Stochastic Systems Group

# Maximum Entropy Relaxation for Multiscale Graphical Model Selection

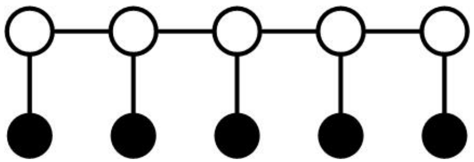
Myung Jin Choi, Venkat Chandrasekaran, and Alan S. Willsky

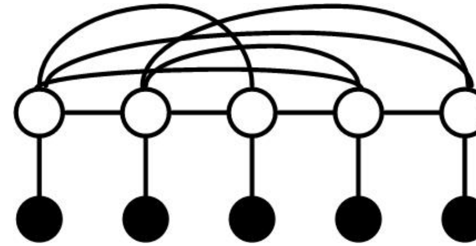
Laboratory for Information and Decision Systems

Massachusetts Institute of Technology

April 3, 2008

# Trade-offs in Model Selection

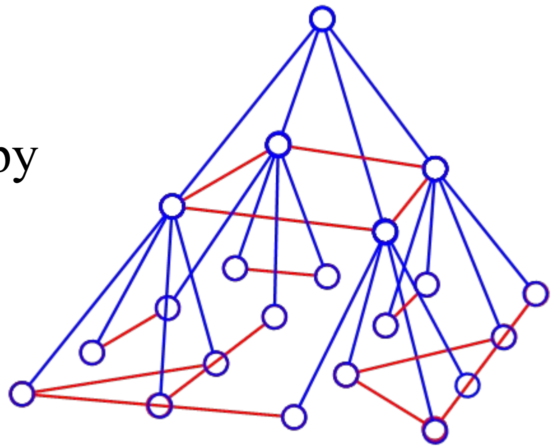





- Model selection for tractable estimation.
- Processes with long-range dependencies?
- We use multiresolution approaches.

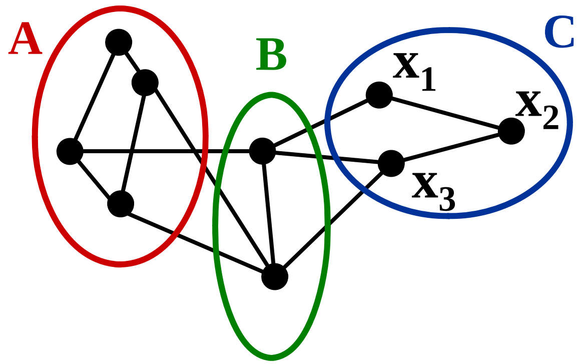
# Outline

- Background
  - Graphical Models
  - Maximum Entropy Modeling
- Multiscale Modeling Using Maximum Entropy
  - Introducing coarser scale variables.
  - Learning multiscale models with loops.
- Simulation Example
  - Modeling and Estimation Performance
- Conclusion and Future Work



# Graphical Models

Graph Separation $\Longleftrightarrow$ Conditional Independence




$$p(x_A, x_C | x_B) = p(x_A | x_B) p(x_C | x_B)$$

e.g. Gaussian distribution $p(x) \propto \exp\{-\frac{1}{2}x^T P^{-1} x\}$ $J = P^{-1}$

The inverse covariance matrix sparse with respect to $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

$$J_{ij} \neq 0 \Leftrightarrow \{i, j\} \in \mathcal{E}$$

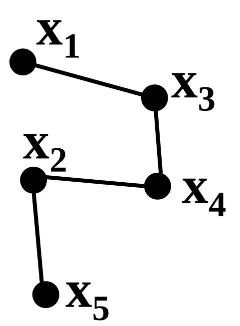
# Maximum Entropy Principle (Gaussian)

$$\boxed{\begin{aligned} \underset{P \succ 0}{\operatorname{argmax}} \quad & H(p(x)) = \log\det P \\ \text{s.t.} \quad & P_v = P_v^*, \quad \forall v \in \mathcal{V} \\ & P_E = P_E^*, \quad \forall E \in \mathcal{E} \end{aligned}}$$

$\hat{J} = \hat{P}^{-1}$ sparse with respect to $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

$$\hat{P} = \begin{pmatrix} \star & \blacksquare & \star & \blacksquare & \blacksquare \\ \blacksquare & \star & \blacksquare & \star & \star \\ \star & \blacksquare & \star & \star & \blacksquare \\ \blacksquare & \star & \star & \star & \blacksquare \\ \blacksquare & \star & \blacksquare & \blacksquare & \star \end{pmatrix}$$

$$\hat{P}^{-1} = \begin{pmatrix} \star & 0 & \star & 0 & 0 \\ 0 & \star & 0 & \star & \star \\ \star & 0 & \star & \star & 0 \\ 0 & \star & \star & \star & 0 \\ 0 & \star & 0 & 0 & \star \end{pmatrix}$$



