Finding Bugs in Short Proofs: The Metamathematics of Resolution Lower Bounds

Jiawei Li
University of Texas at Austin
Austin, Texas, USA
davidlee@cs.utexas.edu

Yuhao Li
Columbia University
New York, New York, USA
yuhaoli@cs.columbia.edu

Hanlin Ren
Institute for Advanced Study
Princeton, New Jersey, USA
h4n1in.r3n@gmail.com

Abstract

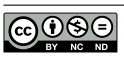
We study the *refuter* problems for proof complexity lower bounds. Suppose φ is a hard tautology that does not admit any length-s proof in some proof system P. In the corresponding refuter problem, we are given (query access to) a purported length-s proof π in P that claims to have proved φ, and our goal is to find an invalid derivation step within π. As suggested by witnessing theorems in bounded arithmetic, the *computational complexity* of these refuter problems is closely tied to the *metamathematics* of the underlying lower bounds.

We focus on refuter problems corresponding to lower bounds for *resolution*, which is arguably the single most studied system in proof complexity. As a warm-up, we show that many refuter problems for resolution *width* lower bounds are PLS-complete. To capture the complexity of refuter problems for resolution *size* lower bounds, we introduce a new class rwPHP(PLS) in decision-tree TFNP, which can be seen as a randomized version of PLS.

- We show that the refuter problems for many resolution size lower bounds can be solved in rwPHP(PLS), including the classic lower bound of Haken [TCS, 1985] for the pigeonhole principle. More generally, we identify a common proof technique that we call "random restriction + width lower bound", and present strong evidence that resolution lower bounds proved by this technique typically have refuter problems in rwPHP(PLS).
- We then show that the refuter problem for *any* resolution size lower bound is rwPHP(PLS)-hard, thereby demonstrating that the rwPHP(PLS) upper bound mentioned above is tight. Informally speaking, this means that "rwPHP(PLS)-reasoning" is *necessary* for proving *all* resolution size lower bounds.

Interpreted in bounded arithmetic, our results show that the theory $\mathrm{T}^1_2(\alpha)+\mathrm{dwPHP}(\mathrm{PV}(\alpha))$ characterizes the "reasoning power" required to prove (the "easiest") resolution size lower bounds.

As a corollary, we obtain surprisingly efficient proofs of resolution lower bounds. In particular, we show that many resolution size lower bounds can be proved in low-width *random resolution* [Pudlák–Thapen, CCC'17].



STOC '26, Salt Lake City, UT, USA

ACM ISBN 979-8-4007-2536-4/2026/06
https://doi.org/10.1145/3798129.3800793

CCS Concepts

• **Theory of computation** → **Complexity classes**; **Proof complexity**; **Complexity theory and logic**; *Problems, reductions and completeness.*

Keywords

Proof complexity, TFNP, Resolution, Refuter Problems, Metamathematics

ACM Reference Format:
Jiawei Li, Yuhao Li, and Hanlin Ren. 2026. Finding Bugs in Short Proofs: The Metamathematics of Resolution Lower Bounds. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing (STOC '26), June 22–26, 2026, Salt Lake City, UT, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3798129.3800793

1 Introduction

One of the earliest lower bounds in proof complexity was Haken's landmark result [47], showing that the pigeonhole principle requires exponential-size proofs in the resolution proof system. Since then, proof complexity has become a vibrant research area with substantial progress in establishing lower bounds for various proof systems, as well as the development of a wide range of lower bound techniques. However, despite decades of efforts, proving nontrivial lower bounds for stronger systems, such as Frege and Extended Frege, remains elusive. It is widely believed that proving lower bounds for Extended Frege is "beyond our current techniques"[1], but what does this even mean? How much, and in which directions, must our techniques expand, to enable us to prove lower bounds for stronger proof systems? These questions call for a study of the *metamathematical* difficulty of proving lower bounds in proof complexity (see, e.g., [82, 93]).

Inspired by recent works on the reverse mathematics of *circuit lower bounds* [23, 25, 26, 59], we propose investigating the metamathematics of proof complexity lower bound through the computational lens of their *refuter* problem. To illustrate, consider the following total search problem: suppose we are given a resolution proof Π that claims to prove the pigeonhole principle, yet its length is shorter than the lower bound established in [47]. By Haken's result, Π cannot be a valid resolution proof; it must contain an invalid derivation. The goal of the search problem is to locate such an error. We refer to this total search problem as the "refuter problem"[2] corresponding to Haken's lower bound:

[1] This belief is partly supported by the intuition that proving strong circuit lower bounds (e.g., $\mathrm{NP} \not\subseteq \mathrm{P}/_{\mathrm{poly}}$) seems to be a prerequisite for proving strong proof complexity lower bounds (e.g., for Extended Frege) [92]. However, formalizing such connections has proven challenging [1, 84].
[2] This term is adopted from [26], as will be discussed later.

Problem 1.1 (Informal). Given (query access to) a subexponential-size resolution proof Π that claims to be a proof of the pigeonhole principle, find an invalid derivation in Π.

For any proof complexity lower bound of the form "the tautology ϕ requires proof length greater than s in the proof system P", we can define an associated search problem: Given a purported P-proof Π of ϕ with length at most s, find an invalid derivation in Π. With the appropriate formalization (see Section 3), these refuter problems are NP *search problems* and are *total* if and only if their underlying lower bounds hold. Therefore, their computational complexity can be studied using the theory of TFNP [73]. As elaborated in Section 2, the complexity of refuter problems reflects the metamathematical difficulty of proving the corresponding lower bounds, thereby providing a purely computational framework for analyzing the difficulty of proving such lower bounds.

In this paper, we initiate the study of refuter problems in proof complexity and take the first step by studying these problems for *resolution* lower bounds. Resolution serves as a natural first step for exploring the metamathematics of proof lower bounds for two main reasons:

(i) First, resolution is a well-studied proof system, largely due to its fundamental connections to SAT-solving and automated theorem provers [36, 37]. Krajíček even estimates that "there are perhaps more papers published about proof complexity of resolution than about all remaining proof complexity topics combined" [68, Chapter 13].
(ii) Second, significant progress has already been made in proving lower bounds against resolution [8, 11, 28, 47, 99], suggesting that investigating the metamathematics of resolution lower bounds is a promising avenue.

We study several important resolution lower bounds, including those for the pigeonhole principle [8, 47], Tseitin tautologies [95, 99], and random CNF formulas [28]. In our study, we introduce a new syntactic subclass of decision-tree TFNP, denoted as rwPHP(PLS), which can be thought of as a randomized version of PLS, residing slightly above PLS in the TFNP hierarchy[3].

At a high level, we show that resolution *width* lower bounds are captured by PLS, while resolution *size* lower bounds are captured by rwPHP(PLS).

Theorem 1.2 (Main Result; Informal). *The refuter problems corresponding to the*

(1) resolution width lower bounds for the pigeonhole principle and Tseitin tautologies are PLS*-complete;*
(2) resolution size lower bounds for the pigeonhole principle (Problem 1.1), Tseitin tautologies, and random CNF formulas are rwPHP(PLS)*-complete.*

Our results are more comprehensive than those stated in Theorem 1.2, and we defer the full formal presentation to Section 4. Here, we highlight a few key insights:

- **Characterizing a common proof technique:** All the aforementioned resolution size lower bound proofs share a common strategy, which we call "random restrictions + width lower bounds". It is often the case that resolution lower bounds proven using this strategy have refuter problems in rwPHP(PLS)[4]. This implies that "rwPHP(PLS)-reasoning" is sufficient for implementing one of the most commonly employed techniques for proving resolution lower bounds.
- **Minimum reasoning for resolution lower bounds:** Complementing the above, we prove that for *any* family of hard tautologies for resolution, the corresponding refuter problem (for size lower bound) is rwPHP(PLS)-hard. This establishes that our rwPHP(PLS) upper bound is indeed tight. Notably, the hardness proof does not rely on the hard tautology being the pigeonhole principle. Consequently, this result carries an intriguing metamathematical implication: "rwPHP(PLS)-reasoning" is necessary for proving *any* resolution lower bound.
- **Consequences in bounded arithmetic:** Theorem 1.2 can also be interpreted as conservativeness results showing that a certain fragment of (relativized) bounded arithmetic, $\mathcal{T}^{1}_{\mathsf{Res}} := \mathsf{T}^{1}_{2}(\alpha) + \mathrm{dwPHP}(\mathrm{PV}(\alpha))$, "captures" the minimum reasoning required for proving resolution size lower bounds. More precisely, $\mathcal{T}^{1}_{\mathsf{Res}}$ is powerful enough to formalize many resolution lower bounds proved in the literature, including Haken's seminal lower bound for PHP [47], while at the same time, it is necessary for proving *any* resolution lower bound. An interesting takeaway of our results is that resolution *appears p-bounded* (i.e., appears to prove every tautology in polynomial size) to every theory weaker than $\mathcal{T}^{1}_{\mathsf{Res}}$.
  We find the *existence* of such a theory quite insightful. It is natural to speculate that there is a very powerful theory $\mathcal{T}^{1}_{\mathsf{EF}}$ such that strong proof systems like Extended Frege "appears p-bounded" to every theory weaker than $\mathcal{T}^{1}_{\mathsf{EF}}$. If true, this speculation would provide a strong barrier result in proof complexity.
- **Very efficient proofs for resolution lower bounds:** Finally, translating our results into proof complexity, we exhibit surprisingly efficient proofs of resolution lower bounds. For instance, we show that low-width *random resolution* [87] can prove many resolution size lower bounds (including Haken's lower bound).[5] Previously, Cook and Pitassi [33] proved Haken's lower bound in the theory IPV^{ω}, which can be interpreted as intuitionistic reasoning using polynomial-time concepts over the purported resolution proof. Our results suggest that AC^0 concepts (and indeed much weaker ones) already suffice to prove the same lower bound.[6]

2 More Background

We now provide further background on bounded reverse mathematics, refuter problems, and the theory of TFNP to justify our methodology: the *metamathematics* of the proof complexity lower

[3] The formal definition and properties of rwPHP(PLS) are presented in the full version.

