ECE521: Lecture 17

16 March 2017

Markov Random Fields, Factor graphs

With thanks to Brendan Frey and others

This week

- Exploring both types of graphical model (directed and undirected)

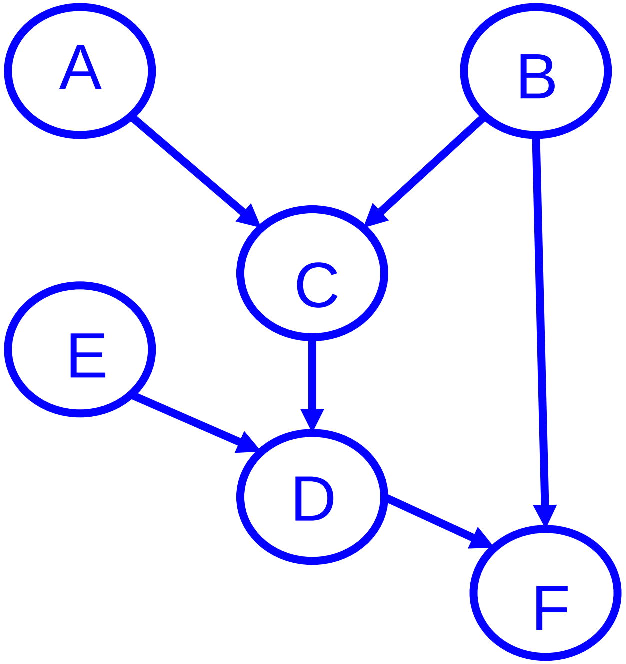
- Examples of additional perspectives:
    - Bishop 2006: parts of chap. 8
    - Murphy 2012: parts of chap. 10
    - Russell & Norvig, 2009: parts of chap. 14 *(Artificial Intelligence: A Modern Approach*)

Outline for today

- Finish: conditional independence in BNs
- Markov Random Fields
- Factor graphs

Finishing Lecture 16

- A true-false quiz:

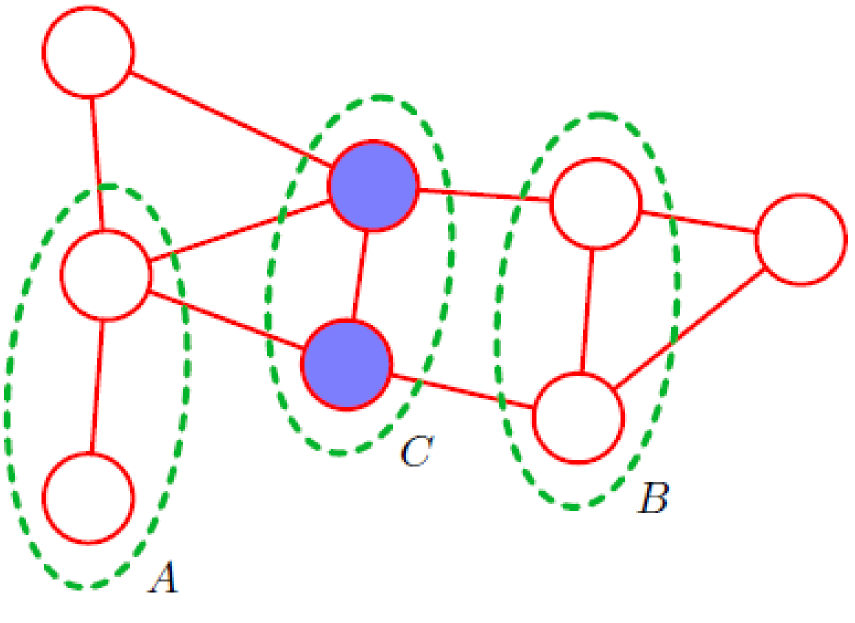


1. $A \perp\!\!\!\perp F$
2. $A \perp\!\!\!\perp D$
3. $A \perp\!\!\!\perp D \mid C$
4. $A \perp\!\!\!\perp D \mid C, F$
5. $A \perp\!\!\!\perp D \mid C, F, B$

