

# Survey-propagation decimation through distributed local computations

Joël Chavas<sup>1</sup>, Cyril Furtlehner<sup>2</sup>, Marc Mézard<sup>2</sup> and Riccardo Zecchina<sup>3</sup>

<sup>1</sup> ISI, Viale Settimio Severo 65, 10133 Torino, Italy

<sup>2</sup> LPTMS, Bâtiment 100, Université Paris-Sud, Centre Scientifique d'Orsay, 15 rue Georges Clémenceau, 91405 Orsay Cedex, France

<sup>3</sup> ICTP, Strada Costiera 11, I-34100 Trieste, Italy

E-mail: [chavas@isiosf.isi.it](mailto:chavas@isiosf.isi.it), [furtlehn@lptms.u-psud.fr](mailto:furtlehn@lptms.u-psud.fr), [mezard@lptms.u-psud.fr](mailto:mezard@lptms.u-psud.fr) and [zecchina@ictp.trieste.it](mailto:zecchina@ictp.trieste.it)

Received 6 June 2005

Accepted 6 October 2005

Published 30 November 2005

Online at [stacks.iop.org/JSTAT/2005/P11016](http://stacks.iop.org/JSTAT/2005/P11016)

[doi:10.1088/1742-5468/2005/11/P11016](https://doi.org/10.1088/1742-5468/2005/11/P11016)

**Abstract.** We discuss the implementation of two *distributed* solvers of the random K-SAT problem, based on some development of the recently introduced survey propagation (SP) algorithm. The first solver, called the 'SP diffusion algorithm', diffuses as dynamical information the maximum bias over the system, so that variable nodes can decide to freeze in a self-organized way, each variable making its decision on the basis of purely local information. The second solver, called the 'SP reinforcement algorithm', makes use of time-dependent external forcing messages on each variable, which are adapted in time in such a way that the algorithm approaches its estimated closest solution. Both methods allow us to find a solution of the random 3-SAT problem in a range of parameters comparable with the best previously described serialized solvers. The simulated time of convergence towards a solution (if these solvers were implemented on a fully parallel device) grows as  $\log(N)$ .

**Keywords:** heuristics, message-passing algorithms

**Contents**

<b>1. Introduction</b>	<b>2</b>
<b>2. The SAT problem and its factor graph representation</b>	<b>4</b>
<b>3. Survey propagation</b>	<b>6</b>
3.1. Warnings . . . . .	6
3.2. The algorithm . . . . .	7
<b>4. Distributed SP I: diffusion algorithm</b>	<b>8</b>
4.1. Simulating SP in parallel . . . . .	8
4.2. Propagation of information . . . . .	10
4.3. Experimental results . . . . .	12
4.3.1. Orders of magnitude for the tuning parameters. . . . .	13
4.3.2. Dependence on the size of the system. . . . .	14
4.3.3. Dependence on $\alpha$ . . . . .	14
<b>5. Distributed SP II: ‘reinforcement algorithm’</b>	<b>15</b>
5.1. SP with external forcing field . . . . .	15
5.2. Presentation of the method . . . . .	17
5.3. Results . . . . .	19
5.3.1. Determination of the optimal intensity of the forcing $\pi$ . . . . .	19
5.3.2. Range of validity of the method. . . . .	20
5.3.3. Which solutions are addressed? . . . . .	21
5.3.4. Logarithmic dependence of time on the size of the graph. . . . .	22
5.3.5. Performance of the reinforcement algorithm on a serialized implementation. . . . .	22
<b>6. Conclusion</b>	<b>23</b>
<b>Acknowledgments</b>	<b>24</b>
<b>References</b>	<b>24</b>

**1. Introduction**

Recent progress in the statistical physics of disordered systems has provided results of algorithmic interest: the connection between the geometrical structure of the space of solutions in random NP-complete constraint satisfaction problems (CSPs) and the onset of exponential regimes in search algorithms [3, 16] has been at least partially clarified. More important, the analytical methods have been converted into a new class of algorithms [17] with very good performance [11].

Similarly to statistical physics models, a generic CSP is composed of many discrete variables which interact through constraints, where each constraint involves a small number of variables. When CSPs are extracted at random from non-trivial ensembles there appear phase transitions as the constraint density (the ratio of constraints to

variables) increases: for small densities the problems are satisfiable with high probability, whereas when the density is larger than a critical value the problems become unsatisfiable with high probability. Close to the phase boundary on the satisfiable side, most of the algorithms are known to take (typically) exponential time to find solutions. The physical interpretation for the onset of such exponential regimes in random hard CSP consists in a trapping of local search processes in metastable states. Depending on the models and on the details of the process the long time behaviour may be dominated by different types of metastable states. However, a common feature which can be observed numerically is that for simulation times which are sub-exponential in the size of the problem there exists an extensive gap in the number of violated constraints which separates the blocking configurations from the optimal ones. Such behaviour can be tested on concrete random instances which therefore constitute a computational benchmark for more general algorithms.

In the last few years there has been a great progress in the study of phase transitions in random CSP which has produced new algorithmic tools: problems which were considered to be algorithmically hard for local search algorithms, for instance random K-SAT close to a phase boundary, turned out to be efficiently solved by the so called survey propagation (SP) algorithm [17] arising from the replica symmetry broken (RSB) cavity approach to CSP. According to the statistical physics analysis, close to the phase transition point the solution space breaks up into many smaller clusters [16, 17]. Solutions in separate clusters are generally far apart. This picture has been confirmed by rigorous results, first on the simple case of the XORSAT problem [2, 5], and more recently for the satisfiability problem [4, 6]. Moreover, the physics analysis indicates that clusters which correspond to partial solutions—which satisfy some but not all of the constraints—are exponentially more numerous than the clusters of complete solutions and act as dynamical traps for local search algorithms. SP turns out to be able to deal efficiently with the proliferation of such clusters of metastable states.

The SP algorithm consists in a message-passing technique which is tightly related to the better known belief propagation algorithm (BP) [21]—recently applied with striking success in the decoding of error-correcting codes based on sparse graph encodings [7, 8], [12]–[15]—but which differs from it in some crucial point. The messages sent along the edges of the graph underlying the combinatorial problem describe in a probabilistic way the cluster-to-cluster fluctuations of the optimal assignment for a variable; while BP performs a ‘white’ average over all the solutions, SP makes averages assuming a decomposition of the global probability measure into intra-cluster measures for each variable. The order parameter computed by SP is a full list of probability distributions telling us what is the probability of picking up a cluster at random and finding a given variable forced to take a specific value within it (‘frozen’ variable). While BP neglects correlations, SP takes care of those correlations that are compatible with the decomposition. Once the iterative equations have reached a fixed-point of such probability distributions (called ‘surveys’ because they capture somehow the distribution of the expectation in the different clusters), it becomes possible in general to identify the variables which can be safely fixed and to simplify the problem accordingly [11, 17]. This procedure, which is intrinsically serial, is known as the SP-inspired decimation algorithm (SID).

From the experimental point of view, the SID has been efficiently used to solve many instances in the hard region of different satisfiability problems and of the graph

colouring problem, including instances too large for any earlier method [17]. For example, for random 3-SAT, instances close to the threshold, up to sizes of order  $10^7$  variables were solved and the computational time in this regime was found experimentally to scale roughly as  $N(\log N)^2$ .

In the present manuscript we show how the SP decimation algorithm can be made fully distributed. This opens the possibility of a parallel implementation which would lead to a further drastic reduction in the computational cost.

The paper is organized as follows: in section 2 we introduce some notation and the K-SAT problem, in section 3 we recall the SP algorithm along with the serial decimation procedure, and sections 4 and 5 are devoted to explaining our new parallel solvers, respectively the first one based on information diffusion and the second one based on external forcing on the variables.

## 2. The SAT problem and its factor graph representation

Though the results we shall discuss are expected to hold for many types of CSP, here we consider just one representative case: the K-SAT problem.

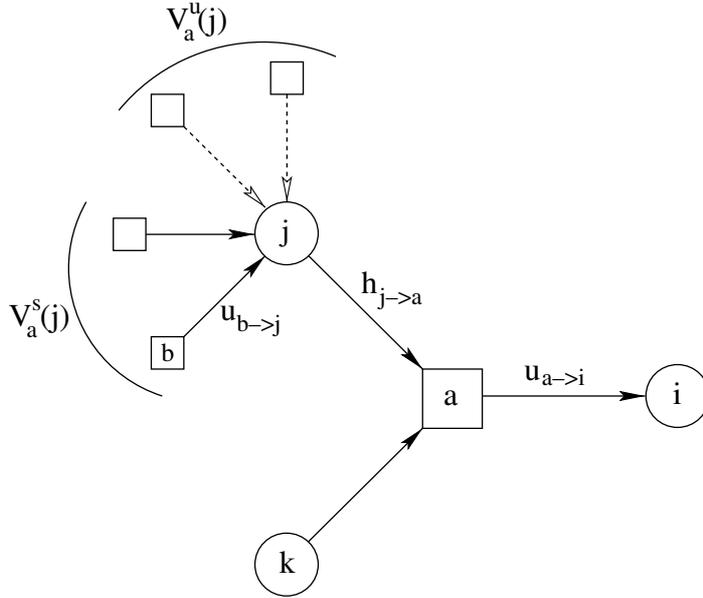
A K-SAT formula consists of  $N$  Boolean variables  $x_i \in \{0, 1\}$ ,  $i \in \{1, \dots, N\}$ , with  $M$  constraints, in which each constraint is a clause, which is the logical OR ( $\vee$ ) of the variables it contains or of their negations.

Solving the K-SAT formula means finding an assignment of the  $x_i \in \{0, 1\} = \{\text{directed, negated}\}$  which is such that all the  $M$  clauses are true. In the literature, such an assignment can indifferently be called the solution, satisfying assignment or ground state if one uses the statistical physics jargon.

In the case of randomly generated K-SAT formulae (variables appearing in the clauses chosen uniformly at random without repetitions and negated with probability  $1/2$ ), a sophisticated phase transition phenomenon sets in: when the number of constraints  $M = \alpha N$  is small, the solutions of the K-SAT formulae are distributed close one to each other over the whole  $N$ -dimensional space, and the problem can easily be solved by the use of classical local search algorithms. Conversely, when  $\alpha$  is included in a narrow region  $\alpha_d < \alpha < \alpha_c$ , the problem is still satisfiable but the now limited solution phase breaks down into an exponential number of clustered components. Solutions become grouped together into clusters which are far apart one from the other. This range of parameters is known as the hard-SAT phase: classical local search algorithms get trapped by strictly positive energy configurations which are exponentially more numerous than the ground state ones.

