Belief Propagation & Approximations

Discrete Tomography

Eric W. Tramel

19 March 2015

Contributors

Florent KRZAKALA
ENS, Univ. P. & M. Curie

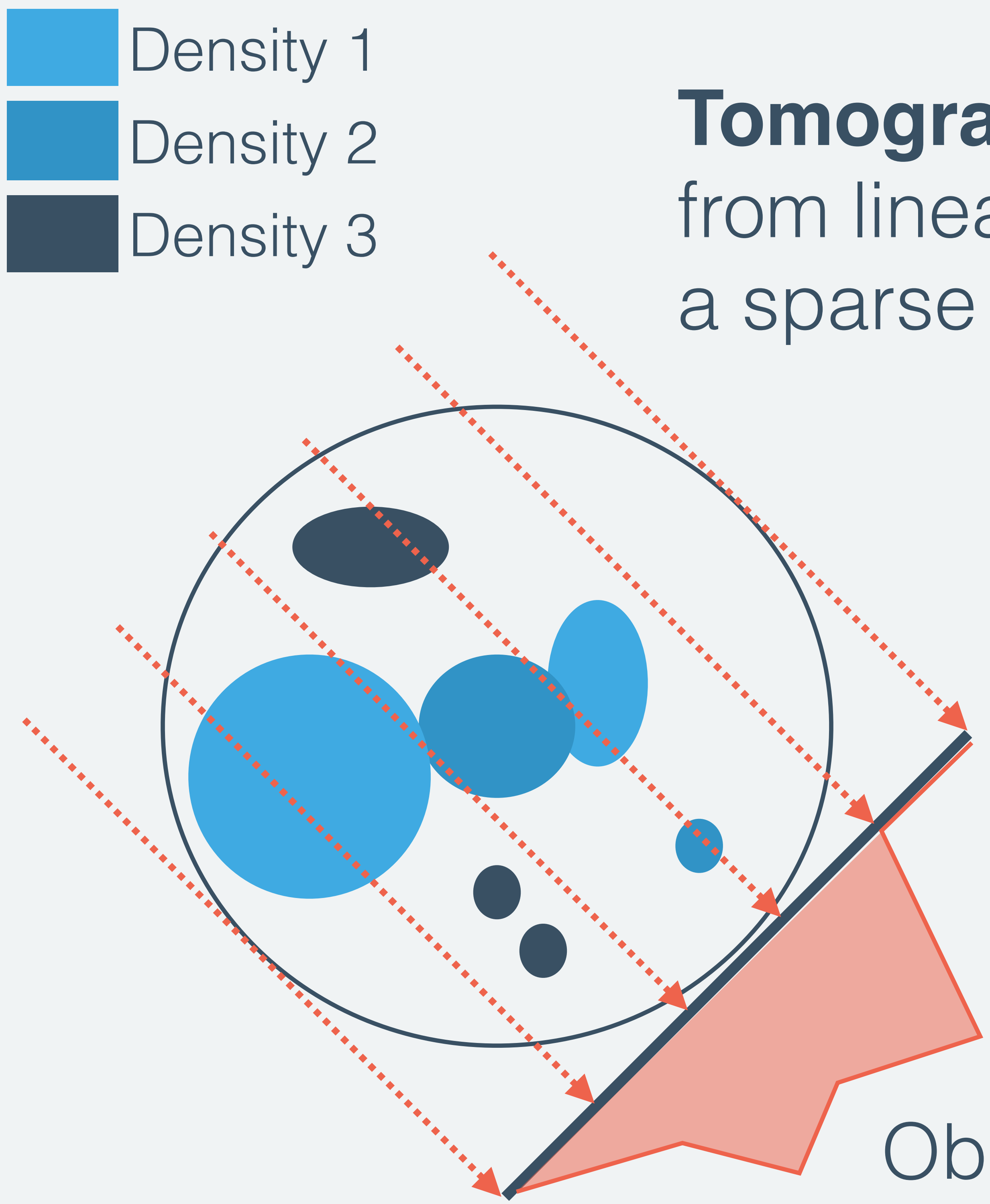
Eric W. TRAMEL
ENS

Federico RICCI-TERSENGHI
Univ. Rome *La Sapienza*

Ovidiu ERSEN
Univ. Strasbourg

Simona MOLDOVAN
Univ. Strasbourg

Discrete Tomography



Tomography Essentially a reconstruction from linear measurements obtained from a sparse set of projections.

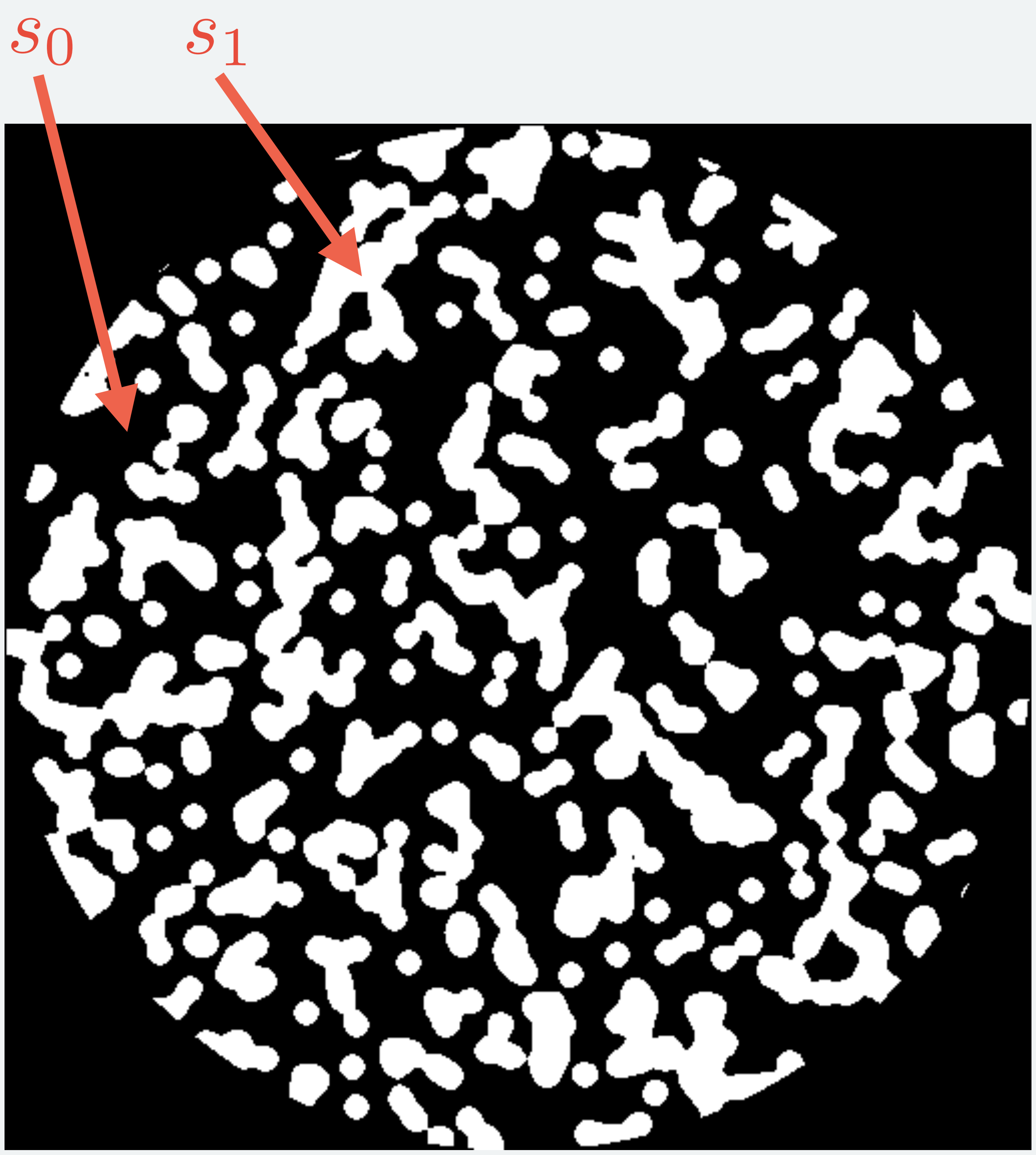
$$\mathbf{y} = F\mathbf{x} + \mathbf{w}$$ *possible noise*

Application
Material Sciences > Biology/Medic.

Observation at angle θ

$$\mathbf{y}_\theta = \langle F_\theta, \mathbf{x} \rangle$$

Binary Tomography



(512x512) Binary Phantom

Binary Simplest case, two possible absorption levels,

$$x_i \in \{s_0, s_1\}$$

for ease, map signal to {0,1} and adjust measurements,

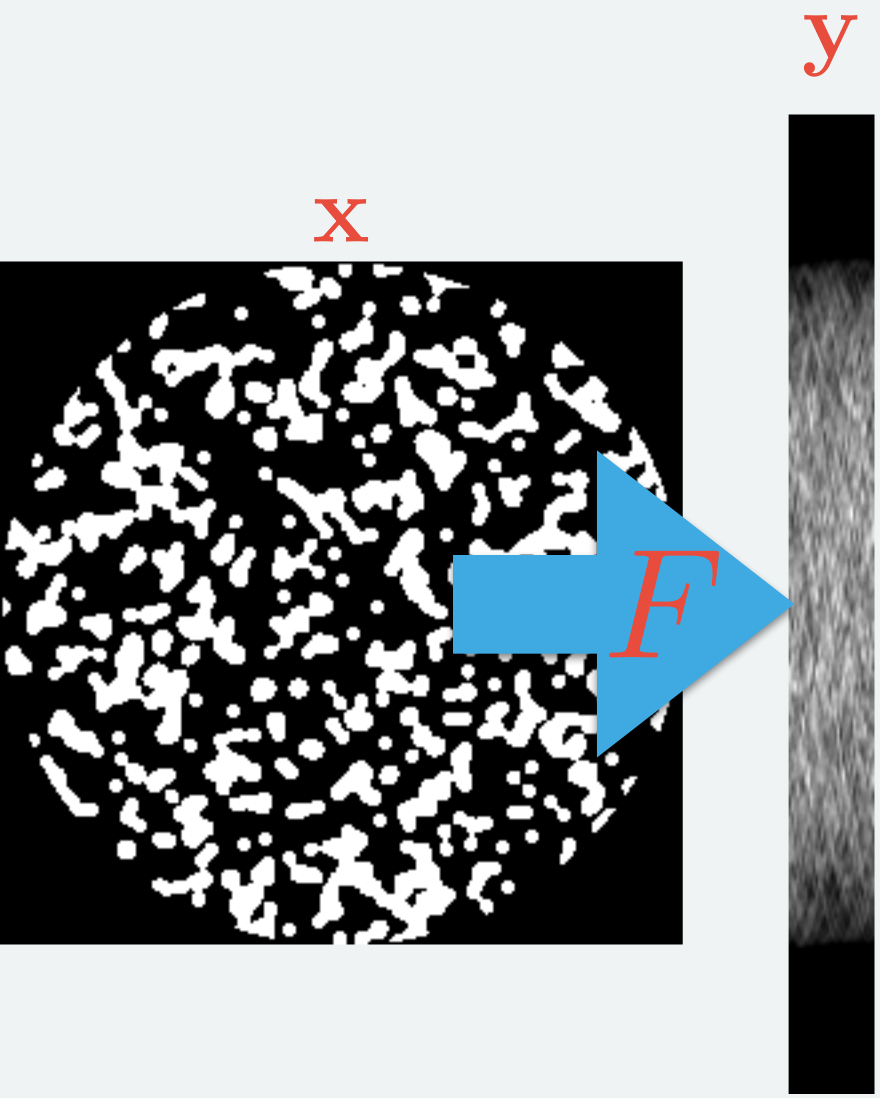
$$y_\mu = \sum_i F_{\mu i} x_i$$

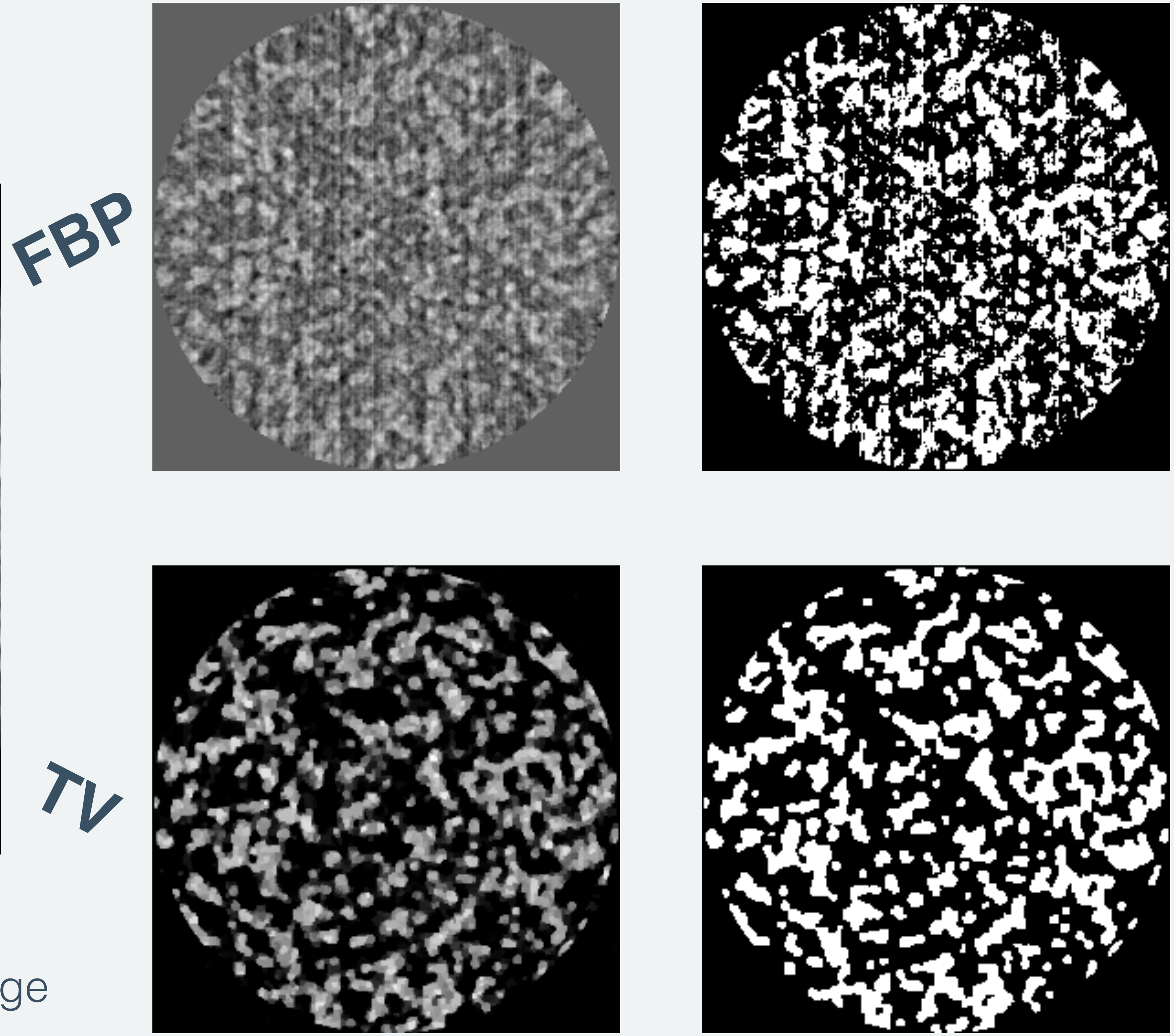
$$\downarrow$$

$$y_\mu^{\mathrm{b}} = \frac{1}{s_1 - s_0}\left(y_\mu - s_0 \sum_i F_{\mu i}\right)$$

Binary Tomography

Reconstruction?





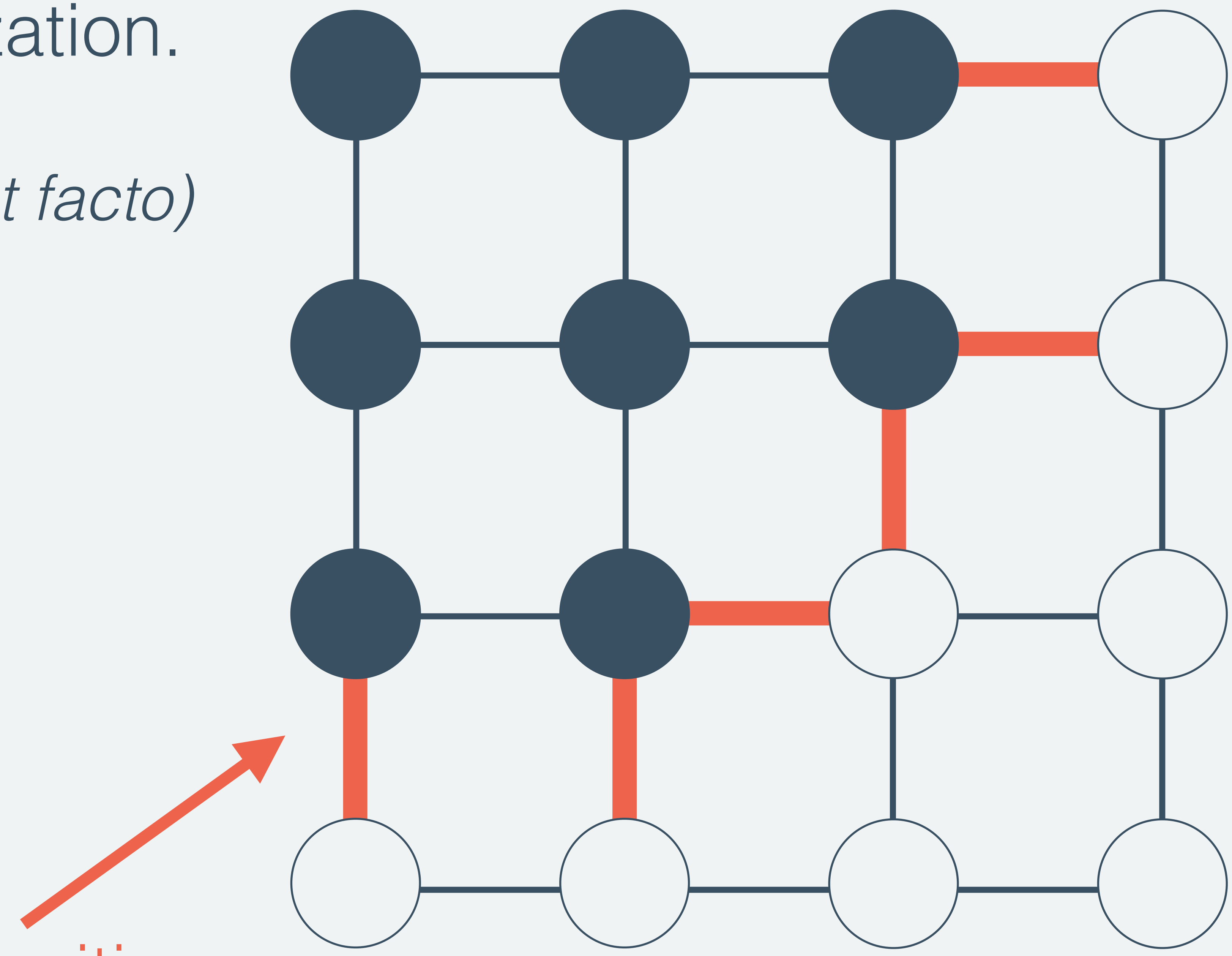
Reconstruction *Thresholding*

Variational Advantage
Leveraging knowledge of image continuity.

Variational Approach

Goal Enforce regularity naturally in the optimization.

(*Not bootstrapped ex post facto*)



e.g. penalize discontinuities in image.

Variational Approach

Total Variation

Convex approach, regularizing to promote a sparse gradient.

$$\mathrm{TV}(\mathbf{x}) = \sum_i |\nabla x_i|$$

Solve…

$$\underset{\mathbf{x}}{\mathrm{argmin}} \quad ||\mathbf{y} - F\mathbf{x}||_2^2 + \beta \mathrm{TV}(\mathbf{x}) + \mathcal{I}_{[0,1]}(\mathbf{x})$$

Match observations…

…while penalizing discontinuities…

…ensuring proper bounds.

Ex. implementations: gen. forward-backward splitting, FISTA, augmented lagrangian/alternating minimization…

Using Belief Prop.

Probabilistic Construction (*Gouillart et al, 2013*)

We desire to estimate the posterior…

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

$$= \frac{1}{\mathcal{Z}} \prod_{\mu} \left[g\left(y_\mu - \sum_{i \in \mu} x_i \right) e^{J_\mu \sum_{(ij)\in\mu} \delta_{x_i, x_j}} \right]$$

For some intractable normalization…

…over the product of factors (measurements/lines)…

…and stochastic output function…

AWGN $e^{-\frac{1}{2\sigma^2}\left(y_\mu - \sum_{i\in\mu} x_i\right)^2}$

Noiseless $\delta\left(y_\mu - \sum_{i\in\mu} x_i \right)$

…promote regularity according to some constant.

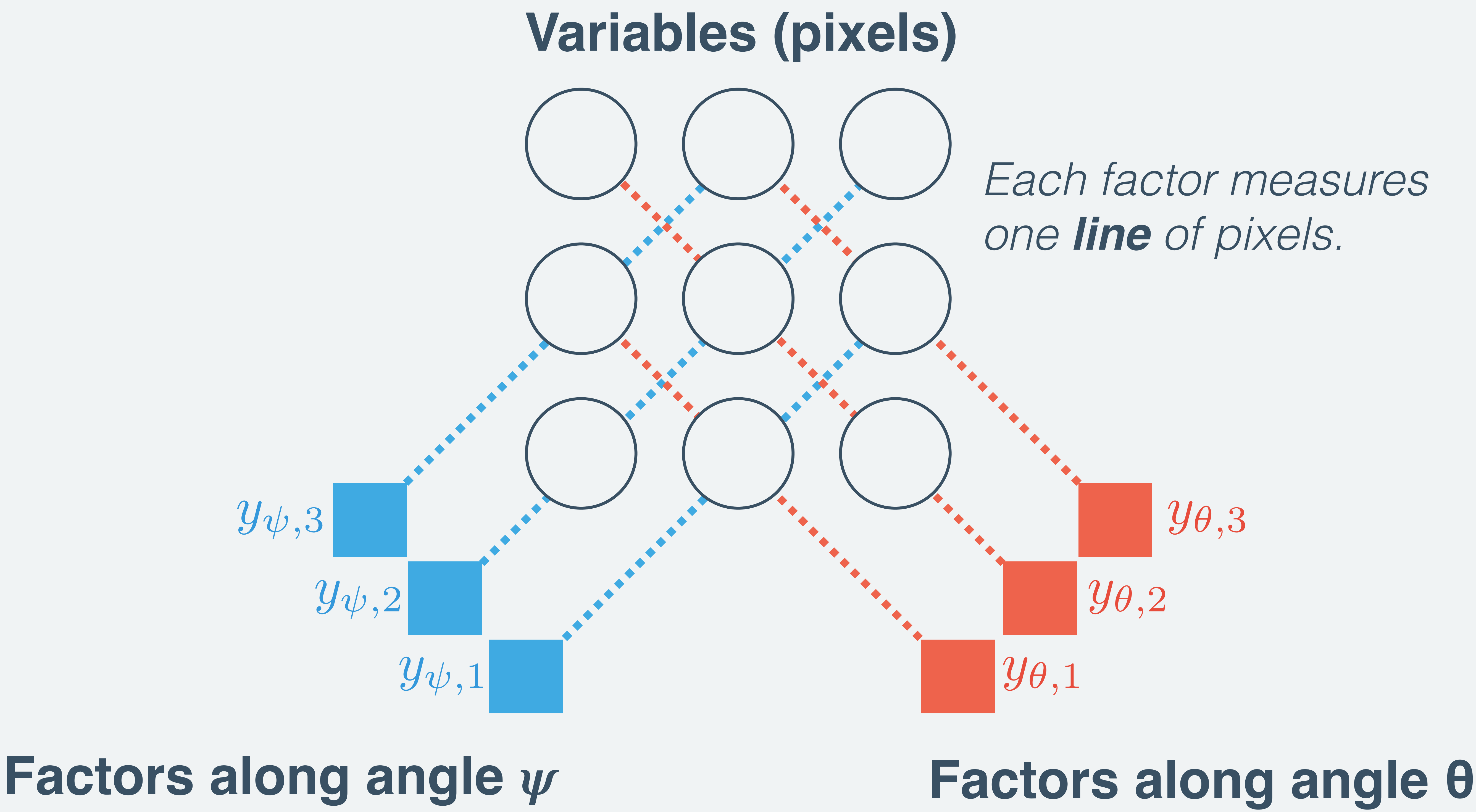
Using Belief Prop.

