The polynomial and linear hierarchies in models where the weak pigeonhole principle fails

Leszek Aleksander Kołodziejczyk*
Neil Thapen†

December 29, 2006

Abstract

We show, under the assumption that factoring is hard, that a model of PV exists in which the polynomial hierarchy does not collapse to the linear hierarchy; that a model of S^1_2 exists in which NP is not in the second level of the linear hierarchy; and that a model of S^1_2 exists in which the polynomial hierarchy collapses to the linear hierarchy and in which the strict version of PH does not collapse to a finite level.

Our methods are model-theoretic. We use the assumption about factoring to get a model in which the weak pigeonhole principle fails in a certain way, and then work with this failure to obtain our results.

One of the main goals of the research into fragments of bounded arithmetic is to understand which relations between computational complexity classes

*Institute of Mathematics, Warsaw University, Banacha 2, 02-097 Warszawa, Poland, lak@mimuw.edu.pl. This work was carried out while the author was visiting the Mathematical Institute of the Academy of Sciences of the Czech Republic in Prague.

†Mathematical Institute, Academy of Sciences of the Czech Republic, Žitná 25, CZ-115 67 Praha 1, Czech Republic, thapen@math.cas.cz. Supported in part by grant AV0Z10190503 and by the Eduard Čech Center grant LC505.

are consistent with these theories. The fundamental question in this area is whether the polynomial hierarchy can be infinite in a model of full bounded arithmetic. But it makes sense to study what weak theories of arithmetic can say about the other important questions from complexity theory which can be meaningfully stated in this setting, especially since many of these questions are interrelated.

In this paper, we consider the problem of whether the polynomial hierarchy PH is equal to the linear time hierarchy LinH. This is a long-standing open problem about which little is known, other than the immediate corollaries of the time hierarchy theorem: a given level of the polynomial hierarchy properly contains the corresponding level of the linear hierarchy, and consequently LinH must be infinite if the two hierarchies are equal.

Under the general assumption that factoring is hard, in the sense that factoring of products of two primes is not possible in probabilistic polynomial time, we prove that:

(1) it is consistent with PV that NP is not contained in LinH,

(2) it is consistent with S^1_2 that NP is not contained in the second level of LinH, but

(3) it is also consistent with S^1_2 that all of PH is contained in LinH.

Hence, the question whether PH = LinH is independent of PV. Additionally, the containment of PH in LinH can be extended to a nonstandard variant of PH, which implies by a typical diagonalization argument that

(4) (the strict, or prenex, version of) PH is infinite consistently with S^1_2.

This proof of this last result bears a strong resemblance to an old theorem of Paris and Wilkie ([PW85]) on the Δ_0 hierarchy in $I\Delta_0$. Nevertheless, the result seems interesting because it shows that the hardness of one specific problem from a low level of the polynomial hierarchy (i.e. factoring) may imply the separation of *all* levels of PH (with parameters) — albeit only in a model of the weak theory S^1_2.

Our methods are model-theoretic and rely on an analysis of which versions of the weak pigeonhole principle (WPHP) for polynomial time functions or

NP multifunctions hold in a given ground model. To obtain results (1) and (2), we use our assumption about factoring to get a model in which the surjective version of WPHP holds but the injective version fails. Similar uses of hardness assumptions from cryptography appear in [Tha02] and [CT06]. The proofs of (3) and (4) are based on the observation that if any version of WPHP fails and the model satisfies a sufficient amount of induction, then, by a modification of an argument in [Tha02], quantifiers over large elements of the model can be translated into quantifiers over smaller elements of the model. The result (4) may actually be stated in a stronger way than above: S_2^1 plus the negation of a relatively strong form of WPHP for Σ_1^b relations proves that the strict version of the polynomial hierarchy does not collapse.

The paper is organized in the following way. After introducing the necessary definitions and notation, and a discussion of variants of WPHP, we prove a simplified version of (1) in section 1, the full version of (1) in section 2, (2) in section 3, (3) in section 4, and (4) in section 5. The final section 6 contains some additional remarks.

Definitions and notation. We assume that the reader is familiar with the basic notions and results of bounded arithmetic as presented in e.g. [Bus86], [Kra95]. In particular, we assume familiarity with the meaning of the symbols $\#$, PV, S_2^n, T_2^n, Σ_n^b and Π_n^b and with notions such as "sharply bounded formula", "length induction", etc.

L_2 denotes the usual language of bounded arithmetic, with the symbols 0, 1, $\leq$, $+$, $\times$, $\#$, $|\cdot|$, and $\lfloor \frac{\cdot}{2} \rfloor$. The language L_1 is L_2 without the smash function symbol $\#$, so that functions definable by L_1-terms grow no faster than polynomials and increase the length of the arguments at most linearly. "Bounded formulae", or Σ_∞^b formulae, are bounded formulae of L_2. "Linearly bounded formulae" are the bounded formulae of L_1.

In the standard model of arithmetic, bounded formulae define exactly the relations in PH, while linearly bounded formulae define the relations in LinH. For this reason, in a nonstandard model of some arithmetical theory it is natural to identify PH with the sets definable by bounded formulae (with parameters from the model), and LinH with the sets definable by linearly bounded formulae (also with parameters). Thus, the precise formulation of

the question whether PH is contained in LinH in a given model is whether each relation on the model definable by a bounded formula is also definable by a linearly bounded formula, with parameters. Note that if parameters were not allowed, then PH $\subseteq$ LinH in a model would have to imply PH $\subseteq$ LinH in the real world.

We will normally work with the strict, or prenex, version of Σ^b_n, which we denote $\hat{\Sigma}^b_n$. A $\hat{\Sigma}^b_n$ formula has the form

$$\exists y_1 < t_1 \, \forall y_2 < t_2 \, \ldots Q y_n < t_n \, \psi,$$

where ψ is sharply bounded. In the standard model, both Σ^b_n and $\hat{\Sigma}^b_n$ formulae define predicates from the n-th level of PH (in particular, $\hat{\Sigma}^b_1$ formulae correspond to NP).

L_{PV} is the usual language of the theory PV, with a function symbol for each polynomial time function (in particular, L_{PV} contains L_2). We will often also use PV as a name for this language. A PV formula is an open formula of L_{PV}. Note that each sharply bounded formula is equivalent in PV to a PV formula. $\hat{\Sigma}^b_n(\mathrm{PV})$ is the class of formulae defined analogously to $\hat{\Sigma}^b_n$, but in the larger language L_{PV} (similarly for the non-strict version $\Sigma^b_n(\mathrm{PV})$). $\mathrm{S}^n_2(\mathrm{PV})$ is the theory axiomatized by PV plus the length induction scheme for $\hat{\Sigma}^b_n(\mathrm{PV})$ formulae. $\mathrm{S}^n_2(\mathrm{PV})$ is conservative over S^n_2, and we will often write "S^n_2", "$\hat{\Sigma}^b_n$", etc., when in fact we mean "$\mathrm{S}^n_2(\mathrm{PV})$", "$\hat{\Sigma}^b_n(\mathrm{PV})$", etc.

For technical reasons, we use a slight redefinition of the usual smash function: $x\#y$ is $2^{|x-1|\cdot|y-1|}$. In this way, we always have $2^\alpha \# 2^\beta = 2^{\alpha\cdot\beta}$. The notation $\#^c a$ is shorthand for $a\# \ldots \# a$, where a appears c times.

If a is an element of a model, $a^{\mathbb{N}}$ denotes the cut given by the standard powers of a, while $\#^{\mathbb{N}} a$ represents the cut consisting of numbers less than $\#^k a$ for some standard k. Also, we often identify a with the initial segment $[0, a)$ of numbers less than a, so that e.g. an "injection from b into a" is simply an injection from $[0, b)$ into $[0, a)$.