# Maximum Entropy Relaxation (MER)

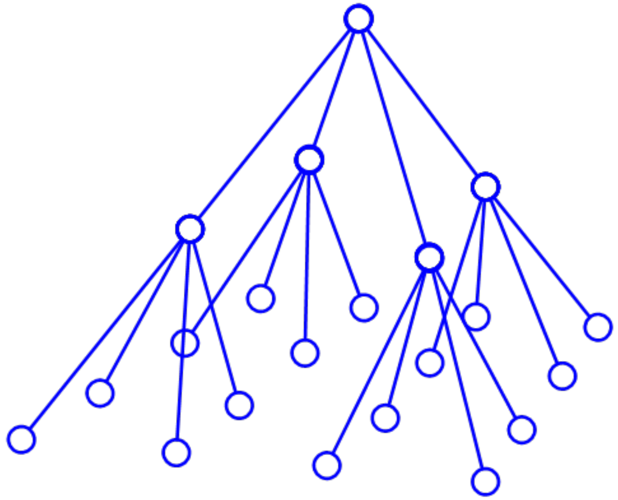
(Johnson, Chandrasekaran, Willsky 07)

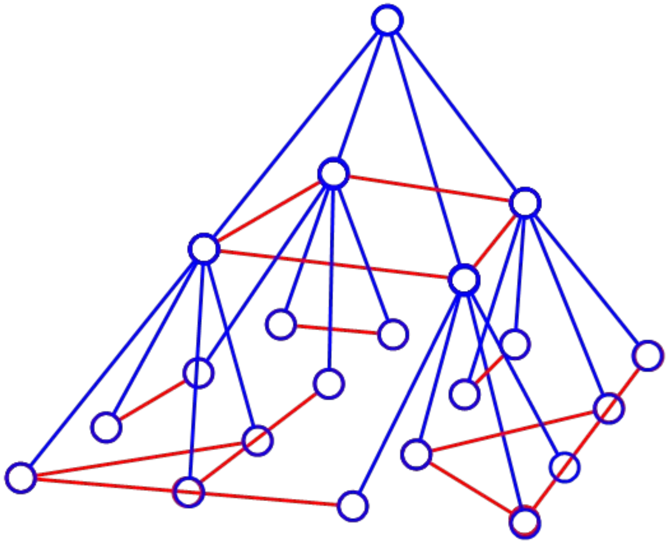
$$\begin{array}{ll} \underset{P \succ 0}{\operatorname{argmax}} & H(p(x)) = \log\det P \\ \text{s.t.} & d_1(P_v, P_v^*) \leq \delta_v, \quad \forall v \in \mathcal{V} \\ & d_2(P_E, P_E^*) \leq \delta_E, \quad \forall E \in \mathcal{E} \end{array}$$

$\hat{J} = \hat{P}^{-1}$ sparse with respect to $\mathcal{G} = (\mathcal{V}, \mathcal{E}_{active})$

$$\mathcal{E}_{active} = \{E \mid d_2(\hat{P}_E, P_E^*) = \delta_E\} \subset \mathcal{E}$$

