“Not just your usual BP”: Making it work in two examples

Florent Krzakala
ESPCI, PCT and Gulliver CNRS

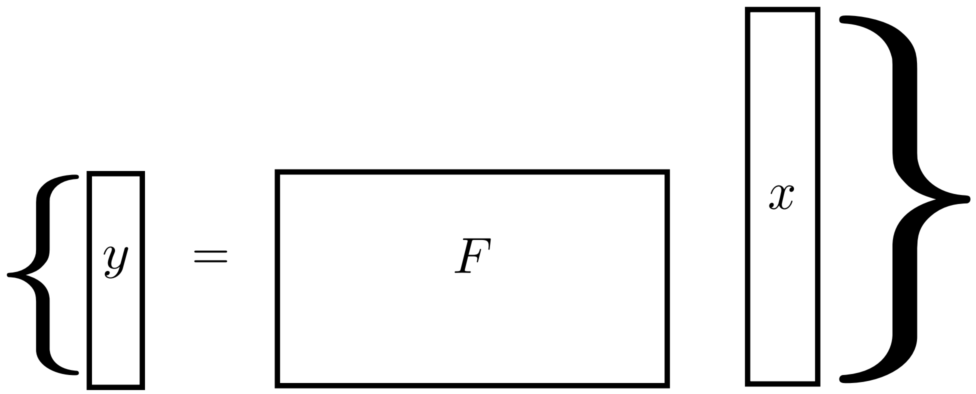
in collaboration with
Emmanuelle Gouillart (St Gobain & CNRS)
Marc Mézard (ENS) & **François Sausset** (St Gobain)
Yifan Sun (ESPCI & Beihang) and **Lenka Zdeborová** (IPhT Saclay)

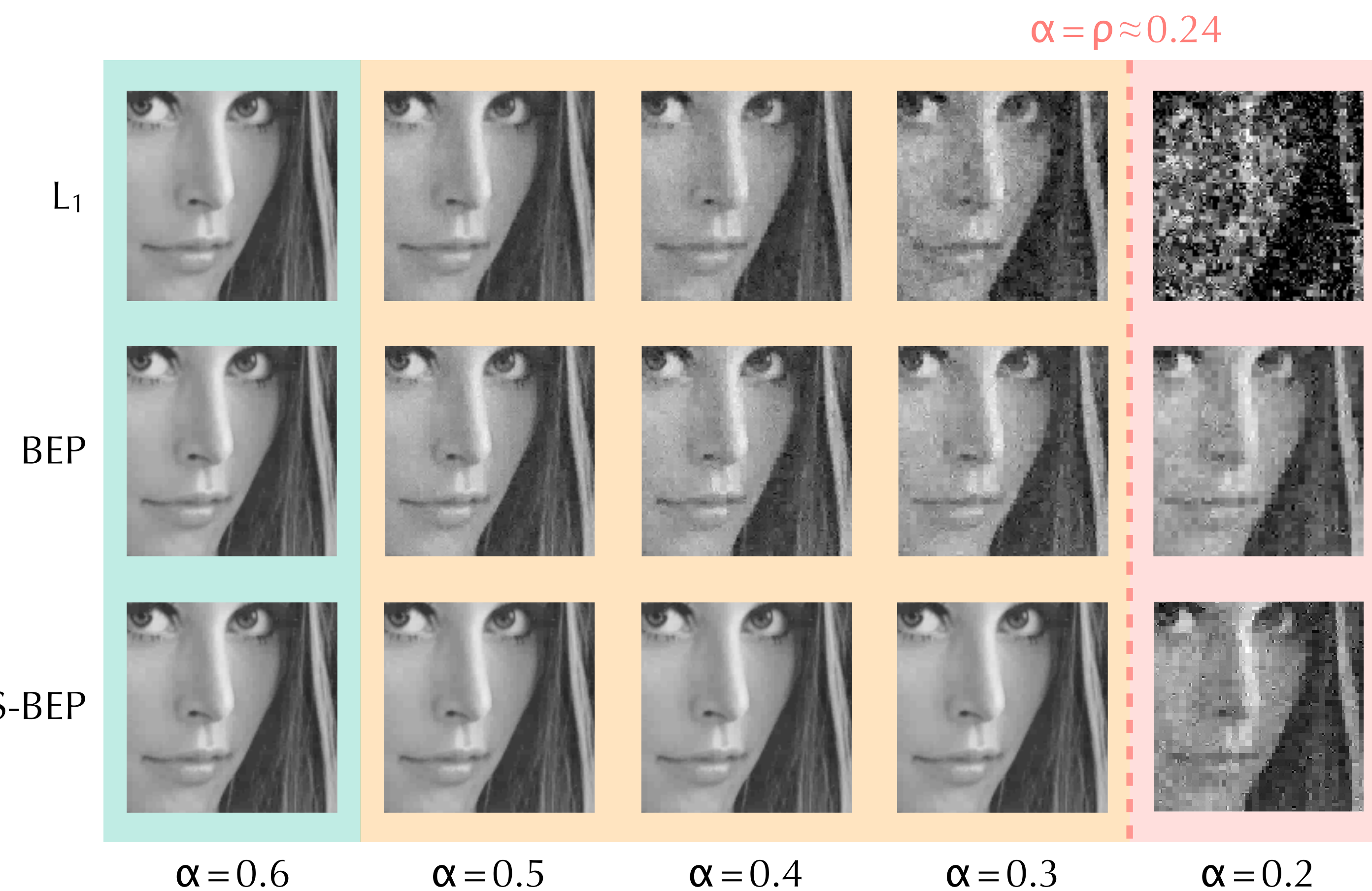
http://aspics.krzakala.org

Phys. Rev. X 2 021005 (2012) (open access)

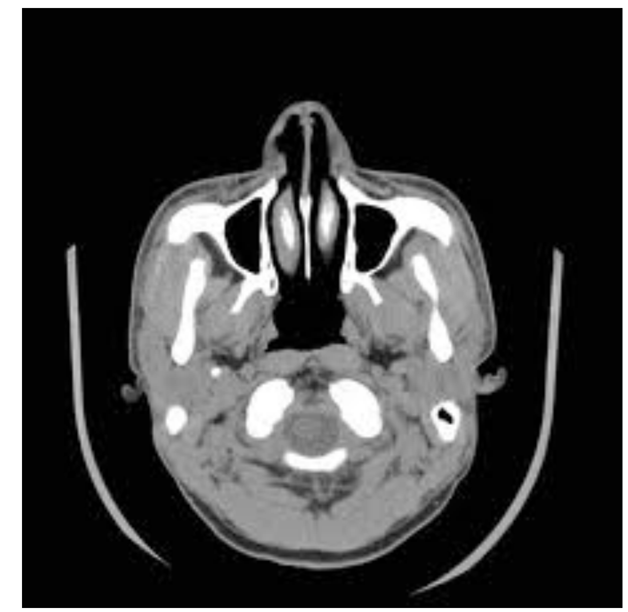
Two problems

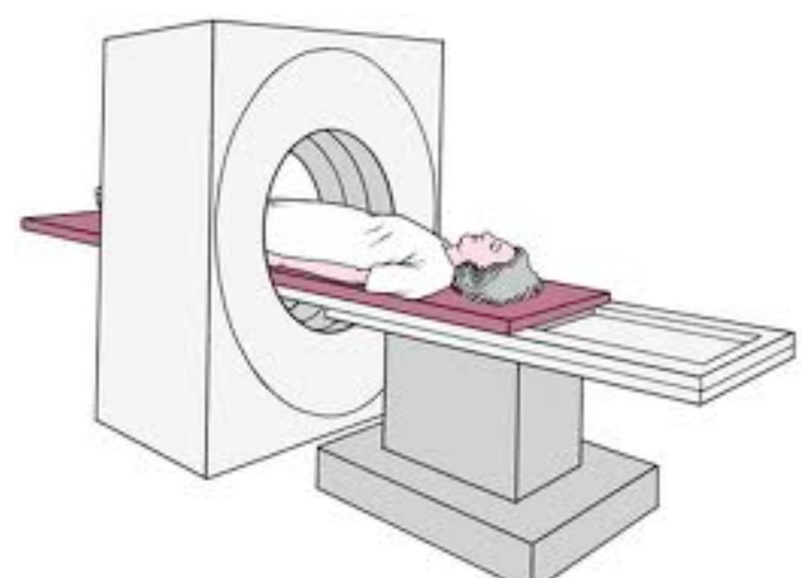
Compressed sensing





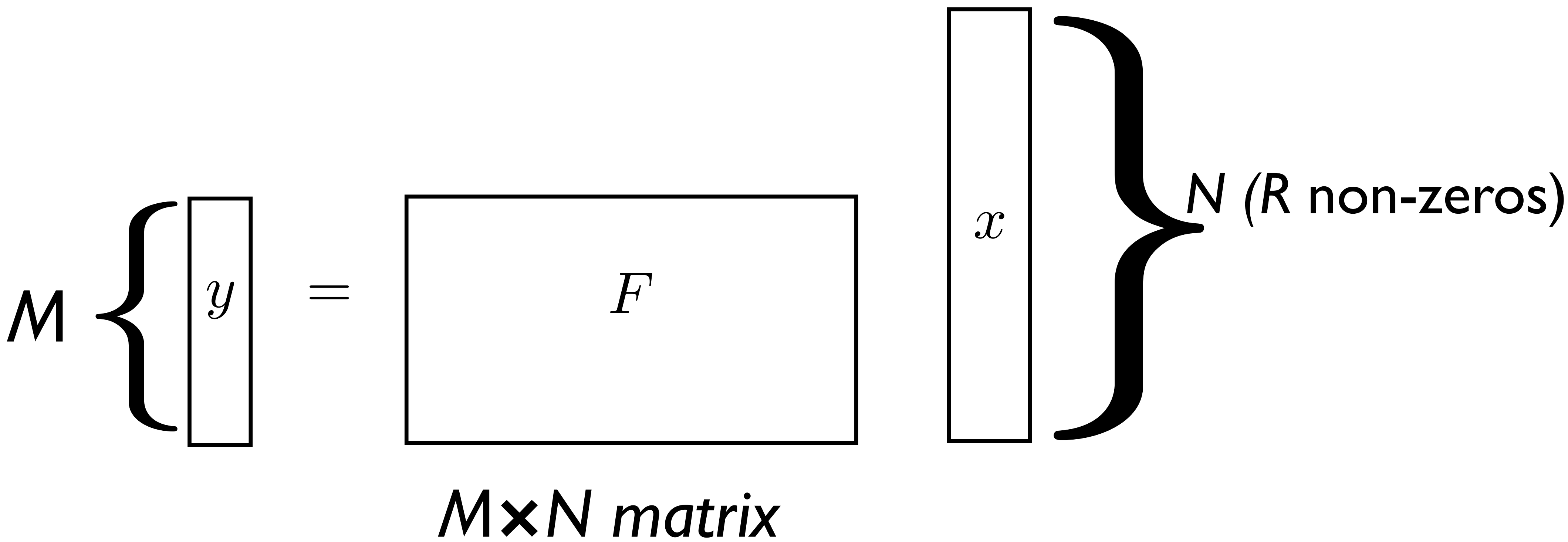
Discrete tomography





Compressed sensing

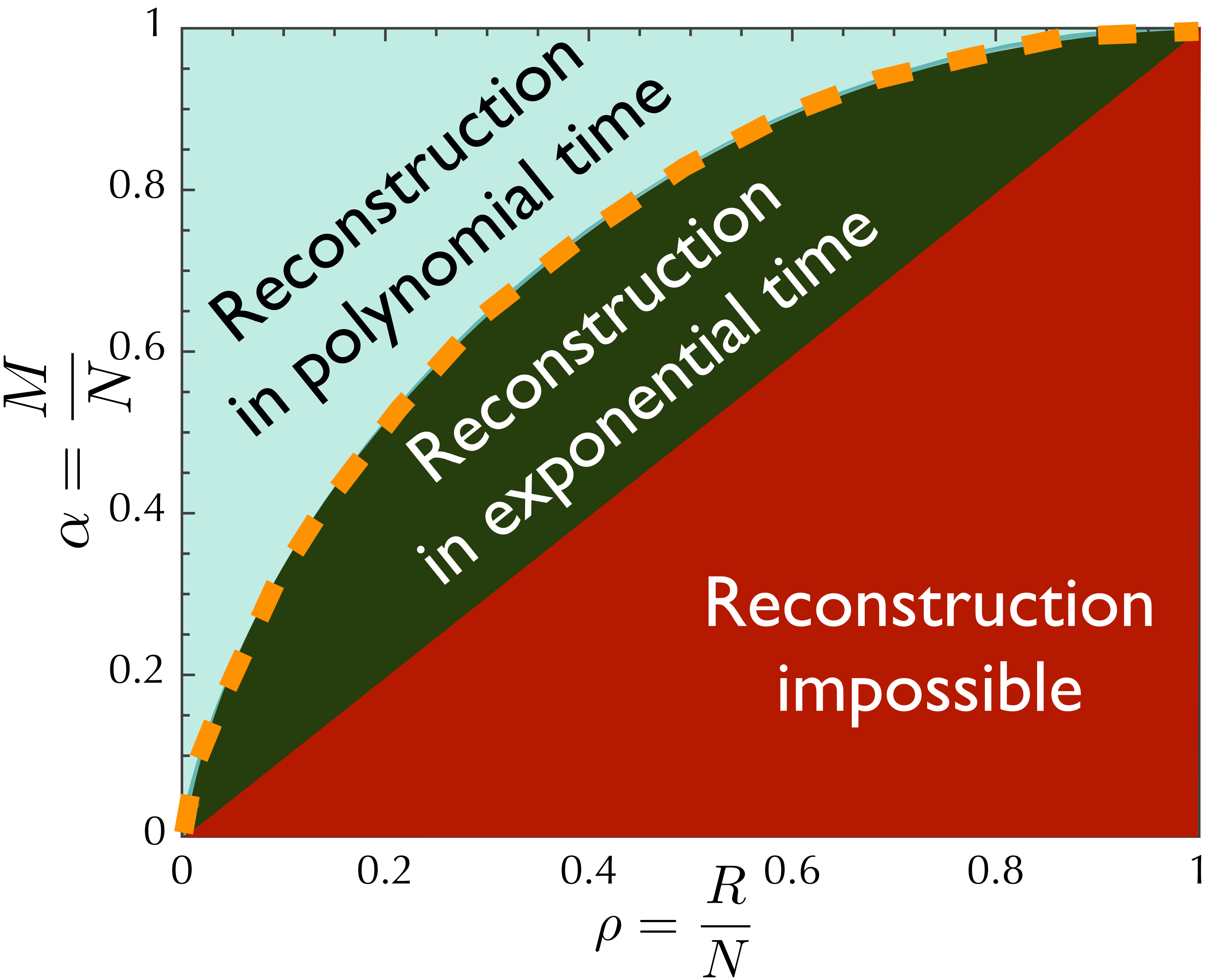
State of the art in CS



- Incoherent samplings (i.e. a random matrix F)
- Reconstruction by minimizing the L1 norm $||\vec{x}||_{L1} = \sum_i |x_i|$

Candès & Tao (2005)
Donoho and Tanner (2005)

State of the art in CS



For a signal with

- $(1-\rho)N$ zeros
- $R=\rho N$ non zeros

and a random iid matrix with $M = \alpha N$

Reconstruction limited by the Donoho-Tanner transition for the L_1 norm minimization

Analysis of the BP/TAP algorithm

(Also known as AMP in compressed sensing, *Montanari et al.*)

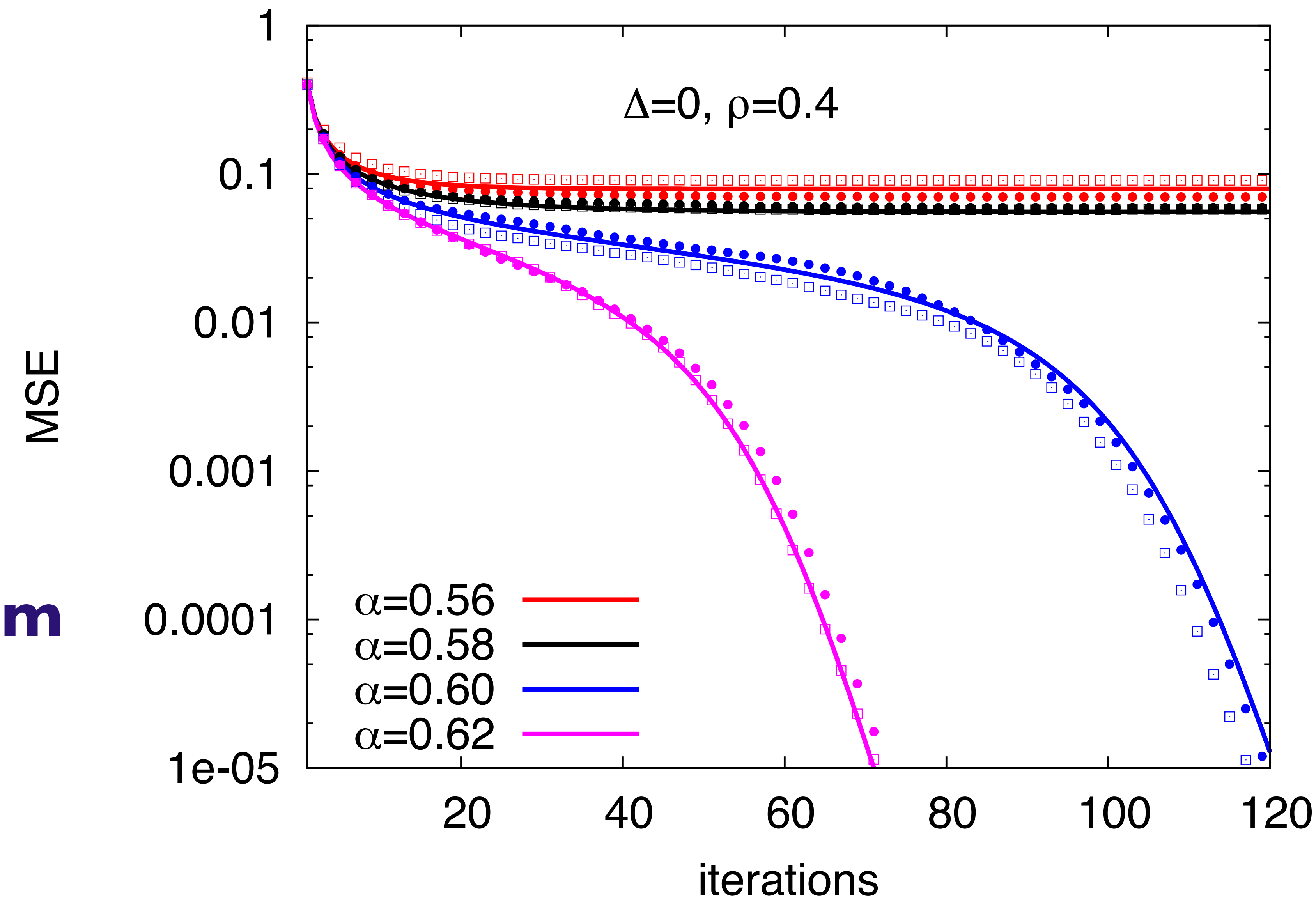
The performance of the algorithm for a given distribution of signals can be analyzed using a method knows as density evolution (coding theory) or replica method (physics)