A bar, as in $\bar{x}$, indicates a tuple, and $\bar{x} < y$ means that all the elements of $\bar{x}$ are smaller than y.

Some useful facts about WPHP. We consider three basic versions of the weak pigeonhole principle. For $a < b$, the injective principle $\mathrm{iWPHP}^b_a(f)$

states that the function f is not an injection from b into a. The surjective principle $\mathrm{sWPHP}^a_b(f)$ states that f is not a surjection from a onto b. The multifunction principle $\mathrm{mWPHP}^b_a(R)$, introduced in [MPW02], states that R is not the graph of a injective multifunction from b into a.

We shall be particularly interested in the following schemes: iWPHP(PV), which is the scheme $\forall x\, \mathrm{iWPHP}^{x^2}_x(f)$, where f ranges over PV functions; sWPHP(PV), which is $\forall x\, \mathrm{sWPHP}^x_{x^2}(f)$ for $f \in \mathrm{PV}$; and $\mathrm{mWPHP}(\Sigma^b_1)$, which is a common strengthening of the previous two, given by $\forall x\, \mathrm{mWPHP}^{x^2}_x(R)$ for $R \in \Sigma^b_1$. All three schemes allow parameters in the definitions of f or R. In addition, we will also consider a parameter-free version of sWPHP(PV). This version is equivalent to sWPHP(PV) in S^1_2 ([Tha02]), but may be strictly weaker in PV.

By [PWW88], WPHP for all Δ_0 definable relations is provable in $\mathrm{I}\Delta_0+\Omega_1$. By [MPW02], all three schemes mentioned in the previous paragraph are provable in T^2_2.

It was shown by Wilkie (published in [Kra95] and [Tha05]) that the Σ^b_1 consequences of $\mathrm{S}^1_2 + \mathrm{sWPHP(PV)}$ can be witnessed in probabilistic polynomial time. On the other hand it was shown in [KP98], using Buss' witnessing theorem for S^1_2, that if iWPHP(PV) is provable in S^1_2 then the cryptosystem RSA is not secure against polynomial time attack. These arguments can be combined to show that $\mathrm{S}^1_2 + \mathrm{sWPHP(PV)}$ does not prove iWPHP(PV) if RSA is secure against randomized polynomial time attack. This was recently strengthened in [Jeř06] to: $\mathrm{S}^1_2 + \mathrm{sWPHP(PV)}$ does not prove iWPHP(PV) if there is no randomized polynomial time algorithm for factoring. Unprovability of iWPHP(PV) is also known to follow from the existence of a certain kind of hashing function ([Kra01]). No analogous assumptions implying unprovability of sWPHP(PV) are known.

An important technical property of weak pigeonhole principles is the possibility of amplifying their failure (see e.g. [Tha02]). Over PV, $\neg\mathrm{iWPHP}^{a^2}_a(\mathrm{PV})$ implies $\neg\mathrm{iWPHP}^b_a(\mathrm{PV})$ for any $b > a$. The only new parameter needed to define the amplified injection is b, and moreover, the additional parameter plays just the role of a size bound, so it may be replaced by any parameter which is at most polynomially smaller. In S^1_2, a similar amplification is possible for sWPHP(PV) and for $\mathrm{mWPHP}(\Sigma^b_1)$. It is open whether sWPHP(PV) can

be amplified in PV, but evidence from relativized theories ([Jeř06]) suggests that this is likely not to be the case.

1 NP and LinH in PV, a weak version

We start by proving a weak version of the statement that it is consistent with PV that NP is not contained in LinH. The weakening is twofold: we do not consider parameters and we exclude only provable equivalence in PV, instead of equivalence in a model. The result below follows immediately from Theorem 2.1, but we give the proof as a simple illustration of the main idea used in the next two sections.

Theorem 1.1. *If* $\mathrm{PV} + \mathrm{sWPHP(PV)} \nvdash \mathrm{iWPHP(PV)}$*, then there exists a* $\hat{\Sigma}^b_1$ *formula* $\varphi(x)$ *which is not equivalent in* PV *to any linearly bounded formula* $\psi^{\mathrm{lin}}(x)$.

Proof. Assume $\mathrm{PV} + \mathrm{sWPHP(PV)} \nvdash \mathrm{iWPHP(PV)}$. Under this assumption we can use amplification to get a model $\mathcal{A} \vDash \mathrm{PV}$ and an element $a \in \mathcal{A}$ such that:

(i) $\mathcal{A} \models \mathrm{sWPHP(PV)}$,

(ii) $f(q, \cdot)$ is an injection from $a\#a$ into a, where f is a PV function symbol and q is a parameter below a,

By (i) and compactness, moving to an elementary extension if necessary we may also assume that there is some element b in $\mathcal{A}$ realizing the type

$$\{b < a\#a\} + \{\forall \bar{x} < a,\ g(\bar{x}) \neq b : g \in \mathrm{PV}\}.$$

To see this, consider the finite fragment involving only the PV functions $g_1, \ldots, g_m$. Let r be the maximal arity of $g_1, \ldots, g_m$. If every element $< a\#a$ is the value of one of the functions g_i on some tuple of parameters $< a$, then we can define a polynomial time surjection from ma^r onto $a\#a$, contradicting sWPHP.

Now consider the following $\hat{\Sigma}^b_1$ formula $\varphi(x, a, q)$:

$$\exists w < a\#a\ f(q, w) = x.$$

Let $\psi^{\mathrm{lin}}(x, y, z)$ be any linearly bounded formula. We will show that

$$\mathrm{PV} \nvdash \forall x, y, z\, (\varphi(x, y, z) \equiv \psi^{\mathrm{lin}}(x, y, z)).$$

Define the model $\mathcal{B}$ to be the closure of $[0, a)$ in $\mathcal{A}$ under all PV functions. Now $b \in [0, a\#a)^{\mathcal{A}} \setminus \mathcal{B}$. Let $\hat{b}$ be the unique element of $[0, a)$ such that $\hat{b} = f(q, b)$.

Assume that in each model of PV, $\varphi(x, y, z)$ is equivalent to $\psi^{\mathrm{lin}}(x, y, z)$. We obtain a contradiction by analyzing the truth values of $\varphi(\hat{b}, a, q)$ and $\psi^{\mathrm{lin}}(\hat{b}, a, q)$ in $\mathcal{A}$ and $\mathcal{B}$. The formula $\psi^{\mathrm{lin}}(\hat{b}, a, q)$ must have the same value in both models, since $b, q < a$ and the models share $[0, a)$, and hence also $[0, a^{\mathbb{N}})$, as a common initial segment. Furthermore, $\varphi(\hat{b}, a, q)$ is clearly true in $\mathcal{A}$, as $b \in [0, a\#a)^{\mathcal{A}}$. Finally, $\varphi(\hat{b}, a, q)$ must be false in $\mathcal{B}$; otherwise the fact that $f(q, \cdot)$ is an injection would imply that $b \in \mathcal{B}$, contrary to our choice of b. □

Corollary 1.2. *If integer factoring is not possible in probabilistic polynomial time, then* PV *does not prove that each* $\hat{\Sigma}^b_1$ *formula is equivalent to a linearly bounded formula.*

Remark. Note that the proof of Theorem 1.1 actually establishes something stronger than the existence of a $\hat{\Sigma}^b_1$ formula φ which is not equivalent in PV to any linearly bounded formula. In fact, the formula φ cannot be equivalent in PV to any formula of the form

$$Q_1 y_1 < s_1\, Q_2 y_2 < s_2 \ldots Q_k y_k < s_k\, \psi,$$

where the Q_i are quantifiers, the s_i are L_1-terms, and ψ is a PV formula (and not just an open L_1-formula). This is because in the model $\mathcal{B}$ appearing in the proof the interpretations of all PV function symbols are inherited from the original model $\mathcal{A}$.

