Generalized Belief Propagation Receiver for Near-Optimal Detection of Two-Dimensional Channels with Memory

Ori Shental[1], Noam Shental[2,3], Yair Weiss[3] and Anthony J. Weiss[1]

1. Department of Electrical Engineering – Systems,
Tel Aviv University

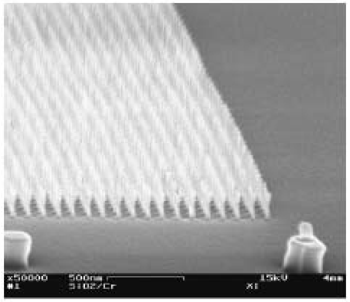
2. Department of Physics of Complex Systems,
Weizmann Institute

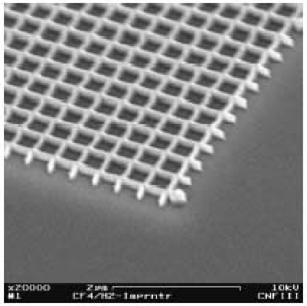
3. School of Computer Science and Engineering and
The Center for Neural Computation,
The Hebrew University of Jerusalem

PASCAL

Examples of 2-D channels

- 2-D inter-symbol interference (ISI) channels
  - Magnetic and optical recording





Science enables:

6 million-million bits/square inch!

from: J.A. O'Sullivan, N. Singla, Y. Wu, R.S. Indeck

Examples of 2-D channels

- Multiple-access (MA) channels – Wyner's model
  - Cellular network's uplink.
  - Indoor Wireless LAN.

Outline

- Examples of 2-D channels

- System model
- Optimal detection

- 2-D channels as undirected graphical models
- Detection = Inference
  - Exact inference
  - Approximate inference
    - Belief propagation (BP)
    - Generalized belief propagation (GBP)

- Experimental results

Outline (cont.)

- Cluster variation method (CVM) for estimation of the information rate of 2D channels.
- Experimental results
- Why is GBP-based CVM correct in this case?

System model

Binary symbols

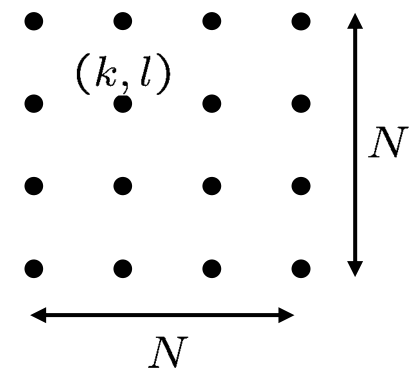
Interference intensity

$$y_{k,l} = d_{k,l} + v_{k,l} + \sum_{(i,j)\in<k,l>} \alpha_{i,j} d_{i,j} \quad \forall k,l = 1,\ldots,N$$

Observed samples

AWGN $\mathcal{N}(0,\sigma^2)$

Short range 2-D interference topology



In vector-matrix form

$$\mathbf{y}_{N^2\times 1} = \mathbf{S}_{N^2\times N^2}\mathbf{d}_{N^2\times 1} + \mathbf{v}_{N^2\times 1}$$

Interference band matrix

- 2-D channels differ in $\mathbf{S}$

System model (cont.)

- Objective
  - Recover $\mathbf{d}$ from $\mathbf{y}$, or
  - Compute $\Pr(\mathbf{x}|\mathbf{y})$, where $\mathbf{x}$ are possible values of $\mathbf{d}$.
  - Assume that d are i.i.d and equiprobable.

$$\Pr(\mathbf{x} \mid \mathbf{y}) \propto \exp(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{S}\mathbf{x}\|^2)$$

- Omitting non-sufficient statistics terms

$$\Pr(\mathrm{x}|\mathrm{y}) \propto \exp\left(-\frac{1}{\sigma^2}\left(\sum_{(i>j)} R_{ij}x_ix_j - \sum_i h_ix_i\right)\right)$$

where

- $\mathrm{R} = \mathrm{S}^T\mathrm{S}$ interference cross-correlation matrix
- $\mathrm{h} = \mathrm{y}^T\mathrm{S}$ output vector of a filter matched to the interference structure

Optimal detection

- MAP decision

$$\hat{d}_i = \arg\max_{x_i} \Pr(x_i|\mathbf{y}) = \arg\max_{x_i} \sum_{\mathbf{x}\backslash x_i} \Pr(\mathbf{x}|\mathbf{y})$$

