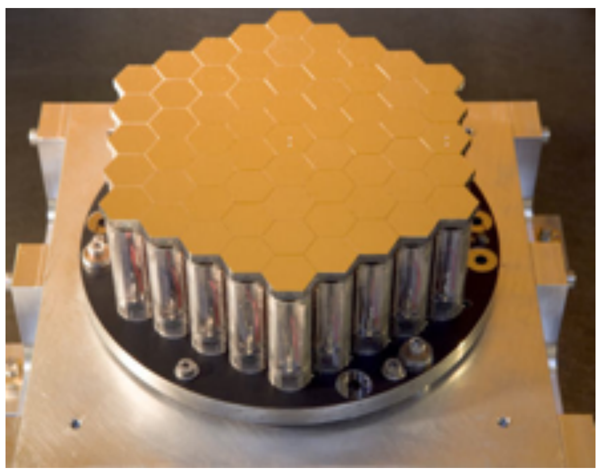
thin layer (about 300 μm)
of white paint
(particle size ≤ 1 μm)





Speckle results from multiple interference
between a multiplicity of random paths

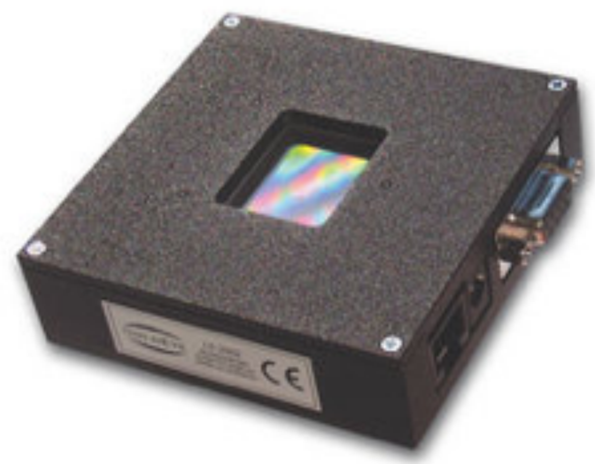
Wavefront shaping : the tool to study scattering



Deformable mirrors

10-100 actuators
moving: 10-20 microns
Speed > kHz

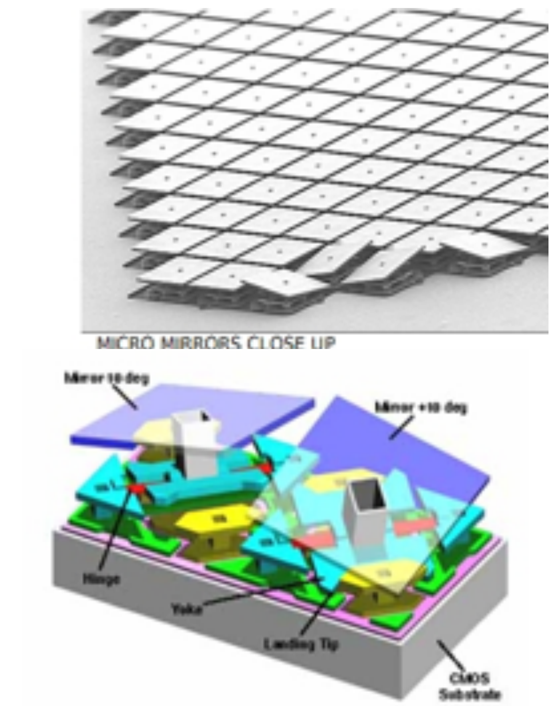
Adaptive optics



Spatial Light Modulators
based on Liquid crystals

>1 million pixels
Phase modulation at: 50Hz

Display

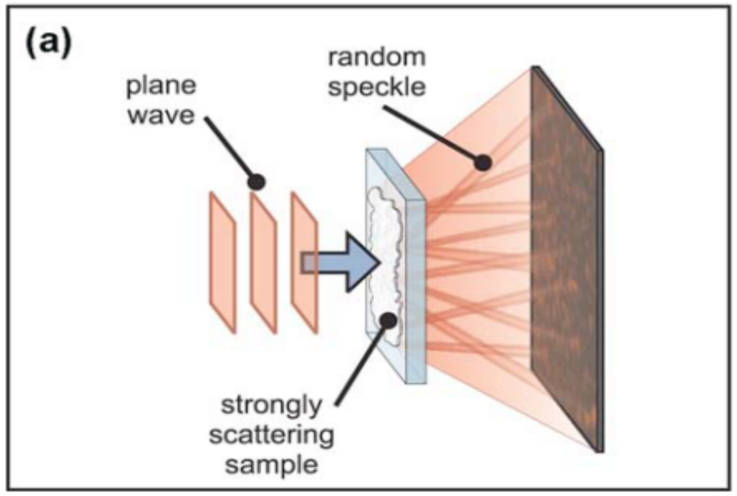


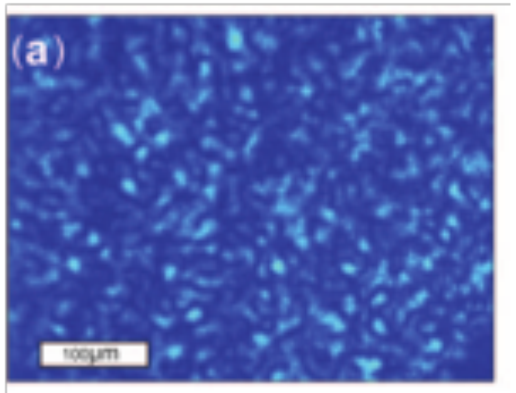
Spatial light modulators based on
MEMS technology
ex: Texas DLP/DMD

>1 million pixels
binary ON/OFF at 20kHz

Display

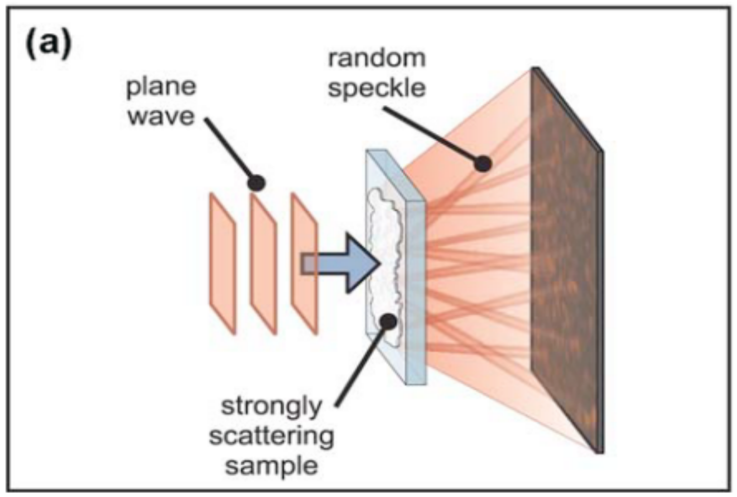
Optimization for focusing through complex media

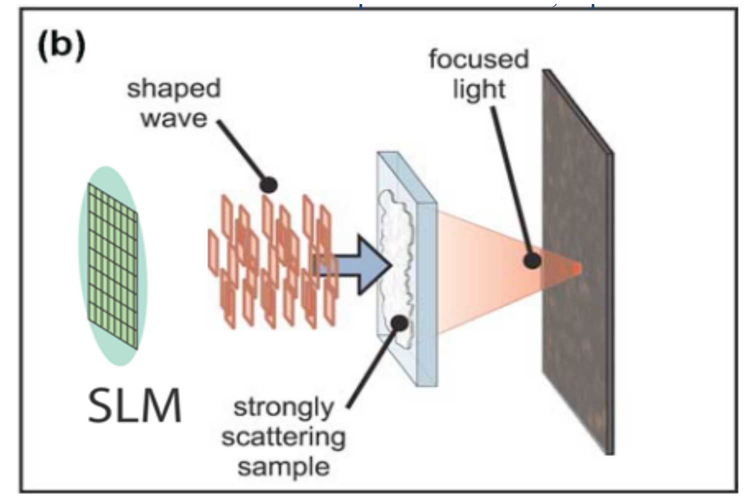


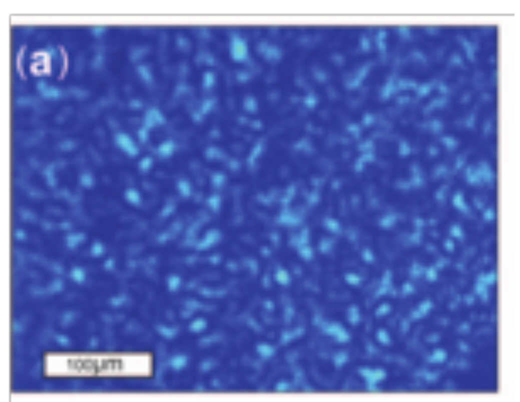


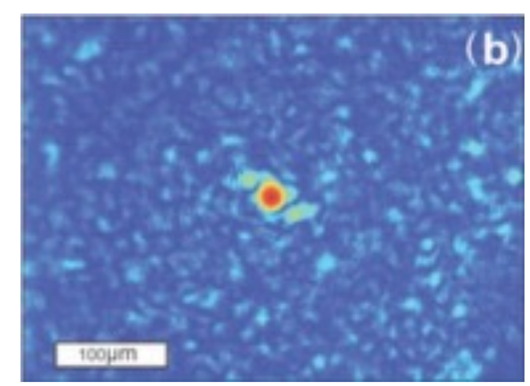
Optimization for focusing through complex media

IM Vellekoop and AP Mosk, Optics Letters, 32(16) 2007





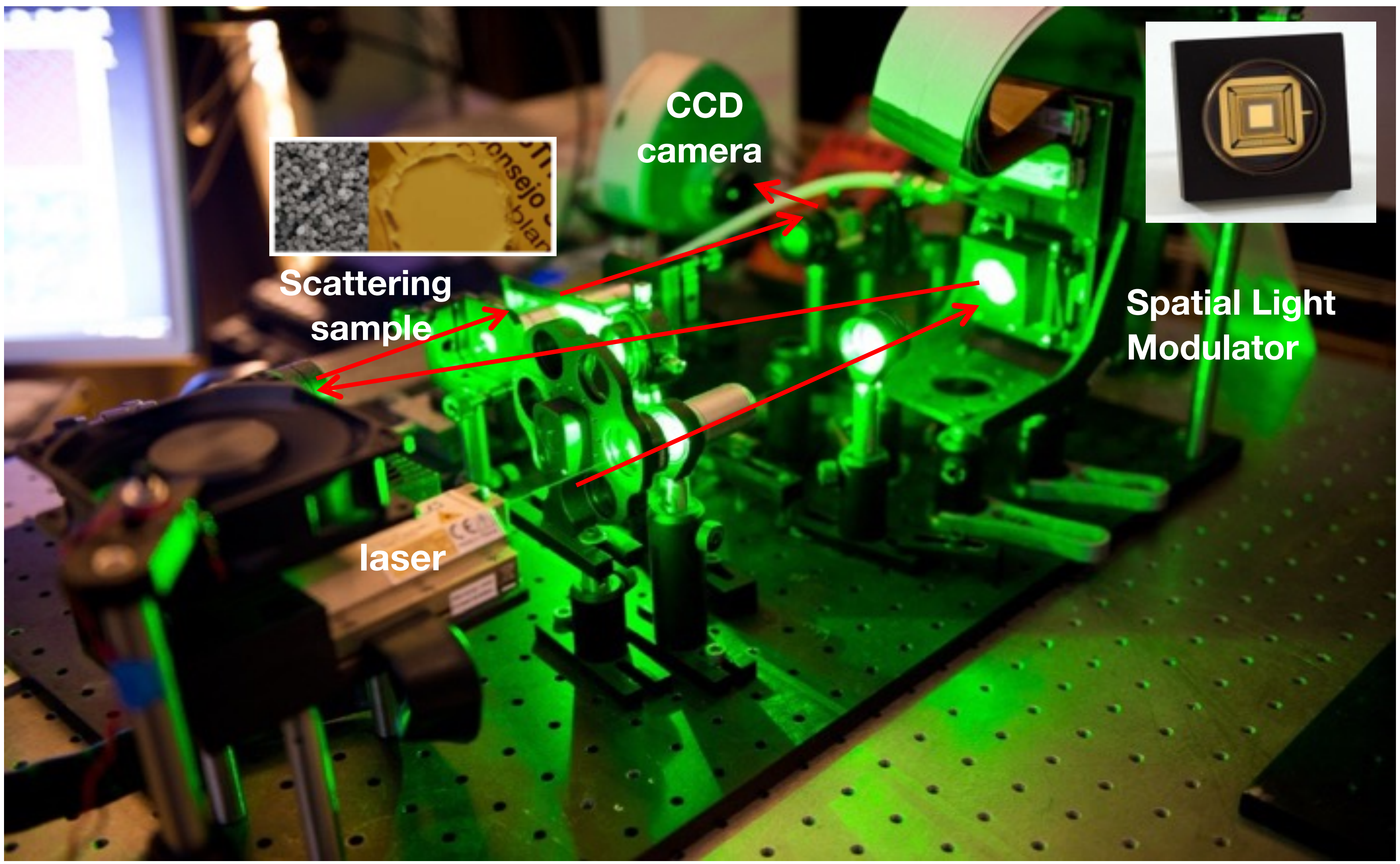




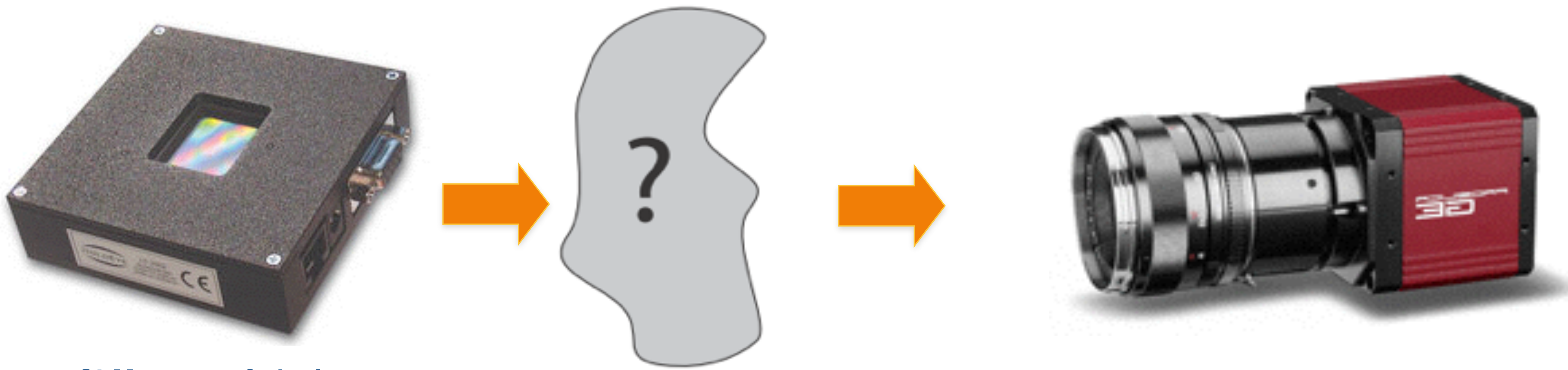
It is possible to shape the incoming wavefront to obtain a constructive interference on a single speckle grain « turn paint into a lens »