2 NP and LinH in PV

The present section is devoted to a proof of the following result, which is a considerable strengthening of Theorem 1.1:

Theorem 2.1. *If* PV + sWPHP(PV) $\nvdash$ iWPHP(PV)*, then there exists a model of* PV *and a* $\hat{\Sigma}^b_1$ *formula* $\varphi(x)$ *such that* $\varphi(x)$ *is not equivalent in the model to any linearly bounded formula* $\psi^{\text{lin}}(x,p)$ *for any parameter* p*.*

As before, assuming PV + sWPHP(PV) $\nvdash$ iWPHP(PV) we can get a model $\mathcal{A} \models$ PV and an element $a \in \mathcal{A}$ such that:

(i) $\mathcal{A} \models$ sWPHP(PV),

(ii) $f(q,\cdot)$ is an injection from $a\#a$ into a, where f is a PV function symbol and q is a parameter below a.

To make some calculations easier, it is not difficult to additionally ensure:

(iii) $a = 2^\alpha$ where $\alpha = |a| - 1$.

In any such model, the function f can be used to define a single PV function $\tilde{f}$ which is an injection from $c\#a$ into c for any c of the form $a\#\ldots\#a$ (where # could occur a nonstandard number of times, although that case will not be needed in this section).

To define $\tilde{f}$, we observe that we can treat any element u of the model as a sequence of numerals $[u]_i$ in base a notation and a sequence of numerals $\langle u\rangle_i$ in base $a\#a$ notation. In other words, $[u]_i$ is the number $< a$ consisting of bits $i\alpha,\ldots,(i+1)\alpha-1$ of u, for $i = 0,\ldots,\lceil\frac{|u|}{\alpha}\rceil - 1$, while $\langle u\rangle_i$ is the number $< a\#a$ consisting of bits $i\alpha^2,\ldots,(i+1)\alpha^2-1$ of u, for $i = 0,\ldots,\lceil\frac{|u|}{\alpha^2}\rceil - 1$. The function $\tilde{f}$ maps u to the unique element $\hat{u}$ such that $\lceil\frac{|\hat{u}|}{\alpha}\rceil = \lceil\frac{|u|}{\alpha^2}\rceil$ and $[\hat{u}]_i = f(p,\langle u\rangle_i)$ for all $i < \lceil\frac{|u|}{\alpha^2}\rceil$.

Note that the definition of $\tilde{f}$ needs no parameters other than q and possibly a (if a cannot be accessed from q by a PV function). Note also that $\tilde{f}$ coincides with f on $[0, a\#a)$. The only case in which $\tilde{f}$ can fail to be an injection is if f maps a number different from 0 to 0. To avoid this we will simply assume that $f(0) = 0$.

In our proof of Theorem 2.1 we will work with the $\hat{\Sigma}^b_1$ formula $\varphi(x,a,q)$ defined as:

$$\exists w \le (a\#a)x^\alpha \, \forall i < \lceil\frac{|x|}{\alpha}\rceil \, (f(q,\langle w\rangle_i) = [x]_i).$$

Thus, φ states that x is in the range of $\tilde{f}$. The number $(a\#a)x^\alpha$ is an upper bound on any w which could possibly be mapped by $\tilde{f}$ to x.

Proof of Theorem 2.1. Let $\mathcal{A}$ be our model satisfying the conditions (i), (ii) and (iii). Expand the language of PV by constant symbols for a, α, q and countably many new constants $c_1, c_2, \ldots$. Let T be the following theory in the expanded language:

$$\begin{aligned}\mathrm{PV} + \{q < a = 2^\alpha\} + \{\neg \mathrm{iWPHP}_a^{a\#a}(f(q,\cdot))\} + \{\#^k a \le c_k < \#^{k+1} a : k \ge 1\} \\ + \{\forall \bar{x} < \#^k a,\ c_k \ne g(\bar{x}, c_{i_1}, \ldots, c_{i_l}) : k \ge 1,\ m \ge 0, \\ g \in \mathrm{PV},\ c_k \text{ not among } c_{i_1}, \ldots, c_{i_l}\}.\end{aligned}$$

We claim that T is finitely consistent. Consider a finite fragment T_0 of T involving only the constants $c_1, \ldots, c_k$ and PV functions $g_1, \ldots, g_m$. We will satisfy T_0 by successively interpreting $c_k, \ldots, c_1$ as suitable elements of $\mathcal{A}$. Let c_k be any element of $[\#^k a, \#^{k+1} a)$ which is not the value of any g_i, $i = 1, \ldots, m$, on any tuple of arguments from $[0, \#^k a)$. Such an element must exist, by the same argument from sWPHP(PV) as in the previous section. Assuming $c_k, \ldots, c_{l+1}$ have already been assigned interpretations, let c_l be any element of $[\#^l a, \#^{l+1} a)$ which is not the value of any g_i, $i = 1, \ldots, m$, on any tuple of arguments from $[0, \#^l a)$ with $c_{l+1}, \ldots, c_k$ allowed as parameters. Again, the existence of such an element follows from sWPHP(PV) (with parameters).

Now take any countable model of T and let $\mathcal{B}$ be the submodel given by closing $\{a, q, c_1, c_2, \ldots\}$ under PV functions. T is still true in $\mathcal{B}$, as it is a universal theory. Note that the elements $c_1, c_2, \ldots$ in $\mathcal{B}$ enjoy a certain independence property: for each k, c_k is not contained in the PV-closure of $[0, \#^k a) \cup \{c_l : l \ne k\}$.

Enumerate all pairs consisting of a parameter from $\mathcal{B}$ and a linearly bounded formula in the variables x, y, z, t as $(p_k, \psi_k^{\mathrm{lin}})_{k \ge 1}$. We will now construct a descending chain $\mathcal{B} = \mathcal{B}_0 \supseteq \mathcal{B}_1 \supseteq \mathcal{B}_2 \ldots$ of substructures of $\mathcal{B}$ and an increasing sequence $0 = m_0 \le m_1 \le m_2 \ldots$ of natural numbers with the following properties, for $k \ge 1$:

1. if $p_k \in \mathcal{B}_k$, then $\mathcal{B}_k \vDash \exists x < \#^{m_k} a\, (\varphi(x, a, q) \not\equiv \psi_k^{\mathrm{lin}}(x, a, q, p_k))$,
2. if $p_k \in \mathcal{B}_k$, then $p_k < \#^{m_k} a$,
3. the initial segment $[0, \#^{m_k+1} a)$ is the same in all models $\mathcal{B}_k, \mathcal{B}_{k+1}, \ldots$,

4. the elements $c_{m_k+1}, c_{m_k+2}, \ldots$ are contained in $\mathcal{B}_k$.

