Source Coding with Low Density Nonlinear Nodes

Marc Mézard

Santa Fe, january 10-12, 2005

Collaboration

M. Mézard (Université Paris Sud and 'Stipco*' EC network)

S. Ciliberti (Université Paris Sud)

R. Zecchina (ICTP Trieste and 'Stipco' EC network)

+ Thanks to David Saad and Jonathan Yedidia

(∗) "Statistical Physics of Information Processing and Combinatorial Optimization", EC supported Research Training Network

General idea

- Success of LDPC codes in channel coding
- Seeking low density graphical codes in source coding
  - Parity Source Coder: theoretically close to optimum, no good encoding algorithm
  - Nonlinear Source Coder: theoretically close to optimum, good encoding with survey propagation

General idea

- Success of LDPC codes in channel coding
- Seeking low density graphical codes in source coding
  - Parity Source Coder: theoretically close to optimum, no good encoding algorithm
  - Nonlinear Source Coder: theoretically close to optimum, good encoding with survey propagation
- Two main issues:
  - Theoretical capacity: approaching the Shannon limit
  - Algorithmic performance: polynomial time encoding and decoding

Lossy Data Compression

$$\{y_1, y_2, \dots y_\mathrm{M}\} \xrightarrow[\text{encoding}]{} \{x_1, x_2 \dots x_\mathrm{N}\} \xrightarrow[\text{decoding}]{} \{y_1^*, y_2^* \dots y_\mathrm{M}^*\}$$

$y_a, y_a^* \in$ alphabet $\mathcal{S}$; $x_i \in \{0,1\}$.

Here $\mathcal{S} = \{0,1\}$, and y_a are iid, $= 0, 1$ with probability $1/2$.

Rate $R = N/M$. Distortion $D = (1/M)\ \sum_{i=1}^{M} \big(1 - \delta(y_a, y_a^*)\big)$

Lossy Data Compression

$$\{y_1, y_2, \ldots y_\mathrm{M}\} \xrightarrow[\text{encoding}]{} \{x_1, x_2 \ldots x_\mathrm{N}\} \xrightarrow[\text{decoding}]{} \{y_1^*, y_2^* \ldots y_\mathrm{M}^*\}$$

$y_a, y_a^* \in$ alphabet $\mathcal{S}$; $x_i \in \{0,1\}$.

Here $\mathcal{S} = \{0,1\}$, and y_a are iid, $= 0, 1$ with probability $1/2$.

Rate $R = N/M$. Distortion $D = (1/M)\ \sum_{i=1}^{M} \big(1 - \delta(y_a, y_a^*)\big)$

