Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 176 (2013)

A Short Implicant of CNFs with Relatively Many Satisfying Assignments*

Daniel Kane
Department of Mathematics
Stanford University
Stanford CA 94305, USA
dankane@math.stanford.edu

Osamu Watanabe
Dept. of Math. and Comput. Sci.,
Tokyo Institute of Technology
Tokyo 152-8552, Japan
watanabe@is.titech.ac.jp

1 Introduction

We consider the following question:

> Consider any Boolean function $F(X_1, \ldots, X_N)$ that has relatively large number of satisfying assignments and that can be expressed by a CNF formula with clauses polynomial in N. Then how many variables do we need to fix in order to satisfy F? In other words, what is the size of the shortest implicant of F?

To state our results precisely, we introduce notation. Throughout this paper, let F be a given Boolean function over N variables, and we assume that it is given as a CNF formula with M clauses and that it has $P2^N$ satisfying assignments, where P will be referred as the *sat. assignment ratio of* F. Furthermore, we introduce two parameters[1] δ, $0 < \delta < 1$, and $d > 0$, and consider the following situation: (i) $P \geq 2^{-N^\delta}$, and (ii) $M \leq N^d$. For such a CNF formula F, we discuss the size of its implicant in terms of δ and d. As our main result, we show that if $\delta < 1$, then one can always find some "short" and "satisfying" partial assignment, where by "short" we mean that it fixes αN variables for some constant $\alpha < 1$ and by "satisfying (or, sat.) partial assignment" we mean that F is evaluated to 1 (i.e., true) under this partial assignment. In other words, F has an implicant of size $\leq \alpha N$. (In this paper, for any partial assignment, by its "size" we will mean the number of variables fixed by this assignment.)

If a function F has a short sat. partial assignment, then it has many sat. assignments. Our result shows that a certain converse relation holds provided that F is expressed as a CNF formula consisting of some fixed polynomial number of clauses. Of course, there should be some limit on the size of such short sat. partial assignments. Clearly, F needs to have many sat. assignments so that F has a short sat. assignment. Also it seems reasonable that the size of the shortest sat.

*This work was started from the discussion at the workshop on Computational Complexity at the Banff International Research Station for Mathematical Innovation and Discovery (BIRS), 2013. The first author is supported in part by an NSF postdoctoral research fellowship. The second author is supported in part by the ELC project (MEXT KAKENHI Grant No. 24106008).

[1]For simplicity, throughout this paper, we assume that these parameters are constants, and whenever necessary we assume that N is sufficiently large w.r.t. these parameters.

ISSN 1433-8092

partial assignment should get close to N when the number of clauses of F gets larger. We justify this intuition by giving some lower bound on the size of the shortest sat. partial assignment.

Hirsch [6] considered a similar problem for k-CNF formulas for small k, in particular, for constant k. He considered k-CNF formulas F with any sat. assignment ratio $P > 0$, and showed a deterministic algorithm that finds, when such an F is given as an input, one of its satisfying assignments quite efficiently, for example, in linear-time when both k and P can be regarded as constants. As a corollary to this analysis, it is also proved that F always has a partial assignment of size $O(2^k \log(1/P))$. Unfortunately, though, his argument does not seem to work for general CNF formulas (i.e., CNF formulas with no clause size restriction). In fact, Hirsch proved the existence of a general CNF formula that does not have a sat. partial assignment that is "very short", i.e., of size $O(\sqrt{N})$ even though it has a large sat. assignment ratio, say, $P \geq 0.5$. We show here that even in the general case, if F satisfies our conditions (i) and (ii), then it indeed has a short sat. partial assignment. The Switching Lemma of Håstad [5] also can be used to discuss the existence of somewhat short satisfying partial assignments. For example, it is not so hard to show that any CNF formulas with some fixed polynomial number of clauses and constant sat. assignment ratio has a satisfying partial assignment of size $\leq (1-1/\log N)N$. But it seems that there is no trivial way to improve this bound. The contribution of this paper is to improve the upper bound to $(1-\Omega(1))N$ (even for much smaller sat. assignment ratio). We believe that this structural property would be of some help for designing algorithms for the general CNF-SAT problem.

We also consider an algorithmic way to get such a short sat. partial assignment, and obtain a deterministic subexponential-time algorithm that finds one of short sat. partial assignments for CNF formulas with subquadratic number of clauses. More precisely, for any δ and ε such that $\delta + \varepsilon < 1$, we can define a deterministic algorithm that takes any F satisfying (i) and (ii) with δ and $d = 1 + \varepsilon$ and computes its sat. partial assignment of size $\leq \alpha N$ in $\widetilde{O}(2^{N^\beta})$-time[2] for some constants $\alpha < 1$ and $\beta < 1$.

Clearly, our deterministic algorithm for computing a short sat. partial assignment can be used for solving the CNF-SAT problem, and it has some advantage over previously known algorithms. An obvious randomized algorithm for the SAT problem for instances with many sat. assignments is to search for a sat. assignment by generating assignments uniformly at random. Such an algorithm finds a sat. assignment with probability $\geq 2^{-N^\delta}$ for any function with sat. assignment ratio $\geq 2^{-N^\delta}$. Then for the CNF-SAT problem, we may design a deterministic algorithm by applying some good pseudo random sequence generator (*prg* in short) against CNF formulas to this randomized algorithm. That is, an algorithm that tries to find a sat. assignment among assignments generated by such a prg from all possible seeds. In order to ensure that this algorithm obtains some sat. assignment for any CNF formula with sat. assignment ratio $\geq 2^{-N^\delta}$, we need to choose the seed length of the prg so that a generated pseudo random sequence (of length N) is $\gamma := O(2^{-N^\delta})$ close to the uniform distribution for any CNF formula (with, say, $N^{O(1)}$ clauses). For this application, the current best upper bound for the seed length is $\tilde{O}(\log(1/\gamma)^2)$ (ignoring minor factors for our discussion) due to the prg proposed by De et al. [3]. For this seed length, the running time of the simple deterministic algorithm becomes $\tilde{O}(2^{N^{2\delta}})$, which is subexponential if $\delta < 1/2$. This is incomparable with our algorithm's time bound $\widetilde{O}(2^{N^\beta})$ with $\beta = 1 - \frac{(1-(\delta+\varepsilon))^2}{3}$. We should also mention that while the algorithm using prg is oblivious to the input, ours does not have this property.

[2]By $\widetilde{O}(t(N))$ we mean $O\left(t(N)(\log t(N))^{O(1)}\right)$.

2 Notation and Results

Throughout this paper, we will fix the usage of the following symbols: Let F be any Boolean function over N Boolean variables $X_1, \ldots, X_N$, where N is our size parameter. We assume that F has $P2^N$ sat. assignments where $P \geq 2^{-N^\delta}$, and that F is given as a CNF formula with $M \leq N^d$ clauses for some[3] $d > 0$. In order to simplify our discussion, we regard parameters δ and d as constants; whenever necessary, we may assume that N is large enough for each choice of δ and d. We use $|F|$ to denote the number of clauses in F, and for any clause C, we use $|C|$ to denote the number of literals in C. The number of elements in a set W is denoted as $\|W\|$. Symbols ρ and σ are used to denote partial assignments over $X_1, \ldots, X_N$. Any partial assignment ρ takes value 0, 1, or X_i on each variable X_i. We say that ρ *fixes* (the value of) X_i if $\rho(X_i) = 0$ or 1, and that ρ *leaves* X_i *unassigned* if $\rho(X_i) = X_i$. By $F|\rho$, we mean a function evaluated by replacing each occurrence of X_i with $\rho(X_i)$. We say that ρ is a *sat. partial assignment* if $F|\rho = 1$; this is a natural generalization of the standard satisfying assignment notion. We use $\mathrm{Fix}(\rho)$ to denote the set of variables that are fixed by ρ.

In this paper, we use symbols α and β for some constants w.r.t. N, which are defined in terms of δ and d (and some other technical parameters). On the other hand, symbol c is used to denote some constants independent of N, δ, and d. For simplifying our notation during the analysis, we will use some concrete constants such as 0.1, 0.5, etc. whenever we can choose them appropriately. On the other hand, we will sometimes use three digit numbers, e.g., 0.99 to denote $1 - o(1)$. We simply write log for $\log_2$ and ln for $\log_\mathrm{e}$. Let $c_\mathrm{e} = \log_2 \mathrm{e}$. When necessary, we write e^x and 2^x as $\exp(x)$ and $\exp_2(x)$ respectively for showing the exponent clearly.