If we succeed in constructing such sequences, the structure $\mathcal{C} = \bigcap_{k\in\mathbb{N}} \mathcal{B}_k$ will satisfy the requirements of the theorem. $\mathcal{C}$ is a model of PV since it is the intersection of a chain of models of PV, which is a universal theory. Furthermore, properties 1, 2, and 3, together with the fact that for x below $\#^{m_k}a$ the only possible witness w for the existential quantifier in $\varphi(x, a, q)$ is below $\#^{m_k+1}a$, will ensure that for each choice of ψ^{lin} and $p \in \mathcal{C}$,

$$\mathcal{C} \vDash \exists x\, (\varphi(x, a, q) \not\equiv \psi^{\text{lin}}(x, a, q, p)),$$

hence also

$$\mathcal{C} \vDash \exists x\, \exists y\, \exists z\, (\varphi(x, y, z) \not\equiv \psi^{\text{lin}}(x, y, z, p)).$$

We construct the sequences $(\mathcal{B}_k)$ and (m_k) inductively. Assume $\mathcal{B}_0, \ldots, \mathcal{B}_{k-1}$ and $m_0, \ldots, m_{k-1}$ have been chosen and consider $(p_k, \psi_k^{\text{lin}})$. If p_k is not an element of $\mathcal{B}_{k-1}$, we have nothing to do: let $\mathcal{B}_k = \mathcal{B}_{k-1}$ and $m_k = m_{k-1}$. If $p_k \in \mathcal{B}_{k-1}$, but there already is some x such that $\varphi(x, a, q)$ is not equivalent in $\mathcal{B}_{k-1}$ to $\psi_k^{\text{lin}}(x, a, q, p_k)$, the only thing we need to do is preserve this inequivalence: let $\mathcal{B}_k = \mathcal{B}_{k-1}$ and let m_k be the least number greater than m_{k-1} such that both x and p_k are below $\#^{m_k}a$.

The final case is when $p_k \in \mathcal{B}_{k-1}$ and

$$\mathcal{B}_{k-1} \vDash \forall x\, (\varphi(x, a, q) \equiv \psi_k^{\text{lin}}(x, a, q, p_k)).$$

Choose m_k to be any number strictly greater than m_{k-1} such that $p_k < \#^{m_k}a$ and let $\mathcal{B}_k$ be the closure of $[0, \#^{m_k}a) \cup \{c_{m_k+1}, c_{m_k+2}, \ldots\}$ in $\mathcal{B}_{k-1}$ under PV functions. Note that the elements $c_{m_k+1}, c_{m_k+2}, \ldots$ are contained in $\mathcal{B}_{k-1}$ by condition 4 for $\mathcal{B}_{k-1}$, and that our choice of $\mathcal{B}_k$ and m_k does not violate conditions 2, 3, and 4. It remains to check that condition 1 is satisfied.

By condition 4 for $\mathcal{B}_{k-1}$, we know that c_{m_k} is an element of $\mathcal{B}_{k-1}$, but by the independence property of the c_is, we also know that c_{m_k} is not an element of $\mathcal{B}_k$. Let $\hat{c}_{m_k}$ denote the image of c_{m_k} under $\tilde{f}$. Since $\hat{c}_{m_k} < \#^{m_k}a$, it must be the case that $\hat{c}_{m_k}$ is in $\mathcal{B}_k$. By the injectivity of $\tilde{f}$, $\varphi(\hat{c}_{m_k}, a, q)$ must be true in $\mathcal{B}_{k-1}$ and false in $\mathcal{B}_k$. On the other hand, $\psi_k^{\text{lin}}(\hat{c}_{m_k}, a, q, p_k)$ must have the same truth value (true) in $\mathcal{B}_{k-1}$ and $\mathcal{B}_k$, because $\hat{c}_{m_k}, a, q, p_k < \#^{m_k}a$ and the two models are the same on $[0, (\#^{m_k}a)^{\mathbb{N}})$. It follows that $\varphi(\hat{c}_{m_k}, a, q)$ is not

equivalent in $\mathcal{B}_k$ to $\psi_k^{\text{lin}}(\hat{c}_{m_k}, a, q, p_k)$, which completes the proof of condition 1 and of the whole theorem. □

Remark. This result can be improved to get a final model $\mathcal{C}$ which also satisfies a weak version of the surjective weak pigeonhole principle with parameters.

For each i, rather than introduce a single constant c_i between $\#^i a$ and $\#^{i+1} a$, we introduce countably many constants $c_i^1, c_i^2, \ldots$ in this interval. As above, we can use compactness to guarantee that each c_i^j is outside the PV closure of $[0, \#^i a) \cup \{c_{i'}^{j'} : (i', j') \neq (i, j)\}$.

When we construct our descending chain of models, it is enough to omit at most one c_i^j at each step. So we may assume that we have countably many c_i^js left in each interval $[\#^i a, \#^{i+1} a)$ in $\mathcal{C}$ and that these are all independent.

Now let $\gamma = |a|^2$. We claim that for every $d \in \mathcal{C}$ and every PV function g, with parameters, g is not a surjection in $\mathcal{C}$ from d onto d^γ.

To see this, notice that if k is the largest number such that $d \leq \#^k a$, then $\#^{k+1} a \leq d^\gamma$. We may assume that the parameters for g are some numbers below a together with a tuple $\bar{c}$ of finitely many of the constants c_i^j. So there is some c_k^l not included in $\bar{c}$, and by construction c_k^l is not in the range of g on the domain $[0, d)$.

3 NP and the second level of LinH in S_2^1

It seems that the technique used to prove Theorem 2.1 cannot be extended to S_2^1. This is because all the models constructed in the proof of Theorem 2.1 satisfy ¬iWPHP(PV), while the proof of Theorem 4.1 in the next section suggests that in S_2^1 failure of any kind of WPHP for Σ_1^b relations actually leads to PH being equal to LinH. Nevertheless, it turns out that a limited extension of Theorem 2.1 to S_2^1 is possible. We prove a result which states roughly that (under our general assumption about pigeonhole principles) it is consistent with S_2^1 that NP is not contained in the second level of the linear time hierarchy.

Definition 3.1. A $\hat{\Sigma}_2^{\text{lin}}$(PV) formula is a formula of the form

$$\exists y_1 < s_1 \, \forall y_2 < s_2 \, \xi,$$

where s_1, s_2 are L_1-terms and ξ is a PV formula.

Theorem 3.2. *If* PV + sWPHP(PV) $\nvdash$ iWPHP(PV)*, then there exists a model of* S^1_2 *and a* $\hat{\Sigma}^b_1$ *formula* $\varphi(x)$ *such that* $\varphi(x)$ *is not equivalent in the model to any* $\hat{\Sigma}^{\mathrm{lin}}_2(\mathrm{PV})$ *formula* $\psi^{\mathrm{lin}}(x,p)$ *for any parameter* p*.*

We will use the function $\tilde{f}$ and the $\hat{\Sigma}^b_1$ formula $\varphi(x,a,q)$ from the previous section. The main tool needed to prove Theorem 3.2 is the following lemma:

Lemma 3.3. *Let* T *be* PV + sWPHP(PV) + $\neg$iWPHP$^{a\#a}_a(f(q,\cdot))$*. Assume* $\mathcal{A}$ *is a countable model of* T*. Let* $p \in \mathcal{A}$ *and let* $\psi^{\mathrm{lin}}(x,y,z,t)$ *be a* $\hat{\Sigma}^{\mathrm{lin}}_2(\mathrm{PV})$ *formula,*

$$\psi^{\mathrm{lin}}(x,y,z,t) = \exists u_1 \le s_1\, \forall u_2 \le s_2\, \xi(x,y,z,t,u_1,u_2).$$

Then there exists a countable $\mathcal{B} \succeq_{\hat{\Sigma}^b_1} \mathcal{A}$ *with* $\mathcal{B} \vDash \mathrm{S}^1_2 + T$ *and* $x \in \mathcal{B}$ *such that one of the following holds in* $\mathcal{B}$*:*

(a) $\varphi(x,a,q)$ *is false and* $\psi^{\mathrm{lin}}(x,a,q,p)$ *is true, or*

(b) $\varphi(x,a,q)$ *is true,* $\psi^{\mathrm{lin}}(x,a,q,p)$ *is false, and there is a* PV *function* h *(with a parameter from* $\mathcal{B}$*) which for each given* $u_1 < s_1$ *outputs some* $u_2 < s_2$ *such that* $\neg\xi(x,a,q,p,u_1,u_2)$*.*