A clause  $a$  is characterized by the set of variables  $i_1, \dots, i_K$  which it contains, and the list of those which are negated, which can be characterized by a set of  $K$  numbers  $J_{i_r}^a \in \{\pm 1\}$  as follows. The clause is written as

$$(z_{i_1} \vee \dots \vee z_{i_r} \vee \dots \vee z_{i_K}) \quad (1)$$



**Figure 1.** A function node  $a$  and its neighbourhood. The survey of the cavity-bias  $u_{a \rightarrow i}$  can be computed from the knowledge of the joint probability distribution for all the cavity-bias in the set  $U$ , i.e. those coming onto all variable node  $j$  neighbours of  $a$  (except  $j = i$ ).

where  $z_{i_r} = x_{i_r}$  if  $J_{i_r}^a = -1$  and  $z_{i_r} = \bar{x}_{i_r}$  if  $J_{i_r}^a = 1$  (note that a positive literal is represented by a negative  $J$ ). The problem is to find whether there exists an assignment of the  $x_i \in \{0, 1\}$  which is such that all the  $M$  clauses are true. We define the total cost  $C$  of a configuration  $\mathbf{x} = (x_1, \dots, x_N)$  as the number of violated clauses.

In the so called ‘factor graph’ representation, the SAT problem can be depicted as follows (see figure 1). Each of the  $N$  variables is associated to a vertex in the graph, called a ‘variable node’ (circles in the graphical representation), and each of the  $M$  clauses is associated to another type of vertex in the graph, called a ‘function node’ (squares in the graphical representation). A function node  $a$  is connected to a variable node  $i$  by an edge whenever the variable  $x_i$  (or its negation) appears in the clause  $a$ . In the graphical representation, we use a full line between  $a$  and  $i$  whenever the variable appearing in the clause is  $x_i$  (i.e.  $J_i^a = -1$ ) and a dashed line whenever the variable appearing in the clause is  $\bar{x}_i$  (i.e.  $J_i^a = 1$ ).

Throughout this paper, the variable node indices are taken in  $i, j, k, \dots$ , while the function node indices are taken in  $a, b, c, \dots$ . For every variable node  $i$ , we denote by  $V(i)$  the set of function nodes  $a$  to which it is connected by an edge, by  $n_i = |V(i)|$  the degree of the node, by  $V_+(i)$  the subset of  $V(i)$  consisting of function nodes  $a$  where the variable appears un-negated—the edge  $(a, i)$  is a full line, and by  $V_-(i)$ , the complementary subset of  $V(i)$ , consisting of function nodes  $a$  where the variable appears negated—the edge  $(a, i)$  is a dashed line.  $V(i) \setminus b$  denotes the set  $V(i)$  without a node  $b$ . Similarly, for each function node  $a$ , we denote by  $V(a) = V_+(a) \cup V_-(a)$  the set of neighbouring variable

nodes, decomposed according to the type of edge connecting  $a$  and  $i$ , and by  $n_a$  its degree. Given a function node  $a$  and a variable node  $j$ , connected by an edge, it is also convenient to define the two sets  $V_a^s(j)$  and  $V_a^u(j)$ , where the indices  $s$  and  $u$  respectively refer to the neighbours which tend to make variable  $j$  respectively satisfy or not satisfy the clause  $a$ , defined as (see figure 1)

$$\text{if } J_j^a = 1 : V_a^u(j) = V_+(j); \quad V_a^s(j) = V_-(j) \setminus a \tag{2}$$

$$\text{if } J_j^a = -1 : V_a^u(j) = V_-(j); \quad V_a^s(j) = V_+(j) \setminus a. \tag{3}$$

### 3. Survey propagation

The survey propagation algorithm (SP) is a message passing algorithm in which the messages passed between function nodes and variable nodes have a probabilistic interpretation. Every message is the probability distribution of some warnings. In this section we provide a short self-contained introduction to SP, starting with the definition of warnings. We refer to [11, 16, 17] for the original derivation of SP from statistical physics tools.

#### 3.1. Warnings

The basic elementary message passed from one function node  $a$  to a variable  $i$  (connected by an edge) is a Boolean number  $u_{a \rightarrow i} \in \{0, 1\}$  called a ‘warning’.

Given a function node  $a$  and one of its variable nodes  $i$ , the warning  $u_{a \rightarrow i}$  is determined from the warnings  $u_{b \rightarrow j}$  arriving on all the variables  $j \in V(a) \setminus i$  according to

$$u_{a \rightarrow i} = \prod_{j \in V(a) \setminus i} \theta \left( -J_j^a \left( \sum_{b \in V(j) \setminus a} J_j^b u_{b \rightarrow j} \right) \right), \tag{4}$$

where  $\theta(x) = 0$  if  $x \leq 0$  and  $\theta(x) = 1$  if  $x > 0$ .

A warning  $u_{a \rightarrow i} = 1$  can be interpreted as a message sent from function node  $a$ , telling the variable  $i$  that it should adopt the correct value in order to satisfy clause  $a$ . This is decided by  $a$  according to the warnings which it has received from all the other variables  $j$  to which it is connected: if  $(\sum_{b \in V(j) \setminus a} J_j^b u_{b \rightarrow j}) J_j^a < 0$ , this means that the tendency for site  $j$  (in the absence of  $a$ ) would be to take a value which does not satisfy clause  $a$ . If all neighbours  $j \in V(a) \setminus i$  are in this situation, then  $a$  sends a warning to  $i$ .

The simplest message passing algorithm one can think of is just the propagation of warnings, in which the rule (4) is used as an update rule which is implemented sequentially. The interest in WP largely comes from the fact that it gives the exact solution for tree-problems. In more complicated problems where the factor graph is not a tree, it turns out that WP is unable to converge whenever the density of constraints is so large that the solution space is clustered. This has prompted the development of the more elaborate SP message passing procedure, which also appears to give good results for locally tree-like graphs with loops.

### 3.2. The algorithm

A message of SP, called a survey, passed from one function node  $a$  to a variable  $i$  (connected by an edge), is a real number  $\eta_{a \rightarrow i} \in [0, 1]$ . The interpretation of the survey  $\eta_{a \rightarrow i}$  is a probability among all clusters of satisfying assignments that a warning is sent from  $a$  to  $i$ .

The SP algorithm uses a random sequential update (see algorithm 1), which calls the local update rule SP-UPDATE (algorithm 2):

---

#### Algorithm 1 : SP

---

**Input:** The factor graph of a Boolean formula in conjunctive normal form; a maximal number of iterations  $t_{max}$ ; a requested precision  $\epsilon$

**Output:** UN-CONVERGED if SP has not converged after  $t_{max}$  sweeps. If it has converged: the set of all messages  $\eta_{a \rightarrow i}^*$

```

1: At time  $t = 0$ :
2: for every edge  $a \rightarrow i$  of the factor graph do
3:   randomly initialize the messages  $\eta_{a \rightarrow i}(t = 0) \in [0, 1]$ 
4: end for
5: for  $t = 1$  to  $t = t_{max}$  do
6:   sweep the set of edges in a random order, and update sequentially the
   warnings on all the edges of the graph, generating the values  $\eta_{a \rightarrow i}(t)$ ,
   using subroutine SP-UPDATE
7:   if  $|\eta_{a \rightarrow i}(t) - \eta_{a \rightarrow i}(t - 1)| < \epsilon$  on all the edges then
8:     the iteration has converged and generated  $\eta_{a \rightarrow i}^* = \eta_{a \rightarrow i}(t)$ 
9:     goto line 12
10:  end if
11: end for
12: if  $t = t_{max}$  then
13:   return UN-CONVERGED
14: else  $\{t < t_{max}\}$ 
15:   return the set of fixed point warnings  $\eta_{a \rightarrow i}^* = \eta_{a \rightarrow i}(t)$ 
16: end if

```

---

Whenever the SP algorithm converges to a fixed-point set of messages  $\eta_{a \rightarrow i}^*$ , one can use it in a decimation procedure in order to find a satisfiable assignment, if such an assignment exists. This procedure, called the survey inspired decimation (SID), is described in algorithm 3.

There exist several variants of this algorithm. In the code which is available at [23], for performance reasons we update simultaneously all  $\eta$  belonging to the same clause. The clauses to be updated are chosen in a random permutation order at each iteration step. The algorithm can also be randomized by fixing, instead of the most biased variables, one variable randomly chosen in the set of the  $x$  per cent variables with the largest bias. This strategy allows us to use some restart in the case where the algorithm has not found a solution. A fastest decimation can also be obtained by fixing in step 12, instead of one variable, a fraction  $f$  of the  $N_t$  variables (the most polarized ones) which have not yet been fixed (going back to one variable when  $fN_t < 1$ ).

**Algorithm 2** : subroutine SP-UPDATE( $\eta_{a \rightarrow i}$ )**Input:** Set of all messages arriving onto each variable node  $j \in V(a) \setminus i$ **Output:** New value for the message  $\eta_{a \rightarrow i}$ 

- 1: **for every**  $j \in V(a) \setminus i$  **do**
- 2:   compute the three numbers:

$$\begin{aligned} \Pi_{j \rightarrow a}^u &= \left[ 1 - \prod_{b \in V_a^u(j)} (1 - \eta_{b \rightarrow j}) \right] \prod_{b \in V_a^s(j)} (1 - \eta_{b \rightarrow j}) \\ \Pi_{j \rightarrow a}^s &= \left[ 1 - \prod_{b \in V_a^s(j)} (1 - \eta_{b \rightarrow j}) \right] \prod_{b \in V_a^u(j)} (1 - \eta_{b \rightarrow j}) \\ \Pi_{j \rightarrow a}^0 &= \prod_{b \in V(j) \setminus a} (1 - \eta_{b \rightarrow j}) \end{aligned} \quad (5)$$

if a set like  $V_a^s(j)$  is empty, the corresponding product takes value 1 by definition.