# Multiscale Models and Algorithms

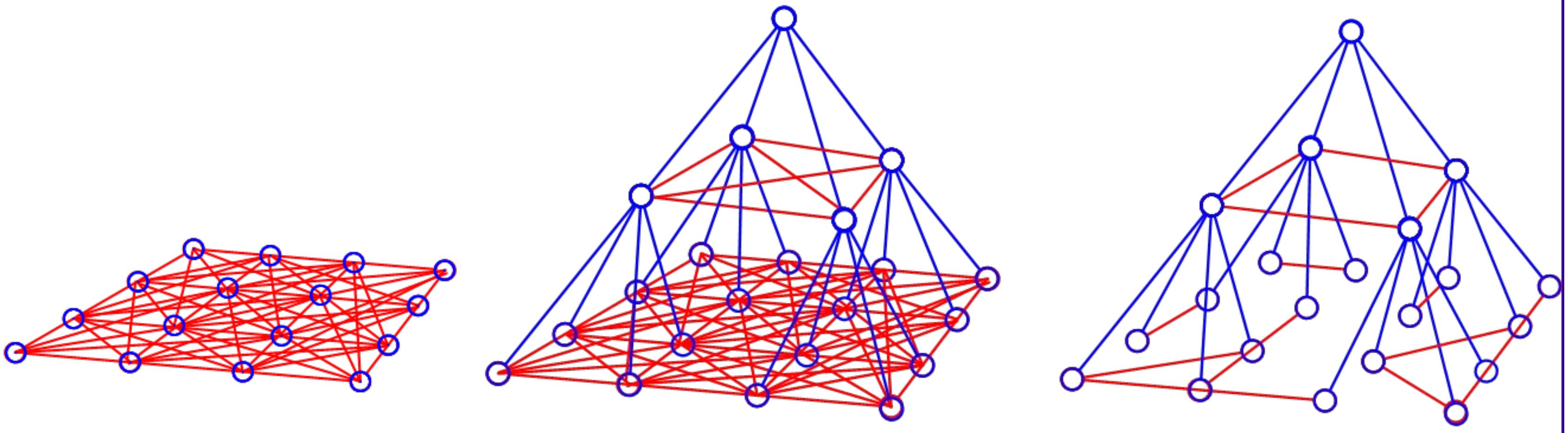




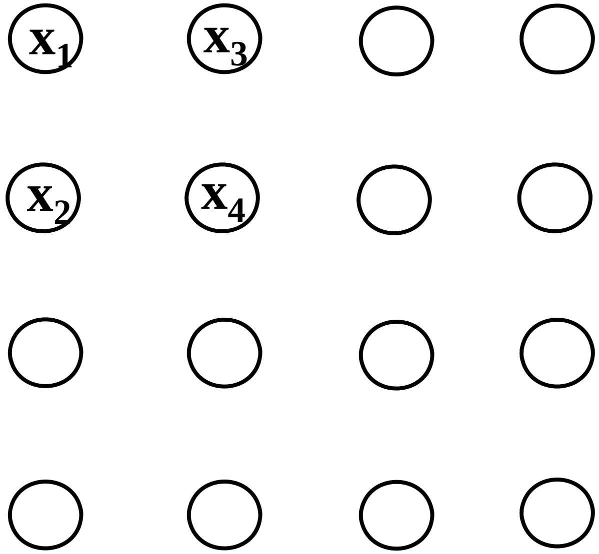
- Multiscale tree models
  - Estimation is easy.
  - Limited modeling capabilities.
- Multiscale models with loops
  - Efficient estimation using hierarchy (Choi and Willsky, 07)

# Overview

Given $x$, $P^*_x$ → Add $z$ Compute $P^*_{(x,z)}$ → Solve Multiscale MER

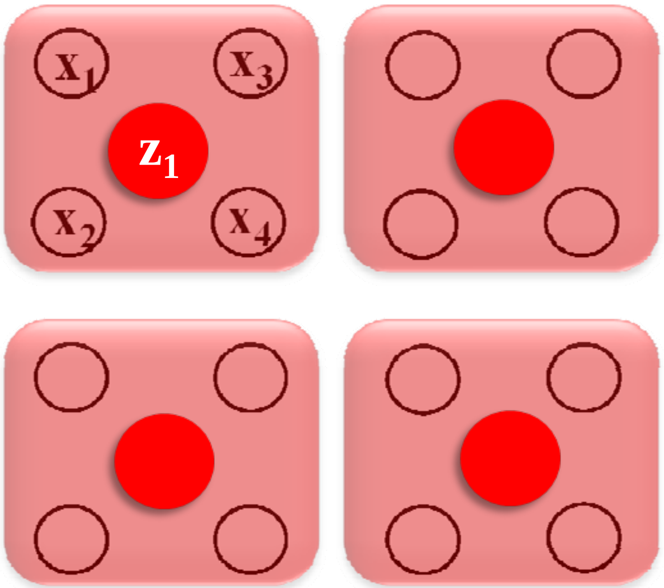


# Introducing Coarse Variables




- Original random variables: $x$
- Target covariance: $P_x^*$

# Introducing Coarse Variables




- Original random variables: $x$
- Target covariance: $P_x^*$

- Coarser scale variables: $z$

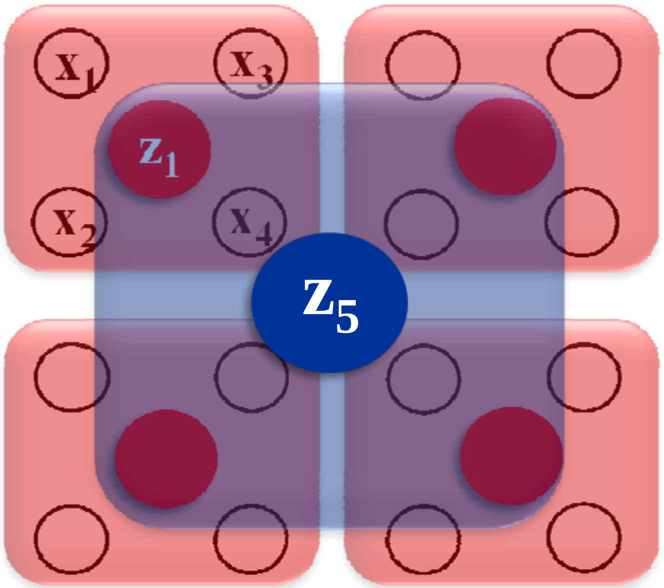
e.g.:

$$z_1 = \frac{1}{4}\sum_{i=1}^{4} x_i + n$$

- Joint target covariance $P^*_{(x,z)}$

  $\rightarrow$ Input to multiscale MER.