We introduce notation to state our results formally. Let $\mathrm{sat}(F)$ denote the set of sat. assignments of F. Then the *sat. assignment ratio of* F (denoetd as $\mathrm{sat.ratio}(F)$) is defined by $\mathrm{sat.ratio}(F) = \|\mathrm{sat}(F)\|/2^N$. This quantity is naturally generalized to $F|\rho$ for any partial assignment ρ, which is denoted as $\mathrm{sat.ratio}(F|\rho)$.

Our main result is the following upper bound for the size of sat. partial assignments.

Theorem 1. *For any δ, $0 < \delta < 1$, and for any $d > 0$, let F be any CNF formula such that (i) it has sat. assignment ratio $P \geq \exp_2(-N^\delta)$, and (ii) it consists of $M \leq N^d$ clauses. Then it has some sat. partial assignment $\widehat{\rho}$ of size $\leq \alpha N$, where α is defined by*

$$\alpha \;=\; 1 - \frac{1-\delta}{cd} \exp_2\left(- \frac{(1+o(1))d}{1-\delta} \right), \tag{1}$$

with some constant $c > 0$.

On the other hand, we have the following lower bound.

Theorem 2. *For any δ, $0 < \delta < 1$, and for any $d \geq 1$, consider α defined by*

$$\alpha \;=\; \frac{d - (1+o(1))}{d - \delta} \;>\; 1 - \frac{1 - \delta + o(1)}{d}. \tag{2}$$

Then we have some CNF formula F such that (i) it has sat. assignment ratio $P \geq \exp_2(-N^\delta)$, (ii) it consists of $M \leq N^d$ clauses, and (iii) it has no sat. partial assignment ρ of size $\leq \alpha N$.

---

[3]Our algorithmic version is shown only for CNF formulas with subquadratic number of clauses.

We also have an algorithmic version of Theorem 1 when the number of clauses is bounded by $N^{1+\varepsilon}$ for some $\varepsilon < 1$ such that $\delta + \varepsilon < 1$ holds.

Theorem 3. *For any $\delta > 0$ and $\varepsilon > 0$ such that $\delta + \varepsilon < 1$, there exists a deterministic algorithm such that for any given CNF formula F satisfying (i) and (ii) of Theorem 1 w.r.t. δ and $d = 1 + \varepsilon$, it runs in $\widetilde{O}(2^{N^\beta})$-time for some $\beta < 1$ and yields some sat. partial assignment $\widehat{\rho}$ for F of size $\le \alpha N$, where α is defined by*

$$\alpha \;=\; 1 - \frac{1-(\delta+\varepsilon)}{c_1} \cdot \exp_2\left(- \frac{c_2}{(1-(\delta+\varepsilon))(1-\delta)} \right) \tag{3}$$

with some constants $c_1, c_2 \ge 1$.
Remark. *We can show that the above time bound holds for any β such that $\beta \ge 1 - \frac{(1-(\delta+\varepsilon))^2}{3}$.*

We recall some common bounds that will be used often in this paper. For any integer $n \ge 1$, we have

$$\left(1 - \frac{1}{n}\right)^n \;\le\; \mathrm{e}^{-1} \;\le\; \left(1 - \frac{1}{n+1}\right)^n, \quad \text{and} \quad \left(1 + \frac{1}{n}\right)^n \;\le\; \mathrm{e} \;\le\; \left(1 + \frac{1}{n}\right)^{n+1}.$$

We also use the following more general ones: for any x, $0 < x < 1$, and any $m \ge 1$, we have

$$\begin{aligned} (1-o(1))\exp(-xm) \;&\le\; (1-x)^m \;\le\; \exp(-xm), \quad \text{and} \\ (1-o(1))\exp(xm) \;&\le\; (1+x)^m \;\le\; \exp(xm), \end{aligned}$$

where both lower bounds are correct (and hence used only) when $x^2 m = o(1)$ w.r.t. size parameter(s), which is N in our case.

3 Upper bound proof

In this section we give a proof of Theorem 1, showing an upper bound on the size of short sat. partial assignments. Throughout this section, for any δ, $0 < \delta < 1$ and any $d > 0$, we consider sufficiently large N and fix any F satisfying (i) and (ii) of the theorem w.r.t. δ and d.

The key tool of our proof is the following lemma, which can be shown as a corollary of the analysis given by Hirsch [6]. By the *width* of a clause, we mean the number of literals appearing in the clause.

Lemma 1. *Consider any CNF formula consisting of clauses of width $\le k$ with sat. assignment ratio $Q > 0$. Then it has a partial satisfying assignment of size $\le 4c_{\mathrm{e}}2^k \log Q^{-1}$.*
Remark. *We may consider k as a function in N. For simplicity, we assume that and $Q < 1/4$.*

Proof. We simply use the main theorem of [6], i.e., Theorem 3.5 (c). In our terminology, the theorem states that any CNF formula consisting clauses of width $\le k$ with sat. assignment ratio Q has a sat. partial assignment of size at most

$$\log_{2/\lambda_0}\left(2Q^{-1}\right) + k - 1 \;<\; (\log 2/\lambda_0)^{-1}\left(1 + \log Q^{-1}\right) + k,$$

where λ_0 is the unique root of $h(x) = 0$ for a function

$$h(x) \;=\; 1 - \sum_{1 \le i \le k} x^{-i} \;\left(\;=\; 1 + \frac{1 - x^{-k}}{1-x} \text{ for any } x > 1 \right).$$

We show that $(\log 2/\lambda_0)^{-1} \le 2c_e 2^k$, which is sufficient for the bound of the lemma since $Q < 1/4$.

Let λ be a number satisfying $(\log 2/\lambda)^{-1} = 2c_e 2^k$. We need to show that $\lambda_0 \le \lambda$, from which $(\log 2/\lambda_0)^{-1} \le 2c_e 2^k$ follows. As mentioned in [6], $h(x)$ is monotone, $h(1) < 0$, and $h(2) > 0$. Thus, $\lambda_0 \in (1,2)$, and $\lambda_0 \le \lambda$ is derived by showing $h(\lambda) > 0$. We confirm this below.

First note that

$$\lambda \;=\; 2 \cdot 2^{-(2c_e 2^k)^{-1}} \;=\; 2 \cdot \mathrm{e}^{-(2^{k+1})^{-1}} \;\ge\; 2\left(1 - \frac{1}{2^{k+1}}\right)^{2^{k+1}\cdot(2^{k+1})^{-1}} \;=\; 2 - 2^{-k}.$$

Then we have $\lambda + \lambda^{-k} > 2$ because $\lambda^{-k} = 2^{-k} \cdot \mathrm{e}^{k(2^{k+1})^{-1}} \ge 2^{-k}$. Then by simple calculation (and using the fact that $\lambda > 1$) we can show that

$$1 + \frac{1-\lambda^{-k}}{1-\lambda} \;>\; 0,$$

which means $h(\lambda) > 0$ (again since $\lambda > 1$). □

Let us first see the outline[4] of our proof. For given F satisfying (i) and (ii) of the theorem, we show the existence of some partial assignment ρ_{12} that assigns some $(1-\eta)N$ variables for some $\eta < 1$ and converts F to a formula consisting of *narrow clauses*, clauses of width, say, $\le 0.99(1-\delta)\log N$, while keeping relatively large sat. assignment ratio $Q \ge 2^{-2N^{\delta}}$. Then the theorem follows from the above lemma. To show the existence of ρ_{12}, we use the idea of Ajtai introduced in [1] and define ρ_{12} in two stages. In the first stage, we define a partial assignment ρ_1 to eliminate all *wide clauses*, clauses of width $\ge Ad\ln N$ where $A \ge 1$ is some parameter defined later. We then define ρ_2 in the second stage that converts all clauses to narrow ones. We will show that $\rho_{12} = \rho_2 \circ \rho_1$ has the desired properties.

Now we explain each stage precisely from the first stage for defining ρ_1. We show a procedure for defining a sequence of partial assignments $\sigma_1, \sigma_2, \ldots, \sigma_T$ so that ρ_1 is defined by $\rho_1 = \sigma_T \circ \cdots \circ \sigma_1$. Intuitively, the main objective of the procedure is to eliminate wide clauses. Consider the situation where we have determined $\sigma_1, \ldots, \sigma_{t-1}$, and let F_{t-1} denote $F|\sigma_{t-1} \circ \cdots \circ \sigma_1$. Also let W denote the set of wide clauses in F_{t-1}. Note that there must be some variable X_i that appears in more than $\frac{\|W\| Ad\log N}{N}$ clauses of W; then either X_i or $\overline{X_i}$ is a literal that appears more than $\frac{\|W\| Ad\log N}{2N}$ clauses of W, which we call a *popular literal* among W. We would like to define σ_t to assign positively to one of such popular literals, thereby killing many wide clauses. But we should be careful not to reduce the sat. assignment ratio too much by this assignment. Here we check whether the assignment reduces the sat. assignment ratio too much; specifically, we check whether the sat. assignment ratio becomes smaller than $1 - p_1$ of the current value, and if so, use the opposite assignment to the popular literal. Note that this opposite assignment increases the sat. assignment ratio by $1 + p_1$. From this, we can show that such opposite assignments do not occur so many times (since the sat. assignment ratio cannot go beyond 1). Though very natural, this is somewhat new technical point for implementing the idea of Ajtai for our problem.