Goal A factorized approximation of the posterior allowing for either MAP or MMSE estimation of **x**.

$$Q(\mathbf{x}) = \prod_{i=1}^{N} q(x_i) \approx P(\mathbf{x}|\mathbf{y}, F)$$

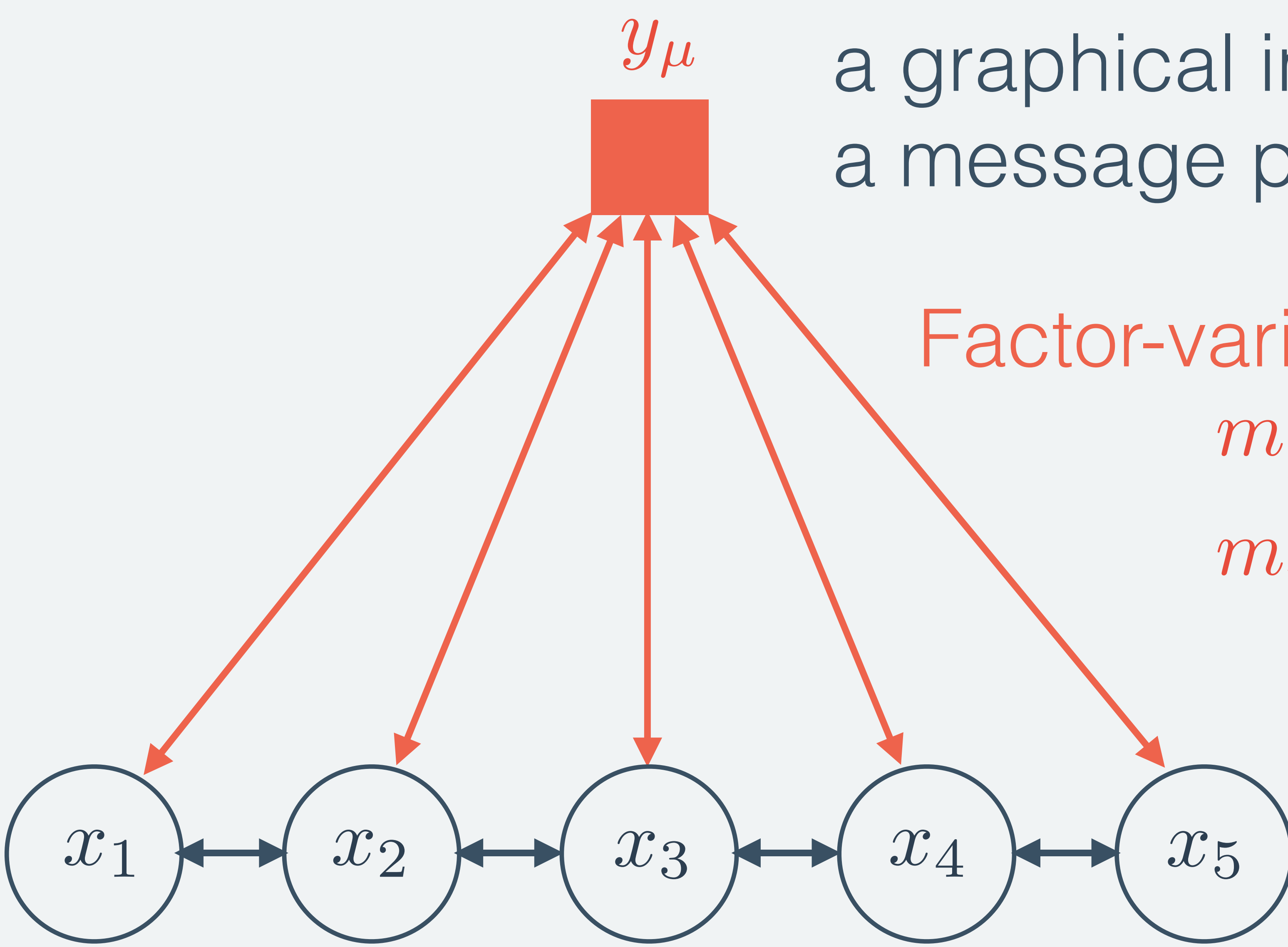
Graphical Representation Key to constructing a message passing to accomplish this factorization.

Using Belief Prop.



Using Belief Prop.

Estimate Posterior via BP Use a graphical interpretation to construct a message passing.



Factor-variable Messages

$$m_{\mu\to i}(x_i)$$

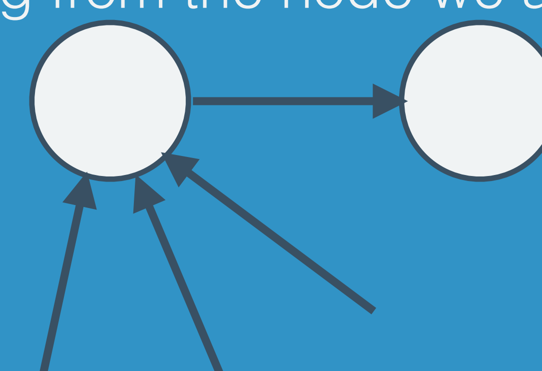
$$m_{i\to\mu}(x_i)$$

Variable-variable Messages

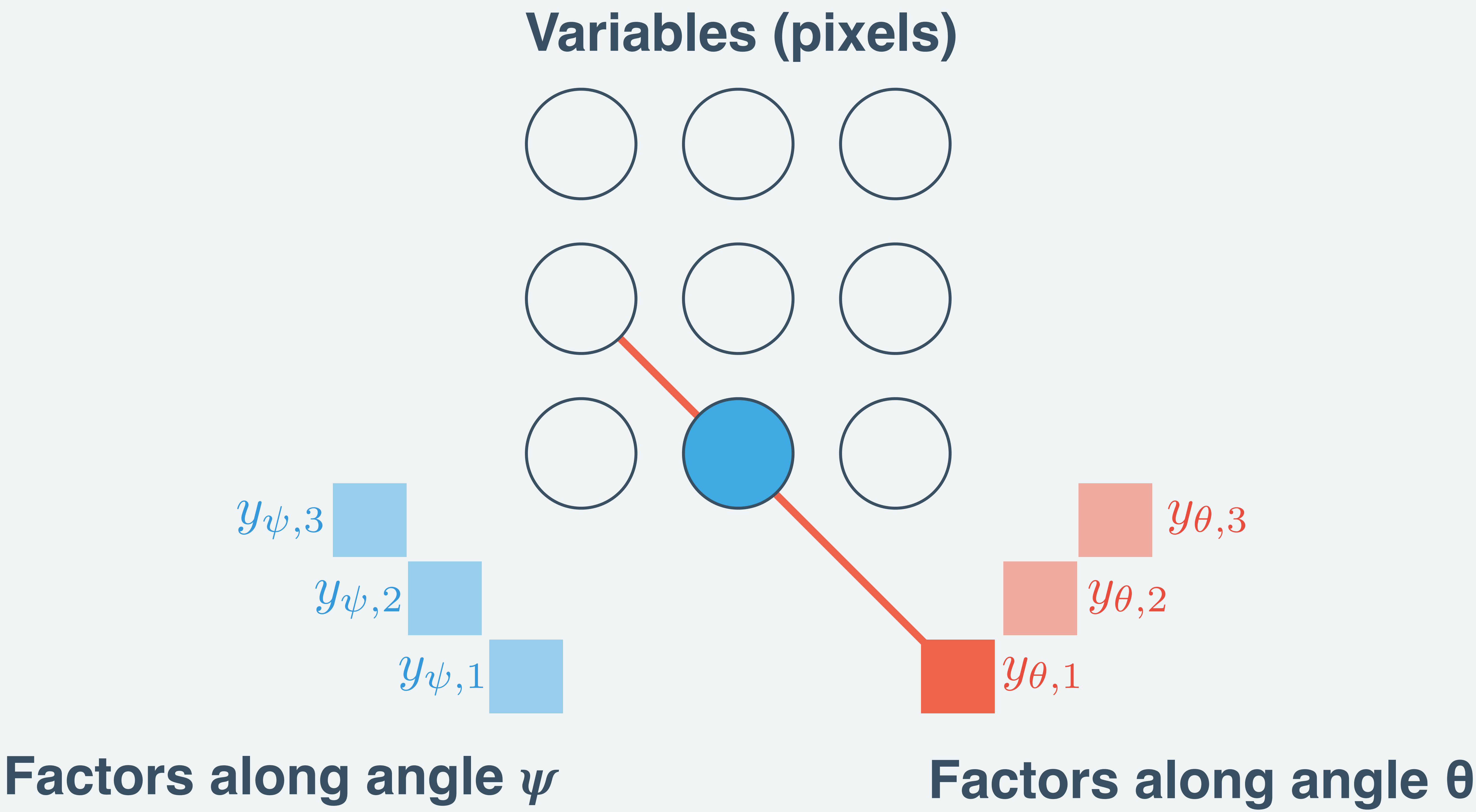
$$\eta^L_{i\to i+1}(x_i)$$

$$\eta^R_{i\to i-1}(x_i)$$

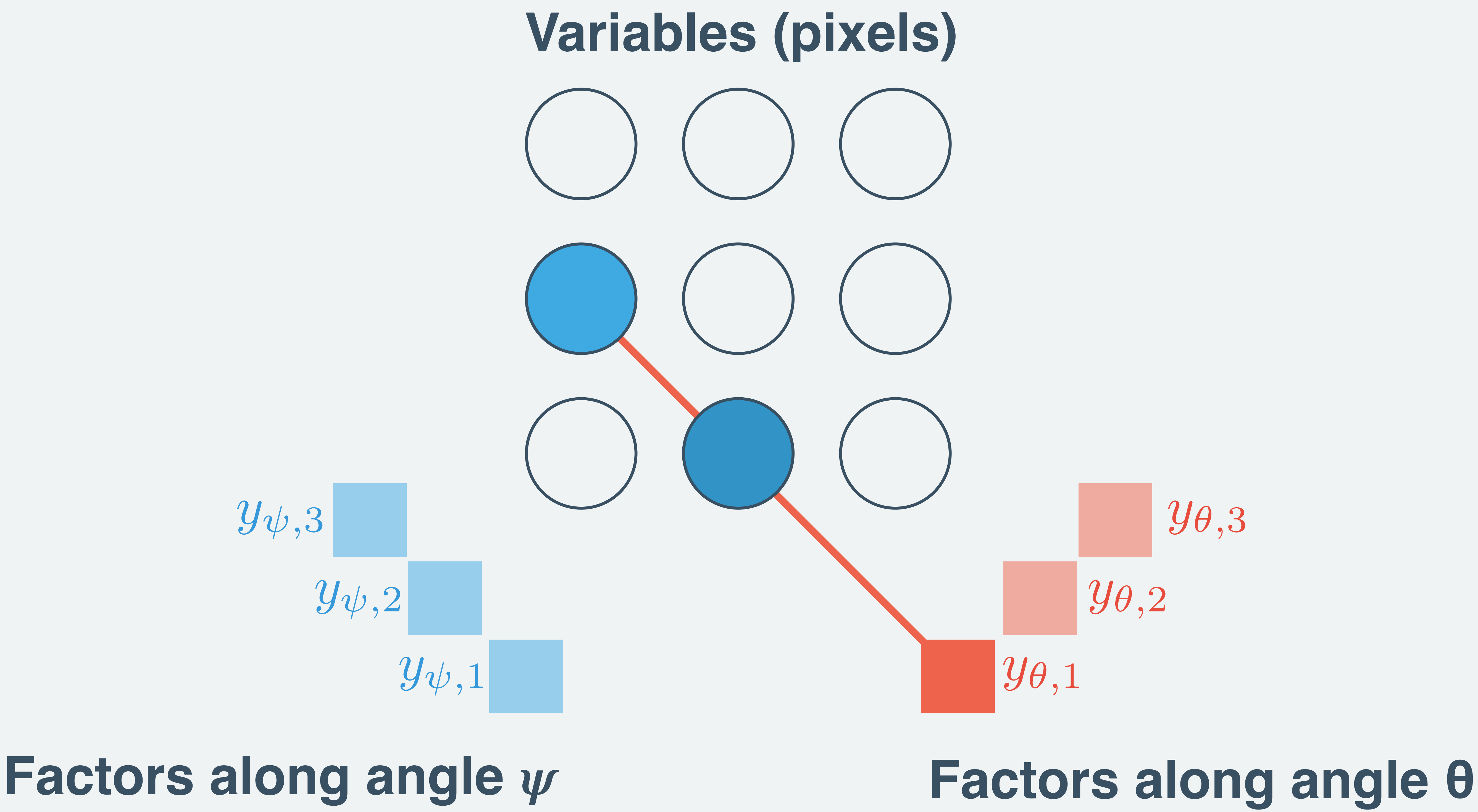
Outgiong messages calculated via ***cavity***: product of all incoming ***sans*** the message coming from the node we are sending to.



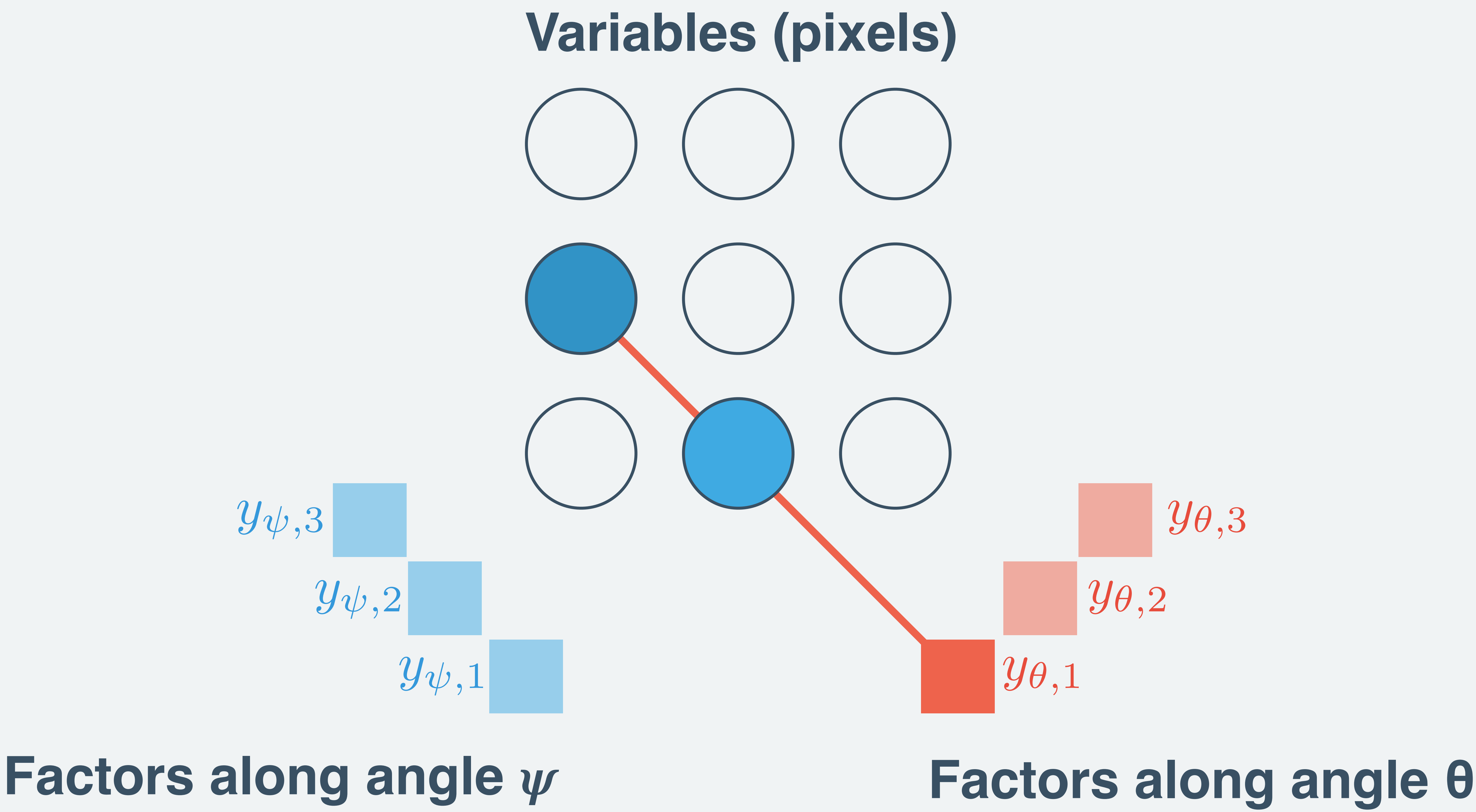
Using Belief Prop.



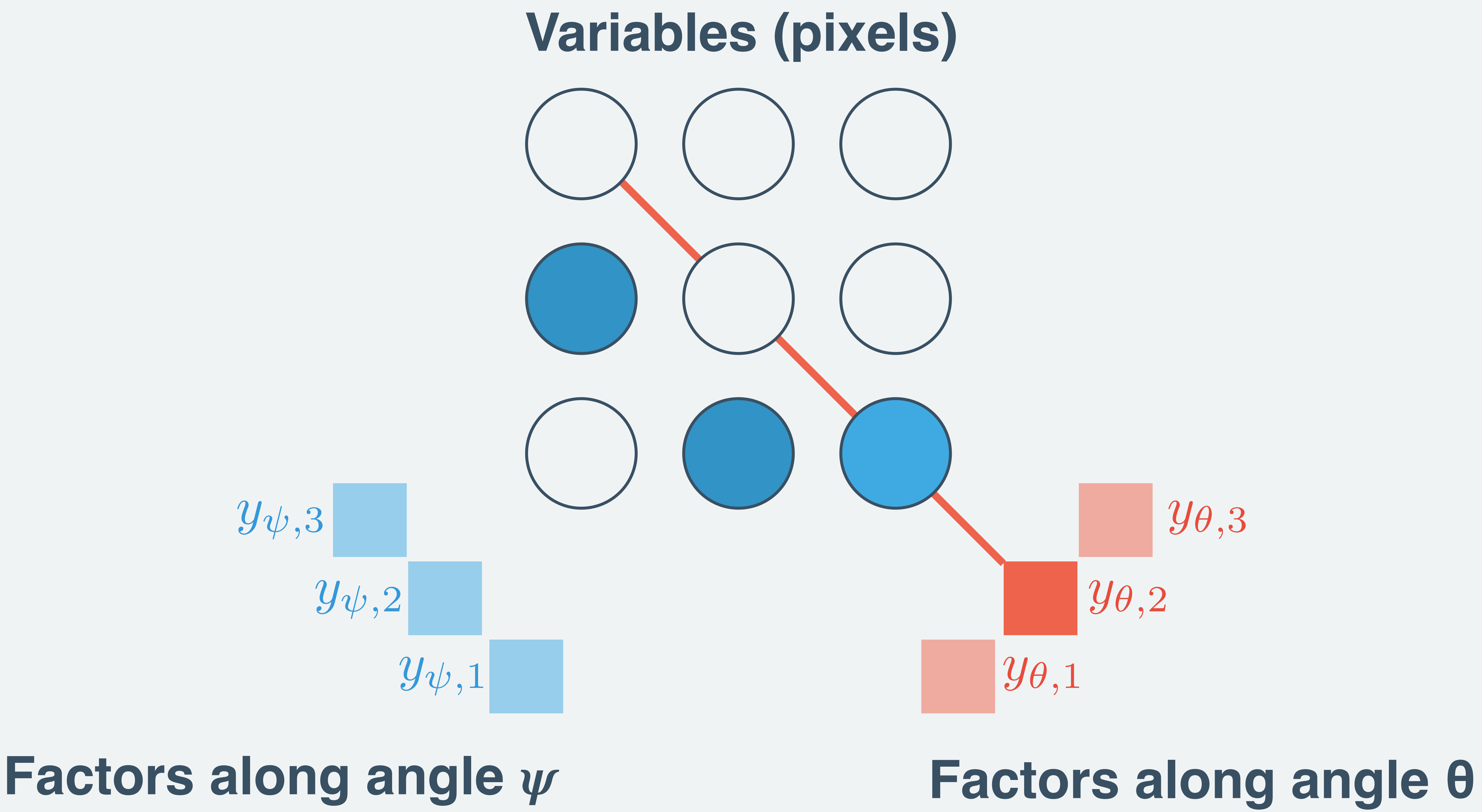
Using Belief Prop.



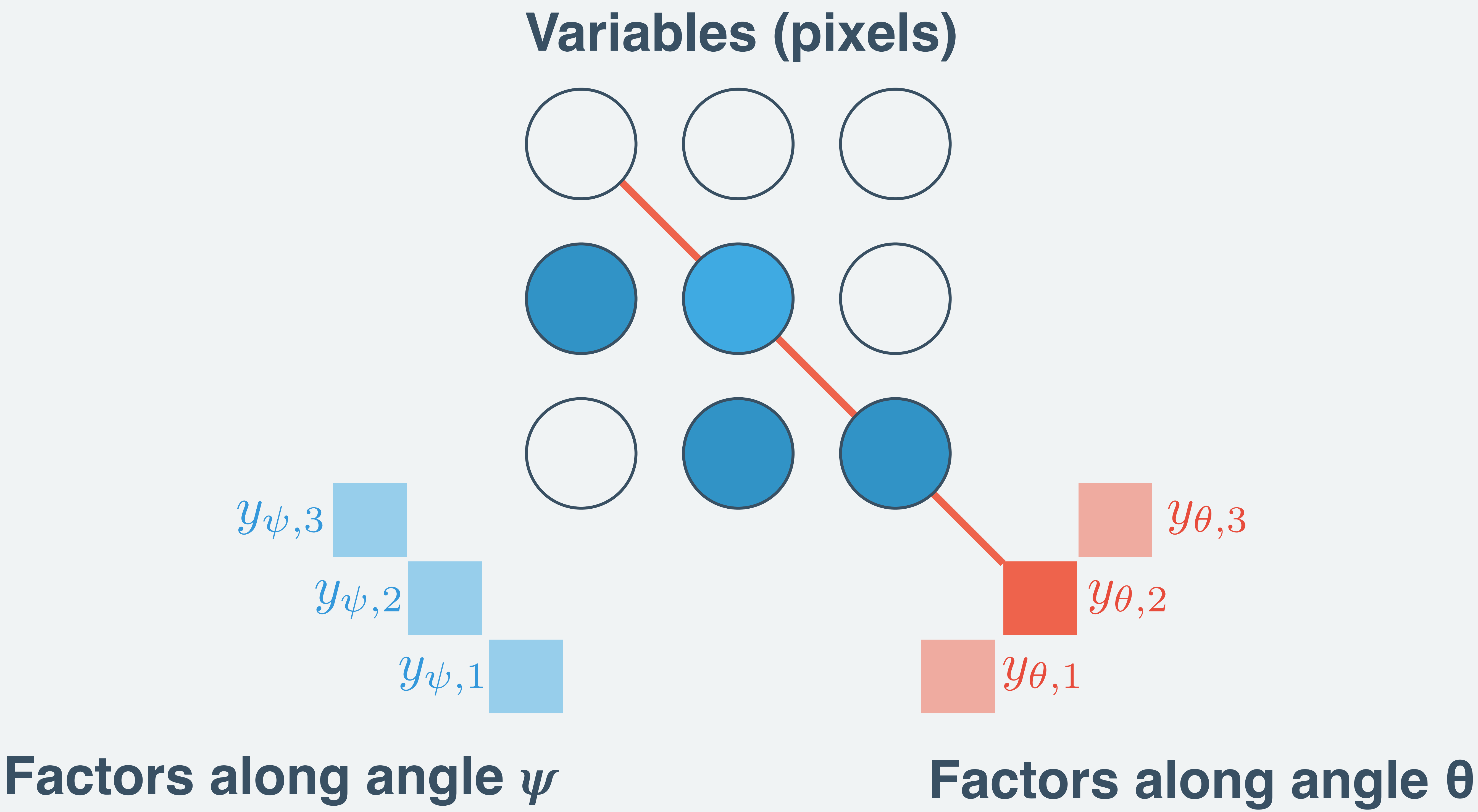
Using Belief Prop.



