Bounded Arithmetic and Lower Bounds in Boolean Complexity

ALEXANDER A. RAZBOROV*

Abstract

We study the question of provability of lower bounds on the complexity of explicitly given Boolean functions in weak fragments of Peano Arithmetic. To that end, we analyze what is the right fragment capturing the kind of techniques existing in Boolean complexity at present. We give both formal and informal arguments supporting the claim that a conceivable answer is V_1^1 (which, in view of $RSUV$-isomorphism, is equivalent to S_2^1), although some major results about the complexity of Boolean functions can be proved in (presumably) weaker subsystems like U_1^1. As a by-product of this analysis, we give a more constructive version of the proof of Håstad Switching Lemma which probably is interesting in its own right.

We also present, in a uniform way, theories which do not involve second order quantifiers and show that they prove the same $\Sigma_0^{1,b}$-theorems as V_k^1, U_k^1 $(k \geq 1)$. Another application of this technique is that the schemes of $\Sigma_0^{1,b}$-replacement, $\Sigma_0^{1,b} - IND$ and $\Sigma_0^{1,b}$ limited iterated comprehension (all of which are given by Boolean combinations of $\Sigma_1^{1,b}$-formulae) together prove all $\mathcal{B}(\Sigma_1^{1,b})$-consequences of the full $\Sigma_1^{1,b} - IND$ scheme.

1. Introduction

Proving lower bounds on the complexity of explicitly given Boolean functions is one of the most challenging tasks in computational complexity. This theory met with remarkable success at least twice: in the 60's (see e.g. [36, 31, 32, 37, 38]) and in more recent time ([9, 1, 28, 11, 33, 34, 29, 3, 27, 30, 35, 24, 4, 12, 18]). A nice survey of many major results known in Boolean complexity at present can be found in [5].

Both times, however, the period of enthusiasm was followed by understanding that it is not quite clear to which extent the methods developed so far can be useful for attacking central open problems in Boolean complexity.

This paper (as well as the earlier paper [20]) mainly stemmed from the author's intention to look at this situation from the logical point of

*Supported by the grant # 93-011-16015 of the Russian Foundation for Fundamental Research

view. Obviously, all methods already developed in Boolean complexity use only a very tiny bit of the power of classical systems like Peano Arithmetic or ZF. We are interested in the question of what is the right "minimal" fragment of PA which suffices for formalizing all these methods in a natural, "straightforward" way.

We carefully present both formal and informal arguments supporting the claim that the desired fragment is V_1^1. Note that, due to $RSUV$-isomorphism [25, 26, 20], this system is equivalent to S_2^1, the latter being considered as the most important among various fragments of Bounded Arithmetic. For several reasons, however, it is more natural and elegant to work directly with second order objects while discussing provability of statements about the complexity of Boolean functions. So, in this paper we almost exclusively deal with second order theories. The interested reader can scale everything down to the first order using $RSUV$-isomorphism, although the outcome of this translation may look somewhat awkward.

The arguments mentioned in the previous paragraph only say that V_1^1 safely contains the algebraic and combinatorial methods existing in Boolean complexity at present. Some of them do not use its full power. We analyze from this point of view several major results in Boolean complexity and see in which natural subtheories of V_1^1 they can be proved. One rather surprising fact discovered during this analysis is that the method of restrictions [1, 9, 28, 11] can be formalized in either $S_2(\alpha)$ or U_1^1. Moreover, the key argument of the method known in its strongest form as *Håstad Switching Lemma* can be carried over already in $I\Delta_0(\alpha)$. This required looking at the proof of this lemma in a rather unusual fashion (see Lemma E.1 below), and this modified proof might be of independent interest.

There exists a powerful witnessing technique for studying provability of $\Sigma_1^{1,b}$-formulae in second order theories with $\Sigma_1^{1,b} - IND$ originated in [6, Theorems 10.12 and 10.16]. This technique, however, does not say anything useful about provability of $\Sigma_0^{1,b}$-formulae which are our main target (as formalizations of statements in Boolean complexity are $\Sigma_0^{1,b}$). It seems that the only known negative results concerning provability of $\Sigma_0^{1,b}$-formulae in systems with $\Sigma_1^{1,b} - IND$ come from Gödel Incompleteness Theorem. But the corresponding formulae implicitly encode large numbers, and the methods used for establishing their unprovability are hardly relevant to the "plain" combinatorial problems from Boolean complexity we consider here.

In the rest of the paper we develop an appropriate framework for studying provability of $\Sigma_0^{1,b}$-formulae. As our approach is purely syntactical, we treat all theories V_k^1, U_k^1 $(k \geq 1)$ in a uniform way, and, in fact, we consider at once more general theories $W_T^{1,\tau}$, where T is a first order theory (obeying some natural restrictions), and $\tau(a)$ is a first order term restricting the range of the eigenvariable in $\Sigma_1^{1,b} - IND$.

We introduce the theories $W_T^{0,\tau}(\delta)$ by adding to the language of $W_T^{1,\tau}$

relationals δ evaluating polynomial size Boolean circuits with the depth constraints specified by the term τ, removing $\Sigma_0^{1,b} - CA$ and restricting $\Sigma_1^{1,b} - IND$ to $\Sigma_0^{1,b} - IND$. The theories $W_T^{0,\tau}(\delta)$ do not involve second order quantifiers at all, and the main result says that they prove the same $\Sigma_0^{1,b}$-formulae in the original language as $W_T^{1,\tau}$.

A by-product of this technique is that $\Sigma_0^{1,b}$-replacement, $\Sigma_0^{1,b} - IND$ and $\Sigma_0^{1,b}$ limited iterated comprehension axioms are altogether powerful enough to prove all $\mathcal{B}(\Sigma_1^{1,b})$-corollaries[1] of the full $\Sigma_1^{1,b} - IND$ scheme. Moreover, if these corollaries do not contain bounded first order quantifiers $\forall x \leq t$ [$\exists x \leq t$] in the scope of quantifiers $\exists\phi$ [$\forall\phi$, respectively] (we call such formulae *strict*) then $\Sigma_0^{1,b}$-replacement can be omitted. This may be interesting since these three schemes, unlike $\Sigma_1^{1,b} - IND$, are given by $\mathcal{B}(\Sigma_1^{1,b})$-formulae themselves and hence may be taken as axioms in free cut free proofs consisting of $\Sigma_1^{1,b} \cup \Pi_1^{1,b}$-formulae.

The paper is organized as follows. In Section 2 we recall some basic facts about second order Bounded Arithmetic, introduce the generalizations $W_T^{1,\tau}$ of the theories V_k^1, U_k^1 and show some simple results concerning their power. In Section 3 we introduce the systems $W_T^{0,\tau}(\delta)$ and construct an interpretation of these systems in a fragment of $W_T^{1,\tau}$. The next section 4 is technical, we show that in the theories $W_T^{0,\tau}(\delta)$ the nested recursive definitions (forbidden in the original axiomatization) are actually admissible. In Section 5 we prove the main witnessing lemma. In the next section 6 we formulate our main results which in fact are plain corollaries of the material contained in the previous sections.

The discussion of connections with Boolean complexity (which is the main motivation for this work) is postponed until Appendix. The reason is that it is convenient to use for this purpose some concepts introduced in the rest of the paper.

1.1. Related results about first order theories

Although, for the reason explained above, we are mainly interested in provability of $\Sigma_0^{1,b}$-formulae, our technique also allows us to view from a single perspective several previously known results on the computational complexity of functions $\Sigma_1^{1,b}$-definable in various first order theories. The basis for this comparison is provided by $RSUV$-isomorphism. Due to this isomorphism, the theory V_1^1 is equivalent to S_2^1 [25, 26, 20], and the theory U_1^1 is equivalent to R_2^1 [26].

It is an immediate corollary of the witnessing lemma 5.2 that $\Sigma_1^{1,b}$-definable functions in $W_1^{1,\tau}$ are exactly those computable by uniform families of polynomial size $\tau(a)^{O(1)}$-depth circuits. For the case of V_1^1 (that is

[1] $\mathcal{B}(\Phi)$ stands for the closure of the class Φ under Boolean operations

when $\tau(a) = a$), the depth constraint becomes unessential (see Theorem 3.9 below), and the corresponding families of circuits can compute exactly functions in P. Thus, our result for this case is analogous to the main result of [6] concerning Σ_1^b-definable in S_2^1 functions.

The first order theory characterizing NC computable functions was in various forms introduced by Allen [2] and Clote [7]. Takeuti [26] later showed that this theory is equivalent to R_2^1 and, via the $RSUV$-isomorphism, to U_1^1. With these equivalencies in mind, our characterization of $\Sigma_1^{1,b}$-definable in U_1^1 functions (note that polynomial size $|n|^{O(1)}$-depth uniform circuits are exactly NC-circuits) provides a new proof of the result by Allen and Clote.

Finally, if we appropriately adjust the languages, the theory $V_1^0(\delta)$ resembles Cook's equational system PV [8], only instead of introducing function symbols for polynomially computable functions, we introduce relationals for evaluating polynomial size circuits. Respectively, the proof of our main result corresponds to the conservation result concerning S_2^1 and PV [6, Chapter 6].

1.2. Recent developments

A purely complexity framework for analyzing the methods developed so far in non-uniform Boolean complexity was proposed by Razborov and Rudich in [22]. Namely, in that paper we introduced the notion of natural proof and argued that the known proofs of lower bounds on the complexity of explicit Boolean functions in non-monotone models fall within this definition of natural. These include e.g. the proofs for bounded-depth circuits analyzed in Appendix E.3, E.4 of this paper.

It was shown in [22], based upon a widely believed hardness assumption, that there is no natural proof of superpolynomial lower bounds for general circuits.

One application of natural proofs to the logical framework developed in this paper was given in [21]. Based upon an interpolation-like theorem, it was proved there that any proof of lower bounds for non-monotone models in the theory $S_2^2(\alpha)$ can be recast as natural. Combined with the main theorem from [22], this leads to the first partial independence result toward the goal of understanding provability of superpolynomial lower bounds for general circuits in V_1^1.

2. Theories $W_T^{1,\tau}$

We assume the familiarity with [6].

Denote by L_k ($k \geq 1$) the first order language with equality which consists of the constant 0, function symbols $S, +, \cdot, \lfloor \frac{1}{2}x \rfloor, |x|, x\#_2 y, \ldots, x\#_k y$

and of the predicate symbol $\leq$. In particular, $L_1 = \langle 0, S, +, \cdot, \lfloor \frac{1}{2}x \rfloor, |x|, \leq \rangle$, and $L_2 = \langle 0, S, +, \cdot, \lfloor \frac{1}{2}x \rfloor, |x|, \#, \leq \rangle$, where $\#$ is used as an abbreviation for $\#_2$.

Let $BASIC_2$ be the set of 32 open axioms in the language L_2 from [6, §2.2] describing basic properties of its symbols. We will denote by $BASIC_1$ the set of axioms in L_1 obtained from $BASIC_2$ by removing the $\#$-related axioms (13)–(18). For $k \geq 3$, the set $BASIC_k$ is obtained by generalizing Buss's axiom (13) to

$$|x\#_j y| = S(|x|\#_{j-1}|y|) \ (2 \leq j \leq k)$$

and adding the new axiom

$$z < x\#_j y \equiv |z| < |x\#_j y| \ (2 \leq j \leq k)$$

(cf. [20]).

2.1. Definition. We say that a first order theory T in a language $L \supseteq L_1$ is *regular* if it possesses the following properties:

a) $BASIC_1 \subseteq T$,

b) all axioms of T are bounded,

c) every function symbol (and hence every term) of the language L can be bounded from above in the theory T by a provably monotone term.

Let in particular T_k be the regular theory in the language L_k which has $BASIC_k$ as its list of axioms.

From now on we fix a first order extension L of the language L_1 and a regular theory T in the language L. Let $\mathcal{L}$ be the second order extension of L obtained by augmenting it with second order variables $\{\alpha_i^r \mid i, r \in \mathbf{N};\ r \geq 1\}$, where r denotes the arity of the variable. The superscript r will be dropped whenever this can not create confusion.

Let $\tau(a)$ be a provably monotone (in T) term such that

$$T \vdash \tau(a) \geq |a|. \tag{1}$$

2.2. Definition. The scheme $\Sigma_1^{1,b} - \tau - IND$ is defined as

$$A(0) \wedge \forall x(A(x) \supset A(x+1)) \ \supset \ \forall x A(\tau(x)),$$

where $A(a)$ is a $\Sigma_1^{1,b}$-formula of the language $\mathcal{L}$.

2.3. Definition. The second order theory $W_T^{1,\tau}$ in the language $\mathcal{L}$ has the following axioms:

a) T,

b) $\Sigma_1^{1,b} - \tau - IND$,

c) $\Sigma_0^{1,b} - CA$.

2.4. Remark. Assumption c) from Definition 2.1 implies that bounding terms in bounded formulae can be w.l.o.g. assumed to be provably monotone. With this remark the proof of [6, Theorems 2.2, 2.4] on the possibility of introducing Σ_1^b-defined function and Δ_1^b-defined predicate symbols into the language of S_2^1 readily extends to the theories $W_T^{1,\tau}$. In particular, all $\Sigma_0^{1,b}$-defined function and predicate symbols can be freely used in the schemes $\Sigma_1^{1,b} - \tau - IND$ and $\Sigma_0^{1,b} - CA$.

If $\tau(a) \equiv a$ then $\Sigma_1^{1,b} - \tau - IND$ is merely $\Sigma_1^{1,b} - IND$. In this case $W_T^{1,\tau}$ for obvious reasons will be denoted by V_T^1.

For $\tau(a) \equiv |a|$, $\Sigma_1^{1,b} - \tau - IND$ becomes $\Sigma_1^{1,b} - LIND$, and $W_T^{1,\tau}$ will be denoted by U_T^1. The latter notation is justified by the following

Theorem 2.5. *U_T^1 is equivalent to $T + \Sigma_1^{1,b} - PIND + \Sigma_0^{1,b} - CA$.*

Proof. A careful inspection of Buss's proof that the $PIND$ and $LIND$ schemes are equivalent reveals that the part $PIND \Rightarrow LIND$ uses only axioms from $BASIC_1$ [6, Theorem 2.6]. The converse result $LIND \Rightarrow PIND$ [6, Theorem 2.11] uses only symbols which can be defined by bounded formulae in $BASIC_1 + \Sigma_1^b - LIND$. Hence, by assumption a) from Definition 2.1 and Remark 2.4, the same proof shows that $U_T^1 \vdash \Sigma_1^{1,b} - PIND$. ∎

Now, if we also abbreviate $U_{T_k}^1, V_{T_k}^1$ to U_k^1, V_k^1, we see that the theories $W_T^{1,\tau}$ form a convenient uniform generalization of Buss's theories U_k^1, V_k^1.

Theorem 2.6. *$W_T^{1,\tau} \vdash \Delta_1^{1,b} - IND$.*

Proof. A careful inspection of the proof of the result due to Dowd and Statman that $S_2^1 \vdash \Delta_1^b - IND$ (see [6, Theorem 2.22]) reveals that it uses only symbols which may be defined already in S_1^1. Hence, the same proof readily shows that $U_1^1 \vdash \Delta_1^{1,b} - IND$. As, in view of assumption a) from Definition 2.1 and (1), $W_T^{1,\tau}$ is an extension of U_1^1, this can be also generalized to $W_T^{1,\tau}$. ∎

We are also interested in the Φ-replacement scheme. It will be convenient to take it in the following form:

2.7. Definition. The *Φ-replacement scheme* is given by

$$\begin{aligned}
&\forall x \le t \exists \phi_1^{r_1} \ldots \exists \phi_l^{r_l} A(x, \phi_1, \ldots, \phi_l) \\
&\quad \supset \ \exists \phi_1^{r_1+1} \ldots \phi_l^{r_l+1} \forall x \le t \\
&\qquad A(x, \{x_1, \ldots, x_{r_1}\}\, \phi_1(x, x_1, \ldots, x_{r_1}), \ldots, \\
&\qquad\quad \{x_1, \ldots, x_{r_l}\}\, \phi_l(x, x_1, \ldots, x_{r_l})),
\end{aligned}$$

where A is in Φ.

Theorem 2.8. $W_T^{1,\tau} \vdash \Sigma_1^{1,b}$*-replacement.*

Proof. By extending the proof of [6, Theorem 9.16] in the same manner as with Theorems 2.5, 2.6. ∎

In the proof of the next theorem and in some other places we will write $\wedge_i x_i \leq y$ and $\wedge_i x_i < y$ in the simplified form $\vec{x} \leq y$ and $\vec{x} < y$ respectively.

Theorem 2.9. *For any fixed integer $k > 0$, $W_T^{1,\tau} \vdash \Sigma_1^{1,b} - \tau^k - IND$.*

Proof. Let $A(a) \in \Sigma_1^{1,b}$. Set $B(a, b_0, \ldots, b_{k-1}) \rightleftharpoons A(b_0 + b_1 \cdot \tau(a) + \cdots + b_{k-1} \cdot \tau^{k-1}(a))$. We show by induction on i that

$$\left.\begin{aligned} &W_T^{1,\tau} \vdash \forall x(A(x) \supset A(x+1)) \ \supset \ [B(a, 0, \ldots, 0, b_i, \ldots, b_{k-1}) \\ &\quad \supset \forall \vec{x} \leq \tau(a) B(a, x_0, \ldots, x_{i-1}, b_i, \ldots, b_{k-1})]. \end{aligned}\right\} \quad (2)$$

Base $i = 0$**.** There is nothing to prove.