Optimization for focusing through complex media

in the lab of Sylvain Gigan - ENS / LKB



A more general approach : the transmission matrix

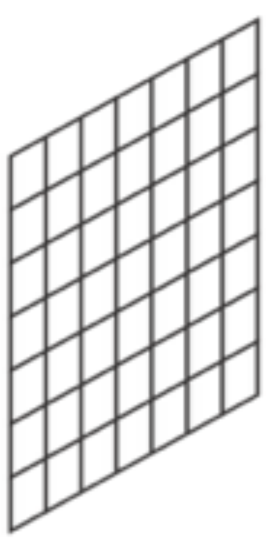


SLM: array of pixels

Linear system

CCD camera: arrays of pixels

=



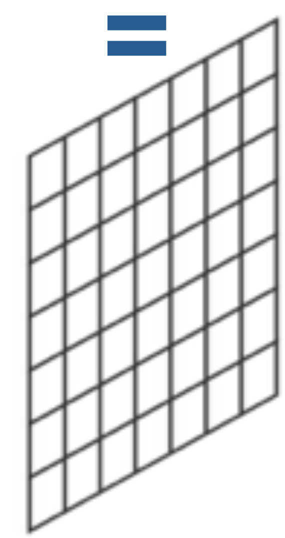
N complex-valued amplitudes

=

$$H = \begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,N} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,N} \\ \vdots & & \ddots & \vdots \\ h_{M,1} & h_{M,2} & \cdots & h_{M,N} \end{pmatrix}$$

MxN complex-valued matrix

=



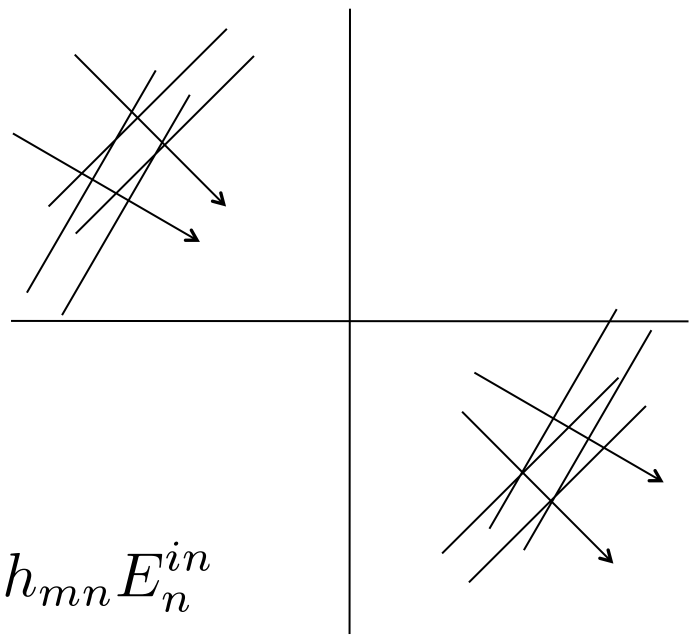
M modulus of complex-valued coefficients

$$\left| E_m^{out} \right| = \left| \sum_n h_{mn} E_n^{in} \right|$$

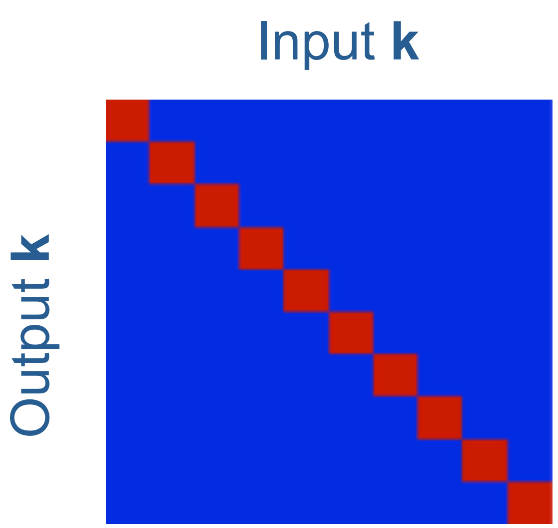
Popoff et al. *Nat. Commun.* 1:81
doi: 10.1038/ncomms1078 (2010)

A more general approach : the transmission matrix

free field

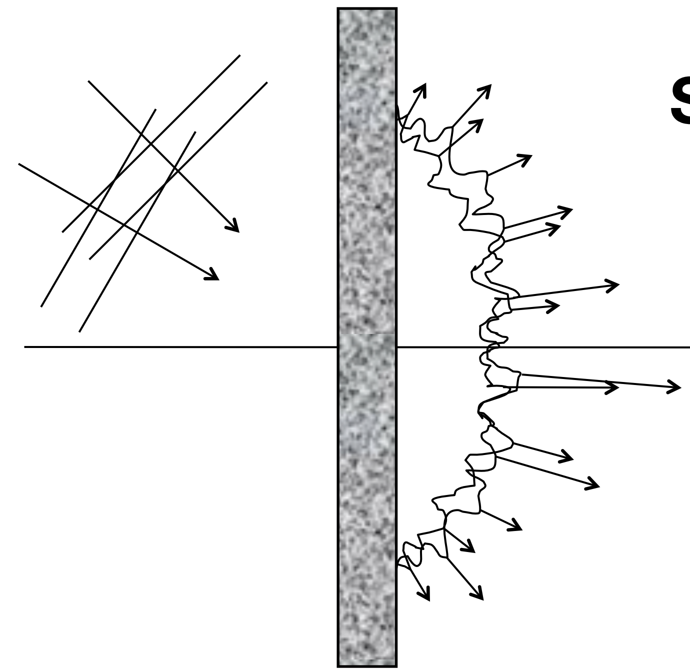


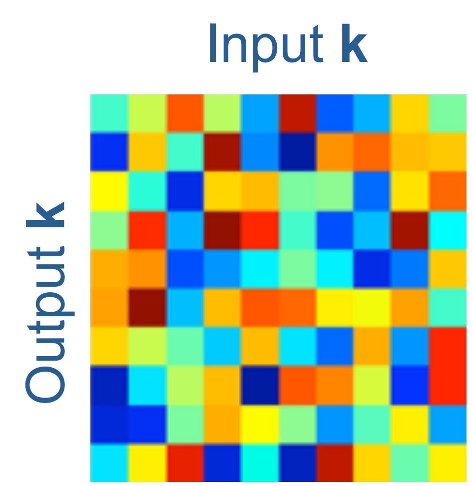
$$E_m^{out} = \sum_n^{1..N} h_{mn} E_n^{in}$$



Identity Matrix

Scattering material

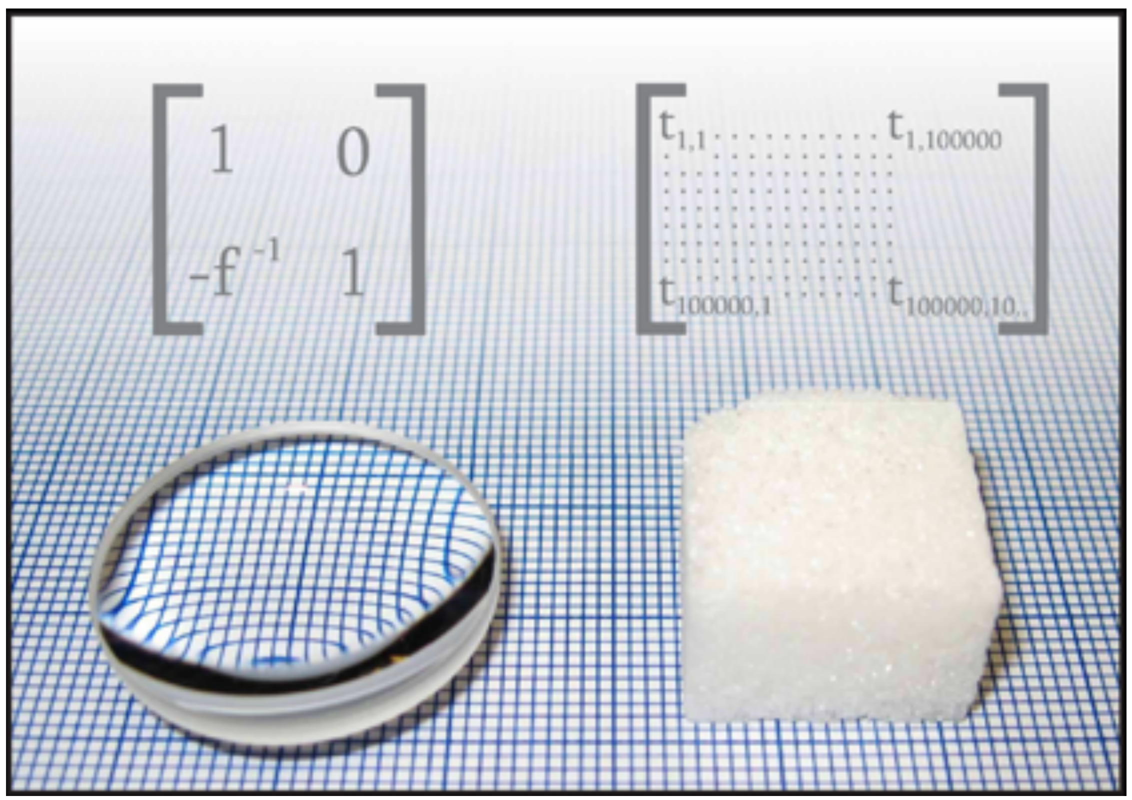




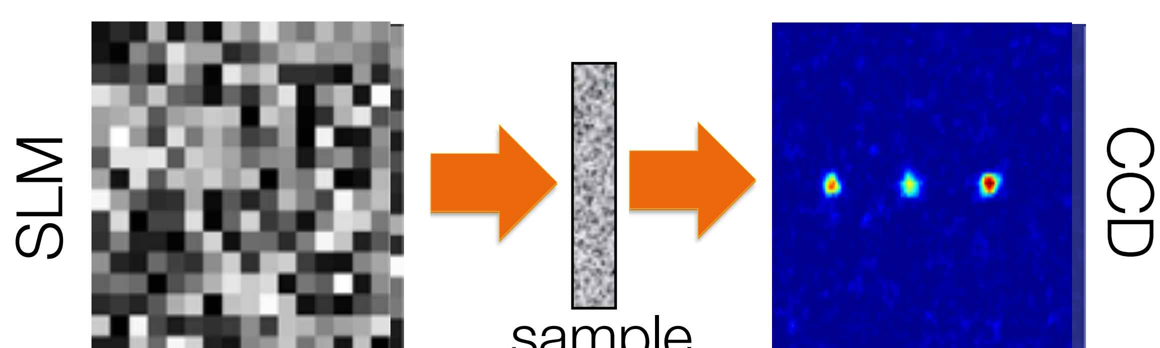
(Seemingly) Random Matrix

A more general approach : the transmission matrix

knowing the transmission matrix turns the scattering material into a « lens » with a very high number of degrees of freedom

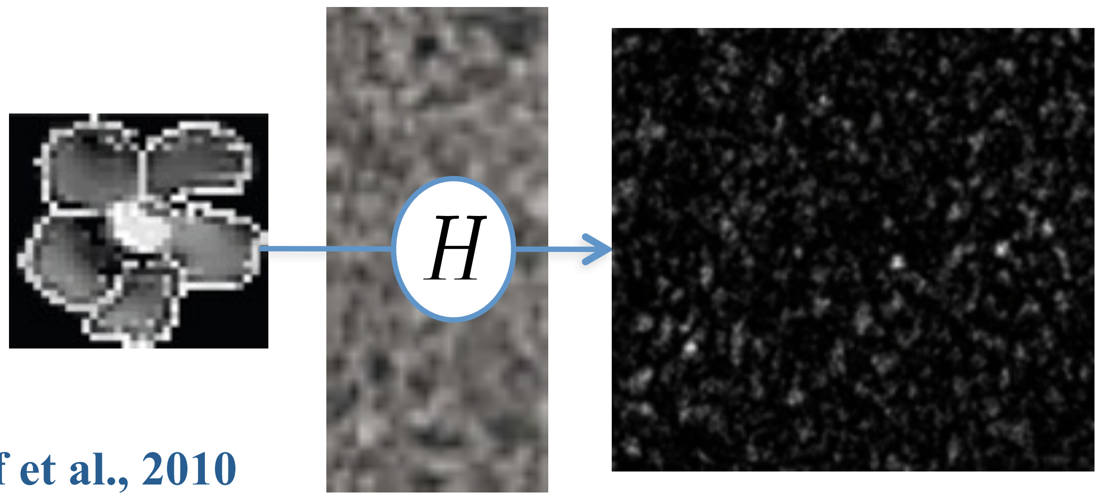


2 applications
- focusing
- imaging



Popoff et al. Phys. Rev. Lett. 104,100601 (2010)

