Analysing Survey Propagation Guided Decimation on Random Formulas[*†]

Samuel Hetterich

Goethe University, Mathematics Institute, Frankfurt, Germany
hetterich@math.uni-frankfurt.de

Abstract

Let $\vec{\Phi}$ be a uniformly distributed random k-SAT formula with n variables and m clauses. For clauses/variables ratio $m/n \le r_{k\text{-SAT}} \sim 2^k \ln 2$ the formula $\vec{\Phi}$ is satisfiable with high probability. However, no efficient algorithm is known to provably find a satisfying assignment beyond $m/n \sim 2k \ln(k)/k$ with a non-vanishing probability. Non-rigorous statistical mechanics work on k-CNF led to the development of a new efficient "message passing algorithm" called *Survey Propagation Guided Decimation* [Mézard et al., Science 2002]. Experiments conducted for $k = 3, 4, 5$ suggest that the algorithm finds satisfying assignments close to $r_{k\text{-SAT}}$. However, in the present paper we prove that the basic version of Survey Propagation Guided Decimation fails to solve random k-SAT formulas efficiently already for $m/n = 2^k(1 + \varepsilon_k)\ln(k)/k$ with $\lim_{k\to\infty} \varepsilon_k = 0$ almost a factor k below $r_{k\text{-SAT}}$.

1998 ACM Subject Classification G.2.1 Combinatorics

Keywords and phrases Survey Propagation Guided Decimation, Message Passing Algorithm, Graph Theory, Random k-SAT

Digital Object Identifier 10.4230/LIPIcs.ICALP.2016.65

1 Introduction

Random k-SAT instances have been known as challenging benchmarks for decades [9, 28, 33]. The simplest and most intensely studied model goes as follows. Let $k \ge 3$ be an integer, fix a density parameter $r > 0$, let n be a (large) integer and let $m = \lceil rn \rceil$. Then $\Phi = \Phi_k(n, m)$ signifies a k-CNF chosen uniformly at random among all $(2n)^{km}$ possible formulas. With k, r fixed the random formula is said to enjoy a property *with high probability* if the probability that the property holds tends to 1 as $n \to \infty$.

The conventional wisdom about random k-SAT has been that the problem of finding a satisfying assignment is computationally most challenging for r below but close to the *satisfiability threshold* $r_{k-\mathrm{SAT}}$ where the random formula ceases to be satisfiable w.h.p. [28]. Whilst the case $k = 3$ may be the most accessible from a practical (or experimental) viewpoint, the picture becomes both clearer and more dramatic for larger values of k. Asymptotically the k-SAT threshold reads $r_{k-\mathrm{SAT}} = 2^k \ln 2 - (1 + \ln 2)/2 + \varepsilon_k$, where $\varepsilon_k \to 0$ in the limit of large k [14]. However, the best current algorithms are known to find satisfying assignments in polynomial time merely up to $r \sim 2^k \ln k/k$ [11]. In fact, standard heuristics such as Unit Clause Propagation bite the dust for even smaller densities, namely $r = c2^k/k$ for a certain

[*] A full version of this paper is available at http://arxiv.org/abs/1602.08519.
[†] The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 278857–PTCC

43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016).
Editors: Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi;
Article No. 65; pp. 65:1–65:12
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

absolute constant $c > 0$ [17]. The same goes (provably) for various DPLL-based solvers [1, 30]. Hence, there is a factor of about $k/\ln k$ between the algorithmic threshold and the actual satisfiability threshold.

In the early 2000s physicists put forward a sophisticated but non-rigorous approach called the *cavity method* to tackle problems such as random k-SAT both analytically and algorithmically. In particular, the cavity method yields a *precise* prediction as to the value of $r_{k-\mathrm{SAT}}$ for any $k \geq 3$ [24, 26], which was recently verified rigorously for sufficiently large values of k [14]. Additionally, the cavity method provided a heuristic explanation for the demise of simple combinatorial or DPLL-based algorithms well below $r_{k-\mathrm{SAT}}$. Specifically, the density $2^k \ln k/k$ marks the point where the geometry of the set of satisfying assignments changes from (essentially) a single connected component to a collection of tiny well-separated clusters [22]. In fact, a typical satisfying assignment belongs to a "frozen" cluster, i.e., there are extensive long-range correlations between the variables. The cluster decomposition as well as the freezing prediction have largely been verified rigorously [29, 3] and we begin to understand the impact of this picture on the performance of algorithms [2].

But perhaps most remarkably, the physics work has led to the development of a new efficient "message passing algorithm" called *Survey Propagation Guided Decimation* to overcome this barrier [6, 21, 27, 31]. More precisely, the algorithm is based on a heuristic that is designed to find whole frozen clusters not only single satisfying assignments by identifying each cluster by the variables determined by long-range correlations and locally "free" variables. Thus, by its very design *Survey Propagation Guided Decimation* is built to work at densities where frozen clusters exist. Although the experimental performance for small k is outstanding this yields no evidence of a relation between the occurrence of frozen clusters and the success of the algorithm. Yet not even the physics methods lead to a precise explanation of these empirical results or to a prediction as to the density up to which we might expect SP to succeed for general values of k. In effect, analysing SP has become one of the most important challenges in the context of random constraint satisfaction problems.

The present paper furnishes the first rigorous analysis of `SPdec` (the basic version of) Survey Propagation Guided Decimation for random k-SAT. We give a precise definition and detailed explanation below. Before we state the result let us point out that two levels of randomness are involved: the choice of the random formula $\vec{\Phi}$, and the "coin tosses" of the randomized algorithm `SPdec`. For a (fixed, non-random) k-CNF Φ let success(Φ) denote the probability that `SPdec`(Φ) outputs a satisfying assignment. Here, of course, "probability" refers to the coin tosses of the algorithm only. Then, if we apply `SPdec` to the *random* k-CNF $\vec{\Phi}$, the success probability success($\vec{\Phi}$) becomes a random variable. Recall that $\vec{\Phi}$ is unsatisfiable for $r > 2^k \ln 2$ w.h.p..