Inductive step. Assume that for some $i \leq k-1$ we already have (2), and we want to prove this for $(i+1)$ instead of i.

First, we have from (2)

$$\left.\begin{aligned} &W_T^{1,\tau} \vdash \forall x(A(x) \supset A(x+1)) \ \supset \ [B(a, 0, \ldots, 0, b_i, \ldots, b_{k-1}) \\ &\quad \supset B(a, \tau(a) \dot{-} 1, \ldots, \tau(a) \dot{-} 1, b_i, \ldots, b_{k-1})]. \end{aligned}\right\} \quad (3)$$

Next,

$$\left.\begin{aligned} &U_1^1 \vdash \forall x(A(x) \supset A(x+1)) \\ &\quad \supset \ [B(a, \tau(a) \dot{-} 1, \ldots, \tau(a) \dot{-} 1, b_i, \ldots, b_{k-1}) \\ &\qquad \supset B(a, 0, \ldots, 0, b_i + 1, b_{i+1}, \ldots, b_{k-1})]. \end{aligned}\right\} \quad (4)$$

From (3) and (4) we conclude

$$\begin{aligned} &W_T^{1,\tau} \vdash \forall x(A(x) \supset A(x+1)) \ \supset \ [B(a, 0, \ldots, 0, b_i, b_{i+1}, \ldots, b_{k-1}) \\ &\quad \supset B(a, 0, \ldots, 0, b_i + 1, b_{i+1}, \ldots, b_{k-1})]. \end{aligned}$$

Applying $\Sigma_1^{1,b} - \tau - IND$ on c to the formula

$$\forall x_i \leq c \ B(a, 0, \ldots, 0, x_i, b_{i+1} \ldots, b_{k-1}),$$

we find

$$\left.\begin{aligned} &W_T^{1,\tau} \vdash \forall x(A(x) \supset A(x+1)) \\ &\quad \supset \ [B(a, 0, \ldots, 0, 0, b_{i+1}, \ldots, b_{k-1}) \\ &\qquad \supset \forall x_i \leq \tau(a) B(a, 0, \ldots, 0, x_i, b_{i+1}, \ldots, b_{k-1})]. \end{aligned}\right\} \quad (5)$$

We use the inductive assumption (2) again and have

$$\left.\begin{aligned} & W_T^{1,\tau} \vdash \forall x(A(x) \supset A(x+1)) \\ & \quad \supset \ [\forall x_i \leq \tau(a) B(a,0,\ldots,0,x_i,b_{i+1},\ldots,b_{k-1}) \\ & \qquad \supset \forall \vec{x} \leq \tau(a) B(a,x_0,\ldots,x_{i-1},x_i,b_{i+1},\ldots,b_{k-1})]. \end{aligned}\right\} \quad (6)$$

(5) and (6) complete the inductive step.

Now, (2) for $i = k$ implies in particular that

$$W_T^{1,\tau} \vdash \forall x(A(x) \supset A(x+1)) \ \supset \ [B(a,0,\ldots,0,0) \supset B(a,0,\ldots,0,\tau(a))]$$

which is what we want to prove. ■

3. Theories $W_T^{0,\tau}(\delta)$

The underlying idea toward defining these theories is to extend the language $\mathcal{L}$ by those "explicit" $\Delta_1^{1,b}$-defined relationals δ in the style of [6, §9.7] which correspond to the predicate analogue of the limited iteration on notation. This in fact is equivalent to declaring the ability of evaluating polynomial size circuits with the depth constraints specified by the term τ.

3.1. Definition. Let $A(a,\vec{\alpha}), B(a,b,\vec{\alpha},\beta^1)$ be $\Sigma_0^{1,b}$-formulae *of the language* $\mathcal{L}$ with all free variables displayed, and $k > 0$ be a positive integer. Then we introduce the relational $\delta(a,b,\vec{\alpha}) = \delta_{A,B}^k(a,b,\vec{\alpha})$ with the *defining axioms*

$$\left.\begin{aligned} \delta(a,0,\vec{\alpha}) &\equiv A(a,\vec{\alpha}), \\ \delta(a,b+1,\vec{\alpha}) &\equiv b < \tau^k(a) \wedge B(a,b,\vec{\alpha},\{x\}(x \leq a \wedge \delta(x,b,\vec{\alpha}))). \end{aligned}\right\} \quad (7)$$

Denote by $\mathcal{L}^\tau(\delta)$ the language obtained from $\mathcal{L}$ by adding to it all relationals $\delta_{A,B}^k$.

3.2. Definition. $W_T^{0,\tau}(\delta)$ is the second order theory in the language $\mathcal{L}^\tau(\delta)$ with the following axioms:

a) T,

b) defining axioms (7) for all relationals $\delta = \delta_{A,B}^k$,

c) $\Sigma_0^{1,b}(\delta) - IND$.

Note that since $W_T^{0,\tau}(\delta)$ always contains $I\Delta_0$, we may freely use in the induction scheme all function and predicate symbols defined in $I\Delta_0$ by bounded formulae. In particular, for each fixed $r > 0$ we have a function symbol $\langle a_1, \ldots, a_r \rangle$ which implements a one-to-one mapping $\mathbf{N}^r \longrightarrow \mathbf{N}$ and r unary symbols $\Pi_1^r, \ldots, \Pi_r^r$ representing the inverse mapping $\mathbf{N} \longrightarrow \mathbf{N}^r$. We choose them in such a way that

$$I\Delta_0 \vdash a_1 + \cdots + a_r < a_1' + \cdots + a_r' \supset \langle a_1, \ldots, a_r \rangle < \langle a_1', \ldots, a_r' \rangle.$$

Notice that this implies that $\langle a_1, \ldots, a_r \rangle$ are provably monotone in each variable, and also that

$$I\Delta_0 \vdash \langle a_1, \ldots, a_r \rangle \geq a_i \ (1 \leq i \leq r). \tag{8}$$

Π_1^2 and Π_2^2 will be abbreviated to Π_1 and Π_2 respectively.

One of the goals of this paper is to show that $W_T^{1,\tau}$ and $W_T^{0,\tau}(\delta)$ are equivalent with respect to $\Sigma_0^{1,b}$-formulae. In this section we will prove one (easier) part of it: every $\Sigma_0^{1,b}$-formula provable in $W_T^{0,\tau}(\delta)$ is also provable in $W_T^{1,\tau}$. It will be convenient, however, to establish at once a stronger result in the form which will be needed in Section 6. For this purpose we define the explicit scheme in the language $\mathcal{L}$ corresponding to Definition 3.1.

3.3. Definition. $\Phi - \tau - LICA$, Φ *τ-limited iterated comprehension axioms* are given by the following axiom scheme:

$$\begin{aligned} &\exists \phi^2 \forall x \leq t[\phi(x,0) \equiv A(x) \\ &\quad \wedge \forall y < \tau^k(s(x)) \\ &\qquad (\phi(x, y+1) \equiv B(x, y, \{x'\}(x' \leq x \wedge \phi(x', y))))], \end{aligned}$$

where $A(a), B(a, b, \alpha')$ are in Φ; $t, s(a)$ are first order terms, and $k > 0$ is a fixed integer.

3.4. Definition ([25, 20]). $\Phi - BCA$, Φ *bounded comprehension axioms* are given by the axiom scheme:

$$\exists \phi^r \forall x_1, \ldots, x_r \left\{ \phi^r(x_1, \ldots, x_r) \equiv (A(x_1, \ldots, x_r) \wedge \vec{x} \leq t) \right\},$$

where A is in Φ.

We have:

Theorem 3.5. *There exists an interpretation of $W_T^{0,\tau}(\delta) + \Sigma_0^{1,b}(\delta) - BCA$ in $T + \Sigma_0^{1,b} - CA + \Sigma_0^{1,b} - \tau - LICA + \Sigma_0^{1,b} - IND$ identical on formulae of the language $\mathcal{L}$.*

Proof. During this proof, it will be convenient (and in fact even necessary to make the word "interpretation" precise) to change our view and, like in [20], treat second order theories as many-sorted first order theories with equality having variables of sorts $0, 1, 2, \ldots$, where 0 is reserved for the sort of first order variables and $r > 0$ is the sort of r-ary second order variables. The equality for variables of sort $r > 0$ is introduced by

$$\alpha^r = \beta^r \rightleftharpoons \forall \vec{x}\,(\alpha^r(\vec{x}) \equiv \beta^r(\vec{x}))\,.$$

Note that it may *not* appear in bounded formulae.

Let δ be the relational of the language $\mathcal{L}^\tau(\delta)$ with the defining axioms (7). Consider the corresponding instance of $\Sigma_0^{1,b} - \tau - LICA$

$$\left.\begin{array}{l} \exists\phi^2 \forall x \le c[\phi(x,0) \equiv A(x,\vec{\alpha}) \\ \quad \wedge\ \forall y < \tau^k(x) \\ \qquad (\phi(x,y+1) \equiv B(x,y,\vec{\alpha},\{x'\}(x' \le x \wedge \phi(x',y))))], \end{array}\right\} \quad (9)$$

where this time all free variables are displayed. Apply $\Sigma_0^{1,b} - CA$ to trim ϕ in (9) to the area $x \le c,\ y \le \tau^k(x)$. We will have

$$\left.\begin{array}{l} T + \Sigma_0^{1,b} - \tau - LICA + \Sigma_0^{1,b} - CA \vdash \exists\phi^2 \\ \quad \{\forall x \forall y(\phi(x,y) \supset (x \le c \wedge y \le \tau^k(x))) \\ \qquad \wedge\ \forall x \le c[\phi(x,0) \equiv A(x,\vec{\alpha}) \\ \qquad\quad \wedge\ \forall y < \tau^k(x)(\phi(x,y+1) \equiv B(x,y,\vec{\alpha}, \\ \qquad\qquad \{x'\}(x' \le x \wedge \phi(x',y))))]\}. \end{array}\right\} \quad (10)$$

$\Sigma_0^{1,b} - IND$ readily shows the uniqueness of ϕ^2 satisfying (10). Hence we may introduce into the theory $T + \Sigma_0^{1,b} - CA + \Sigma_0^{1,b} - \tau - LICA + \Sigma_0^{1,b} - IND$ the function symbol $F_\delta^2(c,\vec{\alpha})$ (the superscript 2 indicates that F_δ takes values in variables of sort 2) with the defining axioms

$$\left.\begin{array}{rcl} F_\delta^2(c,\vec{\alpha})(a,b) & \supset & a \le c \wedge b \le \tau^k(a), \\ a \le c & \supset & \big(F_\delta^2(c,\vec{\alpha})(a,0) \equiv A(a,\vec{\alpha})\big), \\ a \le c \wedge b < \tau^k(a) & \supset & F_\delta^2(c,\vec{\alpha})(a,b+1) \\ & & \quad \equiv B(a,b,\vec{\alpha},\{x\}(x \le a \\ & & \qquad \wedge\ F_\delta^2(c,\vec{\alpha})(x,b))). \end{array}\right\} \quad (11)$$

This extension will be conservative over $T + \Sigma_0^{1,b} - CA + \Sigma_0^{1,b} - \tau - LICA + \Sigma_0^{1,b} - IND$. Throughout the rest of the proof, $\vdash$ will stand for the provability in this extension.

The crucial property of the newly introduced function symbols is their "monotonicity" in the following sense:

$$\vdash c_0 \leq c_1 \ \supset \ \forall x \leq c_0 \, (F_\delta(c_0, \vec{\alpha})(x, b) \equiv F_\delta(c_1, \vec{\alpha})(x, b)) \, . \qquad (12)$$

This is readily proved by $\Sigma_0^{1,b} - IND$ from (11).

Now we interpret the relationals $\delta(a, b, \vec{\alpha})$ as follows:

$$\delta(a, b, \vec{\alpha}) \rightsquigarrow F_\delta(a, \vec{\alpha})(a, b).$$

The interpretations of defining axioms (7) readily follow from (11) using (12).

Let us call a formula in the extended language $\mathcal{L}(F_\delta)$ *simple* if it does not contain nested occurrences of the newly introduced function symbols F_δ. Note that the interpretation of any formula in the language $\mathcal{L}^\tau(\delta)$ is simple. Hence, in order to finish the proof of Theorem 3.5, we only have to show that $T + \Sigma_0^{1,b} - CA + \Sigma_0^{1,b} - \tau - LICA + \Sigma_0^{1,b} - IND$ proves $\Sigma_0^{1,bs}(F_\delta) - IND$ and $\Sigma_0^{1,bs}(F_\delta) - BCA$, where we denoted by $\Sigma_0^{1,bs}(F_\delta)$ the set of all simple $\Sigma_0^{1,b}(F_\delta)$-formulae.

For this we need the following

Claim 3.6. *Let $A(a_1, \ldots, a_r) \in \Sigma_0^{1,bs}$, and $t_1, \ldots, t_r$ be first order terms which do not contain occurrences of $a_1, \ldots, a_r$. Then there exists $A^*(a_1, \ldots, a_r) \in \Sigma_0^{1,bs}$ such that*

$$\vdash \forall x_1 \leq t_1 \ldots \forall x_r \leq t_r (A(x_1, \ldots, x_r) \equiv A^*(x_1, \ldots, x_r)) \qquad (13)$$

and the scope of every F_δ-symbol in A^* contains no variables from the list $a_1, \ldots, a_r$ and no bound variables.

Proof of Claim 3.6. Induction on complexity of A.

If $A \equiv F_\delta(s(a_1, \ldots, a_r), \vec{\alpha})(s_1, u)$, we let

$$A^* \rightleftharpoons s_1 \leq s(a_1, \ldots, a_r) \wedge F_\delta(\bar{s}(t_1, \ldots, t_r), \vec{\alpha})(s_1, u),$$

where $\bar{s}$ is a provably monotone term bounding s from above. Then (13) follows from (12).

If $A \equiv \exists x \leq t(a_1, \ldots, a_r) B(x, a_1, \ldots, a_r)$, we find by inductive assumption a formula $B^*(a, a_1, \ldots, a_r)$ so that $\vdash \forall x \leq \bar{t}(t_1, \ldots, t_r) \forall x_1 \leq t_1 \ldots \forall x_r \leq t_r (B(x, x_1, \ldots, x_r) \equiv B^*(x, x_1, \ldots, x_r))$, where $\bar{t}$ is a provably monotone bound for t, and let $A^* \rightleftharpoons \exists x \leq t(a_1, \ldots, a_r) B^*(x, a_1, \ldots, a_r)$.

All other cases are obvious.

The proof of Claim 3.6 is complete.■

Now, in order to see $\vdash \Sigma_0^{1,bs}(F_\delta) - IND$, we notice first that $\Sigma_0^{1,bs}(F_\delta) - IND$ is equivalent to its bounded version

$$A(0) \wedge \forall x < t(A(x) \supset A(x+1)) \ \supset \ \forall x \leq t A(x). \qquad (14)$$

Applying Claim 3.6 to the formula $A(a)$ and term t, we may assume that the scope of every F_δ in (14) contains no bound variables. But this allows us to derive (14) as a substitutional instance of the $\Sigma_0^{1,b} - IND$ axiom obtained from (14) by replacing all occurrences $F_\delta(s, \vec{\alpha})$, *where s is a term*, with new free variables of sort 2.

The same argument gives us $\vdash \exists\phi\forall\vec{x} \leq t[\phi(\vec{x}) \equiv A(\vec{x})]$, where $A \in \Sigma_0^{1,bs}(F_\delta)$. Now we trim this ϕ to the area $\vec{x} \leq t$ using $\Sigma_0^{1,b} - CA$ to derive $\Sigma_0^{1,bs}(F_\delta) - BCA$.

The proof of Theorem 3.5 is complete. ∎

Corollary 3.7. *Every formula of the language $\mathcal{L}$ provable in $W_T^{0,\tau}(\delta) + \Sigma_0^{1,b}(\delta) - BCA$ is also provable in $T + \Sigma_0^{1,b} - CA + \Sigma_0^{1,b} - \tau - LICA + \Sigma_0^{1,b} - IND$.*

For the most interesting case $\tau(a) \equiv a$ the restriction $b < \tau^k(a)$ in (7) becomes unnecessary, and the set of relationals appended to the language $\mathcal{L}$ can be substantially simplified.

3.8. Definition. Let $A(a, \vec{\alpha}, \beta^1)$ be a $\Sigma_0^{1,b}$-formula in the language $\mathcal{L}$ with all free variables displayed. Then the relational $\bar{\delta} = \bar{\delta}_A(a, \vec{\alpha})$ has the following defining axiom:

$$\bar{\delta}(a, \vec{\alpha}) \equiv A(a, \vec{\alpha}, \{x\}(x < a \wedge \bar{\delta}(x, \vec{\alpha}))). \tag{15}$$

We define $V_T^0(\bar{\delta})$ similarly to $V_T^0(\delta)$ with the difference that this time we use the relationals $\bar{\delta}$ with the defining axioms (15).

