Abstract—Random 3CNF formulas constitute an important distribution for measuring the average-case behavior of propositional proof systems. Lower bounds for random 3CNF refutations in many propositional proof systems are known. Most notable are the exponential-size resolution refutation lower bounds for random 3CNF formulas with \( \Omega(n^{1.5-\epsilon}) \) clauses (Chvátal and Szemerédi [13], Ben-Sasson and Wigderson [9]). On the other hand, the only known non-trivial upper bound on the size of random 3CNF refutations in a non-abstract propositional proof system is for resolution with \( \Omega(n^2 / \log n) \) clauses, shown by Beame et al. [5]. In this paper we show that already standard propositional proof systems, within the hierarchy of Frege proofs, admit short refutations for random 3CNF formulas, for sufficiently large clause-to-variable ratio. Specifically, we demonstrate polynomial-size propositional refutations whose lines are \( TC^0 \) formulas (i.e., \( TC^0 \)-Frege proofs) for random 3CNF formulas with \( n \) variables and \( \Omega(n^{1.4}) \) clauses.

The idea is based on demonstrating efficient propositional correctness proofs of the random 3CNF unsatisfiability witnesses given by Feige, Kim and Ofek [19]. Since the soundness of these witnesses is verified using spectral techniques, we develop an appropriate way to reason about eigenvectors in propositional systems. To carry out the full argument we work inside weak formal systems of arithmetic and use a general translation scheme to propositional proofs.

Index Terms—Proof complexity, random 3-SAT, refutation algorithms, threshold logic, Frege proofs

I. INTRODUCTION

This paper deals with the average complexity of propositional proofs. Our aim is to show that standard propositional proof systems, within the hierarchy of Frege proofs, admit short random 3CNF refutations for a sufficiently large clause-to-variable ratio, and also can outperform resolution for random 3CNF formulas in this ratio. Specifically, we show that most 3CNF formulas with \( n \) variables and at least \( cn^{1.4} \) clauses, for a sufficiently large constant \( c \), have polynomial-size in \( n \) propositional refutations whose proof-lines are constant depth circuits with threshold gates (namely, \( TC^0 \)-Frege proofs). This is in contrast to resolution (that can be viewed as depth-1 Frege) for which it is known that most 3CNF formulas with at most \( n^{1.5-\epsilon} \) clauses (for \( 0 < \epsilon < \frac{1}{2} \)) do not admit sub-exponential refutations [13], [9].

The main technical contribution of this paper is a propositional characterization of the random 3CNF unsatisfiability witnesses given by Feige at al. [19]. In particular we show how to carry out certain spectral arguments inside weak propositional proof systems such as \( TC^0 \)-Frege. The latter should hopefully be useful in further propositional formalizations of spectral arguments. This also places a stream of recent results on efficient refutation algorithms using spectral arguments—beginning in the work of Goerdt and Krivelevich [23] and culminating in Feige et al. [19]—within the framework of propositional proof complexity. Loosely speaking, we show that all these refutation algorithms and witnesses, considered from the perspective of propositional proof complexity, are not stronger than \( TC^0 \)-Frege.

A. Background in proof complexity

Propositional proof complexity is the systematic study of the efficiency of proof systems establishing propositional tautologies (or dually, refuting unsatisfiable formulas). Abstractly one can view a propositional proof system as a deterministic polynomial-time algorithm \( A \) that receives a string \( \pi \) ("the proof") and a propositional formula \( \Phi \) such that there exists a \( \pi \) with \( A(\pi, \Phi) = 0 \) iff \( \Phi \) is a tautology. Such an \( A \) is called an abstract proof system or a Cook-Reckhow proof system due to [17]. Nevertheless, most research in proof complexity is dedicated to more concrete or structured models, in which proofs are sequences of lines, and each line is derived from previous lines by “local” and sound rules.

Perhaps the most studied family of propositional proof systems are those coming from propositional logic, under the name Frege systems, and their fragments (and extensions). In this setting, proofs are written as sequences of Boolean formulas (proof-lines) where each line is either an axiom or was derived from previous lines by means of simple sound derivation rules. The complexity of a proof is just the number of symbols it contains, that is, the total size of formulas in it. And different proof systems are compared via the concept of polynomial simulation: a proof system \( P \) polynomially-simulates another proof system \( Q \) if there is a polynomial-time computable function \( f \) that maps \( Q \)-proofs to \( P \)-proofs of the same tautologies. The definition of Frege systems is sufficiently robust, in the sense that different formalizations can polynomially-simulate each other [37].

It is common to consider fragments (or extensions) of Frege proof systems induced by restricting the proof-lines to contain presumably weaker (or stronger) circuit classes than
Boolean formulas. This stratification of Frege proof systems is thus analogous to that of Boolean circuit classes: Frege proofs consist of Boolean formulas (i.e., NC^1) as proof-lines, TC^0-Frege (also known as Threshold Logic) consists of TC^0-proof-lines, Bounded Depth Frege has AC^0 proof-lines, depth-d Frege has circuits of depth-d proof-lines, etc. In this framework, the resolution system can be viewed as depth-1 Frege. Similarly, one usually considers extensions of the Frege system such as NC^1-Frege, for i > 1, and P/poly-Frege (the latter is polynomially equivalent to the known Extended Frege system, as shown by Jeřábek [29]). Restrictions (and extensions) of Frege proof systems form a hierarchy with respect to polynomial-simulations, though it is open whether the hierarchy is proper.

It thus constitutes one of the main goals of proof complexity to understand the above hierarchy of Frege systems, and to separate different propositional proof systems, that is, to show that one proof system does not polynomially simulate another proof system. These questions also relate in a certain sense to the hierarchy of Boolean circuits (from AC^0, through, AC^0[p], TC^0, NC^1, and so forth; see [15]). Many separations between propositional proof systems (not just in the Frege hierarchy) are known. In the case of Frege proofs there are already known separations between certain fragments of it (e.g., separation of depth-d Frege from depth d + 1 Frege was shown by Krajíček [30]). It is also known that TC^0-Frege is strictly stronger than both resolution and bounded depth Frege proof systems, since, e.g., TC^0-Frege admits polynomial-size proofs of the propositional pigeonhole principle, while resolution and bounded depth Frege do not (see [26] for the resolution lower bound, [1] for the bounded depth Frege lower bound and [16] for the corresponding TC^0-Frege upper bound).

Average-case proof complexity via the random 3CNF model. Much as in algorithmic research, it is important to know the average-case complexity of propositional proof systems, and not just their worst-case behavior. To this end one usually considers the model of random 3CNF formulas, where m clauses with three literals each, out of all possible 2^m · (n^3) clauses with n variables, are chosen independently, with repetitions (however, other possible distributions have also been considered in the literature; for a short discussion on these distributions see Section I-C). When m gets greater than cnl for some sufficiently large c (say, c = 5), it is known that with high probability a random 3CNF is unsatisfiable. (As m gets larger the task of refuting the 3CNF becomes easier since we have more constraints to use.) In average-case analysis of proofs we investigate whether such unsatisfiable random 3CNFs also have short (polynomial-size) refutations in a given proof system. The importance of average-case analysis of proof systems is that it gives us a better understanding of the complexity of a system than merely the worst-case analysis. For example, if we separate two proof systems in the average case—i.e., show that for almost all 3CNF one proof system admits polynomial-size refutations, while the other system does not—we establish a stronger separation.

Until now only weak proof systems like resolution and Res(k) (for k ≤ \sqrt{\log n / \log \log n}, the latter system introduced in [32] is an extension of resolution that operates with kDNF formulas) and polynomial calculus (and an extension of it) were analyzed in the random 3CNF model; for these systems exponential lower bounds are known for random 3CNFs (with varying number of clauses) [13], [5], [9], [4], [38], [2], [8], [3], [22]. For random 3CNFs with n variables and n^{1.5-\epsilon} clauses it is known that there are no sub-exponential size resolution refutations [9]. For many proof systems, like cutting planes (CP) and bounded depth Frege (AC^0-Frege), it is a major open problem to prove random 3CNF lower bounds (even for number of clauses near the threshold of unsatisfiability, e.g., random 3CNFs with n variables and 5n clauses). The results mentioned above only concerned lower bounds. On the other hand, to the best of our knowledge, the only known non-trivial polynomial-size upper bound for random kCNFs refutations in any non-abstract propositional proof system is for resolution. This is a result of Beame et al. [5], and it applies for fairly large number of clauses (specifically, \Omega(n^{k-1} / \log n)).