Problem - intractable

2-D channels as undirected graphical models

- p($\mathbf{x}|\mathbf{y}$) defines an undirected graphical model

$$\Pr(\mathbf{x}|\mathbf{y}) \propto \prod_{(i>j)} \psi_{ij}(x_i, x_j) \prod_i \phi_i(x_i, h_i)$$

where the potentials

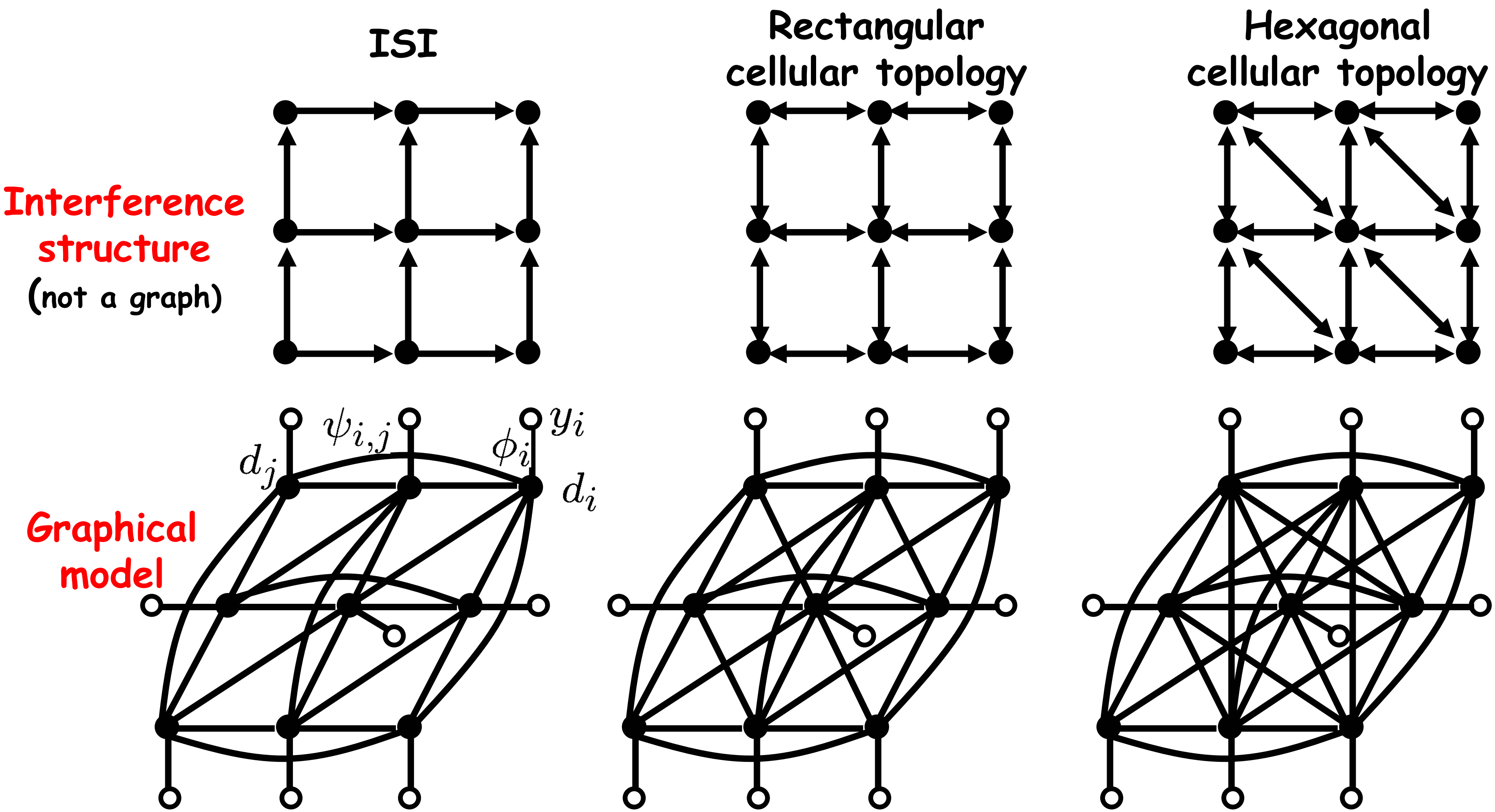
- Compatibility function

$$\psi_{ij}(x_i, x_j) = \exp\left(-\frac{R_{ij} x_i x_j}{\sigma^2}\right)$$

- Evidence, or local likelihood

$$\phi_i(x_i, y_i) = \exp\left(\frac{h_i x_i}{\sigma^2}\right)$$

Examples of 2-D channel representations



Exact inference – junction tree

- Junction tree complexity
  - Exponential in the size of the largest clique
    - For NxN grid-like graphs: N x memory depth ν

- Conclusion
  - We must resort to approximate inference methods

Approximate inference: belief propagation

- Belief propagation often yields good approximations when the cycles in the graph are long
- However – 2-D channels contain many short cycles
- Empirical results show
  - Belief propagation fails to converge
  - When it converges its approximation is poor

