BMS LECTURE SUPPORT SKIDES

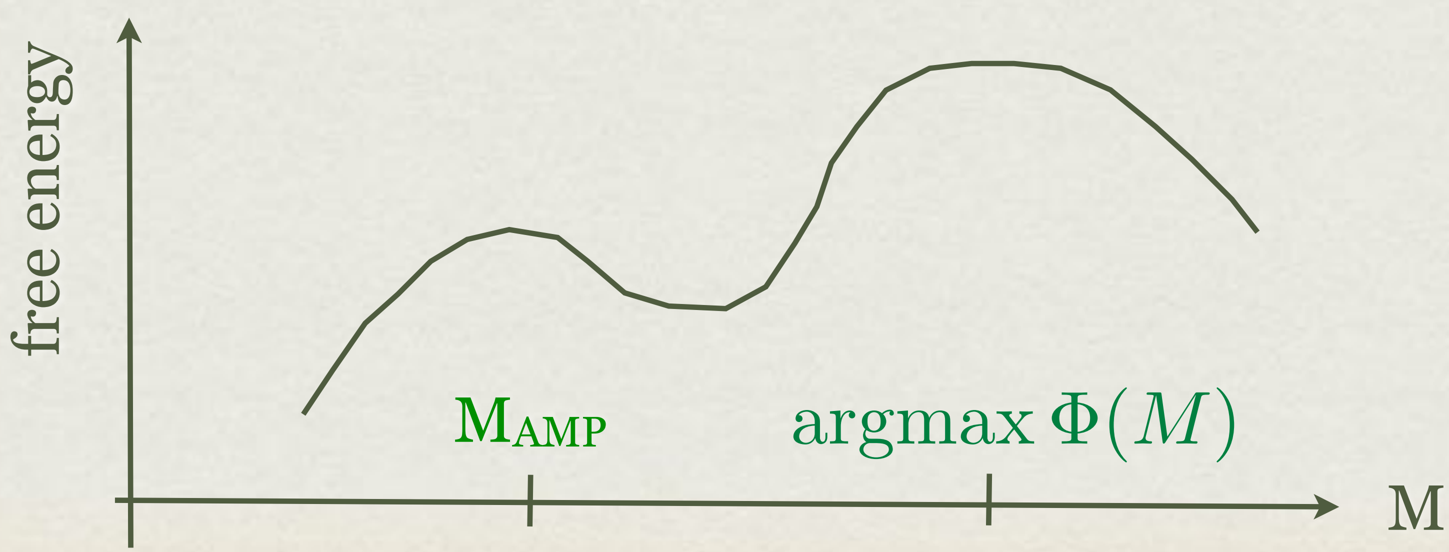
Lenka Zdeborová
(IPhT CEA Saclay)

BOTTOM LINE

$$\Phi(M) = \mathbb{E}_{x,w}\left[\log \mathcal{Z}\left(\frac{M}{\Delta}, \frac{M}{\Delta}x + \sqrt{\frac{M}{\Delta}}w\right)\right] - \frac{\mathrm{Tr}(MM^\top)}{4\Delta}$$

- AMP-MSE given by the local maximum of the free energy reached gradient descent starting from small M/large MSE.

- MMSE is given by the global maximum of the free energy.



$$\mathrm{MMSE} = \mathrm{Tr}[\mathbb{E}_x(xx^\top) - \mathrm{argmax}\,\Phi(M)]$$

$$\mathrm{MSE}_{\mathrm{AMP}} = \mathrm{Tr}[\mathbb{E}_x(xx^\top) - M_{\mathrm{AMP}}]$$

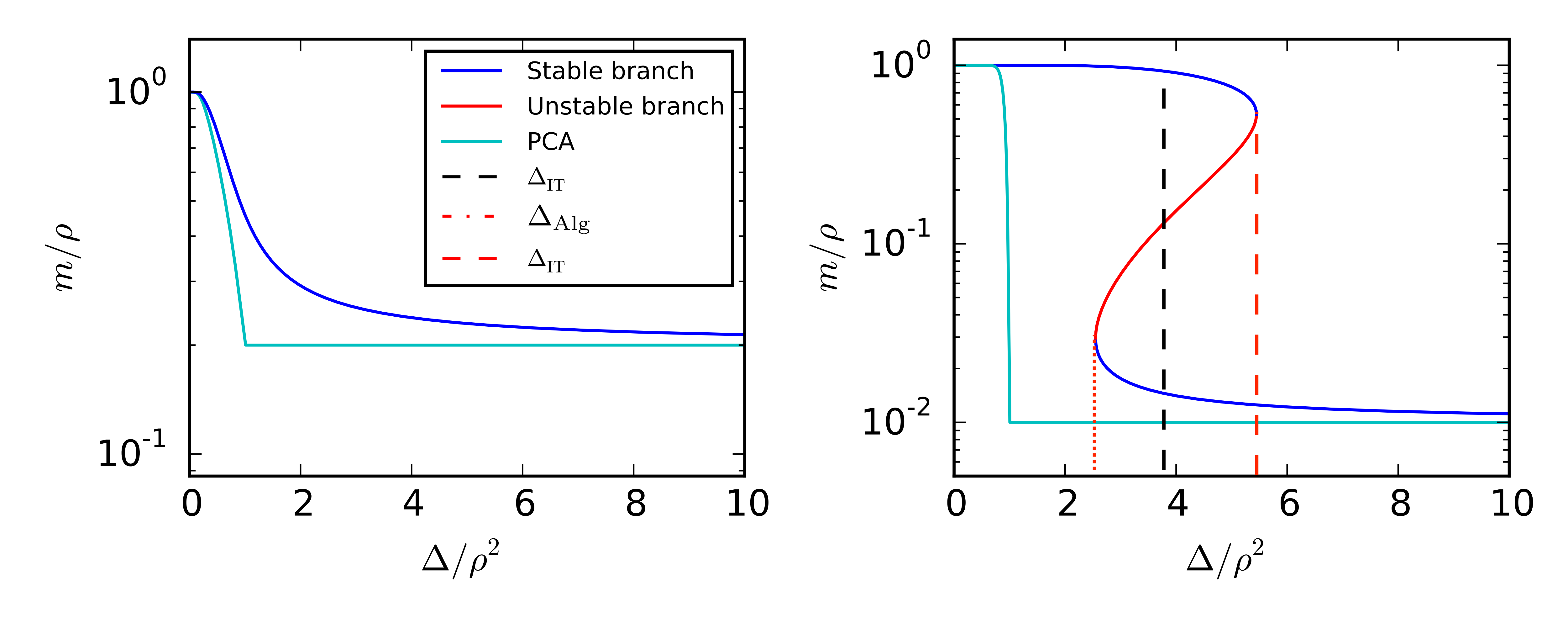
ZOOLOGY OF FIXED POINTS

- Zero mean prior: $\mathbb{E}_X(x) = 0$
  - SE has always a "trivial" fixed point M=0.
  - Stability of the trivial fixed point:

$$M^{t+1} = \frac{\Sigma M^t \Sigma}{\Delta} \qquad\qquad M^{t+1}_{(r=1)} = \frac{[\mathbb{E}_X(x^2)]^2}{\Delta} M^t$$

  - This is the same as the spectral phase transition (BBP transition, known in physics since Edwards'68)
- Non-zero mean priors: $\mathbb{E}_X(x) \neq 0$
  - MMSE always better than random guessing (unlike for spectral methods).
  - Multiple fixed point may still exist.