[4] In the earlier version of our paper [4] submitted to CCC'14, we used an outline similar to our algorithmic version, for which we needed to assume that $M \le N^{1+\varepsilon}$ for some $\varepsilon < 1$. This new proof outline was advised by one of the referees of CCC'14, who also explained us in detail the ideas of Ajtai in [1], which lead us to this improved version. We thank to the CCC'14 PC for their careful reading and many constructive comments to our submission and, in particular, to the anonymous referee for the instructive advice.

We define the probability parameter p_1 by

$$p_1 \; = \; \frac{c_{\rm e}A}{2} N^{-(1-\delta)},$$

and using this p_1, we formally describe our idea as a procedure in Figure 1. We iterate this procedure until no wide clause exists. We show below that the number of iterations is bounded by $6N/A$.

procedure for σ_t (where $t \geq 1$)
// assume that $\sigma_1, \ldots, \sigma_{t-1}$ have been defined, and let F_{t-1} denote $F|\sigma_{t-1} \circ \cdots \circ \sigma_1$.
  if F_{t-1} has no wide clause
    then stop and output the obtained sequence as $\sigma_1, \ldots, \sigma_T$;
  W = the set of wide clauses in F_{t-1};
  Y_i = (any) one of the popular literal (either X_i or $\overline{X_i}$) among W;
  if sat.ratio($F_{t-1}|(Y_i := 1)$) $\leq (1-p_1) \cdot$ sat.ratio(F_{t-1})
    then $\sigma_t = (Y_i := 0)$; (Case I)
    else $\sigma_t = (Y_i := 1)$; (Case II)
    // σ_t leaves the other variables unassigned.

Figure 1: Procedure for defining σ_t

Lemma 2. *Define T be the number of iterations of the above procedure needed until no wide clause exists in F_T. Then we have $T \leq 6N/A$. Also we have* sat.ratio$(F_T) \geq 0.99 \cdot 2^{-2N^\delta}$.

Proof. Let T_1 and T_2 denote respectively the number of iterations such that Case I and Case II occurs. We show that each of them is bounded by $O(N/A)$.

We first show that $T_2 \leq 2N/A$. For any $t \geq 1$, suppose that Case II occurs at the tth iteration of our procedure. That is, the algorithm finds some literal Y_i (either X_i or $\overline{X_i}$) that is popular among the set W of wide clauses in F_{t-1}, and it indeed assigns true to Y_i. This assignment satisfies (and hence removes) more than $\frac{\|W\|Ad\log N}{2N}$ clauses of W, which reduces the number of wide clauses by $(1-\frac{Ad\log N}{2N})$. Thus, since we have initially at most M ($\leq N^d$) wide clauses, if Case II occurs for T' times, then the remaining number of wide clauses becomes at most

$$M\left(1-\frac{Ad\ln N}{2N}\right)^{T'} \;<\; N^d \mathrm{exp}\left(-\frac{Ad\ln N}{2N}T'\right) \;=\; N^d N^{-d\cdot\frac{A}{2N}\cdot T'}.$$

Thus, the remaining number of wide clauses becomes less than 1 (that is, 0) if $T' \geq 2N/A$. Hence, Case II does not occur more than $2N/A$ times, that is, $T_2 \leq 2N/A$.

We can also show here that the sat. ratio does not decrease so much by an assignment defined by this first stage. Note that the sat. ratio may decrease only by assignments defined at Case II. On the other hand, for any iteration t where Y_i is selected as a popular literal, Case II would not be chosen if the sat. ratio is increased by $1+p_1$ by an assignment $Y_i := 0$. Hence, when Case II is chosen, it is guaranteed that the sat. ratio does not decrease by a factor less than $1-p_1$ by assigning Y_i true. Thus, from our bound for T_2, for any tth iteration of the procedure, we have

$$\begin{aligned}
\text{sat.ratio}(F|\sigma_t \circ \cdots \circ \sigma_1) &\geq \text{sat.ratio}(F)\,(1-p_1)^{T_2} \;\geq\; P\left(1 - \frac{c_{\mathrm{e}}A}{2}N^{-(1-\delta)}\right)^{2N/A} \\
&\geq P \cdot 0.99\text{exp}\left(-(c_{\mathrm{e}}A/2)N^{\delta-1}(2N/A)\right) \;=\; 0.99P\text{exp}\left(-c_{\mathrm{e}}N^{\delta}\right) \\
&\geq 0.99 \cdot 2^{-N^{\delta}} \cdot 2^{-N^{\delta}} \;=\; 0.99 \cdot 2^{-2N^{\delta}}.
\end{aligned}$$

In particular, this bound holds when the iteration stops with no wide clause.

Next we give a bound $T_1 \leq 4N/A$. From the above, we know that the sat. ratio cannot be smaller than $0.99 \cdot 2^{-2N^{\delta}}$ by the assignments of Case II. On the other hand, at each step where Case I is chosen, the sat. ratio gets increased by $1 + p_1$. Hence, if Case I occurs T' times by some tth iteration, then the sat. ratio of F_t becomes at least

$$\begin{aligned}
0.99 \cdot 2^{-2N^{\delta}}(1+p_1)^{T'} &= 0.99 \cdot 2^{-2N^{\delta}}\left(1 + \frac{c_{\mathrm{e}}A}{2}N^{-(1-\delta)}\right)^{T'} \\
&\geq 0.99 \cdot 2^{-2N^{\delta}} \cdot 0.99\text{exp}\left(\frac{c_{\mathrm{e}}A}{2}N^{-(1-\delta)}T'\right) \;=\; 0.99^2\text{exp}_2\left(-2N^{\delta} + 2N^{\delta}\frac{AT'}{4N}\right).
\end{aligned}$$

Thus, if $T' > 4N/A$, then the sat. ratio of F_t becomes larger than 1, a contradiction. Therefore we have $T_2 \leq 4N/A$. From these bounds the lemma follows. □

With $A = 12$ use our procedure to define $\rho_1 = \sigma_T \circ \cdots \circ \sigma_1$. Then Lemma 2 guarantees that $F_T = F|\rho_1$ has no wide clause, it has sat. ratio $\geq 0.99 \cdot 2^{-2N^{\delta}}$, and ρ_1 fixes at most $6N/A = N/2$ variables. Next we consider the second stage to define ρ_2 for converting all clauses of F_T to narrow ones. Without loss of generality (by renaming variable indecies) we may assume that, for some $N' \geq N/2$, $\mathbf{X} = \{X_1, \ldots, X_{N'}\}$ is the set of variables of F_T; that is, $X_1, \ldots, X_{N'}$ are variables unassigned by ρ_1. The idea is to show the existence of some subset $\mathbf{S}$ of $\mathbf{X}$ such that (i) each clause of F_T has at most $k = 0.99(1-\delta)\log N$ variables in $\mathbf{S}$, and (ii) $\|\mathbf{S}\| = \Omega(N)$. Then from (i) it follows that any assignment to $\mathbf{X} \setminus \mathbf{S}$ transforms F_T to a formula consisting of only narrow clauses. From such partial assignments, we choose one with the largest sat. ratio as ρ_2.

Lemma 3. *Using the notation above. There exists a subset* $\mathbf{S}$ *of* $\mathbf{X}$ *(= the set of all variables of* F_T*) such that (i) every clause in* F_T *has at most* $k = 0.99(1-\delta)\log N$ *variables in* $\mathbf{S}$*, and (ii)* $\|\mathbf{S}\| \geq 0.99\eta_d N/2$*, where*

$$\eta_d \;=\; \frac{0.99(1-\delta)}{70d}\text{exp}_2\left(-\frac{d}{0.98(1-\delta)}\right).$$

Hence, $F_T|\rho'$ *has only narrow clauses for any partial assignment* ρ' *that fixes all and only variables in* $\mathbf{X}\backslash\mathbf{S}$*. Furthermore, among such partial assignments, there exists some* ρ_2 *such that* $\text{sat.ratio}(F_T|\rho_2) \geq 0.99 \cdot 2^{-2N^{\delta}}$ *holds.*

Proof. We generate $\mathbf{S}$ randomly by selecting each $X_i \in \mathbf{X}$ with probability η_d independently. Then with high probability, we have $\|\mathbf{S}\| \geq 0.99\eta_d N'$ ($\geq 0.99\eta_d N/2$) by Chernoff bound, we can bound the probability that $\|\mathbf{S}\| < 0.99\eta_d N'$ occurs by, say, 0.1 (for sufficiently large N).

