World Scientific
www.worldscientific.com

Computable aspects of the Bachmann–Howard principle

Anton Freund

Fachbereich Mathematik, Technische Universität Darmstadt
Schlossgartenstr. 7, 64289 Darmstadt, Germany
freund@mathematik.tu-darmstadt.de

We have previously established that Π_1^1-comprehension is equivalent to the statement that every dilator has a well-founded Bachmann–Howard fixed point, over $\mathbf{ATR_0}$. In this paper, we show that the base theory can be lowered to $\mathbf{RCA_0}$. We also show that the minimal Bachmann–Howard fixed point of a dilator T can be represented by a notation system $\vartheta(T)$, which is computable relative to T. The statement that $\vartheta(T)$ is well founded for any dilator T will still be equivalent to Π_1^1-comprehension. Thus, the latter is split into the computable transformation $T \mapsto \vartheta(T)$ and a statement about the preservation of well-foundedness, over a system of computable mathematics.

Keywords: Well-Ordering principles; Π_1^1-comprehension; dilators; Bachmann–Howard fixed points; reverse mathematics.

Mathematics Subject Classification 2010: 03B30, 03D60, 03F15

1. Introduction

We begin by recalling the abstract Bachmann–Howard principle, which was introduced in [4] (based on the author's PhD thesis [3] and an earlier arXiv preprint [2]). For this purpose, we consider the category of linear orders, with order embeddings as morphisms. The forgetful functor to the underlying set of an order will be left implicit. Conversely, we will often view a subset of an ordered set as a suborder. Given a set X, we put

$$[X]^{<\omega} := \text{“the set of finite subsets of } X\text{”}.$$

To obtain a functor we map $f : X \to Y$ to the function $[f]^{<\omega} : [X]^{<\omega} \to [Y]^{<\omega}$ with

$$[f]^{<\omega}(a) := \{f(s) \,|\, s \in a\}.$$

It is easy to see that $X \mapsto [X]^{<\omega}$ and $f \mapsto [f]^{<\omega}$ are primitive recursive set functions in the sense of Jensen and Karp [12] (with parameter ω). The same will hold for all class functions considered in the sequel. This allows us to formalize our investigation in primitive recursive set theory with infinity ($\mathbf{PRS}\omega$), as introduced by Rathjen [16]. Extending $\mathbf{PRS}\omega$ by axiom beta and the axiom of countability leads us to $\mathbf{ATR}_0^{\text{set}}$, the set-theoretic version of arithmetical transfinite recursion due to Simpson [22, 23]. A detailed introduction to these theories can be found in [3, Chap. 1]. The theory $\mathbf{PRS}\omega$ cannot quantify over all primitive recursive set functions. It can, however, quantify over a primitive recursive family of class functions, by quantifying over its set-sized parameters. A definition or proposition which speaks about a collection of class functions (e.g. about arbitrary endofunctors of linear orders) should be read as a schema: officially, we have a separate definition or proposition for each primitive recursive family of class functions. Such a family may depend on further sets as parameters (in particular the parameter ω is often required). We will see that the restriction to primitive recursive set functions does not affect the generality of our results. With these methodological remarks in mind we state the following definition, essentially due to Girard [7].

Definition 1.1 ($\mathbf{PRS}\omega$). A prae-dilator consists of

(i) an endofunctor T of linear orders and
(ii) a natural transformation $\operatorname{supp}^T : T \Rightarrow [\cdot]^{<\omega}$ that computes supports, in the following sense: For any linear order X and any element $\sigma \in T_X$ we have $\sigma \in \operatorname{rng}(T_{\iota_\sigma})$, where $\iota_\sigma : \operatorname{supp}_X^T(\sigma) \hookrightarrow X$ is the inclusion.

If T_X is well-founded for any well-order X, then $(T, \operatorname{supp}^T)$ is called a dilator.

Girard's original definition does not include the natural transformation supp^T but demands that T preserves direct limits and pullbacks. It is straightforward to check that the two formulations are equivalent (see [3, Remark 2.2.2]), but we find it very helpful to make the support functions explicit. Our prae-dilators are not quite equivalent to Girard's pre-dilators (hence the different spelling), since [7, Definition 4.4.1] contains an additional monotonicity condition. The latter is automatic in the well-founded case, so that it does not make a difference for dilators. According to Girard's definition the values of a dilator have to be ordinals. This has the advantage that isomorphic dilators become equal. Nevertheless, we want to allow arbitrary well-orders as values: it will be important to represent $T_X \cong \alpha$ by a well-order of rank below α (cf. Remark 2.2). To proceed, we introduce the following notation: if $(X, <_X)$ is a linear order (or just a preorder), then the preorder $<_X^{\text{fin}}$ on $[X]^{<\omega}$ is defined by

$$a <_X^{\text{fin}} b :\Leftrightarrow \text{``for any } s \in a \text{ there is a } t \in b \text{ with } s <_X t\text{''}.$$

In the case of singletons, we will write $s <_X^{\text{fin}} b$ and $a <_X^{\text{fin}} t$ rather than $\{s\} <_X^{\text{fin}} b$ and $a <_X^{\text{fin}} \{t\}$, respectively. The relation $\leq_X^{\text{fin}}$ is defined analogously. In [4], we have introduced the following notion (with precursors in [2, 3]):

Definition 1.2 (PRSω). Consider a prae-dilator (T, supp^T) and a linear order X. A function

$$\vartheta : T_X \to X$$

is called a Bachmann–Howard collapse if the following holds for all $\sigma, \tau \in T_X$:

(i) If we have $\sigma <_{T_X} \tau$ and $\mathrm{supp}^T_X(\sigma) <^{\mathrm{fin}}_X \vartheta(\tau)$, then we have $\vartheta(\sigma) <_X \vartheta(t)$.
(ii) We have $\mathrm{supp}^T_X(\sigma) <^{\mathrm{fin}}_X \vartheta(\sigma)$.

If such a function exists, then X is called a Bachmann–Howard fixed point of T.

In general Bachmann–Howard fixed points do not need to be well founded, but we are particularly interested in the case where they are.

Definition 1.3 (PRSω). The abstract Bachmann–Howard principle is the assertion that every dilator has a well-founded Bachmann–Howard fixed point.

Let us point out that the side condition $\mathrm{supp}^T_X(\sigma) <^{\mathrm{fin}}_X \vartheta(\tau)$ in the definition of a Bachmann–Howard collapse is crucial: It is possible that the order-type of T_X is bigger than the order-type of X, for any well-order X. In this case, the function $\vartheta : T_X \to X$ cannot be fully order preserving. The definition of Bachmann–Howard collapse is inspired by the construction of the Bachmann–Howard ordinal, in particular by the notation system due to Rathjen (see [20, Sec. 1]). Using a strong meta theory, it is standard to show that the abstract Bachmann–Howard principle is sound (e.g. the first uncountable cardinal is a Bachmann–Howard fixed point of any dilator with hereditarily countable parameters, see [4, Sec. 2]). On the other hand, the fact that a Bachmann–Howard collapse $\vartheta : T_X \to X$ is "almost" order preserving requires the existence of rather large ordinals, which suggests that the abstract Bachmann–Howard principle is strong. This is confirmed by the following result.

Theorem 1.1. *The following are equivalent over* $\mathbf{ATR}^{\mathrm{set}}_0$:

(i) *The principle of* Π^1_1*-comprehension.*
(ii) *The statement that every set is an element of some admissible set.*
(iii) *The abstract Bachmann–Howard principle.*

The equivalence between (i) and (ii) is shown in [11, Sec. 7] (see also [3, Sec. 1.4], where the base theory is lowered to $\mathbf{ATR}^{\mathrm{set}}_0$). The equivalence between (ii) and (iii) is established in [4], based on similar results in [2, 3]. Note that Girard, in the unpublished second part of his book on proof theory [9, Sec. 11.6], states a related equivalence but does not give a complete proof. The present author has not determined the relation between his Bachmann–Howard principle and Girard's functor Λ. It would certainly be very illuminating if one could establish a precise connection.

The aim of this paper is to resolve two shortcomings of Theorem 1.1. First, we will show that the equivalence between (i) and (iii) holds over the base theory $\mathbf{RCA_0}$. For this purpose, we formalize dilators in second-order arithmetic. This

was already done by Girard, but we find it worthwhile to give a detailed presentation in terms of support functions (cf. part (ii) of Definition 1.1). The formalization relies on Girard's result that a prae-dilator is determined by its restriction to the category of natural numbers (up to natural isomorphism). In fact, we will see that there is a single primitive recursive set function that reconstructs any prae-dilator from its set-sized restriction. This has two welcome side effects: it will enable us to express the abstract Bachmann–Howard principle by a single formula rather than a schema; and it means that the restriction to primitive recursive set functions is no loss of generality, as promised above. Once statement (iii) of Theorem 1.1 is expressed in second-order arithmetic it is immediate that the equivalence between (i) and (iii) holds over the second-order theory $\mathbf{ATR_0}$, over which $\mathbf{ATR_0^{set}}$ is conservative (due to Simpson [22, 23]). We then prove that the abstract Bachmann–Howard principle implies arithmetical transfinite recursion. It follows that the equivalence between (i) and (iii) holds over $\mathbf{RCA_0}$.

The second shortcoming of Theorem 1.1 results from our "abstract" formulation of the Bachmann–Howard principle: We have merely asserted the existence of a Bachmann–Howard fixed point, without specifying how a concrete fixed point might be constructed. Thus, it is not immediately clear whether the strength of the abstract Bachmann–Howard principle lies in the existence of a Bachmann–Howard fixed point or in the assertion that such a fixed point is well-founded. In [5] (similarly already in [3, Sec. 2.2]) we have shown that any prae-dilator T has a minimal Bachmann–Howard fixed point $\mathrm{BH}(T)$, which can be constructed by a primitive recursive set function: The idea is to define $\mathrm{BH}(T)$ as the direct limit of orders $X_0, X_1, \ldots$. By construction we have almost order preserving functions $\vartheta_{X_n} : T_{X_n} \to X_{n+1}$, which glue to the desired Bachmann–Howard collapse $\vartheta : T_{\mathrm{BH}(T)} \to \mathrm{BH}(T)$. The construction of $\mathrm{BH}(T)$ in [5] is important because it provides a clear semantical understanding of minimal Bachmann–Howard fixed points. However, it is not obvious how the direct limit $\mathrm{BH}(T)$ can be constructed in a weak base theory. In this paper, we show that $\mathrm{BH}(T)$ can be represented by a notation system $\vartheta(T)$. The latter is computable relative to a given prae-dilator T (or rather, relative to the restriction of T to the category of natural numbers), so that its construction can be formalized in $\mathbf{RCA_0}$. We can then state a computable Bachmann–Howard principle, which asserts that $\vartheta(T)$ is well-founded for any dilator T. Due to the minimality of $\vartheta(T)$ the computable Bachmann–Howard principle is equivalent to its abstract counterpart and thus to Π_1^1-comprehension.

Let us explain why the results of this paper are important: a type-one well-ordering principle is a (computable) transformation $X \mapsto T_X$ of linear orders, together with the assertion that T_X is well-founded for any well-order X. Observe that such an assertion is a Π_2^1-statement. The literature contains many equivalences between type-one well-ordering principles and natural Π_2^1-statements that are known from reverse mathematics (see [1, 6, 8, 10, 13, 18, 19, 21, 24, 25]). Rathjen [17, 18] and Montalbán [14, 15] had conjectured that Π_1^1-comprehension,

which is a Π^1_3-statement, has a similar characterization by a type-two well-ordering principle. Such a principle should transform each type-one well-ordering principle into a well-order (or into another type-one well-ordering principle, but the type of the codomain can be lowered by Currying). Theorem 1.1 makes a huge step toward that conjecture: the abstract Bachmann–Howard principle does certainly encapsulate a type-two well-ordering principle. It fails, however, to separate the well-ordering principle into a computable transformation and a statement about the preservation of well-foundedness. The computable Bachmann–Howard principle achieves this separation, so that we finally have a fully satisfactory solution of Rathjen and Montalbán's conjecture.