PASCAL

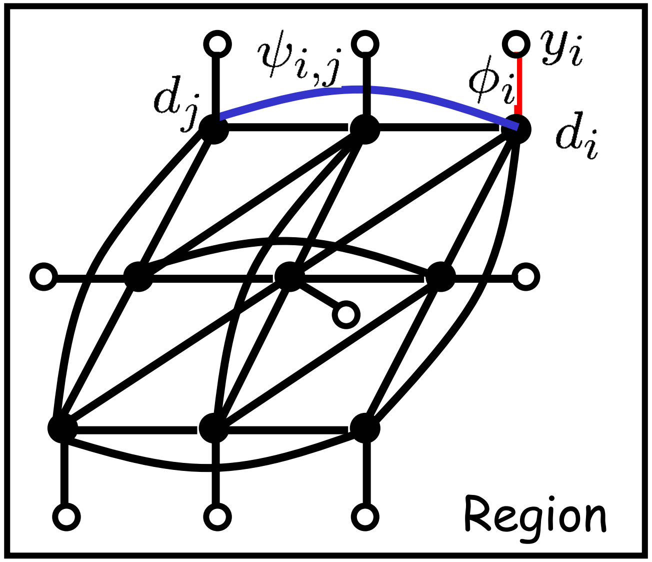
Generalized belief propagation (GBP)

[J. S. Yedidia, W. T. Freeman and Y. Weiss]

- BP <-> Bethe approximation to the free energy
- GBP <-> More complicated free energy approximations
  - Several ways for creating free energy approximations
    - We use the Kikuchi approximation, i.e., the cluster variation method
- GBP is a global name for a family of message passing algorithms
  - Several ways for passing messages
    - We use the 'two-way algorithm'.

Generalized belief propagation (GBP)

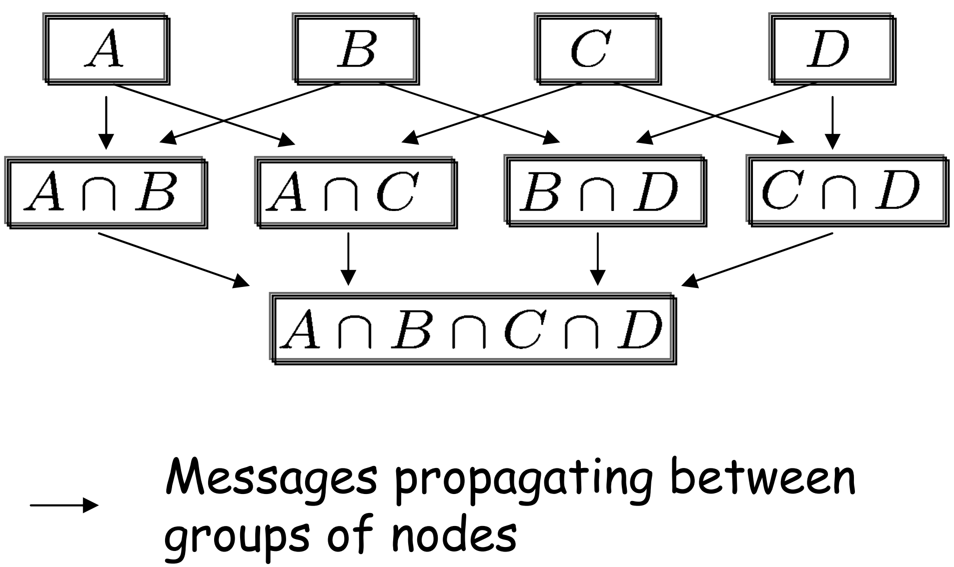
- Rational - choose the basic region to encompass the shortest loops
- We have nearest neighbor and next-nearest neighbor interactions, thus a 3x3 nodes basic cluster is a natural choice
- E.g. ISI

