

Improved mean-field approximations for inferring marginals and model parameters

Federico Ricci-Tersenghi
(Sapienza University)

YITP, Kyoto, July 2015

Outline of the talk

- Two different derivations of mean-field approximations
 - Plefka expansion
 - Cluster Variational Method
- How good these approximations are?
- How one can try to improve it in order to overcome the limitations due to:
 - Loops
 - Ergodicity breaking

Physics & Machine Learning

- This is a Physics talk!
- But useful for Machine Learning (I hope ;-)
- Common problem: compute quickly and accurately the free-energy

$$F(\mathbf{J}, \mathbf{h}) = \log Z(\mathbf{J}, \mathbf{h}) = \log \sum_{\{s_i\}} \exp \left(\sum J_{ij} s_i s_j + \sum_i h_i s_i \right)$$

and the marginals

$$m_i = \langle s_i \rangle \quad C_{ij} = \langle s_i s_j \rangle - m_i m_j$$

Simplifications for this talk

- Ising variables $s_i = \pm 1$...can be extended to Potts
- Pairwise interactions

$$H(\mathbf{s}) = - \sum_{(ij)} J_{ij} s_i s_j - \sum_i h_i s_i$$

...can be extended to more general graphical models

- The measure is

$$P(\mathbf{s}) = \frac{1}{Z(\mathbf{J}, \mathbf{h})} \exp \left[\beta \sum_{(ij)} J_{ij} s_i s_j + \beta \sum_i h_i s_i \right]$$

Often $\beta = 1$

- No hidden variables

Inferring marginals is useful for machine learning

- Willing to maximize the log-likelihood

$$L(\mathbf{J}, \mathbf{h}) = \sum_i h_i \langle s_i \rangle_{\text{data}} + \sum_{ij} J_{ij} \langle s_i s_j \rangle_{\text{data}} - \log Z(\mathbf{J}, \mathbf{h})$$

with respect to $\mathbf{J}$ and $\mathbf{h}$ one gets

$$\langle s_i \rangle_{\text{data}} = \partial_{h_i} F(\mathbf{J}, \mathbf{h}) = m_i(\mathbf{J}, \mathbf{h})$$

$$\langle s_i s_j \rangle_{\text{data}} = \partial_{J_{ij}} F(\mathbf{J}, \mathbf{h}) = C_{ij}(\mathbf{J}, \mathbf{h}) + m_i m_j$$

- Most likely model parameters can be found matching empirical marginals with model marginals

Boltzmann machine learning

- Matching between empirical and model marginals is achieved via a learning

$$\delta h_i = \eta \left(\langle s_i \rangle_{\text{data}} - m_i \right)$$

$$\delta J_{ij} = \eta \left(\langle s_i s_j \rangle_{\text{data}} - (C_{ij} + m_i m_j) \right)$$

- Marginals can be computed by:
 - Monte Carlo → exact but slow...
 - Mean-field approximations
 - faster
 - no learning, if analytic expressions for $m_i(\mathbf{J}, \mathbf{h})$, $C_{ij}(\mathbf{J}, \mathbf{h})$ can be inverted

Alternatively...

- Directly maximize an easy-to-compute approximation to $L(\mathbf{J}, \mathbf{h})$ (e.g. the pseudo-likelihood)
→ see next talk by Aurelien Decelle
- If input data are not configurations but average values for the marginals $\langle s_i \rangle_{\text{data}}, \langle s_i s_j \rangle_{\text{data}}$ then:
 - No pseudo-likelihood maximization
 - Matching empirical and model marginals
 - The maximum entropy probability distribution has only fields and pairwise couplings

Free-energy mean-field expansions

- How to generalize most common mean-field approx?
- Pleška derives an expansion in the couplings intensity for the Gibbs free-energy with given magnetizations
- Cluster Variational Method (Kikuchi, Morita, An) is an expansion of the entropy in terms of correlations up to a given distance (maximal region size)
- Both expansions have the naive mean-field approximation as the first-order approximation
- Beyond naive MF and Bethe approximations the relation among the 2 expansions is not trivial (Yasuda Tanaka)

Plefka's expansion

$\mathcal{F}_{\text{Plefka}}[\mathbf{m}_\Lambda]$

naiveMF

TAP

$$\begin{aligned}
 = & \boxed{-\sum_{i \in \Lambda} h_i m_i + \frac{1}{\beta} \sum_{i \in \Lambda} \left(\frac{1+m_i}{2} \ln \frac{1+m_i}{2} + \frac{1-m_i}{2} \ln \frac{1-m_i}{2} \right) - \sum_{(ij) \in \mathcal{B}} J_{ij} m_i m_j} - \frac{\beta}{2} \sum_{(ij) \in \mathcal{B}} J_{ij}^2 (1-m_i^2)(1-m_j^2) \\
 & - \frac{2\beta^2}{3} \sum_{(ij) \in \mathcal{B}} J_{ij}^3 m_i m_j (1-m_i^2)(1-m_j^2) - \beta^2 \sum_{(ijk) \in \mathcal{T}} J_{ij} J_{jk} J_{ki} (1-m_i^2)(1-m_j^2)(1-m_k^2) \\
 & - \frac{\beta^3}{12} \sum_{(ij) \in \mathcal{B}} J_{ij}^4 (1-m_i^2)(1-m_j^2)(15m_i^2 m_j^2 - 3m_i^2 - 3m_j^2 - 1) \\
 & - 2\beta^3 \sum_{(ijk) \in \mathcal{T}} J_{ij} J_{jk} J_{ki} (1-m_i^2)(1-m_j^2)(1-m_k^2)(J_{ij} m_i m_j + J_{jk} m_j m_k + J_{ki} m_k m_i) \\
 & - \beta^3 \sum_{(ijkl) \in \mathcal{Q}} J_{ij} J_{jk} J_{kl} J_{li} (1-m_i^2)(1-m_j^2)(1-m_k^2)(1-m_l^2) + \mathcal{O}(J_{ij}^5).
 \end{aligned}$$

diagrammatical
representation

$$\begin{aligned}
 F = & \bullet + \bullet\text{---}\bullet + \bullet\text{---}\bullet + \bullet\text{---}\bullet + \bullet\text{---}\bullet + \bullet\text{---}\bullet + \dots \\
 & + \triangle + \triangle + \dots \\
 & + \square + \dots
 \end{aligned}$$

NaiveMF and TAP approximations

$$F_{\text{nMF}} = \sum_i \left[H \left(\frac{1+m_i}{2} \right) + H \left(\frac{1-m_i}{2} \right) \right] + \sum_i h_i m_i + \sum_{i \neq j} J_{ij} m_i m_j \quad H(x) \equiv -x \ln(x)$$

$$\frac{\partial F_{\text{nMF}}}{\partial m_i} = \sum_j J_{ij} m_j + h_i - \text{atanh}(m_i) = 0 \quad \Rightarrow \quad m_i = \tanh \left[h_i + \sum_j J_{ij} m_j \right]$$

$$F_{\text{TAP}} = \sum_i \left[H \left(\frac{1+m_i}{2} \right) + H \left(\frac{1-m_i}{2} \right) \right] + \sum_i h_i m_i + \sum_{i \neq j} \left(J_{ij} m_i m_j + \frac{1}{2} J_{ij}^2 (1-m_i^2)(1-m_j^2) \right)$$

$$m_i = \tanh \left[h_i + \sum_j J_{ij} \left(m_j - \frac{J_{ij}(1-m_j^2)m_i}{J_{ij}} \right) \right]$$

↓
Onsager reaction term

Bethe approximation

- On a tree

$$F = \boxed{\bullet + \bullet - \bullet + \bullet \text{---} \bullet + \bullet \text{---} \bullet + \bullet \text{---} \bullet + \dots} \quad \text{sum to Bethe approx}$$

~~$+ \text{triangle} + \text{triangle} + \dots$
 $+ \text{square} + \dots$~~

- Not easy to write the explicit expression for the Bethe approximation
- Very hard to go beyond the Bethe approximation...

Cluster Variational Method (CVM)

$$F = -\ln Z = \min_p \mathcal{F}(p) = \min_p \sum_{\mathbf{s}} [p(\mathbf{s})H(\mathbf{s}) + p(\mathbf{s}) \ln p(\mathbf{s})] \quad \sum_{\mathbf{s}} p(\mathbf{s}) = 1$$

Pelizzola
review

energy (easy)

entropy (hard) → approximate by
truncating the expansion in cumulants

$$F_{\text{CVM}}(b, J, h) = E(b, J, h) - \frac{1}{\beta} S(b)$$

$$E(b, J, h) = - \sum_{(ij)} J_{ij} \text{Tr}[s_i s_j b_{ij}] - \sum_i h_i \text{Tr}[s_i b_i]$$

local

$$S(b) = - \sum_{r \in R} c_r \text{Tr}[b_r \log b_r]$$

$$b_r = b_r(s_r) \quad \text{local beliefs}$$

$$s_r = \{r_i : i \in r\}$$

sum over regions

counting numbers (Möbius coeff.)

Cluster Variational Method (CVM)

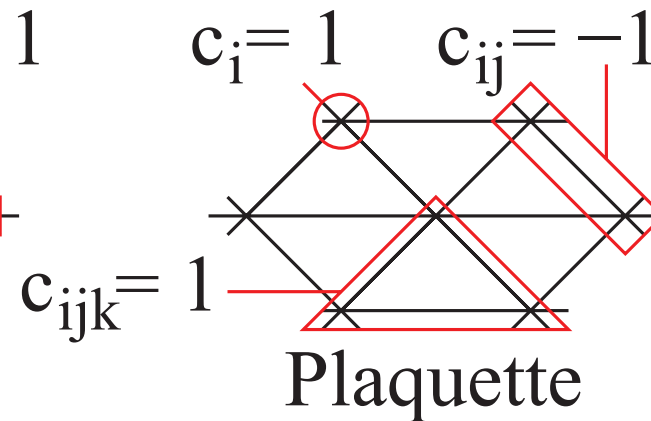
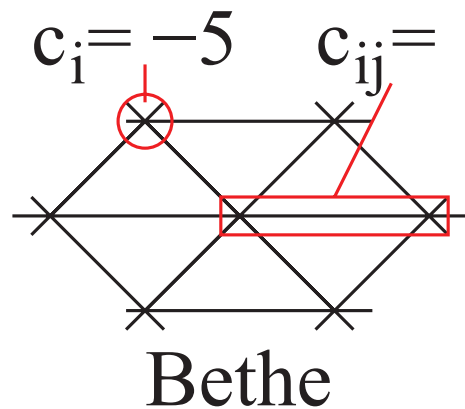
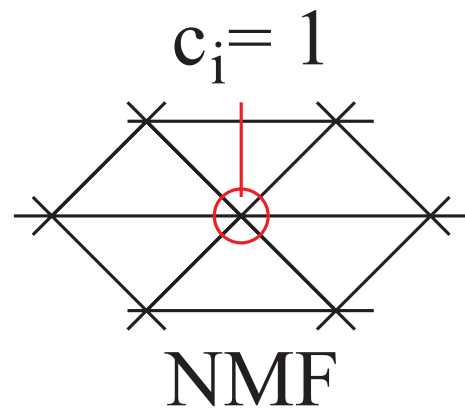
- Beliefs must be normalized and locally consistent

$$\sum_{s_r} b_r(s_r) = 1 \qquad \sum_{s_r \setminus t} b_r(s_r) = b_t(s_t)$$

- Local consistency is not global consistency!
- Beliefs are approximations to true marginals
- Beliefs can be parametrized by magnetizations and connected correlations, e.g.

$$b_i(s_i) = \frac{1 + m_i s_i}{2} \qquad b_{ij}(s_i, s_j) = \frac{1 + m_i s_i + m_j s_j + (c_{ij} + m_i m_j) s_i s_j}{4}$$

Cluster Variational Method (CVM)



$$P(\mathbf{s}) \approx \prod_i b_i(s_i)$$

$$P(\mathbf{s}) \approx \prod_{(ij)} b_{ij}(s_i, s_j) \prod_i b_i(s_i)^{-5} = \prod_{(ij)} \frac{b_{ij}(s_i, s_j)}{b_i(s_i)b_j(s_j)} \prod_i b_i(s_i)$$

Exact on trees

$$\sum_{r \in R: t \subset r} c_r = 1 \quad \forall t$$

How to choose the regions

- Original CVM: all maximal regions and all their intersections (recursively until single site regions)
- Region-based free energy approximation (Yedidia et al.): choose regions at your will as long as each site (variable node) and interaction (factor node) has coefficient $c=1$
- More is better, but computationally ineffective
- Try to include all relevant correlations $\text{diam}(r_{\max}) \approx \xi$
- Easy to derive Bethe and clear how to go beyond Bethe

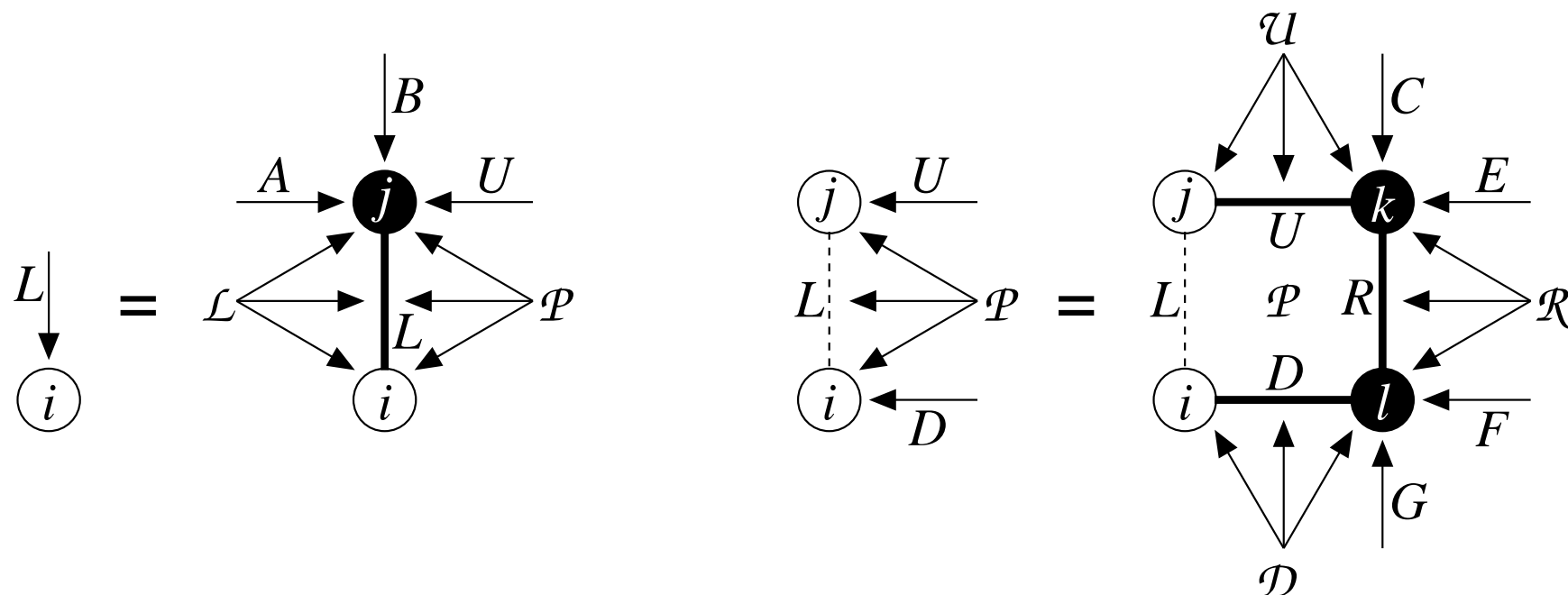
How to find the beliefs

- Introduce Lagrange multipliers (called messages) enforcing the consistency constraint for each pair of regions