Exploiting H for imaging



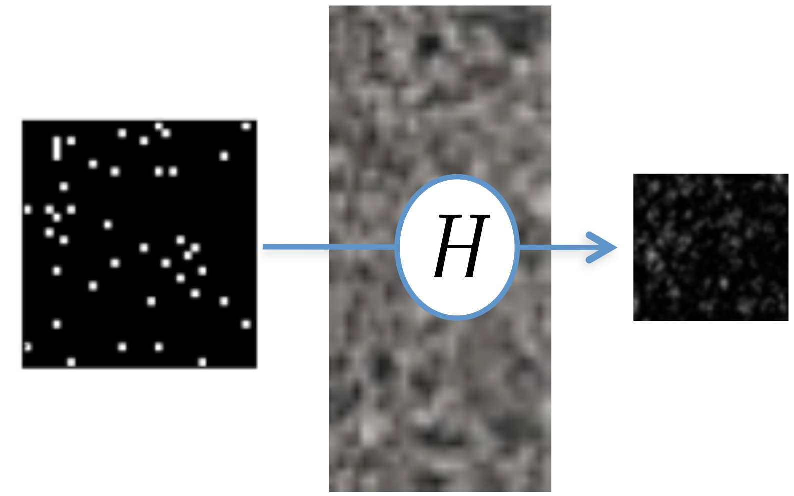
Popoff et al., 2010

Linear Reconstruction

Tikhonov

$$(H^{\dagger}H + \sigma I)^{-1} H^{\dagger}$$

Sparse image



Non-linear Reconstruction

Sparse reconstruction (l1 or l0)

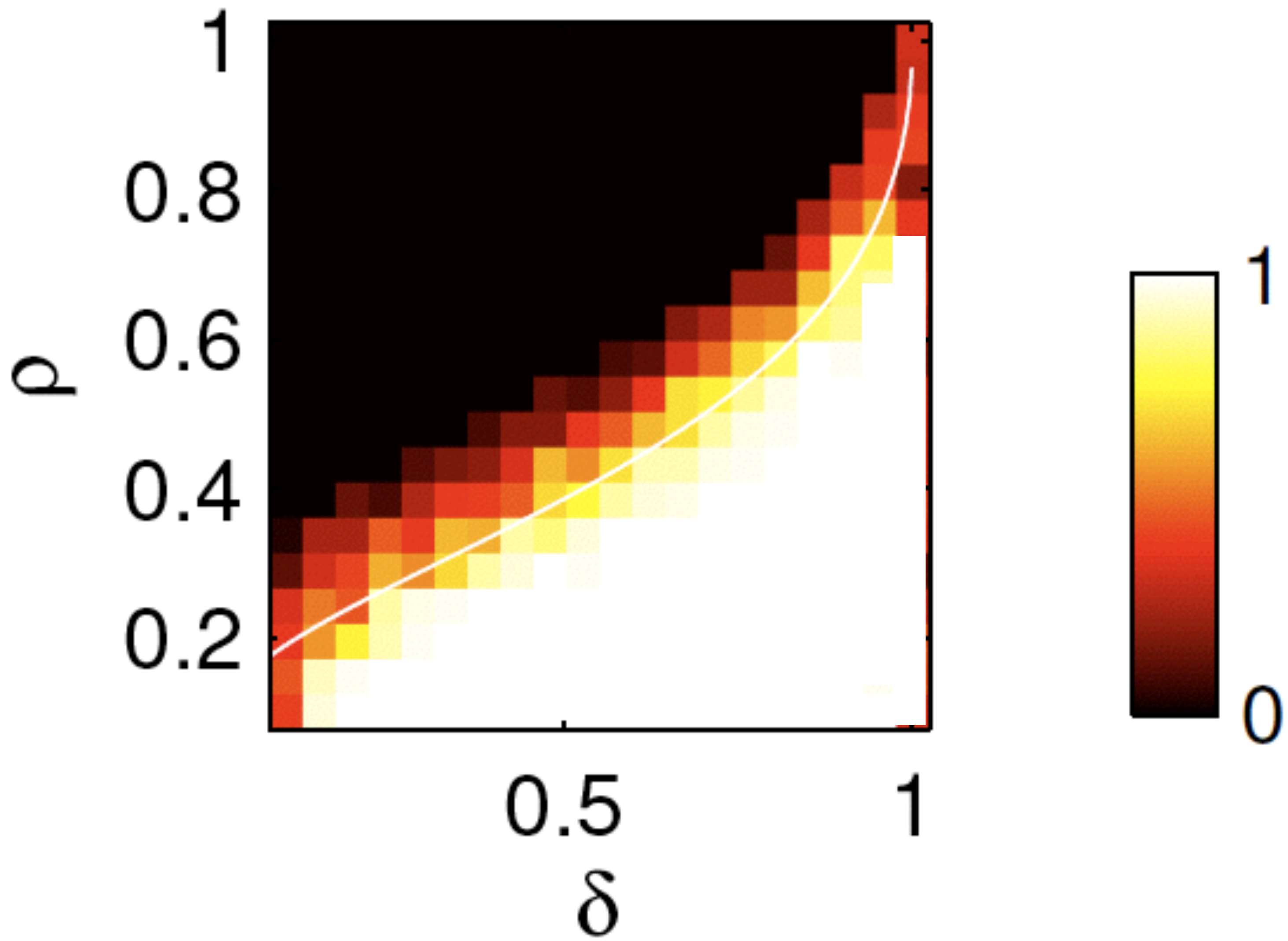
$$u = \arg\min_{u} ||x - H\,u||_2^2 + \lambda \|u\|_1$$

Gaussian iid measurements :
ideal for « compressed sensing » !

Conditions for CS reconstruction

For Gaussian measurement matrices, CS can be performed by L1-minimization, and "universal" behavior is observed (Donoho-Tanner)

sparsity s/m



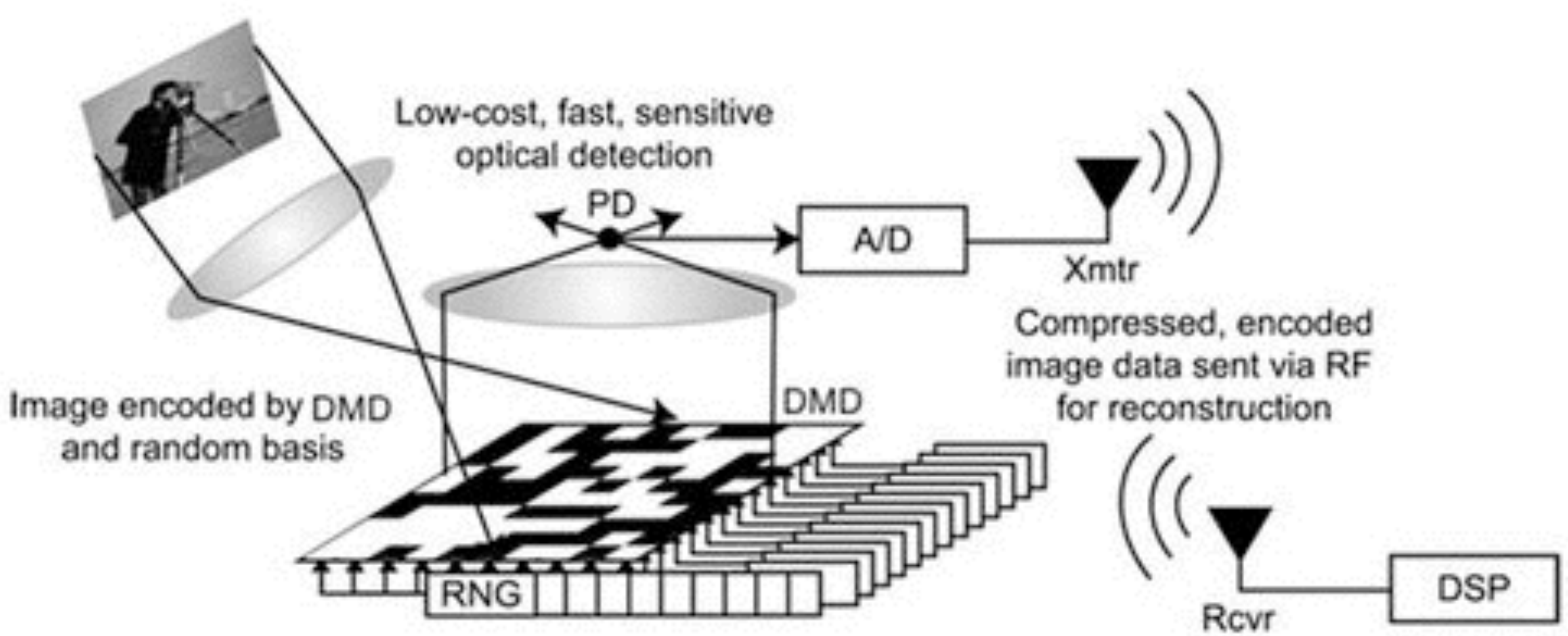
Probability of success

undersampling ratio m/N

Phase transition

The one-pixel camera

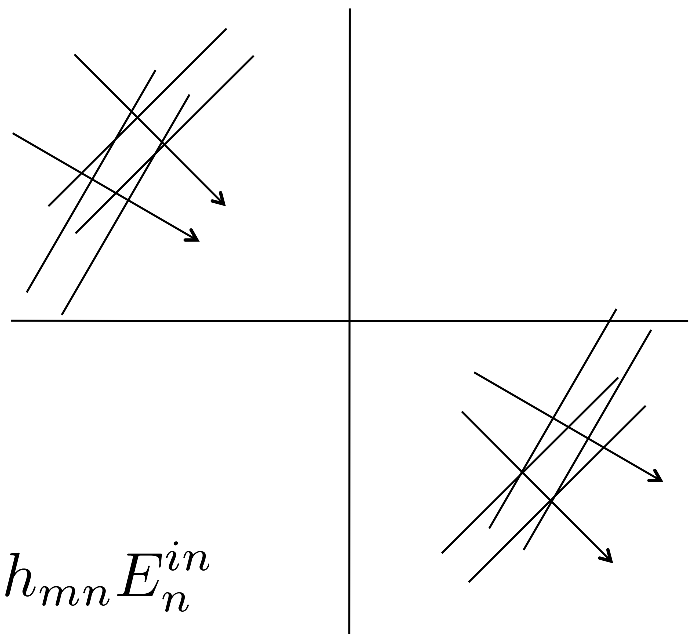
If natural images are sparse, are there smarter sampling schemes than 20 Mpixel regular sensors as in digital cameras ? (where 99% of images end up as JPEGs)



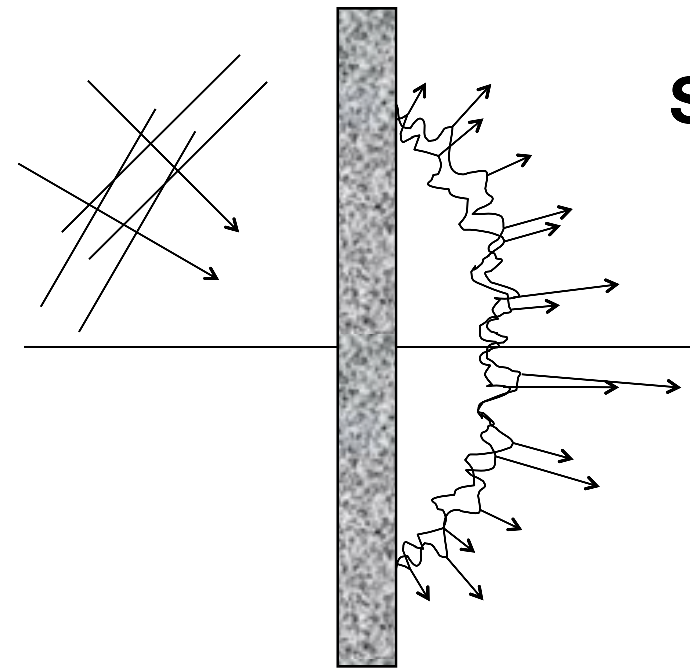
(Rice Univ.)