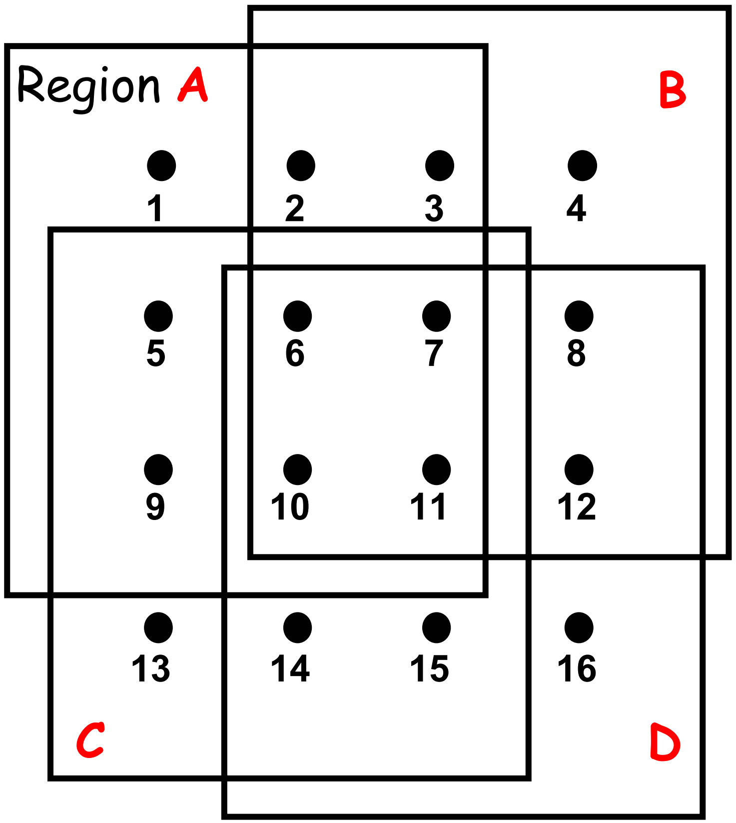


PASCAL

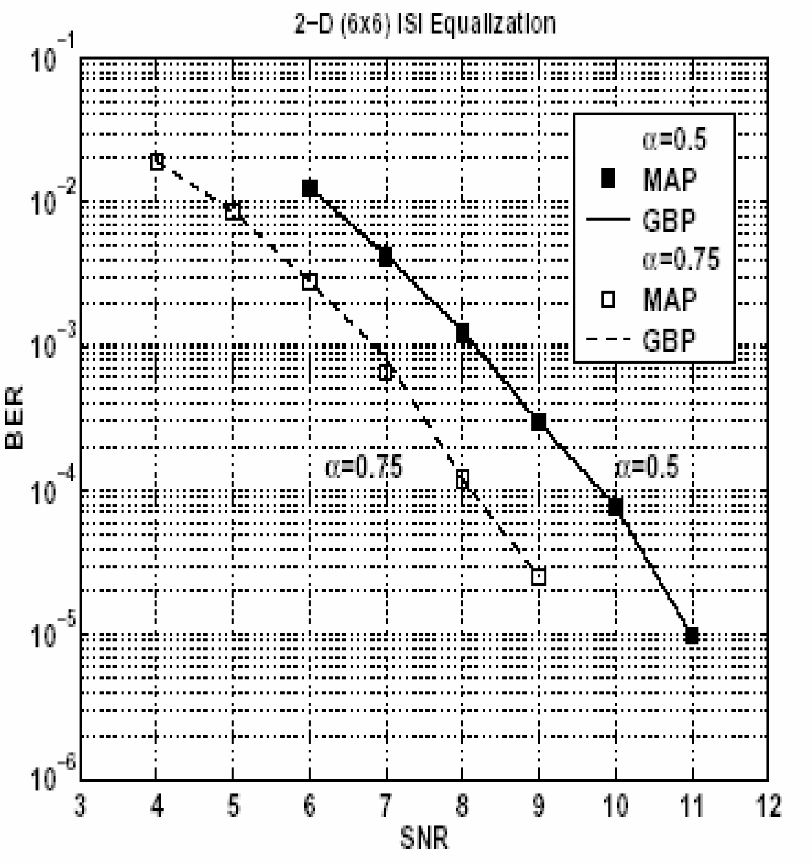
Generalized belief propagation (GBP)

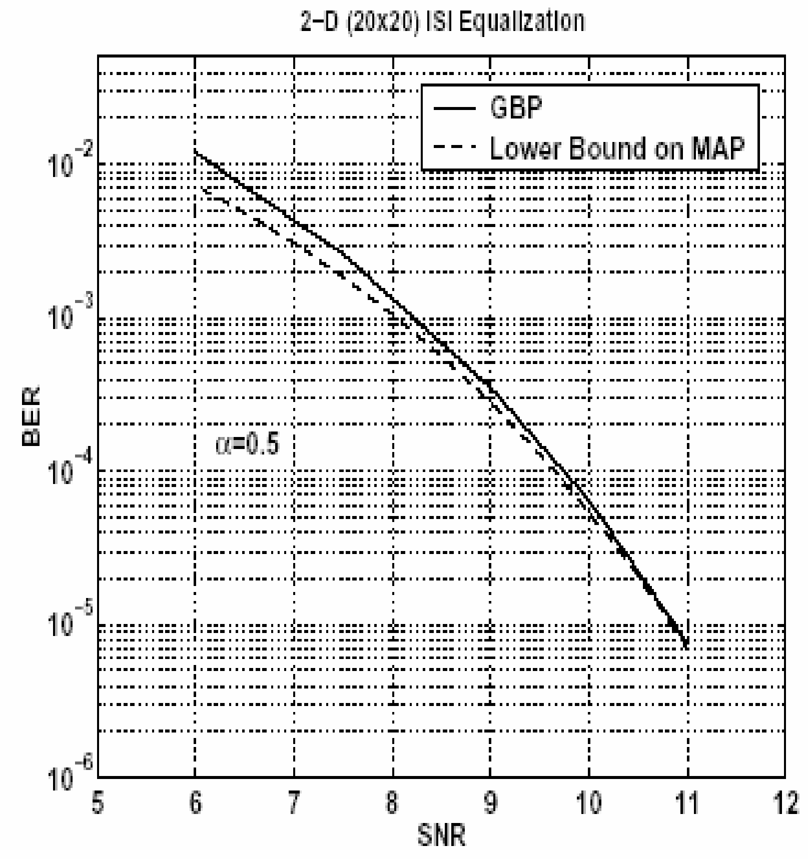
- In GBP messages are propagating between groups (regions) of nodes
- Exact inference is performed within each group





