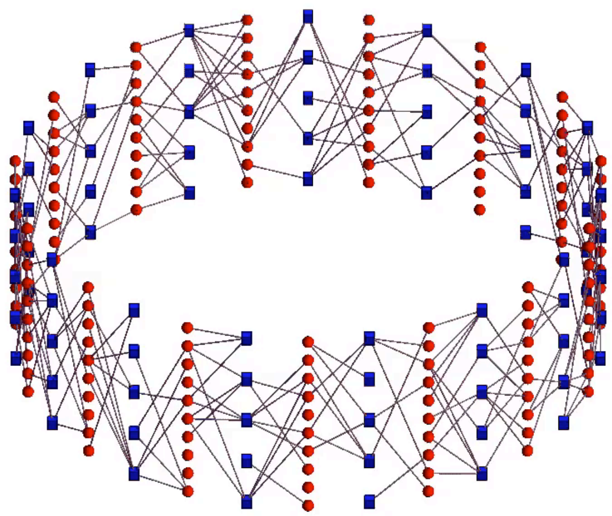
Spatial Coupling as a Proof Technique



ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

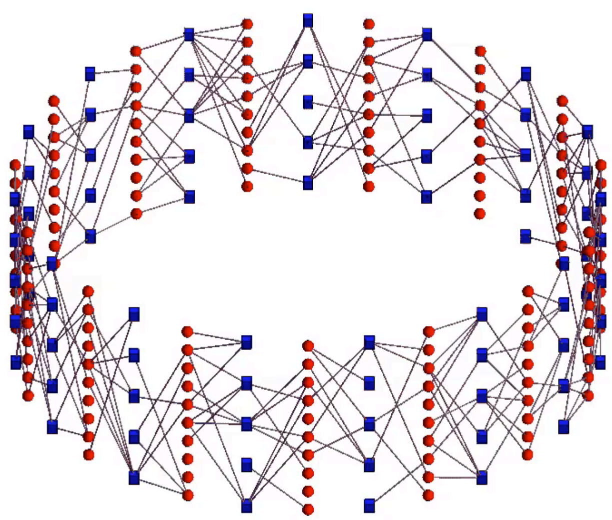
MCM Muenich 2015

Thursday, July 30th, 2015

Based on joint work with D. Achlioptas (UCSD), H. Hassani (EPF-L-Z), and Nicolas Macris (EPFL)

(Some) slides courtesy Nicolas Macris

Spatial Coupling as a Proof Technique



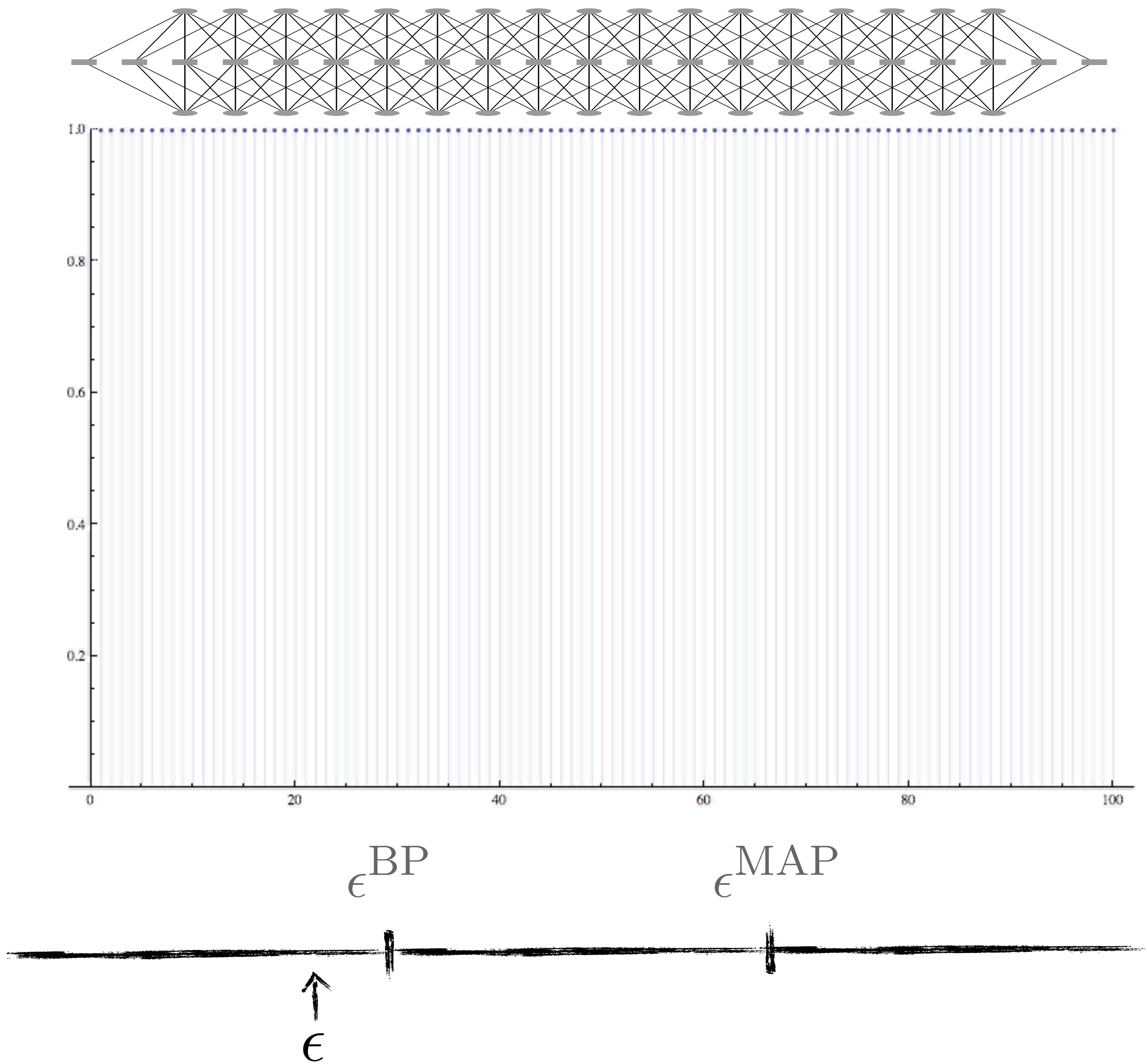
MCM Muenich 2015

Thursday, July 30th, 2015

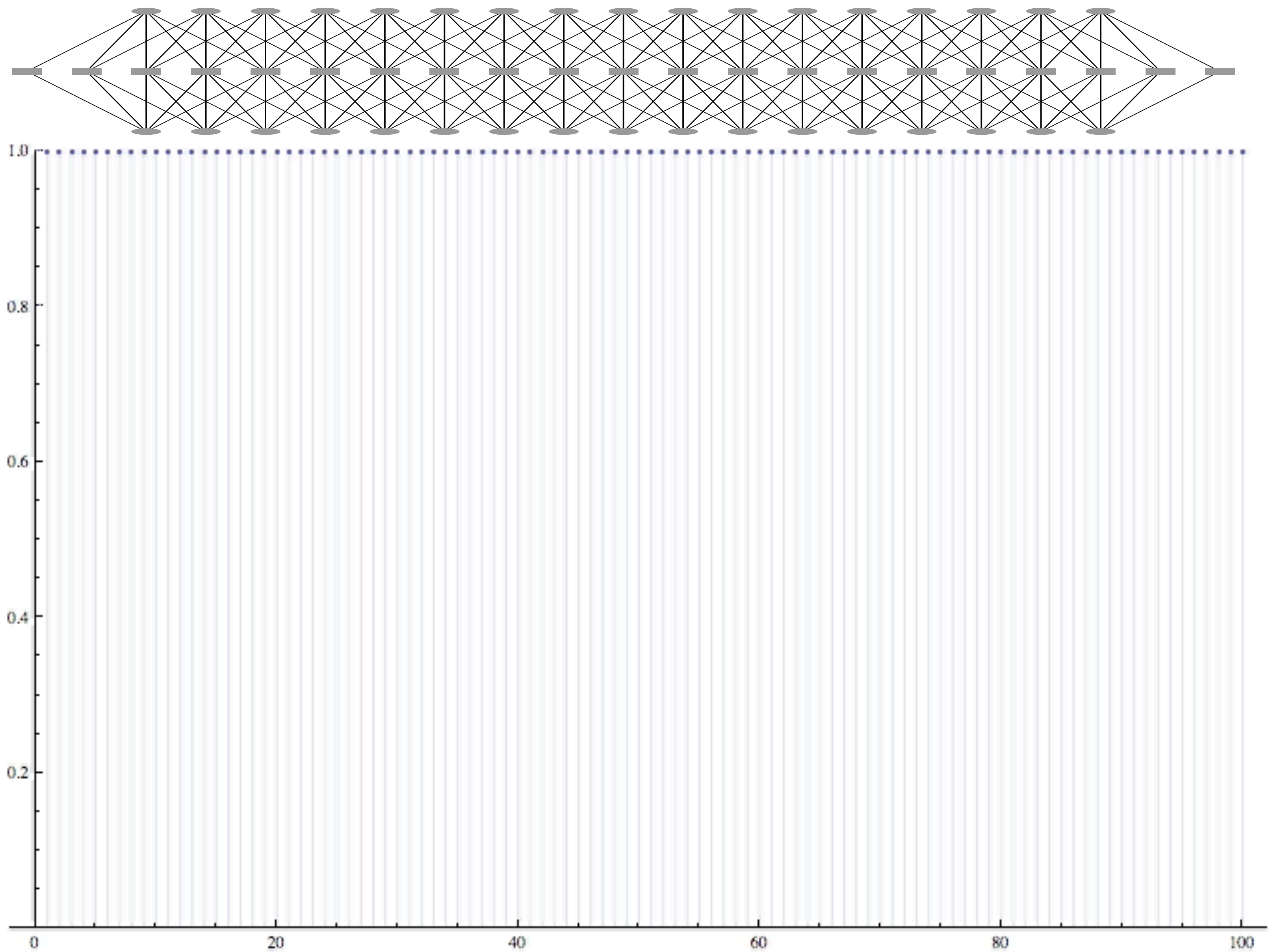
Based on joint work with D. Achlioptas (UCSD), H. Hassani (EPF-L-Z), and Nicolas Macris (EPFL)

