Inferring Sparsity:

Compressed Sensing Using Generalized Restricted Boltzmann Machines

Eric W. Tramel

iTWIST 2016 *— Aalborg, DK*

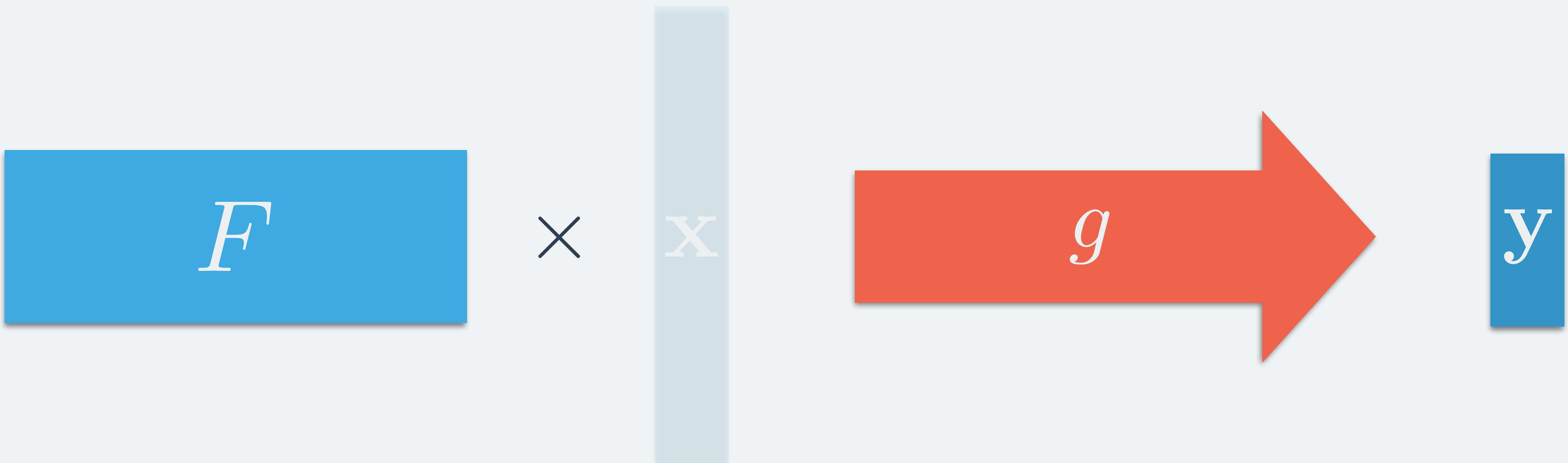
24 August 2016

Andre MANOEL, Francesco CALTAGIRONE, Marylou GABRIE, Florent KRZAKALA

Inverse Problems

General Linear Problem: $\mathbf{y} = g\left(F\mathbf{x}\right)$

$(M \times N)$ $(N \times 1)$ $(M \times 1)$



Projection Matrix **Signal?** **Channel** **Measurements**

Compressed Sensing, Regression, Deconvolution/Debluring, Localization, Super-Resolution, Medical Image Reconstruction (CT/MRI), In-Painting, Denoising, Inference, etc.

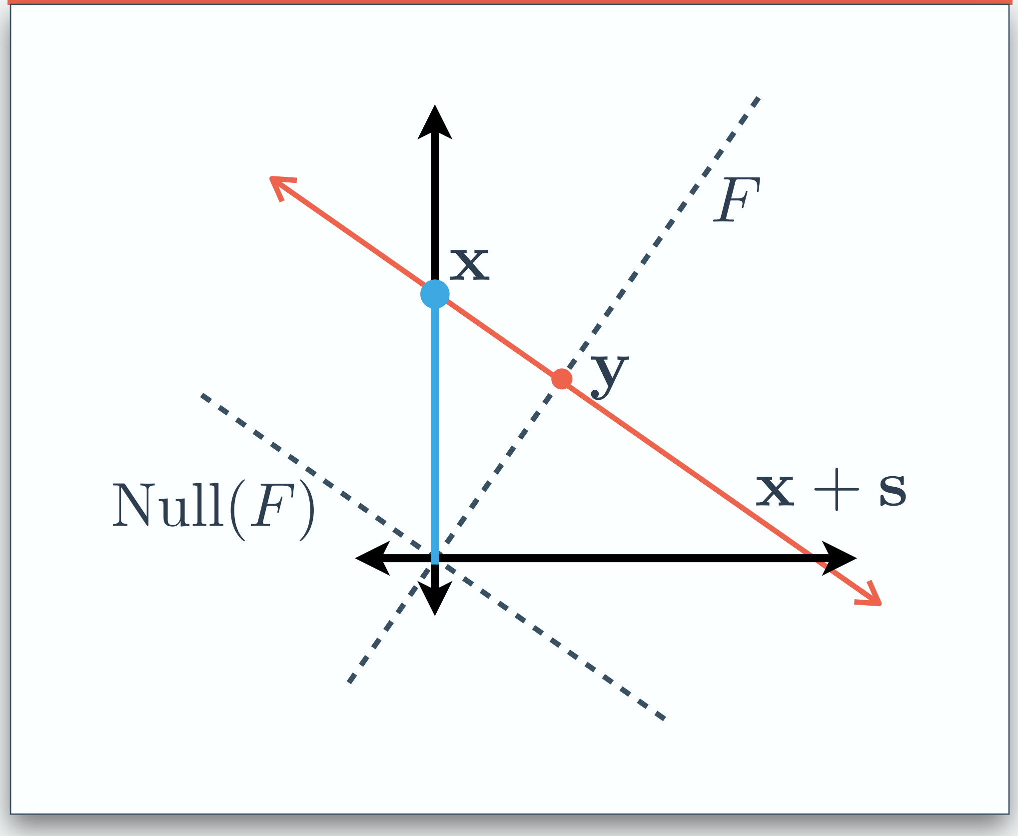
Example — Compressed Sensing

$$\boxed{\mathbf{y} = F\mathbf{x} + \mathbf{w} \qquad w_\mu \sim \mathcal{N}(0, \Delta)}$$

How do we obtain **x** from **y** knowing….

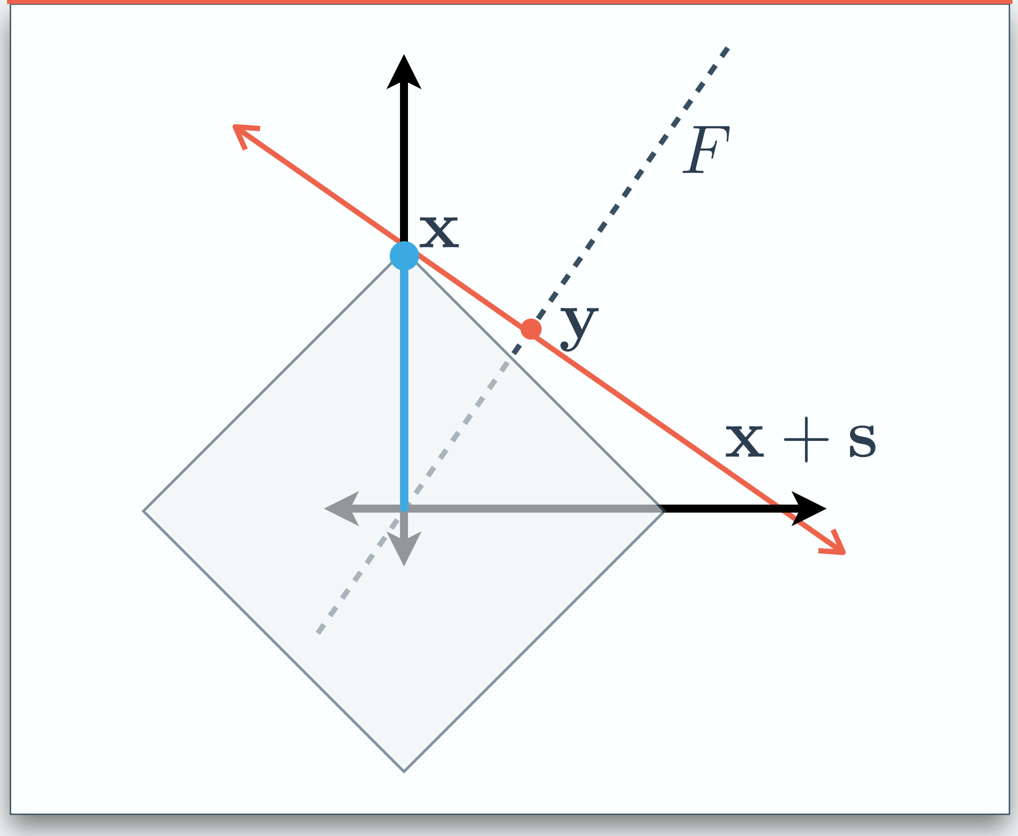
- **g** is AWGN,
- **x** is **K**-Sparse,
- **F** is iid random,
- and **M << N** ?

OLS is under-determined, in general we can't!



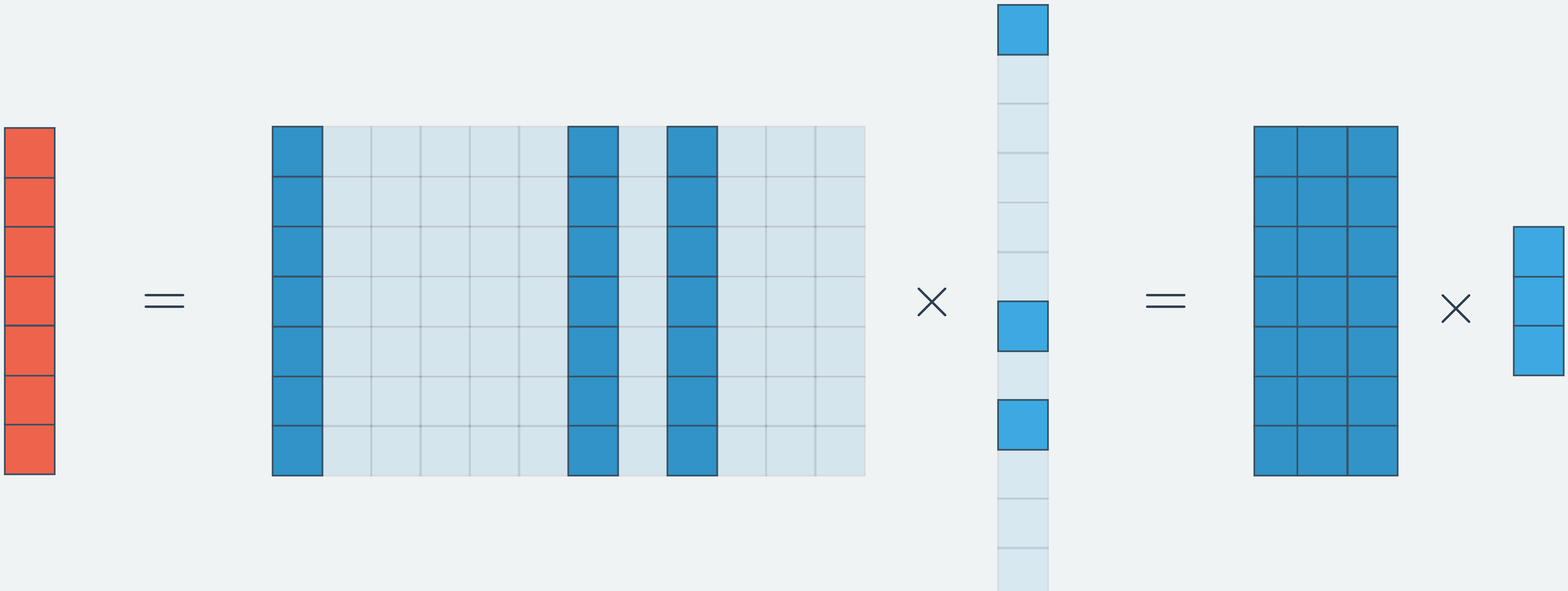
(EC & TT, 2005)
(EC, JR, & TT, 2006)
(EC & MW, 2008)

With **Sparsity**, we can !



Sparsity & Recovery

For **M>=K**, we can recover with OLS, up to noise, if we are given the support locations by an oracle.



However: Without an oracle, finding **S** brute-force is a combinatorial problem!

$$\boxed{\arg\min_{S \in \mathbb{S}} \quad ||\mathbf{y} - F_S \mathbf{x}_S||_2^2}$$

Optimization Approaches

$$\mathbf{y} = F\mathbf{x} + \mathbf{w} \qquad w_\mu \sim \mathcal{N}(0, \Delta)$$

Greedy Approach

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \quad ||\mathbf{x}||_0 \quad \text{s.t.} \quad ||\mathbf{y} - F\mathbf{x}||_2^2 \leq \epsilon$$

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \quad ||\mathbf{y} - F\mathbf{x}||_2^2 \quad \text{s.t.} \quad ||\mathbf{x}||_0 \leq K$$

- Greedily searching for support, solving OLS support.