Efficient refutation algorithms. A different kind of results on refuting random kCNFs were investigated in Goerdt and Krivellevich [23] and subsequent works by Goerdt and Lanka [24], Friedman, Goerdt and Krivellevich [21], Feige and Ofek [20] and Feige [18]. Here, one studies efficient refutation algorithms for kCNFs. Specifically, an efficient refutation algorithm receives a kCNF (above the unsatisfiability threshold) and outputs either “unsatisfiable” or “don’t know”; if the algorithm answers “unsatisfiable” then the kCNF is required to be indeed unsatisfiable; also, the algorithm should output “unsatisfiable” with high probability (which by definition, is also the correct answer). Such refutation algorithms can be viewed as abstract proof systems (according to the definition in Subsection I-A) having short proofs on the average-case: A(\Phi) is a deterministic polytime machine whose input is only kCNFs (we can think of the proposed proof \pi input as being always the empty string). On input \Phi the machine A runs the refutation algorithm and answers 1 iff the refutation algorithm answers “unsatisfiable”; otherwise, A can decide, e.g. by brute-force search, whether \Phi is unsatisfiable or not. (In a similar manner, if the original efficient refutation algorithm is non-deterministic then we also get an abstract proof system for kCNFs; now the proof \pi that A receives is the description of an accepting run of the refutation algorithm.)

Goerdt and Krivellevich [23] initiated the use of spectral methods to devise efficient algorithms for refuting kCNFs. The idea is that a kCNF with n variables can be associated with a graph on n vertices (or directly with a certain matrix). It is possible to show that certain properties of the associated graph witness the unsatisfiability of the original kCNF. One then uses a spectral method to give evidence for the desired graph property, and hence to witness the unsatisfiability of the original kCNF. Now, if we consider a random kCNF then the associated graph essentially becomes random too, and so one may show that the appropriate property witnessing the
unsatisfiability of the \( k \text{CNF} \) occurs with high probability in the graph. The best (with respect to number of clauses) refutation algorithms devised in this way work for \( 3 \text{CNFs} \) with at least \( \Omega(n^{1.5}) \) clauses [20].

Continuing this line of research, Feige, Kim and Ofek [19] considered efficient non-deterministic refutation algorithms (in other words, efficient witnesses for unsatisfiability of \( 3 \text{CNFs} \)). They established the currently best (with respect to the number of clauses) efficient, non-deterministic, refutation procedure; they showed that with probability converging to 1 a random \( 3 \text{CNF} \) with \( n \) variables and at least \( cn^{1.4} \) clauses has a polynomial-size witness, for sufficiently big constant \( c \).

The result in the current paper shows that all the above refutation algorithms, viewed as abstract proof systems, are not stronger (on average) than \( TC^0 \)-Frege. The short \( TC^0 \)-Frege refutations will be based on the witnesses from [19], and so the refutations hold for the same clause-to-variable ratio as in that paper.

**B. Our result**

The main result of this paper is a polynomial-size upper bound on random \( 3 \text{CNF} \) formulas refutations in a proof system operating with constant-depth threshold circuits (known as Threshold Logic or \( TC^0 \)-Frege; see Definition 1). Since Frege and Extended Frege proof systems polynomially simulate \( TC^0 \)-Frege proofs, the upper bound holds for these proof systems as well. (The actual formulation of \( TC^0 \)-Frege is not important since different formulations, given in [12], [34], [10], [36], [16], polynomially simulate each other.)

**Theorem 1.** With probability \( 1 - o(1) \) a random \( 3 \text{CNF} \) formula with \( n \) variables and \( cn^{1.4} \) clauses (for a sufficiently large constant \( c \)) has polynomial-size \( TC^0 \)-Frege refutations.

Beame, Karp, Pitassi, and Saks [5] and Ben-Sasson and Wigderson [9] showed that with probability \( 1 - o(1) \) a random \( 3 \text{CNF} \) formula does not admit sub-exponential refutations for random \( 3 \text{CNF} \) formulas when the number of clauses is at most \( n^{1.5 - \epsilon} \), for any constant \( 0 < \epsilon < 1/2 \). Therefore, Theorem 1 shows that \( TC^0 \)-Frege has an exponential speed-up over resolution for random \( 3 \text{CNFs} \) with at least \( cn^{1.4} \) clauses (when the number of clauses does not exceed \( n^{1.5 - \epsilon} \), for \( 0 < \epsilon < 1/2 \)).

The main result contributes to our understanding (and possibly to the development of) refutation algorithms, by giving an explicit logical characterization of the Feige et al. [19] witnesses. This places a stream of recent results on refutation algorithms using spectral methods, beginning in Goerdt and Krivelevich [23], in the propositional proof complexity setting (showing essentially that these algorithms can be carried out already in \( TC^0 \)-Frege).

**C. Relations to previous works**

For weak proof systems like resolution and \( \text{Res}(k) \) there are known exponential- size lower bounds on random \( 3 \text{CNFs} \) with varying number of clauses; with respect to upper bounds, there are known polynomial size resolution refutations on random \( 3 \text{CNF} \) formulas with \( \Omega(n^2 / \log n) \) number of clauses [5]. Below we shortly discuss several known upper and lower bounds on refutations of different distributions than the random \( 3 \text{CNF} \) model (this is not an exhaustive list of all distributions studied).

Ben-Sasson and Bilu [7] have studied the complexity of refuting random 4-Exactly-Half \( \text{SAT} \) formulas. This distribution is defined by choosing at random \( m \) clauses out of all possible clauses with 4 literals over \( n \) variables. A set of clauses is 4-exactly-half satisfiable if there is an assignment that satisfies exactly two literals in each clause. It is possible to show that when \( m = cn \), for sufficiently large constant \( c \), a random 4-Exactly-Half \( \text{SAT} \) formulas with \( m \) clauses and \( n \) variables is unsatisfiable with high probability. Ben-Sasson and Bilu [7] showed that almost all 4-Exactly-Half \( \text{SAT} \) formulas with \( m = n \cdot \log n \) clauses and \( n \) variables do not have sub-exponential resolution refutations. On the other hand, [7] provided a polynomial-time refutation algorithm for 4-Exactly-Half \( \text{SAT} \) formulas.

Another distribution on unsatisfiable formulas that is worth mentioning is \( 3 \text{-LIN} \) formulas over the two element field \( \mathbb{F}_2 \), or equivalently 3XOR formulas. A 3-LIN formula is a collection of linear equations over \( \mathbb{F}_2 \), where each equation has precisely three variables. When the number of randomly chosen linear equations with 3 variables is large enough, one obtains that with high probability the collection is unsatisfiable (over \( \mathbb{F}_2 \)). It is possible to show that the polynomial calculus proof system (see [14] for a definition), as well as \( TC^0 \)-Frege, can efficiently refute such random instances with high probability, by simulating Gaussian elimination.

A different type of distribution over unsatisfiable CNF formulas can possibly be constructed from the formulas (termed proof complexity generators) in Krajíček [33]. We refer the reader to [33] for more details on this.

**D. Overview of the argument**

Here we outline the proof of the main theorem (Theorem 1). We need to construct certain \( TC^0 \)-Frege proofs. For this purpose we use the theory \( VTC^0 \) introduced in [36] (cf. [16]): any proof of a \( \Sigma^p_2 \) formulas in \( VTC^0 \) can be translated into polynomial-size \( TC^0 \)-Frege proofs. Our construction consists of the following steps:

I. Formalize the following statement as a first-order formula:

\[
\forall i \text{ assignment } A \quad (\text{C is a } 3 \text{CNF and } u \text{ is its } FKO \text{ unsatisfiability witness}) \rightarrow (\exists C_i \text{ in } C \text{ such that } C_i(A) = 0)
\]

where an \( FKO \) witness is a suitable formalization of the unsatisfiability witness defined by Feige, Kim and Ofek [19]. The corresponding predicate is called the \( FKO \) predicate.

II. Prove formula (1) in the theory \( VTC^0 \).

III. Translate the proof in Step II into a family of propositional \( TC^0 \)-Frege proofs (of the family of propositional translations of (1)). By Theorem 2 (proved
in [16]), this will be a polynomial-size propositional proof (in the size of $C$). The translation of (1) will consist of a family of propositional formulas of the form:

$$[\text{C is a 3CNF and } w \text{ is its FKO unsatisfi-}]
\rightarrow \exists \text{ a clause } C_1 \text{ in } C \text{ such that } C_1(A) = 0$$

where $[ ]$ denotes the mapping from first-order formulas to families of propositional formulas. By the nature of the propositional translation (second-sort) variables in the original first-order formula translate into a collection of propositional variables. Thus, (2) will consist of propositional variables derived from the variables in (1).