- 3:   Using these numbers , compute and return the new survey:

$$\eta_{a \rightarrow i} = \prod_{j \in V(a) \setminus i} \left[ \frac{\Pi_{j \rightarrow a}^u}{\Pi_{j \rightarrow a}^u + \Pi_{j \rightarrow a}^s + \Pi_{j \rightarrow a}^0} \right]. \quad (6)$$

If  $V(a) \setminus i$  is empty, then  $\eta_{a \rightarrow i} = 1$ .

- 4: **end for**

In [11], experimental results are reported which show that SID is able to solve efficiently huge SAT instances very close to the threshold. More specifically, the data discussed in [11] show that if SID fixes at each step only one variable it converges in  $\Theta(N^2 \log N)$  operations (the time taken by walksat [18] to solve the simplified sub-formula seems to grow more slowly). When a fraction of variables is fixed at each time step, we get a further reduction of the cost to  $O(N(\log N)^2)$  (the second log comes from sorting the biases).

In the following sections we show how the SP procedures can be made fully distributed and hence amenable to a parallel implementation.

## 4. Distributed SP I: diffusion algorithm

### 4.1. Simulating SP in parallel

The SP algorithm described in the preceding section is by construction a distributed algorithms, since updates of nodes are performed using message passing procedures from nearest neighbours. However, when one uses the surveys in order to find the SAT assignments, in the SID procedure, the decimation process breaks this local information exchange design: it requires global information, namely the maximal polarization field, used to decide which node has to be frozen at first. It is the purpose of this section to define a procedure able to diffuse such a global information, by using the message passing procedure.

**Algorithm 3 : SID**

**Input:** The factor graph of a Boolean formula in conjunctive normal form. A maximal number of iterations  $t_{max}$  and a precision  $\epsilon$  used in SP

**Output:** One assignment which satisfies all clauses, or 'SP UNCONVERGED', or 'probably UNSAT'

```

1: Random initial condition for the surveys
2: run SP
3: if SP does not converge then
4:   return 'SP UNCONVERGED' and stop (or restart, i.e.goto line 1).
5: else {if SP converges}
6:   keep the fixed-point surveys  $\eta_{a \rightarrow i}^*$ ; goto line 9
7: end if
8: Decimate:
9: if non-trivial surveys ( $\{\eta \neq 0\}$ ) are found then
10:  for each variable node  $i$  do
11:    evaluate the three 'biases'  $\{W_i^{(+)}, W_i^{(-)}, W_i^{(0)}\}$  defined by:

```

$$W_i^{(+)} = \frac{\hat{\Pi}_i^+}{\hat{\Pi}_i^+ + \hat{\Pi}_i^- + \hat{\Pi}_i^0} \quad (7)$$

$$W_i^{(-)} = \frac{\hat{\Pi}_i^-}{\hat{\Pi}_i^+ + \hat{\Pi}_i^- + \hat{\Pi}_i^0} \quad (8)$$

$$W_i^{(0)} = 1 - W_i^{(+)} - W_i^{(-)} \quad (9)$$

where  $\hat{\Pi}_i^+, \hat{\Pi}_i^-, \hat{\Pi}_i^0$  are defined by

$$\begin{aligned} \hat{\Pi}_i^+ &= \left[ 1 - \prod_{a \in V_+(i)} (1 - \eta_{a \rightarrow i}^*) \right] \prod_{a \in V_-(i)} (1 - \eta_{a \rightarrow i}^*) \\ \hat{\Pi}_i^- &= \left[ 1 - \prod_{a \in V_-(i)} (1 - \eta_{a \rightarrow i}^*) \right] \prod_{a \in V_+(i)} (1 - \eta_{a \rightarrow i}^*) \\ \hat{\Pi}_i^0 &= \prod_{a \in V(i)} (1 - \eta_{a \rightarrow i}^*) \end{aligned} \quad (10)$$

```

12:  fix the variable with the largest  $|W_i^{(+)} - W_i^{(-)}|$  to the value  $x_i = 1$  if  $W_i^{(+)} > W_i^{(-)}$ , to the value
     $x_i = 0$  if  $W_i^{(+)} < W_i^{(-)}$ 
13:  Clean the graph, which means: to remove the clauses satisfied by this fixing, to reduce the
    clauses that involve the fixed variable with opposite literal, and to update the number of unfixed
    variables.
14:  end for
15: else {if all surveys are trivial ( $\{\eta = 0\}$ ) }
16:  output the simplified sub-formula and run on it a local search process (e.g. walksat [18]).
17: end if
18: if the problem is solved completely by unit clause propagation then
19:  return 'SAT' and stop
20: else if no contradiction is found then
21:  continue the decimation process on the smaller problem (goto line 2)
22: else {if a contradiction is reached}
23:  return 'probably UNSAT'
24:  stop
25: end if

```

In the present distributed implementation of SP (see algorithm 4), each node is responsible for its own information; any sort of centralized information is prohibited. The information stored at a given function node  $a$  is the set of surveys  $\{\eta_{a \rightarrow i}, i \in V(a)\}$ . In turn, a given variable node  $i$  keeps information which is the set of cavity-fields  $\{\Pi_{i \rightarrow a}, a \in V(i)\}$ . The scheme amounts to implementing a message passing procedure such that function nodes send their  $\{\eta\}$  values to neighbouring variable nodes, and variable nodes send their field values in return. Each time a node receives a message from its

neighbours, it has to update its own information, according to the survey propagation equations. The broadcasting of information for each individual node is supposed to occur at random. Our simulation proceeds by random updating of each variable or function node. The update timescale, that is the rate at which nodes in a distributed implementation are able to collect new information and perform their own update, will be denoted by  $\tau$ . Typically, as we shall see, the full decimation process is of the order of  $10^4\tau$ , which means that on average each node has to update  $10^4$  times before the process is completed.

---

**Algorithm 4** : diffusive decimation algorithm

---

**Input:** The factor graph of a Boolean formula in conjunctive normal form; a maximal computational time  $t_{max}$ , a requested precision  $\epsilon$ , a number  $n_{su}$  of successive update with maximum polarization required to freeze a variable, the damping factor  $\delta$  for information decay, a random set of cavity biases  $\eta_{a \rightarrow i}$ .  
**Output:** UN-CONVERGED if the algorithm has not converged after  $t_{max}$ . If it has converged: one assignment which satisfy a fraction of the clauses, and a remaining reduced factor graph corresponding to a paramagnetic state which can be solved calling WALKSAT

```

1: At time  $t = 0$ :
2: Initialize the counter  $n_{su}(i) = 0, i = 1 \dots N$ 
3: parallel update of each variable node using VARIABLE-NODE-UPDATE.
4: for  $t < t_{max}$  AND  $H_{max} > \epsilon$  do
5:   parallel asynchronized update of each variable node using
     VARIABLE-NODE-UPDATE and each function node using FUNCTION-NODE-UPDATE,
     at a mean rate of time  $\tau$  per node.
6: end for
7: if  $H_{max} < \epsilon$  then
8:   return the set of assignments of the frozen variables and the remaining
     reduced factor graph of unfrozen variables.
9: else
10:  return UN-CONVERGED
11: end if

```

---

## 4.2. Propagation of information

Messages can contain additional information to the one needed to run SP. In the present case, the decimation procedure diffuses the information concerning the highest polarization fields  $H_{max} = \max(|W^+ - W^-|)$  in the system, in order that the variable node, being aware of being the most polarized one, can decide to freeze its assignment to the one given by the orientation of the polarization field.

Static information can freely propagate by simply diffusing in the network, and any component of the network should be informed by a time which has to be proportional to  $\log(N)$ ,  $N$  being the size (number of variables nodes) of the system. The difficulty is that the information to be broadcast may still vary with time before the system has reached equilibrium. The difficulty in propagating a signal saying that the system has

**Algorithm 5** : subroutine VARIABLE-NODE-UPDATE

**Input:** Set of all cavity bias surveys  $\{\eta_{a \rightarrow i}, a \in V(i)\}$  arriving at a given variable node  $i$ , information  $H_{max}(a)$  on the maximum polarization coming from each neighbor  $a \in V(i)$   
**Output:** New value for the set of probabilities  $\{\Pi_{i \rightarrow a}^{\pm}, \Pi_{i \rightarrow a}^0, a \in V(i)\}$ , new information  $H_{max}(i)$  on the maximum polarization as estimated by  $i$

```

1: for every  $a \in V(i)$  do
2:   compute the values of  $\Pi_{i \rightarrow a}^{\pm}, \Pi_{i \rightarrow a}^0$  using eq. (5).
3: end for
4: evaluate the two 'biases'  $\{W_i^{(+)}, W_i^{(-)}\}$  using eq. (7,8):
5: if  $H_{max}(a) < |W_i^{(+)} - W_i^{(-)}| + \epsilon$  then
6:    $H_{max}(i) = |W_i^{(+)} - W_i^{(-)}|$ 
7:    $n_{su}(i) = n_{su}(i) + 1$ 
8:   if  $n_{su}(i) > n_{su}$  then
9:     freeze variable  $i$  to the value  $x_i = 1$  if  $W_i^{(+)} > W_i^{(-)}$ , to the value  $x_i = 0$ 
       if  $W_i^{(+)} < W_i^{(-)}$ 
10:  end if
11: else
12:    $H_{max}(i) = \text{Max}(H_{max}(i), H_{max}(a)) * (1 - \delta)$ 
13:    $n_{su}(i) = 0$ 
14: end if

```

**Algorithm 6** : subroutine FUNCTION-NODE-UPDATE

**Input:** Set of all probabilities  $\{\Pi_{i \rightarrow a}^{\pm}, \Pi_{i \rightarrow a}^0, i \in V(a)\}$  arriving at a given function node  $a$ , information  $H_{max}(i)$  on the maximum polarization coming from each neighbor  $i \in V(a)$   
**Output:** New value for the set of cavity bias survey  $\{\eta_{a \rightarrow i}, i \in V(a)\}$ , new information  $H_{max}(a)$  on the maximum polarization as estimated by  $a$

```

1: for every  $i \in V(a)$  do
2:   Compute  $\eta_{a \rightarrow i}$  using eq. (6).
3:    $H_{max}(a) = \text{Max}(H_{max}(i), H_{max}(a)) * (1 - \delta)$ 
4: end for

```

converged might be even more severe. Indeed, this amounts to propagating the maximum displacement from the last update of each individual node, which is in essence transient information.