$$m_{r \rightarrow t}(s_t) \quad \forall r, t \in R : t \subset r$$

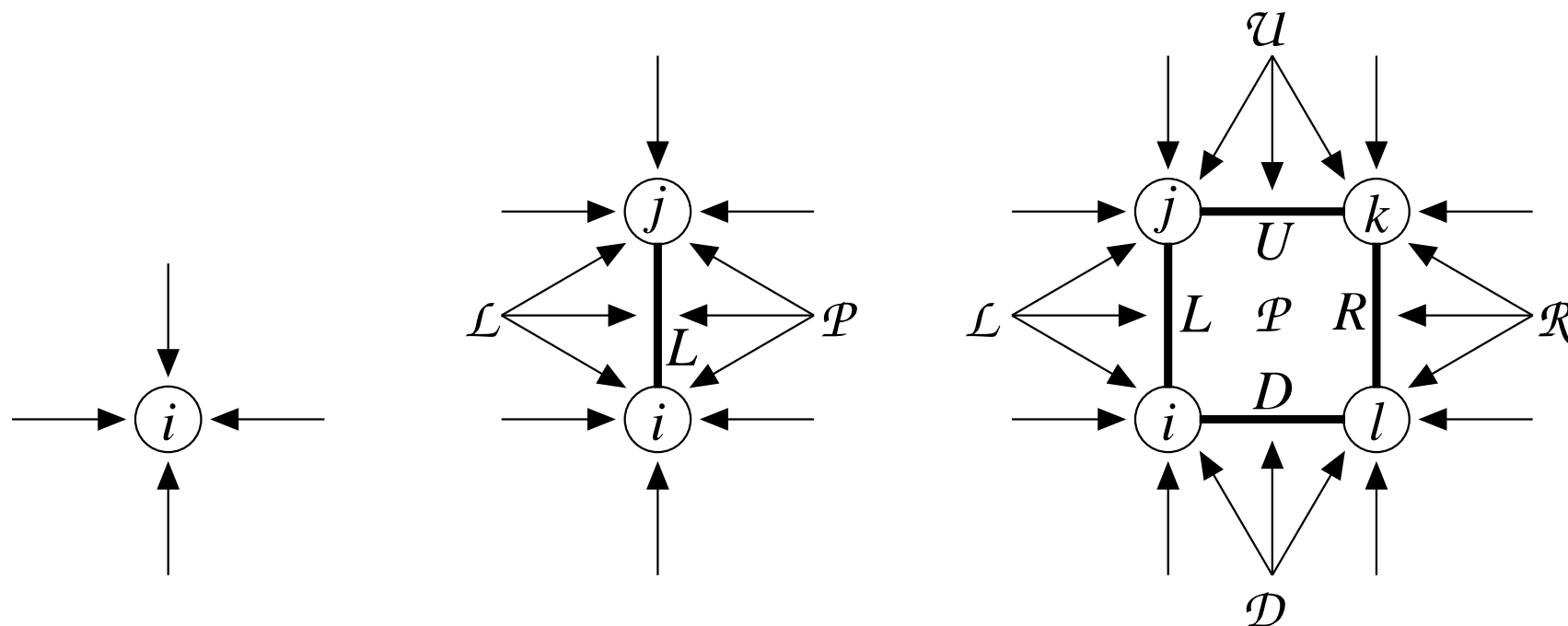
These are generalized cavity probability distributions

- Solve equations for messages iteratively $\rightarrow$ Belief Propagation (BP), Generalized Belief Propagation (GBP)



How to find the beliefs

- After converging to the fixed point of the message passing algorithm, compute beliefs from the messages



- N.B. several equivalent MPA and each may have several equivalent fixed point (gauge invariance).

Plefka's expansion vs. CVM

- Plefka's expansion has N parameters (the magnetizations) nMF , TAP, 3rd order, 4th order,...,Bethe.
- CVM may have much more parameters to optimize over
E.g. on the 2D square lattice:
 - $nMF \rightarrow N$ magnetizations
 - Bethe $\rightarrow N$ magnetizations + $2N$ nn correlations
 - Plaquette $\rightarrow N$ magnetizations + $2N$ nn correlations + $2N$ nnn correlations + $4N$ 3-spin corr. + N 4-spin corr.
- CVM much richer description, but hard to get analytical expressions to estimate model parameters

Bethe approximation

Independent
pair approx.



- The two derivations are equivalent: correlations only depends on magnetizations at the fixed point

$$\frac{\partial F_{\text{Bethe}}}{\partial C_{ij}} = 0 \implies J_{ij} = \frac{1}{4} \ln \left(\frac{((1+m_i)(1+m_j) + c_{ij})((1-m_i)(1-m_j) + c_{ij})}{((1+m_i)(1-m_j) - c_{ij})((1-m_i)(1+m_j) - c_{ij})} \right)$$

$$c_{ij}(m_i, m_j, t_{ij}) = \frac{1}{2t_{ij}} \left(1 + t_{ij}^2 - \sqrt{(1 - t_{ij}^2)^2 - 4t_{ij}(m_i - t_{ij}m_j)(m_j - t_{ij}m_i)} \right) - m_i m_j$$

$$f(m_1, m_2, t) = \frac{1 - t^2 - \sqrt{(1 - t^2)^2 - 4t(m_1 - m_2t)(m_2 - m_1t)}}{2t(m_2 - m_1t)}$$

$$m_i = \tanh \left[h_i + \sum_j \text{atanh} \left(t_{ij} f(m_j, m_i, t_{ij}) \right) \right]$$

Small couplings expansion leads to nMF, TAP, ...

Computing correlations by linear response

- Connected correlations are always null in nMF, TAP
Even in Bethe between non-neighbours spins
- Non trivial (and better) correlations can be obtained via linear response (Kappen Rodriguez, 1998)

$$\chi_{ij} = \frac{\partial m_i}{\partial h_j}$$

$$(\chi_{\text{nMF}}^{-1})_{ij} = \frac{\delta_{ij}}{1 - m_i^2} - J_{ij} ,$$

$$(\chi_{\text{TAP}}^{-1})_{ij} = \left[\frac{1}{1 - m_i^2} + \sum_k J_{ik}^2 (1 - m_k^2) \right] \delta_{ij} - (J_{ij} + 2J_{ij}^2 m_i m_j)$$

$$(\chi_{\text{BA}}^{-1})_{ij} = \left[\frac{1}{1 - m_i^2} - \sum_k \frac{t_{ik} f_2(m_k, m_i, t_{ik})}{1 - t_{ik}^2 f(m_k, m_i, t_{ik})^2} \right] \delta_{ij} - \frac{t_{ij} f_1(m_j, m_i, t_{ij})}{1 - t_{ij}^2 f(m_j, m_i, t_{ij})^2}$$

$$f_1(m_1, m_2, t) \equiv \partial f(m_1, m_2, t) / \partial m_1 \quad f_2(m_1, m_2, t) \equiv \partial f(m_1, m_2, t) / \partial m_2$$

Estimating model parameters via MFA

- Assume $\chi = C$!
- Estimate couplings from matching only off-diagonal elements of χ^{-1} and C_{data}^{-1}

$$J_{ij}^{\text{nMF}} = -(C^{-1})_{ij} \qquad J_{ij}^{\text{TAP}} = \frac{\sqrt{1 - 8m_i m_j (C^{-1})_{ij}} - 1}{4m_i m_j}$$

$$J_{ij}^{\text{BA}} = -\text{atanh} \left[\frac{1}{2(C^{-1})_{ij}} \sqrt{1 + 4(1 - m_i^2)(1 - m_j^2)(C^{-1})_{ij}^2 - m_i m_j} \right. \\ \left. - \frac{1}{2(C^{-1})_{ij}} \sqrt{\left(\sqrt{1 + 4(1 - m_i^2)(1 - m_j^2)(C^{-1})_{ij}^2} - 2m_i m_j (C^{-1})_{ij} \right)^2 - 4(C^{-1})_{ij}^2} \right]$$

- Easier than computing marginals: no need to run MPA!

Estimating model parameters via MFA

- Independent Pair (IP) approximation

$$J_{ij}^{\text{IP}} = \frac{1}{4} \ln \left(\frac{\left((1 + m_i)(1 + m_j) + C_{ij} \right) \left((1 - m_i)(1 - m_j) + C_{ij} \right)}{\left((1 + m_i)(1 - m_j) - C_{ij} \right) \left((1 - m_i)(1 + m_j) - C_{ij} \right)} \right)$$

- Sessak-Monasson (SK) small correlation expansion

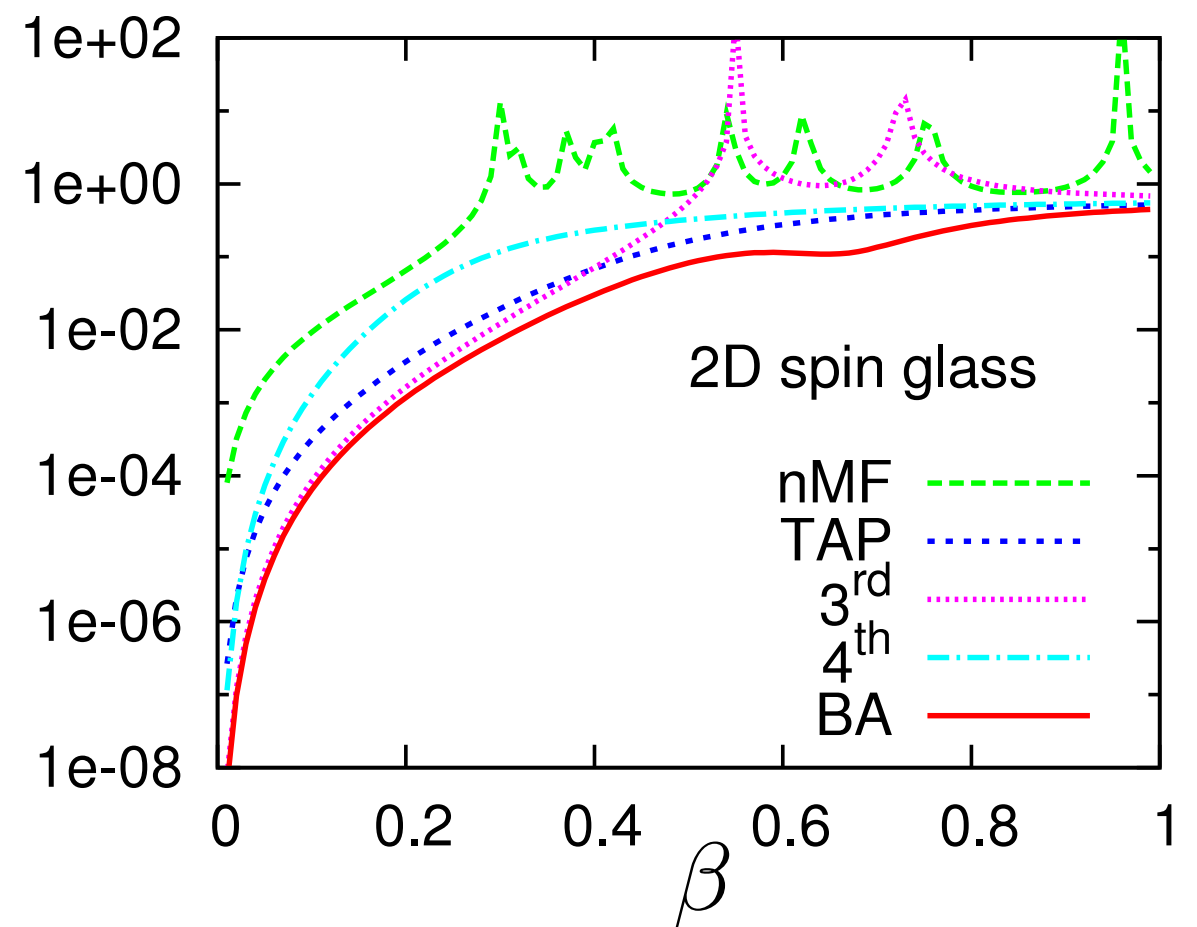
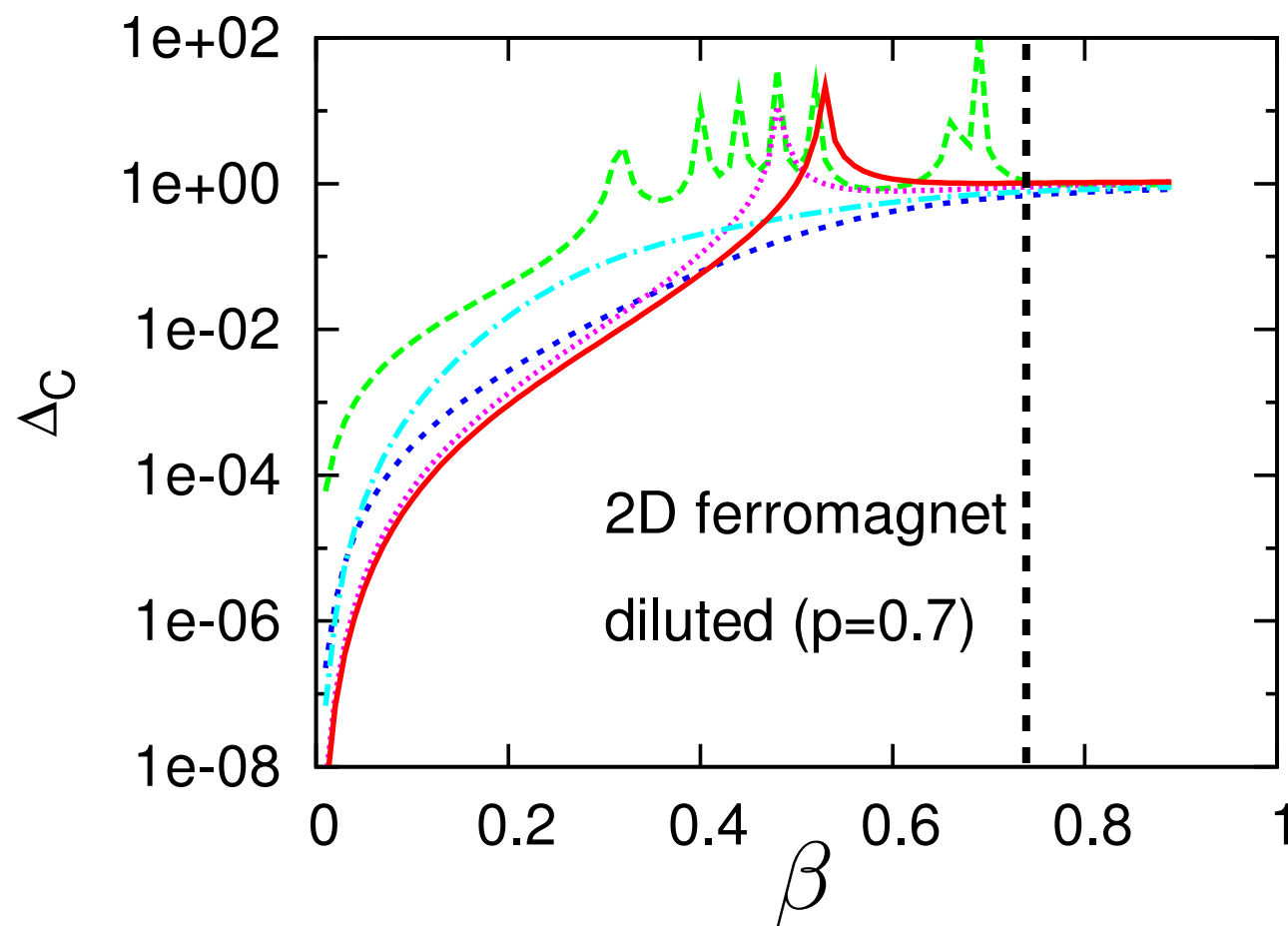
$$J_{ij}^{\text{SM}} = -(C^{-1})_{ij} + J_{ij}^{\text{IP}} - \frac{C_{ij}}{(1 - m_i^2)(1 - m_j^2) - (C_{ij})^2}$$