IV. For the next step we first notice the following two facts:

(i) Assume that $C$ is a random 3CNF with $n$ variables and $cm^{1.4}$ clauses (for a sufficiently large constant $c$). By [19], with high probability there exists an FKO unsatisfiability witness $w$ for $C$. Both $w$ and $C$ can be encoded as finite sets of numbers, as required by the predicate for 3CNF and the FKO predicate in (1). Let us identify $w$ and $C$ with their encodings. Then, assuming (I) was formalized correctly, assigning $w$ and $C$ to (1) satisfies the premise of the implication in (1).

(ii) Now, by the definition of the translation from first-order formulas to propositional formulas, if an object $\alpha$ satisfies the predicate $P(X)$ (i.e., $P(\alpha)$ is true in the standard model), then there is a propositional assignment of 0, 1 values that satisfies the propositional translation of $P(X)$. Thus, by Item (i) above, there exists an 0, 1 assignment $\zeta$ that satisfies the premise of (2) (i.e., the propositional translation of the premise of the implication in (1)).

In the current step we show that after assigning $\zeta$ to the conclusion of (2) (i.e., to the propositional translation of the conclusion in (1)) one obtains precisely $\neg C$ (formally, a renaming of $\neg C$, where $\neg C$ is the 3DNF obtained by negating $C$ and using the de Morgan laws).

V. Take the propositional proof obtained in (III), and apply the assignment $\zeta$ to it. The proof then becomes a polynomial-size $TC^0$-Frege proof of a formula $\phi \rightarrow \neg C$, where $\phi$ is a propositional sentence (without variables) logically equivalent to TRUE (because $\zeta$ satisfies it, by (IV)). From this, one can easily obtain a polynomial-size $TC^0$-Frege refutation of $C$ (or equivalently, a proof of $\neg C$).

The bulk of our work lies in (I) and especially in (II). We need to formalize the necessary properties used in proving the correctness of the FKO witnesses and show that the correctness argument can be carried out in the weak theory. There are two main obstacles in this process. The first obstacle is that the correctness of the witness originally is proved using spectral methods, which assumes that eigenvalues and eigenvectors are over the reals; whereas the reals are not defined in our weak theory. The second obstacle is that one needs to prove the correctness of the witness, and in particular the part related to the spectral method, constructively (formally in our case, inside $VTC^0$). Specifically, linear algebra is not known to be (computationally) in $TC^0$, and (proof-complexity-wise) it is conjectured that $TC^0$-Frege do not admit short proofs of the statements of linear algebra, such as statements relating to inverse matrices and the determinant properties (see [39], [28]).

The first obstacle is solved using rational approximations of sufficient accuracy, and showing how to carry out the proof in the theory with such approximations. The second obstacle is solved basically by constructing the argument in a way that exploits non-determinism (i.e., in a way that enables supplying additional witnesses for the properties needed to prove the correctness of the original witness; e.g., all eigenvectors and all eigenvalues of the appropriate matrices in the original witness).

Organization. Sec. II contains the preliminaries. Sec. III shows how to formalize the main formula in the theory. Sec. IV describes the proof of the main formula in the theory except for the treatment of the spectral argument which is deferred to Sec. V. These constitute Steps (I-II) described above. Sec. VI sketches Steps (III-V) which conclude the argument. We refer the reader to the full version [35] for all missing details.

II. PRELIMINARIES

A random 3CNF is generated by choosing independently, with repetitions, $m$ clauses with three literals each, out of all possible $2^3 \binom{n}{3}$ clauses with $n$ variables $x_1, \ldots, x_n$. We say that a property holds with high probability when the probability is $1 - o(1)$. We now define the notion of $TC^0$ formulas and $TC^0$-Frege proofs. The class of $TC^0$ formulas is built up using unbounded fun-in connectives $\land, \lor, \neg$ and threshold gates $Th_i$, for $i \in \mathbb{N}$, where $Th_i(A_1, \ldots, A_n)$ is true if and only if at least $i$ of the $A_k$’s are true. The depth of a formula is the maximal nesting of connectives in it and the size of a formula is the total number of connectives in it.

Definition 1 ($TC^0$-Frege). A $TC^0$-Frege proof system is a sequent calculus with a set of standard sound derivation rules and axioms. We give only the following rule as an example (see the full version of this paper [35] for the other rules):

$$(Th_i, \text{left}): \text{From the sequents } Th_i(A_1, \ldots, A_n), \Gamma \rightarrow \Delta \text{ and } Th_{i-1}(A_2, \ldots, A_n), A_1, \Gamma \rightarrow \Delta \text{ we may infer the sequent } Th_i(A_1, \ldots, A_n), \Gamma \rightarrow \Delta, \text{ for arbitrary } TC^0 \text{ formulas } A_i \text{ and sets } \Gamma, \Delta \text{ of } TC^0 \text{ formulas.}$$

(The intended meaning of $\Gamma \rightarrow \Delta$ is that the conjunction of the formulas in $\Gamma$ implies the disjunction of the formulas in $\Delta$.)
A TC⁰-Frege proof of a formula φ is a sequence of sequents π = (S₁, ..., Sₖ) such that Sₖ = φ and every sequent in it is either an axiom or was derived from previous lines by a derivation rule. The size of the proof π is the total size of all formulas in its sequents. The depth of the proof π is the maximal depth of a formula in its sequents. A TC⁰-Frege proof of a formula φ is a sequence of sequents in a form S₁ ⊢ ... ⊢ Sₖ = φ, where each Sᵢ is a TC⁰ formula that can be derived from some Sⱼ for j < i using the above rules, such that Sᵢ = ¬φ₁ and there is a common constant c bounding the depth of every formula in all the sequences.

Overview of theories of bounded arithmetic. Here we highlight the theories VTC⁰, as defined by Cook and Nguyen [16]. This is a (first-order) two-sorted theory, having a first sort for natural number variables and a second sort for bit strings (formally, they are finite sets of natural numbers whose characteristic vectors are bit strings). The theory VTC⁰ corresponds to TC⁰-Frege (Theorem 2). (For other treatments of theories of bounded arithmetic see [11], [25], [31].)

The language of two-sorted arithmetic, denoted L₂, consists of the following relation, function, and constant symbols: \{ +, \cdot, \leq, 0, 1, |, =₁, =₂, ∈ \}. The intended semantic of this language is the standard model \(\mathbb{N}_2\) of two-sorted arithmetic consisting of a first-sort universe \(U₁ = \mathbb{N}\) and a second-sort universe \(U₂\) of all finite subsets of \(\mathbb{N}\). 0 and 1 are interpreted in \(\mathbb{N}_2\) as zero and one. The functions + and \cdot are addition and multiplication of numbers, \(≤\) is the less-than relation on numbers. The function | \(|\) maps a finite set of numbers to its largest element plus one. The relation =₁ is interpreted as equality between numbers, =₂ is interpreted as equality between finite sets of numbers. The relation ∈ holds for a number n and a finite set of numbers N if and only if n is an element of N. We denote the first-sorted (number) variables by lower-case letters x, y, z, ..., and the second-sort (string) variables by capital letters X, Y, Z, ... We will build formulas in the usual way, using two sorts of quantifiers, number quantifiers and string quantifiers. A number quantifier \(\exists x (\forall x)\) is polynomially bounded if it is of the form \(\exists x (x ≤ f(n) \land \ldots)\) (\(\forall x (x ≤ f(n) \rightarrow \ldots)\)) for some number term f. Given some function symbol f, a formula φ is in \(\Sigma^0_n(f)\) if it uses no string quantifiers and all number quantifiers are polynomially bounded and it possibly uses the function symbol f. We represent a finite set of natural numbers N by a finite string \(S_N = Sₙ₁ \ldots Sₙ_i \ldots \ldots Sₙ[N-1]\) such that \(Sₙ = 1\) if and only if i ∈ N. We will abuse notation and identify N and \(S_N\).