▶ **Theorem 1.** *There is a sequence* $(\varepsilon_k)_{k\geq 3}$ *with* $\lim_{k\to\infty} \varepsilon_k = 0$ *such that for any* k, r *satisfying* $2^k(1 + \varepsilon_k)\ln(k)/k \leq r \leq 2^k \ln 2$ *we have* $\mathrm{success}(\vec{\Phi}) \leq \exp(-\Omega(n))$ *w.h.p.*

If the success probability is exponential small in n sequentially running `SPdec` a sub-exponential number of times will not find a satisfying assignment w.h.p. rejecting the hypotheses that `SPdec` solves random k-SAT formulas efficiently for considered clauses/variables ratio. Thus, Theorem 1 shows that `SPdec` does not outclass far simpler combinatorial algorithms for general values of k. Even worse, in spite of being designed for this very purpose, the SP algorithm does *not* overcome the barrier where the set of satisfying assignments decomposes into tiny clusters asymptotically. This is even more astonishing since it is possible to *prove* the existence of satisfying assignments up to the satisfiability threshold rigorously based on the cavity method but algorithms designed by insights of this approach fail far below that threshold. Nevertheless, let me note that the insights gained from Theorem 1 is actually in line with some non-rigorous physics work on the SP algorithm. Still, there is some arguing if there is any connection between the failure of algorithms and either the clustering or the so called freezing phenomenon. Both, neither the connection to clustering nor to freezing have been rigorously proven yet.

We are going to describe the SP algorithm in the following section. Let us stress that Theorem 1 pertains to the "vanilla" version of the algorithm. Unsurprisingly, more sophisticated variants with better empirical performance have been suggested, even ones that involve backtracking [23]. Also the first version introduced by Mézard, Parisi and Zecchina [27] contained a bias towards "frozen" variables for the choice of the variable at each decimation step. However, the basic version of the SP algorithm analysed in the present paper arguably (regarding the physicists picture of freezing, correlation decay, replica symmetry assumption [26]) encompasses all the conceptually important features of the SP algorithm.

The only prior rigorous result on the Survey Propagation algorithm is the work of Gamarnik and Sudan [19] on the k-NAESAT problem (where the goal is to find a satisfying assignment whose complement is satisfying as well). However, Gamarnik and Sudan study a "truncated" variant of the algorithm where only a bounded number of message passing iterations is performed. The main result of [19] shows that this version of Survey Propagation fails for densities about a factor of $k/\ln^2 k$ below the NAE-satisfiability threshold and about a factor of $\ln k$ above the density where the set of NAE-satisfying assignments shatters into tiny clusters. Though, experimental data and the conceptional design of the SP algorithm suggest that it exploits its strength in particular by iterating the message passing iterations a unbounded number of times that depends on n. In particular, to gather information from the set of messages they have to converge to a fixed point which turns out to happen only after a number of iterations of order $\ln(n)$.

An in-depth introduction to the cavity method and its impact on combinatorics, information theory and computer science can be found in [25, 26].

2 The SPdec algorithm

The proof of Theorem 1 is by extension of the prior analysis [10] of the much simpler *Belief Propagation Guided Decimation* algorithm. To outline the proof strategy and to explain the key differences, we need to discuss the SP algorithm in detail. For a k-CNF Φ on the variables $V = \{x_1, \dots, x_n\}$ we generally represent truth assignments as maps $\sigma : V \to \{-1, 1\}$, with -1 representing "false" and 1 representing "true". Survey Propagation is an efficient message passing heuristic on the factor graph $G(\Phi)$. The factor graph of Φ is a bipartite graph representation of Φ where each clause and each variable is represented by a vertex. Two vertices are incident if the corresponding variable is contained in the corresponding clause [26].

Before explaining the Survey Propagation heuristic, we explain the simpler Belief Propagation heuristic and emphasize the main extensions later on. To define the messages involved we denote the ordered pair (x, a) with $x \to a$ and similarly (a, x) with $a \to x$ for each $x \in V$ and $a \in N(x)$, where $N(x)$ denotes the neighborhood in the factor graph $G(\Phi)$. The messages are iteratively sent probability distributions $(\mu_{x\to a}(\zeta))_{x\in V_t, a\in N(x), \zeta\in\{-1,1\}}$ over $\{-1, 1\}$. In each iteration messages are sent from variables to adjacent clauses and back. After setting initial messages due to some initialization rule the messages sent are obtained by applying a function to the set of incoming messages at each vertex. Both the initialization and the particular update rules at the vertices are specifying the message passing algorithm. The messages are updated $\omega(n)$ times which may or may not depend on n. A detailed explanation of the Belief Propagation heuristic can be found in [8, p. 519].

It is well known that the Belief Propagation messages on a tree converge after updating the messages two times the depth of the tree to a fixed point. Moreover, in this case for each variable the marginal distribution of the uniform distribution on the set of all satisfying assignments can be computed by the set of the fixed point messages. Since $G(\Phi)$ for constant clauses/variables ratio contains only a small number of short cycles one may expect that on the base of the Belief Propagation messages a good estimate of the marginal distribution of the uniform distribution on the set of all satisfying assignments of Φ could be obtained. Besides the fact that it is not even clear that the messages converge to a fixed point on arbitrary graphs this is of course only a weak heuristic explanation which is refuted by [10]. However, at each decimation step using the Belief Propagation heuristic the Belief Propagation guided decimation algorithm assigns one variable due to the estimated marginal distribution to -1 or 1. Simplifying the formula and running Belief Propagation on the simplified formula and repeating this procedure would lead to a satisfying assignment chosen uniformly at random for sure if the marginals were correct at each decimation step.

Let us now introduce the Survey Propagation heuristic. As mentioned above the geometry of the set of satisfying assignments comes as a collection of tiny well-separated clusters above density $2^k \ln(k)/k$. In that regime a typical solution belongs to a "frozen" cluster. That is all satisfying assignments in such a frozen cluster agree on a linear number of frozen variables. Thus, identifying these frozen variables gives a characterization of the whole cluster. Flipping one of these variables leads to a set of unsatisfied clauses only containing additional frozen variables. Satisfying one of these clauses leads to further unsatisfied clauses of this kind ending up in an avalanche of necessary flippings to obtain a satisfying assignment. This ends only after a linear number of flippings. Given a satisfying assignment with identified frozen variables each satisfying assignment that disagrees on one of these frozen variables has linear distance therefore belonging to a different cluster.