- Fields estimates from self-consistency equation
E.g. for nMF

$$m_i = \tanh[\beta(h_i + \sum_j J_{ij} m_j)] \implies h_i = \frac{\text{atanh}(m_i)}{\beta} - \sum_j J_{ij}^{\text{nMF}} m_j$$

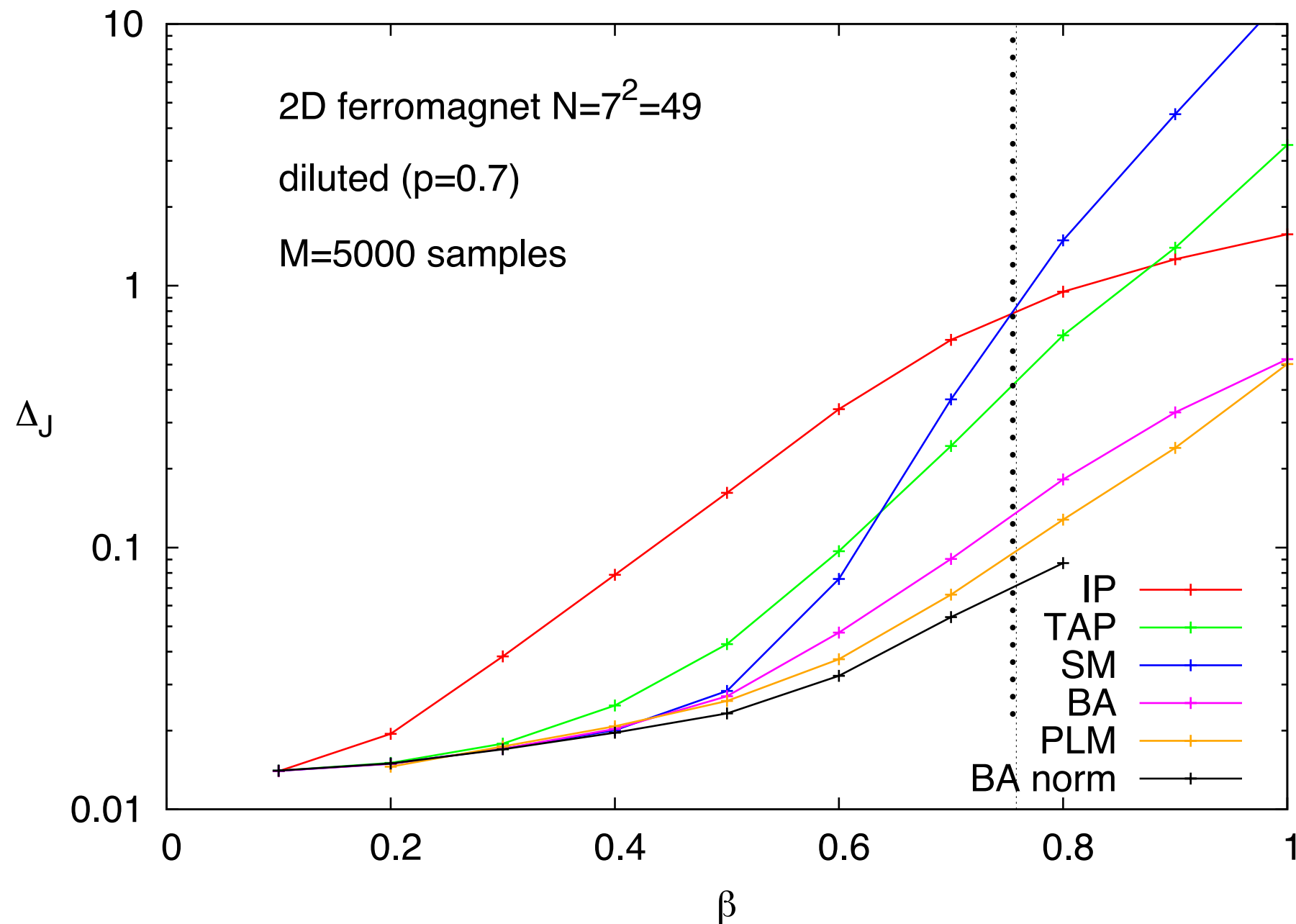
How good are MFA?

- Estimating marginals $\Delta_C = \sqrt{\frac{1}{N^2} \sum_{i,j} (\chi_{ij} - C_{ij}^{\text{true}})^2}$



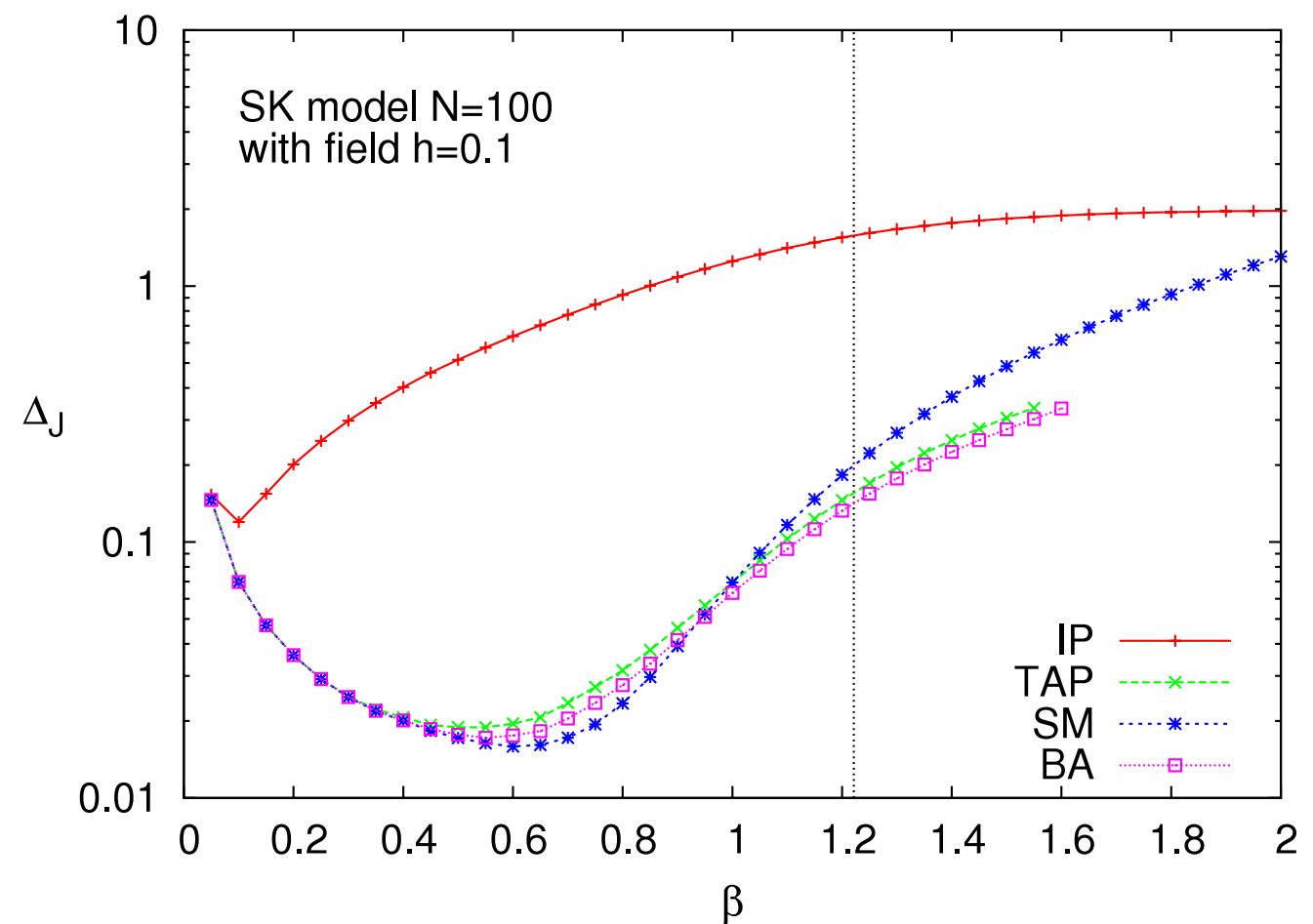
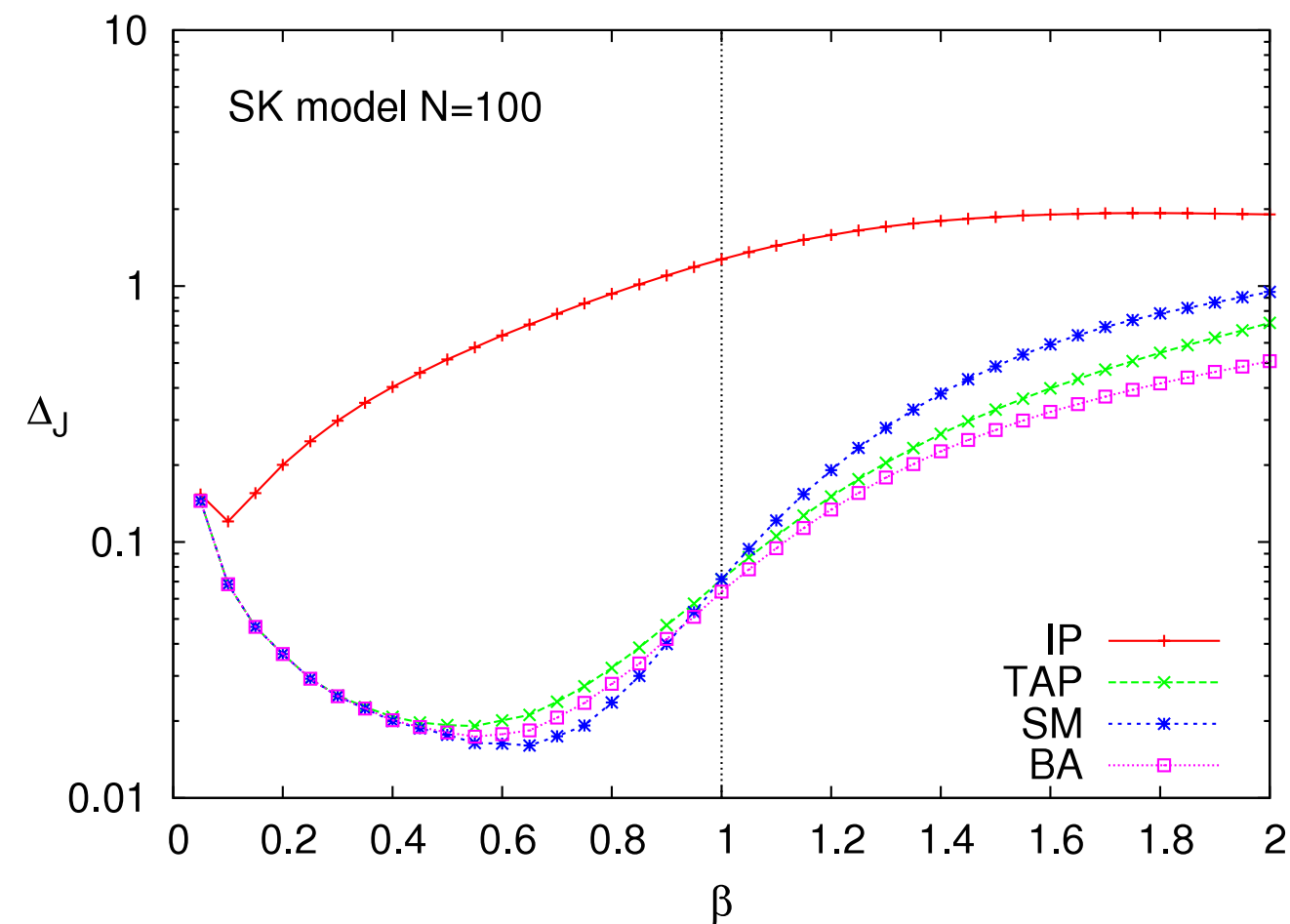
How good are MFA?

- Estimating model parameters



How good are MFA?

- Problems with strongly frustrated models in a field



- Due to negative discriminants in coupling estimates

How good are MFA?

$$\begin{pmatrix} 1 & & C_{ij} \\ & \ddots & \\ C_{ij} & & 1 \end{pmatrix} \stackrel{?}{=} \begin{pmatrix} \chi_{11} & & \chi_{ij} \\ & \ddots & \\ \chi_{ij} & & \chi_{NN} \end{pmatrix}$$

- The linear response correlation matrix (the only we can compute in MFA) has “wrong” element on the diagonal.
- It is different from the true correlation matrix.
- In the Bethe approximation the 2 estimates of nearest neighbor correlations (χ_{ij} and C_{ij}) are different. And linear response estimate is generally better.
- This is due to loops ignored in the MFA

Limitations of MFA

- Ergodicity breaking
 - MFA assume that a single state exists, and that correlations decay fast with distance
 - If many states exist correlations no longer decay, and MF estimates become poor
- Presence of loops
 - Even in presence of a single state, the loops may change a lot the correlations with respect to Susceptibility Propagation estimate, obtained assuming a loopless graph

MFA fail because of loops

- E.g. Bethe approximation in the high temperature phase
- Since $m=0$, at the CVM free-energy minimum

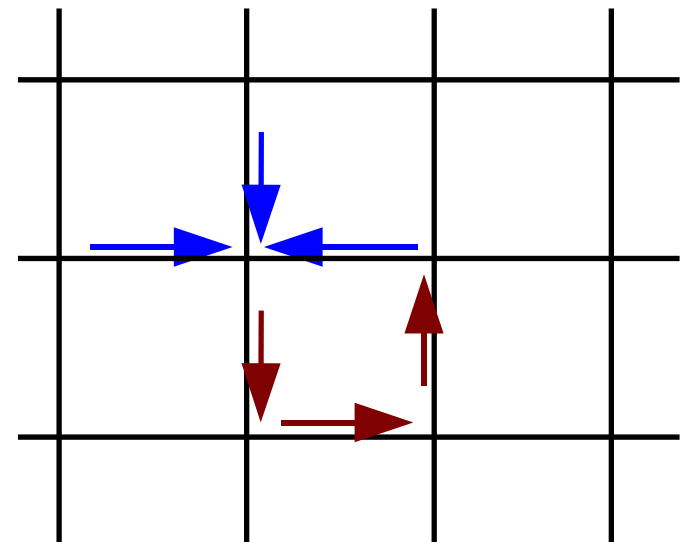
$$\langle \sigma_i \sigma_j \rangle_c^{\text{BA}} = c_{ij}^* = \tanh(\beta J_{ij}) < \langle \sigma_i \sigma_j \rangle_c^{\text{exact}}$$

- Linear response (Susceptibility Propagation)

$$\chi_{ii} = 1 + \sum_{j \in \partial i} u_{j \rightarrow i, i} \neq 1$$

- So in general for a ferromagnet in the high T phase holds

$$C_{ij} < C_{ij}^{\text{true}} < \chi_{ij}$$



Adding loops to Bethe ?

- Several attempts
 - Loop calculus (Chertkov Chernyak)
 - BP + correlations between neighbors (Montanari Rizzo, Mooij Kappen, Rizzo Wammenhove Kappen, Ohzeki)
- All require in some sense the convergence of BP, but loops make BP stop converging...
- None is able to make predictions in a frustrated model with many loops at low enough temperatures

Make MFA & LR consistent

- Choose your preferred MFA free-energy

$$F_{\text{MFA}}(\{m_i\}, \{C_{ij}\}, \dots)$$

- Enforce consistency with linear response estimates

$$\chi_{ii} = 1 - m_i^2 \quad \chi_{ij} = C_{ij}$$

free energy
minimum
curvature

free energy
minimum
location

General framework for MFA + LR

$$F_{\lambda} = F_{\text{MFA}}(\{m_i\}, \{C_{ij}\}, \dots) + \sum_i \lambda_i m_i^2 + \sum_{i < j} \lambda_{ij} C_{ij}$$

Your preferred MFA

can be set to zero to
recover known approx.
or used to satisfy

$$\chi_{ii} = 1 - m_i^2 \quad \chi_{ij} = C_{ij}$$

Other proposals for fixing χ_{ii}

- Kappen Rodriguez (1998)

MF + self-couplings J_{ii}

- Oppen Winter (2001)

Adaptive TAP = TAP + λ_i

- FRT (2012)

Bethe with normalized correlations
useful for the inverse pb.