Rigourous

Bayati and Montanari
Lelarge and Montanari

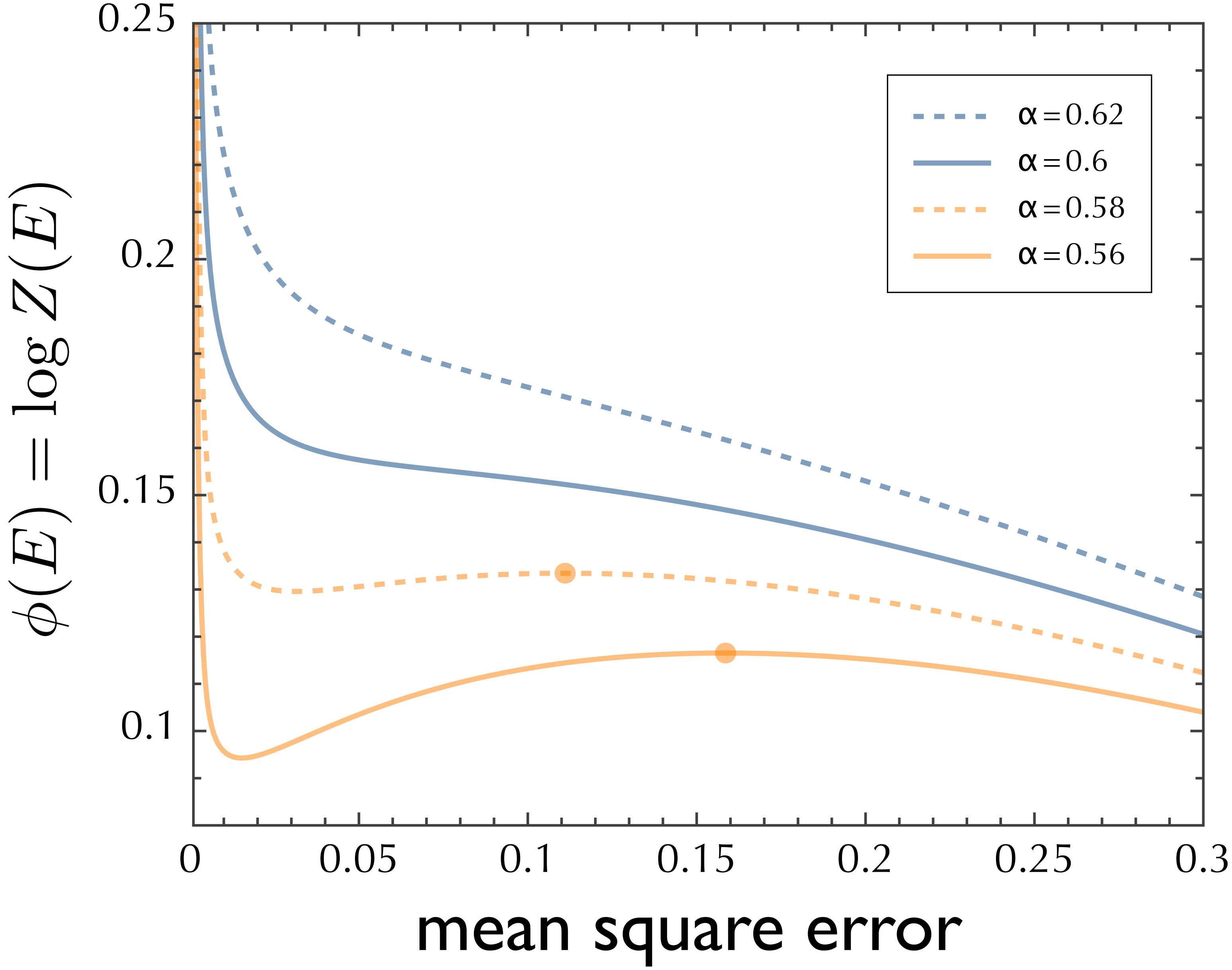
Comparison BP/Algorithm

for discrete and continuous values matrices



Steepest ascent of the free entropy

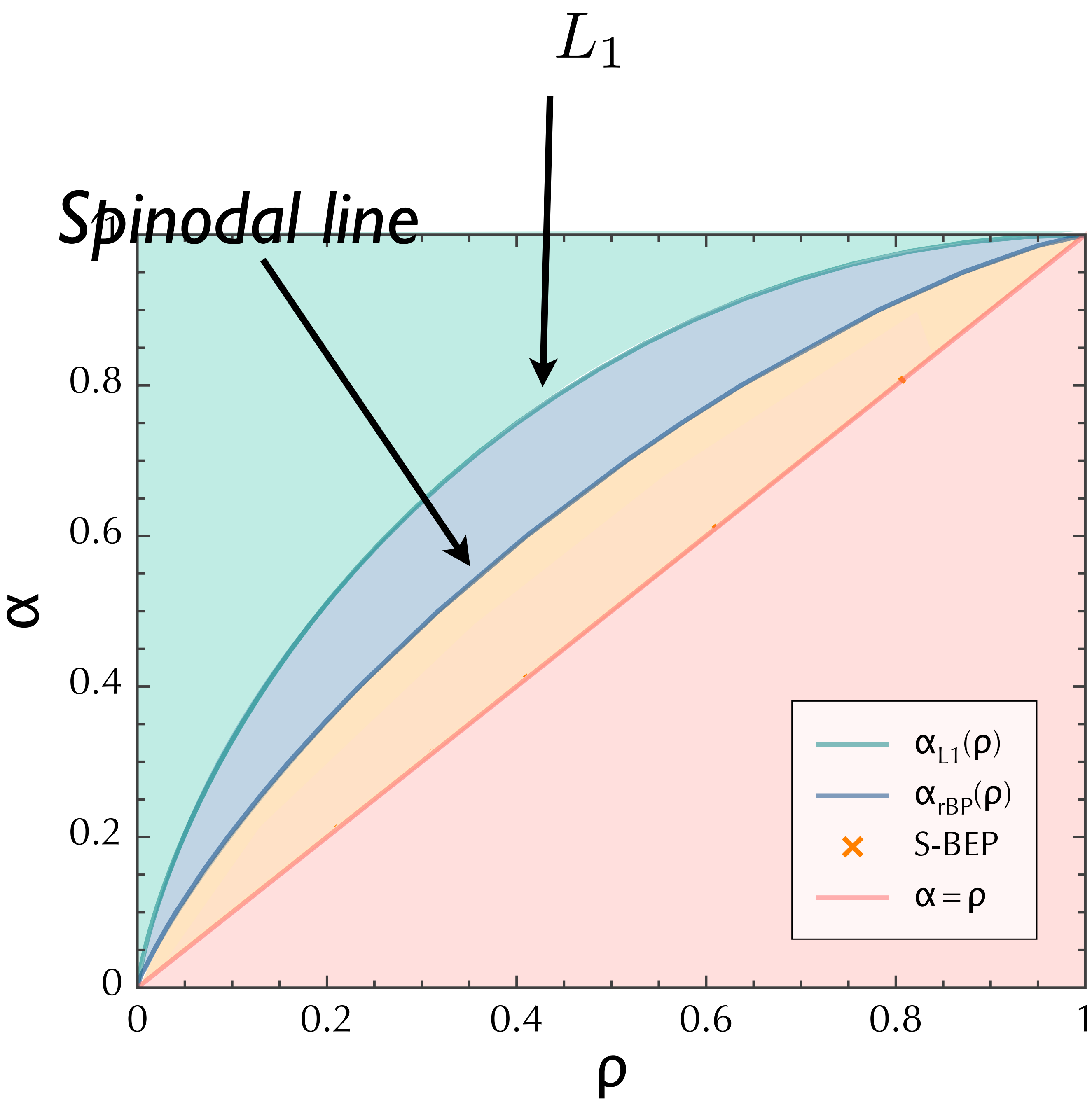
Example with ρ_0=0.4, and Φ_0 a Gaussian distribution with zero mean and unit variance

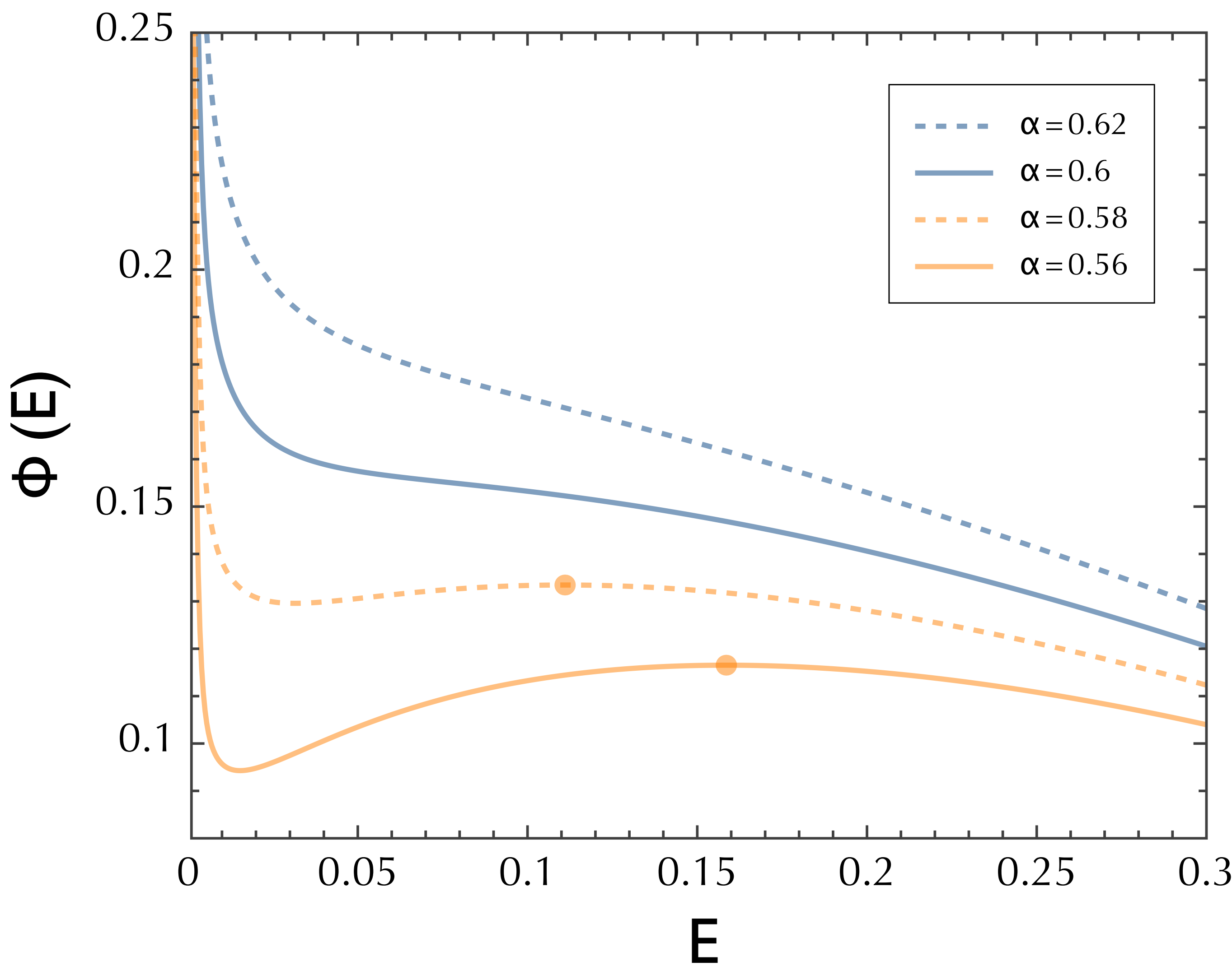


$$E = \frac{1}{N}\sum_i \left(\langle x_i \rangle - x_i^0\right)^2$$

- Maximum is at E=0 (as long as $\alpha>\rho0$): Equilibrium behavior dominated by the original signal
- For $\alpha<0.58$, a secondary maximum appears (meta-stable state): spinodal point
- A steepest ascent dynamics starting from large E would reach the signal for $\alpha>0.58$, but would stay block in the meta-stable state for $\alpha<0.58$, even if the true equilibrium is at E=0.
- Similarity with supercooled liquids

Computing the Phase Diagram

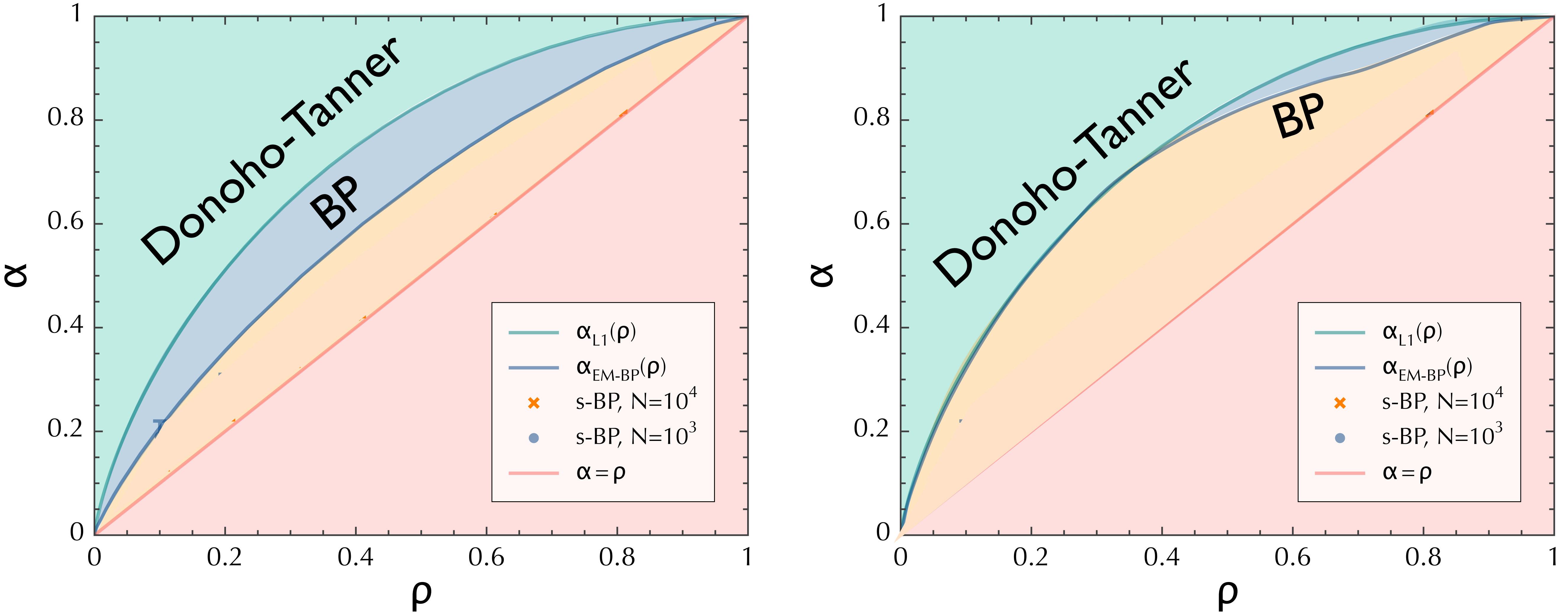




A steepest ascent of the free entropy allows a perfect reconstruction until the spinodal line. This is more efficient than L_1-minimization

Trying different type of signals

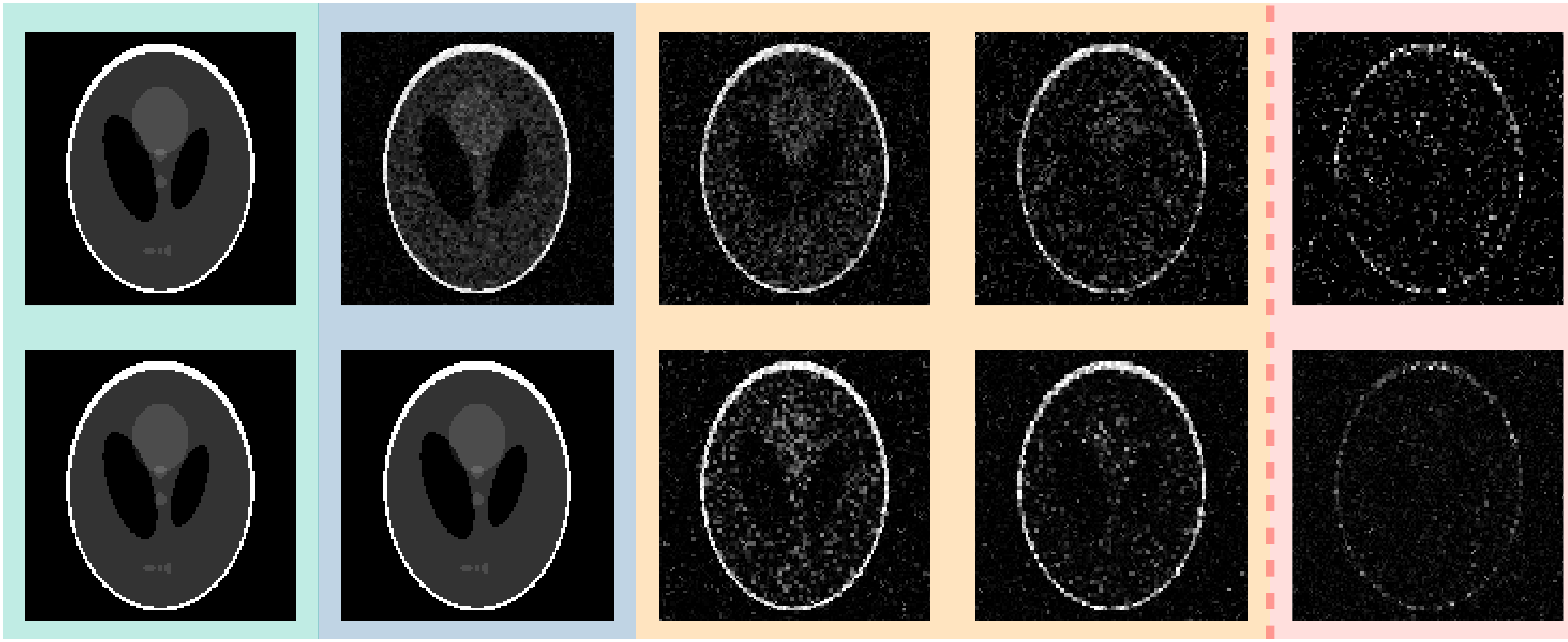
The limit depends on the type of signal
(while the Donoho-Tanner is universal)



Gauss-Bernoulli signal

Binary signals

A more complex signal



Shepp-Logan phantom, in the Haar-wavelet representation

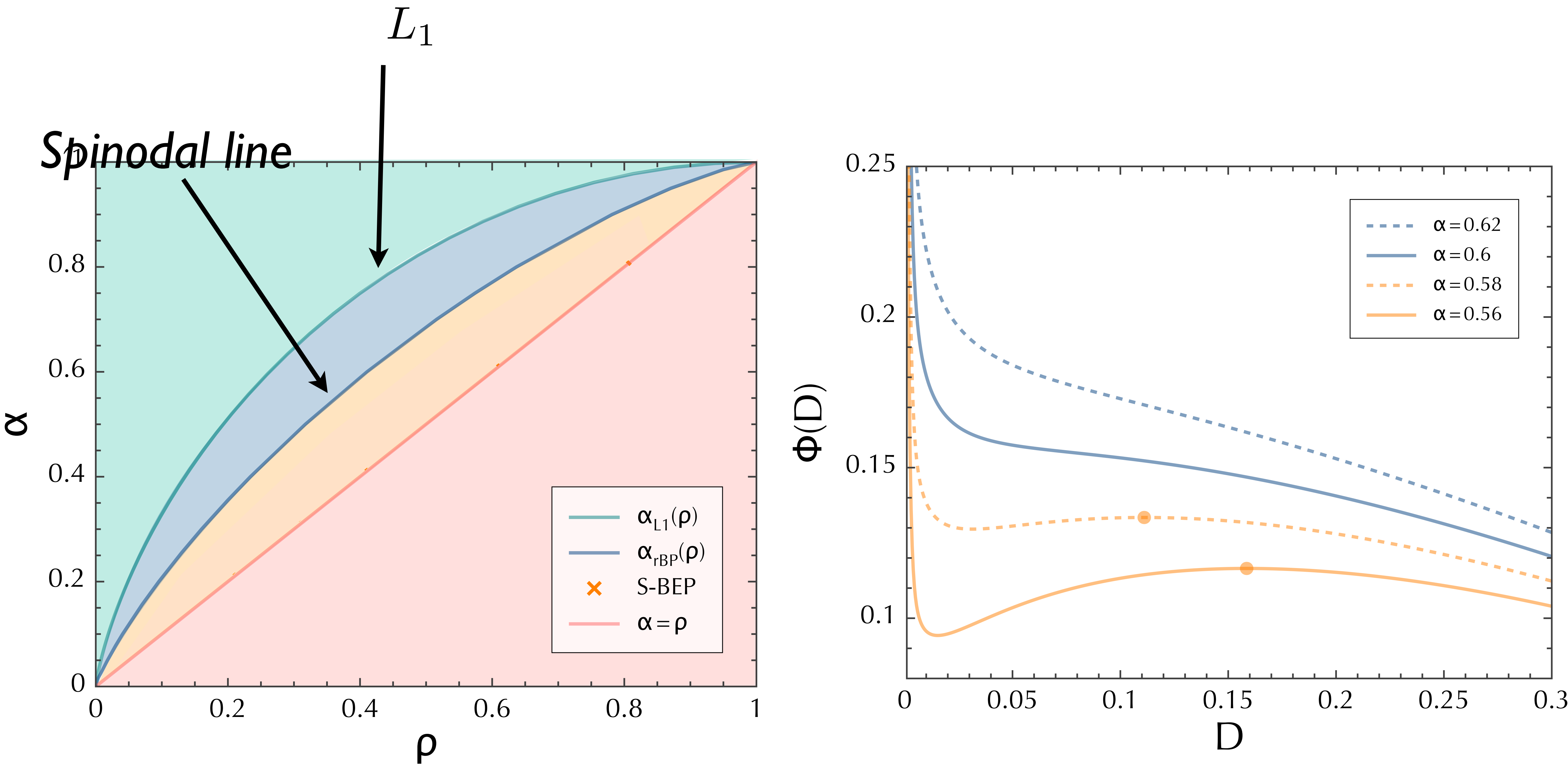
Can we do a better job?