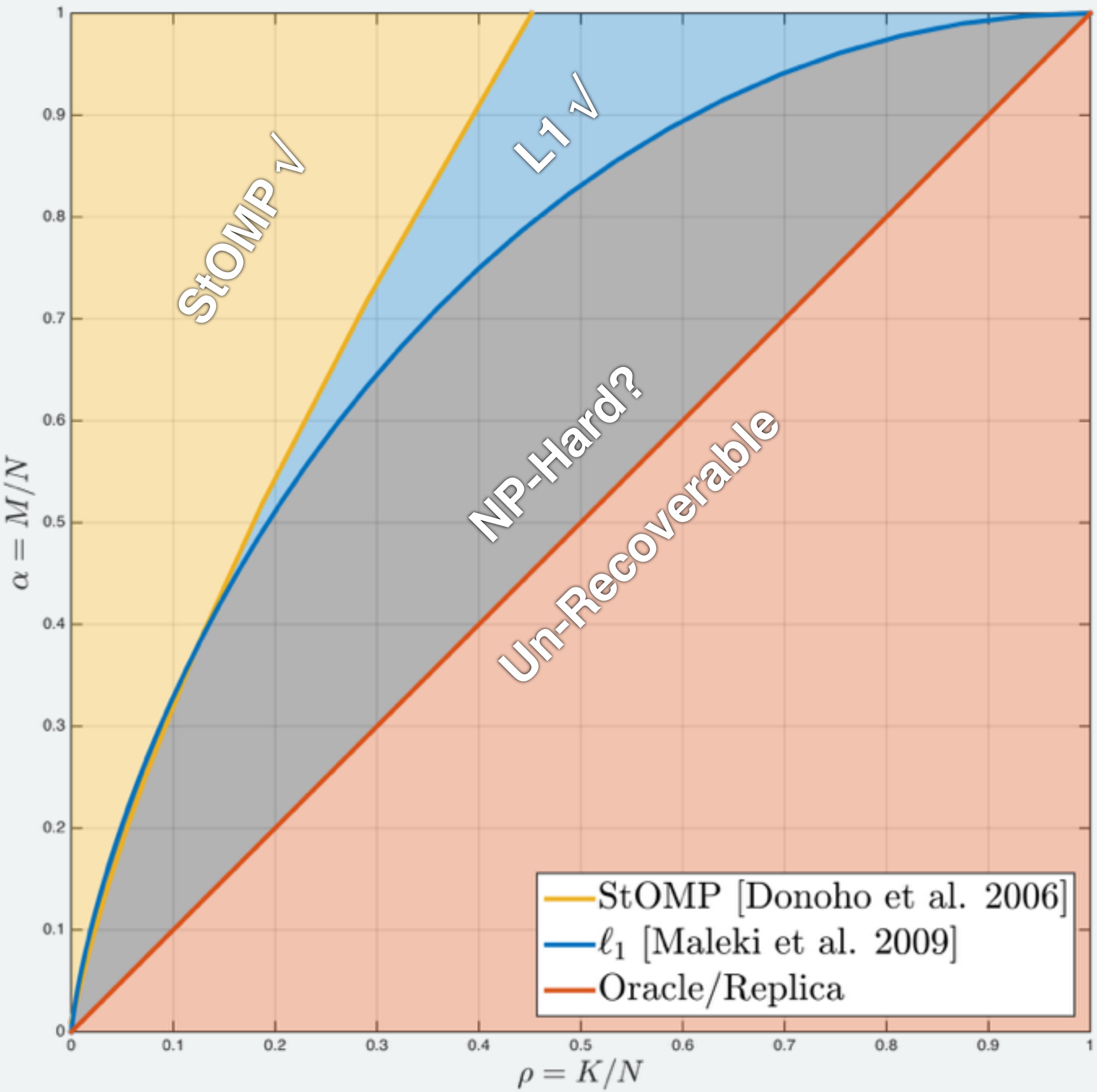
Convex Approach

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \quad ||\mathbf{x}||_1 \quad \text{s.t.} \quad ||\mathbf{y} - F\mathbf{x}||_2^2 \leq \epsilon$$

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \quad ||\mathbf{y} - F\mathbf{x}||_2^2 + \lambda||\mathbf{x}||_1$$

- Relax L0 penalty to convex L1 penalty ("pointiest" convex Lp)

Phase Diagram for CS



Bayesian Approaches

$$\mathbf{y} = F\mathbf{x} + \mathbf{w} \qquad w_\mu \sim \mathcal{N}(0, \Delta)$$

Maximum *a posteriori* (MAP)

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \; P(\mathbf{x}|\mathbf{y}, F)$$

- Find signal to maximize probability.
- Can use unnormalized posterior — minimize negative log prob.
- For some settings — maps to convex optimization.

Minimum Mean Square Error (MMSE)

$$\hat{\mathbf{x}} = \mathbb{E}[\mathbf{x}] = \int \mathrm{d}\mathbf{x} \;\; \mathbf{x} \; P(\mathbf{x}|\mathbf{y}, F)$$

- Average over posterior distribution.

Defining the Posterior

Bayes' Rule

$$P(\mathbf{x}|\mathbf{y}, F) = \frac{1}{Z} P(\mathbf{y}|\mathbf{x}, F) P_0(\mathbf{x})$$

Likelihood defined by stochastic description of **g**.

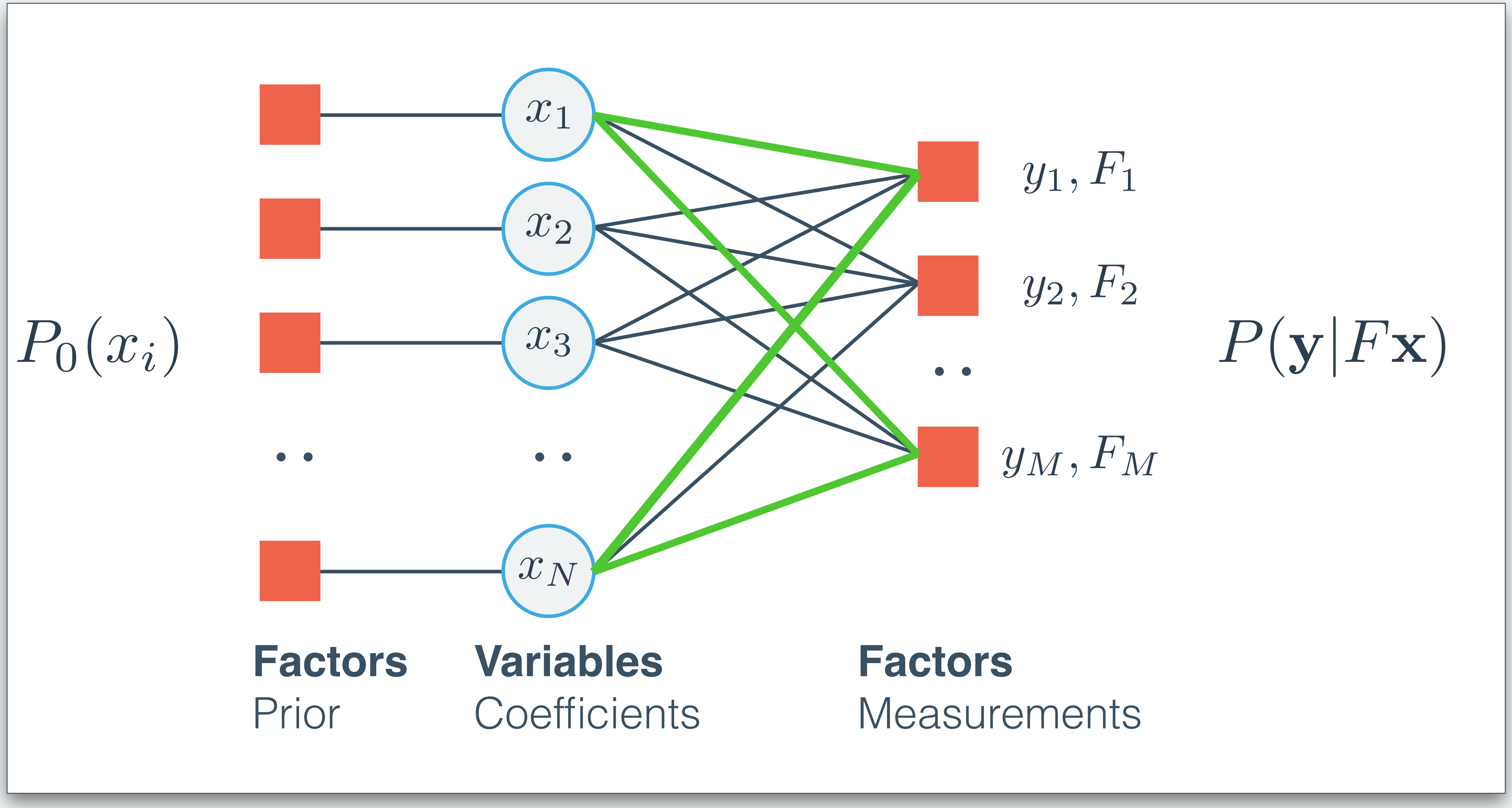
Posterior — Factorized Prior, AWGN Channel

$$P(\mathbf{x}|\mathbf{y}, F) = \frac{1}{Z} \prod_{\mu} \frac{1}{\sqrt{2\pi\Delta}} \exp\left\{ -\frac{1}{2\Delta} \left(y_\mu - \sum_i F_{\mu i} x_i \right)^2 \right\} \prod_i P_0(x_i)$$

For *exact* posterior, we must calculate an intractable **Z**!

Inference: We can approximate it with Belief Propagation.

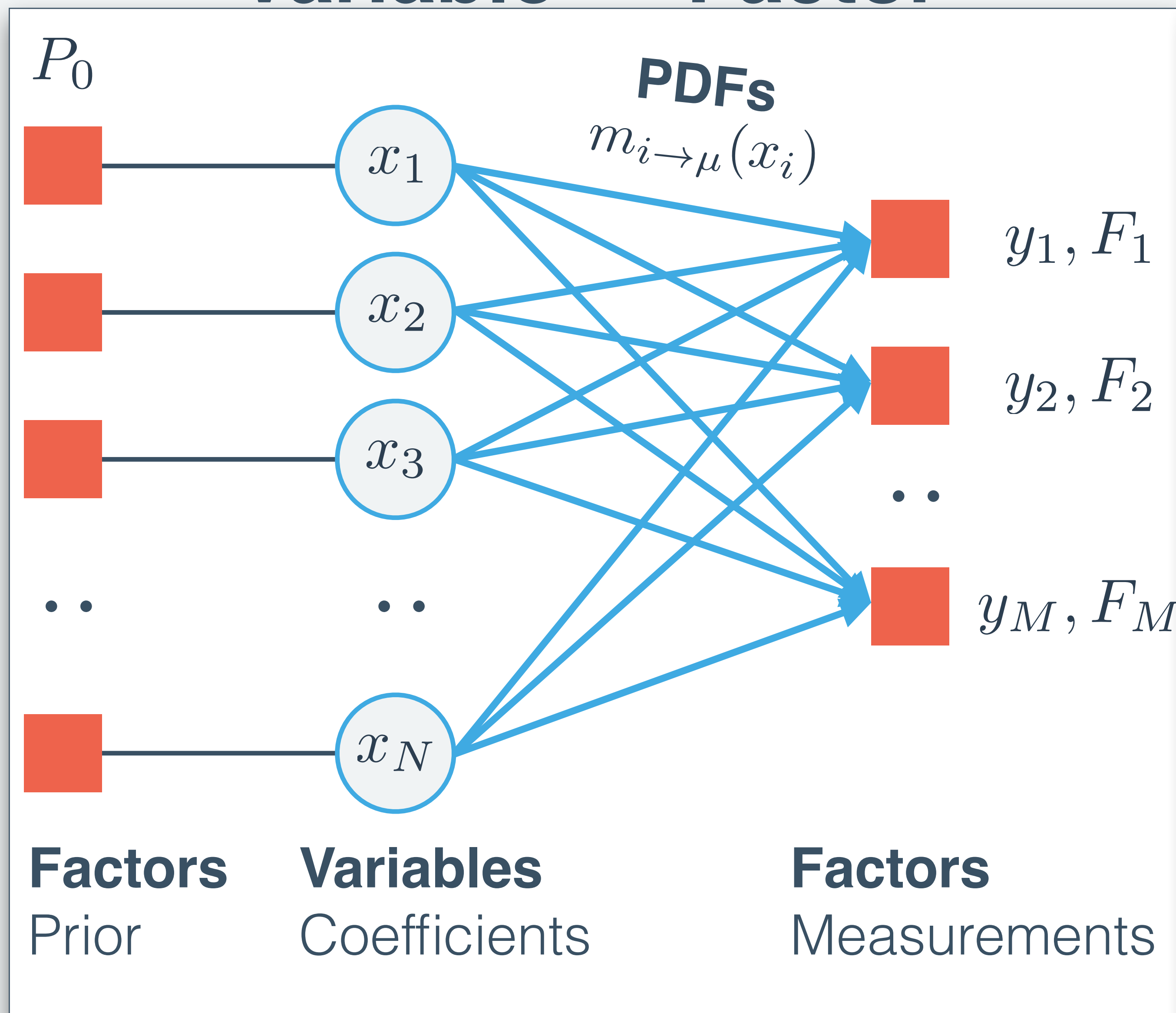
Graphical Model of Posterior



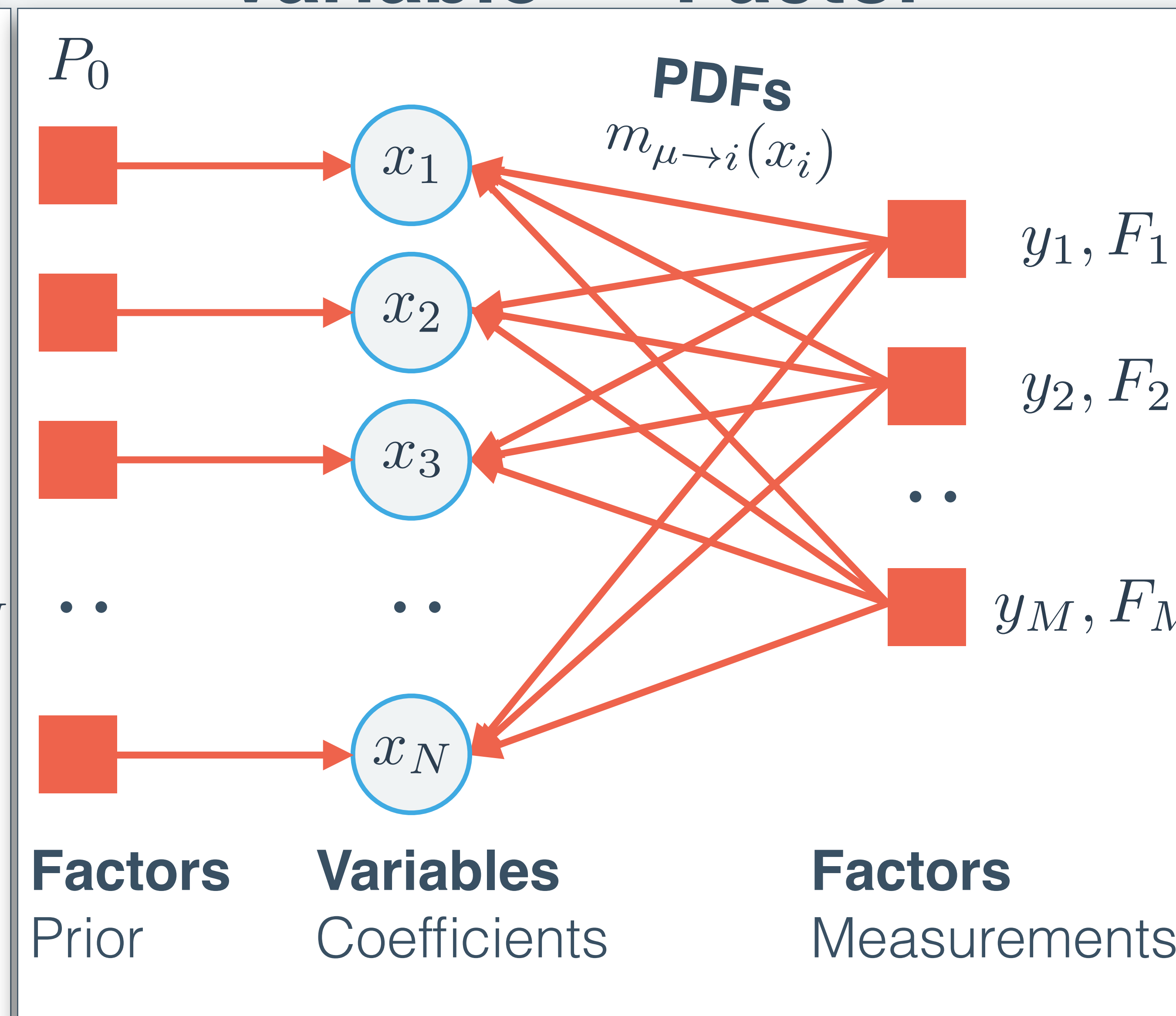
Loopy Belief Propagation — The presence of many loops makes exact inference impossible, but approximate inference may be tractable and "accurate" [Weiss 2000].