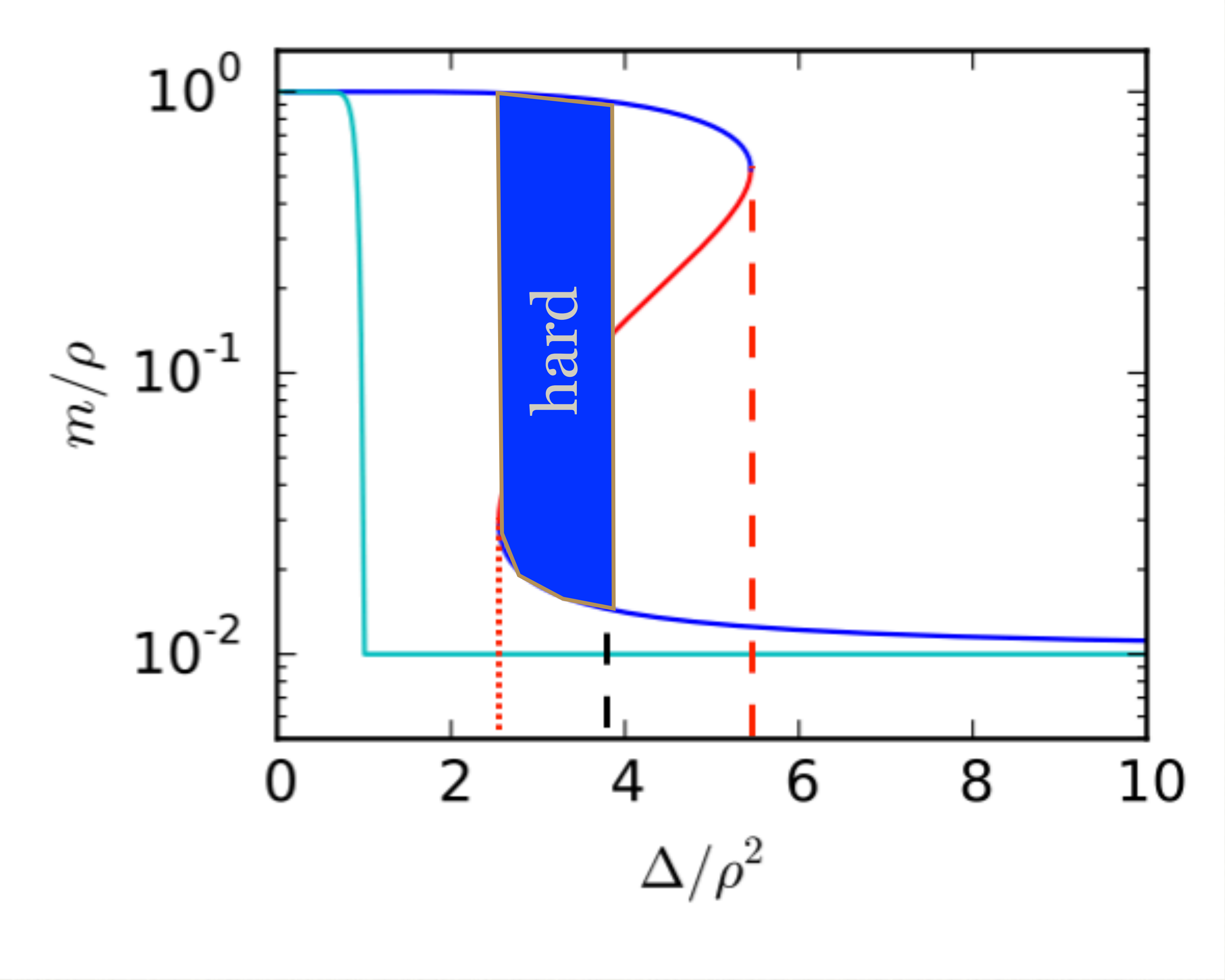
Fixed points:



$\rho = 0.2$ $\rho = 0.01$

$$P_X(x_i) = (1-\rho)\delta(x_i) + \rho\delta(x_i - 1)$$

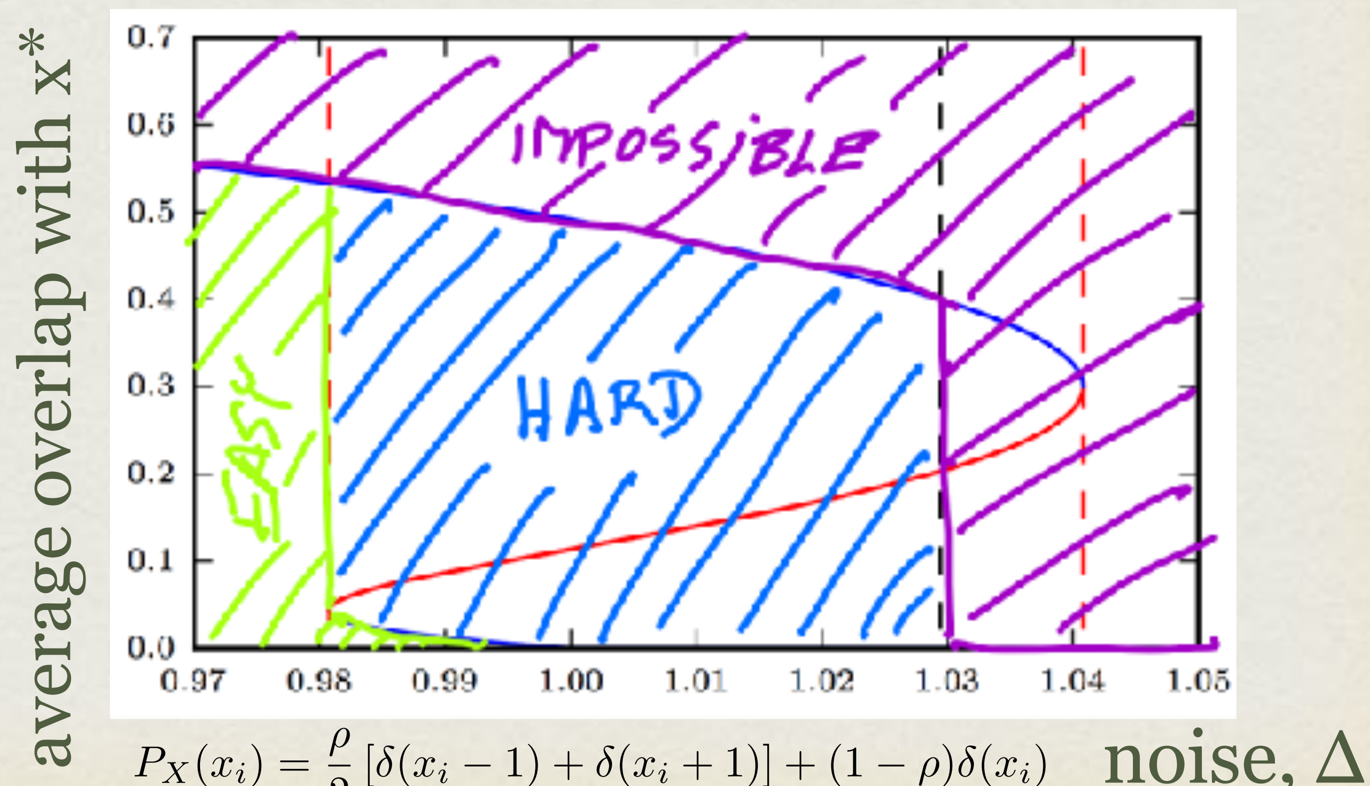
Fixed points:



$$P_X(x_i) = (1-\rho)\delta(x_i) + \rho\delta(x_i - 1)$$

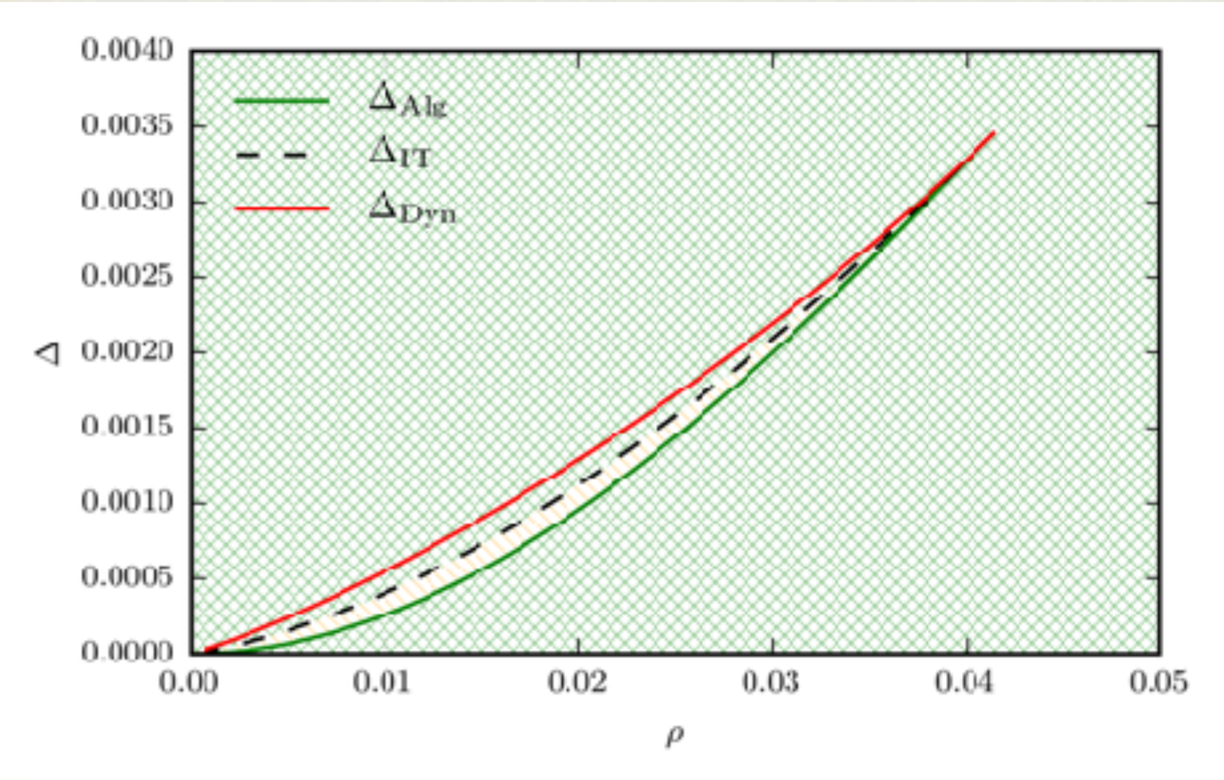
ALGORITHMIC INTERPRETATION

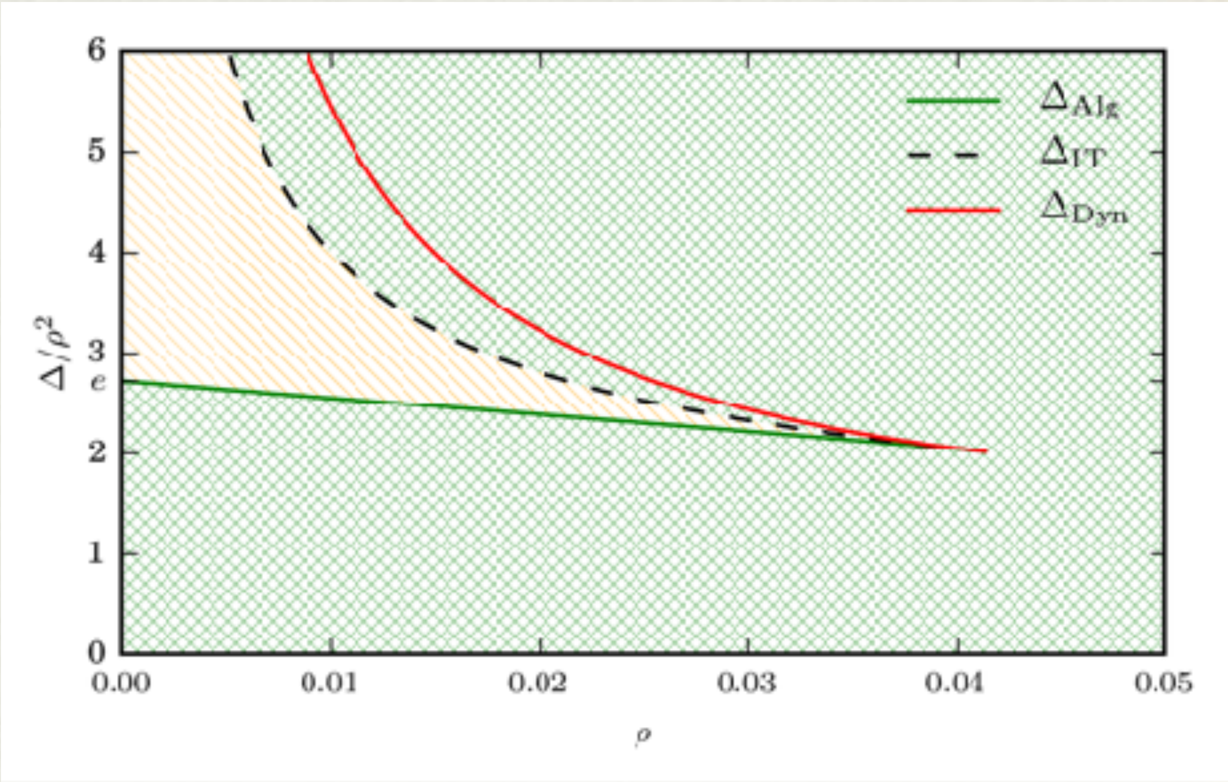
- Easy by message passing algorithms. Proof: Javanmard, Montanari'12.
- Impossible information theoretically. Barbier, Dia, Macris, Krzakala, Lesieur, LZ'16 using Guerra interpolation, AMP and spatial coupling. Different proof by Lelarge, Miolane'16.
- Hard phase conjecture: No polynomial algorithm works.
  Physically sensible.
  Mathematically wide open.