Shannon: the minimum achievable rate R given the distortion D is

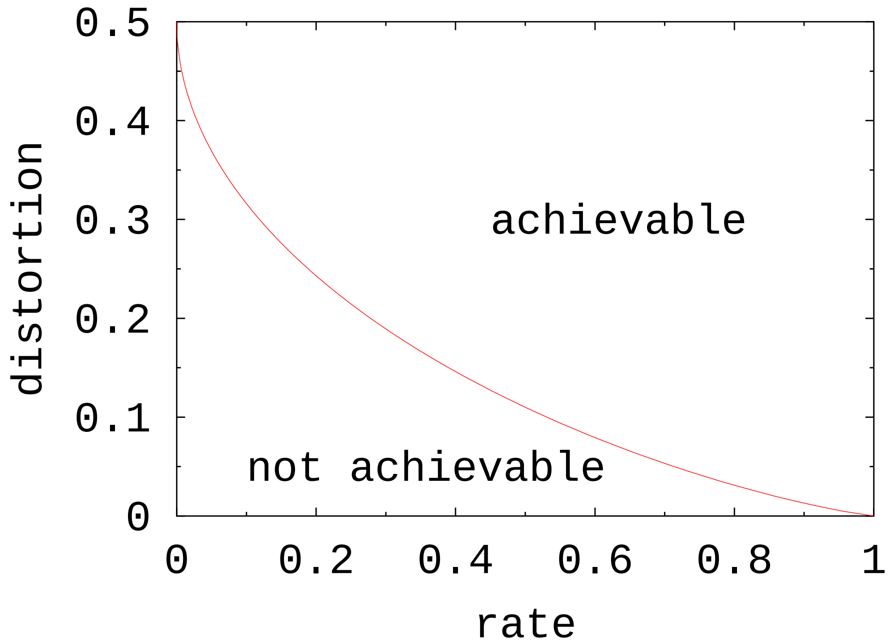
$$R(D) = 1 - h_2(D)$$

Shannon's bound

Uncorrelated unbiased binary source

$$R(D) = 1 - h_2(D)$$

$h_2(\cdot)$ being the binary entropy

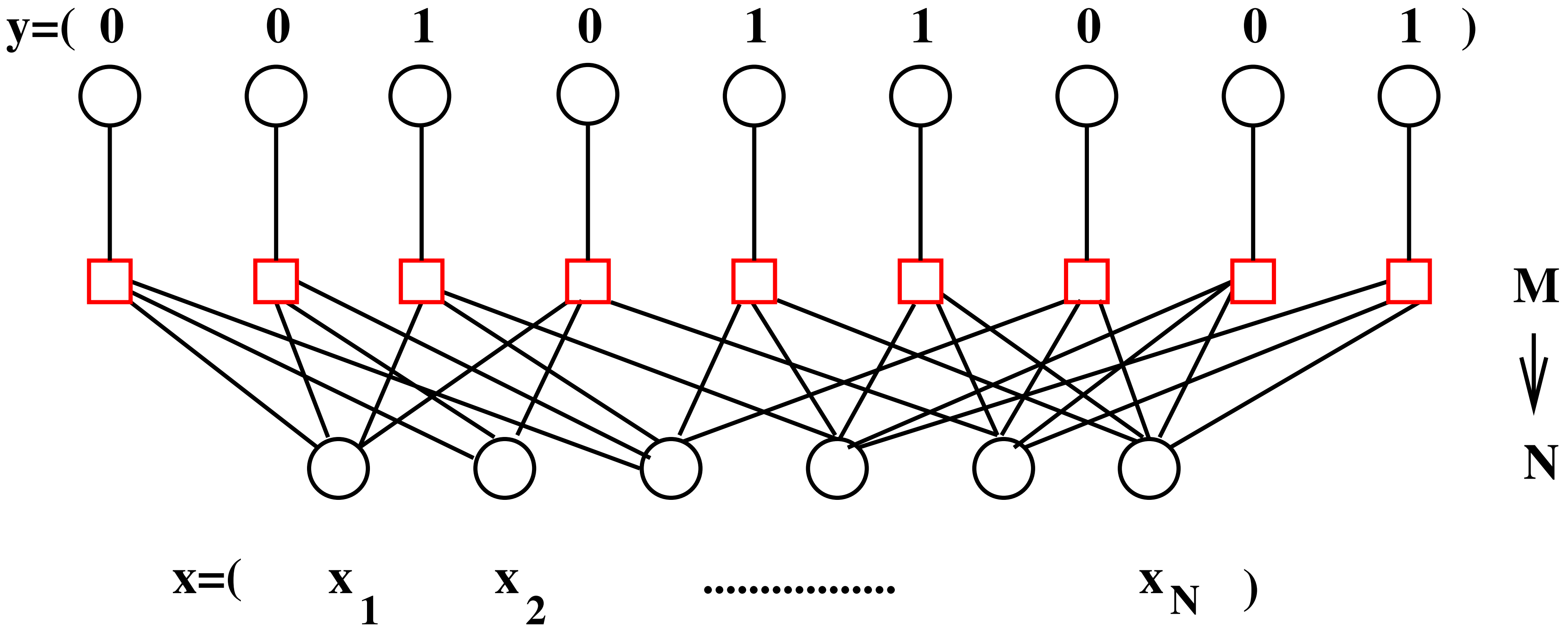


General strategy

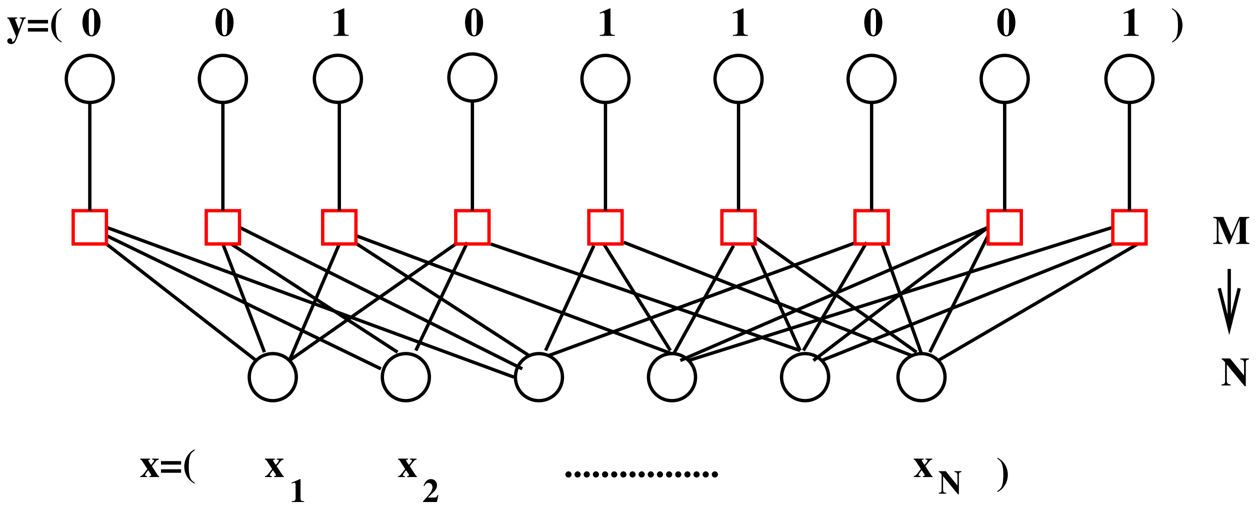
Compressed message: $\{x_1, x_2 ... x_{\mathrm{N}}\} = N$ bits

Initial message: M symbols. $\rightarrow$ M function nodes

Example: 'Parity Source coder': a LDPC code for compression

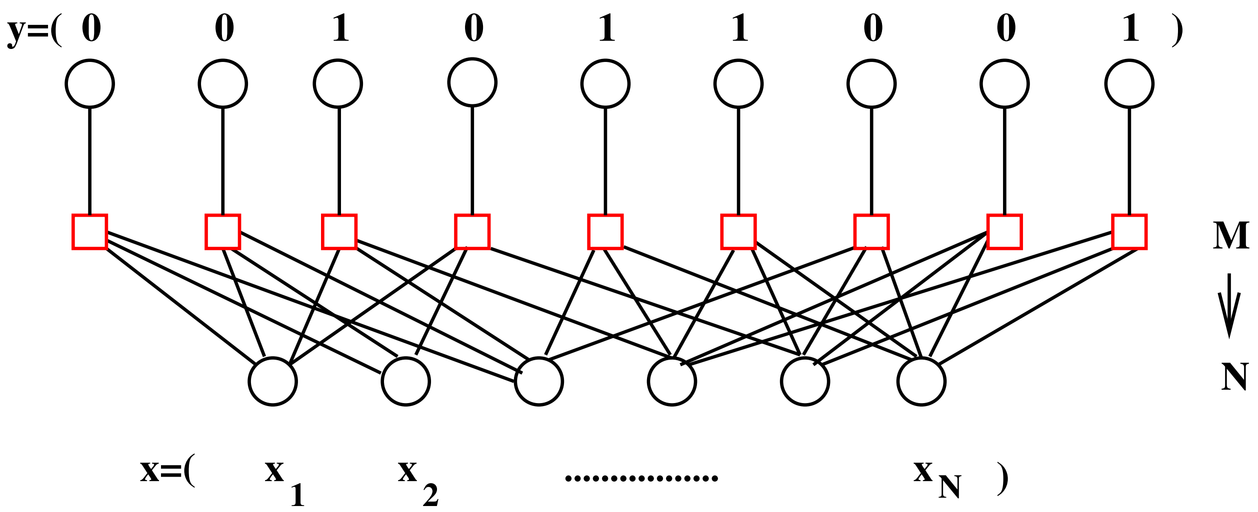


Distortion and energy



Encoding: $\mathbf{y} \to \mathbf{x}$= Find a configuration of x which violates the smallest number of checks. = Find a ground state.

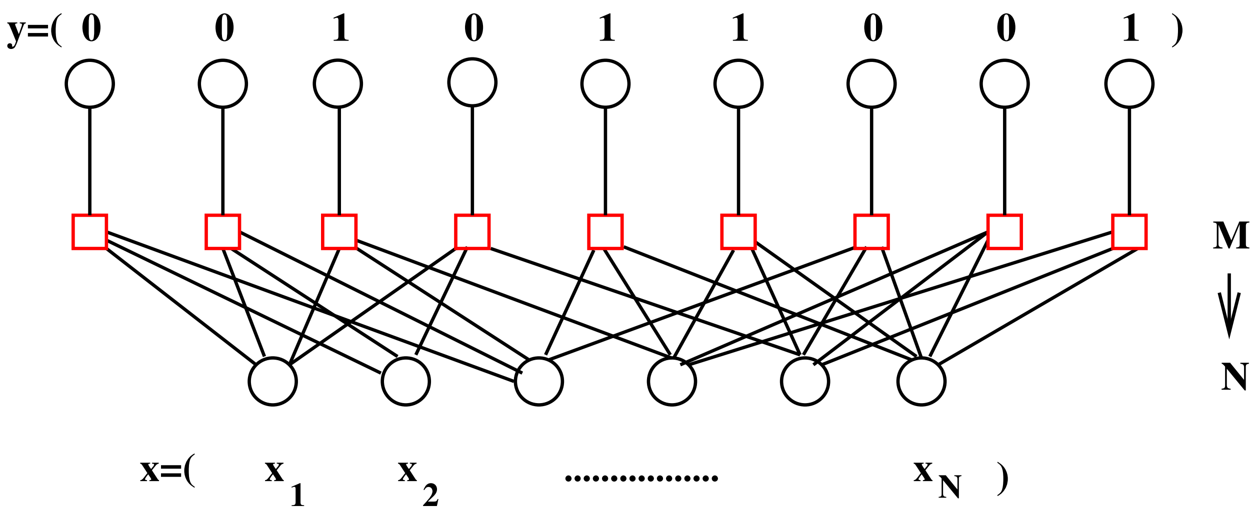
Distortion and energy



Encoding: $\mathbf{y} \to \mathbf{x}$= Find a configuration of x which violates the smallest number of checks. = Find a ground state.