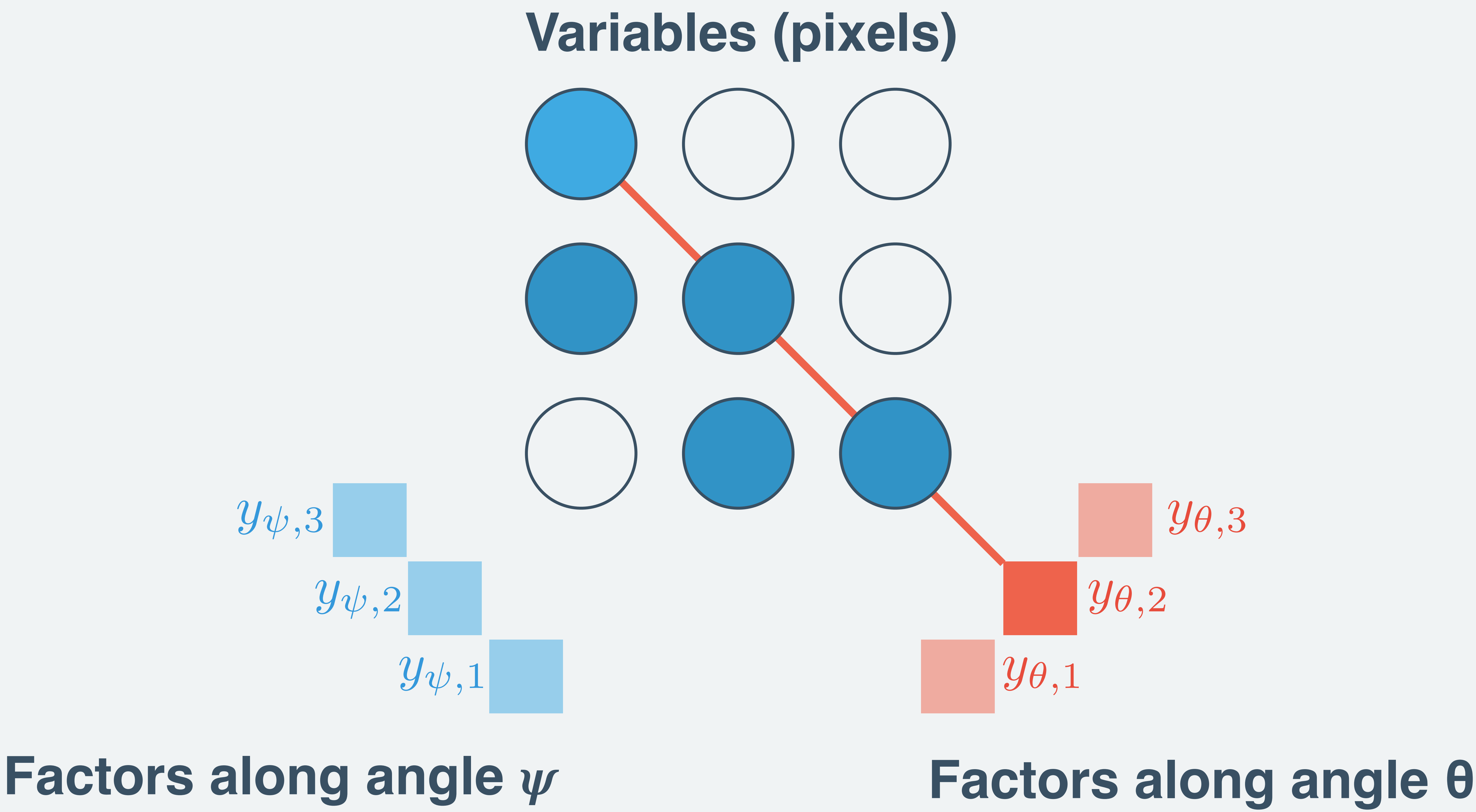
Using Belief Prop.



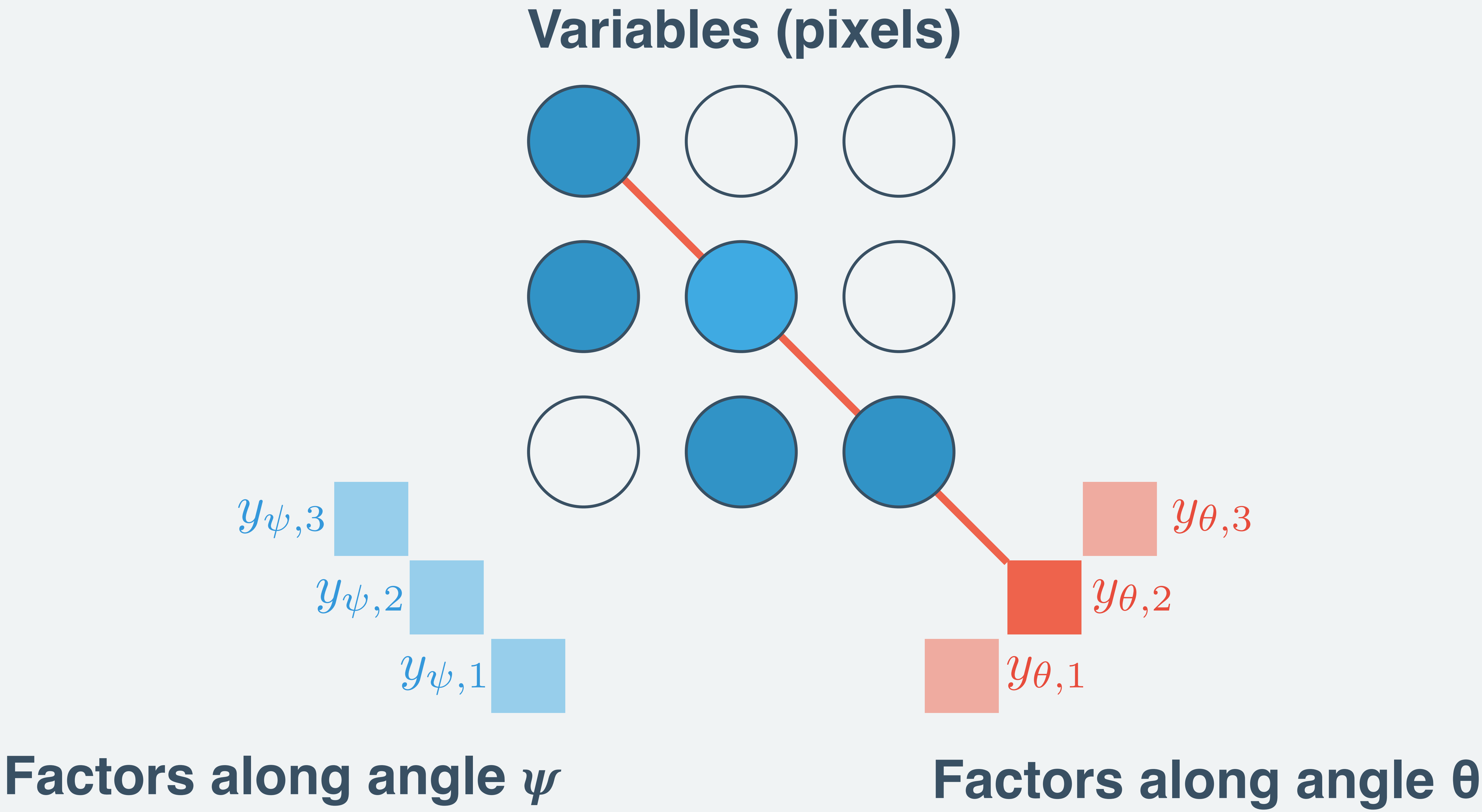
Using Belief Prop.



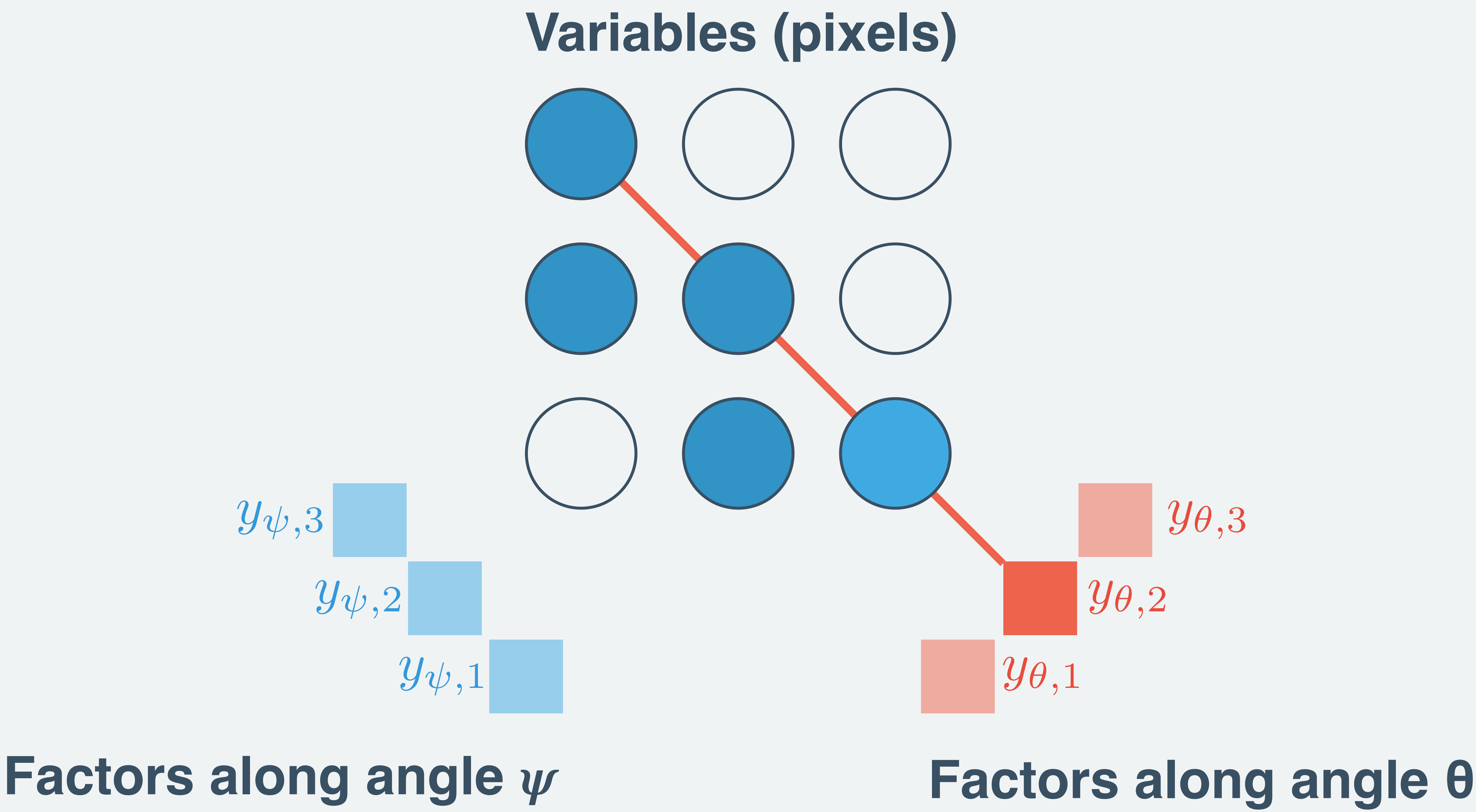
Using Belief Prop.



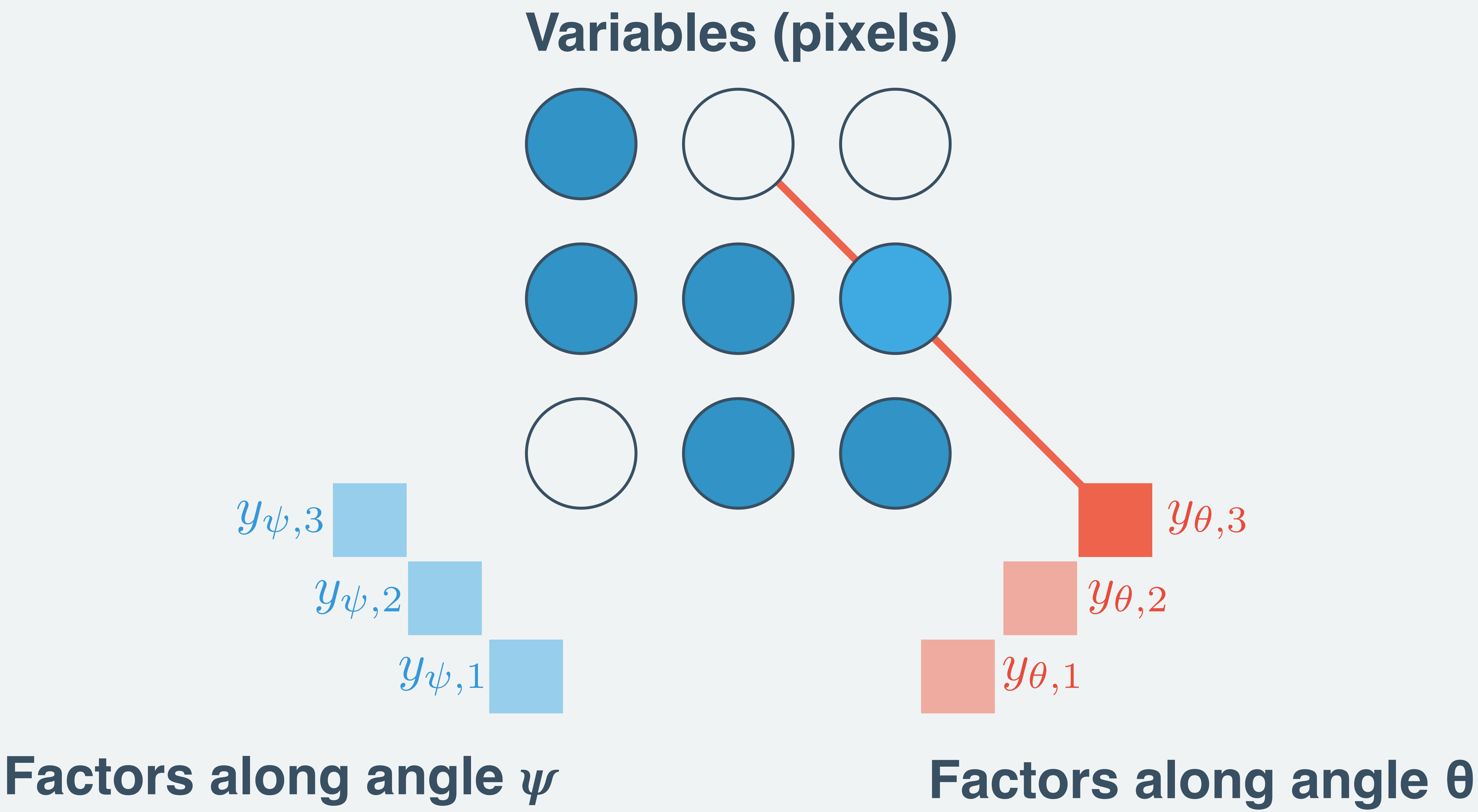
Using Belief Prop.



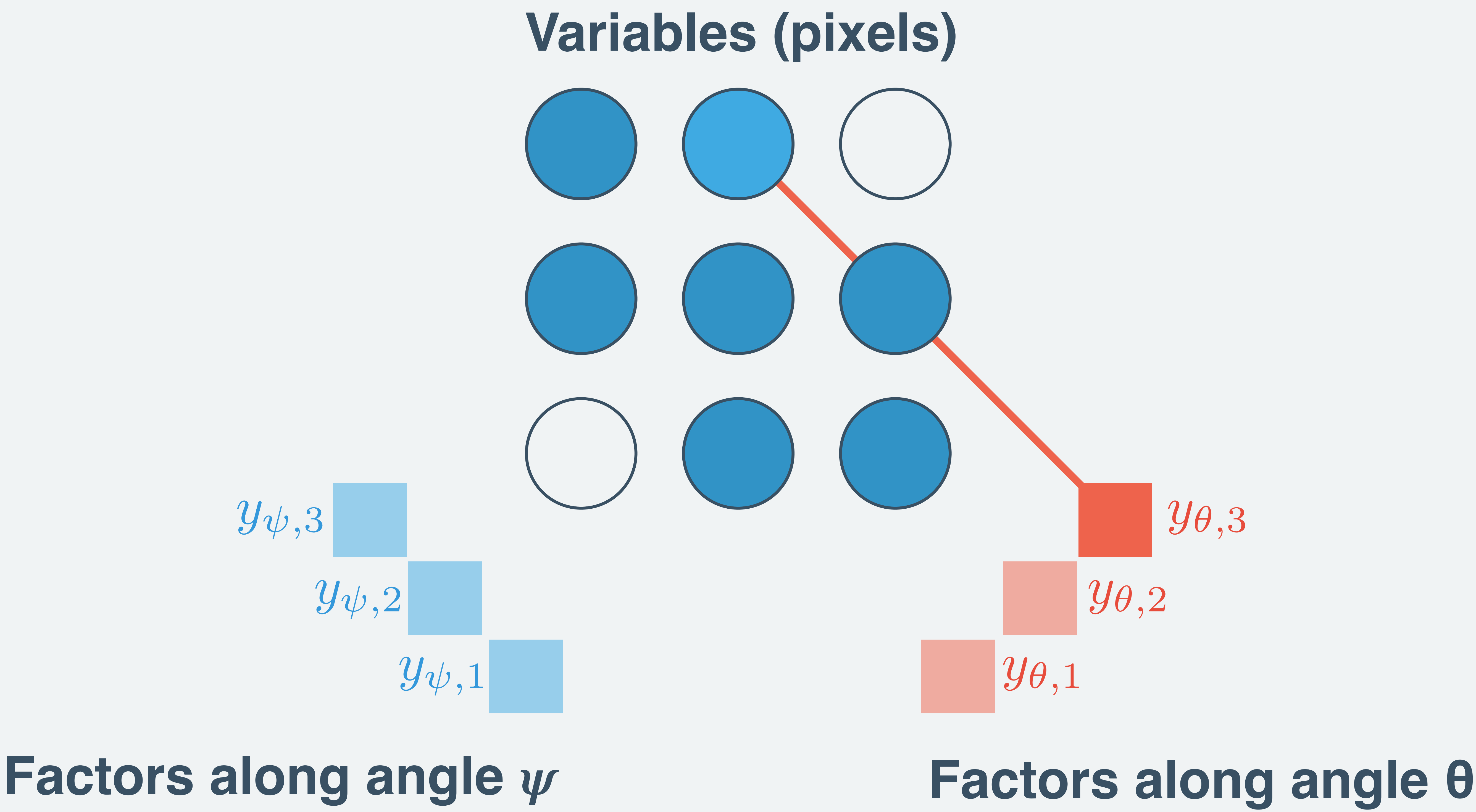
Using Belief Prop.



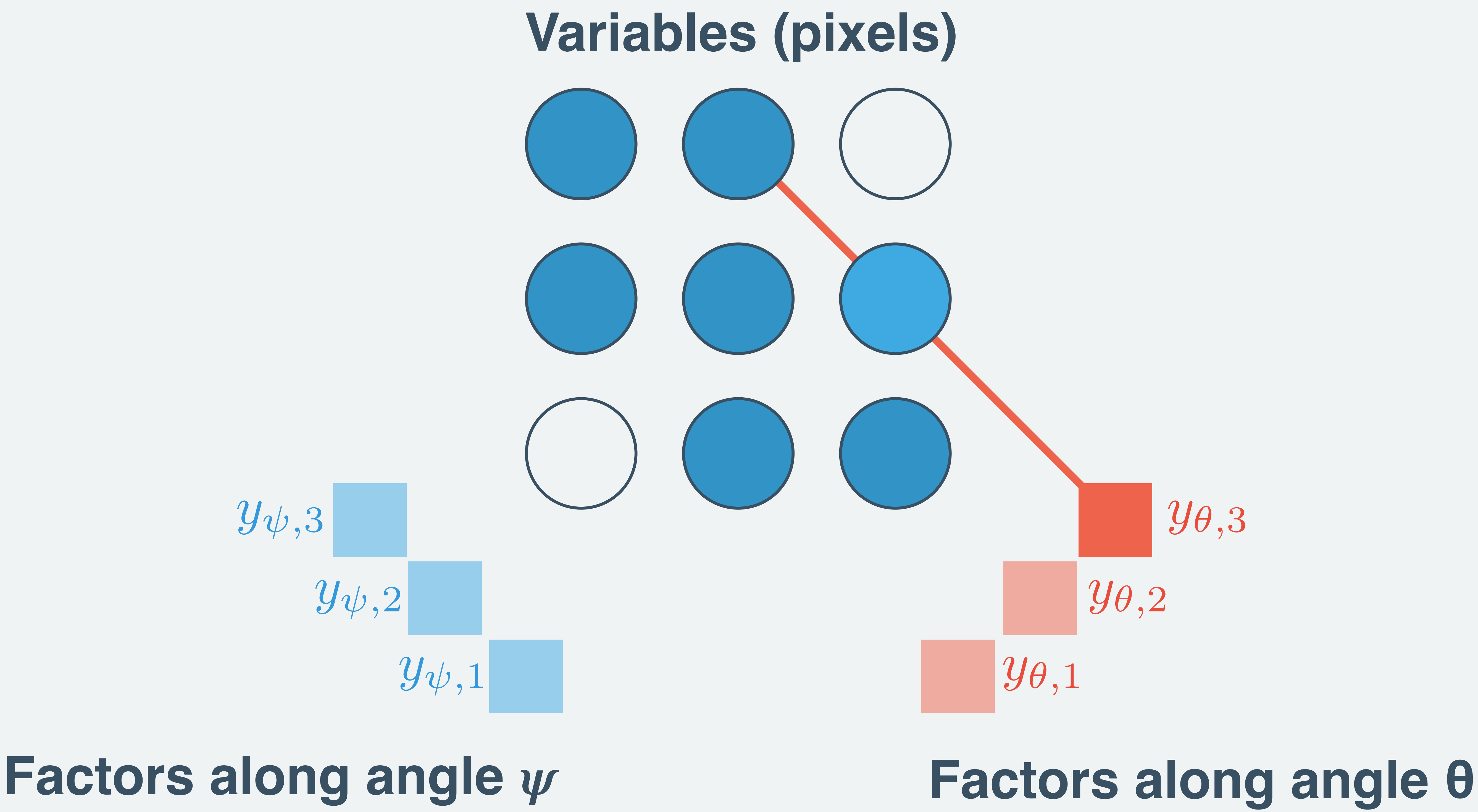
Using Belief Prop.



Using Belief Prop.