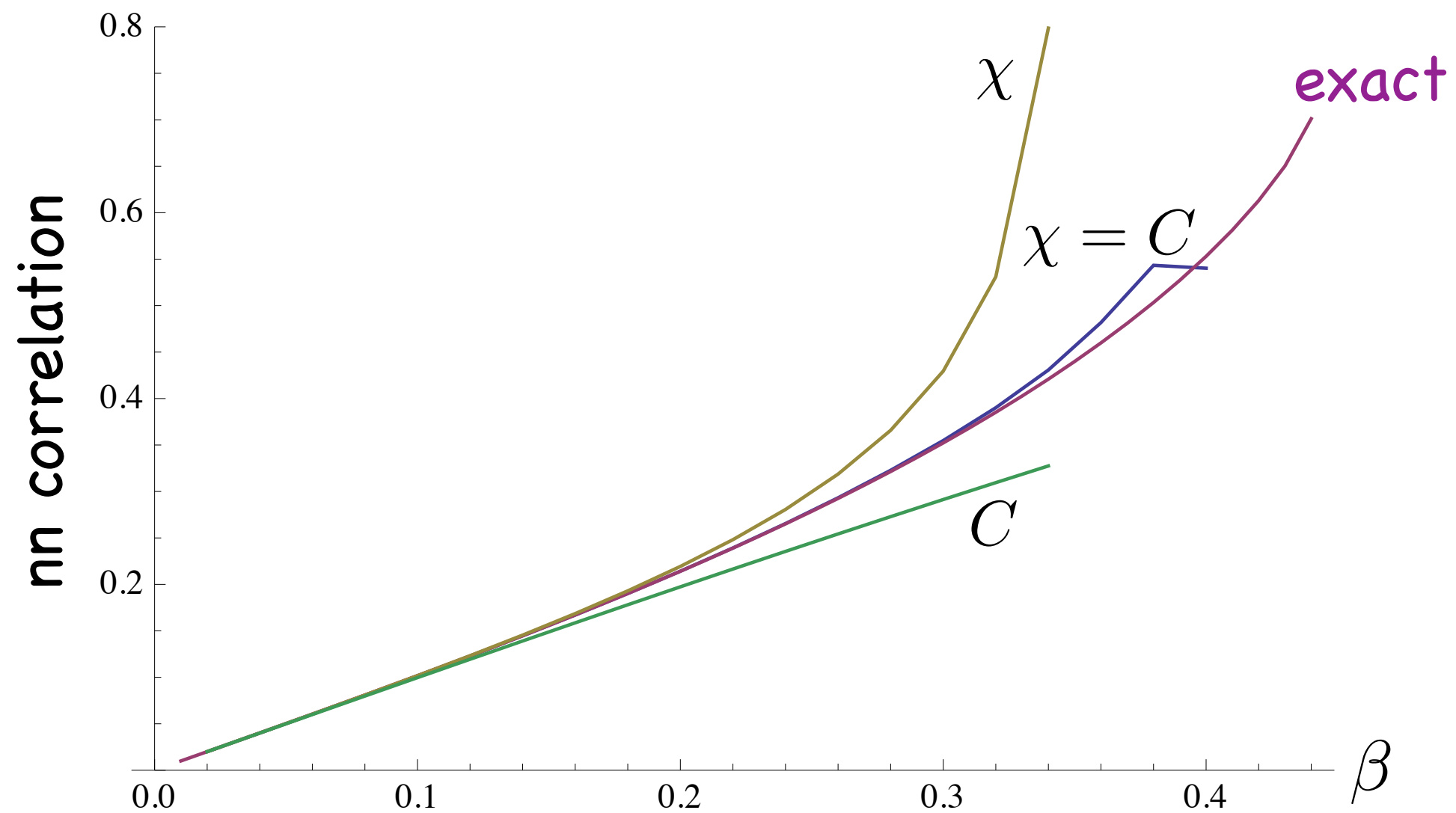
$$\hat{\chi}_{ij} \equiv \frac{\chi_{ij}}{\sqrt{\chi_{ii}\chi_{jj}}}$$

- Yasuda Tanaka (2013)

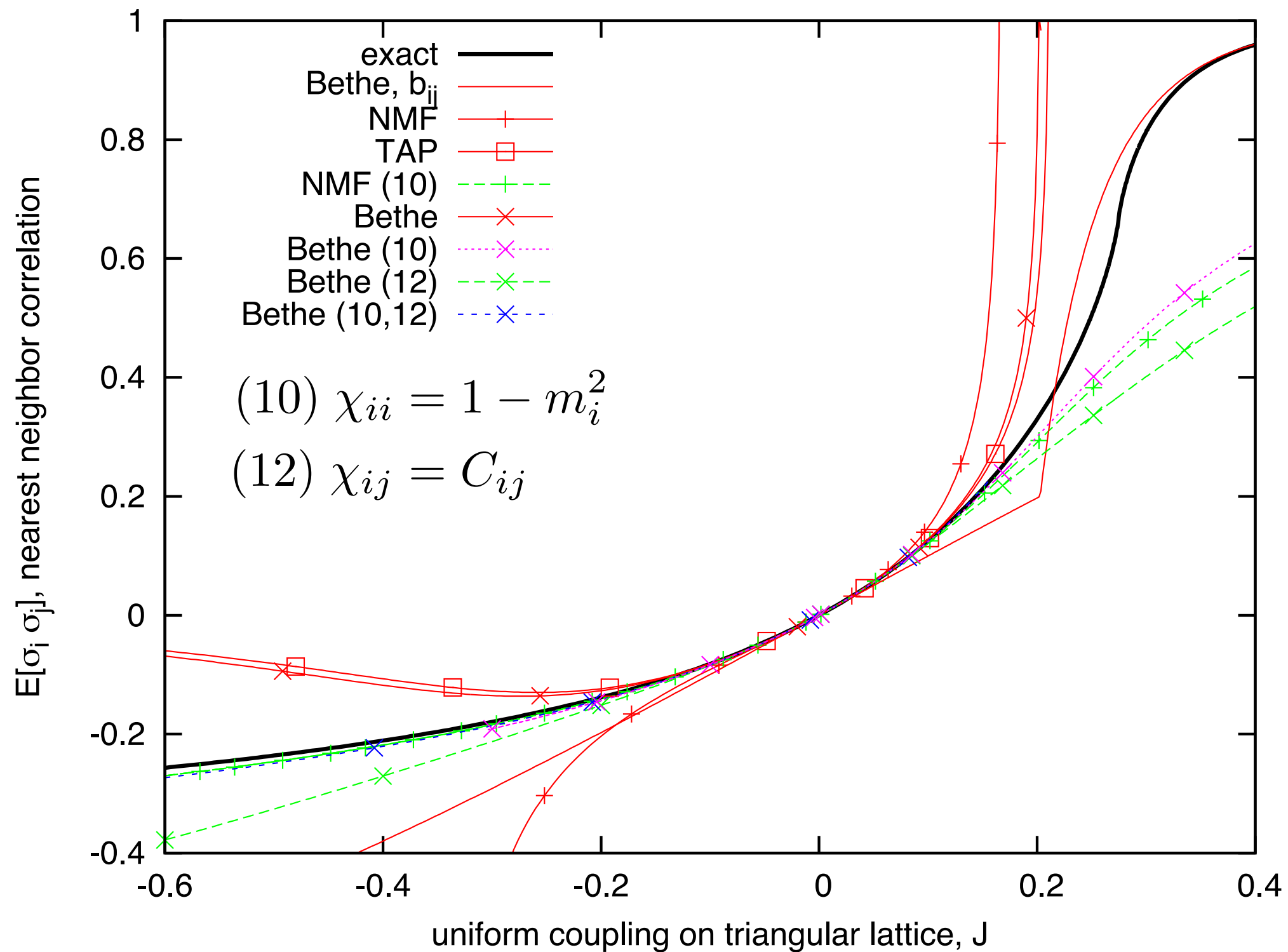
I-SuscP = Bethe + λ_i

Bethe + linear response

- Ising model on a 2D square lattice

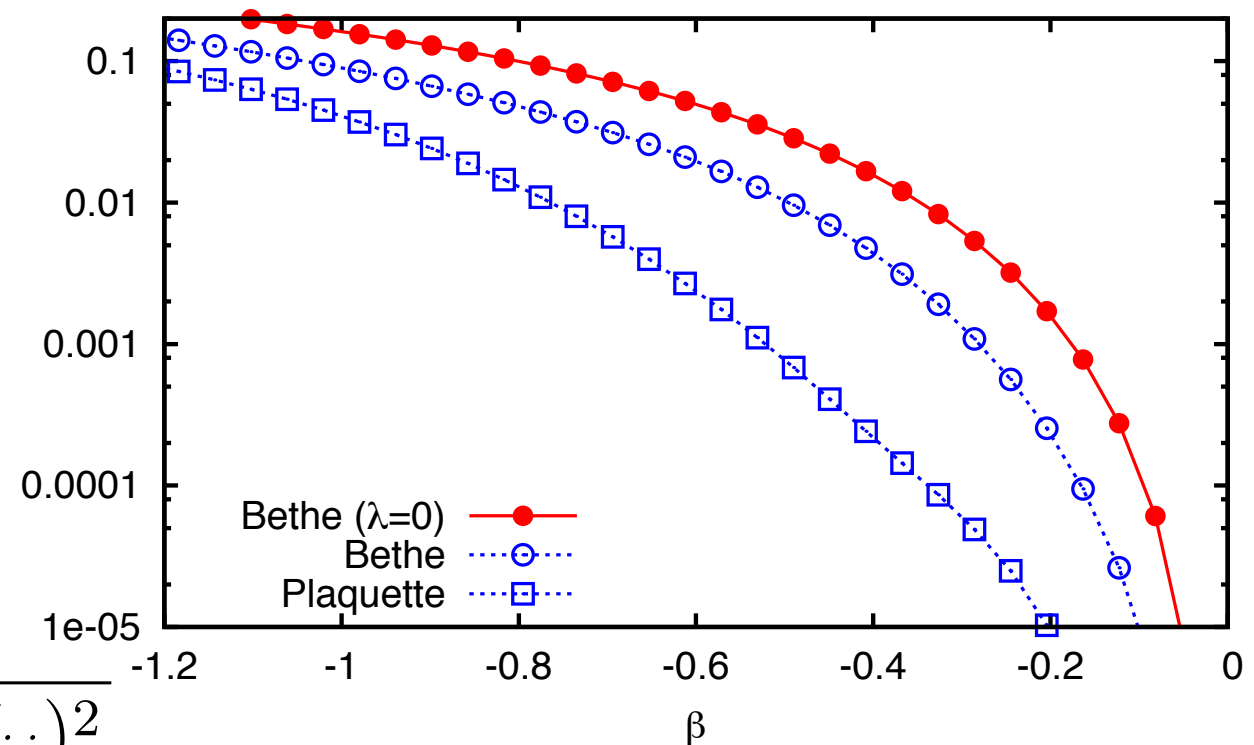


MFA + LR: estimating marginals



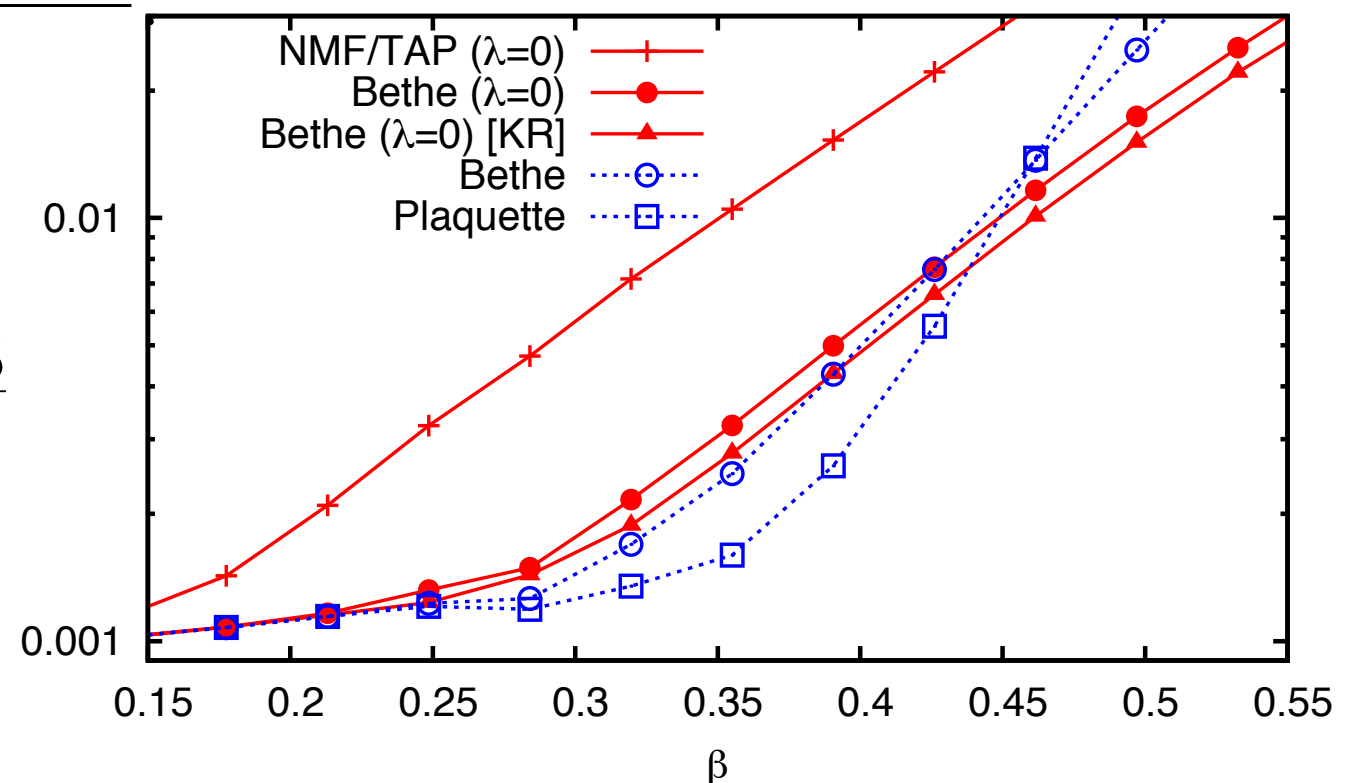
MFA + LR: estimating model param.

- Inferring the couplings of a 2D triangular diluted antiferromagnet from correlations (infinite statistics)



$$\Delta_J = \sqrt{\frac{\sum_{i<j} (J'_{ij} - J_{ij})^2}{\sum_{i<j} J_{ij}^2}}$$

- Inferring the couplings of a 2D diluted Ising model (finite stat.)