Decoding: $\mathbf{x} \to \mathbf{y}^*$= trivial. Example: $y_1^* = x_1 \oplus x_2 \oplus x_3$

Distortion and energy



Encoding: $\mathbf{y} \rightarrow \mathbf{x}$= Find a configuration of x which violates the smallest number of checks. = Find a ground state.

Decoding: $\mathbf{x} \rightarrow \mathbf{y}^*$= trivial. Example: $y_1^* = x_1 \oplus x_2 \oplus x_3$

Distortion = number of violated checks in the encoding = ground state energy of the encoding.

Theoretical performance of the Parity Source Coder I

'XORSAT' problem in combinatorial optimization, or 'p-spin' model in statistical physics.

Instance= $\{$ Geometry, $y_1, \ldots, y_M\}$. $E(x) = \sum_a \left[1 - \delta\left(\oplus_{i\in V(a)} x_i, y_a\right)\right]$

SAT configuration: x with $E(x) = 0$. UNSAT instance: $\forall x : E(x) > 0$.

MAX-XORSAT: ground state energy E_0 (=distortion)

Theoretical performance of the Parity Source Coder I

'XORSAT' problem in combinatorial optimization, or 'p-spin' model in statistical physics.

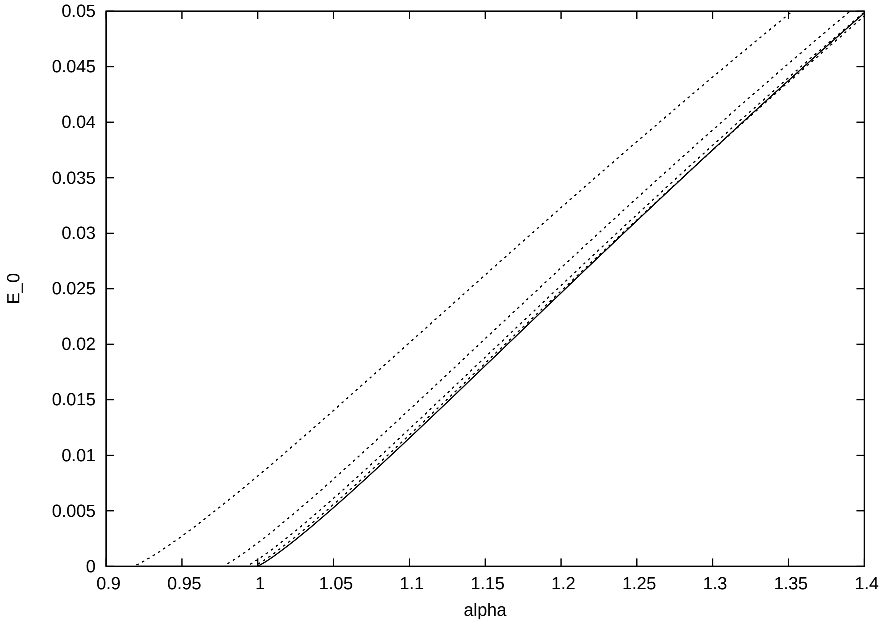
Instance= $\{$ Geometry, $y_1, \ldots, y_M\}$. $E(x) = \sum_a \left[1 - \delta\left(\oplus_{i \in V(a)} x_i, y_a\right)\right]$

SAT configuration: x with $E(x) = 0$. UNSAT instance: $\forall x : E(x) > 0$.

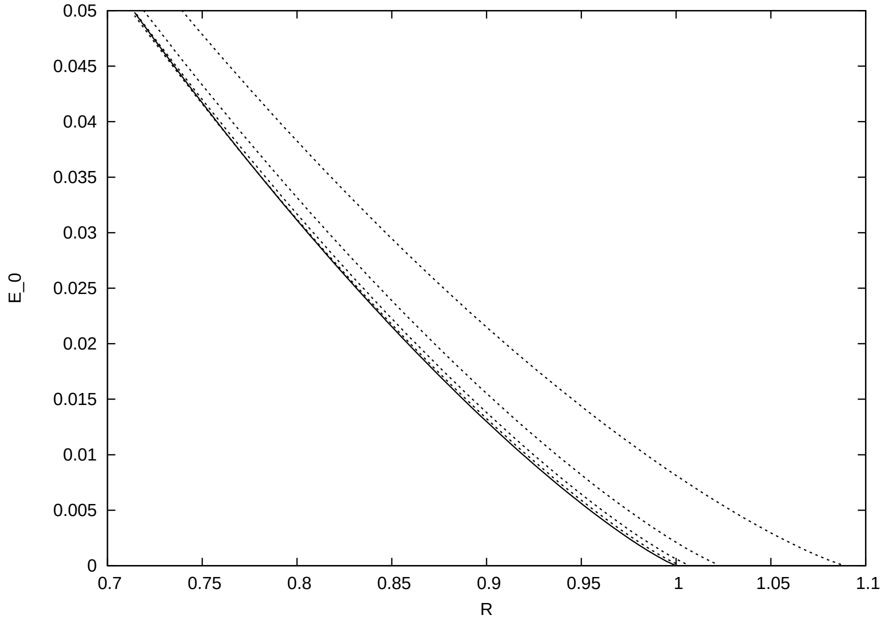
MAX-XORSAT: ground state energy E_0 (=distortion)

Analytic study: "1-step RSB" cavity method (MM,Ricci-Tersenghi,Zecchina 2002). Random $K-$XORSAT problem with $M, N \to \infty$ and $M/N = \alpha\ (= 1/R)$. (Function nodes: degree K. Variable nodes: degree Poisson($K\alpha$)). $\to$ Phase diagram, E_0.

Theoretical performance of the Parity Source Coder II



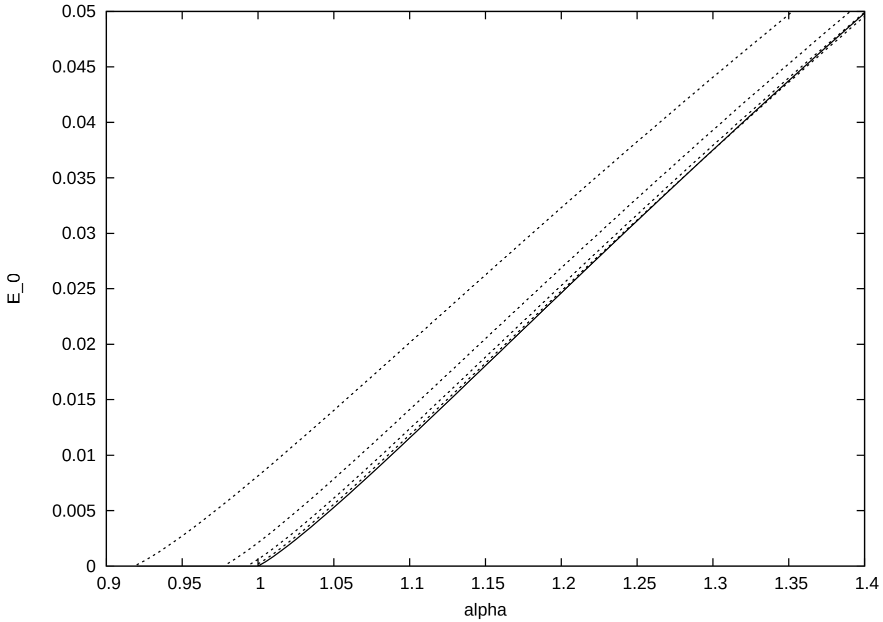
Ground state energy E_0
$K = 3, 4, 5, 6$ and Shannon