Theorem 3.9. *There exist two interpretations $V_T^0(\bar{\delta}) \rightsquigarrow V_T^0(\delta)$ and $V_T^0(\delta) \rightsquigarrow V_T^0(\bar{\delta})$, both identical on the formulae of $\mathcal{L}$.*

Proof. For the relational $\bar{\delta}$ with the defining axiom (15) we define the relational $\delta(a, b, \vec{\alpha})$ by

$$\begin{aligned} \delta(a, 0, \vec{\alpha}) &\equiv a = 0 \wedge A(0, \vec{\alpha}, \{x\}\ -), \\ \delta(a, b+1, \vec{\alpha}) &\equiv b < a \wedge ((a < 2b+2 \wedge \delta(a \dot{-} 1, b, \vec{\alpha})) \vee (a = 2b+2 \\ &\qquad \wedge\ A(b+1, \vec{\alpha}, \{x\}(x \leq b \wedge \delta(b+x, b, \vec{\alpha}))))) . \end{aligned}$$

This definition is easily seen to have the required form (7) with $k := 1$ and $B := (a < 2b+2 \wedge \beta(a \dot{-} 1)) \vee (a = 2b+2 \wedge A(b+1, \vec{\alpha}, \{x\}(x \leq b \wedge \beta(b+x))))$.

Now we prove by $\Sigma_0^{1,b}(\delta)$-induction on b that $V_T^0(\delta) \vdash a \leq b \supset \delta(a+b, b, \vec{\alpha}) \equiv \delta(2a, a, \vec{\alpha})$. With the help of this, it is easy to see that the image of (15) under the translation $\bar{\delta}(a, \vec{\alpha}) \rightsquigarrow \delta(2a, a, \vec{\alpha})$ is provable in $V_T^0(\delta)$. Hence this translation defines the desired interpretation $V_T^0(\bar{\delta}) \rightsquigarrow V_T^0(\delta)$.

For the inverse interpretation, we interpret the relational $\delta(a, b, \vec{\alpha})$ with the defining axioms (7) in the theory $V_T^0(\bar{\delta})$ as

$$\delta(a, b, \vec{\alpha}) \rightsquigarrow \bar{\delta}(\langle a, b\rangle, \vec{\alpha}),$$

where $\bar{\delta}$ has the defining axiom

$$\begin{aligned}\bar{\delta}(a,\vec{\alpha}) \equiv\ & (\Pi_2(a)=0 \wedge A(\Pi_1(a),\vec{\alpha})) \\ & \vee\ (0<\Pi_2(a)\le \Pi_1(a)^k \\ & \quad \wedge\ B(\Pi_1(a),\Pi_2(a)\dot{-}1,\vec{\alpha}, \\ & \qquad \{x\}(x\le \Pi_1(a)\wedge \bar{\delta}(\langle x,\Pi_2(a)\dot{-}1\rangle,\vec{\alpha})))).\end{aligned}$$

■

4. Bootstrapping $W_T^{0,\tau}(\delta)$

In view of our general goal, the theories $W_T^{0,\tau}(\delta)$ were defined in the previous section in the most restricted way. By this I mean that the δ-symbols were a priori forbidden to appear in each other's scope. In this section we show that the most natural constructions of this kind can be in fact simulated in $W_T^{0,\tau}(\delta)$. This amounts to some rather technical work.

Lemma 4.1. *Let $A(a,b,\beta^2) \in \Sigma_0^{1,b}$, where all first order free variables are displayed, and let k be a positive integer. Then there exists a $\Sigma_0^{1,b}(\delta)$-formula $D(a,b)$ such that*

$$W_T^{0,\tau}(\delta) \vdash D(a,b) \ \equiv\ b\le \tau^k(a)\wedge A(a,b,\{x,y\}(x\le a\wedge y<b\wedge D(x,y))).$$

Proof. This lemma is similar to the first part of Theorem 3.9. The main complication is that when $\tau(a)$ is small, we in general are not guaranteed the existence of a term $t(a,b)$ such that $b\le \tau^k(t(a,b))$. We circumvent this by introducing a dummy variable c, like in the proof of Theorem 3.5.

More precisely, let $A(a,b,\beta^2)\equiv A(a,b,\vec{\alpha},\beta^2)$, where this time *all* free variables are displayed. Define a relational $\delta(a,b,\vec{\alpha})$ with the properties

$$\left.\begin{aligned}& W_T^{0,\tau}(\delta)\vdash \delta(\langle a,b,c\rangle,b,\vec{\alpha}) \\ & \quad \equiv\ \Big(b\le \tau^k(a)\ \wedge\ A(a,b,\vec{\alpha}, \\ & \qquad \{x,y\}(x\le a\wedge y<b\wedge\ \delta(\langle x,y,c\rangle,b\dot{-}1,\vec{\alpha})))\Big)\end{aligned}\right\} \quad (16)$$

and

$$W_T^{0,\tau}(\delta)\vdash b\le b'<\tau^k(c)\ \supset\ \delta(\langle a,b,c\rangle,b'+1,\vec{\alpha})\equiv \delta(\langle a,b,c\rangle,b',\vec{\alpha}). \quad (17)$$

This is done straightforwardly, namely we define δ by the axioms

$$\delta(a,0,\vec{\alpha})\equiv A(\Pi_1^3(a),0,\vec{\alpha},\{x,y\}\ -),$$

$$\begin{aligned}\delta(a,b+1,\vec{\alpha}) \equiv b < \tau^k(a) \wedge \big((\Pi_2^3(a) \le b \wedge \delta(a,b,\vec{\alpha}))\\ \vee\ (b+1 = \Pi_2^3(a) \le \tau^k(\Pi_1^3(a)) \wedge A(\Pi_1^3(a), \Pi_2^3(a), \vec{\alpha},\\ \{x,y\}(x \le \Pi_1^3(a) \wedge y < \Pi_2^3(a) \wedge \delta(\langle x,y,\Pi_3^3(a)\rangle, b, \vec{\alpha}))))\big).\end{aligned}$$

This definition has the form required in (7) since $W_T^{0,\tau}(\delta) \vdash x \le \Pi_1^3(a) \wedge y < \Pi_2^3(a) \ \supset\ \langle x,y,\Pi_3^3(a)\rangle \le a$ due to the monotonicity of $\langle a_1,a_2,a_3\rangle$. Note also that $W_T^{0,\tau}(\delta) \vdash \tau^k(a) \ge \tau^k(\Pi_1^3(a))$ and $W_T^{0,\tau}(\delta) \vdash \tau^k(a) \ge \tau^k(\Pi_3^3(a))$ due to (8).

An obvious $\Sigma_0^{1,b}(\delta)$-induction on b' applied to (17) gives us

$$W_T^{0,\tau}(\delta) \vdash b \le b' \le \tau^k(c) \ \supset\ \delta(\langle a,b,c\rangle, b', \vec{\alpha}) \equiv \delta(\langle a,b,c\rangle, b, \vec{\alpha}).$$

Under the assumption $a \le c$ this allows us to replace $\delta(\langle x,y,c\rangle, b\dot{-}1, \vec{\alpha})$ in (16) by $\delta(\langle x,y,c\rangle, y, \vec{\alpha})$ to get first

$$\left.\begin{aligned}W_T^{0,\tau}(\delta) \vdash a \le c \supset \Big(\delta(\langle a,b,c\rangle, b, \vec{\alpha})\\ \equiv \Big(b \le \tau^k(a) \wedge A(a,b,\vec{\alpha},\\ \{x,y\}(x \le a \wedge y < b \wedge \delta(\langle x,y,c\rangle, y, \vec{\alpha})))\Big)\Big)\end{aligned}\right\} \quad (18)$$

and then, as a partial case

$$\left.\begin{aligned}W_T^{0,\tau}(\delta) \vdash \delta(\langle a,b,a\rangle, b, \vec{\alpha})\\ \equiv \Big(b \le \tau^k(a) \ \wedge\ A(a,b,\vec{\alpha},\\ \{x,y\}(x \le a \wedge y < b \wedge \delta(\langle x,y,a\rangle, y, \vec{\alpha})))\Big).\end{aligned}\right\} \quad (19)$$

Using (18), we prove by $\Sigma_0^{1,b}(\delta) - IND$ on b that

$$W_T^{0,\tau}(\delta) \vdash c' \le c \supset \forall x \le c'\,(\delta(\langle x,b,c\rangle, b, \vec{\alpha}) \equiv \delta(\langle x,b,c'\rangle, b, \vec{\alpha}))$$

which allows us to replace $\delta(\langle x,y,a\rangle, y, \vec{\alpha})$ in (19) with $\delta(\langle x,y,x\rangle, y, \vec{\alpha})$ and obtain the desired result with $D(a,b) := \delta(\langle a,b,a\rangle, b, \vec{\alpha})$. ∎

Now we strengthen Lemma 4.1 by enlarging the class of formulae A to which it is applicable.

Lemma 4.2. *Let $A(a,b,\beta^2) \in \Sigma_0^{1,b}(\delta)$, where all first order free variables are displayed, and*

$$\beta \text{ does not appear in the scope of } \delta\text{-symbols.} \quad (20)$$

Let k be a positive integer. Then there exists a $\Sigma_0^{1,b}(\delta)$-formula $D(a,b)$ such that

$$W_T^{0,\tau}(\delta) \vdash D(a,b) \;\equiv\; b \le \tau^k(a) \wedge A(a,b,\{x,y\}(x \le a \wedge y < b \wedge D(x,y))).$$

Proof. Once again, let $\vec{\alpha}$ be the complete list of free second order variables in A other than β. Let $\delta_1(a,b,\vec{\alpha}),\dots,\delta_l(a,b,\vec{\alpha})$ be the complete list of relationals appearing in A. Represent A in the form

$$A(a,b,\vec{\alpha},\beta) \equiv \bar{A}(a,b,\vec{\alpha},\beta,\{x,y\}\delta_i(x,y,\vec{\alpha})),$$

where $\bar{A}(a,b,\vec{\alpha},\beta,\gamma_1,\dots,\gamma_l) \in \Sigma_0^{1,b}$. Similarly to [6, Lemma 10.9] and [20, Theorem 3.6 a)], we have a term $T(a,b)$ (which may be additionally assumed to be provably monotone) such that

$$\left.\begin{aligned} &W_T^{0,\tau}(\delta) \vdash \\ &\quad \forall x \le T(a,b)\forall y\Big(\gamma_1(x,y) \equiv \gamma_1'(x,y) \wedge \dots \\ &\qquad\qquad \wedge\ \gamma_l(x,y) \equiv \gamma_l'(x,y)\Big) \\ &\qquad \supset\ \bar{A}(a,b,\vec{\alpha},\beta,\gamma_1,\dots,\gamma_l) \equiv \bar{A}(a,b,\vec{\alpha},\beta,\gamma_1',\dots,\gamma_l'). \end{aligned}\right\} \quad (21)$$

Let $T_1(a) \rightleftharpoons T(a,\tau^k(a))$, $T_2(a) \rightleftharpoons \langle \boldsymbol{l}, T_1(a)\rangle + 4$ (where $\boldsymbol{l}$ is a closed term representing the integer l), $f(a) \rightleftharpoons \langle 0, \langle a, T_2(a)\rangle\rangle$ and, finally, $g(a,b) \rightleftharpoons \tau^{k'}(T_1(a)) + b + 1$, where k' is the maximal integer among the exponents $k_1,\dots,k_l$ involved in the definitions of $\delta_1,\dots,\delta_l$.

It is straightforward to check that we may define, in accordance with Lemma 4.1 (with $k := \max(k',k)+1$) a $\Sigma_0^{1,b}(\delta)$-formula $D^*(a,b)$ possessing the following properties:

$$W_T^{0,\tau}(\delta) \vdash D^*(\langle \boldsymbol{i}, a\rangle, b) \equiv \delta_i(a,b,\vec{\alpha}) \ (1 \le i \le l), \quad (22)$$

$$\left.\begin{aligned} W_T^{0,\tau}(\delta) \vdash D^*(f(a),g(a,b)) \;\equiv\; & b \le \tau^k(a) \wedge \bar{A}(a,b,\vec{\alpha}, \\ & \{x,y\}(x \le a \wedge y < b \wedge D^*(f(x),g(x,y))), \\ & \{x,y\}(x \le T_1(a) \wedge y \le \tau^{k_i}(x) \wedge D^*(\langle \boldsymbol{i}, x\rangle, y))). \end{aligned}\right\} \quad (23)$$

Due to (22), we may replace $D^*(\langle \boldsymbol{i}, x\rangle, y)$ in (23) by $\delta_i(x,y,\vec{\alpha})$. Then we can drop the term $y \le \tau^{k_i}(x)$ (as $W_T^{0,\tau}(\delta) \vdash \delta_i(a,b,\vec{\alpha}) \supset b \le \tau^{k_i}(a)$) and $x \le T_1(a)$ (due to (21)). These simplifications lead to

$$\begin{aligned} &W_T^{0,\tau}(\delta) \vdash D^*(f(a),g(a,b)) \\ &\quad \equiv (b \le \tau^k(a) \wedge A(a,b,\vec{\alpha}, \\ &\qquad \{x,y\}(x \le a \wedge y < b \ \wedge\ D^*(f(x),g(x,y))))) \end{aligned}$$

which is exactly what we need (with $D(a,b) := D^*(f(a),g(a,b))$). ∎

Corollary 4.3. *Let $A(a), B(a,b,\beta^2) \in \Sigma_0^{1,b}(\delta)$, where all first order free variables are displayed, and k be a positive integer. Assume that* (20) *holds. Then there exists a $\Sigma_0^{1,b}(\delta)$-formula $D(a,b)$ such that*

$$\begin{aligned} W_T^{0,\tau}(\delta) \vdash D(a,0) &\equiv A(a), \\ W_T^{0,\tau}(\delta) \vdash D(a,b+1) &\equiv b < \tau^k(a) \wedge B(a,b,\{x\}(x \le a \wedge D(x,b))). \end{aligned}$$

Now we are in position to define substitutions of arbitrary $\Sigma_0^{1,b}(\delta)$-abstracts into formulae. Namely, let C be a formula of $\mathcal{L}^\tau(\delta)$, $\gamma_1, \ldots, \gamma_r$ be second-order free variables, and $V_1, \ldots, V_r$ be $\Sigma_0^{1,b}(\delta)$-abstracts of the corresponding arities. We define $C\,[V_1/\gamma_1, \ldots, V_r/\gamma_r]$ by induction on complexity of C.

If $C \equiv \delta^k_{A,B}(t,s,\vec{\alpha})$, then we let

$$\delta^k_{A,B}(t,s,\vec{\alpha})\left[\vec{V}/\vec{\gamma}\right] \rightleftharpoons D(t,s),$$

where $D(a,b)$ is the $\Sigma_0^{1,b}(\delta)$-formula such that

$$W_T^{0,\tau}(\delta) \vdash D(a,0) \equiv A(a,\vec{\alpha})\left[\vec{V}/\vec{\gamma}\right] \tag{24}$$

and

$$\left.\begin{aligned} W_T^{0,\tau}(\delta) \vdash D(a,b+1) \equiv \Big(& b < \tau^k(a) \\ & \wedge B(a,b,\vec{\alpha},\beta)\left[\vec{V}/\vec{\gamma}, \{x\}(x \le a \wedge D(x,b))/\beta\right]\Big) \end{aligned}\right\} \tag{25}$$

defined in accordance with Corollary 4.3. We assume here that a, b, β do not occur in $\vec{\gamma}, \vec{V}$ which, in particular, implies (20) for $B(a,b,\vec{\alpha},\beta)\left[\vec{V}/\vec{\gamma}\right]$.

All other cases in the recursive definition of $C\,[V_1/\gamma_1, \ldots, V_r/\gamma_r]$ are treated in the standard way (see e.g. [6, §9.1]).

The axiomatic properties of $C\,[V_1/\gamma_1, \ldots, V_r/\gamma_r]$ needed for our purposes are summarized in the following easy lemma.

Lemma 4.4. **a)** *if C does not contain relationals δ then $C\left[\vec{V}/\vec{\gamma}\right]$ coincides with the usual definition,*

b) *if*

$$W_T^{0,\tau}(\delta) \vdash C_1, \ldots, C_r \longrightarrow D_1, \ldots, D_s \tag{26}$$

then

$$\left.\begin{aligned} W_T^{0,\tau}(\delta) \vdash C_1\left[\vec{V}/\vec{\gamma}\right], \ldots, C_r\left[\vec{V}/\vec{\gamma}\right] \\ \longrightarrow D_1\left[\vec{V}/\vec{\gamma}\right], \ldots, D_s\left[\vec{V}/\vec{\gamma}\right], \end{aligned}\right\} \quad (27)$$

c) $C\left[\vec{V}/\vec{\gamma}\right]\left[\vec{t}/\vec{a}\right] = C\left[\vec{t}/\vec{a}\right]\left[\vec{V}/\vec{\gamma}\right]$, *where variables from the list* $\vec{a}$ *do not occur in* $\vec{V}$,

d) *let* $C, D_1\left(\vec{a}^{(1)}\right), \ldots, D_r\left(\vec{a}^{(r)}\right) \in \Sigma_0^{1,b}(\delta)$; $V_i \rightleftharpoons \left\{\vec{x}^{(i)}\right\} D_i\left(\vec{x}^{(i)}\right)$ $(1 \leq i \leq r)$, $\vec{\eta}$ *be a vector of second order variables not appearing in* C, *and* $\vec{W}$ *be* $\Sigma_0^{1,b}(\delta)$*-abstracts. Then*

$$\left.\begin{aligned} W_T^{0,\tau}(\delta) \vdash C\left[\vec{V}/\vec{\gamma}\right]\left[\vec{W}/\vec{\eta}\right] \\ \equiv C\left[\left\{\vec{x}^{(i)}\right\}\left(D_i\left(\vec{x}^{(i)}\right)\left[\vec{W}/\vec{\eta}\right]\right)/\gamma_i\right]. \end{aligned}\right\} \quad (28)$$

Proof. a) is obvious.

b). Every proof in $W_T^{0,\tau}(\delta)$ of the sequent (26) can be converted into a proof of (27) after substituting $\vec{V}$ for $\vec{\gamma}$ into it if we note that the axioms (7) are taken by this substitution exactly to (24), (25).

c). By obvious induction on the complexity of C.

d). Induction on the complexity of C. In fact, all cases are straightforward (for $C = \exists x \leq t\ E(x)$ use already proven part c)) except for the base case $C = \delta_{A,B}^k(t, s, \vec{\alpha})$. In particular, we already have (28) when $C \in \Sigma_0^{1,b}$.