[4] An interesting exception is the general size-width tradeoff by Ben-Sasson and Wigderson [11]; see the full version for further discussions.

[5] Although random resolution is not a Cook–Reckhow proof system unless P = NP [87], it is possible to define a fragment of random resolution that is Cook–Reckhow and that proves the aforementioned resolution size lower bounds.

[6] Our formalization is different from [33]. Informally speaking, [33] uses "*polynomial-time* reasoning" to handle a *polynomial-size* proof, while we use "PH-*reasoning*" to handle an *exponentially long* proof.

bounds can — and indeed *should* — be understood through the *computational complexity* of their associated refuter problems within TFNP.

Bounded reverse mathematics. Reverse mathematics explores, for each mathematical theorem of interest, the minimal theory required to prove it. In bounded reverse mathematics [30, 32, 77], the theories considered come from *bounded arithmetic*, which (roughly speaking) are logical theories formalizing the idea of "reasoning within a complexity class C". The link between these logical theories and complexity classes makes bounded arithmetic, and hence bounded reverse mathematics, an effective framework for studying the metamathematics of complexity theory.

Indeed, there has been a long history of studying the (un)provability of lower bounds in the context of bounded arithmetic: In 1989, Krajíček and Pudlák investigated the unprovability of proof lower bounds [61], while Razborov studied the unprovability of circuit lower bounds in 1995 [89, 90]. Notably, many lower bounds for weak circuit classes and proof systems can be formalized in weak theories [33, 75, 89], while some strong lower bounds are unprovable within them [24, 61, 62, 64, 69, 81, 83, 90].

We take a different perspective from the aforementioned line of work: rather than asking whether lower bounds are provable in certain theories, our goal is to *characterize* the exact reasoning power required to prove these lower bounds. That is, we seek to identify the *minimal* theory $\mathcal{T}$ that can prove the given lower bound and to establish the minimality of $\mathcal{T}$ by showing that the axioms used in the proof are indeed *necessary*. The necessity of axioms, i.e., deriving the axiom back from the theorem, is called a *reversal* in reverse mathematics.

Example 2.1. Recently, Chen, Li, and Oliveira [25] presented several notable reversals related to complexity lower bounds. In their work, they establish that variants of weak pigeonhole principles are *necessary and sufficient* for proving various classical lower bounds. For instance, the fact that one-tape Turing machines require $\Omega(n^2)$ time to recognize palindromes [70] can be proved using the *weak pigeonhole principle*; Moreover, [25, Theorem 4.9] demonstrates a reversal, proving that this lower bound is, in fact, *equivalent* to the weak pigeonhole principle. The work in [25] serves as one of the main inspirations of this paper.

Refuter problems. To investigate the metamathematics of a lower bound statement, we first write down the statement in forall-exists form:

- Circuit lower bounds: Let L be a hard language and s be a size lower bound for L. The lower bound statement expresses that *for every* circuit C of size s, *there exists* an input x such that $L(x) \neq C(x)$.
- Proof lower bounds: Let ϕ be a tautology that is hard for some proof system P, and s be a size lower bound for ϕ. The lower bound statement expresses that *for every* purported P-proof Π of size s, *there exists* an invalid derivation step in Π.

In general, a statement

$$\forall x\, \exists y\, V(x, y) \tag{1}$$

would define a total search problem of finding a valid y given x such that $V(x, y)$ holds; note that the statement is *true* if and only if the search problem is *total*. In our case, is a total search problem Problem 1.1 precisely because of the Resolution lower bounds against the pigeonhole principle proved by Haken and others [8, 11, 47].

This correspondence can be formally justified by the *witnessing theorems* in bounded arithmetic. A witnessing theorem for a theory $\mathcal{T}$ links it to a syntactic subclass $\mathsf{C}_{\mathcal{T}}$ of TFNP, and the theorem states that if (1) is provable in $\mathcal{T}$, then the corresponding (total) search problem lies in the class $\mathsf{C}_{\mathcal{T}}$.[7] For instance, Buss's witnessing theorem [15] states that if (1) is provable in S^1_2, then the corresponding total search problem can be solved in polynomial time; Buss and Krajíček [19] showed that if (1) is provable in T^1_2, then the corresponding total search problem is solvable in PLS (polynomial local search).

Search problems corresponding to circuit lower bounds have already been studied in the literature [23, 25, 26, 46, 59, 81] and are termed "refuter problems" in [26]. We adopt this terminology and refer to the search problems associated with proof lower bounds as "refuter problems" as well.[8]

Total search problems in NP. The above discussion suggests that the metamathematics of lower bounds can be understood through the computational complexity of their refuter problems. Since these problems are total search problems in NP (as long as the lower bounds are true), it is natural to adopt the methodology of TFNP while studying their complexity.

What is the "methodology of TFNP"? Since the seminal work of Megiddo and Papadimitriou [73], problems in TFNP have been categorized based on their *proof of totality*. For instance, the class PLS captures NP search problems whose totality is provable from the principle "every DAG has a sink" [54], while the class PPAD captures problems whose totality is provable from "every DAG with an unbalanced node has another one" [78]. Moreover, *completeness* results play the same role as reversals in bounded reverse mathematics. For example, a pivotal result in this direction is the PPAD-completeness of finding a Nash equilibrium in two-player games [27, 35]. This result carries an intriguing metamathematical interpretation: Topological arguments (specifically, Brouwer's fixed point theorem [12]) or methods akin to it are *unavoidable* for proving the existence of Nash equilibrium [76], which stands in stark contrast to the linear programming duality methods used for zero-sum games [100].

The attentive reader may have already noticed that the above methodology shares a close resemblance to (bounded) reverse mathematics. This similarity can indeed be formally justified by the witnessing theorems mentioned earlier. (Another formal justification is that provability in (universal variants of) bounded arithmetic is equivalent to reducibility in TFNP; see, e.g., [74, Proposition 3.4].) While reading this paper, it is useful to remember that all TFNP results established here can be translated into results in bounded arithmetic and vice versa, conveying the same underlying conceptual message.

[7] This requires (1) to be a "$\forall\Sigma^b_1$-sentence", meaning that $|x|$ and $|y|$ are polynomially related and $V(x, y)$ is a deterministic polynomial-time relation.

[8] In fact, [26] called these problems "*refutation* problems". We choose to use "*refuter* problems" to avoid confusion with the term "refutation" in proof complexity, which usually refers to a proof showing that a formula is unsatisfiable.

3 Our Settings

Before explaining our results, we first discuss the setting of (decision tree) TFNP and (relativized) bounded arithmetic in which our results take place. This sub-section is of preliminary nature, but we recommend reading (i.e., not skipping) it before proceeding to our results in Section 4. In particular, this sub-section introduces our formalization of lower bounds, which is different from previous works [33, 75] as the purported resolution proof is represented as an *exponentially-long second-order* object.

We consider TFNP problems in the *decision tree* model ($\mathsf{TFNP}^{\mathrm{dt}}$); this model is sometimes called "type-2 TFNP problems" [6] when the decision trees are uniform. In this model, we are given an input x of length N and we think of *decision trees of* $\mathrm{polylog}(N)$ *depth* as "efficient". Each possible solution o can be represented by $\mathrm{polylog}(N)$ bits, and there is an efficient procedure $\phi(x, o)$ that verifies whether o is a valid solution for x. (That is, given the purported solution o, $\phi(x, o)$ makes only $\mathrm{polylog}(N)$ queries to x.) The goal is, of course, to find a solution o such that $\phi(x, o)$ holds.

$\mathsf{TFNP}^{\mathrm{dt}}$ corresponds to *relativized* bounded arithmetic where a new predicate α is added into the language. The predicate α is intuitively treated as an oracle (or an exponentially-long input). For example, $\mathsf{PV}(\alpha)$ captures reasoning using P^{α}-concepts, i.e., *uniform and efficient decision trees over* α.

Remark 1 (Type-1 vs. Type-2 TFNP Problems). In the literature, it is common to define a type-1 TFNP problem in terms of *succinct encodings* of exponentially large objects. For example, a possible definition of a PLS-complete problem is as follows: Given a "neighborhood" circuit $C : \{0,1\}^n \to \{0,1\}^{\mathrm{poly}(n)}$ and a "potential function" circuit $V : \{0,1\}^n \to \{0,1\}^{\mathrm{poly}(n)}$ that together encode a DAG on 2^n nodes, and also an active node (i.e., a node with non-zero out-degree), find a sink of this graph (i.e., a node with non-zero in-degree and zero out-degree). In contrast, the $\mathsf{TFNP}^{\mathrm{dt}}$ / type-2 TFNP problems that we consider simply treat C and V as oracles.

Any separation of type-2 TFNP problems implies a separation of type-1 TFNP problems *in a relativized world* [6]. For example, $\mathsf{PLS}^{\mathrm{dt}} \not\subseteq \mathsf{PPA}^{\mathrm{dt}}$ implies an oracle O under which $\mathsf{PLS}^{O} \not\subseteq \mathsf{PPA}^{O}$.

3.1 Refuter Problems for Resolution Lower Bounds

This subsection formalizes the refuter problem for resolution lower bounds as a $\mathsf{TFNP}^{\mathrm{dt}}$ problem. We assume familiarity with the resolution proof system. In resolution, every line is a *clause* (i.e., the disjunction of literals) and the only inference rule is the *resolution rule*:

$$\frac{C \vee \ell \qquad D \vee \overline{\ell}}{C \vee D},$$

where C, D are clauses and ℓ is a literal. Sometimes, we will also allow the *weakening* rule that replaces a clause with a consequence of it:

$$\frac{C}{C \vee D}.$$

The *size* of a resolution proof is the number of lines (i.e., clauses) in it. The *width* of a resolution proof is the maximum width of any clause in it, where the *width* of a clause is the number of literals in the clause. Basics about resolution can be found in any textbook on proof complexity, e.g., [68, Section 5].