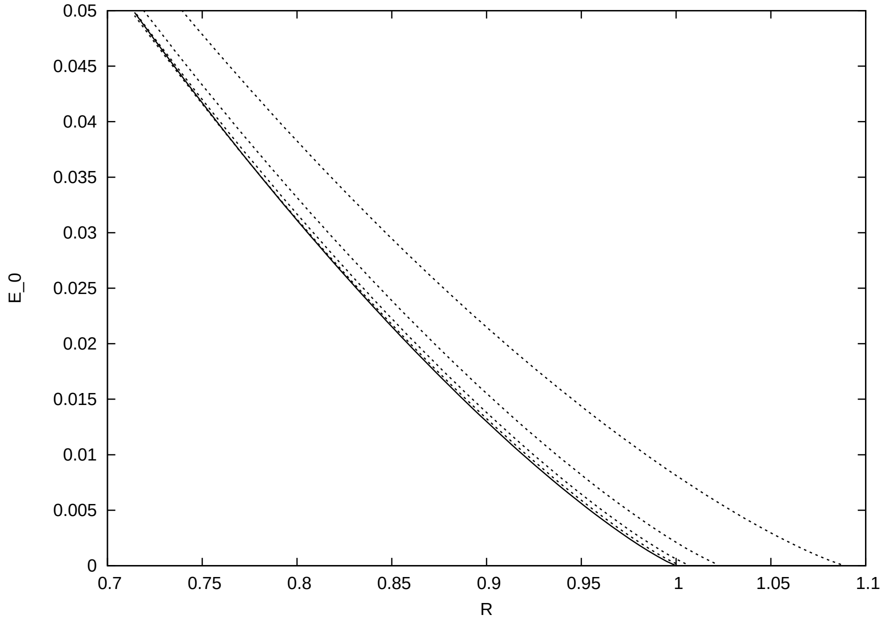
Predicted distortion versus Rate

Rapidly approaching Shannon's bound when K increases. Optimal data compression...

Theoretical performance of the Parity Source Coder II



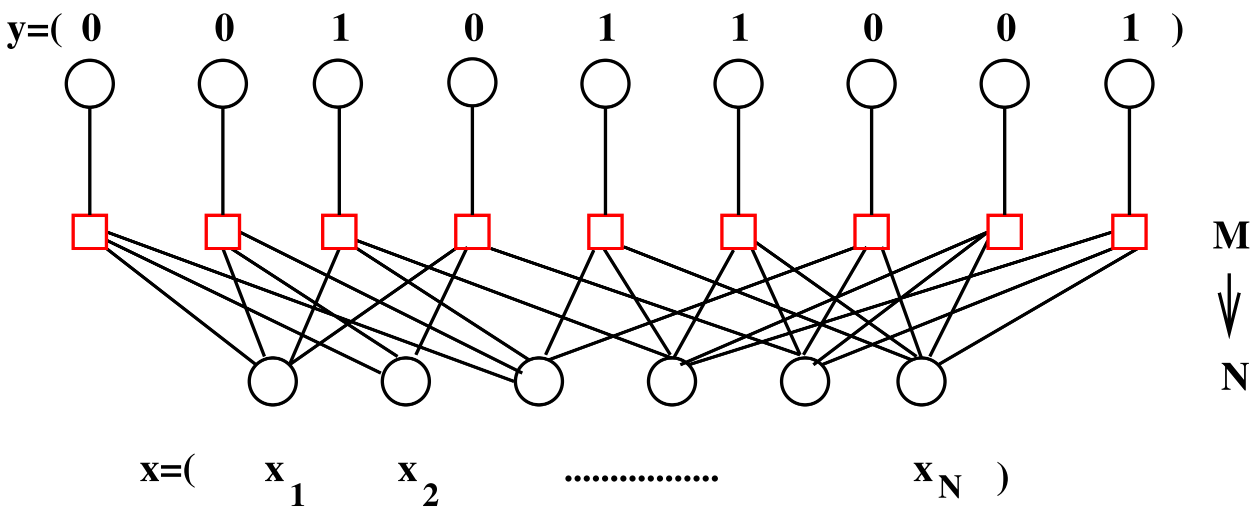
Ground state energy E_0
$K = 3, 4, 5, 6$ and Shannon



Predicted distortion versus Rate

Rapidly approaching Shannon's bound when K increases. Optimal data compression... in theory!

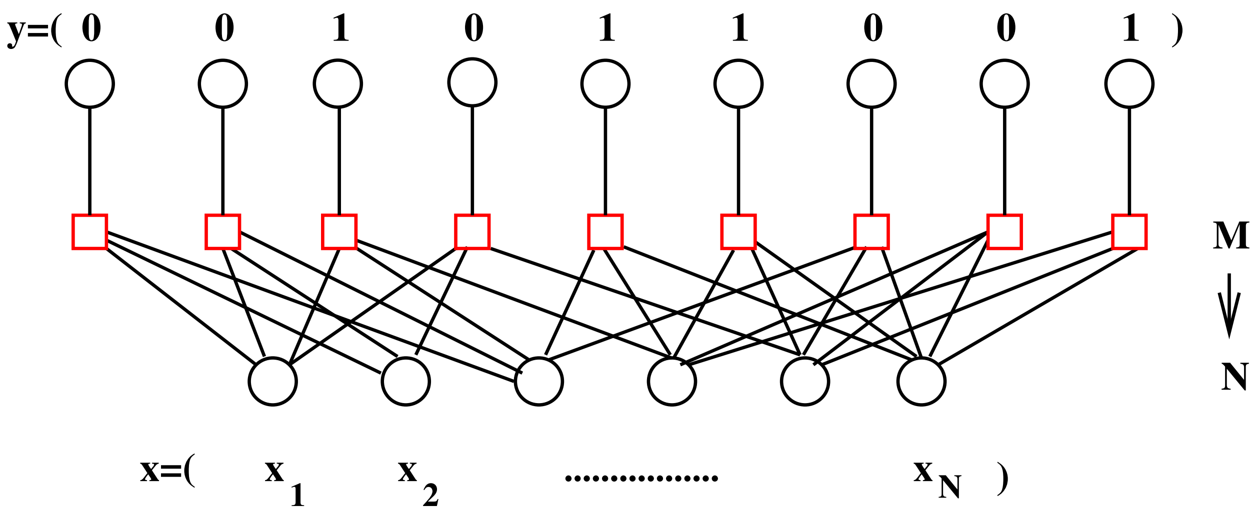
Practical performance of the Parity Source Coder



Encoding: $\mathbf{y} \to \mathbf{x}=$ Find a configuration of x which violates the smallest number of checks. $=$ Find a ground state. DIFFICULT!!!

Belief propagation: does not converge (NB: random initial condition, very different from the case of LDPC codes).