In the remaining case $C = \delta_{A,B}^k(t, s, \vec{\alpha})$ we may assume w.l.o.g. that

$$C(a, b) = \delta_{A,B}^k(a, b, \vec{\alpha}).$$

Then we have from definitions

$$\begin{aligned} &W_T^{0,\tau}(\delta) \vdash C(a, 0)\left[\vec{V}/\vec{\gamma}\right] \equiv A(a, \vec{\alpha})\left[\vec{V}/\vec{\gamma}\right] \\ &W_T^{0,\tau}(\delta) \vdash C(a, b+1)\left[\vec{V}/\vec{\gamma}\right] \\ &\quad \equiv \Big(b < \tau^k(a) \wedge B(a, b, \vec{\alpha}, \beta)\Big[\vec{V}/\vec{\gamma}, \\ &\qquad \{x\}\Big(x \leq a \wedge C(x, b)\left[\vec{V}/\vec{\gamma}\right]\Big)/\beta\Big]\Big) \end{aligned}$$

which, by the already proven part b), implies

$$W_T^{0,\tau}(\delta) \vdash C(a, 0)\left[\vec{V}/\vec{\gamma}\right]\left[\vec{W}/\vec{\eta}\right] \equiv A(a, \vec{\alpha})\left[\vec{V}/\vec{\gamma}\right]\left[\vec{W}/\vec{\eta}\right]$$

$$\begin{aligned} W_T^{0,\tau}(\delta) &\vdash C(a,b+1)\left[\vec{V}/\vec{\gamma}\right]\left[\vec{W}/\vec{\eta}\right] \\ &\equiv \Big(b < \tau^k(a) \wedge B(a,b,\vec{\alpha},\beta)\Big[\vec{V}/\vec{\gamma}, \\ &\qquad \{x\}\left(x \le a \wedge\ C(x,b)\left[\vec{V}/\vec{\gamma}\right]\right)/\beta\Big]\left[\vec{W}/\vec{\eta}\right]\Big). \end{aligned}$$

A and B, however, are in $\Sigma_0^{1,b}$. Hence, since for formulae with this restriction (28) is already established, we have

$$\begin{aligned} W_T^{0,\tau}(\delta) &\vdash C(a,0)\left[\vec{V}/\vec{\gamma}\right]\left[\vec{W}/\vec{\eta}\right] \\ &\equiv A(a,\vec{\alpha})\left[\left\{\vec{x}^{(i)}\right\}\left(D_i\left(\vec{x}^{(i)}\right)\left[\vec{W}/\vec{\eta}\right]\right)/\gamma_i\right] \end{aligned}$$

and

$$\begin{aligned} W_T^{0,\tau}(\delta) &\vdash C(a,b+1)\left[\vec{V}/\vec{\gamma}\right]\left[\vec{W}/\vec{\eta}\right] \\ &\equiv \Big(b < \tau^k(a)\ \wedge\ B(a,b,\vec{\alpha},\beta)\Big[\left\{\vec{x}^{(i)}\right\}\left(D_i\left(\vec{x}^{(i)}\right)\left[\vec{W}/\vec{\eta}\right]\right)/\gamma_i, \\ &\qquad \{x\}\Big(x \le a\ \wedge\ C(x,b)\left[\vec{V}/\vec{\gamma}\right]\left[\vec{W}/\vec{\eta}\right]\Big)/\beta\Big]\Big). \end{aligned}$$

Comparing these with the definitions of

$$C(a,b)\left[\left\{\vec{x}^{(i)}\right\}\left(D_i\left(\vec{x}^{(i)}\right)\left[\vec{W}/\vec{\eta}\right]\right)/\gamma_i\right],$$

we, by a straightforward induction on b, establish

$$\begin{aligned} W_T^{0,\tau}(\delta) &\vdash \forall x \le a\Big(C(x,b)\left[\vec{V}/\vec{\gamma}\right]\left[\vec{W}/\vec{\eta}\right] \\ &\equiv C(x,b)\left[\left\{\vec{x}^{(i)}\right\}\left(D_i\left(\vec{x}^{(i)}\right)\left[\vec{W}/\vec{\eta}\right]\right)/\gamma_i\right]\Big) \end{aligned}$$

which immediately gives the desired result. ∎

Now we can get rid of the restriction (20) in Lemma 4.2.

Lemma 4.5. *Let $A(a,b,\beta^2) \in \Sigma_0^{1,b}(\delta)$, where all first order free variables are displayed, and k be a positive integer. Then there exists a $\Sigma_0^{1,b}(\delta)$-formula $D(a,b)$ such that*

$$W_T^{0,\tau}(\delta) \vdash D(a,b)\ \equiv\ b \le \tau^k(a) \wedge A(a,b,\{x,y\}(x \le a \wedge y < b \wedge D(x,y))).$$

Proof. We use the notation introduced in the proof of Lemma 4.2. The only difference is that this time the relationals $\delta_1,\ldots,\delta_l$ may also depend on β. Write down explicitly their defining axioms:

$$\delta_i(a_1,0,\vec{\alpha},\beta) \equiv A_i(a_1,\vec{\alpha},\beta), \tag{29}$$

$$\left.\begin{aligned}\delta_i(a_1, b_1+1, \vec{\alpha}, \beta) \quad \equiv \quad & b_1 < \tau^{k_i}(a_1) \wedge B_i(a_1, b_1, \vec{\alpha}, \beta, \\ & \{x_1\}(x_1 \leq a_1 \wedge \delta_i(x_1, b_1, \vec{\alpha}, \beta))).\end{aligned}\right\} \quad (30)$$

Let

$$\begin{aligned}\Theta_i(a, a_1) &\rightleftharpoons \langle \boldsymbol{i}, a_1\rangle + T_1(a) + 12 \ (1 \leq i \leq l), \\ \Theta(a) &\rightleftharpoons \langle 0, T_2(a)\rangle + T_1(a) + 12, \\ f_i(a, b, a_1) &\rightleftharpoons \langle a, b, b \cdot \Theta(a) + \Theta_i(a, a_1)\rangle, \\ f(a, b) &\rightleftharpoons \langle a, b, b \cdot \Theta(a) + \Theta(a)\rangle, \\ g_i(a, b, b_1) &\rightleftharpoons b \cdot \left(\tau^{k'}(T_1(a)) + 2\right) + b_1, \\ g(a, b) &\rightleftharpoons (b+1) \cdot \left(\tau^{k'}(T_1(a)) + 2\right) \dot{-} 1.\end{aligned}$$

We define, in accordance with Lemma 4.1 (but this time with $k := k + k' + 1$) a $\Sigma_0^{1,b}(\delta)$-formula $D^*(a, b)$ which has the following properties:

$$\left.\begin{aligned}W_T^{0,\tau}(\delta) \vdash b \leq \tau^k(a) \supset \Big(& D^*(f_i(a, b, a_1), g_i(a, b, 0)) \\ \equiv A_i(a_1, \vec{\alpha}, \{x, y\}(& x \leq a \wedge y < b \\ & \wedge D^*(f(x, y), g(x, y))))\Big) \ (1 \leq i \leq l),\end{aligned}\right\} \quad (31)$$

$$\left.\begin{aligned}W_T^{0,\tau}(\delta) \vdash b \leq \tau^k(a) \supset \Big(& D^*(f_i(a, b, a_1), g_i(a, b, b_1+1)) \\ \equiv \ & b_1 < \tau^{k_i}(a_1) \wedge B_i(a_1, b_1, \vec{\alpha}, \\ & \{x, y\}(x \leq a \wedge y < b \wedge D^*(f(x, y), g(x, y))), \\ & \{x_1\}(x_1 \leq a_1 \wedge \ D^*(f_i(a, b, x_1), g_i(a, b, b_1))))\Big) \\ & (1 \leq i \leq l),\end{aligned}\right\} \quad (32)$$

$$\left.\begin{aligned}W_T^{0,\tau}(\delta) \vdash D^*(f(a, b), g(a, b)) \ \equiv \ & b \leq \tau^k(a) \wedge \bar{A}(a, b, \vec{\alpha}, \\ & \{x, y\}(x \leq a \wedge y < b \wedge D^*(f(x, y), g(x, y))), \\ & \{x_1, y_1\}(x_1 \leq T_1(a) \wedge y_1 \leq \tau^{k_i}(x_1) \\ & \wedge D^*(f_i(a, b, x_1), g_i(a, b, y_1)))).\end{aligned}\right\} \quad (33)$$

Now, using Lemma 4.4 b), we substitute the abstract

$$\{x, y\}(x \leq a \wedge y < b \wedge D^*(f(x, y), g(x, y)))$$

for β into (29), (30). Comparing the result of this substitution with (31), (32), we readily prove by $\Sigma_0^{1,b}(\delta) - IND$ that

$$W_T^{0,\tau}(\delta) \vdash b \leq \tau^k(a) \supset \Big(D^*(f_i(a,b,a_1), g_i(a,b,b_1)) \\ \equiv \delta_i(a_1, b_1, \vec{\alpha}, \{x,y\}(x \leq a \wedge y < b \wedge D^*(f(x,y), g(x,y)))))\Big).$$

This allows us to transform, like in the proof of Lemma 4.2, the right-hand side of (33) to

$$b \leq \tau^k(a) \wedge A(a, b, \vec{\alpha}, \{x,y\}(x \leq a \wedge y < b \wedge D^*(f(x,y), g(x,y))))$$

which immediately gives us the desired result with

$$D(a,b) := D^*(f(a,b), g(a,b)).$$

■

Finally, we convert Lemma 4.5 to the following form of simultaneous induction, which will be convenient in the next section.

Lemma 4.6. *Let*

$$A_i(a_1, \ldots, a_{r_i}), B_i(a_1, \ldots, a_{r_i}, b, \beta_1^{r_1}, \ldots, \beta_l^{r_l}) \in \Sigma_0^{1,b}(\delta) \ (1 \leq i \leq l),$$

and t, $s(b)$ be first order terms, where all occurrences of the variables $\vec{a}, b, \vec{\beta}$ are explicitly displayed. Assume also that k is a positive integer. Then there exist $\Sigma_0^{1,b}(\delta)$-formulae $D_i(a_1, \ldots, a_{r_i}, b)$ $(1 \leq i \leq l)$ such that

$$W_T^{0,\tau}(\delta) \vdash D_i(a_1, \ldots, a_{r_i}, 0) \equiv A_i(a_1, \ldots, a_{r_i}) \ (1 \leq i \leq l),$$
$$W_T^{0,\tau}(\delta) \vdash D_i(a_1, \ldots, a_{r_i}, b+1) \equiv \Big(b < \tau^k(t) \\ \wedge B_i\Big(a_1, \ldots, a_{r_i}, b, \Big\{\vec{x}^{(1)}\Big\}\Big(\vec{x}^{(1)} \leq s(b) \wedge D_1\Big(\vec{x}^{(1)}, b\Big)\Big), \ldots, \\ \Big\{\vec{x}^{(l)}\Big\}\Big(\vec{x}^{(l)} \leq s(b) \wedge D_l\Big(\vec{x}^{(l)}, b\Big)\Big)\Big)\Big) \ (1 \leq i \leq l).$$

Proof. Let $c_1, \ldots, c_n, \vec{\alpha}$ be the complete list of free variables appearing in $A_i, B_i, t, s(b)$ other than $\vec{a}, b, \vec{\beta}$. W.l.o.g. we may assume that $t(\vec{c})$ and $s(b, \vec{c})$ are provably monotone. We set

$$\begin{aligned} T_1(\vec{c}) &\rightleftharpoons s\left(\tau^k(t(\vec{c})), \vec{c}\right), \\ T_2(\vec{c}) &\rightleftharpoons \sum_{i=1}^{l} \langle \boldsymbol{i}, \langle \underbrace{T_1(\vec{c}), \ldots, T_1(\vec{c})}_{r_i} \rangle \rangle, \\ f_i(a_1, \ldots, a_{r_i}, b, \vec{c}) &\rightleftharpoons \langle b \cdot T_2(\vec{c}) + \langle \boldsymbol{i}, \langle a_1, \ldots, a_{r_i} \rangle \rangle, c_1, \ldots, c_n, t(\vec{c}) \rangle. \end{aligned}$$

Then we define, in accordance with Lemma 4.5, a formula $D(a,b)$ with the properties

$$W_T^{0,\tau}(\delta) \vdash D(f_i(a_1,\ldots,a_{r_i},0,\vec{c}),0) \equiv A_i(a_1,\ldots,a_{r_i}) \ (1 \le i \le l),$$
$$W_T^{0,\tau}(\delta) \vdash D(f_i(a_1,\ldots,a_{r_i},b+1,\vec{c}),b+1) \equiv \Big(b < \tau^k(t(\vec{c}))$$
$$\wedge\ B_i\Big(a_1,\ldots,a_{r_i},b,$$
$$\Big\{\vec{x}^{(1)}\Big\}\Big(\vec{x}^{(1)} \le s(b,\vec{c}) \wedge D\Big(f_1\Big(\vec{x}^{(1)},b,\vec{c}\Big),b\Big)\Big),\ldots,$$
$$\Big\{\vec{x}^{(l)}\Big\}\Big(\vec{x}^{(l)} \le s(b,\vec{c}) \wedge D\Big(f_l\Big(\vec{x}^{(l)},b,\vec{c}\Big),b\Big)\Big)\Big)\Big) \ (1 \le i \le l)$$

and let $D_i(a_1,\ldots,a_{r_i},b) := D(f_i(a_1,\ldots,a_{r_i},b,\vec{c}),b)$. ∎

5. The witnessing lemma

In this section we will show that if $W_T^{1,\tau} \vdash A \supset B$, where $A,B \in \Sigma_1^{1,b}$, then this fact can be witnessed *in* $W_T^{0,\tau}(\delta)$ by a family of $\Sigma_0^{1,b}(\delta)$-abstracts. The proof goes more or less along the same lines as the proof of [6, Theorems 10.12, 10.16]. The main difference is that we are interested not only in the computational complexity of the witnessing formulae, but also in removing from the proof second order quantifiers. This will alter our definitions: we prefer to take them as clear as possible syntactically rather than semantically.

Throughout this section second order variables $\gamma_1,\ldots,\gamma_l,\ldots$ will stand for special *witnessing variables*. We will always assume that they do not occur in original $\Sigma_1^{1,b}$-formulae denoted by capital latin letters like A_i, B_i, C, D etc.

5.1. Definition. Given a formula $A \in \Sigma_1^{1,b}$, we define its *witnessing formula* $W_A \in \Sigma_0^{1,b}$ as follows.

a) if A is atomic or negation of an atomic formula, then $W_A \rightleftharpoons A$,

b) if $A = B \wedge C$ or $A = B \vee C$ then we rename if necessary witnessing variables in W_B and W_C so that no variable appears in both of them, and let $W_A \rightleftharpoons W_B \wedge W_C$ [$W_B \vee W_C$, respectively],

c) if $A = \exists x \le tB(x)$ then $W_A \rightleftharpoons \exists x \le tW_{B(a)}[x/a]$,

d) if $A = \forall x \le tB(x)$ then

$$W_A \rightleftharpoons \forall x \le t\ W_{B(a)}\left[\{y_1,\ldots,y_{r_i}\}\gamma_i^{r_i+1}(a,y_1,\ldots,y_{r_i})/\gamma_i^{r_i}\right][x/a],$$

where the second order substitution is extended over all witnessing variables $\gamma_i^{r_i}$ appearing in $W_{B(a)}$,

e) if $A = \exists\phi B(\phi)$ then $W_A \rightleftharpoons W_{B(\alpha)}[\gamma/\alpha]$, where γ is a new witnessing variable which did not appear in $W_{B(\alpha)}$,

f) if $A = \neg B$, where $B \in \Pi_1^{1,b}$, use prenex operations to convert A to the form where negations appear on atomic formulae only and handle the result of this conversion in accordance to cases a)- e) above.

The sole purpose of our variant of the definition of the witnessing formula is to move second order quantifiers in front in the simplest possible way. Note that W_A is defined uniquely up to renaming witnessing variables, and that $W_A = A$ for $A \in \Sigma_0^{1,b}$.

