SIAM J. COMPUT.
Vol. 39, No. 7, pp. 2823–2864

A BETTER ALGORITHM FOR RANDOM k-SAT*

AMIN COJA-OGHLAN†

Abstract. Let $\mathbf{\Phi}$ be a uniformly distributed random k-SAT formula with n variables and m clauses. We present a polynomial time algorithm that finds a satisfying assignment of $\mathbf{\Phi}$ with high probability for constraint densities $m/n < (1-\varepsilon_k)2^k \ln(k)/k$, where $\varepsilon_k \to 0$. Previously no efficient algorithm was known to find satisfying assignments with a nonvanishing probability beyond $m/n = 1.817 \cdot 2^k/k$ [A. Frieze and S. Suen, *J. Algorithms*, 20 (1996), pp. 312–355].

Key words. random structures, efficient algorithms, phase transitions, k-SAT

AMS subject classifications. 68Q87, 68W40

DOI. 10.1137/09076516X

1. Introduction.

The k-SAT problem is well known to be NP-hard for $k \geq 3$. This indicates that no algorithm can solve *all* possible inputs efficiently. Therefore, there has been a significant amount of research on *heuristics* for k-SAT, i.e., algorithms that solve "most" inputs efficiently (where the meaning of "most" varies). While some heuristics for k-SAT are very sophisticated, virtually all of them are based on (at least) one of the following basic paradigms.

Pure literal rule. If a variable x occurs only positively (resp., negatively) in the formula, set it to true (resp., false). Simplify the formula by substituting the newly assigned value for x and repeat.

Unit clause propagation. If there is a clause that contains only a single literal ("unit clause"), then set the underlying variable so as to satisfy this clause. Then simplify the formula and repeat.

Walksat. Initially pick a random assignment. Then repeat the following. While there is an unsatisfied clause, pick one at random, pick a variable occurring in the chosen clause randomly, and flip its value.

Backtracking. Assign a variable x, simplify the formula, and recurse. If the recursion fails to find a satisfying assignment, assign x the opposite value and recurse.

Heuristics based on these paradigms can be surprisingly successful on certain types of inputs (e.g., [10, 16]). However, it remains remarkably simple to generate formulas that seem to elude all known algorithms/heuristics. Indeed, the simplest conceivable type of *random* instance does the trick: let $\mathbf{\Phi}$ denote a k-SAT formula over the variable set $V = \{x_1, \ldots, x_n\}$ that is obtained by choosing m clauses uniformly at random and independently from the set of all $(2n)^k$ possible clauses. Then for a large regime of constraint densities m/n satisfying assignments are known to exist due to nonconstructive arguments, but no algorithm is known to find one in subexponential time with a nonvanishing probability.

1.1. Main result.

To be precise, keeping k fixed and letting $m = \lceil rn \rceil$ for a fixed $r > 0$, we say that $\mathbf{\Phi}$ has some property *with high probability* ("w.h.p.") if the

*Received by the editors July 15, 2009; accepted for publication (in revised form) January 20, 2010; published electronically May 5, 2010. This work was supported by EPSRC grant EP/G039070/1. An extended abstract version of this work appeared in *Proceedings of the International Conference on Automata, Languages and Programming* (ICALP 2009), Springer-Verlag, Berlin, pp. 292–303.
http://www.siam.org/journals/sicomp/39-7/76516.html

†Mathematics Institute and Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK (A.Coja-Oghlan@warwick.ac.uk).

probability that the property holds tends to 1 as $n \to \infty$. Via the (nonalgorithmic) second moment method and the sharp threshold theorem [3, 4, 14], it can be shown that $\mathbf{\Phi}$ has a satisfying assignment w.h.p. if $m/n < (1-\varepsilon_k)2^k \ln 2$. Here ε_k is independent of n but tends to 0 for large k. On the other hand, a first moment argument shows that no satisfying assignment exists w.h.p. if $m/n > 2^k \ln 2$. In summary, the threshold for $\mathbf{\Phi}$ being satisfiable is asymptotically $2^k \ln 2$.

Yet for densities m/n beyond $\mathrm{e} \cdot 2^k/k$ no algorithm has been known to find a satisfying assignment in polynomial time with a probability that remains bounded away from 0 for large n—neither on the basis of a rigorous analysis nor on the basis of experimental or other evidence. In fact, many algorithms, including Pure Literal, Unit Clause, and DPLL, are known to either fail or exhibit an exponential running time beyond $c \cdot 2^k/k$ for certain constants $c < \mathrm{e}$. There is experimental evidence that the same is true of Walksat. Indeed, devising an algorithm to solve random formulas with a nonvanishing probability for densities m/n up to $2^k\omega(k)/k$ for *any* (howsoever slowly growing) $\omega(k) \to \infty$ has been a prominent open problem [3, 4, 8, 22], which the following theorem resolves.

THEOREM 1.1. *There exist a sequence $\varepsilon_k \to 0$ and a polynomial time algorithm* `Fix` *such that* `Fix` *applied to a random formula $\mathbf{\Phi}$ with $m/n \leq (1-\varepsilon_k)2^k \ln(k)/k$ outputs a satisfying assignment w.h.p.*

`Fix` is a combinatorial, local search-type algorithm. It can be implemented to run in time $O((n+m)^{3/2})$.

The recent paper [2] provides evidence that beyond density $m/n = 2^k \ln(k)/k$ the problem of finding a satisfying assignment becomes conceptually significantly more difficult (to say the least). To explain this, we need to discuss a concept that originates in statistical physics.

1.2. A digression: Replica symmetry breaking. For the last decade random k-SAT has been studied by statistical physicists using sophisticated and insightful, but mathematically highly nonrigorous, techniques from the theory of spin glasses. Their results suggest that below the threshold density $2^k \ln 2$ for the existence of satisfying assignments various other phase transitions take place that affect the performance of algorithms.

To us the most important one is the *dynamic replica symmetry breaking* (dRSB) transition. Let $S(\mathbf{\Phi}) \subset \{0,1\}^V$ be the set of all satisfying assignments of the random formula $\mathbf{\Phi}$. We turn $S(\mathbf{\Phi})$ into a graph by considering $\sigma, \tau \in S(\mathbf{\Phi})$ adjacent if their Hamming distance equals 1. Very roughly speaking, according to the dRSB hypothesis, there is a density r_{RSB} such that for $m/n < r_{RSB}$ the correlations that shape the set $S(\mathbf{\Phi})$ are purely local, whereas for densities $m/n > r_{RSB}$ long-range correlations occur. Furthermore, $r_{RSB} \sim 2^k \ln(k)/k$ as k gets large.

Confirming and elaborating on this hypothesis, we recently established a good part of the dRSB phenomenon rigorously [2]. In particular, we proved that there is a sequence $\varepsilon_k \to 0$ such that for $m/n > (1+\varepsilon_k)2^k \ln(k)/k$ the values that the solutions $\sigma \in S(\mathbf{\Phi})$ assign to the variables are mutually heavily correlated in the following sense. Let us call a variable x *frozen* in a satisfying assignment σ if any satisfying assignment τ such that $\sigma(x) \neq \tau(x)$ is at Hamming distance $\Omega(n)$ from σ. Then for $m/n > (1+\varepsilon_k)2^k \ln(k)/k$ in all but a $o(1)$-fraction of all solutions $\sigma \in S(\mathbf{\Phi})$, all but an ε_k-fraction of the variables are frozen w.h.p., where $\varepsilon_k \to 0$.

This suggests that on random formulas with density $m/n > (1+\varepsilon_k)2^k \ln(k)/k$ local search algorithms are unlikely to succeed. To see this, think of the *factor graph*, whose vertices are the variables and the clauses, and where a variable is adjacent to all clauses in which it occurs. Then a local search algorithm assigns a value to a variable x on the basis of the values of the variables that have distance $O(1)$ from x in the factor graph. But in the random formula $\mathbf{\Phi}$ with $m/n > (1+\varepsilon_k)2^k \ln(k)/k$, assigning one variable x is likely to impose constraints on the values that can be assigned to variables at distance $\Omega(\ln n)$ from x. A local search algorithm is unable to catch these constraints. Unfortunately, virtually all known k-SAT algorithms are local.

The above discussion applies to "large" values of k (say, $k \geq 10$). In fact, nonrigorous arguments as well as experimental evidence [5] suggest that the picture is quite different and rather more complicated for "small" k (say, $k = 3$). In this case the various phenomena that occur at (or very near) the point $2^k \ln(k)/k$ for $k \geq 10$ appear to happen at vastly different points in the satisfiable regime. To keep matters as simple as possible we focus on "large" k in this paper. In particular, no attempt has been made to derive explicit bounds on the numbers ε_k in Theorem 1.1 for "small" k. Indeed, `Fix` is designed so as to allow for as easy an analysis as possible for general k rather than to excel for small k. Nevertheless, it would be interesting to see how the ideas behind `Fix` can be used to obtain an improved algorithm for small k as well.

In summary, the dRSB picture leads to the question of whether `Fix` marks the end of the algorithmic road for random k-SAT, up to the precise value of ε_k.

1.3. Related work. Quite a few papers deal with efficient algorithms for random k-SAT, contributing either rigorous results, nonrigorous evidence based on physics arguments, or experimental evidence. Table 1.1 summarizes the part of this work that is most relevant to us. The best rigorous result (prior to this work) is due to Frieze and Suen [15]. They proved that "SCB" succeeds for densities $\eta_k 2^k/k$, where η_k increases to 1.817 as $k \to \infty$. SCB can be considered a (restricted) DPLL algorithm. It combines the shortest clause rule, which is a generalization of Unit Clause, with (very limited) backtracking. Conversely, there is a constant $c > 0$ such that DPLL-type algorithms exhibit an exponential running time w.h.p. for densities beyond $c \cdot 2^k/k$ for large k [1].

TABLE 1.1
Algorithms for random k-SAT.

Algorithm	Density $m/n < \cdots$	Success probability	Ref., year
Pure Literal	$o(1)$ as $k \to \infty$	w.h.p.	[20], 2008
Walksat, rigorous	$\frac{1}{6} \cdot 2^k/k^2$	w.h.p.	[9], 2009
Walksat, nonrigorous	$2^k/k$	w.h.p.	[23], 2003
Unit Clause	$\frac{1}{2}\left(\frac{k-1}{k-2}\right)^{k-2} \cdot \frac{2^k}{k}$	$\Omega(1)$	[7], 1990
Shortest Clause	$\frac{1}{8}\left(\frac{k-1}{k-3}\right)^{k-3} \frac{k-1}{k-2} \cdot \frac{2^k}{k}$	w.h.p.	[8], 1992
SC + backtracking	$\sim 1.817 \cdot \frac{2^k}{k}$	w.h.p.	[15], 1996
BP + decimation (nonrigorous)	$\mathrm{e} \cdot 2^k/k$	w.h.p.	[22], 2007

The term "success probability" refers to the probability with which the algorithm finds a satisfying assignment of a random formula. For all algorithms except Unit Clause this is $1 - o(1)$ as $n \to \infty$. For Unit Clause it converges to a number strictly between 0 and 1.

Montanari, Ricci-Tersenghi, and Semerjian [22] provide evidence that Belief Propagation guided decimation may succeed up to density $\mathrm{e} \cdot 2^k/k$ w.h.p. This algorithm is based on a very different paradigm from the others mentioned in Table 1.1. The basic idea is to run a message passing algorithm (Belief Propagation) to compute for each variable the marginal probability that this variable takes the value true/false in a uniformly random satisfying assignment. Then, the decimation step selects a variable randomly, assigns it the value true/false with the corresponding marginal probability, and simplifies the formula. Ideally, repeating this procedure will yield a satisfying assignment, provided that Belief Propagation keeps yielding the correct marginals. Proving (or disproving) this remains a major open problem.

Survey Propagation is a modification of Belief Propagation that aims to approximate the marginal probabilities induced by a particular nonuniform probability distribution on the set of certain generalized assignments [6, 21]. It can be combined with a decimation procedure as well to obtain a heuristic for *finding* a satisfying assignment. However, there is no evidence that Survey Propagation guided decimation finds satisfying assignments beyond $\mathrm{e} \cdot 2^k/k$ for general k w.h.p.

In summary, various algorithms are known or appear to succeed with either high or a nonvanishing probability for densities $c \cdot 2^k/k$, where the constant c depends on the particulars of the algorithm. But there has been no prior evidence (either rigorous results, nonrigorous arguments, or experiments) that some algorithm succeeds for densities $m/n = 2^k \omega(k)/k$ with $\omega(k) \to \infty$.

The discussion so far concerns the case of general k. In addition, a large number of papers deal with the case $k = 3$. Flaxman [13] provides a survey. Currently the best rigorously analyzed algorithm for random 3-SAT is known to succeed up to $m/n = 3.52$ [17, 19]. This is also the best known lower bound on the 3-SAT threshold. The best current upper bound is 4.506 [11], and nonrigorous arguments suggest that the threshold is ≈ 4.267 [6]. As mentioned in section 1.2, there is nonrigorous evidence that the structure of the set of all satisfying assignments evolves differently in random 3-SAT than in random k-SAT for "large" k. This may be why experiments suggest that Survey Propagation guided decimation for 3-SAT succeeds for densities m/n up to 4.2, i.e., close to the conjectured 3-SAT threshold [6].

1.4. Techniques and outline. Remember the *factor graph* representation of a formula $\mathbf{\Phi}$: the vertices are the variables and the clauses, and each clause is adjacent to all the variables that appear in it. In terms of the factor graph it is easy to point out the key difference between `Fix` and, say, Unit Clause.

The execution of Unit Clause can be described as follows. Initially all variables are unassigned. In each step the algorithm checks for a *unit clause* C, i.e., a clause C that has precisely one unassigned variable x left while the previously assigned variables do not already satisfy C. If there is a unit clause C, the algorithm assigns x so as to satisfy it. If not, the algorithm just assigns a random value to a random unassigned variable.

In terms of the factor graph, every step of Unit Clause merely inspects the *first neighborhood* of each clause C to decide whether C is a unit clause. Clauses or variables that have distance two or more have no immediate impact (cf. Figure 1.1). Thus, one could call Unit Clause a "depth one" algorithm. In this sense most other rigorously analyzed algorithms (e.g., Shortest Clause, Walksat) are depth one as well.

`Fix` is depth three. Initially it sets all variables to true. To obtain a satisfying assignment, in the first phase the algorithm passes over all initially unsatisfied (i.e., all negative) clauses. For each such clause C, `Fix` inspects all variables x in that clause, all clauses D in which these variables occur, and all variables y that occur in those clauses (cf. Figure 1.1). Based on this information, the algorithm selects a variable x

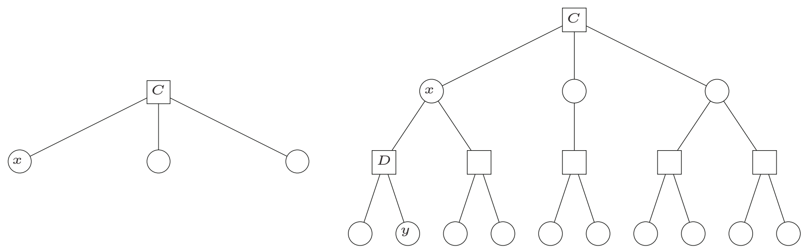

FIG. 1.1. *Depth one versus depth three.*

from C that gets set to false so as to satisfy C. More precisely, `Fix` aims to choose x so that setting it to false does not generate any new unsatisfied clauses. The second and the third phases may reassign (very few) variables once more. We will describe the algorithm precisely in section 3.

In summary, the main reason `Fix` outperforms Unit Clause and the other algorithms is that it bases its decisions on the third neighborhoods in the factor graph, rather than just the first. This entails that the analysis of `Fix` is significantly more involved than that of, say, Unit Clause. The analysis is based on a blend of probabilistic methods (e.g., martingales) and combinatorial arguments. We can employ the *method of deferred decisions* to a certain extent: in the analysis we "pretend" that the algorithm exposes the literals of the random input formula only when it becomes strictly necessary, so that the unexposed ones remain "random." However, the picture is not as clean as in the analysis of, say, Unit Clause. In particular, analyzing `Fix` via the method of differential equations seems prohibitive, at least for general clause lengths k. Section 3 contains an outline of the analysis, the details of which are carried out in sections 4–6. Before we come to this, we summarize a few preliminaries in section 2.

2. Preliminaries and notation. In this section we introduce some notation and present a few basic facts. Although most of them (or closely related ones) are well known, we present some of the proofs for the sake of completeness.

2.1. Balls and bins. Consider a balls and bins experiment where μ distinguishable balls are thrown independently and uniformly at random into n bins. Thus, the probability of each distribution of balls into bins equals $n^{-\mu}$.

LEMMA 2.1. *Let $\mathcal{Z}(\mu, n)$ be the number of empty bins. Let $\lambda = n\exp(-\mu/n)$. Then* $\mathrm{P}\left[\mathcal{Z}(\mu,n) \leq \lambda/2\right] \leq O(\sqrt{\mu}) \cdot \exp(-\lambda/8)$ *as $n \to \infty$.*

The proof is based on the following *Chernoff bound* on the tails of a binomially distributed random variable X with mean λ (see [18, pp. 26–28]): for any $t > 0$

$$\mathrm{P}(X \geq \lambda + t) \leq \exp\left(-\frac{t^2}{2(\lambda + t/3)}\right) \quad \text{and} \quad \mathrm{P}(X \leq \lambda - t) \leq \exp\left(-\frac{t^2}{2\lambda}\right). \tag{2.1}$$

Proof of Lemma 2.1. Let X_i be the number of balls in bin i. In addition, let $(Y_i)_{1 \leq i \leq n}$ be a family of mutually independent Poisson variables with mean μ/n, and let $Y = \sum_{i=1}^{n} Y_i$. Then Y has a Poisson distribution with mean μ. Therefore, Stirling's formula shows that $\mathrm{P}\left[Y = \mu\right] = \Theta(\mu^{-1/2})$. Furthermore, the *conditional* joint distribution of $Y_1, \ldots, Y_n$, given that $Y = \mu$, coincides with the joint distribution of $X_1, \ldots, X_n$ (see, e.g., [12, section 2.6]). As a consequence,

$$\begin{aligned} \mathrm{P}\left[\mathcal{Z}(\mu, n) \le \lambda/2\right] &= \mathrm{P}\left[|\{i \in [n] : Y_i = 0\}| < \lambda/2 | Y = \mu\right] \\ &\le \frac{\mathrm{P}\left[|\{i \in [n] : Y_i = 0\}| < \lambda/2\right]}{\mathrm{P}\left[Y = \mu\right]} \\ &= O(\sqrt{\mu}) \cdot \mathrm{P}\left[|\{i \in [n] : Y_i = 0\}| < \lambda/2\right]. \qquad (2.2) \end{aligned}$$

Finally, since $Y_1, \ldots, Y_n$ are mutually independent and $\mathrm{P}[Y_i = 0] = \lambda/n$ for all $1 \le i \le n$, the number of indices $i \in [n]$ such that $Y_i = 0$ is binomially distributed with mean λ. Thus, the assertion follows from (2.2) and the Chernoff bound (2.1). ∎

2.2. Random k-SAT formulas. Throughout the paper we let $V = V_n = \{x_1, \ldots, x_n\}$ be a set of propositional variables. If $Z \subset V$, then $\bar{Z} = \{\bar{x} : x \in Z\}$ contains the corresponding set of negative literals. Moreover, if l is a literal, then $|l|$ signifies the underlying propositional variable. If μ is an integer, let $[\mu] = \{1, 2, \ldots, \mu\}$.

We let $\Omega_k(n, m)$ be the set of all k-SAT formulas with variables from $V = \{x_1, \ldots, x_n\}$ that contain precisely m clauses. More precisely, we consider each formula an ordered m-tuple of clauses and each clause an ordered k-tuple of literals, allowing both literals to occur repeatedly in one clause and clauses to occur repeatedly in the formula. Thus, $|\Omega_k(n, m)| = (2n)^{km}$. Let $\Sigma_k(n, m)$ be the power set of $\Omega_k(n, m)$, and let $\mathrm{P} = \mathrm{P}_k(n, m)$ be the uniform probability measure.

Throughout the paper we denote a uniformly random element of $\Omega_k(n, m)$ by $\boldsymbol{\Phi}$. In addition, we use Φ to denote specific (i.e., nonrandom) elements of $\Omega_k(n, m)$. If $\Phi \in \Omega_k(n, m)$, then Φ_i denotes the ith clause of Φ, and Φ_{ij} denotes the jth literal of Φ_i.

LEMMA 2.2. *For any $\delta > 0$ and any $k \ge 3$ there is $n_0 > 0$ such that for all $n > n_0$ the following is true. Suppose that $m \ge \delta n$ and that $X_i : \Omega_k(n, m) \to \{0, 1\}$ is a random variable for each $i \in [m]$. Let $\mu = \lceil \ln^2 n \rceil$. For a set $\mathcal{M} \subset [m]$ let $\mathcal{E}_{\mathcal{M}}$ signify the event that $X_i = 1$ for all $i \in \mathcal{M}$. If there is a number $\lambda \ge \delta$ such that for any $\mathcal{M} \subset [m]$ of size μ we have*

$$\mathrm{P}\left[\mathcal{E}_{\mathcal{M}}\right] \le \lambda^{\mu},$$

then

$$\mathrm{P}\left[\sum_{i=1}^{m} X_i \ge (1 + \delta)\lambda m\right] < n^{-10}.$$

Proof. Let $\mathcal{X}$ be the number of sets $\mathcal{M} \subset [m]$ of size μ such that $X_i = 1$ for all $i \in \mathcal{M}$. Then

$$\mathrm{E}\left[\mathcal{X}\right] = \sum_{\mathcal{M} \subset [m] : |\mathcal{M}| = \mu} \mathrm{P}\left[\forall i \in \mathcal{M} : X_i = 1\right] \le \binom{m}{\mu} \lambda^{\mu}.$$

If $\sum_{i=1}^{m} X_i \ge L = \lceil (1 + \delta)\lambda m \rceil$, then $\mathcal{X} \ge \binom{L}{\mu}$. Consequently, by Markov's inequality

$$\begin{aligned} \mathrm{P}\left[\sum_{i=1}^{m} X_i \ge L\right] &\le \mathrm{P}\left[\mathcal{X} \ge \binom{L}{\mu}\right] \le \frac{\mathrm{E}\left[\mathcal{X}\right]}{\binom{L}{\mu}} \le \frac{\binom{m}{\mu}\lambda^{\mu}}{\binom{L}{\mu}} \\ &\le \left(\frac{\lambda m}{L - \mu}\right)^{\mu} \le \left(\frac{\lambda m}{(1 + \delta)\lambda m - \mu}\right)^{\mu}. \end{aligned}$$

Since $\lambda m \geq \delta^2 n$ we see that $(1+\delta)\lambda m - \mu \geq (1+\delta/2)\lambda m$ for sufficiently large n. Hence, for large enough n we have $\mathrm{P}\left[\sum_{i=1}^m X_i \geq L\right] \leq (1+\delta/2)^{-\mu} < n^{-10}$. ∎

Although we allow variables to appear repeatedly in the same clause, the following lemma shows that this occurs very rarely w.h.p.

LEMMA 2.3. *Suppose that $m = O(n)$. Then w.h.p. there are at most $\ln n$ indices $i \in [m]$ such that one of the following is true.*

(1) *There are $1 \leq j_1 < j_2 \leq k$ such that $|\mathbf{\Phi}_{ij_1}| = |\mathbf{\Phi}_{ij_2}|$.*

(2) *There are $i' \neq i$ and indices $j_1 \neq j_2$, $j_1' \neq j_2'$ such that $|\mathbf{\Phi}_{ij_1}| = |\mathbf{\Phi}_{i'j_1'}|$ and $|\mathbf{\Phi}_{ij_2}| = |\mathbf{\Phi}_{i'j_2'}|$.*

Furthermore, w.h.p. no variable occurs in more than $\ln^2 n$ clauses.

Proof. Let X be the number of indices i for which (1) holds. For each $i \in [m]$ and any pair $1 \leq j_1 < j_2 \leq k$, the probability that $|\mathbf{\Phi}_{ij_1}| = |\mathbf{\Phi}_{ij_2}|$ is $1/n$, because each of the two variables is chosen uniformly at random. Hence, by the union bound for any i the probability that there are $j_1 < j_2$ such that $|\mathbf{\Phi}_{ij_1}| = |\mathbf{\Phi}_{ij_2}|$ is at most $\binom{k}{2}/n$. Consequently, $\mathrm{E}[X] \leq m\binom{k}{2}/n = O(1)$ as $n \to \infty$, and thus $X \leq \frac{1}{2}\ln n$ w.h.p. by Markov's inequality.

Let Y be the number of $i \in [m]$ for which (2) is true. For any given $i, i', j_1, j_1', j_2, j_2'$ the probability that $|\mathbf{\Phi}_{ij_1}| = |\mathbf{\Phi}_{i'j_1'}|$ and $|\mathbf{\Phi}_{ij_2}| = |\mathbf{\Phi}_{i'j_2'}|$ is $1/n^2$. Furthermore, there are m^2 ways to choose i, i' and then $(k(k-1))^2$ ways to choose j_1, j_1', j_2, j_2'. Hence, $\mathrm{E}[Y] \leq m^2k^4n^{-2} = O(1)$ as $n \to \infty$. Thus, $Y \leq \frac{1}{2}\ln n$ w.h.p. by Markov's inequality.