Consider any clause C of F_T, and we estimate the probability that it has at least $k = 0.99(1-\delta)\log N$ literals in[5] $\mathbf{S}$. For any fixed k literals in C, the probability that they (i.e., these variables) all selected in $\mathbf{S}$ is η_d^k. Hence, by using the union bound, the probability that some k literals are all selected in $\mathbf{S}$ is at most

$$\begin{aligned}\binom{Ad\log N}{k}\eta_d^k &\le \left(\frac{c_{\mathrm{e}}Ad\log N}{k}\right)^k \eta_d^k = \left(\eta_d\frac{c_{\mathrm{e}}12d\log N}{0.99(1-\delta)\log N}\right)^k \\ &< \left(\eta_d\frac{70d}{0.99(1-\delta)}\right)^k = \exp_2\left(-\frac{dk}{0.98(1-\delta)}\right) \le \exp_2(-1.01d\log N) = N^{-1.01d}.\end{aligned}$$

Thus, again by the union bound, the probability that F_T has some clause that has more than k literals in $\mathbf{S}$ is less than 0.9. Therefore, with some positive probability some $\mathbf{S}$ (among randomly generated ones) satisfies the theorem.

Note that each assignment to variables in $\mathbf{X}-\mathbf{S}$ yields a disjoint partial assignment ρ' of F_T. Thus, among them there should be some ρ' that has at least the sat. ratio of F_T, which is at least $0.99\cdot 2^{-2N^\delta}$. □

We summarize our analysis and prove the theorem.

Proof of Theorem 1. For a given formula F, we define ρ_1 and ρ_2 as stated in Lemma 2 and Lemma 3 respectively. We use $A = 12$ as mentioned above. Then we can guarantee that the resulting formula $F' = F|\rho_2\circ\rho_1$ has at least $N' = 0.99\eta_d N/2$ variables, which are the variables in the set $\mathbf{S}$ that ρ_2 keeps unassigned among variables in $F_T = F|\rho_1$. Note also that F' consists of clauses of width $\le k = 0.99(1-\delta)\log N$ and has sat. assignment ratio $Q \ge 0.99\cdot 2^{-2N^\delta}$. Hence we apply Lemma 1 to this formula to show the existence of some partial assignment ρ_3 (to F') of size at most

$$4c_{\mathrm{e}}2^k\log Q^{-1} \le 4c_{\mathrm{e}}N^{0.99(1-\delta)}\cdot(2N^\delta - \log 0.99) \le 8.01c_{\mathrm{e}}N^{1-0.01(1-\delta)},$$

which is smaller than $N'/2$ for sufficiently large N. Thus, by defining $\widehat{\rho} = \rho_3\circ\rho_2\circ\rho_1$, we have a satisfying partial assignment that keeps at least

$$\frac{N'}{2} = \frac{0.99\eta_d N}{2} = \frac{0.99\cdot 0.99(1-\delta)}{2\cdot 70d}\exp_2\left(-\frac{d}{0.98(1-\delta)}\right)\cdot N \ge \frac{1-\delta}{cd}\exp_2\left(-\frac{(1+o(1))d}{1-\delta}\right)\cdot N$$

variables unassigned for some constant $c > 0$. This gives the desired upper bound to the size of our defined partial assignment $\widehat{\rho}$. □

4 A lower bound

We move on to the proof of Theorem 2. The idea is relatively easy. For any δ, $0<\delta<1$, and $d\ge 1$, consider α satisfying (2) of Theorem 2. To be concrete, let us assume that $\alpha = (d-1.01)/(d-\delta)$. Let Π be the set of partial assignments fixing αN variables. Our goal is to show F that satisfies the conditions (i) and (ii) of the theorem *and* (iii) that is satisfied by no $\rho\in\Pi$.

We define F randomly as the conjunction of at most N^d random clauses chosen independently. Roughly speaking, each clause is a disjunction of approximately $2s$ randomly chosen literals. The parameter s is chosen large enough to guarantee that each clause is satisfiable with a certain

[5] $\mathbf{S}$ is a set of variables; thus, precisely, by "a literal is in $\mathbf{S}$", we mean that its corresponding variable is in $\mathbf{S}$.

probability so that F's sat. assignment ratio exceeds $\exp_2(-N^\delta)$ with probability larger than some p. On the other hand, we keep s small enough so that each clause is satisfied with relatively small probability by fixing values of at most αN variables, thereby ensuring that $F|\rho = 1$ for some partial assignments $\rho \in \Pi$ with probability $<< p$. Then with the probabilistic argument, we can show the existence of our target Boolean formula F.

We start our detailed explanation with a precise way to generate a random clause. For some parameter s that will be defined later, we consider the following way to generate a random clause: For each i, $1 \le i \le N$, independently, with probability s/N, select X_i, with probaiblity s/N, select $\overline{X_i}$, and other wise, discard the variable. The resulting clause is just the disjunction of the selected literals. In the following claim, we assume that C is a random clause obtained by this random clause generation.

Claim 1. *For any assignment $\mathbf{a} \in \{0,1\}^N$ and any partial assignment $\rho \in \Pi$, we have*

$$\Pr_C[\, C(\mathbf{a}) = 0\,] \;\le\; \mathrm{e}^{-s}, \quad \text{and} \quad \Pr_C[\, C|\rho \ne 1\,] \;\ge\; 0.99\mathrm{e}^{-s\alpha}.$$

Proof. Note that there are N literals that are satisfied by $\mathbf{a}$. Thus, the first bound is shown by estimating the probability that none of them are selected for C.

For the second bound, consider any $\rho \in \Pi$, and let $I_{\rho,+}$ (resp., $I_{\rho,-}$) be the set of indecies i such that $\rho(X_i) = 1$ (resp., $\rho(X_i) = 0$). A random clause C is not satisfied by ρ (i.e., $C|\rho \ne 1$ holds) if and only if X_i is not chosen for all $i \in I_{\rho,+}$ *and* $\overline{X_i}$ is not chosen for all $i \in I_{\rho,-}$. Thus, we have

$$\Pr_C[\, C|\rho \ne 1\,] \;=\; \left(1 - \frac{s}{N}\right)^{\alpha N} \;\ge\; 0.99\exp\left(-\frac{s}{N}\cdot \alpha N\right) \;\ge\; 0.99\mathrm{e}^{-s\alpha}.$$

□

Here we fix s by $s = (d-\delta)\ln N + 1$, from which we have

$$\mathrm{e}^{-s}N^d \;=\; N^\delta/\mathrm{e}. \tag{4}$$

Note also that $0 < s/N < 1$; hence, we can use s/N as a parameter for our random clause generation.

For generating a random formula F we iterate this random clause generation procedure independently for N^d times and define F as the conjunction of obtained clauses. In the following analysis, we use F as a random variable denoting a random formula generated in this way. We define $p = \exp_2(-N^\delta)$, and in the following claims, we show the probability that F satisfies the conditions of the theorem is at least, say, $0.9p > 0$, thereby proving the existence of the desired formula.

Claim 2.

$$\Pr_F[\,\text{sat.ratio}(F) \ge p\,] \;\ge\; \exp_2\left(-N^\delta\right) \quad (=\; p). \tag{5}$$

Proof. Consider any assignment $\mathbf{a} \in \{0,1\}^N$. From Claim 1 and from the definition of random formula F, it follows

$$\Pr_F[\, F(\mathbf{a}) = 1\,] \;\ge\; \left(1 - \mathrm{e}^{-s}\right)^{N^d} \;\ge\; 0.99\exp\left(-\mathrm{e}^{-s}N^d\right) \;=\; 0.99\exp\left(-N^\delta/\mathrm{e}\right),$$

where the last derivation is from (4). Let q denote $0.99\exp(-N^\delta/\mathrm{e})$. Note that $q >> p$; in particular, we have $q > 2p$ if N is large enough.