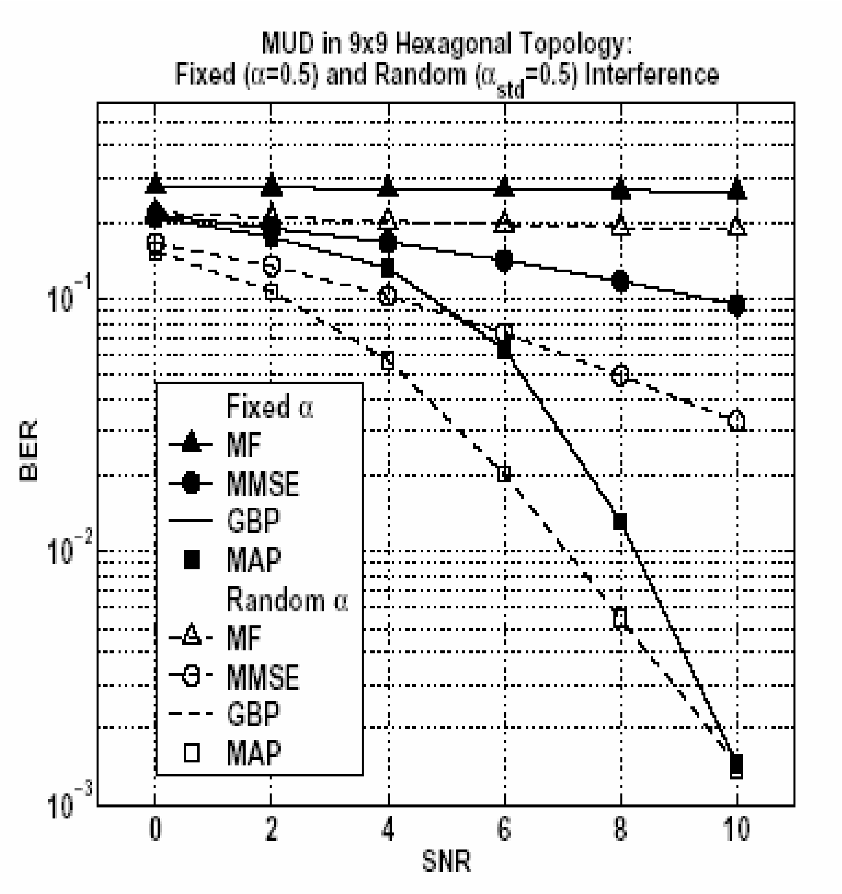
Results: ISI equalization

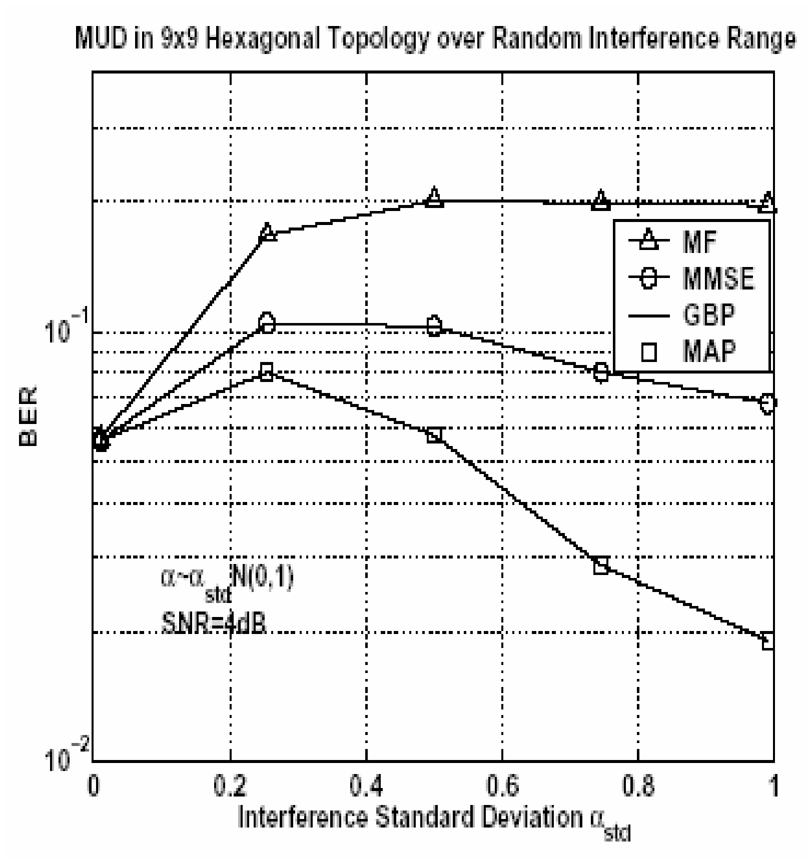




PASCAL

Results:
hexagonal topology cellular network





PASCAL

Conclusions - ITW

- Practically optimal error performance using a fully tractable message passing scheme
  - Consistent both over SNR and interference range
  - The marginal beliefs well approximate the a-posteriori probabilities (APP)
  - superior to other sub-optimal receivers
- A real-life application in which GBP>>BP