To take advantage of the basic message-passing procedure, in the case in which the information to be diffused can vary with time, we have to add a mechanism able to suppress obsolete information. The way we propose to do this is to incorporate some damping factor in the broadcast values.

Let us call  $\delta \ll 1$  this damping factor. The procedure goes as follows (see algorithm 5).

Each node ( $a$  for a function node or  $i$  for a variable node) stores its own estimate of the maximum polarization field  $H_{\max}$  in a local variable  $H_{\max}(i)$  or  $H_{\max}(a)$ . Consider first a variable node  $i$ . When this updates, in addition to the set  $\{\eta_{b \rightarrow i}\}$  from its neighbour function nodes  $b$ , it also collects the information  $H_{\max}(b)$  about the maximum, coming from these nodes, and compares it to its own estimation  $H_{\max}(i)$  and also to its own field value  $W_i = |W_i^+ - W_i^-|$ . If the new  $H_{\max}(i)$  estimation indicates that  $i$  is not the most polarized variable node ( $H_{\max}(i) > W_i$ ),  $H_{\max}(i)$  is multiplied by a factor  $(1 - \delta)$ . If instead it appears that  $i$  might be the most polarized node ( $H_{\max}(i) = W_i$ ), then  $H_{\max}(i)$  is kept as it stands, with no damping correction. For a function node, when it updates, we apply this damping factor anyway. The virtue of this damping effect is to eliminate any false information. Suppose indeed that at a given time a strongly polarized variable node diffuses in the network a very high value of  $H_{\max}$ ; if after some time this value is not confirmed (because this node converges to a less polarized state), which means that all local fields  $W_i$  will be smaller than  $H_{\max}$ , and no unit in the network can pretend to be the most polarized one. Then mechanically the estimation on the maximum will decay at a rate  $\delta$ , until a variable node presents a polarization field higher than the decaying false information.

The counterpart is that the damping takes effect spatially on the network. So typically, since the radius of the network scales like  $\log(N)$ , the distribution of the estimated maximum given by the set  $\{H_{\max}(i), H_{\max}(a)\}$  taken over the network will be enlarged by some factor  $\delta \log(N)$ .

As a result, when playing with the parameter  $\delta$ , two contradictory effects are at work:

- either the information diffuses rapidly with a poor precision (large  $\delta$ ),
- or it takes a long time before precise information is diffused (small  $\delta$ ).

When it comes to the decimation procedure, this can be actually turned into an advantage, since it provides us directly with a parameter able to fix, in a heuristic manner, the rate at which the decimation is performed. Each variable node is equipped with a counting variable  $n_{\text{su}}(i)$  initially set to zero and counting successive updates with the following two conditions:

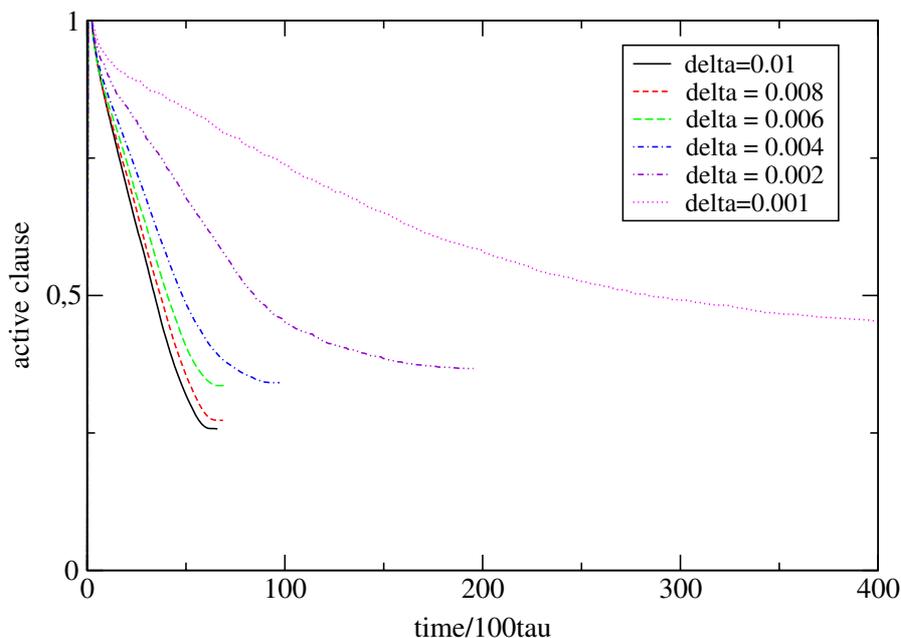
- the variation since last update of the  $\Pi_i$ 's is less than  $\epsilon$  (convergence criteria, typically  $\epsilon \simeq 10^{-3}$ ),
- $i$  may consider itself to be the most polarized variable ( $H_{\max}(i) = |W_i^+ - W_i^-|$ ).

When one of these conditions is not satisfied,  $n_{\text{su}}(i)$  is set to zero. If  $n_{\text{su}}(i)$  is greater than a parameter  $n_{\text{su}}(i)$ , the variable is allowed to freeze in the direction given by  $W_i^+ - W_i^-$ .

### 4.3. Experimental results

We have simulated the behaviour of the algorithm using a standard computer, which allows us to monitor how much progress in computing time would be brought by its fully parallel implementation, and to optimize the choices of parameters.

We use the random 3-SAT formulae as a benchmark problem. Let  $N$  denote the number of variable nodes and  $\alpha = M/N$  the ratio of the number of function nodes  $M$  over the number of variable nodes. The range of interest is  $\alpha \in [4, 4.25]$ . Simulations are performed with samples of  $10^5$  function nodes. Figures 2–5 present some average over



**Figure 2.** Number of active clauses as a function of time for different values of  $\delta$  for a 3-SAT problem of size  $N = 10^5$  at  $\alpha = 4.2$  and  $n_{\text{su}} = 20$ .

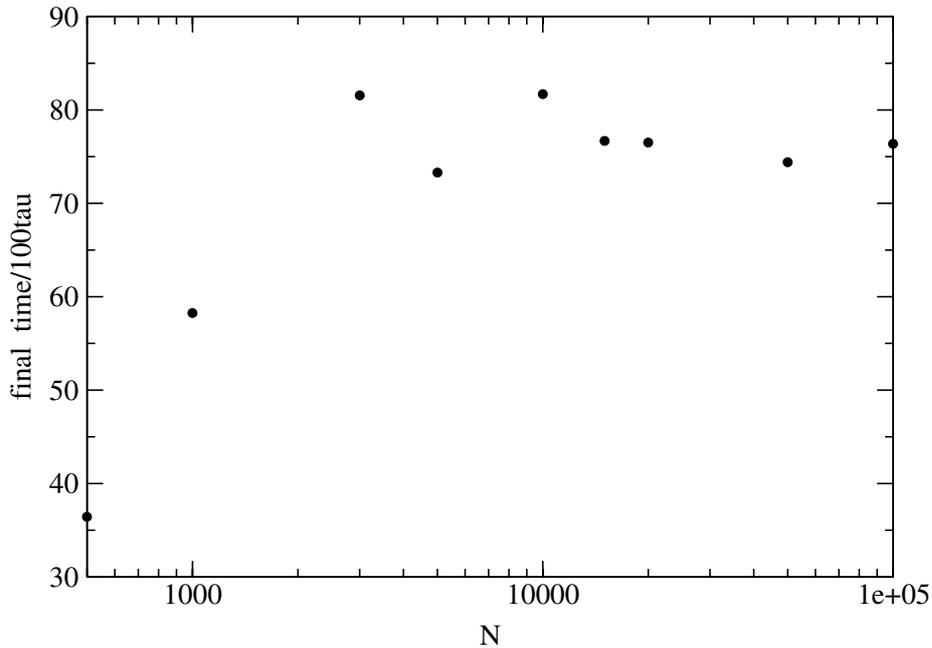
numerous decimation runs, with various parameters. The runs stop when a paramagnetic state is reached, or when the entropy becomes negative. In order to assess the performance of the algorithm, we measure the simulated time (the time it would take if the computation was really distributed) to reach a paramagnetic state, which can be handled afterwards by WALKSAT in order to find a solution.

The observables which we consider are the following.

- $N_a$ , the number of active (unfrozen) variable nodes.
- $M_a$ , the number of active function nodes.
- $\alpha_2$ , the ratio  $M_{a2}/N_a$ , where  $M_{a2}$  is the number of active two-clauses.
- $\alpha_3$ , the ratio  $M_{a3}/N_a$ , where  $M_{a3}$  is the number of active three-clauses. Since we start from a 3-SAT problem we have of course  $M_a = M_{a2} + M_{a3}$ .
- The entropy of the system, computed from the expression given in [11].

*4.3.1. Orders of magnitude for the tuning parameters.* The algorithm depends on two tunable parameters.

- The damping factor  $\delta$  on the diffusion of information. We find that  $\delta \in [0.001, 0.01]$  is roughly the range of effectiveness of this parameter for  $N = 10^5$ . This actually gives essentially the decimation rate between two convergences of SP. For  $\delta \geq 0.01$ , too many variables ( $\simeq 10^3$ ) are frozen at the same time, and the algorithm ceases to find a paramagnetic state. Instead, when  $\delta < 0.001$ , variables are decimated one by one. The time dependence of the decimation process with this parameter is illustrated in figure 2.



