

Upper bounds for metapredicative Mahlo in explicit mathematics and admissible set theory

Gerhard Jäger Thomas Strahm

Abstract

In this article we introduce systems for metapredicative Mahlo in explicit mathematics and admissible set theory. The exact upper proof-theoretic bounds of these systems are established.

1 Introduction

In classical set theory an ordinal κ is called a *Mahlo ordinal* if it is a regular cardinal and if, for every normal function f from κ to κ , there exists a regular cardinal μ less than κ so that $\{f(\xi) : \xi < \mu\} \subset \mu$. The statement that there exists a Mahlo ordinal is a powerful set existence axiom going beyond theories like ZFC. It also outgrows the existence of inaccessible cardinals, hyper inaccessibles, hyperhyperinaccessible and the like.

There is also an obvious recursive analogue of Mahlo ordinal. Typically, an ordinal α is baptized *recursively Mahlo*, if it is admissible and reflects every Π_2 sentence on a smaller admissible ordinal. The corresponding formal theory KPM has been proof-theoretically analyzed by Rathjen [14, 15]. KPM is a highly impredicative theory, and its proof-theoretic strength is significantly beyond that of KPi, the second order theory $(\Delta_2^1\text{-CA}) + (\text{BI})$ and Feferman's theory T_0 , which are all proof-theoretically equivalent (cf. [3, 6, 10]).

This article can be seen as a further contribution to the general program of *metapredicativity*. We have studied other metapredicative theories in Jäger, Kahle, Setzer and Strahm [8], Jäger and Strahm [11], and Strahm [19, 21]; there also some further background material can be found.

One aim here is to look at metapredicative Mahlo in admissible set theory. The corresponding theory, named KPM^0 , is admissible set theory above the natural numbers as urelements plus Π_2 reflection on the admissibles. As induction principles we have complete induction on the natural numbers for sets, but do *not include \in induction*.

A further aim of this paper is to introduce the concept of Mahloness into explicit mathematics and to analyze the proof-theoretic strength of its metapredicative version. An extension of Feferman's theory T_0 by Mahlo axioms is studied in Jäger and Studer [12]. Setzer [18] presents a related formulation in the framework of Martin-Löf type theory.

For the formalization of Mahlo in explicit mathematics we work over the basic theory EETJ which comprises the axioms of applicative theories and has type existence axioms for elementary comprehension and join. To obtain Mahlo, we introduce the concept of universe with the idea that universes act as explicit analogues of admissibles. Then we add Mahlo axioms which require that, for each name a of a type and for every operation f which maps names of types into names of types, there exists a universe with name $m(a, f)$ which contains a and is closed under f . The theory EMA is EETJ plus these Mahlo axioms and induction on the natural numbers for types. It provides one basic formalism for metapredicative Mahlo in explicit mathematics.

In the following we will see that the proof-theoretic ordinals of KPm^0 and EMA are less than or equal to $\varphi\omega 00$, for φ being the ternary Veblen function. Together with the results of Strahm [20] we thus conclude that $\varphi\omega 00$ is the proof-theoretic ordinal of KPm^0 and EMA. If complete induction on the natural numbers for arbitrary formulas is added, we obtain systems of strength $\varphi\varepsilon_0 00$.

2 The theory KPm^0

In this section we introduce the metapredicative theory KPm^0 for a recursively inaccessible Mahlo universe. Basically, KPm^0 is the theory KPi^0 of Jäger [7] augmented by an axiom scheme for Π_2 reflection on the admissibles. It is equivalent to the theory KPM of Rathjen [14] if complete induction on ω is restricted to sets and all other forms of \in induction are omitted.

For the following it will be convenient to work in a framework with the natural numbers as urelements. Therefore, we let \mathcal{L}_1 denote the language of first order arithmetic, which has *number variables* $a, b, c, d, e, f, u, v, w, x, y, z, \dots$ (possibly with subscripts), symbols for all primitive recursive functions and relations, as well as a unary relation symbol Q . Q plays the role of an anonymous relation variable with no specific meaning. Its role will become clear in Definition 1 below. There is also a symbol \sim for forming negative literals.¹

¹This formulation of the language is chosen for the Tait-style reformulation of our systems in the next section.

The *number terms* $(r, s, t, r_1, s_1, t_1, \dots)$ of \mathcal{L}_1 are defined as usual. Positive literals of \mathcal{L}_1 are all expressions $R(s_1, \dots, s_n)$ for R a symbol for an n -ary primitive recursive relation as well as expressions of the form $Q(s)$. The negative literals of \mathcal{L}_1 have the form $\sim E$ so that E is a positive literal. The *formulas* of \mathcal{L}_1 are now generated from the positive and negative literals of \mathcal{L}_1 by closing against disjunction, conjunction, as well as existential and universal number quantification. The negation $\neg A$ of an \mathcal{L}_1 formula A is defined by making use of De Morgan's laws and the law of double negation. Moreover, the remaining logical connectives are abbreviated in the standard way.

KPM^0 is formulated in the extension $\mathcal{L}^* = \mathcal{L}_1(\in, \mathbf{N}, \mathbf{S}, \text{Ad})$ of \mathcal{L}_1 by the membership relation symbol \in , the set constant \mathbf{N} for the set of natural numbers and the unary relation symbols \mathbf{S} and Ad for sets and admissibles, respectively.

The terms $(\mathbf{t}, \mathbf{s}, \mathbf{t}, \mathbf{r}_1, \mathbf{s}_1, \mathbf{t}_1, \dots)$ of \mathcal{L}^* are the terms of \mathcal{L}_1 plus the set constant \mathbf{N} . The formulas $(A, B, C, A_1, B_1, C_1, \dots)$ of \mathcal{L}^* as well as the Δ_0 , Σ , Π , Σ_n and Π_n formulas of \mathcal{L}^* are defined as usual. The notation $\vec{\mathbf{s}}$ is shorthand for a finite string $\mathbf{s}_1, \dots, \mathbf{s}_n$ whose length will be specified by the context. Equality between objects is not represented by a primitive symbol but defined by

$$(\mathbf{s} = \mathbf{t}) := \begin{cases} (\mathbf{s} \in \mathbf{N} \wedge \mathbf{t} \in \mathbf{N} \wedge (\mathbf{s} =_{\mathbf{N}} \mathbf{t})) \vee \\ (\mathbf{S}(\mathbf{s}) \wedge \mathbf{S}(\mathbf{t}) \wedge (\forall x \in \mathbf{s})(x \in \mathbf{t}) \wedge (\forall x \in \mathbf{t})(x \in \mathbf{s})) \end{cases}$$

where $=_{\mathbf{N}}$ is the symbol for the primitive recursive equality on the natural numbers. The formula $A^{\mathbf{s}}$ is the result of replacing each unrestricted quantifier $\exists x(\dots)$ and $\forall x(\dots)$ in A by $(\exists x \in \mathbf{s})(\dots)$ and $(\forall x \in \mathbf{s})(\dots)$, respectively. In addition, we freely make use of all standard set-theoretic notations and write, for example, $\text{Tran}(\mathbf{s})$ for the Δ_0 formula saying that \mathbf{s} is a transitive set.

Let \mathcal{F} be a collection of \mathcal{L}^* formulas. Induction on the natural numbers with respect to \mathcal{F} consist of all formulas

$$(\mathcal{F}\text{-I}_{\mathbf{N}}) \quad A(0) \wedge (\forall x \in \mathbf{N})(A(x) \rightarrow A(x')) \rightarrow (\forall x \in \mathbf{N})A(x)$$

so that $A(a)$ belongs to \mathcal{F} . Below we are particularly interested in the two induction schemas $(\Delta_0\text{-I}_{\mathbf{N}})$ and $(\mathcal{L}^*\text{-I}_{\mathbf{N}})$.

Now we introduce the theory KPM^0 for metapredicative Mahlo. Its logical axioms comprise the usual axioms of classical first order logic with equality. The non-logical axioms of KPM^0 can be divided into the following five groups.

I. **Ontological axioms.** We have for all terms \mathfrak{r} , $\vec{\mathfrak{s}}$ and \mathfrak{t} of \mathcal{L}^* , all function symbols h and relation symbols R of \mathcal{L}_1 and all axioms $A(\vec{a})$ of group III whose free variables belong to the list \vec{a} :

- (1) $\mathfrak{r} \in \mathbf{N} \leftrightarrow \neg \mathbf{S}(\mathfrak{r})$.
- (2) $\vec{\mathfrak{s}} \in \mathbf{N} \rightarrow h(\vec{\mathfrak{s}}) \in \mathbf{N}$.
- (3) $R(\vec{\mathfrak{s}}) \rightarrow \vec{\mathfrak{s}} \in \mathbf{N}$.
- (4) $\mathfrak{r} \in \mathfrak{t} \rightarrow \mathbf{S}(\mathfrak{t})$.
- (5) $\mathbf{Ad}(\mathfrak{t}) \rightarrow (\mathbf{N} \in \mathfrak{t} \wedge \mathbf{Tran}(\mathfrak{t}))$.
- (6) $\mathbf{Ad}(\mathfrak{t}) \rightarrow (\forall \vec{x} \in \mathfrak{t}) A^t(\vec{x})$.
- (7) $\mathbf{Ad}(\mathfrak{r}) \wedge \mathbf{Ad}(\mathfrak{t}) \rightarrow \mathfrak{r} \in \mathfrak{t} \vee \mathfrak{r} = \mathfrak{t} \vee \mathfrak{t} \in \mathfrak{r}$.

II. **Number-theoretic axioms.** We have for all axioms $A(\vec{a})$ of Peano arithmetic PA which are not instances of the schema of complete induction and whose free variables belong to the list \vec{a} :

(Number theory) $(\forall \vec{x} \in \mathbf{N}) A^{\mathbf{N}}(\vec{x})$.

III. **Kripke Platek axioms.** We have for all terms \mathfrak{s} and \mathfrak{t} and all Δ_0 formulas $A(a)$ and $B(a, b)$ of \mathcal{L}^* :

- (Pair) $\exists x(\mathfrak{s} \in x \wedge \mathfrak{t} \in x)$.
- (Tran) $\exists x(\mathfrak{s} \subset x \wedge \mathbf{Tran}(x))$.
- (Δ_0 -Sep) $\exists y(\mathbf{S}(y) \wedge y = \{x \in \mathfrak{s} : A(x)\})$.
- (Δ_0 -Coll) $(\forall x \in \mathfrak{s}) \exists y B(x, y) \rightarrow \exists z(\forall x \in \mathfrak{s})(\exists y \in z) B(x, y)$.

IV. **Mahlo axioms.** We have for all Δ_0 formulas $A(a, b, \vec{c})$ whose parameters belong to the list a, b, \vec{c} :

(M) $(\forall x)(\exists y) A(x, y, \vec{c}) \rightarrow (\exists z)[\mathbf{Ad}(z) \wedge \vec{c} \in z \wedge (\forall x \in z)(\exists y \in z) A(x, y, \vec{c})]$

V. **Induction axioms.** These consist of the schema (Δ_0 -I_N) of complete induction on the natural numbers for Δ_0 formulas.

Observe that \in induction is not available in \mathbf{KPM}^0 . This is the price for obtaining a metapredicative theory. The situation here is analogue to that for theories of iterated admissible sets dealing with recursive inaccessibility. The theory \mathbf{KPi} introduced in Jäger [5] can be considered as a formalized

approach to a recursively inaccessible universe; it contains full induction on the natural numbers and full \in induction. KPi is fairly strong and proof-theoretically equivalent, for example, to the theory $(\Delta_2^1\text{-CA}) + (\text{BI})$, however, if we omit \in induction and restrict induction on the natural numbers to sets, the resulting theory KPi^0 has proof-theoretic ordinal Γ_0 (cf. Jäger [7]).

In the following we will measure the proof-theoretic strength of formal theories in terms of their proof-theoretic ordinals. As usual, for all primitive recursive relations \prec and all \mathcal{L}^* formulas $A(a)$ we set:

$$\begin{aligned} \text{Prog}(\prec, A) &:= (\forall x \in \mathbf{N})[(\forall y \in \mathbf{N})(y \prec x \rightarrow A(y)) \rightarrow A(x)], \\ \text{TI}(\prec, A) &:= \text{Prog}(\prec, A) \rightarrow (\forall x \in \mathbf{N})A(x). \end{aligned}$$

Thus $\text{TI}(\prec, A)$ expresses transfinite induction along the relation \prec for the formula $A(a)$. The proof-theoretic ordinal of a theory \mathbb{T} is defined by referring to transfinite induction for the anonymous relation \mathbf{Q} .