Our work

A statistical physics approach to compressed sensing

- A probabilistic approach to reconstruction
- The Belief Propagation algorithm
- **Seeded measurements matrices**

This is good, but not good enough



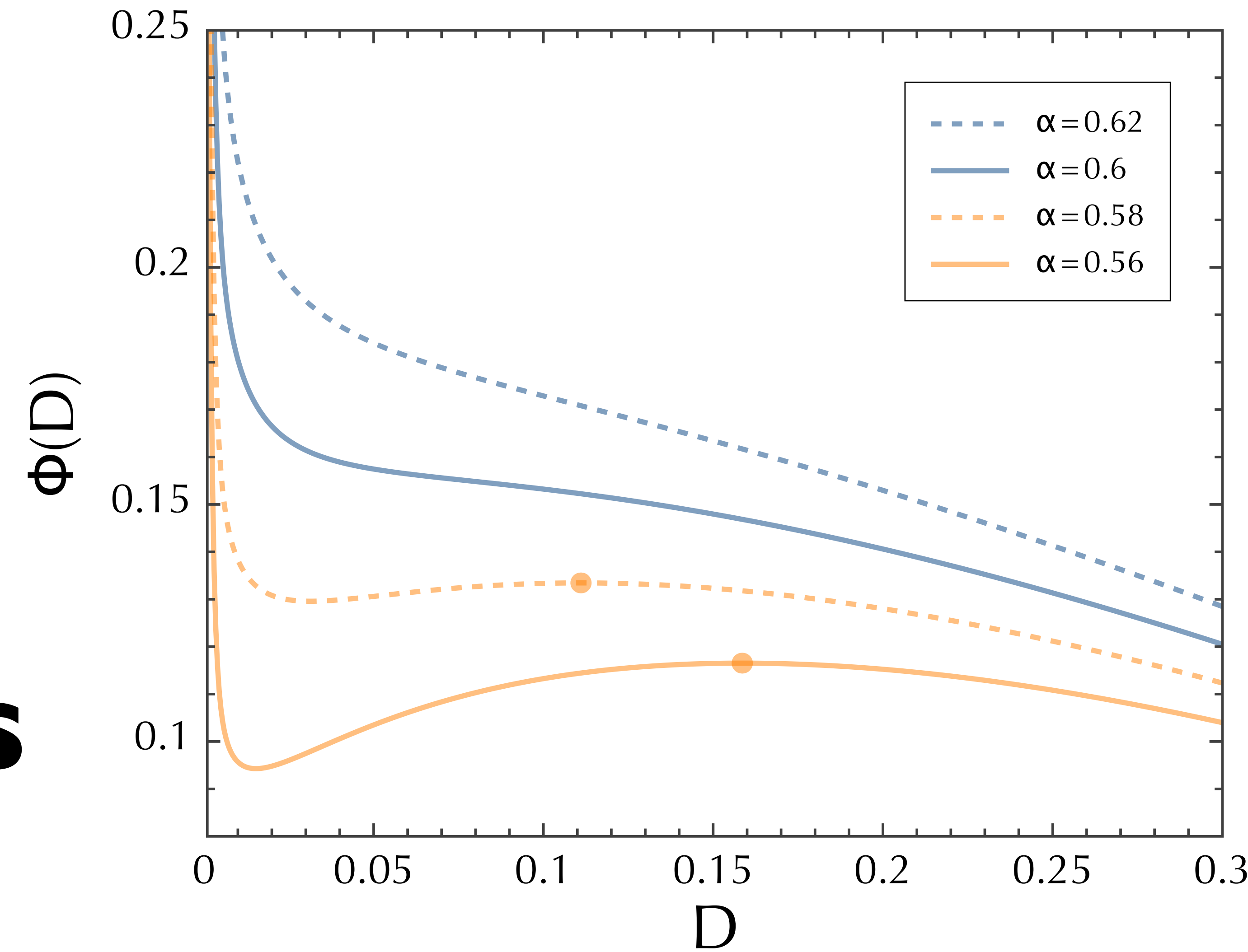
The dynamics is stuck in a metastable state, just as a liquid cooled too fast remains in a supercooled liquid state instead of crystalizing

This is good, but not good enough

How to pass the spinodal point?

By nucleation!

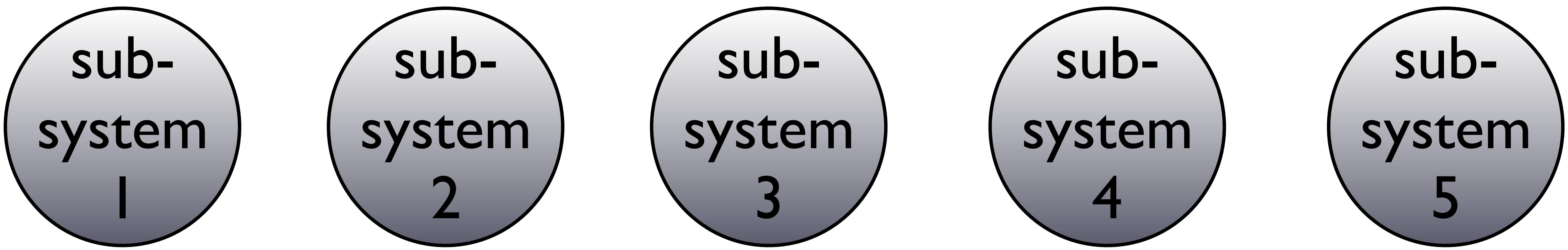
Special design of "seeded" matrices



The dynamics is stuck in a metastable state, just as a liquid cooled too fast remains in a supercooled liquid state instead of crystalizing

Mixed "mean-field" and one-dimensional system:

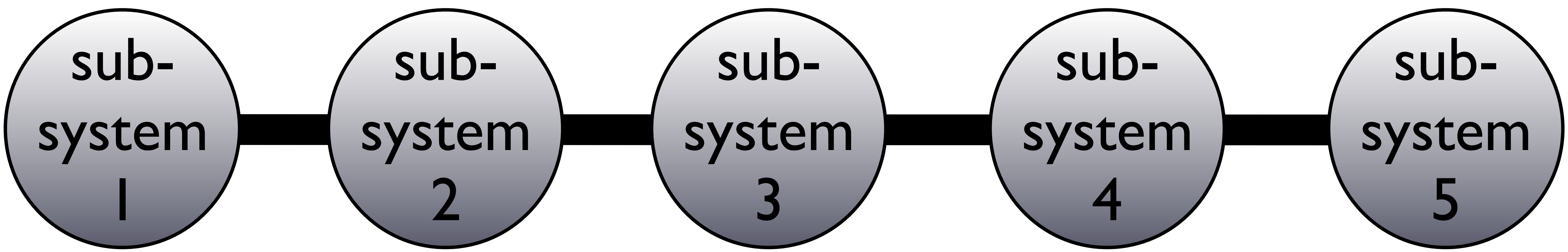
1) Create many "mean-field" sub-systems



A construction inspired by the
"spatially coupled matrices"
developed in coding theory
cf: Urbanke et al.

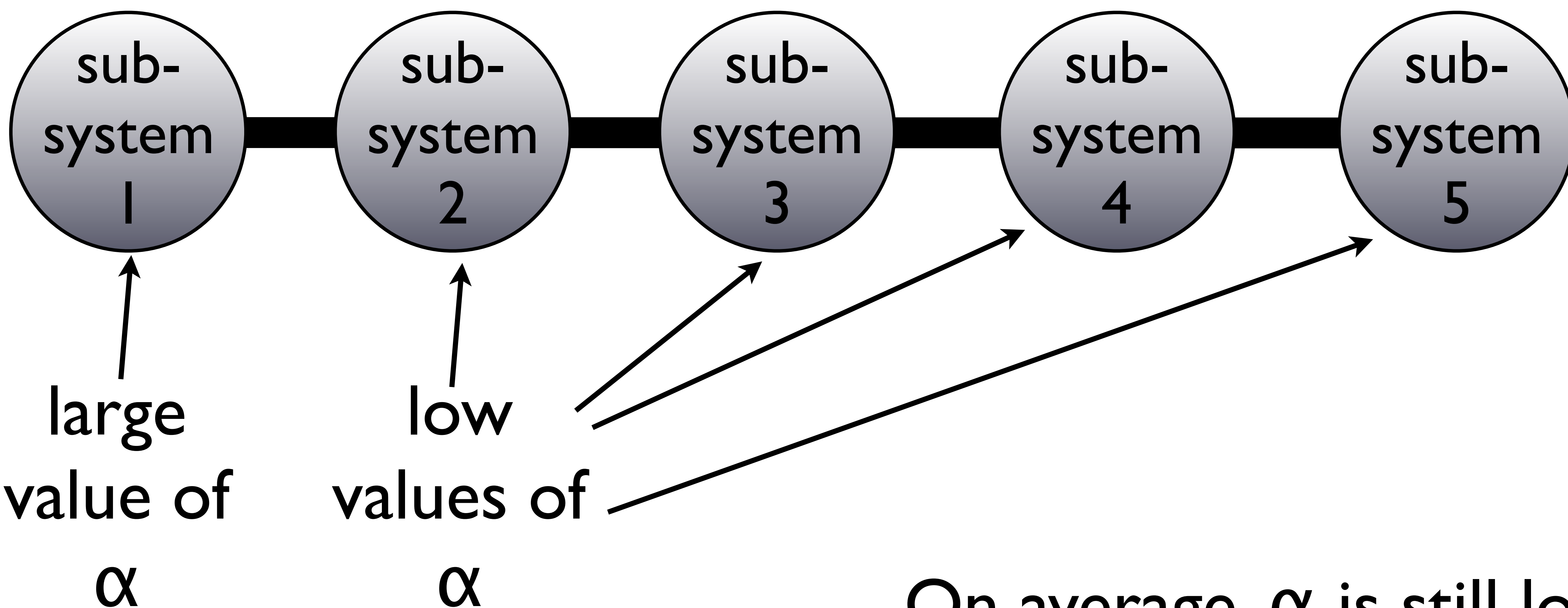
Mixed “mean-field” and one-dimensional system:

2) Add a first neighbor coupling



Mixed "mean-field" and one-dimensional system:

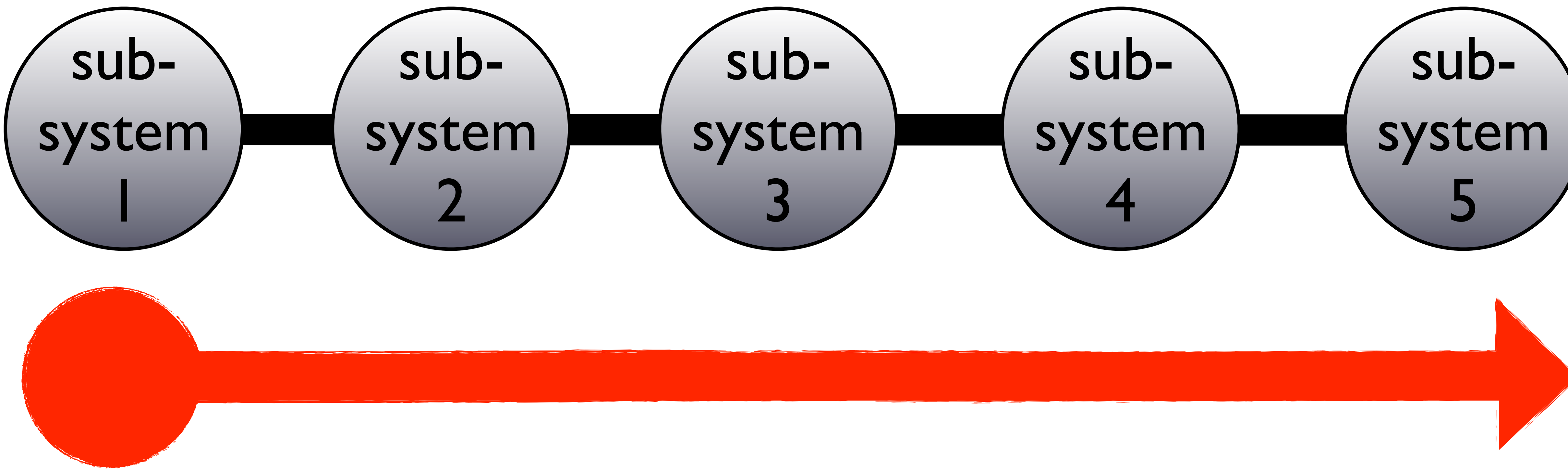
3) Choose parameters such that the first system is in the region of the phase diagram where there is no metastability

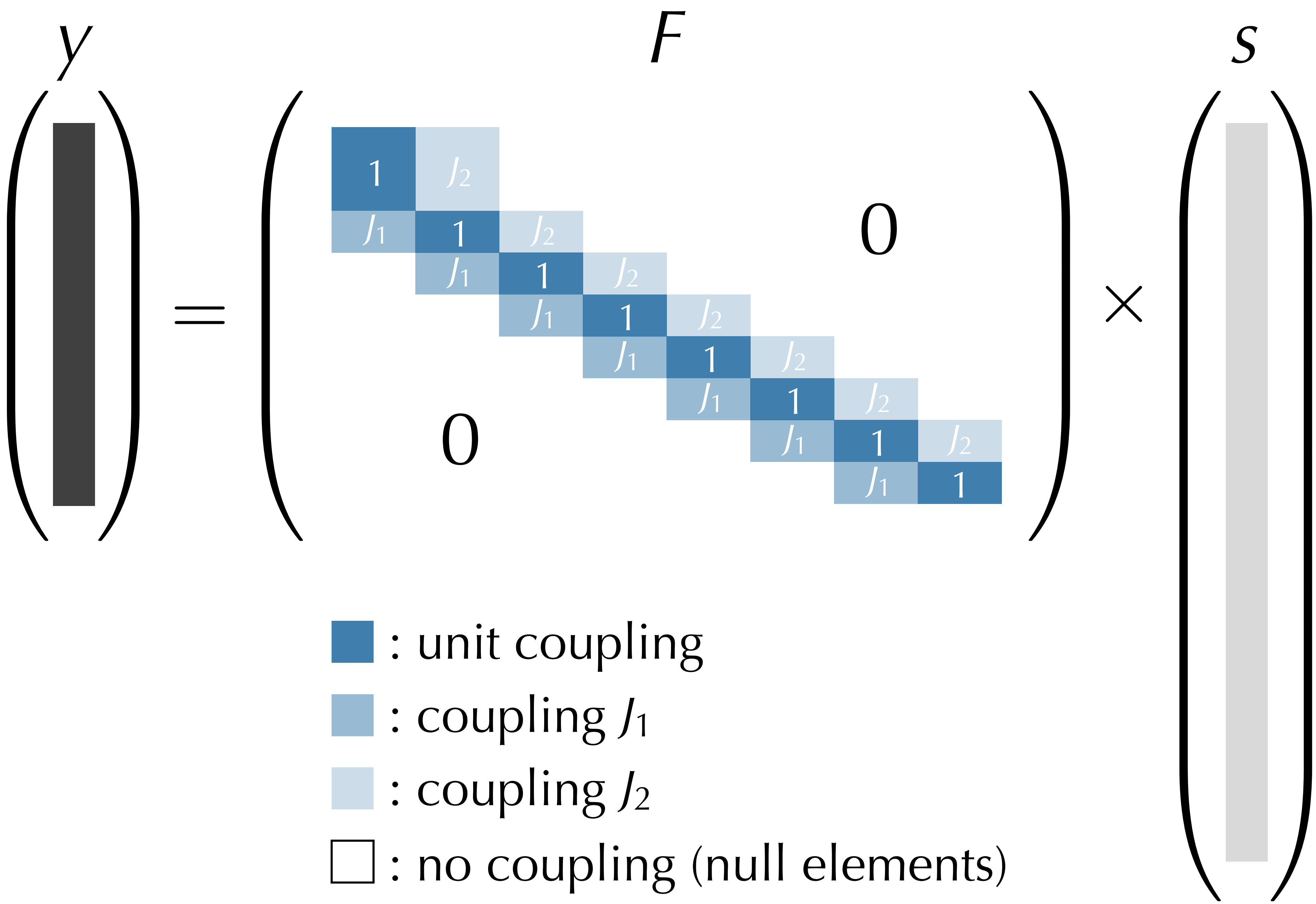


On average, α is still low !

Mixed "mean-field" and one-dimensional system:

4) The solution will appear in the first sub-system (with large α), and then propagate in the system





