Ramsey's theorem for pairs, collection, and proof size

Leszek Aleksander Kołodziejczyk
Institute of Mathematics
University of Warsaw
Banacha 2
02-097 Warszawa, Poland
E-mail: lak@mimuw.edu.pl

Tin Lok Wong
Department of Mathematics
National University of Singapore
10 Lower Kent Ridge Road
Singapore 119076
E-mail: matwong@nus.edu.sg

Keita Yokoyama*
School of Information Science
Japan Advanced Institute of Science and Technology
1-1 Asahidai
Nomi, Ishikawa 923-1292, Japan
E-mail: y-keita@jaist.ac.jp

Abstract

We prove that any proof of a $\forall \Sigma_2^0$ sentence in the theory $\mathrm{WKL}_0 + \mathrm{RT}_2^2$ can be translated into a proof in RCA_0 at the cost of a polynomial increase in size. In fact, the proof in RCA_0 can be obtained by a polynomial-time algorithm. On the other hand, RT_2^2 has non-elementary speedup over the weaker base theory RCA_0^* for proofs of Σ_1 sentences.

We also show that for $n \geqslant 0$, proofs of Π_{n+2} sentences in $\mathrm{B}\Sigma_{n+1} + \exp$ can be translated into proofs in $\mathrm{I}\Sigma_n + \exp$ at a polynomial cost in size. Moreover, the Π_{n+2}-conservativity of $\mathrm{B}\Sigma_{n+1} + \exp$ over $\mathrm{I}\Sigma_n + \exp$ can be proved in PV, a fragment of bounded arithmetic corresponding to polynomial-time computation. For $n \geqslant 1$, this answers a question of Clote, Hájek, and Paris.

Keywords Ramsey's theorem, proof size, proof speedup, forcing interpretation, α-large sets, proof theory, reverse mathematics

MSC classes Primary: 03F20, 03B30, 03F35 Secondary: 05D10, 03F30, 03F25, 03H15

---

*corresponding author

The logical strength of Ramsey's theorem for pairs and two colours, formalized as the second-order arithmetic statement RT^2_2, has been a major topic of interest in reverse mathematics and related areas of logic for over 25 years [24, 4, 5, 6]. Recently, Patey and the third author [22] showed that, as far as proving relatively simple statements is concerned, RT^2_2 is no stronger than the usual base theory considered in reverse mathematics, RCA_0, axiomatized by the comprehension principle for computable properties of numbers and by Σ^0_1 induction. More precisely, the result shown in [22] is that RT^2_2 is $\forall\Sigma^0_2$-$conservative$ over RCA_0: for any sentence γ that consists of a block of universal first- and second-order quantifiers followed by a Σ^0_2 formula, if $\mathrm{RCA}_0 + \mathrm{RT}^2_2$ proves γ, then γ already has a proof in RCA_0. In particular, since any statement expressing the well-foundedness of a given computable ordering is $\forall\Sigma^0_2$, the theories $\mathrm{RCA}_0 + \mathrm{RT}^2_2$ and RCA_0 have the same proof-theoretic strength as measured in terms of provability of the well-foundedness of ordinal notation systems.

This new insight into the strength of RT^2_2 naturally leads to some further questions. One of these is whether the conservation result can be extended to a complete characterization of the Π^1_1 sentences provable from RT^2_2. For example, it is open whether the Π^1_1 consequences of $\mathrm{RCA}_0 + \mathrm{RT}^2_2$ are axiomatized by the Σ^0_2 collection principle (which itself witnesses that RT^2_2 is not Π^1_1-conservative over RCA_0). Here, we consider a different problem, already raised as Question 9.5 in [22]: does RT^2_2 have significant proof speedup over RCA_0 with respect to $\forall\Sigma^0_2$ sentences? A positive answer would indicate, among other things, that the conservation result might have rather limited relevance in practice.

The study of proof size in axiomatic theories and of proof speedup, a topic going back to Gödel [13], was given an excellent (if no longer fully up-to-date) survey by Pudlák [23]. An important phenomenon that has been observed empirically is that, from a quantitative perspective, prominent cases in which an arithmetic theory T is conservative over a theory S for sentences in some class Γ typically fit one of two patterns: either T has iterated exponential ("tower function") speedup over S on proofs of sentences from Γ, or each proof of such a sentence in T can be translated into S with at most polynomial blowup. The former behaviour is illustrated for instance by the conservativity of arithmetical comprehension ACA_0 over PA and the Π_2 conservativity of $\mathrm{I}\Sigma_1$ over primitive recursive arithmetic. The latter is illustrated by the conservativity of RCA_0 over $\mathrm{I}\Sigma_1$ and the Π^1_1-conservativity of Weak König's Lemma over RCA_0. The question we are interested in is whether the conservativity result of [22] also fits one of these patterns and, if so, which one.

Our main result is that the conservativity theorem of [22] is, quantitatively speaking, tame: RT^2_2 has at most polynomial speedup over RCA_0 for $\forall\Sigma^0_2$ sentences. Moreover, this is witnessed by a *polynomial simulation*: there is a polytime procedure that takes a proof of a $\forall\Sigma^0_2$ sentence in $\mathrm{WKL}_0 + \mathrm{RT}^2_2$ (where WKL_0 is RCA_0 extended by Weak König's Lemma) and outputs a proof of the same sentence in RCA_0. By [3, Proposition 4.6], it follows that proofs of purely first-order Π_3 sentences in $\mathrm{WKL}_0 + \mathrm{RT}^2_2$ can be translated in polynomial time into proofs in $\mathrm{I}\Sigma_1$.

To obtain our result, we make use of a general-purpose technique due to Avigad [1]: in order to show polynomial simulation of a theory T by S, build what could be called a "forcing interpretation" of T in S — that is, formalize within S a forcing construction that leads to a model of T — and verify via small proofs in S that sentences of the appropriate class are forced if they are

2

true in the ground model. Interestingly, while most forcing arguments give rise to a generic extension of the ground model, the one we work with produces a *generic cut*, refining an initial segment construction used in [22]. Both the construction of [22] and ours rely on non-trivial finite combinatorics. In the case of [22], this took the form of an upper bound on finite Ramsey's theorem with the size of finite sets expressed in terms of so-called α-largeness. Here, we need a considerable strengthening of that bound with a more constructive proof, which was provided in [19].

We also consider the question whether an analogous polynomial simulation result still holds if we weaken the base theory to RCA_0^*, which differs from RCA_0 in that the Σ_1^0 induction scheme is replaced by Δ_0^0 induction plus an axiom exp guaranteeing totality of the exponential function. It turns out that, perhaps surprisingly, changing the base theory makes a major difference: even though RT_2^2 is $\forall\Sigma_2^0$-conservative over RCA_0^* (as can be shown by the method used to derive Π_2^0-conservativity in [28]), it has iterated exponential speedup over RCA_0^*, already for sentences of very limited syntactic complexity. We give a proof of the speedup for Σ_1 sentences; in fact, it could also be witnessed by finite consistency statements that can are expressible by bounded formulas with exponential terms.

Since $\mathrm{RCA}_0 + \mathrm{RT}_2^2$ proves Σ_2^0 collection, our main theorem immediately implies that the so-called Paris–Friedman theorem, i.e. the Π_{n+2}-conservativity of $\mathrm{B}\Sigma_{n+1}$ over $\mathrm{I}\Sigma_n$, can be strengthened to a polynomial simulation in the case when $n = 1$. It makes sense to ask whether this extends to other values of n, especially because a related question — whether the conservativity can be proved in bounded arithmetic — was asked by Clote et al. [7]. We prove that the answer to both questions is positive for all $n \geqslant 1$, and that it is positive for $n = 0$ if both $\mathrm{B}\Sigma_1$ and $\mathrm{I}\Delta_0$ are extended by exp. We do this by turning one of the model-theoretic proofs of Paris–Friedman into a forcing interpretation. As discussed in slightly greater detail in Section 4, Fedor Pakhomov [private communication] independently devised a completely different approach that proves a more general result; in particular, his argument obviates the need for exp in the $n = 0$ case.

The remainder of this paper has the following structure. After introducing some basic definitions and notational conventions below, we discuss the general concept of forcing interpretations in Section 1. We prove our main theorem in Section 2 and the contrasting speedup result over RCA_0^* in Section 3. The polynomial simulation of $\mathrm{B}\Sigma_{n+1} + \exp$ by $\mathrm{I}\Sigma_n + \exp$ is presented in Section 4.

<center>∗</center>

Basic information on fragments of first- and second-order arithmetic can be found in [15], [25], and [16].

We will write $\mathcal{L}_1$ and $\mathcal{L}_2$ for the languages of first- and second-order arithmetic, respectively. Note that these languages do *not* have a symbol for exponentiation. We use lowercase letters for objects of the numerical ("first-order") sort, and uppercase letters for objects of the set ("second-order") sort. Notation like Σ_n^0, Π_n^0 represents the usual formula classes defined in terms of first-order quantifier alternations, but allowing second-order free variables. On the other hand, notation without the superscript 0, like Σ_n, Π_n, represents analogously defined classes of $\mathcal{L}_1$ formulas — that is, without any second-order variables

<center>3</center>

at all. If we want to specify the second-order parameters appearing in a Σ_n^0 formula, we use notation like $\Sigma_n(\bar{X})$. We extend these conventions to naming theories: thus, for example, $B\Sigma_2^0$ is the $\mathcal{L}_2$-theory axiomatized by Δ_0^0 induction and Σ_2^0 collection, whereas $B\Sigma_2$ is the $\mathcal{L}_1$-theory axiomatized by Δ_0 induction and Σ_2 collection.

Recall that RCA_0^* is the theory defined in [26] which differs from RCA_0 in that Σ_1^0 induction is replaced by Δ_0^0 induction plus an axiom, known as exp, stating the totality of exponentiation. The first-order consequences of RCA_0^* are known to coincide with the theory $B\Sigma_1 + \exp$, which in particular means that RCA_0^* is Π_2-conservative over $I\Delta_0 + \exp$.

The symbol $\mathbb{N}$ stands for the set of natural numbers — both as understood in the metatheory (the *standard* natural numbers) or as formalized in an $\mathcal{L}_1$ or $\mathcal{L}_2$ theory (which, if consistent, will of course have nonstandard models). It should be clear from the context which is meant. A set $X \subseteq \mathbb{N}$ is *finite* if it is bounded, i.e. there is $k \in \mathbb{N}$ such that $\ell \leqslant k$ for all $\ell \in X$; otherwise, X is *infinite*. Write $X \subseteq_{\mathrm{fin}} Y$ for X is a finite subset of Y. Each finite set X is coded in a standard way by the binary representation of some natural number x, and may be identified with x in many contexts. The symbol ω stands for the smallest infinite ordinal — again, we use the same symbol in the metatheory and in formal theories.

We will be interested in comparing the sizes of proofs of statements in various theories, which requires us to fix some terminology and conventions regarding syntax. We use the word *theory* to mean a set of sentences. Every theory T comes with a language $\mathcal{L}(T)$ with the property that each non-logical symbol appearing in T is in $\mathcal{L}(T)$, though not necessarily vice versa. To simplify things, *we assume that $\mathcal{L}(T)$ always contains only finitely many non-logical symbol*s. The *size* of a term, a formula or a proof is the number of symbols in it. All variable, constant, function and relation symbols count as one symbol. We use vertical lines, $|\cdot|$, to denote the size of a given syntactical object.

To measure the sizes of proofs precisely, one needs to fix a proof system. It is known that there are polynomial-time translations between the usual Hilbert-style systems and the sequent calculus *with the cut rule* [11], and between their tree-like or sequence-like versions [23, Theorem 4.1]. Therefore, in principle, it does not matter which proof system we choose here. In our arguments, we have in mind a Hilbert-style system like the one in Enderton [12, Section 2.4], except that instead of allowing all (universal closures of) propositional tautologies as axioms, we derive them from a finite number of propositional axiom schemes. We assume that, like in Enderton's system, the only official connectives are $\neg, \rightarrow$ and the only quantifier is $\forall$. Recall [12, page 113] that a term t is *substitutable* for a variable symbol v in a formula φ if no free occurrence of v in φ is in the scope of a quantifier binding a variable that appears in t.

Definition 0.1. Let T, T' be theories and let Γ be a set of sentences in the language $\mathcal{L}(T) \cap \mathcal{L}(T')$. Then T *polynomially simulates* T' with respect to Γ if there exists a polynomial-time procedure that, given any proof of $\gamma \in \Gamma$ in T' as input, outputs a proof of γ in T'. (In particular, this implies that for every proof of $\gamma \in \Gamma$ in T', there is a proof of γ in T of at most polynomially larger size.)

The theory T' has *non-elementary speedup* over T with respect to Γ if for each elementary recursive function f, there exist $\gamma \in \Gamma$ and a proof π of γ in T'

such that no proof of γ in T has size $\leqslant f(|\pi|)$.

1 Forcing interpretations

The technique we use for proving polynomial simulations was developed in Avigad [1] and relies on a certain kind of formalized forcing argument. In essence, the idea is to define a more general notion of interpretation: while a traditional interpretation gives a uniform way of defining a model $M' \models T'$ inside a model $M \models T$, an Avigad-style interpretation gives a way of describing a generic model $M[G] \models T'$ which is not fully specified until the generic filter G is fixed. The point is that, even though G might not be definable, the properties of $M[G]$ that we care about in the context at hand do not depend on G.

In this section, we discuss this idea, and its connection to questions of proof size, in some generality. We begin with an annoying technical issue.

1.1 Simplification

When defining intepretations, formulas with more than one function symbol in an atomic subformula can cause ambiguities. In this subsection, we explain how to get around the problem by avoiding such formulas as much as possible. This is quite similar in spirit to translating all formulas to a relational language.

Definition 1.1. A *simple* term is one of the form v or $f(\bar{w})$, where $v, \bar{w}$ are variables and f is a function symbol. A *simple* formula is a formula in which every atomic subformula is either

- a formula with no function symbol; or

- an equation with exactly one function symbol.

Constant symbols are regarded as 0-ary function symbols here.

Definition 1.2. For a formula θ, its simple translation θ^* is defined in the following way.

First, to each term t and each variable x not appearing in t, we associate a simple formula $t[x]$, intended to define x to be the term t, by induction on the subterms of t. For a variable y, the formula $y[x]$ is $x = y$. For t of the form $f(t_1, \ldots, t_n)$, the formula $t[x]$ is

$$\forall x_1 \ldots \forall x_n \, (t_1[x_1] \to (t_2[x_2] \to \ldots \to (t_n[x_n] \to x = f(x_1, \ldots, x_n)) \ldots)),$$

where $x_1, \ldots, x_n$ are canonically chosen fresh variables.

The formula θ^* equals θ for θ simple atomic. If θ is $R(t_1, \ldots, t_n)$ and is not simple, then θ^* is

$$\forall x_1 \ldots \forall x_n \, (t_1[x_1] \to (t_2[x_2] \to \ldots \to (t_n[x_n] \to R(x_1, \ldots, x_n)) \ldots))$$

where $x_1, \ldots, x_n$ are canonically chosen fresh variables. The translation commutes with $\neg$, $\to$, and $\forall$.

The following two lemmas mean that when studying questions of proof size, we can largely restrict attention to simple formulas. The lemmas can essentially be read off Visser's exposition in [27, Section 7.3].

Lemma 1.3. There exists a polynomial-time procedure that, given a formula θ, outputs a first-order logic proof of $\theta \leftrightarrow \theta^*$.

Proof. This is essentially Theorem 7.3.6 in Visser [27], with the following differences: (i) Visser defines the translation $\theta \mapsto \theta^*$ (where θ^* would be $\theta^{*\circ}$ in his notation) using $\exists$ and $\wedge$, while we use $\forall$ and $\rightarrow$ to match the official connectives of our proof system, and (ii) Visser only needs a formal proof of $\theta \leftrightarrow \theta^*$ of size polynomial in $|\theta|$, so he does not discuss the time required to construct it. However, these differences have no bearing on the proof, and the reader can verify that all of Visser's constructions can be carried out in polynomial time. $\square$

Lemma 1.4. There is a polynomial-time procedure that, given a proof π in first-order logic, outputs a proof π^* with the following properties:

- π^* is a proof in the same language as π;

- every formula in π^* is simple;

- if $\theta_1, \theta_2, \ldots, \theta_\ell$ are the non-logical axioms in π, then $\theta_1^*, \theta_2^*, \ldots, \theta_\ell^*$ are the non-logical axioms in π^*;

- if η is the conclusion of π, then η^* is the conclusion of π^*.

Proof. See Theorem 7.3.3 and Theorem 7.3.4 in Visser [27], with the same caveats as in Lemma 1.3. $\square$

1.2 Forcing translations and interpretations

A traditionally understood interpretation of a theory T' in a theory T is essentially (i) a translation of $\mathcal{L}(T')$ into $\mathcal{L}(T)$ such that (ii) T proves the translations of axioms of T'. Forcing interpretations of the sort we need here have a similar two-layered structure of "translation" and "interpretation proper", except that now instead of translating an $\mathcal{L}(T')$ formula φ into $\mathcal{L}(T)$, we have to translate "φ is forced" into $\mathcal{L}(T)$. Then we have to verify in T that the axioms of T' are forced.