Inference via BP

Messages
Variable → Factor



Messages
Variable → Factor



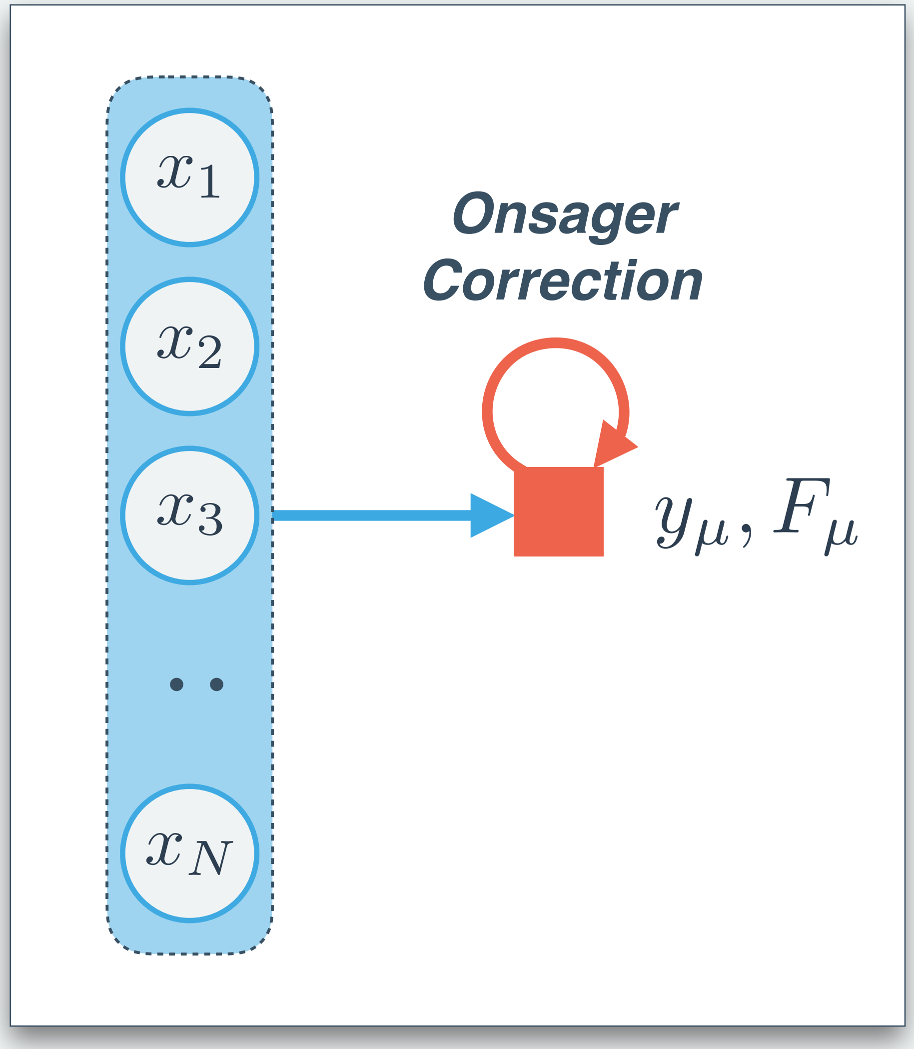
Goal: Produce

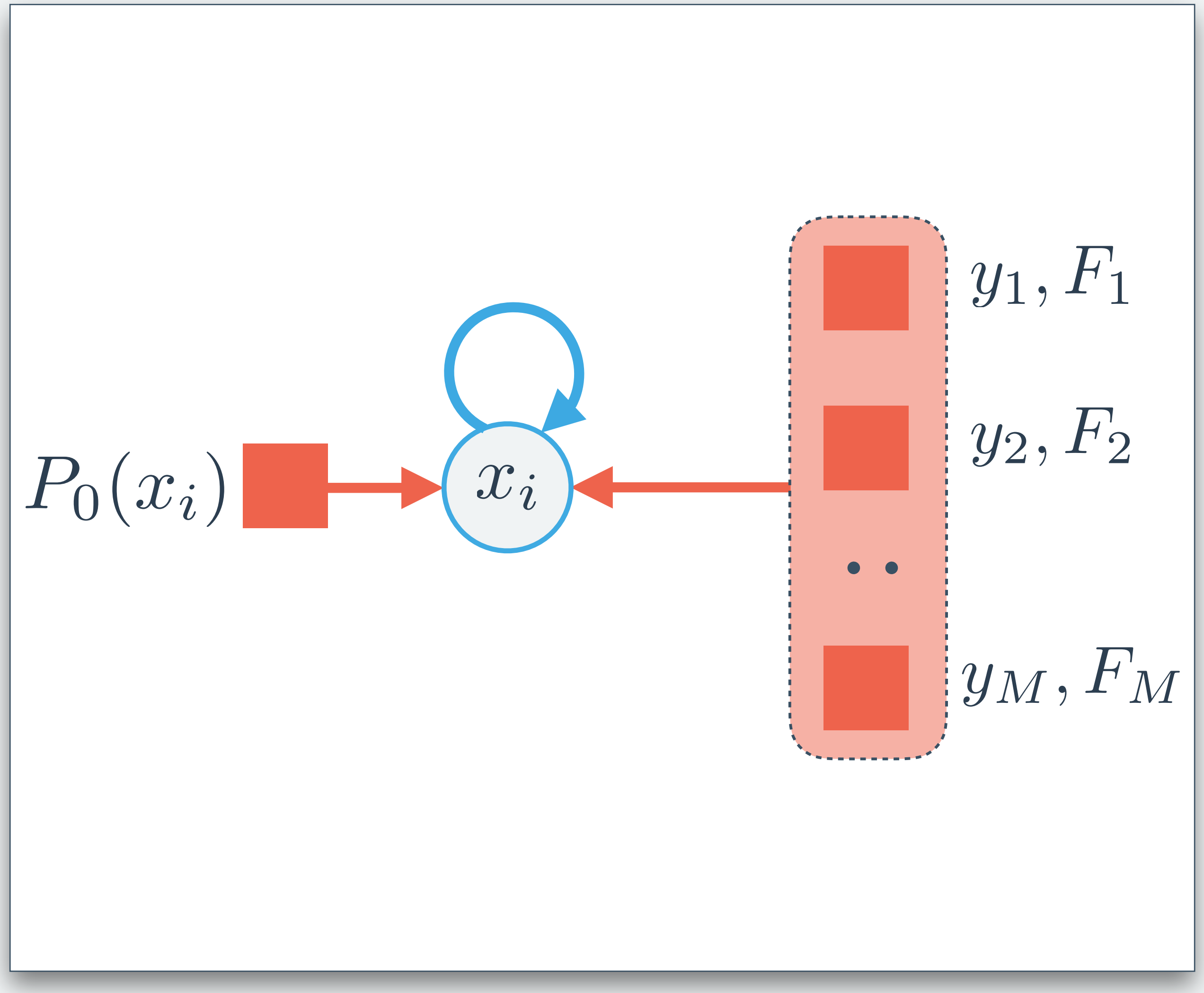
$$P(x_i|\mathbf{y}, F) \propto P_0(x_i) \prod_{\mu} m_{\mu\to i}(x_i)$$

r-BP to AMP via TAP

A Corrected Mean-Field (Donoho, Maleki, Montanari 2009)

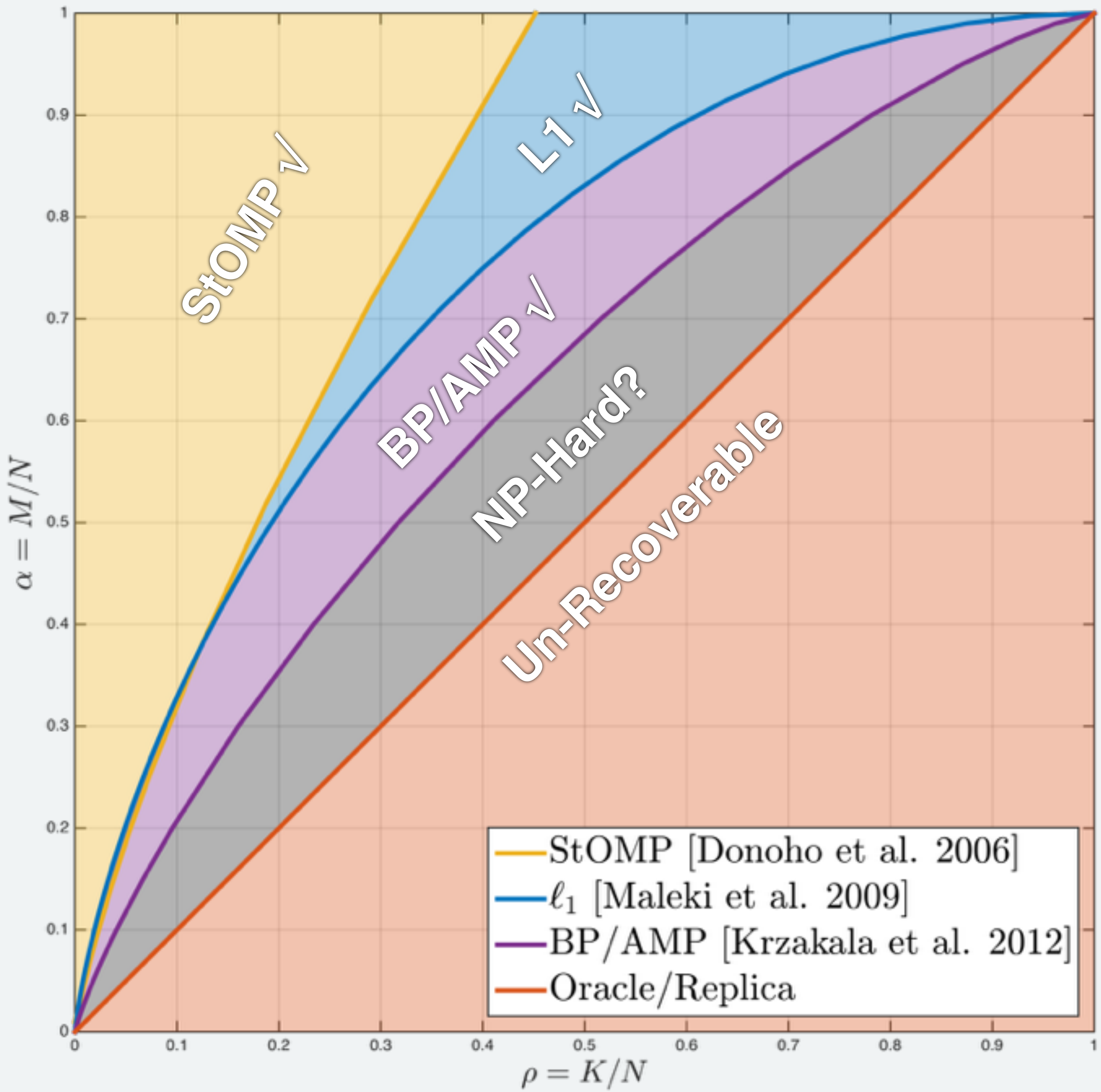
If **F** is ***dense*** and if its entries are ***uncorrelated***, then message means and variances are ***nearly independent*** of any ***single*** edge message in the limit ***N→∞***.





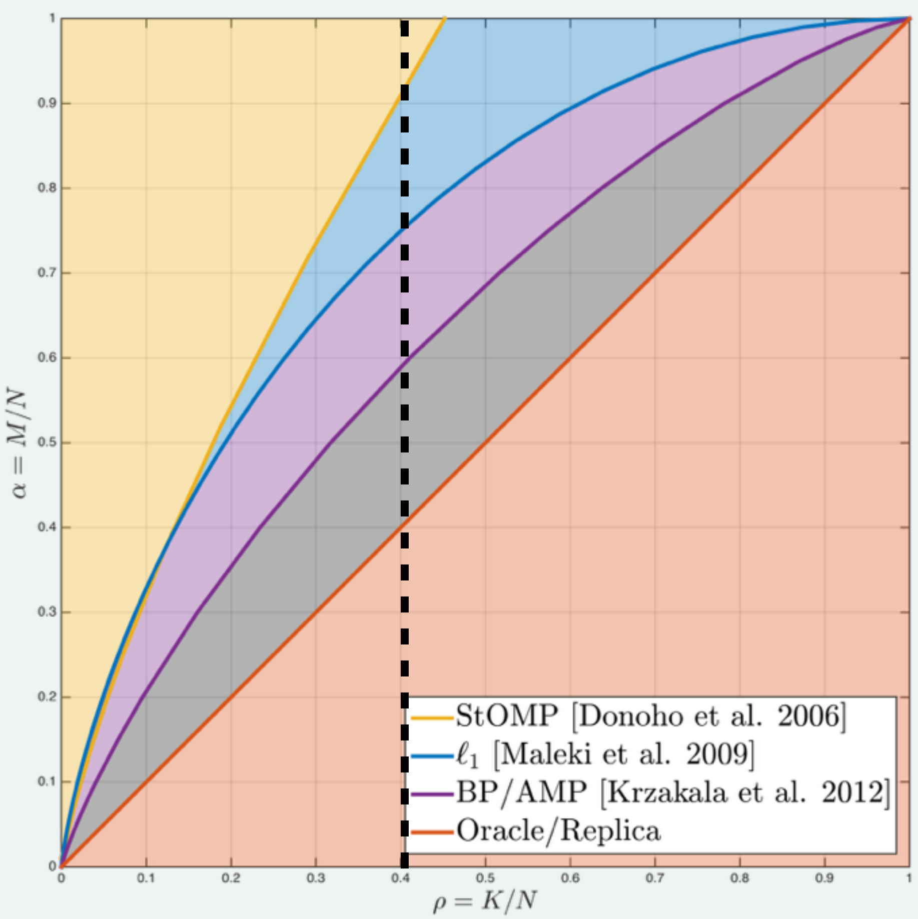
Big Savings: Compute Burden $\mathrm{O}(\alpha N^2) \rightarrow \mathrm{O}((1+\alpha)N)$