Back to the physical scattering processes

free field



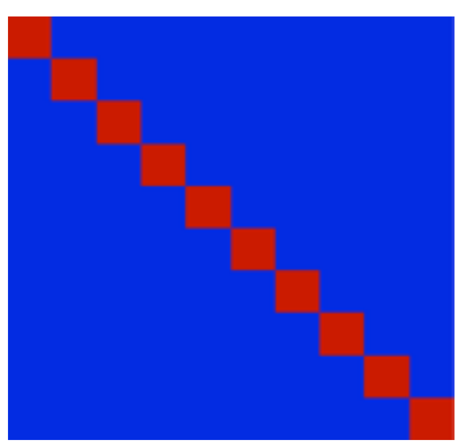
$$E_m^{out} = \sum_n^{1..N} h_{mn} E_n^{in}$$



Scattering material

Input **k**

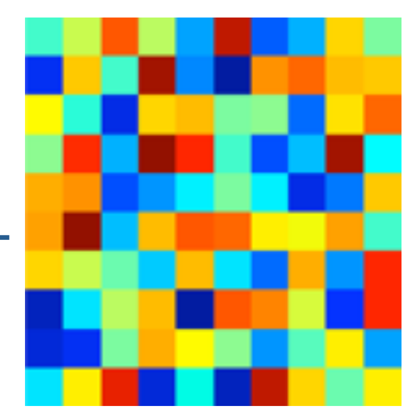
Output **k**



Identity Matrix

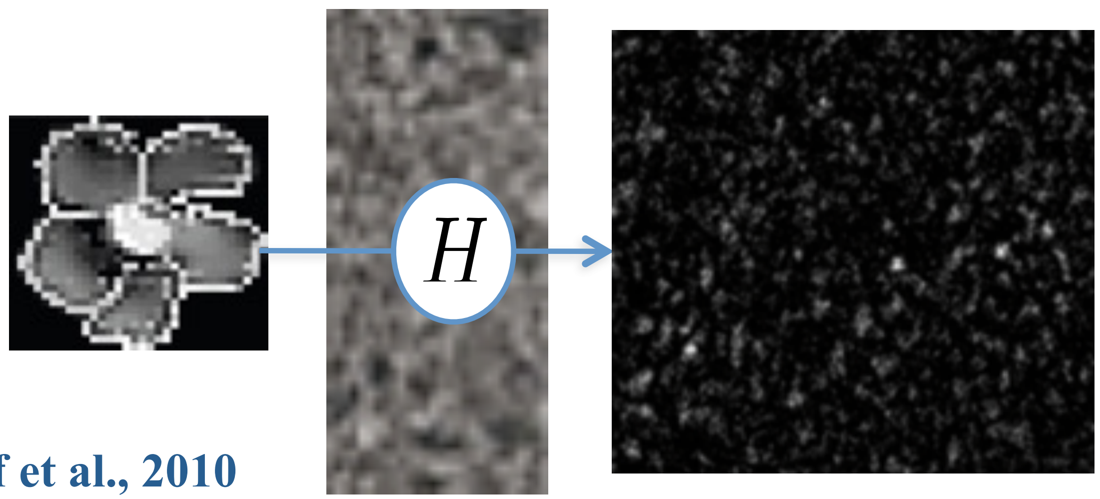
Input **k**

Output **k**



(Seemingly) Random Matrix

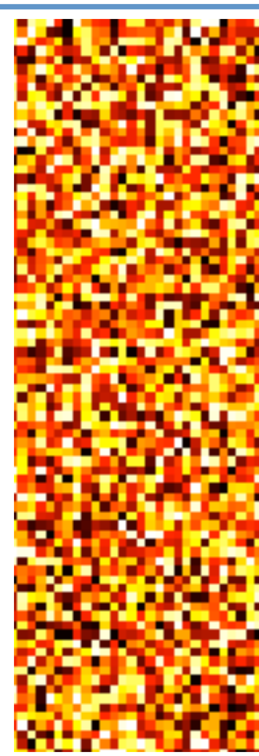
Exploiting H for imaging



Popoff et al., 2010

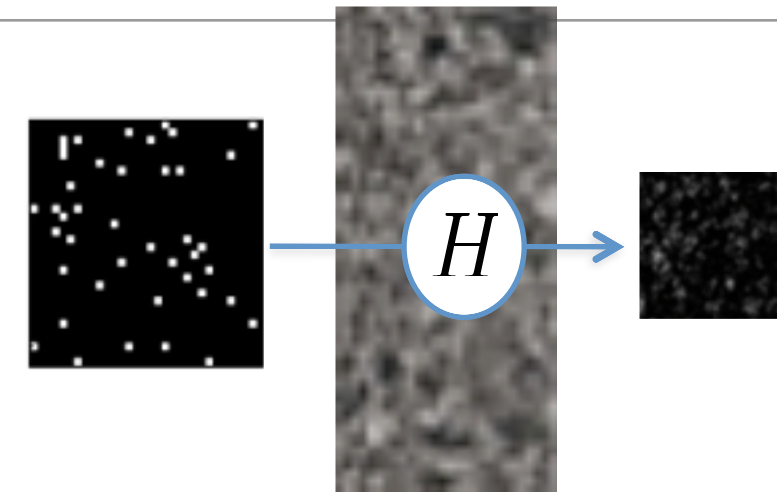
Linear Reconstruction

H



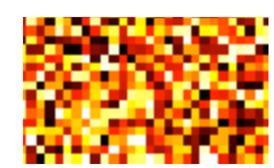
At least as many measurement pixels as input pixels

Sparse image



Non-linear Reconstruction

H

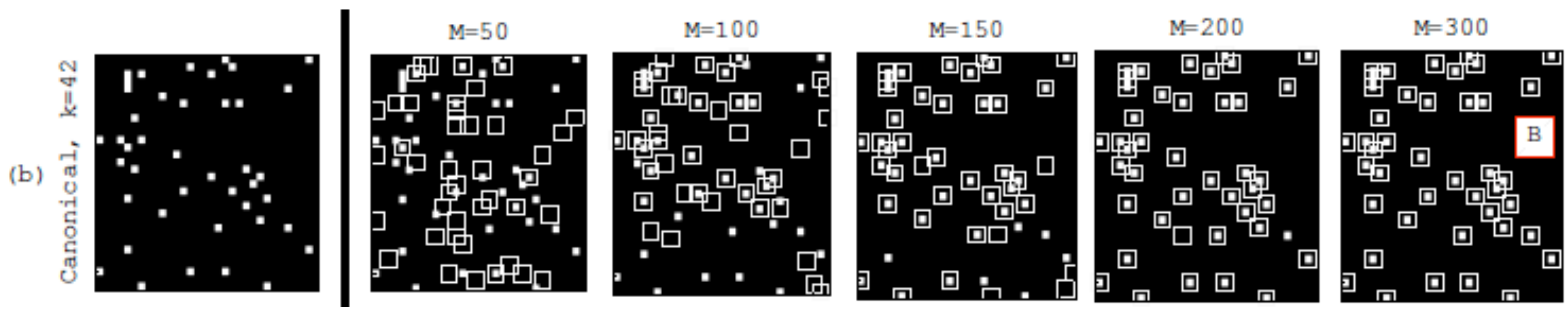


Number of measurement pixels driven by sparsity (<< input pixels)

Compressive imaging with scattering media

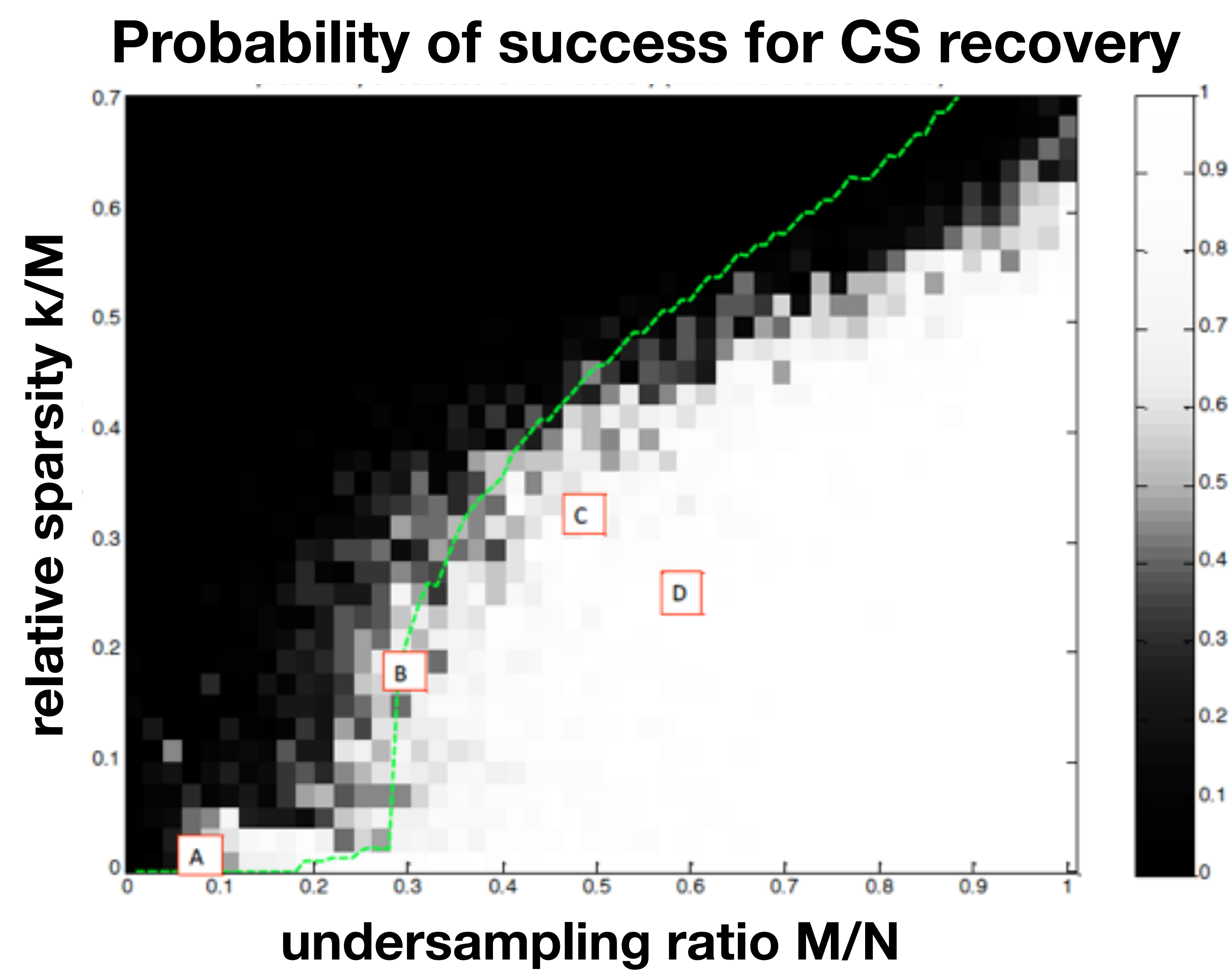
original image
1024 pixels

number of pixels M used for reconstruction



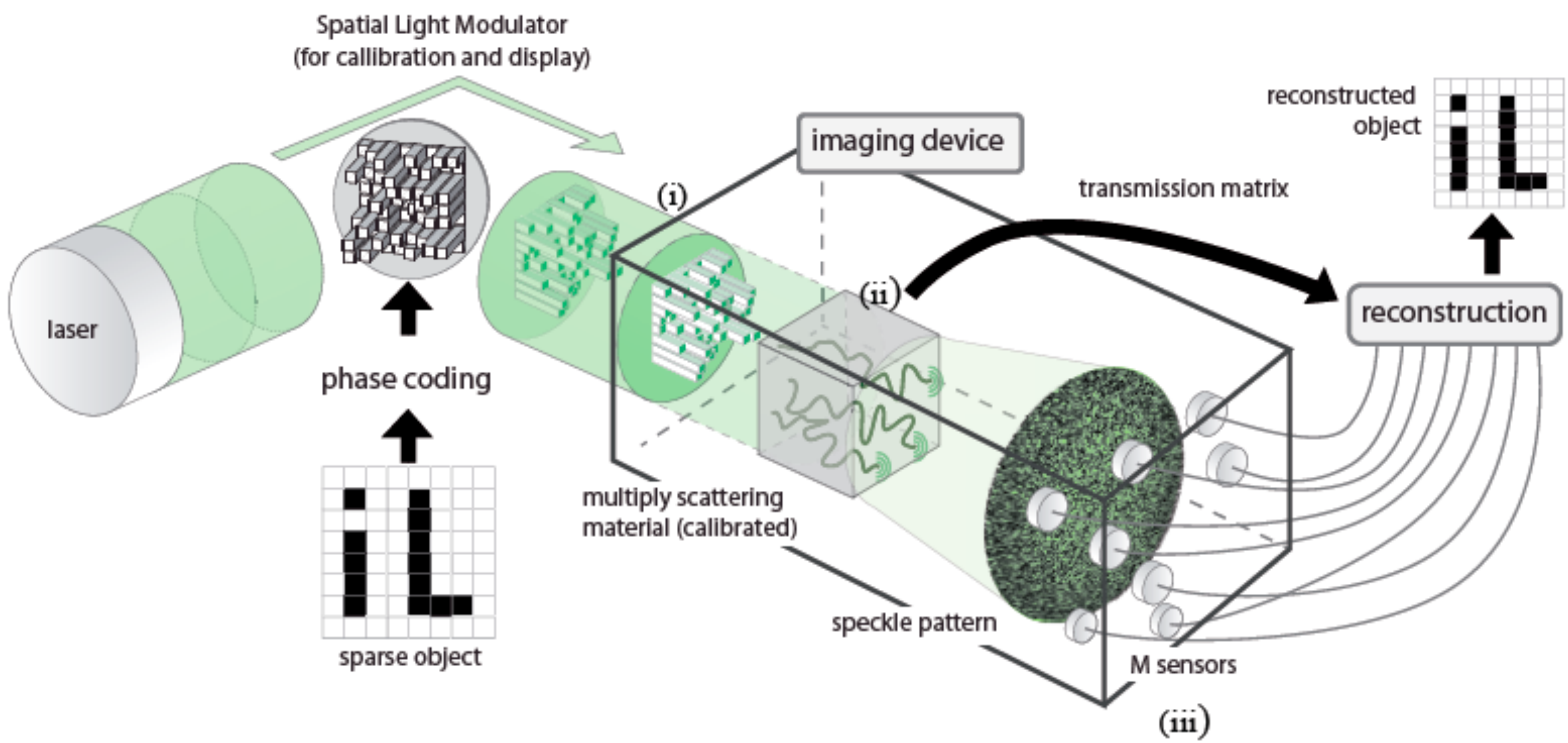
Each pixel provides information about the whole image

Compressive imaging with scattering media



about 10^5 experiments needed !
(measurements are fast !)