For our analysis, we use the standard averaging argument. Note first that

$$\mathrm{Exp}_F\big[\,\mathrm{sat.ratio}(F)\,\big] \;=\; 2^{-N}\sum_{\mathbf{a}}\mathrm{Exp}_F\big[\,F(\mathbf{a})\,\big] \;=\; 2^{-N}\sum_{\mathbf{a}}\Pr_F\big[\,F(\mathbf{a})=1\,\big] \;\geq\; q$$

On the other hand, letting $r = \Pr_F[\,\mathrm{sat.ratio}(F) < p]$, we have

$$\mathrm{Exp}_F\big[\,\mathrm{sat.ratio}(F)\,\big] \;<\; r\cdot p + (1-r)\cdot 1 \;=\; 1 - r(1-p).$$

Then from the above two inequalities, we have $q < 1 - r(1-p)$, from which

$$r \;<\; \frac{1-q}{1-p} \;=\; 1 - \frac{q-p}{1-p} \;<\; 1-p$$

follows since $q > 2p$ and $1-p < 1$. This proves the claim. ☐

Claim 3.

$$\Pr_F\big[\,\exists\rho\in\Pi\,\big[\,F|\rho=1\,\big]\,\big] \;\leq\; \left(\frac{3}{\mathrm{e}^2}\right)^N \quad (<\;0.1p \text{ for sufficiently large } N). \tag{6}$$

Proof. Consider any partial assignment $\rho\in\Pi$. Again from Claim 1 and from the definition of random formula F, it follows

$$\begin{aligned}\Pr_F[\,F|\rho=1\,] \;&\leq\; \big(1-0.99\mathrm{e}^{-s\alpha}\big)^{N^d}\\ &\leq\; \exp\Big(-0.99\mathrm{e}^{-s\alpha}N^d\Big) \;=\; \exp\left(\,-0.99\left(\frac{N^{\delta-d}}{\mathrm{e}}\right)^{\alpha}N^d\right).\end{aligned}$$

We analyze below the last expression and show that it is at most $\exp(-2N)$; then the claim follows from the union bound because there are at most 3^N partial assignments. Consider the argument of the above expression $\exp(\cdots)$; our goal is to show that it is $\leq -2N$. To see this, consider

$$\ln\left(\,0.99\left(\frac{N^{\delta-d}}{\mathrm{e}}\right)^{\alpha}N^d\,\right) \;=\; \big(d-\alpha(d-\delta)\big)\ln N + \ln 0.99 - \alpha. \tag{7}$$

From our choice of α, we have $d-\alpha(d-\delta) = 1.01$. Hence, we have (7) $\geq \ln N + \ln 2$ for sufficiently large N. This gives our desired bound. ☐

5 Algorithmic version

In this section we give a proof of Theorem 3, showing an algorithmic way to define a short partial assignment.

The key tool is to use an algorithmic version of the Lovász Local Lemma, which has been improved greatly [7, 8, 2]. Our idea is simple. We show a subexponential-time deterministic algorithm that reduces our task to the CNF-SAT problem and use an algorithmic version of the Lovász Local Lemma. Here we use the version[6] reported in [2].

[6]In their paper, as a typical application of the lemma, an efficient deterministic algorithm is shown for k-CNF formulas with no variable appearing in many clauses. This may be used in our situation; but here we go back to the original lemma to confirm that our parameter choice works.

We specify our target problem and state the lemma in a slightly simpler way. Consider any sufficiently large N', and let $\mathcal{F}_{N'}$ denote the set of CNF formulas over N' Boolean variables with at most $(N')^2$ clauses. The lemma gives an algorithm that finds a sat. assignment for any formula in $\mathcal{F}_{N'}$ satisfying a certain condition. Let F' be any given formula in $\mathcal{F}_{N'}$. Consider a random assignment to its N' variables, and for each clause C of F', let E_C denote an event that C becomes false by the assignment. Our goal (and the task of our algorithm) is to find an assignment avoiding E_C for all clauses C of F', that is, to find a sat. assignment for this $F' \in \mathcal{F}_{N'}$. Let $\Gamma(C)$ be the set of clauses that shares some variable with C; note that E_C is independent from $E_{C'}$ for any $C' \notin \Gamma(C)$. In the lemma we consider some mapping x (for this F'); by using this x, we also define $x'(E_C)$ by

$$x'(E_C) \;=\; x(E_C) \prod_{C' \in \Gamma(C)} (1 - x(E_{C'}))\,.$$

Now we state the following algorithmic version of the Lovász Local Lemma [2].

Lemma 4. *For any $y > 0$, there exists a deterministic algorithm that takes any $F' \in \mathcal{F}_{N'}$ for any $N' \geq 1$ as input, and runs in time $O((N')^{(c_{\rm LLL}/y)})$ yielding some sat. assignment of F' if we can define some mapping x for F' that satisfies*

$$\Pr[\,E_C\,] \;\leq\; x'(E_C)^{1+y}, \quad x(E_C) \;<\; 1/2, \quad \text{and} \quad x'(E_C) \;\geq\; (N')^{-1} \tag{8}$$

for all its clauses C, where $c_{\rm LLL} > 0$ is some constant independent from y and N'.

In the following, we show some algorithmic way to transform F to another formula F'; we then apply this lemma to F' to obtain its sat. assignment, which can be used to define our desired partial assignment. Here, in order to explain requirements for F', we consider some rough strategy for defining x for F' to achieve the conditions of (8). For any clause C of F', we have $\Pr[E_C] = 2^{-|C|}$. Thus, it is natural to define $x(E_C) \approx 2^{-|C|}$. Then we need to require that $|C|$ is not so small, that is, C is not "very narrow" to satisfy the first and the third conditions of (8). We need, for example, $|C| \geq \log N'$. Also in order to avoid the situation where $x'(E_C)$ gets too small compared with $x(E_C)$, we need to require that $\|\Gamma(C)\|$ is not so large, that is, C is not so "popular." When constructing F', these two requirements are important.

We explain our algorithmic way to obtain a sat. partial assignment $\widehat{\rho}$. First let us fix input related parameters. Let F be any given CNF formula satisfying the condition of the theorem with parameters δ and ε, and let α be the constant defined by (3). Let $\gamma = 1 - (\delta + \varepsilon)$, which may be small[7] but a positive constant. Fix F, δ, ε, γ, and α from now on to the end of this section.

We define $\widehat{\rho}$ in three stages. In the first stage, a partial assignment ρ_1 is defined in a way similar to the first stage in the proof of Theorem 1. In the second stage, we convert F to F' by removing some number of variables randomly from $F|\rho_1$ so that it is still satisfiable and we can use the algorithm of the above lemma to find one of its sat. assignments. Then in the third stage, we use the above algorithm to compute a sat. assignment of F'. Note that this *complete* assignment to F' can be regarded as a sat. partial assignment ρ_2 of $F|\rho_1$ that leaves all (and only) removed variables unassigned; we define our final assignment $\widehat{\rho}$ by $\widehat{\rho} = \rho_2 \circ \rho_1$. The partial assignment ρ_1 is defined to satisfy the following two requirements: (a$'$) $F|\rho_1$ has no "narrow" clause, and (b) $F|\rho_1$ has no "popular" literal. Then from the first requirement we can show that (a) F' has no

[7] For simplicity, we assume that $\gamma < 0.5$. The case where $\gamma \geq 0.5$ can be analyzed similarly with different setting for our technical parameters b and B.

"very narrow" clause with high probability after removing some number of variables. By using this property together with (b), we can satisfy the conditions of (8). (Note that F' clearly satisfies (b) if $F|\rho_1$ does.)

We show that the assignment $\widehat{\rho} = \rho_2 \circ \rho_1$ defined above leaves $\Omega(N)$ variables unassigned as desired. First, we show that ρ_1 leaves some $(1 - o(1))N$ variables unassigned. Like the previous ρ_1, our ρ_1 is defined by using a sequence $\sigma_1, \sigma_2, \ldots$ of very short partial assignments defined step by step. Here we need to eliminate narrow clauses and popular literals. To eliminate each narrow clause, we fix the values of all literals in the clause. We show that the sat. ratio increases a good amount by using an appropriate assignment to those literals. Hence, the number of applying very short partial assignments of this type is limited (because otherwise, the sat. ratio exceeds 1). On the other hand, we eliminate popular literas (here w.r.t. all clauses in the current formula) in the same way as before, and by the same reasoning, we can bound the number of applying this step. Altogether we can show that there exists some ρ_1 that satisfies both (a$'$) and (b) by fixing at most $O(N^\beta)$ variables, where $\beta < 1$ is the constant specified in the theorem. Then we show that with high probability one can remove $\Omega(N)$ variables from $F|\rho_1$ while keeping both (a) and (b) so that we can apply the above lemma to find an assignment satisfying F'.

Consider an algorithmic implementation of these three stages. We show that one can in fact find the best ρ_1 by trying all possible candidate partial assignments in $\widetilde{O}(2^{N^\beta})$-time. By the standard method of conditional probabilities, the random removable of variables can be derandomized in polynomial-time. Thus, F' is obtained deterministically in $\widetilde{O}(2^{N^\beta})$-time. Then the above lemma guarantees that one of its sat. assignment is computed in $O(N^{O(1/y)})$-time. Since the last time bound is subsumed by $\widetilde{O}(2^{N^\beta})$, we can conclude that $\widehat{\rho} = \rho_2 \circ \rho_1$ is deterministically computable in $\widetilde{O}(2^{N^\beta})$-time. This is the outline of our algorithmic way to obtain $\widehat{\rho}$.