$$P_X(x_i) = \frac{\rho}{2}\left[\delta(x_i-1)+\delta(x_i+1)\right]+(1-\rho)\delta(x_i)$$

Phase Diagram:

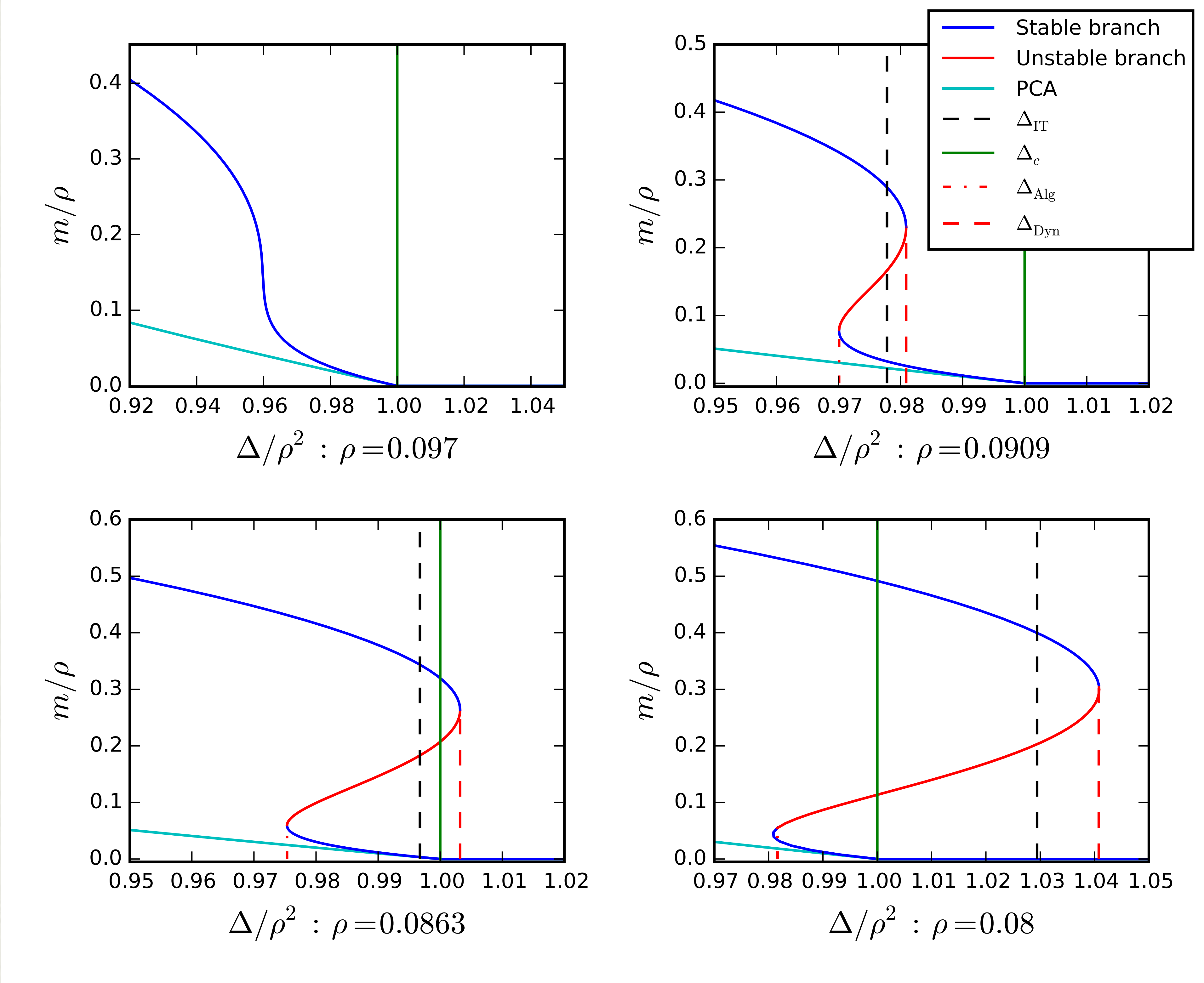




$$\Delta_{\mathrm{Dyn}}(\rho) \sim_{\rho \to 0} \frac{-\rho}{2\log(\rho)}$$

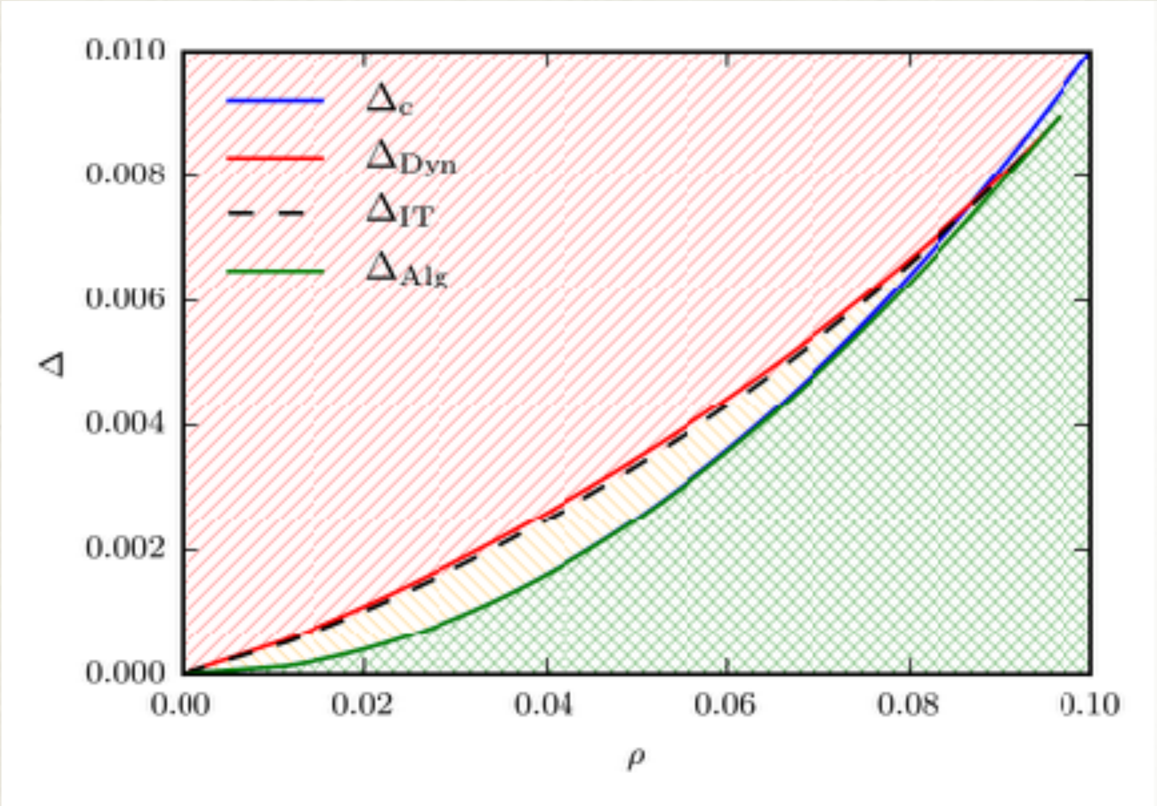
$$\Delta_{\mathrm{IT}}(\rho) \sim_{\rho \to 0} \frac{-\rho}{4\log(\rho)}$$