This picture inspires the definition of *covers* as generalized assignments $\sigma \in \{-1, 0, 1\}^n$ such that

- each clause either contains a true literal or two 0 literals and
- for each variable $x \in V$ that is assigned -1 or 1 exists a clause $a \in N(x)$ such that for all $y \in N(a) \setminus \{x\}$ we have $\text{sign}(y, a) \cdot \sigma(y) = -1$.

These two properties mirrors the situation in frozen clusters where assigning a variable to the value 0 indicates that these variable supposes to be free in the corresponding cluster which is obtained by only flipping 0 variables to one of the values -1 or 1. However, implementing the concept of covers, Survey Propagation is a heuristic of computing the marginals over the set of covers by using the Belief Propagation update rules on covers. This leads to the equations given by Figure 1. For a more detailed explanation of the freezing phenomenon we point the reader to [29]. For a deeper discussion on covers we refer to [12].

We are now ready to state the `SPdec` algorithm.

▶ **Algorithm 2.** `SPdec`(Φ)
Input: A k-CNF Φ on $V = \{x_1, \ldots, x_n\}$. *Output:* An assignment $\sigma : V \to \{-1, 1\}$.

0. Let $\Phi_0 = \Phi$.
1. For $t = 0, \ldots, n-1$ do
2. Use SP to compute $\mu^{[\omega]}_{x_{t+1}}(\Phi_t)$.
3. Assign

$$\sigma(x_{t+1}) = \begin{cases} 1 & \text{with probability } \mu^{[\omega]}_{x_{t+1}}(\Phi_t) \\ -1 & \text{with probability } 1 - \mu^{[\omega]}_{x_{t+1}}(\Phi_t). \end{cases} \tag{7}$$

4. Obtain a formula Φ_{t+1} from Φ_t by substituting the value $\sigma(x_{t+1})$ for x_{t+1} and simplifying.
5. Return the assignment σ.

For real numbers $0 \leq x, y \leq 1$ such that $\max\{x, y\} > 0$ we define

$$\psi_\zeta(x,y) = \begin{cases} xy \cdot \Psi(x,y) & \text{if } \zeta = 0 \\ (1-x)y \cdot \Psi(x,y) & \text{if } \zeta = 1 \\ (1-y)x \cdot \Psi(x,y) & \text{if } \zeta = -1 \end{cases}, \qquad \Psi(x,y) = (x + y - xy)^{-1}$$

If $x = y = 0$ set $\psi_0(0) = 0$ and $\psi_{\pm 1}(0) = \frac{1}{2}$. Define for all $x \in V_t, a, b \in N(x), \zeta \in \{-1, 0, 1\}$ and $\ell \geq 0$

$$\mu^{[0]}_{x \to a}(\pm 1) = \frac{1}{2}, \qquad \mu^{[0]}_{x \to a}(0) = 0, \qquad \mu^{[\ell]}_{b \to x}(0) = 1 - \prod_{y \in N(b) \setminus \{x\}} \mu^{[\ell]}_{y \to b}(-\text{sign}(y, b)) \tag{1}$$

$$\pi^{[\ell+1]}_{x \to a}(\pm 1) = \prod_{b \in N(x, \pm 1) \setminus \{a\}} \mu^{[\ell]}_{b \to x}(0) \tag{2}$$

$$\mu^{[\ell+1]}_{x \to a}(\zeta) = (SP(\mu^{[\ell]}))_{x \to a}(\zeta) = \psi_\zeta(\pi^{[\ell]}_{x \to a}(1), \pi^{[\ell]}_{x \to a}(-1)). \tag{3}$$

Let $\omega = \omega(k, r, n) \geq 0$ be any integer-valued function. Define

$$\pi^{[\omega+1]}_x(\Phi_t, \pm 1) = \prod_{b \in N(x, \pm 1)} \mu^{[\omega]}_{b \to x}(0) \tag{4}$$

$$\mu^{[\omega]}_x(\Phi_t, \zeta) = \psi_\zeta(\pi^{[\omega+1]}_x(\Phi_t, 1) \cdot \pi^{[\omega+1]}_x(\Phi_t, -1)) \tag{5}$$

$$\mu^{[\omega]}_x(\Phi_t) = \frac{\mu^{[\omega]}_x(\Phi_t, 1)}{\mu^{[\omega]}_x(\Phi_t, 1) + \mu^{[\omega]}_x(\Phi_t, -1)} = \mu^{[\omega]}_x(\Phi_t, 1) + \frac{1}{2}\mu^{[\omega]}_x(\Phi_t, 0). \tag{6}$$

Figure 1 The Survey Propagation equations that are the Belief Propagation equations on covers.

Let us emphasize that the value $\mu^{[\omega]}_{x_{t+1}}(\Phi_t)$ in Step 2 of `SPdec` is the estimated marginal probability over the set of covers of variable x_{t+1} in the simplified formula to take the value 1 plus one half the estimated marginal probability over the set of covers in the simplified formula to take the value 0. This makes sense since by the heuristic explanation a variable assigned to the value 0 is free to take either value 1 or -1. Thus, our task is to study the SP operator on the decimated formula Φ_t.

3 Proof of Theorem 1

The probabilistic framework used in our analysis of `SPdec` was introduced in [10] for analysing the *Belief Propagation Guided Decimation* algorithm. The most important technique in analysing algorithms on the random formula $\vec{\Phi}$ is the "method of deferred decisions", which traces the dynamics of an algorithm by differential equations, martingales, or Markov chains. It actually applies to algorithms that decide upon the value of a variable x on the basis of the clauses or variables at small bounded distance from x in the factor graph [5]. Unfortunately, the `SPdec` algorithm at step t explores clauses at distance 2ω from x_t where $\omega = \omega(n)$ may tend to infinity with n. Therefore, the "defered decisions" method does not apply and to prove Theorem 1 a fundamentally different approach is needed.