Definition 1.5. A *forcing translation* τ from a language $\mathcal{L}'$ to a language $\mathcal{L}$ consists of $\mathcal{L}$ formulas

$$s \in \mathrm{Cond}_\tau, \quad s' \trianglelefteq_\tau s, \quad s \Vdash_\tau v{\downarrow}, \quad s \Vdash_\tau \alpha(v_1, v_2, \ldots, v_\ell)$$

for every simple atomic $\mathcal{L}'$ formula $\alpha(v_1, v_2, \ldots, v_\ell)$ such that

(FT1) $s' \trianglelefteq_\tau s$ contains $s' \in \mathrm{Cond}_\tau \wedge s \in \mathrm{Cond}_\tau$ as a conjunct;

(FT2) $s \Vdash_\tau v{\downarrow}$ contains $s \in \mathrm{Cond}_\tau$ as a conjunct;

(FT3) $s \Vdash_\tau \alpha(v_1, v_2, \ldots, v_\ell)$ contains $\bigwedge_{i=1}^{\ell} s \Vdash_\tau v_i{\downarrow}$ as a conjunct, whenever $\alpha(v_1, v_2, \ldots, v_\ell)$ is a simple atomic $\mathcal{L}'$ formula; and

(FT4) if $\alpha(\bar{u}, v)$ is a simple atomic $\mathcal{L}'$ formula and w is a variable, then

$$\big(s \Vdash_\tau \alpha(\bar{u}, v)\big)[v/w] \quad \text{and} \quad s \Vdash_\tau \big(\alpha(\bar{u}, v)[v/w]\big)$$

are the same.

All formulas above have exactly the free variables shown. When there is no risk of ambiguity, we will often omit the subscript τ in the notation. We read '$s \Vdash \ldots$' as 's forces $\ldots$'. By convention, the variable symbols $s, s', s'', \ldots$ are always distinct from $u, v, w, z, \ldots$. In accordance with the usual customs related to forcing, we refer to elements s satisfying $s \in \mathrm{Cond}$ as forcing *conditions*, and think of the objects denoted by the variables $u, v, w, z, \ldots$ as *names* (for the elements of the generic model being described).

Remark. In some contexts, it might make sense to allow forcing conditions to be tuples of elements instead of single elements. However, in this paper we deal exclusively with theories that have a definable pairing function. So to avoid complicating the notation we eschew that kind of generality and continue to write s rather than $\bar{s}$.

The clauses we use to extend a forcing relation to arbitrary simple formulas follow those of a "good strong forcing notion" in the sense of Avigad [1, Definition 4.2]. We deal with non-simple formulas using the $(\cdot)^*$ translation.

Definition 1.6. Let τ be a forcing translation from a language $\mathcal{L}'$ to a language $\mathcal{L}$. We define an $\mathcal{L}$ formula $s \Vdash_\tau \theta(v_1, v_2, \ldots, v_\ell)$ for each $\mathcal{L}'$ formula $\theta(v_1, v_2, \ldots, v_\ell)$ by recursion on θ as follows.

(FT5) If $\theta(v_1, v_2, \ldots, v_\ell)$ is a simple $\mathcal{L}'$ formula, then

$$s \Vdash_\tau \neg\theta(v_1, v_2, \ldots, v_\ell)$$

is defined to be

$$\bigwedge_{i=1}^{\ell} \text{``}s \Vdash_\tau v_i{\downarrow}\text{''} \wedge \forall s' \trianglelefteq_\tau s \ s' \nVdash_\tau \theta(\bar{v}).$$

(FT6) If $\theta(u_1, u_2, \ldots, u_k, w_1, w_2, \ldots, w_m)$ and $\eta(v_1, v_2, \ldots, v_\ell, w_1, w_2, \ldots, w_m)$ are simple $\mathcal{L}'$ formulas, then

$$s \Vdash_\tau \theta(u_1, u_2, \ldots, u_k, w_1, w_2, \ldots, w_m) \to \eta(v_1, v_2, \ldots, v_\ell, w_1, w_2, \ldots, w_m)$$

is defined to be

$$\bigwedge_{i=1}^{k} \text{``}s \Vdash_\tau u_i{\downarrow}\text{''} \wedge \bigwedge_{i=1}^{\ell} \text{``}s \Vdash_\tau v_i{\downarrow}\text{''} \wedge \bigwedge_{i=1}^{m} \text{``}s \Vdash_\tau w_i{\downarrow}\text{''}$$
$$\wedge \ \forall s' \trianglelefteq_\tau s \ \exists s'' \trianglelefteq_\tau s' \ \left(\text{``}s' \Vdash_\tau \theta(\bar{u}, \bar{w})\text{''} \to \text{``}s'' \Vdash_\tau \eta(\bar{v}, \bar{w})\text{''} \right).$$

(FT7) If $\theta(v_1, v_2, \ldots, v_\ell, w)$ is a simple $\mathcal{L}'$ formula, then

$$s \Vdash_\tau \forall w \, \theta(v_1, v_2, \ldots, v_\ell, w)$$

is defined to be

$$\bigwedge_{i=1}^{\ell} \text{``}s \Vdash_\tau v_i{\downarrow}\text{''} \wedge \forall w \, \forall s' \trianglelefteq_\tau s \ \exists s'' \trianglelefteq_\tau s' \ \left( \text{``}s' \Vdash_\tau w{\downarrow}\text{''} \to \text{``}s'' \Vdash_\tau \theta(\bar{v}, w)\text{''} \right).$$

(FT8) If $\theta(\bar{v})$ is an $\mathcal{L}'$ formula that is not simple, then $s \Vdash_\tau \theta(\bar{v})$ is defined to be $s \Vdash_\tau \theta^*(\bar{v})$, where $\theta^*(\bar{v})$ is as defined in Definition 1.2.

(FT9) $\wedge, \vee, \leftrightarrow, \exists$ are defined in terms of $\neg, \rightarrow, \forall$ in the usual way.

Here the quotation marks "$\ldots$" are simply another type of brackets to enhance the readability of formulas. Our convention is that $\Vdash$ has lower precedence than all the logical connectives in first-order logic. All formulas above have exactly the free variables shown. We often abbreviate $\bigwedge\!\!\!\bigwedge_i (s \Vdash_\tau v_i\!\downarrow)$ as $s \Vdash_\tau \bar{v}\!\downarrow$.

The definitions above are designed to make (FT3) and (FT4) hold for all formulas, at least essentially.

Lemma 1.7. Let τ be a forcing translation from a language $\mathcal{L}'$ to a language $\mathcal{L}$. Then:

(i) $s \Vdash_\tau \theta(\bar{v})$ contains $s \Vdash \bar{v}\!\downarrow$ as a conjunct for every $\mathcal{L}'$ formula $\theta(\bar{v})$;

(ii) if $\theta(\bar{u}, v)$ is a $\mathcal{L}'$ formula and w is a variable, then

$$\big(s \Vdash_\tau \theta(\bar{u}, v)\big)[v/w] \quad \text{and} \quad s \Vdash_\tau \big(\theta(\bar{u}, v)[v/w]\big)$$

differ only by a one-to-one renaming of bound variables, and provided θ is simple, the variable w is substitutable for v in $\theta(\bar{u}, v)$ exactly when it is so in $s \Vdash_\tau \theta(\bar{u}, v)$.

Proof. Part (i) is true by construction. Part (ii), which is a special case of Lemma 4.3 in Avigad [1], can be proved by a straightforward induction on θ using the definitions when θ is simple. Our convention that the variable symbols $s, s', s'', \ldots$ are always distinct from $u, v, w, z, \ldots$ is used to establish the substitutability part. The rest follows from (FT8). $\qquad\square$

Definition 1.8. A *forcing interpretation* of a theory T' in a theory T is a forcing translation τ from $\mathcal{L}(T')$ to $\mathcal{L}(T)$ such that T proves

(FI0) $\exists s \, (s \in \mathrm{Cond}_\tau)$;

(FI1) $\forall s \in \mathrm{Cond}_\tau \; s \trianglelefteq_\tau s$;

(FI2) $\forall s, s', s'' \in \mathrm{Cond}_\tau \, \big(s'' \trianglelefteq_\tau s' \wedge s' \trianglelefteq_\tau s \rightarrow s'' \trianglelefteq_\tau s\big)$;

(FI3) $\forall s \in \mathrm{Cond}_\tau \; \exists s' \trianglelefteq_\tau s \; \exists v \, s' \Vdash_\tau v\!\downarrow$;

(FI4) $\forall s, s' \in \mathrm{Cond}_\tau \; \forall v \, \big(s' \trianglelefteq_\tau s \wedge \text{``}s \Vdash_\tau v\!\downarrow\text{''} \rightarrow \text{``}s' \Vdash_\tau v\!\downarrow\text{''}\big)$;

(FI5) for all simple atomic $\mathcal{L}(T')$ formulas $\alpha(\bar{v})$,

$$\forall s, s' \in \mathrm{Cond}_\tau \; \forall \bar{v} \, \big(s' \trianglelefteq_\tau s \wedge \text{``}s \Vdash_\tau \alpha(\bar{v})\text{''} \rightarrow \text{``}s' \Vdash_\tau \alpha(\bar{v})\text{''}\big);$$

(FI6) $\forall s \in \mathrm{Cond}_\tau \; \forall v \, \big(\text{``}s \Vdash_\tau v\!\downarrow\text{''} \rightarrow \text{``}s \Vdash_\tau v = v\text{''}\big)$;

(FI7) $\forall s \in \mathrm{Cond}_\tau \; \forall u, v \, \big(\text{``}s \Vdash_\tau u = v\text{''} \rightarrow \text{``}s \Vdash_\tau v = u\text{''}\big)$;

(FI8) $\forall s \in \mathrm{Cond}_\tau \; \forall u, v, w \, \big(\text{``}s \Vdash_\tau u = v\text{''} \wedge \text{``}s \Vdash_\tau v = w\text{''} \rightarrow \text{``}s \Vdash_\tau u = w\text{''}\big)$;

(FI9) for all $k \in \mathbb{N}$ and all k-ary function symbols f in $\mathcal{L}(T')$,

$$\forall s \in \mathrm{Cond}_\tau \; \forall v_1, v_2, \ldots, v_k \, \Big(\bigwedge\!\!\!\bigwedge_{i=1}^{k} \text{``}s \Vdash_\tau v_i\!\downarrow\text{''}$$
$$\rightarrow \forall s' \trianglelefteq s \; \exists s'' \trianglelefteq s' \; \exists w \, \text{``}s'' \Vdash_\tau w = f(\bar{v})\text{''}$$
$$\wedge \; s \Vdash_\tau \forall w, w' \, \big(w = f(\bar{v}) \rightarrow (w' = f(\bar{v}) \rightarrow w = w')\big)\Big);$$

(FI10) for all simple $\mathcal{L}(T')$ terms $t(v_1, v_2, \ldots, v_k)$ with exactly the free variables shown and all simple atomic $\mathcal{L}(T')$ formulas $\alpha(\bar{u}, w_0)$ such that $\alpha(\bar{u}, t(\bar{v}))$ is simple,

$$\forall s \in \mathrm{Cond}_\tau \, \forall \bar{u}, v_1, v_2, \ldots, v_k, w$$
$$\left(\text{``}s \Vdash_\tau w = t(\bar{v})\text{''} \to \left(\text{``}s \Vdash_\tau \alpha(\bar{u}, w)\text{''} \leftrightarrow \text{``}s \Vdash_\tau \alpha(\bar{u}, t(\bar{v}))\text{''}\right)\right);$$

(FI11) $\forall s \in \mathrm{Cond}_\tau \, \forall v \, (\forall s' \trianglelefteq_\tau s \, \exists s'' \trianglelefteq_\tau s' \, \text{``}s'' \Vdash_\tau v{\downarrow}\text{''} \to s \Vdash_\tau v{\downarrow});$

(FI12) for all simple atomic $\mathcal{L}(T')$ formulas $\alpha(\bar{v})$,

$$\forall s \in \mathrm{Cond}_\tau \, \forall \bar{v} \, \left(\forall s' \trianglelefteq_\tau s \, \exists s'' \trianglelefteq_\tau s' \, \text{``}s'' \Vdash_\tau \alpha(\bar{v})\text{''} \to s \Vdash_\tau \alpha(\bar{v})\right);$$

(FI13) $\forall s \in \mathrm{Cond}_\tau \, s \Vdash_\tau \sigma \quad$ for all $\sigma \in T'$.

We refer to a forcing translation satisfying (FI0)–(FI12) (that is, to a forcing interpretation of pure logic) as a *forcing interpretation of $\mathcal{L}(T')$ in T*.

Clauses (FI0) and (FI3) are technical. (FI1)–(FI2) state that $\trianglelefteq$ is a preorder. Meanwhile, we require neither antisymmetry nor the existence of a maximal element. (FI4)–(FI5) form the base case of the usual requirement that forcing be closed downwards under $\trianglelefteq$. (FI6)–(FI10) say roughly that the equality axioms are forced. The variable symbol w in (FI10) may be syntactically equal to one of $\bar{u}$, but by convention w_0 cannot be. (FI11)–(FI12) express the base case of the connection between forcing $\neg\neg\varphi$ and forcing φ. Finally, the crucial condition (FI13) is what makes a forcing interpretation of pure logic be an interpretation of a theory T': all axioms of T' are forced.

Forcing interpretations are clearly closed under composition and definition by cases. As shown by the following example, our notion of forcing interpretations generalizes the usual notion of interpretations.

Example 1.9. Every interpretation (or, more precisely, every parameter-free one-dimensional global relative interpretation) τ of a theory T' in a theory T gives rise to a forcing interpretation of T' in T as follows.

- Define $s \in \mathrm{Cond}_\tau$ and $s' \trianglelefteq_\tau s$ to be respectively

$$s = s \quad \text{and} \quad s' = s' \wedge s = s.$$

- Define $s \Vdash_\tau v{\downarrow}$ to be $s \in \mathrm{Cond}_\tau \wedge \delta(v)$, where $\delta(v)$ is the defining formula for the domain of τ.

- Define $s \Vdash_\tau \alpha(\bar{v})$, where $\alpha(\bar{v})$ is a simple atomic $\mathcal{L}(T')$ formula, according to the interpretation τ.

1.3 Forcing interpretations and polynomial simulation

We now turn to the question of what is required of a forcing interpretation if it is to imply a polynomial simulation between theories.

We begin by verifying that proofs of some basic facts, including generalizations of (FI5), (FI10), and (FI12) to all formulas, can be found in polynomial time for any forcing interpretation. Our convention that the language of a theory is always finite is implicitly used in the proofs of the lemmas below.

Lemma 1.10. Let τ be a forcing interpretation of $\mathcal{L}(T')$ in a theory T. Then there is a polynomial-time procedure which, given an $\mathcal{L}(T')$ formula $\theta(\bar{v})$, outputs a proof in T of

$$\forall s \in \mathrm{Cond}_\tau \, \forall \bar{v} \, \neg \big(\text{``}s \Vdash_\tau \theta(\bar{v})\text{''} \wedge \text{``}s \Vdash_\tau \neg\theta(\bar{v})\text{''}\big).$$

Proof. Apply (FT5). $\qquad\square$

Lemma 1.11. Let τ be a forcing interpretation of $\mathcal{L}(T')$ in a theory T. Then there is a polynomial-time procedure which, given an $\mathcal{L}(T')$ formula $\theta(\bar{v})$, outputs proofs in T of:

(1) $\forall s, s' \in \mathrm{Cond}_\tau \, \forall \bar{v} \, \big(s' \trianglelefteq_\tau s \wedge \text{``}s \Vdash_\tau \theta(\bar{v})\text{''} \to \text{``}s' \Vdash_\tau \theta(\bar{v})\text{''}\big)$; and

(2) $\forall s \in \mathrm{Cond}_\tau \, \forall \bar{v} \, \big(\forall s' \trianglelefteq_\tau s \, \exists s'' \trianglelefteq_\tau s' \, \text{``}s'' \Vdash_\tau \theta(\bar{v})\text{''} \to \text{``}s \Vdash_\tau \theta(\bar{v})\text{''}\big)$.

Proof. These are part of Lemma 4.3 and Lemma 4.6 in Avigad [1], except that we additionally need to pay attention to the computational complexity of the proof constructions. The proofs are built using induction on the structure of θ, with the inductive step split into cases depending on the outermost logical connective in θ. In each case, the construction is straightforward. $\qquad\square$

Lemma 1.12. Let τ be a forcing interpretation of $\mathcal{L}(T')$ in a theory T. There is a polynomial-time procedure which — given an $\mathcal{L}(T')$ term $t(\bar{v})$ with exactly the free variables shown and a simple $\mathcal{L}(T')$ formula $\theta(\bar{u}, w_0)$ such that $\theta(\bar{u}, t(\bar{v}))$ is also simple, and both w and $t(\bar{v})$ are substitutable for w_0 in $\theta(\bar{u}, w_0)$ — outputs a proof in T of

$$\forall s \in \mathrm{Cond}_\tau \, \forall \bar{u}, \bar{v}, w \, \big(\text{``}s \Vdash_\tau w = t(\bar{v})\text{''} \to (\text{``}s \Vdash_\tau \theta(\bar{u}, w)\text{''} \leftrightarrow \text{``}s \Vdash_\tau \theta(\bar{u}, t(\bar{v}))\text{''})\big).$$

Proof. This is part of Lemma 4.3 in Avigad [1], again with attention paid to the computational complexity of the construction. As previously, one proceeds by induction on the structure of θ. $\qquad\square$

Definition 1.13. A forcing interpretation τ of a theory T' in a theory T is *polynomial* if there is polynomial-time procedure which, given any $\sigma \in T'$, outputs a proof in T of the sentence $\forall s \in \mathrm{Cond}_\tau \, s \Vdash_\tau \sigma$.