Size lower bounds for resolution. Let F be a tautology[9] that is *exponentially*-hard for resolution. For example, take F to be the pigeonhole principle which does not have c^n-size resolution proofs for some absolute constant $c > 1$ [47]. The refuter problem, which we denote as

$$\textsc{Refuter}(s(F \vdash_{\mathrm{Res}} \bot) \le c^n),$$

is defined as follows. The input Π is a purported length-c^n resolution proof of F represented as a list of c^n *nodes*, where each node consists of a clause in the resolution proof and the predecessors of this clause. (For example, if the clause in node i is resolved from the clauses in node j and node k, then the predecessor information would contain two integers (j, k).) A *valid solution* would be the index of any node $i \in [c^n]$ whose derivation is illegal: denoting C_i the clause in node i, then there do not exist clauses C, D and a literal ℓ such that

$$C_i = C \vee D, C_j = C \vee \ell, C_k = D \vee \overline{\ell}.$$

By Haken's lower bound mentioned above [47], every purported resolution proof of length c^n must contain an illegal derivation, thus the above problem is *total*. Let $N := c^n \mathrm{poly}(n)$ denote the bit-length of the input resolution proof, then each node can be described in $\mathrm{poly}(n) \le \mathrm{polylog}(N)$ bits, hence there is an efficient decision tree that verifies whether a node i is illegal and the above refuter problem is indeed in $\mathsf{TFNP}^{\mathrm{dt}}$.

We can also formalize resolution lower bounds in relativized bounded arithmetic as follows. We add a new symbol α into our language that encodes a length-c^n resolution proof, i.e., for each $i \in [c^n]$, $\alpha(i)$ is the i-th bit of the proof. Fixing a hard tautology F, let $\mathsf{mistake}_F(n, \alpha, i)$ be a $\mathsf{PV}(\alpha)$ predicate that is true if $i < c^n$ and α, interpreted as a length-c^n resolution proof for F, makes an invalid derivation in the i-th step. Note that this only depends on a constant number of nodes in the proof, and each node is described in $\mathrm{poly}(n)$ bits, hence $\mathsf{mistake}_F(n, \alpha, i)$ is indeed computable in deterministic polynomial time with oracle access to α. The $\forall\Sigma^b_1(\alpha)$-sentence[10]

$$\forall n \in \mathsf{Log}\, \exists i \le c^n\ \mathsf{mistake}_F(n, \alpha, i)$$

expresses the totality of the refuter problem as defined above; the provability of this sentence in relativized bounded arithmetic corresponds to the complexity of the refuter problem in $\mathsf{TFNP}^{\mathrm{dt}}$.[11]

[9] A DNF D is a *tautology* if and only if the corresponding CNF $\neg D$ is a *contradiction*. A *proof* of D being a tautology is a *refutation* of $\neg D$ being a contradiction. For convenience, we will use the terms "tautology/proof" and "contradiction/refutation" interchangeably.

[10] Roughly speaking, a $\forall\Sigma^b_1$-sentence (resp. $\forall\Sigma^b_1(\alpha)$-sentence) is a sentence of the form

$$\forall x\, \exists y\, \varphi(x, y),$$

where $|x|, |y|$ are polynomially related and φ is a polynomial-time relation (resp. polynomial-time relation with α oracle); these sentences naturally express problems in TFNP (resp. $\mathsf{TFNP}^{\mathrm{dt}}$). The notation $n \in \mathsf{Log}$ means that n is the length of some number, thus allowing one to reason about integers of magnitude $2^{\mathrm{poly}(n)}$ and strings of length $\mathrm{poly}(n)$. In our particular case, it allows the length of the purported proof to be exponential in n. These are standard notations in bounded arithmetic.

[11] As a technical detail, we can also allow α to take *parameters* $\vec{z}$ that can be thought of as non-uniformity. That is, for each $i \in [c^n]$, $\alpha(\vec{z}, i)$ is the i-th bit of the proof. We consider the sentence

$$\forall n \in \mathsf{Log}, \vec{z}\ \exists i \le c^n\ \mathsf{mistake}_F(n, \alpha(\vec{z}, \cdot), i)$$

which expresses that the proof encoded by $\alpha(\vec{z}, \cdot)$ is not a valid length-c^n resolution proof for F. The power of many natural principles with and without parameters are very different (see e.g., [48, Section 4.3]).

Width lower bounds for resolution. In this paper, we also study the refuter problems corresponding to *width* lower bounds for resolution. Let F be a tautology without width-w_F resolution proofs, the refuter problem for this width lower bound would be denoted as

$$\textsc{Refuter}(w(F \vdash_{\text{Res}} \bot) \leq w_F).$$

The formalization of width lower bounds is essentially the same as that of size lower bounds, with the only difference that we now impose that every clause in the input resolution proof contains at most w_F literals. This can be done *syntactically* by only allocating w_F literals to each node.

3.2 Retraction Weak Pigeonhole Principles

This paper demonstrates that the complexity of refuter problems corresponding to resolution size lower bounds is tightly linked to the new complexity class rwPHP(PLS). Therefore, we need to introduce this class before describing our results.

Here, "rwPHP" stands for the *retraction weak pigeonhole principle*:

> For any two functions $f : [N] \to [2N]$ and $g : [2N] \to [N]$, the function $f \circ g : [2N] \to [2N]$ cannot be the identity function.

The term "retraction", borrowed from category theory [52], means that the principle concerns a pair of functions f, g where g is a "retraction"; the term "weak" indicates that the domain of g ($[2N]$) is *much* larger than its range ($[N]$). This principle, along with other variants of weak pigeonhole principles, is widely studied in the context of bounded arithmetic [2, 25, 50, 52, 63, 67, 72, 80, 98] and total search problems [56, 58, 59]; it is sometimes also called the "witnessing weak pigeonhole principle (WPHPWIT)" [25, 51] and "Lossy-Code" [59]. Clearly, rwPHP corresponds to a $\mathsf{TFNP}^{\mathsf{dt}}$ problem: given (query access to) two functions $f : [N] \to [2N]$ and $g : [2N] \to [N]$, find an input $y \in [2N]$ such that $f(g(y)) \neq y$.

Let $\mathcal{P}$ be a problem in $\mathsf{TFNP}^{\mathsf{dt}}$, then one can define a class rwPHP($\mathcal{P}$) capturing the retraction weak pigeonhole principle where, informally speaking, the retraction function g can be computed in $\mathcal{P}$. In the decision tree model, the inputs of rwPHP($\mathcal{P}$) consist of:

(1) (the evaluation table of) a function $f : [N] \to [2N]$, and
(2) $2N$ instances of $\mathcal{P}$, denoted as $\{I_y\}_{y \in [2N]}$, where each valid solution *ans* of each I_y is marked with an integer $g_{y,ans} \in [N]$.

The goal is to find an integer $y \in [2N]$ along with a solution *ans* of I_y such that $f(g_{y,ans}) \neq y$. It is not hard to see that if $\mathcal{P} \in \mathsf{TFNP}^{\mathsf{dt}}$ then rwPHP($\mathcal{P}$) $\in \mathsf{TFNP}^{\mathsf{dt}}$. Furthermore, rwPHP($\mathcal{P}$) can be solved by a simple *randomized* algorithm given oracle access to any solver of $\mathcal{P}$.

The class rwPHP(PLS) is defined as the problems reducible to rwPHP($\mathcal{P}$) for a PLS-complete problem $\mathcal{P}$. It can be shown that rwPHP(PLS) does not depend on the exact choice of the PLS-complete problem $\mathcal{P}$.

Witnessing for T^1_2 + dwPHP(PV). Although rwPHP(PLS) seems to be new to the TFNP community, it already appeared implicitly in the literature of bounded arithmetic. This class captures the TFNP problems whose totality is provable in T^1_2 + dwPHP(PV). In other words, rwPHP(PLS) corresponds to the *witnessing theorem* for T^1_2 + dwPHP(PV) (just like how PLS corresponds to a witnessing theorem for T^1_2 [19]). This was noticed in [18] where they showed every $\forall\Sigma^b_1$-consequence of T^1_2 + dwPHP(PV) *randomly* reduces to PLS; in fact, the same argument implies a deterministic reduction to rwPHP(PLS).

Remark 2 (How Strong is rwPHP(PLS)?).

Since rwPHP(PLS) can be seen as a randomized version of PLS (where the guarantee that "most randomness is good" is provided by the dual weak pigeonhole principle), its position in the $\mathsf{TFNP}^{\mathsf{dt}}$ hierarchy is roughly the same as, but slightly higher than PLS. In particular, in the decision tree setting, it follows from the previous separations (PLS $\not\subseteq$ PPP [44] and PLS $\not\subseteq$ PPA [13]) that rwPHP(PLS) is contained in neither PPP nor PPA. Note that there is already a decision tree separation between PLS and the $\mathsf{TFNP}^{\mathsf{dt}}$ problem corresponding to rwPHP (which follows from a resolution width lower bound for rwPHP [87, Proposition 3.4]), hence in the decision tree setting, rwPHP(PLS) strictly contains PLS.

We also note that $\mathsf{T}^1_2(\alpha)$ + dwPHP(PV(α)) is a relatively weak theory in the realm of relativized bounded arithmetic.[a] This theory is a subtheory of both $\mathsf{T}^2_2(\alpha)$ and Jeřábek's (stronger) fragment for approximate counting $\mathsf{APC}_2(\alpha)$ [53]. It is also "weak" in the sense that unconditional unprovability results are known: it cannot prove the ordering principle [5] and the pigeonhole principle [87].

[a] The reader might have encountered claims in the literature that even weaker theories such as S^1_2 or APC_1 are "strong", so it might be confusing for a reader unfamiliar with bounded arithmetic that we are claiming $\mathsf{T}^1_2(\alpha)$ + dwPHP(PV(α)) as a "weak" theory. The reason is *relativization*: In our formalization, the purported resolution proof α has *exponential* size, and we are only allowed to reason about objects in PH (think of AC^0 circuits over α). This is much weaker than the setting where the proof α has *polynomial* size and we are allowed to reason about polynomial-time concepts. This is roughly analogous to classifying the circuit class AC^0 (i.e., relativized PH) as "weak" and $\mathsf{P}/_{\text{poly}}$ as "strong".