Finally, for any variable x the number of indices $i \in [m]$ such that x occurs in $\mathbf{\Phi}_i$ has a binomial distribution $\mathrm{Bin}(m, 1-(1-1/n)^k)$. Since the mean $m\cdot(1-(1-1/n)^k)$ is $O(1)$, the Chernoff bound (2.1) implies that the probability that x occurs in more than $\ln^2 n$ clauses is $o(1/n)$. Hence, by the union bound there is no variable with this property w.h.p. ∎

Recall that a *filtration* is a sequence $(\mathcal{F}_t)_{0\leq t\leq\tau}$ of σ-algebras $\mathcal{F}_t \subset \Sigma_k(n,m)$ such that $\mathcal{F}_t \subset \mathcal{F}_{t+1}$ for all $0 \leq t < \tau$. For a random variable $X : \Omega_k(n,m) \to \mathbf{R}$ we let $\mathrm{E}[X|\mathcal{F}_t]$ denote the *conditional expectation*. Thus, $\mathrm{E}[X|\mathcal{F}_t] : \Omega_k(n,m) \to \mathbf{R}$ is an $\mathcal{F}_t$-measurable random variable such that for any $A \in \mathcal{F}_t$ we have

$$\sum_{\Phi\in A} \mathrm{E}[X|\mathcal{F}_t](\Phi) = \sum_{\Phi\in A} X(\Phi).$$

Also remember that $\mathrm{P}[\cdot|\mathcal{F}_t]$ assigns a probability measure $\mathrm{P}[\cdot|\mathcal{F}_t](\Phi)$ to any $\Phi \in \Omega_k(n,m)$, namely,

$$\mathrm{P}[\cdot|\mathcal{F}_t](\Phi) : A \in \Sigma_k(n,m) \mapsto \mathrm{E}[\mathbf{1}_A|\mathcal{F}_t](\Phi),$$

where $\mathbf{1}_A(\varphi) = 1$ if $\varphi \in A$ and $\mathbf{1}_A(\varphi) = 0$ otherwise.

LEMMA 2.4. *Let $(\mathcal{F}_t)_{0\leq t\leq\tau}$ be a filtration and let $(X_t)_{1\leq t\leq\tau}$ be a sequence of nonnegative random variables such that each X_t is $\mathcal{F}_t$-measurable. Assume that there are numbers $\xi_t \geq 0$ such that $\mathrm{E}[X_t|\mathcal{F}_{t-1}] \leq \xi_t$ for all $1 \leq t \leq \tau$. Then $\mathrm{E}[\prod_{1\leq t\leq\tau} X_t|\mathcal{F}_0] \leq \prod_{1\leq t\leq\tau} \xi_t$.*

Proof. For $1 \leq s \leq \tau$ we let $Y_s = \prod_{t=1}^s X_t$. Let $s > 1$. Since Y_{s-1} is $\mathcal{F}_{s-1}$-measurable, we obtain

$$\begin{aligned}\mathrm{E}[Y_s|\mathcal{F}_0] &= \mathrm{E}[Y_{s-1}X_s|\mathcal{F}_0] = \mathrm{E}[\mathrm{E}[Y_{s-1}X_s|\mathcal{F}_{s-1}]|\mathcal{F}_0] \\ &= \mathrm{E}[Y_{s-1}\mathrm{E}[X_s|\mathcal{F}_{s-1}]|\mathcal{F}_0] \leq \xi_s\mathrm{E}[Y_{s-1}|\mathcal{F}_0],\end{aligned}$$

whence the assertion follows by induction. ∎

We also need the following tail bound ("Azuma–Hoeffding"; see, e.g., [18, p. 37]).

LEMMA 2.5. *Let $(M_t)_{0\leq t\leq\tau}$ be a martingale with respect to a filtration $(\mathcal{F}_t)_{0\leq t\leq\tau}$ such that $M_0=0$. Suppose that there exist numbers c_t such that $|M_t-M_{t-1}|\leq c_t$ for all $1\leq t\leq\tau$. Then for any $\lambda>0$ we have $\mathrm{P}\left[|M_\tau|>\lambda\right]\leq\exp\left[-\lambda^2/(2\sum_{t=1}^{\tau}c_t^2)\right]$.*

Finally, we need the following bound on the number of clauses that have "few" positive literals in total but contain at least one variable (either positively or negatively) from a "small" set.

LEMMA 2.6. *Suppose that $k\geq 3$ and $m/n\leq 2^k k^{-1}\ln k$. Let $1\leq l\leq\sqrt{k}$ and set $\delta=0.01\cdot k^{-4l}$. For a set $Z\subset V$ let X_Z be the number of indices $i\in[m]$ such that $\mathbf{\Phi}_i$ is a clause with precisely l positive literals that contains a variable from Z. Then $\max\{X_Z:|Z|\leq\delta n\}\leq\sqrt{\delta}n$ w.h.p.*

Proof. Let $\mu=\lceil\sqrt{\delta}n\rceil$. We use a first moment argument. Clearly we just need to consider sets Z of size $\lfloor\delta n\rfloor$. Thus, there are at most $\binom{n}{\delta n}$ ways to choose Z. Once Z is fixed, there are at most $\binom{m}{\mu}$ ways to choose a set $\mathcal{I}\subset[m]$ of size μ. For each $i\in\mathcal{I}$ the probability that $\mathbf{\Phi}_i$ contains a variable from Z and has precisely l positive literals is at most $2^{1-k}k\binom{k}{l}\delta$. Hence, by the union bound

$$\begin{aligned}
\mathrm{P}\left[\max\{X_Z:|Z|\leq\delta n\}\geq\mu\right] &\leq \binom{n}{\delta n}\binom{m}{\mu}\left[2^{1-k}k\binom{k}{l}\delta\right]^{\mu}\\
&\leq \left(\frac{\mathrm{e}}{\delta}\right)^{\delta n}\left(\frac{2\mathrm{e}km\binom{k}{l}\delta}{2^k\mu}\right)^{\mu}\\
&\leq \left(\frac{\mathrm{e}}{\delta}\right)^{\delta n}\left(\frac{2\mathrm{e}\ln(k)\binom{k}{l}\delta n}{\mu}\right)^{\mu} && [\text{as } m/n\leq 2^k k^{-1}\ln k]\\
&\leq \left(\frac{\mathrm{e}}{\delta}\right)^{\delta n}\left(4\mathrm{e}\ln(k)\cdot k^l\cdot\sqrt{\delta}\right)^{\mu} && [\text{because } \mu=\lceil\sqrt{\delta}n\rceil]\\
&\leq \left(\frac{\mathrm{e}}{\delta}\right)^{\delta n}\delta^{\sqrt{\delta}n/8} && [\text{using } \delta=0.01\cdot k^{-4l}]\\
&= \exp\left[n\sqrt{\delta}\left(\sqrt{\delta}(1-\ln\delta)+\frac{1}{8}\ln\delta\right)\right].
\end{aligned}$$

The last expression is $o(1)$ because $\sqrt{\delta}(1-\ln\delta)+\frac{1}{8}\ln\delta$ is negative as $\delta<0.01$. ∎

LEMMA 2.7. *Assume that $k\geq 3$ and $m/n\leq 2^k$. W.h.p. $\mathbf{\Phi}$ does not admit a set $\mathcal{M}\subset[m]$ of $\mu=|\mathcal{M}|\leq\ln n$ clauses and a set $Y\subset V$ of $|Y|\leq\mu$ variables such that for each $i\in\mathcal{M}$ there are at least three $j\in[k]$ such that $|\mathbf{\Phi}_{ij}|\in Y$.*

Proof. We use a first moment argument. For a given $3\leq\mu\leq\ln n$ let X_μ be the number of pairs $(\mathcal{M},Y)$ with $|\mathcal{M}|=|Y|=\mu$ such that $\mathbf{\Phi}_i$ features three variables from Y for each $i\in\mathcal{M}$. Since there are $\binom{m}{\mu}$ ways to choose $\mathcal{M}$ and $\binom{n}{\mu}$ ways to choose Y, and because for a random literal $\mathbf{\Phi}_{ij}$ the probability that $|\mathbf{\Phi}_{ij}|\in Y$ equals μ/n, we obtain

$$\mathrm{E}\left[X_\mu\right]\leq\binom{n}{\mu}\binom{m}{\mu}\binom{k}{3}^{\mu}\left(\frac{\mu}{n}\right)^{3\mu}\leq\left(\frac{\mathrm{e}^2nmk^3\mu^3}{6n^3}\right)^{\mu}\leq(2^{k+1}k^3n^{-1}\ln^3 n)^{\mu}.$$

As $\sum_{3\leq\mu\leq\ln n}\mathrm{E}X_\mu=o(1)$, the assertion follows from Markov's inequality. ∎

3. The algorithm `Fix`. In this section we present the algorithm `Fix`. To establish Theorem 1.1 we will prove the following: for any $0<\varepsilon<0.1$ there is $k_0=k_0(\varepsilon)>10$ such that for all $k\geq k_0$ the algorithm `Fix` outputs a satisfying assignment w.h.p. when applied to $\mathbf{\Phi}$ with $m = \lfloor n \cdot (1-\varepsilon)2^k k^{-1} \ln k \rfloor$. Thus, we assume that k exceeds some large enough number k_0 depending on ε only. In addition, we assume throughout that $n > n_0$ for some large enough $n_0 = n_0(\varepsilon, k)$. We set

$$\omega = (1-\varepsilon)\ln k \text{ and } k_1 = \lceil k/2 \rceil.$$

Let $\Phi \in \Omega_k(n,m)$ be a k-SAT instance. When applied to Φ the algorithm basically tries to "fix" the all-true assignment by setting "a few" variables $Z \subset V$ to false so as to satisfy all clauses. Obviously, the set Z will have to contain one variable from each clause consisting of negative literals only. The key issue is to pick "the right" variables. To this end, the algorithm goes over the all-negative clauses in the natural order. If the present all-negative clause Φ_i does not contain a variable from Z yet, `Fix` (tries to) identify a "safe" variable in Φ_i, which it then adds to Z. Here "safe" means that setting the variable to false does not create new unsatisfied clauses. More precisely, we say that a clause Φ_i is *Z-unique* if Φ_i contains exactly one positive literal from $V \setminus Z$ and no literal from $\bar{Z}$. Moreover, $x \in V \setminus Z$ is *Z-unsafe* if it occurs positively in a Z-unique clause, and *Z-safe* if this is not the case. Then in order to fix an all-negative clause Φ_i, we prefer Z-safe variables.

To implement this idea, `Fix` proceeds in three phases. Phase 1 performs the operation described in the previous paragraph: try to identify a Z-safe variable in each all-negative clause. Of course, it may happen that an all-negative clause does not contain a Z-safe variable. In this case `Fix` just picks the variable in position k_1. Consequently, the assignment constructed in the first phase may not satisfy *all* clauses. However, we will prove that the number of unsatisfied clauses is very small, and the purpose of Phases 2 and 3 is to deal with them. Before we come to this, let us describe Phase 1 precisely.

ALGORITHM 3.1. `Fix`(Φ).
Input: a k-SAT formula Φ. *Output:* either a satisfying assignment or "fail."

1a. Let $Z = \emptyset$.
1b. For $i = 1, \dots, m$ do
1c. If Φ_i is all-negative and contains no variable from Z
1d. If there is $1 \le j < k_1$ such that $|\Phi_{ij}|$ is Z-safe, then pick the least such j and add $|\Phi_{ij}|$ to Z.
1e. Otherwise add $|\Phi_{i\,k_1}|$ to Z.

The following proposition, which we will prove in section 4, summarizes the analysis of Phase 1. Let σ_Z be the assignment that sets all variables in $V \setminus Z$ to true and all variables in Z to false.

PROPOSITION 3.2. *At the end of Phase* 1 *of* `Fix`($\mathbf{\Phi}$) *the following statements are true w.h.p.*

(1) *We have* $|Z| \le 4nk^{-1}\ln\omega$.
(2) *At most* $(1+\varepsilon/3)\omega n$ *clauses are Z-unique.*
(3) *At most* $\exp(-k^{\varepsilon/8})n$ *clauses are unsatisfied under* σ_Z.

Since $k \ge k_0(\varepsilon)$ is "large," we should think of $\exp(-k^{\varepsilon/8})$ as tiny. In particular, $\exp(-k^{\varepsilon/8}) \ll \omega/k$. As the probability that a random clause is all-negative is 2^{-k}, under the all-true assignment, $(1+o(1))2^{-k}m \sim \omega n/k$ clauses are unsatisfied w.h.p. Hence, the outcome σ_Z of Phase 1 is already a lot better than the all-true assignment w.h.p.

Step 1d considers only indices $1 \le j \le k_1$. This is just for technical reasons, namely, to maintain a certain degree of stochastic independence to facilitate (the analysis of) Phase 2.

Phase 2 deals with the clauses that are unsatisfied under σ_Z. The general plan is similar to Phase 1: we (try to) identify a set Z' of "safe" variables that can be used to satisfy the σ_Z-unsatisfied clauses without "endangering" further clauses. More precisely, we say that a clause Φ_i is (Z, Z')-*endangered* if there is no $1 \leq j \leq k$ such that the literal Φ_{ij} is true under σ_Z and $|\Phi_{ij}| \in V \setminus Z'$. Roughly speaking, Φ_i is (Z, Z')-endangered if it relies on one of the variables in Z' to be satisfied. Call Φ_i (Z, Z')-*secure* if it is not (Z, Z')-endangered. Phase 2 will construct a set Z' such that for all $1 \leq i \leq m$ one of the following is true:

- Φ_i is (Z, Z')-secure.
- There are at least three indices $1 \leq j \leq k$ such that $|\Phi_{ij}| \in Z'$.

To achieve this, we say that a variable x is (Z, Z')-*unsafe* if $x \in Z \cup Z'$ or there are indices $(i, l) \in [m] \times [k]$ such that the following two conditions hold:

(a) For all $j \neq l$ we have $\Phi_{ij} \in Z \cup Z' \cup \overline{V \setminus Z}$.

(b) $\Phi_{il} = x$.

(In words, x occurs positively in Φ_i, and all other literals of Φ_i are either positive but in $Z \cup Z'$, or negative but not in Z.) Otherwise we call x (Z, Z')-*safe*. In the course of the process, `Fix` greedily tries to add as few (Z, Z')-unsafe variables to Z' as possible. Phase 2 proceeds as follows.

2a. Let Q consist of all $i \in [m]$ such that Φ_i is unsatisfied under σ_Z. Let $Z' = \emptyset$.
2b. While $Q \neq \emptyset$
2c. Let $i = \min Q$.
2d. If there are indices $k_1 < j_1 < j_2 < j_3 \leq k - 5$ such that $|\Phi_{ij_l}|$ is (Z, Z')-safe for $l = 1, 2, 3$,
pick the lexicographically first such sequence j_1, j_2, j_3 and add $|\Phi_{ij_1}|$, $|\Phi_{ij_2}|, |\Phi_{ij_3}|$ to Z'.
2e. else
let $k - 5 < j_1 < j_2 < j_3 \leq k$ be the lexicographically first sequence such that $|\Phi_{ij_l}| \notin Z'$ and add $|\Phi_{ij_l}|$ to Z' $(l = 1, 2, 3)$.
2f. Let Q be the set of all (Z, Z')-endangered clauses that contain less than 3 variables from Z'.

Note that the While-loop gets executed at most $n/3$ times, because Z' gains three new elements in each iteration. Actually we prove in section 5 below that the final set Z' is fairly small w.h.p.

PROPOSITION 3.3. *The set Z' obtained in Phase 2 of* $\texttt{Fix}(\mathbf{\Phi})$ *has size* $|Z'| \leq nk^{-12}$ *w.h.p.*

After completing Phase 2, `Fix` is going to set the variables in $V \setminus (Z \cup Z')$ to true and the variables in $Z \setminus Z'$ to false. This will satisfy all (Z, Z')-secure clauses. In order to satisfy the (Z, Z')-endangered clauses as well, `Fix` needs to set the variables in Z' appropriately. To this end, we set up a bipartite graph $G(\Phi, Z, Z')$ whose vertex set consists of the (Z, Z')-endangered clauses and the set Z'. Each (Z, Z')-endangered clause is adjacent to the variables from Z' that occur in it. If there is a matching M in $G(\Phi, Z, Z')$ that covers all (Z, Z')-endangered clauses, we construct an assignment $\sigma_{Z,Z',M}$ as follows: for each variable $x \in V$ let

$$\sigma_{Z,Z',M}(x) = \begin{cases} \text{false} & \text{if } x \in Z \setminus Z', \\ \text{false} & \text{if } \{\Phi_i, x\} \in M \text{ for some } i \text{ and } x \text{ occurs negatively in } \Phi_i, \\ \text{true} & \text{otherwise.} \end{cases}$$

To be precise, Phase 3 proceeds as follows.

3. If $G(\Phi, Z, Z')$ has a matching that covers all (Z, Z')-endangered clauses, then compute an (arbitrary) such matching M and output $\sigma_{Z,Z',M}$. If not, output "fail."

The (bipartite) matching computation can be performed in $O((n+m)^{3/2})$ time via the Hopcroft–Karp algorithm. In section 6 we will show that the matching exists w.h.p.

PROPOSITION 3.4. *W.h.p.* $G(\mathbf{\Phi}, Z, Z')$ *has a matching that covers all* (Z, Z')*-endangered clauses.*

Proof of Theorem 1.1. `Fix` is clearly a deterministic polynomial time algorithm. It remains to show that `Fix(`$\mathbf{\Phi}$`)` outputs a satisfying assignment w.h.p. By Proposition 3.4 Phase 3 will find a matching M that covers all (Z, Z')-endangered clauses w.h.p., and thus the output will be the assignment $\sigma = \sigma_{Z,Z',M}$ w.h.p. Assume that this is the case. Then σ sets all variables in $Z \setminus Z'$ to false and all variables in $V \setminus (Z \cup Z')$ to true, thereby satisfying all (Z, Z')-secure clauses. Furthermore, for each (Z, Z')-endangered clause $\mathbf{\Phi}_i$ there is an edge $\{\mathbf{\Phi}_i, |\mathbf{\Phi}_{ij}|\}$ in M. If $\mathbf{\Phi}_{ij}$ is negative, then $\sigma(|\mathbf{\Phi}_{ij}|) = $ false, and if $\mathbf{\Phi}_{ij}$ is positive, then $\sigma(\mathbf{\Phi}_{ij}) = $ true. In either case σ satisfies $\mathbf{\Phi}_i$. ∎

4. Proof of Proposition 3.2. Throughout this section we let $0 < \varepsilon < 0.1$ and assume that $k \geq k_0$ for a sufficiently large $k_0 = k_0(\varepsilon)$ depending on ε only. Moreover, we assume that $m = \lfloor n \cdot (1-\varepsilon)2^k k^{-1} \ln k \rfloor$ and that $n > n_0$ for some large enough $n_0 = n_0(\varepsilon, k)$. Let $\omega = (1-\varepsilon)\ln k$ and $k_1 = \lceil k/2 \rceil$.

4.1. Outline. Before we proceed to the analysis, it is worthwhile to give a brief intuitive explanation as to why Phase 1 "works." Namely, let us consider just the *first* all-negative clause $\mathbf{\Phi}_i$ of the random input formula. Without loss of generality we may assume that $i = 1$. Given that $\mathbf{\Phi}_1$ is all-negative, the k-tuple of variables $(|\mathbf{\Phi}_{1j}|)_{1 \leq j \leq k} \in V^k$ is uniformly distributed. Furthermore, at this point $Z = \emptyset$. Hence, a variable x is Z-unsafe iff it occurs as the unique positive literal in some clause. The expected number of clauses with exactly one positive literal is $k2^{-k}m \sim \omega n$ as $n \to \infty$. Thus, for each variable x the expected number of clauses in which x is the only positive literal is $k2^{-k}m/n \sim \omega$. In fact, for each variable the number of such clauses is asymptotically Poisson. Consequently, the probability that x is Z-safe is $(1+o(1))\exp(-\omega)$. Returning to the clause $\mathbf{\Phi}_1$, we conclude that the *expected* number of indices $1 \leq j \leq k_1$ such that $|\mathbf{\Phi}_{1j}|$ is Z-safe is $(1+o(1))k_1 \exp(-\omega)$. Since $\omega = (1-\varepsilon)\ln k$ and $k_1 \geq \frac{k}{2}$, we have (for large enough n)

$$(1+o(1))k_1 \exp(-\omega) \geq k^{\varepsilon}/3.$$

Indeed, the number of indices $1 \leq j \leq k_1$ so that $|\mathbf{\Phi}_{1j}|$ is Z-safe is binomially distributed, and hence the probability that there is no Z-safe $|\mathbf{\Phi}_{1j}|$ is at most $\exp(-k^{\varepsilon}/3)$. Since we are assuming that $k \geq k_0(\varepsilon)$ for some large enough $k_0(\varepsilon)$, we should think of k^{ε} as "large." Thus, $\exp(-k^{\varepsilon}/3)$ is tiny, and hence it is "quite likely" that $\mathbf{\Phi}_1$ can be satisfied by setting some variable to false without creating any new unsatisfied clauses. Of course, this argument applies only to the first all-negative clause (i.e., $Z = \emptyset$), and the challenge lies in dealing with the stochastic dependencies that arise.

To this end, we need to investigate how the set Z computed in steps 1a–1e evolves over time. Thus, we will analyze the execution of Phase 1 as a stochastic process, in which the set Z corresponds to a sequence $(Z_t)_{t \geq 0}$ of sets. The time parameter t is the number of all-negative clauses for which either step 1d or step 1e has been executed. We will represent the execution of Phase 1 on input $\mathbf{\Phi}$ by a sequence of (random) maps

$$\pi_t : [m] \times [k] \to \{-1, 1\} \cup V \cup \bar{V} = \{\pm 1, x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}.$$

The maps $(\pi_s)_{0\leq s\leq t}$ capture the information that has determined the first t steps of the process. If $\pi_t(i,j)=1$ (resp., $\pi_t(i,j)=-1$), then `Fix` has only taken into account that $\mathbf{\Phi}_{ij}$ is a positive (negative) literal, but not what the underlying variable is. If $\pi_t(i,j)\in V\cup\bar{V}$, `Fix` has revealed the actual literal $\mathbf{\Phi}_{ij}$.

Let us define the sequence $\pi_t(i,j)$ precisely. Let $Z_0=\emptyset$. Moreover, let U_0 be the set of all i such that there is exactly one j such that $\mathbf{\Phi}_{ij}$ is positive. Further, define $\pi_0(i,j)$ for $(i,j)\in[m]\times[k]$ as follows. If $i\in U_0$ and $\mathbf{\Phi}_{ij}$ is positive, then let $\pi_0(i,j)=\mathbf{\Phi}_{ij}$. Otherwise, let $\pi_0(i,j)$ be 1 if $\mathbf{\Phi}_{ij}$ is a positive literal and -1 if $\mathbf{\Phi}_{ij}$ is a negative literal. In addition, for $x\in V$ let

$$U_0(x)=|\{i\in U_0:\exists j\in[k]:\pi_0(i,j)=x\}|$$

be the number of clauses in which x is the unique positive literal. For $t\geq 1$ we define π_t as follows.

PI1 If there is no index $i\in[m]$ such that $\mathbf{\Phi}_i$ is all-negative but contains no variable from Z_{t-1}, the process stops. Otherwise let ϕ_t be the smallest such index.

PI2 If there is $1\leq j<k_1$ such that $U_{t-1}(|\mathbf{\Phi}_{\phi_t j}|)=0$, then choose the smallest such index j; otherwise let $j=k_1$. Let $z_t=\mathbf{\Phi}_{\phi_t j}$ and $Z_t=Z_{t-1}\cup\{z_t\}$.

PI3 Let U_t be the set of all $i\in[m]$ such that $\mathbf{\Phi}_i$ is Z_t-unique. For $x\in V$ let $U_t(x)$ be the number of indices $i\in U_t$ such that x occurs positively in $\mathbf{\Phi}_i$.

PI4 For any $(i,l)\in[m]\times[k]$ let

$$\pi_t(i,l)=\begin{cases}\mathbf{\Phi}_{il} & \text{if } (i=\phi_t\wedge l\leq k_1)\vee|\mathbf{\Phi}_{il}|=z_t\\ & \qquad\qquad\vee(i\in U_t\wedge\pi_0(i,l)=1),\\ \pi_{t-1}(i,l) & \text{otherwise.}\end{cases}$$

Let T be the total number of iterations of this process before it stops, and define $\pi_t=\pi_T$, $Z_t=Z_T$, $U_t=U_T$, $U_t(x)=U_T(x)$, $\phi_t=z_t=0$ for all $t>T$.

Let us discuss briefly how the above process mirrors Phase 1 of `Fix`. Step **PI1** selects the least index ϕ_t such that clause $\mathbf{\Phi}_{\phi_t}$ is all-negative but contains no variable from the set Z_{t-1} of variables that have been selected to be set to false so far. In terms of the description of `Fix`, this corresponds to jumping forward to the next execution of steps 1d–1e. Since $U_{t-1}(x)$ is the number of Z_{t-1}-unique clauses in which variable x occurs positively, Step **PI2** applies the same rule as 1d–1e of `Fix` to select the new element z_t to be included in the set Z_t. Step **PI3** then "updates" the numbers $U_t(x)$. Finally, step **PI4** sets up the map π_t to represent the information that has guided the process so far: we reveal the first k_1 literals of the current clause $\mathbf{\Phi}_{\phi_t}$, all occurrences of the variable z_t, and all positive literals of Z_t-unique clauses.