Lemma 5.2. *Suppose that*

$$W_T^{1,\tau} \vdash A_1, \ldots, A_k, B_1, \ldots, B_l \longrightarrow A_{k+1}, \ldots, A_m, B_{l+1}, \ldots, B_n, \qquad (34)$$

where $A_1, \ldots, A_k, B_{l+1}, \ldots, B_n \in \Sigma_1^{1,b}$ *and* $B_1, \ldots, B_l, A_{k+1}, \ldots, A_m \in \Pi_1^{1,b}$. *Then there exist* $\Sigma_0^{1,b}(\delta)$*-abstracts* $\vec{V}^{(1)}, \ldots, \vec{V}^{(n)}$ *such that*

$$\left.\begin{aligned} &W_T^{0,\tau}(\delta) \vdash W_{A_1}, \ldots, W_{A_k}, W_{\neg A_{k+1}}, \ldots, W_{\neg A_m} \\ &\quad \longrightarrow W_{\neg B_1}\left[\vec{V}^{(1)}/\vec{\gamma}^{(1)}\right], \ldots, W_{\neg B_l}\left[\vec{V}^{(l)}/\vec{\gamma}^{(l)}\right], \\ &\qquad W_{B_{l+1}}\left[\vec{V}^{(l+1)}/\vec{\gamma}^{(l+1)}\right], \ldots, W_{B_n}\left[\vec{V}^{(n)}/\vec{\gamma}^{(n)}\right], \end{aligned}\right\} \qquad (35)$$

where $\vec{\gamma}^{(1)}, \ldots, \vec{\gamma}^{(n)}$ *are complete lists of witnessing variables in* $W_{\neg B_1}, \ldots, W_{B_n}$ *and no witnessing variable appears in any two of the formulae* $W_{A_1}, \ldots, W_{B_n}$.

Proof. (cf. [6, Proof of Theorem 10.12]). Applying cosmetic rules ($\neg$:left), ($\neg$:right), we can assume that our sequent has the form

$$A_1, \ldots, A_m \longrightarrow B_1, \ldots, B_n$$

with $A_1, \ldots, A_m, B_1, \ldots, B_n \in \Sigma_1^{1,b}$. The Cut Elimination Theorem is readily extended to $W_T^{1,\tau}$ hence we may also assume that all formulae in the proof belong to $\Sigma_1^{1,b} \cup \Pi_1^{1,b}$. As usual, we apply induction on the complexity of the proof.

All axioms of $W_T^{1,\tau}$ are in $\Sigma_0^{1,b}$ (see part b) of Definition 2.1), and in this case (35) coincides with (34).

In our analysis of inference rules we omit many cases which either are obvious or can be treated similarly to previously considered cases.

(Weak:left) transforms to (Weak:left) in the theory $W_T^{0,\tau}(\delta)$.

(Weak:right) also transforms to itself (if B_{n+1} is the principal formula then the corresponding abstracts $\vec{V}^{(n+1)}$ can be chosen in an arbitrary way).

(Contraction:right). We have

$$W_T^{0,\tau}(\delta) \vdash W_{A_1}, \ldots, W_{A_m} \longrightarrow W_{B_1}\left[\vec{V}^{(1)}/\vec{\gamma}^{(1)}\right], \ldots,$$
$$W_{B_n}\left[\vec{V}^{(n)}/\vec{\gamma}^{(n)}\right], W_A\left[\vec{V}/\vec{\gamma}\right], W_A\left[\vec{V}'/\vec{\gamma}\right],$$

where, say, $V_i = \left\{\vec{x}^{(i)}\right\} C_i\left(\vec{x}^{(i)}\right)$ and $V_i' = \left\{\vec{x}^{(i)}\right\} C_i'\left(\vec{x}^{(i)}\right)$, and we want to find a single family of abstracts $\vec{V}''$ with the property

$$W_T^{0,\tau}(\delta) \vdash W_{A_1}, \ldots, W_{A_m}$$
$$\longrightarrow W_{B_1}\left[\vec{V}^{(1)}/\vec{\gamma}^{(1)}\right], \ldots, W_{B_n}\left[\vec{V}^{(n)}/\vec{\gamma}^{(n)}\right], W_A\left[\vec{V}''/\vec{\gamma}\right].$$

Clearly, the family defined by

$$V_i'' \rightleftharpoons \left\{\vec{x}^{(i)}\right\}\left(\left(W_A\left[\vec{V}/\vec{\gamma}\right] \wedge C_i\left(\vec{x}^{(i)}\right)\right) \vee \left(\neg W_A\left[\vec{V}/\vec{\gamma}\right] \wedge C_i'\left(\vec{x}^{(i)}\right)\right)\right)$$

will do.

($\wedge$:right). We have

$$W_T^{0,\tau}(\delta) \vdash W_{A_1}, \ldots, W_{A_m}$$
$$\longrightarrow W_{B_1}\left[\vec{V}^{(1)}/\vec{\gamma}^{(1)}\right], \ldots, W_{B_n}\left[\vec{V}^{(n)}/\vec{\gamma}^{(n)}\right], W_A\left[\vec{V}/\vec{\gamma}\right]$$

and

$$W_T^{0,\tau}(\delta) \vdash W_{A_1}, \ldots, W_{A_m}$$
$$\longrightarrow W_{B_1}\left[\vec{V}'^{(1)}/\vec{\gamma}^{(1)}\right], \ldots, W_{B_n}\left[\vec{V}'^{(n)}/\vec{\gamma}^{(n)}\right], W_B\left[\vec{V}'/\vec{\gamma}'\right],$$

and we want to find abstracts $\vec{V}''^{(1)}, \ldots, \vec{V}''^{(n)}$ such that

$$W_T^{0,\tau}(\delta) \vdash W_{A_1}, \ldots, W_{A_m} \longrightarrow W_{B_1}\left[\vec{V}''^{(1)}/\vec{\gamma}^{(1)}\right], \ldots,$$
$$W_{B_n}\left[\vec{V}''^{(n)}/\vec{\gamma}^{(n)}\right], W_A\left[\vec{V}/\vec{\gamma}\right] \wedge W_B\left[\vec{V}'/\vec{\gamma}'\right]$$

(we assume that all variables in $\vec{\gamma}, \vec{\gamma}'$ are pairwise distinct). This is done similarly to the case (Contraction:right). Namely, we let

$$V_j''^{(i)} \rightleftharpoons \left\{\vec{x}^{(ij)}\right\}\Big(\left(W_A\left[\vec{V}/\vec{\gamma}\right] \wedge C_j'^{(i)}\left(\vec{x}^{(ij)}\right)\right)$$
$$\vee \left(\neg W_A\left[\vec{V}/\vec{\gamma}\right] \wedge C_j^{(i)}\left(\vec{x}^{(ij)}\right)\right)\Big),$$

where $C_j^{(i)}, C_j'^{(i)}$ are the $\Sigma_0^{1,b}(\delta)$-formulae defining the abstracts $V_j^{(i)}, V_j'^{(i)}$ respectively.

($\forall \leq$:left). We have

$$\left.\begin{array}{l} W_T^{0,\tau}(\delta) \vdash W_{A(t)}, W_{A_1}, \ldots, W_{A_m} \\ \quad \longrightarrow W_{B_1}\left[\vec{V}^{(1)}/\vec{\gamma}^{(1)}\right], \ldots, W_{B_n}\left[\vec{V}^{(n)}/\vec{\gamma}^{(n)}\right], \end{array}\right\} \quad (36)$$

and we want to show

$$\begin{array}{l} W_T^{0,\tau}(\delta) \vdash t \leq s, \forall x \leq s W_{A(a)}[\{\vec{y}\}\vec{\gamma}(a,\vec{y})/\vec{\gamma}][x/a], W_{A_1}, \ldots, W_{A_m} \\ \quad \longrightarrow W_{B_1}\left[\vec{V}'^{(1)}/\vec{\gamma}^{(1)}\right], \ldots, W_{B_n}\left[\vec{V}'^{(n)}/\vec{\gamma}^{(n)}\right] \end{array}$$

for some $\vec{V}'^{(1)}, \ldots, \vec{V}'^{(n)}$. Here $\vec{\gamma}$ is the complete list of witnessing variables appearing in W_A.

For doing this we, using Lemma 4.4, substitute $\{\vec{y}\}\vec{\gamma}(t,\vec{y})$ for $\vec{\gamma}$ into (36) and find

$$\begin{array}{l} W_T^{0,\tau}(\delta) \vdash W_{A(a)}[\{\vec{y}\}\vec{\gamma}(a,\vec{y})/\vec{\gamma}][t/a], W_{A_1}, \ldots, W_{A_m} \\ \quad \longrightarrow W_{B_1}\left[\vec{V}'^{(1)}/\vec{\gamma}^{(1)}\right], \ldots, W_{B_n}\left[\vec{V}'^{(n)}/\vec{\gamma}^{(n)}\right], \end{array}$$

where

$$V_j'^{(i)} \rightleftharpoons \left\{\vec{x}^{(ij)}\right\}\left(C_j^{(i)}\left(\vec{x}^{(ij)}\right)[\{\vec{y}\}\vec{\gamma}(t,\vec{y})/\vec{\gamma}]\right).$$

Then we apply ($\forall \leq$:left) in the theory $W_T^{0,\tau}(\delta)$.

($\forall \leq$:right). We have

$$\left.\begin{array}{l} W_T^{0,\tau}(\delta) \vdash a \leq t, W_{A_1}, \ldots, W_{A_m} \\ \quad \longrightarrow W_{B_1}\left[\vec{V}^{(1)}(a)/\vec{\gamma}^{(1)}\right], \ldots, W_{B_n}\left[\vec{V}^{(n)}(a)/\vec{\gamma}^{(n)}\right], \\ \qquad W_{A(a)}\left[\vec{V}(a)/\vec{\gamma}\right], \end{array}\right\} \quad (37)$$

where all possible occurrences of the eigenvariable a are displayed, and it suffices to show that

$$\left.\begin{array}{l} W_T^{0,\tau}(\delta) \vdash W_{A_1}, \ldots, W_{A_m} \longrightarrow W_{B_1}\left[\vec{V}'^{(1)}/\vec{\gamma}^{(1)}\right], \ldots, \\ \quad W_{B_n}\left[\vec{V}'^{(n)}/\vec{\gamma}^{(n)}\right], \forall x \leq t W_{A(a)}\left[\vec{V}(a)/\vec{\gamma}\right][x/a]. \end{array}\right\} \quad (38)$$

Indeed,

$$W_{\forall x \leq t A(x)} = \forall x \leq t W_{A(a)}[\{\vec{y}\}\vec{\gamma}(a,\vec{y})/\vec{\gamma}][x/a],$$

hence

$$W_T^{0,\tau}(\delta) \vdash W_{\forall x \le tA(x)} \left[\{x, \vec{y}\} \vec{C}(x, \vec{y}) / \vec{\gamma} \right] \equiv \forall x \le t W_{A(a)} \left[\vec{V}(a) / \vec{\gamma} \right] [x/a]$$

by Lemma 4.4, where, as usual, $V_i(a) = \{\vec{y}\} C_i(a, \vec{y})$.

Let $D(a) \rightleftharpoons \neg W_{A(a)} \left[\vec{V}(a)/\vec{\gamma} \right] \wedge \forall x < a W_{A(a)} \left[\vec{V}(a)/\vec{\gamma} \right] [x/a]$. We modify the abstracts $V_j^{(i)}(a) = \left\{ \vec{y}^{(ij)} \right\} C_j^{(i)} \left(a, \vec{y}^{(ij)} \right)$ to

$$V_j'^{(i)} \rightleftharpoons \left\{ \vec{y}^{(ij)} \right\} \left(\exists x \le t \left(D(x) \wedge C_j^{(i)} \left(x, \vec{y}^{(ij)} \right) \right) \right).$$

In order to see (38), note (arguing informally in $W_T^{0,\tau}(\delta)$) that if

$$\exists x \le t \neg W_{A(a)} \left[\vec{V}(a)/\vec{\gamma} \right] [x/a]$$

then, by $\Sigma_0^{1,b}(\delta) - IND$, there exists the minimal $a \le t$ with the property

$$\neg W_{A(a)} \left[\vec{V}(a)/\vec{\gamma} \right].$$

This a is the unique $x \le t$ satisfying $D(x)$, hence $V_j'^{(i)}$ becomes equivalent to $V_j^{(i)}(a)$ which allows us to apply (37).

(Cut). We have

$$\begin{aligned} & W_T^{0,\tau}(\delta) \vdash W_{A_1}, \ldots, W_{A_m} \\ & \quad \longrightarrow W_{B_1} \left[\vec{V}^{(1)}/\vec{\gamma}^{(1)} \right], \ldots, W_{B_n} \left[\vec{V}^{(n)}/\vec{\gamma}^{(n)} \right], W_A \left[\vec{V}/\vec{\gamma} \right] \end{aligned}$$

and

$$\left. \begin{aligned} & W_T^{0,\tau}(\delta) \vdash W_A, W_{A'_1}, \ldots, W_{A'_{m'}} \\ & \quad \longrightarrow W_{B'_1} \left[\vec{V}'^{(1)}/\vec{\gamma}'^{(1)} \right], \ldots, W_{B'_{n'}} \left[\vec{V}'^{(n')}/\vec{\gamma}'^{(n')} \right], \end{aligned} \right\} \quad (39)$$

and we are going to deduce

$$\begin{aligned} & W_T^{0,\tau}(\delta) \vdash W_{A_1}, \ldots, W_{A_m}, W_{A'_1}, \ldots, W_{A'_{m'}} \longrightarrow W_{B_1} \left[\vec{V}^{(1)}/\vec{\gamma}^{(1)} \right], \ldots, \\ & \quad W_{B_n} \left[\vec{V}^{(n)}/\vec{\gamma}^{(n)} \right], W_{B'_1} \left[\vec{V}''^{(1)}/\vec{\gamma}'^{(1)} \right], \ldots, W_{B'_{n'}} \left[\vec{V}''^{(n')}/\vec{\gamma}'^{(n')} \right] \end{aligned}$$

for some (possibly new) abstracts $\vec{V}''^{(1)}, \ldots, \vec{V}''^{(n')}$. For doing this we substitute $\vec{V}$ for $\vec{\gamma}$ into (39) (using Lemma 4.4, of course), rename witnessing variables if necessary and apply (Cut) in the theory $W_T^{0,\tau}(\delta)$.

(second order $\forall$:left). This case is actually impossible (since, due to our convention, $A_1, \ldots, A_m, B_1, \ldots, B_n$ are in $\Sigma_1^{1,b}$).

(second order $\exists$:left) amounts to declaring the eigenvariable α as a witnessing variable.

(second order $\exists$:right, $\Sigma_0^{1,b} - CA$). We have

$$\begin{aligned} & W_T^{0,\tau}(\delta) \vdash W_{A_1}, \ldots, W_{A_m} \\ & \quad \longrightarrow W_{B_1}\left[\vec{V}^{(1)}/\vec{\gamma}^{(1)}\right], \ldots, W_{B_n}\left[\vec{V}^{(n)}/\vec{\gamma}^{(n)}\right], W_{F(V)}\left[\vec{V}/\vec{\gamma}\right]. \end{aligned}$$

It is easy to see, however, that

$$W_T^{0,\tau}(\delta) \vdash W_{F(V)}\left[\vec{V}/\vec{\gamma}\right] \equiv W_{F(\alpha)}[V/\alpha]\left[\vec{V}/\vec{\gamma}\right] \equiv W_{F(\alpha)}\left[V/\alpha, \vec{V}/\vec{\gamma}\right]$$

(since V itself does not contain witnessing variables). This is exactly what we need.

($\Sigma_1^{1,b} - \tau - IND$). We have

$$\left.\begin{aligned} & W_T^{0,\tau}(\delta) \vdash W_{A_1}, \ldots, W_{A_m}, W_{A(b)}(\vec{\gamma}) \\ & \quad \longrightarrow W_{A(b+1)}(\vec{V}(b,\vec{\gamma})), W_{B_1}\left[\vec{V}^{(1)}(b,\vec{\gamma})/\vec{\gamma}^{(1)}\right], \ldots, \\ & \qquad W_{B_n}\left[\vec{V}^{(n)}(b,\vec{\gamma})/\vec{\gamma}^{(n)}\right], \end{aligned}\right\} \quad (40)$$

where we displayed all occurrences of the eigenvariable b and the witnessing variables $\vec{\gamma}$ of W_A. We wish to deduce

$$\left.\begin{aligned} & W_T^{0,\tau}(\delta) \vdash W_{A_1}, \ldots, W_{A_m}, W_{A(0)}(\vec{\gamma}) \\ & \quad \longrightarrow W_{A(\tau(t))}(\vec{V}'(\vec{\gamma})), W_{B_1}\left[\vec{V}'^{(1)}(\vec{\gamma})/\vec{\gamma}^{(1)}\right], \ldots, \\ & \qquad W_{B_n}\left[\vec{V}'^{(n)}(\vec{\gamma})/\vec{\gamma}^{(n)}\right] \end{aligned}\right\} \quad (41)$$

for some $\vec{V}', \vec{V}'^{(1)}, \ldots, \vec{V}'^{(n)}$.[2]

Let $\vec{\gamma} = \gamma_1^{r_1}, \ldots, \gamma_l^{r_l}$ and $V_i(b,\vec{\gamma}) = \{x_1, \ldots, x_{r_i}\}\, C_i\,(x_1, \ldots, x_{r_i}, b, \vec{\gamma})$.

[2]Note that if A is actually in $\Sigma_0^{1,b}$ then each of the two principal formulae $A(0), A(\tau(t))$ may in fact appear on the list $B_1, \ldots, B_l, A_{k+1}, \ldots, A_m$ in (34) rather than on the list $A_1, \ldots, A_k, B_{l+1}, \ldots, B_n$. However, in this case $W_{A(0)} = A(0)$ and $W_{A(\tau(t))} = A(\tau(t))$ do not contain witnessing variables hence w.l.o.g. $A(0), A(\tau(t))$ can be moved to the list $A_1, \ldots, A_k, B_{l+1}, \ldots, B_n$.