Theorem 3.2 follows from the lemma by a straightforward chain construction. Given a countable $\mathcal{A} \vDash$ PV satisfying (i), (ii) and (iii) from the previous section, we can iterate Lemma 3.3 countably many times, once for each choice of a $\hat{\Sigma}^{\mathrm{lin}}_2(\mathrm{PV})$ formula and a parameter $p \in \mathcal{A}$. Note that by $\hat{\Sigma}^b_1$-elementarity, if at some point we have a witness of type (a) or (b) that φ is not equivalent to a given $\hat{\Sigma}^{\mathrm{lin}}_2(\mathrm{PV})$ formula with a given parameter, then it will remain such a witness in successive steps of the iteration. This is clear in case (a), since the truth of a $\hat{\Sigma}^{\mathrm{lin}}_2(\mathrm{PV})$ formula is preserved upwards under $\hat{\Sigma}^b_1$-elementary extensions. In case (b), we can use the function h to express the falsity of the $\hat{\Sigma}^{\mathrm{lin}}_2(\mathrm{PV})$ formula in a $\hat{\Pi}^b_1$ way (for each u_1, h outputs a u_2 such that $\neg\xi$ holds), so it will be preserved as well.

Taking the union of the chain of models obtained during the iteration, we will get a countable model $\mathcal{A}^* \succeq_{\hat{\Sigma}^b_1} \mathcal{A}$ which satisfies $\neg$iWPHP$^{a\#a}_a(f(q,\cdot))$ and

in which $\varphi(x, a, q)$ is not equivalent to any $\hat{\Sigma}_2^{\text{lin}}(\text{PV})$ formula $\psi^{\text{lin}}(x, a, q, p)$ for any $p \in \mathcal{A}$. Moreover, $\mathcal{A}^*$ satisfies S_2^1 and sWPHP(PV), which can be seen as follows. S_2^1 is $\forall\exists\text{bool}(\hat{\Sigma}_1^b)$-axiomatizable, so it is preserved in unions of $\hat{\Sigma}_1^b$-elementary chains. The parameter-free version of sWPHP(PV) is $\forall\exists\hat{\Pi}_1^b$-axiomatizable, hence also true in $\mathcal{A}^*$, and it is known that in S_2^1 parameter-free sWPHP(PV) implies full sWPHP(PV) [Tha02].

Now let $\mathcal{A}_0$ be a countable model of PV satisfying conditions (i), (ii) and (iii), and for each $n \in \mathbb{N}$, let $\mathcal{A}_{n+1} = (\mathcal{A}_n)^*$. It is not difficult to check that the model $\bigcup_{n\in\mathbb{N}} \mathcal{A}_n$ satisfies the thesis of Theorem 3.2.

Thus, to show that Theorem 3.2 is true it remains to prove Lemma 3.3. In the proof, we will use the following theorem, adapted from Zambella [Zam96]:

Theorem 3.4. *Every countable model $\mathcal{A}$ of* PV *has a $\hat{\Sigma}_1^b$-elementary extension to a cofinal countable model $\mathcal{B} \vDash \text{S}_2^1$ with the following "witnessing property": for each* PV *formula $\xi(x, y, p)$ there is a* PV *function g and a parameter q such that*

$$\mathcal{B} \vDash \forall x{<}u\, \exists y\, \xi(x, y, p) \to \forall x{<}u\, \xi(x, g(x, u, p, q), p).$$

Proof. We sketch how the proof, as presented in section 7.6 of [Kra95], can be modified to give cofinality. Add names for all elements of $\mathcal{A}$ to the language and take a new set of constant symbols $\{c_b : b \in \mathcal{A}\}$ indexed by elements of $\mathcal{A}$. Let T_0 be the universal diagram of $\mathcal{A}$ together with $\{c_b < b : b \in \mathcal{A}\}$. Enumerate all sentences of the form $\forall x < u\, \exists y\, \xi(x, y)$ in the expanded language, with all the new constant symbols. We construct a chain (T_n) of universal theories, beginning with T_0. Suppose we consider the sentence $\forall x < u\, \exists y\, \xi(x, y)$ at stage n in the construction. If $T_n \vdash \forall x < u\, \exists y\, \xi(x, y)$ we put $T_{n+1} = T_n$, otherwise we put $T_{n+1} = T_n \cup \{c_b < u \wedge \forall y\, \neg\xi(c_b, y)\}$ where $b > u$ and c_b has not appeared yet (except in T_0). We eventually obtain a model of the universal theory $\bigcup_{n\in\mathbb{N}} T_n$, and our model $\mathcal{B}$ is the substructure formed by closing $\mathcal{A}$ and the new constant symbols under all PV functions. By the construction, none of the new constants is above $\mathcal{A}$. □

Proof of Lemma 3.3. Let $\mathcal{A}$ satisfy the assumptions of the lemma. Extend $\mathcal{A}$ cofinally and $\hat{\Sigma}_1^b$-elementarily to a model $\mathcal{A}' \vDash \text{S}_2^1$ with the witnessing property of Theorem 3.4. Then $\mathcal{A}'$ is also a model of T: the failure of the

injective WPHP is preserved by $\hat{\Sigma}^b_1$-elementarity. By cofinality and S^1_2 in $\mathcal{A}'$, in order to show that $\mathcal{A}' \vDash$ sWPHP(PV) it is enough to check that in $\mathcal{A}'$ there is no parameter-free PV surjection from c to c^2 for any c from $\mathcal{A}$. But in $\mathcal{A}$ for any function from c to c^2 there is an element outside its range, and this is preserved in $\mathcal{A}'$ by $\hat{\Sigma}^b_1$-elementarity.

Now there are three cases to consider.

Case 1. There exists some $x \in \mathcal{A}'$ such that $\varphi(x, a, q)$ is false and $\psi^{\mathrm{lin}}(x, a, q, p)$ is true. In this case we simply take $\mathcal{B}$ to be $\mathcal{A}'$. Clearly, (a) holds.

Case 2. Case 1 does not hold, but there exists some $x \in \mathcal{A}'$ such that $\varphi(x, a, q)$ is true and $\psi^{\mathrm{lin}}(x, a, q, p)$ is false. Then by the witnessing property we can guarantee that (b) holds in $\mathcal{A}'$. Again we take $\mathcal{B}$ to be $\mathcal{A}'$.

Case 3. For each $x \in \mathcal{A}'$, $\varphi(x, a, q)$ is equivalent to $\psi^{\mathrm{lin}}(x, a, q, p)$. We would like to apply the by now familiar argument. Unfortunately, we have to be careful, since there is no guarantee that a structure obtained by taking the PV-closure of an initial segment will satisfy sWPHP(PV).

By compactness and sWPHP(PV) in $\mathcal{A}'$, we may move to an elementary extension $\mathcal{A}''$ of $\mathcal{A}'$ which contains: an element $b > \mathcal{A}$ of the form $a\# \ldots \# a$; a number $d > \#^{\mathbb{N}} b$; some $e < d^4$ which is not the value of any PV function without parameters on any argument below d^2; and some $w < b\#a$ which is not the value of any PV function with parameters d, e on any tuple of arguments from below b.

Consider the closure $\mathcal{C}$ of $[0, b) \cup \{d, e\}$ in $\mathcal{A}''$ under PV functions. This is a Σ^b_∞-elementary extension of $\mathcal{A}'$, and therefore a $\hat{\Sigma}^b_1$-elementary extension of $\mathcal{A}$. Moreover, we can apply our standard argument. If we let $\hat{w} = \tilde{f}(w)$, then $\varphi(\hat{w}, a, q)$ is true in $\mathcal{A}''$, hence $\psi^{\mathrm{lin}}(\hat{w}, a, q, p)$ is true in $\mathcal{A}''$ (by our equivalence assumption) and so must also be true in $\mathcal{C}$. But $\varphi(\hat{w}, a, q)$ is false in $\mathcal{C}$, so (a) holds.