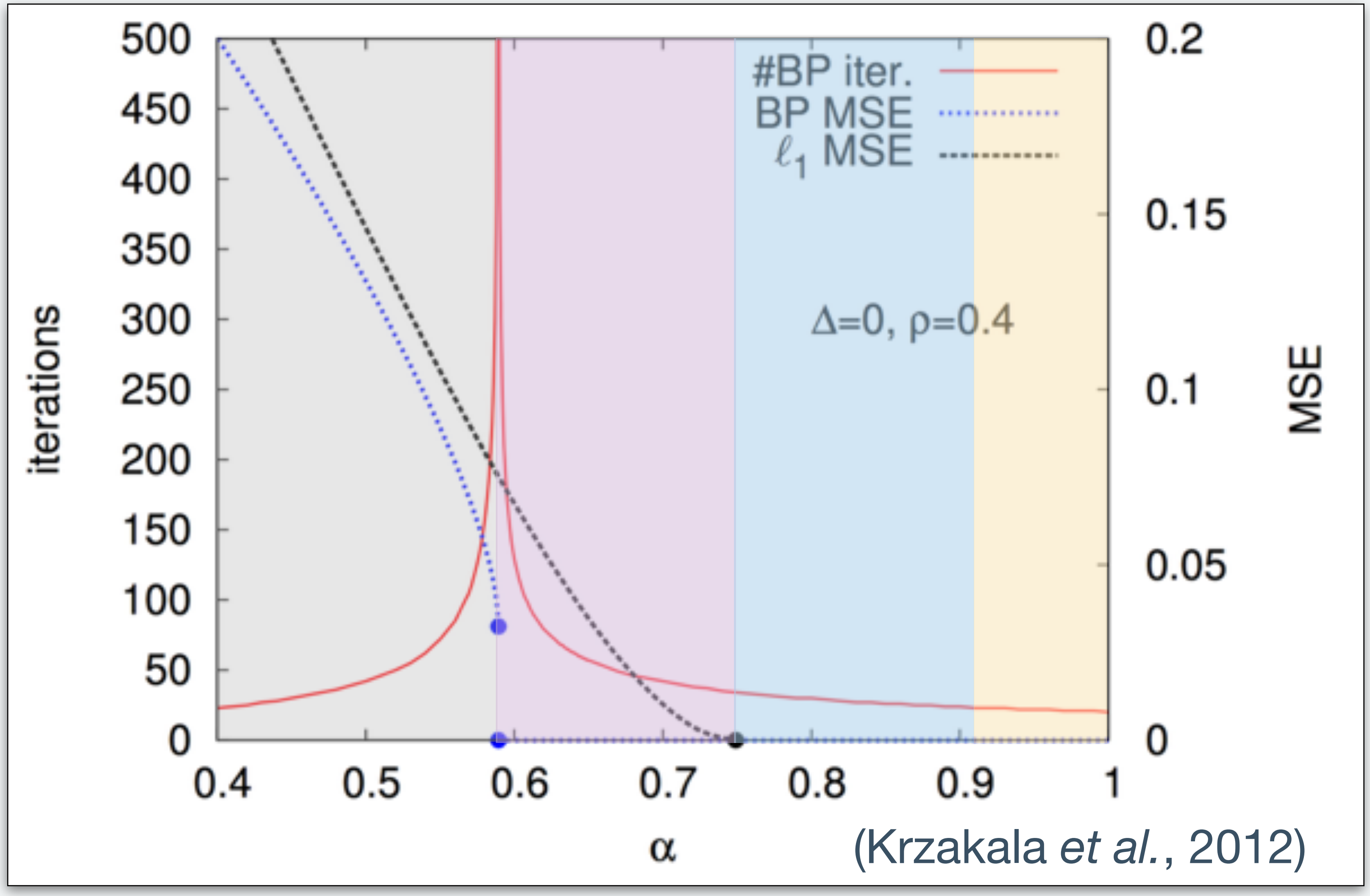
r-BP/AMP with GB Prior



r-BP/TAP Convergence

Small iteration count far from transition provides efficient estimation.

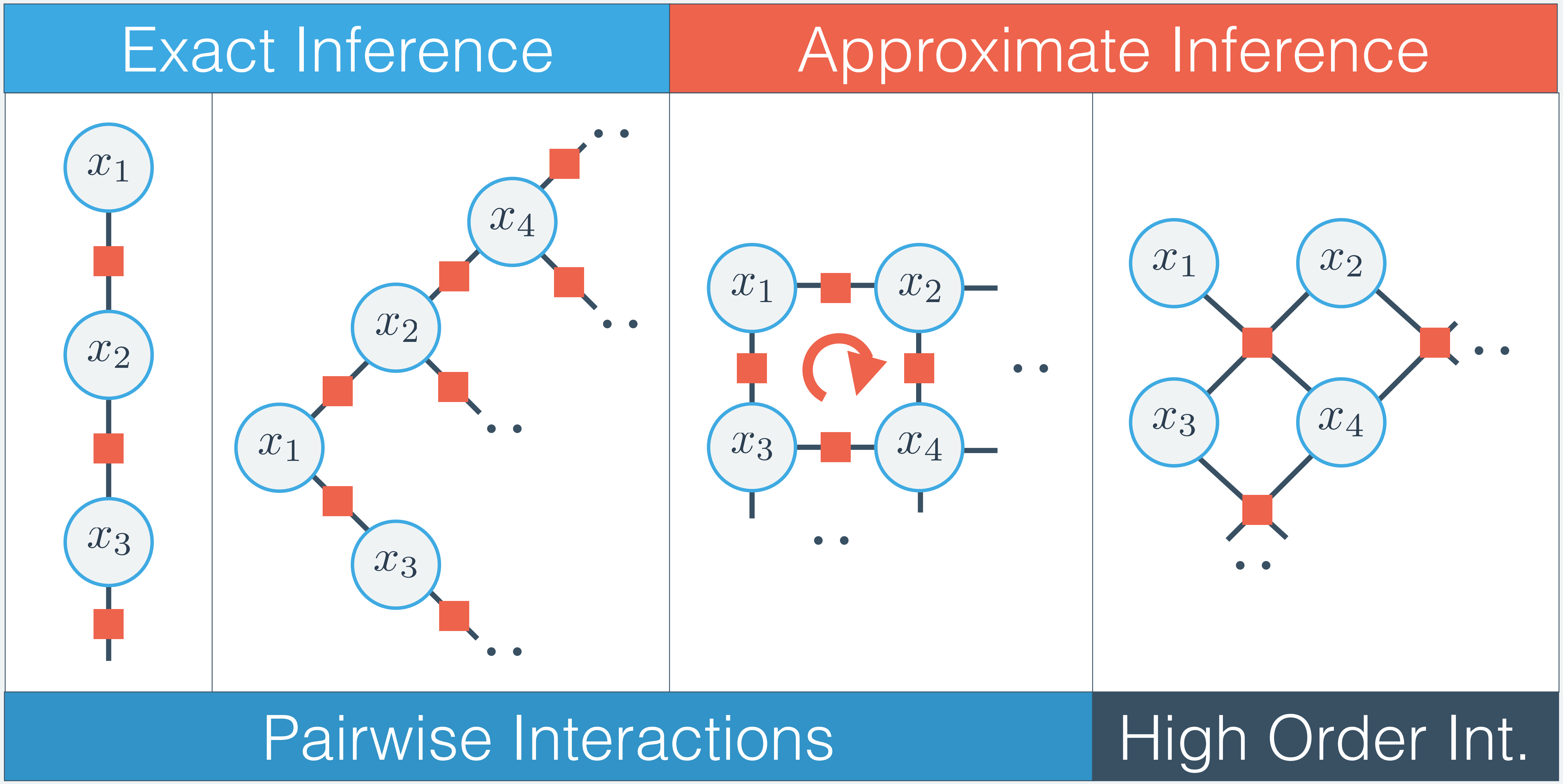




However, the approach encounters critical slowing at or near the transition, as shown via state evolution analysis.

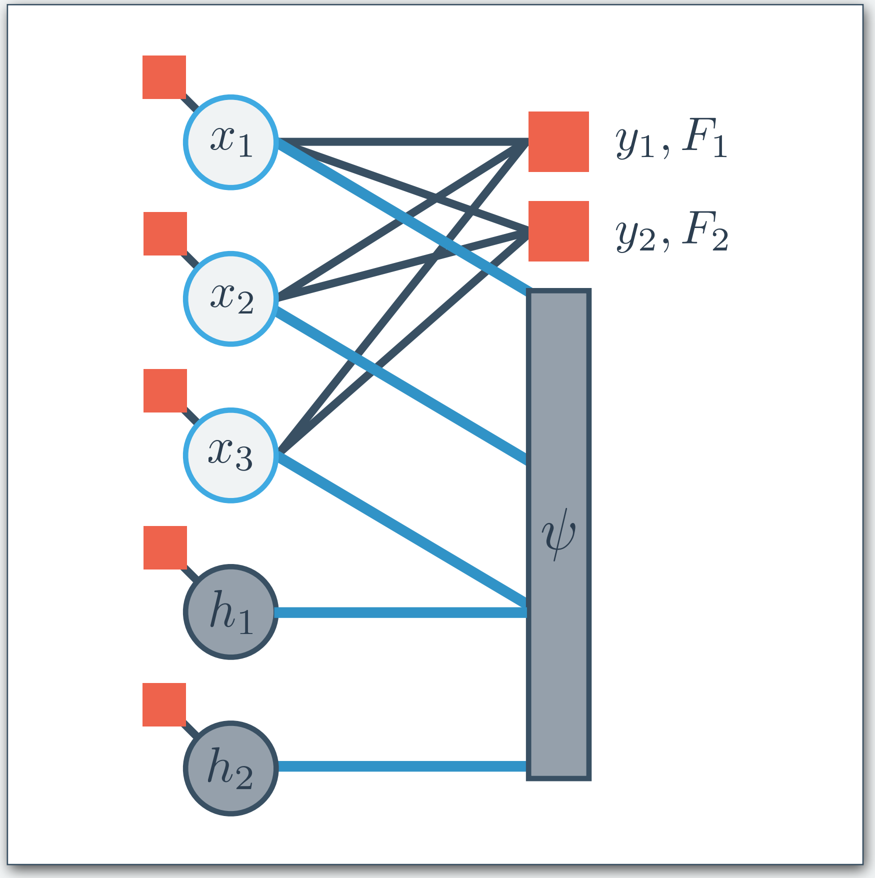
Structured Signal Priors

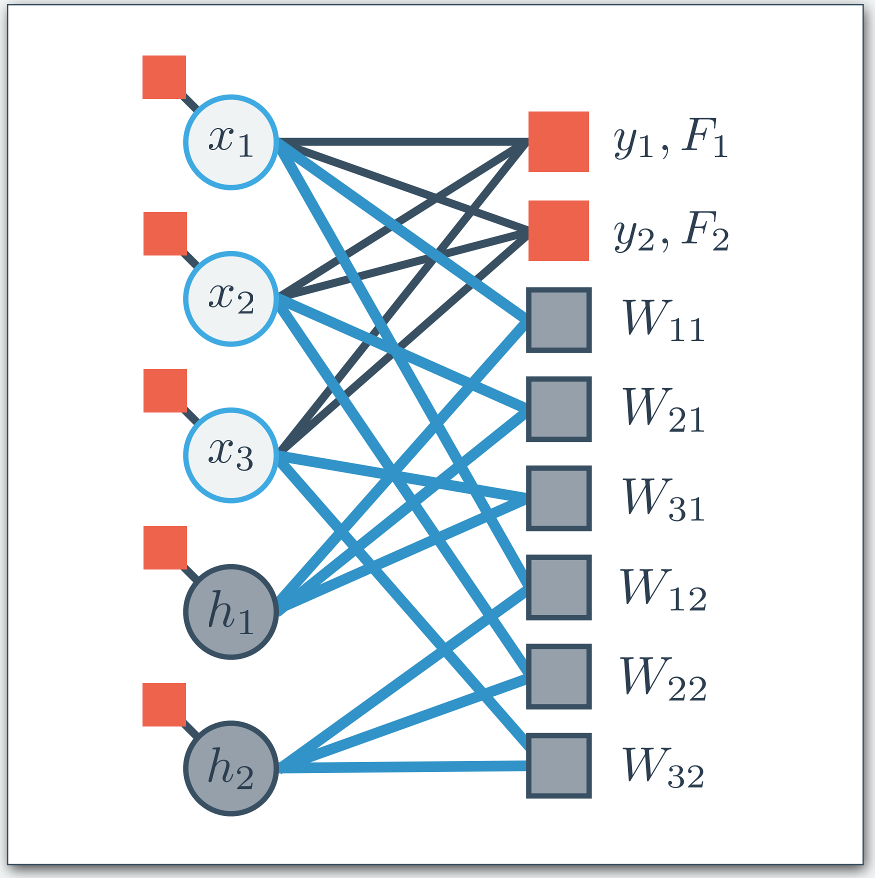
But what if we have more information about our signal? Correlations?



For completely general visible models, see (Rangan *et al*, “Hybrid-GAMP”, 2012).

Regression & Latent Variables





Binary Restricted Boltzmann Machine (RBM)

$$P_0(\mathbf{x}, \mathbf{h}) = \frac{1}{\mathcal{Z}} e^{\sum_{il} x_i W_{il} h_l + \sum_i b_i x_i + \sum_l c_l h_l}$$

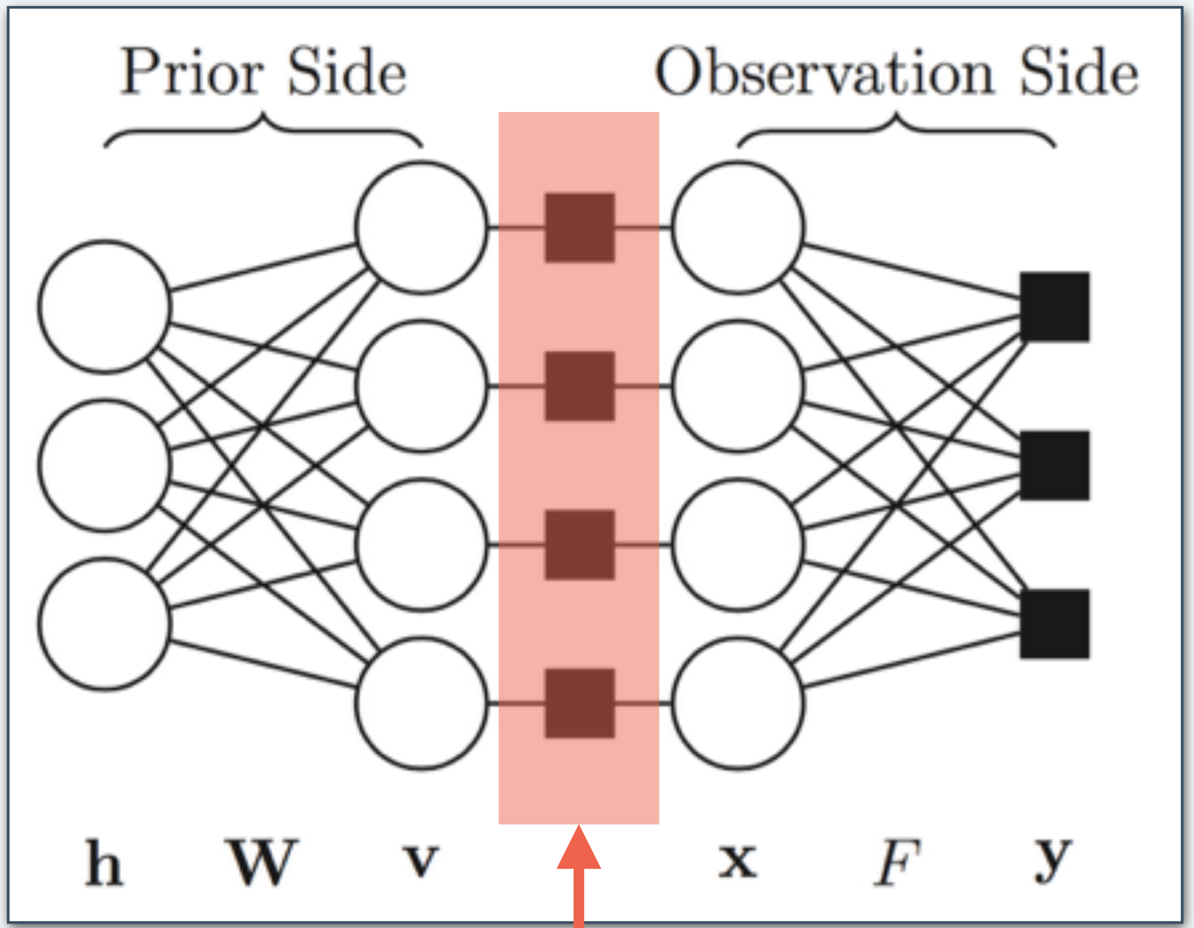
- **Latent Model:** Model data via a nonlinear composition of features.
- **Unsupervised:** Extract features from unlabelled training data.
- **Tuning:** Scale memorization/generalization via number of latent variables.
- **Training:** Sampling (Contrastive Divergence [Hinton 2002])
  Mean-field (NMF [Welling, Hinton 2002], EMF [Gabrié, Tramel, Krzakala 2015])