AND show that #3 would be implied by $A \perp\!\!\!\perp D, E \mid C$

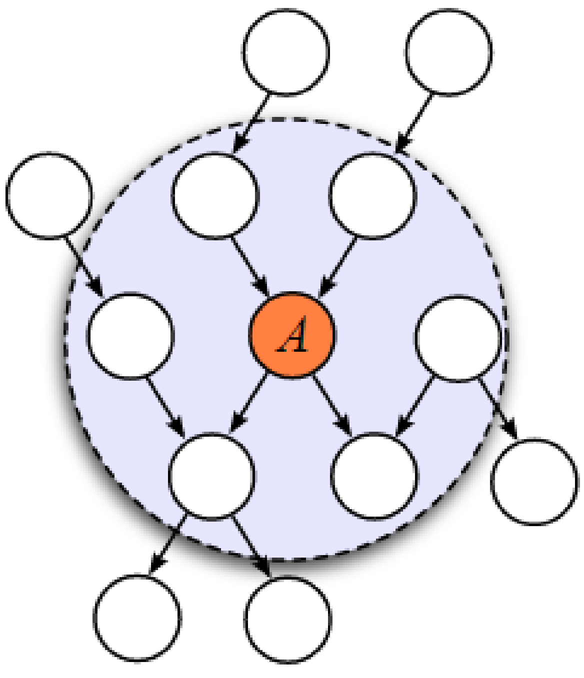
Undirected Graphical Models

- a.k.a. Markov Random Fields
- Conditional independence is easier to assess: e.g. to see if A $\perp\!\!\!\perp$ B | C, remove the given nodes (C) and see if A and B can connect

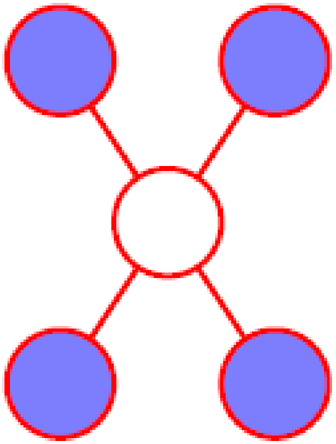


Comparing Markov blankets

for BNs:



for MRFs:



Comparing factorizations

for BNs:

$$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)$$

Easy interpretation!

But not as popular on 2D grids etc

for MRFs:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{C} \psi_C(\mathbf{x}_C)$$

Normalization constant:

$$Z = \sum_{\mathbf{x}} \prod_{C} \psi_C(\mathbf{x}_C)$$

over some *maximal cliques**

* Clique = fully connected subset of nodes

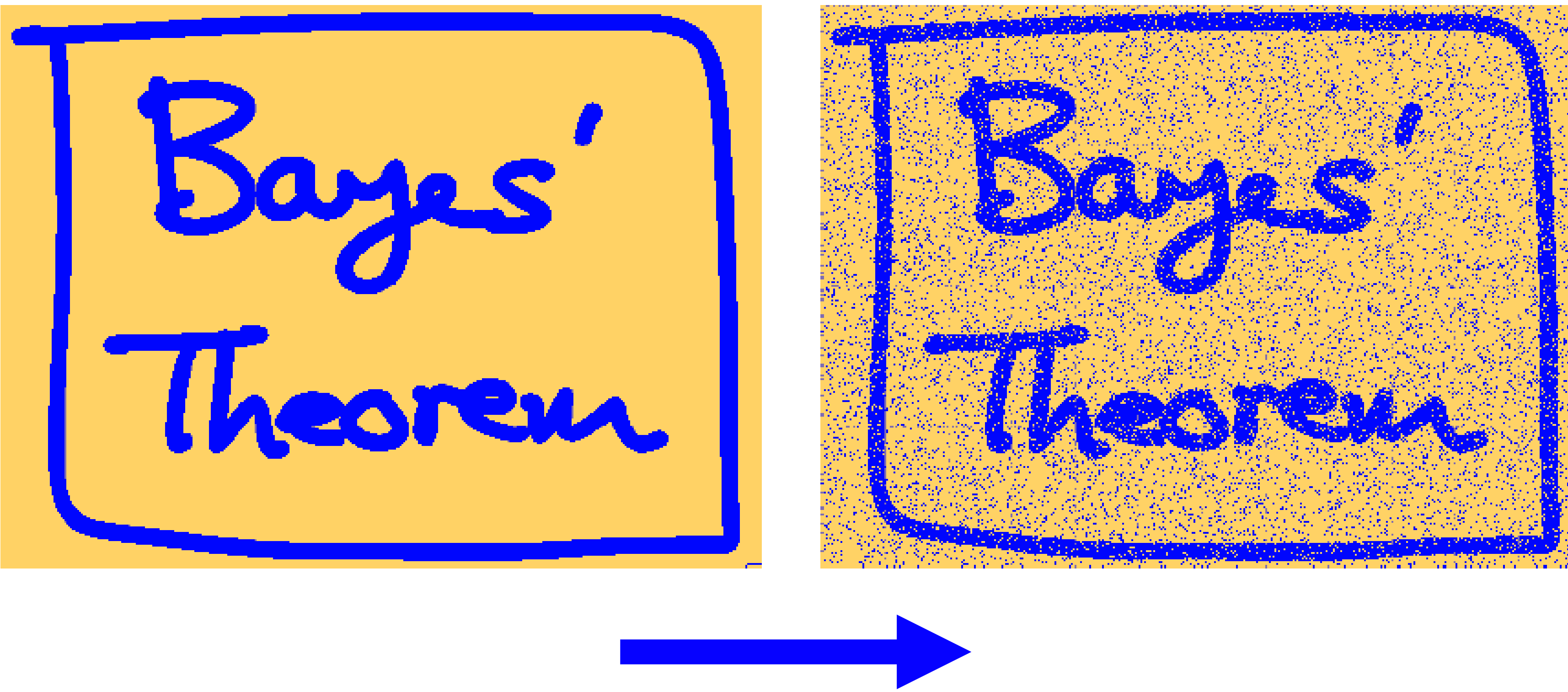
Maximal clique = a clique such that adding any new node from the graph would result in a non-clique