Definition 1 1. An ordinal α is provable in a theory \mathbb{T} , if there is a primitive recursive wellordering \prec of order type α so that $\mathbb{T} \vdash \text{TI}(\prec, \mathbf{Q})$.

2. The least ordinal which is not provable in \mathbb{T} is called the proof-theoretic ordinal of \mathbb{T} and is denoted by $|\mathbb{T}|$.

The proof-theoretic ordinals of the theories which we consider in this paper are most easily expressed by making use of a *ternary* Veblen or φ function which we are now going to define. The usual Veblen hierarchy, generated by the *binary* function φ , starting off with the function $\varphi 0 \beta = \omega^\beta$, is well known from the literature, cf. Pohlers [13] or Schütte [17]. The *ternary* φ function is obtained as a straightforward generalization of the binary case by defining $\varphi \alpha \beta \gamma$ inductively as follows:

- (i) $\varphi 0 \beta \gamma$ is just $\varphi \beta \gamma$;
- (ii) if $\alpha > 0$, then $\varphi \alpha 0 \gamma$ denotes the γ th ordinal which is strongly critical with respect to all functions $\lambda \xi, \eta. \varphi \delta \xi \eta$ for $\delta < \alpha$.
- (iii) if $\alpha > 0$ and $\beta > 0$, then $\varphi \alpha \beta \gamma$ denotes the γ th common fixed point of the functions $\lambda \xi. \varphi \alpha \delta \xi$ for $\delta < \beta$.

For example, $\varphi 1 0 \alpha$ is Γ_α , and more generally, $\varphi 1 \alpha \beta$ denotes a Veblen hierarchy over $\lambda \alpha. \Gamma_\alpha$. It is straightforward how to extend these ideas in order to obtain φ functions of all finite arities, and even further to Schütte's Klammersymbole [16].

In Strahm [20] a primitive recursive notation system based on the ternary Veblen function is used to show that $\varphi\omega 00$ and $\varphi\varepsilon_0 00$ are lower bounds of the proof-theoretic ordinals of KPm^0 and $\text{KPm}^0 + (\mathcal{L}^* - \text{I}_{\mathbb{N}})$, respectively. In this article we will establish that these lower bounds are sharp.

Actually, we do not give a detailed proof-theoretic treatment of KPm^0 ; rather we turn to corresponding systems of explicit mathematics and exhibit their upper proof-theoretic bounds via an interpretation into suitable theories of ordinals. A proof-theoretic analysis of the latter systems will be carried out in detail. An obvious adaptation of the proof-theoretic treatment of these theories for ordinals also works for KPm^0 .

3 The theory EMA

The counterpart of the theory KPm^0 in explicit mathematics is the system **EMA** which we will describe below. The role of admissibles in set theory is now played by (names of) so-called universes. The Mahlo axiom of KPm^0 allows to reflect set-theoretic Π_2 formulas on admissibles. In **EMA** we reflect functions from names to names on universes. The remaining axioms of **EMA** are the usual first order axioms of explicit mathematics, axioms for uniform elementary comprehension and join, as well as complete induction on the natural numbers for types.

EMA is formulated in the second order language \mathbb{L} for individuals and types. It comprises individual variables $a, b, c, f, g, h, u, v, w, x, y, z, \dots$ as well as type variables U, V, W, X, Y, Z, \dots (both possibly with subscripts). \mathbb{L} also includes the individual constants \mathbf{k}, \mathbf{s} (combinators), $\mathbf{p}, \mathbf{p}_0, \mathbf{p}_1$ (pairing and projections), 0 (zero), $\mathbf{s}_{\mathbb{N}}$ (successor), $\mathbf{p}_{\mathbb{N}}$ (predecessor), $\mathbf{d}_{\mathbb{N}}$ (definition by numerical cases) and additional individual constants, called *generators*, which will be used for the uniform naming of types, namely \mathbf{nat} (natural numbers), \mathbf{id} (identity), \mathbf{co} (complement), \mathbf{int} (intersection), \mathbf{dom} (domain), \mathbf{inv} (inverse image), \mathbf{j} (join) and \mathbf{m} (universe generator). There is one binary function symbol \cdot for (partial) application of individuals to individuals. Further, \mathbb{L} has unary relation symbols \downarrow (defined) and \mathbf{N} (natural numbers) as well as three binary relation symbols \in (membership), $=$ (equality) and \mathfrak{R} (naming, representation). Finally, \mathbb{L} has a symbol \sim for forming negative literals.

It is convenient that \mathbb{L} also includes an anonymous unary relation symbol \mathbf{Q} and a corresponding generator \mathbf{q} . As in the previous section, \mathbf{Q} plays the role of an anonymous predicate on the natural numbers with no specific mathematical meaning.

The *individual terms* $(r, s, t, r_1, s_1, t_1, \dots)$ of \mathbb{L} are built up from individual

variables and individual constants by means of our function symbol \cdot for application. In the following we often abbreviate $(s \cdot t)$ simply as (st) , st or sometimes also $s(t)$; the context will always ensure that no confusion arises. We further adopt the convention of association to the left so that $s_1 s_2 \dots s_n$ stands for $(\dots (s_1 \cdot s_2) \dots s_n)$. We also set $t' := s_{\mathbf{N}}t$. Finally, we define general n tupling by induction on $n \geq 2$ as follows:

$$(s_1, s_2) := \mathbf{p}s_1 s_2, \quad (s_1, \dots, s_{n+1}) := ((s_1, \dots, s_n), s_{n+1}).$$

The positive literals of \mathbb{L} are the formulas $\mathbf{N}(s)$, $s \downarrow$, $s = t$, $U = V$, $s \in U$ and $\mathfrak{R}(s, U)$, and the negative literals of \mathbb{L} have the form $\sim E$ so that E is a positive literal. Since we work with a logic of partial terms, it is not guaranteed that all terms have values, and $s \downarrow$ is read as *s is defined* or *s has a value*. Moreover, $\mathbf{N}(s)$ says that s is a natural number, and the formula $\mathfrak{R}(s, U)$ is used to express that the individual s *represents* the type U or is a *name* of U .

The *formulas* $(A, B, C, A_1, B_1, C_1, \dots)$ of \mathbb{L} are generated from the positive and negative literals by closing against disjunction, conjunction, as well as existential and universal quantification for individuals and types. The negation $\neg A$ of an \mathbb{L} formula A is defined as usual. The following table contains a useful list of abbreviations:

$$\begin{aligned} s \simeq t &:= s \downarrow \vee t \downarrow \rightarrow s = t, \\ s \in \mathbf{N} &:= \mathbf{N}(s), \\ (\exists x \in \mathbf{N})A(x) &:= (\exists x)(x \in \mathbf{N} \wedge A(x)), \\ (\forall x \in \mathbf{N})A(x) &:= (\forall x)(x \in \mathbf{N} \rightarrow A(x)), \\ U \subset V &:= (\forall x)(x \in U \rightarrow x \in V), \\ s \dot{\in} t &:= (\exists X)(\mathfrak{R}(t, X) \wedge s \in X), \\ (\exists x \dot{\in} s)A(x) &:= (\exists x)(x \dot{\in} s \wedge A(x)), \\ (\forall x \dot{\in} s)A(x) &:= (\forall x)(x \dot{\in} s \rightarrow A(x)), \\ s \dot{=} t &:= (\exists X)[\mathfrak{R}(s, X) \wedge \mathfrak{R}(t, X)], \\ s \dot{\subset} t &:= (\exists X, Y)[\mathfrak{R}(s, X) \wedge \mathfrak{R}(t, Y) \wedge X \subset Y], \\ \mathfrak{R}(s) &:= (\exists X)\mathfrak{R}(s, X). \end{aligned}$$

The vector notation \vec{U} and \vec{s} is sometimes used to denote finite sequences of type variables U_1, \dots, U_m and individual terms s_1, \dots, s_n , respectively, whose length is given by the context.

The logic of EMA is Beeson's classical *logic of partial terms* (cf. Beeson [2] or Troelstra and Van Dalen [22]) for the individuals and classical logic with

equality for the types. Observe that Beeson's formalization includes the usual strictness axioms.

Before turning to our theory **EMA** for metapredicative Mahlo in explicit mathematics, we introduce the auxiliary theory **EETJ** which provides a framework for explicit elementary types with join. The nonlogical axioms of **EETJ** can be divided into the following groups:

I. **Applicative axioms.** These axioms formalize that the individuals form a partial combinatory algebra, that we have pairing and projection and the usual closure conditions on the natural numbers plus definition by numerical cases.

- (1) $kab = a$,
- (2) $sab\downarrow \wedge sabc \simeq ac(bc)$,
- (3) $p_0(a, b) = a \wedge p_1(a, b) = b$,
- (4) $0 \in \mathbf{N} \wedge (\forall x \in \mathbf{N})(x' \in \mathbf{N})$,
- (5) $(\forall x \in \mathbf{N})(x' \neq 0 \wedge p_{\mathbf{N}}(x') = x)$,
- (6) $(\forall x \in \mathbf{N})(x \neq 0 \rightarrow p_{\mathbf{N}}x \in \mathbf{N} \wedge (p_{\mathbf{N}}x)' = x)$,
- (7) $a \in \mathbf{N} \wedge b \in \mathbf{N} \wedge a = b \rightarrow d_{\mathbf{N}}xyab = x$,
- (8) $a \in \mathbf{N} \wedge b \in \mathbf{N} \wedge a \neq b \rightarrow d_{\mathbf{N}}xyab = y$.

As usual, from axioms (1) and (2), one derives a theorem about λ abstraction and a form of the recursion theorem.

II. **Explicit representation and extensionality.** The following axioms state that each type has a name, that there are no homonyms and that equality of types is extensional.

- (1) $(\exists x)\mathfrak{R}(x, U)$,
- (2) $\mathfrak{R}(a, U) \wedge \mathfrak{R}(a, V) \rightarrow U = V$,
- (3) $(\forall x)(x \in U \leftrightarrow x \in V) \rightarrow U = V$.

III. **Basic type existence axioms.** In the following we provide a finite axiomatization of uniform elementary comprehension plus join.

Natural numbers

$$\mathfrak{R}(\mathbf{nat}) \wedge (\forall x)(x \in \mathbf{nat} \leftrightarrow \mathbf{N}(x)).$$

Representation of Q

$$\mathfrak{R}(\mathbf{q}) \wedge (\forall x)(x \dot{\in} \mathbf{q} \leftrightarrow \mathbf{Q}(x)) \wedge \mathbf{q} \dot{\subset} \mathbf{nat}.$$

Identity

$$\mathfrak{R}(\mathbf{id}) \wedge (\forall x)(x \dot{\in} \mathbf{id} \leftrightarrow (\exists y)(x = (y, y))).$$

Complements

$$\mathfrak{R}(a) \rightarrow \mathfrak{R}(\mathbf{co}(a)) \wedge (\forall x)(x \dot{\in} \mathbf{co}(a) \leftrightarrow x \notin a).$$

Intersections

$$\mathfrak{R}(a) \wedge \mathfrak{R}(b) \rightarrow \mathfrak{R}(\mathbf{int}(a, b)) \wedge (\forall x)(x \dot{\in} \mathbf{int}(a, b) \leftrightarrow x \dot{\in} a \wedge x \dot{\in} b).$$

Domains

$$\mathfrak{R}(a) \rightarrow \mathfrak{R}(\mathbf{dom}(a)) \wedge (\forall x)(x \dot{\in} \mathbf{dom}(a) \leftrightarrow (\exists y)((x, y) \dot{\in} a)).$$

Inverse images

$$\mathfrak{R}(a) \rightarrow \mathfrak{R}(\mathbf{inv}(a, f)) \wedge (\forall x)(x \dot{\in} \mathbf{inv}(a, f) \leftrightarrow fx \dot{\in} a).$$

Joins

$$\mathfrak{R}(a) \wedge (\forall x \dot{\in} a)\mathfrak{R}(fx) \rightarrow \mathfrak{R}(\mathbf{j}(a, f)) \wedge \Sigma(a, f, \mathbf{j}(a, f)).$$

In this last axiom the formula $\Sigma(a, f, b)$ expresses that b names the disjoint union of f over a , i.e.

$$\Sigma(a, f, b) := (\forall x)(x \dot{\in} b \leftrightarrow (\exists y, z)(x = (y, z) \wedge y \dot{\in} a \wedge z \dot{\in} fy)).$$

IV. Uniqueness of generators. These axioms essentially guarantee that different generators create different names. To achieve this we have, for syntactically different generators r_0 and r_1 and arbitrary generators s and t :

- (1) $r_0 \neq r_1$,
- (2) $(\forall x)(sx \neq tx)$,
- (3) $(\forall x, y)(sx = ty \rightarrow s = t \wedge x = y)$.