We will basically reduce the analysis of `SPdec` to the problem of analysing the SP operator on the random formula $\vec{\Phi}^t$ that is obtained from $\vec{\Phi}$ by substituting "true" for the first t variables $x_1, \ldots, x_t$ and simplifying (see Theorem 3 below). In the following sections we will prove that this decimated formula has a number of simple to verify quasirandomness properties with very high probability. Finally, we will show that it is possible to trace the Survey Propagation algorithm on a formula Φ enjoying this properties.

Applied to a fixed, non-random formula Φ on $V = \{x_1, \ldots, x_n\}$, `SPdec` yields an assignment $\sigma : V \to \{-1, 1\}$ that may or may not be satisfying. This assignment is random, because `SPdec` itself is randomized. Hence, for any fixed Φ running `SPdec`(Φ) induces a probability distribution β_Φ on $\{-1, 1\}^V$. With $\mathcal{S}(\Phi)$ the set of all satisfying assignments of Φ, the "success probability" of `SPdec` on Φ is just

$$\text{success}(\Phi) = \beta_\Phi(\mathcal{S}(\Phi)). \tag{8}$$

Thus, to establish Theorem 1 we need to show that in the *random* formula

$$\text{success}(\vec{\Phi}) = \beta_{\vec{\Phi}}(s(\vec{\Phi})) = \exp\left(-\Omega(n)\right) \tag{9}$$

is exponentially small w.h.p. To this end, we are going to prove that the measure $\beta_{\vec{\Phi}}$ is "rather close" to the uniform distribution on $\{-1, 1\}^V$ w.h.p., of which $\mathcal{S}(\vec{\Phi})$ constitutes only an exponentially small fraction. However, to prove Theorem 1 we prove that the entropy of the distribution $\beta_{\vec{\Phi}}$ is large. Let us stress that this is not by Moser's entropy compression argument which works up to far smaller clauses/variables ratios [32].

3.1 Lower bounding the entropy

Throughout the paper we let $\rho_k = (1 + \varepsilon_k)\ln(k)$ where $(\varepsilon_k)_{k\geq 3}$ is the sequence promised by Theorem 1 and let $r = 2^k\rho$ where $\rho \geq \rho_k$.

For a number $\delta > 0$ and an index $i > t$ we say that x_i is (δ, t)-*biased* if

$$\left|\mu_{x_i}^{[\omega]}(\Phi^t, 1) - \frac{1}{2}\left(1 - \mu_{x_i}^{[\omega]}(\Phi^t, 0)\right)\right| > \delta. \tag{10}$$

Moreover Φ is (δ, t)-*balanced* if no more than $\delta(n - t)$ variables are (δ, t)-biased.

If $\vec{\Phi}$ is (δ, t)-balanced, then by the basic symmetry properties of $\vec{\Phi}$ the probability that x_{t+1} is (δ, t)-biased is bounded by δ. Furthermore, given that x_{t+1} is not (δ, t)-biased, the probability that `SPdec` will set it to "true" lies in the interval $[\frac{1}{2} - \delta, \frac{1}{2} + \delta]$. Consequently,

$$\left|\frac{1}{2} - \mathbb{P}\left[\sigma(x_{t+1}) = 1 | \vec{\Phi} \text{ is } (\delta, t)\text{-balanced}\right]\right| \leq 2\delta. \tag{11}$$

Thus, the smaller δ the closer $\sigma(x_{t+1})$ comes to being uniformly distributed. Hence, if (δ, t)-balancedness holds for all t with a "small" δ, then β_Φ will be close to the uniform distribution on $\{-1, 1\}^V$.

To put this observation to work, let $\theta = 1 - t/n$ be the fraction of unassigned variables and define

$$\delta_t = \exp(-c\theta k), \qquad \Delta_t = \sum_{s=1}^{t} \delta_t \qquad \text{and} \qquad \hat{t} = \left(1 - \frac{\ln(\rho)}{c^2 k}\right) n, \tag{12}$$

where $c > 0$ is a small enough absolute constant.

The following result provides the key estimate by providing that at any time t up to $\hat{t}$ with sufficiently high probability $\vec{\Phi}$ is (δ_t, t)-balanced with a sufficiently small δ_t to finally prove Theorem 1.

► **Proposition 3.** *For any k, r satisfying $2^k\rho_k/k < r \leq 2k\ln 2$ there is $\xi = \xi(k, r) \in [0, \frac{1}{k}]$ so that for n large enough the following holds. For any $0 \leq t \leq \hat{t}$ we have*

$$\Pr\left[\vec{\Phi} \textit{ is } (\delta_t, t)\textit{-balanced}\right] \geq 1 - \exp\left[-3\xi n - 10\Delta_t\right]. \tag{13}$$

3.2 Tracing the Survey Propagation Operator

To establish Proposition 3 we have to prove that $\vec{\Phi}$ is (δ_t, t)-balanced with probability very close to one. Thus, our task is to study the SP operator defined in (1) to (3) on $\vec{\Phi}^t$. Roughly speaking, Proposition 3 asserts that with probability very close to one, most of the messages $\mu^{[\ell]}_{x\to a}(\pm 1)$ are close to $\frac{1}{2}(1-\mu^{[\ell]}_{x\to a}(0))$. To obtain this bound, we are going to proceed in two steps: we will exhibit a small number of *quasirandomness properties* and show that these hold in $\vec{\Phi}^t$ with the required probability. Then, we prove that *deterministically* any formula that has these properties is (δ_t, t)-balanced.