Example 4.1. To illustrate the process **PI1**–**PI4** we run it on a 5-CNF $\mathbf{\Phi}$ with $n=10$ variables and $m=9$ clauses. Thus, $k_1=3$. We are going to illustrate the information that the process reveals step by step. Instead of using $+1$ and -1 to indicate positive/negative literals, we just use $+$ and $-$ to improve readability. Moreover, to economize space we let the *columns* correspond to the clauses. Since $\mathbf{\Phi}$ is random, each literal $\mathbf{\Phi}_{ij}$ is positive/negative with probability $\frac{1}{2}$ independently. Suppose the sign pattern of the formula $\mathbf{\Phi}$ is

$$\begin{matrix}- & - & - & + & + & + & + & + & +\\ - & - & - & - & - & - & + & - & +\\ - & - & - & - & - & - & - & - & +\\ - & - & - & - & - & - & - & + & -\\ - & - & - & - & - & - & - & - & -\end{matrix}$$

Thus, the first three clauses $\mathbf{\Phi}_1, \mathbf{\Phi}_2, \mathbf{\Phi}_3$ are all-negative, the three clauses $\mathbf{\Phi}_4, \mathbf{\Phi}_5, \mathbf{\Phi}_6$ have exactly one positive literal, etc. In order to obtain π_0, we need to reveal the variables underlying the unique positive literals of $\mathbf{\Phi}_4, \mathbf{\Phi}_5, \mathbf{\Phi}_6$. Since we have only conditioned on the signs, the positive literals occurring in $\mathbf{\Phi}_4, \mathbf{\Phi}_5, \mathbf{\Phi}_6$ are still uniformly distributed over V. Suppose revealing them yields

$$\pi_0 = \begin{array}{ccccccccc} - & - & - & \mathbf{x_5} & \mathbf{x_2} & \mathbf{x_3} & + & + & + \\ - & - & - & - & - & - & + & - & + \\ - & - & - & - & - & - & - & - & + \\ - & - & - & - & - & - & - & + & - \\ - & - & - & - & - & - & - & - & - \end{array}$$

Thus, we have $U_0 = \{4, 5, 6\}$, $U_0(x_2) = U_0(x_3) = U_0(x_5) = 1$, and $U_0(x) = 0$ for all other variables x. At time $t = 1$ **PI1** looks out for the first all-negative clause, which happens to be $\mathbf{\Phi}_1$. Hence $\phi_1 = 1$. To implement **PI2**, we need to reveal the first $k_1 = 3$ literals of $\mathbf{\Phi}_1$. The underlying variables are unaffected by the conditioning so far; i.e., they are independently uniformly distributed over V. Suppose we get

$$\begin{array}{ccccccccc} \mathbf{\bar{x}_2} & - & - & x_5 & x_2 & x_3 & + & + & + \\ \mathbf{\bar{x}_3} & - & - & - & - & - & + & - & + \\ \mathbf{\bar{x}_1} & - & - & - & - & - & - & - & + \\ - & - & - & - & - & - & - & + & - \\ - & - & - & - & - & - & - & - & - \end{array}$$

The variables x_2, x_3 underlying the first two literals of $\mathbf{\Phi}_1$ are in U_0. This means that setting them to false would produce new violated clauses. Therefore, **PI2** sets $j = k_1 = 3$, $z_1 = x_1$, and $Z_1 = \{x_1\}$. Now, **PI3** checks out which clauses are Z_1-unique. To this end we need to reveal the occurrences of $z_1 = x_1$ all over the formula. At this point each $\pm$-sign still represents a literal whose underlying variable is uniformly distributed over V. Therefore, for each $\pm$-entry (i, j) we have $|\mathbf{\Phi}_{ij}| = x_1$ with probability $1/n$ independently. Assume that the occurrences of x_1 are as follows:

$$\begin{array}{ccccccccc} \bar{x}_2 & - & \mathbf{\bar{x}_1} & x_5 & x_2 & x_3 & + & + & + \\ \bar{x}_3 & - & - & - & - & - & + & - & + \\ \mathbf{\bar{x}_1} & - & - & - & - & - & - & - & \mathbf{x_1} \\ - & - & - & - & - & - & - & \mathbf{x_1} & - \\ - & - & - & \mathbf{\bar{x}_1} & - & - & - & - & - \end{array}$$

As $x_1 \in Z_1$, we consider x_1 assigned false. Since x_1 occurs positively in the second to last clause $\mathbf{\Phi}_8$, this clause has only one "supporting" literal left. As we have already revealed all occurrences of x_1, the variable underlying this literal is uniformly distributed over $V \setminus \{x_1\}$. Suppose it is x_4. As x_4 is needed to satisfy $\mathbf{\Phi}_8$, we "protect" it by setting $U_1(x_4) = 1$. Furthermore, $\mathbf{\Phi}_4$ features x_1 negatively. Hence, this clause is now satisfied by x_1, and therefore x_5 could safely be set to false. Thus, $U_1(x_5) = 0$. Further, we keep $U_1(x_2) = U_2(x_3) = 1$ and let $U_1 = \{5, 6, 8\}$. Summarizing the information revealed at time $t = 1$, we get

$$\pi_1 = \begin{array}{ccccccccc} \bar{x}_2 & - & \bar{x}_1 & x_5 & x_2 & x_3 & + & \mathbf{x_4} & + \\ \bar{x}_3 & - & - & - & - & - & + & - & + \\ \bar{x}_1 & - & - & - & - & - & - & - & x_1 \\ - & - & - & - & - & - & - & x_1 & - \\ - & - & - & \bar{x}_1 & - & - & - & - & - \end{array}$$

At time $t = 2$ we deal with the second clause $\mathbf{\Phi}_2$ whose column is still all-minus. Hence $\phi_2 = 2$. Since we have already revealed all occurrences of x_1, the first $k_1 = 3$ literals of $\mathbf{\Phi}_2$ are uniformly distributed over $V \setminus Z_1 = \{x_2, \ldots, x_{10}\}$. Suppose revealing them gives

$$\begin{array}{ccccccccc} \bar{x}_2 & \bar{\mathbf{x}}_\mathbf{5} & \bar{x}_1 & x_5 & x_2 & x_3 & + & x_4 & + \\ \bar{x}_3 & \bar{\mathbf{x}}_\mathbf{2} & - & - & - & - & + & - & + \\ \bar{x}_1 & \bar{\mathbf{x}}_\mathbf{3} & - & - & - & - & - & - & x_1 \\ - & - & - & - & - & - & - & x_1 & - \\ - & - & - & \bar{x}_1 & - & - & - & - & - \end{array}$$

The first variable of $\mathbf{\Phi}_2$ is x_5, and $U_1(x_5) = 0$. Thus, **PI2** will select $z_2 = x_5$ and let $Z_2 = \{x_1, x_5\}$. To determine U_2, **PI3** needs to reveal all occurrences of x_5. At this time each $\pm$-sign stands for a literal whose variable is uniformly distributed over $V \setminus Z_1$. Therefore, for each $\pm$-sign the underlying variable is equal to x_5 with probability $1/(n-1) = 1/9$. Assume that the occurrences of x_5 are

$$\begin{array}{ccccccccc} \bar{x}_2 & \bar{\mathbf{x}}_\mathbf{5} & \bar{x}_1 & x_5 & x_2 & x_3 & + & x_4 & + \\ \bar{x}_3 & \bar{x}_2 & - & - & - & - & + & - & \mathbf{x}_\mathbf{5} \\ \bar{x}_1 & \bar{x}_3 & - & - & - & - & - & - & x_1 \\ - & - & - & - & - & - & - & x_1 & - \\ \bar{\mathbf{x}}_\mathbf{5} & - & - & \bar{x}_1 & - & - & - & - & - \end{array}$$

Since x_5 occurs positively in the last clause $\mathbf{\Phi}_9$, it has only one plus left. Thus, this clause is Z_2-unique, and we have to reveal the variable underlying the last $+$-sign. As we have already revealed the occurrences of x_1 and x_5, this variable is uniformly distributed over $V \setminus \{x_1, x_5\}$. Suppose it is x_4. Then $U_2 = \{5, 6, 8, 9\}$, $U_2(x_2) = U_2(x_3) = 1$, $U_2(x_4) = 2$, and π_2 reads as

$$\pi_2 = \begin{array}{ccccccccc} \bar{x}_2 & \bar{x}_5 & \bar{x}_1 & x_5 & x_2 & x_3 & + & x_4 & \mathbf{x}_\mathbf{4} \\ \bar{x}_3 & \bar{x}_2 & - & - & - & - & + & - & x_5 \\ \bar{x}_1 & \bar{x}_3 & - & - & - & - & - & - & x_1 \\ - & - & - & - & - & - & - & x_1 & - \\ \bar{x}_5 & - & - & \bar{x}_1 & - & - & - & - & - \end{array}$$

At this point there are no all-minus columns left, and therefore the process stops with $T = 2$. In the course of the process we have revealed all occurrences of variables in $Z_2 = \{x_1, x_5\}$. Thus, the variables underlying the remaining $\pm$-sign are independently uniformly distributed over $V \setminus Z_2$. ∎

Observe that at each time $t \leq T$ the process **PI1–PI4** adds precisely one variable z_t to Z_t. Thus, $|Z_t| = t$ for all $t \leq T$. Furthermore, for $1 \leq t \leq T$ the map π_t is obtained from π_{t-1} by replacing some ± 1's by literals, but no changes of the opposite type are made.

Of course, the process **PI1–PI4** can be applied to any concrete k-SAT formula Φ (rather than the random $\mathbf{\Phi}$). It then yields a sequence $\pi_t[\Phi]$ of maps, variables $z_t[\Phi]$, sets $U_t[\Phi]$, $Z_t[\Phi]$, and numbers $U_t(x)[\Phi]$. For each integer $t \geq 0$ we define an equivalence relation $\equiv_t$ on the set $\Omega_k(n, m)$ of k-SAT formulas by letting $\Phi \equiv_t \Psi$ iff $\pi_s[\Phi] = \pi_s[\Psi]$ for all $0 \leq s \leq t$. Let $\mathcal{F}_t$ be the σ-algebra generated by the equivalence classes of $\equiv_t$. The family $(\mathcal{F}_t)_{t \geq 0}$ is a filtration. Intuitively, a random variable X is $\mathcal{F}_t$-measurable iff its value is determined by time t. Thus, the following is immediate from the construction.

FACT 4.2. *For any $t \geq 0$, the random map π_t, the random variables ϕ_{t+1} and z_t, the random sets U_t and Z_t, and the random variables $U_t(x)$ for $x \in V$ are $\mathcal{F}_t$-measurable.*

If $\pi_t(i,j) = \pm 1$, then up to time t the process **PI1**–**PI4** has taken only the sign of the literal $\mathbf{\Phi}_{ij}$ into account, but has been oblivious to the underlying variable. The only conditioning is that $|\mathbf{\Phi}_{ij}| \notin Z_t$ (because otherwise **PI4** would have replaced the ± 1 by the actual literal). Since the input formula $\mathbf{\Phi}$ is random, this implies that $|\mathbf{\Phi}_{ij}|$ is uniformly distributed over $V \setminus Z_t$. In fact, for all (i,j) such that $\pi_t(i,j) = \pm 1$ the underlying variables are independently uniformly distributed over $V \setminus Z_t$. Arguments of this type are sometimes referred to as the "method of deferred decisions."

FACT 4.3. *Let $\mathcal{E}_t$ be the set of all pairs (i,j) such that $\pi_t(i,j) \in \{-1,1\}$. The conditional joint distribution of the variables $(|\mathbf{\Phi}_{ij}|)_{(i,j)\in\mathcal{E}_t}$ given $\mathcal{F}_t$ is uniform over $(V \setminus Z_t)^{\mathcal{E}_t}$. In symbols, for any formula Φ and for any map $f : \mathcal{E}_t[\Phi] \to V \setminus Z_t[\Phi]$ we have*

$$\mathrm{P}\left[\forall (i,j) \in \mathcal{E}_t[\Phi] : |\mathbf{\Phi}_{ij}| = f(i,j) | \mathcal{F}_t\right](\Phi) = |V \setminus Z_t[\Phi]|^{-|\mathcal{E}_t[\Phi]|}.$$

In each step $t \leq T$ of the process **PI1**–**PI4** one variable z_t is added to Z_t. There is a chance that this variable occurs in several all-negative clauses, and therefore the stopping time T should be smaller than the total number of all-negative clauses. To prove this, we need the following lemma. Observe that by **PI4** clause $\mathbf{\Phi}_i$ is all-negative and contains no variable from Z_t iff $\pi_t(i,j) = -1$ for all $j \in [k]$.

LEMMA 4.4. *W.h.p. the following is true for all $1 \leq t \leq \min\{T, n\}$: the number of indices $i \in [m]$ such that $\pi_t(i,j) = -1$ for all $j \in [k]$ is at most $2n\omega \exp(-kt/n)/k$.*

Proof. The proof is based on Lemma 2.2 and Fact 4.3. Similar proofs will occur repeatedly. We carry out this one at leisure. For $1 \leq t \leq n$ and $i \in [m]$ we define a random variable

$$X_{ti} = \begin{cases} 1 & \text{if } t \leq T \text{ and } \pi_t(i,j) = -1 \, \forall j \in [k], \\ 0 & \text{otherwise.} \end{cases}$$

The goal is to show that w.h.p.

$$\forall 1 \leq t \leq n : \sum_{i=1}^{m} X_{ti} \leq 2n\omega \exp(-kt/n)/k. \tag{4.1}$$

To this end, we are going to prove that

$$\mathrm{P}\left[\sum_{i=1}^{m} X_{ti} > 2n\omega \exp(-kt/n)/k\right] = o(1/n) \qquad \text{for any } 1 \leq t \leq n. \tag{4.2}$$

Then the union bound entails that (4.1) holds w.h.p. Thus, we are left to prove (4.2).

To establish (4.2) we fix $1 \leq t \leq n$. Considering t fixed, we may drop it as a subscript and write $X_i = X_{ti}$ for $i \in [m]$. Let $\mu = \lceil \ln^2 n \rceil$. For a set $\mathcal{M} \subset [m]$ we let $\mathcal{E}_{\mathcal{M}}$ denote the event that $X_i = 1$ for all $i \in \mathcal{M}$. In order to apply Lemma 2.2 we need to bound the probability of the event $\mathcal{E}_{\mathcal{M}}$ for any $\mathcal{M} \subset [m]$ of size μ. To this end, we consider the random variables

$$\mathcal{N}_{sij} = \begin{cases} 1 & \text{if } \pi_s(i,j) = -1 \text{ and } s \leq T, \\ 0 & \text{otherwise} \end{cases} \qquad (i \in [m], \, j \in [k], \, 0 \leq s \leq n).$$

Then $X_i = 1$ iff $\mathcal{N}_{sij} = 1$ for all $0 \le s \le t$ and all $j \in [k]$. Hence, letting $\mathcal{N}_s = \prod_{(i,j)\in\mathcal{M}\times[k]} \mathcal{N}_{sij}$, we have

$$\mathrm{P}\left[\mathcal{E}_{\mathcal{M}}\right] = \mathrm{E}\left[\prod_{i\in\mathcal{M}} X_i\right] = \mathrm{E}\left[\prod_{s=0}^{t} \mathcal{N}_s\right]. \tag{4.3}$$

The expectation of $\mathcal{N}_0$ can be computed easily: for any $i \in \mathcal{M}$ we have $\prod_{j=1}^{k} \mathcal{N}_{0ij} = 1$ iff clause $\mathbf{\Phi}_i$ is all-negative. Since the clauses of $\mathbf{\Phi}$ are chosen uniformly, $\mathbf{\Phi}_i$ is all-negative with probability 2^{-k}. Furthermore, these events are mutually independent for all $i \in \mathcal{M}$. Therefore,

$$\mathrm{E}\left[\mathcal{N}_0\right] = \mathrm{E}\left[\prod_{i\in\mathcal{M}} \prod_{j=1}^{k} \mathcal{N}_{0ij}\right] = 2^{-k|\mathcal{M}|} = 2^{-k\mu}. \tag{4.4}$$

In addition, we claim that

$$\mathrm{E}\left[\mathcal{N}_s|\mathcal{F}_{s-1}\right] \le (1-1/n)^{k\mu} \qquad \text{for any } 1 \le s \le n. \tag{4.5}$$

To see this, fix any $1 \le s \le n$. We consider four cases.

Case 1: $T < s$. Then $\mathcal{N}_s = 0$ by the definition of the variables $\mathcal{N}_{sij}$.

Case 2: $\pi_{s-1}(i,j) \ne -1$ *for some* $(i,j) \in \mathcal{M}\times[k]$. Then $\pi_s(i,j) = \pi_{s-1}(i,j) \ne -1$ by **PI4**, and thus $\mathcal{N}_s = \mathcal{N}_{sij} = 0$.

Case 3: $\phi_s \in \mathcal{M}$. If the index ϕ_s chosen by **PI1** at time s lies in $\mathcal{M}$, then **PI4** ensures that for all $j \le k_1$ we have $\pi_s(\phi_s, j) \ne \pm 1$. Therefore, $\mathcal{N}_s = \mathcal{N}_{s\phi_s j} = 0$.

Case 4: *None of the above occurs.* As $\pi_{s-1}(i,j) = -1$ for all $(i,j) \in \mathcal{M}\times[k]$, given $\mathcal{F}_{s-1}$ the variables $(|\mathbf{\Phi}_{ij}|)_{(i,j)\in\mathcal{M}\times[k]}$ are mutually independent and uniformly distributed over $V \setminus Z_{s-1}$ by Fact 4.3. They are also independent of the choice of z_s, because $\phi_s \notin \mathcal{M}$. Furthermore, by **PI4** we have $\mathcal{N}_{sij} = 1$ only if $|\mathbf{\Phi}_{ij}| \ne z_s$. This event occurs for all $(i,j) \in \mathcal{M}\times[k]$ independently with probability $1 - |V \setminus Z_{s-1}|^{-1} \le 1 - 1/n$. Consequently, $\mathrm{E}\left[\mathcal{N}_s|\mathcal{F}_{s-1}\right] \le (1-1/n)^{k\mu}$, whence (4.5) follows.

For any $0 \le s \le t$ the random variable $\mathcal{N}_s$ is $\mathcal{F}_s$-measurable because π_s is (by Fact 4.2). Therefore, Lemma 2.4 implies in combination with (4.5) that

$$\mathrm{E}\left[\prod_{s=1}^{t} \mathcal{N}_s|\mathcal{F}_0\right] \le (1-1/n)^{kt\mu} \le \exp(-kt\mu/n). \tag{4.6}$$

Combining (4.3) with (4.4) and (4.6), we obtain

$$\begin{aligned}\mathrm{P}\left[\mathcal{E}_{\mathcal{M}}\right] &= \mathrm{E}\left[\prod_{s=0}^{t} \mathcal{N}_s\right] = \mathrm{E}\left[\mathcal{N}_0 \cdot \mathrm{E}\left[\prod_{s=1}^{t} \mathcal{N}_s|\mathcal{F}_0\right]\right] \\ &\le \mathrm{E}\left[\mathcal{N}_0\right] \cdot \exp(-kt\mu/n) = \lambda^{\mu}, \quad \text{where } \lambda = 2^{-k}\exp(-kt/n).\end{aligned}$$

Since this bound holds for any $\mathcal{M} \subset [m]$ of size μ, Lemma 2.2 implies that

$$\mathrm{P}\left[\sum_{i=1}^{m} X_i > 2\lambda m\right] = o(1/n).$$

As $2\lambda m \le 2n\omega\exp(-kt/n)/k$, this yields (4.2) and thus the assertion. ∎

Corollary 4.5. *W.h.p. we have* $T < 4nk^{-1}\ln\omega$.

Proof. Let $t_0 = \lfloor 2nk^{-1}\ln\omega\rfloor$ and let I_t be the number of indices i such that $\pi_t(i,j) = -1$ for all $1 \le j \le k$. Then **PI2** ensures that $I_t \le I_{t-1} - 1$ for all $t \le T$. Consequently, if $T \ge 2t_0$, then $0 \le I_T \le I_{t_0} - t_0$, and thus $I_{t_0} \ge t_0$. Since $\lfloor 2nk^{-1}\ln\omega\rfloor > 3n\omega\exp(-kt_0/n)/k$ for sufficiently large k, Lemma 4.4 entails

$$\begin{aligned}\mathrm{P}\left[T \ge 2t_0\right] &\le \mathrm{P}\left[I_{t_0} \ge t_0\right] = \mathrm{P}\left[I_{t_0} \ge \lfloor 2nk^{-1}\ln\omega\rfloor\right]\\ &\le \mathrm{P}\left[I_{t_0} > 3n\omega\exp(-kt_0/n)/k\right] = o(1).\end{aligned}$$

Hence, $T < 2t_0$ w.h.p. ∎

For the rest of this section we let

$$\theta = \lfloor 4nk^{-1}\ln\omega\rfloor.$$

The next goal is to estimate the number of Z_t-unique clauses, i.e., the size of the set U_t. For technical reasons we will consider a slightly bigger set: let $\mathcal{U}_t$ be the set of all $i \in [m]$ such that there is an index j such that $\pi_0(i,j) \ne -1$, but there exists no l such that $\pi_t(i,l) \in \{1\} \cup \bar{Z}_t$. That is, clause $\mathbf{\Phi}_i$ contains a positive literal, but by time t there is *at most* one positive literal $\mathbf{\Phi}_{ij} \notin Z_t$ left, and there in no l such that $\mathbf{\Phi}_{il} \in \bar{Z}_t$. This ensures that $U_t \subset \mathcal{U}_t$; for $i \in U_t$ iff there is *exactly one* j such that $\mathbf{\Phi}_{ij}$ is positive but not in Z_t and there in no l such that $\mathbf{\Phi}_{il} \in \bar{Z}_t$. In section 4.2 we will establish the following bound.

Lemma 4.6. *W.h.p. we have* $\max_{0\le t\le T}|U_t| \le \max_{0\le t\le T}|\mathcal{U}_t| \le (1+\varepsilon/3)\omega n$.

Additionally, we need to bound the number of Z_t-unsafe variables, i.e., variables $x \in V \setminus Z_t$ such that $U_t(x) > 0$. This is related to an occupancy problem. Let us think of the variables $x \in V \setminus Z_t$ as "bins" and of the clauses $\mathbf{\Phi}_i$ with $i \in U_t$ as "balls." If we place each ball i into the (unique) bin x such that x occurs positively in $\mathbf{\Phi}_i$, then by Lemma 4.6 the average number of balls per bin is

$$\frac{|U_t|}{|V \setminus Z_t|} \le \frac{(1+\varepsilon/3)\omega}{1-t/n} \qquad \text{w.h.p.}$$

Recall that $\omega = (1-\varepsilon)\ln k$. Corollary 4.5 yields $T \le 4nk^{-1}\ln\omega$ w.h.p. Consequently, for $t \le T$ we have $(1+\varepsilon/3)(1-t/n)^{-1}\omega \le (1-0.6\varepsilon)\ln k$ w.h.p., provided that k is large enough. Hence, if the "balls" were uniformly distributed over the "bins," we would expect

$$|V \setminus Z_t|\exp(-|U_t|/|V \setminus Z_t|) \ge (n-t)k^{0.6\varepsilon-1} \ge nk^{\varepsilon/2-1}$$

"bins" to be empty. The next corollary shows that this is accurate. We defer the proof to section 4.3.

Corollary 4.7. *Let* $\mathcal{Q}_t = |\{x \in V \setminus Z_t : U_t(x) = 0\}|$. *Then*

$$\min_{t\le T}\mathcal{Q}_t \ge nk^{\varepsilon/2-1} \ \textit{w.h.p.}$$

Now that we know that for all $t \le T$ there are "a lot" of variables $x \in V \setminus Z_{t-1}$ such that $U_t(x) = 0$ w.h.p., we can prove that it is quite likely that the clause $\mathbf{\Phi}_{\phi_t}$ selected at time t contains one. More precisely, we have the following.

Corollary 4.8. *Let*

$$\mathcal{B}_t = \begin{cases} 1 & \textit{if } \min_{1\le j<k_1} U_{t-1}(|\mathbf{\Phi}_{\phi_t j}|) > 0,\\ & \qquad \mathcal{Q}_{t-1} \ge nk^{\varepsilon/2-1},\ |U_{t-1}| \le (1+\varepsilon/3)\omega n,\ \textit{and } T \ge t,\\ 0 & \textit{otherwise.}\end{cases}$$

Then $\mathcal{B}_t$ *is* $\mathcal{F}_t$*-measurable and* $\mathrm{E}[\mathcal{B}_t|\mathcal{F}_{t-1}] \leq \exp(-k^{\varepsilon/6})$ *for all* $1 \leq t \leq \theta$.

In words, $\mathcal{B}_t = 1$ indicates that the clause $\mathbf{\Phi}_{\phi_t}$ processed at time t does not contain a Z_{t-1}-safe variable ("$\min_{1\leq j<k_1} U_{t-1}(|\mathbf{\Phi}_{\phi_t j}|) > 0$"), although there are plenty such variables ("$\mathcal{Q}_{t-1} \geq nk^{\varepsilon/2-1}$") and the number of Z_{t-1}-unique clauses is small ("$|U_{t-1}| \leq (1+\varepsilon/3)\omega n$").