The Boltzmann distribution

- A convenient choice for the potential functions $\psi_C(\mathbf{x}_C)$ is the Boltzmann distribution: $\psi_C(\mathbf{x}_C) = \exp\{-E(\mathbf{x}_C)\}$

- *E* is the energy function
- **x** describes all states in the network, e.g. **x =** {B=1, E=0, A=1, J=0, M=1}
- $\mathbf{x}_C$ is a subset, e.g. $\mathbf{x}_3$ **=** {J=0, A=1}

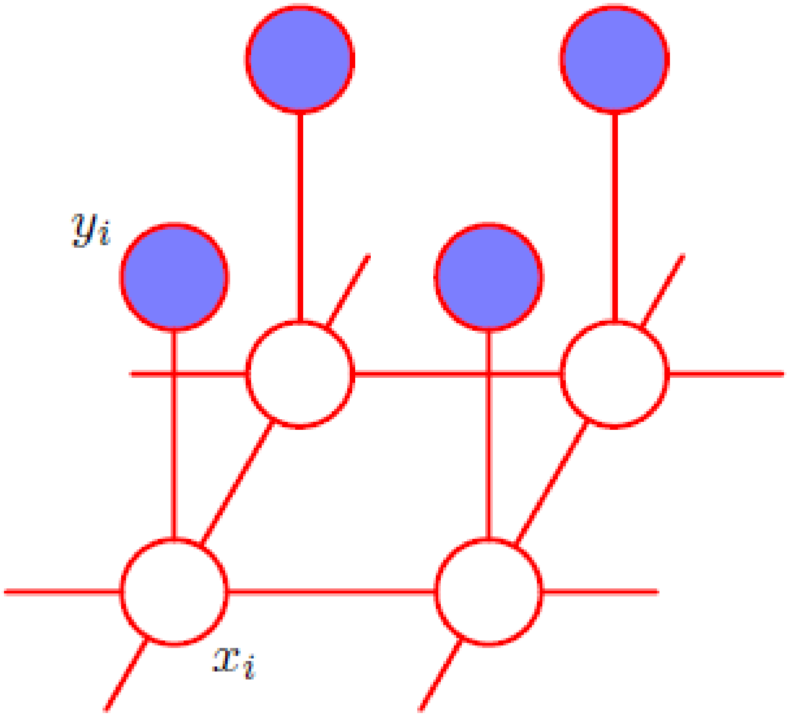
Example: image de-noising



noise = randomly change 10% of pixels

Choice of MRF: the Ising model

- x_i : original image
- y_i : noisy image



$$p(\mathbf{x}, \mathbf{y}) = \frac{1}{Z} \exp\{-E(\mathbf{x}, \mathbf{y})\}$$

where

$$E(\mathbf{x}, \mathbf{y}) = h \sum_i x_i - \beta \sum_{\{i,j\}} x_i x_j - \eta \sum_i x_i y_i$$

Energy function

$$E(\mathbf{x}, \mathbf{y}) = h \sum_i x_i - \beta \sum_{\{i,j\}} x_i x_j - \eta \sum_i x_i y_i$$