Note that if T' is finitely axiomatized, then any forcing interpretation of T' is automatically a polynomial forcing interpretation.

Proposition 1.14. Let τ be a forcing interpretation of a theory T' in a theory T. For all $\mathcal{L}(T')$ formulas $\varphi(\bar{v}), \psi(\bar{v})$, if $T' + \varphi(\bar{v}) \vdash \psi(\bar{v})$, then T proves

$$\forall s \in \mathrm{Cond}_\tau \, \forall \bar{v} \, \big(s \Vdash_\tau \varphi(\bar{v}) \to s \Vdash_\tau \psi(\bar{v})\big). \tag{1}$$

If τ is polynomial, then a proof of (1) in T' can be found in polynomial time given φ, ψ, and a proof π of $\psi(\bar{v})$ from $T' + \varphi(\bar{v})$.

Proof. Let π^* be the proof obtained from π according to Lemma 1.4. By (FT8) and (FI13), for each non-logical axiom σ used in π^* we can find a proof in T of $\forall s \in \mathrm{Cond} \, \forall \bar{v} \, (s \Vdash \varphi(\bar{v}) \to s \Vdash \sigma)$. If τ is a polynomial forcing interpretation of T' in T, these proofs can be found in polynomial time.

We then construct analogous proofs for each line σ in π^*, by a routine induction on the structure of π^*; cf. Proposition 4.8 in Avigad [1]. By (FT8), in the case of the last line this is a proof of $\forall s \in \mathrm{Cond} \, \forall \bar{v} \, (s \Vdash \varphi(\bar{v}) \to s \Vdash \psi(\bar{v}))$. $\qquad\square$

We note the following important special case of Proposition 1.14.

Corollary 1.15. Let τ be a forcing interpretation of T' in T. Then for each $\mathcal{L}(T')$ sentence σ, if T' proves σ, then T proves $\forall s \in \mathrm{Cond}_\tau\,(s \Vdash_\tau \sigma)$. Moreover, if τ is polynomial, then a proof of $\forall s \in \mathrm{Cond}_\tau\,(s \Vdash_\tau \sigma)$ in T can be found algorithmically in polynomial time given a proof of σ in T'.

Definition 1.16. Let T, T' be theories and Γ be a set of sentences in the language $\mathcal{L}(T) \cap \mathcal{L}(T')$. A forcing interpretation τ of T' in T is said to be Γ-*reflecting* if, for all $\gamma \in \Gamma$, T proves

$$\forall s \in \mathrm{Cond}_\tau\,(s \Vdash_\tau \gamma) \to \gamma. \tag{2}$$

The interpretation is *polynomially* Γ-*reflecting* if a T-proof of (2) can be found in polynomial time on input $\gamma \in \Gamma$.

Theorem 1.17 (essentially Avigad [1, Section 10])**.** Let T, T' be theories and Γ be a set of sentences in the language $\mathcal{L}(T) \cap \mathcal{L}(T')$. If there is a Γ-reflecting forcing interpretation τ of T' in T, then T' is Γ-conservative over T. Moreover, if such τ is polynomial and polynomially Γ-reflecting, then T polynomially simulates T' with respect to Γ.

Proof. Combine (FI0), Corollary 1.15, and Definition 1.16. $\qquad\square$

2 Ramsey for pairs: polynomial simulation in RCA_0

In this section, we prove the main theorem of our paper.

Theorem 2.1. RCA_0 polynomially simulates $\mathrm{WKL}_0 + \mathrm{RT}_2^2$ with respect to $\forall \Sigma_2^0$ sentences.

This is a strengthening of the main result of [22] that $\mathrm{WKL}_0 + \mathrm{RT}_2^2$ is $\forall \Sigma_2^0$-conservative over $\mathrm{I}\Sigma_1^0$. Since RCA_0 is polynomially simulated by $\mathrm{I}\Sigma_1$ w.r.t. $\mathcal{L}_1$ sentences (see [17, 3] or consider the obvious interpretation of RCA_0 in $\mathrm{I}\Sigma_1$ and refer to Section 1), Theorem 2.1 has the following immediate corollary.

Corollary 2.2. $\mathrm{I}\Sigma_1$ polynomially simulates $\mathrm{WKL}_0 + \mathrm{RT}_2^2$ with respect to Π_3 sentences.

Our proof of Theorem 2.1 requires a stronger and more explicit version of a result in finite combinatorics used to prove the conservativity theorem of [22]. The combinatorial statement is formulated using the concept of α-largeness (originally introduced in [18] as a tool to study the unprovability of the Paris–Harrington theorem), which provides a framework for measuring the size of finite subsets of $\mathbb{N}$ by means of countable ordinals rather than just natural numbers. We now recall some basic notions related to that framework.

For a given $\alpha < \omega^\omega$ and $m \in \mathbb{N}$, define $0[m] = 0$, $\alpha[m] = \beta$ if $\alpha = \beta + 1$, and $\alpha[m] = \beta + \omega^{n-1} \cdot m$ if $\alpha = \beta + \omega^n$ for some $n \geqslant 1$.

Definition 2.3 ($\mathrm{I}\Delta_0 + \exp$)**.** Let $\alpha < \omega^\omega$. A set $X = \{x_0 < \cdots < x_{\ell-1}\} \subseteq_{\mathrm{fin}} \mathbb{N}$ is said to be α-*large* if $\alpha[x_0] \ldots [x_{\ell-1}] = 0$. In other words, any finite set is 0-large, and X is α-large when

- $X \setminus \{\min X\}$ is β-large if $\alpha = \beta + 1$,

- $X \setminus \{\min X\}$ is $(\beta + \omega^{n-1} \cdot \min X)$-large if $\alpha = \beta + \omega^n$ for $n \geqslant 1$.

The proposition below lists some well-known and simple but important properties of α-largeness. Their provability in $I\Delta_0 + \exp$ was pointed out e.g. in [19].

Proposition 2.4. The following are provable in $I\Delta_0 + \exp$:

(i) If a set X is α-large and $X \subseteq Y \subseteq_{\mathrm{fin}} \mathbb{N}$, then Y is α-large.

(ii) If $\alpha = \omega^{n_k} + \cdots + \omega^{n_0}$ where $n_k \geqslant \cdots \geqslant n_0$, then a set X is α-large if and only if there exist sets $X_0, \ldots, X_k$ such that $X = X_0 \sqcup \cdots \sqcup X_k$, each X_i is ω^{n_i}-large, and for each $i < k$, $\max X_i < \min X_{i+1}$.

As a result, if X is ω^n-large and $k = \min X$, then $X \setminus \{k\} = X_0 \sqcup \cdots \sqcup X_{k-1}$ where each X_i is ω^{n-1}-large, and for each $i < k - 1$, $\max X_i < \min X_{i+1}$.

(iii) If X is $\omega^n \cdot 2$-large and $X = X_0 \cup X_1$, then at least one of X_0, X_1 is ω^n-large.

The following result is crucial in the proof of the main theorem of [22].

Proposition 2.5. [22, Proposition 7.7] For every natural number n there exists a natural number m such that $I\Sigma_1$ proves: for every $X \subseteq_{\mathrm{fin}} \mathbb{N}$ with $\min X \geqslant 3$, if X is ω^m-large, then every colouring $f \colon [X]^2 \to 2$ has an ω^n-large homogeneous set.

As the analysis in the remainder of this section will reveal, proving Theorem 2.1 involves obtaining an upper bound on m in terms of n. The statement of Proposition 2.5 gives no such upper bound, and any bound that could be derived from a simple-minded analysis of the proof in [22] would be very weak — certainly not even elementary recursive. On the other hand, proving Theorem 2.1 requires a bound significantly better than exponential. We will use the bound obtained in [19]:

Theorem 2.6. [19, Theorem 1.6] Provably in $I\Sigma_1$, if $X \subseteq_{\mathrm{fin}} \mathbb{N}$ is ω^{300x}-large, then every colouring $f \colon [X]^2 \to 2$ has an ω^x-large homogeneous set.

Remark. Note that in [19], the theorem includes the additional assumption that $\min X \geqslant 3$. However, using the proof presented in [19] and some very straightforward combinatorics of α-large sets, it is easy to verify that the assumption is in fact unnecessary. To avoid annoying technicalities, we nevertheless assume that any set X mentioned in the context of α-largeness satisfies $\min X \geqslant 1$.

Our proof of Theorem 2.1 will involve an argument by cases, in which the case distinction is based on how much induction is available. We introduce an auxiliary theory, originally motivated by the case in which a certain amount of induction *fails*, which turns out to be useful in the other case as well. The theories and polynomial simulation relationships we consider in the proof are summarized in Figure 1.

Definition 2.7. $\mathrm{RCA}_0 + \mathbb{I}$ is a theory in the language of second-order arithmetic extended by a new unary predicate $\mathbb{I}$ over the first-order sort.

The axioms of $\mathrm{RCA}_0 + \mathbb{I}$ are those of RCA_0 plus the following statements:
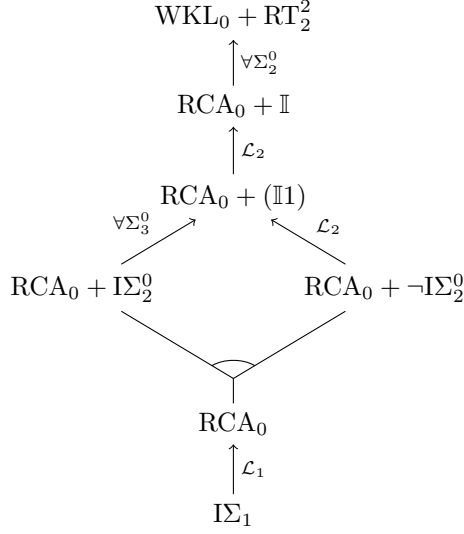
$$\text{WKL}_0 + \text{RT}_2^2$$

$$\uparrow \forall\Sigma_2^0$$

$$\text{RCA}_0 + \mathbb{I}$$

$$\uparrow \mathcal{L}_2$$

$$\text{RCA}_0 + (\mathbb{I}1)$$

$$\forall\Sigma_3^0 \nearrow \qquad \nwarrow \mathcal{L}_2$$

$$\text{RCA}_0 + \text{I}\Sigma_2^0 \qquad\qquad \text{RCA}_0 + \neg\text{I}\Sigma_2^0$$

$$\text{RCA}_0$$

$$\uparrow \mathcal{L}_1$$

$$\text{I}\Sigma_1$$

Figure 1: Polynomial simulations between the various theories in Section 2

($\mathbb{I}1$) $\mathbb{I}$ is a nonempty proper cut in the first-order universe,

($\mathbb{I}2$) $\mathbb{I}$ is closed under addition,

($\mathbb{I}3$) for every infinite set S, there exists finite $X \subseteq S$ which is ω^x-large for some $x > \mathbb{I}$.

For our purposes, the most important property of the theory $\text{RCA}_0 + \mathbb{I}$ will be that it polynomially simulates $\text{WKL}_0 + \text{RT}_2^2$ with respect to proofs of $\forall\Sigma_2^0$ sentences. Before proving that fact, which is Lemma 2.10 below, we verify some more basic properties of $\text{RCA}_0 + \mathbb{I}$.

Lemma 2.8. $\text{RCA}_0 + (\mathbb{I}1)$ polynomially simulates $\text{RCA}_0 + \mathbb{I}$ with respect to $\mathcal{L}_2$ sentences.

Proof. It is enough to show that there is an interpretation of $\text{RCA}_0 + \mathbb{I}$ in $\text{RCA}_0 + (\mathbb{I}1)$ that is the identity interpretation with respect to all symbols of $\mathcal{L}_2$.

So, let $\mathbb{J}$ consist of those numbers x for which every infinite set contains an ω^x-large finite subset; this is easily seen to be a definable cut in RCA_0.

Thus, $\mathbb{I} \cap \mathbb{J}$ is, provably in $\text{RCA}_0 + (\mathbb{I}1)$, a proper definable cut with the property that every infinite set contains an ω^x-large subset for each $x \in \mathbb{I} \cap \mathbb{J}$. It follows that every infinite set X actually contains an ω^x-large subset for some $x > \mathbb{I} \cap \mathbb{J}$, because the formula "$X$ contains an ω^x-large subset" is Σ_1^0, so by $\text{I}\Sigma_1^0$ it cannot define exactly the cut $\mathbb{I} \cap \mathbb{J}$. Let $\mathbb{K}$ be a cut closed under addition obtained from $\mathbb{I} \cap \mathbb{J}$ by applying Solovay's technique of shortening cuts: in other words, put $\mathbb{K} = \{a \in \mathbb{I} \cap \mathbb{J} : \forall x \in \mathbb{I} \cap \mathbb{J}(a + x \in \mathbb{I} \cap \mathbb{J})\}$. Then $\text{RCA}_0 + (\mathbb{I}1)$ proves the axioms of $\text{RCA}_0 + \mathbb{I}$ with $\mathbb{K}$ substituted for $\mathbb{I}$. $\qquad\square$

Lemma 2.9. $\text{RCA}_0 + (\mathbb{I}1)$ is polynomially simulated by:

(a) $\text{RCA}_0 + \text{I}\Sigma_2^0$ with respect to $\forall\Sigma_3^0$ sentences,

(b) $\mathrm{RCA}_0 + \neg\mathrm{I}\Sigma_2^0$ with respect to $\mathcal{L}_2$ sentences.

Proof. The proof of (b) is essentially immediate: by interpreting $\mathbb{I}$ as the intersection of all Σ_2^0-definable cuts, which is a definable proper cut in $\mathrm{RCA}_0 + \neg\mathrm{I}\Sigma_2^0$, we obtain an interpretation of $\mathrm{RCA}_0 + (\mathbb{I}1)$ in $\mathrm{RCA}_0 + \neg\mathrm{I}\Sigma_2^0$ that is the identity with respect to $\mathcal{L}_2$.

In the case of (a), let us first explain the underlying model-theoretic argument, which shows how to construct a model of $\mathrm{RCA}_0 + (\mathbb{I}1)$ from a model of $\mathrm{RCA}_0 + \mathrm{I}\Sigma_2^0$ while preserving the truth of a fixed $\exists\Pi_3^0$ sentence. We then give some hints as to how this is reflected on the syntactic level.

Let $(M, \mathscr{X}) \models \mathrm{RCA}_0 + \mathrm{I}\Sigma_2^0 + \exists X\,\delta(X)$, where $\delta \in \Sigma_4^0$. Take $A \in \mathscr{X}$ such that $(M, \mathscr{X}) \models \delta(A)$. From Beklemishev [2, Theorem 5.1], we know that $\mathrm{I}\Sigma_2$ proves the uniform Π_4 reflection principle for $\mathrm{I}\Sigma_1$. A relativization of this implies $(M, \mathscr{X}) \models \mathrm{Con}(\mathrm{I}\Sigma_1^0 + \delta(P))$, where P is a new predicate. An application of the Low Arithmetized Completeness Theorem [15, Theorem I.4.27] then gives an end-extension $M' \supsetneq_{\mathrm{e}} M$ and $A' \subseteq M'$ such that $(M', A') \models \mathrm{I}\Sigma_1^0 + \delta(P)$. Thus $(M', \Delta_1(P)\text{-Def}(M', A'), M) \models \mathrm{RCA}_0 + (\mathbb{I}1) + \exists X\,\delta(X)$. Moreover, the universe and operations in (M', A') are definable in M, as is the range of a bijection between M and an initial segment of M' closed under $+^{M'}$.