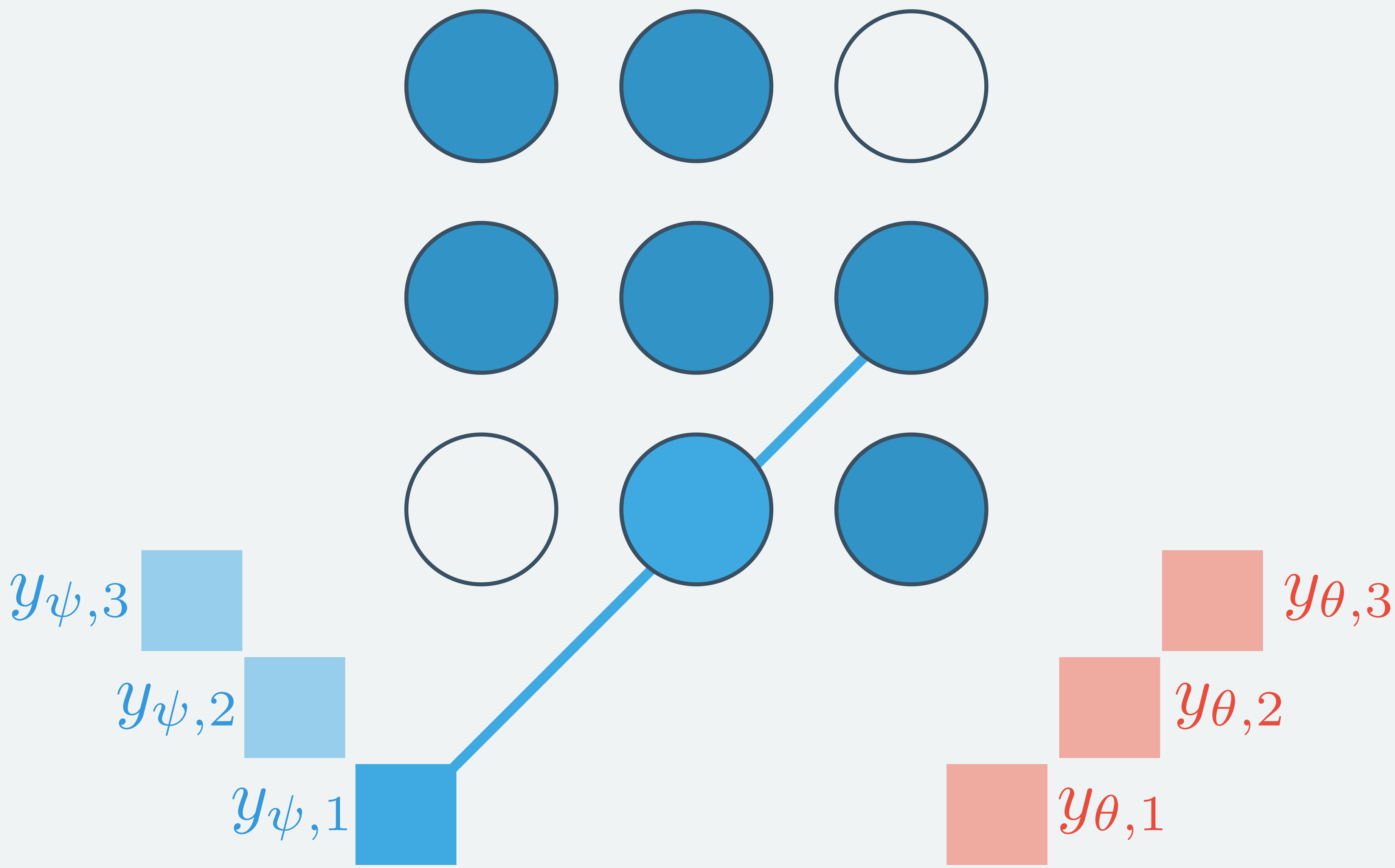


Using Belief Prop.



Using Belief Prop.

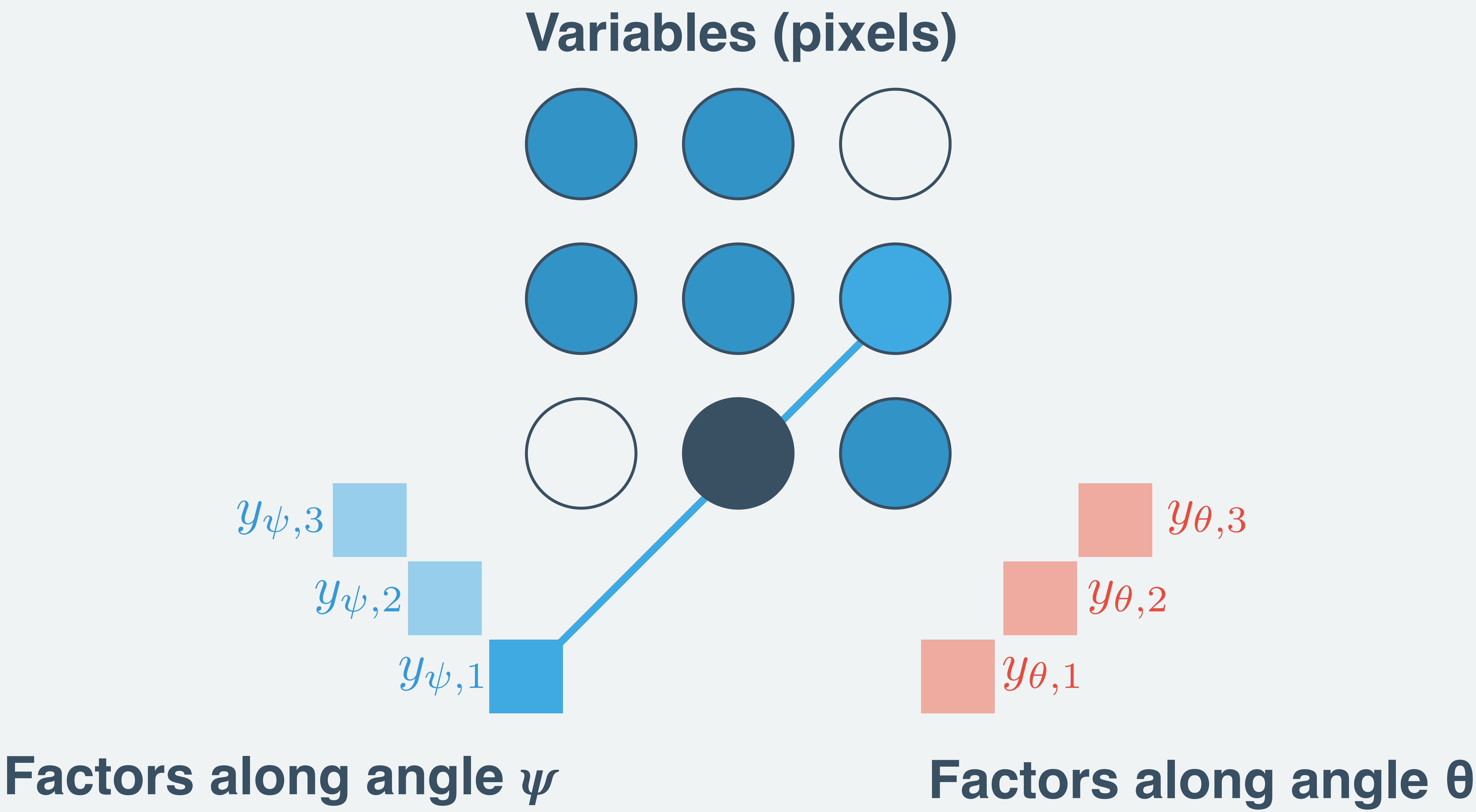
Variables (pixels)



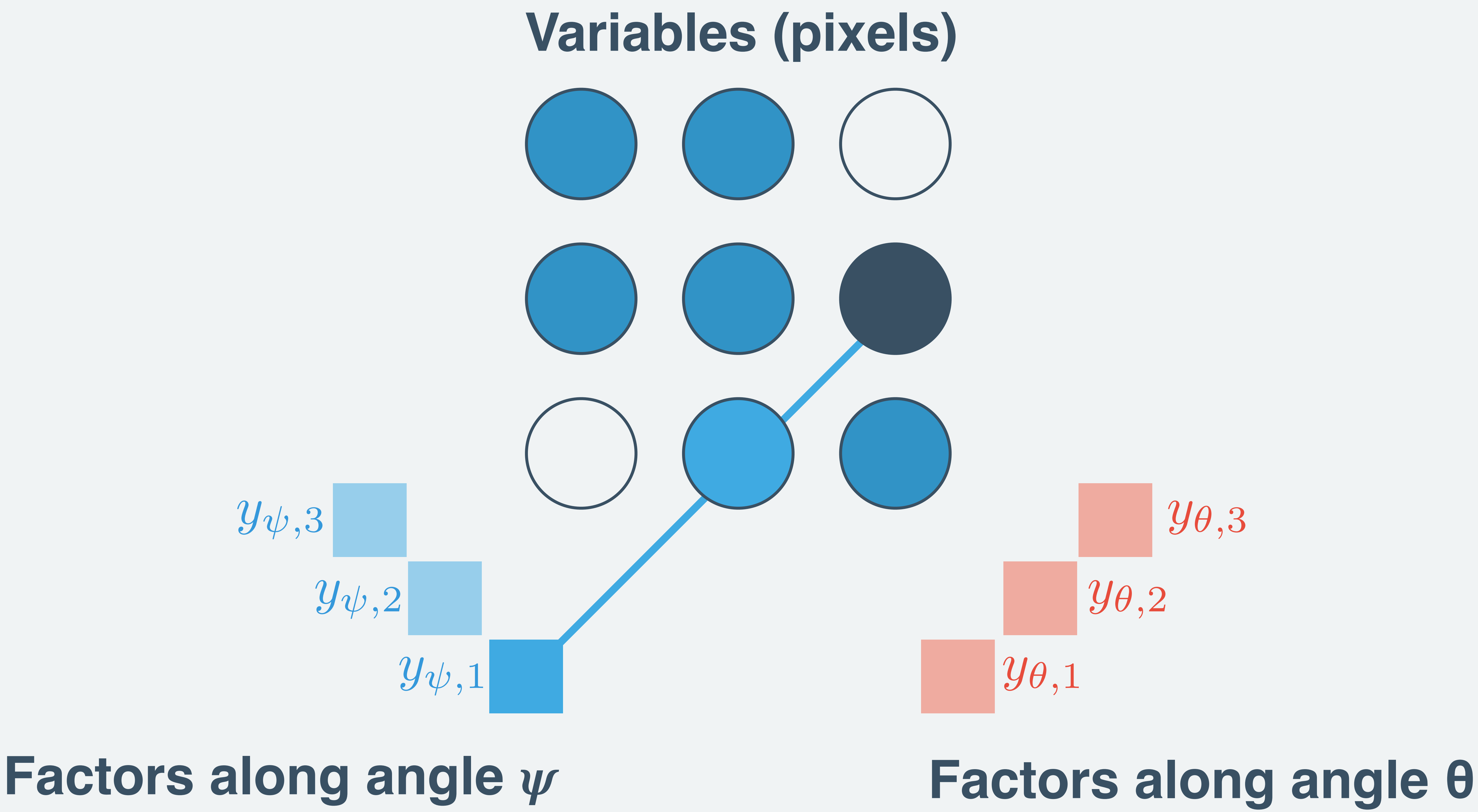
Factors along angle ψ

Factors along angle θ

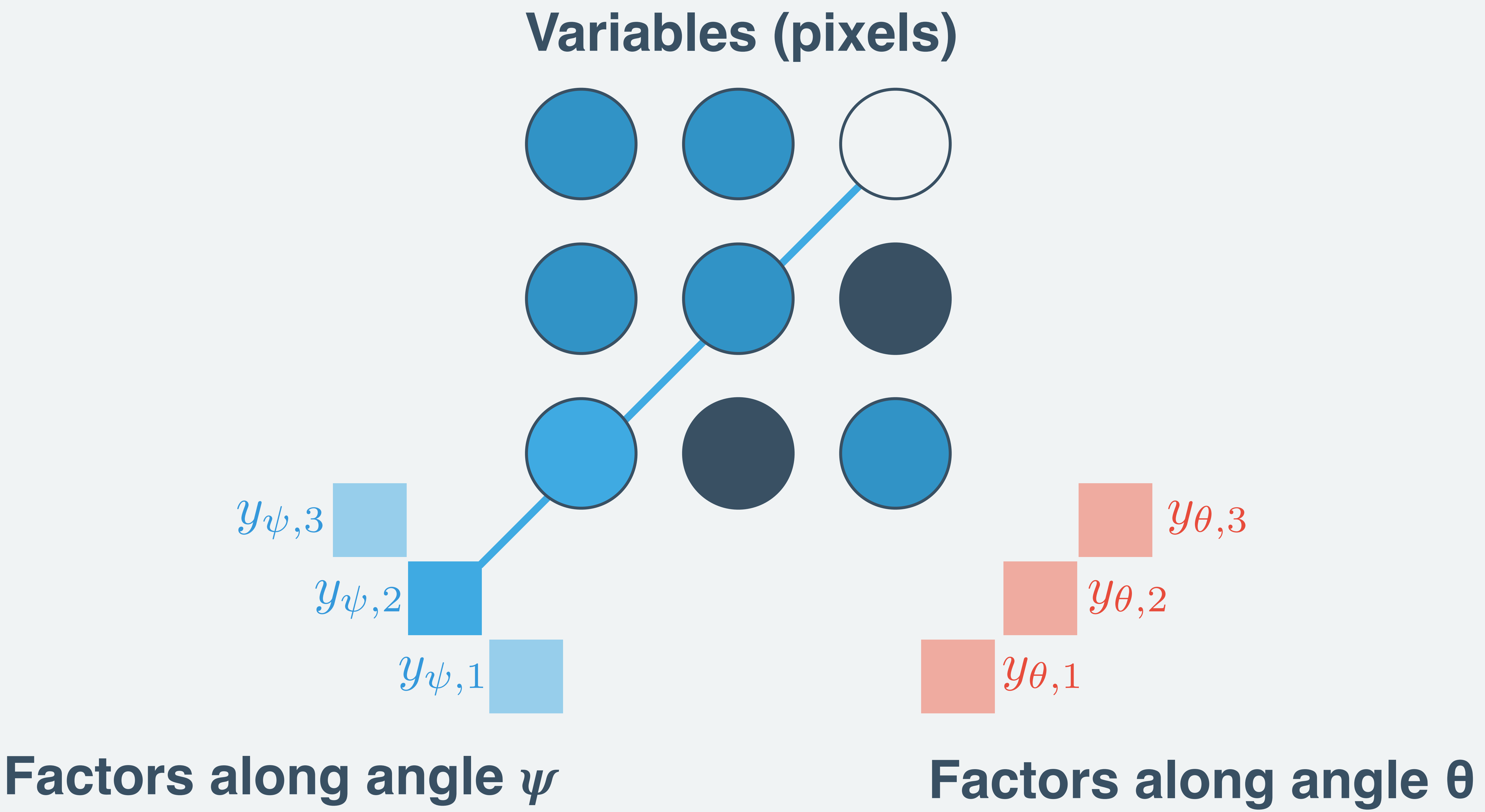
Using Belief Prop.



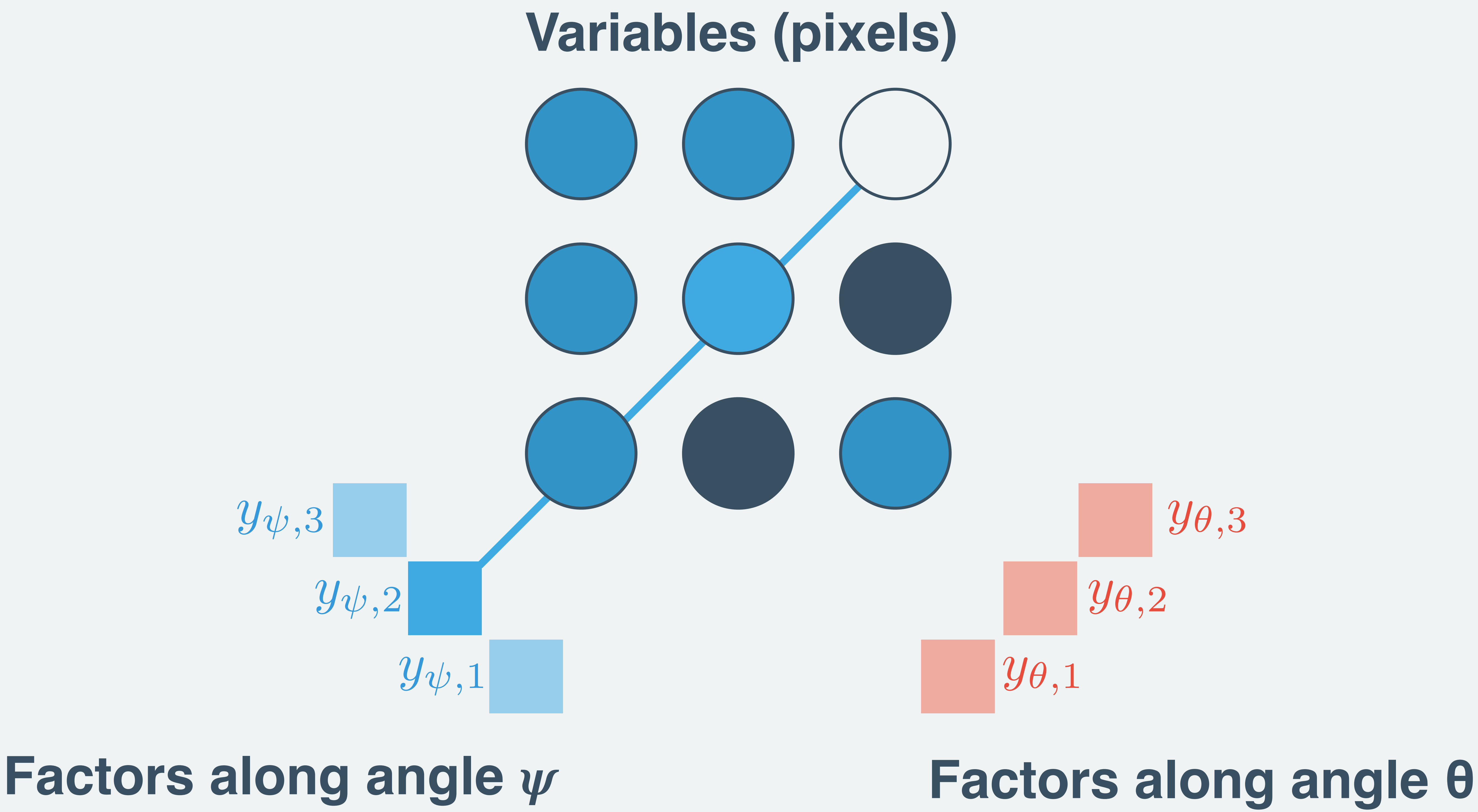
Using Belief Prop.



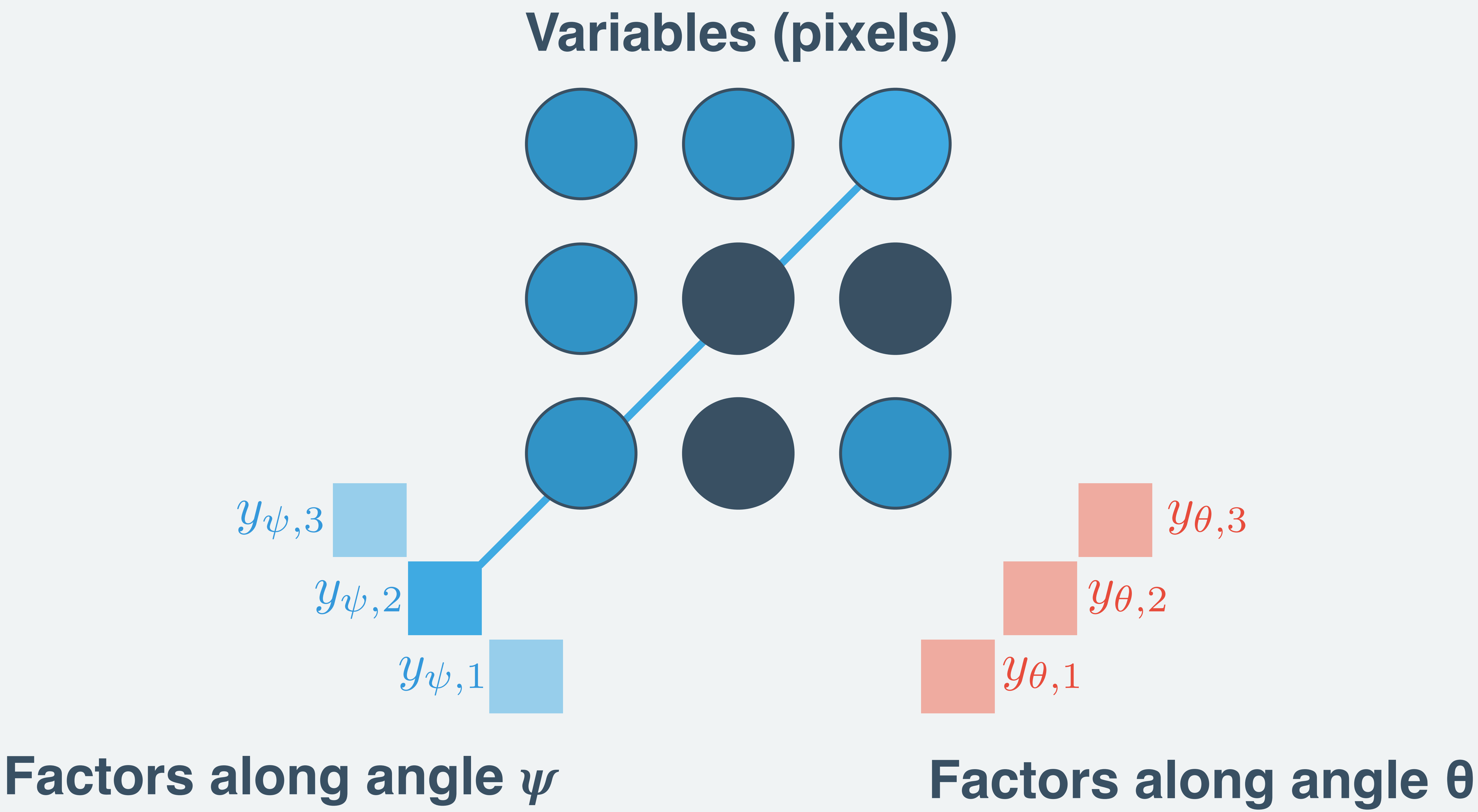
Using Belief Prop.



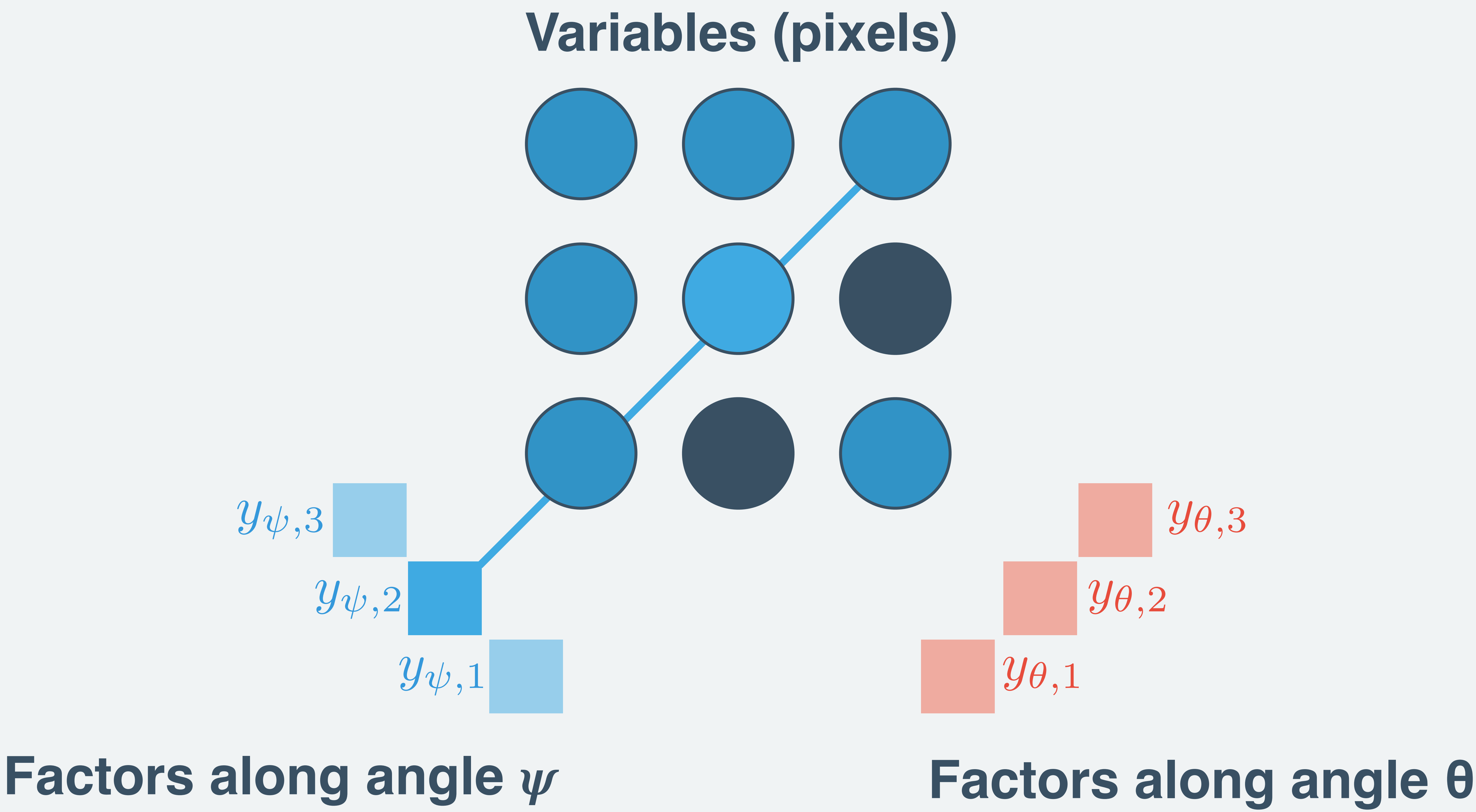
Using Belief Prop.