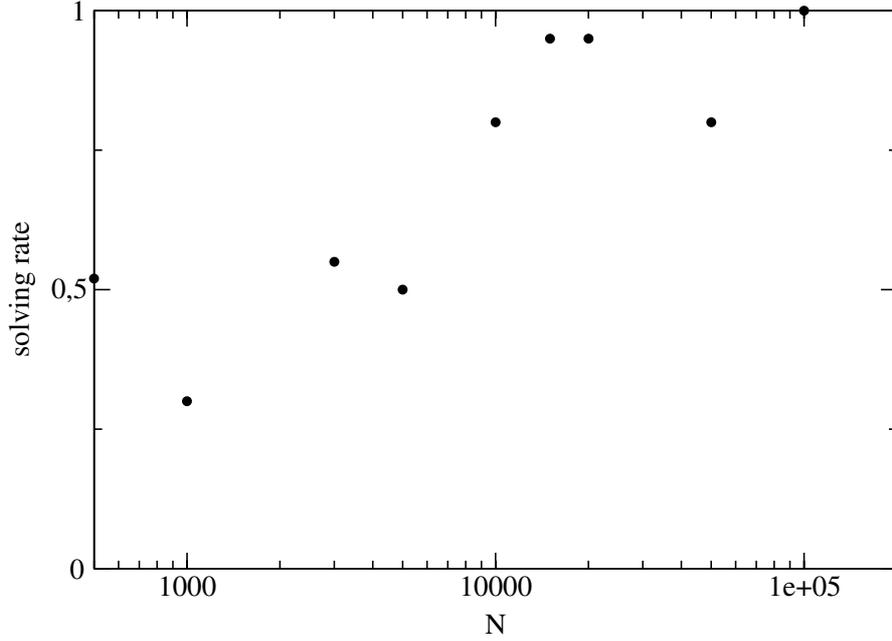
**Figure 3.** Decimation time up to the paramagnetic state as a function of the size of the system, for a 3-SAT problem at  $\alpha = 4.2$ ,  $\delta = 0.005$  and  $n_{\text{su}} = 20$ .

- $n_{\text{su}}$ , the number of successive successful updates needed for a variable to get frozen. It depends on the precision  $\epsilon$  which is required. For  $\epsilon = 10^{-3}$  it is found that the range of validity of this parameter is between 10 and 50. Above 50, SP converges before the next decimation cascade. Below 20 the decimation process is essentially continuous. Below 10 the algorithm ceases to converge to a paramagnetic state.

A reasonable choice for the selection of parameters leading to a fast but still successful decimation procedure is around  $\delta = 0.01$  and  $n_{\text{su}} = 10$ .

*4.3.2. Dependence on the size of the system.* Once the typical values for the tunable parameters are determined, it is possible to study the behaviour of the algorithm with the size of the problem,  $N$ . Figure 3 shows that the dependence on  $N$  is of the order of  $\log N$ , which is the best performance one could have with such an algorithm. Another expected result is that rate of successful decimation should increase when  $N$  increases, since finite size effects are known to alter the efficiency of algorithms for solving 3-SAT formulae. This is what is effectively observed in figure 4.

*4.3.3. Dependence on  $\alpha$ .* Depending on  $\alpha$  we observe that when we approach the transition point  $\alpha_c \simeq 4.26$  the final decimation rate increases (see figure 5), which is consistent with the hypothesis that the size of the backbone of solution clusters increases when we approach the critical value of  $\alpha$ . This can also be visualized when looking at the phase space trajectory of the decimation process (see figure 6). During the process two-clauses are generated, which leads to representation of the trajectory in the plane  $(\alpha_2, \alpha_3)$ . The trajectory is shorter when we get away from  $\alpha_c$ .



**Figure 4.** Rate of successful decimation process as a function of  $N$  for a 3-SAT problem at  $\alpha = 4.2$ ,  $\delta = 0.005$  and  $n_{\text{su}} = 20$ .

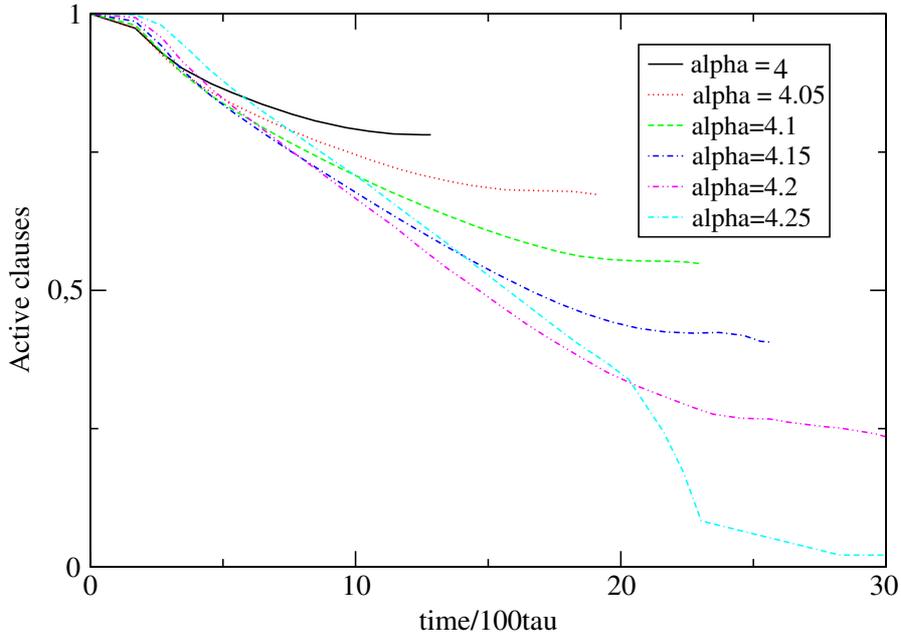
## 5. Distributed SP II: ‘reinforcement algorithm’

### 5.1. SP with external forcing field

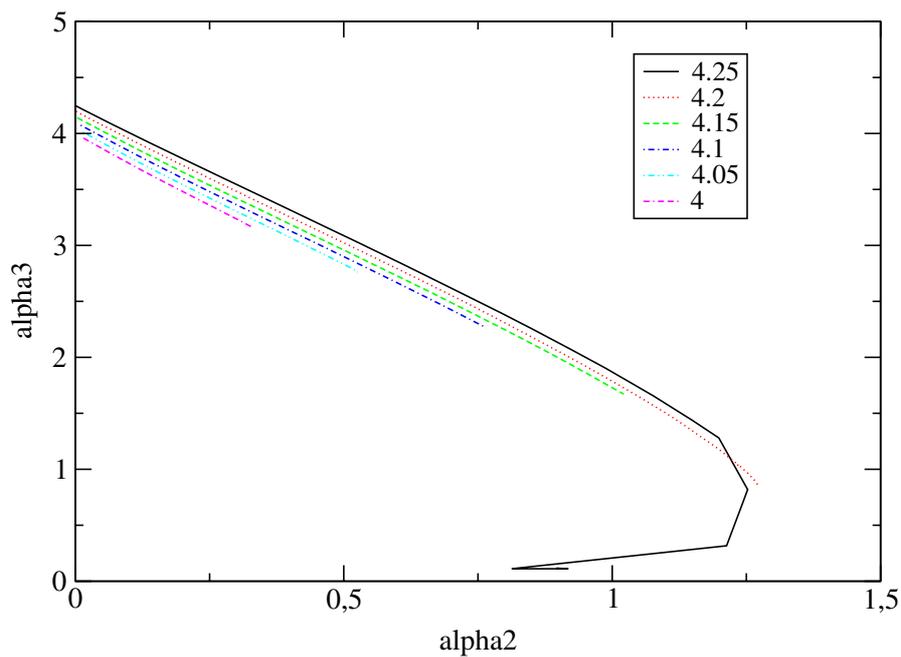
The SP procedure has recently been generalized to allow the retrieval of a large number of different solutions thanks to the introduction of *probability preconditionings* [9, 10]. The original factor graph is modified applying to each variable  $i$  a constant external forcing survey in direction  $\xi_i \in \{-1, 1\}$  and of intensity  $\eta_i^{\text{ext}} = \pi$ . An *a priori* probability  $\pi$  of assuming the value  $\pm 1$  is thus being assigned to the variable  $i$ . The components of the forcing modify the usual SP equation (5) in the following way:

$$\begin{aligned} \Pi_{j \rightarrow a}^{\pm} &= \left[ 1 - (1 - \pi \delta_{\xi_j, \pm}) \prod_{b \in V_{\pm}(j) \setminus a} (1 - \eta_{b \rightarrow j}) \right] (1 - \pi \delta_{\xi_j, \mp}) \prod_{b \in V_{\mp}(j)} (1 - \eta_{b \rightarrow j}) \\ \Pi_{j \rightarrow a}^0 &= (1 - \pi \delta_{\xi_j, +}) (1 - \pi \delta_{\xi_j, -}) \prod_{b \in V(j) \setminus a} (1 - \eta_{b \rightarrow j}). \end{aligned} \quad (11)$$

These relations, together with (6), form the SP equations with external messages (SP-ext). The quantities  $\hat{\Pi}_i^+$ ,  $\hat{\Pi}_i^-$ ,  $\hat{\Pi}_i^0$  used to define the local biases on each variable  $i$  (see equation (10)) are modified in a similar manner and expressed as a function of the new



**Figure 5.** Number of active clauses as a function of time during the decimation process for different values of  $\alpha$ , for a 3-SAT problem of size  $N = 10^5$  at  $\delta = 0.01$  and  $n_{\text{su}} = 10$ .



**Figure 6.** Trajectories in the phase space  $(\alpha_2, \alpha_3)$  during the decimation process for various values of  $\alpha$  for a 3-SAT problem of size  $N = 10^5$  at  $\delta = 0.01$  and  $n_{\text{su}} = 10$ . (Five samples for each value of  $\alpha$ ; all decimation processes were successful up to  $\alpha = 4.2$ , and all failed for  $\alpha = 4.25$ .) The slope of the linear part of the trajectories is  $\simeq -2.4$ .

probabilities  $\eta_{b \rightarrow j}^*$  at the fixed point:

$$\hat{\Pi}_{j \rightarrow a}^{\pm} = \left[ 1 - (1 - \pi \delta_{\xi_j, \pm}) \prod_{b \in V_{\pm}(j)} (1 - \eta_{b \rightarrow j}^*) \right] (1 - \pi \delta_{\xi_j, \mp}) \prod_{b \in V_{\mp}(j)} (1 - \eta_{b \rightarrow j}^*) \quad (12)$$

$$\hat{\Pi}_{j \rightarrow a}^0 = (1 - \pi \delta_{\xi_j, +}) (1 - \pi \delta_{\xi_j, -}) \prod_{b \in V(j)} (1 - \eta_{b \rightarrow j}^*).$$

The external messages drive the value of the surveys  $\eta_{a \rightarrow i}^*$  towards the clusters maximally aligned with the external vector field.  $\pi = 0$  corresponds to the case in which the external messages are absent; in the case  $\pi = 1$ , the system is fully forced in the direction  $\vec{\xi}$ . When  $\pi$  is correctly chosen, this enables us to address more specifically (compared to the SID algorithm, which permits us to retrieve only one solution) clusters close to a given region of the  $N$ -dimensional space. Thus, a right choice of the *a priori* probability  $\pi$ , and the use of alternative random directions  $\vec{\xi}$ , invariably permit us to retrieve many solutions in many different clusters.