Compressive imaging with scattering media

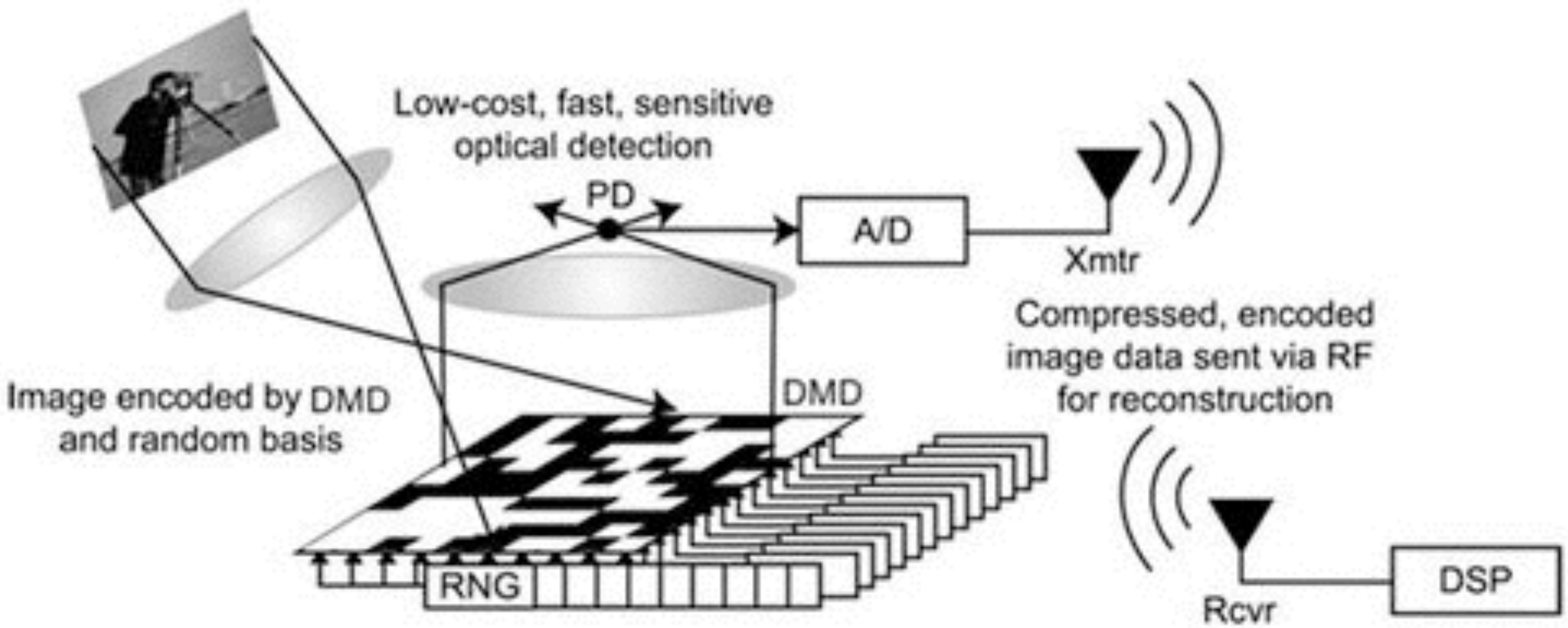


Proof of concept for compressive imaging with simple hardware

 Measurements are made in parallel : extremely fast

 Price to pay : calibration

The single-pixel camera



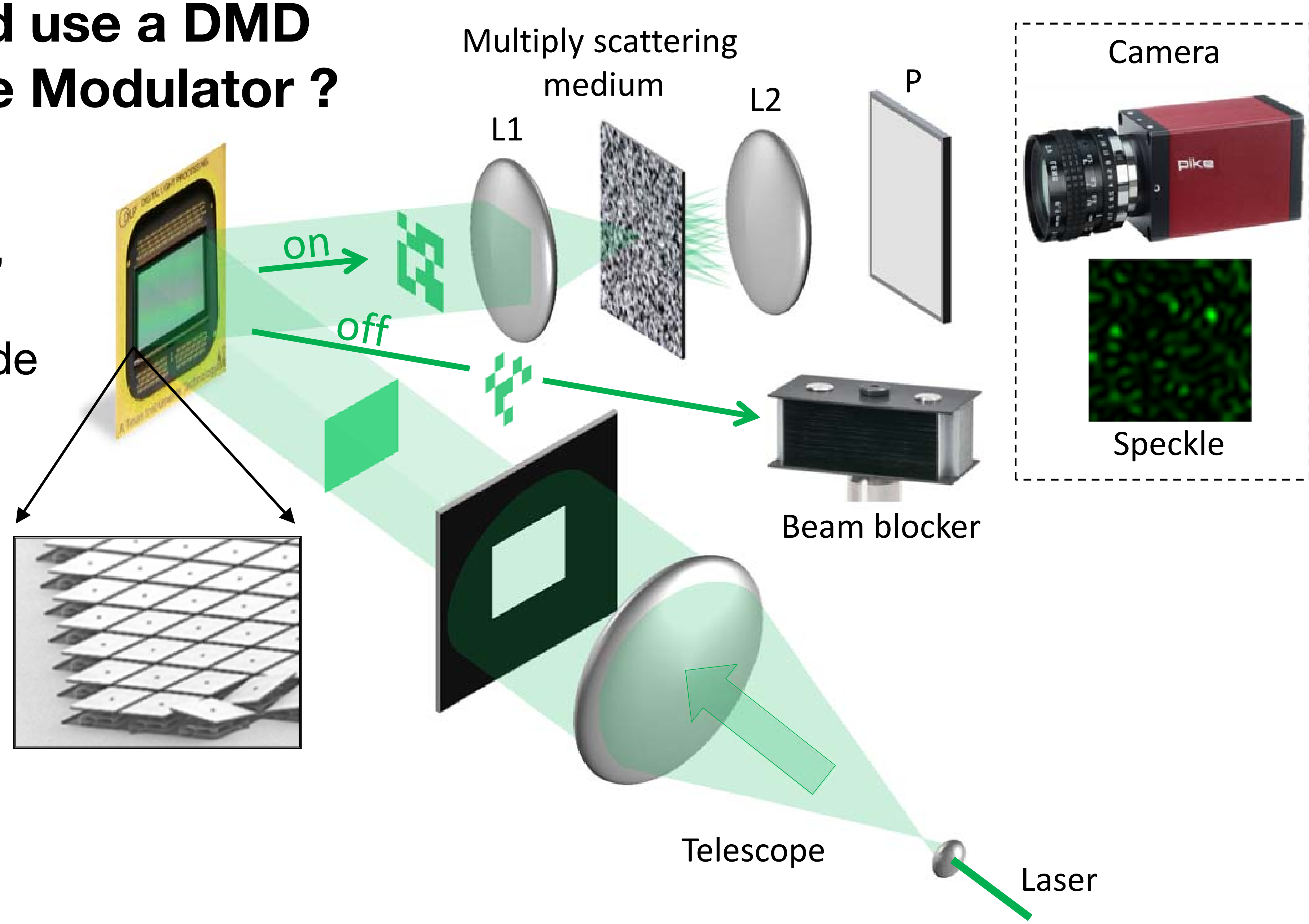
(Baraniuk team, Rice Univ.)

- Measurements are made sequentially : fundamentally slow
- But calibration is easier (pseudo-randomness)

Experimental setup

Can we instead use a DMD Binary Amplitude Modulator ?

High number of pixels, inexpensive, fast, but binary {0,1} amplitude modulation.



No phase control at input —> can only measure intensity | E_{out} |

Compressive imaging with scattering media

A double phase retrieval problem:

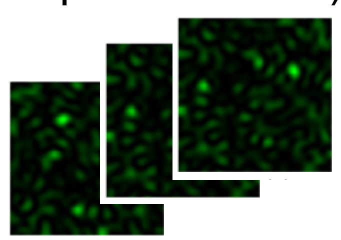
at calibration stage

$$\mathbf{y} = |\mathbf{H}\mathbf{x}|$$

output speckle image intensity (measured, known up to noise)

complex TM (unknown, iid gaussian)

real (binary) input (known)



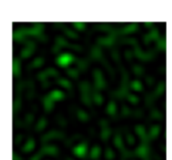
at imaging stage

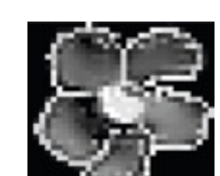
$$\mathbf{y} = |\mathbf{H}\mathbf{x}|$$

output speckle image intensity (measured, known up to noise)

complex TM (estimated at calibration)

real input (unknown)





Compressive imaging with scattering media

A double phase retrieval problem:

A new *single* phase retrieval algorithm
for both problems with different priors

prSAMP : phase retrieval with Swept Approximate Message Passing

- works well with binary {0,1} matrices
- computationally efficient
- flexible signal and noise priors
- Code + demo available (IPOL, Rajaei et al. 2017)

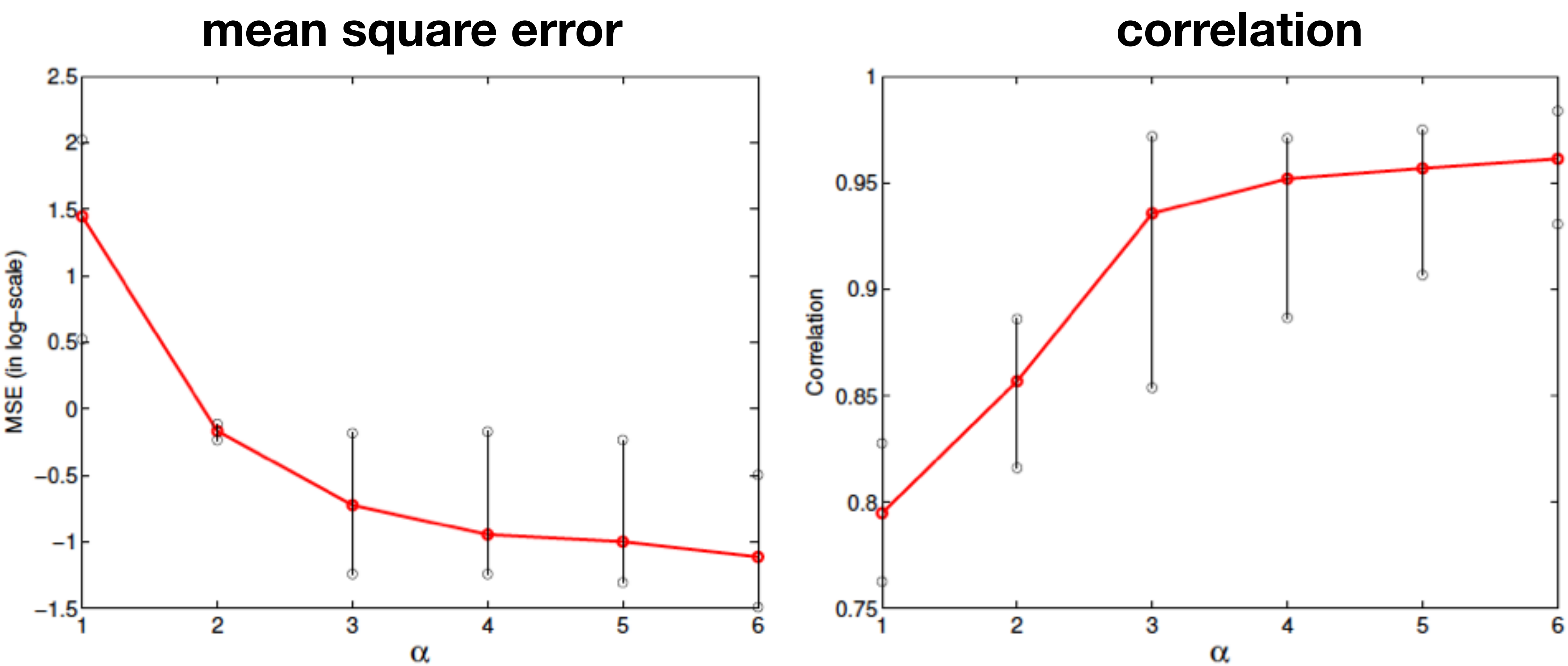
Compressive imaging with scattering media

4 ways to assess the performance of the calibration
(NB ground truth not available !)