To conclude this introduction the author would like to point out that parts of this paper are based on Secs. 2.3 and 2.4 of his PhD thesis [3].

2. Dilators in Primitive Recursive Set Theory and Second-Order Arithmetic

Girard [7] has shown that dilators are determined by their restrictions to the category of natural numbers. This makes it possible to view them as set-sized objects and to represent them in second-order arithmetic. The aim of this section is to give a more explicit presentation of the constructions that are involved (the support functions from Definition 1.1 will turn out very useful). As a result, we will see that the abstract Bachmann–Howard principle can be expressed by a single formula (rather than a schema) in the language of second-order arithmetic (rather than set theory).

In the first half of this section, we work in primitive recursive set theory ($\mathbf{PRS}\omega$). Our goal is to define a primitive recursive set function that reconstructs any praedilator from its restriction to the category of natural numbers. The objects of this category are the natural numbers, each identified with its ordered set of predecessors. The morphisms are the order embeddings

$$n = \{0, \ldots, n-1\} \to \{0, \ldots, m-1\} = m.$$

Note that the resulting category is equivalent to the category of finite linear orders. We write $\mathrm{en}_a : |a| \to a$ for the isomorphism between a finite linear order a and its cardinality $|a|$. If $f : a \to b$ is an embedding of finite linear orders, then $|f| : |a| \to |b|$ denotes the unique order preserving function with

$$\mathrm{en}_b \circ |f| = f \circ \mathrm{en}_a .$$

Thus, $\mathrm{en}_{(\cdot)}$ is a natural isomorphism between the functor $|\cdot|$ and the identity. Let us also fix the notation

$$\iota_X^Y : X \hookrightarrow Y$$

for the inclusion of sets $X \subseteq Y$. We will show that prae-dilators are equivalent to the following set-sized objects:

Definition 2.1 (PRSω). A set-sized prae-dilator consists of

(i) a functor T from natural numbers to linear orders and
(ii) a natural transformation $\operatorname{supp}^T : T \Rightarrow [\cdot]^{<\omega}$ which computes supports, in the following sense: For any n and $\sigma \in T_n$ we have $\sigma \in \operatorname{rng}(T_{\iota_\sigma \circ \operatorname{en}_\sigma})$, with

$$\operatorname{en}_\sigma = \operatorname{en}_{\operatorname{supp}_n^T(\sigma)} : |\operatorname{supp}_n^T(\sigma)| \to \operatorname{supp}_n^T(\sigma),$$
$$\iota_\sigma = \iota_{\operatorname{supp}_n^T(\sigma)}^n : \operatorname{supp}_n^T(\sigma) \hookrightarrow \{0, \dots, n-1\}.$$

The definition of set-sized dilator will be given below, as it requires some preliminary work (it will not be enough to test the well-foundedness of T_n for all natural numbers). Note that we could not demand $\sigma \in \operatorname{rng}(T_{\iota_\sigma})$ in condition (ii) above, since $\operatorname{supp}_n^T(\sigma)$ may not be a natural number. If T is a prae-dilator in the sense of Definition 1.1 (in the following we will speak of class-sized (prae-)dilators), then $\sigma \in \operatorname{rng}(T_{\iota_\sigma})$ implies $\sigma \in \operatorname{rng}(T_{\iota_\sigma \circ \operatorname{en}_\sigma})$, because en_σ is an isomorphism. This gives one direction of our desired equivalence.

Lemma 2.1 (PRSω). *The restriction of a class-sized prae-dilator T to the category of natural numbers yields a set-sized prae-dilator $T \restriction \mathbb{N}$.*

For the other direction, we must reconstruct T from $T \restriction \mathbb{N}$. The idea is to use the pair $\langle a, \sigma \rangle$ with $a \in [X]^{<\omega}$ and $\sigma \in T_{|a|}$ to represent the element $T_{\iota_a^X \circ \operatorname{en}_a}(\sigma) \in T_X$. To get a unique representation we include the minimality condition $\operatorname{supp}_{|a|}^T(\sigma) = |a|$.

Definition 2.2 (PRSω). Let $(T, \operatorname{supp}^T)$ be a set-sized prae-dilator. For each linear order X we define a set D_X^T and a binary relation $<_{D_X^T}$ on D_X^T by

$$D_X^T := \{\langle a, \sigma \rangle \mid a \in [X]^{<\omega} \quad \text{and} \quad \sigma \in T_{|a|} \quad \text{and} \quad \operatorname{supp}_{|a|}^T(\sigma) = |a|\},$$
$$\langle a, \sigma \rangle <_{D_X^T} \langle b, \tau \rangle :\Leftrightarrow T_{|\iota_a^{a \cup b}|}(\sigma) <_{T_{|a \cup b|}} T_{|\iota_b^{a \cup b}|}(\tau).$$

If $f : X \to Y$ is an order embedding, then we put

$$D_f^T(\langle a, \sigma \rangle) := \langle [f]^{<\omega}(a), \sigma \rangle$$

to define a function $D_f^T : D_X^T \to D_Y^T$ (note $|[f]^{<\omega}(a)| = |a|$ to see $D_f^T(\langle a, \sigma \rangle) \in D_Y^T$). Finally, we define a family of functions $\operatorname{supp}_X^{D^T} : D_X^T \to [X]^{<\omega}$ by setting

$$\operatorname{supp}_X^{D^T}(\langle a, \sigma \rangle) := a$$

for each linear order X.

It is straightforward to check that the maps $(T, X) \mapsto (D_X^T, <_{D_X^T})$, $(T, f) \mapsto D_f^T$ and $(T, X) \mapsto \operatorname{supp}_X^{D^T}$ are primitive recursive set functions (see [3] for details). For the other direction of our equivalence we show the following lemma.

Lemma 2.2 (PRSω). *If $(T, \operatorname{supp}^T)$ is a set-sized prae-dilator, then $(D^T, \operatorname{supp}^{D^T})$ is a class-sized prae-dilator.*

Proof. In order to verify that $(D_X^T, <_{D_X^T})$ is a linear order one needs the implication

$$T_{|\iota_a^{a \cup b}|}(\sigma) = T_{|\iota_b^{a \cup b}|}(\tau) \Rightarrow \langle a, \sigma \rangle = \langle b, \tau \rangle.$$

The naturality of supp^T and the condition $\text{supp}_{|a|}^T(\sigma) = |a|$ imply

$$[\text{en}_{a \cup b}]^{<\omega} \circ \text{supp}_{|a \cup b|}^T \circ T_{|\iota_a^{a \cup b}|}(\sigma)$$

$$= [\text{en}_{a \cup b}]^{<\omega} \circ [|\iota_a^{a \cup b}|]^{<\omega} \circ \text{supp}_{|a|}^T(\sigma) = [\iota_a^{a \cup b} \circ \text{en}_a]^{<\omega}(|a|) = a.$$

As a is determined by $T_{|\iota_a^{a \cup b}|}(\sigma)$, the assumption $T_{|\iota_a^{a \cup b}|}(\sigma) = T_{|\iota_b^{a \cup b}|}(\tau)$ yields $a = b$. Then $|\iota_a^{a \cup b}| = |\iota_b^{a \cup b}|$ is the identity on $|a| = |a \cup b| = |b|$, and we also get $\sigma = \tau$. Based on this fact it is straightforward to verify that D^T is an endofunctor of linear orders and that $\text{supp}^{D^T} : D^T \Rightarrow [\cdot]^{<\omega}$ is a natural transformation. It remains to show that supp^{D^T} computes supports: Observe that $\langle a, \sigma \rangle \in D_X^T$ implies $\langle a, \sigma \rangle \in D_a^T$. Writing $\iota_{\langle a, \sigma \rangle} : \text{supp}_X^{D^T}(\langle a, \sigma \rangle) = a \hookrightarrow X$ for the inclusion we have

$$D_{\iota_{\langle a, \sigma \rangle}}^T(\langle a, \sigma \rangle) = \langle [\iota_{\langle a, \sigma \rangle}]^{<\omega}(a), \sigma \rangle = \langle a, \sigma \rangle,$$

which confirms that $\langle a, \sigma \rangle$ lies in the range of $D_{\iota_{\langle a, \sigma \rangle}}^T$. $\qquad\square$

Let us show that we have reconstructed the original dilator.

Proposition 2.1 (PRSω). *For any class-sized prae-dilator (T, supp^T) we can construct a natural equivalence $\eta^T : D^{T \upharpoonright \mathbb{N}} \Rightarrow T$ with $\text{supp}_X^T \circ \eta_X^T = \text{supp}_X^{D^{T \upharpoonright \mathbb{N}}}$.*

Proof. We make the above intuition official and set

$$\eta_X^T(\langle a, \sigma \rangle) := T_{\iota_a^X \circ \text{en}_a}(\sigma).$$

Concerning the formalization in **PRSω**, note that a primitive recursive definition of T is readily transformed into a primitive recursive definition of η^T. We verify that $\eta_X^T : D_X^{T \upharpoonright \mathbb{N}} \to T_X$ is order preserving: Assume that we have $\langle a, \sigma \rangle <_{D_X^{T \upharpoonright \mathbb{N}}} \langle b, \tau \rangle$ and thus $T_{|\iota_a^{a \cup b}|}(\sigma) <_{T_{|a \cup b|}} T_{|\iota_b^{a \cup b}|}(\tau)$. In view of

$$\iota_a^X \circ \text{en}_a = \iota_{a \cup b}^X \circ \iota_a^{a \cup b} \circ \text{en}_a = \iota_{a \cup b}^X \circ \text{en}_{a \cup b} \circ |\iota_a^{a \cup b}|$$

we obtain the desired inequality

$$\eta_X^T(\langle a, \sigma \rangle) = T_{\iota_a^X \circ \text{en}_a}(\sigma) = T_{\iota_{a \cup b}^X \circ \text{en}_{a \cup b}} \circ T_{|\iota_a^{a \cup b}|}(\sigma)$$

$$<_{T_X} T_{\iota_{a \cup b}^X \circ \text{en}_{a \cup b}} \circ T_{|\iota_b^{a \cup b}|}(\tau) = T_{\iota_b^X \circ \text{en}_b}(\tau) = \eta_X^T(\langle b, \tau \rangle).$$

To establish naturality we consider $f : X \to Y$ and observe

$$f \circ \iota_a^X \circ \text{en}_a = \iota_{[f]^{<\omega}(a)}^Y \circ (f \upharpoonright a) \circ \text{en}_a = \iota_{[f]^{<\omega}(a)}^Y \circ \text{en}_{[f]^{<\omega}(a)}.$$

This does indeed yield

$$T_f \circ \eta_X^T(\langle a, \sigma \rangle) = T_{f \circ \iota_a^X \circ \text{en}_a}(\sigma)$$

$$= T_{\iota_{[f]^{<\omega}(a)}^Y \circ \text{en}_{[f]^{<\omega}(a)}}(\sigma)$$

$$= \eta_Y^T(\langle [f]^{<\omega}(a), \sigma \rangle) = \eta_Y^T \circ D_f^{T \upharpoonright \mathbb{N}}(\langle a, \sigma \rangle).$$

Next, we show that the functions $\eta_X^T : D_X^{T\restriction\mathbb{N}} \to T_X$ are surjective: By the definition of prae-dilator any $\sigma \in T_X$ lies in the range of $T_{\iota_a^X}$, for $a := \mathrm{supp}_X^T(\sigma)$. Since the function $\mathrm{en}_a : |a| \to a$ is an isomorphism we obtain a $\sigma_0 \in T_{|a|}$ with $\sigma = T_{\iota_a^X \circ \mathrm{en}_a}(\sigma_0)$. To conclude $\sigma \in \mathrm{rng}(\eta_X^T)$ it remains to verify $\langle a, \sigma_0 \rangle \in D_X^{T\restriction\mathbb{N}}$. The crucial condition $\mathrm{supp}_{|a|}^{T\restriction\mathbb{N}}(\sigma_0) = |a|$ holds in view of

$$[\iota_a^X \circ \mathrm{en}_a]^{<\omega} \circ \mathrm{supp}_{|a|}^T(\sigma_0) = \mathrm{supp}_X^T \circ T_{\iota_a^X \circ \mathrm{en}_a}(\sigma_0) = \mathrm{supp}_X^T(\sigma) = a.$$

So far we have shown that η^T is a natural isomorphism. Finally, for $\langle a, \sigma \rangle \in D_X^{T\restriction\mathbb{N}}$ the condition $\mathrm{supp}_{|a|}^{T\restriction\mathbb{N}}(\sigma) = |a|$ implies

$$\begin{aligned}
\mathrm{supp}_X^T \circ \eta_X^T(\langle a, \sigma \rangle) &= \mathrm{supp}_X^T \circ T_{\iota_a^X \circ \mathrm{en}_a}(\sigma) \\
&= [\iota_a^X \circ \mathrm{en}_a]^{<\omega} \circ \mathrm{supp}_{|a|}^T(\sigma) \\
&= [\iota_a^X \circ \mathrm{en}_a]^{<\omega}(|a|) = a = \mathrm{supp}_X^{D^{T\restriction\mathbb{N}}}(\langle a, \sigma \rangle),
\end{aligned}$$

as was promised in the proposition. $\qquad\square$

Later, we will also need the following lemma.