The criterion used to decimate towards a solution is still the global criterion of the maximum of the magnetization over all variables. So, the decimation based on the SP-ext equations is not yet a distributed solver of the K-SAT formulae.

## 5.2. Presentation of the method

In order to make the decimation procedure fully distributed, we may adopt time-dependent external messages which are updated according to a reinforcement strategy. More precisely, good properties of convergence are obtained adopting the following scheme:

- in one time step (sweep), all the  $\eta$ s are updated in parallel using the SP equations with external messages (6), (11), which in this scheme are time-dependent;
- every second time step, one calculates all temporary local fields  $\{W_i^{(+)}, W_i^{(-)}, W_i^{(0)}\}$  using the equations (7)–(9), (13)—note that these are also functions of the external messages. The directions  $\xi_i, \forall i \in \{1, \dots, N\}$  of the temporary local messages, which are now time dependent, are updated in parallel to equal  $\text{sgn}(W_i^+ - W_i^-)$ , i.e. to align itself with the local bias  $W_i^+ - W_i^-$ .

As we shall see, with an appropriate choice of the intensity  $\pi$  of the forcing, most variables are completely polarized at the end of a single convergence. A solution of the K-SAT formula is finally found by fixing each variable in the direction of its local bias. The reinforcement algorithm (RA) defined above is purely local and hence gives an efficient distributed solver for random K-SAT formulae in the hard-SAT phase.

The algorithm is precisely described in algorithm 7. Its interpretation and a preliminary study of some of its properties are discussed in the subsequent sections.

To understand how the reinforcement strategy is working, one has to look back at the interpretation of the SP equations with an external forcing survey: imposing an external forcing survey of direction  $\xi_i \in \{-1, 1\}$  on the variable  $i$  tries to drive the solution of the cavity bias distributions towards the clusters with a local orientation matching the externally imposed direction  $\xi_i$ .

**Algorithm 7** : reinforcement algorithm

**Input:** The factor graph of a Boolean formula in conjunctive normal form; a maximal number of iterations  $t_{max}$ , a requested precision  $\epsilon$  and the probability  $\pi$  that the external message sends a warning  
**Output:** UN-CONVERGED if the algorithm has not converged after  $t_{max}$  sweeps. If it has converged: one assignment which satisfies all clauses

```

1: At time  $t = 0$ :
2: for every edge  $a \rightarrow i$  of the factor graph do
3:   randomly initialize the cavity biases  $\eta_{a \rightarrow i}(t = 0) \in [0, 1]$ . The additional
   surveys are initially set to  $Q_i(\vec{u}_i, 0)$ 
4: end for
5: for  $t = 1$  to  $t = t_{max}$  do
6:   update in parallel the cavity bias surveys on all the edges of the graph,
   generating the values  $\eta_{a \rightarrow i}(t)$ , using subroutine CBS-UPDATE.
7:   half of the times, update in parallel the reference direction  $\xi_i \in \{-1, 1\}$ 
   for all variables  $x_i$ , using subroutine FORC-UPDATE.
8:   if  $|\eta_{a \rightarrow i}(t) - \eta_{a \rightarrow i}(t - 1)| < \epsilon$  on all the edges then
9:     the iteration has converged and generated  $\eta_{a \rightarrow i}^* = \eta_{a \rightarrow i}(t)$ 
10:    goto line 13
11:   end if
12: end for
13: if  $t = t_{max}$  then
14:   return UN-CONVERGED
15: else {If  $t < t_{max}$ }
16:   return the satisfying assignment which is obtained by fixing the boolean
   variable  $x_i$  parallel to its local field :  $x_i = \text{sign}\{W_i^{(+)} - W_i^{(-)}\}$ , and by
   performing the simplifications over the neighboring variables
17: end if

```

**Algorithm 8** : subroutine CBS-UPDATE( $\eta_{a \rightarrow i}$ )

**Input:** Set of all cavity bias surveys arriving onto each variable node  $j \in V(a) \setminus i$   
**Output:** New value for the cavity bias survey  $\eta_{a \rightarrow i}$

```

1: for every  $j \in V(a) \setminus i$  do
2:   compute the values of  $\Pi_{j \rightarrow a}^{\pm}, \Pi_{j \rightarrow a}^{\mp}, \Pi_{j \rightarrow a}^0$  using eq. (11).
3:   Compute  $\eta_{a \rightarrow i}$  using eq. (6).
4: end for

```

Then, choosing  $\vec{\xi}$  parallel to the vector of the temporary local fields is a strategy which typically guarantees that, if the intensity  $\pi$  of the forcing is small enough, the overlap between the forcing vector field  $\vec{\xi}$  and the closest solution is continuously increasing, in a stronger way than in the case of the usual SP equations.

**Algorithm 9** : subroutine FORC-UPDATE( $\xi_i$ )

**Input:** Set of all cavity bias surveys arriving onto the variable node  $i$ , including the additional survey  $Q_i(\vec{e}_{\xi_i}; \pi)$

**Output:** New value for the direction  $\xi_i$  of the additional survey

- 1: Compute the local fields  $W_i^{(+)}, W_i^{(-)}, W_i^{(0)}$  using eq. (7, 8, 9, 12)
- 2: Compute  $\xi_i = \text{sign}(W_i^{(+)} - W_i^{(-)})$

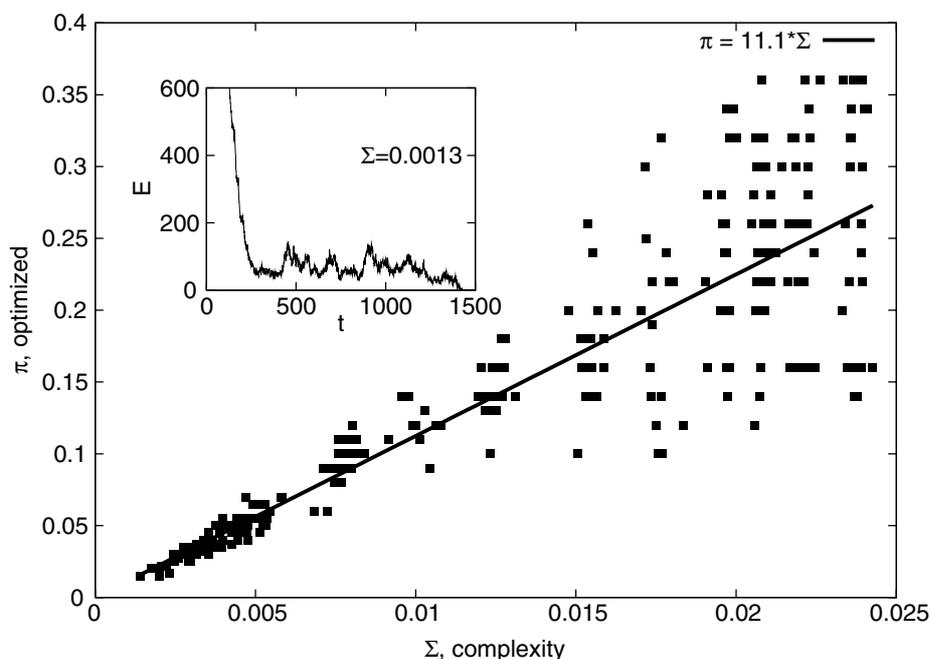
When  $\pi$  is correctly chosen, some satisfying assignments tend to become stable fixed points of the RA dynamics. The optimal value of  $\pi$  is a trade-off between the need for convergence of the SP equations with external forcing survey, the strength of the stabilization induced on the satisfying assignments by the update rule and the amplitude of the perturbation brought by the external forcing. There is *a priori* no guarantee that there exists a particular value of the intensity  $\pi$  for which the algorithm, starting from random initial conditions, gets trapped by a single cluster, and eventually converges to a single solution inside this cluster, without any further decimation. But, our experimental results show that, for  $N$  big enough and for almost all values of the parameter  $\alpha$  corresponding to the SAT phase, there exists an intensity  $\pi$  for which the overlap between the set of local fields and the closest satisfying assignment increases until it eventually reaches unity at the end of a unique convergence.

### 5.3. Results

*5.3.1. Determination of the optimal intensity of the forcing  $\pi$ .* The efficiency of the method is tested on random 3-SAT formulae. One first determines the optimal intensity  $\pi_{\text{opt}}$ , which is the intensity of the forcing which permits us to find a satisfying assignment in the minimum number of steps. When the density of ground state clusters increases (for lower values of  $\alpha$ ),  $\pi_{\text{opt}}$  should be set at a higher value to be able to bias the convergence towards a unique cluster; however, a too high value of  $\pi$  may let the algorithm converge too rapidly in different parts of the graph towards assignments belonging to different clusters. The algorithm would then loop without convergence among these clusters. For these two reasons,  $\pi_{\text{opt}}$  should be positively correlated with the complexity  $\Sigma = \log(N_{\text{Cl}})/N$ .

One observes that the reinforcement process diverges if the parallel update of the forcing messages is performed at a too high frequency with respect to the frequency of the parallel updates of the  $\eta$ ; the local loops in which the algorithm gets trapped are likely due to simultaneous convergence towards different satisfying assignments in different parts of the graph. A good strategy, systematically used for the experiments of the present paper, is to perform the parallel updates of the forcing messages once every second parallel update of the  $\eta$ .

In the hard-SAT phase, one determines  $\pi_{\text{opt}}$  as a function of the calculated complexity [17]. As predicted, the experiments show that, for  $\alpha_D < \alpha < 4.245$ ,  $\pi_{\text{opt}}$  is positively correlated, and on average scales linearly, with  $\Sigma$  ( $\pi_{\text{opt}} = 11.1 \times \Sigma$ , cf figure 7, each point representing a different formula), and that the probability of converging towards a satisfying assignment using this value tends to unity ( $N = 10^5$ ,  $\epsilon = 10^{-3}$ ,  $t_{\text{max}} = 1000$ ; cf figure 8).