3.2.1 The "typical" value of $\pi^{[\ell]}_{x\to a}(\zeta)$

First of all recall that the messages sent from a variable x to a clause $a \in N(x)$ are obtained by

$$\psi_\zeta(\pi^{[\ell]}_{x\to a}(1), \pi^{[\ell]}_{x\to a}(-1)) \qquad \text{for } \zeta \in \{-1, 0, 1\}. \tag{14}$$

This in mind, we claim a strong statement that both $\pi^{[\ell]}_{x\to a}(1)$ and $\pi^{[\ell]}_{x\to a}(-1)$ are very close to a "typical" value $\pi[\ell]$ for most of the variables $x \in V_t$ and clauses $a \in N(x)$ at any iteration step ℓ under the assumption that the set of biased variables is small at time $\ell - 1$. Assuming that

$$\pi^{[\ell]}_{x\to a}(1) = \pi^{[\ell]}_{x\to a}(-1) = \pi[\ell]$$

we of course obtain unbiased messages by

$$\mu^{[\ell]}_{x\to a}(\pm 1) = \psi_1(\pi[\ell]) = \psi_{-1}(\pi[\ell]) = \frac{1}{2}(1 - \mu^{[\ell]}_{x\to a}(0)).$$

The products $\pi^{[\ell]}_{x\to a}(\zeta)$ are nothing else but the product of the messages

$$\mu^{[\ell-1]}_{b\to x}(0) = 1 - \prod_{y\in N(b)\setminus\{x\}} \mu^{[\ell-1]}_{y\to b}(-\mathrm{sign}(y, b))$$

sent from all clauses $b \in N(x, \zeta) \setminus \{a\}$ to x. Therefore, we define inductively $0 \le \pi[\ell] \le 1$ to be the product of this kind over a "typical" neighborhood. The term "typical" refers to the expected number of clauses of all lengths that contain at most one additional biased variable. Focusing on those clauses will suffice to get the tightness result of the biases. Moreover, we assume that all of the messages $\mu^{[\ell-1]}_{y\to b}(-\mathrm{sign}(y, b))$ sent from variables to clauses in such a typical neighborhood are $\psi_{\mathrm{sign}(y,b)}(\pi[\ell-1], \pi[\ell-1])$ which is claimed to be a good estimation of most of the messages sent at time $\ell - 1$. Additionally, define $\tau[\ell] = (1 - \psi_0(\pi[\ell]))$ as the estimate of the sum $\mu^{[\ell]}_{x\to a}(1) + \mu^{[\ell]}_{x\to a}(-1)$. Let us emphasize that there is no "unique" $\pi[\ell]$ and the way it is obtained in the following is in some sense the canonical and convenient choice to sufficiently bound the biases for most of the messages.

Generally, let $T \subset V_t$ and $x \in V_t$. Then the expected number of clauses of length j that contain x and at most one other variable from the set T is asymptotically

$$\mu_{j,\le 1}(T) = 2^j \rho \cdot \Pr\left[\mathrm{Bin}(k-1, \theta) = j - 1\right] \cdot \Pr\left[\mathrm{Bin}\left(j - 1, \frac{|T|}{\theta n}\right) < 2\right]. \tag{15}$$

Indeed, the expected number of clauses of $\vec{\Phi}$ that x appears in equals $km/n = kr = 2^k\rho$. Furthermore, each of these gives rise to a clause of length j in $\vec{\Phi}^t$ iff exactly $j - 1$ among the other $k - 1$ variables in the clauses are from V_t while the $k - j$ remaining variables are in $V \setminus V_t$ and occur with negative signs. (If one of them had a positive sign, the clause would have been satisfied by setting the corresponding variable to true. It would thus not be present in $\vec{\Phi}^t$ anymore.) Moreover, at most one of the $j-1$ remaining variables is allowed to be from the set T. The fraction of variables in T in V_t equals $\frac{|T|}{\theta n}$. Finally, since x appears with a random sign in each of these clauses the expected number of clauses of length j that contain x and at most one other variable from the set T is asymptotically $\mu_{j,\leq 1}(t)/2$.

Additionally let $0 \leq p \leq 1$ and define

$$\tau(p) = 1 - \psi_0\left(p\right) \qquad \text{and} \qquad \pi(T,p) = \prod_{j=0.1\theta k}^{10\theta k} \left(1 - \left(2/\tau(p)\right)^{-j+1}\right)^{\mu_{j,\leq 1}(T)/2}. \tag{16}$$

Moreover, let

$$\Pi(T,p) = \sum_{j=0.1\theta k}^{10\theta k} \frac{\mu_{j,\leq 1}(T)}{2} \cdot \left(2/\tau(p)\right)^{-j+1}$$

be the approximated absolute value of the logarithm of $\pi(T,p)$.

For a fixed variable $x \in V_t$ the expected number of clauses that contain more than one additional variable from a "small" set T for a "typical" clause length $0.1\theta k \leq j \leq 10\theta k$ is very close to the expected number of all clauses of that given length. Thus, the actual size of T will influence $\pi(T,p)$ but this impact is small if T is small and the following bounds on $\pi(T,p)$ can be achieved.

▶ **Lemma 4.** *Let $T \subset V_t$ of size $|T| \leq \delta\theta n$ and $0 \leq p \leq 2\exp(-\rho)$. Then $\exp\left(-2\rho\right) \leq \pi(T,p) \leq 2\exp\left(-\rho\right)$.*

3.2.2 Bias

First of all let us define the bias not only for the 1 and -1 messages but also for the 0 messages. Hence, for $\ell \geq 0, x \in V_t$ and $a \in N(x)$ let

$$\Delta_{x\to a}^{[\ell]} = \mu_{x\to a}^{[\ell]}(1) - \frac{1}{2}\left(1 - \mu_{x\to a}^{[\ell]}(0)\right) \qquad \text{and} \tag{17}$$

$$E_{x\to a}^{[\ell]} = \frac{1}{2}\left(\mu_{x\to a}^{[\ell]}(0) - \psi_0(\pi[\ell])\right). \tag{18}$$

We say that $x \in V_t$ is *ℓ-biased* if

$$\max_{a\in N(x)} |\Delta_{x\to a}^{[\ell]}| > 0.1\delta \qquad \text{or} \qquad \max_{a\in N(x)} |E_{x\to a}^{[\ell]}| > 0.1\delta\pi[\ell] \tag{19}$$

and *ℓ-weighted* if

$$\max_{a\in N(x)} |E_{x\to a}^{[\ell]}| > 10\pi[\ell]. \tag{20}$$

Let $B[\ell]$ be the set of all ℓ-biased variables and $B'[\ell]$ be the set of all ℓ-weighted variables. Obviously, by definition, we have $B'[\ell] \subset B'[\ell]$.