Proof of Corollary 4.8. Since the event $T < t$ and the random variable $\mathcal{Q}_{t-1}$ are $\mathcal{F}_{t-1}$-measurable and as $U_{t-1}(|\mathbf{\Phi}_{\phi_t j}|)$ is $\mathcal{F}_t$-measurable for any $j < k_1$ by Fact 4.2, $\mathcal{B}_t$ is $\mathcal{F}_t$-measurable. Let Φ be such that $T[\Phi] \geq t$, $\mathcal{Q}_{t-1}[\Phi] \geq nk^{\varepsilon/2-1}$, and $|U_{t-1}[\Phi]| \leq (1+\varepsilon/3)\omega n$. We condition on the event $\mathbf{\Phi} \equiv_{t-1} \Phi$. Then at time t the process **PI1**–**PI4** selects ϕ_t such that $\pi_{t-1}(\phi_t, j) = -1$ for all $j \in [k]$. Hence, by Fact 4.3 the variables $|\mathbf{\Phi}_{\phi_t j}|$ are uniformly distributed and mutually independent elements of $V \setminus Z_{t-1}$. Consequently, for each $j < k_1$ the event $U_{t-1}(|\mathbf{\Phi}_{\phi_t j}|) = 0$ occurs with probability $|\mathcal{Q}_{t-1}|/|V \setminus Z_{t-1}| \geq k^{\varepsilon/2-1}$ independently. Thus, the probability that $U_{t-1}(|\mathbf{\Phi}_{\phi_t j}|) > 0$ for *all* $j < k_1$ is at most $(1-k^{\varepsilon/2-1})^{k_1-1}$. Finally, provided that $k \geq k_0(\varepsilon)$ is sufficiently large, we have $(1-k^{\varepsilon/2-1})^{k_1-1} \leq \exp(-k^{\varepsilon/6})$. ∎

Proof of Proposition 3.2. The definition of the process **PI1**–**PI4** mirrors the execution of the algorithm; i.e., the set Z obtained after steps 1a–1d of `Fix` equals the set Z_T. Therefore, the first item of Proposition 3.2 is an immediate consequence of Corollary 4.5 and the fact that $|Z_t| = t$ for all $t \leq T$. Furthermore, the second assertion follows directly from Lemma 4.6 and the fact that $|U_t| \leq |\mathcal{U}_t|$ equals the number of Z_t-unique clauses.

To prove the third claim, we need to bound the number of clauses that are unsatisfied under the assignment σ_{Z_T} that sets all variables in $V \setminus Z_T$ to true and all variables in Z_T to false. By construction any all-negative clause contains a variable from Z_T and is thus satisfied under σ_{Z_T} (cf. **PI1**). We claim that for any $i \in [m]$ such that $\mathbf{\Phi}_i$ is unsatisfied under σ_{Z_T} one of the following is true.

(a) There is $1 \leq t \leq T$ such that $i \in U_{t-1}$ and z_t occurs positively in $\mathbf{\Phi}_i$.

(b) There are $1 \leq j_1 < j_2 \leq k$ such that $\mathbf{\Phi}_{ij_1} = \mathbf{\Phi}_{ij_2}$.

To see this, assume that $\mathbf{\Phi}_i$ is unsatisfied under σ_{Z_T} and (b) does not occur. Let us assume without loss of generality that $\mathbf{\Phi}_{i1}, \ldots, \mathbf{\Phi}_{il}$ are positive and $\mathbf{\Phi}_{il+1}, \ldots, \mathbf{\Phi}_{ik}$ are negative for some $l \geq 1$. Since $\mathbf{\Phi}_i$ is unsatisfied under σ_{Z_T}, we have $\mathbf{\Phi}_{i1}, \ldots, \mathbf{\Phi}_{il} \in Z_T$ and $\mathbf{\Phi}_{il+1}, \ldots, \mathbf{\Phi}_{ik} \notin \bar{Z}_T$. Hence, for each $1 \leq j \leq l$ there is $t_j \leq T$ such that $\mathbf{\Phi}_{ij} = z_{t_j}$. As $\mathbf{\Phi}_{i1}, \ldots, \mathbf{\Phi}_{ik}$ are distinct, the indices $t_1, \ldots, t_l$ are mutually distinct, too. Assume that $t_1 < \cdots < t_l$, and let $t_0 = 0$. Then $\mathbf{\Phi}_i$ contains precisely one positive literal from $V \setminus Z_{t_{l-1}}$. Hence, $i \in U_{t_{l-1}}$. Since $\mathbf{\Phi}_i$ is unsatisfied under σ_{Z_T}, no variable from Z_T occurs negatively in $\mathbf{\Phi}_i$, and thus $i \in U_s$ for all $t_{l-1} \leq s < t_l$. Therefore, $i \in U_{t_l-1}$ and $z_{t_l} = \mathbf{\Phi}_{il}$; i.e., (a) occurs.

Let $\mathcal{X}$ be the number of indices $i \in [m]$ for which (a) occurs. We claim that

$$\mathcal{X} \leq n\exp(-k^{\varepsilon/7}) \qquad \text{w.h.p.} \tag{4.7}$$

Since the number of $i \in [m]$ for which (b) occurs is $O(\ln n)$ w.h.p. by Lemma 2.3, (4.7) implies the third assertion in Proposition 3.2.

To establish (4.7), let $\mathcal{B}_t$ be as in Corollary 4.8 and set

$$\mathcal{D}_t = \begin{cases} U_{t-1}(z_t) & \text{if } \mathcal{B}_t = 1 \text{ and } U_{t-1}(z_t) \leq \ln^2 n, \\ 0 & \text{otherwise.} \end{cases}$$

Then by the definition of the random variables $\mathcal{B}_t, \mathcal{D}_t$, either $\mathcal{X} \leq \sum_{1\leq t\leq\theta} \mathcal{D}_t$ or one of the following events occurs:

(i) $T > \theta$.
(ii) $\mathcal{Q}_t < nk^{\varepsilon/2-1}$ for some $0 \le t \le T$.
(iii) $|U_t| > (1+\varepsilon/3)\omega n$ for some $1 \le t \le T$.
(iv) $|U_{t-1}(z_t)| > \ln^2 n$ for some $1 \le t \le \theta$.

The probability of (i) is $o(1)$ by Corollary 4.5. Moreover, (ii) does not occur w.h.p. by Corollary 4.7, and the probability of (iii) is $o(1)$ by Lemma 4.6. If (iv) occurs, then the variable z_t occurs in at least $\ln^2 n$ clauses for some $1 \le t \le \theta$, which has probability $o(1)$ by Lemma 2.3. Hence, we have shown that

$$\mathcal{X} \le \sum_{1\le t\le\theta} \mathcal{D}_t \quad \text{w.h.p.} \tag{4.8}$$

Thus, we need to bound $\sum_{1\le t\le\theta} \mathcal{D}_t$. By Fact 4.2 and Corollary 4.8 the random variable $\mathcal{D}_t$ is $\mathcal{F}_t$-measurable. Let $\bar{\mathcal{D}}_t = \mathrm{E}\left[\mathcal{D}_t|\mathcal{F}_{t-1}\right]$ and $\mathcal{M}_t = \sum_{s=1}^t \mathcal{D}_s - \bar{\mathcal{D}}_s$. Then $(\mathcal{M}_t)_{0\le t\le\theta}$ is a martingale with $\mathcal{M}_0 = 0$. As all increments $\mathcal{D}_s - \bar{\mathcal{D}}_s$ are bounded by $\ln^2 n$ in absolute value by the definition of $\mathcal{D}_t$, Lemma 2.5 (Azuma–Hoeffding) entails that $\mathcal{M}_\theta = o(n)$ w.h.p. Hence, we have

$$\sum_{1\le t\le\theta} \mathcal{D}_t = o(n) + \sum_{1\le t\le\theta} \bar{\mathcal{D}}_t \quad \text{w.h.p.} \tag{4.9}$$

We claim that

$$\bar{\mathcal{D}}_t \le 2\omega \exp(-k^{\varepsilon/6}) \qquad \forall\, 1 \le t \le \theta. \tag{4.10}$$

For by Corollary 4.8 we have

$$\mathrm{E}\left[\mathcal{B}_t|\mathcal{F}_{t-1}\right] \le \exp(-k^{\varepsilon/6}) \qquad \forall\, 1 \le t \le \theta. \tag{4.11}$$

Moreover, if $\mathcal{B}_t = 1$, then **PI2** sets $z_t = |\mathbf{\Phi}_{\phi_t k_1}|$. The index ϕ_t is chosen so that $\pi_{t-1}(\phi_t, j) = -1$ for all $j \in [k]$. Therefore, given $\mathcal{F}_{t-1}$, the variable $z_t = \mathbf{\Phi}_{\phi_t k_1}$ is uniformly distributed over $V \setminus Z_{t-1}$ by Fact 4.3. Hence,

$$\bar{\mathcal{D}}_t \le \mathrm{E}\left[\mathcal{B}_t|\mathcal{F}_{t-1}\right] \cdot \sum_{x\in V\setminus Z_{t-1}} \frac{U_{t-1}(x)}{|V \setminus Z_{t-1}|} = \frac{|U_{t-1}| \cdot \mathrm{E}\left[\mathcal{B}_t|\mathcal{F}_{t-1}\right]}{|V \setminus Z_{t-1}|}.$$

Furthermore, $\mathcal{B}_t = 1$ implies $|U_{t-1}| \le (1+\varepsilon/3)\omega n$. Consequently, for $k \ge k_0(\varepsilon)$ large enough we get

$$\bar{\mathcal{D}}_t \le \frac{(1+\frac{\varepsilon}{3})\omega n \cdot \mathrm{E}\left[\mathcal{B}_t|\mathcal{F}_{t-1}\right]}{n-t} \le \frac{(1+\frac{\varepsilon}{3})\omega n \cdot \mathrm{E}\left[\mathcal{B}_t|\mathcal{F}_{t-1}\right]}{n-\theta} \le 2\omega \mathrm{E}\left[\mathcal{B}_t|\mathcal{F}_{t-1}\right]. \tag{4.12}$$

Combining (4.11) and (4.12), we obtain (4.10). Further, plugging (4.10) into (4.9) and assuming that $k \ge k_0(\varepsilon)$ is large enough, we get

$$\sum_{1\le t\le\theta} \mathcal{D}_t = 2\omega \exp(-k^{\varepsilon/6})\theta + o(n) \le 3\omega \exp(-k^{\varepsilon/6})\theta \le n \exp(-k^{\varepsilon/7}) \qquad \text{w.h.p.}$$

Thus, (4.7) follows from (4.8). ∎

4.2. Proof of Lemma 4.6. For integers $t \geq 1$, $i \in [m]$, $j \in [k]$, let

$$\mathcal{H}_{tij} = \begin{cases} 1 & \text{if } \pi_{t-1}(i,j) = 1 \text{ and } \pi_t(i,j) = z_t, \\ 0 & \text{otherwise,} \end{cases} \tag{4.13}$$

$$\mathcal{S}_{tij} = \begin{cases} 1 & \text{if } T \geq t \text{ and } \pi_t(i,j) \in \{1,-1\}, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $\mathcal{H}_{tij} = 1$ indicates that the variable underlying the positive literal $\mathbf{\Phi}_{ij}$ is the variable z_t set to false at time t and that $\mathbf{\Phi}_{ij}$ did not get revealed before. Moreover, $\mathcal{S}_{tij} = 1$ means that the variable underlying $\mathbf{\Phi}_{ij}$ has not been revealed up to time t. In particular, it does not belong to the set Z_t of variables set to false.

LEMMA 4.9. *For any two sets $\mathcal{I}, \mathcal{J} \subset [\theta] \times [m] \times [k]$ we have*

$$\mathrm{E}\left[\prod_{(t,i,j)\in\mathcal{I}} \mathcal{H}_{tij} \cdot \prod_{(t,i,j)\in\mathcal{J}} \mathcal{S}_{tij} | \mathcal{F}_0\right] \leq (n-\theta)^{-|\mathcal{I}|} (1-1/n)^{|\mathcal{J}|}.$$

Proof. Let $1 \leq t \leq \theta$. Let $\mathcal{I}_t = \{(i,j) : (t,i,j) \in \mathcal{I}\}$, $\mathcal{J}_t = \{(i,j) : (t,i,j) \in \mathcal{J}\}$, and

$$X_t = \prod_{(i,j)\in\mathcal{I}_t} \mathcal{H}_{tij} \cdot \prod_{(i,j)\in\mathcal{J}_t} \mathcal{S}_{tij}.$$

If $X_t = 1$, then either $\mathcal{I}_t \cup \mathcal{J}_t = \emptyset$ or $t \leq T$; for if $t > T$, then $\mathcal{S}_{tij} = 0$ by definition and $\mathcal{H}_{tij} = 0$ because $\pi_t = \pi_{t-1}$. Furthermore, $X_t = 1$ implies that

$$\pi_{t-1}(i,j) = 1 \quad \forall (i,j) \in \mathcal{I}_t \quad \text{and} \quad \pi_{t-1}(i,j) \in \{-1,1\} \quad \forall (i,j) \in \mathcal{J}_t. \tag{4.14}$$

Thus, let Φ be a k-CNF such that $T[\Phi] \geq t$ and $\pi_{t-1}[\Phi]$ satisfies (4.14). We claim that

$$\mathrm{E}[X_t|\mathcal{F}_{t-1}](\Phi) \leq (n-\theta)^{-|\mathcal{I}_t|}(1-1/n)^{|\mathcal{J}_t|}. \tag{4.15}$$

To show this, we condition on the event $\mathbf{\Phi} \equiv_{t-1} \Phi$. Then at time t steps **PI1–PI2** select a variable z_t from the all-negative clause $\mathbf{\Phi}_{\phi_t}$. Since for any $(i,j) \in \mathcal{I}_t$ the literal $\mathbf{\Phi}_{ij}$ is positive, we have $\phi_t \neq i$. Furthermore, we may assume that if $(\phi_t, j) \in \mathcal{J}_t$, then $j > k_1$, because otherwise $\pi_t(i,j) = \mathbf{\Phi}_{ij}$ and hence $X_t = \mathcal{S}_{t\phi_t j} = 0$ (cf. **PI4**). Thus, due to (4.14) and Fact 4.3 in the conditional distribution $\mathrm{P}[\cdot|\mathcal{F}_{t-1}](\Phi)$, the variables $(|\mathbf{\Phi}_{ij}|)_{(i,j)\in\mathcal{I}_t\cup\mathcal{J}_t}$ are uniformly distributed over $V \setminus Z_{t-1}$ and mutually independent. Therefore, the events $|\mathbf{\Phi}_{ij}| = z_t$ occur independently with probability $1/|V \setminus Z_{t-1}| = 1/(n-t+1)$ for $(i,j) \in \mathcal{I}_t \cup \mathcal{J}_t$, whence

$$\mathrm{E}[X_t|\mathcal{F}_{t-1}](\Phi) \leq (n-t+1)^{-|\mathcal{I}_t|}(1-1/(n-t+1))^{|\mathcal{J}_t|} \leq (n-\theta)^{-|\mathcal{I}_t|}(1-1/n)^{|\mathcal{J}_t|}.$$

This shows (4.15). Finally, combining (4.15) and Lemma 2.4, we obtain

$$\begin{aligned} \mathrm{E}\left[\prod_{(t,i,j)\in\mathcal{I}} \mathcal{H}_{tij} \cdot \prod_{(t,i,j)\in\mathcal{J}} \mathcal{S}_{tij} | \mathcal{F}_0\right] &= \mathrm{E}\left[\prod_{t=1}^{\theta} X_t | \mathcal{F}_0\right] \\ &\leq \prod_{t=1}^{\theta} (n-\theta)^{-|\mathcal{I}_t|}(1-1/n)^{|\mathcal{J}_t|} = (n-\theta)^{-|\mathcal{I}|} (1-1/n)^{|\mathcal{J}|}, \end{aligned}$$

as desired. ∎

Armed with Lemma 4.9, we can now bound the number of indices $i \in \mathcal{U}_t$ such that $\mathbf{\Phi}_i$ has "few" positive literals. Recall that $i \in \mathcal{U}_t$ iff $\mathbf{\Phi}_i$ has $l \geq 1$ positive literals of which (at least) $l-1$ lie in Z_t while no variable from Z_t occurs negatively in $\mathbf{\Phi}_i$.

LEMMA 4.10. *Let $1 \leq l < \sqrt{k}$ and $1 \leq t \leq \theta$. Moreover, let*

$$\Lambda_l(t) = \omega \binom{k-1}{l-1} \left(\frac{t}{n}\right)^{l-1} \left(1 - \frac{t}{n}\right)^{k-l}.$$

With probability $1 - o(1/n)$ either $T < t$ or there are at most $(1+\varepsilon/9)\Lambda_l(t)n$ indices $i \in \mathcal{U}_t$ such that $\mathbf{\Phi}_i$ has precisely l positive literals.

Proof. Fix $1 \leq t \leq \theta$. For $i \in [m]$ let

$$X_i = \begin{cases} 1 & \text{if } T \geq t,\ \mathbf{\Phi}_i \text{ has exactly } l \text{ positive literals, and } i \in \mathcal{U}_t, \\ 0 & \text{otherwise.} \end{cases}$$

Our task is to bound $\sum_{i \in [m]} X_i$. To do so we are going to apply Lemma 2.2. Thus, let $\mu = \lceil \ln^2 n \rceil$, let $\mathcal{M} \subset [m]$ be a set of size μ, and let $\mathcal{E}_{\mathcal{M}}$ be the event that $X_i = 1$ for all $i \in \mathcal{M}$. Furthermore, let $P_i \subset [k]$ be a set of size $l-1$ for each $i \in \mathcal{M}$, and let $\mathcal{P} = (P_i)_{i \in \mathcal{M}}$ be the family of all sets P_i. In addition, let $t_i : P_i \to [t]$ for all $i \in \mathcal{M}$, and let $\mathcal{T} = (t_i)_{i \in \mathcal{M}}$ comprise all maps t_i. Let $\mathcal{E}_{\mathcal{M}}(\mathcal{P}, \mathcal{T})$ be the event that the following statements are true:

(a) $\mathbf{\Phi}_i$ has exactly l positive literals for all $i \in \mathcal{M}$.

(b) $\mathbf{\Phi}_{ij} = z_{t_i(j)}$ and $\pi_{t_i(j)-1}(i,j) = 1$ for all $i \in \mathcal{M}$ and $j \in P_i$.

(c) $T \geq t$, and no variable from Z_t occurs negatively in $\mathbf{\Phi}_i$.

If the event $\mathcal{E}_{\mathcal{M}}$ occurs, then there exist $\mathcal{P}, \mathcal{T}$ such that $\mathcal{E}_{\mathcal{M}}(\mathcal{P}, \mathcal{T})$ occurs. Hence, in order to bound the probability of $\mathcal{E}_{\mathcal{M}}$ we will bound the probabilities of the events $\mathcal{E}_{\mathcal{M}}(\mathcal{P}, \mathcal{T})$ and apply the union bound.

To bound the probability of $\mathcal{E}_{\mathcal{M}}(\mathcal{P}, \mathcal{T})$, let

$$\begin{aligned} \mathcal{I} &= \mathcal{I}_{\mathcal{M}}(\mathcal{P}, \mathcal{T}) = \{(s,i,j) : i \in \mathcal{M}, j \in P_i, s = t_i(j)\}, \\ \mathcal{J} &= \mathcal{J}_{\mathcal{M}}(\mathcal{P}, \mathcal{T}) = \{(s,i,j) \in [t] \times \mathcal{M} \times [k] : \pi_0(i,j) = -1\}. \end{aligned}$$

Let $Y_i = 1$ if clause $\mathbf{\Phi}_i$ has exactly l positive literals, including the $l-1$ literals $\mathbf{\Phi}_{ij}$ for $j \in P_i$ $(i \in \mathcal{M})$. Then $\mathrm{P}[Y_i = 1] = (k-l+1)2^{-k}$ for each $i \in \mathcal{M}$. Moreover, the events $Y_i = 1$ for $i \in \mathcal{M}$ are mutually independent and $\mathcal{F}_0$-measurable. Therefore, by Lemma 4.9

$$\begin{aligned} \mathrm{P}[\mathcal{E}_{\mathcal{M}}(\mathcal{P}, \mathcal{T})] &\leq \mathrm{E}\left[\mathrm{E}\left[\prod_{(t,i,j) \in \mathcal{I}} \mathcal{H}_{tij} \cdot \prod_{(t,i,j) \in \mathcal{J}} \mathcal{S}_{tij} | \mathcal{F}_0\right] \cdot \prod_{i \in \mathcal{M}} Y_i\right] \\ &\leq \left[\frac{k-l+1}{2^k} \cdot (n-t)^{1-l} \left(1 - \frac{1}{n}\right)^{(k-l)t}\right]^{\mu}. \end{aligned} \tag{4.16}$$

For each $i \in \mathcal{M}$ there are $\binom{k}{l-1}$ ways to choose a set P_i and then t^{l-1} ways to choose the map t_i. Therefore, the union bound and (4.16) yield

$$\mathrm{P}[\mathcal{E}_{\mathcal{M}}] \leq \sum_{\mathcal{P}, \mathcal{T}} \mathrm{P}[\mathcal{E}_{\mathcal{M}}(\mathcal{P}, \mathcal{T})] \leq \lambda^{\mu},$$

where

$$\lambda = \binom{k}{l-1} t^{l-1} \cdot \frac{k-l+1}{2^k} \cdot (n-t)^{1-l} \left(1 - \frac{1}{n}\right)^{(k-l)t}.$$

Hence, by Lemma 2.2 with probability $1-o(1/n)$ we have $\sum_{i\in[m]} X_i \leq (1+10^{-6}\varepsilon)\lambda m$. In other words, with probability $1-o(1/n)$ either $T < t$ or there are at most $(1+10^{-6}\varepsilon)\lambda m$ indices $i \in [m]$ such that $\mathbf{\Phi}_i$ has precisely l positive literals and $i \in \mathcal{U}_t$. Thus, the remaining task is to show that

$$\lambda m \leq (1+\varepsilon/10)\Lambda_l(t)n. \tag{4.17}$$

To show (4.17), we estimate

$$\begin{aligned}\lambda m &\leq m \cdot k2^{-k} \cdot \binom{k-1}{l-1}\left(\frac{t}{n-t}\right)^{l-1}\left(1-\frac{1}{n}\right)^{t(k-1-(l-1))} \\ &\leq m \cdot k2^{-k} \cdot \binom{k-1}{l-1}\left(\frac{t}{n}\right)^{l-1}\left(1-\frac{t}{n}\right)^{k-1-(l-1)}\eta, \qquad (4.18)\end{aligned}$$

where we let

$$\eta = \left(\frac{n}{n-t}\right)^{l-1} \cdot \left(\frac{(1-1/n)^t}{1-t/n}\right)^{k-l}.$$

Hence, (4.18) shows that

$$\lambda m \leq n \cdot \Lambda_l(t) \cdot \eta. \tag{4.19}$$

We can bound η as follows:

$$\begin{aligned}\eta &\leq (1+t/(n-t))^l\left(\frac{\exp(-t/n)}{\exp(-t/n-(t/n)^2)}\right)^{k-l} \leq (1+2t/n)^l \exp(k(t/n)^2) \\ &\leq \exp(2l\theta/n + k(\theta/n)^2) \leq \exp(8lk^{-1}\ln\omega + 16k^{-1}\ln^2\omega).\end{aligned}$$

Since $l \leq \sqrt{k}$ and $\omega \leq \ln k$, the last expression is less than $1+\varepsilon/10$ for sufficiently large $k \geq k_0(\varepsilon)$. Hence, $\eta \leq 1+\varepsilon/10$, and thus (4.17) follows from (4.19). ∎

The following lemma deals with $i \in \mathcal{U}_t$ such that $\mathbf{\Phi}_i$ contains "a lot" of positive literals.

LEMMA 4.11. *W.h.p. the following is true for all $l \geq \ln k$. There are at most $n\exp(-l)$ indices $i \in [m]$ such that $\mathbf{\Phi}_i$ has exactly l positive literals among which at least $l-1$ are in Z_θ.*

Proof. For any $i \in [m]$ we let

$$X_i = \begin{cases} 1, & \mathbf{\Phi}_i \text{ has exactly } l \text{ positive literals among which } l-1 \text{ are in } Z_\theta, \\ 0 & \text{otherwise.}\end{cases}$$

Let $\mathcal{M} \subset [m]$ be a set of size $\mu = \lceil \ln^2 n \rceil$, and let $\mathcal{E}_\mathcal{M}$ be the event that $X_i = 1$ for all $i \in \mathcal{M}$. Furthermore, let $P_i \subset [k]$ be a set of size $l-1$ for each $i \in \mathcal{M}$. Let $t_i : P_i \to [\theta]$ for each $i \in \mathcal{M}$, and set $\mathcal{T} = (t_i)_{i\in\mathcal{M}}$. Let $\mathcal{E}_\mathcal{M}(\mathcal{P},\mathcal{T})$ be the event that the following two statements are true for all $i \in \mathcal{M}$:

(a) $\mathbf{\Phi}_i$ has exactly l positive literals.

(b) For all $j \in P_i$ we have $\mathbf{\Phi}_{ij} = z_{t_i(j)}$ and $\pi_{t_i(j)-1}(i,j) = 1$.

If $\mathcal{E}_\mathcal{M}$ occurs, then there are $\mathcal{P},\mathcal{T}$ such that $\mathcal{E}_\mathcal{M}(\mathcal{P},\mathcal{T})$ occurs. Hence, we will use the union bound.