# Introducing Coarse Variables




- Original random variables: $x$
- Target covariance: $P_x^*$

- Coarser scale variables: $z$

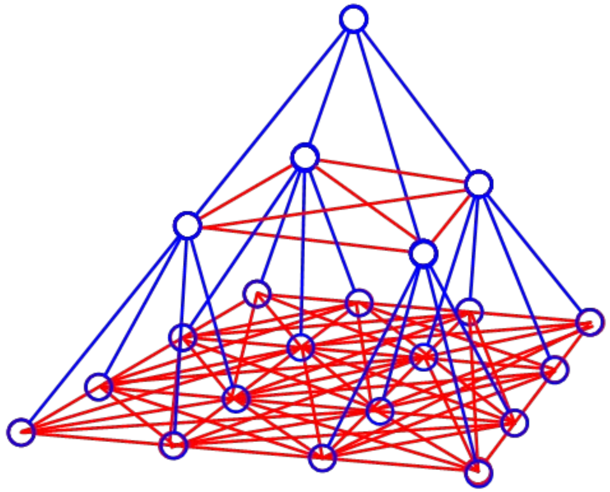
  e.g: $$z_1 = \frac{1}{4}\sum_{i=1}^{4} x_i + n$$

- Joint target covariance $P^*_{(x,z)}$

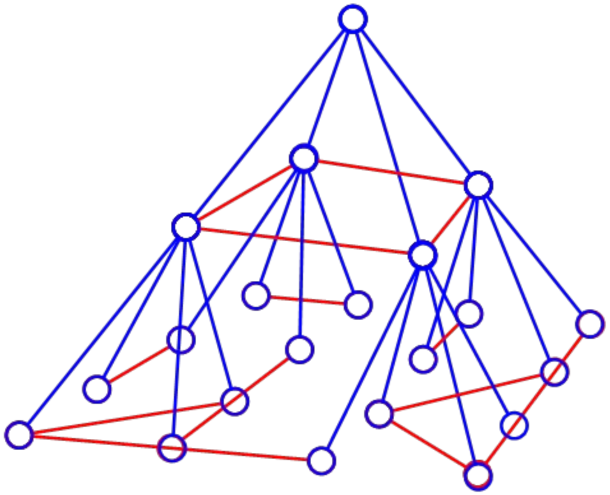
  $\rightarrow$ Input to multiscale MER.

# Multiscale MER

- Target covariance $P^*_{(x,z)}$
- Maximize entropy (s.t. constraints graph)
- Efficient primal-dual interior point method.



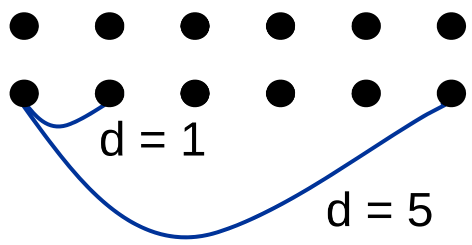
Constraints Graph



Multiscale MER Solution

# Simulation Example

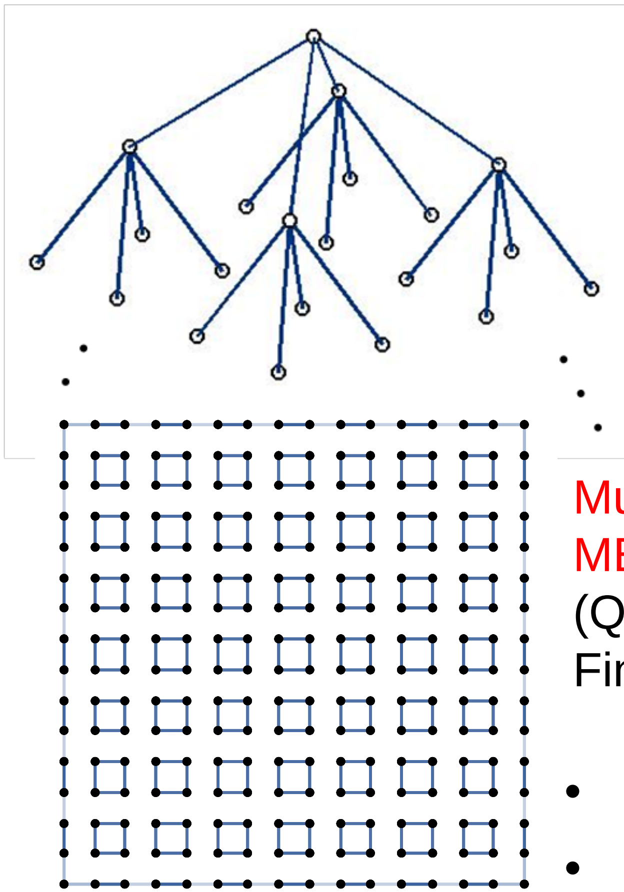
- 256 Gaussian variables arranged spatially on a 16x16 grid.
- Covariance with polynomial decay:



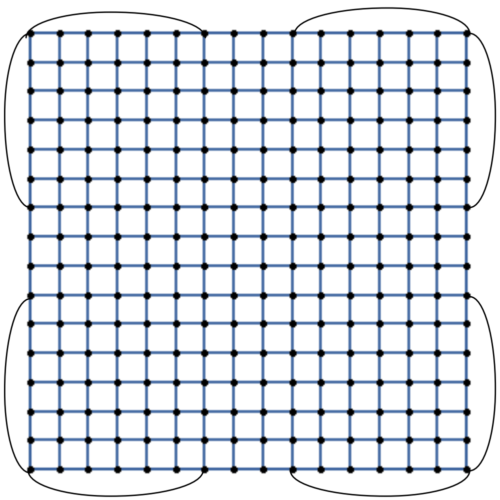

$$P_{x_s} = 1.5$$
$$P_{x_s,x_t} = d(s,t)^{-0.5}$$

- Compare four models
  - The original model (fully connected)
  - Single-scale MER (without coarse variables)
  - Multiscale tree (one coarse variable per four children)
  - Multiscale MER

# Models Learned



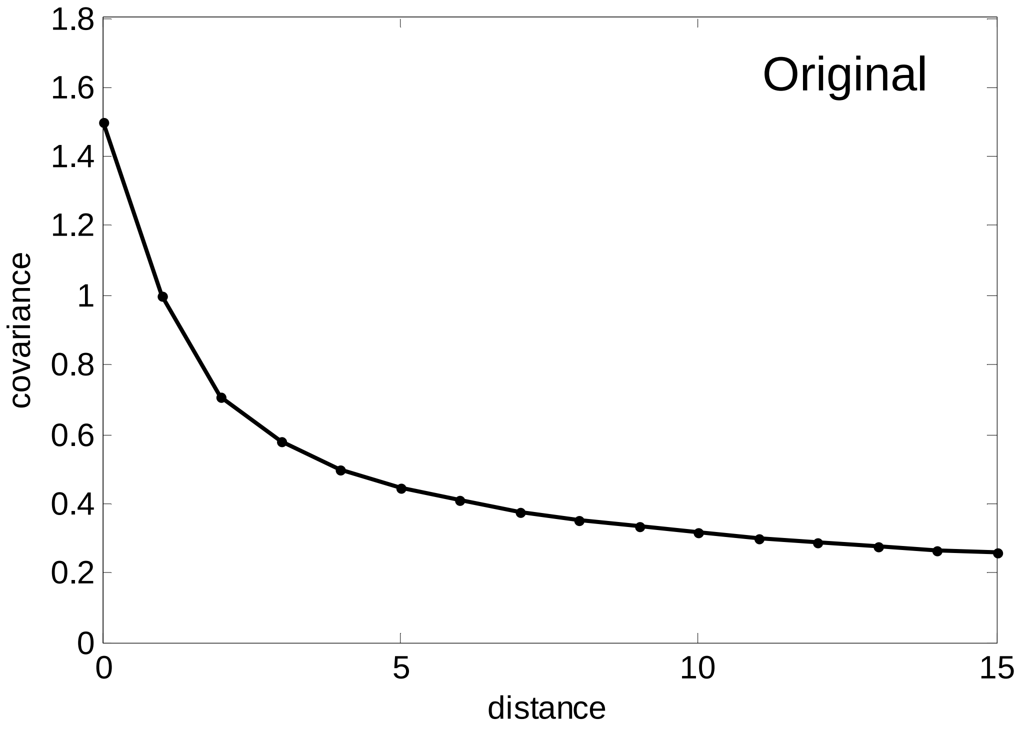
Multiscale MER solution (Quadtree + Finest scale)