bias — neighbours — observed

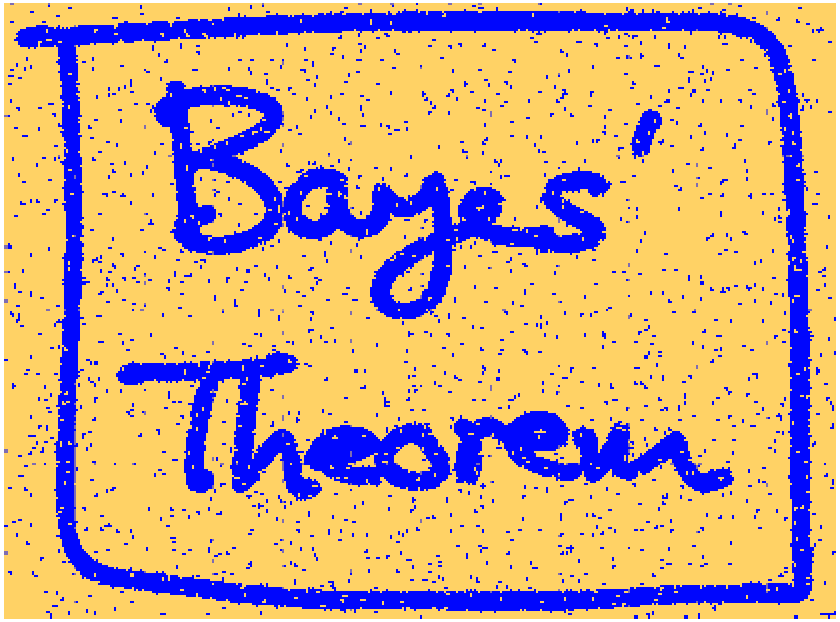
- The relative values of *h*, *β*, and *η* control these three effects
- What are the maximal cliques in an Ising model?

Solving the Ising model

- Select $\beta = 1.0, \eta = 2.1$ and $h = 0$
- Initialize **x** to **y**
- Until convergence:

  for each x_i :

  $$x_i \leftarrow \underset{x_i}{\operatorname{argmin}}\, E(x_i, y_i)$$



Outline for today

- Finish: conditional independence in BNs
- Markov Random Fields
- **Factor graphs**

Factor graphs

- For any subsets of variables $\mathbf{x}_s$:

$$p(\mathbf{x}) = \prod_s f_s(\mathbf{x}_s)$$

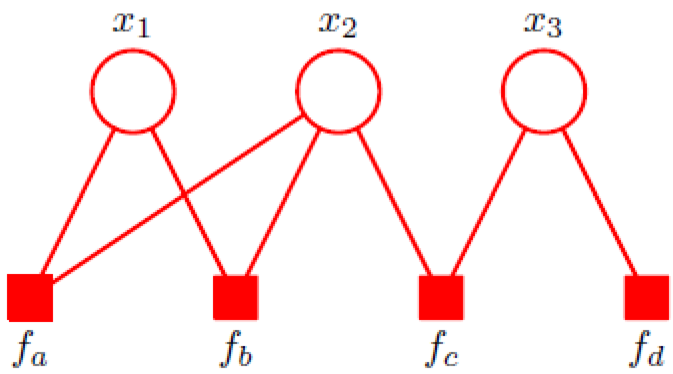
- Special cases:
  - Bayesian networks,
  - MRFs,

$$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)$$

$$p(\mathbf{x}) = \frac{1}{Z} \prod_C \psi_C(\mathbf{x}_C)$$

Factor graph representation

- Bipartite graph consisting of two types of nodes
  - Variable nodes
  - Function nodes



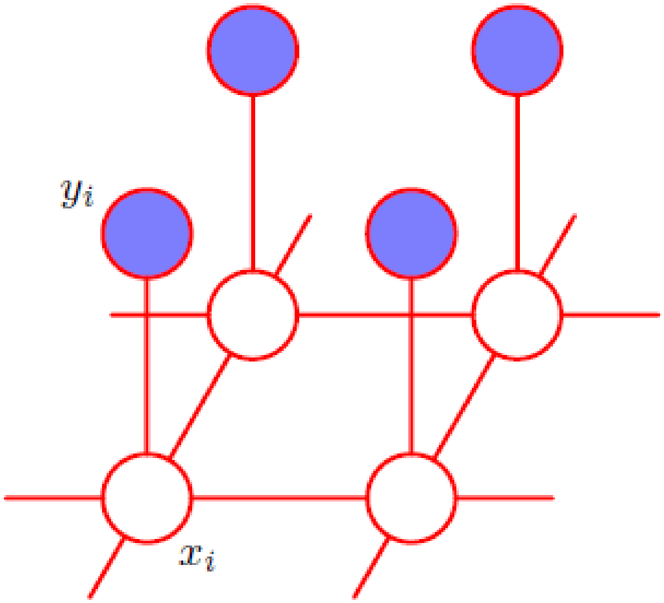
$$p(\mathbf{x}) = \prod_s f_s(\mathbf{x}_s)$$

$$p(\mathbf{x}) = f_a(x_1, x_2) f_b(x_1, x_2) f_c(x_2, x_3) f_d(x_3)$$

How might we draw the factor graph of the Ising model?