Crucially, the formulas used to define (M', A') are not only independent of the model M of $\mathrm{Con}(\mathrm{I}\Sigma_1^0 + \delta(P))$ that we consider, but they are also polynomial-time uniform in δ, in the sense that they can be constructed in polynomial time given δ as input. (Each of the formulas arises essentially by substituting the Gödel number of $\exists X\,\delta(X)$ into a Σ_2 formula $\gamma(x)$ describing a low branch of a binary tree that is Δ_1-definable with first-order parameter x and is infinite as long as x is the Gödel number of a sentence consistent with $\mathrm{I}\Sigma_1^0$.) So, if π is a proof of $\bot$ from $\mathrm{RCA}_0 + (\mathbb{I}1) + \exists X\,\delta(X)$, we can build a proof of $\bot$ from $\mathrm{I}\Sigma_2(P) + \delta(P)$ by taking π^* and "relativizing it to $(M', \Delta_1(P)\text{-Def}(M', P), M)$". That is, we relativize first-order quantifiers to the formula defining the universe of M' and replace first-order atoms by the formulas defining operations in M'; replace second-order quantifiers by quantification over pairs of a $\Sigma_1(P)$ formula and an equivalent $\Pi_1(P)$ formula, changing atoms $x \in X$ (for X different from P) into appropriate instances of the $\Sigma_1(P)$-universal formula; replace $\mathbb{I}$ by the formula defining the range of the inclusion $M \hookrightarrow M'$. Additionally, it is necessary to add polynomially many new proof lines in order to derive the translated axioms of $\mathrm{RCA}_0 + (\mathbb{I}1) + \exists X\,\delta(X)$ from $\mathrm{I}\Sigma_2(P) + \delta(P)$, and to deriving translations of conclusions from translations of premises for each inference in π^*. The details of the latter task are quite standard and somewhat similar to those in the proof that $\mathrm{I}\Sigma_1$ polynomially simulates RCA_0 based on the usual interpretation of RCA_0 in $\mathrm{I}\Sigma_1$, as described in e.g. [17, Chapter 3]. $\square$

Lemma 2.10. $\mathrm{RCA}_0 + \mathbb{I}$ polynomially simulates $\mathrm{WKL}_0 + \mathrm{RT}_2^2$ with respect to $\forall\Sigma_2^0$ sentences.

To prove Lemma 2.10, we now define a particular forcing notion in $\mathrm{RCA}_0 + \mathbb{I}$. We will then use a series of lemmas to show that this provides a polynomial forcing interpretation of $\mathrm{WKL}_0 + \mathrm{RT}_2^2$ in the sense of Section 1 and, moreover, that the forcing interpretation is polynomially $\forall\Sigma_2^0$-reflecting in the sense of Definition 1.16.

14

Definition 2.11. We let a *finite* set s be a forcing condition ($s \in \mathrm{Cond}$) if and only if s is ω^x-large for some $x > \mathbb{I}$. The relation $\trianglelefteq$ is defined simply as $\subseteq$. Note that there is no $\trianglelefteq$-largest condition.

There are two sorts of names, a first- and a second-order sort. A name of the first-order sort is simply a natural number. A name of the second-sort is also simply a natural number, this time viewed as coding a finite set according to the Ackermann interpretation. (Formally, one may think of the first-order names being $\langle 0, v \rangle$ and second-order names being $\langle 1, v \rangle$ for various numbers v.) To avoid notational confusion, we will write v for the first-order and V for the second-order names.

We say that $s \Vdash v\!\downarrow$ if $s \cap [1, v]$ is not a condition, that is, if it is not ω^x-large for any $x > \mathbb{I}$. On the other hand, $s \Vdash V\!\downarrow$ holds always.

For each simple atomic formula $\varphi(\bar{v})$, we define $s \Vdash \varphi$ to be $(s \Vdash \bar{v}\!\downarrow) \wedge \varphi(\bar{v})$. (In the special case of the simple atomic formula $v \in V$, this would be more precisely stated as $(s \Vdash v\!\downarrow \wedge s \Vdash V\!\downarrow) \wedge v \in_{\mathrm{Ack}} V$, where $\in_{\mathrm{Ack}}$ is the standard Ackermann interpretation of $\in$ in arithmetic. We shall ignore this detail from now on.)

Before showing that this forcing notion provides a polynomial forcing interpretation of $\mathrm{WKL}_0 + \mathrm{RT}_2^2$, we first discuss the model-theoretic intuition behind it. Let $(M, \mathscr{X}, I)$ be a countable model of $\mathrm{RCA}_0 + \mathbb{I}$. The set of forcing conditions Cond and the relation $\trianglelefteq$ of Definition 2.11 are both definable in $(M, \mathscr{X}, I)$. Take an $(M, \mathscr{X}, I)$-generic filter G of $(\mathrm{Cond}, \trianglelefteq)$, and put $I_G = \sup_M \{\min s : s \in G\}$. One may then check that $(I_G, \mathrm{Cod}(M/I_G)) \models \mathrm{WKL}_0 + \mathrm{RT}_2^2$ by simulating the inductive construction of a cut in the proof of [19, Theorem 3.3], where various choices made in the inductive steps are now replaced by the genericity of G. One may also check that for any $v \in M$, $v \in I_G$ if and only if $s \cap [1, v] \notin \mathrm{Cond}$ for some $s \in G$, and that $s \cap I_G$ is cofinal in I_G for any $s \in G$.

Thus, in Definition 2.11, $s \Vdash v\!\downarrow$ is intended to hold only if $s \in G$ guarantees that v will be in I_G. On the other hand, the name V is intended to refer to $V \cap I_G$, which will always exist. The structure of $+, \cdot, \leqslant$ on I_G is inherited from the ground model, which motivates the trivial definition of forcing atomic formulas.

Remark. By Proposition 2.4(iii), whenever an ω^x-large set is split into two subsets, one of them is ω^{x-1}-large. As a consequence, if $s \Vdash v\!\downarrow$, then $s \setminus [1, v]$ is a condition.

Lemma 2.12. The set of conditions Cond and the relations $\trianglelefteq, \Vdash$ of Definition 2.11 determine a forcing interpretation of $\mathcal{L}_2$ in the theory $\mathrm{RCA}_0 + \mathbb{I}$.

Proof. Checking that $\mathrm{Cond}, \trianglelefteq, \Vdash$ determine a forcing translation from $\mathcal{L}_2$ into $\mathcal{L}_2 \cup \{\mathbb{I}\}$ is unproblematic, clause (FI13) of the definition of forcing interpretation trivializes, and most of the other clauses are also very easy to verify in $\mathrm{RCA}_0 + \mathbb{I}$. We discuss (FI9) and (FI11), which are perhaps less obvious than the rest.

To verify (FI9), note that if the set $(x, y]$ is ω-large, then $y \geqslant 2x$, and if it is ω^2-large, then $y > x2^x \geqslant x^2$. Assume that s is ω^x-large for some $x > \mathbb{I}$ and that $s \Vdash v\!\downarrow, s \Vdash w\!\downarrow$ with say $v \leqslant w$. Since $s \cap [1, w]$ is not a condition and neither is $s \cap (w, w^2]$ (being a subset of the set $(w, w^2]$, which is not even ω^2-large), it follows from Proposition 2.4(iii) that $s \cap [1, w^2]$ is not a condition. Thus, $s \Vdash (w^2)\!\downarrow$ and a fortiori s forces both $(v + w)\!\downarrow$ and $(vw)\!\downarrow$. This proves

the "existence" part of (FI9). The "uniqueness" part follows easily from (FT7), (FT6), and the definition of $\Vdash$.

To show (FI11), assume $s \nVdash v\!\downarrow$. Then $s' := s \cap [1, v]$ is a condition with $s' \trianglelefteq s$. However, for any condition $s'' \trianglelefteq s'$, we have $s'' \cap [1, v] = s''$, so certainly $s'' \nVdash v\!\downarrow$. $\qquad\square$

We can use the proof of Lemma 2.12 to show that, actually, a slightly stronger version of (FI9) holds.

Lemma 2.13. There is a polynomial-time procedure which, given a term $t(\bar{v})$, outputs an $\mathrm{RCA}_0 + \mathbb{I}$ proof of the statement that, for any condition s, if $s \Vdash \bar{v}\!\downarrow$, then $s \Vdash t\!\downarrow$.

Proof. Using a standard argument, we can prove in RCA_0 that $t < (\max \bar{v})^n$, where n is the number of symbols in t. By the proof of Lemma 2.12, we can also prove in $\mathrm{RCA}_0 + \mathbb{I}$ that for any s and w, if $s \Vdash w\!\downarrow$, then $s \Vdash (w^2)\!\downarrow$. Iterating this reasoning $\log n$ times, we prove that if $s \Vdash w\!\downarrow$, then $s \Vdash (w^n)\!\downarrow$. We apply this to $w := \max \bar{v}$ in order to deduce $s \Vdash (\max \bar{v})^n\!\downarrow$ and hence $s \Vdash t\!\downarrow$. $\qquad\square$

Lemma 2.14. There is a polynomial-time procedure which, given a Δ_0 formula $\theta(\bar{v}, \bar{V})$, outputs an $\mathrm{RCA}_0 + \mathbb{I}$ proof of

$$\forall s \, \forall \bar{v} \, \forall \bar{V} \, [s \Vdash \bar{v}\!\downarrow \to (\theta(\bar{v}, \bar{V}) \leftrightarrow (s \Vdash \theta(\bar{v}, \bar{V})))].$$

Proof. The construction of the proof proceeds by induction on the structure of θ. As the reader will be able to check, it can be carried out in polynomial-time. We give an informal description of how the proof is built for the atomic step and for each case of the inductive step.

Suppose θ is atomic. If θ is a simple formula, then the equivalence between θ and $s \Vdash \theta$ follows immediately from the definition of our forcing relation. Otherwise, if θ is say $t_1(\bar{v}) \leqslant t_2(\bar{v})$ (the other cases are similar or easier), then θ^* is defined to be

$$\forall x_1 \, \forall x_2 \, (t_1[x_1] \to (t_2[x_2] \to x_1 \leqslant x_2)),$$

(see Definition 1.2) and by (FT8) we have to prove that a condition forces θ^* exactly if it forces $\bar{v}\!\downarrow$ and θ in fact holds. To achieve this, one first shows by induction on subterms $t'(\bar{u})$ of t_1 and t_2 (where $\bar{u}$ is a subtuple of $\bar{v}$) that, for the appropriate variable x' and any $s \in \mathrm{Cond}$,

$$s \Vdash t'[x'] \quad \text{iff} \quad (s \Vdash \bar{u}\!\downarrow \wedge x' = t(\bar{u})). \tag{3}$$

The inductive step is proved using (FT7), (FT6), the definition of $\Vdash$, and Lemma 2.13. Once (3) is proved for t' equal to t_1 and t_2, one can prove the equivalence of $s \Vdash \theta^*$ with $(s \Vdash \bar{v}\!\downarrow \wedge t_1 \leqslant t_2)$ by one more series of appeals to (FT7), (FT6), the definition of $\Vdash$, and Lemma 2.13.

Suppose θ is $\neg \eta$. The inductive assumption gives us a proof that whenever $s \Vdash \bar{v}\!\downarrow$, then $s \Vdash \eta(\bar{v}, \bar{V})$ is equivalent to $\eta(\bar{v}, \bar{V})$. This equivalence generalizes to all $s' \trianglelefteq s$, since for each such s' we have $s' \Vdash \bar{v}\!\downarrow$ as well.

So, let $s \Vdash \bar{v}\!\downarrow$. Assuming $\neg \eta$, we have $s' \nVdash \eta$ for any $s' \trianglelefteq s$, and thus $s \Vdash \neg \eta$. On the other hand, if $s \Vdash \neg \eta$, then by the definition of forcing $s \nVdash \eta$ and hence $\neg \eta$.

Suppose θ is $\eta(\bar{v}, \bar{V}) \to \xi(\bar{v}, \bar{V})$. The inductive assumption gives us a proof that whenever $s \Vdash \bar{v}{\downarrow}$ and $s' \trianglelefteq s$, then $s \Vdash \eta(\bar{v}, \bar{V})$ is equivalent to $\eta(\bar{v}, \bar{V})$ and likewise for ξ.

Let $s \Vdash \bar{v}{\downarrow}$. Assume $\eta \to \xi$. If $s' \trianglelefteq s$ forces η, then η holds and by our assumption so does ξ, which implies $s' \Vdash \xi$. This shows $s \Vdash \eta \to \xi$. Conversely, assume $\eta \wedge \neg\xi$. Then $s \Vdash \eta$ but for each $s' \trianglelefteq s$ we have $s' \nVdash \xi$. So, $s \nVdash \eta \to \xi$.

Suppose $\theta(\bar{v}, \bar{V})$ is $\forall w{\leqslant}t\, \eta(w, \bar{v}, \bar{V})$. Recall that $\forall w{\leqslant}t \ldots$ is shorthand for $\forall w\,(w \leqslant t \to \ldots)$. (Here $t = t(\bar{v})$ is a term in the variables $\bar{v}$.) The atomic step and the inductive assumption, respectively, give us proofs that if $s \Vdash \bar{v}{\downarrow}$ and $s \Vdash w{\downarrow}$, then

$$s \Vdash w \leqslant t \text{ is equivalent to } w \leqslant t,$$
$$s \Vdash \eta(w, \ldots) \text{ is equivalent to } \eta(w, \ldots).$$

As in the previous cases, the equivalences generalize to all $s' \trianglelefteq s$.

Moreover, by Lemma 2.13, we have a polynomial-time constructible proof that if $s \Vdash \bar{v}{\downarrow}$, then $s' \trianglelefteq s$ implies $s' \Vdash t(\bar{v}){\downarrow}$ and thus $s' \Vdash w{\downarrow}$ for any $w \leqslant t(\bar{v})$.

Let $s \Vdash \bar{v}{\downarrow}$ and assume $\forall w{\leqslant}t\, \eta(w, \ldots)$. Take $s' \trianglelefteq s$ and w such that $s' \Vdash w{\downarrow}$ and $s' \Vdash w \leqslant t$. Then the atomic step gives $w \leqslant t$, which, by our assumption, implies $\eta(w, \ldots)$. Thus $s' \Vdash \eta(w, \ldots)$, which is what we want.

Conversely, assume $w \leqslant t$ is such that $\neg\eta(w, \ldots)$. Then $s \Vdash w \leqslant t$. Also, for each $s' \trianglelefteq s$ we have $s' \Vdash w{\downarrow}$, so $s' \nVdash \eta(w, \ldots)$. Thus, $s \nVdash \forall w{\leqslant}t\, \eta(w, \ldots)$. □

Lemma 2.15. The relations Cond, $\trianglelefteq$, $\Vdash$ of Definition 2.11 determine a polynomial forcing interpretation of $\mathrm{WKL}_0 + \mathrm{RT}_2^2$ in $\mathrm{RCA}_0 + \mathbb{I}$.

Proof. We have already shown in Lemma 2.12 that Cond, $\trianglelefteq$, $\Vdash$ determine a forcing interpretation of $\mathcal{L}_2$ in $\mathrm{RCA}_0 + \mathbb{I}$. So if we want to argue that this is actually a forcing interpretation of $\mathrm{WKL}_0 + \mathrm{RT}_2^2$, we only have to check (FI13). In other words, we must show that any condition s forces each of the axioms of $\mathrm{WKL}_0 + \mathrm{RT}_2^2$. This is immediate for the axiomatization of non-negative parts of discrete ordered rings, so it remains to deal with Δ_1^0 comprehension, Σ_1^0 induction, Weak König's Lemma, and RT_2^2.

We reason in $\mathrm{RCA}_0 + \mathbb{I}$, sometimes making tacit use of the following simple observation: for any condition s and any k, a condition $s' \subseteq s$ with $s' \Vdash k{\downarrow}$ exists exactly if $s \setminus [1, k]$ is a condition.

We first deal with the Σ_1^0 separation scheme, which implies both Δ_1^0 comprehension and Weak König's Lemma. Consider a Σ_1^0 separation axiom:

$$\forall \bar{V} \, \forall \bar{v} \, [\forall k\, (\sigma_1(k, \bar{v}, \bar{V}) \to \neg\sigma_2(k, \bar{v}, \bar{V}))$$
$$\to \exists W \, \forall k\, ((\sigma_1(k, \bar{v}, \bar{V}) \to k \in W) \wedge (k \in W \to \neg\sigma_2(k, \bar{v}, \bar{V})))],$$

where $\sigma_1 := \exists \ell\, \eta_1(\ell, k, \ldots)$ and $\sigma_2 := \exists \ell\, \eta_2(\ell, k, \ldots)$ are Σ_1^0. Let s be a condition and assume that s forces $\bar{V}{\downarrow}, \bar{v}{\downarrow}$ and that

$$s \Vdash \forall k\, (\sigma_1(k, \bar{v}, \bar{V}) \to \neg\sigma_2(k, \bar{v}, \bar{V})). \tag{4}$$

By the definition of forcing and Lemma 2.14, (4) implies the following: for every $k, \ell_1 \leqslant \max s$ such that $s \setminus [1, \max(k, \ell_1)]$ is a condition, if $\eta_1(\ell_1, k, \ldots)$ holds,

then there is no $\ell_2 \leqslant \max s$ such that $\eta_2(\ell_2, k, \ldots)$ and $s \setminus [1, \ell_2]$ is a condition. In particular, it cannot be the case that $\ell_1 \geqslant \ell_2$. Thus, taking

$$W_0 := \{k \leqslant \max s : \exists \ell_1 \leqslant \max s \, (\eta_1(\ell_1, k, \ldots) \wedge \forall \ell_2 < \ell_1 \, \neg \eta_2(\ell_2, k, \ldots))\},$$

we conclude that $s \Vdash \forall k \, ((\sigma_1(k, \ldots) \to k \in W_0) \wedge (k \in W_0 \to \neg \sigma_2(k, \ldots)))$.