To complete the proof, extend $\mathcal{C}$ $\hat{\Sigma}^b_1$-elementarily to a model $\mathcal{D}$ of S^1_2 and let $\mathcal{B}$ be the cut in $\mathcal{D}$ determined by $\#^{\mathbb{N}} b$. We now need to show that $\mathcal{B}$ has the properties required by the lemma. Certainly, $\mathcal{B}$ is a $\hat{\Sigma}^b_1$-elementary extension of $\mathcal{A}$ to a model of $\mathrm{S}^1_2 + \neg\mathrm{iWPHP}^{a\#a}_a(f(q, \cdot))$. Also certainly, (a) holds in $\mathcal{B}$ since it did in $\mathcal{C}$. It remains to verify that $\mathcal{B} \vDash$ sWPHP(PV).

Otherwise, for some $c \in \mathcal{B}$ there is a PV function g which maps c onto c^2.

By S^1_2, we may assume that the definition of g does not use any parameters. Since c is contained in $\#^{\mathbb{N}} b$ and is thus smaller than d, the function g can be modified in $\mathcal{D}$ to yield a surjection $\tilde{g}$ from c onto d^4. S^1_2 is enough to perform such an amplification, and the only parameter needed to define $\tilde{g}$ is, say, d as a size bound. But this means that if we treat the parameter as part of the argument, $\tilde{g}$ is a surjective map in $\mathcal{D}$ from cd to d^4. This is a contradiction, since $\mathcal{D}$ contains the element $e < d^4$ which is not the value of any PV function on an argument below d^2. □

4 Collapsing PH to LinH

Theorem 4.1. *If* $\mathrm{S}^1_2 \nvdash \mathrm{mWPHP}(\Sigma^b_1)$*, then there exists a model of* S^1_2 *and a parameter* p *such that every bounded formula* $\varphi(x)$ *is equivalent in the model to some linearly bounded formula* $\varphi^{\mathrm{lin}}(x, p)$*.*

To prove the theorem, assume that $\mathrm{S}^1_2 \nvdash \mathrm{mWPHP}(\Sigma^b_1)$. This means that there exists a countable model $\mathcal{A} \vDash \mathrm{S}^1_2$ containing an element $a = 2^\alpha$ such that the Σ^b_1 formula $\zeta(x, y)$ defines an injective multifunction from $a\#a$ into a. The formula ζ may involve a parameter q, but we may assume w.l.o.g. that $q < a$ and that all quantifiers in ζ are bounded by at most $a\#a$. By abuse of notation, we will also refer to the multifunction itself as ζ. We may also assume that $\mathcal{A}$ contains the element $b = \#^c a$ and the element $\#^{3c} a$ for some small nonstandard c. Note that b is also equal to $a^{\alpha^{c-1}}$.

Fix such a model $\mathcal{A}$ for the remainder of this section and let $\mathcal{B}$ be the (proper) cut $\#^{\mathbb{N}} a$ in $\mathcal{A}$. We will show that in $\mathcal{B}$, each bounded formula is equivalent to a linearly bounded formula with parameters $a\#a, c$, and q. Our argument will be based on a construction analogous to the one in [Tha02], which was in turn inspired by [PWW88].

We will show ζ can be used to code each element $u < b$, hence, in particular, each element of $\mathcal{B}$, as a (possibly non-unique) element $\hat{u} < a$. In this way, statements about elements of $\mathcal{B}$ can be translated into statements about their codes. Moreover, the coding can be defined by a linearly bounded formula, which will allow us to perform the translation of bounded into linearly bounded formulae required to obtain Theorem 4.1.

As in earlier sections, we think of each element $u \in \mathcal{A}$ as a sequence of numerals $[u]_i$ in base a notation. If $u < b$, then this sequence will have length at most α^{c-1}. Intuitively, we would like to treat the number $\hat{u} < a$ as a code for u if there exists an α-branching labelled tree of depth $c-1$ with the root labelled by $\hat{u}$, the α^{c-1} leaves labelled by the $[u]_i$s in the correct order, and such that if the sons of some node are labelled by numerals together representing a number $z < a\#a$, then the node itself is labelled by some $y < a$ such that $\zeta(z) = y$.

The natural definition of this coding requires an existential quantifier for the tree, i.e. essentially for a sequence of numerals of length $1 + \alpha + \ldots + \alpha^{c-1}$, or $\frac{\alpha^c - 1}{\alpha - 1}$. This object will typically be larger than b, so there is no hope of referring to it by a linearly bounded formula in $\mathcal{B}$. The way around this obstacle is to speak not about the entire tree, but about the individual branches, requiring each of them to end in the appropriate digit of u. More formally, let $[[\hat{u}]]_i = x$ be the following formula (i is understood to be a sequence $(i_1, \ldots, i_{c-1})$, where each i_j is smaller than α; such a sequence determines a branch in an α-branching tree of depth $c-1$):

$$\exists w = (w_0, \ldots, w_{c-1}),\ \forall j < c\,(w_j < a) \wedge w_0 = \hat{u} \wedge w_{c-1} = x \\ \wedge\ \forall j < c-1\ \exists z < a\#a\,(\zeta(z) = w_j \wedge [z]_{i_{j+1}} = w_{j+1}).$$

The intended sense of the formula $[[\hat{u}]]_i = x$ is that in the coding tree with $\hat{u}$ at the root there is a labelling of the branch given by i, and the leaf at the end of that branch is labelled by x. Let $\mathrm{code}(\hat{u}, u)$, "$\hat{u}$ is a code for u", be $\forall i < \alpha^{c-1}\,([[\hat{u}]]_i = [u]_i)$. Thus, $\mathrm{code}(\hat{u}, u)$ states that each branch of the coding tree with $\hat{u}$ in the root ends in a leaf labelled by the appropriate numeral of u. Finally, let $\mathrm{C}(\hat{u})$, "$\hat{u}$ is a code", be $\forall i < \alpha^{c-1}\ \exists x < a\,[[\hat{u}]]_i = x$. An equivalent formula is obtained by simply deleting the conjunct $w_{c-1} = [u]_i$ in $\mathrm{code}(\hat{u}, u)$.

Note that all three formulae are linearly bounded, assuming $a\#a$ and c are treated as parameters. In fact, $\mathrm{code}(\hat{u}, u)$ is the only one which may refer at all to objects larger than $a\#a$ (we are assuming that c is much smaller than α, so the number a^c needed to bound w is much smaller than $a\#a$). Moreover, all three are (non-strict) Σ^b_1 formulae with parameters in $\mathcal{A}$, which means they can be used in arguments by length induction. We now prove a

lemma which states that our coding apparatus works as it should.