We start detailed explanation by defining necessary parameters for our analysis. Below for simplicity we use one parameter b in several contexts, as a parameter to denote some number in $(0, 1)$ that is close to 1. We will confirm by the end of our analysis that the argument goes through by defining b by, e.g., $b = 1 - 0.4\gamma < 1$. Define ℓ and L by

$$\ell \;=\; b(1-\delta)\log N, \quad \text{and} \quad L \;=\; N^{(1-b\gamma)(1-\delta)}.$$

The motivation for these choices will be given later. We say that a clause is *narrow* if its width is less than ℓ, and a literal is *popular* (in a currently considered CNF formula) if it appears in more than L clauses.

Now consider the first stage, where a partial assignment ρ_1 is defined so that $F|\rho_1$ has no narrow clause nor popular literal. Similar to the proof of Theorem 1 we define ρ_1 by $\rho_1 = \sigma_T \circ \cdots \circ \sigma_1$, where each σ_t is a partial assignment defined by the procedure stated in Figure 2. We iterate this procedure until $(*)$ holds.

Here we give a rough estimate to explain the motivation for our choices of ℓ and L. Suppose that at some tth iteration of the procedure, F_{t-1} has some narrow clause and C is one of such narrow clauses chosen in the procedure. Let σ_t be the partial assignment defined in the procedure w.r.t. C. Note that out of all possible $2^{|C|}$ assignments of C, $2^{|C|} - 1$ of them satisfy C. Hence (see also the proof of the next lemma), sat. assignment ratio sat.ratio(F_t) (= sat.ratio$(F_{t-1}|\sigma_t)$) gets increased from sat.ratio(F_{t-1}) by a factor of

$$\frac{2^{|C|}}{2^{|C|}-1} \;=\; 1 + \frac{1}{2^{|C|}-1} \;\geq\; 1 + \frac{1}{2^{\ell}-1} \;\geq\; 2^{c_e 2^{-\ell}}.$$

procedure for σ_t (where $t \geq 1$)
// assume that $\sigma_1, \ldots, \sigma_{t-1}$ have been defined, and let F_{t-1} denote $F|\sigma_{t-1} \circ \cdots \circ \sigma_1$.
// let $p_2 = 2^{-\ell-2}$.
  if (F_{t-1} has no narrow clause) and (F_{t-1} has no popular literal) — $(*)$
    then stop and output the obtained sequence as $\sigma_1, \ldots, \sigma_T$;
  Case A: (if F_{t-1} has a narrow clause)
    C = (any) one of the narrow clauses;
    σ_t = a satisfying assignment σ of C maximizing sat.ratio(F_{t-1}, σ);
    // σ_t leaves the other variables unassigned.
  Case B: (if F_{t-1} has no narrow clause and F_{t-1} has a popular literal)
    Y_i = (any) one of the popular literal (either X_i or $\overline{X_i}$) in F_{t-1};
    if sat.ratio($F_{t-1}|(Y_i := 1)) \leq (1 - p_2) \cdot$ sat.ratio(F_{t-1})
      then $\sigma_t = (Y_i := 0)$; (Case B.I)
      else $\sigma_t = (Y_i := 1)$; (Case B.II)
      // σ_t leaves the other variables unassigned.

Figure 2: Procedure for defining σ_t

In other words, we have $\log(\text{sat.ratio}(F_t)) - \log(\text{sat.ratio}(F_{t-1})) \geq 2^{-\ell}$. Note here that $\log(\text{sat.ratio}(F_0))$ $(= \log(\text{sat.ratio}(F))) \geq -N^{\delta}$ and $\log(\text{sat.ratio}(F_s)) \leq 0$ for any s. Hence, the number of times that Case A occurs in the iterations of the procedure is at most $N^{\delta}/2^{-\ell} = N^{\delta}2^{\ell} = \exp_2(\delta \log N + \ell)$. In order to bound the number of occurrences of Case A by $o(N)$, we require that $\ell < (1-\delta)\log N$. Suppose next that some popular literal is assigned true at some iteration of the procedure. This partial assignment satisfies at least L clauses, and the number of clauses gets decreased by L at this iteration. Hence, the number of times that (Case B.II) occurs is bounded by $N^{1+\varepsilon}/L$. Again in order to bound this by $o(N)$ we require that $L > N^{\varepsilon}$, which can be achieved by choosing L bit larger than $N^{(1-\gamma)(1-\delta)}$ (since $\varepsilon = 1 - \delta - \gamma$). Finally, as explained later, we also require that L is small enough compared with 2^{ℓ} in order to satisfy the condition (8) of Lemma 4. These requirements determine ℓ and L roughly as defined above.

We formally prove this idea works.

Lemma 5. *Define T be the number of iterations of the above procedure needed until $(*)$ holds for F_T. Then we have $T \leq 10N^{\beta_b}$, where $\beta_b = 1 - \gamma(1-b)$, which is less than 1 by choosing $b < 1$. Also there is a deterministic algorithm simulating these iterations to produce some $\sigma_T \circ \cdots \circ \sigma_1$ satisfying $(*)$ in $\widetilde{O}(2^{N^{\beta_b + o(1)}})$-time.*

Remark. *From our choice of b (see (11)) we may define β directly as*

$$\beta \;=\; 1 - \frac{(1-(\delta+\varepsilon))^2}{3},$$

with which we can bound the running time of the deterministic algorithm by $\widetilde{O}(2^{N^{\beta}})$.

Proof. In order to measure the progress made by each iteration, we introduce the following potential function for a given partial assignment $\overline{\sigma}_s := \sigma_s \circ \cdots \sigma_1$ of F. Below by $|F|\overline{\sigma}_s|$ we denote the number of clauses in $F|\overline{\sigma}_s$.

$$\Phi(\overline{\sigma}_s) \;=\; 2^{\ell} \log\left(\text{sat.ratio}(F|\overline{\sigma}_s)^{-1}\right) + \left|(F|\overline{\sigma}_s)\right| \cdot L^{-1}.$$

Clearly, Φ must be nonnegative for any partial assignment. We show that each σ_t decreases $\Phi(\sigma_{t-1})$ by constant, say, 0.2, whereas its initial value is $2N^{\beta_b}$, thereby proving our upper bound for T.

First estimate the initial potential, that is, $\Phi(\overline{\sigma}_0)$ for the null partial assignment $\overline{\sigma}_0$; that is, $F|\overline{\sigma}_0$ is F itself. Noting that $\text{sat.ratio}(F) = P \geq 2^{-N^\delta}$ and $|F| = M \leq N^{1+\varepsilon}$, we have

$$\begin{aligned}\Phi(\overline{\sigma}_0) &= 2^\ell(-\log P) + M \cdot L^{-1} \;\leq\; N^{b(1-\delta)}N^\delta + N^{1+\varepsilon}N^{-(1-b\gamma)(1-\delta)}\\ &= N^{b(1-\delta)+\delta} + N^{\varepsilon+\delta-b\gamma(1-\delta)} \;=\; N^{b(\gamma+\varepsilon)+1-\varepsilon-\gamma} + N^{1-\gamma(1-b(1-\delta))}\\ &\leq N^{1-\gamma(1-b)} + N^{1-\gamma(1-b)},\end{aligned}$$

where the last bound is derived by using $b < 1$ and $\delta, \varepsilon \geq 0$. Thus, $\Phi(\overline{\sigma}_0) \leq 2N^{\beta_b}$ as desired.