4 Our Results

Our main results can be categorized into three parts: (1) bounded reverse mathematics (TFNP characterizations) for (several) resolution width lower bounds; (2) bounded reverse mathematics (TFNP characterizations) for (several) resolution size lower bounds; and (3) further applications in TFNP and proof complexity. We will describe the results related to width lower bounds first in Section 4.1, not only because they serve as prerequisites for the results regarding size lower bounds (discussed in Section 4.2), but also because the techniques therein find additional applications in TFNP and proof complexity (detailed in Section 4.3).

4.1 Refuters for Resolution Width Lower Bounds

The main message in this subsection is that the refuter problems corresponding to resolution width lower bounds are complete for the well-studied class PLS, the first syntactic subclass of TFNP introduced in the literature [54].

We begin with the results related to the pigeonhole principle. The attentive reader may notice a subtle issue when formulating the refuter problem of width lower bound: $\mathrm{PHP}_{(n+1)\to n}$ already contains an axiom with width n, and the width lower bound for proving it is n as well. Thus, the corresponding width refuter problem becomes trivial. To address this, we instead consider the width

refuter problem for a *constant-width analog* of $\mathrm{PHP}_{(n+1)\to n}$, called $\mathrm{EPHP}_{(n+1)\to n}$, which has constant-width axioms and an $n/3$ width lower bound as shown in [11]. We characterize the complexity of its corresponding refuter problem:

Theorem 4.1. $\textsc{Refuter}(w(\mathrm{EPHP} \vdash_{\mathsf{Res}} \bot) < n/3)$ *is* PLS*-complete.*

A similar PLS-completeness result also holds for Tseitin formulas (on expander graphs), where $e(G)$ below is the *expansion* parameter of the graph G.

Theorem 4.2. $\textsc{Refuter}(w(\text{Tseitin} \vdash_{\mathsf{Res}} \bot) < e(G))$ *is* PLS*-complete.*

The techniques used in these results will be further extended to the refuter problems corresponding to black-box TFNP separations, specifically PLS $\nsubseteq$ PPP and PLS $\nsubseteq$ PPA, as described in Theorem 4.10 below.

To tackle Problem 1.1 though, we have to delve into the proofs of the exponential (size) lower bound. A *monotonized* version of the *width* lower bound plays a crucial role in the simplified proof by Beame and Pitassi [8]. In particular, they show that any resolution refutation of $\mathrm{PHP}_{(n+1)\to n}$ contains a clause C with "monotone width" of at least $2n^2/9$ (see full version). We similarly characterize the complexity of its corresponding refuter problem (where the subscript mono denotes the monotone analog of the width refuter problem; the formal definition is provided in the full version):

Theorem 4.3. $\textsc{Refuter}(w_{\mathsf{mono}}(\mathrm{PHP}_{(n+1)\to n} \vdash_{\mathsf{Res}} \bot) < 2n^2/9)$ *is* PLS*-complete.*

This result serves as a key step toward addressing the size refuter problem for the pigeonhole principle, which will be discussed in the next subsection.

The PLS-hardness parts of all three results above stem from a *unified and simple proof*, detailed in the full version. Conversely, the PLS-membership of these refuter problems is established by carefully analyzing the proofs in [8, 11] and demonstrating that "PLS-reasoning" suffices to prove these lower bounds. (In fact, these proofs can be formalized in the theory $\mathsf{T}^1_2(\alpha)$, and the PLS-membership follows directly from the witnessing theorem in [19].)

A non-uniform universal PLS*-membership.* Finally, we establish a *universal* PLS-membership result with respect to *non-uniform* decision tree reductions: for any resolution width lower bound against every unsatisfiable CNF, *as long as the lower bound is correct*, the corresponding refuter problem can be reduced to PLS under *non-uniform* decision tree reductions.

Theorem 4.4 (Informal). *Let $\mathcal{F}$ be any (possibly non-uniform) family of unsatisfiable CNFs with polynomially many clauses, and let w_0 be any valid resolution width lower bound for $\mathcal{F}$. Then there exists a (non-uniform) decision-tree reduction from* $\textsc{Refuter}(w(\mathcal{F} \vdash_{\mathsf{Res}} \bot) < w_0)$ *to* PLS.

Both the formulation and proof of this result inherently require non-uniformity for at least two reasons: (1) it is computationally hard to check whether an arbitrarily given CNF is unsatisfiable, and (2) even assuming that the given CNF is unsatisfiable, it is hard to calculate the resolution width lower bound. See the full version for further discussion.

Remark 3 (Uniform vs. non-uniform reductions). Note that if one only cares about non-uniform reductions, then (the PLS-membership parts of) Theorem 4.1 and Theorem 4.2 are merely special cases of Theorem 4.4. Nevertheless, we believe that the uniform PLS-membership results in Theorem 4.1 and Theorem 4.2 are informative, as they actually show that the corresponding lower bounds can be formalized in $\mathsf{T}^1_2(\alpha)$; in fact, the *code* of the Turing machine that implementing the uniform reduction to PLS effectively acts as a *proof* of the width lower bound using a *local search* argument. They are also crucial for the uniform rwPHP(PLS)-memberships for the size refuter problems. However, the decision tree reduction in the full version seems to require $\exp(n)$ bits of non-uniformity, making it *highly* non-uniform.

On the other hand, the non-uniform reduction in Theorem 4.4 implies an intriguing proof complexity upper bound: *Small-width resolution can prove width lower bounds for resolution itself*! (See Section 4.3 for more details.) Uniformity is not required for this application, allowing us to derive more proof complexity upper bounds using Theorem 4.4: *every* resolution width lower bound *that is correct* can be proved in low-width resolution. (The size lower bound analog of Theorem 4.4 remains unknown, hence we can only show proof complexity upper bounds for tautologies encoding *specific* resolution size lower bounds.)

4.2 Refuters for Resolution Size Lower Bounds

Our main message in this subsection is that the refuter problems corresponding to many resolution size lower bounds are complete for rwPHP(PLS), the TFNP subclass introduced in Section 3.2. Indeed, the theorems presented in this subsection suggest that rwPHP(PLS) captures the complexity of proving *the easiest-to-prove* size lower bounds for resolution. Our workflow is the same as before:

- First, we show that for many notable resolution size lower bounds proven in the literature, the corresponding refuter problems reduce to rwPHP(PLS). Specifically, we identify a common technique for proving resolution size lower bounds, which we call "random restriction + width lower bounds", and demonstrate that if a resolution size lower bound can be proven using it, then the corresponding refuter problem generally falls within rwPHP(PLS).
- Next, we present a *unified* rwPHP(PLS)-hardness result: the refuter problems for resolution size lower bounds are rwPHP(PLS)-hard, and the hardness proof *does not* depend on the hard tautology considered. Thus, we conclude the rwPHP(PLS)-completeness of many refuter problems for resolution size lower bounds.

The rwPHP(PLS)-hardness of size lower bound refuters turns out to be more challenging than the PLS-hardness of width lower bound refuters, as discussed in the full version.

We begin by showing that Problem 1.1 reduces to rwPHP(PLS):

Theorem 4.5 (Informal). *There exists an absolute constant $c > 1$ and an efficient decision-tree reduction from the problem*

$$\textsc{Refuter}(s(\mathrm{PHP}_{(n+1)\to n} \vdash_{\mathsf{Res}} \bot) \le c^n)$$

to rwPHP(PLS).

In fact, we show that $\mathsf{T}^1_2(\alpha)$+dwPHP(PV(α)) proves the sentence

$$\forall n \in \mathsf{Log}\ \exists i \le c^n\ \mathrm{mistake}_{\mathrm{PHP}}(n, \alpha, i),$$

i.e., α is not a length-c^n resolution proof for PHP, by formalizing the classical proofs in [8, 33, 47]; Theorem 4.5 then follows from the witnessing theorem for $\mathsf{T}^1_2(\alpha)$ + dwPHP(PV(α)). In the full version, we present the reduction from the refuter problem $\textsc{Refuter}(s(\text{PHP} \vdash_{\text{Res}} \bot) \le c^n)$ to rwPHP(PLS) directly, without relying on witnessing theorems.

It turns out that a large variety of resolution size lower bounds can be proven using the paradigm of "random restriction + width lower bounds," including those for XOR-lifted formulas [34], Tseitin formulas [95, 99], and random CNFs [28]. We show that all these lower bounds have corresponding refuter problems in rwPHP(PLS) (see the full version). These results provide strong evidence that rwPHP(PLS) (or $\mathsf{T}^1_2(\alpha)$+dwPHP(PV(α))) captures the "complexity" of this popular proof technique for resolution lower bounds.

We complement the above results by showing that for *every* unsatisfiable family of CNFs $\{F_n\}$ that requires resolution size greater than $s_F(n)$, the corresponding refuter problem $\textsc{Refuter}(s(F_n \vdash_{\text{Res}} \bot) \le s_F(n))$ is hard for rwPHP(PLS).

Theorem 4.6 (Informal). *For every unsatisfiable family of CNF formulas $\{F_n\}$ and parameter $s_F(n)$ such that every resolution refutation of F_n requires more than $s_F(n)$ clauses, there exists a decision tree reduction of depth* poly(n) *from* rwPHP(PLS) *to* $\textsc{Refuter}(s(F_n \vdash_{\text{Res}} \bot) \le s_F(n))$.[12]

Note that Theorem 4.6 holds for *every* hard tautology, whereas the rwPHP(PLS) upper bounds such as Theorem 4.5 are only known to hold for some natural families of hard tautologies. For these natural tautologies, we establish a *reversal* in the bounded reverse mathematics of proof complexity lower bounds: The power of "rwPHP(PLS)-reasoning" is *sufficient* for implementing a popular proof strategy that can prove all these resolution lower bounds and, at the same time, is *necessary* for proving *any* resolution lower bound.

Remark 4. We also note that Theorem 4.6 requires decision tree depth poly(n) regardless of s_F, and is thus only considered "efficient" when $s_F = 2^{n^{\Omega(1)}}$. However, this is merely an artifact of our definition of "efficiency" in the decision tree setting, i.e., if the input length is N, then depth-polylog(N) decision trees are considered "efficient". In fact, even if $s_F = 2^{n^{o(1)}}$, each node in the purported length-s_F resolution proof still requires poly(n) bits to represent, so it takes poly(n) query complexity to *verify* a solution of the refuter problem. Therefore, it still makes sense *in the particular setting of refuter problems* to consider a decision tree reduction *efficient* if its query complexity is at most poly(n). We interpret Theorem 4.6 to mean that "rwPHP(PLS)-reasoning" is necessary for proving *not only subexponential but any moderately large* size lower bound for resolution.