The theory VTC⁰ is meant to allow reasoning that involves counting. Specifically, it enables one to use the function numones(X) whose value is the number of ones in the string X (or equivalently, the number of elements in the set X). We also have the following relation between TC⁰-Frege and VTC⁰:

**Definition 2 (Propositional translation (sketch)).** Let \(\varphi(\vec{X}, \vec{X}')\) be a \(\Sigma^0_n\) formula. The propositional translation of \(\varphi\) is a family \([\varphi]_{\vec{a}, \vec{r}} = \{\varphi\}_{\vec{a}, \vec{r}} | m_i, n_i \in \mathbb{N}\} \) of propositional formulas in variables \(p^X_i\) for every \(X_i \in \vec{X}\). The intended meaning is that \([\varphi]\) is a valid family of formulas if and only if the formula

\[\forall \vec{x} \forall \vec{X} (\bigwedge_i |X_i| = n_i) \rightarrow \varphi(\vec{m}, \vec{X})\]

is true in the standard model \(\mathbb{N}_2\) of two sorted arithmetic. For given \(\vec{m}, \vec{r} \in \mathbb{N}\) we will define \([\varphi]\) by induction on the complexity of the formula \([\varphi]_{\vec{a}, \vec{r}}\).

We can now state the relation between provability of an arithmetical statement \(\varphi\) in VTC⁰ to the provability of the family \([\varphi]\) in TC⁰-Frege as follows.

**Theorem 2** (Section X.4.3. [16]). Let \(\varphi(\vec{x}, \vec{X})\) be a \(\Sigma^0_n\) formula. Then, if VTC⁰ proves \(\varphi(\vec{x}, \vec{X})\) then there is a polynomial size family of TC⁰-Frege proofs of \([\varphi]\).

III. FEIGE-KIM-OFEK WITNESSES AND THE MAIN FORMULA

We now describe the main two-sorted first ordered formula we are going to prove in the theory VTC⁰. Basically, it will formalize the correctness of the Feige et al. witnesses: it will state that if there exists a certain witness with certain properties then there exists a clause in C that is not satisfied by any assignment A (one can think of all the free variables in the formula as universally quantified). To actually construct the main formula, we need to define several predicates.

The formula we construct will speak about 3CNFs \(C = \bigwedge_{\alpha=1}^m C_\alpha\), where \(m\) is the number of variables and \(n\) is the number of clauses. Each clause \(C_\alpha\) is of the form \(x^1 \lor x^2 \lor x^3\), for \(\ell_1, \ell_2, \ell_3 \in \{0, 1\}\), where \(x^1\) abbreviates \(x_i\) and \(x^3\) abbreviates \(\neg x_i\). It is easy to give a \(\Sigma^0_n\)-definition for the predicate Clause \((x, n, m)\) which says that \(x\) is a clause in a 3CNF with \(n\) variables and \(m\) clauses totally. Moreover, for some clause \(C\) and a string variable \(A\) (interpreted as a Boolean assignment), we define by a \(\Sigma^0_n\) formula the predicate NotSAT \((C, A)\), stating that \(C\) is not satisfied under the assignment \(A\).

The following concepts were defined in [19], and are all \(\Sigma^0_n\) (numones)-definable predicates. The imbalance of a variable \(x_i\) is the absolute value of its positive occurrences and negative occurrences in \(C\). The imbalance of \(C\), denoted \(I\), is the sum over the imbalances of all variables; this predicate is denoted Imb \((C, I)\). We define the predicate \(Mat(M, C)\) that holds if \(M\) is an \(n \times n\) rational matrix such that \(M_{i,j}\) equals \(\frac{1}{2}\) times the number of clauses in \(C\) where \(x_i\) and \(x_j\) appear with a different polarity minus \(\frac{1}{2}\) times the number of clauses where they appear with the same polarity. The predicate EigenBound \((M, \lambda, V)\) ensures that given the matrix \(M, \lambda\), \(V\) is a collection of \(n\) rational approximations of the normalized eigenvalues of \(M\) and that \(V\) is the rational matrix whose rows are the rational approximations of the eigenvectors of \(M\). We say that a tuple of \(k\) clauses is an even \(k\)-tuple iff every variable appears an even number of times in the tuple. An even \(k\)-tuple is said to be inconsistent if the total number of negations in its clauses is odd. The predicate COLL \((t, k, d, m, n, M, C, D)\) states that \(D\) is a \((t, k, d)\)-collection of \(t\) inconsistent \(k\)-tuples of \(C\).
such that every single clause appears in at most \( d \) inconsistent \( k \)-tuple (the notation \( o(1) \) below stands for a specific rational number \( b/n^c \), for \( c \) a constant and \( b \) a positive integer).

**Definition 3 (Main Formula).** The Main Formula is the following formula (\( \lambda \) denotes \( n \) distinct number parameters \( \lambda_1, \ldots, \lambda_n \)):

\[
(3\text{CNF}(\mathbf{C}, n, m) \land \text{COLL}(t, k, d, n, m, \mathbf{C}, \mathcal{D}) \land \\
\text{IMB}(\mathbf{C}, I) \land \text{MAT}(M, C) \land \text{EIGVALBOUND}(M, \lambda, V) \land \\
\lambda = \max \{\lambda \} \land t > \frac{d \cdot (I + \lambda_1)}{2} + o(1)
\]

\[
\rightarrow \exists i \leq m \text{ NOTSAT}(\mathbf{C}[i], A).
\]

**IV. PROOF OF THE MAIN FORMULA**

The following is our key theorem:

**Theorem 3 (Key).** The theory \( \text{VTC}^0 \) proves the Main Formula (Definition 3).

The proof in the theory follows the proof of correctness of the unsatisfiability witnesses introduced in Feige et al. [18]. Showing how to carry out this proof in \( \text{VTC}^0 \) constitutes our main body of work. The full details of the proof can be found in the full version of this paper [35].

**Proof of Key Theorem:** We reason inside \( \text{VTC}^0 \). Assume by a way of contradiction that the premise of the implication in the Main Formula holds and that there is an assignment \( A \in \{0, 1\}^m \) (construed as a string variable of length \( n \)) that satisfies every clause in \( \mathbf{C} \). We define the following sets and functions in the theory. Let \text{satLit}(A, C)\ be the string function that outputs the set of all positions of literals in \( \mathbf{C} \) that are satisfied by \( A \). If the literals of a clause are not all true or not all false under \( A \), then we say that the clause is satisfied as NAE (standing for “not all equal”) by \( A \). Let \text{satNAE}(A, C)\ be the string function that returns the set of all clauses in \( \mathbf{C} \) that are satisfied as NAE by \( A \).

**Lemma 1.** (In \( \text{VTC}^0 \)) \( \text{numones} (\text{satLit}(A, C)) \leq \frac{3m + I}{2} \).

This lemma is proved by basic counting in \( \text{VTC}^0 \). We now bound the number of clauses in \( \mathbf{C} \) that contain exactly two literals satisfied by \( A \).

**Lemma 2.** (In \( \text{VTC}^0 \)) Assume the premise of the Main Formula and let \( h \) be the number of clauses in \( \mathbf{C} \) that contain exactly two literals satisfied by \( A \). Then

\[
h \leq \frac{3m + I}{2} - 3m + 2 \cdot \text{numones} (\text{satNAE}(A, C)).
\]

Similar to Lemma 1, Lemma 2 is proved by basic counting in \( \text{VTC}^0 \) as follows. Let \( \ell, h \), and \( g \) be the number of clauses in \( \mathbf{C} \) that have precisely one, two, and three literals satisfied under \( A \), respectively. By Lemma 1 we have \((3m + I)/2 \geq 1 \cdot \ell + 2 \cdot h + 3 \cdot g\). Let \( f = \text{numones} (\text{satNAE}(A, C)) \). Then by definition, \( \ell + 2h = 2f - (f - h) = f + h \). Also, \( g = m - f \) (since by assumption every clause has at least one literal set to true under \( A \)). We therefore get

\[
(3m + I)/2 \geq \ell + 2h + 3g = f + h + 3m - 3f = h + 3m - 2f,
\]

which implies that \( h \leq (3m + I)/2 - 3m + 2f \), concluding the lemma. (See the full version of this paper [35] for the precise formalization of this proof in the theory.)

We now wish to provide an upper bound on the number of clauses in \( \mathbf{C} \) that can be satisfied as NAE by the assignment \( A \), that is, \( \text{numones} (\text{satNAE}(A, C)) \). First we need the technical claim below.

**Notation:** For an assignment \( A \) we define its associated vector \( a \in \{-1, 1\}^n \) such that \( a(i) = 1 \) if \( A(i) = 1 \) and \( a(i) = -1 \) if \( A(i) = 0 \).