For analyzing the decrement of Φ, consider first Case A of the above procedure for defining σ_t when $\overline{\sigma}_{t-1} = \sigma_{t-1} \circ \cdots \circ \sigma_1$ have been defined. We estimate the difference between $\Phi(\overline{\sigma}_t)$ and $\Phi(\overline{\sigma}_{t-1})$. Let C be the narrow clause satisfied by σ_t. We first estimate the sum of $\text{sat.ratio}(F|\sigma \circ \overline{\sigma}_{t-1})$ over all satisfying assignments σ of C, which we write by $\Sigma_{\sigma:\text{sat}.C}\ \text{sat.ratio}(F|\sigma \circ \overline{\sigma}_{t-1})$. From the definition of the satisfying assignment ratio, we have

$$\Sigma_{\sigma:\text{sat}.C}\ \text{sat.ratio}(F|\sigma \circ \overline{\sigma}_{t-1}) \;=\; \frac{\Sigma_{\sigma:\text{sat}.C}\ \|\text{sat}(F|\sigma \circ \overline{\sigma}_{t-1})\|}{\|\{0,1\}^N|\sigma \circ \overline{\sigma}_{t-1}\|},$$

where by "$\{0,1\}^N|\rho$" we denote the set of assignments consistent with ρ. Note here that

$$\Sigma_{\sigma:\text{sat}.C}\ \|\text{sat}(F|\sigma \circ \overline{\sigma}_{t-1})\| \;=\; \|\text{sat}(F|\overline{\sigma}_{t-1})\|,$$

and

$$\|\{0,1\}^N|\sigma \circ \overline{\sigma}_{t-1}\| \;=\; \Sigma_{\sigma:\text{sat}.C}\ 2^{-|C|}\|\{0,1\}^N|\overline{\sigma}_{t-1}\|.$$

Hence, we have

$$\begin{aligned}\Sigma_{\sigma:\text{sat}.C}\ \text{sat.ratio}(F|\sigma \circ \overline{\sigma}_{t-1}) &= \frac{\|\text{sat}(F|\overline{\sigma}_{t-1})\|}{\Sigma_{\sigma:\text{sat}.C}\ 2^{-|C|}\|\{0,1\}^N|\overline{\sigma}_{t-1}\|}\\ &= 2^{|C|}\Sigma_{\sigma:\text{sat}.C}\ \frac{\|\text{sat}(F|\overline{\sigma}_{t-1})\|}{\|\{0,1\}^N|\overline{\sigma}_{t-1}\|} \;=\; 2^{|C|}\Sigma_{\sigma:\text{sat}.C}\ \text{sat.ratio}(F|\overline{\sigma}_{t-1}).\end{aligned}$$

Note that there are $2^{|C|} - 1$ satisfying assignments for C. Thus, the above estimation shows that $\text{sat.ratio}(F|\sigma \circ \overline{\sigma}_{t-1})$ on average is $2^{|C|}/(2^{|C|}-1) \times \text{sat.ratio}(F|\overline{\sigma}_{t-1})$. Since we choose for σ_t an assignment maximizing the ratio, we have

$$\begin{aligned}\text{sat.ratio}(F|\sigma_t \circ \overline{\sigma}_{t-1}) &\geq \frac{2^{|C|}}{2^{|C|}-1} \cdot \text{sat.ratio}(F|\overline{\sigma}_{t-1}) \;\geq\; \left(1 + \frac{1}{2^\ell - 1}\right) \cdot \text{sat.ratio}(F|\overline{\sigma}_{t-1})\\ &\geq 2^{c_e 2^{-\ell}} \cdot \text{sat.ratio}(F|\overline{\sigma}_{t-1}),\end{aligned}$$

since $|C| < \ell$ because C is a narrow clause. Then $2^\ell \log(\text{sat.ratio}(F|\overline{\sigma}_{t-1})^{-1})$, the first term of $\Phi(\overline{\sigma}_{t-1})$, gets decreased at least by $c_e = \log \mathrm{e} > 1$ (while the second term of $\Phi(\overline{\sigma}_{t-1})$ does not increase). Thus, we have $\Phi(\sigma_t \circ \overline{\sigma}_{t-1}) \leq \Phi(\overline{\sigma}_{t-1}) - 1$.

Consider next Case B. In this case, we want to assign Y_i positively (i.e., to use the assignment $Y_i := 1$) to satisfy at least L clauses. Here we make use of the previous technique. In order to avoid the situation where too many satisfying assignments are lost by this assignment, we first check whether the satisfying assignment ratio gets increased by assigning Y_i negatively. If the ratio is increased by a factor $\geq 1 + p_2 = 1 + 2^{-\ell-2}$, then we simply use this negative assignment for σ_t, by which the Φ value gets decreased by at least 0.2 (for sufficiently large N), since $2^\ell \log(1 + p_2)^{-1} \approx c_{\mathrm{e}}/4 > 0.2$ (or more precisely, $2^\ell \log(1+p_2)^{-1} \geq c_{\mathrm{e}}/4 - 0.1 > 0.2$ for sufficiently large N). Otherwise, the ratio would not get decreased by a factor $< (1 - 2^{-\ell-2})$ by assigning Y_i positively, which means that the increment of the first term of the Φ value is at most $c_{\mathrm{e}}/4 < 0.4$. On the other hand, by assigning Y_i positively, we can satisfy (and hence eliminate) at least L clauses from $F|\overline{\sigma}_{t-1}$, by which the Φ value is decreased by at least 1. Thus, we have $\Phi(\sigma_t \circ \overline{\sigma}_{t-1}) \leq \Phi(\overline{\sigma}_{t-1}) - 0.6$.

Finally, we show that the whole computation can be executed *deterministically* in subexponetial-time, more precisely, within the time bound stated in the theorem. Here we simply consider all possible sequence of partial assignments $\sigma_t \circ \cdots \circ \sigma_1$ that could be chosen during (at most) $10N^{\beta_b}$ iterations of the procedure to find one yielding a CNF formula satisfying $(*)$, that is, a CNF formula with no narrow clause nor popular literal. Since there are at most

$$\left(2^\ell + 2\right)^{10N^{\beta_b}+1} \;=\; \exp_2\left((\ell+1)(10N^{\beta_b}+1)\right) \;\leq\; \exp_2\left(O\left((\log N)(N^{\beta_b})\right)\right) \;\leq\; \exp_2\left(N^{\beta_b+o(1)}\right)$$

possible choices and one can check whether a given sequence of partial assignments satisfies $(*)$ in polynomial-time in N, the whole computation can be done in $\widetilde{O}(2^{N^{\beta_b+o(1)}})$-time. □

Define $\rho_1 = \sigma_T \circ \cdots \circ \sigma_1$ by using the sequence $\sigma_1, \sigma_2, \ldots, \sigma_T$ of partial assignments that our algorithm produces. This is the partial assignment obtained in the first stage. As guaranteed by the above lemma, $F|\rho_1$ has no narrow clause nor popular literal. Also note that it fixes at most $\ell T \leq (\log N) \cdot 10N^{\beta_b}$ variables for some $\beta_b < 1$; hence, there are at least, say, $0.99N$ unassigned variables in $F|\rho_1$. Let $\mathbf{X}$ denote the set of such remaining variables of $F|\rho_1$.

We consider the second stage. We choose here (at least) $(1-\alpha)N$ variables from $\mathbf{X}$ to keep as free variables of our final partial assignment, and construct a CNF formula F' by removing these variables from $F|\rho_1$. We first consider a randomized way to select removed variables. We use $\mathbf{R}$ to denote the set of selected variables. We introduce two parameters B and p_*, where p_* is defined as a constant (that may depend on δ and ε) so that

$$1 - Bp_* \;=\; b$$

holds with constant B defined below. Our method is simply to select each variable $\mathbf{X}$ with probability p_* to $\mathbf{R}$ independently. We show below that with high probability the random set $\mathbf{R}$ satisfies desired properties for constructing F'.

Lemma 6. *Use notation as above. We define B by*

$$B \;=\; \mathrm{e} \cdot \exp_2\left(\frac{2.1}{(1-b)b(1-\delta)}\right). \tag{9}$$

Then with probability $1 - o(1)$, $\mathbf{R}$ satisfies the following conditions: (i) $(p_*/2)N \leq \|\mathbf{R}\| \leq 0.1N$, and (ii) every clause of $F|\rho_1$ has at least $\ell' := b\ell$ $(= (1 - Bp_*)\ell)$ literals in $\mathbf{X} \setminus \mathbf{R}$.

Proof. Though small, since p_* is constant, we can use the standard Chernoff bound to show that the probability that (i) fails to hold is less than, e.g., N^{-1}.