Single-scale MER solution

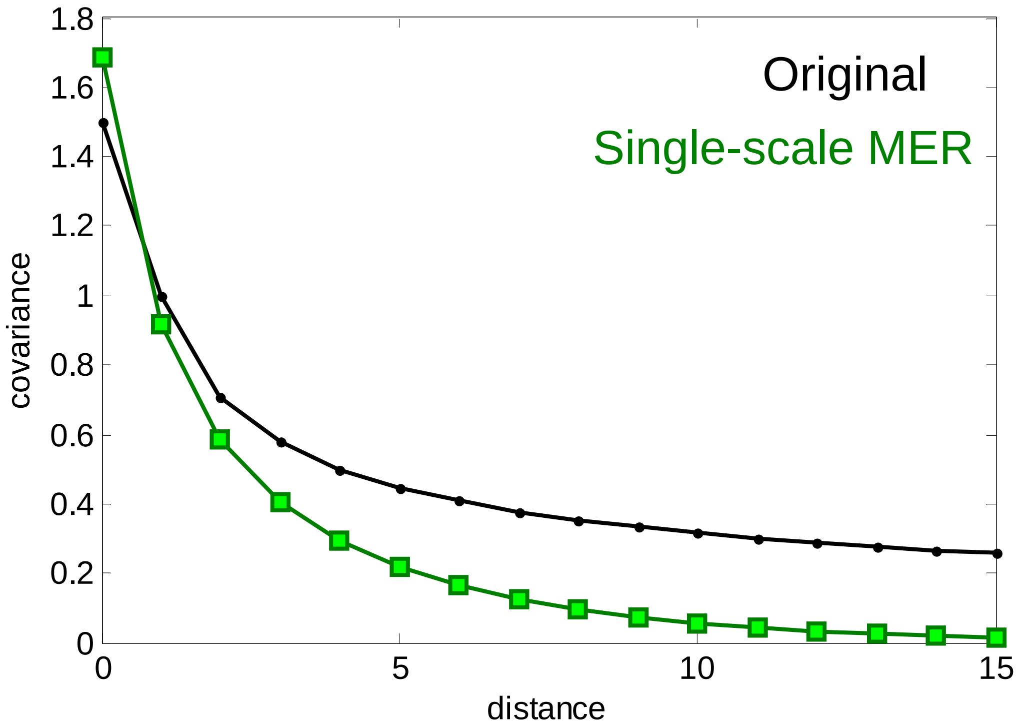
- More number of variables yet simple structure.
- Efficient algorithms available.

# Modeling Performance




- Covariance decay in distance.

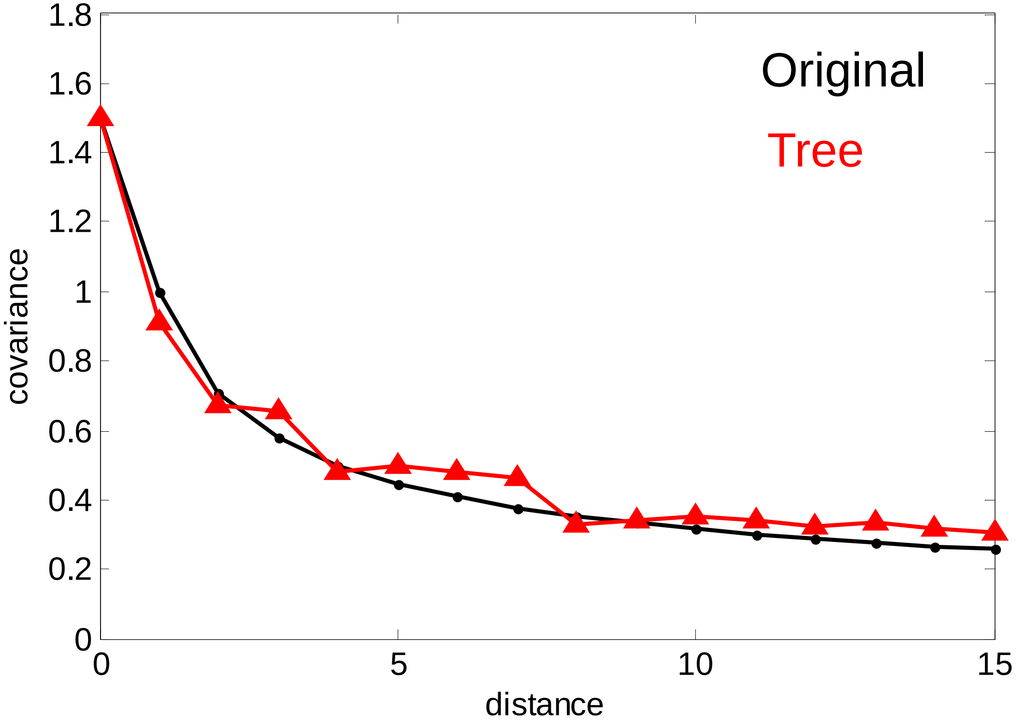
# Modeling Performance




- Covariance decay in distance.