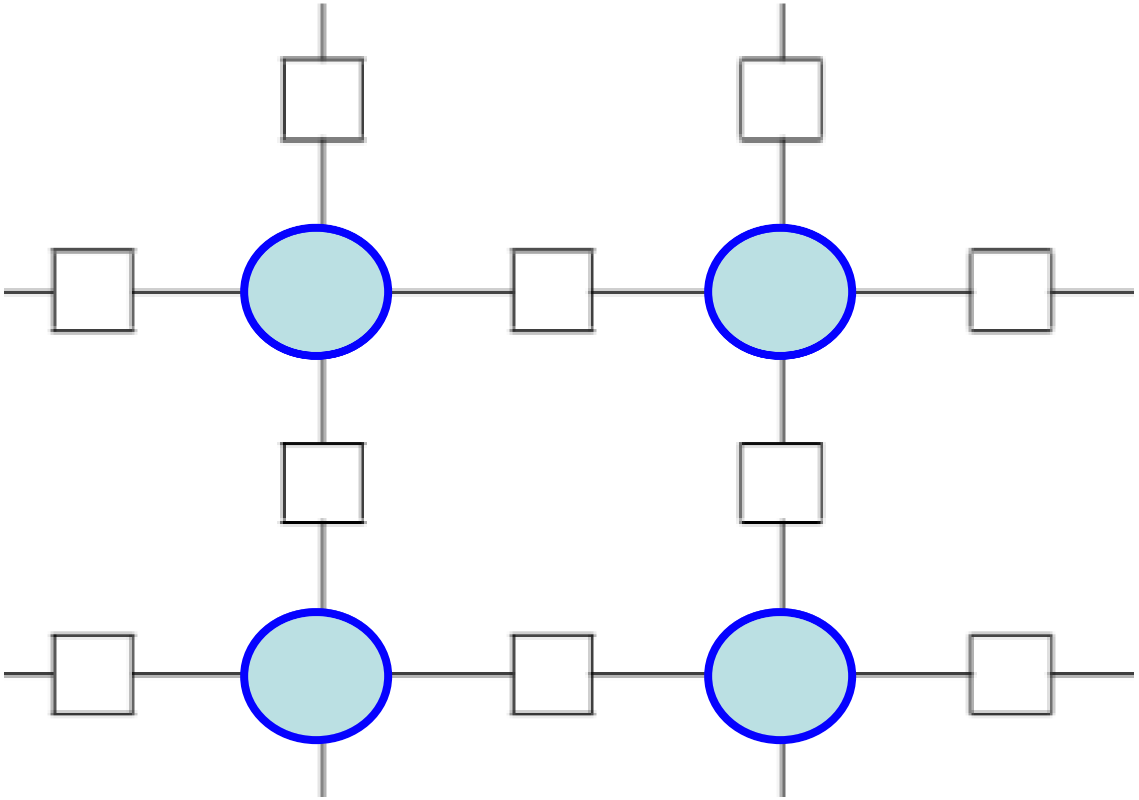
$$p(\mathbf{x}) = \frac{1}{Z} \prod_C \psi_C(\mathbf{x}_C)$$

. . .