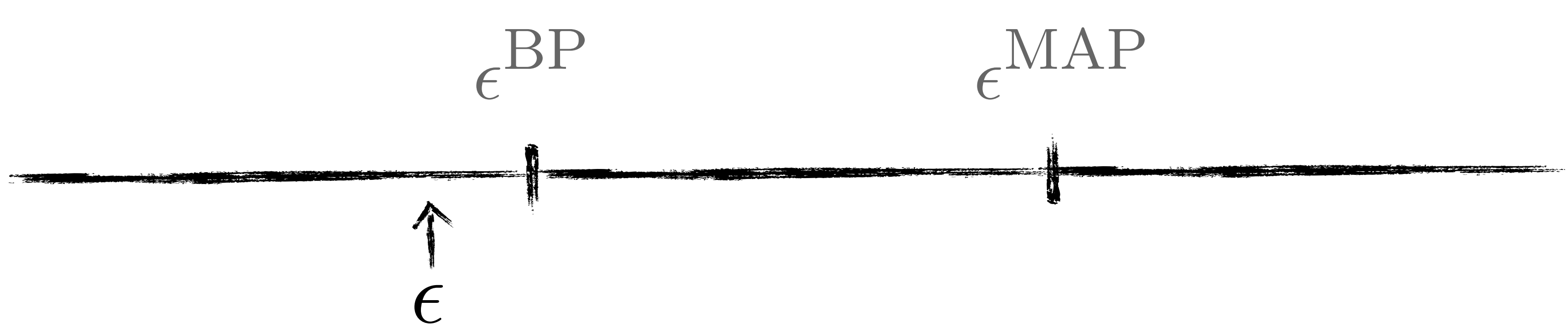
(Some) slides courtesy Nicolas Macris

DE for Coupled Ensemble

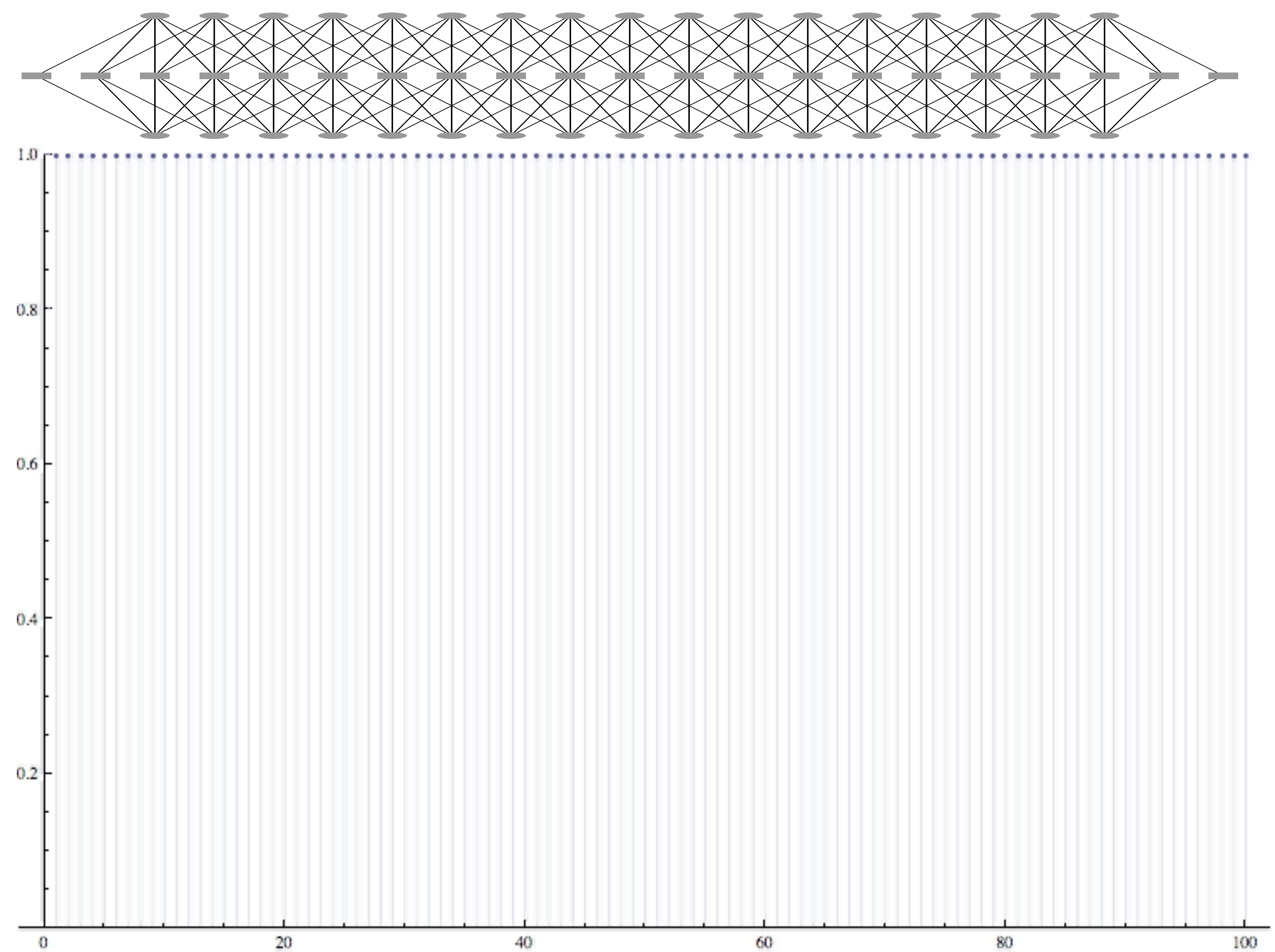


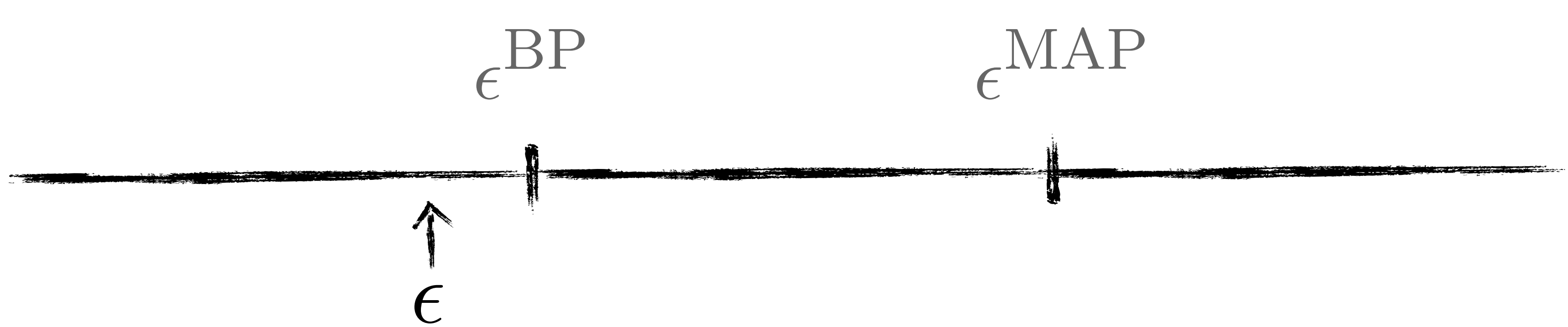
DE for Coupled Ensemble



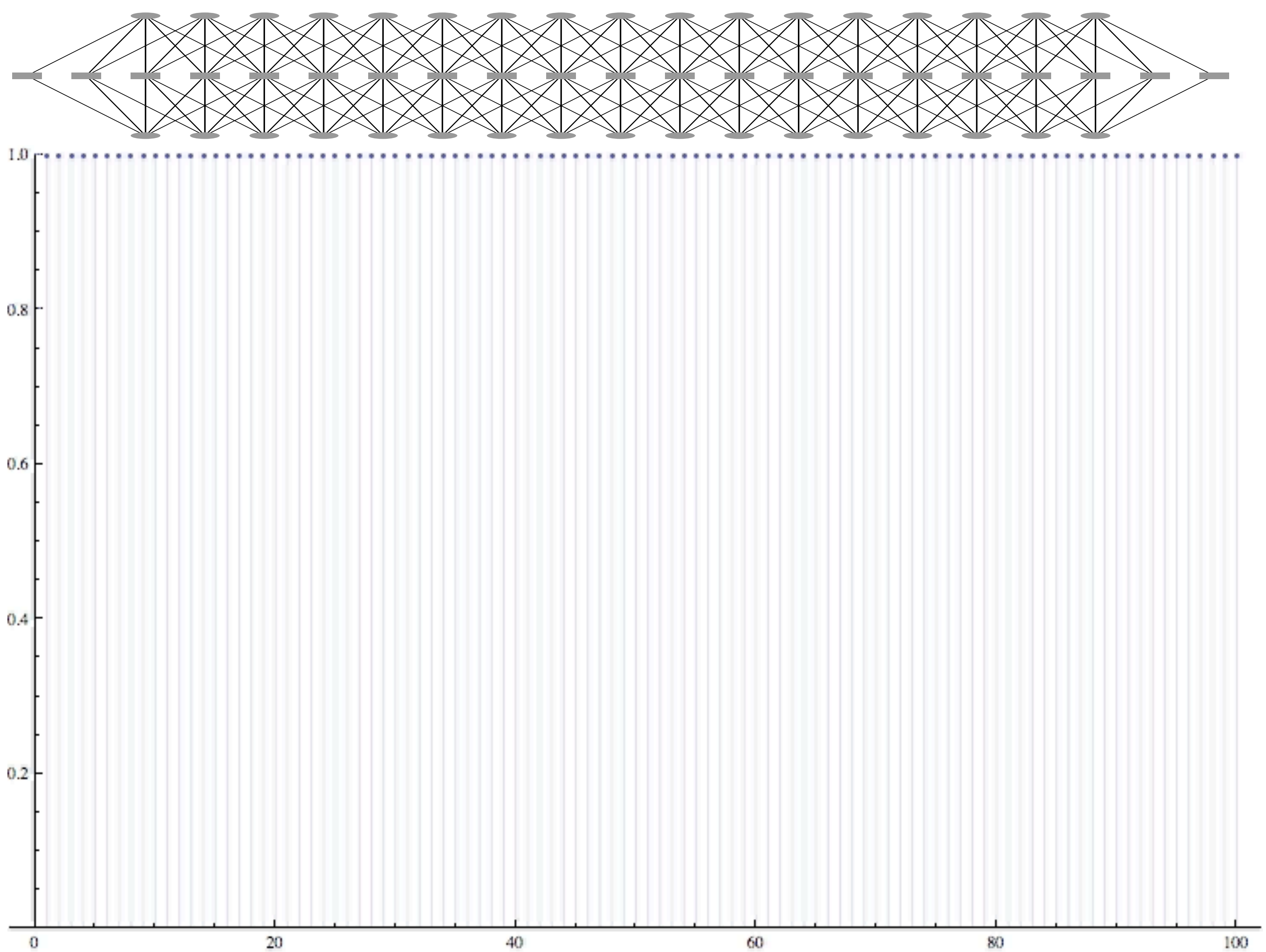


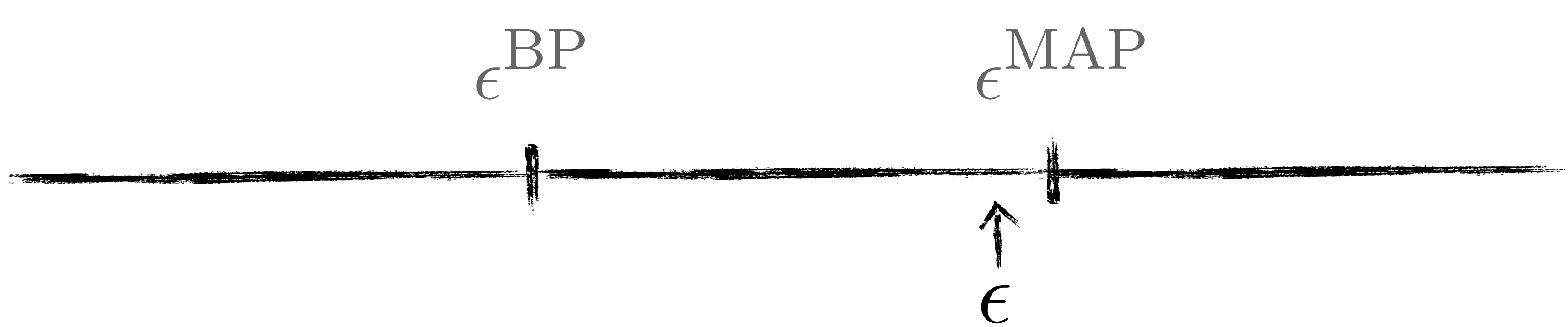
DE for Coupled Ensemble



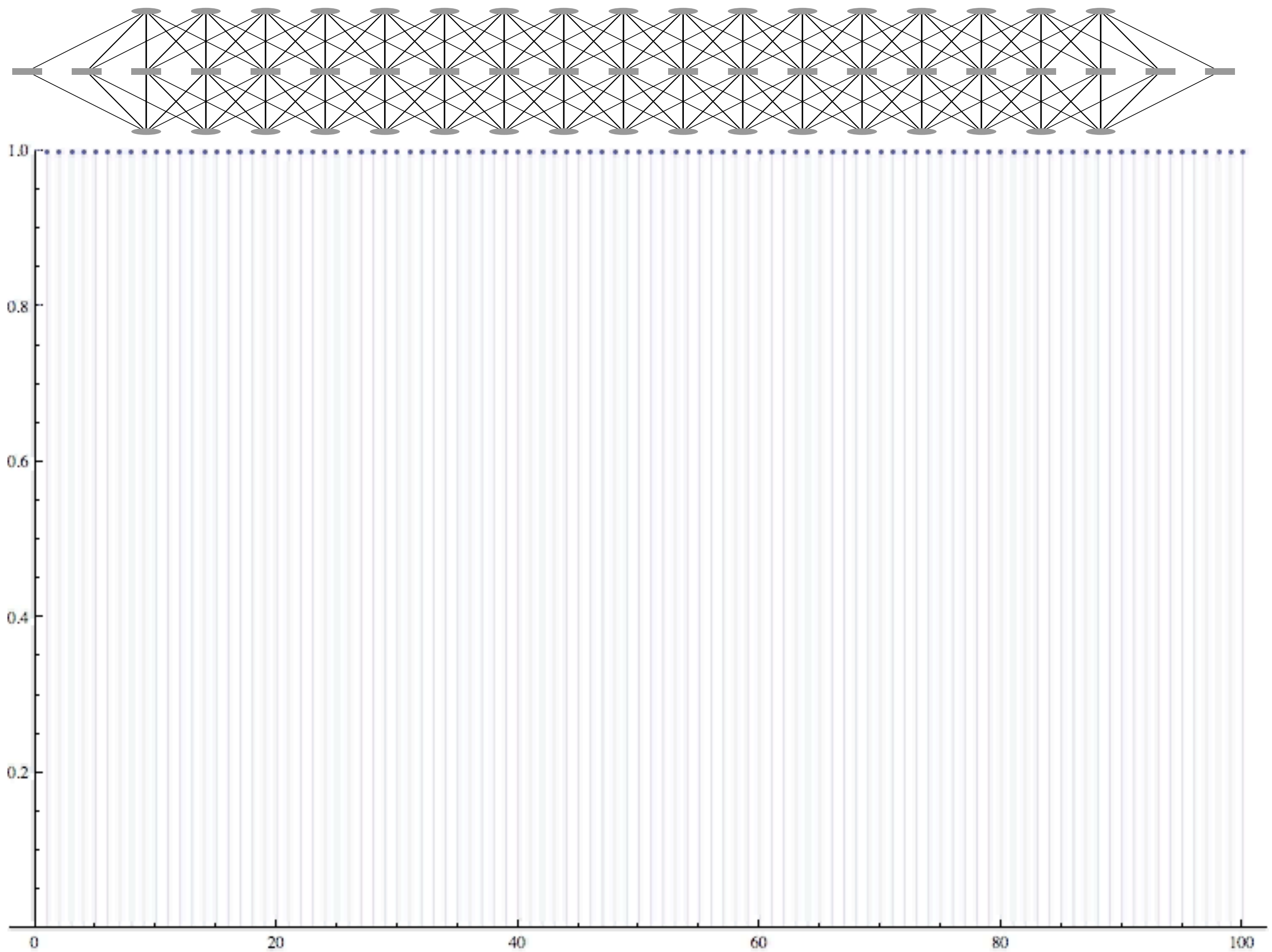


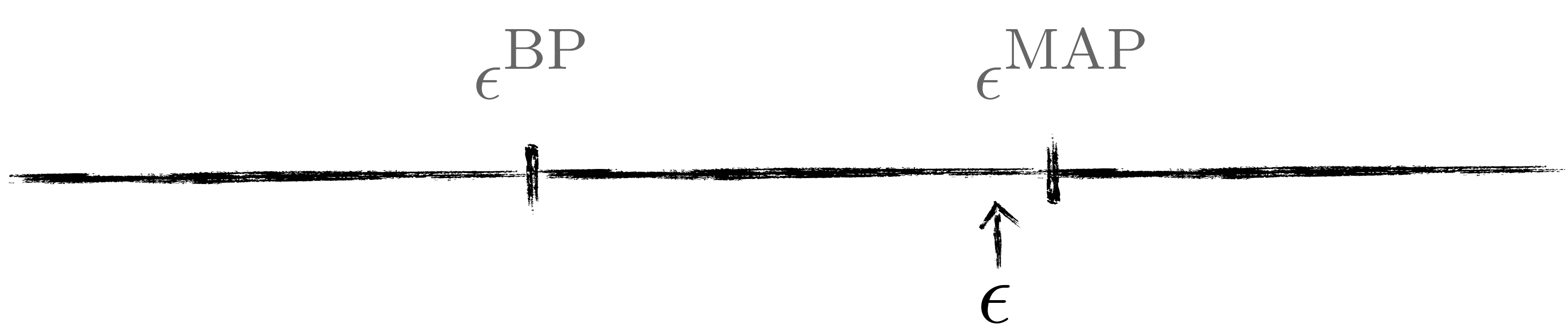
DE for Coupled Ensemble



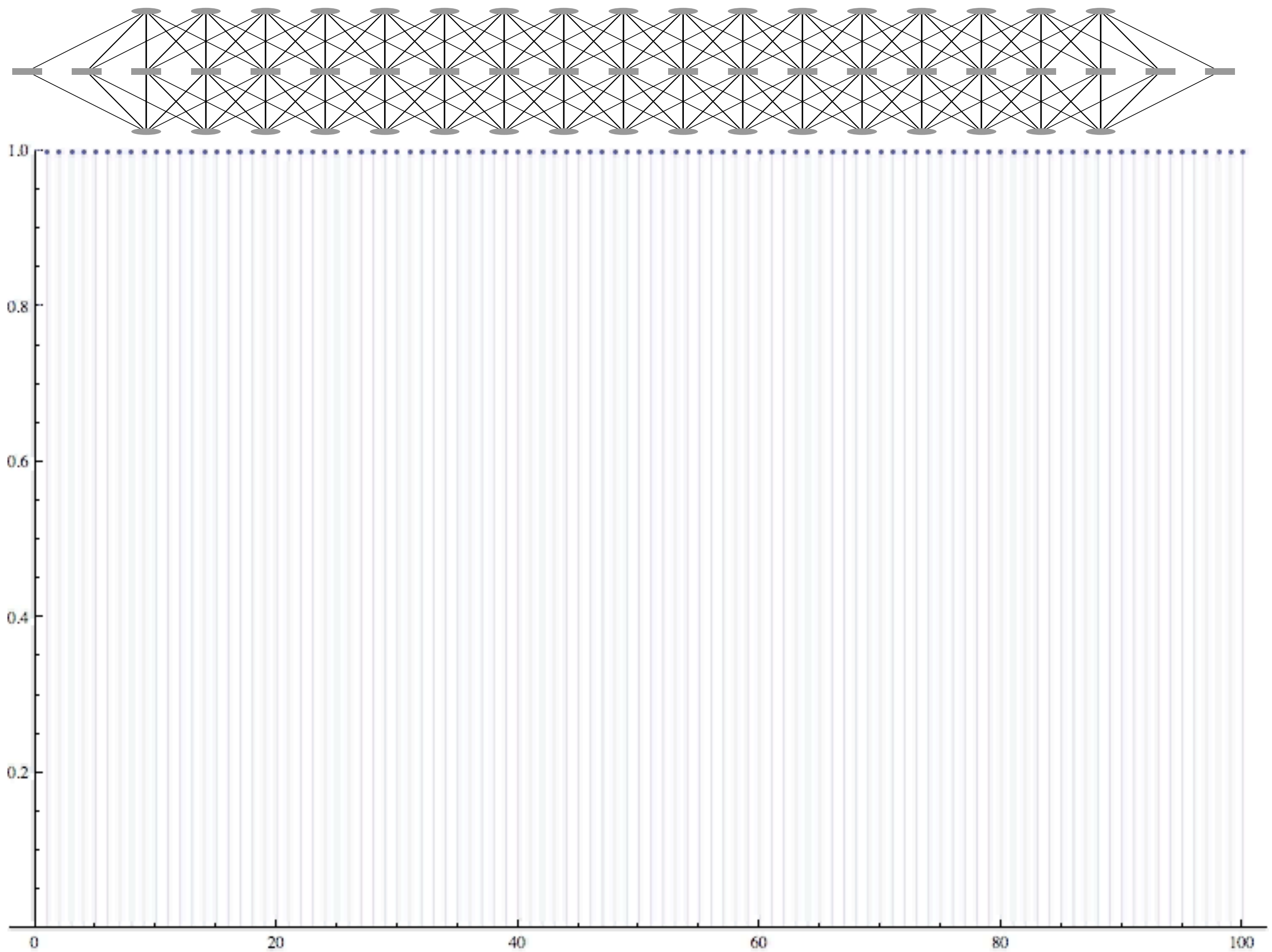


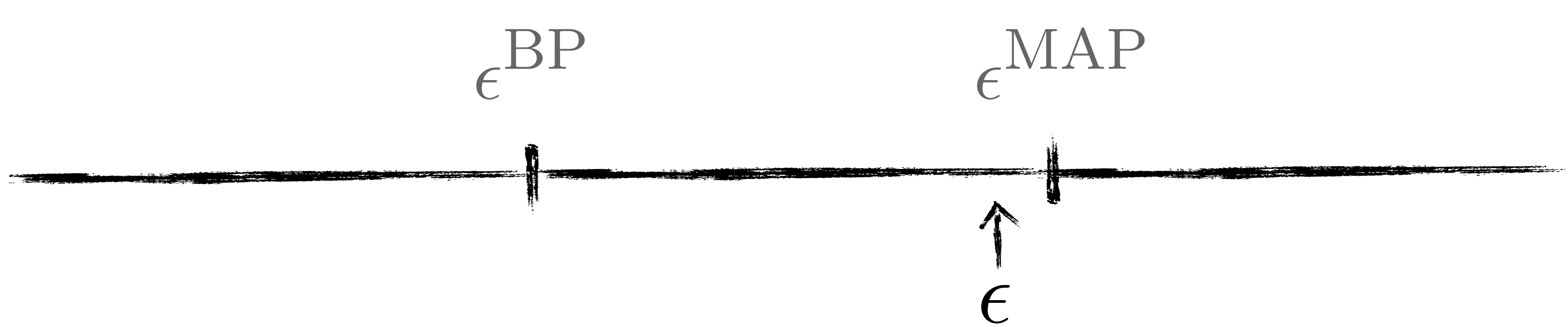
DE for Coupled Ensemble



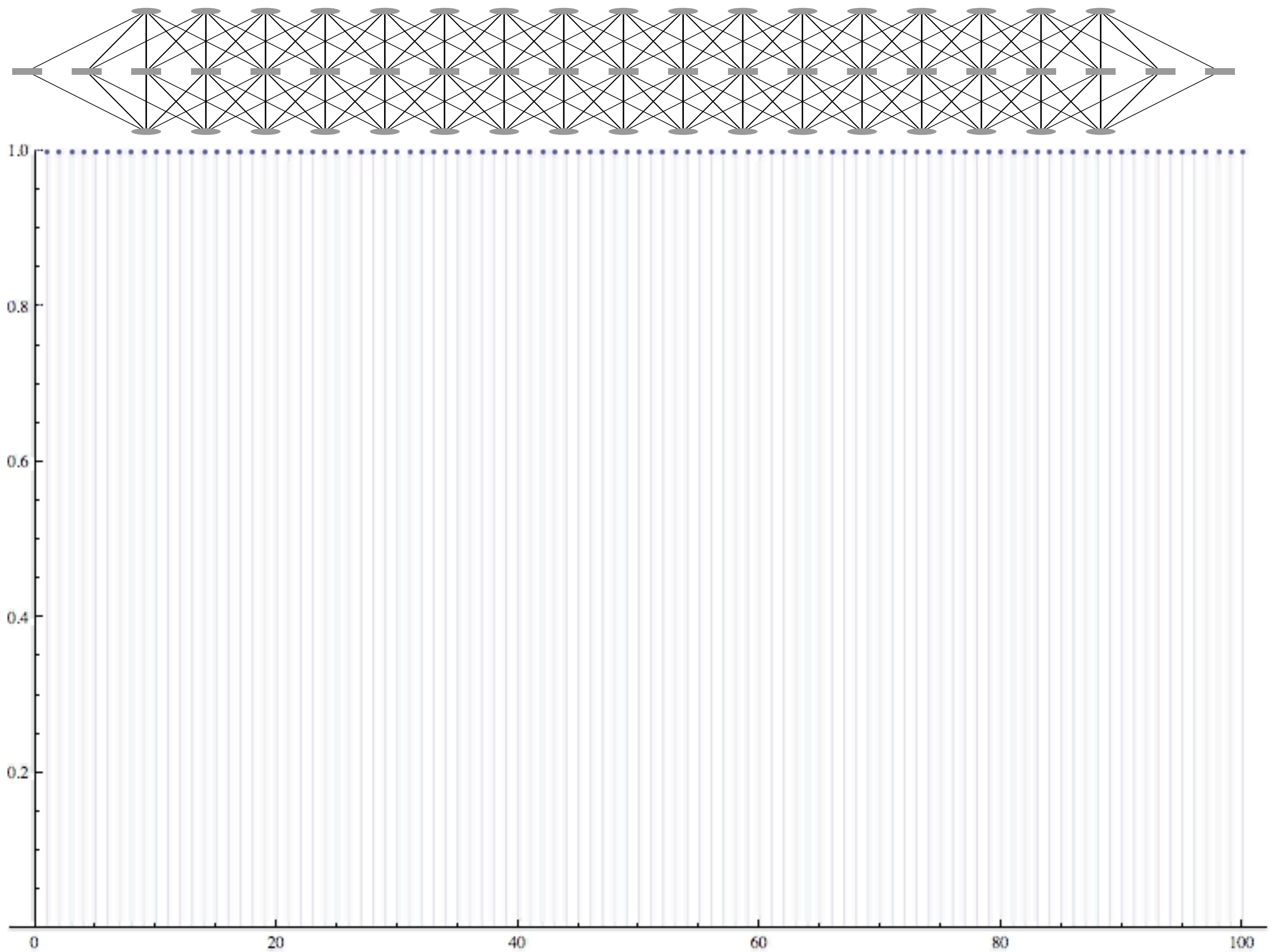


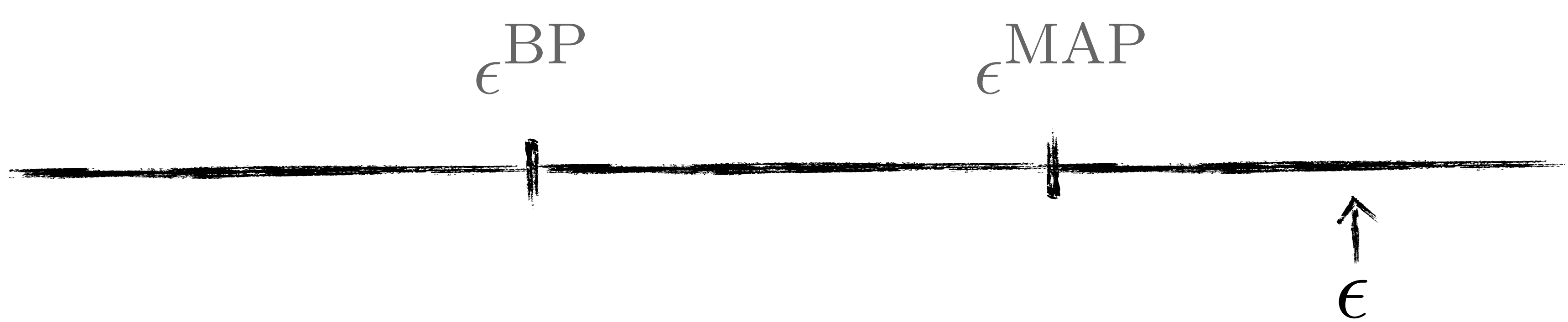
DE for Coupled Ensemble



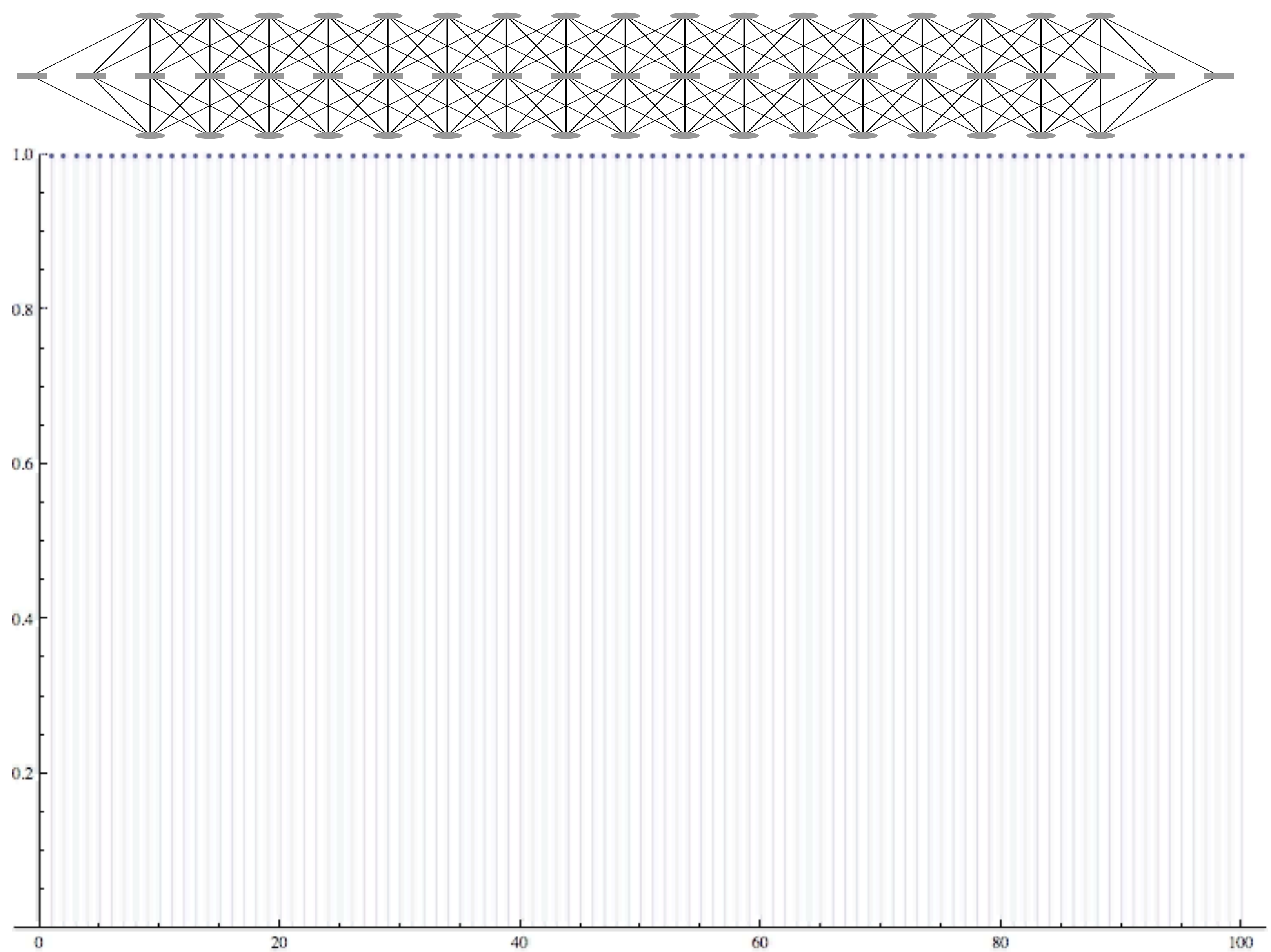


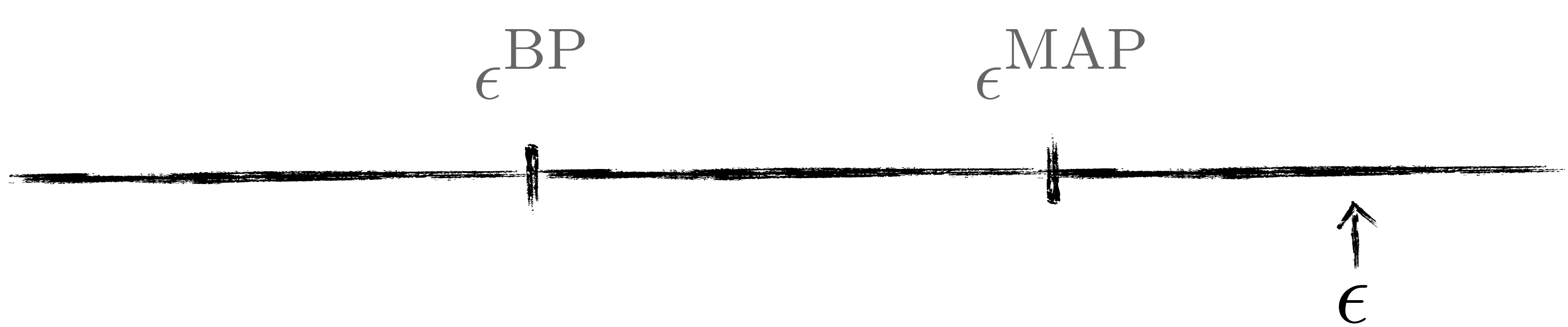
DE for Coupled Ensemble





DE for Coupled Ensemble





Recall

Recall