Lemma 2.3 (PRSω). *Consider set-sized prae-dilators S and T. Given a natural equivalence $\eta^0 : S \Rightarrow T$ with $\mathrm{supp}_n^T \circ \eta_n^0 = \mathrm{supp}_n^S$, we can construct a natural equivalence $\eta : D^S \Rightarrow D^T$ with $\mathrm{supp}_X^{D^T} \circ \eta_X = \mathrm{supp}_X^{D^S}$.*

Proof. It is straightforward to verify that

$$\eta_X(\langle a, \sigma \rangle) := \langle a, \eta_{|a|}^0(\sigma) \rangle$$

defines the desired family of functions. $\qquad\square$

Whether a given set is a set-sized prae-dilator can be decided by a primitive recursive set function. In contrast, the notion of dilator retains its logical complexity.

Definition 2.3 (PRSω). A set-sized prae-dilator (T, supp^T) is called a set-sized dilator if the order $(D_X^T, <_{D_X^T})$ is well-founded for any well-order X.

Proposition 2.1 does, in particular, tell us that $D_X^{T\restriction\mathbb{N}}$ is well-founded if T_X is. Together with Lemma 2.1 we get the following corollary.

Corollary 2.1 (PRSω). *If T is a class-sized dilator, then its restriction $T \restriction \mathbb{N}$ is a set-sized dilator.*

The converse is trivial, based on the corresponding result for prae-dilators:

Corollary 2.2 (PRSω). *If (T, supp^T) is a set-sized dilator, then $(D^T, \mathrm{supp}^{D^T})$ is a class-sized dilator.*

The previous results show that (prae-)dilators are essentially set-sized objects. As promised, this allows us to express the abstract Bachmann–Howard principle by

a single formula (recall that Definition 1.3 is a schema, because we cannot quantify over all class-sized dilators).

Proposition 2.2 (PRSω). *The following consequence of the abstract Bachmann–Howard principle implies each of its instances*: *For every set-sized dilator T there is a well-order X with a Bachmann–Howard collapse $\vartheta : D_X^T \to X$.*

Proof. Let us first establish that the given statement follows from the Bachmann–Howard principle: If T is a set-sized dilator, then D^T is a class-sized dilator. Using the abstract Bachmann–Howard principle (for the primitive recursive family of functions $X \mapsto D_X^T$ with parameter T) we obtain the desired collapse $\vartheta : D_X^T \to X$ for a well-order X. Conversely, we deduce an arbitrary instance of the abstract Bachmann–Howard principle: If T is a class-sized dilator, then $T \upharpoonright \mathbb{N}$ is a set-sized dilator. Thus, the statement from the proposition yields a well-order X with a Bachmann–Howard collapse $\vartheta : D_X^{T \upharpoonright \mathbb{N}} \to X$. Proposition 2.1 provides an isomorphism $\eta_X^T : D_X^{T \upharpoonright \mathbb{N}} \to T_X$ with $\operatorname{supp}_X^T \circ \eta_X^T = \operatorname{supp}_X^{D^{T \upharpoonright \mathbb{N}}}$. One can check that

$$\vartheta \circ (\eta_X^T)^{-1} : T_X \to X$$

is a Bachmann–Howard collapse as well. □

As promised, we can also deduce that the restriction to primitive recursive set functions does not mean a loss of generality.

Remark 2.1. If one works in a stronger theory, then one may want to consider a dilator T that is not given by a primitive recursive set function. The corresponding dilator $D^{T \upharpoonright \mathbb{N}}$ will still be primitive recursive (even though stronger separation axioms may be needed to show that the parameter $T \upharpoonright \mathbb{N}$ exists as a set). Furthermore, the dilators T and $D^{T \upharpoonright \mathbb{N}}$ will still be equivalent (even though the equivalence $\eta^T : D^{T \upharpoonright \mathbb{N}} \Rightarrow T$ may no longer be primitive recursive). As in Proposition 2.2, a Bachmann–Howard collapse for $D^{T \upharpoonright \mathbb{N}}$ can be transformed into a Bachmann–Howard collapse for T. This shows that the abstract Bachmann–Howard principle does not become stronger if we admit dilators which are not primitive recursive.

While any dilator is equivalent to a primitive recursive one, the statement that "all dilators are primitive recursive set functions" can also be very misleading.

Remark 2.2. For each dilator T we can consider the corresponding function

$$\alpha \mapsto \operatorname{otyp}(T_\alpha)$$

of order types. It is important to realize that this function is not primitive recursive in general. By induction on the primitive recursive definition of F we find a number n such that $\vec{x} \in \mathbb{V}_\beta$ implies $F(\vec{x}) \in \mathbb{V}_{\varphi_n(\beta)}$. Thus $\alpha \mapsto \varphi_\alpha(0)$ cannot be a primitive recursive set function. On the other hand, one can construct primitive recursive notation systems T_X such that $\operatorname{otyp}(X) = \alpha$ implies $\operatorname{otyp}(T_X) = \varphi_\alpha 0$ (see [13, 21]).

To explain this phenomenon we recall that the order type of a given well-order cannot be computed by a primitive recursive set function (axiom beta is not provable in **PRS**ω and not even in Kripke–Platek set theory). For this reason it is important to admit arbitrary well-orders as values of dilators. If one only allowed ordinals as values (as in the formulation of Girard [7, Definition 2.3.1]), then it would not be true that any dilator is equivalent to a primitive recursive one.

In the second half of this section, we show that (prae-)dilators can be formalized in the subsystem **RCA$_0$** of second-order arithmetic. This is due to Girard [7], but we know of no explicit presentation. It is well known that finite sets of natural numbers and functions between such sets can be coded by natural numbers. Basic relations and operations on the codes are primitive recursive (in the usual number-theoretic sense). This allows us to express the following by an arithmetical formula (with parameters $T^0, T^1, \mathrm{supp}^T \subseteq \mathbb{N}$).

Definition 2.4 (RCA$_0$). A coded prae-dilator consists of

(i) a functor T from natural numbers to linear orders with fields $T_n \subseteq \mathbb{N}$, represented by the sets

$$T^0 = \{\langle 0, n, \sigma \rangle \,|\, \sigma \in T_n\} \cup \{\langle 1, n, \sigma, \tau \rangle \,|\, \sigma <_{T_n} \tau\},$$
$$T^1 = \{\langle f, \sigma, \tau \rangle \,|\, T_f(\sigma) = \tau\},$$

(ii) a natural transformation $\mathrm{supp}^T : T \Rightarrow [\cdot]^{<\omega}$ that computes supports (in the sense of Definition 2.1), represented by the set

$$\mathrm{supp}^T = \{\langle n, \sigma, a \rangle \,|\, \mathrm{supp}_n^T(\sigma) = a\}.$$

When we work in **RCA$_0$** we can only refer to the sets $T^0, T^1, \mathrm{supp}^T$. In this context, we use $\sigma \in T_n$ as an abbreviation for the Δ_1^0-formula $\langle 0, n, \sigma \rangle \in T^0$. The same applies to the expressions $\sigma <_{T_n} \tau$, $T_f(\sigma) = \tau$ and $\mathrm{supp}_n^T(\sigma) = a$. Formulas of second-order arithmetic have a natural translation into the language of set theory (cf. [23, Theorem VII.3.9]). If T^0 and $n \mapsto T_n$ are related as in the definition, then the set-theoretic translation of the second-order formula $\sigma \in T_n \equiv \langle 0, n, \sigma \rangle \in T^0$ is equivalent to the set-theoretic formula $\sigma \in T_n$ (even though the two formulas are not literally equal). When we work in **PRS**ω we may thus identify coded prae-dilators and those set-sized prae-dilators (T, supp^T) with the property that the field of each linear order T_n is a subset of the natural numbers.

Our next goal is to single out the coded dilators in **RCA$_0$**. For this purpose, we reconstruct the orders $(D_X^T, <_{D_X^T})$ from Definition 2.2, for each order $(X, <_X)$ with field $X \subseteq \mathbb{N}$: The set

$$D_X^T = \{\langle a, \sigma \rangle \,|\, \text{``}a \text{ codes a finite subset of } X\text{''} \land \sigma \in T_{|a|} \land \mathrm{supp}_{|a|}^T(\sigma) = |a|\} \subseteq \mathbb{N}$$

exists by Δ_1^0-comprehension. Recall the functions $|\iota_c^d| : |c| \to |d|$ for finite subsets (in fact suborders) c, d of X. It is straightforward to see that the operation $(c, d) \mapsto |\iota_c^d|$

on the codes is primitive recursive relative to $(X, <_X)$. In view of

$$\langle a, \sigma \rangle <_{D_X^T} \langle b, \tau \rangle \equiv \exists_{\sigma', \tau'} (T_{|\iota_a^{a \cup b}|}(\sigma) = \sigma' \wedge T_{|\iota_b^{a \cup b}|}(\tau) = \tau' \wedge \sigma' <_{T_{|a \cup b|}} \tau')$$

$$\equiv \forall_{\sigma', \tau'} (T_{|\iota_a^{a \cup b}|}(\sigma) = \sigma' \wedge T_{|\iota_b^{a \cup b}|}(\tau) = \tau' \rightarrow \sigma' <_{T_{|a \cup b|}} \tau')$$

the relation $<_{D_X^T}$ can be defined by Δ_1^0-comprehension as well. As in the proof of Lemma 2.2 one can show the following lemma.

Lemma 2.4 (RCA$_0$). *If T is a coded prae-dilator and $(X, <_X)$ is a linear order, then $(D_X^T, <_{D_X^T})$ is a linear order as well.*

Coded dilators can now be defined by a Π_2^1-formula of second-order arithmetic.

Definition 2.5 (RCA$_0$). A coded prae-dilator T is a coded dilator if $(D_X^T, <_{D_X^T})$ is well-founded for every well-order $(X, <_X)$ with field $X \subseteq \mathbb{N}$.

There is one subtlety when we translate back into set theory: Definition 2.5 does only probe well-orders $(X, <_X)$ with $X \subseteq \mathbb{N}$. In a set theory with choice this does not make a difference: using the definition of well-foundedness in terms of descending sequences, one can show that D_Y^T is well-founded if $D_{Y_0}^T$ is well-founded for all countable suborders $Y_0 \subseteq Y$ (due to Girard [7, Theorem 2.1.15]). All countable orders are covered by Definition 2.5, since Lemma 2.2 tells us that $Y_0 \cong X \subseteq \mathbb{N}$ implies $D_{Y_0}^T \cong D_X^T$. In the base theory $\mathbf{ATR_0^{set}}$ of Theorem 1.1 the issue disappears for a rather different reason: This theory includes the axiom of countability, which implies that any set is in bijection with a subset of the natural numbers (axiom beta, which is also included in $\mathbf{ATR_0^{set}}$, is not needed here).