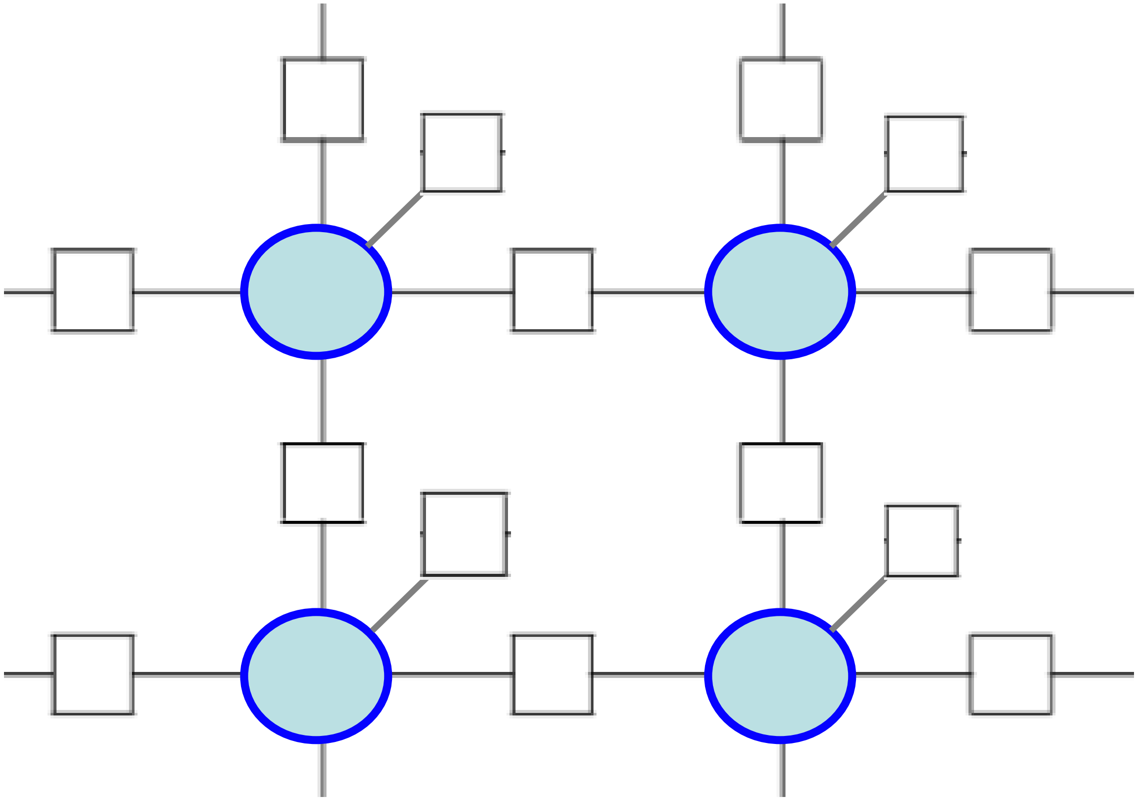
$$E(\mathbf{x}, \mathbf{y}) = h \sum_i x_i - \beta \sum_{\{i,j\}} x_i x_j - \eta \sum_i x_i y_i$$

Factor graph of the Ising model



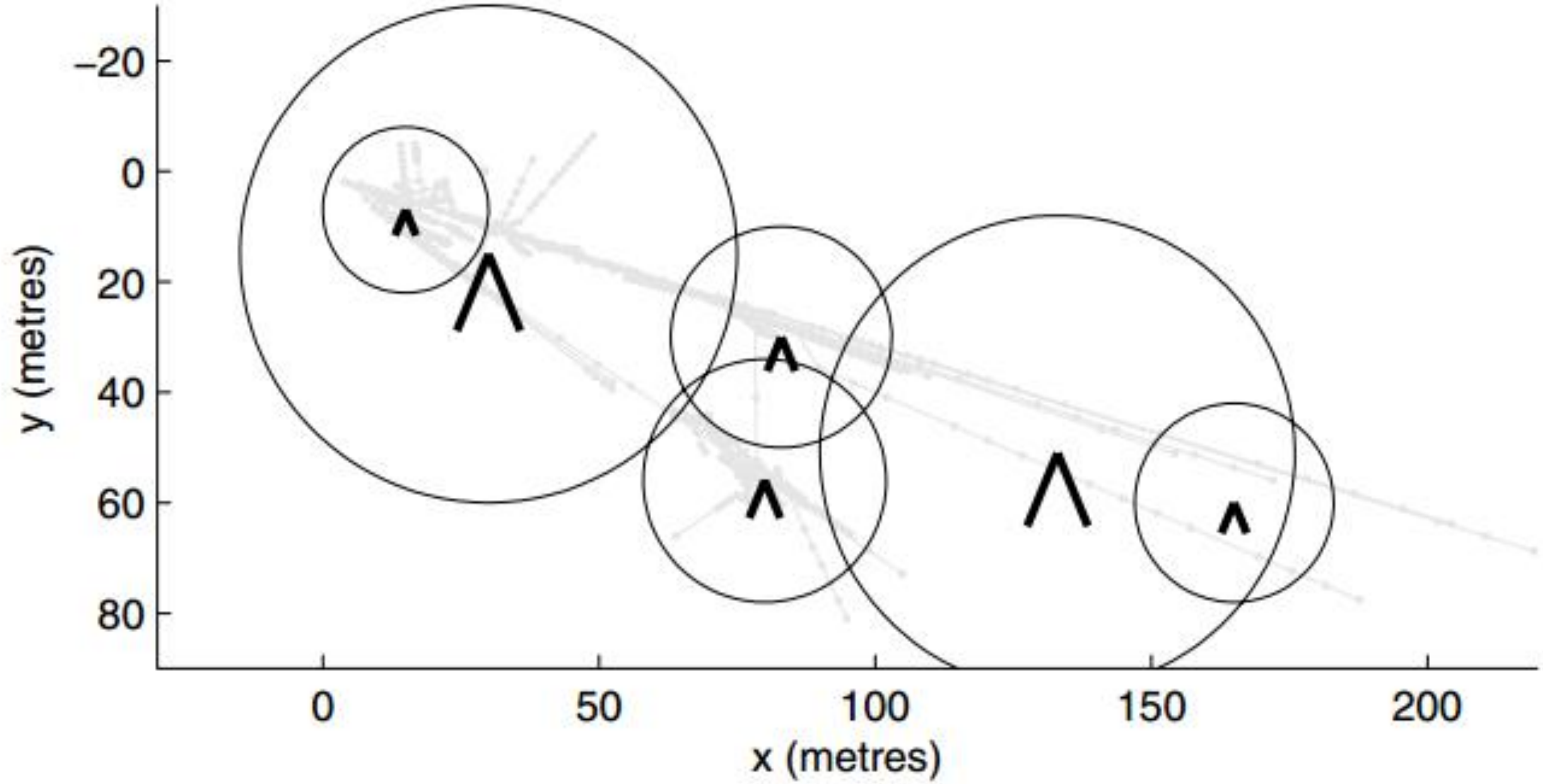
$$E(\mathbf{x}, \mathbf{y}) = h \sum_i x_i - \beta \sum_{\{i,j\}} x_i x_j - \eta \sum_i x_i y_i$$