In the original formulation of explicit mathematics, elementary comprehension is not dealt with by a finite axiomatization but directly as an infinite axiom scheme. An \mathbb{L} formula A is called *elementary* if it contains neither the relation symbol \mathfrak{R} nor bound type variables. The following theorem of Feferman and Jäger [4] shows that this scheme of uniform elementary comprehension is provable from our finite axiomatization. Join and uniqueness of generators are not needed for this argument.

Theorem 2 For every elementary formula $A(u, \vec{v}, W_1, \dots, W_n)$ with at most the indicated free variables there exists a closed term t so that one can prove in EETJ:

1. $\bigwedge_{i=1}^n \mathfrak{R}(w_i, W_i) \rightarrow \mathfrak{R}(t(\vec{v}, w_1, \dots, w_n)),$
2. $\bigwedge_{i=1}^n \mathfrak{R}(w_i, W_i) \rightarrow (\forall x)(x \in t(\vec{v}, w_1, \dots, w_n) \leftrightarrow A(x, \vec{v}, W_1, \dots, W_n)).$

In the following we employ two forms of induction on the natural numbers, type induction and formula induction. Type induction is the axiom

$$(\mathbb{T}\text{-I}_{\mathbb{N}}) \quad (\forall X)(0 \in X \wedge (\forall x \in \mathbb{N})(x \in X \rightarrow x' \in X) \rightarrow (\forall x \in \mathbb{N})(x \in X)).$$

Formula induction, on the other hand, is the schema

$$(\mathbb{L}\text{-I}_{\mathbb{N}}) \quad A(0) \wedge (\forall x \in \mathbb{N})(A(x) \rightarrow A(x')) \rightarrow (\forall x \in \mathbb{N})A(x)$$

for each \mathbb{L} formula A . According to Feferman [3], one has the following theorem about the proof-theoretic strength of EETJ plus type and formula induction, respectively.

Theorem 3 The theory EETJ + $(\mathbb{T}\text{-I}_{\mathbb{N}})$ is proof-theoretically equivalent to Peano arithmetic PA and to the system $(\Sigma_1^1\text{-AC})_0$; the theory EETJ + $(\mathbb{L}\text{-I}_{\mathbb{N}})$ is proof-theoretically equivalent to $(\Sigma_1^1\text{-AC})$.²

The next step is to introduce the concept of a universe in explicit mathematics. To put it very simply, a universe is supposed to be a type which consists of names only and reflects the theory EETJ.

For the detailed definition of a universe we introduce some auxiliary notation and let $\mathcal{C}(W, a)$ be the closure condition which is the disjunction of the following \mathbb{L} formulas:

- (1) $a = \text{nat} \vee a = \mathbf{q} \vee a = \text{id},$
- (2) $(\exists x)(a = \text{co}(x) \wedge x \in W),$
- (3) $(\exists x, y)(a = \text{int}(x, y) \wedge x \in W \wedge y \in W),$
- (4) $(\exists x)(a = \text{dom}(x) \wedge x \in W),$
- (5) $(\exists x, f)(a = \text{inv}(x, f) \wedge x \in W),$
- (6) $(\exists x, f)[a = \mathbf{j}(x, f) \wedge x \in W \wedge (\forall y \in x)(fy \in W)].$

² $(\Sigma_1^1\text{-AC})_0$ is $(\Sigma_1^1\text{-AC})$ with induction restricted to sets.

Thus the formula $(\forall x)(\mathcal{C}(W, x) \rightarrow x \in W)$ states that W is a type which is closed under the type constructions of **EETJ**, i.e. elementary comprehension and join. If, in addition, all elements of W are names, we call W a universe.

Definition 4 1. We write $\mathbf{U}(W)$ to express that the type W is a universe,

$$\mathbf{U}(W) := (\forall x)(\mathcal{C}(W, x) \rightarrow x \in W) \wedge (\forall x \in W)\mathfrak{R}(x).$$

2. We write $\mathcal{U}(a)$ to express that the individual a is a name of a universe,

$$\mathcal{U}(a) := (\exists X)(\mathfrak{R}(a, X) \wedge \mathbf{U}(X)).$$

Based on (names of) universes we can now introduce the Mahlo axiom for explicit mathematics. Given a name a and an operation f from names to names one simply claims that there exists (a name of) a universe $\mathfrak{m}(a, f)$ which contains a and reflects f . Taking up the analogy that regular cardinals in classical set theory correspond to universes in explicit mathematics, our formulation of Mahlo in explicit mathematics may be regarded as a uniform version of Mahlo in classical set theory.

The following shorthand notations are useful for obtaining a compact form of our Mahlo axiom:

$$\begin{aligned} (f : \mathfrak{R} \rightarrow \mathfrak{R}) &:= (\forall x)(\mathfrak{R}(x) \rightarrow \mathfrak{R}(fx)), \\ (f : s \rightarrow s) &:= (\forall x \dot{\in} s)(fx \dot{\in} s). \end{aligned}$$

Obviously, $(f : \mathfrak{R} \rightarrow \mathfrak{R})$ and $(f : s \rightarrow s)$ means that f maps names to names and elements of (the type named by) s to elements of (the type named by) s , respectively. Mahloness in explicit mathematics is now expressed by the axioms

$$(M.1) \quad \mathfrak{R}(a) \wedge (f : \mathfrak{R} \rightarrow \mathfrak{R}) \rightarrow \mathcal{U}(\mathfrak{m}(a, f)) \wedge a \dot{\in} \mathfrak{m}(a, f),$$

$$(M.2) \quad \mathfrak{R}(a) \wedge (f : \mathfrak{R} \rightarrow \mathfrak{R}) \rightarrow (f : \mathfrak{m}(a, f) \rightarrow \mathfrak{m}(a, f)).$$

It is an interesting topic to see what kind of ordering principles for universes can be consistently added to the previous axioms. This question is discussed at full length in Jäger, Kahle and Studer [9], and it is shown there that one must not be too liberal. As a consequence of these considerations we do not claim linearity and connectivity for arbitrary universes, but only for so-called *normal universes*, i.e. universes which are named by means of the type generator \mathfrak{m} ,

$$\mathcal{U}_{\mathfrak{m}}(a) := (\exists x, f)[a = \mathfrak{m}(x, f) \wedge \mathcal{U}(a)].$$

Linearity and connectivity of normal universes are then given by the following two axioms:

$$(\mathcal{U}_m\text{-Lin}) \quad (\forall x, y)[\mathcal{U}_m(x) \wedge \mathcal{U}_m(y) \rightarrow x \dot{\in} y \vee x \dot{=} y \vee y \dot{\in} x],$$

$$(\mathcal{U}_m\text{-Con}) \quad (\forall x, y)[\mathcal{U}_m(x) \wedge \mathcal{U}_m(y) \rightarrow x \dot{\subset} y \vee y \dot{\subset} x].$$

It is shown in [9] that connectivity of normal universes also implies transitivity of normal universes in its most general form.

The theory **EMA** of explicit mathematics, whose universe is Mahlo, comprises the theory **EETJ** plus type induction ($\mathbb{T}\text{-I}_{\mathbb{N}}$), the Mahlo axioms (M.1) and (M.2) as well as the ordering principles ($\mathcal{U}_m\text{-Lin}$) and ($\mathcal{U}_m\text{-Con}$).

Strahm [20] provides a proof that $\varphi_{\omega 0 0}$ is a lower bound to the proof-theoretic ordinal of **EMA** and that $\varphi_{\varepsilon_0 0 0}$ is a lower bound to the proof-theoretic ordinal of **EMA** + ($\mathbb{L}\text{-I}_{\mathbb{N}}$). In the next two sections we will prove that these ordinals are also upper bounds of the proof-theoretic ordinal of the theories **EMA** and **EMA** + ($\mathbb{L}\text{-I}_{\mathbb{N}}$), respectively, thus establishing the exact proof-theoretic strength of these systems.

No methods of impredicative proof theory are used in our analysis of **EMA** and **EMA** + ($\mathbb{L}\text{-I}_{\mathbb{N}}$) so that the metapredicativity of both systems is established. Impredicative Mahlo in explicit mathematics is obtained by adding the principle of inductive generation to **EMA**, cf. Jäger and Studer [12].

4 The theory **OMA**

In this section we introduce the ordinal theory **OMA** for the Mahlo axiom. It is a first order theory with ordinals tailored for dealing with certain non-monotone inductive definitions which provides the appropriate framework for modelling our theory **EMA**. In the next section we will determine the upper proof-theoretic bounds of **OMA**.

In the following we make use of the usual primitive recursive coding machinery in \mathcal{L}_1 : $\langle \dots \rangle$ is a standard primitive recursive function for forming n -tuples $\langle t_1, \dots, t_n \rangle$; Seq is the primitive recursive set of sequence numbers; $lh(t)$ denotes the length of (the sequence number coded by) t ; $(t)_i$ is the i th component of (the sequence coded by) t if $i < lh(t)$, i.e. $t = \langle (t)_0, \dots, (t)_{lh(t)-1} \rangle$ if t is a sequence number.

Now let \mathbf{P} be a fresh n -ary relation symbol and write $\mathcal{L}_1(\mathbf{P})$ for the extension of \mathcal{L}_1 by \mathbf{P} . An $\mathcal{L}_1(\mathbf{P})$ formula which contains at most a_1, \dots, a_n free is called an n -ary *operator form*, and we let $\mathcal{A}(\mathbf{P}, a_1, \dots, a_n)$ range over such forms.

The theory **OMA** is formulated in the language \mathcal{L}_0 which extends \mathcal{L}_1 by adding a new sort of ordinal variables $\sigma, \tau, \eta, \xi, \dots$ (possibly with subscripts), new binary relation symbols $<$ and $=$ for the less and equality relation on the ordinals, respectively, and a unary relation symbol **Ad** to express that an ordinal is admissible. Moreover, \mathcal{L}_0 includes an $(n + 1)$ -ary relation symbol $P_{\mathcal{A}}$ for each operator form $\mathcal{A}(P, a_1, \dots, a_n)$.

The number terms of \mathcal{L}_0 are the number terms of \mathcal{L}_1 , and the ordinal terms of \mathcal{L}_0 are the ordinal variables. The positive literals of \mathcal{L}_0 are the positive literals of \mathcal{L}_1 plus all expressions $(\sigma < \tau)$, $(\sigma = \tau)$ and $P_{\mathcal{A}}(\sigma, \vec{s})$ for each n -ary operator form $\mathcal{A}(P, \vec{a})$. We write $P_{\mathcal{A}}^{\sigma}(\vec{s})$ for $P_{\mathcal{A}}(\sigma, \vec{s})$. The negative literals of \mathcal{L}_0 are the expressions $\sim E$ with E a positive literal of \mathcal{L}_0 .

The *formulas* $(A, B, C, A_1, B_1, C_1, \dots)$ of \mathcal{L}_0 are generated from the positive and negative literals by closing under conjunction and disjunction, quantification over natural numbers, and the bounded ordinal quantifiers $(\exists \xi < \sigma)$ and $(\forall \xi < \sigma)$ as well as the unbounded ordinal quantifiers $(\exists \xi)$ and $(\forall \xi)$.

An \mathcal{L}_0 formula is called Σ^0 if it does not contain ordinal quantifiers of the form $(\forall \xi)$; it is called Π^0 if it does not have ordinal quantifiers of the form $(\exists \xi)$. Finally, the Δ_0^0 formulas of \mathcal{L}_0 are those formulas which are both Σ^0 and Π^0 ; the Σ_1^0 formulas of \mathcal{L}_0 are the Δ_0^0 formulas plus all formulas of the form $(\exists \xi)A(\xi)$ with A a Δ_0^0 formula and accordingly for Π_1^0 formulas. Further, we write A^{σ} to denote the \mathcal{L}_0 formula which is obtained from A by replacing all unbounded ordinal quantifiers $(Q\xi)$ in A by bounded ordinal quantifiers $(Q\xi < \sigma)$. Additional abbreviations are

$$P_{\mathcal{A}}^{\leq \sigma}(\vec{s}) := (\exists \xi < \sigma)P_{\mathcal{A}}^{\xi}(\vec{s}) \quad \text{and} \quad P_{\mathcal{A}}(\vec{s}) := (\exists \xi)P_{\mathcal{A}}^{\xi}(\vec{s}).$$

The theory **OMA** is formulated in classical two sorted predicate logic with equality in both sorts, containing the axioms of Peano arithmetic **PA**, linearity axioms for the ordinals, operator axioms and certain reflection principles.