Coupled ensembles under BP decoding behave like uncoupled ensembles under MAP decoding.

Recall

Coupled ensembles under BP decoding behave
like uncoupled ensembles under MAP decoding.

Since coupled ensemble achieve the highest threshold they can achieve (namely the MAP threshold) under BP we speak of the threshold saturation phenomenon.

Recall

Coupled ensembles under BP decoding behave
like uncoupled ensembles under MAP decoding.

Since coupled ensemble achieve the highest threshold they can achieve (namely the MAP threshold) under BP we speak of the threshold saturation phenomenon.

By using spatial coupling we can construct codes which are capacity-achieving *universally* across the whole set of BMS channels.

Recall

Coupled ensembles under BP decoding behave
like uncoupled ensembles under MAP decoding.

Since coupled ensemble achieve the highest threshold they can achieve (namely the MAP threshold) under BP we speak of the threshold saturation phenomenon.

By using spatial coupling we can construct codes which are capacity-achieving *universally* across the whole set of BMS channels.

The basic principle is applicable to a wide range of graphical models.

Recall

Coupled ensembles under BP decoding behave
like uncoupled ensembles under MAP decoding.

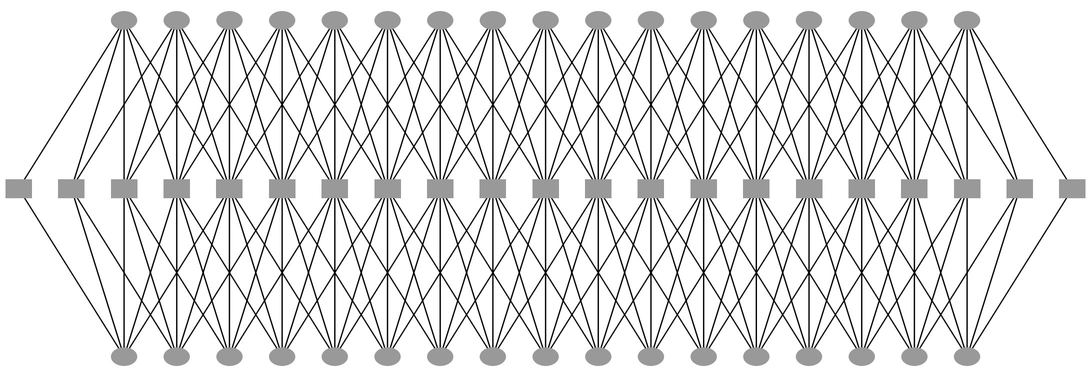
Since coupled ensemble achieve the highest threshold they can achieve (namely the MAP threshold) under BP we speak of the threshold saturation phenomenon.

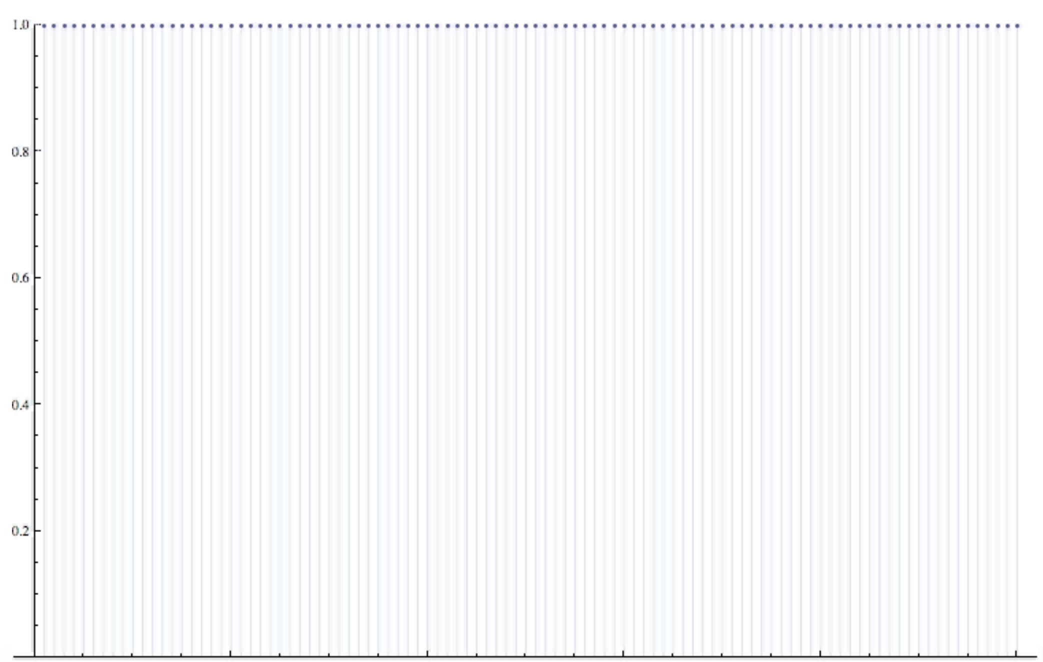
By using spatial coupling we can construct codes which are capacity-achieving *universally* across the whole set of BMS channels.