Practical performance of the Parity Source Coder



Encoding: $\mathbf{y} \to \mathbf{x}=$ Find a configuration of x which violates the smallest number of checks. $=$ Find a ground state. DIFFICULT!!!

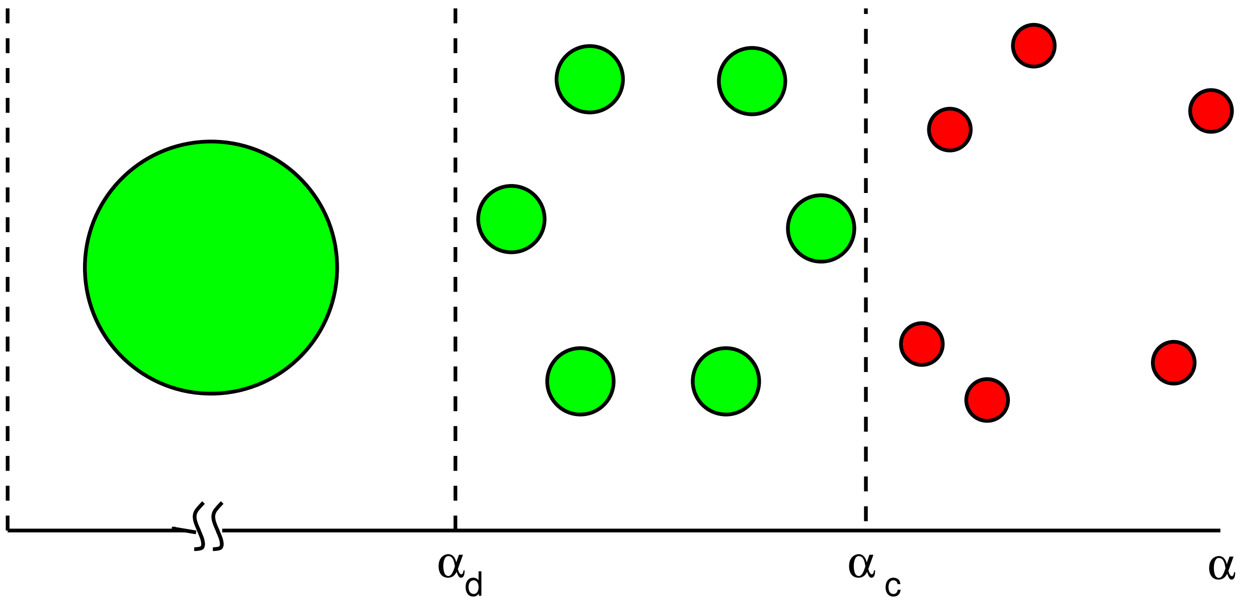
Belief propagation: does not converge (NB: random initial condition, very different from the case of LDPC codes).

Clusters of solutions

Clusters in the XORSAT problem

Random parity checks (graph, y_a); $N, M \to \infty$, $\alpha = M/N$ fixed.

Topology of configurations with $E = E_0$:



With probability one:

$\alpha < \alpha_c$: SAT

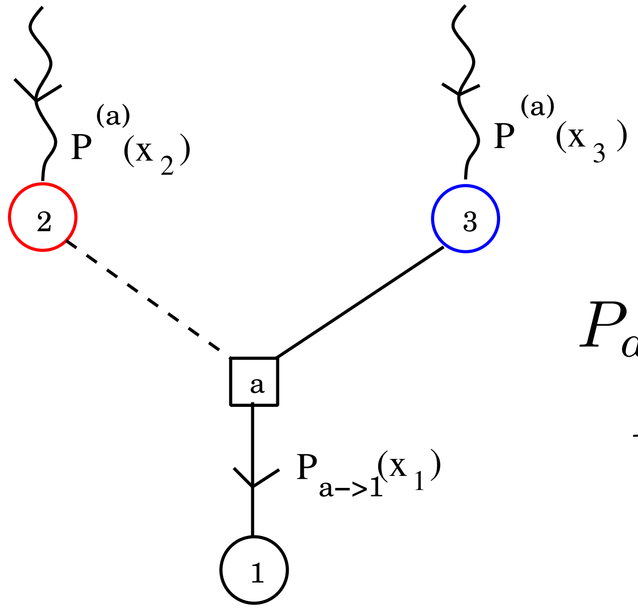
$\alpha > \alpha_c$: UNSAT

But three phases:

Easy SAT, Hard SAT, UNSAT.

K	3	4	5	6
α_c	.92	.97	.99	1.00
α_d	.82	.77	.70	.63

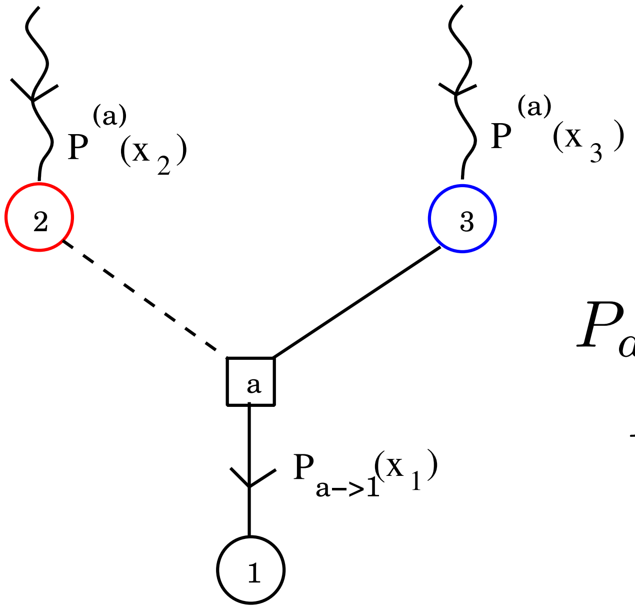
Why clusters 'kill' belief propagation



Belief propagation:

$$P_{a\to 1}(x_1) = \sum_{x_2,x_3} C_a(x_1,x_2,x_3)P^{(a)}(x_2)P^{(a)}(x_3)$$
$$P^{(b)}(x_1) \propto \prod_{a\in V(1)\backslash b} P_{a\to 1}(x_1)$$

Why clusters 'kill' belief propagation



Belief propagation:

$$P_{a\to 1}(x_1) = \sum_{x_2,x_3} C_a(x_1,x_2,x_3) P^{(a)}(x_2) P^{(a)}(x_3)$$
$$P^{(b)}(x_1) \propto \prod_{a\in V(1)\setminus b} P_{a\to 1}(x_1)$$

Basic underlying idea: $P^{(a)}(x_2,x_3) \sim P^{(a)}(x_2)P^{(a)}(x_3)$. Correct if

1) x_2, x_3 distant (OK)

2) Measure restricted to one cluster (Wrong in the hard SAT phase).

From belief propagation to survey propagation

Hard SAT phase: Message = Survey of the elementary messages in the clusters of SAT configurations. Project: Belief $\rightarrow$ Warning $\rightarrow$ Survey

Belief $P_{a\rightarrow i}(x_i), P^{(a)}(x_i)$ are probabilities, in $[0,1]$.

For each belief, e.g. $P_{a\rightarrow i}(x_i)$, construct the warning $\rho_{a\rightarrow i}(x_i) \in \{0,*\}$.

$$\rho = I(P) = \begin{cases} 0 & \text{if} \quad P_{a\rightarrow i}(x_i) = 0 \\ * & \text{if} \quad P_{a\rightarrow i}(x_i) > 0 \end{cases}$$

Warning propagation: focuses on forced variables.