$$L = 8$$

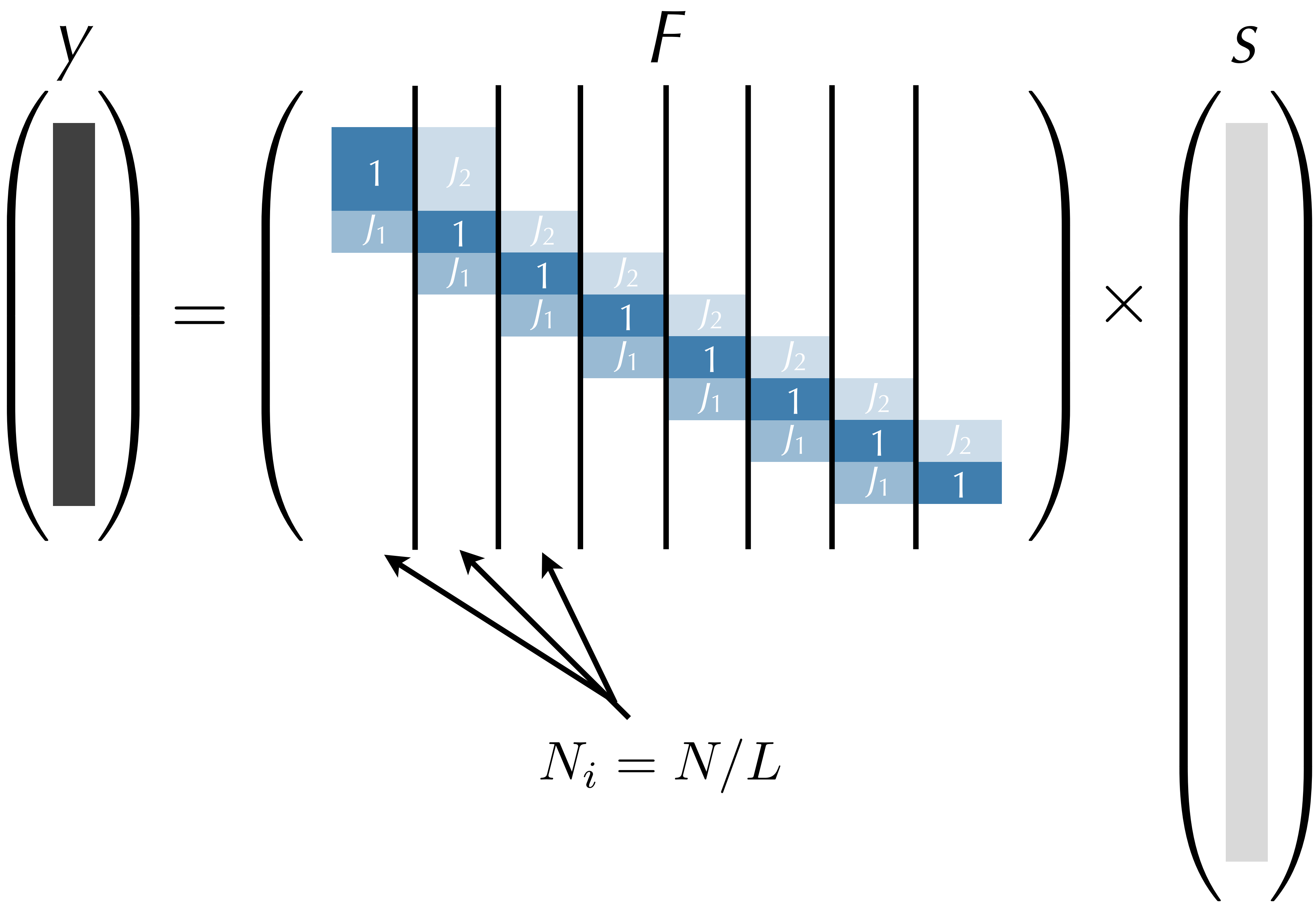
$$N_i = N/L$$

$$M_i = \alpha_i N/L$$

$$\alpha_1 > \alpha_{BP}$$

$$\alpha_j = \alpha' < \alpha_{BP} \qquad j \geq 2$$

$$\alpha = \frac{1}{L}\left(\alpha_1 + (L-1)\alpha'\right)$$



$$L = 8$$

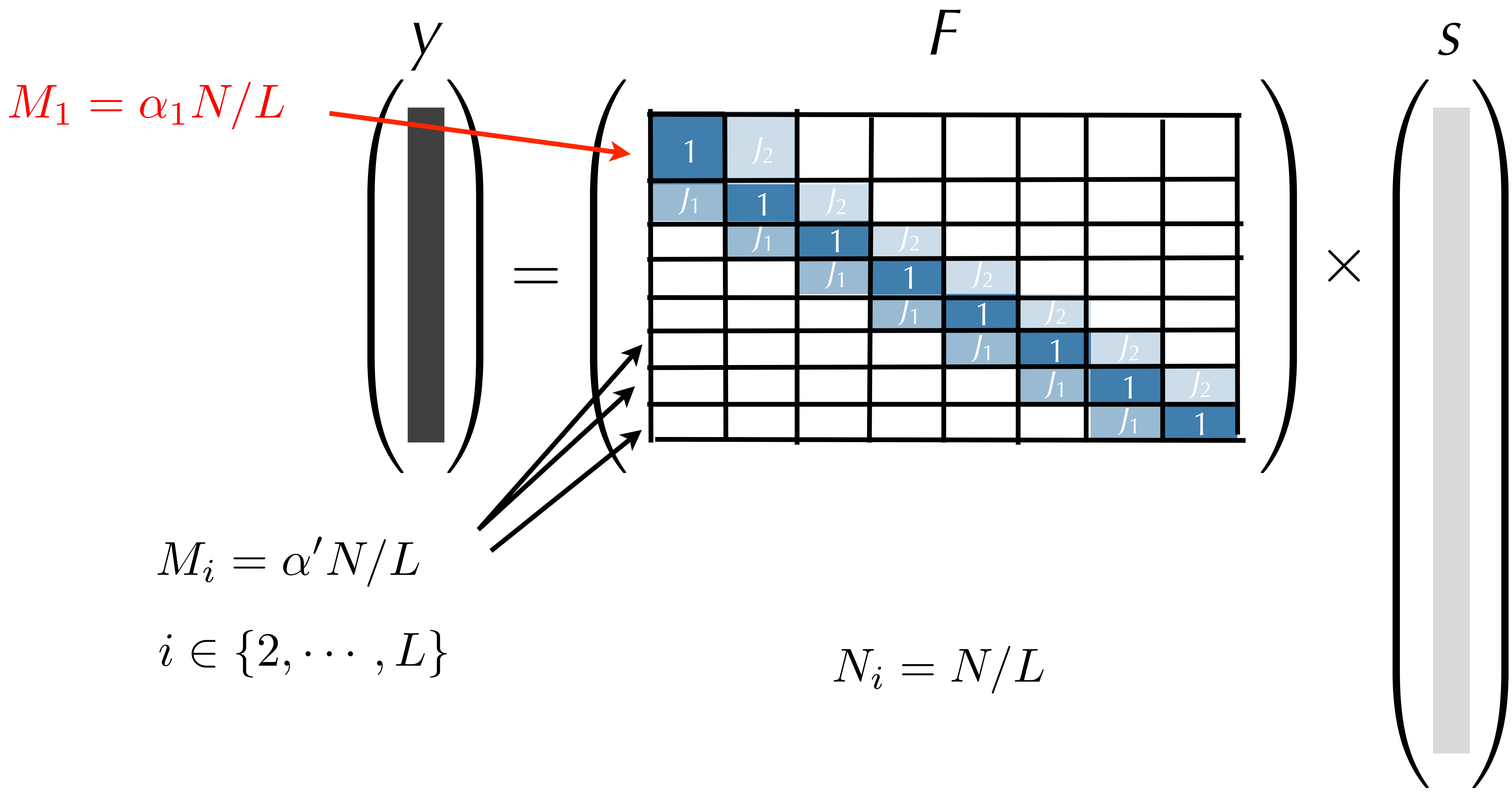
$$N_i = N/L$$

$$M_i = \alpha_i N/L$$

$$\alpha_1 > \alpha_{BP}$$

$$\alpha_j = \alpha' < \alpha_{BP} \qquad j \geq 2$$

$$\alpha = \frac{1}{L}\left(\alpha_1 + (L-1)\alpha'\right)$$



$L = 8$

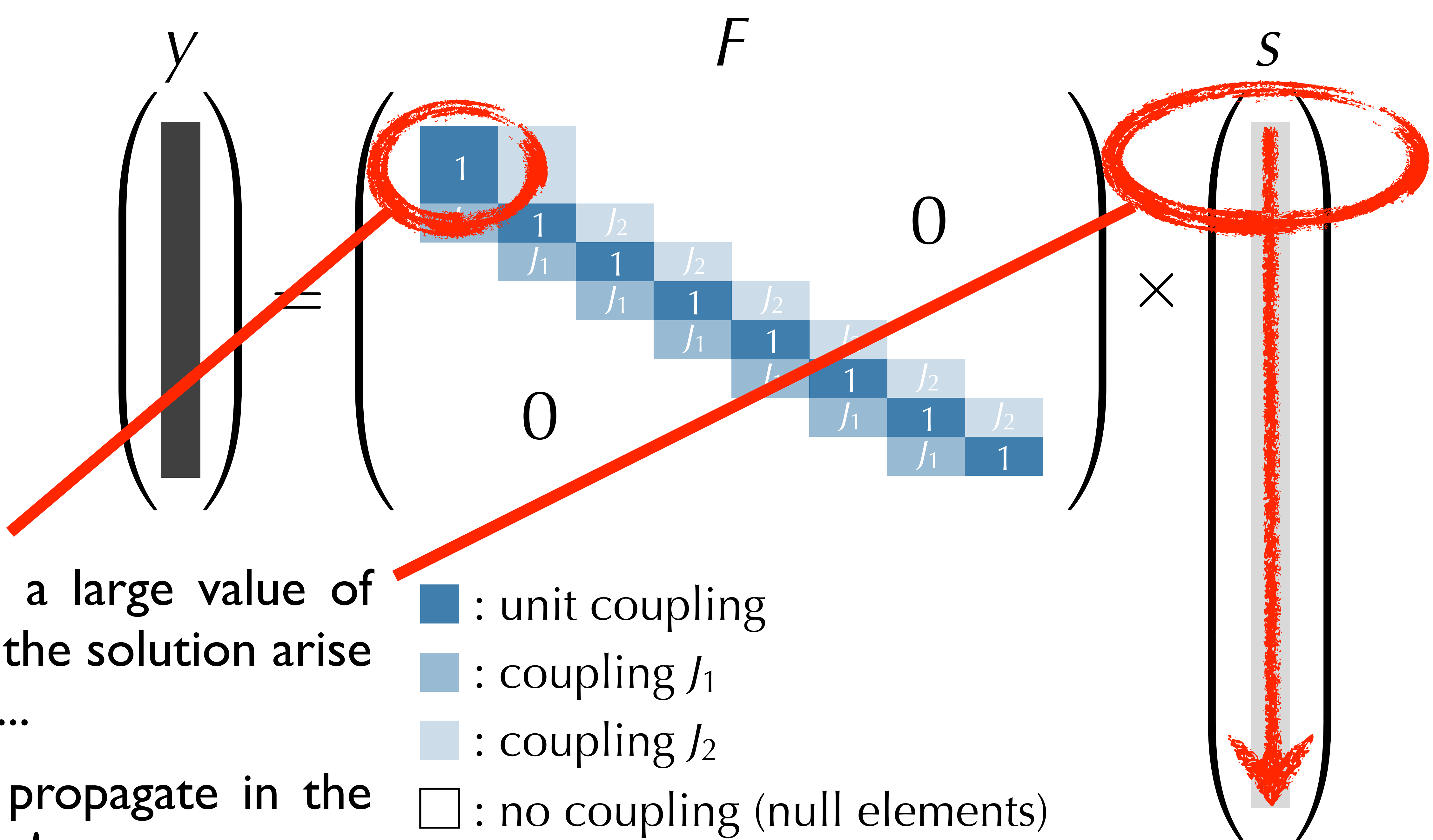
$N_i = N/L$

$M_i = \alpha_i N/L$

$\alpha_1 > \alpha_{BP}$

$\alpha_j = \alpha' < \alpha_{BP} \qquad j \geq 2$

$$\alpha = \frac{1}{L}\left(\alpha_1 + (L-1)\alpha'\right)$$



Block 1 has a large value of M such that the solution arise in this block...

... and then propagate in the whole system!

$$L = 8$$

$$N_i = N/L$$

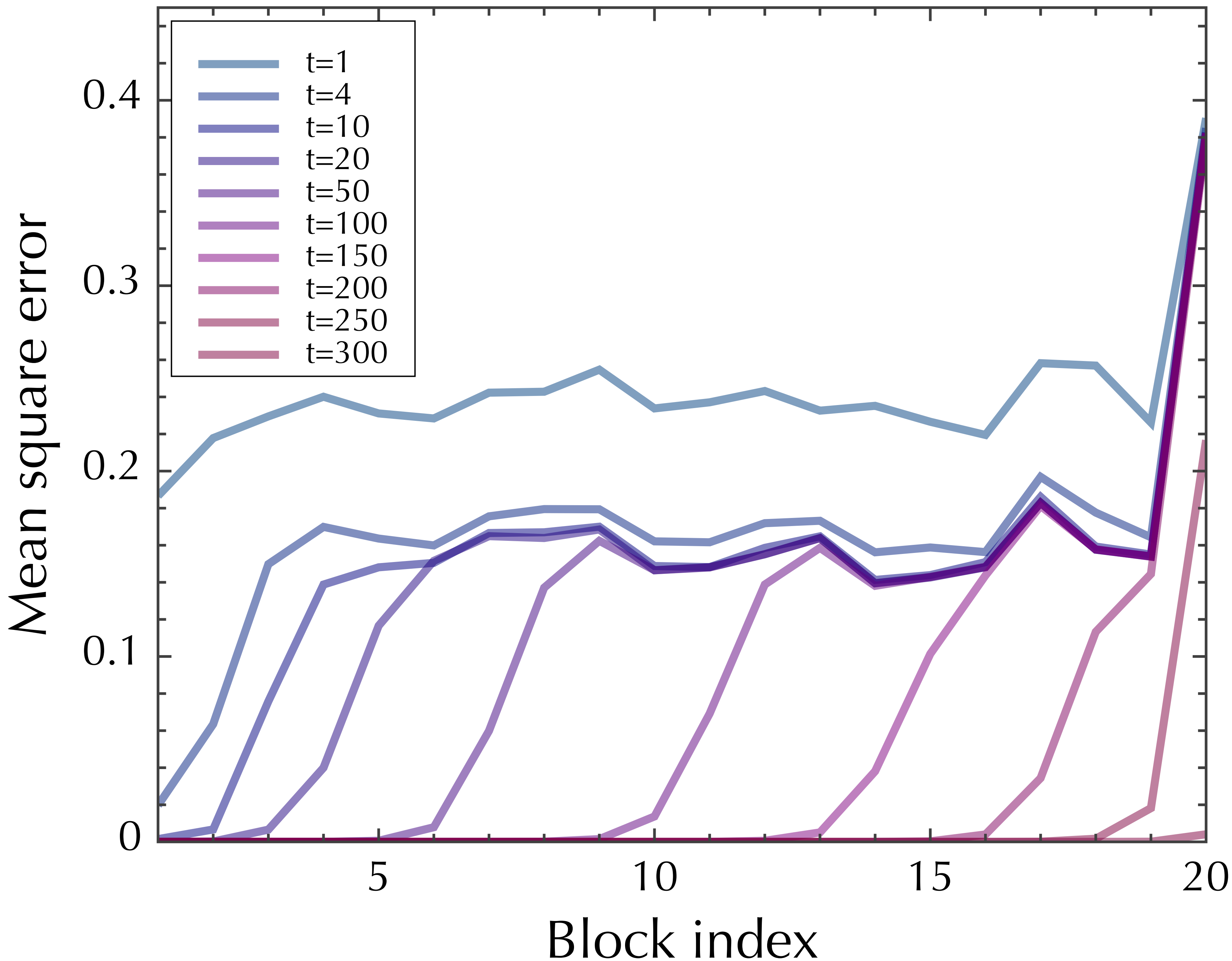
$$M_i = \alpha_i N/L$$

$$\alpha_1 > \alpha_{BP}$$

$$\alpha_j = \alpha' < \alpha_{BP} \qquad j \geq 2$$

$$\alpha = \frac{1}{L}\left(\alpha_1 + (L-1)\alpha'\right)$$

Example with ρ_0=0.4, and Φ_0 a Gaussian distribution with 0 mean and unit variance



$L = 20$ $N = 50000$ $\rho = .4$ $J_1 = 20$ $\alpha_1 = 1$

$J_2 = .2$ $\alpha = .5$

Analytical results for seeding matrices

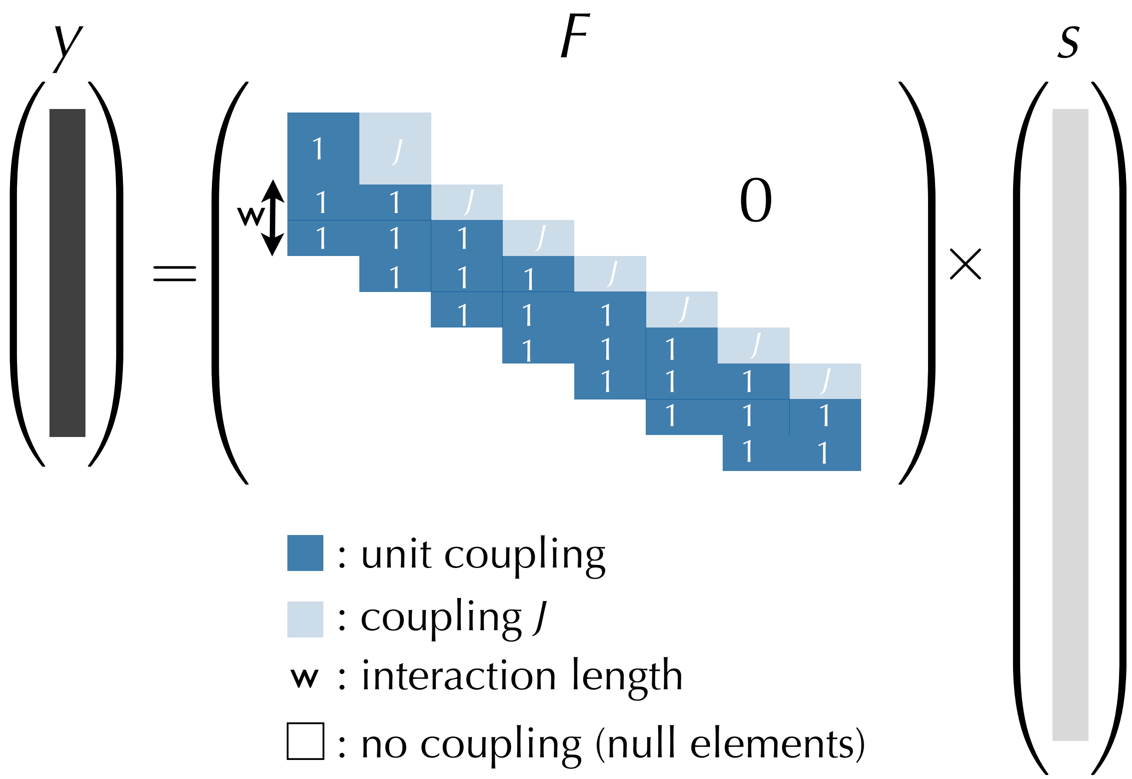
•One can repeat the replica/density evolution analysis for the seeded matrices, and the performance of the algorithm can be studied analytically, leading to $\alpha > \rho$ in the large N limit:

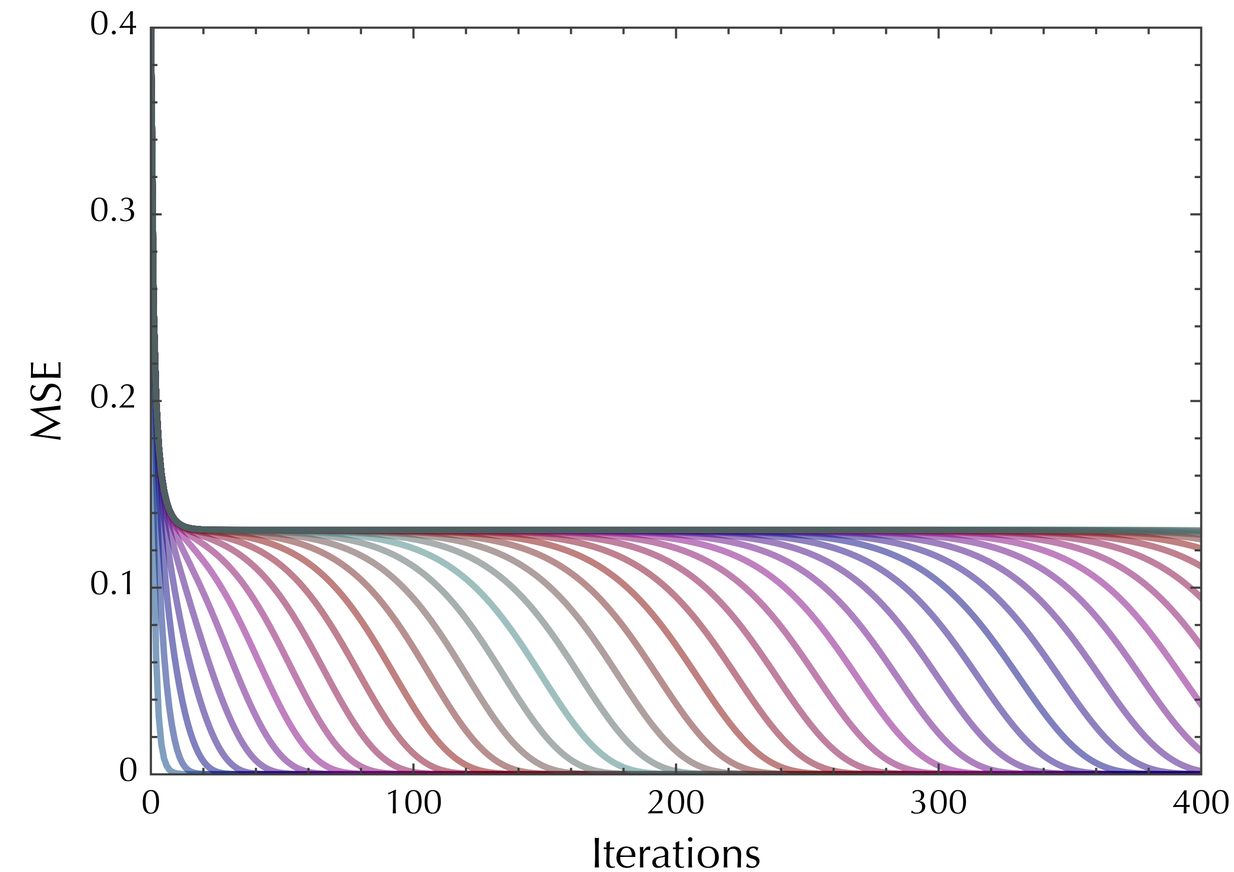
Asymptotically optimal measurements

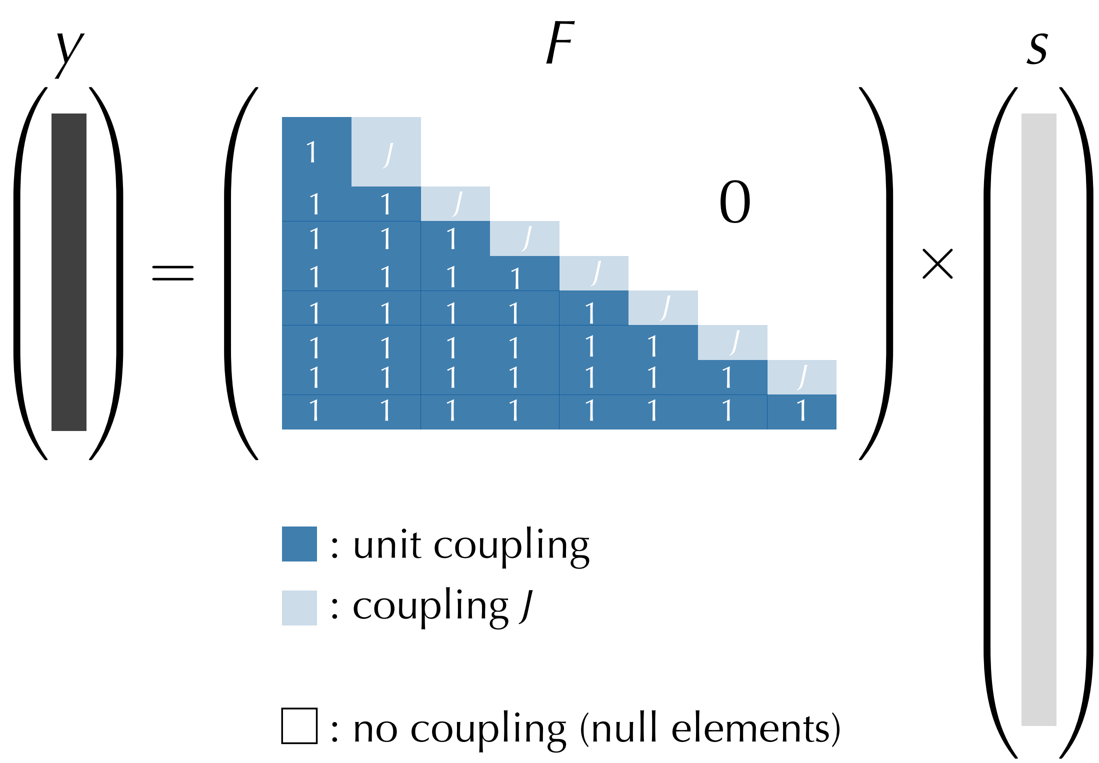
•A special case (zero of vanishing noise, and with prior matching the signal) have been recently confirmed by a rigorous analysis by Donoho, Montanari and Javanmard `(arXiv:1112.0708)`