Using Belief Prop.

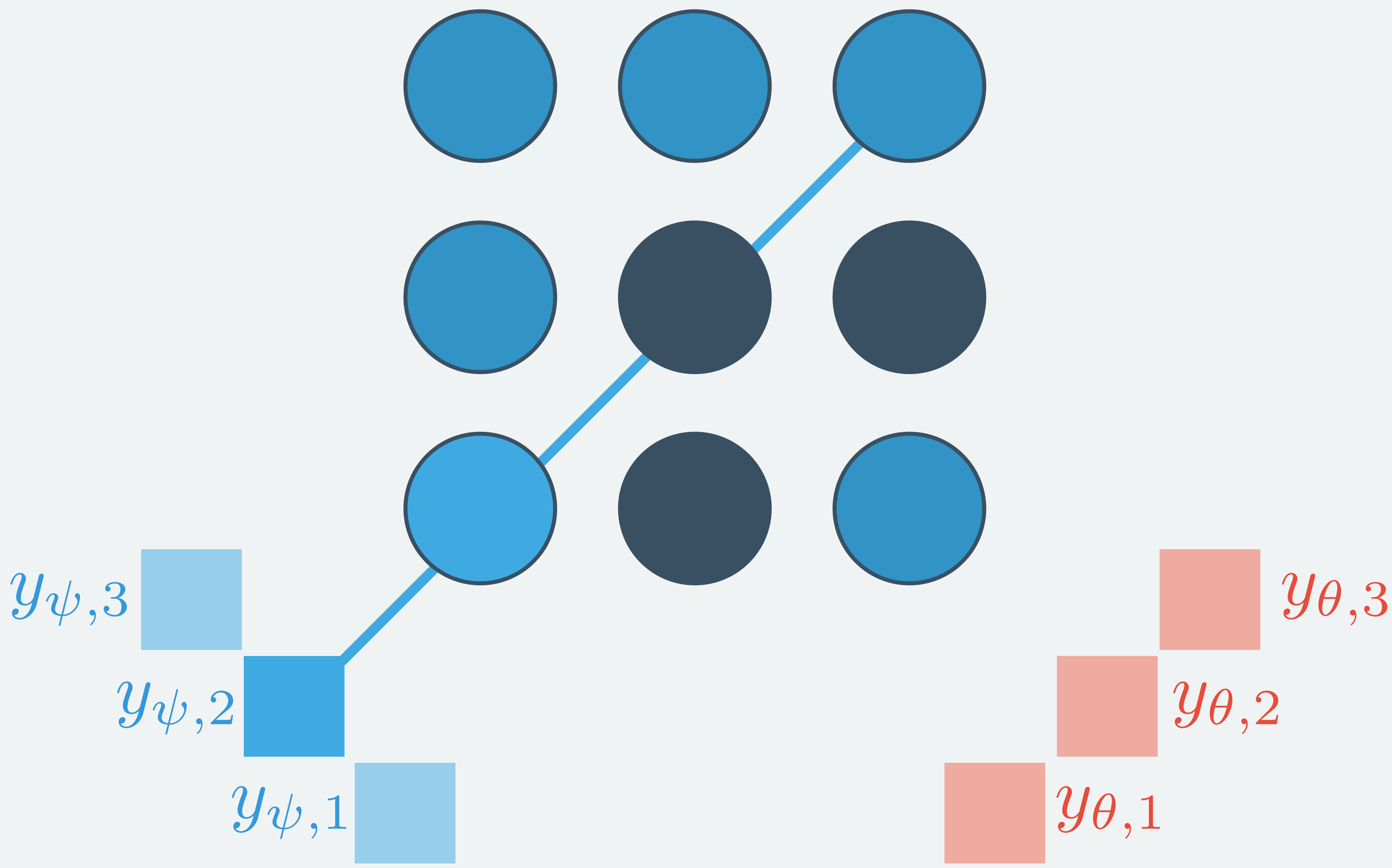


Using Belief Prop.



Using Belief Prop.

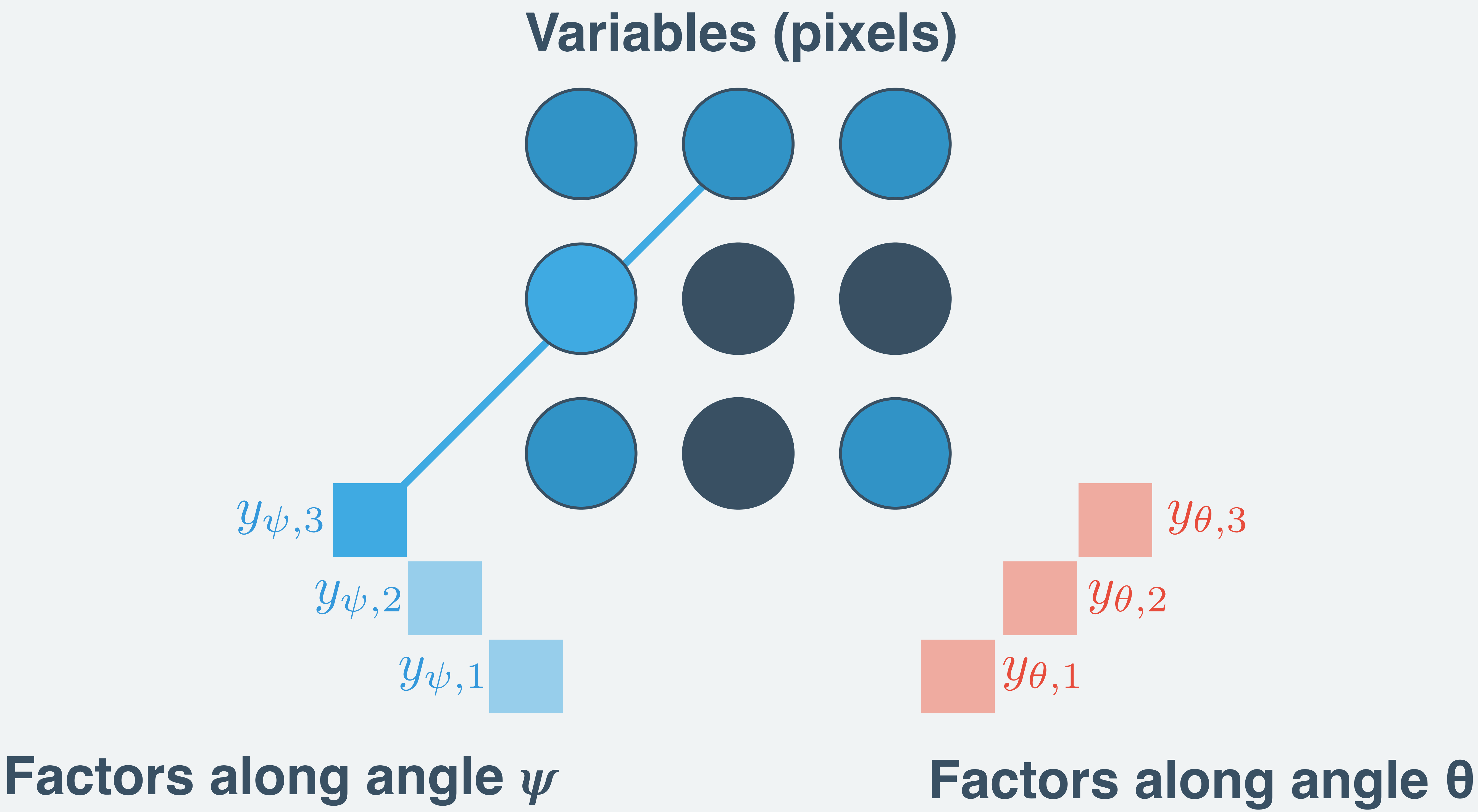
Variables (pixels)



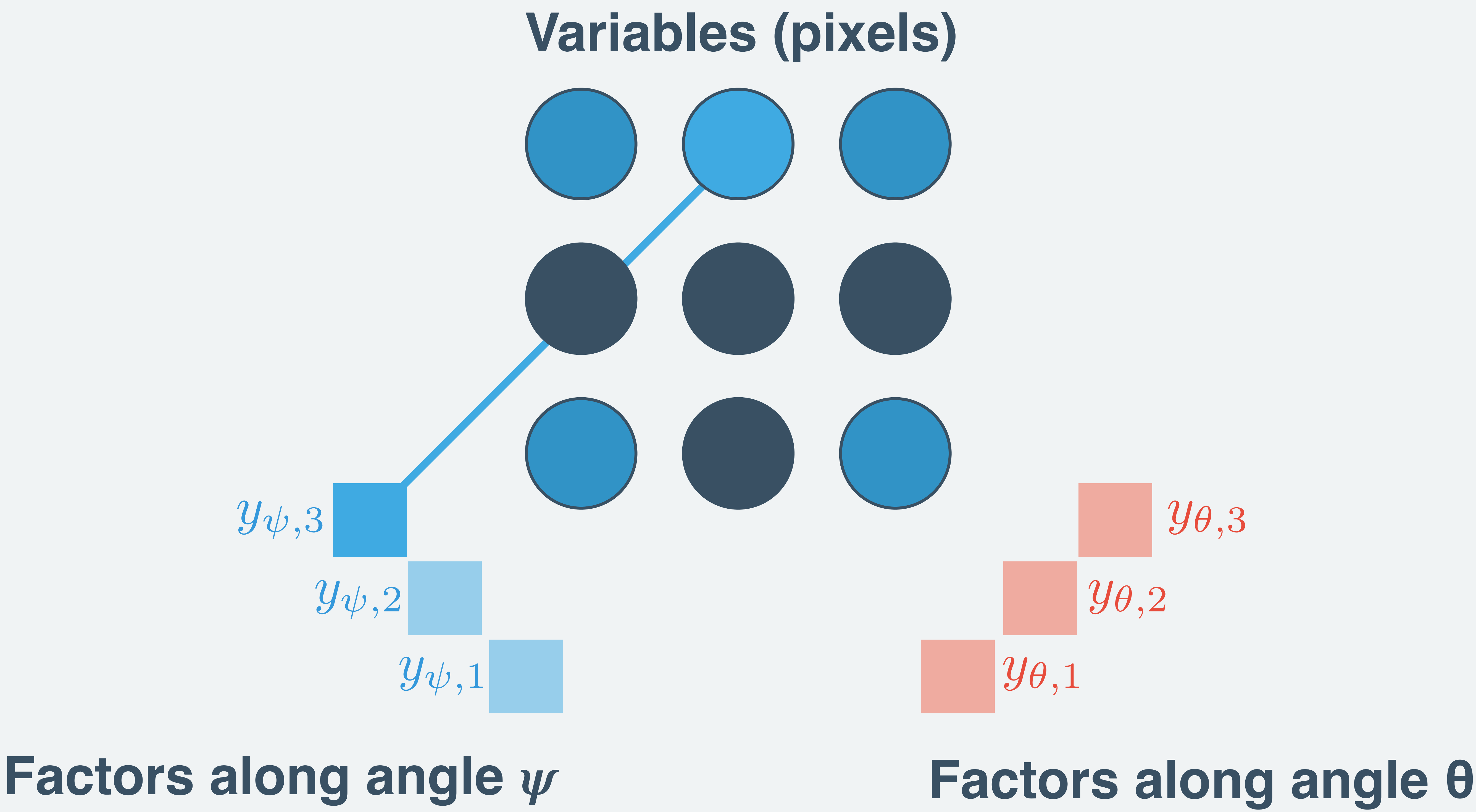
Factors along angle ψ

Factors along angle θ

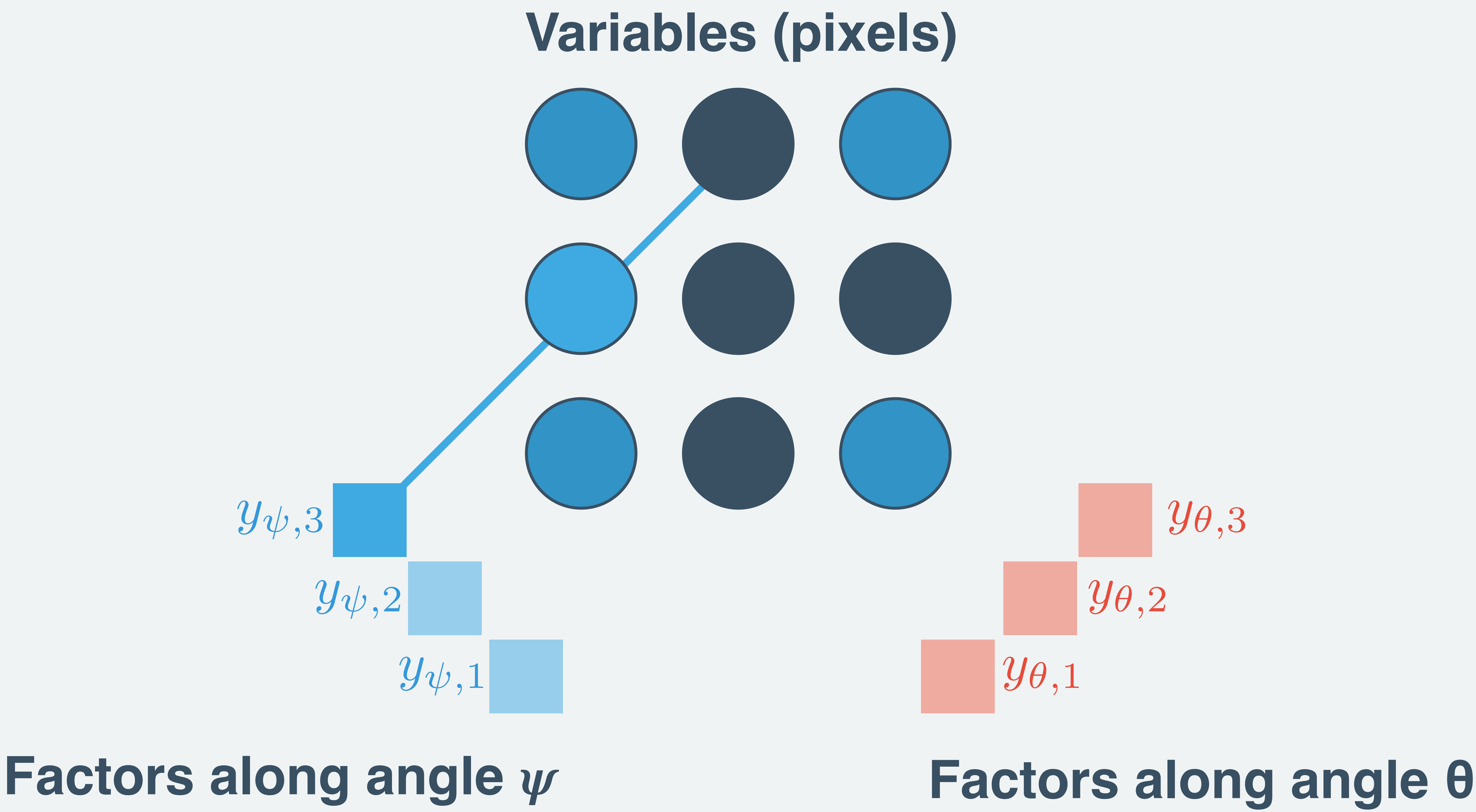
Using Belief Prop.



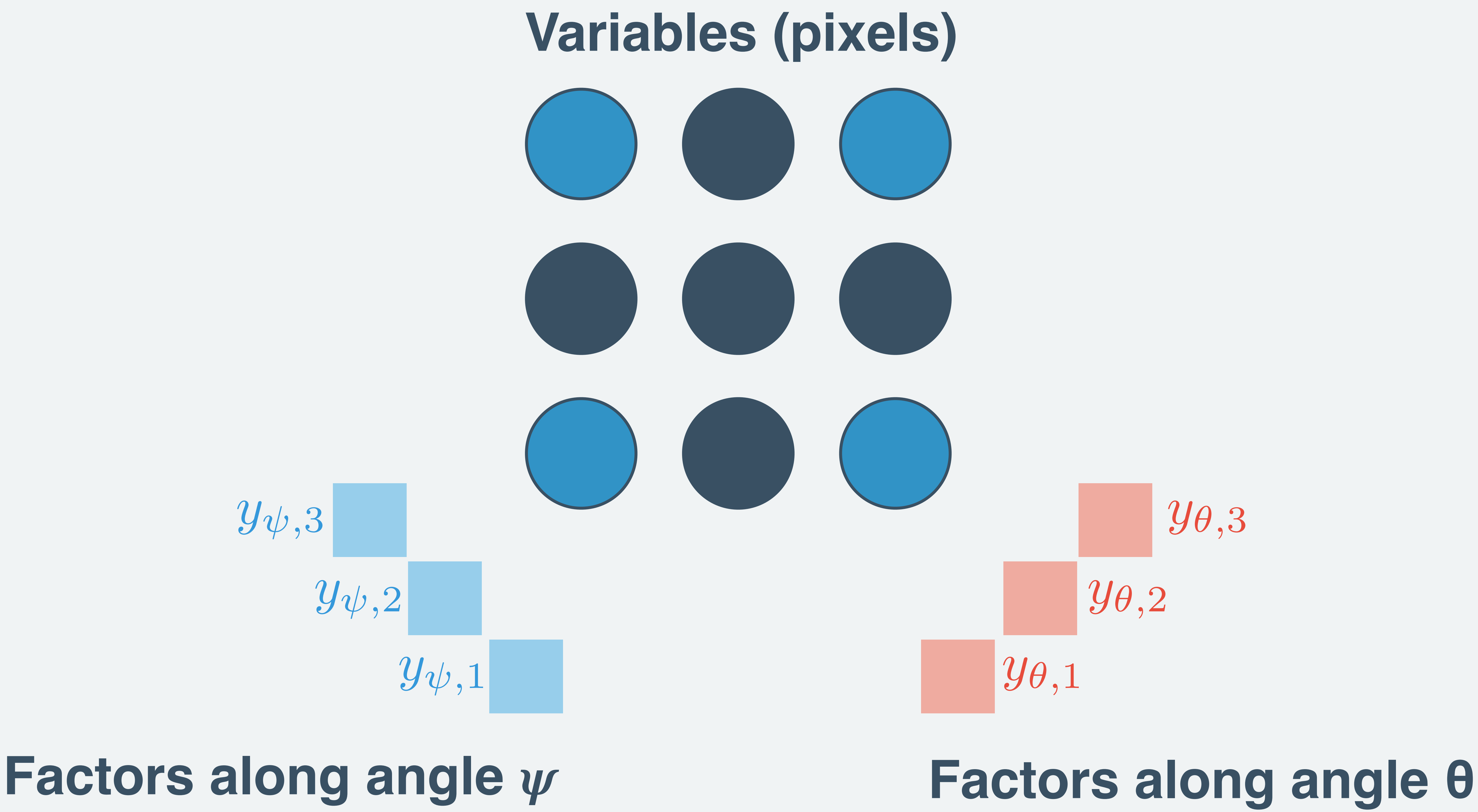
Using Belief Prop.



Using Belief Prop.



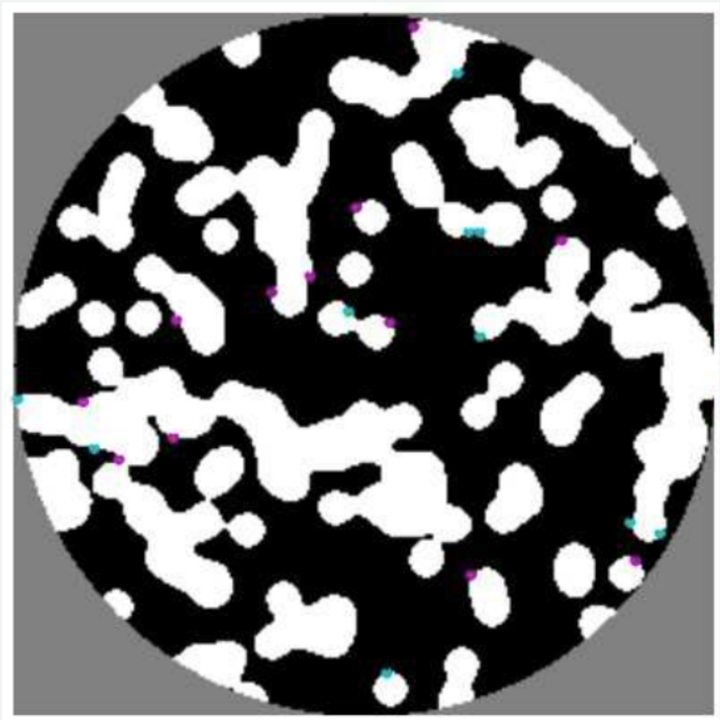
Using Belief Prop.



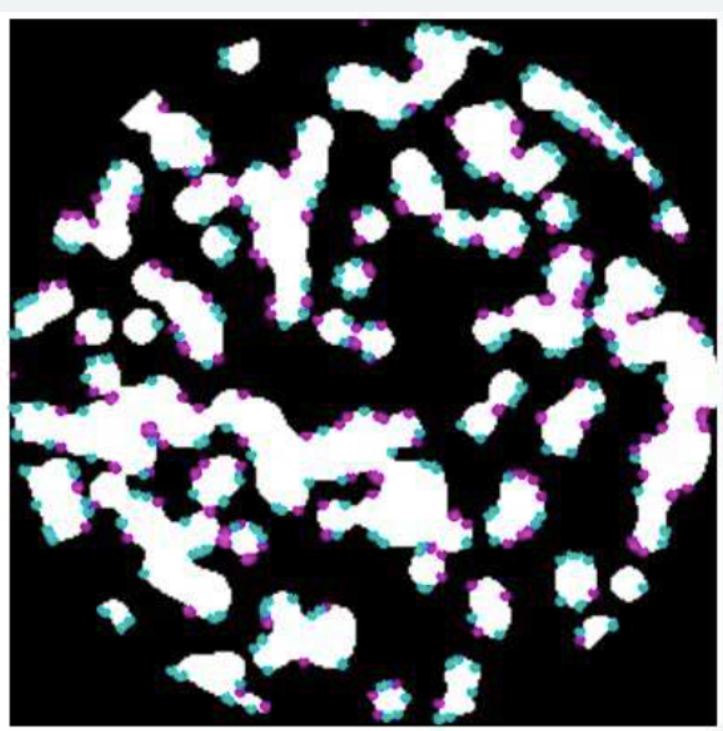
Using Belief Prop.