Back in $\mathbf{RCA_0}$, a Bachmann–Howard collapse $\vartheta : D_X^T \rightarrow X$ for an order $X \subseteq \mathbb{N}$ can be represented by a set $\vartheta \subseteq \mathbb{N}$. The conditions from Definition 1.2 are readily expressed by an arithmetical formula (with the appropriate set parameters). Thus, the following amounts to a Π_3^1-statement of second-order arithmetic:

Definition 2.6 (RCA$_0$). The second-order version of the abstract Bachmann–Howard principle is the following statement: For every coded dilator T there is a well-founded Bachmann–Howard fixed point X, i.e. a well-order $X \subseteq \mathbb{N}$ with a Bachmann–Howard collapse $\vartheta : D_X^T \rightarrow X$.

We will speak of the set-theoretic version of the abstract Bachmann–Howard principle in order to refer to Definition 1.3. In an appropriate meta theory one can construct prae-dilators with uncountable parameters that only have uncountable Bachmann–Howard fixed points: It is straightforward to show that any Bachmann–Howard collapse must be injective (cf. [3, Lemma 2.1.7]). Now, consider the constant prae-dilator $T_X = Y$ for an uncountable order Y. On the other hand, the axiom of countability ensures the following lemma.

Lemma 2.5 (ATR$_0^{set}$). *The second-order version of the abstract Bachmann–Howard principle is equivalent to the set-theoretic version.*

Proof. To deduce the second-order version from the set-theoretic version we consider a coded dilator T. In the presence of countability we may view T as a set-sized dilator, as discussed above. Then, the set-theoretic version of the Bachmann–Howard principle yields a Bachmann–Howard collapse $\vartheta : D_X^T \to X$ for some well-order X. The axiom of countability yields a well-order $Y \cong X$ with field $Y \subseteq \mathbb{N}$. It is straightforward to transform ϑ into a Bachmann–Howard collapse of D_Y^T into Y, as demanded by the second-order version of the abstract Bachmann–Howard principle. To deduce the set-theoretic version from the second-order version we invoke Proposition 2.2 and consider a set-sized dilator T. The axiom of countability yields a bijection between $\bigcup\{T_n \mid n \in \mathbb{N}\}$ and a subset of the natural numbers. This allows us to construct a coded prae-dilator S and a natural equivalence $\eta^0 : S \Rightarrow T$ with $\operatorname{supp}_n^T \circ \eta_n^0 = \operatorname{supp}_n^S$ (cf. the proof of [3, Proposition 2.3.21]). By Lemma 2.3, we get a natural equivalence $\eta : D^S \Rightarrow D^T$ with $\operatorname{supp}_X^{D^T} \circ \eta_X = \operatorname{supp}_X^{D^S}$. In particular $D_X^S \cong D_X^T$ ensures that S is a coded dilator. Now, the second-order version of the abstract Bachmann–Howard principle yields a well-order X with a Bachmann–Howard collapse $\vartheta : D_X^S \to X$. We can use η to transform ϑ into the required Bachmann–Howard collapse of D_X^T into X, as in the proof of Proposition 2.2. $\qquad\square$

In view of the previous result, Theorem 1.1 implies that Π_1^1-comprehension is equivalent to the second-order version of the abstract Bachmann–Howard principle, over the base theory $\mathbf{ATR_0^{set}}$. Now that we have an equivalence between second-order statements we can immediately conclude that it holds over the second-order theory $\mathbf{ATR_0}$, due to the conservativity result of Simpson [22, 23]. In Sec. 3 we will show that the base theory can be lowered to $\mathbf{RCA_0}$.

3. Bootstrapping the Bachmann–Howard Principle

Via a series of intermediate steps we show that (the second-order version of) the abstract Bachmann–Howard principle implies arithmetical transfinite recursion. This will immediately allow us to lower the base theory in Theorem 1.1.

To initiate our bootstrapping process, we prove that the abstract Bachmann–Howard principle makes ordinal exponentiation available (on the level of notation systems). By a result of Girard [8, Sec. II.5] (see also the proof by Hirst [10]) this brings us up to $\mathbf{ACA_0}$. Writing $X^{<\omega}$ for the set of finite sequences with entries in a set $X \subseteq \mathbb{N}$ (coded by natural numbers), we consider the following structure.

Definition 3.1 ($\mathbf{RCA_0}$). For each linear order $(X, <_X)$ the set $\omega^X \subseteq X^{<\omega}$ and the relation $<_{\omega^X} \subseteq \omega^X \times \omega^X$ are given by

$$\langle x_0, \ldots, x_{n-1} \rangle \in \omega^X \Leftrightarrow x_{n-1} \leq_X \cdots \leq_X x_0,$$

$$\langle x_0, \ldots, x_{n-1} \rangle <_{\omega^X} \langle x_0', \ldots, x_{m-1}' \rangle \Leftrightarrow \begin{cases} \text{either } n < m \quad \text{and} \quad x_i = x_i' \text{ for } i < n, \\ \text{or there is } j < \min\{n, m\} \text{ with} \\ x_j <_X x_j' \quad \text{and} \quad x_i = x_i' \text{ for } i < j. \end{cases}$$

On an informal level, assume that X is isomorphic to an ordinal α. If the elements $x_{n-1} \leq_X \cdots \leq_X x_0$ correspond to $\alpha_{n-1} \leq \cdots \leq \alpha_0 < \alpha$, then $\langle x_0, \ldots, x_{n-1} \rangle \in \omega^X$ represents the ordinal $\omega^{\alpha_0} + \cdots + \omega^{\alpha_{n-1}} < \omega^\alpha$ in Cantor normal form. Thus ω^X is a notation system for the ordinal ω^α. The following is standard.

Lemma 3.1 (RCA$_0$). *If $(X, <_X)$ is a linear order, then so is $(\omega^X, <_{\omega^X})$.*

The aforementioned result of Girard and Hirst implies that **RCA$_0$** cannot show the well-foundedness of ω^X. In the following, we refer to the second-order version of the abstract Bachmann–Howard principle, as formulated in Definition 2.6.

Proposition 3.1 (RCA$_0$). *The abstract Bachmann–Howard principle implies that ω^X is well-founded for any well-order X.*

Proof. For a fixed well-order X we consider the dilator (T, supp^T) given by

$$T_Y = X \times (\{\perp\} \cup Y),$$

$$(x, y) <_{T_Y} (x', y') \Leftrightarrow \begin{cases} \text{either } x <_X x', \\ \text{or } x = x' \quad \text{and} \quad y <_Y y' \text{ (with } \perp <_Y y' \text{ for any } y' \in Y), \end{cases}$$

$$T_f(x, y) = \begin{cases} (x, f(y)) & \text{if } y \in Y \text{ (where } f : Y \to Y' \text{ is an embedding)}, \\ (x, \perp) & \text{if } y = \perp, \end{cases}$$

$$\text{supp}_Y^T(x, y) = \begin{cases} \{y\} & \text{if } y \in Y, \\ \emptyset & \text{if } y = \perp. \end{cases}$$

Officially, we must work with the representations T^0, T^1 and supp^T from Definition 2.4: it is straightforward to see that they exist as sets (by Δ_1^0-comprehension) and that they represent a coded prae-dilator. Instead of T_Y we must consider

$$D_Y^T = \{\langle \emptyset, (x, \perp) \rangle \mid x \in X\} \cup \{\langle \{y\}, (x, 0) \rangle \mid x \in X, y \in Y\}.$$

Note that $(x, \perp) \in T_{|\emptyset|}$ satisfies $\text{supp}_{|\emptyset|}^T(x, \perp) = \emptyset = |\emptyset|$ while $(x, 0) \in T_{|\{y\}|}$ satisfies $\text{supp}_{|\{y\}|}^T(x, 0) = \{0\} = 1 = |\{y\}|$, as demanded by the definition of D_Y^T in Sec. 2. Since the interpretation

$$D_Y^T \ni \langle \emptyset, (x, \perp) \rangle \mapsto (x, \perp) \in T_Y, \quad D_Y^T \ni \langle \{y\}, (x, 0) \rangle \mapsto (x, y) \in T_Y$$

is an isomorphism (with respect to the order $<_{D_Y^T}$ from Sec. 2), we may work with T_Y rather than D_Y^T after all. To see that T is a coded dilator we must show that $D_Y^T \cong T_Y$ is well-founded for any well-order Y. Aiming at a contradiction, assume that $(x_n, y_n)_{n \in \mathbb{N}}$ is a strictly decreasing sequence in T_Y (the two obvious definitions of well-foundedness are equivalent over **RCA$_0$**, see e.g. [3, Lemma 2.3.12]). If the sequence $(x_n)_{n \in \mathbb{N}}$ does not become constant, then we can transform it into a strictly decreasing sequence in X. If we have $x_n = x_N$ for all $n \geq N$, then $(y_n)_{n \geq N}$ is a strictly increasing sequence in $\{\perp\} \cup Y$. Both possibilities contradict the assumption

that X and Y are well-founded. Since T is a coded dilator the abstract Bachmann–Howard principle yields a well-order Y with a Bachmann–Howard collapse

$$\vartheta : T_Y \cong D_Y^T \to Y.$$

To deduce that ω^X is well-founded, we show that the function $f : \omega^X \to \{\bot\} \cup Y$ with recursive clauses

$$f(\langle\rangle) := \bot,$$

$$f(\langle x_0, \ldots, x_n \rangle) := \vartheta(x_0, f(\langle x_1, \ldots, x_n \rangle))$$

is order-preserving (it is worth observing that ϑ cannot be fully order-preserving: if X and Y have order-type $\alpha > 1$ and β, then T_Y has order-type $(1 + \beta) \cdot \alpha > \beta$). So assume that we have

$$\langle x_0, \ldots, x_{n-1} \rangle <_{\omega^X} \langle x_0', \ldots, x_{m-1}' \rangle.$$

To see that we can cancel equal entries at the beginning of the sequences it suffices to observe that $y <_Y y'$ (possibly with $y = \bot$) implies $\vartheta(x, y) <_Y \vartheta(x, y')$: By the definition of Bachmann–Howard collapse we have $\{y'\} = \operatorname{supp}_Y^T(x, y') <_Y^{\mathrm{fin}} \vartheta(x, y')$, in other words $y' <_Y \vartheta(x, y')$. This implies $\operatorname{supp}_Y^T(x, y) <_Y^{\mathrm{fin}} \vartheta(x, y')$. Together with $(x, y) <_{T_Y} (x, y')$ we get $\vartheta(x, y) <_Y \vartheta(x, y')$, again by the definition of Bachmann–Howard collapse. It remains to consider the cases $0 = n < m$ and $x_0 <_X x_0'$. In the first case, we observe

$$f(\langle x_0, \ldots, x_{n-1} \rangle) = \bot <_Y f(\langle x_0', \ldots, x_{m-1}' \rangle) \in Y.$$

In case $x_0 <_X x_0'$, we set $y := f(\langle x_1', \ldots, x_{m-1}' \rangle)$ and prove

$$f(\langle x_{n-i}, \ldots, x_{n-1} \rangle) <_Y \vartheta(x_0', y) = f(\langle x_0', \ldots, x_{m-1}' \rangle)$$

by induction on $i \leq n$: For $i = 0$, we have $f(\langle x_{n-i}, \ldots, x_{n-1} \rangle) = \bot$ and the claim follows as before. In the step the induction hypothesis provides

$$\operatorname{supp}_Y^T(x_{n-(i+1)}, f(\langle x_{n-i}, \ldots, x_{n-1} \rangle)) \subseteq \{f(\langle x_{n-i}, \ldots, x_{n-1} \rangle)\} <_Y^{\mathrm{fin}} \vartheta(x_0', y).$$

Invoking the definition of ω^X we have $x_{n-(i+1)} \leq_X x_0 <_X x_0'$ and thus

$$(x_{n-(i+1)}, f(\langle x_{n-i}, \ldots, x_{n-1} \rangle)) <_{T_Y} (x_0', y).$$

By the definition of Bachmann–Howard collapse this implies

$$f(\langle x_{n-(i+1)}, \ldots, x_{n-1} \rangle) = \vartheta(x_{n-(i+1)}, f(\langle x_{n-i}, \ldots, x_{n-1} \rangle)) <_Y \vartheta(x_0', y),$$

which completes the induction step. $\qquad\square$

Recall that ε_α denotes the αth ordinal with $\omega^\gamma = \gamma$. By another application of the abstract Bachmann–Howard principle we want to establish the well-foundedness of these ε-numbers. Using a result of Marcone and Montalbán [13] (see also the proof by Afshari and Rathjen [1]) this will secure arithmetical recursion along the natural numbers, the defining principle of the theory $\mathbf{ACA_0^+}$. First, we define a notation

system ε_X for ε_α, relative to a notation system X for α. The ε-numbers below ε_α are represented by terms ε_x with $x \in X$. The gaps are filled with terms of the form $\omega^{t_0} + \cdots + \omega^{t_n}$, which correspond to ordinals in Cantor normal form.