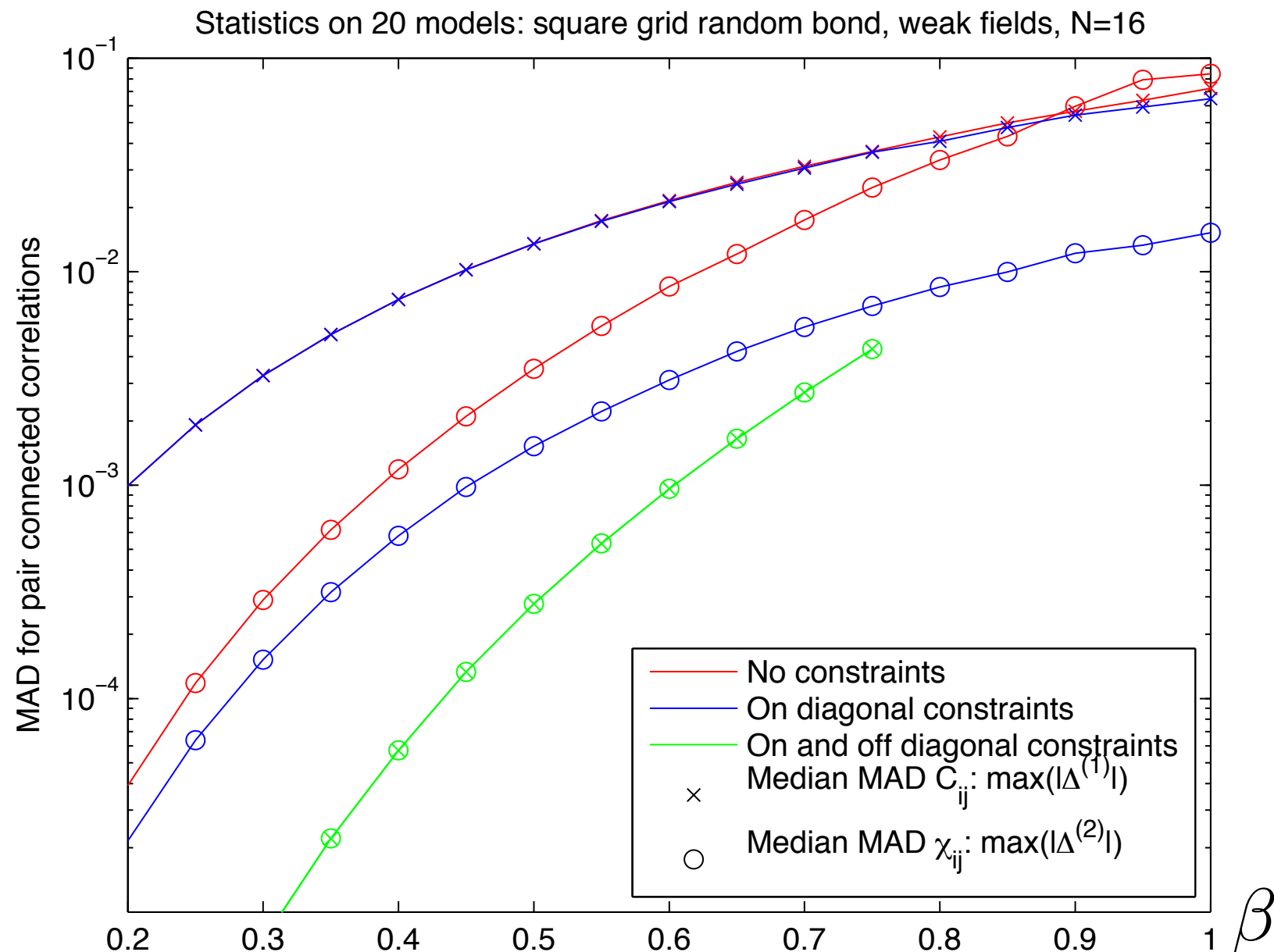
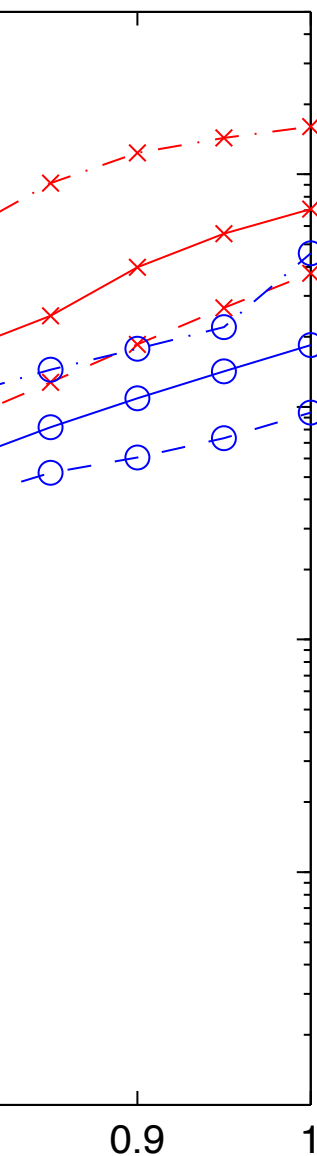


MFA + LR: disordered models

- 2D spin glass in a random field

$$J_{ij} \in [-1, 1], \quad h_i \in [-0.25, 0.25]$$

x fields, N=16

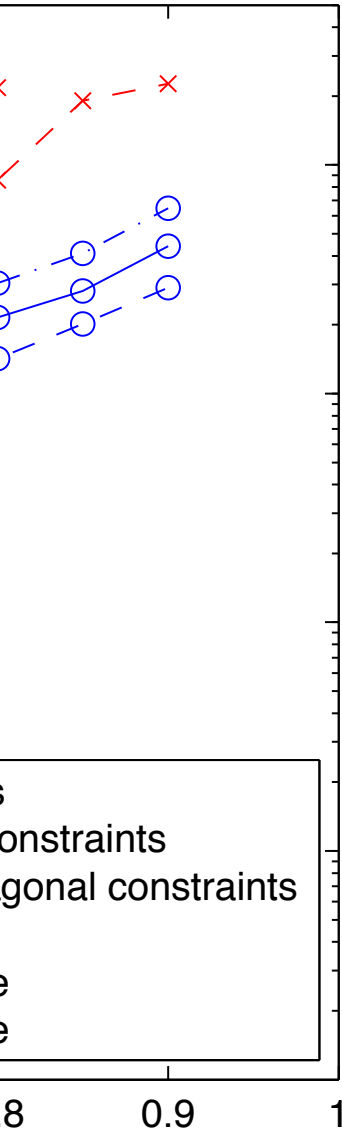


MFA + LR: disordered models

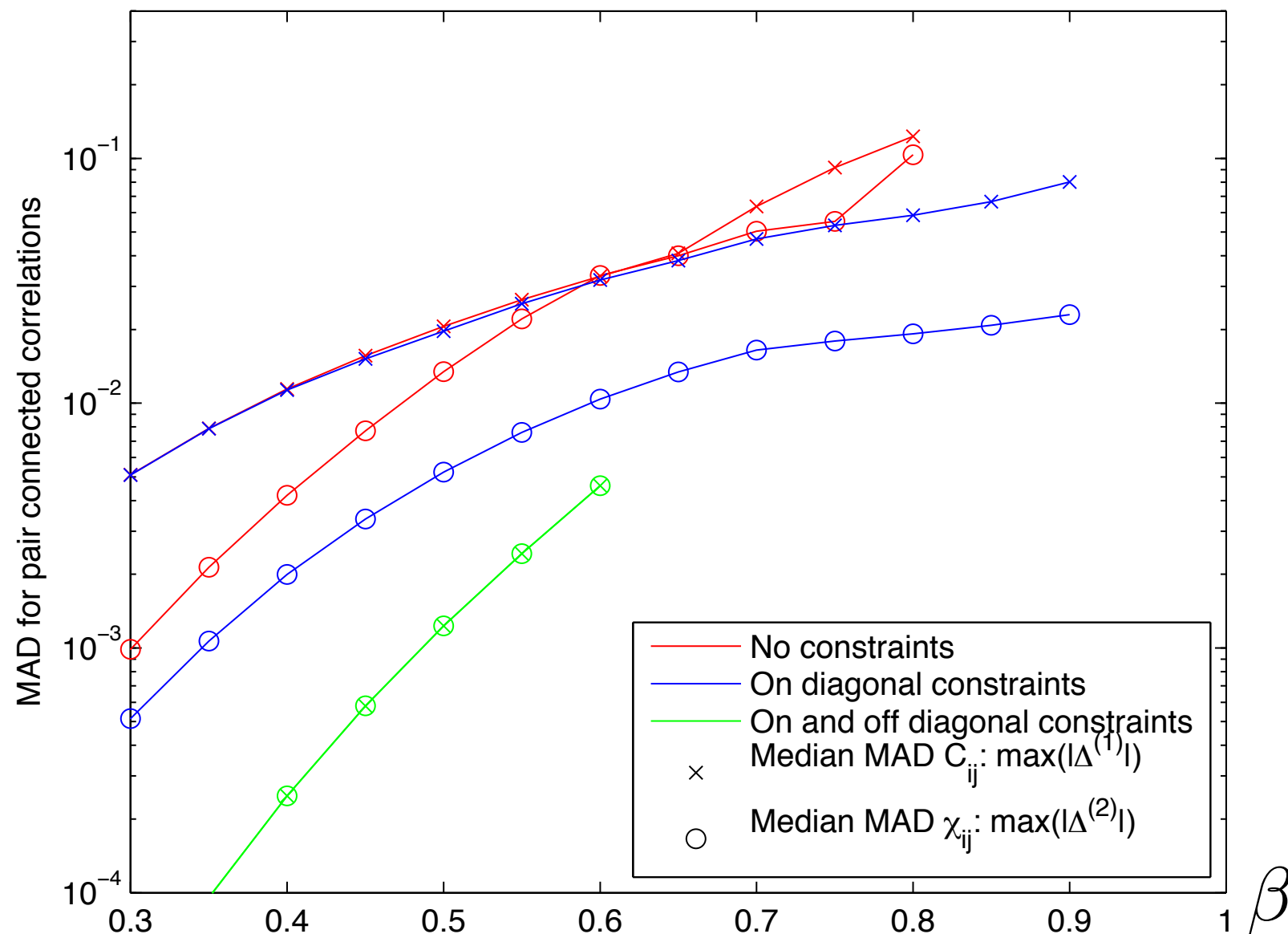
- 2D spin glass in a random field

$$J_{ij} \in [-1, 1], \quad h_i \in [-0.25, 0.25]$$

Weak fields, N=49



Statistics on 20 models: square grid random bond, weak fields, N=49

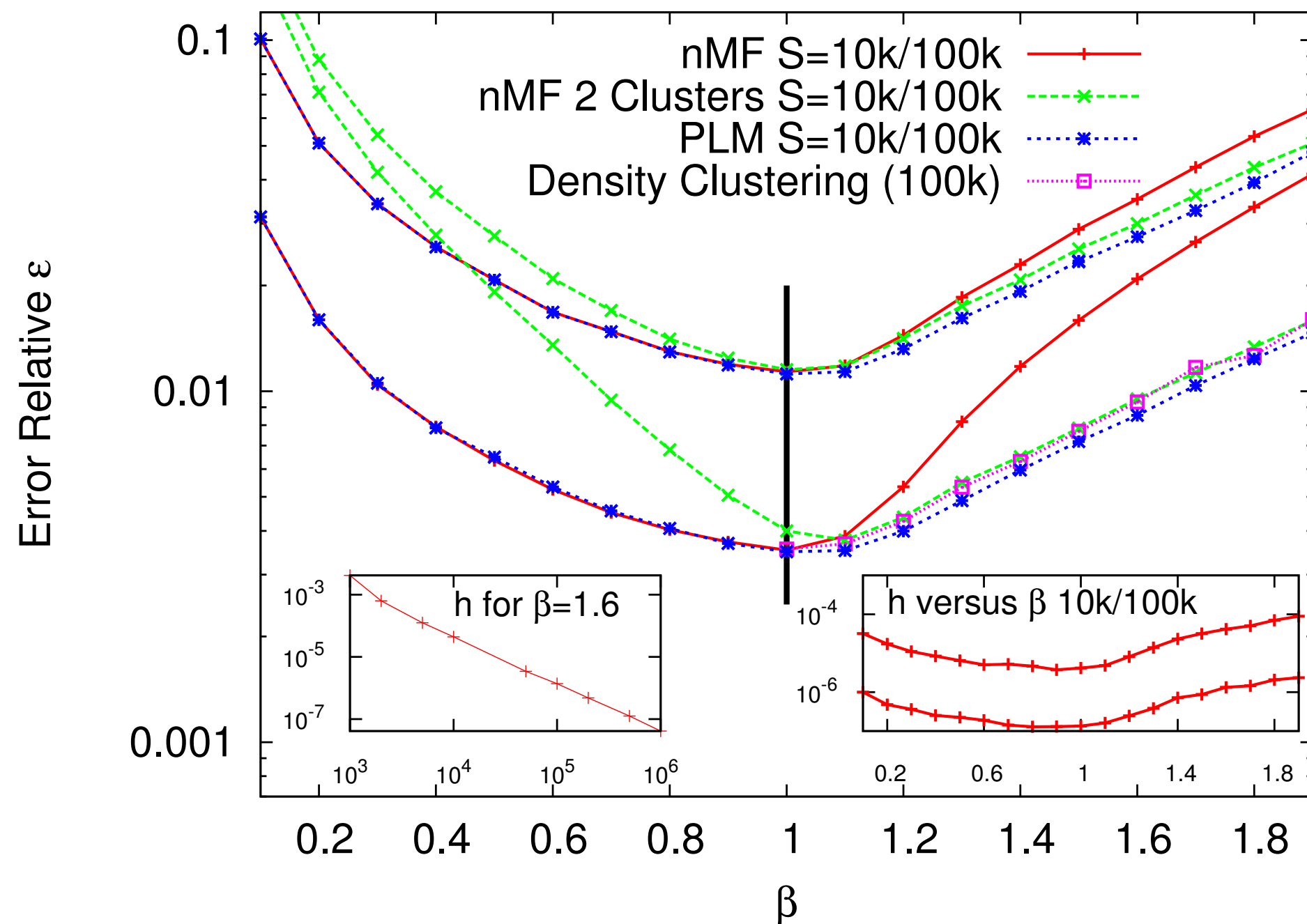


Ergodicity breaking & MFA

- **Estimate model parameters** in a phase with many states
 - Pseudo-likelihood based methods are rather insensitive to ergodicity breaking (see Aurelien's talk)
 - However also MFA can be used if data are properly clustered
 - Each cluster of data returns comparable estimates for couplings and fields