The proof of Theorem 4.6 is heavily inspired by the NP-hardness of automating resolution [4] and the exposition of this result in [38]. In these proofs, it was crucial to show that resolution cannot prove lower bounds against itself; in particular, [38, Section 5] showed that resolution requires a large (block-)width to prove resolution lower bounds. Notably, the proof in [38] is by a reduction from rwPHP, i.e., resolution cannot prove lower bounds against itself because resolution cannot prove rwPHP. We strengthen these results by reducing a stronger problem —rwPHP(PLS) instead of rwPHP — to the refuter problems, thereby obtaining a *tight* characterization of these refuter problems.

Finally, our results provide an intriguing characterization of the provably total NP search problems in T^1_2 + dwPHP(PV). That is:

> Just as "every DAG has a sink" characterizes the $\forall\Sigma^b_1$-consequences of T^1_2 [19], "resolution requires $2^{\Omega(n)}$ size to prove PHP" characterizes the $\forall\Sigma^b_1$-consequences of T^1_2 + dwPHP(PV).

4.3 Applications

Besides being interesting in itself, our study of refuter problems also reveals several new insights into these well-studied proof complexity lower bounds and TFNP separations. More specifically, we translate our results into different languages using the generic connection between $\mathsf{TFNP}^{\mathsf{dt}}$ and proof complexity via the *false clause search* problem (see, e.g., [39]): For an unsatisfiable CNF $F = C_1 \wedge \cdots \wedge C_M$, the false clause search problem Search(F) is a $\mathsf{TFNP}^{\mathsf{dt}}$ problem where, given oracle access to an input $x \in \{0,1\}^N$, the goal is to find a clause C_i such that $C_i(x) =$ false. Any $\mathsf{TFNP}^{\mathsf{dt}}$ problem can be written as a false clause search problem for a family of low-width CNFs, and vice versa. In particular, a family of unsatisfiable CNFs has low-width resolution refutations if and only if the corresponding false clause search problem reduces to PLS [90] (see also [55, Section 8.2.2] for an exposition). See the full version for a diagram that summarizes the translations of our main results in different languages.

Proof complexity of proof lower bounds. We first use our results to provide surprisingly efficient proofs for proof complexity lower bounds. Note that a proof complexity lower bound can be expressed by a family of CNFs $\mathcal{F}_{\text{LB}}$ by formulating the corresponding refuter problem as a false clause search problem Search($\mathcal{F}_{\text{LB}}$) (see, e.g., Section 7.1 in our full version).

In particular, there exists a family of $\tilde{O}(w)$-width CNFs that encodes a width-w resolution lower bound. Then, since PLS and low-width resolution are equivalent, Theorem 4.4 implies the following *upper bound* on the resolution width required to prove resolution lower bounds.

Theorem 4.7 (Informal). *Any width-w resolution lower bound can be proved in resolution width $\tilde{O}(w)$.*

We also use our rwPHP(PLS) upper bounds to show that poly(n)-width *random resolution* [18, 87] can prove exponential-size resolution lower bounds (encoded as poly(n)-width CNFs). In fact, using our results on random k-CNFs (see Section 6.4 of the full version), we can show that *most* resolution size lower bounds are provable in low-width random resolution:

Theorem 4.8 (Informal). *With high probability over a random k-CNF F, the resolution size lower bound $s(F \vdash_{\text{Res}} \bot) > 2^{\Omega(n)}$ can be proved in* random resolution *width of* poly(n).

These results stand in stark contrast with Garlík's result [41] that tautologies encoding any resolution size lower bounds are hard for resolution: We show that either switching to width lower bounds or considering *random* resolution makes these lower bound

[12] This theorem requires a mild technical condition that $s_F(n)$ should be moderately larger than the size of the rwPHP(PLS) instance; see the formal statement in the full version for details.

tautologies easy to prove! See Section 7.1 of our full version for details.[13]

Complexity of refuting black-box TFNP *separations.* We also consider the refuter problem for *black-box* TFNP *separations*. Let A, B be two $\mathsf{TFNP}^{\mathsf{dt}}$ classes such that $\mathsf{A} \not\subseteq \mathsf{B}$. Informally, $\mathsf{Ref}(\mathsf{A} \subseteq \mathsf{B})$ is the class of problems reducible to the following kind of "refuter" problems: The input is a purported decision tree reduction from A to B, and the solution is a short witness showing that the reduction is wrong. The refuter problems for $\mathsf{TFNP}^{\mathsf{dt}}$ separations also lie in $\mathsf{TFNP}^{\mathsf{dt}}$, as their totality follows from the correctness of the black-box separation $\mathsf{A} \not\subseteq \mathsf{B}$.

The complexity of such refuter problems measures the strength of the arguments used for black-box separation results. For example, the following corollary conveys a simple but often overlooked fact: *when separating a syntactic* $\mathsf{TFNP}^{\mathsf{dt}}$ *subclass* A *from* B*, it is necessary to incur the totality principle of* B.

Theorem 4.9. *For any two* $\mathsf{TFNP}^{\mathsf{dt}}$ *classes* A, B *such that* $\mathsf{A} \not\subseteq \mathsf{B}$, $\mathsf{B} \subseteq \mathsf{Ref}(\mathsf{A} \subseteq \mathsf{B})$.

Due to the connection between $\mathsf{TFNP}^{\mathsf{dt}}$ and proof complexity, the refuter problem for each black-box TFNP separation naturally aligns with a corresponding refuter problem for a proof complexity lower bound. In particular, we build a uniform reduction from the refuter problems for separations from PLS to the refuter problems for resolution width lower bounds, because showing a $\mathsf{TFNP}^{\mathsf{dt}}$ subclass A is not in PLS is essentially showing a resolution width lower bound for the formula expressing the totality of A.

Note that the false clause search problem for EPHP and Tseitin are in PPP and PPA respectively. Therefore, using our characterization of the resolution width refuter for EPHP (Theorem 4.1) and Tseitin (Theorem 4.2), we conclude that *it is necessary and sufficient to use local search principle to separate* PPP *and* PPA *from* PLS *in the black-box setting.*

Theorem 4.10 (Informal). $\mathsf{Ref}(\mathsf{PPP} \subseteq \mathsf{PLS}) = \mathsf{Ref}(\mathsf{PPA} \subseteq \mathsf{PLS}) = \mathsf{PLS}$.

5 Discussions, Speculations, and Future Directions

This paper initiates a research program that attempts to understand, for every proof system $\mathcal{P}$ of interest, the metamathematics of proving lower bounds against $\mathcal{P}$ through the lens of refuter problems.[14] Our results on resolution suggest that this is a promising direction. There are a plethora of future research directions, both regarding "weak" systems (where we already know strong lower bounds against $\mathcal{P}$-proofs) and "strong" ones (where we are still struggling to prove non-trivial lower bounds against $\mathcal{P}$).

Weak proof systems. It might be feasible to characterize the complexity of refuter problems for weak proof systems. How does the complexity of refuting lower bounds for $\mathcal{P}$ compare with $\mathcal{P}$ itself (or, more precisely, the $\mathsf{TFNP}^{\mathsf{dt}}$ subclass corresponding to $\mathcal{P}$ [14])? In the case that $\mathcal{P}$ is resolution, our work shows that the complexity of refuting *width* lower bounds for $\mathcal{P}$ is exactly $\mathcal{P}$ itself (i.e., PLS), and the complexity of refuting *size* lower bounds is a randomized version of $\mathcal{P}$ (i.e., rwPHP(PLS)). Thus, it seems reasonable to conjecture that for "weak" proof systems, the complexity of proving lower bounds against them is not much higher than themselves.

Moreover, the proof complexity of proof complexity lower bounds is intimately connected to the hardness of automatability of proof systems, see e.g., [4, 10, 38, 42, 45, 49, 79]. We expect that a thorough understanding of the former would help make progress on the latter as well.

Strong proof systems. The situation for strong proof systems seems much more mysterious. For strong proof systems $\mathcal{P}$ (think of $\mathcal{P}$ being Frege or Extended Frege), it is even unclear whether there should be an "easiest-to-prove" lower bound for $\mathcal{P}$ (which would correspond to a syntactic subclass $C(\mathcal{P}) \subseteq \mathsf{TFNP}^{\mathsf{dt}}$ that characterizes the complexity of proving lower bounds for $\mathcal{P}$). Even if such a $C(\mathcal{P})$ exists, it is unclear if it is captured within our current landscape of $\mathsf{TFNP}^{\mathsf{dt}}$.[15]

This suggests the following possibility: The reason that we have not been able to prove lower bounds for $\mathcal{P}$ is that $C(\mathcal{P})$ is a very complicated class, far beyond our current understanding of $\mathsf{TFNP}^{\mathsf{dt}}$ and bounded arithmetic. An even more speculative hypothesis would be that the proof systems $\mathcal{P}$ for which we are able to prove lower bounds are exactly those where $C(\mathcal{P})$ is not "much" higher than $\mathcal{P}$ themselves. We hope that future work will determine to what extent these hypotheses are correct.

The case of $\mathsf{AC}^0[p]$-Frege (where p is a prime) is of particular interest. Although strong lower bounds for $\mathsf{AC}^0[p]$ *circuits* have been known for decades [88, 97], we have not yet succeeded in turning these circuit lower bounds into proof complexity lower bounds against $\mathsf{AC}^0[p]$-Frege (see, e.g., [17, 71]). The paper [16] laid out a research program towards $\mathsf{AC}^0[p]$-Frege lower bounds by studying weaker algebraic proof systems such as the Nullstellensatz [7] and Polynomial Calculus [29, 91]. After a few decades, we have become proficient at proving lower bounds against such algebraic proof systems, but lower bounds against $\mathsf{AC}^0[p]$-Frege remain elusive. Is it because the refuter problems corresponding to $\mathsf{AC}^0[p]$-Frege lower bounds are *fundamentally different* from those for the weaker algebraic proof systems? Does our metamathematical $\mathsf{TFNP}^{\mathsf{dt}}$ perspective bring new insights to this long-standing open question?