$\alpha = 1/10, \sigma/L = 0.006$

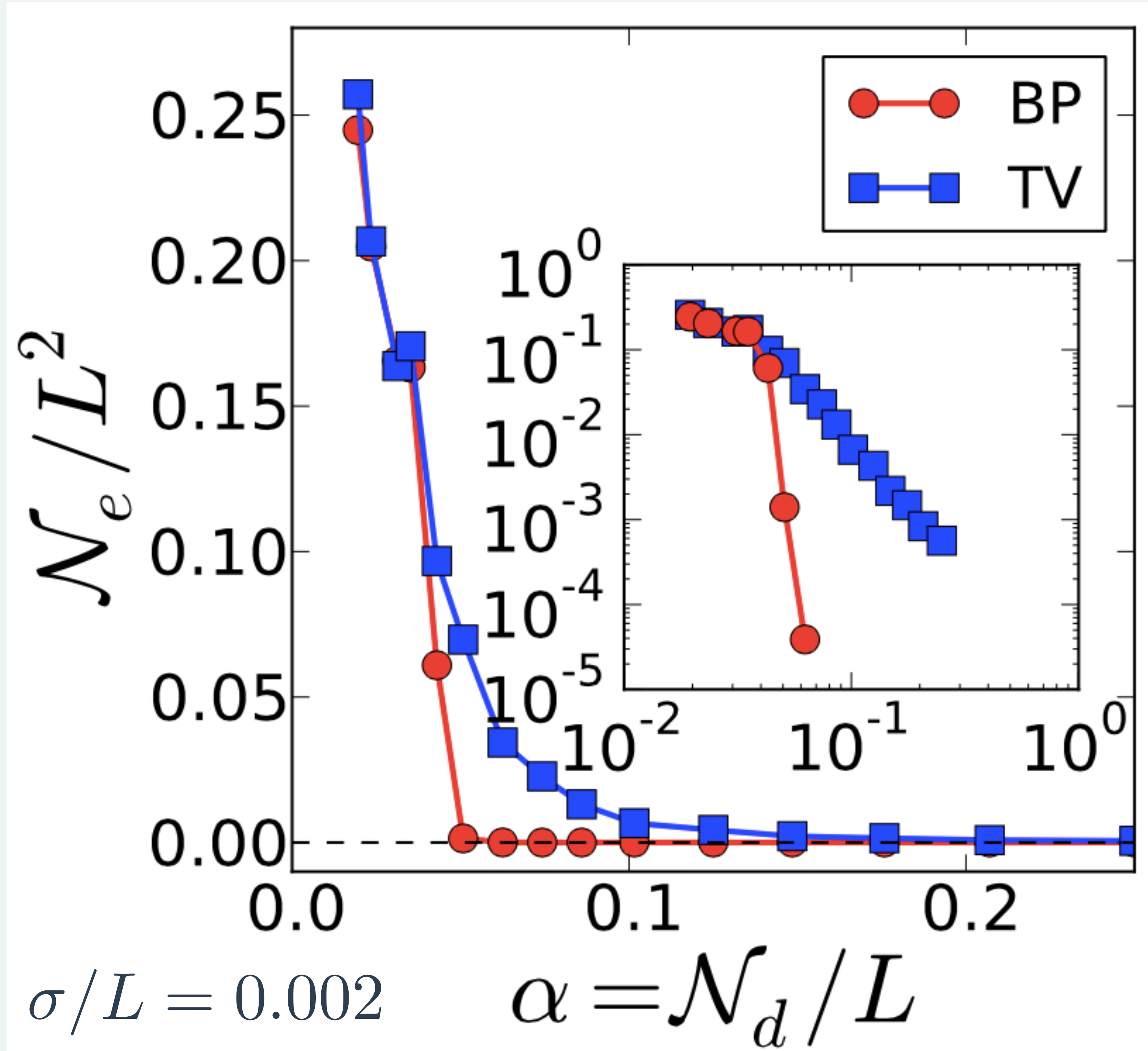
BP



TV



(Gouillart et al, 2013)

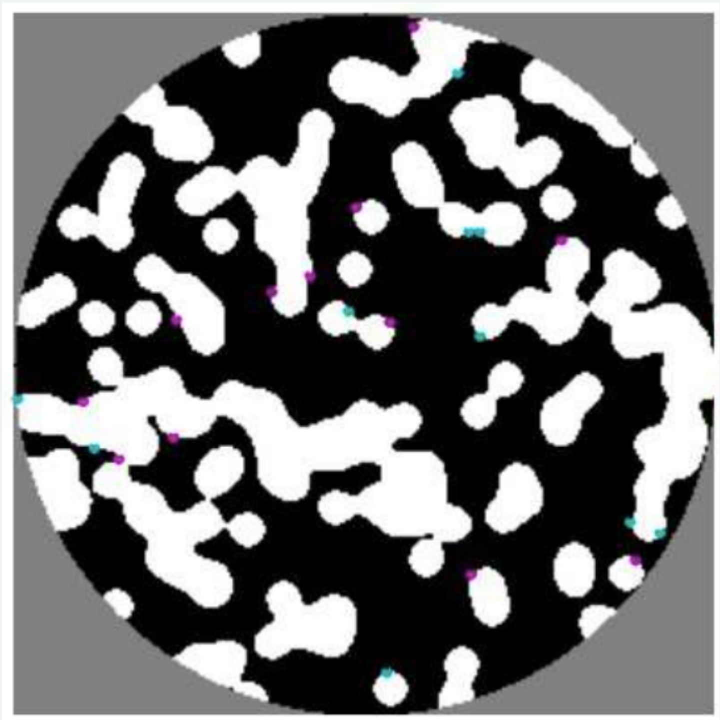


(Gouillart et al, 2013)

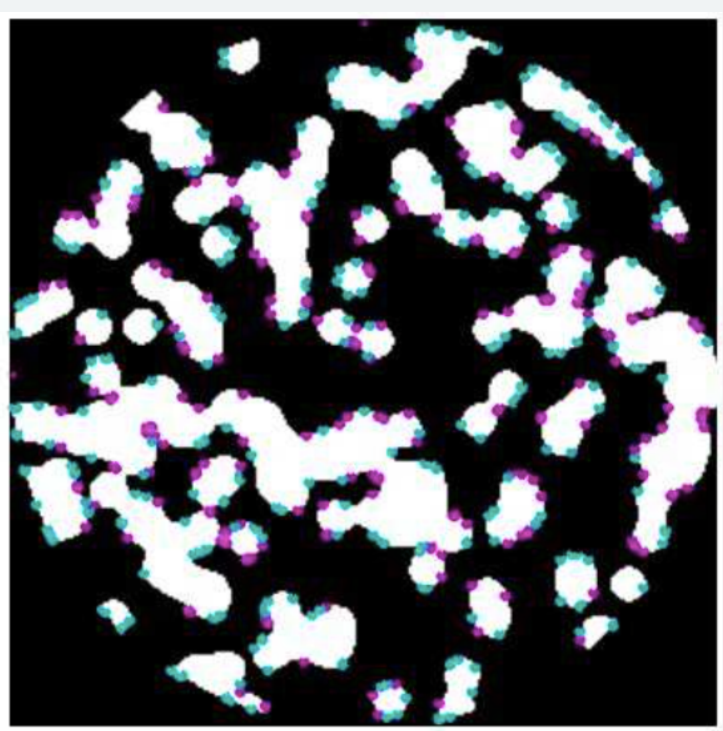
Using Belief Prop.

$\alpha = 1/10, \sigma/L = 0.006$

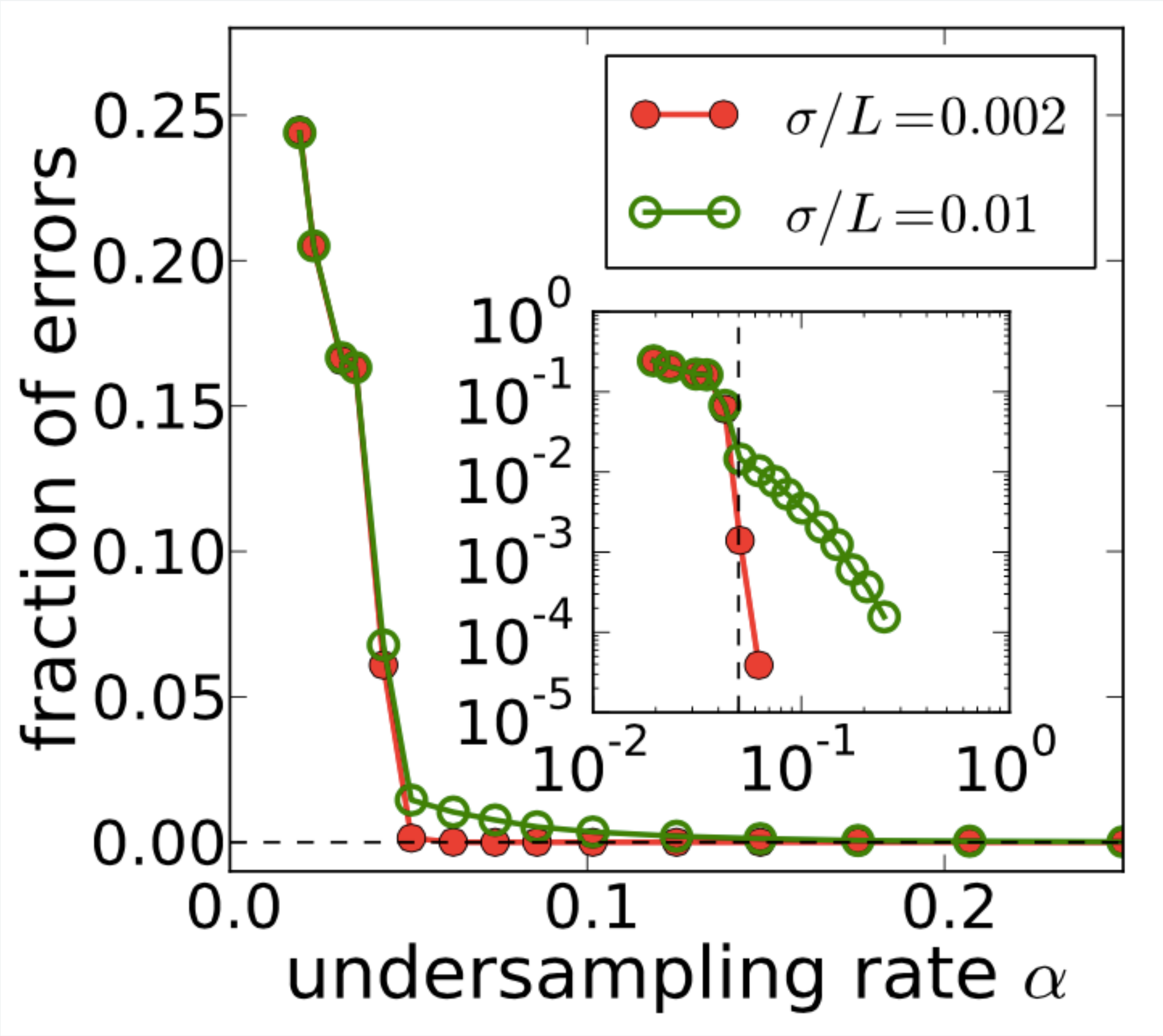
BP



TV



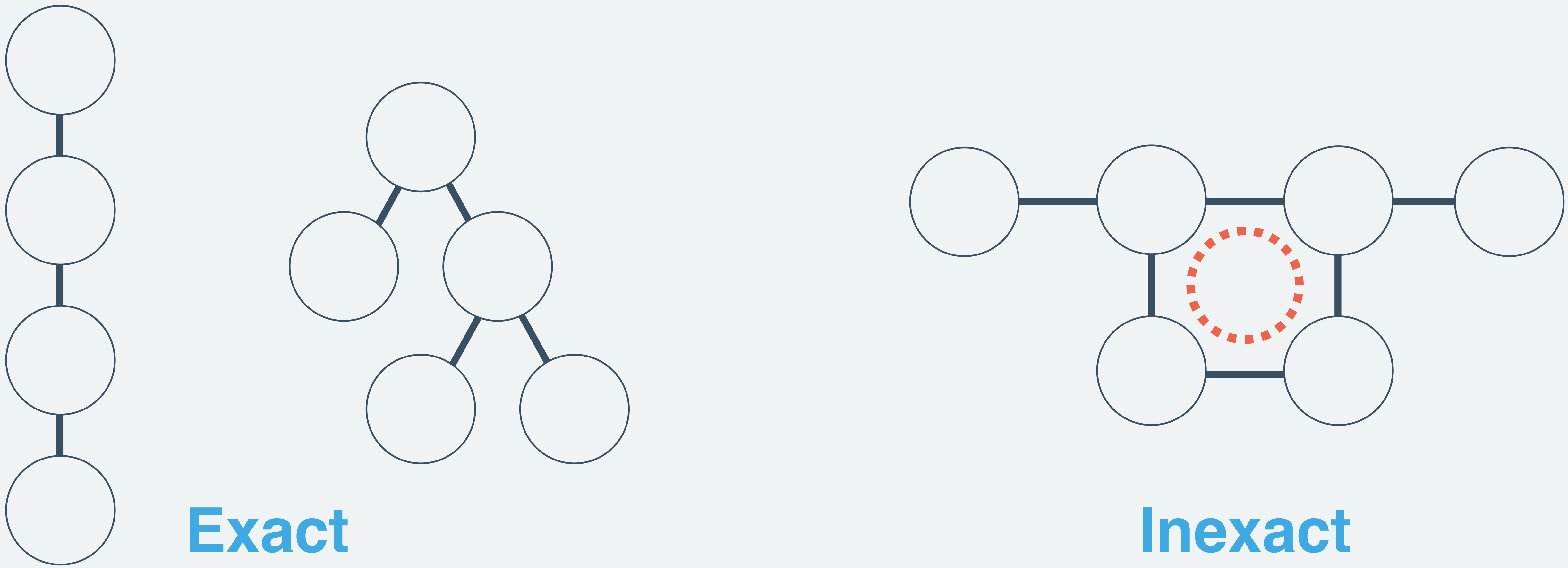
(Gouillart et al, 2013)



(Gouillart et al, 2013)

From Lines to Lattices…

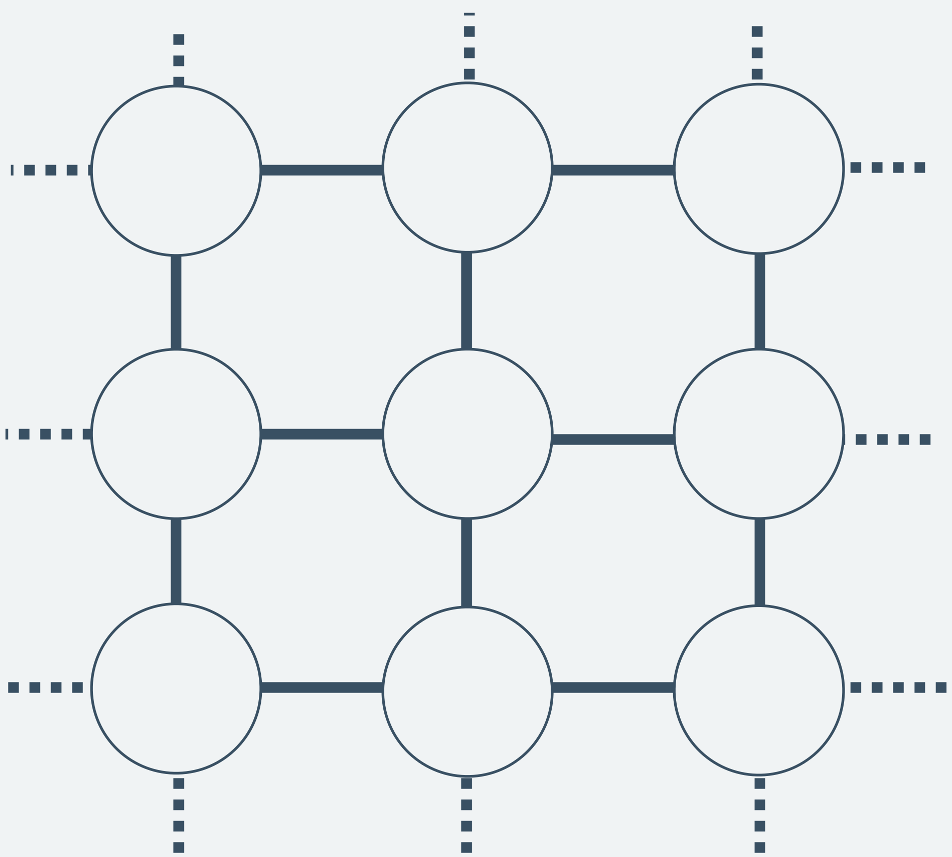
Why Lines? BP known to be exact on trees. Nice properties!



However each line tied to a factor, resulting in many inner BP calculations and a sequential update.

From Lines to Lattices…

A Lattice? A full model of the entire signal that incorporates local correlations. *(related: MRFs)*



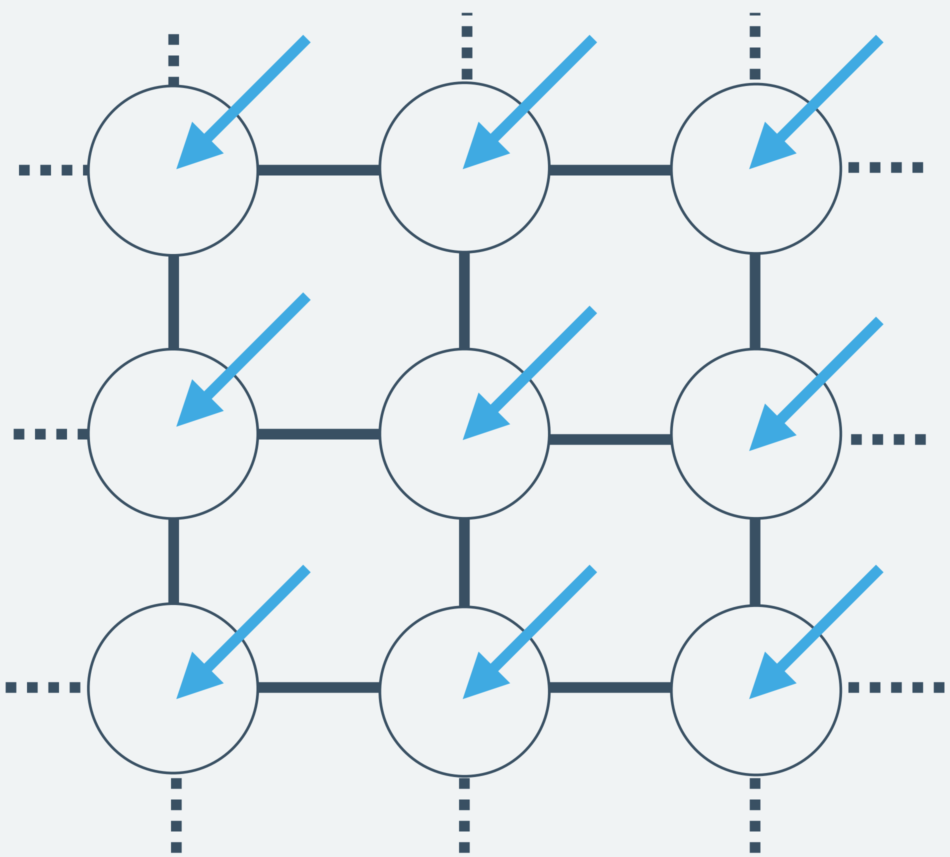
Caution Many tight loops, we cannot expect perfection.

Advantages

- Prior model not tied to the sampling procedure
- Perhaps a more accurate image model
- Adaptable correlation model (edges & weights) that can possibly be trained to exemplars
- Known results from familiar models
- Potentially fewer messages than line model

From Lines to Lattices…

An Ising Model For binary images, we can see that this prior is just a mapping of the square-lattice Ising Model.



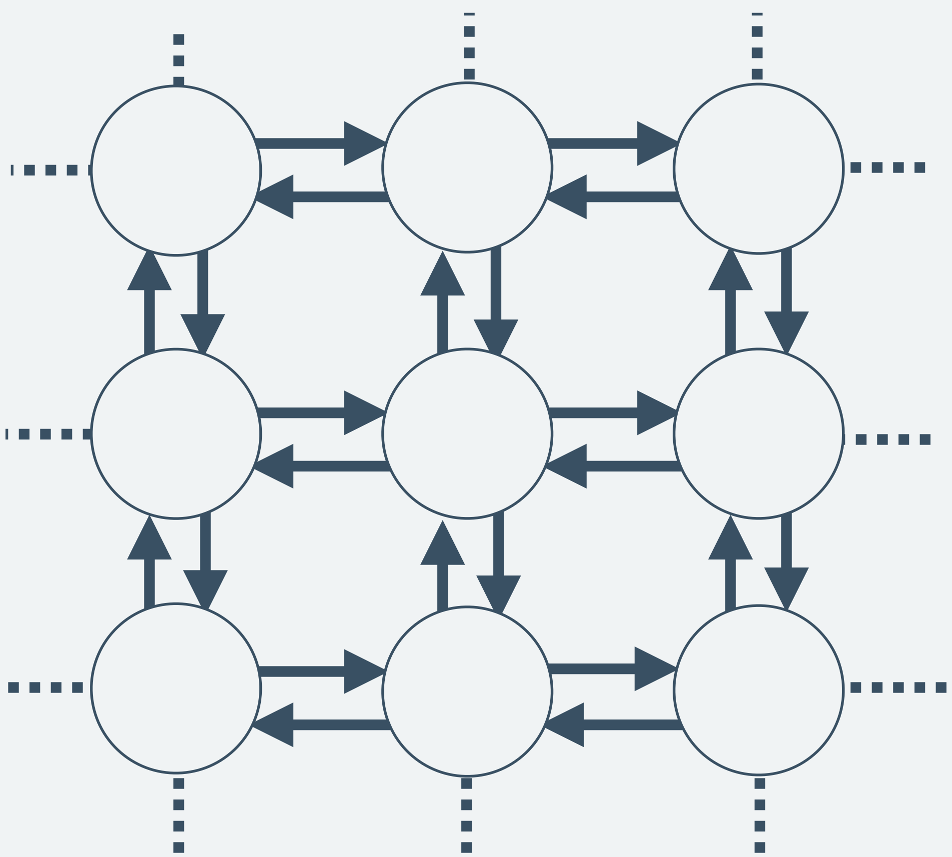
$$P(\mathbf{x}) = \frac{1}{\mathcal{Z}} e^{-\mathcal{H}(\mathbf{x})} \qquad x_i \in \pm 1$$

$$-\mathcal{H}(\mathbf{x}) = \sum_{<i,j>} J_{ij} x_i x_j + \sum_i h_i x_i$$

Some local biasing

Edges & correlation weights encoded in **J.**

…and messages to marginals.