Writing $\mu_{x\to a}^{[\ell]}(\mathrm{sign}(x,a))$ in terms of the biases we obtain

$$\begin{aligned}\mu_{x\to a}^{[\ell]}(\mathrm{sign}(x,a)) &= \frac{1}{2}(1 - \psi_0(\pi[\ell])) - \left(E_{x\to a}^{[\ell]} + \mathrm{sign}(x,a)\Delta_{x\to a}^{[\ell]}\right) \\ &= \tau[\ell]/2 - \left(E_{x\to a}^{[\ell]} + \mathrm{sign}(x,a)\Delta_{x\to a}^{[\ell]}\right)\end{aligned} \tag{21}$$

We are going to prove that $|\Delta_{x\to a}^{[\ell]}|$ and $|E_{x\to a}^{[\ell]}|$ are small for most x and $a \in N(x)$. That is, given the $\Delta_{x\to a}^{[\ell]}$ and $E_{x\to a}^{[\ell]}$ we need to prove that the biases $\Delta_{x\to a}^{[\ell+1]}$ and $E_{x\to a}^{[\ell+1]}$ do not "blow up". The proof is by induction where the hypothesis is that at most $\delta_t \theta n$ variables are ℓ-biased and at most $\delta^2 \theta n$ variables are ℓ-weighted and our goal is to show that the same holds true for $\ell + 1$.

3.2.3 The quasirandomness property

We will now exhibit a few simple quasirandomness properties that $\vec{\Phi}^t$ is very likely to possess. Based only on these graph properties we identify potentially ℓ-biased or ℓ-weighted variables. In turn, we prove that variables in the complement of these sets are surely not ℓ-biased resp. ℓ-weighted. Moreover, we show that these sets are small enough with sufficiently high probability.

To state the quasirandomness properties, fix a k-CNF Φ. Let Φ^t denote the CNF obtained from Φ by substituting "true" for $x_1, \ldots, x_t$ and simplifying ($1 \le t \le n$). Let $V_t = \{x_{t+1}, \ldots, x_n\}$ be the set of variables of Φ^t. Let $\delta = \delta_t$. With $c > 0$ we let $k_1 = \sqrt{c}\theta k$. For a variable $x \in V_t, \zeta \in \{1, -1\}$ and a set $T \subset V_t$ let

$$
\begin{aligned}
\mathcal{N}(x,\zeta) &= \{b \in N(x,\zeta) : 0.1\theta k \le |N(b)| \le 10\theta k\}, \\
\mathcal{N}_{\le 1}(x,T,\zeta) &= \{b \in \mathcal{N}(x,\zeta) : |N(b) \cap T \setminus \{x\}| \le 1\}, \\
\mathcal{N}_i(x,T,\zeta) &= \{b \in \mathcal{N}(x,\zeta) : |N(b) \cap T \setminus \{x\}| = i\} \text{ for } i \in \{0,1\}, \\
N_1(x,T,\zeta) &= \{b \in N(x,\zeta) : |N(b) \setminus T| \ge k_1 \wedge |N(b) \cap T \setminus \{x\}| = 1\}, \\
N_{>1}(x,T,\zeta) &= \{b \in N(x,\zeta) : |N(b) \setminus T| \ge k_1 \wedge |N(b) \cap T \setminus \{x\}| > 1\}.
\end{aligned}
$$

Thus, $\mathcal{N}_{\le 1}(x,T,\zeta)$ is the set of all clauses a that contain x with $\mathrm{sign}(x,a) = \zeta$ (which may or may not be in T) and at most one other variable from T. In addition, there is a condition on the length $|N(b)|$ of the clauses b in the decimated formula Φ^t. Having assigned the first t variables, we should "expect" the average clause length to be θk. The sets $\mathcal{N}_i(x,T,\zeta)$ are a partition of $\mathcal{N}_{\le 1}(x,T,\zeta)$ separating clauses that contain exactly one additional variable from $T \setminus \{x\}$ and clauses that contain none.

Q1. No more than $10\delta\theta n$ variables occur in clauses of length less than $\theta k/10$ or greater than $10\theta k$ in Φ_t. Moreover, there are at most $10^{-4}\delta\theta n$ variables $x \in V_T$ such that
$$(\theta k)^3 \delta \cdot \sum_{b \in N(x,\zeta)} 2^{-|N(b)|} \quad > \quad 1.$$

Q2. For any set $T \subset V_t$ of size $|T| \le s\theta n$ such that $\delta^5 \le s \le 10\delta$ and any $p \in (0,1]$ there are at most $10^{-3}\delta^2\theta n$ variables x such that for one $\zeta \in \{-1,1\}$ either
$$
\begin{aligned}
\left| \Pi(T,p) - \sum_{b \in \mathcal{N}_{\le 1}(x,T,\zeta)} (2/\tau(p))^{1-|N(b)|} \right| &> 2\delta/1000 \qquad \text{or} \\
\sum_{b \in \mathcal{N}_1(x,T,\zeta)} 2^{-|N(b)|} &> 10^4 \rho\theta k s \qquad \text{or} \\
\sum_{b \in \mathcal{N}_{\le 1}(x,T,\zeta)} 2^{-|N(b)|} &> 10^4 \rho.
\end{aligned}
$$

Q3. If $T \subset V_t$ has size $|T| \le \delta\theta n$, then there are no more than $10^{-4}\delta\theta n$ variables x such that at least for one $\zeta \in \{-1,1\}$
$$\sum_{b \in N_{>1}(x,T,\zeta)} 2^{|N(b) \cap T \setminus \{x\}| - |N(b)|} > \delta/(\theta k).$$

Q4. For any $0.01 \le z \le 1$ and any set $T \subset V_t$ of size $|T| \le 100\delta\theta n$ we have
$$\sum_{b : |N(b) \cap T| \ge z|N(b)|} |N(b)| \le \frac{1.01}{z}|T| + 10^{-4}\delta\theta n.$$

Q5. For any set $T \subset V_t$ of size $|T| \le 10\delta\theta n$, any $p \in (0,1]$ and any $\zeta \in \{-1,1\}$ the linear operator $\Lambda(T,\mu,\zeta) : \mathbb{R}^{V_t} \to \mathbb{R}^{V_t}$,

$$\Gamma = (\Gamma_y)_{y \in V_t} \mapsto \left\{ \sum_{b \in \mathcal{N}_{\le 1}(x,T,\zeta)} \sum_{y \in N(b)\setminus\{x\}} (2/\tau(p))^{-|N(b)|} \operatorname{sign}(y,b)\Gamma_y \right\}$$

has norm $\| \ \Lambda(T,\mu,\zeta) \ \|_\square \le \ \delta^4\theta n$, where for a real $b \times a$ matrix Λ we let $\|\Lambda\|_\square = \max_{\zeta \in \mathbb{R}^a\setminus\{0\}} \frac{\|\Lambda\zeta\|_1}{\|\zeta\|_\infty}$.