6 Further Related Works

Refuter problems for circuit lower bounds. Our study of the refuter problems for proof lower bounds is strongly influenced by the line of work on refuter problems for circuit lower bounds. Chen, Jin,

[13] In our formalization, the lower bound tautologies use *binary encoding*, where (e.g.) the predecessors of every node are encoded by $O(\log N)$ bits. In contrast, Garlík [41] uses *unary encoding* where for every pair of nodes (i, j) (a minor detail is that [41] requires i to be "one level above" j), there is a Boolean variable $x_{i,j}$ indicating whether i is a predecessor of j. As [41] pointed out, a resolution lower bound for the unary-encoded refutation statements implies a similar lower bound for the binary-encoded refutation statements. On the other hand, since we are proving *width* upper bounds and the unary encoding already results in large-width CNFs, we can only afford to use binary encoding.

[14] We believe the similar research program for circuit lower bounds would also be fruitful, which has already started since [23, 25, 59] if not earlier. We limit our discussions to proof lower bounds here.

[15] Note that the question of where $C(\mathcal{P})$ sits in the $\mathsf{TFNP}^{\mathsf{dt}}$ hierarchy is merely a restatement of the open problem of determining the proof complexity of proof complexity lower bounds for $\mathcal{P}$. For example, $C(\mathcal{P})$ is a subclass of PTFNP [43] if and only if Q-EFF (the proof system underlying the definition of PTFNP) can prove lower bounds for $\mathcal{P}$.

Santhanam, and Williams [23] call a lower bound *constructive* if the corresponding refuter problem can be solved in deterministic polynomial time, and they argued that constructivity is a desirable aspect of lower bounds. Chen, Tell, and Williams [26] showed that for many lower bounds against randomized computational models, their refuter problems characterize derandomizing pr-BPP. The main result of Korten [59] can also be seen as the WPHPWIT-hardness of refuter problems for one-tape Turing machine lower bounds. Pich and Santhanam [84] showed how to turn proof complexity lower bounds into circuit lower bounds, assuming the refuter problem for the (conjectured) lower bound $\mathsf{SAT} \notin \mathsf{P}/_{\mathrm{poly}}$ is "provably easy" in a certain sense. Finally, the results of Chen, Li, and Oliveira [25] can be interpreted as the PWPP- and WPHPWIT-completeness of various refuter problems.

It is also worth mentioning that Ebtehaj [40] studied the refuter problems for $\mathcal{A} \not\subseteq \mathsf{BPP}$ for each (type-1) subclass $\mathcal{A} \subseteq \mathsf{TFNP}$ that is indeed hard. However, [40] did not obtain any completeness results for such refuter problems.

Unprovability of complexity upper bounds. In parallel to the investigation of unprovability of complexity lower bounds, there is another line of work showing the unprovability of complexity *upper* bounds in fragments of bounded arithmetic [3, 20–22, 31, 65]. For example, Krajíček and Oliveira [65] proved that Cook's theory PV cannot prove $\mathsf{P} \subseteq \mathsf{SIZE}[n^k]$, and Atserias, Buss, and Müller [3] proved that the theory V^0_2 cannot prove $\mathsf{NEXP} \subseteq \mathsf{P}/_{\mathrm{poly}}$. These results are equivalent to the *consistency* of lower bounds with fragments of bounded arithmetic, thus in some sense representing progress towards proving circuit lower bounds.[16] Indeed, [22] presented a general framework for showing such consistency results by proving lower bounds against circuits with a certain uniformity condition called "LEARN-uniformity", and the techniques employed in many of these papers are inspired by uniform circuit lower bounds such as [94].

Witnessing theorems. TFNP and bounded arithmetic are connected through *witnessing theorems*: each theory is associated with the class of TFNP problems whose totality is provable in this theory. Perhaps the best-known witnessing theorem is Buss's one [15]: every NP search problem provably total in S^1_2 can be solved in deterministic polynomial time. The class PLS and its generalizations such as CPLS capture the NP search problems provably total in higher levels of bounded arithmetic hierarchy [19, 66, 86, 96]; in this sense, witnessing theorems also provide a systematic method for defining new syntactic subclasses of TFNP. Other witnessing theorems considered in the literature include [9, 57, 60]. Our paper contributes to this line of research by characterizing the class of NP search problems provably total in $\mathsf{T}^1_2 + \mathrm{dwPHP}(\mathsf{PV})$ by the refuter problems corresponding to many resolution lower bounds, in particular the problem $\textsc{Refuter}(s(\mathrm{PHP}_{(n+1)\to n} \vdash_{\mathsf{Res}} \bot) < c^n)$.

[16] The "conventional wisdom" seems to believe that the complexity lower bounds are true (for discussions, see https://rjlipton.com/conventional-wisdom-and-pnp/, accessed Mar 14, 2026). Hence, unprovability of complexity lower bounds can be seen as the difficulty for proving this "conventional wisdom", while unprovability of complexity upper bounds represents progress towards proving it. One should keep in mind that the opposite opinion makes equal sense: for a believer of complexity *upper* bounds, the unprovability of these upper bounds indicates the difficulty of confirming their belief, while the unprovability of lower bounds implies progress towards it!

Comparison with the consistency search problem. We note that the refuter problem looks superficially similar to WrongProof, the *consistency search* problem for proof systems [9, 43, 85]. Let $\mathcal{P}$ be a proof system, $\textsc{WrongProof}(\mathcal{P})$ is the $\mathsf{TFNP}^{\mathrm{dt}}$ problem that given as input a purported $\mathcal{P}$-proof Π of *an incorrect statement*, asks for the location of an invalid derivation in Π.

Although both WrongProof and our refuter problems take a purported proof as input and ask for an invalid derivation in the proof, we think that these two problems are fundamentally different, because they have different *reasons of totality*. Roughly speaking, the totality of WrongProof is proved by the *soundness* of $\mathcal{P}$, and the totality of Refuter is guaranteed by *lower bound proofs*. We elaborate on this in the full version.

Another (superficial) similarity between these two problems is that both problems are used to characterize the provably total NP problems in bounded arithmetic. The consistency search problems for Frege and Extended Frege characterize the $\forall\Sigma^b_1$-consequences of U^1_2 and V^1_2 respectively [9], while in this paper we show that the refuter problem for resolution (with a suitable hard tautology) characterizes the $\forall\Sigma^b_1$-consequences of $\mathsf{T}^1_2 + \mathrm{dwPHP}(\mathsf{PV})$.

Acknowledgments

Jiawei thanks Igor C. Oliveira for introducing him to refuters and their connections with TFNP and thanks Robert Robere and Noah Fleming for knowledge of proof complexity.

Yuhao thanks Toniann Pitassi for the knowledge and guidance in the field of proof complexity and thanks Robert Robere for introducing him to beneficial intuition about proof systems and $\mathsf{TFNP}^{\mathrm{dt}}$ classes.

Hanlin thanks Svyatoslav Gryaznov and Iddo Tzameret for helpful discussions regarding [38] and proof complexity in general, and Rahul Santhanam and Ján Pich for beneficial conversations.

We thank Lijie Chen, Jiatu Li, and Igor C. Oliveira for sending us a preliminary version of [25]. We thank Michal Garlík for helpful discussions on [41]. We thank Ján Pich and anonymous referees for their helpful suggestions that improve the presentation of this paper.

References

[1] Noel Arteche, Erfan Khaniki, Ján Pich, and Rahul Santhanam. 2024. From Proof Complexity to Circuit Complexity via Interactive Protocols. In *ICALP (LIPIcs, Vol. 297)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 12:1–12:20. doi:10.4230/LIPICS.ICALP.2024.12

[2] Albert Atserias. 2003. Improved bounds on the Weak Pigeonhole Principle and infinitely many primes from weaker axioms. *Theor. Comput. Sci.* 295 (2003), 27–39. doi:10.1016/S0304-3975(02)00394-8

[3] Albert Atserias, Sam Buss, and Moritz Müller. 2023. On the Consistency of Circuit Lower Bounds for Non-deterministic Time. In *STOC*. ACM, 1257–1270. doi:10.1145/3564246.3585253

[4] Albert Atserias and Moritz Müller. 2020. Automating Resolution is NP-Hard. *J. ACM* 67, 5 (2020), 31:1–31:17. doi:10.1145/3409472

[5] Albert Atserias and Neil Thapen. 2014. The Ordering Principle in a Fragment of Approximate Counting. *ACM Trans. Comput. Log.* 15, 4 (2014), 29:1–29:11. doi:10.1145/2629555

[6] Paul Beame, Stephen A. Cook, Jeff Edmonds, Russell Impagliazzo, and Toniann Pitassi. 1998. The Relative Complexity of NP Search Problems. *J. Comput. Syst. Sci.* 57, 1 (1998), 3–19. doi:10.1006/JCSS.1998.1575

[7] Paul Beame, Russell Impagliazzo, Jan Krajíček, Toniann Pitassi, and Pavel Pudlák. 1994. Lower Bound on Hilbert's Nullstellensatz and propositional proofs. In *FOCS*. IEEE Computer Society, 794–806. doi:10.1109/SFCS.1994.365714

[8] Paul Beame and Toniann Pitassi. 1996. Simplified and improved resolution lower bounds. In *FOCS*. IEEE, 274–282. doi:10.1109/SFCS.1996.548486

[9] Arnold Beckmann and Sam Buss. 2017. The NP Search Problems of Frege and Extended Frege Proofs. *ACM Trans. Comput. Log.* 18, 2 (2017), 11:1–11:19. doi:10.1145/3060145

[10] Zoë Bell. 2020. Automating Regular or Ordered Resolution is NP-Hard. *Electron. Colloquium Comput. Complex.* TR20-105 (2020). https://eccc.weizmann.ac.il/report/2020/105

[11] Eli Ben-Sasson and Avi Wigderson. 2001. Short proofs are narrow - resolution made simple. *J. ACM* 48, 2 (2001), 149–169. doi:10.1145/375827.375835

[12] L. E. J. Brouwer. 1911. Über Abbildung von Mannigfaltigkeiten. *Math. Ann.* 71 (1911), 97–115. Issue 1. doi:10.1007/BF01456931 In German.