$d = 4$

Potential Efficiency Lattice model involves fewer messages between pixels…

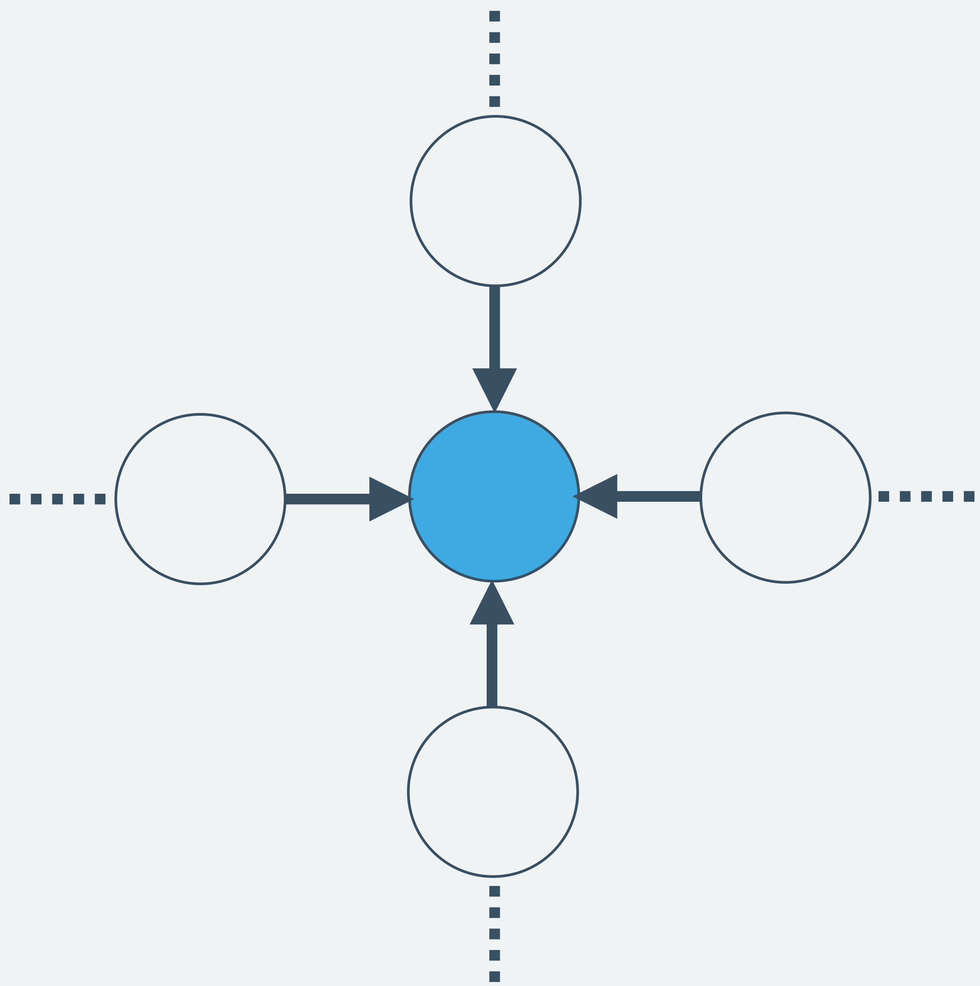
$$O(2dN) < O(4NN_\theta) \quad \text{for} \quad d < N_\theta$$

However, we cannot use the nice Transfer Matrix approach of the linear model.

Already approximate *(LBP)*, **why not approximate more?**

…and messages to marginals.

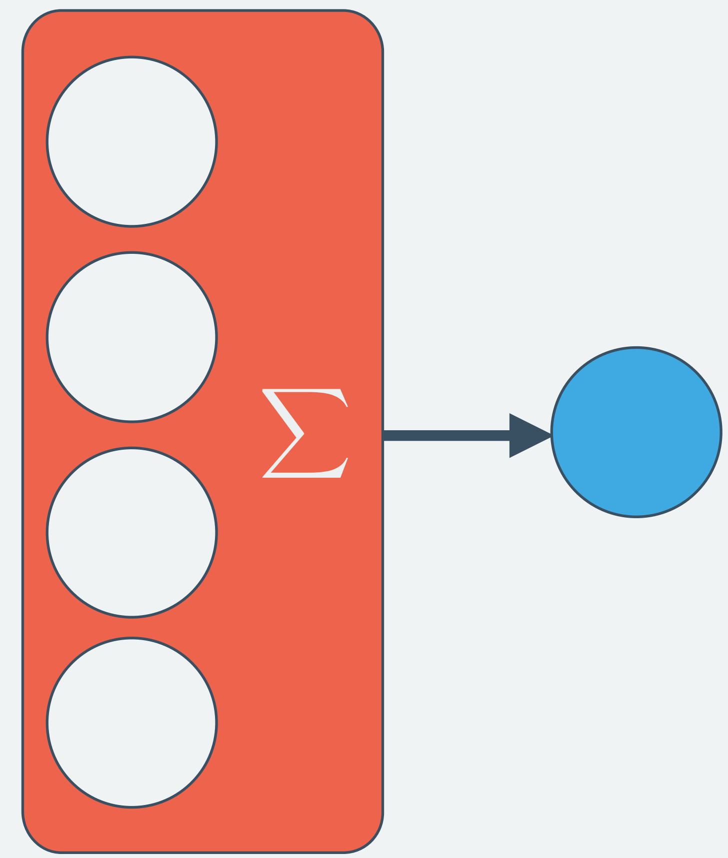
Passing on Edges



$$m_i(x_i) \propto \prod_{j \in \partial_i} m_{j \to i}(x_i)$$

Mean-Field Approximation

Passing on Variables (Marginals)



$$m_i(x_i) \propto \frac{1}{|\partial_i|} \sum_{j \in \partial_i} m_j(x_j)$$

...and messages to marginals.

MFA as an approximation of Partition

When applying the MFA, we are approximating the intractable partition (via its *Free Energy*)...

$$\mathcal{F} = -\log \mathcal{Z} = \sum_{x_1 \in \{0,1\}} \sum_{x_2 \in \{0,1\}} \cdots \sum_{x_N \in \{0,1\}} -\mathcal{H}(\mathbf{x})$$

...by minimizing

$$\begin{aligned}\mathcal{F}^{\text{nmf}} &= -\mathcal{S}(\mathbf{m}) - \mathcal{H}(\mathbf{m}) \\ &= \sum_i \{m_i \ln m_i + (1-m_i)\ln(1-m_i)\} + \sum_i h_i m_i + \sum_{<i,j>} J_{ij} m_i m_j\end{aligned}$$

Promoting greater entropy
(more general)

$$m_i \triangleq \langle x_i \rangle_{m_i(x_i)}$$

…and messages to marginals.

Finding the Factorization

Factorize lattice by minimizing MFA Free Energy…

… leading to a *fixed point iteration*.

$$m_i^{(t+1)} = \text{sigmoid}(h_i + \sum_j J_{ij} m_j^{(t)})$$

$$\therefore m_i^* = \text{sigmoid}(h_i + \sum_j J_{ij} m_j^*)$$

Well-known MFA result leaves much to be desired in terms of accuracy.

…and messages to marginals.

More moments -> More Accurate
Can use the *Thouless-Anderson-Palmer* (TAP)-type approach, tracking variance, also. Via Pfleka expansion assuming small coupling…

$$\mathcal{F}^{\text{TAP}} = -\mathcal{S}(\mathbf{m}) + \sum_i h_i m_i + \sum_{<i,j>} J_{ij} m_i m_j + \frac{1}{2} \sum_{<i,j>} J_{ij}^2 v_i v_j$$

$$v_i = m_i - m_i^2$$

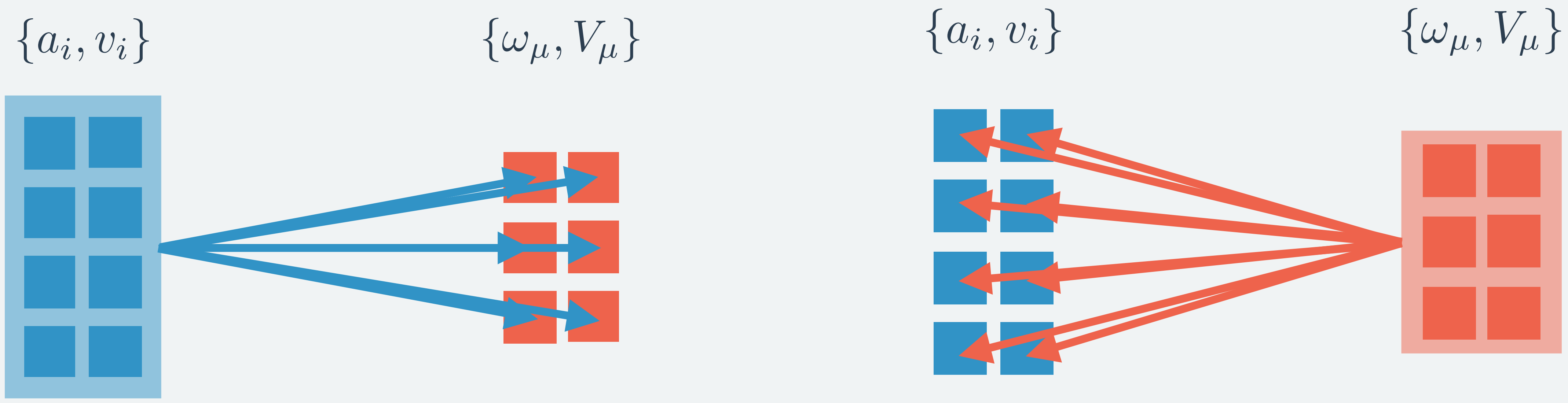
Which gives the FPI…

$$\therefore m_i^{(t+1)} = \text{sigmoid}(h_i + \sum_j J_{ij} m_j^{(t)} + (0.5 - m_i^{(t-1)}) \sum_j J_{ij}^2 v_i^{(t)})$$

...and messages to marginals.

One Step Further

Can we compute the variable-factor messages on the marginals as well?



Approximate Message Passing (AMP)

Used with great success for Compressed Sensing problems and general inference, as well.

- "Simple" FPI
- Direct application of same TAP approximations (but for real variables) to CS factor graph.

TAP-BI + AMP

Bringing it Together

The full iteration including the BI model factorization…

$$V_\mu^{t+1} = \sum_i F_{\mu i}^2 v_i^t$$

$$\omega_\mu^{t+1} = \sum_i F_{\mu i} a_i^t - \frac{V_\mu^{t+1}}{\Delta + V_\mu^t}(y_\mu - \omega_\mu^t)$$

$$(\Sigma_i^{t+1})^2 = \left[\sum_\mu \frac{F_{\mu i}^2}{\Delta + V_\mu^{t+1}}\right]^{-1}$$

$$R_i^{t+1} = a_i^t + (\Sigma_i^{t+1})^2 \sum_\mu F_{\mu i} \frac{(y_\mu - \omega_\mu^{t+1})}{\Delta + V_\mu^{t+1}}$$

Standard AMP Iteration

$$h_i^{t+1} = \frac{(R^{t+1} - 0.5)}{(\Sigma_i^{t+1})^2}$$

$$a_i^{t+1} = \mathrm{sigmoid}(h_i^{t+1} + \sum_j J_{ij} a^t - (0.5 - a^{t-1}) \sum_j J_{ij}^2 v^t)$$

$$v_i^{t+1} = a^{t+1} - (a^{t+1})^2$$

TAP-BI + AMP

Bringing it Together

The full iteration including the BI model factorization…

$$V_\mu^{t+1} = \sum_i F_{\mu i}^2 v_i^t$$

$$\omega_\mu^{t+1} = \sum_i F_{\mu i} a_i^t - \frac{V_\mu^{t+1}}{\Delta + V_\mu^t}(y_\mu - \omega_\mu^t)$$

$$(\Sigma_i^{t+1})^2 = \left[\sum_\mu \frac{F_{\mu i}^2}{\Delta + V_\mu^{t+1}}\right]^{-1}$$

$$R_i^{t+1} = a_i^t + (\Sigma_i^{t+1})^2 \sum_\mu F_{\mu i} \frac{(y_\mu - \omega_\mu^{t+1})}{\Delta + V_\mu^{t+1}}$$

$$h_i^{t+1} = \frac{(R^{t+1} - 0.5)}{(\Sigma_i^{t+1})^2}$$

$$a_i^{t+1} = \mathrm{sigmoid}(h_i^{t+1} + \sum_j J_{ij} a^t - (0.5 - a^{t-1}) \sum_j J_{ij}^2 v^t)$$

$$v_i^{t+1} = a^{t+1} - (a^{t+1})^2$$

Calculate fields from AMP

TAP-BI + AMP

Bringing it Together

The full iteration including the BI model factorization…

$$V_\mu^{t+1} = \sum_i F_{\mu i}^2 v_i^t$$

$$\omega_\mu^{t+1} = \sum_i F_{\mu i} a_i^t - \frac{V_\mu^{t+1}}{\Delta + V_\mu^t}(y_\mu - \omega_\mu^t)$$

$$(\Sigma_i^{t+1})^2 = \left[\sum_\mu \frac{F_{\mu i}^2}{\Delta + V_\mu^{t+1}}\right]^{-1}$$

$$R_i^{t+1} = a_i^t + (\Sigma_i^{t+1})^2 \sum_\mu F_{\mu i} \frac{(y_\mu - \omega_\mu^{t+1})}{\Delta + V_\mu^{t+1}}$$

$$h_i^{t+1} = \frac{(R^{t+1} - 0.5)}{(\Sigma_i^{t+1})^2}$$

$$a_i^{t+1} = \text{sigmoid}(h_i^{t+1} + \sum_j J_{ij} a^t - (0.5 - a^{t-1}) \sum_j J_{ij}^2 v^t)$$

$$v_i^{t+1} = a^{t+1} - (a^{t+1})^2$$

Update

Binary Ising Factorization

TAP-BI + AMP

Bringing it Together

A full iteration including the BI model factorization…

$$V_\mu^{t+1} = \sum_i F_{\mu i}^2 v_i^t$$

$$\omega_\mu^{t+1} = \sum_i F_{\mu i} a_i^t - \frac{V_\mu^{t+1}}{\Delta + V_\mu^t}(y_\mu - \omega_\mu^t)$$

$$(\Sigma_i^{t+1})^2 = \left[\sum_\mu \frac{F_{\mu i}^2}{\Delta + V_\mu^{t+1}}\right]^{-1}$$

$$R_i^{t+1} = a_i^t + (\Sigma_i^{t+1})^2 \sum_\mu F_{\mu i} \frac{(y_\mu - \omega_\mu^{t+1})}{\Delta + V_\mu^{t+1}}$$

$$h_i^{t+1} = \frac{(R^{t+1} - 0.5)}{(\Sigma_i^{t+1})^2}$$

$$a_i^{t+1} = \mathrm{sigmoid}(h_i^{t+1} + \sum_j J_{ij} a^t - (0.5 - a^{t-1}) \sum_j J_{ij}^2 v^t)$$

$$v_i^{t+1} = a^{t+1} - (a^{t+1})^2$$

Repeat until some criterion met, like

- Uncertainty
- Residual
- Convergence of factorization

TAP-BI + AMP

Bringing it Together

A full iteration including the BI model factorization…

$$V_\mu^{t+1} = \sum_i F_{\mu i}^2 v_i^t$$

$$\omega_\mu^{t+1} = \sum_i F_{\mu i} a_i^t - \frac{V_\mu^{t+1}}{\Delta + V_\mu^t}(y_\mu - \omega_\mu^t)$$

$$(\Sigma_i^{t+1})^2 = \left[\sum_\mu \frac{F_{\mu i}^2}{\Delta + V_\mu^{t+1}}\right]^{-1}$$

$$R_i^{t+1} = a_i^t + (\Sigma_i^{t+1})^2 \sum_\mu F_{\mu i} \frac{(y_\mu - \omega_\mu^{t+1})}{\Delta + V_\mu^{t+1}}$$

$$h_i^{t+1} = \frac{(R^{t+1} - 0.5)}{(\Sigma_i^{t+1})^2}$$

$$a_i^{t+1} = \mathrm{sigmoid}(h_i^{t+1} + \sum_j J_{ij} a^t - (0.5 - a^{t-1}) \sum_j J_{ij}^2 v^t)$$

$$v_i^{t+1} = a^{t+1} - (a^{t+1})^2$$

Some Nuance One can update noise variance to improve convergence…

$$\Delta^t = \frac{1}{M} ||\mathbf{y} - F\mathbf{a}^t||_2^2$$

TAP-BI + AMP

Bringing it Together

A full iteration including the BI model factorization…

$$V_\mu^{t+1} = \sum_i F_{\mu i}^2 v_i^t$$

$$\omega_\mu^{t+1} = \sum_i F_{\mu i} a_i^t - \frac{V_\mu^{t+1}}{\Delta + V_\mu^t}(y_\mu - \omega_\mu^t)$$

$$(\Sigma_i^{t+1})^2 = \left[\sum_\mu \frac{F_{\mu i}^2}{\Delta + V_\mu^{t+1}}\right]^{-1}$$

$$R_i^{t+1} = a_i^t + (\Sigma_i^{t+1})^2 \sum_\mu F_{\mu i} \frac{(y_\mu - \omega_\mu^{t+1})}{\Delta + V_\mu^{t+1}}$$

$$h_i^{t+1} = \frac{(R^{t+1} - 0.5)}{(\Sigma_i^{t+1})^2}$$

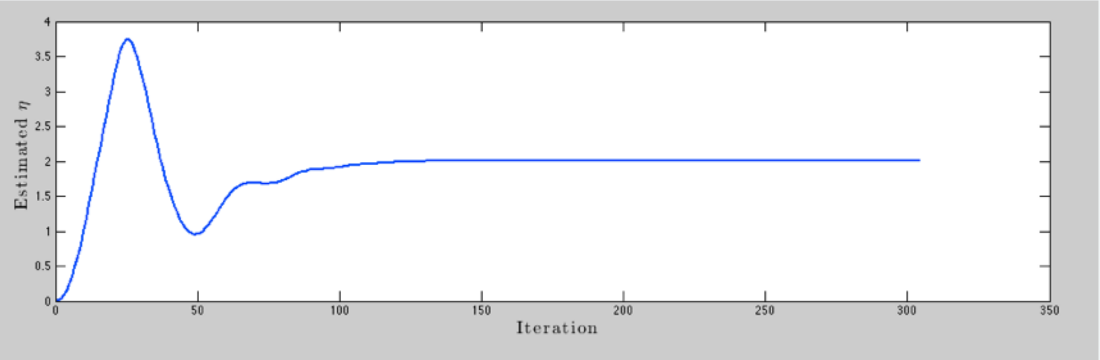
$$a_i^{t+1} = \mathrm{sigmoid}(h_i^{t+1} + \sum_j J_{ij} a^t - (0.5 - a^{t-1}) \sum_j J_{ij}^2 v^t)$$

$$v_i^{t+1} = a^{t+1} - (a^{t+1})^2$$

Some Nuance Also, one can update the coupling strength…

$$J_{ij}^t \triangleq \eta^t E_{ij}$$

$$\eta^{t+1} = \frac{1}{N} \sum_{<i,j>} J_{i,j}^t a_i^t a_j^t$$



TAP-BI + AMP

Bringing it Together

A full iteration including the BI model factorization…

$$V_\mu^{t+1} = \sum_i F_{\mu i}^2 v_i^t$$

$$\omega_\mu^{t+1} = \sum_i F_{\mu i} a_i^t - \frac{V_\mu^{t+1}}{\Delta + V_\mu^t}(y_\mu - \omega_\mu^t)$$

$$(\Sigma_i^{t+1})^2 = \left[\sum_\mu \frac{F_{\mu i}^2}{\Delta + V_\mu^{t+1}}\right]^{-1}$$

$$R_i^{t+1} = a_i^t + (\Sigma_i^{t+1})^2 \sum_\mu F_{\mu i} \frac{(y_\mu - \omega_\mu^{t+1})}{\Delta + V_\mu^{t+1}}$$

$$h_i^{t+1} = \frac{(R^{t+1} - 0.5)}{(\Sigma_i^{t+1})^2}$$

$$a_i^{t+1} = \text{sigmoid}(h_i^{t+1} + \sum_j J_{ij} a^t - (0.5 - a^{t-1}) \sum_j J_{ij}^2 v^t)$$

$$v_i^{t+1} = a^{t+1} - (a^{t+1})^2$$

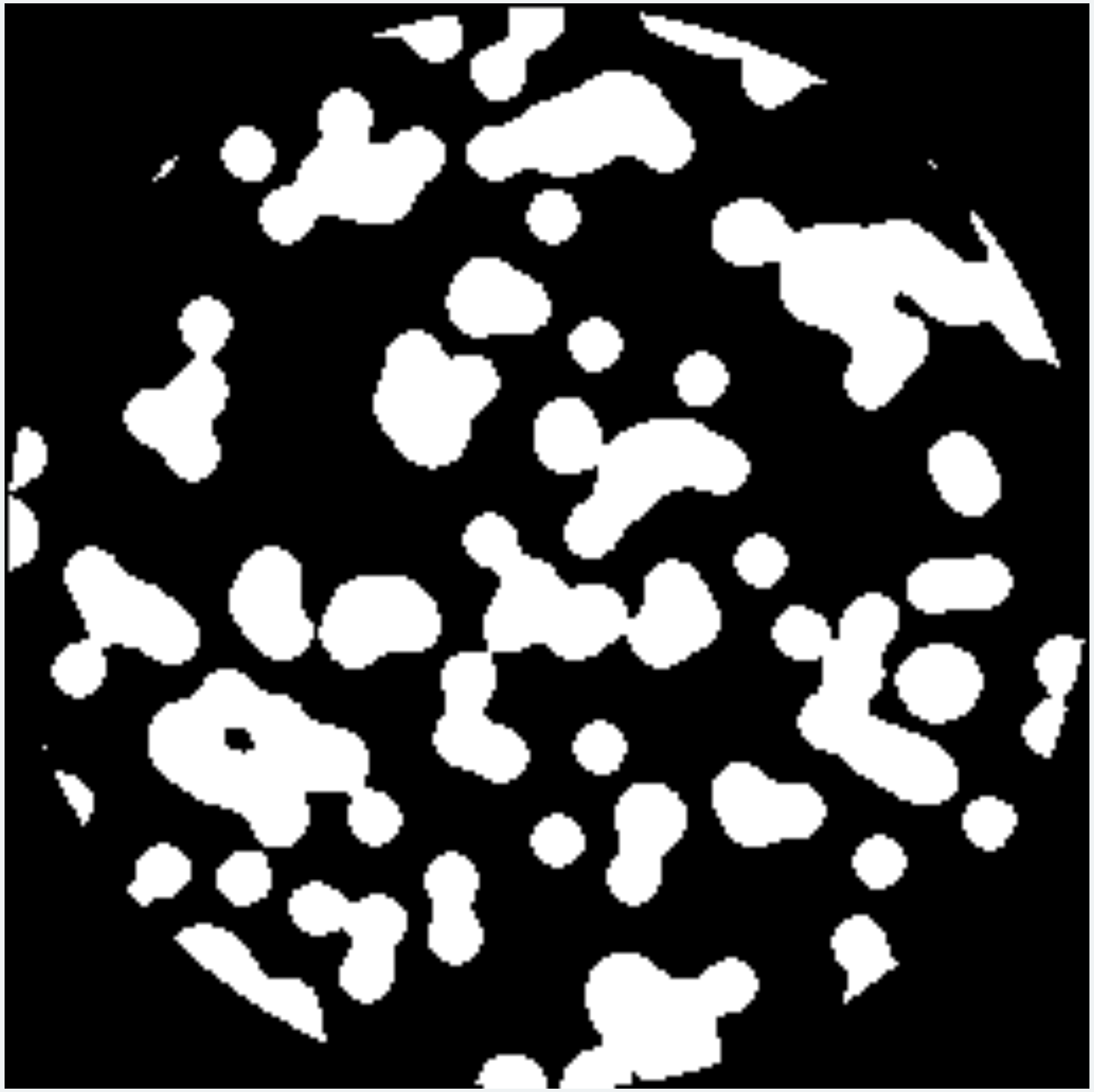
Some Nuance Damping can be necessary…

$$a^{t+1} = \beta a^t + (1-\beta) a^{t+1}$$

TAP-BI + AMP

A Small Experiment…

Target Image



256

256

Parameters

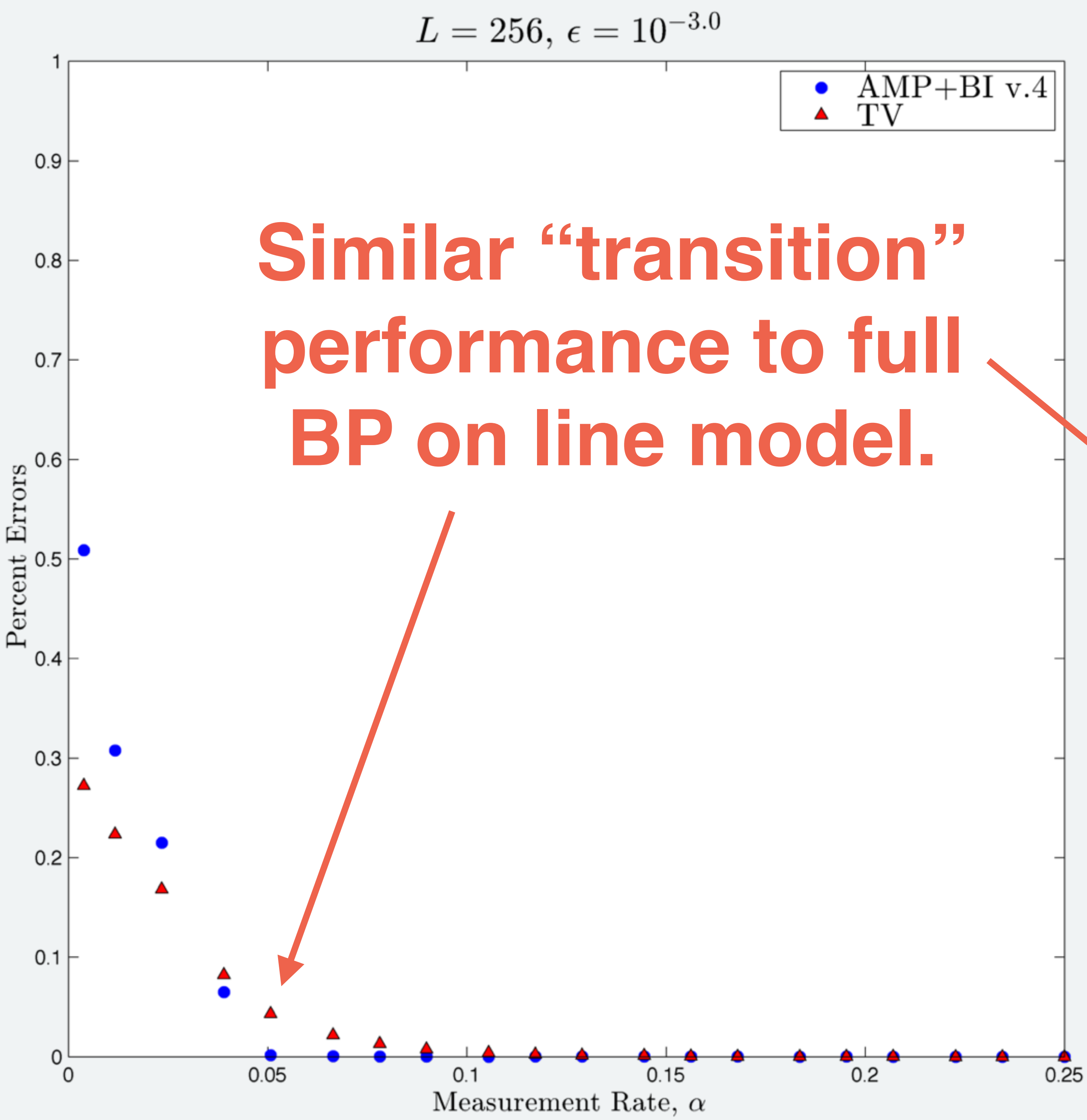
$$\epsilon = \Delta/L \in [10^{-2}, 10^{-3}]$$