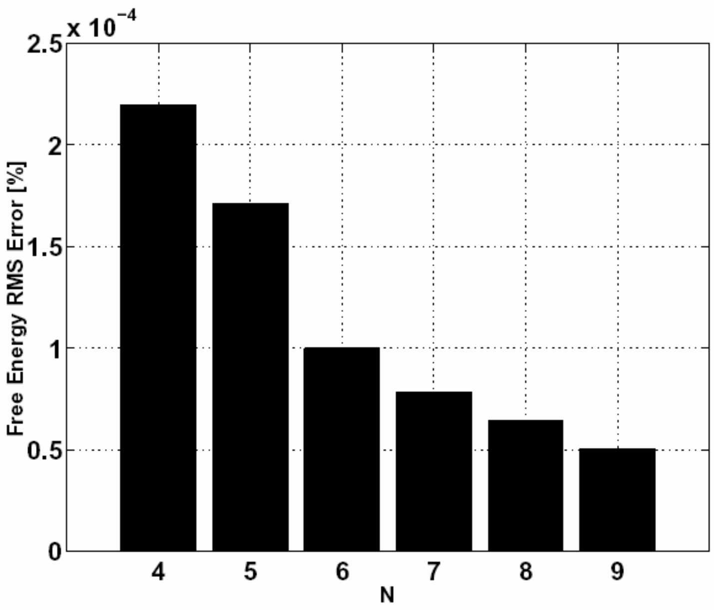
Current work

- Joint work with Shlomo Shamai (Shitz), Technion.
- Cluster variation method (CVM) for the estimation of the information rate of 2D channels.
- Experimental results.
- Why is GBP-based CVM correct in this case?

Approximate free energy

NxN system

Estimate the exact and approximate free energies per site



Hexagonal cellular network, SNR=0dB, alpha=0.5

PASCAL

The connection between the free energy and the information rate

Assume a system of size $N^2=k$

$$x^k \equiv x_1,\dots,x_k \qquad y^k \equiv y_1,\dots,y_k$$

are stationary random processes.

Information rate:

$$I(x;y) = h(y) - h(y\,|\,x)$$

$$h(y) \equiv \lim_{k\to\infty} h(y^k)/k \qquad h(y\,|\,x) \equiv \lim_{k\to\infty} h(y^k\,|\,x^k)/k$$

are differential entropy rates

As far as we know, only bounds exist for I(x;y) in the 2D case

The connection between the free energy and the information rate (cont.)

h(y|x): $$h(y \mid x) \square \lim_{k\to\infty} h(y^k \mid x^k)/k = \lim_{k\to\infty} h(v^k)/k$$

$$= \frac{1}{2}(1+\log 2\pi\sigma^2)$$

---

h(y): Assuming a stationary ergodic process, using the Shannon-McMillan-Breiman theorem

$$-\frac{1}{k}\log(p(y^k)) \to h(y)$$ with probability one.

The connection between free energy and symmetric information rate

Estimating $p(y^k)$:

Assume $x_1,\ldots,x_k$ are i.i.d. and equiprobable.

$$p(y^k) = \frac{1}{2^k}\sum p(y^k \mid x^k) = \frac{1}{2^k}(2\pi\sigma^2)^{-k/2} \underbrace{\sum \exp\left(-\frac{1}{2\sigma^2}\left\|y^k - Sx^k\right\|^2\right)}_{Z}$$