Definition 3.2 ($\mathbf{RCA_0}$). For each linear order $(X, <_X)$ the set ε_X and the relation $<_{\varepsilon_X} \subseteq \varepsilon_X \times \varepsilon_X$ are defined by the following simultaneous recursion:

(i) The term 0 is an element of ε_X.
(ii) For each $x \in X$ the term ε_x is an element of ε_X.
(iii) If $t_0, \ldots, t_n$ are terms in ε_X, then so is $\omega^{t_0} + \cdots + \omega^{t_n}$, provided that

- either $n = 0$ and t_0 is not of the form ε_x,
- or $n > 0$ and $t_n \leq_{\varepsilon_X} \cdots \leq_{\varepsilon_X} t_0$ (note that $s \leq_{\varepsilon_X} t$ abbreviates the disjunction $s <_{\varepsilon_X} t \vee s = t$, where the second disjunct refers to equality as terms).

For $t, t' \in \varepsilon_X$ we have $t <_{\varepsilon_X} t'$ precisely if one of the following holds:

(i') We have $t = 0$ and $t' \neq 0$.
(ii') We have $t = \varepsilon_x$ and

- either $t' = \varepsilon_{x'}$ with $x <_X x'$,
- or $t' = \omega^{t'_0} + \cdots + \omega^{t'_m}$ with $t \leq_{\varepsilon_X} t'_0$.

(iii') We have $t = \omega^{t_0} + \cdots + \omega^{t_n}$ and

- either $t' = \varepsilon_x$ with $t_0 <_{\varepsilon_X} t'$,
- or $t' = \omega^{t'_0} + \cdots + \omega^{t'_m}$ such that one of the following holds:
  — Either we have $n < m$ and $t_i = t'_i$ for all $i \leq n$,
  — or there is $j \leq \min\{n, m\}$ with $t_j <_{\varepsilon_X} t'_j$ and $t_i = t'_i$ for $i < j$.

To formalize the definition in $\mathbf{RCA_0}$ one first defines a term system $\varepsilon_X^0 \supseteq \varepsilon_X$ by ignoring the condition $t_n \leq_{\varepsilon_X} \cdots \leq_{\varepsilon_X} t_0$ in clause (iii). Writing $\ulcorner t \urcorner$ for the Gödel number of the term $t \in \varepsilon_X^0$, the length function $L_{\varepsilon_X} : \varepsilon_X^0 \to \mathbb{N}$ is given as

$$L_{\varepsilon_X}(t) := \begin{cases} \ulcorner t \urcorner & \text{if } t = 0 \text{ or } t = \varepsilon_x, \\ \max\{\ulcorner t \urcorner, L_{\varepsilon_X}(t_0) + \cdots + L_{\varepsilon_X}(t_n) + 1\} & \text{if } t = \omega^{t_0} + \cdots + \omega^{t_n}. \end{cases}$$

The Gödel numbers are included to ensure that $\forall_t (L_{\varepsilon_X}(t) \leq n \to \cdots)$ is a bounded quantifier. For $s, t, t' \in \varepsilon_X^0$ one can now decide $s \in \varepsilon_X$ and $t <_{\varepsilon_X} t'$ by simultaneous recursion on $L_{\varepsilon_X}(s)$ and $L_{\varepsilon_X}(t) + L_{\varepsilon_X}(t')$. It is standard to show the following lemma.

Lemma 3.2 ($\mathbf{RCA_0}$). *If $(X, <_X)$ is a linear order, then so is $(\varepsilon_X, <_{\varepsilon_X})$.*

The crucial point is, once again, the preservation of well-foundedness.

Proposition 3.2 ($\mathbf{RCA_0}$). *The abstract Bachmann–Howard principle implies that ε_X is well-founded for any well-order X.*

Proof. For a fixed well-order X we consider the dilator (T, supp^T) with

$$T_Y = (\{\bot\} \cup X) \times \omega^Y,$$

$$(x, \sigma) <_{T_Y} (x', \sigma') \Leftrightarrow \begin{cases} \text{either } x <_X x' \quad \text{(with } \bot <_X x' \text{ for any } x' \in X\text{)}, \\ \text{or } x = x' \text{ and } \sigma <_{\omega^Y} \sigma', \end{cases}$$

$$T_f(x, \langle y_0, \ldots, y_{n-1} \rangle) = (x, \langle f(y_0), \ldots, f(y_{n-1}) \rangle) \quad \text{(with } f : Y \to Y'\text{)},$$

$$\mathrm{supp}^T_Y(x, \langle y_0, \ldots, y_{n-1} \rangle) = \{y_0, \ldots, y_{n-1}\}.$$

As in the proof of Proposition 3.1 one must officially work with the representations from Definition 2.4: it is straightforward to see that they exist as sets and that they represent a coded prae-dilator. By Proposition 3.1, the abstract Bachmann–Howard principle ensures that ω^Y is well-founded for any well-order Y. We conclude that T_Y is well-founded, so that (T, supp^T) is indeed a dilator. Another application of the abstract Bachmann–Howard principle yields a well-order Y with a Bachmann–Howard collapse

$$\vartheta : T_Y \to Y.$$

To each term $t \in \varepsilon_X$ we associate its "ε-degree" $t^* \in \{\bot\} \cup X$ by the recursion

$$0^* = \bot, \quad \varepsilon_x^* = x, \quad (\omega^{t_0} + \cdots + \omega^{t_n})^* = t_0^*.$$

In order to conclude we show that the function $f : \varepsilon_X \to Y$ with

$$f(0) := \vartheta(\bot, \langle \rangle),$$

$$f(\varepsilon_x) := \vartheta(x, \langle \rangle),$$

$$f(\omega^{t_0} + \cdots + \omega^{t_n}) := \begin{cases} \vartheta(t_0^*, \langle f(t_0), \ldots, f(t_n) \rangle) & \text{if } f(t_n) \leq_Y \cdots \leq_Y f(t_0), \\ \vartheta(\bot, \langle \rangle) & \text{otherwise} \end{cases}$$

is order preserving (thus the alternative in the case distinction will never apply). Let us argue by induction on $L_{\varepsilon_X}(s) + L_{\varepsilon_X}(t)$ to establish the implication

$$s <_{\varepsilon_X} t \Rightarrow f(s) <_Y f(t).$$

As a first interesting case we assume that $s = \omega^{s_0} + \cdots + \omega^{s_n} <_{\varepsilon_X} \varepsilon_x = t$ holds because of $s_0 <_{\varepsilon_X} t$. By an auxiliary induction on s_0 we see $s_0^* <_X x = t^*$ (note that we could only infer $s_0^* \leq_X t^*$ if t was not of the form ε_x). The induction hypothesis yields $f(s_n) \leq_Y \cdots \leq_Y f(s_0) <_Y f(t) = \vartheta(x, \langle \rangle)$. Thus, we have

$$(s_0^*, \langle f(s_0), \ldots, f(s_n) \rangle) <_{T_Y} (x, \langle \rangle),$$

$$\mathrm{supp}^T_Y(s_0^*, \langle f(s_0), \ldots, f(s_n) \rangle) = \{f(s_0), \ldots, f(s_n)\} <^{\mathrm{fin}}_Y \vartheta(x, \langle \rangle).$$

By the definition of Bachmann–Howard collapse we get

$$f(s) = \vartheta(s_0^*, \langle f(s_0), \ldots, f(s_n) \rangle) <_Y \vartheta(x, \langle \rangle) = f(t),$$

as desired. Let us also consider $s = \omega^{s_0} + \cdots + \omega^{s_n} <_{\varepsilon_X} \omega^{t_0} + \cdots + \omega^{t_m} = t$. The induction hypothesis yields $f(s_n) \leq_Y \cdots \leq_Y f(s_0)$ and $f(t_m) \leq_Y \cdots \leq_Y f(t_0)$, as

well as $\langle f(s_0), \ldots, f(s_n) \rangle <_{\omega^Y} \langle f(t_0), \ldots, f(t_m) \rangle$. In view of $s_0 \leq_{\varepsilon_X} t_0$ we also have $s_0^* \leq_X t_0^*$ and thus

$$(s_0^*, \langle f(s_0), \ldots, f(s_n) \rangle) <_{T_Y} (t_0^*, \langle f(t_0), \ldots, f(t_m) \rangle).$$

Furthermore, we can observe

$$\mathrm{supp}_Y^T(s_0^*, \langle f(s_0), \ldots, f(s_n) \rangle)$$
$$\leq_Y^{\mathrm{fin}} \mathrm{supp}_Y^T(t_0^*, \langle f(t_0), \ldots, f(t_m) \rangle) <_Y^{\mathrm{fin}} \vartheta(t_0^*, \langle f(t_0), \ldots, f(t_m) \rangle),$$

where the second inequality relies on the definition of Bachmann–Howard collapse. Also by the latter we finally get

$$f(s) = \vartheta(s_0^*, \langle f(s_0), \ldots, f(s_n) \rangle) <_Y \vartheta(t_0^*, \langle f(t_0), \ldots, f(t_m) \rangle) = f(t).$$

The remaining cases are straightforward. Thus $f : \varepsilon_X \to Y$ is an order embedding and ε_X is well-founded. $\square$

As the final step of our bootstrapping process we use the abstract Bachmann–Howard principle to show that a certain notation system ϑ_X is well-founded for any well-order X. The latter implies that every set is contained in a (countable coded) ω-model of bar induction, as shown by Rathjen and Valencia Vizcaíno [19]. We thus reach a statement which is stronger than the base theory **ATR₀** of Theorem 1.1. The following coincides with [19, Definition 2.6].

Definition 3.3. For each linear order $(X, <_X)$ the set ϑ_X of terms, the relation $<_{\vartheta_X} \subseteq \vartheta_X \times \vartheta_X$ and the function $\cdot^* : \vartheta_X \to \vartheta_X$ are defined by the following simultaneous recursion:

(i) The term 0 is an element of ϑ_X.
(ii) The term Ω is an element of ϑ_X.
(iii) For each $x \in X$ we have a term $\mathfrak{E}_x \in \vartheta_X$.
(iv) If s is a term in ϑ_X, then so is ϑs.
(v) If $s_0, \ldots, s_n$ are terms in ϑ_X, then so is $\omega^{s_0} + \cdots + \omega^{s_n}$, provided that
   - either $n = 0$ and s_0 is not of the form Ω, $\mathfrak{E}_x$ or $\vartheta s'$,
   - or $n > 0$ and $s_n \leq_{\vartheta_X} \cdots \leq_{\vartheta_X} s_0$.