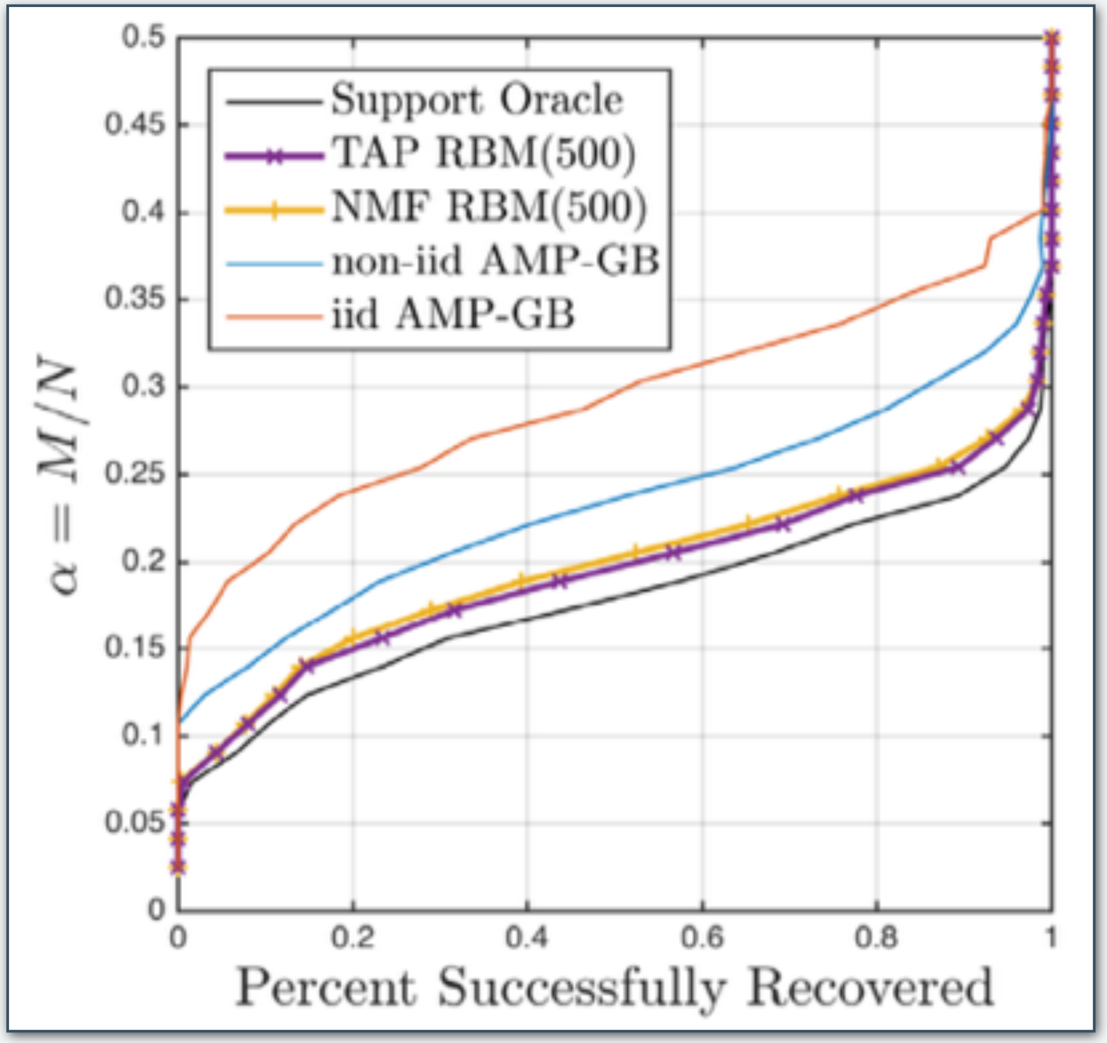
AMP with RBM Support Prior

*E.W.T, Drémeau, Krzakala, "**AMP with Boltzmann machine priors**," JSTAT 2016.*

*Gabrié, E.W.T, Krzakala, "**Training RBMs with the TAP FE**," NIPS 2015.*

- For structured sparse signals, we can train a binary RBM to model the signal *support*, subsequently use it in AMP.





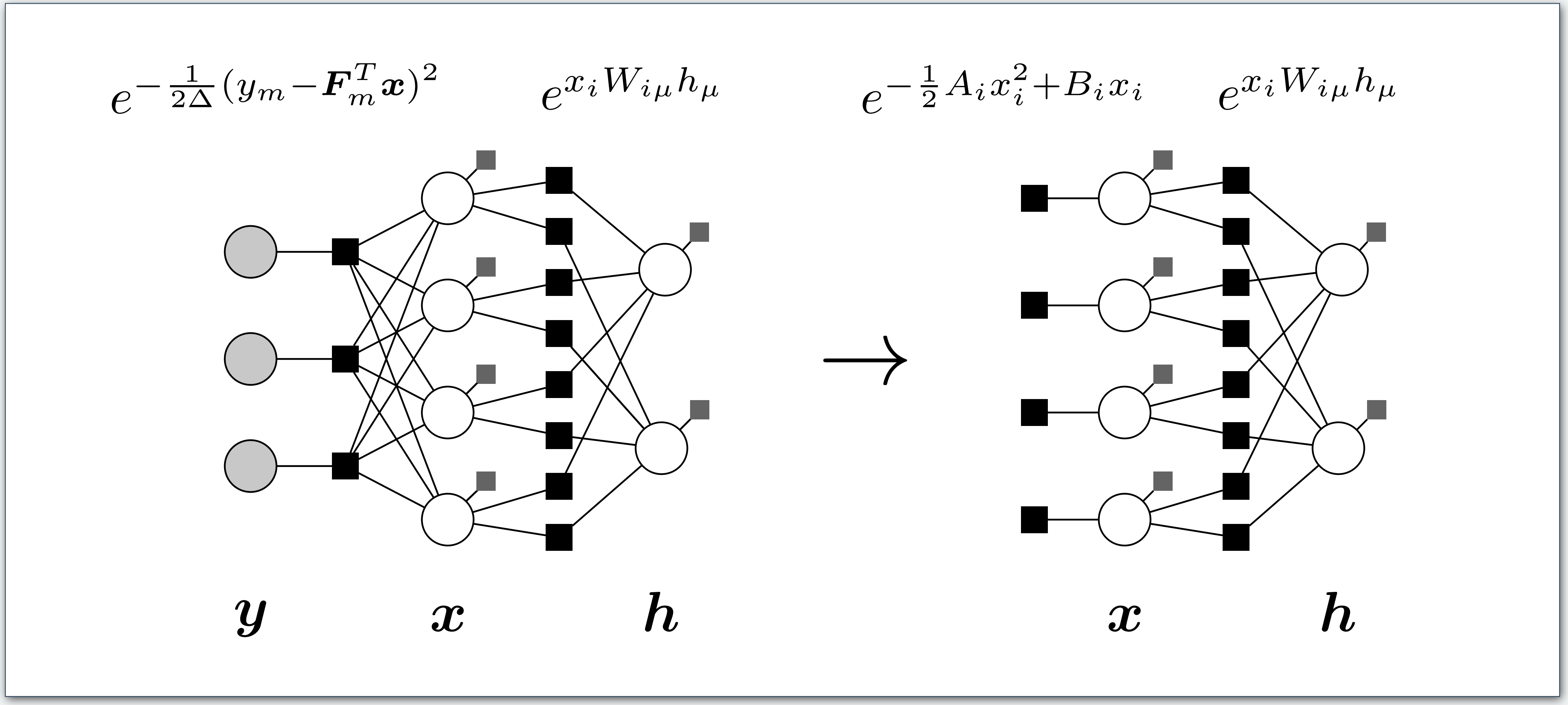
Demonstrates: The RBM and AMP interact only via local biases…inference on each essentially agnostic to the other.

AMP with General RBM Prior

A General RBM

$$P_0(\mathbf{x}, \mathbf{h}|\boldsymbol{\theta}) = \frac{1}{\mathcal{Z}} e^{\mathbf{x}^T \mathbf{W} \mathbf{h}} \prod_i P_0(x_i|\boldsymbol{\theta}_{\mathbf{x}}) \prod_l P_0(h_l|\boldsymbol{\theta}_{\mathbf{h}})$$

- Using this generalized Boltzmann prior, we can model signals *directly*, without invoking sparsity.



AMP with General RBM Prior

A General RBM

$$P_0(\mathbf{x}, \mathbf{h}|\boldsymbol{\theta}) = \frac{1}{\mathcal{Z}} e^{\mathbf{x}^T \mathbf{W} \mathbf{h}} \prod_i P_0(x_i|\boldsymbol{\theta}_{\mathbf{x}}) \prod_l P_0(h_l|\boldsymbol{\theta}_{\mathbf{h}})$$

- Exact knowledge of **P(x)** is intractable and sampling impractical for AMP.

TAP Approximation of GRBM

$$-\ln \mathcal{Z} \approx \mathbb{F}(\mathbf{a}^*, \mathbf{c}^*; \theta, W) \triangleq \sum_i \ln Z_i(B_i^*, A_i^*; \theta_i) - \sum_i B_i^* a_i^* + \frac{1}{2} \sum_i A_i^* ((a_i^*)^2 + c_i^*)$$
$$+ \sum_{(i,j)} W_{ij} a_i^* a_j^* + \frac{1}{2} \sum_{(i,j)} W_{ij}^2 c_i^* c_j^*$$

Inference: Moments at each visible variable can be approximated by fixed-point iteration.

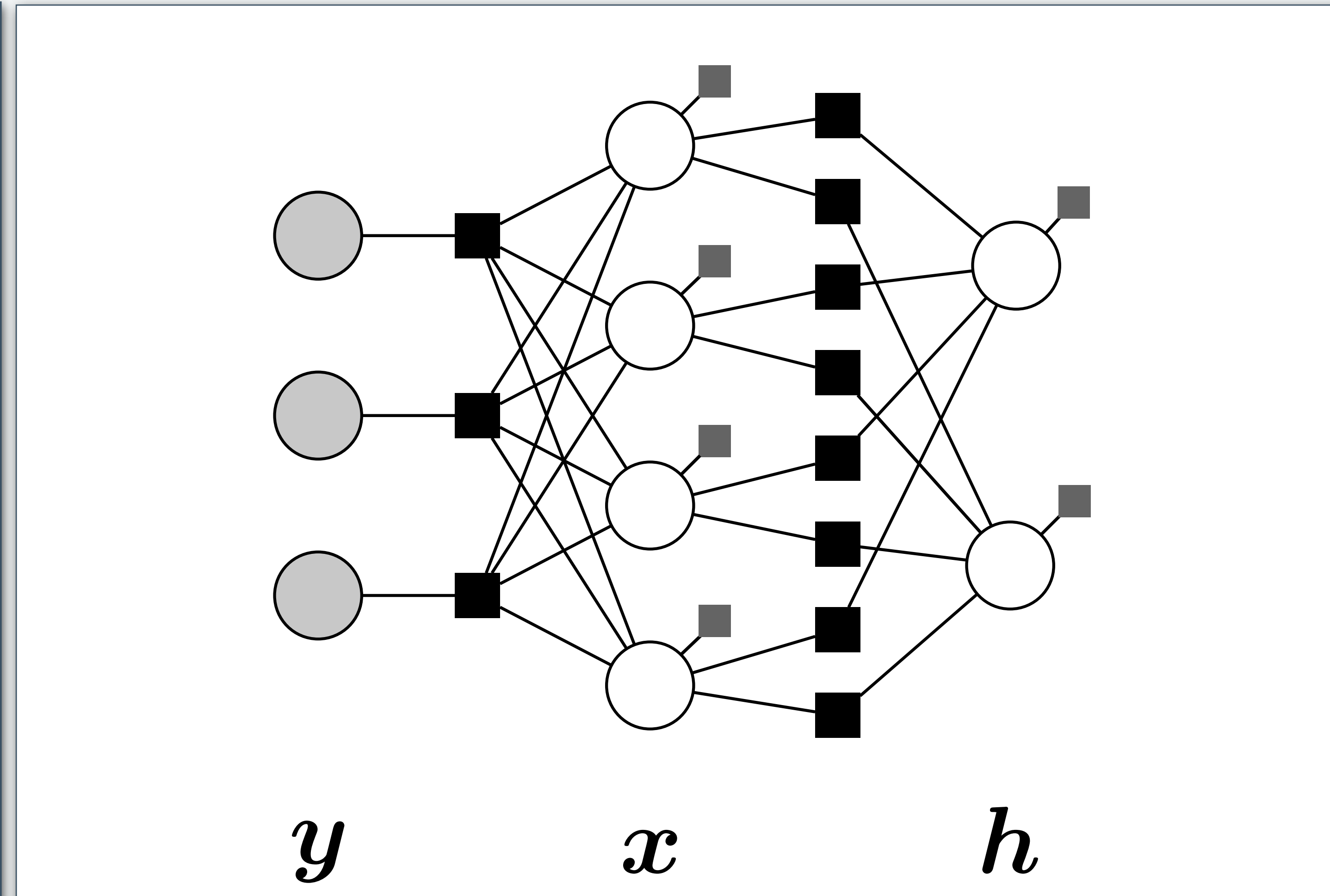
$$A_\mu^{\mathrm{h}} = -\sum_{i \in V} W_{i\mu}^2 c_i^{\mathrm{v}}, \quad B_\mu^{\mathrm{h}} = a_\mu^{\mathrm{h}} A_\mu^{\mathrm{h}} + \sum_{i \in V} W_{i\mu} a_i^{\mathrm{v}},$$
$$a_\mu^{\mathrm{h}} = f_{\mathrm{a}}^{\mathrm{h}}(A_\mu^{\mathrm{h}}, B_\mu^{\mathrm{h}}), \quad c_\mu^{\mathrm{h}} = f_{\mathrm{c}}^{\mathrm{h}}(A_\mu^{\mathrm{h}}, B_\mu^{\mathrm{h}}),$$
$$A_i^{\mathrm{v}} = -\sum_{\mu \in H} W_{i\mu}^2 c_\mu^{\mathrm{h}}, \quad B_i^{\mathrm{v}} = a_i^{\mathrm{v}} A_i^{\mathrm{v}} + \sum_{\mu \in H} W_{i\mu} a_\mu^{\mathrm{h}},$$
$$a_i = f_{\mathrm{a}}^{\mathrm{v}}(A_i^{\mathrm{AMP}} + A_i^{\mathrm{v}}, B_i^{\mathrm{AMP}} + B_i^{\mathrm{v}}),$$
$$c_i = f_{\mathrm{c}}^{\mathrm{v}}(A_i^{\mathrm{AMP}} + A_i^{\mathrm{v}}, B_i^{\mathrm{AMP}} + B_i^{\mathrm{v}}).$$

AMP with General RBM Prior

What does this algorithm look like?