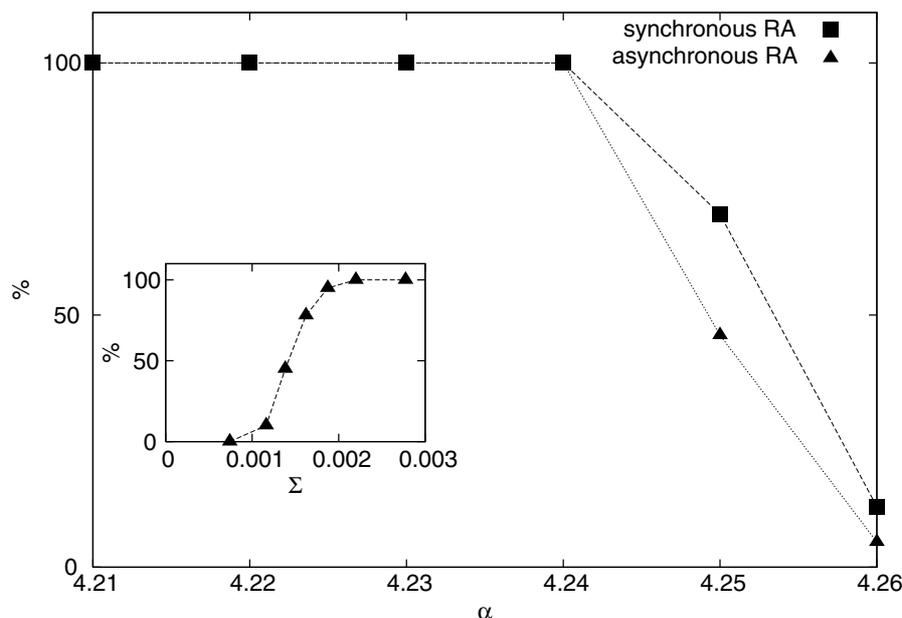


**Figure 7.** Scaling of the optimized  $\pi$  against the complexity  $\Sigma$ . The optimized  $\pi$ , which is the intensity of the forcing permitting us to find a satisfying assignment in the minimum number of steps, is recalculated for each formula: it depends on average linearly on the complexity:  $\pi = 11.1 \times \Sigma$  ( $N = 10^5$ ,  $K = 3$ ,  $4.0 < \alpha < 4.24$ ). Each point corresponds to a different formula; note that the relationship  $\pi_{\text{opt}} = f(\Sigma)$  becomes slightly sub-linear for  $\Sigma < 0.003$ . The sub-plot shows the evolution of the energy of the dump solution (in which each variable is fixed parallel to the temporary local field) that would be obtained if the convergence was stopped at time  $t$  ( $\alpha = 4.24$ ).

*5.3.2. Range of validity of the method.* In this section we compare RA with the so-called serialized survey propagation-inspired decimation (SID).

Figure 8 plots the percentage of successful runs (i.e. of convergence towards a satisfying assignment after a single run) of both the synchronous and the asynchronous RAs as a function of  $\alpha$ . One calls the algorithm previously described (algorithm 7), in which the updates of all the  $\eta$ , as well as respectively the updates of all forcing messages, are performed simultaneously, synchronous RA. Conversely, in the asynchronous RA, the updates are performed sequentially in a random permutation order (note that, in the latter case, forcing messages can be updated as frequently as the  $\eta$ ). For  $\alpha = 4.24$ , both algorithms invariably permit us to retrieve a solution. This is in accordance with the result presented in [11], in which the classical SID has been shown to always find a solution by fixing 0.125% of variables after each convergence of the SP equations.

The cut-off value of the complexity  $\Sigma$ , for which the decimation converges in half of the cases, is  $\Sigma_{1/2} = 0.0013$  (sub-plot of figure 8).  $\Sigma_{1/2}$  was constant for all sizes of the graph  $N > 2 \times 10^4$ . When the complexity decreases, the decimation progressively fails. Depending on the value of  $\pi$ , either it converges towards a set of local fields, which corresponds to a weighted average of the contributions of numerous clusters of solutions,



**Figure 8.** Percentage of successful runs (convergence leading to a satisfying assignment) as a function of  $\alpha$ . The sub-plot shows the same percentage as a function of the calculated complexity  $\Sigma$ . The cut-off complexity, for which the RA is successful in half of the cases tested, is  $\Sigma_{1/2} = 0.0013$  ( $N = 10^5$ ,  $\epsilon = 10^{-3}$ ,  $t_{\max} = 1500$ ). The range of validity of the reinforcement algorithm thus matches the one of the SID (see text).

or it loops among low energy configurations (usually, for  $N = 10^5$ ,  $\alpha = 4.25$ , one finally finds assignments of energy  $E = 0-50$ ).

Moreover, asymptotically, the numerical critical  $\alpha$  above which the SID, without backtracking, should not permit us to find a satisfying assignment has been estimated to be  $\alpha_A = 4.252$  [19], substantially smaller than the theoretical critical value  $\alpha_C = 4.2667$  (for the improvement brought by the backtracking procedure, see [11, 20]).

One may study the behaviour of the RA at  $\alpha_A$  for  $N = 10^6$ . The calculated complexity of the formulae with  $N = 10^6$  and  $\alpha = 4.252$  equals  $\Sigma = 0.00133 \pm 0.00013$ , in agreement with the cut-off complexity determined for formulae of size  $N = 10^5$  (figure 8). For  $N = 10^6$  and  $\alpha = 4.252$ , one then sets  $\pi_{\text{opt}} = 10.5 \times \Sigma$  (in agreement with the curve of figure 7 for  $N = 10^5$ ). The ‘reinforcement algorithm’ using these parameters nicely permits us to converge towards a solution in 10 out of the 15 formulae tested.

One concludes that the present algorithm is valid at least in the same range of parameters as the classical SID [11].

*5.3.3. Which solutions are addressed?* The satisfying assignments found by the present algorithm are highly dependent on the initial conditions, i.e. on the initial random values imposed on the  $\eta$ . For different random initial conditions, one converges towards solutions belonging to different clusters.

To know where the found solutions are located, one chooses an instance from a random graph ensemble in which the number of clauses where any given variable appears

negated or directed are kept strictly equal (*bar-balanced formulae*). For such an instance ( $N = 30\,000$ ,  $\alpha = 3.33$ ), the clusters of solutions are expected (and have been observed) to be uniformly distributed over the whole  $N$ -dimensional space [10]. The RA is launched in such an instance 800 times starting from different random conditions. The histogram of the Hamming distances  $d(S_1, S_2) = \frac{1}{2}(1 - (1/N) \sum_i x_i^{(S_1)} x_i^{(S_2)})$  between any two of the 800 solutions peaks at 0.4995 (with a standard deviation of 0.0118).

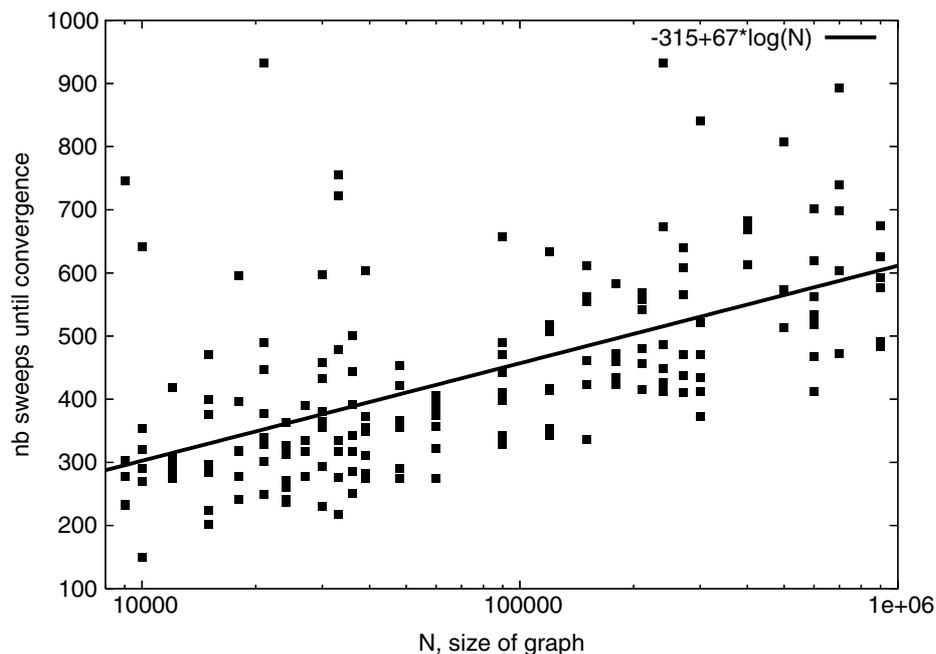
There is no guarantee that the present algorithm is able to address all clusters of solutions, but the fact that the distance distribution among the found solutions of a bar-balanced formula peaks at 0.5 with a standard deviation expected from a random distribution (i.e.  $2/\sqrt{N} = 0.0115$ ) suggests that the addressable solutions could be homogeneously distributed over the space of the clusters of solutions.

*5.3.4. Logarithmic dependence of time on the size of the graph.* One now analyses the evolution of the convergence time as a function of the size of the graph. For this, one determines, for a fixed value of  $\alpha = 4.22$ , the simulated time that the convergence would take if implemented on a distributed device, as a function of the size  $N$  of the graph (figure 9). For  $N$  ranging from 9000 to  $9 \times 10^5$ , the time of convergence is well approximated by a logarithmic fit ( $t = -315 + 67 * \log(N)$ ). The unit of time is one parallel update of all  $\eta$ . Each point corresponds to a different random formula. Note that, above  $N > 3 \times 10^4$ , all trials were successful, and the RA invariably leads to a satisfying assignment. For smaller graphs, due to finite-size effects, the algorithm either fails to converge or converges to an assignment with a few contradictions (generally one to four) in a few cases: at most 5% of the experiments for  $N = 9000$  and none for  $N > 30\,000$ .

As local messages need a logarithmic time to reach all parts of the graph, and as, in the hard-SAT phase, the assignment of a value to a given variable  $i$  does not depend on purely local constraints but on the state of the whole graph, one expects that, by the use of purely local messages on a distributed device, the problem cannot be solved faster than in a logarithmic time.