$$-\frac{1}{k}\log(p(y^k)) = \log(2) + \log\sqrt{2\pi\sigma^2} + F/k$$

PASCAL

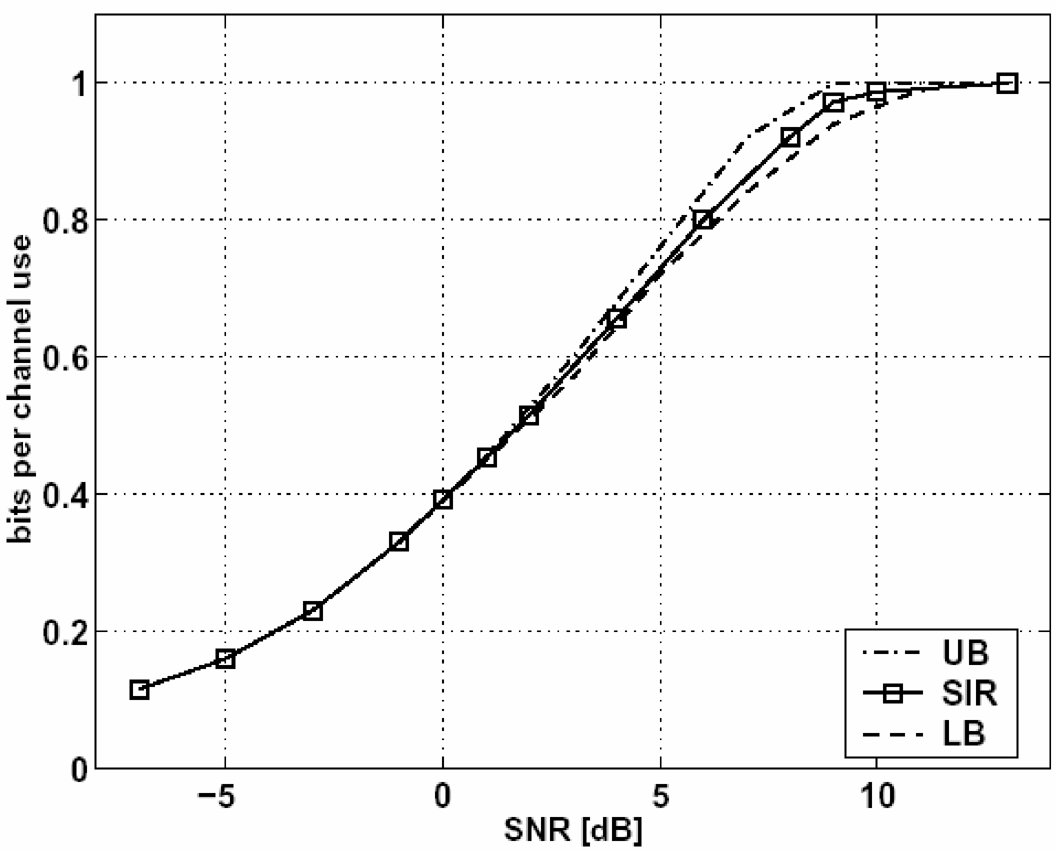
The connection between free energy and the symmetric information rate (cont.)

Thus,

$$\begin{aligned} I(x;y) &= h(y) - h(y \mid x) \\ &= (\log(2) + \log\sqrt{2\pi\sigma^2} + F/k) - (1/2 + \log\sqrt{2\pi\sigma^2}) \\ &= \log 2 - 1/2 + F/k \end{aligned}$$

Idea: Use the CVM of a large system to approximate the free energy F.