Algorithm 1 AMP with GRBM Signal Prior

Input: $\mathbf{F}$, $\mathbf{y}$, $\mathbf{W}$, θ^{v}, θ^{h}
Initialize: $\mathbf{a}$,$\mathbf{c}$, $\mathrm{t} = 1$
repeat
  AMP Update on $\{V_m, \omega_m\}$, as in [12]
  AMP Update on $\{R_i, \Sigma_i^2\}$, as in [12]
  Set $A_i^{\mathrm{AMP}} = 1/\Sigma_i^2$, $B_i^{\mathrm{AMP}} = R_i/\Sigma_i^2 \; \forall i$
  (Re)Initialize: $a_i = f_{\mathrm{a}}^{\mathrm{v}}(A_i^{\mathrm{AMP}}, B_i^{\mathrm{AMP}}) \; \forall i$, $a_\mu^{\mathrm{h}} = c_\mu^{\mathrm{h}} = 0 \; \forall \mu$
  repeat
    Update $\{A_i^{\mathrm{v}}, B_i^{\mathrm{v}}\}$ as in (17)
    Update $\{a_i, c_i\}$ as in (18), (19)
    Update $\{A_\mu^{\mathrm{h}}, B_\mu^{\mathrm{h}}\}$ as in (15)
    Update $\{a_\mu^{\mathrm{h}}, c_\mu^{\mathrm{h}}\}$ as in (16)
  until Convergence
  $\mathbf{a}^{(\mathrm{t})} = \gamma \cdot \mathbf{a}^{(\mathrm{t}-1)} + (1-\gamma) \cdot \mathbf{a}$
  $\mathbf{c}^{(\mathrm{t})} = \gamma \cdot \mathbf{c}^{(\mathrm{t}-1)} + (1-\gamma) \cdot \mathbf{c}$
  $\mathrm{t} \leftarrow \mathrm{t} + 1$
until Convergence on $\mathbf{a}$

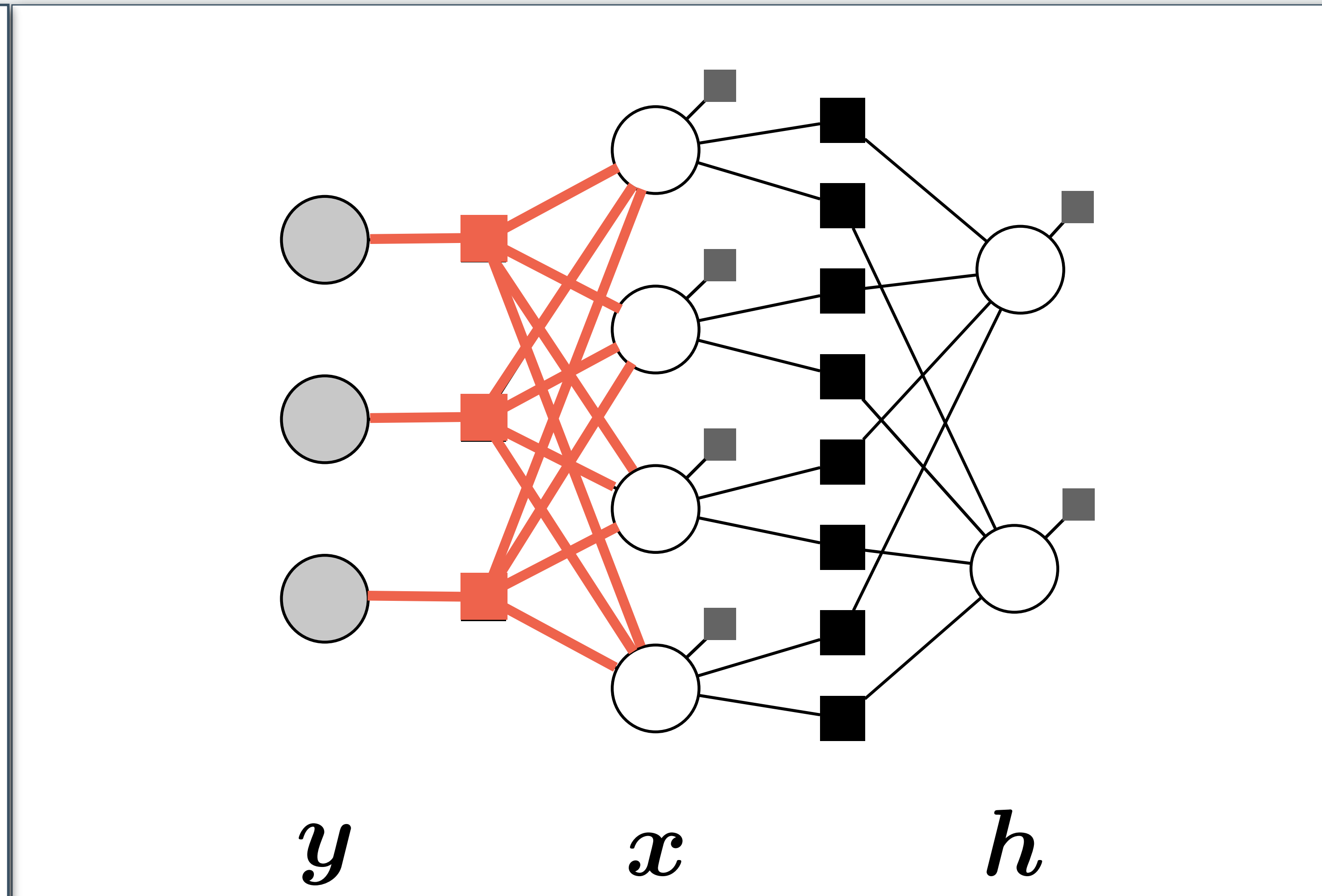


AMP with General RBM Prior

1. Update information from observation factors given current state.

Algorithm 1 AMP with GRBM Signal Prior

Input: $\mathbf{F}$, $\mathbf{y}$, $\mathbf{W}$, θ^{v}, θ^{h}
Initialize: $\mathbf{a}$,$\mathbf{c}$, $\mathrm{t} = 1$
repeat
  AMP Update on $\{V_m, \omega_m\}$, as in [12]
  AMP Update on $\{R_i, \Sigma_i^2\}$, as in [12]
  Set $A_i^{\mathrm{AMP}} = 1/\Sigma_i^2$, $B_i^{\mathrm{AMP}} = R_i/\Sigma_i^2 \ \forall i$
  (Re)Initialize: $a_i = f_{\mathrm{a}}^{\mathrm{v}}(A_i^{\mathrm{AMP}}, B_i^{\mathrm{AMP}}) \ \forall i$, $a_\mu^{\mathrm{h}} = c_\mu^{\mathrm{h}} = 0 \ \forall \mu$
  repeat
    Update $\{A_i^{\mathrm{v}}, B_i^{\mathrm{v}}\}$ as in (17)
    Update $\{a_i, c_i\}$ as in (18), (19)
    Update $\{A_\mu^{\mathrm{h}}, B_\mu^{\mathrm{h}}\}$ as in (15)
    Update $\{a_\mu^{\mathrm{h}}, c_\mu^{\mathrm{h}}\}$ as in (16)
  until Convergence
  $\mathbf{a}^{(\mathrm{t})} = \gamma \cdot \mathbf{a}^{(\mathrm{t}-1)} + (1-\gamma) \cdot \mathbf{a}$
  $\mathbf{c}^{(\mathrm{t})} = \gamma \cdot \mathbf{c}^{(\mathrm{t}-1)} + (1-\gamma) \cdot \mathbf{c}$
  $\mathrm{t} \leftarrow \mathrm{t} + 1$
until Convergence on $\mathbf{a}$

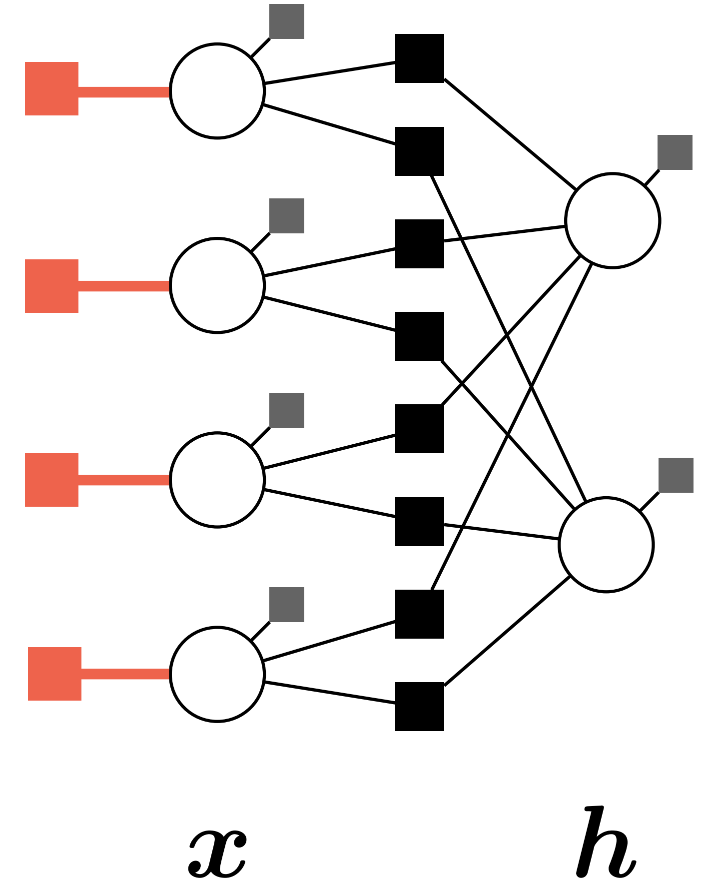


AMP with General RBM Prior

2. Calculate local fields from AMP to use as a bias during GRBM inference.

Algorithm 1 AMP with GRBM Signal Prior

Input: $\mathbf{F}$, $\mathbf{y}$, $\mathbf{W}$, θ^{v}, θ^{h}
Initialize: $\mathbf{a}$,$\mathbf{c}$, $\mathrm{t} = 1$
repeat
  AMP Update on $\{V_m, \omega_m\}$, as in [12]
  AMP Update on $\{R_i, \Sigma_i^2\}$, as in [12]
  Set $A_i^{\mathrm{AMP}} = 1/\Sigma_i^2$, $B_i^{\mathrm{AMP}} = R_i/\Sigma_i^2\ \forall i$
  (Re)Initialize: $a_i = f_{\mathrm{a}}^{\mathrm{v}}(A_i^{\mathrm{AMP}}, B_i^{\mathrm{AMP}})\ \forall i$, $a_\mu^{\mathrm{h}} = c_\mu^{\mathrm{h}} = 0\ \forall \mu$
  repeat
    Update $\{A_i^{\mathrm{v}}, B_i^{\mathrm{v}}\}$ as in (17)
    Update $\{a_i, c_i\}$ as in (18), (19)
    Update $\{A_\mu^{\mathrm{h}}, B_\mu^{\mathrm{h}}\}$ as in (15)
    Update $\{a_\mu^{\mathrm{h}}, c_\mu^{\mathrm{h}}\}$ as in (16)
  until Convergence
  $\mathbf{a}^{(\mathrm{t})} = \gamma \cdot \mathbf{a}^{(\mathrm{t}-1)} + (1-\gamma)\cdot \mathbf{a}$
  $\mathbf{c}^{(\mathrm{t})} = \gamma \cdot \mathbf{c}^{(\mathrm{t}-1)} + (1-\gamma)\cdot \mathbf{c}$
  $\mathrm{t} \leftarrow \mathrm{t} + 1$
until Convergence on $\mathbf{a}$