I. **Number-theoretic axioms.** The axioms of Peano arithmetic **PA** with the exception of complete induction on the natural numbers.

II. **Linearity axioms.**

$$\sigma \not< \sigma \wedge (\sigma < \tau \wedge \tau < \eta \rightarrow \sigma < \eta) \wedge (\sigma < \tau \vee \sigma = \tau \vee \tau < \sigma).$$

III. **Operator axioms.** For all operator forms $\mathcal{A}(P, \vec{a})$:

$$P_{\mathcal{A}}^{\sigma}(\vec{s}) \leftrightarrow \mathcal{A}(P_{\mathcal{A}}^{\leq \sigma}, \vec{s}).$$

IV. Σ^0 **reflection axioms.** For all Σ^0 formulas A :

$$A \rightarrow (\exists \xi)A^{\xi}.$$

V. Axioms for Π_2^0 reflection on Ad. For all Δ_0^0 formulas $A(\xi, \eta, \vec{\tau})$ whose free ordinal variables are from the list $\xi, \eta, \vec{\tau}$:

$$(\forall \xi)(\exists \eta)A(\xi, \eta, \vec{\tau}) \rightarrow (\exists \sigma)[\text{Ad}(\sigma) \wedge \vec{\tau} < \sigma \wedge (\forall \xi < \sigma)(\exists \eta < \sigma)A(\xi, \eta, \vec{\tau})].$$

VI. Axioms for Ad. For all Σ^0 formulas $A(\vec{\tau})$ whose free ordinal variables are from the list $\vec{\tau}$:

$$\text{Ad}(\sigma) \wedge \vec{\tau} < \sigma \wedge A^\sigma(\vec{\tau}) \rightarrow (\exists \xi < \sigma)A^\xi(\vec{\tau}).$$

VII. Δ_0^0 induction on the natural numbers. For all Δ_0^0 formulas $A(a)$:

$$A(0) \wedge (\forall x)(A(x) \rightarrow A(x')) \rightarrow (\forall x)A(x).$$

Observe that there are no induction principles for ordinals. In this respect the theory **OMA** is related to **KPm**⁰ in which no \in induction is present. Even very weak induction along the ordinals, for example for Δ_0^0 formulas, would make our theory much stronger than the theory **EMA**.

Later in this paper we will sometimes use the fact that, for each ordinal σ in **OMA**, there exists an ordinal τ greater than σ . This is an immediate consequence of Σ^0 reflection.

The next step is to build a model of **EMA** in **OMA**. The crucial idea is to choose a suitable operator form $\mathcal{A}(\mathcal{P}, a, b, c)$ so that the relation symbol $\mathcal{P}_{\mathcal{A}}$ can then be used to single out the numbers which name types, and to define elementhood in the names of types. Before doing this, we have to translate term application and the individual constants of the language \mathbb{L} into \mathcal{L}_1 .

We interpret application \cdot of \mathbb{L} in the sense of ordinary recursion theory so that $(a \cdot b)$ in \mathbb{L} is translated into $\{a\}(b)$ in \mathcal{L}_1 , where $\{n\}$ for $n = 0, 1, 2, 3, \dots$ is a standard enumeration of the partial recursive functions. Then it is possible to assign pairwise different numerals to the constants $\mathbf{k}, \mathbf{s}, \mathbf{p}, \mathbf{p}_0, \mathbf{p}_1, \mathbf{s}_N, \mathbf{p}_N$ and \mathbf{d}_N so that the applicative axioms (1)–(8) of **EMA** are satisfied. We also require that the constant 0 of \mathbb{L} is interpreted as the 0 of \mathcal{L}_1 and the term $\mathbf{s}_N a$ of \mathbb{L} as $a + 1$ in \mathcal{L}_1 . In addition, we let pairing and projections of \mathbb{L} go over into the primitive recursive pairing and unpairing machinery introduced above.

For each \mathbb{L} term t there also exists an \mathcal{L}_1 formula $\text{Val}_t(a)$ expressing that a is the value of t under the interpretation described above. Accordingly, the atomic formulas $t \downarrow$, $(s = t)$ and $\mathbf{N}(t)$ are given their obvious interpretations in \mathcal{L}_1 with the translation of \mathbf{N} ranging over all natural numbers.

For dealing with the generators we choose, again by ordinary recursion theory, numerals nat, q, id, co, int, dom, inv, j and m so that we have the following properties:

$$\begin{aligned} \underline{\text{nat}} &= \langle 0, 0 \rangle, \quad \underline{\text{q}} = \langle 1, 0 \rangle, \quad \underline{\text{id}} = \langle 2, 0 \rangle, \quad \{\underline{\text{co}}\}(a) = \langle 3, a \rangle, \\ \{\underline{\text{int}}\}(\langle a, b \rangle) &= \langle 4, a, b \rangle, \quad \{\underline{\text{dom}}\}(a) = \langle 5, a \rangle, \quad \{\underline{\text{inv}}\}(\langle a, b \rangle) = \langle 6, a, b \rangle, \\ \{\underline{\text{j}}\}(\langle a, b \rangle) &= \langle 7, a, b \rangle, \quad \{\underline{\text{m}}\}(\langle a, b \rangle) = \langle 8, a, b \rangle, \quad \{e_0\}(a) \neq e_1 \end{aligned}$$

for all natural numbers a, b and all e_0 and e_1 from the set ranging over nat, q, id, co, int, dom, inv, j and m.

It is our strategy to define a specific operator form $\mathcal{A}(\mathbf{P}, a, b, c)$ and use the corresponding relation symbol $\mathbf{P}_{\mathcal{A}}$ for dealing with codes for types and elements of types. Later our interpretation will be so that

$$\begin{aligned} \mathfrak{R}(a) &\text{ translates into } (\exists \xi) \mathbf{P}_{\mathcal{A}}^{\xi}(a, 0, 0) \text{ and} \\ b \dot{\in} a &\text{ translates into } (\exists \xi) \mathbf{P}_{\mathcal{A}}^{\xi}(a, b, 1). \end{aligned}$$

Before turning to our final operator form $\mathcal{A}(\mathbf{P}, a, b, c)$ we introduce the auxiliary ternary operator form $\mathcal{A}_0(\mathbf{P}, a, b, c)$ which is the disjunction of the following formulas (1)–(16):

- (1) $a = \langle 0, 0 \rangle \wedge b = 0 \wedge c = 0,$
- (2) $a = \langle 0, 0 \rangle \wedge c = 1,$
- (3) $a = \langle 1, 0 \rangle \wedge b = 0 \wedge c = 0,$
- (4) $a = \langle 1, 0 \rangle \wedge \mathbf{Q}(b) \wedge c = 1,$
- (5) $a = \langle 2, 0 \rangle \wedge b = 0 \wedge c = 0,$
- (6) $a = \langle 2, 0 \rangle \wedge (\exists x)(b = \langle x, x \rangle) \wedge c = 1,$
- (7) $(\exists u)[a = \langle 3, u \rangle \wedge \mathbf{P}(u, 0, 0)] \wedge b = 0 \wedge c = 0,$
- (8) $(\exists u)[a = \langle 3, u \rangle \wedge \mathbf{P}(u, 0, 0) \wedge \neg \mathbf{P}(u, b, 1)] \wedge c = 1,$
- (9) $(\exists u, v)[a = \langle 4, u, v \rangle \wedge \mathbf{P}(u, 0, 0) \wedge \mathbf{P}(v, 0, 0)] \wedge b = 0 \wedge c = 0,$
- (10) $(\exists u, v)[a = \langle 4, u, v \rangle \wedge \mathbf{P}(u, 0, 0) \wedge \mathbf{P}(v, 0, 0) \wedge \mathbf{P}(u, b, 1) \wedge \mathbf{P}(v, b, 1)]$
 $\wedge c = 1,$
- (11) $(\exists u)[a = \langle 5, u \rangle \wedge \mathbf{P}(u, 0, 0)] \wedge b = 0 \wedge c = 0,$

- (12) $(\exists u, x)[a = \langle 5, u \rangle \wedge \mathbf{P}(u, 0, 0) \wedge \mathbf{P}(u, \langle b, x \rangle, 1)] \wedge c = 1,$
- (13) $(\exists u, f)[a = \langle 6, u, f \rangle \wedge \mathbf{P}(u, 0, 0)] \wedge b = 0 \wedge c = 0,$
- (14) $(\exists u, f)[a = \langle 6, u, f \rangle \wedge \mathbf{P}(u, 0, 0) \wedge \mathbf{P}(u, \{f\}(b), 1)] \wedge c = 1,$
- (15) $(\exists u, f)[a = \langle 7, u, f \rangle \wedge \mathbf{P}(u, 0, 0) \wedge (\forall x)(\mathbf{P}(u, x, 1) \rightarrow \mathbf{P}(\{f\}(x), 0, 0))] \wedge b = 0 \wedge c = 0,$
- (16) $(\exists u, f)[a = \langle 7, u, f \rangle \wedge \mathbf{P}(u, 0, 0) \wedge (\forall x)(\mathbf{P}(u, x, 1) \rightarrow \mathbf{P}(\{f\}(x), 0, 0)) \wedge (\exists y, z)(b = \langle y, z \rangle \wedge \mathbf{P}(u, y, 1) \wedge \mathbf{P}(\{f\}(y), z, 1))] \wedge c = 1.$

This operator form takes care of all generators except the generator for the Mahlo axiom. In order to deal with the generator \mathbf{m} we have to make sure that $\mathbf{m}(a, f)$ is only made a name provided that the codes generated so far constitute a universe, contain a and reflect f . The following shorthand notation expresses that the names given by \mathbf{P} form a universe:

$$\mathbf{Univ}(\mathbf{P}) := (\forall a, b, c)[\mathcal{A}_0(\mathbf{P}, a, b, c) \rightarrow \mathbf{P}(a, b, c)].$$

$\mathcal{A}_1(\mathbf{P}, a, b, c)$ is the operator form given by the disjunction of $\mathcal{A}_0(\mathbf{P}, a, b, c)$ and the following formulas (17) and (18):

- (17) $(\exists x, f)[a = \langle 8, x, f \rangle \wedge \mathbf{P}(x, 0, 0) \wedge (\forall y)(\mathbf{P}(y, 0, 0) \rightarrow \mathbf{P}(\{f\}(y), 0, 0))] \wedge \mathbf{Univ}(\mathbf{P}) \wedge b = 0 \wedge c = 0,$
- (18) $(\exists x, f)[a = \langle 8, x, f \rangle \wedge \mathbf{P}(x, 0, 0) \wedge (\forall y)(\mathbf{P}(y, 0, 0) \rightarrow \mathbf{P}(\{f\}(y), 0, 0))] \wedge \mathbf{Univ}(\mathbf{P}) \wedge \mathbf{P}(b, 0, 0) \wedge c = 1.$

If we had foundation on the ordinals, this operator form $\mathcal{A}_1(\mathbf{P}, a, b, c)$ would be sufficient for our model construction. By induction on the ordinals we could show for example that $(\exists \xi)\mathbf{P}_{\mathcal{A}_1}^\xi(a, 0, 0)$ implies that there is a least such ξ . In our context, however, induction on the ordinals is not available. Thus, in order to have a “unique time stamp” for triples (a, b, c) to get into stages generated, we work with the following operator form $\mathcal{A}(\mathbf{P}, a, b, c)$:

$$\mathcal{A}(\mathbf{P}, a, b, c) := \mathcal{A}_1(\mathbf{P}, a, b, c) \wedge \neg \mathbf{P}(a, 0, 0).$$

Given this careful definition of the operator form $\mathcal{A}(\mathbf{P}, a, b, c)$, the following lemma concerning the stages of $\mathcal{A}(\mathbf{P}, a, b, c)$ is trivially provable in OMA:

Lemma 5 *The following assertions are provable in OMA:*

1. $\mathbf{P}_{\mathcal{A}}^\sigma(a, 0, 0) \wedge \mathbf{P}_{\mathcal{A}}^\tau(a, 0, 0) \rightarrow \sigma = \tau,$
2. $\mathbf{P}_{\mathcal{A}}^\sigma(a, b, 1) \rightarrow \mathbf{P}_{\mathcal{A}}^\sigma(a, 0, 0),$

$$3. \mathsf{P}_{\mathcal{A}}^{\sigma}(a, 0, 0) \rightarrow (\forall b)[\mathsf{P}_{\mathcal{A}}(a, b, 1) \leftrightarrow \mathsf{P}_{\mathcal{A}}^{\sigma}(a, b, 1)].$$

Before turning to the interpretation of the types, the \in relation and the naming relation we introduce the following definition:

$$\mathsf{Rep}(a) := (\exists \xi)\mathsf{P}_{\mathcal{A}}^{\xi}(a, 0, 0), \quad \mathsf{E}(b, a) := (\exists \xi)\mathsf{P}_{\mathcal{A}}^{\xi}(a, b, 1).$$

In our embedding of **EMA** into **OMA** we first assume that the number and types variables of \mathbb{L} are mapped into the number variables of \mathcal{L}_{\circ} so that no conflicts arise; to simplify the notation we often identify the type variables with their translations in \mathcal{L}_{\circ} . Then we let the type variables of **EMA** range over **Rep** and the translation of the atomic formulas of \mathbb{L} involving types is as follows:

$$\begin{aligned} \mathfrak{R}(t, U)^* &:= (\exists x)[\mathsf{Val}_t(x) \wedge \mathsf{Rep}(x) \wedge \mathsf{Rep}(U) \wedge (\forall y)(\mathsf{E}(y, x) \leftrightarrow \mathsf{E}(y, U))], \\ (t \in U)^* &:= (\exists x)[\mathsf{Val}_t(x) \wedge \mathsf{E}(x, U)], \\ (U = V)^* &:= (\forall x)(\mathsf{E}(x, U) \leftrightarrow \mathsf{E}(x, V)). \end{aligned}$$

On the basis of these basic cases the translation of arbitrary \mathbb{L} formulas A into \mathcal{L}_{\circ} formulas A^* should be obvious. The embedding of **EMA** into **OMA** is given by the following theorem.

Theorem 6 *We have for all \mathbb{L} formulas $A(\vec{U}, \vec{a})$ with all its free variables indicated that*

$$\mathsf{EMA} \vdash A(\vec{U}, \vec{a}) \implies \mathsf{OMA} \vdash \mathsf{Rep}(\vec{U}) \rightarrow A^*(\vec{U}, \vec{a}).$$

PROOF The proof proceeds by induction on the length of the derivation of the formula A . If A is an applicative axiom or an axiom concerning the uniqueness of generators then its translation is provable in **OMA** by our assumptions about the coding of the first order part of **EMA**. The translations of the axioms about explicit representation and extensionality as well as linearity ($\mathcal{U}_m\text{-Lin}$) and connectivity ($\mathcal{U}_m\text{-Con}$) of normal universes are easily verified. In the case of the basic type existence axioms we confine ourselves to showing the translation of the axioms about *Intersection*.

Assume we are given two natural numbers a and b so that $\mathsf{Rep}(a)$ and $\mathsf{Rep}(b)$. Hence, there exist ordinals σ and τ with $\mathsf{P}_{\mathcal{A}}^{\sigma}(a, 0, 0)$ and $\mathsf{P}_{\mathcal{A}}^{\tau}(b, 0, 0)$. Choose an ordinal η greater than σ and τ and carry through the following distinction by cases.

Case 1: $\neg \mathsf{P}_{\mathcal{A}}^{<\eta}(\langle 4, a, b \rangle, 0, 0)$. Then our operator form $\mathcal{A}(\mathsf{P}, a, b, c)$ yields $\mathsf{P}_{\mathcal{A}}^{\eta}(\langle 4, a, b \rangle, 0, 0)$. Moreover, we also have

$$(\forall x)[\mathsf{P}_{\mathcal{A}}^{\eta}(\langle 4, a, b \rangle, x, 1) \leftrightarrow \mathsf{P}_{\mathcal{A}}^{<\eta}(a, x, 1) \wedge \mathsf{P}_{\mathcal{A}}^{<\eta}(b, x, 1)].$$

In view of Lemma 5 we thus have $\text{Rep}(\langle 4, a, b \rangle)$ and for all natural numbers x that $\text{E}(x, \langle 4, a, b \rangle)$ if and only if $\text{E}(x, a)$ and $\text{E}(x, b)$.

Case 2: $\text{P}_{\mathcal{A}}^{<\eta}(\langle 4, a, b \rangle, 0, 0)$. Because of Lemma 5 there exists a unique ξ less than η so that $\text{P}_{\mathcal{A}}^{\xi}(\langle 4, a, b \rangle, 0, 0)$. Hence, the operator form $\mathcal{A}(\text{P}, a, b, c)$ forces $\text{P}_{\mathcal{A}}^{<\xi}(a, 0, 0)$, $\text{P}_{\mathcal{A}}^{<\xi}(b, 0, 0)$ and $\neg\text{P}_{\mathcal{A}}^{<\xi}(\langle 4, a, b \rangle, 0, 0)$. Now we proceed as in the previous case.

Hence, the intersection axiom of EMA is verified. The other basic type existence axioms are treated accordingly. Observe that Σ^{\circledast} reflection is essential for handling *Join*. By carrying through these proofs in detail one obtains even more: the basic type existence axioms can already be seen to be valid at the admissible stages of our construction, i.e.

$$(1) \quad \text{Ad}(\sigma) \rightarrow \text{Univ}(\text{P}_{\mathcal{A}}^{<\sigma}).$$

We finally turn to the verification of the axioms (M.1) and (M.2). To this end assume that we have a and f so that $\text{Rep}(a)$ and $(\forall x)(\text{Rep}(x) \rightarrow \text{Rep}(\{f\}(x)))$. Hence, there exists a τ so that

$$(2) \quad \text{P}_{\mathcal{A}}^{\tau}(a, 0, 0).$$

A simple transformation of our second assumption yields, in addition, that

$$(3) \quad (\forall \xi)(\forall x)(\exists \eta)[\text{P}_{\mathcal{A}}^{<\xi}(x, 0, 0) \rightarrow \text{P}_{\mathcal{A}}^{<\eta}(\{f\}(x), 0, 0)],$$

and, therefore, Σ^{\circledast} reflection gives

$$(4) \quad (\forall \xi)(\exists \eta)(\forall x)[\text{P}_{\mathcal{A}}^{<\xi}(x, 0, 0) \rightarrow \text{P}_{\mathcal{A}}^{<\eta}(\{f\}(x), 0, 0)].$$

Hence, Π_2^{\circledast} reflection on Ad provides an admissible σ so that

$$(5) \quad \tau < \sigma \wedge (\forall \xi < \sigma)(\exists \eta < \sigma)(\forall x)[\text{P}_{\mathcal{A}}^{<\xi}(x, 0, 0) \rightarrow \text{P}_{\mathcal{A}}^{<\eta}(\{f\}(x), 0, 0)].$$

In view of (2) and by a simple transformation we derive

$$(6) \quad \text{P}_{\mathcal{A}}^{<\sigma}(a, 0, 0) \wedge (\forall x)[\text{P}_{\mathcal{A}}^{<\sigma}(x, 0, 0) \rightarrow \text{P}_{\mathcal{A}}^{<\sigma}(\{f\}(x), 0, 0)].$$

As we have remarked above the admissibility of σ forces $\text{Univ}(\text{P}_{\mathcal{A}}^{<\sigma})$. As before we proceed by a distinction by cases.

Case 1: $\neg\text{P}_{\mathcal{A}}^{<\sigma}(\langle 8, a, f \rangle, 0, 0)$. Then our operator form $\mathcal{A}(\text{P}, a, b, c)$ gives $\text{P}_{\mathcal{A}}^{\sigma}(\langle 8, a, f \rangle, 0, 0)$ and, therefore, $\text{Rep}(\langle 8, a, f \rangle)$. Our operator form also gives for all x

$$(7) \quad \text{E}(x, \langle 8, a, f \rangle) \leftrightarrow \text{P}_{\mathcal{A}}^{<\sigma}(x, 0, 0).$$

Combining (6) and (7) shows that $\langle 8, a, f \rangle$ names a type which contains a and is closed under f . In view of $\text{Univ}(\mathbb{P}_A^{<\sigma})$ we indeed have by (7) that (the translation of) $\mathcal{U}(\langle 8, a, f \rangle)$ is true in our model.

Case 2: $\mathbb{P}_A^{<\sigma}(\langle 8, a, f \rangle, 0, 0)$. Because of Lemma 5 there exists a unique ξ less than σ so that $\mathbb{P}_A^\xi(\langle 8, a, f \rangle)$. Hence, we have that the operator form $\mathcal{A}(\mathbb{P}, a, b, c)$ forces $\mathbb{P}_A^{<\xi}(a, 0, 0)$ and $(\forall x)[\mathbb{P}_A^{<\xi}(x, 0, 0) \rightarrow \mathbb{P}_A^{<\xi}(\{f\}(x), 0, 0)]$ as well as $\text{Univ}(\mathbb{P}_A^{<\xi})$. The rest is as in the previous case.

Therefore, our Mahlo axioms (M.1) and (M.2) are shown to be valid in our model, and this completes the proof of the embedding of **EMA** into **OMA**. \square

5 Proof-theoretic analysis of OMA

Our next goal is to establish the upper proof-theoretic bound of **OMA**. First we show that the proof-theoretic strength of **OMA** is already exhausted by suitable theories for n -inaccessible ordinals for all finite $n < \omega$. Afterwards, we compute the upper proof-theoretic bounds of these auxiliary systems.

5.1 Reduction of OMA to n -inaccessibility

We begin our proof-theoretic analysis of **OMA** by eliminating the axiom of Π_2° reflection in favor of n -inaccessible ordinals for sufficiently large n less than ω . To this end, we first reformulate **OMA** in a Tait-style manner.

The Tait-style version OMA^\top of **OMA** is formulated for finite sets Γ, Λ, \dots (possibly with subscripts) of \mathcal{L}_\circ formulas. If A is an \mathcal{L}_\circ formula, then Γ, A is a shorthand for $\Gamma \cup \{A\}$, and similar for expressions of the form Γ, A, B . The system OMA^\top contains the following axioms and rules of inference.

I. **Axioms.** For all finite sets Γ of \mathcal{L}_\circ formulas, all Δ_0° formulas A and all Δ_0° formulas B which are axioms of **OMA**:

$$\Gamma, \neg A, A \quad \text{and} \quad \Gamma, B.$$

II. **Propositional and quantifier rules.** These include the usual Tait-style inference rules for the propositional connectives and all sorts of quantifiers.

III. Σ° **reflection rules.** For all finite sets Γ of \mathcal{L}_\circ formulas and for all Σ° formulas A :

$$\frac{\Gamma, A}{\Gamma, (\exists \xi)A^\xi}.$$

IV. Π_2° reflection on Ad rules. For all finite sets Γ of \mathcal{L}_{\circ} formulas and for all Δ_0° formulas $A(\xi, \eta, \vec{\tau})$ whose free ordinal variables are from the list $\xi, \eta, \vec{\tau}$:

$$\frac{\Gamma, (\forall\xi)(\exists\eta)A(\xi, \eta, \vec{\tau})}{\Gamma, (\exists\sigma)[\text{Ad}(\sigma) \wedge \vec{\tau} < \sigma \wedge (\forall\xi < \sigma)(\exists\eta < \sigma)A(\xi, \eta, \vec{\tau})]}.$$

V. Cut rules. For all finite sets Γ of \mathcal{L}_{\circ} formulas and all \mathcal{L}_{\circ} formulas A :

$$\frac{\Gamma, A \quad \Gamma, \neg A}{\Gamma}.$$

The notion $\text{OMA}^{\top} \stackrel{n}{\vdash} \Gamma$ is used to express that the set Γ is provable in OMA^{\top} by a proof of depth less than or equal to n ; we write $\text{OMA}^{\top} \stackrel{n}{\vdash}_{\star} \Gamma$ if Γ is provable in OMA^{\top} by a proof of depth less than or equal to n so that all its cut formulas are Σ_1° or Π_1° formulas. In addition, $\text{OMA}^{\top} \vdash \Gamma$ or $\text{OMA}^{\top} \vdash_{\star} \Gamma$ means that there exists a natural number n so that $\text{OMA}^{\top} \stackrel{n}{\vdash} \Gamma$ or $\text{OMA}^{\top} \stackrel{n}{\vdash}_{\star} \Gamma$, respectively.