[13] Josh Buresh-Oppenheim and Tsuyoshi Morioka. 2004. Relativized NP Search Problems and Propositional Proof Systems. In *CCC*. IEEE Computer Society, 54–67. doi:10.1109/CCC.2004.1313795

[14] Sam Buss, Noah Fleming, and Russell Impagliazzo. 2023. TFNP Characterizations of Proof Systems and Monotone Circuits. In *ITCS (LIPIcs, Vol. 251)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 30:1–30:40. doi:10.4230/LIPIcs.ITCS.2023.30

[15] Samuel R. Buss. 1985. *Bounded arithmetic*. Princeton University.

[16] Samuel R. Buss, Russell Impagliazzo, Jan Krajíček, Pavel Pudlák, Alexander A. Razborov, and Jirí Sgall. 1997. Proof Complexity in Algebraic Systems and Bounded Depth Frege Systems with Modular Counting. *Comput. Complex.* 6, 3 (1997), 256–298. doi:10.1007/BF01294258

[17] Samuel R. Buss, Leszek Aleksander Kołodziejczyk, and Konrad Zdanowski. 2015. Collapsing modular counting in bounded arithmetic and constant depth propositional proofs. *Trans. Amer. Math. Soc.* 367, 11 (2015), 7517–7563. doi:10.1090/S0002-9947-2015-06233-3

[18] Samuel R. Buss, Leszek Aleksander Kołodziejczyk, and Neil Thapen. 2014. Fragments of Approximate Counting. *J. Symb. Log.* 79, 2 (2014), 496–525. doi:10.1017/JSL.2013.37

[19] Samuel R. Buss and Jan Krajíček. 1994. An Application of Boolean Complexity to Separation Problems in Bounded Arithmetic. *Proceedings of the London Mathematical Society* s3-69, 1 (1994), 1–21. doi:10.1112/plms/s3-69.1.1

[20] Jan Bydzovsky, Jan Krajícek, and Igor C. Oliveira. 2020. Consistency of circuit lower bounds with bounded theories. *Log. Methods Comput. Sci.* 16, 2 (2020). doi:10.23638/LMCS-16(2:12)2020

[21] Jan Bydzovsky and Moritz Müller. 2020. Polynomial time ultrapowers and the consistency of circuit lower bounds. *Arch. Math. Log.* 59, 1-2 (2020), 127–147. doi:10.1007/S00153-019-00681-Y

[22] Marco Carmosino, Valentine Kabanets, Antonina Kolokolova, and Igor C. Oliveira. 2021. LEARN-Uniform Circuit Lower Bounds and Provability in Bounded Arithmetic. In *FOCS*. IEEE, 770–780. doi:10.1109/FOCS52979.2021.00080

[23] Lijie Chen, Ce Jin, Rahul Santhanam, and Ryan Williams. 2024. Constructive Separations and Their Consequences. *TheoretiCS* volume 3 (Feb. 2024). doi:10.46298/theoretics.24.3

[24] Lijie Chen, Jiatu Li, and Igor Carboni Oliveira. 2024. On the Unprovability of Circuit Size Bounds in Intuitionistic S_2^1. *Electron. Colloquium Comput. Complex.* (2024), TR24–083. https://eccc.weizmann.ac.il/report/2024/083

[25] Lijie Chen, Jiatu Li, and Igor C. Oliveira. 2024. Reverse Mathematics of Complexity Lower Bounds. In *FOCS*. IEEE, 505–527. doi:10.1109/FOCS61266.2024.00040

[26] Lijie Chen, Roei Tell, and Ryan Williams. 2023. Derandomization vs Refutation: A Unified Framework for Characterizing Derandomization. In *FOCS*. IEEE, 1008–1047. doi:10.1109/FOCS57990.2023.00062

[27] Xi Chen, Xiaotie Deng, and Shang-Hua Teng. 2009. Settling the complexity of computing two-player Nash equilibria. *J. ACM* 56, 3 (2009), 1–57. doi:10.1145/1516512.1516516

[28] Vasek Chvátal and Endre Szemerédi. 1988. Many Hard Examples for Resolution. *J. ACM* 35, 4 (1988), 759–768. doi:10.1145/48014.48016

[29] Matthew Clegg, Jeffery Edmonds, and Russell Impagliazzo. 1996. Using the Groebner basis algorithm to find proofs of unsatisfiability. In *STOC*. 174–183. doi:10.1145/237814.237860

[30] Stephen A. Cook. 2007. Bounded Reverse Mathematics. Plenary lecture for CiE 2007.

[31] Stephen A. Cook and Jan Krajíček. 2007. Consequences of the provability of NP $\subseteq$ $P/_{poly}$. *J. Symb. Log.* 72, 4 (2007), 1353–1371. doi:10.2178/JSL/1203350791

[32] Stephen A. Cook and Phuong Nguyen. 2010. *Logical Foundations of Proof Complexity*. Vol. 11. Cambridge University Press. doi:10.1017/CBO9780511676277

[33] Stephen A. Cook and Toniann Pitassi. 1990. A Feasibly Constructive Lower Bound for Resolution Proofs. *Inf. Process. Lett.* 34, 2 (1990), 81–85. doi:10.1016/0020-0190(90)90141-J

[34] Stefan S. Dantchev and Søren Riis. 2003. On Relativisation and Complexity Gap for Resolution-Based Proof Systems. In *CSL (Lecture Notes in Computer Science, Vol. 2803)*. Springer, 142–154. doi:10.1007/978-3-540-45220-1_14

[35] Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. 2009. The Complexity of Computing a Nash Equilibrium. *SIAM J. Comput.* 39, 1 (2009), 195–259. doi:10.1137/070699652

[36] Martin Davis, George Logemann, and Donald W. Loveland. 1962. A machine program for theorem-proving. *Commun. ACM* 5, 7 (1962), 394–397. doi:10.1145/368273.368557

[37] Martin Davis and Hilary Putnam. 1960. A Computing Procedure for Quantification Theory. *J. ACM* 7, 3 (1960), 201–215. doi:10.1145/321033.321034

[38] Susanna F. de Rezende, Mika Göös, Jakob Nordström, Toniann Pitassi, Robert Robere, and Dmitry Sokolov. 2021. Automating algebraic proof systems is NP-hard. In *STOC*. ACM, 209–222. doi:10.1145/3406325.3451080

[39] Susanna F. de Rezende, Mika Göös, and Robert Robere. 2022. Proofs, Circuits, and Communication. *SIGACT News* 53, 1 (2022), 59–82. doi:10.1145/3532737.3532746

[40] Mohammad Hossein Ebtehaj. 2023. *Variants of Pseudo-deterministic Algorithms and Duality in* TFNP. Master's thesis. University of Waterloo. http://hdl.handle.net/10012/19721

[41] Michal Garlík. 2019. Resolution Lower Bounds for Refutation Statements. In *MFCS (LIPIcs, Vol. 138)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 37:1–37:13. doi:10.4230/LIPICS.MFCS.2019.37

[42] Michal Garlík. 2024. Failure of Feasible Disjunction Property for k-DNF Resolution and NP-Hardness of Automating It. In *CCC (LIPIcs, Vol. 300)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 33:1–33:23. doi:10.4230/LIPICS.CCC.2024.33

[43] Paul W. Goldberg and Christos H. Papadimitriou. 2018. Towards a unified complexity theory of total functions. *J. Comput. Syst. Sci.* 94 (2018), 167–192. doi:10.1016/J.JCSS.2017.12.003

[44] Mika Göös, Alexandros Hollender, Siddhartha Jain, Gilbert Maystre, William Pires, Robert Robere, and Ran Tao. 2022. Separations in Proof Complexity and TFNP. In *FOCS*. IEEE, 1150–1161. doi:10.1109/FOCS54457.2022.00111

[45] Mika Göös, Sajin Koroth, Ian Mertz, and Toniann Pitassi. 2020. Automating cutting planes is NP-hard. In *STOC*. ACM, 68–77. doi:10.1145/3357713.3384248

[46] Dan Gutfreund, Ronen Shaltiel, and Amnon Ta-Shma. 2007. If NP Languages are Hard on the Worst-Case, Then it is Easy to Find Their Hard Instances. *Comput. Complex.* 16, 4 (2007), 412–441. doi:10.1007/S00037-007-0235-8

[47] Armin Haken. 1985. The Intractability of Resolution. *Theor. Comput. Sci.* 39 (1985), 297–308. doi:10.1016/0304-3975(85)90144-6

[48] Rahul Ilango, Jiatu Li, and R. Ryan Williams. 2023. Indistinguishability Obfuscation, Range Avoidance, and Bounded Arithmetic. In *STOC*. ACM, 1076–1089. doi:10.1145/3564246.3585187

[49] Dmitry Itsykson and Artur Riazanov. 2022. Automating OBDD proofs is NP-hard. In *MFCS (LIPIcs, Vol. 241)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 59:1–59:15. doi:10.4230/LIPICS.MFCS.2022.59

[50] Emil Jeřábek. 2004. Dual weak pigeonhole principle, Boolean complexity, and derandomization. *Ann. Pure Appl. Log.* 129, 1-3 (2004), 1–37. doi:10.1016/j.apal.2003.12.003

[51] Emil Jeřábek. 2007. Approximate counting in bounded arithmetic. *J. Symb. Log.* 72, 3 (2007), 959–993. doi:10.2178/JSL/1191333850

[52] Emil Jeřábek. 2007. On Independence of Variants of the Weak Pigeonhole Principle. *J. Log. Comput.* 17, 3 (2007), 587–604. doi:10.1093/LOGCOM/EXM017

[53] Emil Jeřábek. 2009. Approximate counting by hashing in bounded arithmetic. *J. Symb. Log.* 74, 3 (2009), 829–860. doi:10.2178/JSL/1245158087

[54] David S. Johnson, Christos H. Papadimitriou, and Mihalis Yannakakis. 1988. How Easy is Local Search? *J. Comput. Syst. Sci.* 37, 1 (1988), 79–100. doi:10.1016/0022-0000(88)90046-3