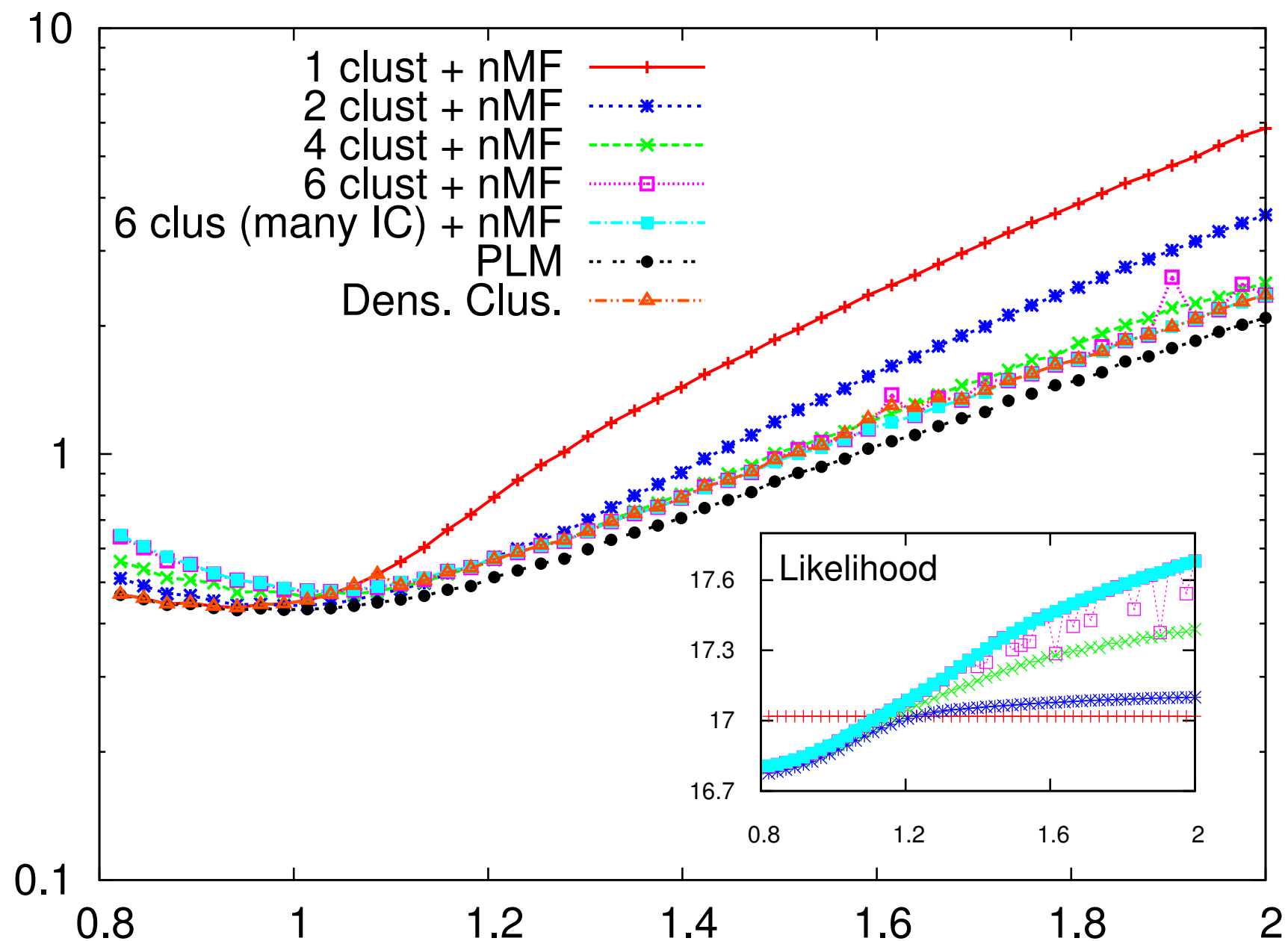
Ergodicity breaking & MFA

Curie-Weiss model $N=100$



Ergodicity breaking & MFA

Hopfield model $P=3$ (6 minima)



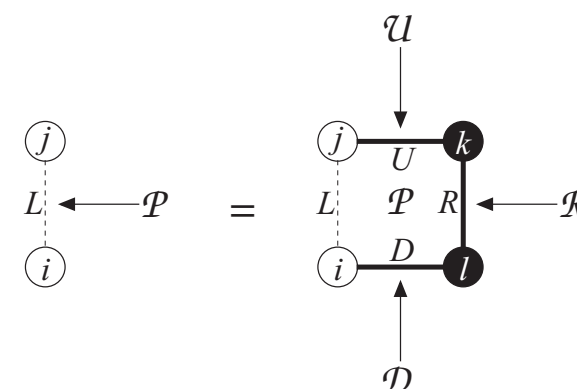
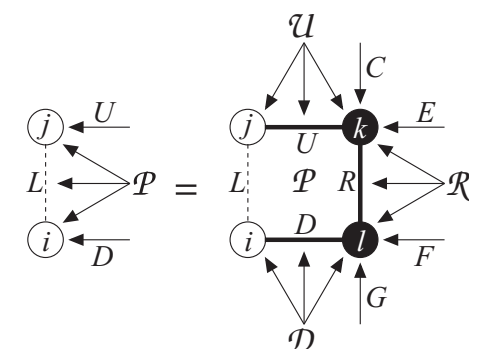
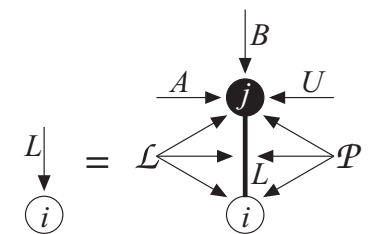
Ergodicity breaking & MFA

- The problem of **estimating marginals** is much harder in presence of ergodicity breaking
- What happens when we use MFA in a disordered model with many states? (relevant for multimodal models)
- On random graphs: replica symmetry breaking (RSB) and Survey Propagation with Parisi parameter m a.k.a. $SP(m)$ is ok to describe 1RSB solutions
With a little effort one can obtain 2RSB solutions...
- On finite dimensional lattices our understanding is still very limited :-(

CVM on finite dimensional spin glasses

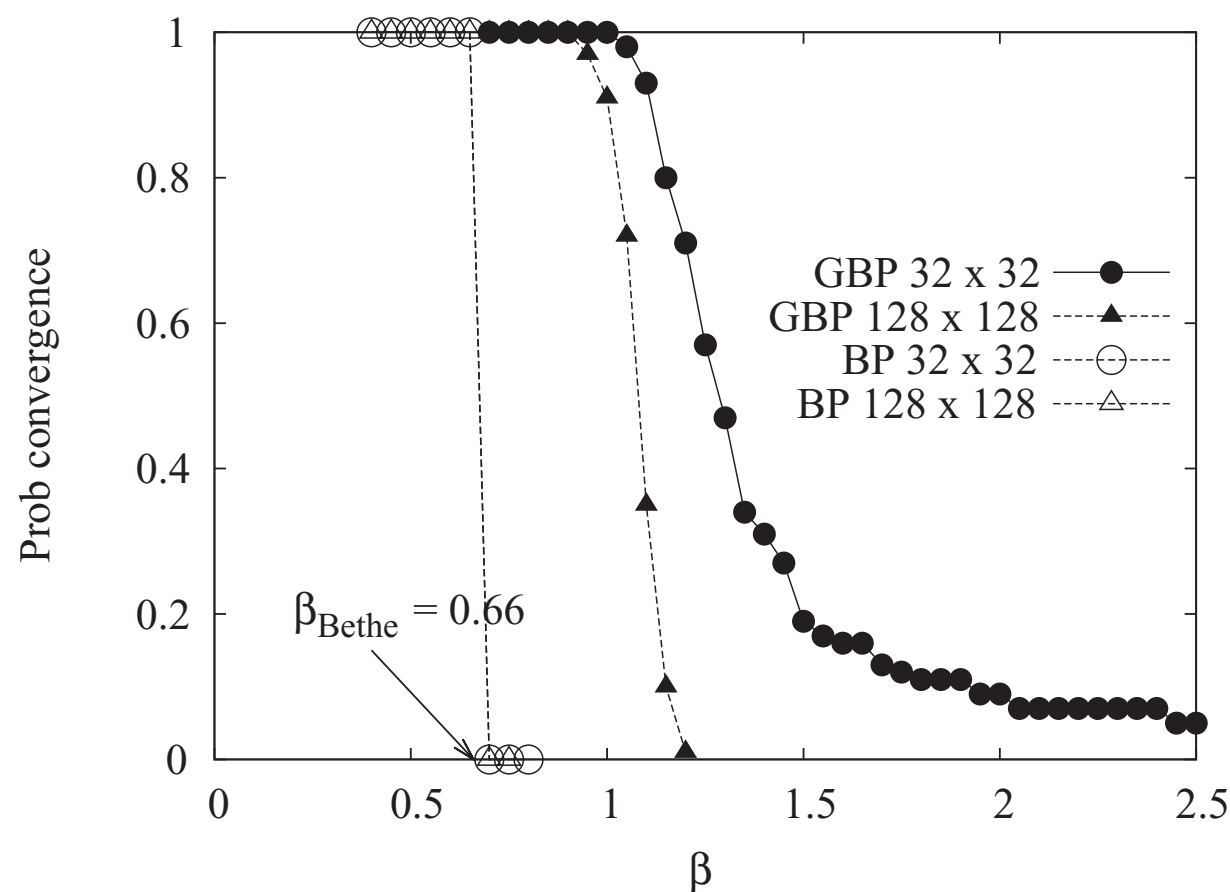
- Edwards-Anderson (EA) model in $d=2$ with symmetric couplings ($J_{ij} = \pm 1$) is the most difficult situation
- For $d=2$ the EA model has no phase transition (in the thermodynamical limit!) but low temperature physics is still dominated by many different local minima in the free-energy

- Algorithms to optimize CVM free-energy:
 - BP
 - plaquette GBP (parent-to-child), HAK (2-ways)
 - Double loop
 - MPA on the dual lattice ($m=0$)



MPA convergence on 2d spin glasses

- Double loop, HAK and MPA on the dual converge at any temperature
- BP and GBP only for high enough temperatures

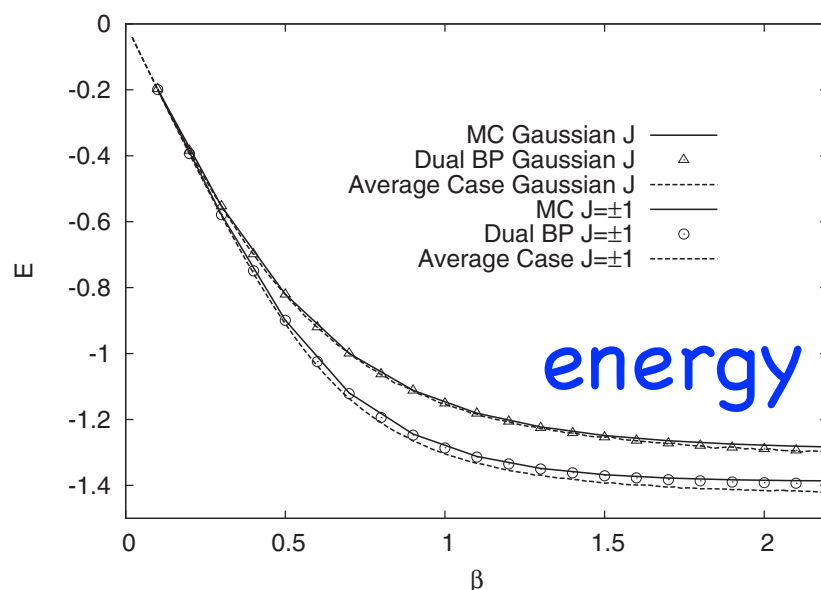
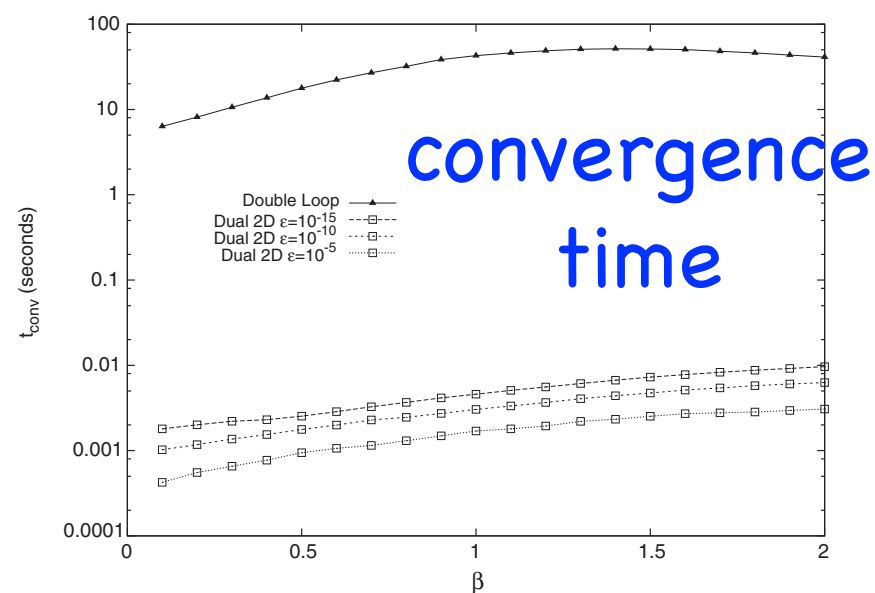


Tricks for
making GBP
converge

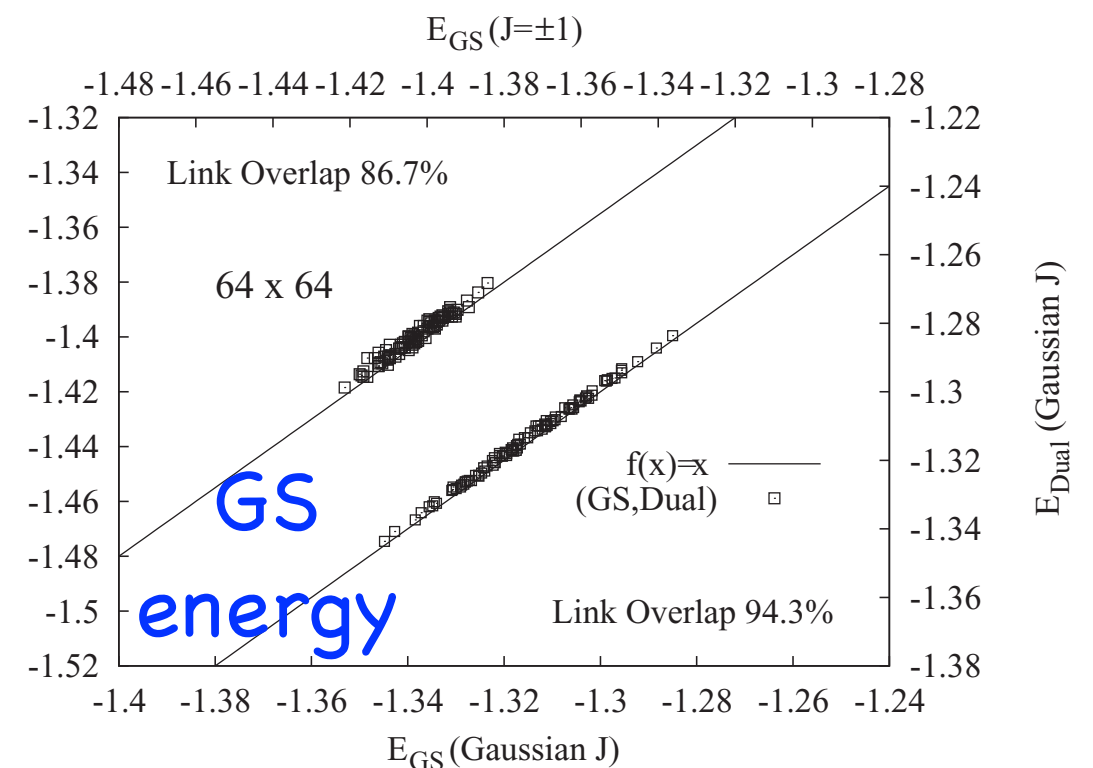
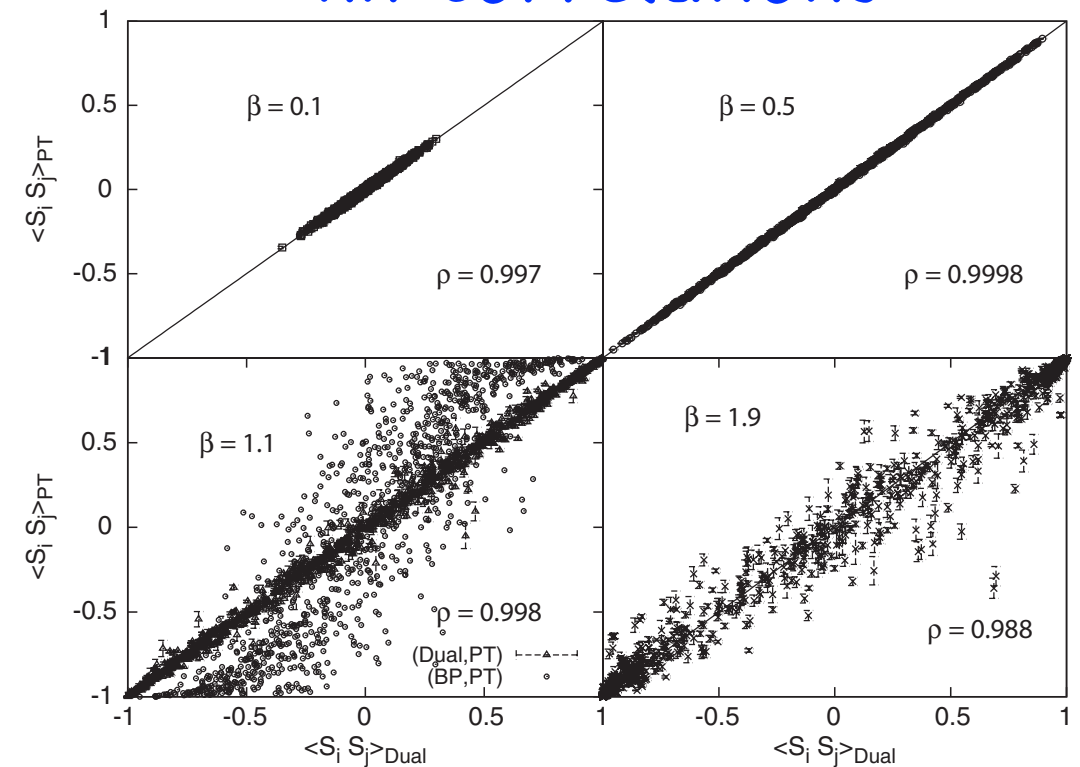
BP critical point on
RRG of degree 4 is
 $\text{atanh}(1/\sqrt{3}) \simeq 0.658$