We now turn to Σ_1^0 induction. Consider the induction axiom for the Σ_1^0 formula $\sigma := \exists \ell \, \eta(\ell, k, \bar{v}, \bar{V})$ with respect to the variable k. Let s be a condition forcing $\bar{v}{\downarrow}, \bar{V}{\downarrow}$ and let j be such that $s \setminus [1, j]$ is a condition. Let $m = \min(s \setminus [1, j])$ and let $x > \mathbb{I}$ be such that $s \setminus [1, j]$ is ω^x-large. It follows that $s \setminus [1, m]$ is $\omega^{x-1}m$-large, which by Proposition 2.4(ii) means that we can write $s \setminus [1, m]$ as a disjoint union

$$s \setminus [1, m] = s_0 \sqcup \cdots \sqcup s_{m-1},$$

where each set s_i is ω^{x-1}-large (in particular, it is a condition) and, for each $i < m - 1$, $\max s_i < \min s_{i+1}$. Define:

$$W_0 := \{\ell : \exists k < j \, (\eta(\ell, k, \ldots) \wedge \forall \ell' < \ell \, \neg \eta(\ell', k, \ldots))\}.$$

Since W_0 has at most j elements and $j < m$, the finite pigeonhole principle implies that $W_0 \cap [\min s_{i_0}, \max s_{i_0}] = \emptyset$ for some $i_0 < m$. Note that for every $k < j$ we have

$$\exists \ell < \min s_{i_0} \, \eta(\ell, k, \ldots) \text{ iff } \exists \ell \leqslant \max s_{i_0} \, \eta(\ell, k, \ldots). \tag{5}$$

By bounded induction on k, either $\neg \exists \ell \leqslant \max s_{i_0} \, \eta(\ell, 0, \ldots)$ or there is a maximal $k < j$ such that $\exists \ell \leqslant \max s_{i_0} \, \eta(\ell, k, \ldots)$. By the property of s_{i_0} stated in (5), it follows that s_{i_0} forces one of the three statements $\neg \sigma(0, \ldots)$, $\sigma(j-1, \ldots)$, or $\exists k < j-1 \, (\sigma(k, \ldots) \wedge \neg \sigma(k+1, \ldots))$. In each case, the existence of s_{i_0} implies that s forces induction for σ below j.

Finally, we deal with RT_2^2. Assume that s is a condition such that $s \Vdash f{\downarrow}$ and $s \Vdash f \colon [\mathbb{N}]^2 \to 2$. This means that f is defined on each pair $\langle i, j \rangle$ such that $i < j < \max s$ and $s \setminus [1, j]$ is a condition. Writing $s = s_0 \sqcup s_1$ where $\max s_0 < \min s_1$ and s_0, s_1 are both conditions, we can conclude that f is defined on all arguments below $\max s_0$.

Let $x > \mathbb{I}$ be such that s_0 is ω^x-large. By Theorem 2.6, there exists an $\omega^{x/300}$-large set $s' \subseteq s_0$ which is homogeneous w.r.t f. By (I2), $\mathbb{I}$ is closed under addition, so $x/300 > \mathbb{I}$, and hence s' is a forcing condition.

It is easily verified that s' forces "s' is homogeneous w.r.t f". However, we also have:

$$s' \Vdash \forall k \, \exists \ell \, (\ell > k \wedge \ell \in s').$$

To see this, take a condition $s'' \subseteq s'$ and k such that $s'' \Vdash k{\downarrow}$. Then $s'' \setminus [1, k]$ is a condition, and so is $s'' \setminus [1, \ell]$ for $\ell := \min(s'' \setminus [1, k])$. However, $s'' \setminus [1, \ell] \Vdash \ell{\downarrow}$, and since $\ell > k \wedge \ell \in s'$ is a true Δ_0 statement, Lemma 2.14 implies that it is forced by $s'' \setminus [1, \ell]$ as well.

This completes the proof that $\mathrm{Cond}, \trianglelefteq, \Vdash$ give a forcing interpretation of $\mathrm{WKL}_0 + \mathrm{RT}_2^2$ in $\mathrm{RCA}_0 + \mathbb{I}$. Polynomiality of the interpretation is immediate if we assume that $\mathrm{WKL}_0 + \mathrm{RT}_2^2$ is finitely axiomatized. We may make this assumption w.l.o.g. because both WKL and RT_2^2 are single axioms, while RCA_0 can be axiomatized by using a finite number of instances of Δ_1^0 comprehension and Σ_1^0 induction in such a way that a proof of any of the other instances can be constructed in polynomial time. $\qquad\square$

Lemma 2.16. The forcing interpretation of $\mathrm{WKL}_0 + \mathrm{RT}_2^2$ in $\mathrm{RCA}_0 + \mathbb{I}$ given by Cond, $\trianglelefteq$, $\Vdash$ of Definition 2.11 is polynomially $\forall \Sigma_2^0$-reflecting.

Proof. Let $\varphi := \exists W \, \exists w \, \forall y \, \exists z \, \theta(W, w, y, z)$ be an $\exists \Pi_2^0$ sentence. We sketch a proof in $\mathrm{RCA}_0 + \mathbb{I}$ that, assuming φ, there is a condition s such that $s \not\Vdash \forall W \, \forall w \, \exists y \, \forall z \, \neg\theta$. On the basis of the sketch, it will be routine to verify that the proof can be constructed in polynomial time on input φ.

Let A, k be such that $\forall y \, \exists z \, \theta(A, k, y, z)$. By Σ_1^0 collection, for each ℓ there exists m such that $\forall y {\leqslant} \ell \, \exists z {\leqslant} m \, \theta(A, k, y, z)$. Use primitive recursion to define a sequence of numbers by:

$$\begin{aligned} k_0 &:= \ k, \\ k_{n+1} &:= \ \text{smallest } m > k_n \text{ such that } \forall y {\leqslant} k_n \, \exists z {\leqslant} m \, \theta(A, k, y, z). \end{aligned}$$

The axioms of RCA_0 imply the existence of the set $S = \{k_n : n \in \mathbb{N}\}$, which is clearly infinite. Use the axiom ($\mathbb{I}$3) of $\mathrm{RCA} + \mathbb{I}$ to obtain a finite set $s \subseteq S$ which is a forcing condition. Below, we abuse notation and write A for the finite set $A{\restriction}_{\max s}$.

We claim that $s \Vdash \forall y \, \exists z \, \theta(A, k, y, z)$, which is enough to imply $s \not\Vdash \forall W \, \forall w \, \exists y \, \forall z \, \neg\theta$. Take $s' \trianglelefteq s$ and ℓ such that $s' \Vdash \ell{\downarrow}$. Since $s' \cap [1, \ell]$ is not a condition, $s' \setminus [1, \ell]$ must be one. Let $m_1 < m_2$ be the two smallest elements of $s' \setminus [1, \ell]$. By Proposition 2.4(iii), $s' \setminus [1, m_2]$ is also a condition and $(s' \setminus [1, m_2]) \Vdash m_2{\downarrow}$. It follows from the definition of s that $\forall y {\leqslant} m_1 \, \exists z {\leqslant} m_2 \, \theta(A, k, y, z)$. In particular, there is some $m \leqslant m_2$ such that $(s' \setminus [1, m_2]) \Vdash m{\downarrow}$ and $\theta(A, k, \ell, m)$. By Lemma 2.14, we get $(s' \setminus [1, m_2]) \Vdash \theta(A, k, \ell, m)$. This implies $(s' \setminus [1, m_2]) \Vdash \exists z \, \theta(A, k, \ell, z)$, which is what we wanted. $\qquad\square$

Proof of Lemma 2.10. This follows directly from Lemma 2.15, Lemma 2.16, and Theorem 1.17. $\qquad\square$

Proof of Theorem 2.1. By Lemma 2.10 there is a polynomial-time procedure which, given a proof π of a $\forall \Sigma_2^0$ sentence ψ in $\mathrm{WKL}_0 + \mathrm{RT}_2^2$, outputs a proof π' of ψ in $\mathrm{RCA}_0 + \mathbb{I}$. By Lemma 2.9, a further polynomial-time procedure outputs a proof π'' of ψ in $\mathrm{RCA}_0 + \mathrm{I}\Sigma_2^0$ and a proof π''' of ψ in $\mathrm{RCA}_0 + \neg\mathrm{I}\Sigma_2^0$. Combine π'', π''', and a case distinction to obtain a proof of ψ in RCA_0. $\qquad\square$

3 Ramsey for pairs: speedup over RCA_0^*

We will now prove that Theorem 2.1 breaks down completely in the absence of Σ_1^0 induction, that is, if the base theory RCA_0 is replaced by RCA_0^*, although $\mathrm{WKL}_0^* + \mathrm{RT}_2^2$ is also $\forall \Sigma_2^0$-conservative over RCA_0^* [28],

Theorem 3.1. RT_2^2 has non-elementary speed-up over RCA_0^* with respect to Σ_1 sentences.

We note that the speedup of RT_2^2 over RCA_0^* can in fact be witnessed by $\Delta_0(\exp)$ sentences, though the proof involves a slightly larger amount of background than for the Σ_1 case. The idea of the argument is explained in the remark after the proof of Theorem 3.1.

The basic reason why Theorem 3.1 holds is expressed in the following lemma. Once the lemma is proved, the upper and lower bounds used to derive the

theorem are obtained by more or less standard techniques described e.g. in the survey [23].

Lemma 3.2. $\mathrm{RCA}_0^* + \mathrm{RT}_2^2$ proves the following statement: "for every k, if each infinite set has a finite subset with k elements, then each infinite set has a finite subset with 2^k elements".

Proof. Clearly, it is enough to prove the statement with $2^{k/2}$ substituted for 2^k. Working in $\mathrm{RCA}_0^* + \mathrm{RT}_2^2$, let k be such that the infinite set A does not have a subset with $2^{k/2}$ elements. W.l.o.g., we may assume that $0 \in A$. We will use the fact that there is a 2-colouring of $[2^{k/2}]^2$ with no homogeneous set of size k, which has a well-known probabilistic proof that easily formalizes in RCA_0^*, to define a 2-colouring of $[\mathbb{N}]^2$ such that every infinite homogeneous sets for the new colouring fails to have a k-element subset.

So, let $f \colon [2^{k/2}]^2 \to 2$ have no homogeneous set of size k. Let $\{a_i : i \in I\}$, for some cut I, be an increasing enumeration of A. Note that $2^{k/2} > I$. Define $g : [\mathbb{N}]^2 \to 2$ as follows:

$$g(x,y) = \begin{cases} b & \text{if there are } i < j \text{ s.t. } x \in [a_i, a_{i+1}), y \in [a_j, a_{j+1}), f(i,j) = b, \\ 0 & \text{if there is } i \text{ s.t. } x, y \in [a_i, a_{i+1}). \end{cases}$$

Let H be an infinite g-homogeneous set. By possibly thinning out H, we may assume w.l.o.g. that for each $i \in I$, there is at most one $x \in H \cap [a_i, a_{i+1})$. This means that for any two $x, y \in H$, the value $g(x, y)$ is determined by the first clause of the definition of g.

Let $x_1, \ldots, x_\ell$ be the first ℓ elements of H, and let $i_1, \ldots, i_\ell$ be such that $x_j \in [a_{i_j}, a_{i_j+1})$. By our choice of H and the definition of g, we see that $f(i_j, i_{j'}) = g(x_j, x_{j'})$, so the set $\{i_1, \ldots, i_\ell\}$ is homogeneous for f. So $\ell < k$ by our choice of f. $\qquad\square$

Consider the simple variant of the iterated exponential function defined by $2_0 = 1$ and $2_{n+1} = 2^{2_n}$ for $n \in \mathbb{N}$. To prove Theorem 3.1, we will make use of a family of sentences stating the existence of $2_{2_{2_n}}$, for n a natural number. The relation $y = 2_x$ has a Δ_0 definition which is well-behaved in $\mathrm{I}\Delta_0$ — see [10, §3.1], or see [15, Chapt. V.3(c)] for a similar idea applied to the more difficult case of $y = 2^x$). Thus, $\exists y \, (y = 2_{2_{2_n}})$ can be expressed as a Σ_1 sentence of size $O(\log n)$, for instance, by substituting the numeral for n into the fixed formula $\exists y \, (y = 2_{2_{2_x}})$.

Lemma 3.3. In $\mathrm{RCA}_0^* + \mathrm{RT}_2^2$, the sentence $\exists y \, (y = 2_{2_{2_n}})$ has a proof of size polynomial in n.

Proof. Consider the definable set

$$I = \{k : \text{every infinite set has a finite subset with } 2_k \text{ elements}\}.$$

In the absence of Σ_1^0 induction, this is a proper initial segment of $\mathbb{N}$ and, in general, not a set (i.e. a second-order object). However, Lemma 3.2 implies that I is a definable cut provably in $\mathrm{RCA}_0^* + \mathrm{RT}_2^2$.

Consider also the set

$$A = \{\ell : \exists k \, (\ell = 2_k)\}.$$

20

Provably in RCA_0^*, this is indeed a set, because the $\exists k$ quantifier can be bounded by ℓ. By the totality of the exponential function, A is infinite. Moreover, since $k \mapsto 2_k$ is an increasing monotone operation with a Δ_0-definable graph, A has a subset with k elements exactly for those k for which 2_k exists. It follows that for each k, if $I(k)$, then 2_{2_k} exists.

Now fix a natural number n. By [23, Theorem 3.4.1], there are formulas $I_0, I_1, \ldots, I_n$ such that I_0 is I and for each $j < n$, the theory $\mathrm{RCA}_0^* + \mathrm{RT}_2^2$ has a poly(n)-size proof that I_{j+1} is a cut and that $\forall k\, (I_{j+1}(k) \to I_j(2^k))$. By putting these proofs together, we get a poly(n)-size proof of $I(2_n)$ and hence of the statement that $2_{2_{2_n}}$ exists. $\qquad\square$

Lemma 3.4. The size of the smallest RCA_0^* proof of $\exists y\, (y = 2_{2_{2_n}})$ grows faster than any elementary recursive function of n.

Proof. In view of Theorem 4.9 in the next section, it suffices to show that the size of the shortest $\mathrm{I}\Delta_0 + \exp$ proof of $\exists y\, (y = 2_{2_{2_n}})$ grows nonelementarily in n. To this end, it is clearly enough to show the existence of a polynomial p such that, if the sentence $\exists z\, \varphi(z)$ for a bounded formula φ has a proof in $\mathrm{I}\Delta_0 + \exp$ of size n, then $\exists z \leqslant 2_{p(n)}\, \varphi(z)$ holds in ω.

So, assume that $\mathrm{I}\Delta_0 + \exp$ has a size-n proof of $\exists z\, \varphi(z)$ for φ bounded. We may assume that $\mathrm{I}\Delta_0 + \exp$ is finitely axiomatized, because it has a finite axiomatization that polynomially simulates the more usual one in which Δ_0 induction is a scheme. By standard upper bounds on the cut elimination procedure and its connection to Herbrand's theorem [23, Section 5], there is a fixed polynomial p such that some Herbrand disjunction for $\neg(\mathrm{I}\Delta_0 + \exp) \vee \exists z\, \varphi(z)$ of size at most $2_{p(n)}$ is logically valid.

In more detail, what this means is as follows. Consider the conjunction of:

- $\forall v_1 \ldots \forall v_k\, \delta(v_1, \ldots, v_k)$, a statement that induction holds for finitely many fixed Δ_0 formulas with parameters among the $\bar{v}$; we may assume w.l.o.g. that all the quantifiers in δ are bounded by one of the v_i,

- $\forall x\, \exists y\, (y = 2^x)$, with $y = 2^x$ a Δ_0 formula in the language of first-order arithmetic,

- $\forall z\, \neg\varphi(z)$.

Put this conjunction in prenex normal form and skolemize it, obtaining the sentence $\forall \bar{v}\, \forall x\, \forall z\, \forall \bar{w}\, \xi(\bar{v}, x, z, \bar{w})$, where the $\forall \bar{w}$ quantifiers are bounded and ξ is quantifier-free (but contains the Skolem function symbols corresponding to the original existential quantifiers). Then, for some s, there is a list of closed terms

$$t_{v_1,1}, \ldots, t_{v_k,1}, t_{x,1}, t_{z,1}, t_{w_1,1}, \ldots, t_{w_\ell,1}, \ldots, t_{v_1,s}, \ldots, t_{v_k,s}, t_{x,s}, t_{z,s}, t_{w_1,s}, \ldots, t_{w_\ell,s}$$

(of the arithmetical language extended by the Skolem functions) such that the conjunction

$$\Xi := \bigwedge_{i=1}^{s} \xi(t_{\bar{v},i}, t_{x,i}, t_{z,i}, t_{\bar{w},i})$$

has size at most $2_{p(n)}$ and is unsatisfiable.

We now interpret the function symbols appearing in Ξ as operations on ω. The symbols $+, \cdot, 0, 1$ are interpreted in the usual way. If $f(\ldots, x, \ldots)$ is

21

the Skolem function symbol corresponding to the $\exists y$ quantifier (f may have more arguments as a result of the move to prenex normal form), then we put $f(\ldots, m, \ldots) = 2^m$. If f is the Skolem function symbol corresponding to a bounded existential quantifier $\exists u$ in front of a subformula ψ of δ, $\neg\varphi$, or $y = 2^x$, we let f pick out the smallest witness to $\exists u \, \psi$ on those arguments for which there is such a witness; otherwise, f returns 0.