For $i \in \mathcal{M}$ we let $Y_i = 1$ if clause $\mathbf{\Phi}_i$ has exactly l positive literals, including the literals $\mathbf{\Phi}_{ij}$ for $j \in P_i$. Set $\mathcal{I} = \{(s,i,j) : i \in \mathcal{M},\ j \in P_i,\ s = t_i(j)\}$. If $\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})$ occurs, then

$$\prod_{(s,i,j)\in\mathcal{I}} \mathcal{H}_{sij} \cdot \prod_{i\in\mathcal{M}} Y_i = 1.$$

As in the proof of Lemma 4.10 we have $\mathrm{E}\left[\prod_{i\in\mathcal{M}} Y_i\right] \le ((k-l+1)/2^k)^{\mu}$. Moreover, bounding $\mathrm{E}\left[\prod_{(s,i,j)\in\mathcal{I}} \mathcal{H}_{sij}|\mathcal{F}_0\right]$ via Lemma 4.9 and taking into account that $\prod_{i\in\mathcal{M}} Y_i$ is $\mathcal{F}_0$-measurable, we obtain

$$\mathrm{P}\left[\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})\right] \le \mathrm{E}\left[\prod_{i\in\mathcal{M}} Y_i\right] \cdot \left\|\mathrm{E}\left[\prod_{(s,i,j)\in\mathcal{I}} \mathcal{H}_{sij}|\mathcal{F}_0\right]\right\|_{\infty} \le \left[\frac{k-l+1}{2^k}\cdot(n-\theta)^{1-l}\right]^{\mu}.$$

Hence, by the union bound

$$\mathrm{P}\left[\mathcal{E}_{\mathcal{M}}\right] \le \mathrm{P}\left[\exists \mathcal{P},\mathcal{T} : \mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T}) \text{ occurs}\right] \le \sum_{\mathcal{P},\mathcal{T}} \mathrm{P}\left[\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})\right] \le \lambda^{\mu},$$

where

$$\lambda = \binom{k}{l-1}\theta^{l-1}\cdot\frac{k-l+1}{2^k}\cdot(n-\theta)^{1-l}.$$

Lemma 2.2 implies that $\sum_{i\in[m]} X_i \le 2\lambda m$ w.h.p. That is, w.h.p. there are at most $2\lambda m$ indices $i \in [m]$ such that $\mathbf{\Phi}_i$ has exactly l positive literals of which $l-1$ lie in Z_θ. Thus, the estimate

$$\begin{aligned}
2\lambda m &\le \frac{2^{k+1}\omega n}{k}\cdot\binom{k}{l-1}\cdot\frac{k-l+1}{2^k}\cdot\left(\frac{\theta}{n-\theta}\right)^{l-1}\\
&\le 2\omega n\cdot\left(\frac{\mathrm{e}k\theta}{(l-1)(n-\theta)}\right)^{l-1} \le 2\omega n\left(\frac{12\ln\omega}{l}\right)^{l-1} \qquad (\text{as } \theta = 4nk^{-1}\ln\omega)\\
&\le n\exp(-l) \qquad (\text{because } l \ge \ln k \ge \omega)
\end{aligned}$$

completes the proof. $\square$

Proof of Lemma 4.6. Since $T \le \theta$ w.h.p. by Corollary 4.5, it suffices to show that w.h.p. for all $0 \le t \le \min\{T,\theta\}$ the bound $|\mathcal{U}_t| \le (1+\varepsilon/3)\omega n$ holds. Let $\mathcal{U}_{tl}$ be the number of indices $i \in \mathcal{U}_t$ such that $\mathbf{\Phi}_i$ has precisely l positive literals. Then Lemmas 4.10 and 4.11 imply that w.h.p. for all $t \le \min\{T,\theta\}$ and all $1 \le l \le k$ simultaneously

$$\mathcal{U}_{tl} \le \begin{cases} n\exp(-l) & \text{if } l \ge \sqrt{k},\\ (1+\varepsilon/9)\Lambda_l(t) & \text{otherwise.}\end{cases}$$

Therefore, assuming that $k \geq k_0(\varepsilon)$ is sufficiently large, we see that w.h.p.

$$\begin{aligned}\max_{0\leq t\leq\min\{T,\theta\}}|\mathcal{U}_t| &\leq \max_{0\leq t\leq\min\{T,\theta\}}\sum_{l=1}^{k}\mathcal{U}_{tl}\\ &\leq nk\exp(-\sqrt{k}) + \max_{0\leq t\leq\min\{T,\theta\}}\sum_{1\leq l\leq\sqrt{k}}\left(1+\frac{\varepsilon}{9}\right)\Lambda_l(t)n\\ &\leq n+\left(1+\frac{\varepsilon}{9}\right)\omega n\\ &\qquad\cdot\max_{0\leq t\leq\min\{T,\theta\}}\sum_{1\leq l\leq\sqrt{k}}\binom{k-1}{l-1}\left(\frac{t}{n}\right)^{l-1}\left(1-\frac{t}{n}\right)^{(k-1)-(l-1)}\\ &\leq\left(1+\frac{\varepsilon}{3}\right)\omega n,\end{aligned}$$

as desired. ∎

4.3. Proof of Corollary 4.7. Define a map $\psi_t : \mathcal{U}_t \to V$ as follows. For $i \in \mathcal{U}_t$ let s be the least index such that $i \in \mathcal{U}_s$; if there is j such that $\mathbf{\Phi}_{ij} \in V \setminus Z_s$, let $\psi_t(i) = \mathbf{\Phi}_{ij}$, and otherwise let $\psi_t(i) = z_s$. The idea is that $\psi_t(i)$ is the unique positive literal of $\mathbf{\Phi}_i$ that is not assigned false at the time s when the clause became Z_s-unique. The following lemma shows that the (random) map ψ_t is not too far from being "uniformly distributed."

LEMMA 4.12. *Let $t \geq 0$, $\hat{\mathcal{U}}_t \subset [m]$, and $\hat{\psi}_t : \hat{\mathcal{U}}_t \to V$. Then*

$$\mathrm{P}\left[\psi_t = \hat{\psi}_t | \mathcal{U}_t = \hat{\mathcal{U}}_t\right] \leq (n-t)^{-|\hat{\mathcal{U}}_t|}.$$

The precise proof of Lemma 4.12 is a little intricate, but the lemma itself is very plausible. If clause $\mathbf{\Phi}_i$ becomes Z_s-unique at time s, then there is a unique index j such that $\mathbf{\Phi}_{ij} \in V \setminus Z_s$. Moreover, $\pi_{s-1}(i,j) = 1$; i.e., the literal $\mathbf{\Phi}_{ij}$ has not been "revealed" before time s. Therefore, Fact 4.3 implies that $\mathbf{\Phi}_{ij}$ is *uniformly* distributed over $V \setminus Z_s$ (given $\mathcal{F}_{s-1}$). Thus, $\psi_t(i) = \mathbf{\Phi}_{ij}$ attains each of $|V \setminus Z_s| = n - s \geq n - t$ possible values with equal probability. Hence, we can think of $\mathbf{\Phi}_i$ as a ball that gets tossed into a uniformly random "bin" $\psi_s(i)$ at time s. But this argument alone does not quite establish Lemma 4.12, because our "ball" may disappear from the game at a later time $s < u \leq t$: if $\mathbf{\Phi}_{il} = \bar{z}_u$ for some $l \in [k]$, then $\mathbf{\Phi}_i$ is not Z_u-unique anymore. However, this event is independent of the bin $\psi_s(i)$ into which the ball was tossed, as it depends only on literals $\mathbf{\Phi}_{il}$ such that $\pi_{u-1}(i,l) = -1$. Let us now give the detailed proof.

Proof of Lemma 4.12. Set $Z_{-1} = \emptyset$. Moreover, define random variables

$$\gamma_t(i,j) = \begin{cases} \pi_t(i,j) & \text{if } \pi_t(i,j) \in \{-1,1\} \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } (i,j) \in [m] \times [k].$$

Thus, γ_t is obtained by just recording *which positions* the process **PI1–PI4** has revealed up to time t, without taking notice of the actual literals $\pi_t(i,j) \in V \cup \bar{V}$ in these positions. We claim that for any $i \in [m]$

$$i \in \mathcal{U}_t \Leftrightarrow \max_{j\in[k]}\gamma_0(i,j) \geq 0 \wedge (\forall j \in [k] : \gamma_t(i,j) = \min\{\gamma_0(i,j), 0\}). \tag{4.20}$$

For $\mathcal{U}_t$ is the set of all $i \in [m]$ such that $\mathbf{\Phi}_i$ contains none of the variables in Z_t negatively and has at most one positive occurrence of a variable from $V \setminus Z_t$. Hence, $i \in \mathcal{U}_t$ iff the following hold:

(a) For any $j \in [k]$ such that $\mathbf{\Phi}_{ij}$ is negative we have $\mathbf{\Phi}_{ij} \notin Z_t$; by **PI4** this is the case iff $\pi_t(i,j) = -1$, and then $\gamma_t(i,j) = -1$.

(b) For any $j \in [k]$ such that $\mathbf{\Phi}_{ij}$ is positive we have $\pi_t(i,j) = \mathbf{\Phi}_{ij}$ and hence $\gamma_t(i,j) = 0$. To see this, assume that $i \in \mathcal{U}_t$. If $\mathbf{\Phi}_{ij} \in Z_t$, then $\pi_t(i,j) = \mathbf{\Phi}_{ij}$ by **PI4**, and hence $\gamma_t(i,j) = 0$. Moreover, if $\mathbf{\Phi}_{ij}$ is the only positive literal of $\mathbf{\Phi}_i$ that does not belong to Z_t, then $i \in U_t$ and hence $\pi_t(i,j) = \mathbf{\Phi}_{ij}$ by **PI4**. Thus, $\gamma_t(i,j) = 0$. Conversely, if $\gamma_t(i,j) = 0$ for all positive $\mathbf{\Phi}_{ij}$, then $\mathbf{\Phi}_i$ has at most one occurrence of a positive variable from $V \setminus Z_t$.

Thus, we have established (4.20).

Fix a set $\hat{\mathcal{U}}_t \subset [m]$, let Φ be any formula such that $\mathcal{U}_t[\Phi] = \hat{\mathcal{U}}_t$, and let $\hat{\gamma}_s = \gamma_s[\Phi]$ for all $s \le t$. Moreover, for $s \le t$ let Γ_s be the event that $\gamma_u = \hat{\gamma}_u$ for all $u \le s$. The goal is to prove that

$$\mathrm{P}\left[\psi_t = \hat{\psi}_t | \Gamma_t\right] \le (n-t)^{-|\hat{\mathcal{U}}_t|}. \tag{4.21}$$

Let $\tau : \hat{\mathcal{U}}_t \to [0,t]$ assign to each $i \in \hat{\mathcal{U}}_t$ the least s such that $i \in \hat{\mathcal{U}}_s$. Intuitively this is the first time s when Φ_i becomes either Z_s-unique or unsatisfied under the assignment σ_{Z_s} that sets the variables in Z_s to false and all others to true. We claim that

$$\mathrm{P}\left[\forall i \in \hat{\mathcal{U}}_t : \psi_t(i) = \hat{\psi}_t(i) | \Gamma_t\right] \le \prod_{i \in \hat{\mathcal{U}}_t} (n - \tau(i))^{-1}. \tag{4.22}$$

Since $\tau(i) \le t$ for all $i \in \hat{\mathcal{U}}_t$, (4.22) implies (4.21) and thus the assertion.

Let τ_s be the event that $\psi_u(i) = \hat{\psi}_t(i)$ for all $0 \le u \le s$ and all $i \in \tau^{-1}(u)$, and let $\tau_{-1} = \Omega_k(n,m)$ be the trivial event. In order to prove (4.22), we will show that for all $0 \le s \le t$

$$\mathrm{P}\left[\tau_s | \tau_{s-1} \cap \Gamma_s\right] \le (n-s)^{-|\tau^{-1}(s)|} \tag{4.23}$$

and

$$\mathrm{P}\left[\tau_s | \tau_{s-1} \cap \Gamma_s\right] = \mathrm{P}\left[\tau_s | \tau_{s-1} \cap \Gamma_t\right]. \tag{4.24}$$

Combining (4.23) and (4.24) yields

$$\begin{aligned}\mathrm{P}\left[\forall i \in \hat{\mathcal{U}}_t : \psi_t(i) = \hat{\psi}_t(i) | \Gamma_t\right] &= \mathrm{P}\left[\tau_t | \Gamma_t\right] = \prod_{0 \le s \le t} \mathrm{P}\left[\tau_s | \tau_{s-1} \cap \Gamma_t\right] \\ &= \prod_{0 \le s \le t} \mathrm{P}\left[\tau_s | \tau_{s-1} \cap \Gamma_s\right] \le \prod_{0 \le s \le t} (n-s)^{-|\tau^{-1}(s)|},\end{aligned}$$

which shows (4.22). Thus, the remaining task is to establish (4.23) and (4.24).

To prove (4.23) it suffices to show that

$$\frac{\mathrm{P}\left[\tau_s \cap \Gamma_s | \mathcal{F}_{s-1}\right](\varphi)}{\mathrm{P}\left[\tau_{s-1} \cap \Gamma_s | \mathcal{F}_{s-1}\right](\varphi)} \le (n-s)^{-|\tau^{-1}(s)|} \qquad \forall\, \varphi \in \tau_{s-1} \cap \Gamma_s. \tag{4.25}$$

Note that the left-hand side is just the conditional probability of τ_s given $\tau_{s-1} \cap \Gamma_s$ with respect to the probability measure $\mathrm{P}\left[\cdot | \mathcal{F}_{s-1}\right](\varphi)$. Thus, let us condition on the event $\mathbf{\Phi} \equiv_{s-1} \varphi \in \tau_{s-1} \cap \Gamma_s$. Then $\mathbf{\Phi} \in \Gamma_s$, and therefore $\gamma_0 = \hat{\gamma}_0$ and $\gamma_s = \hat{\gamma}_s$. Hence, (4.20) entails $\mathcal{U}_s[\mathbf{\Phi}] = \mathcal{U}_s[\varphi] = \mathcal{U}_s[\Phi]$ and thus $\tau^{-1}(s) \subset \mathcal{U}_s[\mathbf{\Phi}]$. Let $i \in \tau^{-1}(s)$, and let $J_i \ne \emptyset$ be the set of indices $j \in [k]$ such that $\gamma_{s-1}(i,j) = 1$. Recall that

$\psi_s(i)$ is defined as follows: if $\mathbf{\Phi}_{ij} = z_s$ for all $j \in J_i$, then $\psi_s(i) = z_s$; otherwise $\psi_s(i) = \mathbf{\Phi}_{ij}$ for the (unique) $j \in J_i$ such that $\mathbf{\Phi}_{ij} \neq z_s$. By Fact 4.3 in the measure $\mathrm{P}\left[\cdot|\mathcal{F}_{s-1}\right](\varphi)$, the variables $(\mathbf{\Phi}_{ij})_{i\in\tau^{-1}(s),j\in J_i}$ are independently uniformly distributed over $V \setminus Z_{s-1}$ (because $\pi_{s-1}(i,j) = \gamma_{s-1}(i,j) = 1$). Hence, the events $\psi_s(i) = \hat{\psi}_t(i)$ occur independently for all $i \in \tau^{-1}(s)$. Thus, letting

$$p_i = \mathrm{P}\left[\psi_s(i) = \hat{\psi}_t(i) \wedge \forall j \in J_i : \gamma_s(i,j) = 0|\mathcal{F}_{s-1}\right](\varphi),$$
$$q_i = \mathrm{P}\left[\forall j \in J_i : \gamma_s(i,j) = 0|\mathcal{F}_{s-1}\right](\varphi)$$

for $i \in \tau^{-1}(s)$, we have

$$\frac{\mathrm{P}\left[\tau_s \cap \Gamma_s|\mathcal{F}_{s-1}\right](\varphi)}{\mathrm{P}\left[\tau_{s-1} \cap \Gamma_s|\mathcal{F}_{s-1}\right](\varphi)} = \prod_{i\in\tau^{-1}(s)} \frac{p_i}{q_i}. \tag{4.26}$$

Observe that the event $\forall j \in J_i : \gamma_s(i,j) = 0$ occurs iff $\mathbf{\Phi}_{ij} = z_s$ for at least $|J_i| - 1$ elements $j \in J_i$ (cf. **PI4**). Therefore,

$$q_i = |J_i| \cdot |V \setminus Z_{s-1}|^{-(|J_i|-1)}(1 - |V \setminus Z_{s-1}|^{-1}) + |V \setminus Z_{s-1}|^{-|J_i|}.$$

To bound p_i for $i \in \tau^{-1}(s)$ we consider three cases.

Case 1: $\hat{\psi}_t(i) \notin V \setminus Z_{s-1}$**.** As $\mathbf{\Phi}_{ij} \in V \setminus Z_{s-1}$ for all $j \in J_i$ the event $\psi_s(i) = \hat{\psi}_t(i)$ has probability 0.

Case 2: $\hat{\psi}_t(i) = z_s$**.** The event $\psi_s(i) = \hat{\psi}_t(i)$ occurs iff $\mathbf{\Phi}_{ij} = z_s$ for all $j \in J_i$, which happens with probability $|V \setminus Z_{s-1}|^{-|J_i|}$ in the measure $\mathrm{P}\left[\cdot|\mathcal{F}_{s-1}\right](\varphi)$. Hence, $p_i = (n - s + 1)^{-|J_i|}$.

Case 3: $\hat{\psi}_t(i) \in V \setminus Z_s$**.** If $\psi_s(i) = \hat{\psi}_t(i)$, then there is $j \in J_i$ such that $\mathbf{\Phi}_{ij} = \hat{\psi}_t(i)$ and $\mathbf{\Phi}_{ij'} = z_s$ for all $j' \in J_s \setminus \{j\}$. Hence, $p_i = |J_i| \cdot |V \setminus Z_{s-1}|^{-|J_i|} = |J_i|(n - s + 1)^{-|J_i|}$.

In all three cases we have

$$\frac{q_i}{p_i} \geq \frac{|J_i|(n - s + 1)^{1-|J_i|}(1 - 1/(n - s + 1))}{|J_i|(n - s + 1)^{-|J_i|}} = n - s.$$

Thus, (4.25) follows from (4.26). This completes the proof of (4.23).

In order to prove (4.24) we will show that for any $0 \leq b \leq c < a$

$$\mathrm{P}\left[\Gamma_a|\tau_b \cap \Gamma_c\right] = \mathrm{P}\left[\Gamma_a|\Gamma_c\right]. \tag{4.27}$$

This implies (4.24) as follows:

$$\begin{aligned}\mathrm{P}\left[\tau_s|\tau_{s-1} \cap \Gamma_t\right] &= \frac{\mathrm{P}\left[\tau_s \cap \Gamma_t\right]}{\mathrm{P}\left[\tau_{s-1} \cap \Gamma_t\right]} = \frac{\mathrm{P}\left[\Gamma_t|\tau_s \cap \Gamma_s\right]\mathrm{P}\left[\tau_s \cap \Gamma_s\right]}{\mathrm{P}\left[\Gamma_t|\tau_{s-1} \cap \Gamma_s\right]\mathrm{P}\left[\tau_{s-1} \cap \Gamma_s\right]} \\ &\overset{(4.27)}{=} \frac{\mathrm{P}\left[\tau_s \cap \Gamma_s\right]}{\mathrm{P}\left[\tau_{s-1} \cap \Gamma_s\right]} = \mathrm{P}\left[\tau_s|\tau_{s-1} \cap \Gamma_s\right].\end{aligned}$$

To show (4.27) it suffices to consider the case $a = c+1$, because for $a > c+1$ we have

$$\begin{aligned}\mathrm{P}\left[\Gamma_a|\tau_b \cap \Gamma_c\right] &= \mathrm{P}\left[\Gamma_a|\tau_b \cap \Gamma_{c+1}\right]\mathrm{P}\left[\tau_b \cap \Gamma_{c+1}|\tau_b \cap \Gamma_c\right] \\ &= \mathrm{P}\left[\Gamma_a|\tau_b \cap \Gamma_{c+1}\right]\mathrm{P}\left[\Gamma_{c+1}|\tau_b \cap \Gamma_c\right].\end{aligned}$$

Thus, suppose that $a = c+1$. At time $a = c+1$ **PI1** selects an index $\phi_a \in [m]$. This is the least index i such that $\gamma_c(i,j) = -1$ for all j; thus, ϕ_a is determined once we condition on Γ_c. Then, **PI2** selects a variable $z_a = |\mathbf{\Phi}_{\phi_a j_a}|$ with $j_a \le k_1$. Now, γ_a is obtained from γ_c by setting to 0 the entries for some (i,j) such that $\gamma_c(i,j) \in \{-1,1\}$ (cf. **PI4**). More precisely, we have $\gamma_a(\phi_a, j) = 0$ for all $j \le k_1$. Furthermore, for $i \in [m] \setminus \{\phi_a\}$ let $\mathcal{J}_i$ be the set of all $j \in [k]$ such that $\pi_c(i,j) = \gamma_c(i,j) \in \{-1,1\}$, and for $i = \phi_a$ let $\mathcal{J}_i$ be the set of all $k_1 < j \le k$ such that $\pi_c(i,j) = \gamma_c(i,j) \in \{-1,1\}$. Then for any $i \in [m]$ and any $j \in \mathcal{J}_i$ the event $\gamma_a(i,j) = 0$ depends only on the events $|\mathbf{\Phi}_{ij'}| = z_a$ for $j' \in \mathcal{J}_i$. By Fact 4.3 the variables $(|\mathbf{\Phi}_{ij'}|)_{i \in [m], j \in \mathcal{J}_i}$ are independently uniformly distributed over $V \setminus Z_c$. Therefore, the events $|\mathbf{\Phi}_{ij'}| = z_a$ for $j' \in \mathcal{J}_i$ are independent of the choice of z_a and of the event τ_b. This shows (4.27) and thus (4.24). ◻

Proof of Corollary 4.7. Let $\mu \le (1+\varepsilon/3)\omega n$ be a positive integer, and let $\hat{\mathcal{U}}_t \subset [m]$ be a set of size μ. Suppose that $t \le \theta$. Let $\nu = nk^{-\varepsilon/2}$, and let B be the set of all maps $\psi : \hat{\mathcal{U}}_t \to [n]$ such that there are less than $\nu + t$ numbers $x \in [n]$ such that $\psi^{-1}(x) = \emptyset$. Furthermore, let $\mathcal{B}_t$ be the event that there are less than ν variables $x \in V \setminus Z_t$ such that $\mathcal{U}_t(x) = 0$. Since $|Z_t| = t$, we have

$$\begin{aligned} \mathrm{P}\left[\mathcal{B}_t | \mathcal{U}_t = \hat{\mathcal{U}}_t\right] &\le \sum_{\psi \in B} \mathrm{P}\left[\psi_t = \psi | \mathcal{U}_t = \hat{\mathcal{U}}_t\right] \le |B|(n-t)^{-\mu} \qquad \text{(by Lemma 4.12)} \\ &= \frac{|B|}{n^\mu} \cdot \left(1 + \frac{t}{n-t}\right)^\mu \le \frac{|B|}{n^\mu} \cdot \exp\left(2\theta \frac{\mu}{n}\right) \\ &\le \frac{|B|}{n^\mu} \cdot \exp(9nk^{-1} \ln^2 k). \end{aligned} \tag{4.28}$$

Furthermore, $|B|/n^\mu$ is just the probability that there are less than ν empty bins if μ balls are thrown uniformly and independently into n bins. Hence, we can use Lemma 2.1 to bound $|B|n^{-\mu}$. To this end, observe that because we are assuming $\varepsilon < 0.1$ the bound

$$\exp\left(-\frac{\mu}{n}\right) \ge \exp\left(-\left(1 + \frac{\varepsilon}{3}\right)\omega\right) = k^{\alpha - 1} \quad \text{holds, where } \alpha = \frac{2\varepsilon}{3} - \frac{\varepsilon^2}{3} \ge 0.6\varepsilon.$$

Therefore, Lemma 2.1 entails that

$$\begin{aligned} |B|n^{-\mu} &\le \mathrm{P}\left[\mathcal{Z}(\mu, n) \le \exp(-\mu/n)n/2\right] \\ &\le O(\sqrt{n}) \exp\left[-\exp(-\mu/n)n/8\right] \le \exp\left[-k^{\alpha-1}n/9\right]. \end{aligned} \tag{4.29}$$

Combining (4.28) and (4.29), we see that for $k \ge k_0(\varepsilon)$ large enough

$$P_t = \mathrm{P}\left[\mathcal{B}_t | \mathcal{U}_t = \hat{\mathcal{U}}_t : \hat{\mathcal{U}}_t \subset [m], |\hat{\mathcal{U}}_t| = \mu\right] \le \exp\left[nk^{-1}\left(9\ln^2 k - k^\alpha/9\right)\right] = o(1/n).$$

Thus, Corollary 4.5 and Lemma 4.6 imply that

$$\begin{aligned} &\mathrm{P}\left[\exists t \le T : |\{x \in V \setminus Z_t : \mathcal{U}_t(x) = 0\} < \nu|\right] \\ &\qquad \le \mathrm{P}\left[T > \theta\right] + \mathrm{P}\left[\max_{0 \le t \le T} |\mathcal{U}_t| > (1+\varepsilon/3)\omega n\right] + \sum_{0 \le t \le \theta} P_t = o(1), \end{aligned}$$

as desired. ◻

5. Proof of Proposition 3.3. Let $0 < \varepsilon < 0.1$. Throughout this section we assume that $k \ge k_0$ for a large enough $k_0 = k_0(\varepsilon) \ge 10$, and that $n > n_0$ for some large enough $n_0 = n_0(\varepsilon, k)$. Let $m = \lfloor n \cdot (1-\varepsilon)2^k k^{-1} \ln k \rfloor$, $\omega = (1-\varepsilon)\ln k$, and $k_1 = \lceil k/2 \rceil$. In addition, we keep the notation introduced in section 4.1.