*5.3.5. Performance of the reinforcement algorithm on a serialized implementation.* Numerical experiments show that RA scales as  $N \times \log(N)$  when implemented on a serialized computer. In the scheme of the classical SID, a slightly slower scaling, in  $N \times (\log(N))^2$ , can be achieved by fixing after each convergence a given percentage of the total amount of variable [11].

Moreover, Braunstein *et al* [11] reports that, for  $\alpha = 4.24$  and  $N = 10^5$ , by fixing at least  $f = 0.125\%$  of the  $N$  variables at the end of each convergence, the SID reaches a paramagnetic state after an average of 7460 sweeps, i.e. 7460 updates of all remaining  $\eta$ . Conversely, the synchronous RA only needs 600 parallel updates of the  $\eta$  plus 300 parallel updates of the forcing messages to converge towards a solution. Its asynchronous version requires 450–500 complete updates of the  $\eta$  and of the forcing messages to arrive at a solution. This means that, in a serialized computer, close to the critical point, the present algorithm performs faster than the classical SID by a factor of five. This can be checked by running the two algorithms on a 2.4 GHz PC: the formula ( $N = 10^5$  and  $\alpha = 4.24$ ) is solved on average in 30 min using the SID with  $f = 0.125\%$  and in 7 min using the ‘reinforcement algorithm’ (either asynchronous or synchronous).



**Figure 9.** The number of parallel sweeps needed to find a solution depends logarithmically on  $N$  ( $\alpha = 4.22$ ,  $\pi = 0.04$ ). The examples in which the convergence fails (due to finite size effects, which have not been observed for  $N > 30\,000$ ) are not represented.

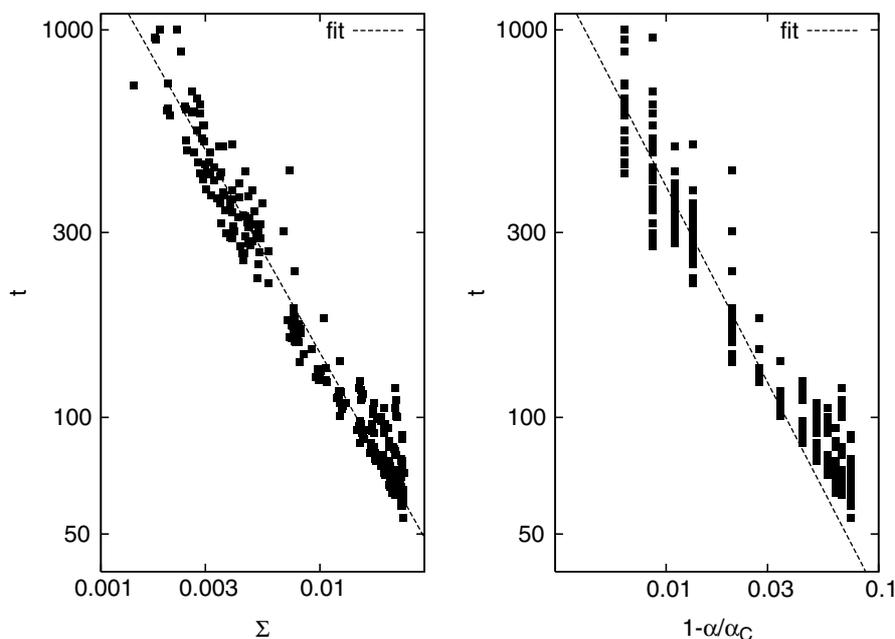
We finally characterize the slowing down of the RA when approaching the critical value  $\alpha_C = 4.2667$  (figure 10). Far away from the critical point, the time of convergence is highly reliable. Close to the critical point, the algorithm gets slower and the time of convergence becomes more variable. Setting the intensity of the forcing at its optimal value, one finds that, for  $\alpha < \alpha_A$ , the time of convergence is inversely proportional to the complexity  $\Sigma$ :  $t = t_0^\Sigma / \Sigma$  ( $t_0^\Sigma = 1.474$ , figure 10, left panel); moreover, as expected from the pseudo-linear relationship between  $\Sigma$  and  $\alpha$ , the time of convergence also scales with  $1/(\alpha - \alpha_C)$ :  $t = t_0^\alpha / (\alpha - \alpha_C)$  ( $t_0^\alpha = 3.02$ , figure 10, right panel).

Very close to  $\alpha_C$  ( $\alpha > \alpha_A$ ), the RA is unable to find ground states and only explores quasi-optimal assignments.

## 6. Conclusion

This study aimed at making fully distributed the survey propagation decimation algorithm.

The first distributed solver, the ‘SP diffusion algorithm’, diffuses as dynamical information the maximum bias over the system, so that variable nodes can decide to freeze in a self-organized way. Its properties (solving rate, final number of active clauses when it reaches the paramagnetic state) are comparable with the previously described serialized ‘SP inspired decimation’. The new feature is that the simulated time of convergence towards a solution, if this was implemented on a fully distributed device, goes as  $\log(N)$ , i.e. scales optimally with the size of the graph.



**Figure 10.** Time of convergence as a function of  $\Sigma$  (left) or  $\alpha$  (right), using for  $\pi$  the optimal intensity of the forcing ( $N = 10^5$ ). Close to the critical point, the time of convergence is shown to be inversely proportional either to  $\Sigma$  or to  $1 - \alpha/\alpha_C$ . The fit of the data  $t = t_0^\Sigma \times \Sigma^{-\beta}$  over  $\Sigma \in [0, 0.008]$  gives  $t_0^\Sigma = 1.474$  and  $\beta = 0.9994$ . The fit  $t = t_0^\alpha \times (1 - \alpha/\alpha_C)^{-\gamma}$  over  $\alpha \in [4.14, 4.25]$  gives  $t_0^\alpha = 3.02$  and  $\gamma = 1.057$ .

The second solver, the ‘SP reinforcement algorithm’, makes use of time-dependent local external forcing messages, which let the variables get completely polarized in the direction of a solution at the end of a single convergence. The estimated time of convergence of this solver, when implemented on a distributed device, also goes as  $\log(N)$ .

The present algorithm is the fastest existing solver of the K-SAT formulae, when implemented either on a distributed device or on a serialized computer.

Moreover, the strategies proposed in the present manuscript are not specific to the K-SAT problem, but are likely to apply to many types of NP-complete constraint satisfaction problem (CSP) with direct applications (e.g. data compression algorithm [1]).

The different codes are available at [22, 23].

## Acknowledgments

This work was supported by EVERGROW, integrated project No 1935 in the complex systems initiative of the Future and Emerging Technologies Directorate of the IST Priority, EU Sixth Framework. We thank Demian Battaglia and Alfredo Braunstein for very helpful discussions.

## References

- [1] Ciliberti S, Mézard M and Zecchina R, *Lossy data compression with random gates*, 2005 *Phys. Rev. Lett.* **95** 038701 [[cond-mat/0504509](#)]
- [2] Cocco S, Dubois O, Mandler J and Monasson R, *Rigorous decimation-based construction of ground pure states for spin glass models on random lattices*, 2003 *Phys. Rev. Lett.* **90** 047205

- [3] Dubois O, Monasson R, Selman B and Zecchina R (ed), *Phase transitions in combinatorial problems*, 2001 *Theor. Comput. Sci.* **265** 1
- [4] Mézard M, Mora T and Zecchina R, *Clustering of solutions in the random satisfiability problem*, 2005 *Phys. Rev. Lett.* **94** 197205 [[cond-mat/0504070](#)]
- [5] Mézard M, Ricci-Tersenghi F and Zecchina R, *Two solutions to diluted  $p$ -spin models and XORSAT problems*, 2003 *J. Stat. Phys.* **111** 505
- [6] Achlioptas D and Ricci-Tersenghi F, 2005 in preparation
- [7] Richardson T and Urbanke R, *The capacity of low-density parity-check codes under message-passing decoding*, 2001 *IEEE Trans. Inf. Theory* **47** 599
- [8] Luby M G, Mitzenmacher M, Shokrollahi M A and Spielman D A, 2001 *IEEE Trans. Inf. Theory* **47** 585
- [9] Battaglia D, Braunstein A, Chavas J and Zecchina R, 2005 *Phys. Rev. E* **72** 015103
- [10] Battaglia D, Braunstein A, Chavas J and Zecchina R, 2005 *Prog. Theor. Phys. Suppl.* **157** 330
- [11] Braunstein A, Mézard M and Zecchina R, 2005 *Random Struct. Algorithms* **27** 201
- [12] Gallager R G, *Low density parity check codes*, 1962 *IRE Trans. Inf. Theory* **8** 21
- [13] Kschischang F R, Frey B J and Loeliger H-A, *Factor graph and the sum-product algorithm*, 2001 *IEEE Trans. Inf. Theory* **47** 498
- [14] Levine B, Taylor R R and Schmit H, *Implementation of near Shannon limit error-correcting codes using reconfigurable hardware*, 2000 *Proc. IEEE Symp. on Field Programmable Computing Machines* pp 217–26
- [15] MacKay D J C, *Good error-correcting codes based on very sparse matrices*, 1999 *IEEE Trans. Inf. Theory* **45** 399
- [16] Mézard M, Parisi G and Zecchina R, *Analytic and algorithmic solution of random satisfiability problems*, 2002 *Science* **297** 812
- [17] Mézard M and Zecchina R, *The Random  $K$ -satisfiability problem: from an analytic solution to an efficient algorithm*, 2002 *Phys. Rev. E* **66** 056126
- [18] Selman B, Kautz H and Cohen B, 1994 *Proc. AAAI-94 (Seattle, WA)* pp 337–43
- [19] Parisi G, *Some remarks on the survey decimation algorithm for  $K$ -satisfiability*, 2003 Preprint [cs.CC/0301015](#)
- [20] Parisi G, *A backtracking survey-propagation algorithm for  $K$ -satisfiability*, 2003 Preprint [cond-mat/0308510](#)
- [21] Yedidia J S, Freeman W T and Weiss Y, *Generalized belief propagation*, 2001 *Advances in Neural Information Processing Systems* vol 13 (Cambridge, MA: MIT Press) pp 689–95
- [22] The diffusion code: <http://www-rocq.inria.fr/who/Cyril.Furtlehner/>
- [23] The asynchronous and the synchronous reinforcement code: <http://www.ictp.trieste.it/~zecchina/SP/>