- prediction error : for a known x_{in} , compare y_{out} to $| D_{est}\, x_{in} |$
- eigenvalue distribution
- focusing results
- imaging results

Compressive imaging with scattering media

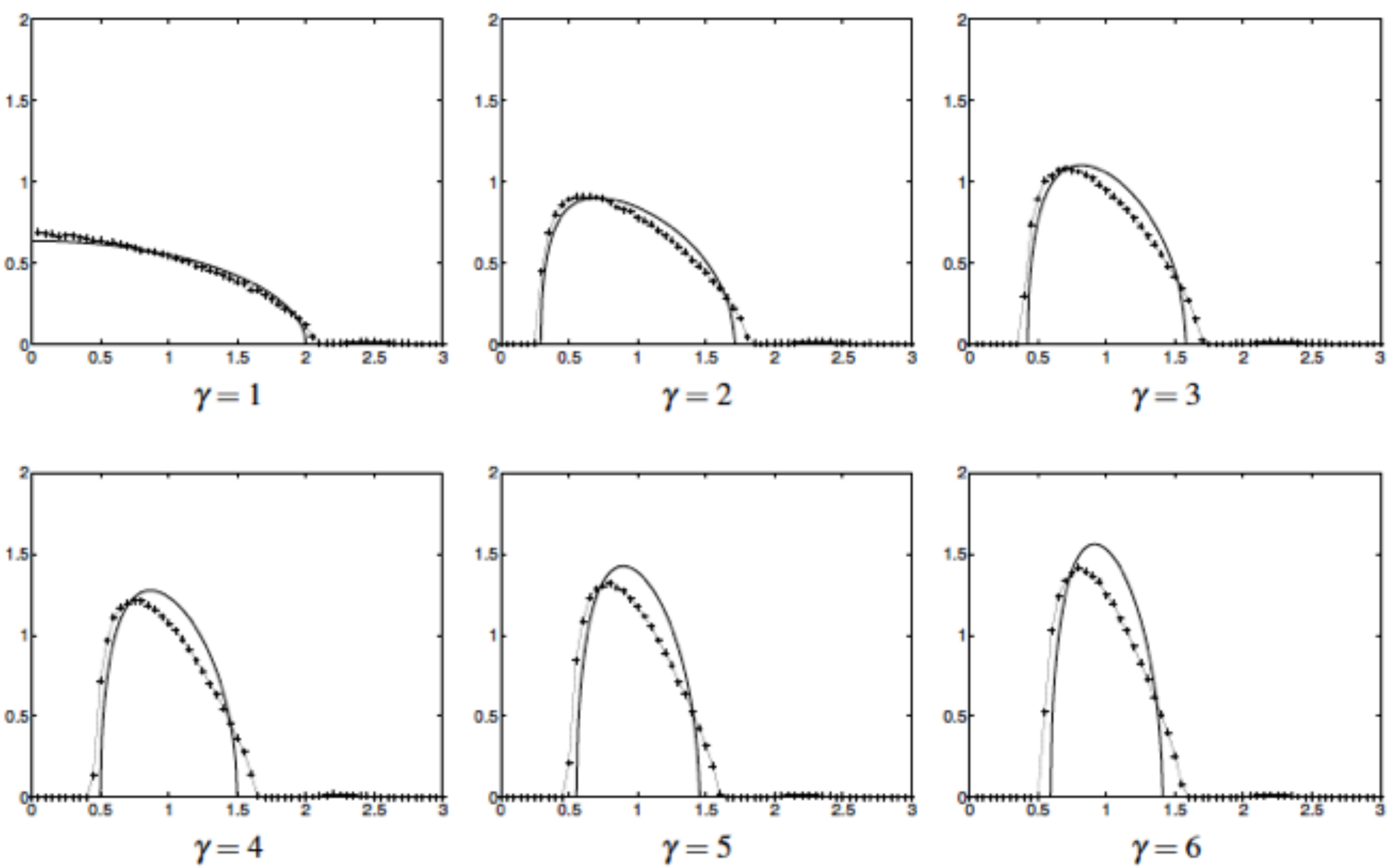
prediction error



number of training samples $P = \alpha N$

Compressive imaging with scattering media

Distribution of eigenvalues



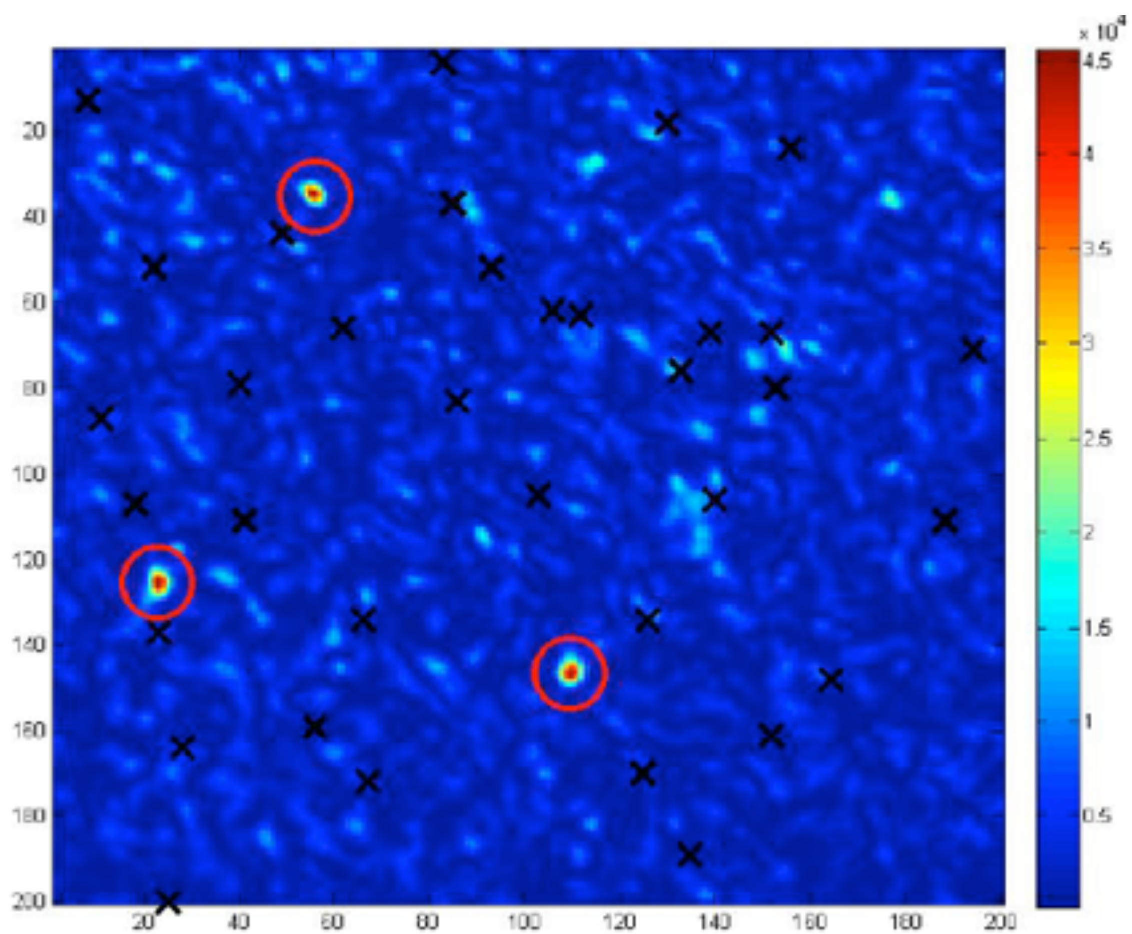
Distribution of normalized eigenvalues, comparison with random matrix theory (Marcenko-Pastur), for different undersampling ratio

Compressive imaging with scattering media

Light focusing

Goal : use estimated D to focus light on output plane

What is the best *binary* x as input, to focus on only a few target output pixels ?



one can use :

- binarized phase conjugation

$$\hat{\mathbf{x}} = \left[\Re(\mathbf{D}^H \mathbf{y}) > 0\right]$$

- or the same bayesian model used for calibration, here particularized for the estimation of x, with a binary prior.

Compressive imaging with scattering media

Imaging results

original

reconstructed

- Better images can be obtained with more precise signal priors
- Larger images raise significant computational issues

ex for 128x128 images: H is 10^5 x 10^5 (5 GB in memory)

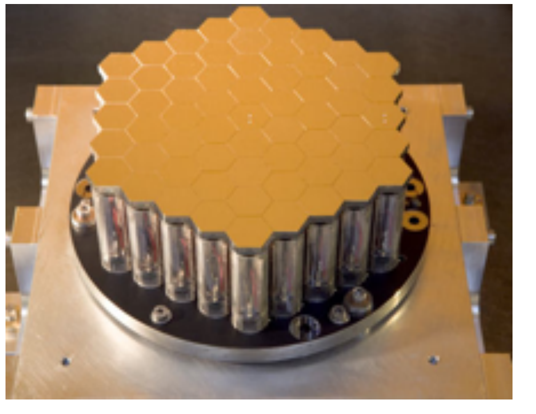
Phase retrieval algorithms do not scale well : see preprint
Fast phase retrieval in high dimension : a block-based approach,
B. Rajaei et al, arXiv:1602.02944

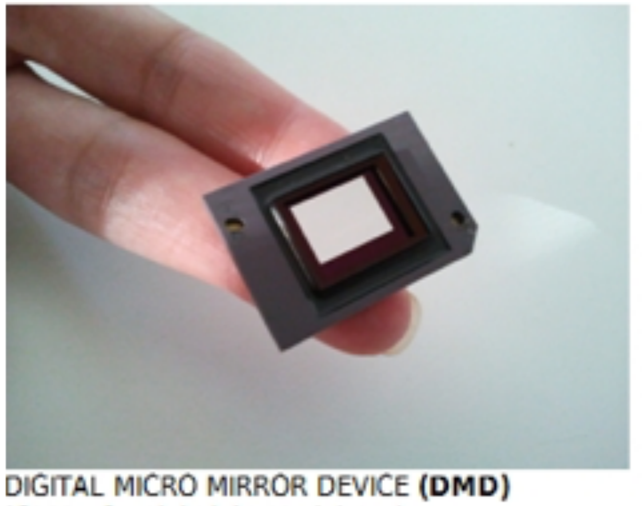
Compressive imaging with scattering media

(Compressive) imaging through strongly scattering material is possible thanks to **wavefront shaping**

with expensive & low res. SLM,
8 shots / sparse image

(our previous studies)





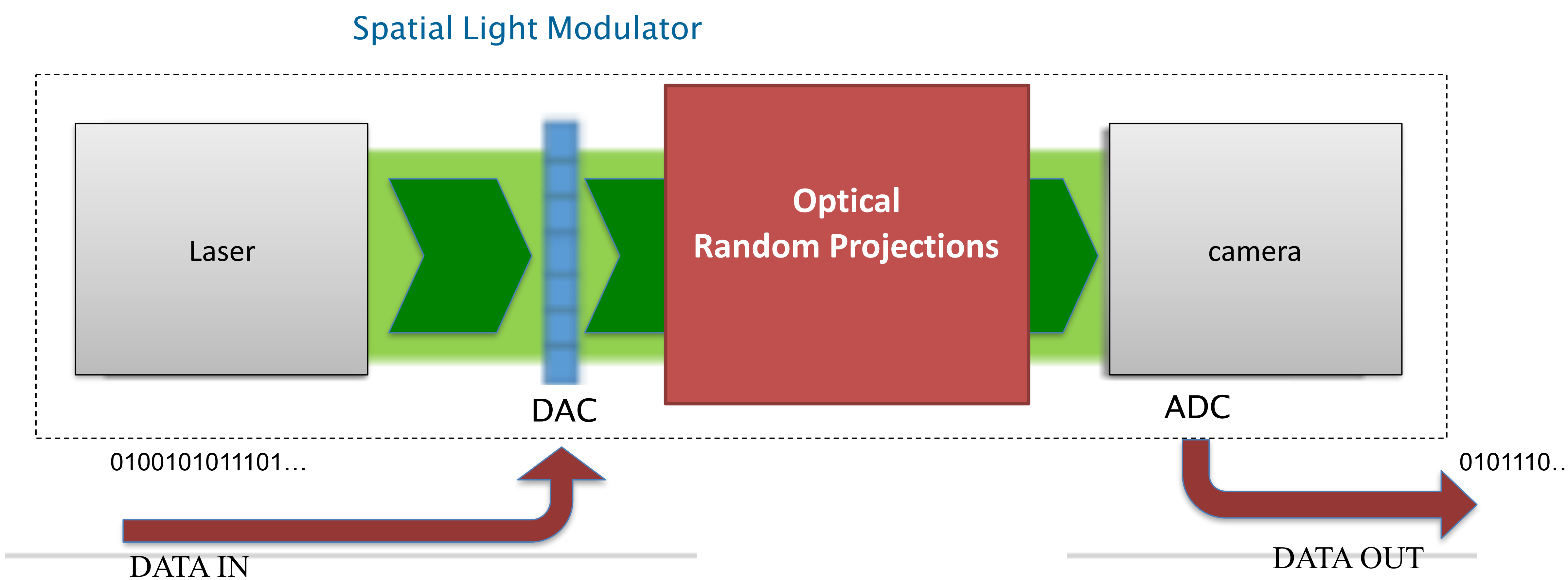
DIGITAL MICRO MIRROR DEVICE **(DMD)**
(**SLM** - Spatial Light Modulator)

with cheap & high res. DMD,
1 shot / image !
price : robust and scalable PR
new algorithm : prSwAMP

« Ask not what computing can do for optics – ask what optics can do for computing »

Towards optical computing

Now, let us just only consider the previous experiment as a "black box" with input in the SLM and output on the CCD



Towards optical computing

- These components can be very fast (kHz), with high pixel counts (10 Mpix)
  ▷ potentially 10 Gbs total throughput

- This simulates the operation y = |Mx| with M a *complex* random iid matrix **of size potentially 10^7 x 10^7** (TBs of memory)

- The key idea : if you just want to *compare* outputs **you do not have to know (calibrate) and store this matrix**

- This study presents a very simple proof-of-concept of image classification based on kernel ridge regression, where the random features are obtained with the optical experiment.