Choose a term $s(b)$ such that

$$\left.\begin{aligned} W_T^{0,\tau}(\delta) \vdash \bigwedge_{i=1}^{l} \forall \vec{x}^{(i)} \leq s(b) \left(\gamma_i\left(\vec{x}^{(i)}\right) \equiv \gamma_i'\left(\vec{x}^{(i)}\right)\right) \\ \supset \Big(W_{A(b)}(\vec{\gamma}) \equiv W_{A(b)}(\vec{\gamma}') \\ \wedge \bigwedge_{j=1}^{n} \Big(W_{B(j)}[\vec{V}^{(j)}(b,\vec{\gamma})/\vec{\gamma}^{(j)}] \\ \equiv W_{B(j)}[\vec{V}^{(j)}(b,\vec{\gamma}')/\vec{\gamma}^{(j)}]\Big)\Big). \end{aligned}\right\} \quad (42)$$

Now we apply Lemma 4.6 and find $\Sigma_0^{1,b}(\delta)$-formulae $D_i(a_1,\ldots,a_{r_i},b,\vec{\gamma})$ such that

$$W_T^{0,\tau}(\delta) \vdash D_i(a_1,\ldots,a_{r_i},0,\vec{\gamma}) \equiv \gamma_i(a_1,\ldots,a_{r_i}) \ (1 \leq i \leq l), \quad (43)$$

$$\left.\begin{aligned} W_T^{0,\tau}(\delta) \vdash D_i(a_1,\ldots,a_{r_i},b+1,\vec{\gamma}) \\ \equiv \Big(b < \tau(t) \wedge C_i\Big(a_1,\ldots,a_{r_i},b, \\ \left\{\vec{x}^{(1)}\right\}\left(\vec{x}^{(1)} \leq s(b) \wedge D_1\left(\vec{x}^{(1)},b,\vec{\gamma}\right)\right),\ldots, \\ \left\{\vec{x}^{(l)}\right\}\left(\vec{x}^{(l)} \leq s(b) \wedge D_l\left(\vec{x}^{(l)},b,\vec{\gamma}\right)\right)\Big)\Big) \ (1 \leq i \leq l). \end{aligned}\right\} \quad (44)$$

We let

$$V_i'(\vec{\gamma}) \rightleftharpoons \left\{\vec{x}^{(i)}\right\} D_i\left(\vec{x}^{(i)},\tau(t),\vec{\gamma}\right), \quad (45)$$

$$\begin{aligned} D(b,\vec{\gamma}) \rightleftharpoons \\ \forall y \leq b\ W_{A(b)}\Big[\left\{\vec{x}^{(1)}\right\} D_1\left(\vec{x}^{(1)},b,\vec{\gamma}\right),\ldots, \\ \left\{\vec{x}^{(l)}\right\} D_l\left(\vec{x}^{(l)},b,\vec{\gamma}\right)\Big][y/b] \\ \wedge \neg W_{A(b+1)}\Big(\left\{\vec{x}^{(1)}\right\} D_1\left(\vec{x}^{(1)},b+1,\vec{\gamma}\right),\ldots, \\ \left\{\vec{x}^{(l)}\right\} D_l\left(\vec{x}^{(l)},b+1,\vec{\gamma}\right)\Big) \end{aligned}$$

and

$$\begin{aligned} V_i'^{(j)}(\vec{\gamma}) \rightleftharpoons \left\{\vec{x}^{(ij)}\right\}\Big(\exists y \leq \tau(t)\Big(D(y,\vec{\gamma}) \\ \wedge C_i^{(j)}\Big(\vec{x}^{(ij)},y,\left\{\vec{z}^{(1)}\right\} D_1\left(\vec{z}^{(1)},y,\vec{\gamma}\right),\ldots, \\ \left\{\vec{z}^{(l)}\right\} D_l\left(\vec{z}^{(l)},y,\vec{\gamma}\right)\Big)\Big)\Big), \end{aligned}$$

where, as usual, $V_i^{(j)}(b,\vec{\gamma}) \equiv \left\{\vec{x}^{(ij)}\right\} C_i^{(j)}\left(\vec{x}^{(ij)}, b, \vec{\gamma}\right)$.

In order to show (41), we first substitute

$$\left\{\vec{x}^{(i)}\right\}\left(\vec{x}^{(i)} \leq s(b) \wedge D_i\left(\vec{x}^{(i)}, b, \vec{\gamma}\right)\right)$$

for γ_i into (40), apply (44) for converting

$$W_{A(b+1)}\left(\vec{V}\left(b, \left\{\vec{x}^{(i)}\right\}\left(\vec{x}^{(i)} \leq s(b) \wedge D_i\left(\vec{x}^{(i)}, b, \vec{\gamma}\right)\right)\right)\right)$$

into $W_{A(b+1)}\left(\left\{\vec{x}^{(i)}\right\} D_i\left(\vec{x}^{(i)}, b+1, \vec{\gamma}\right)\right)$ (under the assumption $b < \tau(t)$) and drop the restriction $\vec{x}^{(i)} \leq s(b)$ using (42). We will have

$$\begin{aligned}
&W_T^{0,\tau}(\delta) \vdash b < \tau(t), W_{A_1}, \ldots, W_{A_m}, W_{A(b)}\left(\left\{\vec{x}^{(i)}\right\}\left(D_i\left(\vec{x}^{(i)}, b, \vec{\gamma}\right)\right)\right) \\
&\quad \longrightarrow W_{A(b+1)}\left(\left\{\vec{x}^{(i)}\right\} D_i\left(\vec{x}^{(i)}, b+1, \vec{\gamma}\right)\right), \\
&\qquad W_{B_1}\left[\vec{V}^{(1)}\left(b, \left\{\vec{x}^{(i)}\right\} D_i\left(\vec{x}^{(i)}, b, \vec{\gamma}\right)\right) / \vec{\gamma}^{(1)}\right], \ldots, \\
&\qquad W_{B_n}\left[\vec{V}^{(n)}\left(b, \left\{\vec{x}^{(i)}\right\} D_i\left(\vec{x}^{(i)}, b, \vec{\gamma}\right)\right) / \vec{\gamma}^{(n)}\right].
\end{aligned}$$

Now, the same argument as in the case ($\forall \leq$:right) (implicitly involving $\Sigma_0^{1,b}(\delta) - IND$ on b applied to the formula $W_{A(b)}\left(\left\{\vec{x}^{(i)}\right\}\left(D_i\left(\vec{x}^{(i)}, b, \vec{\gamma}\right)\right)\right)$) gives us

$$\begin{aligned}
&W_T^{0,\tau}(\delta) \vdash W_{A_1}, \ldots, W_{A_m}, W_{A(0)}\left(\left\{\vec{x}^{(i)}\right\} D_i\left(\vec{x}^{(i)}, 0, \vec{\gamma}\right)\right) \\
&\quad \longrightarrow W_{A(\tau(t))}\left(\left\{\vec{x}^{(i)}\right\} D_i\left(\vec{x}^{(i)}, \tau(t), \vec{\gamma}\right)\right), \\
&\qquad W_{B_1}\left[\vec{V}'^{(1)}(\vec{\gamma}) / \vec{\gamma}^{(1)}\right], \ldots, W_{B_n}\left[\vec{V}'^{(n)}(\vec{\gamma}) / \vec{\gamma}^{(n)}\right].
\end{aligned}$$

In order to get from here (41), we only have to note that

$$W_T^{0,\tau}(\delta) \vdash W_{A(0)}\left(\left\{\vec{x}^{(i)}\right\} D_i\left(\vec{x}^{(i)}, 0, \vec{\gamma}\right)\right) \equiv W_{A(0)}(\vec{\gamma})$$

(by (43)) and

$$W_{A(\tau(t))}\left(\left\{\vec{x}^{(i)}\right\} D_i\left(\vec{x}^{(i)}, \tau(t), \vec{\gamma}\right)\right) = W_{A(\tau(t))}\left(\vec{V}'(\vec{\gamma})\right)$$

by the definition (45) of $\vec{V}'$.

The proof of Lemma 5.2 is complete. ∎

6. The results

We say that a formula A is *strictly* $\Sigma_1^{1,b}$ if $A \in \Sigma_1^{1,b}$ and first order quantifiers $\forall x \leq t$ $[\exists x \leq t]$ never precede in A second order quantifiers $\exists \phi$

[$\forall\phi$, respectively]. The formal inductive definition of strictly $\Sigma_1^{1,b}$-formulae (and strictly $\Pi_1^{1,b}$-formulae) is obtained from the definition of $\Sigma_1^{1,b}$- and $\Pi_1^{1,b}$-formulae (see [6, §9.1]) by dropping the case "if A is in $\Sigma_1^{1,b}$ so is $(\forall x \le t)A$" and the dual case for $\Pi_1^{1,b}$-formulae. We will denote the set of strictly $\Sigma_1^{1,b}$-formulae by $\Sigma_1^{s1,b}$.

Lemma 6.1. **a)** *if $A \in \Sigma_1^{1,b}$ then $\Sigma_0^{1,b}-BCA+\Sigma_0^{1,b}-$replacement $\vdash A \equiv \exists\vec{\phi}W_A[\vec{\phi}/\vec{\gamma}]$, where the substitution is extended over all witnessing variables,*

b) *if $A \in \Sigma_1^{s1,b}$ then the equivalence $A \equiv \exists\vec{\phi}W_A[\vec{\phi}/\vec{\gamma}]$ can be already proved in pure second order logic.*

Proof. Obvious induction on the complexity of A. The only nontrivial case $A = \forall x \le t\ B(x)$ (which, unless $B \in \Sigma_0^{1,b}$, may occur in part a) only) is handled by $\Sigma_0^{1,b} - BCA$ for proving

$$\begin{aligned}&\exists\vec{\phi}\forall x \le t\ W_{B(a)}\left[\left\{\vec{y}^{(i)}\right\}\phi_i^{r_i+1}(a,\vec{y}^{(i)})/\gamma_i^{r_i}\right][x/a]\\ &\qquad\supset \forall x \le t\exists\vec{\phi}\ W_{B(a)}[\vec{\phi}/\vec{\gamma}][x/a]\end{aligned}$$

and by $\Sigma_0^{1,b}$-replacement in the opposite direction. ∎

Recall that for a class $\mathcal{S}$ of formulae, $\mathcal{B}(\mathcal{S})$ denotes the set of Boolean combinations of formulae from the class $\mathcal{S}$.

Theorem 6.2. *Let T be a regular theory in a first-order language $L \supseteq L_1$, and τ be a provably monotone (in T) first order term of the language L such that $T \vdash \tau(a) \ge |a|$. Then the three theories $W_T^{1,\tau}$, $W_T^{0,\tau}(\delta) + \Sigma_0^{1,b} - BCA + \Sigma_0^{1,b} -$ replacement, $T + \Sigma_0^{1,b} - CA + \Sigma_0^{1,b} - \tau - LICA + \Sigma_0^{1,b} - IND + \Sigma_0^{1,b} -$ replacement have the same set of $\mathcal{B}\left(\Sigma_1^{1,b}\right)$-theorems.*

Proof. Let $A \in \mathcal{B}\left(\Sigma_1^{1,b}\right)$.

1. $W_T^{1,\tau} \vdash A \Rightarrow W_T^{0,\tau}(\delta) + \Sigma_0^{1,b} - BCA + \Sigma_0^{1,b} -$ replacement $\vdash A$.

Represent A in the equivalent form $\bigwedge_{i=1}^k\left(\bigvee_{j=1}^{m_i}\neg A_{ij} \vee \bigvee_{j=1}^{n_i} B_{ij}\right)$ with $A_{ij}, B_{ij} \in \Sigma_1^{1,b}$. It suffices to show that $W_T^{1,\tau} \vdash A_1,\ldots,A_m \longrightarrow B_1,\ldots,B_n$ implies

$$\left.\begin{aligned}&W_T^{0,\tau}(\delta) + \Sigma_0^{1,b} - BCA + \Sigma_0^{1,b} - \text{replacement} \vdash A_1,\ldots,A_m\\ &\quad\longrightarrow B_1,\ldots,B_n,\end{aligned}\right\}\quad(46)$$

where $A_j, B_j \in \Sigma_1^{1,b}$ (to be applied afterwards to the sequents

$$A_{i1},\ldots,A_{im_i} \longrightarrow B_{i1},\ldots,B_{in_i}$$

).

If $W_T^{1,\tau} \vdash A_1, \ldots, A_m \longrightarrow B_1, \ldots, B_n$ then, by Lemma 5.2,

$$W_T^{0,\tau}(\delta) \vdash W_{A_1}, \ldots, W_{A_m} \longrightarrow W_{B_1}\left[\vec{V}^{(1)}/\vec{\gamma}^{(1)}\right], \ldots, W_{B_n}\left[\vec{V}^{(n)}/\vec{\gamma}^{(n)}\right]$$

for some $\Sigma_0^{1,b}(\delta)$-abstracts $\vec{V}^{(1)}, \ldots, \vec{V}^{(n)}$. Arguing as in [6, Lemma 10.9] and [20, Theorem 3.6 a)], we may replace the abstracts

$$V_j^{(i)} = \left\{\vec{x}^{(ij)}\right\} C_j^{(i)}\left(\left\{\vec{x}^{(ij)}\right\}\right)$$

by their bounded versions $\left\{\vec{x}^{(ij)}\right\}\left(\vec{x}^{(ij)} \leq T_{ij} \wedge C_j^{(i)}\left(\left\{\vec{x}^{(ij)}\right\}\right)\right)$ for suitable terms T_{ij}. Then $\Sigma_0^{1,b} - BCA$ yields

$$\begin{aligned} &W_T^{0,\tau}(\delta) + \Sigma_0^{1,b} - BCA \vdash W_{A_1}, \ldots, W_{A_m} \\ &\quad \longrightarrow \exists\vec{\phi}^{(1)} W_{B_1}\left[\phi^{(1)}/\vec{\gamma}^{(1)}\right], \ldots, \exists\vec{\phi}^{(n)} W_{B_n}\left[\phi^{(n)}/\vec{\gamma}^{(n)}\right], \end{aligned}$$

and (second order $\exists$:left) yields

$$\left.\begin{aligned} &W_T^{0,\tau}(\delta) + \Sigma_0^{1,b} - BCA \vdash \exists\vec{\psi}^{(1)} W_{A_1}\left[\psi^{(1)}/\vec{\gamma}'^{(1)}\right], \ldots, \\ &\quad \exists\vec{\psi}^{(m)} W_{A_m}\left[\psi^{(m)}/\vec{\gamma}'^{(m)}\right] \\ &\longrightarrow \exists\vec{\phi}^{(1)} W_{B_1}\left[\phi^{(1)}/\vec{\gamma}^{(1)}\right], \ldots, \exists\vec{\phi}^{(n)} W_{B_n}\left[\phi^{(n)}/\vec{\gamma}^{(n)}\right]. \end{aligned}\right\} \quad (47)$$

The proof of (46) is completed by applying Lemma 6.1 a).

2. $W_T^{0,\tau}(\delta) + \Sigma_0^{1,b} - BCA + \Sigma_0^{1,b} -$ replacement $\vdash A \;\Rightarrow\; T + \Sigma_0^{1,b} - CA + \Sigma_0^{1,b} - \tau - LICA + \Sigma_0^{1,b} - IND + \Sigma_0^{1,b} -$ replacement $\vdash A$.

Immediately follows from Corollary 3.7.

3. $T + \Sigma_0^{1,b} - CA + \Sigma_0^{1,b} - \tau - LICA + \Sigma_0^{1,b} - IND + \Sigma_0^{1,b} -$ replacement $\vdash A \;\Rightarrow\; W_T^{1,\tau} \vdash A$.

Note that $W_T^{1,\tau} \vdash \Sigma_0^{1,b} - IND$ and $W_T^{1,\tau} \vdash \Sigma_0^{1,b} -$ replacement by Theorems 2.6 and 2.8 respectively. Also, $W_T^{1,\tau} \vdash \Sigma_0^{1,b} - \tau - LICA$ (apply, using Theorem 2.9, $\Sigma_1^{1,b} - \tau^k - IND$ up to $\tau^k(\bar{s}(t))$ on a to the formula $\exists\phi^2\forall x \leq t[\phi(x,0) \equiv A(x) \;\wedge\; \forall y < a(\phi(x,y+1) \equiv B(x,y,\{x'\}(x' \leq x \wedge \phi(x',y))))]$, where $\bar{s}$ is a provably monotone term bounding the term s in Definition 3.3). The statement follows. ∎

Theorem 6.3. *Under the same assumptions as in Theorem* 6.2, *the theories* $W_T^{1,\tau}$, $W_T^{0,\tau}(\delta) + \Sigma_0^{1,b} - BCA$, $T + \Sigma_0^{1,b} - CA + \Sigma_0^{1,b} - \tau - LICA + \Sigma_0^{1,b} - IND$ *prove the same* $\mathcal{B}\left(\Sigma_1^{s\,1,b}\right)$*-theorems.*

Proof. In addition to Theorem 6.2 we only have to show that

$$W_T^{1,\tau} \vdash A \;\Rightarrow\; W_T^{0,\tau}(\delta) + \Sigma_0^{1,b} - BCA \vdash A \tag{48}$$

and

$$\left.\begin{aligned} & W_T^{0,\tau}(\delta) + \Sigma_0^{1,b} - BCA \vdash A \\ & \quad \Rightarrow\; T + \Sigma_0^{1,b} - CA + \Sigma_0^{1,b} - \tau - LICA \\ & \qquad + \Sigma_0^{1,b} - IND \vdash A, \end{aligned}\right\} \tag{49}$$

where $A \in \mathcal{B}\left(\Sigma_1^{s\,1,b}\right)$.