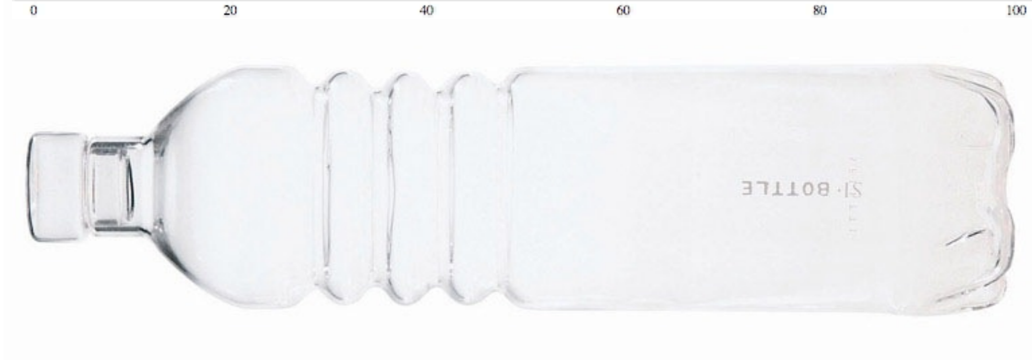
The basic principle is applicable to a wide range of graphical models.

The phenomenon of threshold saturation is closely connected to the way of how crystals grow.

Physics Interpretation



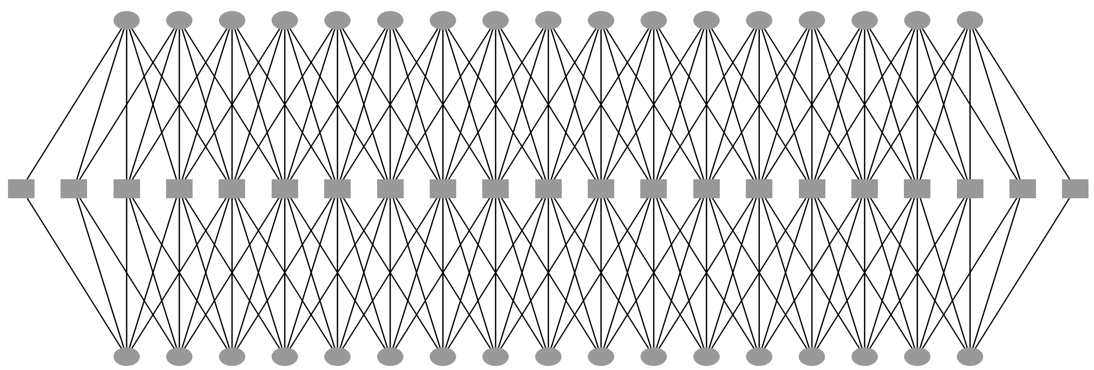


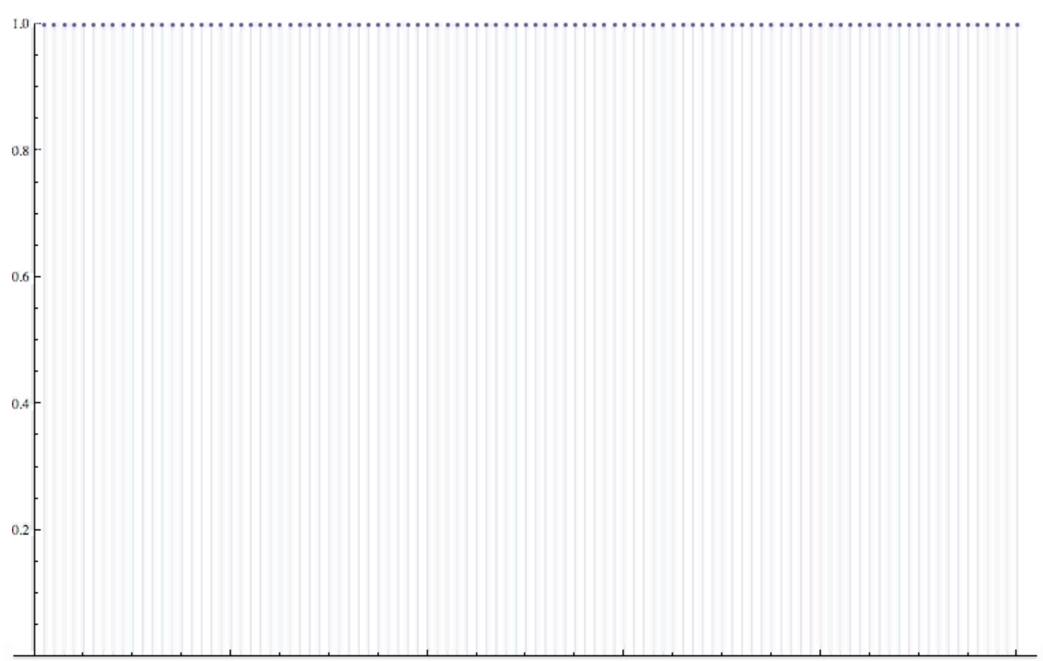


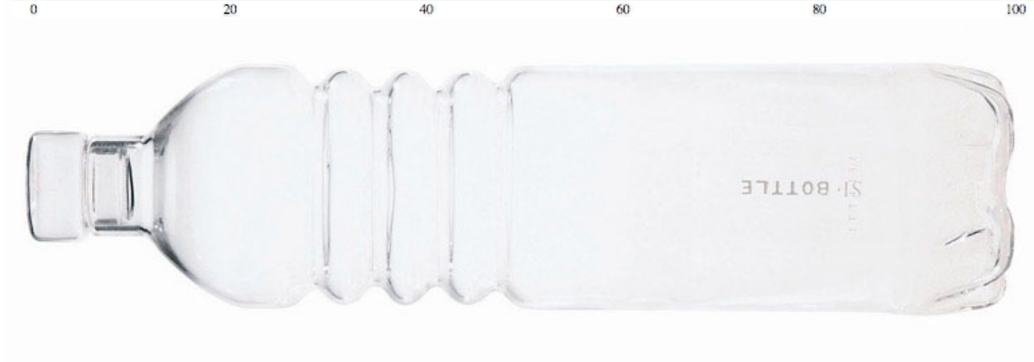
Krzakala, Mezard, Sausset, Sun, and Zdeborova

metastability and nucleation

Physics Interpretation



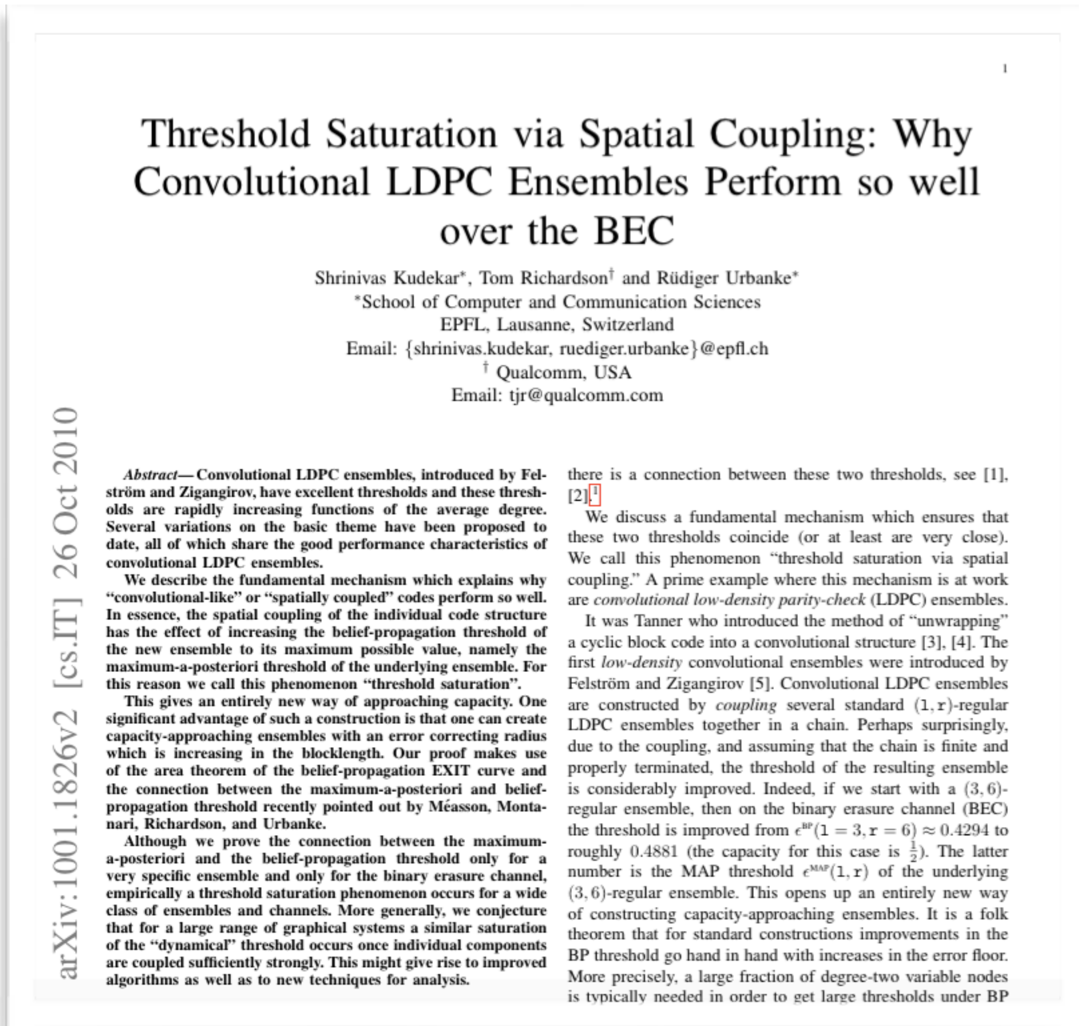




Krzakala, Mezard, Sausset, Sun, and Zdeborova

metastability and nucleation

Spatial Coupling as a Proof Technique (coding)



1

Threshold Saturation via Spatial Coupling: Why Convolutional LDPC Ensembles Perform so well over the BEC

Shrinivas Kudekar*, Tom Richardson† and Rüdiger Urbanke*
*School of Computer and Communication Sciences
EPFL, Lausanne, Switzerland
Email: {shrinivas.kudekar, ruediger.urbanke}@epfl.ch
† Qualcomm, USA
Email: tjr@qualcomm.com

arXiv:1001.1826v2 [cs.IT] 26 Oct 2010

***Abstract*— Convolutional LDPC ensembles, introduced by Felström and Zigangirov, have excellent thresholds and these thresholds are rapidly increasing functions of the average degree. Several variations on the basic theme have been proposed to date, all of which share the good performance characteristics of convolutional LDPC ensembles.**

We describe the fundamental mechanism which explains why "convolutional-like" or "spatially coupled" codes perform so well. In essence, the spatial coupling of the individual code structure has the effect of increasing the belief-propagation threshold of the new ensemble to its maximum possible value, namely the maximum-a-posteriori threshold of the underlying ensemble. For this reason we call this phenomenon "threshold saturation".