5.1. Outline. Similarly as in section 4, we will describe the execution of Phase 2 of $\texttt{Fix}(\mathbf{\Phi})$ via a stochastic process. Roughly speaking, the new process starts where the process **PI1**–**PI4** from section 4 (i.e., Phase 1 of Fix) stopped. More precisely, recall that T denotes the stopping time of **PI1**–**PI4**. Let $Z'_0 = \emptyset$ and $\pi'_0 = \pi_T$. Let $U'_0 = U_T$, and let $U'_0(x)$ be the number of indices $i \in U'_0$ such that x occurs positively in $\mathbf{\Phi}_i$ for any variable x. Moreover, let Q'_0 be the set of indices $i \in [m]$ such that $\mathbf{\Phi}_i$ is unsatisfied under the assignment σ_{Z_T} that sets the variables in Z_T to false and all others to true. For $t \geq 1$ we proceed as follows.

PI1′ If $Q'_{t-1} = \emptyset$, the process stops. Otherwise let $\psi_t = \min Q'_{t-1}$.

PI2′ If there are three indices $k_1 < j \leq k-5$ such that $\pi'_{t-1}(\psi_t, j) \in \{1, -1\}$ and $U'_{t-1}(|\mathbf{\Phi}_{\psi_t j}|) = 0$, then let $k_1 < j_1 < j_2 < j_3 \leq k-5$ be the lexicographically first sequence of such indices. Otherwise let $k-5 < j_1 < j_2 < j_3 \leq k$ be the lexicographically first sequence of indices $k-5 < j \leq k$ such that $|\mathbf{\Phi}_{\psi_t j}| \notin Z'_{t-1}$. Let $Z'_t = Z'_{t-1} \cup \{|\mathbf{\Phi}_{\psi_t j_l}| : l = 1, 2, 3\}$.

PI3′ Let U'_t be the set of all $i \in [m]$ that satisfy the following condition. There is exactly one $l \in [k]$ such that $\mathbf{\Phi}_{il} \in V \setminus (Z'_t \cup Z_T)$ and for all $j \neq l$ we have $\mathbf{\Phi}_{ij} \in Z_T \cup Z'_t \cup \overline{V \setminus Z_T}$. Let $U'_t(x)$ be the number of indices $i \in U'_t$ such that x occurs positively in $\mathbf{\Phi}_i$ $(x \in V)$.

PI4′ Let

$$\pi'_t(i,j) = \begin{cases} \mathbf{\Phi}_{ij} & \text{if } (i = \psi_t \wedge j > k_1) \vee \\ & \qquad |\mathbf{\Phi}_{ij}| \in Z'_t \cup Z_T \vee (i \in U'_t \wedge \pi_0(i,j) = 1), \\ \pi'_{t-1}(i,j) & \text{otherwise.} \end{cases}$$

Let Q'_t be the set of all (Z_T, Z'_t)-endangered clauses that contain less than three variables from Z'_t.

Let T' be the stopping time of this process. For $t > T'$ and $x \in V$ let $\pi'_t = \pi'_{T'}$, $U'_t = U'_{T'}$, $Z'_t = Z'_{T'}$, and $U'_t(x) = U'_{T'}(x)$.

The process **PI1′**–**PI4′** models the execution of Phase 2 of $\texttt{Fix}(\mathbf{\Phi})$ since in the terminology of section 3, a variable x is (Z_T, Z'_t)-safe iff $U'_t(x) = 0$. Hence, the set Z' computed in Phase 2 of Fix coincides with $Z'_{T'}$. Thus, our task is to prove that $|Z'_{T'}| \leq nk^{-12}$ w.h.p.

The process **PI1′**–**PI4′** can be applied to any concrete k-SAT formula Φ (rather than the random $\mathbf{\Phi}$). It then yields a sequence $\pi'_t[\Phi]$ of maps, variables $z'_t[\Phi]$, etc. In analogy to the equivalence relation $\equiv_t$ from section 4, we define an equivalence relation $\equiv'_t$ by letting $\Phi \equiv'_t \Psi$ iff $\Phi \equiv_s \Psi$ for all $s \geq 0$, and $\pi'_s[\Phi] = \pi'_s[\Psi]$ for all $0 \leq s \leq t$. Thus, intuitively $\Phi \equiv'_t \Psi$ means that the process **PI1**–**PI4** behaves the same on both Φ, Ψ, and the process **PI1′**–**PI4′** behaves the same on Φ, Ψ up to time t. Let $\mathcal{F}'_t$ be the σ-algebra generated by the equivalence classes of $\equiv'_t$. Then $(\mathcal{F}'_t)_{t \geq 0}$ is a filtration.

FACT 5.1. *For any $t \geq 0$ the map π'_t, the random variable ψ'_{t+1}, the random sets U'_t and Z'_t, and the random variables $U'_t(x)$ for $x \in V$ are $\mathcal{F}'_t$-measurable.*

In analogy to Fact 4.3 we have the following (by "deferred decisions").

FACT 5.2. *Let $\mathcal{E}'_t$ be the set of all pairs (i,j) such that $\pi'_t(i,j) \in \{\pm 1\}$. The conditional joint distribution of the variables $(|\mathbf{\Phi}_{ij}|)_{(i,j) \in \mathcal{E}'_t}$ given $\mathcal{F}'_t$ is uniform over $(V \setminus Z'_t)^{\mathcal{E}'_t}$.*

Let

$$\theta' = \lfloor \exp(-k^{\varepsilon/16}) n \rfloor,$$

and recall that $\theta = \lfloor 4nk^{-1} \ln \omega \rfloor$, where $\omega = (1-\varepsilon) \ln k$. To prove Proposition 3.3 it is sufficient to show that $T' \leq \theta'$ w.h.p., because $|Z'_t| = 3t$ for all $t \leq T'$. To this end, we follow a similar program as in section 4.1: we will show that $|U'_t|$ is "small" w.h.p. for all $t \le \theta'$, and therefore that for $t \le \theta'$ there are plenty of variables x such that $U'_t(x) = 0$. This implies that for $t \le \theta'$ the process will "generate" only a very few (Z_T, Z'_t)-endangered clauses. This then entails a bound on T', because each step of the process removes (at least) one (Z_T, Z'_t)-endangered clause from the set Q'_t. In section 5.2 we will infer the following bound on $|U'_t|$.

LEMMA 5.3. *W.h.p. for all $t \le \theta'$ we have $|U'_t \setminus U_T| \le n/k$.*

COROLLARY 5.4. *W.h.p. the following is true for all $t \le \theta'$: there are at least $nk^{\varepsilon/3-1}$ variables $x \in V \setminus (Z'_t \cup Z_T)$ such that $U'_t(x) = 0$.*

Proof. By Corollary 4.7 there are at least $nk^{\varepsilon/2-1}$ variables $x \in V \setminus Z_T$ such that $U_T(x) = 0$ w.h.p. Hence,

$$u_1 = |\{x \in V \setminus Z_T : U_T(x) = 0\}| \ge nk^{\varepsilon/2-1}.$$

If $x \in V \setminus (Z'_t \cup Z_T)$ has the property $U'_t(x) > 0$ but $U_T(x) = 0$, then there is an index $i \in U'_t \setminus U_T$ such that x is the unique positive literal of $\mathbf{\Phi}_i$ in $V \setminus (Z'_t \cup Z_T)$. Therefore, by Lemma 5.3 w.h.p.

$$u_2 = |\{x \in V \setminus (Z'_t \cup Z_T) : U_T(x) = 0 < U'_t(x)\}| \le |U'_t \setminus U_T| \le n/k.$$

Finally, by **PI2′** we have $|Z'_t| \le 3t$ for all t. Hence,

$$|\{x \in V \setminus (Z'_t \cup Z_T) : U'_t(x) = 0\}| \ge u_1 - u_2 - |Z'_t| \ge nk^{\varepsilon/2-1} - n/k - 3\theta' \ge nk^{\varepsilon/3-1},$$

provided that $k \ge k_0(\varepsilon)$ is sufficiently large. □

COROLLARY 5.5. *Let $\mathcal{Y}$ be the set of all $t \le \theta'$ such that there are less than 3 indices $k_1 < j \le k - 5$ such that $\pi'_{t-1}(\psi_t, j) \in \{-1, 1\}$ and $U'_{t-1}(|\mathbf{\Phi}_{\psi_t j}|) = 0$. Then $|\mathcal{Y}| \le 3\theta' \exp(-k^{0.3\varepsilon})$ w.h.p.*

We defer the proof of Corollary 5.5 to section 5.3, where we also prove the following.

COROLLARY 5.6. *Let $\kappa = \lfloor k^{\varepsilon/4} \rfloor$. There are at most $2k \exp(-\kappa)n$ indices $i \in [m]$ such that $\mathbf{\Phi}_i$ contains more than κ positive literals, all of which lie in $Z_{\theta'} \cup Z_T$.*

COROLLARY 5.7. *W.h.p. the total number of $(Z_T, Z'_{\theta'})$-endangered clauses is at most θ'.*

Proof. Recall that a clause $\mathbf{\Phi}_i$ is $(Z_T, Z'_{\theta'})$-endangered if for any j such that the literal $\mathbf{\Phi}_{ij}$ is true under σ_{Z_T} the underlying variable $|\mathbf{\Phi}_{ij}|$ lies in $Z'_{\theta'}$. Let $\mathcal{Y}$ be the set from Corollary 5.5, and let $\mathcal{Z} = \bigcup_{s \in \mathcal{Y}} Z'_s \setminus Z'_{s-1}$. We claim that if $\mathbf{\Phi}_i$ is $(Z_T, Z'_{\theta'})$-endangered, then one of the following statements is true:

(a) There are two indices $1 \le j_1 < j_2 \le k$ such that $|\mathbf{\Phi}_{ij_1}| = |\mathbf{\Phi}_{ij_2}|$.

(b) There are indices $i' \ne i$, $j_1 \ne j_2$, $j'_1 \ne j'_2$ such that $|\mathbf{\Phi}_{ij_1}| = |\mathbf{\Phi}_{i'j'_1}|$ and $|\mathbf{\Phi}_{ij_2}| = |\mathbf{\Phi}_{i'j'_2}|$.

(c) $\mathbf{\Phi}_i$ is unsatisfied under σ_{Z_T}.

(d) $\mathbf{\Phi}_i$ contains more than $\kappa = \lfloor k^{\varepsilon/4} \rfloor$ positive literals, all of which lie in $Z'_{\theta'} \cup Z_T$.

(e) $\mathbf{\Phi}_i$ has at most κ positive literals, is satisfied under σ_{Z_T}, and contains a variable from $\mathcal{Z}$.

To see this, assume that $\mathbf{\Phi}_i$ is $(Z_T, Z'_{\theta'})$-endangered and (a)–(d) do not hold. Observe that $\mathcal{Z} \supset Z_T \cap Z'_{\theta'}$ by construction (cf. **PI2′**). Hence, if there is j such that $\mathbf{\Phi}_{ij} = \bar{x}$ for some $x \in Z_T$, then $x \in \mathcal{Z}$ and thus (e) holds. Thus, assume that no variable from Z_T occurs negatively in $\mathbf{\Phi}_i$. Then $\mathbf{\Phi}_i$ contains $l \ge 1$ positive literals from $V \setminus Z_T$, and we may assume without loss of generality that these are just the first l literals $\mathbf{\Phi}_{i1}, \ldots, \mathbf{\Phi}_{il}$. Furthermore, $\mathbf{\Phi}_{i1}, \ldots, \mathbf{\Phi}_{il} \in Z'_{\theta'}$. Hence, for each $1 \le j \le l$

there is $1 \le t_j \le \theta'$ such that $\mathbf{\Phi}_{ij} \in Z'_{t_j} \setminus Z'_{t_j-1}$. Since $\mathbf{\Phi}_i$ satisfies neither (a) nor (b), the numbers $t_1, \ldots, t_l$ are mutually distinct. (Indeed, if, say, $t_1 = t_2$, then either $\mathbf{\Phi}_{i1} = \mathbf{\Phi}_{i2}$, or $\mathbf{\Phi}_i$ and $\mathbf{\Phi}_{\psi_{t_1}}$ have at least two variables in common.) Thus, we may assume without loss of generality that $t_1 < \cdots < t_l$. Then $i \in U'_{t_l-1}$ by the construction in step **PI3'**, and thus $\mathbf{\Phi}_{il} \in \mathcal{Z}$. Hence, (e) holds.

Let $X_a, \ldots, X_e$ be the numbers of indices $i \in [m]$ for which (a)–(e) above hold. W.h.p. $X_a + X_b = O(\ln n)$ by Lemma 2.3. Furthermore, $X_c \le \exp(-k^{\varepsilon/8})n$ w.h.p. by Proposition 3.2. Moreover, Corollary 5.6 yields $X_d \le 2k\exp(-\kappa/2)n$ w.h.p. Finally, since $\mathcal{Y} \le 3\theta' \exp(-k^{0.3\varepsilon})$ w.h.p. by Corollary 5.5 and as $|\mathcal{Z}| = 3|\mathcal{Y}|$, Lemma 2.6 shows that w.h.p. for $k \ge k_0(\varepsilon)$ large enough

$$X_e \le \kappa \cdot \sqrt{|\mathcal{Z}|/n} \cdot n \le \kappa \cdot \sqrt{9\exp(-k^{\varepsilon/4})\theta'/n} < \theta'/2 \qquad (\text{as } \theta' = \lfloor \exp(-k^{\varepsilon/16})n \rfloor).$$

Combining these estimates, we obtain $X_a + \cdots + X_e \le \theta'$ w.h.p., provided that $k \ge k_0(\varepsilon)$ is large. ◻

Proof of Proposition 3.3. We claim that $T' \le \theta'$ w.h.p. This implies the proposition because $|Z_{T'}| = 3T'$ and $3\theta' = 3\lfloor \exp(-k^{\varepsilon/16})n \rfloor \le nk^{-12}$ if $k \ge k_0(\varepsilon)$ is sufficiently large. To see that $T' \le \theta'$ w.h.p., let X_0 be the total number of $(Z_T, Z'_{\theta'})$-endangered clauses, and let X_t be the number of $(Z_T, Z'_{\theta'})$-endangered clauses that contain less than 3 variables from Z'_t. Since **PI2′** adds 3 variables from a $(Z_T, Z'_{\theta'})$-endangered clause to Z'_t at each time step, we have $0 \le X_t \le X_0 - t$ for all $t \le T'$. Hence, $T' \le X_0$, and thus the assertion follows from Corollary 5.7. ◻

5.2. Proof of Lemma 5.3. As in (4.13) we let

$$\mathcal{H}_{tij} = \begin{cases} 1 & \text{if } \pi_{t-1}(i,j) = 1 \text{ and } \pi_t(i,j) = z_t, \\ 0 & \text{otherwise,} \end{cases}$$
$$\mathcal{S}_{tij} = \begin{cases} 1 & \text{if } T \ge t \text{ and } \pi_t(i,j) \in \{1, -1\}, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\mathcal{H}_{tij}, \mathcal{S}_{tij}$ refer to the process **PI1**–**PI4** from section 4. With respect to **PI1′**–**PI4′**, we let

$$\mathcal{H}'_{tij} = \begin{cases} 1 & \text{if } \pi'_{t-1}(i,j) = 1,\ \pi'_t(i,j) \in Z'_t, \text{ and } T \le \theta, \\ 0 & \text{otherwise.} \end{cases}$$

In analogy to Lemma 4.9 we have the following.

LEMMA 5.8. *For any $\mathcal{I}' \subset [\theta'] \times [m] \times [k]$ we have*

$$\mathrm{E}\left[\prod_{(t,i,j)\in\mathcal{I}'} \mathcal{H}'_{tij} | \mathcal{F}'_0\right] \le (3/(n - \theta - 3\theta'))^{|\mathcal{I}'|}.$$

Proof. Let $\mathcal{I}'_t = \{(i,j) : (t,i,j) \in \mathcal{I}'\}$ and $X_t = \prod_{(i,j)\in\mathcal{I}'_t} \mathcal{H}'_{tij}$. Due to Lemma 2.4 it suffices to show that

$$\mathrm{E}\left[X_t | \mathcal{F}'_{t-1}\right] \le (3/(n - \theta - 3\theta'))^{|\mathcal{I}'_t|} \quad \forall\, t \le \theta'. \tag{5.1}$$

To see this, let $1 \le t \le \theta'$ and consider a formula Φ such that $T[\Phi] \le \theta$, $t \le T'[\Phi]$, and $\pi'_{t-1}(i,j)[\Phi] = 1$ for all $(i,j) \in \mathcal{I}'_t$. We condition on the event $\mathbf{\Phi} \equiv'_{t-1} \Phi$. Then at time t steps **PI1′**–**PI2′** obtain Z'_t by adding three variables that occur in clause $\mathbf{\Phi}_{\psi_t}$, which is (Z_T, Z'_{t-1})-endangered. Let $(i,j) \in \mathcal{I}'_t$. Since $\mathbf{\Phi} \equiv_{t-1} \Phi$ and $\pi'_{t-1}(i,j)[\Phi] = 1$, we have $\pi'_{t-1}(i,j)\,[\mathbf{\Phi}] = 1$. By **PI4′** this means that $\mathbf{\Phi}_{ij} \notin Z_T \cup Z'_{t-1}$ is a positive literal. Thus, $\mathbf{\Phi}_i$ is not (Z_T, Z'_{t-1})-endangered. Hence, $\psi_t \neq i$. Furthermore, by Fact 5.2 in the conditional distribution $\mathrm{P}\left[\cdot|\mathcal{F}'_{t-1}\right](\Phi)$, the variables $(\mathbf{\Phi}_{ij})_{(i,j)\in\mathcal{I}'_t}$ are independently uniformly distributed over the set $V \setminus (Z_T \cup Z'_{t-1})$. Hence,

$$\mathrm{P}\left[\mathbf{\Phi}_{ij} \in Z'_t|\mathcal{F}'_{t-1}\right][\Phi] \leq 3/|V \setminus (Z_T \cup Z'_{t-1})| \qquad \text{for any } (i,j) \in \mathcal{I}'_t, \tag{5.2}$$

and these events are mutually independent for all $(i,j) \in \mathcal{I}'_t$. Since $|Z_T| = n - T$ and $T = T\,[\Phi] \leq \theta$, and because $|Z'_{t-1}| = 3(t-1)$, (5.2) implies (5.1) and hence the assertion. □

LEMMA 5.9. *Let $2 \leq l \leq \sqrt{k}$, $1 \leq l' \leq l-1$, $1 \leq t \leq \theta$, and $1 \leq t' \leq \theta'$. For each $i \in [m]$ let $X_i = X_i(l,l',t,t') = 1$ if $\theta \geq T \geq t$, $T' \geq t'$, and the following four events occur:*

(a) *$\mathbf{\Phi}_i$ has exactly l positive literals.*
(b) *l' of the positive literals of $\mathbf{\Phi}_i$ lie in $Z'_{t'} \setminus Z_T$.*
(c) *$l - l' - 1$ of the positive literals of $\mathbf{\Phi}_i$ lie in Z_t.*
(d) *No variable from Z_t occurs in $\mathbf{\Phi}_i$ negatively.*

Let

$$B(l,l',t) = 4\omega n \cdot \left(\frac{6\theta' k}{n}\right)^{l'} \cdot \binom{k-l'-1}{l-l'-1}\left(\frac{t}{n}\right)^{l-l'-1}(1-t/n)^{k-l}. \tag{5.3}$$

Then $\mathrm{P}\left[\sum_{i=1}^m X_i > B(l,l',t)\right] = o(n^{-3})$.

Proof. We are going to apply Lemma 2.2. Set $\mu = \lceil \ln^2 n \rceil$, and let $\mathcal{M} \subset [m]$ be a set of size μ. Let $\mathcal{E}_{\mathcal{M}}$ be the event that $X_i = 1$ for all $i \in \mathcal{M}$. Let $P_i \subset [k]$ be a set of size l, and let $H_i, H'_i \subset P_i$ be disjoint sets such that $|H_i \cup H'_i| = l-1$ and $|H'_i| = l'$ for each $i \in \mathcal{M}$. Let $\mathcal{P} = (P_i, H_i, H'_i)_{i\in\mathcal{M}}$. Furthermore, let $t_i : H_i \to [t]$ and $t'_i : H'_i \to [t']$ for all $i \in \mathcal{M}$, and set $\mathcal{T} = (t_i, t'_i)_{i\in\mathcal{M}}$. Let $\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})$ be the event that $\theta \geq T \geq t$, $T' \geq t'$, and the following four statements are true for all $i \in \mathcal{M}$:

(a′) The literal $\mathbf{\Phi}_{ij}$ is positive for all $j \in P_i$ and negative for all $j \in [k] \setminus P_i$.
(b′) $\mathbf{\Phi}_{ij} \in Z'_{t'_i(j)}$ and $\pi'_{t'_i(j)-1}(i,j) = 1$ for all $i \in \mathcal{M}$ and $j \in H'_i$.
(c′) $\mathbf{\Phi}_{ij} = z_{t_i(j)}$ for all $i \in \mathcal{M}$ and $j \in H_i$.
(d′) No variable from Z_t occurs negatively in $\mathbf{\Phi}_i$.

If $\mathcal{E}_{\mathcal{M}}$ occurs, then there exist $(\mathcal{P},\mathcal{T})$ such that $\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})$ occurs. Hence, we are going to use the union bound. For each $i \in \mathcal{M}$ there are

$$\binom{k}{1, l', l-l'-1} \text{ ways to choose the sets } P_i,\, H_i,\, H'_i.$$

Once these are chosen, there are

$$t'^{l'} \text{ ways to choose the map } t'_i, \text{ and } t^{l-l'-1} \text{ ways to choose the map } t_i.$$

Thus,

$$\begin{aligned}\mathrm{P}\left[\mathcal{E}_{\mathcal{M}}\right] &\leq \sum_{\mathcal{P},\mathcal{T}} \mathrm{P}\left[\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})\right]\\ &\leq \left[\binom{k}{1, l', l-l'-1} t'^{l'} t^{l-l'-1}\right]^{\mu} \max_{\mathcal{P},\mathcal{T}} \mathrm{P}\left[\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})\right].\end{aligned} \tag{5.4}$$

Hence, we need to bound $\mathrm{P}[\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})]$ for any given $\mathcal{P},\mathcal{T}$. To this end, let

$$\begin{aligned}\mathcal{I}=\mathcal{I}(\mathcal{M},\mathcal{P},\mathcal{T})&=\{(s,i,j):i\in\mathcal{M},j\in H_i,s=t_i(j)\},\\ \mathcal{I}'=\mathcal{I}'(\mathcal{M},\mathcal{P},\mathcal{T})&=\{(s,i,j):i\in\mathcal{M},j\in H_i',s=t_i'(j)\},\\ \mathcal{J}=\mathcal{J}(\mathcal{M},\mathcal{P},\mathcal{T})&=\{(s,i,j):i\in\mathcal{M},j\in[k]\setminus P_i,s\le t\}.\end{aligned}$$

If $\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})$ occurs, then the positive literals of each clause $\mathbf{\Phi}_i$, $i\in\mathcal{M}$, are precisely $\mathbf{\Phi}_{ij}$ with $j\in P_i$, which occurs with probability 2^{-k} independently. In addition, we have $\mathcal{H}_{sij}=1$ for all $(s,i,j)\in\mathcal{I}$, $\mathcal{H}'_{sij}=1$ for all $(s,i,j)\in\mathcal{I}'$, and $\mathcal{S}_{sij}=1$ for all $(s,i,j)\in\mathcal{J}$. Hence,

$$\mathrm{P}[\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})]\le 2^{-k\mu}\cdot\left\|\mathrm{E}\left[\prod_{(t,i,j)\in\mathcal{I}'}\mathcal{H}'_{tij}\cdot\prod_{(t,i,j)\in\mathcal{I}}\mathcal{H}_{tij}\cdot\prod_{(t,i,j)\in\mathcal{J}}\mathcal{S}_{tij}|\mathcal{F}_0\right]\right\|_\infty.$$

Since the variables $\mathcal{H}_{tij}$ and $\mathcal{S}_{tij}$ are $\mathcal{F}_0'$-measurable, Lemmas 4.9 and 5.8 yield

$$\begin{aligned}\mathrm{P}[\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})]&\le 2^{-k\mu}\cdot\left\|\mathrm{E}\left[\mathrm{E}\left[\prod_{(t,i,j)\in\mathcal{I}'}\mathcal{H}'_{tij}|\mathcal{F}_0'\right]\cdot\prod_{(t,i,j)\in\mathcal{I}}\mathcal{H}_{tij}\cdot\prod_{(t,i,j)\in\mathcal{J}}\mathcal{S}_{tij}|\mathcal{F}_0\right]\right\|_\infty\\ &\le 2^{-k\mu}\cdot\left(\frac{3}{n-\theta-3\theta'}\right)^{l'\mu}\cdot\left\|\mathrm{E}\left[\prod_{(t,i,j)\in\mathcal{I}}\mathcal{H}_{tij}\cdot\prod_{(t,i,j)\in\mathcal{J}}\mathcal{S}_{tij}|\mathcal{F}_0\right]\right\|_\infty\\ &\le 2^{-k\mu}\cdot\left(\frac{3}{n-\theta-3\theta'}\right)^{l'\mu}\cdot(n-\theta)^{-(l-l'-1)\mu}(1-1/n)^{(k-l)t\mu}.\end{aligned}\tag{5.5}$$

Combining (5.4) and (5.5), we see that $\mathrm{P}[\mathcal{E}_{\mathcal{M}}]\le\lambda^\mu$, where

$$\lambda=2^{-k}\binom{k}{1,l',l-l'-1}\left(\frac{3t'}{n-\theta-3\theta'}\right)^{l'}\left(\frac{t}{n-\theta}\right)^{l-l'-1}(1-1/n)^{(k-l)t},\tag{5.6}$$

whence Lemma 2.2 yields

$$\mathrm{P}\left[\sum_{i=1}^m X_i>2\lambda m\right]=o(n^{-3}).\tag{5.7}$$

Thus, the remaining task is to estimate λm: by (5.6) and since $m\le n\cdot 2^k\omega/k$, we have

$$\begin{aligned}\lambda m&=mk2^{-k}\binom{k-1}{l'}\left(\frac{3t'}{n-\theta-3\theta'}\right)^{l'}\cdot\binom{k-l'-1}{l-l'-1}\left(\frac{t}{n-\theta}\right)^{l-l'-1}(1-1/n)^{(k-l)t}\\ &\le\omega n\cdot\left(\frac{6\theta' k}{n}\right)^{l'}\cdot\binom{k-l'-1}{l-l'-1}\left(\frac{t}{n}\right)^{l-l'-1}(1-t/n)^{k-l}\cdot\eta,\qquad\text{where}\\ \eta&=\left(\frac{n}{n-\theta}\right)^{l-l'-1}\cdot\left(\frac{(1-1/n)^t}{1-t/n}\right)^{k-l}\\ &\le\left(1+\frac{\theta}{n-\theta}\right)^{l-l'-1}\exp(kt^2/n^2)\le\exp(2\theta l/n+k\theta^2/n^2).\end{aligned}\tag{5.8}$$

Since $\theta \le 4k^{-1}n\ln k$ and $l \le \sqrt{k}$, we have $\eta \le 2$ for large enough $k \ge k_0(\varepsilon)$. Thus, $2\lambda m \le B(l,l',t)$, whence the assertion follows from (5.7) and (5.8). ∎

LEMMA 5.10. *Let $\ln k \le l \le k$, $1 \le l' \le l$, $1 \le t \le \theta$, and $1 \le t' \le \theta'$. For each $i \in [m]$ let $Y_i = 1$ if $\theta \ge T \ge t$, $T' \ge t'$, and the following three events occur:*

(a) *$\mathbf{\Phi}_i$ has exactly l positive literals.*

(b) *l' of the positive literals of $\mathbf{\Phi}_i$ lie in $Z'_{t'} \setminus Z_T$.*

(c) *$l-l'-1$ of the positive literals of $\mathbf{\Phi}_i$ lie in Z_t.*

Then $\mathrm{P}\left[\sum_{i=1}^m Y_i > n\exp(-l)\right] = o(n^{-3})$.