**Claim 4 (In \( \text{VTC}^0 \)).** Assuming the premise of the Main Formula holds and let \( f = \text{numones} (\text{satNAE}(A, C)) \). Then, \( a^t A = f - 3(m - f) = 4f - 3m \).

**Proof of claim:** First note that by symmetry of \( M \) we have \( a^t M a = \sum_{i,j \in [n]} a(i)a(j)M_{ij} = \sum_{i,j \in [n]} 2a(i)a(j)M_{ij} \)

The proof follows by the definition of the matrix \( M \). To be able to carry out the proof in the theory we define \( M \) in the following way (in the premise of the Main Formula): \( M_{ij} := \sum_{k < m} E_{ij}^{(k)} \), where \( E_{ij}^{(k)} \) is the contribution of clause \( C_k \) to \( M_{ij} \) as defined in Sec. III. Then,

\[
\sum_{i,j \in [n]} 2a(i)a(j)M_{ij} = \sum_{i,j \in [n]} 2a(i)a(j) \sum_{k < m} E_{ij}^{(k)} = \sum_{k < m} \sum_{i,j \in [n]} 2a(i)a(j)E_{ij}^{(k)}.
\]

Fix some \( 0 \leq k < m \), and consider \( \sum_{i,j \in [n]} 2a(i)a(j)E_{ij}^{(k)} \). If variable \( x_i \) does not occur in clause \( C_k \), then by definition, \( E_{ij}^{(k)} = 0 \) for all \( j \). Thus, considering all six possible cases where \( 2a(i)a(j)E_{ij}^{(k)} \) is nonzero, it is not hard to prove (in \( \text{VTC}^0 \)) that \( 2a(i)a(j)E_{ij}^{(k)} \) is 1 if \( C_k \) is satisfied as NAE by \( A \), and otherwise it is -3. Thus, we can prove that \( a^t M a = 1 \cdot f - 3(m - f) = 4f - 3m \).

**Claim**

**Lemma 3 (In \( \text{VTC}^0 \)).** If the premise of the Main Formula holds then

\[
\text{numones} (\text{satNAE}(A, C)) \leq (n\lambda + 3m)/4 + o(1).
\]

To prove this lemma we reason as follows: let \( f = \text{numones} (\text{satNAE}(A, C)) \). Then, \( 4f - 3m = a^t M a \) by the previous claim. Thus, \( f = (a^t M a + 3m)/4 \). Hence, it suffices to prove the following basic spectral inequality:

**Lemma 4 (Main spectral bound).** (In \( \text{VTC}^0 \)) If \( \text{EIGVALBOUND}(M, \lambda, V) \) holds, then for any assignment \( A \) to \( n \) variables:

\[
a^t M a \leq \lambda n + o(1) .
\]

Proving this spectral bound in the theory is fairly difficult because one has to cope with rational approximations (as the
eigenvalues and eigenvectors might be irrationals, and so undefined in the theory) and further the proof must be sufficiently constructive, in the sense that it would be formalized in the theory \( VTC^0 \). We defer the proof of this lemma to the next section.

We can now finish the proof of the key theorem. In \( VTC^0 \) (and assuming the premise of the Main Formula), let \( h \) be the number of clauses in \( C \) that contain exactly two literals satisfied by \( A \). By Lemmata 2 and 3, we have:

\[
\begin{align*}
  h & \leq \frac{3m + I}{2} - 3m + 2 \cdot \text{numones}(\text{sat} \neg\neg (A, C)) \\
  & \leq \frac{3m + I}{2} - 3m + \frac{3m + \lambda n}{2} + o(1) = \frac{I + \lambda n}{2} + o(1). 
\end{align*}
\]

Since we assumed that \( A \) satisfies \( C \), then every clause in \( C \) has at least one literal satisfied by \( A \). Thus, the clauses in \( C \) that are not satisfied as 3XOR by \( A \) are precisely the clauses that have exactly two literals satisfied by \( A \). By (5), the number of clauses that have exactly two literals satisfied by \( A \) is at most \( \frac{I + \lambda n}{2} + o(1) \). We now use our witness, assumed to exist in the premise of the Main Formula, to show that:

**Lemma 5 (In \( VTC^0 \)).** (Assuming the premise of the Main Formula) the number of clauses in \( C \) that are not satisfied as 3XOR by \( A \) is at least \( \lceil t/d \rceil \).

This concludes the key theorem, since the number of clauses in \( C \) not satisfied as 3XOR by \( A \) is at most \( \frac{I + \lambda n}{2} + o(1) \), and so by Lemma 5 we get that \( t = d \cdot \frac{I}{2} \leq \frac{d \cdot I}{2} \leq \frac{d \cdot I + \lambda n}{2} + o(1) \), which contradicts our assumption (in the Main Formula) that \( t > \frac{d \cdot I + \lambda n}{2} + o(1) \).

To prove Lemma 5 we need the following lemma:

**Lemma 6.** (In \( VTC^0 \)) If \( S \) is an inconsistent (even) \( k \)-tuple, then for every assignment \( A \) to its variables there exists a clause in \( S \) that is not satisfied as 3XOR.

The proof of this lemma is again by a basic counting argument: assume by a way of contradiction that all the clauses in the \( k \)-tuple are satisfied as 3XOR under \( A \). We consider the sum modulo 2 of all literals in the \( k \)-tuple under \( A \). First we consider this sum by summing over clauses; since \( k \) is even, and by assumption all clauses are satisfied as 3XOR under \( A \), we get that summing over clauses evaluates to 0.

Now we consider summing the literals under \( A \) over variables. It is not hard to show that since every number occurs even times in the \( k \)-tuple and since the number of negative literals (and hence also the number of positive literals) is odd, summing the literals over variables evaluates to 1, modulo 2, and we obtain a contradiction.

It is important to note that the reason why we can carry out the above argument in \( VTC^0 \) is that we can define (and reason about) the parity of big sums of numbers in the theory.

**V. The Spectral Bound**

In this section we show how to prove inside \( VTC^0 \) the desired spectral inequality in Lemma 4. Since the original matrix associated to a 3CNF is a real symmetric matrix, and its eigenvectors and eigenvalues might also be real, and thus cannot be represented in our theory \( VTC^0 \), we shall need to work with rational approximations of real numbers. We will work with polynomially small approximations. Specifically, a real number \( r \) in the real interval \([-1, 1]\) is represented with precision \( 1/n^c \), where \( n \) is the number of variables in the 3CNF and \( c \) is a constant natural number independent of \( n \) (that is, if \( \bar{r} \) is the approximation of \( r \), we shall have \(|r - \bar{r}| \leq 1/n^c \)). For convenience we will assume that all rational numbers have in fact the same denominator \( n^c \) for some specific global constant \( c \).

Let us explain informally how to prove the bound \( a^t M a \leq \lambda n + o(1) \), for any \( a \in \{-1, 1\}^n \), in the theory \( VTC^0 \), assuming that \( \text{EigValBound}(M, \lambda, \bar{V}) \) (and \( \text{MAT}(M, C) \)) hold (see Lemma 4). The idea is as follows: in the predicate \( \text{EigValBound}(M, \lambda, \bar{V}) \) we certify that the rows of a given matrix \( V \) are rational approximations of the normalized eigenvector basis of \( M \). Since \( M \) is symmetric and real, \( V \) will approximate an orthonormal matrix, and \( V^-1 \) will approximate \( V^{-1} \) (this is where we circumvent the need to prove the correctness of inverting a matrix in the theory \( VTC^0 \): instead of proving the existence of an inverse matrix, we simply assume that there exists an object which [approximates] the inverse matrix of \( V \)). Thus, \( V^-1 \) approximates the matrix of the basis transformation from the standard basis to the eigenvector basis. Note that \( a \) (as a \( \{-1, 1\} \) vector) is already almost described in the standard basis. Hence, it will be possible to prove in the theory that \( V^-1 a \) is the representation of \( a \) in the (approximate) eigenvector basis, i.e., we shall have an equality \( a = \sum^n_{i=1} \gamma_i v_i + o(1) \), for \( v_i \)'s the approximate eigenvectors of \( M \) and some rationals \( \gamma_i \)'s. After plugging-in this equality in \( a^t M a \), to prove \( a^t M a \leq \lambda n \) we only need to validate computations—using also the fact that we know the inequalities \( M v_i \leq \lambda v_i + o(1) \), for any \( i \in [n] \), hold (since this will be stated in the predicate \( \text{EigValBound}(M, \lambda, \bar{V}) \)).