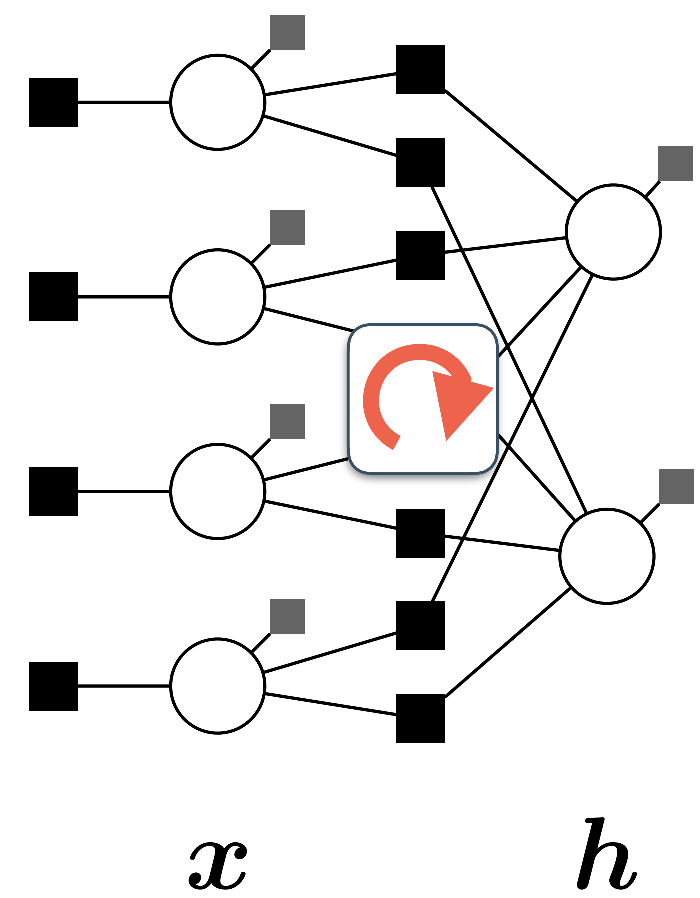


AMP with General RBM Prior

3. Run GRBM inference in order to obtain marginal estimate of moments at each signal coefficient.

Algorithm 1 AMP with GRBM Signal Prior

Input: $\mathbf{F}$, $\mathbf{y}$, $\mathbf{W}$, θ^{v}, θ^{h}
Initialize: $\mathbf{a}$,$\mathbf{c}$, $\mathrm{t} = 1$
repeat
  AMP Update on $\{V_m, \omega_m\}$, as in [12]
  AMP Update on $\{R_i, \Sigma_i^2\}$, as in [12]
  Set $A_i^{\mathrm{AMP}} = 1/\Sigma_i^2$, $B_i^{\mathrm{AMP}} = R_i/\Sigma_i^2 \; \forall i$
  (Re)Initialize: $a_i = f_{\mathrm{a}}^{\mathrm{v}}(A_i^{\mathrm{AMP}}, B_i^{\mathrm{AMP}}) \; \forall i$, $a_\mu^{\mathrm{h}} = c_\mu^{\mathrm{h}} = 0 \; \forall \mu$
  repeat
    Update $\{A_i^{\mathrm{v}}, B_i^{\mathrm{v}}\}$ as in (17)
    Update $\{a_i, c_i\}$ as in (18), (19)
    Update $\{A_\mu^{\mathrm{h}}, B_\mu^{\mathrm{h}}\}$ as in (15)
    Update $\{a_\mu^{\mathrm{h}}, c_\mu^{\mathrm{h}}\}$ as in (16)
  until Convergence
  $\mathbf{a}^{(\mathrm{t})} = \gamma \cdot \mathbf{a}^{(\mathrm{t}-1)} + (1-\gamma) \cdot \mathbf{a}$
  $\mathbf{c}^{(\mathrm{t})} = \gamma \cdot \mathbf{c}^{(\mathrm{t}-1)} + (1-\gamma) \cdot \mathbf{c}$
  $\mathrm{t} \leftarrow \mathrm{t} + 1$
until Convergence on $\mathbf{a}$

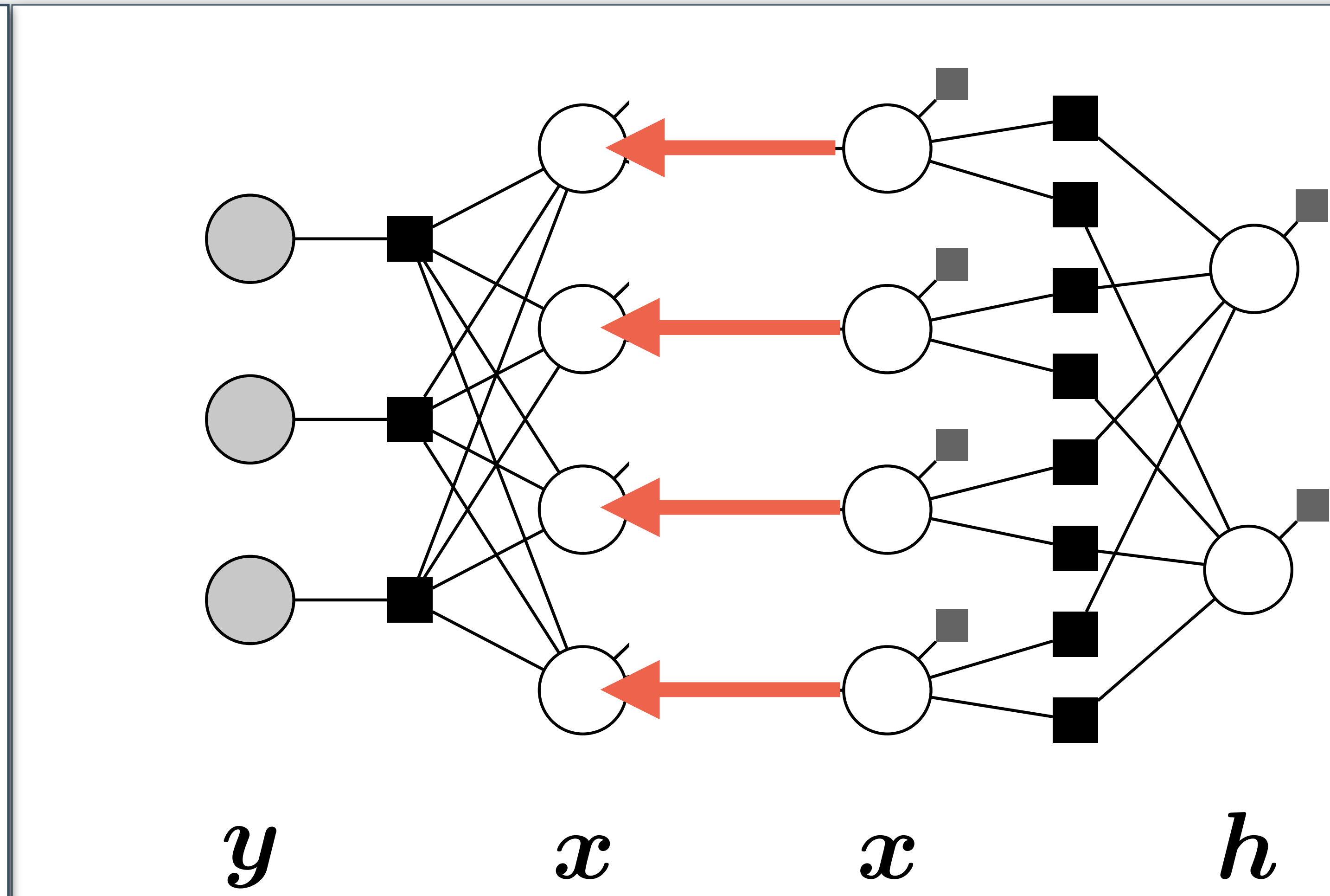


AMP with General RBM Prior

4. Apply light damping in order to avoid thrashing between GRBM modes and repeat.

Algorithm 1 AMP with GRBM Signal Prior

Input: $\mathbf{F}$, $\mathbf{y}$, $\mathbf{W}$, θ^{v}, θ^{h}
Initialize: $\mathbf{a}$,$\mathbf{c}$, $\mathrm{t}=1$
repeat
  AMP Update on $\{V_m, \omega_m\}$, as in [12]
  AMP Update on $\{R_i, \Sigma_i^2\}$, as in [12]
  Set $A_i^{\mathrm{AMP}} = 1/\Sigma_i^2$, $B_i^{\mathrm{AMP}} = R_i/\Sigma_i^2\ \forall i$
  (Re)Initialize: $a_i = f_a^{\mathrm{v}}(A_i^{\mathrm{AMP}}, B_i^{\mathrm{AMP}})\ \forall i$, $a_\mu^{\mathrm{h}} = c_\mu^{\mathrm{h}} = 0\ \forall \mu$
  repeat
    Update $\{A_i^{\mathrm{v}}, B_i^{\mathrm{v}}\}$ as in (17)
    Update $\{a_i, c_i\}$ as in (18), (19)
    Update $\{A_\mu^{\mathrm{h}}, B_\mu^{\mathrm{h}}\}$ as in (15)
    Update $\{a_\mu^{\mathrm{h}}, c_\mu^{\mathrm{h}}\}$ as in (16)
  until Convergence
  $\mathbf{a}^{(\mathrm{t})} = \gamma \cdot \mathbf{a}^{(\mathrm{t}-1)} + (1-\gamma)\cdot \mathbf{a}$
  $\mathbf{c}^{(\mathrm{t})} = \gamma \cdot \mathbf{c}^{(\mathrm{t}-1)} + (1-\gamma)\cdot \mathbf{c}$
  $\mathrm{t} \leftarrow \mathrm{t}+1$
until Convergence on $\mathbf{a}$



Extra Cost: Proportional to the number of interior steps you take, though can be set < 10.

Experimental Framework

Offline Training

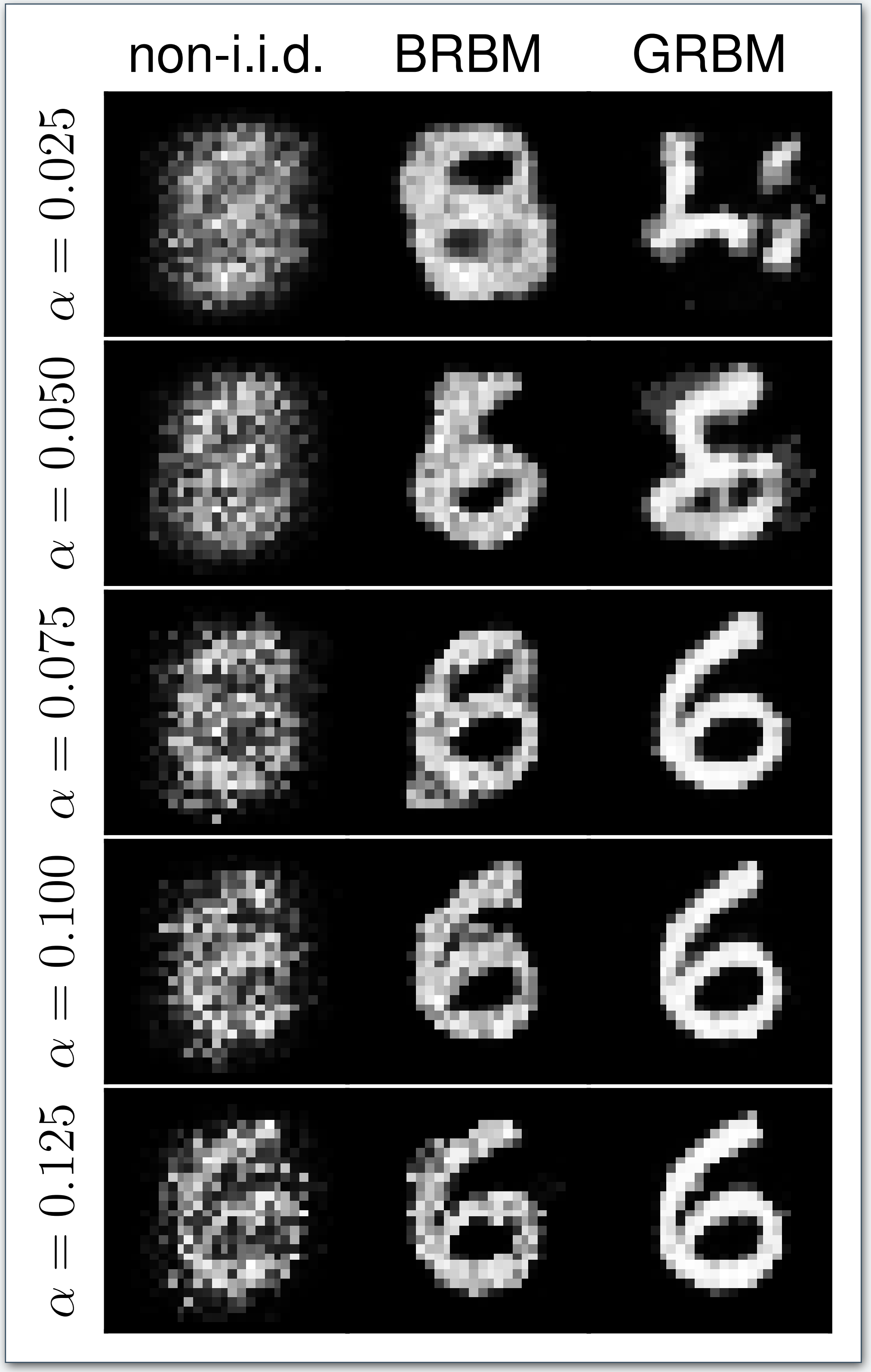
- 60k real-valued MNIST training samples
- 784 (28x28) Trunc. Gauss-Bernoulli Visible
- 500 Binary Hidden
- 100 sample mini-batches
- 150 Epochs (90k parameter updates)

Reconstruction

- 1k real-valued MNIST test samples (h.o.)
- Noise Variance: 10^{-8}
- IID Random Projection Matrix **F**

Methods

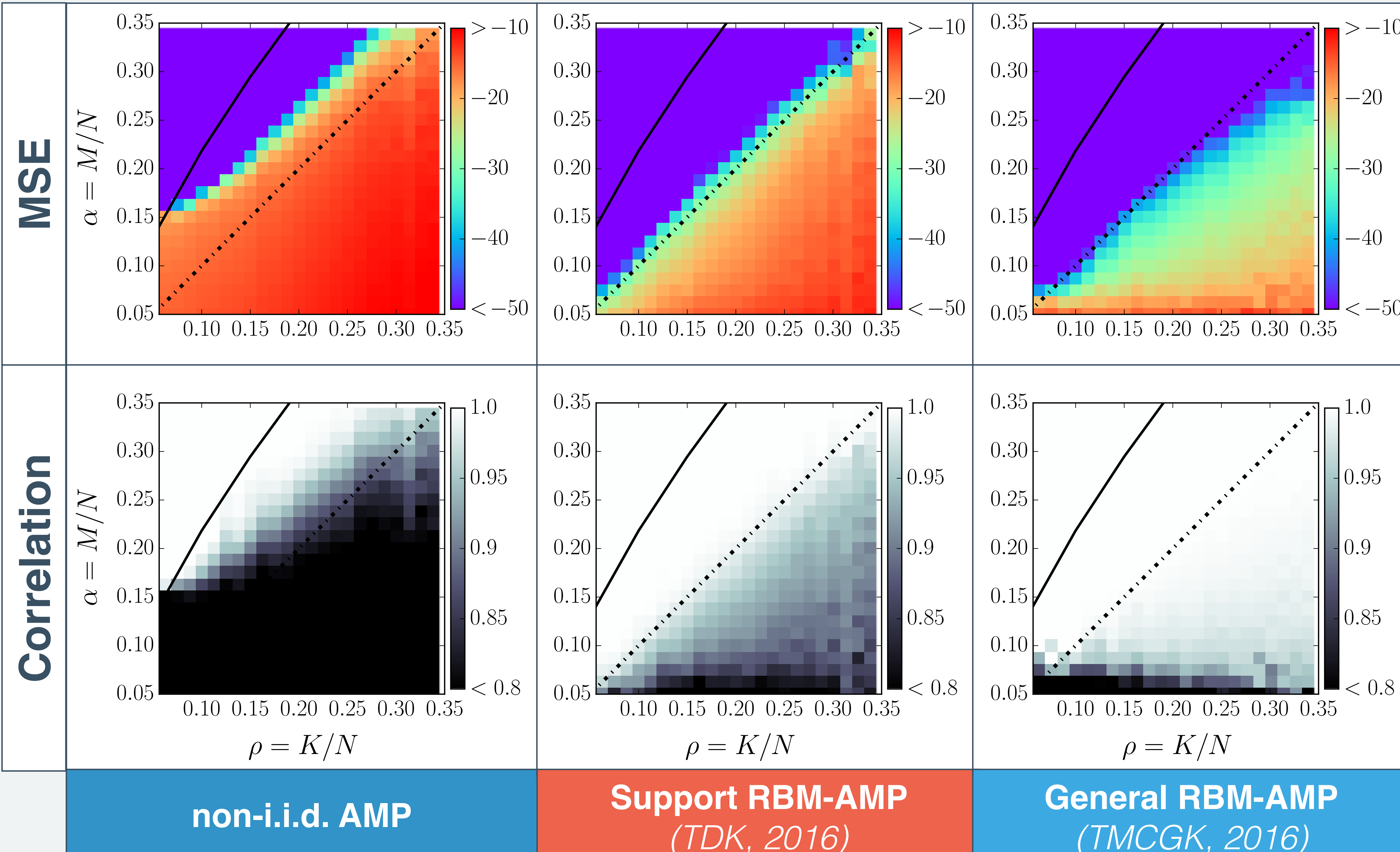
- ***non-.i.i.d.*** — Use empirical support probability with GB prior AMP.
- ***BRBM*** — Binary RBM to model support location.
- ***GRBM*** — Generalized RBM to model entire signal.