The map $s \mapsto s^*$ is given by

$$0^* = 0, \quad \Omega^* = 0, \quad \mathfrak{E}_x^* = 0, \quad (\vartheta s)^* = \vartheta s, \quad (\omega^{s_0} + \cdots + \omega^{s_n})^* = \max_{i \leq n} s_i^*,$$

where the maximum is taken with respect to $<_{\vartheta_X}$. For $s, t \in \vartheta_X$ we have $s <_{\vartheta_X} t$ if and only if one of the following holds:

(i') We have $s = 0$ and $t \neq 0$.
(ii') We have $s = \Omega$ and
   - either t is of the form $\mathfrak{E}_x$,
   - or $t = \omega^{t_0} + \cdots + \omega^{t_n}$ with $s \leq_{\vartheta_X} t_0$.

(iii') We have $s = \mathfrak{E}_x$ and

- either $t = \mathfrak{E}_y$ with $x <_X y$,
- or $t = \omega^{t_0} + \cdots + \omega^{t_n}$ with $s \leq_{\vartheta_X} t_0$.

(iv') We have $s = \vartheta s'$ and

- either $t = \vartheta t'$ with $s' <_{\vartheta_X} t'$ and $(s')^* <_{\vartheta_X} \vartheta t'$,
- or $t = \vartheta t'$ with $\vartheta s' \leq_{\vartheta_X} (t')^*$,
- or t is of the form Ω or $\mathfrak{E}_x$,
- or $t = \omega^{t_0} + \cdots + \omega^{t_n}$ with $s \leq_{\vartheta_X} t_0$.

(v') We have $s = \omega^{s_0} + \cdots + \omega^{s_n}$ and

- either t is of the form $\Omega, \mathfrak{E}_x$ or $\vartheta t'$ and $s_0 <_{\vartheta_X} t$,
- or $t = \omega^{t_0} + \cdots + \omega^{t_m}$ and one of the following holds:
  — Either we have $n < m$ and $s_i = t_i$ for all $i \leq n$,
  — or there is a $j \leq \min\{n, m\}$ with $s_j <_{\vartheta_X} t_j$ and $s_i = t_i$ for $i < j$.

To formalize this in **RCA$_0$** one starts with a term system $\vartheta_X^0 \supseteq \vartheta_X$ that ignores the condition $s_n \leq_{\vartheta_X} \cdots \leq_{\vartheta_X} s_0$ in clause (v). One then defines $L_{\vartheta_X} : \vartheta_X^0 \to \mathbb{N}$ by

$$L_{\vartheta_X}(t) := \begin{cases} \max\{\ulcorner 0 \urcorner, \ulcorner t \urcorner\} & \text{if } t = 0, \ t = \Omega \text{ or } t = \mathfrak{E}_x, \\ \max\{\ulcorner t \urcorner, L_{\vartheta_X}(t') + 1\} & \text{if } t = \vartheta t', \\ \max\{\ulcorner t \urcorner, L_{\vartheta_X}(t_0) + \cdots + L_{\vartheta_X}(t_n) + 1\} & \text{if } t = \omega^{t_0} + \cdots + \omega^{t_n}. \end{cases}$$

The occurrence of $\ulcorner 0 \urcorner$ in the first case ensures $L_{\vartheta_X}(s^*) \leq L_{\vartheta_X}(s)$. To decide $s \in \vartheta_X$ and $t <_{\vartheta_X} t'$ and to compute s^* one proceeds by simultaneous recursion on $L_{\vartheta_X}(s)$ resp. $L_{\vartheta_X}(t) + L_{\vartheta_X}(t')$. By [19, Lemma 2.7], we have the following lemma.

Lemma 3.3 (RCA$_0$). *If $(X, <_X)$ is a linear order, then so is $(\vartheta_X, <_{\vartheta_X})$.*

Now, we come to the main technical result of this section.

Theorem 3.1 (RCA$_0$). *The abstract Bachmann–Howard principle implies that ϑ_X is well-founded for any well-order X.*

Proof. For a fixed well-order X and an arbitrary linear order Y we put

$$T_Y := \varepsilon_{Y \cup \{\Omega\} \cup X}.$$

Here, $Y \cup \{\Omega\} \cup X$ is ordered as written: any element of Y is smaller than the constant Ω, which is in turn smaller than any element of X. By Definition 3.2, we obtain an order $<_{T_Y}$ on T_Y. For each embedding $f : Y \to Y'$ we define an embedding $T_f : T_Y \to T_{Y'}$ by the recursion

$$T_f(t) = \begin{cases} t & \text{if } t = 0, \ t = \varepsilon_\Omega \text{ or } t = \varepsilon_x \text{ with } x \in X, \\ \varepsilon_{f(y)} & \text{if } t = \varepsilon_y \text{ with } y \in Y, \\ \omega^{T_f(t_0)} + \cdots + \omega^{T_f(t_n)} & \text{if } t = \omega^{t_0} + \cdots + \omega^{t_n}. \end{cases}$$

It is straightforward to verify that $Y \mapsto T_Y$ and $f \mapsto T_f$ form an endofunctor of linear orders. To obtain a prae-dilator we define $\mathrm{supp}_Y^T : T_Y \to [Y]^{<\omega}$ by

$$\mathrm{supp}_Y^T(t) = \begin{cases} \emptyset & \text{if } t = 0, \ t = \varepsilon_\Omega \text{ or } t = \varepsilon_x \text{ with } x \in X, \\ \{y\} & \text{if } t = \varepsilon_y \text{ with } y \in Y, \\ \bigcup_{i \leq n} \mathrm{supp}_Y^T(t_i) & \text{if } t = \omega^{t_0} + \cdots + \omega^{t_n}. \end{cases}$$

By Proposition 3.2, the abstract Bachmann–Howard principle ensures that $\varepsilon_{Y \cup \{\Omega\} \cup X}$ is well-founded for any well-order Y. Thus T is a dilator. Another application of the abstract Bachmann–Howard principle yields a well-order Y with a Bachmann–Howard collapse

$$\vartheta : T_Y \to Y.$$

In order to conclude we show that the function $f : \vartheta_X \to T_Y = \varepsilon_{Y \cup \{\Omega\} \cup X}$ with

$$f(0) = 0,$$
$$f(\Omega) = \varepsilon_\Omega,$$
$$f(\mathfrak{E}_x) = \varepsilon_x,$$
$$f(\vartheta t) = \varepsilon_{\vartheta(f(t))},$$
$$f(\omega^{t_0} + \cdots + \omega^{t_n}) = \omega^{f(t_0)} + \cdots + \omega^{f(t_n)}$$

is order preserving. To establish the implication

$$s <_{\vartheta_X} t \Rightarrow f(s) <_{T_Y} f(t)$$

we argue by induction on $L_{\vartheta_X}(s) + L_{\vartheta_X}(t)$. Simultaneously, one must verify

$$f(r^*) = \max\nolimits_{<_{T_Y}} (\{0\} \cup \{\varepsilon_y \mid y \in \mathrm{supp}_Y^T(f(r))\})$$

by induction on $L_{\vartheta_X}(r)$. The only interesting case is

$$s = \vartheta s' <_{\vartheta_X} \vartheta t' = t.$$

We have to consider two possibilities: First assume $s' <_{\vartheta_X} t'$ and $(s')^* <_{\vartheta_X} \vartheta t'$. By the simultaneous induction hypothesis we get

$$\{\varepsilon_y \mid y \in \mathrm{supp}_Y^T(f(s'))\} \leq_{T_Y}^{\mathrm{fin}} f((s')^*) <_{T_Y} f(\vartheta t') = \varepsilon_{\vartheta(f(t'))},$$

which implies $\mathrm{supp}_Y^T(f(s')) <_Y^{\mathrm{fin}} \vartheta(f(t'))$. The induction hypothesis also provides the inequality $f(s') <_{T_Y} f(t')$. By the definition of Bachmann–Howard collapse, we obtain $\vartheta(f(s')) <_Y \vartheta(f(t'))$ and then

$$f(s) = \varepsilon_{\vartheta(f(s'))} <_{T_Y} \varepsilon_{\vartheta(f(t'))} = f(t).$$

Now, assume that $s = \vartheta s' <_{\vartheta_X} \vartheta t' = t$ holds because of $s \leq_{\vartheta_X} (t')^*$. By the definition of Bachmann–Howard collapse, we have $\mathrm{supp}_Y^T(f(t')) <_Y^{\mathrm{fin}} \vartheta(f(t'))$. Together with the simultaneous induction hypothesis and $0 <_{T_Y} \varepsilon_{\vartheta(f(t'))}$ we can infer

$$f(s) \leq_{T_Y} f((t')^*) = \max\nolimits_{<_{T_Y}} (\{0\} \cup \{\varepsilon_y \mid y \in \mathrm{supp}_Y^T(f(t'))\}) <_{T_Y} \varepsilon_{\vartheta(f(t'))} = f(t),$$

as desired. $\qquad\square$

Note that the case $X = \emptyset$ of Theorem 3.1 yields the well-foundedness of the usual Bachmann–Howard ordinal. The following result completes our bootstrapping: It allows us to lower the base theory to $\mathbf{RCA_0}$ (cf. Theorem 4.3).

Corollary 3.1 ($\mathbf{RCA_0}$). *The abstract Bachmann–Howard principle implies all axioms of $\mathbf{ATR_0}$.*

Proof. According to Proposition 3.1, the abstract Bachmann–Howard principle implies that ω^X is well-founded for any well-order X. By the aforementioned result of Girard [8, Sec. II.5] and Hirst [10] this secures arithmetical comprehension. The latter allows us to argue in terms of countable coded ω-models and valuations of formulas in these models. Theorem 3.1 tells us that ϑ_X is well-founded for any well-order X. As shown by Rathjen and Valencia Vizcaíno [19] this implies that any set is contained in a countable coded ω-model of bar induction. To conclude we recall that the axioms of $\mathbf{ATR_0}$ are provable by bar induction and have complexity Π^1_2 (see [23, Corollary VII.2.19]). □

In view of Theorem 1.1, the abstract Bachmann–Howard principle implies the well-foundedness of even stronger notation systems below the proof-theoretic ordinal of Π^1_1-$\mathbf{CA_0}$ (and even some way beyond, if the base theory contains suitable induction principles). It would certainly be instructive to give direct well-ordering proofs that witness this fact. We do not do so in this paper, since Corollary 3.1 suffices to lower the base theory of Theorem 1.1, which was the main goal of this section.

4. Computing a Bachmann–Howard Fixed Point

In this section, we construct a notation system $\vartheta(T)$ for the smallest Bachmann–Howard fixed point of a coded prae-dilator T. The point is that $\vartheta(T)$ is computable relative to T, so that $\mathbf{RCA_0}$ proves its existence as a set and indeed a linear order. The statement that $\vartheta(T)$ is well-founded for any coded dilator T will be called the computable Bachmann–Howard principle. We will show that it is equivalent to the abstract Bachmann–Howard principle and thus to Π^1_1-comprehension.

To understand the construction of $\vartheta(T)$, assume that we have a Bachmann–Howard collapse $\vartheta : D^T_{\vartheta(T)} \to \vartheta(T)$. Any element of $D^T_{\vartheta(T)}$ is of the form $\langle a, \sigma \rangle$, where a is a finite subset of $\vartheta(T)$ and $\sigma \in T_{|a|}$ satisfies $\mathrm{supp}^T_{|a|}(\sigma) = |a|$. Write $a = \{s_0, \ldots, s_{n-1}\}$ with $s_0 <_{\vartheta(T)} \cdots <_{\vartheta(T)} s_{n-1}$. The idea is to represent the collapsed element $\vartheta(\langle a, \sigma \rangle) \in \vartheta(T)$ by the term $\vartheta^{s_0, \ldots, s_{n-1}}_\sigma$.

Definition 4.1 ($\mathbf{RCA_0}$). For each coded prae-dilator T the set $\vartheta(T)$ and the relation $<_{\vartheta(T)} \subseteq \vartheta(T) \times \vartheta(T)$ are defined by the following simultaneous recursion:

(i) If we have elements $s_0 <_{\vartheta(T)} \cdots <_{\vartheta(T)} s_{n-1}$ of $\vartheta(T)$ and an element $\sigma \in T_n$ with $\mathrm{supp}^T_n(\sigma) = n$, then the term $\vartheta^{s_0, \ldots, s_{n-1}}_\sigma$ is an element of $\vartheta(T)$ as well.