BP on the dual lattice

- It is very fast and accurate enough to find a very good approximation to ground states



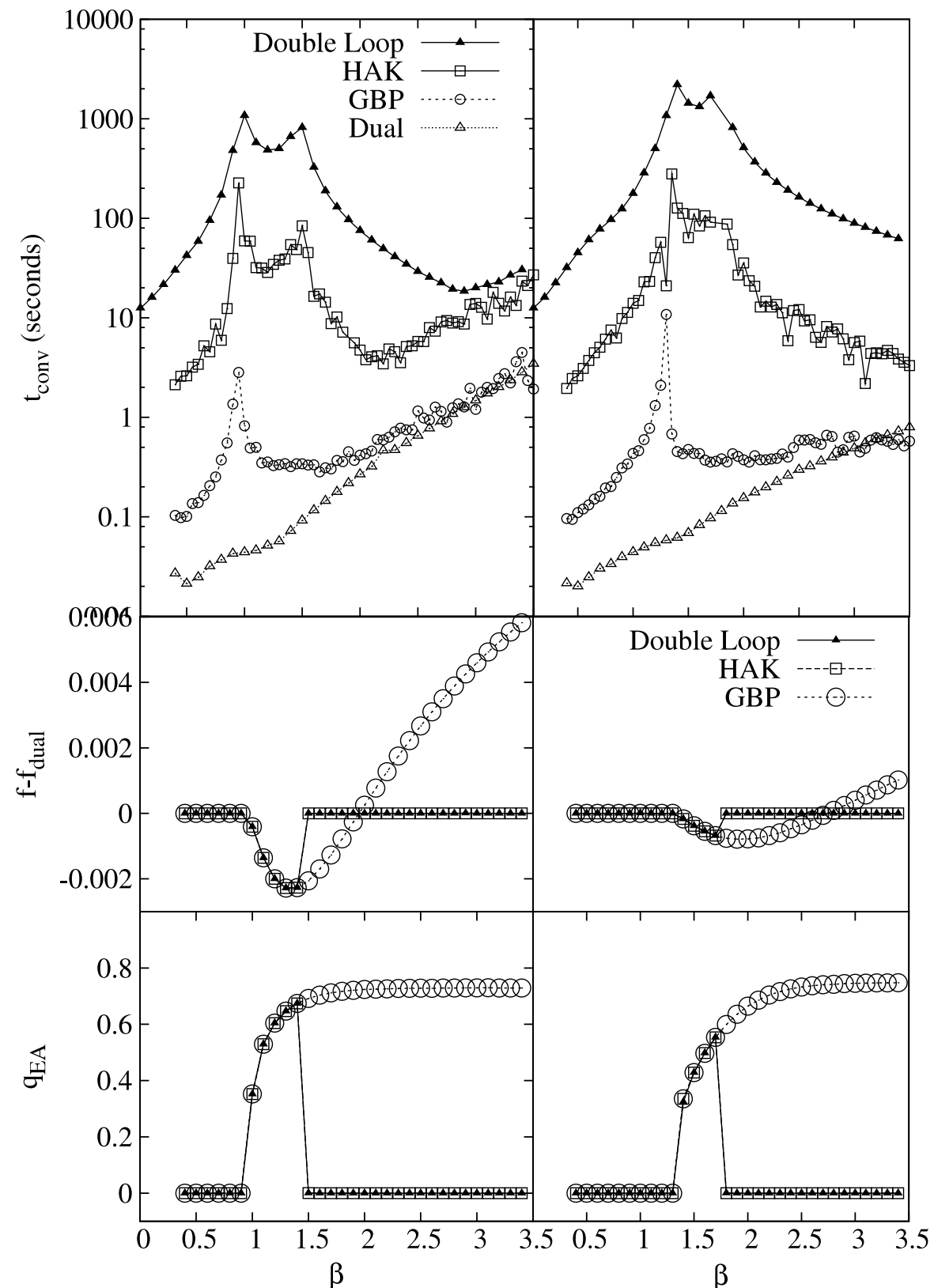
nn correlations



Multiple minima of CVM free-energy

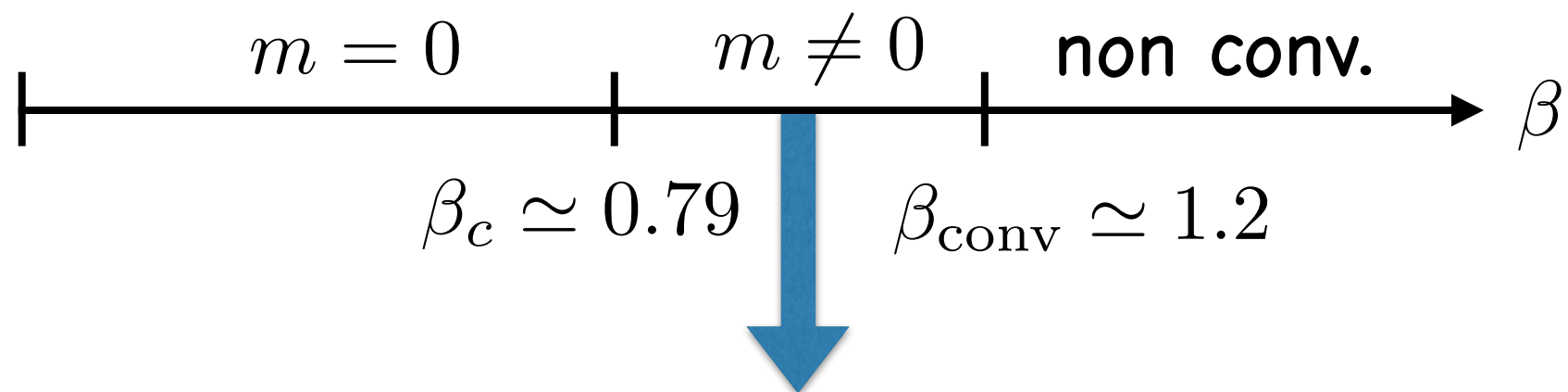
- 2 samples
- 4 algorithms
- convergence time
- free-energy
- order parameter

$$q_{\text{EA}} = \frac{1}{N} \sum_i m_i^2$$

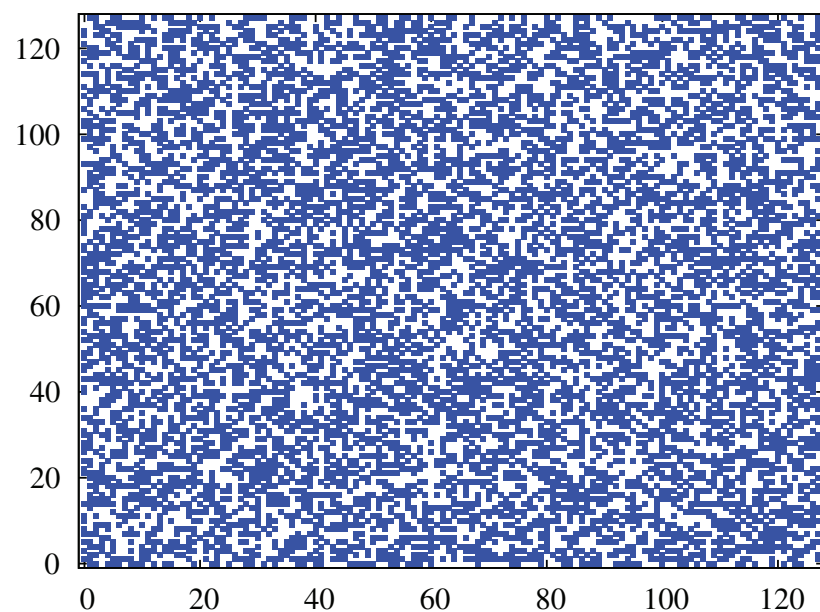


Multiple minima of CVM free-energy

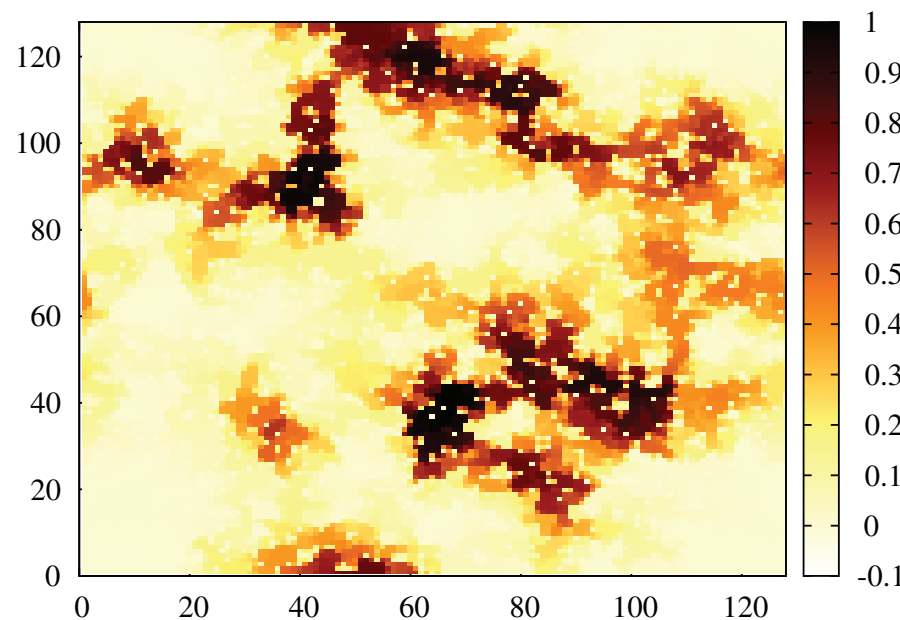
- Running GBP on samples of the 2d EA model the general scenario is the following