Proof. The proof is similar to (and less involved than) the proof of Lemma 5.9. We are going to apply Lemma 2.2 once more. Set $\mu = \lceil \ln^2 n \rceil$, and let $\mathcal{M} \subset [m]$ be a set of size μ. Let $\mathcal{E}_{\mathcal{M}}$ be the event that $Y_i = 1$ for all $i \in [M]$. Let $P_i \subset [k]$ be a set of size l, and let $H_i, H'_i \subset P_i$ be disjoint sets such that $|H_i \cup H'_i| = l-1$ and $|H'_i| = l'$ for each $i \in \mathcal{M}$. Let $\mathcal{P} = (P_i, H_i, H'_i)_{i\in\mathcal{M}}$. Furthermore, let $t_i : H_i \to [t]$ and $t'_i : H'_i \to [t']$ for all $i \in \mathcal{M}$, and set $\mathcal{T} = (t_i, t'_i)_{i\in\mathcal{M}}$. Let $\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})$ be the event that $\theta \ge T \ge t$, $T' \ge t'$, and that the following statements are true for all $i \in \mathcal{M}$:

(a$'$) $\mathbf{\Phi}_{ij}$ is positive for all $j \in P_i$ and negative for all $j \notin P_i$.

(b$'$) $\mathbf{\Phi}_{ij} \in Z'_{t'_i(j)}$ and $\pi'_{t'_i(j)-1}(i,j) = 1$ for all $i \in \mathcal{M}$ and $j \in H'_i$.

(c$'$) $\mathbf{\Phi}_{ij} = z_{t_i(j)}$ for all $i \in \mathcal{M}$ and $j \in H_i$.

If $\mathcal{E}_{\mathcal{M}}$ occurs, then there are $(\mathcal{P},\mathcal{T})$ such that $\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})$ occurs. Using the union bound as in (5.4), we get

$$\begin{aligned}\mathrm{P}\left[\mathcal{E}_{\mathcal{M}}\right] &\le \sum_{\mathcal{P},\mathcal{T}} \mathrm{P}\left[\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})\right]\\ &\le \left[\binom{k}{1,l',l-l'-1} t'^{l'} t^{l-l'-1}\right]^{\mu} \max_{\mathcal{P},\mathcal{T}} \mathrm{P}\left[\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})\right]. \qquad (5.9)\end{aligned}$$

Hence, we need to bound $\mathrm{P}\left[\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})\right]$ for any given $\mathcal{P},\mathcal{T}$. To this end, let

$$\begin{aligned}\mathcal{I} = \mathcal{I}(\mathcal{M},\mathcal{P},\mathcal{T}) &= \left\{(s,i,j) : i \in \mathcal{M}, j \in H_i, s = t_i(j)\right\},\\ \mathcal{I}' = \mathcal{I}'(\mathcal{M},\mathcal{P},\mathcal{T}) &= \left\{(s,i,j) : i \in \mathcal{M}, j \in H'_i, s = t'_i(j)\right\}.\end{aligned}$$

If $\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})$ occurs, then the positive literals of each clause $\mathbf{\Phi}_i$ are precisely $\mathbf{\Phi}_{ij}$ with $j \in P_i$ ($i \in \mathcal{M}$). In addition, $\mathcal{H}_{sij} = 1$ for all $(s,i,j) \in \mathcal{I}$ and $\mathcal{H}'_{sij} = 1$ for all $(s,i,j) \in \mathcal{I}'$. Hence, by Lemmas 4.9 and 5.8

$$\begin{aligned}\mathrm{P}\left[\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})\right] &\le 2^{-k\mu}\left\|\mathrm{E}\left[\prod_{(t,i,j)\in\mathcal{I}'} \mathcal{H}'_{tij} \prod_{(t,i,j)\in\mathcal{I}} \mathcal{H}_{tij} | \mathcal{F}_0\right]\right\|_{\infty}\\ &\le \left[2^{-k}\left(\frac{3}{n-\theta-3\theta'}\right)^{l'}\left(\frac{1}{n-\theta}\right)^{l-l'-1}\right]^{\mu}. \qquad (5.10)\end{aligned}$$

Combining (5.9) and (5.10), we see that $\mathrm{P}\left[\mathcal{E}_{\mathcal{M}}\right] \le \lambda^{\mu}$, where

$$\begin{aligned}\lambda &= 2^{-k}\binom{k}{1,l',l-l'-1}\left(\frac{3t'}{n-\theta-3\theta'}\right)^{l'}\left(\frac{t}{n-\theta}\right)^{l-l'-1}\\ &\le k2^{-k}\binom{k-1}{l'}\left(\frac{3t'}{n-\theta-3\theta'}\right)^{l'}\cdot\binom{k-l'-1}{l-l'-1}\left(\frac{t}{n-\theta}\right)^{l-l'-1}\\ &\le k2^{-k}\cdot\left(\frac{6k\theta'}{n}\right)^{l'}\left(\frac{\mathrm{e}(k-l'-1)\theta}{(l-l'-1)n}\right)^{l-l'-1}. \qquad (5.11)\end{aligned}$$

Invoking Lemma 2.2, we get $\mathrm{P}\left[\sum_{i=1}^{m} Y_i > 2\lambda m\right] = o(n^{-3})$. Thus, we need to show that $2\lambda m < \exp(-l)n$.

Case 1: $l' \geq l/2$**.** Since $\theta/n \leq 4k^{-1}\ln\omega$ and $\theta'/n < k^{-2}$, (5.11) yields

$$\lambda m \leq \omega n \left(4\mathrm{e}\ln\omega \cdot \theta'/n\right)^{l'/2} \leq \exp(-l)n/2.$$

Case 2: $l' < l/2$**.** Then (5.11) entails

$$\lambda m \leq \omega n \exp(-2l') \left(10\mathrm{e}\ln\omega/l\right)^{l-l'-1} \leq \exp(-l)n/2.$$

Hence, in either case we obtain the desired bound. ∎

Proof of Lemma 5.3. For $1 \leq t' \leq \theta'$ and $1 < l \leq k$, let $I_l(t')$ be the set of indices $i \in U'_{t'} \setminus U_T$ such that $\mathbf{\Phi}_i$ has precisely l positive literals. Then

$$U'_{t'} \setminus U_T = \bigcup_{l=2}^{k} I_l(t'). \tag{5.12}$$

To bound the size of the set on the right-hand side, we define (random) sets $X(l, l', t, t')$ for $1 \leq l' \leq l-1$, and $t \geq 1$ as follows. If $t > T$ or $t' > T'$, we let $X(l, l', t, t') = \emptyset$. Otherwise, $X(l, l', t, t')$ is the set of all $i \in [m]$ such that $\mathbf{\Phi}_i$ satisfies the following conditions (cf. Lemma 5.9):

(a) $\mathbf{\Phi}_i$ has exactly l positive literals.
(b) l' of the positive literals of $\mathbf{\Phi}_i$ lie in $Z'_{t'} \setminus Z_T$.
(c) $l - l' - 1$ of the positive literals of $\mathbf{\Phi}_i$ lie in Z_t.
(d) No variable from Z_t occurs in $\mathbf{\Phi}_i$ negatively.

We claim that

$$I_l(t') \subset \bigcup_{l'=1}^{l-1} X(l, l', T, \min\{T', t'\}). \tag{5.13}$$

To see this, recall that U_T contains all $i \in [m]$ such that $\mathbf{\Phi}_i$ has precisely one positive literal $\mathbf{\Phi}_{ij} \in V \setminus Z_T$ and no negative literal from $\bar{Z}_T$. Moreover, $U'_{t'}$ is the set of all $i \in [m]$ such that $\mathbf{\Phi}_i$ features precisely one positive literal $\mathbf{\Phi}_{ij} \notin Z'_{t'} \cup Z_T$ and no negative literal from $\bar{Z}_T$. Now, let $i \in I_l$. Then (a) follows directly from the definition of I_l. Moreover, as $i \in I_l \subset U'_{t'}$ clause $\mathbf{\Phi}_i$ has no literal from $\bar{Z}_T$; this shows (d). Further, if $i \in I_l(t')$, then at least one positive literal of $\mathbf{\Phi}_i$ lies in $Z'_{t'} \setminus Z_T$, as otherwise $i \in U_T$. Let $l' \geq 1$ be the number of these positive literals. Then $l' < l$, because there is exactly one j such that $\mathbf{\Phi}_{ij} \notin Z_T \cup Z'_{t'}$ is positive (by the definition of $U'_{t'}$). Furthermore, as there is *exactly* one such j, the remaining $l - l' - 1$ positive literals of $\mathbf{\Phi}_i$ are in Z_T. Hence, (b) and (c) hold as well.

With $B(l, l', t)$ as in Lemma 5.9, let $\mathcal{E}_1$ be the event that

$$\forall 2 \leq l \leq \sqrt{k}, 1 \leq l' \leq l-1, 1 \leq t \leq \theta, 1 \leq t' \leq \theta' : |X(l, l', t, t')| \leq B(l, l', t).$$

Further, let $\mathcal{E}_2$ be the event that

$$\forall \sqrt{k} < l \leq k, 1 \leq l' \leq l-1, 1 \leq t \leq \theta, 1 \leq t' \leq \theta' : |X(l, l', t, t')| \leq n\exp(-l).$$

Let $\mathcal{E}$ be the event that $T \leq \theta$ and that both $\mathcal{E}_1$, $\mathcal{E}_2$ occur. Then by Corollary 4.5 and Lemmas 5.9 and 5.10,

$$\mathrm{P}\left[\neg\mathcal{E}\right] \leq \mathrm{P}\left[T > \theta\right] + \mathrm{P}\left[\neg\mathcal{E}_1\right] + \mathrm{P}\left[\neg\mathcal{E}_2\right] \leq o(1) + 2k^2\theta\theta' \cdot o(n^{-3}) = o(1). \tag{5.14}$$

Furthermore, if $\mathcal{E}$ occurs, then (5.13) entails that for all $t' \le \theta'$

$$\sum_{2\le l\le\sqrt{k}} |I_l(t')| \le \sum_{2\le l\le\sqrt{k}}\sum_{l'=1}^{l-1} |X(l,l',T,\min\{T',t'\})| \le \sum_{l=1}^{k}\sum_{l'=1}^{l-1} B(l,l',T)$$

$$\le 4\omega n \sum_{l'=1}^{k}\left(\frac{6\theta' k}{n}\right)^{l'} \sum_{j=0}^{k-l'-1}\binom{k-l'-1}{j}\left(\frac{T}{n}\right)^{j}\left(1-\frac{T}{n}\right)^{k-l'-1-j}$$

$$= 4\omega n \sum_{l'=1}^{k}\left(\frac{6\theta' k}{n}\right)^{l'} \le 5\omega n\cdot\frac{6\theta' k}{n} \le \frac{n^2}{k}, \tag{5.15}$$

because $\theta' < n/k^4$ for $k \ge k_0(\varepsilon)$ large. Moreover, if $\mathcal{E}$ occurs, then (5.13) yields that for all $t' \le \theta'$

$$\sum_{\sqrt{k}<l\le k} |I_l(t')| \le \sum_{\sqrt{k}<l\le k} \exp(-l)n \le n/k^2, \tag{5.16}$$

provided that $k \ge k_0(\varepsilon)$ is large enough. Thus, the assertion follows from (5.12) and (5.14)–(5.16). ∎

5.3. Proof of Corollaries 5.5 and 5.6. As a preparation we need to estimate the number of clauses that contain a huge number of literals from Z_t for some $t \le \theta$. Note that the following lemma refers solely to the process **PI1**–**PI4** from section 4.

LEMMA 5.11. *Let $t \le \theta$. With probability at least $1 - o(1/n)$ there are no more than $n\exp(-k)$ indices $i \in [m]$ such that $|\{j : k_1 < j \le k,\, |\mathbf{\Phi}_{ij}| \in Z_t\}| \ge k/4$.*

Proof. For any $i \in [m]$, $j \in [k]$, and $1 \le s \le \theta$, let

$$\mathcal{Z}_{sij} = \begin{cases} 1 & \text{if } |\mathbf{\Phi}_{ij}| = z_s,\ \pi_{s-1}(i,j) \in \{-1,1\}, \text{ and } s \le T,\\ 0 & \text{otherwise.}\end{cases}$$

We claim that for any set $\mathcal{I} \subset [t] \times [m] \times ([k] \setminus [k_1])$ we have

$$\mathrm{E}\left[\prod_{(s,i,j)\in\mathcal{I}} \mathcal{Z}_{sij}\right] \le (n-\theta)^{-|\mathcal{I}|}. \tag{5.17}$$

To see this, let $\mathcal{I}_s = \{(i,j) : (s,i,j) \in \mathcal{I}\}$, and set $\mathcal{Z}_s = \prod_{(i,j)\in\mathcal{I}_s} \mathcal{Z}_{sij}$. Then for all $s \le \theta$ the random variable $\mathcal{Z}_s$ is $\mathcal{F}_s$-measurable by Fact 4.2. Moreover, we claim that

$$\mathrm{E}\left[\mathcal{Z}_s|\mathcal{F}_{s-1}\right] \le (n-\theta)^{-|\mathcal{I}_s|} \tag{5.18}$$

for any $s \le \theta$. To prove this, consider any formula Φ such that $s \le T[\Phi]$ and $\pi_{s-1}(i,j)[\Phi] \in \{-1,1\}$ for all $(i,j) \in \mathcal{I}_s$. Then by Proposition 4.3 in the probability distribution $\mathrm{P}[\cdot|\mathcal{F}_{s-1}](\Phi)$ the variables $(\mathbf{\Phi}_{ij})_{(i,j)\in\mathcal{I}_s}$ are mutually independent and uniformly distributed over $V \setminus Z_{s-1}$. They are also independent of the choice of the variable z_s, because $j > k_1$ for all $(i,j) \in \mathcal{I}_s$ and the variable z_s is determined by the first k_1 literals of some clause $\mathbf{\Phi}_{\phi_s}$ (cf. **PI2**). Therefore, for all $(i,j) \in \mathcal{I}_s$ the event $\mathbf{\Phi}_{ij} = z_s$ occurs with probability $1/|V \setminus Z_{s-1}|$ independently. As $|Z_{s-1}| = s-1$, this shows (5.18), and (5.17) follows from Lemma 2.4 and (5.18).

For $i \in [m]$ let $X_i = 1$ if $t \le T$ and there are at least $\kappa = \lceil k/4 \rceil$ indices $j \in [k]\setminus[k_1]$ such that $|\mathbf{\Phi}_{ij}| \in Z_t$, and set $X_i = 0$ otherwise. Let $\mathcal{M} \subset [m]$ be a set of size $\mu = \lceil \ln^2 n \rceil$, and let $\mathcal{E}_{\mathcal{M}}$ be the event that $X_i = 1$ for all $i \in \mathcal{M}$. Furthermore, let

$P_i \subset [k] \setminus [k_1]$ be a set of size $\kappa - 1$ for each $i \in \mathcal{M}$, and let $t_i : P_i \to [t]$ be a map. Let $\mathcal{P} = (P_i)_{i\in\mathcal{M}}$ and $\mathcal{T} = (t_i)_{i\in\mathcal{M}}$, and let $\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})$ be the event that $t \le T$ and $\mathcal{Z}_{t_i(j)ij} = 1$ for all $i \in \mathcal{M}$ and all $j \in P_i$. Let

$$\mathcal{I} = \mathcal{I}_{\mathcal{M}}(\mathcal{P},\mathcal{T}) = \{(t_i(j), i, j) : i \in \mathcal{M}, j \in P_i\}.$$

Then (5.17) entails that for any $\mathcal{P},\mathcal{T}$

$$\mathrm{P}\left[\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})\right] \le \mathrm{E}\left[\prod_{(s,i,j)\in\mathcal{I}} \mathcal{Z}_{sij}\right] \le (n-\theta)^{-|\mathcal{I}|} \le (n-\theta)^{-\mu(\kappa-1)}. \tag{5.19}$$

Moreover, if $\mathcal{E}_{\mathcal{M}}$ occurs, then there exist $\mathcal{P},\mathcal{T}$ such that $\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})$ occurs. For any $i \in \mathcal{M}$ there are $\binom{k-k_1}{\kappa-1}$ ways to choose P_i and $t^{\kappa-1}$ ways to choose t_i. Hence, by the union bound

$$\mathrm{P}\left[\mathcal{E}_{\mathcal{M}}\right] \le \sum_{\mathcal{P},\mathcal{T}} \mathrm{P}\left[\mathcal{E}_{\mathcal{M}}(\mathcal{P},\mathcal{T})\right] \le \lambda^{\mu}, \qquad \text{where}$$

$$\lambda = \binom{k-k_1}{\kappa-1} t^{\kappa-1} \cdot (n-\theta)^{1-\kappa} \le \left(\frac{\mathrm{e}kt}{(\kappa-1)(n-\theta)}\right)^{\kappa-1} \le \left(\frac{12\theta}{n}\right)^{\kappa-1}.$$

Finally, Lemma 2.2 implies that for sufficiently large k we have with probability $1 - o(n^{-1})$

$$\sum_{i=1}^{m} X_i \le 2m\lambda \le n \cdot 2^k (12\theta/n)^{\kappa-1} \le n \exp(-k),$$

because $\theta = \lfloor 4nk^{-1} \ln \omega \rfloor \le 4nk^{-1} \ln\ln k$. ∎

Proof of Corollary 5.5. The goal is to bound the number $|\mathcal{Y}|$ of times $t \le \theta'$ such that the clause $\mathbf{\Phi}_{\psi_t}$ chosen by **PI1′** features less than three literals $\mathbf{\Phi}_{\psi_t j}$ such that $\pi'_{t-1}(\psi_t, j) \in \{-1, 1\}$ and $U'_{t-1}(|\mathbf{\Phi}_{\psi_t j}|) = 0$ ($k_1 < j \le k-5$). We use a similar argument as in the proof of Corollary 4.8. Let

$$\mathcal{Q}'_t = \left|\{x \in V \setminus (Z_T \cup Z'_t) : U'_t(x) = 0\}\right|,$$

and define 0/1 random variables $\mathcal{B}'_t$ for $t \ge 1$ by letting $\mathcal{B}'_t = 1$ iff the following four statements hold:

(a) $T' \ge t$.
(b) $\mathcal{Q}'_{t-1} \ge nk^{\varepsilon/3-1}$.
(c) There are less than $k/4$ indices $k_1 < j \le k$ such that $|\mathbf{\Phi}_{\psi_t j}| \in Z_T$.
(d) At most two indices $k_1 < j \le k-5$ satisfy

$$\pi'_{t-1}(\psi_t, j) = -1 \quad \text{and} \quad U'_{t-1}(|\mathbf{\Phi}_{\psi_t j}|) = 0.$$

This random variable is $\mathcal{F}'_t$-measurable by Fact 5.1. Let $\delta = \exp(-k^{\varepsilon/3}/6)$. We claim

$$\mathrm{E}\left[\mathcal{B}'_t | \mathcal{F}_{t-1}\right] \le \delta \qquad \text{for any } t \ge 1. \tag{5.20}$$

To see this, let Φ be a formula for which (a)–(c) hold. We condition on the event $\mathbf{\Phi} \equiv'_{t-1} \Phi$. Then at time t the process **PI1′**–**PI4′** chooses $\psi_t = \psi_t[\Phi]$ such that $\mathbf{\Phi}_{\psi_t}$ is (Z_T, Z'_{t-1})-endangered and contains less than three variables from Z'_{t-1}. If $\pi'_{t-1}(\psi_t, j) \ne -1$, then either $\pi'_{t-1}(\psi_t, j) = 1$ or $\mathbf{\Phi}_{\psi_t j} \in Z_T \cup Z'_{t-1}$. Due to (c) there are less than $k/4$ indices $j > k_1$ such that $|\mathbf{\Phi}_{\psi_t j}| \in Z_T$. Further, since $\mathbf{\Phi}_{\psi_t}$ is (Z_T, Z'_{t-1})-endangered, there is in fact no j such that $\pi'_{t-1}(\psi_t, j) = 1$. Consequently, there are at least $(k-k_1-5)-\frac{1}{4}k-2$ indices $k_1 < j \leq k-5$ such that $\pi'_{t-1}(\psi_t, j) = -1$. Let $\mathcal{J}$ be the set of all these indices. Assuming $k \geq k_0(\varepsilon)$ is sufficiently large, we have

$$|\mathcal{J}| \geq (k - k_1 - 5) - k/4 - 2 \geq k/5. \tag{5.21}$$

By Fact 5.2 the variables $(|\mathbf{\Phi}_{\psi_t j}|)_{j \in \mathcal{J}}$ are independently uniformly distributed over $V \setminus (Z_T \cup Z'_{t-1})$. Therefore, the number of $j \in \mathcal{J}$ such that $U'_{t-1}(|\mathbf{\Phi}_{\psi_t j}|) = 0$ is binomial $\text{Bin}(|\mathcal{J}|, \mathcal{Q}'_{t-1}/|V \setminus (Z_T \cup Z'_{t-1})|)$. Since (b) requires $\mathcal{Q}'_{t-1} \geq nk^{\varepsilon/3-1}$, (5.21) and the Chernoff bound (2.1) yield

$$\begin{aligned}\mathrm{E}\left[\mathcal{B}'_t|\mathcal{F}'_{t-1}\right](\Phi) &\leq \mathrm{P}\left[\text{Bin}\left(|\mathcal{J}|, \frac{\mathcal{Q}'_{t-1}}{|V \setminus (Z_T \cup Z'_{t-1})|}\right) < 3\right] \\ &\leq \mathrm{P}\left[\text{Bin}\left(\lceil k/5 \rceil, k^{\varepsilon/3-1}\right) < 3\right] \leq \delta,\end{aligned}$$

provided that k is sufficiently large. Thus, we have established (5.20).

Let $\mathcal{Y}' = |\{t \in [\theta'] : \mathcal{B}'_t = 1\}|$. We are going to show that

$$\mathcal{Y}' \leq 2\theta'\delta \quad \text{w.h.p.} \tag{5.22}$$

To this end, letting $\mu = \lceil \ln n \rceil$, we will show that

$$\mathrm{E}\left[(\mathcal{Y}')_\mu\right] \leq (\theta'\delta)^\mu, \qquad \text{where } (\mathcal{Y}')_\mu = \prod_{j=0}^{\mu-1} \mathcal{Y}' - j. \tag{5.23}$$

This implies (5.22). For if $\mathcal{Y}' > 2\theta'\delta$, then for large n we have $(\mathcal{Y}')_\mu > (2\theta'\delta - \mu)^\mu \geq (1.9 \cdot \theta'\delta)^\mu$, whence Markov's inequality entails $\mathrm{P}\left[\mathcal{Y}' > 2\theta'\delta\right] \leq \mathrm{P}\left[(\mathcal{Y}')_\mu > (1.9\theta'\delta)^\mu\right] \leq 1.9^{-\mu} = o(1)$.