One readily notes that the main formulas of all axioms and rules of OMA^{\top} are Σ_1° formulas. As a consequence, we obtain the following weak cut elimination theorem for OMA^{\top} .

Theorem 7 (Weak cut elimination) *We have for all finite sets Γ of \mathcal{L}_{\circ} formulas that*

$$\text{OMA}^{\top} \vdash \Gamma \quad \Longrightarrow \quad \text{OMA}^{\top} \vdash_{\star} \Gamma.$$

Of course, the axioms and rules of OMA^{\top} are tailored so that the OMA can be embedded into OMA^{\top} in a straightforward manner. Thus we obtain the following corollary.

Corollary 8 *If the \mathcal{L}_{\circ} formula A is provable in OMA, then there exists a natural number n so that $\text{OMA}^{\top} \stackrel{n}{\vdash}_{\star} A$.*

Our next immediate aim is to introduce the semiformal system \mathbf{H} , which will be used to interpret quasi-normalized OMA^{\top} derivations. Essentially, \mathbf{H} is OMA^{\top} without reflection rules and with complete induction on the natural numbers replaced by the ω rule. Moreover, in \mathbf{H} we no longer have *unbounded* ordinal quantifiers, and since \mathbf{H} is a semiformal system, free number variables are not present.

The language \mathcal{L} of \mathbf{H} is obtained from \mathcal{L}_{\circ} by omitting free number variables and *unbounded* quantifiers for ordinals. In addition, we assume that \mathcal{L} includes a new constant 0 for the least ordinal. Therefore, the ordinal terms

of \mathcal{L} are the constant 0 and the ordinal variables. We call two literals of \mathcal{L} *numerically equivalent*, if they are syntactically identical modulo number subterms which have the same value. The axioms and rules of inference of \mathbf{H} are now given as follows.

I. **Axioms, group 1.** For all finite sets Γ of \mathcal{L} formulas, all numerically equivalent \mathcal{L} literals A and B , and all true \mathcal{L}_1 literals C :

$$\Gamma, \neg A, B \quad \text{and} \quad \Gamma, C.$$

II. **Axioms, group 2.** For all finite sets Γ of \mathcal{L} formulas, all literals $A(\sigma)$ of \mathcal{L} , all ordinal terms μ, ν of \mathcal{L} and all (instances of) axioms B of OMA from the groups II, III and VI:

$$\Gamma, 0 = \mu, 0 < \mu \quad \text{and} \quad \Gamma, \mu \neq \nu, \neg A(\mu), A(\nu) \quad \text{and} \quad \Gamma, B.$$

III. **Propositional rules.** The usual Tait-style rules for disjunction and conjunction.

IV. **Number quantifier rules.** For all finite sets Γ of \mathcal{L} formulas and all \mathcal{L} formulas $A(s)$:

$$\frac{\Gamma, A(s)}{\Gamma, (\exists x)A(x)}, \quad \frac{\Gamma, A(t) \text{ for all closed number terms } t}{\Gamma, (\forall x)A(x)} \quad (\omega).$$

V. **Ordinal quantifier rules.** For all finite sets Γ of \mathcal{L} formulas, all \mathcal{L} formulas A , all ordinal terms μ, ν of \mathcal{L} and all ordinal variables σ so that the usual variable conditions are satisfied:

$$\frac{\Gamma, \mu < \nu \wedge A(\mu)}{\Gamma, (\exists \xi < \nu)A(\xi)}, \quad \frac{\Gamma, \sigma < \nu \rightarrow A(\sigma)}{\Gamma, (\forall \xi < \nu)A(\xi)}.$$

VI. **Cut rules.** For all finite sets Γ of \mathcal{L} formulas and all \mathcal{L} formulas A :

$$\frac{\Gamma, A \quad \Gamma, \neg A}{\Gamma}.$$

Similarly as before, $\mathbf{H} \stackrel{\alpha}{\vdash} \Gamma$ means that the finite set Γ of \mathcal{L} formulas has an \mathbf{H} proof of depth less than or equal to α . Furthermore, we write $\mathbf{H} \stackrel{\alpha}{\vdash}_0 \Gamma$ if Γ has a cut-free proof in \mathbf{H} of depth less than or equal to α . Moreover, we write $\mathbf{H} \stackrel{\leq \alpha}{\vdash} \Gamma$ and $\mathbf{H} \stackrel{< \alpha}{\vdash} \Gamma$ if there exists a $\beta < \alpha$ such that $\mathbf{H} \stackrel{\beta}{\vdash} \Gamma$ and $\mathbf{H} \stackrel{\beta}{\vdash}_0 \Gamma$, respectively.

For the following theorem, the notion of n -inaccessibility is crucial. By recursion on $n < \omega$ we define a formula $\mathbf{la}_n(\sigma)$ to express that σ is an n -inaccessible ordinal as follows:

$$\begin{aligned} \mathbf{la}_0(\sigma) &:= \text{Ad}(\sigma), \\ \mathbf{la}_{n+1}(\sigma) &:= \text{Ad}(\sigma) \wedge (\forall \xi < \sigma)(\exists \eta < \sigma)[\xi < \eta \wedge \mathbf{la}_n(\eta)]. \end{aligned}$$

We observe that each formula $\mathbf{la}_n(\sigma)$ is a Δ_0^{\circledast} formula without free number variables and therefore also an \mathcal{L} formula.

Unfortunately we need some further terminology before we turn to the main result of this section. If $\vec{\tau}$ is the sequence of ordinal variables τ_1, \dots, τ_m , then $(\vec{\tau} \not\prec \sigma)$ stands for the set

$$\{\tau_1 \not\prec \sigma, \dots, \tau_m \not\prec \sigma\}.$$

A finite set of $\mathcal{L}_{\circledast}$ formulas Λ is called an *instance* of the finite set of $\mathcal{L}_{\circledast}$ formulas Γ if it results from Γ by replacing all free number variables of formulas in Γ by closed number terms of \mathcal{L}_1 .

For an $\mathcal{L}_{\circledast}$ formula A we use the notation $A[\vec{\tau}]$ to express that all its free ordinal variables belong to the list $\vec{\tau}$; the analogous convention is employed for finite sets of $\mathcal{L}_{\circledast}$ formulas. Finally, we write Γ^σ for the finite set of $\mathcal{L}_{\circledast}$ formulas which is obtained from Γ by replacing each formula A in Γ by its restriction A^σ .

Theorem 9 (Reduction of OMA^{T}) *Assume that $\Gamma[\vec{\tau}]$ is a finite set of Σ^{\circledast} formulas of $\mathcal{L}_{\circledast}$. Then we have for all instances $\Lambda[\vec{\tau}]$ of $\Gamma[\vec{\tau}]$ and all natural numbers n that*

$$\text{OMA}^{\text{T}} \stackrel{\mu}{\star} \Gamma[\vec{\tau}] \implies \text{H} \stackrel{\omega(n+2)}{\star} \neg \mathbf{la}_n(\sigma), (\vec{\tau} \not\prec \sigma), \Lambda^\sigma[\vec{\tau}].$$

PROOF This theorem is proved by induction on n . In the following we exemplarily treat the cases of cut and Π_2^{\circledast} reflection on admissible ordinals. We note that complete induction on the natural numbers is dealt with as usual by making use of the ω rule. In all other cases the claim is immediate from the induction hypothesis.

Let us first look at the case where $\Gamma[\vec{\tau}]$ is the conclusion of a cut. Then there are natural numbers $n_0, n_1 < n$ and a Δ_0^{\circledast} formula $A[\xi, \vec{\eta}]$ so that

$$(1) \quad \text{OMA}^{\text{T}} \stackrel{\mu_0}{\star} \Gamma[\vec{\tau}], (\exists \xi) A[\xi, \vec{\eta}] \quad \text{and} \quad \text{OMA}^{\text{T}} \stackrel{\mu_1}{\star} \Gamma[\vec{\tau}], (\forall \xi) \neg A[\xi, \vec{\eta}].$$

Suppose that $\Lambda[\vec{\tau}], (\exists \xi) B[\xi, \vec{\eta}]$ is an instance of $\Gamma[\vec{\tau}], (\exists \xi) A[\xi, \vec{\eta}]$. Then inversion applied to the second premise and the induction hypothesis yield

$$(2) \quad \text{H} \stackrel{\omega(n_0+2)}{\star} \neg \mathbf{la}_{n_0}(\sigma), (\vec{\tau}, \vec{\eta} \not\prec \sigma), \Lambda^\sigma[\vec{\tau}], (\exists \xi < \sigma) B[\xi, \vec{\eta}],$$

$$(3) \quad \text{H} \stackrel{\omega(n_1+2)}{\star} \neg \mathbf{la}_{n_1}(\sigma), (\vec{\tau}, \vec{\eta}, \eta_0 \not\prec \sigma), \Lambda^\sigma[\vec{\tau}], \neg B[\eta_0, \vec{\eta}],$$

where η_0 is a fresh ordinal variable. From (2) and (3) we obtain

$$(4) \quad \text{H} \stackrel{<\omega(n+2)}{\star} \neg \mathbf{la}_n(\sigma), (\vec{\tau}, \vec{\eta} \not\prec \sigma), \Lambda^\sigma[\vec{\tau}], (\exists \xi < \sigma) B[\xi, \vec{\eta}],$$

$$(5) \quad \text{H} \stackrel{<\omega(n+2)}{\star} \neg \mathbf{la}_n(\sigma), (\vec{\tau}, \vec{\eta} \not\prec \sigma), \Lambda^\sigma[\vec{\tau}], (\forall \xi < \sigma) \neg B[\xi, \vec{\eta}].$$

Here we have used the obvious fact that $\mathbf{H} \mid^{\leq \omega} \neg \mathbf{la}_n(\sigma), \mathbf{la}_k(\sigma)$ for each natural number k less than n . A cut applied to (4) and (5) reveals

$$(6) \quad \mathbf{H} \mid^{\omega(n+2)} \neg \mathbf{la}_n(\sigma), (\vec{\tau} \not\prec \sigma), \Lambda^\sigma[\vec{\tau}]$$

since superfluous ordinal variables can be easily eliminated. This is as desired and completes the treatment of the cut rule.

Let us now turn to the heart of the reduction, namely the interpretation of Π_2° reflection on admissible ordinals. Assume that $\Gamma[\vec{\tau}]$ is the conclusion of the corresponding rule of \mathbf{OMA}^\top . Hence, there exist an $n_0 < n$ and a Δ_0° formula $A[\xi, \eta, \vec{\tau}]$ so that

$$(7) \quad \mathbf{OMA}^\top \mid_{\star}^{n_0} \Gamma[\vec{\tau}], (\forall \xi)(\exists \eta)A[\xi, \eta, \vec{\tau}].$$

An application of inversion to (7) with a fresh ordinal variable τ_0 forces

$$(8) \quad \mathbf{OMA}^\top \mid_{\star}^{n_0} \Gamma[\vec{\tau}], (\exists \eta)A[\tau_0, \eta, \vec{\tau}].$$

Next we choose an instance $\Lambda[\vec{\tau}], (\exists \eta)B[\tau_0, \eta, \vec{\tau}]$ of $\Gamma[\vec{\tau}], (\exists \eta)A[\tau_0, \eta, \vec{\tau}]$ and apply the induction hypothesis to (8) in order to obtain

$$(9) \quad \mathbf{H} \mid^{\omega(n_0+2)} \neg \mathbf{la}_{n_0}(\sigma_0), (\tau_0, \vec{\tau} \not\prec \sigma_0), \Lambda^{\sigma_0}[\vec{\tau}], (\exists \eta < \sigma_0)B[\tau_0, \eta, \vec{\tau}].$$

But from (9) we can immediately derive by bounded universal ordinal quantification that

$$(10) \quad \mathbf{H} \mid^{\leq \omega(n+2)} \neg \mathbf{la}_{n_0}(\sigma_0), (\vec{\tau} \not\prec \sigma_0), \Lambda^{\sigma_0}[\vec{\tau}], (\forall \xi < \sigma_0)(\exists \eta < \sigma_0)B[\xi, \eta, \vec{\tau}].$$

Moreover, it is an easy task to check that we also have

$$(11) \quad \mathbf{H} \mid^{\leq \omega} \neg \mathbf{la}_n(\sigma), (\vec{\tau} \not\prec \sigma), (\exists \sigma_0 < \sigma)[\mathbf{la}_{n_0}(\sigma_0) \wedge \vec{\tau} < \sigma_0].$$