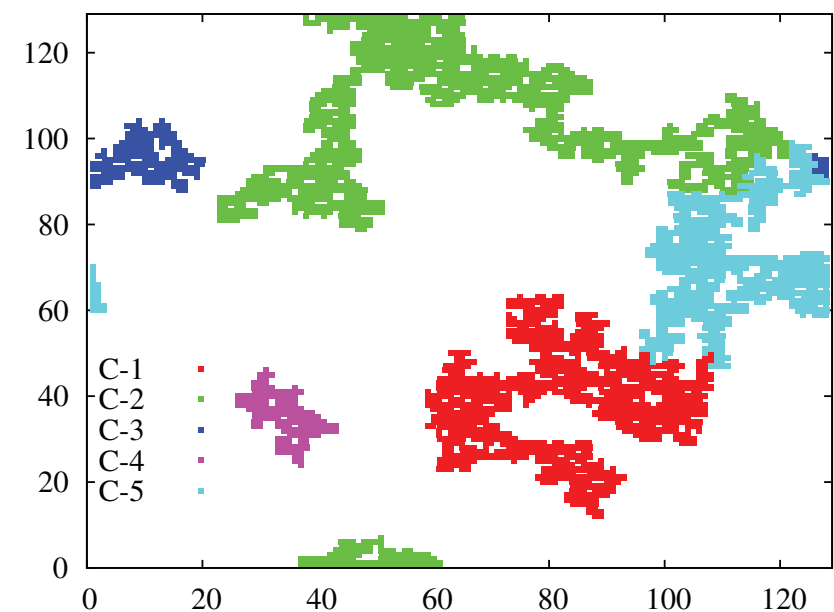
Unfrustrated plaquettes



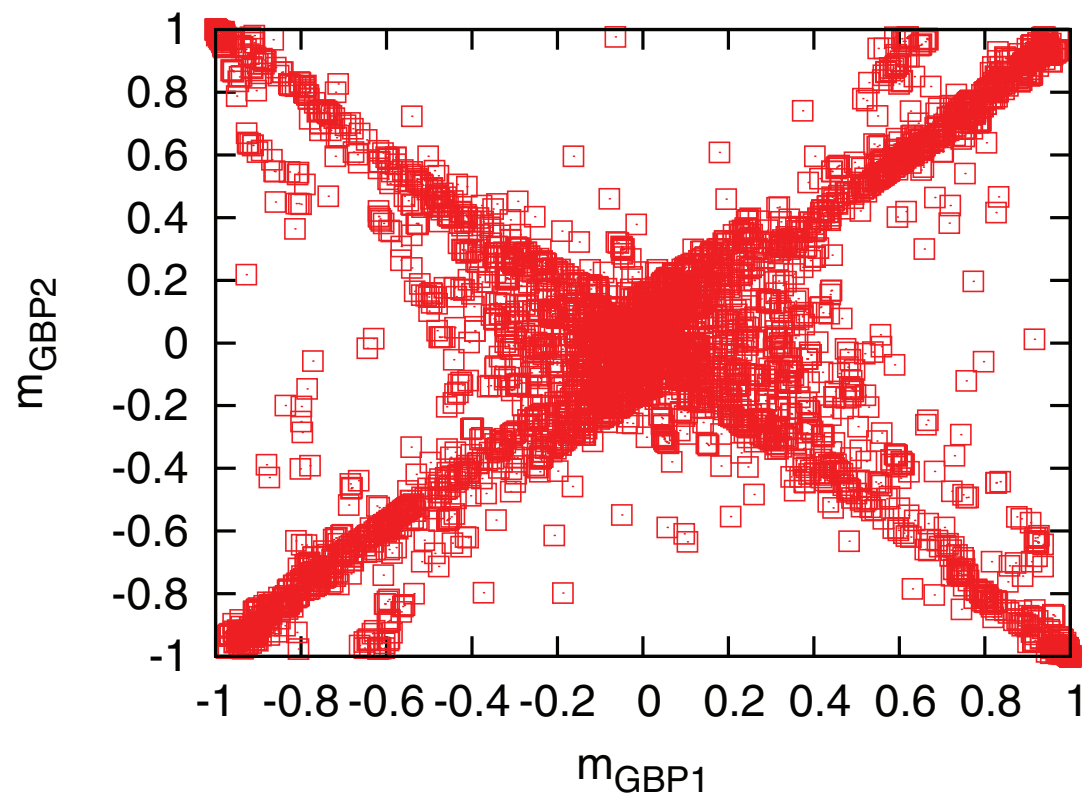
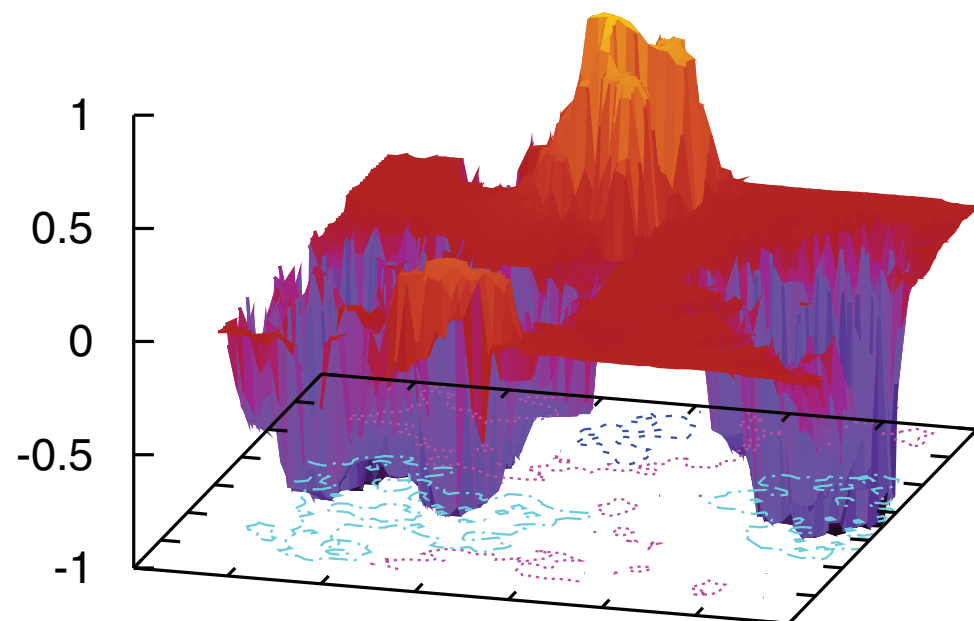
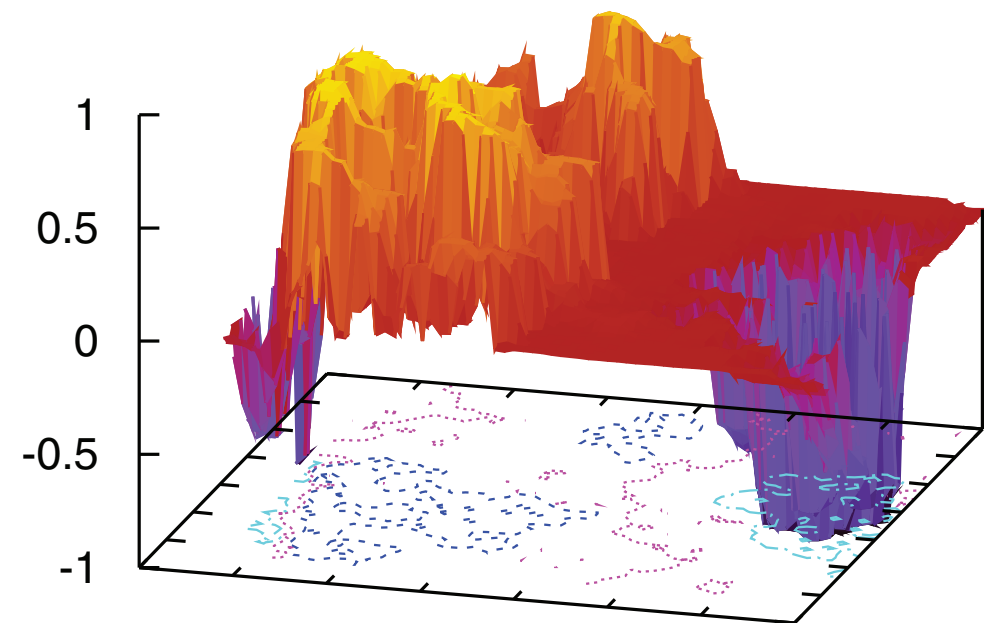
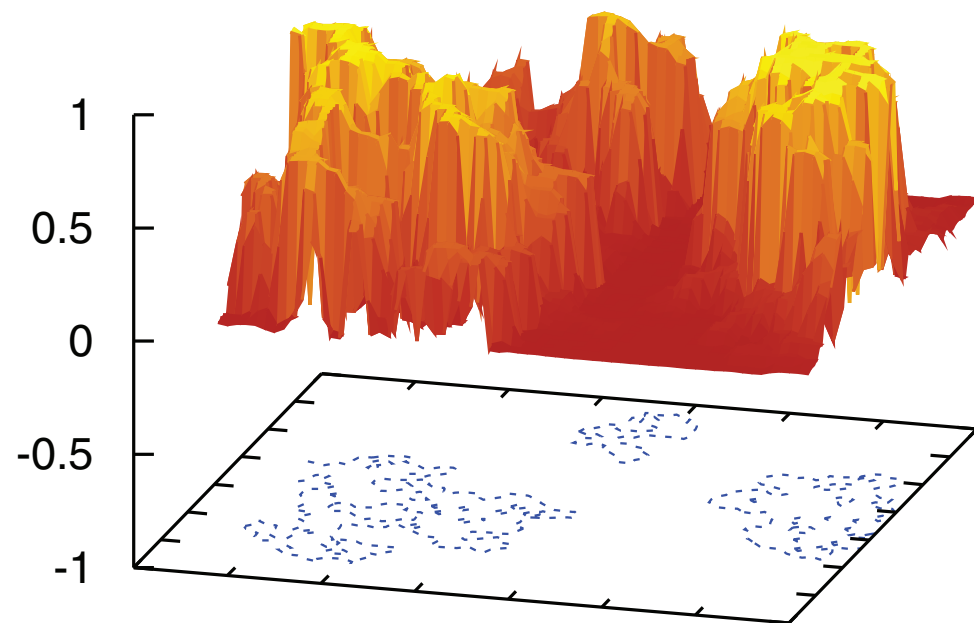
Full GBP configuration



Strongly polarized clusters



Multiple minima of CVM free-energy

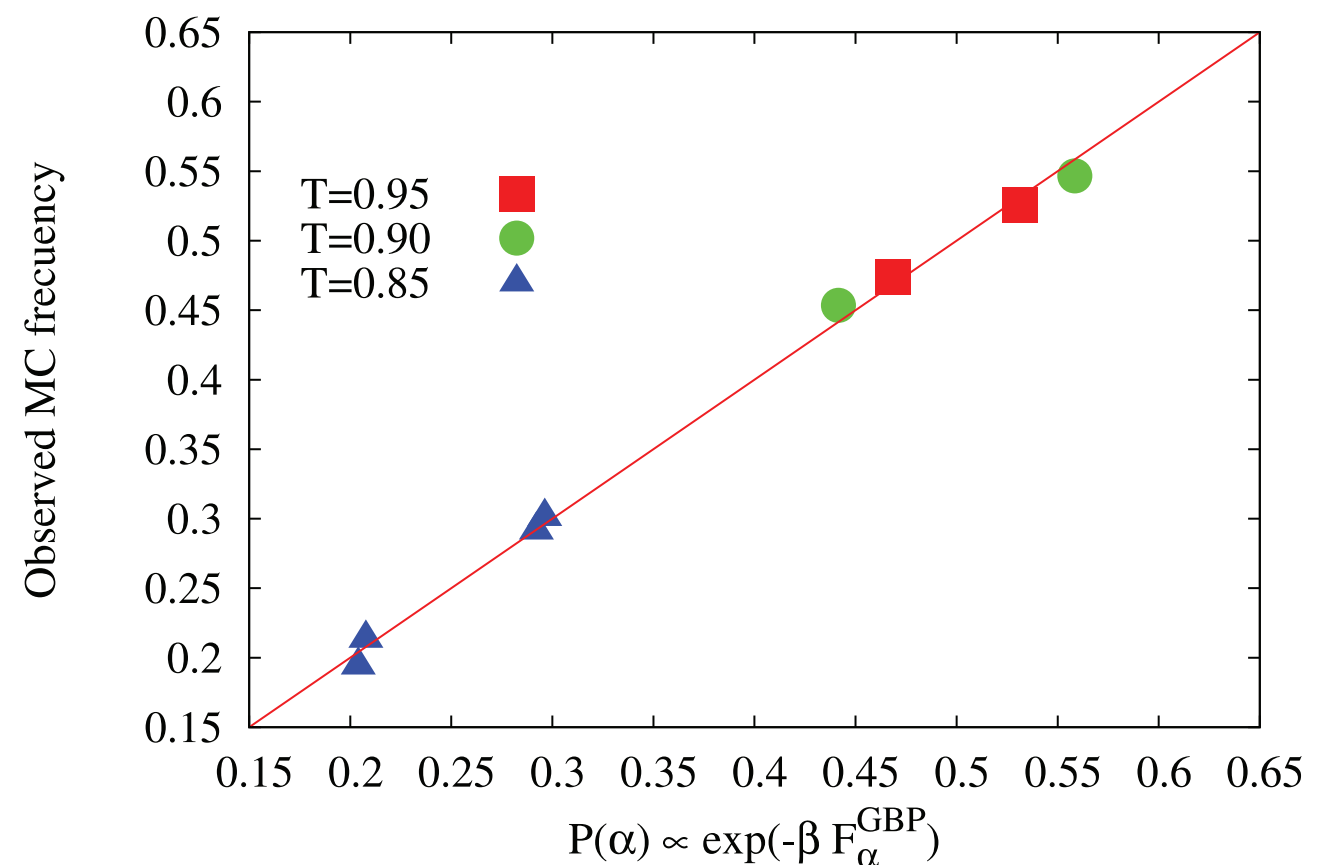


CVM vs Monte Carlo

- Do have the many CVM free-energy minima a physical meaning and role?
- Comparison with Monte Carlo dynamics
- Time spent close to a CVM free-energy minimum is equal to the CVM approximated weight of such a state

$$T_\alpha \propto w_\alpha = \frac{\exp(-\beta F_\alpha)}{\sum_{\alpha'=1}^n \exp(-\beta F_{\alpha'})}$$

time s.t. α was
the closest state



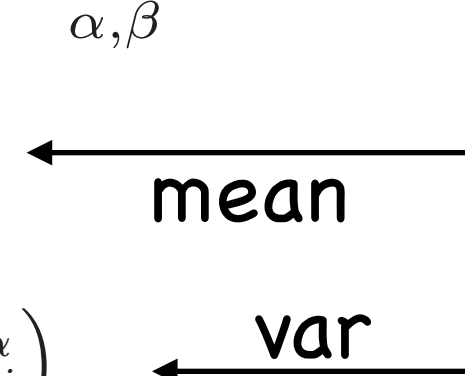
CVM vs Monte Carlo

- Approximating the overlap distribution from CVM states (free-energy minima)

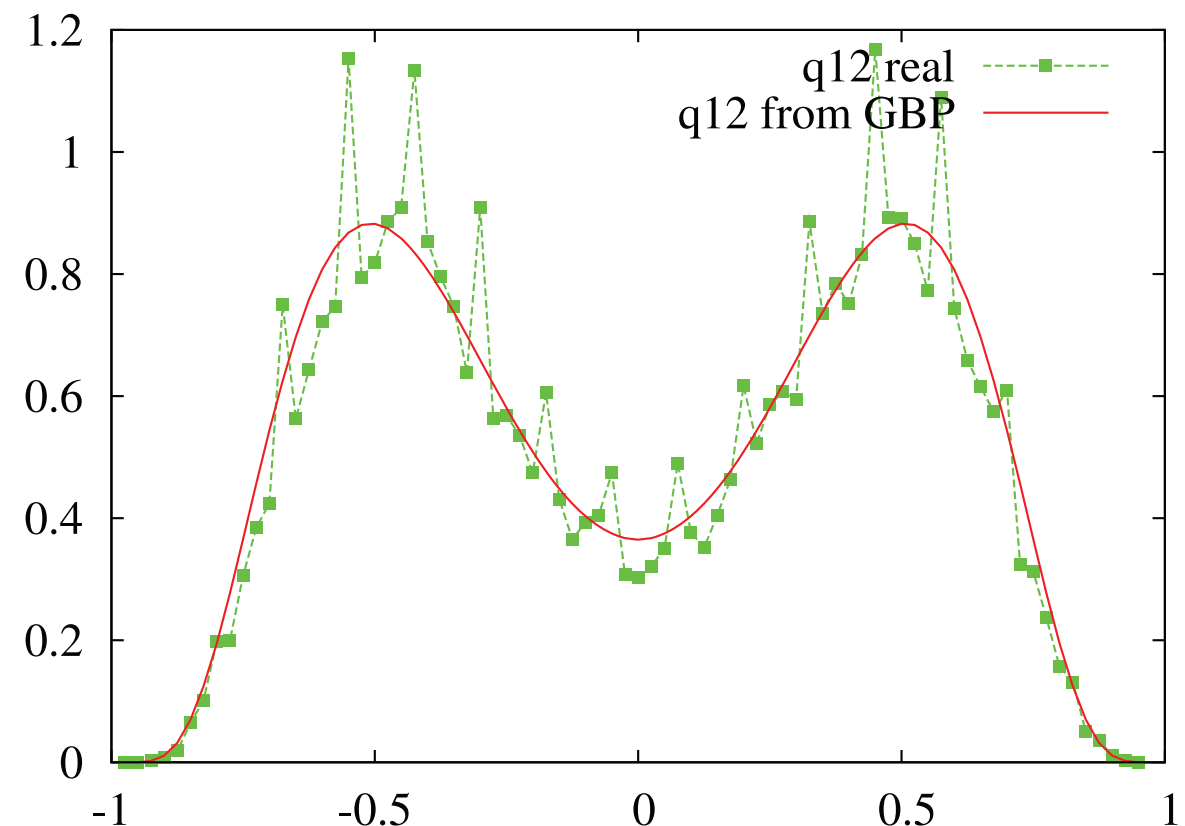
$$P(q) = \sum_{\alpha, \beta} w_{\alpha} w_{\beta} P_{\alpha\beta}(q)$$

$$q_{\alpha\beta} = \left\langle \frac{1}{N} \sum_i s_i^{\alpha} s_i^{\beta} \right\rangle = \frac{1}{N} \sum_i m_i^{\alpha} m_i^{\beta}$$

$$\sigma_{\alpha\beta}^2 = \frac{1}{N^2} \sum_{ij} \left(C_{ij}^{\alpha} C_{ij}^{\beta} + m_i^{\alpha} m_j^{\alpha} C_{ij}^{\beta} + m_i^{\beta} m_j^{\beta} C_{ij}^{\alpha} \right)$$



estimation of
would require
generalized
SuscProp...



Summary & open problems

- General framework for making MFA consistent with linear response
 - Recover several previous approximation (apparently unrelated): e.g. adaptive-TAP and SM approx.
 - Improves inference, but has some limitations (to be overcome...)
- GBP can converge to many non-trivial fixed point with physical relevance
 - Still missing a broad-purpose region-based algorithm that can deal with the many free-energy minima...

Figures in this talk are from

- *Inference algorithm for finite-dimensional spin glasses: Belief Propagation on the dual lattice*
A. Lage-Castellanos, R. Mulet, F. Ricci-Tersenghi and T. Rizzo
Phys. Rev. E 84, 046706 (2011)
- *Characterizing and Improving Generalized Belief Propagation Algorithms on the 2D Edwards-Anderson Model*
E. Dominguez, A. Lage-Castellanos, R. Mulet, F. Ricci-Tersenghi and T. Rizzo
J. Stat. Mech. P12007 (2011)
- *The Bethe approximation for solving the inverse Ising problem: a comparison with other inference methods*
F. Ricci-Tersenghi
J. Stat. Mech. P08015 (2012)
- *Mean-field method with correlations determined by linear response*
J. Raymond and F. Ricci-Tersenghi
Phys. Rev. E 87, 052111 (2013)
- *Message passing and Monte Carlo algorithms: Connecting fixed points with metastable states,*
A. Lage-Castellanos, R. Mulet and F. Ricci-Tersenghi,
Europhys. Lett. 107, 57011 (2014)
- *Solving the inverse Ising problem by mean-field methods in a clustered phase space with many states*
Aurélien Decelle, Federico Ricci-Tersenghi,
arxiv:1501.03034 (2015)
- *Correction of variational methods with pairwise linear response identities*
J. Raymond and F. Ricci-Tersenghi,
in preparation (2015)