(48) is proved in the same way as the implication $W_T^{1,\tau} \vdash A \;\Rightarrow\; W_T^{0,\tau}(\delta) + \Sigma_0^{1,b} - BCA + \Sigma_0^{1,b} -$ replacement $\vdash A$ in the proof of Theorem 6.2 with the difference that for inferring (46) from (47) we use part b) of Lemma 6.1 instead of part a).

(49) again immediately follows from Corollary 3.7. ∎

Theorem 6.4. *Let T be a regular theory in a first-order language $L \supseteq L_1$, and τ be a provably monotone (in T) first order term of the language L such that $T \vdash \tau(a) \geq |a|$. Then $W_T^{1,\tau}$ and $W_T^{0,\tau}(\delta)$ prove the same $\Sigma_0^{1,b}$-theorems.*

Proof. In addition to Theorem 6.2 we only have to show $W_T^{1,\tau} \vdash A \;\Rightarrow\; W_T^{0,\tau}(\delta) \vdash A,\; A \in \Sigma_0^{1,b}$. This immediately follows from Lemma 5.2 since $W_A = A$ for $A \in \Sigma_0^{1,b}$. ∎

6.5. Remark. Note that actually we have proved that all *six* theories involved in the statements of Theorems 6.2, 6.3, 6.4 have the same set of $\Sigma_0^{1,b}$-theorems.

7. Appendix

As we noted in the introduction, the main motivation for developing this logical formalism is the author's feeling that V_1^1 is the right theory to capture that part of reasoning in Boolean complexity which led to actual lower bounds for explicitly given Boolean functions. Here I will try to present both formal and informal arguments in favour of this thesis. For our analysis I have chosen the following results in Boolean complexity (often considered among the major results in the area):

- *lower bounds for constant-depth circuits over the standard basis* [1, 9, 28, 11],
- *lower bounds for monotone circuits* [33, 34, 29, 3, 27, 30],

- *lower bounds for constant-depth circuits with $MOD-q$ gates* [35, 24, 4],
- *lower bounds for monotone formulae based on communication complexity* [12, 19, 18].

The reader wishing to learn more about these and other results is referred to the excellent survey [5].

A. The formalization

Apparently, $\mathcal{L}_1$ is the most natural and elegant formal language for formalizing the statements of results and open problems in Boolean complexity. Using second order languages allows one to capture the common practice in the area which tends to treat Boolean inputs and functions separately, as two different kinds of objects. Throughout the Appendix, we will let n denote the number of variables of the Boolean function in question, and $N \rightleftharpoons 2^n \dot{-} 1$ will have the property $|N| = n$ (N approximately equals the overall number of Boolean inputs).

In the proposed framework, first order objects are integers of order $2^{O(n)} = N^{O(1)}$, they are used for encoding Boolean inputs, restrictions etc. Second order objects correspond to Boolean functions, circuits, protocols of their computations (given simply as a collection of truth-tables of intermediate results) etc. Formally this is achieved by introducing N into our formulae as a "dummy" free first-order variable.

To give you just one example of this encoding, we explicitly write down the $\Sigma_0^{1,b}(\alpha)$-formula $MonCircuit(t, N, \alpha^2)$ asserting that α encodes the protocol of computation by a monotone circuit of size t in $|N|$ variables:

$$\begin{aligned}
&MonCircuit(t, N, \alpha^2) \rightleftharpoons \forall u < t \\
&\quad \Big(\exists i < |N| \Big(\forall x < 2^{|N|}(\alpha(u, x) \equiv Bit(i, x))\Big) \\
&\qquad \vee \exists u_1 < u \exists u_2 < u \\
&\qquad\quad \Big(\forall x < 2^{|N|}(\alpha(u, x) \;\equiv\; (\alpha(u_1, x) \wedge \;\alpha(u_2, x))) \\
&\qquad\qquad \vee \forall x < 2^{|N|}(\alpha(u, x) \;\equiv\; (\alpha(u_1, x) \vee \alpha(u_2, x)))\Big)\Big).
\end{aligned}$$

Then, say, the statement "every monotone circuit in n variables computing SATISFIABILITY must have size at least $\tau(n)$" is formalized by the following $\Sigma_0^{1,b}(\alpha)$-formula:

$$\left.\begin{aligned}
&MonCircuit(t, N, \alpha) \wedge \forall x < 2^{|N|}(\alpha(t \dot{-} 1, x) \equiv SAT(x, N)) \\
&\quad \supset \; t \geq \tau(|N|),
\end{aligned}\right\} \quad (50)$$

where $SAT(a, N)$ is a bounded formula representing the predicate SATISFIABILITY on strings of length $|N|$.

B. V_1^1 proves what we want it to prove

Having written our problems in the formal language $\mathcal{L}_1$ in the style of (50), the next question is what is the minimal natural theory in this language which can carry out the amount of combinatorial and algebraic technique developed in Boolean complexity so far. I argue that V_1^1 is exactly the required theory. By this I mean in particular that it proves all lower bounds mentioned above and, moreover, these formal proofs are obtained from their informal counterparts in a very natural and straightforward way[3]. This thesis becomes especially clear if we take V_1^1 in the equivalent form $V_1^0(\delta)$ or $V_1^0(\bar{\delta})$ (see Theorem 3.9). For an illustration I have chosen the proof of lower bounds on the monotone circuit size of the clique function [33, 3] binding to the notation used in the nice presentation of this result in [5, Section 4].

We encode clique indicators $\lceil X \rceil$ by integers of bit length $O(l \log n)$, and approximator circuits by second order variables (outputting 1 on exactly those clique indicators which appear in the approximator).

A subtle point is that sunflowers can be also coded by integers. Indeed, their bit length is $O(pl \log n)$, and pl must not exceed $O(n)$ (irrelevantly of exact values of parameters specified in [5, Section 4.3]!) e.g. because otherwise Erdös-Rado Lemma [5, Lemma 4.1] becomes meaningless. Recall that we have access to integers of bit length $O(n^2)$ as we have $\binom{n}{2}$ variables.

Everything beyond this point is fairly straightforward. The relation "$(Z_1, \ldots, Z_p)$ is a sunflower in a collection $\mathcal{L}$" as well as "Z is a petal of the minimal (in some explicit order) sunflower in $\mathcal{L}$" and "C is the center of the minimal sunflower in $\mathcal{L}$" are in $\Sigma_0^{1,b}(\mathcal{L})$. Hence we may introduce in $V_1^0(\delta)$ a relational $\delta(a, \alpha)$ such that $\delta(\langle N, l, p, h\rangle, \mathcal{L})$ encodes the result of h consecutive pluckings of minimal sunflowers from $\mathcal{L}$, and then find a $\Sigma_0^{1,b}(\delta)$-formula representing the final result of the plucking procedure.

The proof of Erdös-Rado Lemma [5, Lemma 4.1] is also straightforwardly carried over in $V_1^0(\delta)$. The only thing to be noticed is that the set $\mathcal{M}$ in the proof can be also treated as a first order object.

Now, using the possibility to iterate δ-symbols in $V_1^0(\delta)$ (see Lemma 4.5), we can $\Sigma_0^{1,b}(\delta)$-define the approximator $\tilde{C}$ of a monotone circuit C. All combinatorics from [5, Section 4.3] is then easily formalized in $V_1^0(\delta)$ if we recall that V_1^1 (and hence $V_1^0(\delta)$) can easily count first order objects.

I do not know of any lower bound in Boolean complexity for an NP-

[3] The question of minimality of V_1^1 will be thoroughly discussed in Section E.

function whose proof could not be carried over in V_1^1.

C. V_1^1 apparently does not prove what we do not want

In Boolean complexity one should cope with the phenomenon that the inherently hard problems making the core of the field become trivial when one considers "random" functions. Unfortunately, the powerful Shannon counting arguments (see e.g. [23]) used for dealing with random functions are hardly relevant to proving lower bounds for "explicit" functions. For this reason it is customary in the modern Boolean complexity to distance oneself from the realm of those arguments either by considering only explicit functions (which is somewhat ill-defined) or only functions from the class NP (which is not quite adequate since e.g. superlinear lower bounds on the circuit size of a *natural* function in, say, $PSPACE$ would be also of extreme interest). All this is a little bit annoying.

I take it as a strong argument in support of our main thesis that V_1^1 *can not formalize Shannon counting arguments, at least in an obvious way.* As a consequence, it is open whether V_1^1 can prove even the *existence* of Boolean functions f_n with circuit size $\geq 10n$.

In order to understand the reasons for this, it is convenient to scale V_1^1 down to S_2^1 using $RSUV$-isomorphism (see [25, 26, 20]). Then we have Δ_1^b-function $Value(C, F)$, where F is the dummy variable with the meaning "$F \rightleftharpoons 2^{2^n}$ is the number of Boolean functions", which expresses the function computed by the circuit C. The number of circuits of size $10n$ is $2^{\Theta(n \log n)}$ which is $|F|^{\Theta(\log\log\log F)}$ in terms of F. Hence, the mapping $Value$ restricted to circuits of size $10n$, gives rise to the Δ_1^b-mapping

$$|F|^{\Theta(\log\log\log F)} \to F,$$ [4]

and our question on the provability of the existence of complex Boolean functions in V_1^1 becomes equivalent to whether S_2^1 can rule out that this mapping is surjective.

However, for any unbounded $\tau(a)$, $S_2^1(f)$ can *not* disprove that $f : |a|^{\tau(a)} \longrightarrow a$ is surjective (see [6, Corollary 5.13]). Hence, due to the "universal" nature of the predicate $Value$, it is hardly conceivable that its internal properties will help us to show that it is not surjective by some modified counting arguments.

The overall conclusion is that the theory V_1^1 can do exactly the right amount of counting. That is, this theory is capable of arbitrary "slice-and-measure" arguments on the Boolean cube, but becomes absolutely helpless when asked to count Boolean functions. Note for comparison that V_2^1

[4] we will identify here and in other appropriate places an integer a with the set $\{0, 1, \ldots, a-1\}$

already can prove the existence of Boolean functions whose circuit size is $\geq n^{10}$. V_1^3 can prove the existence of Boolean functions with exponential circuit size (since S_2^3 proves $WPHP$, the *weak pigeon hole principle* stating that $2a$ pigeons can not sit in a holes [17]).

D. On the mathematics and metamathematics of V_1^1

Many people believe that the theory S_2^1 is the most important and natural theory among various fragments of Bounded Arithmetic. The same applies to V_1^1 which is equivalent to it via the $RSUV$-isomorphism.

If we take V_1^1 in the equivalent form $V_1^0(\bar{\delta})$ then it becomes very transparent that every proof in V_1^1 can be viewed as evaluating a polynomial size (in N) circuit followed by proving certain "plain" (that is $\Sigma_0^{1,b}(\delta)$) facts about the protocol of this evaluation. In other words, the same concept of a polynomial size circuit making the core of Boolean complexity is also responsible for the metamathematics of the formal theory V_1^1. This close link between mathematics and metamathematics already has turned out important for research on the possibility of solving major open problems in Boolean complexity by means of Bounded Arithmetic (see [21]).

E. Proofs in subsystems of V_1^1

In Section B we formulated the thesis that V_1^1 supports proofs of lower bounds for explicit functions existing in Boolean complexity at the moment. It is conceivable, however, that easier proofs do not use the full power of V_1^1 and can be carried over in (presumably) weaker fragments of it. In this section we analyze from this point of view the proofs listed in the beginning of this Appendix. As a reward, we will see in Section E.4 a constructive version of the proof of Håstad Switching Lemma which is probably interesting in its own right.

E.1. Lower bounds for monotone circuits

As we saw in Section B, carrying these proofs out in $V_1^0(\delta)$ involves, besides the relationals for counting first-order objects, two other kinds of relationals δ. The first relational is used to evaluate the result of the plucking procedure, and another to construct the circuit approximator. Both of these can be evaluated within $t(n)$ steps of iterations, where $t(n)$ is the bound on the circuit size we are proving. Hence these proofs can be for-

malized in $W_1^{1,\tau}$, where $\tau(a) \rightleftharpoons t(|a|)$ [5] (e.g. for [3, Theorem 3.9] we have $\tau(a) = \lfloor 2^{\left(\frac{|a|}{\|a\|}\right)^{1/6}} \rfloor$). Whereas it is not quite unlikely that the number of iterations in the plucking procedure can be decreased by using some clever combinatorics, it would be very astonishing if constructing the approximator $\tilde{C}$ for a general circuit C could be implemented within any number of steps substantially less than the size of C. Hence it is plausible that $W_1^{1,\tau}$ is a "tight" upper bound on the strength of a theory which captures this kind of argument.

E.2. Lower bounds for monotone formulae based on communication complexity, excluding [19]

Here the situation is apparently also very clear. The minimal fragment of V_1^1 capable of formalizing these proofs seems to be U_1^1. Indeed, they use only induction up to $O(n)$ (which is an obvious upper bound on the circuit depth) and plain counting arguments. The latter still can be carried out in U_1^1 (see [6, proof of Proposition 10.2]). The only additional remark which should be made in this respect concerns probability distributions.

All distributions over a set of first order objects $\mathcal{S}$ [over second order objects] should be reduced to the form $f(\boldsymbol{a})$ [$V(\boldsymbol{a})$ respectively], where $\boldsymbol{a}$ is a random integer taken *uniformly* from a set of cardinality $N^{O(1)}$, f is a $\Sigma_0^{1,b}(\delta)$-definable function symbol, and V is a $\Sigma_0^{1,b}(\delta)$-abstract. This allows one to comfortably perform all usual operations with distributions (like taking the product) within $U_1^0(\delta)$; on the other hand, all distributions actually used in the proofs can be always reduced to this form.

E.3. Lower bounds using algebraic arguments

The situation with lower bounds for constant-depth circuits using $MOD-q$ gates [35, 24, 4], as well as with the bound for the monotone formula size of MINIMUM COVER [19] is far less clear. Trying to carry out these proof in $U_1^0(\delta)$, we are stuck in two places. Firstly, the proofs of [35, Lemma 1] and [24, Lemma 1] require the following

Fact 1. *For some $\epsilon > 0$, the following is true. Let α be an $L \times M$ matrix over $\mathbb{F}_q$ such that every column has at least one non-zero position. Then there exists a row vector β of length L such that $\beta\alpha$ has at least ϵM non-zero positions.*

The standard probabilistic argument (which in V_1^1 can be replaced by an easy induction on L) gives here $\epsilon = 1 - \frac{1}{q}$ but we would be satisfied with

[5] strictly speaking, $\tau(a)$ is not necessarily a *term* of the language L_1. However, for all τ which are definable in $I\Delta_0$ by a bounded formula, we can always append them to L_1 (along with natural defining axioms)

any ϵ since this affects only the multiplicative constant in the exponent of the bound.

The second part of the proofs in [35, 24, 4] (that is bounding from below the distance between a symmetric Boolean function and the set of polynomials of a low degree) in both known methods [35] and [24] requires the following

Fact 2. *Let α be an $L \times M$ matrix over $\mathbb{F}_q$ such that $L > M$ (or $L \gg M$). Then there exists a non-zero row vector β of length L such that $\beta\alpha = 0$.*

Now, recall from Section 1.1 that $U_1^0(\delta)$ is exactly the theory capturing NC-computations. Examples of βs needed in Facts 1, 2 are known to be NC-computable from α. For Fact 1 one could apply the standard derandomization procedure [16]; the NC-algorithm for Fact 2 is based upon computing the matrix rank [15]. Hence $U_1^0(\delta)$ can define relationals δ witnessing Facts 1, 2 in the real world.

It is not clear, however, to which extent $U_1^0(\delta)$ can *prove* the desired properties of these relationals. The point is that the standard proofs often involve highly consecutive concepts well beyond the reach of the class NC. E.g. even the most basic subroutine of computing the determinant of a matrix is designed by parallelising a large-depth circuit which we can not even evaluate in $U_1^0(\delta)$. I have found a number of similar examples but, since the careful study of provability in U_1^1 does not quite match the main topic of this paper, I did not put that much effort into checking how essential these difficulties are.

E.4. Lower bounds for constant-depth circuits over the standard basis

In this subsection we again stick to the notation from [5]. Our final goal will be to demonstrate that [5, Theorem 3.8] can be proven in both $S_2(\alpha)$ and U_1^1. As the first step, we show that a variant of the Håstad Switching Lemma [5, Lemma 3.3] can be proven already in $S_1(\alpha)$. Note that this theory is equivalent to $I\Delta_0(\alpha)$ since the latter can define originally missing symbols $\lfloor \frac{x}{2} \rfloor$ and (in a very nontrivial way, see e.g. [10, Chapter 5.3]) $|x|$.

Unfortunately, the standard proof of this lemma involves conditioning on second-order objects (see e.g. the quantifier on F in [5, Lemma 3.3′]) and hence it is not clear a priori how to place it even into V_1^1. Woods (see [14], [13, Chapter 15]) gave a variant of this proof which avoids such conditioning. Apparently, this is already formalizable in V_1^1. For our purposes, however, we need the following more constructive version of Håstad's proof which probably is interesting in its own right.

We denote by $\mathcal{R}^\ell$ the set of all restrictions assigning exactly ℓ stars. Let $\mathrm{Bad}(f, s)$ consist of all restrictions ρ for which $\min(f|_\rho) \geq s$, and let $\mathrm{Bad}^\ell(f, s) \rightleftharpoons \mathcal{R}^\ell \cap \mathrm{Bad}(f, s)$.