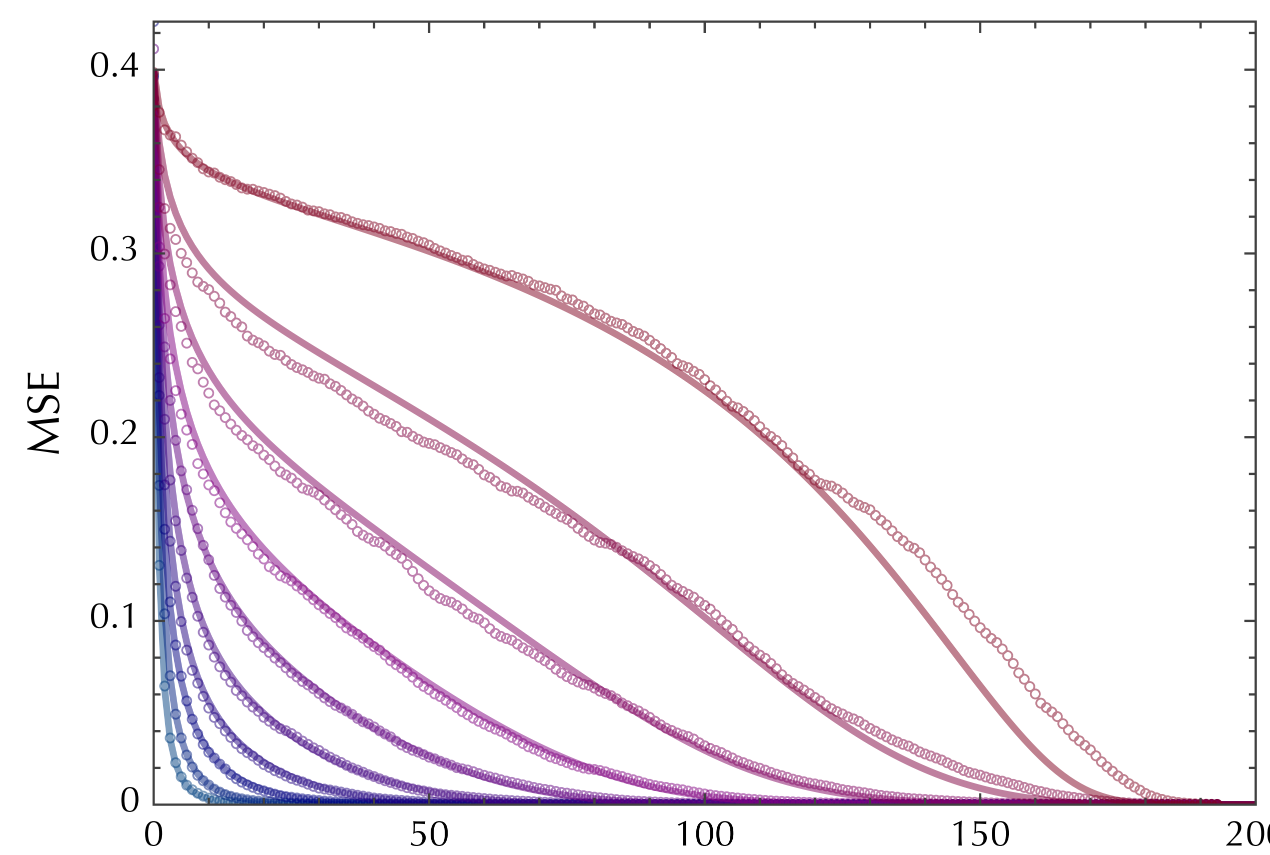
•But note that the analysis of the density evolution shows that our construction works *even* when the prior is not the correct one, and also with large noise (although with noise the performances depend on the prior)

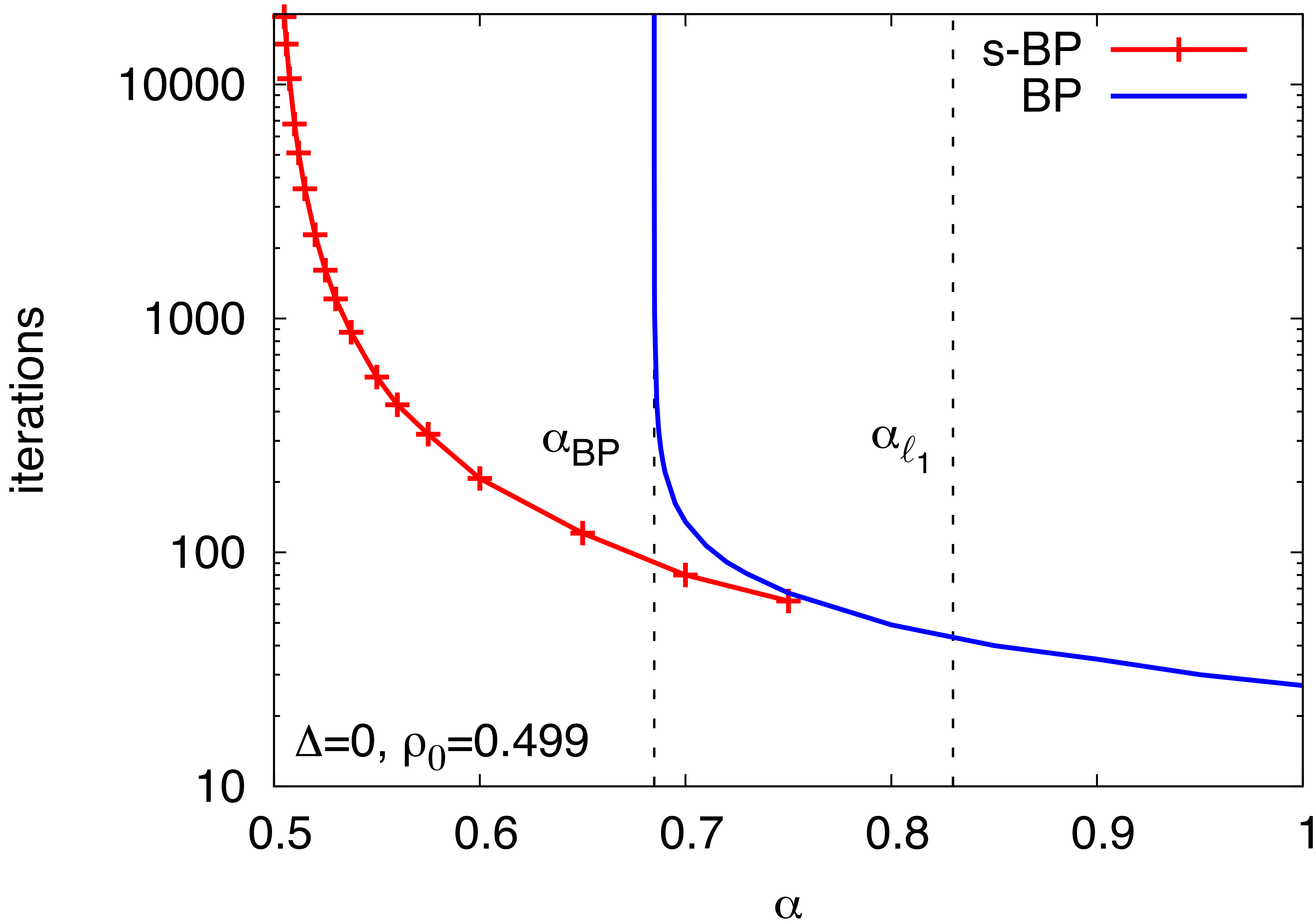
Many way to design seeding matrices





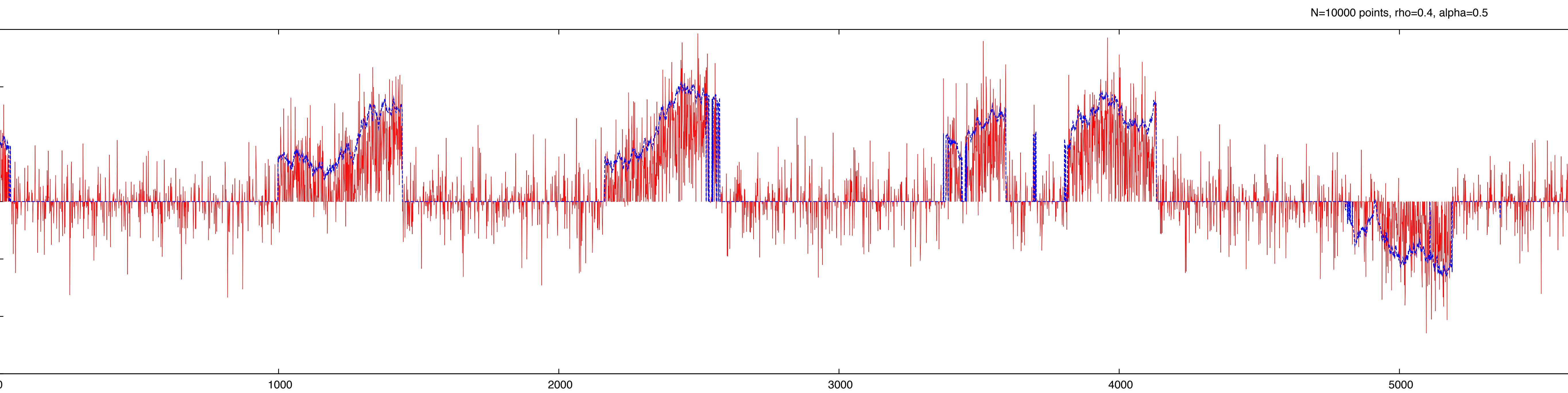






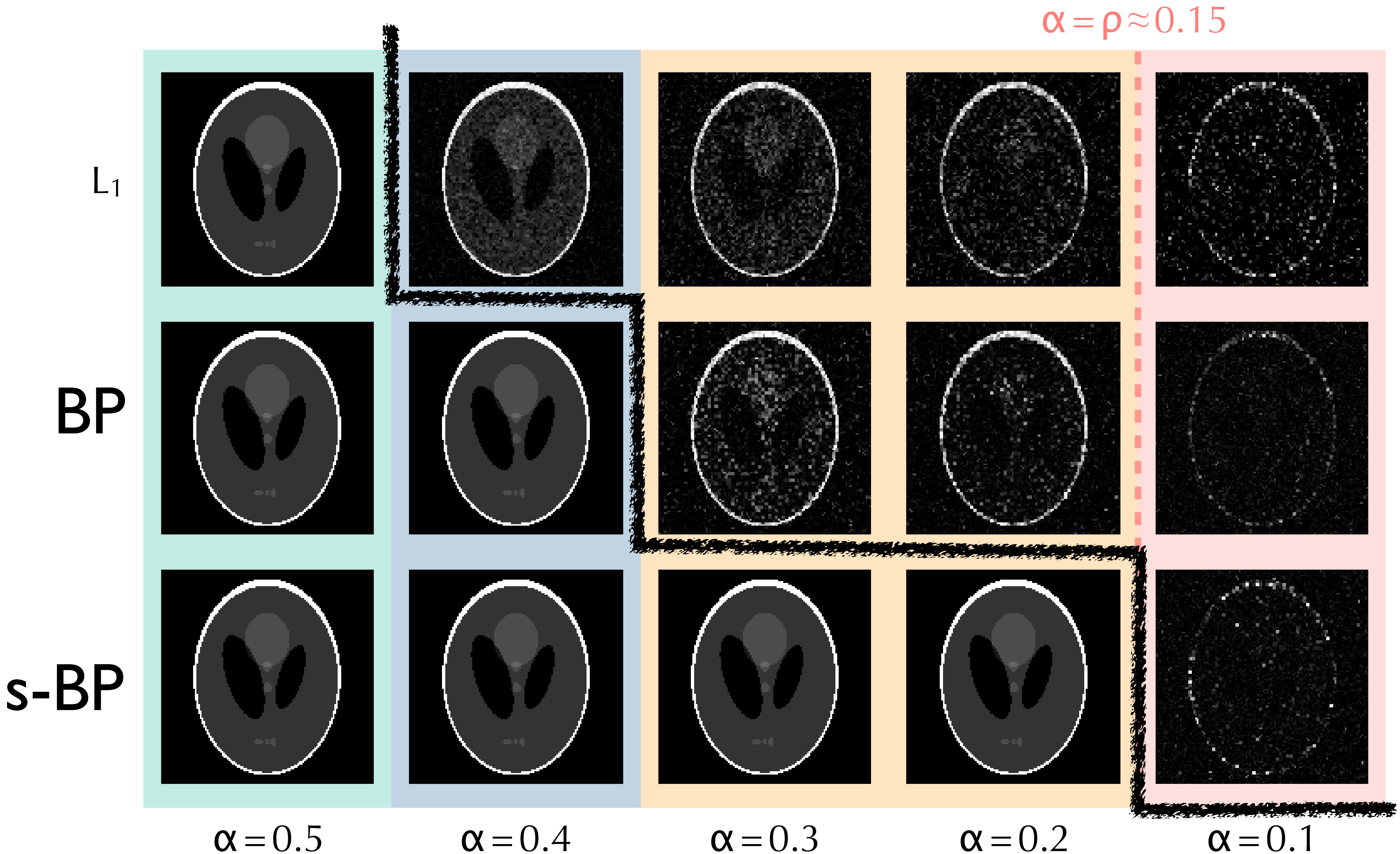
s-BP
BP
10000
1000
100
10
iterations
α_{BP}
α_{ℓ_1}
Δ=0, ρ_0=0.499
0.5
0.6
0.7
0.8
0.9
1
α

A signal with α=0.5 and ρ=0.4



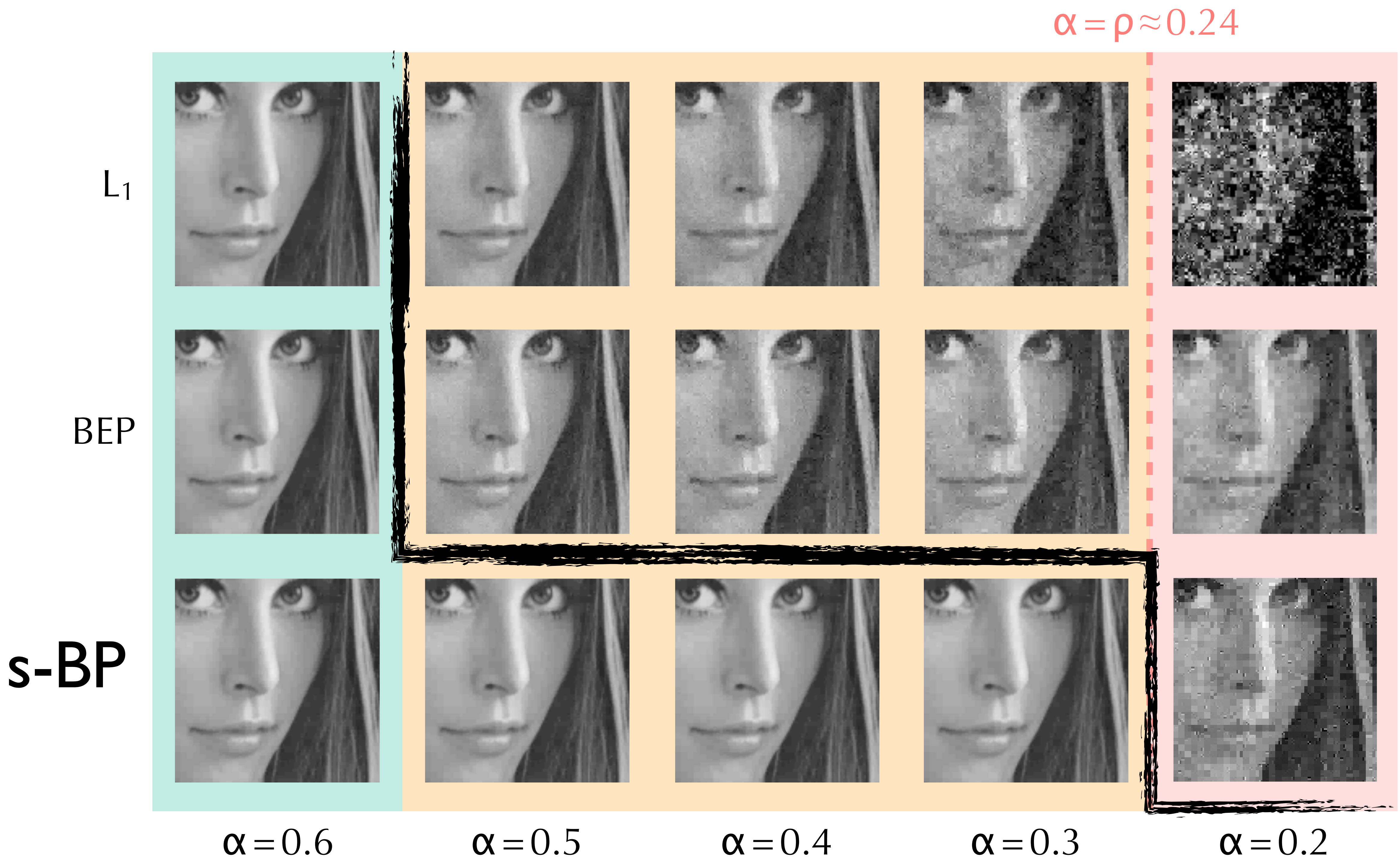
Blue is the true signal reconstructed by s-BP
Red is the signal found by L_1

A more interesting example



Shepp-Logan phantom, in the Haar-wavelet representation

A EVEN more interesting example



The Lena picture in the Haar-wavelet representation

Conclusions...