**Notation:** We denote by \( e_1, \ldots, e_n \) the standard basis vectors spanning \( \mathbb{Q}^n \). That is, for any \( 1 \leq i \leq n \) the vector \( e_i \in \mathbb{Q}^n \) is 1 in the \( i \)-th coordinate and all other coordinates are 0. For a vector \( v \) we denote by \( v(j) \) the \( j \)-th entry in \( v \). Given a real symmetric matrix \( M \) we denote by \( u_1, \ldots, u_n \in \mathbb{R}^n \) the normalized eigenvectors of \( M \). It is known that the collection of normalized eigenvectors of a symmetric \( n \times n \) real matrix \( M \) forms an orthonormal basis for \( \mathbb{R}^n \), called the eigenvector basis of \( M \) (cf. [27]). The (rational) approximation of the eigenvectors will be denoted \( v_1, \ldots, v_n \in \mathbb{Q}^n \) and we define \( v_i := v_i(j) \). Recall that for a real or rational vector \( v = (v_1, \ldots, v_n) \) we denote by \( \|v\|^2 = v_1^2 + \ldots + v_n^2 \). We also define \( \|v\|_\infty := \max\{v_i : 1 \leq i \leq n\} \).

**A. Rational approximations of real numbers, vectors and matrices**

**Definition 4 (Rational \( \varepsilon \)-approximation of a real number).** For \( r \in \mathbb{R} \), we say that \( q \in \mathbb{Q} \) is a rational \( \varepsilon \)-approximation of \( r \)
(or just $\varepsilon$-approximation), if $|r - q| \leq \varepsilon$.

Claim 5. For any real number $r \in [-1, 1]$ and any natural number $m$ there exists a $1/m$-approximation of $r$ whose numerator and denominator have values linearly bounded in $m$.

In a similar fashion we have:

**Definition 5** (Rational $\varepsilon$-approximation of (sets of) real vectors). Let $0 < \varepsilon < 1$. For $u \in \mathbb{R}^n$, we say that $v \in \mathbb{Q}^n$ is an $\varepsilon$-approximation of $u$, if $v(i)$ is an $\varepsilon$-approximation of $u(i)$, for all $i = 1, \ldots, n$. Accordingly, for a set $U = \{u_1, \ldots, u_k\} \subseteq \mathbb{R}^n$, we say that $V = \{v_1, \ldots, v_k\} \subseteq \mathbb{Q}^n$ is a (rational) $\varepsilon$-approximation of $U$ if every $v_i \in \mathbb{Q}^n$ is an $\varepsilon$-approximation of the vector $u_i$, $i = 1, \ldots, n$.

**B. The predicate $\text{EIGVALBOUND}$**

We define the predicate $\text{EIGVALBOUND}(M, \bar{\lambda}, V)$ which is meant to express the properties needed for the main proof. Basically, $\text{EIGVALBOUND}(M, \bar{\lambda}, V)$ expresses the fact that $V$ is a rational $1/n^c$-approximation of $M$, whose $1/n^c$-approximate eigenvalues (in decreasing order with respect to value) are $\bar{\lambda}$, for a sufficiently large constant $c \in \mathbb{N}$.

**Definition 6** (EIGVALBOUND predicate). The predicate $\text{EIGVALBOUND}(M, \bar{\lambda}, V)$ is a $\Sigma_0^B$-definable relation in $\text{VTC}^0$ that holds (in the standard two-sorted model) iff all the following properties hold (where $c \in \mathbb{N}$ is a sufficiently large global constant):

1. $V$ is a sequence of $n$ vectors $v_1, \ldots, v_n \in \mathbb{Q}^n$ with polynomially small entries. That is, for any $1 \leq i, j \leq n$, the rational number $v_{ij} := v_i(j) \in \mathbb{Q}$ is polynomial in $n$ (meaning that both its denominator and numerator are polynomially bounded in $n$).

2. For any $1 \leq i, j \leq n$, $v_{ij}$ has the absolute value $|v_{ij}| \leq 1$.

3. For any $1 \leq i \leq n$, define $\tilde{c}_i := \sum_{j=1}^n v_{ij} \cdot v_j$. Then, there exists $r_i \in \mathbb{Q}^n$ for which $\tilde{c}_i = r_i$ and $\|r_i\|_\infty = O(1/n^{c-1})$. To formalize the existence of such an $r_i$, we do not use an existential second-sort quantifier here; instead, we simply assert that for any $\ell = 1, \ldots, n$: $|\tilde{c}_i(\ell) - e_i(\ell)| = O(1/n^{c-1}).$

4. The vectors in $V$ are “almost” orthonormal, in the following sense: $\langle v_i, v_j \rangle = O(1/n^{c-1})$, for all $1 \leq i \neq j \leq n$ and $\langle v_i, v_i \rangle = 1 + O(1/n^{c-1})$, for all $1 \leq i \leq n$.

5. The parameter $\bar{\lambda}$ is a sequence $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ of rational numbers such that for every $1 \leq i \leq n$, there exists a vector $t_i \in \mathbb{Q}^n$ for which $\|t_i\|_\infty = O(1/n^{c-2})$, and $M \cdot t_i = \lambda_i t_i$.

It is not hard to check that $\text{EIGVALBOUND}(M, \bar{\lambda}, V)$ is a $\Sigma_0^B$-definable relation in $\text{VTC}^0$.

Now we show that there exist objects $M, \bar{\lambda}, V$ that satisfy the predicate $\text{EIGVALBOUND}(M, \bar{\lambda}, V)$.

**Proposition 6** (Suitable approximations of eigenvector bases exist). Let $M$ be an $n \times n$ real symmetric matrix whose entries are quadratic in $n$. Let $U = \{u_1, \ldots, u_n\} \subseteq \mathbb{R}^n$ be the orthonormal basis consisting of the eigenvectors of $M$, let $c \in \mathbb{N}$ be positive and constant (independent of $n$). If $V = \{v_1, \ldots, v_n\} \subseteq \mathbb{Q}^n$ is an $1/n^c$-approximation of $U$ (Definition 5), $\bar{\lambda} = \{\lambda_1, \ldots, \lambda_n\}$ is the collection of rational $1/n^c$-approximations of the real eigenvalues of $M$ such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, then $\text{EIGVALBOUND}(M, \bar{\lambda}, V)$ holds (as before, the predicate holds in the standard two-sorted model, for the appropriate encodings of its parameters).

The proof of this proposition proceeds by checking that all the conditions in Definition 6 hold for $V$ and $\bar{\lambda}$. This can be done in a somewhat straightforward manner (note that we do not need to certify the existence of $\bar{\lambda}$ and $V$ in the theory).

**C. Certifying the spectral inequality**

In this section we show that the theory $\text{VTC}^0$ can prove that, if $\text{EIGVALBOUND}(M, \bar{\lambda}, V)$ holds, then the desired spectral inequality also holds.

For an assignment $A \in \{0, 1\}^n$ we define its associated vector $a \in \{-1, 1\}^n$ such that $a(i) = 1$ if $A(i) = 1$ and $a(i) = -1$ if $A(i) = 0$. We define

$$\tilde{a} := \sum_{i=1}^n a(i) \cdot \tilde{c}_i,$$

and recall that $\tilde{c}_i := \sum_{j=1}^n v_{ij} \cdot v_j$ is a rational approximation of $c_i$ (Definition 6). We let $a' M a$ abbreviate $\langle a, M a \rangle$ (which is $\Sigma_0^B$-definable in $\text{VTC}^0$).

**Lemma 7** (Main spectral bound). The theory $\text{VTC}^0$ proves that if $A$ is an assignment to $n$ variables (that is, $A$ is a string variable of length $n + 1$) and $\text{EIGVALBOUND}(M, \bar{\lambda}, V)$ holds, then

$$a' M a \leq \lambda n + o(1).$$

This is a corollary of Lemma 8 and Lemma 9 that follow.

**Lemma 8.** The theory $\text{VTC}^0$ proves that for any assignment $A$ to $n$ variables, $\text{EIGVALBOUND}(M, \bar{\lambda}, V)$ implies:

$$a' M a \leq \tilde{a} M \tilde{a} + O(1/n^{c-5}),$$

where $c$ is the constant from the $\text{EIGVALBOUND}(M, \bar{\lambda}, V)$ predicate.