In order to establish (5.23), we define a random variable $\mathcal{Y}'_\mathcal{T}$ for any tuple $\mathcal{T} = (t_1, \ldots, t_\mu)$ of mutually distinct integers $t_1, \ldots, t_\mu \in [\theta']$ by letting $\mathcal{Y}'_\mathcal{T} = \prod_{i=1}^{\mu} \mathcal{B}'_{t_i}$. Since $(\mathcal{Y}')_\mu$ equals the number of μ-tuples $\mathcal{T}$ such that $\mathcal{Y}'_\mathcal{T} = 1$, we obtain

$$\mathrm{E}\left[(\mathcal{Y}')_\mu\right] \leq \sum_{\mathcal{T}} \mathrm{E}\left[\mathcal{Y}'_\mathcal{T}\right] \leq \theta'^\mu \max_{\mathcal{T}} \mathrm{E}\left[\mathcal{Y}'_\mathcal{T}\right]. \tag{5.24}$$

To bound the last expression, we may assume that $\mathcal{T}$ is such that $t_1 < \cdots < t_\mu$. As $\mathcal{B}'_t$ is $\mathcal{F}'_t$-measurable, we have for all $l \leq \mu$

$$\begin{aligned}\mathrm{E}\left[\prod_{i=1}^{l} \mathcal{B}'_{t_i}\right] &\leq \mathrm{E}\left[\mathrm{E}\left[\prod_{i=1}^{l} \mathcal{B}'_{t_i}|\mathcal{F}'_{t_l-1}\right]\right] \\ &= \mathrm{E}\left[\prod_{i=1}^{l-1} \mathcal{B}'_{t_i} \cdot \mathrm{E}\left[\mathcal{B}'_{t_l}|\mathcal{F}'_{t_l-1}\right]\right] \overset{(5.20)}{\leq} \delta \cdot \mathrm{E}\left[\prod_{i=1}^{l-1} \mathcal{B}'_{t_i}\right].\end{aligned}$$

Proceeding inductively from $l = \mu$ down to $l = 1$, we obtain $\mathrm{E}\left[\mathcal{Y}'_\mathcal{T}\right] \leq \delta^\mu$, and thus (5.23) follows from (5.24).

To complete the proof, let $\mathcal{Y}''$ be the number of indices $i \in [m]$ such that $|\mathbf{\Phi}_{ij}| \in Z_T$ for at least $k/4$ indices $k_1 < j \leq k$. Combining Corollary 4.5 (which shows that $|Z_T| = T \leq \theta$ w.h.p.) with Lemma 5.11, we see that $\mathcal{Y}'' \leq n\exp(-k) \leq \theta\delta$

w.h.p. As $|\mathcal{Y}| \leq \mathcal{Y}' + \mathcal{Y}''$, the assertion thus follows from Corollary 5.4 (showing that $\mathcal{Q}'_{t-1} \geq nk^{\varepsilon/3-1}$ for all t w.h.p.), (5.22), and the fact that $\theta\delta + 2\theta'\delta \leq \exp(-k^{0.3\varepsilon})n$ for $k \geq k_0(\varepsilon)$ large enough. ∎

Proof of Corollary 5.6. Let $\kappa = \lfloor k^{\varepsilon/4} \rfloor$. The goal is to bound the number of $i \in [m]$ such that $\mathbf{\Phi}_i$ contains at least κ positive literals, all of which end up in $Z_T \cup Z'_{\theta'}$. Since $T \leq \theta$ w.h.p. by Corollary 4.5, we just need to bound the number $\mathcal{V}$ of $i \in [m]$ such that $\mathbf{\Phi}_i$ has at least κ positive literals among which at least κ lie in $Z_\theta \cup Z'_{\theta'}$. Let $\mathcal{V}_{ll'}$ be the number of $i \in [m]$ such that $\mathbf{\Phi}_i$ has exactly l positive literals among which exactly l' lie in $Z'_{\theta'} \setminus Z_\theta$, while exactly $l - l'$ of them lie in Z_θ. Then w.h.p.

$$\sum_{l=\kappa}^{k} \sum_{l'=1}^{l} \mathcal{V}_{ll'} \leq nk \exp(-\kappa) \quad \text{by Lemma 5.10, and}$$

$$\sum_{l=\kappa}^{k} \mathcal{V}_{l0} \leq nk \exp(-\kappa) \quad \text{by Lemma 4.11.}$$

Thus, $\mathcal{V} \leq 2nk \exp(-\kappa)$ w.h.p., as desired. ∎

6. Proof of Proposition 3.4. As before, we let $0 < \varepsilon < 0.1$. We assume that $k \geq k_0$ for a large enough $k_0 = k_0(\varepsilon)$, and that $n > n_0$ for some large enough $n_0 = n_0(\varepsilon, k)$. Furthermore, we let $m = \lfloor n \cdot (1-\varepsilon) 2^k k^{-1} \ln k \rfloor$, $\omega = (1-\varepsilon) \ln k$, and $k_1 = \lceil k/2 \rceil$. We keep the notation introduced in section 4.1. In particular, recall that $\theta = \lfloor 4nk^{-1} \ln \omega \rfloor$.

In order to prove that the graph $G(\mathbf{\Phi}, Z, Z')$ has a matching that covers all (Z, Z')-endangered clauses, we are going to apply the marriage theorem. Basically we are going to argue as follows. Let $Y \subset Z'$ be a set of variables. Since Z' is "small" by Proposition 3.3, Y is small, too. Furthermore, Phase 2 ensures that any (Z, Z')-endangered clause contains three variables from Z'. To apply the marriage theorem, we thus need to show that w.h.p. for any $Y \subset Z'$ the number of (Z, Z')-endangered clauses that contain only variables from $Y \cup (V \setminus Z')$ (i.e., the set of all (Z, Z')-endangered clauses whose neighborhood in $G(\mathbf{\Phi}, Z, Z')$ is a subset of Y) is at most $|Y|$.

To establish this, we will use a first moment argument (over sets Y). This argument does not actually take into account that $Y \subset Z'$, but is over all "small" sets $Y \subset V$. Thus, let $Y \subset V$ be a set of size yn. We define a family $(y_{ij})_{i \in [m], j \in [k]}$ of random variables by letting

$$y_{ij} = \begin{cases} 1 & \text{if } |\mathbf{\Phi}_{ij}| \in Y, \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, define for each integer $t \geq 0$ an equivalence relation $\equiv_t^Y$ on $\Omega_k(n, m)$ by letting $\Phi \equiv_t^Y \Phi'$ iff $\pi_s[\Phi] = \pi_s[\Phi']$ for all $0 \leq s \leq t$ *and* $y_{ij}[\Phi] = y_{ij}[\Phi']$ for all $(i, j) \in [m] \times [k]$. In other words, $\Phi \equiv_t^Y \Phi'$ means that the variables from Y occur in the same places, and that the process **PI1**–**PI4** from section 4 behaves the same up to time t. Thus, $\equiv_t^Y$ is a refinement of the equivalence relation $\equiv_t$ from section 4.1. Let $\mathcal{F}_t^Y$ be the σ-algebra generated by the equivalence classes of $\equiv_t^Y$. Then the family $(\mathcal{F}_t^Y)_{t \geq 0}$ is a filtration. Since $\mathcal{F}_t^Y$ contains the σ-algebra $\mathcal{F}_t$ from section 4.1, all random variables that are $\mathcal{F}_t$-measurable are $\mathcal{F}_t^Y$-measurable as well. In analogy to Fact 4.3 we have the following ("deferred decisions").

FACT 6.1. *Let $\mathcal{E}_t^Y$ be the set of all pairs (i, j) such that $\pi_t(i, j) \in \{1, -1\}$ and $y_{ij} = 0$. The conditional joint distribution of the variables $(|\mathbf{\Phi}_{ij}|)_{(i,j) \in \mathcal{E}_t^Y}$ given $\mathcal{F}_t^Y$*

is uniform over $(V \setminus (Z_t \cup Y))^{\mathcal{E}_t^Y}$.

For any $t \geq 1$, $i \in [m]$, and $j \in [k]$, we define a random variable

$$\mathcal{H}_{tij}^Y = \begin{cases} 1 & \text{if } y_{ij} = 0,\ \pi_{t-1}(i,j) = 1, \text{ and } \pi_t(i,j) = z_t, \\ 0 & \text{otherwise.} \end{cases}$$

LEMMA 6.2. *For any set* $\mathcal{I} \subset [\theta] \times [m] \times [k]$ *we have*

$$\mathrm{E}\left[\prod_{(t,i,j)\in\mathcal{I}} \mathcal{H}_{tij}^Y | \mathcal{F}_0^Y\right] \leq (n - \theta - |Y|)^{-|\mathcal{I}|}.$$

Proof. Due to Fact 6.1 the proof of Lemma 4.9 carries over directly. ∎

For a given set Y we would like to bound the number of $i \in [m]$ such that $\mathbf{\Phi}_i$ contains at least three variables from Y and $\mathbf{\Phi}_i$ has no positive literal in $V \setminus (Y \cup Z_T)$. If for any "small" set Y the number of such clauses is less than $|Y|$, then we can apply this result to $Y \subset Z'$ and use the marriage theorem to show that $G(\mathbf{\Phi}, Z, Z')$ has the desired matching. We proceed in several steps.

LEMMA 6.3. *Let* $t \leq \theta$ *and* $y \leq 0.1$. *Let* $\mathcal{M} \subset [m]$, *and set* $\mu = |\mathcal{M}|$. *Furthermore, let* L, Λ *be maps that assign a subset of* $[k]$ *to each* $i \in \mathcal{M}$ *such that*

$$L(i) \cap \Lambda(i) = \emptyset \quad \text{and} \quad |\Lambda(i)| \geq 3 \quad \forall\, i \in \mathcal{M}. \tag{6.1}$$

Let $\mathcal{E}(Y, t, \mathcal{M}, L, \Lambda)$ *be the event that the following statements are true for all* $i \in \mathcal{M}$:

(a) $|\mathbf{\Phi}_{ij}| \in Y$ *for all* $j \in \Lambda(i)$.

(b) $\mathbf{\Phi}_{ij} \in \overline{V \setminus (Y \cup Z_t)}$ *for all* $j \in [k] \setminus (L(i) \cup \Lambda(i))$.

(c) $\mathbf{\Phi}_{ij} \in Z_t \setminus Y$ *for all* $j \in L(i)$.

Let $l = \sum_{i\in\mathcal{M}} |L(i)|$ *and* $\lambda = \sum_{i\in\mathcal{M}} |\Lambda(i)|$. *Then*

$$\mathrm{P}\left[\mathcal{E}(Y, t, \mathcal{M}, L, \Lambda)\right] \leq 2^{-k\mu} n^{\mu} (2t/n)^l (2y)^{\lambda}.$$

Proof. Let $\mathcal{E} = \mathcal{E}(Y, t, \mathcal{M}, L, \Lambda)$. Let t_i be a map $L(i) \to [t]$ for each $i \in \mathcal{M}$, let $\mathcal{T} = (t_i)_{i\in\mathcal{M}}$, and let $\mathcal{E}(\mathcal{T})$ be the event that (a) and (b) hold and $\mathbf{\Phi}_{ij} = z_{t_i(j)} \notin Y$ for all $i \in \mathcal{M}$ and $j \in L(i)$. If $\mathcal{E}$ occurs, then there is $\mathcal{T}$ such that $\mathcal{E}(\mathcal{T})$ occurs. Hence, by the union bound

$$\mathrm{P}\left[\mathcal{E}\right] \leq \sum_{\mathcal{T}} \mathrm{P}\left[\mathcal{E}(\mathcal{T})\right] \leq t^l \max_{\mathcal{T}} \mathrm{P}\left[\mathcal{E}(\mathcal{T})\right]. \tag{6.2}$$

To bound (6.2) fix any $\mathcal{T}$. For $i \in \mathcal{M}$ we let $l_i = \max t_i^{-1}(\max t_i(L(i)))$; intuitively, this is the last index $j \in L(i)$ such that $\mathbf{\Phi}_{ij}$ gets added to Z_t. Let

$$\mathcal{I} = \left\{(s, i, j) : i \in \mathcal{M}, j \in L(i) \setminus \{l_i\}, s = t_i(j)\right\}.$$

We claim that if $\mathcal{E}(\mathcal{T})$ occurs, then $\mathcal{H}_{sij}^Y = 1$ for all $(s, i, j) \in \mathcal{I}$. For if $\mathcal{E}(\mathcal{T})$ occurs and $(s, i, j) \in \mathcal{I}$, then $s = t_i(j)$ and $\pi_s(i, j) = \mathbf{\Phi}_{ij} = z_s \notin Y$. In addition, by the choice of $l_i \neq j$ both $\mathbf{\Phi}_{ij}$ and $\mathbf{\Phi}_{il_i}$ are positive but not in Z_{s-1}, and consequently $\pi_{s-1}(i, j) = \pi_{s-1}(i, l_i) = 1$. Therefore, $\mathcal{H}_{sij}^Y = 1$, and thus Lemma 6.2 shows that

$$\begin{aligned} \mathrm{P}\left[\mathcal{E}(\mathcal{T}) | \mathcal{F}_0^Y\right] &\leq \mathrm{E}\left[\prod_{(s,i,j)\in\mathcal{I}} \mathcal{H}_{sij}^Y | \mathcal{F}_0^Y\right] \\ &\leq ((1 - y)n - \theta)^{-|\mathcal{I}|} \leq ((1 - y)n - \theta)^{\mu - l}. \end{aligned} \tag{6.3}$$

Furthermore, the event for all $i \in \mathcal{M}$ that

(a′) $|\mathbf{\Phi}_{ij}| \in Y$ for all $j \in \Lambda(i)$,
(b′) $\mathbf{\Phi}_{ij}$ is negative for all $j \notin L(i) \cup \Lambda(i)$,
(c′) $\mathbf{\Phi}_{ij}$ is positive for all $j \in L(i)$

is $\mathcal{F}_0^Y$-measurable. Since the literals $\mathbf{\Phi}_{ij}$ are chosen independently, we have

$$\mathrm{P}\left[(\mathrm{a}'), (\mathrm{b}'), \text{ and } (\mathrm{c}') \text{ hold } \forall\, i \in \mathcal{M}\right] \leq y^{\lambda} 2^{\lambda - k\mu} = (2y)^{\lambda}\, 2^{-k\mu}. \tag{6.4}$$

Combining (6.3) and (6.4), we obtain $\mathrm{P}\left[\mathcal{E}(\mathcal{T})\right] \leq 2^{-k\mu}((1-y)n-\theta)^{\mu-l}\,(2y)^{\lambda}$. Finally, plugging this bound into (6.2), we get for $k \geq k_0(\varepsilon)$ sufficiently large

$$\mathrm{P}\left[\mathcal{E}\right] \leq 2^{-k\mu} t^l \left((1-y)n-\theta\right)^{\mu-l} (2y)^{\lambda} \leq 2^{-k\mu} n^{\mu} \left(\frac{2t}{n}\right)^l (2y)^{\lambda},$$

because $y \leq 0.1$ and $\theta = \lfloor 4nk^{-1}\ln\omega \rfloor < n/3$. ∎

Corollary 6.4. *Let $t \leq \theta$. Let $\mathcal{M} \subset [m]$, and set $\mu = |\mathcal{M}|$. Let l, λ be integers such that $\lambda \geq 3\mu$. Let $\mathcal{E}(Y,t,\mathcal{M},l,\lambda)$ be the event that there exist maps L, Λ that satisfy (6.1) such that $l = \sum_{i\in\mathcal{M}} |L(i)|$, $\lambda = \sum_{i\in\mathcal{M}} |\Lambda(i)|$, and the event $\mathcal{E}(Y,t,\mathcal{M},L,\Lambda)$ occurs. Then*

$$\mathrm{P}\left[\mathcal{E}(Y,t,\mathcal{M},l,\lambda)\right] \leq 2^{-l-k\mu} n^{\mu} (2k^2 y)^{\lambda}.$$

Proof. Given l, λ there are at most $\binom{k\mu}{l,\lambda}$ ways to choose the maps L, Λ (because the clauses in $\mathcal{M}$ contain a total of $k\mu$ literals). Therefore, by Lemma 6.3 and the union bound

$$\begin{aligned} 2^{k\mu} n^{-\mu} \mathrm{P}\left[\mathcal{E}(Y,t,\mathcal{M},l,\lambda)\right] &\leq \binom{k\mu}{l,\lambda} \left(\frac{2t}{n}\right)^l (2y)^{\lambda} \\ &\leq 2^{-l} \left(\frac{4\mathrm{e}\theta k\mu}{ln}\right)^l \left(\frac{2\mathrm{e}k\mu y}{\lambda}\right)^{\lambda} \\ &\leq 2^{-l} \left(\frac{50\mu\ln\omega}{l}\right)^l (2ky)^{\lambda} \\ &= 2^{-l}(2ky)^{\lambda} \cdot \omega^{-50\mu\cdot\alpha\ln\alpha}, \quad \text{where } \alpha = \frac{l}{50\mu\ln\omega}. \end{aligned} \tag{6.5}$$

Since $-\alpha\ln\alpha \leq 1/2$, we obtain $\omega^{-50\mu\cdot\alpha\ln\alpha} \leq \omega^{25\mu} \leq (\ln k)^{25\mu} \leq k^{\lambda}$. Plugging this last estimate into (6.5) yields the desired bound. ∎

Corollary 6.5. *Let $t \leq \theta$, and let $\mathcal{E}(t)$ be the event that there are sets $Y \subset V$, $\mathcal{M} \subset [m]$ of size $\ln n \leq |Y| = |\mathcal{M}| = \mu \leq nk^{-12}$, and integers $l \geq 0$, $\lambda \geq 3\mu$ such that the event $\mathcal{E}(Y,t,\mathcal{M},l,\lambda)$ occurs. Then $\mathrm{P}\left[\mathcal{E}(t)\right] = o(1/n)$.*

Proof. Let us fix an integer $1 \leq \mu \leq nk^{-12}$, and let $\mathcal{E}(t,\mu)$ be the event that there exist sets $Y, \mathcal{M}$ of the given size $\mu = yn$ and numbers l, λ such that $\mathcal{E}(Y,t,\mathcal{M},l,\lambda)$ occurs. Then the union bound and Corollary 6.4 yield

$$\begin{aligned} \mathrm{P}\left[\mathcal{E}(t,\mu)\right] &\leq \sum_{\lambda\geq 3\mu} \sum_{Y,\mathcal{M}:|Y|=|\mathcal{M}|=\mu} \sum_{l\geq 0} \mathrm{P}\left[\mathcal{E}(Y,t,\mathcal{M},l,\lambda)\right] \\ &\leq n^{\mu} \binom{n}{\mu}\binom{m}{\mu} 2^{2-k\mu}(2k^2y)^{3\mu} \\ &\leq \left(\frac{\mathrm{e}^2 n 2^k \ln\omega}{ky^2}\right)^{\mu} \cdot 2^{2-k\mu}(2k^2y)^{3\mu} \\ &\leq 4\left[ynk^6\right]^{\mu}. \end{aligned}$$

Summing over $\ln n \leq \mu \leq nk^{-12}$, we obtain $\mathrm{P}[\mathcal{E}(t)] \leq \sum_{\mu} \mathrm{P}[\mathcal{E}(t,\mu)] = o(1/n)$. ◻

Proof of Proposition 3.4. Assume that the graph $G(\mathbf{\Phi}, Z, Z')$ does not have a matching that covers all (Z, Z')-endangered clauses. Then by the marriage theorem there are a set $Y \subset Z'$ and a set $\mathcal{M}$ of (Z, Z')-endangered clauses such that $|\mathcal{M}| = |Y| > 0$ and all neighbors of indices $i \in \mathcal{M}$ in the graph $G(\mathbf{\Phi}, Z, Z')$ lie in Y. Therefore, for each clause $i \in \mathcal{M}$ the following three statements are true:

(a) There is a set $\Lambda(i) \subset [k]$ of size at least 3 such that $|\mathbf{\Phi}_{ij}| \in Y$ for all $j \in \Lambda(i)$.

(b) There is a (possibly empty) set $L(i) \subset [k] \setminus \Lambda(i)$ such that $\mathbf{\Phi}_{ij} \in Z$ for all $j \in L(i)$.

(c) For all $j \in [k] \setminus (L(i) \cup \Lambda(i))$ we have $\mathbf{\Phi}_{ij} \in \overline{V \setminus (Y \cup Z)}$.

As a consequence, at least one of the following events occurs:

(1) $T > \theta = \lfloor 4k^{-1} \ln \omega \rfloor$.

(2) $|Z'| > nk^{-12}$.

(3) The conclusion of Lemma 2.7 is violated.

(4) There is $t \leq \theta$ such that $\mathcal{E}(t)$ occurs.

The probability of the first event is $o(1)$ by Proposition 3.2, and the second event has probability $o(1)$ by Proposition 3.3, as does the third due to Lemma 2.7. Finally, the probability of the last event is $\theta \cdot o(n^{-1}) = o(1)$ by Corollary 6.5. ◻

REFERENCES

[1] D. Achlioptas, P. Beame, and M. Molloy, *Exponential bounds for DPLL below the satisfiability threshold*, in Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 2004, pp. 132–133.

[2] D. Achlioptas and A. Coja-Oghlan, *Algorithmic barriers from phase transitions*, in Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2008, pp. 793–802.

[3] D. Achlioptas and C. Moore, *Random k-SAT: Two moments suffice to cross a sharp threshold*, SIAM J. Comput., 36 (2006), pp. 740–762.

[4] D. Achlioptas and Y. Peres, *The threshold for random k-SAT is* $2^k \log 2 - O(k)$, J. Amer. Math. Soc., 17 (2004), pp. 947–973.

[5] J. Ardelius and L. Zdeborova, *Exhaustive enumeration unveils clustering and freezing in random* 3*-satisfiability problem*, Phys. Rev. E, 78 (2008), 040101(R).

[6] A. Braunstein, M. Mézard, and R. Zecchina, *Survey propagation: An algorithm for satisfiability*, Random Structures Algorithms, 27 (2005), pp. 201–226.

[7] M.-T. Chao and J. Franco, *Probabilistic analysis of a generalization of the unit-clause literal selection heuristic for the k-satisfiability problem*, Inform. Sci., 51 (1990), pp. 289–314.

[8] V. Chvátal and B. Reed, *Mick gets some (the odds are on his side)*, in Proceedings of the 33rd Annual IEEE Symposium on Foundations of Computer Science (FOCS), 1992, pp. 620–627.

[9] A. Coja-Oghlan, U. Feige, A. Frieze, M. Krivelevich, and D. Vilenchik, *On smoothed k-CNF formulas and the Walksat algorithm*, in Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 2009, pp. 451–460.

[10] M. Davis, G. Longemann, and D. Loveland, *A machine program for theorem proving*, Comm. ACM, 5 (1962), pp. 394–397.

[11] O. Dubois, Y. Boufkhad, and J. Mandler, *Typical random* 3*-SAT formulae and the satisfiability threshold*, in Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 2000, pp. 126–127.

[12] R. Durrett, *Probability and Examples*, 3rd ed., Thomson/Brooks/Cole, Belmont, CA, 2005.

[13] A. Flaxman, *Algorithms for random* 3*-SAT*, in Encyclopedia of Algorithms, M.-Y. Kao, ed., Springer-Verlag, Heidelberg, 2008, pp. 742–744.

[14] E. Friedgut, *Sharp thresholds of graph properties, and the k-SAT problem*, J. Amer. Math. Sci., 12 (1999), pp. 1017–1054.

[15] A. Frieze and S. Suen, *Analysis of two simple heuristics on a random instance of k-SAT*, J. Algorithms, 20 (1996), pp. 312–355.

[16] E. Goldberg and Y. Novikov, *BerkMin: A fast and robust SAT-solver*, Discrete Appl. Math., 155 (2007), pp. 1549–1561.

[17] M. Hajiaghayi and G. Sorkin, *The Satisfiability Threshold of Random* 3-*SAT Is at Least* 3.52, Research report RC22942, IBM, T. J. Watson Research Center, Yorktown Heights, NY, 2003.

[18] S. Janson, T. Łuczak, and A. Ruciński, *Random Graphs*, Wiley-Interscience, New York, 2000.

[19] A. Kaporis, L. Kirousis, and E. Lalas, *The probabilistic analysis of a greedy satisfiability algorithm*, Random Structures Algorithms, 28 (2006), pp. 444–480.

[20] J. H. Kim, *Poisson Cloning Model for Random Graphs*, preprint, 2008; available online from http://www.arxiv.org/abs/0805.4133v1.

[21] E. Maneva, E. Mossel, and M. J. Wainwright, *A new look at survey propagation and its generalizations*, J. ACM, 54 (2007), article 17.

[22] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian, *Solving constraint satisfaction problems through belief propagation-guided decimation*, in Proceedings of the 45th Annual Allerton Conference on Communication, Control, and Computing, University of Illinois at Urbana-Champaign, Urbana, IL, 2007.

[23] G. Semerjian and R. Monasson, *A study of pure random walk on random satisfiability problems with "physical" methods*, in Proceedings of the 6th International Conference on Theory and Applications of Satisfiability Testing (SAT), Springer-Verlag, Berlin, 2003, pp. 120–134.