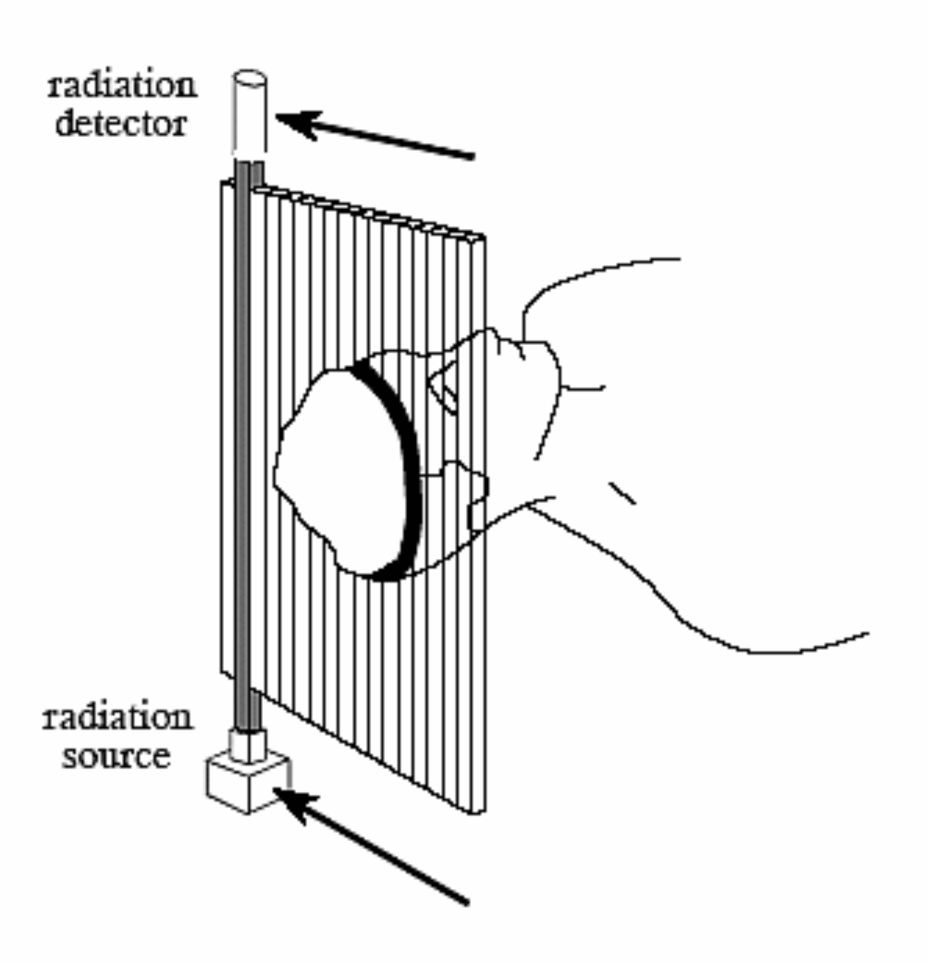
- Probabilistic approach to reconstruction in compressed sensing...
- ... with a Belief Propagation algorithm.
- Seeded measurements matrices allows to perform optimally

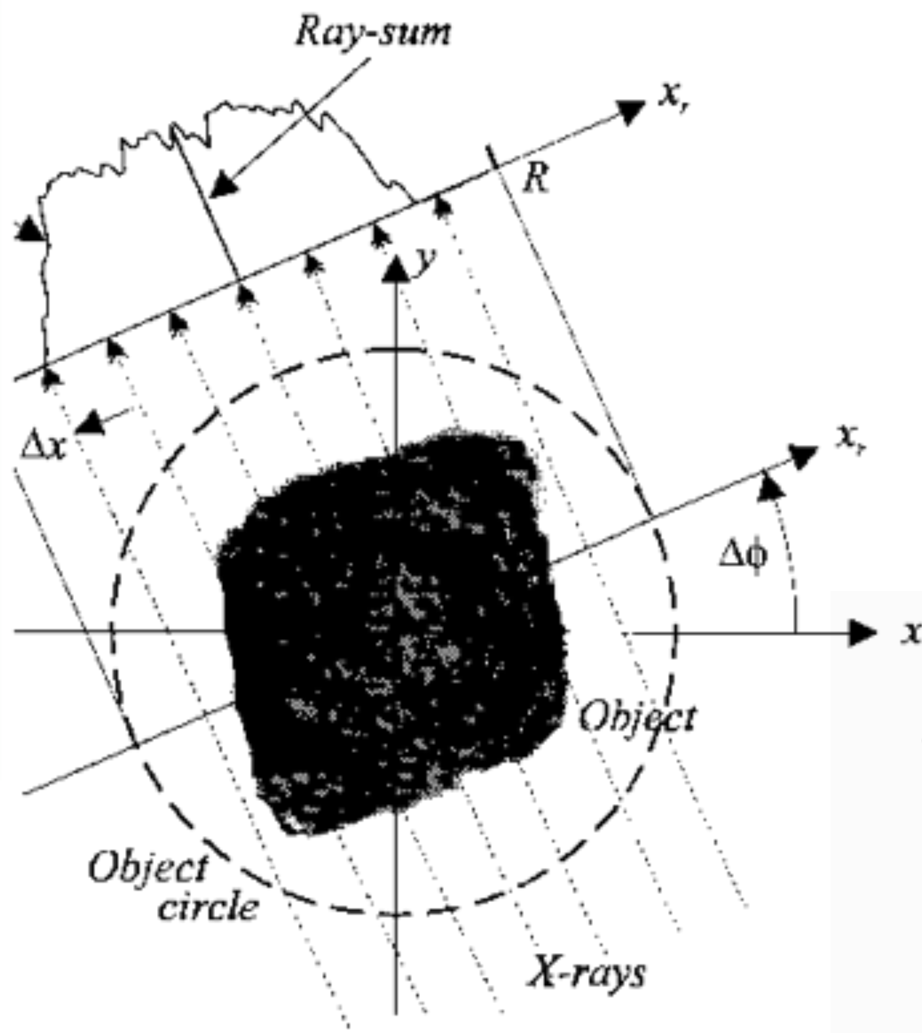
... and perspectives:

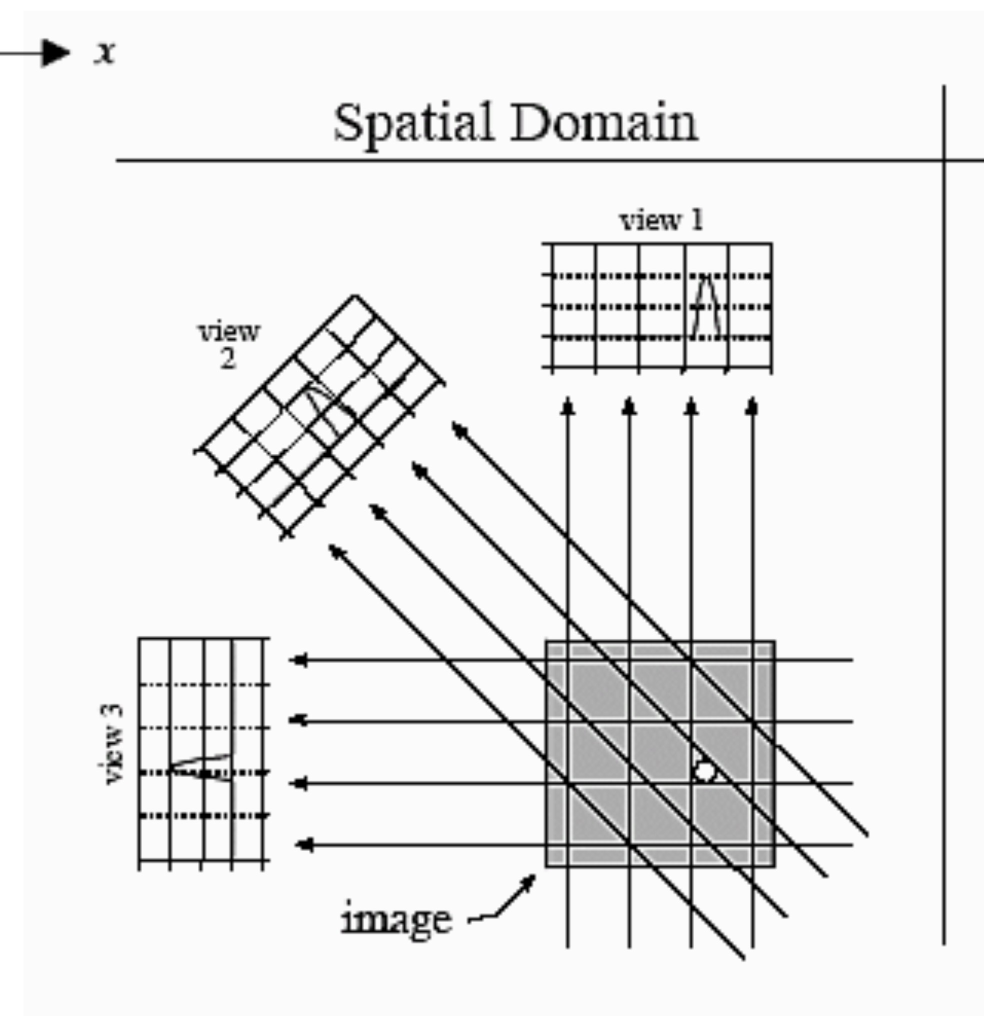
- More information in the prior ? Calibration noise, additive noise, approximated-sparsity, structure sparsity, etc... ?
- Dictionary learning? Sparse PCA? Fast data compression? Quantum tomography? Group testing? etc...

Discrete Tomography

X-ray computed tomography

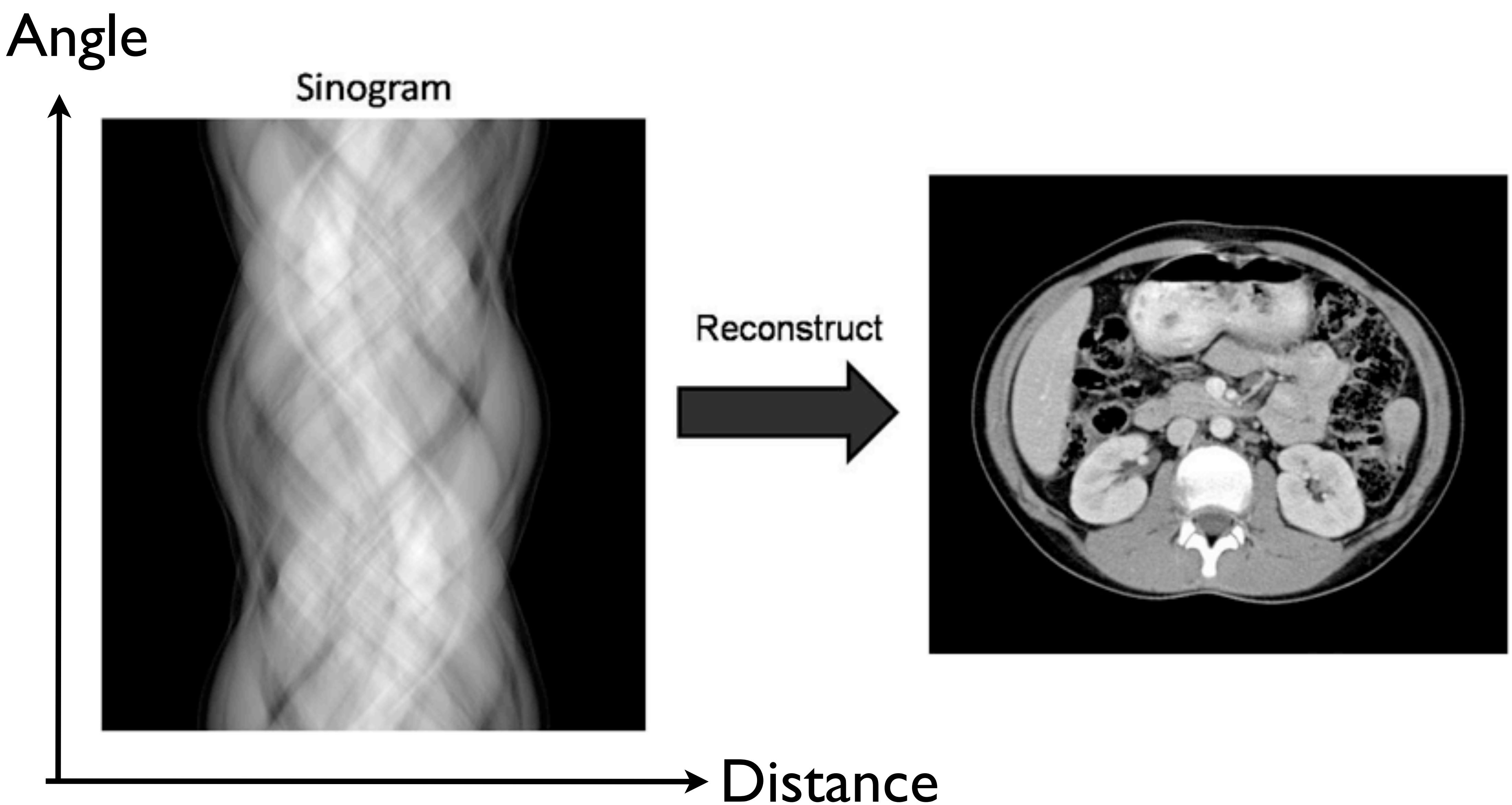




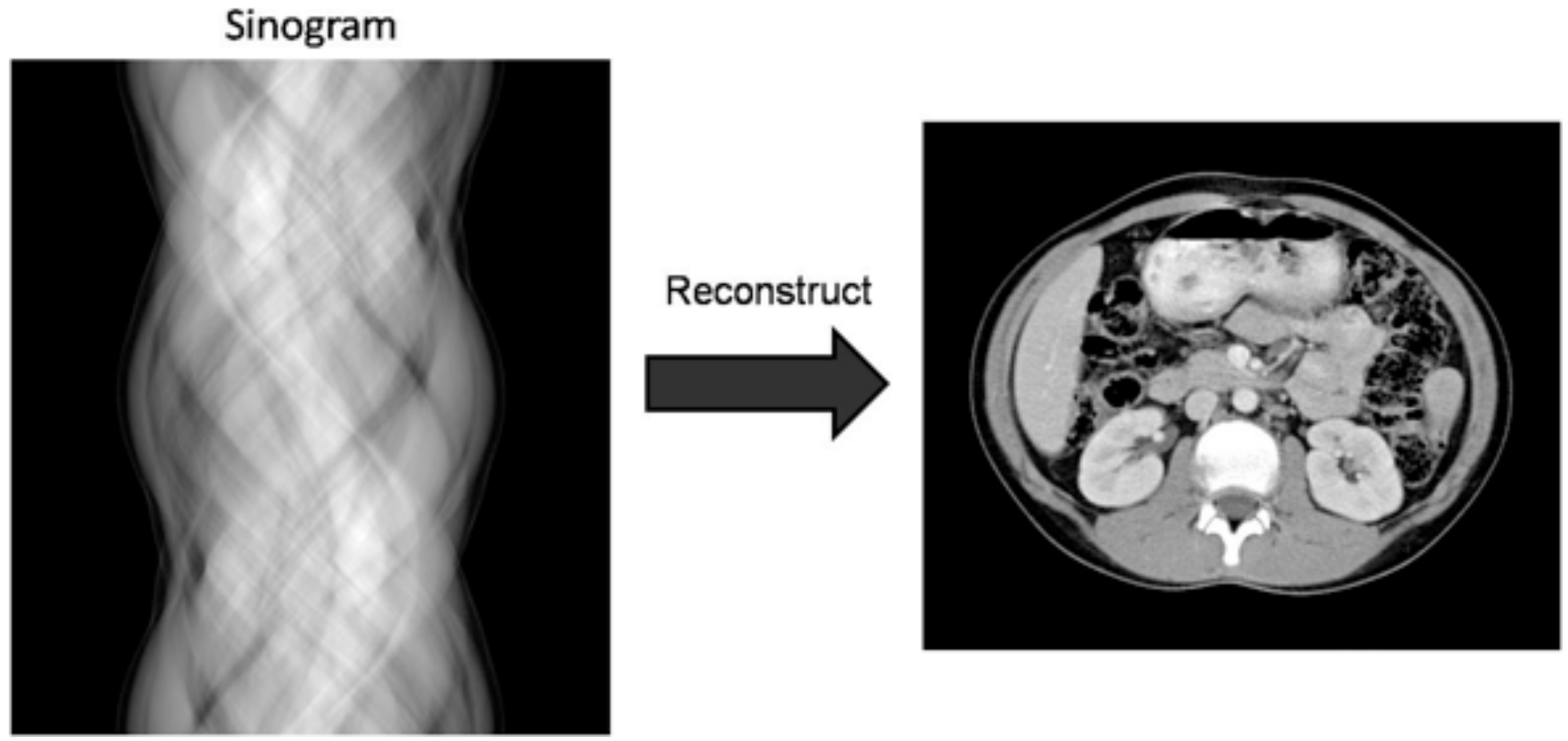


X-ray computed tomography

The reconstruction problem



Radon and inverse Radon Transform



$$
\begin{aligned}
Rf(\alpha, s) &= \int_{-\infty}^{\infty} f(x(t), y(t))\, dt \\
&= \int_{-\infty}^{\infty} f\big((t\sin\alpha + s\cos\alpha), (-t\cos\alpha + s\sin\alpha)\big)\, dt
\end{aligned}
$$

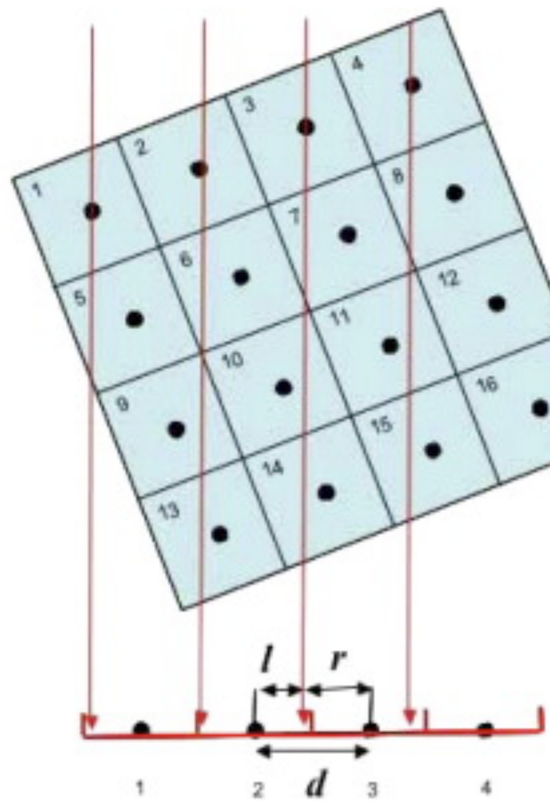
Direct

Inverse

$$
f(\mathbf{r}) = \frac{1}{2\pi} \int_{0}^{\pi} \int_{-\infty}^{+\infty} |k| \widetilde{(\widehat{f})}(k, \mathbf{u_t}) e^{+i(\mathbf{r}\cdot\mathbf{u_t})k}\, \mathrm{d}k\, \mathrm{d}\Phi
$$

Works well, but need the knowledge of all possible projections!

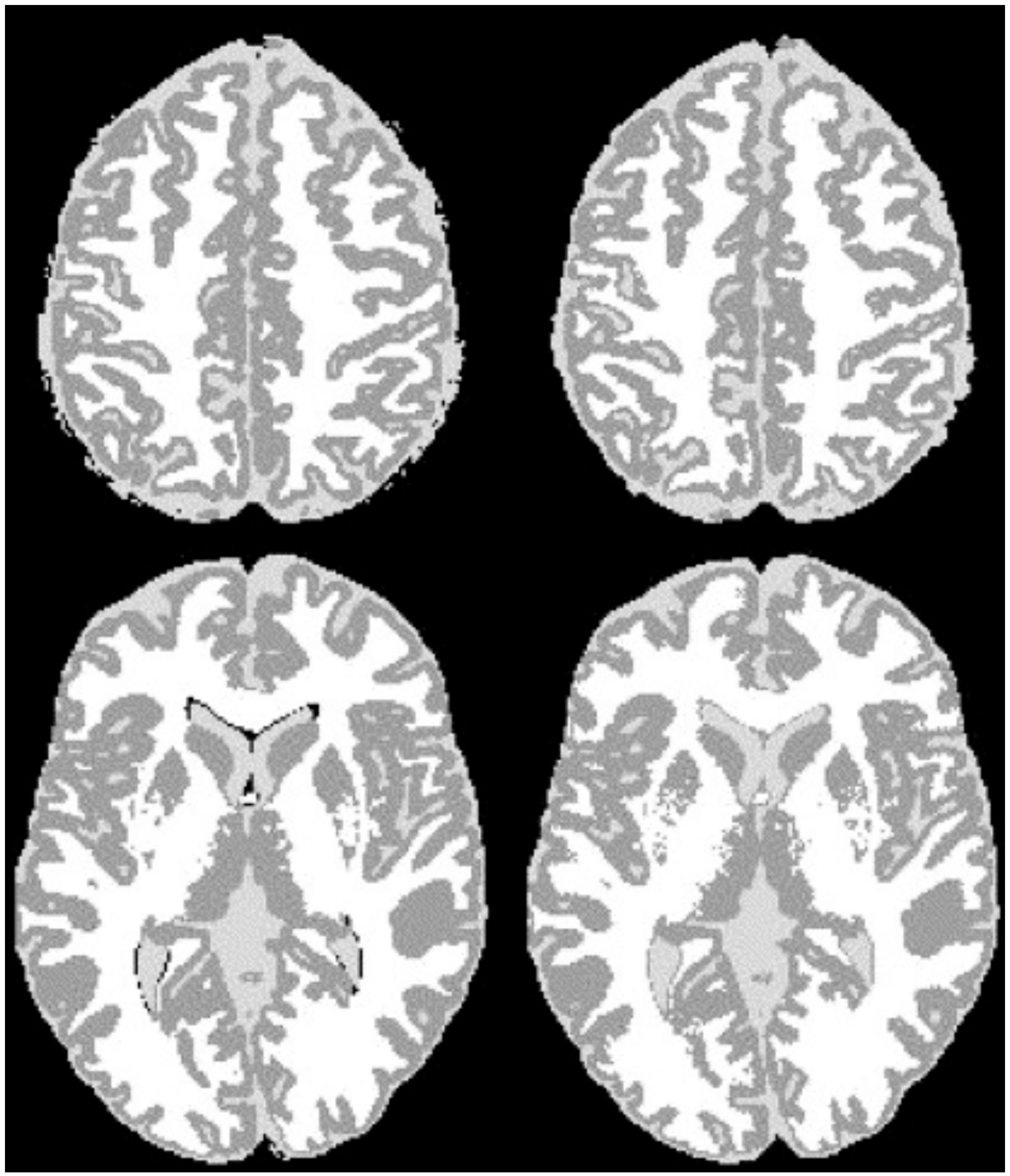
Algebraic methods: Inverting the matrix!



$$\underbrace{\begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ g_5 \\ g_6 \end{bmatrix}}_{g} = \underbrace{\begin{bmatrix} 001001001 \\ 010010010 \\ 100100100 \\ 111000000 \\ 000111000 \\ 000000111 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \\ f_9 \end{bmatrix}}_{f} .$$

Can one reconstruct when M (number of measurements) is smaller than N (number of "pixel") ?