Since each function interpreting a symbol in Ξ increases its arguments at most exponentially, and Ξ has size at most $2_{p(n)}$, it follows that under our interpretation each term in Ξ has a value bounded by $2_{2_{p(n)}}$. Since we interpreted the Skolem function symbols as actual Skolem functions, and $I\Delta_0 + \exp$ is a true theory, each substitution instance of the skolemizations of $\forall v_1 \ldots \forall v_k \, \delta(v_1, \ldots, v_k)$ and $\forall x \, \exists y \, (y = 2^x)$ appearing in Ξ is true under our interpretation. However, Ξ is unsatisfiable, which implies that some substitution instance of the skolemization of $\forall z \, \neg\varphi(z)$ must be false. Due to the way we defined our interpretation, this means that $\exists z \leqslant 2_{2_{p(n)}} \, \varphi(z)$ holds. $\qquad\square$

Proof of Theorem 3.1. Immediate from Lemmas 3.3 and 3.4. $\qquad\square$

Remark. As already mentioned, it is possible to witness the speedup of $RCA_0^* + RT_2^2$ over RCA_0^* by a family of $\Delta_0(\exp)$ sentences. The sentences in question take the form $\mathrm{Con}_{2_n}(I\Delta_0 + \exp)$, where $\mathrm{Con}_x(T)$ is a formula stating in a natural way that there is no proof of contradiction in the theory T of size at most x. Clearly, $\mathrm{Con}_{2_n}(I\Delta_0 + \exp)$ can be expressed by a $\Delta_0(\exp)$ sentence of size polynomial in n.

The so-called finitistic Gödel theorem, proved independently by Friedman and Pudlák (see [23, Theorem 6.3.2] and the references therein), implies that for a sufficiently strong finitely axiomatized T, the size of the smallest proof of $\mathrm{Con}_n(T)$ in T is $n^{\Omega(1)}$. As a consequence, the size of the smallest proof of $\mathrm{Con}_{2_n}(I\Delta_0 + \exp)$ in $I\Delta_0 + \exp$ is $(2_n)^{\Omega(1)}$. By Theorem 4.9, a proof of this sentence in RCA_0^* also requires size $(2_n)^{\Omega(1)}$.

On the other hand, essentially by formalizing the proof of Lemma 3.4 but with $\forall z \, \neg\varphi(z)$ replaced by $0 = 0$, one can prove in $I\Delta_0 + \exp$ that, for a certain fixed polynomial p, the existence of $2_{2_{p(x)}}$ implies $\mathrm{Con}_x(I\Delta_0 + \exp)$. By Lemma 3.3, $RCA_0^* + RT_2^2$ can prove the existence of $2_{2_{2_{p(n)}}}$, a number greater than $2_{2_{p(2_n)}}$, in size $\mathrm{poly}(n)$. Thus, it can also prove $\mathrm{Con}_{2_n}(I\Delta_0 + \exp)$ in size $\mathrm{poly}(n)$.

4 Induction versus collection

Since RT_2^2 proves the collection principle $B\Sigma_2$, Corollary 2.2 implies that $I\Sigma_1$ polynomially simulates $B\Sigma_2$ w.r.t. proofs of Π_3 sentences. In this context, it is natural to ask whether the Π_{n+2}-conservativity of $B\Sigma_{n+1}$ over $I\Sigma_n$ is also witnessed by a polynomial simulation for $n \neq 1$. A related problem was raised by Clote et al. [7, page 216]: is the conservativity provable in bounded arithmetic? Note that if a Π_2 sentence of the form $\forall x \, \exists y \, \delta(x, y)$, such as a conservativity statement, is provable in bounded arithmetic, then the least witness for the $\exists y$ quantifier can be bounded by a polynomial-time function of x.

In this section, we present, for each $n \in \mathbb{N}$, a polynomial forcing interpretation $\tau(n)$ of $B\Sigma_{n+1} + \exp$ in $I\Sigma_n + \exp$ that is polynomially Π_{n+2}-reflecting.

The existence of these forcing interpretations, of course, implies that $I\Sigma_n$ polynomially simulates $B\Sigma_{n+1}$ w.r.t. Π_{n+2} sentences whenever $n \geqslant 1$; for $n = 0$, we get a polynomial simulation of $B\Sigma_1 + \exp$ by $I\Delta_0 + \exp$. Our interpretations can be formalized in (a fragment of) bounded arithmetic, which partially solves the problem from [7] as well: completely for $n \geqslant 1$, and over exp for $n = 0$.

The only role of exp is to provide us with the usual universal Σ_{n+1} formula $\Sigma_{n+1}\text{-Sat}(v, x)$. Recall that $I\Delta_0 + \exp$ proves $\theta(\bar{x}) \leftrightarrow \Sigma_{n+1}\text{-Sat}(\underline{\theta}, \langle \bar{x} \rangle)$ for $\theta \in \Sigma_{n+1}$, where $\underline{\theta}$ denotes the numeral for (the code/Gödel number of) θ and $\langle \bar{x} \rangle$ denotes the code for the finite sequence $\langle x_1, \dots, x_l \rangle$. It also proves that $\Sigma_{n+1}\text{-Sat}(v, x)$ satisfies the usual recursive conditions for a satisfaction relation restricted to Σ_{n+1} formulas; cf. [15, Chapts. I.1(d), V.5(b)].

Remark. The question from [7] as stated concerned provability in $I\Delta_0 + \Omega_1$, which coincides with the $\mathcal{L}_1$-consequences of Buss' theory S_2. The related Question 35 from the Clote-Krajíček list of open questions [9] — whether the stronger theory $I\Delta_0 + \exp$ proves the mere equiconsistency of $B\Sigma_{n+1}$ and $I\Sigma_n$ — was already answered positively in Hájek's paper [14, Section 2] in the same volume where the list appeared.

Remark. While we were writing up our paper, we learnt that Fedor Pakhomov [private communication] independently proved that $I\Sigma_n$ polynomially simulates $B\Sigma_{n+1}$ w.r.t. Π_{n+2} (and that bounded arithmetic proves this) by a different argument that avoids the need for exp in the case $n = 0$. Moreover, in an email discussion with us, Pakhomov found a parametric interpretation of $B\Sigma_{n+1} + \exp + \neg\sigma$ in $I\Sigma_n + \exp + \neg\sigma$ for each Π_{n+2} sentence σ, at least for $n > 0$. It would be interesting to know whether his interpretation can be made independent of the Π_{n+2} sentence in question, and whether the use of parameters can be avoided.

The model-theoretic idea behind our forcing interpretations is simple and appeared implicitly already in [7, Lemma 12]: given $M \models I\Sigma_n + \exp$, if we construct a non-cofinal extension $K \succcurlyeq_{\Sigma_{n+1}} M$, then $M \preccurlyeq_{\Sigma_{n+1}} \sup_K M \models B\Sigma_{n+1} + \exp$. If the aim is only to show the Π_{n+2}-conservativity of $B\Sigma_{n+1} + \exp$ over $I\Sigma_n + \exp$, one can build the extension K by any method, but to get a forcing interpretation a forcing construction is needed. The forcing construction we choose is that of a generic Σ_{n+1} ultrapower [20, pages 181f.]; cf. the proof of Theorem B in [21]. In particular, the construction employs a maximal filter $\mathcal{V}$ in the lattice consisting of the Σ_{n+1}-definable subsets of M. Elements of the Σ_{n+1} ultrapower K are represented by Σ_{n+1}-definable partial functions $M \rightharpoonup M$ whose domain is an element of $\mathcal{V}$. To ensure that the ultrapower is a non-cofinal extension, we require every element of $\mathcal{V}$ to be cofinal in M. It would be possible to first define a forcing relation $\Vdash'_{\tau(n)}$ for the ultrapower K and then define a second forcing relation $\Vdash_{\tau(n)}$ for the truncation $\sup_K M$ in terms of $\Vdash'_{\tau(n)}$. To save some notation, we introduce the second one directly.

An alternative forcing interpretation, much closer in spirit to the one used to prove Theorem 2.1 and therefore perhaps easier to follow for the reader acquainted with Section 2, is mentioned in a remark after the proof of Theorem 4.6. However, that interpretation only works for $n > 0$.

In the remainder of this section, $\exists^\infty x$ ("for infinitely many x" or "for cofinally many x") stands for $\forall y\, \exists x \geqslant y$ and $\forall^\infty x$ ("for almost all x") stands for $\exists y\, \forall x \geqslant y$.

For $n \in \mathbb{N}$, we will define a forcing translation $\tau(n)$ from the language of first-order arithmetic to itself. The definition of the set of conditions is relatively

straightforward.

Definition 4.1. Let $n \in \mathbb{N}$, and let $x \in \mathrm{Ext}_{n+1}(s)$ stand for $\Sigma_{n+1}\text{-Sat}(s, \langle x \rangle)$. Then a (code for a) Σ_{n+1} formula with one free variable s is in $\mathrm{Cond}_{\tau(n)}$ if $\exists^{\infty} x \, (x \in \mathrm{Ext}_{n+1}(s))$. For two conditions s, s', we let $s' \trianglelefteq_{\tau(n)} s$ if $\mathrm{Ext}_{n+1}(s') \subseteq \mathrm{Ext}_{n+1}(s)$.

In other words, conditions are infinite Σ_{n+1}-definable sets. Our names will be (codes for) Σ_{n+1} formulas with two variables. The definition of $s \Vdash_{\tau(n)} v \downarrow$ is somewhat subtle, but the idea is that names are to be viewed as Σ_{n+1}-definable functions whose ranges are bounded. To capture this intuition, we introduce some auxiliary concepts. We will write $v(x) \downarrow$ if $\exists! y \, \Sigma_{n+1}\text{-Sat}(v, \langle x, y \rangle)$, and simply write $v(x)$ for the unique such y. We write $s \Vdash^{+}_{\tau(n)} v \downarrow$ if there exists some d such that $\forall x {\in} \mathrm{Ext}_{n+1}(s) \, (v(x) \downarrow \wedge v(x) \leqslant d)$. Note that if $v_1, \ldots, v_\ell$ are names, then under the assumption $v_1(x) \downarrow, \ldots, v_\ell(x) \downarrow$ the assertion $\theta(\bar{v}(x))$ can be formulated either as

$$\exists y_1, y_2, \ldots, y_\ell \, \Big(\bigwedge_{i=1}^{\ell} \Sigma_{n+1}\text{-Sat}(v_i, \langle x, y_i \rangle) \wedge \theta(\bar{y}) \Big)$$

or as

$$\forall y_1, y_2, \ldots, y_\ell \, \Big(\bigwedge_{i=1}^{\ell} \Sigma_{n+1}\text{-Sat}(v_i, \langle x, y_i \rangle) \to \theta(\bar{y}) \Big).$$

For $\theta \in \Sigma_n \cup \Pi_n$, these statements are Σ_{n+1} and Π_{n+1}, respectively. Below, we sometimes write $\bar{v}(x) \downarrow$ for the conjunction of $v_1(x) \downarrow, \ldots, v_\ell(x) \downarrow$.

The actual definition of $s \Vdash_{\tau(n)} v \downarrow$ says that $s' \Vdash^{+}_{\tau(n)} v \downarrow$ happens densely below s.

Definition 4.2. For a condition s and a name v, we say that $s \Vdash_{\tau(n)} v \downarrow$ if the following hold:

(i) $\forall^{\infty} x {\in} \mathrm{Ext}_{n+1}(s) \, \exists^{\leqslant 1} y \, \Sigma_{n+1}\text{-Sat}(v, \langle x, y \rangle)$,

(ii) $\exists d \, \forall^{\infty} x {\in} \mathrm{Ext}_{n+1}(s) \, \forall y \, (\Sigma_{n+1}\text{-Sat}(v, \langle x, y \rangle) \to y \leqslant d)$,

(iii) there is no $s' \trianglelefteq_{\tau(n)} s$ such that $\forall^{\infty} x {\in} \mathrm{Ext}_{n+1}(s') \, \forall y \, \neg \Sigma_{n+1}\text{-Sat}(v, \langle x, y \rangle)$.

For a simple atomic formula $\alpha(v_1, \ldots, v_\ell)$, we say that $s \Vdash_{\tau(n)} \alpha(\bar{v})$ if $s \Vdash_{\tau(n)} \bar{v} \downarrow$ and

$$\forall^{\infty} x {\in} \mathrm{Ext}_{n+1}(s) \, \Big(\bigwedge_{i=1}^{\ell} v_i(x) \downarrow \to \alpha(v_1(x), \ldots, v_\ell(x)) \Big).$$

When there is no risk of ambiguity, we will often omit the subscript $\tau(n)$.

Lemma 4.3. For each $n \in \mathbb{N}$, the relations $\mathrm{Cond}_{\tau(n)}, \trianglelefteq_{\tau(n)}, \Vdash_{\tau(n)}$ determine a forcing interpretation of $\mathcal{L}_1$ in $\mathrm{I}\Sigma_n + \exp$. Moreover, there is a polynomial-time procedure which, given $n \in \mathbb{N}$ in unary and an instance of (FI0)–(FI12) for $\tau(n)$, outputs a proof of that instance in $\mathrm{I}\Sigma_n + \exp$.

In this and later similar statements on the existence of polynomial-time procedures that output some proofs, the dependence of the running time on n rather than just on the formula to be proved cannot be avoided: it comes for instance from the universal Σ_{n+1} formula, which is already used in the definition of Ext_{n+1}, and thus of $\mathrm{Cond}_{\tau(n)}$ and $\Vdash_{\tau(n)}$.

Proof. The only conditions that are not completely trivial to prove are (FI11) and (FI12). Below, we think of $n \in \mathbb{N}$ as fixed and explain how to prove (FI11) and (FI12) for $\tau(n)$ in $\mathrm{I}\Sigma_n + \exp$. The verification that the proofs can be constructed in polynomial time is left to the reader.

Condition (FI11) states that if $s' \Vdash v{\downarrow}$ happens densely below s, then in fact $s \Vdash v{\downarrow}$. So assume that $s \nVdash v{\downarrow}$. If item (iii) from the definition of $s \Vdash v{\downarrow}$ is violated, this immediately gives some $s' \trianglelefteq s$ such that $v(x)$ is undefined for each $x \in \mathrm{Ext}_{n+1}(s')$, so clearly $s'' \nVdash v{\downarrow}$ whenever $s'' \trianglelefteq s'$. If (i) is violated, let s' be a (code for a) formula such that $x \in \mathrm{Ext}_{n+1}(s')$ iff $x \in \mathrm{Ext}_{n+1}(s') \wedge \exists^{\geqslant 2} y\, \Sigma_{n+1}\text{-Sat}(v, \langle x, y \rangle)$. Then $s' \in \mathrm{Cond}$, and any $s'' \trianglelefteq s'$ violates (i) with respect to v. Finally, if (ii) is violated but (i) holds, first let b be such that $\exists^{\leqslant 1} y\, \Sigma_{n+1}\text{-Sat}(v, \langle x, y \rangle)$ holds for $x \in \mathrm{Ext}_{n+1}(s)$ with $x > b$. Let $s' \trianglelefteq s$ be such that $x \in \mathrm{Ext}_{n+1}(s')$ iff $x \in \mathrm{Ext}_{n+1}(s) \wedge x > b$. The failure of (ii) means that for each d, there exists $x \in \mathrm{Ext}_{n+1}(s')$ with $x > d$ and $v(x){\downarrow} \wedge v(x) > d$. Let H be the set of all finite sequences of the form $(x_i, y_i, w_i)_{i \leqslant k}$ such that for each $i \leqslant k$, the triple $\langle x_i, y_i, w_i \rangle$ is the smallest one satisfying:

- w_i is a witness for (the outermost existential quantifier block of) the Σ_{n+1} statement $x_i \in \mathrm{Ext}_{n+1}(s') \wedge \Sigma_{n+1}\text{-Sat}(v, \langle x_i, y_i \rangle)$,

- $x_i > \langle x_j, y_j, w_j \rangle$ for each $j < i$,

- $y_i > \langle x_j, y_j, w_j \rangle$ for each $j < i$.

It can be observed that the set H is $\Sigma_0(\Sigma_n)$-definable; in fact, in the presence of $\mathrm{B}\Sigma_n$ it can be defined by the conjunction of a Π_n and a Σ_n formula. Moreover, H has no maximal element, by our assumption on the failure of (ii). Thus, H is cofinal by $\mathrm{I}\Sigma_n$. Now define s'' so that $x \in \mathrm{Ext}_{n+1}(s'')$ iff there is a finite sequence $(x_i, y_i, w_i)_{i \leqslant k}$ in H such that $x = x_k$. Then $s'' \in \mathrm{Cond}$, $s'' \trianglelefteq s'$, and for each d, if $x \in \mathrm{Ext}_{n+1}(s'')$ and $x \geqslant d$, then $v(x') > d$ for each $x' \in \mathrm{Ext}_{n+1}(s'')$ with $x' > x$. So, any $s''' \trianglelefteq s''$ violates (ii) with respect to v.