By combining (10) and (11) and applying persistency, we can finally derive

$$(12) \quad \mathbf{H} \mid^{\omega(n+2)} \neg \mathbf{la}_n(\sigma), (\vec{\tau} \not\prec \sigma), \Lambda^\sigma[\vec{\tau}], C[\sigma, \vec{\tau}]$$

for $C[\sigma, \vec{\tau}]$ denoting the formula

$$(\exists \sigma_0 < \sigma)[\mathbf{Ad}(\sigma_0) \wedge \vec{\tau} < \sigma_0 \wedge (\forall \xi < \sigma_0)(\exists \eta < \sigma_0)B[\xi, \eta, \vec{\tau}]].$$

Since $C[\sigma, \vec{\tau}]$ is in fact an element of $\Lambda^\sigma[\vec{\tau}]$ we have indeed established that $\mathbf{H} \mid^{\omega(n+2)} \neg \mathbf{la}_n(\sigma), (\vec{\tau} \not\prec \sigma), \Lambda^\sigma[\vec{\tau}]$ as desired. Observe that we have made crucial use of the fact that $\Lambda[\vec{\tau}]$ contains Σ° formulas only in order to be able to apply persistency to $\Lambda^{\sigma_0}[\vec{\tau}]$. Altogether this completes the reduction of \mathbf{OMA}^\top to n -inaccessibility. \square

5.2 The semiformal systems $\mathbf{H}[S, n, \alpha]$

Let \mathcal{L}' be some extension of \mathcal{L} by additional constants for ordinals and let S be a (finite or infinite) set of \mathcal{L}' formulas. The final part of this section is devoted to the proof-theoretic analysis of semiformal systems $\mathbf{H}[S, n, \alpha]$ for each $n < \omega$ and each ordinal α . The crucial axioms of $\mathbf{H}[S, n, \alpha]$ claim: (i) all formulas of S ; (ii) the existence of α many n -inaccessible ordinals which are ordered in an increasing chain and are greater than all the ordinal constants occurring in S .

The language $\mathcal{L}[S, n, \alpha]$ of $\mathbf{H}[S, n, \alpha]$ is the extension of the language \mathcal{L} generated by the constants occurring in S plus additional *new* constants $\mathfrak{c}[S, n, \beta]$ for each $\beta < \alpha$. The semiformal system $\mathbf{H}[S, n, \alpha]$ includes the axioms and rules of inference of \mathbf{H} (extended to the language $\mathcal{L}[S, n, \alpha]$) plus the following axioms:

- (i) Γ, A and $\Gamma, \mathfrak{d} < \mathfrak{c}[S, n, 0]$,
- (ii) $\Gamma, \mathfrak{I}_n(\mathfrak{c}[S, n, \beta])$ and $\Gamma, \mathfrak{c}[S, n, \gamma] < \mathfrak{c}[S, n, \beta]$,

for all finite sets Γ of $\mathcal{L}[S, n, \alpha]$ formulas, all elements A of S , all ordinal constants \mathfrak{d} from S , and all ordinals $\gamma < \beta < \alpha$. The deducibility relation $\mathbf{H}[S, n, \alpha] \vdash^\beta \Gamma$ is understood as before.

We call a finite set $\Gamma[\tau_1, \dots, \tau_n]$ of $\mathcal{L}[S, n, \alpha]$ formulas *quasi closed* if there exist $\beta_1, \dots, \beta_n < \alpha$ so that Γ is of the form

$$\tau_1 \not\prec \mathfrak{c}[S, n, \beta_1], \dots, \tau_n \not\prec \mathfrak{c}[S, n, \beta_n], \Lambda[\tau_1, \dots, \tau_n].$$

Hence, in a quasi closed set of $\mathcal{L}[S, n, \alpha]$ formulas all occurring free ordinal variables are bound by some ordinal constant $\mathfrak{c}[S, n, \beta]$ with $\beta < \alpha$.

The following main result of this paragraph is the natural generalization of Main Lemma II in Jäger, Kahle, Setzer and Strahm [8] to the context of n -inaccessible ordinals.

Theorem 10 (Reduction of $\mathbf{H}[S, n, \beta + \omega^{1+\rho}]$) *Let Γ be a quasi closed set of $\mathcal{L}[S, n, \beta + \omega^{1+\rho}]$ formulas with the property that*

$$\mathbf{H}[S, n, \beta + \omega^{1+\rho}] \vdash^\alpha \Gamma.$$

Then we have for all ordinals γ less than $\omega^{1+\rho}$ which are big enough for Γ being a quasi closed set of $\mathcal{L}[S, n, \beta + \gamma]$ formulas that

$$\mathbf{H}[S, n, \beta + \gamma] \vdash^{\varphi(n+1)\rho\alpha} \Gamma.$$

PROOF This theorem is proved by induction on $n < \omega$. In the case of $n = 0$ one essentially proceeds as in the proof of Main Lemma II in [8], the only difference being that instead of fixed points one eliminates admissible ordinals. For the induction step we assume that our theorem holds for some natural number n . Then we show our claim for $n + 1$ by main induction on ρ and side induction on α . Again the main steps of the argument are similar to the proof of Main Lemma II in [8], but for definiteness we spell out the details in the sequel. We distinguish cases whether $\rho = 0$, ρ is a successor, or ρ is a limit ordinal.

(a) $\rho = 0$. Assume that Γ is a finite and quasi closed set of $\mathcal{L}[S, n+1, \beta+k]$ formulas for some natural number k so that $\text{H}[S, n+1, \beta+\omega] \models^{\alpha} \Gamma$. If Γ is an axiom of $\text{H}[S, n+1, \beta+\omega]$, then the claim is trivial. Furthermore, if Γ is the conclusion of a rule different from the cut rule, the claim is immediate from the induction hypothesis. Hence, the only critical case comes up if Γ is the conclusion of a cut rule. Then there exist a natural number $l \geq k$, $\alpha_0, \alpha_1 < \alpha$ and an $\mathcal{L}[S, n+1, \beta+l]$ formula A so that

$$(1) \quad \text{H}[S, n+1, \beta+\omega] \models^{\alpha_0} \Gamma, A \quad \text{and} \quad \text{H}[S, n+1, \beta+\omega] \models^{\alpha_1} \Gamma, \neg A.$$

Let A_{Γ} be the formula which results from A by replacing all free ordinal variables of A which do not occur in Γ by the ordinal constant 0. Then we also have

$$(2) \quad \text{H}[S, n+1, \beta+\omega] \models^{\alpha_0} \Gamma, A_{\Gamma} \quad \text{and} \quad \text{H}[S, n+1, \beta+\omega] \models^{\alpha_1} \Gamma, \neg A_{\Gamma}.$$

By the induction hypothesis we can conclude that

$$(3) \quad \text{H}[S, n+1, \beta+l] \models^{\varphi(n+2)0\alpha_0} \Gamma, A_{\Gamma},$$

$$(4) \quad \text{H}[S, n+1, \beta+l] \models^{\varphi(n+2)0\alpha_1} \Gamma, \neg A_{\Gamma}.$$

Hence, by a cut we yield

$$(5) \quad \text{H}[S, n+1, \beta+l] \models^{\delta} \Gamma,$$

where δ denotes the ordinal $\max(\varphi(n+2)0\alpha_0, \varphi(n+2)0\alpha_1)+1$. If $l = k$ then we are done; therefore, let us assume that $l = l' + 1 > k$. In order to get rid of the $(n+1)$ -inaccessible $\text{c}[S, n+1, l']$ one uses standard partial cut elimination and asymmetric interpretation in order to show

$$(6) \quad \text{H}[S', n, \delta^+] \models^{\delta^+} \Gamma,$$

where S' denotes the set of axioms of the system $\mathbf{H}[S, n+1, \beta+l']$ and δ^+ is the least ε number greater than δ . We know by induction hypothesis that the claim of our theorem is true for n and, hence, we can conclude from (6)

$$(7) \quad \mathbf{H}[S', n, 0] \mid_{\varphi(n+1)\delta^+\delta^+} \Gamma$$

since $\Gamma \subset \mathcal{L}[S', n, 0]$ by hypothesis. But in fact the system $\mathbf{H}[S', n, 0]$ is just $\mathbf{H}[S, n+1, \beta+l']$ and, moreover, we have that $\varphi(n+1)\delta^+\delta^+ < \varphi(n+2)0\alpha$. Thus (7) immediately reveals

$$(8) \quad \mathbf{H}[S, n+1, \beta+l'] \mid_{\leq \varphi(n+2)0\alpha} \Gamma.$$

Repeating this whole step $l - k$ times enables us to get rid of finitely many $(n + 1)$ -inaccessibles and we finally obtain

$$(9) \quad \mathbf{H}[S, n+1, \beta+k] \mid_{\varphi(n+2)0\alpha} \Gamma.$$

This concludes our proof in the case $\rho = 0$.

(b) $\rho = \rho_0 + 1$. Let $\gamma < \omega^{1+\rho_0+1}$ and Γ be a finite and quasi closed set of $\mathcal{L}[S, n+1, \beta+\gamma]$ formulas so that

$$(10) \quad \mathbf{H}[S, n+1, \beta+\omega^{1+\rho_0+1}] \mid_{\alpha} \Gamma.$$

Note that $\gamma = \omega^{1+\rho_0} \cdot k + \gamma'$ for some natural number k and some γ' less than $\omega^{1+\rho_0}$. Again the only crucial case occurs if Γ is the conclusion of a cut. Then there exist a natural number $l > k$, $\alpha_0, \alpha_1 < \alpha$ and an $\mathcal{L}[S, n+1, \beta+\omega^{1+\rho_0} \cdot l]$ formula A so that

$$(11) \quad \mathbf{H}[S, n+1, \beta+\omega^{1+\rho_0+1}] \mid_{\alpha_0} \Gamma, A_{\Gamma},$$

$$(12) \quad \mathbf{H}[S, n+1, \beta+\omega^{1+\rho_0+1}] \mid_{\alpha_1} \Gamma, \neg A_{\Gamma},$$

where A_{Γ} is defined as before. By applying the side induction hypothesis to (11) and (12) we derive

$$(13) \quad \mathbf{H}[S, n+1, \beta+\omega^{1+\rho_0} \cdot l] \mid_{\varphi^1 \rho \alpha_0} \Gamma, A_{\Gamma},$$

$$(14) \quad \mathbf{H}[S, n+1, \beta+\omega^{1+\rho_0} \cdot l] \mid_{\varphi^1 \rho \alpha_1} \Gamma, \neg A_{\Gamma},$$

and, hence, we also have $\mathbf{H}[S, n+1, \beta+\omega^{1+\rho_0} \cdot l] \mid_{\delta} \Gamma$, for δ being the ordinal $\max(\varphi(n+2)\rho\alpha_0, \varphi(n+2)\rho\alpha_1) + 1$. If we inductively define a sequence of ordinals δ_i by $\delta_0 := \delta$ and $\delta_{i+1} := \varphi(n+2)\rho_0\delta_i$, then by applying the main induction hypothesis $l - k$ times one readily obtains:

$$(15) \quad \begin{aligned} & \mathbf{H}[S, n+1, \beta+\omega^{1+\rho_0} \cdot (l-1)] \mid_{\delta_1} \Gamma, \\ & \quad \vdots \\ & \mathbf{H}[S, n+1, \beta+\omega^{1+\rho_0} \cdot (k+1)] \mid_{\delta_{l-k-1}} \Gamma, \\ & \mathbf{H}[S, n+1, \beta+\omega^{1+\rho_0} \cdot k + \gamma'] \mid_{\delta_{l-k}} \Gamma. \end{aligned}$$

Here we have successively replaced β by

$$\beta + \omega^{1+\rho_0} \cdot (l-1), \quad \dots, \quad \beta + \omega^{1+\rho_0} \cdot (k+1), \quad \beta + \omega^{1+\rho_0} \cdot k$$

in the main induction hypothesis. Since $\delta_{l-k} < \varphi(n+2)\rho\alpha$, we have indeed established

$$(16) \quad \mathbf{H}[S, n+1, \beta+\gamma] \mid_{\varphi(n+2)\rho\alpha} \Gamma$$

as desired. This finishes the treatment of the successor case.