Factor graph of the Ising model



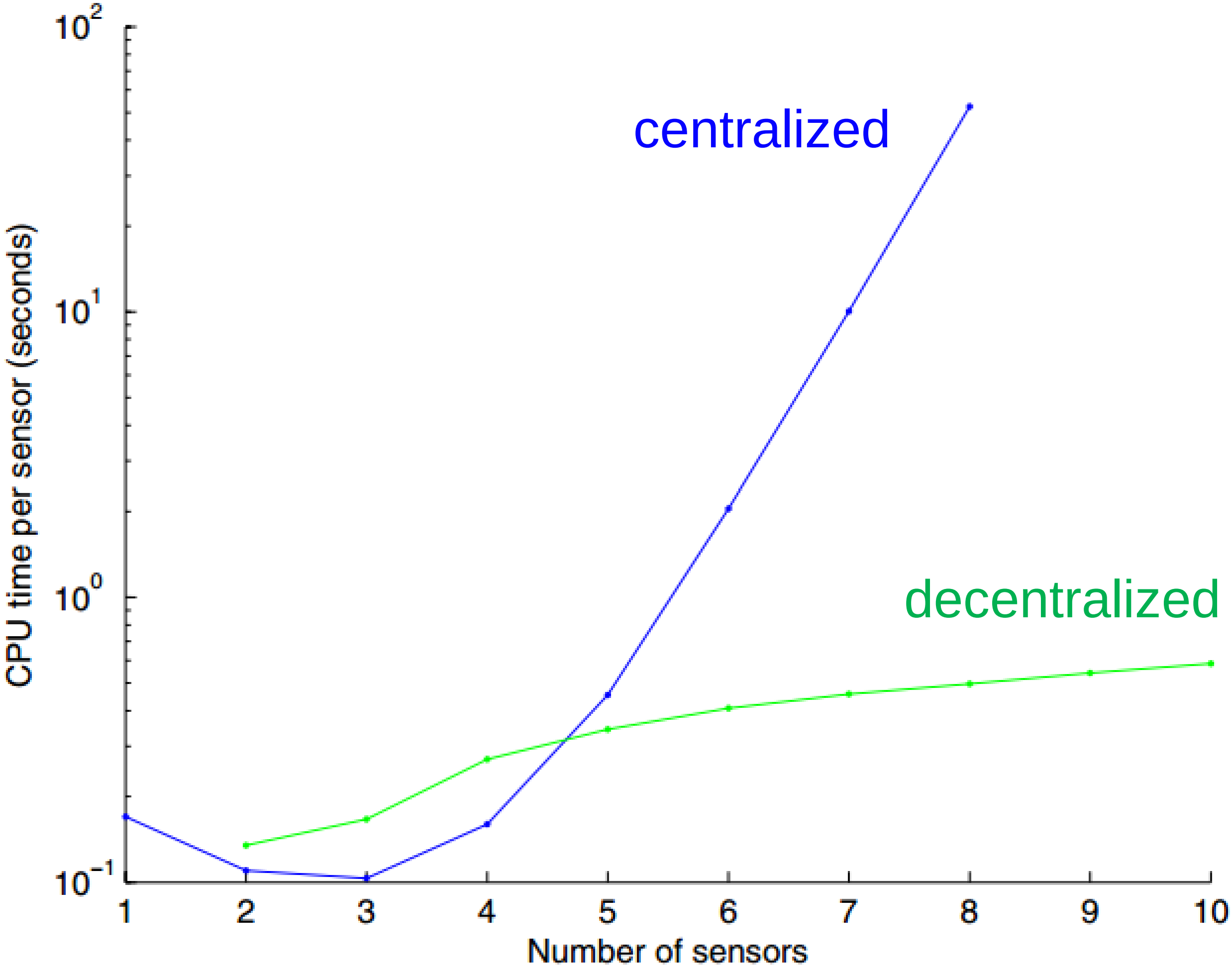
$$E(\mathbf{x}, \mathbf{y}) = h \sum_i x_i - \beta \sum_{\{i,j\}} x_i x_j - \eta \sum_i x_i y_i$$