Example : classification with ridge regression on random features

training

U : data Y: labels

$$\underset{\beta \in \mathbb{R}^{p \times q}}{\text{argmin}} \ ||\mathbf{U}\beta - \mathbf{Y}||_2^2 + \gamma ||\beta||_2^2$$

$$\beta = (\mathbf{U}^T\mathbf{U} + \gamma \mathbf{I}_p)^{-1}\mathbf{U}^T\mathbf{Y} = \mathbf{U}^T(\mathbf{U}\mathbf{U}^T + \gamma \mathbf{I}_n)^{-1}\mathbf{Y}$$

regression

$$\begin{aligned} \tilde{\mathbf{Y}} = \tilde{\mathbf{U}}\beta &= \tilde{\mathbf{U}}(\mathbf{U}^T\mathbf{U} + \gamma \mathbf{I}_p)^{-1}\mathbf{U}^T\mathbf{Y} \\ &= \tilde{\mathbf{U}}\mathbf{U}^T(\mathbf{U}\mathbf{U}^T + \gamma \mathbf{I}_n)^{-1}\mathbf{Y} \end{aligned}$$

These are only inner products

inverting this N x N matrix can be hard

use a **kernel** for these inner products

$$\mathbf{K}_{i,j} = k(\mathbf{U}_i, \mathbf{U}_j) \ \text{ and } \ \tilde{\mathbf{K}}_{i,j} = k(\tilde{\mathbf{U}}_i, \mathbf{U}_j)$$

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{K}}(\mathbf{K} + \gamma \mathbf{I}_n)^{-1}\mathbf{Y}$$

Kernel ridge regression

Consider the following elliptic kernel (EK)

$$k(\mathbf{U}_i, \mathbf{U}_j) = \frac{\sqrt{\mathbf{U}_i^T \mathbf{U}_i \, \mathbf{U}_j^T \mathbf{U}_j}}{2} \left\{ -(\sin^2\theta)\mathcal{E}_K\left[\cos^2\theta\right] + 2\mathcal{E}_E\left[\cos^2\theta\right] + |\sin\theta| \left(2\mathcal{E}_E\left[-\frac{\cos^2\theta}{\sin^2\theta}\right] - \mathcal{E}_K\left[-\frac{\cos^2\theta}{\sin^2\theta}\right] \right) \right\}$$

$\mathcal{E}_K[.]$ and $\mathcal{E}_E[.]$ are the complete elliptic integrals of the first / second kind

θ is the angle between $\mathrm{U}i$ and $\mathrm{U}j$

<u>Example</u> : classifying the MNIST database

training set of 60000 training pictures (28x28) of handwritten digits

test set of 10000 digits

Using the EK, one obtains a 1.31 % error rate

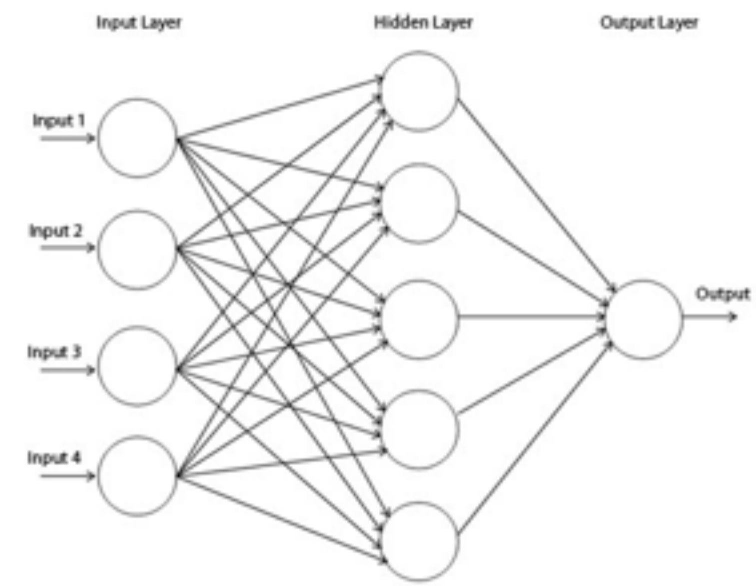
(baseline 12 % with plain ridge regression)

Approximating kernels with random projections

In the spirit of Rahimi-Recht / ELMs :

$$\mathbf{X}_{i,j} = \phi((\mathbf{W}\mathbf{U}_i)_j + \mathbf{b}_j)$$

where W is a random complex matrix with gaussian i.i.d. entries, and ϕ a non-linearity



$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T + \gamma \mathbf{I}_n)^{-1}\mathbf{Y} = \tilde{\mathbf{X}}(\mathbf{X}^T\mathbf{X} + \gamma \mathbf{I}_N)^{-1}\mathbf{X}^T\mathbf{Y}$$

of size N x N

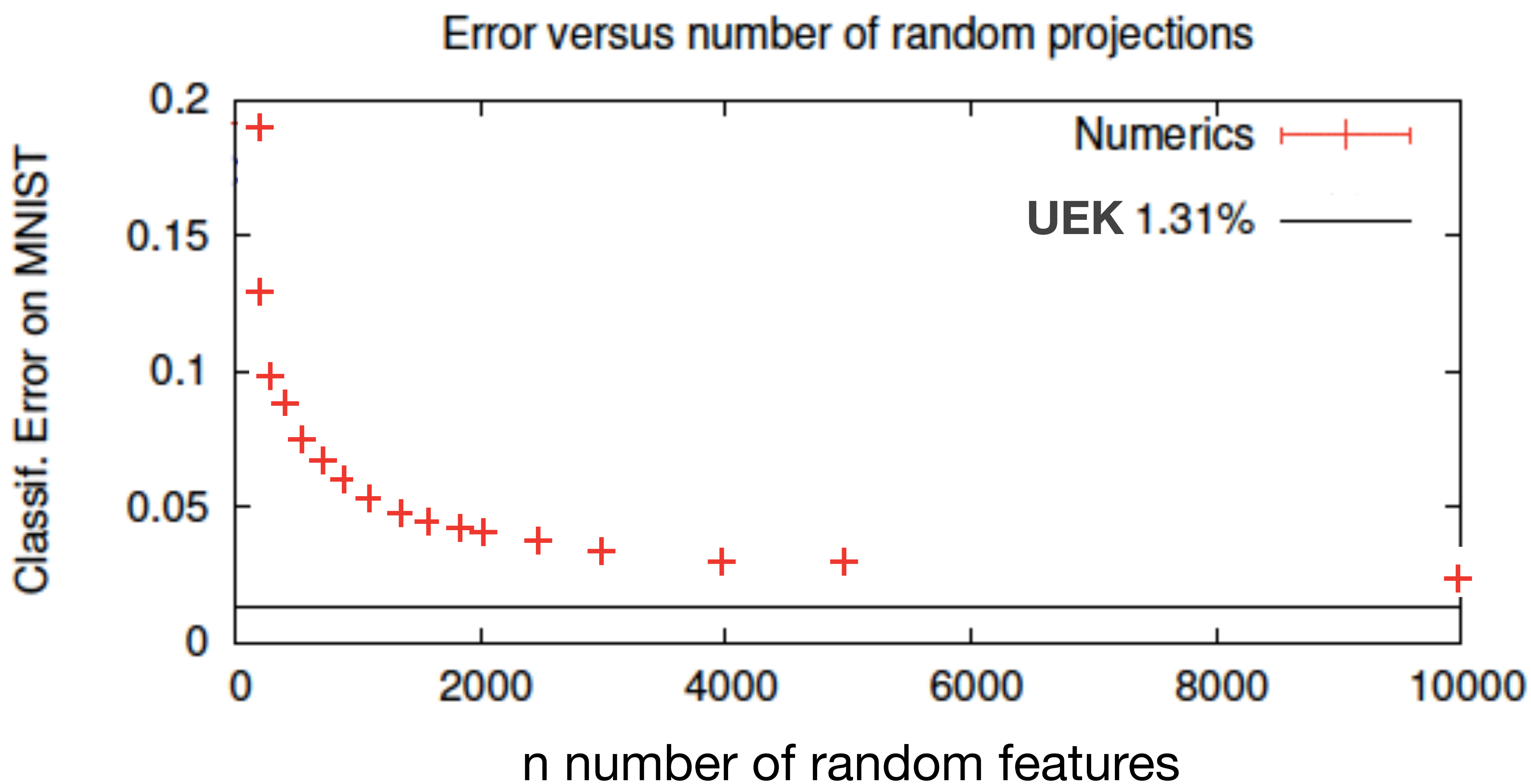
of size n x n

where N is the number of training examples

where n is the number of random features
no dependency on N !

For b = 0, ϕ | |, as n —> infinity, this tends to the Elliptic Kernel above !

Random projections



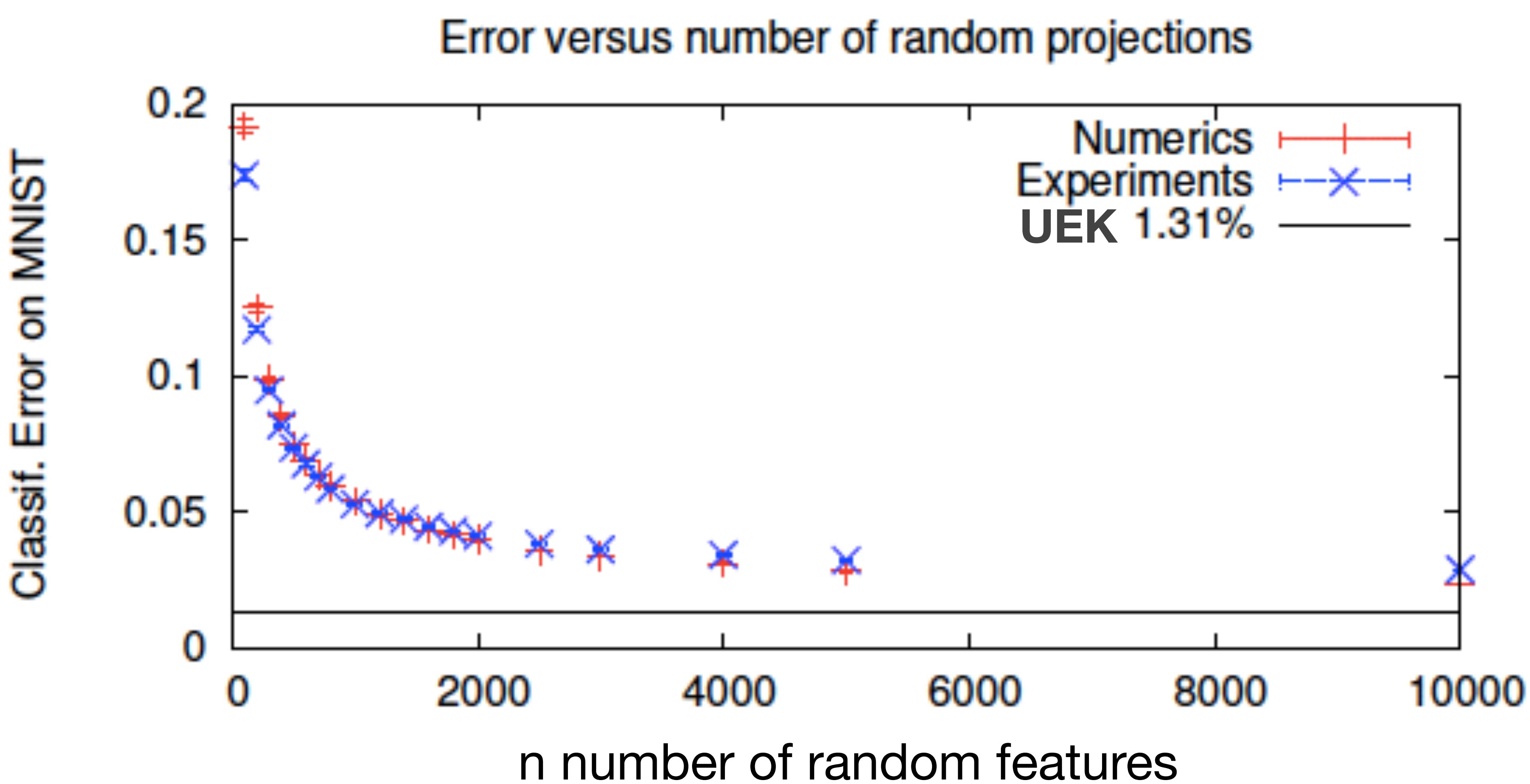
example : at n = 10000 random features, error rate is about 2%

empirical scaling law in $N^{-2/3}$

Random projections

$$\mathbf{X}_{i,j} = \phi((\mathbf{W}\mathbf{U}_i)_j + \mathbf{b}_j)$$

this is precisely what is performed by our optical experiment with $\phi = |\ \ |$



Random projections

We optically perform an operation that approximates the norm of random projections with complex-valued iid Gaussian entries

$$\mathbf{X}_{i,j} = \phi((\mathbf{W}\mathbf{U}_i)_j + \mathbf{b}_j)$$

when the number of output random features tends to infinity, this amounts to computing a kernel   (ugly but well behaved !)

that efficiently feeds a simple linear classifier (kernel ridge regression)

1st experimental proof of concept on a small dataset : needs to be confirmed in large-scale experiments