[55] Pritish Kamath. 2019. *Some hardness escalation results in computational complexity theory*. Ph. D. Dissertation. Massachusetts Institute of Technology. https://hdl.handle.net/1721.1/128290

[56] Robert Kleinberg, Oliver Korten, Daniel Mitropolsky, and Christos H. Papadimitriou. 2021. Total functions in the polynomial hierarchy. In *ITCS (LIPIcs, Vol. 185)*. 44:1–44:18. doi:10.4230/LIPIcs.ITCS.2021.44

[57] Leszek Aleksander Kołodziejczyk, Phuong Nguyen, and Neil Thapen. 2011. The provably total NP search problems of weak second order bounded arithmetic. *Ann. Pure Appl. Log.* 162, 6 (2011), 419–446. doi:10.1016/J.APAL.2010.12.002

[58] Oliver Korten. 2021. The Hardest Explicit Construction. In *FOCS*. IEEE, 433–444. doi:10.1109/FOCS52979.2021.00051

[59] Oliver Korten. 2022. Derandomization from Time-Space Tradeoffs. In *CCC (LIPIcs, Vol. 234)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 37:1–37:26. doi:10.4230/LIPIcs.CCC.2022.37

[60] Leszek Aleksander Kołodziejczyk and Neil Thapen. 2022. Approximate counting and NP search problems. *J. Math. Log.* 22, 3 (2022), 2250012:1–2250012:31. doi:10.1142/S021906132250012X

[61] Jan Krajíček and Pavel Pudlák. 1989. Propositional Provability and Models of Weak Arithmetic. In *CSL (Lecture Notes in Computer Science, Vol. 440)*. Springer, 193–210. doi:10.1007/3-540-52753-2_40

[62] Jan Krajíček. 1997. Interpolation Theorems, Lower Bounds for Proof Systems, and Independence Results for Bounded Arithmetic. *J. Symb. Log.* 62, 2 (1997), 457–486. doi:10.2307/2275541

[63] Jan Krajíček. 2004. Dual weak pigeonhole principle, pseudo-surjective functions, and provability of circuit lower bounds. *J. Symb. Log.* 69, 1 (2004), 265–286. doi:10.2178/jsl/1080938841

[64] Jan Krajíček. 2011. On the Proof Complexity of the Nisan-Wigderson Generator based on a Hard NP $\cap$ coNP function. *J. Math. Log.* 11, 1 (2011). doi:10.1142/S0219061311000979

[65] Jan Krajíček and Igor C. Oliveira. 2017. Unprovability of circuit upper bounds in Cook's theory PV. *Log. Methods Comput. Sci.* 13, 1 (2017). doi:10.23638/LMCS-13(1:4)2017

[66] Jan Krajíček, Alan Skelley, and Neil Thapen. 2007. NP search problems in low fragments of bounded arithmetic. *J. Symb. Log.* 72, 2 (2007), 649–672. doi:10.2178/JSL/1185803628

[67] Jan Krajíček. 2001. On the weak pigeonhole principle. *Fundamenta Mathematicae* 170, 1-2 (2001), 123–140. http://eudml.org/doc/282141

[68] Jan Krajíček. 2019. *Proof Complexity*. Cambridge University Press. doi:10.1017/9781108242066

[69] Jiatu Li and Igor C. Oliveira. 2023. Unprovability of Strong Complexity Lower Bounds in Bounded Arithmetic. In *STOC*. ACM, 1051–1057. doi:10.1145/3564246.3585144

[70] Wolfgang Maass. 1984. Quadratic Lower Bounds for Deterministic and Nondeterministic One-Tape Turing Machines (Extended Abstract). In *STOC*. ACM, 401–408. doi:10.1145/800057.808706

[71] Alexis Maciel and Toniann Pitassi. 1996. Towards lower bounds for bounded-depth Frege proofs with modular connectives. In *Proof Complexity and Feasible Arithmetics (DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 39)*. DIMACS/AMS, 195–227. doi:10.1090/DIMACS/039/12

[72] Alexis Maciel, Toniann Pitassi, and Alan R. Woods. 2002. A New Proof of the Weak Pigeonhole Principle. *J. Comput. Syst. Sci.* 64, 4 (2002), 843–872. doi:10.1006/JCSS.2002.1830

[73] Nimrod Megiddo and Christos H. Papadimitriou. 1991. On Total Functions, Existence Theorems and Computational Complexity. *Theor. Comput. Sci.* 81, 2 (1991), 317–324. doi:10.1016/0304-3975(91)90200-L

[74] Moritz Müller. 2021. Typical forcings, NP search problems and an extension of a theorem of Riis. *Ann. Pure Appl. Log.* 172, 4 (2021), 102930. doi:10.1016/J.APAL.2020.102930

[75] Moritz Müller and Ján Pich. 2020. Feasibly constructive proofs of succinct weak circuit lower bounds. *Ann. Pure Appl. Log.* 171, 2 (2020). doi:10.1016/j.apal.2019.102735

[76] John Nash. 1951. Non-cooperative games. *Annals of mathematics* 54, 2 (1951), 286–295. doi:10.2307/1969529

[77] Phuong Nguyen. 2008. *Bounded reverse mathematics*. Ph. D. Dissertation. University of Toronto.

[78] Christos H. Papadimitriou. 1994. On the Complexity of the Parity Argument and Other Inefficient Proofs of Existence. *J. Comput. Syst. Sci.* 48, 3 (1994), 498–532. doi:10.1016/S0022-0000(05)80063-7

[79] Theodoros Papamakarios. 2024. Depth-d Frege Systems Are Not Automatable Unless P = NP. In *CCC (LIPIcs, Vol. 300)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 22:1–22:17. doi:10.4230/LIPICS.CCC.2024.22

[80] Jeff B. Paris, A. J. Wilkie, and Alan R. Woods. 1988. Provability of the Pigeonhole Principle and the Existence of Infinitely Many Primes. *J. Symb. Log.* 53, 4 (1988), 1235–1244. doi:10.1017/S0022481200028061

[81] Ján Pich. 2015. Circuit lower bounds in bounded arithmetics. *Ann. Pure Appl. Log.* 166, 1 (2015), 29–45. doi:10.1016/J.APAL.2014.08.004

[82] Ján Pich and Rahul Santhanam. 2019. Why are Proof Complexity Lower Bounds Hard?. In *FOCS*. IEEE Computer Society, 1305–1324. doi:10.1109/FOCS.2019.00080

[83] Ján Pich and Rahul Santhanam. 2021. Strong co-nondeterministic lower bounds for NP cannot be proved feasibly. In *STOC*. ACM, 223–233. doi:10.1145/3406325.3451117

[84] Ján Pich and Rahul Santhanam. 2023. Towards P $\neq$ NP from Extended Frege Lower Bounds. *arXiv math.LO* (2023). doi:10.48550/arXiv.2312.08163

[85] Pavel Pudlák. 2020. Reflection principles, propositional proof systems, and theories. *arXiv* abs/2007.14835 (2020). doi:10.48550/arXiv.2007.14835

[86] Pavel Pudlák and Neil Thapen. 2012. Alternating minima and maxima, Nash equilibria and Bounded Arithmetic. *Ann. Pure Appl. Log.* 163, 5 (2012), 604–614. doi:10.1016/J.APAL.2011.06.014

[87] Pavel Pudlák and Neil Thapen. 2019. Random resolution refutations. *Comput. Complex.* 28, 2 (2019), 185–239. doi:10.1007/S00037-019-00182-7

[88] Alexander A. Razborov. 1987. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical Notes of the Academy of Sciences of the USSR* 41, 4 (1987), 333–338.

[89] Alexander A. Razborov. 1995. Bounded Arithmetic and Lower Bounds in Boolean Complexity. In *Feasible Mathematics II*. Birkhäuser Boston, 344–386. doi:10.1007/978-1-4612-2566-9_12

[90] Alexander A. Razborov. 1995. Unprovability of lower bounds on circuit size in certain fragments of bounded arithmetic. *Izvestiya: mathematics* 59, 1 (1995), 205. doi:10.1070/IM1995v059n01ABEH000009

[91] Alexander A. Razborov. 1998. Lower Bounds for the Polynomial Calculus. *Comput. Complex.* 7, 4 (1998), 291–324. doi:10.1007/S000370050013

[92] Alexander A. Razborov. 2015. Pseudorandom generators hard for k-DNF Resolution and Polynomial Calculus Resolution. *Annals of Mathematics* 181, 2 (2015), 415–472. doi:10.4007/annals.2015.181.2.1

[93] Rahul Santhanam and Iddo Tzameret. 2021. Iterated lower bound formulas: a diagonalization-based approach to proof complexity. In *STOC*. ACM, 234–247. doi:10.1145/3406325.3451010

[94] Rahul Santhanam and Ryan Williams. 2014. On Uniformity and Circuit Lower Bounds. *Comput. Complex.* 23, 2 (2014), 177–205. doi:10.1007/S00037-014-0087-Y

[95] Uwe Schöning. 1997. Resolution Proofs, Exponential Bounds, and Kolmogorov Complexity. In *MFCS (Lecture Notes in Computer Science, Vol. 1295)*. Springer, 110–116. doi:10.1007/BFB0029954

[96] Alan Skelley and Neil Thapen. 2011. The provably total search problems of bounded arithmetic. *Proceedings of the London Mathematical Society* 103, 1 (2011), 106–138. doi:10.1112/plms/pdq044

[97] Roman Smolensky. 1987. Algebraic Methods in the Theory of Lower Bounds for Boolean Circuit Complexity. In *STOC*. ACM, 77–82. doi:10.1145/28395.28404

[98] Neil Thapen. 2002. *The weak pigeonhole principle in models of bounded arithmetic*. Ph. D. Dissertation. University of Oxford.

[99] Alasdair Urquhart. 1987. Hard examples for resolution. *J. ACM* 34, 1 (1987), 209–219. doi:10.1145/7531.8928

[100] Bernhard Von Stengel. 2023. Zero-sum games and linear programming duality. *Mathematics of Operations Research* (2023). doi:10.1287/moor.2022.0149