(c) ρ is limit. Assume that $\gamma < \omega^{1+\rho}$ and Γ is a finite and quasi closed set of $\mathcal{L}[S, n+1, \beta+\gamma]$ formulas so that

$$(17) \quad \mathbf{H}[S, n+1, \beta+\omega^\rho] \mid_{\alpha} \Gamma.$$

Again assume that Γ is the conclusion of cut. Then there exists $\rho_0 < \rho$ with $\gamma \leq \omega^{1+\rho_0}$, $\alpha_0, \alpha_1 < \alpha$ and an $\mathcal{L}[S, n+1, \beta+\omega^{1+\rho_0}]$ formula A so that

$$(18) \quad \mathbf{H}[S, n+1, \beta+\omega^\rho] \mid_{\alpha_0} \Gamma, A_\Gamma \quad \text{and} \quad \mathbf{H}[S, n+1, \beta+\omega^\rho] \mid_{\alpha_1} \Gamma, \neg A_\Gamma.$$

The side induction hypothesis applied to (18) produces

$$(19) \quad \mathbf{H}[S, n+1, \beta+\omega^{1+\rho_0}] \mid_{\varphi(n+2)\rho\alpha_0} \Gamma, A_\Gamma,$$

$$(20) \quad \mathbf{H}[S, n+1, \beta+\omega^{1+\rho_0}] \mid_{\varphi(n+2)\rho\alpha_1} \Gamma, \neg A_\Gamma,$$

and, hence, we also have $\mathbf{H}[S, n+1, \beta+\omega^{1+\rho_0}] \mid_{\delta} \Gamma$, for δ being the ordinal $\max(\varphi(n+2)\rho\alpha_0, \varphi(n+2)\rho\alpha_1)+1$. From this, we conclude by the main induction hypothesis

$$(21) \quad \mathbf{H}[S, n+1, \beta+\gamma] \mid_{\varphi^1\rho_0\delta} \Gamma.$$

Since $\varphi^1\rho_0\delta < \varphi^1\rho\alpha$, this is our claim. This finishes the proof of (c) and also the verification of the theorem. \square

5.3 Bounds for the arithmetic fragment of OMA

In this subsection we focus on the arithmetic part of OMA, and it only remains to piece together our previous results in order to establish the appropriate upper bounds. In the following we simply write $\mathbf{H}[n, \alpha]$ instead of $\mathbf{H}[S, n, \alpha]$ if S is the empty set. Similarly $\mathbf{c}[n, \alpha]$ stands for $\mathbf{c}[S, n, \alpha]$ with empty S .

Theorem 11 *Let A be a closed \mathcal{L}_1 formula and assume $\text{OMA} \vdash A$. Then we also have that $\mathbf{H} \frac{\leq \varphi\omega 00}{0} A$.*

PROOF Assume that the \mathcal{L}_1 sentence A is provable in OMA . By Corollary 8 there exists a natural number n so $\text{OMA}^\top \frac{n}{\star} A$. This enables us to invoke the reduction theorem for OMA^\top , i.e. Theorem 9, in order to derive

$$(1) \quad \mathbf{H} \frac{\leq \omega^2}{} \neg \text{la}_n(\sigma), A.$$

A substitution of the ordinal constant $\text{c}[n, 0]$ for the ordinal variable σ followed by a cut on the formula $\text{la}_n(\text{c}[n, 0])$ reveals

$$(2) \quad \mathbf{H}[n, \omega] \frac{\leq \omega^2}{} A.$$

Now we apply Theorem 10 with S the empty set and $\beta = \gamma = \rho = 0$ and obtain

$$(3) \quad \mathbf{H}[n, 0] \frac{\leq \varphi(n+1)0\omega^2}{} A.$$

According to our definition, the theory $\mathbf{H}[n, 0]$ does not contain constants for n -inaccessibles and therefore is identical to \mathbf{H} . Moreover, a standard predicative cut elimination for \mathbf{H} finally yields

$$(4) \quad \mathbf{H} \frac{\leq \varphi(n+1)0\omega^2}{0} A.$$

Since obviously $\varphi(n+1)0\omega^2 < \varphi\omega 00$ this is as desired and completes our argument. \square

We end this section with some remarks concerning interesting proper subsystems of OMA . For this purpose we introduce the limit axiom

$$(n\text{-Lim}) \quad (\forall \xi)(\exists \eta)(\xi < \eta \wedge \text{la}_n(\eta))$$

for n -inaccessible ordinals. $(n+1)\text{-INAC}$ is defined to be the theory which results from OMA if we replace Π_2° reflection on Ad by the axiom $(n\text{-Lim})$. Hence, $(n+1)\text{-INAC}$ formalizes an $(n+1)$ -inaccessible universe of ordinals. However, observe that $(n+1)\text{-INAC}$, as OMA , does not include induction on the ordinals and that induction on the natural numbers is restricted to Δ_0° formulas. We mention that theories of explicit mathematics and admissible set theory corresponding to $(n+1)\text{-INAC}$ can be easily defined.

A standard reduction shows that the strength of $(n+1)\text{-INAC}$ is already exhausted by finite chains of n -inaccessible ordinals. Further, by gradually unfolding the topmost n -inaccessible into transfinitely many $n-1$ -inaccessibles and applying Theorem 10 one obtains the following theorem.

Theorem 12 *Let A be a closed \mathcal{L}_1 formula and assume $(n+1)\text{-INAC} \vdash A$. Then we also have that $\mathbf{H} \frac{\leq \varphi(n+1)00}{0} A$.*

6 Concluding remarks

Finally we turn to the proof-theoretic ordinals of KPm^0 , EMA , OMA and the other systems mentioned in this paper. In principle, all work is already done in Theorem 6, Theorem 11 and Theorem 12.

The formula $\text{TI}(\prec, A)$ has been defined in the context of the language \mathcal{L}^* . From now on we take the liberty to also write $\text{TI}(\prec, A)$ for the corresponding formulas in the languages \mathbb{L} and \mathcal{L}_0 . It only remains to apply one of the usual boundedness theorems (cf. e.g. Beckmann [1] or Schütte [17]) stating that

$$\text{H} \frac{\alpha}{0} \text{TI}(\prec, \mathbf{Q}) \implies |\prec| \leq \omega \cdot \alpha,$$

for all α and all primitive recursive wellorderings \prec ; here $|\prec|$ is the ordertype of \prec as usual. Together with the lower bound computations of [20] we thus have the following main theorem of this article.

Theorem 13 *We have the following proof-theoretic ordinals:*

1. $|\text{KPm}^0| = |\text{EMA}| = |\text{OMA}| = \varphi\omega 00$;
2. $|(n+1)\text{-INAC}| = \varphi(n+1)00$.

In the theories mentioned in the previous theorem, complete induction on the natural numbers is restricted to sets, types and Δ_0^0 formulas, respectively. The methods applied before also provide bounds for theories with complete induction on the natural numbers for arbitrary formulas.

The pattern of the argument for $\text{EMA} + (\mathbb{L}\text{-I}_{\mathbb{N}})$ and $\text{OMA} + (\mathcal{L}_0\text{-I}_{\mathbb{N}})$ is as follows; $\text{KPm}^0 + (\mathcal{L}^*\text{-I}_{\mathbb{N}})$ is treated accordingly: $\text{EMA} + (\mathbb{L}\text{-I}_{\mathbb{N}})$ is interpreted into $\text{OMA} + (\mathcal{L}_0\text{-I}_{\mathbb{N}})$ following the proof of Theorem 6. Then $\text{OMA} + (\mathcal{L}_0\text{-I}_{\mathbb{N}})$ is embedded into OMA^{T} plus ω rule, thus getting rid of full complete induction in favor of infinite derivation lengths. Weak cut elimination for OMA^{T} plus ω rule is proved as before, but because of the infinite derivations we now have

$$\text{OMA} + (\mathcal{L}_0\text{-I}_{\mathbb{N}}) \vdash A \implies \text{OMA}^{\text{T}} + (\omega) \stackrel{\leq \varepsilon_0}{\star} A$$

for each \mathcal{L}_0 sentence A . From now on we can proceed as before, but always with families $(\text{Ia}_{\alpha}(\sigma) : \alpha < \varepsilon_0)$ instead of families $(\text{Ia}_n(\sigma) : n < \omega)$. Carrying through everything in detail finally gives the following results for theories with full induction.

Theorem 14 *We have the following proof-theoretic ordinals:*

1. $|\text{KPm}^0 + (\mathcal{L}^*\text{-I}_{\mathbb{N}})| = |\text{EMA} + (\mathbb{L}\text{-I}_{\mathbb{N}})| = |\text{OMA} + (\mathcal{L}_0\text{-I}_{\mathbb{N}})| = \varphi\varepsilon_0 00$;
2. $|(n+1)\text{-INAC} + (\mathcal{L}_0\text{-I}_{\mathbb{N}})| = \varphi(n+1)\varepsilon_0 0$.

References

- [1] BECKMANN, A. *Separating fragments of bounded arithmetic*. PhD thesis, Universität Münster, 1996.
- [2] BEESON, M. J. *Foundations of Constructive Mathematics: Metamathematical Studies*. Springer, Berlin, 1985.
- [3] FEFERMAN, S. Constructive theories of functions and classes. In *Logic Colloquium '78*, M. Boffa, D. van Dalen, and K. McAloon, Eds. North Holland, Amsterdam, 1979, pp. 159–224.
- [4] FEFERMAN, S., AND JÄGER, G. Systems of explicit mathematics with non-constructive μ -operator. Part II. *Annals of Pure and Applied Logic* 79, 1 (1996), 37–52.
- [5] JÄGER, G. *Die konstruktible Hierarchie als Hilfsmittel zur beweistheoretischen Untersuchung von Teilsystemen der Mengenlehre und Analysis*. PhD thesis, Universität München, 1979.
- [6] JÄGER, G. A well-ordering proof for Feferman's theory T_0 . *Archiv für mathematische Logik und Grundlagenforschung* 23 (1983), 65–77.
- [7] JÄGER, G. The strength of admissibility without foundation. *The Journal of Symbolic Logic* 49, 3 (1984), 867–879.
- [8] JÄGER, G., KAHLE, R., SETZER, A., AND STRAHM, T. The proof-theoretic analysis of transfinitely iterated fixed point theories. *Journal of Symbolic Logic* 64, 1 (1999), 53–67.
- [9] JÄGER, G., KAHLE, R., AND STUDER, T. Universes in explicit mathematics. Preprint.
- [10] JÄGER, G., AND POHLERS, W. Eine beweistheoretische Untersuchung von $(\Delta_2^1\text{-CA}) + (\text{BI})$ und verwandter Systeme. In *Sitzungsberichte der Bayerischen Akademie der Wissenschaften, Mathematisch-naturwissenschaftliche Klasse*. 1982, pp. 1–28.
- [11] JÄGER, G., AND STRAHM, T. Fixed point theories and dependent choice. *Archive for Mathematical Logic*. To appear.
- [12] JÄGER, G., AND STUDER, T. Extending the system T_0 of explicit mathematics: the limit and Mahlo axioms. In preparation.

- [13] POHLERS, W. *Proof Theory: An Introduction*, vol. 1407 of *Lecture Notes in Mathematics*. Springer, Berlin, 1989.
- [14] RATHJEN, M. Proof-theoretic analysis of KPM. *Archive for Mathematical Logic* 30 (1991), 377–403.
- [15] RATHJEN, M. Collapsing functions based on recursively large ordinals: A wellordering proof for KPM. *Archive for Mathematical Logic* 33 (1994).
- [16] SCHÜTTE, K. Kennzeichnung von Ordnungszahlen durch rekursiv erklärte Funktionen. *Mathematische Annalen* 127 (1954), 15–32.
- [17] SCHÜTTE, K. *Proof Theory*. Springer, Berlin, 1977.
- [18] SETZER, A. Extending Martin-Löf type theory by one Mahlo universe. *Archive for Mathematical Logic*. To appear.
- [19] STRAHM, T. Autonomous fixed point progressions and fixed point transfinite recursion. In *Logic Colloquium '98*, S. Buss, P. Hájek, and P. Pudlák, Eds., *Lecture Notes in Logic*. Springer. To appear.
- [20] STRAHM, T. Wellordering proofs for metapredicative Mahlo. In preparation.
- [21] STRAHM, T. First steps into metapredicativity in explicit mathematics. In *Sets and Proofs*, S. B. Cooper and J. Truss, Eds. Cambridge University Press, 1999, pp. 383–402.
- [22] TROELSTRA, A., AND VAN DALEN, D. *Constructivism in Mathematics*, vol. I. North-Holland, Amsterdam, 1988.

Address

Gerhard Jäger, Thomas Strahm
 Institut für Informatik und angewandte Mathematik, Universität Bern
 Neubrückstrasse 10, CH-3012 Bern, Switzerland
 {jaeger, strahm}@iam.unibe.ch