Once (FI11) is shown, the proof of (FI12) becomes simple. If $s \nVdash \alpha(\bar{v})$, then there are two cases to consider. Either $s \nVdash \bar{v}{\downarrow}$, and then by the above argument $s' \Vdash \bar{v}{\downarrow}$ does not occur densely below s. Otherwise, we have

$$\exists^{\infty} x {\in} \mathrm{Ext}_{n+1}(s) \left(\bigwedge_{i=1}^{\ell} v_i(x){\downarrow} \wedge \neg\alpha(v_1(x), \ldots, v_\ell(x)) \right).$$

Let b be such that $\exists^{\leqslant 1} y\, \Sigma_{n+1}\text{-Sat}(v, \langle x, y \rangle)$ holds for $x \in \mathrm{Ext}_{n+1}(s)$ with $x > b$. Let $s' \trianglelefteq s$ be such that $x \in \mathrm{Ext}_{n+1}(s')$ iff $x \in \mathrm{Ext}_{n+1}(s) \wedge x > b$. Let $s'' \trianglelefteq s'$ be such that

$$x \in \mathrm{Ext}_{n+1}(s'') \leftrightarrow x \in \mathrm{Ext}_{n+1}(s') \wedge \bigwedge_{i=1}^{\ell} v_i(x){\downarrow} \wedge \neg\alpha(v_1(x), \ldots, v_\ell(x)).$$

Then $s''' \nVdash \alpha(\bar{v})$ whenever $s''' \trianglelefteq s''$. $\qquad\square$

Lemma 4.4. For any $s \in \mathrm{Cond}$ and names $v_1 \ldots, v_\ell$ such that $s \Vdash \bar{v}{\downarrow}$, there exists $\hat{s} \trianglelefteq s$ such that:

- $\hat{s} \Vdash^+ \bar{v}$,

- $\forall^\infty x \in \mathrm{Ext}_{n+1}(s)\,(\bar{v}(x){\downarrow} \to x \in \mathrm{Ext}_{n+1}(\hat{s}))$,

- $s \Vdash \theta(\bar{v}) \leftrightarrow \hat{s} \Vdash \theta(\bar{v})$ for any $\mathcal{L}_1$-formula θ.

Proof. Take any $s \in \mathrm{Cond}$ and $\bar{v}$ such that $s \Vdash \bar{v}{\downarrow}$. By (i) and (ii) of the definition of $s \Vdash \bar{v}{\downarrow}$, let b, d be such that for each $x \in \mathrm{Ext}_{n+1}(s)$ with $x > b$ and each $i = 1, \dots, \ell$,

$$\exists^{\leqslant 1} y\, \Sigma_{n+1}\text{-}\mathrm{Sat}(v_i, \langle x, y \rangle) \wedge \forall y\,(\Sigma_{n+1}\text{-}\mathrm{Sat}(v_i, \langle x, y \rangle) \to y \leqslant d)$$

holds. Let $\hat{s} \trianglelefteq s$ be such that $x \in \mathrm{Ext}_{n+1}(\hat{s}) \leftrightarrow x \in \mathrm{Ext}_{n+1}(s) \wedge x > b \wedge \bar{v}(x){\downarrow}$. Then we have $\forall^\infty x \in \mathrm{Ext}_{n+1}(s)\,(\bar{v}(x){\downarrow} \to x \in \mathrm{Ext}_{n+1}(\hat{s}))$.

We show $s \Vdash \theta(\bar{v}) \leftrightarrow \hat{s} \Vdash \theta(\bar{v})$ for any $\mathcal{L}_1$-formula θ. The implication from left to right is immediate from Lemma 1.11. To show the converse implication, assume that $\hat{s} \Vdash \theta(\bar{v})$ for an $\mathcal{L}_1$-formula θ. By Lemma 1.11, it is enough to show that $\theta(\bar{v})$ is forced dense below s. Let $s' \trianglelefteq s$. Take $s'' \trianglelefteq s'$ so that $x \in \mathrm{Ext}_{n+1}(s'') \leftrightarrow x \in \mathrm{Ext}_{n+1}(s') \wedge x > b \wedge \bar{v}(x){\downarrow}$. Then, $s'' \trianglelefteq \hat{s}$, and hence $s'' \Vdash \theta(\bar{v})$. $\qquad\square$

It is known that Σ_{n+1} ultrapowers of models of $\mathrm{I}\Sigma_n$ satisfy Łoś's Theorem for Σ_{n+1} and Π_{n+1} formulas [20, Lemma 5.13]. This ensures Σ_{n+2}-elementarity between the base model M and the ultrapower K, although, as remarked earlier, Σ_{n+1}-elementarity is already enough to run our argument. We transform these ideas into a syntactic proof in the following.

Lemma 4.5. There is a polynomial-time procedure which, given $n \in \mathbb{N}$ in unary and a formula $\theta(y_1, y_2, \dots, y_\ell)$ in $\Sigma_n \cup \Pi_n$, outputs a proof of the following in $\mathrm{I}\Sigma_n + \exp$:

for any $s \in \mathrm{Cond}$ and names $v_1 \dots, v_\ell$ such that $s \Vdash_{\tau(n)} \bar{v}{\downarrow}$, $s \Vdash_{\tau(n)} \theta(\bar{v}) \leftrightarrow \forall^\infty x \in \mathrm{Ext}_{n+1}(s)\,(\bar{v}(x){\downarrow} \to \theta(\bar{v}(x)))$.

Proof. Like the usual proof of Łoś's Theorem for Σ_{n+1} ultrapowers [20, Lemma 5.13], the construction splits into two parts.

(1) For Δ_0 formulas θ, we build the required proof by induction on the structure of θ.

(2) Beyond that, we build the proof using induction on the number of unbounded quantifiers in θ.

As in the proof of Lemma 4.3, we think of $n \in \mathbb{N}$ as fixed, and leave it to the reader to verify that the proofs in $\mathrm{I}\Sigma_n + \exp$ described below can be found in time polynomial in the given parameters.

In the construction for (1), the base step is for an atomic formula $\theta(\bar{v})$. If θ is simple atomic, the required equivalence follows immediately from the definition of $\Vdash_{\tau(n)}$. For a non-simple atomic θ, the argument is similar to the one in Lemma 2.14. By induction on the structure of a term $t(\bar{u})$ ($\bar{u}$ a subtuple of $\bar{v}$) appearing in θ, we build a proof that, if w is a name, then $s \Vdash t(\bar{u}) = w$ if and only if the following happens: $s \Vdash \bar{u}{\downarrow}$, $s \Vdash w{\downarrow}$, and for almost all $x \in \mathrm{Ext}_{n+1}(s)$, if $\bar{u}(x){\downarrow}$ and $w(x){\downarrow}$, then $t(\bar{u}(x)) = w(x)$. Once this is proved for each term in θ, the required equivalence is proved using (FT7), (FT6), and the definition of $\Vdash_{\tau(n)}$.

The inductive steps of the construction for (1) are more or less based on the usual proof of Łoś's Theorem with the satisfaction relation for the Σ_{n+1} ultrapower replaced by the forcing relation. As a demonstration, we show the case for negation. Lemma 4.4 lets us restrict attention to the case when $s \Vdash^+ \bar{v}{\downarrow}$, which in particular implies $\bar{v}(x){\downarrow}$ for all $x \in \mathrm{Ext}_{n+1}(s)$.

Suppose θ is $\neg\eta$ for $\eta \in \Delta_0$. The inductive assumption gives us a proof that if s and $\bar{v}$ are such that $s \Vdash^+ \bar{v}{\downarrow}$, then $s \Vdash \eta(\bar{v})$ exactly if for almost all $x \in \mathrm{Ext}_{n+1}(s)$ the values of the v_i's at x satisfy η.

Take any $s \in \mathrm{Cond}$ and any $\bar{e}$ such that $s \Vdash^+ \bar{e}{\downarrow}$. If $s \not\Vdash \neg\eta(\bar{e})$, then (FT5) gives us $s' \trianglelefteq s$ such that $s' \Vdash \eta(\bar{e})$, and so, by the proof from the induction hypothesis, $\forall^\infty x{\in}\mathrm{Ext}_{n+1}(s')\,\eta(\bar{e}(x))$, hence also $\neg\forall^\infty x{\in}\mathrm{Ext}_{n+1}(s)\,\neg\eta(\bar{e}(x))$.

Conversely, suppose $s \Vdash \neg\eta(\bar{e})$. By (FT5) and the proof from the induction hypothesis,

$$\forall s'{\trianglelefteq}s \; \exists^\infty x{\in}\mathrm{Ext}_{n+1}(s')\,\neg\eta(\bar{e}(x)).$$

Take a (code for a) Σ_{n+1} formula s' so that $x \in \mathrm{Ext}_{n+1}(s') \leftrightarrow x \in \mathrm{Ext}_{n+1}(s) \wedge \eta(\bar{e}(x))$. By the above, s' cannot be a member of Cond, and thus $\mathrm{Ext}_{n+1}(s')$ is bounded. Hence, we have $\neg\exists^\infty x{\in}\mathrm{Ext}_{n+1}(s)\,\eta(\bar{e}(x))$. So, given that $s \Vdash^+ \bar{e}{\downarrow}$, we have $\forall^\infty x{\in}\mathrm{Ext}_{n+1}(s)\,\neg\eta(\bar{e}(x))$.

If $n = 0$, then (1) already suffices. So suppose $n > 0$. In the construction for (2), we also imitate the usual proof of Łoś's Theorem, but additionally we need to take care of the fact that our anticipated ultrapower is truncated. The bounds involved will be provided by strong Σ_n collection, which is the principle that given a Σ_n formula $\varphi(x,y)$ and a bound a, there is a bound b such that for all $x < a$, if any y satisfying $\varphi(x,y)$ exists, then some such y may be found below b. It is well-known that for $n > 0$ this is equivalent to $\mathrm{I}\Sigma_n$, cf. [15, Theorem I.2.23(a)]. The only part of (2) that requires special attention is the $\to$ direction for the universal quantifier. Again, by Lemma 4.4 we only need to consider the case when $s \Vdash^+ \bar{v}{\downarrow}$.

So, suppose $\theta \in \Sigma_n \cup \Pi_n$ is of the form $\forall y\,\eta(\bar{x},y)$. Note that $\eta \in \Pi_n$ in such a case. Work over $\mathrm{I}\Sigma_n + \exp$. The inductive assumption gives us a proof that if $s,\bar{v},w$ are such that $s \Vdash^+ \bar{v}{\downarrow},w{\downarrow}$, then $s \Vdash \eta(\bar{v},w)$ exactly if for almost all $x \in \mathrm{Ext}_{n+1}(s)$, it holds that $\eta(\bar{v}(x),w(x))$.

Take any $s \in \mathrm{Cond}$ and any $\bar{e}$ such that $s \Vdash^+ \bar{e}{\downarrow}$. Assume

$$\exists^\infty x{\in}\mathrm{Ext}_{n+1}(s)\,\neg\forall y\,\eta(\bar{e}(x),y).$$

Take a (code for a) Σ_{n+1} formula s' so that

$$x \in \mathrm{Ext}_{n+1}(s') \leftrightarrow x \in \mathrm{Ext}_{n+1}(s) \wedge \exists y\,\neg\eta(\bar{e}(x),y).$$

Then, by our assumption, $s' \in \mathrm{Cond}$ and $s' \trianglelefteq s$. Take a (code for a) Σ_{n+1} formula e' with free variables x,z so that

$$\Sigma_{n+1}\text{-}\mathrm{Sat}(e', \langle x,z\rangle) \leftrightarrow x \in \mathrm{Ext}_{n+1}(s') \wedge \neg\eta(\bar{e}(x),z) \wedge \forall z'{<}z\,\eta(\bar{e}(x),z').$$

Since $\mathrm{I}\Sigma_n$ holds and $s' \Vdash^+ \bar{e}$, we know $\forall x{\in}\mathrm{Ext}_{n+1}(s')\,\exists!z\,\Sigma_{n+1}\text{-}\mathrm{Sat}(e', \langle x,z\rangle)$.

Recall that $s \Vdash^+ \bar{e}{\downarrow}$ and $s' \trianglelefteq s$. Find d such that

$$\forall x{\in}\mathrm{Ext}_{n+1}(s') \bigwedge_{i=1}^{\ell} e_i(x) \leqslant d.$$

Use strong Σ_n collection to obtain d' such that

$$\forall \bar{y}{\leqslant}d \left(\exists z \, \neg\eta(\bar{y},z) \rightarrow \exists z{\leqslant}d' \, \neg\eta(\bar{y},z) \right),$$

which ensures

$$\forall x{\in}\mathrm{Ext}_{n+1}(s') \, \exists z{\leqslant}d' \, \Sigma_{n+1}\text{-}\mathrm{Sat}(e', \langle x, z \rangle).$$

Thus $s' \Vdash^{+} \bar{e}'{\downarrow}$, and the proof from the induction hypothesis gives us $s' \Vdash \neg\eta(\bar{e}, \bar{e}')$. Hence (FT7) tells us $s \nVdash \forall y \, \eta(\bar{e}, y)$. $\qquad\square$

Remark. For $n > 0$, the argument in the proof above relies on induction axioms in two places. Speaking in model-theoretic terms, first we use at most $\mathrm{I}\Sigma_{n-1}$ to guarantee the existence of sufficiently many Skolem functions when proving Łoś's Theorem for the ultrapower K, and then we use strong Σ_n collection to pass on the elementarity from K to $\sup_K M$. For $n = 0$, the latter application is not needed (and we do not have strong collection anyway), but we use full $\mathrm{I}\Sigma_n$ to get the Łoś Theorem we want.

We can now complete the argument that $\tau(n)$ is a forcing interpretation of $\mathrm{B}\Sigma_{n+1} + \exp$ in $\mathrm{I}\Sigma_n + \exp$. In the usual argument proving Π_{n+1} conservativity of $\mathrm{B}\Sigma_{n+1}$ of $\mathrm{I}\Sigma_n$ [15, page 229] it is easy to deduce from the analogue of Lemma 4.5 that the truncated ultrapower $\sup_K M$ is a Σ_{n+1}-elementary extension of the base model M. In the model-theoretic setting, this directly implies $\sup_K M \models \mathrm{I}\Delta_0$. When formulated in terms of forcing interpretations, elementarity becomes reflection, but being reflecting in the sense of Definition 1.16 is not enough to ensure that *every* condition forces $\mathrm{I}\Delta_0$, which is required by (FI13). So we strengthen the notion of reflection in the statement below. From this, one quickly derives $\exp$ using the cofinality of M inside $\sup_K M$ and the $\mathrm{I}\Delta_0$-provable monotonicity of the exponential function.

Theorem 4.6. For every $n \in \mathbb{N}$, the forcing translation $\tau(n)$ is a polynomial forcing interpretation of $\mathrm{B}\Sigma_{n+1} + \exp$ in $\mathrm{I}\Sigma_n + \exp$ that is polynomially Π_{n+2}-reflecting.

In fact the $\tau(n)$ are polynomially Π_{n+2}-reflecting in the following strengthened and uniform sense: there is a single polynomial-time procedure which, given $n \in \mathbb{N}$ in unary and a Π_{n+2} sentence γ, outputs a proof of

$$\exists s{\in}\mathrm{Cond}_{\tau(n)} \, (s \Vdash_{\tau(n)} \gamma) \rightarrow \gamma$$

in $\mathrm{I}\Sigma_n + \exp$.

Proof. As in earlier arguments in this section, we think of $n \in \mathbb{N}$ as fixed and leave it to the reader to check that the running time of a procedure constructing the proofs in $\mathrm{I}\Sigma_n + \exp$ described below can be bounded by a polynomial in n and $|\gamma|$. We only have to consider (FI13) and the strengthened reflection property, as the other conditions from the definition of forcing interpretation have been dealt with in Lemma 4.3.

Strengthened reflection. Let γ be a Π_{n+2} sentence of the form $\forall \bar{x} \, \exists \bar{y} \, \theta(\bar{x}, \bar{y})$, where $\theta(x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_\ell)$ is Π_n. Let s be such that $s \Vdash \forall \bar{x} \, \exists \bar{y} \, \theta(\bar{x}, \bar{y})$. Take any $a_1, a_2, \ldots, a_k$. Then, as one can verify using (FT7) and the definition of $\Vdash_{\tau(n)}$, we have $s \Vdash \exists \bar{y} \, \theta(\breve{\bar{a}}, \bar{y})$, where each $\breve{a}_i$ is the (code for the) Σ_{n+1} formula $x = x \wedge y = \underline{a_i}$, which defines the graph of the constant function with value

a_i. From this, we deduce that $\forall s' \trianglelefteq s\ \exists s'' \trianglelefteq s'\ \exists \bar{v}\ \left(s'' \Vdash \theta(\bar{\bar{a}}, \bar{v})\right)$. We then apply Lemma 4.5 to get

$$\forall s' \trianglelefteq s\ \exists s'' \trianglelefteq s'\ \exists \bar{v}\ \left(s'' \Vdash \bar{v}\downarrow \wedge\ \forall^{\infty} x {\in} \mathrm{Ext}_{n+1}(s'')\ (\bar{v}(x)\downarrow\ \to\ (\theta(\bar{a}, \bar{v}(x))))\right).$$

In particular, we know $\exists \bar{y}\ \theta(\bar{a}, \bar{y})$ holds.