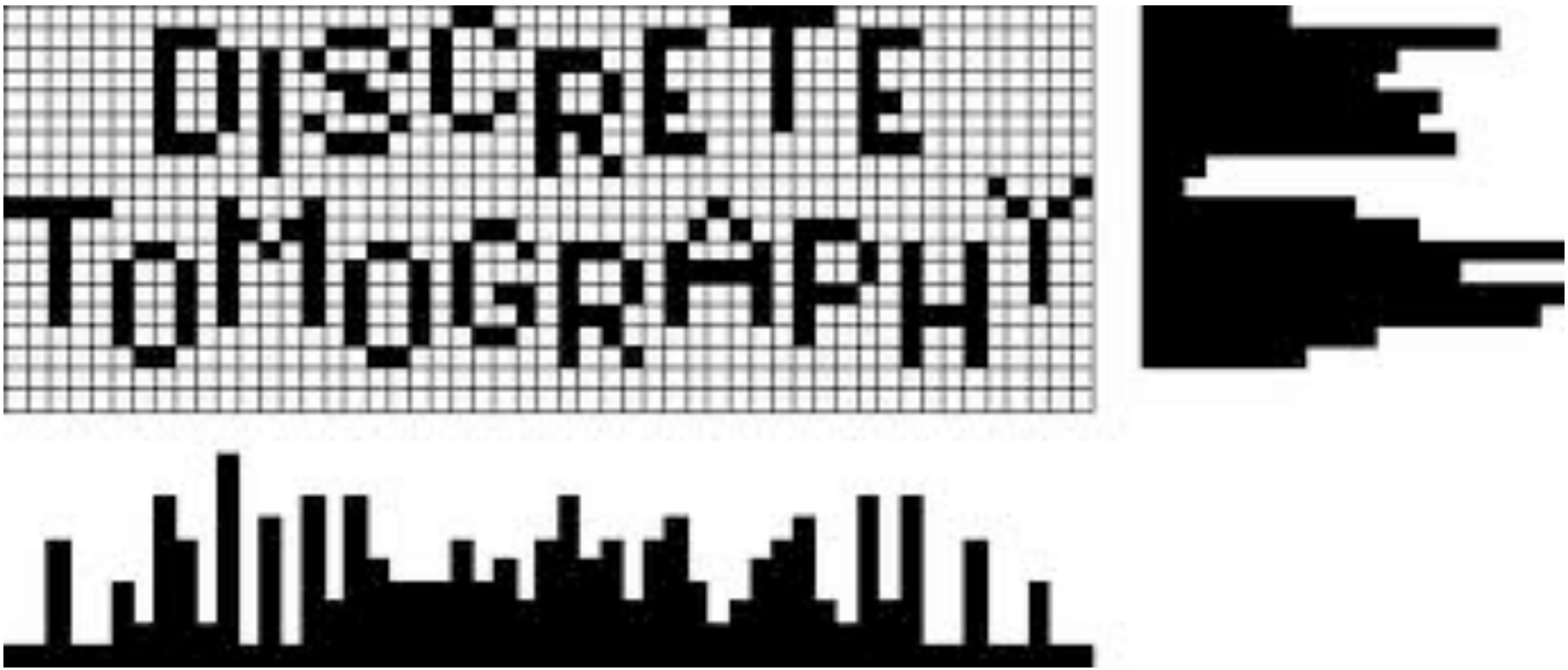
Discrete tomography



1) The image take discrete values: discrete tomography

2) Interfaces are rare

The problem:
Example with two angles



In general: NP-hard problem for 3 angles and more

☞Popular game known as PICROSS

Our work

A probabilistic approach to X-ray tomography

$$P(\{\vec{S}\}) \propto \prod_{\mu=1}^{M} \delta(y_\mu - \sum_{i\in\mu} S_i) \prod_{\mu=1}^{M} e^{J\sum_{i\in\mu} S_i S_{i+1}}$$

Solution of the linear system

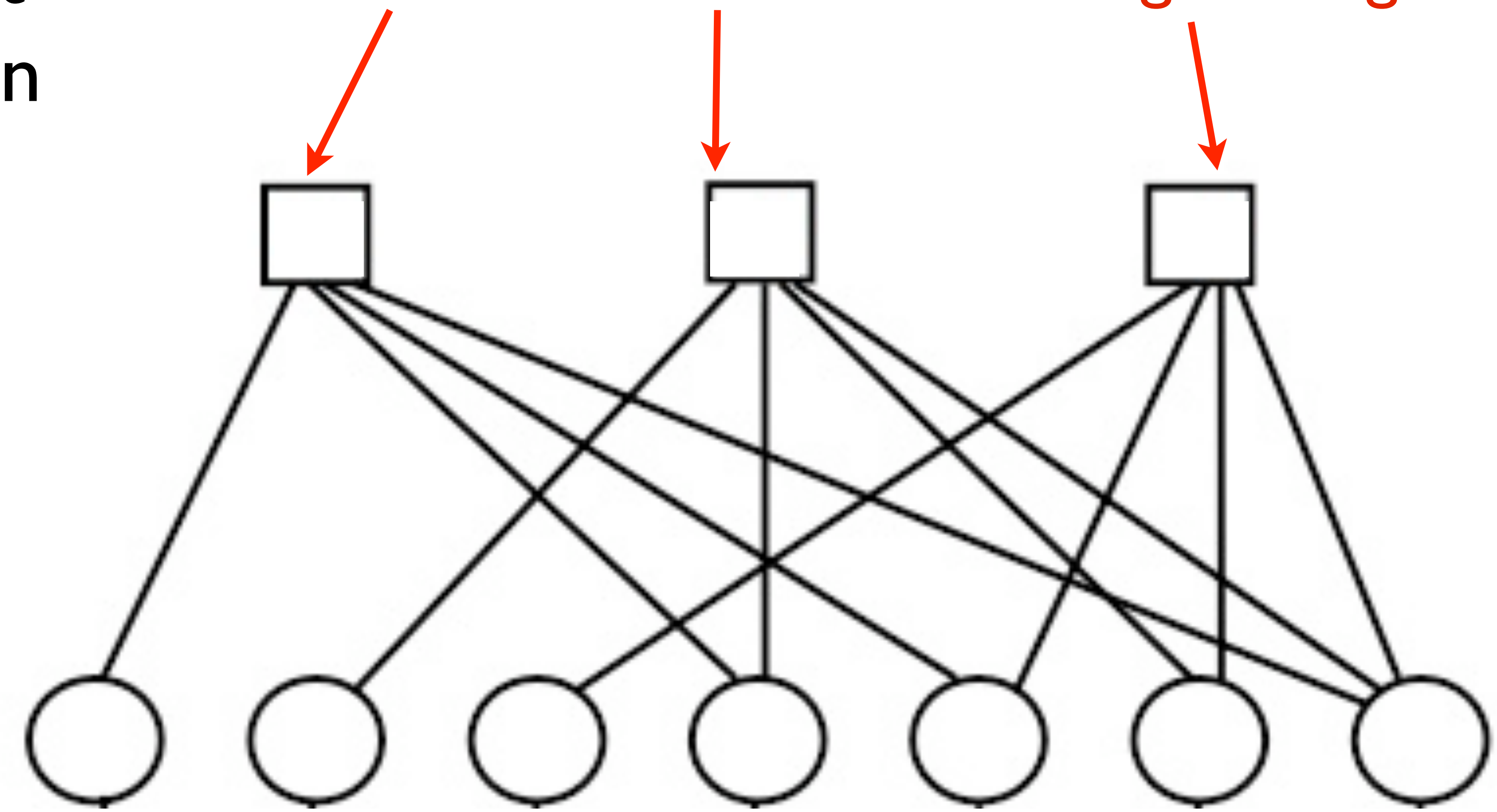
Prior on the images

BP for Discrete Tomography

Fix the sum of the spins for a given projection ...

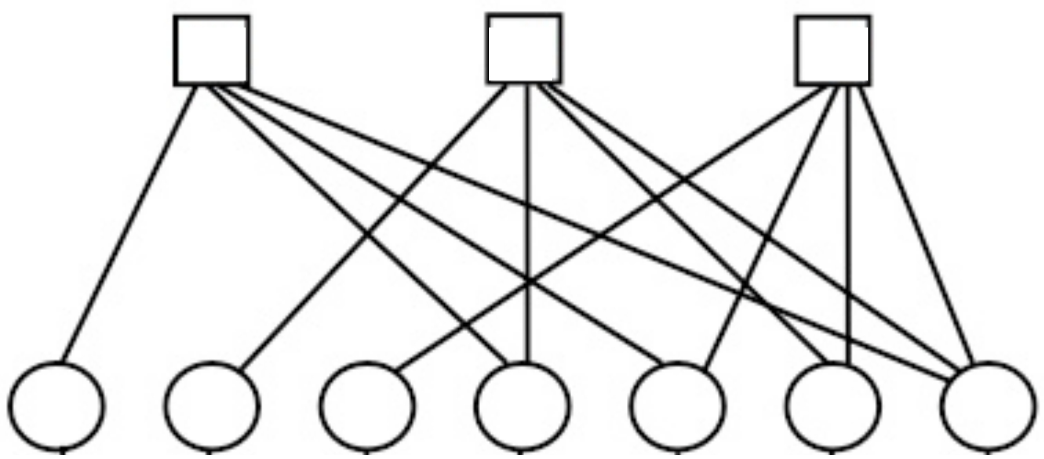
1 constraint by projection

... and take care of the first-neighboring interaction



Pixels: $S_i = \pm 1$

BP for Discrete Tomography



The usual BP equations

From spin to factor nodes:

$$m_{i\to\gamma}(\sigma_i) \propto \prod_{\mu\in i\neq\gamma} \tilde{m}_{\mu\to i}(\sigma_i)$$

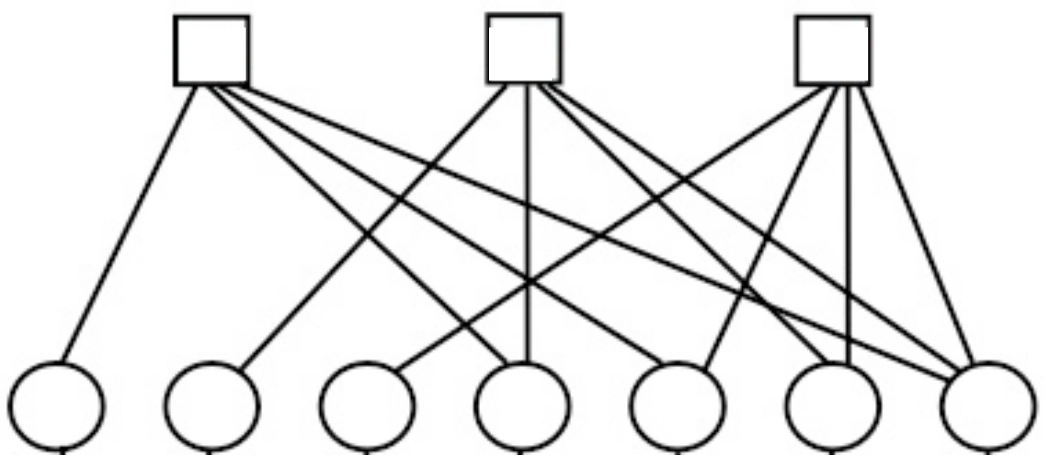
$$m = \frac{e^h}{\cosh h}$$

$$h_{i\to\gamma} = \sum_{\mu\in i\neq\gamma} \tilde{h}_{\mu\to i}$$

From nodes to spins:

$$\tilde{m}_{\mu\to i}(\sigma_i) \propto \sum_{\sigma\in\mu\neq i} \delta(y_\mu - \sum_{j\in\mu} \sigma_j) e^{J_\mu \sum_{j\in\mu} \sigma_j \sigma_{j+1}} \prod_{j\in\mu\neq i} m_{j\to\mu}(\sigma_j)$$

BP for Discrete Tomography

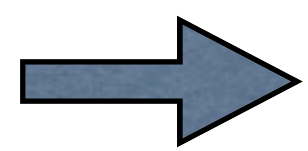


The usual BP equations

From spin to factor nodes:

$$m_{i\to\gamma}(\sigma_i) \propto \prod_{\mu\in i\neq\gamma} \tilde{m}_{\mu\to i}(\sigma_i)$$

$$m = \frac{e^h}{\cosh h}$$

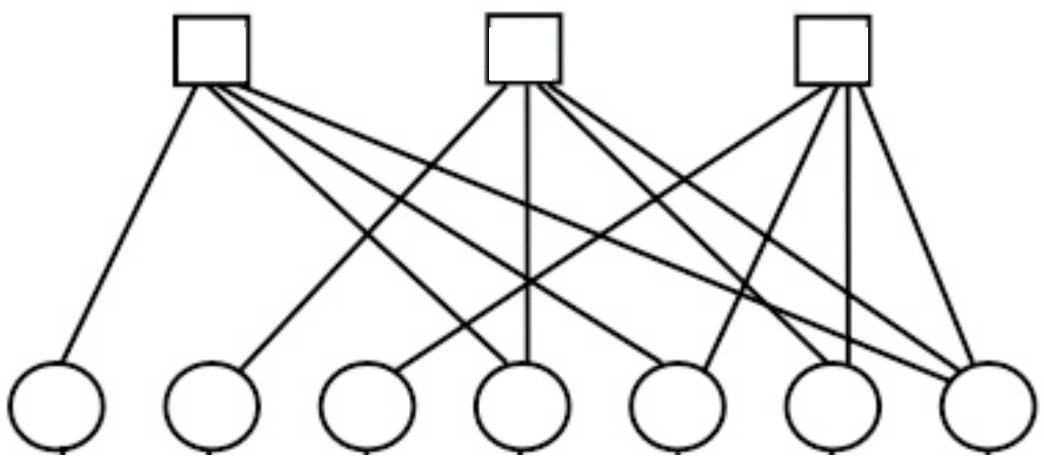


$$h_{i\to\gamma} = \sum_{\mu\in i\neq\gamma} \tilde{h}_{\mu\to i}$$

From nodes to spins:

$$\tilde{m}_{\mu\to i}(\sigma_i) \propto \sum_{\sigma\in\mu\neq i} \delta(y_\mu - \sum_{j\in\mu}\sigma_j) e^{J_\mu \sum_{j\in\mu}\sigma_j\sigma_{j+1} + \sum_{j\in\mu\neq i} h_{j\to\mu}}$$

BP for Discrete Tomography



2^L operations!

☞Intractable

From spin to factor nodes:

$$m_{i\to\gamma}(\sigma_i) \propto \prod_{\mu\in i\neq\gamma} \tilde{m}_{\mu\to i}(\sigma_i)$$

$$m = \frac{e^h}{\cosh h}$$

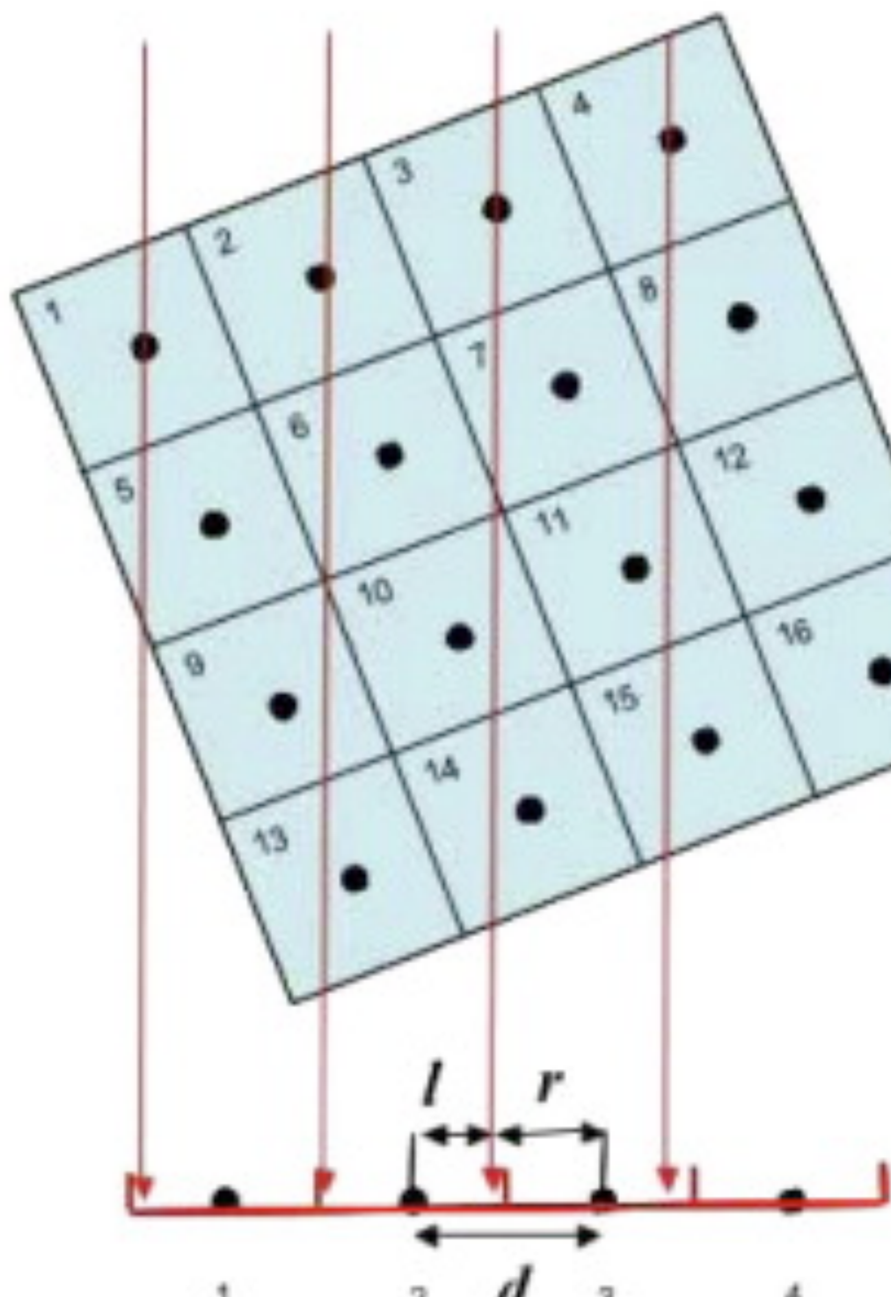
$$h_{i\to\gamma} = \sum_{\mu\in i\neq\gamma} \tilde{h}_{\mu\to i}$$

From nodes to spins:

$$\tilde{m}_{\mu\to i}(\sigma_i) \propto \sum_{\sigma\in\mu\neq i} \delta(y_\mu - \sum_{j\in\mu}\sigma_j) e^{J_\mu \sum_{j\in\mu}\sigma_j\sigma_{j+1} + \sum_{j\in\mu\neq i} h_{j\to\mu}}$$

In each constraint: BP in BP!

$$\tilde{m}_{\mu\to i}(\sigma_i) \propto \sum_{\sigma\in\mu\neq i} \delta(y_\mu - \sum_{j\in\mu}\sigma_j) e^{J_\mu \sum_{j\in\mu}\sigma_j\sigma_{j+1} + \sum_{j\in\mu\neq i} h_{j\to\mu}}$$



All the spin involved in one given constraints are just neighboring spins on a line

One needs to estimates the marginal of variables on a one-dimension chain in random field

Use BP!

In each constraint: BP in BP!