For bounding the probability that (ii) fails to hold, we consider any clause C of $F|\rho_1$. Let Z denote the number of variables of C that are selected to $\mathbf{R}$. Note that $\mathrm{Exp}[Z] = p_*|C|$. Hence, if

C has less than $\ell' = b\ell$ literals in $\mathbf{X} \setminus \mathbf{R}$, then we have $Z > |C| - b\ell > (1-b)|C| = Bp_*|C|$. Again by the Chernoff bound, this probability is bounded by

$$
\begin{aligned}
\Pr[\, Z > Bp_*|C|\,] \;&\le\; \left(\frac{\mathrm{e}^{(B-1)}}{B^B}\right)^{p_*|C|} \;\le\; \left(\frac{\mathrm{e}^{B}}{B^B}\right)^{p_*\ell} \\
&=\; \exp_2\big(\, (-B\log B + c_{\mathrm{e}}B)p_*\ell\,\big) \;=\; \exp_2\big(\, (-\log B + c_{\mathrm{e}})Bp_*b(1-\delta)\log N\,\big) \\
&=\; \exp_2\big(\, (-\log B + c_{\mathrm{e}})(1-b)b(1-\delta)\log N\,\big) \\
&=\; \exp_2\big(\, -2.1\log N\,\big) \;=\; N^{-2.1}.
\end{aligned}
$$

Since there are at most N^2 clauses in $F|\rho_1$, by the union bound, the probability that some clause has less than ℓ' literals in $\mathbf{X} \setminus \mathbf{R}$ is bounded by $N^{-0.1}$. $\square$

We can derandomize this selection procedure by using the standard method of conditional probabilities. That is, in some fixed order of variables in $\mathbf{X}$, we determine whether each variable is selected to $\mathbf{R}$ or not based on the following conditional probabilities under the selection made so far: the probability that $\|\mathbf{R}\| > 0.1N$ occurs, the probability that $\|\mathbf{R}\| \le (p_*/2)N$ occurs, and the probability that the width of C becomes less than ℓ' for each clause C of $F|\rho_1$. Though tedious, these conditional probabilities are polynomial-time computable, and following the method of conditional probabilities, a deterministic algorithm computes some desired set $\mathbf{R}$ by selecting variables while keeping the sum of these conditional probabilities < 1. Now for defining F', we remove variables in $\mathbf{R}$ from $F|\rho_1$. Also for simplicity we reduce the width of each clause to exactly ℓ' by removing literals (in any fixed way) from clauses with more than ℓ' literals. Let F' be the resulting CNF formula, which is the CNF formula constructed in the second stage. Let N' denote the number of variables of F'; we may assume that $N' \ge 0.99N - 0.1N = 0.89N$.

In the third stage, we apply the algorithmic version of the Lovász Local Lemma stated as Lemma 4 to obtain some *complete* sat. assignment to F'. Then as explained before, this can be regarded as a sat. partial assignment ρ_2 of $F|\rho_1$ that leaves all variables in $\mathbf{R}$ unassigned, and we define $\widehat{\rho} = \rho_2 \circ \rho_1$ as our final partial assignment. We first confirm that Lemma 4 is applicable.

Lemma 7. *Use notation as above. Formula F' obtained in the second stage satisfies the condition of Lemma 4. In particular, we can define a mapping x satisfying (8) w.r.t. some constant $y > 0$.*
Remark. *By "constant y" we mean some number determined by δ and ε but independent from N, F, and F'.*

Proof. Recall that every clause C of F' is of width ℓ' and hence $\Pr[E_C] = 2^{-\ell'}$. We define $x(E_c) = 2^{-b\ell'}$ and show that the first condition of (8) is satisfied w.r.t. some y. The second condition is clear, and the third condition is immediate from the following analysis.

Consider any clause C of F'. Since there is no popular literal, every variable appears at most $2L$ clauses; hence, we have $\|\Gamma(C)\| \le 2L\ell'$. By using this we have

$$
\begin{aligned}
x'(E_C) \;&=\; x(E_C)\prod_{C'\in\Gamma(C)}(1 - x(E_{C'})) \;=\; 2^{-b\ell'}\prod_{C'\in\Gamma(C)}\left(1 - 2^{-b\ell'}\right) \\
&\ge\; 2^{-b\ell'}\left(1 - 2^{-b\ell'}\right)^{2L\ell'} \;\ge\; 2^{-b\ell'}\cdot 0.99\exp\left(-2^{-b\ell'}2L\ell'\right) \\
&=\; 0.99\exp_2\left(-\ell'\left(b + c_{\mathrm{e}}2^{-(b\ell' - \log 2L)}\right)\right).
\end{aligned} \tag{10}
$$

Below we show (if b is chosen appropriately) that $(b\ell' - \log 2L) \to \infty$ (when $N \to \infty$) and hence we may ignore $c_{\rm e}2^{-(b\ell' - \log 2L)}$ (for sufficiently large N). But before moving on to the detailed analysis, we point out that the requirement that L is smaller than 2^ℓ we mentioned before is due to more precise requirement that $\log 2L < b\ell'$ $(= b^2\ell)$.

Let us focus on $b\ell' - \log 2L$. By definition, we have

$$b\ell' - \log 2L \;=\; b^2(1-\delta)\log N - (1-b\gamma)(1-\delta)\log N - 1 \;=\; (b^2 + b\gamma - 1)(1-\delta)\log N - 1.$$

We choose b so that $b^2 + b\gamma - 1 > 0$ holds. Here we define it by

$$b \;=\; 1 - 0.4\gamma. \tag{11}$$

Then it is easy to check that $b^2 + b\gamma - 1 > 0$ if γ is small enough, say, $\gamma < 0.5$. From this it follows that $(b\ell' - \log 2L) \to \infty$. Thus, for sufficiently large N, we have

$$(10) \;\geq\; 0.99\mathrm{exp}_2\big(-\ell'(1 - 0.4\gamma + 0.1\gamma)\big) \;\geq\; \mathrm{exp}_2\big(-\ell'(1-0.2\gamma)\big).$$

Hence, by using $y = 0.2\gamma$, we have

$$\begin{aligned}
(x'(E_C))^{1+y} \;&\geq\; \big(\mathrm{exp}_2(-\ell'(1-0.2\gamma))\big)^{1+y} \;=\; \mathrm{exp}_2\big(-\ell'(1-0.2\gamma)(1+0.2\gamma)\big)\\
&=\; \mathrm{exp}_2\big(-\ell'(1-0.04\gamma^2)\big) \;=\; \mathrm{exp}_2(-\ell')\mathrm{exp}_2(0.04\gamma^2)\\
&\geq\; \mathrm{exp}_2(-\ell') \;=\; \Pr[\,E_C\,].
\end{aligned}$$

□

Finally, we summarize our analysis and prove Theorem 3.

Proof of Theorem 3. For a given formula F, we define $\widehat{\rho} = \rho_2 \circ \rho_2$ as explained just before Lemma 7 by using our technical parameters b and B defined in (11) and (9) respectively. It has been shown in our discussion that $\widehat{\rho}$ is a sat. partial assignment and that it is deterministically computable in $\widetilde{O}(2^{N^\beta})$-time with $\beta = 1 - (1-(\delta+\varepsilon))^2/3$. Also as guaranteed by Lemma 6 it leaves at least $(p_*/2)N$ variables unassigned. On the other hand, we have (by using the assumption $\gamma < 0.5$)

$$\begin{aligned}
\frac{p_*}{2} \;&=\; \frac{1-b}{2}B^{-1} \;=\; \frac{0.4\gamma}{2}\mathrm{e}^{-1}\cdot\mathrm{exp}_2\left(-\frac{2.1}{0.4\gamma(1-0.4\gamma)(1-\delta)}\right)\\
&\leq\; \frac{0.4(1-(\delta+\varepsilon))}{2\mathrm{e}}\mathrm{exp}_2\left(-\frac{2.1}{0.4(1-(\delta+\varepsilon))0.8(1-\delta)}\right)\\
&\leq\; \frac{1-(\delta+\varepsilon)}{c_1}\mathrm{exp}_2\left(-\frac{c_2}{(1-(\delta+\varepsilon))(1-\delta)}\right)
\end{aligned}$$

with some appropriate constants $c_1, c_2 \geq 1$. We use these constants to define α by (3) of the theorem. Then from the above, we have $\alpha N \geq (1-p_*/2)N$, where $(1-p_*/2)N$ bounds the number of fixed variables in $F|\widehat{\rho}$, which proves the theorem. □

References

[1] M. Ajtai, Σ^1_1-formula on finite structures, *Ann. Pure. Appl. Logic* 24, 1–48, 1983.

[2] K. Chandrasekaran, N. Goyal, and B. Haeupler, Deterministic algorithms for the Lovász Local Lemma, in *Proc. of the 21st Annual ACM-SIAM Sympos. on Discrete Algorithms* (SODA'10), SIAM, 992–1004, 2010.

[3] A. De, O. Etesami, L. Trevisan, and M. Tulsiani, Improved pseudo-random generators for depth 2 circuits, in *Proc. the 14th. International Workshop on Randomization and Computation*, (APPROX-RANDOM'10), Lecture Notes in Computer Science 6302, 504–517, 2010.

[4] D. Kane and O. Watanabe, A short implicant of CNFs with relatively many satisfying assignments, *ECCC* TR13-176, 2013.

[5] J. Håstad, *Computational limitations of small depth circuits*, Ph.D. thesis, Massachusetts Institute of Technology, 1987.

[6] E. Hirsch, A fast deterministic algorithm for formulas that have many satisfying assignments, *Logic Journal of the IGPL*, 6(1): 59–71, 1998.

[7] R. Moser, A constructive proof of the Lovász local lemma, in *Proc. of the 41st Annual ACM Sympos. on Theory of Computing* (STOC'09), 343–350, 2009.

[8] R. Moser and G. Tardos, A constructive proof of the general Lovász local lemma, *J. ACM*, 57(2), 2010.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de