Lemma 4.2. *Let* $[[\hat{u}]]_i = x$, $\mathrm{code}(\hat{u}, u)$, $\mathrm{C}(\hat{u})$ *be as above. Then the following hold in* $\mathcal{A}$*:*

(a) for each $u < b$*, there exists some (not necessarily unique)* $\hat{u} < a$ *such that* $\mathrm{code}(\hat{u}, u)$*;*

(b) for each $y < a$ *such that* $\mathrm{C}(y)$ *and each sequence* i*, there is exactly one* $x < a$ *such that* $[[y]]_i = x$*;*

(c) for each $y < a$ *such that* $\mathrm{C}(y)$*, there is exactly one* $u < b$ *such that* $\mathrm{code}(y, u)$*.*

Proof. To prove part (a), one may show that each $u < b$ is coded in the intuitive sense, i.e. that there is a labelled tree with leaves labelled by successive numerals of u and all branches labelled as required. Obviously, any element $\hat{u}$ which is in the root of some such tree will also satisfy $\mathrm{code}(\hat{u}, u)$. The existence of a coding tree is proved inductively level by level. The base step is for the leaves, which are simply numerals of u. In the inductive step, we assume that there is a correct labelling for levels $j+1, \ldots, c-1$ of a potential coding tree and we need to extend the labelling to level j. We label the nodes on level j also inductively, say from left to right, and the inductive step consists in simply choosing some value of ζ on the number represented by the labels of the sons of a given node.

In part (b), the existence condition follows from the definition of the formula $\mathrm{C}(y)$. Now assume that $[[y]]_i = x$ and $[[y]]_i = x'$, and that w and w' are the respective witnessing strings. Then induction on $j < c$ shows that $w_j = w'_j$ for each j: w_0 and w'_0 are both equal to y, while the inductive step uses the injectivity of ζ. Since $w_{c-1} = x$ and $w'_{c-1} = x'$, we obtain that $x = x'$.

In part (c), existence is an easy consequence of the definition of $\mathrm{C}(y)$ and (non-strict) Σ^b_1 replacement. Uniqueness follows from part (b). □

The following is the main technical lemma needed in this and the next section:

Lemma 4.3. *Let $\psi(x_1, \ldots, x_n)$ be an L_2-formula with all quantifiers bounded by b and with all function symbols appearing as relations (i.e. exclusively in atomic formulae of the form $y_1 + y_2 = y_3$, $y_1 \# y_2 = y_3$, etc., where y_1, y_2, y_3 are variables). Then there exists a linearly bounded formula $\tilde{\psi}$, with free variables $\hat{x}_1, \ldots, \hat{x}_n$ and $a \# a, c, q$ as parameters, with the following property: for any $u_1, \ldots, u_n < b$ and any $\hat{u}_1, \ldots \hat{u}_n$ such that $\mathrm{code}(\hat{u}_i, u_i)$ for each i, $\psi(u_1, \ldots, u_n)$ is equivalent in $\mathcal{A}$ to $\tilde{\psi}(\hat{u}_1, \ldots \hat{u}_n, a \# a, c, q)$.*

Proof. The translation required to obtain the lemma was essentially presented in the proof of Theorem 3.7 in [Tha02]. In that paper, it was assumed that $b = \#^l a$ where l is standard, but this does not have a significant influence on the translation. An additional difference between our translation and that of [Tha02] is that we are not concerned about quantifier complexity, so there is no need to have a separate translation for sharply bounded quantifiers.

The translation is defined by induction on the structure of the formula, with the step for atomic formulae requiring the most effort. The details are straightforward but somewhat tedious to describe, so we only sketch a few cases.

If ψ is $x < y$, then $\tilde{\psi}$ is (omitting some existential quantifiers):

$$\exists i < \alpha^{c-1} \left([[\hat{x}]]_i < [[\hat{y}]]_i \wedge \forall i < j < \alpha^{c-1} \, [[\hat{x}]]_j = [[\hat{y}]]_j \right).$$

A similar, simpler, translation is needed for $x = y$.

To translate $x + y = z$, we think of computing the sum of x and y in base a. To express this in terms of $\hat{x}, \hat{y}$, and $\hat{z}$, we need to introduce an existential quantifier for an auxiliary number $w < a$ coding the values of the carry function appearing during the computation, and then state that the relations between $[[x]]_i$, $[[y]]_i$, $[[w]]_{i-1}$ and $[[z]]_i$, $[[w]]_i$ are as required. The case of multiplication is similar except that here the auxiliary object we need to encode has to be some form of multiplication table, i.e. a number whose length in base a notation is roughly quadratic in c. The translation in [Tha02] uses two such tables, each consisting of $(2\alpha^c + 1)\alpha^{2c}$ numerals. The value $(2\alpha^c + 1)\alpha^{2c}$ is smaller than $3\alpha^{3c}$, so we may use the fact that $\#^{3c} a$ exists to define a separate new encoding of numbers consisting of up to $3c$ numerals, completely analogously to the encoding described above, and

then state that the numerals encoded by $\hat{x}, \hat{y}, \hat{z}$ (in the old encoding) and the entries in the multiplication tables (in the new encoding) are related as they should be. Note that since we already have $a\#a$, c as parameters and since a^{3c} is much smaller than $a\#a$, the definition of the new encoding does not require any new parameters.

Once addition and multiplication are translated, the cases of the remaining function symbols $|x| = y$, $x\#y = z$ and $\lfloor \frac{x}{2} \rfloor = y$ are relatively unproblematic.

Finally, if ψ is $\exists x_0 < b\, \chi(x_0, x_1, \ldots, x_n)$, then $\tilde{\psi}$ is:

$$\exists \hat{x}_0 < a\, (\mathrm{C}(\hat{x}_0) \wedge \tilde{\chi}(\hat{x}_0, \hat{x}_1, \ldots, \hat{x}_n, a\#a, c, q)).$$

The case of the universal quantifier is handled similarly.

The correctness of the translation is proved inductively in a straightforward way. (Non-strict) Σ^b_1 length induction is needed for the atomic formulae; the inductive step for the quantifiers uses Lemma 4.2. □

Once we have Lemma 4.3, the rest of the proof of Theorem 4.1 is straightforward. Let $\varphi(x)$ be any bounded formula. Form $\varphi_b(x)$ by first "unwinding" all terms in φ so that all function symbols appear as relations, which will lead to the introduction of some new quantifiers, and then relativizing all quantifiers to b. The resulting formula is equivalent to $\varphi(x)$ for all arguments from $\mathcal{B}$. By Lemma 4.3, there is a linearly bounded formula $\widetilde{\varphi}_b(x, a\#a, c, q)$ (which no longer has b as a parameter) such that $\varphi_b(u)$ is equivalent to $\widetilde{\varphi}_b(\hat{u}, a\#a, c, q)$ for any $u < b$ and any $\hat{u}$ coding u. Let $\varphi^{\mathrm{lin}}(x, a\#a, c, q)$ be:

$$\exists \hat{x} < a\, (\mathrm{code}(\hat{x}, x) \wedge \widetilde{\varphi}_b(\hat{x}, a\#a, c, q)).$$

This is a linearly bounded formula with parameters $a\#a, c, q$. Moreover, it must be equivalent to $\varphi(x)$ in $\mathcal{B}$, since every element of $\mathcal{B}$ has a code below a by Lemma 4.2 part (a). As $\varphi(x)$ was arbitrary, Theorem 4.1 is now proved.

5 Failure of WPHP implies non-collapse of the polynomial hierarchy

Theorem 5.1. $\mathrm{S}^1_2 + \neg\mathrm{mWPHP}(\Sigma^b_1)$ *proves that the strict version of* PH *does not collapse, even allowing parameters. That is, for each model* $\mathcal{A}$ *of*

$S_2^1 + \neg\text{mWPHP}(\Sigma_1^b)$ *and each natural number* m*, there is a bounded formula* $\varphi(x)$ *which is not equivalent in* $\mathcal{A}$ *to any* $\hat{\Sigma}_m^b$ *formula* $\psi(x,p)$ *for any choice of a parameter* $p \in \mathcal{A}$*.*

Corollary 5.2. *If factoring is not in probabilistic polynomial time, then there is a model of* $S_2^1 + \text{sWPHP(PV)}$ *in which the strict version of* PH *does not collapse, even allowing parameters. The same holds in a model of* S_2^1 *if RSA is secure against deterministic polynomial time attack.*

The remainder of this section contains a proof of Theorem 5.1. The proof is by diagonalization and is based on a similar argument for $I\Delta_0$ from [PW85]. We will also employ the machinery of the previous section, especially Lemma 4.3.