**Proof:** First note that $A$ is a string variable of length $n$. By Definition 6 for any $1 \leq j \leq n$ there exists a vector $r_j \in \mathbb{Q}^n$ such that $\tilde{c}_j = c_j + r_j$, and where $\|r_j\|_\infty = O(1/n^{c-1})$. Therefore, by (6):

$$\tilde{a} = \sum_{i=1}^n a(i) \tilde{c}_i = \sum_{i=1}^n a(i) (c_i + r_i) = \sum_{i=1}^n a(i) c_i + \sum_{i=1}^n a(i) r_i.$$

Note that $\sum_{i=1}^n a(i) c_i = a$, and let $r := \sum_{i=1}^n a(i) r_i$. Then, $\tilde{a} = a + r$, and since $a(i) \in \{-1, 1\}$, we have $\|r\|_\infty = O(1/n^{c-2})$. Now, proceed as follows:

$$a' M a = (\tilde{a} - r') M (\tilde{a} - r) = \tilde{a} M \tilde{a} - r' M \tilde{a} + r' M r.$$

508
We now claim that (provably in \(VTC^0\)) the three right terms in (8) are \(o(1)\):

**Claim 7.** The theory \(VTC^0\) proves that for any assignment \(A\) to \(n\) variables, \(EIGVALBOUND(M, \vec{\lambda}, V)\) implies \(-\vec{a}^tM\vec{r} - r^tMa + r^tMr = O(1/n^{c-\delta})\).

Claim 7 concludes the proof of Lemma 8.

**Claim 8.** There is a constant \(c'\) such that the theory \(VTC^0\) proves that \(EIGVALBOUND(M, \vec{\lambda}, V)\) implies that:

\[
\langle \vec{\lambda}_i, \vec{\lambda}_j \rangle = O(1/n^{c'}), \quad \text{for any } 1 \leq i < n, \text{ and} \quad \langle \vec{\lambda}_i, \vec{\lambda}_j \rangle = O(1/n^{c'}), \quad \text{for any } 1 \leq i \neq j \leq n.
\]

**Lemma 9.** The theory \(VTC^0\) proves that for any assignment \(A\) to \(n\) variables, \(EIGVALBOUND(M, \vec{\lambda}, V)\) implies:

\[
\vec{a}^tM\vec{a} \leq \lambda n + o(1). \tag{9}
\]

**Proof:** We have:

\[
\vec{a}^tM\vec{a} = \vec{a}^tM\sum_{i=1}^{n}a(i)\vec{e}_i = \vec{a}^tM\sum_{i=1}^{n}(a(i) \cdot \sum_{j=1}^{n}v_{ij}v_j) = \vec{a}^t\sum_{i=1}^{n}(a(i) \cdot \sum_{j=1}^{n}v_{ij}Mv_j).
\]

\[
= \vec{a}^t\sum_{i=1}^{n}(a(i) \cdot \sum_{j=1}^{n}v_{ij}(\lambda_jv_j + r_j)) = \vec{a}^t\sum_{i=1}^{n}(a(i) \cdot \sum_{j=1}^{n}\lambda_jv_jv_j) + \vec{a}^t\sum_{i=1}^{n}(a(i) \cdot \sum_{j=1}^{n}v_{ij}r_j) \tag{10}
\]

Similar to the proof of Claim 7, the second term in (10) above can be proved to be of size \(o(1)\).

It remains to bound the first term in (10):

\[
\vec{a}^t\sum_{i=1}^{n}(a(i) \cdot \sum_{j=1}^{n}\lambda_jv_jv_j). \tag{11}
\]

By the definition of \(\vec{a}\) in (6) and the definition of the \(\vec{e}_i\)'s, we get that (11) equals:

\[
\sum_{i=1}^{n}(a(i) \cdot \sum_{j=1}^{n}v_{ij}v_j) \cdot \sum_{i=1}^{n}(a(i) \cdot \sum_{j=1}^{n}\lambda_jv_jv_j). \tag{12}
\]

We can prove in \(VTC^0\) that for any vectors \(b_1, \ldots, b_\ell \in Q^n\) and any rational numbers \(c_1, \ldots, c_\ell\) such that \(\ell = \max\{\xi_i : 1 \leq i \leq \ell\}\), we have

\[
\sum_{i=1}^{\ell}c_i b_i, \sum_{i=1}^{\ell}c_i b_i, \leq \xi : \sum_{i=1}^{\ell}c_i b_i, \sum_{i=1}^{\ell}c_i b_i).
\]

Therefore, we can prove in \(VTC^0\) that (12) is at most:

\[
\lambda \cdot \sum_{i=1}^{n}(a(i) \cdot \sum_{j=1}^{n}v_{ij}v_j) \cdot \sum_{i=1}^{n}(a(i) \cdot \sum_{j=1}^{n}v_{ij}v_j) = \lambda \cdot \sum_{i=1}^{n}(a(i)\vec{e}_i) \cdot \sum_{i=1}^{n}(a(i)\vec{e}_i)
\]

This concludes the proof of Lemma 9.

**VI. WRAPPING UP THE PROOF**

In this section we finally establish Theorem 1. Note that the Main Formula is a \(\Sigma_0^B(\mathcal{L})\) formula, where the language \(\mathcal{L}\) contains function symbols not in \(L^\mathcal{L}_n\), and in particular it contains the *numones* function. To use Theorem 2 we need to convert the Main Formula into a \(\Sigma_0^B\) formula (in the language \(L^\mathcal{L}_n\)). It suffices to show that \(VTC^0\) proves that the main formula is equivalent to a \(\forall\Sigma_0^B\) formula, since if \(VTC^0\) proves a \(\forall\Sigma_0^B\) formula \(\forall\Phi\), it also proves the \(\Sigma_0^B\) formula obtained by discarding all the universal quantifiers in \(\forall\Phi\). We omit the details of this conversion (see the full version [35]), and assume from now on that the Main Formula is in fact a \(\Sigma_0^B\) formula.

The following is a restatement of the main theorem in [19]:

**Theorem 9 (19)].** Let \(C\) be random 3CNF with \(n\) variables and \(m = c \cdot n^{1.4}\) clauses where \(c\) is sufficiently large constant. Then, with probability converging to 1, the following holds:

1) There exists an \(I = O(n^{1.2})\) such that IMB\((C, I)\).
2) There exists an \(1/n^{c'}\)-rational approximation \(V\) of the eigenvector matrix of \(M\) and an \(1/n^{c'}\)-rational approximations \(\vec{\lambda}\) of the eigenvalues of \(M\), for some constant \(c' > 6\); in other words, \(EIGVALBOUND(M, \vec{\lambda}, V)\) and \(MAT(M, C)\) hold. And the \(1/n^{c'}\)-rational approximation \(\lambda\) of the largest eigenvalue of \(M\) satisfies \(\lambda = O(n^{1/2})\).
3) There are natural numbers \(k = O(n^{0.2}), t = \Omega(n^{1.4})\), \(d = O(k) = O(n^{0.2})\) and a sequence \(Z\) of \(t\) inconsistent \(Z\)-tuples such that \(COLL(t, k, d, n, m, C, \mathcal{D})\) holds, and such that: \(t > d(I + \lambda n)^2 + o(1)\).

Recall the premise in the implication in the main formula:

\(3CNF(C, n, m) \land COLL(t, k, d, n, m, C, \mathcal{D}) \land IMB(C, I) \land MAT(M, C) \land EIGVALBOUND(M, \vec{\lambda}, V) \land \lambda = \max(\vec{\lambda}) \land t > d(I + \lambda n)^2 + o(1)\).

Let \(PREM(C, n, m, t, k, d, \mathcal{D}, I, \vec{\lambda}, V, M, \lambda, \vec{Z})\) be the formula obtained after transforming the main formula into a \(\forall\Sigma_0^B\) formula (where \(\vec{Z}\) is a sequence of string variables for “counting sequences” added for technical reasons after the transformation). The following is a simple claim about the propositional translation (given without a proof):

\[ \lambda = \max(\vec{\lambda}) \land t > d(I + \lambda n)^2 + o(1). \]
Claim 10. If a $\sum_2^B$ formula $\varphi(\bar{x}, \bar{X})$ can be evaluated to a true sentence in $\mathbb{N}_2$ by assigning numbers $\bar{x}$ and sets $\bar{X}$ to the appropriate variables, then the translation $\overrightarrow{\varphi(\bar{x}, \bar{X})}$ is satisfiable.