From belief propagation to survey propagation

Hard SAT phase: Message = Survey of the elementary messages in the clusters of SAT configurations. Project: Belief $\rightarrow$ Warning $\rightarrow$ Survey

Belief $P_{a\rightarrow i}(x_i), P^{(a)}(x_i)$ are probabilities, in $[0,1]$.

For each belief, e.g. $P_{a\rightarrow i}(x_i)$, construct the warning $\rho_{a\rightarrow i}(x_i) \in \{0,*\}$.

$$\rho = I(P) = \begin{cases} 0 & \text{if } \; P_{a\rightarrow i}(x_i) = 0 \\ * & \text{if } \; P_{a\rightarrow i}(x_i) > 0 \end{cases}$$

Warning propagation: focuses on forced variables.

Survey: $\eta_{a\rightarrow i}$ = probability of a warning being sent from constraint a to variable i, when a cluster is picked up at random. Propagate the surveys along the graph, then use them to decimate. OK in satisfiability/colouring problem.

From belief propagation to survey propagation

Hard SAT phase: Message = Survey of the elementary messages in the clusters of SAT configurations. Project: Belief $\longrightarrow$ Warning $\longrightarrow$ Survey

Belief $P_{a\to i}(x_i), P^{(a)}(x_i)$ are probabilities, in $[0,1]$.

For each belief, e.g. $P_{a\to i}(x_i)$, construct the warning $\rho_{a\to i}(x_i) \in \{0,*\}$.

$$\rho = I(P) = \begin{cases} 0 & \text{if} \quad P_{a\to i}(x_i) = 0 \\ * & \text{if} \quad P_{a\to i}(x_i) > 0 \end{cases}$$

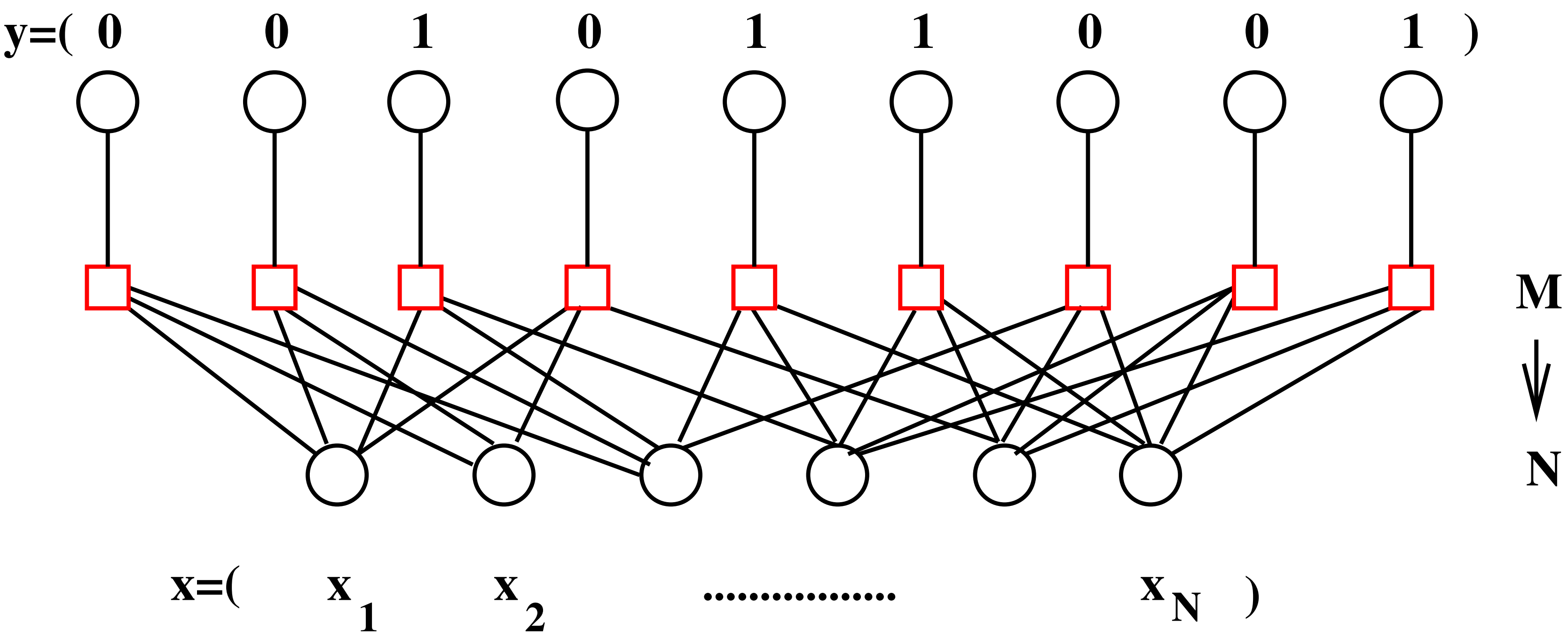
Warning propagation: focuses on forced variables.

Survey: $\eta_{a\to i}$ = probability of a warning being sent from constraint a to variable i, when a cluster is picked up at random. Propagate the surveys along the graph, then use them to decimate. OK in satisfiability/colouring problem.Not in XORSAT (totally symmetric messages)...

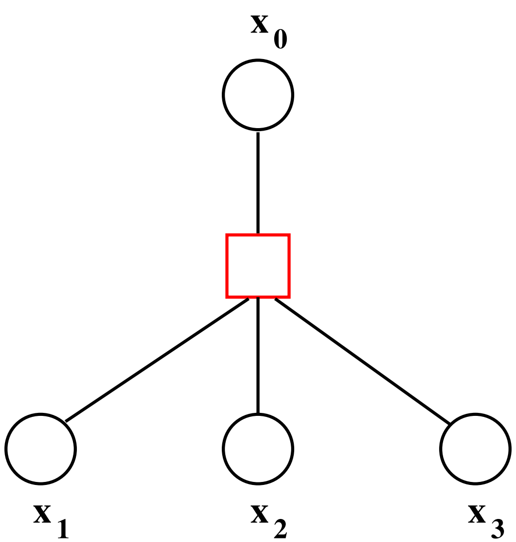
Nonlinear nodes

Introduce generalized function nodes, different from parity checks

- Keep theoretical performances nearly as good as in parity checks
- Break the symmetry $\rightarrow$ SP converges and allows to encode a message



Random Nonlinear nodes



x_1	x_2	x_3	Parity check $x_0=0$	Random check $x_0=0$	Random check $x_0=1$
0	0	0	S	S	U
0	0	1	U	S	U
0	1	0	U	U	S
1	0	0	U	S	U
0	1	1	S	U	S
1	0	1	S	U	S
1	1	0	S	U	S
1	1	1	U	S	U