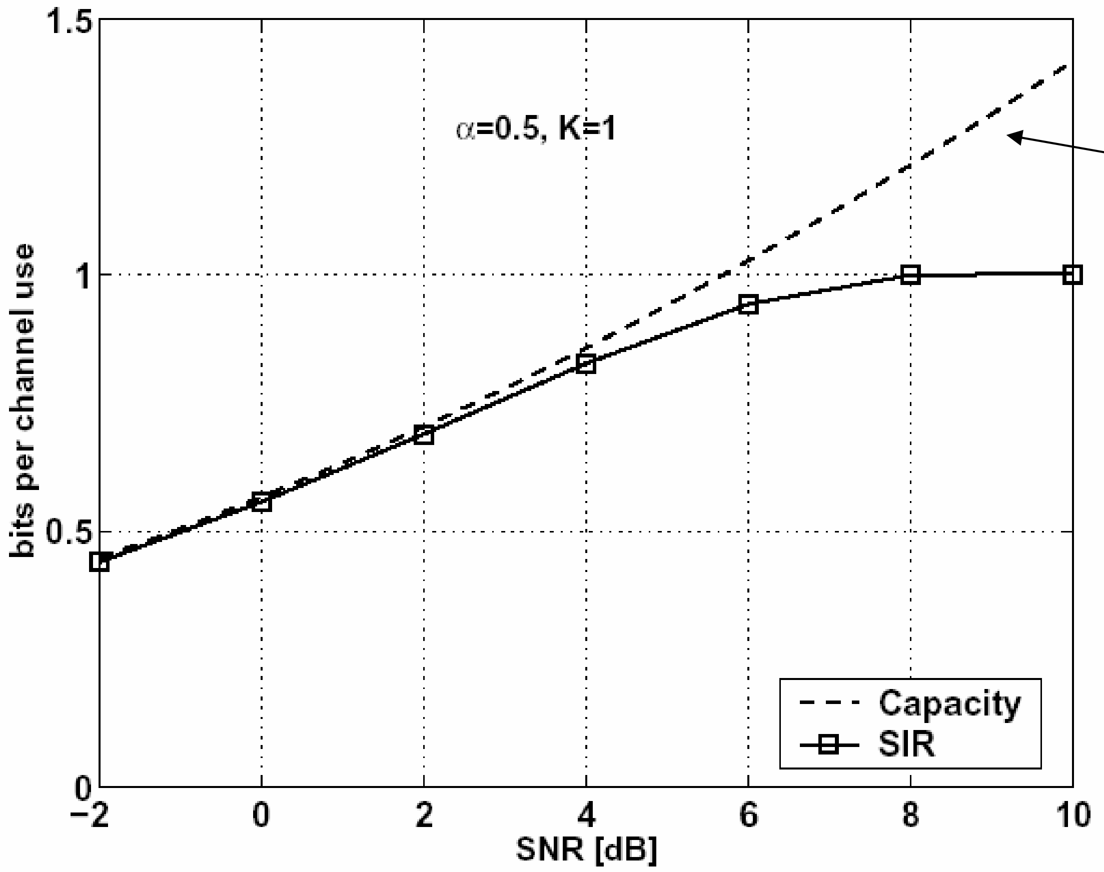
Experimental results ISI



Upper and lower bounds – Chen and Siegel, "On the symmetric information rate of two-dimensional finite state ISI channels", ITW 2003.

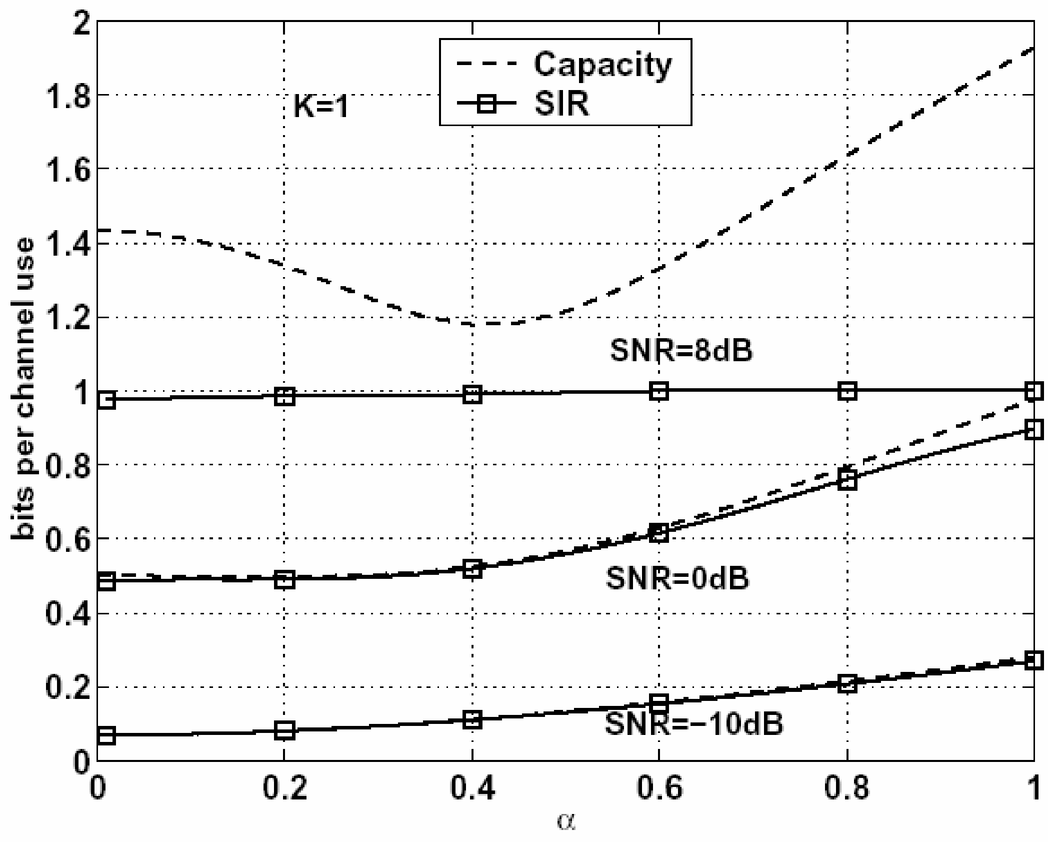
PASCAL

Experimental results: Wyner's model



Wyner's result

Experimental results: Wyner's model



Why do GBP-based CVM serve so remarkably?

- Currently we do not have a rigorous answer.
- Maybe the following empirical evidence may shed light on this issue:
  - The CVM results are not exact.
  - When does not it work? In a different setting : homogeneous antiferromagnetic interactions with random fields.
  - GBP converges to the same solution under different initial conditions (convexity of the Kikuchi free energy?).
  - 'Local' estimates may also perform well.

PASCAL

Local estimates and GBP

Count the number of sites for which $|h_i| > \sum_j |R_{ij}|$

Local exact inference: neighborhood size $\pm 2, \pm 3$