Using this claim and the fact that the Main Formula is provable in $VTC^0$ by Theorem 3, we obtain the following:

Lemma 10. For every $m, n \in \mathbb{N}$ and every unsatisfiable 3CNF formula $C$ with $m$ clauses and $n$ variables such that $PREM(C, n, m, \ldots)$ is true for some assignment to the remaining variables (i.e. to the unspecified variables denoted by “...”), this also implies that $\overrightarrow{PREM(C, n, m, \ldots)}$ is satisfiable, there exists a polynomially bounded $TC^0$-Frege proof of $\neg C$ (i.e. the sequent $\rightarrow\neg C$ can be derived).

Due to lack of space we omit the proof of this lemma (see [35]). Finally, we can conclude Theorem 1:

Corollary. With probability converging to 1, a random 3CNF $C$ with $n$ variables and $m \geq c \cdot n^{1.4}$ clauses, $c$ a sufficiently large constant, $\neg C$ has polynomially bounded $TC^0$-Frege proofs, while $C$ has no sub-exponential size resolution refutations (as long as $mn = O(n^{1.5-\epsilon})$, for $0 < \epsilon < 1/2$).

ACKNOWLEDGMENTS

We wish to thank Jan Krajíček for very helpful discussions concerning the topic of this paper and for commenting on an earlier manuscript, and Emil Jeřábek and Neil Thapen for answering many of our questions about theories of weak arithmetic and for many other insightful comments. We also wish to thank the anonymous referees for greatly improving the exposition of this paper. The first author was supported by the Marie Curie Initial Training Network in Mathematical Logic - MALOA - From Mathematical Logic to Applications, PITN-GA-2009-238381. The second author was supported in part by the National Basic Research Program of China Grant 2011CBA00300, 2011CBA00301, the National Natural Science Foundation of China Grant 61033001, 61061130540, 61073174, and by The Eduard Čech Center for Algebra and Geometry and The John Templeton Foundation.

REFERENCES

[1] Miklós Ajtai. The complexity of the pigeonhole principle. In Proc. 29th Ann. IEEE Sympos. Found. Comp., 346–355, 1988.
[2] Michael Alekhnovich. Lower bounds for $k$-DNF resolution on random 3-CNFs. In Proceedings of the Annual ACM Symposium on the Theory of Computing, 251–256, 2005.
[3] Michael Alekhnovich and Alexander A. Razborov. Lower bounds for polynomial calculus: non-bimonial case. In Proc. 42nd Ann. IEEE Symp. Found. Comp., 190-199, 2001.
[4] A. Atserias, Maria Luisa Bonet, and J. Esteban. Lower bounds for the weak pigeonhole principle and random formulas beyond resolution. Inform. and Comp., 176:152–136, 2002.
[5] Paul Beame, Richard Karp, Toniann Pitassi, and Michael Saks. The efficiency of resolution and Davis-Putnam procedures. SIAM J. Comput., 31(4):1048–1075, 2002.
[6] Eli Ben-Sasson. Expansion in Proof Complexity. PhD thesis, Hebrew University, Jerusalem, Israel, September 2001.
[7] Eli Ben-Sasson and Yonatan Bilu. A gap in average proof complexity. Elec. Colloq. Comput. Complexity (ECCC), 2002. TR02-003.
[8] Eli Ben-Sasson and Russell Impagliazzo. Random CNFs are hard for the polynomial calculus. Comput. Complexity, pages 1–19, 2010.
[9] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow—resolution made simple. J. ACM, 48(2):149–169, 2001.
[10] Maria Luisa Bonet, Toniann Pitassi, and Ran Raz. On interpolation and automatization for Frege systems. SIAM J. Comput., 29(6):1939–1967, 2000.
[11] Samuel R. Buss. Bounded Arithmetic, volume 3 of Studies in Proof Theory. Bibliopolis, 1986.
[12] Samuel R. Buss and Peter Clote. Cutting planes, connectivity, and threshold logic. Arch. Math. Logic, 35(1):33–62, 1996.
[13] Vašek Chvátal and Endre Szemerédi. Many hard examples for resolution. J. Assoc. Comput. Mach., 35(4):759–768, 1988.
[14] Matthew Clegg, Jeffery Edmonds, and Russell Impagliazzo. Using the Groebner basis algorithm to find proofs of unsatisfiability. In Proceedings of the 28th Annual ACM Symposium on the Theory of Computing pages 174–183, 1996.
[15] Stephen Cook. Theories for Complexity Classes and Their Propositional Interpolations. Translations, pages 175–227. Complexity of computations and proofs, Jan Krajíček, ed. Quaderni di Matematica, 2005.
[16] Stephen Cook and Phuong Nguyen. Logical Foundations of Proof Complexity. ASL Perspectives in Logic. Cambridge, 2010.
[17] Stephen A. Cook and Robert A. Reckhow. The relative efficiency of propositional proof systems. J. Symbolic Logic, 44(1):36–50, 1979.
[18] Uriel Feige. Refuting smoothed 3CNF formulas. In Proc. 48th Ann. IEEE Sympos. Found. Comp., 407–417, 2007.
[19] Uriel Feige, Jeong Han Kim, and Eran Ofek. Witnesses for non-satisfiability of dense random 3CNF formulas. In Proc. 47th Ann. IEEE Sympos. Found. Comp., 2006.
[20] Uriel Feige and Eran Ofek. Easily refutable subformulas of large random 3CNF formulas. Theory of Computing, 3(1):25–43, 2007.
[21] Joel Friedman, Andreas Goerdt, and Michael Krivelevich. Recognizing more unsatisfiable random $k$-SAT instances efficiently. SIAM J. Comput., 35(2):408–430, 2005.
[22] Nicola Galesi and Massimo Lauria. Optimality of size-degree trade-offs for polynomial calculus. ACM Trans. Comput. Log., 12(1), 2011.
[23] A. Goerdt and M. Krivelevich. Efficient recognition of random unsatisfiable $k$-SAT instances by spectral methods. In Annual Symposium on Theoretical Aspects of Computer Science, pages 294–304, 2001.
[24] Andreas Goerdt and André Lanka. Recognizing more random unsatisfiable 3-SAT instances efficiently. Electronic Notes in Discrete Mathematics, 16:21–46, 2003.
[25] P. Hajek and P. Pudlak. Metamathematics of First-order Arithmetic. Perspectives in Mathematical Logic. Springer-Verlag, 1993.
[26] Armin Haken. The intractability of resolution. Theoret. Comput. Sci., 39(2-3):297–308, 1985.
[27] Roger A. Horn and Charles R. Johnson. Matrix Analysis. Cambridge University Press, 1985.
[28] Pavel Hrubes and Iddo Tzameret. Short proofs for the determinant identity. In STACS’11, 33-44, 2011. TR11-006.
[29] Phuong Nguyen and Stephen A. Cook. Theories for Complexity Classes and Their Propositional Interpolations. Translations, pages 175–227. Complexity of computations and proofs, Jan Krajíček, ed. Quaderni di Matematica, 2005.
[30] Jan Krajíček. Bounded arithmetic, propositional logic, and complexity theory, volume 60 of Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1995.
[31] Jan Krajíček. On the weak pigeonhole principle. Fund. Math., 170(1-2):123–140, 2001.
[32] Jan Krajíček. A proof complexity generator. Proc. 15th Intern. Congress of Logic, Methodology and Philosophy of Science, 2007. Stud. Logic and the Found. Math. 2009.
[33] Alexei Maciel and Toniann Pitassi. On ACC$^0[p^k]$ Frege proofs. In Proceedings of the Annual ACM Symposium on the Theory of Computing 1997, pages 720–729, 1999.
[34] Sebastian Müller and Iddo Tzameret. Short Propositional Refutations for Dense Random 3CNF Formulas. Elect. Colloq. Comput. Complexity (ECCC), 2011. TR11-006.
[35] Phuong Nguyen and Stephen A. Cook. Theories for $TC^0$ and other small complexity classes. Log. Meth.in Comp. Sci., 2(1), 2006.
[36] Robert Reckhow. On the lengths of proofs in the propositional calculus. PhD thesis. University of Toronto, 1976.
[37] Nathan Segerlind, Sam Buss, and Russell Impagliazzo. A switching lemma for small complexity classes. J. Assoc. Comput. Mach., 2(1), 2006.
[38] Michael Soltys and Stephen Cook. The proof complexity of linear algebra. Ann. Pure Appl. Logic, 130(1-3):277–323, 2004.