Suppose that the theorem is not true, i.e. that there exists a number m and a model $\mathcal{A} \models S_2^1 + \neg\text{mWPHP}(\Sigma_1^b)$ in which each bounded formula is equivalent to a $\hat{\Sigma}_m^b$ formula with some parameter.

By compactness and a standard argument based on amplifying the failure of mWPHP, we may pass to an elementary extension $\mathcal{A}'$ of $\mathcal{A}$ which contains the following: a number $a = 2^\alpha$ such that $a > \mathcal{A}$ and there is a Σ_1^b injective multifunction ζ from $a\#a$ into a; a number $t > \#^{\mathbb{N}}a$; a number b of the form $\#^c a$ such that $b > \#^{\mathbb{N}}t$; and the number $\#^{3c}a$. Let $\mathcal{B}$ be the initial segment $\#^{\mathbb{N}}a$ in $\mathcal{A}'$. Note that the relation between $\mathcal{A}'$ and $\mathcal{B}$ is exactly as between $\mathcal{A}$ and $\mathcal{B}$ in the previous section.

Since $\mathcal{A}' \succeq \mathcal{A}$, it remains true in both $\mathcal{A}'$ and $\mathcal{B}$ that each bounded formula ϕ is equivalent to a $\hat{\Sigma}_m^b$ formula with a parameter p_ϕ from $\mathcal{A}$. We will now show that this leads to a contradiction.

Since $t > \mathcal{B}$, there is a universal $\hat{\Sigma}_m^b$ formula U_m such that for all $x, y \in \mathcal{B}$ and all $\hat{\Sigma}_m^b$ formulae ψ, $\psi(x,y)$ is equivalent to $U_m(x, (\ulcorner\psi\urcorner, y), t)$.

U_m is bounded, so for $x, y \in \mathcal{B}$ and for standard ψ the quantifiers in $U_m(x, (\ulcorner\psi\urcorner, y), t)$ range only over numbers well below b, by our choice of b. Therefore, we may equivalently present $U_m(x, (\ulcorner\psi\urcorner, y), t)$ in a form to which Lemma 4.3 is applicable. As a result, for all $x, y \in \mathcal{B}$ and all ψ, $U_m(x, (\ulcorner\psi\urcorner, y), t)$ is equivalent to $U_m^{\text{lin}}(x, (\ulcorner\psi\urcorner, y), \hat{t}, a\#a, c, q)$, where $\hat{t}$ is a number below a coding t and U_m^{lin} is a linearly bounded formula.

A linearly bounded formula is, in particular, a bounded formula. It follows that $\neg U_m^{\text{lin}}(x, x, \hat{t}, a\#a, c, q)$ must be equivalent in $\mathcal{B}$ to a $\hat{\Sigma}_m^b$ for-

mula with parameters $\hat{t}, a\#a, c, q$, and p_U, or more briefly, to a $\hat{\Sigma}^b_m$ formula $\varphi(x, p)$ where p is some parameter from $\mathcal{B}$. Consider $\varphi((\ulcorner\varphi\urcorner, p), p)$. By the properties of U_m, this is equivalent to $U_m((\ulcorner\varphi\urcorner, p), (\ulcorner\varphi\urcorner, p), t)$, hence to $U^{\mathrm{lin}}_m((\ulcorner\varphi\urcorner, p), (\ulcorner\varphi\urcorner, p), \hat{t}, a\#a, c, q)$, hence to $\neg\varphi((\ulcorner\varphi\urcorner, p), p)$. This is a contradiction, which completes the proof of Theorem 5.1.

6 Concluding remarks

All of our main theorems require the assumption that some version of the weak pigeonhole principle is unprovable in S^1_2 (since the unprovability of iWPHP(PV) in PV + sWPHP(PV) is actually equivalent to its unprovability in S^1_2 + sWPHP(PV)). Thus, it might be interesting to look for other natural and plausible statements from cryptography or proof complexity, perhaps weaker than the ones about factoring or RSA, which would imply some such unprovability result. It also seems worthwile to search for a natural computational assumption which would imply that some variant of WPHP for Σ^b_1 relations is not provable in S^2_2. By Theorem 5.1, such an assumption would allow us to conclude that it is consistent with S^2_2 that the strict version of PH does not collapse.

It should also be noted that all of our results (formulated in terms of unprovability of WPHP) relativize to higher levels of the bounded arithmetic hierarchy. Thus, if T^n_2 plus the surjective WPHP for $\square^p_{n+1}$ functions does not prove the injective WPHP for $\square^p_{n+1}$ functions, then T^n_2 does not prove that Σ^p_{n+1} is contained in LinH, and S^{n+1}_2 does not prove that Σ^p_{n+1} is contained in the $(n+2)$-nd level of LinH; if S^{n+1}_2 does not prove mWPHP(Σ^b_{n+1}), then S^{n+1}_2 does not prove that PH is not contained in LinH; finally, S^{n+1}_2 + $\neg$mWPHP(Σ^b_{n+1}) implies that the strict version of PH does not collapse.

Unlike the case of $n = 0$ and factoring, for higher n no natural computational assumptions implying the unprovability of appropriate versions of WPHP are known. Nevertheless, it seems plausible that iWPHP($\square^p_{n+1}$) is unprovable in T^n_2 + sWPHP($\square^p_{n+1}$). If this is indeed the case for all n, then the question whether LinH = PH is independent of each finite fragment of bounded arithmetic.

Acknowledgements. The authors would like to thank Pavel Pudlák, for

inspiring one of them to think about when initial segments of a model determine the whole model; Jan Krajíček, for numerous discussions and helpful remarks; and Emil Jeřábek, for pointing out an error in the original version of one of the proofs.

References

[Bus86] S. Buss, *Bounded Arithmetic*, Bibliopolis, 86.

[CT06] S. Cook and N. Thapen, *The strength of replacement in weak arithmetic*, ACM Transactions on Computational Logic **7** (2006), no. 4.

[Jeř06] E. Jeřábek, *On independence of variants of the weak pigeonhole principle*, 2006, preprint.

[KP98] J. Krajíček and P. Pudlák, *Some consequences of cryptographical conjectures for* S_2^1 *and* EF, Information and Computation **140** (1998), 82–89.

[Kra95] J. Krajíček, *Bounded arithmetic, propositional logic, and complexity theory*, Cambridge University Press, 1995.

[Kra01] ———, *On the weak pigeonhole principle*, Fundamenta Mathematicae **170** (2001), 123–140.

[MPW02] A. Maciel, T. Pitassi, and A. R. Woods, *A new proof of the weak pigeonhole principle*, Journal of Computer and System Sciences **64** (2002), 843–872.

[PW85] J. B. Paris and A. J. Wilkie, *Counting problems in bounded arithmetic*, Methods in Mathematical Logic, Lecture Notes in Mathematics, vol. 1130, Springer-Verlag, 1985, pp. 317–340.

[PWW88] 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** (1988), 1235–1244.

[Tha02] N. Thapen, *A model-theoretic characterization of the weak pigeonhole principle*, Annals of Pure and Applied Logic **118** (2002), 175–195.

[Tha05] ———, *Structures interpretable in models of bounded arithmetic*, Annals of Pure and Applied Logic **136** (2005), 247–266.

[Zam96] D. Zambella, *Notes on polynomially bounded arithmetic*, Journal of Symbolic Logic **61** (1996), 942–966.