SCRAMBLE

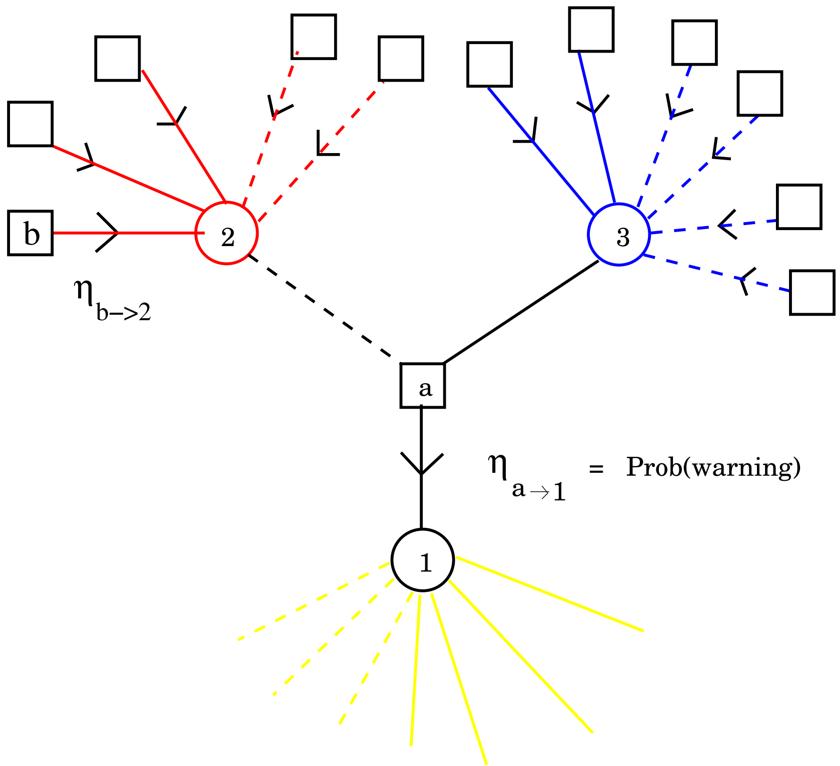
Symmetry

Cavity analysis and Message Passing Algorithms

Warning propagation = hard constraint limit of belief propagation.

Cavity analysis and Message Passing Algorithms

Warning propagation = hard constraint limit of belief propagation.Survey Propagation (SP): On each edge $a \to i$: survey = Proba(warning), when a cluster of ground states is chosen at random.



$\eta_{a\to 1}$: known exactly from

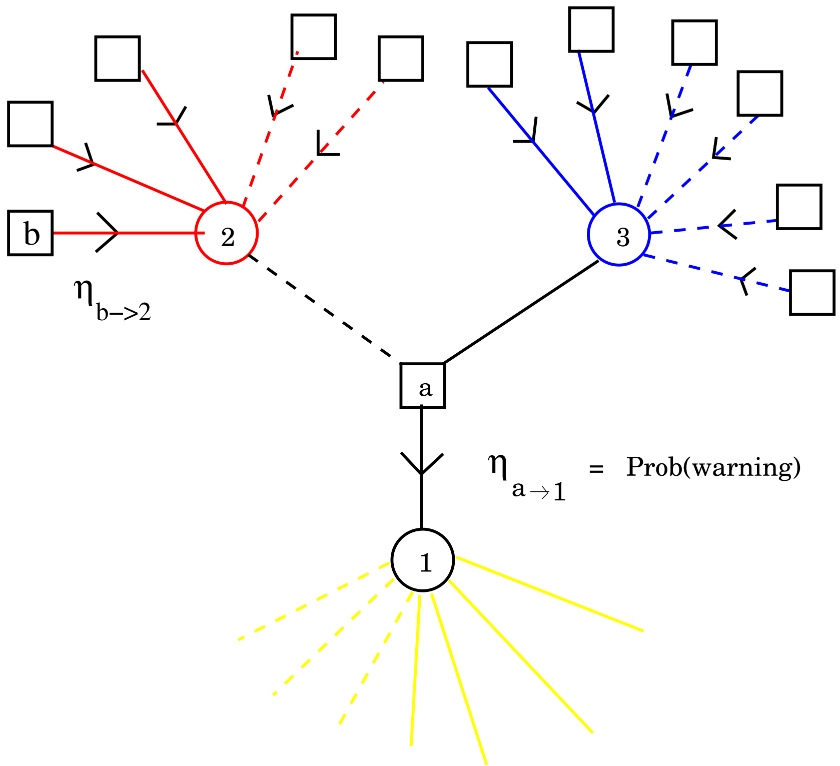
$\eta_{b\to 2}$ and $\eta_{c\to 3}$.

Penalty when conflicting warnings

$\to$ works also in the UNSAT phase

Cavity analysis and Message Passing Algorithms

Warning propagation = hard constraint limit of belief propagation.Survey Propagation (SP): On each edge $a \longrightarrow i$: survey = Proba(warning), when a cluster of ground states is chosen at random.



$\color{blue}{\eta_{a\rightarrow 1}}$: known exactly from

$\eta_{b\rightarrow 2}$ and $\eta_{c\rightarrow 3}$.