(Tramel *et al.*, 2016)

Measure — Reconstruction quality as a function of number of measurements.

AMP with General RBM Prior



Learning GRBMs — Open Work

1. **What can be gained with deep architectures?**
   (a) Can we push the boundaries even further with DBMs?
2. **Is there an upper limit?**
   (a) What can we learn from density evolution in random case?
3. **Further experiments on non-sparse data.**
   (a) Can an RBM be trained to a sufficient level to allow for CS-like reconstruction of non-sparse structured signals?
4. **Time Indexing.**
   (a) Parallel time indexing for BP on pairwise graphical models not set in stone…
   (b) Need to “re-derive” time indices via pairwise r-BP.
5. **Comparison with r-BP learning.**
6. **The Influence of Hidden Unit Distribution**
   (a) E.g. Gauss-Bernoulli hidden units to allow for modeling of visible covariance structure (ala SS-RBM).

SPHINX @ENS

Statistical **PH**ysics of **IN**formation e**X**traction

«OU»

Statistical **PH**ysics of **IN**verse comple**X** sysems

ENS

Questions?

Merci!

Collaborators

Andre MANOEL	*LPS-ENS*
Francesco CALTAGIRONE	*INRIA Paris*
Marylou GABRIE	*LPS-ENS*
Florent KRZAKALA	*LPS-ENS*

Supplements

Learning GRBMs

How do we train the necessary GRBM models?

Iterating the same equations allows us to evaluate the TAP f.e. approximation of the GRBM.

=> Stochastic Gradient Ascent on dataset likelihood.

$$A_\mu^{\mathrm{h}} = -\sum_{i\in V} W_{i\mu}^2 c_i^{\mathrm{v}}, \quad B_\mu^{\mathrm{h}} = a_\mu^{\mathrm{h}} A_\mu^{\mathrm{h}} + \sum_{i\in V} W_{i\mu} a_i^{\mathrm{v}},$$
$$a_\mu^{\mathrm{h}} = f_{\mathrm{a}}^{\mathrm{h}}(A_\mu^{\mathrm{h}}, B_\mu^{\mathrm{h}}), \quad c_\mu^{\mathrm{h}} = f_{\mathrm{c}}^{\mathrm{h}}(A_\mu^{\mathrm{h}}, B_\mu^{\mathrm{h}}),$$
$$A_i^{\mathrm{v}} = -\sum_{\mu\in H} W_{i\mu}^2 c_\mu^{\mathrm{h}}, \quad B_i^{\mathrm{v}} = a_i^{\mathrm{v}} A_i^{\mathrm{v}} + \sum_{\mu\in H} W_{i\mu} a_\mu^{\mathrm{h}},$$
$$a_i = f_{\mathrm{a}}^{\mathrm{v}}(A_i^{\mathrm{AMP}} + A_i^{\mathrm{v}}, B_i^{\mathrm{AMP}} + B_i^{\mathrm{v}}),$$
$$c_i = f_{\mathrm{c}}^{\mathrm{v}}(A_i^{\mathrm{AMP}} + A_i^{\mathrm{v}}, B_i^{\mathrm{AMP}} + B_i^{\mathrm{v}}).$$

Biggest Issue…

Inescapable negative variances! Signed definition required to make messages Gaussian, but now it bites us. What to do?

$$A_\mu^{\mathrm{h}} = -\sum_{i\in V} W_{i\mu}^2 c_i^{\mathrm{v}},$$

$$A_i^{\mathrm{v}} = -\sum_{\mu\in H} W_{i\mu}^2 c_\mu^{\mathrm{h}},$$

Current sol'n: Truncated Distributions, allow neg. vars!

Learning GRBMs

For a Gaussian distributed unit…

$$a = \frac{B+U}{A+V} - \sqrt{\frac{2}{\pi|A+V|}} \cdot \begin{cases} \frac{e^{-\phi_\omega^2} - e^{-\phi_\alpha^2}}{\mathrm{Erf}[\phi_\omega] - \mathrm{Erf}[\phi_\alpha]}, & \text{for} \quad A+V>0 \\ \frac{e^{\phi_\omega^2} - e^{\phi_\alpha^2}}{\mathrm{Erfi}[\phi_\omega] - \mathrm{Erfi}[\phi_\alpha]}, & \text{for} \quad A+V<0 \end{cases} .$$

$$\begin{aligned} \langle x^2 \rangle_{Q(x)} = {} & \frac{1}{A+V} + \frac{(B+U)^2}{(A+V)^2} \\ & + \sqrt{\frac{2}{\pi(A+V)}} \cdot \frac{(\omega + \frac{B+U}{A+V}) \exp^{-\frac{A+V}{2} \cdot (\omega - \frac{B+U}{A+V})^2} - (\alpha + \frac{B+U}{A+V}) \exp^{-\frac{A+V}{2} \cdot (\alpha - \frac{B+U}{A+V})^2}}{\mathrm{Erf}\left[\sqrt{\frac{A+V}{2}}(\omega - \frac{B+U}{A+V})\right] - \mathrm{Erf}\left[\sqrt{\frac{A+V}{2}}(\alpha - \frac{B+U}{A+V})\right]}, \end{aligned} \tag{37}$$

The \`ERF - ERF\` in the denominators make for terrible numerical issues for sufficiently likely arguments.

- We “solved” via high-order Taylor approximation.

Not themes elegant solution…

Relaxed BP (r-BP)

Problems

- Analytically intractable messages.
- Messages are continuous objects (PDFs), not fit for a computable algorithm.

Assumption

- All values of **F** scale as **O(1/N)**.

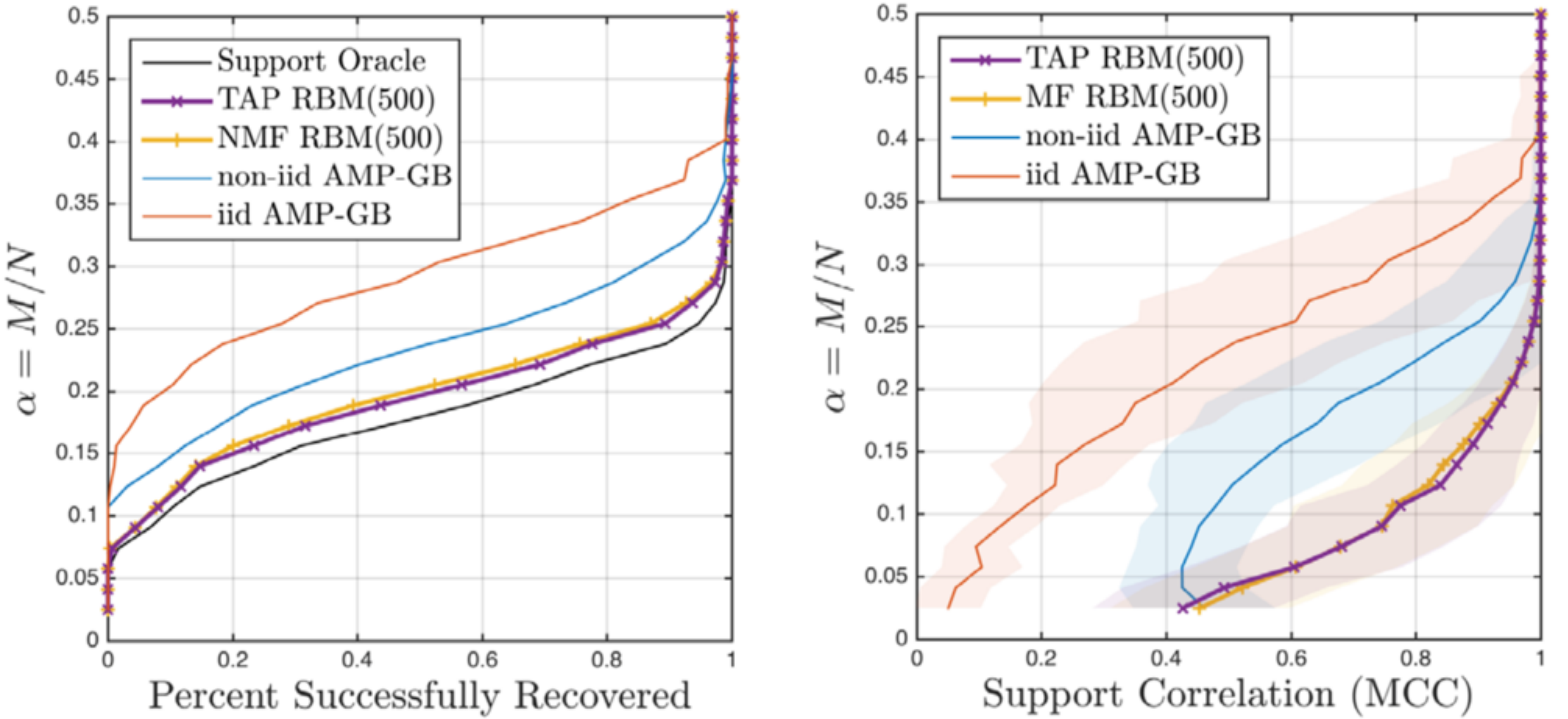
Remedy (Rangan, 2010), (Krzakala *et al.*, 2012)

- Given the above, we can perform a small-weight expansion on the messages, allowing for all messages to be written as **Normal Distributions** via the CLT, and are parameterized by…

$$a_{i\to\mu} \triangleq \int \mathrm{d}x_i \;\; x_i m_{i\to\mu}(x_i), \;\; v_{i\to\mu} \triangleq -(a_{i\to\mu})^2 + \int \mathrm{d}x_i \;\; x_i^2 m_{i\to\mu}(x_i)$$

AMP with RBM Support Prior

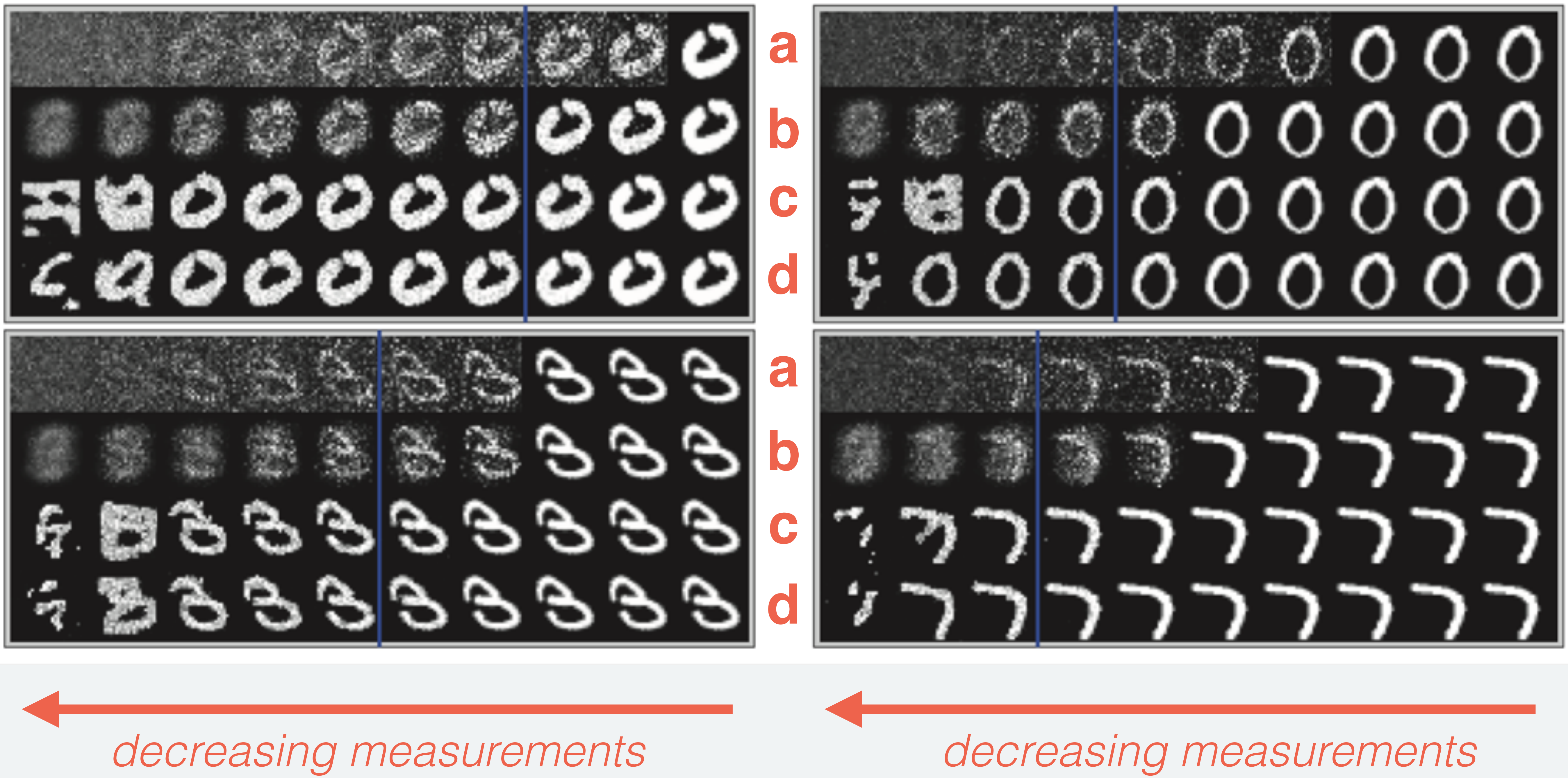
(Tramel, Drémeau, Krzakala, 2016)



MNIST Experiments — Goal: CS reconstruction of 300 test set digits given training set of 60,000 samples.

AMP with RBM Support Prior

(Tramel, Drémeau, Krzakala, 2016)



(a) i.i.d. GB-AMP, (b) non-i.i.d. GB-AMP,
(c) naive mean-field RBM-AMP
(d) TAP mean-field RBM-AMP