Given elements $s = \vartheta_\sigma^{s_0,\ldots,s_{n-1}}$ and $t = \vartheta_\tau^{t_0,\ldots,t_{m-1}}$ of $\vartheta(T)$, we have $s <_{\vartheta(T)} t$ precisely if one of the following holds:

(i') We have $T_f(\sigma) <_{T_k} T_g(\tau)$ for some strictly increasing functions

$$f : n \to k := |\{s_0,\ldots,s_{n-1},t_0,\ldots,t_{m-1}\}| \quad \text{and} \quad g : m \to k$$

with $f(i) < g(j) \Leftrightarrow s_i <_{\vartheta(T)} t_j$ and $g(j) < f(i) \Leftrightarrow t_j <_{\vartheta(T)} s_i$. Furthermore, we have $s_{n-1} <_{\vartheta(T)} t$ or $n = 0$.

(ii') We have $m > 0$ and $s \leq_{\vartheta(T)} t_{m-1}$.

Note that $n = 0$ is permitted in clause (i), leading to initial terms $\vartheta_\sigma^{\langle\rangle}$ with empty upper index. Thus the set $\vartheta(T)$ is empty if and only if T_0 is. The formulation of clause (i') is somewhat awkward because we do not yet know that $<_{\vartheta(T)}$ is a linear order. Once this fact is established we see that f and g are the unique functions that make the following diagram commute, where the vertical arrows are the increasing enumerations with respect to $<_{\vartheta(T)}$:

$$
\begin{array}{ccc}
n & \longrightarrow & \{s_0,\ldots,s_{n-1}\} \\
\big\downarrow{\scriptstyle f} & & \big\downarrow \\
k & \longrightarrow & \{s_0,\ldots,s_{n-1},t_0,\ldots,t_{m-1}\} \\
\big\uparrow{\scriptstyle g} & & \big\uparrow \\
m & \longrightarrow & \{t_0,\ldots,t_{m-1}\}\,.
\end{array}
$$

To formalize Definition 4.1 in **RCA$_0$** one starts with a set $\vartheta^0(T) \supseteq \vartheta(T)$ of terms that ignores the condition $s_0 <_{\vartheta(T)} \cdots <_{\vartheta(T)} s_{n-1}$ in clause (i). Then consider the length function $L_{\vartheta(T)} : \vartheta^0(T) \to \mathbb{N}$ defined by

$$
L_{\vartheta(T)}(s) = \begin{cases} \ulcorner s \urcorner & \text{if } s = \vartheta_\sigma^{\langle\rangle}, \\ \max\{\ulcorner s \urcorner, 2 \cdot L_{\vartheta(T)}(s_0) + \cdots + 2 \cdot L_{\vartheta(T)}(s_n) + 1\} & \text{if } s = \vartheta_\sigma^{s_0,\ldots,s_n}. \end{cases}
$$

Now, one can decide $r \in \vartheta(T)$ and $s <_{\vartheta(T)} t$ by simultaneous induction on $L_{\vartheta(T)}(r)$ and $L_{\vartheta(T)}(s) + L_{\vartheta(T)}(t)$. As in Sec. 3, we have included the Gödel number $\ulcorner s \urcorner$ in order to ensure that $\forall_s(L_{\vartheta(T)}(s) \leq n \to \cdots)$ amounts to a bounded quantifier. The significance of the factor 2 becomes clear in the following proof.

Proposition 4.1 (RCA$_0$). *For any coded prae-dilator T the relation $<_{\vartheta(T)}$ is a linear order on $\vartheta(T)$.*

Proof. By simultaneous induction on n one shows that

$$2 \cdot L_{\vartheta(T)}(s) \leq n \to s \not<_{\vartheta(T)} s,$$

$$L_{\vartheta(T)}(s) + L_{\vartheta(T)}(t) \leq n \to s <_{\vartheta(T)} t \vee s = t \vee t <_{\vartheta(T)} s,$$

$$L_{\vartheta(T)}(r) + L_{\vartheta(T)}(s) + L_{\vartheta(T)}(t) \leq n \to (r <_{\vartheta(T)} s \wedge s <_{\vartheta(T)} t \to r <_{\vartheta(T)} t)$$

holds for all $r, s, t \in \vartheta(T)$. To establish antisymmetry we write $s = \vartheta_\sigma^{s_0,\ldots,s_{n-1}}$. Aiming at a contradiction, assume first that $s <_{\vartheta(T)} s$ holds by clause (i') of Definition 4.1. The functions f and g can only be the identity on $n = k = m$. Thus, we would have to have $\sigma = T_f(\sigma) <_{T_n} T_g(\sigma) = \sigma$, contradicting the antisymmetry of $<_{T_n}$. Now assume that $s <_{\vartheta(T)} s$ holds by clause (ii'), which means that we have $n > 0$ and $s \leq_{\vartheta(T)} s_{n-1}$. On the other hand, we have $s_{n-1} <_{\vartheta(T)} s$, by clause (ii') and the trivial inequality $s_{n-1} \leq_{\vartheta(T)} s_{n-1}$. In view of

$$L_{\vartheta(T)}(s_{n-1}) + L_{\vartheta(T)}(s) + L_{\vartheta(T)}(s_{n-1}) < 2 \cdot L_{\vartheta(T)}(s)$$

we can invoke the induction hypothesis to get $s_{n-1} <_{\vartheta(T)} s_{n-1}$ by transitivity. This contradicts the antisymmetry available by induction hypothesis.

To establish trichotomy we write $s = \vartheta_\sigma^{s_0,\ldots,s_{n-1}}$ and $t = \vartheta_\tau^{t_0,\ldots,t_{m-1}}$. The induction hypothesis implies that $<_{\vartheta(T)}$ is linear on $\{s_0,\ldots,s_{n-1},t_0,\ldots,t_{m-1}\}$ (in particular the induction hypothesis covers $s_i <_{\vartheta(T)} t_j \wedge t_j <_{\vartheta(T)} s_i \rightarrow s_i <_{\vartheta(T)} s_i$, due to the factor 2 in the definition of $L_{\vartheta(T)}$). Thus, we can consider the unique functions f and g that make the above diagram commute. First, assume $T_f(\sigma) = T_g(\tau)$. Since supp^T is a natural transformation we get

$$[f]^{<\omega}(n) = [f]^{<\omega}(\operatorname{supp}_n^T(\sigma)) = \operatorname{supp}_k^T(T_f(\sigma))$$
$$= \operatorname{supp}_k^T(T_g(\tau)) = [g]^{<\omega}(\operatorname{supp}_m^T(\tau)) = [g]^{<\omega}(m).$$

Together with $[f]^{<\omega}(n) \cup [g]^{<\omega}(m) = k$ this implies that f and g must be the identity on $n = k = m$. Thus, we obtain $\sigma = \tau$ and $\langle s_0,\ldots,s_{n-1} \rangle = \langle t_0,\ldots,t_{m-1} \rangle$, which means $s = t$. Now, let us assume $T_f(\sigma) <_{T_k} T_g(\tau)$. If we have $n = 0$, then we get $s <_{\vartheta(T)} t$ by clause (i'). If we have $n > 0$, then the induction hypothesis yields $s_{n-1} <_{\vartheta(T)} t$ or $t \leq_{\vartheta(T)} s_{n-1}$. In the first case, we get $s <_{\vartheta(T)} t$ by clause (i'), while the second case yields $t <_{\vartheta(T)} s$ by clause (ii'). For $T_g(\tau) <_{T_k} T_f(\sigma)$ the argument is symmetric.

Finally, we establish transitivity: Consider terms $r = \vartheta_\rho^{r_0,\ldots,r_{l-1}}$, $s = \vartheta_\sigma^{s_0,\ldots,s_{n-1}}$ and $t = \vartheta_\tau^{t_0,\ldots,t_{m-1}}$ with $r <_{\vartheta(T)} s$ and $s <_{\vartheta(T)} t$. First assume that $s <_{\vartheta(T)} t$ holds by clause (ii'), i.e. that we have $m > 0$ and $s \leq_{\vartheta(T)} t_{m-1}$. Then the induction hypothesis yields $r <_{\vartheta(T)} t_{m-1}$, so that we get $r <_{\vartheta(T)} t$ by clause (ii'). Now, assume that $s <_{\vartheta(T)} t$ holds by clause (i') while $r <_{\vartheta(T)} s$ holds by clause (ii'). Then, we have $n > 0$ and $r \leq_{\vartheta(T)} s_{n-1} <_{\vartheta(T)} t$, so that the induction hypothesis yields $r <_{\vartheta(T)} t$. Finally, assume that both inequalities hold by clause (i'). Then, we have $l = 0$ or $r_{l-1} <_{\vartheta(T)} s$, which yields $r_{l-1} <_{\vartheta(T)} t$ by induction hypothesis. To see that the remaining condition of clause (i') is transitive one completes the above diagram by the inclusions into $\{r_0,\ldots,r_{l-1},s_0,\ldots,s_{n-1},t_0,\ldots,t_{m-1}\}$. □

We can now show a central result of this paper: The theory **RCA₀** proves the existence of Bachmann–Howard fixed points (but in general it will not prove their well-foundedness).

Theorem 4.1 (RCA₀). *Given any coded prae-dilator T, the linear order $\vartheta(T)$ is a Bachmann–Howard fixed point of T.*

Proof. We must construct a Bachmann–Howard collapse $\vartheta : D^T_{\vartheta(T)} \to \vartheta(T)$. In view of Definition 2.2 we set

$$\vartheta(\langle a, \sigma \rangle) = \vartheta^{s_0, \dots, s_{n-1}}_\sigma \quad \text{for } a = \{s_0, \dots, s_{n-1}\} \text{ with } s_0 <_{\vartheta(T)} \cdots <_{\vartheta(T)} s_{n-1}.$$

Let us verify the conditions from Definition 1.2: aiming at condition (i) we assume $\langle a, \sigma \rangle <_{D^T_{\vartheta(T)}} \langle b, \tau \rangle$. By Definition 2.2 this means $T_{|\iota^{a \cup b}_a|}(\sigma) <_{T_{|a \cup b|}} T_{|\iota^{a \cup b}_b|}(\tau)$, where $\iota^{a \cup b}_a$ and $\iota^{a \cup b}_b$ are the inclusion maps from a and b into $a \cup b$. Write $a = \{s_0, \dots, s_{n-1}\}$ and $b = \{t_0, \dots, t_{m-1}\}$ in increasing order, and observe that $|\iota^{a \cup b}_a|$ and $|\iota^{a \cup b}_b|$ coincide with the functions f and g from Definition 4.1(i'). Assuming the side condition of Definition 1.2(i) we also get

$$\{s_0, \dots, s_{n-1}\} = \operatorname{supp}^{D^T}_{\vartheta(T)}(\langle a, \sigma \rangle) <^{\text{fin}}_{\vartheta(T)} \vartheta(\langle b, \tau \rangle) = \vartheta^{t_0, \dots, t_{m-1}}_\tau.$$

Thus, we must have $s_{n-1} <_{\vartheta(T)} \vartheta^{t_0, \dots, t_{m-1}}_\tau$ or $n = 0$. Now, Definition 4.1(i') yields

$$\vartheta(\langle a, \sigma \rangle) = \vartheta^{s_0, \dots, s_{n-1}}_\sigma <_{\vartheta(T)} \vartheta^{t_0, \dots, t_{m-1}}_\tau = \vartheta(\langle b, \tau \rangle),$$

as required by condition (i) of Definition 1.2. To establish condition (ii) we consider $\langle a, \sigma \rangle \in D^T_{\vartheta(T)}$ and write $a = \{s_0, \dots, s_{n-1}\}$ in increasing order. In case $n > 0$, we observe

$$s_0 <_{\vartheta(T)} \cdots <_{\vartheta(T)} s_{n-1} <_{\vartheta(T)} \vartheta^{s_0, \dots, s_{n-1}}_\sigma = \vartheta(\langle a, \sigma \rangle),$$

where the last inequality holds by clause (ii') of Definition 4.1. This implies

$$\operatorname{supp}^{D^T}_{\vartheta(T)}(\langle a, \sigma \rangle) = \{s_0, \dots, s_{n-1}\} <^{\text{fin}}_{\vartheta(T)} \vartheta(\langle a, \sigma \rangle),$$

just as condition (ii) of Definition 1.2 demands. $\qquad\square$

In view of the theorem, the following assertion is at least as strong as (the second-order version of) the abstract Bachmann–Howard principle:

Definition 4.2 (RCA$_0$). The computable Bachmann–Howard principle is the statement that $\vartheta(T)$ is well-founded for any coded dilator T.