Example application: surveillance



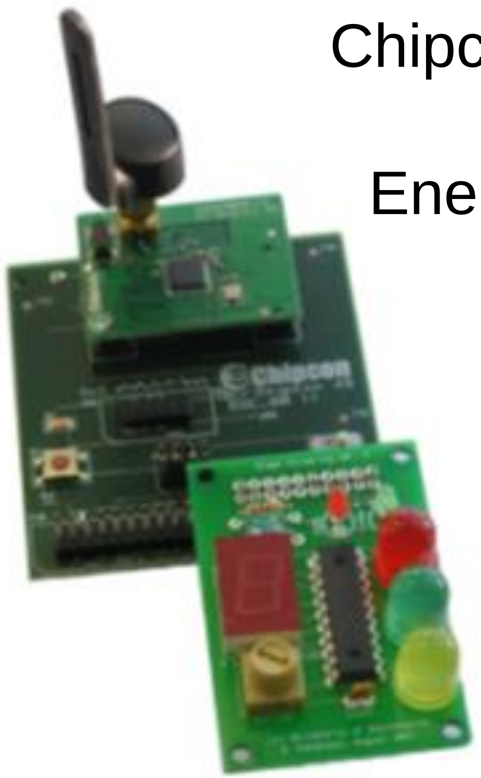
http://mebden.com/reports/AdHocNetworks.pdf

Example application: surveillance



http://mebden.com/reports/AdHocNetworks.pdf

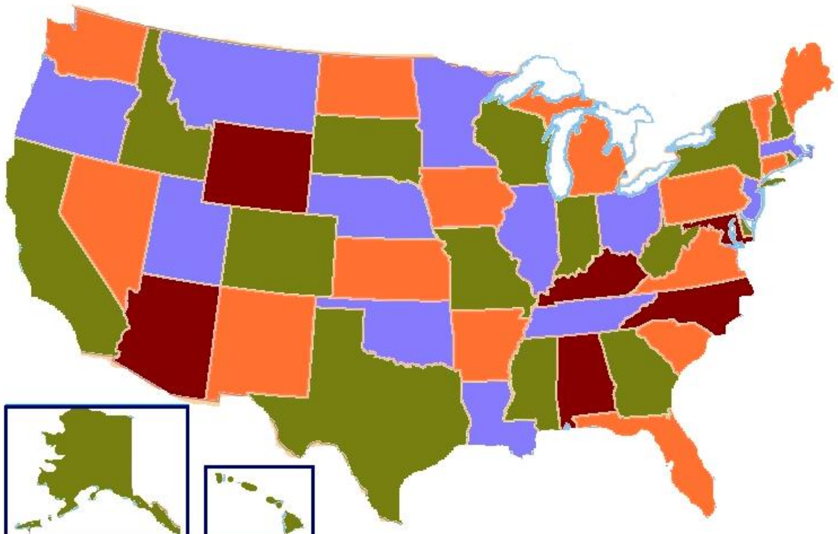
Hardware toys for decentralized coordination



Chipcon CC2431 System-on-Chip (SoC)

Energy function penalizes neighbours
who have the same colour

Four-colour theorem:

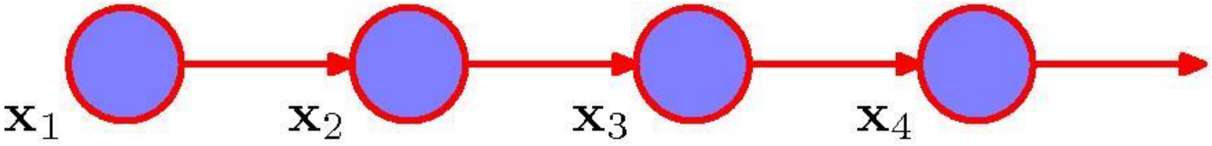


Example conference deadlines

- ICML (February, held in July)
- UAI (March, held in July)
- NIPS (June, held in December)
- AISTATS (October, held in May)

Upcoming topics

- Sequential models



- Sum-product algorithm