▶ **Definition 5.** Let $\delta > 0$. We say that Φ is (δ,t)-*quasirandom* if ***Q0-Q5*** are satisfied.

Condition **Q0** simply bounds the number of redundant clauses and the number of variables of very high degree; it is well-known to hold for random k-CNFs w.h.p. Apart from a bound on the number of very short/very long clauses, **Q1** provides a bound on the "weight" of clauses in which variables $x \in V_t$ typically occur, where the weight of a clause b is $2^{-|N(b)|}$. Moreover, **Q2** and **Q3** provide that there is no small set T for which the total weight of the clauses touching that set is very big. In addition, **Q2** (essentially) requires that for most variables x the weights of the clauses where x occurs positively/negatively should approximately cancel. Further, **Q4** provides a bound on the lengths of clauses that contain many variables from a small set T. Finally, the most important condition is **Q5**, providing a bound on the cut norm of a signed, weighted matrix, representation of Φ^t.

▶ **Proposition 6.** *There is a sequence $(\varepsilon_k)_{k\ge3}$ with $\lim_{k\to\infty}\varepsilon_k = 0$ such that for any k, r satisfying $2^k(1+\varepsilon_k)\ln(k)/k \le r \le 2^k \ln 2$ there is $\xi = \xi(k,r) \in [0, \frac{1}{k}]$ so that for n large and $\delta_t, \hat{t}$ as in (12) for any $1 \le t \le \hat{t}$ we have*

$$P\left[\Phi \textit{ is } (\delta_t,t)\textit{-quasirandom}\right] \ge 1 - \exp\left(-10(\xi n + \Delta_t)\right)$$

▶ **Theorem 7.** *There is a sequence $(\varepsilon_k)_{k\ge3}$ with $\lim_{k\to\infty}\varepsilon_k = 0$ such that for any k, r satisfying $2^k(1+\varepsilon_k)\ln(k)/k \le r \le 2^k \ln 2$ and n sufficiently large the following is true.*

Let Φ be a k-CNF with n variables and m clauses that is (δ_t,t)-quasirandom for some $1 \le t \le \hat{t}$. Then Φ is (δ_t,t)- balanced.

The proof of Proposition 6 is a necessary evil: it is long, complicated and based on standard arguments. Theorem 7 together with Proposition 6 yields Proposition 3.

3.2.4 Setting up the induction

To prove Theorem 7 we proceed by induction over ℓ. In particular we define sets $T[\ell]$ and $T'[\ell]$ that contain variables that are potentially ℓ-biased or ℓ-weighted only depending on the graph structure and the size of the sets $T[\ell-1]$ and $T'[\ell-1]$. The exact definition of the sets $T[\ell]$ and $T'[\ell]$ which inspired the quasirandomness properties are omitted in this extended abstract. It actually will turn out that $T[\ell] \subset B\ell$ and $T'[\ell] \subset B'\ell$. Since we are going to trace the SP operator on Φ^t iterated from the initial set of messages $\mu^{[0]}_{x\to a}(\pm1) = \frac{1}{2}$ and $\mu^{[0]}_{x\to a}(0) = 0$ for all $x \in V_t$ and $a \in N(x)$ we set $T[\ell] = T'[\ell] = \emptyset$ and $\pi[0] = 0$ such that $\tau[0] = 1$. Now we define inductively

$$\pi[\ell+1] = \pi\left(T[\ell],\pi[\ell]\right), \quad \Pi[\ell+1] = \Pi\left(T[\ell],\pi[\ell]\right) \quad \text{and} \quad \tau[\ell+1] = \tau\left(\pi[\ell+1]\right).$$

▶ **Proposition 8.** *Assume that $\pi[\ell] \le 2\exp(-\rho)$. We have $B[\ell] \subset T[\ell]$ and $B'[\ell] \subset T'[\ell]$ for all $\ell \ge 0$.*

Furthermore, we establish the following bounds on the size of $T[\ell]$ and $T'[\ell]$. Since the sets are defined by graph properties independent from the actual state of the algorithm the quasirandomness properties suffice to obtain

▶ **Proposition 9.** *If Φ is (δ_t, t)-quasirandom, we have $T[\ell] < \delta\theta n, T'[\ell] < \delta^2\theta n$ and $\pi[\ell] \le 2\exp(-\rho)$ for all $\ell \ge 0$.*

Finally, let us give an idea how this is actually proved. We aim to prove that for most variables $x \in V_t$ for all $a \in N(x)$ simultaneously for both $\zeta \in \{-1,1\}$ the values $\pi^{[\ell]}_{x\to a}(\zeta)$ are close to a typical value which is estimated by $\pi[\ell]$ for each iteration. Let us define for $x \in V_t, a \in N(x)$ and $\zeta \in \{1,-1\}$

$$P^{[\ell+1]}_{\le 1}(x \to a,\zeta) = \prod_{b\in\mathcal{N}_{\le 1}(x,T[\ell],\zeta)\setminus\{a\}} \mu^{[\ell]}_{b\to x}(0)$$
$$P^{[\ell+1]}_{>1}(x \to a,\zeta) = \prod_{b\in N(x,\zeta)\setminus(\{a\}\cup\mathcal{N}_{\le 1}(x,T[\ell],\zeta))} \mu^{[\ell]}_{b\to x}(0).$$

We obtain

$$\pi^{[\ell]}_{x\to a}(\zeta) = P^{[\ell]}_{\le 1}(x\to a,\zeta)\cdot P^{[\ell]}_{>1}(x\to a,\zeta). \tag{22}$$

We show that the first factor representing the product over messages sent by clauses of typical length (regarding the decimation time t) and exposed to at most one additional variable from $T[\ell]$ is close to $\pi[\ell+1]$ simultaneously for $\zeta \in \{-1,1\}$ for all variables $x \in V \setminus T'[\ell+1]$ and all $a \in N(x)$. Additionally, we prove that the second factor representing the product over messages sent by clauses of atypical length or exposed to at least two additional variables from $T[\ell]$ is close to one simultaneously for $\zeta \in \{-1,1\}$ for all variables $x \in V \setminus T[\ell+1]$ and all $a \in N(x)$.