To see that the abstract Bachmann–Howard principle implies its computable counterpart we show that the Bachmann–Howard fixed point $\vartheta(T)$ is minimal:

Theorem 4.2 (RCA$_0$). *Consider a coded prae-dilator T. The order $\vartheta(T)$ can be embedded into any Bachmann–Howard fixed point of T.*

Proof. Let $\vartheta : D^T_X \to X$ be a Bachmann–Howard collapse. The desired embedding $f : \vartheta(T) \to X$ can be recursively defined by

$$f(\vartheta^{s_0, \dots, s_{n-1}}_\sigma) = \vartheta(\langle \{f(s_0), \dots, f(s_{n-1})\}, \sigma \rangle).$$

Inductively, we may assume that f is order preserving on $\{s_0, \dots, s_{n-1}\}$. In particular, we have $|\{f(s_0), \dots, f(s_{n-1})\}| = n$, so that $\langle \{f(s_0), \dots, f(s_{n-1})\}, \sigma \rangle$ is indeed

an element of D_X^T. The implication

$$s <_{\vartheta(T)} t \Rightarrow f(s) <_X f(t)$$

can be established by induction on $L_{\vartheta(T)}(s) + L_{\vartheta(T)}(t)$: Let us write $s = \vartheta_\sigma^{s_0,\ldots,s_{n-1}}$ and $t = \vartheta_\tau^{t_0,\ldots,t_{m-1}}$, as well as $a = \{s_0,\ldots,s_{n-1}\}$ and $b = \{t_0,\ldots,t_{m-1}\}$. First, assume that $s <_{\vartheta(T)} t$ holds by clause (i') of Definition 4.1. This means that we have $T_{|\iota_a^{a\cup b}|}(\sigma) <_{T_{|a\cup b|}} T_{|\iota_b^{a\cup b}|}(\tau)$, using the notation from Sec. 2. The induction hypothesis ensures that f is order preserving on $a \cup b$, which implies

$$|\iota_a^{a\cup b}| = \left| \iota_{[f]^{<\omega}(a)}^{[f]^{<\omega}(a)\cup[f]^{<\omega}(b)} \right| \quad \text{and} \quad |\iota_b^{a\cup b}| = \left| \iota_{[f]^{<\omega}(b)}^{[f]^{<\omega}(a)\cup[f]^{<\omega}(b)} \right|.$$

In view of Definition 2.2, we can infer

$$\langle [f]^{<\omega}(a), \sigma \rangle <_{D_X^T} \langle [f]^{<\omega}(b), \tau \rangle.$$

In case $n > 0$ we observe $s_0 <_{\vartheta(T)} \cdots <_{\vartheta(T)} s_{n-1} <_{\vartheta(T)} s <_{\vartheta(T)} t$, where the penultimate inequality holds by clause (ii') of Definition 4.1. By induction hypothesis we get $f(s_0) <_X \cdots <_X f(s_{n-1}) <_X f(t)$ and thus

$$\mathrm{supp}_X^{D^T}(\langle [f]^{<\omega}(a), \sigma \rangle) = [f]^{<\omega}(a) <_X^{\mathrm{fin}} f(t) = \vartheta(\langle [f]^{<\omega}(b), \tau \rangle).$$

Then condition (i) of Definition 1.2 yields

$$f(s) = \vartheta(\langle [f]^{<\omega}(a), \sigma \rangle) <_X \vartheta(\langle [f]^{<\omega}(b), \tau \rangle) = f(t),$$

as desired. Now assume that $s <_{\vartheta(T)} t$ holds by clause (ii') of Definition 4.1. This means that we have $m > 0$ and $s \leq_{\vartheta(T)} t_{m-1}$. By induction hypothesis, we can infer $f(s) \leq_X f(t_{m-1})$. Using clause (ii) of Definition 1.2 we also get

$$f(t_{m-1}) \in [f]^{<\omega}(b) = \mathrm{supp}_X^{D^T}(\langle [f]^{<\omega}(b), \tau \rangle) <_X^{\mathrm{fin}} \vartheta(\langle [f]^{<\omega}(b), \tau \rangle) = f(t),$$

which yields $f(t_{m-1}) <_X f(t)$. By transitivity we get $f(s) <_X f(t)$. $\qquad\square$

Putting results together, we obtain the following refinement of Theorem 1.1.

Theorem 4.3. *The following are equivalent over* $\mathbf{RCA_0}$:

(i) *The principle of* Π_1^1-*comprehension.*
(ii) *The abstract Bachmann–Howard principle.*
(iii) *The computable Bachmann–Howard principle.*

We remark that statement (ii) refers to the second-order version of the abstract Bachmann–Howard principle (cf. Definition 2.6). Over $\mathbf{ATR_0^{set}}$ the latter is equivalent to the set-theoretic version of the abstract Bachmann–Howard principle (cf. Definition 1.3), as we have shown in Lemma 2.5.

Proof. By Theorem 1.1 (and the aforementioned Lemma 2.5) the equivalence between (i) and (ii) can be proved in Simpson's set-theoretic version of $\mathbf{ATR_0}$. By conservativity (see [22, 23]) it is provable in $\mathbf{ATR_0}$ itself. Since Π_1^1-comprehension

implies arithmetical transfinite recursion, the implication (i) $\Rightarrow$ (ii) is already provable in **RCA$_0$**. In view of Corollary 3.1, the same holds for the implication (ii) $\Rightarrow$ (i). Aiming at (ii) $\Rightarrow$ (iii), we invoke the abstract Bachmann–Howard principle to get a well-founded Bachmann–Howard fixed point X of a given coded dilator T. By Theorem 4.2 there is an order embedding of $\vartheta(T)$ into X. Thus $\vartheta(T)$ is well-founded as well, as required by the computable Bachmann–Howard principle. Finally, the implication (iii) $\Rightarrow$ (ii) follows from the fact that $\vartheta(T)$ is a Bachmann–Howard fixed point of T, as established in Theorem 4.1. □

The merit of Theorem 4.3 is that it pinpoints the computational content: It shows that the strength of the Bachmann–Howard principle lies uniquely in the preservation of well-foundedness, not in the existence of a linearly ordered Bachmann–Howard fixed point as such. It would be interesting to use methods from computability theory (similar to those in [13]) to analyze the computable Bachmann–Howard principle, or indeed its contrapositive: can one describe a computable prae-dilator T with a computable descending sequence in $\vartheta(T)$, such that the hyperjump is computable from any witness to the fact that T fails to be a dilator?

References

[1] B. Afshari and M. Rathjen, Reverse mathematics and well-ordering principles: A pilot study, *Ann. Pure Appl. Logic* **160** (2009) 231–237.

[2] A. Freund, A higher Bachmann–Howard Principle, unpublished manuscript (2017), arXiv:1704.01662.

[3] A. Freund, Type-two well-ordering principles, admissible sets, and Π^1_1-comprehension, PhD thesis, University of Leeds (2018), http://etheses.whiterose.ac.uk/20929/.

[4] A. Freund, Π^1_1-comprehension as a well-ordering principle, *Adv. Math.* 355 (to appear), doi:10.1016/j.aim.2019.106767.

[5] A. Freund, A categorical construction of Bachmann–Howard fixed points, Bulletin of the London Mathematical Society (to appear), doi:10.1112/blms.12285.

[6] H. Friedman, A. Montalbán and A. Weiermann, A characterization of **ATR$_0$** in terms of a Kruskal-like tree theorem, unpublished.

[7] J.-Y. Girard, Π^1_2-*logic, part* 1: *Dilators*, *Ann. Pure Appl. Logic* **21** (1981) 75–219.

[8] J.-Y. Girard, *Proof Theory and Logical Complexity*, Studies in Proof Theory, Vol. 1 (Bibliopolis, Napoli, 1987).

[9] J.-Y. Girard, *Proof Theory and Logical Complexity*, Vol. 2, http://girard.perso. math.cnrs.fr/Archives4.html (accessed on November 21, 2017), 1982, Manuscript.

[10] J. L. Hirst, Reverse mathematics and ordinal exponentiation, *Ann. Pure Appl. Logic* **66** (1994) 1–18.

[11] G. Jäger, *Theories for Admissible Sets. A Unifying Approach to Proof Theory*, Studies in Proof Theory (Bibliopolis, Napoli, 1986).

[12] R. B. Jensen and C. Karp, Primitive recursive set functions, Axiomatic set theory, *Proc. Sympos. Pure Mathematics*, ed. Dana S. Scott, (Providence, Rhode Island) Vol. XIII, Part I (American Mathematical Society, 1971), pp. 143–176.

[13] A. Marcone and A. Montalbán, The Veblen functions for computability theorists, *J. Symbolic Logic* **76** (2011) 575–602.

[14] A. Montalbán, *Ordinal functors and* $\Pi_1^1 - \mathbf{CA_0}$, December 2009, draft.

[15] A. Montalbán, Open questions in reverse mathematics, *Bull. Symbolic Logic* **17** (2011) 431–454.

[16] M. Rathjen, A proof-theoretic characterization of the primitive recursive set functions, *J. Symbolic Logic* **57** (1992) 954–969.

[17] M. Rathjen, *Omega Models and Well-Ordering Principles, Beta Models and Functors*, Talk at the Reverse Mathematics Workshop, September 2011, University of Chicago.

[18] M. Rathjen, *ω-Models and Well-Ordering Principles*, Foundational Adventures: Essays in Honor of Harvey M. Friedman, ed. Neil Tennant (College Publications, 2014), pp. 179–212.

[19] M. Rathjen and P. F. Valencia Vizcaíno, *Well Ordering Principles and Bar Induction*, Gentzen's centenary: The quest for consistency, eds. Reinhard Kahle and Michael Rathjen (Springer, Berlin, 2015), pp. 533–561.

[20] M. Rathjen and A. Weiermann, Proof-theoretic investigations on Kruskal's theorem, *Ann. Pure Appl. Logic* **60** (1993) 49–88.

[21] M. Rathjen and A. Weiermann, *Reverse Mathematics and Well-Ordering Principles*, Computability in Context: Computation and Logic in the Real World, ed. S. Barry Cooper and Andrea Sorbi (Imperial College Press, 2011), pp. 351–370.

[22] S. G. Simpson, Set Theoretic Aspects of $\mathbf{ATR_0}$, Logic Colloquium '80, eds. Dirk van Dalen, Daniel Lascar and Timothy J. Smiley, Studies in Logic and the Foundations of Mathematics (North Holland, 1982), pp. 255–271.

[23] S. G. Simpson, *Subsystems of Second Order Arithmetic*, Perspectives in Logic (Cambridge University Press, 2009).

[24] I. A. Thomson, Well-ordering principles and Π_1^1-comprehension + bar induction, PhD thesis, University of Leeds (2017), http://etheses.whiterose.ac.uk/22206/.

[25] I. A. Thomson and M. Rathjen, *Well-ordering principles, ω-models and* Π_1^1-*comprehension*, The Legacy of Kurt Schütte, eds. Reinhard Kahle and Michael Rathjen (Springer), to appear.