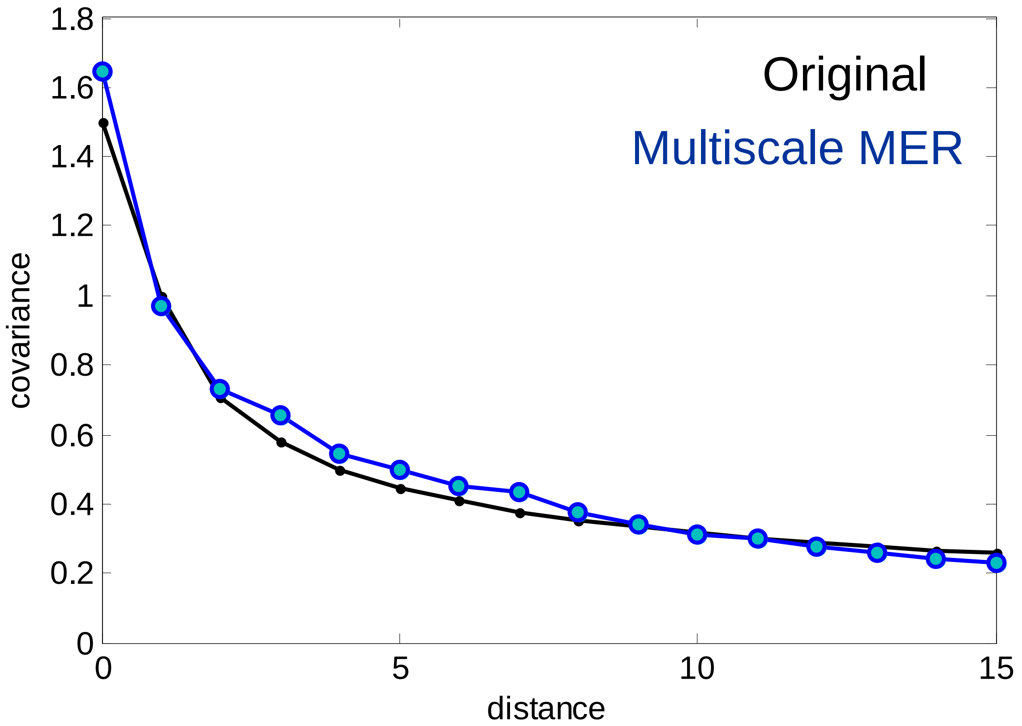
Stochastic Systems Group

# Modeling Performance



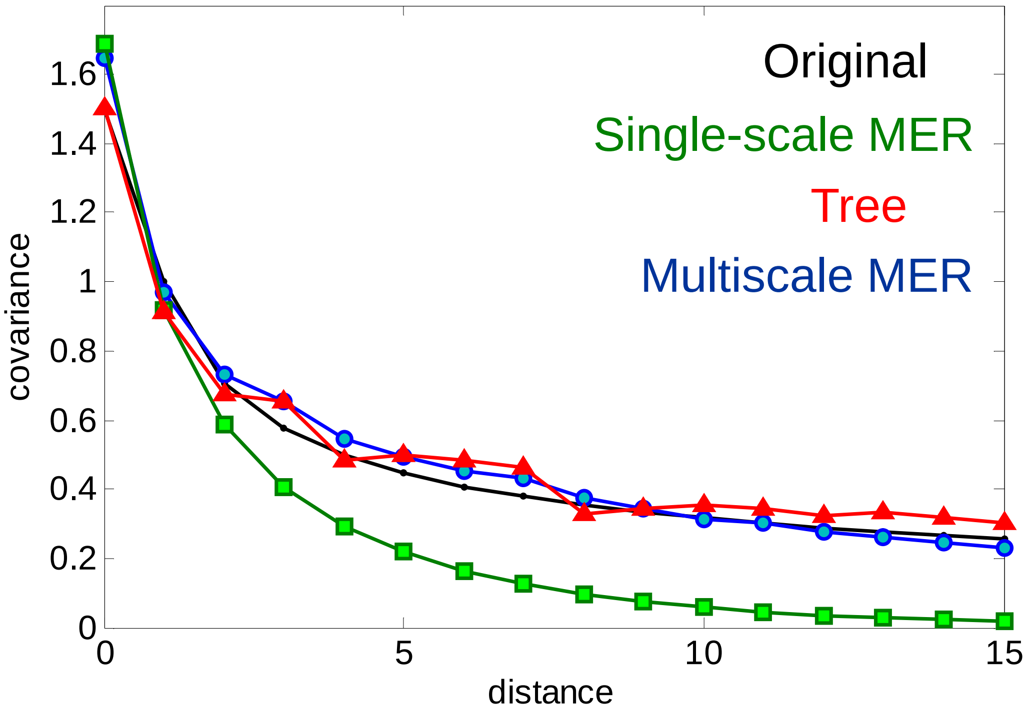

- Covariance decay in distance.