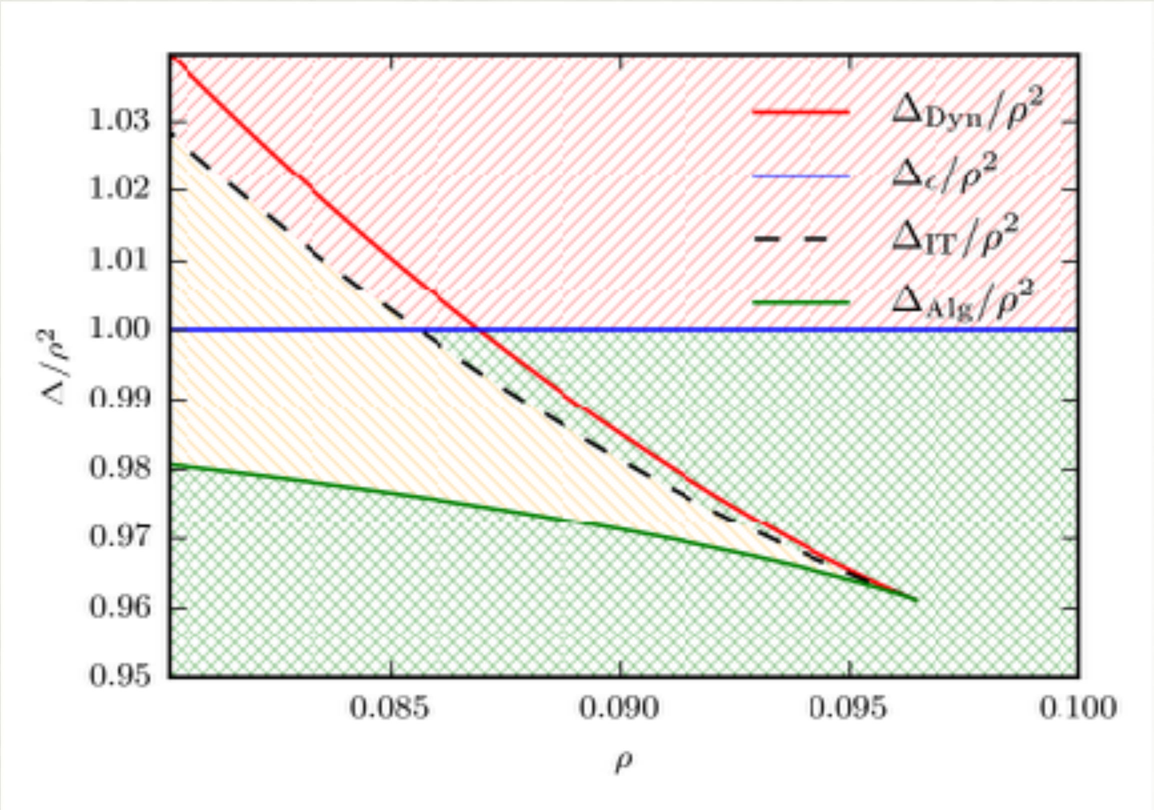
$$P_X(x_i) = (1-\rho)\delta(x_i) + \rho\delta(x_i - 1)$$



$$P_X(x_i) = \frac{\rho}{2}\left[\delta(x_i - 1) + \delta(x_i + 1)\right] + (1-\rho)\delta(x_i)$$

Phase Diagram:

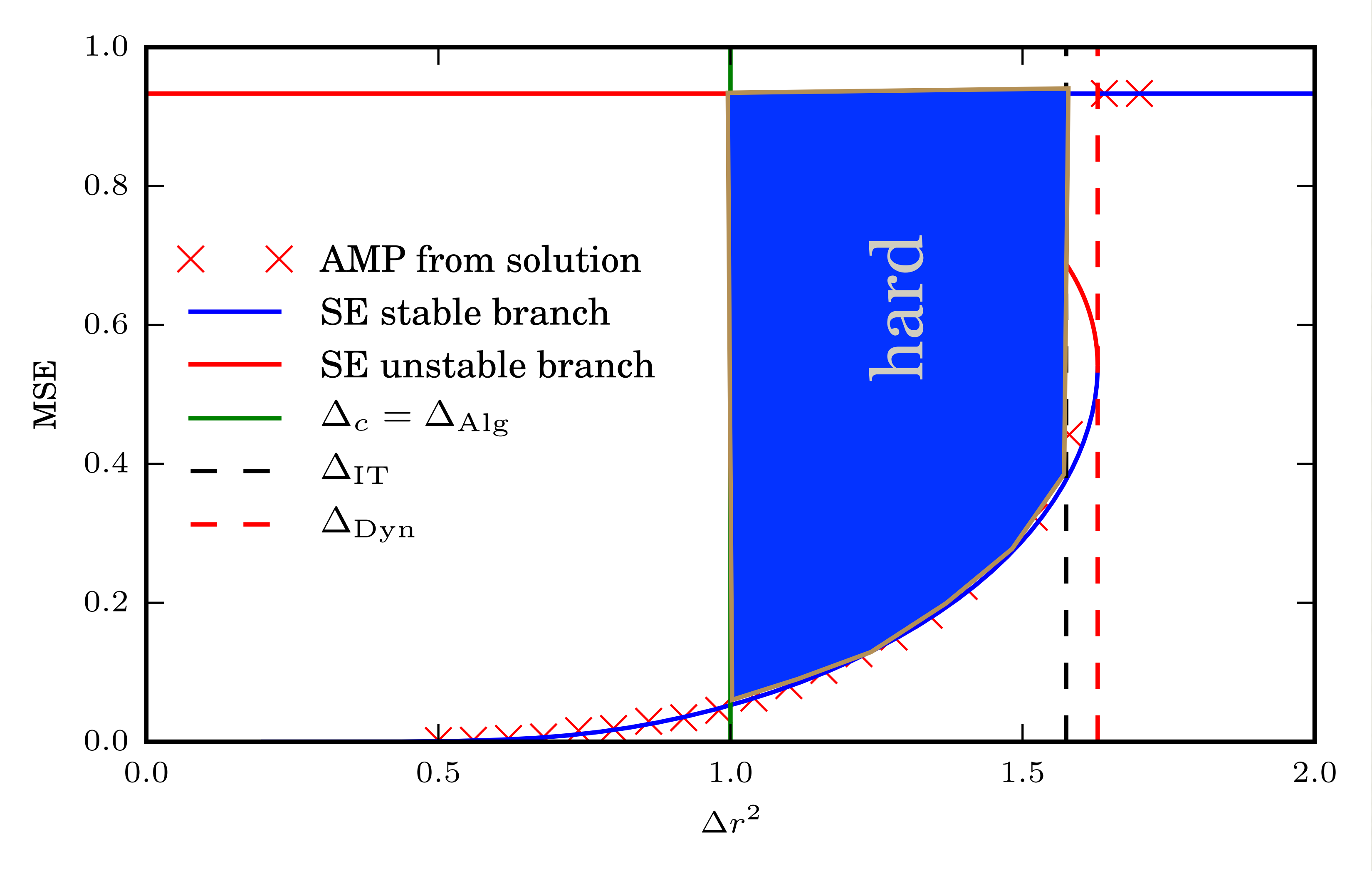




$$P_X(x_i) = \frac{\rho}{2}\left[\delta(x_i - 1) + \delta(x_i + 1)\right] + (1 - \rho)\delta(x_i)$$

Stochastic block model, r groups

$$P_{\rm out}(Y_{ij}=1|\frac{x_i^T x_j}{\sqrt{N}}) = p_{\rm out} + \frac{\mu}{\sqrt{N}} x_i^T x_j \qquad \Delta = \frac{p_{\rm out}(1-p_{\rm out})}{\mu^2}$$



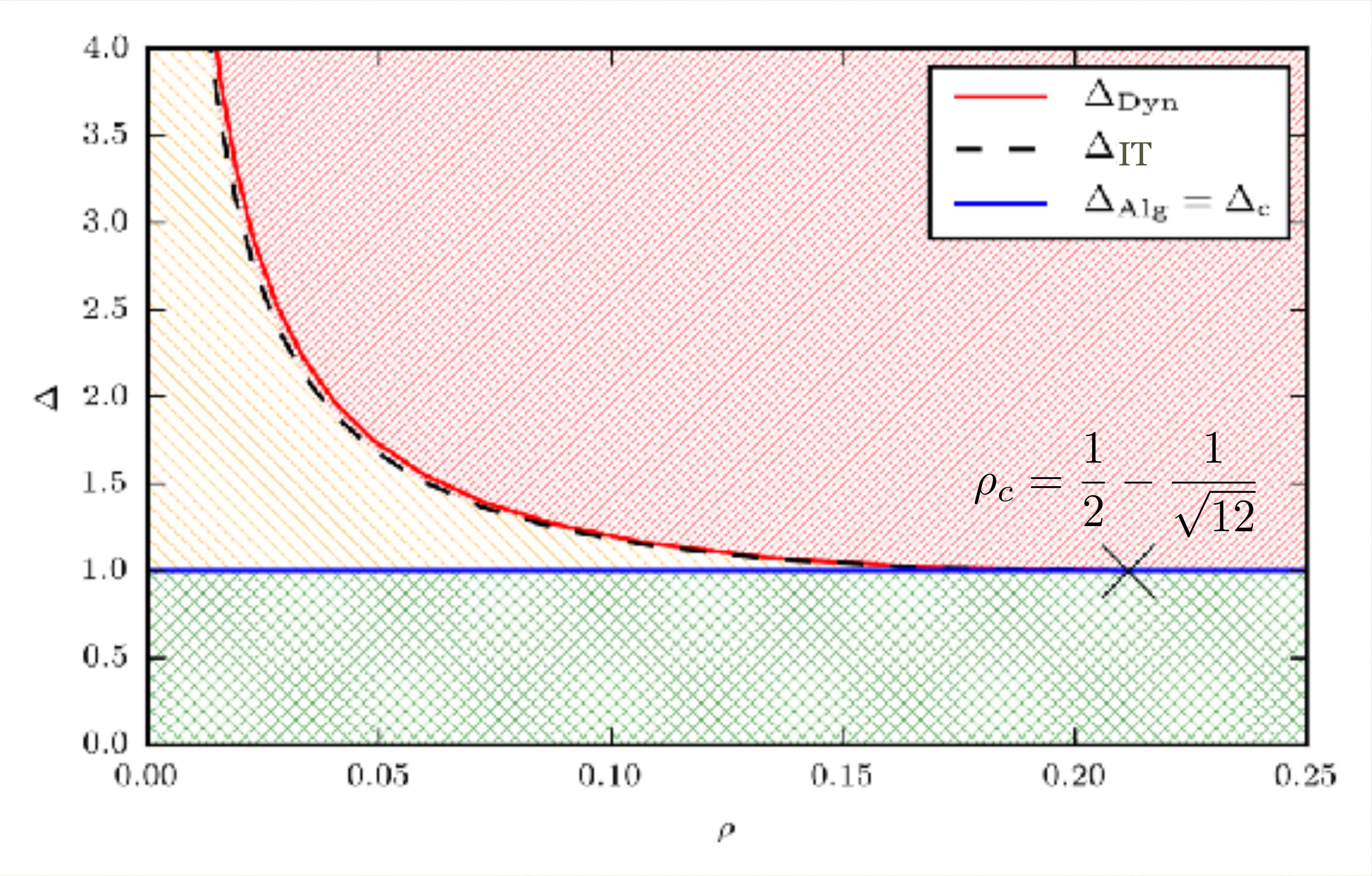
r=15

r>4 hard phase exists. r<4 hard phase does not exist.

2 groups, different sizes, same average degree

$$\begin{pmatrix} p_{\rm out} & p_{\rm out} \\ p_{\rm out} & p_{\rm out} \end{pmatrix} + \frac{\mu}{\sqrt{N}} \begin{pmatrix} \frac{1-\rho}{\rho} & -1 \\ -1 & \frac{\rho}{1-\rho} \end{pmatrix} \qquad \Delta = \frac{p_{\rm out}(1-p_{\rm out})}{\mu^2}$$

$$P_X(x) = \rho\delta\left(x - \sqrt{\frac{1-\rho}{\rho}}\right) + (1-\rho)\delta\left(x + \sqrt{\frac{\rho}{1-\rho}}\right)$$



small ρ

$$k_{\rm Alg} = \sqrt{N}\sqrt{\frac{p_{\rm out}}{1-p_{\rm out}}}\,,$$

$$k_{\rm IT} = \log(N)\frac{4p_{\rm out}}{1-p_{\rm out}}\,.$$

As in balanced planted clique.