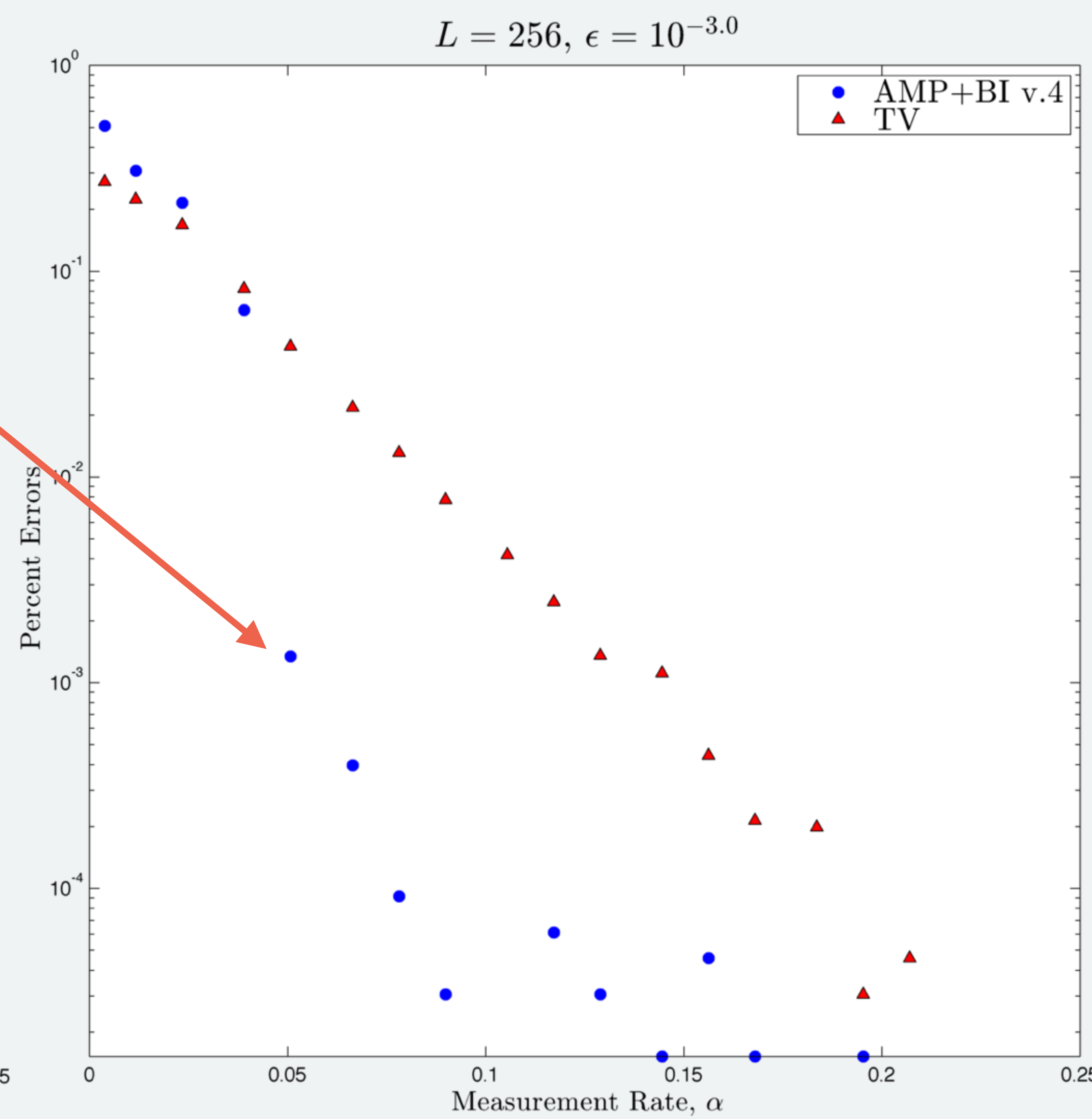
$$\alpha = M/N = N_\theta/L \in (0, 0.25]$$

$$d = 8$$

TAP-BI + AMP

Small Noise Variance

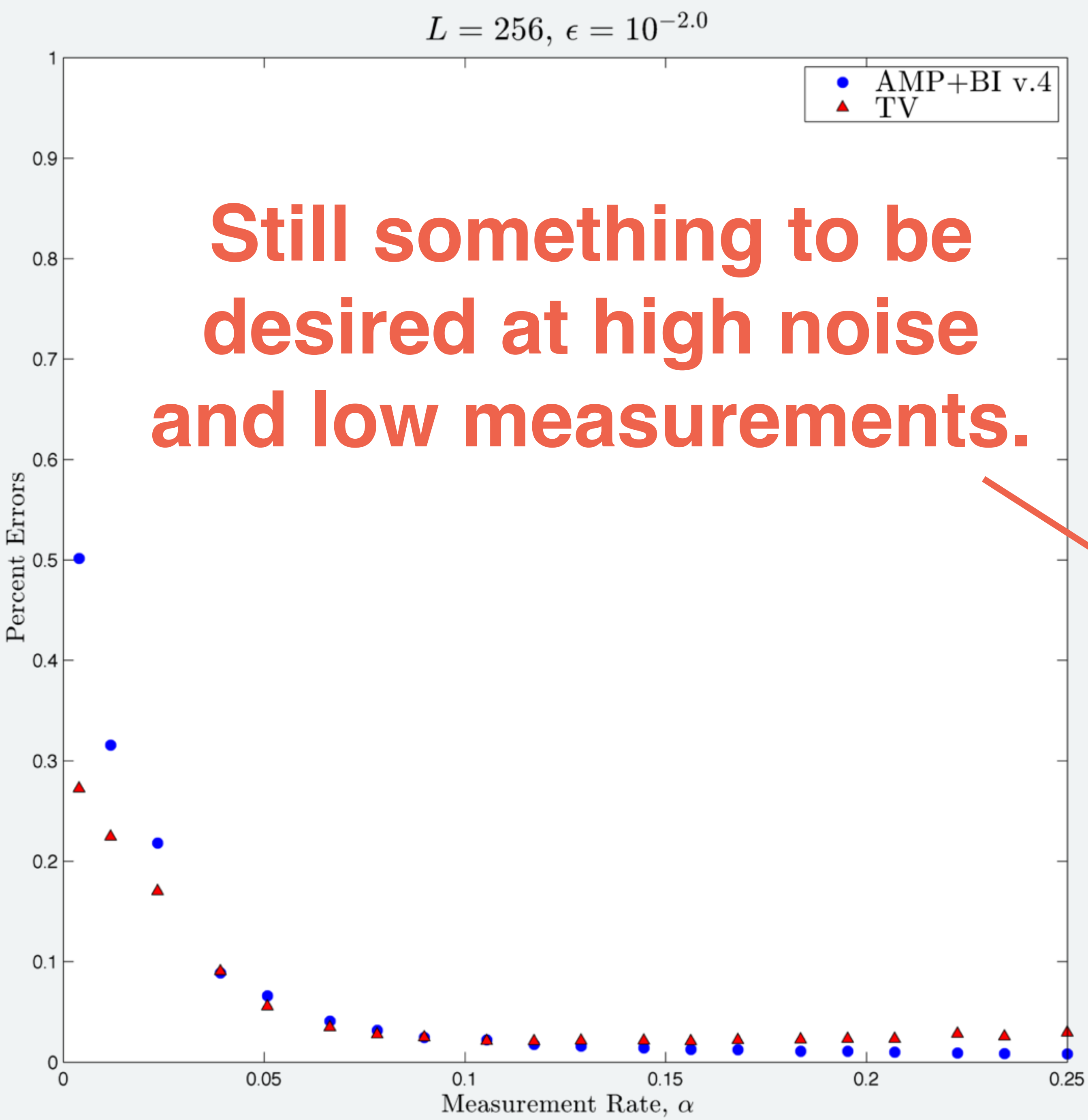




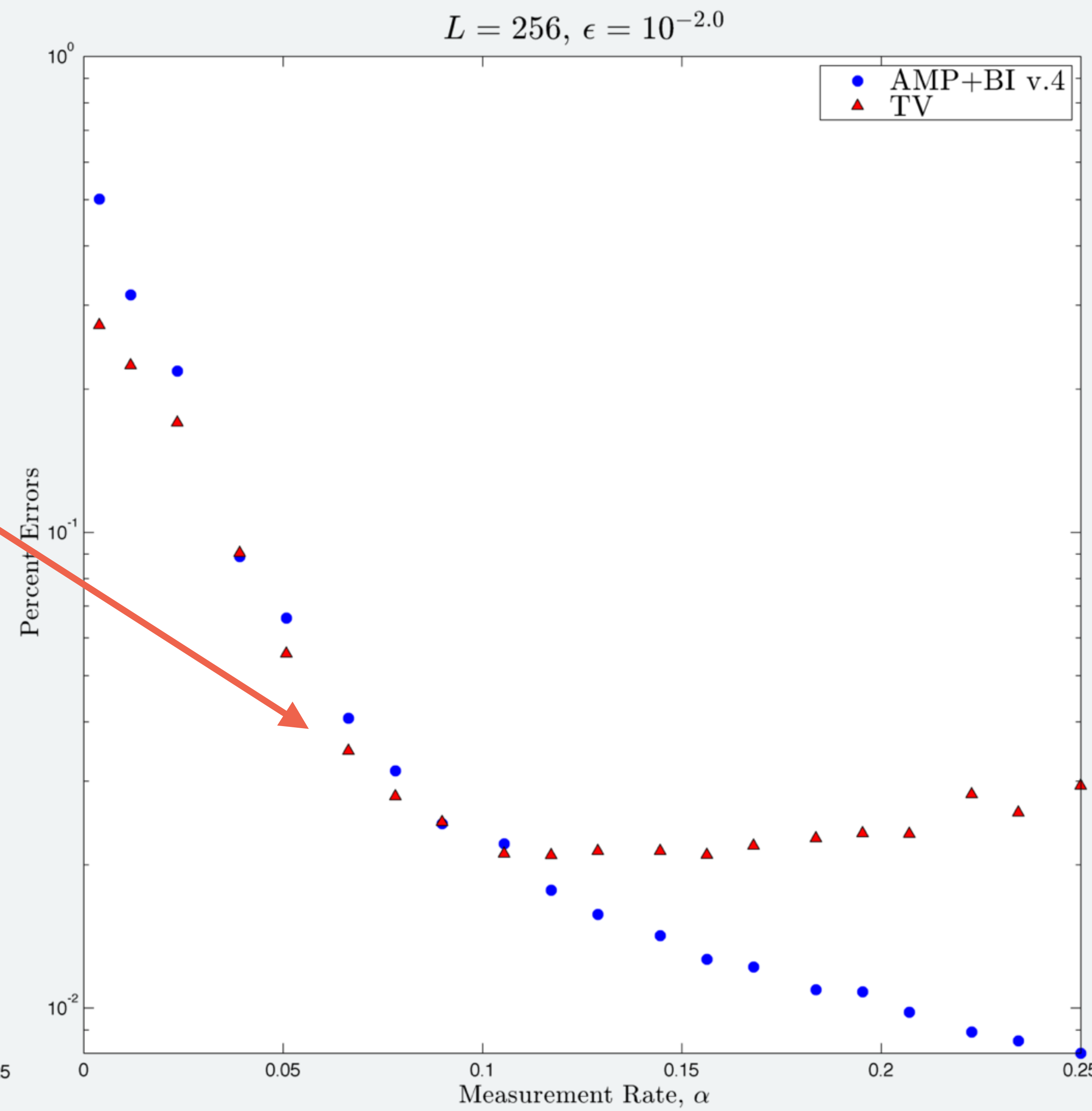
Similar “transition” performance to full BP on line model.

TAP-BI + AMP

Large Noise Variance



Still something to be desired at high noise and low measurements.



Looking Forward

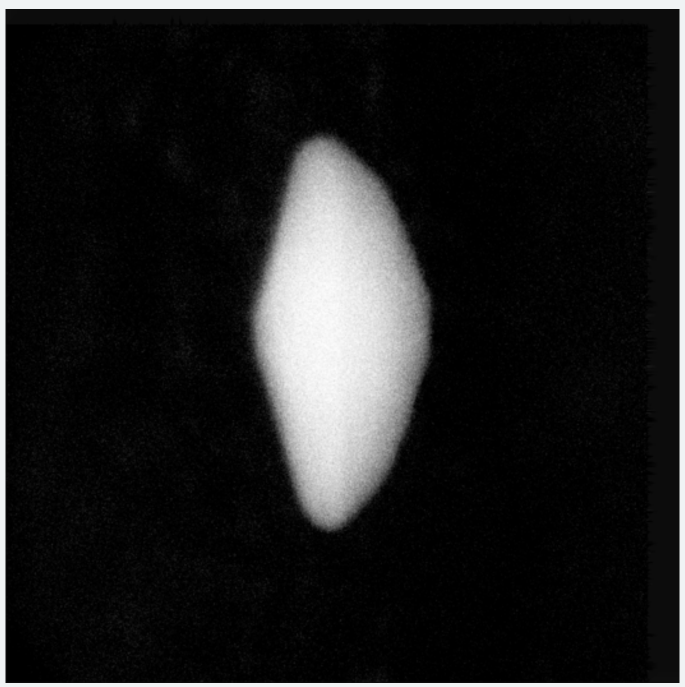
Much work to do…

- What are the effects of learning free parameters?
  - Is there a better update scheme for these?
- Optimal stopping criterion?
- How to choose damping? Adaptive scheme based on free energy?
- Will these changes help high-noise performance?

Looking Forward

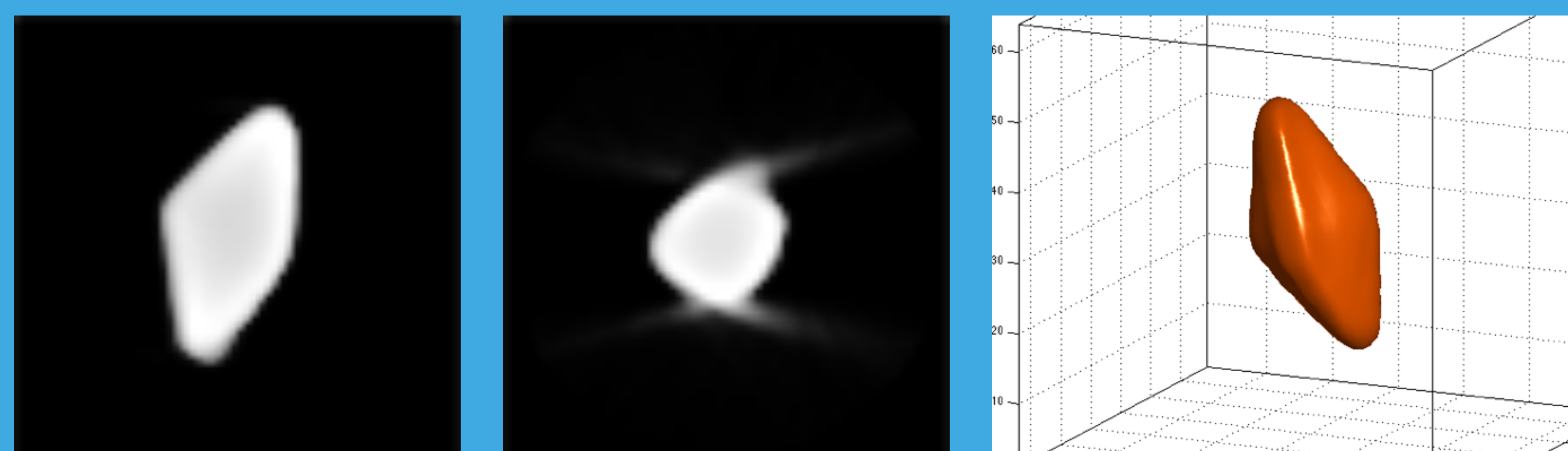
Much work to do…

- Extension to Potts…some preliminary work applied to limited-angle electron tomography



Au Bipyramid

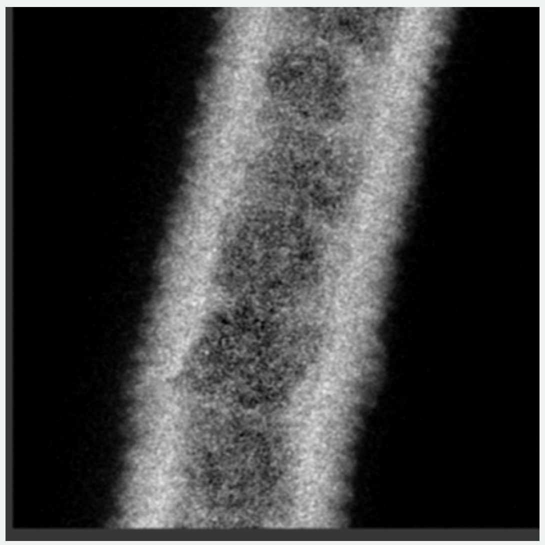
Potts + AMP



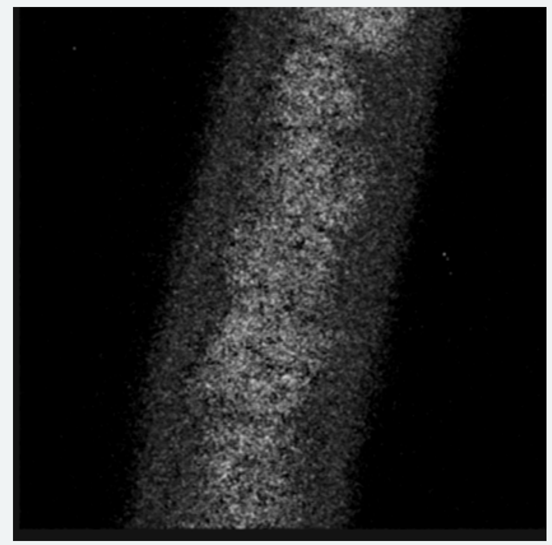
Looking Forward

Much work to do…

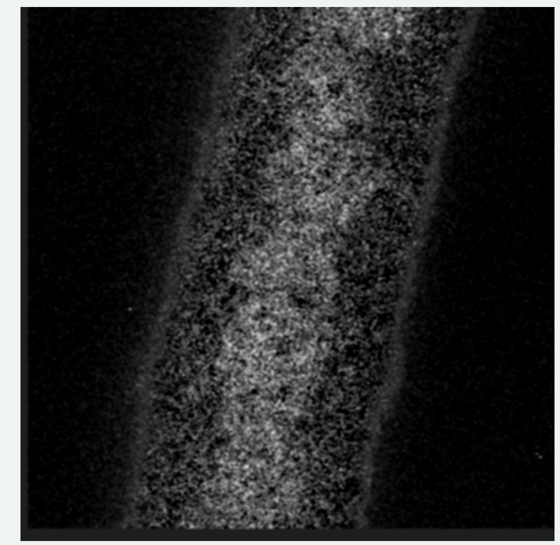
- Extension to Potts…some preliminary work applied to limited-angle analytic electron tomography



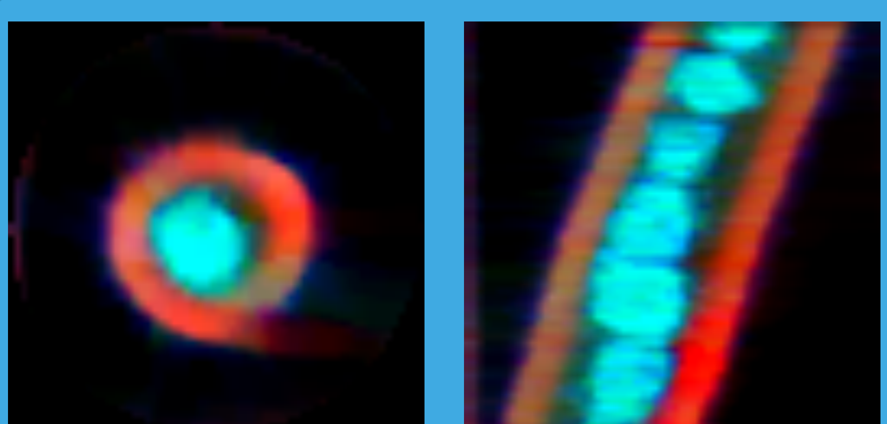
Carbon



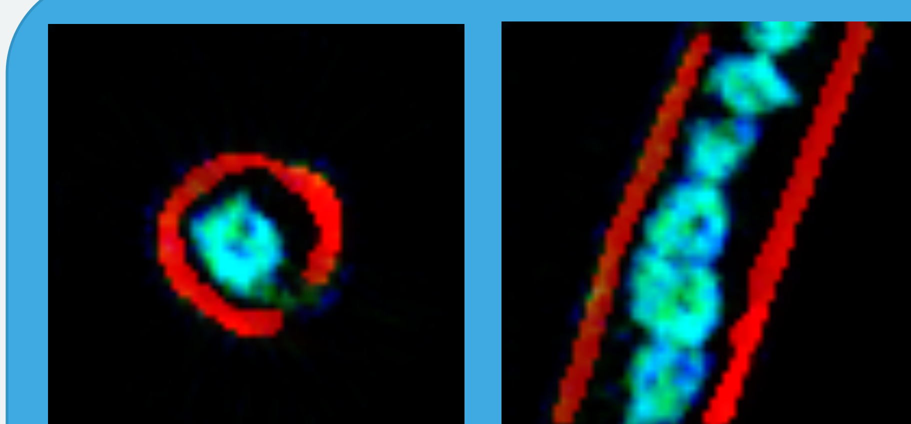
Cobalt



Oxygen



Total Variation



Potts + AMP

Questions?

Thanks!

Scratch