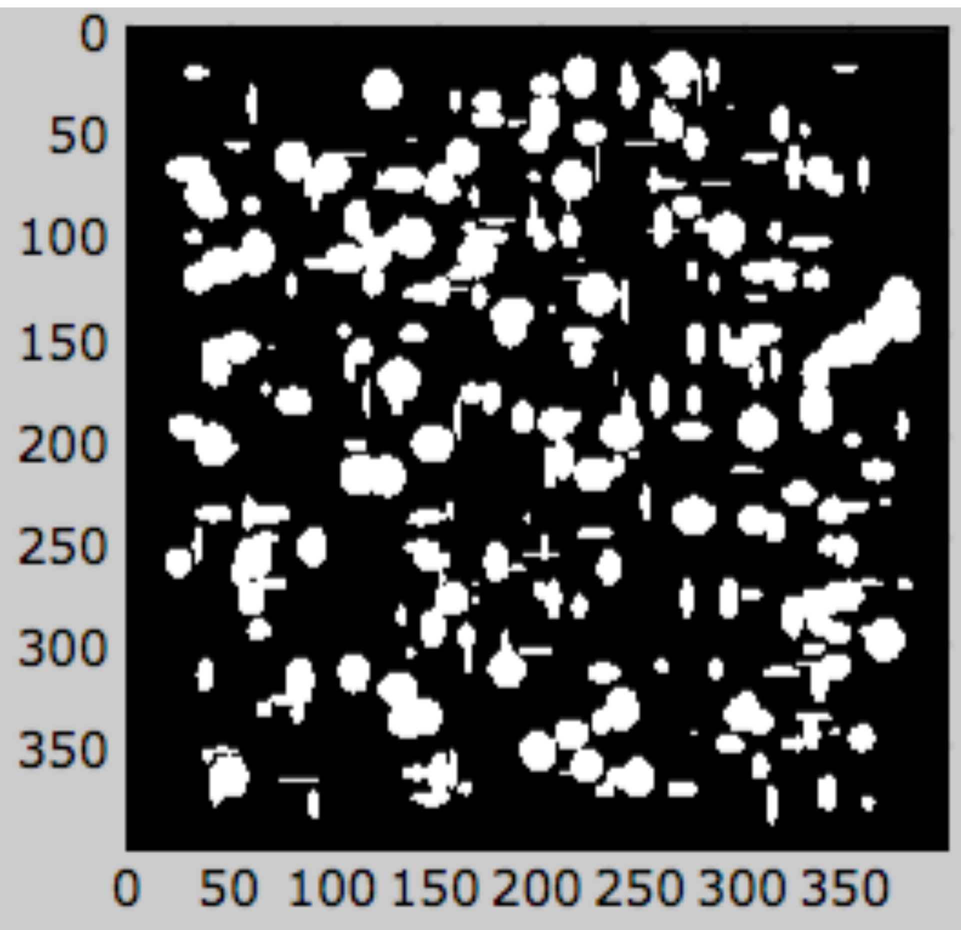
$$\tilde{m}_{\mu\to i}(\sigma_i) \propto \sum_{\sigma\in\mu\neq i} \delta(y_\mu - \sum_{j\in\mu}\sigma_j) e^{J_\mu \sum_{j\in\mu}\sigma_j\sigma_{j+1} + \sum_{j\in\mu\neq i} h_{j\to\mu}}$$

Replace the delta by a
Lagrange multiplier (magnetic field)

$$\tilde{m}_{\mu\to i}(\sigma_i) \propto \sum_{\sigma\in\mu\neq i} e^{H\sum_i S_i} e^{J_\mu \sum_{j\in\mu}\sigma_j\sigma_{j+1} + \sum_{j\in\mu\neq i} h_{j\to\mu}}$$

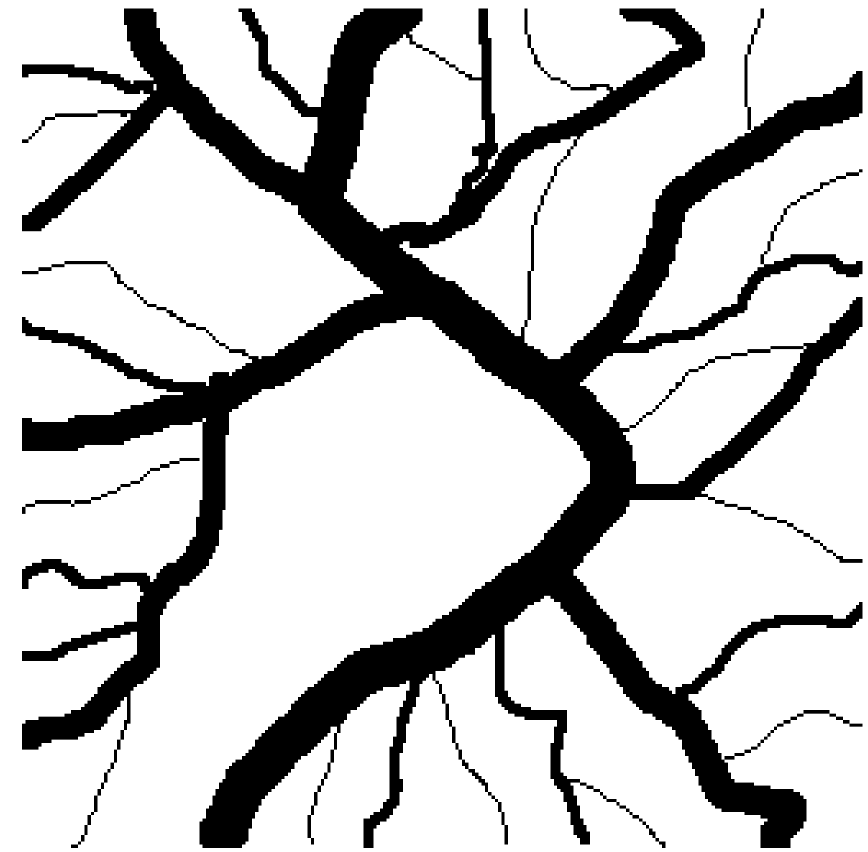
Find H such that the delta constraint is satisfied
(by Dichotomy or using Newton method)

BP at works...



30 angles

14 angles



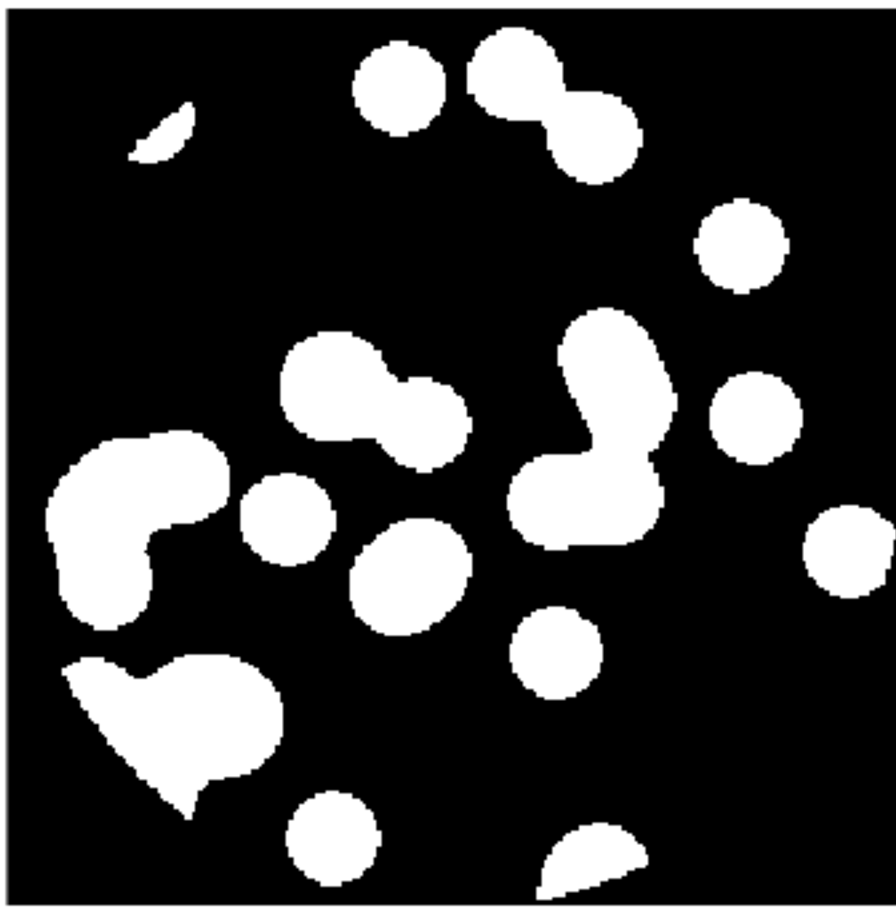
17 angles

Fast, and need for only few projections

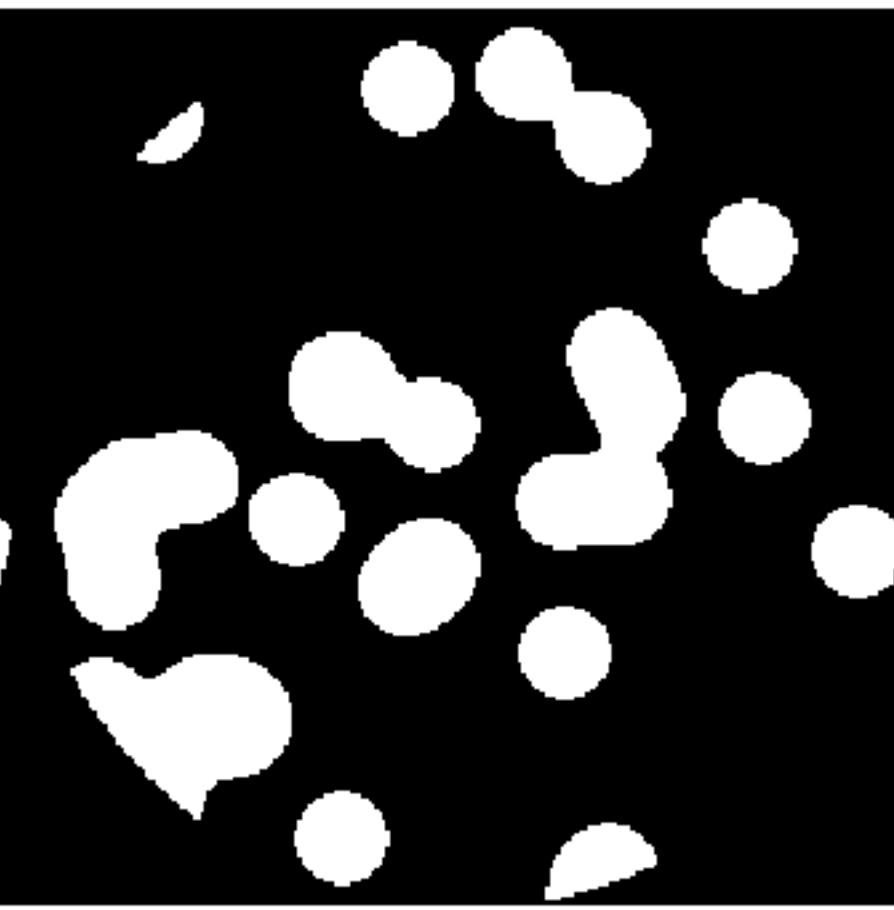
Robust to noise!

Adding a noise to the projections

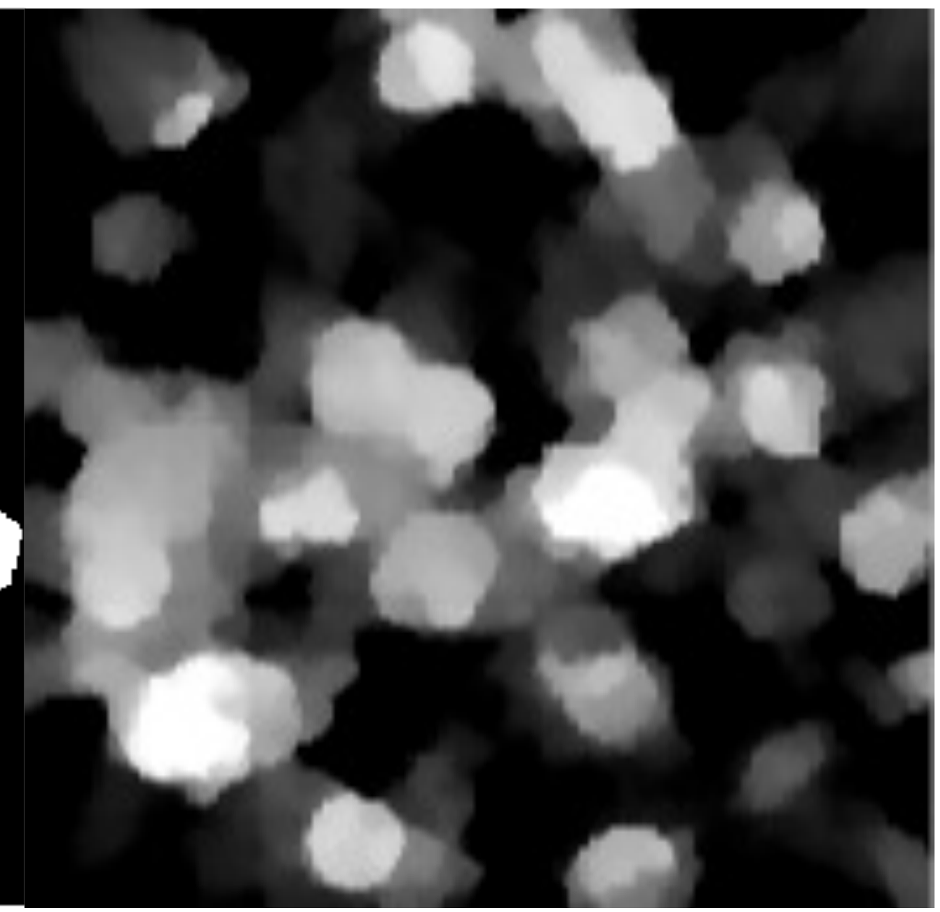
From 6 angles...



Original



BP



Continuous
+Total Variation

(i.e. LASSO-type problem)

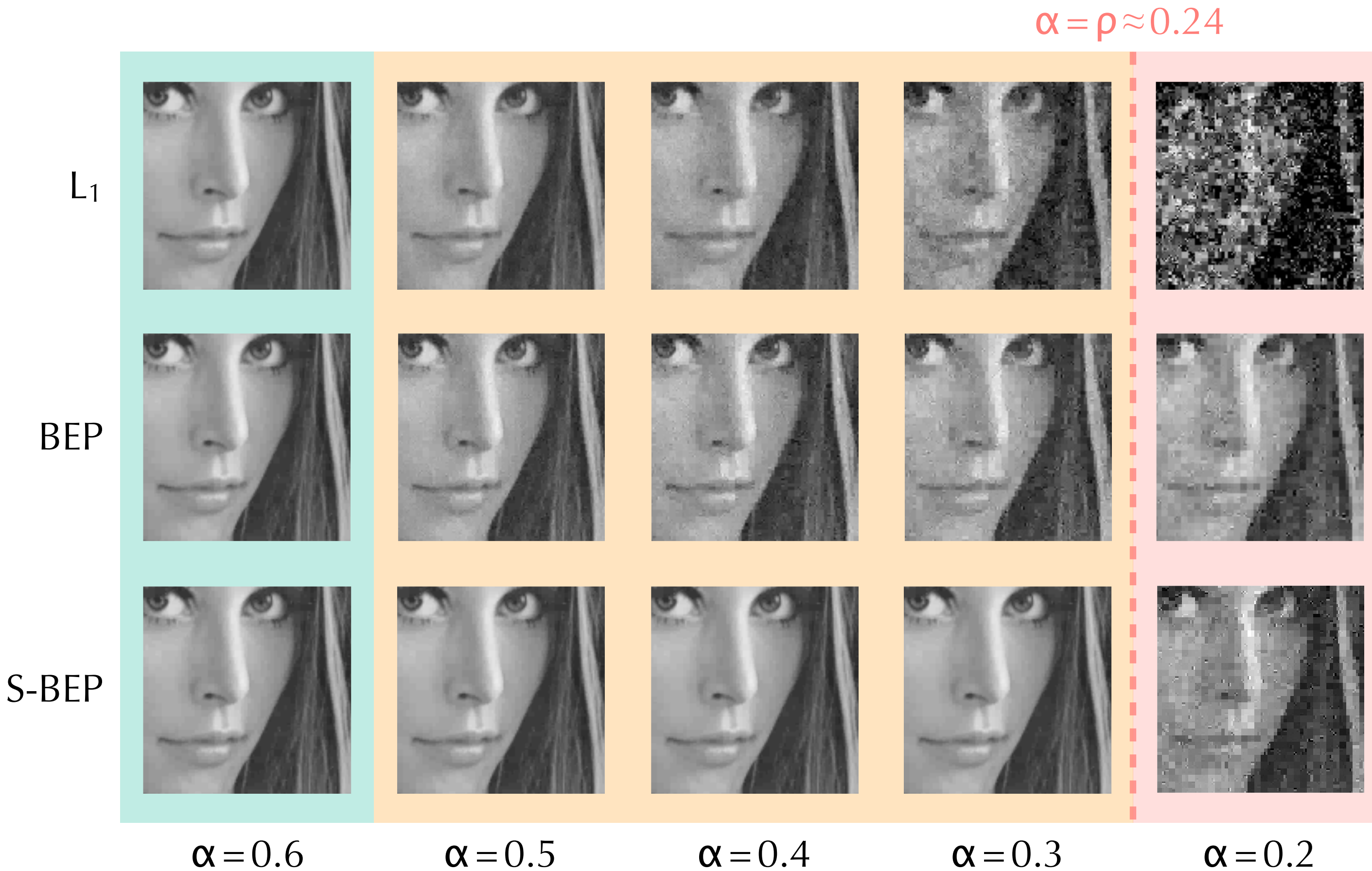
Conclusions...

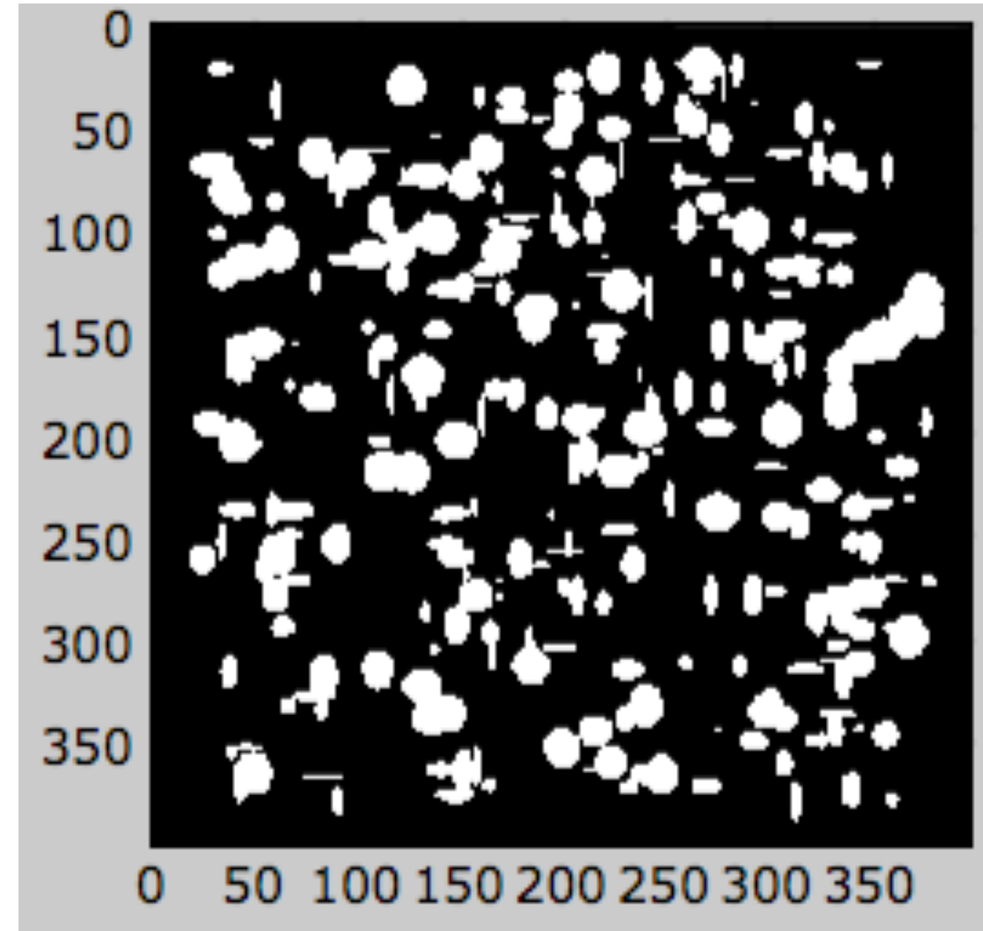
- Probabilistic approach to reconstruction in tomography...
- ... with a Belief Propagation algorithm
- 1d ising model (BP in BP)

... and perspectives:

- Q-state and continuous tomographic reconstruction?
- Multi-scale approach?
- Generic message: LASSO type problem are often better to be replace by a probabilistic approach with a BP algorithm.
- ☞Toward Applications!

Thanks for your attention





<u>COMING SOON:</u> Post-doc and Ph.d openings on these topics:

If you work in Statistical physics, Information science, Signal processing, etc...

ASPICS *Project*

Applying Statistical Physics to Inference in Compressed Sensing

<u>COMING SOON:</u> An interdisciplinary school on these topics:

Les Houches, October 2013, Organizers F. Krzakala & L. Zdeborová