GAUSSIAN MIXTURE R=2 CLUSTERS

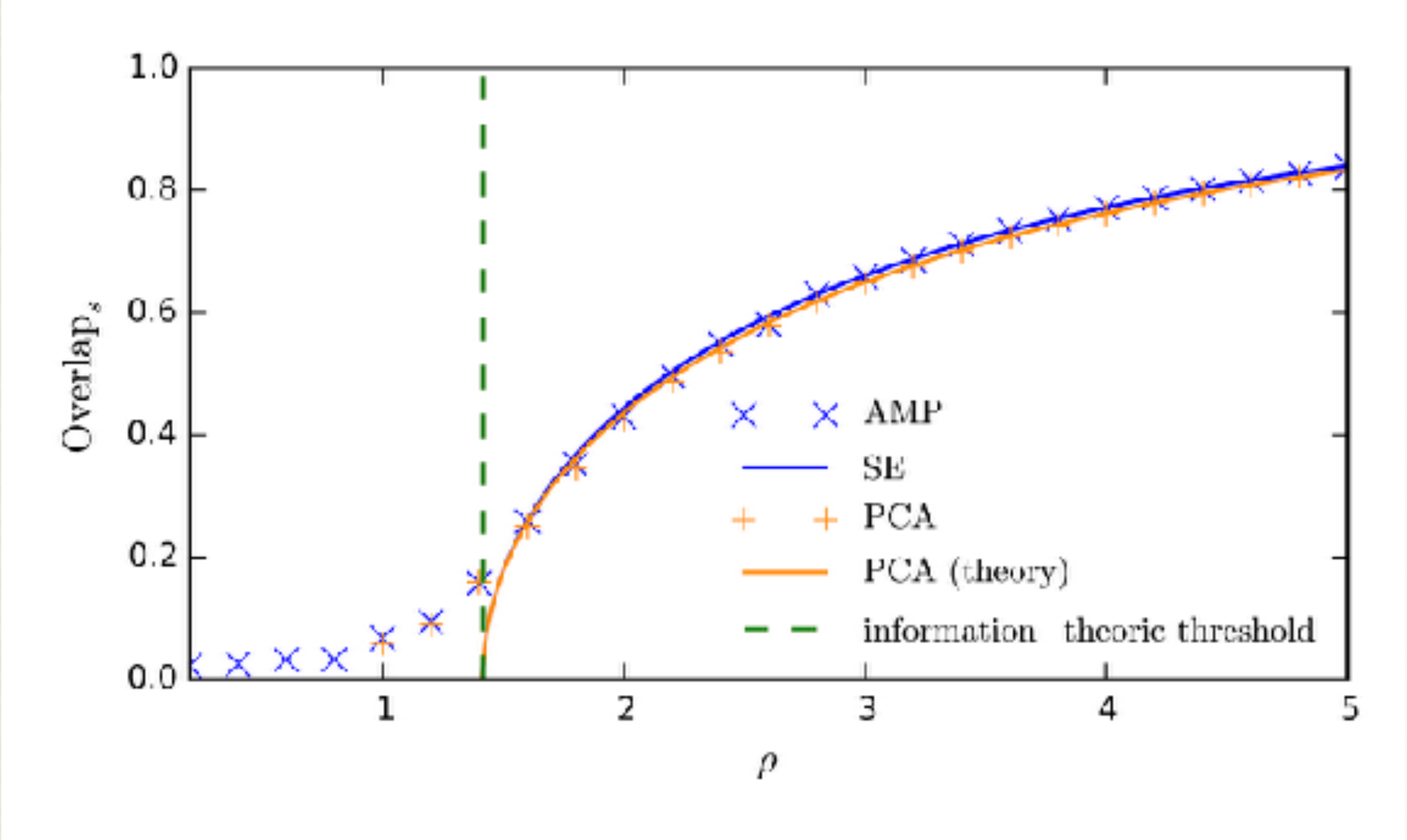
N points, M dimensions, distance between clusters ρ

$$\alpha = \frac{N}{M} = 2$$

Phase transition:

$$\rho_c = \frac{r}{\sqrt{\alpha}}$$

AMP is Bayes-optimal for R=2, and only very slightly better than PCA.