Interpretation of $\mathrm{I}\Delta_0 + \exp$. The $\mathrm{I}\Delta_0$ part follows easily from strengthened reflection for Σ_1 sentences. For exp, simply prove that for any $s \in \mathrm{Cond}$ and any e, if $s \Vdash e\downarrow$, then $s \Vdash e' = 2^e$, where e' is a (code for a) Σ_{n+1} formula such that

$$\Sigma_{n+1}\text{-}\mathrm{Sat}(e', \langle x, z\rangle) \leftrightarrow \exists z\ \left(\Sigma_{n+1}\text{-}\mathrm{Sat}(e, \langle x, y\rangle) \wedge y = 2^z\right).$$

Interpretation of $\mathrm{B}\Sigma_{n+1}$. By induction on $m \leqslant n$, we construct proofs of

$$\forall s {\in} \mathrm{Cond}\ (s \Vdash \mathrm{B}\Pi_m).$$

This is sufficient by Proposition 1.14 and the well-known equivalence between $\mathrm{B}\Pi_n$ and $\mathrm{B}\Sigma_{n+1}$. Take any $m \leqslant n$. In the case when $m \neq 0$, suppose that we already have a proof of $\forall s {\in} \mathrm{Cond}\ (s \Vdash \mathrm{B}\Pi_{m-1})$.

Consider a Π_m formula $\theta(x, y, \bar{z})$. Work over $\mathrm{I}\Sigma_n + \exp$. Given names $t, \bar{u}$ and a condition $s \in \mathrm{Cond}$ such that $s \Vdash^+ t\downarrow$, $s \Vdash^+ \bar{u}\downarrow$, and $s \Vdash \forall x {\leqslant} t\ \exists y\ \theta(x, y, \bar{u})$, it is enough to show that $s \Vdash \exists y^*\ \forall x {\leqslant} t\ \exists y {\leqslant} y^*\ \theta(x, y, \bar{u})$.

For the sake of contradiction, assume that $s \nVdash \exists y^*\ \forall x {\leqslant} t\ \exists y {\leqslant} y^*\ \theta(x, y, \bar{u})$. In what follows, we will construct a condition $s' \trianglelefteq s$ and a name e so that $s' \Vdash^+ e\downarrow$, $s' \Vdash e \leqslant t$ and $s' \Vdash \neg \exists y\ \theta(e, y, \bar{u})$, which will lead to a contradiction with $s \Vdash \forall x {\leqslant} t\ \exists y\ \theta(x, y, \bar{u})$. We first make the following claim.

Claim. For any a, there exist $p \geqslant a$ and q such that

$$(\dagger)\ p \in \mathrm{Ext}_{n+1}(s) \wedge q \leqslant t(p) \wedge \forall y {\leqslant} a\ \neg\theta(q, y, \bar{u}(p)).$$

To prove the claim, argue as follows. By our assumptions on s, we have $s \nVdash \forall x {\leqslant} t\ \exists y {\leqslant} \check{a}\ \theta(x, y, \bar{u})$ for any a, where $\check{a}$ is defined as above. The formula $\forall x {\leqslant} t\ \exists y {\leqslant} \check{a}\ \theta(x, y, \bar{u})$ is either a Π_m formula outright if $m = 0$, or equivalent to a Π_m formula over $\mathrm{B}\Pi_{m-1}$ otherwise, and in the latter case we know that $s \Vdash \mathrm{B}\Pi_{m-1}$ by the proof from the induction hypothesis. So, we may apply Lemma 4.5 to this formula, deriving the existence of arbitrarily large p for which there exists $q \leqslant t(p)$ such that $(\dagger)$ holds. To conclude the proof of the claim, take some such $p \geqslant a$.

Notice that $(\dagger)$ is a Σ_{n+1} statement. We now reason somewhat like in the proof of Lemma 4.3. Let H be the set of all finite sequences of the form $(p_i, q_i, w_i)_{i \leqslant k}$ such that for each $i \leqslant k$, the triple $\langle p_i, q_i, w_i\rangle$ is the smallest one satisfying:

- w_i is a witness that p_i, q_i satisfy $(\dagger)$ with $a = \langle p_{i-1}, q_{i-1}, w_{i-1}\rangle$, or with $a = 0$ if $i = 0$,

- $p_i > \langle p_j, q_j, w_j\rangle$ for each $j < i$.

Then H is $\Sigma_0(\Sigma_n)$-definable; in fact, in the presence of $\mathrm{B}\Sigma_n$ it can be defined by the conjunction of a Π_n and a Σ_n formula. Moreover, H has no maximal element by the Claim, thus it is cofinal. Define s' so that $p \in \mathrm{Ext}_{n+1}(s')$ iff

there is a finite sequence $(p_i, q_i, w_i)_{i \leqslant k}$ in H such that $p = p_k$. Define the Σ_{n+1} formula e so that Σ_{n+1}-$\mathrm{Sat}(e, \langle p, q \rangle)$ holds iff there is a finite sequence in H such that $p = p_k, q = q_k$. Then $s' \in \mathrm{Cond}$, and clearly $s' \trianglelefteq s$. Moreover, by construction we have $s' \Vdash^+ e{\downarrow}$ and $s' \Vdash e \leqslant t$. Note in particular that the existence of a number bounding all values of e on arguments from $\mathrm{Ext}_{n+1}(s')$ follows from $s \Vdash^+ t{\downarrow}$.

It remains to show that $s' \Vdash \neg \exists y\, \theta(e, y, \bar{u})$. Otherwise, there would be a condition $s'' \trianglelefteq s'$ and a name e' such that $s'' \Vdash^+ e'{\downarrow}$ and $s'' \Vdash \theta(e, e', \bar{u})$. By Lemma 4.5, we have

$$\forall^{\infty} p{\in}\mathrm{Ext}_{n+1}(s'')\, \theta(e(p), e'(p), \bar{u}(p)).$$

Since $s'' \Vdash^+ e'{\downarrow}$, there is some number d such that $\forall p{\in}\mathrm{Ext}_{n+1}(s'')\, (e'(p) \leqslant d)$, and thus

$$\forall^{\infty} p{\in}\mathrm{Ext}_{n+1}(s'')\, \exists y{\leqslant}d\, \theta(e(p), y, \bar{u}(p)).$$

Taking large enough $(p_i, q_i, w_i)_{i \leqslant k}$ in H such that $\langle p_{k-1}, q_{k-1}, w_{k-1} \rangle \geqslant d$ and $p_k \in \mathrm{Ext}_{n+1}(s'')$, we get $\exists y{\leqslant}d\, \theta(q_k, y, \bar{u}(p_k))$, but this contradicts the definition of p_k and q_k. $\qquad\square$

Remark. For $n > 0$, there is an alternative polynomially Π_{n+2}-reflecting forcing interpretation of $\mathrm{B}\Sigma_{n+1}$ in $\mathrm{I}\Sigma_n$ based on a generic cut construction similar to the one from Section 2. We work in a version $\mathrm{I}\Sigma_n + (\mathbb{I})$ in which the axiom $(\mathbb{I}3)$ is changed to "every infinite Σ_{n-1}-definable set contains an ω^x-large subset for some $x > \mathbb{I}$". (The axioms $(\mathbb{I}1)$–$(\mathbb{I}3)$ can then be eliminated by some variant of Lemma 2.9 as in Section 2). The role of infinite sets is now played by infinite Σ_{n-1}-definable sets. The forcing conditions are finite sets s (which now include e.g. all bounded Δ_n-definable sets) that are ω^x-large for some $x > \mathbb{I}$ and, for $n \geqslant 2$, have the property that if $\ell_1 < \ell_2$ are elements of s, then the Skolem function for the first $\exists$ quantifier of the Σ_{n-1} universal formula takes values below ℓ_2 on inputs below ℓ_1. The rest of the argument is along the lines of Section 2, except that the step for RT_2^2 in the proof of Lemma 2.15 is replaced by the considerably easier step for the infinite Δ_0 pigeonhole principle (which is well known to be equivalent to $\mathrm{B}\Sigma_2$) relative to $0^{(n-1)}$.

In an analogue of this argument for $n = 0$, the conditions would be finite sets with more than $\mathbb{I}$ elements that are at least exponentially far apart from each other. However, the proof of Lemma 2.16 (reflection) no longer works. The proof of the reflection lemma makes use of the fact that for any element k there is a condition s such that $\min s \geqslant k$. But in a model of $\mathrm{I}\Delta_0 + \exp$ there might be an element above which the exponential function may be iterated only a standard (*a fortiori*, no greater than $\mathbb{I}$) number of times.

For each fixed $n \in \mathbb{N}$, the proof of Theorem 4.6 can be formalized in the theory PV, a fragment of bounded arithmetic corresponding to polynomial-time computation. As already mentioned, this provides a solution to a problem of Clote et al. [7] for $n \geqslant 1$ and a partial solution for $n = 0$.

Corollary 4.7. For each $n \in \mathbb{N}$, the theory $\mathrm{I}\Sigma_n + \exp$ polynomially simulates $\mathrm{B}\Sigma_{n+1} + \exp$ with respect to Π_{n+2} sentences. Moreover, the Π_{n+2}-conservativity of $\mathrm{B}\Sigma_{n+1} + \exp$ over $\mathrm{I}\Sigma_n + \exp$ is provable in PV.

Proof. The first part follows directly from Theorem 4.6 and Theorem 1.17. The second part is obtained by formalizing the proof of Theorem 4.6 for a fixed $n \in \mathbb{N}$ in PV. $\qquad\square$

In fact, the proof of Theorem 4.6 is witnessed by a single polynomial-time algorithm that takes n (in unary) as input, and this algorithm can also be formalized in PV. As a result, we can conclude that the Π_{n+2}-conservativity of $B\Sigma_{n+1} + \exp$ over $I\Sigma_n + \exp$ holds provably in PV for any n for which the axioms of $B\Sigma_{n+1} + \exp$ over $I\Sigma_n + \exp$ actually exist as finite strings — that is, for any n in the definable cut Log, the domain of the exponential function.

Corollary 4.8. PV proves that for every $n \in \mathrm{Log}$, the theory $B\Sigma_{n+1} + \exp$ is Π_{n+2}-conservative over $I\Sigma_n + \exp$.

As the reader may have noticed, one can actually extract from the truncated generic Σ_{n+1} ultrapower construction (alternatively, from the generic cut construction for $n > 0$) a polynomial forcing interpretation of $\mathrm{WKL}_0^* + B\Sigma_{n+1}^0$ in $I\Sigma_n + \exp$ that is polynomially Π_{n+2}-reflecting.

Theorem 4.9. For each $n \in \mathbb{N}$, the theory $I\Sigma_n + \exp$ polynomially simulates $\mathrm{WKL}_0^* + B\Sigma_{n+1}^0$ with respect to Π_{n+2} sentences.

Proof. First, notice there exists a polynomial forcing interpretation of $\mathrm{WKL}_0^* + B\Sigma_{n+1}^0$ in $B\Sigma_{n+1} + \exp$ that polynomially reflects $\mathcal{L}_1$ sentences. For $n \geqslant 1$, this follows from Theorem 3.13(a) in Hájek [14]. If $n = 0$, then such a forcing interpretation can be extracted from Sections 5–8 in Avigad [1]. Then compose with $\tau(n)$ and invoke Theorem 1.17. $\square$

Acknowledgements

References

[1] Jeremy Avigad. Formalizing forcing arguments in subsystems of second-order arithmetic. *Annals of Pure and Applied Logic*, 82(2):165–191, 1996.

[2] Lev D. Beklemishev. A proof-theoretic analysis of collection. *Archive for Mathematical Logic*, 37(5–6):275–296, July 1998.

[3] Patrick Caldon and Aleksandar Ignjatović. On mathematical instrumentalism. *J. Symbolic Logic*, 70(3):778–794, 2005.

[4] Peter A. Cholak, Carl G. Jockusch, Jr., and Theodore A. Slaman. On the strength of Ramsey's theorem for pairs. *J. Symbolic Logic*, 66(1):1–55, 2001.

[5] C. T. Chong, Theodore A. Slaman, and Yue Yang. The metamathematics of Stable Ramsey's Theorem for Pairs. *J. Amer. Math. Soc.*, 27(3):863–892, 2014.

[6] C. T. Chong, Theodore A. Slaman, and Yue Yang. The inductive strength of Ramsey's Theorem for Pairs. *Adv. Math.*, 308:121–141, 2017.

[7] Peter G. Clote, Petr Hájek, and Jeff B. Paris. On some formalized conservation results in arithmetic. *Archive for Mathematical Logic*, 30(4):201–218, 1990.

[8] Peter G. Clote and Jan Krajíček, editors. *Arithmetic, Proof Theory, and Computational Complexity*, volume 23 of *Oxford Logic Guides*. Clarendon Press, Oxford, 1993. Papers from the conference held in Prague, July 2–5, 1991.

[9] Peter G. Clote and Jan Krajíček. Open problems. In *Arithmetic, Proof Theory, and Computational Complexity* [8], pages 1–19.

[10] Paola D'Aquino. A sharpened version of McAloon's theorem on initial segments of models of IΔ_0. *Annals of Pure and Applied Logic*, 61(1–2):49–62, May 1993.

[11] Elmar Eder. *Relative complexities of first order calculi*. Artificial Intelligence. Friedr. Vieweg & Sohn, Braunschweig, 1992.

[12] Herbert B. Enderton. *A Mathematical Introduction to Logic*. Harcourt/Academic Press, San Diego, second edition, 2001.

[13] Kurt Gödel. Über die Länge von Beweisen. In *Ergebnisse eines mathematischen Kolloquiums, Heft 7*, pages 23–24. Reprinted with English translation in Kurt Gödel, *Collected Works*, Vol. 1, pp. 396-399, Oxford University Press, 1986.

[14] Petr Hájek. Interpretability and fragments of arithmetic. In Clote and Krajíček [8], pages 185–196.

[15] Petr Hájek and Pavel Pudlák. *Metamathematics of First-Order Arithmetic*. Perspectives in Mathematical Logic. Springer-Verlag, Berlin, 1993.

[16] Denis R. Hirschfeldt. *Slicing the truth. On the computable and reverse mathematics of combinatorial principles*, volume 28 of *Lecture Notes Series. Institute for Mathematical Sciences. National University of Singapore*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2015.

[17] Aleksandar Djordje Ignjatović. *Fragments of First and Second Order Arithmetic and Lengths of Proofs*. PhD thesis, University of California at Berkeley, 1990.

[18] Jussi Ketonen and Robert Solovay. Rapidly growing Ramsey functions. *Ann. of Math. (2)*, 113(2):267–314, 1981.

[19] Leszek Aleksander Kołodziejczyk and Keita Yokoyama. Some upper bounds on ordinal-valued Ramsey numbers for colourings of pairs. *Selecta Math. (N.S.)*, 26(4):paper No. 56, 18 pages, 2020.

[20] Thomas G. McLaughlin. Sub-arithmetical ultrapowers: a survey. *Annals of Pure and Applied Logic*, 49(2):143–191, October 1990.

[21] J. B. Paris and L. A. S. Kirby. Σ_n-collection schemas in arithmetic. In *Logic Colloquium '77 (Proc. Conf., Wrocław, 1977)*, volume 96 of *Stud. Logic Foundations Math.*, pages 199–209, 1978.

[22] Ludovic Patey and Keita Yokoyama. The proof-theoretic strength of Ramsey's theorem for pairs and two colors. *Adv. Math.*, 330:1034–1070, 2018.

[23] Pavel Pudlák. The lengths of proofs. In Samuel R. Buss, editor, *Handbook of Proof Theory*, volume 137 of *Studies in Logic and the Foundations of Mathematics*, chapter VIII, pages 547–637. Elsevier, Amsterdam, 1998.

[24] David Seetapun and Theodore A. Slaman. On the strength of Ramsey's theorem. *Notre Dame Journal of Formal Logic*, 36(4):570–582, 1995.

[25] Stephen G. Simpson. *Subsystems of Second Order Arithmetic*. Perspectives in Mathematical Logic. Springer-Verlag, 1999. XIV + 445 pages; Second Edition, Perspectives in Logic, Association for Symbolic Logic, Cambridge University Press, 2009, XVI+ 444 pages.

[26] Stephen G. Simpson and Rick L. Smith. Factorization of polynomials and Σ_1^0 induction. *Annals of Pure and Applied Logic*, 31:289–306, 1986.

[27] Albert Visser. An inside view of EXP; or, the closed fragment of the provability logic of $I\Delta_0 + \Omega_1$ with a propositional constant for EXP. *The Journal of Symbolic Logic*, 57(1):131–165, March 1991.

[28] Keita Yokoyama. On the strength of Ramsey's theorem without Σ_1-induction. *Math. Log. Q.*, 59(1-2):108–111, 2013.