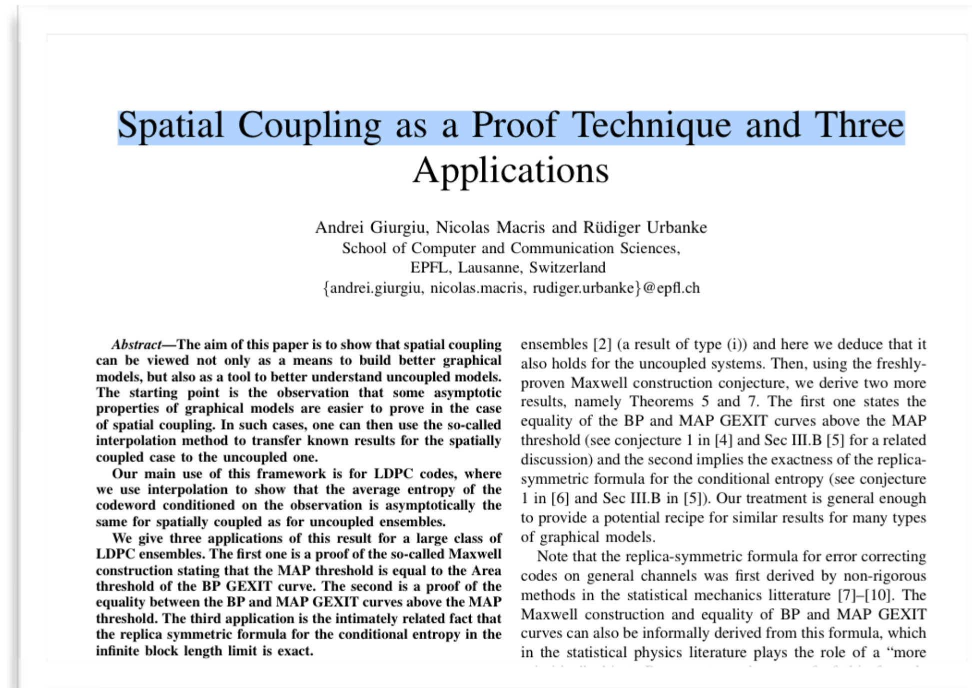
This gives an entirely new way of approaching capacity. One significant advantage of such a construction is that one can create capacity-approaching ensembles with an error correcting radius which is increasing in the blocklength. Our proof makes use of the area theorem of the belief-propagation EXIT curve and the connection between the maximum-a-posteriori and belief-propagation threshold recently pointed out by Méasson, Montanari, Richardson, and Urbanke.

Although we prove the connection between the maximum-a-posteriori and the belief-propagation threshold only for a very specific ensemble and only for the binary erasure channel, empirically a threshold saturation phenomenon occurs for a wide class of ensembles and channels. More generally, we conjecture that for a large range of graphical systems a similar saturation of the "dynamical" threshold occurs once individual components are coupled sufficiently strongly. This might give rise to improved algorithms as well as to new techniques for analysis.

there is a connection between these two thresholds, see [1], [2].[1]

We discuss a fundamental mechanism which ensures that these two thresholds coincide (or at least are very close). We call this phenomenon "threshold saturation via spatial coupling." A prime example where this mechanism is at work are *convolutional low-density parity-check* (LDPC) ensembles.

It was Tanner who introduced the method of "unwrapping" a cyclic block code into a convolutional structure [3], [4]. The first *low-density* convolutional ensembles were introduced by Felström and Zigangirov [5]. Convolutional LDPC ensembles are constructed by *coupling* several standard $(\mathtt{l}, \mathtt{r})$-regular LDPC ensembles together in a chain. Perhaps surprisingly, due to the coupling, and assuming that the chain is finite and properly terminated, the threshold of the resulting ensemble is considerably improved. Indeed, if we start with a $(3, 6)$-regular ensemble, then on the binary erasure channel (BEC) the threshold is improved from $\epsilon^{\text{BP}}(\mathtt{l}=3, \mathtt{r}=6) \approx 0.4294$ to roughly 0.4881 (the capacity for this case is $\frac{1}{2}$). The latter number is the MAP threshold $\epsilon^{\text{MAP}}(\mathtt{l}, \mathtt{r})$ of the underlying $(3, 6)$-regular ensemble. This opens up an entirely new way of constructing capacity-approaching ensembles. It is a folk theorem that for standard constructions improvements in the BP threshold go hand in hand with increases in the error floor. More precisely, a large fraction of degree-two variable nodes is typically needed in order to get large thresholds under BP



Spatial Coupling as a Proof Technique and Three Applications

Andrei Giurgiu, Nicolas Macris and Rüdiger Urbanke
School of Computer and Communication Sciences,
EPFL, Lausanne, Switzerland
{andrei.giurgiu, nicolas.macris, rudiger.urbanke}@epfl.ch

***Abstract*—The aim of this paper is to show that spatial coupling can be viewed not only as a means to build better graphical models, but also as a tool to better understand uncoupled models. The starting point is the observation that some asymptotic properties of graphical models are easier to prove in the case of spatial coupling. In such cases, one can then use the so-called interpolation method to transfer known results for the spatially coupled case to the uncoupled one.**

Our main use of this framework is for LDPC codes, where we use interpolation to show that the average entropy of the codeword conditioned on the observation is asymptotically the same for spatially coupled as for uncoupled ensembles.

We give three applications of this result for a large class of LDPC ensembles. The first one is a proof of the so-called Maxwell construction stating that the MAP threshold is equal to the Area threshold of the BP GEXIT curve. The second is a proof of the equality between the BP and MAP GEXIT curves above the MAP threshold. The third application is the intimately related fact that the replica symmetric formula for the conditional entropy in the infinite block length limit is exact.

ensembles [2] (a result of type (i)) and here we deduce that it also holds for the uncoupled systems. Then, using the freshly-proven Maxwell construction conjecture, we derive two more results, namely Theorems 5 and 7. The first one states the equality of the BP and MAP GEXIT curves above the MAP threshold (see conjecture 1 in [4] and Sec III.B [5] for a related discussion) and the second implies the exactness of the replica-symmetric formula for the conditional entropy (see conjecture 1 in [6] and Sec III.B in [5]). Our treatment is general enough to provide a potential recipe for similar results for many types of graphical models.

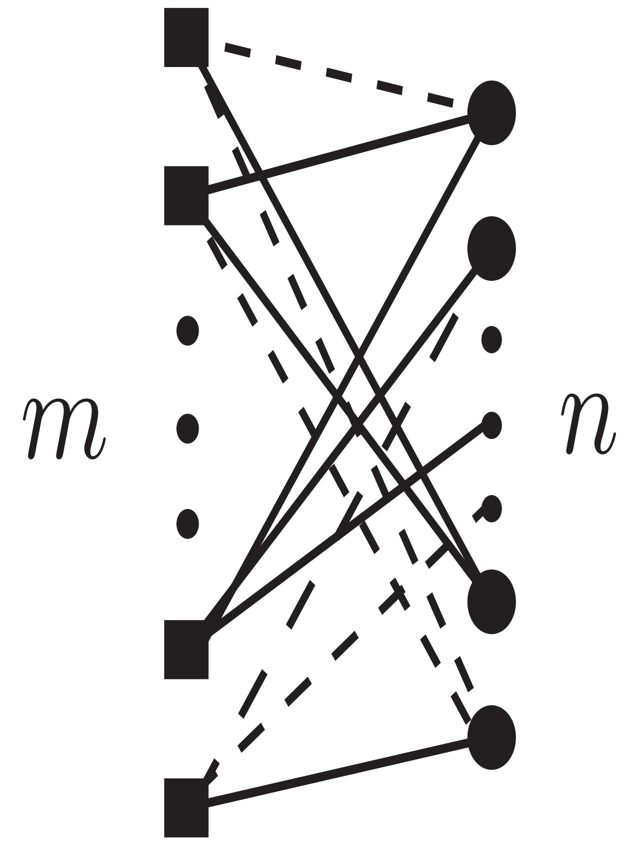
Note that the replica-symmetric formula for error correcting codes on general channels was first derived by non-rigorous methods in the statistical mechanics litterature [7]–[10]. The Maxwell construction and equality of BP and MAP GEXIT curves can also be informally derived from this formula, which in the statistical physics literature plays the role of a "more

shows that MAP threshold is given by Maxwell conjecture

Paradigmatic CSP: random K-SAT

- Random graph with n variable nodes and m clauses.
- Each clause is connected to K variables u.a.r.
- Edge is dashed or full with probability $1/2$. Degree of variable nodes is Poisson(αK).

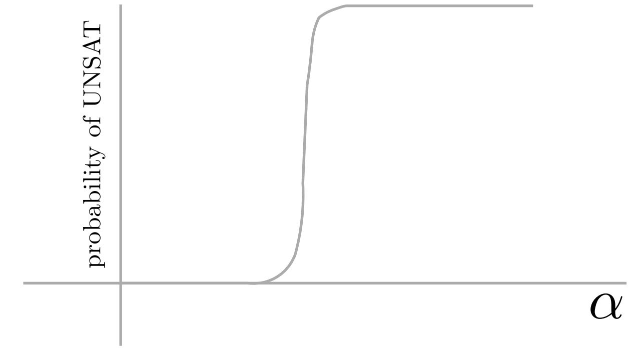
- Boolean variables: $x_i \in \{T, F\}$ or $\in \{0, 1\}$, $i = 1, \cdots, n$
- Clauses: $(\vee_{i=1}^{K} x_{a_i}^{n(a_i)})$, $a = 1, \cdots, m$
- $F_{n,\alpha,K} = \wedge_{a=1}^{M} (\vee_{i=1}^{K} x_{a_i}^{s(a_i)})$



Control parameter $\alpha = \frac{\#(\text{clauses})}{\#(\text{variables})} = \frac{m}{n}$**: Phase Transitions.**

- Friedgut 1999: $\exists \alpha_s(n,K)$ s.t $\forall \epsilon > 0$

$$\lim_{n\to\infty} \Pr\{F_{n,\alpha,K} \text{is SAT}\} = \begin{cases} 1 & \text{if } \alpha < (1-\epsilon)\alpha_s(n,K), \\ 0 & \text{if } \alpha > (1-\epsilon)\alpha_s(n,K). \end{cases}$$



Existence of $\lim_{n\to+\infty} \alpha_s(n,K)$ is still an open problem.

- This talk: MAX-SAT or Hamiltonian version of the problem:

$$H_F(\underline{x}) = \sum_{a=1}^{m} \left(1 - \mathbf{1}\left(\vee_{i=1}^{K} x_{a_i}^{s(a_i)}\right)\right),$$

the MAX-SAT/UNSAT threshold is defined as:

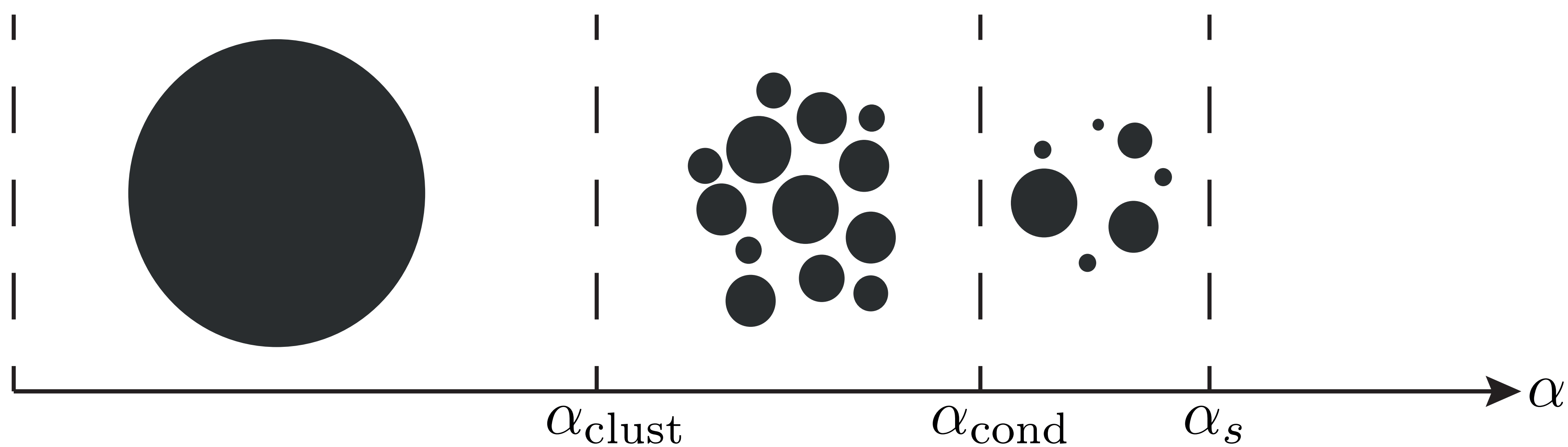
$$\alpha_s(K) \equiv \inf\{\ \alpha \ \mid \ \underbrace{\lim_{n\to+\infty} \frac{1}{n}\mathbb{E}[\min_{\underline{x}} H_F(\underline{x})]}_{\text{exists and continuous function of } \alpha} > 0\ \}$$

In particular α_s exists. [Interpolation methods: Franz-Leone, Panchenko, Gamarnik-Bayati-Tetali].