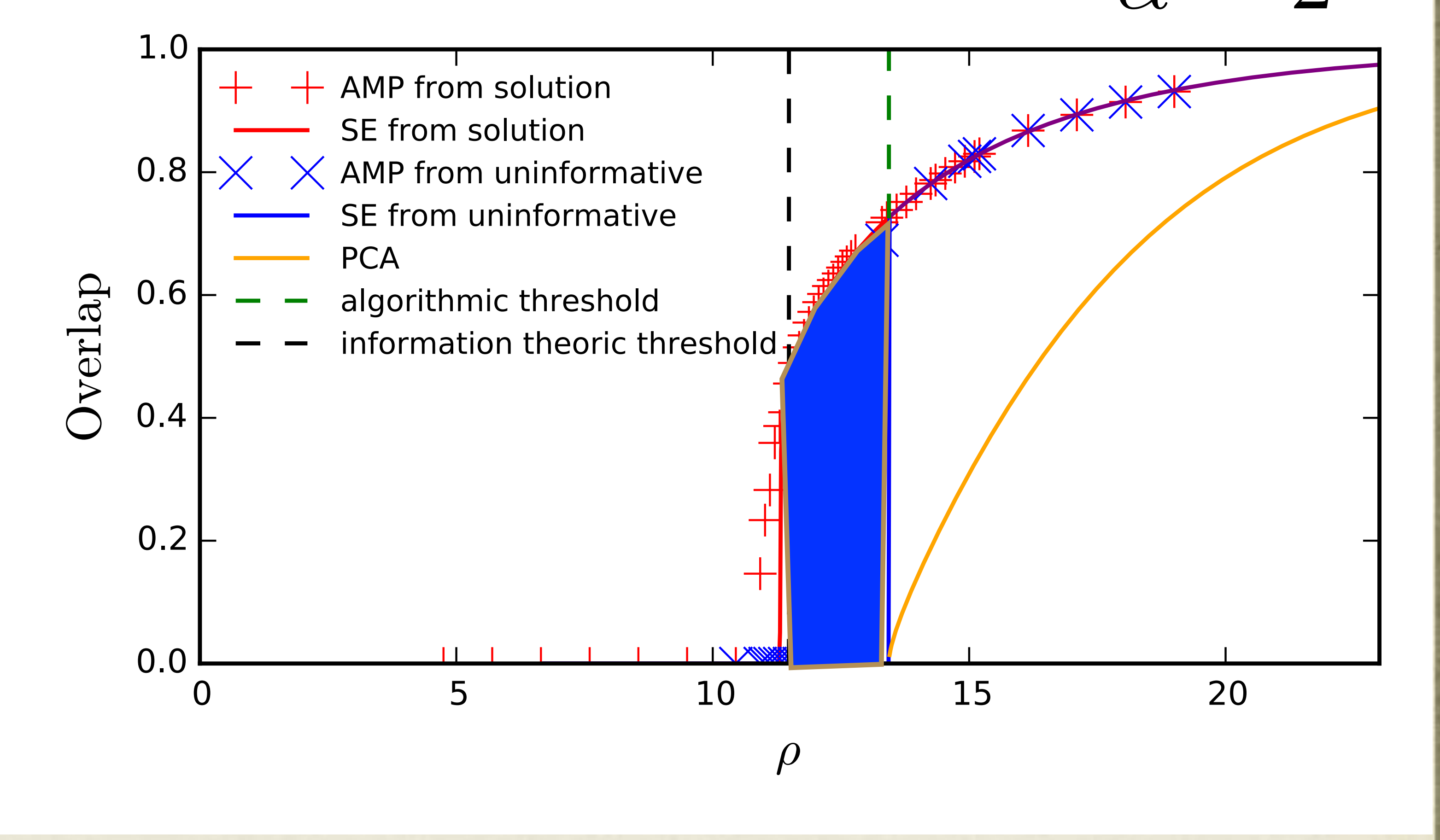
GAUSSIAN MIXTURE R=20 CLUSTERS

PCA and AMP estimate better than random guessing for: $\rho_c = \frac{r}{\sqrt{\alpha}}$

Discontinuous phase transition implies a gap between statistical and computational performance.

$\alpha = 2$

Performance of AMP is considerably better than PCA.



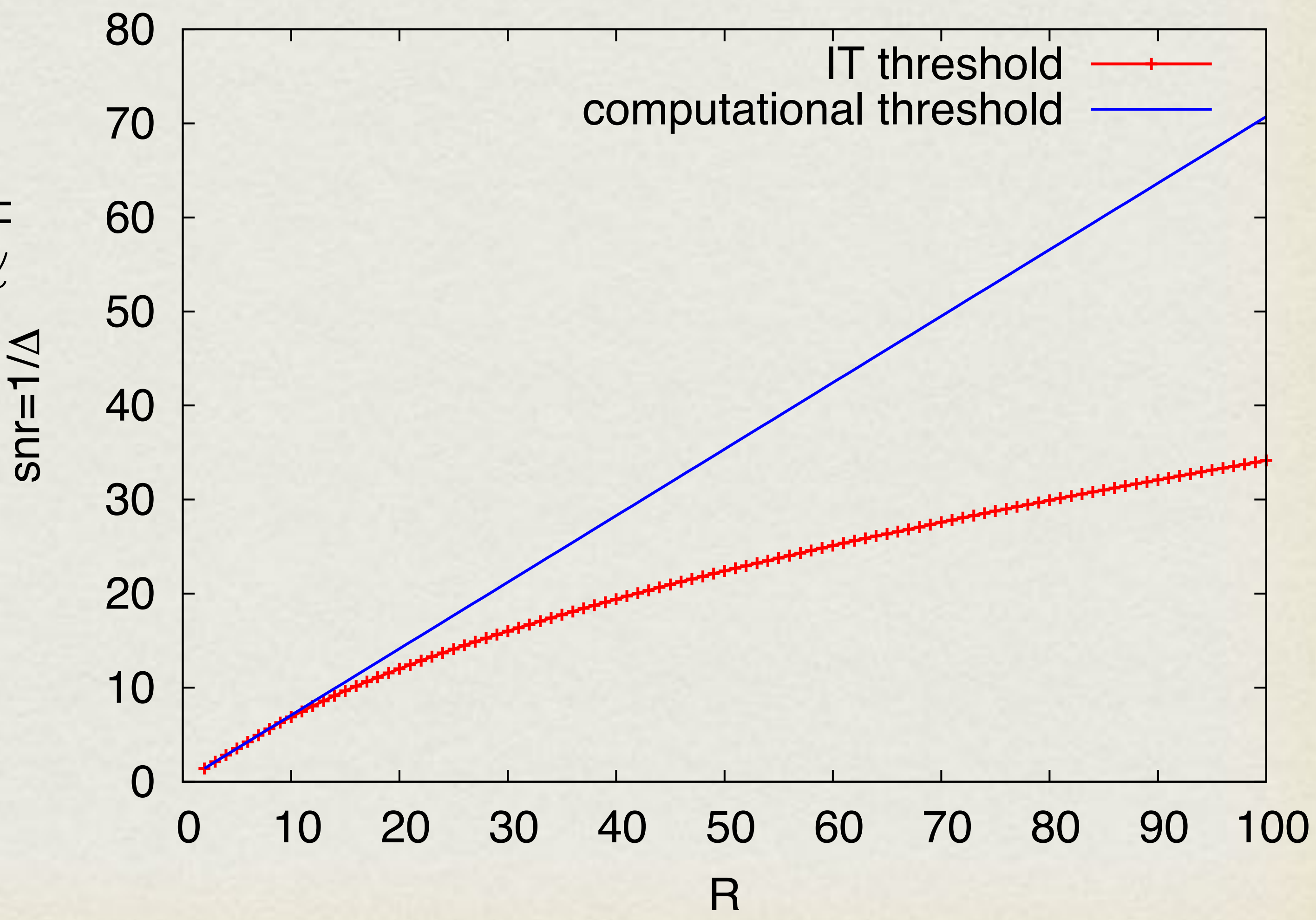
GAUSSIAN MIXTURE GENERAL # OF CLUSTERS

- Computational gap exists for: $r > 4 + 2\sqrt{\alpha}$
- Hard phase:

$$\left.\frac{2\sqrt{r\log r}}{\sqrt{\alpha}}\right|_{r\to\infty} < \rho < \frac{r}{\sqrt{\alpha}}$$

information theoretic threshold for large r.

(bound with this scaling in Banks, Moore, Xu'16)



LECTURE BASED ON

- Krzakala, LZ, *Statistical Physics of Inference: Thresholds and Algorithms*, Advances of Physics'16
- Decelle, Krzakala, Moore, LZ, *Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications,* Phys. Rev. E'11
- Krzakala, Moore, Mossel, Neeman, Sly, LZ, Zhang, *Spectral redemption in clustering sparse networks,* PNAS'13
- Saade, Lelarge, Krzakala, LZ, *Spectral detection in the censored block model,* ISIT'15
- Lesieur, Krzakala, LZ, *MMSE of probabilistic low-rank matrix estimation: Universality with respect to the output channel*, Allerton'15.
- Lesieur, Krzakala, LZ, *Constrained Low-rank Matrix Estimation: Phase Transitions, Approximate Message Passing and Applications*, J. Stat. Mech.'17
- Krzakala, Xu, LZ, *Mutual information in rank-one matrix estimation*, ITW'16
- Barbier, Dia, Macris, Krzakala, Lesieur, LZ *Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula*, NIPS'16
- Lesieur, Miolane, Lelarge, Krzakala, LZ, *Statistical and computational phase transitions in spiked tensor estimation*, ISIT'17