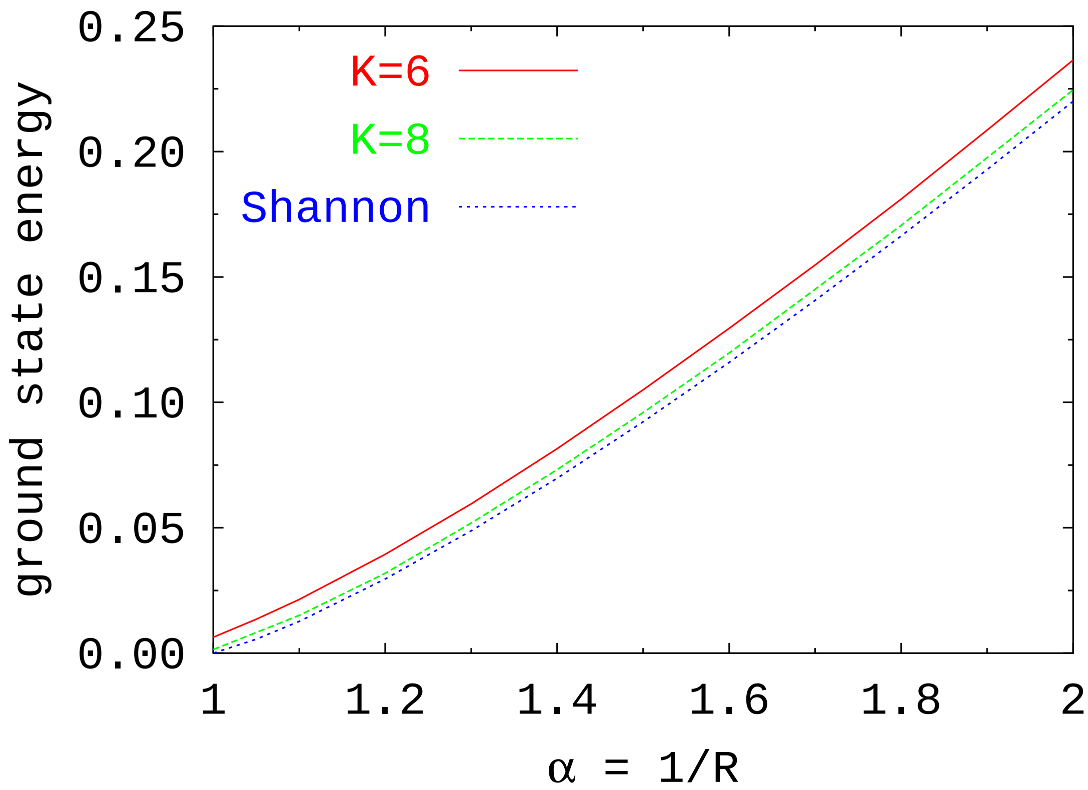
Penalty when conflicting warnings

$\longrightarrow$ works also in the UNSAT phase

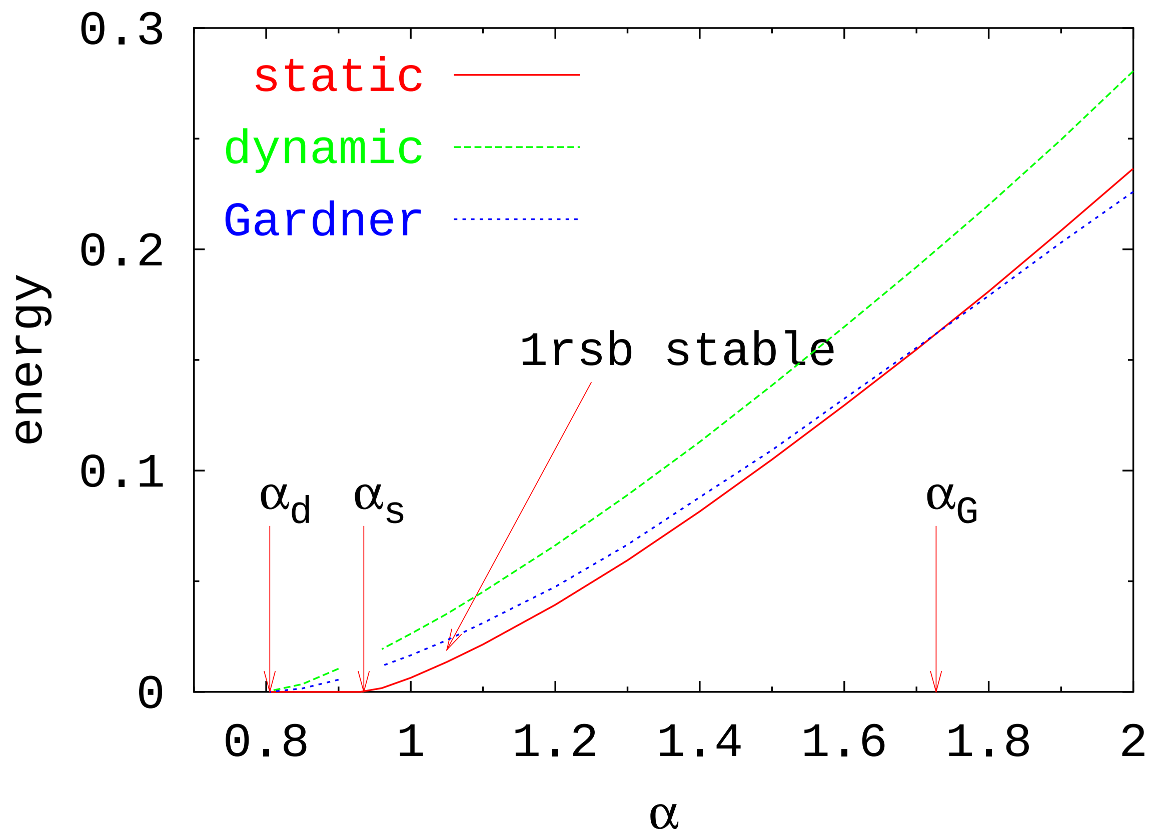
Statistical analysis $\longrightarrow$ phase diagram.

Single sample $\longrightarrow$ SP algorithm

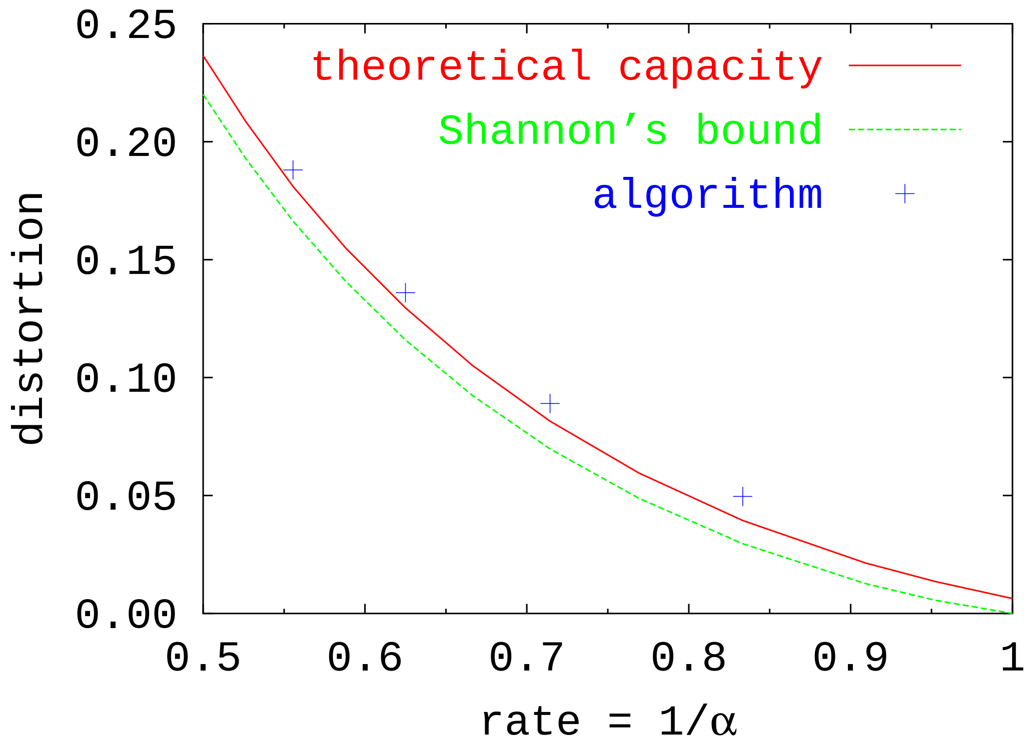
Theoretical capacity (20 types of random nodes)



Phase diagram



Performance



Conclusions

- New approach for lossy data compression based on low density constraint satisfaction problems

- Theoretical capacity $\approx$ Shannon's bound

- Message passing algorithms converge on CSP with non-linear nodes and stop just above the ground state energy $\Leftrightarrow$ approaching the Shannon's bound

Conclusions

- New approach for lossy data compression based on low density constraint satisfaction problems

- Theoretical capacity $\approx$ Shannon's bound

- Message passing algorithms converge on CSP with non-linear nodes and stop just above the ground state energy $\Leftrightarrow$ approaching the Shannon's bound

- To be improved: Encoding works, but still slow (limited to $N < 10000$, using general purpose SP software...

Conclusions

- New approach for lossy data compression based on low density constraint satisfaction problems
- Theoretical capacity $\approx$ Shannon's bound
- Message passing algorithms converge on CSP with non-linear nodes and stop just above the ground state energy $\Leftrightarrow$ approaching the Shannon's bound
- To be improved: Encoding works, but still slow (limited to $N < 10000$, using general purpose SP software...
- Perspectives: Generalize this algorithm in order to compress sequences of real numbers. Revisit nonlinear function nodes in channel coding.