# Modeling Performance



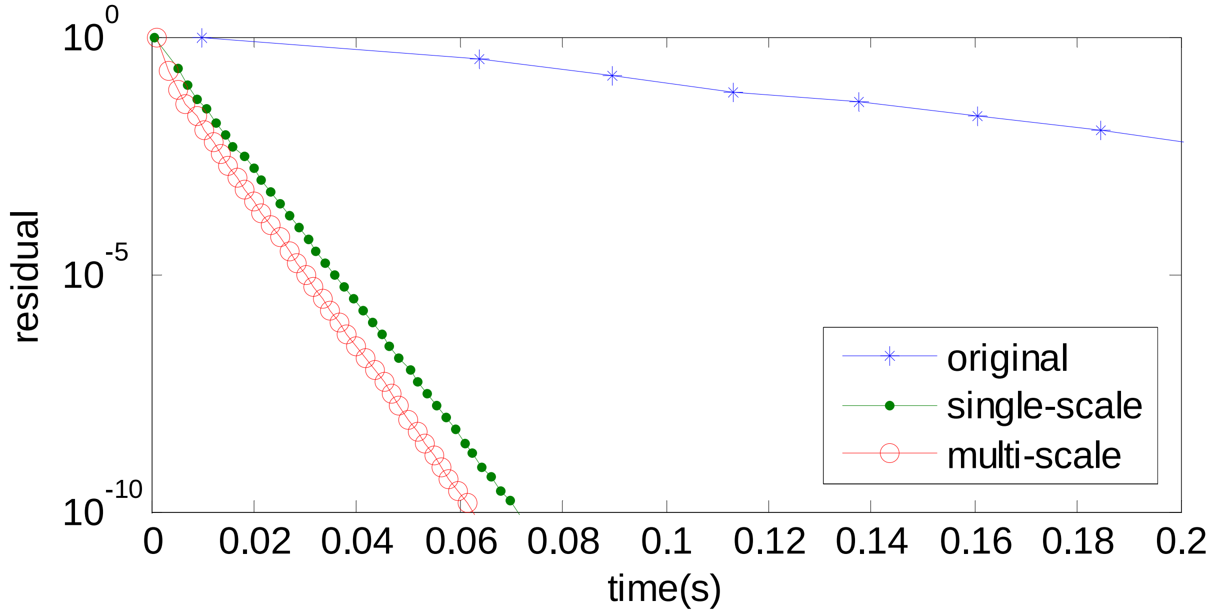

- Covariance decay in distance.
- Multiscale model captures long-range dependencies without blocky artifacts.

# Modeling Performance




- Covariance decay in distance.
- Multiscale model captures long-range dependencies without blocky artifacts.

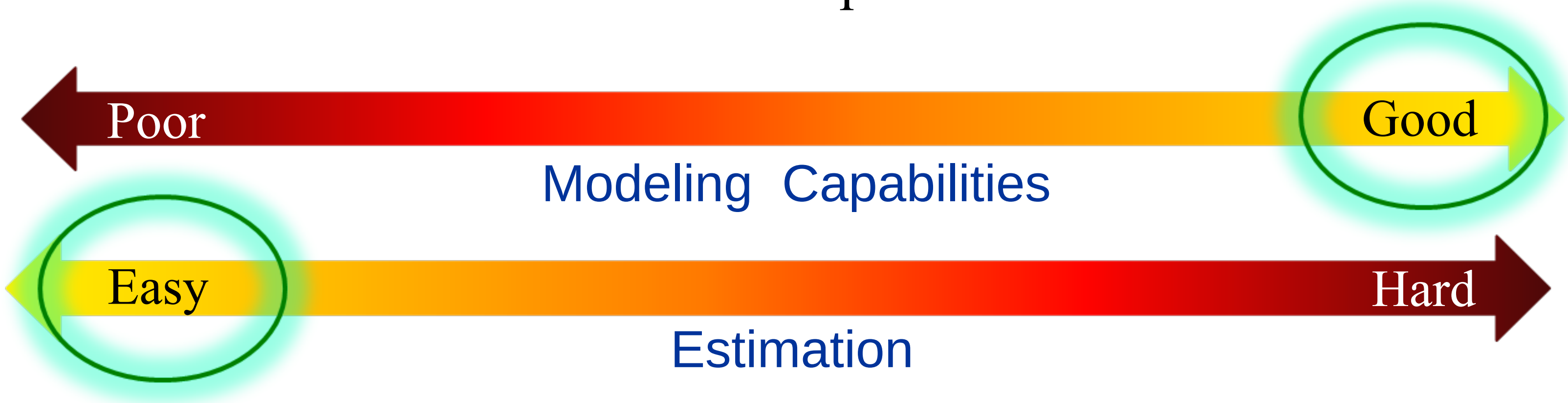
# Estimation Performance




- Estimation given sparse noisy measurements.
- Residual error vs. computation time.
- Adaptive embedded trees, multipole-motivated estimation methods.
- Multiscale MER solution 10% faster than single-scale MER.

# Conclusion and Future Work

- Multiscale modeling based on maximum entropy relaxation.
- Coarse hidden variables as a computational tool.




- Future work
  - Adaptive introduction of hidden variables.