The Physics Picture

Parisi-Mezard-Zechina 2001

Semerjian-RicciTersenghi-Montanari, Krazkala-Zdeborova 2008



Known Lower bounds on the SAT-UNSAT threshold

- Algorithmic lower bounds: find analyzable algorithm and find solutions for $\alpha_{\mathrm{alg}}(K) < \alpha_s(K)$. [long history ...]
- Second Moment lower bounds, weighted s.m with cavity inspired weights [long history, ... Achlioptas - Coja Oghlan].

K	3	4	$\cdots$	large K
best lower bound	3.52^{alg}	$7.91^{\mathrm{s.m}}$	$\cdots$	$2^K \ln 2 - \frac{3}{2}\ln 2 + o(1)^{\mathrm{s.m}}$
best algor bound	3.52	5.54	$\cdots$	$\frac{2^K \ln K}{K}(1+o(1))$
α_{dyn}	3.86	9.38	$\cdots$	$\frac{2^K \ln K}{K}(1+o(1))$
α_{cond}	3.86	9.55	$\cdots$	$2^K \ln 2 - \frac{3}{2}\ln 2 + o(1)$
α_{s}	4.26	9.93	$\cdots$	$2^K \ln 2 - \frac{1}{2}(1+\ln 2) + o(1)$

New Lower bounds by the Spatial Coupling Method

Recall:

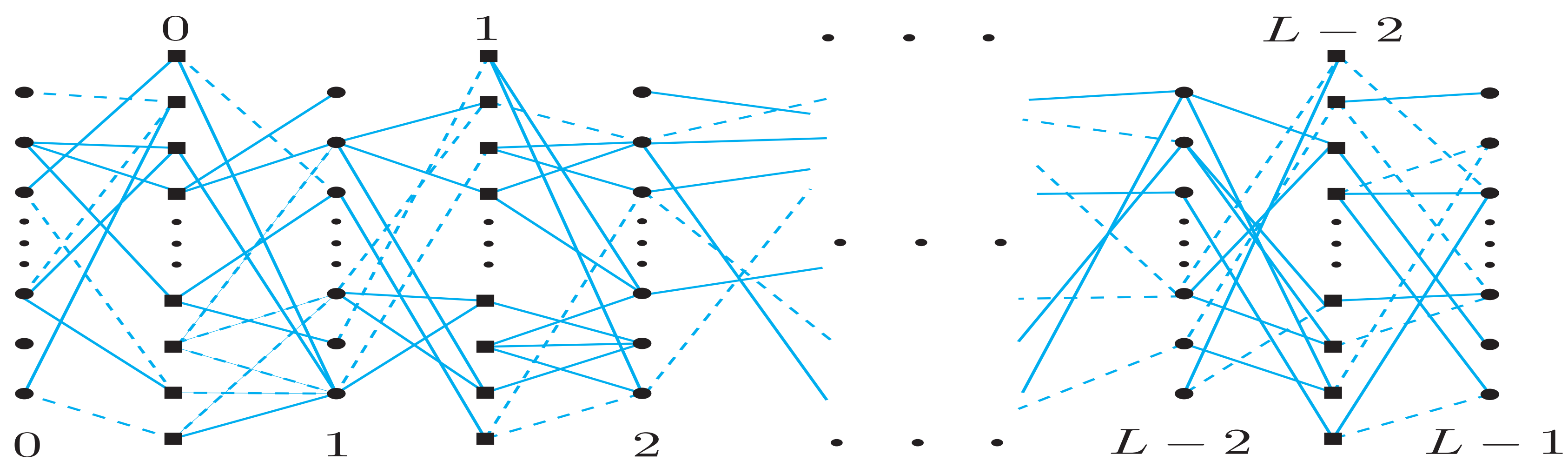
$H_F(\underline{x})$ = number of UNSAT clauses of F for $\underline{x} \in \{0,1\}^n$

and $\alpha_s = \inf\{\alpha \mid \lim_{n\to+\infty} \frac{1}{n}\mathbb{E}[\min_{\underline{x}} H_F(\underline{x})] > 0\}$

K	3	4	$\cdots$	large K
α_{new}	3.67	7.81	$\cdots$	$2^K \times \frac{1}{2}$
best algor bound	3.52	5.54	$\cdots$	$\frac{2^K \ln K}{K}(1+o(1))$
best lower bound	3.52^{alg}	$7.91^{\text{s.m}}$	$\cdots$	$2^K \ln 2 - \frac{3}{2}\ln 2 + o(1)^{\text{s.m}}$
α_{dyn}	3.86	9.38	$\cdots$	$\frac{2^K \ln K}{K}(1+o(1)$
α_{cond}	3.86	9.55	$\cdots$	$2^K \ln 2 - \frac{3}{2}\ln 2 + o(1)$
α_{s}	4.26	9.93	$\cdots$	$2^K \ln 2 - \frac{1}{2}(1+\ln 2) + o(1)$

Strategy

construct spatially coupled model



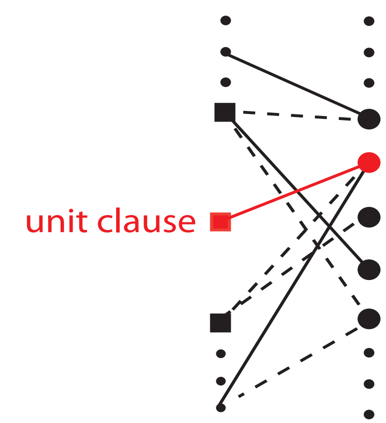
$$\alpha_{SAT}^{\text{coupled}} = \alpha_{SAT}^{\text{uncoupled}}$$

$$\alpha_{alg}^{\text{uncoupled}} \leq \alpha_{alg}^{\text{coupled}} \leq \alpha_{SAT}^{\text{(un)coupled}}$$

Unit Clause Propagation algorithm

1. Repeat until all variables are set:

2. Forced Step: If F contains unit clauses choose one at random and satisfy it by setting unique variable. Remove or shorten other clauses that contain this variable.



3. Free Step: If there are no unit clauses choose a variable at random and set it at random. Remove or shorten clauses that contain this variable.

Analysis (Wormald Method)

$$E[\text{ change} \mid \text{state}] = f(\text{state})$$

convert this to differential equation

show that "typical" instances closely follow solution of diff equation

state for our case: C_3 – number of clauses of degree 3

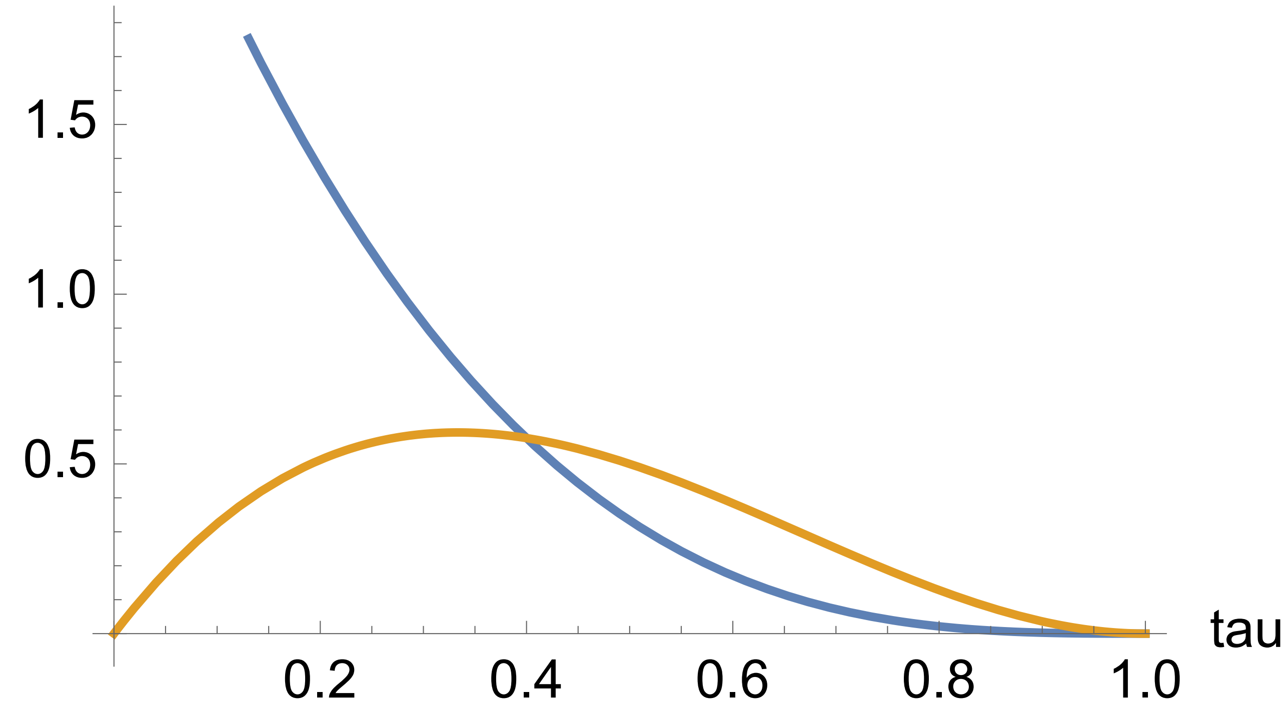
C_2 – number of clauses of degree 2

C_1 – number of clauses of degree 1

t – time; number of peeled-off variables

Analysis (Wormald Method)

c3 (decreasing curve), c2 (unimodal)



$$\frac{dc_3(\tau)}{d\tau} = -\frac{3c_3(\tau)}{1-\tau},$$
$$\frac{dc_2(\tau)}{d\tau} = \frac{3c_3(\tau)}{2(1-\tau)} - \frac{2c_2(\tau)}{1-\tau}$$

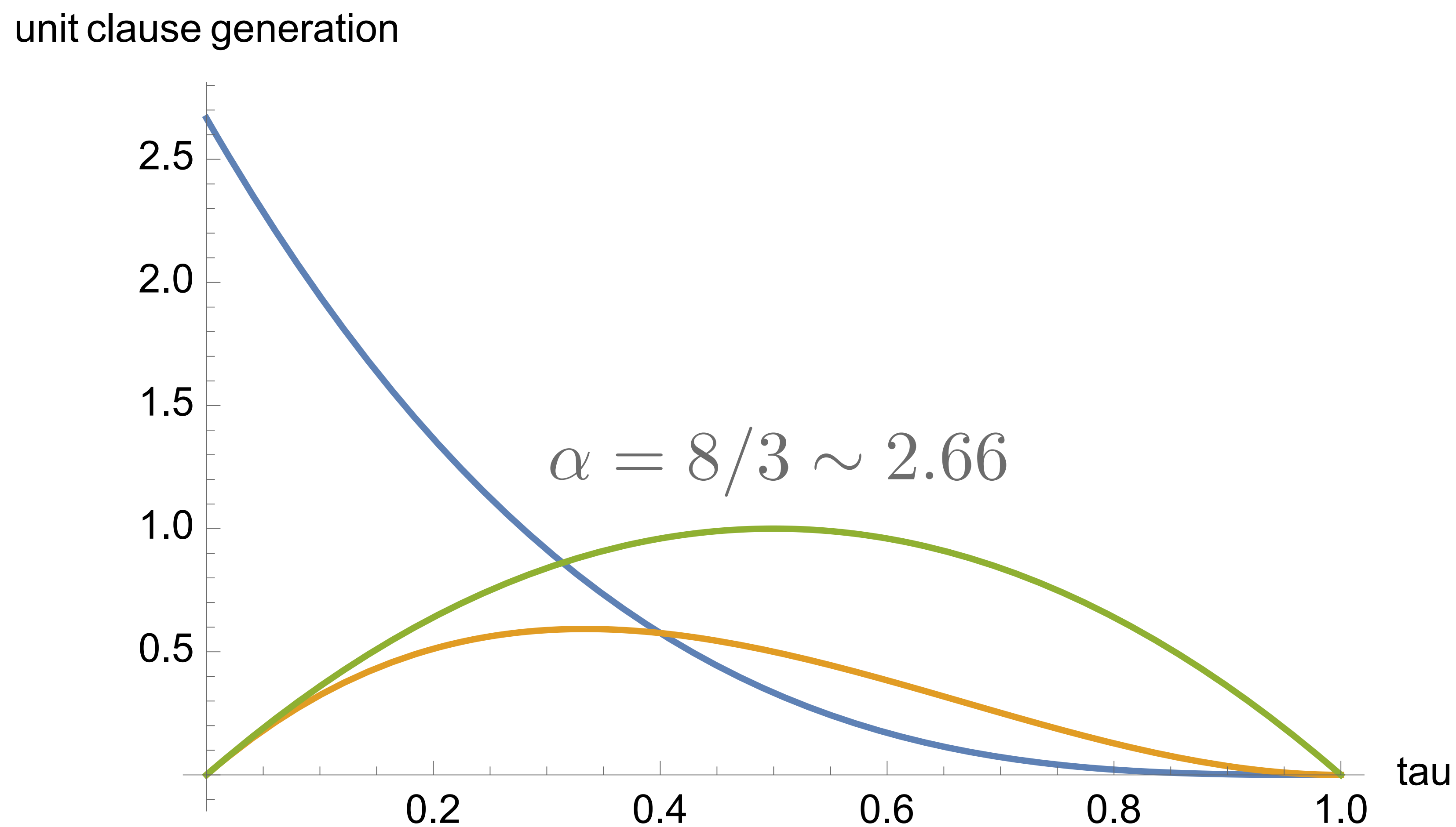
$$c_3(\tau=0) = \alpha, c_2(\tau=0) = 0$$