Acknowledgements I thank my supervisor Amin Coja-Oghlan for supportive conversations and helpful comments on the final version of this paper.

References

1 D. Achlioptas, P. Beam, M. Molloy Exponential bounds for DPLL below the satisfiability threshold. Proc. SODA (2004) .
2 D. Achlioptas, A. Coja-Oghlan: Algorithmic barriers from phase transitions. Proc. 49th FOCS (2008) 793–802.
3 D. Achlioptas, A. Coja-Oghlan, F. Ricci-Tersenghi: On the solution-space geometry of random constraint satisfaction problems. Random Structures and Algorithms **38** (2011) 251–268.
4 D. Achlioptas, Y. Peres: The threshold for random k-SAT is $2^k \ln 2 - O(k)$. Journal of the AMS **17** (2004) 947–973.
5 D. Achlioptas, G. Sorkin: Optimal myopic algorithms for random 3-SAT. Proc.41st FOCS (2000) 590–600.
6 A. Braunstein, M. Mézard, R. Zecchina: Survey propagation: an algorithm for satisfiability. Random Structures and Algorithms **27** (2005) 201–226.
7 A. Braunstein, R. Mulet, A. Pagnani, M. Weigt, R. Zecchina: Polynomial iterative algorithms for coloring and analysing random graphs. Phys. Rev. E **68** (2003) 036702.
8 A. Braunstein, R. Zecchina: Survey and Belief Propagation on Random k-SAT. Lecture Notes in Computer Science, Springer Berlin (2003).

9 P. Cheeseman, B. Kanefsky, W. Taylor: Where the *really* hard problems are. Proc. IJCAI (1991) 331–337.

10 A. Coja-Oghlan: On belief propagation guided decimation for random k-SAT. Proc. 22nd SODA (2011) 957–966.

11 A. Coja-Oghlan: A better algorithm for random k-SAT. SIAM J. Computing **39** (2010) 2823–2864.

12 A. Coja-Oghlan: The asymptotic k-SAT threshold. Proc. 46th STOC (2014) 804–813.

13 A. Coja-Oghlan, A. Y. Pachon-Pinzon: The decimation process in random k-SAT. SIAM Journal on Discrete Mathematics **26** (2012) 1471–1509.

14 J. Ding, A. Sly, N. Sun: Proof of the satisfiability conjecture for large k. Proc. 47th STOC (2015) 59–68.

15 U. Feige, E. Mossel, D. Vilenchik: Complete convergence of message passing algorithms for some satisfiability problems. Theory of Computing **9** (2013) 617–651.

16 E. Friedgut: Sharp thresholds of graph properties, and the k-SAT problem. J. AMS **12** (1999) 1017–1054.

17 A. Frieze, S. Suen: Analysis of two simple heuristics on a random instance of k-SAT. Journal of Algorithms **20** (1996) 312–355.

18 D. Gamarnik, M. Sudan: Limits of local algorithms over sparse random graphs. Proc. 5th ITCS (2014) 369–376.

19 D. Gamarnik, M. Sudan: Performance of the Survey Propagation-guided decimation algorithm for the random NAE-K-SAT problem. arXiv 1402.0052 (2014).

20 S. Janson, T. Łuczak, A. Ruciński: Random Graphs, Wiley 2000.

21 L. Kroc, A. Sabharwal, B. Selman: Message-passing and local heuristics as decimation strategies for satisfiability. Proc 24th SAC (2009) 1408–1414.

22 F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, L. Zdeborova: Gibbs states and the set of solutions of random constraint satisfaction problems. Proc. National Academy of Sciences **104** (2007) 10318–10323.

23 R. Marino, G. Parisi, F. Ricci-Tersenghi: The backtracking Survey Propagation algorithm for solving random K-SAT problems. arXiv 1508.05117 (2015).

24 S. Mertens, M. Mézard, R. Zecchina: Threshold values of random K-SAT from the cavity method. Random Struct. Alg. **28** (2006) 340–373.

25 S. Mertens, C. Moore: The Nature of Computation. Oxford University Press 2011.

26 M. Mézard, A. Montanari: Information, physics and computation. Oxford University Press 2009.

27 M. Mézard, G. Parisi, R. Zecchina: Analytic and algorithmic solution of random satisfiability problems. Science **297** (2002) 812–815.

28 D. Mitchell, B. Selman, H. Levesque: Hard and easy distribution of SAT problems. Proc. 10th AAAI (1992) 459–465.

29 M. Molloy: The freezing threshold for k-colourings of a random graph. Proc. 43rd STOC (2012) 921–930.

30 R. Monasson: A Generating Function Method for the Average-Case Analysis of DPLL. Proc. APPROX-RANDOM (2005) 402–413.

31 A. Montanari, F. Ricci-Tersenghi, G. Semerjian: Solving constraint satisfaction problems through Belief Propagation-guided decimation. Proc. 45th Allerton (2007).

32 R. Moser A constructive proof of the Lovász local lemma. Proc. 41st STOC (2009) 343–350.

33 J. Pearl: Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

34 F. Ricci-Tersenghi, G. Semerjian: On the cavity method for decimated random constraint satisfaction problems and the analysis of belief propagation guided decimation algorithms. J. Stat. Mech. (2009) P09001.