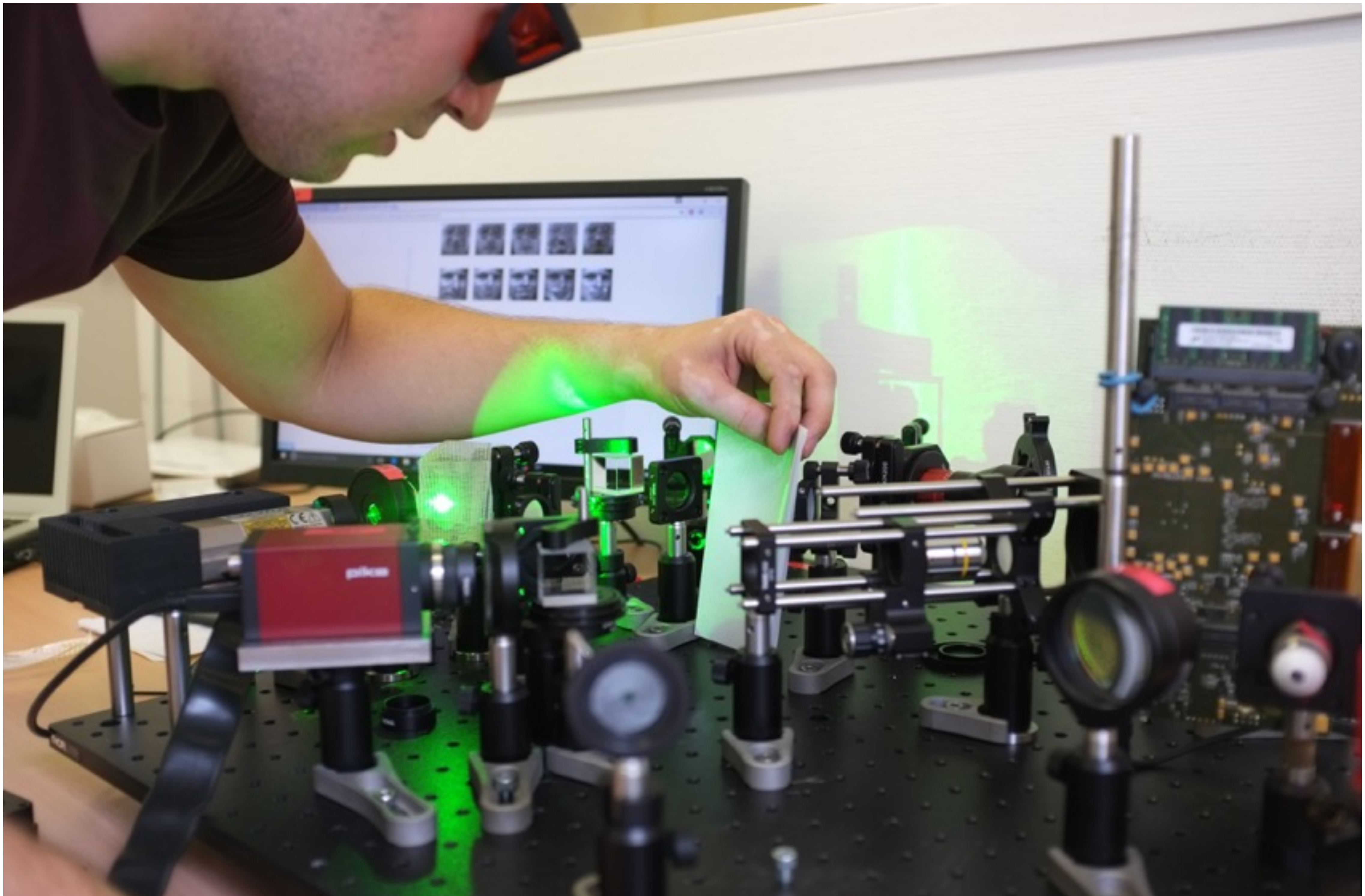
Conclusion

Multiply scattering media provides «optimal / universal» scrambling of information in a fully scalable *analog* way.

- **computational imaging:**
  turning a layer of scattering material into a « super-lens »
  From *imaging through scattering media* (challenge)
  to *using scattering media to better image* (opportunity)

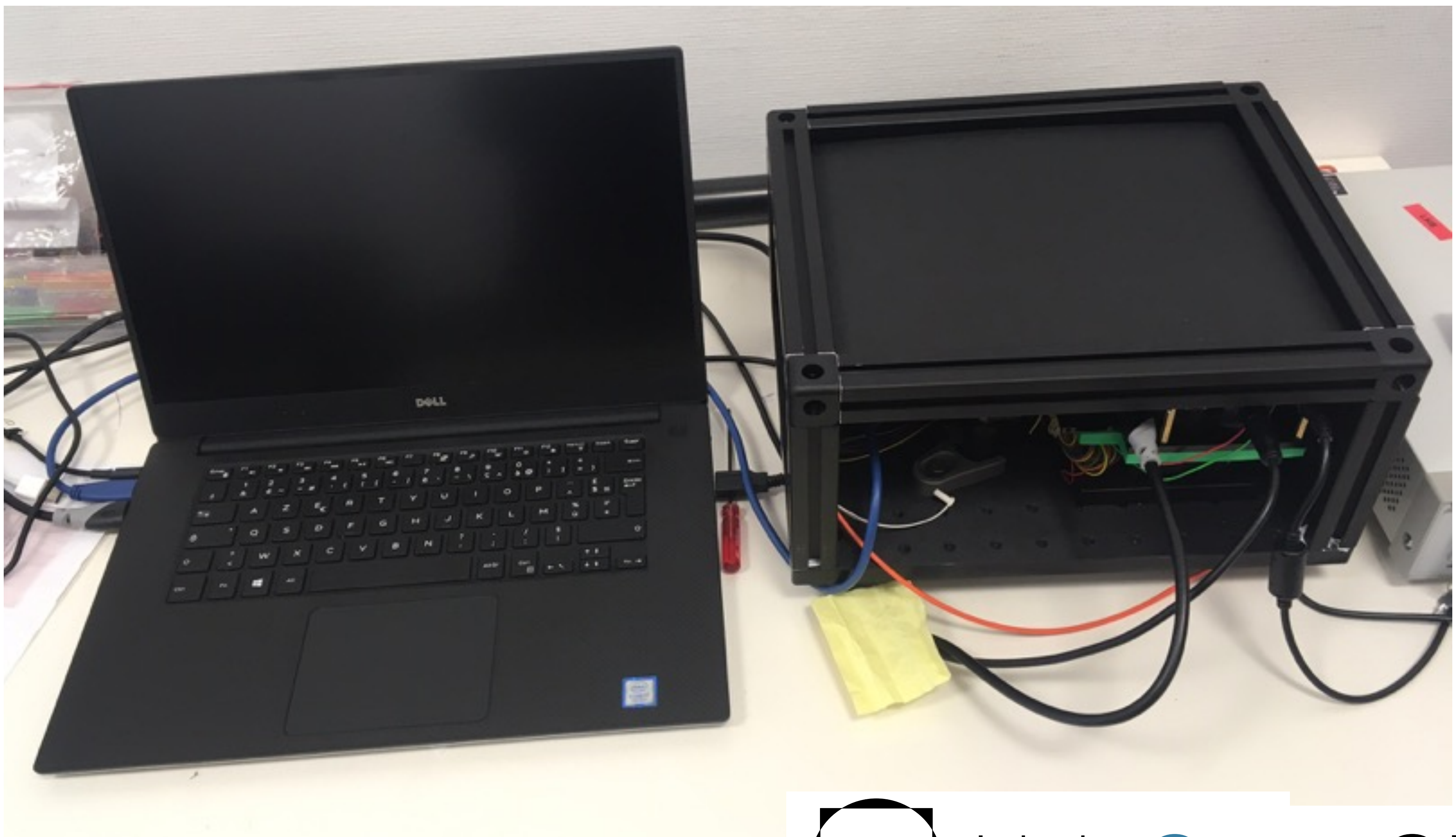
- **optical computing**
  turning a layer of scattering material into a « computer »

From lab experiment to prototype



From lab experiment to prototype

Using only off-the-shelf components



LightOn « OPU »

From lab experiment to prototype

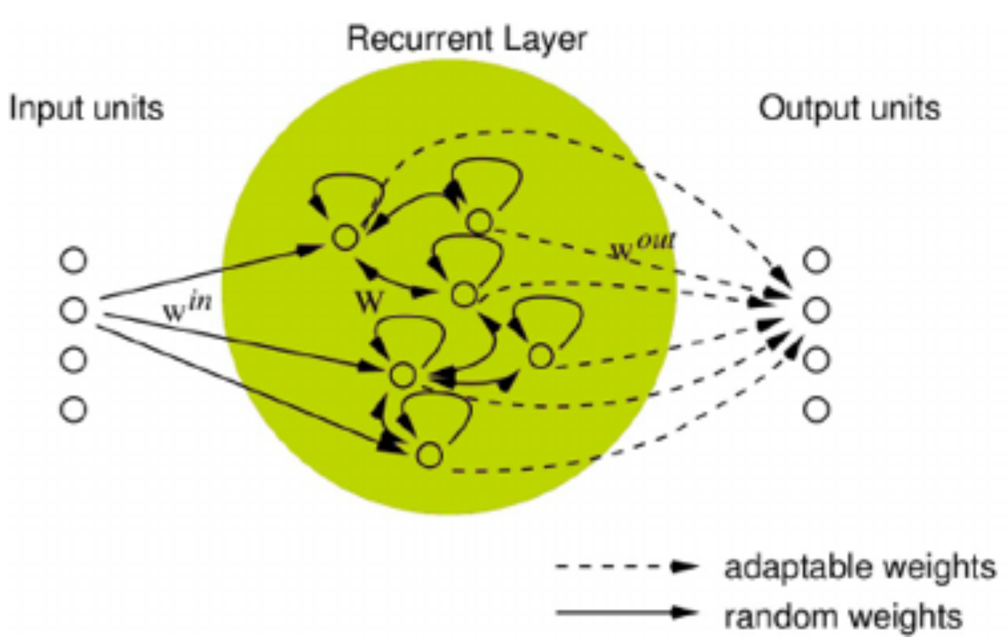
Current prototype performs random projections 800 times faster than CPU at the max size handled by RAM, it can also go at much larger sizes

LightOn « OPU »

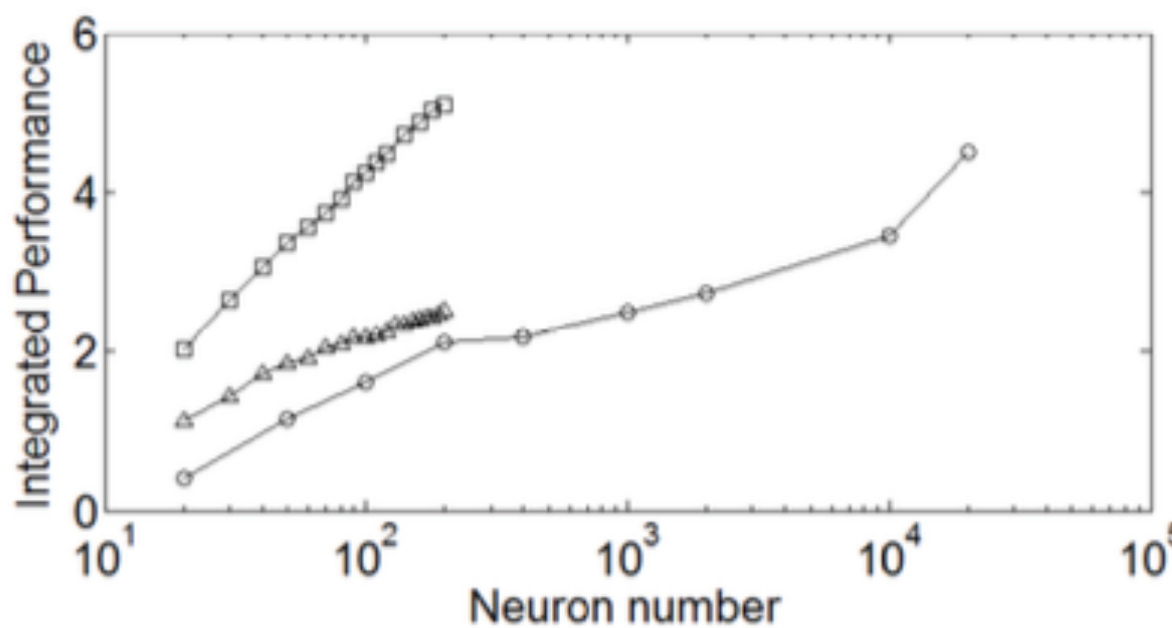
Reservoir computing

Our prototype can be used as a physical implementation of large-scale echo-state networks

POC experiment with the XOR operation *[Dong et al., arXiv:1609.05204]*



[diagram from Obst et al. 2013]



Can make ESNs at sizes not reachable by standard PCs

From lab experiment to prototype

We are investigating random projections for a number of Machine Learning schemes <u>at scale</u>:

- Supervised / unsupervised schemes
- Feed-forward / recurrent
- dimensionality reduction
- …

From lab experiment to prototype

Soon (Q1 2018) available in the cloud with CPU/GPU for the beta-users (you !) to play with.

Conclusion

- Large random matrices can be found in Nature : « easily » harvested !
- « What the fly actually does is, instead of reducing it, it expands the dimension into much larger than it was [using Random Projection], and it creates a very sparse point in a high-dimensional space » (Brian Gallagher blog post, on a study by Navlakha *et al.*)

LightOn

- We are hiring (jobs/internships) - located in the center of Paris !
- Can provide remote access to our cloud-computing platform

www http://www.LightOn.io

laurent@LightOn.io

@Laurent_Daudet

Selected references

- "Imaging With Nature: Compressive Imaging Using a Multiply Scattering Medium", A. Liutkus et al., *Scientific Reports 4 (july 2014)*
- "Reference-less measurement of the transmission matrix of a highly scattering material using a DMD and phase retrieval techniques", A. Drémeau et al., *Optics Express 23(9), 2015*
- "Random Projections through multiple optical scattering: Approximating kernels at the speed of light", A. Saade et al., *Proc. ICASSP (2016)*
- « Scaling up Echo-State Networks with multiple light scattering », J. Dong et al., *arXiv: 1609.05204*