$$c_3(\tau) = \alpha(1-\tau)^3, c_2(\tau) = \frac{3\alpha}{2}\tau(1-\tau)^2$$

Analysis (Wormald Method)

unit clause generation rate

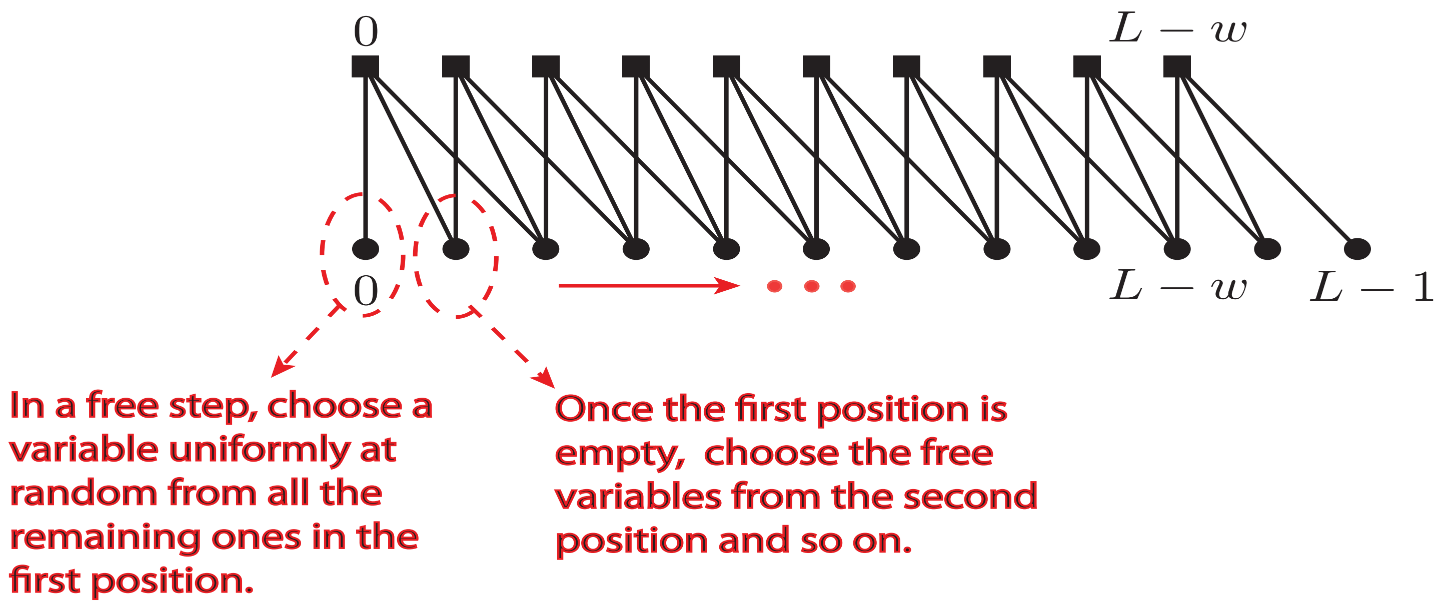
$$\frac{1}{2}\frac{2c_2(\tau)}{2c_2(\tau)+3c_3(\tau)}\frac{2c_2(\tau)+3c_3(\tau)}{1-\tau} = \frac{c_2(\tau)}{1-\tau} < 1$$



Unit Clause Propagation for coupled Formulas:

- Forced step: as long as $\exists$ unit clause, then satisfy it by setting the variable. Remove or shorten clauses containing this variable.

- Free step:



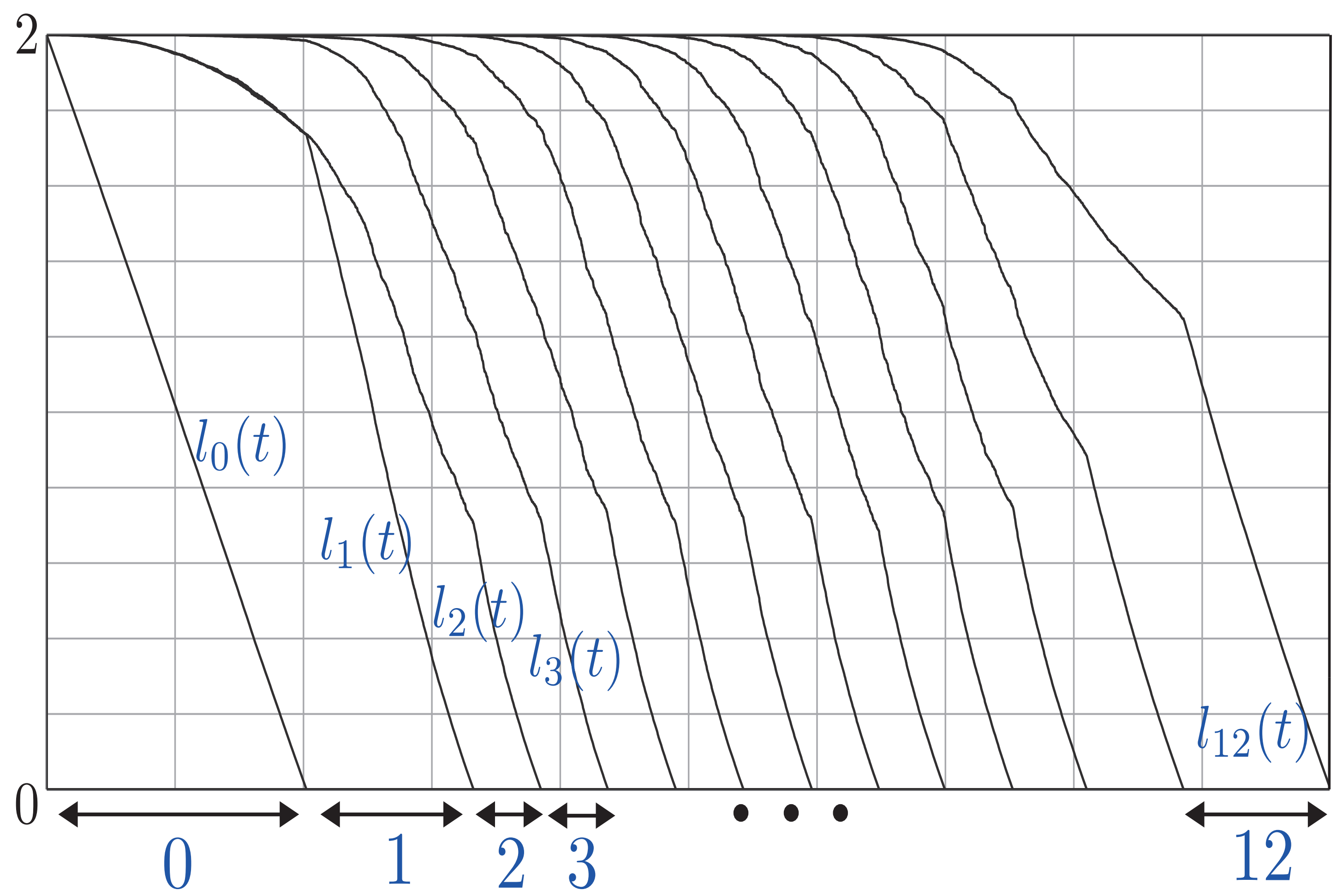
Differential Equations for Coupled-UC

Phase p ($i \geq p$). Round $\equiv$ free step followed by forced steps.

$$\frac{d\ell_i(t)}{dt} \equiv -2\beta_i(t) = -2 \text{ rate of removal of nodes at pos } i$$

$$\begin{cases} \frac{dc_i^{(3)}(t,\vec{\tau})}{dt} = -2\sum_{d=0}^{w-1} \beta_{i+d}(t)\frac{\tau_d c_i^{(3)}(t,\vec{\tau})}{\ell_{i+d}(t)} \\ \frac{dc_i^{(2)}(t,\vec{\tau})}{dt} = -2\sum_{d=0}^{w-1} \beta_{i+d}(t)\frac{\tau_d c_i^{(2)}(t,\vec{\tau})}{\ell_{i+d}(t)} + \sum_{d=0}^{w-1}(1+\tau_d)\beta_{i+d}(t)\frac{c_i^{(3)}(t,\vec{\tau}^d)}{\ell_{i+d}(t)} \end{cases}$$

Evolution of number of variables per position



Algorithm runs in "phases" $p = 0, 1, 2, 3, \dots$ which terminate each time all variables have been set in a position p.

At $\alpha \approx 3.67$ the curves develop vertical slopes: explosion of unit clauses.

Proposition: Let $\alpha_{\mathrm{UC}}^{\mathrm{coupled}}(K) \equiv \lim_{w\to\infty} \lim_{L\to\infty} \alpha_{\mathrm{UC}}^{\mathrm{coupled}}(K, L, w)$

K	3	4	...	large K
$\alpha_{\mathrm{UC}(K)}$	2.67	4.50	...	$\frac{e}{K}2^{K-1}$
$\alpha_{\mathrm{UC}}^{\mathrm{coupled}}(K)$	3.67	7.81	...	$2^{K-1} + \cdots$

Exact formula:

$$\alpha_{\mathrm{UC}}^{\mathrm{coupled}}(K) = \max\{\alpha \geq 0 | \min_{\ell\in[0,2]} \Phi_{\alpha,K}(\ell)\}$$

with

$$\Phi_{\alpha,K}(\ell) = 2 - \ell(1 - \frac{\ln \ell}{2}) - \frac{\alpha}{2^{K-2}}(1 - \frac{\ell}{2})^K$$

Conclusion

- Lower bounds for CSP's by algorithmic lower bounds on coupled-CSP's.
- Applies to many problems: K-SAT, COL, XORSAT, Error Correcting LDPC codes, Rate-Distortion theory.
- For XORSAT and Error Correcting codes it gives optimal lower bounds $\alpha_{\rm alg} < \alpha_{\rm coupled-alg} = \alpha_s$.
- For SAT, COL, can we perform better with more sophisticated local rule instead of free step ?
- Above some K we find that $\alpha_{\rm UC}^{\rm coupled} > \alpha_{\rm dyn}^{\rm uncoupled}$.
- Sometimes we go above condensation threshold. E.g coloring with $Q \geq 4$.

Summary

Spatial coupling can be used in two different ways.

Algorithmic: spatially coupled graphs are particularly suited for message passing

Proof technique: extend problem to spatially coupled version
proof desired property for this version
show that original problem is equivalent to spatially coupled
with respect to this property;