We define the auxiliary set $\mathrm{Code}(t,s)$ which consists of those sequences

$$\left(\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(k)}\right),$$

where $\sigma^{(i)} \in \{0,1,*\}^t$, for which:

a) every $\sigma^{(i)}$, $1 \le i \le k$ contains at least one 1,

b) the whole sequence has exactly s positions occupied by 0,1.

Lemma E.1. *Let f be a t-closed function. Then there exists an injective mapping $F : \mathrm{Bad}^\ell(f,s) \longrightarrow \mathcal{R}^{\ell-s} \times \mathrm{Code}(t,s)$.*

Proof. Fix a representation $f = \bigwedge_{i=1}^{H} C_i$, where C_is are ORs of fan-in $\le t$. Let $\rho \in \mathrm{Bad}^\ell(f,s)$. Fix a minterm π of $f|_\rho$ whose size is at least s. We will recursively define assignments $\pi_1, \pi_2, \ldots, \pi_k, \ldots$ breaking up π into pieces.

Assume that we already have $\pi_1, \pi_2, \ldots, \pi_{i-1} \subseteq \pi$ with mutually disjoint domains and such that $\pi_{i-1} \ldots \pi_1$ is still different from π. Apply $\pi_{i-1} \ldots \pi_1 \rho$ to the OR gates $C_1, \ldots, C_H$ and find the minimal ν_i with the property $C_{\nu_i}|_{\pi_{i-1}\ldots\pi_1\rho} \not\equiv 1$. Such ν_i must exist since $\pi_{i-1} \ldots \pi_1 \neq \pi$ and hence $f|_{\pi_{i-1}\ldots\pi_1\rho} \not\equiv 1$. Let T_i be the collection of variables of C_{ν_i}, and let Y_i be those variables from T_i which are set by π but not by $\pi_{i-1} \ldots \pi_1$. Note that $Y_i \neq \emptyset$ since $f|_{\pi\rho} \equiv 1$ and thus $C_{\nu_i}|_{\pi\rho} \equiv 1$. We let $\pi_i \rightleftharpoons \pi|_{Y_i}$.

Now, let k be minimal with the property that $\pi_1, \pi_2, \ldots, \pi_{k-1}, \pi_k$ altogether assign at least s variables. We trim π_k in an arbitrary way so that it still sets C_{ν_k} to 1 and $\pi_1, \pi_2, \ldots, \pi_{k-1}, \pi_k$ assign *exactly* s variables.

Let $\bar{\pi}_i$ be the uniquely determined assignment which has the same domain as π_i and does *not* set the gate C_{ν_i} to 1. We let

$$F_1(\rho) \rightleftharpoons \bar{\pi}_k \bar{\pi}_{k-1} \ldots \bar{\pi}_1 \rho;$$

note that $F_1(\rho) \in \mathcal{R}^{\ell-s}$.

For $1 \le i \le k$ we define $\sigma^{(i)} \in \{0,1,*\}^t$ as follows. Let $T_i = \left\{x_{\theta(i,1)}, \ldots, x_{\theta(i,t_i)}\right\}$, where $t_i \le t$ and $\theta(i,1) < \ldots < \theta(i,t_i)$. We set

$$\sigma_j^{(i)} \rightleftharpoons \begin{cases} \pi_i(x_{\theta(i,j)}) \oplus \bar{\pi}_i(x_{\theta(i,j)}) \text{ if } j \le t_i \\ \qquad \text{and } x_{\theta(i,j)} \text{ is in the domain of } \pi_i \\ * \text{ if either } j > t_i \text{ or } x_{\theta(i,j)} \text{ is not in the domain of } \pi_i. \end{cases}$$

Note that π_i and $\bar{\pi}_i$ have different effect on the gate C_{ν_i} hence $\forall i \exists j\ \sigma_j^{(i)} = 1$ which implies $\left(\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(k)}\right) \in \mathrm{Code}(t,s)$. We let

$$F_2(\rho) \rightleftharpoons \left(\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(k)}\right).$$

The desired mapping $F : \mathrm{Bad}^\ell(f,s) \longrightarrow \mathcal{R}^{\ell-s} \times \mathrm{Code}(t,s)$ takes ρ to $(F_1(\rho), F_2(\rho))$. In order to complete the proof we only have to show how

to retrieve ρ from $(F_1(\rho), F_2(\rho)) = \left(\bar\pi_k\bar\pi_{k-1}\dots\bar\pi_1\rho, \sigma^{(1)}, \dots, \sigma^{(k)}\right)$. It is sufficient to retrieve all $(\bar\pi_i, \pi_i)$ $(1 \le i \le k)$.

We do it by induction on i. Assume that we already know $(\bar\pi_1, \pi_1), \dots, (\bar\pi_{i-1}, \pi_{i-1})$. Then we also know $\bar\pi_k \dots \bar\pi_i\pi_{i-1}\dots\pi_1\rho$. The crucial observation is that *ν_i is the minimal index ν such that* $C_\nu|_{\bar\pi_k\dots\bar\pi_i\pi_{i-1}\dots\pi_1\rho} \not\equiv 1$. Indeed, $C_{\nu_i}|_{\pi_{i-1}\dots\pi_1\rho} \not\equiv 1$, $C_{\nu_i}|_{\bar\pi_i} \not\equiv 1$ from definitions, and $\bar\pi_{i+1}, \dots, \bar\pi_k$ do not assign variables from T_i at all. Hence $C_{\nu_i}|_{\bar\pi_k\dots\bar\pi_i\pi_{i-1}\dots\pi_1\rho} \not\equiv 1$. On the other hand, all C_ν with $\nu < \nu_i$ are already set to 1 by $\pi_{i-1}\dots\pi_1\rho$ and hence by $\bar\pi_k\dots\bar\pi_i\pi_{i-1}\dots\pi_1\rho$.

Now, when we know ν_i, the rest is easy. Having ν_i, we have T_i. From $\sigma^{(i)}$ and T_i we know the domain of $\bar\pi_i$, then we retrieve from $F_1(\rho)$ the actual value of $\bar\pi_i$ and consult $\sigma^{(i)}$ again to get π_i. ∎

Note that if $\ell, t, s \le |N|/||N||$ (and the applications of Håstad Switching Lemma appeal only to parameters from this range), then S_1 is capable of coding finite sequences in the amount which is sufficient for formalizing the proof of Lemma E.1. The key observation to this is that if a is a binary string of length $n^{O(1)}$ *known to contain at most $O(n/|n|)$ ones* then we can list in S_1 all positions where a has ones in increasing order and freely switch from one representation of a to another, whichever is more appropriate at the moment. We leave the details of formalizing the above proof of Håstad Switching Lemma in $S_1(\alpha)$ to the reader.

Now we can complete the proof of [5, Theorem 3.8] in either $S_2(\alpha)$ or U_1^1. We will not be too fussy about the exact value of the multiplicative constant in the exponent of the bound (which is a common practice in Boolean complexity) and estimate the cardinality of $\mathrm{Code}(t, s)$ in Lemma E.1 by $\mathrm{Code}(t, s) \le (4t)^s$. This means that we construct a straightforward injective mapping $\mathrm{Code}(t, s) \longrightarrow (4t)^s$ defined by a bounded formula.

Then, for a given $\Sigma_d^{S,s}$ circuit C, where $S = 2^{(1/20)n^{1/d}}$ and $s = \log S$, we show by induction on k, $1 \le k \le d$, that there exists $\rho \in \mathcal{R}^{n^{(1-(k-1)/d)}}$ which makes all functions computed at the kth level either s-close or s-open depending on whether $d - k$ is odd or not.

For the inductive step we first apply Lemma E.1 with

$$n := n^{(1-(k-1)/d)},\ \ell := n^{(1-k/d)},\ t := s$$

to $f_1|_{\rho_k}, \dots, f_S|_{\rho_k}$, where $f_1, \dots, f_S$ is the complete list of the functions computed at the k-th level, and ρ_k is the restriction found at the previous step. Then, assuming that the desired ρ_{k+1} does not exist, we glue the resulting mappings together and get an injective mapping

$$\mathcal{R}^\ell \longrightarrow \mathcal{R}^{\ell-s} \times (4s)^s \times S. \tag{51}$$

Up to this point we still were in the theory $S_1(\alpha)$.

It is not clear if S_1 can define binomial coefficients $\binom{|N|}{a}$ and prove their primary properties. But any of the two theories S_2^1, U_1^1 certainly

can do that. S_2^1 merely exploits the definition $\binom{n}{a} \rightleftharpoons \frac{n!}{a!(n-a)!}$, and in U_1^1 we can naturally enumerate the set $[n]^a$. In particular, both theories prove the bound $\binom{|N|}{\ell} \geq \binom{|N|}{\ell - s} \cdot \left(\frac{|N|-\ell+s}{\ell}\right)^s$ (for U_1^1 we also need the assumption $s \leq O(|N|/||N||)$ since otherwise the bit size of the right-hand side can be too large). This allows us to rewrite (51) as an injective mapping

$$\binom{n}{\ell} \cdot 2^{n-\ell} \longrightarrow \binom{n}{\ell - s} \times 2^{n-\ell+s} \times (4s)^s \times S.$$

But the existence of a mapping of this form contradicts the weak pigeon hole principle which is provable in both $S_2(\alpha)$ and U_1^1 (see [17] for the case of $S_2(\alpha)$). This contradiction completes the inductive step.

The rest of the proof of [5, Theorem 3.8] does not present any problems.

Acknowledgement. I am indebted to many people, Mauricio Karchmer, Jan Krajíček, Mike Sipser, Avi Wigderson being among them, for many useful remarks, and especially for their often critical assessment of the material contained in Appendix. This highly helped me in elaborating the concepts. I would also like to thank anonymous referees for many valuable suggestions.

REFERENCES

[1] M. Ajtai. Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24:1–48, May 1983.

[2] B. Allen. Arithmetizing uniform NC. *Annals of Pure and Applied Logic*, 53(1):1–50, 1991.

[3] N. Alon and R. Boppana. The monotone circuit complexity of Boolean functions. *Combinatorica*, 7(1):1–22, 1987.

[4] D. A. Barrington. A note on a theorem of Razborov. Technical report, University of Massachusetts, 1986.

[5] R. B. Boppana and M. Sipser. The complexity of finite functions. In Jan van Leeuwen, editor, *Handbook of Theoretical Computer Science, vol. A* (*Algorithms and Complexity*), chapter 14, pages 757–804. Elsevier Science Publishers B.V. and The MIT Press, 1990.

[6] S. R. Buss. *Bounded Arithmetic.* Bibliopolis, Napoli, 1986.

[7] P. Clote. A first-order theory for the parallel complexity class NC. Technical Report BCCS-89-01, Boston College, January 1989. Published in expanded form jointly with G. Takeuti in "Arithmetics for NC, $ALOGTIME$, L and NL", *Annals of Pure and Applied Logic*, 56(1992), 73-117.

[8] S. Cook. Feasibly constructive proofs and the propositional calculus. In *Proceedings of the 7th Annual ACM Symposium on the Theory of Computing*, pages 83–97, 1975.

[9] M. Furst, J. B. Saxe, and M. Sipser. Parity, circuits and the polynomial time hierarchy. *Math. Syst. Theory*, 17:13–27, 1984.

[10] P. Hájek and P. Pudlák. *Metamathematics of First-Order Arithmetic.* Springer-Verlag, 1993.

[11] J. Håstad. *Computational limitations on Small Depth Circuits.* PhD thesis, Massachusetts Institute of Technology, 1986.

[12] M. Karchmer and A. Wigderson. Monotone circuits for connectivity require super-logarithmic depth. *SIAM J. on Disc. Math.*, 3(2):255–265, May 1990.

[13] J. Krajíček. *Bounded arithmetic, propositional logic and complexity theory.* Cambridge University Press, 1994.

[14] J. Krajíček, P. Pudlák, and A. R. Woods. Exponential lower bounds to the size of bounded depth frege proofs of the pigeonhole principle. Submitted to *Random Structures and Algorithms*, 1993.

[15] K. Mulmuley. A fast parallel algorithm to compute the rank of a matrix over an arbitrary field. *Combinatorica*, 7(1):101–104, 1987.

[16] J. Naor and M. Naor. Small-bias probability spaces: Efficient constructions and applications. In *Proceedings of the 22th ACM STOC*, pages 213–223, 1990.

[17] J. B. Paris, A. J. Wilkie, and A. R. Woods. Provability of the pigeonhole principle and the existence of infinitely many primes. *Journal of Symbolic Logic*, 53(4):1235–1244, 1988.

[18] R. Raz and A. Wigderson. Monotone circuits for matching require linear depth. In *Proceedings of the 22th Ann. ACM Symposium on the Theory of Computing*, pages 287–292, 1990.

[19] A. Razborov. Applications of matrix methods to the theory of lower bounds in computational complexity. *Combinatorica*, 10(1):81–93, 1990.

[20] A. Razborov. An equivalence between second order bounded domain bounded arithmetic and first order bounded arithmetic. In P. Clote and J. Krajíček, editors, *Arithmetic, Proof Theory and Computational Complexity*, pages 247–277. Oxford University Press, 1992.

[21] A. Razborov. Unprovability of lower bounds on circuit size in certain fragments of Bounded Arithmetic. To appear in *Izvestiya of the RAN*, 1994.

[22] A. Razborov and S. Rudich. Natural proofs. In *Proceedings of the 26th ACM Symposium on Theory of Computing*, pages 204–213, 1994.

[23] J. Riordan and C. E. Shannon. The number of two-terminal series parallel networks. *J. Math. Phys.*, 21(2):83–93, 1942.

[24] R. Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the 19th ACM Symposium on Theory of Computing*, pages 77–82, 1987.

[25] G. Takeuti. S_3^i and $\overset{\circ}{V}{}_2^i(BD)$. *Archive for Math. Logic*, 29:149–169, 1990.

[26] G. Takeuti. *RSUV* isomorphism. In P. Clote and J. Krajíček, editors, *Arithmetic, Proof Theory and Computational Complexity*, pages 364–386. Oxford University Press, 1992.

[27] É. Tardos. The gap between monotone and nonmonotone circuit complexity is exponential. *Combinatorica*, 8:141–142, 1988.

[28] A. Yao. Separating the polynomial-time hierarchy by oracles. In *Proceedings of the 26th IEEE FOCS*, pages 1–10, 1985.

[29] А.Е. Андреев. Об одном методе получения нижних оценок сложности индивидуальных монотонных функций. *ДАН СССР*, 282(5):1033–1037, 1985. A.E. Andreev, On a method for obtaining lower bounds for the complexity of individual monotone functions. *Soviet Math. Dokl.* 31(3):530-534, 1985.

[30] А.Е. Андреев. Об одном методе получения эффективных нижних оценок монотонной сложности. *Алгебра и логика*, 26(1):3–21, 1987. A.E. Andreev, On one method of obtaining effective lower bounds of monotone complexity. *Algebra i logika*, 26(1):3-21, 1987. In Russian.

[31] А. А. Марков. О минимальных контактно-вентильных двухполюсниках для монотонных симметрических функций. In *Проблемы кибернетики*, volume 8, pages 117–121. Наука, 1962.

A. A. Markov, On minimal switching-and-rectifier networks for monotone symmetric functions, *Problems of Cybernetics*, vol. 8, 117-121 (1962).

[32] Э. И. Нечипорук. Об одной булевской функции. *ДАН СССР*, 169(4):765–766, 1966. E. I. Nečiporuk, On a Boolean function, *Soviet Mathematics Doklady* 7:4, pages 999-1000.

[33] А. А. Разборов. Нижние оценки монотонной сложности некоторых булевых функций. *ДАН СССР*, 281(4):798–801, 1985. A. A. Razborov, Lower bounds for the monotone complexity of some Boolean functions, *Soviet Math. Dokl.*, 31:354-357, 1985.

[34] А. А. Разборов. Нижние оценки монотонной сложности логического перманента. *Матем. Зам.*, 37(6):887–900, 1985. A. A. Razborov, Lower bounds of monotone complexity of the logical permanent function, *Mathem. Notes of the Academy of Sci. of the USSR*, 37:485-493, 1985.

[35] А. А. Разборов. Нижние оценки размера схем ограниченной глубины в полном базисе, содержащем функцию логического сложения. *Матем. Зам.*, 41(4):598–607, 1987. A. A. Razborov, Lower bounds on the size of bounded-depth networks over a complete basis with logical addition, *Mathem. Notes of the Academy of Sci. of the USSR*, 41(4):333-338, 1987.

[36] Б. А. Субботовская. О реализации линейных функций формулами в базисе $\&, \vee, -$. *ДАН СССР*, 136(3):553–555, 1961. B.A.Subbotovskaya, Realizations of linear functions by formulas using $+, *, -$, *Soviet Mathematics Doklady* 2(1961), 110-112.

[37] В. М. Храпченко. О сложности реализации линейной функции в классе П-схем. *Матем. заметки*, 9(1):35–40, 1971. V.M. Khrapchenko, Complexity of the realization of a linear function in the class of π-circuits, *Math. Notes Acad. Sciences USSR* 9(1971), 21-23.

[38] В. М. Храпченко. Об одном методе получения нижних оценок сложности П-схем. *Матем. заметки*, 10(1):83–92, 1971. V.M. Khrapchenko, A method of determining lower bounds for the complexity of П-schemes, *Math. Notes Acad. Sciences USSR* 10(1971), 474-479.

Alexander A. Razborov
Steklov Mathematical Institute
Vavilova 42, 117966, GSP–1, Moscow, RUSSIA
razborov@class.mian.su